



**Kauno technologijos universitetas**  
Matematikos ir gamtos mokslų fakultetas

# **Europos Sąjungos šalių konkurencingumo vertinimas ir klasterizavimas**

Baigiamasis magistro studijų projektas

---

**Ieva Vytuvienė**  
Projekto autorė

Doc. dr. Paulius Palevičius

Vadovas

Prof. dr. Jurgita Bruneckienė

Vadovė

**Kaunas, 2021**



**Kauno technologijos universitetas**  
Matematikos ir gamtos mokslų fakultetas

# **Europos Sąjungos šalių konkurencingumo vertinimas ir klasterizavimas**

Baigiamasis magistro studijų projektas  
Didžiųjų verslo duomenų analitika (6213AX001)

---

**Ieva Vytuvienė**  
Projekto autorė

**Doc. dr. Paulius Palevičius**  
Vadovas

**Prof. dr. Jurgita Bruneckienė**  
Vadovė

**Doc. dr. Mindaugas Šnipas**  
Recenzentas

**Doc. dr. Akvilė Čibinskienė**  
Recenzentė

**Kaunas, 2021**



**Kauno technologijos universitetas**

Matematikos ir gamtos mokslų fakultetas

Ieva Vytuvienė

## **Europos Sąjungos šalių konkurencingumo vertinimas ir klasterizavimas**

Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdama(s) kitų asmenų autoriaus ar kitų teisių, laikydamasi(s) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs (-usi);
4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalinta(s) iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Ieva Vytuvienė

*Patvirtinta elektroniniu būdu*

Vytuvienė Ieva. Europos Sąjungos šalių konkurencingumo vertinimas ir klasterizavimas. Magistro krypties studijų baigiamasis projektas / vadovai prof. dr. Jurgita Bruneckienė ir doc. dr. Paulius Palevičius; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų kryptių grupė): Taikomoji matematika.

Reikšminiai žodžiai: konkurencingumas, klasterinė analizė, MissForest, VSURF.

Kaunas, 2021. 117 p.

### Santrauka

Baigiamajame magistro projekte siekiama suklasterizuoti Europos Sąjungos šalis ir įvertinti jų konkurencingumą. Konkurencingumo vertinimas yra svarbus siekiant nustatyti skirtumus tarp šalių narių bei siekiant kryptingos Europos Sąjungos plėtros. Atlikus mokslinės literatūros analizę buvo sudarytas šalių konkurencingumą lemiančių veiksnių modelis, susidedantis iš 4 dalių – ekonominio, socialinio, šalies aplinkos bei aplinkosauginio konkurencingumo. Naudojant viešai prieinamas duomenų bazines buvo surinkti rodikliai, atspindintys sudarytą modelį. Sudarytas šalių konkurencingumą lemiančių veiksnių modelis gali būti panaudojamas atliekant ir kitus tyrimus, skirtus kitų šalių konkurencingumo vertinimui. Trūkstatų reikšmių užpildymui buvo pritaikytas *missForest* algoritmas, kuris leido išsaugoti vertingus duomenis ir panaudoti juos tolimesniame tyrime. Pagal šalių konkurencingumą lemiančių veiksnių modelį atrinkti rodikliai leido suklasterizuoti šalis pagal jų panašumus. Tokiu būdu buvo identifikuoti konkurencingiausi ir mažiausiai konkurencingi klasteriai bei išryškinti skirtumai tarp klasterių. Daugeliu atvejų klasterizavimo rezultatams įtakos turėjo geografinė šalies vieta, nors tiesiogiai į tyrimą ji nebuvo įtraukta. Įvertinus Europos Sąjungos šalių 2008–2019 m. rangus, gautus pagal apskaičiuotą bendrą konkurencingumo indeksą, paaiškėjo, jog didžiausią konkurencinį pranašumą analizuojamu laikotarpiu įgijo Švedija, Liuksemburgas, Nyderlandai, Danija ir Suomija. Mažiausias konkurencinis pranašumas analizuojamu laikotarpiu buvo šių šalių – Rumunijos, Bulgarijos, Lenkijos, Graikijos. Siekiant padidinti Europos Sąjungos, kaip vieno teritorinio vieneto, konkurencingumą pasaulyje, turėtų būti mažinami skirtumai tarp šalių narių, ypač atsižvelgiant į žemiausiose pozicijose esančias šalis. Padarius prielaidą, jog konkurencingumas gali būti vertinamas vienu makroekonominio rodikliu – bendruoju vidaus produktu, tenkančiu vienam gyventojui – buvo atrinkti svarbiausi kintamieji, turintys reikšmingos įtakos jam. Svarbiausių kintamųjų atrankai buvo panaudotas mašininio mokymosi algoritmas *VSURF*. Iš viso buvo atrinkta 11 rodiklių, iš kurių daugiausia – ekonominių. Palyginus 2008–2019 m. Europos Sąjungos šalių rangus buvo pastebėta, jog didžiausias rangų suderinamumas, vertinant pagal Kendall'o konkordancijos koeficientą, buvo gautas lyginant bendro konkurencingumo indekso (naudojant visus rodiklius) rangą su konkurencingumo indekso (naudojant *VSURF* atrinktus rodiklius) rangą ir šis rodiklis analizuojamu laikotarpiu buvo didesnis už 0,96. Europos Sąjungos šalių konkurencingumui vertinti galėtų būti naudojamas ir mažesnis skaičius kintamųjų, o bendrasis vidaus produktas buvo svarbus rodiklis identifikuojant juos.

Vytuvienė, Ieva. Competitiveness assessment and clustering of European Union countries. Master's Final Degree Project / supervisors assoc. prof. Paulius Palevičius and prof. Jurgita Bruneckienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Applied Mathematics.

Keywords: competitiveness, cluster analysis, MissForest, VSURF.

Kaunas, 2021. 117.

### Summary

The aim of the master's project is to cluster the countries of the European Union and to assess their competitiveness. Competitiveness assessment is important for identifying disparities between Member States and for seeking the targeted enlargement of the European Union. After the analysis of the scientific literature, the model of the factors determining the competitiveness of the countries was created, consisting of 4 parts - economic, social, country's environment and environmental competitiveness. Using publicly available databases, indicators, which reflected the developed model, were collected. The model of the factors determining the competitiveness of the countries can also be used in other researches. The *MissForest* algorithm was applied in order to fill in the missing values, which allowed to save valuable data and to use it in further research. The indicators selected according to the model of the factors determining the competitiveness of the countries allowed to cluster countries according to their similarities. In this way, the most competitive and least competitive clusters were identified and differences between clusters were highlighted. In most cases, the clustering results were influenced by the geographical location of the country, although it was not directly included in the study. According to the rankings obtained from the calculated competitiveness index showed that Sweden, Luxembourg, the Netherlands, Denmark and Finland gained the greatest competitive advantage during the period under analysis. The lowest competitive advantage during the analyzed period was in Romania, Bulgaria, Poland, Greece. In order to increase the European Union's as a single territorial unit's global competitiveness, disparities between Member States should be reduced, especially with regard to the lowest-ranking countries. Assuming that competitiveness can be measured by one macroeconomic indicator – gross domestic product per capita, the most important variables that have a significant impact on it have been identified. The machine learning algorithm *VSURF* was used in order to select the most important variables. 11 indicators were selected, most of them were economic. Comparing the ranks of the European Union countries in the analyzed period, it was noticed that the highest agreement of ranks, measured by the Kendall's coefficient of concordance, was obtained by comparing the ranking of the competitiveness index (using all indicators) with the competitiveness index (using *VSURF* selected indicators) and this measure was above 0.96. A smaller set of variables could also be used to assess the competitiveness of European Union countries, and gross domestic product has been an important indicator in identifying them.

## Turinys

<b>Lentelių sąrašas .....</b>	<b>8</b>
<b>Paveikslų sąrašas .....</b>	<b>9</b>
<b>Santrumpų sąrašas .....</b>	<b>10</b>
<b>Įvadas.....</b>	<b>11</b>
<b>1. Literatūros apžvalga .....</b>	<b>12</b>
1.1. Konkurencingumo samprata.....	12
1.2. Konkurencingumo vertinimo metodai.....	14
1.3. Teritorijos konkurencingumą lemiantys veiksniai .....	17
1.3.1. Ekonominiai veiksniai .....	17
1.3.2. Socialiniai veiksniai.....	19
1.3.3. Šalies aplinkos veiksniai.....	21
1.3.4. Aplinkosauginiai veiksniai .....	22
1.3.5. Šalių konkurencingumą lemiančių veiksnių modelis .....	23
1.4. Matematinė dalis .....	24
1.4.1. Trūkstumų reikšmių užpildymas .....	24
1.4.2. Klasterizavimas .....	25
1.4.3. Duomenų analizės programinė įranga .....	26
<b>2. Metodologija .....</b>	<b>28</b>
2.1. Programinė įranga .....	28
2.2. Duomenys.....	28
2.3. Trūkstumų reikšmių užpildymas .....	29
2.4. Klasterizavimas .....	31
2.4.1. K-vidurkių algoritmas .....	32
2.4.2. Hierarchinis klasterizavimas .....	34
2.4.3. Klasterių skaičiaus nustatymas ir klasterizavimo kokybės vertinimas.....	36
2.5. Kintamųjų atranka, algoritmas VSURF .....	40
2.6. Kendall'o konkordancijos koeficientas .....	41
2.7. Indeksų skaičiavimas.....	42
<b>3. Rezultatai.....</b>	<b>44</b>
3.1. Trūkstamos reikšmės .....	44
3.1.1. Trūkstumų reikšmių apžvalga.....	44
3.1.2. Trūkstumų reikšmių užpildymas .....	46
3.2. Europos Sąjungos šalių klasterizavimas.....	47
3.2.1. 2008 m. klasterizavimo rezultatai.....	48
3.2.2. 2013 m. klasterizavimo rezultatai.....	51
3.2.3. 2019 m. klasterizavimo rezultatai.....	55
3.2.4. Klasterizavimo rezultatų apibendrinimas .....	58
3.3. Europos Sąjungos šalių konkurencingumo vertinimas.....	59
3.3.1. Vertinimas pagal apskaičiuotus konkurencingumo indeksus.....	59
3.3.2. Kintamųjų atranka algoritmu VSURF .....	64
3.3.3. ES šalių rangų, gautų skirtingais metodais, suderinamumas.....	65
<b>Išvados .....</b>	<b>68</b>

<b>Rekomendacijos</b> .....	<b>69</b>
<b>Literatūros sąrašas</b> .....	<b>70</b>
<b>Informacijos šaltinių sąrašas</b> .....	<b>76</b>
<b>Priedai</b> .....	<b>77</b>
1 priedas. Konkurencingumą lemiančios veiksnių grupės ir rodikliai .....	77
2 priedas. Trūkstamos reikšmės pagal požymius ir šalis.....	80
3 priedas. Požymių, kurie turėjo trūkstamų reikšmių, skaitinės charakteristikos prieš ir po užpildymo .....	94
4 priedas. Klasteriuose esančių šalių požymių vidutinės reikšmės 2008 m. ....	95
5 priedas. Klasteriuose esančių šalių požymių vidutinės reikšmės 2013 m. ....	97
6 priedas. Klasteriuose esančių šalių požymių vidutinės reikšmės 2019 m. ....	99
7 priedas. Pirmosios ir antrosios pagrindinės komponentės svoriai 2008 m. ....	101
8 priedas. Pirmosios, antrosios, trečiosios pagrindinės komponentės svoriai 2013 m. ....	103
9 priedas. Pirmosios, antrosios, trečiosios pagrindinės komponentės svoriai 2019 m. ....	106
10 priedas. ES šalių ekonominis konkurencingumo indeksas 2008–2019 m. ....	109
11 priedas. ES šalių socialinis konkurencingumo indeksas 2008–2019 m. ....	110
12 priedas. ES šalių aplinkosauginis konkurencingumo indeksas 2008–2019 m. ....	111
13 priedas. ES šalių šalies aplinkos konkurencingumo indeksas 2008–2019 m. ....	112
14 priedas. ES šalių bendrasis konkurencingumo indeksas 2008–2019 m. ....	113
15 priedas. ES šalių rangai pagal tris rodiklius 2008–2019 m. ....	114
16 priedas. ES šalių konkurencingumo ranginio vertinimo sutapimo lygiai.....	117

## Lentelių sąrašas

<b>1 lentelė.</b> Regiono konkurencingumo sąvoka pagal skirtingus autorius .....	13
<b>2 lentelė.</b> Regiono konkurencingumą lemiantys veiksniai, ekonominė veiksnių grupė .....	17
<b>3 lentelė.</b> Regiono konkurencingumą lemiantys veiksniai, socialinių veiksnių grupė .....	20
<b>4 lentelė.</b> Regiono konkurencingumą lemiantys veiksniai, šalies aplinkos veiksnių grupė.....	21
<b>5 lentelė.</b> Regiono konkurencingumą lemiantys veiksniai, aplinkosauginių veiksnių grupė .....	22
<b>6 lentelė.</b> Klasterizavimo metodai pagal skirtingus autorius .....	25
<b>7 lentelė.</b> <i>MissForest</i> algoritmas, [61 p. 133] .....	31
<b>8 lentelė.</b> Cronbach alpha reikšmės interpretavimui, [88 p. 60].....	43
<b>9 lentelė.</b> Cronbach alpha reikšmės atitinkamoms rodiklių grupėms 2008–2019 m. ....	59
<b>10 lentelė.</b> Cronbach alpha reikšmės 4-iems konkurencingumo indeksams 2008–2019 m.....	60



## Paveikslų sąrašas

<b>1 pav.</b> Konkurencingumo vertinimo objektai, [3 p. 346] .....	12
<b>2 pav.</b> PKI 4.0 struktūra, [5 p. 2] .....	15
<b>3 pav.</b> ES 2019 m. PKI 4.0 balai, sudaryta autorės, remiantis šaltiniu [5].....	15
<b>4 pav.</b> Šalių konkurencingumą lemiančių veiksnių modelis .....	23
<b>5 pav.</b> Struktūrinė k-vidurkių algoritmo diagrama, [67] .....	33
<b>6 pav.</b> Jungimo metodai: a) vienetinės jungties, b) pilnosios jungties, c) vidutinės jungties, d) centroidų, [80 p. 5596]. .....	35
<b>7 pav.</b> Kvadratinių nuokrypių sumos, [66 p. 243] .....	36
<b>8 pav.</b> Klasterių skaičiaus parinkimas pagal $R^2$ , [66 p. 245] .....	37
<b>9 pav.</b> Duomenų rinkinio trūkstamos reikšmės pagal metus, proc. ....	44
<b>10 pav.</b> Duomenų rinkinio trūkstamos reikšmės pagal ES šalis, proc. ....	45
<b>11 pav.</b> Trūkstamos reikšmės pagal požymius, proc.....	45
<b>12 pav.</b> OOB klaidos prie skirtingų <i>mtry</i> reikšmių .....	46
<b>13 pav.</b> Pasirinktų kintamųjų esamos (žalia) ir užpildytos (raudona) reikšmės, viršuje kairėje Lietuvos, viršuje dešinėje Danijos, apačioje kairėje Kroatijos, apačioje dešinėje Švedijos .....	47
<b>14 pav.</b> Alkūnės metodas klasterių skaičiaus parinkimui .....	47
<b>15 pav.</b> $R^2$ prie skirtingo klasterių skaičiaus .....	47
<b>16 pav.</b> Silueto koeficientas prie skirtingo klasterių skaičiaus .....	48
<b>17 pav.</b> K-vidurkių 2008 m. klasterizavimo rezultatų vizualizavimas naudojant dvi pirmas pagrindines komponentes .....	49
<b>18 pav.</b> Hierarchinio klasterizavimo dendograma pagal 2008 m. duomenis .....	51
<b>19 pav.</b> K-vidurkių 2013 m. klasterizavimo rezultatų vizualizavimas naudojant dvi pirmas pagrindines komponentes .....	52
<b>20 pav.</b> K-vidurkių 2013 m. klasterizavimo rezultatai atvaizdavius pagal tris pirmas pagrindines komponentes.....	52
<b>21 pav.</b> Hierarchinio klasterizavimo dendograma pagal 2013 m. duomenis .....	55
<b>22 pav.</b> K-vidurkių klasterizavimo rezultatai pagal 2019 m. duomenis naudojant dvi pirmas pagrindines komponentes .....	56
<b>23 pav.</b> K-vidurkių klasterizavimo rezultatai 2019 m. atvaizdavius pagal tris pirmas pagrindines komponentes.....	56
<b>24 pav.</b> Hierarchinio klasterizavimo dendograma pagal 2019 m. duomenis .....	57
<b>25 pav.</b> 2019 m. k-vidurkių (kairėje) ir hierarchinio (dešinėje) klasterizavimo rezultatai .....	58
<b>26 pav.</b> 2008, 2013 ir 2019 m. k-vidurkių klasterizavimo rezultatai žemėlapyje.....	58
<b>27 pav.</b> ES šalių ekonominio konkurencingumo rangai 2008–2019 m. ....	61
<b>28 pav.</b> ES šalių socialinio konkurencingumo rangai 2008–2019 m. ....	61
<b>29 pav.</b> ES šalių aplinkosauginio konkurencingumo rangai 2008–2019 m. ....	62
<b>30 pav.</b> ES šalių šalies aplinkos konkurencingumo rangai 2008–2019 m. ....	63
<b>31 pav.</b> ES šalių bendrojo konkurencingumo rangai 2008–2019 m. ....	63
<b>32 pav.</b> OOB klaidos prie skirtingų <i>mtry</i> reikšmių .....	64
<b>33 pav.</b> VSURF trečiame etape atrinkti kintamieji pagal mažėjančią svarbą.....	65
<b>34 pav.</b> ES šalių rangų, gautų pagal 3 būdus, palyginimas 2019 m.....	66
<b>35 pav.</b> Kendall'o konkordancijos koeficientų kitimo dinamika 2008–2019 m.....	67

## Santrumpų sąrašas

### Santrumpos:

BVP – bendrasis vidaus produktas;

ES – Europos Sąjunga;

PKI – Pasaulinis konkurencingumo indeksas;

TUI – tiesioginės užsienio investicijos;

VI – kintamojo svarba (angl. *variable importance*).

## Įvadas

**Tyrimų aktualumas, naujumas.** Konkurencingumo vertinimas yra aktuali problematika tarp mokslininkų, politikų, verslininkų. Tai padeda nustatyti įmonės, šalies, regiono esamą poziciją, identifikuoti silpnybes, stiprybes ir nustatyti strateginius tikslus, siekti kryptingos plėtros. Europos Sąjungoje konkurencingumo vertinimas yra aktualus siekiant regioninės plėtros, mažinant skirtumus tarp šalių ir siekiant įgyti visos sąjungos, kaip vieno teritorinio vieneto, konkurencinį pranašumą pasaulyje. Šio darbo tyrimo naujumas yra tas, jog be įprastų ekonominių, socialinių, šalies aplinkos rodiklių, Europos Sąjungos šalių konkurencingumas bus vertinamas ir pagal aplinkosauginius rodiklius. Šiuo metu aplinkosaugos klausimas yra itin aktualus, nes Europos Sąjunga įgyvendina „Europos žaliąjį kursą“ ir iki 2050 m. siekia tapti neutralaus poveikio klimatui žemynu. Tyrimui atlikti bus panaudoti ne tik tradiciniai metodai kaip, pavyzdžiui, šalių klasterizavimas k-vidurkių algoritmu, tačiau apdorojant duomenis bei siekiant identifikuoti svarbiausius konkurencingumo rodiklius, bus panaudoti ir mašininio mokymosi algoritmai *missForest* ir *VSURF*.

**Tyrimo problema.** Vertinant konkurencingumą mokslininkai itin daug dėmesio skiria ekonominiams, socialiniams arba aplinkosauginiams veiksniams, tačiau pasigendama metodikos, kuri konkurencingumą leistų vertinti apjungiant minėtus veiksniai. Taip pat analizuojant mokslinius straipsnius buvo pastebėta, jog mokslininkai, atlikdami konkurencingumo vertinimo tyrimus, atsisako duomenų, kuriuose yra trūkstamų reikšmių ir dažnai naudoja tiesinės regresijos modelius, koreliacinę analizę, siekdami išsiaiškinti įtakingiausius kintamuosius. Šiame darbe bus stengiamasi išsaugoti vertingus duomenis, kuriuose yra trūkstamų reikšmių jas užpildant naudojant *missForest* algoritmą, o reikšmingų kintamųjų atrankai bus panaudotas mašininio mokymosi algoritmas *VSURF*, kuris geba atpažinti ir netiesinius ryšius duomenyse.

**Tyrimo tikslas** – sudarius šalių konkurencingumą lemiančių veiksnių modelį, pasitelkus statistinius bei mašininio mokymosi metodus ir panaudojus viešai prieinamus duomenis, įvertinti Europos Sąjungos šalių konkurencingumą.

### Tyrimo uždaviniai:

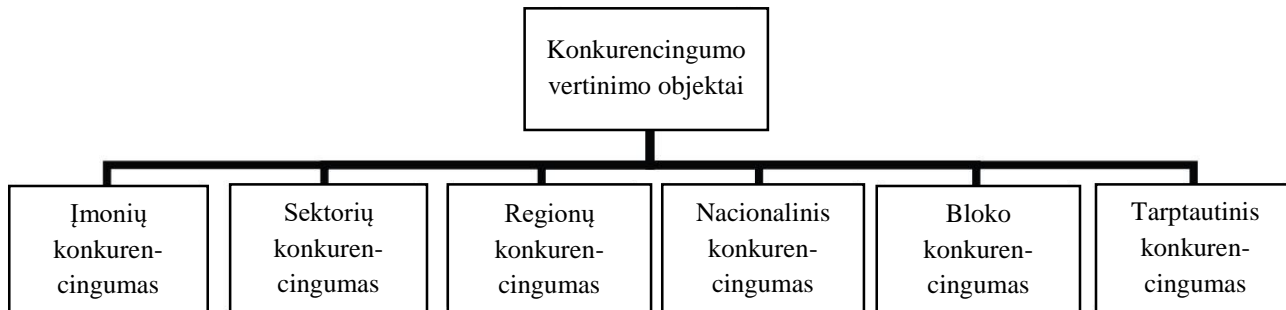
1. susisteminius naujausią mokslinės literatūros analizę, išskirti pagrindinius konkurencingumą lemiančius veiksniai ir sudaryti šalių konkurencingumą lemiančių veiksnių modelį;
2. įvertinti duomenų rinkinio trūkstamas reikšmes ir panaudojus mašininio mokymosi algoritmą užpildyti jas;
3. atlikti Europos Sąjungos šalių 2008, 2013 ir 2019 m. klasterinę analizę pagal šalių konkurencingumą lemiančių veiksnių modelį atspindinčius rodiklius;
4. sudaryti konkurencingumo indeksą ir įvertinti Europos Sąjungos šalių konkurencingumą;
5. atrinkti svarbiausius konkurencingumą lemiančius rodiklius pritaikius mašininio mokymosi algoritmą;
6. nustatyti Europos Sąjungos šalių konkurencingumo rangų, gautų pagal skirtingus būdus, suderinamumą 2008–2019 m.

## 1. Literatūros apžvalga

Šiame skyriuje bus apžvelgiama konkurencingumo samprata, akcentuojama, kodėl svarbu vertinti konkurencingumą. Taip pat bus pristatomi konkurencingumo vertinimo metodai. Vėliau bus išskiriami esminiai šalies konkurencingumą lemiantys veiksniai bei sudaromas šalių konkurencingumą lemiančių veiksnių modelis. Šiame skyriuje taip pat bus pateikiama bendra matematinių algoritmų, skirtų spręsti analizuojamai problemai, apžvalga. Bus trumpai pristatomi trūkstumų reikšmių užpildymo metodai, aptariami klasterizavimo metodai bei pristatoma duomenų analizės programinė įranga.

### 1.1. Konkurencingumo samprata

Konkurencingumo konceptas vis dar nėra iki galo apibrėžtas, skirtingi autoriai konkurencingumo sąvoką apibrėžia skirtingai. Egzistuoja kelios konkurencingumo taikymo sritys. Cellini ir Soci [1] išskiria makro-lygmens (šalies), mikro-lygmens (įmonės) ir meso-lygmens (vietinės ekonominės sistemos – pramonės rajonų ar regionų) konkurencingumą. Įmonės konkurencingumą galima vertinti pagal jos pelningumą, užimamos rinkos dalį ar akcijų kainą. Pramonės šakos konkurencingumas analizuojamas pagal bendrojo vidaus produkto (BVP) dalį ar eksporto rodiklius. Šalies konkurencingumo sąvoka vis dar ginčytina, skirtingi ekonomistai gali skirtingai ją suvokti [2]. Balkytė ir Tvaronavičienė [3] konkurencingumo vertinimo objektus suskirstė į 6-ias dalis: įmonių konkurencingumą, sektorių konkurencingumą, regionų konkurencingumą (srities, vietos, vietovės, teritorijos, miesto), nacionalinį konkurencingumą, bloko (pavyzdžiui, BRICS ar ES) konkurencingumą, tarptautinį konkurencingumą (žr. 1 pav.).



1 pav. Konkurencingumo vertinimo objektai, [3 p. 346]

Pastaruoju metu šalies konkurencingumas yra vienas iš labiausiai stebimų požymių, atspindintis jos klestėjimą, gerovę, gyvenimo lygį [4]. Pasaulio ekonomikos forumo (angl. *The World Economic Forum - WEF*) 2019 m. ataskaitoje šalies konkurencingumas apibrėžiamas kaip institucijų, strategijų ir faktorių rinkinys, kuris nusako šalies produktyvumą [5]. Didesnis konkurencingumas siejamas su didesniu šalies ekonomikos produktyvumu, o tai turėtų lemti didesnę ir pastovesnę ekonomikos augimą, kuris siejamas su pagerėjusiu šalies piliečių gyvenimo lygiu [2].

Europos Sąjungoje (ES) konkurencingumo klausimas įgijo ypatingą reikšmę, kai buvo priimta Lisabonos strategija, kurios ambicingas tikslas buvo panaikinti konkurencingumo atotrūkį tarp JAV ir tapti dinamiškiausia ir konkurencingiausia žiniomis paremta ekonomika iki 2010 m. [6]. Vertinant ES regionų konkurencingumą, svarbu apibrėžti regiono ir regiono konkurencingumo sąvokas. Simanavičienė ir kt. [7 p. 69] regioną apibrėžia kaip „teisiškai nustatytą teritorijos vienetą, turintį savivaldą ir autonomiją bei išsiskiriantį iš gretimų teritorijų pagal tam tikrus rodiklius

(ekonominius, socialinius, demografinius, kultūros-istorijos, gamtinius ir politinius)“. Šiame darbe regionas bus suprantamas kaip teritorinė ES dalis, atitinkanti ES šalį. Nepaisant to, kad konkurencingumo sąvoka plačiai vartojama moksliniuose tyrimuose, vis dar nėra ne tik vienintelio konkurencingumo, bet ir regionų konkurencingumo apibrėžimo [8]. Autoriai [6], [1], [9], [10] pripažįsta, kad sunku apibrėžti regionų konkurencingumo sąvoką. Toliau pateikiama lentelė, kaip skirtingi autoriai nusako regiono konkurencingumo sąvoką (žr. 1 lent.).

**1 lentelė.** Regiono konkurencingumo sąvoka pagal skirtingus autorius

<b>Regiono konkurencingumo sąvoka</b>	<b>Autoriai</b>
Regiono konkurencingumas – tai regiono gebėjimas sudaryti patrauklią ir tvarią aplinką įmonėms ir žmonėms gyventi ir dirbti tiek trumpuoju, tiek ilguoju laikotarpiu.	Dijkstra ir kt. [11]
Regiono konkurencingumas – tai požymiai, apibūdinantys visas vieno regiono įmones, įskaitant institucijas, infrastruktūrą, švietimą, darbo jėgos įgūdžius, technologijas, inovacijas, kitaip tariant, viską, kas gali padėti įmonei veikti jos plėtrai palankioje verslo aplinkoje.	Clipa ir Ifrim [10]
Regiono konkurencingumas – tai regiono gebėjimas vidaus ir tarptautinės konkurencijos sąlygomis sukurti patrauklią verslo aplinką naujoms investicijoms ir verslo plėtrai ir, tuo remiantis, gyventojams užtikrinti santykinai dideles pajamas, užimtumą ir visą apimančią gerovę.	Nikolic ir kt. [12]
Regiono konkurencingumas yra gebėjimas būti produktyviu ir išlaikyti aukštą pragyvenimo lygį.	Vukovic ir kt. [13]
Konkurencingumas vertinamas kaip regiono gebėjimas sudaryti palankesnes sąlygas ekonominei veiklai ir gyvenimui nei konkurentai ir deklaruojamas kaip plėtros tikslas, turintis tikimybinį pobūdį.	Žitkus [14]
Regiono konkurencingumas yra labai sudėtingas klausimas, kurio negalima vertinti tik mikro- ar makro-požiūriu, nes jis turi apimti vietinius ir kitus specifinius veiksnius (socialinius, kultūrinius, „minkštus veiksnius“), kurie daro regioną konkurencingą.	Vukovic ir Wei [9]
„ [...] regiono konkurencingumas – tai jo (regionas) gebėjimas geriau nei kiti regionai prisitaikyti prie kintančių ekonominių, socialinių ir politinių sąlygų“.	Žitkus ir Mickevičienė [15 , p. 436]
Regiono konkurencingumas gali būti apibrėžtas kaip sėkmė konkuruojant vienas su kitu kažkokių būdu. Pavyzdžiui, tai gali būti matuojama eksporto rinkos dalimi ar kapitalo, darbuotojų pritraukimu.	Kitson ir kt. [6]

Svarbu akcentuoti, kodėl yra svarbus regionų konkurencingumas ir jo vertinimas. Susidomėjimas vertinti regionų konkurencingumą susijęs su pranašumais, kuriuos tai gali sukelti. Regiono konkurencingumo vertinimas pateikia regiono dabartinės ekonominės aplinkos įvertinimą, nustato regiono ekonomikos silpnybes, skatina ilgalaikę ekonomikos plėtros perspektyvą, skatina kurti naujas rinkodaros ir reklamos programas regione [10]. Be to, tai yra būdas įvertinti praeityje gautus rezultatus, kurie taip pat gali būti laikomi kaip potencialas būsimai konkurencijai [14]. Regionų konkurencingumas yra vienas svarbiausių komponentų siekiant nuolatinės plėtros vietos lygmeniu, o tam, kad ji būtų pasiekta, reikalingos žinios apie regiono ekonomiką (įmones, sektorius, žmoniškuosius išteklius, žaliavas, infrastruktūrą ir t. t.), taip pat ir specifinius regioninius faktorius, siekiant padaryti regioną patrauklią investicijoms, kurios padidina konkurencingumą, ekonominę plėtrą ir regiono klestėjimą [12]. Petrylė [2] atliko tyrimą ir įrodė, jog šalių, kurios yra konkurencingesnės (vertinant pagal Pasaulinį konkurencingumo indeksą – PKI), ekonomikos auga stabiliau ir susiduria su švelnesniais svyravimais palyginus su tomis šalimis, kurių

konkurencingumas yra mažesnis. Tahir ir Tahir'is [16] taip pat teigia, jog konkurencingumas yra svarbus nuolatiniam (angl. *sustained*) augimui, o konkurencinga šalis gali daugiau parduoti. Autoriai pažymi, jog šalys konkuruoja viena su kita suteikdamos palankesnę aplinką verslui, jog galėtų pritraukti konkurencingas įmones [17]. Bayar'is ir kt. [18] taip pat pripažįsta, jog konkurencingumas yra svarbus faktorius tiesioginėms užsienio investicijoms pritraukti.

Atlikus mokslinės literatūros analizę galima teigti, jog skirtingi autoriai pateikia savitas regiono konkurencingumo apibrėžimo sąvokas, tačiau jose yra panašumų. Apibendrinant autorių mintis apie regiono konkurencingumą, konkurencingas regionas gali būti apibrėžiamas kaip dinamiškas regionas, kuris išsiskiria iš kitų regionų tam tikromis savybėmis ir gali suteikti palankią aplinką piliečiams gyventi ir įmonėms veikti tiek trumpuoju, tiek ilguoju laikotarpiu. Svarbu vertinti regionų konkurencingumą tam, kad būtų nustatytos silpnosios vietos ir kad būtų aišku, kas turėtų būti tobulinama, kad regionas klestėtų ir užsitikrintų konkurencinį pranašumą. Konkurencingas regionas – tai tarsi žinutė, padedanti pritraukti inovatyvių darbuotojų, investicijų.

## **1.2. Konkurencingumo vertinimo metodai**

Prieš pradėdant analizuoti konkurencingumą lemiančius veiksnius, svarbu pristatyti jau egzistuojančią šalių konkurencingumo vertinimo metodiką. Ji ir bus trumpai pristatoma šiame skyrelyje.

Vienas labiausiai žinomų sudėtinių indeksų šalių konkurencingumui vertinti yra Pasaulinis konkurencingumo indeksas (PKI) (angl. *Global Competitiveness Index – GCI*), kurį kiekvienais metais skelbia Pasaulio ekonomikos forumas (angl. *The World Economic Forum – WEF*). Pasaulio ekonomikos forumas skelbia Pasaulinį konkurencingumo indeksą nuo 2004 m. [19]. Naujausia PKI metodikos versija, kuri buvo pristatyta 2018 m., yra PKI 4.0 (angl. *GCI 4.0*) [5]. PKI 4.0 susideda iš 12 pagrindinių rodiklių grupių (angl. *pillars*), nusakančių produktyvumą. Bendras PKI 4.0 balas yra 12 grupių balų vidurkis, o šios 12 grupių apima 103 rodiklius, iš kurių 47 yra gaunami apklausos būdu. Likę rodikliai gaunami iš tarptautinių organizacijų, akademinė institucijų ir nevyriausybinių organizacijų. 12 rodiklių grupių yra suskirstytos į 4 didesnes grupes – įgalinančios aplinkos (angl. *Enabling Environment*), žmogiškojo kapitalo (angl. *Human Capital*), rinkų (angl. *Markets*) ir inovacijų ekosistemos (angl. *Innovation Ecosystem*) (žr. 2 pav.).

Verta atkreipti dėmesį, jog ankstesnėje šio indekso metodikoje rodikliai buvo suskirstyti į 3 grupes – bazinius, efektyvumo ir inovacijų [20]. Kiekvienai iš trijų dimensijos grupių (bazinei, efektyvumo, inovacijų) buvo suskaičiuojamas subindeksas, kuris buvo tai grupei priklausančių normalizuotų rodiklių grupių vidurkis. Pasaulinis konkurencingumo indeksas buvo apskaičiuojamas kaip svertinis visų trijų subindeksų rodiklių vidurkis, kur svorių suma buvo lygi 1. Svoriai buvo suteikiami atsižvelgiant į analizuojamos šalies išsivystymo lygį, kuris buvo vertinamas BVP, tenkančiu viename gyventojui, rodikliu. Kuo šalis labiau išsivysčiusi (t. y. didesnis BVP gyventojui), tuo didesnis svoris buvo suteikiamas inovacijų subindeksui ir mažesnis svoris buvo suteikiamas baziniam subindeksui. Esminis ankstesnio PKI nuo PKI 4.0 skirtumas yra tas, jog dabar kiekvienai rodiklių grupei yra suteikiamas vienodas svoris.

### Įgalinanti aplinka

1 rodiklių grupė <b>Institucijos</b>
2 rodiklių grupė <b>Infrastruktūra</b>
3 rodiklių grupė <b>IRT įsisavinimas</b>
4 rodiklių grupė <b>Makroekonominis stabilumas</b>

### Žmogiškasis kapitalas

5 rodiklių grupė <b>Sveikata</b>
6 rodiklių grupė <b>Įgūdžiai</b>

### Rinkos

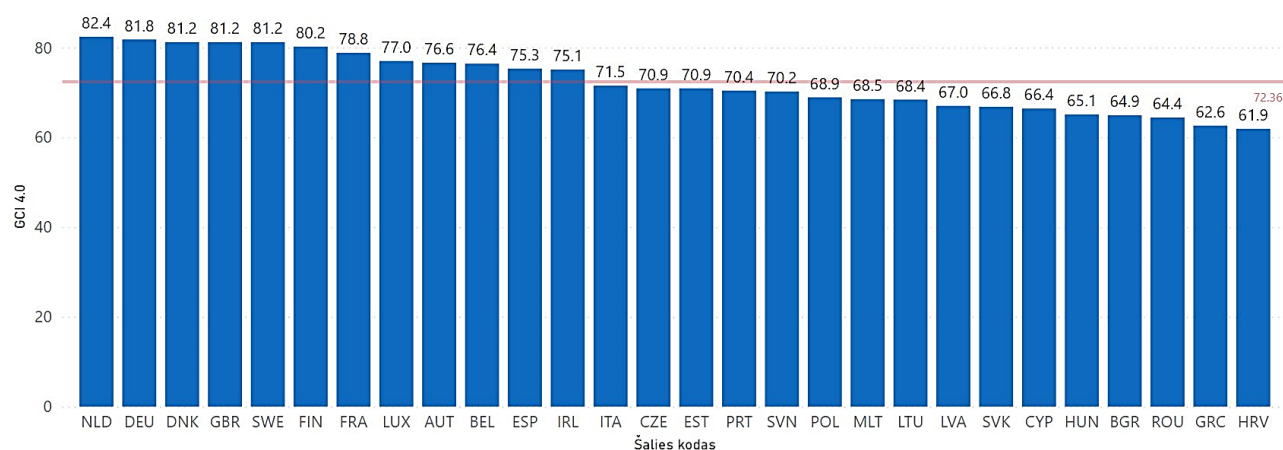
7 rodiklių grupė <b>Produktų rinka</b>
8 rodiklių grupė <b>Darbo rinka</b>
9 rodiklių grupė <b>Finansų sistema</b>
10 rodiklių grupė <b>Rinkos dydis</b>

### Inovacijų ekosistema

11 rodiklių grupė <b>Verslo dinamiškumas</b>
12 rodiklių grupė <b>Inovacijų galimybės</b>

2 pav. PKI 4.0 struktūra, [5 p. 2]

PKI 4.0 skaičiavimas yra atliekamas agreguojant balus, pradedant nuo rodiklio lygio ir baigiant bendru PKI 4.0 balu. Kiekviename agregavimo lygyje kiekvienas agreguotas matas apskaičiuojamas imant jo komponentų aritmetinį vidurkį, o bendras PKI 4.0 balas gaunamas apskaičiuojant 12 grupių aritmetinį vidurkį. Prieš agregavimą rodiklių reikšmės yra transformuojamos į pažangos balą, kurio intervalas yra nuo 0 iki 100, kur 100 yra ideali, geriausia reikšmė, o 0 – blogiausia.



3 pav. ES 2019 m. PKI 4.0 balai, sudaryta autorės, remiantis šaltiniu [5]

Remiantis 2019 m. Pasaulio ekonomikos forumo ataskaitos duomenimis, buvo palygintos ES šalys (žr. 3 pav.). Pagal PKI 4.0 aukščiausiu konkurencingumu balu ES įvertinti Nyderlandai ir Vokietija, Danija, Jungtinė Karalystė, Švedija. Galima pastebėti, jog 2019 m. PKI 4.0 ES šalyse svyruoja nuo 61,90 iki 82,40, o vidurkis yra 72,36. Žemiausi PKI 4.0 balai yra Kroatijos, Graikijos, Rumunijos, Bulgarijos ir Vengrijos. 12-os šalių PKI 4.0 balas yra didesnis už ES šalių vidurkį ir 16-os šalių mažesnis už ES šalių vidurkį.

Aptariant konkurencingumo vertinimo metodus svarbu paminėti M. Porter'į, kuris sukūrė „Nacionalinio deimanto“ modelį. Porter'io [21] sukurtas modelis apibrėžia 4-is svarbiausius veiksnius, lemiančius šalies konkurencingumą, tai: 1) veiksnų sąlygos; 2) paklausos sąlygos; 3) susijusios ir palaikančios pramonės šakos; 4) įmonių strategija, struktūra ir konkurencija. Veiksnų sąlygos – tai gamybos veiksniai, reikalingi konkuruojant pramonės šakoje. Šie veiksniai grupuojami į: žmogiškuosius išteklius, fizinius išteklius, žinių išteklius, kapitalo išteklius, infrastruktūrą. Konkurencinis pranašumas priklauso nuo to, kaip efektyviai šie veiksniai bus panaudoti. Paklausos sąlygos – tai pramonės produkto ar paslaugos paklausos šalyje pobūdis. Susijusios ir palaikančios pramonės šakos – tai veiksnys, kuris apima pramonės šakų, kurios papildo viena kitą, sistemą (pavyzdžiui, pramonės šaka – chemikalai, susijusi pramonės šaka – spausdinimo rašalas). Įmonių strategija, struktūra ir konkurencija – tai sąlygos šalyje, reguliuojančios įmonių steigimą, organizavimą ir valdymą, ir vidaus konkurencijos pobūdis. Porter'is [21] į modelį taip pat įtraukė ir atsitiktinius įvykius (tokius, kurių dažniausiai negalima paveikti, pavyzdžiui, karai, užsienio vyriausybės sprendimai) ir vyriausybę. Simanavičienė ir kt. [7] pripažįsta, jog Porter'is padarė svarų indėlį, sukurdamas šalies konkurencingumo modelį, tačiau teigia, jog abejotinas šio modelio pritaikymas mažų šalių ar regionų konkurencingumui vertinti.

Analizuojant mokslinius straipsnius buvo pastebėta, jog yra nevengiama šalies konkurencingumo vertinti ir atskirais makroekonominiais rodikliais, vienas dažniausiai sutinkamų – BVP. Rucinska ir Rucinsky [22] išskiria dvi sąvokas – veiksnius (angl. *drivers*), kurie apibrėžia regiono gebėjimą konkuruoti su kitais regionais ir rezultatus (angl. *outcomes*), kurie apibūdina konkurencingumo pasėkmę. Minėti autoriai pažymi, jog konkurencingumas gali būti išmatuojamas tokiais regiono rodikliais kaip BVP, tenkančiu gyventojui, nedarbo lygiu, vidutiniu atlyginimu. Ginevicius ir kt. [23] teigia, jog BVP atspindi nacionalinį konkurencingumą ir išskiria tokio rodiklio privalumus – BVP apima daugelį šalies veiksnų, yra lengvai prieinamas tiek šalies, tiek tarptautiniuose statistikos šaltiniuose, šis rodiklis yra palyginamas tarp šalių, nes jo apskaičiavimui naudojama vieninga metodika.

Dauguma autorių sutinka, jog konkurencingumo vertinimas yra sudėtingas uždavinys. Analizuojant mokslinę literatūrą buvo pastebėta, jog dažnai konkurencingumas vertinamas naudojant ne vieną rodiklį, o naudojant rodiklių rinkinius, kuriant modelius ir konkurencingumo indeksus. Vienas tokių modelių – „Snaigės“ (angl. *Snowflake*) modelis, kuris sudarytas iš 8-ių veiksnų grupių [13]. Šis modelis buvo sukurtas siekiant įvertinti Serbijos regionų konkurencingumą. Minėtas modelis ypatingas tuo, jog apima ne tik daug rodiklių, tačiau jis turi ir vizualinę interpretaciją. Analizuojant konkurencingumo indeksus svarbu paminėti lietuvių mokslininkus V. Sniešką ir J. Bruneckienę, kurie padarė svarų indėlį sukurdami konkurencingumo indeksą (RKI), kurį naudojo Lietuvos apskričių konkurencingumui vertinti [24].

Apibendrinant galima teigti, jog vienas dažniausiai naudojamų rodiklių šalies konkurencingumui vertinti yra Pasaulinis konkurencingumo indeksas. Sutinkama nemažai mokslininkų, kurie savo darbuose tiria ekonominių, socialinių rodiklių ryšį su Pasauliniu konkurencingumo indeksu. Nors yra tyrėjų, kurie nevengia vertinti konkurencingumo makroekonominiais rodikliais, tačiau analizuojant mokslinę literatūrą tokių tyrėjų nebuvo sutikta daug. Mokslininkai sutinka, jog konkurencingumo vertinimas yra kompleksinis uždavinys, todėl dažniausiai yra naudojami rodiklių rinkiniai, sudaromas konkurencingumo indeksas.



### 1.3. Teritorijos konkurencingumą lemiantys veiksniai

Analizuojant regionų konkurencingumo sąvokas buvo galima pastebėti, jog jų apibrėžimai yra gana panašūs. Galima teigti, jog yra priimtas konsensusas dėl regiono konkurencingumo apibrėžimo, tačiau autorių nuomonių skirtumai išryškėja kuriant regionų konkurencingumo modelį ir parenkant konkurencingumą lemiančius veiksniai [13]. Analizuojant mokslinę literatūrą, buvo pastebėta, jog nėra vienintelės regionų konkurencingumo vertinimo metodologijos, o regiono konkurencingumą vertinti tik pagal vieną pasirinktą rodiklį (pavyzdžiui, eksportą, BVP) nėra korektiška. Petrylė [2] tvirtina, jog konkurencingumas negali būti suprantamas vien tik kaip produktyvumas ar eksporto rezultatai, nes tokie rodikliai neparodo šalies ekonomikos pastovumo (angl. *sustainability*) ar piliečių gyvenimo lygio. Ruzekova ir kt. [25] taip pat pripažįsta, jog eksporto rezultatai nėra universalus konkurencingumo matas, reikia naudoti daugiau konkurencingumą apibūdinančių rodiklių. Kiekybiniai indikatoriai, tokie kaip BVP, negali visiškai įvertinti regionų konkurencingumo, vertinimui reikia naudoti ir kokybinius kintamuosius, kurie padeda paaiškinti skirtumus tarp regionų [10]. Autoriai sutinka, jog BVP svarbus rodiklis matuojant konkurencingumą, tačiau jis neatspindi kitų svarbių konkurencingumo veiksnių, tokių kaip laiko praleidimo po darbo kokybės, turimų natūralių išteklių, nusikalstamumo lygio, BVP suteikia tik dalinę informaciją, susijusią su konkurencingumu.

Kiekvieno regiono siekis yra būti konkurencingu, bet tai pasiekti nėra lengva, kiekvienas regionas yra unikalus ir turi savo kontekstinę situaciją, kurios negalima nepaisyti [26]. Įvertinti regionų konkurencingumą nėra paprasta. Regionų konkurencingumas ir tyrėjams, ir politikams kelia iššūkių siekiant jį išmatuoti, nėra regionų konkurencingumo mato, kuris apimtų visus faktorius [10]. Konkurencingumas neturi tiesioginių išraiškos formų ar požymių, dėl kurių galėtų būti tiesiogiai matuojamas [14]. Išties, analizuojant mokslinę literatūrą buvo pastebėta, jog skirtingi autoriai išskiria nevienodus konkurencingumo veiksniai. Kaip ir pati konkurencingumo sąvoka, taip ir konkurencingumo vertinimas yra itin platus uždavinys, o išrinkti visus konkurencingumą lemiančius veiksniai, tiesiog neįmanoma. Toliau skyreliuose bus pateikiama apibendrinta informacija apie konkurencingumui įtaką darančius veiksniai, kurie buvo suskirstyti į 4-ias didesnes grupes – **ekonominiai, socialiniai, šalies aplinkos, aplinkosauginiai**.

#### 1.3.1. Ekonominiai veiksniai

Atlikus mokslinės literatūros analizę buvo pastebėta, jog didelė dalis autorių dėmesį skiria ekonominiams veiksniams, kurie daro poveikį konkurencingumui. Ekonominiai veiksniai yra pateikiami 2 lentelėje. Ekonominius veiksniai būtų galima apibūdinti kaip regiono ekonominę situaciją, kuri apima įmones, darbuotojus, makroekonominius rodiklius, infrastruktūrą ir inovacijas.

**2 lentelė.** Regiono konkurencingumą lemiantys veiksniai, ekonominė veiksnių grupė

Veiksniai	Autoriai
Investicijos:	-
<i>TUI</i>	Vukovic ir Wei [9], Bayar ir kt. [18], Danilevičienė ir Lukšytė [27], Skare ir Cvek [28]
<i>Investicijos</i>	Nikolic ir kt. [12], Vukovic ir Wei [9], Ginevicius ir kt. [23], Russu [29], Vukovic ir kt. [13], Drozd ir Volkov [30]
Ekonomikos atvirumas:	Nikolic ir kt. [12], Zinovyeva ir kt. [31]
<i>Eksportas</i>	Simanavičienė ir kt. [7], Drozd ir Volkov [30], Russu [29]

<i>Importas</i>	Simanavičienė ir kt. [7], Drozd ir Volkov [30], Cheba ir kt. [32]
Ekonominiai rezultatai	Balkytė ir Tvaronavičienė [3], Wenzel ir Wolf [33], Ginevicius ir kt. [23]
<i>Verslumas ir įmonių rezultatai</i>	Vukovic ir kt. [13], Russu [29], Clipa ir Ifrim [10], Szymańska [34], Zinovyeva ir kt. [31], Ignatov [35], Popa [36], Drozd ir Volkov [30], Rucinska ir Rucinsky [22]
<i>Užimtumas</i>	Vukovic ir kt. [13], Nikolic ir kt. [12], Cheba ir kt. [32], Russu [29]
<i>Darbo užmokestis</i>	Vukovic ir kt. [13]
<i>Produktyvumas</i>	Nikolic ir kt. [12], Ginevicius ir kt. [23], Vukovic ir Wei [9], Gardiner ir kt. [37]
<i>Turizmas</i>	Vukovic ir kt. [13], Čibinskienė [38], Drozd ir Volkov [30]
Inovacijos ir technologijos	Nikolic ir kt. [12], Vukovic ir Wei [9], Aiginger ir Vogel [39], Zinovyeva ir kt. [31], Terzić [40], Kruk [41], Denkowska ir kt. [42], Banelienė [43], Valodkienė ir Snieška [44], Ignatov [35], Popa [36], Cheba ir kt. [32], Kiseľáková ir kt. [45], Vukovic ir kt. [13], Russu [29], Drozd ir Volkov [30], Androniceanu ir kt. [46], Balkytė ir Tvaronavičienė [3], Szymańska [34], Skare ir Cvek [28]
Infrastruktūra	Rucinska ir Rucinsky [22], Simanavičienė ir kt. [7], Szymańska [34], Zinovyeva ir kt. [31], Kruk [41], Drozd ir Volkov [30], Vukovic ir kt. [13], Vukovic ir Wei [9], Russu [29], Wenzel ir Wolf [33], Carvalho ir Fonseca [17]

Vienas iš regionų konkurencingumą lemiančių veiksnių – **investicijos**. Danilevičienė ir Lukšytė [27] teigia, jog tiesioginės užsienio investicijos yra vienas esminių faktorių, didinančių produktyvumą ir padedančių vykdyti nuolatinę šalies plėtrą. Minėtos autorės su Lietuvos 2006–2015 m. duomenimis atliko tyrimą ir įrodė, jog tarp TUI ir BVP, eksporto ir darbo užmokesčio egzistuoja stiprus teigiamas ryšys, todėl autorės priėjo prie išvados, jog TUI įplaukos daro teigiamą poveikį šalies ekonomikos konkurencingumui. Ginevicius ir kt. [23] pažymi, jog šalies konkurencingumas priklauso nuo ekonominės plėtros, kurią lemia daug veiksnių, vienas jų – investicijos. Nikolic’as [12] teigia, jog regiono gebėjimas pritraukti investicijas, ypač užsienio, užtikrinti papildomą užimtumą, eksporto rezultatus ir pažangumą yra tai, kas turėtų būti regiono politikos dėmesio centras. Lengyel’is and Rechnitzer’as [47] pažymi, jog regionų ekonominė plėtra yra glaudžiai susijusi su jų gebėjimu pritraukti ir palaikyti sėkmingą gamybinę veiklą, o gaunamos TUI didina užimtumą.

Kitas svarbus veiksnys, kurį autoriai išskiria kaip darantį poveikį konkurencingumui – **ekonomikos atvirumas**, kuris suprantamas kaip tarptautinė prekyba, vertinant eksportą ir importą. Jei regionas geba užtikrinti pakankamą eksporto lygį, vadinasi, regione yra sukuriamas užimtumas, kuris padeda palaikyti gyventojų pajamų lygį [12]. Vilpišauskas [48] pažymi, jog tarptautinė prekyba padeda išnaudoti šalių pranašumus plėtojant darbo pasidalinimą ir specializaciją, taip pat ji skatina ekonomikos augimą.

Dar viena svarbi veiksnių grupė, kuri lemia regionų konkurencingumą, tai – **ekonominiai rezultatai**. Remiantis moksline literatūra ir autorių išskirtais veiksniais, į šią grupę buvo įtrauktas verslumas ir įmonių rezultatai, užimtumas, darbo užmokestis, produktyvumas ir turizmas. Clipa ir Ifrim [10] pažymi, jog konkurencingumo vertinimas turėtų būti orientuotas į įmonių rezultatus, nes būtent įmonės yra pagrindiniai ekonomikos veikėjai ir ekonominio turto kūrėjai. Užimtumas, darbo užmokestis buvo išskirti kaip veiksniai, kurie atspindi regiono gebėjimą įdarbinti gyventojus ir suteikti adekvatų užmokestį [13]. Produktyvumas buvo pasirinktas, nes, kaip nurodo Gardiner’is ir kt. [37], produktyvumo augimas yra esminis regioninės ekonomikos klestėjimo veiksnys. Dar vienas veiksnys, prisidedantis prie ekonominės regiono gerovės – turizmas. Kaip teigia Čibinskienė [38], turizmas yra reikšminga paslaugų ekonomikos dalis.

Daugiausiai dėmesio analizuotoje mokslinėje literatūroje autoriai skiria **inovacijoms ir technologijoms**. Europos Centrinis Bankas [49] inovacijas apibūdina kaip itin svarbų veiksnį, kuris skatina ekonomikos pažangą, o inovacijų sąvoką apibrėžia kaip sukurtas ir taikomas naujas idėjas bei technologijas, pagerinančias prekes ir paslaugas arba padidinančias jų gamybos ir teikimo našumą. Valodkienė ir Snieška [44] pabrėžia, jog analizuojant konkurencingumą ir jį lemiančius veiksnius, vis dažniau akcentuojama inovacijų svarba. Martin'as [50] teigia, jog žinios ir inovacijos – tai pagrindiniai veiksniai, lemiantys ekonomikos plėtrą. Terzić [40] teigia, jog besivystančių šalių ekonomikos augimas ir konkurencingumas yra itin susijęs su jų inovacijų lygiu. Banelienė [43] pažymi, jog inovacijos užtikrina šalies vystymąsi, ekonomikos augimą ir pažangą, konkurencingumą ir sėkmę ilguoju laikotarpiu. Vukovic'as ir kt. [13] akcentuoja, jog šiuolaikinių technologijų naudojimas ir naujų žinių pritaikymas padidina našumą, o Zinovyeva ir kt. [31] pažymi, jog taikant inovacijas neišsivystę regionai gali sumažinti atotrūkį nuo pirmaujančių regionų ir netgi juos aplenksti. Androniceanu ir kt. [46] atliktas tyrimas parodė, jog egzistuoja stiprus ryšys tarp mokslinių tyrimų ir technologinės plėtros išlaidų, patentų paraiškų, žmogiškųjų išteklių mokslo ir technologijų srityse ir konkurencingumo (konkurencingumą vertinant PKI). Kiseľáková ir kt. [45] remiantis 2006–2016 m. 24 ES šalių duomenimis atliko tyrimą ir priėjo prie išvados, jog pirmaujančios šalys daugiausia dėmesio skyrė žinių ekonomikai, tyrimams, inovacijoms, technologijoms ir IT sektoriui, todėl autoriai teigia, jog šie veiksniai yra svarbiausi siekiant konkurencingumo ir nuolatinio augimo. Tyrimų ir technologijų plėtra skatina nuolatinį regiono konkurencingumo augimą [47]. Autoriai sutinka, jog inovacijos yra bene vienas iš svarbiausių konkurencingumo veiksnių, kuris skatina ekonomikos augimą ir užtikrina regiono konkurencinę poziciją ilguoju laikotarpiu.

Kitas svarbus veiksnys, kuris lemia regiono konkurencingumą, tai – **infrastruktūra**. Šis veiksnys apima svarbiausias infrastruktūros dalis, matuojant kelių, geležinkelių, oro transporto ir elektros energijos infrastruktūros kokybę. Infrastruktūros kokybė daro įtaką ne tik piliečių gyvenimo kokybei, bet ir daugelio investicinių sprendimų priėmimui [13]. Carvalho ir Fonseca [17] teigia, jog logistikos struktūra (angl. *logistics structure*) yra svarbus konkurencinio pranašumo veiksnys, ji yra svarbi sudarant palankias sąlygas verslui ir prisidedant prie ekonomikos augimo, konkurencingumo skatinimo. Minėti autoriai atlikę tyrimą įrodė, jog šalys, kuriose geresnės logistikos struktūros sąlygos, vertinant autorių apskaičiuotu indeksu (jame tokie rodikliai kaip motorinių transporto priemonių skaičius 1000 gyventojų, kelio tankis, geležinkelio tankis, oro uostų tankis, vandens kelių tankis, dyzelino kaina, benzino kaina ir kt.), pasiekia geresnių prekybos rezultatų ir taip pat jų PKI yra aukštesnis. Autoriai, kurie išskyrė infrastruktūrą kaip vieną iš veiksnių, darančių įtaką konkurencingumui, yra pateikiami 2 lentelėje.

### 1.3.2. Socialiniai veiksniai

Kaip jau buvo pastebėta anksčiau, vertinant regionų konkurencingumą neužtenka vertinti tik ekonominius rodiklius, nes konkurencingumas nėra susijęs tik su finansiniais rezultatais. Svarbu atsižvelgti ir į tame regione gyvenančius piliečius, jų gyvenimo kokybę. Todėl sudarant socialinių veiksnių grupę į ją buvo įtraukti šie veiksniai – žmogiškieji išteklių ir piliečių gyvenimo kokybė (vertinant sveikatą ir socialinę apsaugą, saugumą, gyvenamos erdvės kokybę, materialaus turto lygį bei kultūrą). Socialiniai veiksniai yra pateikiami 3 lentelėje.

### 3 lentelė. Regiono konkurencingumą lemiantys veiksniai, socialinių veiksmų grupė

Veiksniai	Autoriai
Žmogiškieji ištekliai	Vukovic ir Wei [9], Szymańska [34], Aiginger ir Vogel [39], Kruk [41], Godelytė Korsakienė [51], Ignatov [35], Popa [36], Januškaitė ir Užienė [26], Nikolic ir kt. [12], Wenzel ir Wolf [33], Terzić [40], Rucinska ir Rucinsky [22], Ginevicius ir kt. [23], Denkowska ir kt. [42], Bayar ir kt. [18], Valodkienė ir Snieška [44], Gižienė ir Simanavičienė [52], Skare ir Cvek [28], Cheba ir kt. [32], Vukovic ir kt. [13], Drozd ir Volkov [30]
Gyvenimo kokybė:	Balkytė ir Tvaronavičienė [3], Cheba ir kt. [32]
<i>Sveikata ir socialinė apsauga</i>	Szymańska [34], Gavurova ir kt. [53], Wenzel ir Wolf [33], Drozd ir Volkov [30], Cheba ir kt. [32], Russu [29], Vukovic ir kt. [13]
<i>Saugumas</i>	Drozd ir Volkov [30], Rucinska ir Rucinsky [22], Cheba ir kt. [32]
<i>Gyvenamos erdvės kokybė</i>	Vukovic ir kt. [13], Drozd ir Volkov [30]
<i>Materialaus turto lygis, skurdas ir socialinė atskirtis</i>	Drozd ir Volkov [30], Cheba ir kt. [32]
<i>Kultūra</i>	Vukovic ir kt. [13], Drozd ir Volkov [30]

Vertinant regiono konkurencingumą, **žmogiškasis kapitalas** yra vienas iš esminių autorių išskiriamų veiksnių, nes nuo jo kokybės – turimo išsilavinimo ir įgūdžių – gali priklausyti regionų sėkmė. Pasauliui vis labiau globalėjant, svarbiu ekonomikos veiksniu tampa ne materialiniai ar finansiniai ištekliai, o žmogiškasis kapitalas (žinios, gebėjimai, įgūdžiai) [52]. Januškaitė ir Užienė [26] pažymi, jog žmogiškasis kapitalas susijęs su žiniomis, jis lemia didesnę socialinį sąmoningumą apie socialines ir ekologines problemas, jis yra vienas pagrindinių veiksnių, lemiančių regionų konkurencingumą. Denkowska ir kt. [42], Valodkienė ir Snieška [44] žmogiškąjį kapitalą sieja su inovacijų augimu, kuris taip pat yra vienas esminių konkurencingumą lemiančių veiksnių. Denkowska ir kt. [42] atliko tyrimą, kurio tikslas buvo ištirti, koks ryšys egzistuoja tarp 25–64 m. žmonių išsilavinimo tipo ir ES šalių inovacijų lygio. Tyrimo rezultatai atskleidė, jog papildomas išsilavinimas daro ženkliai ir teigiamą įtaką inovacijoms ir konkurencingumui, ypač tų žmonių papildomas išsilavinimas, kurie jau turi aukštąjį išsilavinimą. Denkowska ir kt. [42] taip pat pažymėjo, jog šalys, kuriose ženkliai darbo jėgos dalis, kuri mokosi visą gyvenimą (angl. *lifelong learning*), pasiekia aukštesnę konkurencingumą. Ekonomikoje vykstantys spartūs pokyčiai reiškia, kad darbuotojų žinios ir įgūdžiai nuolat nuvertėja, tad reikia juos atnaujinti, todėl autoriai [42] pabrėžia neformaliojo švietimo svarbą. Investicijos į mokslą yra vienas svarbiausių šalies sėkmės garantų, nes tai yra vienas iš veiksnių, dėl kurių vyksta inovacijų plėtra [44]. Gižienė ir Simanavičienė [52] teigia, jog valstybė, kuri skiria daugiau lėšų švietimui, vystosi greičiau nei valstybė, skirianti mažiau lėšų.

**Gyvenimo kokybės** veiksnys apibūdina regiono gebėjimą suteikti sveikatos ir socialinę apsaugą, sukurti saugią aplinką gyvenimui, atsižvelgti į gyvenamąją erdvę, išlaikyti aukštą materialaus turto lygį, išlaikyti žemą socialinę atskirtį ir suteikti kultūrinių paslaugų, kas neabejotinai prisideda prie piliečių gyvenimo kokybės. Galima tikėtis, jog sveiki piliečiai sugebės sukurti konkurencingus ekonominius rezultatus, kuriais būtų galima konkuruoti su kitomis ekonomikomis. Gavurova ir kt. [53] vertino sveikatos rodiklių poveikį išsivysčiusių šalių konkurencingumui. Minėti autoriai akcentavo, jog sveikata yra vienas iš esminių šalies ekonominio gyvenimo aspektų. Šalys turėtų sutelkti dėmesį į savo gyventojų sveikatą, kad padidintų konkurencingumą, o tai galima pasiekti pagerinus šalies sveikatos ir socialinę sistemą [53]. Autoriai, kurie atsižvelgė į socialinius konkurencingumo veiksnius, tokius kaip saugumą – Drozd ir Volkov’as [30], Rucinska ir

Rucinsky [22], Cheba ir kt. [32]; gyvenamos erdvės kokybę – Vukovic’as ir kt. [13], Drozd ir Volkov’as [30], materialaus turto lygį ir socialinę atskirtį - Drozd ir Volkov’as [30], Cheba ir kt. [32]; kultūrą – Vukovic’as ir kt. [13], Drozd ir Volkov’as [30]. Kultūra neturi didelės įtakos ekonominiams regiono rezultatams, tačiau ji daro įtaką daugelio toje teritorijoje gyvenančių žmonių pasitenkinimui [13].

### 1.3.3. Šalies aplinkos veiksniai

Šalies aplinkos veiksniai apima verslo aplinką, makroekonominį stabilumą bei institucinę ir politinę aplinką. Šie veiksniai yra pateikiami 4 lentelėje.

**4 lentelė.** Regiono konkurencingumą lemiantys veiksniai, šalies aplinkos veiksmų grupė

Veiksniai	Autoriai
Verslo aplinka	Vukovic ir Wei [9], Vukovic ir kt. [13], Ginevicius ir kt. [23]
Makroekonominis stabilumas	Simanavičienė ir kt. [7], Skare ir Cvek [28], Wenzel ir Wolf [33]
Institucinė ir politinė aplinka	Skare ir Cvek [28], Cheba ir kt. [32], Zinovyeva ir kt. [31], Ruzekova ir kt. [25], Vukovic ir Wei [9], Januškaitė ir Užienė [26], Vukovic ir kt. [13], Ginevicius ir kt. [23], Bayar ir kt. [18], Wenzel ir Wolf [33], Ignatov [35], Rucinska ir Rucinsky [22], Cheba ir kt. [32], Kruk [41]

Kadangi įmonės yra svarbūs dalyviai kuriant šalies konkurencingumą, vienas iš autorių išskiriamų veiksmų, darančių įtaką konkurencingumui, yra **verslo aplinka**. Tam, kad būtų pradėtas verslas, pritraukiamos investicijos turi būti sukurta palanki aplinka verslui. Geresnė aplinka verslui suteikia galimybes plėtoti verslumo potencialą [9], o verslumo prioritizavimas yra svarbus, nes jis skatina žmogiško kapitalo kūrimą [35].

Taip pat kaip atskiras veiksnys buvo išskirtas ir **makroekonominis stabilumas**, nes anot Simanavičienės ir kt. [7], vertinant regionų konkurencingumą, daugiau dėmesio turėtų būti skiriama faktiniams baziniams ekonomikos pagrindams, vienas iš jų – makroekonomikos stabilumas. Skare ir Cvek [28] taip pat pripažįsta, jog makroekonominis stabilumas yra būtina sąlyga ekonominiam konkurencingumui. Martin’as [50] atliko šaltinių analizę ir padarė išvadą, jog makroekonominis stabilumas dažnai traktuojamas kaip išankstinė konkurencingumo sąlyga.

Daugiausia dėmesio šalies aplinkos veiksmų grupėje analizuotos mokslinės literatūros autoriai skyrė **institucinei ir politinei aplinkai**. Ginevicius ir kt. [23] tyrime analizavo šešėlinės ekonomikos įtaką šalies ekonominei plėtrai ir priėjo prie išvados, jog šešėlinė ekonomika yra kliūtis šalies ekonominei plėtrai. Kaip būdus sumažinti šešėlinę ekonomiką minėti autoriai pateikė – mokesčių naštos mažinimą, mokesčių moralės skatinimą, įstaigų kokybės gerinimą. Bayar’is ir kt. [18] taip pat tyrė šešėlinės ekonomikos įtaką TUI (kaip galimam konkurencingumo veiksmui). Minėtų autorių empirinis tyrimas atskleidė, jog šešėlinė ekonomika neigiamai veikia TUI. Instituciškai silpnose šalyse, kurios susiduria su biurokratija, korupcija, TUI pritraukimas nėra toks didelis [18]. Ignatov’as [35] taip pat pripažįsta, jog pagerinta institucinė kokybė (mažinant korupciją ir biurokratiją) prisidėtų prie geresnės aplinkos verslui. Institucinė aplinka ( tiek formali, tiek neformali) yra svarbus faktorius konkurencingumui, todėl jis turėtų būti vertinamas, vyriausybės politika gali daryti teigiamą arba neigiamą poveikį ekonominei aplinkai [25].

### 1.3.4. Aplinkosauginiai veiksniai

Vertinant regiono konkurencingumą svarbu atsižvelgti ir į aplinkosauginius veiksnius, nes konkurencingumas yra susijęs su ilguoju laikotarpiu. Tam, jog būtų užtikrinta būsimų kartų ateitis, būtina atsižvelgti į gamtos išsaugojimą, atsinaujinančius energijos šaltinius, užterštumo mažinimą, atliekų perdirbimą. Į aplinkosauginių veiksnių grupę buvo įtraukti tokie veiksniai, kaip aplinka ir užterštumas, atsinaujinančioji energetika bei žiedinė ekonomika. Aplinkosauginiai veiksniai yra pateikiami 5 lentelėje.

**5 lentelė.** Regiono konkurencingumą lemiantys veiksniai, aplinkosauginių veiksnių grupė

Veiksniai	Autoriai
Aplinka ir užterštumas	Cheba ir kt. [32], Drozd ir Volkov [30], Aiginger ir Vogel [39], Kasztelan [54], Vukovic ir kt. [13], Kruk [41]
Atsinaujinančioji energetika	Cheba ir kt. [32], Simionescu ir kt. [55], Aiginger ir Vogel [39]
Žiedinė ekonomika	Europos Komisija [56]

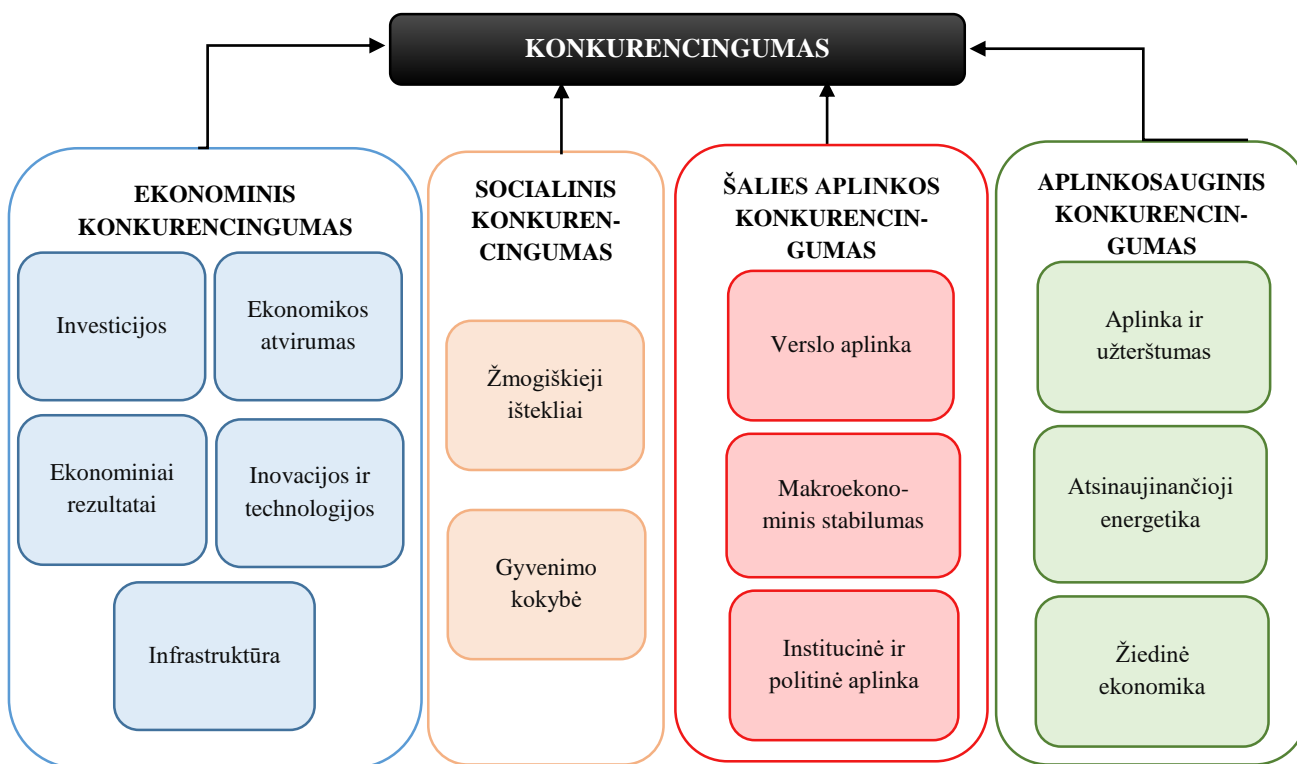
Buvo išskirta **aplinkos ir užterštumo** veiksnių grupė. Autorių, išskiriančių šį veiksnių, analizuotoje mokslinėje literatūroje nebuvo daug. Kasztelan'as [54] teigė, jog gamtinis kapitalas (angl. *natural capital*) turėtų būti laikomas vienu svarbiausių regionų konkurencingumo veiksnių, tačiau šiam veiksniai kol kas daug dėmesio nėra skiriama. Minėtas autorius atliko tyrimą ir siekė įvertinti Lenkijos vaivadijų aplinkos konkurencingumą (angl. *environmental competitiveness*), naudodant tokius rodiklius, kaip miško žemės dalis visoje vaivadijos teritorijoje, dirbtinių trąšų sunaudojimas, susidariusių pramoninių atliekų kiekis, perdirbtų atliekų kiekis ir kitus. Kasztelan'as [54] išskyrė tinkamo aplinkos potencialo panaudojimo privalumus – papildomų ekonominių rezultatų sukūrimas, tolimesnis aplinkos būklės gerinimas, socialinių tikslų skatinimas mažinant nedarbą ir gerinant gyvenimo kokybę. Taip pat aukšta aplinkos kokybė, palaikoma žemu taršos lygiu, gali būti laikoma tam tikru prekės ženklu, galinčiu įtakoti regioninę plėtrą. Aktyvios aplinkos apsaugos priemonės skatina „žaliąją“ regionų specializaciją ir taip padidina jų konkurencingumą [54]. Drozd ir Volkov'as [30] kaip vieną iš regiono konkurencingumo veiksnių išskyrė aplinkos kokybės veiksnių, susijusių su vandens ir oro užterštumu bei natūraliais ištekliais. Cheba ir kt. [32] vertinant tvarų konkurencingumą be naudojamų ekonominių, socialinių, institucinių ir politinių veiksnių atsižvelgė ir į aplinkos veiksnius. Aiginger'is ir Vogel [39] pažymi, jog ekologinės ir socialinės investicijos, jei teisingai įgyvendintos ir panaudotos, neturėtų būti laikomomis išlaidomis, bet labiau konkurencingumo įgalintojomis. Kruk [41] ieškojo panašumų tarp tvarios plėtros ir konkurencingumo ir vienas iš galimų konkurencingumo veiksnių buvo išskirtas aplinkos veiksnys.

Atsinaujinančioji energetika yra alternatyva iškastiniam kurui. Simionescu ir kt. [55] akcentavo **atsinaujinančios energetikos** svarbą ir atliko ES šalių 2007–2019 m. tyrimą, kurio tikslas buvo nustatyti ryšį tarp BVP, PKI ir atsinaujinančios energijos suvartojimo. Rezultatai atskleidė, jog atsinaujinančios energijos suvartojimas teigiamai veikia tiek BVP, tiek PKI, atsinaujinančių energijos šaltinių vartojimas yra ekonomikos augimo ir ekonominio konkurencingumo veiksnys. Autoriai išskyrė atsinaujinančių energijos šaltinių naudą – užterštumo sumažėjimas, taip pat ekonominė nauda dėl naujų darbo vietų sukūrimo. Remiantis Europos Parlamento [57] pateikiama informacija, iki 2030 m. ES yra iškeltas privalomas tikslas, jog atsinaujinančiųjų išteklių energija sudarytų nemažiau 30 proc. viso suvartojamo energijos kiekio.

ES iki 2050 m. siekia neutralizuoti poveikį klimatui ir tapti pirmuoju neutralaus poveikio klimatui žemynui. Strategija yra žinoma kaip Europos žalioji paktas (angl. *Green Deal*). Žalioji paktas yra atsakas į iššūkius, su kuriais susiduria pasaulis – klimato kaita, rūšių nykimu, tarša. Šia strategija siekiama taip pertvarkyti ES, jog ji pasižymėtų modernia, efektyviai išteklius naudojančia ir konkurencinga ekonomika [56]. Atsižvelgiant į šios strategijos siekį užtikrinti ES ekonomikos tvarumą, į aplinkosauginių veiksnių grupę buvo įtraukti **žiedinės ekonomikos** rodikliai. Žiedinė ekonomika siekiama kaip įmanoma labiau sumažinti atliekų kiekį ir išteklių naudojimą pažangiu produktų projektavimu, pakartotiniu produktų naudojimu ir taisymu, perdirbimu, darniu vartojimu [58]. Europos Parlamentas [58] išskiria šiuos žiedinės ekonomikos privalumus – sušvelnintas poveikis aplinkai, padidintas konkurencingumas, paskatintos inovacijos ir naujų darbo vietų sukūrimas, galimybė vartotojams įsigyti daugiau patvarių ir pažangių produktų. Kaip pažymi Europos Komisija [56], yra būtina labiau vertinti gamtinių ekosistemų apsaugą ir atkūrimą, tausų išteklių naudojimą ir žmonių sveikatos gerinimą, nes tai gali duoti didelę naudą ES ekonomikai, visuomenei ir gamtinei aplinkai.

### 1.3.5. Šalių konkurencingumą lemiančių veiksnių modelis

Remiantis anksčiau skyreliuose analizuota mokslinė literatūra buvo sudarytas šalių konkurencingumą lemiančių veiksnių modelis (žr. 4 pav.). Jis sukurtas atsižvelgiant į autorių išskiriamus regionų konkurencingumą lemiančius veiksniai. Modelio naujumas yra tas, jog į veiksniai, kurie daro įtaką konkurencingumui, yra įtraukiamos aplinkosauginės veiksnių grupės. Jos yra ypač svarbios, nes, kaip jau buvo minėta 1.3.4. skyrelyje, ES iki 2050 m. siekia neutralizuoti poveikį klimatui ir tapti pirmuoju neutralaus poveikio klimatui žemynui. Konkurencingumas yra susijęs ne tik su trumpuoju, bet ir su ilguoju laikotarpiu, todėl svarbu atsižvelgti į šalies gebėjimą užtikrinti tvarią aplinką, o aplinkosauginės veiksnių grupės padės tai įvertinti.



4 pav. Šalių konkurencingumą lemiančių veiksnių modelis

Siekiant suklasterizuoti ES šalis naudojant sukurtą šalių konkurencingumą lemiančių veiksmių modelį, buvo surinkti juos atspindintys 70 rodiklių, iš kurių: ekonominių 22 rodikliai, socialinių 20 rodiklių, šalies aplinkos 10 rodiklių, aplinkosauginių 18 rodiklių. Tikslūs rodiklių pavadinimai, jų matavimo vienetai, šaltinis yra pateikiami 1 priede.

#### **1.4. Matematinė dalis**

Šiame skyrelyje bus pateikiama bendra matematinių algoritmų apžvalga. Bus trumpai pristatomi trūkstumų reikšmių užpildymo metodai, aptariami klasterizavimo metodai bei pristatoma duomenų analizės programinė įranga.

##### **1.4.1. Trūkstumų reikšmių užpildymas**

Analizuojant duomenis praktikoje yra susiduriama su iššūkiais. Duomenų rinkinys, su kuriuo yra pradama dirbti, retai būna tvarkingas. Dažnai pasitaiko, jog duomenyse yra trūkstumų reikšmių. Trūkstumų reikšmių užpildymas dažnu atveju yra svarbus etapas. Dauguma duomenų analizės metodų reikalauja, jog duomenų rinkinyje nebūtų trūkstumų reikšmių. Kaip vienas iš sprendimo būdų yra stebėjimų, kuriuose yra trūkstumų reikšmių, atsisakymas. Tačiau toks sprendimo būdas yra netinkamas, kai trūkstumų reikšmių yra daug. Dar vienas sprendimo būdas – trūkstumų reikšmių pakeitimas vidurkiu. Nors šis būdas apsaugo nuo imties sumažėjimo, tačiau jis turi tam tikrų neigiamų savybių. Trūkstumų reikšmių pakeitimas vidurkiu turi šališkumo dispersijos atžvilgiu, nes yra pašalinamas natūralus kintamumas (angl. *natural variability*), kuris būtų trūkstamosiose reikšmėse, jei jos nebūtų trūkstamos [59]. Trūkstamas reikšmes užpildžius vidurkiu, gauti dispersijos įverčiai yra per maži.

Siekiant užpildyti trūkstamas reikšmes yra taikomi du pagrindiniai būdai. Pirmasis būdas, kai pakartotinai susisiekiama su respondentu ar panaudojamos dalykinės žinios tam, kad būtų užpildyta trūkstama reikšmė. Antrasis būdas apima trūkstumų reikšmių apskaičiavimą, kuris yra paremtas statistiniais metodais, pavyzdžiui, regresijos modeliais [60].

Trūkstumų reikšmių prognozės yra gaunamos iš parametrinių ar neparametrinių modelių. Užpildymo modelis prognozuoja trūkstumą reikšmę panaudojant funkciją iš pagalbinių kintamųjų. Pagalbiniai kintamieji gali būti pasirenkami iš esamo duomenų rinkinio arba gali būti naudojama istorinė informacija. Paplitę duomenų užpildymo tipai yra įvairūs regresijos modeliai, sudaryti naudojant stebėtus duomenis (angl. *observed data*) [60].

Kaip labiausiai paplitusius trūkstumų duomenų užpildymo metodus, *missForest* algoritmo kūrėjai Stekhoven'as ir Buhlmann'as [61] įvardija k-artimiausių kaimynų (tolydiniam duomenims), daugianario modelį (angl. *saturated multinomial model*) (kategoriniams duomenims) bei MICE (angl. *multivariate imputation by chained equations*) algoritmą, kuris skirtas mišriems duomenų tipams. Minėti autoriai teigia, jog visiems šiems metodams yra būtinos prielaidos apie duomenų ar kintamųjų pogrupių pasiskirstymą. K-artimiausių kaimynų trūkstumų reikšmių užpildymo metodo idėja yra ta, kad k kaimynai yra parenkami pagal tam tikrą atstumo matą ir jų vidurkis yra naudojamas trūkstamos reikšmės užpildymui, t. y. trūkstama reikšmė yra nustatoma objektų, kurie yra arčiausiai, reikšmių pagalba [62]. MICE algoritmas sukuria kelis užpildytų trūkstumų reikšmių duomenų rinkinius. Be to, šis algoritmas naudoja lygtis (angl. *chained equations*), kas suteikia lankstumo ir minėtas metodus gali susidoroti su įvairių tipų kintamaisiais (pavyzdžiui, tolydiniais ar binariniais). MICE algoritmo metu yra vykdoma regresijos modelių serija, kai kiekvienas



kintamasis su trūkstantomis reikšmėmis yra modeliuojamas atsižvelgiant į kitus kintamuosius, esančius duomenų rinkinyje. Tai reiškia, jog kiekvieną kintamąjį galima modeliuoti pagal jo pasiskirstymą, pavyzdžiui, binariniai kintamieji yra modeliuojami naudojant logistinę regresiją, o tolydiniai kintamieji modeliuojami naudojant tiesinę regresiją [63]. Kadangi daugianario modelis (angl. *saturated multinomial model*) skirtas kategoriniams duomenis, jis plačiau nebuvo apibūdinamas, detalesnė informacija gali būti randama [64] šaltinyje.

#### 1.4.2. Klasterizavimas

Objektų grupavimas yra reikalingas įvairiose srityse – inžinerijoje, medicinoje, biologijoje, ekonomikoje ir kitose. Klasterizavimas – tai duomenų tyrybos būdas, naudojamas atrasti tam tikras svarbias struktūras duomenyse [65]. Klasterizavimas – tai statistinė operacija, skirta sugrupuoti objektus į tam tikrą skaičių grupių, vadinamų klasteriais [66]. Klasterizavimas padalina duomenis taip, jog panašios duomenų struktūros yra klasterizuojamos kartu [67]. Išanalizavus mokslinę literatūrą, galima teigti, jog klasterizavimas – tai metodas, skirtas sugrupuoti objektus pagal juos apibūdinančius požymius į grupes – klasterius – taip, kad objektai, esantys tame pačiame klasteryje būtų kuo homogeniškesni, o objektai, esantys skirtinguose klasteriuose, kuo nepanašesni. Akivaizdu, jog panašumas (angl. *similarity*) – tai svarbiausias faktorius klasterizavimo procese. Dažniausiai panašumas nusakomas artumo matu (pavyzdžiui, Euklido atstumu), kuo maženis atstumas tarp objektų – tuo objektai panašesni ir atvirkščiai [67]. Euklido atstumas yra vienas populiariausių artumo matų, kiti artumo matai – Manheteno, Minkovskio.

Klasterizavimas priskiriamas prie mokymosi be mokytojo (angl. *unsupervised learning*). Tai reiškia, jog objektų priklausomybė klasteriui nėra žinoma iš anksto, ji paaiškėja tik atlikus klasterizavimą. Taip pat verta paminėti, jog dažnu atveju klasterių skaičių pasirenka duomenų tyrėjas, atsižvelgdamas į tam tikrus kriterijus. Beje, klasterizuojamas duomenų rinkinys – tai matrica, kurios eilutės yra objektai, o jos stulpeliai – tai požymiai, apibūdinantys objektus [66]. Požymiai yra išreiškiami skaitinėmis reikšmėmis [67].

Literatūroje egzistuoja daug įvairių klasterizavimo metodų. Sudėtinga pateikti aiškų klasterizavimo metodų skirstymą į kategorijas, nes kategorijos gali persidengti ir metodas gali turėti bruožų iš kelių kategorijų [68]. Toliau pateikiama lentelė, kaip skirtinguose šaltiniuose yra skirstomi klasterizavimo metodai (žr. 6 lent.).

**6 lentelė.** Klasterizavimo metodai pagal skirtingus autorius

Autoriai	Klasterizavimo metodai
Tuffe'ry [66]	Padalinimo, hierarchiniai, persidengiantys (angl. <i>fuzzy</i> ).
Han ir kt. [68]	Padalinimo, hierarchiniai, tankiu grįsti, tinkleliu grįsti (angl. <i>grid-based</i> ).
Saxena ir kt. [67]	Padalinimo, hierarchiniai.
Xu ir Tian [69]	Padalinimo, hierarchiniai, persidengiantys, pasiskirstymu grįsti, tankiu grįsti, grafų teorija grįsti, tinkleliu grįsti, fraktalų teorija grįsti, modeliu grįsti.
Govender ir Sivakumar [70]	Hierarchiniai, nehierarchiniai

Apžvelgus klasterizavimo metodų kategorijas galima teigti, jog klasterizavimo metodai skirstomi į hierarchinius ir nehierarchinius metodus. Toliau bus trumpai pristatomi padalinimo ir tankiu grįsti metodai, kurie priklauso nehierarchinių klasterizavimo metodų kategorijai, taip pat bus aptariami ir hierarchiniai metodai.

Dauguma padalinimo metodų yra paremti artumu. Tarkime turime duomenų rinkinį, kuris sudarytas iš  $n$  objektų, o algoritmui nurodomas klasterių skaičius  $k$ , padalinimo algoritmas suskirsto  $n$  objektų aibę į  $k$  dalių ( $k \leq n$ ), kur kiekviena dalis reprezentuoja klasterį. Kiekvienas objektas priklauso tik vienam klasteriui, o klasteryje turi būti mažiausiai vienas objektas. Klasteriai formuojami atsižvelgiant į tam tikrą tikslo funkciją, paremtą atstumu taip, kad objektai, esantys tame pačiame klasteryje būtų panašūs (atstumas mažas) vienas į kitą ir nepanašūs (atstumas didelis) į objektus, esančius kituose klasteriuose [68]. Dažniausiai naudojamas kriterijus – Euklido atstumas. Pagrindinė padalinimo algoritmų idėja yra duomenų taškų centrą laikyti atitinkamo klasterio centru [69]. Literatūroje išskiriami du populiariausi padalinimo metodų algoritmai –  $k$ -vidurkių (angl. *k-means*) ir  $k$ -medoidų (angl. *k-medoids*). Esminis skirtumas tarp šių dviejų algoritmų yra tas, jog  $k$ -vidurkių atveju klasterio centras yra centroidas, kuris yra klasteryje esančių objektų požymių vidurkis, o  $k$ -medoidų atveju, klasterio centrą reprezentuoja medoidas, kuris yra labiausiai klasterio centre esantis objektas.

Pagrindinė hierarchinių metodų idėja yra suklasterizuoti objektus, remiantis hierarchiniu principu. Galimi du hierarchinio klasterizavimo variantai – jungimo (angl. *agglomerative*) arba skaidymo (angl. *divisive*). Jungimo metodo pradžioje laikoma, kad kiekvienas objektas yra atskiras klasteris, vėliau jie yra apjungiami į vis didesnius klasterius, kol gaunamas vienas klasteris, o skaidymo metodo pradžioje turimas vienas didelis klasteris, kuris yra skaidomas į vis mažesnius klasterius iki tol, kol kiekvienas objektas tampa klasteriu [67].

Tam, kad būtų rasti sudėtingų formų klasteriai, yra naudojami tankiu grįsti metodai. Klasteriai yra formuojami pagal tankias sritis duomenų erdvėje, kurios yra atskirtos nuo netankių regionų, t.y. kaimyniniai objektai yra grupuojami į klasterius pagal objektų tankį. Objekto  $o$  tankis gali būti matuojamas kaip objektų, esančių šalia objekto  $o$ , skaičius [68]. Pagrindinė idėja – didinti klasterį tol, kol tankis kaimynystėje viršija tam tikrą slenkstį.

### 1.4.3. Duomenų analizės programinė įranga

Dažnai sutinkamos duomenų analizės programavimo kalbos – *Python* ir *R*. Tiek *Python*, tiek *R* yra atviro kodo ir nemokamos programavimo kalbos. Nors tokios programos kaip *Excel*, *SPSS* gali būti lengviau ir greičiau išmokstamos, tačiau jų skaičiavimo galimybės yra daug prastesnės lyginant su *R* ar *Python*, kuriems reikalingos tik bazinės programavimo žinios. Lyginant šias dvi kalbas, *Python* gali būti geresnis pasirinkimas, nes jo naudojama sintaksė yra panašesnė į kitas programavimo kalbas, tačiau daugelis programuotojų mano, kad *R* naudojama sintaksė yra lengvai išmokstama ir suprantama [71].

Remiantis Python internetinio puslapio pateikiama informacija, *Python* – tai interaktyvi, į objektinį programavimą orientuota (angl. *object-oriented*) programavimo kalba, sukurta G. V. Rossum'o 1991 m. Python nesudėtinga išmokti, naudoti, suprasti. Ši programavimo kalba naudojama ne tik statistiniams skaičiavimams, bet ir žiniatinklio ir interneto vystymui (angl. *Web and Internet development*), grafinėms naudotojo sąsajoms (angl. *Desktop GUIs*), programinės įrangos kūrimui (angl. *Software Development*), žaidimams ir t. t. *Python* turi skirtingas bibliotekas, kurios yra skirtos skirtingiems tikslams pasiekti. Pavyzdžiui, populiariausios bibliotekos, skirtos statistinei duomenų analizei yra – *Scikit Learn* (klasifikavimas, regresija, klasterizavimas, dimensionalumo mažinimas ir kt.), *Pandas* (duomenų struktūra), *Matplotlib* (duomenų vizualizavimas), *NumPy* (matematinės funkcijos, atsitiktinių skaičių generatorius ir kt.).

*R* sukūrė R. Gentleman'as and R. Ihaka 1993 m. *R* buvo sukurta, nes kūrėjai suprato, jog jų klasėms reikia geresnės programavimo aplinkos [71]. Remiantis *R* internetiniame puslapyje [72] pateikiama informacija, *R* – tai programavimo kalba ir aplinka, skirta statistiniams skaičiavimams ir grafikų pateikimui. *R* yra vektorinė programavimo kalba, teikianti platų statistinių (tiesinio ir netiesinio modeliavimo, klasikinių statistinių testų, laiko eilučių analizės, klasifikavimo, klasterizavimo ir t. t.) ir grafinių metodų įvairovę [72]. *R* yra nemokamas ir atviro kodo. Įdiegus *R* galima rasti pagrindines funkcijas, skirtas statistinei analizei. Esant poreikiui, gali būti įdiegiami papildomi paketai (angl. *packages*). Kiekvienas paketas yra skirtas tam tikram tikslui. Pavyzdžiui, kai kurie paketai naudojami analizuojant vaizdinius ar tekstinius duomenis, kiti skirti mašiniam mokymui ir t. t. Kaip pavyzdį būtų galima pateikti *ggplot2* paketą, kuris yra skirtas duomenų vizualizavimui arba *dplyr* paketą, kuris naudojamas duomenų manipuliavimui, *rpart* paketą, kuris skirtas klasifikavimui ir regresijai. *R* yra nuolat plėtojamas ir atnaujinamas. Šiuo metu (2021-04-27) CRAN paketų saugykloje yra 17489 galimi paketai.

## 2. Metodologija

Šiame skyriuje bus pristatoma programinė įranga, duomenys, tyrimo metodai, kurie buvo naudojami siekiant suklasterizuoti Europos Sąjungos šalis ir įvertinti jų konkurencingumą.

### 2.1. Programinė įranga

Šiam projektui atlikti buvo pasirinkta *R* programavimo kalba. Be naudojamų bazinių paketų, toliau pateikiami *R* paketai, kurie buvo panaudoti tyrime:

- *dplyr* – duomenų manipuliavimui;
- *DataExplorer* – žvalgomajai duomenų analizei;
- *ggplot2* – duomenų vizualizavimui;
- *missForest* – trūkstamų reikšmių užpildymui;
- *factoextra* – klasterizavimo analizei ir vizualizavimui;
- *NbClust* – indeksų skaičiavimui, kurie padeda apsispręsti dėl klasterių skaičiaus parinkimo;
- *Heatmaply* – duomenų normalizavimui;
- *tidyverse* – duomenų manipuliavimui;
- *randomForest* – parametrų derinimui, atsitiktinio miško modelio sudarymui;
- *VSURF* – svarbių kintamųjų atrankai;
- *ggpubr* – kelių grafikų, kurtų su *ggplot2*, vaizdavimui viename lange;
- *DescTools* – Kendall'o konkordancijos koeficiento skaičiavimui;
- *rgl* – vizualizavimui trimatėje erdvėje;
- *ltm* – Cronbach alpha skaičiavimui.

Vizualizavimui buvo naudotas ne tik *R*, buvo panaudotas ir *Power BI* (klasterizavimo rezultatai Europos žemėlapyje, rangų vizualizacijos).

### 2.2. Duomenys

Siekiant atlikti šį projektą, buvo renkami įvairūs rodikliai apie ES šalis 2008–2019 m. laikotarpiu ir atspindintys sukurtą modelį, kuris pateiktas 1.3.5. skyrelyje. Buvo renkami tokie rodikliai, kurie išmatuoti tokiais vienetais, kuriais remiantis būtų galima lyginti šalis tarpusavyje, t. y. buvo naudojami santykiniai rodikliai, pavyzdžiui, turizmo pajamos, tenkančios vienam gyventojui. 2020 m. rodikliai buvo nenaudoti, nes rengiant šį darbą dar nebuvo didžiosios dalies 2020 m. rodiklių, naujausia informacija buvo 2019 m. Toliau pateikiami šaltiniai, iš kurių buvo gauti duomenys.

1. *Eurostat*. Šioje duomenų bazėje pateikiami Europos šalių statistiniai duomenys. Rodikliai apie ekonomiką ir finansus, populiaciją ir socialines sąlygas, pramonę, prekybą, paslaugas, mokslą, technologijas ir daug kitų sričių. Duomenys nemokami.
2. Ekonominio bendradarbiavimo ir plėtros organizacija – EBPO – (angl. *The Organisation for Economic Co-operation and Development, OECD*). Pateikiami rodikliai apie žemdirbystę, plėtrą, ekonomiką, mokslą, aplinką, finansus, sveikatą ir kitas sritis. Duomenys nemokami.

3. *The Global Economy*. Šioje duomenų bazėje pateikiami virš 300 įvairių sričių rodiklių, kurie yra surinkti iš oficialių šaltinių, tokių kaip Pasaulio bankas (angl. *The World Bank*, Tarptautinis valiutos fondas (angl. *The International Monetary Fund*), Jungtinės Tautos (angl. *the United Nations*), Pasaulio ekonomikos forumas (angl. *the World Economic Forum*). Pateikiama informacija apie daugiau kaip 200 šalių. Duomenys yra mokami, tačiau parsisiunčiamų duomenų kiekis nėra ribojamas. Mėnesio kaina 7,95 USD, metų – 29,95 USD.
4. Pasaulio bankas (angl. *The World Bank*). Šioje duomenų bazėje pateikiami rodikliai įvairiomis temomis. Duomenys yra nemokami.

Aplinkosauginiam konkurencingumui apibūdinti buvo surinkta 18 rodiklių, ekonominį konkurencingumą apibūdinti – 22 rodikliai, šalies aplinkos konkurencingumą apibūdinti – 10 rodiklių, socialinį konkurencingumą apibūdinti – 20 rodiklių, viso surinkta 70 rodiklių. Tam, kad būtų gautas bendras duomenų rinkinys, apjungiant visas šalis ir kintamuosius 2008–2019 m. laikotarpiu, buvo panaudota *Microsoft Excel*. Apjungus duomenis buvo gautas panelinių (angl. *panel*) duomenų rinkinys. Paneliniai duomenys – tai tų pačių vienetų stebėjimai (pavyzdžiui, asmenų, miestų, šalių) ir fiksuoti keliais laiko momentais [73]. Informacija apie šiame darbe naudojamą duomenų rinkinį, tikslus rodiklių pavadinimus, sutrumpintus rodiklių pavadinimus, matavimo vienetus bei jų šaltinį pateikiama 1 priede.

### 2.3. Trūkstumų reikšmių užpildymas

Šiame darbe analizuotas duomenų rinkinys turėjo trūkstumų reikšmių. Ištrinti stebėjimus, kur trūksta bent vienos jų apibūdinančio požymio reikšmės, nebuvo galimybės, nes būtų prarasta daug vertingos informacijos. Ištrynus stebėjimus, kai trūksta bent vienos reikšmės, duomenų rinkinyje būtų likę tik 25 stebėjimai iš 336. Todėl toks stebėjimų pašalinimo būdas buvo atsisakytas. Trūkstumų reikšmių užpildymui buvo nuspręsta panaudoti *missForest* algoritmą, kuris turimų duomenų pagalba užpildo trūkstamas reikšmes. Šis algoritmas buvo pasirinktas todėl, jog yra sąlyginai naujas (2012 m.), neparametrinis ir analizuojant mokslinius straipsnius, pastebėta, jog gali pasiekti didelį tikslumą.

Pasirinktas taikyti trūkstumų reikšmių užpildymo algoritmas *missForest* veikia atsitiktinio miško (angl. *Random Forest*) algoritmo pagrindu, detalesnė informacija apie atsitiktinio miško algoritmą gali būti rasta [74] šaltinyje. Prieš pradėdant trumpai pristatyti atsitiktinį mišką, būtina paminėti sprendimų medžius. Jeigu atsako kintamasis yra kategorinis, tuomet naudojami klasifikavimo medžiai (angl. *classification trees*), jeigu atsako kintamasis yra tolydus, tuomet naudojami regresijos medžiai (angl. *regression trees*). Šiame darbe atsako kintamieji yra tolydūs. Sprendimų medžiai yra sudaryti iš šių dalių: šaknies mazgo (angl. *root node*), kuri yra medžio viršūnė; mazgų (angl. *nodes*); lapų mazgų (angl. *leaf node*), kurie daugiau neskaidomi. Sprendimų medžio tikslas yra padalinti duomenis į mažesnes, homogeniškas grupes (vertinant pagal atsako kintamąjį). Esminis sprendimų medžio uždavinys tampa kintamojo ir slenksčio parinkimas, kuris padeda geriausiai suskaidyti duomenis.

Yra naudojamas kriterijus, kuriuo remiantis yra priimamas sprendimas dėl skaidymo. Jis matuoja mazgo „užterštumą“ (angl. *node impurity*), t. y. vertina atsako kintamojo homogeniškumą mazge. Regresijos medžiams yra naudojamas regresijos nuokrypis (angl. *regression deviance*) ir tai yra apskaičiuojama pagal toliau pateikiamą formulę [75]:

$$D_{\text{node}} = \sum_{i=1}^n (y_i - \bar{y})^2; \quad (1)$$

Čia  $D_{\text{node}}$  – regresijos nuokrypis mazge;

$n$  – stebėjimų skaičius mazge;

$y_i$  – mazgo stebėtos atsako kintamojo reikšmės;

$\bar{y}$  – mazgo atsako kintamojo reikšmių vidurkis.

Tam, kad būtų rastas geriausias kintamasis pirminiam stebėjimų suskaidymui į dvi dalis, yra vertinami nuokrypiai prie skirtingų kintamųjų ir jų slenksčių. Tas kintamasis ir jo slenkstis, su kuriuo nuokrypis yra mažiausias, bus naudojamas šaknies mazge (angl. *root node*). Maža nuokrypio reikšmė parodo mazgo homogeniškumą ir mazgo švarumą (angl. *node purity*). Vėliau procedūra kartojama tęsiant stebėjimų dalinimą iki tol, kol skaidymas nebegalimas. Reikšmė kiekviename atskirame lapo mazge (angl. *leaf node*) yra apskaičiuojama kaip visų stebėjimų, einančių per medį ir patekusių į lapo mazgą, atsako kintamojo vidurkis [76].

Atsitiktinis miškas yra sudarytas iš daug apmokytų sprendimų medžių, kurių kiekvienam sudaryti buvo naudota atsitiktinai atrinkta imtis. Apmokytam atsitiktinio miško modeliui pateikus naują požymių vektorių skirtingi sprendimų medžiai, esantys miške, duoda skirtingą atsako kintamojo reikšmę, o prognozuojama atsitiktinio miško atsako kintamojo reikšmė – tai visų sprendimų medžių priklausomo kintamojo reikšmės vidurkis. Atsitiktinio miško metodas priskiriamas prie mašininio mokymosi *bagging* metodų, kurie geba susidoroti su permokymu [77].

Atsitiktinio miško modelio tikslumas (kiek modelis paaiškina priklausomo kintamojo dispersijos) gali būti vertinamas naudojant imtį, kuri nebuvo naudota sprendimų medžių sudarymui (angl. *out-of-bag sample*). Remiantis  $R$  dokumentacija [78] toks tikslumas (angl. *explained variance*) būtų apskaičiuojamas kaip  $1 - \text{mse} / \text{Var}(y)$ , kur  $\text{mse}$  – vidutinė kvadratinė paklaida, o  $\text{Var}(y)$  – stebėto priklausomo kintamojo  $y$  dispersija.

*MissForest* algoritmo aprašymas pateikiamas naudojant [61] literatūros šaltinį. Šis algoritmas yra mašininio mokymosi pagrįstas duomenų užpildymo algoritmas, kuris veikia atsitiktinio miško (angl. *Random Forest*) algoritmo pagrindu. *MissForest* algoritmo kūrėjai D. J. Stekhoven'as ir P. Bühlmann'as 2011 m. atliko tyrimą, kuriame buvo lyginami duomenų užpildymo algoritmai. Buvo naudojami duomenų rinkiniai, kur ištrintos reikšmės sudarė nuo 10 iki 30 proc. Autoriai įrodė, jog *MissForest* gali sėkmingai užpildyti trūkstamas reikšmes, net ir tuose duomenų rinkiniuose, kuriuose yra įvairių tipų kintamųjų.

Tarkime, jog  $X = (X_1, X_2, \dots, X_p)$  yra  $n \times p$  formato matrica, kur  $n$  – stebėjimų skaičius,  $p$  – kintamųjų skaičius. *MissForest* algoritmas apskaičiuoja trūkstamas reikšmes, naudojant atsitiktinį mišką, apmokytą ant netrūkstamų (angl. *observed*) duomenų rinkinio dalių. Atitinkamam (angl. *arbitrary*) kintamajam  $X_s$ , įskaitant trūkstamas reikšmes prie įrašų  $i_{\text{miss}}^{(s)} \subseteq \{1, \dots, n\}$ , duomenis galima suskirstyti į keturias dalis:

1. netrūkstamas kintamojo  $X_s$  reikšmės, žymimas kaip  $y_{\text{obs}}^{(s)}$ ;
2. trūkstamas kintamojo  $X_s$  reikšmės, žymimas kaip  $y_{\text{miss}}^{(s)}$ ;
3. kitus kintamuosius (be  $X_s$ ) su stebėjimais  $i_{\text{obs}}^{(s)} = \{1, \dots, n\} \setminus i_{\text{miss}}^{(s)}$ , žymimus kaip  $x_{\text{obs}}^{(s)}$ , beje,  $x_{\text{obs}}^{(s)}$  tipiškai nėra visiškai užpildyta;
4. kitus kintamuosius (be  $X_s$ ) su stebėjimais  $i_{\text{miss}}^{(s)}$ , žymimus kaip  $x_{\text{miss}}^{(s)}$ , beje,  $x_{\text{miss}}^{(s)}$  tipiškai nėra visiškai trūkstama.

Pavyzdžiui, turima  $100 \times 5$  formato matrica  $X = (X_1, X_2, X_3, X_4, X_5)$ , kur 100 – stebėjimų skaičius ir 5 – kintamųjų skaičius. Tarkime, jog 5 ir 10 stebėjimai turi trūkstamas reikšmes kintamajame  $X_2$ . Tuomet  $y_{\text{obs}}^{(2)}$  bus visos kintamojo  $X_2$  reikšmės be 5 ir 10 stebėjimo,  $y_{\text{miss}}^{(2)}$  bus kintamojo  $X_2$  5 ir 10 stebėjimų reikšmės,  $x_{\text{obs}}^{(2)}$  bus kintamųjų  $X_1, X_3, X_4, X_5$  reikšmės be 5 ir 10 stebėjimo,  $x_{\text{miss}}^{(2)}$  bus kintamųjų  $X_1, X_3, X_4, X_5$  reikšmės prie 5 ir 10 stebėjimo.

Iš pradžių įvykdomas pradinis trūkstamų  $X$  reikšmių užpildymas, panaudojant vidurkį ar kitą užpildymo metodą. Vėliau rūšiuojami kintamieji  $X_s$ ,  $s = 1, \dots, p$  pagal trūkstamų reikšmių kiekį, pradedant nuo mažiausiai turinčio. Kiekvienam kintamajam  $X_s$  trūkstamos reikšmės yra užpildomos apmokant atsitiktinį mišką su atsaku  $y_{\text{obs}}^{(s)}$  ir kintamaisiais  $x_{\text{obs}}^{(s)}$  ir tuomet prognozuojant trūkstamas reikšmes  $y_{\text{miss}}^{(s)}$ , pritaikius apmokytą atsitiktinį mišką  $x_{\text{miss}}^{(s)}$ . Užpildymo procedūra kartojama tol, kol yra pasiektas stabdymo kriterijus.

**7 lentelė.** *MissForest* algoritmas, [61 p. 133]

<p>Reikalauja: matricos <math>X</math> (formatas <math>n \times p</math>) ir stabdymo kriterijaus <math>\gamma</math>.</p> <ol style="list-style-type: none"> <li>1. Pradinis trūkstamų reikšmių apskaičiavimas.</li> <li>2. <math>k \leftarrow</math> matricos <math>X</math> surūšiuotų stulpelių (pagal didėjantį trūkstamų reikšmių skaičių) indeksų vektorius.</li> <li>3. Kol nepasiektas <math>\gamma</math>, atlikti (angl. <i>while not <math>\gamma</math>, do</i>):             <ol style="list-style-type: none"> <li>4. <math>X_{\text{old}}^{\text{imp}} \leftarrow</math> saugoti anksčiau užpildytą (angl. <i>imputed</i>) matricą;</li> <li>5. kiekvienam <math>s</math>, esančiam <math>k</math>, atlikti (angl. <i>for <math>s</math> in <math>k</math> do</i>):                 <ol style="list-style-type: none"> <li>6. apmokyti (angl. <i>fit</i>) atsitiktinį mišką: <math>y_{\text{obs}}^{(s)} \sim x_{\text{obs}}^{(s)}</math>;</li> <li>7. apmokytą atsitiktinį mišką panaudoti apskaičiuojant <math>y_{\text{miss}}^{(s)}</math> naudojant <math>x_{\text{miss}}^{(s)}</math>;</li> <li>8. <math>X_{\text{new}}^{\text{imp}} \leftarrow</math> atnaujinti matricą, naudojant apskaičiuotas <math>y_{\text{miss}}^{(s)}</math> reikšmes;</li> </ol> </li> <li>9. užbaigti (angl. <i>end for</i>)</li> <li>10. atnaujinti <math>\gamma</math>.</li> </ol> </li> <li>11. Užbaigti ciklą (angl. <i>end while</i>).</li> <li>12. Grąžinti užpildytą matricą <math>X^{\text{imp}}</math>.</li> </ol>
--

Svarbu išskirti analizuoto algoritmo privalumus. *MissForest* algoritmo kūrėjai pažymi, jog jų tikslas buvo pristatyti tokį duomenų užpildymo algoritmą, kuris gali susidoroti su mišraus tipo duomenimis. Kūrėjai pasirinko atsitiktinį mišką, nes jis ne tik susidoroja su įvairaus tipo duomenimis, tai pat atsitiktinis miškas yra itin tikslus ir puikiai veikia esant didelei dimensijai (netgi tada, kai požymių skaičius viršija stebėjimų skaičių) ir netiesiniams ryšiams duomenyse. Galima išskirti dar vieną *MissForest* privalumą – tai neparametrisinis metodas, nereikia tenkinti prielaidų apie duomenų pasiskirstymą. Taip pat galima naudoti kintamuosius, išmatuotus skirtingose skalėse, jų nereikia standartizuoti. Beje, algoritmas leidžia įvertinti duomenų užpildymo kokybę nereikalaujant testavimo imties ar kryžminio patikrinimo. Minėtas algoritmas apskaičiuoja klaidas ant imties (angl. *out-of-bag sample*), kuri nebuvo naudota sprendimų medžių, esančių atsitiktiniame miške, sudarymui.

**2.4. Klasterizavimas**

Prieš atliekant klasterinę analizę svarbu įsitikinti, jog duomenyse yra kažkokia struktūra, kuri galėtų būti atskleista. Tam buvo skaičiuota Hopkins'o statistika. Tarkime turime  $m \times n$  formato matricą  $M$ , kur  $m$  – objektų skaičius ir  $n$  – dimensijų (požymių) skaičius. Hopkins'o statistika parodo, ar  $m$  objektų pasiskirstymas  $n$ -dimensionalumo erdvėje reikšmingai skiriasi nuo atsitiktinio objektų pasiskirstymo toje erdvėje. Iš esmės tai atliekama vertinant du matavimus, kurie pateikiami toliau.

1. Atsitiktinai atrenkami  $k$  ( $k < m$ ) eilučių vektoriai iš matricos  $M$  ir kiekvienam iš tų  $k$  vektorių yra apskaičiuojamas atstumas iki artimiausio kaimyno  $M$  matricoje, šie atstumai yra susumuojami ir ta suma yra lygi  $p$ .
2. Sugeneruojami  $k$  vektoriai, kurie yra atsitiktinai pasiskirstę duomenų erdvėje ir apskaičiuojami artimiausio kaimyno atstumai tarp kiekvieno iš tų  $k$  vektorių ir  $M$  matricos vektorių, jie susumuojami ir šis dydis yra lygus  $q$ .

Tuomet Hopkins'o statistika yra apskaičiuojama kaip:

$$H = \frac{q}{p+q}; \quad (2)$$

Čia  $H$  – Hopkins'o statistika;

$q$  – kiekvieno iš sugeneruotų vektorių mažiausių atstumų (iki  $M$  matricos vektorių) suma

$p$  – kiekvieno iš atsitiktinai atrinkto vektoriaus iš matricos  $M$  mažiausių atstumų (iki  $M$  matricos vektorių) suma

Jeigu duomenų pasiskirstymas nėra atsitiktinis, tuomet  $p$  turėtų būti mažesnis už  $q$ , nes atstumai iki artimiausių kaimynų tarp klasterizuojamų objektų yra mažesni nei tarp atsitiktinai pasiskirsčiusių.  $H$  interpretacija paprasta, jei  $H \approx 0,5$ , tuomet duomenų rinkinys yra atsitiktinis. Kuo  $H$  yra arčiau 1, tuo duomenyse yra didesnė klasterizavimo struktūra. Kuo  $H$  yra arčiau 0, tuo taškai erdvėje tolygiau išsidėstę (t. y. nėra klasterizavimo struktūros) [79].

ES šalių klasterizavimui atlikti buvo pasirinkta naudoti  $k$ -vidurkių ir hierarchinį klasterizavimą. Šie metodai buvo pasirinkti todėl, jog turimas nedidelis klasterizuojamų objektų (šalių) skaičius. Beje, šie metodai yra vieni geriausiai žinomų ir nesudėtingai įgyvendinami praktiškai. Verta paminėti, jog prieš taikant klasterizavimą (tiek  $k$ -vidurkių algoritmo atveju, tiek hierarchinio klasterizavimo atveju), duomenyse neturėtų būti trūkstamų reikšmių bei duomenys turėtų būti normalizuoti. Duomenų normalizavimas yra svarbus todėl, kad požymiai, kurių reikšmės yra didelės, nebūtų įtakingesni už požymius, kurių reikšmės yra mažos, t. y. normalizavimas padeda suteikti visiems požymiams vienodą svorį [68].

$$\hat{v}_i = \frac{v_i - \bar{A}}{\sigma_A}; \quad (3)$$

čia  $\hat{v}_i$  – normalizuota požymio  $A$  reikšmė  $z$ -score būdu;

$v_i$  – požymio  $A$  reikšmė;

$\bar{A}$  – požymio  $A$  vidurkis;

$\sigma_A$  – požymio  $A$  standartinis nuokrypis.

Šiame darbe buvo naudojamas  $z$ -score normalizavimas, kuris apskaičiuojamas remiantis požymio vidurkiu ir standartiniu nuokrypiu (3 formulė).

#### 2.4.1. K-vidurkių algoritmas

$K$ -vidurkių algoritmas bus aprašomas remiantis [67] ir [68] literatūros šaltiniais.  $K$ -vidurkių algoritmas – vienas geriausiai žinomų ir paprasčiausių klasterizavimo algoritmų. Atliekant klasterizavimo uždavinį ir siekiant įvertinti panašumą tarp objektų, buvo naudojamas Euklido atstumas, kuris yra vienas populiariausių artumo matų.



$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}; \quad (4)$$

čia  $d(i, j)$  – Euklido atstumas tarp objektų  $i$  ir  $j$ ;

$x_{i1}, x_{i2}, \dots, x_{ip}$  – objektą  $i$  apibūdinantys požymiai;

$x_{j1}, x_{j2}, \dots, x_{jp}$  – objektą  $j$  apibūdinantys požymiai;

$p$  – objektus apibūdinančių požymių skaičius.

Tarkime turimas duomenų rinkinys  $D$ , kurį sudaro  $n$  objektų. Duomenų rinkinyje  $D$  esantys objektai yra paskirstomi į  $k$  klasterius,  $C_1, \dots, C_k$ , tai yra  $C_i \subset D$  ir  $C_i \cap C_j = \emptyset, 1 \leq i, j \leq k$ . Padalinimo kokybei įvertinti naudojama tikslo funkcija. Siekiama, jog objektai, esantys tame pačiame klasteryje būtų kuo panašesni, o objektai, esantys skirtinguose klasteriuose, kuo nepanašesni. Tai reiškia, jog tikslo funkcija siekia kuo didesnio panašumo klasterio viduje ir mažo tarpklasterinio panašumo. Klasteris yra reprezentuojamas centroidu  $c_i$ , kur centroidas – tai klasteryje esančių objektų požymių vidurkis, tai klasterio centrinis taškas. Skirtumas tarp objekto  $p \in C_i$  ir klasterio centroido  $c_i$  yra matuojamas  $\text{dist}(p, c_i)$ , kur  $\text{dist}(x, y)$  yra Euklido atstumas tarp dviejų taškų  $x$  ir  $y$ . Klasterio  $C_i$  klasterizavimo kokybė gali būti išmatuota remiantis tarpklasterine variacija, kuri yra kvadratinių paklaidų tarp objektų, esančių klasteryje  $C_i$ , ir to klasterio centroido  $c_i$ , suma. Toliau pateikiama tikslo funkcija siekia klasterius padaryti kuo kompaktiškesnius ir kuo labiau atskiriamus.

$$E = \sum_{i=1}^k \sum_{p \in C_i} \text{dist}(p, c_i)^2; \quad (5)$$

čia  $E$  – viso duomenų rinkinio objektų nuokrypių kvadratų suma;

$k$  – klasterių skaičius;

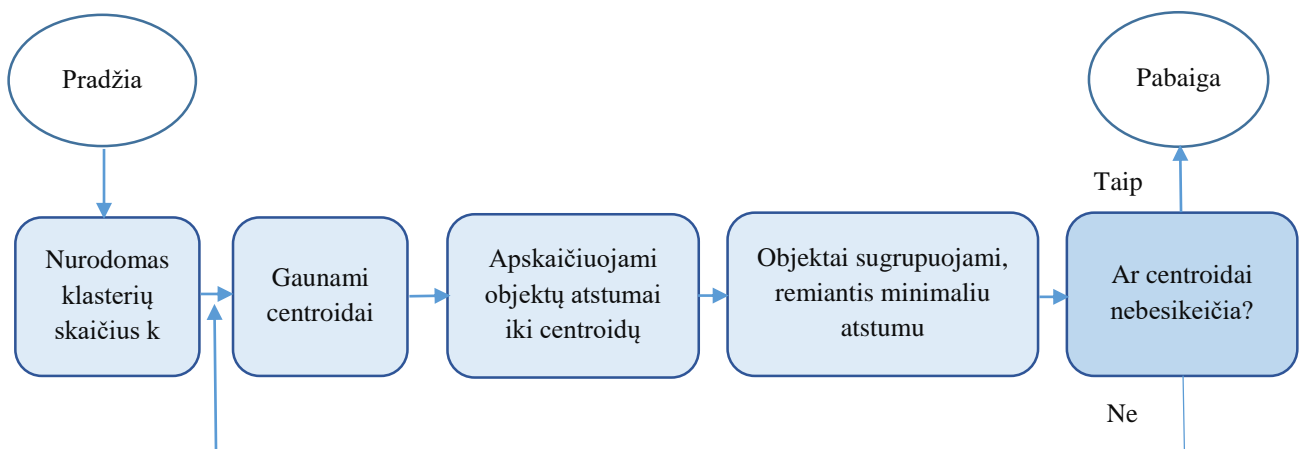
$p$  – taškas duomenų erdvėje, kuris reprezentuoja objektą;

$c_i$  – klasterio centroidas;

$C_i$  – klasteris,

$\text{dist}(p, c_i)$  – atstumas tarp objekto  $p$  ir klasterio centroido  $c_i$ .

Svarbu apibrėžti kaip veikia minėtas algoritmas (žr. 5 pav.). Pirmiausiai tyrėjas nurodo klasterių skaičių  $k$ . Atsitiktinai parenkami  $k$  objektai, esantys duomenų rinkinyje  $D$ , kurių kiekvienas reprezentuoja pradinį klasterio centrą.



5 pav. Struktūrinė k-vidurkių algoritmo diagrama, [67]

Tuomet kiekvienas iš likusių objektų yra priskiriamas atitinkamam klasteriui, remiantis Euklido atstumu tarp objekto ir pradinio klasterio centro. K-vidurkių algoritmas kiekvienam klasteriui apskaičiuoja naują, tame klasteryje esančių objektų, požymių vidurkį. Tuomet visi objektai yra perskirstomi atsižvelgiant į atnaujintus klasterio centrus. Iteracijos kartojasi tol, kol klasterio centroidai nebesikeičia, t. y. objektų priskyrimas atitinkamiems klasteriams lieka pastovus. Struktūrinė k-vidurkių algoritmo diagramą galima matyti 5 pav.

Apibendrinant k-vidurkių algoritmą būtų verta paminėti, jog šis algoritmas yra nesudėtingas, reikalauja santykinai mažų laiko sąnaudų, jo vykdymo laikas proporcingas objektų skaičiui  $n$ . Priešingai nei hierarchiniame klasterizavime, klasteriai gali pasikeisti po jų suformavimo. Kaip vienas šio metodo trūkumų gali būti išskiriama tai, jog reikia iš anksto nurodyti klasterių skaičių  $k$ . Taip pat k-vidurkių algoritmo trūkumai – jis yra tinkamas atrasti sferines formas (angl. *nonconvex*) bei yra jautrus išskirtims, nes išskirtys gali reikšmingai paveikti klasterio centro reikšmę.

#### 2.4.2. Hierarchinis klasterizavimas

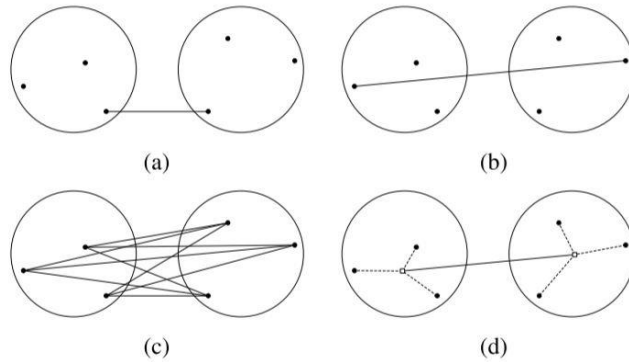
Pagrindinė hierarchinių klasterizavimo algoritmų idėja yra suklastezuoti objektus, remiantis hierarchiniu principu. Galimi du hierarchinio klasterizavimo variantai – jungimo (angl. *agglomerative*) arba skaidymo (angl. *divisive*). Jungimo metodo pradžioje laikoma, kad kiekvienas objektas yra atskiras klasteris, vėliau jie yra apjungiami į vis didesnius klasterius, kol gaunamas vienas klasteris, o skaidymo metodo pradžioje turimas vienas didelis klasteris, kuris yra skaidomas į vis mažesnius klasterius iki tol, kol kiekvienas objektas tampa klasteriu [67]. Šiame darbe buvo naudojamas jungimo metodas, todėl detaliau skaidymo metodas nebus analizuojamas.

Hierarchinis jungimo algoritmas yra vykdomas sekančiai [70]:

- 1 žingsnis: kiekvienas objektas yra laikomas pradinio klasteriu.
- 2 žingsnis: apskaičiuojami atstumai tarp klasterių.
- 3 žingsnis: du klasteriai, kurie yra arčiausiai, apjungiami ir tampa vienu klasteriu. Perskaičiuojama artumų matrica tarp klasterių tam, kad būtų atspindėtas šis susijungimo procesas.
- 4 žingsnis. Kartojamas 2 ir 3 žingsniai kol gaunamas vienas klasteris, kuriame yra visi objektai.

Hierarchinio klasterizavimo išvestis – dendograma, kuri yra dvimatė medžio pavidalo struktūra, vaizduojanti klasterių hierarciją. Kiekvieno jungimo atstumas taip pat atvaizduojamas dendogramoje. Panašumui tarp objektų nustatyti yra naudojamas artumo matas. Po to, kai pasirinktas artumo matas, reikia pasirinkti susiejimo (angl. *linkage*) kriterijų, kuris naudojamas klasterių sujungimui. Naudojami vienetinės jungties (angl. *single link*), pilnosios jungties (angl. *complete link*), vidutinės jungties (angl. *average link*), centroidų (angl. *centroid method*) ir Vordo (angl. *Ward's method*) metodai [70]. Jungimo metodų pavyzdžiai pateikiami 6 pav.

**Vienetinės jungties metodas.** Šis metodas dar vadinamas minimalaus atstumo, artimiausio kaimyno metodu. Naudojant šį metodą, atstumas tarp klasterių apibrėžiamas kaip atstumas tarp dviejų artimiausių objektų, priklausančių skirtingiems klasteriams [67]. Toks susiejimo metodas tinkamas aptikti ne elipsės formas, yra jautrus išskirtims [70]. Šis būdas mažiau tinka sferinių klasterių aptikimui, tačiau šis metodas gan plačiai naudojamas, nes yra efektyvus aptikti ištęstus, netaisyklingos formos, vingiuotus klasterius [66].



**6 pav.** Jungimo metodai: a) vienetinės jungties, b) pilnosios jungties, c) vidutinės jungties, d) centroidų, [80 p. 5596].

**Pilnosios jungties metodas.** Šis metodas dar vadinamas maksimalaus atstumo, tolimiausio kaimyno metodu. Jis yra priešingas vienetinės jungties metodui, nes naudojant pilnosios jungties metodą, atstumas tarp klasterių apibrėžiamas kaip atstumas tarp dviejų tolimiausių objektų, priklausančių skirtingiems klasteriams. Šis metodas yra mažiau jautrus išskirtims [70].

**Vidutinės jungties metodas.** Tai metodas, kuris yra tarpinis tarp vienetinės ir pilnosios jungties. Naudojant šį metodą atstumas tarp klasterių yra atstumo vidurkis tarp visų objektų, priklausančių skirtingiems klasteriams, porų. Šis metodas yra geresnis ir už vienetinės jungties, ir už pilnosios jungties, nes jis leidžia minimizuoti dispersiją klasterio viduje ir maksimizuoti dispersiją tarp klasterių [70].

**Centroidų metodas.** Atstumas tarp klasterių apibrėžiamas kaip atstumas tarp klasterio centroidų. Pridėjus, ar pašalinus objektą, centroidas yra perskaičiuojamas. Šis metodas yra labiau atsparus išskirtims ir geriau nei kiti metodai tvarkosi su skirtingų dydžių klasteriais [70].

**Vordo metodas.** Tai metodas, kuris labiausiai atspindi klasterizavimo principą, yra efektyvus, nes remiasi dispersijos minimizavimu. Tai dažnai naudojamas hierarchinio klasterizavimo metodas. Jis beveik neefektyvus būdas pailgiems klasteriams (jiems geriau tinka vienetinės jungties) ir labai jautrus išskirtims, bet vis dėl to tai vienas populiariausių metodų, nes jis efektyvus taikant realiose problemose [66]. Vordo atstumas tarp dviejų klasterių A ir B, apibrėžiamas pagal 6 formulę.

$$d(A, B) = \frac{d(a, b)^2}{n_A^{-1} + n_B^{-1}}; \quad (6)$$

čia  $d(A, B)$  – Vordo atstumas tarp A ir B klasterių;

$d(a, b)$  – atstumas tarp klasterių centrų a ir b;

$n_A$  – objektų skaičius A klasteryje;

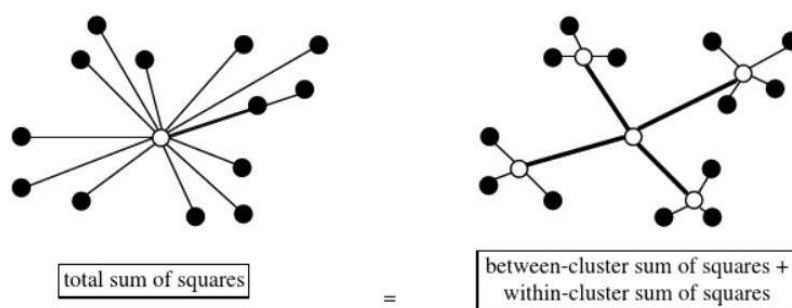
$n_B$  – objektų skaičius B klasteryje.

Atliekant šį tyrimą buvo naudojamas Vordo metodas. Apibendrinant hierarchinį klasterizavimą galima teigti, jog tokiu būdu klasterizuojami objektai yra hierarchinė dekompozicija. Šis būdas leidžia aptikti skirtingų formų klasterius pagal pasirinktą susiejimo būdą. Padalijimo kriterijus nėra globalus, du objektai, priskirti skirtingiems klasteriams, daugiau nebebus lyginami (priešingai nei k-

vidurkių atveju). Šio metodo trūkumai – palyginti didelės laiko sąnaudos, reikia pasirinkti klasterių skaičių  $k$ .

### 2.4.3. Klasterių skaičiaus nustatymas ir klasterizavimo kokybės vertinimas

Prieš pradėdant pristatyti šio skyrelio metodus, būtina apibūdinti kvadratines nuokrypių sumas. Jos yra trys: klasterio kvadratinių nuokrypių suma (angl. *within-cluster sum of squares* - *WSS*), kuri yra apskaičiuojama kaip kiekvieno to klasterio objekto nuokrypio nuo klasterio centro kvadratų suma; tarpklasterinė kvadratinių nuokrypių suma (angl. *between-cluster sum of squares* - *BSS*), kuri yra apibrėžiama kaip kiekvieno klasterio centro atstumo nuo globalaus centro kvadratų suma; ir bendra kvadratinių nuokrypių suma (angl. *total sum of squares* - *TSS*), kuri yra lygi anksčiau minėtų nuokrypių sumai. Toliau pateikiamas šias kvadratinių nuokrypių sumas iliustruojantis paveikslas (žr. 7 pav.).



7 pav. Kvadratinių nuokrypių sumos, [66 p. 243]

Vienas svarbiausių sprendimų atliekant klasterinę analizę – klasterių skaičiaus pasirinkimas, kas neabejotinai turi neigiamą įtaką klasterizavimo kokybei, jei pasirinktas klasterių skaičius neatspindi tikrosios duomenų struktūros. Reikalaujama, jog klasterizavimo metu gauti klasteriai būtų lengvai interpretuojami nei kad būtų teoriškai idealūs [66]. Literatūroje aptariami įvairiausi indeksai optimaliam klasterių skaičiaus nustatymui. *R* pakete *NbClust* įgyvendinti 30 indeksų, kurie skirti klasterių skaičiaus nustatymui. Šį paketą galima naudoti su  $k$ -vidurkių ir hierarchiniu klasterizavimu. *NbClust* užduodama funkcija, kuri atlieka  $k$ -vidurkių ir hierarchinį klasterizavimą su skirtingais atstumo ir jungimo būdais bei pageidaujama minimaliu ir maksimaliu klasterių skaičiumi. Šis paketas leidžia tyrėjui įvertinti duomenų rinkinio klasterizavimo kokybę pagal 30 indeksų prie skirtingo klasterių skaičiaus (intervale nuo minimalaus iki maksimalaus nurodyto klasterių skaičiaus) ir naudotojas gali pasirinkti, kuris klasterių skaičius yra geriausias. Visi šie klasterizavimo vertinimo indeksai apima informaciją apie klasterio kompaktiškumą ir tarpklasterinį atskyrimą, taip pat apima geometrines ir statistines duomenų ypatybes, duomenų objektų skaičių bei nepanašumo ar panašumo matavimus [81].

Šiame darbe pasirenkant optimalų klasterių skaičių bei vertinant klasterizavimo kokybę buvo atsižvelgiama į **Silueto** (angl. *Silhouette*) **indeksą**. Šis matas parodo, kiek klasteris yra kompaktiškas ir kaip gerai atsiskyręs nuo kitų klasterių. Tarkime turimas duomenų rinkinys  $D$ , sudarytas iš  $n$  objektų.  $D$  yra padalinama į  $k$  klasterius  $C_1, \dots, C_k$ . Kiekvienam objektui  $o \in D$  yra apskaičiuojamas  $a(o)$  kaip vidutinis atstumas tarp  $o$  ir kitų to klasterio, kuriam priklauso  $o$ , objektų. Analogiškai, yra apskaičiuojamas ir  $b(o)$  kaip minimalus vidutinis atstumas nuo  $o$  iki visų klasterių, kuriems  $o$  nepriklauso, objektų. Tuomet Silueto indeksas apskaičiuojamas pagal formulę:

$$s(o) = \frac{b(o)-a(o)}{\max\{a(o),b(o)\}}; \quad (7)$$

čia  $s(o)$  – objekto  $o$  Silueto koeficientas;

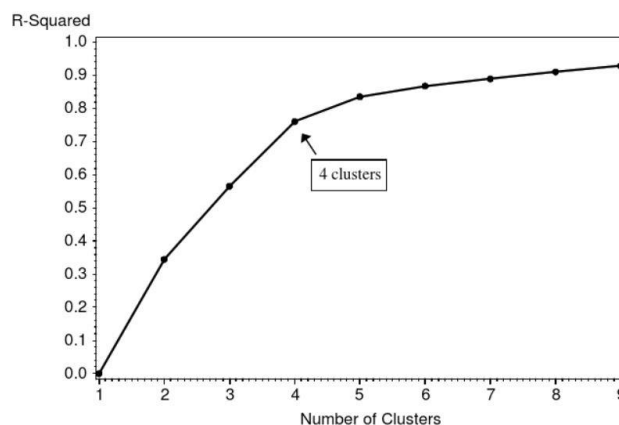
$a(o)$  – vidutinis atstumas tarp objekto  $o$  ir kitų klasterio, kuriam priklauso  $o$ , objektų;

$b(o)$  – minimalus vidutinis atstumas tarp  $o$  ir kitų klasterių, kuriems  $o$  nepriklauso, objektų.

Silueto indeksas gali kisti nuo -1 iki 1. Klasterio kompaktiškumą atspindi  $a(o)$  reikšmė, kuo ji mažesnė, tuo klasteris kompaktiškesnis. Kaip  $o$  atsiskyręs nuo kitų klasterių parodo  $b(o)$  reikšmė, kuo ji didesnė, tuo  $o$  yra labiau atsiskyręs nuo kitų klasterių. Kai Silueto koeficientas artėja prie 1, klasteris, kuriame yra  $o$ , yra kompaktiškas ir  $o$  yra toli nuo kitų klasterių. Kai Silueto koeficientas yra neigiamas (t. y.  $b(o) < a(o)$ ), tai reiškia, jog  $o$  yra arčiau kitų klasterių objektų. Norint įvertinti klasterio kokybę, galima apskaičiuoti vidutinį Silueto koeficientą visiems objektams, esantiems tame klasteryje. Analogiškai, siekiant išmatuoti viso klasterizavimo kokybę, skaičiuojamas vidutinis Silueto indeksas visiems suklasterizuotiems objektams [68].

**Alkūnės metodas** (angl. *elbow method*) gali padėti apsispręsti pasirenkant teisingą klasterių skaičių. Galima įvykdyti klasterizavimą nurodant skirtingą klasterių skaičių (pavyzdžiui, nuo 1 iki 10) ir apskaičiuoti bendrą visų klasterių vidaus (angl. *within-cluster*) kvadratinių nuokrypių sumą prie kiekvieno klasterio skaičiaus. Tuomet grafike atvaizduojama ši suma prie skirtingų  $k$  reikšmių. Pirmasis ar reikšmingiausias kreivės pasikeitimas (alkūnė) laikomas teisingu klasterių skaičiumi [68].

Dar vienas būdas klasterių skaičiaus nustatymui yra grafikas, kuriame vaizduojamas  $R^2$  prie skirtingo klasterių skaičiaus.  $R^2$  – tai kvadratų sumos dalis, paaiškinama klasteriais (kvadratų suma tarp klasterių / visa kvadratų suma), t. y. kiek sudaryti klasteriai paaiškina dispersijos. Kuo  $R^2$  yra arčiau 1, tuo klasterizavimas yra geresnis, nes tai reiškia, jog kvadratų suma tarp klasterių yra didelė, o kvadratų suma klasterių viduje yra maža. Tačiau vis tik tikslas neturėtų būti maksimizuoti  $R^2$ , nes didesnis  $R^2$  reikš ir didesnį klasterių skaičių. Tikslas yra turėti tokį  $R^2$ , kuris yra artimas 1, tačiau su ne per daug klasterių. Galioja taisyklė: jei paskutinis žymus  $R^2$  padidėjimas įvyksta tada, kai pereinama nuo  $k$  prie  $k+1$  klasterių, tuomet padalinimas į  $k+1$  klasterius yra korektiškas [66]. Toliau pateikiamas tokio grafiko pavyzdys, kuriame matyti, jog atsižvelgiant į  $R^2$  optimaliausias klasterių skaičius yra 4 (žr. 8 pav.).



**8 pav.** Klasterių skaičiaus parinkimas pagal  $R^2$ , [66 p. 245]

Gali būti atliekamas ir intuityvus klasterizavimo kokybės vertinimas pasitelkus **pagrindines komponentes**. Objektai gali būti atvaizduojami plokštumoje pagal pagrindines komponentes ir duomenų taškai gali būti nuspalvinami pagal klasterius [66]. Tuomet vizualiai galima matyti, ar aiškiai atsiskiria klasteriai. Pagrindinių komponentių analizė (angl. *principal components analysis* – *PCA*) – tai būdas sumažinti duomenų dimensionalumą. Sukuriami nauji nekoreliuoti kintamieji (pagrindinės komponentės), kurie maksimizuoja duomenų dispersiją [82]. Pagrindinių komponentių analizė siekia paaiškinti kintamųjų koreliacijos struktūrą naudodama mažesnę tų kintamųjų tiesinių derinių rinkinį (pagrindines komponentes) [83]. Didžiausią dalį dispersijos nusako pirmoji pagrindinė komponentė, tada antroji pagrindinė komponentė ir t. t. Svarbu paminėti, jog antroji pagrindinė komponentė yra ortogonalinė pirmajai. Trečioji pagrindinė komponentė yra ortogonalinė pirmoms dviem pagrindinėms komponentėms ir t. t. Toliau bus pateikiamas detalesnis principinių komponentių apibūdinimas remiantis [83] šaltiniu. Tarkime turime  $X_1, X_2, \dots, X_m$  kintamuosius. Prieš atliekant dimensionalumo mažinimą, pirmiausia duomenis reikia standartizuoti, kad kiekvieno kintamojo vidurkis būtų lygus 0 ir standartinis nuokrypis būtų lygus 1. Tarkime, jog kiekvienas kintamasis  $X_i$  yra  $n \times 1$  vektorius, kur  $n$  – įrašų skaičius, tuomet standartizuotas kintamojo vektorius bus apskaičiuojamas pagal toliau pateikiamą formulę:

$$Z_i = \frac{(X_i - \mu_i)}{\sigma_{ii}}, \quad (8)$$

čia  $Z_i$  – standartizuotas kintamojo  $X_i$  vektorius;

$X_i$  – kintamojo vektorius;

$\mu_i$  – kintamojo  $X_i$  vidurkis;

$\sigma_{ii}$  – kintamojo  $X_i$  standartinis nuokrypis.

Standartizuota duomenų matrica  $Z = (V^{1/2})^{-1}(X - \mu)$ , kur laipsnio rodiklis „-1“ žymi atvirkštinę matricą, o  $V^{1/2}$  yra diagonalioji matrica (nenuliniai elementai yra tik matricos įstrižainėje). Tuomet  $m \times m$  formato standartinių nuokrypių matrica  $V^{1/2}$  atrodytų taip [83]:

$$V^{1/2} = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{mm} \end{bmatrix}$$

Kovariacija  $\sigma_{ij}^2$  (čia  $i \neq j$ ) tarp kintamųjų  $X_i$  ir  $X_j$  apskaičiuojama pagal toliau esančią formulę:

$$\sigma_{ij}^2 = \frac{\sum_{k=1}^n (x_{k1} - \mu_i)(x_{kj} - \mu_j)}{n}, \quad (9)$$

čia  $\sigma_{ij}^2$  – kovariacija tarp kintamųjų  $X_i$  ir  $X_j$ ;

$n$  – stebėjimų skaičius;

$\mu_i$  – kintamojo  $X_i$  vidurkis;

$\mu_j$  – kintamojo  $X_j$  vidurkis.

Tegul  $\Sigma$  bus simetrinė kovariacijų (dispersijų-kovariacijų) matrica, kur  $\sigma_{ij}^2$  ( $i \neq j$ ) yra kovariacija tarp  $X_i$  ir  $X_j$ , o  $\sigma_{ii}^2$  bus kintamojo  $X_i$  dispersija, tuomet  $\Sigma$  [83]:

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1m}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 & \cdots & \sigma_{2m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1m}^2 & \sigma_{2m}^2 & \cdots & \sigma_{mm}^2 \end{bmatrix}$$

Svarbu atkreipti dėmesį, jog pakeitus matavimo vienetus pasikeistų ir kovariacijos reikšmės. Koreliacijos koeficientas  $r_{ij}$  padeda išvengti šio trūkumo ir yra apskaičiuojamas pagal toliau pateikiamą formulę:

$$r_{ij} = \frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}}; \quad (10)$$

čia  $r_{ij}$  – koreliacijos koeficientas tarp kintamųjų  $X_i$  ir  $X_j$ ;

$\sigma_{ij}^2$  – kovariacija tarp kintamųjų  $X_i$  ir  $X_j$ ;

$\sigma_{ii}$  – kintamojo  $X_i$  standartinis nuokrypis;

$\sigma_{ij}$  – kintamojo  $X_j$  standartinis nuokrypis.

Tuomet koreliacijų matrica  $\rho$  atrodytų taip [83]:

$$\rho = \begin{bmatrix} \frac{\sigma_{11}^2}{\sigma_{11}\sigma_{11}} & \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} & \cdots & \frac{\sigma_{1m}^2}{\sigma_{11}\sigma_{mm}} \\ \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} & \frac{\sigma_{22}^2}{\sigma_{22}\sigma_{22}} & \cdots & \frac{\sigma_{2m}^2}{\sigma_{22}\sigma_{mm}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1m}^2}{\sigma_{11}\sigma_{mm}} & \frac{\sigma_{2m}^2}{\sigma_{22}\sigma_{mm}} & \cdots & \frac{\sigma_{mm}^2}{\sigma_{mm}\sigma_{mm}} \end{bmatrix}$$

Standartizuotų duomenų matrica yra  $Z = (V^{1/2})^{-1}(X-\mu)$ . Kai kiekvienas kintamasis yra standartizuotas, tuomet  $E(Z) = 0$ , kur 0 žymi  $n \times m$  formato matricą, sudarytą iš 0, o  $Z$  kovariacijų matrica yra  $\text{Cov}(Z) = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1} = \rho$ . Standartizuotų duomenų kovariacijų ir koreliacijų matricos yra lygios. Standartizuotų duomenų matricos  $Z = [Z_1, Z_2, \dots, Z_m]$  i-tąją pagrindinę komponentę galima būtų užrašyti kaip  $Y_i = \acute{e}_i Z$ , kur  $e_i$  yra i-tasis tikrinis vektorius, o  $\acute{e}_i$  yra transponuotas  $e_i$  vektorius. Pagrindinės komponentės  $Y_1, Y_2, \dots, Y_k$  yra tiesinės standartizuotų kintamųjų  $Z$  kombinacijos taip, jog  $Y_i$  dispersija yra kaip įmanoma maksimalesnė ir  $Y_i$  tarpusavyje yra nekoreliuotos. Pirmoji pagrindinė komponentė yra tiesinė kombinacija  $Y_1 = \acute{e}_1 Z = e_{11}Z_1 + e_{21}Z_2 + \dots + e_{m1}Z_m$ , kuri nusako didžiausią  $Z$  kintamųjų sklaidą, tad:

- pirmoji pagrindinė komponentė yra tiesinė kombinacija  $Y_1 = \acute{e}_1 Z$ , kuri maksimuoja  $\text{Var}(Y_1) = \acute{e}_1 \rho e_1$ ;
- antroji pagrindinė komponentė yra tiesinė kombinacija  $Y_2 = \acute{e}_2 Z$  ir kuri yra nepriklausoma nuo  $Y_1$  ir maksimuoja  $\text{Var}(Y_2) = \acute{e}_2 \rho e_2$ ;
- i-oji pagrindinė komponentė yra tiesinė kombinacija  $Y_i = \acute{e}_i Z$ , kuri nepriklauso nuo visų kitų pagrindinių komponentių  $Y_j, j < i$ , ir maksimuoja  $\text{Var}(Y_i) = \acute{e}_i \rho e_i$ .

Tegul  $B$  bus  $m \times m$  formato matrica, o  $I$  – vienetinė matrica (diagonalioji, kurios įstrižainėje 1). Tuomet skaliarai  $\lambda_1, \lambda_2, \dots, \lambda_m$  yra tikrinės matricos  $B$  reikšmės, jeigu tenkinama sąlyga  $|B - \lambda I| = 0$ ,

kur  $|Q|$  žymi  $Q$  determinantą. Tegul  $B$  bus  $m \times m$  formato matrica, o  $\lambda$  – matricos  $B$  tikrinė reikšmė. Tuomet nenulinis  $m \times 1$  vektorius  $e$  yra matricos  $B$  tikrinis vektorius, jei  $Be = \lambda e$ .

Visa standartizuotų duomenų sklaida yra lygi  $Z$ -vektorių dispersijų sumai, kuri yra lygi pagrindinių komponentių dispersijų sumai, kuri yra lygi tikrinių reikšmių sumai, kuri yra lygi kintamųjų skaičiui t. y. [83]:

$$\sum_{i=1}^m Var(Z_i) = \sum_{i=1}^m Var(Y_i) = \sum_{i=1}^m \lambda_i = m$$

Dalinė koreliacija tarp pagrindinės komponentės ir standartizuoto kintamojo yra tikrinio vektoriaus ir tikrinės reikšmės funkcija, t. y.  $Corr(Y_i, Z_j) = e_{ij} \sqrt{\lambda_i}$ , kur  $i, j = 1, 2, \dots, m$  ir  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_m, e_m)$  yra tikrinės reikšmės ir tikrinio vektoriaus poros koreliacijų matricai  $\rho$  ir  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_m$  [83].

Standartizuotų duomenų  $Z$  dispersijos dalis, kuri yra paaiškinama  $i$ -tają pagrindine komponente, yra  $i$ -tosios tikrinės reikšmės ir kintamųjų skaičiaus santykis, t. y.  $\frac{\lambda_i}{m}$ .

## 2.5. Kintamųjų atranka, algoritmas VSURF

Šiame skyrelyje bus aprašomas algoritmas VSURF (angl. *Variable Selection Using Random Forests*) remiantis šaltiniu [84]. Kaip galima suprasti iš pačio pavadinimo, šis metodas atranka svarbiausius kintamuosius remiantis atsitiktiniu mišku. Algoritmas turi du tikslus – interpretavimo (angl. *interpretation*) ir prognozės (angl. *prediction*). Pirmasis tikslas reiškia svarbių kintamųjų (gali būti ir perteklinių), kurie labai susiję su atsako kintamuoju, radimą. Antrasis tikslas yra rasti mažesnę rinkinį kintamųjų (vengiant perteklinių), kurių pakanka gana tiksliai prognozuoti atsako kintamąjį. Algoritmo principas yra toks, kad pirmiausia kintamieji yra ranguojami atsižvelgiant į kintamųjų svarbą (angl. *variable importance - VI*) ir nesvarbūs yra pašalinami.

VSURF remiasi kintamojo svarbos (VI) matu, tad būtina jį apibrėžti. Toliau pateikiama kintamojo  $X^j$  svarbos skaičiavimo formulė.

$$VI(X^j) = \frac{1}{ntree} \sum_t (\widetilde{errOOB}_t^j - errOOB_t); \quad (11)$$

čia  $VI(X^j)$  – kintamojo  $X^j$  svarba;

$ntree$  – atsitiktinio miško medžių skaičius;

$t$  – medis;

$\widetilde{errOOB}_t^j$  – paklaida (MSE), gauta ant duomenų, kurie nebuvo naudojami sudarant medį  $t$  ir kuriuose sukeista kintamojo  $X^j$  reikšmių tvarka;

$errOOB_t$  – paklaida (MSE), gauta ant duomenų, kurie nebuvo naudojami sudarant medį  $t$ .

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2; \quad (12)$$

čia  $MSE$  – vidutinė kvadratinė paklaida;

$n$  – stebėjimų skaičius;



$y_i$  – stebėta reikšmė;

$\hat{y}_i$  – prognozuota reikšmė.

Reikia atkreipti dėmesį, jog šiame algoritme naudojamas būtent šis kintamųjų svarbos (VI) apibrėžimas. Algoritmas yra apibūdinamas dviejų žingsnių procedūra. Atkreiptinas dėmesys, jog paprastai atsitiktinis miškas yra sudaromas iš 2000 medžių.

1 žingsnis. Pradinis kintamųjų šalinimas ir reitingavimas:

- kintamieji vertinami pagal VI (paprastai gavus vidutinę reikšmę iš 50 atsitiktinių miškų) ir išrikiuojami mažėjimo tvarka;
- pašalinami mažiausios svarbos kintamieji. Atsižvelgiama į atitinkamų VI standartinius nuokrypius, kurie panaudojami nustatant kintamųjų šalinimo slenkstį. Svarbių kintamųjų (angl. *true variables*) VI kintamumas (angl. *variability*) yra didesnis lyginant su nenaudingų kintamųjų VI kintamumu, todėl slenkstis yra nustatomas atsižvelgiant į nenaudingų kintamųjų VI standartinį nuokrypį.

2 žingsnis. Kintamųjų atranka.

- Interpretavimui. Sudaromas atsitiktinių miškų modelių rinkinys įtraukiant  $k$  pirmuosius kintamuosius, kai  $k=1$  iki  $m$ , kur  $m$  – kintamųjų skaičius likęs po pradinio kintamųjų šalinimo. Yra apskaičiuojamos atsitiktinių miškų klaidos (angl. *out-of-bag errors - OOB*) ant imties, kuri nebuvo naudota medžių sudarymui (paprastai suvidurkintos per 25 paleidimus) pradėdant nuo modelio, kuriame yra tik vienas svarbiausias kintamasis ir baigiant modeliu, kuriame yra visi svarbūs kintamieji. Pasirenkamas tas modelis, kuris duoda mažiausią klaidą ir kintamieji, esantys šiame modelyje, yra vadinami interpretavimo rinkiniu.
- Prognozei. Šio etapo esmė yra pašalinti perteklinius kintamuosius, gautus interpretavimo etape, ir turėti mažesnę kintamųjų rinkinį, kuris vadinamas prognozės rinkiniu. Tai pažingsninis būdas, kai kiekviename žingsnyje yra patikrinama, ar naujai įtrauktas į modelį kintamasis padeda pakankamai sumažinti klaidų lygį. Procedūra remiasi išrikiuotais kintamaisiais, gautais interpretavimo etape, ir su jais sudaromi atsitiktinio miško modeliai didėjančia tvarka. Kintamojo įtraukimas yra paremtas testu – kintamasis yra įtraukiamas tik tuo atveju, jei klaidos sumažėjimas yra didesnis nei slenkstis.

## 2.6. Kendall'o konkordancijos koeficientas

Aprašant Kendall'o konkordancijos koeficientą buvo remiamasi [85] ir [86] šaltiniais. Kendall'o konkordancijos koeficientas  $W$  (angl. *Kendall's coefficient of concordance*) yra naudojamas įvertinti ekspertų vertinimų sutapimą. Tam, kad būtų apskaičiuotas šis koeficientas, ekspertų vertinimai turi būti suranguoti. Tarkime  $p$  ekspertų įvertino  $n$  objektų ir vertinami objektai pateikiami eilutėse, o ekspertų įvertinimai yra stulpeliuose. Kiekviename stulpelyje esančios ekspertų reikšmės turi būti pakeistos rangais. Tai reiškia, jog geriausiam objektui (konkrečiu atveju stipriausiai šaliai) suteikiamas rangas 1, antram pagal gerumą – rangas 2 ir t. t., o paskutiniam pagal gerumą – rangas  $n$ , kur  $n$  – lyginamų objektų skaičius. Koeficientas  $W$  gali įgyti reikšmes nuo 0 (vertinimai tarp ekspertų visiškai nesutampa) iki 1 (vertinimai tarp ekspertų visiškai sutampa), t. y.  $0 \leq W \leq 1$ . Tarpinės reikšmės parodo didesnę (artėjant prie 1) ar mažesnę (artėjant prie 0) ekspertų vertinimo sutapimo lygį. Kendall'o konkordancijos hipotezės [85]:

$H_0$ : nėra sutarimo tarp ekspertų (t. y. Kendall'o konkordancijos koeficientas  $W=0$ );

$H_A$ : yra sutarimas tarp ekspertų (t. y. Kendall'o konkordancijos koeficientas  $W>0$ ).

Kendall'o konkordancijos koeficientas yra apskaičiuojamas pagal toliau pateikiamą formulę:

$$W = \frac{12 \sum_{i=1}^n (R_i - \bar{R})^2}{p^2(n^3 - n)}; \quad (13)$$

čia  $W$  – Kendall'o konkordancijos koeficientas;

$i$  – ranguojamas objektas ( $i = 1, \dots, n$ );

$R_i$  – objekto  $i$  rangų suma, suteikta  $p$  teisėjų:  $R_i = \sum_{j=1}^p R_{ij}$ ;

$\bar{R}$  – vidutinė rangų reikšmė:  $\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i$ ;

$p$  – ekspertų skaičius;

$n$  – objektų skaičius.

Taip pat svarbu įvertinti konkordancijos koeficiento reikšmingumą. Pagal nulinę hipotezę,  $C$ , kur  $C = p(n-1)W$  (čia  $p$  – ekspertų skaičius,  $n$  – objektų skaičius,  $W$  – Kendall'o konkordancijos koeficientas), yra pasiskirstęs pagal  $\chi^2$  skirstinį su  $n-1$  laisvės laipsnių [86]. Nulinė hipotezė (nėra sutarimo tarp ekspertų) yra atmetama, kai  $C \geq \chi^2_{n-1, 1-\alpha}$ , kur  $\chi^2_{n-1, 1-\alpha}$  žymi  $\chi^2$  skirstinio su  $n-1$  laisvės laipsnių  $(1-\alpha)\%$  procentilį [86].

## 2.7. Indeksų skaičiavimas

Skaičiuojant indeksus buvo remtasi Kasztelan'o [54] tyrime naudota metodika. Kintamieji yra padalinami į dvi grupes. Pirmai grupei priklauso kintamieji, kurių padidėjimas lemia konkurencingumo padidėjimą. Šių kintamųjų normalizuotos reikšmės yra apskaičiuojamos pagal formulę:

$$Z = \frac{(x - x_{min})}{(x_{max} - x_{min})}; \quad (14)$$

čia  $z$  – normalizuotas kintamasis;

$x$  – kintamasis prieš normalizavimą;

$x_{min}$  – minimali kintamojo reikšmė duotame rinkinyje;

$x_{max}$  – maksimali kintamojo reikšmė duotame rinkinyje.

Antrai grupei priklauso kintamieji, kurių padidėjimas lemia konkurencingumo sumažėjimą. Šių kintamųjų normalizuotos reikšmės yra apskaičiuojamos pagal formulę:

$$Z = \frac{(x_{max} - x)}{(x_{max} - x_{min})}; \quad (15)$$

čia  $z$  – normalizuotas kintamasis;

$x$  – kintamasis prieš normalizavimą;

$x_{min}$  – minimali kintamojo reikšmė duotame rinkinyje;

$x_{max}$  – maksimali kintamojo reikšmė duotame rinkinyje.

Šiame tyrime naudoti kintamieji, kurių padidėjimas lemia mažesnę konkurencingumą – *ammon\_emis\_agr*, *greenhouse\_gas*, *munic\_waste\_kg\_pc*, *start\_business\_days*, *cost\_start\_business*,

*infl\_cpi*, *shadow\_ec\_gdp*, *prisoners\_nbr*, *thefts*, *homicides*, *matt\_depr\_perc*. Kintamieji, normalizuoti anksčiau nurodytu būdu (t. y. remiantis 14 ir 15 formule), įgyja reikšmę intervale [0, 1]. Kuo reikšmė artimesnė 1, tuo situacija, atsižvelgiant į tiriamą požymį, yra geresnė. Kuo vertė yra arčiau 0, tuo situacija yra blogesnė.

Siekiant išsiaiškinti, ar atitinkami kintamieji gali sudaryti indeksą, t. y. ar atitinkami kintamieji gali būti derinami tarpusavyje, buvo skaičiuojama Cronbach alpha. Cronbach alpha parodo vidinį darnumą (angl. *internal consistency*), t. y. kokių mastu rodikliai matuoja tą pačią savybę. Cronbach alpha buvo skaičiuojama naudojant normalizuotą duomenų rinkinį pagal aukščiau pateiktas 14 ir 15 formules. Toliau pateikiama Cronbach alpha formulė, remiantis [87] šaltiniu.

$$\alpha = \frac{N}{N-1} \left( \frac{\sigma_X^2 - \sum_{i=1}^N \sigma_{Y_i}^2}{\sigma_X^2} \right); \quad (16)$$

čia N – rodiklių skaičius;

$\sigma_X^2$  – susumuočių pagal stebėjimus rodiklių dispersija;

$\sigma_{Y_i}^2$  – rodiklio i dispersija.

Yprastai  $\alpha$  kinta nuo 0 iki 1. Tiesa, gali būti ir neigiama  $\alpha$ , kai kintamieji tarpusavyje nėra teigiamai koreliuoti. Aukštesnė Cronbach alpha reikšmė parodo didesnę vidinę darną (angl. *internal consistency*). Toliau pateikiama lentelė, kurioje galima matyti Cronbach alpha interpretavimą.

**8 lentelė.** Cronbach alpha reikšmės interpretavimui, [88 p. 60]

Cronbach alpha ( $\alpha$ )	Vidinė darną (angl. <i>internal consistency</i> )
$\alpha < 0,5$	Nepriimtina
$0,5 \leq \alpha \leq 0,6$	Nepakankama
$0,6 \leq \alpha \leq 0,7$	Abejotina
$0,7 \leq \alpha \leq 0,8$	Priimtina
$0,8 \leq \alpha \leq 0,9$	Gera
$\alpha \geq 0,9$	Puiki

Iš atitinkamų normalizuotų požymių reikšmių yra apskaičiuojamas bendras agreguotas įvertis (t. y. atitinkamas konkurencingumo indeksas) atitinkamai šaliai pagal formulę:

$$z_i^t = \frac{1}{m} \sum_{k=1}^m z_{ik}^t; \quad (17)$$

čia  $z_i^t$  – agreguotas įvertis šaliai i laiko momente t;

m – požymių skaičius;

$z_{ik}^t$  – normalizuota požymio k reikšmė šaliai i laiko momente t.

Kiekvienai ES šaliai kiekvienais analizuojamais 2008–2019 m. buvo apskaičiuoti 4 agreguoti įverčiai – ekonominio, socialinio, šalies aplinkos ir aplinkosauginio konkurencingumo pagal 17 formulę ir pagal atitinkamus rodiklius, kurie pateikiami 1 priede (išskyrus rodiklius, kurie turėjo per daug trūkstamų reikšmių ir buvo nenaudoti). Bendras konkurencingumo indeksas buvo apskaičiuotas kaip anksčiau minėtų konkurencingumo indeksų vidurkis, t. y. kiekvienam iš jų suteikiant vienodą 0,25 svorį.

### 3. Rezultatai

Šiame skyriuje pateikiami rezultatai, gauti siekiant suklasterizuoti Europos Sąjungos šalis ir įvertinti jų konkurencingumą, pasitelkiant statistinius ir mašininio mokymosi metodus.

#### 3.1. Trūkstamos reikšmės

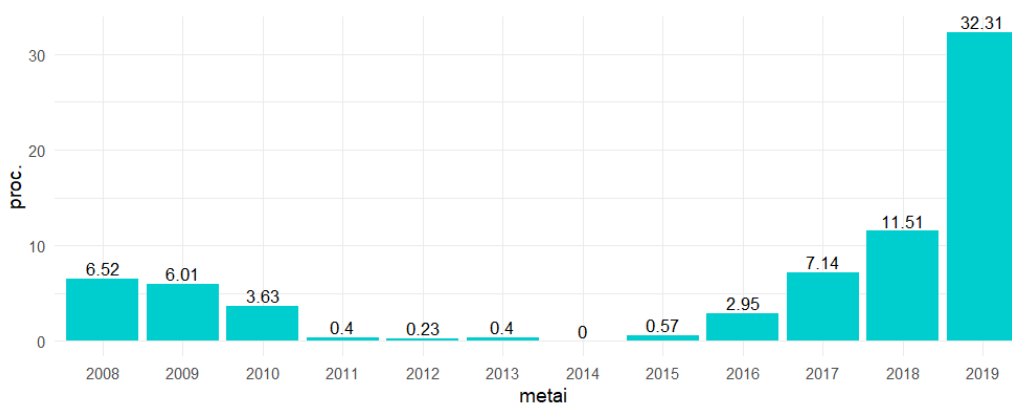
Šiame skyrelyje bus aptariamos trūkstamos reikšmės bei pristatomi rezultatai trūkstamų reikšmių užpildymui panaudojus *missForest* algoritimą.

##### 3.1.1. Trūkstamų reikšmių apžvalga

Prieš pradėdant užpildyti trūkstamas analizuojamo duomenų rinkinio reikšmes svarbu išanalizuoti duomenų rinkinį, įvertinti, kiek ir kokių reikšmių trūksta. Duomenų matricą sudaro 336 stebėjimai ir 72 kintamieji (neįskaičiuoti tokie stulpeliai kaip šalis, šalies kodas, metai). Analizuojami 28 ES šalių duomenys 2008–2019 m. laikotarpiu. Kadangi Jungtinė Karalystė išstojo iš ES 2020-01-31, tad šiame darbe ji analizuota kaip ES šalis. Duomenų rinkinyje trūkstamos reikšmės sudaro 8,44 proc. ir yra 25 stebėjimai, kai netrūksta nė vienos reikšmės.

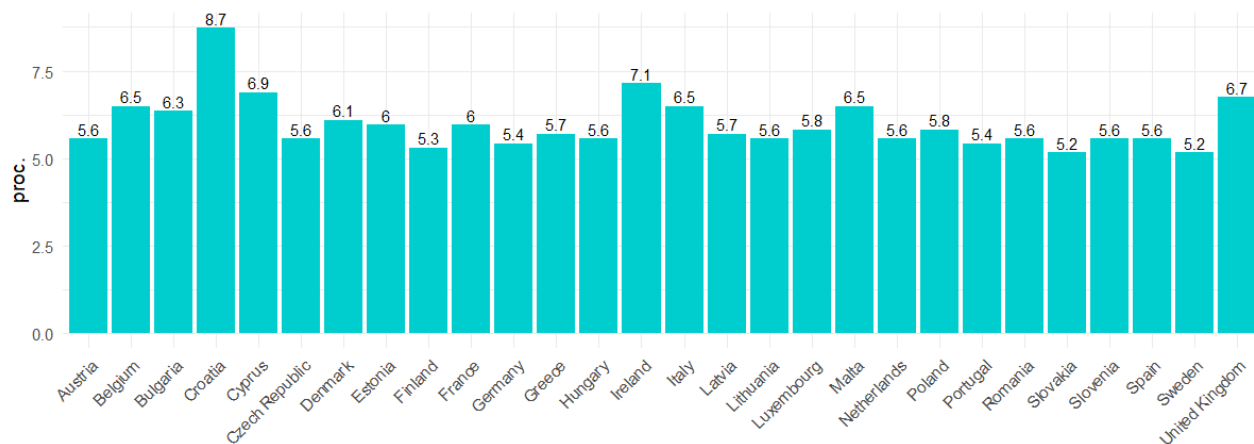
Svarbu įvertinti, ar nėra tokių kintamųjų, kurių reikšmių trūksta ištiesai per visą šalį (šalis) visu analizuojamu laikotarpiu. Tai iliustruojantys paveikslai pateikiami 2 priede. Kadangi duomenų rinkinyje buvo kritinių kintamųjų, kurių reikšmių trūko ištiesai per visą šalį, buvo nuspręsta jų atsisakyti, iš duomenų rinkinio buvo pašalinti 9 kintamieji (*waste\_perc\_dmc*, *doctors\_pthc*, *hospital\_beds\_pthc*, *researchers\_per\_ths*, *young\_self\_empl*, *railroad\_quality*, *infrastr\_invest\_gdp*, *gva\_ce\_eur\_pc*, *dis\_income\_pc*). Toliau analizuojami duomenys be kritinių kintamųjų.

Daugiausia trūkstamų reikšmių pastebima 2019 m. ir jos sudaro 32,3 proc. visų 2019 m. duomenų (žr. 9 pav.). Kitais analizuojamais 2008–2018 m. trūkstamos reikšmės svyruoja nuo 0 iki 11,51 proc.



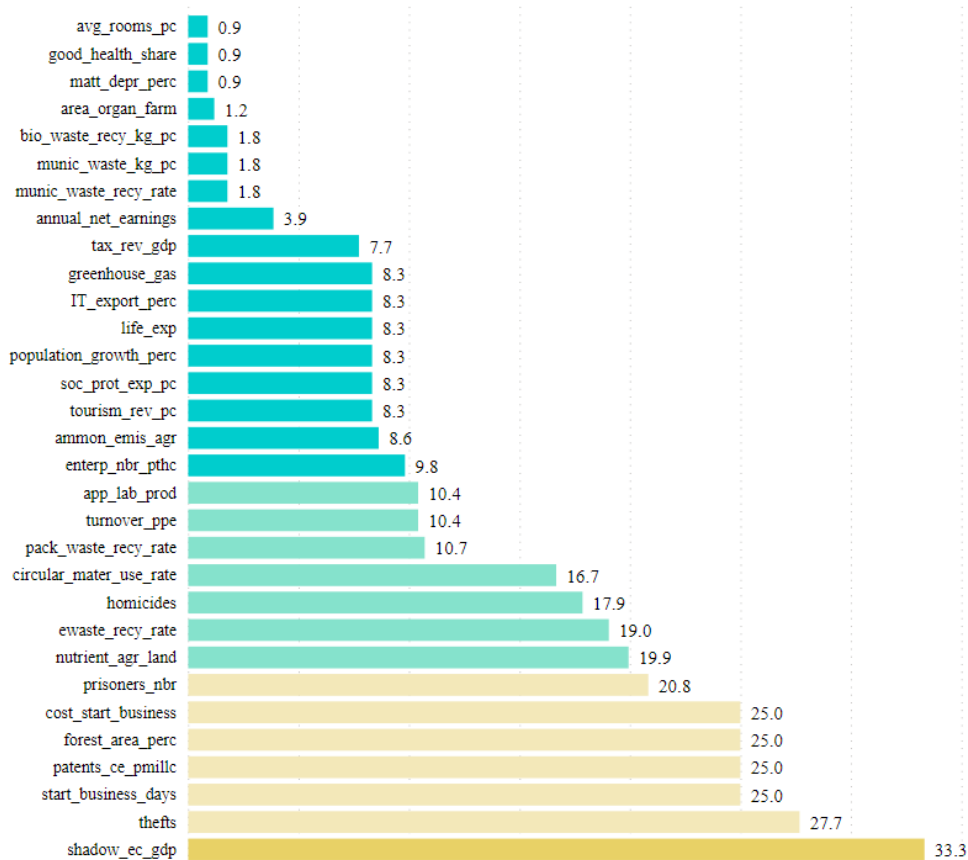
9 pav. Duomenų rinkinio trūkstamos reikšmės pagal metus, proc.

Taip pat svarbu išsiaiškinti, kiek trūkstamų reikšmių yra pagal analizuojamas ES šalis. Nėra nė vienos ES šalies duomenų rinkinyje, kuri neturėtų trūkstamų reikšmių, tačiau galima pastebėti, jog trūkstamos reikšmės neviršija 9 proc. (žr. 10 pav.). Daugiausia trūkstamų reikšmių yra Kroatijos duomenyse, šios šalies trūkstamos reikšmės visuose šios šalies duomenyse sudaro 8,7 proc. Yra 19 ES šalių, kurių trūkstami duomenys svyruoja nuo 5 iki 6 proc., 7 šalys, kurių trūkstamos reikšmės sudaro nuo 6 iki 7 proc. ir dvi šalys, kurių trūkstami duomenys sudaro nuo 7 iki 9 proc.



10 pav. Duomenų rinkinio trūkstamos reikšmės pagal ES šalis, proc.

Taip pat svarbu išsiaiškinti, kiek trūkstamų reikšmių yra pagal požymius. Iš 63 kintamųjų, 32 kintamieji analizuojamame duomenų rinkinyje neturi trūkstamų reikšmių. 31 kintamojo trūkstamos reikšmės svyruoja nuo 0,9 proc. iki 33,3 proc. (žr. 11 pav.). 17 kintamųjų trūkstamos reikšmės sudaro iki 10 proc., 7 kintamųjų – nuo 10 iki 20 proc., 6 kintamųjų – nuo 20 iki 30 proc., ir 1 kintamojo – nuo 30 iki 40 proc.



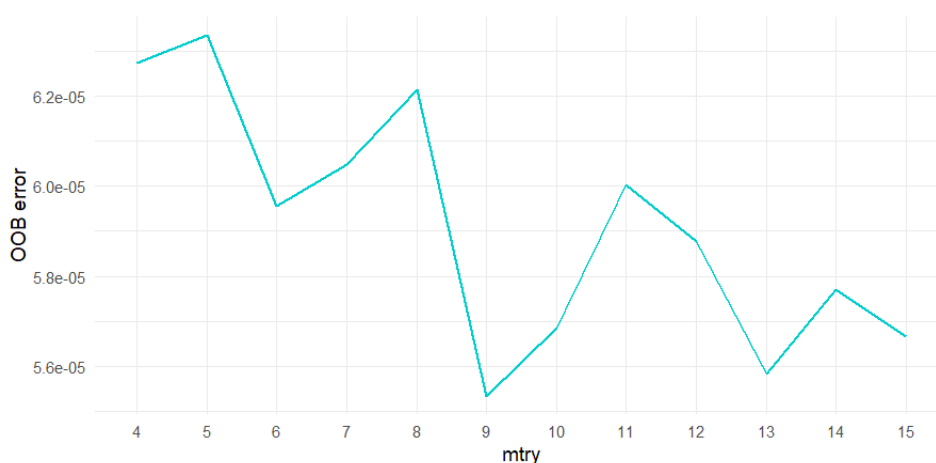
11 pav. Trūkstamos reikšmės pagal požymius, proc.

Buvo nuspręsta tolimesnėje analizėje naudoti kintamuosius, kurių trūkstamos reikšmės duomenų rinkinyje neviršija 35 proc. (žr. 11 pav.). Nors kintamojo *shadow\_ec\_gdp* trūkstamos reikšmės

sudarė 33,3 proc., jis buvo naudotas tolimesniame tyrime, nes šis kintamasis yra svarbus, jis buvo akcentuojamas literatūroje. Tolimesnėje analizėje bus naudojamas aptartas duomenų rinkinys – 32 kintamieji, kurie neturi trūkstamų reikšmių ir 31 kintamasis, kuriems reikia užpildyti trūkstamas reikšmes.

### 3.1.2. Trūkstamų reikšmių užpildymas

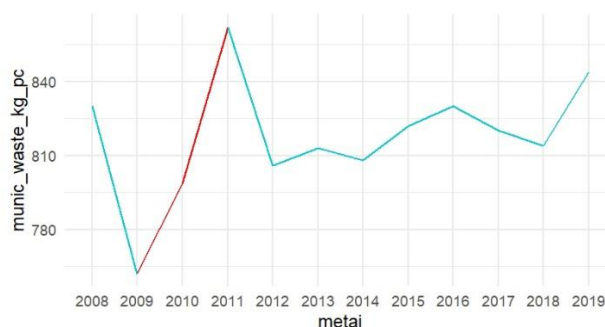
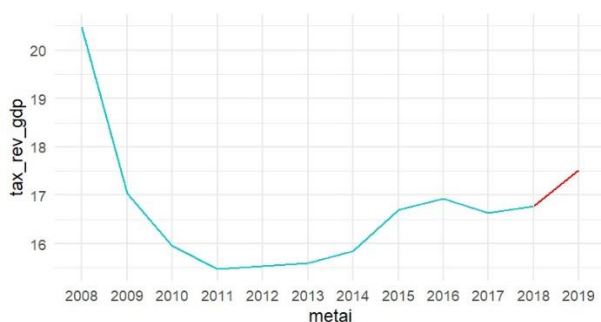
Trūkstamų reikšmių užpildymui buvo panaudotas *missForest* algoritmas. Medžių skaičius kiekviename miške buvo nurodytas 100 (*n tree* parametras), o atsitiktinai parinktų kintamųjų skaičius kiekviename padalijime buvo nurodytas 9 (*mtry* parametras). Numatytas (angl. *default*) *mtry* parametras yra  $\sqrt{p}$ , kur  $p$  – tai trūkstamų kintamųjų skaičius. Minėtas parametras buvo pasirinktas derinant algoritmą prie skirtingo *mtry* skaičiaus nuo 4 iki 15, viena mažiausių klaidų buvo gauta, kai *mtry* buvo lygus 9 (žr. 12 pav.).

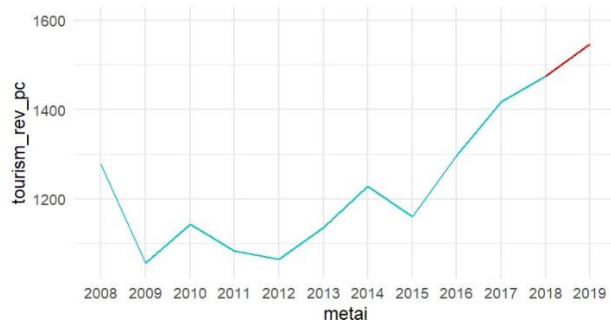
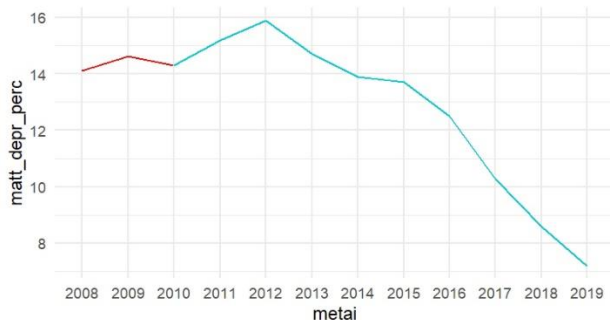


12 pav. OOB klaidos prie skirtingų *mtry* reikšmių

Algoritmui prireikė 5 iteracijų ir apie 0,5 min, buvo gauta paklaida, kuri yra lygi 5.532308e-05 (OOB NRMSE). Analizuojant duomenų skaitines charakteristikas (duomenis laikant kaip viena laikius, angl. *cross-sectional*) pastebėta, jog kintamųjų, kurių trūkstamos reikšmės buvo užpildytos, minimalios ir maksimalios reikšmės nepakito, jos liko tokios pat kaip ir pradiname duomenų rinkinyje (žr. 3 priedą).

Įvertinti, ar algoritmas teisingai užpildė trūkstamas reikšmes – sudėtinga, nes nėra žinomos tikrosios trūkstamų kintamųjų reikšmės. Svarbu atsižvelgti, ar užpildytos reikšmės yra logiškos, pavyzdžiui, ar užpildytos kintamojo reikšmės nėra neigiamos, kai jis gali įgyti tik teigiamas reikšmes, ir ar užpildytos kintamųjų reikšmės atitinka bendrą tendenciją. Išanalizavus skaitines užpildyto duomenų rinkinio charakteristikas, galima teigti, jog užpildytos reikšmės yra korektiškos.



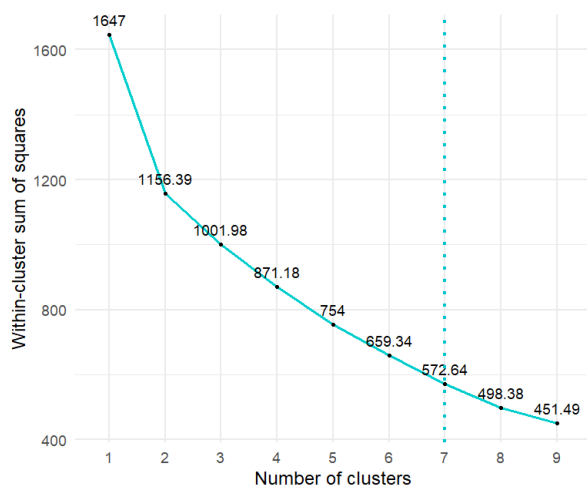


**13 pav.** Pasirinktų kintamųjų esamos (žalia) ir užpildytos (raudona) reikšmės, viršuje kairėje Lietuvos, viršuje dešinėje Danijos, apačioje kairėje Kroatijos, apačioje dešinėje Švedijos

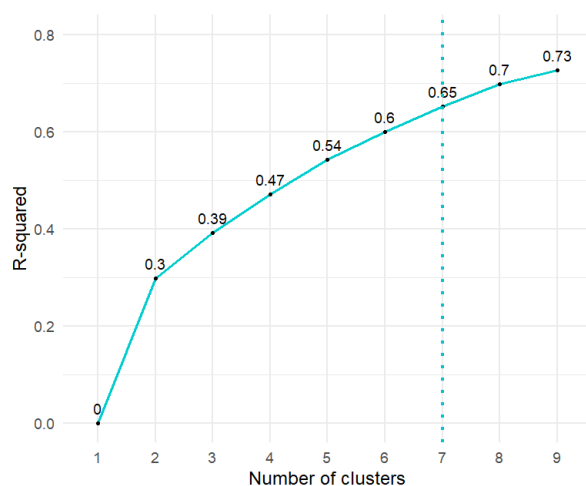
Siekiant įvertinti, ar užpildytos reikšmės atitinka bendrą tendenciją, buvo vertinami kintamųjų grafikai pagal šalis, keletas iš jų yra pateikiami 13 pav. Galima pastebėti, jog užpildytos reikšmės atitinka bendrą tendenciją.

### 3.2. Europos Sąjungos šalių klasterizavimas

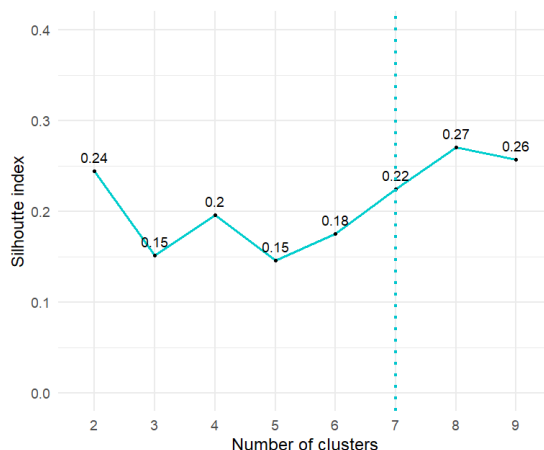
Siekiant įsitikinti, jog duomenys gali būti klasterizuojami, buvo skaičiuojama Hopkins'o statistika 2008, 2013 ir 2019 m. ir šios reikšmės buvo lygios 0,663, 0,640 ir 0,636 atitinkamai. Kadangi gautos minėtos statistikos reikšmės buvo didesnės už 0,5, todėl galima teigti, jog duomenys gali būti klasterizuojami, juose yra tam tikra struktūra, kurią galima atskleisti. ES šalių klasterizavimas buvo atliekamas su k-vidurkių ir jungimo hierarchiniu klasterizavimu. Klasterizacija buvo atliekama 2008, 2013 ir 2019 m. pagal 61 rodiklį, kurių reikšmės buvo normalizuojamos (*z-score* būdu). K-vidurkių algoritmo atveju klasterių skaičius buvo parinktas atsižvelgiant į Silueto indeksą ir taikant alkūnės metodo bei  $R^2$  grafikus. 2008 m. pasirinktas šalių padalinimas į 7 klasterius (žr. 14-16 pav.).



**14 pav.** Alkūnės metodas klasterių skaičiaus parinkimui



**15 pav.**  $R^2$  prie skirtingo klasterių skaičiaus



**16 pav.** Silueto koeficientas prie skirtingo klasterių skaičiaus

Klasterių skaičių nustatyti iš 14 ar 15 pav. buvo sudėtinga, nes ženkliausias lūžio taškas buvo nuo 1 iki 2 klasterio ir, atrodytų, 2 klasteriai būtų optimaliausias pasirinkimas, tačiau buvo nuspręsta, jog 2 klasteriai šiame tyrime yra per mažai, todėl atsižvelgiant į 14-16 pav. bei į gautus klasterizavimo rezultatus, kurie buvo interpretuojami, buvo nuspręsta pasirinkti klasterių skaičių, kuris yra lygus 7.

Siekiant palyginamumo, 2013 m. ir 2019 m. šalys taip pat buvo klasterizuojamos nurodant 7 klasterius. Klasterizavimo vizualizavimui buvo panaudotos principinės komponentės. Klasterių interpretavimas buvo atliekamas pagal k-vidurkių rezultatus, o interpretavimui buvo panaudoti klasteryje esančių šalių požymių vidurkiai, kurie pateikiami 4, 5 ir 6 priede. Toliau skyreliuose bus pateikiama informacija apie klasterizavimo rezultatus 2008, 2013 ir 2019 m. atitinkamai.

### 3.2.1. 2008 m. klasterizavimo rezultatai

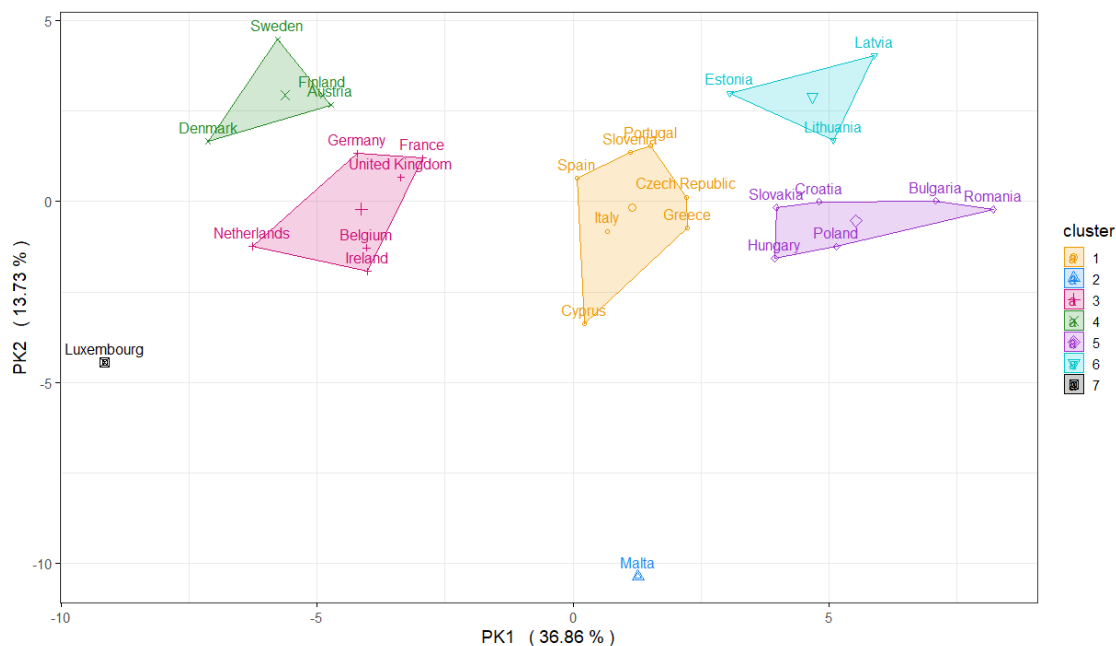
2008 m. k-vidurkių klasterizavimo rezultatus galima matyti 17 pav. Pastebima, jog klasteriai aiškiai atsiskyrę juos atvaizdavo dvimatėje erdvėje panaudojant pirmas dvi pagrindines komponentes, kurios nusako apie 50,6 proc. duomenų sklaidos. Verta atkreipti dėmesį, jog vienu klasterių (1, 2, 5, 6) šalių pirmosios pagrindinės komponentės (PK1) reikšmės yra didesnės už 0, o kitų klasterių (3, 4, 7) šalių PK1 reikšmės yra mažesnės už 0. Atsižvelgiant į pirmosios komponentės didžiausius svorius (žr. 7 priedą), galima teigti, jog šalių, kurių PK1 reikšmės yra didesnės už 0, gyventojams tenka mažiau socialinės apsaugos ir sveikatos išlaidų, taip pat tenka mažesnė bendroji pridėtinė vertė, šiose šalyse vyrauja mažesnis metinis grynasis uždarbis, būdinga didesnė korupcija, skiriamos mažesnės išlaidos moksliniams tyrimams ir plėtrai palyginus su šalimis, kurių PK1 reikšmės yra mažesnės už 0. Bulgarija ir Rumunija yra silpniausios šalys pagal išvardintus kintamuosius lyginant su kitomis ES šalimis. Vertinant antrosios komponentės (PK2) didžiausius svorius (žr. 7 priedą), galima teigti, jog kuo šalies PK2 reikšmė didesnė, tuo žemės ūkyje išmetamas amoniakas mažesnis, mažesnės TUI, didesnė atsinaujinančios energetikos dalis bendrame galutiniame energijos suvartojime, didesnė aktyvios populiacijos dalis, didesnis miškingumas, mažesnis importas, mažesnis bendras maistinių medžiagų balansas žemės ūkio paskirties žemėje ir t. t.

Pirmą klasterį sudaro Kipras, Čekija, Graikija, Italija, Portugalija, Slovėnija ir Ispanija (žr. 17 pav.). Šį klasterį būtų galima įvardinti kaip vieną mažiausiai konkurencingų. Jis pasižymi žemu aplinkosauginiu konkurencingumu. 1 klasteris iš kitų išsiskiria mažiausiu eksporto lygiu. Verta



atkreipti dėmesį, jog kituose analizuojamuose klasteriuose prekių ir paslaugų eksportas vidutiniškai sudaro nuo 50,36 iki 187,11 proc. nuo BVP, tuo tarpu kai 1 klasterio vidutinis eksportas yra 40,92 proc. nuo BVP. Taip pat 1 klasteriui būdingi ir kiti žemi ekonominiai rodikliai, tokie kaip – importas, apyvarta, bendroji pridėtinė vertė, turizmo pajamos. Analizuojant šalies aplinkos rodiklius galima pastebėti, jog šiam klasteriui būdingas aukštas šešėlinės ekonomikos lygis bei žemas politinis stabilumas. Vertinant socialinį konkurencingumą galima pastebėti, jog šio klasterio socialinis konkurencingumas nėra aukštas, jį būtų galima vertinti kaip vidutinį, tačiau 1 klasteris iš kitų klasterių išsiskiria mažiausiu žmogžudysčių skaičiumi.

Antrą klasterį sudaro Malta (žr. 17 pav.). Ši šalis pasižymi žemu aplinkosauginiu konkurencingumu, nes jos atliekų perdirbimo lygis ir atsinaujinančios energetikos dalis bendrame galutiniame energijos suvartojime yra mažiausi lyginant su kitais analizuojamais klasteriais. Maltai būdingos didžiausios tiesioginės užsienio investicijos, eksporto augimas bei IT eksportas, kuris yra išskirtinai didesnis nei kitų nagrinėjamų klasterių. Maltos IT eksportas visame eksporte 2008 m. sudarė 43,19 proc. tuo tarpu kai kituose klasteriuose šis dydis svyravo nuo 3,03 iki 9,3 proc. Šiam klasteriui būdingos nepalankiausios sąlygos verslui bei didžiausia šešėlinės ekonomikos dalis ir mažiausias užimtumo lygis. Beje, Maltoje 2008 m. lyginant su kitais klasteriais buvo mažiausia aktyvios populiacijos dalis, kuri buvo lygi 62,3 proc. tuo tarpu kai kituose klasteriuose šis dydis svyravo nuo 70,98 iki 80,7 proc.



**17 pav.** K-vidurkių 2008 m. klasterizavimo rezultatų vizualizavimas naudojant dvi pirmas pagrindines komponentes

Trečią klasterį sudaro Belgija, Prancūzija, Vokietija, Airija, Nyderlandai ir Jungtinė Karalystė (žr. 17 pav.). Šį klasterį būtų galima apibūdinti kaip vieną konkurencingiausių. Nors šiame klasteryje esančių šalių vidutinė atsinaujinančios energetikos dalis bendrame energijos suvartojime yra viena mažiausių ir komunalinių atliekų kiekis, tenkantis gyventojui, yra vienas didžiausių, tačiau šio klasterio aplinkosauginį konkurencingumą galima būtų vertinti kaip vieną aukštesnių, nes šiam klasteriui būdingi aukšti atliekų perdirbimo rodikliai bei aukštas žiedinis medžiagų panaudojimo lygis. Vertinant kitus rodiklius, galima pastebėti, jog šis klasteris iš kitų išsiskiria mažiausiu eksporto augimu, mažiausia infliacija ir mažiausiu laiku reikalingu verslui pradėti. 3 klasteriui

būdinga didžiausia aukštąjį išsilavinimą turinčių dalis bei didžiausia žmonių, kurių sveikatos būklė gerai ar labai gera, dalis.

Ketvirtą klasterį sudaro Austrija, Danija, Suomija ir Švedija (žr. 17 pav.). 4 klasterį būtų galima apibūdinti kaip vieną konkurencingiausių. Šis klasteris pasižymi aukštu aplinkosauginiu konkurencingumu, nes atsinaujinančios energetikos dalis bendrame galutiniame energijos suvartojime, komunalinių ir elektroninių atliekų perdirbimo lygis, miško plotas bei ekologinio ūkininkavimo plotas yra didžiausi lyginant su kitais klasteriais. Verta atkreipti dėmesį, jog šio klasterio vidutinė atsinaujinančios energetikos dalis bendrame galutiniame energijos suvartojime 2008 m. vidutiniškai sudarė 30,71 proc. ir tai buvo išskirtinai geras rodiklis, lyginant su kitais klasteriais, kuriuose atitinkamo rodiklio reikšmės svyravo nuo 0,2 iki 22,1 proc. Analizuojant ekonominius rodiklius, galima pastebėti, jog 4 klasteris pasižymi mažiausiomis tiesioginėmis užsienio investicijomis bei importu, tačiau šiam klasteriui būdingas didžiausias užimtumo lygis bei geriausi transporto infrastruktūros rodikliai. Taip pat šiam klasteriui būdingi ir kiti aukšti ekonominiai rodikliai, tokie kaip apyvarta, produktyvumas, mokslinių tyrimų ir plėtros išlaidos, patentų paraiškų skaičius bei dirbančių moksle ir technologijose dalis. 4 klasteriui būdingas aukščiausias šalies aplinkos konkurencingumas, dauguma rodiklių yra geresni nei kitų klasterių. Šis klasteris pasižymi aukštu socialiniu konkurencingumu, nes jam būdinga didžiausia aktyvios populiacijos dalis, besimokančių suaugusiųjų dalis bei mažiausias kalinių skaičius. Taip pat 4 klasteriui būdingi ir kiti geri socialiniai rodikliai, tokie kaip – didelės išlaidos švietimui, sveikatai ir socialinei apsaugai, mažas žmogžudysčių skaičius bei maža itin nepasiturinčių asmenų dalis.

Penktą klasterį sudaro Bulgarija, Kroatija, Vengrija, Lenkija, Rumunija ir Slovakija (žr. 17 pav.). Šį klasterį būtų galima apibūdinti kaip vieną iš mažiausiai konkurencingų. Jam būdingi vieni mažiausių atliekų perdirbimo rodikliai bei mažiausias žiedinis medžiagų panaudojimo lygis. Taip pat 5 klasteris pasižymi mažiausiu grynuoju uždarbiu, mažiausiomis išlaidomis moksliniams tyrimams ir plėtrai, mažiausiu patentų skaičiumi, mažiausia dirbančių dalimi moksle ir technologijose bei žemiausiais transporto infrastruktūros rodikliais. Šiam klasteriui būdingas mažas šalies aplinkos konkurencingumas bei socialinis konkurencingumas. 5 klasteris pasižymi mažiausia aukštąjį išsilavinimą turinčių bei besimokančių suaugusių dalimi ir didžiausia nepasiturinčių dalimi.

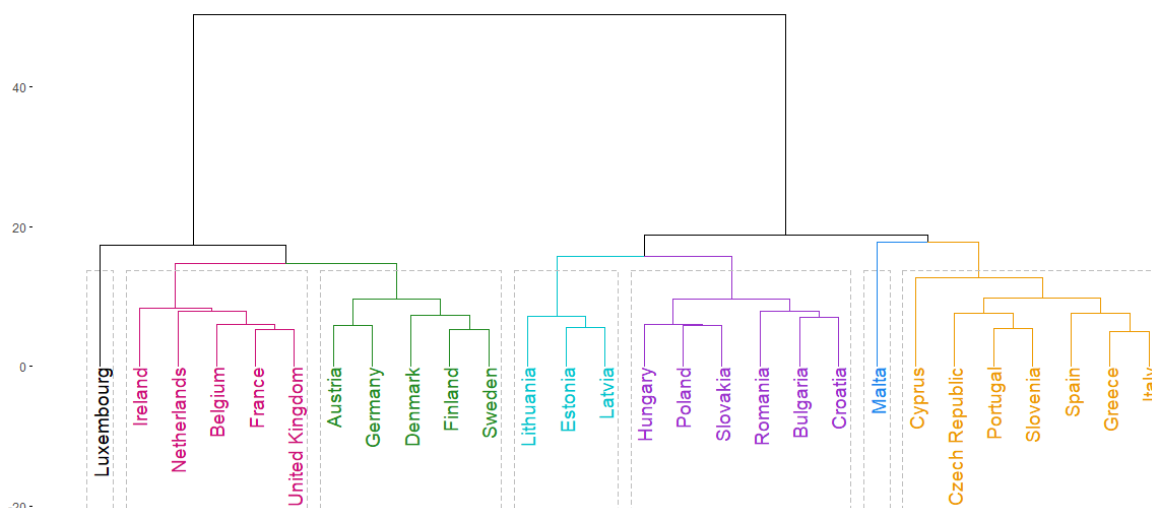
Šeštam klasteriui priklauso Baltijos šalys – Lietuva, Latvija, Estija (žr. 17 pav.). Šis klasteris yra vienas mažiausiai konkurencingų. 6 klasteris iš kitų klasterių labiausiai išsiskiria dėl žemo socialinio konkurencingumo. Šiam klasteriui būdinga mažiausia tikėtina gyvenimo trukmė, mažiausia žmonių, kurių sveikatos būklė gerai ar labai gera, dalis, mažiausios išlaidos sveikatai ir socialinei apsaugai bei didžiausias kalinių ir žmogžudysčių skaičius. Taip pat šis klasteris pasižymi didžiausia infliacija bei mažiausiu politiniu stabilumu. 6 klasteriui tenka mažiausios turizmo pajamos, žemiausia bendroji pridėtinė vertė bei mažiausia apyvarta, tenkanti vienam dirbančiam. Analizuojant aplinkosauginius rodiklius galima pastebėti, jog šiam klasteriui būdingas mažiausias komunalinių atliekų kiekis bei aukšta atsinaujinančios energetikos dalis bendrame energijos suvartojime.

Septintą klasterį sudaro Liuksemburgas (žr. 17 pav.). Ši šalis nėra panaši nei į vieną iš likusių nagrinėjamų šalių, todėl ji tokia atsiskyrusi. Šiai šaliai tenka didžiausias šiltnamio efekta sukeliančių dujų emisijos bei komunalinių atliekų kiekis. Liuksemburgas pasižymi aukštais atliekų perdirbimo rodikliais bei didžiausiu žiediniu medžiagų panaudojimo lygiu ir su perdirbimu ir antrinėmis žaliavomis susijusių patentų skaičiumi. 7 klasteriui būdingi didžiausi lyginant su kitais

klasteriais ekonominiai rodikliai – eksporto, importo, apyvartos, bendrosios pridėtinės vertės, metinio grynojo uždarbio, turizmo pajamų, mokslinių tyrimų ir plėtros išlaidų, patentų paraiškų, dirbančių moksle ir technologijose dalies. Taip pat Liuksemburge mažiausia šešėlinė ekonomika bei aukščiausias politinis stabilumas. Šiam klasteriui būdingas didžiausias populiacijos prieaugis, tikėtina gyvenimo trukmė, sveikatos ir socialinės apsaugos išlaidos bei mažiausia nepasiturinčių dalis.

Apibendrinant aukščiau aprašytus 2008 m. ES klasterizavimo k-vidurkių algoritmo rezultatus būtų galima teigti, jog konkurencingiausi klasteriai yra 7 (Liuksemburgas), 4 (Austrija, Danija, Suomija, Švedija) ir 3 (Belgija, Prancūzija, Vokietija, Airija, Nyderlandai, Jungtinė Karalystė), nekonkurencingiausi – 2 (Malta), 5 (Bulgarija, Kroatija, Vengrija, Lenkija, Rumunija, Slovakija), 6 (Lietuva, Latvija, Estija) ir 1 (Kipras, Čekija, Graikija, Italija, Portugalija, Slovėnija, Ispanija).

ES šalių klasterizavimas su 2008 m. duomenimis buvo atliktas panaudojus ir hierarchinį klasterizavimą, jo rezultatai yra pateikiami 18 pav. dendogramoje. Siekiant palyginamumo, klasterių skaičius taip pat buvo parinktas 7.



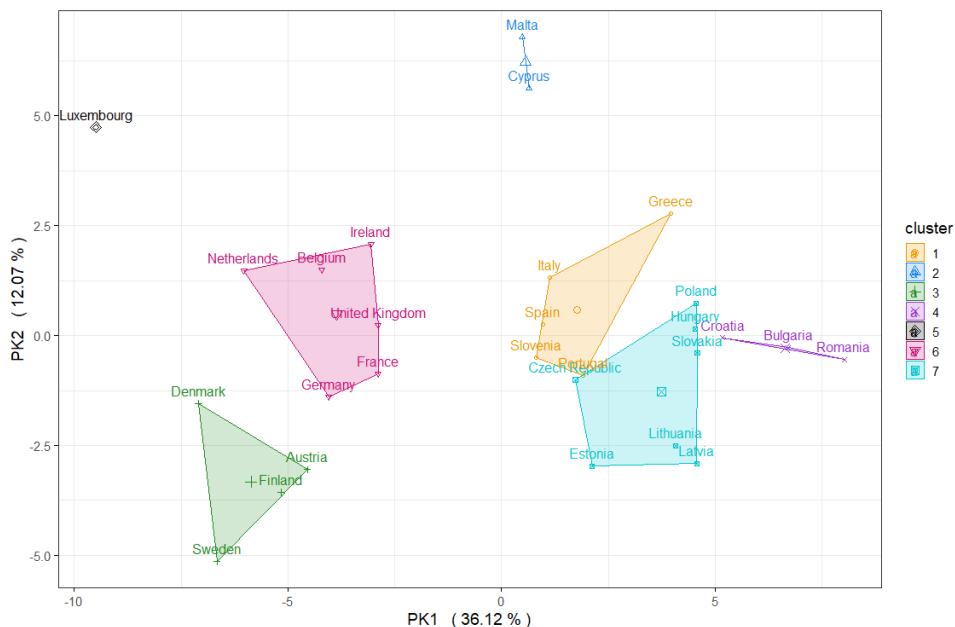
18 pav. Hierarchinio klasterizavimo dendograma pagal 2008 m. duomenis

Hierarchinio klasterizavimo būdu gautas šalių suskirstymas į klasterius labai panašus į tą, kuris buvo gautas panaudojus k-vidurkių algoritimą. Įvyko tik vienas pasikeitimas. Prie klasterio, kuris buvo gautas k-vidurkių būdu ir kurį sudarė Danija, Austrija, Suomija, Švedija, prisijungė Vokietija. Kiti klasteriai išliko nepakitę. Hierarchinio klasterizavimo atveju vidutinis Silueto indeksas yra 0,14867, o k-vidurkių šis indeksas yra didesnis ir yra lygus 0,152902.

### 3.2.2. 2013 m. klasterizavimo rezultatai

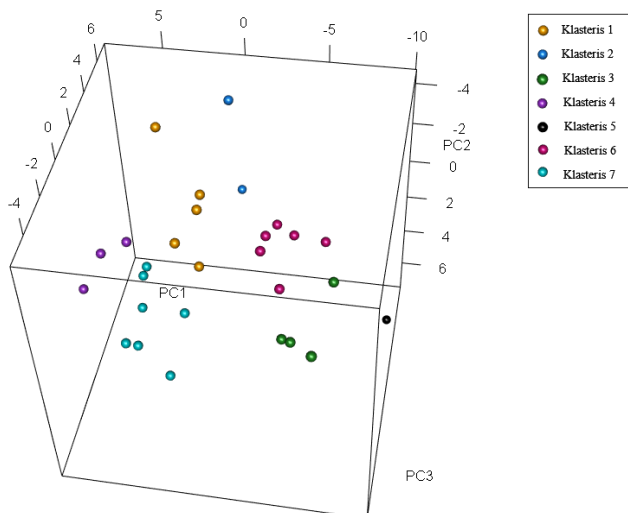
2013 m. k-vidurkių klasterizavimo rezultatus galima matyti 19 pav. Kaip ir 2008 m. taip ir 2013 m. didžiausi pirmosios pagrindinės komponentės svoriai buvo šių kintamųjų: metinio grynojo uždarbio, socialinės apsaugos ir sveikatos išlaidų, korupcijos kontrolės indekso, bendrosios pridėtinės vertės bei mokslinės plėtros ir tyrimų išlaidų (žr. 8 priedą). Vėlgi šalys, kurių pirmosios pagrindinės komponentės (PK1) reikšmės yra didesnės už 0, yra mažiau konkurencingos už tas šalis, kurių PK1 reikšmės yra mažesnės už 0. Galime pastebėti, jog vertinant pagal PK1 kaip ir 2008 m., taip ir 2013 m. Bulgarija ir Rumunija yra mažiausiai konkurencingos. Didžiausi antrosios

pagrindinės komponentės (PK2) svariai buvo šių kintamųjų – atsinaujinančios energetikos dalies, žemės ūkyje išmetamo amoniako, miško ploto, bendrojo maistinių medžiagų balanso, kapitalo investicijų bei ekologinio ūkininkavimo ploto (žr. 8 priedą). Galima pastebėti, jog atvaizdavirus k-vidurkių klasterizavimo rezultatus pagal dvi pirmas pagrindines komponentes, kurios apibūdina apie 48,2 proc. duomenų sklaidos, nėra itin aiškiai 1 ir 7 klasterio atskyrimo (žr. 19 pav.).



**19 pav.** K-vidurkių 2013 m. klasterizavimo rezultatų vizualizavimas naudojant dvi pirmas pagrindines komponentes

Klasterizavimo rezultatai buvo atvaizduoti trimatėje erdvėje panaudojus pirmas tris pagrindines komponentes, kurios apibūdina apie 57,4 proc. duomenų sklaidos (žr. 20 pav.). Trečiosios pagrindinės komponentės didžiausi svariai buvo šių kintamųjų – importo, eksporto, prekybos perdirbamomis žaliavomis, su perdirbimu ir antrinėmis žaliavomis susijusių patentų bei išlaidų kultūrai (žr. 8 priedą). Atvaizdavirus k-vidurkių klasterizavimo rezultatus trimatėje erdvėje panaudojus tris pirmas pagrindines komponentes galima matyti, jog visi klasteriai gana aiškiai atsiskiria (žr. 20 pav.).



**20 pav.** K-vidurkių 2013 m. klasterizavimo rezultatai atvaizdavirus pagal tris pirmas pagrindines komponentes

Pirmą klasterį sudaro Graikija, Italija, Portugalija, Slovėnija, Ispanija (žr. 19 pav.). Šį klasterį būtų galima įvardinti kaip vieną mažiausiai konkurencingų. Jis pasižymi vidutiniu aplinkosauginiu konkurencingumu. 1 klasteris iš kitų išsiskiria mažiausiu eksporto, importo ir užimtumo lygiu. Analizuojant šalies aplinkos rodiklius galima pastebėti, jog šiam klasteriui būdingas aukštas šešėlinės ekonomikos lygis bei žemas politinis stabilumas. Vertinant socialinį konkurencingumą galima pastebėti, jog šio klasterio socialinis konkurencingumas nėra aukštas, jį būtų galima vertinti kaip vidutinį, tačiau 1 klasteris iš kitų klasterių išsiskiria viena aukščiausių tikėtina gyvenimo trukme.

Antrą klasterį sudaro Kipras, Malta (žr. 19 pav.). Šis klasteris iš kitų išsiskiria vienais prasčiausių aplinkosauginių rodiklių – žemas atliekų perdirbimo lygis, maža atsinaujinančios energetikos dalis, žemas žiedinis medžiagų panaudojimo lygis. Vertinant ekonominius rodiklius galima pastebėti, jog šis klasteris iš kitų išsiskiria didžiausiomis tiesioginėmis užsienio investicijomis bei didžiausiu IT eksportu. Taip pat šiam klasteriui būdingas didelis eksportas ir importas bei aukštos turizmo pajamos. Vertinant šalies aplinkos rodiklius galima pastebėti, jog šiam klasteriui būdinga nepalanki aplinka verslui bei didžiausia šešėlinės ekonomikos dalis (net 30,9 proc. nuo BVP, kai kituose klasteriuose šis dydis svyruoja nuo 10,65 iki 24 proc.). 2 klasteriui taip pat būdinga didžiausia žmonių, turinčių aukštąjį išsilavinimą, dalis bei didžiausia žmonių, kurių sveikatos būklė gera ar labai gera dalis.

Trečią klasterį sudaro Austrija, Danija, Suomija, Švedija (žr. 19 pav.). 3 klasterį būtų galima apibūdinti kaip vieną konkurencingiausių. Šis klasteris kaip ir 2008 m., taip ir 2013 m. pasižymi aukštu aplinkosauginiu konkurencingumu – atsinaujinančios energetikos dalis bendrame galutiniame energijos suvartojime, komunalinių ir elektroninių bei biologinių atliekų perdirbimo lygis, miško bei ekologinio ūkininkavimo plotas yra didžiausi lyginant su kitais klasteriais. Verta atkreipti dėmesį, jog šio klasterio vidutinė atsinaujinančios energetikos dalis bendrame galutiniame energijos suvartojime 2013 m. vidutiniškai sudarė 36,84 proc. ir tai buvo išskirtinai geras rodiklis, lyginant su kitais klasteriais, kuriuose atitinkamo rodiklio reikšmės svyravo nuo 3,5 iki 23,61 proc. Analizuojant ekonominius rodiklius, galima pastebėti, jog 3 klasteris pasižymi mažiausiomis tiesioginėmis užsienio investicijomis bei mažiausiu eksporto augimu, tačiau šiam klasteriui būdingas didžiausias užimtumo lygis, didžiausios išlaidos moksliniams tyrimams ir plėtrai bei vieni geriausių transporto infrastruktūros rodiklių. Taip pat šiam klasteriui būdingi ir kiti aukšti ekonominiai rodikliai, tokie kaip kapitalo investicijos, bendroji pridėtinė vertė, patentų paraiškų skaičius bei dirbančių moksle ir technologijose dalis. 3 klasteriui būdingas aukštas šalies aplinkos konkurencingumas, dauguma rodiklių yra geresni nei kitų klasterių. Šis klasteris pasižymi aukštu socialiniu konkurencingumu – jam būdinga didžiausia aktyvios populiacijos dalis, besimokančių suaugusiųjų dalis, didžiausios išlaidos švietimui bei mažiausias kalinių skaičius. Taip pat 3 klasteriui būdingi ir kiti geri socialiniai rodikliai, tokie kaip – išlaidos sveikatai ir socialinei apsaugai bei maža itin nepasiturinčių asmenų dalis.

Ketvirtą klasterį sudaro Bulgarija, Kroatija, Rumunija (žr. 19 pav.). Šį klasterį būtų galima apibūdinti kaip vieną mažiausiai konkurencingų. Analizuojant ekonominius šiam klasteriui priklausančių šalių rodiklių vidurkius galima pastebėti, jog jie yra vieni mažiausių. Šio klasterio gyventojams tenka mažiausiai tyrimų ir plėtros išlaidų. 4 klasteriui būdingas vienas žemiausių užimtumo lygių bei mažiausias metinis grynasis uždarbis, mažiausias patentų skaičius, mažiausia dalis dirbančių moksle ir technologijose. Taip pat šio klasterio transporto infrastruktūros kokybė yra prasčiausia. 4 klasteris pasižymi nekonkurencingiausia iš visų klasterių šalies aplinka – didelė

infliacija ir šešėlinės ekonomikos dalis, korupcija, žemas politinis stabilumas. Vertinant socialinius rodiklius galima matyti, jog šio klasterio socialinis konkurencingumas yra žemas – būdingas neigiamas populiacijos prieaugis, mažiausia aktyvios populiacijos dalis, mažiausia žmonių, turinčių aukštąjį išsilavinimą, dalis, mažiausia besimokančių suaugusių dalis, mažiausia tikėtina gyvenimo trukmė, mažiausios išlaidos sveikatai ir socialinei apsaugai bei didžiausia itin nepasiturinčių žmonių dalis. Vertinant aplinkosauginius rodiklius galima pastebėti, jog šis klasteris turi rodiklių, kurie yra geresni nei kitų klasterių, pavyzdžiui, mažiausia šiltnamio efektą sukeliančių dujų emisija, tenkanti vienam gyventojui bei vienas mažiausių komunalinių atliekų kiekis, tenkantis vienam gyventojui.

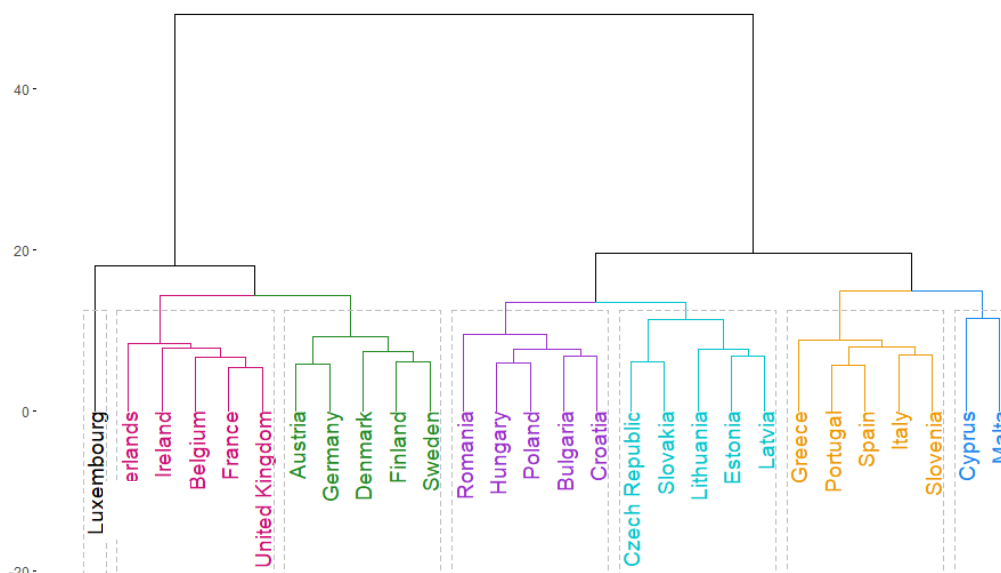
Penktą klasterį sudaro Liuksemburgas (žr. 19 pav.). Kaip ir 2008, taip ir 2013 m. Liuksemburgas yra išskirtinė šalis, nepanaši į likusius klasterius. Jam būdingi vieni geriausių ekonominių rodiklių ir kai kurie iš jų net kelis kartus viršija kitų klasterių tuos pačius rodiklius, pavyzdžiui, eksportas, apyvarta vienam dirbančiam, metinis grynasis uždarbis, turizmo pajamos, patentų paraiškų skaičius. Vertinant politinę situaciją šioje šalyje galima pastebėti, jog jai būdingas aukštas politinis stabilumas, nedidelis šešėlinės ekonomikos lygis. Dauguma šio klasterio socialinių rodiklių yra geriausi arba didesni nei vidutiniai, pavyzdžiui, šiai šaliai būdingas didžiausias populiacijos prieaugis ir didžiausia tikėtina gyvenimo trukmė, aukščiausios išlaidos sveikatai ir socialinei apsaugai. Nors Liuksemburgo aukštas atliekų ar žiedinis medžiagų panaudojimo lygis ir didžiausias su perdurbimu ir antrinėmis žaliavomis susijęs patentų skaičius, tačiau ši šalis išskiria daugiausia šiltnamio efektą sukeliančių dujų, sugeneruoja daugiausia komunalinių atliekų ir jos atsinaujinančios energetikos dalis bendrame energijos suvartojime yra mažiausia.

Šeštą klasterį, kurį sudaro Belgija, Prancūzija, Vokietija, Airija, Nyderlandai, Jungtinė Karalystė, galima apibūdinti kaip vieną konkurencingiausių (žr. 19 pav.). Vertinant aplinkosauginius rodiklius galima pastebėti, jog šis klasteris daugiausiai iš visų išleidžia aplinkos apsaugai, taip pat šiam klasteriui būdingas aukščiausias komunalinių bei pakuočių atliekų perdurbimo lygis ir didžiausias žiedinis medžiagų panaudojimo lygis. 6 klasteris pasižymi aukštu užimtumu bei aukštu metiniu grynuoju uždarbiu. Taip pat šiam klasteriui būdingos didelės išlaidos moksliniams tyrimams ir plėtrai bei aukšti transporto infrastruktūros rodikliai. Šiam klasteriui būdinga viena mažiausių šešėlinė ekonomika bei aukšti korupcijos kontrolės, valdžios efektyvumo rodikliai. Vertinant socialinius rodiklius galima pastebėti, jog dauguma jų yra aukšti ir geresni nei vidutiniai, pavyzdžiui, žmonių, turinčių aukštąjį išsilavinimą, dalis, žmonių, kurių sveikatos būklė gera ar labai gera, dalis, išlaidos sveikatai ir aplinkos apsaugai.

Septintą klasterį sudaro Čekija, Estija, Vengrija, Latvija, Lietuva, Lenkija, Slovakija (žr. 19 pav.). Šį klasterį būtų galima vertinti kaip konkurencingesnį už 4 klasterį (Bulgarija, Kroatija, Rumunija), tačiau jį vis tiek būtų galima priskirti prie mažiausiai konkurencingų. 7 klasteris pasižymi žemu socialiniu konkurencingumu. Jam būdingas neigiamas populiacijos prieaugis, viena mažiausių tikėtina gyvenimo trukmė, mažiausia žmonių, kurių sveikatos būklė gera ar labai gera, dalis, vienos mažiausių išlaidų sveikatai ir socialinei apsaugai, didžiausias kalinių ir žmogžudysčių skaičius. Vertinant šalies aplinkos konkurencingumo rodiklius galima pastebėti, jog jie yra vidutiniai ar žemesni nei vidutiniai. Analizuojant ekonominius rodiklius galima matyti, jog dauguma jų yra prastesni nei vidutiniai. Vertinant aplinkosauginį konkurencingumą verta pastebėti, jog šiam klasteriui būdingas mažiausias komunalinių atliekų kiekis, tenkantis vienam gyventojui bei vienas aukščiausių ekologinio ūkininkavimo ploto rodiklių, tačiau dauguma atliekų perdurbimo rodiklių yra prastesni nei vidutiniai.

Apibendrinant aukščiau aptartus 2013 m. klasterizavimo rezultatus, galima prieiti prie išvados, jog konkurencingiausi klasteriai – 5 (Liuksemburgas), 3 (Austrija, Danija, Suomija, Švedija), 6 (Belgija, Prancūzija, Vokietija, Airija, Nyderlandai, Jungtinė Karalystė). Mažiausiai konkurencingi klasteriai – 4 (Bulgarija, Kroatija, Rumunija), 2 (Kipras, Malta), 7 (Čekija, Estija, Vengrija, Latvija, Lietuva, Lenkija, Slovakija) ir 1 (Graikija, Italija, Portugalija, Slovėnija, Ispanija).

ES šalių klasterizavimas panaudojant 2013 m. duomenis buvo atliktas panaudojus ir hierarchinį klasterizavimą, jo rezultatai yra pateikiami 21 pav. dendogramoje.



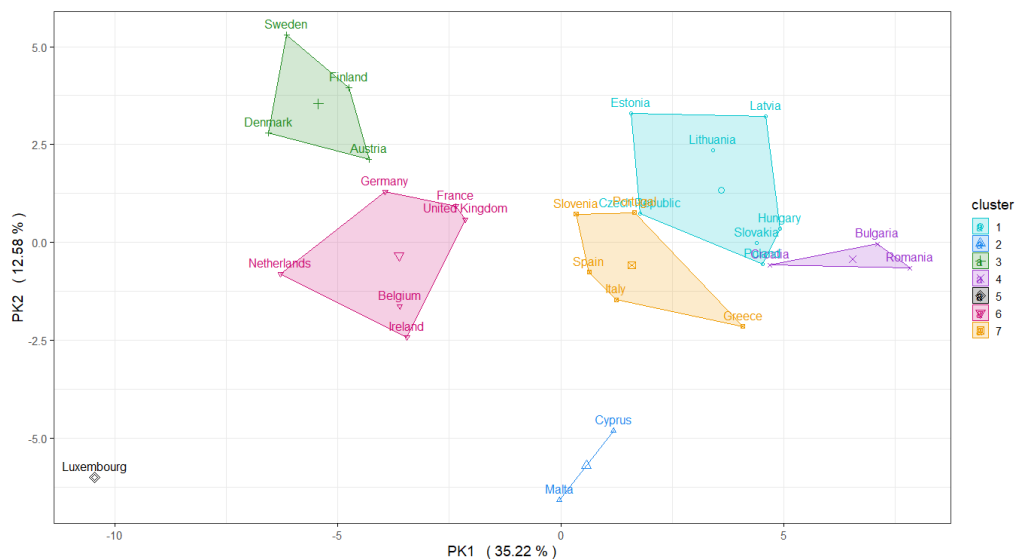
21 pav. Hierarchinio klasterizavimo dendograma pagal 2013 m. duomenis

Hierarchinio klasterizavimo būdu gautas šalių suskirstymas į klasterius labai panašus į tą, kuris buvo gautas panaudojus k-vidurkių algoritimą. Galima pastebėti, jog pakitimai įvyko su šių šalių priskyrimu klasteriams – Vokietija, Lenkija ir Vengrija. Prie klasterio, kuris buvo gautas k-vidurkių algoritmu ir kurį sudarė Danija, Švedija, Suomija ir Austrija, atlikus hierarchinį klasterizavimą buvo prijungta Vokietija. Taip pat prie klasterio, kurį sudarė Kroatija, Rumunija ir Bulgarija, atlikus hierarchinį klasterizavimą buvo prijungtos Lenkija ir Vengrija. Hierarchinio klasterizavimo atveju vidutinis Silueto indeksas buvo 0,12009, o k-vidurkių šis indeksas buvo didesnis ir buvo lygus 0,13823.

### 3.2.3. 2019 m. klasterizavimo rezultatai

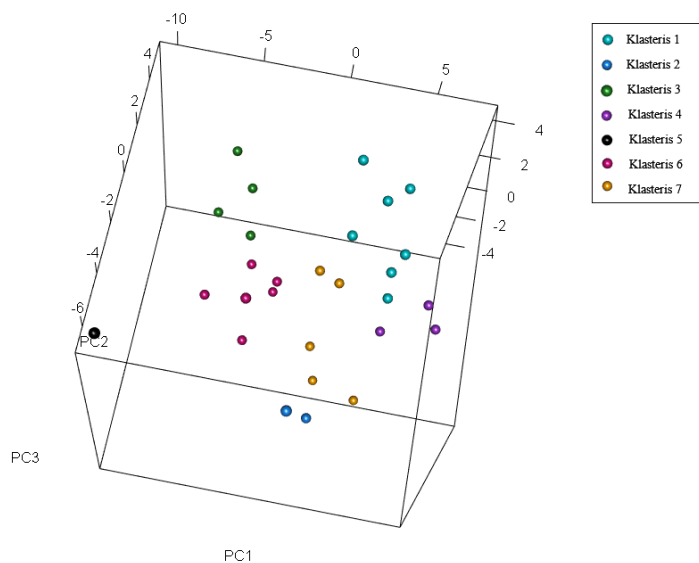
K-vidurkių klasterizavimo rezultatai su 2019 m. duomenimis yra pateikti 22 pav. Didžiausi pirmosios pagrindinės komponentės (PK1) svoriai buvo šių kintamųjų – socialinės apsaugos išlaidų, metinio grynojo uždarbio, patentų skaičiaus, dirbančių moksle ir technologijose, korupcijos kontrolės indekso (žr. 9 priedą). Kuo šalys silpnesnės pagal minėtus rodiklius, tuo jų PK1 reikšmės yra didesnės. Galima ir vėl pastebėti, jog vertinant pagal PK1 reikšmės, Bulgarija ir Rumunija yra silpniausios, o Liuksemburgas – stipriausias. Antrosios pagrindinės komponentės (PK2) svoriai 2019 m. buvo didžiausi šių kintamųjų – atsinaujinančios energetikos dalies, miškingumo, žemės ūkyje išmetamo amoniako, bendrojo maistinių medžiagų balanso, ekologinio ūkininkavimo ploto (žr. 9 priedą). Vertinant tik pagal PK2 reikšmes galima teigti, jog kuo šalies PK2 reikšmė didesnė, tuo šalis konkurencingesnė.

Galima pastebėti, jog atvaizdavus k-vidurkių klasterizavimo rezultatus pagal pirmas dvi pagrindines komponentes, kurios apibūdina apie 47,8 proc. duomenų sklaidos, nėra itin aiškus 1, 7 ir 4 klasterių atskyrimo (žr. 22 pav.).



**22 pav.** K-vidurkių klasterizavimo rezultatai pagal 2019 m. duomenis naudojant dvi pirmas pagrindines komponentes

Siekiant pamatyti aiškesnį klasterių atskyrimą, klasterizavimo rezultatai buvo atvaizduoti trimatėje erdvėje panaudojus pirmas tris pagrindines komponentes, kurios apibūdina apie 57,2 proc. duomenų sklaidos (žr. 23 pav.). Trečiosios pagrindinės komponentės (PK3) didžiausi svoriai buvo šių kintamųjų – išlaidų kultūrai, importo, eksporto, IT eksporto, politinio stabilumo, užimtumo lygio. Atvaizdavus k-vidurkių klasterizavimo rezultatus trimatėje erdvėje panaudojus pirmas tris pagrindines komponentes matyti, jog visi klasteriai gana aiškiai atsiskiria (žr. 23 pav.).



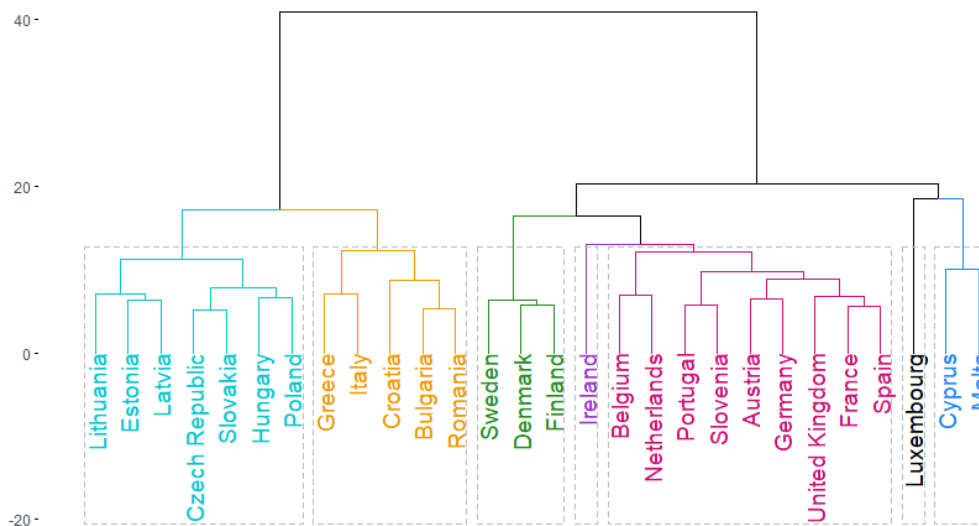
**23 pav.** K-vidurkių klasterizavimo rezultatai 2019 m. atvaizdavus pagal tris pirmas pagrindines komponentes

Atlikus ES šalių klasterinę analizę su k-vidurkių algoritmu pagal 2019 m. duomenis, paaiškėjo, jog šalių suskirstymas į klasterius buvo analogiškas 2013 m., todėl 2019 m. klasterizavimo rezultatai



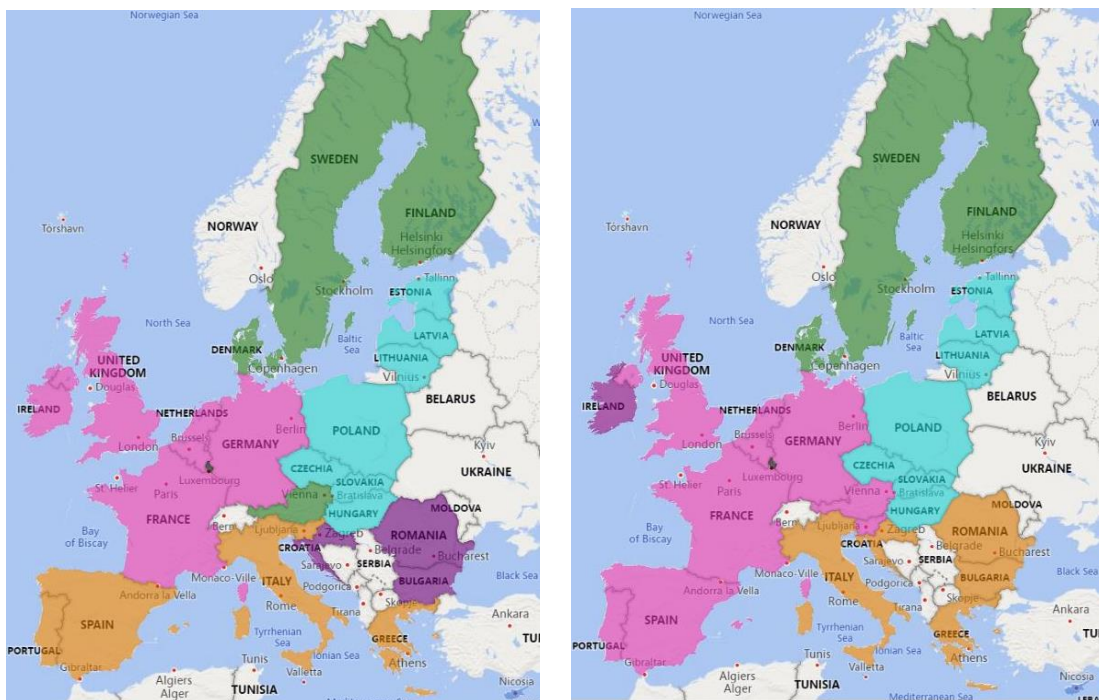
nebus detaliai aptariami, bus akcentuojami tik svarbiausi klasterių išskirtinimai (žr. 22 pav.). 5 klasteris, kuriam priskiriamas Liuksemburgas, kaip ir 2008, 2013, taip ir 2019 m. yra išskirtinai konkurencinga šalis. 3 klasteris, kurį sudaro Austrija, Danija, Suomija ir Švedija, yra taip pat vienas konkurencingiausių. Šis klasteris pasižymi ne tik aukštu aplinkosauginiu konkurencingumu, bet ir aukštais socialiniais, ekonominiais, šalies aplinkos rodikliais. 6 klasteris, kurį sudaro Belgija, Prancūzija, Vokietija, Airija, Nyderlandai, Jungtinė Karalystė, yra taip pat konkurencingas, bet nuo 5 ir 3 klasterio skiriasi žemesniais aplinkosauginiais, ekonominiais bei socialiniais rodikliais. 4 klasterį, kuris sudarytas iš Bulgarijos, Kroatijos, Rumunijos, galima įvardinti kaip mažiausiai konkurencingą. Nuo kitų klasterių šis klasteris labiausiai išsiskiria prastais ekonominiais ir ypač šalies aplinkos rodikliais – pastebima didelė šešėlinės ekonomikos dalis, korupcija, politinis nestabilumas – bei socialiniais rodikliais – neigiamas populiacijos prieaugis, maža besimokančių suaugusių dalis, mažiausia tikėtina gyvenimo trukmė, mažiausios išlaidos sveikatai ir socialinei apsaugai, didžiausia nepasiturinčių dalis, mažiausia aktyvios populiacijos dalis ir kt. Kipras ir Malta, kurie sudaro 2 klasterį, nuo kitų klasterių labiausiai išsiskiria prastesniais aplinkosauginiais rodikliais. 1 klasteris, kurį sudaro Čekija, Estija, Vengrija, Latvija, Lietuva, Lenkija, Slovakija, nuo kitų klasterių (išskyrus 4 klasterį) labiausiai išsiskiria prastesniais socialiniais rodikliais, pavyzdžiui, žmonių, kurių sveikatos būklė gera ar labai gera šiame klasteryje 2019 m. vidutiniškai buvo 56,47 proc., tuo tarpu kai kituose klasteriuose šis dydis svyravo nuo 66,3 iki 76 proc. 7 klasterį, kurį sudaro Graikija, Italija, Portugalija, Slovėnija, Ispanija, galima įvardinti kaip vidutinio konkurencingumo, lyginant su kitais klasteriais.

ES šalių klasterizavimas su 2019 m. duomenimis buvo atliktas panaudojus ir hierarchinį klasterizavimą, jo rezultatai yra pateikiami 24 pav. dendogramoje.



**24 pav.** Hierarchinio klasterizavimo dendograma pagal 2019 m. duomenis

Hierarchinio klasterizavimo būdu gautas šalių suskirstymas į klasterius bei k-vidurkių klasterizavimo rezultatų palyginimas yra pateikiamas toliau esančiame 25 pav. Lyginant k-vidurkių ir hierarchinio klasterizavimo klasterius galima pastebėti, jog yra trys klasteriai, kurie išliko nepakitę – Liuksemburgo (juoda); Baltijos šalių, Lenkijos, Čekijos, Slovakijos, Vengrijos (žalsvai mėlyna); Kipro, Maltos (mėlyna). Kituose klasteriuose pastebimi pakitimai.

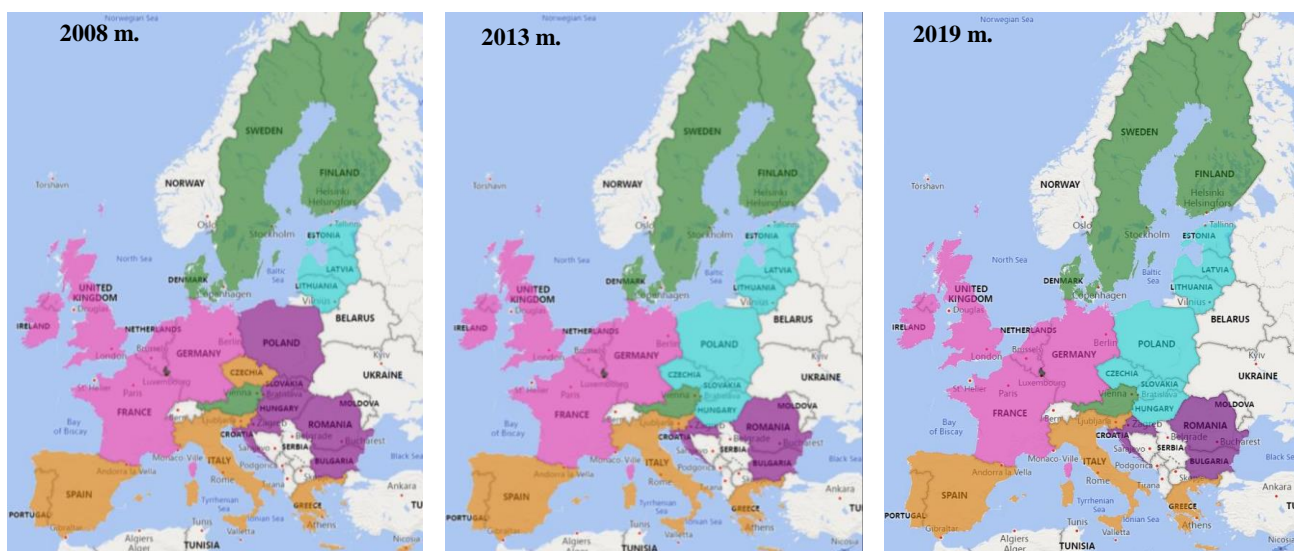


25 pav. 2019 m. k-vidurkių (kairėje) ir hierarchinio (dešinėje) klasterizavimo rezultatai

Hierarchinio klasterizavimo atveju vidutinis Silueto indeksas buvo 0,1291, o k-vidurkių šis indeksas buvo didesnis ir buvo lygus 0,14006.

### 3.2.4. Klasterizavimo rezultatų apibendrinimas

Šiame skyrelyje bus apibendrinti rezultatai gauti atlikus ES šalių klasterizavimą 2008, 2013 ir 2019 m. su k-vidurkių algoritmu. Šalys, žemėlapyje žymimos tomis pačiomis spalvomis, yra panašios pagal analizei pasirinktus rodiklius ir priklauso tam pačiam klasteriui (žr. 26 pav.).



26 pav. 2008, 2013 ir 2019 m. k-vidurkių klasterizavimo rezultatai žemėlapyje

Galima pastebėti, jog daugeliu atveju kaimyninės šalys priklauso tam pačiam klasteriui. Atlikus 2008, 2013 ir 2019 m. ES šalių klasterizavimą pastebėta, jog nepakito šie klasteriai: Liuksemburgo

(26 pav. žymima juoda spalva); Švedijos, Suomijos, Danijos ir Austrijos (26 pav. žymima žalia spalva); Belgijos, Prancūzijos, Vokietijos, Airijos, Nyderlandų, Jungtinės Karalystės (26 pav. žymima rožine spalva). Minėti klasteriai visą analizuojamą laikotarpį išlaikė konkurencinį pranašumą lyginant su kitais klasteriais. 2008 m. mažiausiai konkurencingi klasteriai buvo: Maltos (26 pav. žymima mėlyna spalva); Bulgarijos, Kroatijos, Vengrijos, Lenkijos, Rumunijos, Slovakijos (26 pav. žymima violetine spalva); Baltijos šalių (26 pav. žymima žalsvai mėlyna spalva). 2008 m. klasterį (26 pav. žymima oranžine spalva), kuris sudarytas iš Kipro, Čekijos, Graikijos, Italijos, Portugalijos, Slovėnijos, Ispanijos, būtų galima įvardinti kaip vidutinio konkurencingumo.

2013 m. ir 2019 m. gautas analogiškas šalių suskirstymas į klasterius. Remiantis vidutinėmis klasteryje esančių šalių požymių reikšmėmis, galima teigti, jog klasterį, kuris sudarytas iš Bulgarijos, Kroatijos, Rumunijos galima įvardinti kaip turintį mažiausią konkurencinį pranašumą. Dauguma šio klasterio rodiklių yra prastesni nei kitų klasterių, o ypač šalies aplinkos ir socialiniai rodikliai. Klasterį, kuris sudarytas iš Čekijos, Estijos, Vengrijos, Latvijos, Lietuvos, Lenkijos, Slovakijos, taip pat būtų galima įvardinti kaip vieną mažiau konkurencingų. Dar vienas klasteris, kuris yra mažiau konkurencingas – Graikijos, Italijos, Portugalijos, Slovėnijos, Ispanijos. Maltos ir Kipro klasterį būtų galima įvardinti kaip vidutinio konkurencingumo ir kurio žemas aplinkosauginis konkurencingumas.

### 3.3. Europos Sąjungos šalių konkurencingumo vertinimas

Šiame skyrelyje ES šalys bus vertinamos pagal apskaičiuotus konkurencingumo indeksus, išskiriant šalis, kurios įgyja didžiausią ir mažiausią konkurencinį pranašumą, taip pat akcentuojant svarbiausius rangų pokyčius analizuojamu 2008–2019 m. laikotarpiu. Vėliau bus vertinamas rangų sutapimo lygis tarp ES šalių rangavimo pagal apskaičiuotą bendrąjį konkurencingumo indeksą ir tarp ES šalių rangavimo pagal apskaičiuotą konkurencingumo indeksą, kuriame atrinkti svarbiausi požymiai, kai priklausomas kintamasis buvo BVP, tenkantis vienam gyventojui (*kintamasis gdp\_pc*).

#### 3.3.1. Vertinimas pagal apskaičiuotus konkurencingumo indeksus

Kiekvienų 2008–2019 metų kiekvieno kintamojo reikšmė buvo nomalizuojama. Vėliau kaip atitinkamų kintamųjų vidurkiai buvo apskaičiuojami ekonominis, socialinis, aplinkosauginis ir šalies aplinkos konkurencingumo indeksai. Bendrasis konkurencingumo indeksas buvo apskaičiuotas kaip anksčiau minėtų indeksų vidurkis, t. y. kiekvienam jų suteikiant vienodą 0,25 svorį.

Siekiant išsiaiškinti, ar atitinkami rodikliai gali būti derinami tarpusavyje sudarant aplinkosauginį, ekonominį, socialinį ir šalies aplinkos konkurencingumo indeksą, buvo skaičiuojams Cronbach alpha. Toliau pateikiama lentelė su gautais rezultatais.

**9 lentelė.** Cronbach alpha reikšmės atitinkamoms rodiklių grupėms 2008–2019 m.

Metai	Cronbach alpha reikšmės			
	Aplinkosauginių rodiklių	Ekonominių rodiklių	Socialinių rodiklių	Šalies aplinkos rodiklių
2008	0,4943	0,7796	0,7411	0,8602

2009	0,4027	0,8378	0,7642	0,8811
2010	0,3997	0,8266	0,7632	0,8613
2011	0,4457	0,8047	0,7565	0,8579
2012	0,4016	0,7994	0,7617	0,8510
2013	0,3800	0,8074	0,7428	0,8243
2014	0,2176	0,8483	0,7447	0,7551
2015	0,1147	0,8551	0,6958	0,7676
2016	0,3499	0,8339	0,6639	0,7493
2017	0,3382	0,7984	0,6818	0,8049
2018	0,3054	0,7460	0,6663	0,8340
2019	0,3404	0,7710	0,7124	0,8279

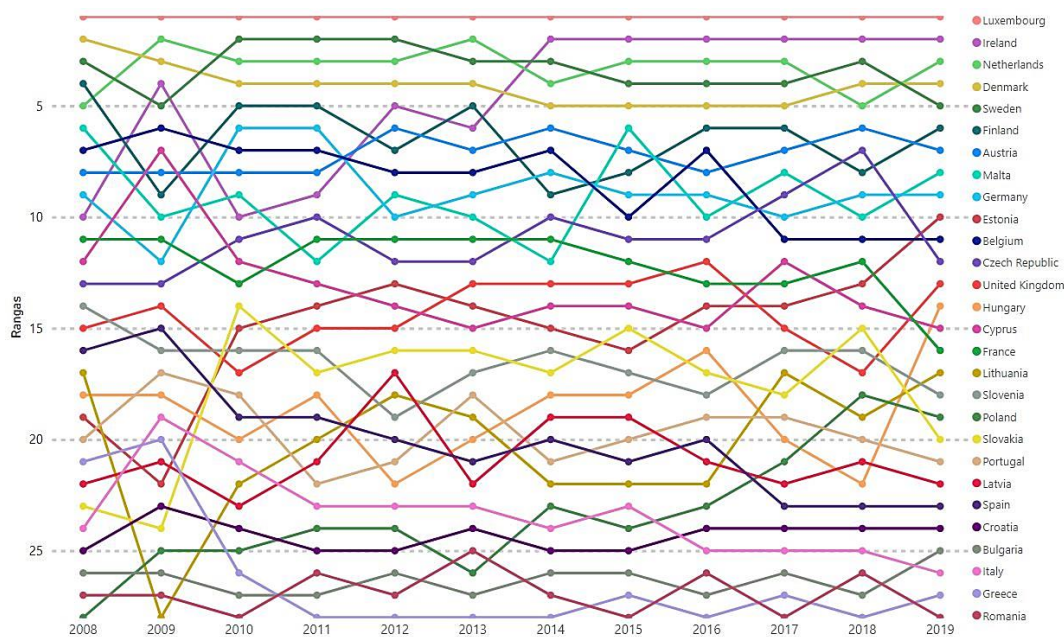
Iš 8 lentelės galime pastebėti, jog sudarant atitinkamus konkurencingumo indeksus, ekonominiai, socialiniai ir šalies aplinkos rodikliai visą analizuojamą laikotarpį gali būti derinami tarpusavyje, nes Cronbach alpha reikšmės yra didesnės už 0,7. Verta atkreipti dėmesį, jog aplinkosauginių rodiklių derinimas sudarant aplinkosauginio konkurencingumo indeksą yra abejotinas ir diskutuotinas, nes Cronbach alpha reikšmės yra mažesnės už 0,7, bet, kadangi šie rodikliai buvo surinkti atsižvelgiant į šalių konkurencingumą lemiančių veiksnių modelį, jie bus naudojami tolimesniame tyrime ir bus skaičiuojamas aplinkosauginis konkurencingumo indeksas. Kaip jau buvo minėta anksčiau, bendras konkurencingumo indeksas buvo apskaičiuotas naudojant ekonominį, socialinį, aplinkosauginį ir šalies aplinkos konkurencingumo indeksus kiekvienam jų suteikiant 0,25 svorį, tad Cronbach alpha buvo paskaičiuotas ir jiems, siekiant išsiaiškinti, kaip minėti indeksai dera tarpusavyje (žr. 10 lent.). Visais analizuotais metais Cronbach alpha reikšmės yra didesnės už 0,7.

**10 lentelė.** Cronbach alpha reikšmės 4-iems konkurencingumo indeksams 2008–2019 m.

Metai	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
<b>Cronbach alpha reikšmė</b>	0,865	0,869	0,857	0,872	0,858	0,843	0,842	0,821	0,825	0,824	0,83	0,829

Kiekvienais analizuojamais metais ES šalys buvo ranguojamos pagal anksčiau minėtus konkurencingumo indeksus suteikiant 1 rangą šaliai, kurios indekso reikšmė yra didžiausia ir suteikiant 28 rangą šaliai, kurios indekso reikšmė yra mažiausia. Toliau pateikiama informacija apie ES šalių rangus pagal apskaičiuotus konkurencingumo indeksus 2008–2019 m. Apskaičiuoti konkurencingumo indeksai yra pateikiami 10–14 prieduose.

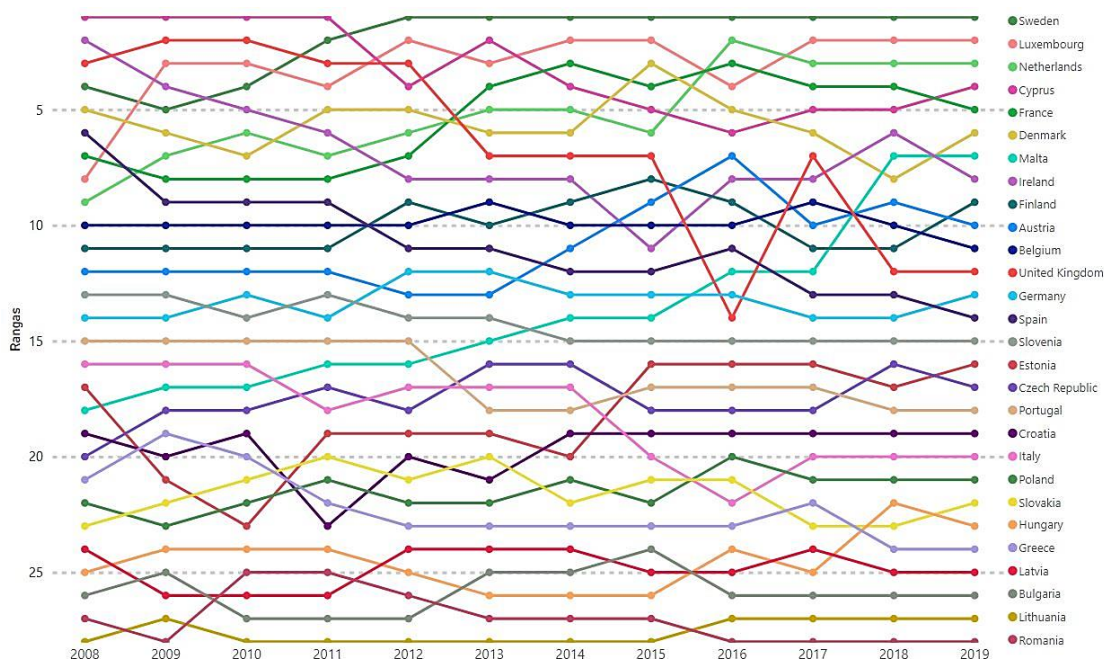
Analizuojamu 2008–2019 m. laikotarpiu konkurencingiausios šalys pagal apskaičiuotą ekonominį konkurencingumo indeksą buvo Liuksemburgas, Airija, Nyderlandai, Danija ir Švedija, o nekonkurencingiausios – Rumunija, Graikija, Italija, Bulgarija ir Kroatija (žr. 27 pav.).



27 pav. ES šalių ekonominio konkurencingumo rangai 2008–2019 m.

Liuksemburgo rangas visą analizuojamą laikotarpį išliko pastovus, šios šalies apskaičiuotas ekonominis konkurencingumo indeksas buvo didžiausias 2008–2019 m. lyginant jį su likusiomis ES šalimis (žr. 27 pav.). Galima pastebėti, jog įvyko ženklus Estijos, Airijos ir Lenkijos rangų pagerėjimas lyginant 2008 m. su 2019 m. Estijos teigiamas rango pokytis sudarė 9 (iš 19 į 10), Airijos – 8 (iš 10 į 2), Lenkijos – 9 (iš 28 į 19). Svarbu pastebėti ir neigiamą rangų pokytį. Ženkiausi neigiami rangų pokyčiai įvyko šių šalių: Ispanijos (iš 16 į 23), Graikijos (iš 21 į 27), Prancūzijos (iš 11 į 16).

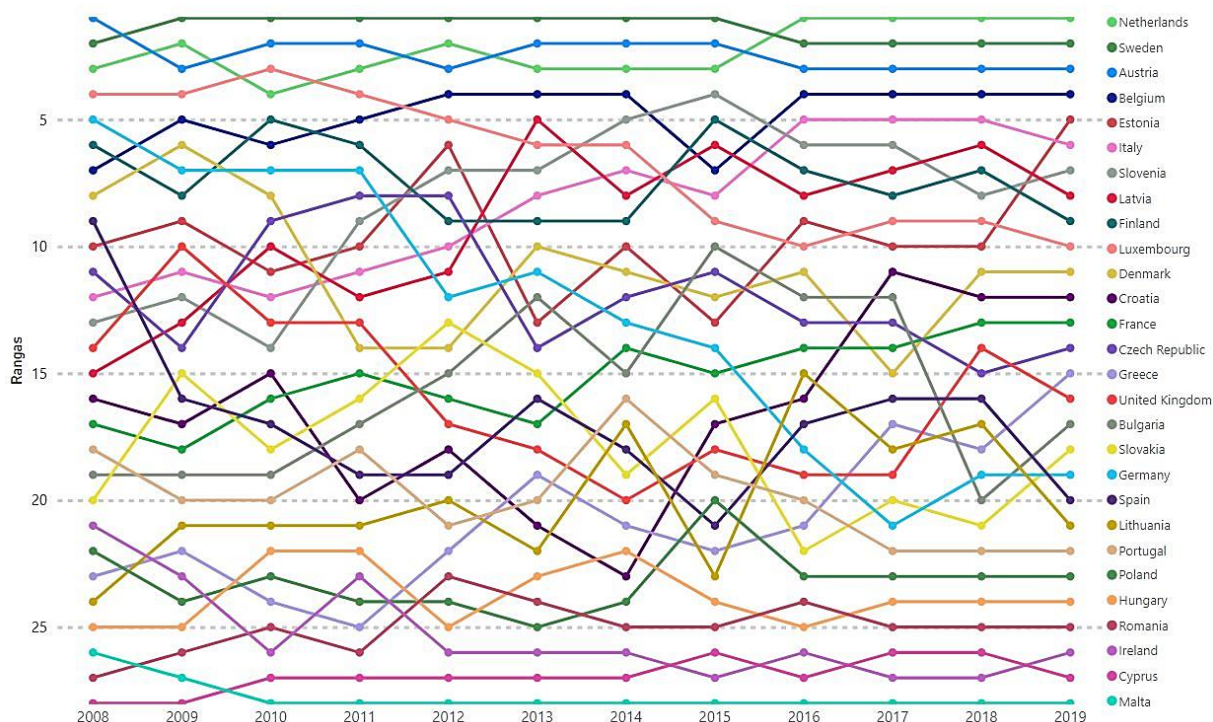
Galima pastebėti, jog stipriausios šalys, vertinant jas pagal apskaičiuotą socialinį konkurencingumo indeksą, yra Švedija, Liuksemburgas, Nyderlandai, Kipras, o silpniausios – Latvija, Bulgarija, Lietuva ir Rumunija (žr. 28 pav.).



28 pav. ES šalių socialinio konkurencingumo rangai 2008–2019 m.

Lyginant 2008 m. su 2019 m. verta atkreipti dėmesį, jog Maltos rangas pagal apskaičiuotą socialinį konkurencingumo indeksą pagerėjo per 11, iš 18 į 7 (žr. 28 pav.). Taip pat svarbu pastebėti, jog buvo šalių, kai šis rangas pasikeitė ženkliai neigiamai – Airijos, Ispanijos, Jungtinės Karalystės.

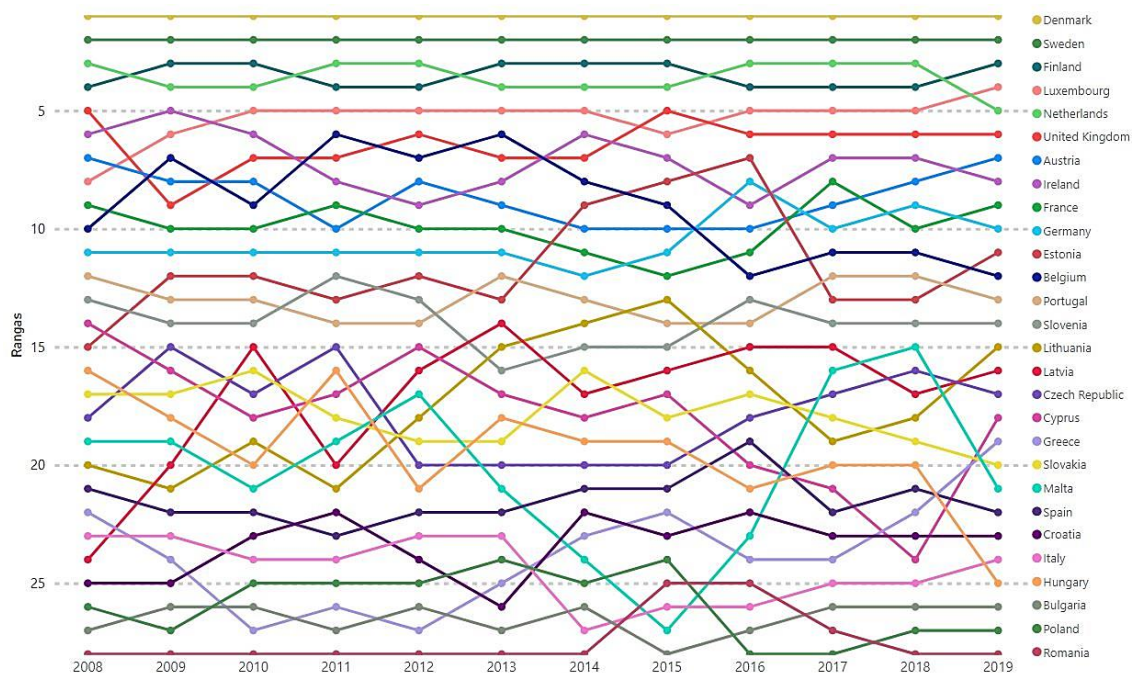
Vertinant ES šalis pagal apskaičiuotą aplinkosauginį konkurencingumo indeksą, galime matyti, jog 2019 m. stipriausios šalys yra Nyderlandai, Švedija, Austrija, Belgija ir Estija, o silpniausios – Vengrija, Rumunija, Airija, Kipras ir Malta (žr. 29 pav.).



**29 pav.** ES šalių aplinkosauginio konkurencingumo rangai 2008–2019 m.

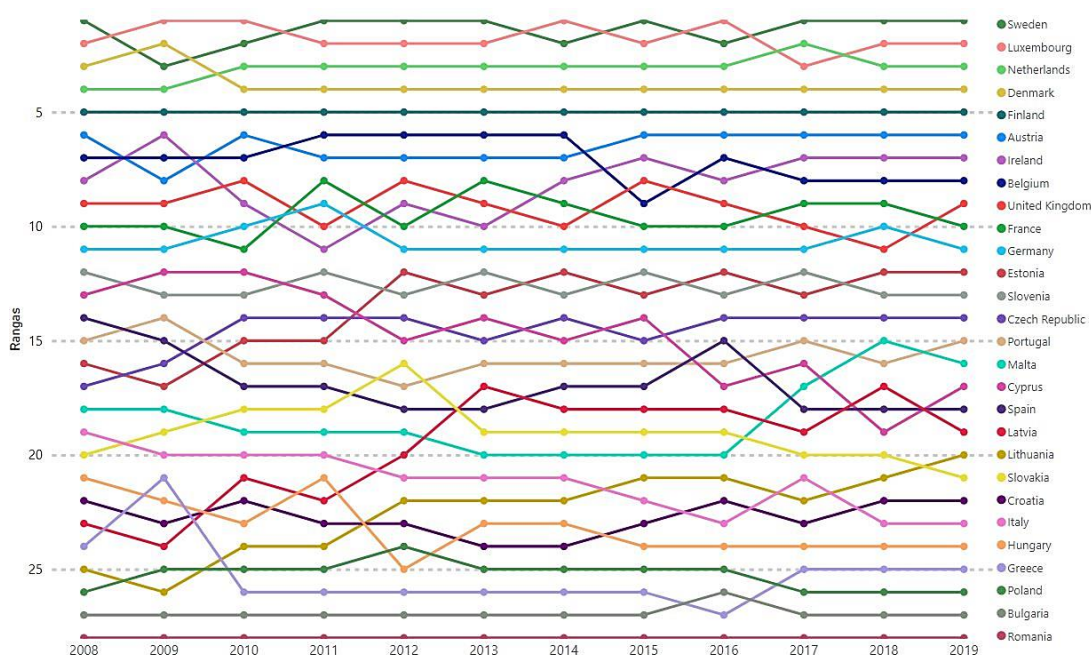
Pastebimas šių šalių ženklus rangų pagerėjimas vertinant jas pagal apskaičiuotą aplinkosauginį konkurencingumo indeksą ir lyginant 2008 m. su 2019 m.: Graikijos (iš 23 į 15), Slovėnijos (iš 13 į 7), Italijos iš (iš 12 į 6) bei Latvijos (iš 15 į 8). (žr. 29 pav.).

ES šalys 2008–2019 m. buvo ranguojamos ir pagal apskaičiuotą šalies aplinkos konkurencingumo indeksą. Šiuos rezultatus galime matyti 30 pav. Vertinant ES šalis pagal šį apskaičiuotą konkurencingumo indeksą pastebėta, jog didžiausias šalies aplinkos konkurencingumas buvo Danijoje, Švedijoje, Suomijoje, Liuksemburge ir Nyderlanduose. Atitinkamai galime atkreipti dėmesį, jog mažiausias šalies aplinkos konkurencingumas buvo Bulgarijoje ir Rumunijoje. Lyginant 2008 m. su 2019 m. matyti, jog ženkliausias teigiamas rangų pasikeitimas buvo Latvijos (iš 24 į 16) bei Lietuvos (iš 20 į 15). Taip pat verta atkreipti dėmesį, jog ir trečioje Baltijos šalyje – Estijoje – įvyko teigiamas rango pasikeitimas (iš 15 į 11). Vengrijoje analizuojamas rangas lyginant 2008 m. su 2019 m. ženkliai pablogėjo ir pasikeitė iš 16 į 25.



30 pav. ES šalių šalies aplinkos konkurencingumo rangai 2008–2019 m.

Kaip jau buvo minėta anksčiau, buvo sudarytas bendras konkurencingumo indeksas, kuris buvo apskaičiuotas remiantis ekonominiu, socialiniu, aplinkosauginiu ir šalies aplinkos konkurencingumo indeksais. ES šalys buvo ranguojamos pagal bendrąjį konkurencingumo indeksą (žr. 31 pav.). Galima pastebėti, jog analizuojamu laikotarpiu konkurencingiausios šalys buvo Švedija, Liuksemburgas, Nyderlandai, Danija ir Suomija (žr. 31 pav.). Mažiausiai konkurencingos ES šalys 2008–2019 m. – Rumunija, Bulgarija, Lenkija, Graikija ir Vengrija. Rumunija visą analizuojamą laikotarpį išliko paskutinėje pozicijoje. Lyginant 2008 m. su 2019 m. verta atkreipti dėmesį, jog įvyko ženklus teigiamas Baltijos šalių konkurencingumo rangų pokytis. 2008 m. Estija pagal bendrąjį konkurencingumo indeksą buvo 16 vietoje, o 2019 m. – 12 vietoje. 2008 m. Latvija užėmė 23 vietą, o 2019 m. – 19 vietą, o Lietuva 2008 m. – 25 vietą, o 2019 m. – 20 vietą.



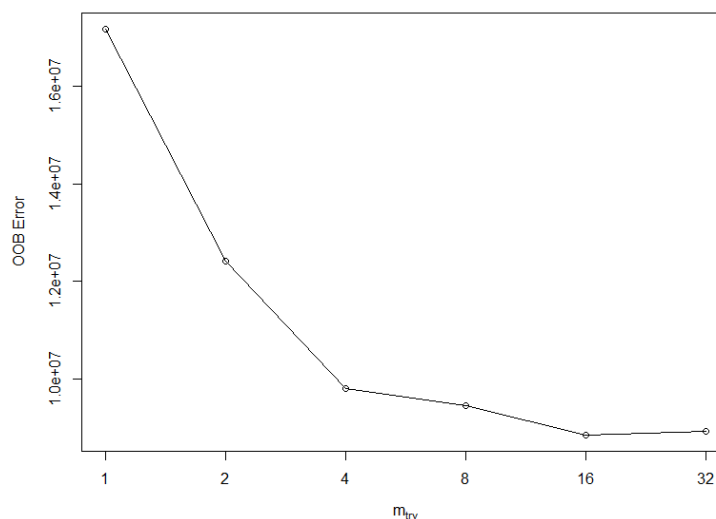
31 pav. ES šalių bendrojo konkurencingumo rangai 2008–2019 m.

Verta atkreipti dėmesį, jog buvo ir neigiamų ES šalių rangų pasikeitimų lyginant 2008 m. su 2019 m. (žr. 31 pav.). Tokie ženkliausi pasikeitimai buvo Kipro, Italijos ir Ispanijos.

### 3.3.2. Kintamųjų atranka algoritmu VSURF

Buvo daroma prielaida, jog konkurencingumas gali būti vertinamas vienu rodikliu – BVP, tenkančiu vienam gyventojui (kintamasis *gdp\_pc*). Šiame etape bus atrenkami svarbiausi kintamieji, kurie turi reikšmingos įtakos priklausomam kintamajam *gdp\_pc*. Svarbu paminėti, jog atrenkant svarbiausius rodiklius nebuvo naudojami kintamieji, kurie yra išreikšti per BVP. Atrinkti svarbiausi kintamieji vėliau bus naudojami sekančiame etape – konkurencingumo indekso apskaičiavimui ir ES šalių rangavimui. Bendrasis konkurencingumo indeksas, kuriam sudaryti buvo panaudoti visi kintamieji, bus lyginamas su konkurencingumo indeksu, kuriam sudaryti buvo panaudoti svarbiausi kintamieji, kurie turi reikšmingos įtakos kintamajam *gdp\_pc*. Svarbiausių kintamųjų atrankai bus panaudotas VSURF algoritmas (angl. *Variable Selection Using Random Forests*). Vėliau kiekvienais analizuojamais 2008–2019 m. bus vertinamas rangų sutapimo lygis, panaudojant Kendall'o konkordancijos koeficientą.

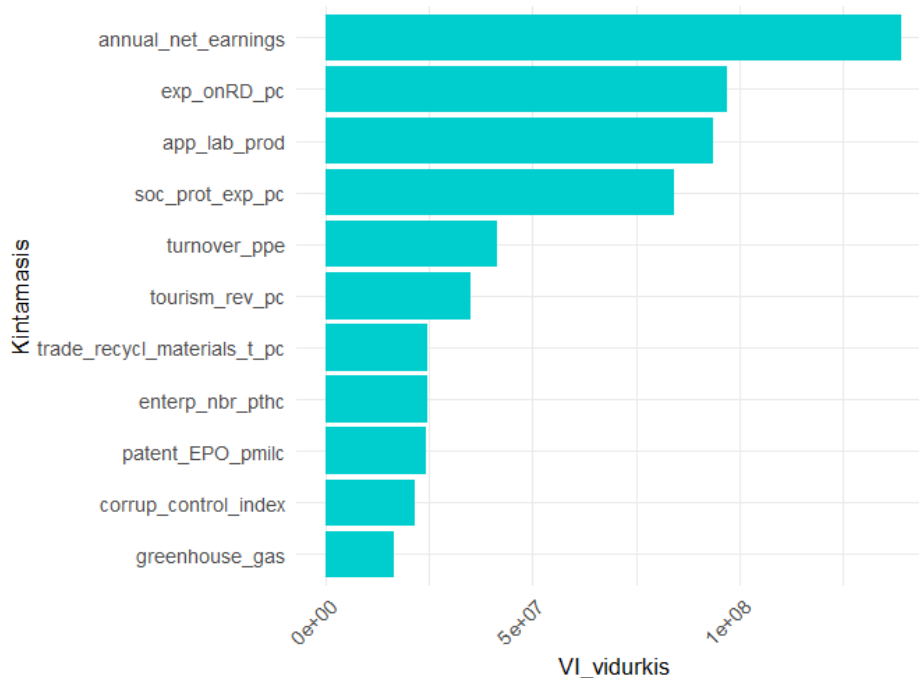
Buvo daroma prielaida, jog duomenys yra vienalaikiai (angl. *cross-sectional*). Siekiant atrinkti kintamuosius VSURF funkcijai reikėjo nurodyti parametrus. Vienas svarbiausių – *mtry*, kuris yra atsitiktinai atrinktų kintamųjų skaičius kiekviename medžio padalijime. Numatytas (angl. *default*) skaičius yra  $p/3$ , kur  $p$  – požymių skaičius. Konkrečiu atveju numatytasis *mtry* būtų buvęs 16. Siekiant parinkti optimalią reikšmę, buvo nuspręsta panaudoti *tuneRF* funkciją ir buvo gautas optimalus *mtry*, kuris taip pat yra lygus 16 (žr. 32 pav.). Šis parametras buvo nurodytas VSURF funkcijai.



32 pav. OOB klaidos prie skirtingų *mtry* reikšmių

Pirmame etape buvo atrinkti 47 kintamieji (iš 50), antrame – 11 ir trečiame – 11 kintamųjų, algoritmo sugaištas laikas 2,9 min. Svarbiausi kintamieji yra tie, kurie yra atrenkami trečiame etape, nes šio etapo metu surandamas mažesnis kiekis kintamųjų, kuris yra pakankamas priklausomo kintamojo prognozei. Toliau pateikiamas 33 pav., kuriame galima matyti atrinktus kintamuosius pagal svarbą.





33 pav. VSURF trečiame etape atrinkti kintamieji pagal mažėjančią svarbą

Atrinkti svarbiausi požymiai, turintys reikšmingos įtakos kintamajam *gdp\_pc* – metinis grynasis uždarbis (*annual\_net\_earnings*), mokslinių tyrimų ir plėtros išlaidos, tenkančios vienam gyventojui (*exp\_onRD\_pc*), bendroji pridėtinė vertė, tenkanti vienam dirbančiam (*app\_lab\_prod*), socialinės apsaugos išlaidos, tenkančios vienam gyventojui (*soc\_prot\_exp\_pc*), apyvarta vienam dirbančiam (*turnover\_ppe*), tarptautinio turizmo pajamos vienam gyventojui (*tourism\_rev\_pc*), prekyba perdirbamomis žaliavomis, tonos gyventojui (*trade\_recycl\_materials\_t\_pc*), įmonių skaičius, tenkantis 1000 gyventojų (*enterp\_nbr\_pthc*), patentų paraiškų Europos patentų tarnybai skaičius, tenkantis 1 mln. gyventojų (*patent\_EPO\_pmilc*), korupcijos kontrolės indeksas (*corrup\_control\_index*), šiltnamio efektą sukeliančių dujų emisija, tenkanti vienam gyventojui (*greenhouse\_gas*). Panaudojus šiuos svarbiausius kintamuosius atsitiktinio miško modelio, kuris galėtų prognozuoti *gdp\_pc*, sudarymui, buvo gauta, jog šis modelis paaiškina 97,76 proc. stebėto priklausomo kintamojo sklaidos (vertinant ant imties, kuri nebuvo naudota sprendimų medžių sudarymui).

VSURF algoritmo pagalba daugiausia buvo atrinkta rodiklių, kurie apibūdina ekonominį konkurencingumą ir jų skaičius yra 7. Socialinį konkurencingumą apibūdinti buvo atrinktas 1 rodiklis, šalies aplinkos konkurencingumą apibūdinančių – 1 ir aplinkosauginį konkurencingumą apibūdinančių – 2. Kaip jau buvo minėta anksčiau, šie rodikliai bus naudojami konkurencingumo indekso apskaičiavimui.

### 3.3.3. ES šalių rangų, gautų skirtingais metodais, suderinamumas

Kiekvienais analizuojamais 2008–2019 m. ES šalys buvo ranguojamos (aukščiausią reikšmę surinkusiai šaliai buvo suteikiamas 1 rangas, antrai pagal didumą – 2 ir t. t.) pagal tris rodiklius:

- bendrąjį konkurencingumo indeksą, kuriam apskaičiuoti buvo panaudoti visi kintamieji;
- konkurencingumo indeksą, kuriam apskaičiuoti buvo panaudoti kintamieji, atrinkti VSURF algoritmo pagalba;

- BVP, tenkantį vienam gyventojui.

ES šalių rangavimas pagal aukščiau paminėtus rodiklius yra pateikiamas 15 priede. Toliau pateikiamas 34 pav., kuris iliustruoja ES šalių rangavimą 2019 m. pagal aukščiau minėtus 3 rodiklius.



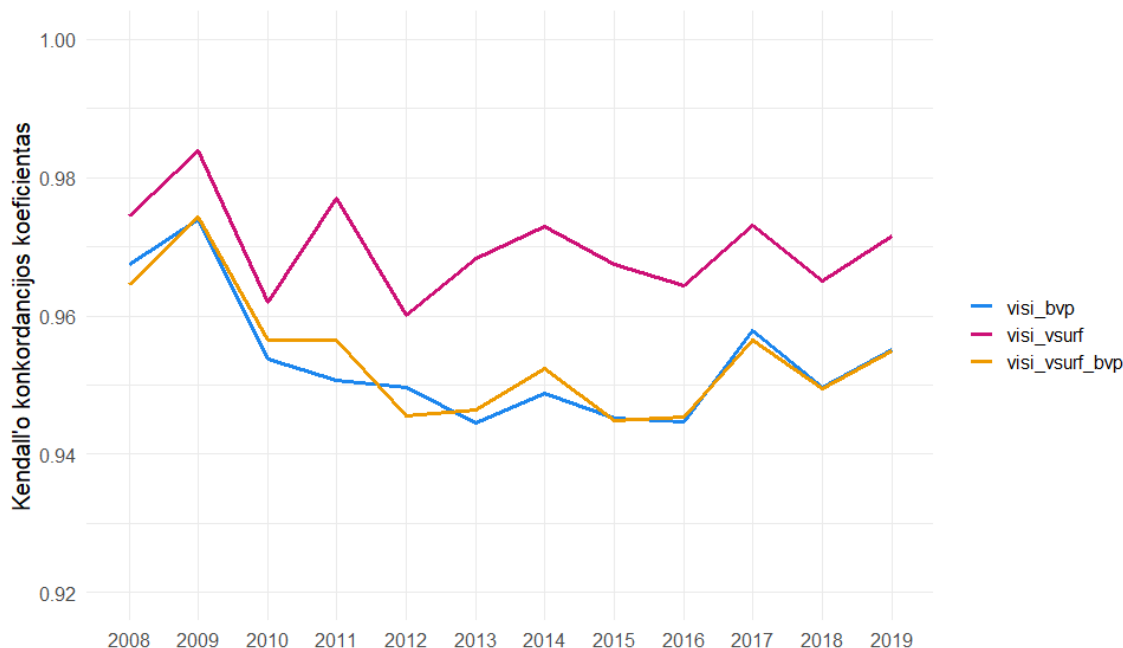
**34 pav.** ES šalių rangų, gautų pagal 3 būdus, palyginimas 2019 m.

Galima pastebėti, jog yra rangų sutapimų (žr. 34 pav.). Austrijos rangas pagal visus tris rodiklius yra vienodas ir yra lygus 6. Apskaičiavus bendrąjį konkurencingumo indeksą naudojant visus kintamuosius, Belgijai buvo suteiktas 8 rangas, toks pat rangas šiai šaliai buvo suteiktas apskaičiavus konkurencingumo indeksą naudojant rodiklius, atrinktus VSURF metodu pagalba. Analogiška situacija yra ir Prancūzijos, Graikijos, Slovakijos, Jungtinės Karalystės. Apskaičiavus bendrąjį konkurencingumo indeksą naudojant visus kintamuosius, Lietuvai buvo suteiktas 20 rangas, toks pat rangas šiai šaliai buvo suteiktas ir pagal BVP, tenkantį gyventojui.

Buvo vertinamas ES šalių rangų, gautų pagal skirtingus būdus, sutapimo lygis. Lyginimui buvo naudotas Kendall'o konkordancijos koeficientas, apskaičiuotos koeficientų reikšmės ir p-reikšmės yra pateikiamos 16 priede. Lyginimas buvo atliekamas 2008–2019 m. Buvo vertinamas rangų sutapimo lygis, kai:

- bendrojo konkurencingumo indekso (naudojant visus rodiklius) rangas buvo lyginamas su konkurencingumo indekso (naudojant VSURF atrinktus rodiklius) rangą;

- bendrojo konkurencingumo indekso (naudojant visus rodiklius) rangas buvo lyginamas su BVP, tenkančiu gyventojui, rangui.
- bendrojo konkurencingumo indekso (naudojant visus rodiklius) rangas buvo lyginamas su konkurencingumo indekso (naudojant VSURF atrinktus rodiklius) rangui ir su BVP, tenkančiu gyventojui, rangui.



**35 pav.** Kendall'o konkordancijos koeficientų kitimo dinamika 2008–2019 m.

Galima pastebėti, jog didžiausias rangų sutapimo lygis pastebimas lyginant bendrojo konkurencingumo indekso (naudojant visus rodiklius) rangą su konkurencingumo indekso (naudojant VSURF atrinktus rodiklius) rangui (žr. 35 pav.). Šis konkordancijos koeficientas (35 pav. žymima *visi\_vsurf*) analizuojamą 2008–2019 m. laikotarpį yra didesnis nei 0,96 ir tai reiškia, jog rangų sutapimo lygis yra ištis aukštas. Mažesnis rangų sutapimo lygis pastebimas lyginant bendrojo konkurencingumo indekso (naudojant visus rodiklius) rangą su BVP, tenkančiu gyventojui, rangui. Šis konkordancijos koeficientas (35 pav. žymimas *visi\_bvp*) analizuojamą 2008–2019 m. laikotarpį yra didesnis nei 0,94 ir tai taip rodo, jog rangų sutapimo lygis taip pat yra aukštas, tačiau mažesnis nei prieš tai minėtos rangų poros. Lyginant visus tris rangus (35 pav. žymima *visi\_vsurf\_bvp*), galima pastebėti, jog analizuojamą laikotarpį Kendall'o konkordancijos koeficientas taip pat buvo aukštas ir viršijo 0,94.

Apibendrinant būtų galima teigti, jog vertinant ES šalių konkurencingumą galima naudoti ir mažesnių skaičių rodiklių, nes rangų suderinamumas tarp konkurencingumo indeksų, kai naudojami visi rodikliai ir, kai naudojami VSURF atrinkti rodikliai, yra ištis aukštas. Taip pat būtų galima paminėti, jog konkurencingumą vertinti vien tik pagal BVP, tenkantį gyventojui, nėra taip tikslu, nes vertinant konkurencingumą pagal indeksą, kuriam buvo panaudoti VSURF atrinkti kintamieji, analizuojamą 2008-2019 m. laikotarpį buvo gaunamas aukštesnis Kendall'o konkordancijos koeficientas.

## Išvados

1. Atlikus mokslinės literatūros analizę paaiškėjo, jog nėra vieningos konkurencingumo vertinimo metodikos. Remiantis mokslininkų išskirtais veiksniais buvo sudarytas šalių konkurencingumą lemiančių veiksnių modelis, kuris susideda iš 4 dalių – ekonominio, socialinio, šalies aplinkos bei aplinkosauginio konkurencingumo. Sudarytas šalių konkurencingumą lemiančių veiksnių modelis gali būti panaudojamas atliekant ir kitus tyrimus, vertinant kitų šalių konkurencingumą.
2. Siekiant išsaugoti vertingus duomenis, kuriuose buvo trūkstamų reikšmių, buvo pritaikytas *missForest* algoritmas. Analizuojant užpildytą duomenų rinkinį buvo pastebėta, jog kintamųjų, kurių trūkstamos reikšmės buvo užpildytos, minimalios ir maksimalios reikšmės nepakito, užpildytos reikšmės atitiko bendrą tendenciją.
3. Atrinkti rodikliai pagal šalių konkurencingumą lemiančių veiksnių modelį leido suklasterizuoti šalis pagal jų panašumus. Tokiu būdu buvo identifikuoti konkurencingiausi ir mažiausiai konkurencingi klasteriai bei išryškinti skirtumai tarp klasterių. Buvo pastebėta, jog daugeliu atvejų klasterizavimo rezultatams įtakos turėjo geografinė šalies vieta, nors tiesiogiai į tyrimą ji nebuvo įtraukta. Konkurencingiausi klasteriai 2008, 2013 ir 2019 m. išliko nepakitę.
4. Išrangavus Europos Sąjungos šalis pagal apskaičiuotą bendrąjį konkurencingumo indeksą 2008–2019 m., paaiškėjo, jog didžiausią konkurencinį pranašumą analizuojamu laikotarpiu įgijo Švedija, Liuksemburgas, Nyderlandai, Danija ir Suomija. Mažiausias konkurencinis pranašumas analizuojamu laikotarpiu buvo šių šalių – Rumunijos, Bulgarijos, Lenkijos, Graikijos. Siekiant padidinti Europos Sąjungos, kaip vieno teritorinio vieneto, konkurencingumą pasaulyje, turėtų būti mažinami skirtumai tarp šalių narių, ypač atsižvelgiant į žemiausiose pozicijose esančias šalis.
5. Padarius prielaidą, jog konkurencingumas gali būti vertinamas vienu makroekonominiu rodikliu – bendruoju vidaus produktu, tenkančiu vienam gyventojui – buvo atrinkti svarbiausi kintamieji, kurie turi reikšmingos įtakos jam. Svarbiausių kintamųjų atrankai buvo panaudotas mašininio mokymosi algoritmas *VSURF*. Iš viso buvo atrinkta 11 rodiklių, iš kurių daugiausia – ekonominių.
6. Palyginus 2008–2019 m. Europos Sąjungos šalių rangus buvo pastebėta, jog didžiausias rangų suderinamumas, vertinant pagal Kendall'o konkordancijos koeficientą, buvo gautas lyginant bendrojo konkurencingumo indekso (naudojant visus rodiklius) rangą su konkurencingumo indekso (naudojant *VSURF* atrinktus rodiklius) rangą ir šis rodiklis analizuojamu laikotarpiu buvo didesnis už 0,96. Europos Sąjungos šalių konkurencingumui vertinti galėtų būti naudojamas ir mažesnis skaičius kintamųjų, o bendrasis vidaus produktas buvo svarbus rodiklis juos identifikuojant.

## Rekomendacijos

1. Kadangi Europos Sąjunga siekia neutralizuoti poveikį klimatui, ypač svarbu atsižvelgti į tas šalis, kurių apskaičiuotas aplinkosauginio konkurencingumo indeksas šiame tyrime buvo mažiausias visą analizuojamą 2008–2019 m. laikotarpį. Siekiant užtikrinti, jog Europos Sąjungos perėjimas prie tvarumo būtų sklandus, reikėtų didesnę dėmesį skirti Maltai, Kipriui, Airijai.
2. Šis tyrimas atskleidė, jog Bulgarija ir Rumunija buvo mažiausiai konkurencingos šalys vertinant jas pagal apskaičiuotą bendrąjį konkurencingumo indeksą. Buvo pastebėta, jog šios šalys buvo silpnos tiek pagal ekonominius, tiek pagal socialinius ir šalies aplinkos rodiklius. Būtina imtis priemonių, kurios padėtų šioms šalims sustiprėti. Atsakingiems asmenims priėmus teisingus sprendimus galima tikėtis sėkmingesnės visos Europos Sąjungos plėtros.
3. Trūkstatų reikšmių užpildymui panaudojus *missForest* algoritmą nebuvo galimybės įvertinti, ar užpildytos reikšmės atitinka realius duomenis. Nors užpildyti duomenys buvo korektiški (pavyzdžiui, požymiai neįgijo neigiamų reikšmių, kai jie galėjo įgyti tik teigiamas reikšmes) ir atitiko tendenciją, atsiradus galimybei gauti 2019 m. trūkstatamus duomenis, būtų naudinga įvertinti *missForest* algoritmo užpildytų reikšmių tikslumą, nuokrypį nuo tikrosios reikšmės. Taip pat būtų svarbu atlikti pakartotinę 2019 m. Europos Sąjungos šalių klasterinę analizę su realiais duomenimis ir patikrinti, ar pasikeistų šalių suskirstymas į klasterius, ar jis visgi išliktų toks pat.
4. Siekiant išskirti svarbiausius konkurencingumo rodiklius, kaip priklausomas kintamasis šiame tyrime buvo panaudotas bendrasis vidaus produktas. Atliekant tolimesnius tyrimus būtų galima panaudoti ir kitus indikatorius tokius, kaip, pavyzdžiui, eksportas ar tiesioginės užsienio investicijos, kurie taip pat galėtų būti svarbūs kintamieji identifikuojant esminius konkurencingumą lemiančius rodiklius.
5. Šis tyrimas patvirtino, jog egzistuoja stipresnių ir silpnesnių Europos Sąjungos šalių klasteriai. Konkurencingumo rangų, gautų pagal skirtingus metodus, suderinamumas buvo vertinamas bendrai visose Europos Sąjungos šalyse. Būtų naudinga atlikti detalesnę tyrimą, kuris leistų įvertinti rangų sutapimo lygį atskirose dvejose šalių grupėse – mažiau konkurencingose ir daugiau konkurencingose.

## Literatūros sąrašas

1. CELLINI, R. ir A. SOCI. Pop competitiveness. *PSL Quarterly Review* [interaktyvus]. 2002, **55**(220), 71–101 [žiūrėta 2021-01-26]. ISSN 2037-3635. Prieiga per: DOAJ
2. PETRYLĖ, V. Does the Global Competitiveness Index Demonstrate the Resilience of Countries to Economic Crises? *Ekonomika* [interaktyvus]. 2017, **95**(3), 28–36 [žiūrėta 2021-01-12]. ISSN 2424-6166. Prieiga per: doi:10.15388/Ekon.2016.3.10326
3. BALKYTĖ, A. ir M. TVARONAVIČIENĖ. Perception of competitiveness in the context of sustainable development: Facets of “sustainable competitiveness”. *Journal of Business Economics and Management* [interaktyvus]. 2010, **11**(2), 341–365 [žiūrėta 2021-02-16]. ISSN 1611-1699. Prieiga per: doi:10.3846/jbem.2010.17
4. JAN, N. Visegrad Four Countries – Case Study of Econometric Panel Data Model for Regional Competitiveness Evaluation. *Journal of competitiveness* [interaktyvus]. 2012, **4**(4), 3–15 [žiūrėta 2021-01-19]. ISSN 1804-1728. Prieiga per: doi:10.7441/joc.2012.04.01
5. SCHWAB, K. *The Global Competitiveness Report 2019* [interaktyvus]. 2019. Prieiga per: [http://www3.weforum.org/docs/WEF\\_TheGlobalCompetitivenessReport2019.pdf](http://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf)
6. KITSON, M., R. MARTIN ir P. TYLER. Regional Competitiveness: An Elusive yet Key Concept? *Regional Studies* [interaktyvus]. 2004, **38**(9), 991–999 [žiūrėta 2021-01-22]. ISSN 0034-3404. Prieiga per: doi:10.1080/0034340042000320816
7. SIMANAVIČIENĖ, Ž., J. BRUNECKIENĖ ir I. ŠIMBEROVA. Regionų konkurencingumo vertinimo teoriniai aspektai. *Ekonomika* [interaktyvus]. 2007, **77**, 68–81 [žiūrėta 2021-01-26]. ISSN 2424-6166. Prieiga per: doi:10.15388/Ekon.2007.17601
8. GRYNIUK, D. Social-geographical essence and content of the competitiveness of the region. *Časopis social'no-ekonomičnoï geografii* [interaktyvus]. 2017, **22**(1), 141–147 [žiūrėta 2021-01-22]. ISSN 2076-1333. Prieiga per: doi:10.26565/2076-1333-2017-22-21
9. VUKOVIC, D. ir L. WEI. Regional competitiveness: The case of Western China. *Journal of the Geographical Institute „Jovan Cvijic“, SASA* [interaktyvus]. 2010, **60**(1), 107–124 [žiūrėta 2021-01-19]. ISSN 0350-7599. Prieiga per: doi:10.2298/IJGI1001107V
10. CLIPA, R.I. ir M. IFRIM. Measuring regional competitiveness. The case of Romania. *Analele Universității din Oradea. Științe economice* [interaktyvus]. 2016, **25**(1), 103–111 [žiūrėta 2021-01-18]. ISSN 1222-569X. Prieiga per: <http://anale.steconomieuoradea.ro/volume/2016/n1/11.pdf>
11. DIJKSTRA, L., P. ANNONI ir K. KOZOVSKA. *A New Regional Competitiveness Index: Theory, Methods and Findings* [interaktyvus]. 2011. Prieiga per: [https://ec.europa.eu/regional\\_policy/en/information/publications/working-papers/2011/a-new-regional-competitiveness-index-theory-methods-and-findings](https://ec.europa.eu/regional_policy/en/information/publications/working-papers/2011/a-new-regional-competitiveness-index-theory-methods-and-findings)
12. NIKOLIC, M., M. FILIPOVIC ir S. POKRAJAC. Regional competitiveness for achieving sustainable development: The case of Serbia. *Industrija* [interaktyvus]. 2016, **44**(3), 7–26 [žiūrėta 2021-01-18]. ISSN 0350-0373. Prieiga per: doi:10.5937/industrija44-10615
13. VUKOVIC, D., D. MARKOVIC ir A. HANIC. Snowflake model of regional competitiveness: Evidence from Serbia. *International Review* [interaktyvus]. 2016, **2016**(1–2), 59–73 [žiūrėta 2021-01-18]. ISSN 2217-9739. Prieiga per: doi:10.5937/intrev1602059V
14. ŽITKUS, L. New Approach to Regional Competitiveness as Instrument of Regional Development. *European Integration Studies* [interaktyvus]. 2015, (9), 114–126 [žiūrėta 2021-01-19]. ISSN 2335-8831. Prieiga per: doi:10.5755/j01.eis.0.9.12800
15. ŽITKUS, L. ir M. MICKEVIČIENĖ. Konkurencingumas kaip regiono plėtros siekinys. *VIEŠOJI POLITIKA IR ADMINISTRATIVAS* [interaktyvus]. 2013, **12**(3), 430–441 [žiūrėta 2021-01-21]. ISSN 2029-2872. Prieiga per: doi:10.5755/j01.ppa.12.3.4008
16. TAHIR, N. ir P. TAHIR. Does competition explain growth in OECD and BRICS countries? *Competitiveness Review: An International Business Journal* [interaktyvus]. 2019, **29**(5), 515–533 [žiūrėta 2021-02-12]. ISSN 1059-5422. Prieiga per: doi:10.1108/CR-10-2018-0063
17. CARVALHO, H.D. ir H.T. FONSECA. A multivariate data analysis of logistics structure as

- a factor of competitiveness. *Brazilian Journal of Operations & Production Management* [interaktyvus]. 2017, **14**(1), 90–100 [žiūrėta 2021-XX-XX]. ISSN 2237-8960. Prieiga per: 10.14488/BJOPM.2017.v14.n1.a10
18. BAYAR, Y., R. REMEIKIENE, A. ANDRONICEANU, L. GASPARENIENE ir R. JUCEVICIUS. The Shadow Economy, Human Development and Foreign Direct Investment Inflows. *Journal of Competitiveness* [interaktyvus]. 2020, **12**(1), 5–21 [žiūrėta 2021-02-17]. ISSN EISSN: 1804-1728. Prieiga per: doi:10.7441/joc.2020.01.01
  19. GALGÁNKOVÁ, V. Competitiveness of V4 Countries Using the Global Competitiveness Index. *SHS Web of Conferences* [interaktyvus]. 2020, **74**, 1–9 [žiūrėta 2021-02-12]. ISSN 2261-2424. Prieiga per: doi:10.1051/shsconf/20207406007
  20. SCHWAB, K. *The Global Competitiveness Index Report 2017-2018* [interaktyvus]. 2017. Prieiga per: <https://www.weforum.org/reports/the-global-competitiveness-report-2017-2018>
  21. PORTER, M.E. *The competitive advantage of nations : with a new introduction*. New York: The Free Press, 1998. ISBN 0-684-84147-9.
  22. RUCINSKA, S. ir R. RUCINSKY. Factors of regional competitiveness. *2nd Central European Conference in Regional Science – CERS* [interaktyvus]. 2007, p. 902–911. Prieiga per: Research Gate
  23. GINEVICIUS, R., T. KLIESTIK, A. STASIUKYNAS ir K. SUHAJDA. The Impact of National Economic Development on the Shadow Economy. *Journal of Competitiveness* [interaktyvus]. 2020, **12**(4), 39–55 [žiūrėta 2021-02-19]. ISSN 1804171X. Prieiga per: doi:10.7441/joc.2020.04.03
  24. SNIEŠKA, V. ir J. BRUNECKIENĖ. Measurement of Lithuanian regions by regional competitiveness index. *Engineering economics* [interaktyvus]. 2009, **61**(1), 45–57 [žiūrėta 2021-02-08]. ISSN 1392-2785. Prieiga per: doi:10.5755/j01.ee.61.1.11584
  25. RUZEKOVA, V., Z. KITTOVA ir D. STEINHAUSER. Export Performance as a Measurement of Competitiveness. *Journal of Competitiveness* [interaktyvus]. 2020, **12**(1), 145–160 [žiūrėta 2021-02-10]. ISSN 1804171X. Prieiga per: doi:10.7441/joc.2020.01.09
  26. JANUŠKAITĖ, V. ir L. UŽIENĖ. Intellectual Capital as a Factor of Sustainable Regional Competitiveness. *Sustainability* [interaktyvus]. 2018, **10**(12), 4848 [žiūrėta 2021-01-26]. ISSN 2071-1050. Prieiga per: doi:10.3390/su10124848
  27. DANILEVIČIENĖ, I. ir V. LUKŠYTĖ. The Assessment of Foreign Direct Investment Influence on The Country’s Economic Competitiveness. *Mokslas - Lietuvos ateitis / Science – Future of Lithuania* [interaktyvus]. 2017, **9**(2), 183–196 [žiūrėta 2021-02-20]. ISSN 2029-2341. Prieiga per: doi:10.3846/mla.2017.1017
  28. SKARE, M. ir D. CVEK. A Vector Autoregression Analysis of Foreign Direct Investment and Its Link to Competitiveness. *Journal of Competitiveness* [interaktyvus]. 2020, **12**(4), 127–142 [žiūrėta 2021-02-17]. ISSN 1804171X. Prieiga per: doi:10.7441/joc.2020.04.08
  29. RUSSU, C. Determinants of Regional Competitiveness. Status of Factors Determining Comparative Advantages to Romanian Development Regions. *Economic Insights - Trends & Challenges* [interaktyvus]. 2019, (1), 1–10 [žiūrėta 2021-01-18]. ISSN 22848576. Prieiga per: EBSCO
  30. DROZDZ, J. ir A. VOLKOV. Regional Economic Systems: Competitiveness and Innovative Development. *The 7th International Scientific Conference „Business and Management 2012“. Selected papers* [interaktyvus]. Vilnius, Lithuania: Vilnius Gediminas Technical University Publishing House Technika, 2012, p. 351–358. ISBN 9786094571169. Prieiga per: doi:10.3846/bm.2012.046
  31. ZINOVYEVA, I.S., Y.A. KOZENKO, K.B. GERASIMOV, Y.I. DUBOVA ir M.S. IRIZEPOVA. Regional Innovation Development as a Feature of Competitiveness in the XXI Century. *Contemporary Economics* [interaktyvus]. 2016, **10**(4), 333–342 [žiūrėta 2021-02-02]. ISSN 20840845. Prieiga per: doi:10.5709/ce.1897-9254.220
  32. CHEBA, K., I. BAŲ ir K. SZOPIK-DEPCZYŃSKA. SUSTAINABLE COMPETITIVENESS AS A NEW ECONOMIC CATEGORY – DEFINITION AND

- MEASUREMENT ASSESSMENT. *Technological and Economic Development of Economy* [interaktyvus]. 2020, **26**(6), 1399–1421 [žiūrėta 2021-02-05]. ISSN 2029-4913. Prieiga per: doi:10.3846/tede.2020.13528
33. WENZEL, L. ir A. WOLF. Towards a new measure of a country's competitiveness: applying canonical correlation. *Competitiveness Review* [interaktyvus]. 2016, **26**(1), 87–107 [žiūrėta 2021-02-12]. ISSN 1059-5422. Prieiga per: doi:10.1108/CR-09-2014-0030
  34. SZYMAŃSKA, K. Factors shaping competitiveness of the region. *Megatrend revija : međunarodni časopis za primenjenu ekonomiju* [interaktyvus]. 2014, **11**(3), 239–248 [žiūrėta 2021-01-31]. ISSN 1820-3159. Prieiga per: doi:10.5937/MegRev1403239S
  35. IGNATOV, A. Institutional Efficiency, Entrepreneurship, and the Premises of Economic Development in the Eastern European Countries. *Studia Universitatis Babes-Bolyai Oeconomica* [interaktyvus]. 2019, **64**(2), 12–32 [žiūrėta 2021-02-10]. ISSN 2065-9644. Prieiga per: doi:10.2478/subboec-2019-0007
  36. POPA, F. Regional competitiveness, conceptual elements. *Studies and Scientific Researches. Economics Edition (University of Bacau)* [interaktyvus]. 2015, (22), 14–22 [žiūrėta 2021-01-26]. ISSN 2344-1321. Prieiga per: doi:10.29358/sceco.v0i22.334
  37. GARDINER, B., R. MARTIN ir P. TYLER. Competitiveness, Productivity and Economic Growth across the European Regions. *Regional Studies* [interaktyvus]. 2004, **38**(9), 1045–1067 [žiūrėta 2021-01-27]. ISSN 0034-3404. Prieiga per: doi:10.1080/0034340042000292638
  38. ČIBINSKIENĖ, A. Miesto turizmo konkurencingumo vidinės aplinkos veiksniai [interaktyvus]. 2013, **2**(30), 57–66 [žiūrėta 2021-02-20]. Prieiga per: <https://www.lituanistika.lt/content/52289>
  39. AIGINGER, K. ir J. VOGEL. Competitiveness: from a misleading concept to a strategy supporting Beyond GDP goals. *Competitiveness Review* [interaktyvus]. 2015, **25**(5), 497–523 [žiūrėta 2021-02-18]. ISSN 1059-5422. Prieiga per: doi:10.1108/CR-06-2015-0052
  40. TERZIĆ, L. The Role of Innovation in Fostering Competitiveness and Economic Growth: Evidence from Developing Economies. *Comparative Economic Research. Central and Eastern Europe* [interaktyvus]. 2017, **20**(4), 65–81 [žiūrėta 2021-02-15]. ISSN 2082-6737. Prieiga per: doi:10.1515/cer-2017-0028
  41. KRUK, H. Similarities between sustainable development and competitiveness of the regions - theoretical approach. *Modern Management Review* [interaktyvus]. 2017, **XXII**(2), 71–86 [žiūrėta 2021-01-31]. ISSN 23006366. Prieiga per: doi:10.7862/rz.2017.mmr.17
  42. DENKOWSKA, S., K. FIJOREK ir G. WEGRZYN. Formal and Non-Formal Education and Training As an Instrument Fostering Innovation and Competitiveness in EU Member Countries. *Journal of Competitiveness* [interaktyvus]. 2020, **12**(3), 82–98 [žiūrėta 2021-02-15]. ISSN 1804171X. Prieiga per: doi:10.7441/joc.2020.03.05
  43. BANELIENĖ, R. Inovacinė veikla : šiuolaikiniai požūriai ir rodikliai. *Public Administration* [interaktyvus]. 2019, **1**(56), 56–70 [žiūrėta 2021-02-20]. Prieiga per: EBSCO
  44. VALODKIENĖ, G. ir V. SNIEŠKA. Tarptautinis konkurencingumas ir jį lemiantys veiksniai ekonomikos nuosmukio laikotarpiu. *ECONOMICS AND MANAGEMENT* [interaktyvus]. 2012, **17**(2), 602–608 [žiūrėta 2021-02-19]. ISSN 1822-6515. Prieiga per: doi:10.5755/j01.em.17.2.2188
  45. KISELÁKOVÁ, D., B. ŠOFRANKOVÁ, V. ČABINOVÁ ir E. ONUFEROVÁ. Competitiveness and sustainable growth analysis of the EU countries with the use of Global Indexes' methodology. *Entrepreneurship and Sustainability Issues* [interaktyvus]. 2018, **5**(3), 581–599 [žiūrėta 2021-02-15]. ISSN 2345-0282. Prieiga per: doi:10.9770/jesi.2018.5.3(13)
  46. ANDRONICEANU, A.-M., J. KINNUNEN, I. GEORGESCU ir A. ANDRONICEANU. A Multidimensional Approach to Competitiveness, Innovation and Well-Being in the EU Using Canonical Correlation Analysis. *Journal of Competitiveness* [interaktyvus]. 2020, **12**(4), 5–21 [žiūrėta 2021-02-17]. ISSN 1804171X. Prieiga per: doi:10.7441/joc.2020.04.01



47. LENGYEL, I. ir J. RECHNITZER. THE COMPETITIVENESS OF REGIONS IN THE CENTRAL EUROPEAN TRANSITION COUNTRIES. *The MacrotHEME Review* [interaktyvus]. 2013, 2(4), 106–121 [žiūrėta 2021-02-08]. Prieiga per: [http://macrotheme.com/yahoo\\_site\\_admin/assets/docs/9LengyelMR24.13191544.pdf](http://macrotheme.com/yahoo_site_admin/assets/docs/9LengyelMR24.13191544.pdf)
48. VILPIŠAUSKAS, R. Tarptautinis konkurencingumas ir Lietuvos eksporto politika. *Pinigų studijos* [interaktyvus]. 2004, 1, 54–69 [žiūrėta 2021-03-12]. Prieiga per: [https://www.lb.lt/uploads/documents/docs/publications/vilpisauskas\\_2.pdf](https://www.lb.lt/uploads/documents/docs/publications/vilpisauskas_2.pdf)
49. EUROPEAN CENTRAL BANK. *How does innovation lead to growth?* [interaktyvus]. 2017. Prieiga per: <https://www.ecb.europa.eu/explainers/tell-me-more/html/growth.en.html>
50. MARTIN, R. *A study on the Factors of Regional Competitiveness: Report for the European Commission DG Regio.* [interaktyvus]. 2003. Prieiga per: [https://ec.europa.eu/regional\\_policy/sources/docgener/studies/pdf/3cr/competitiveness.pdf](https://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/3cr/competitiveness.pdf)
51. GODELYTĖ, L. ir R. KORSAKIENĖ. Mažų ir vidutinių įmonių internacionalizacija ir žmogiškasis kapitalas. *Mokslas – Lietuvos ateitis* [interaktyvus]. 2015, 7(2), 157–162 [žiūrėta 2021-02-10]. ISSN 2029-2341. Prieiga per: doi:10.3846/mla.2015.739
52. GIŽIENĖ, V. ir Ž. SIMANAVIČIENĖ. Žmogiškojo kapitalo vertinimo koncepcija. *Business Systems and Economics* [interaktyvus]. 2012, 2(1), 116–131 [žiūrėta 2021-02-20]. Prieiga per: <https://repository.mruni.eu/handle/007/13236?show=full>
53. GAVUROVA, B., V. IVANKOVA, M. RIGELSKY ir I. KMECOVA. How Do Gender Inequalities in Health Relate to the Competitiveness of Developed Countries? An Empirical Study. *Journal of Competitiveness* [interaktyvus]. 2020, 12(3), 99–118 [žiūrėta 2021-02-17]. ISSN 1804171X. Prieiga per: doi:10.7441/joc.2020.03.06
54. KASZTELAN, A. Natural capital as a factor in regional competitiveness. *Miscellanea Geographica* [interaktyvus]. 2015, 19(3), 63–69 [žiūrėta 2021-01-31]. ISSN 2084-6118. Prieiga per: doi:10.1515/mgrsd-2015-0010
55. SIMIONESCU, M., C.B. PĂUNA ir T. DIACONESCU. Renewable Energy and Economic Performance in the Context of the European Green Deal. *Energies* [interaktyvus]. 2020, 13(23), 6440 [žiūrėta 2021-02-18]. ISSN 1996-1073. Prieiga per: doi:10.3390/en13236440
56. EUROPOS KOMISIJA. *Europos žaliąsias kursus* [interaktyvus]. 2019. Prieiga per: [https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0011.02/DOC\\_1&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0011.02/DOC_1&format=PDF)
57. EUROPOS PARLAMENTAS. *Atsinaujinančioji energetika* [interaktyvus]. [žiūrėta 2021-03-13]. Prieiga per: <https://www.europarl.europa.eu/factsheets/lt/sheet/70/atsinaujinancioji-energija>
58. EUROPOS PARLAMENTAS. *Žiedinė ekonomika: kas tai ir kodėl ji svarbi?* [interaktyvus]. 2021. Prieiga per: <https://www.europarl.europa.eu/news/lt/headlines/economy/20151201STO05603/ziedine-ekonomika-kas-tai-ir-kodel-ji-svarbi>
59. FISCHETTI, T. *Data Analysis with R.* Birmingham, UNITED KINGDOM: Packt Publishing, Limited, 2015. ISBN 9781785286445.
60. DE WAAL, T., J. PANNEKOEK, S. SCHOLTUS ir T. De WAAL. *Handbook of Statistical Data Editing and Imputation.* Hoboken, UNITED STATES: John Wiley & Sons, Incorporated, 2011. ISBN 9780470904831.
61. STEKHOVEN, D.J. ir P. BUHLMANN. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* [interaktyvus]. 2012, 28(1), 112–118 [žiūrėta 2021-03-03]. ISSN 1367-4803. Prieiga per: doi:10.1093/bioinformatics/btr597
62. ISPIROVA, G., T. EFTIMOV ir B.K. SELJAK. Evaluating missing value imputation methods for food composition databases. *Food and Chemical Toxicology* [interaktyvus]. 2020, 141(May), 111368 [žiūrėta 2021-04-28]. ISSN 02786915. Prieiga per: doi:10.1016/j.fct.2020.111368
63. AZUR, M.J., E.A. STUART, C. FRANGAKIS ir P.J. LEAF. Multiple imputation by chained

- equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research* [interaktyvus]. 2011, **20**(1), 40–49 [žiūrėta 2021-04-28]. ISSN 10498931. Prieiga per: doi:10.1002/mpr.329
64. SCHAFFER, J.L. *Analysis of Incomplete Multivariate Data*. New York: CRC Press, 1997. ISBN 9781439821862.
  65. DAI, J., M.A. VASARHELYI ir A.F. MEDINETS. *Audit Analytics in the Financial Industry*. Bingley, UNITED KINGDOM: Emerald Publishing Limited, 2019. ISBN 9781787431737.
  66. TUFFÉRY, S. *Data Mining and Statistics for Decision Making*. New York, UNITED KINGDOM: John Wiley & Sons, Incorporated, 2011. ISBN 9780470979167.
  67. SAXENA, A., M. PRASAD, A. GUPTA, N. BHARILL, O.P. PATEL, A. TIWARI, M.J. ER, W. DING ir C.-T. LIN. A review of clustering techniques and developments. *Neurocomputing* [interaktyvus]. 2017, **267**, 664–681 [žiūrėta 2021-03-30]. ISSN 09252312. Prieiga per: doi:10.1016/j.neucom.2017.06.053
  68. HAN, J., J. PEI ir M. KAMBER. *Data Mining : Concepts and Techniques*. 3-asis leid. Saint Louis, UNITED STATES: Elsevier Science & Technology, 2011. ISBN 9780123814807.
  69. XU, D. ir Y. TIAN. A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science* [interaktyvus]. 2015, **2**(2), 165–193 [žiūrėta 2021-03-30]. ISSN 2198-5804. Prieiga per: doi:10.1007/s40745-015-0040-1
  70. GOVENDER, P. ir V. SIVAKUMAR. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research* [interaktyvus]. 2020, **11**(1), 40–56 [žiūrėta 2021-03-30]. ISSN 13091042. Prieiga per: doi:10.1016/j.apr.2019.09.009
  71. OZGUR, C., T. COLLIAU, G. ROGERS ir Z. HUGHES. MatLab vs. Python vs. R. *Journal of Data Science* [interaktyvus]. 2017, **15**(3), 355–372 [žiūrėta 2021-04-28]. ISSN 1680-743X. Prieiga per: doi:10.6339/JDS.201707\_15(3).0001
  72. *What is R?* [interaktyvus]. Prieiga per: <https://www.r-project.org/about.html>
  73. WORRALL, J.L. A User-Friendly Introduction to Panel Data Modeling. *Journal of Criminal Justice Education* [interaktyvus]. 2010, **21**(2), 182–196 [žiūrėta 2021-03-27]. ISSN 1051-1253. Prieiga per: doi:10.1080/10511251003693702
  74. BREIMAN, L. Random forests. *Machine Learning* [interaktyvus]. 2001, **45**, 5–32 [žiūrėta 2021-05-03]. Prieiga per: Springer
  75. LEDOLTER, J. *Data Mining and Business Analytics with R*. Somerset, UNITED STATES: John Wiley & Sons, Incorporated, 2013. ISBN 9781118593745.
  76. ROIGER, R.J. *Data Mining : A Tutorial-Based Primer, Second Edition*. Boca Raton, UNITED STATES: CRC Press LLC, 2016. ISBN 9781498763981.
  77. RAGHAVAN, V. V., V.N. GUDIVADA, V. GOVINDARAJU ir C.R. RAO. *Cognitive Computing: Theory and Applications*. Oxford, NETHERLANDS, THE: Elsevier Science & Technology, 2016. ISBN 9780444637512.
  78. *Package 'randomForest'* [interaktyvus]. Prieiga per: <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
  79. MOISL, H. *Cluster Analysis for Corpus Linguistics*. Berlin/Boston, GERMANY: De Gruyter, Inc., 2015. ISBN 9783110363814.
  80. JEON, Y., J. YOO, J. LEE ir S. YOON. NC-link: A New Linkage Method for Efficient Hierarchical Clustering of Large-Scale Data. *IEEE Access* [interaktyvus]. 2017, **5**, 5594–5608 [žiūrėta 2021-03-30]. ISSN 2169-3536. Prieiga per: doi:10.1109/ACCESS.2017.2690987
  81. CHARRAD, M., N. GHAZZALI, V. BOITEAU ir A. NIKNAFS. NbClust : An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software* [interaktyvus]. 2014, **61**(6), 1–36 [žiūrėta 2021-03-31]. ISSN 1548-7660. Prieiga per: doi:10.18637/jss.v061.i06
  82. JOLLIFFE, I.T. ir J. CADIMA. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical*

- and Engineering Sciences* [interaktyvus]. 2016, **374**(2065), 1–16 [žiūrėta 2021-03-31]. ISSN 1364-503X. Prieiga per: doi:10.1098/rsta.2015.0202
83. LAROSE, D.T. ir C.D. LAROSE. *Data Mining and Predictive Analytics*. New York, UNITED STATES: John Wiley & Sons, Incorporated, 2015. ISBN 9781118868676.
  84. GENUER, R., J.-M. POGGI ir C. TULEAU-MALOT. VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal* [interaktyvus]. 2015, **7**(2), 19–33 [žiūrėta 2021-04-15]. ISSN 2073-4859. Prieiga per: doi:10.32614/RJ-2015-018
  85. NAGHSHPOUR, S. *A Primer on Non-Parametric Analysis*. New York, UNITED STATES: Business Expert Press, 2016. ISBN 9781631575518.
  86. MAROZZI, M. Testing for concordance between several criteria. *Journal of Statistical Computation and Simulation* [interaktyvus]. 2014, **84**(9), 1843–1850 [žiūrėta 2021-04-15]. ISSN 0094-9655. Prieiga per: doi:10.1080/00949655.2013.766189
  87. VASKE, J.J., J. BEAMAN ir C.C. SPONARSKI. Rethinking Internal Consistency in Cronbach's Alpha. *Leisure Sciences* [interaktyvus]. 2017, **39**(2), 163–173 [žiūrėta 2021-05-25]. ISSN 0149-0400. Prieiga per: doi:10.1080/01490400.2015.1127189
  88. VERMA, J.P. ir A.-S.G. ABDEL-SALAM. *Testing Statistical Assumptions in Research*. Newark, UNITED STATES: John Wiley & Sons, Incorporated, 2019. ISBN 9781119528401.

### Informacijos šaltinių sąrašas

1. R, prieiga per: <https://www.r-project.org/about.html>
2. Python, prieiga per: <https://www.python.org/about/>
3. Eurostat rodiklių duomenų bazė, prieiga per: <https://ec.europa.eu/eurostat/data/database>
4. OECD rodiklių duomenų bazė, prieiga per:  
<https://data.oecd.org/searchresults/?r=+f/type/datasets>
5. The Global Economy rodiklių duomenų bazė, prieiga per:  
[https://www.theglobaleconomy.com/indicators\\_list.php](https://www.theglobaleconomy.com/indicators_list.php)
6. The World Bank rodiklių duomenų bazė, prieiga per: <https://data.worldbank.org/>

## Priedai

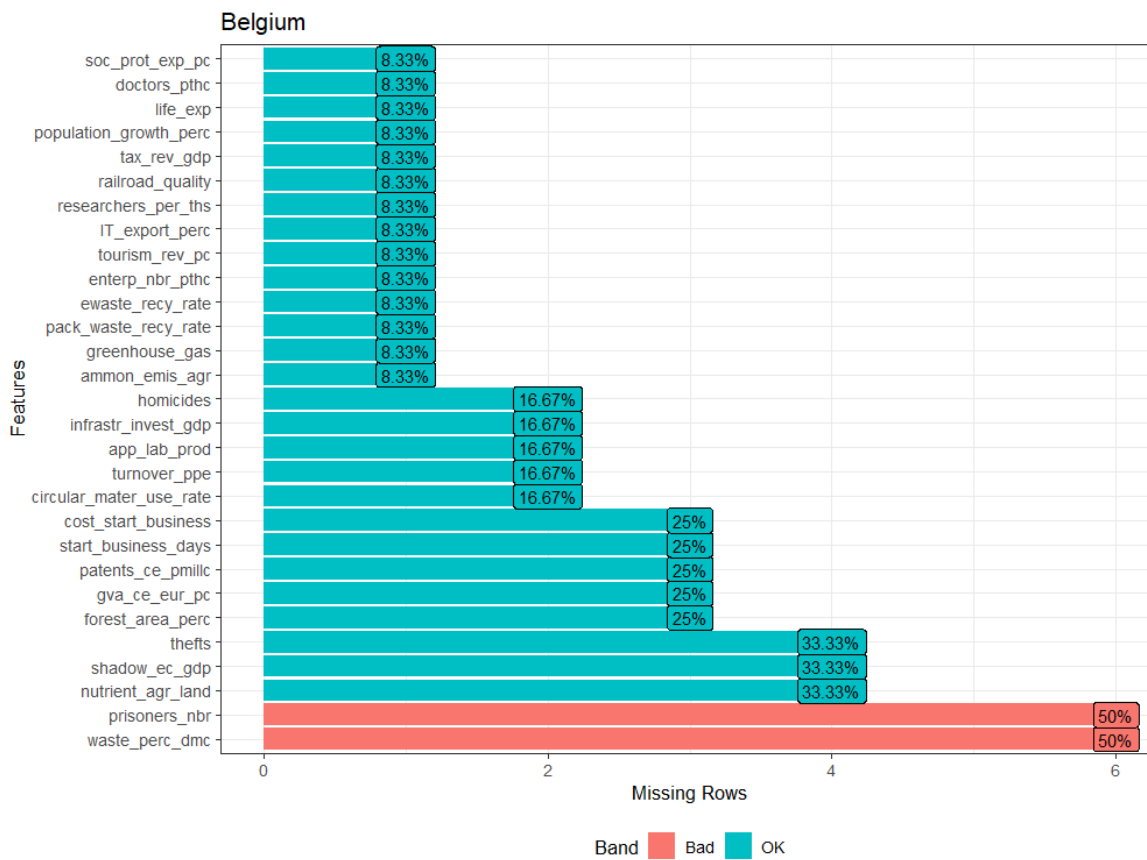
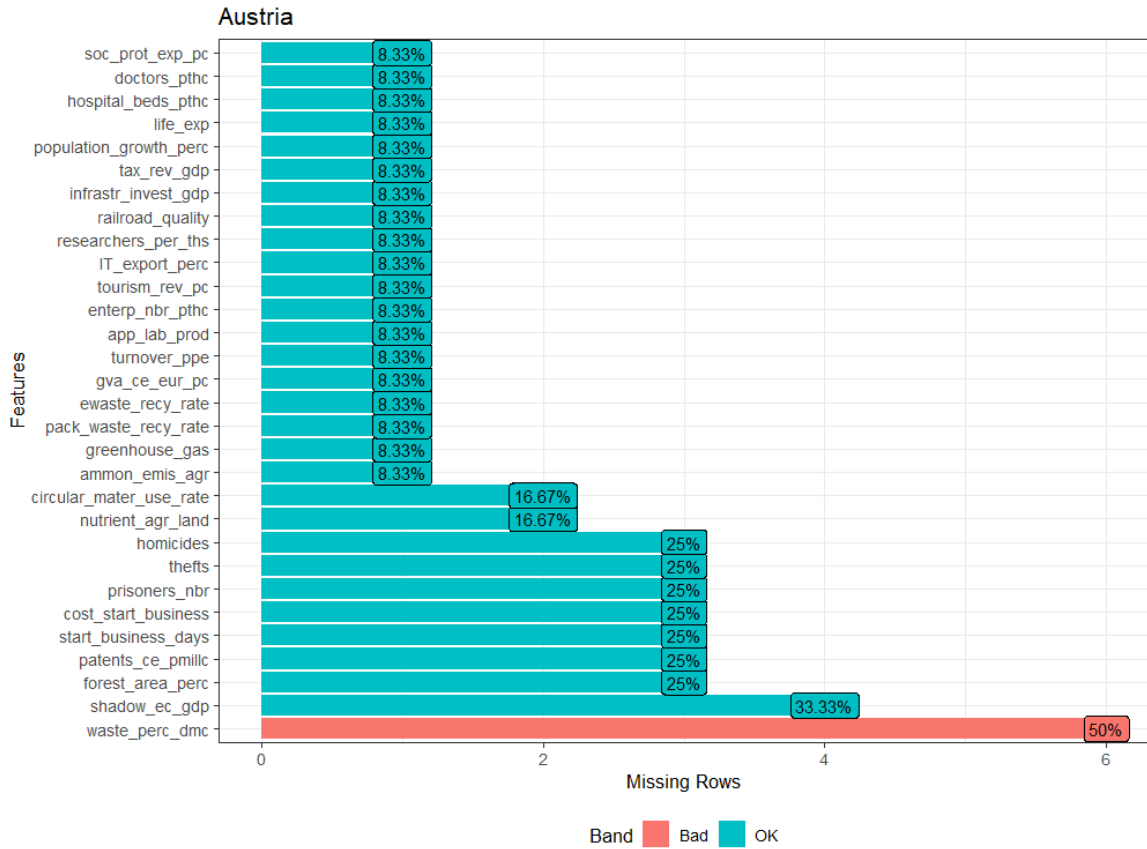
### 1 priedas. Konkurencingumą lemiančios veiksmų grupės ir rodikliai

EKONOMINIS KONKURENCINGUMAS	Veiksmų grupė	Rodiklis ir jo matavimo vienetai	Rodiklio šaltinis	Rodiklio trumpinys, naudotas R'e
	Investicijos	TUI, proc. nuo BVP		The Global Economy
Kapitalo investicijos, proc. nuo BVP			The Global Economy	cap_gdp
Ekonomikos atvirumas	Prekių ir paslaugų eksportas, proc. nuo BVP		The Global Economy	export_gdp
	Prekių ir paslaugų eksportas, metinis augimas proc.		The Global Economy	export_growth
	Prekių ir paslaugų importas, proc. nuo BVP		The Global Economy	import_gdp
Ekonominiai rezultatai	Apyvarta vienam dirbančiam, tūkst. EUR		Eurostat	turnover_ppe
	Bendroji pridėtinė vertė vienam dirbančiam, tūkst. EUR		Eurostat	app_lab_prod
	Įmonių skaičius, tenkantis 1000 gyventojų. Rodiklis išskaičiuotas remiantis OECD ir Eurostat duomenimis		OECD, Eurostat	enterp_nbr_pthc
	Savarankiškai dirbantys asmenys, proc. nuo visų 20-29 m. dirbančių		OECD	young_self_empl
	Užimtumo lygis 20-64 m., proc. nuo visų gyventojų		Eurostat	empl_rate
	Metinis grynasis uždarbis (atėmus mokesčius), eur		Eurostat	annual_net_earnings
	Darbo našumas vienam asmeniui, metinis proc. pokytis		Eurostat	lab_prod_perc
	Tarptautinio turizmo pajamos vienam gyventojui, USD. Rodiklis išskaičiuotas remiantis The Global Economy ir Eurostat duomenimis		The Global Economy, Eurostat	tourism_rev_pc
Inovacijos ir technologijos	Mokslinių tyrimų ir plėtros išlaidos, tenkančios vienam gyventojui EUR		Eurostat	exp_onRD_pc
	Patentų paraiškos Europos patentų tarnybai, tenkančios 1 mln. gyventojų		Eurostat	patent_EPO_pmilc
	IT eksportas, proc. nuo viso eksporto		The Global Economy	IT_export_perc
	Tyrėjai, tenkantys 1000 dirbančiųjų		OECD	researchers_per_ths
	Dirbantys 25-64 m. moksle ir technologijose, proc. nuo aktyvios populiacijos		Eurostat	pers_scienc_techn
Infrastruktūra	Kelių kokybės indeksas, 1 (žema) - 7 (aukšta)		The Global Economy	roads_quality
	Geležinkelio infrastruktūros kokybės indeksas, 1 (žema) - 7 (aukšta)		The Global Economy	railroad_quality
	Oro transporto infrastruktūros kokybės indeksas, 1 (žema) - 7 (aukšta)		The Global Economy	airtransp_quality
	Investicijos į infrastruktūrą (transporto), proc. nuo BVP		OECD	infrastr_invest_gdp

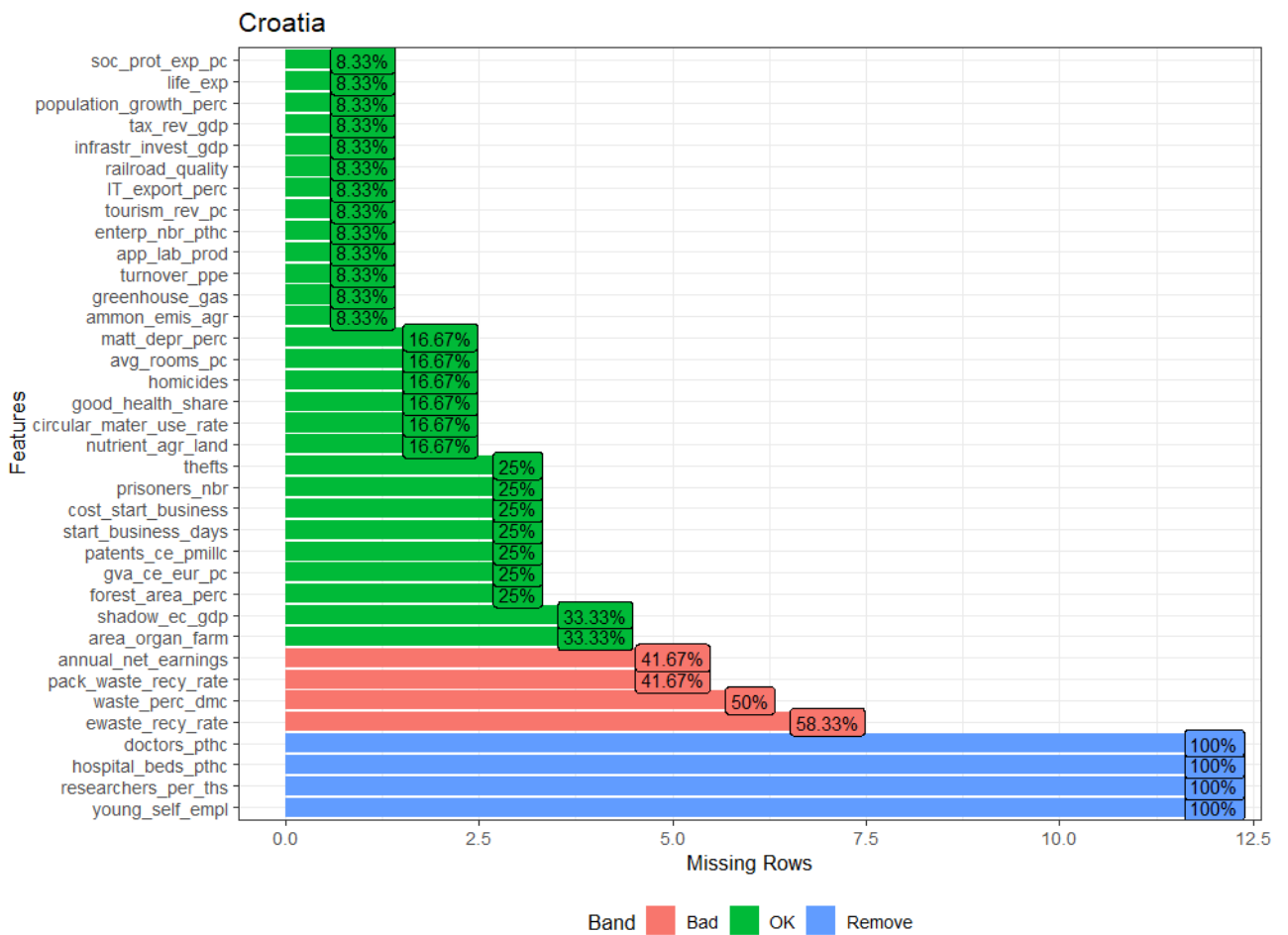
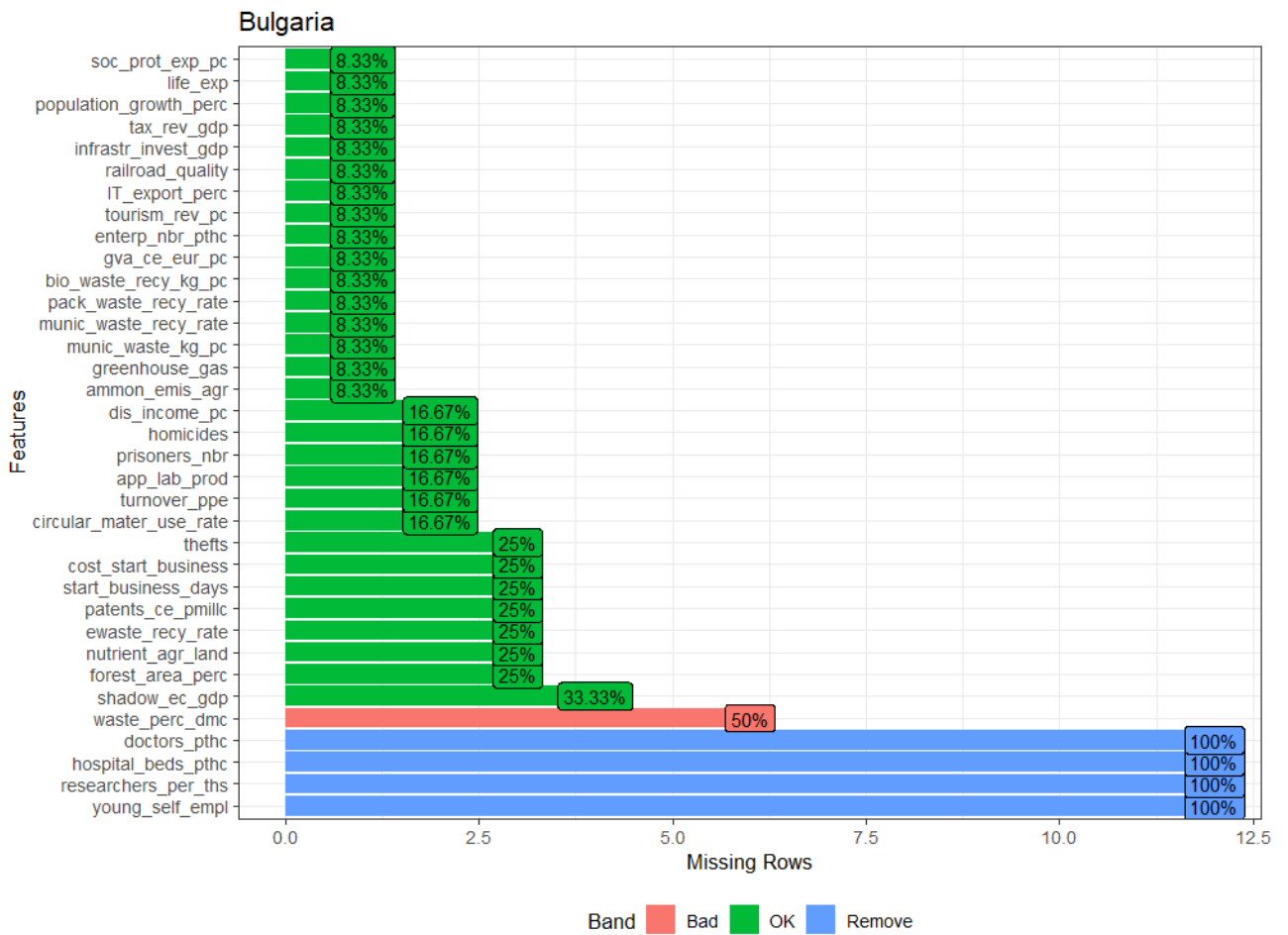
	<b>Veiksnių grupė</b>	<b>Rodiklis ir jo matavimo vienetai</b>	<b>Rodiklio šaltinis</b>	<b>Rodiklio trumpinys, naudotas R'e</b>
<b>SOCIALINIS KONKURENCINGUMAS</b>	Žmogiškieji ištekliai	Gyventojų skaičius (rodiklis naudotas tik tam tikrų rodiklių paskaičiavimui, į klasterizavimą netraukta)	Eurostat	population_nbr
		Populiacijos metinis augimo tempas, proc.	The Global Economy	population_growth_perc
		Aktyvi populiacija 15-64 m., proc. nuo visos populiacijos	Eurostat	active_pop_perc
		Aukštasis išsilavinimas tarp 20-29 m., proc. nuo visos populiacijos	Eurostat	tertiary_educ_perc
		Besimokantys suaugusieji 25-64 m., proc. nuo populiacijos	Eurostat	adult_learn_perc
		Valstybės išlaidos švietimui, proc. nuo BVP	Eurostat	exp_educ_gdp
	Gyvenimo kokybė	Vidutinė tikėtina gyvenimo trukmė, metais	The Global Economy	life_exp
		Žmonių, kurių sveikatos būklė gerai ar labai gera, dalis	Eurostat	good_health
		Sveikatos išlaidos gyventojui, USD	OECD	health_spend_pc
		Ligoninės lovų skaičius, tenkantis 1000 žmonių	The Global Economy	hospital_beds_pthc
		Gydytojų skaičius, tenkantis 1000 žmonių	The Global Economy	doctors_pthc
		Socialinės apsaugos išlaidos vienam gyventojui, EUR	Eurostat	soc_prot_exp_pc
		Kalinių skaičius 100 000 gyventojų	The Global Economy	prisoners_nbr
		Vagystės 100 000 gyventojų	The Global Economy	thefts
		Žmogžudystės 100 000 gyventojų	The Global Economy	homicides
		Valstybės išlaidos viešajai tvarkai ir saugumui, proc. nuo BVP	Eurostat	exp_publ_safety_gdp
		Vidutinis kambarių skaičius gyventojui	Eurostat	avg_rooms_pc
		Valstybės išlaidos būstui ir bendruomenės patogumams, proc. nuo BVP	Eurostat	exp_housing_gdp
		Namų ūkių disponuojamos pajamos, tenkančios vienam gyventojui	Eurostat	dis_income_pc
Labai nepasituričių dalis, proc. nuo visos populiacijos	Eurostat	matt_depr_perc		
Valstybės išlaidos kultūrai, proc. nuo visų valstybės išlaidų	Eurostat	gov_exp_cult_perc		
<b>ŠALIES APLINKOS KONKURENCINGUMAS</b>	Verslo aplinka	Laikas, reikalingas verslui pradėti, dienomis	The World Bank	start_business_days
		Verslo pradžios procedūrų išlaidos, proc. nuo bendrųjų nacionalinių pajamų vienam gyventojui	The World Bank	cost_start_business
	Makroekonominis stabilumas	Valstybės išlaidos, proc. nuo BVP	The Global Economy	gov_spend_gdp
		Mokesčių pajamos, proc. nuo BVP	The Global Economy	tax_rev_gdp
		Infliacija: vartotojų kainų indekso pokytis proc.	The Global Economy	infl_cpi
	Institucinė ir politinė aplinka	Šešėlinė ekonomika, proc. nuo BVP	The Global Economy	shadow_ec_gdp
		Korupcijos kontrolės indeksas, -2,5 silpnas; 2,5 stiprus	The Global Economy	corrup_control_index
		Reguliavimo kokybės indeksas, -2,5 silpnas; 2,5 stiprus	The Global Economy	reg_quality_index
		Politinio stabilumo indeksas, -2,5 silpnas; 2,5 stiprus	The Global Economy	pol_stability_index
	Vyriausybės efektyvumo indeksas, -2,5 silpnas; 2,5 stiprus	The Global Economy	gov_eff_index	

APLINKOSAUGINIS KONKURENCINGUMAS	Veiksnių grupė	Rodiklis ir jo matavimo vienetai	Rodiklio šaltinis	Rodiklio trumpinys, naudotas R'e
	Aplinka ir užterštumas	Miško plotas, proc. nuo visos teritorijos ploto	The Global Economy	forest_area_perc
		Ekologinio ūkininkavimo plotas, proc. nuo viso naudojamo žemės ūkio ploto	Eurostat	area_organ_farm
		Bendras maistinių medžiagų balansas žemės ūkio paskirties žemėje, kg/ha	Eurostat	nutrient_agr_land
		Žemės ūkyje išmetamas amoniakas, kg/ha	Eurostat	ammon_emis_agr
		Šiltnamio efektą sukeliančių dujų emisija, tenkanti vienam gyventojui	Eurostat	greenhouse_gas
		Valstybės išlaidos aplinkos apsaugai, proc. nuo BVP	Eurostat	exp_envir_prot_gdp
		Aplinkosauginiai mokesčiai, proc. nuo visų mokesčių pajamų	Eurostat	envir_tax_rev_perc
	Atsinaujinančioji energetika	Atsinaujinančios energijos dalis bendrame galutiniame energijos suvartojime	Eurostat	share_renew_energy
	Žiedinė ekonomika	Komunalinės atliekos, tenkančios vienam gyventojui, kg	Eurostat	munic_waste_kg_pc
Atliekų susidarymas, išskyrus pagrindines mineralines atliekas, proc. nuo visų vidaus medžiagų suvartojimo (angl. domestic material consumption (DMC) )		Eurostat	waste_perc_dmc	
Komunalinių atliekų perdirbimo lygis, proc. nuo visų komunalinių atliekų		Eurostat	munic_waste_recy_rate	
Perdirbtų pakuočių atliekų dalis visose susidariusiose pakuočių atliekose, proc.		Eurostat	pack_waste_recy_rate	
Elektroninių atliekų perdirbimo lygis, proc.		Eurostat	ewaste_recy_rate	
Biologinių atliekų perdirbimas, tenkantis gyventojui, kg		Eurostat	bio_waste_recy_kg_pc	
Žiedinis medžiagų panaudojimo lygis		Eurostat	circular_mater_use_rate	
Prekyba perdirbamomis žaliavomis, tonos gyventojui		Eurostat	trade_recycl_materials_t_pc	
Pridėtinė vertė žiedinės ekonomikos sektoriuose, tenkanti gyventojui, eur		Eurostat	gva_ce_eur_pc	
Su perdirbimu ir antrinėmis žaliavomis susiję patentai, tenkantys 1 mln. gyventojų		Eurostat	patents_ce_pmillc	

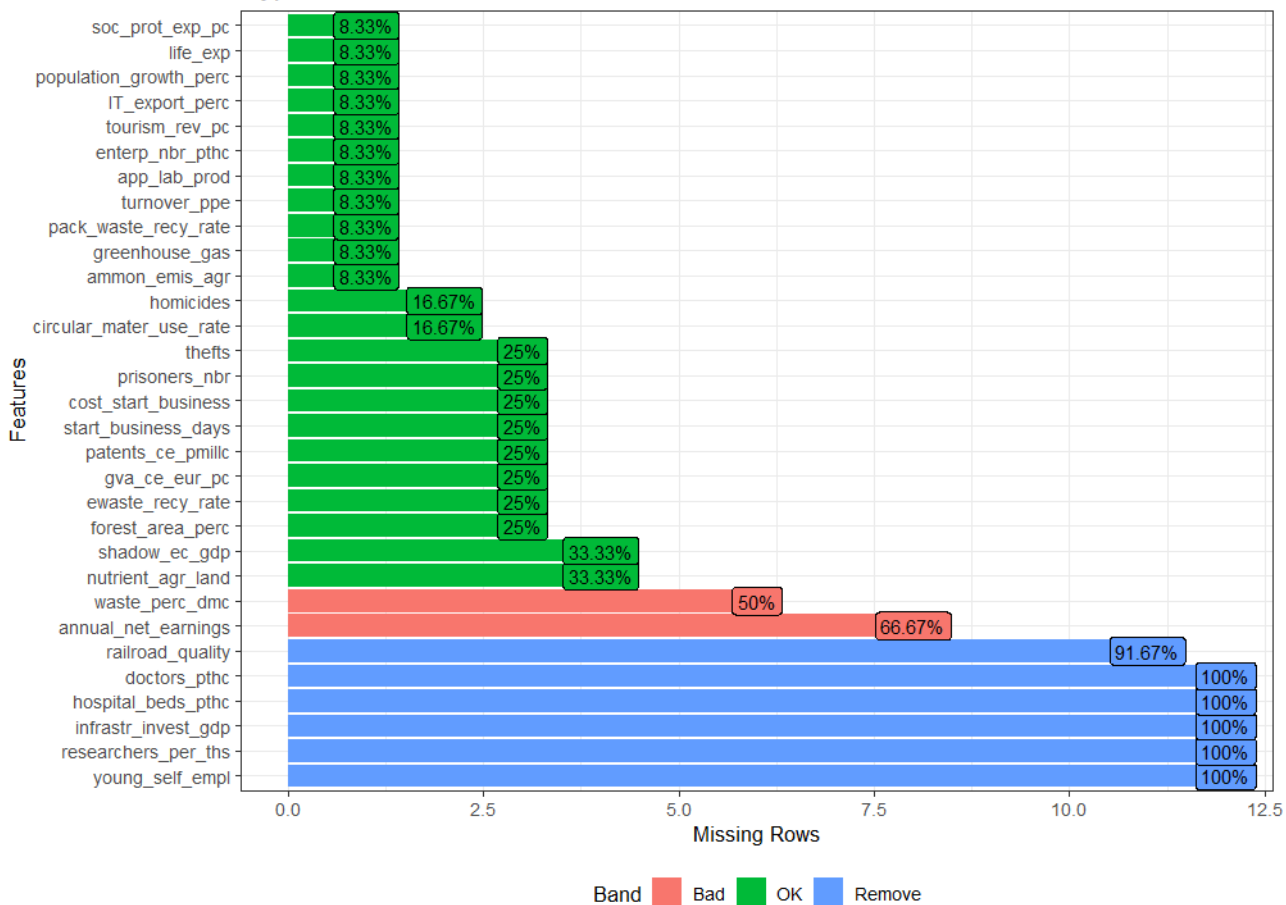
## 2 priedas. Trūkstamos reikšmės pagal požymius ir šalis



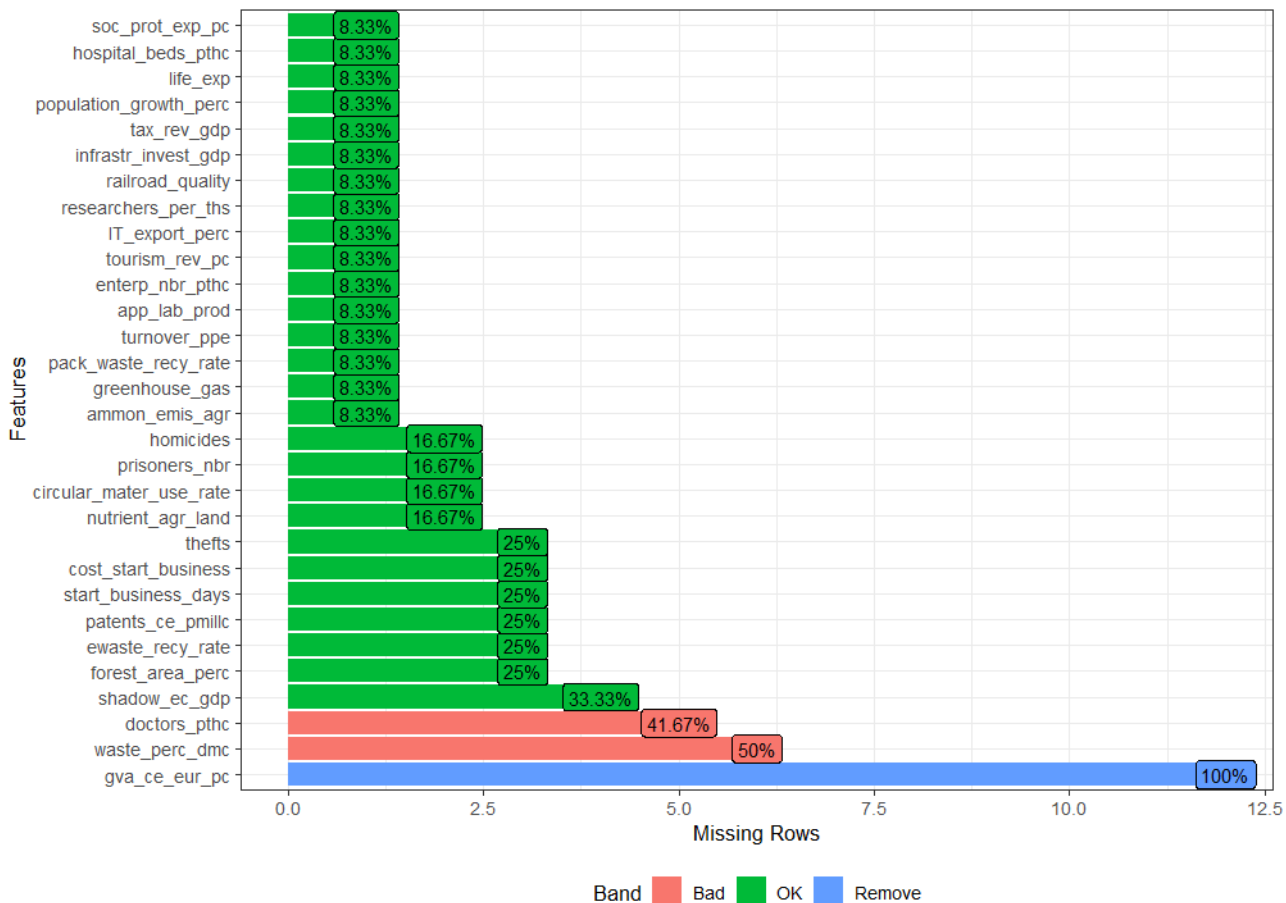




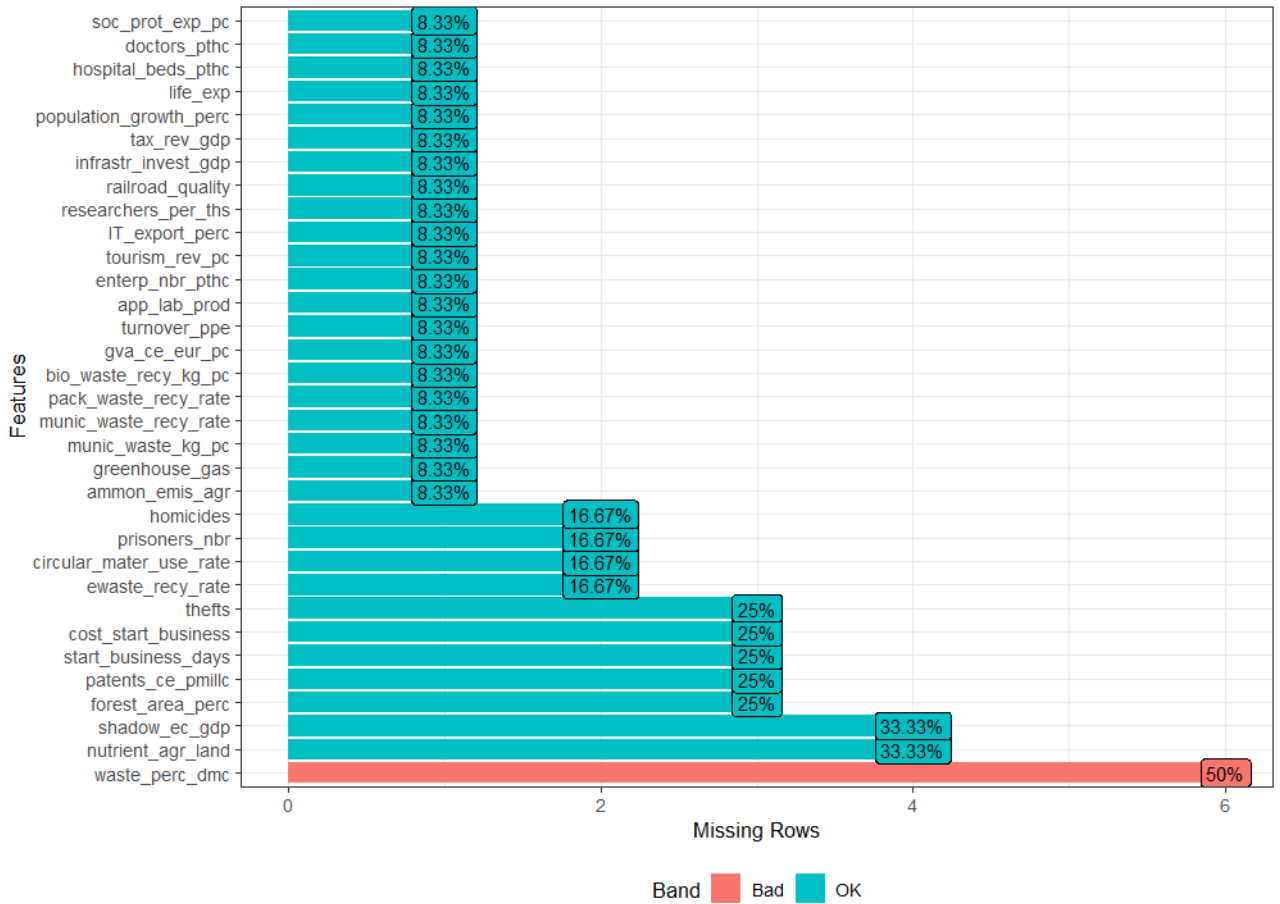
### Cyprus



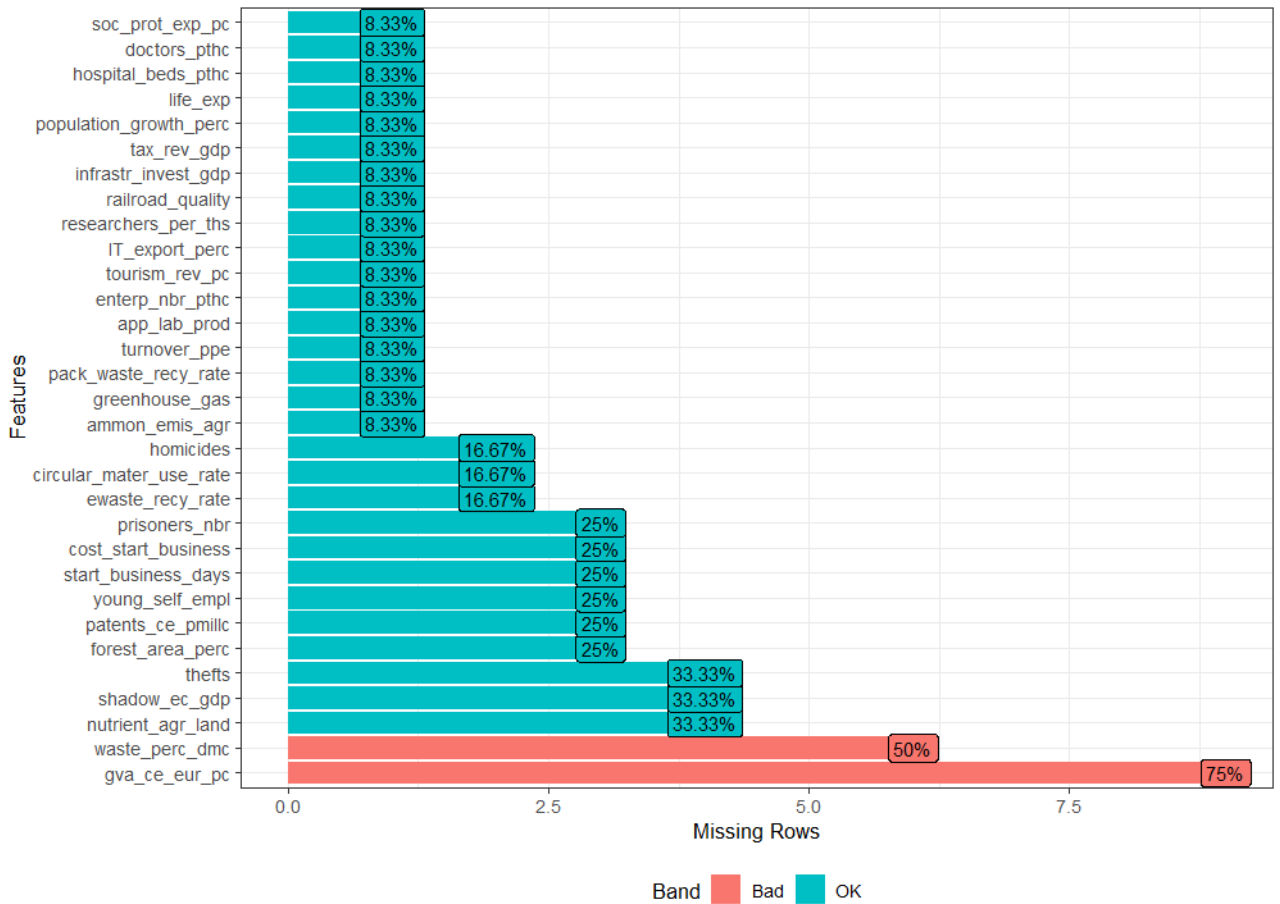
### Czechia



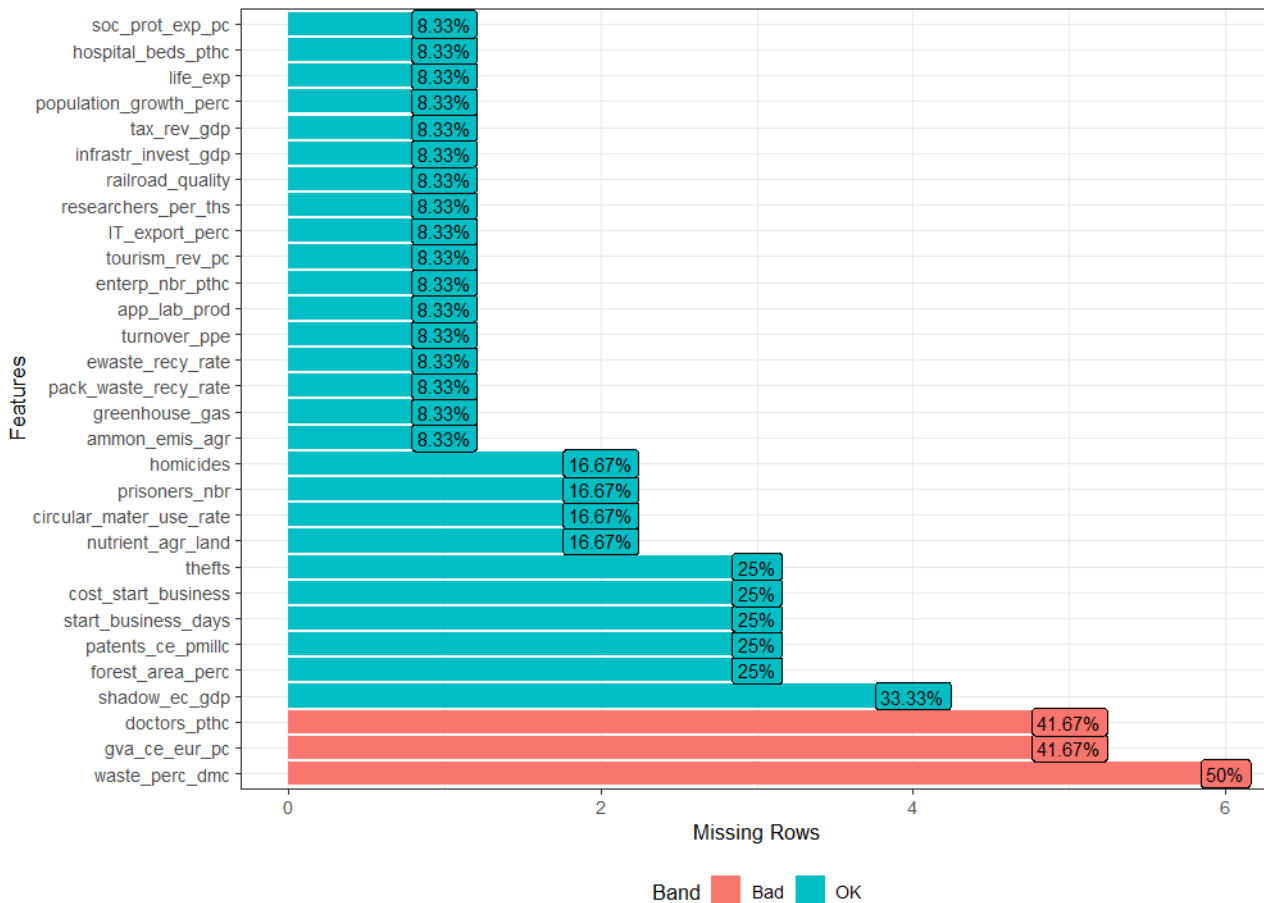
### Denmark



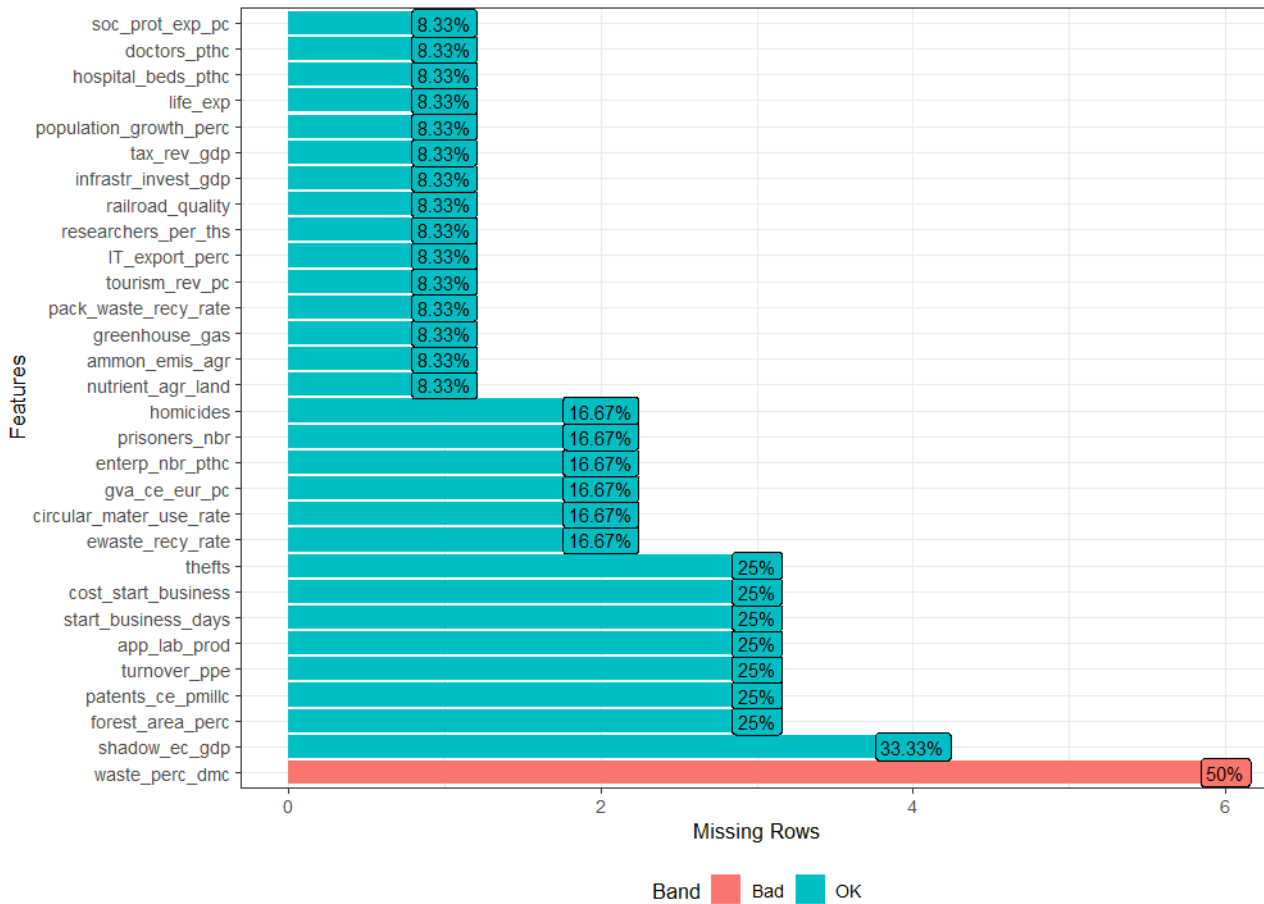
### Estonia



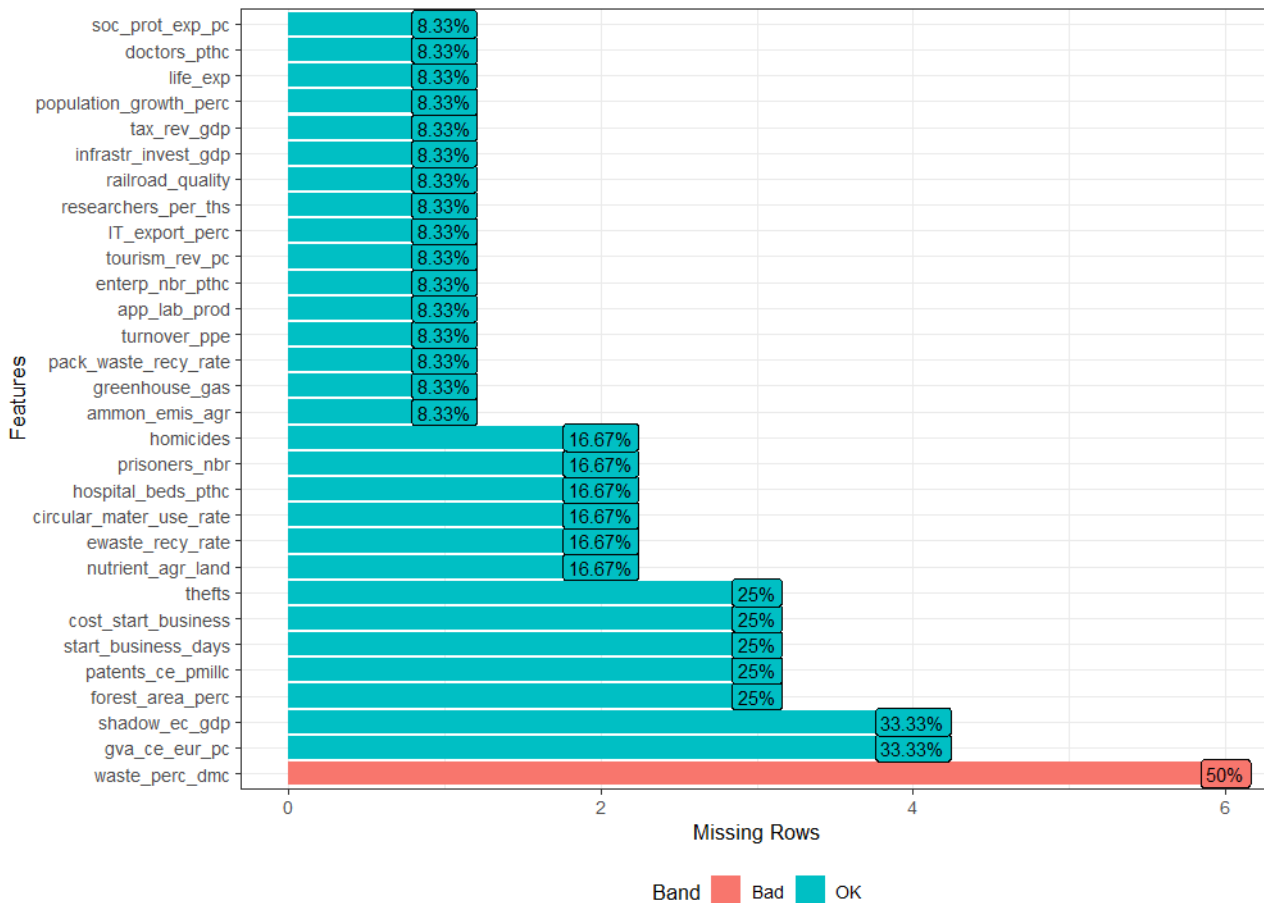
### Finland



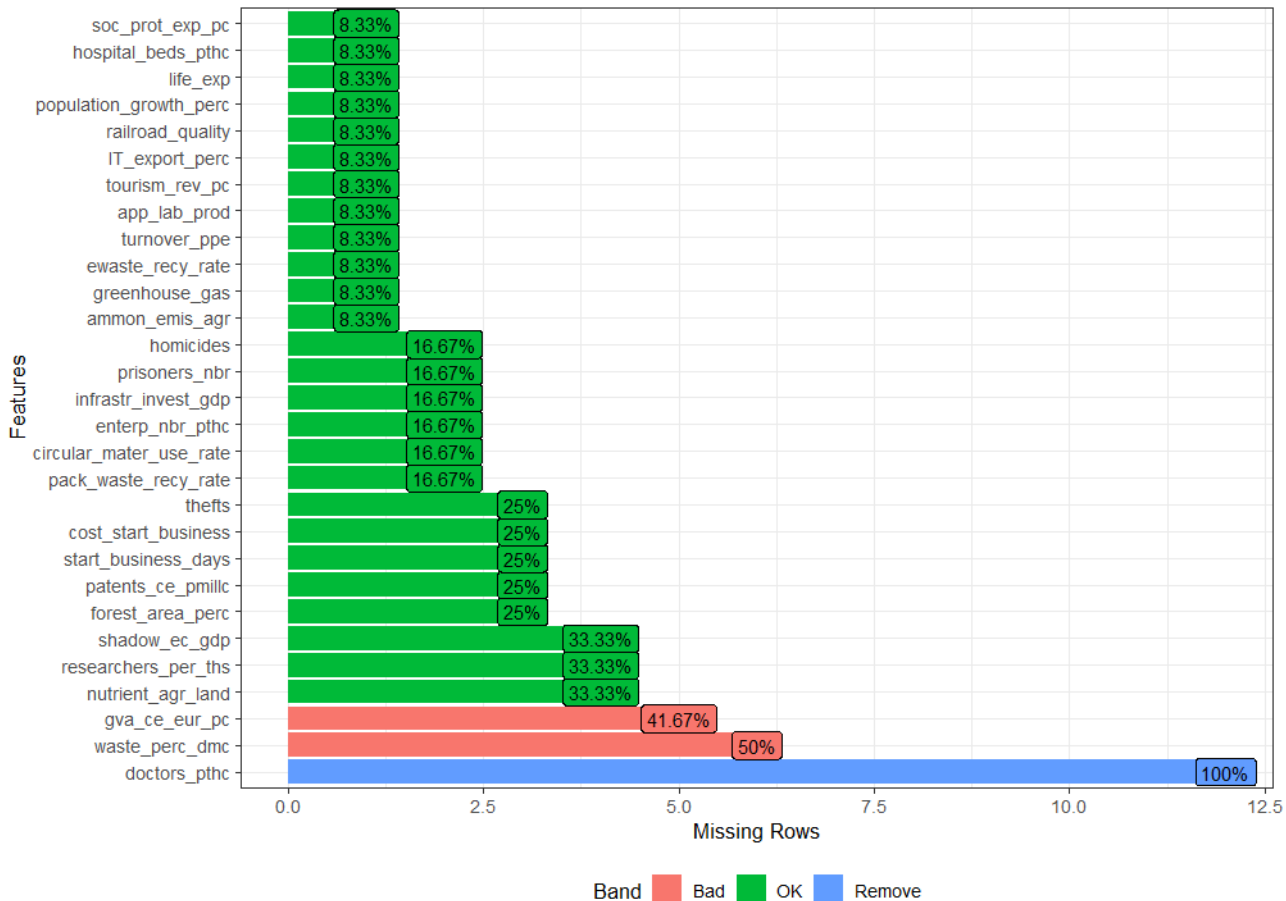
### France

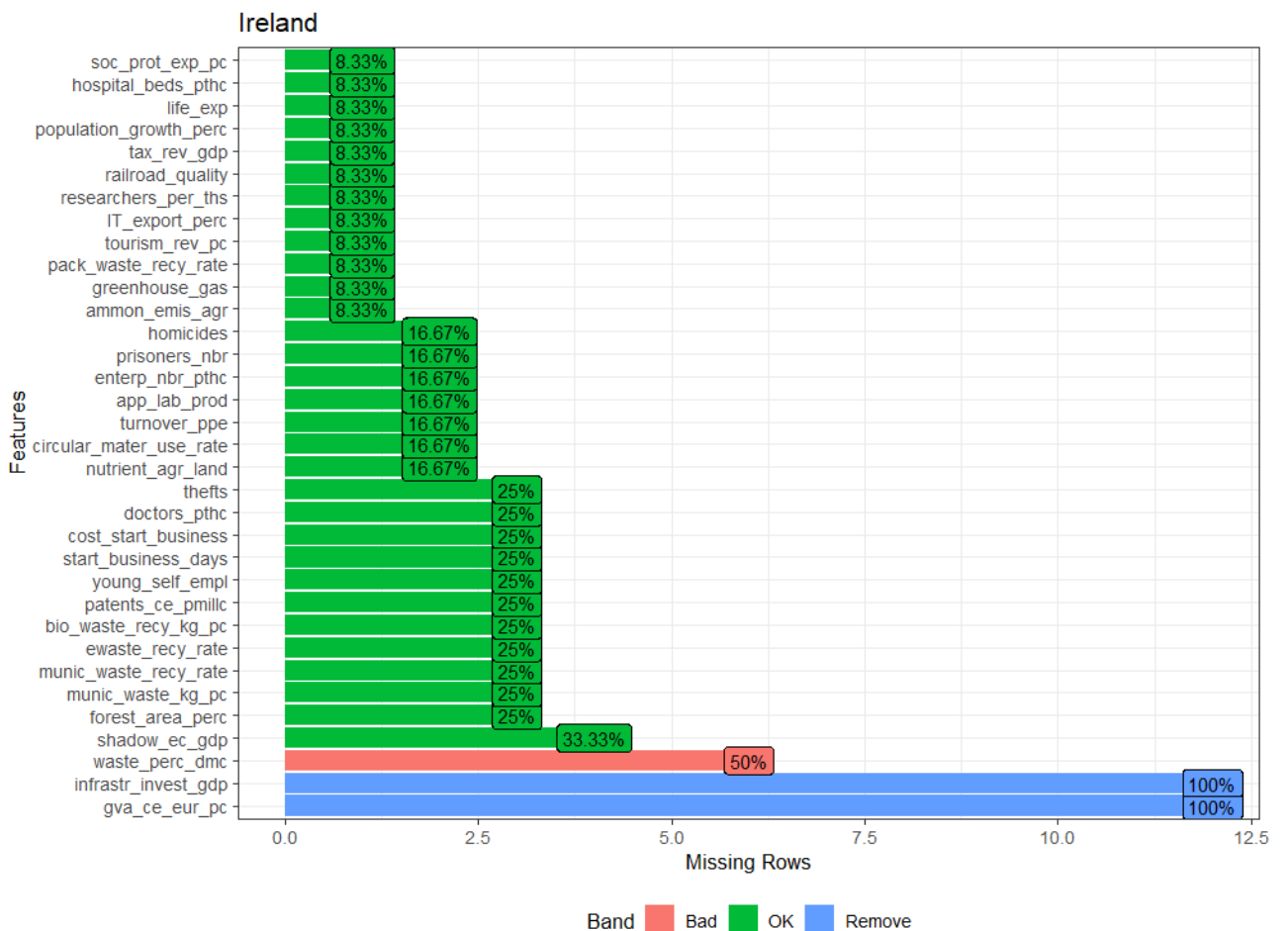
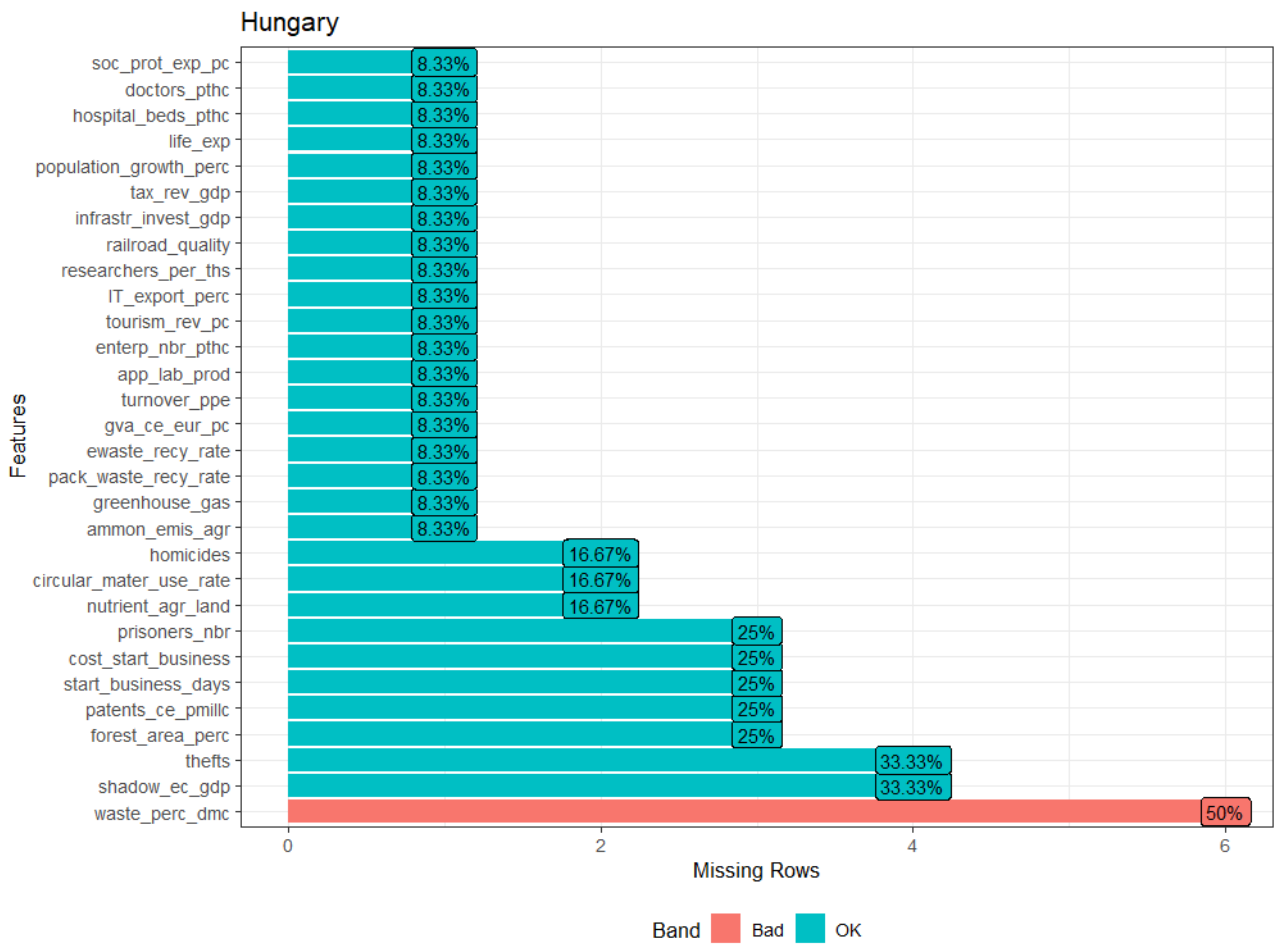


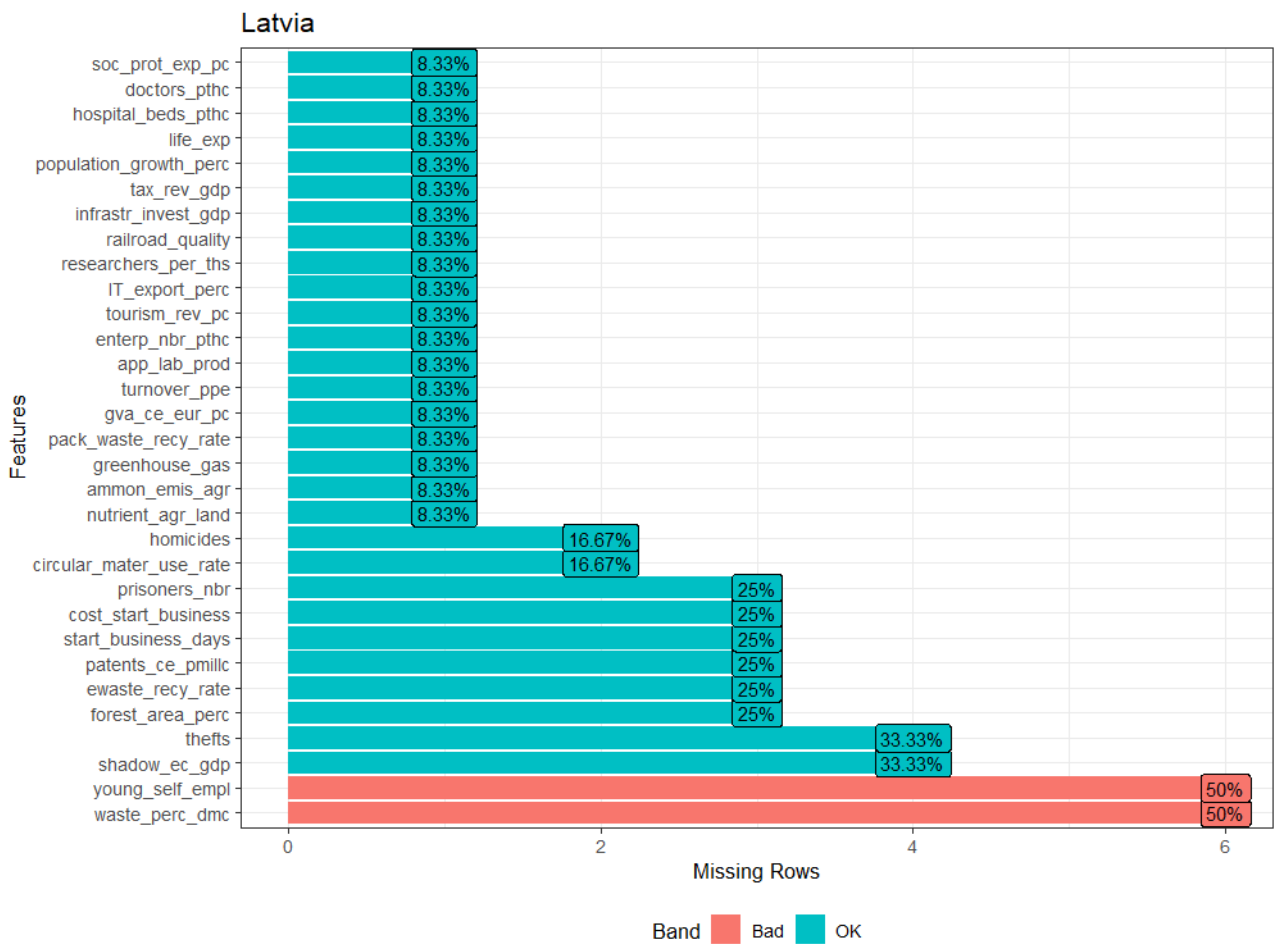
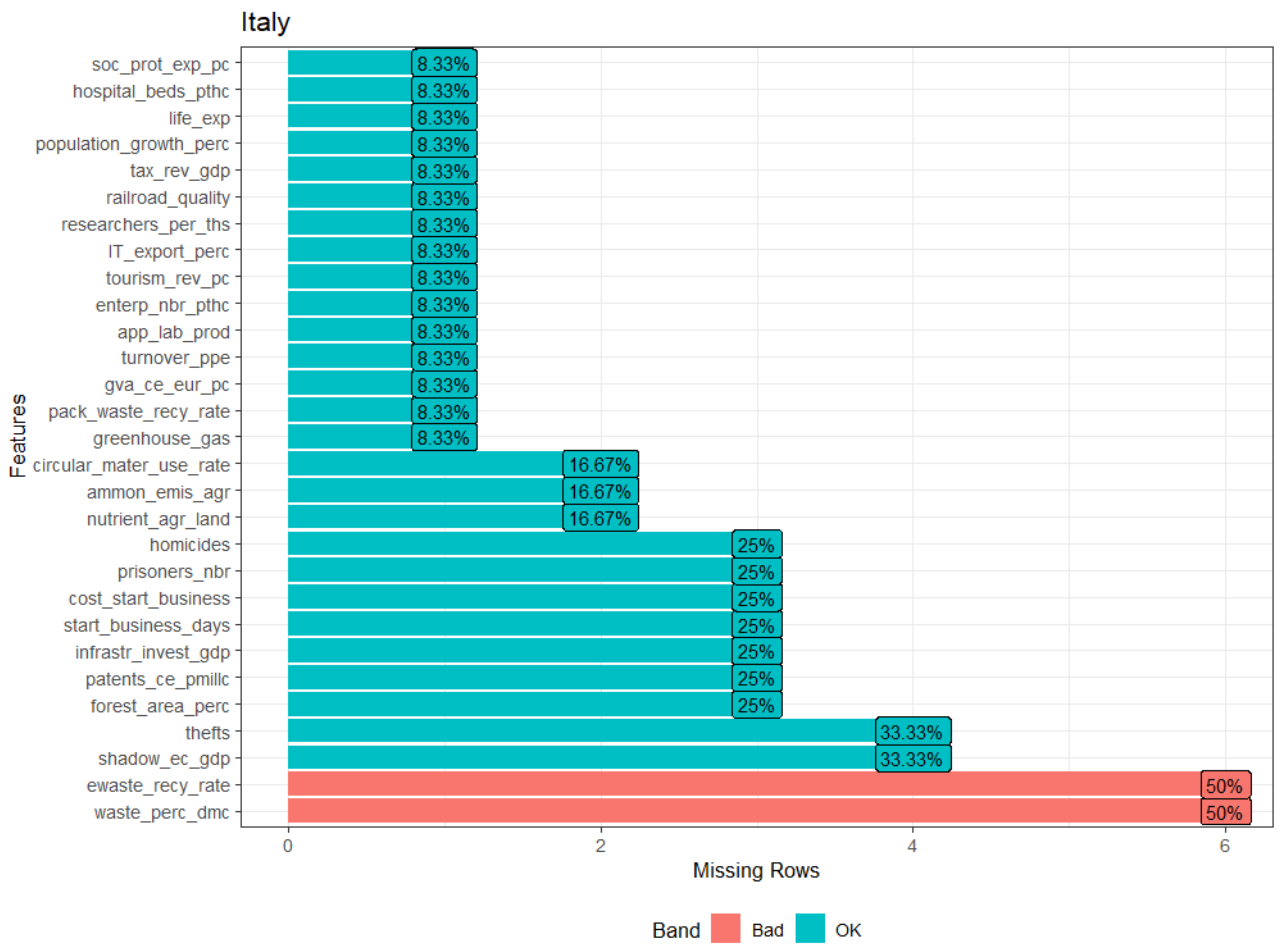
### Germany

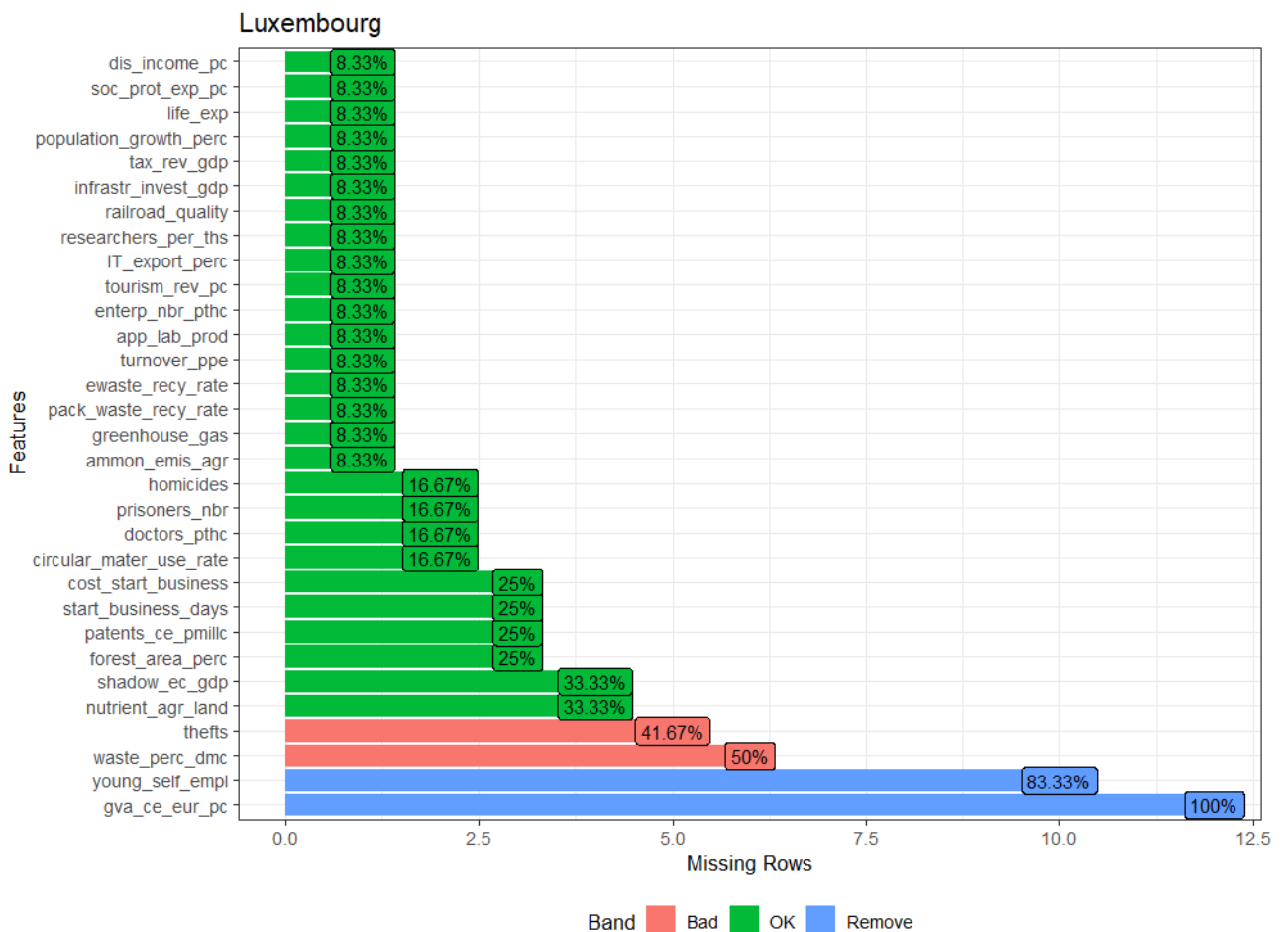
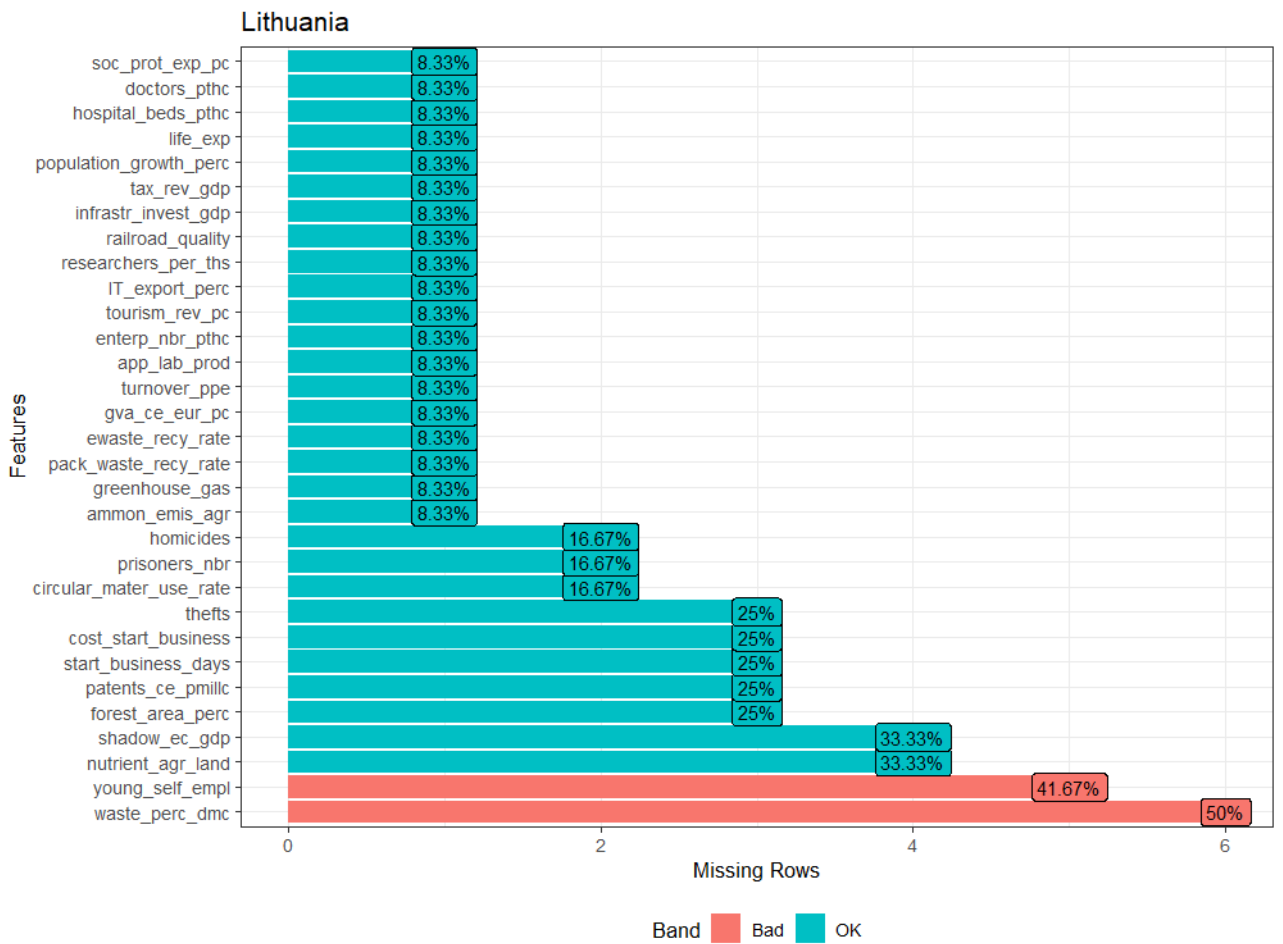


### Greece

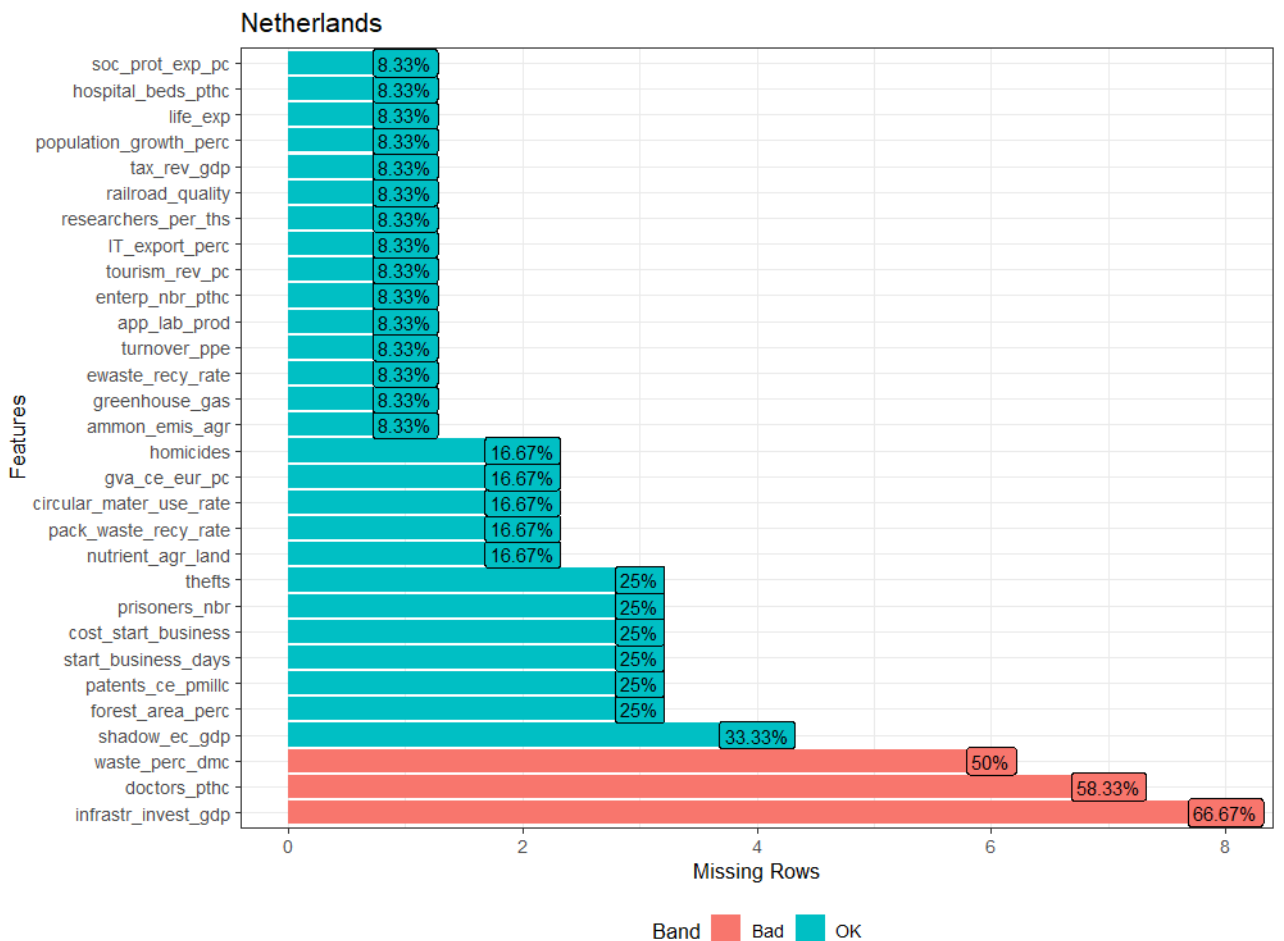
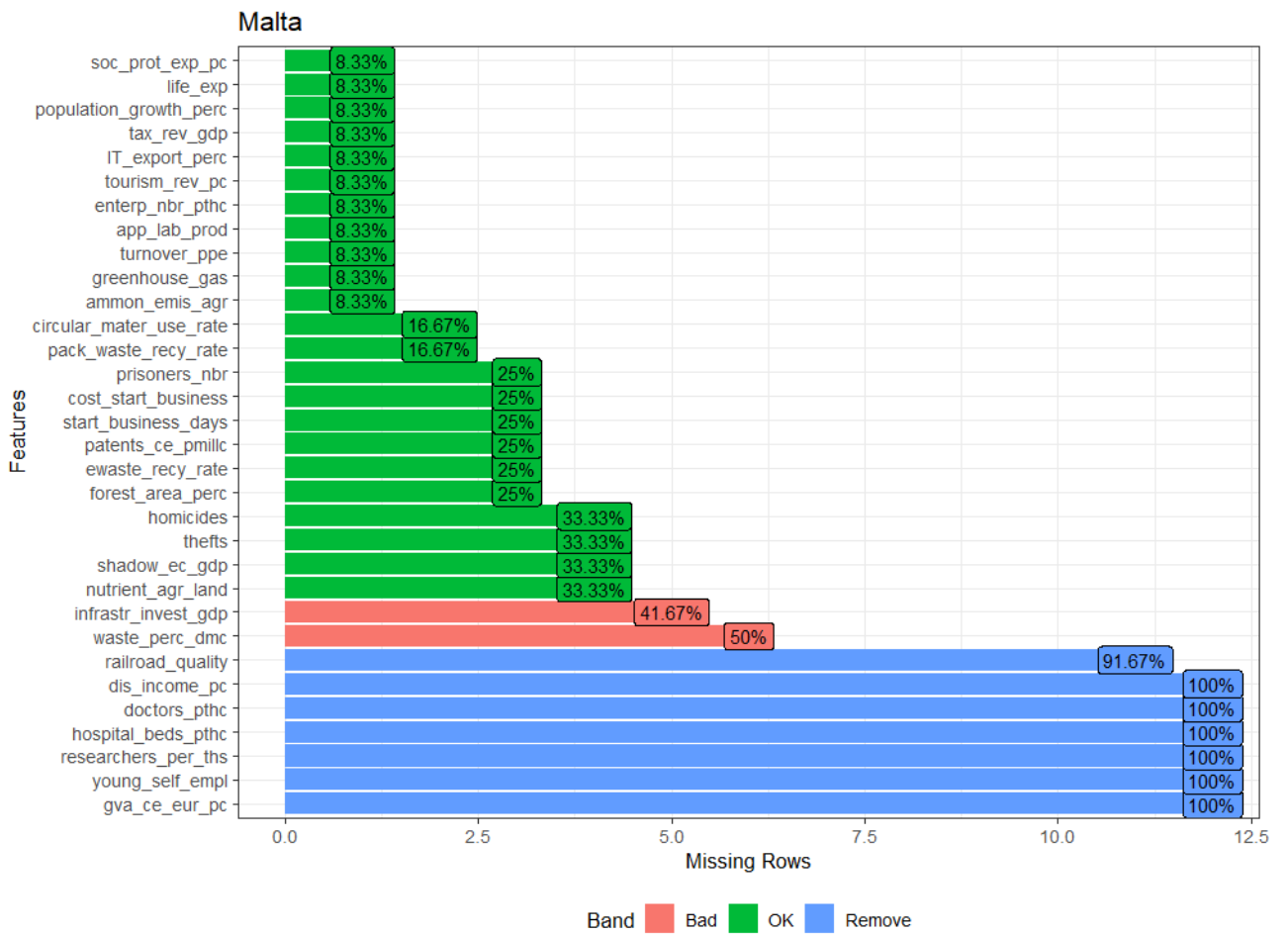


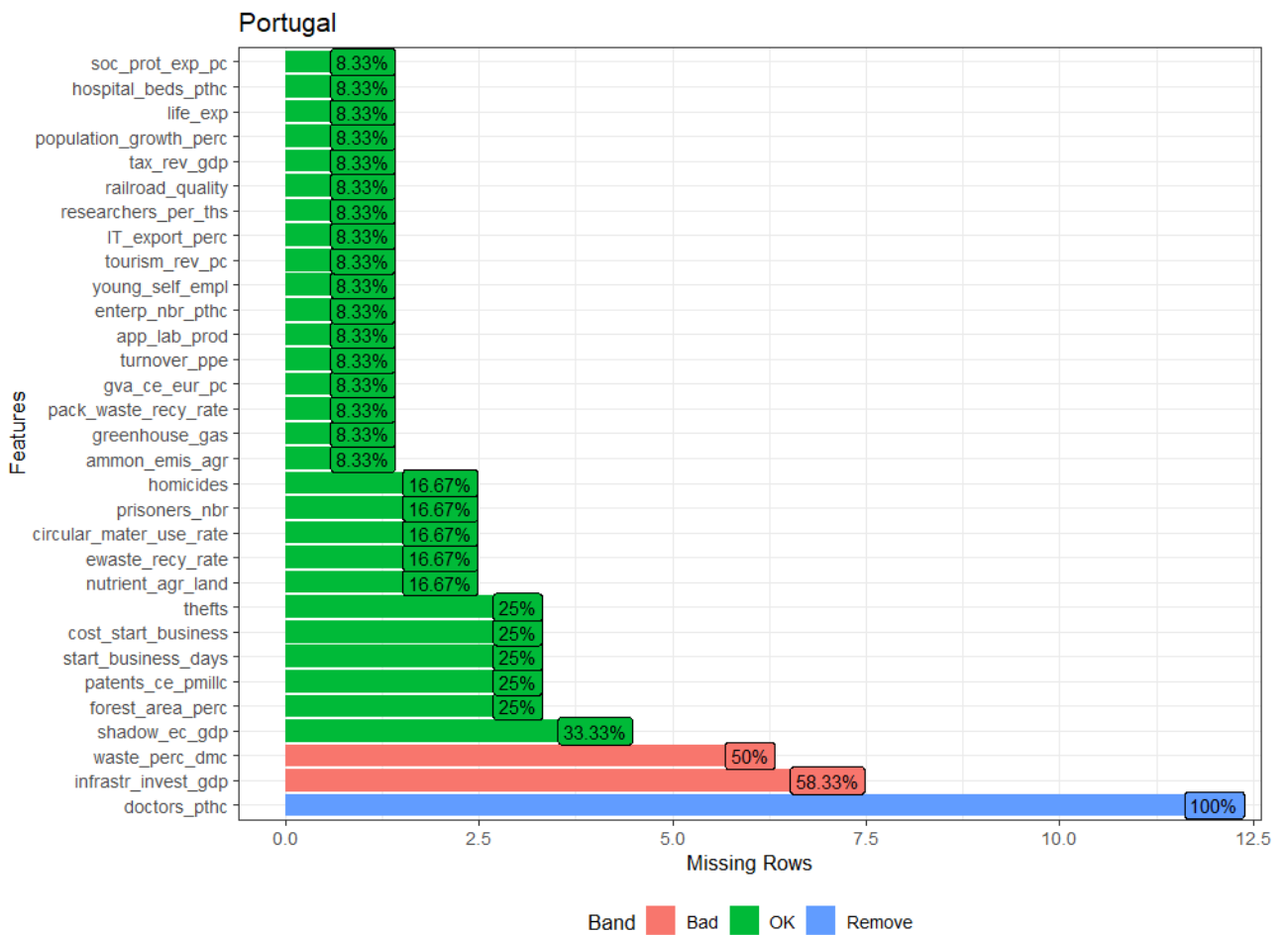
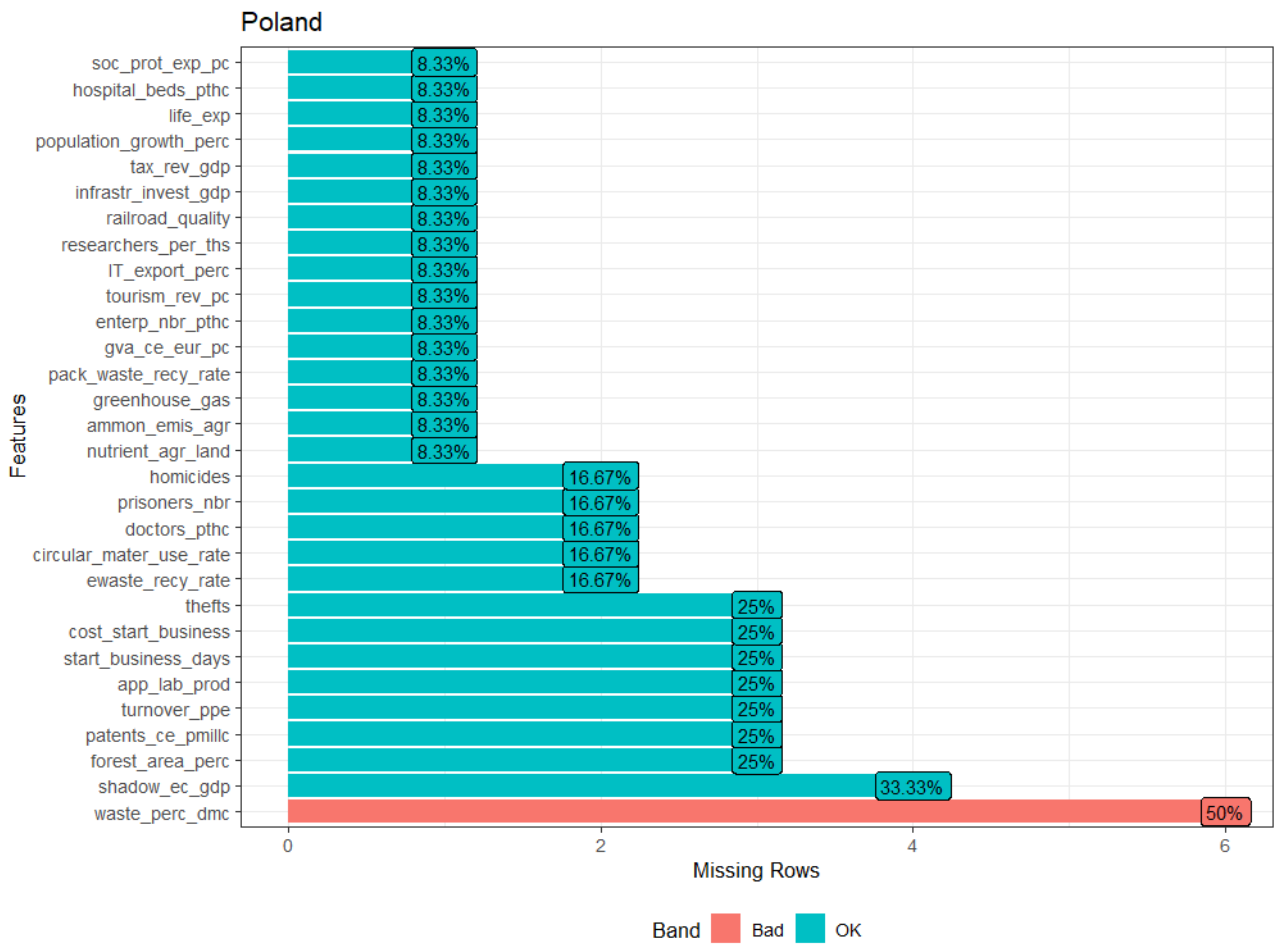


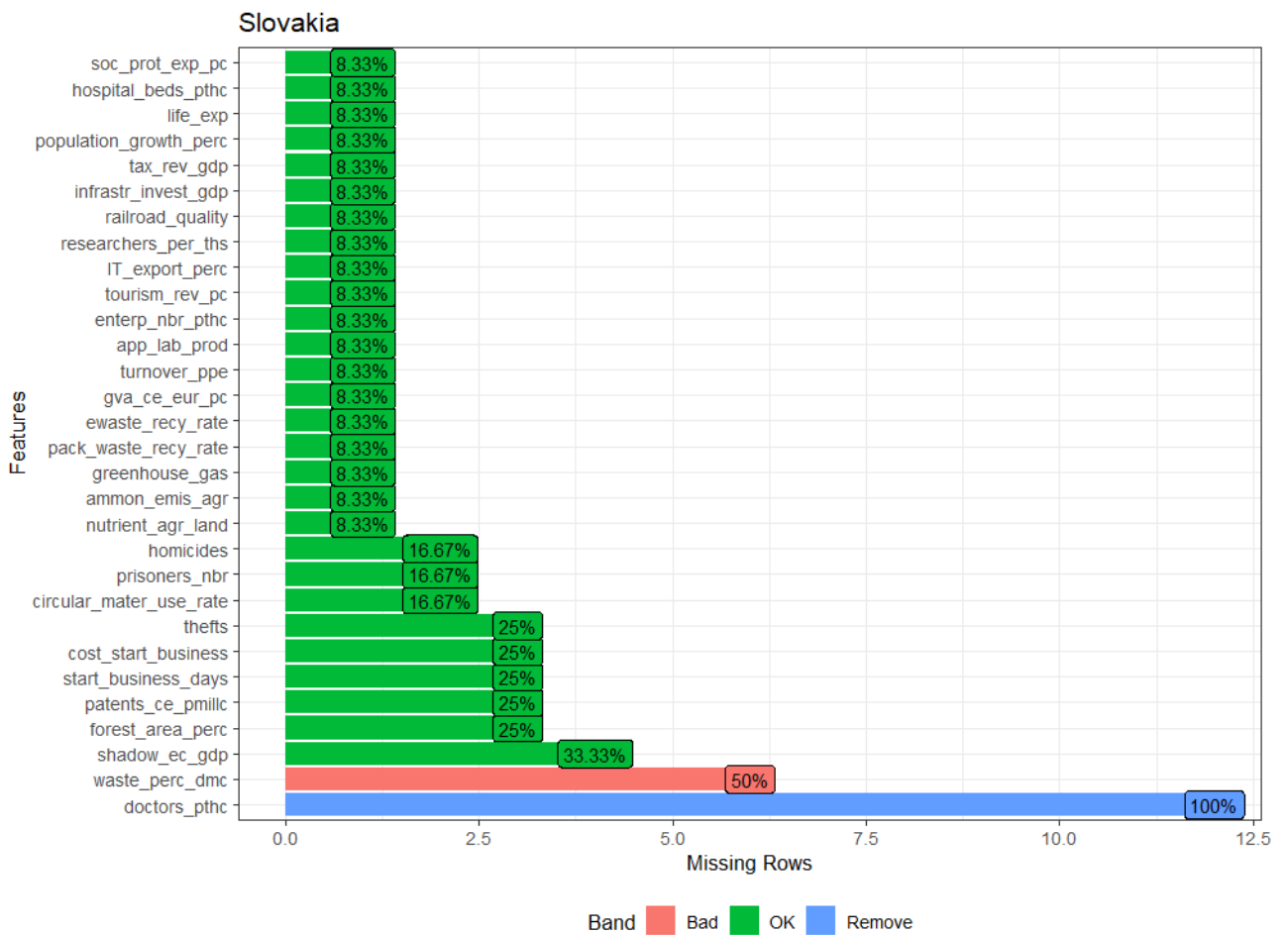
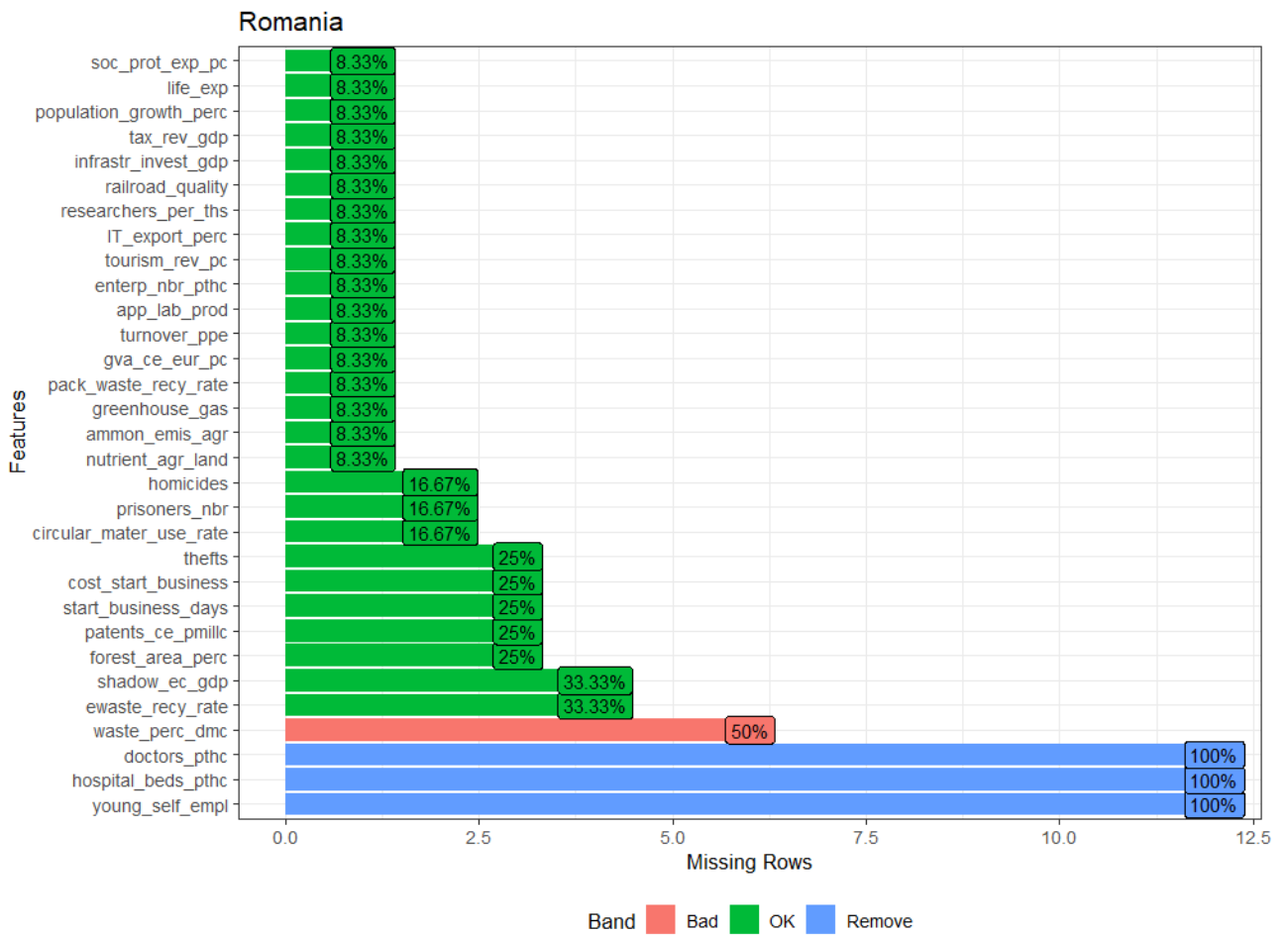




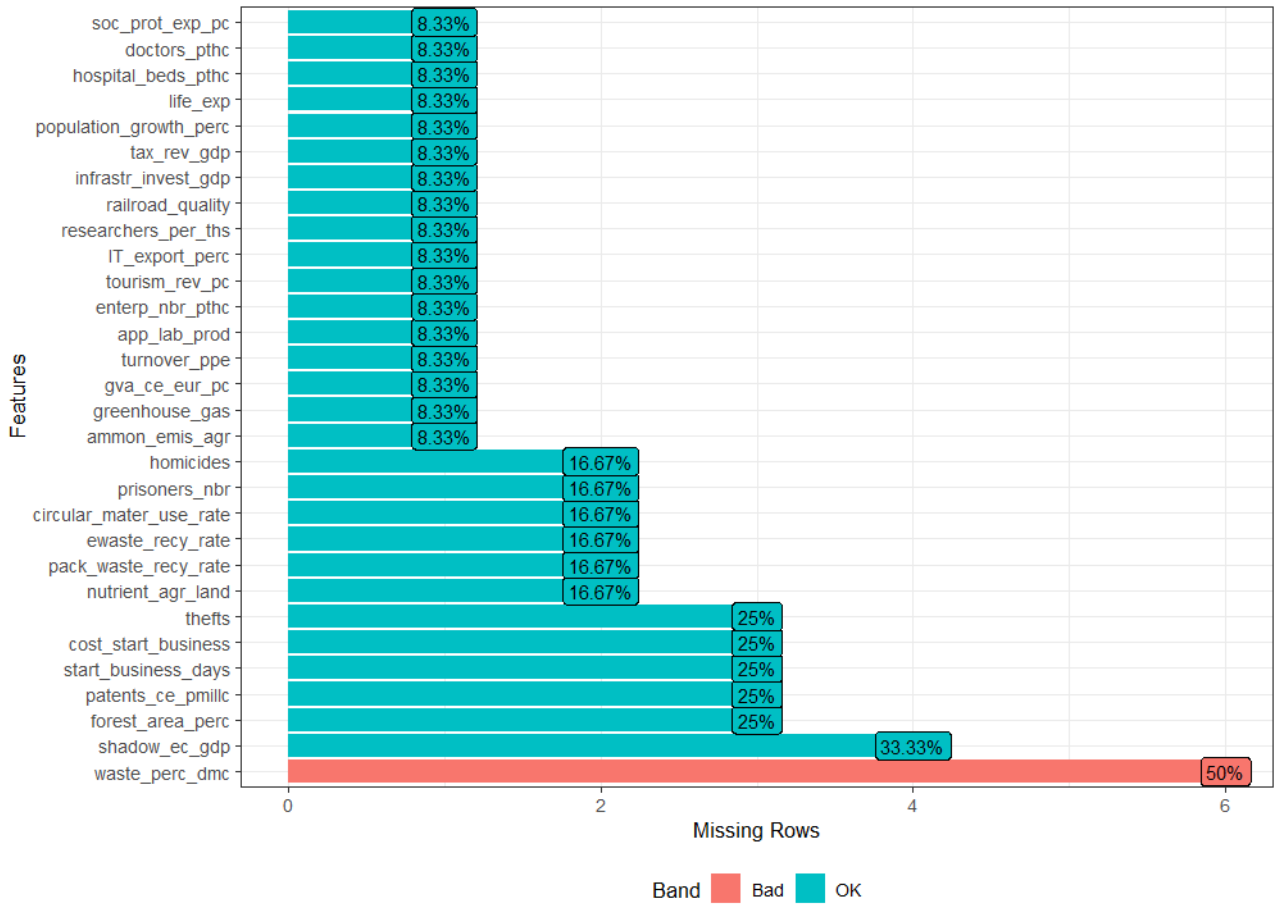




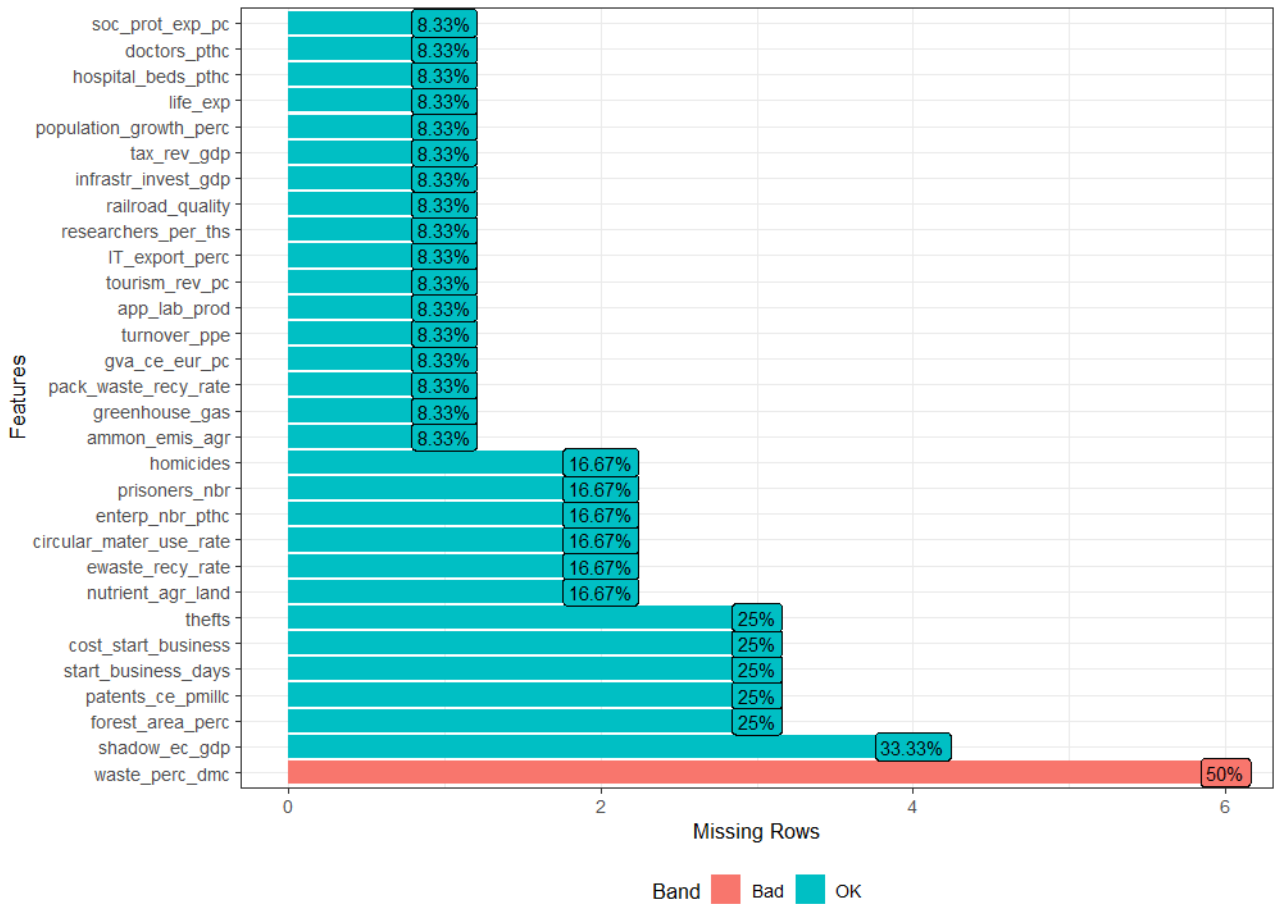


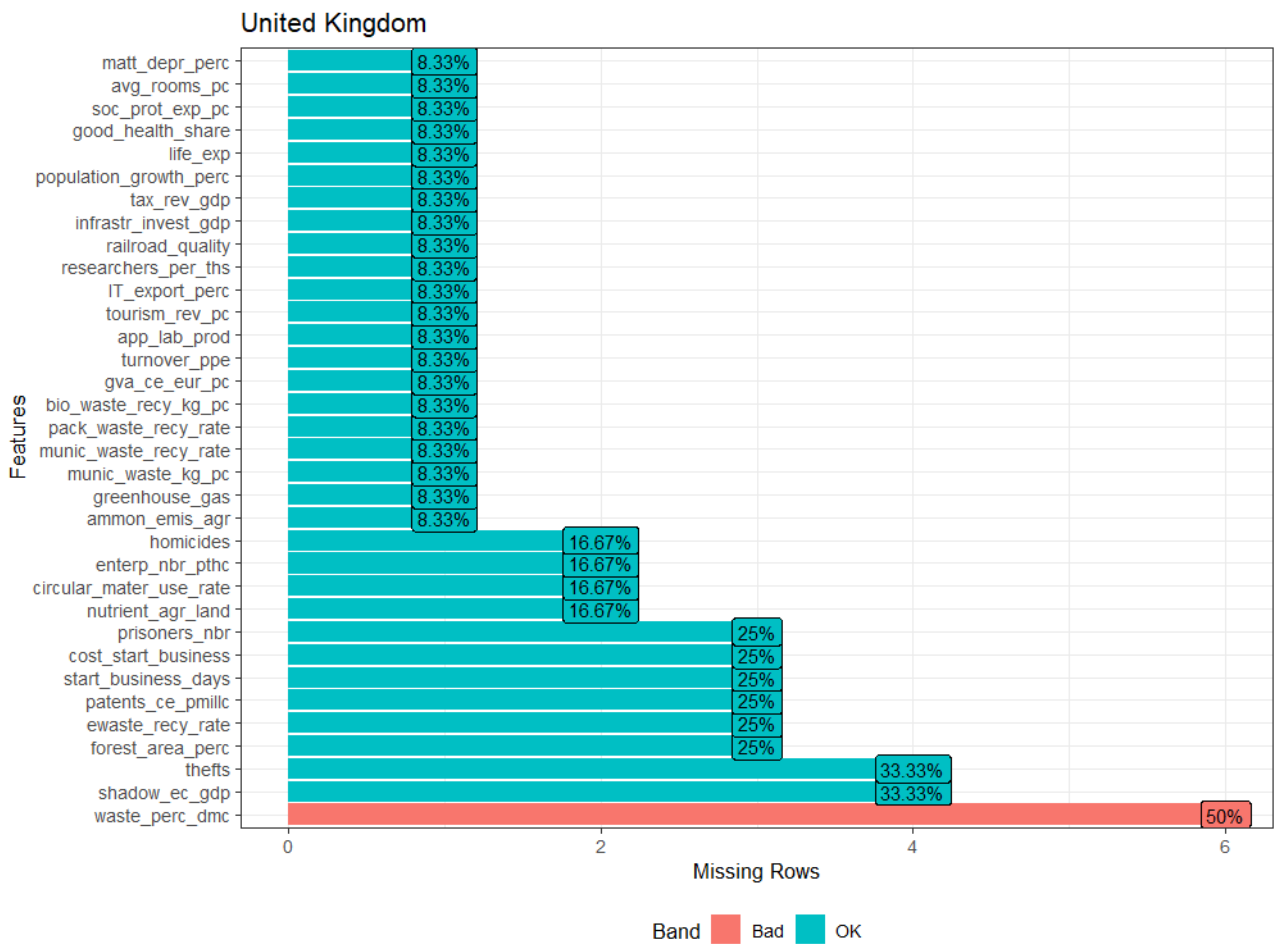
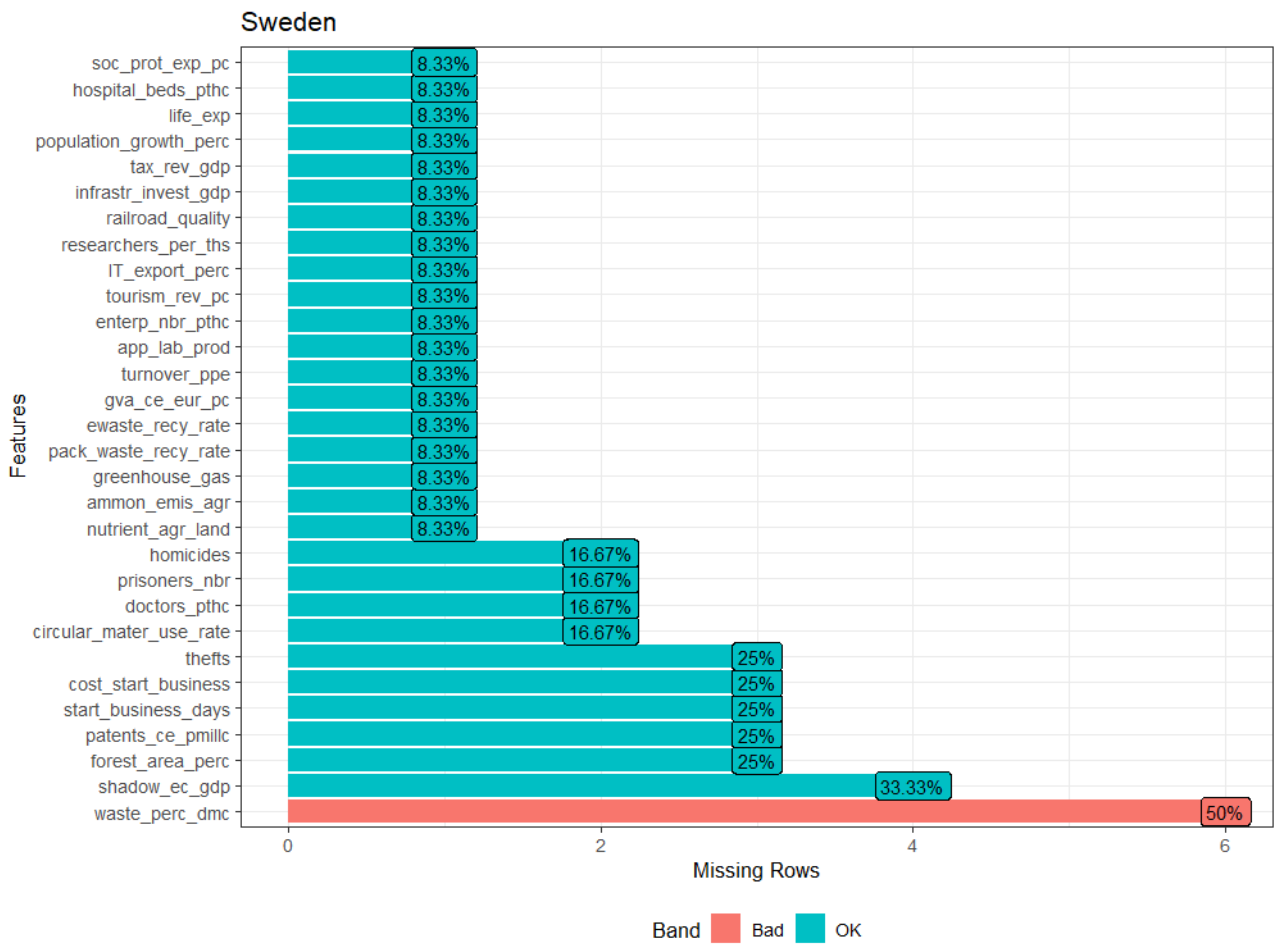


### Slovenia



### Spain





### 3 priedas. Požymių, kurie turėjo trūkstaų reikšmių, skaitinės charakteristikos prieš ir po užpildymo

forest_area_perc	area_organ_farm	nutrient_agr_land	ammon_emis_agr	greenhouse_gas	munic_waste_kg_pc	
Min. : 1.10	Min. : 0.060	Min. : -11.20	Min. : 6.20	Min. : 5.100	Min. : 247.0	
1st Qu.: 22.60	1st Qu.: 2.975	1st Qu.: 33.30	1st Qu.: 12.20	1st Qu.: 7.200	1st Qu.: 405.0	
Median : 33.40	Median : 6.045	Median : 47.80	Median : 18.80	Median : 9.000	Median : 482.0	
Mean : 33.59	Mean : 7.019	Mean : 64.49	Mean : 26.07	Mean : 9.797	Mean : 488.7	
3rd Qu.: 37.75	3rd Qu.: 9.293	3rd Qu.: 81.80	3rd Qu.: 31.25	3rd Qu.: 11.600	3rd Qu.: 564.8	
Max. : 73.10	Max. : 25.330	Max. : 214.90	Max. : 156.80	Max. : 27.500	Max. : 862.0	
NA's : 84	NA's : 4	NA's : 67	NA's : 29	NA's : 28	NA's : 6	
munic_waste_recy_rate	pack_waste_recy_rate	ewaste_recy_rate	bio_waste_recy_kg_pc	circular_mater_use_rate		
Min. : 0.90	Min. : 28.50	Min. : 9.70	Min. : 0.00	Min. : 1.200		
1st Qu.: 18.32	1st Qu.: 56.70	1st Qu.: 26.38	1st Qu.: 13.00	1st Qu.: 3.875		
Median : 33.25	Median : 62.35	Median : 35.25	Median : 43.50	Median : 7.050		
Mean : 32.94	Mean : 61.67	Mean : 35.60	Mean : 57.85	Mean : 8.739		
3rd Qu.: 46.75	3rd Qu.: 68.30	3rd Qu.: 43.00	3rd Qu.: 87.00	3rd Qu.: 11.600		
Max. : 67.20	Max. : 85.30	Max. : 105.20	Max. : 202.00	Max. : 29.700		
NA's : 6	NA's : 36	NA's : 64	NA's : 6	NA's : 56		
patents_ce_pmillic	turnover_ppe	app_lab_prod	enterp_nbr_pthc	annual_net_earnings	tourism_rev_pc	
Min. : 0.0000	Min. : 48.6	Min. : 8.60	Min. : 1.266	Min. : 4659	Min. : 80.37	
1st Qu.: 0.1375	1st Qu.: 103.8	1st Qu.: 21.50	1st Qu.: 3.067	1st Qu.: 15628	1st Qu.: 649.47	
Median : 0.4750	Median : 139.1	Median : 34.90	Median : 4.487	Median : 34549	Median : 1041.46	
Mean : 0.7648	Mean : 182.4	Mean : 42.51	Mean : 5.088	Mean : 35427	Mean : 1531.90	
3rd Qu.: 1.0100	3rd Qu.: 240.8	3rd Qu.: 61.70	3rd Qu.: 5.854	3rd Qu.: 54302	3rd Qu.: 1632.14	
Max. : 12.1000	Max. : 670.4	Max. : 152.70	Max. : 16.983	Max. : 84647	Max. : 10975.48	
NA's : 84	NA's : 35	NA's : 35	NA's : 33	NA's : 13	NA's : 28	
IT_export_perc	start_business_days	cost_start_business	tax_rev_gdp	shadow_ec_gdp	population_growth_perc	
Min. : 1.070	Min. : 3.50	Min. : 0.000	Min. : 10.23	Min. : 7.75	Min. : -2.2600	
1st Qu.: 2.542	1st Qu.: 6.00	1st Qu.: 0.900	1st Qu.: 17.55	1st Qu.: 11.67	1st Qu.: -0.2525	
Median : 4.110	Median : 11.50	Median : 2.100	Median : 20.93	Median : 18.28	Median : 0.2600	
Mean : 6.519	Mean : 13.85	Mean : 4.222	Mean : 20.99	Mean : 17.87	Mean : 0.2345	
3rd Qu.: 8.935	3rd Qu.: 20.62	3rd Qu.: 5.625	3rd Qu.: 24.62	3rd Qu.: 23.00	3rd Qu.: 0.6525	
Max. : 43.190	Max. : 39.00	Max. : 20.500	Max. : 46.05	Max. : 34.66	Max. : 3.4900	
NA's : 28	NA's : 84	NA's : 84	NA's : 26	NA's : 112	NA's : 28	
life_exp	good_health_share	soc_prot_exp_pc	prisoners_nbr	thefts	homicides	avg_rooms_pc
Min. : 71.81	Min. : 42.8	Min. : 731.8	Min. : 46.00	Min. : 62	Min. : 0.000	Min. : 1.000
1st Qu.: 76.89	1st Qu.: 60.0	1st Qu.: 2327.3	1st Qu.: 84.25	1st Qu.: 660	1st Qu.: 0.900	1st Qu.: 1.200
Median : 80.39	Median : 68.3	Median : 4683.7	Median : 112.00	Median : 1370	Median : 1.100	Median : 1.700
Mean : 79.17	Mean : 66.3	Mean : 6679.9	Mean : 131.61	Mean : 1453	Mean : 1.541	Mean : 1.659
3rd Qu.: 81.39	3rd Qu.: 74.0	3rd Qu.: 10486.6	3rd Qu.: 162.50	3rd Qu.: 1876	3rd Qu.: 1.600	3rd Qu.: 2.000
Max. : 83.43	Max. : 84.5	Max. : 22329.1	Max. : 359.00	Max. : 4246	Max. : 9.300	Max. : 2.400
NA's : 28	NA's : 3	NA's : 28	NA's : 70	NA's : 93	NA's : 60	NA's : 3
matt_depr_perc						
Min. : 0.500						
1st Qu.: 4.000						
Median : 6.300						
Mean : 9.489						
3rd Qu.: 11.800						
Max. : 45.700						
NA's : 3						
forest_area_perc	area_organ_farm	nutrient_agr_land	ammon_emis_agr	greenhouse_gas	munic_waste_kg_pc	
Min. : 1.10	Min. : 0.060	Min. : -11.20	Min. : 6.20	Min. : 5.100	Min. : 247.0	
1st Qu.: 22.70	1st Qu.: 3.015	1st Qu.: 35.48	1st Qu.: 12.40	1st Qu.: 7.200	1st Qu.: 406.5	
Median : 33.55	Median : 5.960	Median : 50.20	Median : 18.90	Median : 9.000	Median : 483.0	
Mean : 33.87	Mean : 6.985	Mean : 67.13	Mean : 26.02	Mean : 9.772	Mean : 490.4	
3rd Qu.: 37.15	3rd Qu.: 9.283	3rd Qu.: 83.21	3rd Qu.: 30.48	3rd Qu.: 11.525	3rd Qu.: 568.5	
Max. : 73.10	Max. : 25.330	Max. : 214.90	Max. : 156.80	Max. : 27.500	Max. : 862.0	
munic_waste_recy_rate	pack_waste_recy_rate	ewaste_recy_rate	bio_waste_recy_kg_pc	circular_mater_use_rate		
Min. : 0.90	Min. : 28.50	Min. : 9.70	Min. : 0.00	Min. : 1.200		
1st Qu.: 18.85	1st Qu.: 56.77	1st Qu.: 26.90	1st Qu.: 13.75	1st Qu.: 3.800		
Median : 33.40	Median : 62.65	Median : 35.70	Median : 45.00	Median : 7.150		
Mean : 33.06	Mean : 61.78	Mean : 35.94	Mean : 57.90	Mean : 8.678		
3rd Qu.: 46.60	3rd Qu.: 68.22	3rd Qu.: 43.52	3rd Qu.: 87.00	3rd Qu.: 11.625		
Max. : 67.20	Max. : 85.30	Max. : 105.20	Max. : 202.00	Max. : 29.700		
patents_ce_pmillic	turnover_ppe	app_lab_prod	enterp_nbr_pthc	annual_net_earnings	tourism_rev_pc	
Min. : 0.0000	Min. : 48.6	Min. : 8.60	Min. : 1.266	Min. : 4659	Min. : 80.37	
1st Qu.: 0.2000	1st Qu.: 104.6	1st Qu.: 22.02	1st Qu.: 3.122	1st Qu.: 15552	1st Qu.: 671.72	
Median : 0.4904	Median : 141.5	Median : 35.60	Median : 4.492	Median : 34092	Median : 1049.71	
Mean : 0.7753	Mean : 184.1	Mean : 43.01	Mean : 5.081	Mean : 35076	Mean : 1548.67	
3rd Qu.: 0.9841	3rd Qu.: 241.5	3rd Qu.: 62.10	3rd Qu.: 5.868	3rd Qu.: 53964	3rd Qu.: 1678.35	
Max. : 12.1000	Max. : 670.4	Max. : 152.70	Max. : 16.983	Max. : 84647	Max. : 10975.48	
IT_export_perc	start_business_days	cost_start_business	tax_rev_gdp	shadow_ec_gdp	population_growth_perc	
Min. : 1.070	Min. : 3.500	Min. : 0.000	Min. : 10.23	Min. : 7.75	Min. : -2.2600	
1st Qu.: 2.600	1st Qu.: 6.553	1st Qu.: 1.000	1st Qu.: 17.51	1st Qu.: 11.77	1st Qu.: -0.2425	
Median : 4.110	Median : 12.500	Median : 2.795	Median : 20.95	Median : 18.27	Median : 0.2650	
Mean : 6.462	Mean : 14.343	Mean : 4.483	Mean : 20.99	Mean : 17.66	Mean : 0.2422	
3rd Qu.: 8.906	3rd Qu.: 21.000	3rd Qu.: 6.500	3rd Qu.: 24.60	3rd Qu.: 22.74	3rd Qu.: 0.6425	
Max. : 43.190	Max. : 39.000	Max. : 20.500	Max. : 46.05	Max. : 34.66	Max. : 3.4900	
life_exp	good_health_share	soc_prot_exp_pc	prisoners_nbr	thefts	homicides	avg_rooms_pc
Min. : 71.81	Min. : 42.80	Min. : 731.8	Min. : 46.0	Min. : 62.0	Min. : 0.0000	Min. : 1.000
1st Qu.: 77.02	1st Qu.: 59.85	1st Qu.: 2417.0	1st Qu.: 83.5	1st Qu.: 695.5	1st Qu.: 0.8875	1st Qu.: 1.200
Median : 80.45	Median : 68.30	Median : 4696.1	Median : 112.0	Median : 1371.0	Median : 1.1000	Median : 1.700
Mean : 79.25	Mean : 66.25	Mean : 6744.1	Mean : 130.0	Mean : 1448.4	Mean : 1.4979	Mean : 1.657
3rd Qu.: 81.41	3rd Qu.: 74.00	3rd Qu.: 10555.4	3rd Qu.: 164.4	3rd Qu.: 1866.3	3rd Qu.: 1.6000	3rd Qu.: 2.000
Max. : 83.43	Max. : 84.50	Max. : 22329.1	Max. : 359.0	Max. : 4246.0	Max. : 9.3000	Max. : 2.400
matt_depr_perc						
Min. : 0.500						
1st Qu.: 4.000						
Median : 6.350						
Mean : 9.504						
3rd Qu.: 11.950						
Max. : 45.700						

#### 4 priedas. Klasteriuose esančių šalių požymių vidutinės reikšmės 2008 m.

	1 klasteris	2 klasteris	3 klasteris	4 klasteris	5 klasteris	6 klasteris	7 klasteris
<b>2008 m. k-vidurkių</b>	Kipras, Čekija, Graikija, Italija, Portugalija, Slovėnija, Ispanija	Malta	Belgija, Prancūzija, Vokietija, Airija, Nyderlandai, Jungtinė Karalystė	Austrija, Danija, Suomija, Švedija	Bulgarija, Kroatija, Vengrija, Lenkija, Rumunija, Slovakija	Lietuva, Latvija, Estija	Liuksem- burgas
forest_area_perc	35,2	1,1	19,78	50,48	31,6	46,93	35,7
area_organ_farm	6,14	0,4	2,95	10,1	2,9	7,7	2,7
nutrient_agr_land	74,16	214,9	91,68	61,58	41,83	28,33	121,6
ammon_emis_agr	26,27	156,8	33,17	18,45	15,98	9,67	42,2
greenhouse_gas	11,07	8,2	12,37	11,03	8,47	9,43	27,5
exp_envir_prot_gdp	0,77	1,5	0,98	0,45	0,65	0,9	0,8
envir_tax_rev_perc	6,99	10,16	6,54	6,82	7,8	6,71	7,01
share_renew_energy	12,26	0,2	5,88	30,71	12,76	22,1	2,81
munic_waste_kg_pc	522,29	674	577,67	609	418,67	388,33	697
munic_waste_recy_rate	19,33	3,6	44,78	47,75	9,37	11,7	46
pack_waste_recy_rate	53,86	45,9	66,7	60,7	45,97	47,33	63,6
ewaste_recy_rate	23,53	14,31	29,48	47,56	33,95	22,96	36,6
bio_waste_recy_kg_pc	35	0	85,83	104,25	5,5	9,67	138
circular_mater_use_rate	6,77	5,31	13,81	9,41	4,7	5,84	16,73
trade_recycl_materials_t_pc	0,11	0	0,18	0,17	0,02	0,06	6,02
patents_ce_pmille	0,31	0	0,74	0,81	0,19	0,11	2,27
fdi_gdp	8,37	163,22	12,98	4,22	14,6	5,24	12,74
cap_invest_gdp	27,36	22,13	22,65	24,57	29,74	31,76	20,69
export_gdp	40,92	148,5	55,62	50,36	52,05	54,43	187,11
export_growth	0,97	19,47	0,56	3,62	5,14	4,95	5,47
import_gdp	46,69	150,32	52,19	46,06	60,31	63,83	156,45
turnover_ppe	134,63	107,7	258,37	237,3	86,27	79,13	455,4
app_lab_prod	32,5	26	62,33	64,9	17,33	15,9	75,5
enterp_nbr_pthc	8,31	4,57	2,89	4,15	4,08	4,14	1,78
empl_rate	70,39	59,2	72,92	77,17	65,88	74,83	68,8
annual_net_earnings	27969,2	23694,81	51168,23	49151,96	10329,29	11677,27	68807,83
lab_prod_perc	-0,06	1,2	-1,07	-1,15	3,32	-1,13	-5,8
tourism_rev_pc	1678,22	2653,05	1083,01	1486,92	855,07	660,49	10165,38
exp_onRD_pc	232,89	80,1	639,67	1192,4	63,3	100,07	1279
patent_EPO_pmilc	37,73	65,95	213,41	287	4,81	9,32	560,73
IT_export_perc	5,22	43,19	7,57	7,8	9,3	4,35	3,03
pers_science_techn	26,97	27,8	31,72	36,75	24,25	27,33	41,4
roads_quality	4,64	3,28	5,54	6,02	2,96	4,02	5,74
airtransp_quality	5,2	5,61	6,13	6,28	3,94	5,1	5,04

start_business_days	14,87	32,73	10,22	14,66	22,1	13,79	16
cost_start_business	7,42	12,05	3,48	2,05	6,4	2,73	3,32
gov_spend_gdp	19,03	19,82	21,04	22,64	18,24	18,99	15,73
tax_rev_gdp	22,16	25,81	20,43	26,86	19,55	20,17	24,71
infl_cpi	4,41	4,3	3,33	3,52	6,85	12,23	3,4
shadow_ec_gdp	20,93	27,3	11,89	10,51	20,59	19,32	9,65
corrup_control_index	0,74	1	1,68	2,2	0,15	0,46	1,97
reg_quality_index	1,08	1,2	1,61	1,68	0,82	1,19	1,66
pol_stability_index	0,61	1,27	0,78	1,25	0,64	0,51	1,51
gov_eff_index	0,94	1,28	1,55	2,01	0,37	0,78	1,63
population_growth_per c	0,76	0,65	0,73	0,54	-0,41	-0,78	1,79
active_pop_perc	75,19	62,3	77,52	80,7	70,98	79,77	72,3
tertiary_educ_perc	22,1	19,8	29,08	18,3	15,67	24,8	24,9
adult_learn_perc	8,09	6,3	11	22,2	2,95	7,17	8,7
exp_educ_gdp	4,96	5,1	5,12	5,78	4,53	6,17	4,5
life_exp	79,42	79,64	80,1	79,89	74,23	72,67	80,54
good_health_share	65,33	73,8	74,97	72,47	59,72	48,27	74,1
health_spend_pc	1664,65	1306,39	2942,99	2876,68	943,69	876,26	4742,93
soc_prot_exp_pc	4646,5	2724,23	9043,49	10732,95	1666,19	1598,98	16339,19
prisoners_nbr	113,71	163	96,5	76,25	146,5	275,67	157
thefts	958	2000	2234,17	2986,5	541,5	1206,67	1377
homicides	0,97	1,5	1,27	1,27	1,75	6,77	1,6
exp_publ_safety_gdp	1,81	1,4	1,77	1,2	2,28	2,23	0,9
avg_rooms_pc	1,49	2	1,98	1,88	1,05	1,1	2
exp_housing_gdp	0,99	0,7	0,93	0,45	1,27	0,73	0,6
matt_depr_perc	7,8	4,3	4,67	3,3	22,57	12,23	0,7
gov_exp_cult_perc	1,07	1	1	1,12	1,37	2,8	1,4



## 5 priedas. Klasteriuose esančių šalių požymių vidutinės reikšmės 2013 m.

2013 m. k-vidurkių	1 klasteris	2 klasteris	3 klasteris	4 klasteris	5 klasteris	6 klasteris	7 klasteris
	Graikija, Italija, Portugalija, Slovėnija, Ispanija	Kipras, Malta	Austrija, Danija, Suomija, Švedija	Bulgarija, Kroatija, Rumunija	Liuksemburgas	Belgija, Prancūzija, Vokietija, Airija, Nyderlandai, Jungtinė Karalystė	Čekija, Estija, Vengrija, Latvija, Lietuva, Lenkija, Slovakija
forest_area_perc	39,16	9,9	50,77	32,8	35,7	20,1	38,34
area_organ_farm	7,64	2,04	12,6	2,11	3,39	3,58	8,58
nutrient_agr_land	52,34	162,85	53,72	21,3	127,3	95,2	41,89
ammon_emis_agr	20,72	86,7	18,4	12,87	40,5	32,25	13,54
greenhouse_gas	7,94	8,8	9,47	6,43	22,7	10,87	9,44
exp_envir_prot_gdp	0,98	0,8	0,42	0,8	0,8	1	0,76
envir_tax_rev_perc	8,17	8,4	6,67	8,53	5,66	6,71	7,59
share_renew_energy	19,26	6,11	36,84	23,61	3,5	8,85	19,54
munic_waste_kg_pc	456,2	598,5	584,75	363,33	616	528,34	337,43
munic_waste_recy_rate	29,66	11,2	45,42	18,87	46,6	47,67	21,16
pack_waste_recy_rate	63,24	47,35	66,58	59,1	62,8	70,35	54,86
ewaste_recy_rate	24,7	11,55	44,1	42,87	29,3	30,6	33,96
bio_waste_recy_kg_pc	50,2	4,5	114,5	15	111	89,66	16,29
circular_mater_use_rate	7,7	4,35	8,53	2,9	15,3	14,72	7,3
trade_recycl_materials_t_p c	0,15	0	0,11	0,01	4,48	0,17	0,09
patents_ce_pmillc	0,44	0	1,38	0,04	6,53	0,79	0,8
fdi_gdp	2,56	56,73	-0,32	2,4	25,93	11,06	1,59
cap_invest_gdp	15,99	14,55	22,06	21,98	19,02	19,72	22,53
export_gdp	41,15	108,66	47,2	48,39	190,63	61,27	74,82
export_growth	3,3	1,05	0,43	10,9	5,26	1,68	3,81
import_gdp	39,3	104,07	44,05	49,52	158,61	55,73	72,04
turnover_ppe	142,2	123,45	262,73	65,17	605,4	274,3	106,91
app_lab_prod	32,08	32,85	67,38	14,67	80,7	65,3	20,36
enterp_nbr_pthc	6,17	5,43	3,81	3,77	1,56	2,96	7,42
empl_rate	60,76	66,7	75,5	61,8	71,1	71,78	68,33
annual_net_earnings	29587,12	32303,44	57283,09	9608,41	75724,68	54046,02	13479,53
lab_prod_perc	0,16	0,3	0,18	2,67	1,8	0,22	0,74
tourism_rev_pc	1351,31	3351,06	1437,76	995,69	10481,55	1157,99	665,94
exp_onRD_pc	286,56	120,4	1310,88	49,23	1127,9	754,2	151,29
patent_EPO_pmilc	34,92	74,26	327,95	2,39	789,53	198,18	16,89
IT_export_perc	1,71	11,34	4,12	2,96	2	4,98	10,56
pers_science_techn	25,58	29,05	41,73	22,93	55,1	36,52	28,64
roads_quality	5,19	4,42	5,82	3,51	5,79	5,74	3,8

airtransp_quality	5,12	5,52	5,75	3,97	5,59	5,96	4,38
start_business_days	12,5	23,25	16	18,5	16,5	9,5	19,36
cost_start_business	6,84	13,95	1,62	4,47	1,9	2,8	4,96
gov_spend_gdp	19,72	18,61	24,17	17,32	17,3	21,82	18,44
tax_rev_gdp	20,49	24,66	26,84	19,25	25,88	20,03	18,14
infl_cpi	0,76	0,4	1,07	2,37	1,7	1,47	1,33
shadow_ec_gdp	24	30,9	12,33	23,87	10,65	11,6	16,71
corrup_control_index	0,52	1,12	2,11	-0,11	2,12	1,68	0,45
reg_quality_index	0,76	1,11	1,77	0,54	1,78	1,52	1,09
pol_stability_index	0,39	0,8	1,21	0,33	1,34	0,81	0,9
gov_eff_index	0,86	1,31	1,92	0,26	1,63	1,57	0,82
population_growth_perc	-0,06	1,08	0,58	-0,4	2,31	0,46	-0,38
active_pop_perc	74,5	74,95	80,9	70,33	75,4	78,27	76,33
tertiary_educ_perc	24,18	35,95	21,27	18,8	31,6	32,17	25,73
adult_learn_perc	8,6	7,45	24,72	2,37	14,6	12,45	6,56
exp_educ_gdp	4,94	6	6,2	3,77	4,9	5,28	5,06
life_exp	81,62	80,88	81,09	75,68	81,8	81,09	76,03
good_health_share	64,62	74,1	71,42	60,6	71,9	72,93	55,07
health_spend_pc	1788,43	1341,82	3581,04	897,94	3852,67	3497,03	1302,35
soc_prot_exp_pc	5363,23	4106,81	13144,52	1429,34	19764,45	10327,76	2187,94
prisoners_nbr	113,8	98	74,25	131	120	98,17	218,71
thefts	1145,2	1070,5	2808	443	1897	2130	912,29
homicides	0,96	1,35	1,02	1,43	0,2	1,08	2,5
exp_publ_safety_gdp	1,92	1,75	1,23	2,37	1	1,75	1,99
avg_rooms_pc	1,56	2,05	1,9	1,07	2,1	2,07	1,31
exp_housing_gdp	0,52	1,1	0,4	1,57	0,5	0,67	0,7
matt_depr_perc	11,28	13,15	3,05	29,17	1,8	6,02	14,87
gov_exp_cult_perc	0,74	1,1	1,07	1,23	1,1	0,97	1,89

## 6 priedas. Klasteriuose esančių šalių požymių vidutinės reikšmės 2019 m.

2019 m. k-vidurkių	1 klasteris	2 klasteris	3 klasteris	4 klasteris	5 klasteris	6 klasteris	7 klasteris
	Čekija, Estija, Vengrija, Latvija, Lietuva, Lenkija, Slovakija	Kipras, Malta	Austrija, Danija, Suomija, Švedija	Bulgarija, Kroatija, Rumunija	Liuksemburgas	Belgija, Prancūzija, Vokietija, Airija, Nyderlandai, Jungtinė Karalystė	Graikija, Italija, Portugalija, Slovėnija, Ispanija
forest_area_perc	39,19	13,91	52,08	33,69	33,8	24,39	38,63
area_organ_farm	11,42	2,73	17,53	4,13	4,42	5,05	10,72
nutrient_agr_land	45,46	160,04	55,47	41,04	124,69	89,26	53,96
ammon_emis_agr	14,98	71,22	19,2	13,62	42,9	32,73	21,05
greenhouse_gas	9,6	8,92	8,91	6,88	21,18	10,48	8,06
exp_envir_prot_gdp	0,61	0,85	0,38	0,7	0,9	0,88	0,88
envir_tax_rev_perc	7,4	7,78	5,98	9,21	4,4	6,24	7,77
share_renew_energy	22,32	11,14	42,58	24,77	7,05	12,93	21,76
munic_waste_kg_pc	417,71	668	611,75	385,11	791	527,3	504,2
munic_waste_recy_rate	37,57	11,95	49,95	23,78	48,9	51,5	39,02
pack_waste_recy_rate	60,21	53,44	68,34	59,01	68	70,57	65,18
ewaste_recy_rate	43,84	27,92	50,62	53,09	44,07	42,79	37,25
bio_waste_recy_kg_pc	43,71	4,5	121,25	15,52	152	97,92	76,6
circular_mater_use_rate	7,83	5	8,12	2,93	11,9	17,17	9,26
trade_recycl_materials_t_pc	0,08	0	0,13	0,04	3,78	0,18	0,17
patents_ce_pmillc	0,72	0,53	1,03	0,2	4,84	0,85	0,35
fdi_gdp	6,17	62,71	0,96	2,29	-16,06	-1,5	2,22
cap_invest_gdp	23,7	21,12	24,4	22,24	17,43	25,49	18,22
export_gdp	73,58	107,84	50,27	52,17	208,75	67,04	46,17
export_growth	4,4	2,86	4,73	4,88	0,84	3,78	3,14
import_gdp	70,35	101,41	46,62	52,47	172,76	62,18	43,22
turnover_ppe	121,99	148,44	264,76	83,06	603,55	294,76	149,64
app_lab_prod	25,78	38,85	72,28	19,53	87,61	76,82	35,91
enterp_nbr_pthc	7,98	5,57	3,7	4,53	1,38	3	6,24
empl_rate	76,83	76,25	78,6	70,87	72,8	76,2	69,04
annual_net_earnings	19929,92	35458,87	61656,73	13859,86	84646,68	61247,59	31811,66
lab_prod_perc	3,03	-0,4	0,5	2,43	-1,3	0,5	0,44
tourism_rev_pc	892,33	3451,33	1670,26	1161,49	9231,26	1412,35	1501,66
exp_onRD_pc	224,86	162,3	1453,33	91,87	1233,1	926,02	349
patent_EPO_pmilc	15,39	83,56	353,8	3,86	695,56	226,64	42,32
IT_export_perc	9,71	9,24	4,58	3,28	2,71	5,46	2,63
pers_science_techn	31,91	32,2	45,15	26,57	55,7	40,02	29,22
roads_quality	4,19	4,2	5,55	4	5,5	5,13	5,12

airtransp_quality	4,77	5,3	5,75	4,63	5,6	5,63	5,1
start_business_days	14,93	13,25	11,25	20,83	16,5	6	8,4
cost_start_business	3,03	6,15	1,53	2,5	1,6	2,77	4,22
gov_spend_gdp	18,97	16,87	23,04	17,95	17,1	20,31	18,47
tax_rev_gdp	19,02	24,76	26,41	19,13	26,18	19,86	20,8
infl_cpi	2,63	0,95	1,27	2,57	1,7	1,52	0,7
shadow_ec_gdp	17,55	27,68	11,74	22,52	10,51	11,56	21,6
corrup_control_index	0,59	0,42	1,98	-0,05	2,11	1,66	0,51
reg_quality_index	1,12	0,98	1,67	0,53	1,7	1,59	0,9
pol_stability_index	0,7	0,8	0,99	0,61	1,36	0,62	0,6
gov_eff_index	0,85	0,92	1,8	0,16	1,73	1,42	0,82
population_growth_perc	-0,11	1,55	0,65	-0,63	2,22	0,59	0,08
active_pop_perc	80,41	80,4	83	74,4	76,8	79,82	76,94
tertiary_educ_perc	27,06	41,35	29,3	21,03	40,6	35,7	28,18
adult_learn_perc	8,13	8,9	25,82	2,27	19,1	13,8	8,86
exp_educ_gdp	5,03	5,35	5,9	4,1	4,7	4,8	4,36
life_exp	77,64	81,41	81,58	76,11	82,14	81,58	81,93
good_health_share	56,47	76	71,38	66,3	71,9	72,89	68,86
health_spend_pc	1820,45	2019,77	4382,54	1469,42	4697,04	4429,81	2212,08
soc_prot_exp_pc	3044,9	4789,39	13563,78	2140,5	20409,79	11303,74	5764,5
prisoners_nbr	186,66	97,5	72,41	119,31	123,79	102,01	107,39
thefts	930,74	1029,99	2717,07	533,3	1859,45	1907,05	1130,03
homicides	1,99	1,14	1,05	1,4	0,63	1,15	0,85
exp_publ_safety_gdp	1,97	1,5	1,2	2,4	1,2	1,57	1,8
avg_rooms_pc	1,43	2,2	1,95	1,2	2,2	2,09	1,6
exp_housing_gdp	0,63	1	0,38	1,43	0,6	0,62	0,4
matt_depr_perc	6,2	6,35	2,35	13,87	1,3	4,06	7,3
gov_exp_cult_perc	2,14	1,25	1,03	1,27	1,3	0,93	0,82

## 7 priedas. Pirmosios ir antrosios pagrindinės komponentės svoriai 2008 m.

Kintamasis	PK1	Kintamasis	PK1
soc_prot_exp_pc	-0,2025	greenhouse_gas	-0,1212
annual_net_earnings	-0,2001	prisoners_nbr	0,1116
health_spend_pc	-0,1986	pol_stability_index	-0,1064
app_lab_prod	-0,1985	gov_spend_gdp	-0,101
corrup_control_index	-0,1973	empl_rate	-0,0961
exp_onRD_pc	-0,1896	tourism_rev_pc	-0,09
turnover_ppe	-0,1876	trade_recycl_materials_t_pc	-0,0894
gov_eff_index	-0,1873	start_business_days	0,0841
patent_EPO_pmilc	-0,1817	active_pop_perc	-0,0822
avg_rooms_pc	-0,1808	homicides	0,0804
reg_quality_index	-0,1807	nutrient_agr_land	-0,08
munic_waste_recy_rate	-0,1785	gov_exp_cult_perc	0,0717
life_exp	-0,1661	exp_housing_gdp	0,0684
matt_depr_perc	0,1637	ewaste_recy_rate	-0,067
pers_science_techn	-0,1624	tertiary_educ_perc	-0,0664
bio_waste_recy_kg_pc	-0,1601	export_gdp	-0,0662
exp_publ_safety_gdp	0,1578	cost_start_business	0,0597
airtransp_quality	-0,1577	tax_rev_gdp	-0,0588
roads_quality	-0,1567	enterp_nbr_pthc	0,0562
thefts	-0,1506	export_growth	0,0547
patents_ce_pmilc	-0,1489	ammon_emis_agr	-0,0412
circular_mater_use_rate	-0,1465	exp_educ_gdp	-0,0362
adult_learn_perc	-0,1458	envir_tax_rev_perc	0,0322
pack_waste_recy_rate	-0,1441	area_organ_farm	-0,0301
shadow_ec_gdp	0,1441	import_gdp	-0,0244
cap_invest_gdp	0,1428	forest_area_perc	0,0116
infl_cpi	0,1408	fdi_gdp	0,0096
population_growth_perc	-0,1404	exp_envir_prot_gdp	-0,0082
good_health_share	-0,1387	share_renew_energy	0,0054
munic_waste_kg_pc	-0,1326	IT_export_perc	0,0051
lab_prod_perc	0,1234		

<b>Kintamasis</b>	<b>PK2</b>	<b>Kintamasis</b>	<b>PK2</b>
ammon_emis_agr	-0,2782	tax_rev_gdp	-0,0855
fdi_gdp	-0,2713	avg_rooms_pc	-0,0837
share_renew_energy	0,2558	exp_onRD_pc	0,0836
active_pop_perc	0,2411	munic_waste_recy_rate	0,0827
forest_area_perc	0,2378	roads_quality	0,0814
import_gdp	-0,2376	airtransp_quality	0,0622
nutrient_agr_land	-0,2296	thefts	0,062
export_gdp	-0,2194	life_exp	-0,0601
empl_rate	0,2137	pol_stability_index	-0,0578
area_organ_farm	0,2097	tertiary_educ_perc	-0,0538
IT_export_perc	-0,208	exp_housing_gdp	-0,0491
cost_start_business	-0,1964	pack_waste_recy_rate	0,0475
exp_envir_prot_gdp	-0,1547	bio_waste_recy_kg_pc	0,0466
tourism_rev_pc	-0,1536	exp_publ_safety_gdp	0,0405
envir_tax_rev_perc	-0,1453	turnover_ppe	-0,0389
export_growth	-0,1438	corrup_control_index	0,0345
shadow_ec_gdp	-0,1369	pers_science_techn	0,0304
ewaste_recy_rate	0,1298	gov_eff_index	0,0295
population_growth_perc	-0,1214	prisoners_nbr	0,0268
start_business_days	-0,1206	annual_net_earnings	-0,0227
munic_waste_kg_pc	-0,12	lab_prod_perc	-0,0213
good_health_share	-0,1187	reg_quality_index	0,0168
gov_exp_cult_perc	0,1179	matr_depr_perc	0,0117
adult_learn_perc	0,1168	patents_ce_pmille	-0,0112
exp_educ_gdp	0,1036	soc_prot_exp_pc	0,0081
cap_invest_gdp	0,098	enterp_nbr_pthc	0,0056
trade_recycl_materials_t_pc	-0,0973	health_spend_pc	-0,0048
greenhouse_gas	-0,0961	circular_mater_use_rate	-0,0044
homicides	0,0887	patent_EPO_pmilc	-0,004
gov_spend_gdp	0,0886	app_lab_prod	0,0027
infl_cpi	0,0863		

## 8 priedas. Pirmosios, antrosios, trečiosios pagrindinės komponentės svoriai 2013 m.

Kintamasis	PK1	Kintamasis	PK1
annual_net_earnings	-0,2046	greenhouse_gas	-0,1157
soc_prot_exp_pc	-0,2039	active_pop_perc	-0,1147
corrup_control_index	-0,1997	pack_waste_recy_rate	-0,1128
health_spend_pc	-0,1993	export_growth	0,1121
app_lab_prod	-0,1961	exp_educ_gdp	-0,1098
exp_onRD_pc	-0,196	nutrient_agr_land	-0,106
gov_eff_index	-0,1941	tourism_rev_pc	-0,0958
pers_science_techn	-0,1919	trade_recycl_materials_t_pc	-0,0924
turnover_ppe	-0,1846	tax_rev_gdp	-0,0907
patent_EPO_pmilc	-0,1828	exp_housing_gdp	0,0866
reg_quality_index	-0,1817	envir_tax_rev_perc	0,0821
avg_rooms_pc	-0,1765	homicides	0,076
thefts	-0,167	enterp_nbr_pthc	0,0735
adult_learn_perc	-0,1662	ammon_emis_agr	-0,0624
matt_depr_perc	0,1657	cost_start_business	0,0608
bio_waste_recy_kg_pc	-0,1638	gov_exp_cult_perc	0,0576
munic_waste_recy_rate	-0,1637	lab_prod_perc	0,0561
exp_publ_safety_gdp	0,1598	tertiary_educ_perc	-0,0486
airtransp_quality	-0,157	export_gdp	-0,048
roads_quality	-0,154	start_business_days	0,047
life_exp	-0,1517	IT_export_perc	0,0378
munic_waste_kg_pc	-0,1503	area_organ_farm	-0,0347
population_growth_perc	-0,1494	import_gdp	-0,0312
empl_rate	-0,1378	fdi_gdp	-0,0288
gov_spend_gdp	-0,1351	exp_envir_prot_gdp	0,0198
shadow_ec_gdp	0,1246	ewaste_recy_rate	-0,0119
prisoners_nbr	0,1224	infl_cpi	0,0117
pol_stability_index	-0,1191	share_renew_energy	0,0071
circular_mater_use_rate	-0,1179	cap_invest_gdp	0,0061
patents_ce_pmille	-0,1163	forest_area_perc	0,0015
good_health_share	-0,1161		

<b>Kintamasis</b>	<b>PK2</b>	<b>Kintamasis</b>	<b>PK2</b>
share_renew_energy	-0,2933	envir_tax_rev_perc	0,0733
ammon_emis_agr	0,2618	bio_waste_recy_kg_pc	-0,0725
forest_area_perc	-0,2548	prisoners_nbr	-0,0717
nutrient_agr_land	0,2515	tax_rev_gdp	0,0683
cap_invest_gdp	-0,2351	turnover_ppe	0,0674
area_organ_farm	-0,2297	thefts	-0,0658
cost_start_business	0,2195	reg_quality_index	-0,063
ewaste_recy_rate	-0,205	start_business_days	0,0623
tourism_rev_pc	0,1943	pol_stability_index	-0,0583
fdi_gdp	0,1936	infl_cpi	-0,0538
export_gdp	0,1874	IT_export_perc	0,0432
import_gdp	0,1853	exp_educ_gdp	-0,0423
active_pop_perc	-0,1683	exp_housing_gdp	0,042
population_growth_perc	0,1669	health_spend_pc	-0,0413
shadow_ec_gdp	0,1667	annual_net_earnings	0,0395
exp_envir_prot_gdp	0,157	roads_quality	-0,0372
empl_rate	-0,1551	gov_eff_index	-0,037
tertiary_educ_perc	0,1519	corrup_control_index	-0,0313
good_health_share	0,1409	patents_ce_pmille	0,0311
adult_learn_perc	-0,1359	enterp_nbr_pthc	-0,031
gov_exp_cult_perc	-0,125	matr_depr_perc	0,0281
trade_recycl_materials_t_pc	0,1174	airtransp_quality	0,0263
munic_waste_kg_pc	0,11	exp_publ_safety_gdp	0,0253
exp_onRD_pc	-0,1095	pers_science_techn	-0,0229
life_exp	0,1043	app_lab_prod	0,0187
homicides	-0,1039	patent_EPO_pmilc	0,0141
pack_waste_recy_rate	-0,0992	lab_prod_perc	-0,0134
munic_waste_recy_rate	-0,0985	circular_mater_use_rate	0,0125
gov_spend_gdp	-0,0938	export_growth	0,0018
greenhouse_gas	0,0923	soc_prot_exp_pc	0,0018
avg_rooms_pc	0,0828		



<b>Kintamasis</b>	<b>PK3</b>	<b>Kintamasis</b>	<b>PK3</b>
import_gdp	0,3155	population_growth_perc	0,0763
export_gdp	0,3131	fdi_gdp	-0,0686
trade_recycl_materials_t_pc	0,2496	exp_housing_gdp	-0,0671
patents_ce_pmille	0,2419	avg_rooms_pc	-0,0569
gov_exp_cult_perc	0,238	tertiary_educ_perc	-0,0564
prisoners_nbr	0,2182	annual_net_earnings	-0,0535
cap_invest_gdp	0,2171	exp_educ_gdp	-0,0529
pol_stability_index	0,216	envir_tax_rev_perc	-0,051
greenhouse_gas	0,2015	forest_area_perc	0,0502
tourism_rev_pc	0,1853	adult_learn_perc	-0,0464
infl_cpi	0,1813	munic_waste_recy_rate	-0,0446
IT_export_perc	0,1785	tax_rev_gdp	-0,0418
life_exp	-0,1755	share_renew_energy	-0,0411
lab_prod_perc	0,1688	bio_waste_recy_kg_pc	-0,0394
gov_spend_gdp	-0,1662	circular_mater_use_rate	0,0381
shadow_ec_gdp	-0,1392	active_pop_perc	-0,0364
roads_quality	-0,1371	app_lab_prod	-0,033
patent_EPO_pmilc	0,1334	gov_eff_index	-0,032
airtransp_quality	-0,1215	ammon_emis_agr	0,0318
good_health_share	-0,1203	exp_envir_prot_gdp	-0,0266
pers_science_techn	0,1198	ewaste_recy_rate	0,0245
homicides	0,115	enterp_nbr_pthc	0,0243
empl_rate	0,1148	thefts	-0,0222
export_growth	0,1139	health_spend_pc	-0,0213
munic_waste_kg_pc	-0,1125	nutrient_agr_land	-0,0171
start_business_days	0,1069	area_organ_farm	0,0108
pack_waste_recy_rate	-0,1038	exp_onRD_pc	-0,0092
turnover_ppe	0,102	matt_depr_perc	-0,0078
cost_start_business	-0,0889	corrup_control_index	-0,0051
reg_quality_index	0,0871	soc_prot_exp_pc	-0,0029
exp_publ_safety_gdp	-0,0816		

## 9 priedas. Pirmosios, antrosios, trečiosios pagrindinės komponentės svoriai 2019 m.

Kintamasis	PK1	Kintamasis	PK1
soc_prot_exp_pc	-0,208	enterp_nbr_pthc	0,107
annual_net_earnings	-0,2059	trade_recycl_materials_t_pc	-0,1065
patent_EPO_pmilc	-0,1974	pol_stability_index	-0,1061
pers_science_techn	-0,1967	circular_mater_use_rate	-0,102
corrup_control_index	-0,1966	nutrient_agr_land	-0,0984
health_spend_pc	-0,1953	tertiary_educ_perc	-0,0974
exp_onRD_pc	-0,1922	good_health_share	-0,0923
gov_eff_index	-0,191	tax_rev_gdp	-0,0903
turnover_ppe	-0,1863	gov_spend_gdp	-0,0872
reg_quality_index	-0,1849	homicides	0,0866
app_lab_prod	-0,1828	exp_housing_gdp	0,0859
thefts	-0,1751	ammon_emis_agr	-0,0836
exp_publ_safety_gdp	0,1705	exp_educ_gdp	-0,0822
avg_rooms_pc	-0,1695	infl_cpi	0,0801
adult_learn_perc	-0,168	active_pop_perc	-0,0795
life_exp	-0,1613	empl_rate	-0,0727
matt_depr_perc	0,157	export_gdp	-0,0715
population_growth_perc	-0,1569	start_business_days	0,0681
bio_waste_recy_kg_pc	-0,1549	gov_exp_cult_perc	0,0675
airtransp_quality	-0,1532	import_gdp	-0,0574
shadow_ec_gdp	0,1482	fdi_gdp	0,0453
munic_waste_kg_pc	-0,1465	IT_export_perc	0,0351
patents_ce_pmilc	-0,1388	area_organ_farm	-0,0311
roads_quality	-0,136	cap_invest_gdp	-0,0297
lab_prod_perc	0,1299	export_growth	0,0249
munic_waste_recy_rate	-0,1282	cost_start_business	0,0239
pack_waste_recy_rate	-0,1259	ewaste_recy_rate	-0,0139
envir_tax_rev_perc	0,1248	forest_area_perc	-0,0111
prisoners_nbr	0,1193	share_renew_energy	-0,0098
greenhouse_gas	-0,1161	exp_envir_prot_gdp	-0,0055
tourism_rev_pc	-0,1102		

<b>Kintamasis</b>	<b>PK2</b>	<b>Kintamasis</b>	<b>PK2</b>
share_renew_energy	0,2826	exp_onRD_pc	0,1066
forest_area_perc	0,2607	turnover_ppe	-0,0911
ammon_emis_agr	-0,2436	patents_ce_pmille	-0,0898
nutrient_agr_land	-0,2344	start_business_days	-0,0818
area_organ_farm	0,229	gov_exp_cult_perc	0,0781
import_gdp	-0,2146	avg_rooms_pc	-0,0774
tourism_rev_pc	-0,2137	life_exp	-0,0742
export_gdp	-0,2123	infl_cpi	0,0709
active_pop_perc	0,1857	roads_quality	0,0683
gov_spend_gdp	0,1825	pack_waste_recy_rate	0,0649
exp_envir_prot_gdp	-0,1758	bio_waste_recy_kg_pc	0,0631
population_growth_perc	-0,1724	export_growth	0,0614
good_health_share	-0,1628	matt_depr_perc	-0,0545
empl_rate	0,1545	exp_housing_gdp	-0,0544
ewaste_recy_rate	0,1541	pers_science_techn	0,0473
munic_waste_recy_rate	0,147	cap_invest_gdp	0,0453
adult_learn_perc	0,1462	health_spend_pc	0,0416
trade_recycl_materials_t_pc	-0,1455	enterp_nbr_pthc	0,0394
shadow_ec_gdp	-0,1452	prisoners_nbr	0,0325
exp_educ_gdp	0,1409	annual_net_earnings	-0,032
homicides	0,1327	tax_rev_gdp	-0,0313
cost_start_business	-0,1251	app_lab_prod	-0,0279
tertiary_educ_perc	-0,1229	envir_tax_rev_perc	-0,0181
thefts	0,1159	pol_stability_index	-0,0161
munic_waste_kg_pc	-0,1148	airtransp_quality	0,0147
greenhouse_gas	-0,1141	patent_EPO_pmilc	-0,0101
lab_prod_perc	0,1118	IT_export_perc	-0,01
gov_eff_index	0,1118	circular_mater_use_rate	-0,0056
reg_quality_index	0,1107	exp_publ_safety_gdp	-0,001
fdi_gdp	-0,1105	soc_prot_exp_pc	0
corrup_control_index	0,1075		

<b>Kintamasis</b>	<b>PK3</b>	<b>Kintamasis</b>	<b>PK3</b>
gov_exp_cult_perc	0,2951	shadow_ec_gdp	-0,0776
import_gdp	0,2916	bio_waste_recy_kg_pc	-0,0749
export_gdp	0,2845	reg_quality_index	0,0715
IT_export_perc	0,2542	turnover_ppe	0,0649
pol_stability_index	0,2408	exp_educ_gdp	0,0557
empl_rate	0,2373	ewaste_recy_rate	0,0535
infl_cpi	0,2366	exp_publ_safety_gdp	-0,0526
prisoners_nbr	0,2009	exp_onRD_pc	-0,0523
life_exp	-0,1715	annual_net_earnings	-0,0507
homicides	0,1698	gov_eff_index	0,0477
good_health_share	-0,1694	health_spend_pc	-0,0447
lab_prod_perc	0,1692	munic_waste_recy_rate	-0,0433
patents_ce_pmille	0,1673	patent_EPO_pmilc	0,0433
greenhouse_gas	0,1572	forest_area_perc	0,0418
pack_waste_recy_rate	-0,1565	soc_prot_exp_pc	-0,0415
active_pop_perc	0,1559	adult_learn_perc	0,0332
cap_invest_gdp	0,1552	corrup_control_index	0,0274
trade_recycl_materials_t_pc	0,1383	thefts	-0,0263
cost_start_business	-0,1338	ammon_emis_agr	0,0236
roads_quality	-0,1334	fdi_gdp	-0,0199
tourism_rev_pc	0,1308	area_organ_farm	-0,018
export_growth	0,125	envir_tax_rev_perc	-0,0156
gov_spend_gdp	-0,1221	tax_rev_gdp	-0,0148
circular_mater_use_rate	-0,1199	avg_rooms_pc	0,0148
exp_envir_prot_gdp	-0,1044	munic_waste_kg_pc	0,0145
start_business_days	0,1034	tertiary_educ_perc	0,0132
population_growth_perc	0,0991	nutrient_agr_land	-0,0076
enterp_nbr_pthc	0,0945	share_renew_energy	-0,0058
pers_science_techn	0,0922	app_lab_prod	-0,0012
matt_depr_perc	-0,0826	exp_housing_gdp	-0,0002
airtransp_quality	-0,0815		

**10 priedas. ES šalių ekonominis konkurencingumo indeksas 2008–2019 m.**

Šalis\Metai	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Austria	0,416	0,433	0,440	0,444	0,433	0,421	0,428	0,397	0,412	0,411	0,459	0,406
Belgium	0,430	0,450	0,447	0,445	0,425	0,412	0,425	0,374	0,412	0,377	0,396	0,363
Bulgaria	0,196	0,186	0,168	0,176	0,184	0,190	0,177	0,154	0,197	0,180	0,216	0,196
Croatia	0,212	0,208	0,210	0,196	0,188	0,220	0,193	0,194	0,238	0,224	0,257	0,230
Cyprus	0,365	0,436	0,399	0,379	0,347	0,317	0,334	0,301	0,340	0,370	0,360	0,320
Czech Republic	0,340	0,361	0,405	0,394	0,384	0,360	0,396	0,364	0,374	0,399	0,429	0,359
Denmark	0,492	0,502	0,475	0,488	0,468	0,457	0,455	0,432	0,471	0,463	0,503	0,479
Estonia	0,258	0,219	0,328	0,352	0,357	0,322	0,326	0,265	0,341	0,340	0,380	0,364
Finland	0,485	0,428	0,455	0,450	0,428	0,430	0,418	0,391	0,425	0,435	0,425	0,425
France	0,374	0,399	0,378	0,394	0,388	0,376	0,371	0,345	0,353	0,345	0,386	0,315
Germany	0,411	0,399	0,448	0,448	0,419	0,403	0,424	0,386	0,404	0,398	0,423	0,373
Greece	0,239	0,225	0,178	0,140	0,144	0,138	0,156	0,140	0,127	0,174	0,216	0,178
Hungary	0,271	0,240	0,257	0,280	0,227	0,265	0,285	0,250	0,302	0,274	0,302	0,327
Ireland	0,393	0,501	0,413	0,413	0,437	0,427	0,524	0,592	0,542	0,562	0,589	0,548
Italy	0,227	0,227	0,256	0,248	0,225	0,225	0,222	0,205	0,230	0,224	0,237	0,196
Latvia	0,239	0,224	0,211	0,265	0,310	0,246	0,273	0,242	0,258	0,261	0,308	0,265
Lithuania	0,274	0,150	0,231	0,269	0,296	0,270	0,229	0,219	0,252	0,311	0,312	0,313
Luxembourg	0,627	0,658	0,706	0,668	0,664	0,697	0,731	0,640	0,703	0,599	0,613	0,594
Malta	0,447	0,413	0,430	0,386	0,424	0,394	0,367	0,402	0,398	0,400	0,397	0,396
Netherlands	0,463	0,510	0,499	0,490	0,479	0,498	0,492	0,500	0,514	0,507	0,501	0,500
Poland	0,161	0,204	0,206	0,198	0,209	0,201	0,224	0,200	0,250	0,262	0,319	0,282
Portugal	0,251	0,289	0,278	0,260	0,261	0,275	0,257	0,241	0,270	0,283	0,309	0,272
Romania	0,191	0,160	0,162	0,187	0,146	0,217	0,174	0,120	0,199	0,173	0,217	0,174
Slovakia	0,231	0,207	0,336	0,302	0,333	0,302	0,290	0,281	0,302	0,291	0,346	0,278
Slovenia	0,340	0,297	0,326	0,316	0,274	0,294	0,310	0,259	0,291	0,316	0,345	0,302
Spain	0,275	0,308	0,273	0,275	0,262	0,261	0,260	0,234	0,263	0,246	0,272	0,231
Sweden	0,488	0,475	0,525	0,520	0,482	0,478	0,497	0,471	0,476	0,470	0,516	0,474
United Kingdom	0,319	0,322	0,305	0,338	0,346	0,336	0,339	0,334	0,355	0,331	0,342	0,328

## 11 priedas. ES šalių socialinis konkurencingumo indeksas 2008–2019 m.

Šalis\Metai	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Austria	0,534	0,548	0,557	0,555	0,552	0,535	0,567	0,586	0,595	0,575	0,565	0,560
Belgium	0,603	0,613	0,617	0,622	0,602	0,589	0,578	0,579	0,584	0,575	0,560	0,558
Bulgaria	0,343	0,343	0,337	0,331	0,312	0,341	0,352	0,380	0,357	0,365	0,332	0,334
Croatia	0,452	0,444	0,460	0,441	0,435	0,417	0,433	0,432	0,451	0,428	0,420	0,424
Cyprus	0,696	0,697	0,713	0,696	0,646	0,651	0,617	0,613	0,596	0,602	0,589	0,599
Czech Republic	0,440	0,457	0,466	0,469	0,450	0,455	0,447	0,435	0,454	0,438	0,475	0,469
Denmark	0,631	0,645	0,639	0,642	0,629	0,617	0,608	0,620	0,618	0,597	0,576	0,583
Estonia	0,466	0,441	0,435	0,457	0,447	0,439	0,427	0,449	0,489	0,478	0,465	0,476
Finland	0,588	0,599	0,607	0,609	0,605	0,582	0,585	0,588	0,585	0,567	0,552	0,563
France	0,623	0,625	0,631	0,630	0,616	0,639	0,630	0,618	0,627	0,613	0,603	0,597
Germany	0,502	0,528	0,550	0,532	0,554	0,543	0,540	0,544	0,549	0,549	0,536	0,545
Greece	0,437	0,454	0,451	0,442	0,419	0,399	0,389	0,384	0,401	0,397	0,389	0,368
Hungary	0,366	0,360	0,358	0,362	0,332	0,325	0,327	0,344	0,388	0,370	0,391	0,369
Ireland	0,689	0,650	0,649	0,640	0,613	0,595	0,592	0,572	0,587	0,582	0,580	0,575
Italy	0,472	0,479	0,488	0,460	0,452	0,450	0,441	0,420	0,422	0,425	0,412	0,416
Latvia	0,382	0,336	0,346	0,344	0,362	0,371	0,353	0,364	0,388	0,377	0,379	0,343
Lithuania	0,319	0,331	0,297	0,305	0,298	0,283	0,289	0,291	0,316	0,304	0,301	0,315
Luxembourg	0,621	0,672	0,664	0,659	0,673	0,643	0,659	0,639	0,626	0,627	0,627	0,661
Malta	0,465	0,473	0,488	0,492	0,470	0,490	0,515	0,522	0,549	0,555	0,577	0,578
Netherlands	0,617	0,634	0,641	0,639	0,628	0,621	0,610	0,612	0,646	0,623	0,611	0,616
Poland	0,425	0,423	0,440	0,445	0,429	0,414	0,408	0,413	0,430	0,403	0,404	0,386
Portugal	0,492	0,501	0,508	0,516	0,470	0,448	0,436	0,441	0,456	0,445	0,437	0,434
Romania	0,320	0,316	0,355	0,358	0,318	0,314	0,292	0,319	0,310	0,299	0,294	0,296
Slovakia	0,392	0,437	0,444	0,448	0,434	0,425	0,393	0,418	0,427	0,389	0,390	0,383
Slovenia	0,528	0,547	0,549	0,554	0,535	0,518	0,504	0,506	0,525	0,518	0,509	0,516
Spain	0,624	0,623	0,626	0,624	0,588	0,560	0,547	0,550	0,565	0,552	0,541	0,540
Sweden	0,651	0,646	0,650	0,670	0,677	0,673	0,668	0,673	0,683	0,678	0,671	0,683
United Kingdom	0,666	0,685	0,694	0,666	0,646	0,610	0,603	0,592	0,546	0,586	0,548	0,548

## 12 priedas. ES šalių aplinkosauginis konkurencingumo indeksas 2008–2019 m.

Šalis\Metai	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Austria	0,576	0,533	0,566	0,587	0,526	0,533	0,518	0,520	0,503	0,524	0,515	0,523
Belgium	0,462	0,461	0,493	0,511	0,500	0,499	0,491	0,469	0,493	0,490	0,501	0,516
Bulgaria	0,369	0,375	0,401	0,410	0,413	0,434	0,404	0,432	0,438	0,436	0,404	0,422
Croatia	0,393	0,381	0,419	0,395	0,385	0,393	0,371	0,398	0,418	0,439	0,447	0,449
Cyprus	0,232	0,236	0,279	0,329	0,263	0,282	0,285	0,332	0,276	0,312	0,307	0,285
Czech Republic	0,409	0,397	0,444	0,465	0,470	0,432	0,424	0,432	0,433	0,435	0,416	0,426
Denmark	0,432	0,451	0,463	0,421	0,419	0,438	0,429	0,428	0,443	0,429	0,453	0,460
Estonia	0,419	0,439	0,433	0,451	0,480	0,433	0,445	0,420	0,458	0,454	0,465	0,495
Finland	0,465	0,445	0,514	0,507	0,464	0,467	0,448	0,484	0,472	0,466	0,469	0,473
France	0,393	0,380	0,416	0,418	0,396	0,411	0,414	0,414	0,427	0,430	0,433	0,442
Germany	0,467	0,448	0,474	0,481	0,434	0,435	0,418	0,416	0,413	0,402	0,406	0,418
Greece	0,326	0,350	0,367	0,359	0,366	0,400	0,395	0,383	0,396	0,416	0,408	0,424
Hungary	0,320	0,324	0,378	0,374	0,338	0,366	0,376	0,359	0,343	0,350	0,333	0,355
Ireland	0,345	0,328	0,350	0,365	0,336	0,337	0,327	0,303	0,296	0,311	0,299	0,309
Italy	0,402	0,412	0,432	0,448	0,458	0,470	0,468	0,465	0,477	0,487	0,489	0,495
Latvia	0,396	0,398	0,437	0,440	0,451	0,487	0,468	0,477	0,465	0,477	0,474	0,473
Lithuania	0,324	0,355	0,386	0,386	0,381	0,378	0,402	0,371	0,418	0,412	0,410	0,411
Luxembourg	0,519	0,519	0,541	0,524	0,491	0,486	0,479	0,462	0,446	0,460	0,466	0,473
Malta	0,284	0,257	0,260	0,244	0,242	0,236	0,250	0,277	0,210	0,227	0,253	0,263
Netherlands	0,531	0,534	0,538	0,547	0,528	0,524	0,515	0,514	0,523	0,532	0,529	0,552
Poland	0,343	0,327	0,370	0,362	0,338	0,338	0,368	0,393	0,376	0,360	0,365	0,366
Portugal	0,379	0,372	0,395	0,405	0,378	0,396	0,402	0,395	0,400	0,401	0,389	0,399
Romania	0,259	0,292	0,356	0,334	0,346	0,350	0,355	0,352	0,345	0,334	0,332	0,343
Slovakia	0,365	0,392	0,402	0,413	0,422	0,418	0,398	0,402	0,380	0,409	0,401	0,419
Slovenia	0,400	0,398	0,425	0,459	0,478	0,478	0,483	0,484	0,473	0,483	0,469	0,484
Spain	0,420	0,390	0,403	0,401	0,382	0,416	0,402	0,391	0,416	0,423	0,415	0,413
Sweden	0,544	0,539	0,575	0,595	0,558	0,573	0,534	0,528	0,521	0,528	0,529	0,537
United Kingdom	0,399	0,416	0,432	0,423	0,396	0,407	0,396	0,395	0,410	0,410	0,429	0,423

### 13 priedas. ES šalių šalies aplinkos konkurencingumo indeksas 2008–2019 m.

Šalis\Metai	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Austria	0,714	0,665	0,679	0,675	0,699	0,673	0,627	0,636	0,647	0,667	0,680	0,695
Belgium	0,659	0,669	0,667	0,700	0,707	0,706	0,656	0,638	0,600	0,621	0,630	0,635
Bulgaria	0,282	0,295	0,315	0,301	0,339	0,337	0,359	0,321	0,382	0,359	0,357	0,325
Croatia	0,361	0,340	0,384	0,419	0,384	0,352	0,410	0,407	0,422	0,419	0,420	0,429
Cyprus	0,550	0,459	0,474	0,497	0,511	0,484	0,478	0,488	0,474	0,455	0,417	0,501
Czech Republic	0,472	0,461	0,476	0,526	0,465	0,463	0,452	0,457	0,486	0,529	0,548	0,507
Denmark	0,902	0,883	0,885	0,917	0,888	0,888	0,848	0,844	0,852	0,875	0,893	0,891
Estonia	0,532	0,574	0,556	0,544	0,575	0,577	0,644	0,658	0,663	0,604	0,623	0,644
Finland	0,784	0,788	0,802	0,802	0,810	0,805	0,771	0,775	0,768	0,813	0,798	0,817
France	0,698	0,624	0,660	0,685	0,669	0,652	0,620	0,628	0,624	0,669	0,635	0,663
Germany	0,644	0,606	0,623	0,661	0,651	0,627	0,604	0,629	0,655	0,665	0,667	0,660
Greece	0,408	0,360	0,293	0,313	0,336	0,368	0,404	0,412	0,405	0,410	0,443	0,478
Hungary	0,516	0,451	0,449	0,505	0,442	0,474	0,458	0,461	0,455	0,478	0,455	0,418
Ireland	0,717	0,730	0,718	0,687	0,698	0,681	0,697	0,666	0,655	0,689	0,697	0,672
Italy	0,396	0,362	0,375	0,395	0,394	0,383	0,342	0,359	0,389	0,398	0,402	0,420
Latvia	0,395	0,407	0,489	0,438	0,501	0,539	0,488	0,544	0,565	0,534	0,544	0,530
Lithuania	0,431	0,406	0,452	0,421	0,481	0,529	0,535	0,589	0,550	0,497	0,533	0,547
Luxembourg	0,710	0,715	0,726	0,752	0,751	0,733	0,707	0,721	0,748	0,752	0,792	0,792
Malta	0,446	0,417	0,447	0,476	0,493	0,429	0,395	0,359	0,419	0,531	0,550	0,470
Netherlands	0,795	0,768	0,796	0,831	0,822	0,773	0,769	0,770	0,804	0,824	0,830	0,791
Poland	0,338	0,291	0,339	0,346	0,372	0,374	0,361	0,394	0,378	0,336	0,349	0,316
Portugal	0,603	0,571	0,541	0,537	0,551	0,580	0,558	0,568	0,579	0,619	0,624	0,631
Romania	0,262	0,212	0,239	0,255	0,311	0,275	0,288	0,359	0,397	0,351	0,189	0,262
Slovakia	0,493	0,454	0,481	0,480	0,477	0,467	0,496	0,487	0,517	0,520	0,477	0,476
Slovenia	0,576	0,544	0,540	0,602	0,557	0,521	0,534	0,555	0,584	0,574	0,566	0,572
Spain	0,426	0,375	0,388	0,413	0,408	0,401	0,434	0,452	0,475	0,436	0,452	0,470
Sweden	0,840	0,819	0,832	0,850	0,883	0,863	0,825	0,839	0,837	0,858	0,860	0,873
United Kingdom	0,741	0,651	0,710	0,694	0,714	0,692	0,686	0,753	0,725	0,703	0,703	0,736



**14 priedas. ES šalių bendrasis konkurencingumo indeksas 2008–2019 m.**

Šalis\Metai	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Austria	0,560	0,545	0,560	0,565	0,552	0,540	0,535	0,535	0,539	0,544	0,555	0,546
Belgium	0,539	0,548	0,556	0,570	0,558	0,552	0,537	0,515	0,522	0,516	0,522	0,518
Bulgaria	0,297	0,300	0,305	0,304	0,312	0,325	0,323	0,322	0,344	0,335	0,327	0,319
Croatia	0,355	0,343	0,368	0,363	0,348	0,345	0,352	0,358	0,382	0,377	0,386	0,383
Cyprus	0,461	0,457	0,466	0,475	0,442	0,433	0,428	0,434	0,422	0,435	0,418	0,426
Czech Republic	0,416	0,419	0,448	0,464	0,442	0,427	0,430	0,422	0,437	0,450	0,467	0,440
Denmark	0,614	0,620	0,615	0,617	0,601	0,600	0,585	0,581	0,596	0,591	0,606	0,603
Estonia	0,419	0,418	0,438	0,451	0,465	0,443	0,460	0,448	0,488	0,469	0,483	0,495
Finland	0,581	0,565	0,594	0,592	0,577	0,571	0,556	0,559	0,563	0,571	0,561	0,569
France	0,522	0,507	0,521	0,532	0,517	0,520	0,509	0,501	0,508	0,514	0,514	0,504
Germany	0,506	0,495	0,524	0,531	0,515	0,502	0,497	0,494	0,505	0,503	0,508	0,499
Greece	0,353	0,347	0,322	0,314	0,316	0,326	0,336	0,330	0,332	0,349	0,364	0,362
Hungary	0,368	0,344	0,360	0,380	0,335	0,358	0,361	0,353	0,372	0,368	0,371	0,367
Ireland	0,536	0,552	0,533	0,526	0,521	0,510	0,535	0,533	0,520	0,536	0,541	0,526
Italy	0,374	0,370	0,388	0,388	0,382	0,382	0,368	0,362	0,380	0,383	0,385	0,382
Latvia	0,353	0,341	0,371	0,372	0,406	0,411	0,395	0,407	0,419	0,412	0,426	0,403
Lithuania	0,337	0,311	0,342	0,345	0,364	0,365	0,364	0,367	0,384	0,381	0,389	0,397
Luxembourg	0,619	0,641	0,659	0,651	0,645	0,640	0,644	0,615	0,631	0,609	0,624	0,630
Malta	0,411	0,390	0,406	0,400	0,407	0,387	0,382	0,390	0,394	0,428	0,444	0,427
Netherlands	0,602	0,612	0,618	0,627	0,614	0,604	0,597	0,599	0,622	0,622	0,618	0,615
Poland	0,317	0,311	0,339	0,338	0,337	0,332	0,340	0,350	0,358	0,340	0,359	0,337
Portugal	0,431	0,433	0,430	0,430	0,415	0,425	0,413	0,411	0,426	0,437	0,440	0,434
Romania	0,258	0,245	0,278	0,284	0,280	0,289	0,277	0,288	0,313	0,289	0,258	0,269
Slovakia	0,370	0,373	0,416	0,411	0,417	0,403	0,394	0,397	0,406	0,402	0,403	0,389
Slovenia	0,461	0,446	0,460	0,483	0,461	0,453	0,458	0,451	0,468	0,473	0,472	0,469
Spain	0,436	0,424	0,423	0,428	0,410	0,409	0,411	0,407	0,430	0,414	0,420	0,414
Sweden	0,631	0,620	0,646	0,659	0,650	0,647	0,631	0,628	0,629	0,633	0,644	0,642
United Kingdom	0,531	0,518	0,535	0,530	0,526	0,511	0,506	0,519	0,509	0,507	0,505	0,509

## 15 priedas. ES šalių rangai pagal tris rodiklius 2008–2019 m.

15 priede pateikiamose lentelėse stulpelyje „visi“ – pateikiamas ES šalies rangas pagal apskaičiuotą bendrąjį konkurencingumo indeksą, stulpelyje „vsurf“ – ES šalies rangas pagal apskaičiuotą konkurencingumo indeksą, kuriam buvo panaudoti VSURF atrinkti kintamieji, stulpelyje „bvp“ – ES šalies rangas pagal bendrąjį vidaus produktą, tenkantį gyventojui, 1 – geriausias rangas, 28 – blogiausias rangas.

Metai	2008			2009			2010			2011			2012		
	visi	vsurf	bvp	visi	vsurf	bvp	visi	vsurf	bvp	visi	vsurf	bvp	visi	vsurf	bvp
Austria	6	6	7	8	6	5	6	6	6	7	7	6	7	7	6
Belgium	7	10	8	7	8	8	7	8	8	6	6	8	6	6	8
Bulgaria	27	28	28	27	28	28	27	28	28	27	28	28	27	28	28
Croatia	22	25	23	23	26	22	22	25	22	23	26	22	23	26	24
Cyprus	13	13	13	12	16	13	12	16	13	13	16	13	15	13	13
Czech Republic	17	21	18	16	20	19	14	19	19	14	19	19	14	20	19
Denmark	3	2	2	2	2	2	4	2	2	4	2	2	4	2	2
Estonia	16	20	21	17	19	21	15	21	21	15	18	21	12	18	20
Finland	5	4	6	5	5	6	5	5	7	5	5	7	5	4	7
France	10	9	11	10	9	9	11	9	10	8	8	10	10	9	11
Germany	11	7	10	11	10	10	10	7	9	9	9	9	11	8	9
Greece	24	18	15	21	18	15	26	18	15	26	22	15	26	25	17
Hungary	21	19	24	22	21	23	23	22	23	21	20	24	25	19	26
Ireland	8	8	3	6	7	4	9	10	5	11	10	5	9	11	5
Italy	19	15	12	20	15	12	20	15	12	20	15	12	21	16	12
Latvia	23	24	22	24	23	24	21	26	26	22	24	25	20	23	23
Lithuania	25	26	25	26	25	25	24	23	25	24	25	23	22	24	22
Luxembourg	2	1	1	1	1	1	1	1	1	2	1	1	2	1	1
Malta	18	17	19	18	17	18	19	17	18	19	17	18	19	17	16
Netherlands	4	5	4	4	4	3	3	4	4	3	4	4	3	5	4
Poland	26	22	26	25	24	26	25	24	24	25	23	26	24	21	25
Portugal	15	14	17	14	13	17	16	13	17	16	13	17	17	14	18
Romania	28	27	27	28	27	27	28	27	27	28	27	27	28	27	27
Slovakia	20	23	20	19	22	20	18	20	20	18	21	20	16	22	21
Slovenia	12	16	16	13	14	16	13	14	16	12	14	16	13	15	15
Spain	14	12	14	15	12	14	17	12	14	17	12	14	18	12	14
Sweden	1	3	5	3	3	7	2	3	3	1	3	3	1	3	3
United Kingdom	9	11	9	9	11	11	8	11	11	10	11	11	8	10	10

<b>Metai</b>	<b>2013</b>			<b>2014</b>			<b>2015</b>			<b>2016</b>			<b>2017</b>		
<b>Šalis \ Rangai</b>	<b>visi</b>	<b>vsurf</b>	<b>bvp</b>	<b>visi</b>	<b>vsurf</b>	<b>bvp</b>	<b>visi</b>	<b>vsurf</b>	<b>bvp</b>	<b>visi</b>	<b>vsurf</b>	<b>bvp</b>	<b>visi</b>	<b>vsurf</b>	<b>bvp</b>
Austria	7	6	6	7	6	6	6	7	7	6	6	6	6	6	6
Belgium	6	7	8	6	7	9	9	9	10	7	7	9	8	8	9
Bulgaria	27	28	28	27	28	28	27	28	28	26	28	28	27	28	28
Croatia	24	25	26	24	24	26	23	22	26	22	24	26	23	24	26
Cyprus	14	12	14	15	15	14	14	16	15	17	17	15	16	17	15
Czech Republic	15	22	19	14	20	20	15	21	19	14	20	18	14	19	18
Denmark	4	3	2	4	2	2	4	3	3	4	3	3	4	2	3
Estonia	13	18	20	12	18	19	13	18	20	12	18	19	13	18	19
Finland	5	4	7	5	4	7	5	4	8	5	4	7	5	4	7
France	8	9	11	9	10	11	10	10	11	10	10	11	9	10	11
Germany	11	8	9	11	8	8	11	8	9	11	8	8	11	7	8
Greece	26	26	17	26	26	18	26	26	18	27	26	20	25	26	20
Hungary	23	19	25	23	25	25	24	25	24	24	25	24	24	25	24
Ireland	10	11	5	8	11	4	7	11	2	8	11	2	7	11	2
Italy	21	17	12	21	17	12	22	17	12	23	16	12	21	15	12
Latvia	17	20	23	18	21	23	18	20	23	18	22	23	19	20	23
Lithuania	22	21	22	22	19	22	21	19	22	21	19	22	22	21	22
Luxembourg	2	1	1	1	1	1	2	1	1	1	1	1	3	1	1
Malta	20	15	15	20	16	15	20	12	14	20	14	14	17	13	14
Netherlands	3	5	4	3	5	5	3	5	5	3	5	5	2	5	5
Poland	25	24	24	25	23	24	25	24	25	25	21	25	26	22	25
Portugal	16	14	18	16	12	17	16	13	17	16	12	17	15	14	17
Romania	28	27	27	28	27	27	28	27	27	28	27	27	28	27	27
Slovakia	19	23	21	19	22	21	19	23	21	19	23	21	20	23	21
Slovenia	12	16	16	13	13	16	12	14	16	13	13	16	12	12	16
Spain	18	13	13	17	14	13	17	15	13	15	15	13	18	16	13
Sweden	1	2	3	2	3	3	1	2	4	2	2	4	1	3	4
United Kingdom	9	10	10	10	9	10	8	6	6	9	9	10	10	9	10

Metai	2018			2019		
	visi	vsurf	bvp	visi	vsurf	bvp
Austria	6	6	6	6	6	6
Belgium	8	8	9	8	8	9
Bulgaria	27	28	28	27	28	28
Croatia	22	24	26	22	24	26
Cyprus	19	18	15	17	18	15
Czech Republic	14	19	18	14	20	18
Denmark	4	2	3	4	2	3
Estonia	12	17	19	12	14	17
Finland	5	4	7	5	4	7
France	9	10	11	10	10	11
Germany	10	7	8	11	7	8
Greece	25	26	20	25	25	21
Hungary	24	25	24	24	26	24
Ireland	7	11	2	7	11	2
Italy	23	14	12	23	15	12
Latvia	17	23	23	19	22	23
Lithuania	21	21	22	20	19	20
Luxembourg	2	1	1	2	1	1
Malta	15	16	13	16	17	13
Netherlands	3	5	5	3	5	4
Poland	26	22	25	26	23	25
Portugal	16	13	17	15	16	19
Romania	28	27	27	28	27	27
Slovakia	20	20	21	21	21	22
Slovenia	13	12	16	13	12	16
Spain	18	15	14	18	13	14
Sweden	1	3	4	1	3	5
United Kingdom	11	9	10	9	9	10

## 16 priedas. ES šalių konkurencingumo ranginio vertinimo sutapimo lygiai

Metai	Visi-VSURF <sup>1</sup>		Visi-BVP <sup>2</sup>		Visi-VSURF-BVP <sup>3</sup>	
	Kendall'o koef.	p-reikšmė	Kendall'o koef.	p-reikšmė	Kendall'o koef.	p-reikšmė
2008	0,974548	0,002233	0,967433	0,002482	0,964483	7,20E-07
2009	0,983853	0,001942	0,974001	0,002251	0,974336	5,50E-07
2010	0,96196	0,002692	0,953749	0,003038	0,956456	9,10E-07
2011	0,977012	0,002152	0,950739	0,003175	0,956577	9,00E-07
2012	0,960044	0,002769	0,949644	0,003227	0,945509	1,23E-06
2013	0,968254	0,002452	0,944444	0,003481	0,94636	1,20E-06
2014	0,972906	0,002288	0,948823	0,003266	0,952442	1,01E-06
2015	0,967433	0,002482	0,945266	0,00344	0,944779	1,25E-06
2016	0,964423	0,002595	0,944718	0,003467	0,945387	1,23E-06
2017	0,97318	0,002279	0,957854	0,00286	0,956577	9,00E-07
2018	0,96497	0,002574	0,949644	0,003227	0,949401	1,10E-06
2019	0,971538	0,002335	0,955118	0,002978	0,954996	9,40E+00

<sup>1</sup> – konkurencingumo indekso (naudojant visus rodiklius) rangas buvo lyginamas su konkurencingumo indekso (naudojant VSURF atrinktus rodiklius) rangu.

<sup>2</sup> – konkurencingumo indekso (naudojant visus rodiklius) rangas buvo lyginamas su BVP, tenkančiu gyventojui, rangu.

<sup>3</sup> – konkurencingumo indekso (naudojant visus rodiklius) rangas buvo lyginamas su konkurencingumo indekso (naudojant VSURF atrinktus rodiklius) rangu ir su BVP, tenkančiu gyventojui, rangu.