

Raporty statystyczne

OMÓWIENIE RAPORTÓW STATYSTYCZNYCH

Etapy realizacji badań

Problem badawczy



Cel badań

Pytanie (zbiór pytań) na które szukamy odpowiedzi na drodze badań naukowych.

- Określa co chcemy się dowiedzieć.
- Formułowany po przeanalizowaniu literatury na dany temat (Co jeszcze nie zostało zrobione?)

Polega na uświadomieniu sobie, po co podejmujemy badania i do czego mogą być przydatne uzyskane w nich wyniki.

- Określenie
- Stworzenie
- Wyjaśnienie

Etapy realizacji badań - przykład

Problem
badawczy



Cel

Zmiany segregacji rasowej w Chicago w latach 1910-2010.

Jak zmieniał się poziom segregacji rasowej w latach 1910-2010?

Jaki trend odnotowano? Czy poziom segregacji rośnie czy maleje na przestrzeni lat?

- Scharakteryzowanie struktury rasowo-etnicznej miasta Chicago w latach 1910-2010
- Określenie poziomu segregacji rasowej w latach 1910-2010

Etapy realizacji badań

Dane

Jakie dane będą potrzebne do rozwiązania problemu badawczego?



Metody

Jakie metody trzeba zastosować, aby zrealizować założone cele?

Dane

- Jakie dane będą potrzebne do rozwiązania problemu badawczego?
- W jakiej formie dane są dostępne?
- Czy dane, którymi dysponujemy są porównywalne między latami, obszarami?
- Czy mogę użyć wartości bezwzględnych czy muszą zostać użyte wartości względne?
 - Długość sieci kolejowej w krajach Europy czy gęstość sieci kolejowej
 - Liczba przestępstw czy liczba przestępstw na 1000 (100 000) mieszkańców
 - Liczba zgonów czy wskaźnik śmiertelności

Metody badawcze

- Jakie metody zostaną użyte do rozwiązania problemu badawczego?
- Czy zastosowano wystarczające metody do rozwiązania problemu?
- W jakie formie najlepiej przedstawić wyniki?
 - Tabele
 - Mapy
 - Wykresy

Etapy realizacji badań

Wyniki

Wyniki wraz z ich krótkim opisem (bez interpretacji)



Podsumowanie

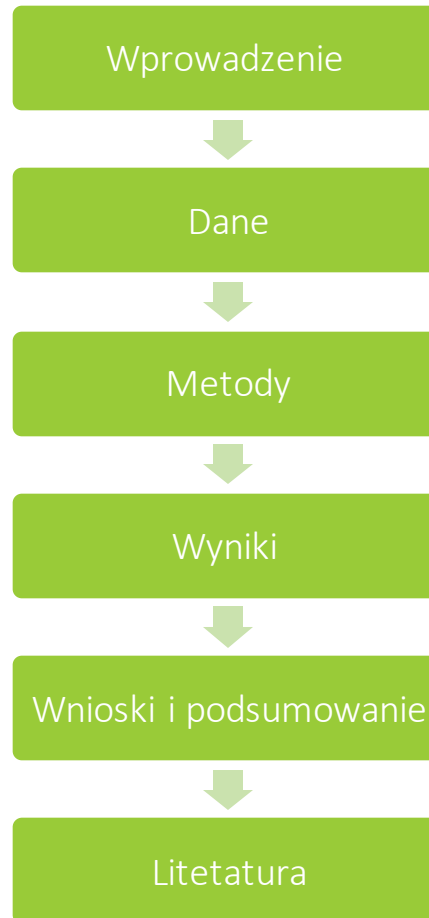
Zawiera odpowiedź na pytania postawione w problemie badawczym

Co wynika z przedstawionych wyników?

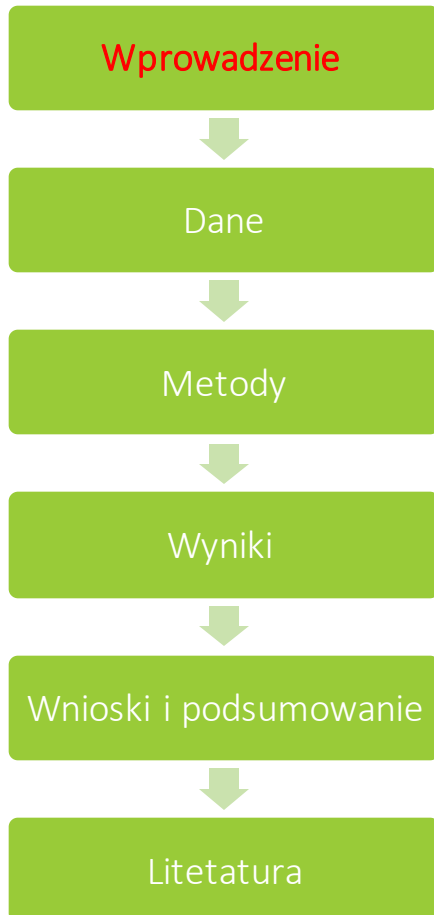
Wyniki

- W jakiej formie przedstawiono wyniki? (tabele, wykresy, tekst, mapy)
- Czy tabele/mapy/wykresy nie przedstawiają tej samej informacji?
- Czy każdy z celów analizy znajduje swoje odzwierciedlenie w wynikach?
- Czy została wybrana najlepsza możliwa forma prezentacji wyników? (tabela, odpowiedni wykres, itd.)
- Czy do każdego wykresu, tabeli, mapy jest odwołanie w tekście.
 - Każda tabela/wykres/mapa musi być podpisana.
 - Mapy i wykresy podpisujemy na dole jako Rycina 1. Opis. W tekście odwołujemy się do nich jako ryc.1 (Ryc.1 przedstawia albo Zróżnicowanie struktury rasowej (ryc.1) wskazuje...)
 - Tabele podpisujemy powyżej (nad tabelą) jako Tabela 1, a w tekście odwołujemy się do niej jako tabela 1 lub tab.1

Struktura raportu

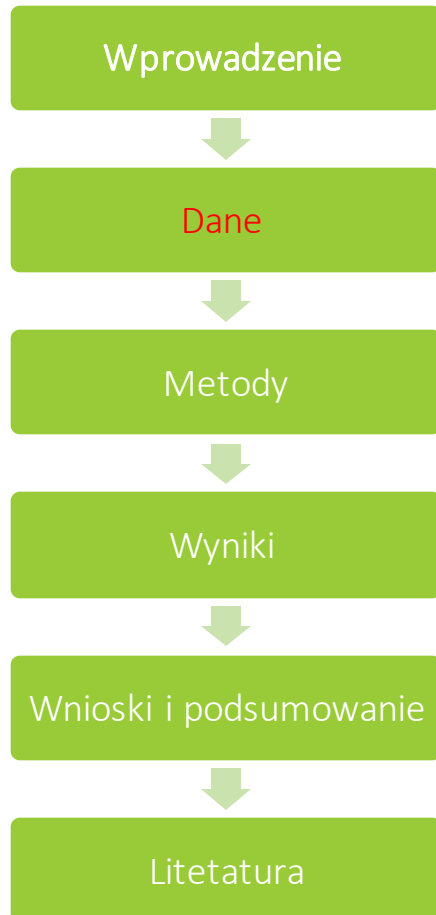


Struktura raportu - Wprowadzenie



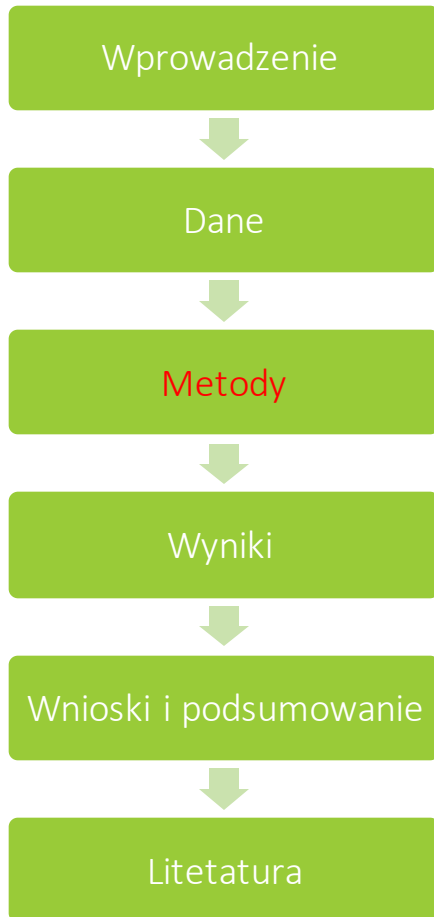
- stanowi wprowadzenie do problematyki
- Powinno nakreślać co w danym temacie zostało zrobione
- Powinno zawierać odwołania do literatury
- powinien być zakończony sformułowaniem problemu badawczego oraz postawieniem celu badań

Struktura raportu - Dane



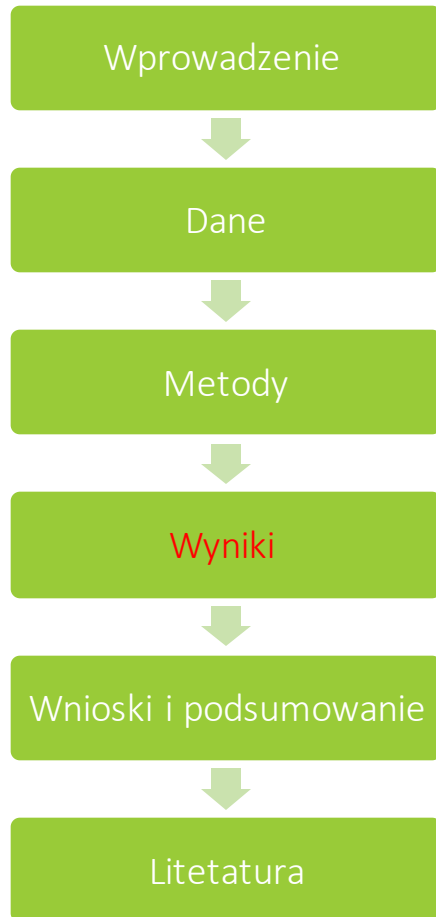
- informacje o wykorzystanym zbiorze danych (co mierzono? kiedy? gdzie? po co? jak?)
- Źródła danych
- Sekcja Dane powinna zawierać szczegółowe informacje na temat zbioru danych – Tak aby na ich podstawie można było odtworzyć analizę.
- Nie zamieszczamy tutaj opisu czym jest EUROSTAT
- Podajemy jak nazywał się zbiór danych (np. Tytuł tabeli)
- Co zawiera zbiór danych

Struktura raportu - Metody



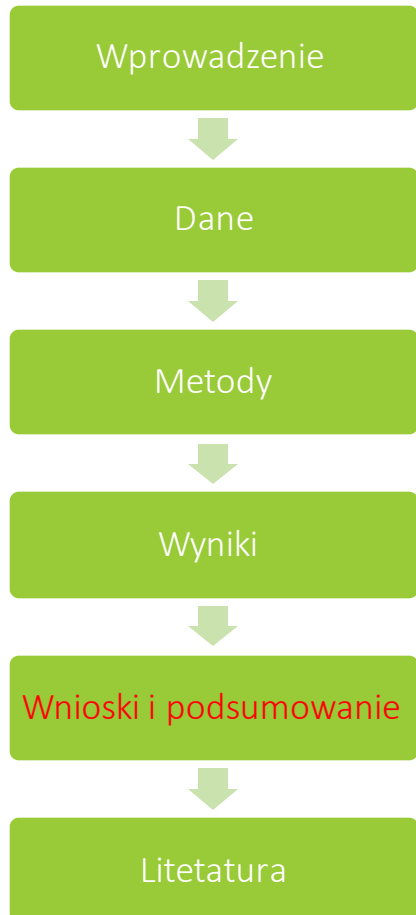
- opis wykorzystanych metod do eksploracji danych
- opis wykorzystanych metod do analizy danych
- określenie wykorzystanego oprogramowania

Struktura raportu - Wyniki



- Co odkryliśmy w wyniku analizy danych?
- Wyniki eksploracji danych
- Wyniki zastosowanych testów
- Ta część zawiera zestawienia tabelaryczne, wykresy.
- Ta część **nie zawiera** analizy i interpretacji wyników.
- Tabele, wykresy, mapy muszą być opisane
- Musimy się odwołać w tekście do zamieszczonych tabel, wykresów.
- Sekcja ta **musi zawierać** opis wyników.
 - Np. Poziom segregacji rasowej wynosił między ... a , średnio. ...
 - Tabela 1 przedstawia zestawienie statystyk. Możemy zauważyć, że...

Struktura raportu – wnioski i podsumowanie



- Podsumowanie analizy
- Odniesienie się do postawionego celu
- Stwierdzenie braków czy problemów w analizie
- **Czy odpowiedzieliśmy na pytanie postawione w problemie badawczym?**

Struktura raportu - Literatura

Wprowadzenie



Dane



Metody



Wyniki



Wnioski i podsumowanie



Litetatura

Spis literatury, do której odwołano się w tekście raportu umieszczony w kolejności alfabetycznej.

W tekście:

Analizę wykonano wykorzystując R (**R Core Team, 2020**).

Odwołując się do oprogramowania R w sekcji Literatura podajemy:

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL

<https://www.R-project.org/>.

Błędy jakich
należy unikać

Projekt Raport Statystyczny - na podstawie danych
zawartych na stronie Eurostat



Nieprawidłowy tytuł

**Tytuł opracowania musi nawiązywać do
problemu badawczego oraz zawartości raportu**

Tytuł
opracowania

Regionalna statystyka turystyczna

WprowadzenieDaneWynikiPodsumowanie

Wprowadzenie

Turystyka to jedna z najważniejszych usług gospodarczych w naszym kraju. Nasze plany wyjazdowe zaczynają się od wyboru noclegu. Co roku każdy chce wypocząć na polskim morzem lub też w naszych polskich górach. Na długie kilkuniedniowe wyjazdy potrzebujemy miejsca od spania czy też relaksu jaki oferują miejsca noclegowe

Prowadzi to do rozwoju branży hotelarskiej, czyli powstawania nowych hoteli, hosteli, burs oraz innych miejsc noclegowych.

Nasz raport przedstawia dane dotyczące liczby łóżek oraz miejsc noclegowych w całej Polsce w 2019 roku. Za pomocą uzyskanych danych postaramy się sprawdzić jaki udział procentowy ma każde województwo.

Drugim aspektem „na którym warto się skupić jest liczba łóżek w danych województwach to pozwoli na określenie największych kompleksów w całej Polsce. Dane dotyczące ilości łóżek również pokażemy w udziale [%] w całym naszym kraju.

Jak można zauważyć na początkowej stronie znajduje się kod do pobrania danych użytych w projekcie. Kod ten został przefiltrowany do obszaru Polski w 2019 roku.

Cel badania

- Uzyskanie informacji na temat procentowego udziału miejsc noclegowych w każdym województwie
- Uzyskanie informacji na temat procentowego udziału łóżek w każdym województwie
- Zbadanie proporcji ilości łóżek do miejsc noclegowych
- Teoretyczne określenie najbardziej odwiedzanych miejsc w Polsce
- Teoretyczne określenie największych kompleksów w Polsce

Kod do pobrania oraz wyświetlenia danych użytych w projekcie

```
id = search_eurostat(pattern = "Number of establishments and bed-places",
                     type = "table")$code[1]
# dane dot. liczba zakładów i miejsc noclegowych wg. regionów
print(id)
#pobranie danych
dane = get_eurostat(id, time_format = "num", type = "code")
#struktura danych
str(dane)
#nazwy tabel
table_eurostat_vars(names(dane))
# dodanie skrótów
dane$cntry = substr(dane$geo, 1,2)
# dodanie nazwy państw
dane$countryname = countrycode(dane$cntry, "iso2c", "country.name")
str(dane)
# estymacja danych
pl2019 = filter(dane, cntry == "PL", time == 2019)
pl2019 <- droplevels(pl2019)
pl2019 = mutate(pl2019, geo = label_eurostat(pl2019)[["geo"]])
pl_2019 = group_by(pl2019, geo)
# zmiana nazw kolumn
colnames(pl2019) = c("Rodzaj zakwaterowania", "Jednostka miary",
                    "Klasyfikacja działalności", "Województwo",
                    "Rok", "wartosc", "cntry", "countryname")
```

Forma prezentacji wyników

Nieodpowiedni wybór formy prezentacji wyników dla raportu pisemnego.

2 Dane

Analizę przeprowadzono w oparciu o dane pochodzące z Eurostatu (<https://ec.europa.eu/eurostat>), czyli z Europejskiego Urzędu Statystycznego. Pochodzą z pakietu o nazwie "**Death due to alcoholic abuse, by sex**", a kod do tabeli danych to "**tps00140**". Po przefiltrowaniu wartości poddawanych analizie w tabeli znalazły się cztery kolumny opisujące: płeć, kraj, rok i wartość wskaźnika standaryzowanego.



**Przykład prawidłowego odwołania do danych
pozwalający na powtórzenie analiz**

Dane

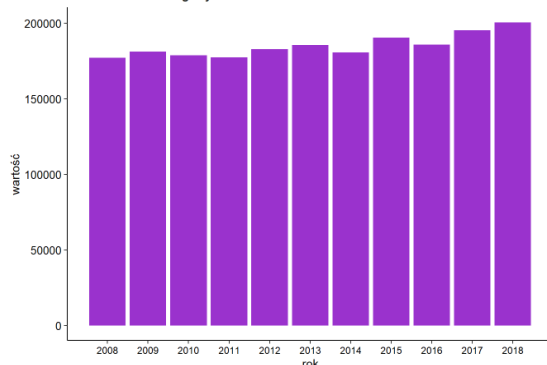
Które z poszczególnych państw europejskich miało najdłuższą sieć kolejową w 2017 roku?



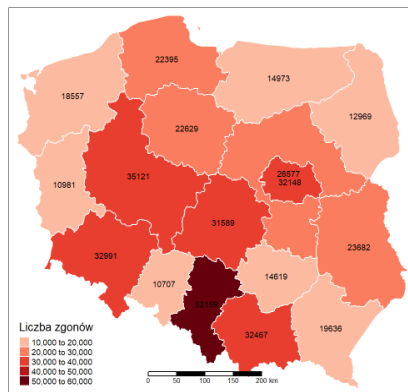
Długość czy gęstość sieci kolejowej?

Wartości bezwzględne a wartości względne

Zgony kobiet w Polsce w latach 2008-2018



Liczba czy wskaźnik śmiertelności (liczba zgonów/liczba ludności)?



Niepoprawny nagłówek

5 pobranie i przetworzenie danych z EUROSTAT i tabela

Code

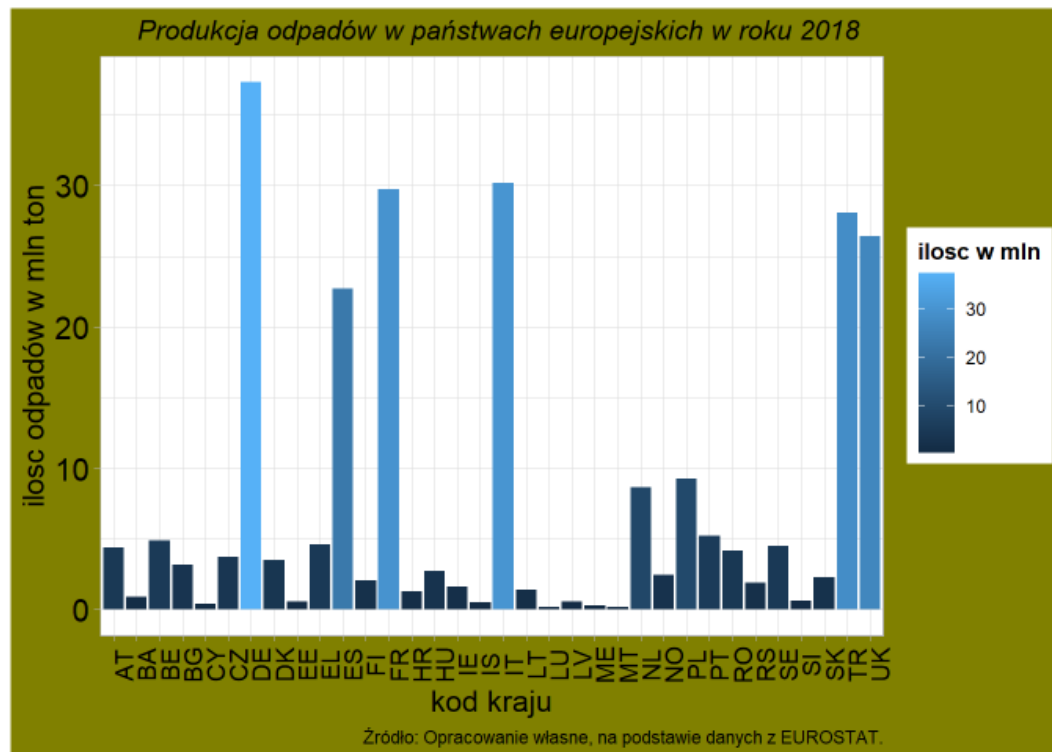
Dane o odpadach w roku 2018

geo	time	ilosc_w_mln
AT	2018	4.407376
BA	2018	0.920541
BE	2018	4.885123
BG	2018	3.145709
CY	2018	0.386668
CZ	2018	3.724150
DE	2018	37.308892
DK	2018	3.499512

Nagłówki sekcji

*Nagłówki poszczególnych części raportu to Wprowadzenie,
Dane i metody, Wyniki, Wnioski i podsumowanie*

7 wykres produkcji odpadów



- Nie poprawny nagłówek sekcji.
- Czy tło pod wykresem w raporcie jest potrzebne?
- Skróty nazw powinny być wyjaśnione w opisie pod ryciną lub w tekście raportu,
- Przesunięte etykiety w opisie osi x

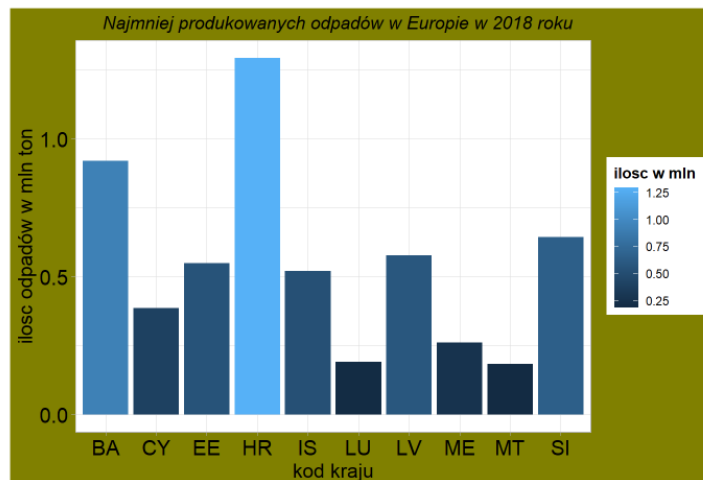
9 najmniej odpadów w Europie w roku 2018 (pierwsze 10 krajów)

Code

10 krajów produkujących najmniej odpadów w Europie w 2018 roku

geo	time	ilosc_w mln
MT	2018	0.183880
LU	2018	0.190568
ME	2018	0.261154
CY	2018	0.386668
IS	2018	0.519927
EE	2018	0.548242
LV	2018	0.577587
SI	2018	0.643342
BA	2018	0.920541
HR	2018	1.293535

Code



Informacje w
tabelach i na
wykresach

Tabela i wykres nie powinny przedstawiać tych
samyh informacji

Ryc. 1 Procentowy udział miejsc noclegowych w każdym województwie.

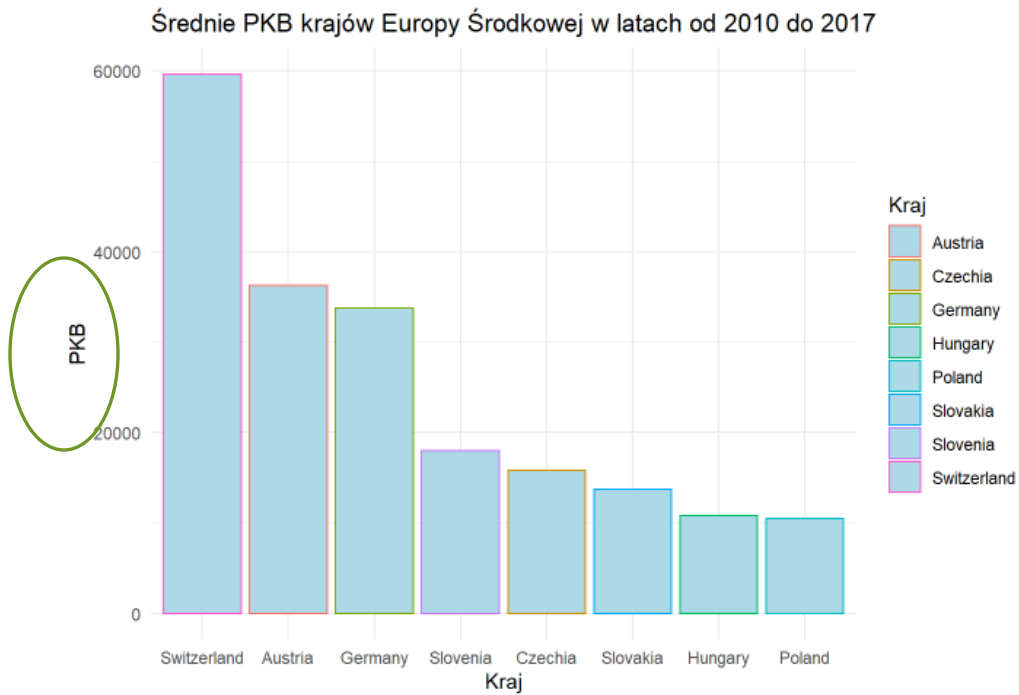


Podpisy rycin

Ilościowy udział czego?

Niekompletny opis osi.

Podpis ryciny nie zgodny z tym co wykres przedstawia

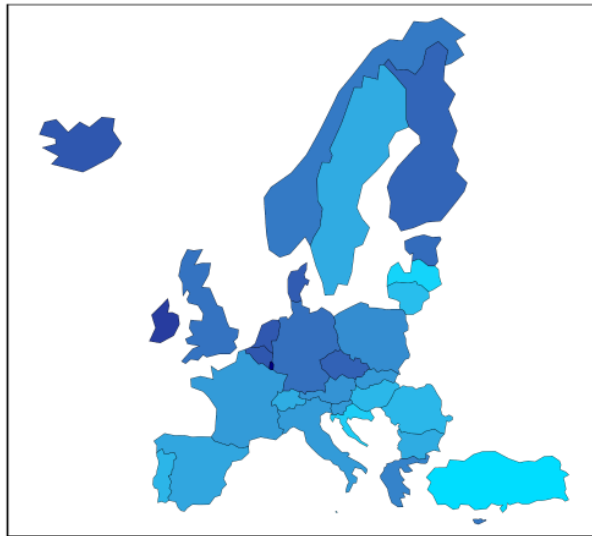


Ryc. 2. Rozkład wartości średniego PKB na jednego mieszkańca dla państw Europy Środkowej w latach 2010-2017.

Podpisy rycin

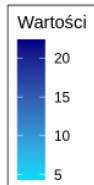
Podpis ryciny nie zgodny z tym co wykres przedstawia

Wartość GHG per capita w krajach Europy w roku 1998



Źródło: Opracowanie własne na podstawie danych z Eurostatu

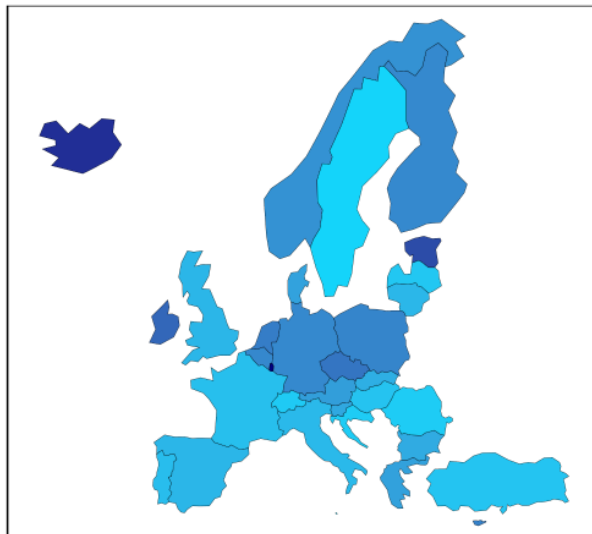
Ryc.6 Mapa uwalniania GHG w 1998 w krajach europejskich



Ten sam kolor może
oznaczać różne
wartości.

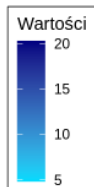
Obie mapy muszą
mieć te same
przedziały

Wartość GHG per capita w krajach Europy w roku 2018



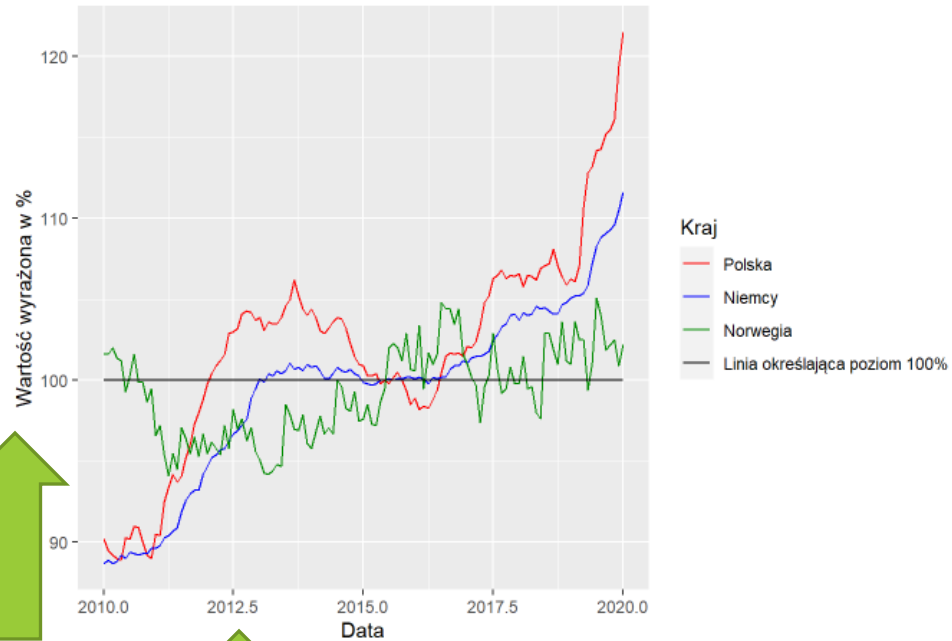
Źródło: Opracowanie własne na podstawie danych z Eurostatu

Ryc.7 Mapa uwalniania GHG w 2018 w krajach europejskich



Porównywalność
map

Cena mięsa w latach 2010 do 2020 w Polsce, Norwegii i Niemczech



Opisy osi
wykresów

Podstawowe statystyki

Średnia ilość wszystkich popełnionych przestępstw w krajach europejskich w 2007 roku.

[1] 912380.9

Średnia ilość kradzieży zmotoryzowanych pojazdów lądowych w krajach europejskich w 2007 roku.

[1] 38735.59

Średnia ilość włamań do prywatnych lokali mieszkalnych w Europie w 2007 roku.

[1] 41752.68

Średnia liczba celowych zabójstw popełnionych w krajach europejskich w 2007 roku.

[1] 284.0769

Średnia ilość napadów rabunkowych w państwach europejskich w 2007 roku.

[1] 17966.64



**Statystyki umieszczamy w
tabeli lub w tekście
raportu**

Format
wyników

4 Wyniki eksploracji danych

4.1 Rycina 1

Zestawiając liczbę pasażerów, korzystających z lotnisk w danych krajach Europy, nie sposób nie zauważyć ewidentnej tendencji zniżkowej w czerwcu 2020 względem czerwca 2019.

Praktycznie każdy kraj miał bardzo duże spadki popularności lotów pasażerskich, co jest widoczne na ryc.1.

Są tylko trzy kraje, które zanotowały ubytki pasażerów transportu lotniczego mniejsze od 90% względem czerwca 2019 i jest to Norwegia (85.27%), Serbia (87.57%) i Bułgaria (89.22%). Są również trzy kraje, które zanotowały spadek liczby pasażerów o ponad 99% i jest to Macedonia (99.98%), Malta (99.52%) i Słowacja (99.11%).

Nie ma jednak żadnego kraju, gdzie w czerwcu 2020 nie odbyłby się przynajmniej jeden lot pasażerski.

Code

Liczba pasażerów, obsłużonych przez porty lotnicze poszczególnych krajów Europy w czerwcu 2019 i czerwcu 2020.

Kod kraju	Liczba Pasażerów czerwiec 2020	Liczba Pasażerów czerwiec 2019	Bezwzględny spadek liczby pasażerów	Względny spadek liczby pasażerów w %
AT	141631	3314289	-3172658	95.73
BE	109685	3257996	-3148311	96.63
BG	156857	1455157	-1298300	89.22
CH	276354	5160769	-4884415	94.65
CY	40640	1293838	-1253198	96.86
CZ	37395	1925827	-1888432	98.06
DE	1323571	21595042	-20271471	93.87
DK	171539	3401375	-3229836	94.96
EE	24788	316964	-292176	92.18
EL	388740	7632190	-7243450	94.91
ES	663360	22809002	-22145642	97.09
FI	84496	2142627	-2058131	96.06
FR	896556	15951619	-15055063	94.38

Czy tabela z danymi jest potrzebna w raporcie?

Nagłówki sekcji

Tabela nie jest ryciną.

Rycina jest częścią raportu, nie stanowi osobnej sekcji

Wyniki = Minimalne, srednie i maksymalne wartosci dla poszczególnych krajów w przedziałach czasowych:

W formie tabeli



```
by_country_1 = group_by(dane_eu2008_2010, countryname)
stat_by_country_1 = summarise(by_country_1,
  srednia = mean(values),
  min = min(values),
  max = max(values))

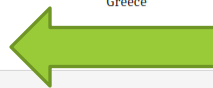
colnames(stat_by_country_1) = c("Nazwa kraju", "Średnia", "Minimum", "Maksimum")

table_1 = kbl(stat_by_country_1, digits = 2, caption = "Średnia wartość emisji gazów cieplarnianych dla przedziału czasowego 2008 - 2010")
kable_classic(table_1, full_width = F, html_font = "Cambria")
```

Średnia wartość emisji gazów cieplarnianych dla przedziału czasowego 2008 - 2010

Nazwa kraju	Średnia	Minimum	Maksimum
Austria	251.22	244.47	259.86
Belgium	330.08	318.40	351.90
Bulgaria	1641.60	1515.38	1848.51
Croatia	585.30	573.51	606.17
Cyprus	506.73	473.20	532.45
Czechia	843.96	827.78	855.97
Denmark	461.80	435.58	482.14
Estonia	1414.29	1315.60	1598.99
Finland	405.64	387.96	431.41
France	216.28	211.61	222.69
Germany	355.49	347.10	367.60
Greece	494.47	480.22	512.01

W formie wykresu

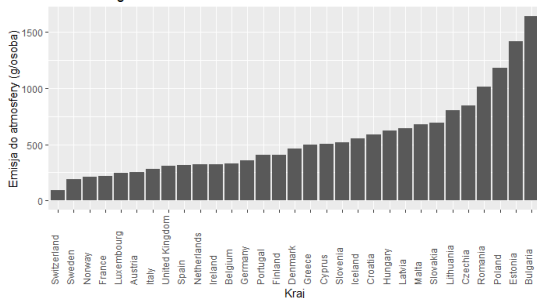


```
#dane
wykres_1 = dane_eu2008_2010 %>% group_by(countryname) %>% summarise(srednia = mean(values))
wykres_2 = dane_eu2011_2014 %>% group_by(countryname) %>% summarise(srednia = mean(values))
wykres_3 = dane_eu2015_2018 %>% group_by(countryname) %>% summarise(srednia = mean(values))

#wykresy

ggplot(data = wykres_1, aes(x = reorder(countryname, srednia), srednia)) +
  geom_col() +
  labs(title = "Średnia wartość emisji gazów cieplarnianych w 31 krajach europejskich w gramach na osobę w latach 2008 - 2010", x = "Kraj", y = "Emisja do atmosfery (g/osoba)") + theme(axis.text.x = element_text(angle = 90, vjust = 0, hjust = 0))
```

Średnia wartość emisji gazów cieplarnianych w 31 krajach europejskich w gramach na osobę w latach 2008 - 2010



Format
prezentacji
wyników

Kod w
raporcie