

Rzeszów 01.06.2022

Języki Obce Nowożytne - Liczba Uczniów w Roku Szkolnym 2018/2019 Według Szkół Statystyczna Analiza Danych

Złotek Sebastian Inżynieria i Analiza Danych





Spis treści

1. vvstęp	
2. Dane do analizy	
3. Obróbka danych	
4. Parametry opisowe	
5. Graficzna prezentacja danych	10
6. Hipotezy statystyczne	22
1. Hipoteza pierwsza	22
2. Hipoteza druga	23
7. Wnioski	24
8. Kod źródłowy	25
9. Źródła	3(





1. Wstęp

W szkołach w Polsce wymagane jest uczenie się najpierw jednego, później dwóch (lub więcej) języków obcych, z racji tego każdy uczeń w większym, lub mniejszym stopniu ma niejaką wiedzę i znajomość danego języka(najczęściej angielskiego lub niemieckiego). Nauczanie języków obcych jest coraz ważniejszym elementem edukacji w systemie kształcenia. W roku szkolnym 2005/2006 w szkołach dla dzieci i młodzieży - podstawowych, gimnazjach, szkołach ponadgimnazjalnych i policealnych realizowana jest nauka następujących języków obcych: angielskiego, francuskiego, niemieckiego, rosyjskiego oraz w mniejszym stopniu innych, m.in. łaciny, hiszpańskiego i włoskiego.

Polscy uczniowie znajdują się w europejskiej czołówce pod względem liczby języków obcych uczonych w szkołach. Nie przekłada się to jednak na ich znajomość

W polskich szkołach statystyczny uczeń uczy się niemal dwóch języków obcych (współczynnik 1,7). To ósmy wynik w Unii Europejskiej. Najwięcej uczniów - jak wynika z danych Eurostatu za 2017 r. - uczy się języka angielskiego. To od 90,4 proc. (szkoły zawodowe) do 98,2 proc. (dawnym gimnazjum) uczniów.

Patrząc na dane Eurostatu, można także dostrzec, że język niemiecki staje się powszechnie obowiązującym dopiero na poziomie dawnego gimnazjum (uczniowie w wieku 12-14 lat wg. Eurostatu). Wówczas wzrasta także odsetek uczniów chodzących na lekcje jęz. rosyjskiego - 6,4 proc., hiszpańskiego - 3,7 proc. oraz francuskiego - 3,3 proc., by na poziomie liceum ogólnokształcącego osiągnąć odpowiednio 8,2, 7,2 i 8,6 proc.

Posługując się danymi statystycznymi, postaram się sprawdzić jak te dane wyglądały w latach 2018-2019.

2. Dane do analizy

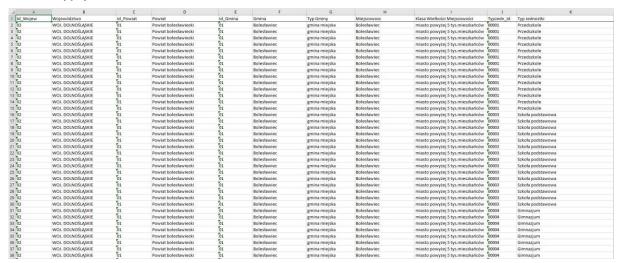
Dane na temat języków obcych uczonych w szkołach w latach 2018/2019 dotyczą uczniów, uczących się danego języka w danym województwie, dalej powiecie i gminie, oraz rozdzielają wszystko na dane o miastach powyżej 5tys. mieszkańców i dane o uczniach na wsi. Uwzględniają one także rodzaj szkoły (publiczna i niepubliczna) oraz posiadają wszystkie ważne dane tych placówek. W analizie będę uwzględniał wszystkie 83170 wierszy dostępnych w pliku .xlsx (Excel).





Dane zostały pobrane ze strony dane.gov.pl, są to otwarte dane publiczne instytucji, urzędów, z których każdy może korzystać. Na bazie otwartych danych publicznych w Europie i na świecie powstaje coraz więcej nowoczesnych produktów i usług. Otwarte dane są źródłem realnych oszczędności pieniędzy i czasu dla administracji i obywateli. Obywatele, w tym przedsiębiorcy, mogą korzystać z zasobów danych publicznych realizując własne cele, rozwijając swoją działalność gospodarczą lub badania.

Głównym przedmiotem badań w ramce będą języki dodatkowe i obowiązkowe, które wybierają szkoły oraz dzieci w danych województwach i niższych jednostkach administracyjnych.



Rysunek 1 Dane uruchomione w MS Excel po pobraniu ze strony dane.gov.pl

3. Obróbka danych

Do obróbki danych użyłem graficznego środowiska R Studio, dzięki któremu możemy w szybki i prosty sposób obrabiać dane używając oczywiście wymaganego w projekcie języka R.

Do samej obróbki danych potrzebny jest jedynie pakiet "readxl", dzięki któremu w języku R możemy uruchomić plik z programu MS EXCEL. Potrzebujemy także polecenia, dzięki któremu możemy ustawić główny katalog z danymi:

- setwd("ścieżka pliku")
- install.packages("readxl") instalacja paczki do wczytania plików xls,xlsx
- library("readxl") uruchomienie paczki do wczytania plików xls,xlsx





Dalej uruchamiamy samą ramkę danych poniższym poleceniem i wczytujemy ją do zmiennej "jezyki" :

 read_excel("Jezyki_obce_nowozytne_liczba_uczniow_rokszkolny20182019_w g_szkol.xlsx")

*	‡ Id_Wojew	‡ Województwo	‡ Id_Powiat	\$ Powiat	‡ Id_Gmina	\$ Gmina	‡ Typ Gminy	† Miejscowosc	‡ Klasa Wielkości Miejscowości	‡ TypJedn_ld
1	02	WOJ. DOLNOŚLĄSKIE	01	Powiat bolesławiecki	01	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
2	02	WOJ. DOLNOŚLĄSKIE	01	Powiat bolesławiecki	01	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
3	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
4	02	WOJ. DOLNOŚLĄSKIE	01	Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
5	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
6	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
7		WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
8	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
9	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
10	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
11	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
12	02	WOJ. DOLNOŚLĄSKIE	01	Powiat bolesławiecki	01	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
13	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
14	02	WOJ. DOLNOŚLĄSKIE	01	Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00001
15	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00003
16	02	WOJ. DOLNOŚLĄSKIE	01	Powiat bolesławiecki	01	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00003
17		WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00003
18	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00003
19		WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00003
20	02	WOJ. DOLNOŚLĄSKIE		Powiat bolesławiecki		Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	00003

Rysunek 2 Ramka danych po wczytaniu pliku xlsx

Następnie przechodzimy do właściwej obróbki danych, gdzie po uruchomieniu i wczytaniu pliku w środowisku R Studio otrzymujemy kilka niepotrzebnych do analizy kolumn takich jak:

- Id_Wojew
- Id Powiat
- Id Gmina

I FIZYKI STOSOWANEJ

- Późniejsze kolumny od 10 do 26 zawierające adresy, numery telefonów numery jednostki czy nazwy placówki, które nie są potrzebne do analizy Używamy do tego polecenia:
- jezyki <- jezyki[, -c(1,3,5,10:26)], gdzie w funkcji c wybieram kolumny do usunięcia wymienione powyżej.

Na sam koniec obróbki danych podzieliłem dane z ramki na dwa typy, dane z miast i ze wsi. Dzięki temu zabiegowi w łatwy sposób będę mógł porównywać ilu uczniów względem miast i wsi w Polsce uczy się danych języków i jak



duża jest przewaga miast, jeśli chodzi o dzieci i szkoły wybierające dane języki dodatkowe i obowiązkowe. Przewaga miast przyjmuję, że wynika z oczywistej większej populacji obiektów miejskich nad wsiami. Do podzielenia ramki użyłem danych funkcji:

- miasto <- jezyki[jezyki\$`Klasa Wielkości Miejscowości`=="miasto powyżej 5 tys.mieszkańców",]
- wies <-jezyki[jezyki\$`Klasa Wielkości Miejscowości`=="wieś",]

W obu przypadkach wybieram z kolumny "Klasa Wielkości Miejscowości" dany typ aglomeracji i przypisuje go do nowej ramki danych.

Otrzymuję dzięki temu ramki:





*	÷ Województwo	\$ Powiat	‡ Gmina	‡ Typ Gminy	‡ Miejscowosc	‡ Klasa Wielkości Miejscowości	‡ Jezyk Obcy	Liczba [‡] Uczniów Język Obowiązkowy	Liczba [‡] Uczniów Język Dodatkowy
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	83	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	183	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	25	54
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski		94
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski		62
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski		67
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	68	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	125	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	niemiecki	122	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	125	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	niemiecki	125	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski		
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	99	
14	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	134	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	268	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	16	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	206	
18	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	niemiecki	127	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	608	
20	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	niemiecki	166	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina miejska	Bolesławiec	miasto powyżej 5 tys.mieszkańców	angielski	397	

Rysunek 3 Ramka danych miasto utworzona do analizy z ramki jezyki

•	* Województwo	÷ Powiat	\$ Gmina	‡ Typ Gminy	† Miejscowosc	‡ Klasa Wielkości Miejscowości	‡ Jezyk Obcy	Liczba [‡] Uczniów Język Obowiązkowy	Liczba ‡ Uczniów Język Dodatkowy
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Bożejowice	wieś	angielski	48	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Bożejowice	wieś	angielski	97	18
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Bożejowice	wieś	niemiecki	35	80
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Brzeźnik	wieś	angielski		40
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Brzeźnik	wieś	angielski	107	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Brzeźnik	wieś	niemiecki	25	82
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Dąbrowa Bolesławiecka	wieś	angielski	35	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Dąbrowa Bolesławiecka	wieś	angielski	32	25
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Dąbrowa Bolesławiecka	wieś	niemiecki	37	20
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Kraśnik Dolny	wieś	angielski	50	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Kraśnik Dolny	wieś	angielski	47	89
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Kraśnik Dolny	wieś	niemiecki	120	18
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Kruszyn	wieś	angielski	74	
14	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Kruszyn	wieś	angielski		182
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Kruszyn	wieś	niemiecki	182	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Łaziska	wieś	angielski		67
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Ocice	wieś	angielski		20
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Ocice	wieś	angielski		100
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Ocice	wieś	niemiecki	101	
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Trzebień	wieś	angielski		45
	WOJ. DOLNOŚLĄSKIE	Powiat bolesławiecki	Bolesławiec	gmina wiejska	Trzebień	wieś	angielski	103	

Rysunek 4 Ramka danych miasto utworzona do analizy z ramki jezyki

4. Parametry opisowe





Wszystkie użyte parametry opisowe bazują na dwóch kolumnach, jakimi są "Liczba uczniów język obowiązkowy" oraz "Liczba uczniów język dodatkowy", które są widoczne na kolejno rysunku 3 i rysunku 4.

Do obliczenia parametrów opisowych potrzebna jest jedna biblioteka:

- install.packages("moments") instalacja paczki do miary koncentracji i skośności
- library("moments") wywołanie paczki do miary koncentracji i skośności
 Odchylenia standardowe:

```
> #odchylenia standardowe dla miast i wsi według języka obowiazkowego
> sdmiasto <- sd(miasto$`Liczba Uczniów Język Obowiązkowy`)
> sdmiasto
[1] 162.0149
> sdwies <- sd(wies$`Liczba Uczniów Język Obowiązkowy`)
> sdwies
[1] 75.43071
>
> #odchylenia standardowe dla miast i wsi według języka dodatkowego
> sdmiastododatkowy <- sd(miasto$`Liczba Uczniów Język Dodatkowy`)
> sdmiastododatkowy
[1] 37.25847
> sdwiesdodatkowy <- sd(wies$`Liczba Uczniów Język Dodatkowy`)
> sdwiesdodatkowy
[1] 23.25572
```

Rysunek 5 Odchylenia standardowe dla miast i wsi

Średnia:

```
> #Srednia dla miast i wsi według języka obowiązkowego
> meanmiasto <- mean(miasto$`Liczba Uczniów Język Obowiązkowy`)
> meanwiasto
[1] 121.0427
> meanwies <- mean(wies$`Liczba Uczniów Język Obowiązkowy`)
> meanwies
[1] 58.24585
> #Srednia dla miast i wsi według języka dodatkowego
> meanmiastododatkowy <- mean(miasto$`Liczba Uczniów Język Dodatkowy`)
> meanmiastododatkowy
[1] 9.247241
> meanwiesdodatkowy <- mean(wies$`Liczba Uczniów Język Dodatkowy`)
> meanwiesdodatkowy
[1] 6.416396
```

Rysunek 6 Średnie dla miast i wsi



Mediana:



```
> #Mediana dla miast i wsi według języka obowiązkowego
> medianmiasto <- median(miasto$`Liczba Uczniów Język Obowiązkowy`)
> medianmiasto
[1] 64
> medianwies <- median(wies$`Liczba Uczniów Język Obowiązkowy`)
> medianwies
[1] 34
>
> #Mediana dla miast i wsi według języka dodatkowego
> medianmiastododatkowy <- median(miasto$`Liczba Uczniów Język Dodatkowy`)
> medianmiastododatkowy
[1] 0
> medianwiesdodatkowy <- median(wies$`Liczba Uczniów Język Dodatkowy`)
> medianwiesdodatkowy
[1] 0
```

Rysunek 7 Mediana dla miast i wsi

Kwartyle:

```
> #Kwartyle dla miast i wsi według języka obowiązkowego
> quantilemiasto <- quantile(miasto$\text{Liczba Uczniów Język Obowiązkowy}\)
> quantilemiasto
0% 25% 50% 75% 100%
0 23 64 141 2371
> quantilewies <- quantile(wies\text{Liczba Uczniów Język Obowiązkowy}\)
> quantilewies
0% 25% 50% 75% 100%
0 16 34 71 1553
> #Kwartyle dla miast i wsi według języka dodatkowego
> quantilemiastododatkowy <- quantile(miasto\text{Liczba Uczniów Język Dodatkowy}\)
> quantilemiastododatkowy
0% 25% 50% 75% 100%
0 0 0 0 830
> quantilewiesdodatkowy
0% 25% 50% 75% 100%
0 0 0 0 604
```

Rysunek 8 Kwartyle dla miast i wsi

Wariancja:





```
> #Wariancja dla miast i wsi według języka obowiązkowego
> varmiasto <- var(miasto$`Liczba Uczniów Język Obowiązkowy`)
> varmiasto
[1] 26248.84
> varwies <- var(wies$`Liczba Uczniów Język Obowiązkowy`)
> varwies
[1] 5689.791
>
> #Wariancja dla miast i wsi według języka dodatkowego
> varmiastododatkowy <- var(miasto$`Liczba Uczniów Język Dodatkowy`)
> varmiastododatkowy
[1] 1388.194
> varwiesdodatkowy <- var(wies$`Liczba Uczniów Język Dodatkowy`)
> varwiesdodatkowy
[1] 540.8287
```

Rysunek 9 Wariancja dla miast i wsi

Miara koncentracji (kurtoza):

```
> #Miara koncentracji dla miast i wsi według języka obowiązkowego
> kurtosismiasto <- kurtosis(miasto$`Liczba Uczniów Język obowiązkowy`)
> kurtosismiasto
[1] 12.1668
> kurtosiswies <- kurtosis(wies$`Liczba Uczniów Język Obowiązkowy`)
> kurtosiswies
[1] 29.97703
>
> #Miara koncentracji dla miast i wsi według języka dodatkowego
> kurtosismiastododatkowy <- kurtosis(miasto$`Liczba Uczniów Język Dodatkowy`)
> kurtosismiastododatkowy
[1] 71.4941
> kurtosiswiesdodatkowy <- kurtosis(wies$`Liczba Uczniów Język Dodatkowy`)
> kurtosiswiesdodatkowy <- kurtosis(wies$`Liczba Uczniów Język Dodatkowy`)
> kurtosiswiesdodatkowy
[1] 86.95064
```

Rysunek 10 Miara koncentracji dla miast i wsi

Skośność:





```
> #Skośność dla miast i wsi według języka obowiązkowego
> skewnessmiasto <- skewness(miasto$`Liczba Uczniów Język Obowiązkowy`)
> skewnessmiasto
[1] 2.643239
> skewnesswies <- skewness(wies$`Liczba Uczniów Język Obowiązkowy`)
> skewnesswies
[1] 3.841255
>
> #Skośność dla miast i wsi według języka dodatkowego
> skewnessmiastododatkowy <- skewness(miasto$`Liczba Uczniów Język Dodatkowy`)
> skewnessmiastododatkowy
[1] 6.920203
> skewnesswiesdodatkowy <- skewness(wies$`Liczba Uczniów Język Dodatkowy`)
> skewnesswiesdodatkowy
[1] 7.154551
```

Rysunek 11 Skośność dla miast i wsi

Współczynnik zmienności:

```
> #współczynnik zmienności dla miast i wsi według języka obowiązkowego
> wzmiasto <- sd(miasto$`Liczba Uczniów Język Obowiązkowy`) / mean(miasto$`Liczba Uczniów Język Obowiązkowy`)
> wzmiasto
[1] 1. 338494
> wzwies <- sd(wies$`Liczba Uczniów Język Obowiązkowy`) / mean(wies$`Liczba Uczniów Język Obowiązkowy`)
> wzwies
[1] 1. 29504
> #współczynnik zmienności dla miast i wsi według języka dodatkowego
> wzmiastododatkowy <- sd(miasto$`Liczba Uczniów Język Dodatkowy`) / mean(miasto$`Liczba Uczniów Język Dodatkowy`)
> wzmiastododatkowy
[1] 4. 029145
> wzwiesdodatkowy <- sd(wies$`Liczba Uczniów Język Dodatkowy`) / mean(wies$`Liczba Uczniów Język Dodatkowy`)
> wzwiesdodatkowy <- sd(wies$`Liczba Uczniów Język Dodatkowy`) / mean(wies$`Liczba Uczniów Język Dodatkowy`)
[1] 3. 624422</pre>
```

Rysunek 12 Współczynnik zmienności dla miast i wsi

Błąd standardowy:

```
> #Błąd standardowy dla miast i wsi według języka obowiązkowego
> bsmiasto <- sd(miasto$`Liczba Uczniów Język Obowiązkowy`) / sqrt(length(miasto$`Liczba Uczniów Język Obowiązkowy`))
> bswies <- sd(wies$`Liczba Uczniów Język Obowiązkowy`) / sqrt(length(wies$`Liczba Uczniów Język Obowiązkowy`))
> bswies
[1] 0.4029294
> #Błąd standardowy dla miast i wsi według języka dodatkowego
> bsmiastododatkowy <- sd(miasto$`Liczba Uczniów Język Dodatkowy`) / sqrt(length(miasto$`Liczba Uczniów Język Dodatkowy`))
> bsmiastododatkowy
| 0.1755737
> bswiesdodatkowy <- sd(wies$`Liczba Uczniów Język Dodatkowy`) / sqrt(length(wies$`Liczba Uczniów Język Dodatkowy`))
> bswiesdodatkowy <- sd(wies$`Liczba Uczniów Język Dodatkowy`) / sqrt(length(wies$`Liczba Uczniów Język Dodatkowy`))
> bswiesdodatkowy
| 0.1242255
```

Rysunek 13 Błąd standardowy dla miast i wsi

5. Graficzna prezentacja danych





Do graficznej prezentacji danych potrzebne będą 3 pakiety, które pomogą nam utworzyć wykresy dopasowane do danych, które analizujemy oraz zapewnić im lepszą graficzną reprezentacje na wykresach, są to:

- ggplot2
- RColorBrewer
- plotrix

Paczki instalujemy i wczytujemy za pomocą poleceń:

- install.packages("ggplot2") instalacja paczki do bardziej zaawansowanych wykresów
- install.packages("RColorBrewer") instalacja paczki do kolorów na wykresach
- install.packages("plotrix") instalacja paczki do wykresu pie3D
- library("ggplot2") biblioteka do wykresów
- library("RColorBrewer") biblioteka do kolorów w wykresach
- library("plotrix") biblioteka do wykresu trójwymiarowego pie3D

Przed uruchomieniem wykresów używam także funkcji options(scipen=999), która ma na celu wyłączyć notację wykładniczą na wykresach, a także dzięki:

- miastosuma <- aggregate(miasto\$`Liczba Uczniów Język Obowiązkowy`, by=list(miasto\$Województwo), FUN = sum)
- wiessuma <- aggregate(wies\$`Liczba Uczniów Język Obowiązkowy`, by=list(wies\$Województwo), FUN = sum)

Zliczam ile w każdym województwie jest uczniów uczących się języków obowiązkowych. Funkcje wypisują uczniów dla miast i wsi, dzięki czemu, na wykresach i w ramce danych będziemy widzieć dane do porównania bardziej przejrzyście. Wyniki z uruchomienia obu poleceń możemy zobaczyć na rysunkach poniżej.







Rysunek 14 Liczba uczniów w danym województwie dla miast

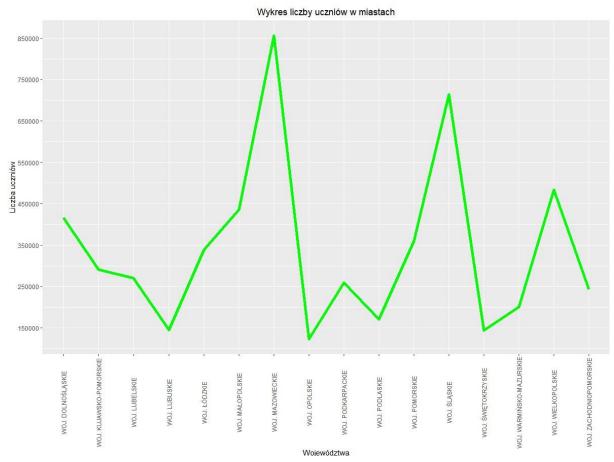


Rysunek 15 Liczba uczniów w danych województwie dla wsi





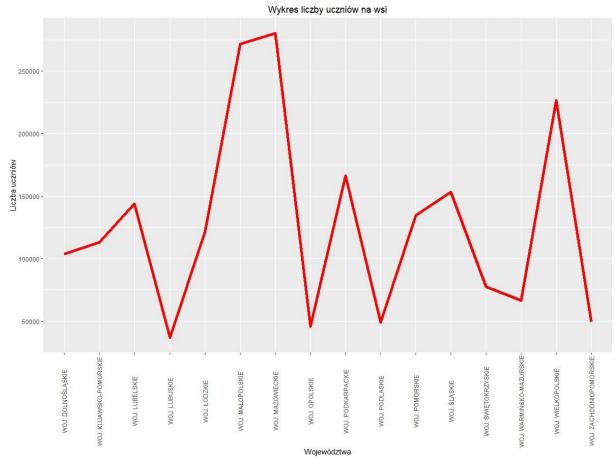
Przechodzę do wizualizacji danych, by były one widoczne bardziej przejrzyście do porównania liczby uczniów.



Rysunek 16 Liczba uczniów dla województwa w miastach







Rysunek 17 Liczba uczniów dla województwa na wsi

Oba powyższe wykresy są wykresami liniowymi zrobionymi za pomocą ggplot i pokazują liczbę uczniów uczących się języków na dane województwo. Podzielone są one oczywiście na miasta i wsie. Z wykresów można jasno odczytać, że uczniów w mieście jest oczywiście więcej. W obu przypadkach największą liczbę uczniów możemy znaleźć w województwie Mazowieckim, co nie powinno dziwić, ze względu na to, że w tym województwie znajduje się stolica naszego kraju – Warszawa. Najmniej uczniów w miastach można znaleźć w województwie Opolskim, jest to około 120000 dzieci, natomiast biorąc pod uwagę wieś jest to województwo Lubuskie (około 37000 uczniów). W województwach, gdzie urbanizacja jest naprawdę wysoka tj. Mazowieckie, Śląskie, Wielkopolskie można zauważyć, że uczniów jest najwięcej, z racji tego można również wywnioskować, że populacja tych województw jest o wiele wyższa od pozostałych. Nieco inaczej sprawa ma się na wsi, gdzie najwięcej dzieci uczęszcza na lekcje językowe na Podkarpaciu, w Wielkopolsce, w Mazowieckim i Małopolskim (gdzie liczba uczniów jest niemal równa Mazowieckiemu). Patrząc przez pryzmat tego widać też różnice w woj. Podkarpackim, gdzie w miastach jest ono notowane bardzo nisko, a na wsiach tak jak wspomniałem wyżej jest jednak w górnej części "tabeli". Biorąc dane ze wsi można wywnioskować z



wykresu, że w Małopolskim i Mazowieckim uczy się na wsi około 270-280tys. dzieci, Podkarpackie to około 166tys. a Wielkopolskie 226tys. Dokładniejsze dane z wykresów widać również na rysunkach 14 i 15, gdzie co do jednego ucznia mamy pokazane jak wygląda liczba uczniów języków obcych w oświacie.

Do stworzenia wykresów liniowych użyłem ww. funkcji ggplot, a ich kod wygląda następująco:

```
#wykres liczby uczniów w województwach w mieście
ggplot(data.frame(miastosuma), aes(Group.1, x, group=1)) + geom_line(color="green", size = 2) + theme(axis.text.x = element_text(angle=90,
vjust=0.5)) + scale_y_continuous(breaks=seq(50000,900000,by=1000000)) + ylab("Liczba uczniów") + xlab("województwa") + ggtitle("wykres
liczby uczniów w miastach") + theme(plot.title = element_text(hjust = 0.5))

#wykres liczby uczniów w województwach na wsi
ggplot(data.frame(wiessuma), aes(Group.1, x, group=1)) + geom_line(color="red", size = 2) + theme(axis.text.x = element_text(angle=90,
vjust=0.5)) + scale_y_continuous(breaks=seq(50000,300000,by=50000)) + ylab("Liczba uczniów") + xlab("Województwa") + ggtitle("wykres
liczby uczniów na wsi") + theme(plot.title = element_text(hjust = 0.5))
```

Rysunek 18 Polecenia do wywołania wykresów liniowych

Dla obu wykresów przypisujemy ramkę (dataframe(wiessuma/miastosuma)), którą chcemy na wykresie, następnie podajemy jak mają być ułożone kolumny (aes), geom_line opisuje grubość linii, theme jak mają być ułożone wartości x (pionowo 90 stopni), scale_y_continous wypisuje sekwencje liczb dla wartości osi y, dalej funkcjami ylab oraz xlab nazywamy osie, a ostatnia opcja ggtitle nazywa nam wykres.

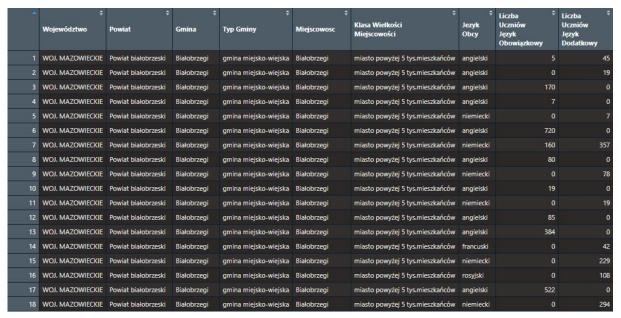
Następnym wykresem w kodzie jest wykres liczby uczniów uczących się danego języka podstawowego lub dodatkowego w województwie Mazowieckim, aby go wykonać najpierw z ramki jezyki wybieramy jedynie szkoły z ww. województwa. i bierzemy 6 najpopularniejszych, dla lepszego porównania wyników. Polecenia wykonujące to wygląda następująco:

- wojewodztwouczniowiemazowieckie <- jezyki[jezyki\$Województwo=="WOJ. MAZOWIECKIE",]
- wums <- (tail(sort(table(wojewodztwouczniowiemazowieckie\$`Jezyk Obcy`))))

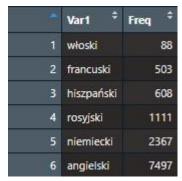
Pierwsza funkcja wybiera szkoły jedynie z mazowsza, a druga 6 najczęściej uczonych języków.







Rysunek 19 Szkoły z woj. Mazowieckiego

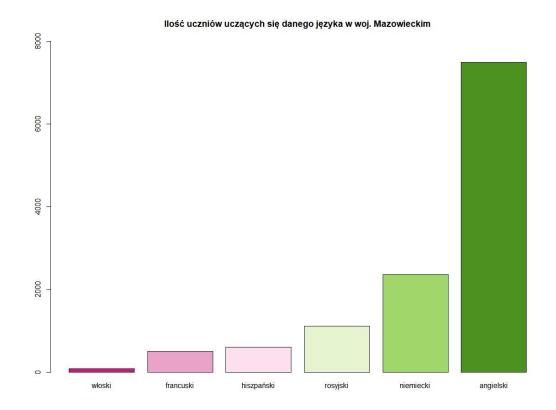


Rysunek 20 Najczęściej uczone języki w woj. Mazowieckim

W drugiej ramce, możemy odczytać dokładne wartości, które następnie zostaną przeniesione na wykres. Typ wykresu, który użyjemy do graficznej prezentacji tego typu danych to wykres typu słupkowego.







Rysunek 21 Wykres ilości uczniów w woj. Mazowieckim

Wcześniej wymieniony wykres słupkowy pokazuje bardziej przejrzyście, że język angielski nie ma sobie równych na Mazowszu. Uczących się języka angielskiego jest niemal 7500, trzy razy więcej niż drugiego języka niemieckiego (2400 dzieci). Trzecim na liście jest rosyjski (1100 uczniów), dalej są to prawie równo hiszpański i francuski. Najmniej popularny jest włoski, będący "najniższą" kolumną na wykresie. Obecność języka angielskiego nie powinna dziwić, gdyż jest to język używany przez większość osób na świecie jako język do komunikacji międzynarodowej.

Do stworzenia histogramu użyłem funkcji barplot, a jej wartości widać poniżej:

Rysunek 22 Kod źródłowy histogramu

Przypisujemy do barplot'a ramkę danych (w tym przypadku wums), ograniczamy oś y ylim'em do liczb od 0 do 8000, funkcją main podpisujemy nazwę wykresu, a col=brewer.pal wybieramy dla kolumn paletę kolorów.

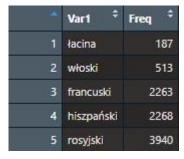




Przechodząc dalej wygenerowałem wykres języków oferowanych w szkołach bez najpopularniejszych języków (angielski i niemiecki), po to aby sprawdzić jakich jeszcze języków uczą się dzieci i młodzież w szkołach w Polsce. Użyłem do tego funkcji:

 jezykisortowanie <- head(tail((sort(table(jezyki\$'Jezyk Obcy'),DECREASING=F)), n=7), n=5)

Wybierającej 5 najczęściej wybieranych języków do nauki w Polsce, dane niemal całkowicie pokrywają się z wcześniejszą analizą województwa Mazowieckiego, przez co można stwierdzić, że języki te nie tylko na Mazowszu ale i w całym kraju są tak samo popularne.



Rysunek 23 Najczęściej wybierane języki poza angielskim i niemieckim

W tabeli widzimy dokładne wartości języków wybieranych w Polsce.

Tworzymy wykres za pomocą funkcji z rysunku 24:

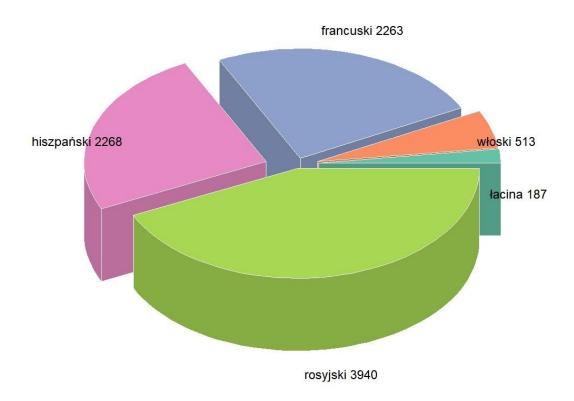
Rysunek 24 Funkcja do wykresu kołowego 3D

Dzięki lbls możemy przypisać nazwy jak i wartości wykresowi, a paleta barw tym razem odpowiada za kolory. Wywołujemy pie3D, wczytujemy ramke jezykisortowanie, określamy kolor linii za pomocą border, mainem podpisujemy wartości jak we wcześniejszych wykresach, labels przypisujemy lbls, dzięki czemu wyświetlam nazwy języków wraz z wartościami. Theta, explode oraz high pomagają nam w określeniu wymiarów wykresu, aby był on czytelny nawet dla mniejszych wartości takich jak łacina, a col odwołuje się do wartości kolorów co nadaje wykresowi przejrzystość.





Wykres języków oferowanych w szkołach bez najpopularniejszych języków (angielski i niemiecki)



Rysunek 25 Języki wybierane najczęściej w naszym kraju

Do przestawienia najbardziej popularnych języków do nauki w Polsce użyłem wykresu pie3D, jest to typ wykresu kołowego w trójwymiarze z ww. biblioteki plotrix. Jedyną odskocznią od analizy województwa Mazowieckiego jest tutaj 187 uczniów, którzy uczą się łaciny, wynika to po prostu z obcięcia ramki z angielskiego i niemieckiego. Zabieg usunięcia tych języków był celowy, gdyż obecność i popularność tych wykresów w statystykach jest pewna jako, iż są one bardzo w naszym kraju popularne. Poprzez usunięcie ich mamy lepszą perspektywę ile osób wolało wybrać jako język obowiązkowy lub dodatkowy coś innego niż te dwa. Ponad połowę liczby języków stanowią języki bardziej kojarzone z południem Europy wywodzące się z łaciny takie jak hiszpański, włoski, francuski. Najczęściej uczonym językiem w Polsce poza nimi jest Rosyjski z niemalże czterema tysiącami uczniów.





Patrząc na poprzednie wykresy, można się zastanowić jakie są najmniej popularne, "egzotyczne" języki wybierane w naszym kraju, dzięki funkcji

• najmniejuczacych <- head(sort(table(jezyki\$'Jezyk Obcy'),DECREASING=F), n = 13)

Policzyłem owe języki, i zapisałem do tabeli:



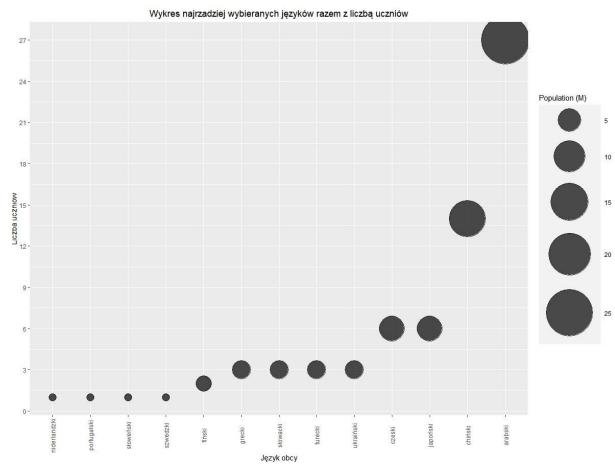
Rysunek 26 Najrzadziej wybierane języki

Z analizy wynika, że są 4 języki w Polsce, gdzie zainteresowanych ich nauką jest po jednej osobie, fińskiego uczy się dwie osoby, greckiego, słowackiego, tureckiego i ukraińskiego jest po trzy osoby. Wyżej w rankingu są czeski i japoński (6 uczniów). Czternaście osób uczy się chińskiego a 27 arabskiego. Mała liczba uczniów owych języków może wynikać z tego, że oprócz niderlandzkiego i szwedzkiego kraje używające tych języków nie są ani popularnym celem pracy, a oprócz Grecji i Turcji nie cieszą się one wśród Polaków tak dużym zainteresowaniem turystycznym. Mała liczba uczniów może też zależeć od ilości nauczycieli, potrafiących prowadzić lekcje o, oraz w danym języku w naszym kraju.





Informacje i wartości zostały zawarte w wykresie bąbelkowym poniżej, który wybrałem ze względu, iż dobrze ukazuje on małe wartości przy okazji porównując je w bardzo wygodnym dla oka stylu. Wykres opisuje wszystkie potrzebne informacje z tabeli i poprzez większy okrąg reprezentuje większą populacje (liczbę uczniów)



Rysunek 27 Wykres bąbelkowy najrzadziej wybieranych języków

Do stworzenia wykresu użyłem kodu opisanego poniżej:





Rysunek 28 Tworzenie wykresu bąbelkowego

Wektor pop zapisuje nam wszystkie wartości potrzebne do utworzenia okręgów, bąbelków, następnie tworzymy tabele z rysunku 26, konwertujemy do ramki danych. W ggplocie wypisujemy ramke, wartości kolumn i, geom_point ma za zadanie utrzymać transparentność okręgów, theme wypisuje nazwy języków pod kątem 90 stopni, scale_size dobiera wielkość bąbelków, natomiast następne wartości działają dokładnie tak samo jak w wykresach liniowych z rysunków 16 i 17.

6. Hipotezy statystyczne

Do weryfikacji hipotez użyję funkcji t.test, przeprowadzająca testy na wektorach języku R. Jeśli hipoteza jest prawdziwa to wartość p musi być większa niż 0.05

1. Hipoteza pierwsza

Średnia uczniów uczących się języka obowiązkowego w województwach w Polsce jest równa 93

```
> hipoteza1 <- t.test(x=jezyki$`Liczba Uczniów Język Obowiązkowy`, mu=93)
> hipoteza1

One Sample t-test

data: jezyki$`Liczba Uczniów Język Obowiązkowy`
t = 1.1892, df = 83169, p-value = 0.2344
alternative hypothesis: true mean is not equal to 93
95 percent confidence interval:
    92.64143 94.46508
sample estimates:
mean of x
    93.55325
```

Rysunek 29 Hipoteza nr.1 języki obowiązkowe

Jak widać hipoteza jest prawdziwa, z racji, że wartość p jest równa 0.2344, a średnia wynosi 93.55 czyli niemal 93, dlatego nie mamy podstaw do odrzucenia hipotezy statystycznej.





2. Hipoteza druga

Średnia uczniów uczących się języka dodatkowego w województwach w Polsce jest równa niemal 1/3 średniej uczniów języka obowiązkowego

```
> hipoteza2 <- t.test(x=jezyki$`Liczba Uczniów Język Dodatkowy`, mu=30)
> hipoteza2

One Sample t-test

data: jezyki$`Liczba Uczniów Język Dodatkowy`
t = -196.95, df = 83169, p-value < 0.00000000000000022
alternative hypothesis: true mean is not equal to 30
95 percent confidence interval:
7.853237 8.289699
sample estimates:
mean of x
8.071468</pre>
```

Po wyniku wartości p < 0.05 widzimy, że hipotezę możemy odrzucić, i średnia nie jest równa 1/3 średniej uczniów języka obowiązkowego.

H2.1:

Średnia uczniów uczących się języka dodatkowego w województwach w Polsce jest równa 8

Wartość p = 0.521, przez co hipotezę można uznać za prawdziwą, i średnia rzeczywiście jest równa prawie 8, jest to około 8.07





7. Wnioski

Z przeprowadzonych analiz możemy powiedzieć wprost, że uczniów w szkołach uczących się języków obcych najwięcej jest w miastach, ponad to najwięcej dzieci uczęszcza na lekcje językowe w woj. Mazowieckim, oraz prawie żadne województwo nie może się z nim równać w tym wyniku oprócz woj. Małopolskiego biorąc pod uwagę statystyki ze wsi. Najczęściej wybieranymi językami bez dwóch zdań są język angielski i język niemiecki, a oprócz nich jest to język rosyjski. Najmniej osób uczy się niderlandzkiego, portugalskiego, słoweńskiego i szwedzkiego (1 osoba). Najbardziej popularnym z języków o najmniejszym stopniu zainteresowania jest język arabski (27 osób). Średnio na wsi języka obowiązkowego uczy się 58 dzieci, natomiast w mieście 121 uczniów, języków dodatkowych kolejno 6 i 9 uczniów. Wszystkie te dane mają odchylenie standardowe na poziomie kolejno 75, 162, 23, 37. Analiza tych danych pomogła mi lepiej zrozumieć użycie statystyki w życiu codziennym oraz pracy analityka, a także poszerzyła moją wiedze z zakresu tego, jakich języków uczą się polacy.





8. Kod źródłowy

Statystyczna Analiza Danych, Projekt
#
Ustawienie glownego katalogu setwd("C:/Users/Sebastian/Desktop/Studia/II rok 2
semestr/Statystyczna_Analiza_Danych_LAB") getwd()
#
#Instalacja potrzebnych paczek install.packages("readxl") #instalacja paczki do
wczytania plików xls,xlsx install.packages("ggplot2") #instalacja paczki do
wykresów install.packages("RColorBrewer") #instalacja paczki do kolorów
install.packages("plotrix") #instalacja paczki do wykresu pie3D
install.packages("moments") #instalacja paczki do miary koncentracji i skośności
#
Mozliwosc wczytania plików xls,xlsx library("readxl")
Biblioteka do wykresów library("ggplot2")
#Biblioteka do kolorów w wykresach library("RColorBrewer")
#Biblioteka do wykresu trójwymiarowego pie3D
library("plotrix")
#Biblioteka do miary koncentracji i skośności library("moments")
#
#Wczytanie pliku z MS EXCEL jezyki <-
read_excel("Jezyki_obce_nowozytne_liczba_uczniow_rokszkolny20182019_wg_szkol.xlsx")
#Obróbka danych (Usuniecie niepotrzebnych kolumn) jezyki
<- jezyki[, -c(1,3,5,10:26)]
#Obrobka danych (podzielenie ramki na dwie mniejsze według miast i wsi)
(miasto <-jezyki[jezyki\$`Klasa Wielkości Miejscowości`=="miasto powyżej 5 tys.mieszkańców",])
(wies <-jezyki[jezyki\$`Klasa Wielkości Miejscowości`=="wieś",]) #
#Parametry opisowe
#Odchylenia standardowe dla miast i wsi według języka obowiazkowego
sdmiasto <- sd(miasto\$`Liczba Uczniów Język Obowiązkowy`) sdmiasto
sdwies <- sd(wies\$`Liczba Uczniów Język Obowiązkowy`) sdwies
#Odchylenia standardowe dla miast i wsi według języka dodatkowego
sdmiastododatkowy <- sd(miasto\$`Liczba Uczniów Język Dodatkowy`)





sdwiesdodatkowy <- sd(wies\$`Liczba Uczniów Język Dodatkowy`) sdwiesdodatkowy

#Srednia dla miast i wsi według języka obowiązkowego meanmiasto
<- mean(miasto\$`Liczba Uczniów Język Obowiązkowy`) meanmiasto
meanwies <- mean(wies\$`Liczba Uczniów Język Obowiązkowy`)
meanwies

#Srednia dla miast i wsi według języka dodatkowego meanmiastododatkowy
-- mean(miasto\$`Liczba Uczniów Język Dodatkowy`) meanmiastododatkowy
meanwiesdodatkowy -- mean(wies\$`Liczba Uczniów Język Dodatkowy`)
meanwiesdodatkowy

#Mediana dla miast i wsi według języka obowiązkowego medianmiasto
<- median(miasto\$`Liczba Uczniów Język Obowiązkowy`) medianmiasto
medianwies <- median(wies\$`Liczba Uczniów Język Obowiązkowy`)
medianwies

#Mediana dla miast i wsi według języka dodatkowego medianmiastododatkowy <- median(miasto\$`Liczba Uczniów Język Dodatkowy`) medianmiastododatkowy medianwiesdodatkowy <- median(wies\$`Liczba Uczniów Język Dodatkowy`) medianwiesdodatkowy

#Kwartyle dla miast i wsi według języka obowiązkowego quantilemiasto
<- quantile(miasto\$`Liczba Uczniów Język Obowiązkowy`) quantilemiasto
quantilewies <- quantile(wies\$`Liczba Uczniów Język Obowiązkowy`) quantilewies
#Kwartyle dla miast i wsi według języka dodatkowego quantilemiastododatkowy
<- quantile(miasto\$`Liczba Uczniów Język Dodatkowy`) quantilemiastododatkowy
quantilewiesdodatkowy <- quantile(wies\$`Liczba Uczniów Język Dodatkowy`)
quantilewiesdodatkowy

#Wariancja dla miast i wsi według języka obowiązkowego varmiasto <- var(miasto\$`Liczba Uczniów Język Obowiązkowy`) varmiasto varwies <- var(wies\$`Liczba Uczniów Język

Obowiązkowy') varwies

#Wariancja dla miast i wsi według języka dodatkowego
varmiastododatkowy <- var(miasto\$`Liczba Uczniów Język Dodatkowy`)
varmiastododatkowy varwiesdodatkowy <- var(wies\$`Liczba Uczniów
Język Dodatkowy`) varwiesdodatkowy

#Miara koncentracji dla miast i wsi według języka obowiązkowego





kurtosis(miasto\$`Liczba Uczniów Język Obowiązkowy`) kurtosismiasto

kurtosiswies <- kurtosis(wies\$`Liczba Uczniów Język Obowiązkowy`)

kurtosiswies

#Miara koncentracji dla miast i wsi według języka dodatkowego

kurtosismiastododatkowy <- kurtosis(miasto\$`Liczba Uczniów Język Dodatkowy`)

kurtosismiastododatkowy kurtosiswiesdodatkowy <- kurtosis(wies\$`Liczba

Uczniów Język Dodatkowy`) kurtosiswiesdodatkowy

#Skośność dla miast i wsi według języka obowiązkowego skewnessmiasto

<- skewness(miasto\$`Liczba Uczniów Język Obowiązkowy`) skewnessmiasto

skewnesswies <- skewness(wies\$`Liczba Uczniów Język Obowiązkowy`)

skewnesswies

#Skośność dla miast i wsi według języka dodatkowego skewnessmiastododatkowy <-

skewness(miasto\$`Liczba Uczniów Język Dodatkowy`) skewnessmiastododatkowy

skewnesswiesdodatkowy <- skewness(wies\$`Liczba Uczniów Język Dodatkowy`) skewnesswiesdodatkowy

#Współczynnik zmienności dla miast i wsi według języka obowiązkowego wzmiasto <- sd(miasto\$`Liczba Uczniów

Język Obowiązkowy`) / mean(miasto\$`Liczba Uczniów Język Obowiązkowy`) wzmiasto wzwies <- sd(wies\$`Liczba

Uczniów Język Obowiązkowy`) / mean(wies\$`Liczba Uczniów Język Obowiązkowy`) wzwies

#Współczynnik zmienności dla miast i wsi według języka dodatkowego wzmiastododatkowy <- sd(miasto\$`Liczba Uczniów

Język Dodatkowy`) / mean(miasto\$`Liczba Uczniów Język Dodatkowy`) wzmiastododatkowy wzwiesdodatkowy <-

sd(wies\$`Liczba Uczniów Język Dodatkowy`) / mean(wies\$`Liczba Uczniów Język Dodatkowy`) wzwiesdodatkowy

#Błąd standardowy dla miast i wsi według języka obowiązkowego bsmiasto <- sd(miasto\$`Liczba Uczniów Język

Obowiązkowy`) / sqrt(length(miasto\$`Liczba Uczniów Język Obowiązkowy`)) bsmiasto bswies <- sd(wies\$`Liczba Uczniów

Język Obowiązkowy`) / sqrt(length(wies\$`Liczba Uczniów Język Obowiązkowy`)) bswies

#Błąd standardowy dla miast i wsi według języka dodatkowego

bsmiastododatkowy <- sd(miasto\$`Liczba Uczniów Język Dodatkowy`) / sqrt(length(miasto\$`Liczba Uczniów Język

Dodatkowy`)) bsmiastododatkowy bswiesdodatkowy <- sd(wies\$`Liczba Uczniów Język Dodatkowy`) /

sqrt(length(wies\$`Liczba Uczniów Język Dodatkowy`)) bswiesdodatkowy

#-----

#Wizualizacja danych options(scipen=999)

#Suma uczniów w miastach i wsiach dla poszczególnych województw miastosuma <- aggregate(miasto\$`Liczba

Uczniów Język Obowiązkowy`, by=list(miasto\$Województwo), FUN = sum) wiessuma <- aggregate(wies\$`Liczba

Uczniów Język Obowiązkowy`, by=list(wies\$Województwo), FUN = sum)

#Wykres liczby uczniów w województwach w mieście

ggplot(data.frame(miastosuma), aes(Group.1, x, group=1)) + geom_line(color="green", size = 2) + theme(axis.text.x = element_text(angle=90, vjust=0.5)) +





scale_y_continuous(breaks=seq(50000,900000,by=100000)) + ylab("Liczba uczniów") + xlab("Województwa") + ggtitle("Wykres liczby uczniów w miastach") + theme(plot.title = element_text(hjust = 0.5))

#Wykres liczby uczniów w województwach na wsi

```
ggplot(data.frame(wiessuma), aes(Group.1, x, group=1)) + geom_line(color="red", size = 2) + theme(axis.text.x =
element_text(angle=90, vjust=0.5)) + scale_y_continuous(breaks=seq(50000,300000,by=50000)) + ylab("Liczba uczniów") +
xlab("Województwa") + ggtitle("Wykres liczby uczniów na wsi") + theme(plot.title = element_text(hjust = 0.5)) #Wykres liczby
uczniów uczących się danego języka podstawowego lub dodatkowego w województwie mazowieckim
wojewodztwouczniowiemazowieckie <- jezyki[jezyki$Województwo=="WOJ. MAZOWIECKIE",] wums <-
(tail(sort(table(wojewodztwouczniowiemazowieckie$`Jezyk Obcy`)))) wojewodztwawykres <- barplot(wums,
                                ylim = c(0,8000),
                                                                                                main = "Ilość uczniów uczących się
danego języka w woj. Mazowieckim",
                                                                                                    col=brewer.pal(n = 6, name =
"PiYG")
)
#Wykres języków oferowanych w szkołach bez najpopularniejszych (niemieckiego i angielskiego)
jezykisortowanie <- head(tail((sort(table(jezyki$'Jezyk Obcy'),DECREASING=F)), n=7), n=5) lbls
<- paste(names(jezykisortowanie), jezykisortowanie, sep=" ") paletabarw <- brewer.pal(8, "Set2")</pre>
(wykres_jezykow <- pie3D(jezykisortowanie,
                          border="white",
                                                                                 main = "Wykres języków oferowanych w szkołach bez najpopularniejszych
języków \n (angielski i niemiecki)",
                                                                                        labels = lbls,
                                                                                                                                         theta = 0.95,
                                                                                                                                                                                           explode = 0.15.
height = 0.2,
                                                 col = paletabarw))
#Wykres najrzadziej wybieranych języków razem z liczbą uczniów pop <-
c(1,1,1,1,2,3,3,3,3,6,6,14,27) #przypisanie ilości kropek do wektora
najmniejuczacych <- head(sort(table(jezyki$'Jezyk Obcy'),DECREASING=F), n = 13) #wybranie 13 najrzadziej wybieranych
najmniejuczacych <- as.data.frame(najmniejuczacych) #zamiana na ramke danych ggplot(najmniejuczacych,
aes(Var1,Freq, size = pop)) +
  geom\_point(alpha=0.7) + theme(axis.text.x = element\_text(angle=90, vjust=0.5)) + scale\_size(range = c(5, 33), name="Population of the context of the conte
scale y continuous(breaks=seq(0,40,by=3)) + ylab("Liczba uczniów") + xlab("Język obcy") + ggtitle("Wykres najrzadziej
wybieranych języków razem z liczbą uczniów") + theme(plot.title = element_text(hjust = 0.5)) #Hipotezy
#Hipoteza 1
#Średnia uczniów uczących się języka obowiązkowego w województwach w Polsce jest równa 93
hipoteza1 <- t.test(x=jezyki$`Liczba Uczniów Język Obowiązkowy`, mu=93) hipoteza1
```



#Hipoteza 2



#Średnia uczniów uczących się języka dodatkowego w województwach w Polsce jest równa niemal 1/3 średniej uczniów języka obowiązkowego

hipoteza2 <- t.test(x=jezyki\$`Liczba Uczniów Język Dodatkowy`, mu=30) hipoteza2

#Średnia uczniów uczących się języka dodatkowego w województwach w Polsce jest równa 8

hipoteza3 <- t.test(x=jezyki\$`Liczba Uczniów Język Dodatkowy`, mu=8) hipoteza3





9. Źródła

- https://dane.gov.pl/pl/dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szko/table?page=1&per page=20&q=&sort="">https://dane.gov.pl/pl/dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szko/table?page=1&per page=20&q=&sort="">https://dane.gov.pl/pl/dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szko/table?page=1&per page=20&q=&sort="">https://dane.gov.pl/pl/dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szko/table?page=1&per page=20&q=&sort="">https://dane.gov.pl/pl/dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szko/table?page=1&per page=20&q=&sort="">https://dane.gov.pl/pl/dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szko/table?page=1&per page=20&q=&sort="">https://dane.gov.pl/pl/dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szko/table?page=1&per page=20&q=&sort="">https://dane.gov.pl/pl/dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szko/table?page=1&per page=20&q=&sort="">https://dane.gov.pl/pl/dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szko/table?page=1&per page=20&q=&sort="">https://dataset/220/resource/18517,jezyki-obce-nowozytne-liczba-ucznioww-roku-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym-20182019-wg-szkolnym
- https://www.cran.r-project.org
- https://www.gov.pl
- https://www.geeksforgeeks.org/r-charts-and-graphs/
- https://krakow.stat.gov.pl/cps/rde/xbcr/krak/ASSETS 2006 11 inf syg.pdf?fbclid=IwAR1Ht UmT GOSMUwRi1mALMUwjTVKsHpmBmPNUCA1ewaJa9lG1aCnMh3V3jM
- https://www.bankier.pl/wiadomosc/Co-trzeci-Polak-nie-zna-zadnego-jezykaobcego7744683.html





