



**POLITECHNIKA
RZESZOWSKA**

Demografia

LICZBA ZMARŁYCH 2010-2020

PROWADZĄCY: DR MARIUSZ STARTEK

PRZEDMIOT: STATYSTYCZNA ANALIZA DANYCH – ZAJĘCIA
LABORATORYJNE

AUTOR: ADRIAN DEREŃ

KIERUNEK: INŻYNIERIA I ANALIZA DANYCH

NR. INDEKSU 166724

Dane wejściowe.

Dane wykorzystywane w projekcie pochodzą ze strony Głównego Urzędu Statystycznego a dokładniej z działu Demografii. Następnie spośród dostępnych tablic wybrałem tablice trwania życia 1990-2020. Pobrany plik w formacie .xls zawiera 30 arkuszy odpowiadających każdemu rokowi, z których każdy posiada 8 kolumn oraz ponad 200 wierszy. Dane zostały opracowane na podstawie 100 tys. osób oraz zostały podzielone ze względu na płeć i wiek. Kolumny to kolejno: płeć, wiek, liczba dożywających, prawdopodobieństwo zgonu, liczba zmarłych, ludność stacjonarna, ludność stacjonarna skumulowana oraz przeciętne dalsze trwanie życia. Kolumna płeć zawiera wyłącznie wartości 1 i 2 odpowiadające płci tzn. 1 – mężczyźni, 2 – kobiety, natomiast wiek wartości od 0 do 100. Do realizacji projektu wybrałem wartości z kolumny „liczba zmarłych” równocześnie ograniczając zakres lat z 1990 - 2020 na 2010 – 2020 oraz dzieląc na oddzielne tablice dla kobiet i mężczyzn.

Płeć 1-mężczyźni 2-kobiety	Wiek	Liczba dożywa- jących	Prawdo- podobień- stwo zgonu	Liczba zmarłych	Ludność stacjonarna	Ludność stacjonarna skumulo- wana	Przeciętne dalsze trwanie życia
sex	x	lx	qx	dx	Lx	Tx	ex
1	0	100000	0,00389	389	99650	7260673	72,61
1	1	99611	0,00028	28	99597	7161024	71,89
1	2	99583	0,00019	18	99574	7061427	70,91
1	3	99565	0,00015	15	99558	6961853	69,92
1	4	99550	0,00012	12	99544	6862295	68,93
1	5	99538	0,00010	10	99533	6762751	67,94
1	6	99528	0,00009	9	99524	6663218	66,95
1	7	99519	0,00009	9	99515	6563695	65,95
1	8	99510	0,00009	9	99506	6464180	64,96
1	9	99501	0,00009	8	99497	6364675	63,97
1	10	99493	0,00009	9	99489	6265178	62,97
1	11	99484	0,00010	10	99479	6165689	61,98
1	12	99474	0,00011	11	99469	6066210	60,98
1	13	99463	0,00014	14	99456	5966742	59,99
1	14	99449	0,00017	17	99441	5867286	59,00
1	15	99432	0,00023	23	99421	5767845	58,01
1	16	99409	0,00030	30	99394	5668425	57,02
1	17	99379	0,00040	39	99360	5569031	56,04
1	18	99340	0,00050	50	99315	5469671	55,06
1	19	99290	0,00061	61	99260	5370356	54,09
1	20	99229	0,00072	71	99194	5271097	53,12
1	21	99158	0,00081	80	99118	5171903	52,16
1	22	99078	0,00088	88	99034	5072785	51,20
1	23	98990	0,00094	93	98944	4973751	50,24
1	24	98897	0,00099	98	98848	4874808	49,29
1	25	98799	0,00104	103	98748	4775960	48,34
1	26	98696	0,00109	107	98643	4677212	47,39
1	27	98589	0,00114	112	98533	4578570	46,44
1	28	98477	0,00119	118	98418	4480037	45,49
1	29	98359	0,00126	124	98297	4381619	44,55
1	30	98235	0,00135	132	98169	4283322	43,60

Rysunek 1 Urywek danych z roku 2020 po pobraniu ze strony statgov.pl

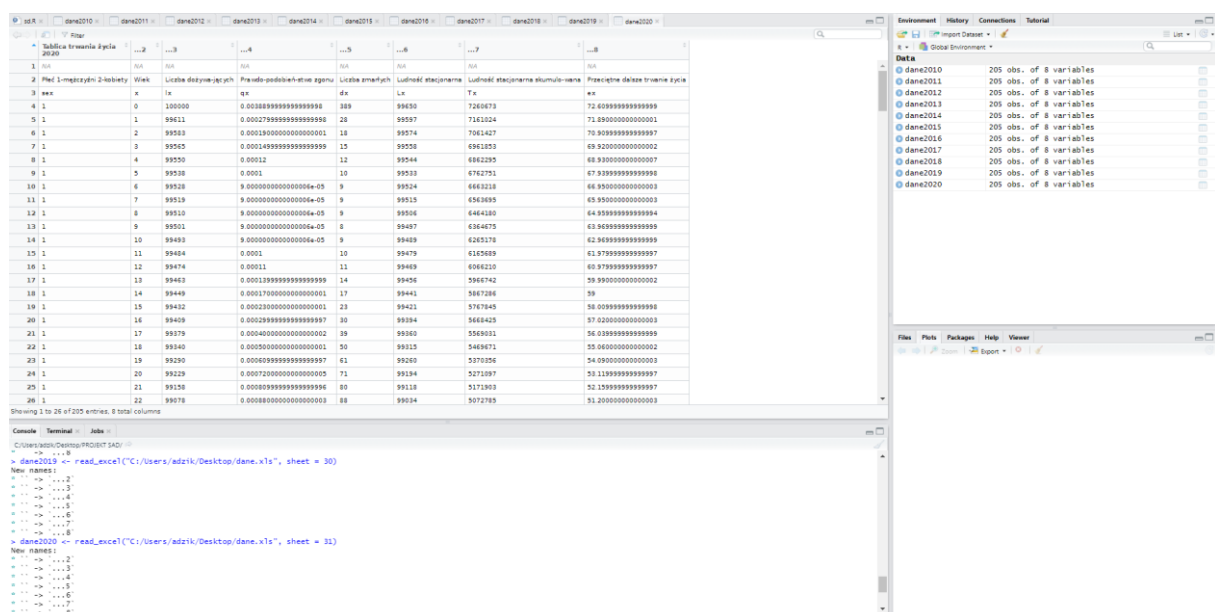
Wczytanie danych.

Po wyborze danych nadszedł czas na kolejny krok jakim było wczytanie danych do środowiska R. Dane udało się wczytać przy pomocy biblioteki „readxl”, która umożliwia odczytanie danych z formatem m.in. .xls i dodatkowo sprecyzować, który dokładnie arkusz z pliku zostanie wczytany. Wczytałem zatem dane od 2010 do 2020 roku, tworząc kolejno 11 tablic, z których każdą przypisałem do zmiennych.

```
1 ##### W CZYTANIE DANYCH
2 library(readxl)
3 dane2010 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 21)
4 dane2011 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 22)
5 dane2012 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 23)
6 dane2013 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 24)
7 dane2014 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 25)
8 dane2015 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 26)
9 dane2016 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 27)
10 dane2017 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 28)
11 dane2018 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 29)
12 dane2019 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 30)
13 dane2020 <- read_excel("C:/Users/adzik/Desktop/dane.xls", sheet = 31)
14
```

Rysunek 2 Wczytanie danych do środowiska R

Po wczytaniu dane prezentowały się dla każdego roku w następujący sposób :



	m	w	t	p	d	l	t	p
1	1	1	100000	0.000000000000000000	389	99430	7240673	72.409999999999999
2	2	1	99611	0.00027999999999999999	28	99597	7161024	71.890000000000001
3	3	1	99583	0.00013000000000000001	18	99574	7061427	70.909999999999997
4	4	1	99585	0.00014999999999999999	15	99558	6961193	69.920000000000002
5	5	1	99550	0.00012	12	99544	6862395	68.930000000000007
6	6	1	99530	0.0001	10	99533	6762751	67.939999999999998
7	7	1	99528	9.0000000000000000e-05	9	99524	6662218	66.950000000000003
8	8	1	99519	9.0000000000000000e-05	9	99515	6563895	65.950000000000003
9	9	1	99510	9.0000000000000000e-05	9	99506	6464180	64.959999999999994
10	10	1	99501	9.0000000000000000e-05	8	99497	6364075	63.969999999999999
11	11	1	99493	9.0000000000000000e-05	9	99489	6263276	62.969999999999999
12	12	1	99484	0.0001	10	99479	6163089	61.979999999999997
13	13	1	99474	0.00011	11	99469	6062210	60.979999999999997
14	14	1	99463	0.00011999999999999999	14	99456	5966742	59.990000000000002
15	15	1	99449	0.00017000000000000001	17	99441	5867286	58
16	16	1	99432	0.00023000000000000001	23	99421	5767149	58.039999999999998
17	17	1	99409	0.00029999999999999997	30	99394	5664625	57.030000000000003
18	18	1	99379	0.00040000000000000002	39	99380	5563031	56.039999999999999
19	19	1	99340	0.00050000000000000001	50	99315	5463671	55.060000000000002
20	20	1	99290	0.00060999999999999997	61	99260	5370356	54.090000000000003
21	21	1	99229	0.00072000000000000003	71	99134	5271197	53.119999999999997
22	22	1	99126	0.00080999999999999996	80	99118	5171303	52.159999999999997
23	23	1	99178	0.00088000000000000003	88	99034	5072789	51.200000000000003

Rysunek 3 Dane po wczytaniu do środowiska R

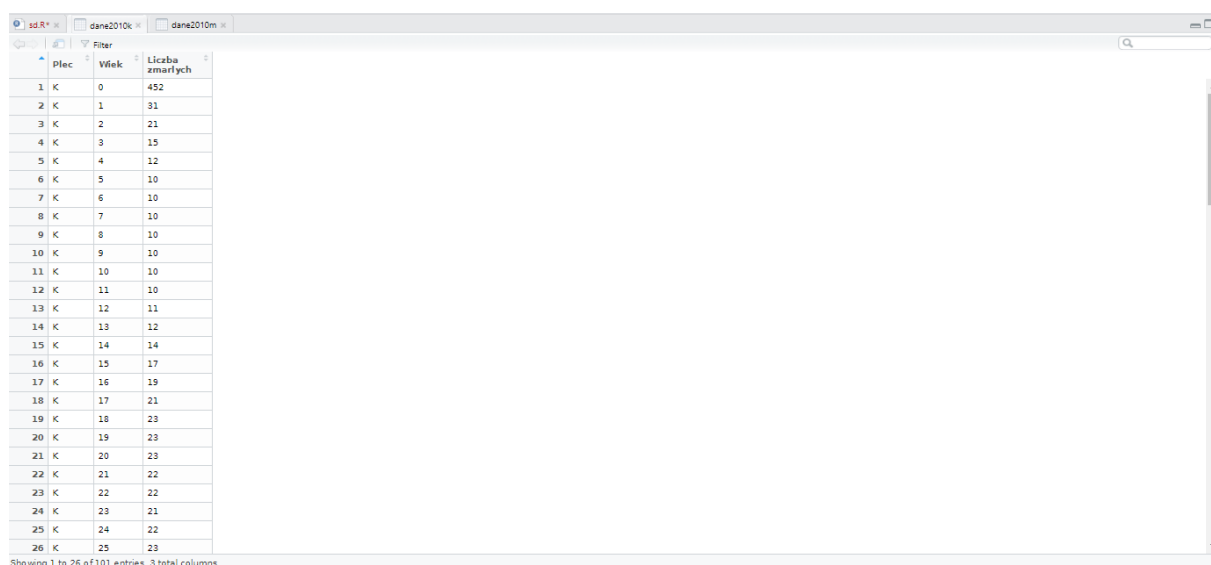
Jak możemy zauważyć, pojawiły się wartości NA, oraz kolumny które nie będą potrzebne w dalszej realizacji projektu.

Opracowanie danych

Kontynuując pracę nad projektem w kolejnym etapie dane musiały zostać zmodyfikowane, aby operacje na poszczególnych tabelach były możliwe a środowisko nie zwracało błędów.

W pierwszej kolejności zmieniłem format wczytanych danych na ramkę danych używając komendy `as.data.frame`, następnie usunąłem kolumny tj. liczba dożywających, prawdopodobieństwo zgonu, ludność stacjonarna, ludność stacjonarna skumulowana i przeciętne dalsze trwanie życia. Pierwsze trzy wiersze, które były nazwami kolumn, zastąpiłem jednym wierszem z nazwami kolejno: płeć, wiek oraz liczba zmarłych. Na końcu wartości 1 i 2 odpowiadające płci zastąpiłem literami M i K oraz ramkę z każdego roku podzieliłem na dwie ramki względem płci.

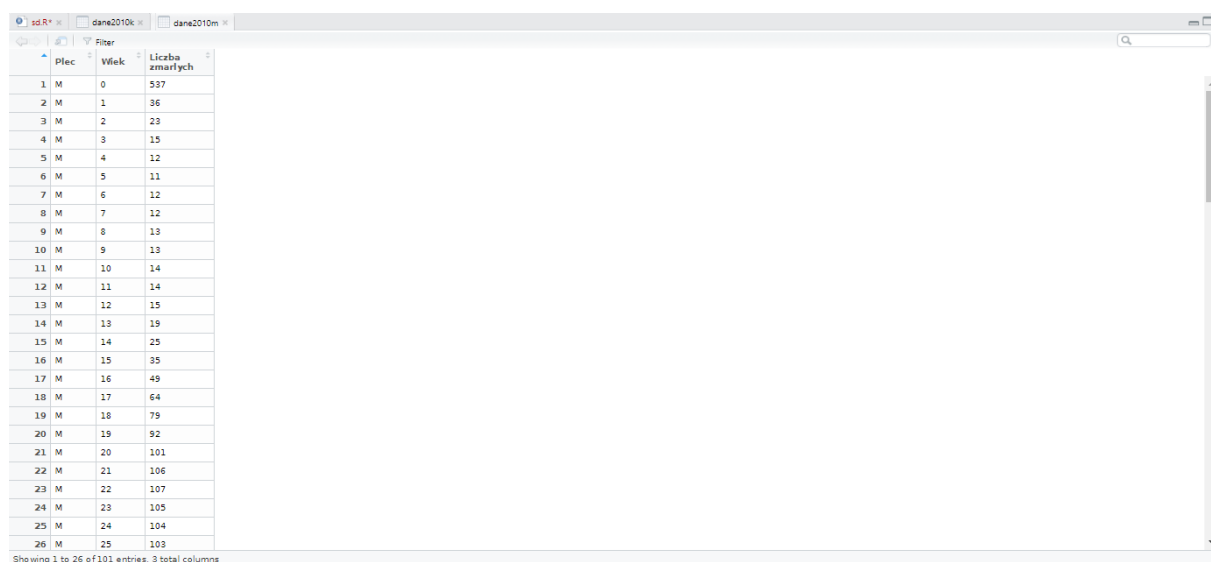
Dane przybrały wówczas dla każdego roku postać:



	Płeć	Wiek	Liczba zmarłych
1	K	0	452
2	K	1	31
3	K	2	21
4	K	3	15
5	K	4	12
6	K	5	10
7	K	6	10
8	K	7	10
9	K	8	10
10	K	9	10
11	K	10	10
12	K	11	10
13	K	12	11
14	K	13	12
15	K	14	14
16	K	15	17
17	K	16	19
18	K	17	21
19	K	18	23
20	K	19	23
21	K	20	23
22	K	21	22
23	K	22	22
24	K	23	21
25	K	24	22
26	K	25	23

Showing 1 to 26 of 101 entries, 3 total columns

Rysunek 4 Dane dla roku 2010 - kobiety



	Płeć	Wiek	Liczba zmarłych
1	M	0	537
2	M	1	36
3	M	2	23
4	M	3	15
5	M	4	12
6	M	5	11
7	M	6	12
8	M	7	12
9	M	8	13
10	M	9	13
11	M	10	14
12	M	11	14
13	M	12	15
14	M	13	19
15	M	14	25
16	M	15	35
17	M	16	49
18	M	17	64
19	M	18	79
20	M	19	92
21	M	20	101
22	M	21	106
23	M	22	107
24	M	23	105
25	M	24	104
26	M	25	103

Showing 1 to 26 of 101 entries, 3 total columns

Rysunek 5 Dane dla roku 2010 - mężczyźni

Wraz z kolejnym krokiem zmieniłem typ danych kolumny „Liczba zmarłych” na numeric w celu poprawnego działania przyszłych operacji na tych właśnie kolumnach.

```
175 ▶ #####
176 ▶ ##### LICZBA ZMARLYCH KAZDY ROK #####
177 ▶ #####
178
179 lzm2010m <- as.numeric(dane2010m$Liczba zmarlych)
180 lzm2010k <- as.numeric(dane2010k$Liczba zmarlych)
181 lzm2011m <- as.numeric(dane2011m$Liczba zmarlych)
182 lzm2011k <- as.numeric(dane2011k$Liczba zmarlych)
183 lzm2012m <- as.numeric(dane2012m$Liczba zmarlych)
184 lzm2012k <- as.numeric(dane2012k$Liczba zmarlych)
185 lzm2013m <- as.numeric(dane2013m$Liczba zmarlych)
186 lzm2013k <- as.numeric(dane2013k$Liczba zmarlych)
187 lzm2014m <- as.numeric(dane2014m$Liczba zmarlych)
188 lzm2014k <- as.numeric(dane2014k$Liczba zmarlych)
189 lzm2015m <- as.numeric(dane2015m$Liczba zmarlych)
190 lzm2015k <- as.numeric(dane2015k$Liczba zmarlych)
191 lzm2016m <- as.numeric(dane2016m$Liczba zmarlych)
192 lzm2016k <- as.numeric(dane2016k$Liczba zmarlych)
193 lzm2017m <- as.numeric(dane2017m$Liczba zmarlych)
194 lzm2017k <- as.numeric(dane2017k$Liczba zmarlych)
195 lzm2018m <- as.numeric(dane2018m$Liczba zmarlych)
196 lzm2018k <- as.numeric(dane2018k$Liczba zmarlych)
197 lzm2019m <- as.numeric(dane2019m$Liczba zmarlych)
198 lzm2019k <- as.numeric(dane2019k$Liczba zmarlych)
199 lzm2020m <- as.numeric(dane2020m$Liczba zmarlych)
200 lzm2020k <- as.numeric(dane2020k$Liczba zmarlych)
201
```

Rysunek 6 Zmiana typu danych kolumn "Liczba zmarłych" z każdego roku na numeric

Idąc dalej, powyższe kolumny złączyłem w jedną ramkę danych odpowiednio dla kobiet i mężczyzn oraz za pomocą polecenia rowmeans obliczyłem średnie wartości liczby zgonów z lat 2010-2020 dla każdego wieku.

```
203 ▶ ##### ŚREDNIA LICZBA ZMARLYCH 2010-2020 WZGLEDN WIEKU #####
204 ▶ #####
205
206
207
208 lzm <- data.frame(lzm2010m,
209                 lzm2011m,
210                 lzm2012m,
211                 lzm2013m,
212                 lzm2014m,
213                 lzm2015m,
214                 lzm2016m,
215                 lzm2017m,
216                 lzm2018m,
217                 lzm2019m,
218                 lzm2020m)
219 colnames(lzm) <- c("Liczba zmarlych 2010", "Liczba zmarlych 2011", "Liczba zmarlych 2012", "Liczba zmarlych 2013",
220                  "Liczba zmarlych 2014", "Liczba zmarlych 2015", "Liczba zmarlych 2016", "Liczba zmarlych 2017",
221                  "Liczba zmarlych 2018", "Liczba zmarlych 2019", "Liczba zmarlych 2020")
222 avg1zm <- rowMeans(lzm)
223 srednia_zmarlych_mezczyzn_2010_do_2020_wzgledn_wieku <- data.frame(Plec=dane2010m$Plec, Wiek=dane2010m$Wiek, Srednia_zmarlych_2010_2020=avg1zm)
224
225
226 lzk <- data.frame(lzk2010k,
227                 lzk2011k,
228                 lzk2012k,
229                 lzk2013k,
230                 lzk2014k,
231                 lzk2015k,
232                 lzk2016k,
233                 lzk2017k,
234                 lzk2018k,
235                 lzk2019k,
236                 lzk2020k)
237 colnames(lzk) <- c("Liczba zmarlych 2010", "Liczba zmarlych 2011", "Liczba zmarlych 2012", "Liczba zmarlych 2013",
238                  "Liczba zmarlych 2014", "Liczba zmarlych 2015", "Liczba zmarlych 2016", "Liczba zmarlych 2017",
239                  "Liczba zmarlych 2018", "Liczba zmarlych 2019", "Liczba zmarlych 2020")
240 srednia_zmarlych_kobiet_2010_do_2020_wzgledn_wieku <- data.frame(Plec=dane2010k$Plec, Wiek=dane2010k$Wiek, Srednia_zmarlych_2010_2020=avg1zk)
241
242 ▶
```

Rysunek 7 Średnie liczby zgonów z lat 2010-2020

Finalnie dane są podzielone ze względu na płeć, tym samym otrzymujemy dwie tabele posiadające kolumny wyżej wspomnianej płci, wieku oraz średniej liczby zgonów w latach 2010-2020

	Plec	Wiek	Srednia_zmarlych_2010_2020
1	K	0	384.454545
2	K	1	24.363636
3	K	2	16.727273
4	K	3	11.636364
5	K	4	9.545455
6	K	5	8.727273
7	K	6	8.363636
8	K	7	8.363636
9	K	8	8.272727
10	K	9	8.363636
11	K	10	8.636364
12	K	11	9.363636
13	K	12	10.727273
14	K	13	12.545455
15	K	14	14.909091
16	K	15	17.272727
17	K	16	19.818182
18	K	17	22.181818
19	K	18	23.727273
20	K	19	24.272727
21	K	20	24.272727
22	K	21	23.545455
23	K	22	23.272727
24	K	23	22.909091
25	K	24	23.181818
26	K	25	23.818182

Rysunek 8 Dane po modyfikacjach – kobiety

	Plec	Wiek	Srednia_zmarlych_2010_2020
1	M	0	457.72727
2	M	1	28.81818
3	M	2	20.54545
4	M	3	15.09091
5	M	4	12.18182
6	M	5	10.63636
7	M	6	10.36364
8	M	7	10.36364
9	M	8	10.54545
10	M	9	10.36364
11	M	10	10.90909
12	M	11	11.54545
13	M	12	12.72727
14	M	13	15.72727
15	M	14	20.18182
16	M	15	28.09091
17	M	16	39.54545
18	M	17	53.36364
19	M	18	67.81818
20	M	19	80.00000
21	M	20	88.18182
22	M	21	92.54545
23	M	22	94.54545
24	M	23	95.72727
25	M	24	97.36364
26	M	25	99.54545

Rysunek 9 Dane po modyfikacjach - mężczyźni

Podstawowe wartości

1. Średnia.

Za pomocą polecenia `mean` obliczyłem średnią roczną liczbę zgonów.

```
246 > ##### SREDNIA #####
247 srednia_zmarlych_mezczyzn <- mean(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
248
249 srednia_zmarlych_kobiet <- mean(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
250
```

Rysunek 10 Kod Średniej

Wynik:

Kobiety:

```
> srednia_zmarlych_kobiet
[1] 974.0972
```

Rysunek 11 Średnia zmarłych kobiet w R

Mężczyźni:

```
> srednia_zmarlych_mezczyzn
[1] 985.0522
```

Rysunek 12 Średnia zmarłych mężczyzn w R

2. Moda.

W celu poprawnego zwrócenia wartości występującej najczęściej konieczne było napisanie funkcji przeszukującej zbiór wartości. Wbudowana funkcja `mode` zwracała wartość „vector”, nie działała zatem poprawnie

```
251 > ##### MODA #####
252
253 MODA <- function(v) {
254   uniqv <- unique(v)
255   uniqv[which.max(tabulate(match(v, uniqv)))]
256 }
257
258 moda_zmarlych_kobiet <- MODA(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
259
260 moda_zmarlych_mezczyzn <- MODA(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
261
```

Rysunek 13 Funkcja mody

Wynik:

Kobiety:

```
> moda_zmarlych_kobiet
[1] 8.363636
```

Rysunek 14 Moda liczby zgonów kobiet

Mężczyźni:

```
> moda_zmarlych_mezczyzn
[1] 10.36364
```

Rysunek 15 Moda liczby zgonów mężczyzn

3. Mediana

Za pomocą funkcji median obliczyłem medianę.

```
262- #### MEDIANA ####
263
264 mediana_zmarlych_kobiet <- median(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
265 srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku[srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku['Srednia_zmarlych_2010_2020'] == mediana_zmarlych_kobiet]
266
267 mediana_zmarlych_mezczyzn <- median(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
268 srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku[srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku['Srednia_zmarlych_2010_2020'] == mediana_zmarlych_mezczyzn]
```

Rysunek 16 Funkcja mediany w R

Wynik:

Kobiety:

```
> mediana_zmarlych_kobiet
[1] 284.6364
```

Rysunek 17 Mediana zgonów kobiet

Mężczyźni:

```
> mediana_zmarlych_mezczyzn
[1] 497.5455
```

Rysunek 18 Mediana zgonów mężczyzn

4. Odchylenie standardowe od średniej.

Za pomocą funkcji sd obliczyłem odchylenie standardowe.

```
270- #### ODCHYLENIE STD OD ŚREDNIEJ ####
271
272 odchylenie_std_sredniej_liczby_zmarlych_kobiet <- sd(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
273
274 odchylenie_std_sredniej_liczby_zmarlych_mezczyzn <- sd(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
275
```

Rysunek 19 Funkcja odchylenia standardowego od średniej w R

Wynik:

Kobiety:

```
> odchylenie_std_sredniej_liczby_zmarlych_kobiet
[1] 1284.712
```

Rysunek 20 Odchylenie standardowe od średniej dla zgonów kobiet

Mężczyźni:

```
> odchylenie_std_sredniej_liczby_zmarlych_mezczyzn
[1] 1017.397
```

Rysunek 21 Odchylenie standardowe od średniej dla zgonów mężczyzn

5. Odchylenie standardowe od mediany.

```
276 ▾ #### ODCHYLENIE OD MEDIANY ####
277
278 odchylenie_od_mediany_k <- mad(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
279
280 odchylenie_od_mediany_m <- mad(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
281 |
```

Rysunek 22 Funkcja odchylenia standardowego od mediany

Za pomocą funkcji mad obliczyłem odchylenie standardowe od mediany

Wynik:

Kobiety:

```
> odchylenie_od_mediany_k
[1] 406.0976
```

Rysunek 23 Odchylenie standardowe od mediany dla zgonów kobiet

Mężczyźni:

```
> odchylenie_od_mediany_m
[1] 714.3436
```

Rysunek 24 Odchylenie standardowe od mediany dla zgonów mężczyzn

6. Kwantyle.

Za pomocą funkcji quantile obliczyłem kwantyle.

```
283 ▾ #### KWANTYLE #####
284 kwantyle__m <- quantile(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
285 kwantyle_k <- quantile(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
286
```

Rysunek 25 Funkcja kwantyli w R

Wyniki:

Kobiety:

```
> kwantyle_k
      0%      25%      50%      75%     100%
8.272727 24.454545 284.636364 1550.090909 4082.636364
```

Rysunek 26 Kwantyle zgonów kobiet

Mężczyźni:

```
> kwantyle__m
      0%      25%      50%      75%     100%
10.36364 102.36364 497.54545 1866.63636 2959.18182
```

Rysunek 27 Kwantyle zgonów mężczyzn

7. Odchylenie ćwiartkowe

Przy pomocy funkcji IQR obliczyłem odchylenie ćwiartkowe

```
288 ##### ODCHYLENIE CWIARTKOWE #####
289 odchylenie_cwiartkowe_m <- IQR(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
290 odchylenie_cwiartkowe_k <- IQR(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020)
291
```

Rysunek 28 Kod odchylenia ćwiartkowego

Wyniki:

Kobiety:

```
> odchylenie_cwiartkowe_k
[1] 1525.636
```

Rysunek 29 Odchylenie ćwiartkowe dla zgonów kobiet

Mężczyźni:

```
> odchylenie_cwiartkowe_m
[1] 1764.273
```

Rysunek 30 Odchylenie ćwiartkowe dla zgonów mężczyzn

8. Współczynnik zmienności

W celu obliczenia współczynnika zmienności załadowałem bibliotekę formattable aby przedstawić wynik w procentach. Aby wyznaczyć współczynnik zmienności skorzystałem ze wzoru:

$$cv = \frac{s}{\bar{x}} \cdot 100\%,$$

gdzie:

cv – współczynnik zmienności,

S – odchylenie standardowe od średniej,

\bar{x} – średnia

```
294 #install.packages("formattable")
295 library(formattable) ## biblioteka wczytana w celu przedstawienia danych w % ##
296 wsp_zmienosci_k <- percent(sd(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020) /
297 mean(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020))
298 wsp_zmienosci_m <- percent(sd(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020) /
299 mean(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020))
300
```

Rysunek 31 Kod do obliczenia współczynnika zmienności

Wynik:

Kobiety:

```
> wsp_zmienosci_k
[1] 131.89%
```

Rysunek 32 Współczynnik zmienności dla zgonów kobiet

Mężczyźni:

```
> wsp_zmienosci_m
[1] 103.28%
```

Rysunek 33 Współczynnik zmienności dla zgonów mężczyzn

9. Wskaźnik asymetrii

W celu obliczenia wskaźnika asymetrii wykorzystałem wzór:

$$A_s = \frac{\bar{x} - D}{S}$$

gdzie:

A_s – wskaźnik asymetrii,

D – dominanta,

\bar{x} – średnia,

S – odchylenie standardowe od średniej

```
#### WSKAZNIK ASYMETRII ####
```

```
wa_k <- (srednia_zmarlych_kobiet-moda_zmarlych_kobiet)-moda_zmarlych_kobiet/odchylenie_std_sredniej_liczby_zmarlych_kobiet
```

```
wa_m <- (srednia_zmarlych_mezczyzn-moda_zmarlych_mezczyzn)-moda_zmarlych_mezczyzn/odchylenie_std_sredniej_liczby_zmarlych_mezczyzn
```

Rysunek 34 Kod zwracający wartość współczynnika asymetrii

Wynik:

Kobiety:

```
> wa_k  
[1] 965.7271
```

Rysunek 35 Współczynnik asymetrii dla zgonów kobiet

Mężczyźni:

```
> wa_m  
[1] 974.6784
```

Rysunek 36 Współczynnik asymetrii dla zgonów mężczyzn

Graficzna reprezentacja danych

Komendy:

col() – pozwala wybrać kolor słupków

cex.axis – odstęp podpisów osi X

cex.names – wielkość czcionki podpisów osi X

legend – tworzy legendę

abline – wyznacza linie na wykresie (w tym przypadku średnią)

labels - podpisy

labelcex = rozmiar podpisów

radius – wysokość podpisów

theta – kąt nachylenia wykresu

explode – szerokość pomiędzy wartościami wykresu

main – tytuł wykresu

ylab – podpis osi Y

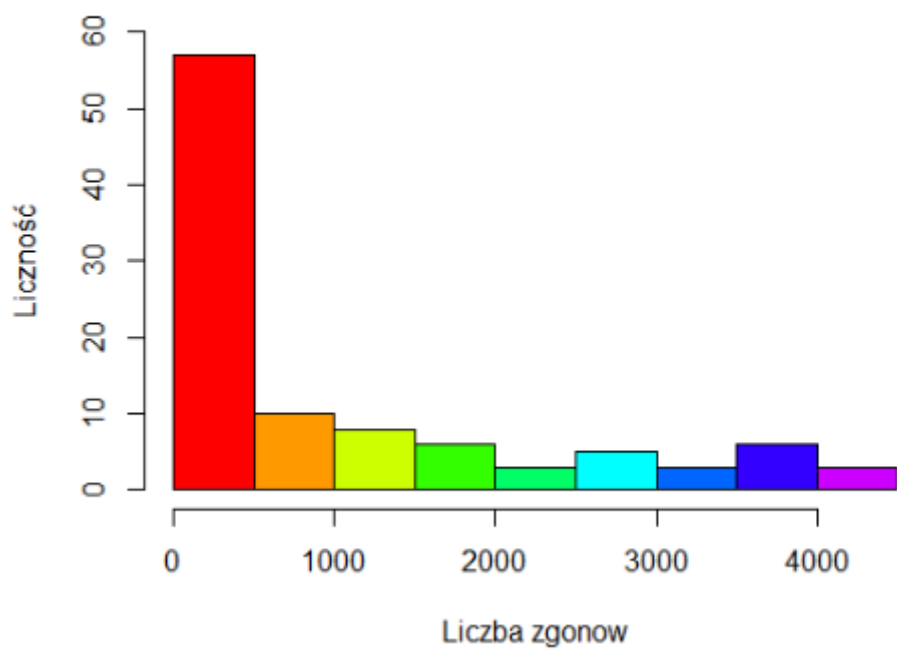
xlabel – podpis osi X

1. Histogramy

```
312 # histogram średniej liczby zgonów kobiet w latach 2010-2020
313
314 hist_k <- hist(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020,
315               main="Histogram średniej liczby zgonów kobiet 2010-2020",
316               ylab="Liczność",
317               xlab="Liczba zgonów",
318               col=rainbow(10),
319               ylim=c(0,60)
320 )
321
322 # histogram średniej liczby zgonów mężczyzn w latach 2010-2020
323
324 hist_m <- hist(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020,
325               main="Histogram średniej liczby zgonów mężczyzn 2010-2020",
326               ylab="Liczność",
327               xlab="Liczba zgonów",
328               col=rainbow(6),
329               ylim=c(0,55)
330 )
331
332
```

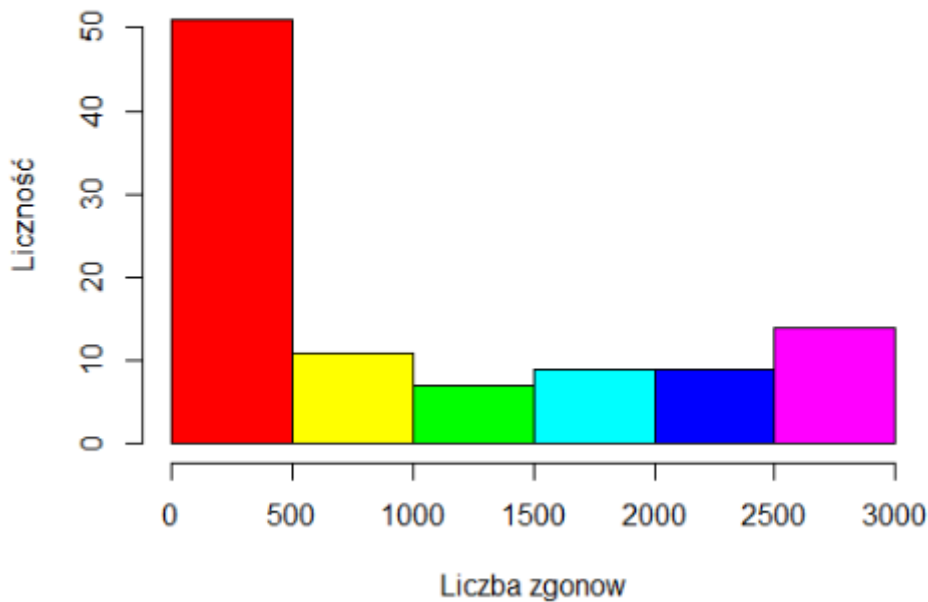
Rysunek 37 Kod tworzący histogramy dla liczby zgonów kobiet i mężczyzn

Histogram sredniej liczby zgonow kobiet 2010-2020



Rysunek 38 Histogram dla zgonów kobiet

Histogram średniej liczby zgonów mężczyzn 2010-2020



Rysunek 39 Histogram dla zgonów mężczyzn

2. Wykresy słupkowe

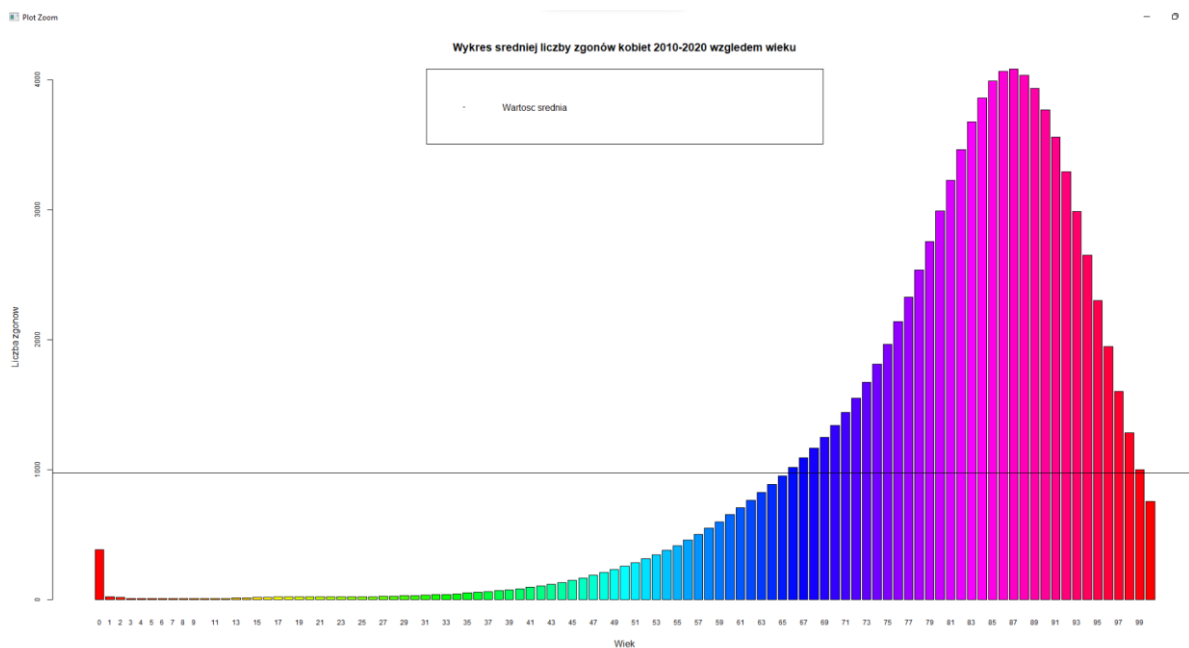
Biblioteka ggplot2 umożliwia tworzenie różnego rodzaju wykresów, w tym wykresów słupkowych dlatego też postanowiłem wykorzystać ją do realizacji mojego projektu. Wykresy słupkowe przedstawiają średnią liczbę zgonów w okresie 2010-2020 dla wieku od 0-100. Czarna linia na wykresie ukazuje wartość średnią obliczoną we wcześniejszym etapie projektu.

```

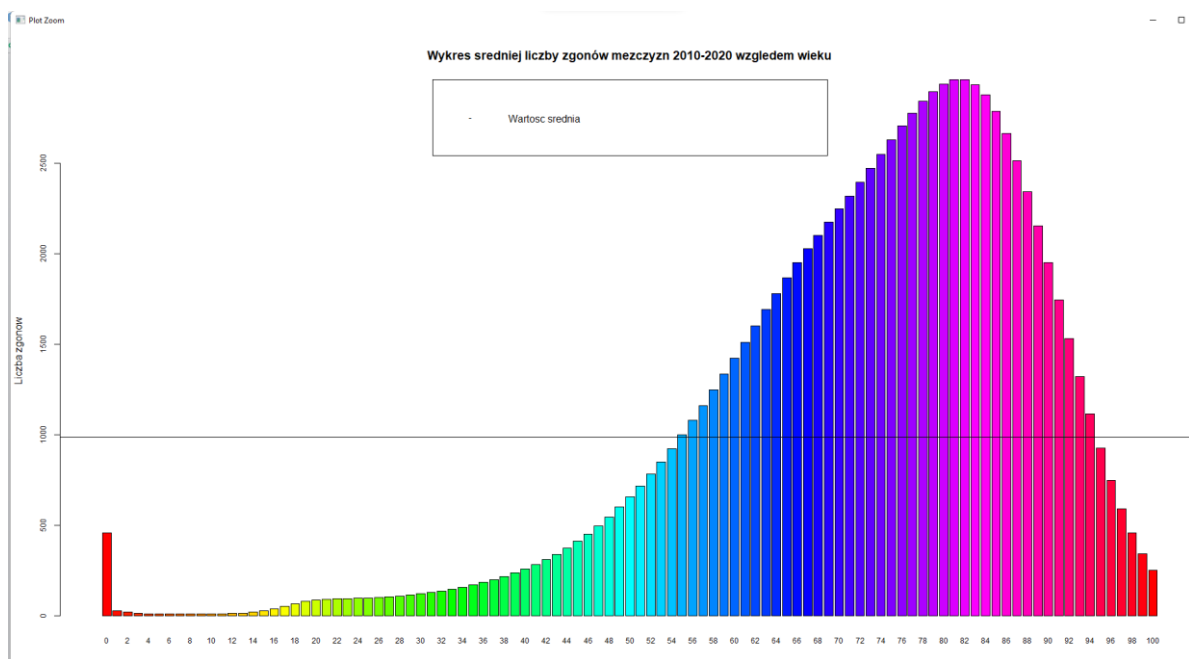
334 # wykresy słupkowy
335
336 library(ggplot2) ## biblioteka pozwalająca na utworzenie wykresow
337
338 # LICZBA ZGONOW KOBIET WZGLEDEM WIEKU #
339
340 wykres_slup_k <- barplot(srednia_zmarlych_kobiet_2010_do_2020_wzgleciem_wieku$Srednia_zmarlych_2010_2020,
341   names.arg = srednia_zmarlych_kobiet_2010_do_2020_wzgleciem_wieku$Wiek,
342   col=rainbow(100),
343   cex.axis=0.75,
344   cex.names=0.75,
345   ylab="Liczba zgonow",
346   xlab="Wiek",
347   main = "Wykres sredniej liczby zgonow kobiet 2010-2020 wzgleciem wieku",
348 )
349 abline(h=srednia_zmarlych_kobiet, col="black")
350 legend("top", legend = "Wartosc srednia", pch = "-", col = "black")
351
352 # LICZBA ZGONOW MEZCZYZN WZGLEDEM WIEKU #
353
354 barplot(srednia_zmarlych_mezczyzn_2010_do_2020_wzgleciem_wieku$Srednia_zmarlych_2010_2020,
355   names.arg = srednia_zmarlych_mezczyzn_2010_do_2020_wzgleciem_wieku$Wiek,
356   col=rainbow(100),
357   cex.axis=0.75,
358   cex.names=0.75,
359   ylab="Liczba zgonow",
360   xlab="Wiek",
361   main = "Wykres sredniej liczby zgonow mezczyzn 2010-2020 wzgleciem wieku",
362 )
363 abline(h=srednia_zmarlych_mezczyzn, col="black")
364 legend("top", legend = "Wartosc srednia", pch = "-", col = "black")
365

```

Rysunek 40 Kod tworzący wykresy słupkowe ukazujące liczbe zgonów w danym wieku



Rysunek 41 Wykres słupkowy przedstawiający średnią liczbę zgonów kobiet w poszczególnym wieku wraz z wartością średnią



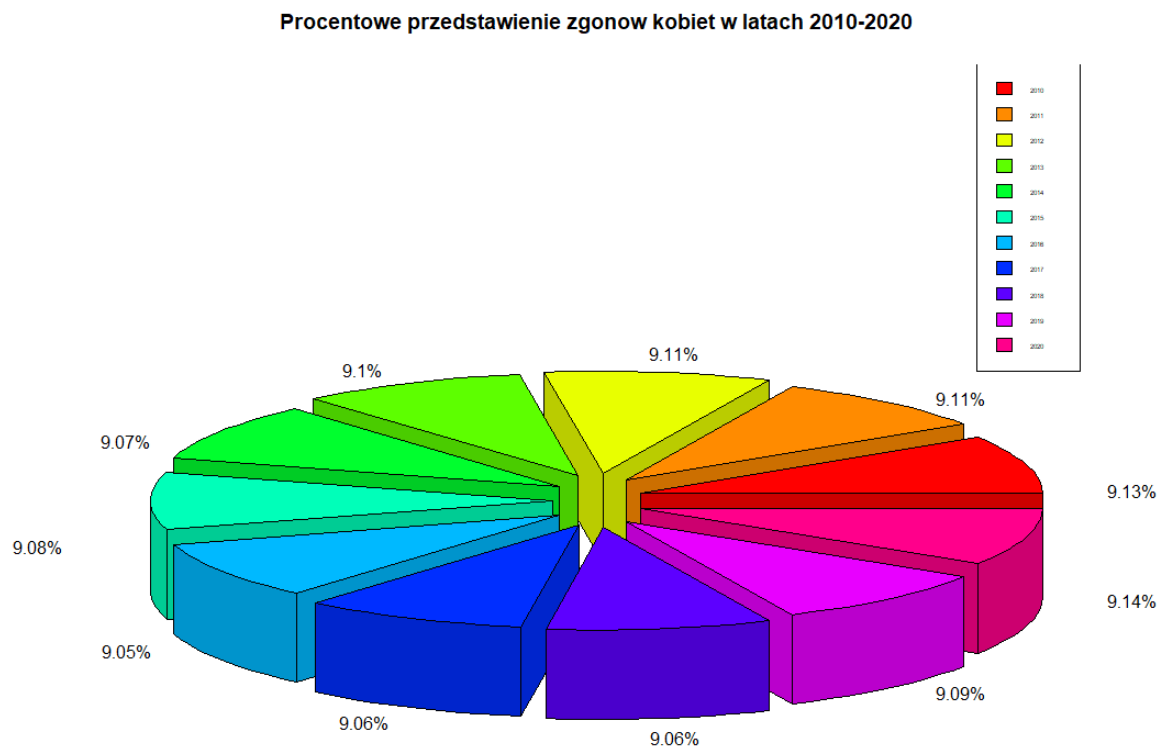
Rysunek 42 Wykres słupkowy przedstawiający średnią liczbę zgonów mężczyzn w poszczególnym wieku wraz z wartością średnią

3. Wykresy kołowe

Wykresy kołowe zostały przygotowane z wykorzystaniem biblioteki plotrix, która umożliwia wiele modyfikacji w wykresach m.in. dostosować ich kąt nachylenia, kształt, skalę lub wielkości podpisów. Utworzone w projekcie wykresy kołowe ukazują procentową wartość zgonów rocznych na przestrzeni lat 2010-2020.

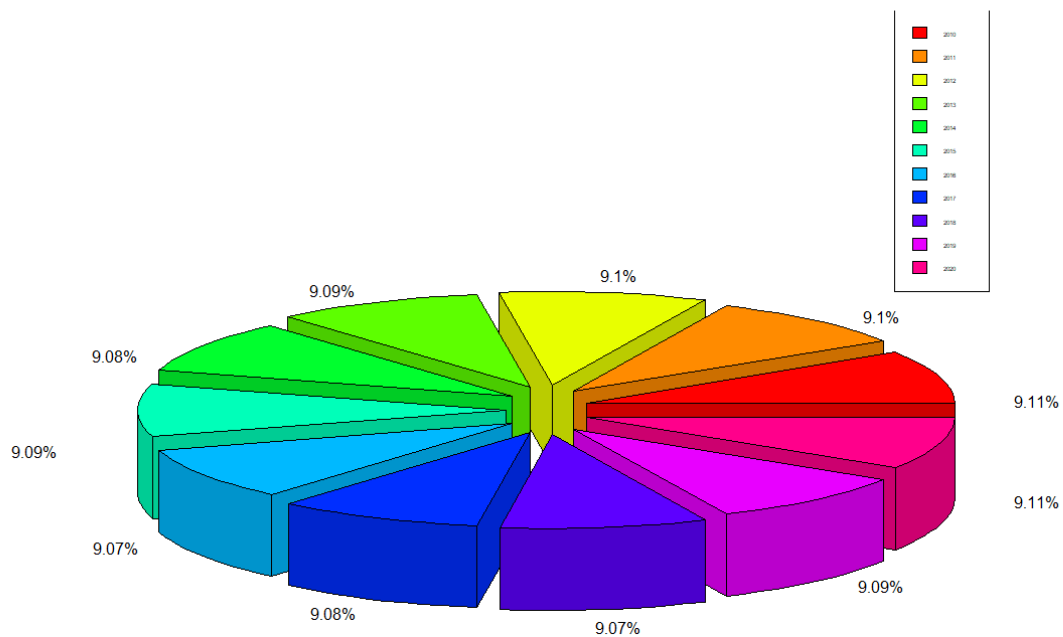
```
371 ##### PROCENTOWE PRZEDSATWIENIE ZGONOW KOBIECI W LATACH 2010-2020 / WYKRES KOŁOWY #####
372
373 lzk
374 zmarli_2010_2020_k <- colMeans(lzk)
375
376 pie_labels_k <- paste0(round(100 * zmarli_2010_2020_k/sum(zmarli_2010_2020_k), 2), "%")
377
378 procent_kobiety <- pie3D(zmarli_2010_2020_k, labels = pie_labels_k, labelcex=1, radius=0.9, theta=0.4, explode= 0.1,
379   main = "Procentowe przedstawienie zgonow kobiet w latach 2010-2020", col = rainbow(length(zmarli_2010_2020_k)))
380 legend("topright", c("2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020"), cex = 0.38,
381   fill = rainbow(length(zmarli_2010_2020_k)))
382
383 ##### PROCENTOWE PRZEDSATWIENIE ZGONOW MEZCZYZN W LATACH 2010-2020 / WYKRES KOŁOWY #####
384
385 lzm
386 zmarli_2010_2020_m <- colMeans(lzm)
387
388 pie_labels_m <- paste0(round(100 * zmarli_2010_2020_m/sum(zmarli_2010_2020_m), 2), "%")
389
390 procent_mezczyzni <- pie3D(zmarli_2010_2020_m, labels = pie_labels_m, labelcex=1, radius=0.9, theta=0.4, explode= 0.1,
391   main = "Procentowe przedstawienie zgonow mezczyzn w latach 2010-2020", col = rainbow(length(zmarli_2010_2020_m)))
392 legend("topright", c("2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020"), cex = 0.38,
393   fill = rainbow(length(zmarli_2010_2020_m)))
394
395
396
```

Rysunek 43 Kod tworzący wykresy kołowe ukazujących procentową wartość zgonów 2010-2020



Rysunek 44 Wykres kołowy przedstawiający procentowy wykres zgonów kobiet na przestrzeni lat 2010-2020

Procentowe przedstawienie zgonów mężczyzn w latach 2010-2020



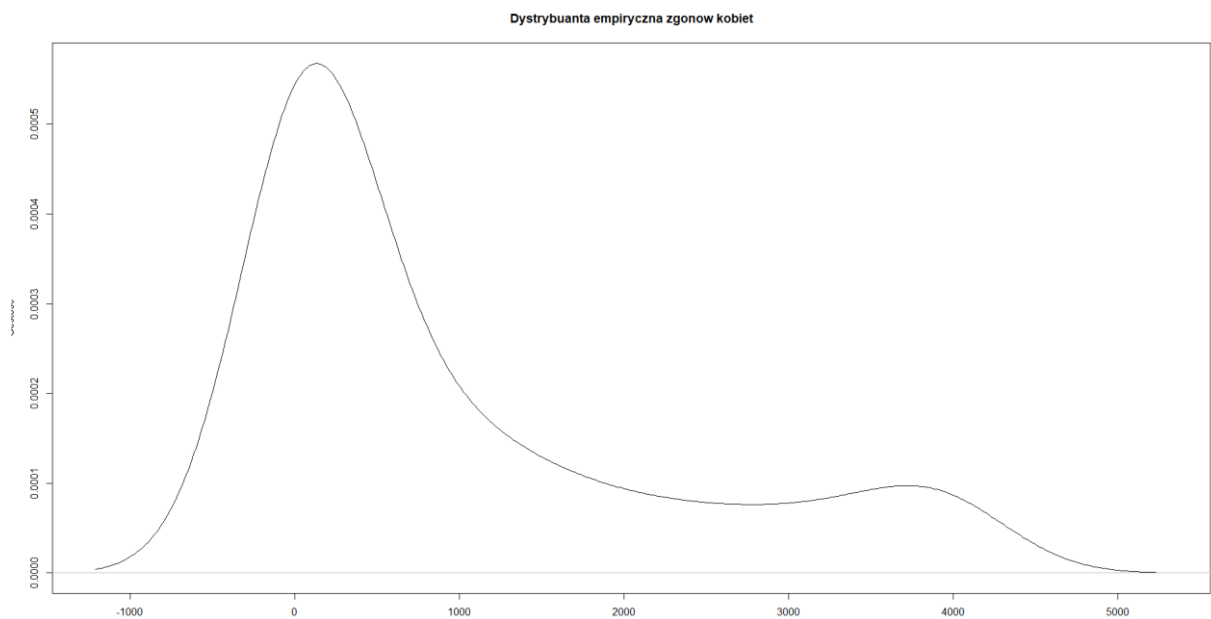
Rysunek 45 Wykres kołowy przedstawiający procentowy wykres zgonów mężczyzn na przestrzeni lat 2010-2020

4. Graficzna prezentacja dystrybuant empirycznych

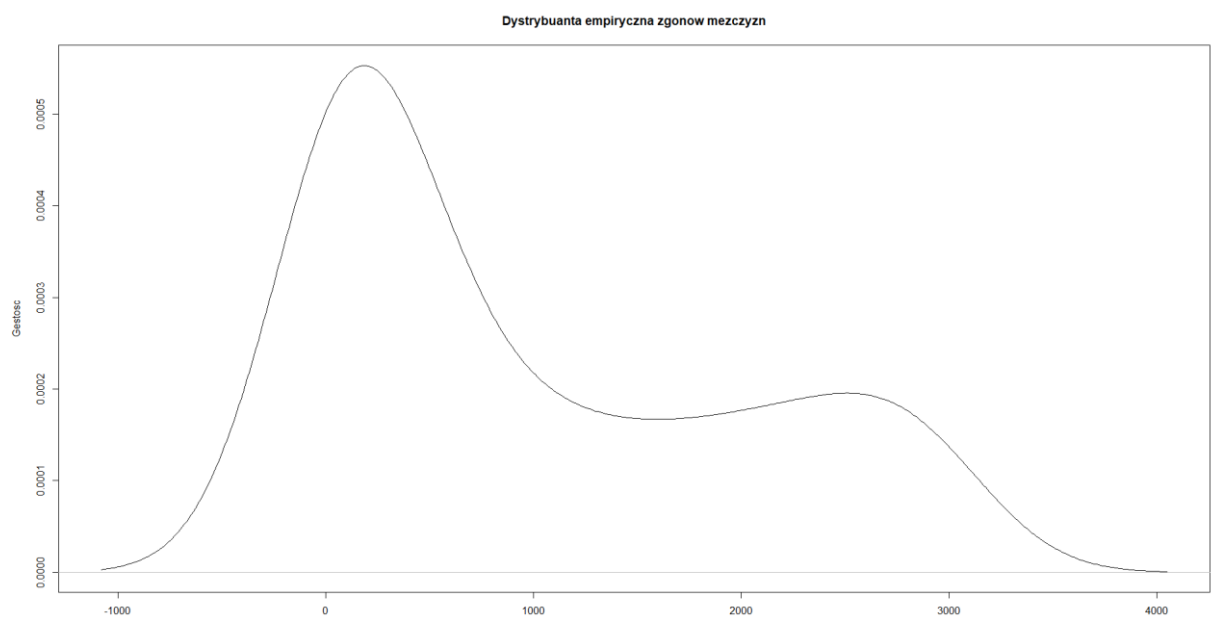
Dystrybuanty zostały wykonane przy pomocy domyślnie wbudowanych funkcji w środowisku R a mianowicie mowa tu o `plot` oraz `density`, gdzie `density` wyznacza dystrybuantę a `plot` tworzy wykres.

```
399  
400- ### DYSTRYBUANTA ZGONÓW KOBIEŃ ###  
401 dystrybuanta_empiryczna_k <- plot(density(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020),  
402 ylab= "Gestosc",  
403 main = "Dystrybuanta empiryczna zgonow kobiet")  
404  
405  
406- ### DYSTRYBUANTA ZGONÓW MEZCZYŹN ###  
407 dystrybuanta_empiryczna_M <- plot(density(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020),  
408 ylab= "Gestosc",  
409 main = "Dystrybuanta empiryczna zgonow mezczyzn")  
410
```

Rysunek 46 Kod tworzący dystrybuanty dla średnich zgonów kobiet i mężczyzn w latach 2010-2020



Rysunek 47 Dystrybuanta empiryczna średniej liczby zgonów kobiet



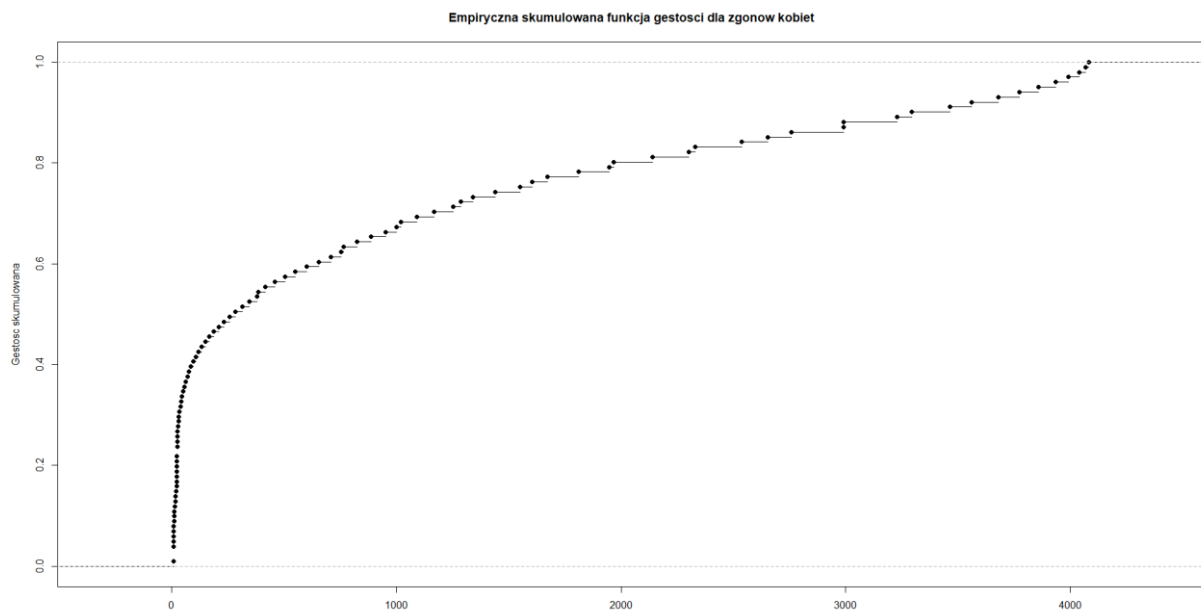
Rysunek 48 Dystrybuanta empiryczna średniej liczby zgonów mężczyzn

5. Wykresy empirycznej skumulowanej funkcji gęstości

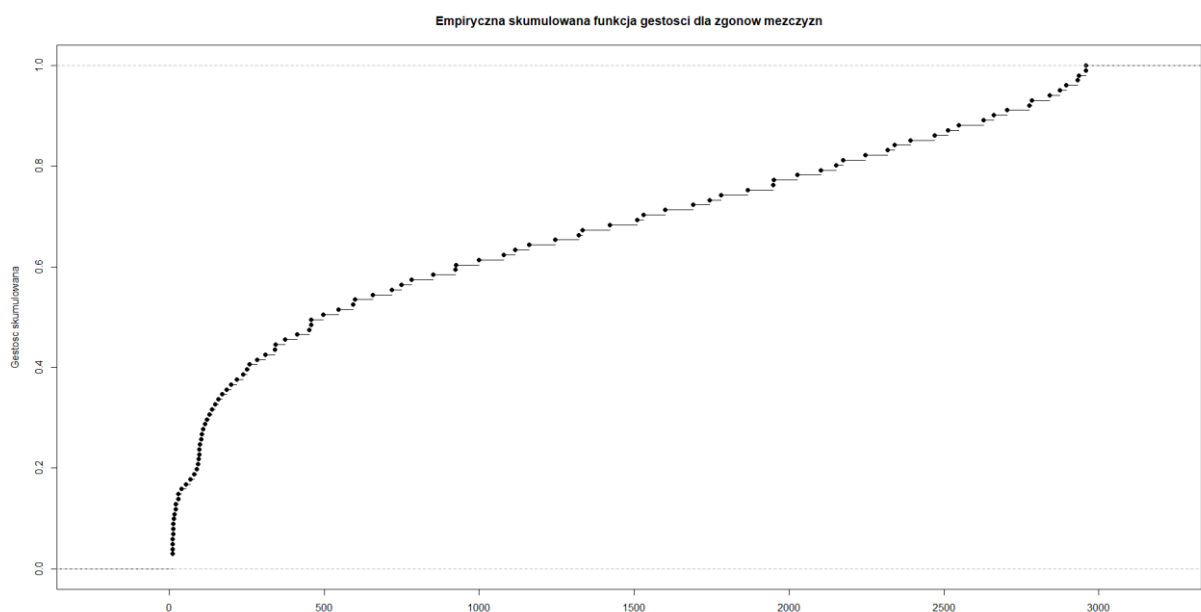
Podobnie jak dystrybuenta empiryczna, wykres skumulowanej funkcji gęstości został utworzony przy wykorzystaniu wbudowanych domyślnie funkcji w środowisku R z tą różnicą że funkcję density zastępuje funkcja ecdf

```
414 ##### EMPIRYCZNA SKUMULOWANA FUNKCJA GESTOSCI DLA ZGONOW KOBIET #####
415 skumulowana_gestosc_k <- plot(ecdf(srednia_zmarlych_kobiet_2010_do_2020_wzglede_wieku$Srednia_zmarlych_2010_2020),
416   ylab= "Gestosc skumulowana",
417   main= "Empiryczna skumulowana funkcja gestosci dla zgonow kobiet",
418   xlab=" ")
419
420 ##### EMPIRYCZNA SKUMULOWANA FUNKCJA GESTOSCI DLA ZGONOW MEZCZYZN #####
421 skumulowana_gestosc_m <- plot(ecdf(srednia_zmarlych_mezczyzn_2010_do_2020_wzglede_wieku$Srednia_zmarlych_2010_2020),
422   ylab= "Gestosc skumulowana",
423   main= "Empiryczna skumulowana funkcja gestosci dla zgonow mezczyzn",
424   xlab=" ")
425
```

Rysunek 49 Kod wykresów empirycznych skumulowanych funkcji gęstości



Rysunek 50 Wykres empirycznej skumulowanej funkcji gęstości dla średniej liczby zgonów kobiet 2010-2020



Rysunek 51 Wykres empirycznej skumulowanej funkcji gęstości dla średniej liczby zgonów mężczyzn 2010-2020

Hipotezy

1. Czy średnia ilość zgonów kobiet na przestrzeni lat 2010-2020 jest równa 1000?

```
430- ### HIPOTEZA : CZY ŚREDNIA IŁOŚĆ ZGONÓW KOBIET NA PRZESTRZENI LAT 2010-2020 JEST RÓWNA 1000 ###
431-
432- hip1 <- t.test(srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020,mu=1000) # NIE MA PODSTAW ZEBY ODRZUCIC HIPOTEZE ZEROWA
433-
434-
```

Rysunek 52 Kod hipotezy nr 1

```
One Sample t-test

data: srednia_zmarlych_kobiet_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020
t = -0.20263, df = 100, p-value = 0.8398
alternative hypothesis: true mean is not equal to 1000
95 percent confidence interval:
 720.479 1227.715
sample estimates:
mean of x
 974.0972
```

Rysunek 53 Wynik hipotezy nr 1

$p > \alpha$

Nie ma podstaw do odrzucenia hipotezy zerowej.

2. Czy średnia ilość zgonów mężczyzn na przestrzeni lat 2010-2020 jest większa od 1000?

```
### HIPOTEZA : CZY ŚREDNIA IŁOŚĆ ZGONÓW MEZCZYŻYN JEST WIEKSZA OD 1000
hip2 <- t.test(srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020,mu=1500, alternative = "less") ## ODRZUCAMY HIPOTEZE ZEROWA NA KORZYSC ALTERNATYWNEJ
|
```

Rysunek 54 Kod hipotezy nr 2

```
One Sample t-test

data: srednia_zmarlych_mezczyzn_2010_do_2020_wzgledem_wieku$Srednia_zmarlych_2010_2020
t = -5.0867, df = 100, p-value = 0.0000008538
alternative hypothesis: true mean is less than 1500
95 percent confidence interval:
 -Inf 1153.126
sample estimates:
mean of x
 985.0522
```

Rysunek 55 Wynik hipotezy nr 2

$p < \alpha$

Odrzucamy hipotezę zerową na korzyść alternatywnej