

Projekt zaliczenowy 1 - pierwszy raport statystyczny

Denys Vitiuk

Celem zadania jest statystyczna analiza danych znajdujących się w pliku people

1. 1)Ilość obserwacji jest równa 499

 data	499 obs. of 10 variables
--	--------------------------

Zmienny ilościowe:wiek,waga,wzrost,liczba_dzieci,wydatki,wydatki_zywnosc,oszczednosci Więc,7
Jakościowe:plec,budynek,stan_cywilny.Więc,3 W niektórych komórkach kolumny 'plec' nie ma danych, stąd występują braki danych

Ten zbiór danych opisuje wydatki i oszczędności osoby w zależności od wieku, płci, wagi, wzrostu, liczby dzieci, domu, czy osoba jest rozwiedziona(zamężna).

2. Najpierw dodaliśmy biblioteki,które ułatwią nam pracę nad projektem

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(ggplot2)
```

```
library(fitdistrplus)
```

```
## Loading required package: MASS
```

```
## Loading required package: survival
```

A potem dodaliśmy nasz dataset i sprawdziliśmy cor.test

```
data <- read.csv("people.csv")
```

```
cor(data[,c("wiek", "waga", "wzrost", "wydatki", "wydatki_zywnosc", "oszczednosci")])
```

```
##          wiek      waga      wzrost  wydatki wydatki_zywnosc
## wiek      1.0000000 -0.06888485 -0.03957449  0.1836638      0.1813799
## waga     -0.06888485  1.00000000  0.69855112 -0.1163270     -0.1169035
## wzrost   -0.03957449  0.69855112  1.00000000 -0.1883054     -0.1898849
## wydatki    0.18366379 -0.11632696 -0.18830541  1.0000000      0.9993912
## wydatki_zywnosc 0.18137987 -0.11690354 -0.18988492  0.9993912      1.0000000
## oszczednosci  0.89297022  0.01955825  0.04145199 -0.1073919     -0.1099041
##
##          oszczednosci
## wiek      0.89297022
## waga      0.01955825
## wzrost    0.04145199
## wydatki   -0.10739191
## wydatki_zywnosc -0.10990406
## oszczednosci  1.00000000
```

Widzimy, że największą korelację mają wiek i wzrost, co wynika z faktu, że osoby starsze zazwyczaj są wyższe niż młodsze osoby. Współczynnik korelacji między wiekiem a oszczędnościami jest również dość wysoki, co może oznaczać, że starsze osoby zazwyczaj mają większe oszczędności niż młodsze osoby.

Coś, zrobimy niektóre działania, żeby sprawdzić zależności pomiędzy zmiennymi

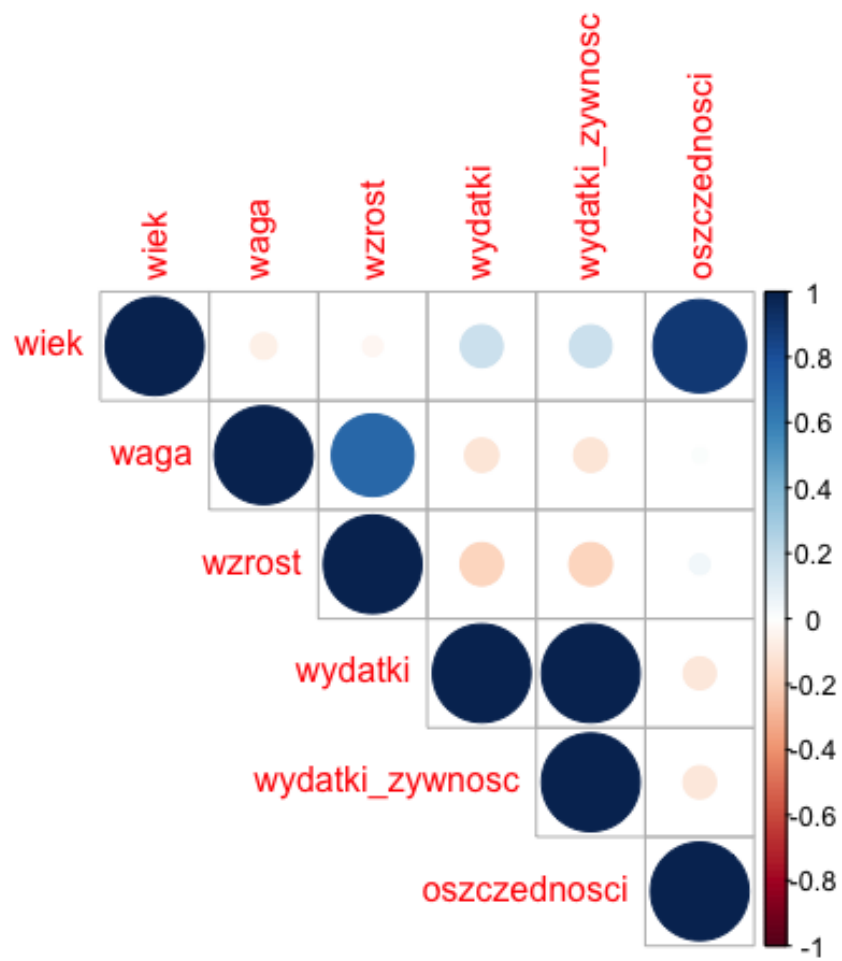


Figure 1: cortest

zastąpienie NA wartością “Nieznana” w kolumnie “plec”

```
data$plec[is.na(data$plec)] <- "Nieznana"
```

wyodrębnienie zmiennych jakościowych

```
zmienne_jakosciowe <- data[, c("plec", "stan_cywilny", "budynek")]
```

wygenerowanie tabeli krzyżowej

```
tabela_krzyzowa <- table(zmienne_jakosciowe)
```

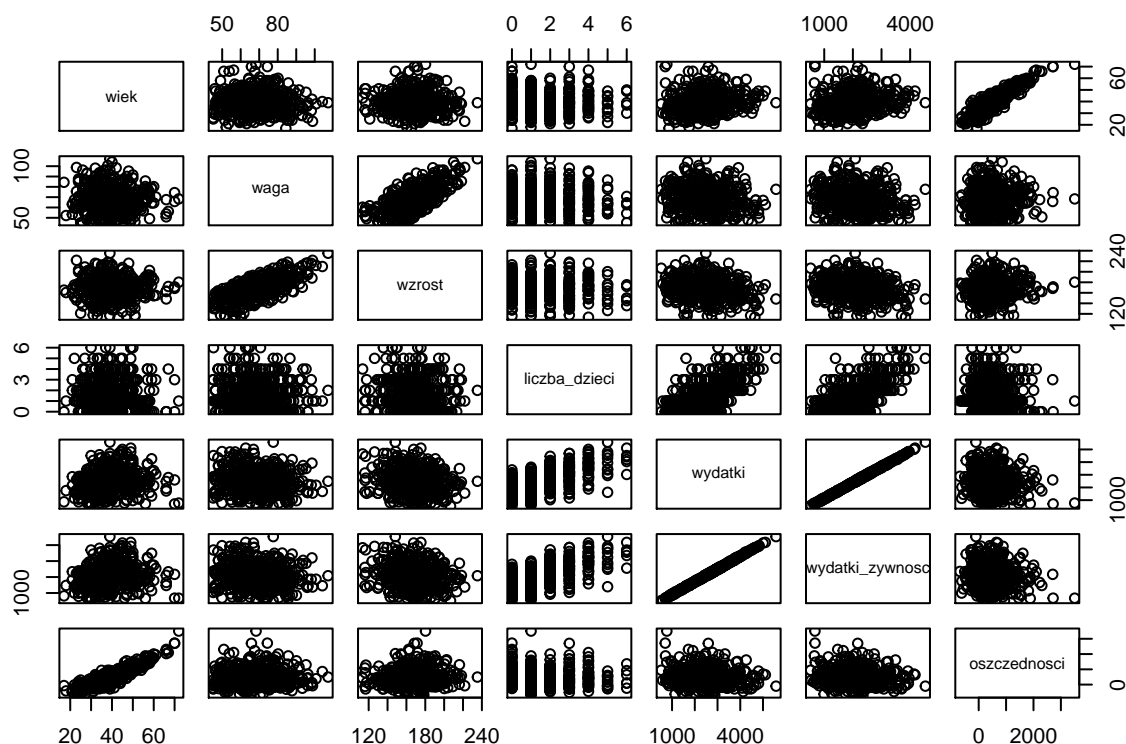
przeprowadzenie testu chi-kwadrat

```
chisq.test(tabela_krzyzowa)
```

wykres typu scatter-plot:

```
data_num <- data[, c("wiek", "waga", "wzrost", "liczba_dzieci", "wydatki", "wydatki_zywnosc", "oszczednosc")]
```

```
pairs(data_num)
```



```
ggplot(data, aes(x=wiek, y=wzrost, color=stan_cywilny)) +  
  geom_point(size=3) +  
  geom_smooth(method=lm, se=FALSE) +  
  labs(title="Zależność między wiekiem a wzrostem, z uwzględnieniem stanu cywilnego",  
        x="Wiek", y="Wzrost")
```

[illegible]

[illegible]

[illegible]

[illegible]

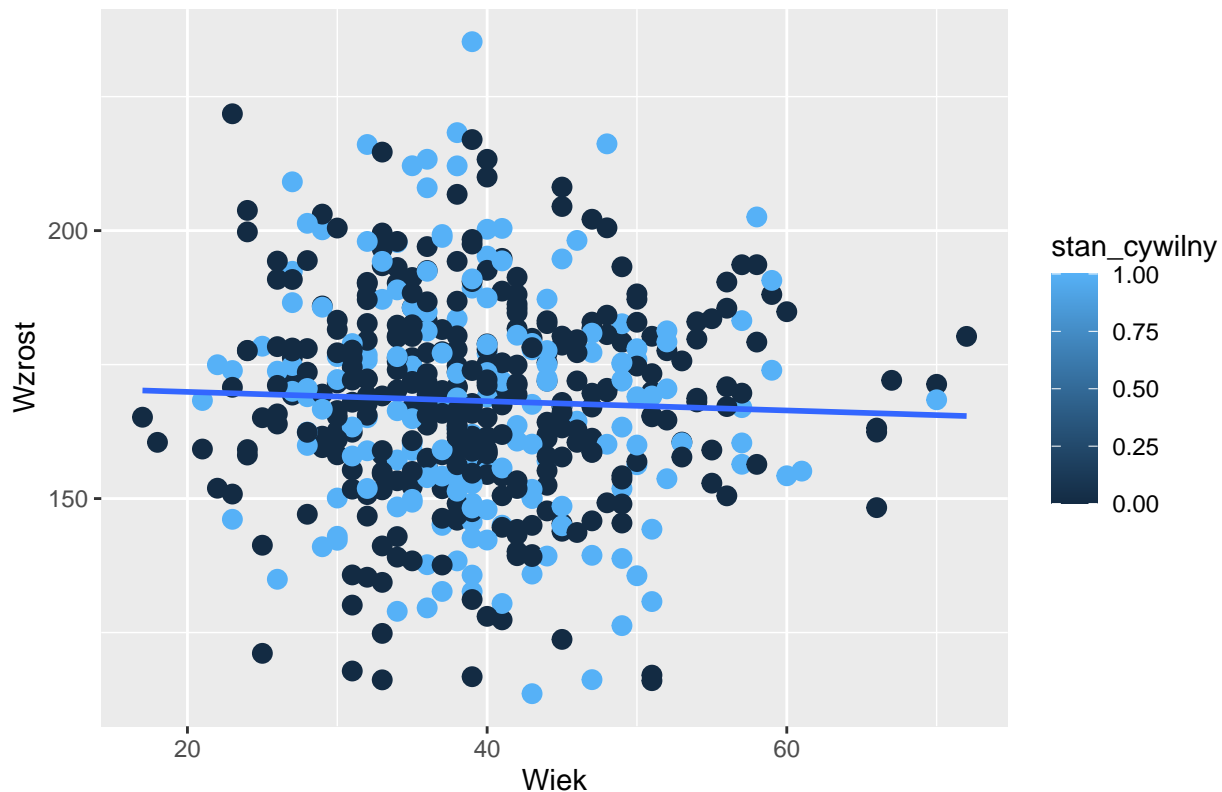

```
## conversion failure on 'Zależność między wiekiem a wzrostem, z uwzględnieniem
## stanu cywilnego' in 'mbsToSbcs': dot substituted for <c4>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wzrostem, z uwzględnieniem
## stanu cywilnego' in 'mbsToSbcs': dot substituted for <99>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wzrostem, z uwzględnieniem
## stanu cywilnego' in 'mbsToSbcs': dot substituted for <c4>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wzrostem, z uwzględnieniem
## stanu cywilnego' in 'mbsToSbcs': dot substituted for <99>
```

Zale..no.... mi..dzy wiekiem a wzrostem, z uwzgl..dnieniem stanu cywilnegc



```
ggplot(data, aes(x=stan_cywilny, y=waga)) +
  geom_boxplot(fill="lightblue", color="darkblue") +
  labs(title="Rozkład wagi w zależności od stanu cywilnego",
        x="Stan cywilny", y="Waga")
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbsToSbcs': dot substituted for <82>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
```

[illegible]

```

## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <82>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <bc>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <82>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <82>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <bc>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <82>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <bc>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

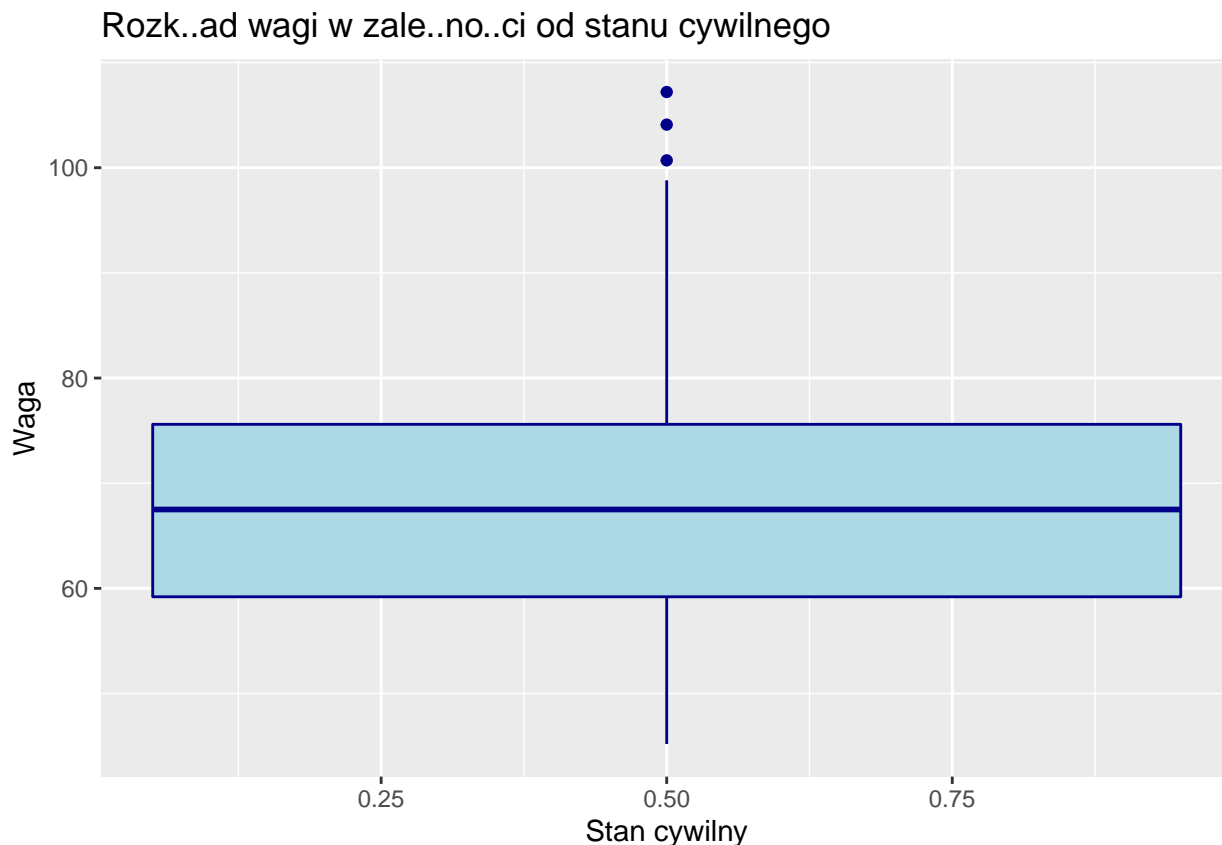
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <82>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Rozkład wagi w zależności od stanu cywilnego' in
## 'mbcsToSbcs': dot substituted for <9b>
```



Wykres przedstawia rozkład wagi w zależności od stanu cywilnego. Możemy zauważyć, że średnia waga jest nieco wyższa dla osób w związku małżeńskim i znacznie niższa dla osób rozwiedzionych.

```
ggplot(data, aes(x="", fill=budynek)) +
  geom_bar(width=1, color="white") +
  coord_polar("y", start=0) +
  labs(title="Udział poszczególnych typów budynków",
        fill="Typ budynku") +
  theme_void()
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':
## dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':
## dot substituted for <82>

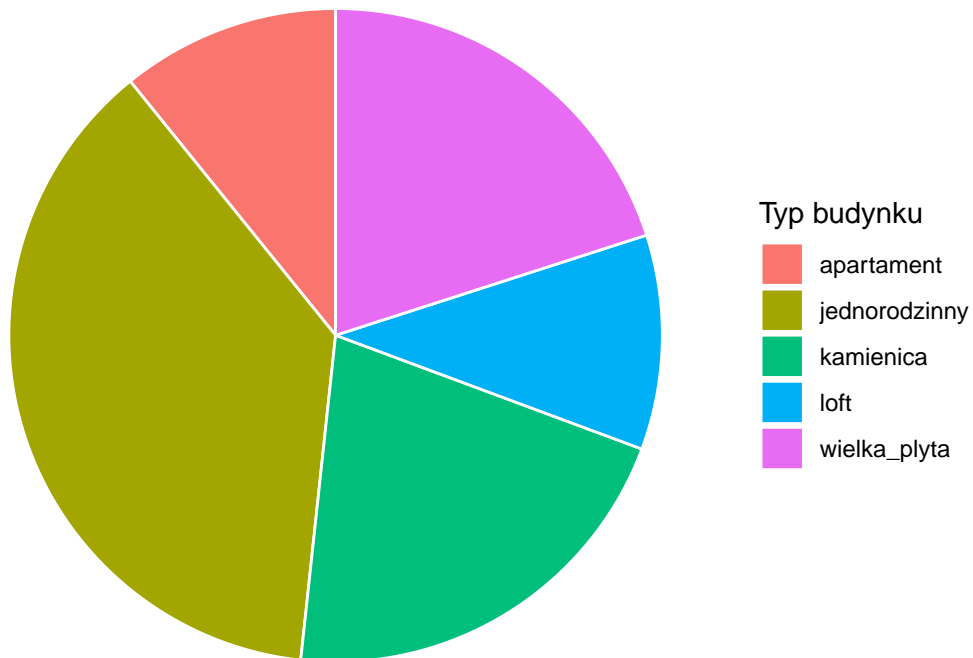
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':
## dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':
## dot substituted for <82>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':
## dot substituted for <c5>
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':  
## dot substituted for <82>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':  
## dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':  
## dot substituted for <82>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':  
## dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':  
## dot substituted for <82>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':  
## dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':  
## dot substituted for <82>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':  
## dot substituted for <c5>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on 'Udział poszczególnych typów budynków' in 'mbcsToSbcs':  
## dot substituted for <82>
```

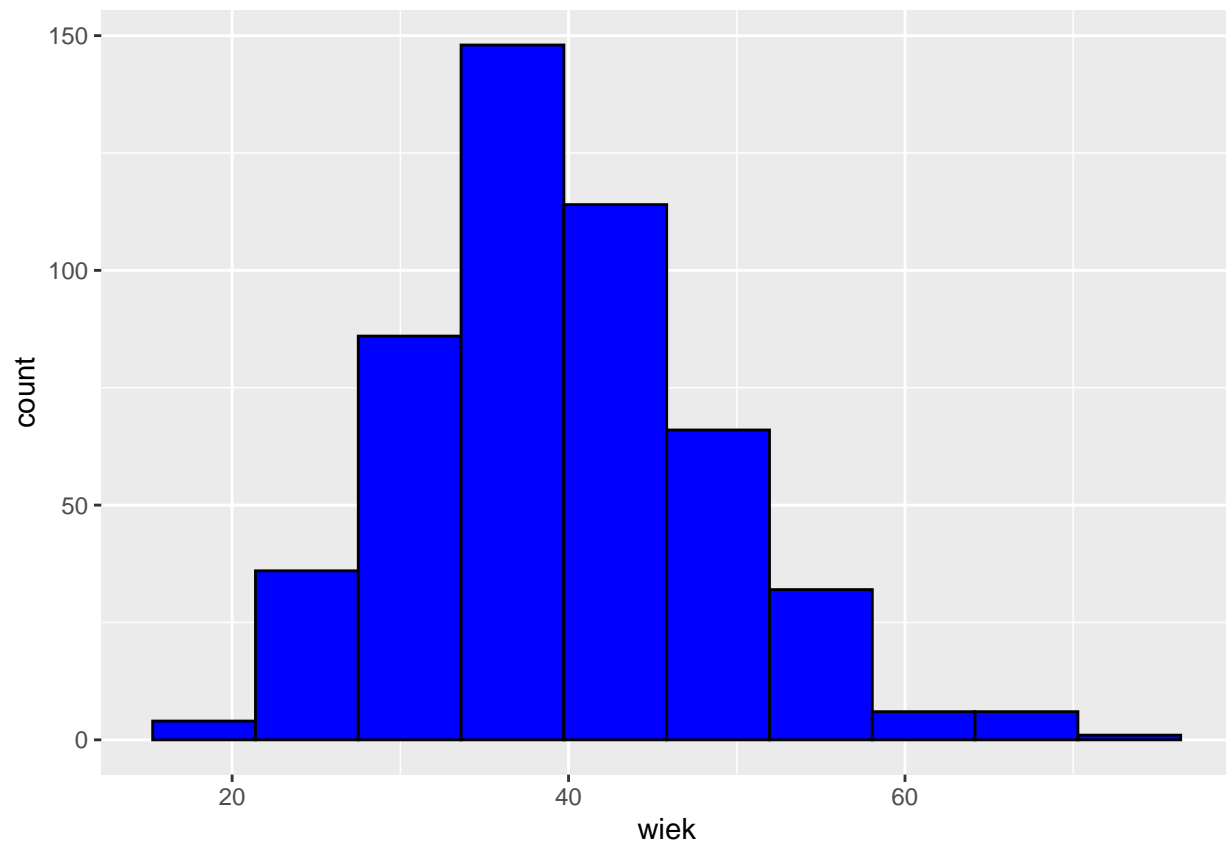
Udzia.. poszczególnych typów budynków



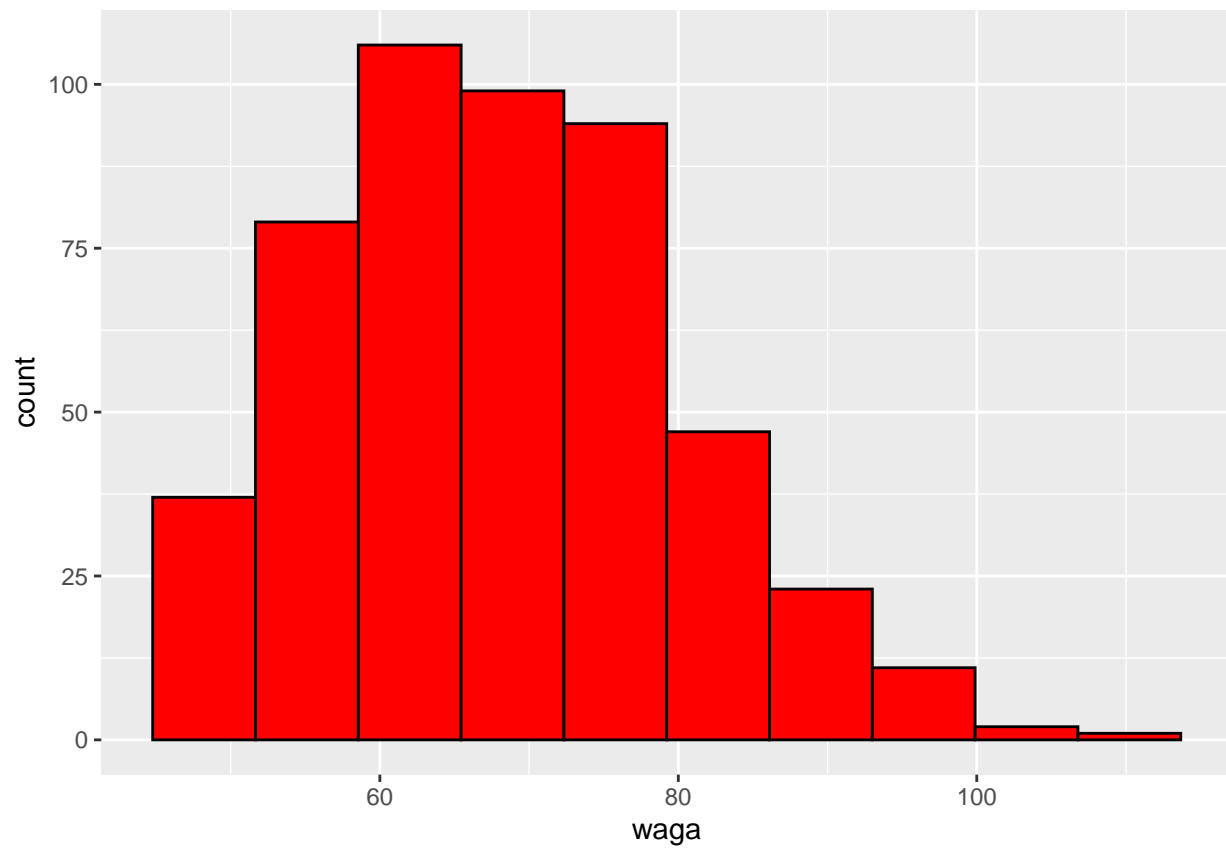
Wykres przedstawia procentowy udział poszczególnych typów budynków. Możemy zauważyć, że najwięcej respondentów mieszka w bloku wielorodzinnym, a najmniej w luksusowym apartamencie.

I to diagramy dla naszych danych:

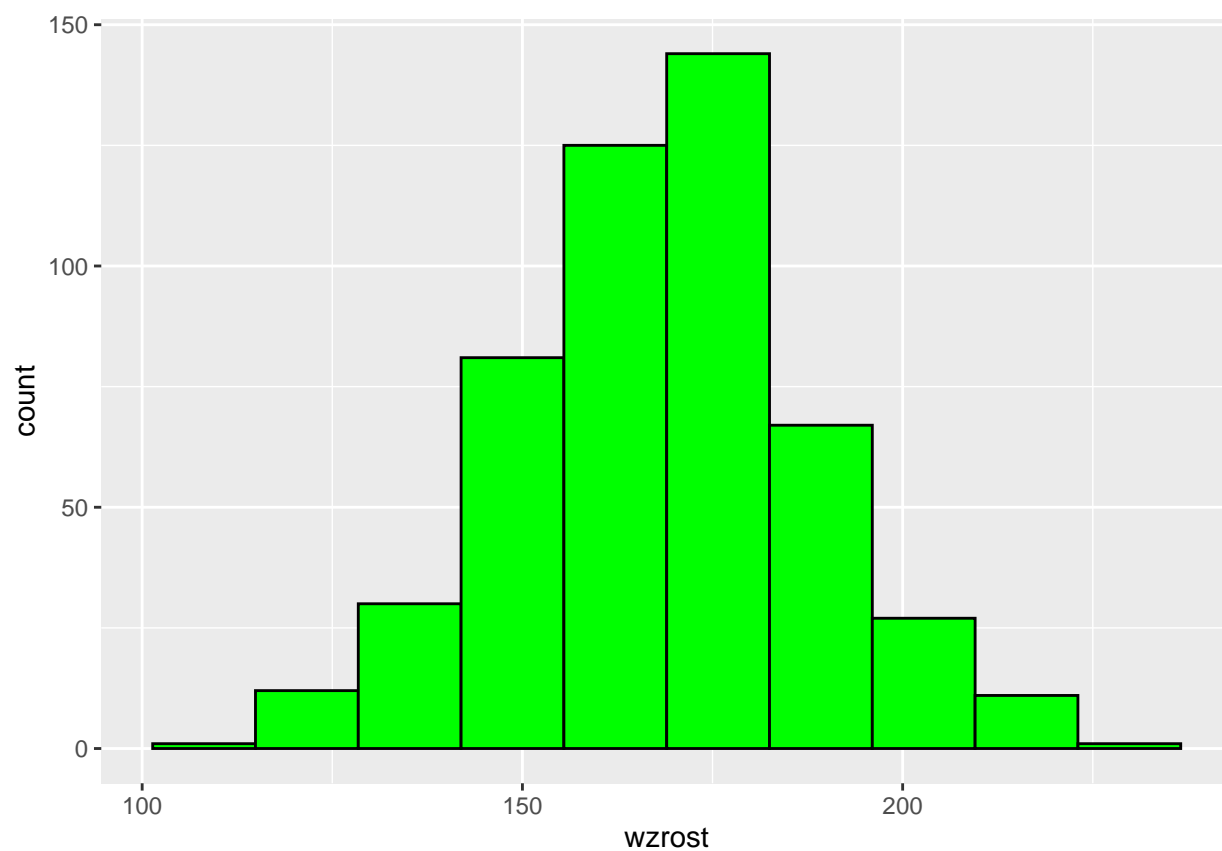
```
ggplot(data, aes(x = wiek)) +  
  geom_histogram(color = "black", fill = "blue", bins = 10)
```



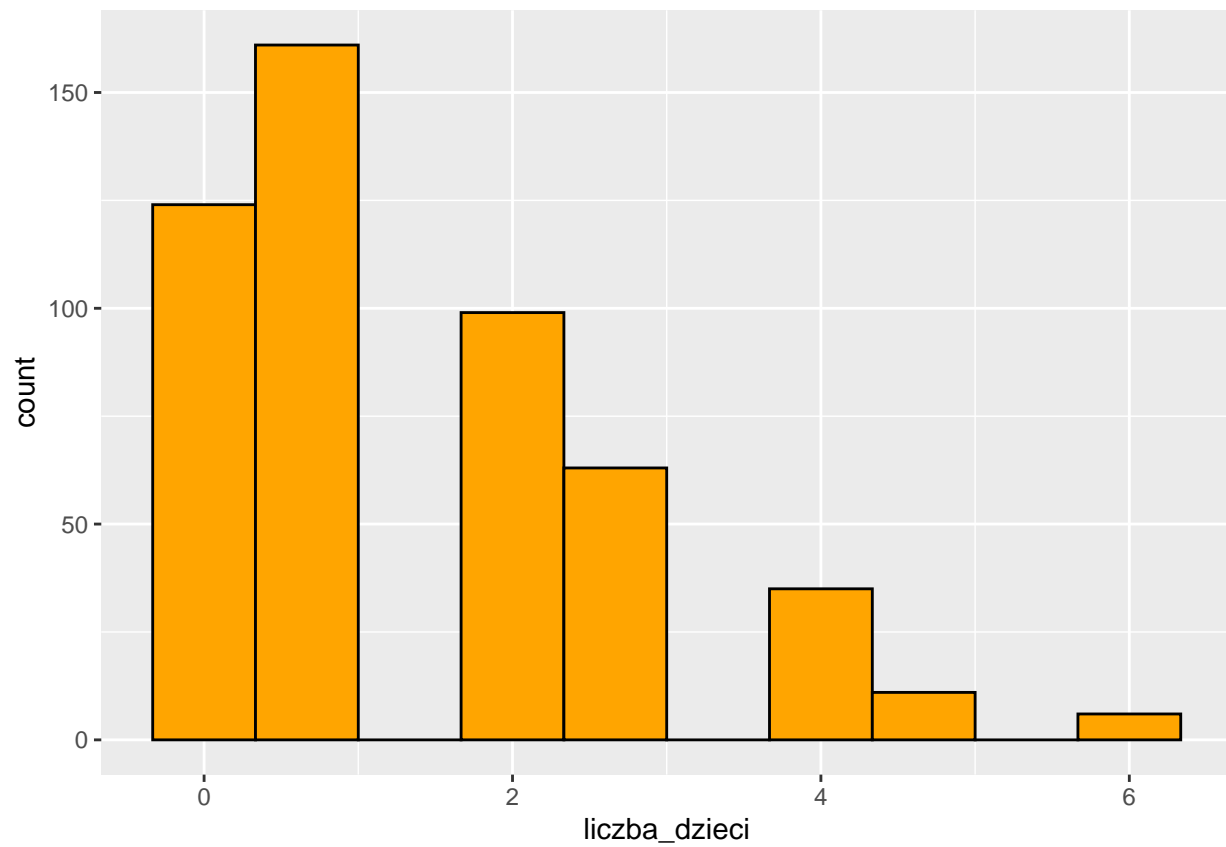
```
ggplot(data, aes(x = waga)) +  
  geom_histogram(color = "black", fill = "red", bins = 10)
```

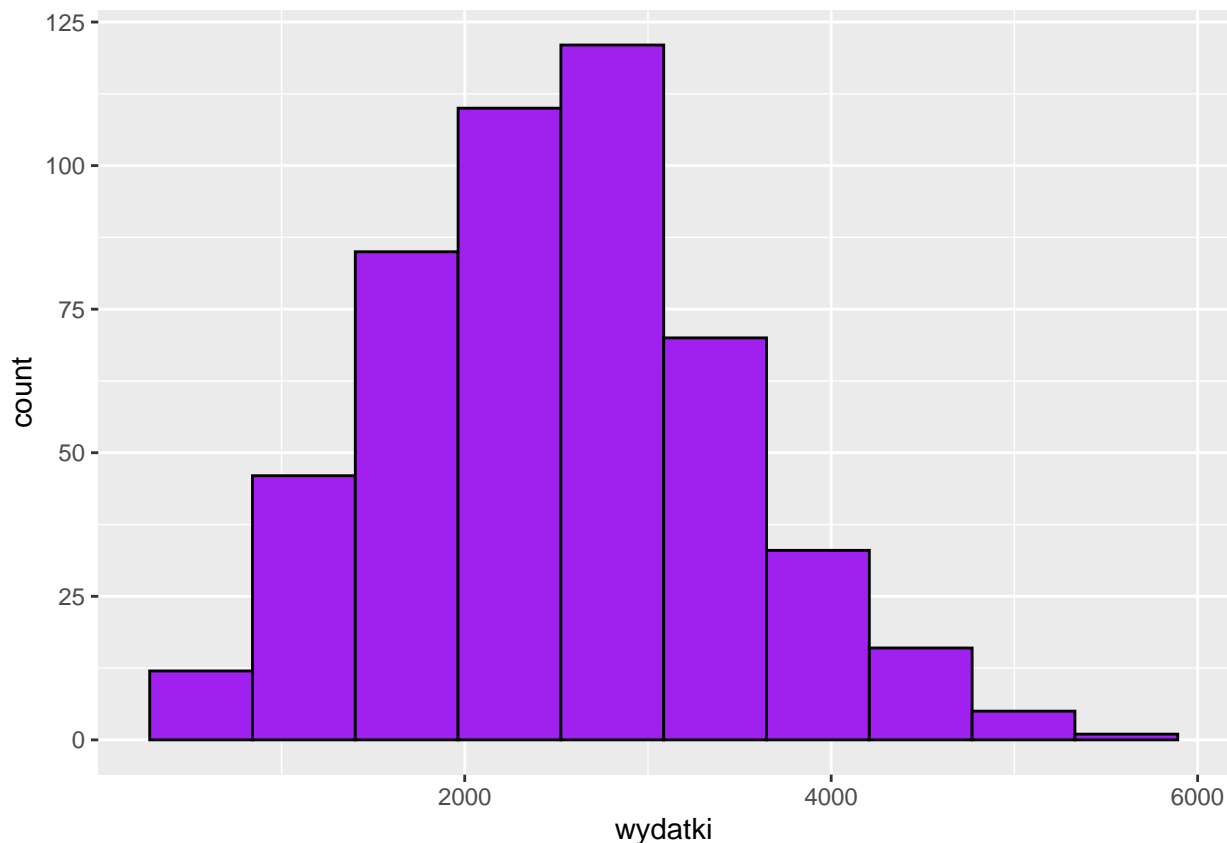
```
ggplot(data, aes(x = wzrost)) +  
  geom_histogram(color = "black", fill = "green", bins = 10)
```



```
ggplot(data, aes(x = liczba_dzieci)) +  
  geom_histogram(color = "black", fill = "orange", bins = 10)
```



```
ggplot(data, aes(x = wydatki)) +  
  geom_histogram(color = "black", fill = "purple", bins = 10)
```



```
ggplot(data, aes(x = wiek, y = wydatki_zywnosc, group = 1)) +
  geom_line(color = "green") +
  labs(title = "Zależność między wiekiem a wydatkami na jedzenie",
       x = "Wiek", y = "Wydatki na jedzenie")
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <bc>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
```

[illegible]

[illegible]

[illegible]

```

## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <bc>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <c4>

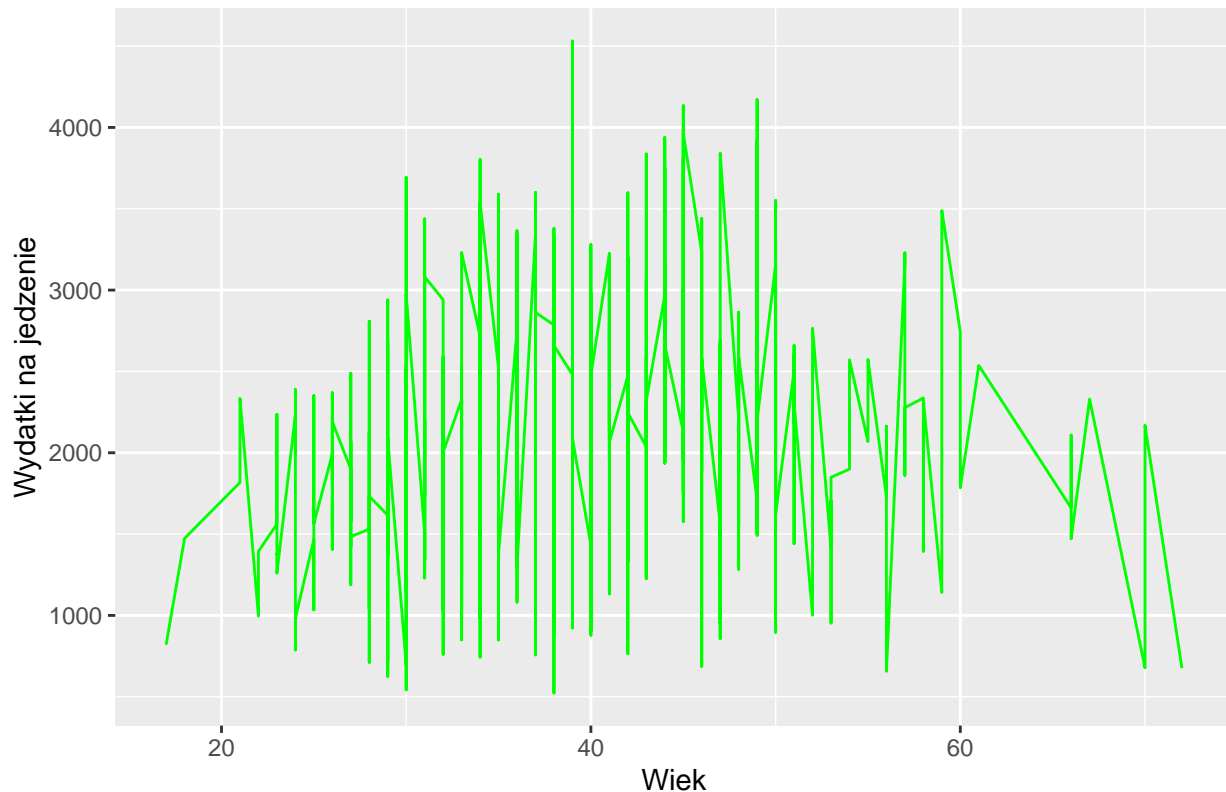
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wiekiem a wydatkami na jedzenie' in
## 'mbcsToSbcs': dot substituted for <99>

```


Zależność między wiekiem a wydatkami na jedzenie



```
ggplot(data, aes(x = wydatki, y = oszczednosci)) +
  geom_point(color = "blue") +
  labs(title = "Zależność między wydatkami a oszczędnościami",
       x = "Wydatki", y = "Oszczędności")
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
```

[illegible]

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Oszczędności' in 'mbcsToSbcs': dot substituted for <99>

```

[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <87>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <99>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c5>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <9b>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c4>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
```

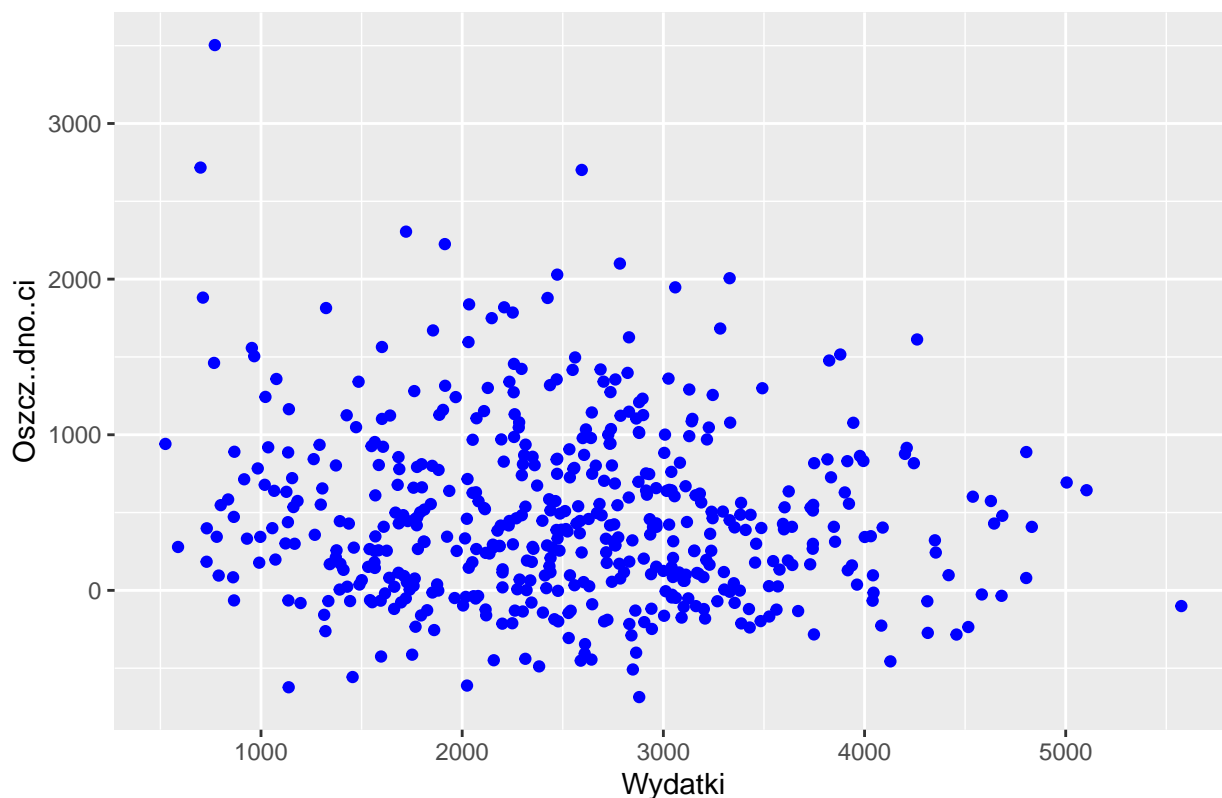
[illegible]

[illegible]

[illegible]


```
## 'mbcsToSbcs': dot substituted for <c4>
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <99>
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c4>
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <99>
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <c5>
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on 'Zależność między wydatkami a oszczędnościami' in
## 'mbcsToSbcs': dot substituted for <9b>
```

Zale..no.... mi..dzy wydatkami a oszcz..dno..ciami



#p-wartości dla hipotez o wartości średniej $\mu = 70$ i medianie $me = 65$ (kg) dla #zmiennej waga, osobno w podpróbach kobiet i mężczyzn.

Zacnimy z kobiet

```
#Test jednostronny t-Studenta dla wartości średniej
t.test(data$waga[data$plec == "K"], mu = 70, alternative = "less")
```

```
##
```

```
## One Sample t-test
##
## data: data$waga[data$plec == "K"]
## t = -3.8787, df = 237, p-value = 6.804e-05
## alternative hypothesis: true mean is less than 70
## 95 percent confidence interval:
##      -Inf 68.40365
## sample estimates:
## mean of x
## 67.22017
```

```
# Test Wilcozona-Manna-Whitneya dla wartości mediany
wilcox.test(data$waga[data$plec == "K"], mu = 65, alternative = "less")
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: data$waga[data$plec == "K"]
## V = 16834, p-value = 0.9952
## alternative hypothesis: true location is less than 65
```

Otrzymały wynik dla testu t-Studento bardzo mały, więc nie możemy odrzucić hipotezę. A dla Wilcozona na odwrót (p-value=0,99). Więc, odrzucamy.

I nie zapominajmy o mężczyznach

```
# Test jednostronny t-Studenta dla wartości średniej
t.test(data$waga[data$plec == "M"], mu = 70, alternative = "less")
```

```
##
## One Sample t-test
##
## data: data$waga[data$plec == "M"]
## t = -1.0265, df = 222, p-value = 0.1529
## alternative hypothesis: true mean is less than 70
## 95 percent confidence interval:
##      -Inf 70.52058
## sample estimates:
## mean of x
## 69.14529
```

```
# Test Wilcozona-Manna-Whitneya dla wartości mediany
wilcox.test(data$waga[data$plec == "M"], mu = 65, alternative = "less")
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: data$waga[data$plec == "M"]
## V = 16620, p-value = 1
## alternative hypothesis: true location is less than 65
```

Otrzymały wynik dla testu t-Studento większy za poziom istotności (0,05) i to oznacza, że możemy odrzucić hipotezę. Dla Wilcozona p-value=1. Więc, też odrzucamy. Warto zauważyć, że test t-studenta zakłada normalność rozkładu zmiennej, a test Wilcozona nie zakłada takiego założenia. W przypadku, gdy rozkład zmiennej jest normalny, test t-studenta jest bardziej efektywny (ma większą moc), a w przypadku braku normalności, test Wilcozona może być bardziej odpowiedni.

Teraz policzymy dwustoronne przedziały ufności:

```
#Średnia i odchylenie standardowe
n <- nrow(data)
mean_wiek <- mean(data$wiek)
sd_wiek <- sd(data$wiek)
z <- qt(0.995, df = n - 1)
blad_sredniej <- z * (sd_wiek / sqrt(n))
przedzial_ufnosci <- c(mean_wiek - blad_sredniej, mean_wiek + blad_sredniej)
przedzial_ufnosci
```

```
## [1] 38.42779 40.50608
```

```
#Kwantyle
q1 <- quantile(data$wiek, 0.25)
q2 <- quantile(data$wiek, 0.5)
q3 <- quantile(data$wiek, 0.75)
n <- length(data$wiek)
t <- qt(0.995, df = n - 1)
se <- IQR(data$wiek) / (sqrt(n) * qt(0.25, df = n - 1))
lower_ci <- q2 - t * se
upper_ci <- q2 + t * se
przedzial_ufnosci <- c(lower_ci, upper_ci)
przedzial_ufnosci
```

```
##      50%      50%
## 41.05789 36.94211
```

W obu przypadkach korzystamy z rozkładu t-Studenta z (n-1) stopniami swobody, ponieważ nie znamy wartości populacyjnych parametrów. Nasze wyniki są (50% 50% 41.05789 36.94211) Dla poziomu ufności 0.99 korzystamy z kwantyla rozkładu t-Studenta dla 0.995, ponieważ to dwustronny przedział ufności.

Hipoteza zerowa: Średnie wartości wybranej zmiennej są równe dla osób zamężnych/zonatych i panów/kawalerów w podpróbie osób w wieku poniżej 40 lat. Hipoteza alternatywna: Średnie wartości wybranej zmiennej są różne dla osób zamężnych/zonatych i panów/kawalerów w podpróbie osób w wieku poniżej 40 lat. Wybieramy test t-Studenta dla dwóch niezależnych prób. Założenia: normalność rozkładu zmiennych i równość wariancji w obu podgrupach.

```
data_sub <- subset(data, wiek < 40, select = c("wydatki", "oszczednosci", "stan_cywilny"))
```

Test t dla dwóch grup w zmiennej “stan__cywilny”

```
t.test(wydatki ~ stan_cywilny, data = data_sub)
```

Różnica dla zmiennej |2109.180- 2880.377| jest istotna #6.2

```
data_sub1 <- subset(data, wiek < 25, select = c("wydatki", "wydatki_zywnosc"))
```

Wartość p jest mniejsza niż poziom istotności 0,01(p=1,563e-06), możemy stwierdzić, że istnieją istotne różnice między średnimi wydatkami ogółem a średnimi wydatkami na żywność.

```
t.test(data_sub1$wydatki, data_sub1$wydatki_zywnosc, paired = TRUE)
```

```
##
## Paired t-test
##
## data: data_sub1$wydatki and data_sub1$wydatki_zywnosc
## t = 7.3764, df = 16, p-value = 1.563e-06
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
## 185.8780 335.8055
## sample estimates:
## mean of the differences
## 260.8418
```

Wartość p jest większa niż poziom istotności 0,01 ($p=0,06$), nie możemy stwierdzić, że istnieje istotna korelacja między udziałem wydatków na żywność w wydatkach ogółem a oszczędnościami. Współczynnik korelacji wskazuje na siłę i kierunek korelacji: wartości dodatnie wskazują na pozytywną korelację (im większy udział wydatków na żywność, tym większe oszczędności), a wartości ujemne na negatywną korelację (im większy udział wydatków na żywność, tym mniejsze oszczędności).

```
data_sub2 <- data[, c("wydatki_zywnosc", "wydatki", "oszczednosci")]

data_sub2$udzial_zywnosci <- data_sub2$wydatki_zywnosc / data_sub2$wydatki
cor.test(data_sub2$udzial_zywnosci, data_sub2$oszczednosci, method = "pearson")
```

```
##
## Pearson's product-moment correlation
##
## data: data_sub2$udzial_zywnosci and data_sub2$oszczednosci
## t = 1.8247, df = 497, p-value = 0.06865
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.006248875 0.168148817
## sample estimates:
## cor
## 0.08157436

library(fitdistrplus)
```

Fit an exponential distribution to the data

```
fit_exp <- fitdist(data$wydatki_zywnosc/100, "exp")
```

Test the goodness of fit of the exponential distribution

```
test_exp <- gofstat(fit_exp, fitnames = "exp")
```

Calculate the p-value of the Chi-squared test

```
p_val1 <- test_exp[1, "pvalue"]
```

Calculate Kolmogorov-Smirnov test

```
ks.test(data$wydatki, pexp, 10)
```

Compare the p-value with the significance level (e.g., $\alpha = 0.05$)

```
if(p_val1 < 0.05){ cat("Hipoteza zerowa zostaje odrzucona na poziomie istotności 0.05.") }else{ cat("Nie ma podstaw do odrzucenia hipotezy zerowej na poziomie istotności 0.05.") }
```

```
#7 #7.1
```

```
model <- lm(oszczednosci ~ wiek + waga + wzrost + plec + stan_cywilny + liczba_dzieci + budynek + wydatki_zywnosc, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = oszczednosci ~ wiek + waga + wzrost + plec + stan_cywilny +
##     liczba_dzieci + budynek + wydatki_zywnosc, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -306.77  -60.38   -2.82   59.76  484.19
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -801.89926    57.75419  -13.885 < 2e-16 ***
## wiek           64.02584     0.54344  117.815 < 2e-16 ***
## waga           3.67455     0.55280   6.647 8.04e-11 ***
## wzrost        -2.38245     0.34594  -6.887 1.77e-11 ***
## plecM          1.27206     9.76642   0.130  0.896
## plecNieznana   12.06944    18.17896   0.664  0.507
## stan_cywilny   -4.67292    12.54050  -0.373  0.710
## liczba_dzieci  154.11866     5.97938  25.775 < 2e-16 ***
## budynekjednorodzinny -184.16970    16.06441 -11.464 < 2e-16 ***
## budynekkamienica  -307.31625    17.45922 -17.602 < 2e-16 ***
## budyneklift     -344.24376    23.80913 -14.458 < 2e-16 ***
## budynekwielka_plyta -566.51557    19.72552 -28.720 < 2e-16 ***
## wydatki_zywnosc   -0.49901     0.01253 -39.841 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 103.4 on 486 degrees of freedom
## Multiple R-squared:  0.9675, Adjusted R-squared:  0.9666
## F-statistic: 1204 on 12 and 486 DF,  p-value: < 2.2e-16
```

Otrzymały taki wyniki, że trzy zmienni można odrzucić jako niewystarczającą w tłumaczeniu zmienności zmiennej zależnej. ($p > 0,05$)

plecM 0.896

plecNieznana 0.507

stan_cywilny 0.710

można ich odrzucić jako niewystarczającą w tłumaczeniu zmienności zmiennej zależnej.

#.2

```
model <- lm(oszczednosci ~ wiek + waga + wzrost + plec + stan_cywilny + liczba_dzieci + budynek + wydatki_zywnosc, data = data)
reduced_model <- lm(oszczednosci ~ . - liczba_dzieci, data = data)
```

Aby obliczyć wartość R^2 dla modelu bez danej zmiennej, można użyć funkcji `lm()` w R, ale z pominięciem tej zmiennej w formule modelu. Następnie można wykorzystać funkcję `summary()` i odczytać wartość R^2 z podsumowania modelu.

Na przykład, jeśli zmienna `liczba_dzieci` jest zmienną, którą chcemy usunąć z pełnego modelu, to otrzymujemy wynik: 0,044 Po wykonaniu tego testu wiele razy dla różnych zmiennych otrzymaliśmy tę samą wartość 0,044

wartości R^2

```
R2_full <- summary(model)$r.squared
R2_reduced <- summary(reduced_model)$r.squared
```

```

# różnica R^2
R2_diff_liczba_dzieci <- R2_full - R2_reduced

reduced_model_waga <- summary(lm(oszczednosci ~ . - waga, data = data))$r.squared
R2_diff_waga <- R2_full - reduced_model_waga

model <- lm(oszczednosci ~ wiek + waga + wzrost + plec + stan_cywilny + liczba_dzieci + budynek + wydatki)
# Usuń zmienną "plec" z modelu
model_reduced <- lm(oszczednosci ~ wiek + waga + wzrost + stan_cywilny + liczba_dzieci + budynek + wydatki)

# Oblicz RSS dla pełnego modelu i modelu z usuniętą zmienną "plec"
RSS_full <- sum(resid(model)^2)
RSS_reduced <- sum(resid(model_reduced)^2)

Po usunięciu zmiennej 'plec' RSS wzrósł o 52636.82

# Oblicz różnicę RSS
RSS_diff <- RSS_full - RSS_reduced

# Wyświetl wynik
cat("Po usunięciu zmiennej 'plec' RSS wzrósł o", RSS_diff)

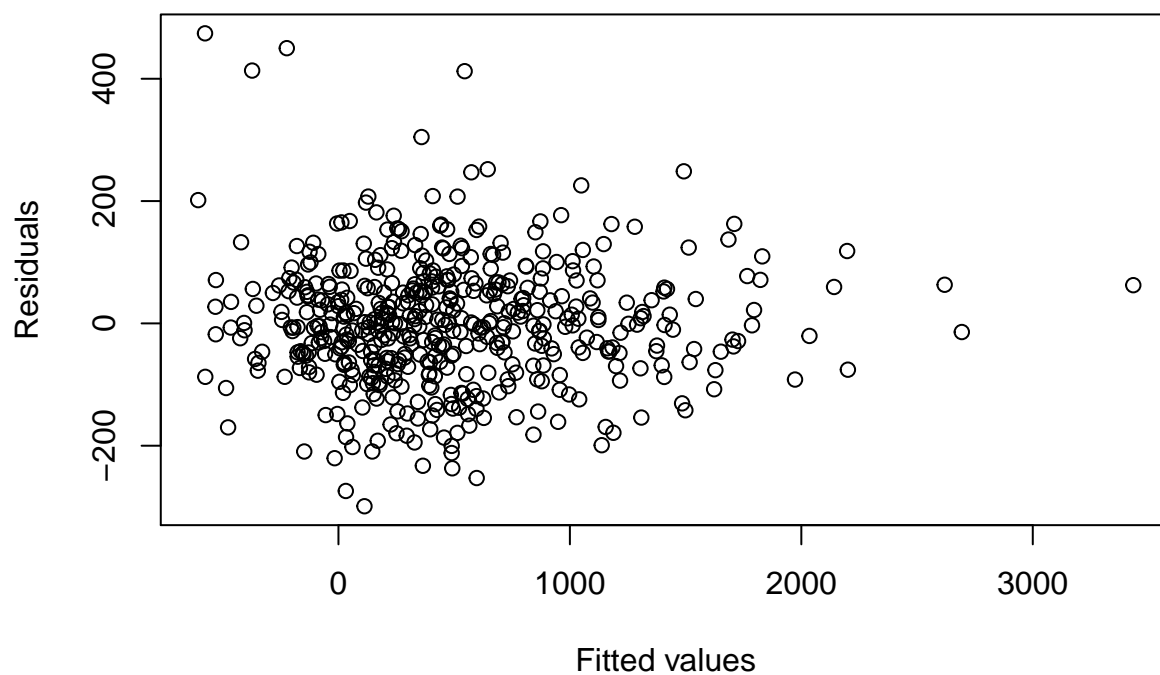
## Po usunięciu zmiennej 'plec' RSS wzrósł o 52636.82

# .4 Remove the column corresponding to the selected variable from the independent variables matrix X
X_reduced <- data[, !(names(data) %in% c("wzrost"))]

# Fit a linear regression model with the reduced set of independent variables
model_reduced_wzrost <- lm(oszczednosci ~ ., data = X_reduced)
# Plot the residuals against the dependent variable
plot(model_reduced_wzrost$fitted.values, model_reduced$residuals,
      xlab = "Fitted values", ylab = "Residuals",
      main = "Residuals vs Fitted values plot")

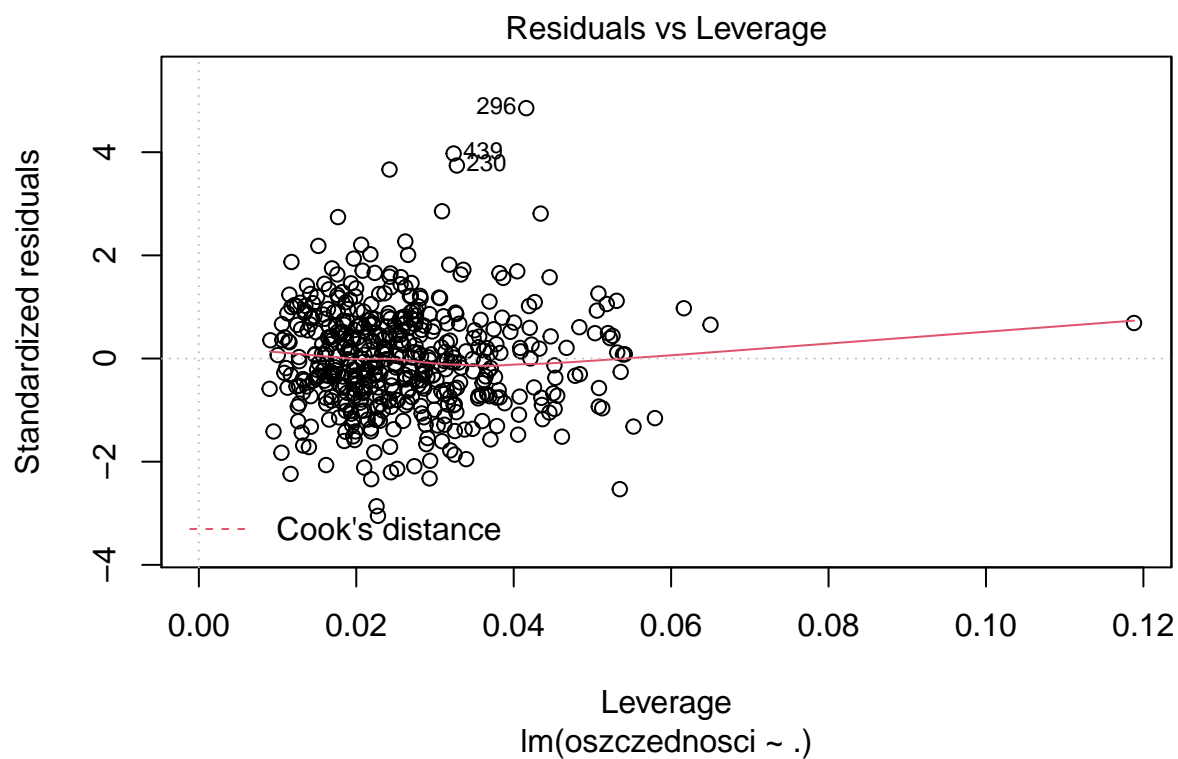
```

Residuals vs Fitted values plot



```
# Plot the leverage
plot(model_reduced_wzrost, which = 5, main = "Leverage plot")
```

Leverage plot



Wykres zależności reszt od zmiennej objaśnianej pozwala nam sprawdzić, czy warunek homoskedastyczności

jest spełniony (czyli czy wariancja reszt jest stała wzdłuż całego zakresu wartości zmiennej objaśnianej). Na wykresie powinniśmy zobaczyć losowo rozproszone punkty bez żadnego widocznego trendu. Jeśli punkty układają się w sposób określony, np. w formie leja, to wskazuje to na niehomoskedastyczność. W naszym przypadku wykres sugeruje, że warunek homoskedastyczności jest spełniony.