Raport 1

Romana Żmuda

1 04 2020

Wprowadzenie zmiennych z pliku dane1.txt do Tabeli "Moje Dane", jak również poprawne ustawienie i nazwanie zmiennych.

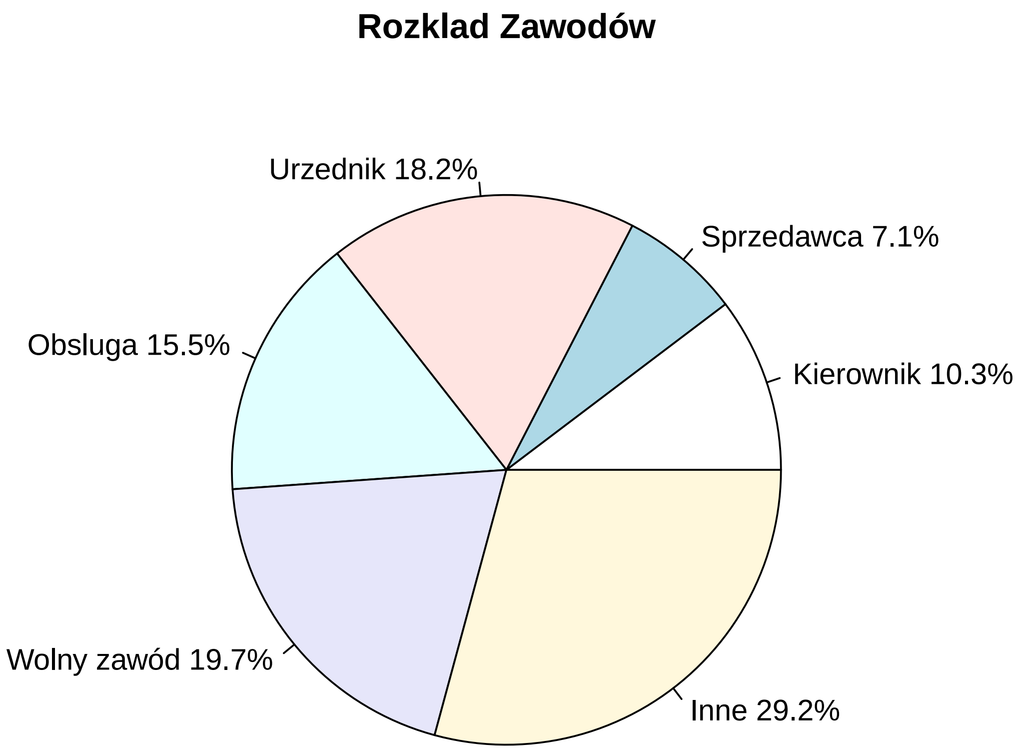
```
setwd(datapath)
MojeDane<-read.table("dane1.txt",skip = '13' ,blank.lines.skip = TRUE, fill = TRUE, header = FALSE)
colnames(MojeDane)<-c('edukacja','miejscezam','płeć','doświadczenie','związek','zarobki','wiek','rasa',
```

Zadanie 1

W tej sekcji mamy do analizy dane z tabeli Moje Dane o nazwie typ, w której to przenalizowałam 6 typów zawodów. Następnie na wykresie kołowym umieściłam dane procentowe, jak również nazwy 6 zawodów: kierownik, sprzedawca/marketing, urzędnik, obsługa, wolny zawód, inne.

Po wstępnej analizie danych widzimy, jak rozkłada się ilościowy podział (534 osoób) na konkretne zawody: 55 kiernownik, 38 sprzedawca, 97 urzędnik, 83 obsługa, 105 wolny zawód, 156 inne. Wykres kołowy prezentujący procentowy udział każdego z zawodów.

```
zawody<-c(praca1,praca2,praca3,praca4,praca5,praca6)
suma_zawody<-sum(zawody)
procenty_zawody<-round(100*zawody/suma_zawody,1)
labels<-c("Kierownik","Sprzedawca","Urzędnik","Obsługa","Wolny zawód","Inne")
labels<-paste(labels,procenty_zawody)
labels<-paste(labels,"%", sep="")
pie(table(MojeDane$praca),labels=labels,main="Rozkład Zawodów")
```



Zadanie 2

Tabela odpowiednich wartości dla danych pensja, czyli : minimum, pierwszy kwantyl, mediana, średnia, trzeci kwantyl, maksimum.

```
pensja_wartosci<-select(MojeDane,pensja)
summary(pensja_wartosci)
```

```
##      pensja
##  Min.   :  4.03
##  1st Qu.: 21.16
##  Median : 31.35
##  Mean   : 36.37
##  3rd Qu.: 45.34
##  Max.   :179.34
```

Zadanie 3

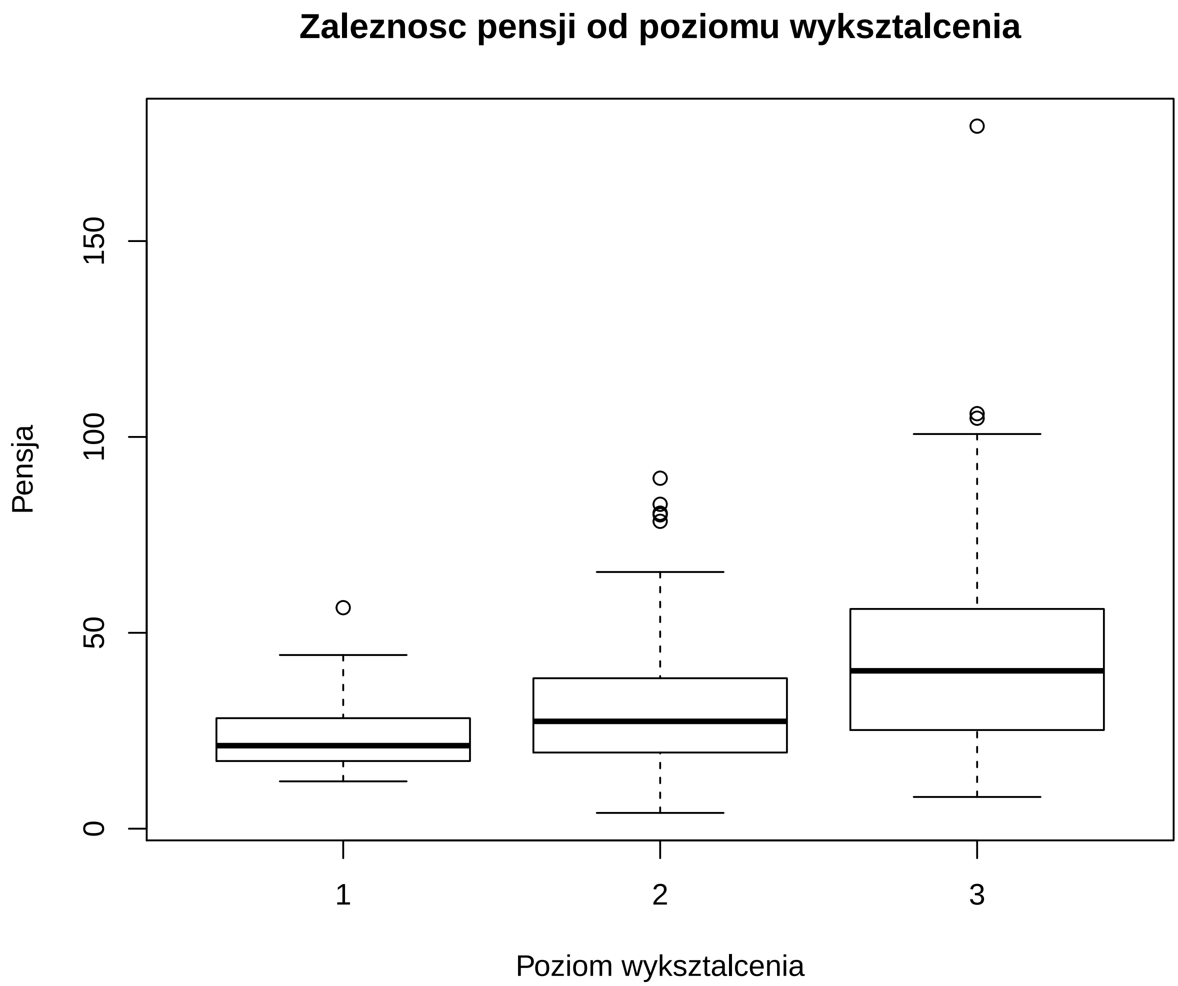
W tym zadaniu mamy do przeanalizowania dane z kolumny wykształcenie, które dzieli się na 3 podgrupy: 1. Osoba ucząca się do 8 lat. 2.Osoba ucząca się od 8 do 12 lat. 3.Osoba ucząca się ponad 12 lat. Jak również korelację między wykształceniem, a pensją.

Tabela odpowiednich wartości dla danych wykształcenie po odpowiednim podziale, czyli : minimum, pierwszy kwantyl, mediana, średnia, trzeci kwantyl, maksimum.

```
edu_1<-summary(edukacja_1_P)
edu_2<-summary(edukacja_2_P)
edu_3<-summary(edukacja_3_P)
total<-cbind(edu_1,edu_2,edu_3)
colnames(total)<-c('edukacja<=8','8<edukacja<=12','edukacja>12')
total
```

```
##  edukacja<=8       8<edukacja<=12    edukacja>12
##  "Min.   :12.09  " "Min.   : 4.03  " "Min.   :  8.10  "
##  "1st Qu.:17.27  " "1st Qu.:19.44  " "1st Qu.: 25.19  "
##  "Median :21.20  " "Median :27.40  " "Median : 40.30  "
##  "Mean   :24.28  " "Mean   :30.71  " "Mean   : 44.48  "
##  "3rd Qu.:28.21  " "3rd Qu.:38.41  " "3rd Qu.: 56.04  "
##  "Max.   :56.42  " "Max.   :89.47  " "Max.   :179.34  "
```

```
wyksz_pensja<-MojeDane %>% select(wykszt,pensja)
boxplot(pensja ~ wykszt, data = wyksz_pensja, xlab = "Poziom wykształcenia",
        ylab = "Pensja", main = "Zależność pensji od poziomu wykształcenia")
```



Zauważmy, iż pensja rośnie wraz z poziomem edukacji, jak również rośnie zakres kwotowy zarobków, więc istnieją spore dysproporcje w średnich zarobkach. Jeśli uczyliśmy się do 8 lat to zarabiamy podobnie w całej naszej grupie, a samo odchylenie nie odbiega od normy. Największe różnice są gdy przekroczymy 12 lat nauki, wtedy istnieje spora szansa wysokich zarobków., jednak spora grupa zarabia poniżej średniej. ## Zadanie 4

```
d<-CrossTable(MojeDane$rasa, MojeDane$wykszt, expected = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  534
```

```
##
##
##                | MojeDane$wykszt
## MojeDane$rasa |         1 |         2 |         3 | Row Total |
## --------------|-----------|-----------|-----------|-----------|
##             1 |         5 |        37 |        25 |        67 |
##               |     0.767 |     0.181 |     0.580 |           |
##               |     0.075 |     0.552 |     0.373 |     0.125 |
##               |     0.185 |     0.135 |     0.108 |           |
##               |     0.009 |     0.069 |     0.047 |           |
## --------------|-----------|-----------|-----------|-----------|
##             2 |         8 |         8 |        11 |        27 |
##               |    32.246 |     2.507 |     0.045 |           |
##               |     0.296 |     0.296 |     0.407 |     0.051 |
##               |     0.296 |     0.029 |     0.047 |           |
##               |     0.015 |     0.015 |     0.021 |           |
## --------------|-----------|-----------|-----------|-----------|
##             3 |        14 |       230 |       196 |       440 |
##               |     3.057 |     0.051 |     0.122 |           |
##               |     0.032 |     0.523 |     0.445 |     0.824 |
##               |     0.519 |     0.836 |     0.845 |           |
##               |     0.026 |     0.431 |     0.367 |           |
## --------------|-----------|-----------|-----------|-----------|
##  Column Total |        27 |       275 |       232 |       534 |
##               |     0.051 |     0.515 |     0.434 |           |
## --------------|-----------|-----------|-----------|-----------|
##
##
```

```
e<-CrossTable(MojeDane$rasa, MojeDane$wykszt, expected=FALSE, prop.r=FALSE, prop.c=FALSE,
              prop.t=TRUE, prop.chisq=FALSE, chisq = FALSE, fisher=FALSE, mcnemar=FALSE,
              resid=FALSE, sresid=FALSE, asresid=FALSE,
              missing.include=FALSE,format=c("SAS","SPSS"))
```

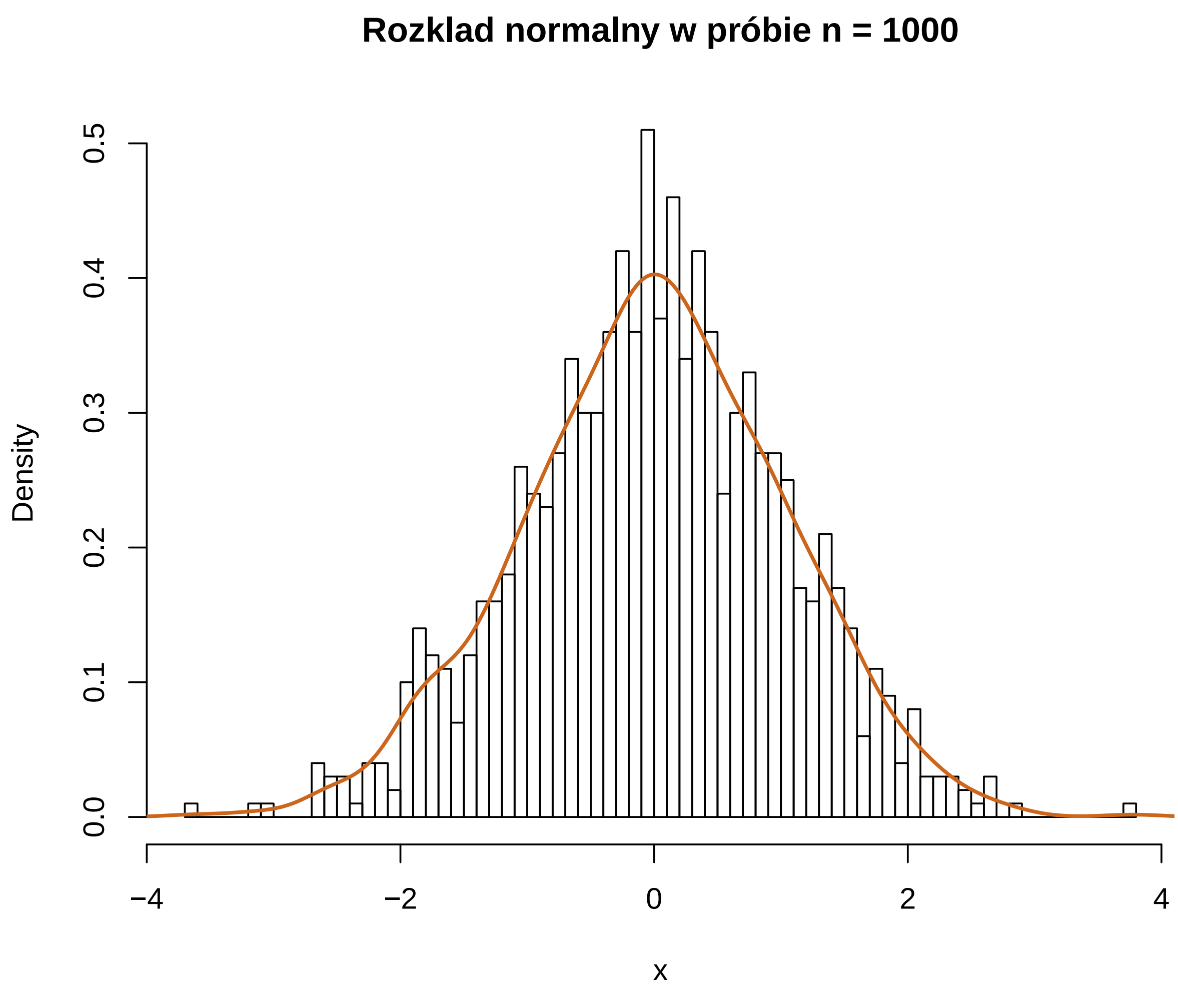
```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  534
##
##
##                | MojeDane$wykszt
## MojeDane$rasa |         1 |         2 |         3 | Row Total |
## --------------|-----------|-----------|-----------|-----------|
##             1 |         5 |        37 |        25 |        67 |
##               |     0.009 |     0.069 |     0.047 |           |
## --------------|-----------|-----------|-----------|-----------|
##             2 |         8 |         8 |        11 |        27 |
##               |     0.015 |     0.015 |     0.021 |           |
## --------------|-----------|-----------|-----------|-----------|
```

```
##              3 |        14 |       230 |       196 |       440 |
##                |     0.026 |     0.431 |     0.367 |           |
## ---------------|-----------|-----------|-----------|-----------|
##   Column Total |        27 |       275 |       232 |       534 |
## ---------------|-----------|-----------|-----------|-----------|
##
##
```

W Pivocie zsumowano liczbę osób różnych nacji z stosunku z poziomem edukacji, w każdej komórce umieszczono odpowiednie podsumowania. W drugiej tabeli mamy samą wartość procentową sumującą się do jedynki, więc aby uzyskać procentowy udział należy pomnożyć razy 100.

Zadanie 5

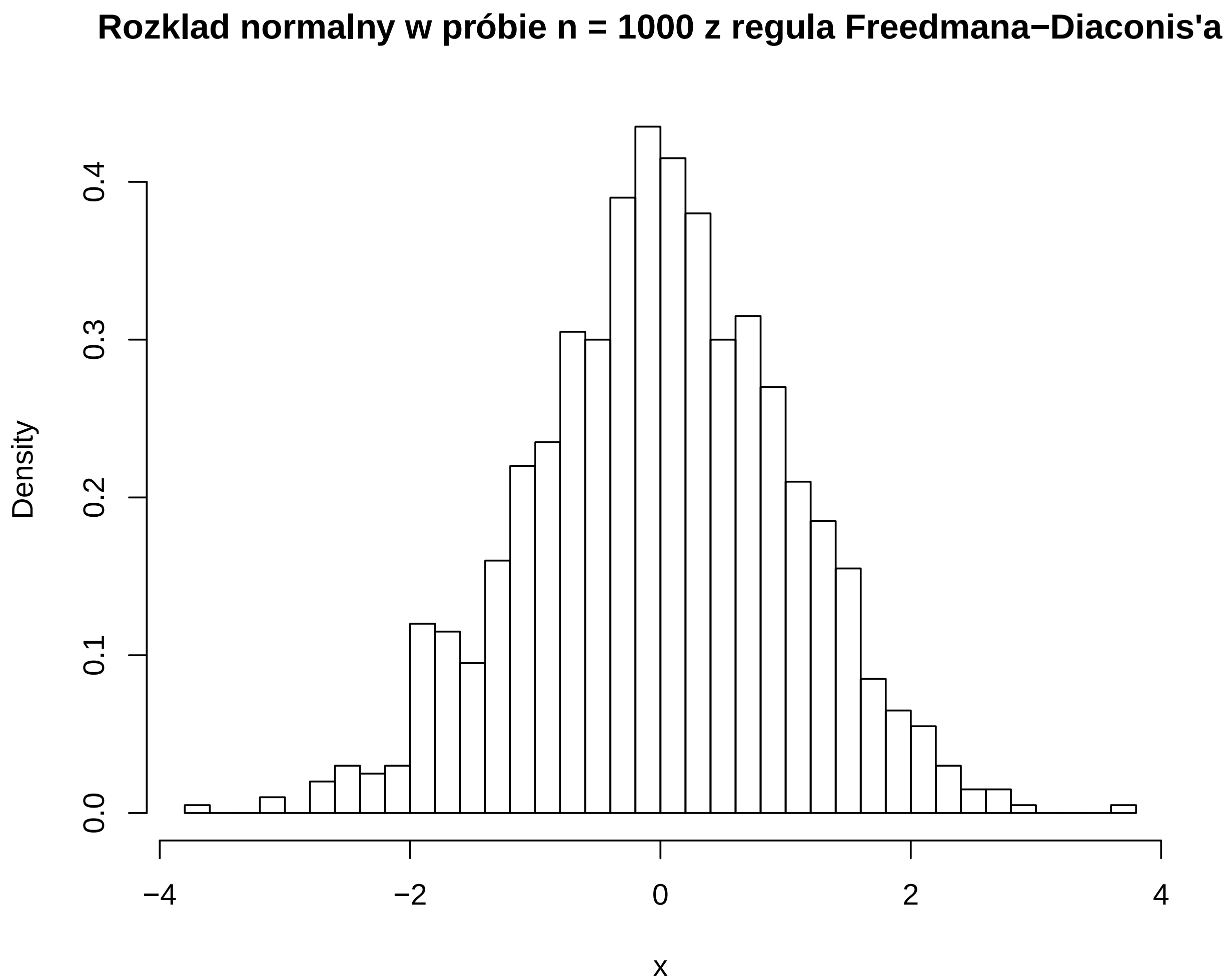
Narysowałam histogram dla próby losowej rozmiaru n = 1000 z rozkładu N(0,1), umieszczając na nim także wykres gęstości tego rozkładu



Jednak ten wykres nie przybliża najlepiej dlatego obliczamy liczbę klas według reguły Freedmana-Diaconisa

```
max<-max(x)
min<-min(x)
iqr<-IQR(x) # obliczanie mianownika, czyli funkcję h
```

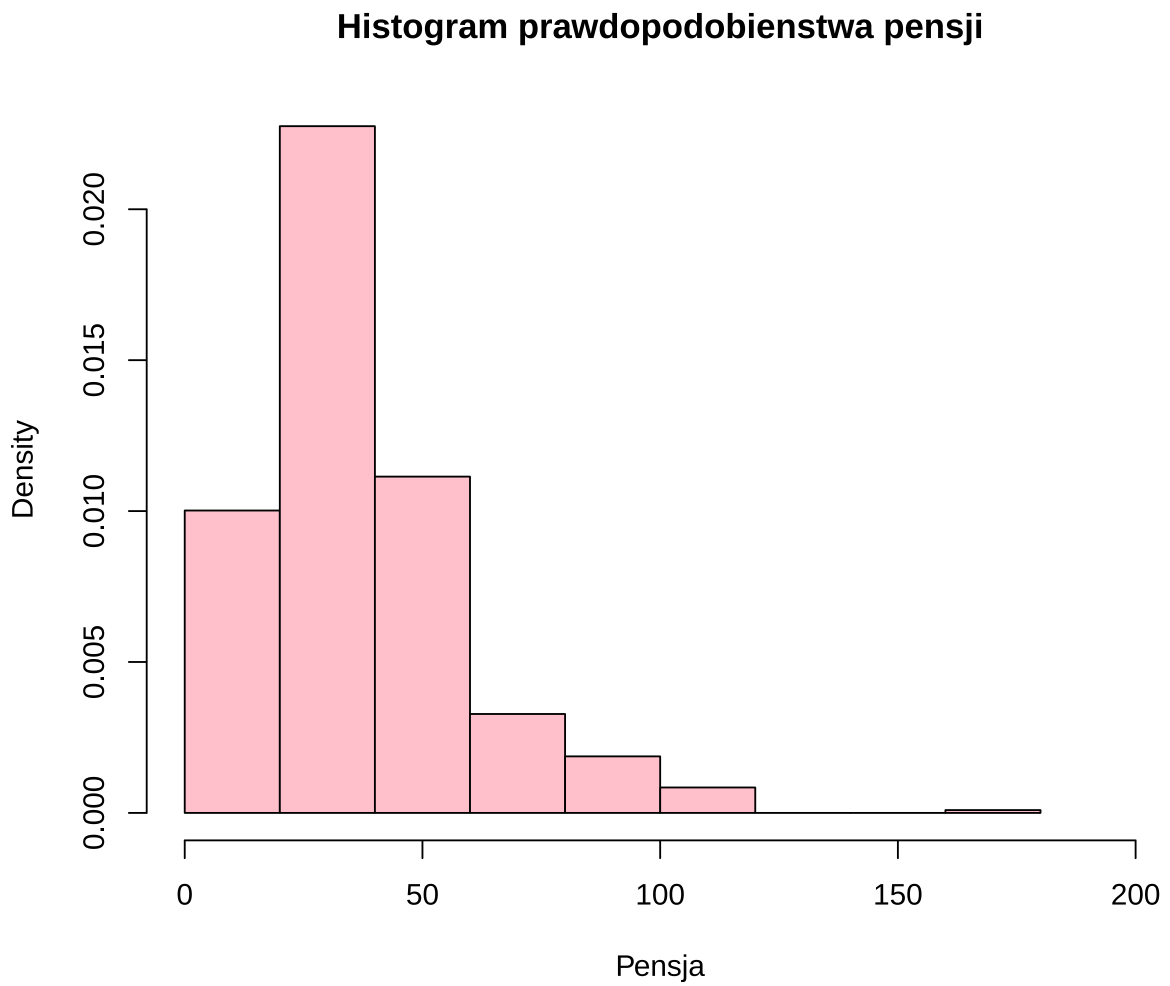
```r
liczba_klas<-ceiling(((max - min)*(1000)^(1/3))/2*iqr) # liczba klas,
#czyli na ile podzielić rozkład,
#aby był zbliżony do rozkładu normalnego
hist(x,main="Rozkład normalny w próbie n = 1000 z regułą Freedmana-Diaconis'a",
     freq = FALSE, breaks = liczba_klas)
```



Zadanie 6

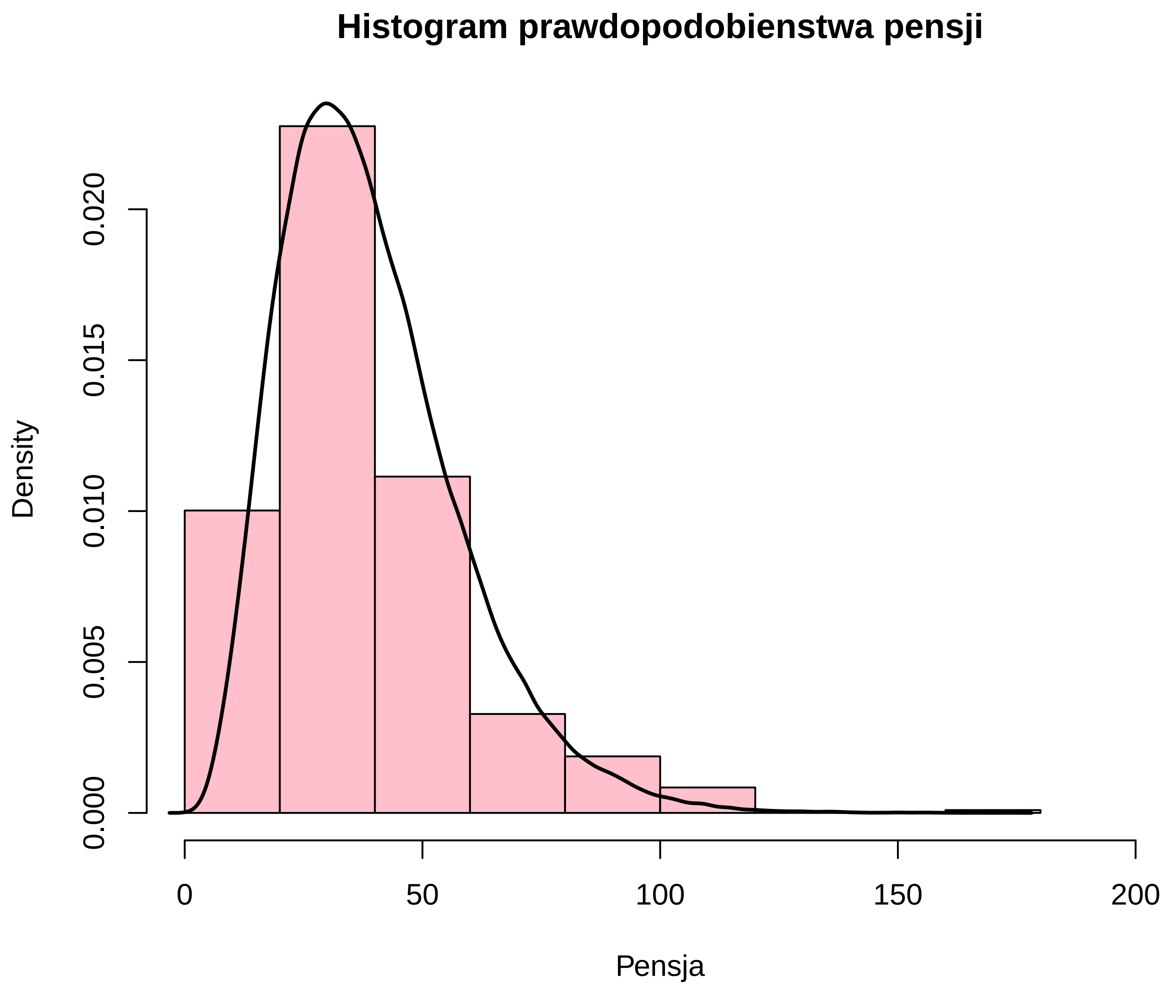
Rozkład pensji:

```r
hist(MojeDane$pensja,main="Histogram prawdopodobieństwa pensji",
     xlab="Pensja",freq = FALSE,xlim=c(0,200),col="pink")
```



Zadanie 7

```
k<-rgamma(100000, shape=4.35, scale =9)
hist(MojeDane$pensja,main="Histogram prawdopodobieństwa pensji",
     xlab="Pensja",freq = FALSE,xlim=c(0,200),col="pink")
lines(density(k),
      lwd = 2)
```



Po obserwacji wykresu z zadania 6 możemy zauważyć, że histogram przypomina rozkład gammę. Po pewnych próbach mogłam ustalić wartości alfy oraz bety są one odpowiednio równe:

1. $\alpha = shape = 4.35$
2. $\beta = \frac{1}{scale} = 0.(1)$