Report #1

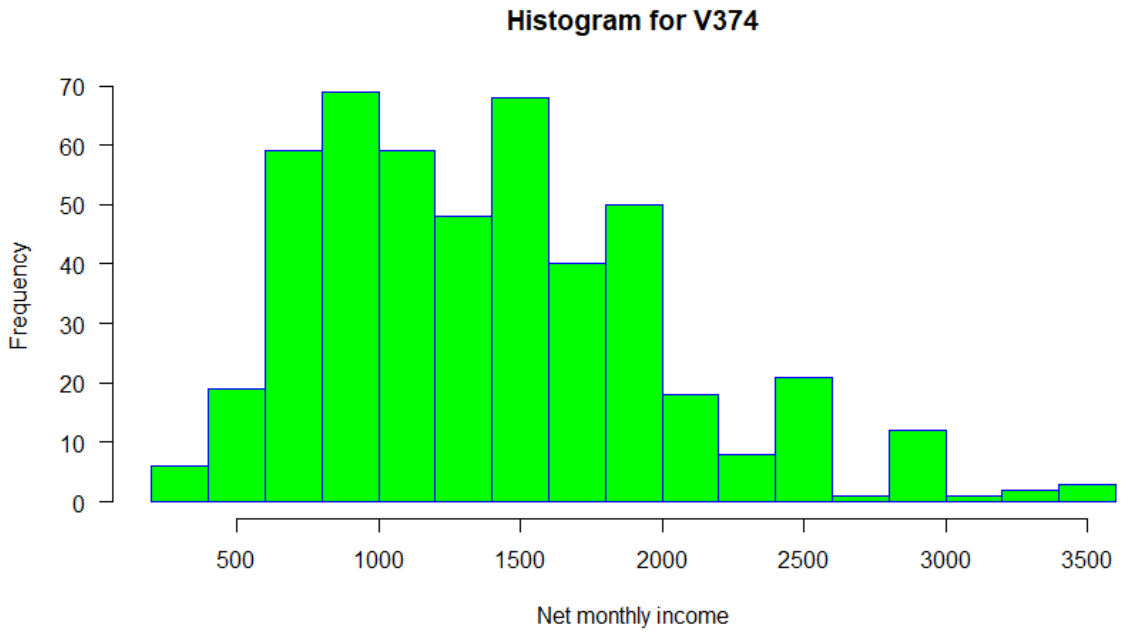
1. Introduction

This report gives an overview of net monthly income in a "WARMIŃSKO-MAZURSKIE" voivodship. The overall report is based on a full set of micro-data from CBOS questionnaires. The data was cleaned and outliers was delated.

2. Body

I. Statistics for all answers in question V39

*"**V374** M18A_1. Miesięczne dochodów NETTO źródła resp: Stała praca najemna (etat, kontrakt itp.)"*

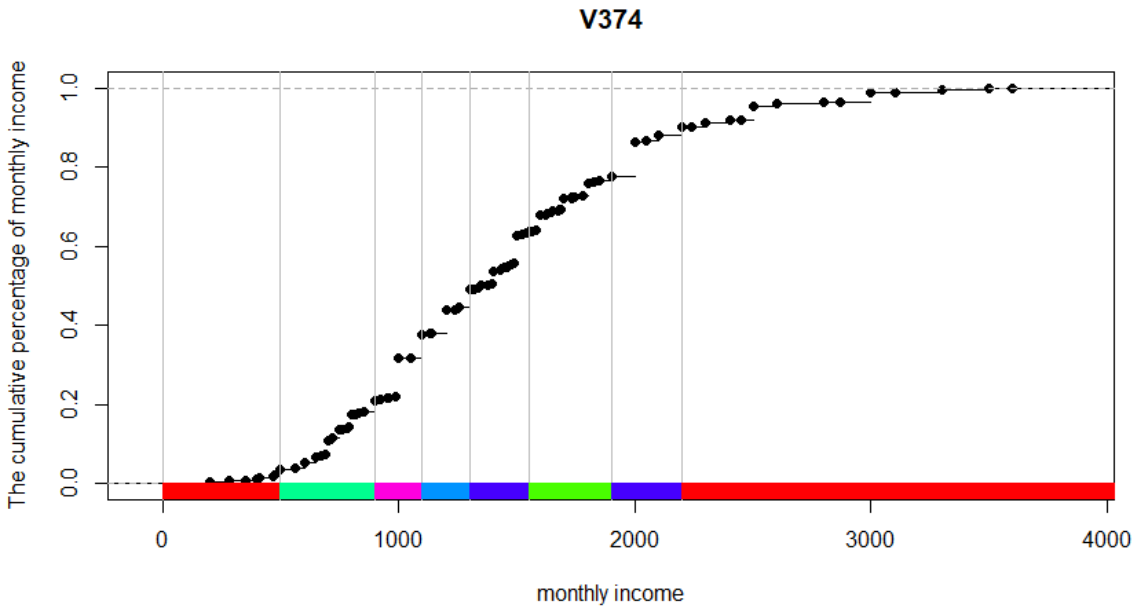


We can see that our population has diversified net monthly income. The majority of people earn 800 – 2000 per month. To future analysis we have to divide the data for samples.

Intervals	n	%
[0,500]	17	3,512397
(500,900]	85	17,56198
(900,1100]	80	16,52893
(1100,1300]	55	11,36364
(1300,1550]	71	14,66942
(1550,1900]	68	14,04959
(1900,2200]	60	12,39669
(2200,3600]	48	9,917355

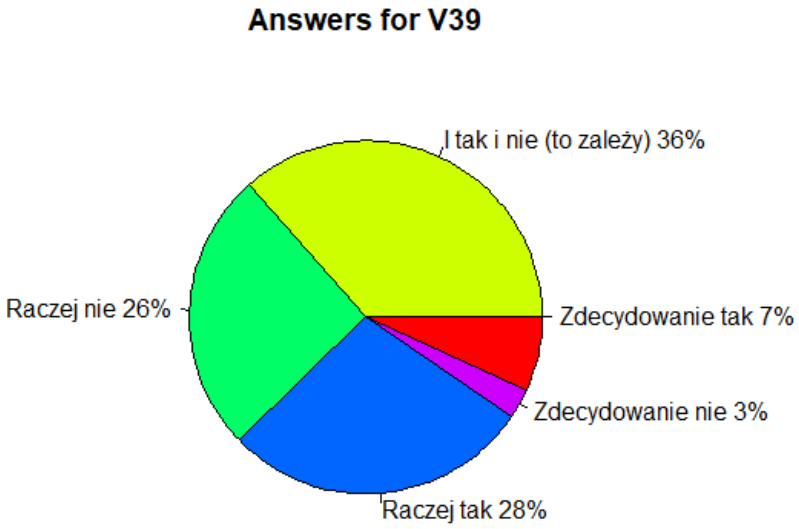
Goodness of fit = 0.9554584 Tabular accuracy = 0.8236688

We grouped the data into mutually exclusive classes showing the number of observations in each interval. Next, we converted it to relative class frequencies to show the fraction of the total number of observations in each interval.



The plot shows us the frequency distribution. Cumulative percentage is calculated by dividing the cumulative frequency by the total number of observations.

*"**V39** P6_6. Czy zgadza się z: … Obecnie w naszym kraju każdy przedsiębiorczy człowiek może się dorobić."*



The chart shows the percent that each answer represents of the total number of frequencies. We can make a conclusion that our population is unsure about this topic because only 10% answers are certainly yes/no. The most frequently answer is "it depends".

Report #1

Answer	Arithmetic mean[1]	Median[2]	Minimum	Maximum	1st quartile[3]	3rd quartile[4]	IQR[5]	Standard deviation[6]
All	1436.326	1365	200	3600	1000	1800	800	622.0755
I tak i nie (to zależy)	1484.738	1400	200	3600	1000	2000	1000	648.5559
Raczej nie	1396.664	1200	350	3500	980.0	1767.5	787.5	663.9321
Raczej tak	1423.738	1300	500	3300	1000	1800	800	576.1581
Zdecydowanie nie	1499	1487	700	2500	1000	1600	600	555.2894
Zdecydowanie tak	1388.387	1500	200	2600	950	1800	850	602.0083

We can make some conclusions from the data:

- The majority of people answered “I tak i nie (to zależy)”.
- People who answered „Zdecydowanie nie” and „Zdecydowanie tak” statisticaly earn more than others.
- The lowest salaries (statisticaly) are connected with answer “Raczej nie”.
- However the lowest salary and the highest one are connected with answer “I tak i nie (to zależy)”.
- We do not have clear dependence between salary of people and their answer on question V39.

[1] Arithmetic mean or simply the average is the sum of a collection of numbers divided by the number of numbers in the collection.

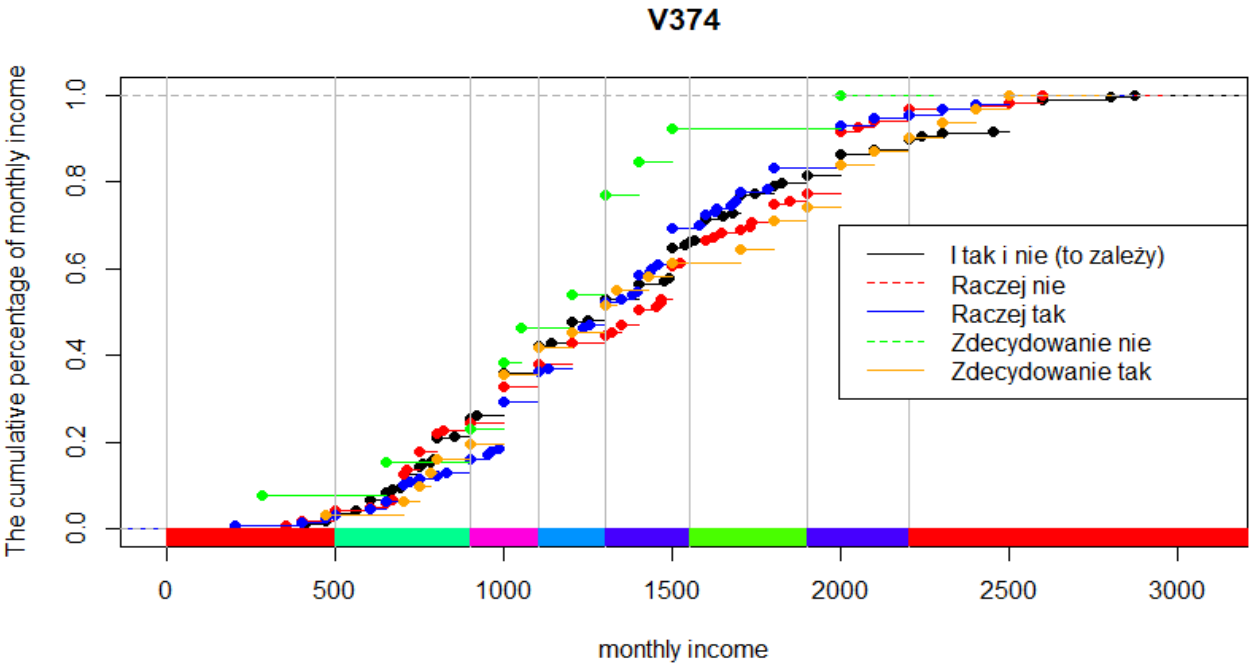
[2] The median is the value separating the higher half of a data sample from the lower half. It may be thought of as the "middle" value.

[3] The first quartile is defined as the middle number between the smallest number and the median of the data set.

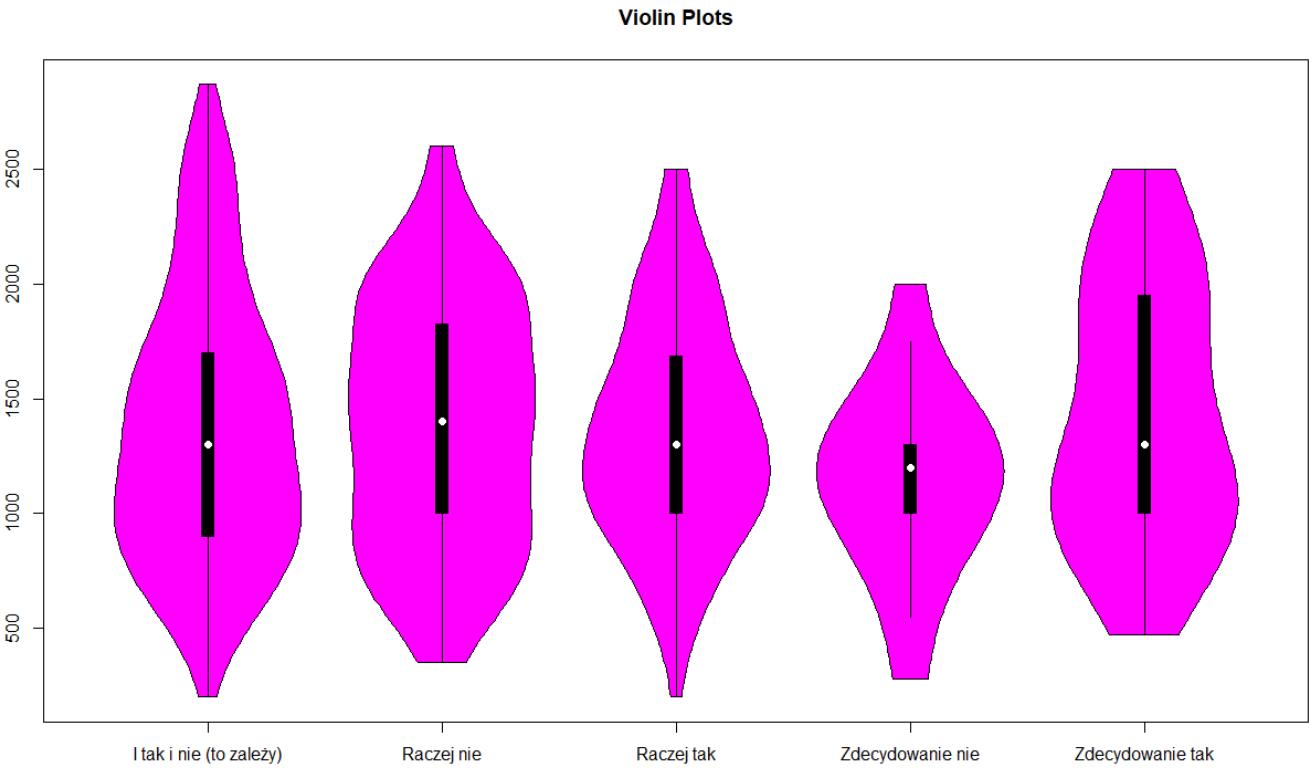
[4] The third quartile is the middle value between the median and the highest value of the data set.

[5] The IQR is the first quartile subtracted from the third quartile.

[6] The standard deviation is a measure that is used to quantify the amount of variation or dispersion of a set of data values. A low standard deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values.



The plot shows us the frequency distribution in each class.

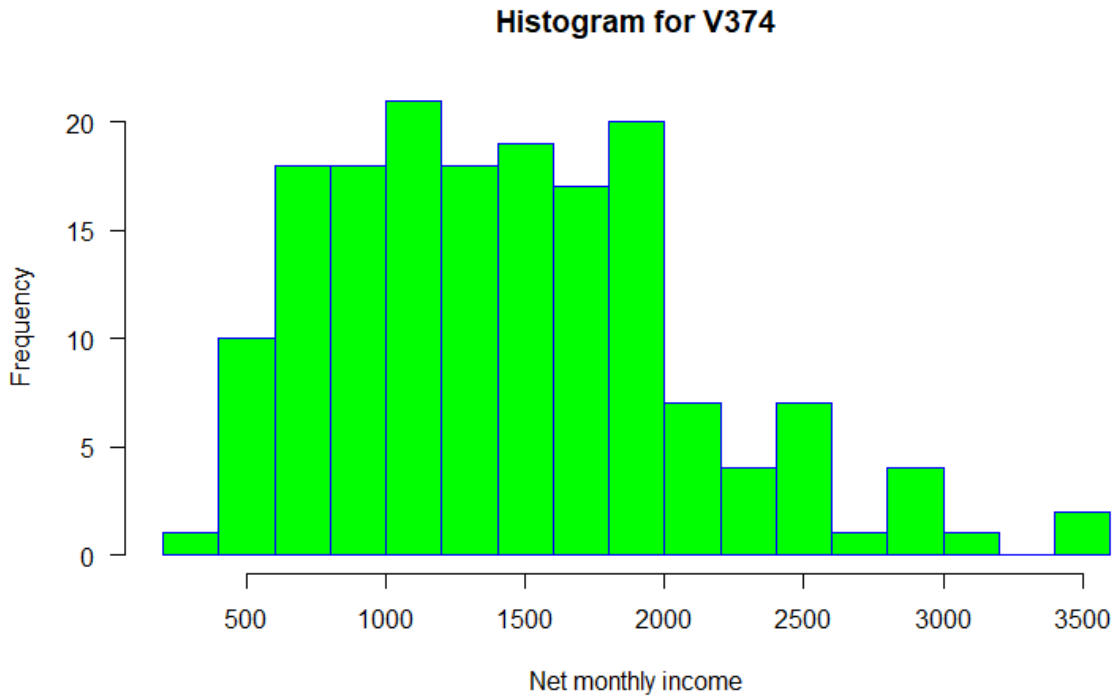


A violin plot is a combination of a box plot and a kernel density plot. Specifically, it starts with a box plot. It then adds a rotated kernel density plot to each side of the box plot. The plot show information which was described before.

II. Statistics for answer: "I tak i nie (to zależy)" in question V39

Intervals	n	%
[0,500]	7	4,166667
(500,900]	28	16,66667
(900,1100]	23	13,69048
(1100,1300]	18	10,71429
(1300,1550]	23	13,69048
(1550,1900]	24	14,28571
(1900,2200]	26	15,47619
(2200,5000]	19	11,30952

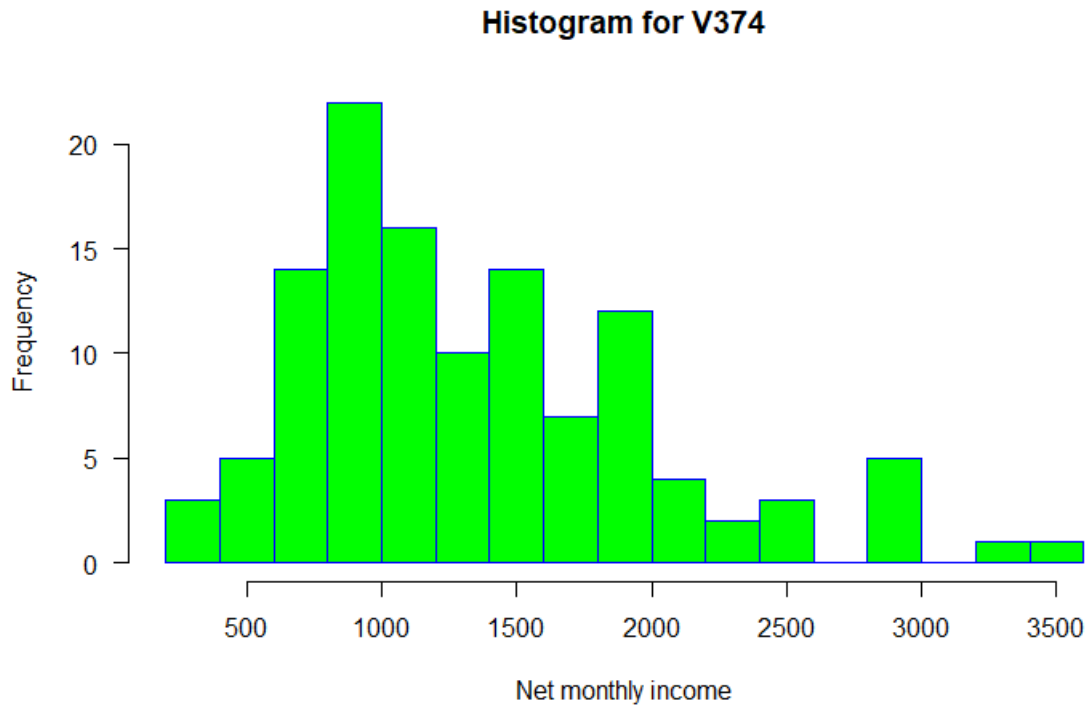
Goodness of fit = 0.9477837 Tabular accuracy = 0.8204318



III. Statistics for answer: "Raczej nie" in question V39

Intervals	n	%
[0,500]	6	5,042017
(500,900]	23	19,32773
(900,1100]	23	19,32773
(1100,1300]	14	11,76471
(1300,1550]	16	13,44538
(1550,1900]	11	9,243697
(1900,2200]	14	11,76471
(2200,5000]	12	10,08403

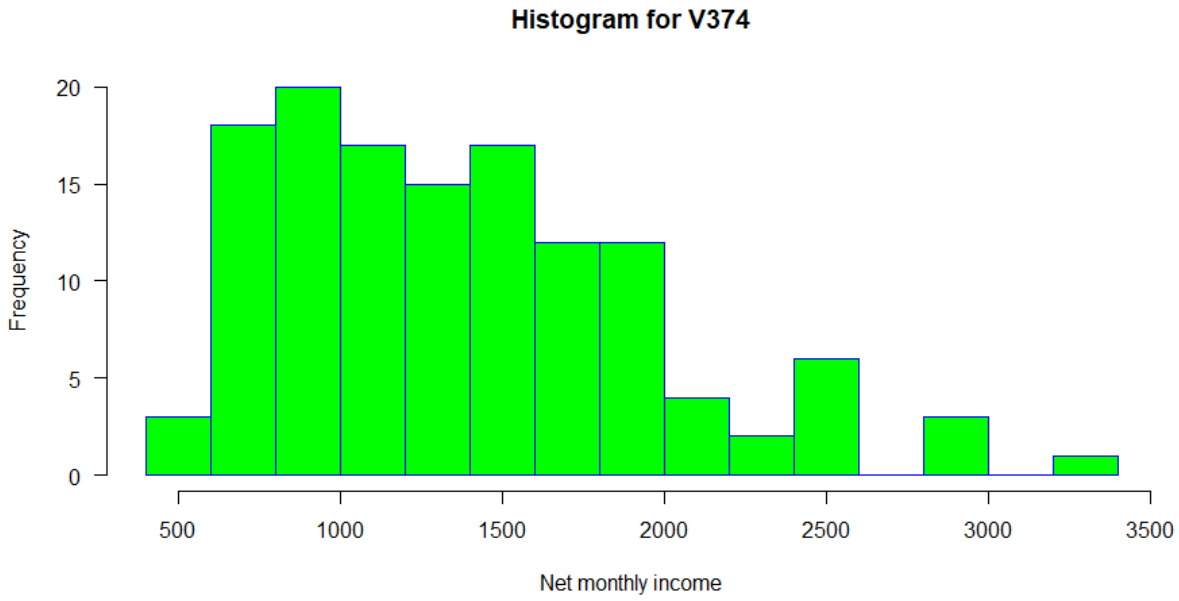
Goodness of fit = 0.9588010 Tabular accuracy = 0.8339149



IV. Statistics for answer: "Raczej tak " in question V39

Intervals	n	%
[0,500]	1	0,769231
(500,900]	23	17,69231
(900,1100]	24	18,46154
(1100,1300]	18	13,84615
(1300,1550]	17	13,07692
(1550,1900]	23	17,69231
(1900,2200]	12	9,230769
(2200,5000]	12	9,230769

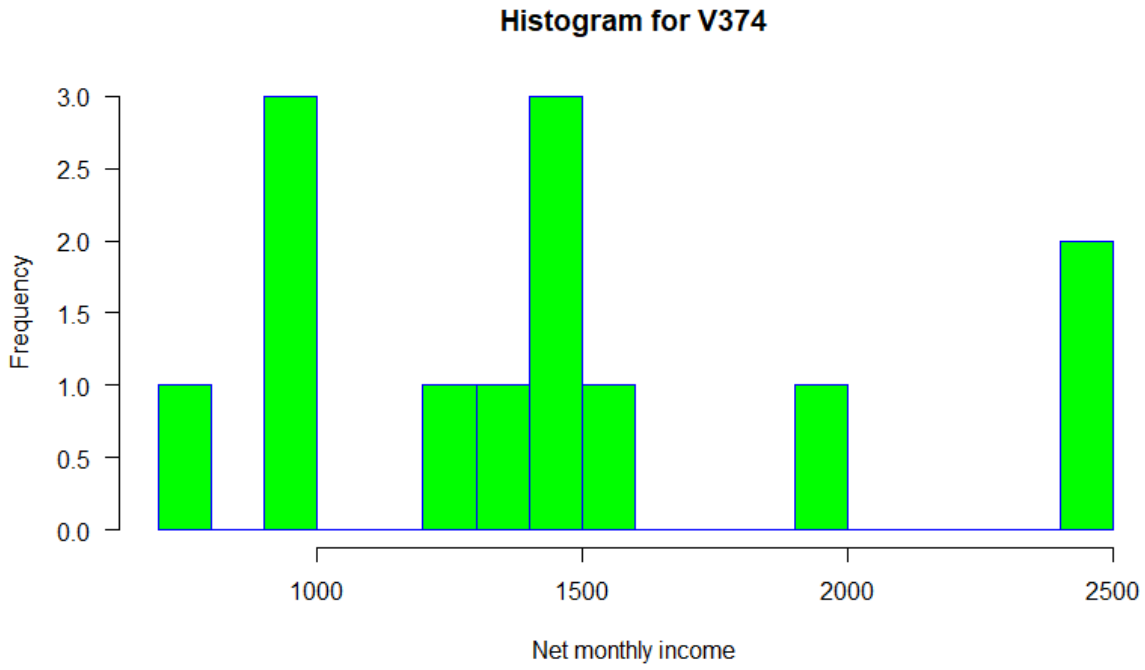
Goodness of fit = 0.9601527 Tabular accuracy = 0.8259087



V. Statistics for answer: "Zdecydowanie nie " in question V39

Intervals	n	%
[0,500]	0	0
(500,900]	1	7,692308
(900,1100]	3	23,07692
(1100,1300]	1	7,692308
(1300,1550]	4	30,76923
(1550,1900]	1	7,692308
(1900,2200]	1	7,692308
(2200,5000]	2	15,38462

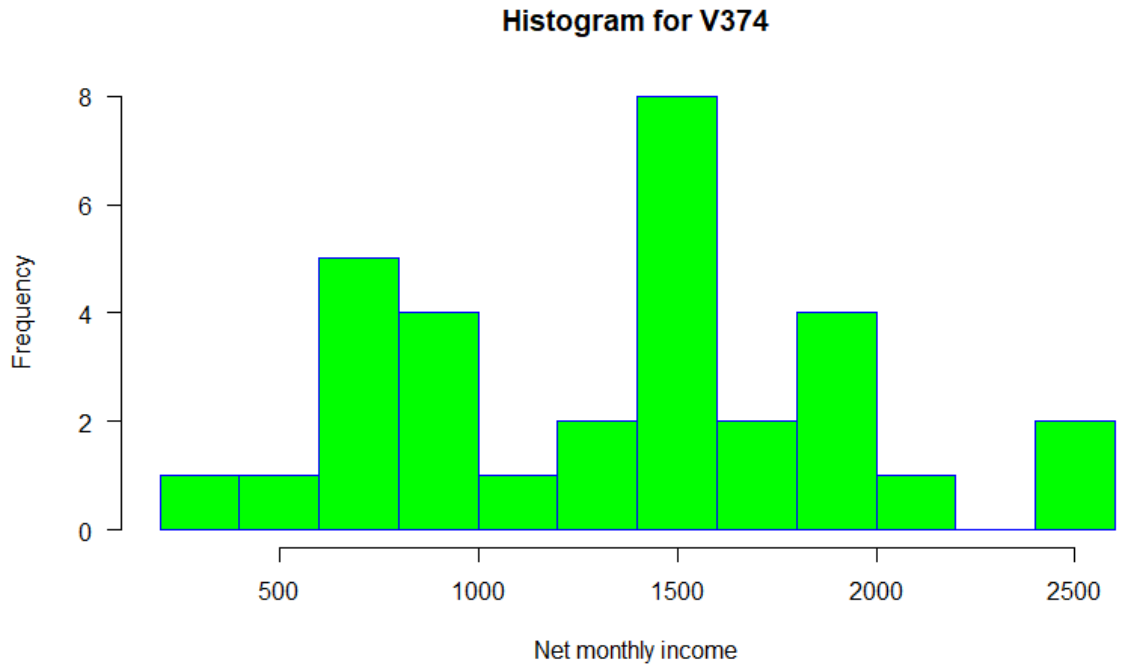
Goodness of fit = 0.9981145 Tabular accuracy = 0.9724674



VI. Statistics for answer: "Zdecydowanie tak " in question V39

Intervals	n	%
[0,500]	2	6,451613
(500,900]	6	19,35484
(900,1100]	4	12,90323
(1100,1300]	1	3,225806
(1300,1550]	8	25,80645
(1550,1900]	3	9,677419
(1900,2200]	5	16,12903
(2200,5000]	2	6,451613

Goodness of fit = 0.9981145 Tabular accuracy = 0.9724674



3. Appendices (R-code)

```
data=read.csv("C:\\Users\\rober\\Desktop\\CBOS_ogolnopolski.csv", sep=";", dec=",")
woj = data[,1]
levels(woj)
data = data[woj=="WARMIŃSKO-MAZURSKIE",]
V374 = data[,371]
#(V374)I changed all not numeric values on NA in Excel
      V374=V374[which(V374!='NA')]
      data = data[which(V374!='NA'),]
      hist(V374)
      V374=V374[V374<quantile(V374,0.98)]
      data=data[V374<quantile(V374,0.98),]
      hist(V374, main="Histogram for V374", xlab="Net monthly income", border="blue", col="green",las=1, breaks=15)
      plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The cumulative percentage of monthly income')
      library(classInt)
      tabelka=classIntervals(V374, style='fixed', fixedBreaks = c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
      jenks.tests(tabelka)
      mean(V374)
      median(V374)
      min(V374)
      max(V374)
      quantile(V374,0.25)
      quantile(V374,0.75)
      IQR(V374)
      sd(V374)
      pyt = data[,36]
      levels(pyt)
      pyt=pyt[which(pyt!='')]
      prop.table(summary(pyt))*100
      p = c("7%","36%","26%","28%","3%")
      lbl = c("Zdecydowanie tak","I tak i nie (to zależy)", "Raczej nie" , "Raczej tak","Zdecydowanie nie")
      pie(summary(pyt), main="Answers for V39",labels = lb,col = rainbow(length(lb)))

      tabelka=classIntervals(V374_odp1, style='fixed', fixedBreaks = c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
      plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The cumulative percentage of monthly income', col = "black")

      tabelka=classIntervals(V374_odp2, style='fixed', fixedBreaks = c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
      plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The cumulative percentage of monthly income',col = "red",add =T)

      tabelka=classIntervals(V374_odp3, style='fixed', fixedBreaks = c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
      plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The cumulative percentage of monthly income',col="blue",add=T)

      tabelka=classIntervals(V374_odp4, style='fixed', fixedBreaks = c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
      plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The cumulative percentage of monthly income',col="green",add=T)

      tabelka=classIntervals(V374_odp5, style='fixed', fixedBreaks = c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
```

```
49.      plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The
   cumulative percentage of monthly income',col="orange",add=T)
50.
51.      legend(y=0,'right', legend=c("I tak i nie (to zależy)", "Raczej
   nie","Raczej tak","Zdecydowanie nie","Zdecydowanie tak"),col=c("black","red",
   "blue","green","orange"), lty=1:2)
52.
53.      vioplot::vioplot(V374_odp1,V374_odp2,V374_odp3,V374_odp4,V374_odp5, names =
   c("I tak i nie (to zależy)", "Raczej nie","Raczej tak","Zdecydowanie
   nie","Zdecydowanie tak"))
54.      title("Violin Plots")
55.
56.      odp1 = which(pyt=="I tak i nie (to zależy)")
57.      V374_odp1 = V374[odp1]
58.      tabelka=classIntervals(V374_odp1, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
59.      tabelka
60.      jenks.tests(tabelka)
61.      hist(V374_odp1, main="Histogram for V374", xlab="Net monthly income",
   border="blue", col="green",las=1, breaks=15)
62.      mean(V374_odp1)
63.      median(V374_odp1)
64.      min(V374_odp1)
65.      max(V374_odp1)
66.      quantile(V374_odp1)
67.      quantile(V374_odp1)
68.      IQR(V374_odp1)
69.      sd(V374_odp1)
70.
71.      odp2 = which(pyt=="Raczej nie")
72.      V374_odp2 = V374[odp2]
73.      tabelka=classIntervals(V374_odp2, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
74.      tabelka
75.      jenks.tests(tabelka)
76.      hist(V374_odp2, main="Histogram for V374", xlab="Net monthly income",
   border="blue", col="green",las=1, breaks=15)
77.      mean(V374_odp2)
78.      median(V374_odp2)
79.      min(V374_odp2)
80.      max(V374_odp2)
81.      quantile(V374_odp2)
82.      quantile(V374_odp2)
83.      IQR(V374_odp2)
84.      sd(V374_odp2)
85.
86.      odp3 = which(pyt=="Raczej tak")
87.      V374_odp3 = V374[odp3]
88.      tabelka=classIntervals(V374_odp3, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
89.      tabelka
90.      jenks.tests(tabelka)
91.      hist(V374_odp3, main="Histogram for V374", xlab="Net monthly income",
   border="blue", col="green",las=1, breaks=15)
92.      mean(V374_odp3)
93.      median(V374_odp3)
94.      min(V374_odp3)
95.      max(V374_odp3)
96.      quantile(V374_odp3)
97.      IQR(V374_odp3)
98.      sd(V374_odp3)
```

```
99.
100.     odp4 = which(pyt=="Zdecydowanie nie")
101.     V374_odp4 = V374[odp4]
102.     tabelka=classIntervals(V374_odp4, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
103.     tabelka
104.     jenks.tests(tabelka)
105.     hist(V374_odp4, main="Histogram for V374", xlab="Net monthly income",
   border="blue", col="green",las=1, breaks=15)
106.     mean(V374_odp4)
107.     median(V374_odp4)
108.     min(V374_odp4)
109.     max(V374_odp4)
110.     quantile(V374_odp4)
111.     IQR(V374_odp4)
112.     sd(V374_odp4)
113.
114.     odp5 = which(pyt=="Zdecydowanie tak")
115.     V374_odp5 = V374[odp5]
116.     tabelka=classIntervals(V374_odp5, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
117.     tabelka
118.     jenks.tests(tabelka)
119.     hist(V374_odp5, main="Histogram for V374", xlab="Net monthly income",
   border="blue", col="green",las=1, breaks=15)
120.     mean(V374_odp5)
121.     median(V374_odp5)
122.     min(V374_odp5)
123.     max(V374_odp5)
124.     quantile(V374 odp5)
125.     IQR(V374_odp5)
126.     sd(V374_odp5)
127.
128.     tabelka=classIntervals(V374_odp1, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
129.     plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The
   cumulative percentage of monthly income', col = "black")
130.     tabelka=classIntervals(V374_odp2, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
131.     plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The
   cumulative percentage of monthly income',col = "red",add =T)
132.     tabelka=classIntervals(V374_odp3, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
133.     plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The
   cumulative percentage of monthly income',col="blue",add=T)
134.     tabelka=classIntervals(V374_odp4, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
135.     plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The
   cumulative percentage of monthly income',col="green",add=T)
136.     tabelka=classIntervals(V374_odp5, style='fixed', fixedBreaks =
   c(0,500,900,1100,1300,1550,1900,2200,5000), intervalClosure='right')
137.     plot(tabelka, pal=colcode , main='V374', xlab='monthly income', ylab='The
   cumulative percentage of monthly income',col="orange",add=T)
138.     legend(y=0,'right', legend=c("I tak i nie (to zależy)", "Raczej
   nie","Raczej tak","Zdecydowanie nie","Zdecydowanie tak"),col=c("black","red",
   "blue","green","orange"), lty=1:2)
139.     vioplot::vioplot(V374_odp1,V374_odp2,V374_odp3,V374_odp4,V374_odp5, names =
   c("I tak i nie (to zależy)", "Raczej nie","Raczej tak","Zdecydowanie
   nie","Zdecydowanie tak"))
140.     title("Violin Plots")
```