# Homework1

Darian-Florian Voda

2022-10-11

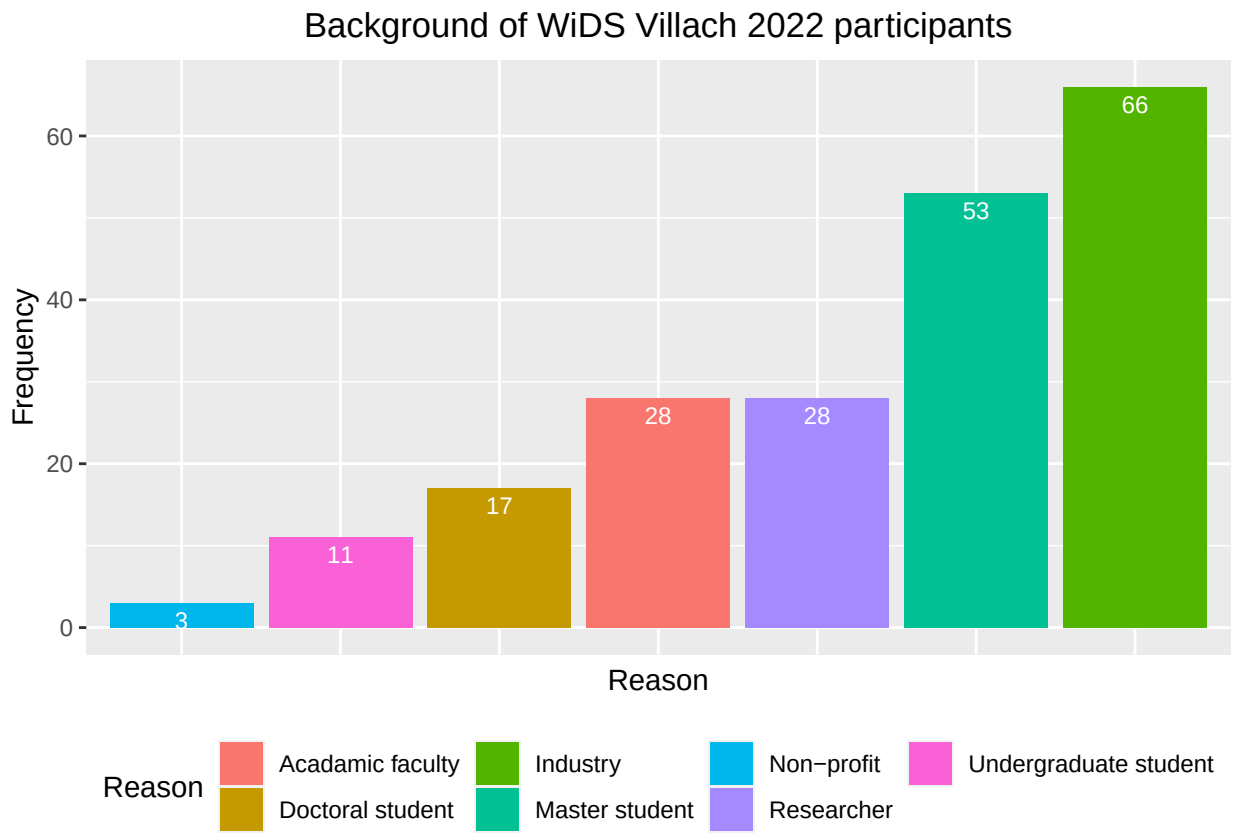
## R Markdown

## Exercise 6:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
ex6 = data.frame(Frequency= c(66, 53, 28, 28, 17, 11, 3),
                 Reason=c("Industry", "Master student",
                          "Acadamic faculty", "Researcher",
                          "Doctoral student", "Undergraduate student",
                          "Non-profit"))

plot1 = ggplot(ex6, aes(x=reorder(Reason, Frequency),
                y=Frequency, fill=Reason),
               geom_text(aes(label=ex6$Reason), vjust=1.6, color="white",
                         position = position_dodge(0.9), size=3.5))

plot1 + geom_col() + geom_text(
                       aes(label = Frequency),
                       colour = "white", size = 3,
                       vjust = 1.5, position = position_dodge(.9)) +
                       ggtitle("Background of WiDS Villach 2022 participants") +
                       theme(plot.title = element_text(hjust = 0.5),
                       axis.text.x = element_blank(),
                       axis.ticks.x = element_blank(),
                       legend.position="bottom") + labs(x="Reason")
```




## Exercise 7:

```
# import data
students<-read.delim("C:/Users/daria/OneDrive/Desktop/Master - AppDS/Statistics/Datasets-20221007/stude
                     stringsAsFactors=F)

# import libraries
library(mosaic)
```

```
## Warning: package 'mosaic' was built under R version 4.1.3
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2
```

```
##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.
```

```
##
## Attaching package: 'mosaic'
```

```
## The following objects are masked from 'package:dplyr':
##
##     count, do, tally

## The following object is masked from 'package:Matrix':
##
##     mean

## The following object is masked from 'package:ggplot2':
##
##     stat

## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##     quantile, sd, t.test, var

## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```

```
library(ggplot2)
library(ggrepel)
```

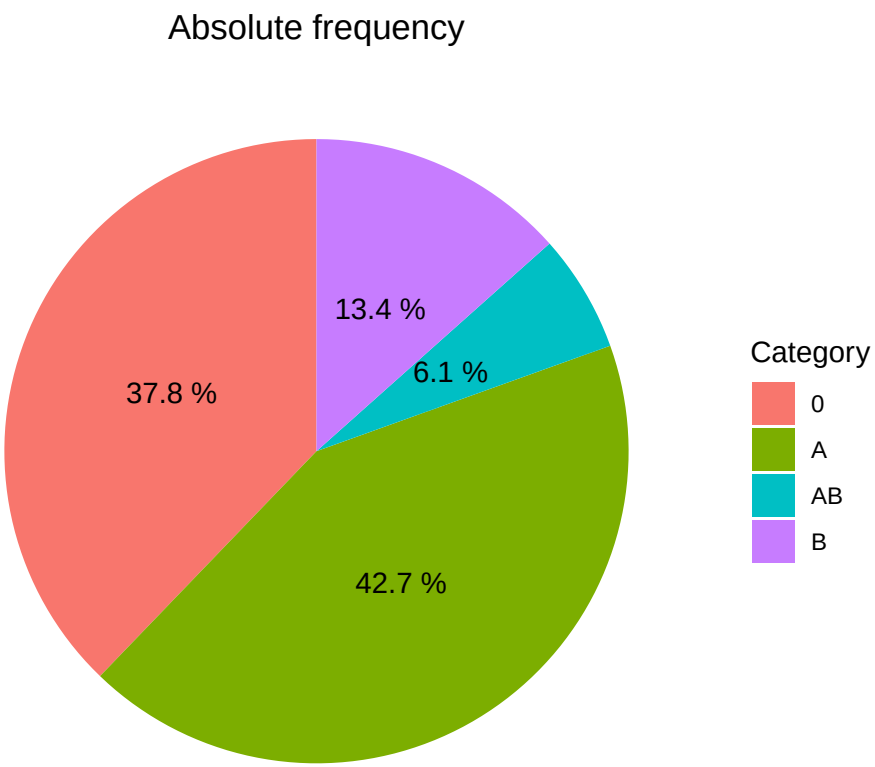
```
## Warning: package 'ggrepel' was built under R version 4.1.3
```

```
# absolute frequency
tally(~Blood_group, data=students)
```

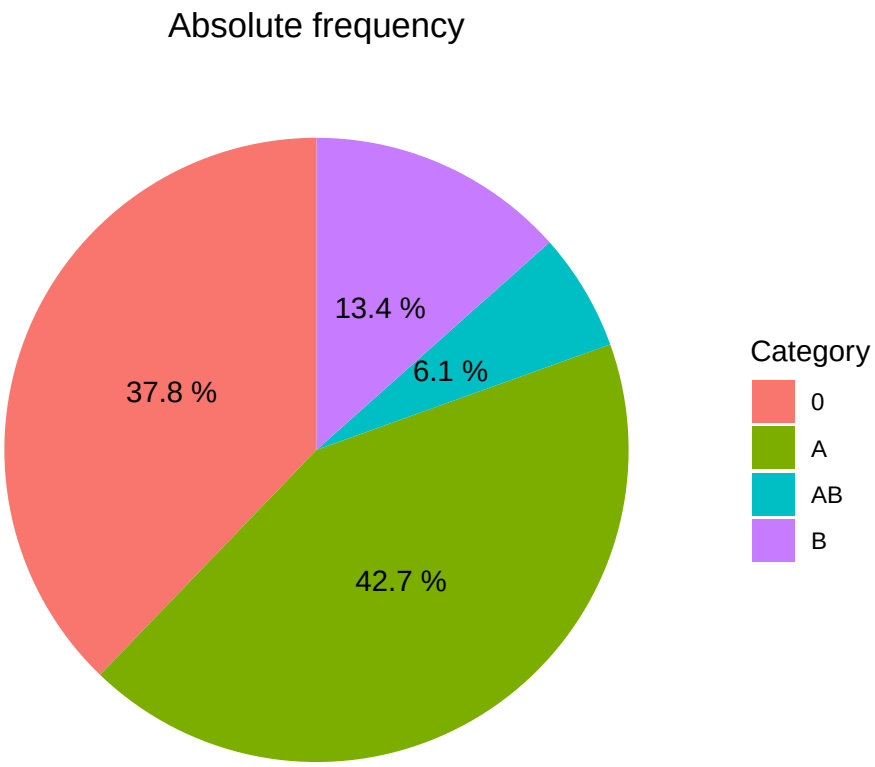
```
## Blood_group
##  O  A AB  B
## 31 35  5 11
```

```
# relative frequency
p0 = prop(~Blood_group, success = "O", data = students)
pA = prop(~Blood_group, success = "A", data = students)
pAB = prop(~Blood_group, success = "AB", data = students)
pB = prop(~Blood_group, success = "B", data = students)

# create plot
blood_pie <- c(31,35,5,11)
df = data.frame(value = blood_pie, group = c("O", "A", "AB", "B"))
ggplot(df, aes (x="", y = value, fill = factor(group))) +
  geom_bar(width = 1, stat = "identity") +
  geom_text(aes(label = paste(round(value / sum(value) * 100, 1), "%")),
            position = position_stack(vjust = 0.5)) +
  theme_classic() +
  theme(plot.title = element_text(hjust=0.5),
        axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank()) +
```

```
  labs(fill = "Category",
       x = NULL,
       y = NULL,
       title = "Absolute frequency") +
  coord_polar("y")
```

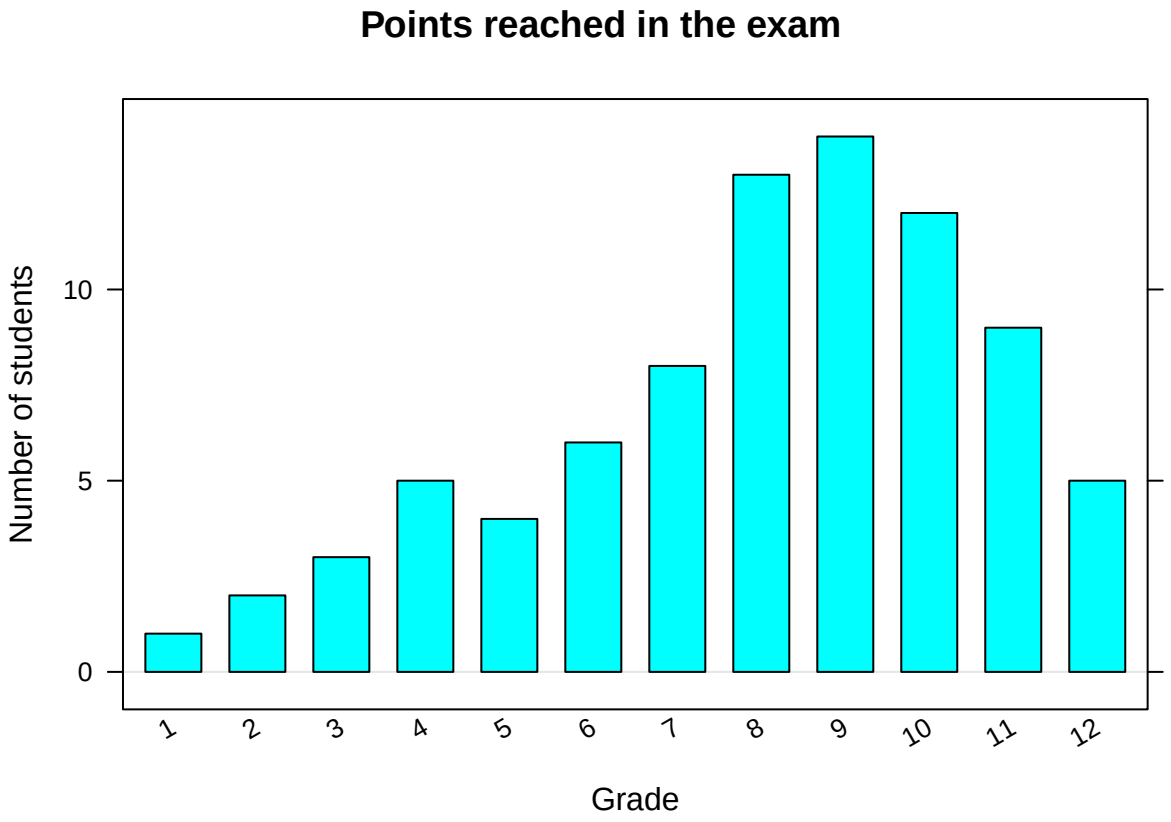



```
# second plot
relative_freq = c(p0, pA, pAB, pB)
df2 = data.frame(value = relative_freq, group = c("0", "A", "AB", "B"))
ggplot(df2, aes (x="", y = value, fill = factor(group))) +
  geom_bar(width = 1, stat = "identity") +
  geom_text(aes(label = paste(round(value / sum(value) * 100, 1), "%")),
            position = position_stack(vjust = 0.5)) +
  theme_classic() +
  theme(plot.title = element_text(hjust=0.5),
        axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank()) +
  labs(fill = "Category",
       x = NULL,
       y = NULL,
       title = "Absolute frequency") +
  coord_polar("y")
```



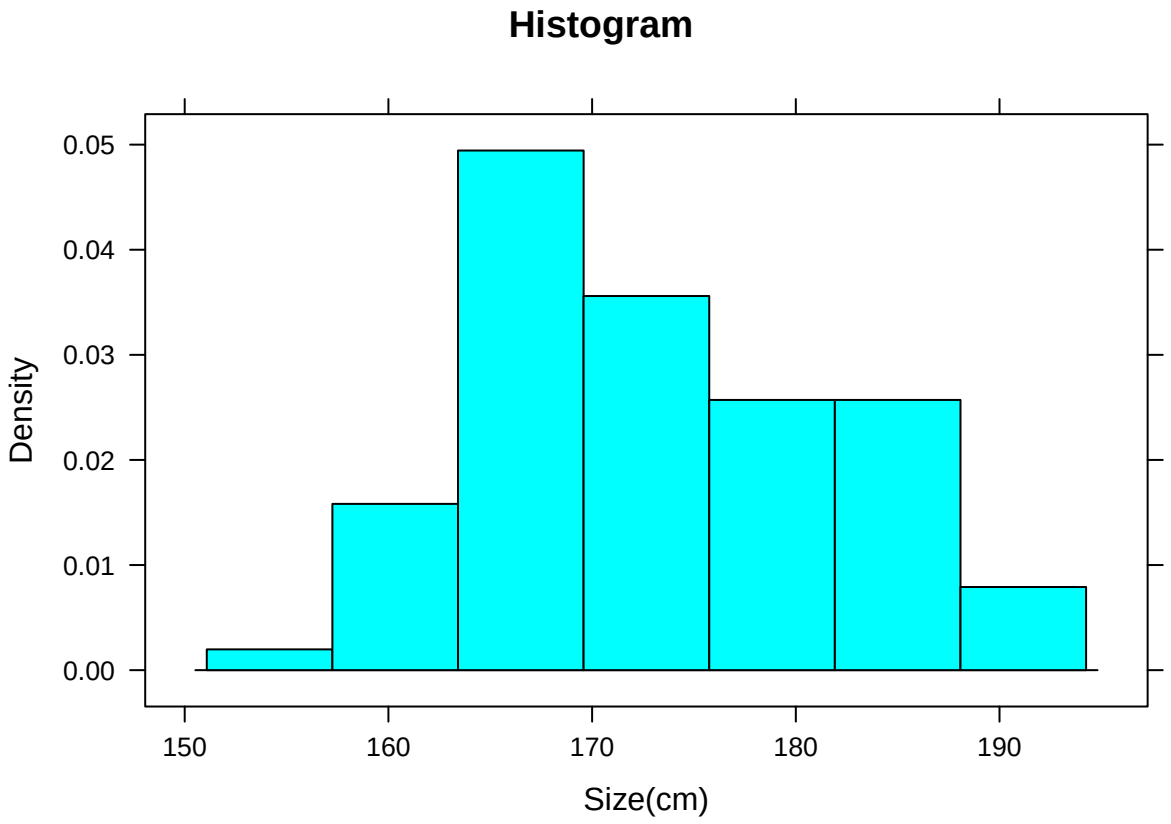

## Exercise 8:

```
students<-read.delim("C:/Users/daria/OneDrive/Desktop/Master - AppDS/Statistics/Datasets-20221007/stude
library(mosaic)
bargraph(~Points_exam, data=students,
         main="Points reached in the exam",
         xlab="Grade",
         ylab="Number of students")
```




## Exercise 9:

```
students<-read.delim("C:/Users/daria/OneDrive/Desktop/Master - AppDS/Statistics/Datasets-20221007/stude
library(mosaic)
histogram(~Size_cm, data = students,
          main="Histogram",
          xlab="Size(cm)",
          ylab="Density")
```



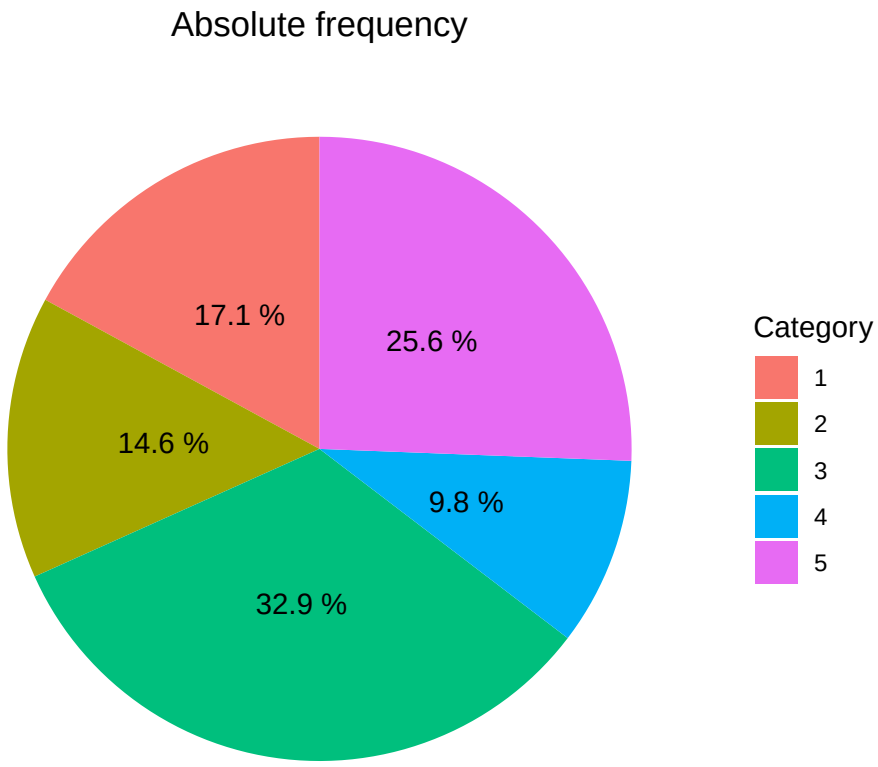

## Exercise 10:

```
# Exercise 10
students<-read.delim("C:/Users/daria/OneDrive/Desktop/Master - AppDS/Statistics/Datasets-20221007/stude
library(mosaic)

# absolute frequency
tally(~Grade, data=students)
```

```
## Grade
##  1  2  3  4  5
## 14 12 27  8 21
```

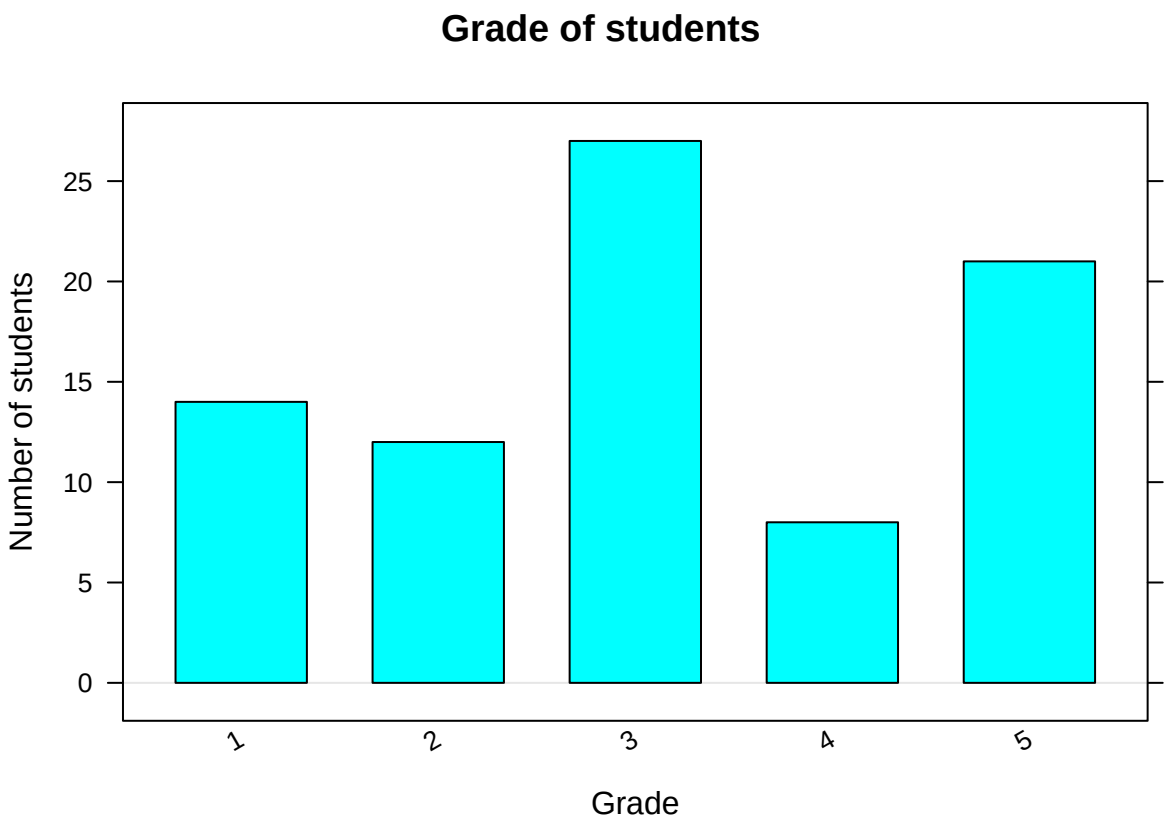
```
# relative frequency
p1 = prop(~Grade, success = "1", data = students)
p2 = prop(~Grade, success = "2", data = students)
p3 = prop(~Grade, success = "3", data = students)
p4 = prop(~Grade, success = "4", data = students)
p5 = prop(~Grade, success = "5", data = students)


# Pie chart
```

```
blood_pie <- c(p1,p2,p3,p4,p5)
df = data.frame(value = blood_pie, group = c("1", "2", "3", "4", "5"))
ggplot(df, aes (x="", y = value, fill = factor(group))) +
  geom_bar(width = 1, stat = "identity") +
  geom_text(aes(label = paste(round(value / sum(value) * 100, 1), "%")),
            position = position_stack(vjust = 0.5)) +
  theme_classic() +
  theme(plot.title = element_text(hjust=0.5),
        axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank()) +
  labs(fill = "Category",
       x = NULL,
       y = NULL,
       title = "Absolute frequency") +
  coord_polar("y")
```



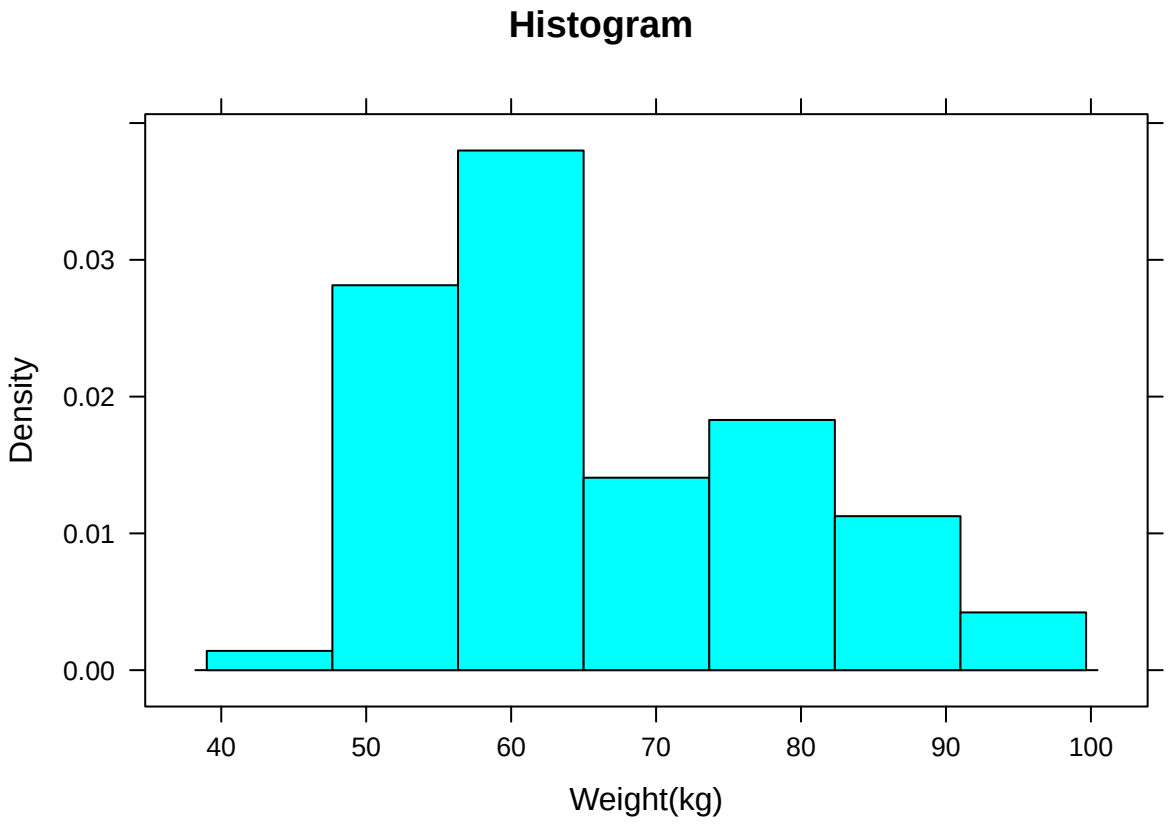

Bar graph

```
bargraph(~Grade, data=students,
         main="Grade of students",
         xlab="Grade",
         ylab="Number of students")
```




## Exercise 11

```
# Exercise 11
students<-read.delim("C:/Users/daria/OneDrive/Desktop/Master - AppDS/Statistics/Datasets-20221007/stude
library(mosaic)
histogram(~Weight_kg, data = students,
          main="Histogram",
          xlab="Weight(kg)",
          ylab="Density")
```


Histogram
Density
0.00
0.01
0.02
0.03
40
50
60
70
80
90
100
Weight(kg)