Assignment 1

Majeti Arbuda Sivani

Jan 30, 2022

Contents


```
library(knitr)
library(rmarkdown)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tinytex)
```

Import data

```
Data <- read.csv("~/R/Dataset/Assignment_1/CD_additional_balanced-1.csv")
```

Descriptive Statistics for Quantitative Varibles

```
summary(Data$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    17.0    31.0    38.0    40.4    48.0    98.0
```

```
mean(Data$age)
```

```
## [1] 40.40345
```

Descriptive Statistics for Categorical Variables

```
table(Data$marital)
```

```
##
## divorced  married   single  unknown
##     1021     5338     2900       21
```

```
prop.table(table(Data$marital))
```

```
##
##    divorced     married      single     unknown
## 0.110021552 0.575215517 0.312500000 0.002262931
```

```
table(Data$job)
```

```
##
##        admin.   blue-collar  entrepreneur     housemaid    management
##          2517          1769           308           216           651
##       retired self-employed      services       student    technician
##           595           306           773           358          1459
##    unemployed       unknown
##           248            80
```

```
prop.table(table(Data$job))
```

```
##
##        admin.   blue-collar  entrepreneur     housemaid    management
##    0.27122845    0.19062500    0.03318966    0.02327586    0.07015086
##       retired self-employed      services       student    technician
##    0.06411638    0.03297414    0.08329741    0.03857759    0.15721983
##    unemployed       unknown
##    0.02672414    0.00862069
```

Data Transformation

```r
# Grouping by level of education to see avg scores
group_by_scores <- Data %>% group_by(marital) %>%
  summarize(
    Avg_age = mean(age)
  )
group_by_scores
```

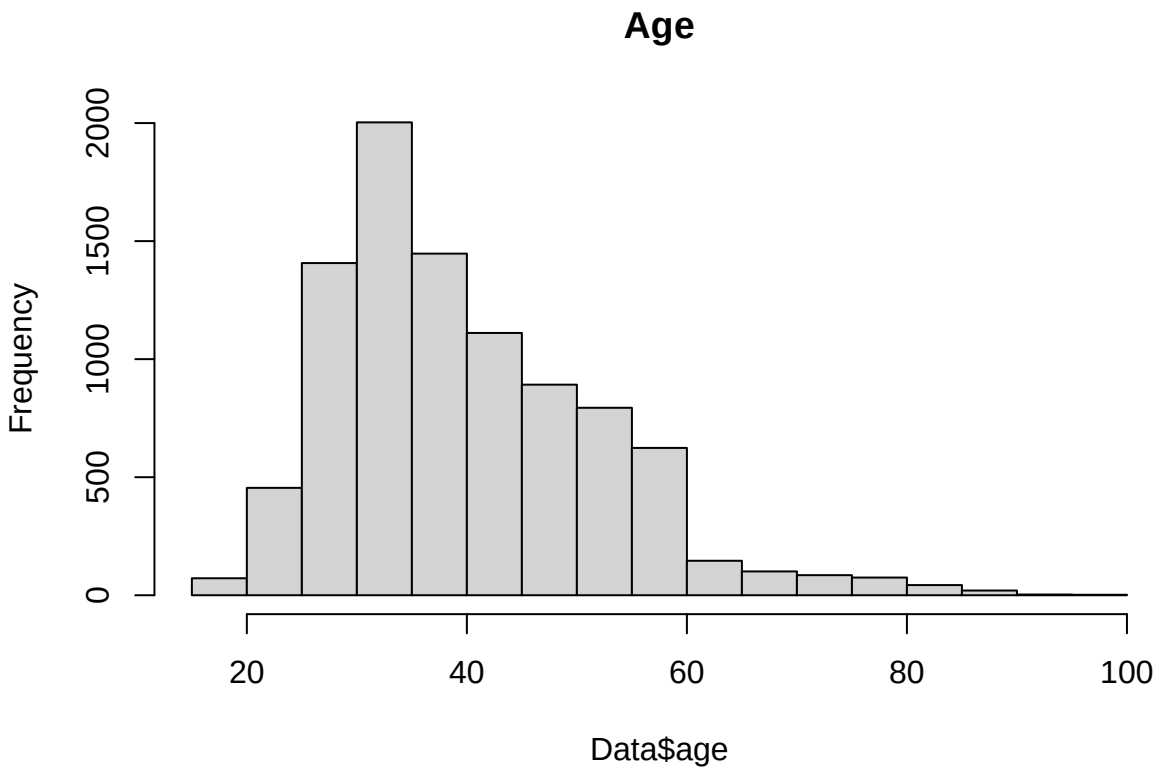
```
## # A tibble: 4 x 2
##   marital  Avg_age
##   <chr>      <dbl>
## 1 divorced    47.3
## 2 married     43.5
## 3 single      32.2
## 4 unknown     39.7
```

Plotting one quantitative variable and one scatterplot

```r
library(ggplot2)

# Plotting a histogram for the reading score variable

hist(Data$age, main = "Age")
```



```
# Scatter plot of math score variable

ggplot(Data,aes(x=Data$age, y=Data$duration)) + geom_point()
```

```
## Warning: Use of ‘Data$age‘ is discouraged. Use ‘age‘ instead.

## Warning: Use of ‘Data$duration‘ is discouraged. Use ‘duration‘ instead.
```