Your turn

- Load the data in file `sp500_data.csv` into your statistical software package
- This file contains data on companies in the *S&P 500 Index* from *Bloomberg*

- Get an overview about the available data
- Prepare *descriptive statistics* about the data at hand
- Try to prepare summary statistics by industry/sector
- Try to prepare figures describing one or more variables in the data set you find interesting

- Make sure to use a `script` and *document* how you solved the task

Your turn

```r
data <- read.csv("data/sp500_data.csv")
```

```r
str(data)
```

```
## 'data.frame':    505 obs. of  21 variables:
##  $ LAST_UPDATE_DT              : Factor w/ 1 level "2019-11-18": 1 1 1 1 1 1 1 1 1 1 ...
##  $ NAME                        : Factor w/ 505 levels "3M CO","ABBOTT LABORATORIES",..: 13 32 8 49 3 39
##  $ INDUSTRY_SECTOR             : Factor w/ 9 levels "Basic Materials",..: 7 3 3 8 4 4 4 4 8 8 ...
##  $ INDUSTRY_GROUP              : Factor w/ 60 levels "Advertising",..: 20 4 51 14 47 47 27 27 14 55 ...
##  $ INDUSTRY_SUBGROUP           : Factor w/ 183 levels "Advertising Agencies",..: 68 7 149 40 106 110 114
##  $ PX_LAST                     : num  77.9 28.6 163.7 267.1 88.7 ...
##  $ CHG_NET_YTD                 : num  10.48 -3.51 6.21 109.36 -3.46 ...
##  $ EQY_DVD_YLD_IND             : num  0.842 1.399 0.147 1.153 5.32 ...
##  $ EQY_BETA                    : num  1.111 1.261 0.927 1.065 1.157 ...
##  $ PE_RATIO                    : num  28.11 5.95 21.62 22.53 11.43 ...
##  $ CUR_MKT_CAP                 : num  2.41e+10 1.25e+10 1.13e+10 1.19e+12 1.31e+11 ...
##  $ BOARD_SIZE                  : int  10 12 10 8 11 10 7 13 12 11 ...
##  $ PCT_WOMEN_ON_BOARD          : num  20 16.7 20 25 27.3 ...
##  $ NUMBER_OF_WOMEN_ON_BOARD    : int  2 2 2 2 3 3 2 5 4 2 ...
##  $ BOARD_AVERAGE_TENURE        : num  7.5 6.6 3.77 10.71 5.8 ...
##  $ BOARD_AVERAGE_AGE           : num  64.2 64 59.5 65.1 63.9 ...
##  $ NUMBER_OF_DIRECTORS_ON_BOARD: int  11 10 11 7 11 9 7 14 12 11 ...
##  $ PERCENTAGE_OF_FEMALE_EXECUTIVES : num  0.33 3 0.40 18.2
```

Your turn

```
summary(data)
```

```
##      LAST_UPDATE_DT                    NAME                      INDUSTRY_SECTOR
##  2019-11-18:505    3M CO                  :  1   Consumer, Non-cyclical:106
##                    ABBOTT LABORATORIES    :  1   Financial             : 99
##                    ABBVIE INC             :  1   Consumer, Cyclical    : 70
##                    ABIOMED INC            :  1   Industrial            : 67
##                    ACCENTURE PLC-CL A     :  1   Technology            : 51
##                    ACTIVISION BLIZZARD INC:  1   Communications        : 37
##                    (Other)                :499   (Other)               : 75
##           INDUSTRY_GROUP              INDUSTRY_SUBGROUP    PX_LAST
##  REITS               : 31   Electric-Integrated   : 22   Min.   :   8.95
##  Retail              : 29   Oil Comp-Explor&Prodtn: 13   1st Qu.:  49.25
##  Electric            : 25   Electronic Compo-Semicon: 11  Median :  88.73
##  Banks               : 23   Commercial Serv-Finance : 10  Mean   : 134.82
##  Insurance           : 22   Medical Products      : 10   3rd Qu.: 147.63
##  Diversified Finan Serv: 21  Medical-Biomedical/Gene :  9  Max.   :3652.64
##  (Other)             :354   (Other)               :430
##   CHG_NET_YTD       EQY_DVD_YLD_IND      EQY_BETA          PE_RATIO
##  Min.   :-146.950   Min.   : 0.01963   Min.   :0.3954   Min.   :  5.499
##  1st Qu.:   5.395   1st Qu.: 1.42585   1st Qu.:0.8485   1st Qu.: 13.780
##  Median :  15.170   Median : 2.19592   Median :1.0167   Median : 20.662
##  Mean   :  28.076   Mean   : 2.44812   Mean   :0.9887   Mean   : 27.016
##  3rd Qu.:  33.490   3rd Qu.: 3.13591   3rd Qu.:1.1327   3rd Qu.: 29.485
##  Max.   :1215.650   Max.   :10.72961   Max.   :1.7004   Max.   :708.164
##  NA's   :6          NA's   :81         NA's   :11       NA's   :15
##   CUR_MKT_CAP        BOARD_SIZE    PCT_WOMEN_ON_BOARD NUMBER_OF_WOMEN_ON_BOARD
##  Min.   :3.950e+09  Min.   : 5.00  Min.   : 0.00      Min.   :0.000
```

Your turn

```
mean(data$CUR_MKT_CAP)
```

```
## [1] 56348498967
```

```
mean(data$BOARD_SIZE, na.rm=TRUE)
```

```
## [1] 10.98207
```

```
mean(data$AVERAGE_BOD_TOTAL_COMPENSATION, na.rm=TRUE)
```

```
## [1] 315443.6
```

```
sd(data$CUR_MKT_CAP)
```

```
## [1] 116900588955
```

```
sd(data$BOARD_SIZE, na.rm=TRUE)
```

```
## [1] 2.069086
```

```
sd(data$AVERAGE_BOD_TOTAL_COMPENSATION, na.rm=TRUE)
```

```
## [1] 402246.7
```

Your turn

```r
library(tidyverse)

data %>%
  group_by(INDUSTRY_SECTOR) %>%
  summarise(mean(BOARD_SIZE, na.rm=TRUE), min(BOARD_SIZE, na.rm=TRUE), max(BOARD_SIZE, na.rm=TRUE))
```

```
## # A tibble: 9 x 4
##   INDUSTRY_SECTOR   `mean(BOARD_SIZE, na.~ `min(BOARD_SIZE, na.~ `max(BOARD_SIZE, na.~
##   <fct>                             <dbl>                 <int>                 <int>
## 1 Basic Materials                    11.1                     6                    16
## 2 Communications                     10.7                     7                    15
## 3 Consumer, Cyclical                 11.0                     5                    14
## 4 Consumer, Non-cyc~                 11.0                     6                    16
## 5 Energy                             10.6                     7                    16
## 6 Financial                          11.6                     7                    21
## 7 Industrial                         10.7                     6                    14
## 8 Technology                         10.1                     5                    15
## 9 Utilities                          11.8                     8                    15
```
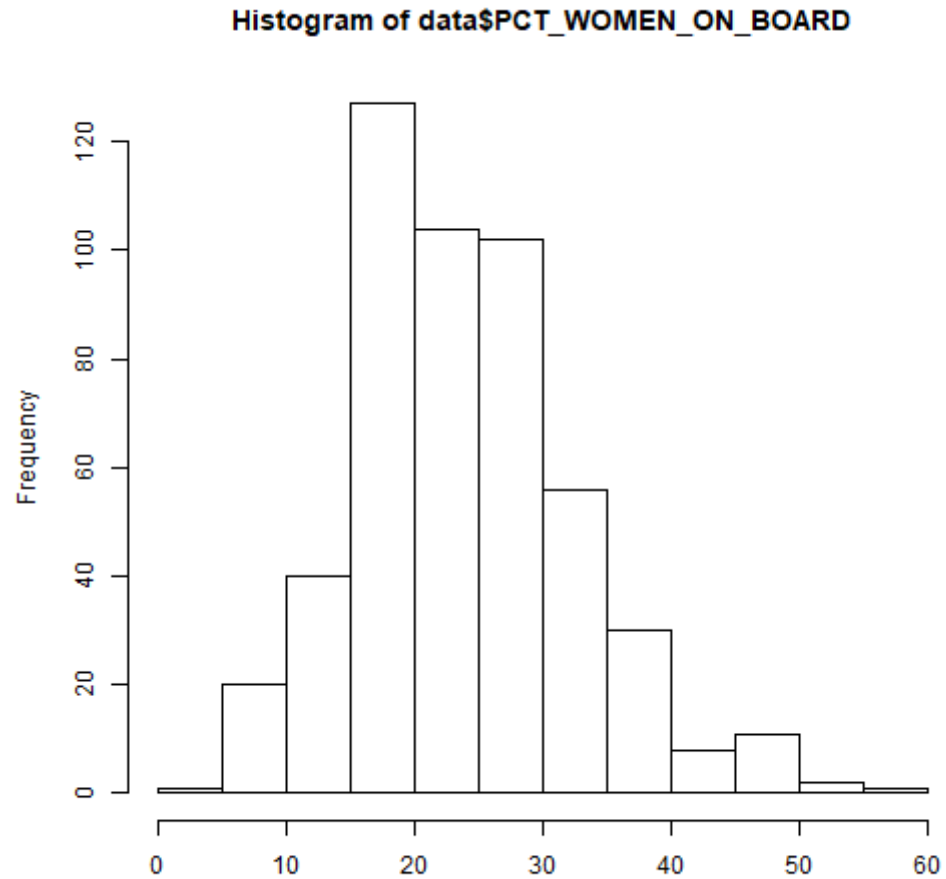
universität
innsbruck

Your turn

```r
library(tidyverse)

data %>%
  group_by(INDUSTRY_SECTOR) %>%
  summarise(mean(PCT_WOMEN_ON_BOARD, na.rm=TRUE), min(PCT_WOMEN_ON_BOARD, na.rm=TRUE), max(PCT_WOMEN_O
```

```
## # A tibble: 9 x 4
##   INDUSTRY_SECTOR    `mean(PCT_WOMEN_ON_BO~ `min(PCT_WOMEN_ON_BOA~ `max(PCT_WOMEN_ON_BO~
##   <fct>                              <dbl>                  <dbl>                  <dbl>
## 1 Basic Materials                     25.9                   15.4                   41.7
## 2 Communications                      24.6                   0                      55.6
## 3 Consumer, Cyclic~                   26.1                   8.33                   50
## 4 Consumer, Non-cy~                   25.0                   9.09                   45.5
## 5 Energy                              22.0                   9.09                   36.4
## 6 Financial                           25.4                   10                     45.5
## 7 Industrial                          23.0                   7.69                   37.5
## 8 Technology                          22.8                   8.33                   44.4
## 9 Utilities                           25.5                   14.3                   50
```
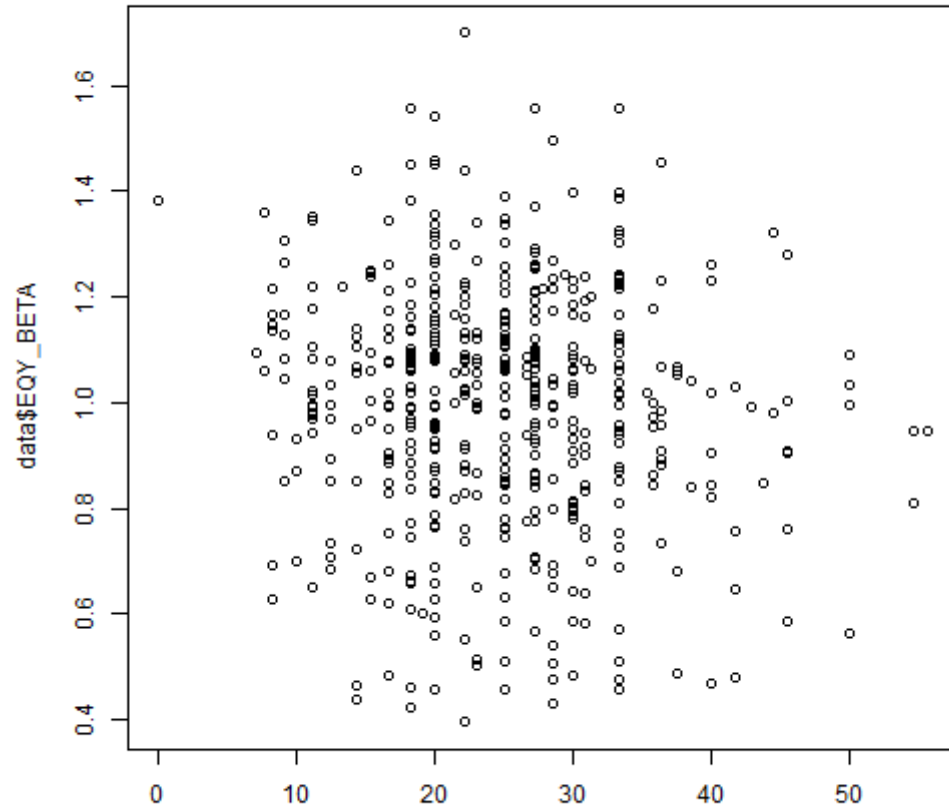
Your turn

```
hist(data$PCT_WOMEN_ON_BOARD)
```


Histogram of data$PCT_WOMEN_ON_BOARD
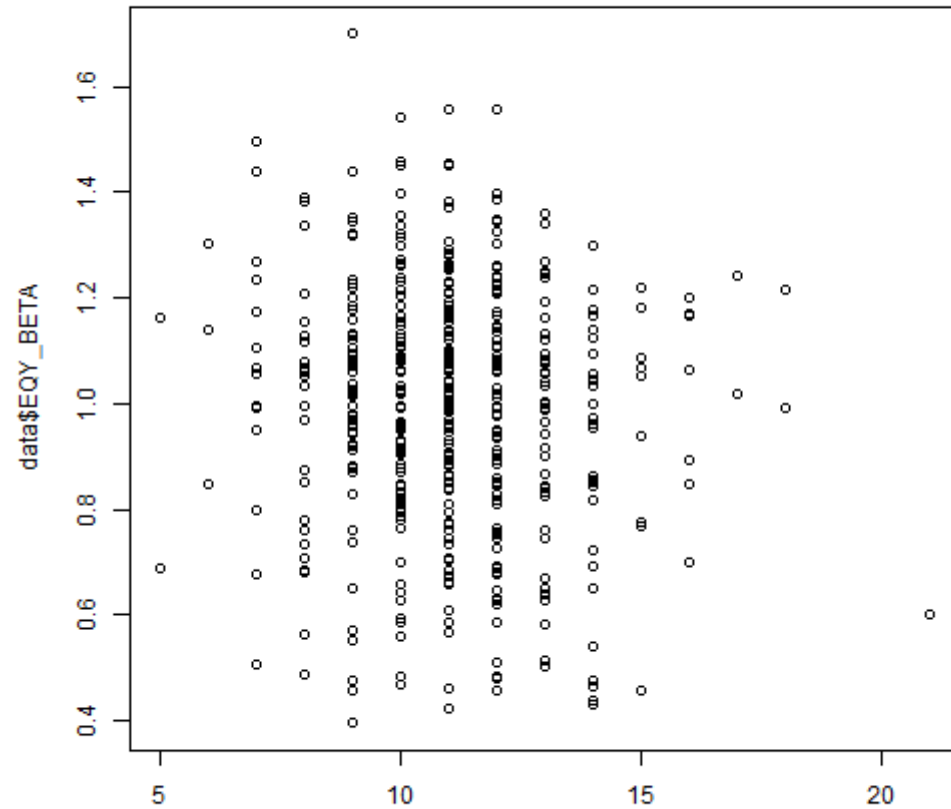
Your turn

```
plot(data$PCT_WOMEN_ON_BOARD, data$EQY_BETA)
```

Your turn

```
plot(data$BOARD_SIZE, data$EQY_BETA)
```

Your turn

```
grouped <- data %>%
  group_by(INDUSTRY_SECTOR) %>%
  summarise(mean=mean(PCT_WOMEN_ON_BOARD, na.rm=TRUE), min=min(PCT_WOMEN_ON_BOARD, na.rm=TRUE), max=max(PCT_WOMEN_ON_BOAF
```

```
grouped %>% ggplot(aes(x = INDUSTRY_SECTOR, y = mean)) +
  geom_col(width=0.7)
```