NateVelarde_ChetGutwein_Lab1

Nate Velarde & Chet Gutwein

May 30, 2017

```
rm(list = ls())
```

```
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

Introduction

Is CEO salary related to company performance?

We will attempt to answer this question utilizing a dataset of CEO salaries from 1990. We will use exploratory data analysis techniques to look at each variable by itself and also in relation to other variables such as company performance.

```
getwd()
```

```
## [1] "C:/Users/Gutwein/Google Drive/MIDS/W203 Statistics for Data Science/lab_01/w203/lab_01"
```

loading CEO object from workspace file...

```
load("ceo_w203.RData")
objects()
```

```
## [1] "CEO"
```

loading standard EDA libraries...

```
library(car)
library(dplyr)
```

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

```
## The following object is masked from 'package:car':
##
##     recode
```

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

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

```
library(Hmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     combine, src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

```
library(ggplot2)
str(CEO)
```

```
## 'data.frame':    185 obs. of  8 variables:
##  $ salary : num  1033 879 971 567 1336 ...
##  $ age    : num  62 63 72 56 60 59 46 59 51 56 ...
##  $ college: num  1 1 1 1 1 1 1 1 1 1 ...
##  $ grad   : num  1 1 1 0 1 1 1 1 0 1 ...
##  $ comten : num  30 21 33 31 21 2 7 3 8 9 ...
##  $ ceoten : num  1 9 24 10 13 2 3 3 8 3 ...
##  $ profits: num  478 212 69 65 562 401 44 257 13 34 ...
##  $ mktval : num  7300 4900 609 1700 4300 10700 533 3900 458 6700 ...
```

```
summary(CEO)
```

```
##      salary            age           college            grad
##  Min.   : 100.0   Min.   :21.00   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 467.0   1st Qu.:51.00   1st Qu.:1.0000   1st Qu.:0.0000
##  Median : 697.0   Median :57.00   Median :1.0000   Median :1.0000
##  Mean   : 852.9   Mean   :55.78   Mean   :0.9622   Mean   :0.5514
##  3rd Qu.:1101.0   3rd Qu.:61.00   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :5299.0   Max.   :86.00   Max.   :1.0000   Max.   :1.0000
##      comten          ceoten          profits           mktval
##  Min.   : 2.00   Min.   : 0.000   Min.   :-463.0   Min.   :   -1
##  1st Qu.: 9.00   1st Qu.: 3.000   1st Qu.:  33.0   1st Qu.:  567
##  Median :21.00   Median : 5.000   Median :  57.0   Median : 1200
##  Mean   :21.66   Mean   : 7.681   Mean   : 199.2   Mean   : 3450
##  3rd Qu.:33.00   3rd Qu.:11.000   3rd Qu.: 195.0   3rd Qu.: 3200
##  Max.   :58.00   Max.   :37.000   Max.   :2700.0   Max.   :45400
```

```
head(CEO)
```

```
##     salary age college grad comten ceoten profits mktval
## 154   1033  62       1    1     30      1     478   7300
## 79     879  63       1    1     21      9     212   4900
## 19     971  72       1    1     33     24      69    609
## 115    567  56       1    0     31     10      65   1700
## 36    1336  60       1    1     21     13     562   4300
## 153   1444  59       1    1      2      2     401  10700
```

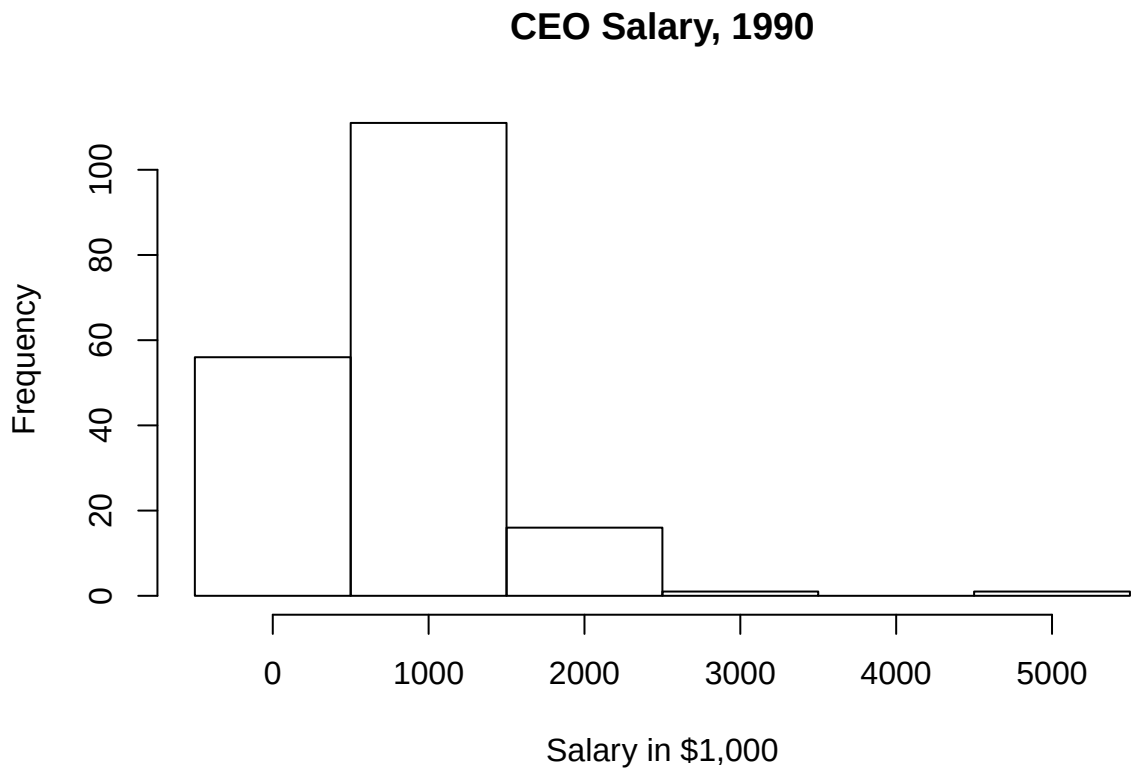
Univariate Analysis

We will take a look at each variable

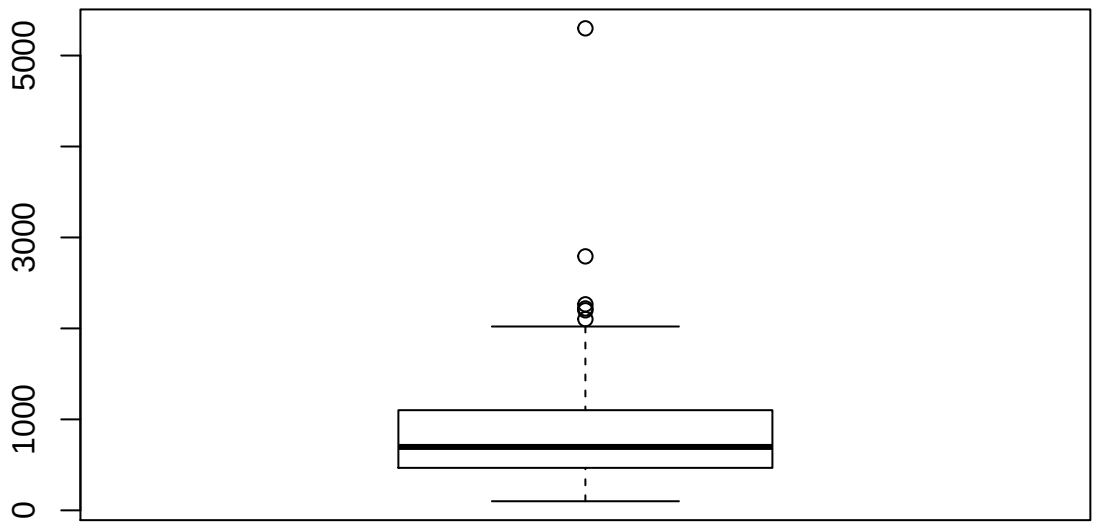
```
describe(CEO$salary)
```

```
## CEO$salary
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      185        0      171        1    852.9    559.5      276      358
##      .25      .50      .75      .90      .95
##      467      697     1101     1495     1750
##
## lowest :  100  129  173  174  185, highest: 2199 2220 2265 2792 5299
```

```
hist(CEO$salary, breaks = c(-500, 500, 1500, 2500, 3500, 4500,
    5500), main = "CEO Salary, 1990", xlab = "Salary in $1,000")
```




```
boxplot(CEO$salary)
```



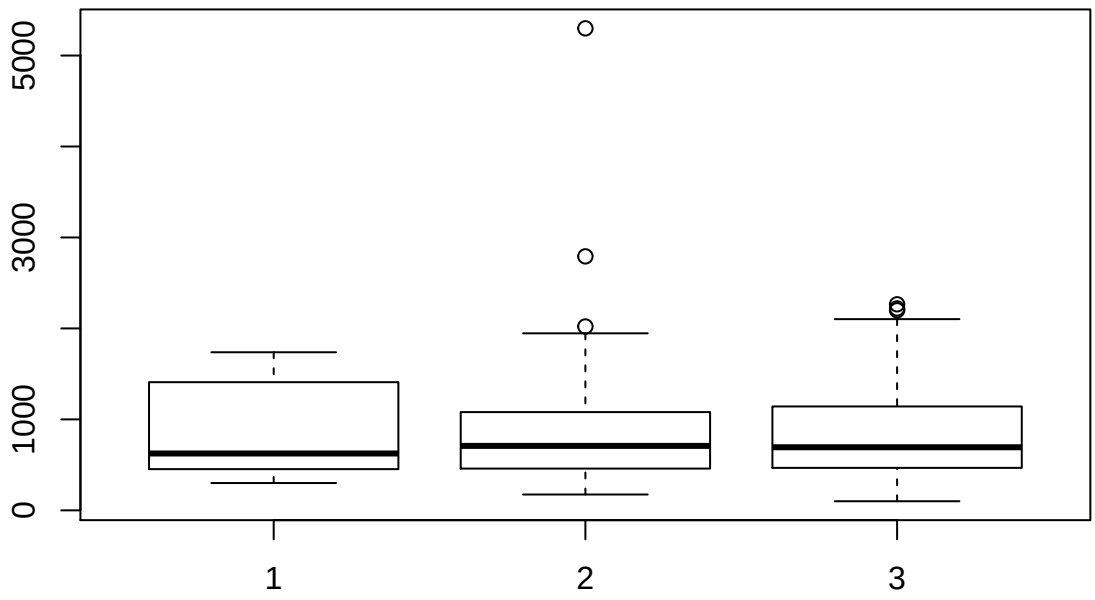
We want to see how salary compares for CEO's across different education levels. We will sort this into three groups:

- Not college educated
- College educated but no graduate school
- College educated and graduate school

```
no_college <- CEO$college == 0
college <- CEO$college == 1 & CEO$grad == 0
graduate <- CEO$grad == 1
```

```
CEO_nc <- CEO[no_college, ]
CEO_c <- CEO[college, ]
CEO_m <- CEO[graduate, ]
```

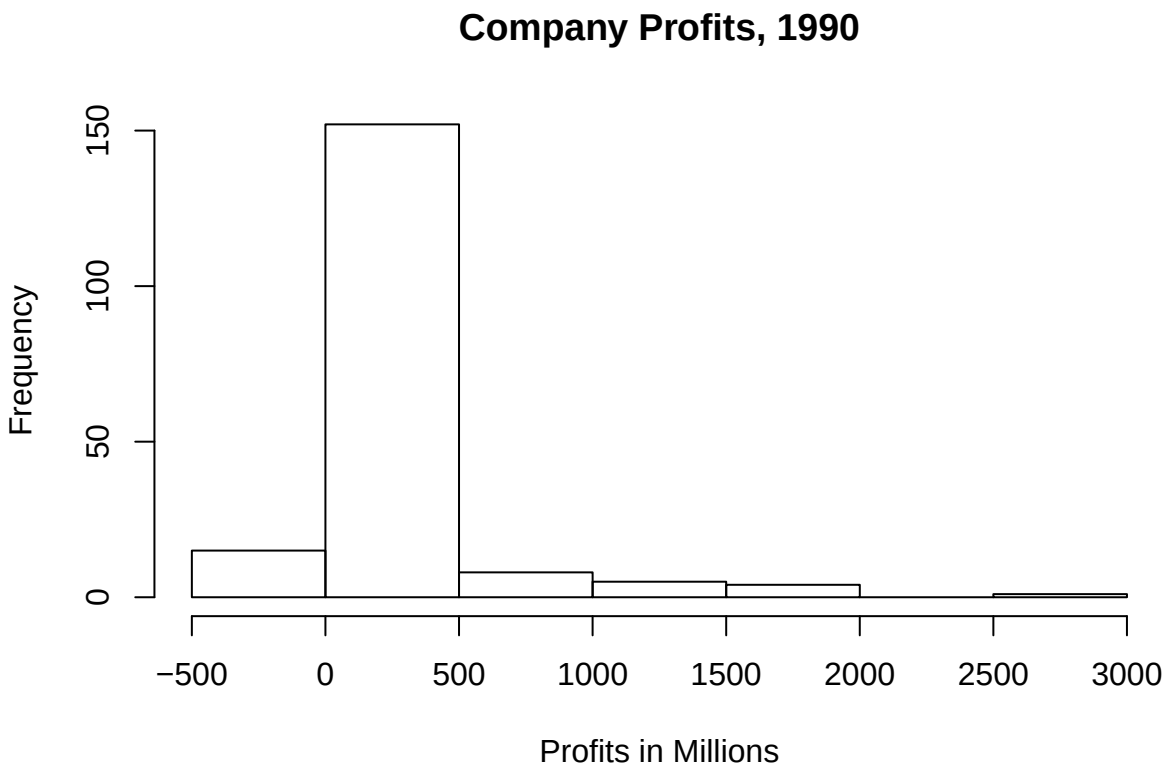
```
boxplot(CEO_nc$salary, CEO_c$salary, CEO_m$salary)
```



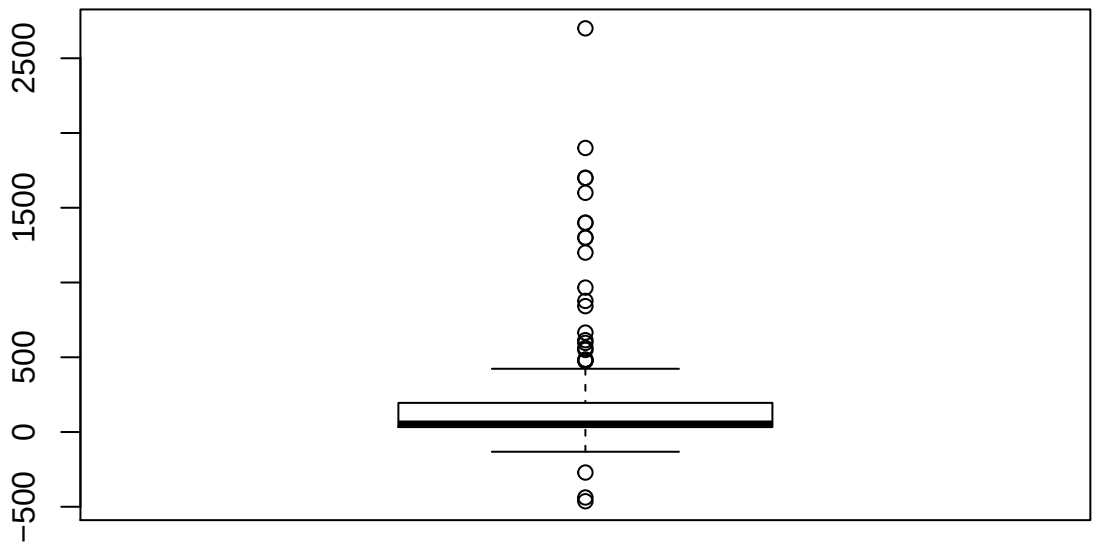

```
describe(CEO$profits)
```

```
## CEO$profits
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      185        0      134        1    199.2      310     -2.6      7.4
##      .25      .50      .75      .90      .95
##     33.0     57.0    195.0    483.6   1153.2
##
## lowest : -463 -438 -271 -132  -80, highest: 1400 1600 1700 1900 2700
```

```
hist(CEO$profits, main = "Company Profits, 1990", xlab = "Profits in Millions")
```



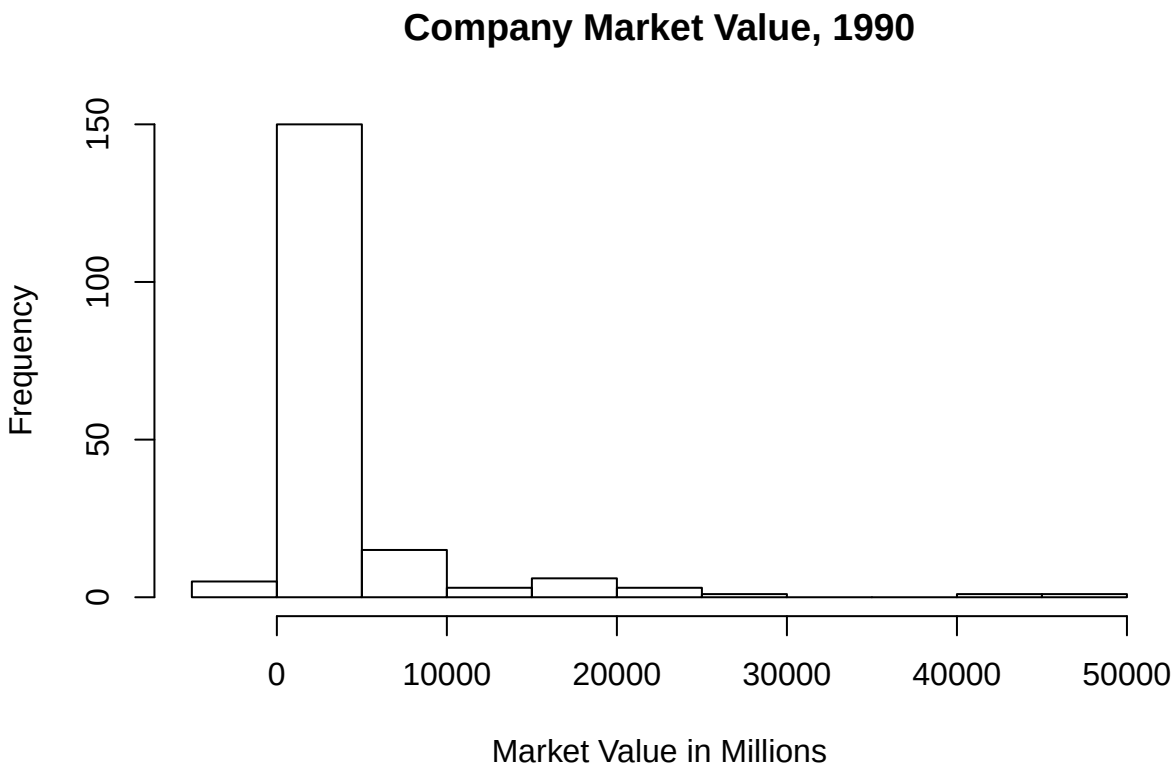

```
boxplot(CEO$profits)
```



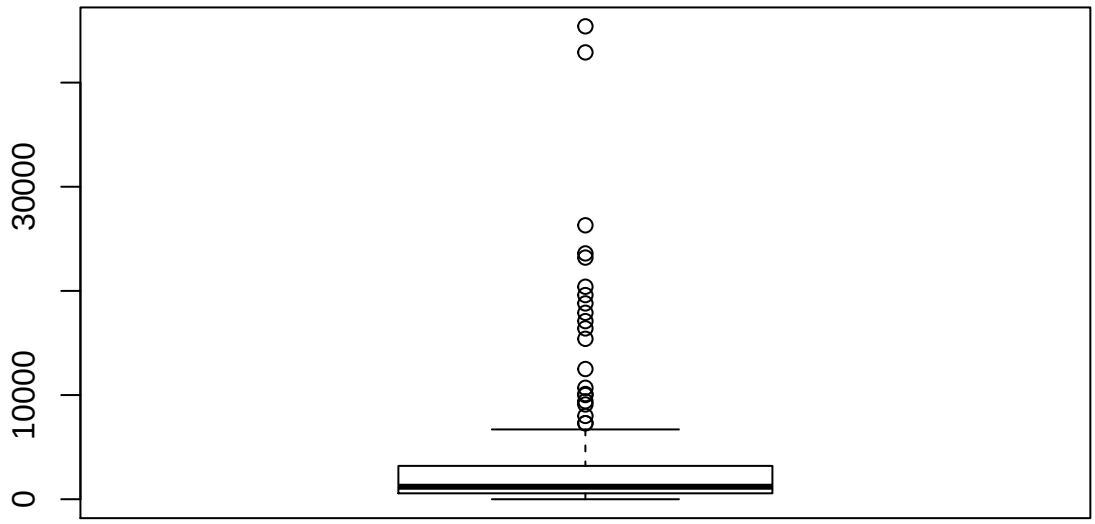

```
describe(CEO$mktval)
```

```
## CEO$mktval
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      185        0      131        1     3450     4605    391.6    434.6
##      .25      .50      .75      .90      .95
##    567.0   1200.0   3200.0   7720.0  16960.0
##
## lowest :    -1   303   344   387   390, highest: 23200 23600 26300 42900 45400
```

```
hist(CEO$mktval, main = "Company Market Value, 1990", xlab = "Market Value in Millions")
```




```
boxplot(CEO$mktval)
```

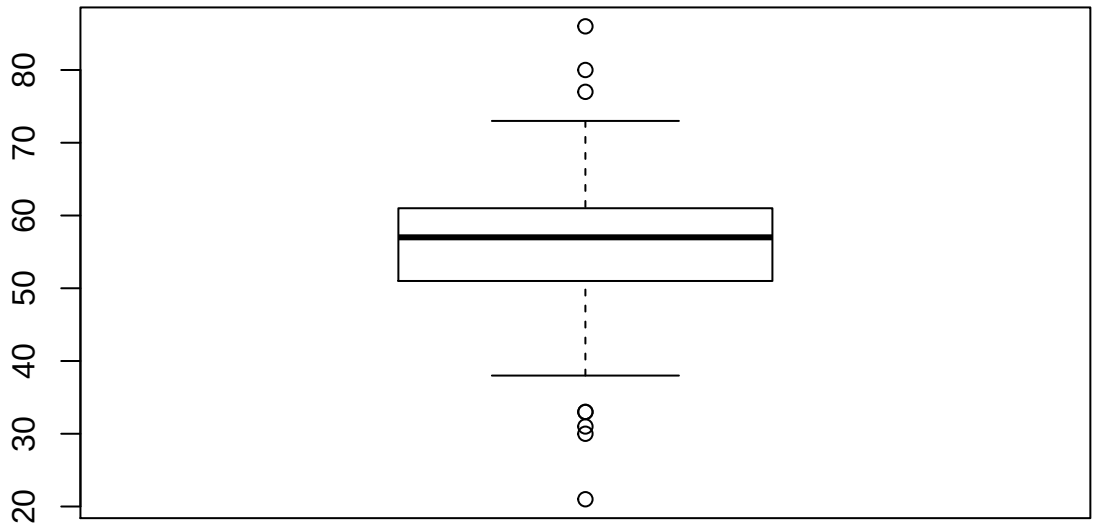



```
describe(CEO$age)
```

```
## CEO$age
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      185        0       42    0.998    55.78    10.09       40       44
##      .25      .50      .75      .90      .95
##       51       57       61       66       69
##
## lowest : 21 30 31 33 38, highest: 72 73 77 80 86
```

```
hist(CEO$age, main = "Age of CEOs in 1990")
```

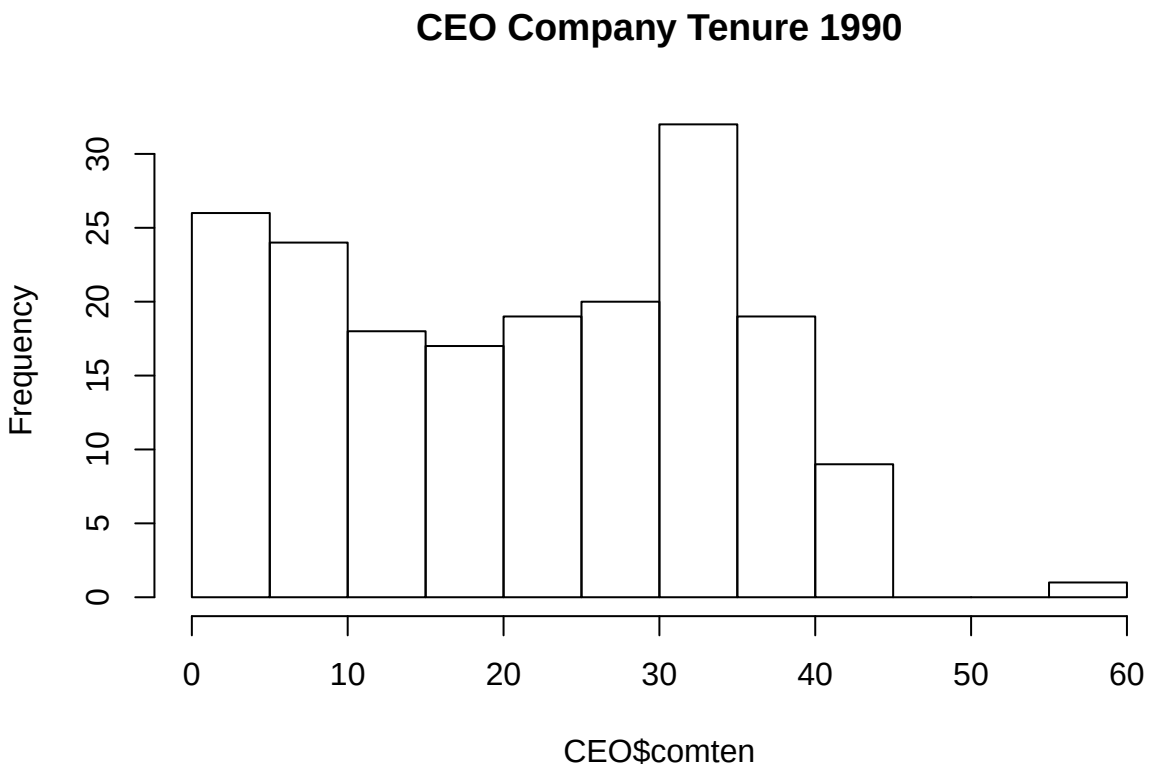

```
boxplot(CEO$age)
```



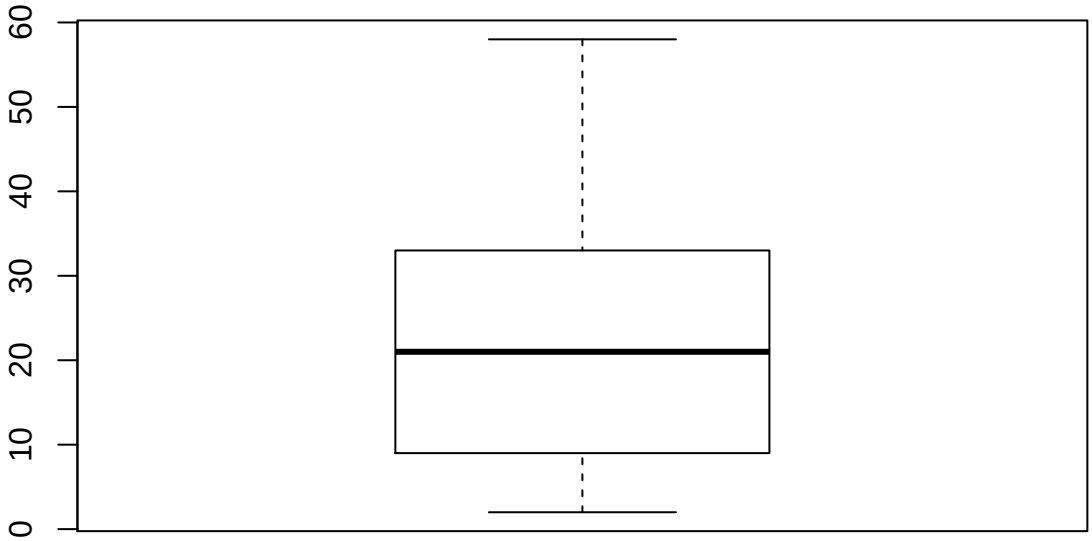

```
describe(CEO$comten)
```

```
## CEO$comten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      185        0       45    0.999    21.66    14.54      3.0      4.0
##      .25      .50      .75      .90      .95
##      9.0     21.0     33.0     36.6     40.8
##
## lowest :  2  3  4  5  6, highest: 42 43 44 45 58
```

```
hist(CEO$comten, main = "CEO Company Tenure 1990")
```



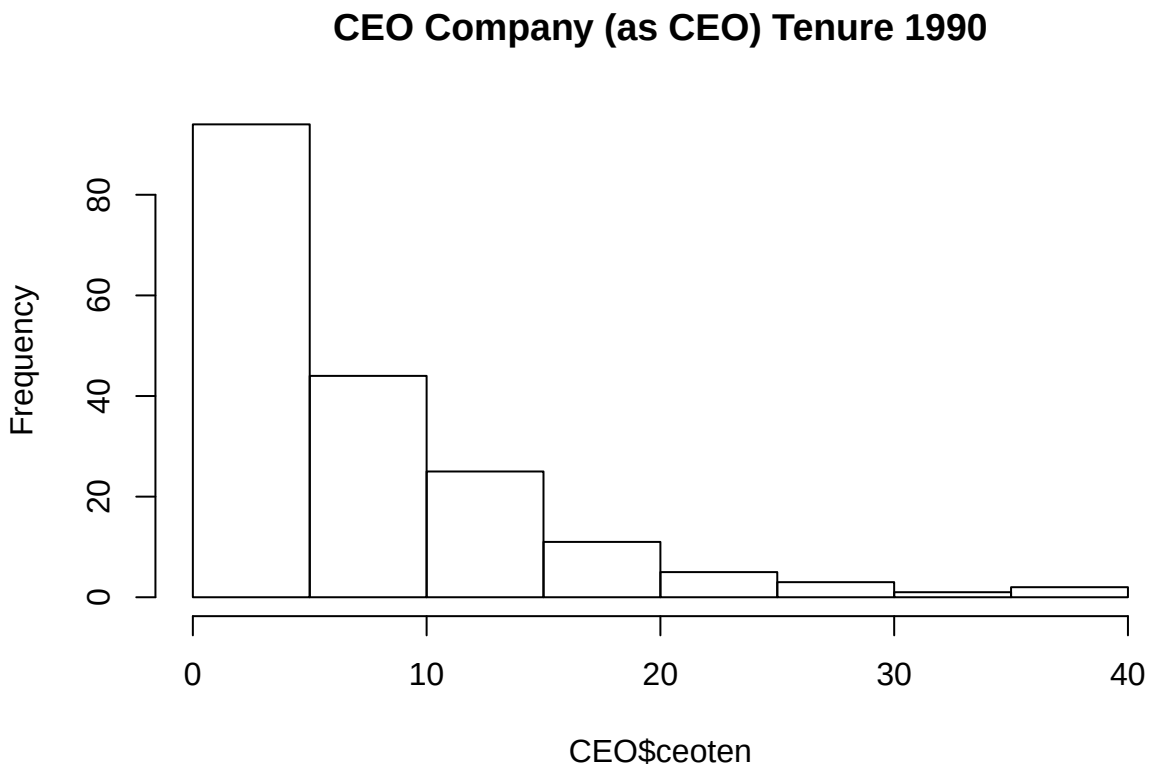

```
boxplot(CEO$comten)
```



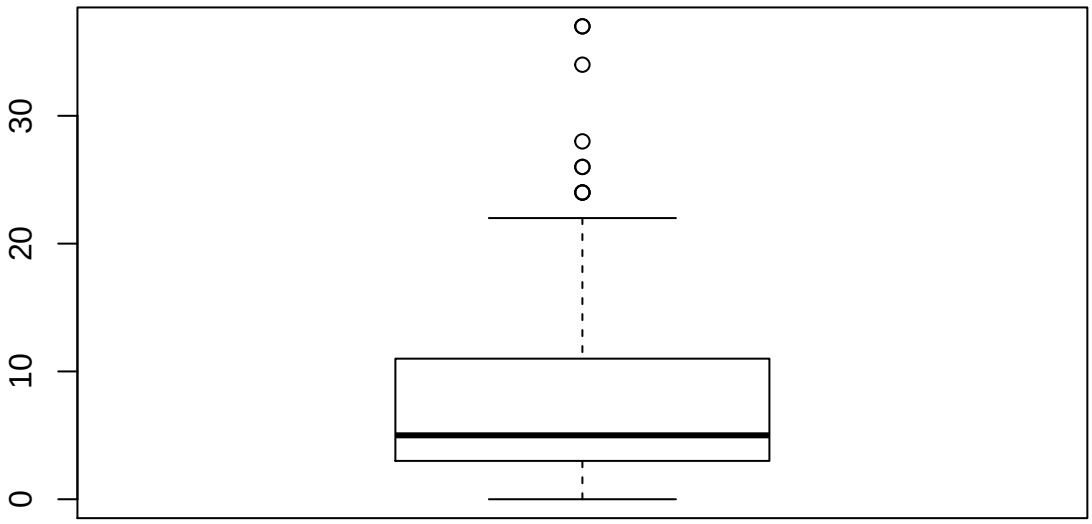

```
describe(CEO$ceoten)
```

```
## CEO$ceoten
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      185        0       28    0.993    7.681    7.285      1.0      1.0
##      .25      .50      .75      .90      .95
##      3.0      5.0     11.0     17.0     21.8
##
## lowest :  0  1  2  3  4, highest: 24 26 28 34 37
```

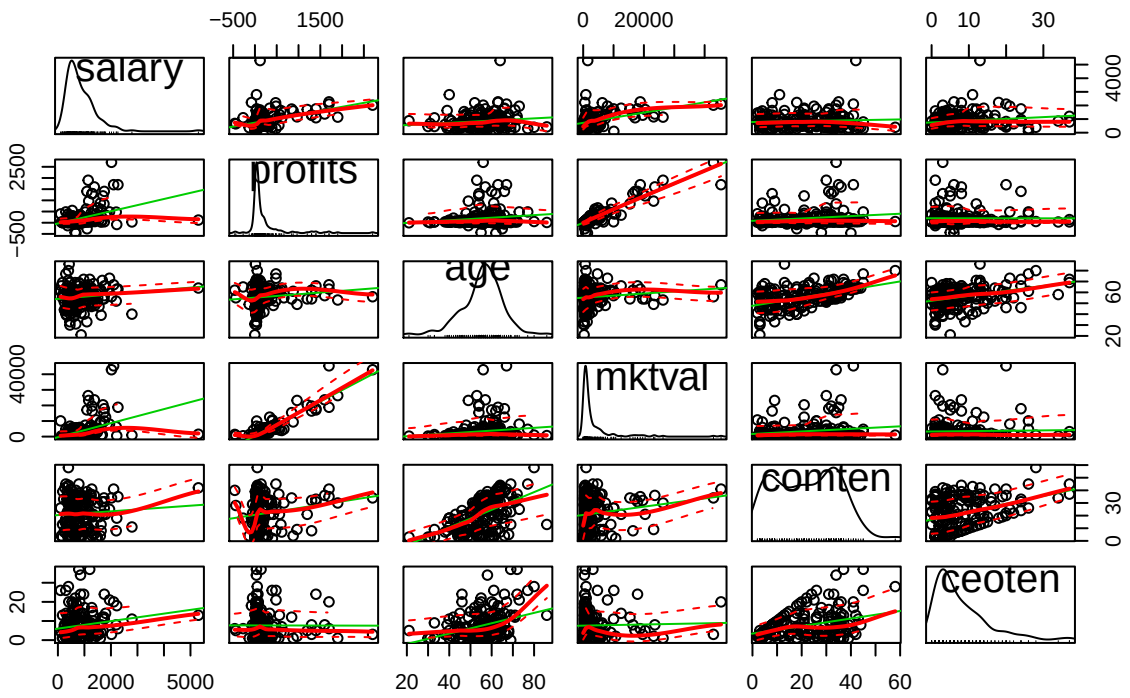
```
hist(CEO$ceoten, main = "CEO Company (as CEO) Tenure 1990")
```




```
boxplot(CEO$ceoten)
```




```
scatterplotMatrix(~salary + profits + age + mktval + comten +
    ceoten, data = CEO, main = "Scatterplot Matrix for Key CEO Variables")
```



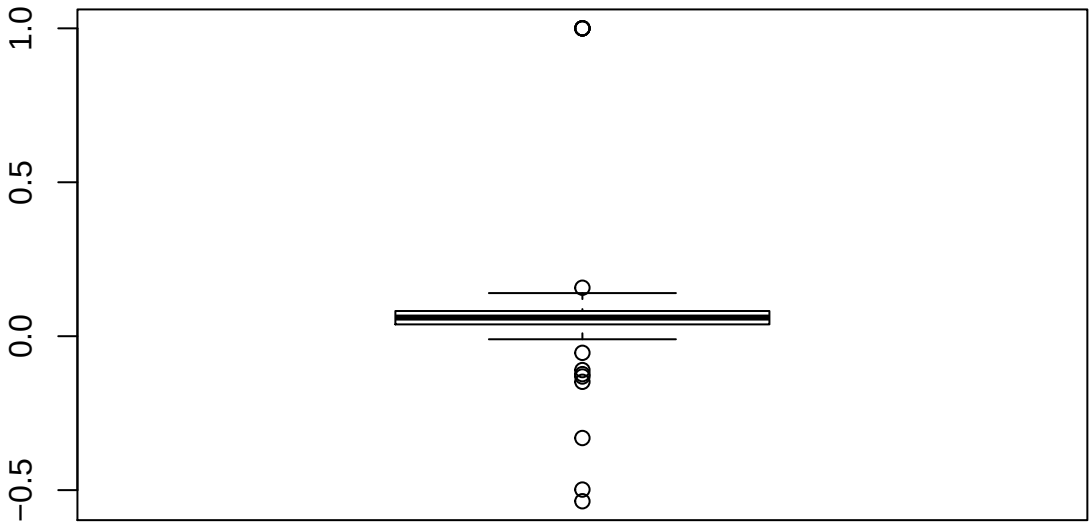
Secondary Variables

Because the size of each company varies significantly, we have decided to "normalize" the performance measure of each CEO by dividing the profits by the market value. Typically a company would measure its percentage of profits with the gross revenue as the denominator, however, since we do not have revenue data available we will use the market value of the company.

```
CEO$prof_per <- CEO$profits/CEO$mktval
describe(CEO$prof_per)
```

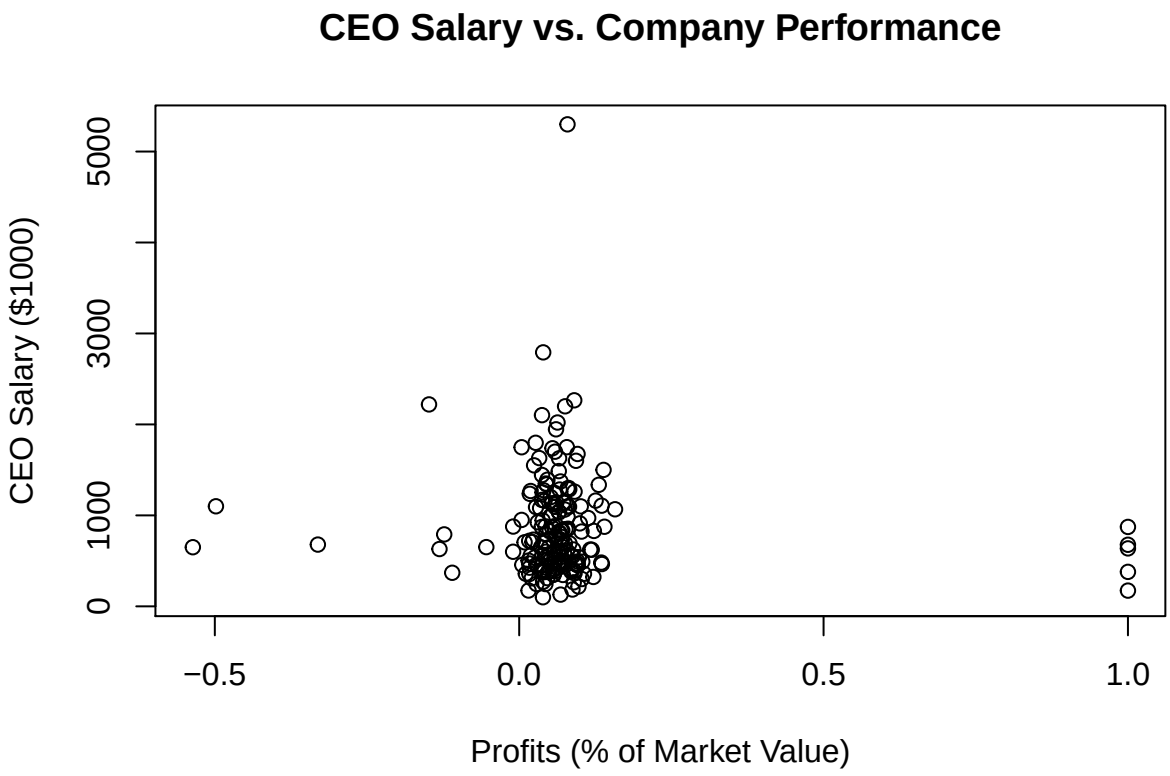
```
## CEO$prof_per
##          n    missing   distinct       Info       Mean        Gmd        .05
##        185          0        176          1    0.07473     0.1059 -0.007121
##        .10        .25        .50        .75        .90        .95
##  0.016591   0.038378   0.060500   0.081871   0.105289   0.135350
##
## -0.54 (1, 0.005), -0.5 (1, 0.005), -0.34 (1, 0.005), -0.14 (2, 0.011),
## -0.12 (2, 0.011), -0.0599999999999997 (1, 0.005), -0.0199999999999997 (1,
## 0.005), 3.33066907387547e-16 (5, 0.027), 0.0200000000000004 (19, 0.103),
## 0.0400000000000004 (36, 0.195), 0.0600000000000004 (49, 0.265),
## 0.0800000000000004 (30, 0.162), 0.1 (19, 0.103), 0.12 (6, 0.032), 0.14 (6,
## 0.032), 0.16 (1, 0.005), 1 (5, 0.027)
```

```
boxplot(CEO$prof_per)
```



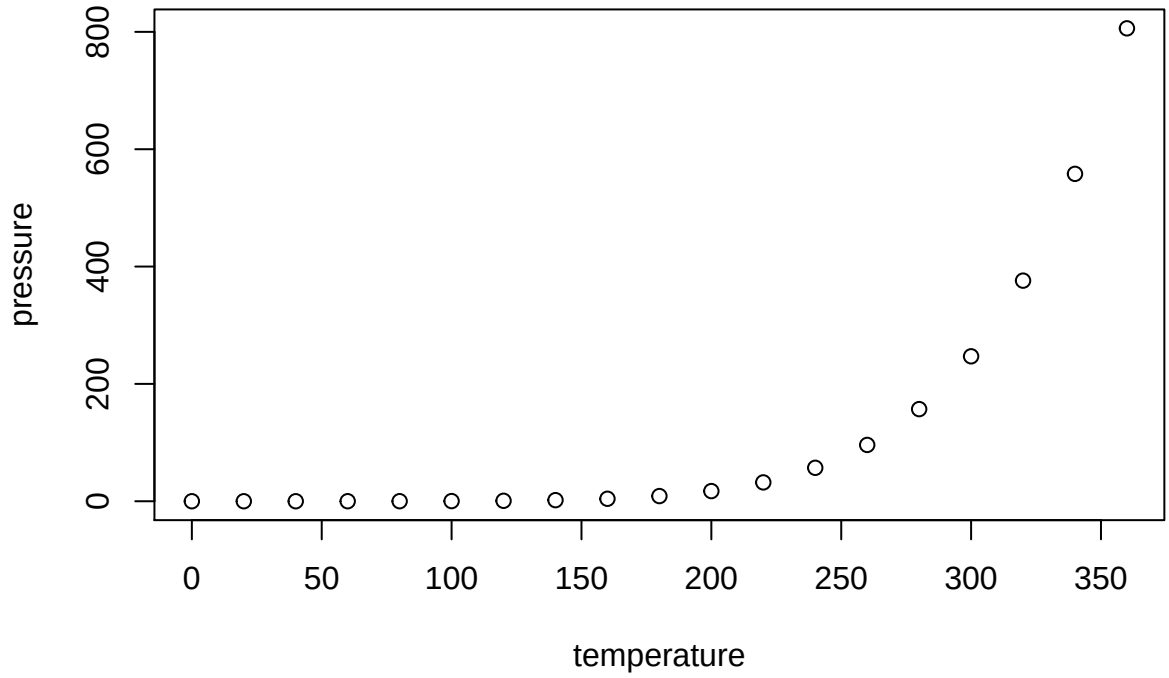
Now we will take a look at the relationship of CEO salary vs profits as a percentage of market value.

```
plot(CEO$prof_per, CEO$salary, xlab = "Profits (% of Market Value)",
    ylab = "CEO Salary ($1000)", main = "CEO Salary vs. Company Performance")
```



Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.