

# 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?

This section is reserved for Intro...

```
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)
```

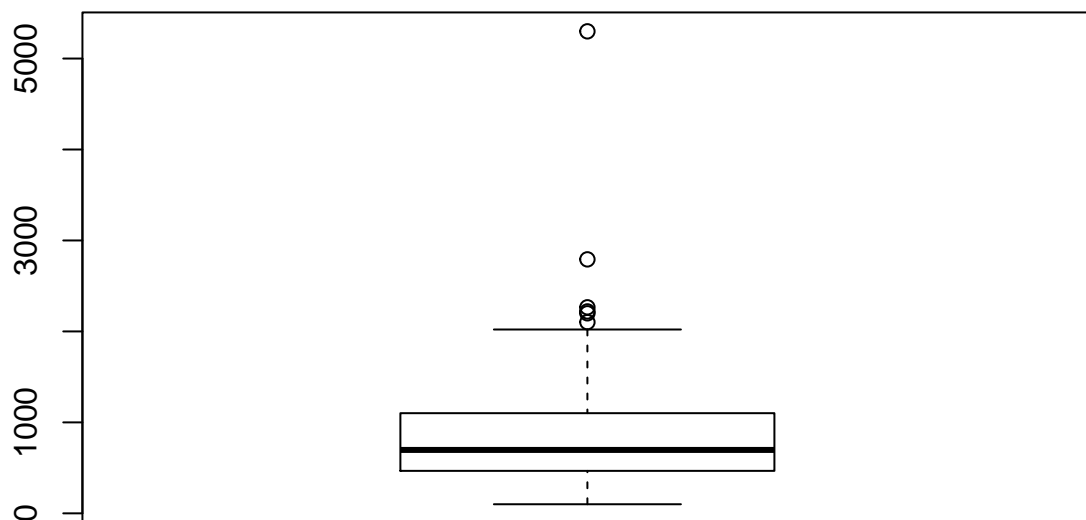
```
## CEO$salary
##      n missing distinct    Info    Mean    Gmd    .05    .10
##    185      0        171      1  852.9  559.5  276   358
```

```
##      .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")
```

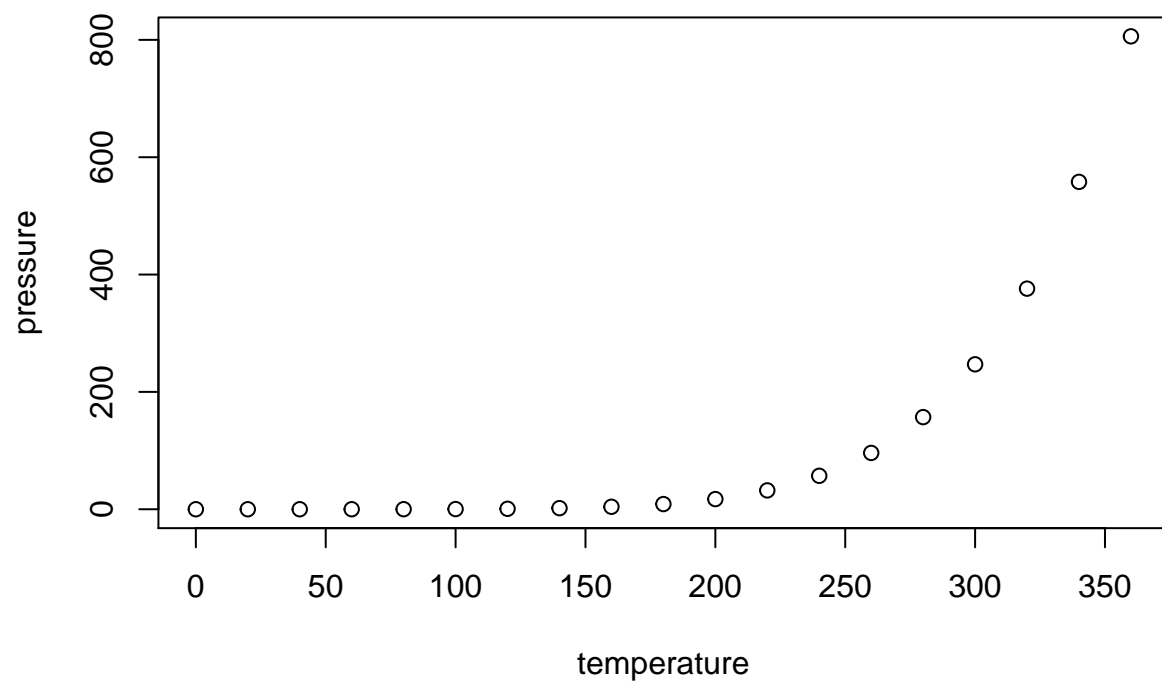


```
boxplot(CEO$salary)
```



## 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.