

Final Project

MuthukumarSrinivasan

Friday, May 27, 2016

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

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

```
require(ggplot2)

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.2.3

library(ggplot2)

finalProject<-read.csv('c:/temp/SFSalries-Cleaned.csv',header=TRUE)
finalProject100Rows<-read.csv('c:/temp/SFSalaries-sliced100Rows.csv',header=TRUE)
summary(finalProject)

##          Id            EmployeeName
##  Min.    : 1   Kevin Lee    : 13
##  1st Qu.: 37162 Richard Lee : 11
##  Median : 74324 Steven Lee  : 11
##  Mean   : 74324 William Wong: 11
##  3rd Qu.:111485 John Chan   : 9
##  Max.   :148646 KEVIN LEE  : 9
##                (Other)     :148582
##          JobTitle        BasePay
##  Transit Operator      : 7036  Min.   : -166
##  Special Nurse         : 4389  1st Qu.: 33590
##  Registered Nurse      : 3736  Median : 65009
##  Public Svc Aide-Public Works: 2518  Mean   : 66327
##  Police Officer 3       : 2421  3rd Qu.: 94691
##  Custodian             : 2416  Max.   :319275
##  (Other)                :126130 NA's    :605
##          OvertimePay      OtherPay        Benefits
##  Min.   : -0.01  Min.   : -7058.6  Min.   : -33.89
##  1st Qu.:  0.00  1st Qu.:     0.0  1st Qu.:11540.26
##  Median :  0.00  Median :  811.4  Median :28628.70
##  Mean   : 5066.20  Mean   : 3648.9  Mean   :25008.78
##  3rd Qu.: 4658.45  3rd Qu.: 4236.6  3rd Qu.:35566.93
##  Max.   :245131.88  Max.   :400184.2  Max.   :96570.66
##                NA's   :36159
##          TotalPay        TotalPayBenefits      Year        Notes
```

```

## Min. : 0 Min. :-33.9 Min. :2011 Mode:logical
## 1st Qu.: 36173 1st Qu.: 44081.4 1st Qu.:2012 NA's:148646
## Median : 71428 Median : 92406.3 Median :2013
## Mean : 74772 Mean : 93697.6 Mean :2013
## 3rd Qu.:105846 3rd Qu.:132879.8 3rd Qu.:2014
## Max. :567595 Max. :567595.4 Max. :2014
##
##          Agency      Status
## San Francisco:148646    :110531
##                           FT: 22334
##                           PT: 15781
##
```

```
str(finalProject)
```

```

## 'data.frame': 148646 obs. of 13 variables:
## $ Id           : int 1 2 3 4 5 6 7 8 9 10 ...
## $ EmployeeName : Factor w/ 110809 levels "A Bernard Fatooh",...: 77634 34712 1560 17232 81100 23...
## $ JobTitle     : Factor w/ 2158 levels "Account Clerk",...: 836 298 298 2148 594 135 246 609 246 ...
## $ BasePay      : num 167411 155966 212739 77916 134402 ...
## $ OvertimePay  : num 0 245132 106088 56121 9737 ...
## $ OtherPay     : num 400184 137811 16453 198307 182235 ...
## $ Benefits     : num NA NA NA NA NA NA NA NA NA ...
## $ TotalPay     : num 567595 538909 335280 332344 326373 ...
## $ TotalPayBenefits: num 567595 538909 335280 332344 326373 ...
## $ Year         : int 2011 2011 2011 2011 2011 2011 2011 2011 ...
## $ Notes        : logi NA NA NA NA NA NA ...
## $ Agency        : Factor w/ 1 level "San Francisco": 1 1 1 1 1 1 1 1 1 ...
## $ Status        : Factor w/ 3 levels "", "FT", "PT": 1 1 1 1 1 1 1 1 1 ...

```

```
str(finalProject100Rows)
```

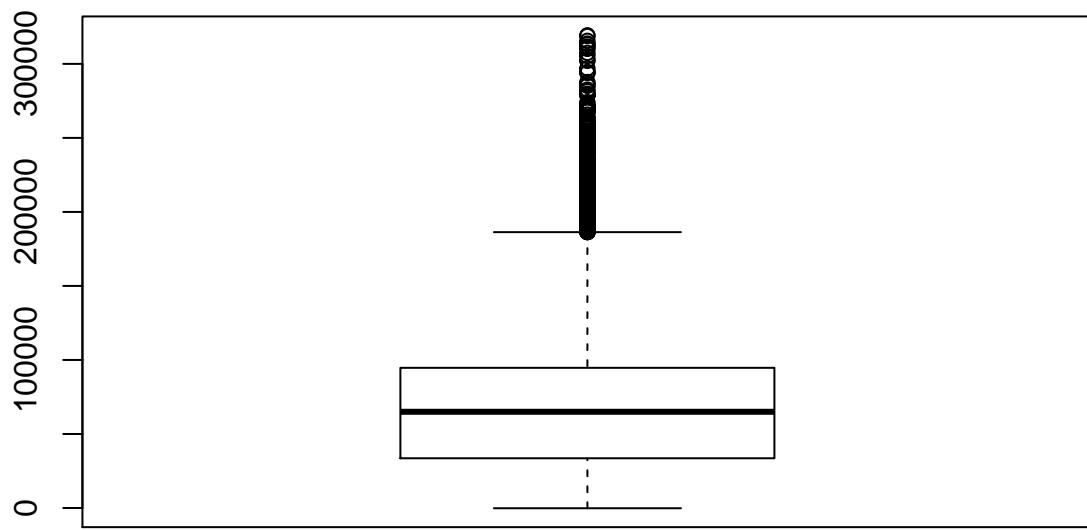
```

## 'data.frame': 99 obs. of 12 variables:
## $ Id           : int 1 2 3 4 5 6 7 8 9 10 ...
## $ EmployeeName : Factor w/ 98 levels "AI-KYUNG CHUNG",...: 75 31 2 14 77 19 3 17 69 42 ...
## $ JobTitle     : Factor w/ 37 levels "ADMINISTRATOR, SFGH MEDICAL CENTER",...: 23 7 7 37 15 4 6 10...
## $ BasePay      : num 167411 155966 212739 77916 134402 ...
## $ OvertimePay  : num 0 245132 106088 56121 9737 ...
## $ OtherPay     : num 400184 137811 16453 198307 182235 ...
## $ Benefits     : logi NA NA NA NA NA NA ...
## $ TotalPay     : num 567595 538909 335280 332344 326373 ...
## $ TotalPayBenefits: num 567595 538909 335280 332344 326373 ...
## $ Year         : int 2011 2011 2011 2011 2011 2011 2011 2011 ...
## $ Notes        : logi NA NA NA NA NA NA ...
## $ Agency        : Factor w/ 1 level "San Francisco": 1 1 1 1 1 1 1 1 1 ...

```

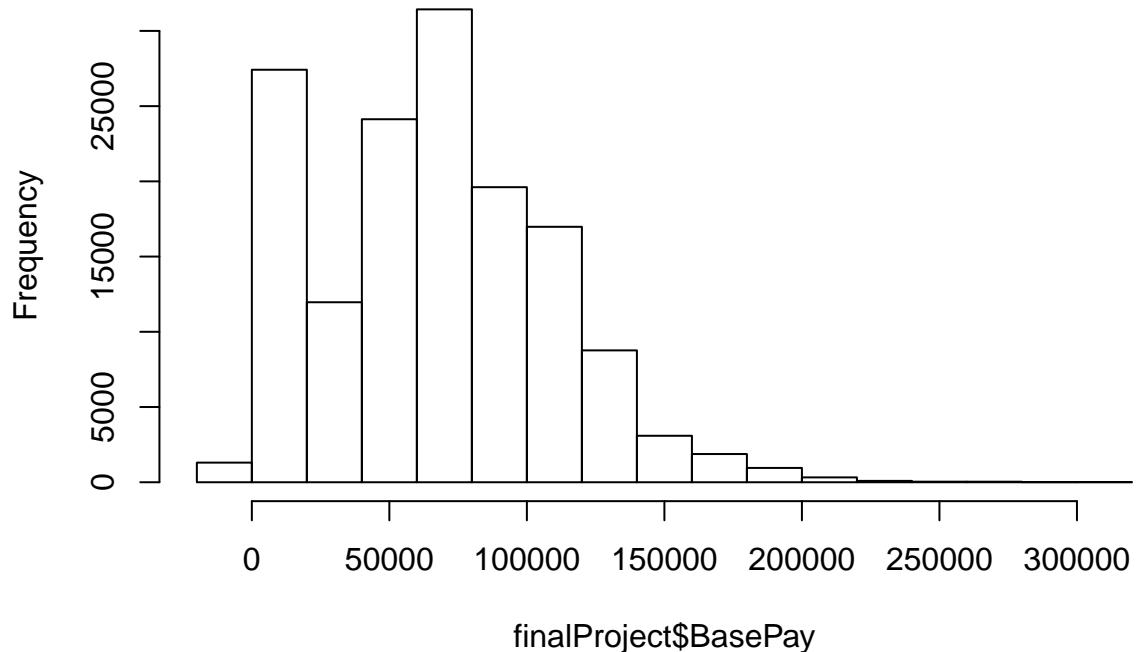
```
#### Chunck 2 in R Studie is about ALL Numeric colum boxplots. Box Plots for Base Pay, OvertimePay, OtherPay
```

```
boxplot(finalProject$BasePay)
```

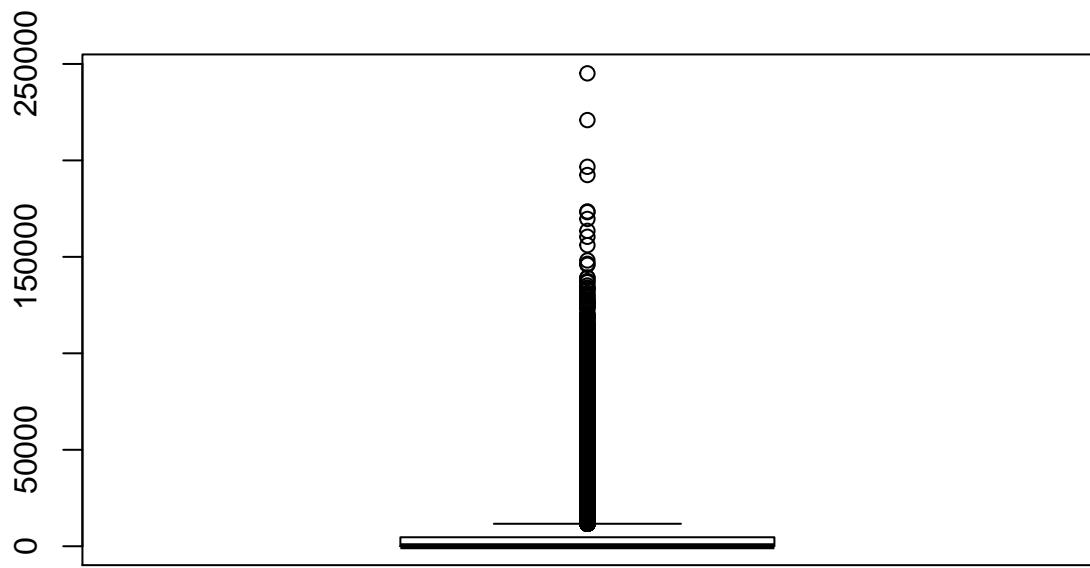


```
hist(finalProject$BasePay)
```

Histogram of finalProject\$BasePay

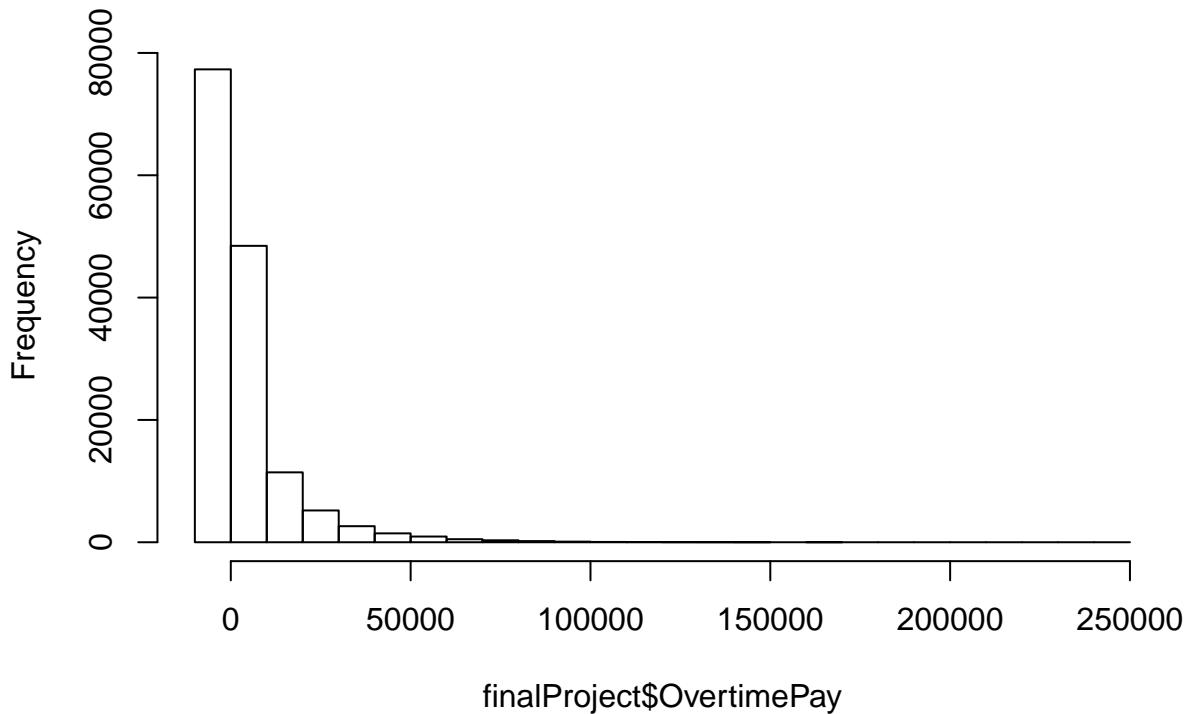


```
boxplot(finalProject$OvertimePay)
```

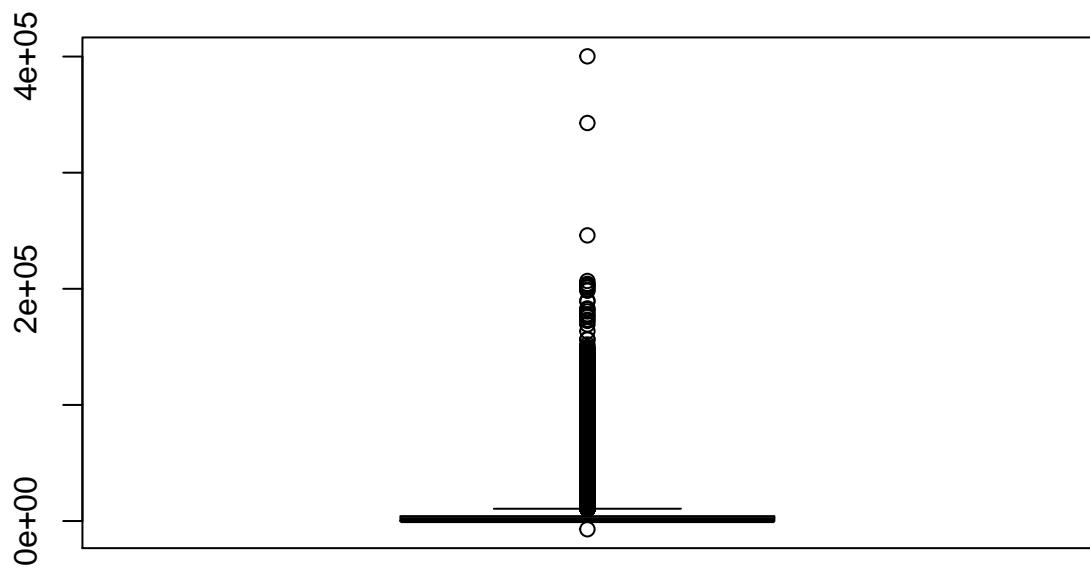


```
hist(finalProject$OvertimePay)
```

Histogram of finalProject\$OvertimePay

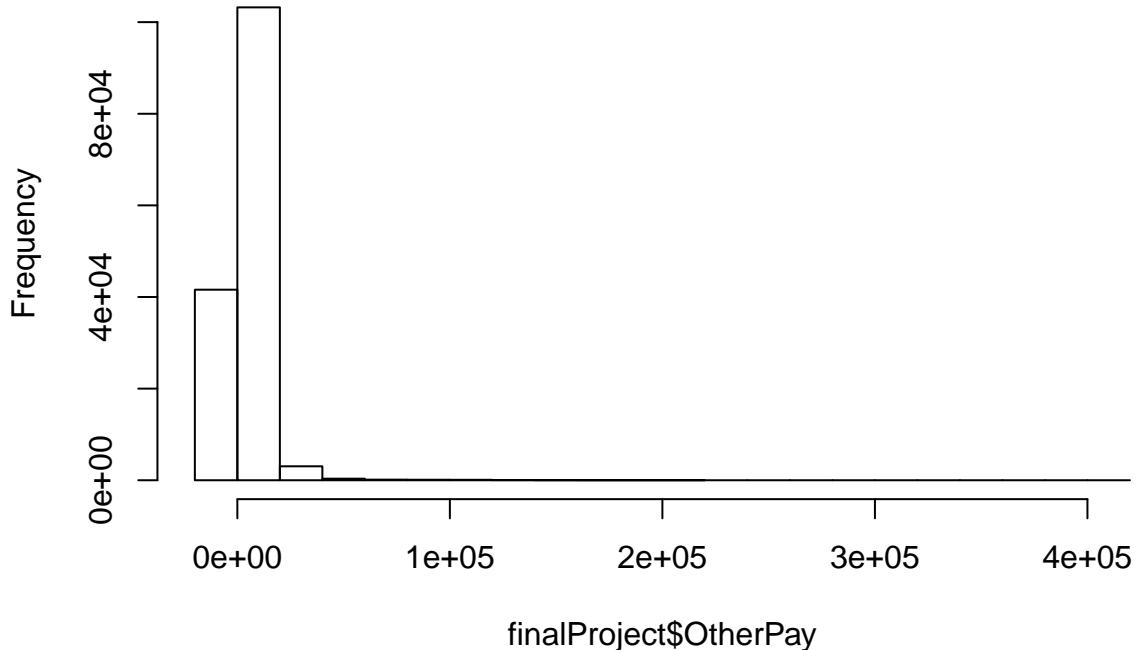


```
boxplot(finalProject$OtherPay)
```

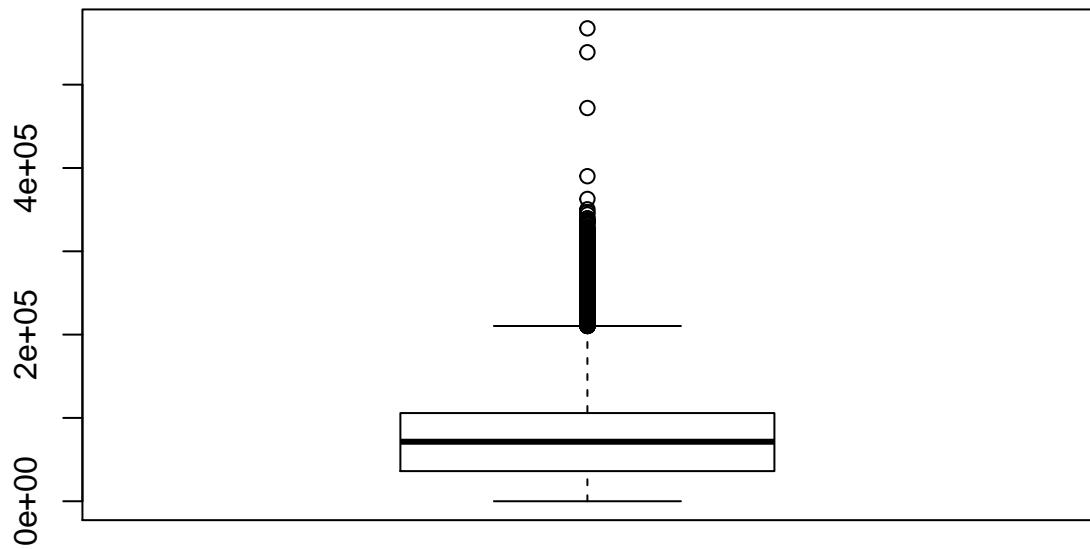


```
hist(finalProject$OtherPay)
```

Histogram of finalProject\$OtherPay

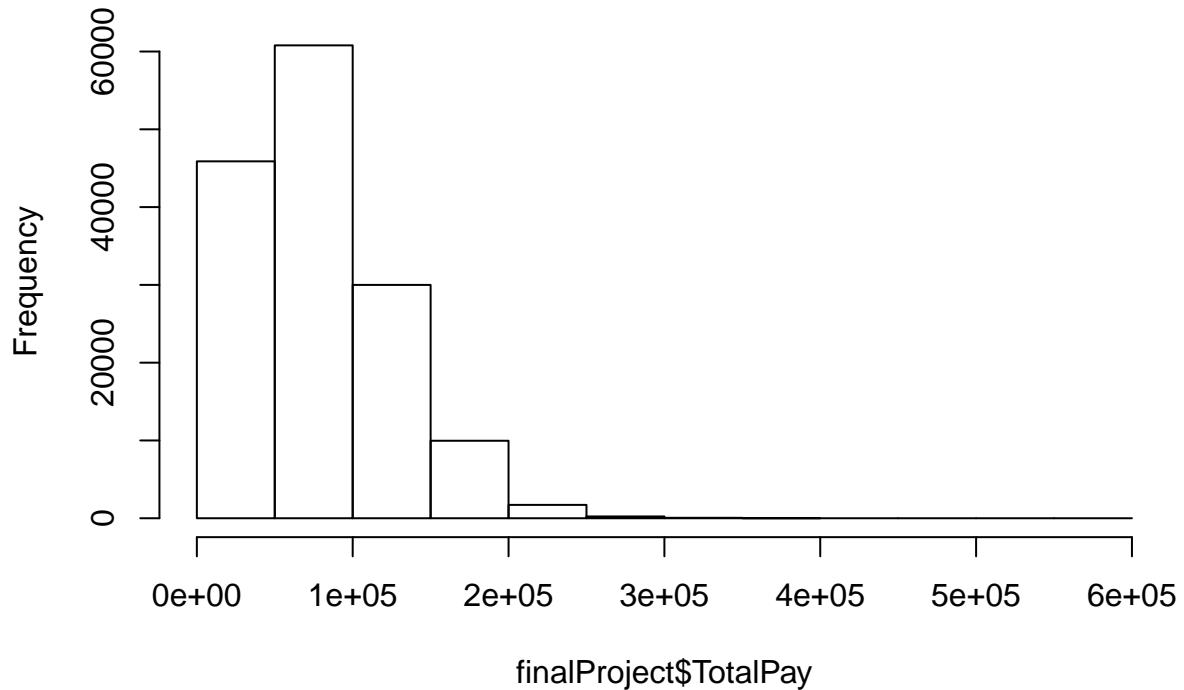


```
boxplot(finalProject$TotalPay)
```

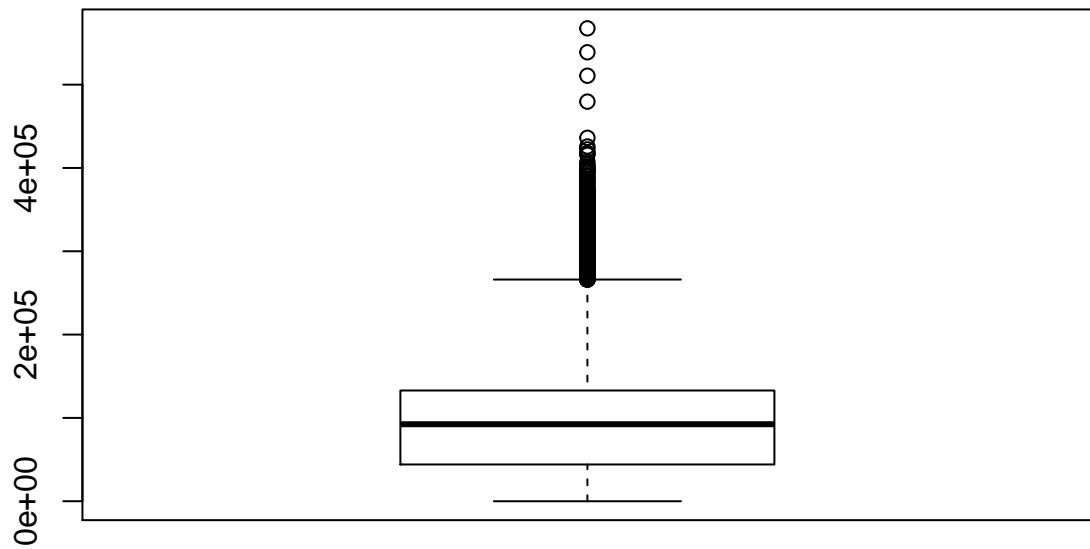


```
hist(finalProject$TotalPay)
```

Histogram of finalProject\$TotalPay

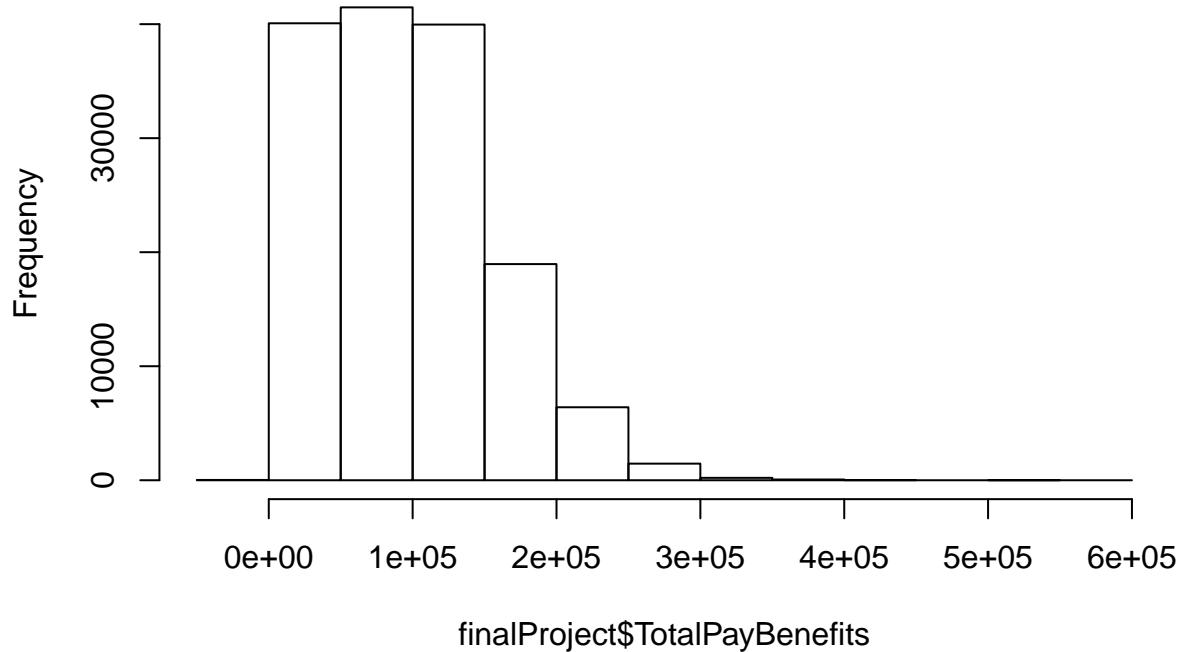


```
boxplot(finalProject$TotalPayBenefits)
```



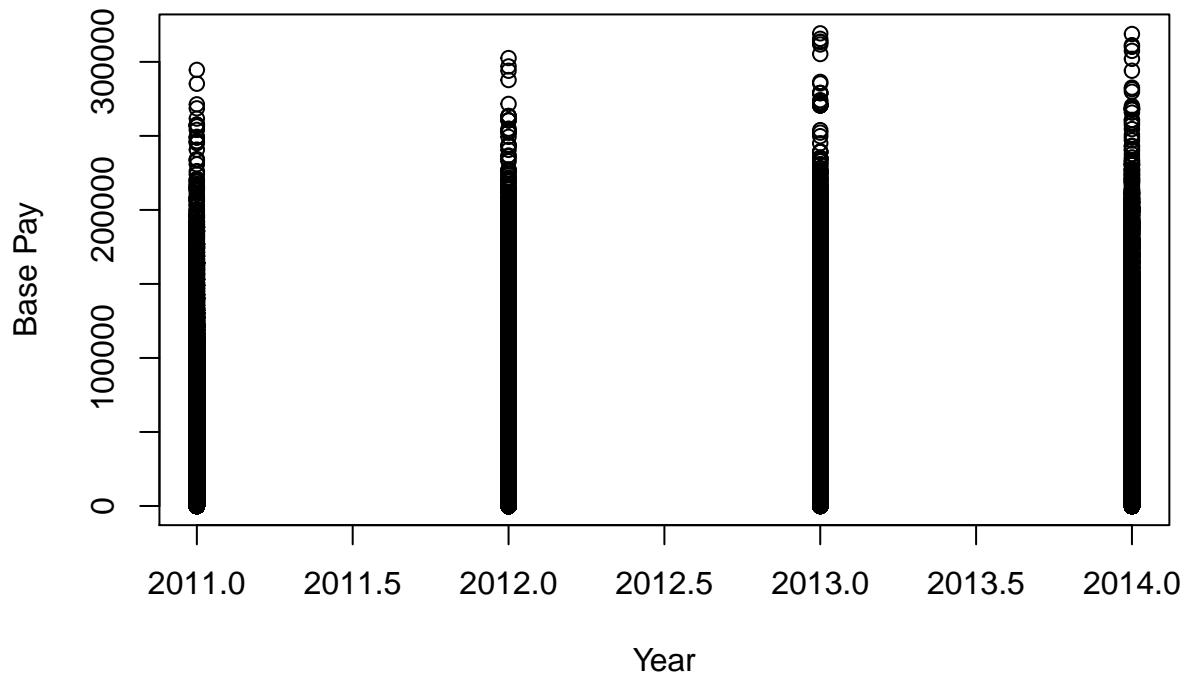
```
hist(finalProject$TotalPayBenefits)
```

Histogram of finalProject\$TotalPayBenefits



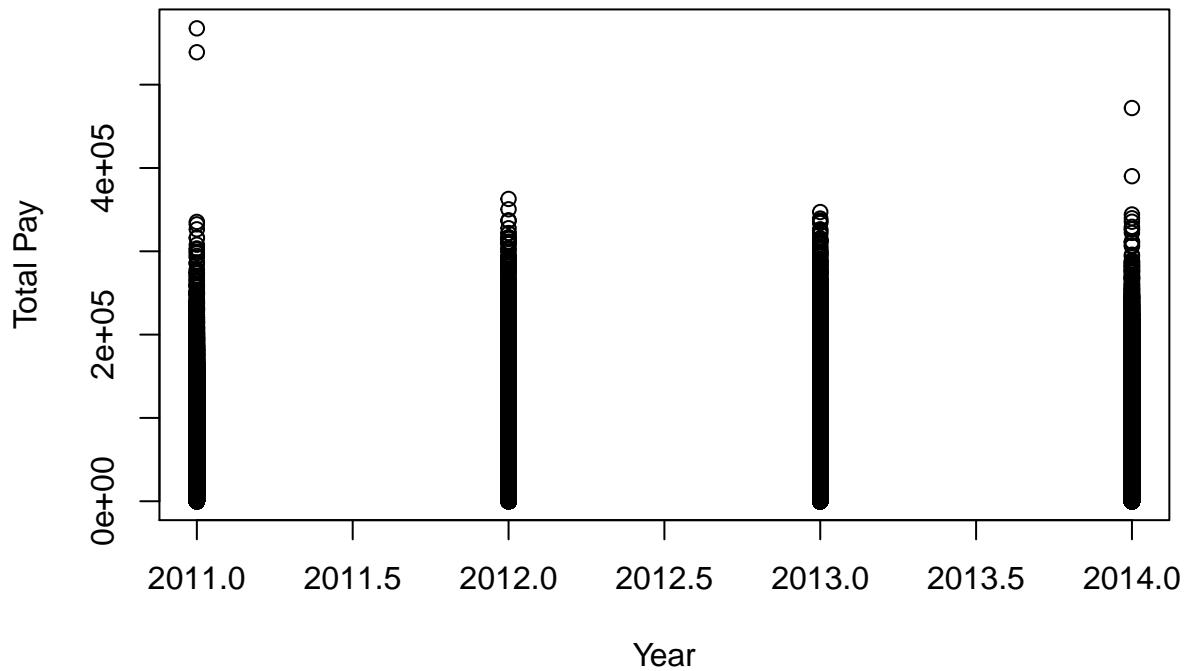
```
x1<-finalProject$Year  
y1<-finalProject$BasePay  
plot(x1,y1, main="Scatter Plot of SF Salries of DATA", xlab="Year", ylab="Base Pay",)
```

Scatter Plot of SF Salries of DATA



```
x1<-finalProject$Year  
y1<-finalProject$TotalPay  
plot(x1,y1, main="Scatter Plot of SF Salries of DATA", xlab="Year", ylab="Total Pay",)
```

Scatter Plot of SF Salries of DATA



```
x1<-finalProject$Year  
y1<-finalProject$TotalPayBenefits  
plot(x1,y1, main="Scatter Plot of SF Salries of DATA", xlab="Year", ylab="Total Pay + Benefits",)
```

Scatter Plot of SF Salries of DATA

