HW1

Hanao Li

September 10, 2019

Homework 1

i)

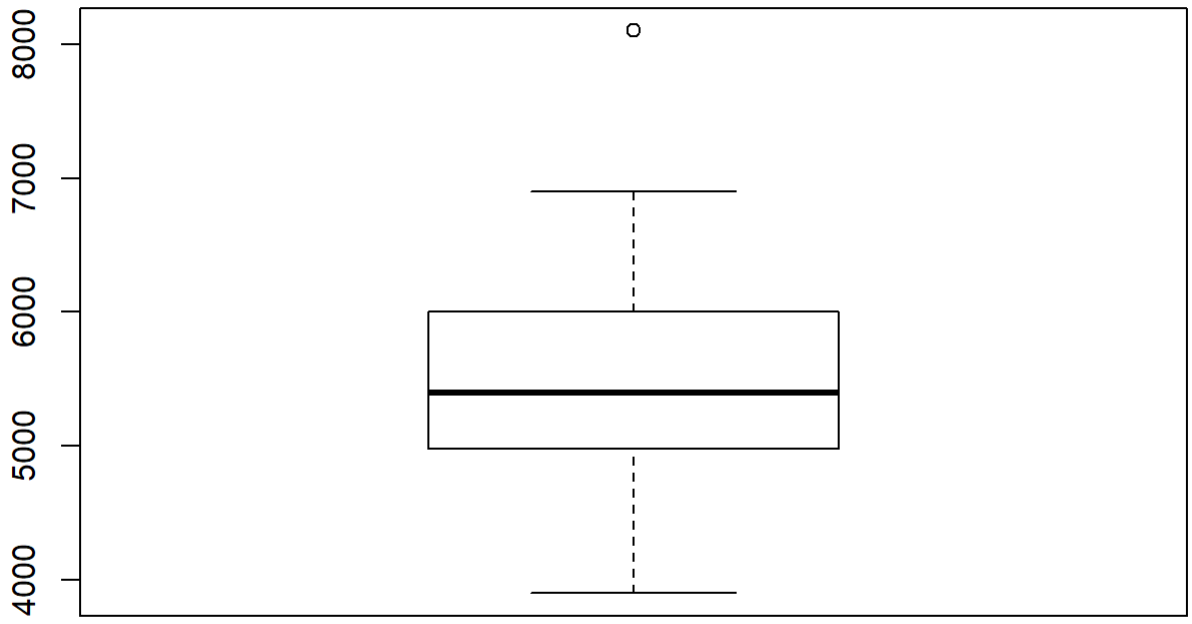
```
if(!require("pacman")) install.packages("pacman")
```

```
## Loading required package: pacman
```

```
p_load(Sleuth3, bootstrap)

#Load Data
data <- case0102

#Box Plot
boxplot(data$Salary)
```



```
#There is an outlier for the combined data

#Density
```

```
#Check missing
any(is.na(data))
```

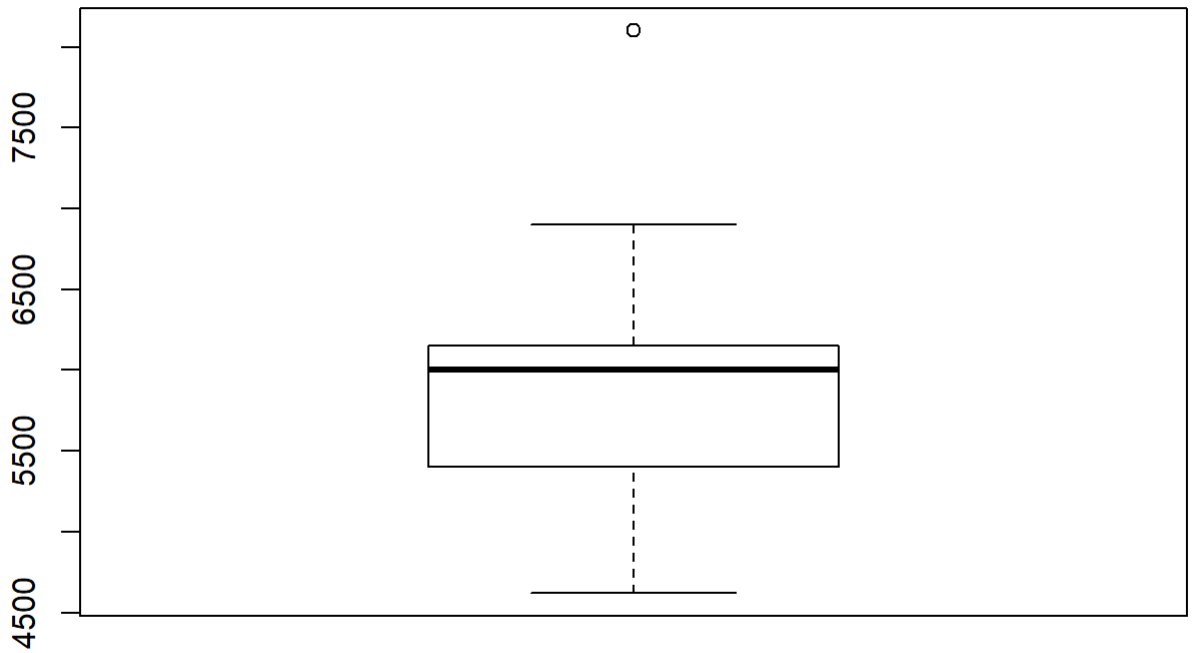
```
## [1] FALSE
```

```
#There is no missing values for salaries and gender
```

ii)

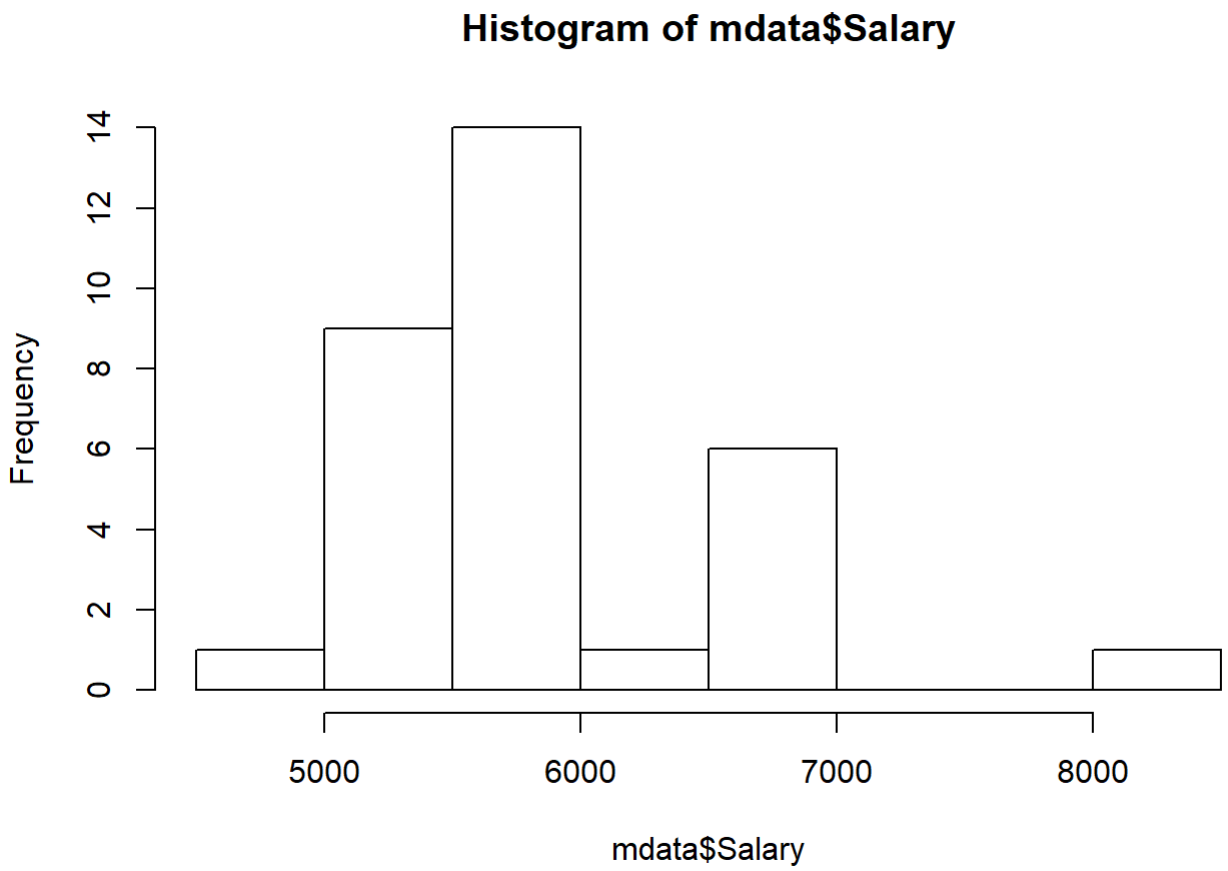
```
mdata <- data[data$Sex == "Male",]
fdata <- data[data$Sex == "Female",]

#Male Boxplot
boxplot(mdata$Salary)
```



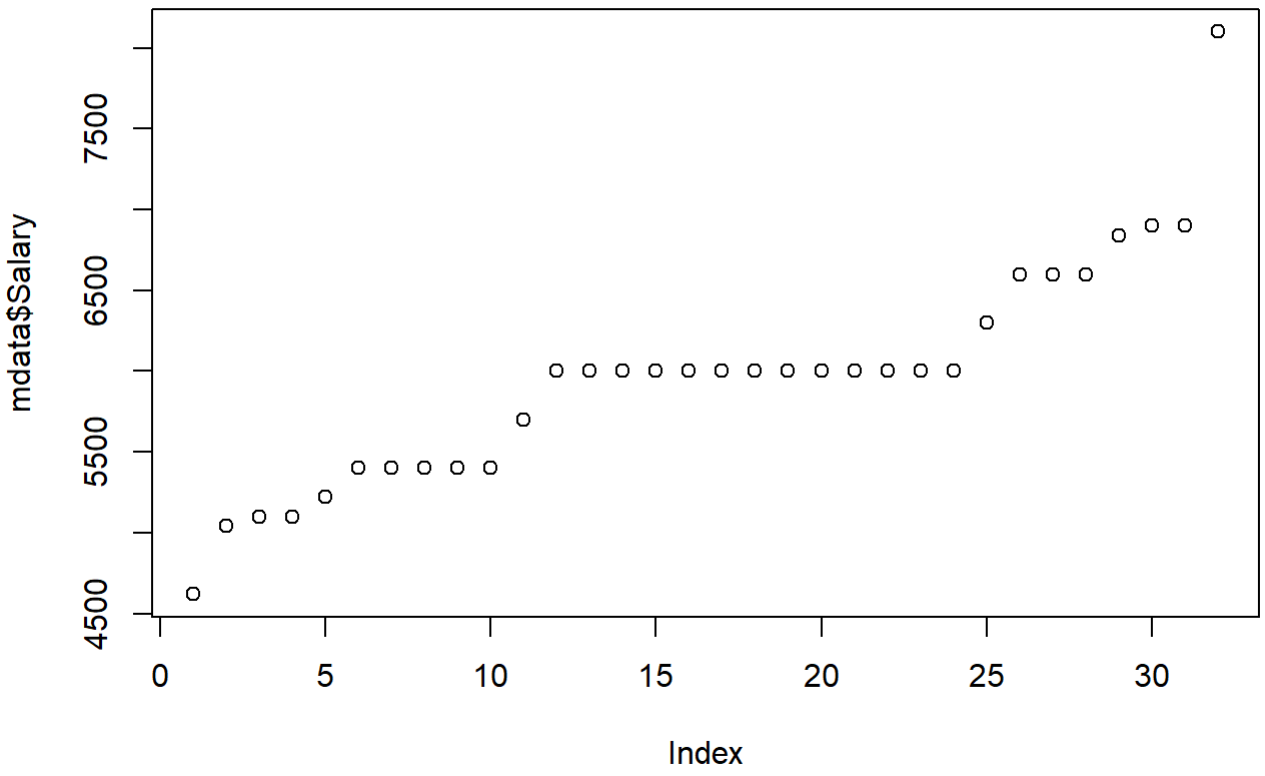
```
#There is an outlier for the male data

#Histogram
hist(mdata$Salary)
```



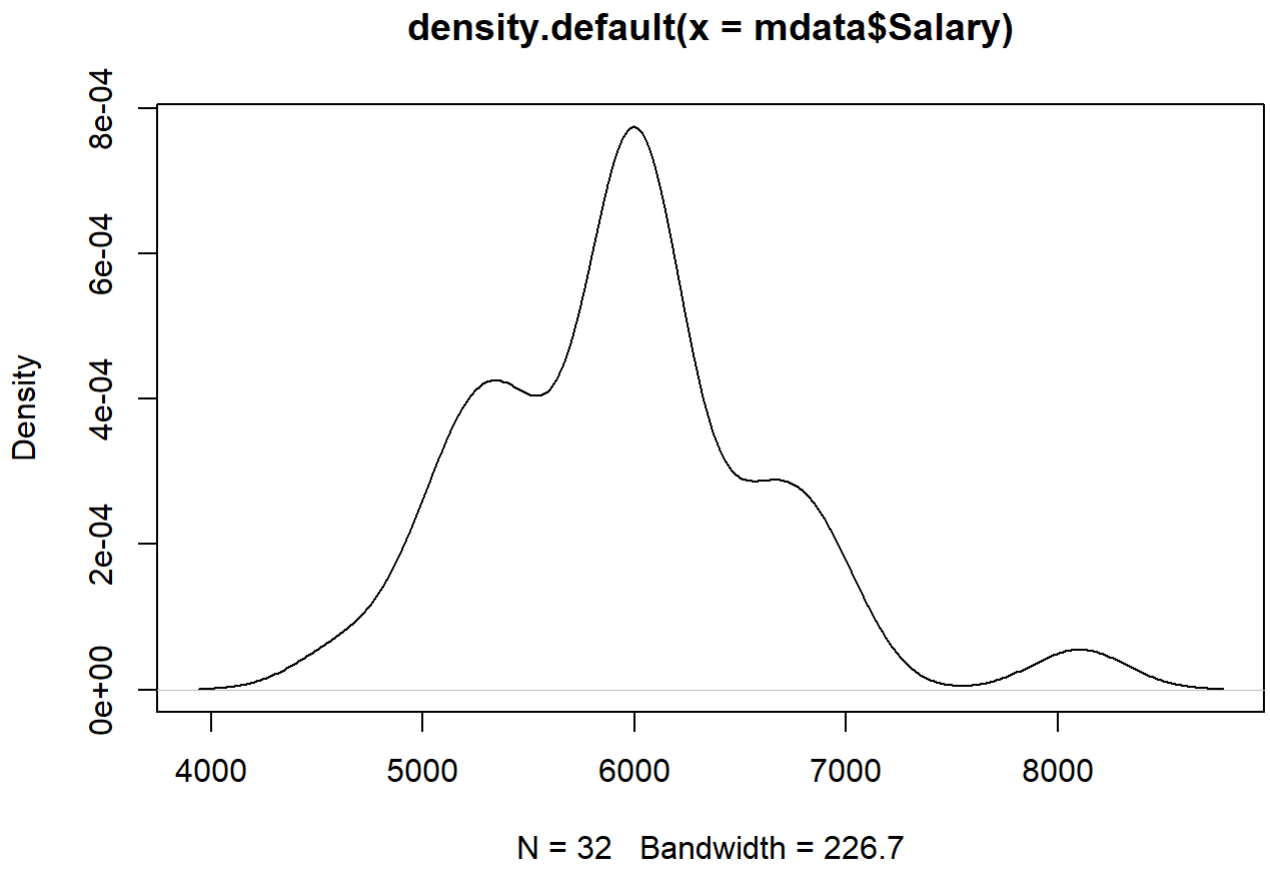

```
#Hard to tell the distribution of Male Data Salary

#Scatterplot
plot(mdata$Salary)
```

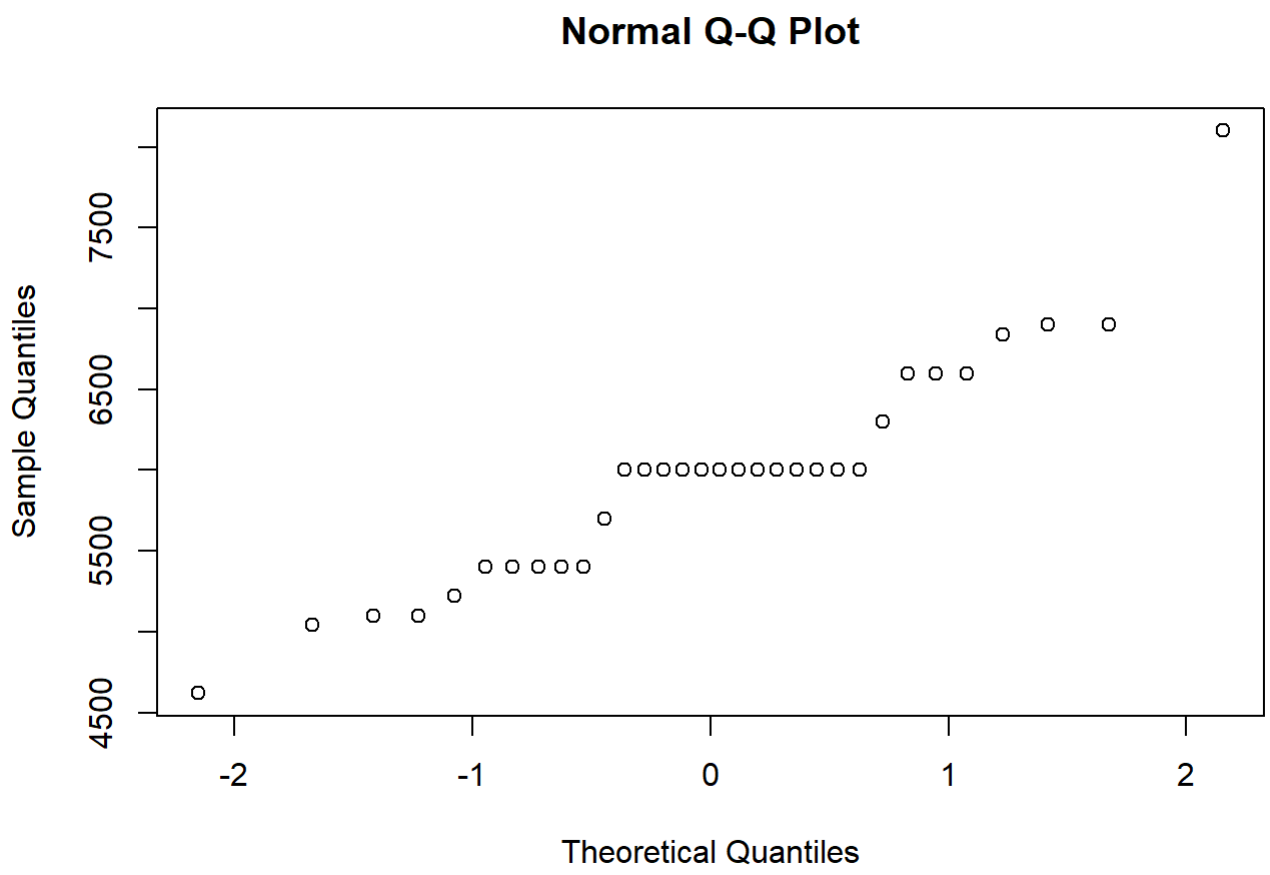



```
#Salray Increasing as index increases

#Density
plot(density(mdata$Salary))
```




```
#QQ
qqnorm(mdata$Salary)
```




```
#SD
sd(mdata$Salary)
```

```
## [1] 690.7333
```

```
#Var
var(mdata$Salary)
```

```
## [1] 477112.5
```

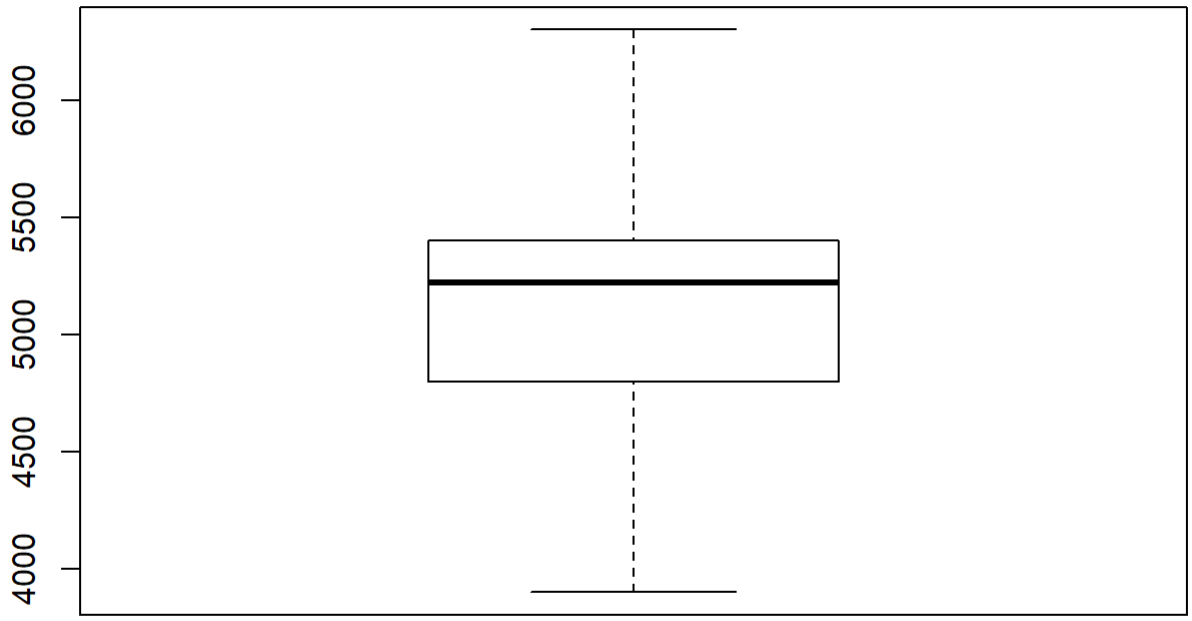
```
#IQR
IQR(mdata$Salary)
```

```
## [1] 675
```

```
#Summary
summary(mdata$Salary)
```

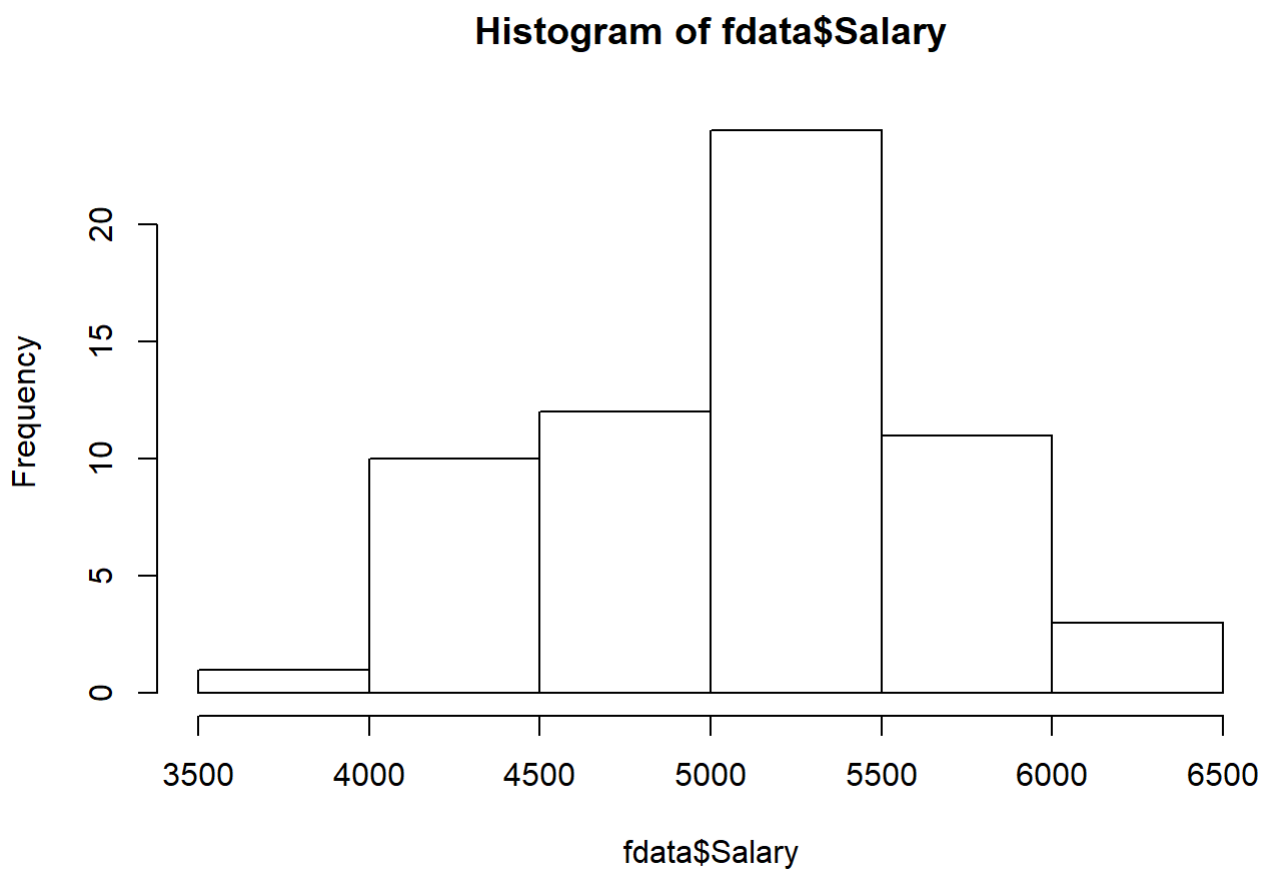
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4620    5400    6000    5957    6075    8100
```

```
#Female Boxplot
boxplot(fdata$Salary)
```



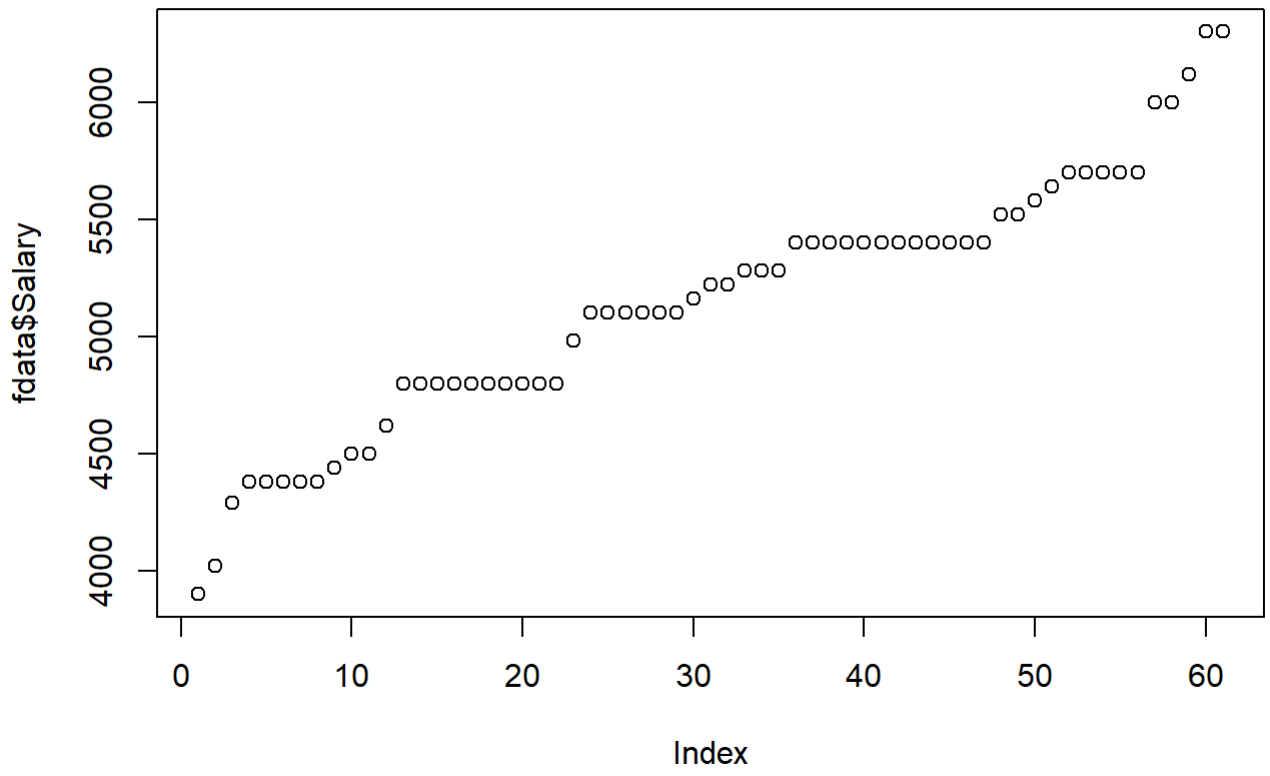

```
#There is no outlier for the female data

#Histogram
hist(fdata$Salary)
```



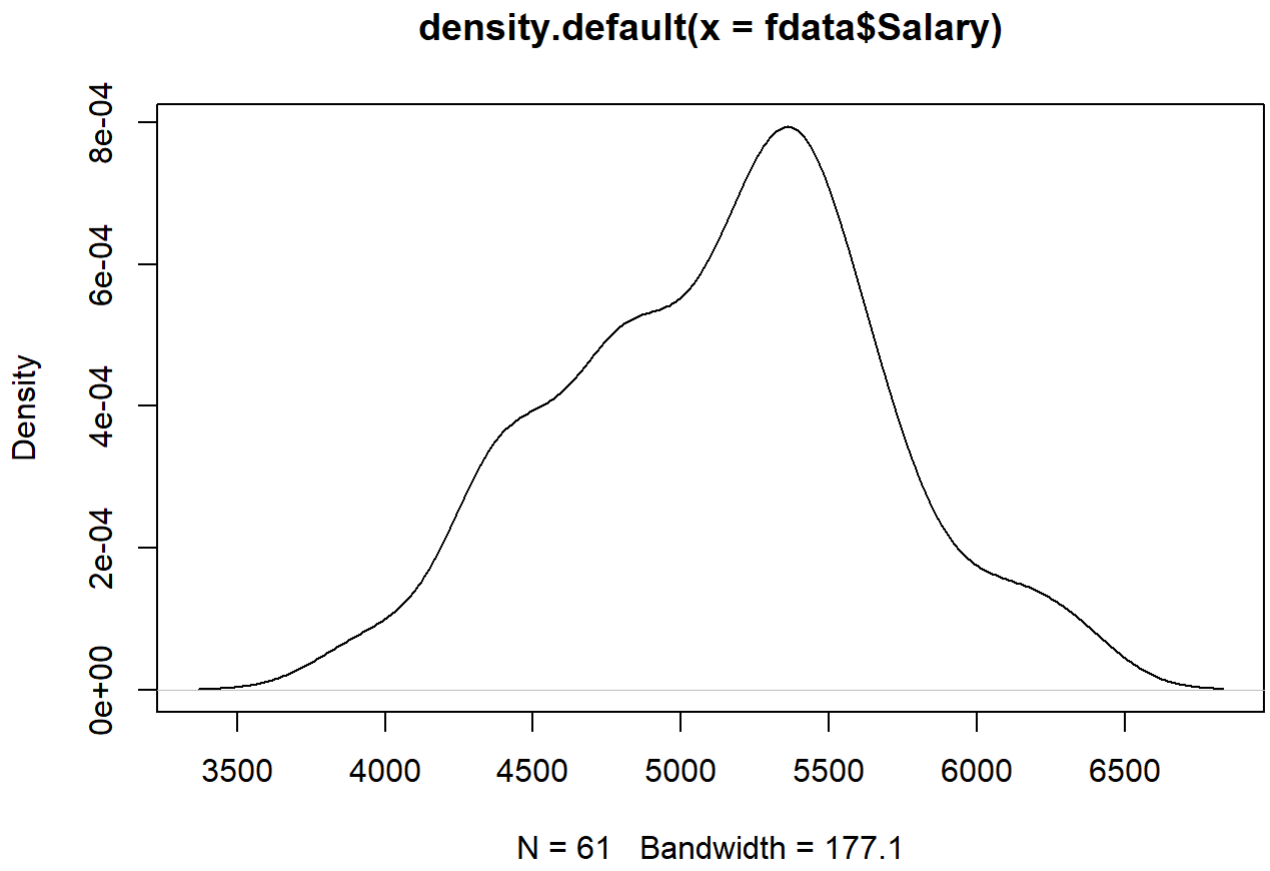

```
#Distribution seems to be normal

#Scatterplot
plot(fdata$Salary)
```

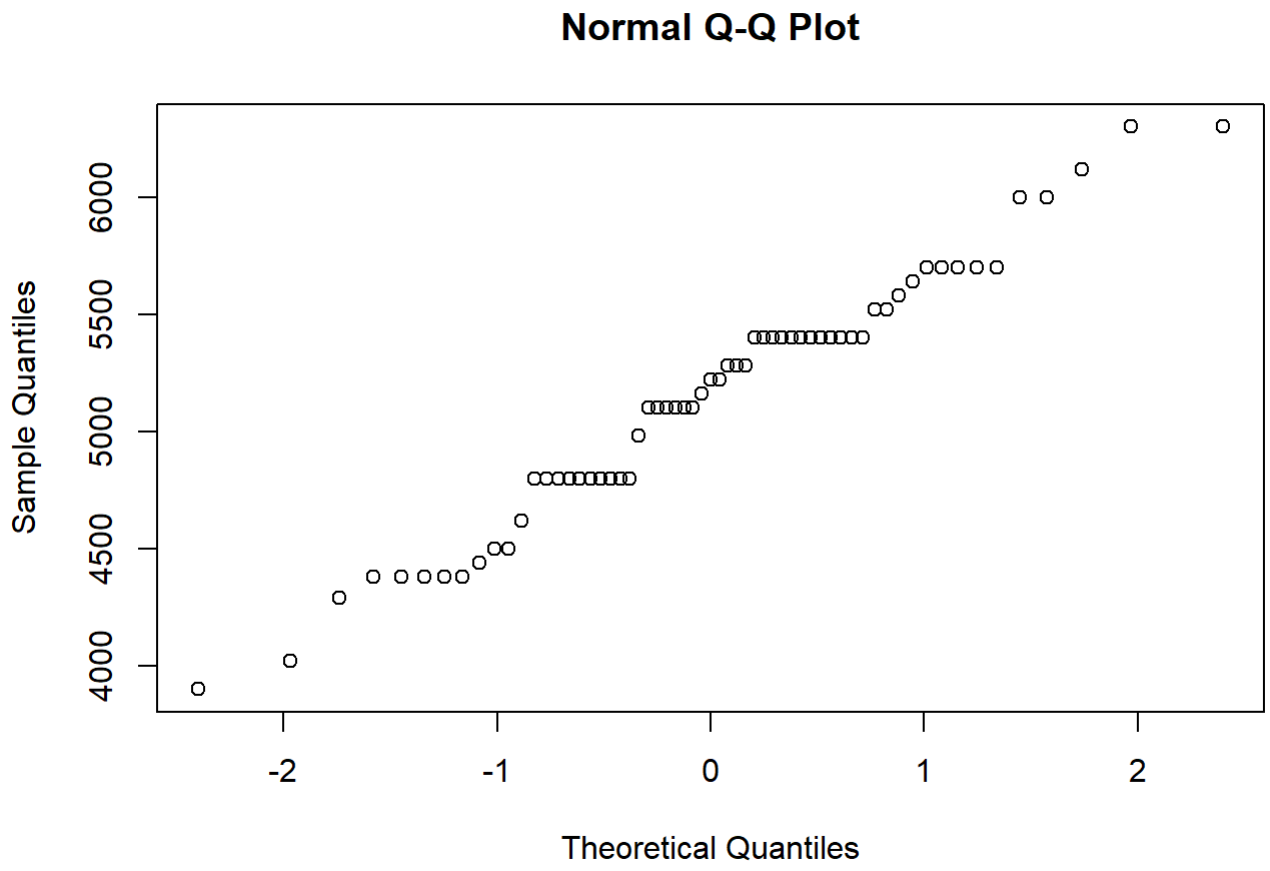



```
#Salary increasing as index increases

#Density
plot(density(fdata$Salary))
```




```
#QQ
qqnorm(fdata$Salary)
```




```
#SD
sd(fdata$Salary)
```

```
## [1] 539.8707
```

```
#Var
var(fdata$Salary)
```

```
## [1] 291460.3
```

```
#IQR
IQR(fdata$Salary)
```

```
## [1] 600
```

```
#Summary
summary(fdata$Salary)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3900    4800    5220    5139    5400    6300
```

iii)

```
#Male Jackknife

#Mean estimator
mj_mean <- jackknife(mdata$Salary, mean)
mj_mean
```

```
## $jack.se
## [1] 122.1056
##
## $jack.bias
## [1] 0
##
## $jack.values
##  [1] 6000.000 5986.452 5984.516 5984.516 5980.645 5974.839 5974.839
##  [8] 5974.839 5974.839 5974.839 5965.161 5955.484 5955.484 5955.484
## [15] 5955.484 5955.484 5955.484 5955.484 5955.484 5955.484 5955.484
## [22] 5955.484 5955.484 5955.484 5945.806 5936.129 5936.129 5936.129
## [29] 5928.387 5926.452 5926.452 5887.742
##
## $call
## jackknife(x = mdata$Salary, theta = mean)
```

```
#SD estimator
mj_sd <- jackknife(mdata$Salary, sd)
mj_sd
```

```
## $jack.se
## [1] 124.8158
##
## $jack.bias
## [1] -11.28011
##
## $jack.values
##  [1] 656.9018 681.2418 683.9242 683.9242 688.7183 694.5112 694.5112
##  [8] 694.5112 694.5112 694.5112 700.5325 702.1056 702.1056 702.1056
## [15] 702.1056 702.1056 702.1056 702.1056 702.1056 702.1056 702.1056
## [22] 702.1056 702.1056 702.1056 699.2604 691.9426 691.9426 691.9426
## [29] 682.7742 680.0076 680.0076 578.7729
##
## $call
## jackknife(x = mdata$Salary, theta = sd)
```

```
#Median estimator
mj_median <- jackknife(mdata$Salary,median)
mj_median
```

```
## $jack.se
## [1] 0
##
## $jack.bias
## [1] 0
##
## $jack.values
##  [1] 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000
## [15] 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000 6000
## [29] 6000 6000 6000 6000
##
## $call
## jackknife(x = mdata$Salary, theta = median)
```

```
#IQR estimator
mj_iqr <- jackknife(mdata$Salary, IQR)
mj_iqr
```

```
## $jack.se
## [1] 361.6369
##
## $jack.bias
## [1] 1162.5
##
## $jack.values
##  [1] 750 750 750 750 750 750 750 750 750 750 750 750 750 750 750 750 750 750
## [18] 750 750 750 750 750 750 750 600 600 600 600 600 600 600 600
##
## $call
## jackknife(x = mdata$Salary, theta = IQR)
```

```
#Female Jackknife

#Mean estimator
fj_mean <- jackknife(fdata$Salary, mean)
fj_mean
```

```
## $jack.se
## [1] 69.12335
##
## $jack.bias
## [1] 0
##
## $jack.values
##  [1] 5159.5 5157.5 5153.0 5151.5 5151.5 5151.5 5151.5 5151.5 5150.5 5149.5
## [11] 5149.5 5147.5 5144.5 5144.5 5144.5 5144.5 5144.5 5144.5 5144.5 5144.5
## [21] 5144.5 5144.5 5141.5 5139.5 5139.5 5139.5 5139.5 5139.5 5139.5 5138.5
## [31] 5137.5 5137.5 5136.5 5136.5 5136.5 5134.5 5134.5 5134.5 5134.5 5134.5
## [41] 5134.5 5134.5 5134.5 5134.5 5134.5 5134.5 5134.5 5132.5 5132.5 5131.5
## [51] 5130.5 5129.5 5129.5 5129.5 5129.5 5129.5 5124.5 5124.5 5122.5 5119.5
## [61] 5119.5
##
## $call
## jackknife(x = fdata$Salary, theta = mean)
```

```
#SD estimator
fj_sd <- jackknife(fdata$Salary, sd)
fj_sd
```

```
## $jack.se
## [1] 45.84659
##
## $jack.bias
## [1] -1.946738
##
## $jack.values
##  [1] 519.5710 524.2416 532.9016 535.2358 535.2358 535.2358 535.2358
##  [8] 535.2358 536.6419 537.9289 537.9289 540.1495 542.6065 542.6065
## [15] 542.6065 542.6065 542.6065 542.6065 542.6065 542.6065 542.6065
## [22] 542.6065 544.0271 544.4027 544.4027 544.4027 544.4027 544.4027
## [29] 544.4027 544.4195 544.3224 544.3224 544.1112 544.1112 544.1112
## [36] 543.3463 543.3463 543.3463 543.3463 543.3463 543.3463 543.3463
## [43] 543.3463 543.3463 543.3463 543.3463 543.3463 542.1227 542.1227
## [50] 541.3380 540.4374 539.4204 539.4204 539.4204 539.4204 539.4204
## [57] 532.5615 532.5615 528.9729 522.6543 522.6543
##
## $call
## jackknife(x = fdata$Salary, theta = sd)
```

```
#Median estimator
fj_median <- jackknife(fdata$Salary,median)
fj_median
```

```
## $jack.se
## [1] 116.1739
##
## $jack.bias
## [1] -914.7541
##
## $jack.values
##  [1] 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220
## [15] 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220 5220
## [29] 5220 5220 5190 5190 5190 5190 5190 5190 5190 5190 5190 5190 5190 5190
## [43] 5190 5190 5190 5190 5190 5190 5190 5190 5190 5190 5190 5190 5190 5190
## [57] 5190 5190 5190 5190 5190
##
## $call
## jackknife(x = fdata$Salary, theta = median)
```

```
#IQR estimator
fj_iqr <- jackknife(fdata$Salary, IQR)
fj_iqr
```

```
## $jack.se
## [1] 0
##
## $jack.bias
## [1] 0
##
## $jack.values
##  [1] 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600
## [18] 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600
## [35] 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600 600
## [52] 600 600 600 600 600 600 600 600 600 600
##
## $call
## jackknife(x = fdata$Salary, theta = IQR)
```

```
#Male Bootstrap

#Mean estimator
mb_mean <- bootstrap(mdata$Salary,10000,mean)
mean(mb_mean$thetastar)
```

```
## [1] 5956.172
```

```
mb_meanbias <-mean(mb_mean$thetastar)-mean(mdata$Salary)
mb_meanbias
```

```
## [1] -0.7025625
```

```
mb_meanvar <- var(mb_mean$thetastar)
mb_meanvar
```

```
## [1] 14341.66
```

```
#Median estimator
mb_median <- bootstrap(mdata$Salary,10000,median)
mean(mb_median$thetastar)
```

```
## [1] 5985.33
```

```
mb_medianbias <-mean(mb_median$thetastar)-median(mdata$Salary)
mb_medianbias
```

```
## [1] -14.67
```

```
mb_medianvar <- var(mb_median$thetastar)
mb_medianvar
```

```
## [1] 6004.392
```

```
#SD estimator
mb_sd <- bootstrap(mdata$Salary,10000,sd)
mean(mb_sd$thetastar)
```

```
## [1] 671.322
```

```
mb_sdbias <-mean(mb_sd$thetastar)-sd(mdata$Salary)
mb_sdbias
```

```
## [1] -19.41129
```

```
mb_sdvar <- var(mb_sd$thetastar)
mb_sdvar
```

```
## [1] 12325.97
```

```
#IQR estimator
mb_iqr <- bootstrap(mdata$Salary,10000,IQR)
mean(mb_iqr$thetastar)
```

```
## [1] 715.563
```

```
mb_iqrbias <-mean(mb_iqr$thetastar)-IQR(mdata$Salary)
mb_iqrbias
```

```
## [1] 40.563
```

```
mb_iqrvar <- var(mb_iqr$thetastar)
mb_iqrvar
```

```
## [1] 84183.91
```

```
#Female Bootstrap

#Mean estimator
fb_mean <- bootstrap(fdata$Salary,10000,mean)
mean(fb_mean$thetastar)
```

```
## [1] 5138.837
```

```
fb_meanbias <-mean(fb_mean$thetastar)-mean(fdata$Salary)
fb_meanbias
```

```
## [1] -0.0157377
```

```
fb_meanvar <- var(fb_mean$thetastar)
fb_meanvar
```

```
## [1] 4642.869
```

```
#Median estimator
fb_median <- bootstrap(fdata$Salary,10000,median)
mean(fb_median$thetastar)
```

```
## [1] 5201.316
```

```
fb_medianbias <-mean(fb_median$thetastar)-median(fdata$Salary)
fb_medianbias
```

```
## [1] -18.684
```

```
fb_medianvar <- var(fb_median$thetastar)
fb_medianvar
```

```
## [1] 12772.75
```

```
#SD estimator
fb_sd <- bootstrap(fdata$Salary,10000,sd)
mean(fb_sd$thetastar)
```

```
## [1] 533.4943
```

```
fb_sdbias <-mean(fb_sd$thetastar)-sd(fdata$Salary)
fb_sdbias
```

```
## [1] -6.376407
```

```
fb_sdvar <- var(fb_sd$thetastar)
fb_sdvar
```

```
## [1] 1960.916
```

```
#IQR estimator
fb_iqr <- bootstrap(fdata$Salary,10000,IQR)
mean(fb_iqr$thetastar)
```

```
## [1] 676.764
```

```
fb_iqrbias <-mean(fb_iqr$thetastar)-IQR(fdata$Salary)
fb_iqrbias
```

```
## [1] 76.764
```

```
fb_iqrvar <- var(fb_iqr$thetastar)
fb_iqrvar
```

```
## [1] 14928.3
```