HW3.R

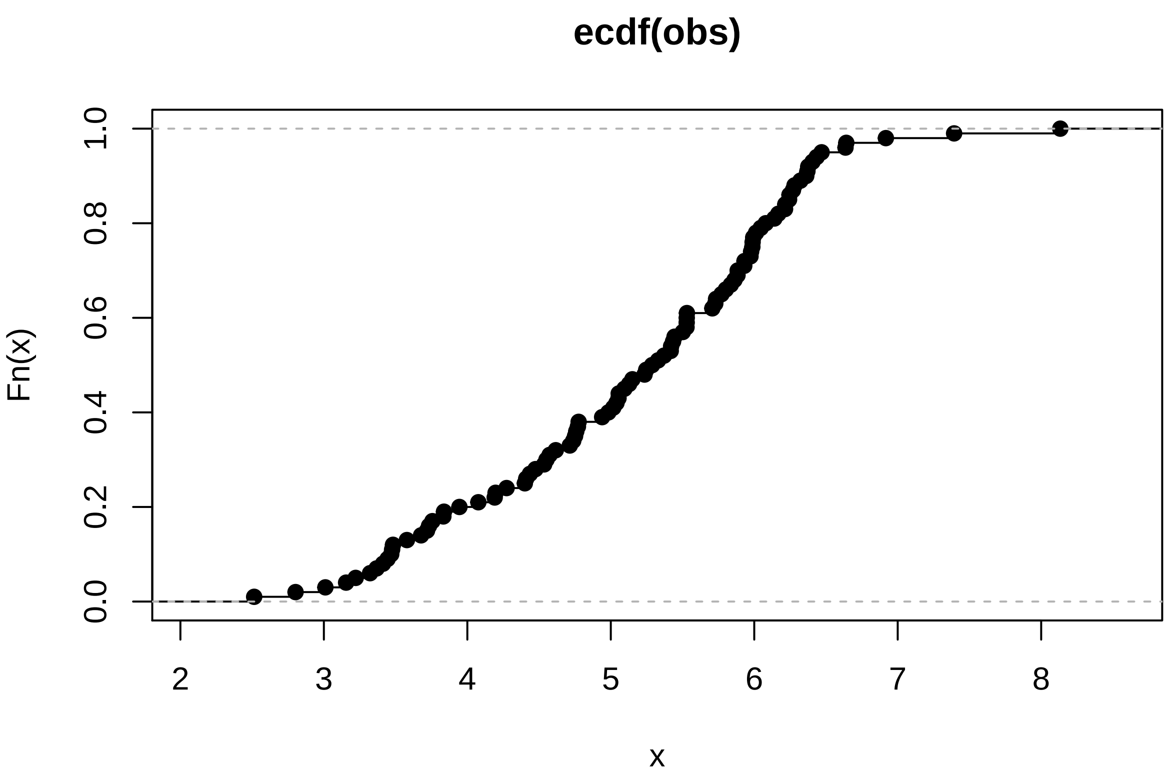
jiayuan

Fri Sep 25 11:26:07 2015

```
#p117.6
#a
setwd("/Users/jiayuan/Documents/MA681")
options(digits=2)
obs<-rnorm(n=100, mean = 5, sd = 1)
obs
```

```
##   [1] 2.8 6.2 3.7 5.9 6.1 5.9 5.0 4.2 3.4 3.3 3.9 6.0 5.8 4.8 6.1 3.7 4.4
##  [18] 5.1 6.4 5.4 4.8 5.9 4.6 4.4 5.2 3.2 4.1 6.0 6.6 5.0 2.5 4.2 4.3 5.1
##  [35] 6.3 4.6 4.7 5.5 4.6 3.6 6.4 5.8 5.2 5.3 3.4 5.4 5.8 5.0 6.2 3.8 5.7
##  [52] 6.0 6.4 5.9 5.9 5.5 3.2 4.4 6.5 5.4 6.0 5.5 3.5 5.3 5.7 4.8 3.4 3.0
##  [69] 5.4 5.5 6.3 3.8 5.7 6.9 5.4 5.1 5.2 6.2 5.1 6.2 6.6 8.1 6.4 7.4 4.7
##  [86] 3.7 6.3 4.9 3.5 6.2 6.0 5.5 6.0 4.8 6.0 6.4 3.5 4.5 3.8 4.5
```

```
th.hat<-exp(mean(obs))
plot(ecdf(obs))
```




```
obs.n <- length(obs)
B <- 1000
Tboot <- rep(0,B)
for(i in 1:B){
```

```r
  obs.s <- sample(obs, obs.n, replace=TRUE)
  Tboot[i] <- exp(mean(obs.s))
}
se <- sqrt(var(Tboot))
se
```

```
## [1] 19
```

```r
qnorm(.025,0,1) #-2
```

```
## [1] -2
```

```r
qnorm(.975,0,1) #2
```

```
## [1] 2
```

```r
Normal <- c(th.hat - 2*se, th.hat + 2*se)
Percentile <- c(quantile(Tboot,.025),quantile(Tboot,.975))
pivotal <- c((2*th.hat - quantile(Tboot, .975)),(2*th.hat - quantile(Tboot, .025)))

cat("Method       95% Interval\n")
```

```
## Method       95% Interval
```

```r
cat("Normal      (", Normal[1], ",   ",Normal[2], ")\n")
```

```
## Normal      ( 132 ,    209 )
```

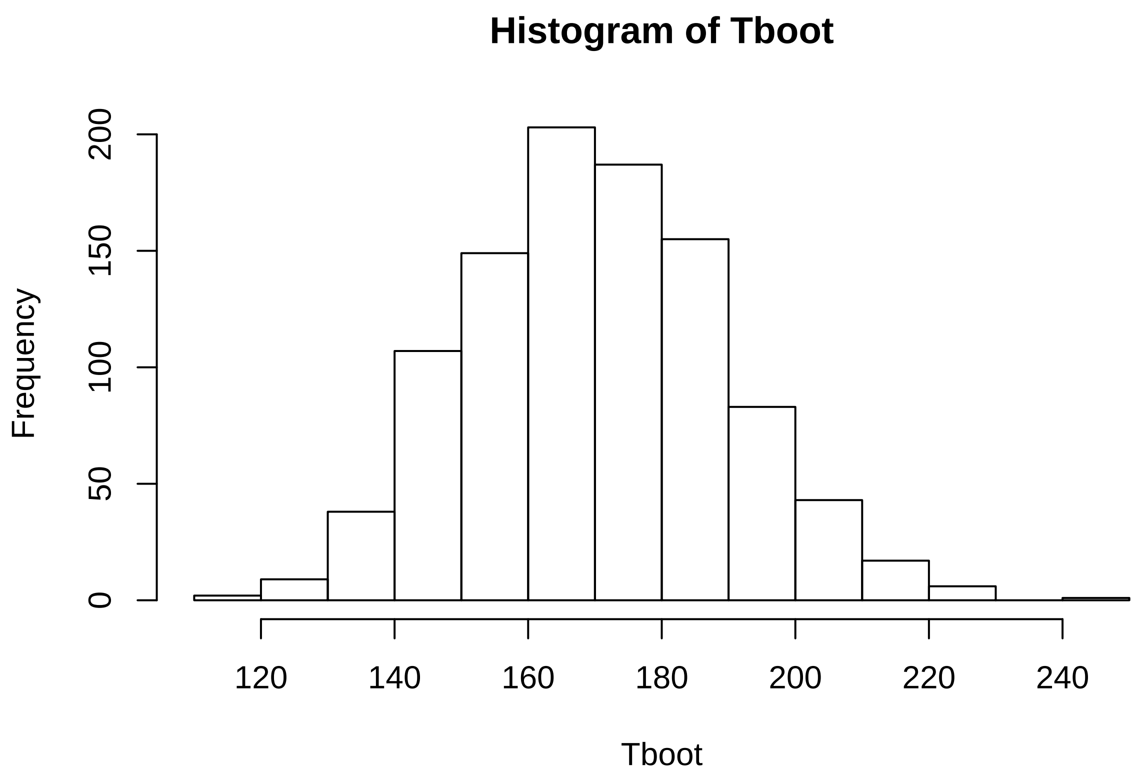
```r
cat("Pivotal     (", pivotal[1], ",     ", pivotal[2], ") \n")
```

```
## Pivotal     ( 131 ,      206 )
```

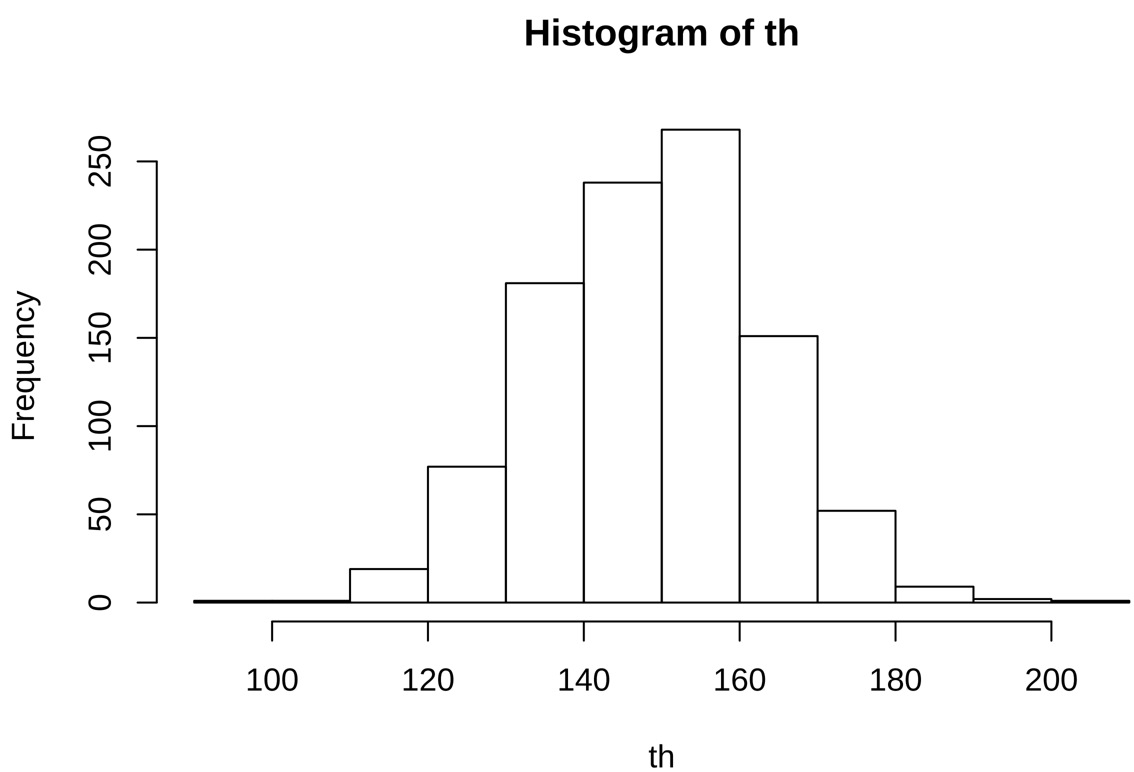
```r
cat("Percentile  (", Percentile[1], ",    ", Percentile[2], ") \n")
```

```
## Percentile  ( 134 ,     210 )
```

```r
#b
hist(Tboot) #estimate of the distribution of theta hat
```



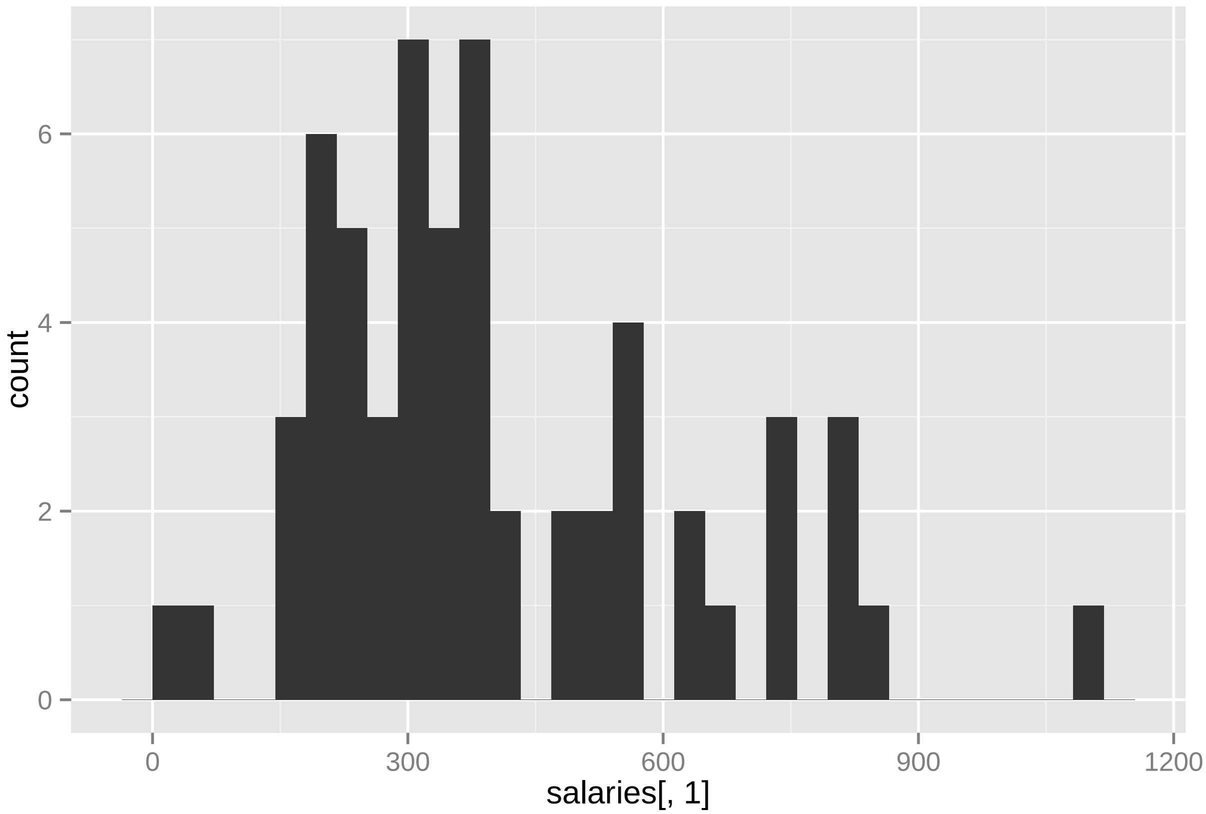

```r
th.mean<-exp(5)
th.var<-((exp(5))^2)*1/100
th<-rnorm(1000, mean = th.mean, sd = sqrt(th.var))
hist(th)
```



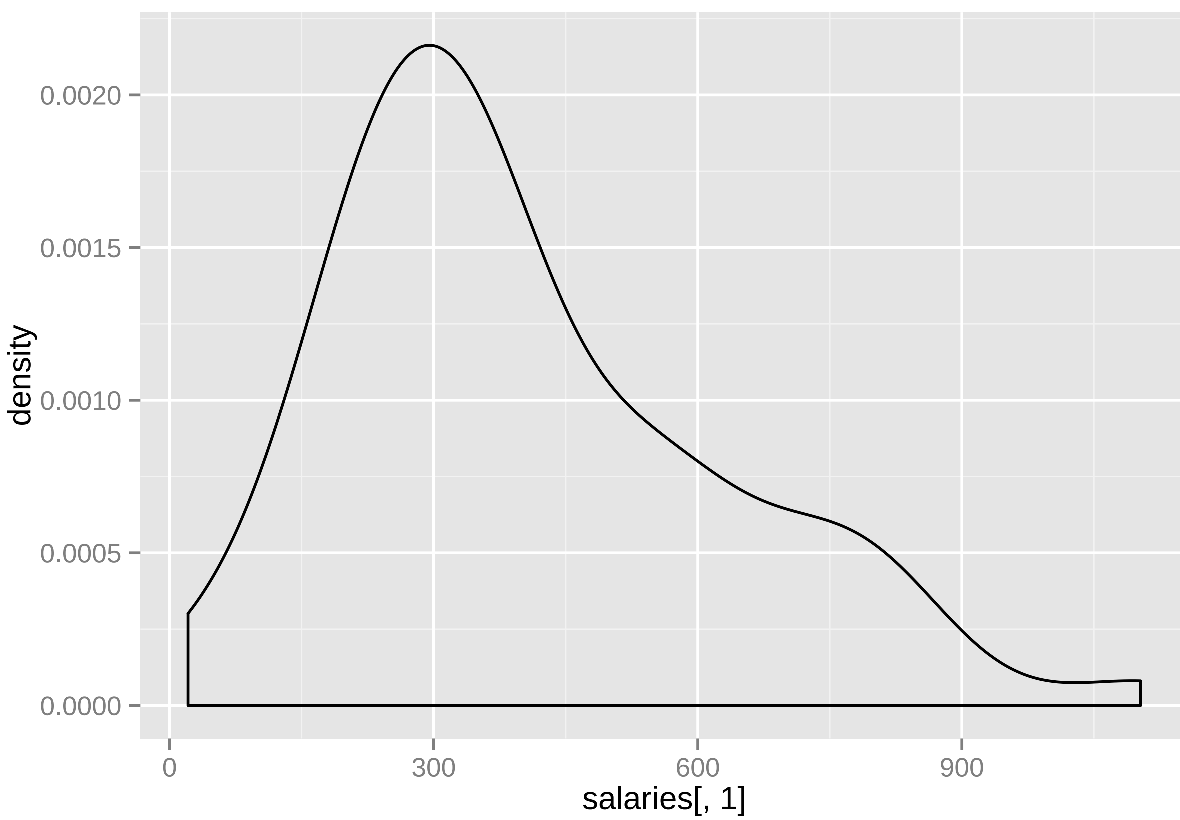
```r
#true sampling distribution of theta hat
#The histogram of the bootstrap replications and the true sampling distribution
#have a similar shape (normal distribution).

#2
#a
salaries <- read.csv("CEO compensation.csv", header=TRUE)
hist(salaries[,1])
library(ggplot2)
```


```r
qplot(salaries[,1],data=salaries,geom = "histogram")
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```




```
qplot(salaries[,1],data=salaries,geom = "density")
```



```r
#From the above plots, I observe that most of the CEOs' salaries are from 200 to 400.
#I think the confidence interval for the median small company CEO salary will fall
#into the interval of (200,400)

#b
sal <- salaries[,1]
sal.n <- length(sal)
th.hat<-median(sal)
B <- 1000
Tboot <- rep(0,B)
for(i in 1:B){
  sal.s <- sample(sal, sal.n, replace=TRUE)
  Tboot[i] <- median(sal.s)
}
se <- sqrt(var(Tboot))
se
```

```
## [1] 28
```

```r
qnorm(.05,0,1) #-1.6
```

```
## [1] -1.6
```

```r
qnorm(.95,0,1) #1.6
```

```
## [1] 1.6
```

```r
Normal <- c(th.hat - 1.6*se, th.hat + 1.6*se)
Percentile <- c(quantile(Tboot,.05),quantile(Tboot,.95))
pivotal <- c((2*th.hat - quantile(Tboot, .95)),(2*th.hat - quantile(Tboot, .05)))

cat("Method       90% Interval\n")
```

```
## Method       90% Interval
```

```r
cat("Normal      (", Normal[1], ",   ",Normal[2], ")\n")
```

```
## Normal      ( 306 ,    394 )
```

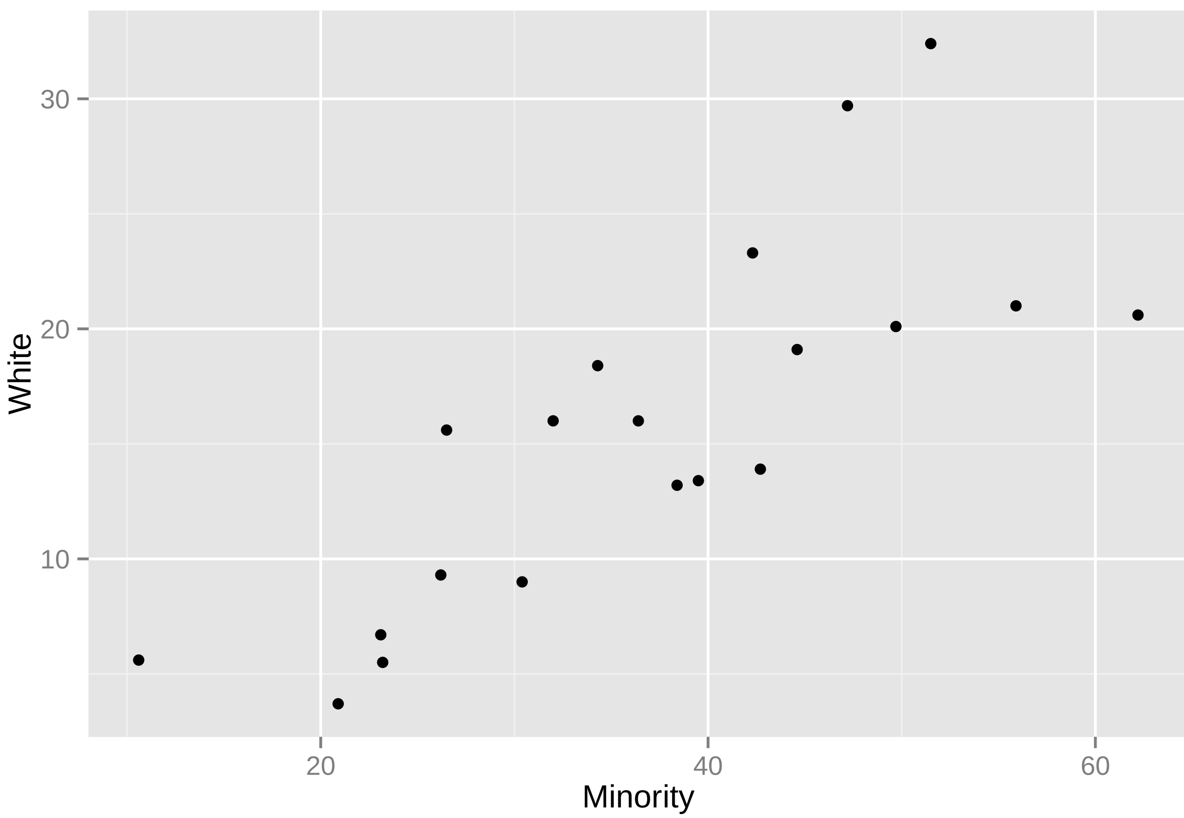
```r
cat("Pivotal     (", pivotal[1], ",     ", pivotal[2], ") \n")
```

```
## Pivotal     ( 310 ,      402 )
```

```r
cat("Percentile  (", Percentile[1], ",     ", Percentile[2], ") \n")
```

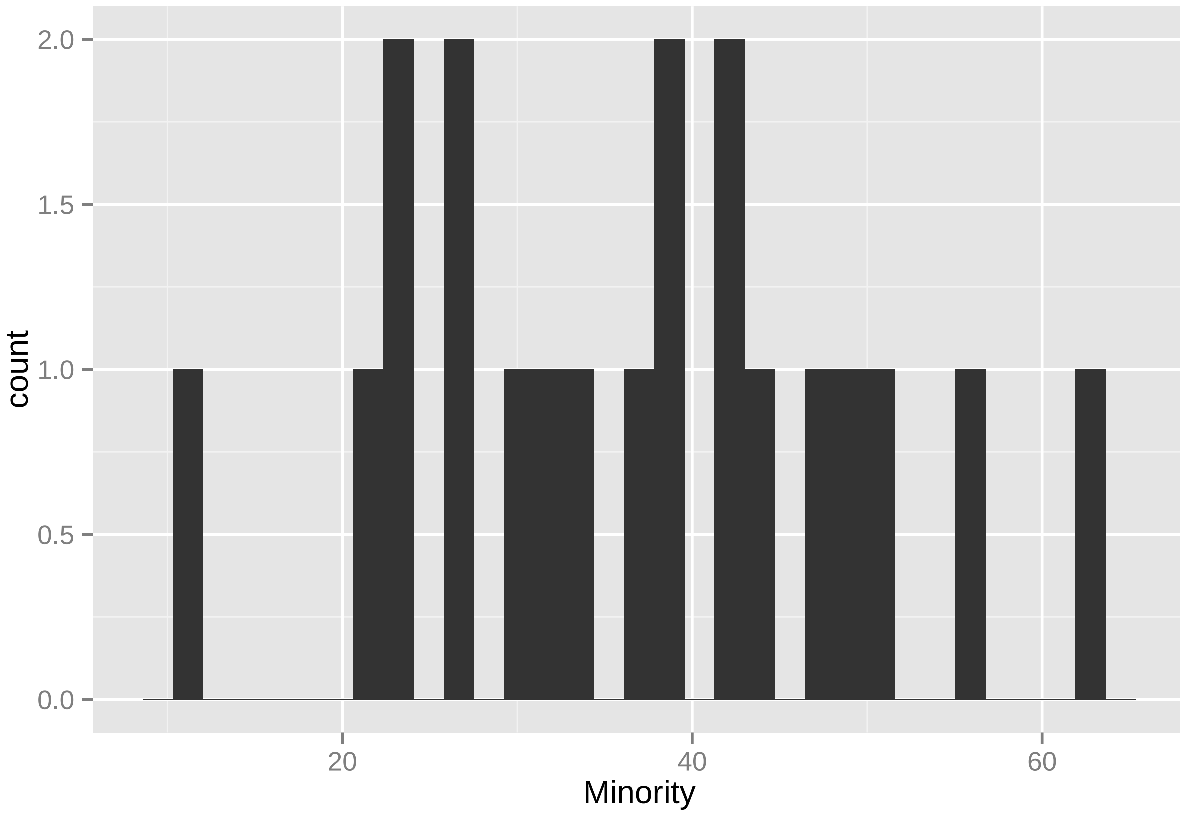
```
## Percentile  ( 298 ,     390 )
```

```
#3
#a
options(digits=3)
refusals <- read.csv("Mortgage Refusals.csv", header=TRUE)
attach(refusals)
M<-as.matrix(Loan.Refusals[2:21])
W<-as.matrix(X.1[2:21])
M<-as.numeric(M[,1])
W<-as.numeric(W[,1])
library(ggplot2)
qplot(M,W,xlab="Minority",ylab="White")
```



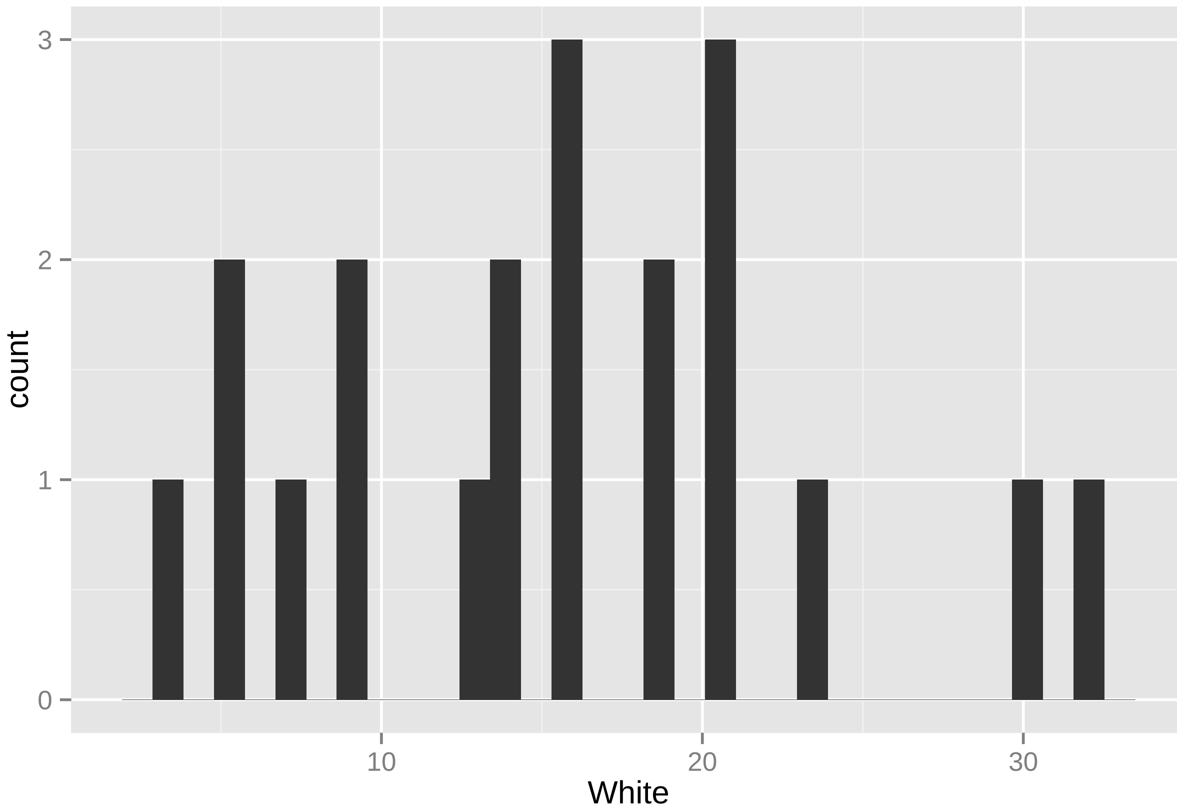

```
#From the plot, we can see that for most of the banks, the refusal rate might be higher
#for minority applicants than white applicants.
qplot(M,geom="histogram",xlab="Minority")
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```




```
qplot(W,geom="histogram",xlab="White")
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```




```
#Comaping with the two hisgrams,
#we can see that most of minority applicants has the the refusal rate from 20 to 60.
#In contrast, most of white applicants only has the the refusal rate from 5 to 20.
#We can also conclude that the refusal rate might be higher
#for minority applicants than white applicants.

#b
t.test(M, W, alternative = c("two.sided"), paired = TRUE, var.equal = TRUE,conf.level = 0.95)
```

```
##
##  Paired t-test
##
## data:  M and W
## t = 10, df = 20, p-value = 6e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  17.4 25.1
## sample estimates:
## mean of the differences
##                    21.3
```

```
#H0:difference=0; H1:H0 is not true
#The t statistics is 10, and the p-value is 6e-10>0.05.
#We reject the null hypothesis that the true difference in means is equal to 0.
#So we can not conclude that the means of the refusal rates for the minority
#and white applicants are significantly similar.
#Then the saying that,"It appears that the refusal rate might be higher
```

```r
#for minority applicants than white applicants." is true.
#95% conference interval for the difference in the population means is (17.4, 25.1).

#c
M.n <- length(M)
W.n <- length(W)
th.hat <- mean(M)-mean(W)
B <- 1000
Tboot <- rep(0,B)
for(i in 1:B){
  M.s <- sample(M, M.n, replace=TRUE)
  W.s <- sample(W, W.n, replace=TRUE)
  Tboot[i] <- mean(M.s) - mean(W.s)
}
se <- sqrt(var(Tboot))
se
```

```
## [1] 3.4
```

```r
qnorm(.025,0,1) #-2
```

```
## [1] -1.96
```

```r
qnorm(.975,0,1) #2
```

```
## [1] 1.96
```

```r
Normal <- c(th.hat - 2*se, th.hat + 2*se)
Percentile <- c(quantile(Tboot,.025),quantile(Tboot,.975))
pivotal <- c((2*th.hat - quantile(Tboot, .975)),(2*th.hat - quantile(Tboot, .025)))

cat("Method       95% Interval\n")
```

```
## Method       95% Interval
```

```r
cat("Normal      (", Normal[1], ",   ",Normal[2], ")\n")
```

```
## Normal      ( 14.5 ,    28.1 )
```

```r
cat("Pivotal     (", pivotal[1], ",     ", pivotal[2], ") \n")
```

```
## Pivotal     ( 14.6 ,      27.8 )
```

```r
cat("Percentile  (", Percentile[1], ",    ", Percentile[2], ") \n")
```

```
## Percentile  ( 14.7 ,     27.9 )
```

```
#95% confidence interval using the bootstrap standard error is
#(14.9, 27.6), (14.9, 27.4), (15.1, 27.6)  by differnt ways.
#These result are wider than the interval I calculated in part (b), which is (17.4, 25.1).
```