Homework 6

Joshua Oswari - A14751270

4/27/2019

Problem 1

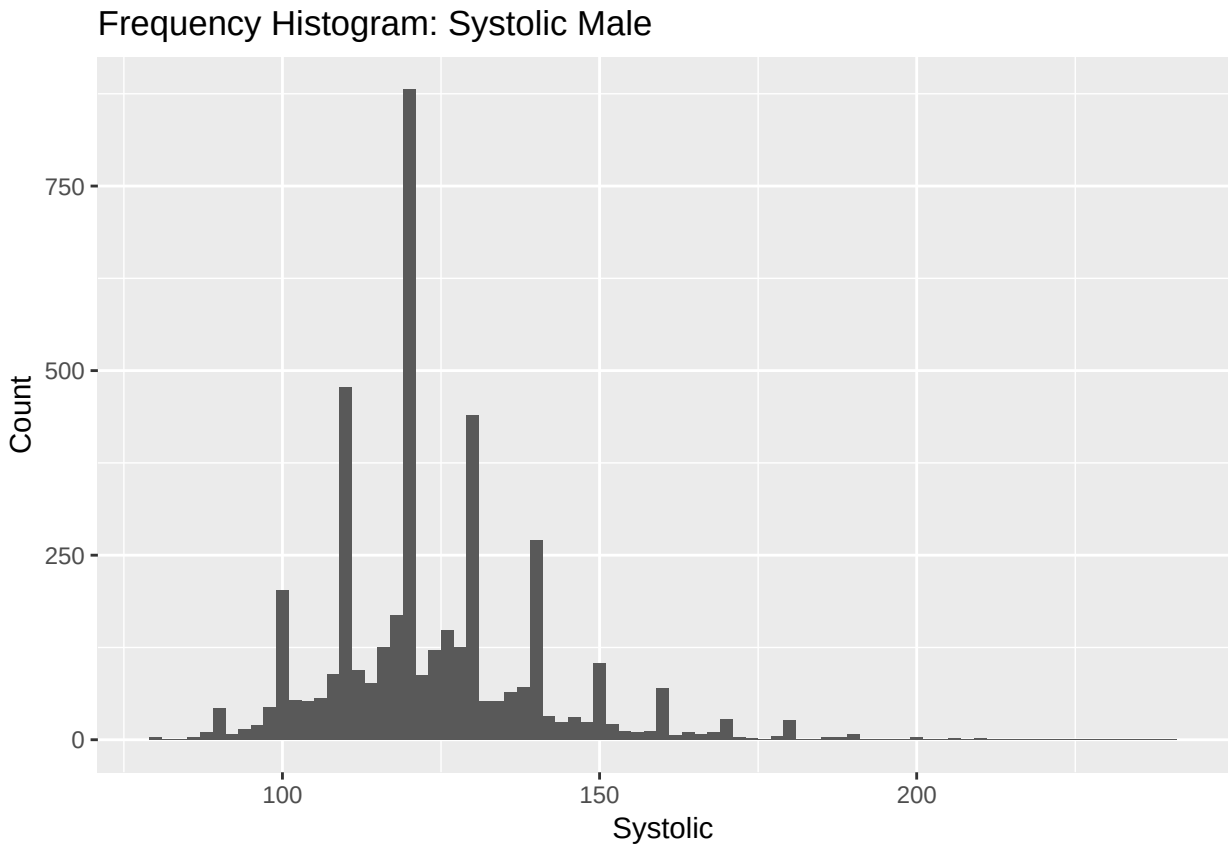
```
#used from lecture
dat = read.csv("GSE7621.csv", sep ="", header = TRUE)

dat = dat[,-1] # remove the ID
m = nrow(dat)
pval = numeric(m)
for (i in 1:m) {
    pval[i] = t.test(dat[i,1:9], dat[i,-(1:9)])$p.value
    }

# corrected p-values
pval.bh = p.adjust(pval, "BH")
pval.by = p.adjust(pval, "BY")

# rejections at the 20% level without adjustement for multiple testing
reject = (pval <= 0.20)
R = sum(reject) # total number of rejections

# rejections at the 20% FDR level
reject.bh = (pval.bh <= 0.20)
R.bh = sum(reject.bh)
reject.by = (pval.by <= 0.20)
R.by = sum(reject.by)
```

Problem 2

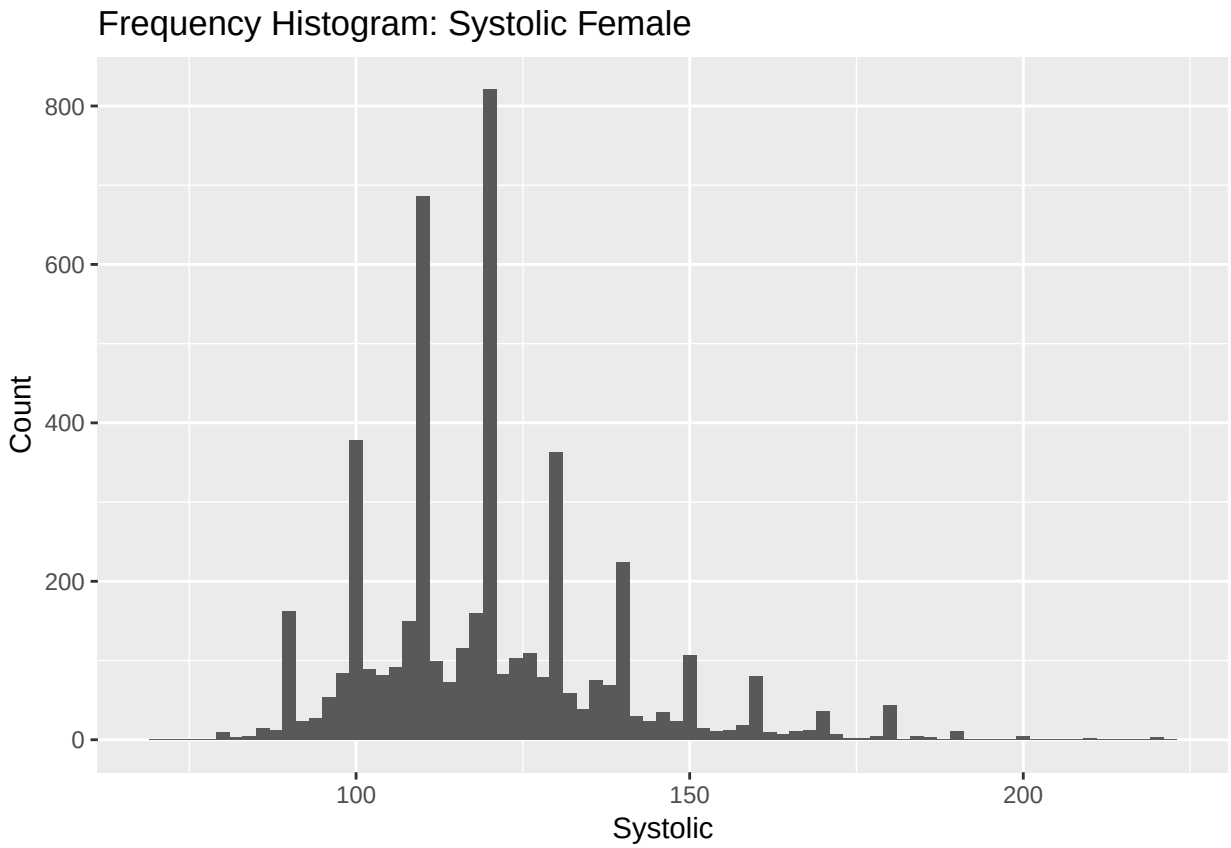
```
load("~/Documents/Math185/dataset-chns-2006-subset3.Rdata")
```

Part A

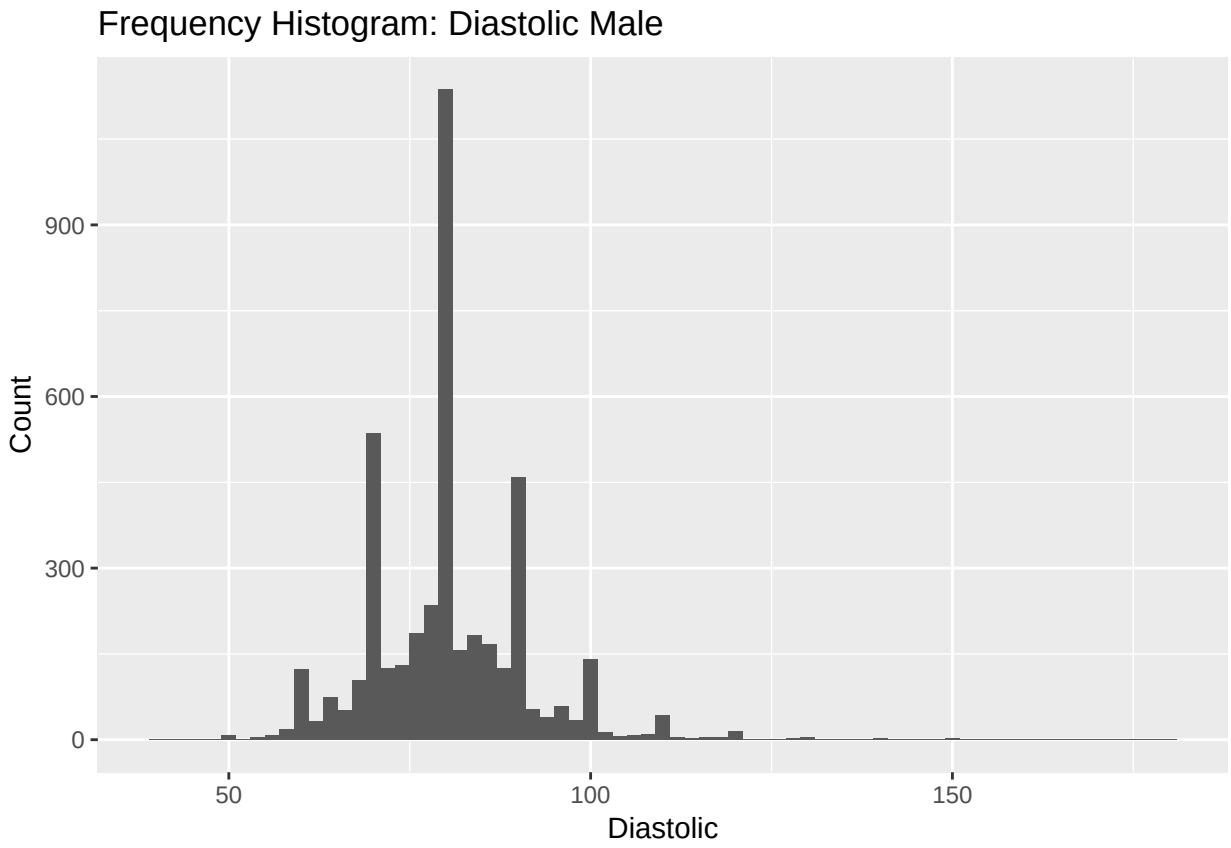
```
library("ggplot2")
male = data[1:4314,]
female = data [4315:9178,]

qplot(male$systolic, xlab = 'Systolic', ylab = 'Count', binwidth = 2,
      main='Frequency Histogram: Systolic Male')
```



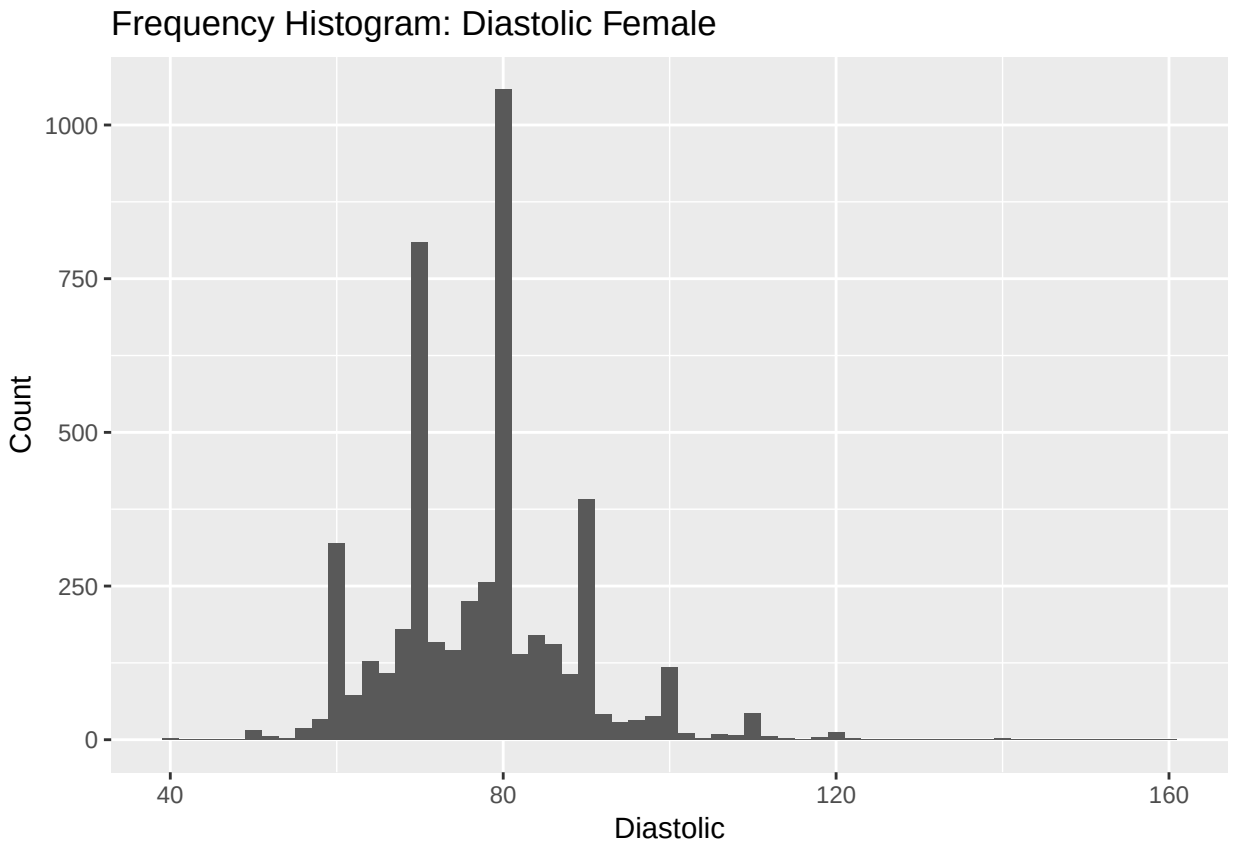

```
qplot(female$systolic, xlab = 'Systolic', ylab = 'Count', binwidth = 2,
      main='Frequency Histogram: Systolic Female')
```



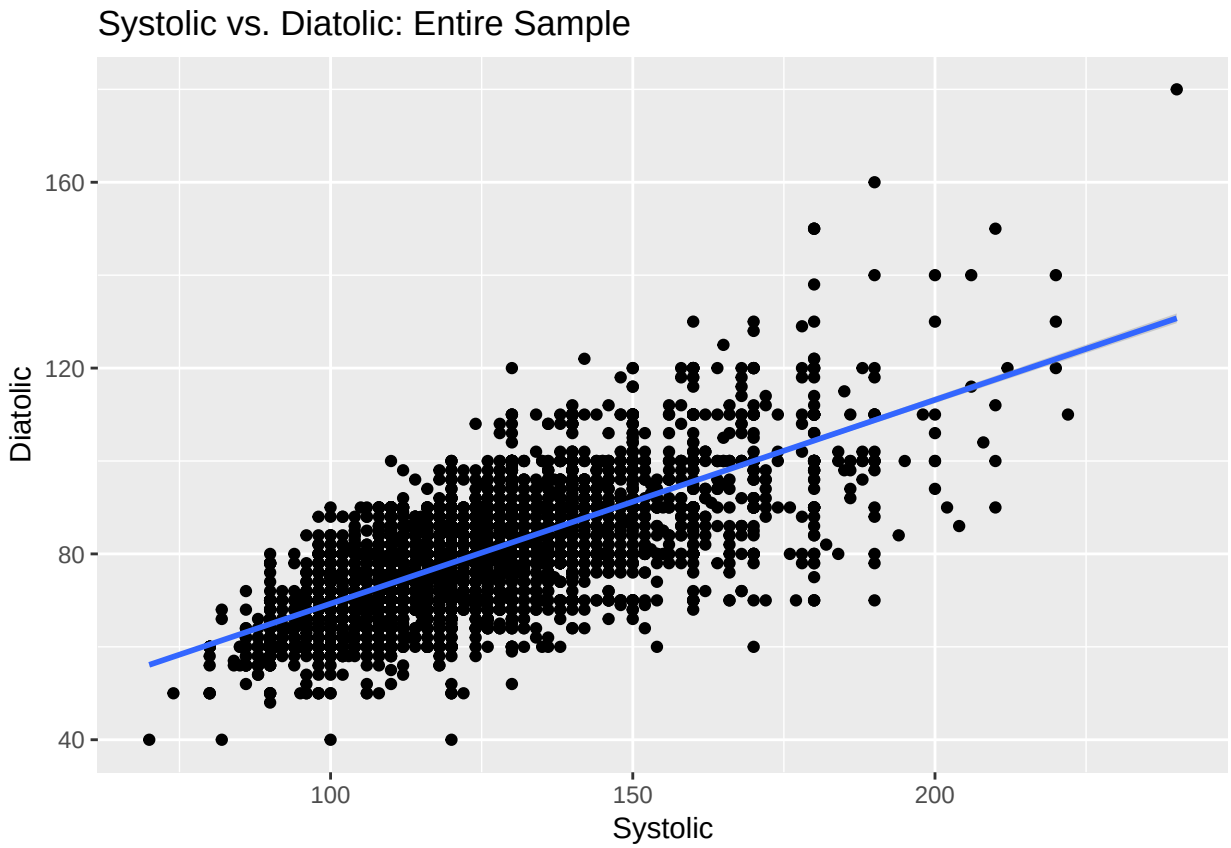

```
qplot(male$diastolic, xlab = 'Diastolic', ylab = 'Count', binwidth = 2,
      main='Frequency Histogram: Diastolic Male')
```




```
qplot(female$diastolic, xlab = 'Diastolic', ylab = 'Count', binwidth = 2,
      main='Frequency Histogram: Diastolic Female')
```




```
ggplot(data = data, aes(x = systolic, y = diastolic)) +
  geom_point() +
  geom_smooth(method='lm') +
  xlab('Systolic') +
  ylab('Diatolic') +
  ggtitle('Systolic vs. Diatolic: Entire Sample')
```



Part B

```
#do the paired test
t.test(data$systolic, data$diastolic, alternative = "two.sided", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$systolic and data$diastolic
## t = 193.35, df = 15221, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  42.50951 43.38022
## sample estimates:
## mean of x mean of y
## 121.79375  78.84888
```

Part C

```
#product 90% CI
t.test(data$systolic, data$diastolic, alternative = "two.sided", var.equal = FALSE, conf.level = 0.90)
```

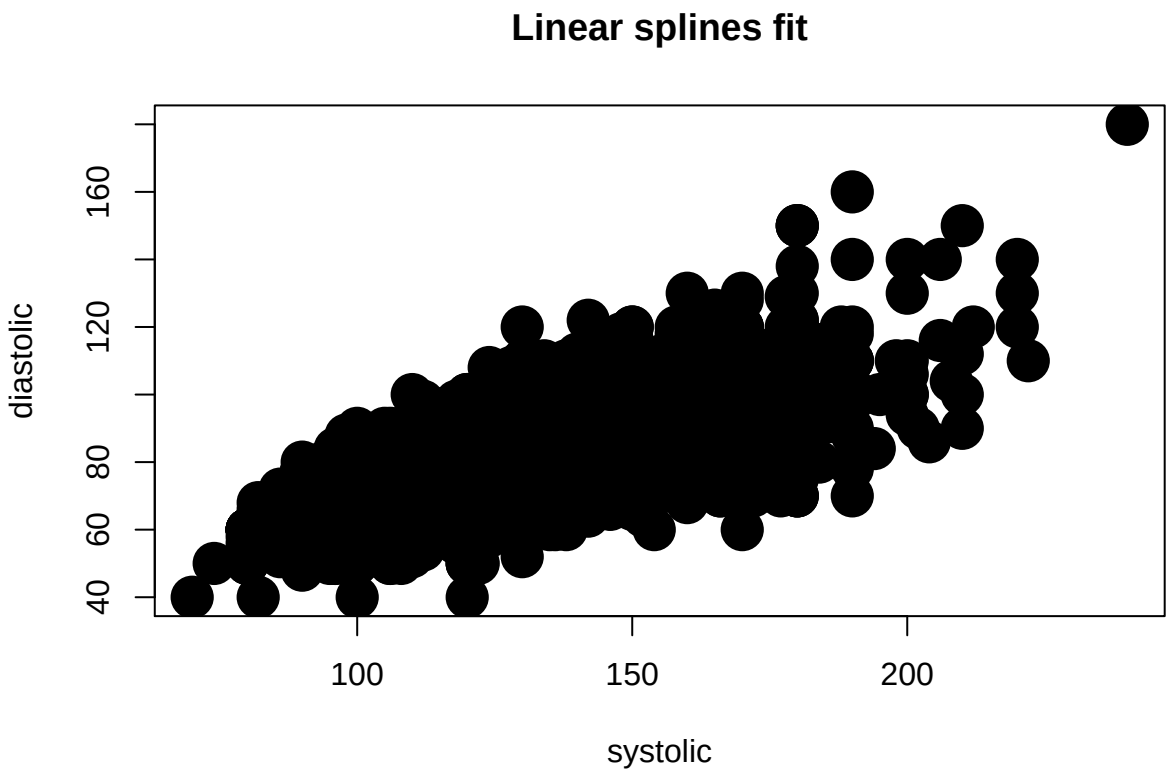
```
##
##  Welch Two Sample t-test
##
## data:  data$systolic and data$diastolic
## t = 193.35, df = 15221, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
```

```
##  42.57951 43.31022
## sample estimates:
## mean of x mean of y
## 121.79375  78.84888
```

```
attach(data)
require(splines)
```

```
## Loading required package: splines
```

```
plot(systolic, diastolic, pch = 16, main="Linear splines fit", cex=3)
```




```
fit = lm(systolic ~ bs(diastolic, degree=2))
pts = seq(0, 600, len=100)
val = predict(fit, data.frame(disp = pts))
```

```
## Warning: 'newdata' had 100 rows but variables found have 9178 rows
```

```
#lines(pts, val, col="red", lwd = 4)
```

Part D

```
attach(data)
```

```
## The following objects are masked from data (pos = 4):
##
##     age, diastolic, gender, systolic, weight
```

```
fit = lm(diastolic ~ systolic + I(systolic^2))
summary(fit)
```

```
##
## Call:
## lm(formula = diastolic ~ systolic + I(systolic^2))
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.513  -3.632   0.607   4.368  64.859
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.5314237  2.3852993   1.061    0.289
## systolic       0.7962714  0.0365551  21.783   <2e-16 ***
## I(systolic^2) -0.0013628  0.0001383  -9.851   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.725 on 9175 degrees of freedom
## Multiple R-squared:  0.5175, Adjusted R-squared:  0.5174
## F-statistic:  4921 on 2 and 9175 DF,  p-value: < 2.2e-16
```