Exam_2

Jonathan De Los Santos

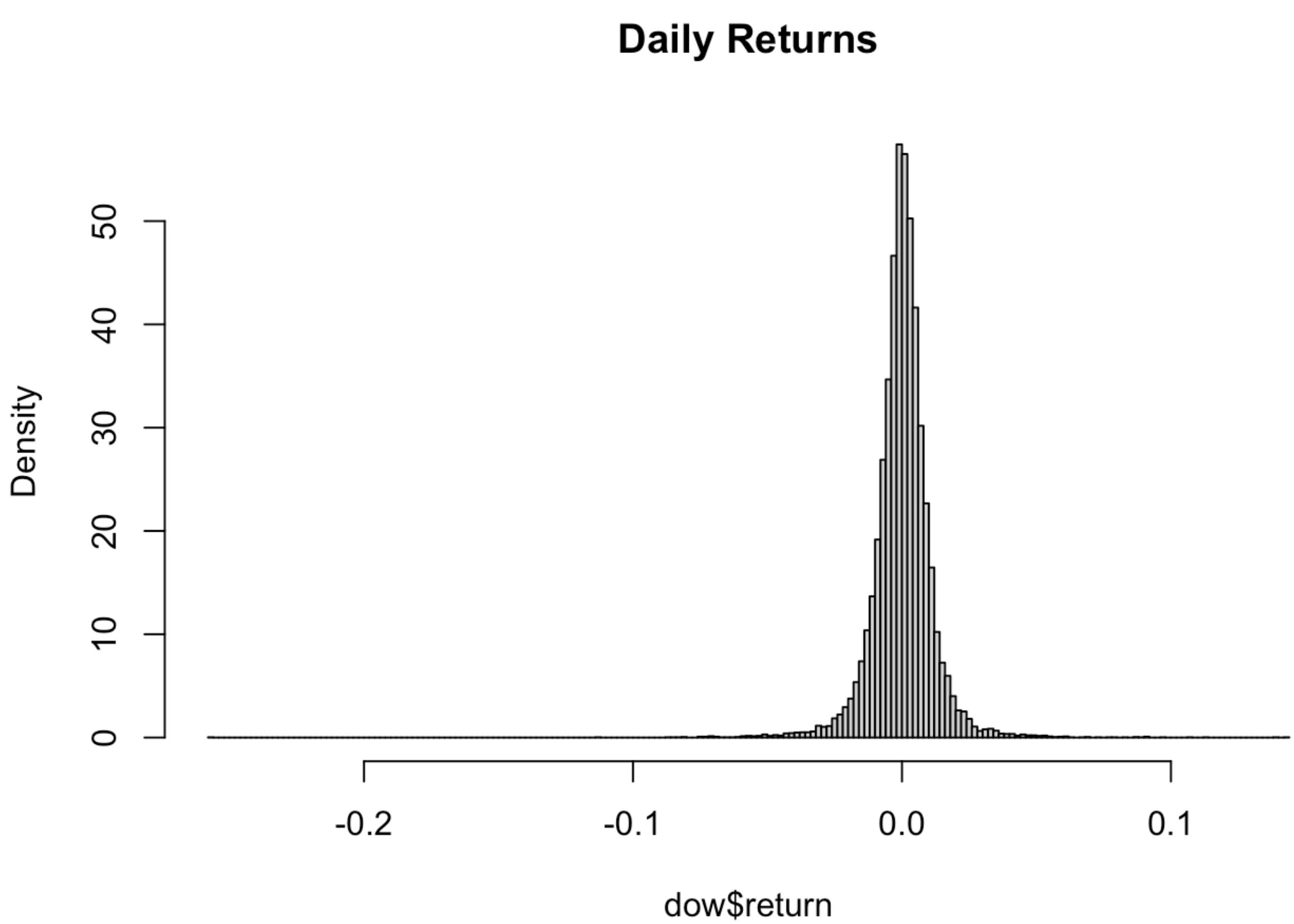
Part A

a) Create a histogram and q-q plot for daily return variable; dow$return and state what is the distribution of daily returns?

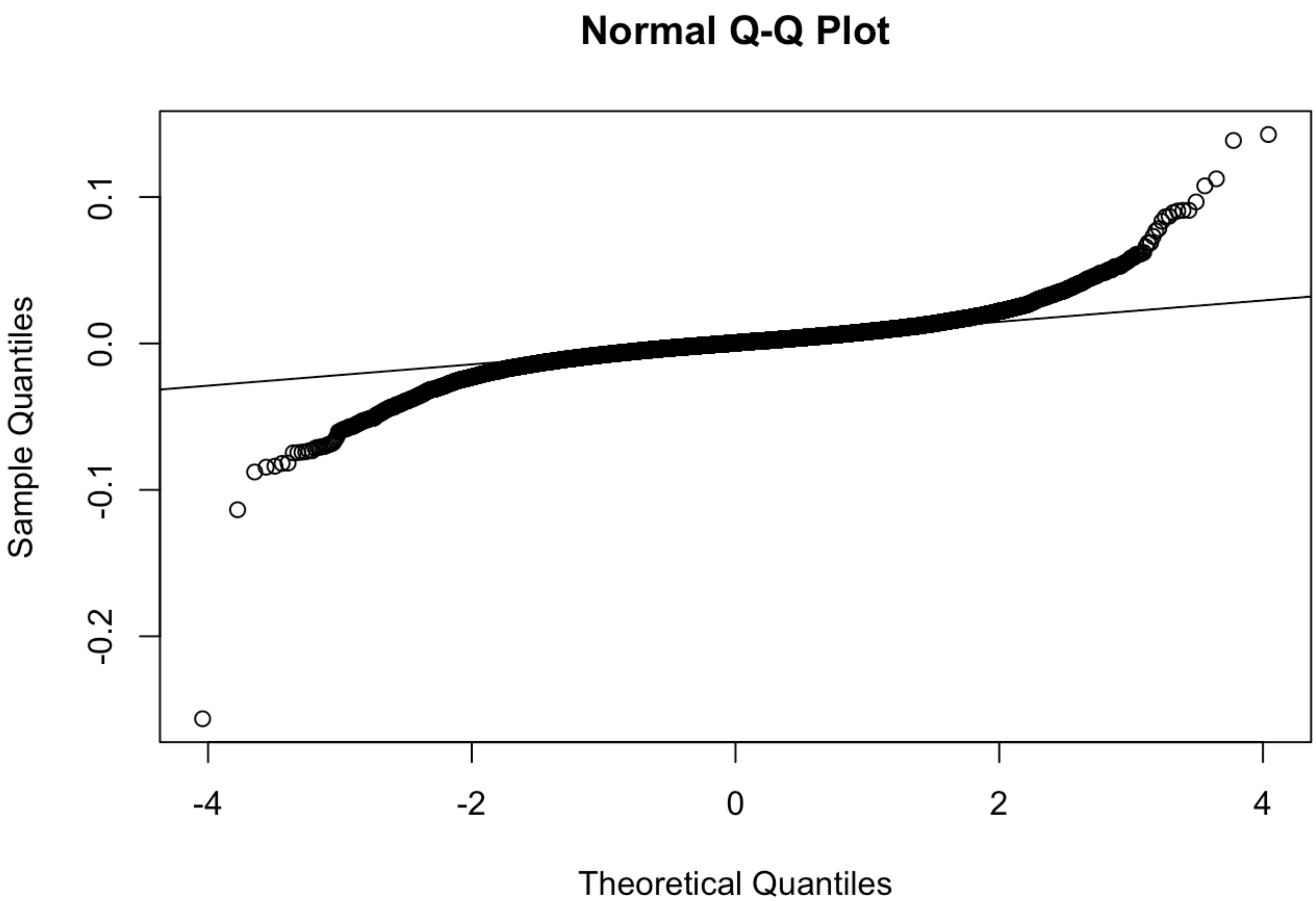
The long tails make this difficult to read, but this distribution does appear to skew left. Given the dataset, it wouldn't be surprising that recessions/depressions cause more outliers to the left than booms due to the right.

```
dow = read.csv("http://tiny.cc/djia")

# Tried different break options for more visibility
hist(dow$return, breaks = 'scott', freq = FALSE, main = paste('Daily Returns'))
```



```
qqnorm(dow$return)
qqline(dow$return)
```



Part B

b) Find a 95% confidence interval for the daily return. What does this show?

If my calculations are correct, the 95% confidence interval is between 4.04 x 10^{-5} and 3.6 x 10^{-4} which indicates that with 95% confidence a random day's return will average between .0000404 and .00036. This is an extremely narrow window likely due to the high sample size.

My result seems a little off because the histogram indicates values out to well past 0 and .01, but it could be that the density of values with that range is so high that it still applies.

```
# Starting with t.test to compare manual results
# Also gives me df without doing a count
dow = read.csv("http://tiny.cc/djia")

# Having issues when I use "dow" so I will proceed calling the column from the data
t.test(dow$return, alternative = 'two.sided' ,conf.level = .95)
```

```
##
##  One Sample t-test
##
## data:  dow$return
## t = 2.4558, df = 18833, p-value = 0.01407
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  4.041003e-05 3.599686e-04
## sample estimates:
##    mean of x
## 0.0002001893
```

```
n = length(dow$return)
# Manually calculate df to be sure
df = length(dow$return) - 1
alpha = .05
sd = sd(dow$return)
Xbar = mean(dow$return)

# Get upper and lower bounds with t-score and standard error
lb = Xbar - qt(alpha/2, df) * sd/sqrt(n)
ub = Xbar + qt(alpha/2, df) * sd/sqrt(n)
c(lb, ub)
```

```
## [1] 3.599686e-04 4.041003e-05
```

Part C

c) Perform a hypothesis test at alpha = 0.01 assuming that H0: Mean Daily Return >= 0 Ha: Mean Daily return < 0

The resulting T-stat is not less than the critical value, therefore we fail to reject the null hypothesis that mean daily return >= 0.

```
# Null hypothesis
mu0 = 0

n = length(dow$return)

df = length(dow$return) - 1
alpha = .01
sd = sd(dow$return)
xbar = mean(dow$return)

# Null distribution parameters
muXbar = mu0
sigmaXbar = sd/sqrt(n)

# Find test statistic
Tstat = (xbar - mu0)/(sd/sqrt(n))
print(paste('Tstat: ', Tstat ))
```

```
## [1] "Tstat:  2.45581945104101"
```

```
# Check significance
# Use < because of alt hypothesis, looking at left tail
print(paste('Critical value: ', qt(1-alpha, df) ))
```

```
## [1] "Critical value:  2.326545897615"
```

```
print(paste('Is Tstat < critical value? ', Tstat < qt(1-alpha, df) ))
```

```
## [1] "Is Tstat < critical value?  FALSE"
```

```
print(paste('Probability that T < Tstat: ',  pt(Tstat, df) ))
```

```
## [1] "Probability that T < Tstat:  0.992967329026086"
```