HW 2 Key

Name: Will Schievelbein

Q1: The total market return will have a beta of exactly one. This is because the market is the benchmark of comparison.

Q2A: The null hypothesis is that the gains or losses in the overall market has no correlation with the stock's gains or losses. The alternative hypothesis is that the gains or losses of the market are correlated with the stock.

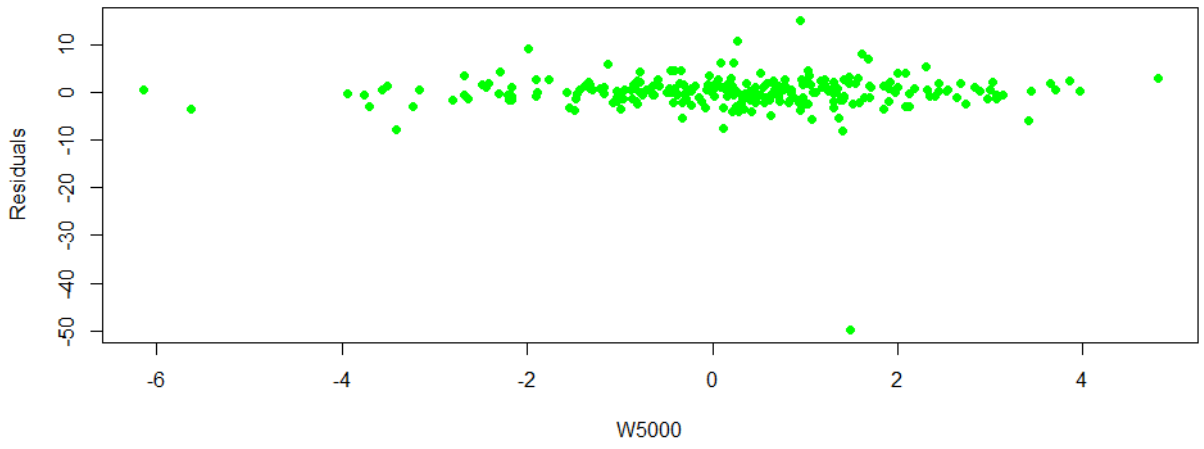
$$Null: \beta = 0; \text{Alternative: } \beta \neq 0$$

Q2B: Starbucks beta is equal to .70. Bank of America's beta is equal to 1.4. Proctor & Gamble's beta is equal to .45.

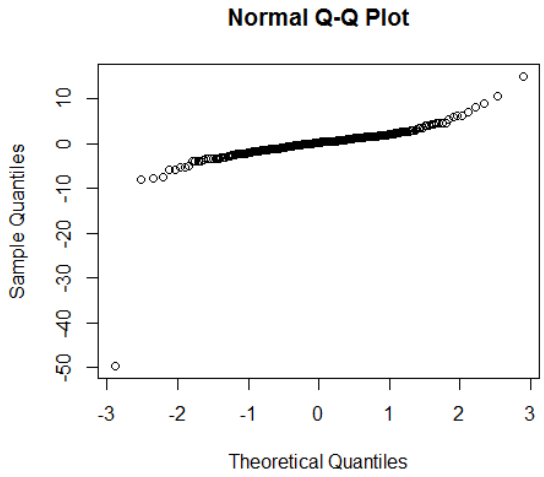
```
dat <- read.csv('stock_market_returns.csv', header=TRUE)
Smod <- lm(SBUX ~ W5000, dat)
Bmod <- lm(BAC ~ W5000, dat)
Pmod <- lm(PG ~ W5000, dat)
summary(Smod)
summary(Bmod)
summary(pmod)
```

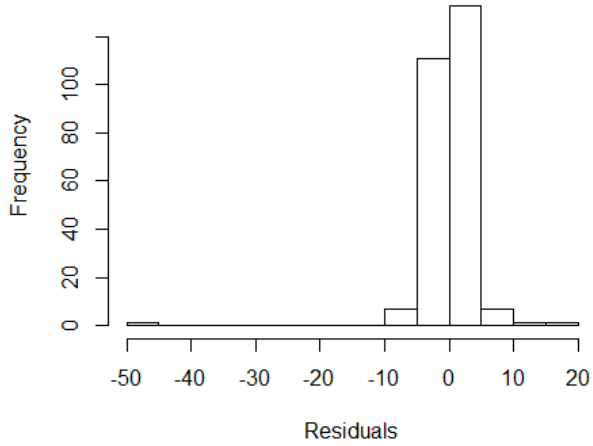
Q2C: Assumptions for linear regression include error independence. This is not as graphically testable and could be violated as this is a time series. Violations in error independence does not induce bias into the coefficient estimation, but could disrupt the estimate of the estimator variance.

The residuals are plotted below. The residuals are approximately linear and no signs of heteroscedasticity are present.



Normality of the residuals are demonstrated with either a Q-Q plot or a histogram of the residuals. The histogram is difficult to read due to the extreme outlier. The extreme outlier in this case is ok, as it does not have much leverage, though we have not taught that yet in the class.




Q2D: The riskiest stock according to Beta is Bank of America. The least risky stock according to Beta is Proctor & Gamble.

Q2E: All three regressions were found to be statistically significant as the p-values were very small (<.05). A difference in the p-values is not significant in interpretation, just that the fit is not only due to chance.

Q2F: A low $R^2$ demonstrates that the beta is unreliable and not practically significant no matter its value. Starbucks has the lowest $R^2$ so the beta regression is the least useful.

Q3A: The correlation between BOA and the market is .604. This is found by looking at the regression R^2 (not adjusted) and taking the square root, or using the cor() function.

Q3B: The prediction for BOA's return would be 2.37% if the market increased 1.5%
1.5*1.442+.2105

Q3C: The p value is 2e-16 for both. The model is composed of a single predictor, thus it is the same either way.

Q3D: The 95 % confidence interval for Bank of America's beta is bounded by 1.21 and 1.67.

```
Bmod <- lm(BAC ~ W5000, dat)
confint(Bmod)
```

Q3E: The confidence interval gives an idea of how much the true value for beta is likely different from the estimate.

Q3F: The y intercept represents the return of Bank of America if the market was to not move in one day. A value of .2105 means that the best prediction of the model for Bank of America's gain on a day the market remains at its previous value would be a gain of .2105%. Note that this is NOT the average gain for the stock as the market average return is likely not zero.

Q3G: The beta for BOA on yahoo is 1.31, and on Reuters it is 1.43. This is very close to the estimate we found (1.40) but likely different due to a different base or by looking over a different time period. Additionally, the confidence interval gives a sense of the area where the true value may be. Other possible reasons are that Yahoo could be normalizing BOA to other similar companies using mixture models and weighted to more recent data.

Q4A: One could measure how we the index tracks by plotting a regression of gains and losses and checking the R^2 which measures the fit.

Q4B: VINIX vs W5000 leads to a model with a r squared of .7609 which is not considered a very good fit for an index fund to the index that it is tracking. The reason for this is that VINIX is designed to track the S&P 500 (large cap index), not the Wilshire 5000 (total market index).

```
setwd("C:\\Users\\wills\\OneDrive\\Documents\\Regression\\h2")

dat <- read.csv('stock_market_returns.csv', header=TRUE)
#For some reason I loaded a null row at the end. It is deleted in the code below
dat <- dat[1:261,]

#Checking regression for SBUX, BAC and PG
Smod <- lm(SBUX ~ W5000, dat)
par(mfrow=c(2,2))
plot(Smod)
dev.off()
Bmod <- lm(BAC ~ W5000, dat)
par(mfrow=c(2,2))
plot(Bmod)
dev.off()
Pmod <- lm(PG ~ W5000, dat)
par(mfrow=c(2,2))
plot(Pmod)
dev.off()

#Betas of stocks
summary(Smod)
summary(Bmod)
summary(Pmod)

#test linearity assumption
plot(dat$W5000, resid(Smod), pch=16, col='green', xlab='W5000', ylab='Residuals')
#Test normality of residuals
hist(resid(Smod), xlab='Residuals', main='')
qqnorm(resid(Smod))
```