

Analysis to Predict Closing Stock Prices from Company AT&T

Victor Yung

Benson Ou-yang

Ivan Cao

Abstract

In our project, our goal was to develop a suitable prediction model by assessing the closing stock prices of the company AT&T using linear regression analysis. To determine the regressors we used stepwise analysis, which discovered various sets of variables that resulted in having a significant relationship to the closing price of AT&T stocks which includes volume, capital surplus, gross margin, and liabilities. By fitting each regressor into the model, we were able to produce a model that explained roughly 66.4% of variability of predicting the closing price. Furthermore, we used visual graphs such as normal q-q plots and Residual plots to help identify any underlying issues with patterns or outliers we came across in our model. The overall analysis of our model helped us subset and distinguish the effectiveness of each regressor we found to our response variable.

Introduction

The telecommunication space is an inevitably growing industry dependent on the advancement of technology. Therefore, there will be stock investment opportunities for consumers to take part in as we notice that the amount of potential this industry presents, however it is not risk free. In our analysis, we designate our efforts toward one of the most well-known telecommunication companies around the world, AT&T. Our question of interest is to predict the closing stock price of the years 2012 to 2016 for AT&T. In order to tackle this problem, we first collected and constructed the data set with categories related to closing stock prices and tested to see if there is a strong significance to those specific years. We then performed a stepwise procedure to ensure the categories we chose had a strong significance to the closing price. Therefore, we introduced those variables as the regressor we will use for our final model for prediction. The variables include volume measuring the number of shares traded during a specific time, capital surplus which is the excess remaining after common stock sold, gross margin as the percentage of the difference between revenue and cost of goods sold divided by revenue, and lastly liabilities being how much a company owes. The model will be further dissected through visual plots that will explain the different patterns and possible outliers that may affect the results of our final model.

Data Description

We are using three datasets: securities.csv, fundamentals.csv, and prices-split-adjusted.csv.

Securities.csv contains information on the stock companies such as the company's name and ticker symbol, the type of sector they are in, location of headquarters and others.

Fundamentals.csv contains information of yearly reports of fundamental information of each company such as total revenue, accounts payable, liabilities and many more.

Prices-split-adjusted.csv contains information of the stocks adjusted prices after splitting. The columns included are the date, ticker symbol, close, open, low, high, and volume.

For our linear regression model, we have selected to predict close prices of AT&T's stock. Our regressor variables are volume from the prices-split-adjusted.csv and capital surplus, gross margin, and liabilities from the fundamentals.csv.

$$close = \beta_0 + \beta_1(volume) + \beta_2(CapitalSurplus) + \beta_3(GrossMargin) + \beta_4(Liabilities) + \epsilon$$

Since the data from fundamentals.csv is yearly, we applied the previous year data into the next year since the yearly reports are at the end of the year so we use that information for the next year. For example , the total revenue for 2013 is \$1,000,000 so we made a column for total revenue and made every row that is in 2014 to be \$1,000,000. We ran a nested for loop to apply this for all years and columns. Tprices is the dataset with the columns of interest for our linear model.

close	volume	date	year	Capital.Surplus	Gross.Margin	Liabilities
35.00	38323500	2013-01-02	2013	9.1038e+10	57	1.798e+09
35.02	28932700	2013-01-03	2013	9.1038e+10	57	1.798e+09
35.23	21136600	2013-01-04	2013	9.1038e+10	57	1.798e+09
35.39	27500500	2013-01-07	2013	9.1038e+10	57	1.798e+09
34.35	29210300	2013-01-08	2013	9.1038e+10	57	1.798e+09

Close corresponds to the price of the stock when the market closes.

Volume is the number of trades that occurred that day.

Date is the date of the trading day.

Year is the year of the trading day.

Capital Surplus or share premium, most commonly refers to the surplus resulting after common stock is sold for more than its par value.

Gross Margin is a company's net sales revenue minus its cost of goods sold. The higher the gross margin, the more capital a company retains on each dollar of sales, which it can then use to pay other costs or satisfy debt obligations.

Liabilities are the debts and obligations of a company.

Figure @ref(fig:c-plot) represents the market closing price of the stock for AT&T over the years. The red points on the line is just the indicator of the beginning of the year. Over the years, the closing price is around \$35 starting the year in 2013 and 2014. Around spring of 2013, the stock shot up to about \$39 which is the highest closing price until 2016. The stock drops to about \$32 in the beginning of the year and before 2015. In 2016, the stock was starting to rise and in mid 2016, the stock got a new record of about \$43.

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    30.00   33.96   35.10   35.77   36.74   43.47
```

In the five number summary of the close prices, the minimum is \$30 but that is the one data point that we have added. So without that data point, the lowest is \$31.80. The max close price is \$43.47. The mean close price is \$35.77 since the mean wouldn't change that much from one data point that is \$30.

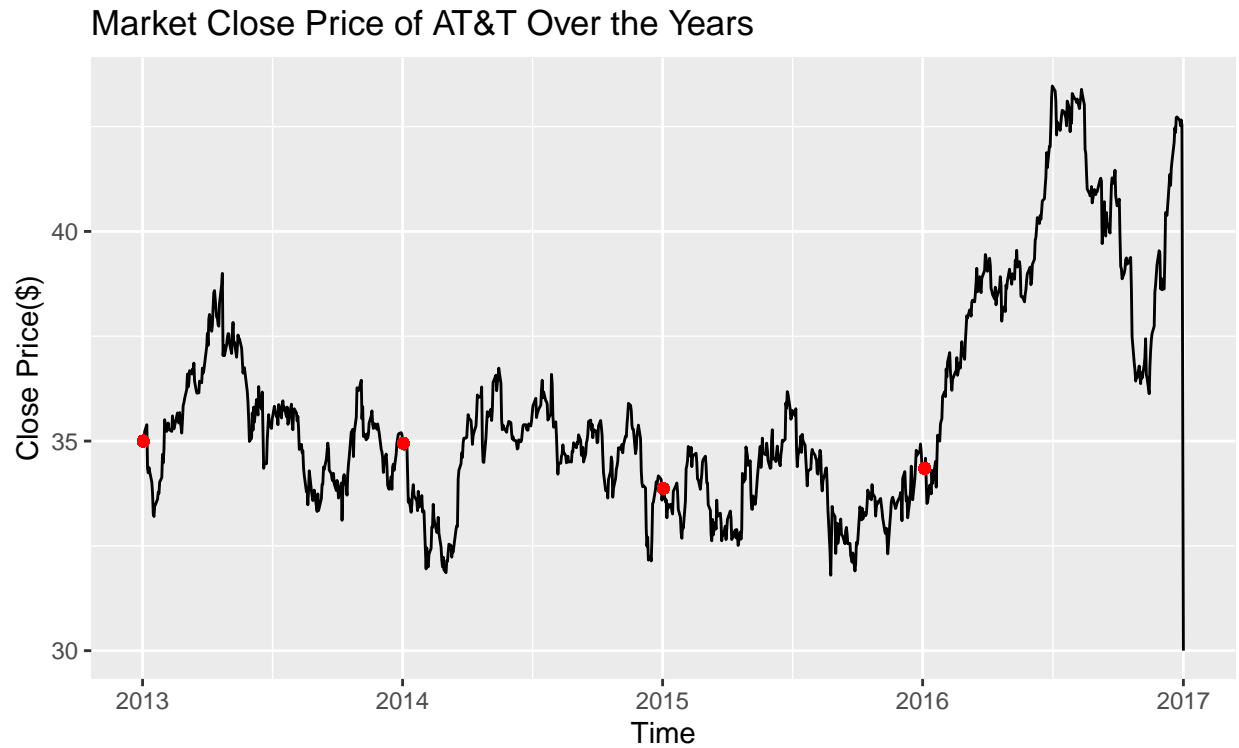
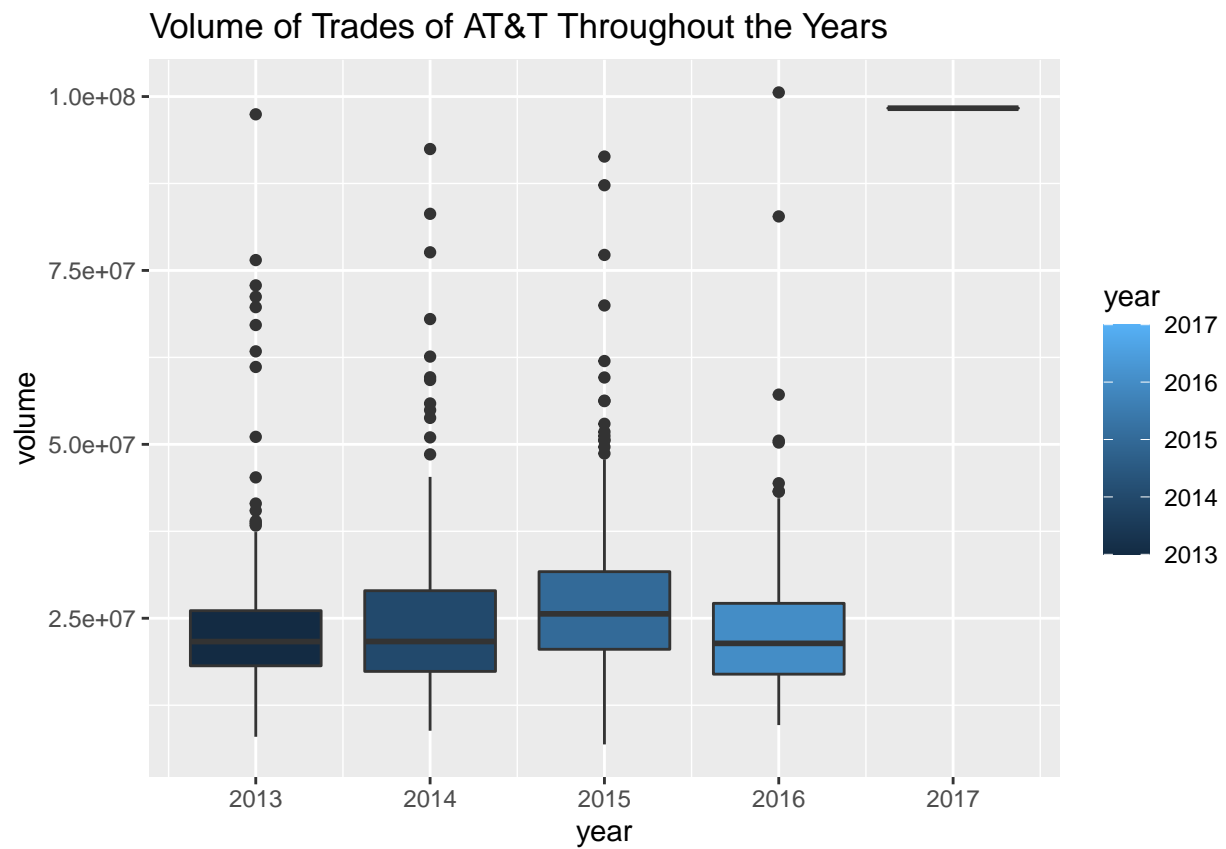


Figure 1: Close Price of ATNT 2013-2016



This boxplot shows the volume traded of each year. The mean volume traded is about the same for every year except for 2015. The dots of the boxplot represent the outliers of each year. There are days where the stock is traded more often than usual such as when the stock is low, more people are buying and when the stock is high, more are selling. The max volume traded was in 2016, we can assume a lot of people were selling when AT&T stock was at its highest in this data.

```
## [1] "Summary of Volume in 2013"
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 7960000 18167000 21646500 23940484 26088200 97444100
```

```
## [1] "Summary of Volume in 2014"
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 8831900 17360725 21657950 24690541 28960075 92453000
```

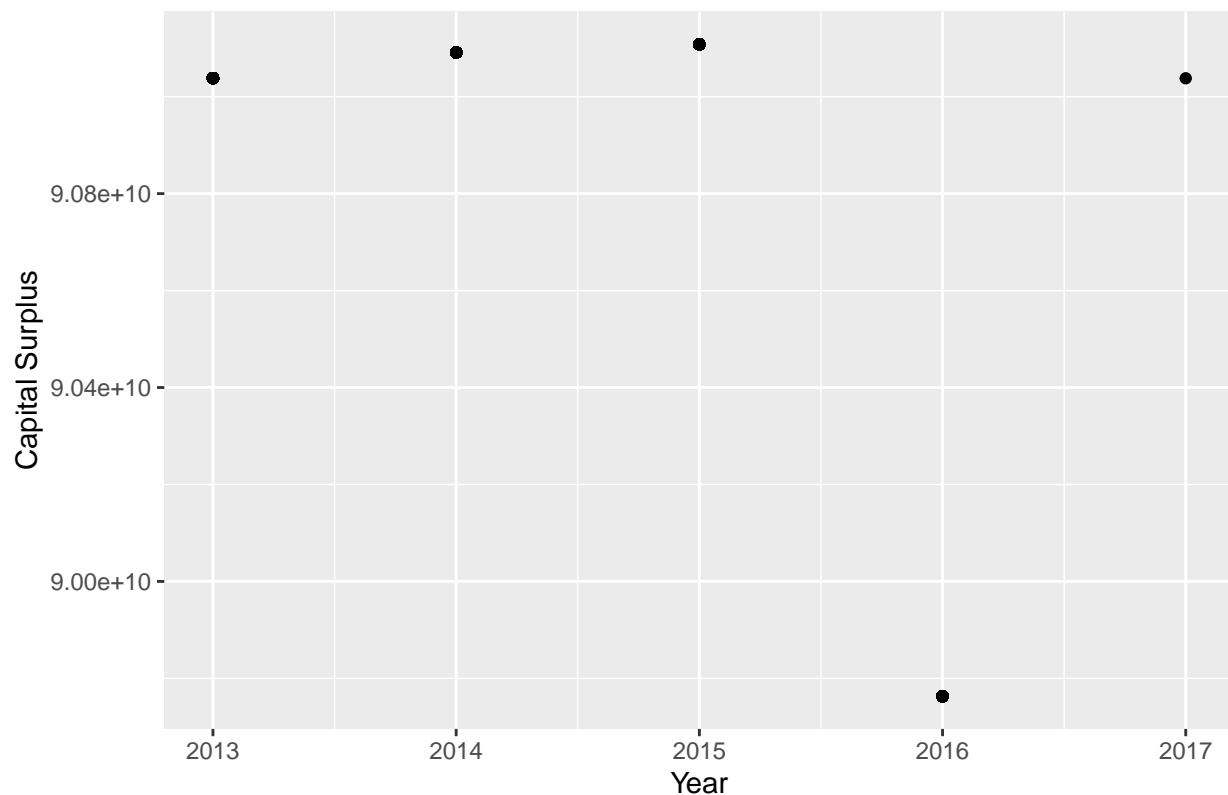
```
## [1] "Summary of Volume in 2015"
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 6862400 20536525 25633950 28282645 31696250 91372900
```

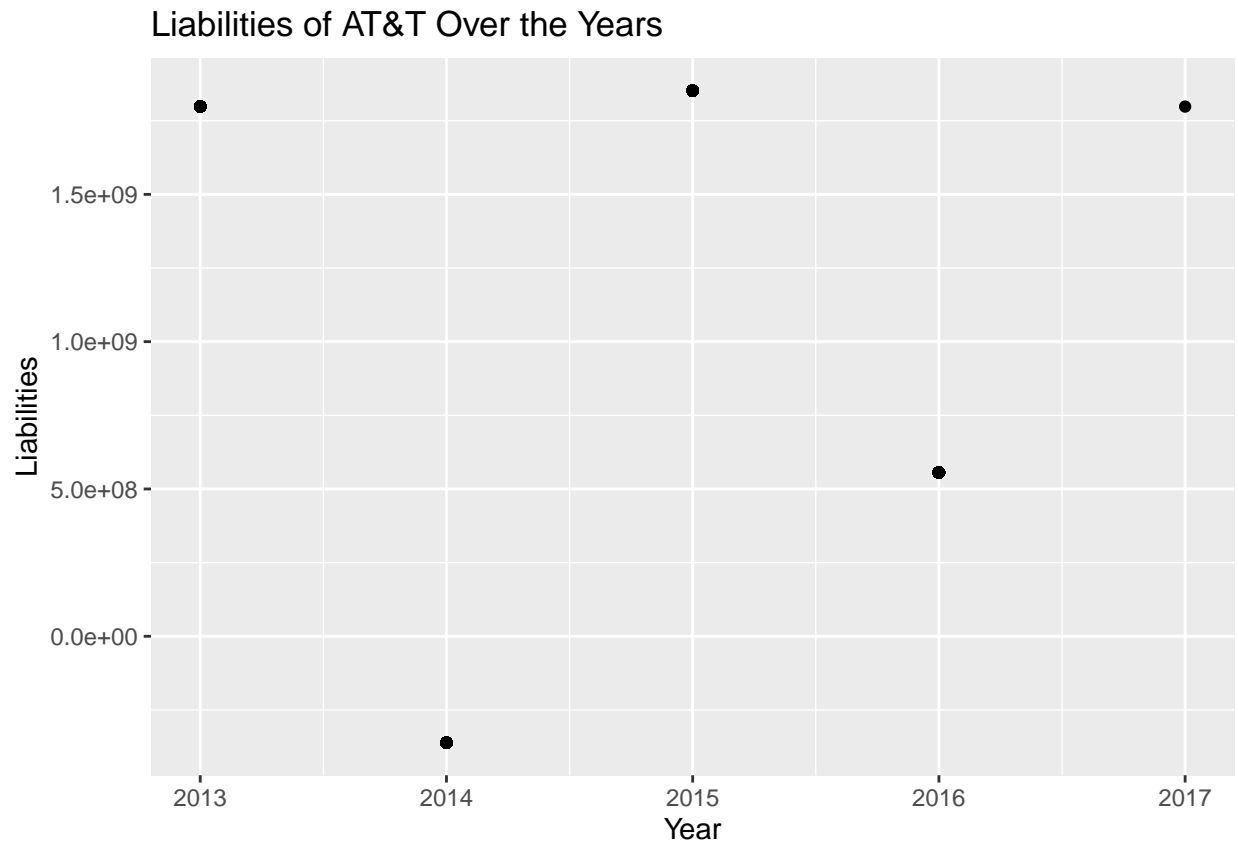
```
## [1] "Summary of Volume in 2016"
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 9645400 16962575 21388750 23596375 27142400 100586200
```

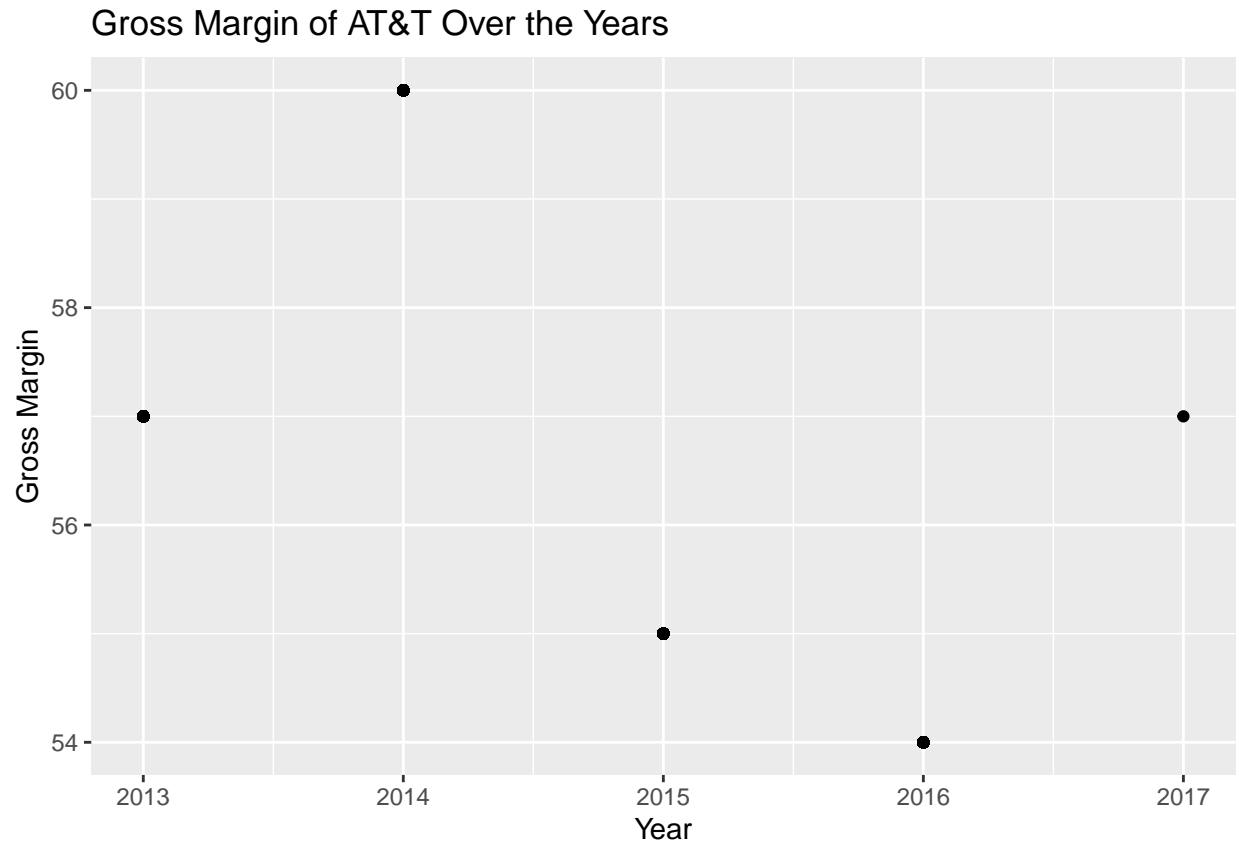
Capital Surplus of AT&T Over the Years



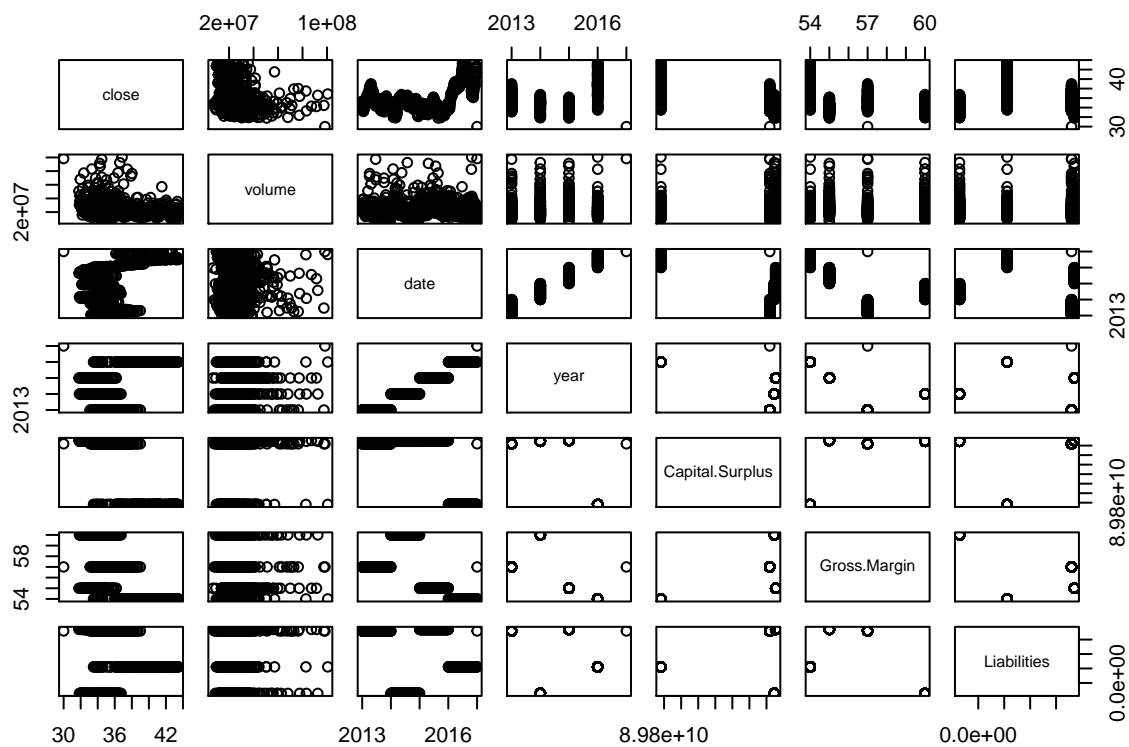
This plot shows the capital surplus of each year. It was rising up until 2016 where it dropped by a lot.



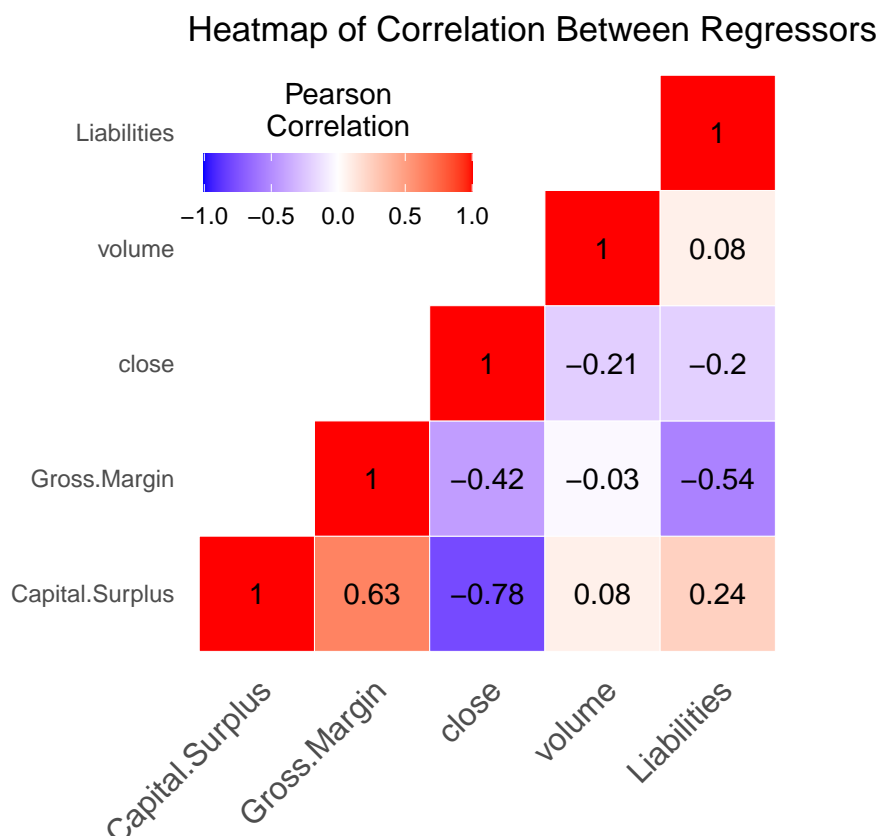
This plot shows the liabilities of each year. In 2013 and 2015, AT&T had the highest liabilities. In 2014, they had the lowest. In 2016, it was between the highest and lowest liabilities.



This plot shows the gross margin of each year. From 2013 to 2014, it rose up to the highest of 60. In 2015 it fell down to 55 and 2016 dropped to 54.



This pairs plot shows the columns plotted against each other. Since the columns from the fundamentals.csv is yearly data, when plotted against other columns, they are shown as separate lines due to the values being the same daily for the year.



The heatmap shows the correlation between each column.

Methods

Datasets

To begin, we chose to combine two of the four datasets offered, “Price-split-adjusted” (PSA) and “Fundamentals” (Fund). The PSA dataset accounts for all stocks traded in the NYSE daily from 2010-2016 and the Fund dataset accounts for the 10-K filing from 2012-2016, an annual comprehensive report required by the U.S. Securities and Exchanges Commission (SEC). Since the dates in the datasets varies from daily in PSA and annually in Fund, we attach the previous year’s filing of the 10-K report to help predict the next year over.

Subsetting Dataset

With many predictors in our dataset, we used stepwise regression to identify which predictors deemed significant and insignificant, with the insignificant predictors being removed from the model. Since our datasets covers different years, we kept only the years that overlap in both datasets, 2013-2016. Lastly, we removed all companies other than our company of focus, AT&T. Figure 4.1 presents data for subsetting data.

close	volume	date	year	Capital.Surplus	Gross.Margin	Liabilities
35.00	38323500	2013-01-02	2013	9.1038e+10	57	1.798e+09
35.02	28932700	2013-01-03	2013	9.1038e+10	57	1.798e+09

close	volume	date	year	Capital.Surplus	Gross.Margin	Liabilities
35.23	21136600	2013-01-04	2013	9.1038e+10	57	1.798e+09
35.39	27500500	2013-01-07	2013	9.1038e+10	57	1.798e+09
34.35	29210300	2013-01-08	2013	9.1038e+10	57	1.798e+09

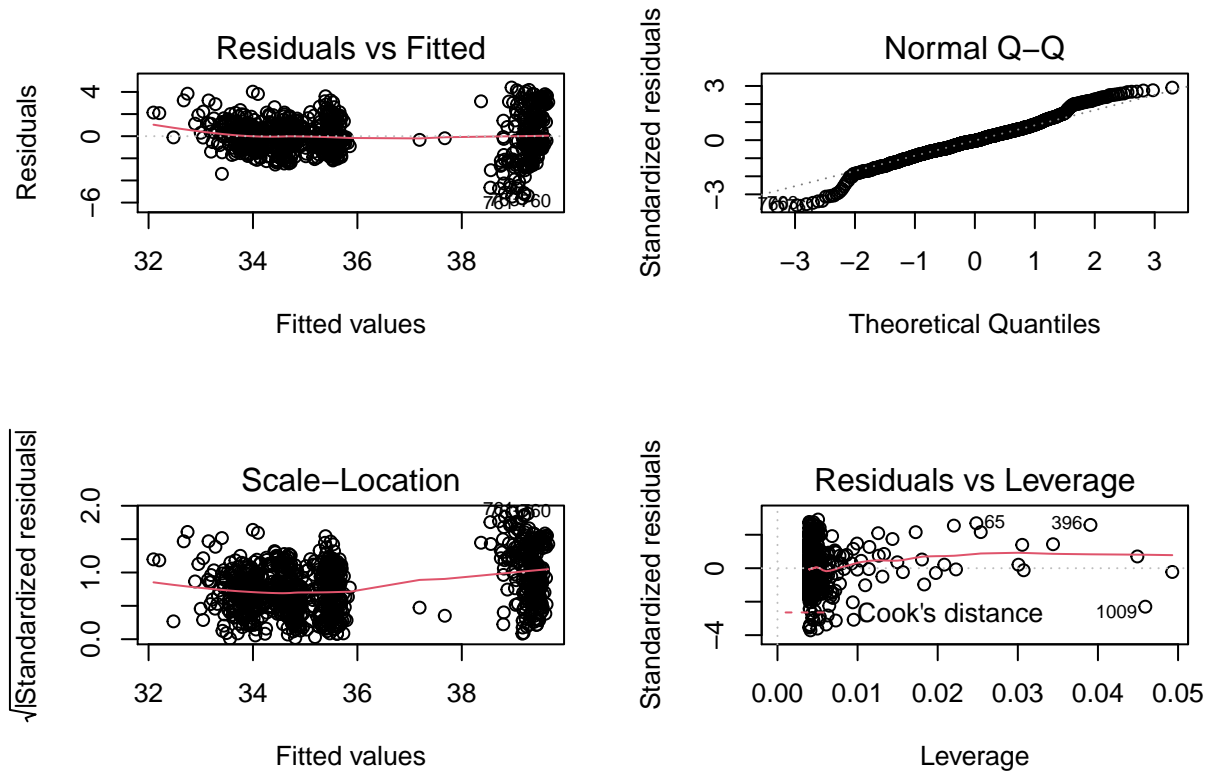
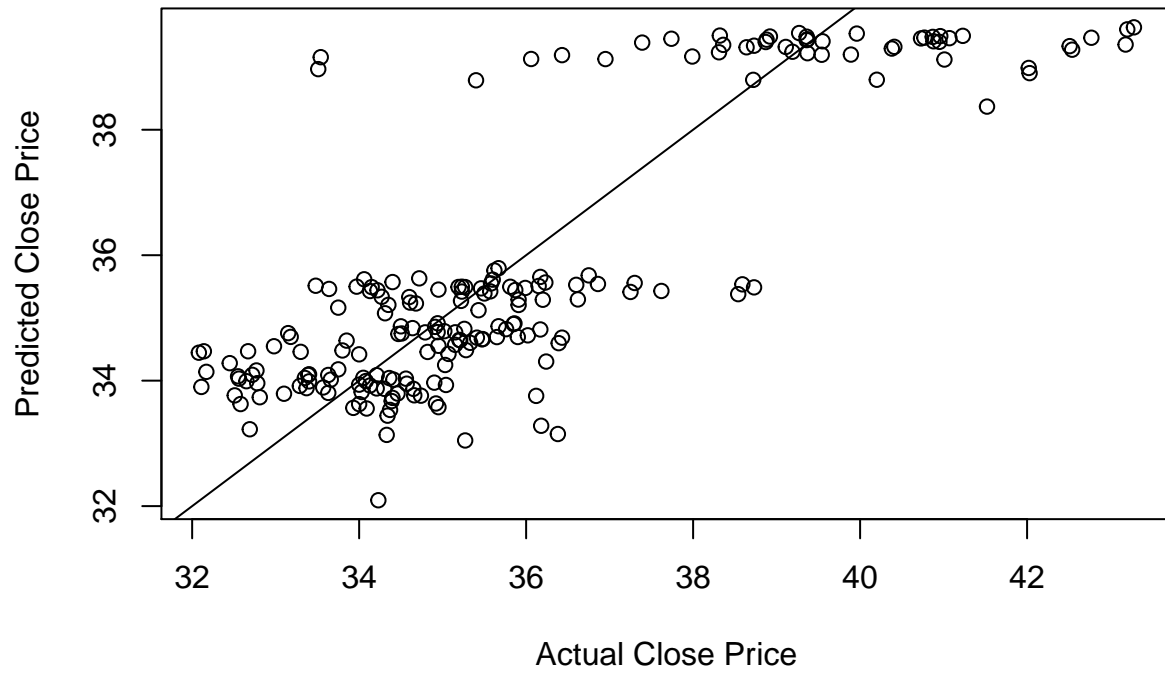


Figure 2: Markdownvellous!

Model Selecting

After establishing a model with stepwise selection, we also tried a handful of transformations to try to improve the accuracy of our model. Using transformations on the response variables did little to improve our model (Figure 4.2) whereas transformations performed on the regressor variables worsened our model (Figure 4.3). In the end, we stuck to our original model, keeping close as our response variables, and volume, capital surplus, gross margin, and liabilities as our predictors (Figure 4.3).

Actual vs Predicted Close Price



AT&T Close Price Over Time

