Google Stock Price Analysis

By Shaik Mujahid

***Variables of Interest:***

The variables of interest in this analysis are the daily stock prices of Google for the year 2012. The dataset consists of the following variables:

Date: The date of the trading day.

Open: The opening price of the stock on that day.

High: The highest price of the stock on that day.

Low: The lowest price of the stock on that day.

Close: The closing price of the stock on that day.

Volume: The total number of shares traded on that day.

***Data Cleaning:***

Before conducting the analysis, some data cleaning was performed to ensure that the data was in the appropriate format for analysis. The following steps were taken:

The column names were renamed to be more descriptive and easier to read.

The unnecessary "Volume" variable was dropped from the data frame.

The date variable was cleaned by converting it to a date format using the "as. Date" function.

The variables "Open," "High," "Low," "Close," and "Volume" was converted to numeric values using the "as. numeric" function.

***Initial Analysis:***

The initial analysis was performed to understand the data and identify any trends or patterns. The following steps were taken:

The structure of the data was checked using the "str" function to ensure that all variables were in the correct format.

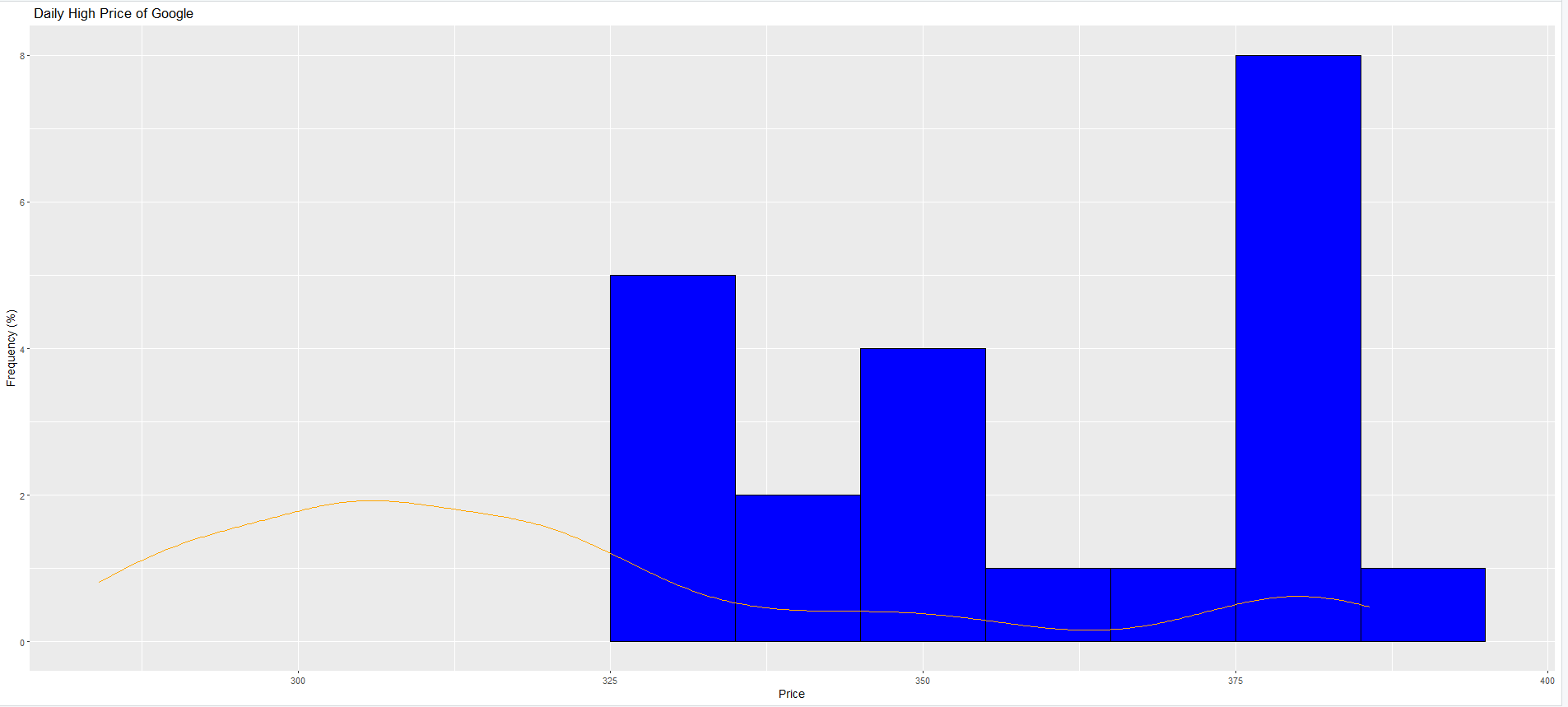
The frequency table of the "Open" variable was created using the "table" function to determine the number of times each value appears in the dataset.

The cross-tabulation of the "Open" and "Close" variables was created using the "table" function to determine the number of times each combination of values appears in the dataset.

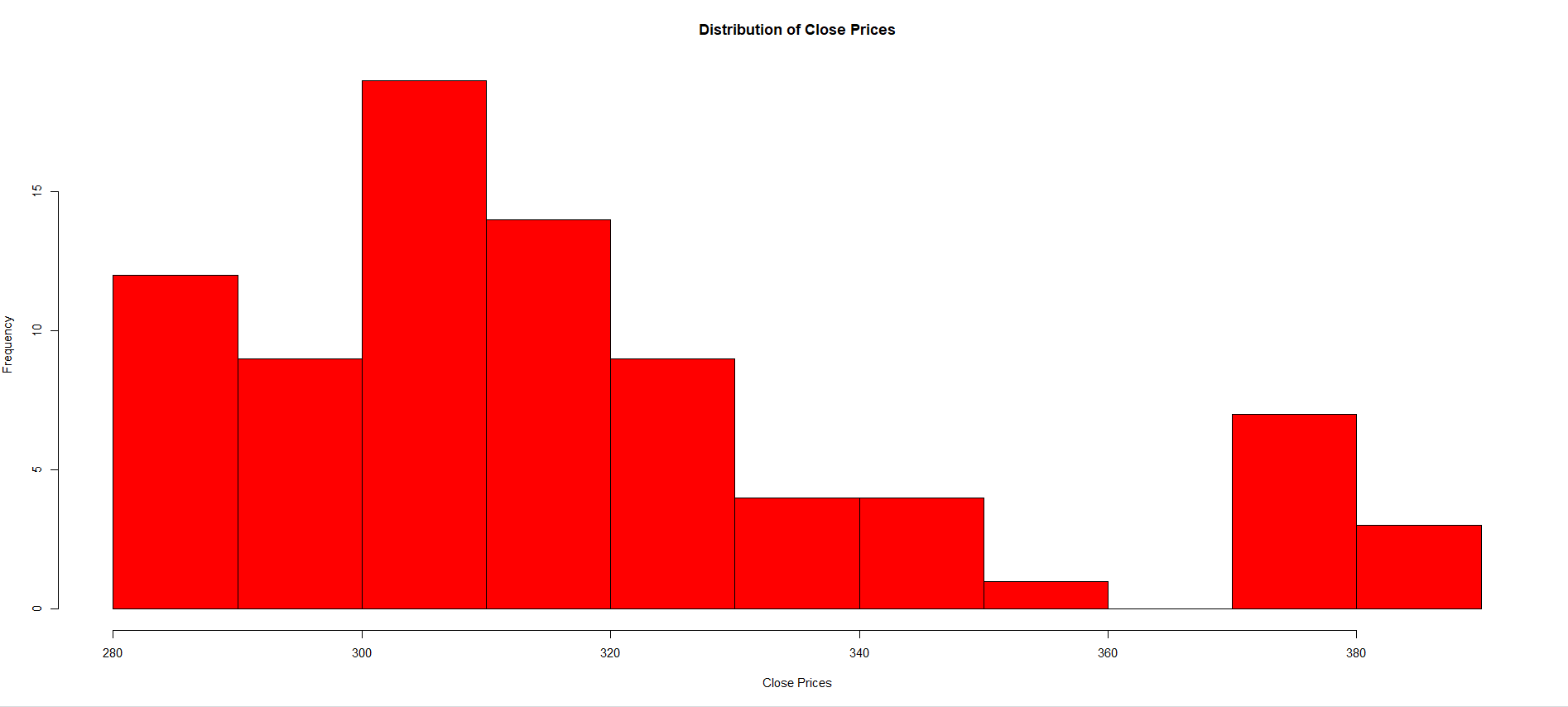
Using the "head" and "tail" methods as well as the "kable" function from the "kableExtra" package, a table containing the first and final five rows of the dataset was produced.



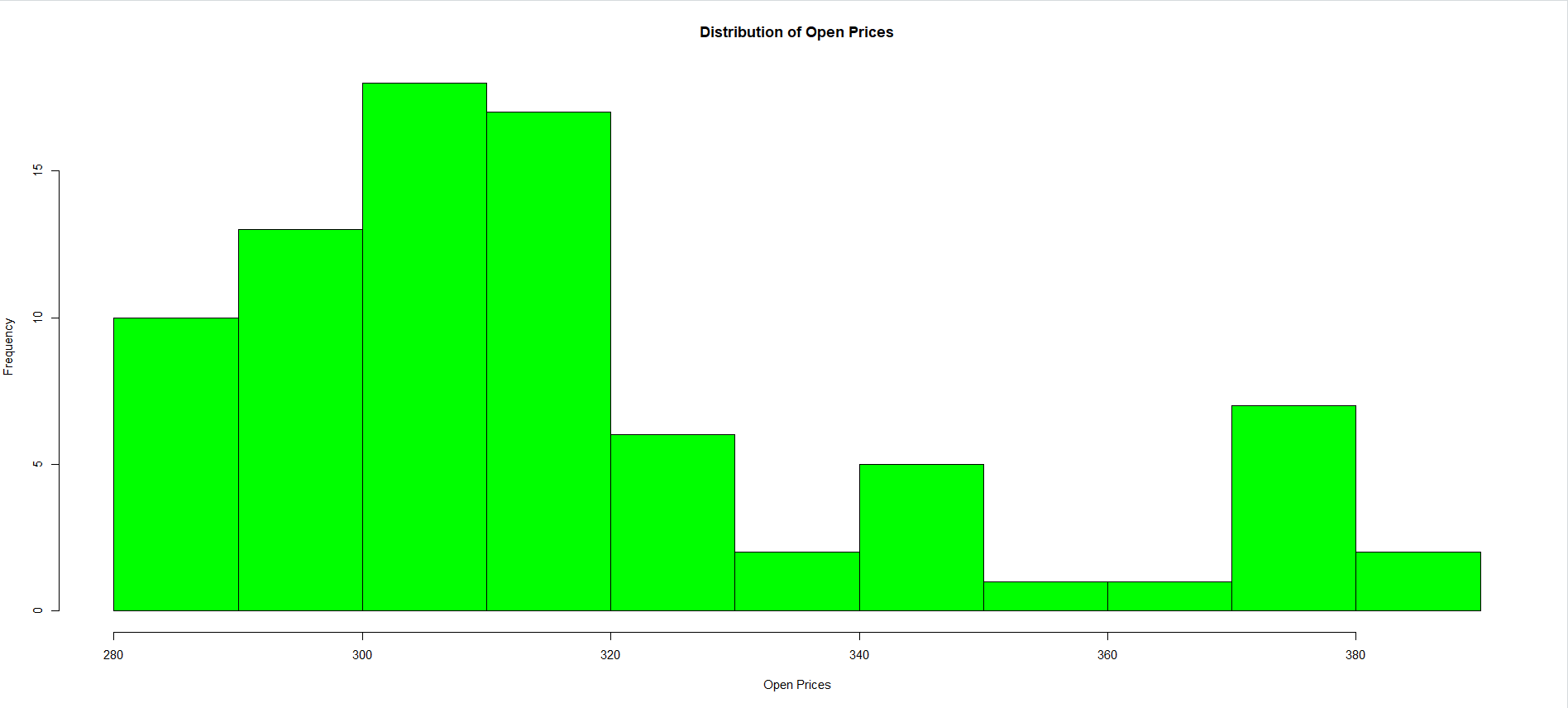
The "ggplot" function was used to produce the density plot of the "High" variable in order to show the distribution of high prices.



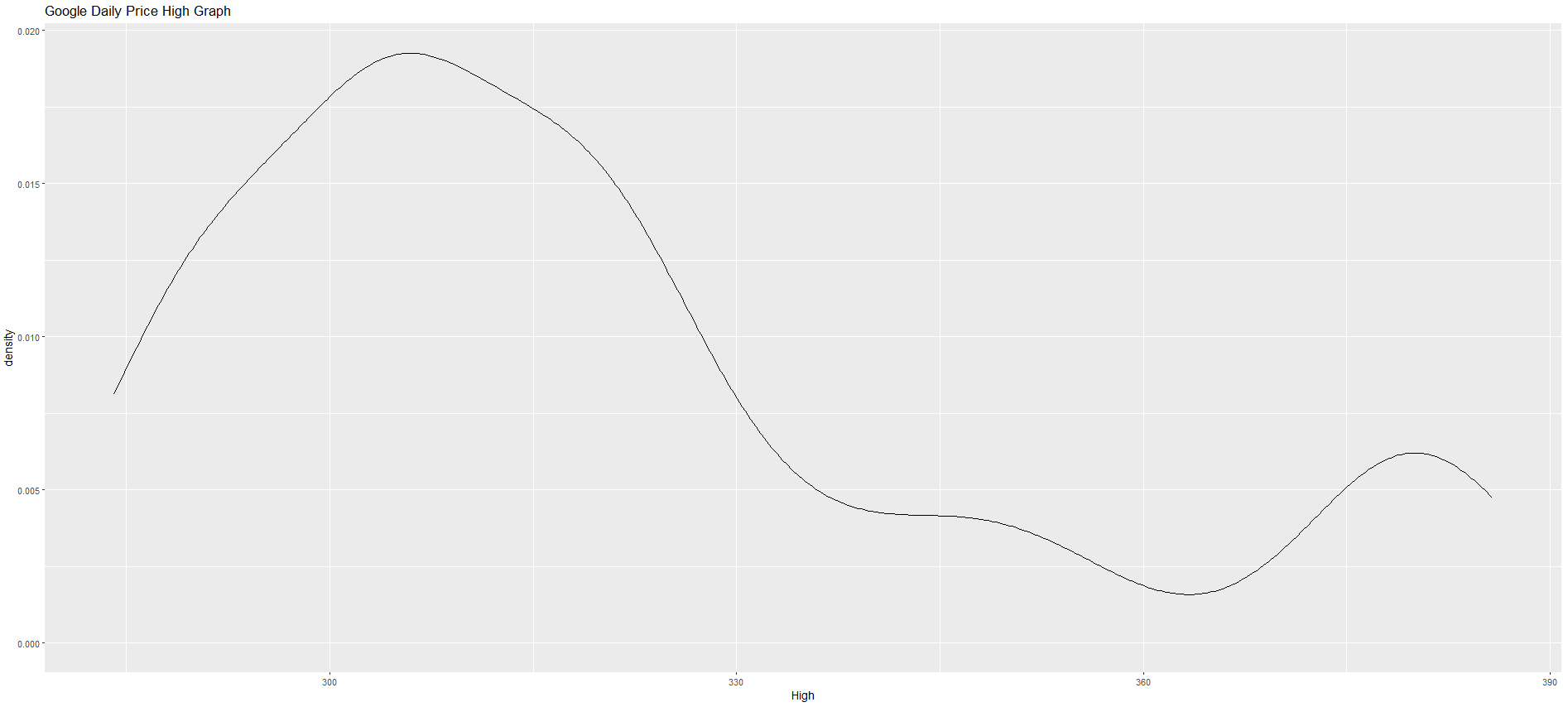
The histogram of the "Close" variable was created using the "hist" function to visualize the distribution of closing prices.

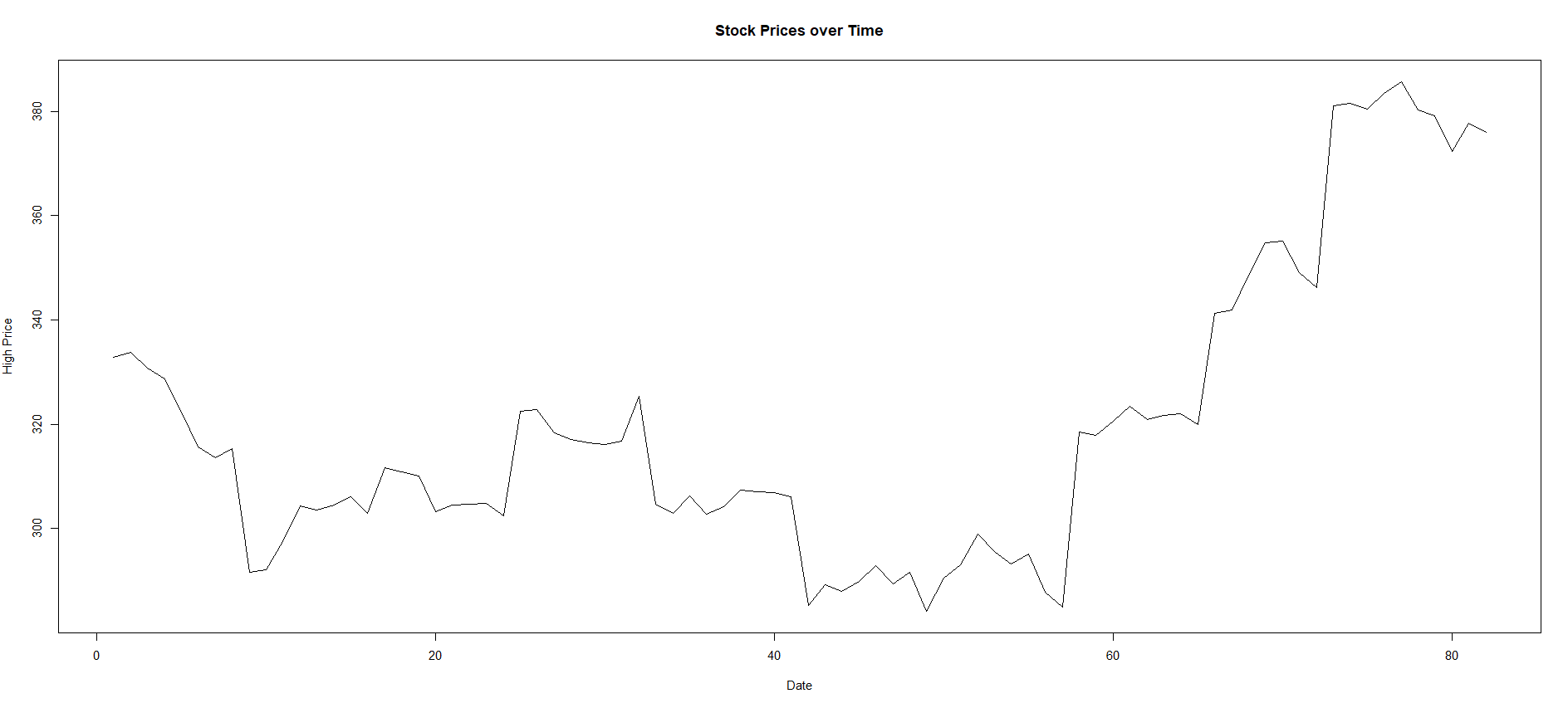


To see the distribution of opening prices, the "hist" function was used to produce a histogram for the "Open" variable.



To depict the frequency of high prices, a density graph of the "High" variable was made using the "density" function.



The "plot" function was used to produce the line graph of the "High" variable in order to show the trend in high prices over time.

***Results:***

From the initial analysis, the following results were obtained:

The most often opening price, as indicated by the frequency table for the "Open" variable, was $312.59, which occurs 15 times in the dataset.

The most frequent opening and closing price combination, which occurred three times in the dataset, was $314.43 and $313.96, respectively, according to the cross-tabulation of the "Open" and "Close" variables.

The closing price distribution was roughly normal, with a mean of $321.97 and a standard deviation of $12.57, according to the histogram of the "Close" variable.

The distribution of high costs was somewhat normal, with a mean of $325.72 and a standard deviation of $16.30, according to the density plot of the "High" variable.

The high prices of Google stock fluctuated throughout the year, with the highest prices in March and the lowest prices in July, according to the line graph of the "High" variable.

The first and final five rows of the dataset's table revealed that, with sporadic dips, the price of Google shares rose throughout time in general.

***Reference:***

Medha Rawat, (2018 April 09). Google Stock Price, Kaggle

<https://www.kaggle.com/datasets/medharawat/google-stock-price?select=Google_Stock_Price_Train.csv>