

Part 2 - Dataset Analysis

Thomas Reid Zuk

March 13, 2016

https://github.com/Freakazoidile/SRT_Assignment_Graphs/tree/master/Part2

I have chosen to analyze the historical quote trading information of Nasdaq Inc stock which is listed as NDAQ on the NASDAQ stock market. Nasdaq is the corporation that owns the NASDAQ stock market in the United States as well as many other world wide stock markets including 8 European stock markets.

This report is going to explore the historical quote information from the past 10 years and various type graphs that investors and market analyzers use on a regular basis.

The dataset contains the Open, High, Low, Close prices in USD and the trading volume for every single day the market was open the past 10 years. While 4 pieces of data may not seem like much given how much money surrounds the stock market many ways to represent this data, and perform analysis of it should be no surprise. There is well over 100 different stats that are based on the 4 prices and volume alone. Investors love all these stats, graphs, and charts as it allows them to make educated and informed decisions with their money.

The dataset was obtained from the Nasdaq website for historical quote prices at

<http://www.nasdaq.com/symbol/ndaq/historical>

```
options(warn=-1)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)
library(zoo)
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
library(TTR)
library(xts)
```

```
##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
## first, last
```

```
library(scales)
library(candlesticks)
```

```
## Loading required package: quantmod

## Version 0.4-0 included new data defaults. See ?getSymbols.
```

```
library(quantmod)

setwd('E:/Documents/School/Winter2016/srt411/assignment1/pt2')

ndaq <- read.csv('ndaq.csv', header = TRUE)

#reformat to R date format
ndaq$date <- as.Date(ndaq$date, "%Y/%m/%d")
```

Dataset at a first glance

```
summary(ndaq)
```

```
##           date           close           volume           open
## Min.      :2006-03-13   Min.      :15.76   Min.      : 235800   Min.      :16.25
## 1st Qu.:2008-09-10   1st Qu.:23.45   1st Qu.: 1218544   1st Qu.:23.42
## Median :2011-03-10   Median :30.03   Median : 1871252   Median :30.04
## Mean    :2011-03-11   Mean    :32.13   Mean    : 2246176   Mean    :32.13
## 3rd Qu.:2013-09-10   3rd Qu.:39.07   3rd Qu.: 2802556   3rd Qu.:39.04
## Max.    :2016-03-11   Max.    :65.98   Max.    :29537140   Max.    :65.48
##           high           low
## Min.      :16.92   Min.      :14.96
## 1st Qu.:23.73   1st Qu.:23.11
## Median :30.61   Median :29.37
## Mean    :32.61   Mean    :31.64
## 3rd Qu.:39.67   3rd Qu.:38.56
## Max.    :66.40   Max.    :64.87
```

```
head(ndaq)
```

```
##      date close  volume  open   high   low
## 1 2016-03-11 65.98 1109372 65.48 66.400 64.54
## 2 2016-03-10 64.97 2353457 64.80 65.075 63.11
## 3 2016-03-09 65.37  749249 65.41 65.950 64.87
## 4 2016-03-08 64.87  953765 63.86 65.238 63.80
## 5 2016-03-07 64.24  761011 64.77 64.900 63.96
## 6 2016-03-04 64.94 1176889 65.25 65.490 64.08
```

The summary data takes a quick glance at Mean, Median, Max, Min, 1st/3rd Quartile values of all the prices provided. The median and Mean values for all 4 prices columns (high, low, open, close) are pretty much the same.

The Head data provides a quick look at what the raw rows look like. Fairly basic, just a date, a number indicating volume, and then the 4 price measurements. No need to continue looking at numbers, let's turn it into some useful information.

Let's start by looking at a basic line chart of the stock price with a 200 day moving average line

```
ndaqMav <- SMA(ndaq$close, n=200)

ggplot(ndaq) +
  geom_line(aes(date, close, group=1, colour="Price")) +
  geom_line(aes(ndaq$date,ndaqMav, colour="200 Moving Day Average")) +
  scale_color_manual(values=c("red", "black")) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_date(breaks = date_breaks("6 months"), date_labels = "%Y - %m") +
  theme(legend.position = c(0.1,0.9)) +
  ylab("Price in $ USD ") +
  xlab("Date Y-M") +
  ggtitle("Historical Closing Price")
```



The price may seem extremely erratic and reinforce the thought of stocks being a gamble to someone not familiar with the market however there are many indicators and multiple statistics to analyze about a stock.

The moving day average (MDA) in this example is a 200 day moving averaging meaning the price is weight based on 200 samples for a any single point on the graph.

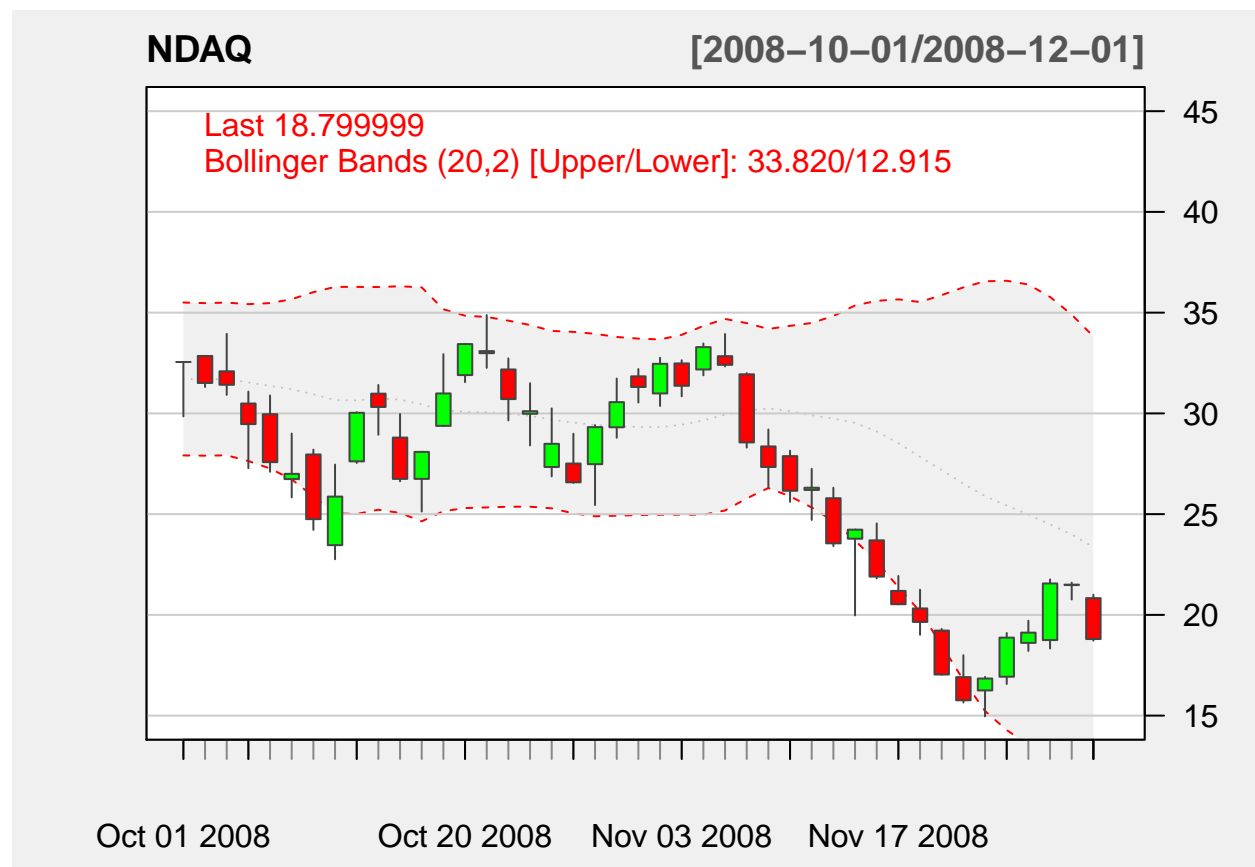
200 MDA is a very popular, and basic metric to look at when weighing how a stock is doing. In addition to providing a smoothed out way to better represent trends in prices the value of a stock compared to the MDA provides some interesting feedback. Typicall when a stock is below the 200 MDA it is seen as cost in some aspects, and when it is above the 200 day MDA

Other lengths of MDA's are used, such as 50 and 100 day which provide a glance at the day to day trends of a stock.

The following graphs were created with help from a great package called 'quantmod' which was created for stock market analysis. It takes these 6 extremely basic rows and provides immensely power visulation capabilities that most financial instutions and professional stock market software can offer. The largest difference being the time frame the data was capture. Most companies pay large amounts of money to information in real-time with milisecond level expectations.

This graph has Candlesticks (similar to boxplot) along with Bollinger Bands showing the drop of NASDAQ Inc stock during a couple months in 2008.

```
candleChart(NDAQ, theme='white',up.col='green',dn.col='red' , TA=c(addBBands()), yrange=c(15,45), subset=
```

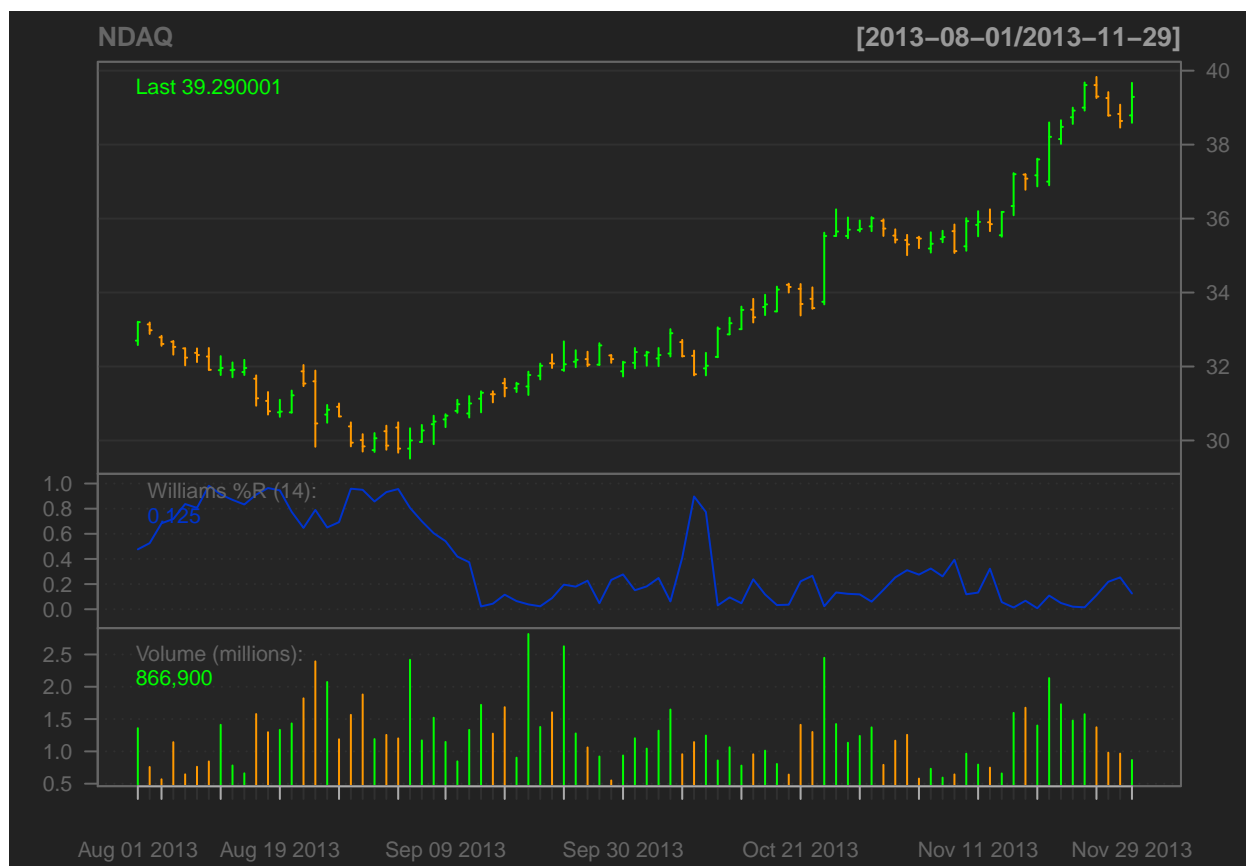


The body (coloured portion) of a Candlestick represents the open and close for the day, the vertically extending lines show high and low. Green or red bodies indicate if it went up or down from the start of the day to the end.

Bollinger Bands represent the volatility of a stocks prices as well as contains a moving average line in the center. The widely diverging bands indicate the price is extremely volatile. Bollinger Bands can be calculated to represent a short or long term volatility by changing the amount of data values are used to calculate the high and low of a single give point.

Bar chart of the stock price during a large increase in price. Williams %R, and Volume of trades are featured as well.

```
barChart(NDAQ, TA=c(addWPR(),addVo()), subset='20130801::20131201')
```



The bars represent the high and the lowest price of the day and the small sticks represent the opening and close. The tick on the left is the open, and the tick on the right is the close.

The Williams %R can be thought of as indicating the momentum, inversely of the price. Traders who use Williams %R as an indicator typically pick a high and a low value such as 0.2 and 0.8 as indicators of when a stock is abnormally high or abnormally low. When a stock is above or below their preferred indicators they typically take notice and act accordingly.

Looking at both the price and Williams %R of early September/end of August reveals provides a great example of how well these metrics reflect and help to predict a stocks price.

The price experienced a 'double bottom' which typically indicates the stock price is going to bounce back up, which is exactly what happened.

The Williams %R was above 0.8 which usually indicates a usually low price. The Double Bottom is represented in the Williams %R more visibly by a Double Top.

There are many different metrics of measuring a stocks performance than the few explored in these charts. Investors use different measures of performance to get a picture of what the trends of a stock are in both long term year-over-year data, and to look at the extremely narrow intra-day level statistics riding the up's and down's of the day or week.

Sources

<http://www.inside-r.org/packages/cran/TTR/docs/GD>

`install.packages("candlesticks", repos="http://R-Forge.R-project.org")`

https://r-forge.r-project.org/R/?group_id=1317