So what can r markdown do?

Josh Hanna

September 14, 2016

Getting started

To play with r markdown, I have downloaded 5 years worth of data on monthly United States Brent Oil Fund prices, Horizons SP/TSX 60 Canadian dollar fund prices, and the conversion of USD to CAD. I will run correlational analyses, and demonstrate the use of *ggplot2* to create some pretty line graphs.

The first step is opening up the libraries we will need. These include *ggplot2*, the grammar of graphics plot, and *apaTables*, a library for the creation of pretty apa correlational, regression, and ANOVA tables created by Guelph's very own David Stanley.

```
library(ggplot2)
library(apaTables)
```

```
## Warning: package 'apaTables' was built under R version 3.2.5
```

The next step is **importing data** into a new data frame

```
#It is worth noting that using a hashtag before text in script will keep the code from executing.
#also, creating text, followed by a little arrow, and a command will create a new data fram, in
#this case a whole data set
#but I will also use this command later to build a graph
BNO.HXT.5.year.monthly<-read.csv("BNO HXT 5 year monthly.csv", header=TRUE)
```

And then creating pretty **APA correlation tables** to get an idea of the relationships between our variables. BNO is a United States Brent Crude Oil fund, and HXT is a Canadian large market capitalization stock index. CAD USD is the exchange rate(the price in Canadian Dollars of one American dollar). All are reported in CAD, and prices of the indices are adjusted for stock splits where appropriate.

```
apa.cor.table(BNO.HXT.5.year.monthly,filename="apacorrelationtable.doc", show.conf.interval=TRUE)
```

```
##
##
## Means, standard deviations, and correlations with confidence intervals
##
##
##   Variable    M     SD    1            2            3
##   1. Date     30.50 17.46
##
##   2. BNO_Open 35.64 10.03 -.67**
##                           [-.79, -.50]
##
##   3. HXT_Open 24.04 3.01  .87**        -.38**
##                           [.79, .92]   [-.58, -.14]
##
##   4. CAD_USD  1.12  0.13  .91**        -.84**       .73**
##                           [.86, .95]   [-.90, -.75] [.59, .83]
```

```
##
##
## Note. * indicates p < .05; ** indicates p < .01.
## M and SD are used to represent mean and standard deviation, respectively.
## Values in square brackets indicate the 95% confidence interval for each correlation.
## The confidence interval is a plausible range of population correlations
## that could have caused the sample correlation (Cumming, 2014).
##
```

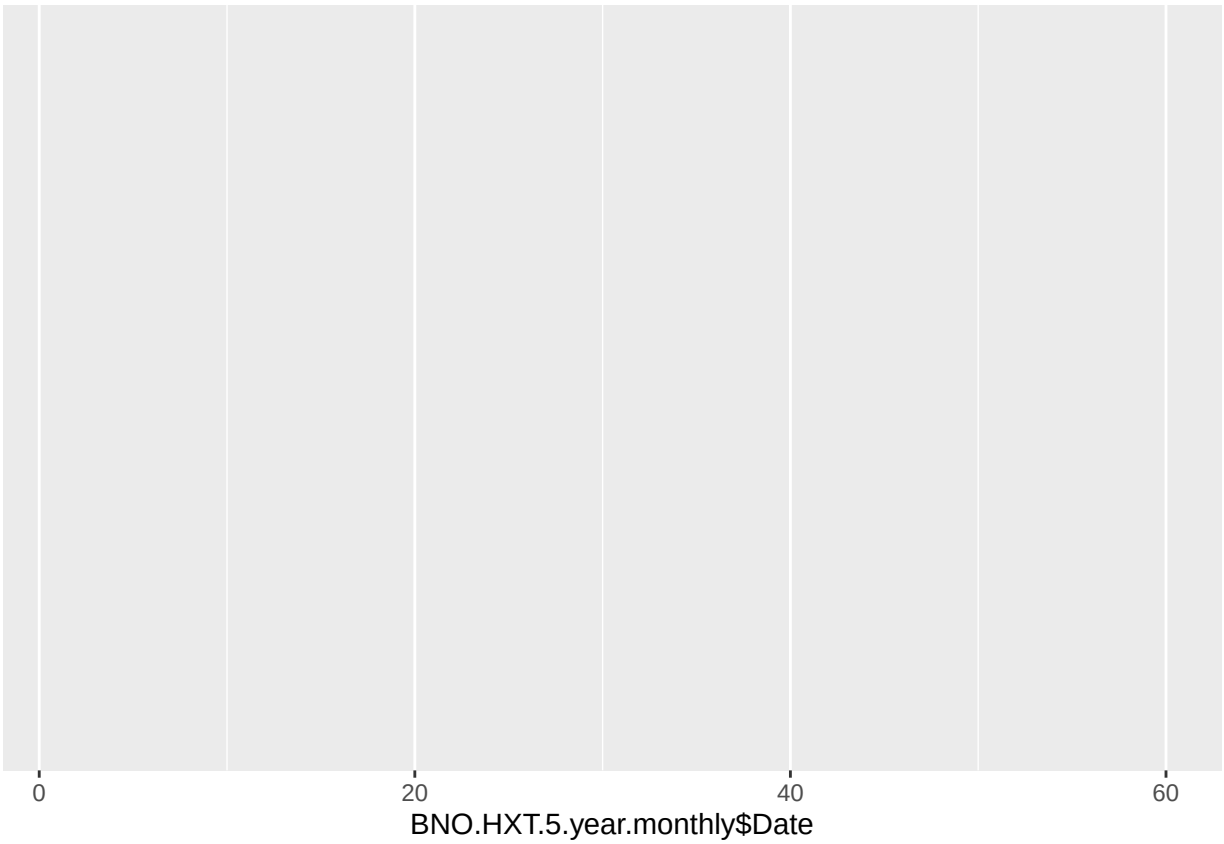
Results

Now those results are interesting. APATables made a cute little table with significance tests, and confidence intervals for the effect sizes of relationships between our variables. It looks like the price of oil has declined significantly since 2011, the Canadian stock index has increased in value, and the cost of American money has gone up. There is a smaller, though still sizable negative impact of oil on the value of the Canadian stock index, which makes sense given that 13 of the largest companies in Canada are energy companies with interest in oil. The strength of the relationship between oil prices and the exchange to American dollars is quite astounding really, the relationships accounts for about 70% of variance, even grater than the approximately 53% accounted for by the performance of Canada's largest companies.

Graphing using GGplot2

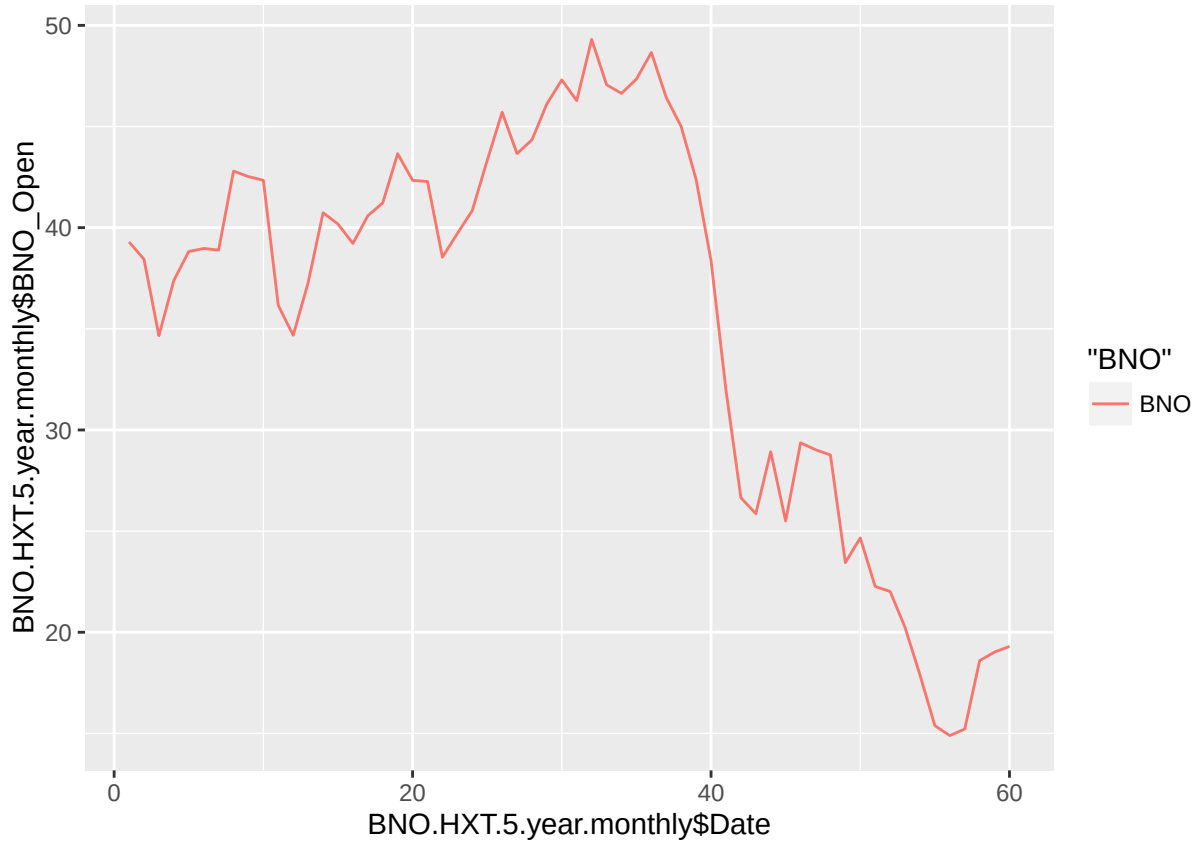
GGplot2 functions kind of like a series of transparancies to build a pretty looking graph. Each layer of code adds something to the graph, like labels, new data representations, changes colours, or scales. I will demonstrate how it works with a series of different codes to, hopefully, add some progressively more pretty graphs to this document.First, I am going to create a new data frame I will label 'Line1' which is going to serve as the base for the line graph.

```
Line1<-ggplot(BNO.HXT.5.year.monthly, aes(BNO.HXT.5.year.monthly$Date))
print(Line1)
```

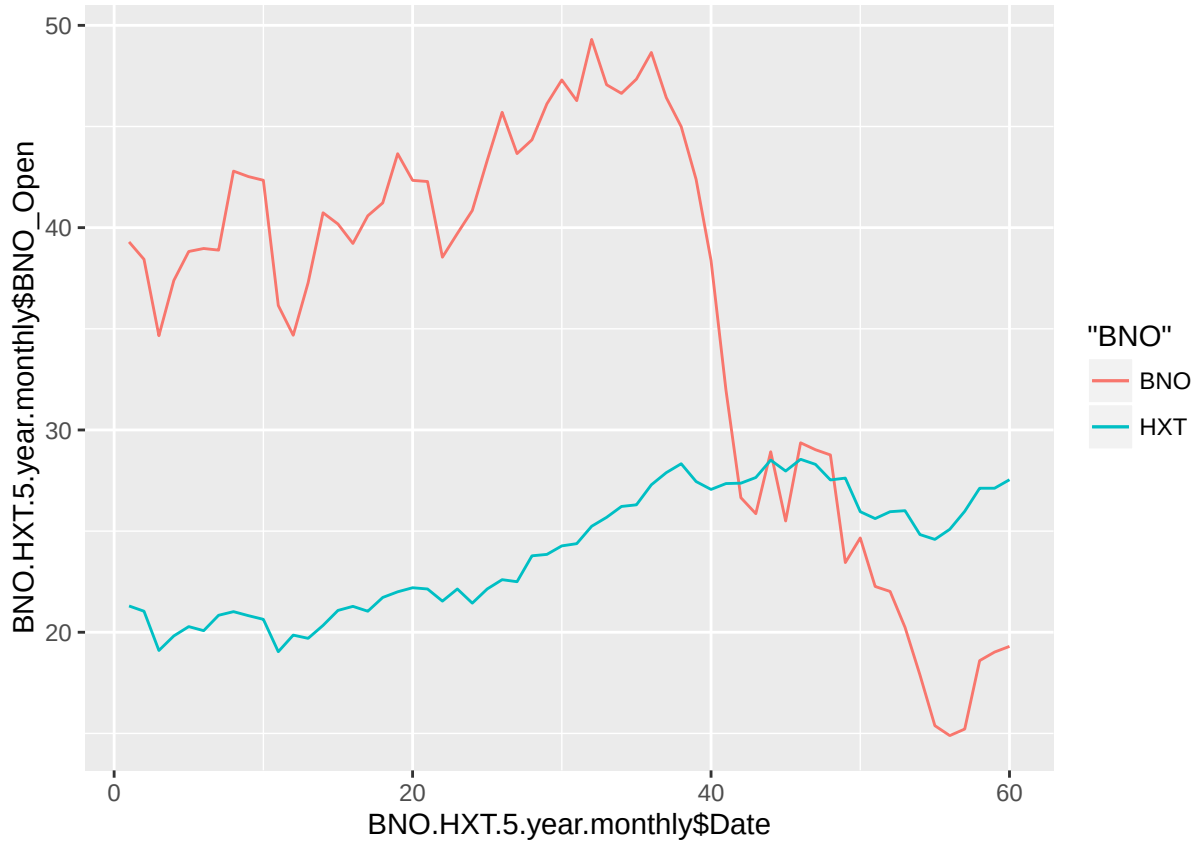


Alright, so that created a plot, but it is lacking actual data lines, so the next few chunks of code are going to create lines for the three variables.

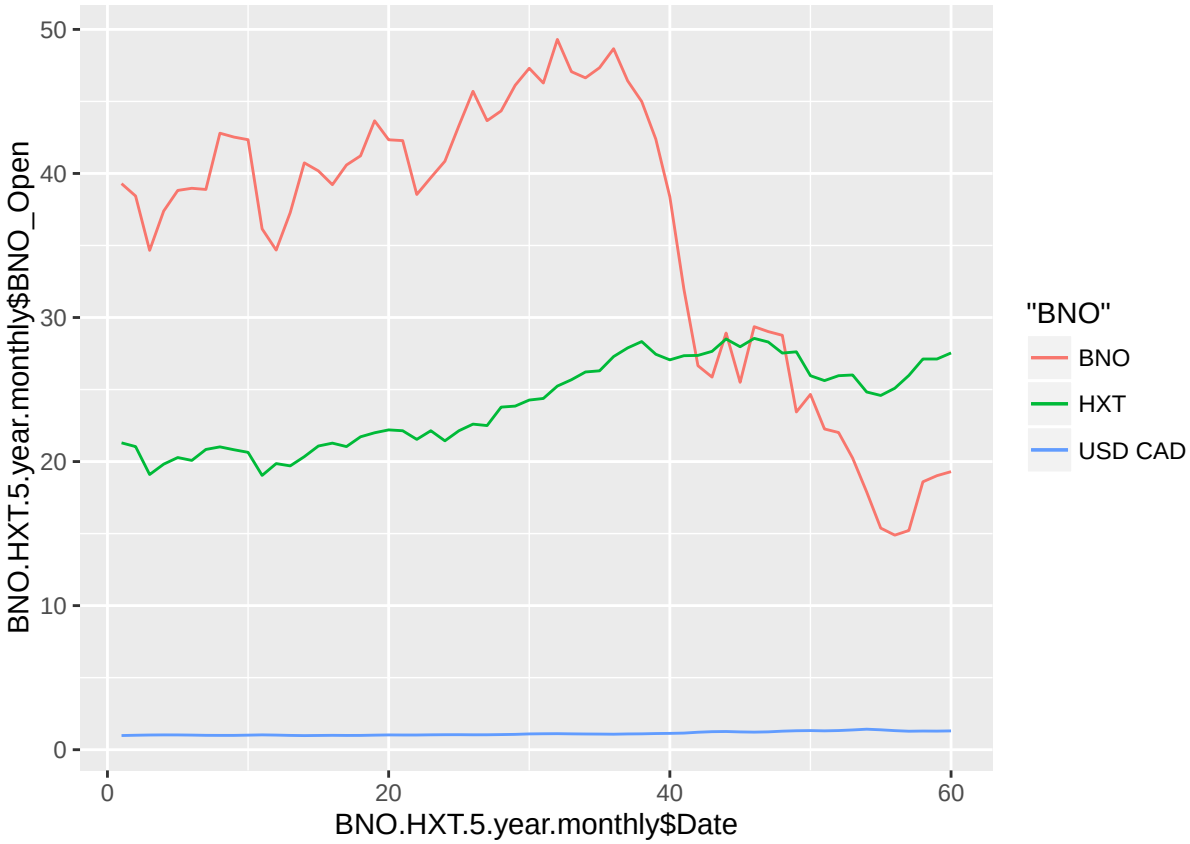
```
#It's important to note here that typing the name of your data set followed by a dollar sign,
#and the name of a variable selects a single variable for use in the graph. This is also useful
#in more in depth analyses like regression where you will want to regress one variable on another.
Line1<-Line1+geom_line(aes(y=BNO.HXT.5.year.monthly$BNO_Open, colour="BNO"))
print(Line1)
```




```
Line1<-Line1+geom_line(aes(y=BNO.HXT.5.year.monthly$HXT_Open, colour="HXT"))
print(Line1)
```

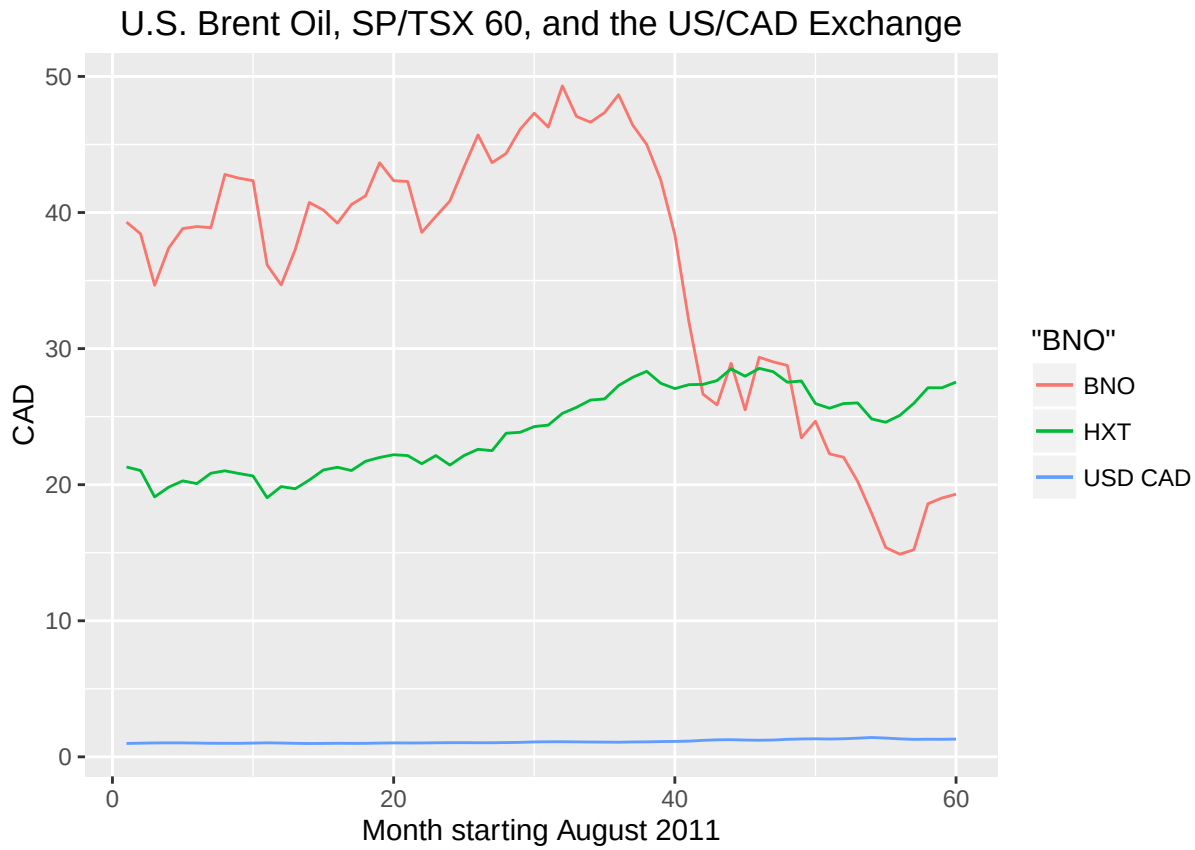



```
Line1<-Line1+geom_line(aes(y=BNO.HXT.5.year.monthly$CAD_USD, colour="USD CAD"))
print(Line1)
```



Okay, we have something that resembles a line graph, but the labels are a bit messy, so we will replace them with plainer language using the following code chunk.

```
Line1<-Line1+labs(title="U.S. Brent Oil, SP/TSX 60, and the US/CAD Exchange",
                  x = "Month starting August 2011", y = "CAD")
print(Line1)
```



That is that, creating a line graph and APA format correlation table is pretty easy with the right code and some clean data. ANd about three hours of work, but who is counting.