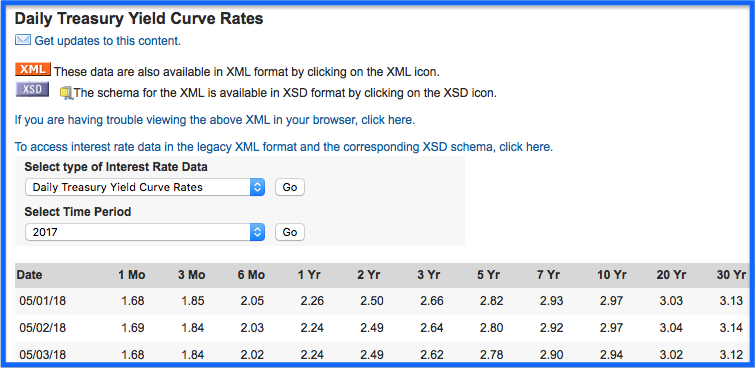
**Unit 1 Capstone**

by John Foxworthy

<https://github.com/sarilacivert/Data-Science/blob/master/Unit1capstone.ipynb>

The dataset that I will be using are the Daily U.S. Treasury Yield Curve Rates and its significance is the global importance in the financial trading industry, data quality and its statistical significance. The figures below have daily dates as rows with maturity points as columns. For example, a ten year is 2.97 percent as of the 1st of May, 2018, which is the amount of the interest rate on a decade long loan. The data goes back to 1990 up until to the current date in 2018.

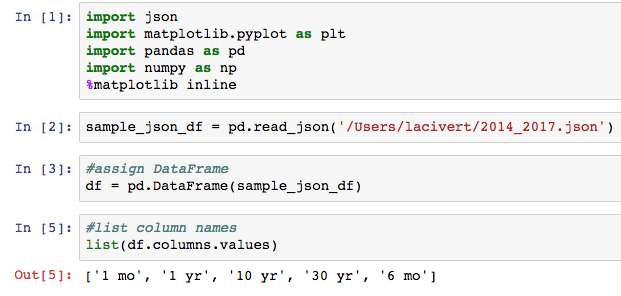


[Daily\_Treasury\_Yield\_Curve\_Rates](https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield)

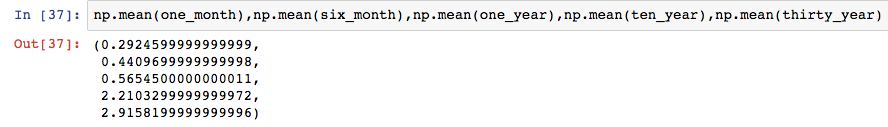
Each of the figures represents hundreds of billions of U.S. dollar trades from participants from dozens of countries and have the impressive data attribute of a complete absence of stale data. Not some, but all figures, on each business day of the year, has at least one person and likely many people trying to influence the interest rate values to change. Specifically, the strong liquidity, in other words, the quantity of buyers finding sellers and vice versa is very high, in every trading day, in the year going back decades since the U.S. Dollar is the world’s reserve currency. According to organizations like the International Monetary Fund (IMF), 6 out of 10 or 7 out of 10 transactions by physical hand or electronically, involve the U.S. Dollar. There are more than 200 countries in the world and each have some type of exporting and importing business that does involve the U.S. Dollar. The more a country exports, the more their currency, not the U.S. Dollar, rises in value because of increased demand, thereby making the value of their exports more expensive. To counteract, countries buy U.S. Dollars to dampen their local currency appreciation and the most common format is in the U.S. loan or fixed income market. Borrow a month at 1.68% or a year at 2.26% in U.S. dollars on the 1st of May for your exporting business from the Daily Treasury Yield Curve, for instance. Not to mention, the U.S. economy is the largest in the world so there are plenty of individuals and institutions that depend on this data to raise money for countless causes from a mortgage loan for a residential property to purchasing a new office building for an expanding company or local government institution.

Lastly, the source of this data is the U.S. Treasury department of the U.S. government. There are no impediments for the use of this data as it public from a legal perspective and officially published by the U.S. government on every business day. Daily Treasury Yield Curve is widely used by countless people and institutions for many causes and I choose it to forward my career in Data Science.

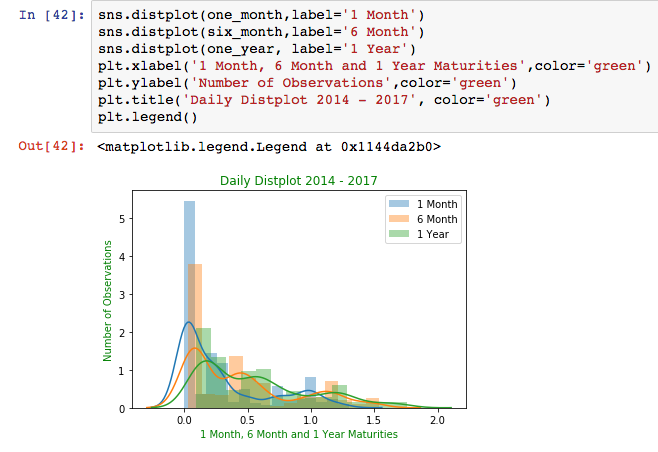
There are roughly 250 business days in a given year so beyond the demonstration below there are about 6,750 business days for this single column only from 1990 to 2017. Furthermore, the 11 columns from the 1 month to the 30 year would be more than 75,000 interest rate daily data from 1990 to the current day. For this report, we will use daily yield curve rates from 2014 to 2017, which is 1,000 data points per column maturity point for a total of 5,000.



Furthermore, there is an expectation that the longer maturities have less noise than the shorter maturities, which contribute selling more stable products in the financial services industry. A car loan with a ten year debt or a house for a family with a thirty mortgage are stable products compared to shorter term loans by small business owners. Cars and houses are conservatives investments compared to an entrepreneurs’ ideas. Conversely, the null hypothesis is the opposite of our expectation of shorter term maturities that would be more stable and less noisy than longer term maturities. The sufficient sample size of 1,000 data points over a 4 year period from 2014 to 2017 for each maturity point of the 6,750 data points available for each maturity point from 1990 to 2017.

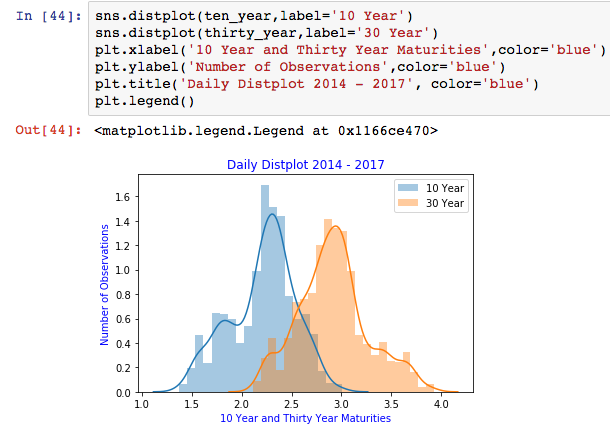


If we take the average of each of the 1 month, 6 month, 1 year, 10 year and 30 year maturity points, then we see a progression above , but how stable is it?



Furthermore, there is an expectation that the longer maturities have less noise than the shorter maturities, which contribute selling more stable products in the financial services industry. A car loan with a ten year debt or a house for a family with a thirty mortgage are stable products compared to shorter term loans by small business owners. Cars and houses are conservatives investments compared to an entrepreneurs’ ideas.

The distribution have a pattern of shorter maturities clustering together with more observations skewed to one side then smoothing out away from zero. The longer maturities mirror each other and overlap so there is some implication of stability long term.

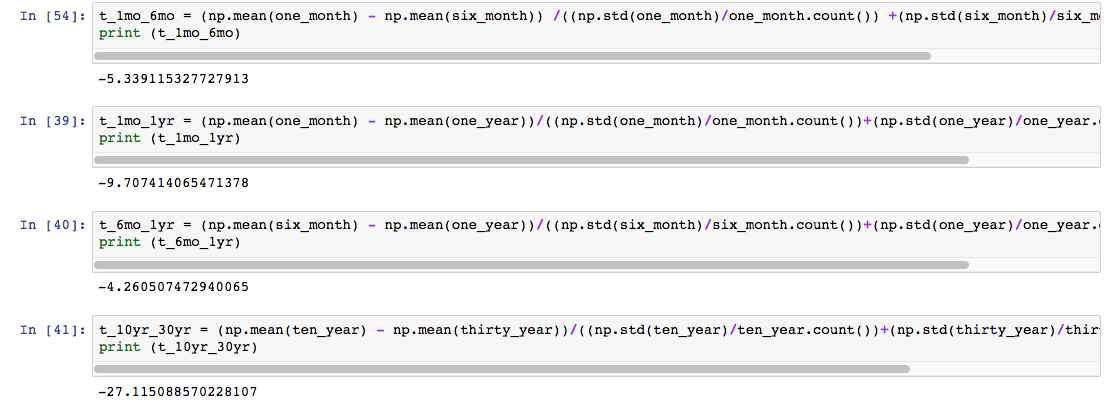


The distplot demonstrates more observations on one side with the shorter maturities more than the longer maturities, which adds to a picture of statistical noise smoothing out over time. If we take a look at means of each maturity, then we can use the t - statistic to analytically demonstrate the statistical noise pattern.



t\_1mo\_6mo = (np.mean(one\_month) - np.mean(six\_month)) /((np.std(one\_month)/one\_month.count()) +(np.std(six\_month)/six\_month.count()))\*\*0.5

If we compare the means of each of the shorter maturities, and then separately with the longer maturities, then we will thereby answer the hypothesis of noise smoothing.



We can reject null hypothesis that shorter maturities are more stable, less noisy than longer maturities because the t - statistic grows larger from short term maturity t - statistic to long term t - statistic. The 1 month and 6 month t - statistic is -5.33 and as you t - stat the others you end with a -27.11 with the 10 year - 30 year t - statistic. However, we cannot accept the hypothesis that longer maturities are more stable as this is one iteration of sampling or just a 4 year period in almost 3 decades of daily data.