1. **Do you want to work as an Application User, Scripter, or Hacker? Read**[**Enterprise Data Analysis and Visualization: An Interview Study**](http://vis.stanford.edu/files/2012-EnterpriseAnalysisInterviews-VAST.pdf)**by Kandel, Paepcke, Hellerstein, Heer**

**Solution:**

* Being a software engineer, programming and coding always excites and influences me.
* Since hackers are supposed to be the most proficient programmers of the three groups and the most comfortable manipulating data, I would like to work as a hacker since this is something which would motivate me to work on.
* I am always keen to learn new languages and multitask things.
* I can work on SQL queries to perform different operations on data and analyze it and pull it from any analytical tool of choice.

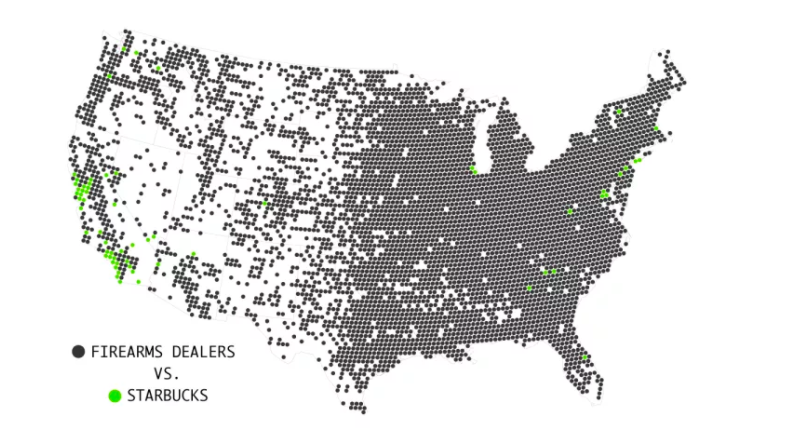
1. **From the FollowingData website read a**[**Project**](http://flowingdata.com/category/projects/)**that uses maps. Explain what the project was about and what data was used. Summarize the information being presented in the maps.**

**Solution:**

I have selected the project “Firearms Dealers vs. Burgers, Pizza, and Coffee”.

Based on listings from the Bureau of Alcohol, Tobacco, Firearms and Explosives, the graphic drew attention to the large number of firearms dealers and pawnbrokers — about 6 for 1 Starbucks.

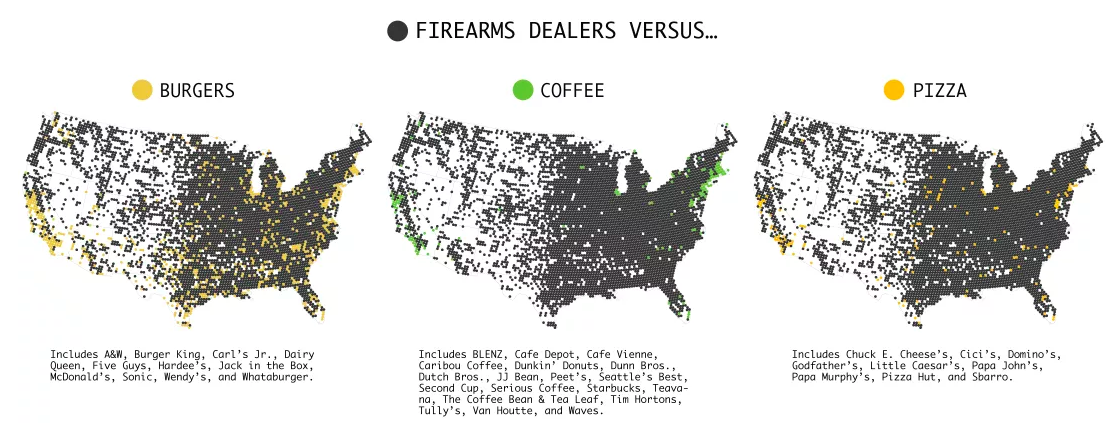
The ratio is 1:6 which sounds horrible.



In the map above Each circle represents an area with a 10-mile radius. The black circle represents Firearms dealers and green one represents Starbucks.

While for many, it seems like Starbucks is present nationwide, the coffee shop tends to concentrate in [densely populated areas](http://flowingdata.com/2014/03/18/coffee-place-geography/).

The following graphs represents the number of firearms dealers with Burger shops, Coffee dealers and Pizza dealers:



In all the above graphs, it is evident that the number of firearms dealer is much more than burgers, coffee and pizza shops.

In the following graph, Firearms dealers are plotted against all dealers on burgers, coffee and pizza dealers together.

# /Users/nikki/Desktop/Screen Shot 2018-01-29 at 3.58.52 PM.png

There are a lot of dealers. It’s hard to say how many of these licensees actively sell firearms, but still. It isn’t until I sum all the burger, pizza, and coffee chains together against firearms dealers that there’s a noticeable change.

But still it seems the number of firearm dealers is greater than the food dealers in USA.

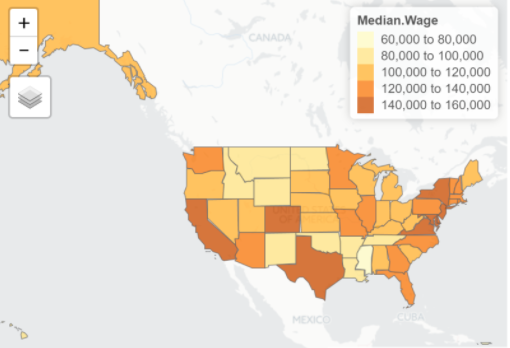
1. **From the**[**Revolution Analytics**](http://blog.revolutionanalytics.com/)**blog find a post that uses maps. Click on the graphics Category. Explain what the project was about and what data was used. Summarize the information being presented in the maps.**

**Solution:**

Following shown map is an example of geographic data maps with.

While it's been possible to create maps in R for a long time, some recent packages and data APIs have made the process much simpler. The tutorial is based on the following R packages- sf, tmap, tmaptools, tigris and rio.

[rio](http://mran.microsoft.com/package/rio/), a package to streamline the import of flat files from third-party data sources like the [**U.S. Bureau of Labor Statistics**](https://data.bls.gov/oes/#/occGeo/One%20occupation%20for%20multiple%20geographical%20areas)



The above map shows the median wages of people in USA ranging from 60,000 to 160,000.

The range has been divided in the window of $20,000 starting from $60,000 and ending to $160,000 and hence we can see 5 divisions.

The colors show incremental trend of wages which go from lighter color to darker presenting lightest color to represent lowest income and the dark most color to represent highest wages.

1. **Find a dataset of interest to you. First find a data source. Then download the data. Open the data in the visualization software. Make a map if you have the location or geolocation data. Or make some other kind of visualization. Comment about what your visualization is trying to communicate.**
   * **Make a plot of your data on a map using Google Charts (this is from Google Sheets in Google Docs).**
   * **Make the same plot of your data in Tableau.**

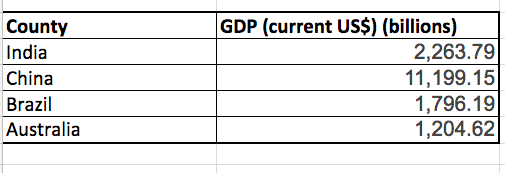
**Solution:**

My Data source:

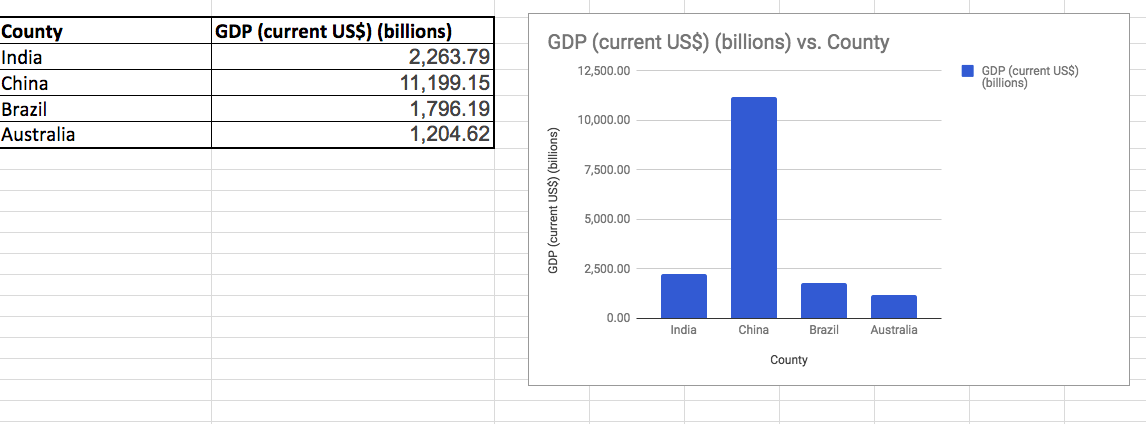
<https://data.worldbank.org/country/>

Then after entering the country names, I referred to Country profile, and got values for GDP (current US$) (billions) for every country.

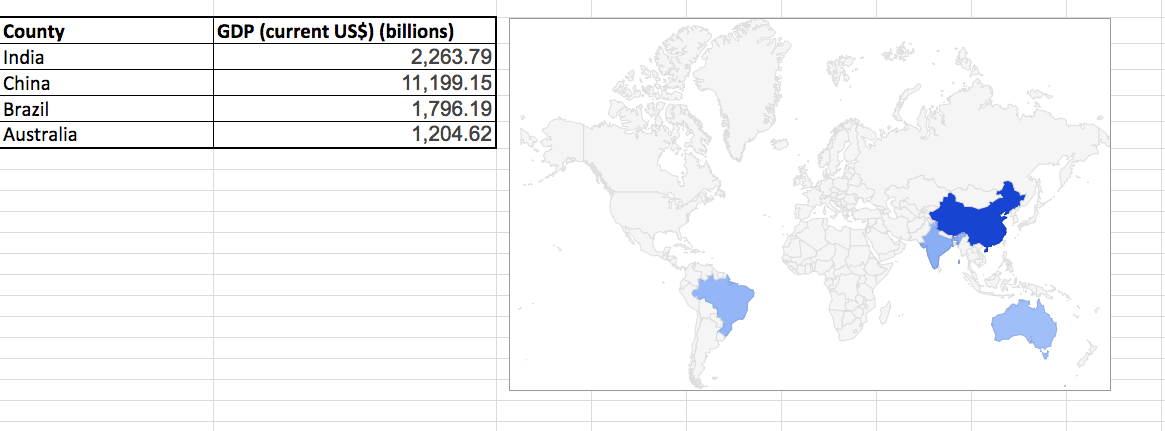
The data was inserted in MS Excel:



**PART 1:** The excel sheet is been uploaded in the Google sheets in Google docs.The bar chart is displayed below for the data considered:

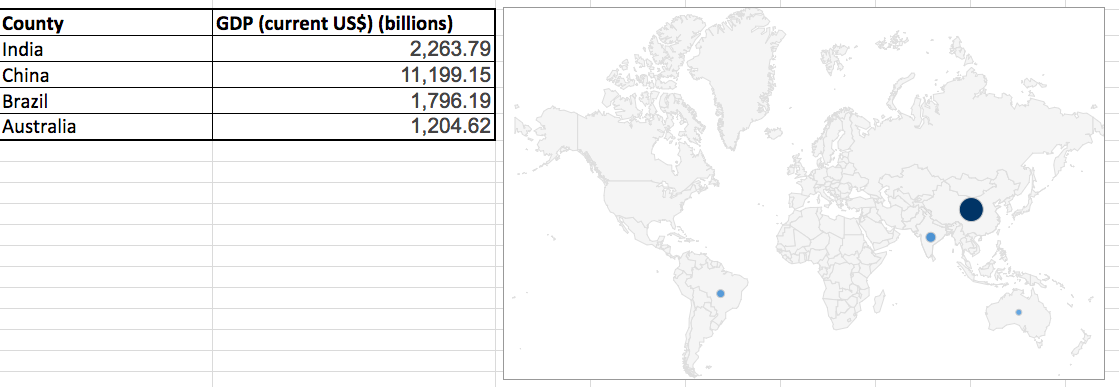


Another way is to display data by selecting Map – Geo Chart in the chart section:

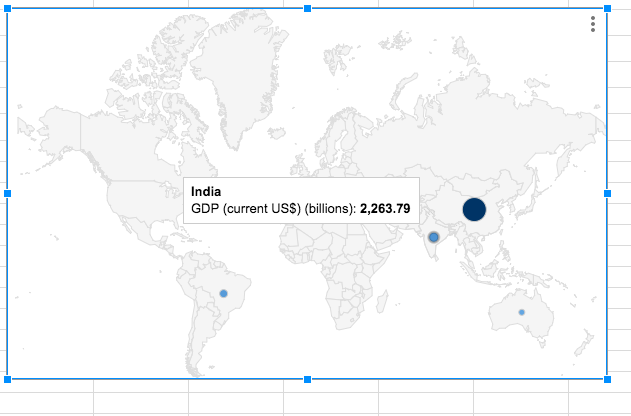


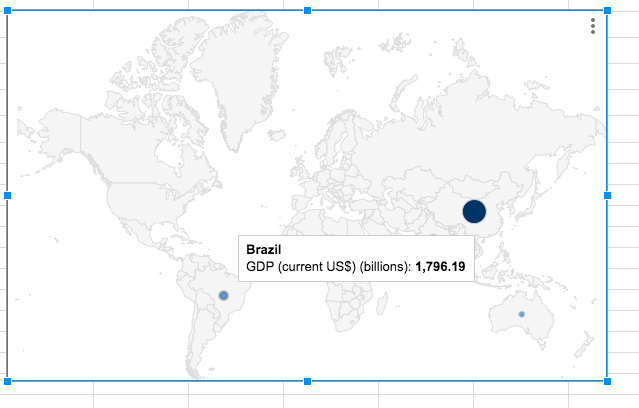
Here the colors represent the intensity of the data, which means the minimum value is $1,204.62 represented with lightest Blue color and which is in Australia, then medium blue color for Brazil and then India for medium values $1,796.19 and $2,263.79 respectively and the darkest blue showing that highest GDP is of China which is Max $11,199.15 values among all four countries.

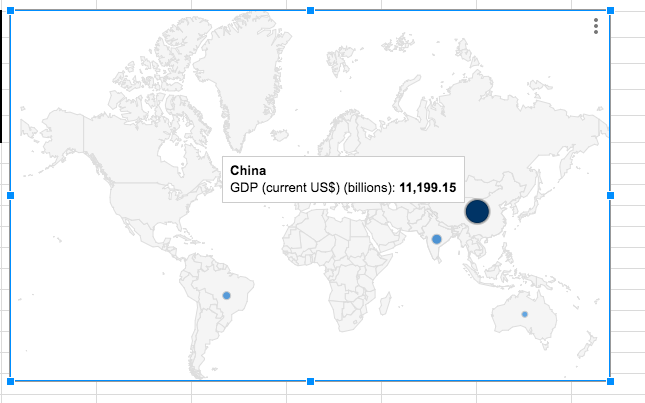
Also, one more way is to display data by selecting Map – Geo Chart with markers in the chart section:



The markers are represented in shades of blue colors - Here the colors represent the intensity of the data, which means the minimum value is $1,204.62 represented with lightest Blue color and which is in Australia, then medium blue color for Brazil and then India for medium values $1,796.19 and $2,263.79 respectively and the darkest blue showing that highest GDP is of China which is Max $11,199.15 values among all four countries.

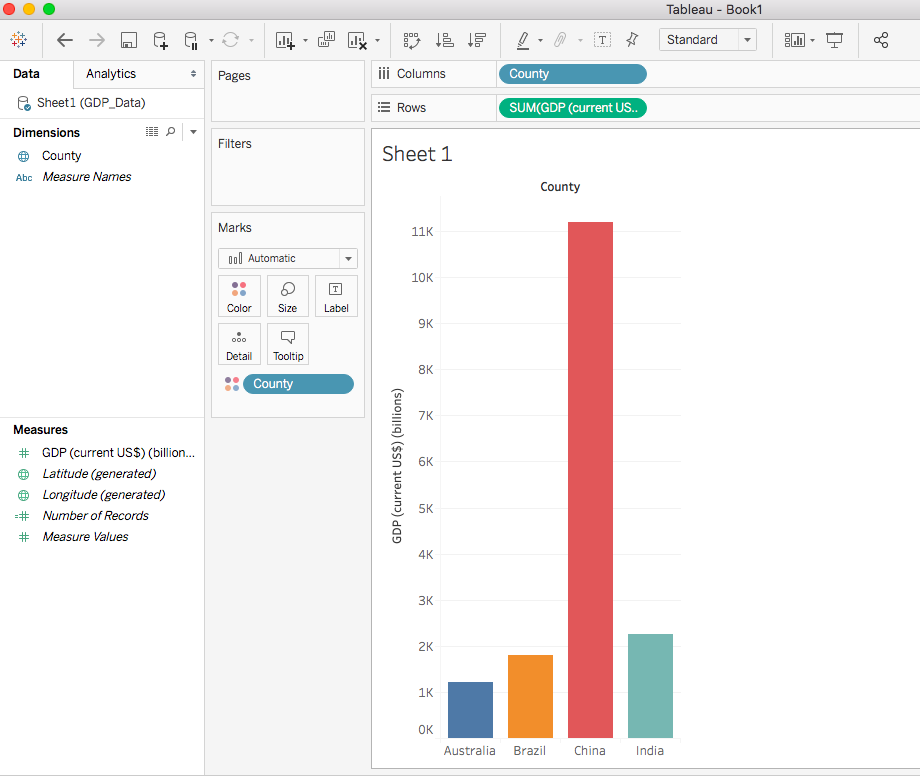






Part 2:

The similar Excel sheet is imported in Tableau. In the Dashboard, the dimensions are given as Columns – Country values and Row – GDP:

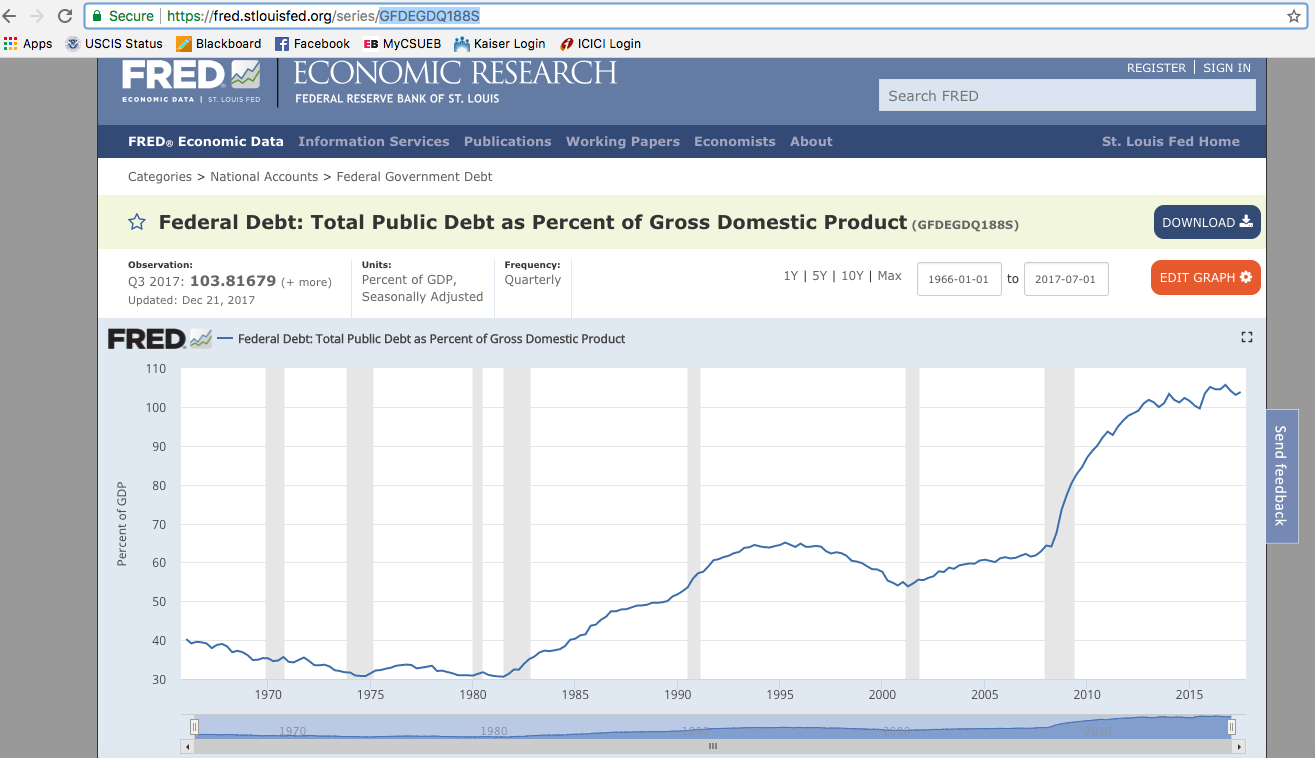


1. **Install the Quandl library in R. Find an interesting dataset from Quandl and make a time series plot in R.**

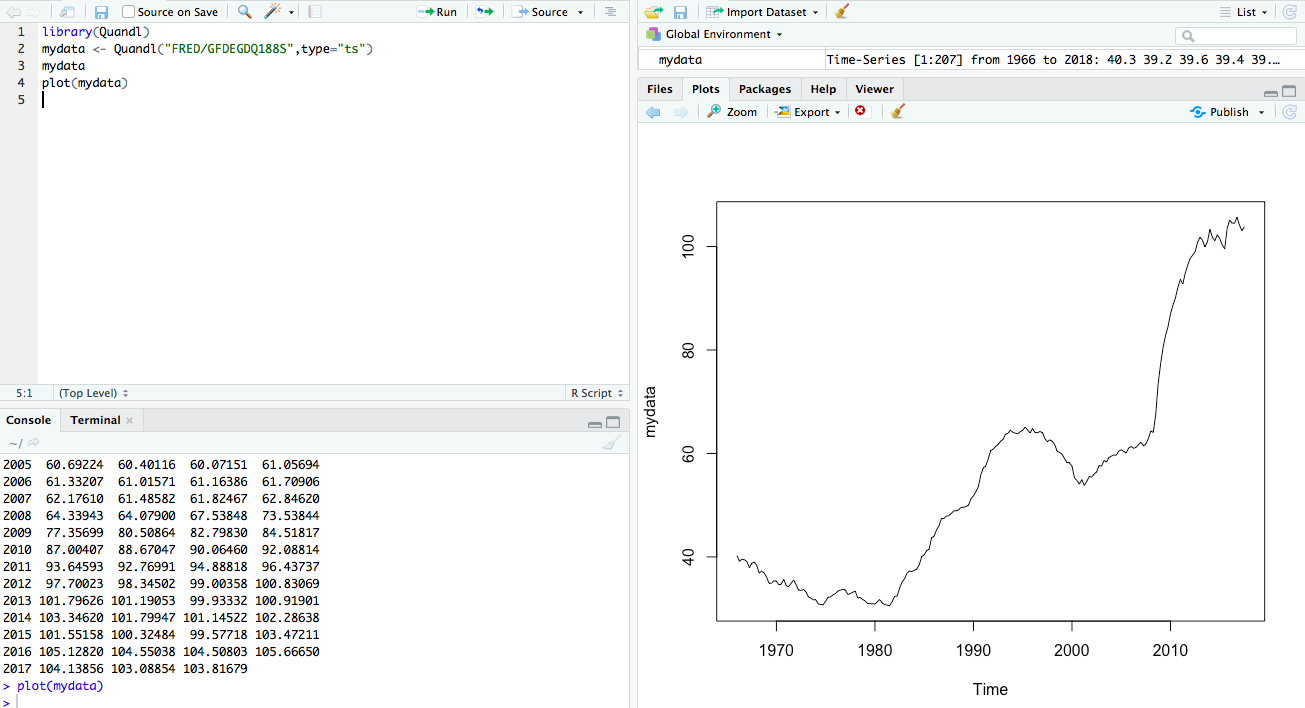
**Solution:**

I have selected data from FRED - Federal Debt: Total Public Debt as Percent of Gross Domestic Product (GFDEGDQ188S)

“GFDEGDQ188S” is used in the R code to extract the graph from the FRED using Quandl.



The code in R and the Time series plot for the above data in R is as follows:



1. **Using R run the mtcars examples from the end of the**[**Tools**](http://www.sci.csueastbay.edu/~esuess/classes/Statistics_6610/viz/viz5/Tools.html#/)**presentations. Show the plots you have created in your homework document.**

**Solution:**

