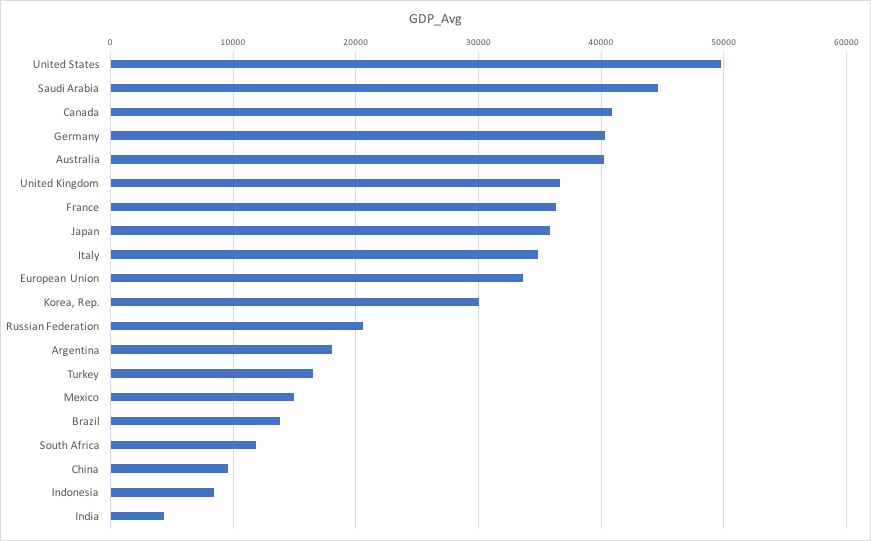
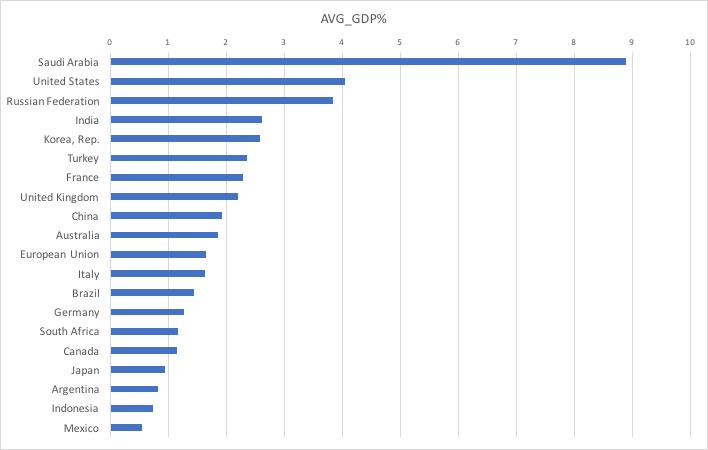
Outline - Military Spending by the G20:

DATS 6103 – Individual Project 1 – Nelson Foster

* Data Sources and Tools Used
  + The World Bank’s [Open Data](http://data.worldbank.org/) website was used to obtain datasets for [Gross Domestic Product (GDP) per capita](http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD) and [military expenditures by percentage of GDP](http://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS).
  + To create a data frame consisting of only the G20, a list of current G20 member nations was obtained from the [G20 website](https://www.g20.org/Webs/G20/EN/G20/Participants/participants_node.html) hosted by the Federal Government of Germany, who will be hosting the 2017 G20 summit in Hamburg.
  + Utilized Anaconda’s Jupyter Notebook and Microsoft Excel to conduct the analysis.
  + Referenced DataCamp training materials, class demonstration examples, and external public resources such as Github and Stack Overflow to navigate through challenges with mining the datasets.
* Exploratory Data Analysis
  + Loaded an assortment of libraries in the course of exploring various methodologies, however, Numpy and Pandas were used for the overwhelming majority of the analysis.
  + To preserve the integrity of the original files, copies of the datasets were made, and extraneous rows were removed from those copies. Initial exploratory data analysis was conducted on the files to determine average overall GDP and military expenditures by percentage of GDP. Two files, one for the GDP Per Capita (GDP.xls) and military expenditures by percentage of GDP (MilSpnd.xls) were read into the Jupyter notebook as data frames (“MS” and “GDP”).
  + Exploratory data analysis was conducted on each dataset by looking at the head, tail, descriptive statistics. This revealed that the datasets contained information on GDP and percentage of GDP dedicated to military spending for every nation from 1960 to present. Two initial dataframes were created from the originals:
    - **Military Expenditures by Percentage of GDP Dataframe (“MS”)**
      * To filter for just G20 member nations, the aforementioned list of G20 nations was referenced to create a new data frames by subsetting the original data frame (“MS”), using the **loc** function. The new dataframe was named “G20”
      * The format of the dataset has the year variable as columnar. Since many of the entries contained null values prior to 1988, and the scope of this project is to look at 5+ years of data for the G20, a **drop** function was used to remove every year prior to 2005. This function was also used to remove 2016, as this column had null values as well.
      * Because of the format of the dataset, it didn’t lend itself to a time series analysis. In researching this problem, the **groupby** function demonstrated in class for the terrorism dataset was used to create a new data frame, named “YearSum.” A time series plot was then created using this new dataframe. In researching different types of visualizations I discovered how to create a heatmap, however I didn’t find this to be a very useful visualization for this exercise.
    - **Gross Domestic Product (GDP) DataFrame (“GDP”)**
      * To filter for just G20 member nations, the aforementioned list of G20 nations was referenced to create a new data frames by subsetting the original data frame (“MS”), using the **loc** function. The new dataframe was named “GDPG20”
      * The format of the dataset has the year variable as columnar. Since many of the entries contained null values prior to 1988, and the scope of this project is to look at 5+ years of data for the G20, a **drop** function was used to remove every year prior to 2005. This function was also used to remove 2016, as this column had null values as well.
      * Because of the format of the dataset, it didn’t lend itself to a time series analysis. In researching this problem, I utilized the **groupby** function demonstrated in class for the terrorism dataset to create a new data frame, named “GDPYearSum.” A time series plot was then created using this new dataframe.
* Comparing GDP and Military Expenditures by Percentage of GDP
  + To determine the actual per person military spending (in dollars) amongst the G20 nations, a dataframe was created that multiplies the percentage of GDP in military expenditures (“YearSum”) by the total GDP (“GDPYearSum”), named “frames.” Then, the **pd.concat** formula was used on the “frames” dataframe to create a joined dataframe, named “results.” This new dataframe displays the G20 nations’ per person military expenditures in whole dollars.
  + To calculate and visualize the rate of growth in order to identify the top ten fastest growing G20 nations in terms of military expenditures, a new variable was needed for the “results” dataframe. The new variable, *Growth,* was created by combining the **np.exp , np.diff, and np.log** functions in a “[growth](http://stackoverflow.com/questions/26146759/calculating-growth-rate)” formula (**np.exp(np.diff(np.log(results))) – 1)**
* **Findings**
  + As stated earlier, initial exploratory data analysis was conducted in excel to determine basic summary statistics on the datasets. The distribution of average GDP and average percentage of GDP spent on military expenditures per G20 nation are as follows:

****

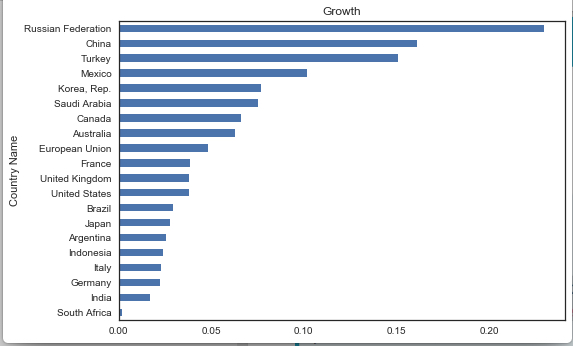
****

Among the G20 nations, the top ten in terms of overall GDP are as follows:

1. **United States**
2. Saudi Arabia
3. Canada
4. Germany
5. Australia
6. United Kingdom
7. France
8. Japan
9. Italy
10. European Union

However, the top ten in terms of military expenditure by GDP percentage are as follows:

1. **Saudi Arabia**
2. United States
3. Russian Federation
4. India
5. Korea, Rep.
6. Turkey
7. France
8. United Kingdom
9. China
10. Australia
    * According to the growth formula in python, however, from 2005 to 2015, the top ten G20 Member nations are as follows, with a few caveats:
11. **Russian Federation**
12. China
13. Turkey
14. Mexico
15. Korean Republic
16. Saudi Arabia
17. Canada
18. Australia
19. European Union\*
20. France
21. United States & United Kingdom\*



\*The European Union is a confounder because, though it is included in the G20 (G19 + 1), its government is constituted by a confederation of member states, many of which are included in the G20 individually. If the EU is dropped from the list, The United States and United Kingdom would be tied for tenth place.

References:

* Military Expenditure by GDP data:
  + <http://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS>
* GDP per capita data:
  + <http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD>
* Listing of G20 Member Nations
  + <https://www.g20.org/Webs/G20/EN/G20/Participants/participants_node.html>
* Research on compound annual growth rate, LOGEST, and Growth crosschecks in excel
  + <http://www.igetit.net/newsletters/y06_08/calculategrowth.aspx>
  + <http://feliperego.github.io/blog/2016/08/10/CAGR-Function-In-Python>
* Growth Rate Formula Research
  + <http://stackoverflow.com/questions/26146759/calculating-growth-rate>
  + <https://docs.scipy.org/doc/numpy-1.10.1/reference/generated/numpy.log.html>
  + <https://docs.scipy.org/doc/numpy/reference/generated/numpy.diff.html>
* G20 Image Credit
  + <http://www.de.digital/DIGITAL/Redaktion/EN/Dossier/g20-shaping-digitalisation-at-global-level.html>
  + <https://www.g20.org/Webs/G20/EN/G20/Participants/participants_node.html>