CASE STUDY 034
[Python]
Interactive Visualizations with Plot.ly

Difficulty Level: 2 of 3

Disclaimer: the dataset used in this case study was downloaded from The World Bank's website. The dataset was modified and prepared to simplify the analysis in this case study, but it can be downloaded in its original form https://data.worldbank.org/.

World Bank Open Data provides free and open access to global development data. I encourage you to explore the data and look for interesting indicators. In this case study we received data for the percentage of the population using Internet (API_IT.NET.USER.ZS_DS2) and the GDP (API_NY.GDP.PCAP.CD_DS2). There is a third dataset with information about continents.

You are a Data Scientist investigating how these two indicators correlate with each other along the years. Your objetive is to build an interactive tool that allows users to select a year an visualize a scatter plot.

Your analysis should follow these steps:

1. Install Plot.ly on Anaconda
2. Load the datasets

   a) Explore the datasets;

   b) Change the column names of Internet Users dataset including in the years the prefix 'iu_';

   c) Do the same with the GDP per Capita dataset, with the prefix 'pc_';
3. Merge the datasets

   a) Merge Internet Users and GDP per Capita dataset;

   b) Merge the internet users dataset and the continent dataset;

   c) Explore the new dataset;
4. Plot the scatter plot to investigate the relationship between GDP and Internet Users

   a) Plot the relationship between variables in the year of 2000. Make sure to set the title containing the year, and the axis labels.

   b) Improve the plot making annotation of the country name and changing the color accordingly to the continent color for each data point

   c) Plot again for the year of 2016
5. Build a function to receive a year and plot the graph

6. Make a list of available years
7. Create a Select widget with all available years and call the function created passing the selected year as a parameter

Good luck!

Difficulty note: this is a difficult assignment. Do not be surprised that there will be lots of nuances we have not covered off in the courses. But just like in the Real Life – there will be things training has not prepared you for and you will need to do research to find how to solve the problems at hand. If you get stuck, check the clues file.