7PAM2000 Applied Data Science 1 - Assignment 1: Visualization

**Student’s Name:Ratan Srivasthav Reddy Vullinkala**

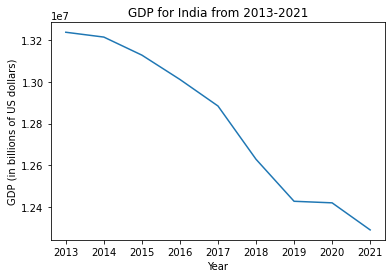
**Student’s ID:22031426**

# DatasetLink: <https://www.kaggle.com/code/loveakai/notebook9464449f05/input>

Github link: **https://github.com/Srivasthav09/srivasthav/upload/main**

# Visualization 1: GPD For India from 2013 - 2021

# Line Plot



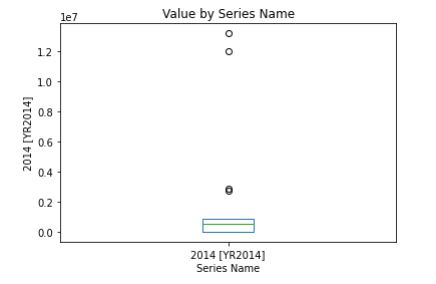
A line plot, also known as a line chart or a line graph, is a type of chart that displays data points connected by a straight line. It is a simple and commonly used way to represent data over time, with the x-axis typically representing time and the y-axis representing the values being measured.

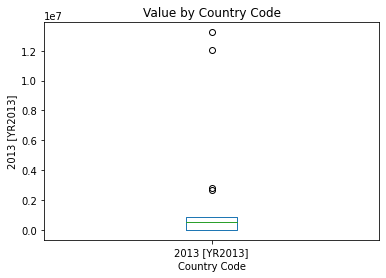
Line plots are useful for showing trends in data and identifying patterns, making them a popular choice in many fields, such as economics, finance, and science. They can be used to show the change in a variable over time, compare multiple data series, and identify relationships between variables.

Line plots can also be customized to display different types of information, such as using different colors and styles to distinguish between multiple data series or adding labels and annotations to provide additional context for the data.

Overall, line plots are a versatile and widely used tool for visualizing and analyzing data over time, and can be a helpful way to communicate complex information in a clear and intuitive way.

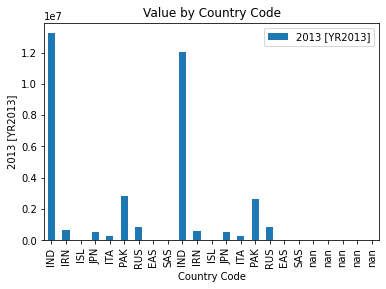
# Visualization 2:





The minimal, initial percentile, middle, and interquartile range, as well as the highest values, should all be shown in the same summary of any sort of collection of information values. Inside this line graph, a box distinguishing both the initial with third quartiles is being created, as well as a line graph also exists that traverses the box just in the middle. The y-coordinate in this graph shows the frequency distribution, and the x-axis denotes the data being displayed. Arrays, Language sequences, or multivalued arrays are all acceptable data values for the ax. Plot method. Let's create a line graph using NumPy.random.normal, which also takes inputs for mean, standard error, and necessary using NumPy.random.normal(), which takes arguments for the mean, deviation, as well as the necessary amount of variables to generate some random data. An observation that clearly differs from those other measurements in the dataset is an outlier. Anomalies or errors can be considered outliers. So, we must first identify the outliers before deciding whether to ignore them. An effective analysis tool for displaying outliers is a box plot. (Lähnemann, et, al, 2020).

# Visualization 3:

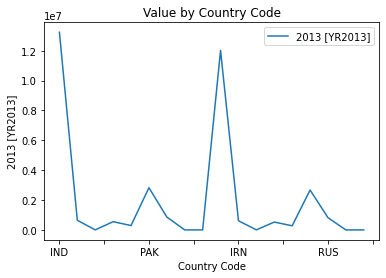


A sort of information visualization called a bar plot uses rectangular bars to represent data. Each bar's width denotes the type of data, and each bar's height denotes the number associated with that data point. ( Sarker, et, al, 2019).

Bar plots are frequently made using Python's mat plot lib module. For the purpose of producing bar graphs, the Mat plot lib bar method accepts information in the forms of lists, arrays, as well as python data frames.

Bar plots come in a variety of forms, such as horizontal bar plots, error column plots, bar charts plot lines, grouped bar plots, and basic bar graphs. The comparison of the data set is done using multiple bar plots while one parameter is increasing. We can quickly transform it into an industrial complex line graph, where each segment is shown one of these on the upper edge of the other. By adjusting the bars' thickness and placement, it can be plotted. (Caicedo, et, al, 2020).

# Visualization 4:



Python's matplotlib library is used for data visualization. A set of methods called py plot, which is a library of matplotlib, aid in the creation of several charts. The relationship between two pieces of data, X and Y, is shown using line charts on a distinct axis. Data visualization is the process of presenting data graphically in order to communicate information. It basically serves as a visual to aid in the interpretation of the data and in-depth research of its nature. While working with large amounts of data row-by-row is a very time-consuming activity, data visualization is a perfect substitute.

Data visualization is accomplished with the help of the Python module Seaborn, which itself is founded on matplotlib. It offers a way to exhibit data in the form of a statistical graph as an interesting and appealing way to communicate certain information. Functionalities occasionally need that data to be compared to one another, and under these circumstances, numerous plots could be created. (Vicaro, et, al, 2020).

A numerous-line plot makes it easier to distinguish among information so that it may be examined and interpreted in light of other data. Every line plot essentially adheres to the single-line plot concept, although each one has a different screen presentation. Each data line plot can be customized by altering its color, line type, size, or all of them, and it can be read using a scale.