**ANALYSIS ON VIDEO GAMES SALES IN R**

**Data Description:**

This dataset contains a list of video games with sales greater than 100000 copies. The attributes in this data includes:

* *Rank:* Ranking of overall sales
* *Name:* The games names
* *Platform:* Platform of the games release (i.e., PC, PS4 etc.)
* *Year:* Year of the game’s release
* *Genre:* Genre of the game
* *Publisher:* Publisher of the game
* *NA\_Sales:* Sales in North America (in millions)
* *EU\_Sales:* Sales in Europe (in millions)
* *JP\_Sales:* Sales in Japan (in millions)
* *Other\_Sales:* Sales in the rest of the world (in millions)
* *Global\_Sales:* Total worldwide sales

**Data Analysis:**

* For analyzing the data, we are taking R as the platform. Initially the csv file is to be loaded. Read this file into a variable named ‘vg’.
* The column names are displayed and the dimension of the data is obtained by using their respective commands. Dimension of the dataset includes the number of rows and columns in the dataset.
* Before starting the analysis process, we need to import some packages using ‘library’ function as pre-requisites.
* The summary of the data is displayed by using the ‘summary’ command.
* Now by using the ggplot, we display the video games sales distribution. This plot is with respect to the global sales.
* Some basis analysis is made on the attributes in the data by using the functions like ‘mean’, ‘median’, ‘sum’.
* The video games sales distribution is further assessed by changing the binwidth to make the analysis more clear and deep.
* The maximum of the global sales is displayed by sub setting the data using ‘subset’.
* By using the piping feature, we group year and publisher and display the sales.
* The further analysis is made using the ggplot. In ggplot, we used the geom\_boxplot on the attributes in the dataset by making some combinations as shown in the R notebook.