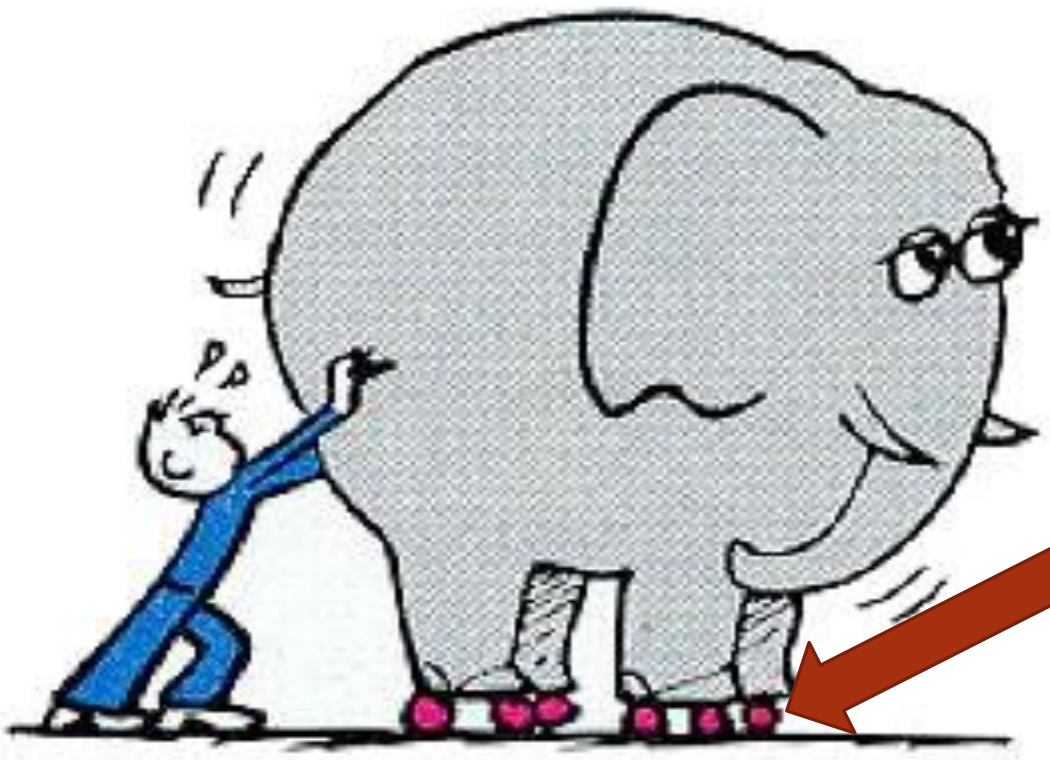
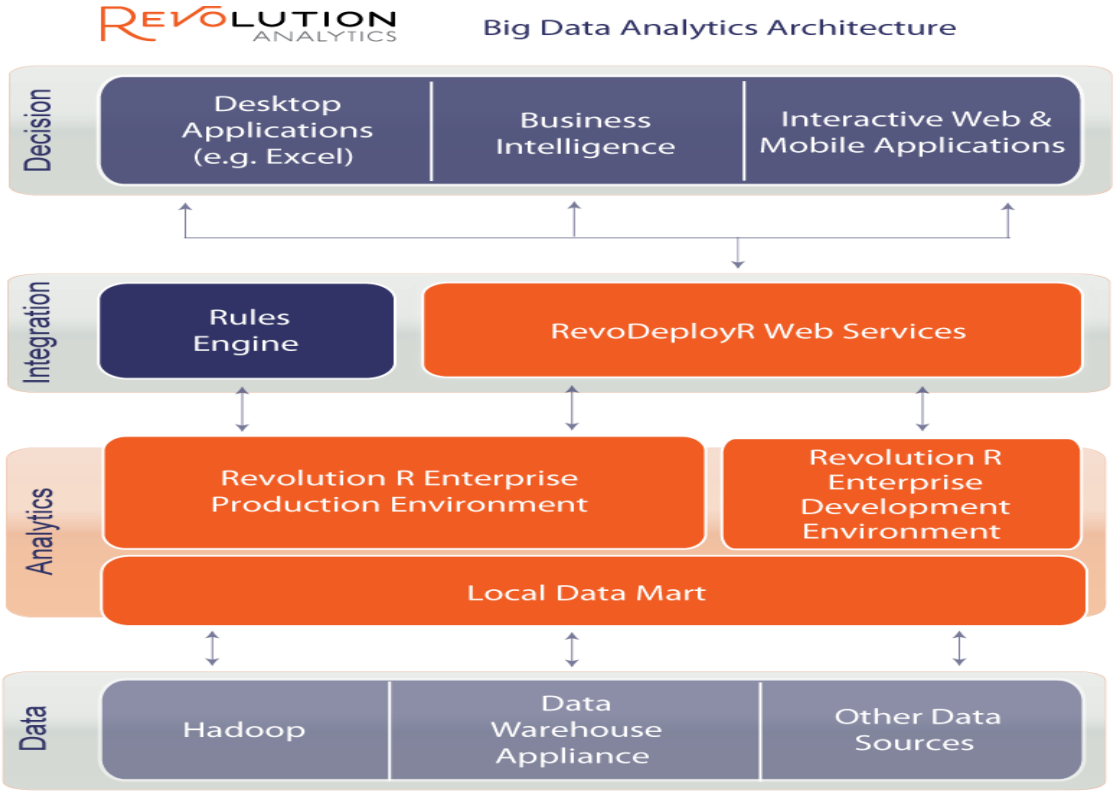
In-Depth Analysis & Representation of Airline Data (Big Data) – *Using R Language*

NIIT technologies



PROBLEM STATEMENT

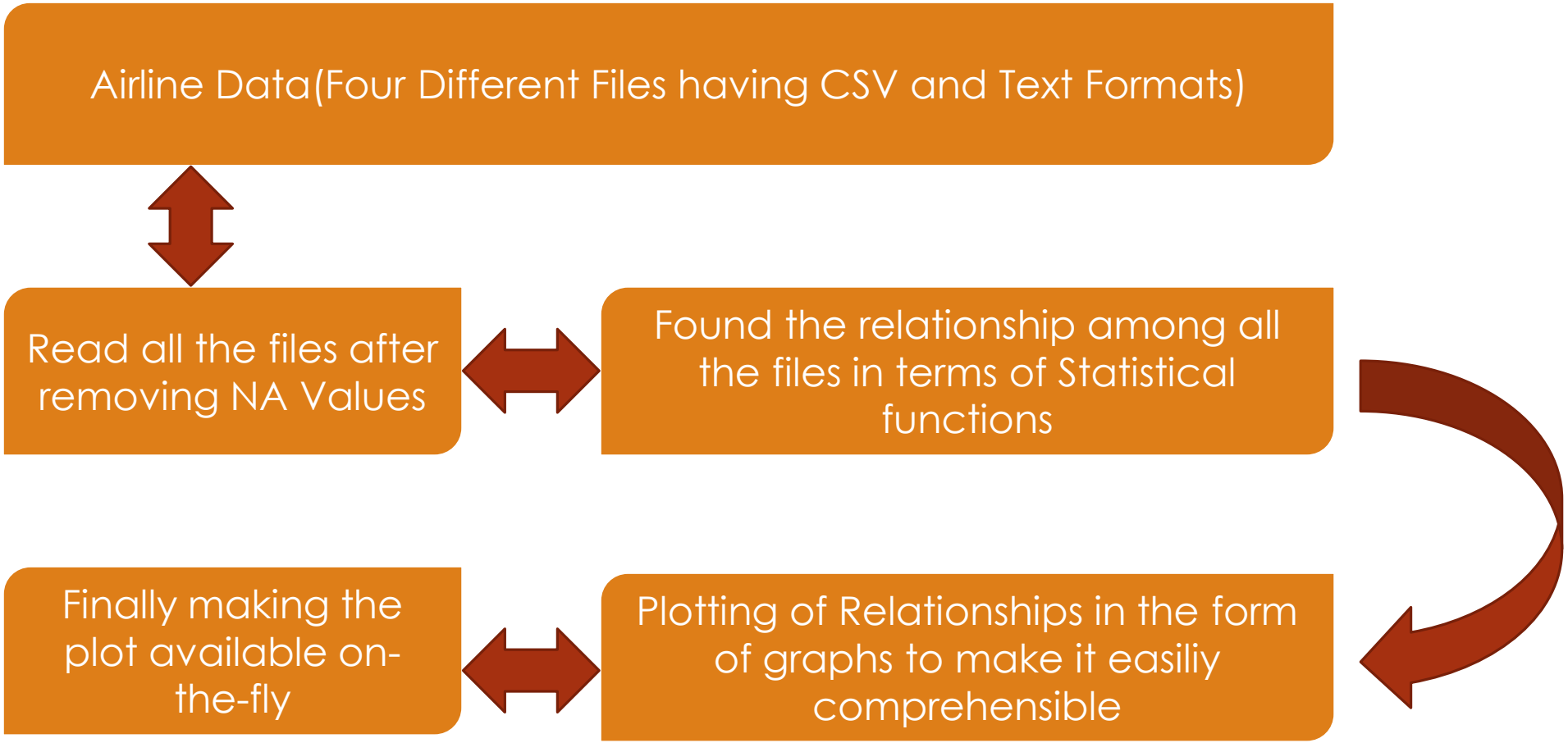
- Big data analysis using R for Airline Operations



APPROACH FOLLOWED

- Reading of different formats of data (e.g. CSV, Text).
- Statistical analysis.
- Plotting of data using wide-range of plotting schemes available (Scatter, Box, Perspective).
- On-the-fly availability of plots (Using Plotly to create 3D plots and making it available on internet).

APPROACH FOLLOWED



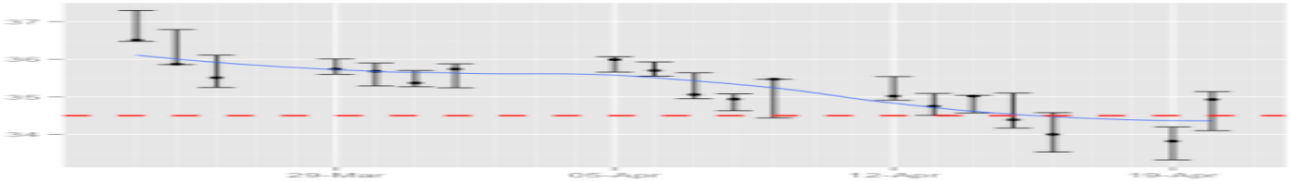
PLATFORMS & ADD-ON'S USED

NIIT technologies

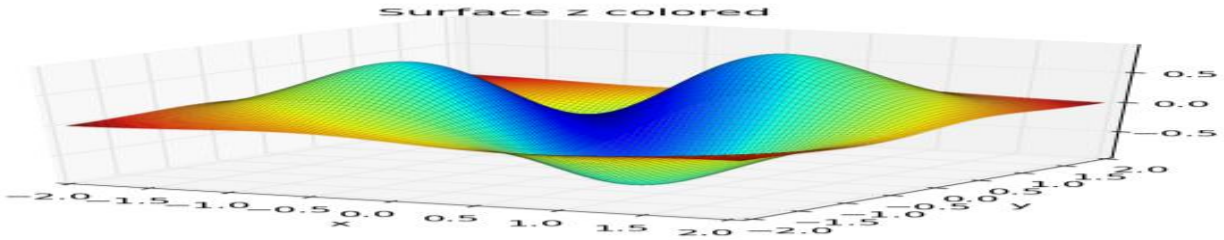
- R Language as statistical analysis tool.

REVOLUTION ANALYTICS

- GGPLOT2 Library to plot data along with error bars.



- Plotly (d3.js based library) For R used to make highly-effective and on-the-fly modifiable plots.



CODING TECHNIQUE USED

- Separation of Concern.
  - R files containing ready-to-use functions.

ReadOperations.R File

This file contains functions for reading Text and CSV files with and without NA Values as well as to read columns quarterly and fully.

CODING TECHNIQUE USED

- Separation of Concern.
  - R files containing ready-to-use console commands to see the output shown.

ConsoleCommandsForBasicOperations.R File	This file contains console commands for reading files, quarterly data, and columns (containing NULL values also).
StatisticsAndPlotConsoleCommands.R File	This file contains console commands of all the Objective Solutions which can be directly used by just changing the file path as per requirement.

SOLUTION STEPS DIVIDED INTO OBJECTIVES

- Plotting Velocity Passenger Flow and Overall Traffic Flow With Respect To Months depending on the factor (Quarter) and thus finding out the peak flow month.
- Plotting of Quarterly Standard Deviation of Passengers using VA Flights with respect to mean (Diagrammatic Prediction and Representation of deviation from normal behavior). Plotting Method Used:-Scatter Plot.
- Plotting Loyalty Traffic and General Traffic based on factor (here quarters) with respect to months and thus visually depicting the peak loyalty member and general member flows.

SOLUTION STEPS DIVIDED INTO OBJECTIVES

- Plotting of Quarterly Standard Deviation Of Loyalty Members of Virgin Australia with respect to mean using Box Plot and analyzing the Outlier behavior of the plot.
- Out of the Loyalty Members of Virgin Australia plotting loyalty benefit availing and non-benefit availing loyalty members with respect to months and analyzing the trend.
- Plotting of Quarterly Standard Deviation of Loyalty Benefit Availing Customers with respect to mean with factor being Quarter using MATPLOT.

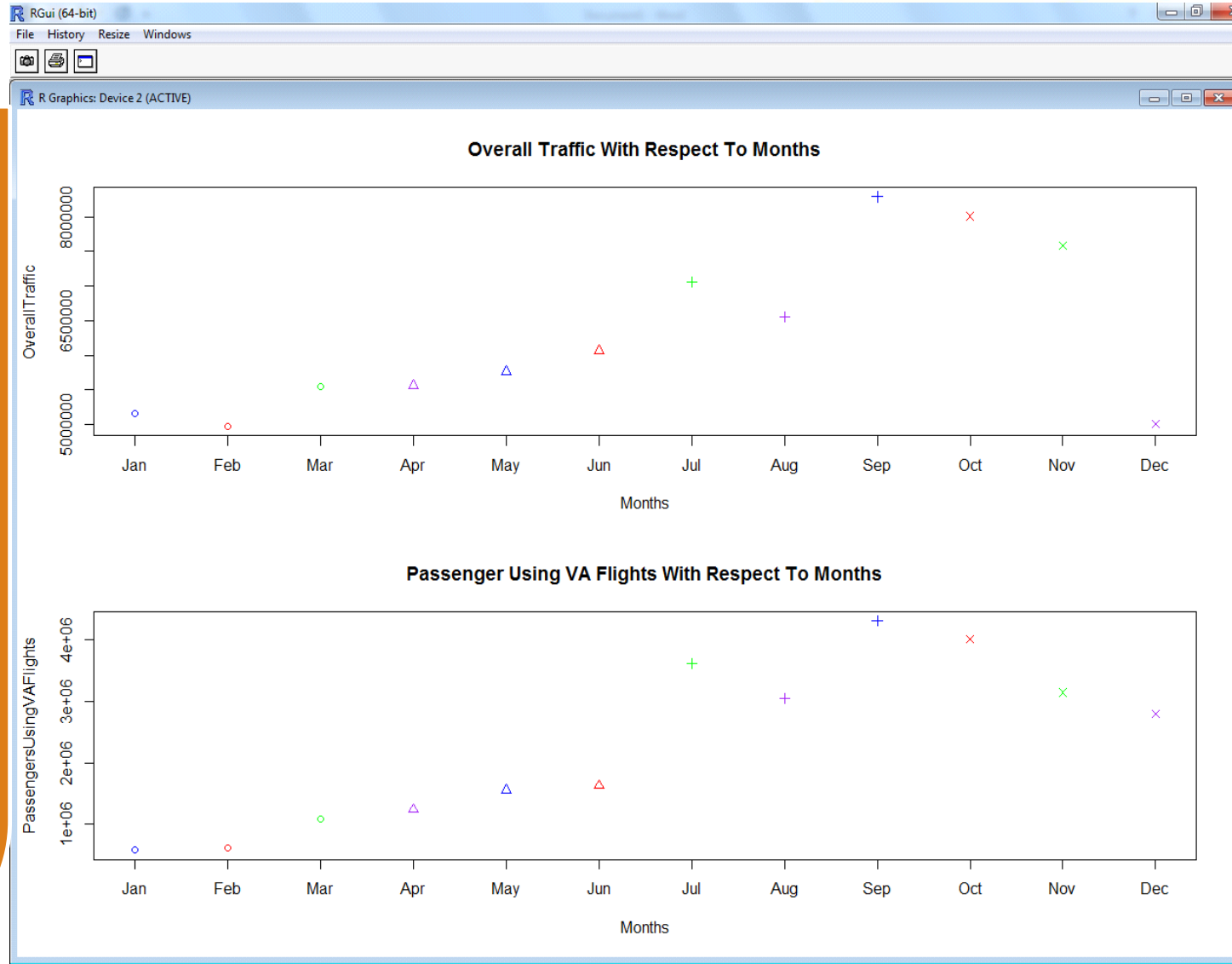
SOLUTION STEPS DIVIDED INTO OBJECTIVES

- Plotting a 3D view of Revenue Generated and Expenditure with Loyalty Revenue as depth element in order to determine how loyalty revenue has played a role in the overall profit.
- Plotting of Standard Deviation of Loyalty Revenue from Mean with Quarter as a factor using Plotly (d3.js based library package).
- Plotting of Correlation between Quarterly Expenditure and Quarterly Profit using Plotly Library.

OBJECTIVE #1

COMMANDS USED

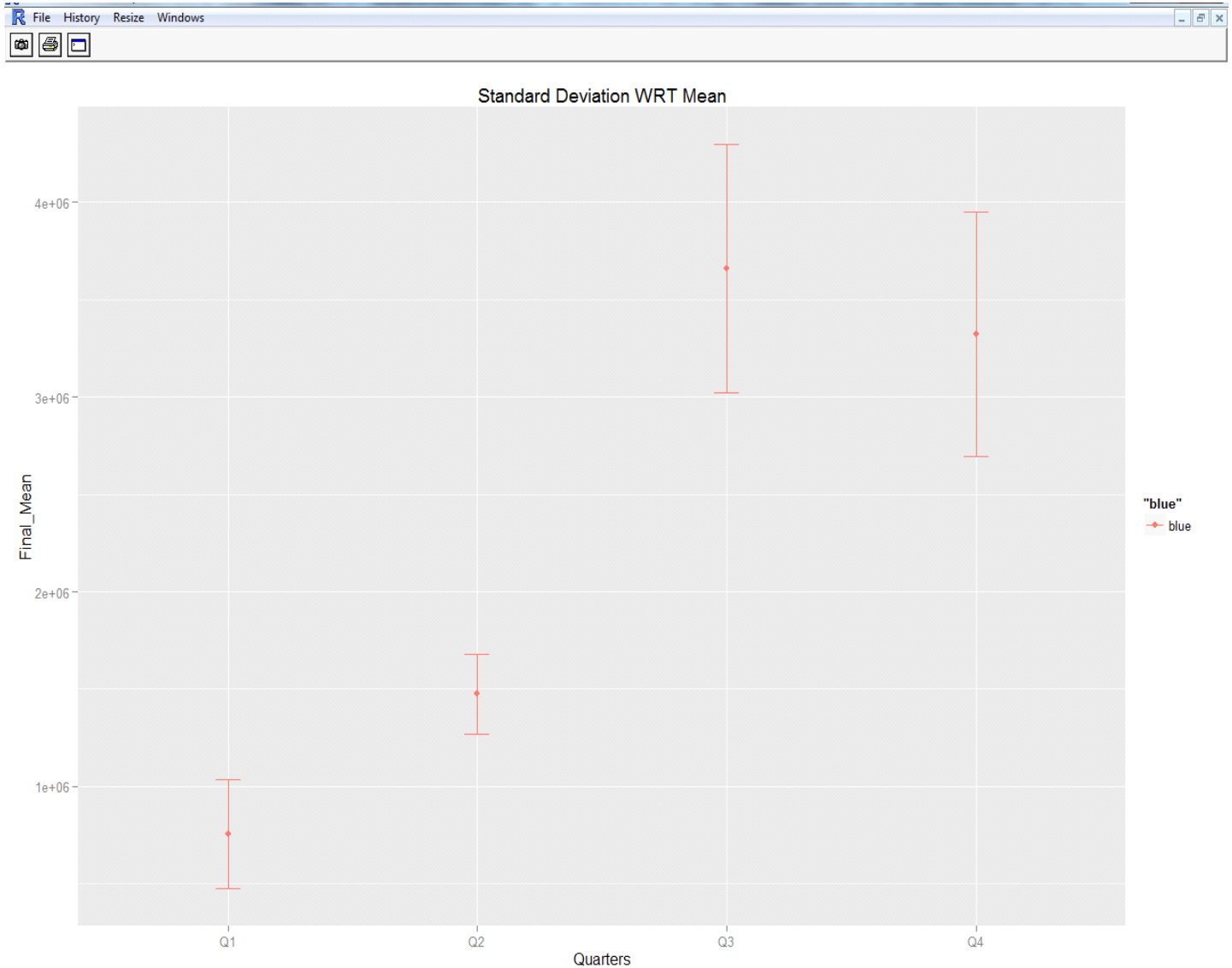
- Source (R command)
- ReadTextFile() (User-Defined Function)
- ReadGeneralColumn() (User-Defined Function)
- Par() (R's library function for multi-dim plotting)
- Plot() (R's library function for scatter plot)
- Axis() (R's library function for defining the axes of plot)



OBJECTIVE #2

COMMANDS USED

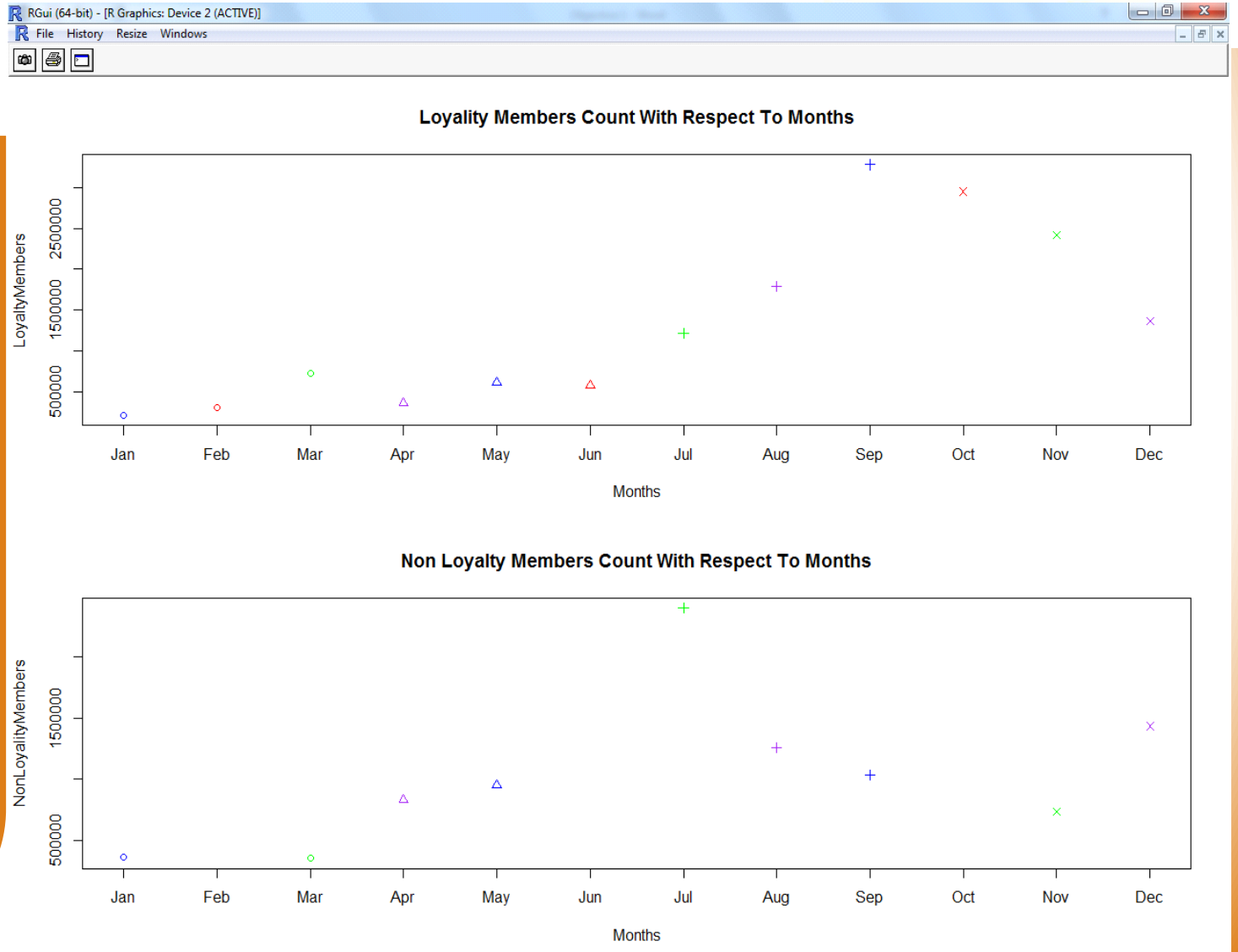
- Source (R command)
- ReadTextFile() (User-Defined Function)
- ReadQuarterColumn() (User-Defined Function)
- Sd() (R function to compute Standard Deviation)
- Mean() (R function to compute Mean)
- Library(ggplot2) (R function)
- Ggplot() (Ggplot library function)



OBJECTIVE #3

COMMANDS USED

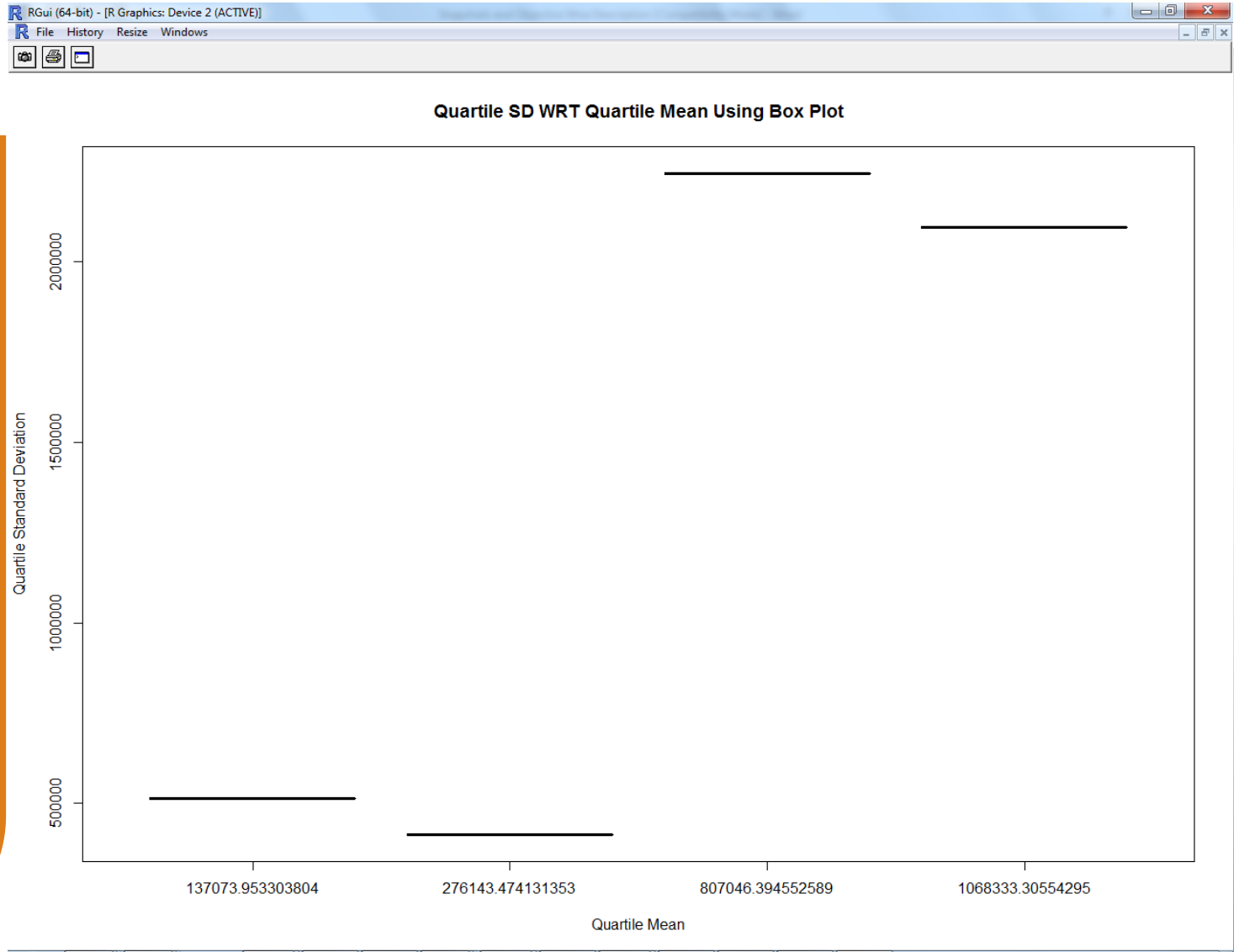
- Source (R command)
- ReadCSVFile() (User-Defined Function)
- ReadGeneralColumn() (User-Defined Function)
- Par() (R's library function for multi-dim plotting)
- Plot() (R's library function for scatter plot)
- Axis() (R's library function for defining the axes of plot)



OBJECTIVE #4

COMMANDS USED

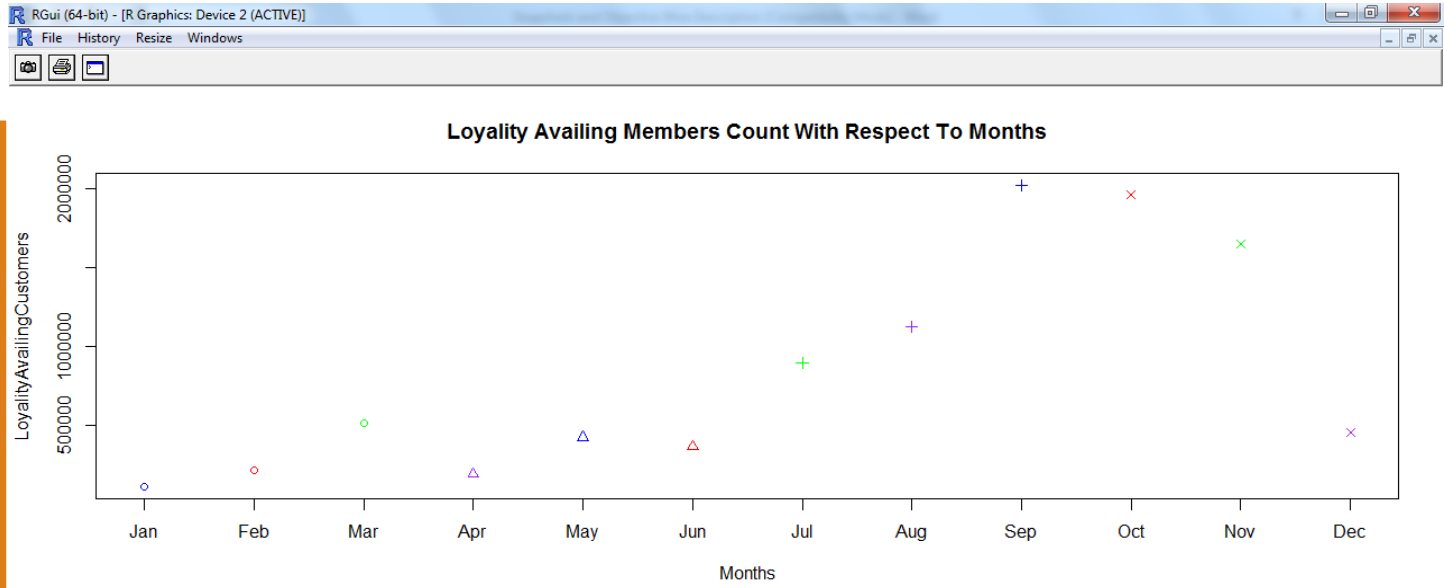
- Source (R command)
- ReadCSVFile() (User-Defined Function)
- ReadQuarterColumn() (User-Defined Function)
- Sd() (R function to compute Standard Deviation)
- Mean() (R function to compute Mean)
- Data.frame() (R function)
- Boxplot() (R function)

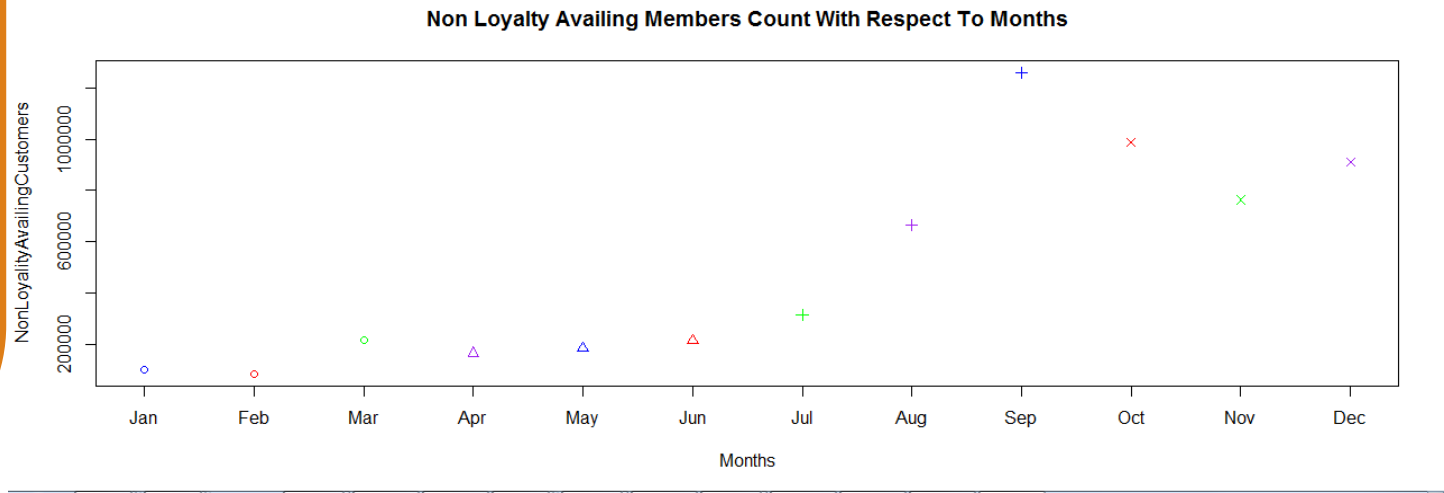


OBJECTIVE #5

COMMANDS USED

- Source (R command)
- ReadTextFile() (User-Defined Function)
- ReadGeneralColumn() (User-Defined Function)
- Par() (R's library function for multi-dim plotting)
- Plot() (R's library function for scatter plot)
- Axis() (R's library function for defining the axes of plot)

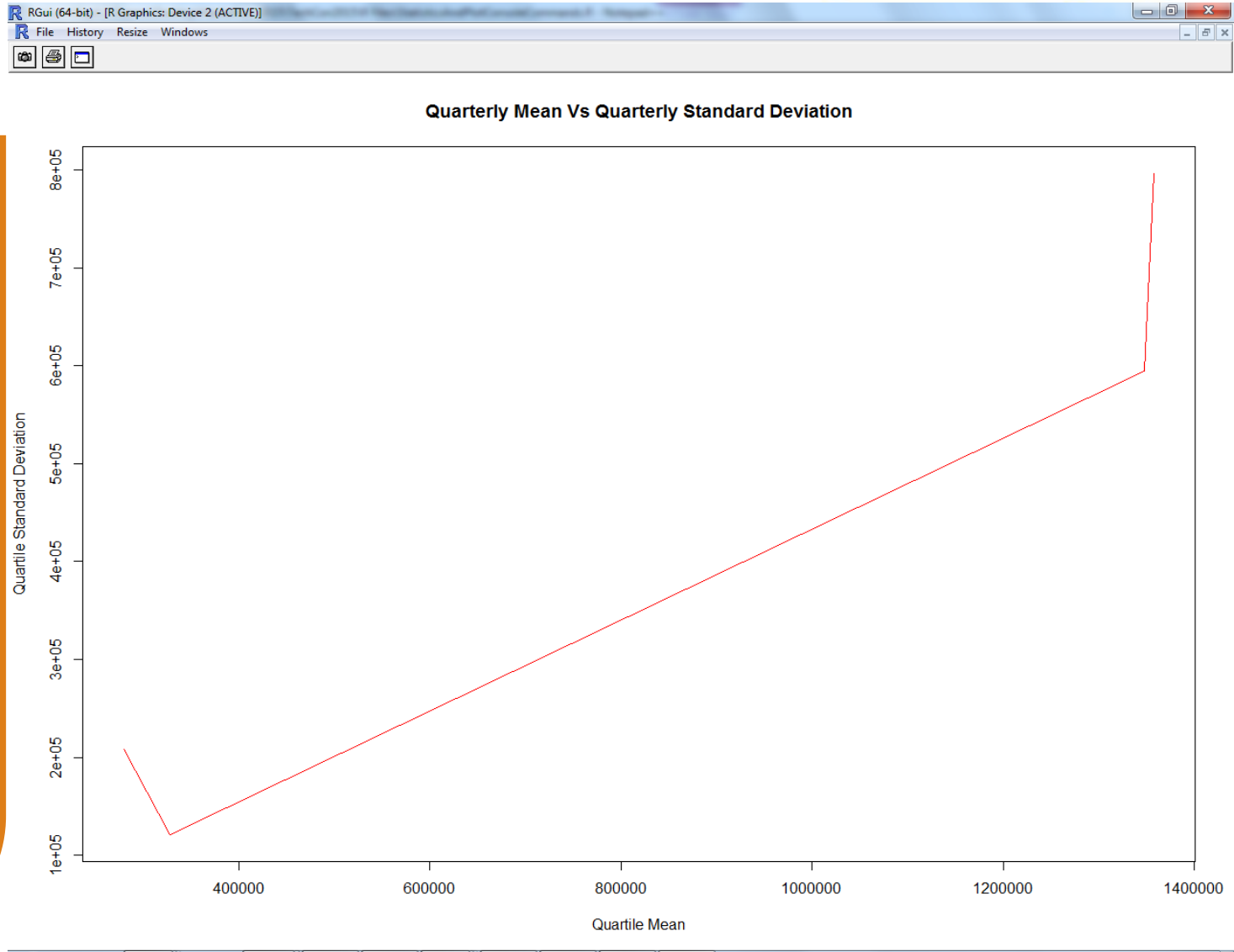




OBJECTIVE #6

COMMANDS USED

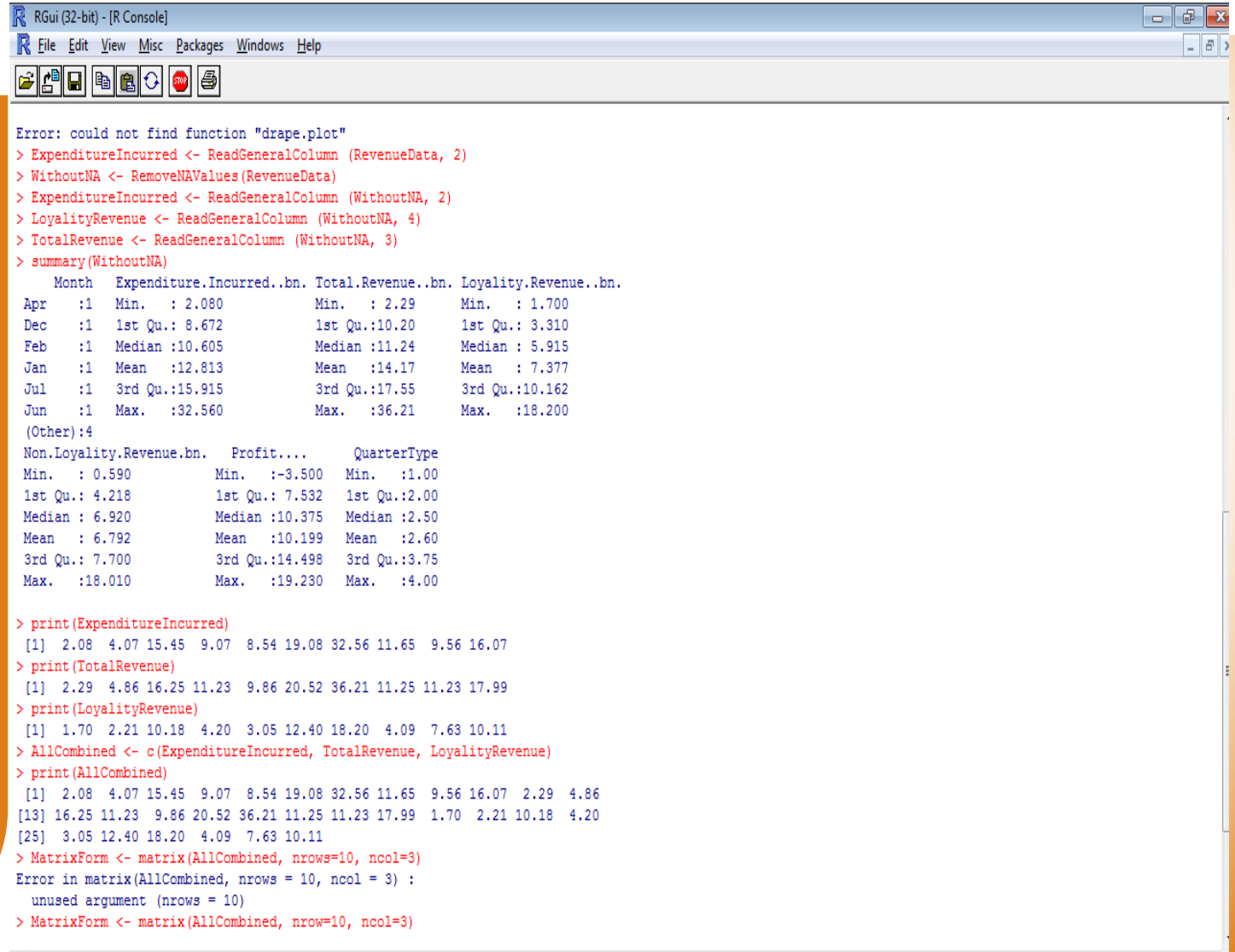
- Source (R command)
- ReadTextFile() (User-Defined Function)
- ReadQuarterColumn() (User-Defined Function)
- Sd() (R function to compute Standard Deviation)
- Mean() (R function to compute Mean)
- matplot() (R function)



OBJECTIVE #7.1

COMMANDS USED

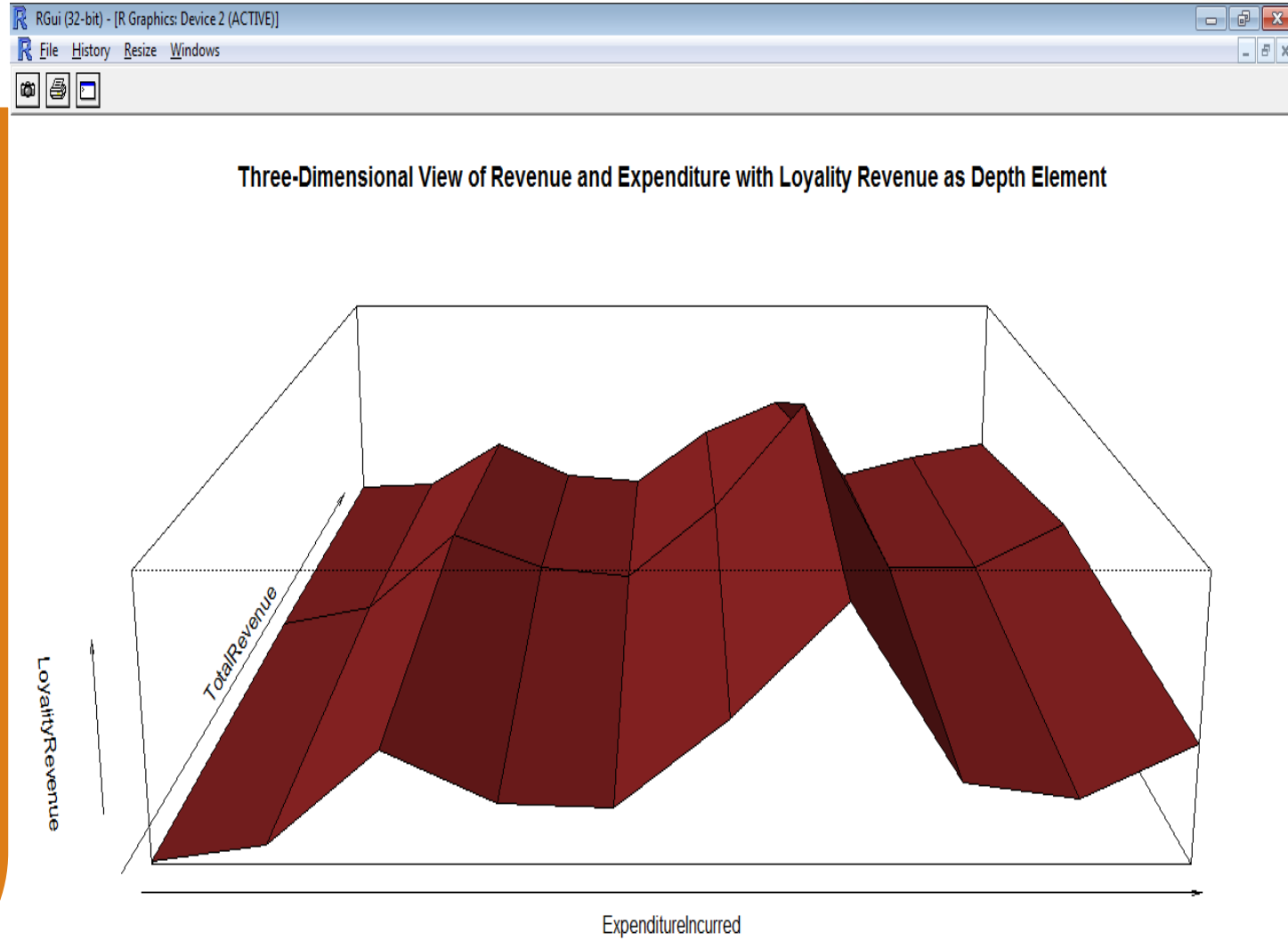
- Source (R command)
- ReadCSVFile() (User-Defined Function)
- Summary() (R function to print detailed statistical values of any data frame or vector)



OBJECTIVE #7.2

COMMANDS USED

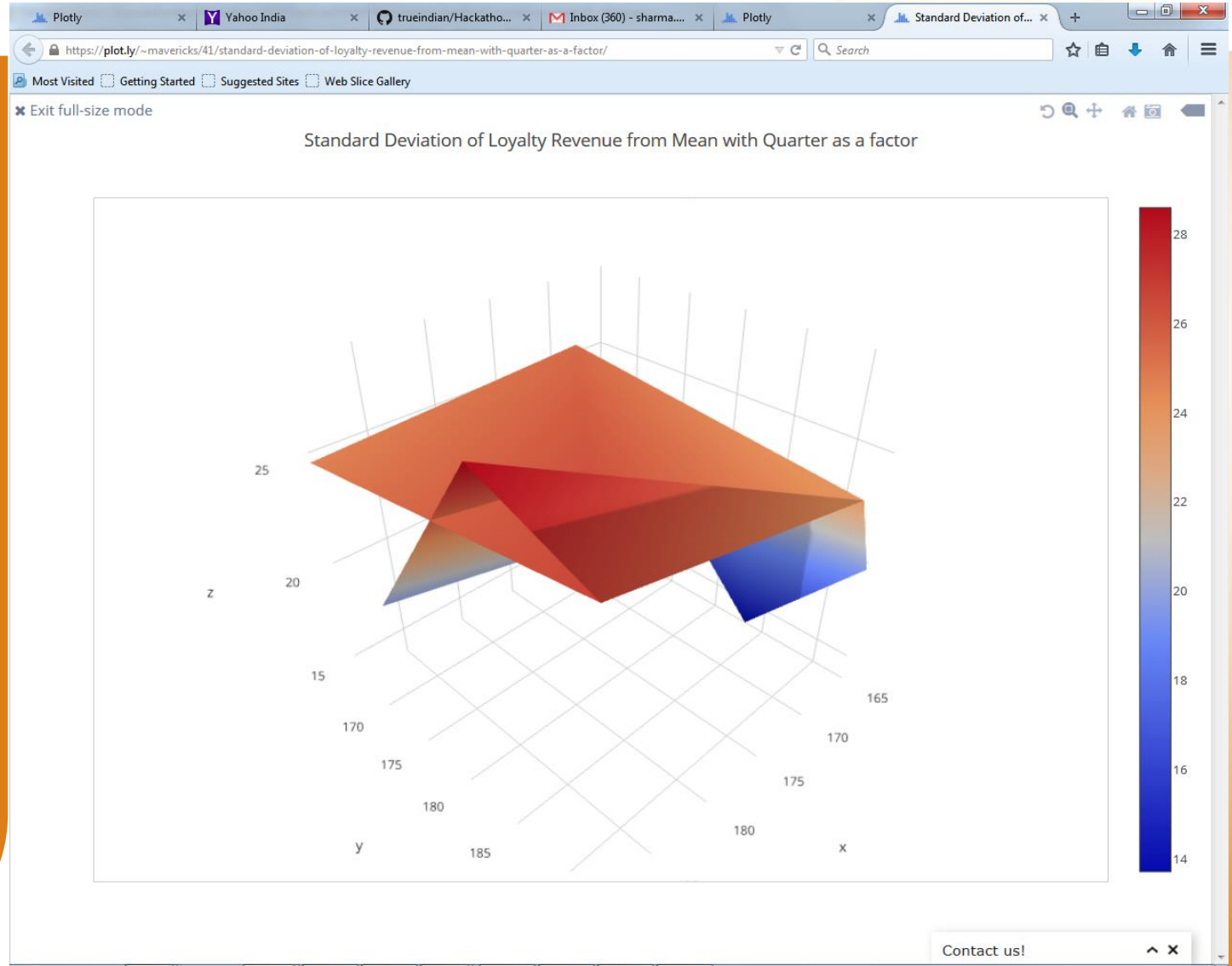
- Source (R command)
- ReadGeneralColumn() (User-Defined Function)
- Matrix() (R function)
- Persp() (R function for 3D plot)



OBJECTIVE #8

COMMANDS USED

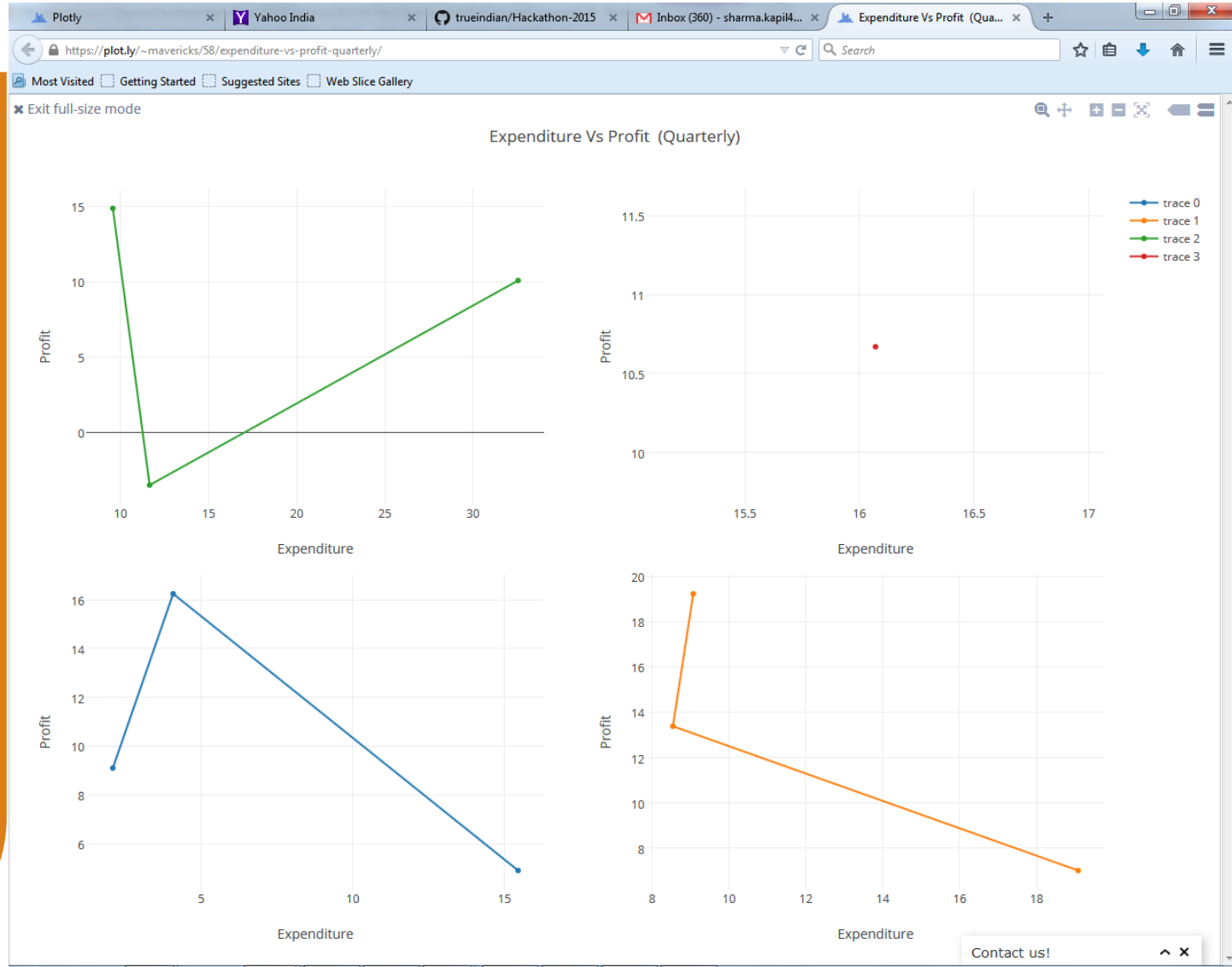
- install.packages("devtools") (R function to install Plotly Packages)
- devtools::install_github("ropensci/plotly") (R function to register on Github)
- library(plotly) (R function to get Library plotly)
- set_credentials_file("mavericks", "azaijets3m") (R function to set our credentials on Plotly)
- List() (R function to make a list)
- Plotly() (R function to initialize plot)
- Response() (R function to send response to Plotly)
- Url() (R function to get hosted URL of our plot)



OBJECTIVE #9

COMMANDS USED

- Source (R command)
- ReadCSVFile() (User-Defined Function)
- RemoveNAValues() (User-Defined Fn)
- ReadQuarterColumn() (User-Defined Function)
- library(plotly) (R function to get Library plotly)
- plotly(username="mavericks", key="azaijets3m")
- List() (R function to make a list)
- Plotly() (R function to initialize plot)
- Response() (R function to send response to Plotly)
- Url() (R function to get hosted URL of our plot)



ONLINE LINKS FOR CODE AND PLOTLY GRAPHS

- **GITHUB Link:- https://github.com/trueindian/Hackathon-2015 (All the code along with data files and R files is available at this link)**

- **Plotly Graph Links:-**
  - https://plot.ly/~mavericks/58
  - https://plot.ly/~mavericks/41


Thank you!