

**Department of DATA ANALYTICAL ENGINEERING**  
**George Mason University**

**A**  
**Mid Term Project Report on**  
**“International Football Results from 1872-2019”**

**Submitted by:**

**Vineel Vishwanth Busi**  
**G#01222602**

**Instructor:**

**Dr. Sterling Thomas**

## Table of Contents

1	About Dataset:.....	3
1.1	About the fields of Dataset:.....	3
2	Aim of the project.....	4
3	Tools used .....	4
3.1	RStudio .....	4
3.2	Tableau .....	4
4	Data Exploration in RStudio: .....	4
5	Visualizations.....	5
5.1	In Tableau .....	5
5.1.1	Country wise scores – Homescore and away score .....	5
5.1.2	Away team and Away score in form of pie chart .....	6
5.1.3	Home Team and Home Score in form of bubble chart.....	7
5.1.4	Tree map designed to depict city country and away score .....	8
5.1.5	Scatter plot to visualize away team home team away score home score.....	9
5.2	In RStudio .....	10
5.2.1	Total football matches held since 1872 .....	10
6	References.....	10

## 1 About Dataset:

This Dataset [1] includes 40495 records of international first official football matches starting from 1872 to 2019. These football tournaments are of various types right from friendly matches to FIFA world cup. All the countries are getting played in their home country and also the other country. These tournaments are Men's football tournaments.

### 1.1 About the fields of Dataset:

1. Date: This is the date information when the football match has conducted.
2. Home\_Team: This is the name of the Home team.
3. Away\_team: The name of the away team.
4. Home\_score: home team score information with not including penalty-shootouts and includes extra time.
5. Away\_score: away team score information with not including penalty-shootouts and includes extra time.
6. Tournament: The name of the tournament
7. City: The location name where the football match held
8. Country: the ame of the country where the football match played
9. Neutral: TRUE or FALSE information giving the information whether the match played at neutral venue.

## 2 Aim of the project

To analyse the football dataset and extract some visualizations so to understand the trend of the Football matches held.

## 3 Tools used

I explored the data using following tools –

### 3.1 RStudio

### 3.2 Tableau

## 4 Data Exploration in RStudio:

1. I read the dataset file in the RStudio tool.
2. The summary of the dataset is given below –

```
summary(football)
  date           home_team      away_team      home_score      away_score      tournament
Min.   :1872-11-30 Length:40945 Length:40945 Min.   : 0.000 Min.   : 0.000 Length:40945
1st Qu.:1884-02-28 Class :character Class :character 1st Qu.: 1.000 1st Qu.: 0.000 Class :character
Median :1889-02-26 Mode  :character Mode  :character Median : 1.000 Median : 1.000 Mode  :character
Mean   :1888-10-15                      Mean   : 1.745 Mean   : 1.189
3rd Qu.:1894-03-29                      3rd Qu.: 2.000 3rd Qu.: 2.000
Max.   :1899-04-08                      Max.   :31.000 Max.   :21.000
NA's   :40819

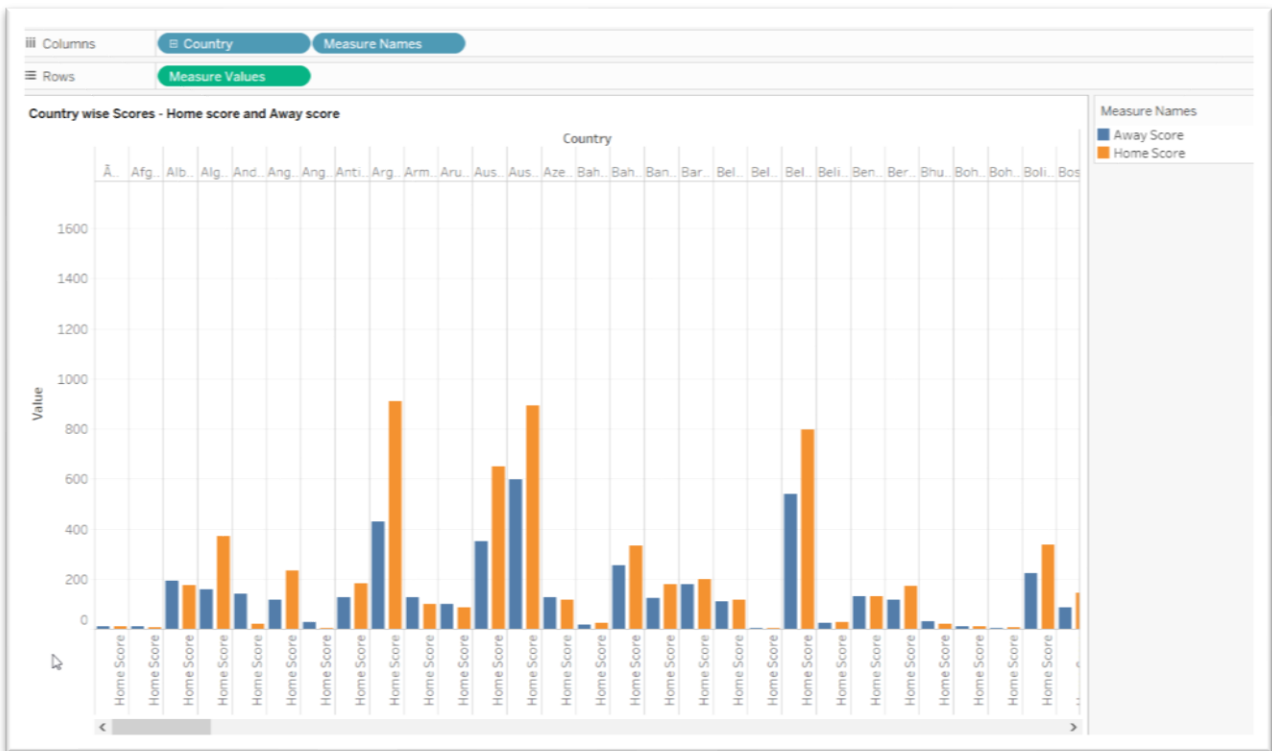
  city           country      neutral      year      winner
Length:40945 Length:40945 Mode :logical Min.   :1872 away:11577
Class :character Class :character FALSE:30756 1st Qu.:1884 draw: 9456
Mode  :character Mode  :character TRUE :10189 Median :1889 home:19912
                      Mean   :1889
                      3rd Qu.:1894
                      Max.   :1899
                      NA's   :40819
```

## 5 Visualizations

### 5.1 In Tableau

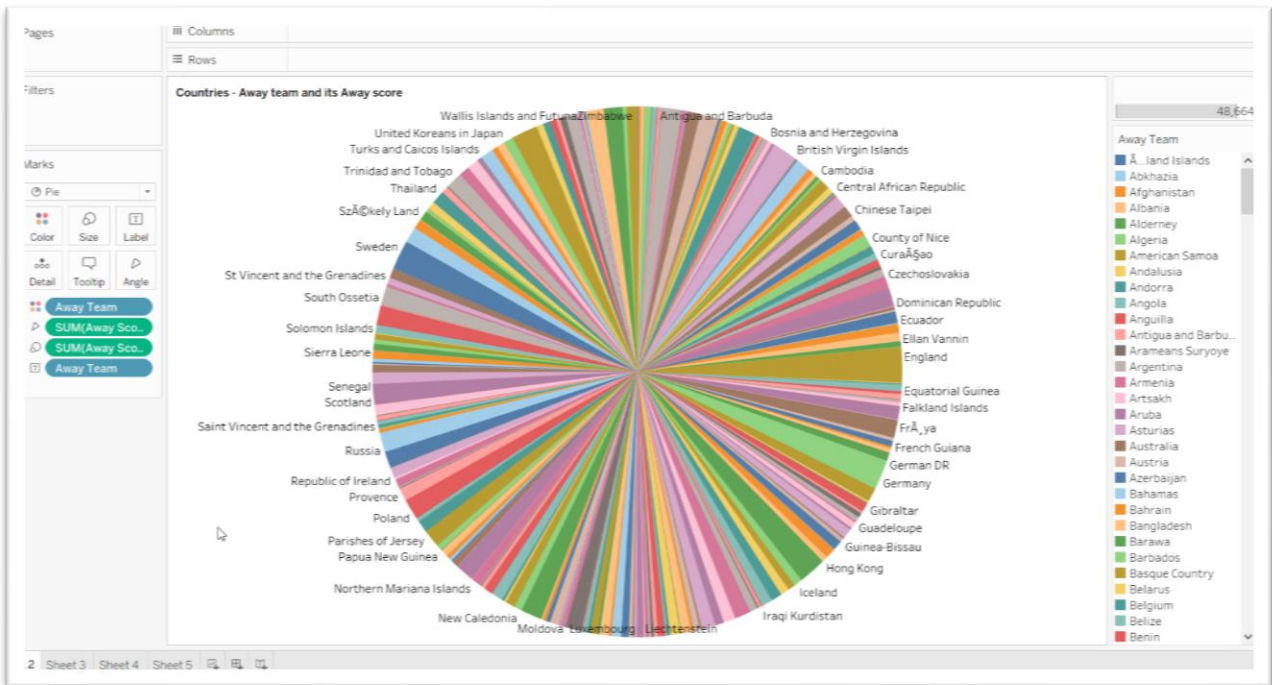
I used Tableau tool to view the Visualizations.

#### 5.1.1 Country wise scores – Homescore and away score



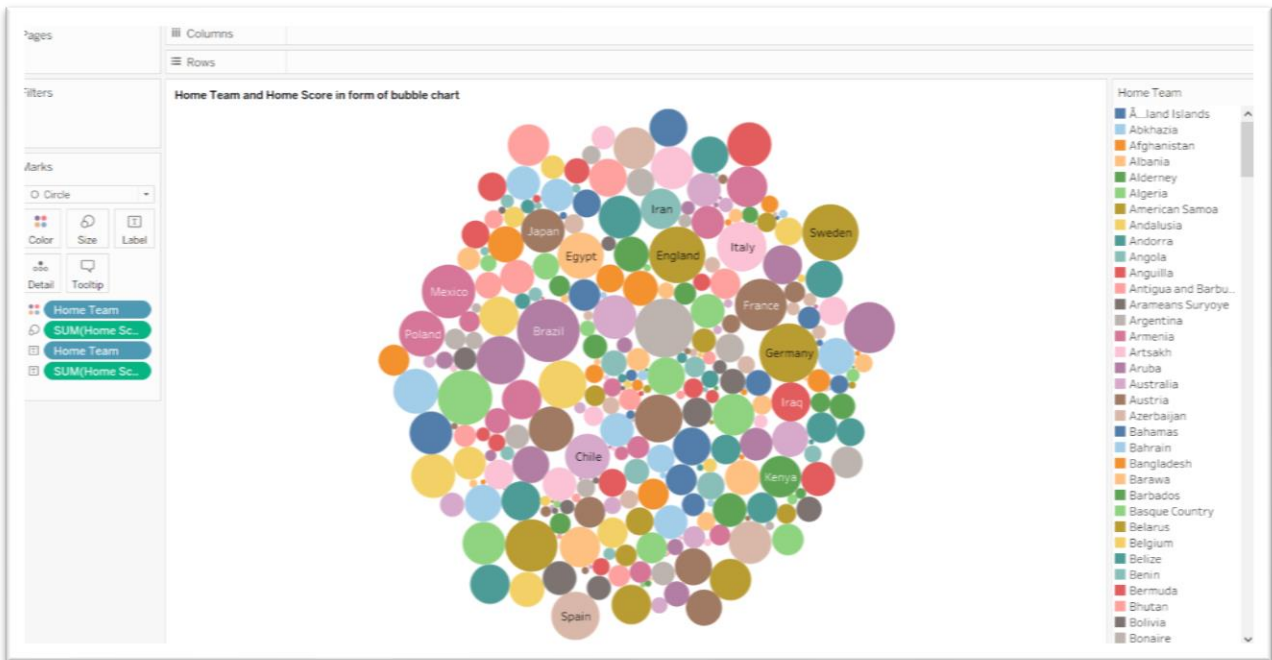
Analysis: The graph is giving the information of the countries played the football match – Home score and Away score summing it up. Mostly we could see that the Home score is huge compared to away scoring.

### 5.1.2 Away team and Away score in form of pie chart



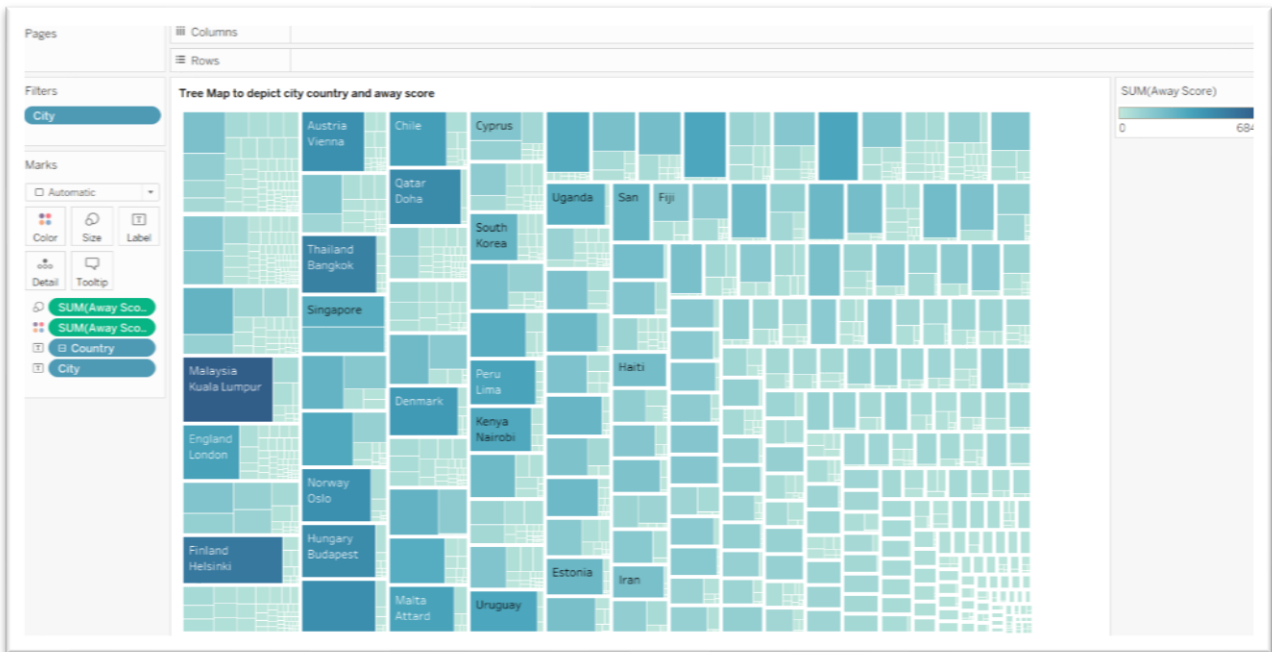
Analysis: This Pie chart gives the information of the countries that not played at the home country. On hovering the curse over the pie chart, the information of the count of the away scores are also been shown. The information is not that clear as the dataset is very huge.

### 5.1.3 Home Team and Home Score in form of bubble chart



Analysis: I depicted this graph considering the graph which country is doing good at scoring at the home location. I wanted to show it in the form of the bubble graph and from the image it is clear that the country Brazil is at the leading. Germany, England, Sweden and Mexico will come next.

#### 5.1.4 Tree map designed to depict city country and away score



Analysis: From the above figure, it is clear that Malaysia and Kuala Lumpur are in huge scores at the away location.



### 5.1.5 Scatter plot to visualize away team home team away score home score



Analysis: The above image is one of the scatter plot diagram I have used and by hovering over the image will give the counts of the scores along with the country names. It will be a bit tough to come to a conclusion with the above image.

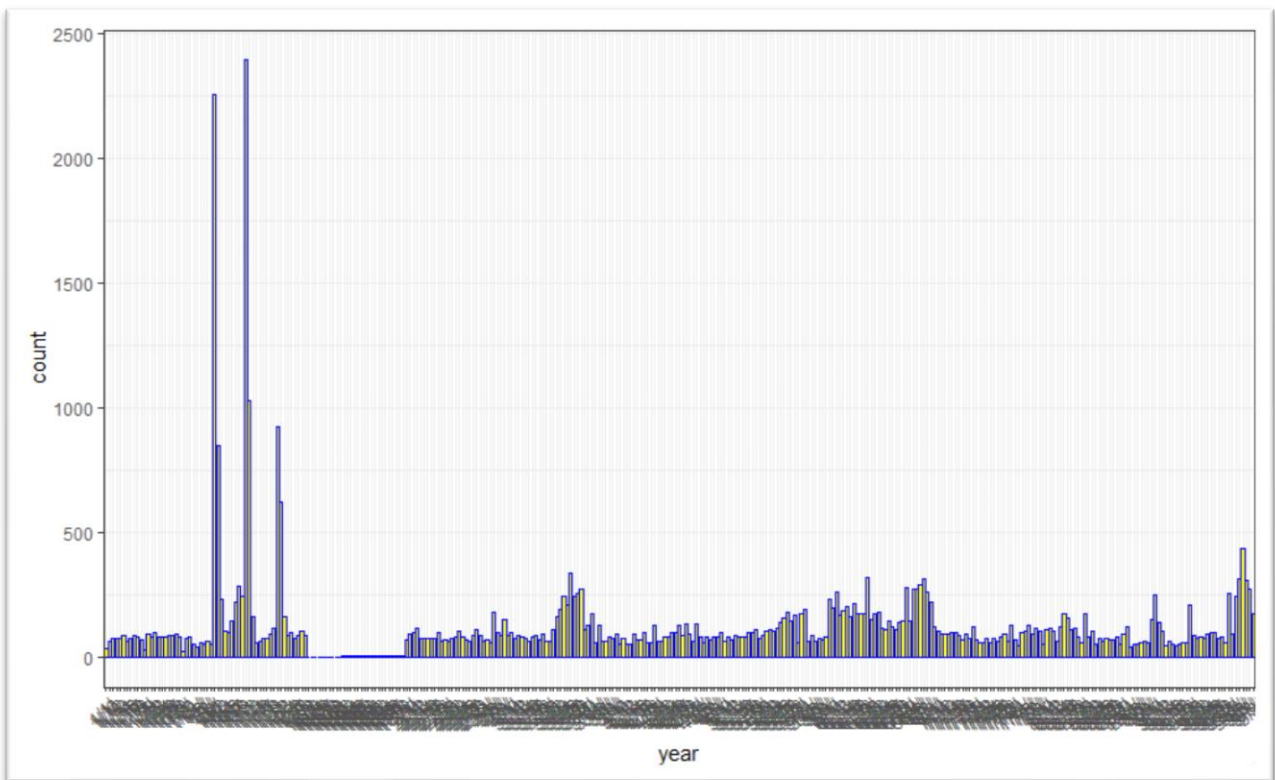
## 5.2 In RStudio

### 5.2.1 Total football matches held since 1872

Input [2] [3]:

```
ggplot(fb,aes(year)) + geom_bar(colour = 'blue', fill = 'yellow') + theme_bw() + theme(axis.text.x =  
element_text(angle = 45, hjust = 1))
```

Output:



Analysis: This graph depicts that the years on the x-axis and count of matches on the y-axis. The figure clearly explains that the matches held in the years from 1872 to 2019. The X-axis has overlapped the years.

## 6 References

- [1] "International football results from 1872 to 2019," Kaggle, 21 11 2011. [Online]. Available: <https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017>. [Accessed 26 11 2019].

- [2] R. Siemieniuk. [Online]. Available: <https://www.kaggle.com/rafalsiemieniuk/football-history-analysis>.
- [3] Gpoudel. [Online]. Available: <https://www.kaggle.com/gpoudel/150-years-of-the-beautiful-game>. [Accessed 26 11 2019].