Created by: Gurpreet Singh Master's student UCD

Analysis of Worldwide Natural Disaster

In this report i have analysed the disasters i.e Drought, Earthquake, Epidemic, Extreme temperature, Flood Landslide and Storm from 1904 to 2016. For the analysis I used Tableau to get pictorial images and with it i further generated different charts and did my analysis.

The first and foremost thing that I needed was the relevant data for natural disaster. For this i used a famous website i.e. *Gapminder* to extract data. After getting the required data I did data profiling that includes the following:

- Merging of data: As the data that i received was in various excel files, so I had to merge it in to a single file.
- Data Cleaning: As there are many countries whose data was not available, i.e sparse data thereby data cleaning becomes an important aspect.
- Normalisation: Since the data is now very much diverged and to find out the relations between the data, the normalisation is done.
- Arranging: After the merging the data in to a single sheet there are multiple data for the same country, hence arranging of the data had to be done for clear understanding.
- Conversion to CSV: It stands for 'Comma Separated Values' since it is very easy and free from macros and formatting. Hence I find it more user friendly and quick to use. For more benefits please refer *here.*

Visualisation:

For visualisation I used the software *Tableau* which has a plotting functionality, and categorises the data into Dimensions and Measures.

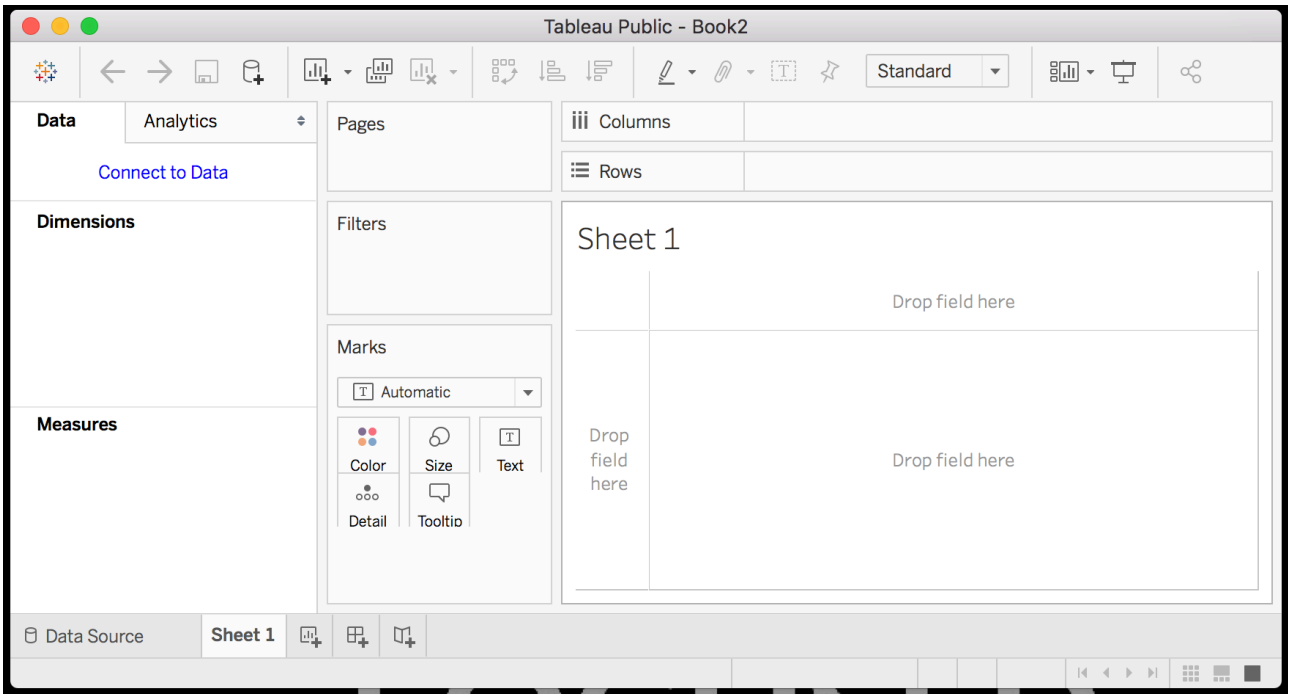

Figure 1

For plotting the data its further subdivided in to Rows and columns which a user is very user-friendly, as user has the option to decide which all fields to put in column and rows. (Figure 1).

Once data is imported the software itself analyses and segregates it in to different features of dimensions and measures. (Figure 2)

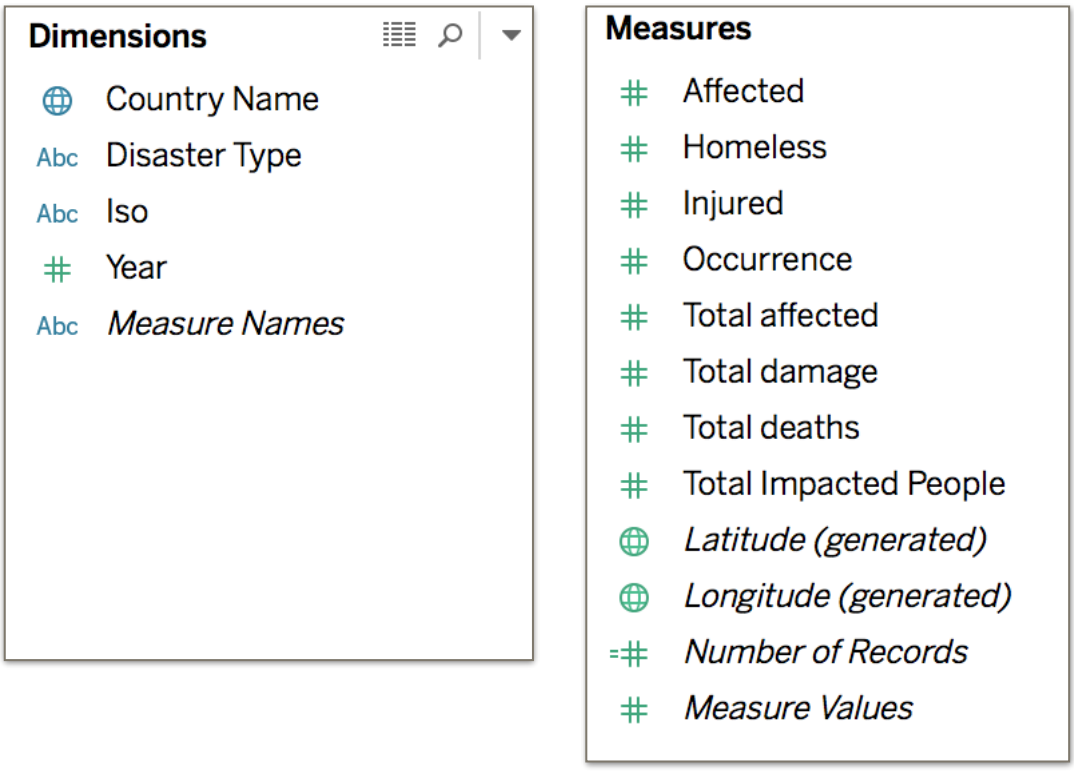

Figure 2

Then for creating the bar chart i added the Latitude and Longitude to the sheet and selected the visualisation format as bar chart. (Figure 3)

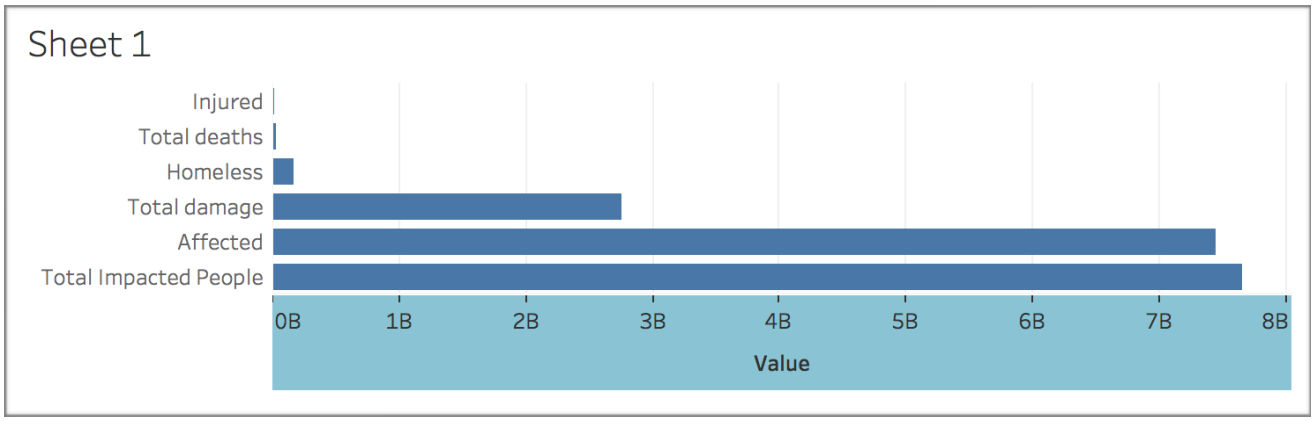

Figure 3

In x-axis the the total number of affected people, i.e. the count is present in billions (B). In y-axis the people who got injured, became homeless, people affected, impacted people, etc are plotted.

Now If we see the 2 countries which are the most populated, i.e. *China and India*. Then we can clearly see the most impacted people density wise are the most populated country. (Figure 4)

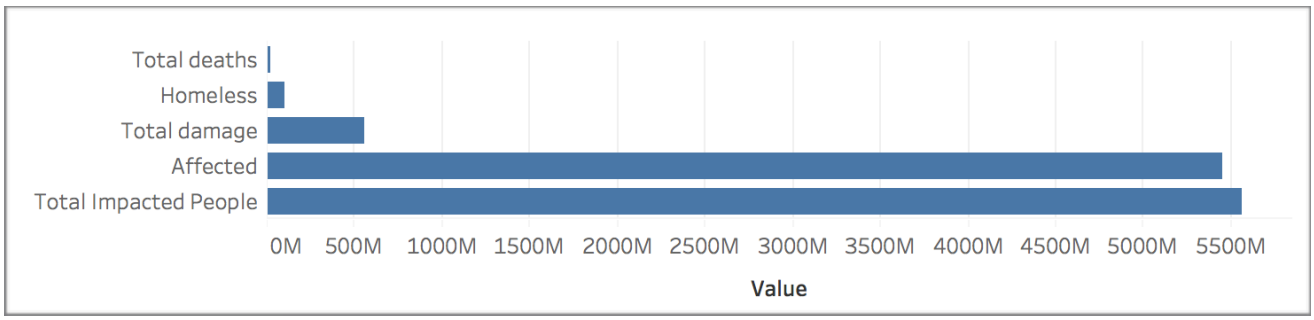

Figure 4

That is 7659578706 (Total affected people all over the world) - 5567059118(Total affected people in 2 majorly populated countries) gives us 2092519588*.

* *Since data is a sparse data, hence these are not 100% actual numbers.*

The beauty of this software is to visualise the data in a very smooth and understanding manner. As we can see below:

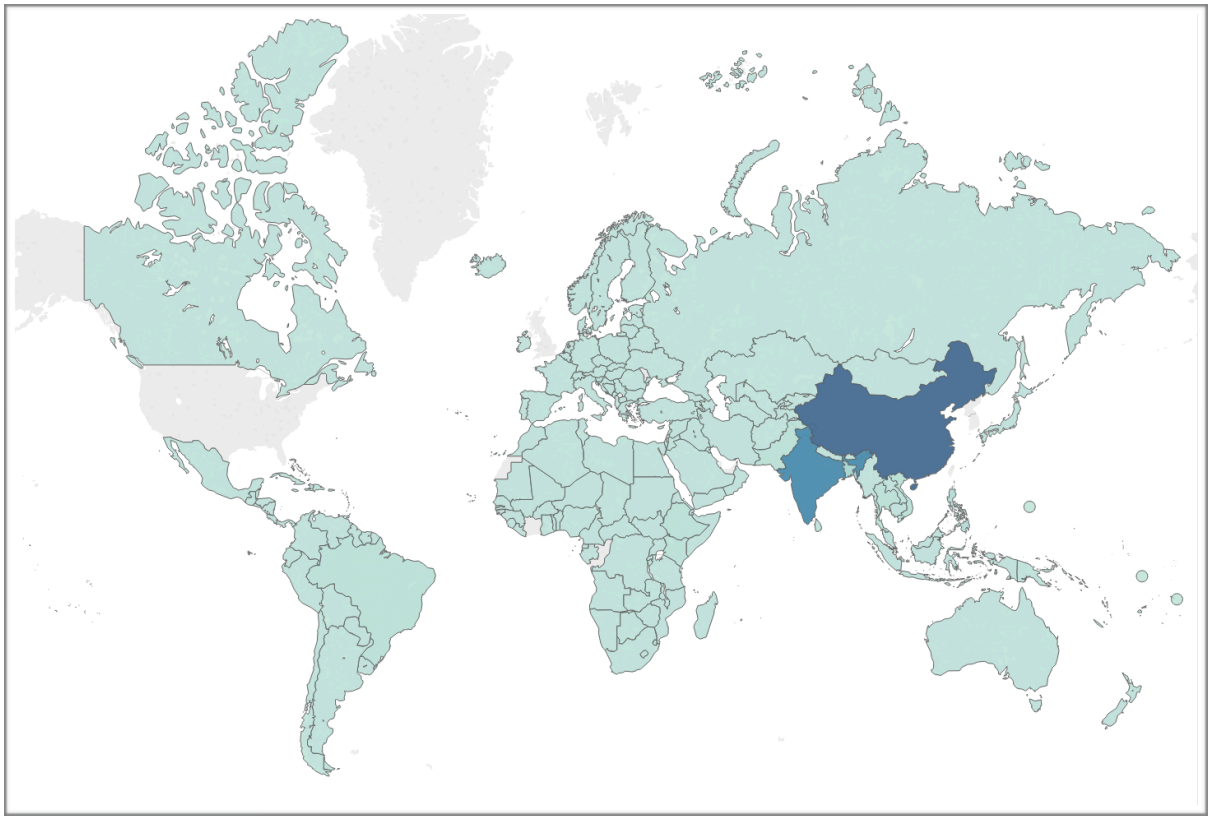

Figure 5

The figure 5 is the representation of all the disasters, but here we can clearly see the most impacted people with the help of colour depth that its China and India.

Now on further analysing i found Kenya was most affected by ‘Epidemic’ (Figure 6). But we can further see that countries like Libya, Poland, Austria, Czech republic, Slovakia, Green land, U.S.A and some part of South America are greyed out, this is so because in the data for these countries are not available.

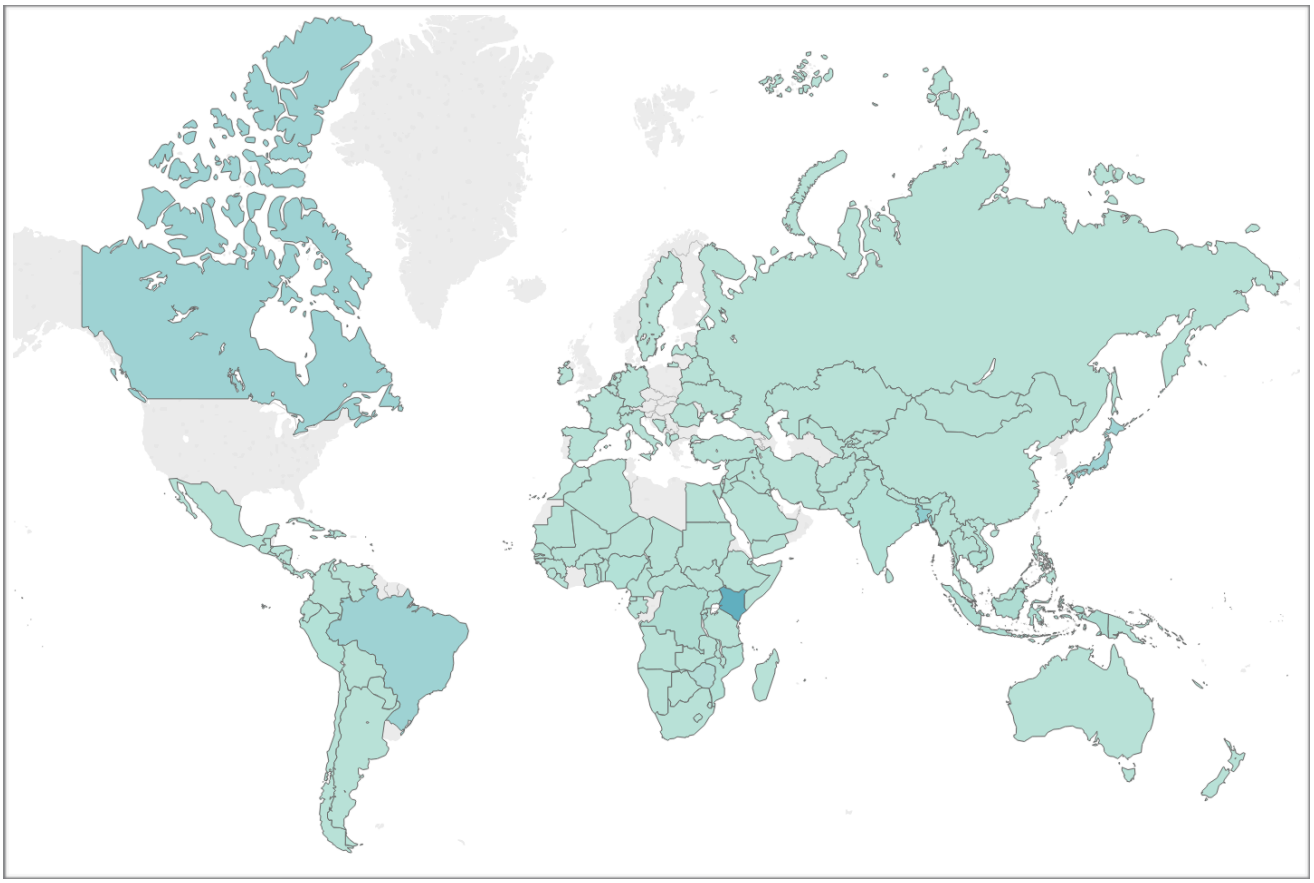

Figure 6

We can further make various types of charts such as pie chart, tree maps, packed bubbles and many more.

Below is a visualisation for occurrence of storms using Tableau. (Figure 7)

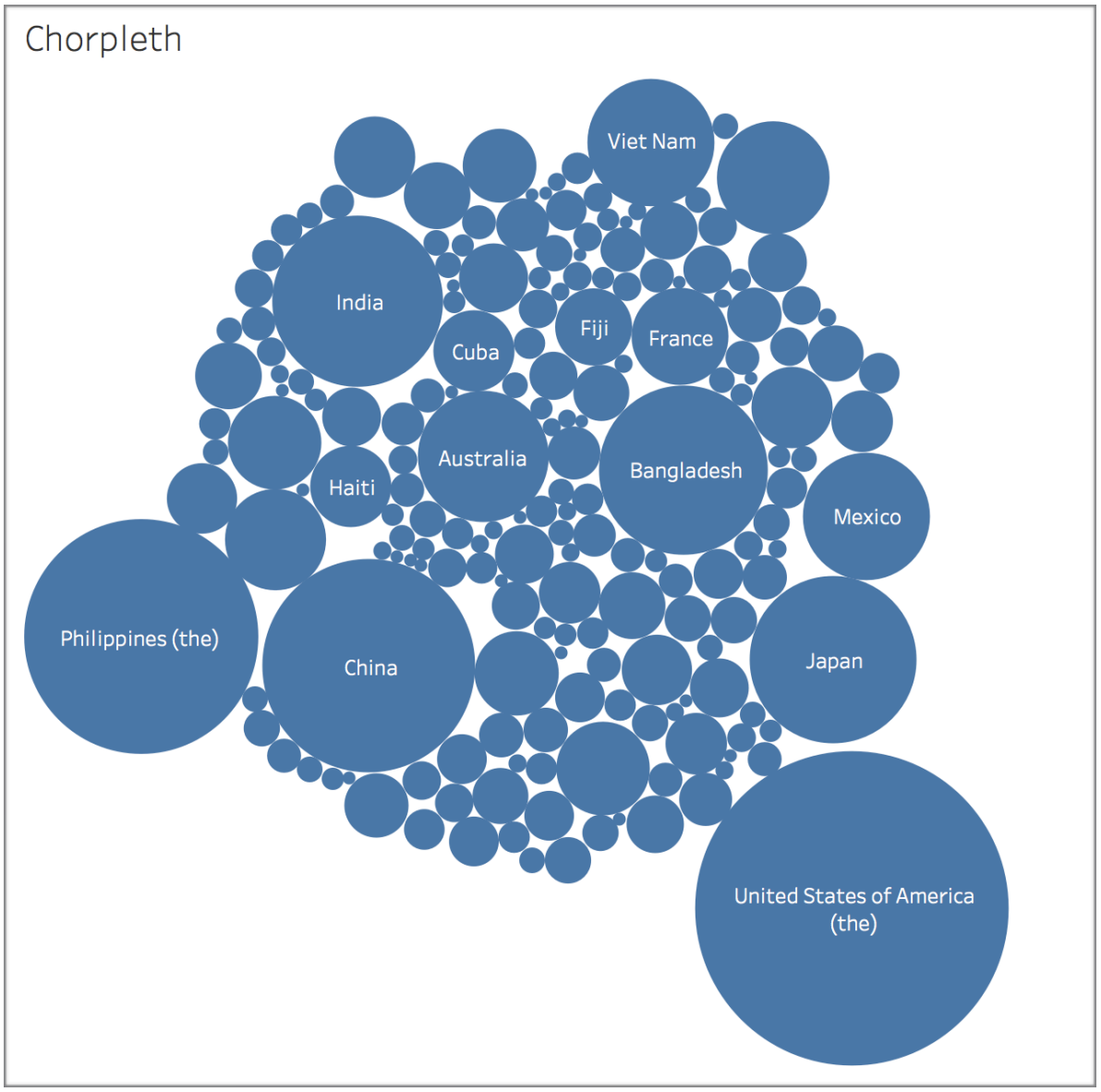
Chorpleth
Viet Nam
India
Fiji
France
Cuba
Australia
Bangladesh
Haiti
Mexico
Philippines (the)
China
Japan
United States of America (the)

Figure 7