

## Declaration on Plagiarism

*This form must be filled in and completed by the student(s) submitting an assignment*

<b>Name:</b>	Nivedita Mahato
<b>Student Number:</b>	19210398
<b>Programme:</b>	MCM1
<b>Module Code:</b>	CA682
<b>Assignment Title:</b>	Data Visualisation
<b>Submission Date:</b>	15 Dec 2019
<b>Module Coordinator:</b>	Dr Suzanne Little
<b>Tableau Link</b>	<a href="#">Tableau</a>

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found at <http://www.dcu.ie/info/regulations/plagiarism.shtml>, <https://www4.dcu.ie/students/az/plagiarism> and/or recommended in the assignment guidelines

Name: Nivedita Mahato

Date: 15 – 12 -2019

# Visualizing Road Accident in India.

## Abstract:

India is the second-most populous country (with over 1.2 billion people), there has been an enormous increase in the number of vehicles used, also the number of accidents in the past decade. World Life Expectancy conducted a study which states that road traffic accidents are amongst the top 10 reasons for deaths causing more than 2.5 lacs in India. Road accidents are affecting people's livelihood and push them towards poverty. The study also shows that poor people go into debt by borrowing money to cope with the additional medical expenses, also to losing their income after an accident. The human cost in this is enormous, and so is the impact on the economy. A World Bank study has found that if India were to halve road deaths and injuries between 2014 and 2038 successfully, it could potentially add 14 per cent to its GDP per capita. Hence, it is necessary to analyse road accident data. This paper deals with the road security of people of the nation and dwells into great depths to decipher what are the trends that are causing all the possible accidents.

## 1.Dataset:

The data for this project has been taken from data sets available on <https://data.gov.in/>, an authorised portal by the Government of India. The data is made available for research purpose for innovative uses of Government Data to give a different perspective.

The data consist of 3 CSV sheets. The data consist of 5 Million rows having time series from 1970 to 2017.

### Attributes:

The data refers to Cause of Accident, type of Vehicles, monthly death report, states in India, the Death count of Male, Death count of Female, Total Death, Total Number of Road Accidents, Persons Killed and Injured, Number of Accidents per Ten Thousand Vehicles, Number of Accidents/ Lakh Population, Number of Persons Killed / Ten Thousand Vehicles, Number of Persons Killed / Lakh Population, Number of Persons Killed/ Ten Thousand Kms of Roads, Number of Persons Injured / Lakh Population. Transport Research Wing (TRW) of Ministry of Road Transport & Highways collected the Data of road accidents in India.

### Datatype:

The different data types present are:

1. temporal: Year.
2. Spatial: State in India
3. Ordinal: Cause of Accident
4. Nominal: monthly death report, states in India, the Death count of Male, Death count of Female, Total Death, Total Number of Road Accidents, Persons Killed and Injured
5. Ratio: Number of Accidents/ Lakh Population, Number of Persons Killed / Ten Thousand Vehicles, Number of Persons Killed / Lakh Population, Number of Persons Killed/ Ten Thousand Kms of Roads, Number of Persons Injured / Lakh Population.

## 2.Data Exploration, Processing, Cleaning and/or Integration:

### Data Exploration:

Tableau was used to explore the data obtained from data.gov.in website and discovered various relations which can be chosen to visualize the data like Number of killed vs Number of inquired in an accident over the period of time. Number of State Vs Number of Accident,

Population Vs Accident over the period of Time and many more making the data a rich source for visualizers.

Abc Accident data.csv CAUSE	# Accident data.... Female	# Accident data.csv Male	# road accident - killed and injur... Number of Accide...	# road accident - killed and injur... Number of Accide...	# road accident - killed and injur... Number of Accide...	# road accident - killed and injur... Number of Person...	# road accident - killed and injur... Number of Person...
Truck/Lorry	26	103	39.40000	1,202.400	73.8000	39.40000	1,201.200
Truck/Lorry	414	2,171	39.40000	1,202.400	73.8000	39.40000	1,201.200
Bus	119	488	39.40000	1,202.400	73.8000	39.40000	1,201.200
Bus	76	289	39.40000	1,202.400	73.8000	39.40000	1,201.200
Tempo/Vans	3	13	39.40000	1,202.400	73.8000	39.40000	1,201.200
Tempo/Vans	70	310	39.40000	1,202.400	73.8000	39.40000	1,201.200
Jeep	9	31	39.40000	1,202.400	73.8000	39.40000	1,201.200
Jeep	111	473	39.40000	1,202.400	73.8000	39.40000	1,201.200

For my visualization I am going to consider the different state and accident over the period of time and number of deaths of male and female in these states and the cause of these death.

### Data Processing and Cleaning and Integration:

For visualizing the above relation, the data needed to be processed and joined. All the 3 datasets where combined together using Google BigQuery to get the data which is required.

*select a.Year, b.State/UT,c.Cause, c.Male,c.Female,b.Total from data1 a join data2 b on a.Year = b.Year join data3 on b.Year = c.Year  
where a.Year between 2006 and 2017;*

Above Query was generated in BigQuery using join function to extract data consisting of below attributes:

STATE/UT	Year	CAUSE	Male	Female	Total Death
String	Integer	String	Integer	Integer	Integer

The data from the year 2006 -2017 was taken into consideration for our visualisation because the accident count was meagre and will not impact our analysis, the other attributes considered are the cause of death (the type of vehicles used), location of the accident (in which state the accident occurred). The data has also shown each vehicle to Male and Female death count from the year (2006-2017). The data cleaning was done using excel, all the NULL values were replaced by 0 in the data.

### 3. Visualisation:

As India has many states hence knowing which states have a greater number of accidents is crucial. It will help to analyse which state require more attention related to road safety. The considered data shows relation and different information details related to the all the 34 states in India displaying all the state in any other graph would have been difficult to understand. **Map graph** helps to analyse the data as locations on an interactive map using the map visualization. Map areas are coloured based on the value of a metric of Total Accident. Each state displays the state name and total number of accident when hovered over

it. The dark colour shows the state having a greater number of accidents to the lighter colour for a smaller number of accidents over the period 2006 – 2017.

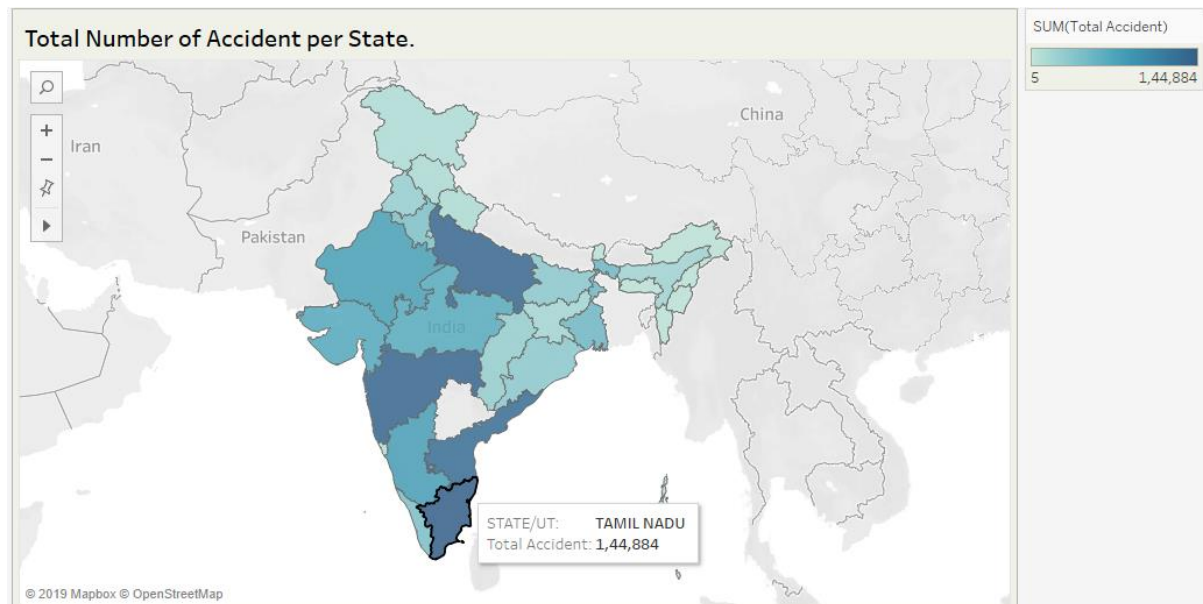


Figure 1: Number of accident per state Map graph Colour = Total Accident

From the above graph we analyzed that the State of TamilNadu a state in southern India has the most number of accident compared to other state of India. It makes us to digging more to know what is the trend of accident in the state TamilNadu over the years so I created a trend graph. The graph shows percentage difference (by creating calculated filed in tableau) of total accident over the years of graphs from previous year for a particular state in India. The below graph shows us in which year the accident was more and what is the trend of accident when compared to the previous year. Considering the benefits, **Animated Trend ine Chart** has been chosen in visualizing this accident data from the Year 2006 to 2017 which helps the government bodies to understand the trend of various accident in various state over the years and identify key trends and recognizable patterns which eventually helps in better deciding which state is showing improvement over year and which state requires more attention towards road safety measures. The below trend shows that there was hike of death during the year 2010 where as comparitvely less accident in 2006 and no major dip in trend after 2012. The red colour was choosen as it is the color of blood and graph is showing accident over the period of time.

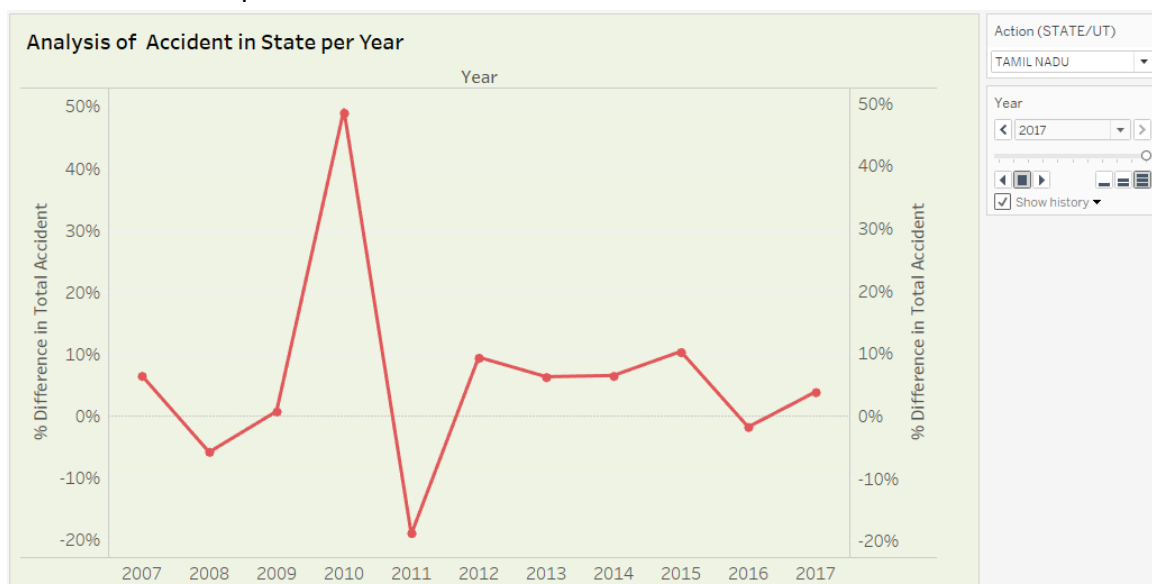


Figure 2: line graph, Colour = percentatge difference of total accident

To analyze what are Cause of these accidents(vehicle used) click on any point in the graph it will route to the other graph showing the perecent of death Male and female because of the Cause for that year. To visualize the data a **Butterfly Chart** (also called Tornado **Chart**) is a type of bar **chart** is used where two sets of data series can be displayed side by side.

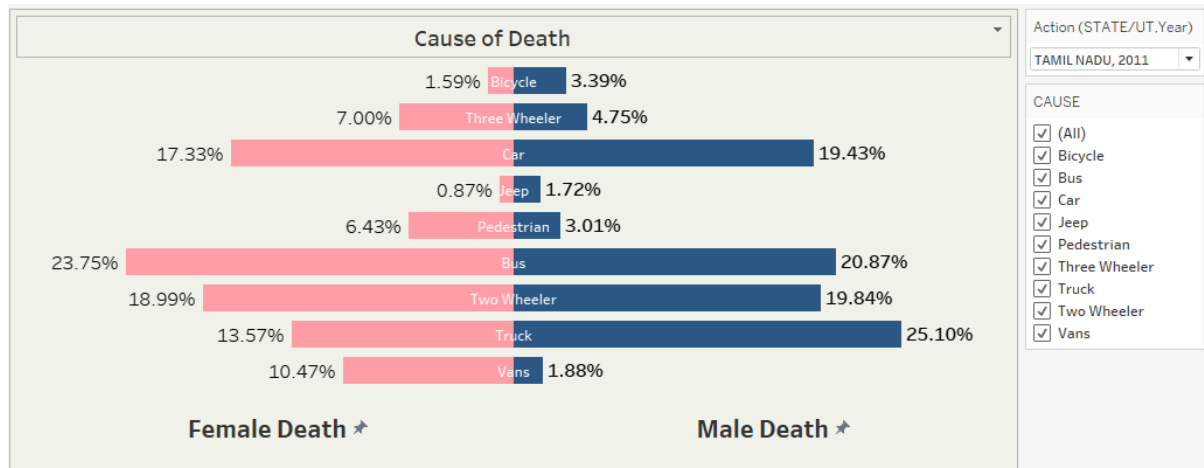


Figure3: Cause of Death butterfly graph, Colour = Male / Female, Size: Percentage of Death

To create the butterfly graph two zero measures was created. The measure was then combined with Sum of female death and sum of male death measures using dual axis function present in tableau. The colour pink is used for female death blue to denote male death because [4] after WWI, blue was used extensively for men's uniforms. which is why it got associated as more of a masculine colour. From the 1940's onward, pink was pushed as a woman's colour. "Think Pink" was the marketing slogan to convince women to embrace their femininity. The above graphs shows the vehicles which caused the accident and the percentage of death of both Male and female by the vehicle for the state of Tamilnadu in year 2011. This shows us that major accident and death causing vehicles are Trucks and Buses. Hence, helping us know which vechile is causing more fatalities.

#### Tool Used for Visualization:

Tableau – I used because support complex computations, data exploration for creating beautiful visualizations that deliver insights that cannot easily be derived.

#### 4. Conclusion:

India has one of the largest road networks in the world. However, Road safety is still an issue of national concern. I tried to visualize multiple dimensions and magnitude of road accidents in India. It gave helped us visualized which vehicle is causing more fatalities. The dip in accidents over the year shows a good sign that initiatives have been take my government, local bodies and public towards road safety. It also helped us know which states requires more attention in case of road safety and awareness. The Visualization will help to create awareness and decision making in the area of road safety. I hope, this interactive visualization would be useful for policy makers, citizens, and civil society organizations working in areas that have an influence on road safety.

Though the visualization provided us with good amount of information regarding accidents and their causes, we realized that data will be never to be enough to make a strong decision. If more data, like time when accident happen, state of driver (drunk or not) and so on are available, more test could be performed thus more analysis could be made from the data.

## **5.References:**

1. Amit (MARCH 6, 2016) Analysis of Road Accidents – India (2008-12) [online]  
Available at: <https://amitman3.wordpress.com/2016/03/06/analysis-of-road-accidents-india-2008-12/>
2. Bindu ((MAY 7, 2019) Road accidents still cause most deaths: WHO report [online]  
Available at: <https://www.thehindu.com/news/national/road-accidents-still-cause-most-deaths-who-report/article27061224.ece>
3. Prachi (March 28, 2017) Overview of Road Accidents in India [online] Available at:  
<http://prsindia.org/policy/vital-stats/overview-road-accidents-india>
4. Shraddha (July 28, 2015) This Is How Pink Came To Be A ‘Girly’ Colour [online]  
Available at: <https://www.scoopwhoop.com/inothernews/how-pink-became-the-girly-colour/>