# How would you visualize your Data?

Vinit Ravichandran Iyer
*Visual Analytics*
*The University of Sydney*
Sydney, NSW, Australia
viye2035@uni.sydney.edu.au

***Abstract*— This paper is for the assignment on how we would visualize data for the unit COMP5048. This report describes a certain dataset and its background. Furthermore, the report explores the different types of visualizations, graphical representations and derivation of new graphs based on the impositions of usual visualizations.**

***Keywords*— *Symbolic representations, data visualizations, tableau.***

## I. Introduction

A graph or a visualization is nothing but a graphical representation of data. The data in this case is numerical information collected through observation. This paper shows the different visualizations and their symbolic representations of the chose dataset. This paper uses the data obtained from the Government of India which is used to highlight the foreign tourist arrivals and trends in foreign exchange earnings. It also highlights the most visited monuments by International tourists (Foreign Tourist Arrivals in India, 2022).

## II. About the Data

### A. Nature of the Data

The data has been taken from the repository dataset of the Government of India. This data was collected for the purpose of understanding the trends of Foreign tourist arrivals in India and the trend in Foreign Exchange Earnings. The data is split into 4 different datasets, a preview for each of which has been attached in the appendix.

The datasets are:

- State-wise Domestic and Foreign tourist visit
- Foreign Exchange Earnings tourism India
- Foreign Tourist arrivals based on Age Group
- Number of visitors for centrally protected monuments.

The dataset was collected through multiple sources such as the logs and registers at the monuments in question, logs from immigration offices across the nation and many more.

### B. Consumers of such Data

These data are of utmost importance to the Reserve Bank of India due to the Foreign Exchange trend provided by this data. This data is also very useful for the tourism sector of India along with the hospitality section. The particular report which has been a resource for this paper was published by the OGD (Open Government Data) PMU (Project Management Unit) Team of India.

### C. Data Types involved

The visualizations have been created from a dataset which has been split into 4 sub-datasets. The attributes, data type and the description of the data is as follows:

**About the Data**

| *Attribute* | *Data Type* | *Description* |
|---|---|---|
| State/UTs | Qualitative | Represents the name of the State or Union Territory in question. |
| Domestic-2019 | Ratio | Represents the sum of domestic tourist which visited the state/uts in 2019. |
| Domestic-2020 | Ratio | Represents the sum of domestic tourists which visited the state/uts in 2020. |
| Foreign- 2019 | Ratio | Represents the sum of foreign tourists which visited the state/uts in 2019. |
| Foreign- 2020 | Ratio | Represents the sum of foreign tourists which visited the state/uts in 2020. |
| Growth Rate Domestic | Interval | Represents the growth rate of domestic tourist arrivals. |
| Growth Rate Foreign | Interval | Represents the growth rate of foreign tourist arrivals. |
| FTAs | Ratio | Represents the sum of foreign tourists for a particular year. |
| FEE in terms of crores | Interval | Represents the Foreign Exchange Earnings from tourism |
| Name of the monument | Qualitative | Represents the name of the monument being visited |

## III. Original Representation of Data

The data was presented to the users in a form of an infographic report. The following visualizations were created for the representation of each of the dataset.

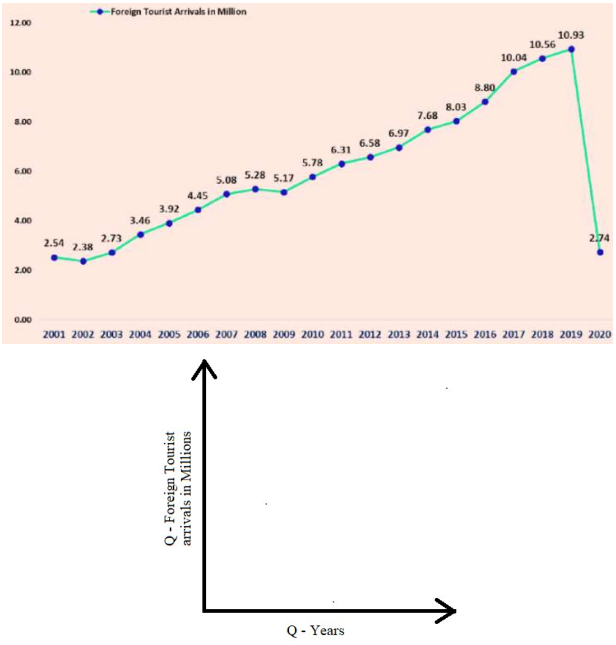



*Figure 1- Foreign Tourist Arrivals in India Trend and its symbolic representation.*

The data is aptly presented in Figure 1. The line chart captures the drastic drop in the number of foreign arrivals perfectly. The attributes used in this is of the numeric data type and are best represented by either a line chart or an area chart.

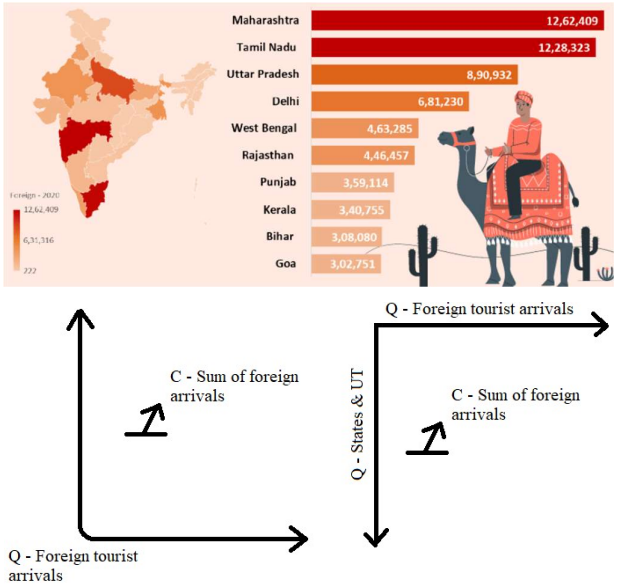



*Figure 2- Top 10 Most visited states of India by Foreign Tourists in 2020 and their symbolic representations.*

The data is presented in an appropriate format but the color representation is very wrong. The dark red hue usually is used for depicting something in the negative or something which is supposed to be opposite of expectations. In Figure 2 the color red has been used to depict the highest visitations by foreign tourists. This data can also be depicted on a pie graph.

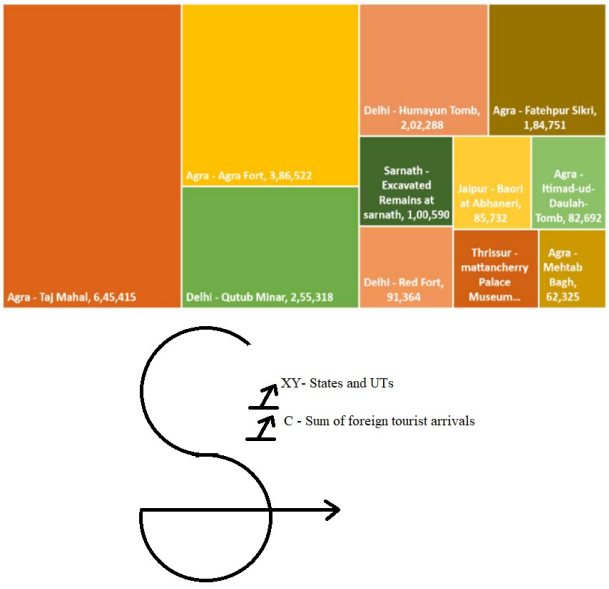



*Figure 3- Most visited Ticketed monuments of India in 2019-2020 and its symbolic representation.*

The data is presented in a tree map which makes it difficult to understand the purpose at first sight. The data is represented in a very wrong color format and needs to be in another color format. The tree map is used for making the report look a bit fancy and hence can be substituted with a bar graph or a pie chart which can easily represent the data in a much better and easy to understand format.

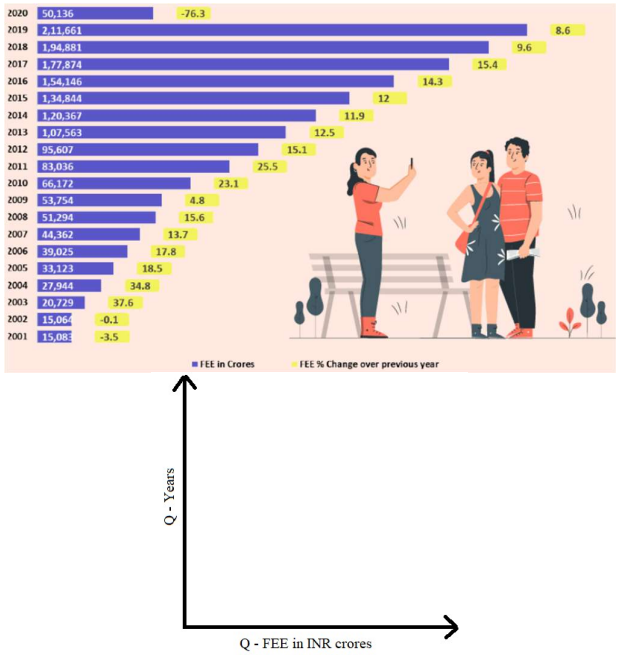



*Figure 4- Trend in Foreign Exchange Earnings from Tourism in India*

The visualization represents the data in an appropriate way but at first sight gives a misinterpretation that it has been sorted. Rather than representing the data in this format a pie chart, tree map or any other visualization which does not display an orderliness in the data can be used to make sure misinterpretations are avoided.

## IV. Questions Asked through this Data

The questions asked for the visualizations are very clear. The visualizations answer them quite easily but fail to answer any in-depth, analytical questions. A few questions the data answers are:

- How many foreign tourists have arrived in India every year?
- What is the trend of the foreign tourist arrivals in India?
- What are the top 10 most visited states in India by Foreign tourists?
- Which monument has the most ticketed visits in India?
- How much has India earned through Foreign Exchange Earnings from tourism in the past 20 years?
- How did the trends shift due to COVID-19?

## V. Typical Mistakes Committed during Visualizations

The visualizations are great for answering direct questions but fail to answer analytical or in-depth questions.

Figure 1 depicts information of the sum of foreign tourist's arrival to India but does injustice to the data collected on the age bins. Figure 2 displays the top 10 most visited states of India by foreign tourists but fails to display the domestic tourists which also contribute significantly towards the Indian tourism industry. The visualization also fails to provide information about the least visited states, knowledge of which could be used for increasing the tourism rate of the state through several reforms. The same problem persists with Figure 3 where the most visited ticketed monuments statistics are given but information for the least ticketed monuments is not. Another problem is the absence of details. The numeric value present is an unknown value and not known whether it is of foreign tourists, domestic tourists or a sum of both.

## VI. My Visualizations

The visualizations provided can be further improved. The following are improved visualizations of the previously given figures. The symbolic representation of the old visualization, symbolic representation of the new visualization and the new visualization based on the new symbolic representation are as given below:

For figure 1, I would have included a trend line to depict the supposed information and the reality to depict the massive drop in the foreign traveller arrivals. The new visualization would be figure 5.

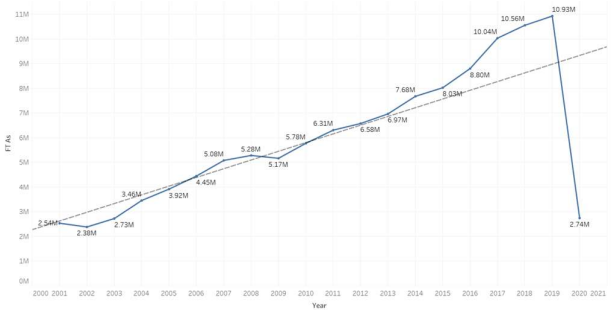



*Figure 5- Trend of Foreign Tourist arrivals vs actual statistics.*

Figure 2 is made up of 2 visualizations, one of the country map representing the number of foreign arrivals along with a bar graph giving the exact numbers for the top 10 states.

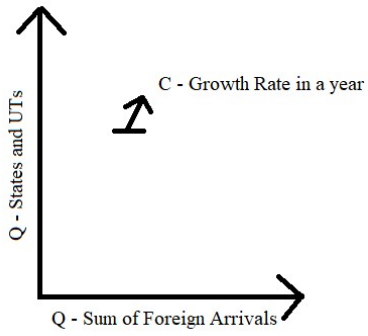



*Figure 6- Symbolic representation of Improvised figure 2*

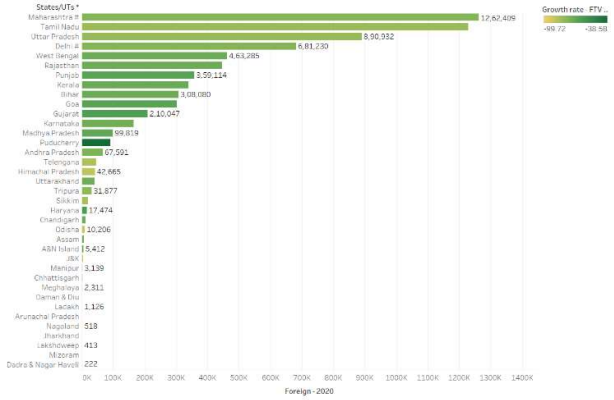
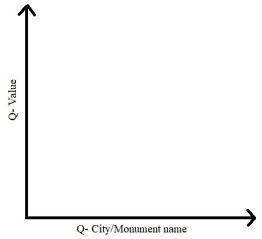



*Figure 7- Improvised visualization of sum of foreign arrivals in India according to State and UTs.*

In the improvised visualization of Figure 2, the visualization retains the bar graph but removes the geographic map. The visualization uses the bar graph to visualize all of the states and union territories and also incorporates an imposition of the growth rate into the visualization as color. The years and names of the states and union territories form the XY axis of the plot, the sum of foreign arrivals in each state as the main constituent of the graph and the growth rate of the arrivals as a color component. This visualization was derived by adding the imposition of color for growth rate while keeping all the other impositions constant.

Figure 3 is a tree map which visualizes the most ticketed monuments in India. From this visualization I created a side-by-side bar chart which visualizes the top 10 ticketed monument locations in India on the basis of Domestic and Foreign arrivals. The symbolic representation of the visualization is as follows:




The visualization is as follows:

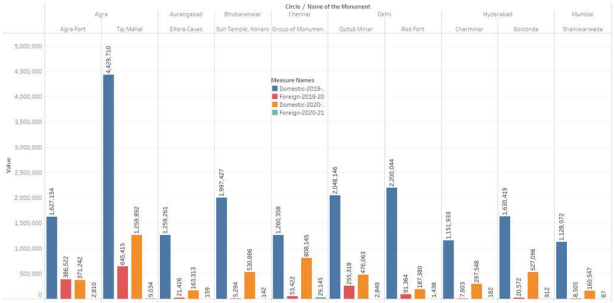



*Figure 8- Top 10 monuments city-wise based on their domestic and foreign arrival values of the year 2019-2021.*

The visualization puts forward all the information regarding the top 10 monuments which have observed the most ticketed arrivals, both of domestic and foreign. The tree map gives an overview regarding the information whereas the newly created visualization gives in detail information of the top 10 monuments. For the visualization the symbolic representation was easily derived from the tree map's symbolic representation. The XY imposition became the axes of the new visualization while removing the color imposition for demarcating the sum of foreign arrivals and instead being used to represent the 4 different sums of ticketed arrivals instead.

## VII. Other Visualizations

Symbolic representations allow us to freely convert any visualization to another using the same parameters. In this section of the assignment, we would use symbolic representations to create a new visualization from the symbolic representations already created for the improvised visualizations.


*Figure 9- State-wise sum of foreign arrivals and growth rate*

The above symbolic representation is an equivalent of the original alternate visualization (figure 7). This visualization used the States as the axes, the sum of foreign arrivals as the size of the blocks in the tree map and the color imposition is the representation of the growth rate of the foreign arrivals in 2020.

## References

[1] Foreign Torusit Arrivlas in India, 2022 Retrieved Spetember 17, 2022 from https://community.data.gov.in/foreign-tourist-arrivals-in-india/

## VIII. APPENDIX

Datasets:

*Table 1- Domestic and Foreign tourist visit state-wise*

| S. No. | States/UTs * | Domestic -2019 | Foreign - 2019 | Domestic -2020 | Foreign - 2020 | Growth rate - DTV 2020/19 | Growth rate - FTV 2020/19 |
|---|---|---|---|---|---|---|---|
| 1 | A&N Island | 505398 | 16206 | 191207 | 5412 | -62.17 | -66.6 |
| 2 | Andhra Pradesh | 237051508 | 280356 | 70828590 | 67591 | -70.12 | -75.89 |
| 3 | Arunachal Pradesh | 555639 | 7825 | 42871 | 961 | -92.28 | -87.72 |
| 4 | Assam | 5447805 | 26878 | 1266898 | 7285 | -76.74 | -72.9 |
| 5 | Bihar | 33990038 | 1093141 | 5638024 | 308080 | -83.41 | -71.82 |
| 6 | Chandigarh | 1563795 | 44132 | 417953 | 12218 | -73.27 | -72.31 |
| 7 | Chhattisgarh | 17304506 | 6817 | 2810227 | 2322 | -83.76 | -65.94 |
| 8 | Dadra & Nagar Haveli | 618330 | 1666 | 104959 | 222 | -83.03 | -86.67 |
| 9 | Daman & Diu | 897804 | 5703 | 297436 | 1382 | -66.87 | -75.77 |
| 10 | Delhi # | 36467598 | 2983436 | 9583671 | 681230 | -73.7 | -77.2 |

Source: https://data.gov.in/resource/foreign-tourist-arrivals-ftas-india-according-age-group-2001-2020

*Table 2- Foreign Exchange Earnings tourism India*

| Year | FEE in ` terms - `Crore | FEE in ` terms - % Change over previous year | FEE in US$ terms - US $ Million | FEE in US$ terms - % Change over previous year |
|---|---|---|---|---|
| 1991 | 4318 | NA | 1861 | NA |
| 2001 | 15083 | -3.5 | 3198 | -7.6 |
| 2002 | 15064 | -0.1 | 3103 | -3 |
| 2003 | 20729 | 37.6 | 4463 | 43.8 |
| 2004 | 27944 | 34.8 | 6170 | 38.2 |
| 2005 | 33123 | 18.5 | 7493 | 21.4 |
| 2006 | 39025 | 17.8 | 8634 | 15.2 |
| 2007 | 44362 | 13.7 | 10729 | 24.3 |

Source: https://data.gov.in/resource/foreign-exchange-earnings-tourism-india-during-1991-2020

*Table 3- Foreign tourist arrivals based on age groups*

| Year | FTAs | % distribution by Age- Group (in years) - 0-14 | % distribution by Age- Group (in years) - 15-24 | % distribution by Age- Group (in years) - 25-34 | % distribution by Age- Group (in years) - 35-44 | % distribution by Age- Group (in years) - 45-54 |
|---|---|---|---|---|---|---|
| 2001 | 2537282 | 7 | 10.8 | 20.1 | 21.1 | 19.4 |
| 2002 | 2384364 | 9.2 | 10 | 19.4 | 21.6 | 19.4 |
| 2003 | 2726214 | 7.2 | 10 | 19.5 | 21.6 | 19.4 |
| 2004 | 3457477 | 8.5 | 9.8 | 18.8 | 21.3 | 19.4 |
| 2005 | 3918610 | 8.6 | 9.6 | 18.8 | 21.3 | 19.5 |
| 2006 | 4447167 | 9.6 | 8.7 | 18.3 | 21.4 | 19.6 |
| 2007 | 5081504 | 9.2 | 8.6 | 18 | 21 | 19.6 |
| 2008 | 5282603 | 7.5 | 7.1 | 15.2 | 17.1 | 16.1 |

Source: https://data.gov.in/resource/stateut-wise-domestic-and-foreign-tourist-visits-2019-20

*Table 4- Number of visitors centrally protected ticketed monuments*

| Circle | Name of the Monument | Domestic-2019-20 | Foreign-2019-20 | Domestic-2020-21 | Foreign-2020-21 | % Growth 2021-21/2019-20-Domestic | % G |
|---|---|---|---|---|---|---|---|
| Agra | Taj Mahal | 4429710 | 645415 | 1259892 | 9034 | -71.56 | |
| Agra | Agra Fort | 1627154 | 386522 | 371242 | 2810 | -77.18 | |
| Agra | Fatehpur Sikri | 454376 | 184751 | 107835 | 574 | -76.27 | |
| Agra | Akbar Tomb Sikandra | 229270 | 19625 | 99509 | 321 | -56.6 | |
| Agra | Mariam tomb Sikandra | 22517 | 414 | 9765 | 31 | -56.63 | |
| Agra | Itimad-ud-Daulah-Tomb | 132800 | 82692 | 41016 | 410 | -69.11 | |
| Agra | Ram Bagh | 84051 | 355 | 18599 | 54 | -77.87 | |
| Agra | Mehtab Bagh | 178574 | 62325 | 62652 | 544 | -64.92 | |
| Agra | Mausoleum | 474462 | 12536 | 91185 | 321 | -80.78 | |
| Total | Total | 7632914 | 1394635 | 2061695 | 14099 | -72.99 | |
| Lucknow | Site of Sahet mahet | 74597 | 13628 | 27201 | 35 | -63.54 | |
| Lucknow | Residency Building | 116348 | 3367 | 57194 | 37 | -50.84 | |

Source: https://data.gov.in/resource/number-visitors-centrally-protected-tickted-monuments-2019-20-2020-21