



ISSN: 0975-766X  
CODEN: IJPTFI  
Research Article

Available Online through  
[www.ijptonline.com](http://www.ijptonline.com)

## DATA VISUALIZATION OF VARIOUS INCLINATIONS IN HEALTH CARE SYSTEM

Aniketh Bondada, Harsha Vardhan M, Sai Teja S, Jabanjalin Hilda J<sup>1,2,3,4</sup>

<sup>1,2,3,4</sup>School of Computer Science and Engineering, VIT University, Vellore, TamilNadu.

Email: [jabanjalin.hilda@vit.ac.in](mailto:jabanjalin.hilda@vit.ac.in)

Received on: 15-02-14

Accepted on: 25-03-17

### Abstract

**Background and Objective:** In the current society with increasing need of health care facilities, it has been a pivotal sector in the society. So with such importance and demands comes the requirement for efficient health care and with that in picture we need reliability of these services being provided to us. Therefore faster cure and efficient cure is the need of the hour, which adversely needs information regarding as in where and which disease is being cured with utmost efficient and ease. So in this paper we provide a solution to meet this by automatically monitoring the data containing the death rates at various parts of India pertaining to various prominent diseases. **Method:** The visualization is done by using the Tableau Software. Some key advantages of using Tableau are it works with bigger data set for visualization, and also it has intuitive interface which makes it a user friendly tool. Tableau provides us quick and highly efficient visualization of research data without the maximum need to build sophisticated applications. **Result:** Death rate comparisons for various diseases are made and visualized for several states in India. The outcome of these visualizations help the general public to be informed and take knowledgeable decisions regarding their health ensuring safety and good health.

**Keywords:** Tableau, Data Visualization, Health Care, Death Rate

### 1. Introduction:

In this study, we will be collecting the data from the hospital database from various parts of the country. This data will include deaths of the people at various states with respect to various diseases. The prominent diseases, which will be analyzed, are cancer, tuberculosis, cholera, diarrhea and typhoid. In the data we will be having deaths pertaining to all the states in India pertaining to those diseases. The objective of this analysis study is at the end of the day we need visualizations which tell us which disease is being cured the most at a particular area and at the same time it tells us which diseases are being affected the most in a particular area accordingly. So based on our need and requirement we

get to know the desired information from the visualizations. When we need to make a decision as in where to head when we are attacked with a particular disease, by referring to these visualizations an efficient decision can be made accordingly. The software which we will be using to visualize the data is Tableau which is an open source visualization tool of various data sets. In the domain of healthcare, analytics is an excellent decision support technology. It enables physicians, nurses, health officials, health policy makers, and other healthcare entities to make better and faster health decisions. We show in a form of trend analysis how the various death rates tell us about those diseases at those areas. Relationships between indicators are also established using the advanced statistical modeling technique. We obtain all these visualizations using such data sets.

**2. Methodology:** In this study we can divide the complete analysis into majorly three steps which being, data collection, data visualization and data presentation. This basically includes firstly taking all the data into excels and then by using tableau to visualize and then we take that data properly analyze with respect to the need and the factors to be compared.

### **2.1 Data Collection:**

The various data which we need include for each state the number of death which are required and the same the different diseases which are predominant in those areas. The major challenge during this is the legitimacy of the data being obtained because many further decisions depend on this data and proper care should be taken to produce reliable data. So in order to obtain reliable data we have outsourced data from the government databases available online which include data regarding the numbers of deaths being reported at each hospital for every year. So when someone needs to make a decision to select the hospital instead of looking over short term performance of that hospital, they need to know on a long term basis, how these hospitals are performing and in order to do that we need to analyze for at least three years and we are referring to 2013, 2014 and 2015[1].

With all these data in excel sheets categorized into columns we can further move on to visualize, the need for the data to be in columns is because only then Tableau can recognize those columns to be a metric to be visualized.

### **2.2 Data Visualization:**

The main challenge we face is during the visualization part where we have many metrics but we should be aware to compare the correct metrics so that we have proper analytics based on which we can take a decision, comparing unrelated metrics cannot be used to make any decision in that aspect. So firstly under Tableau we need to connect a data source which in our case is the death rates at various states for a span of 3 years. This dataset can be in any form

be it excel, notepad or word. Tableau automatically segregates the various sectors and columns to form separate metrics and put them in groups. Now we have all such groups at our exposure and we can use any one of them to compare and analyze. We take 2 or 3 metrics and put them along either X or Y axis and then those are further put up to explore the different types of graphs which are possible. Based on our need and requirement we take the particular metrics and then the visualizations are brought up. Further if we need to analyze discrete data then we can change the data into discrete and then we can perform the visualizations which will give us the distributed data in form of discretion. With such visualizations we can evaluate the deaths and other metrics in a much deeper way giving us in depth knowledge which adversely gives us effective decisions at the end of the day.

### 3. Discussions:

Cancer and Tuberculosis are two commonly highly infected diseases in India. In this graph we will plot the data available as year against disease state wise in India. Column wise we will have the states plotted along their values by the Vertical bar graph. Comparing the two dreadful diseases statistically and biologically we can depict the required information which provides the state with highest caused deaths caused by either of the diseases , state with least caused deaths and other pivotal information that may lead to further study of the diseases and to explore remedy options for any particular state[2]. We will start the visualization by the first graph which is nothing but visualizing the data regarding cancer year wise. We have the bar graph starting from the year 2012 -2014. The X-axis indicates the year plotting against the Y-axis which provides the value from the dataset. Each State and Union Territory is shown with respect to the above axes. We can derive from the graph that Uttar Pradesh stands among the highest effected state related to deaths with current value as per 2015 standing at 186638 deaths. There is an increase of 3.15% from the year 2014 to 2015 in that particular state itself. Bihar and Maharashtra stand at the next two highly caused deaths states. In the graph we can find some states and UT's negligible as the deaths occurred over there are relatively less than compared to top 3 cancer causing states.

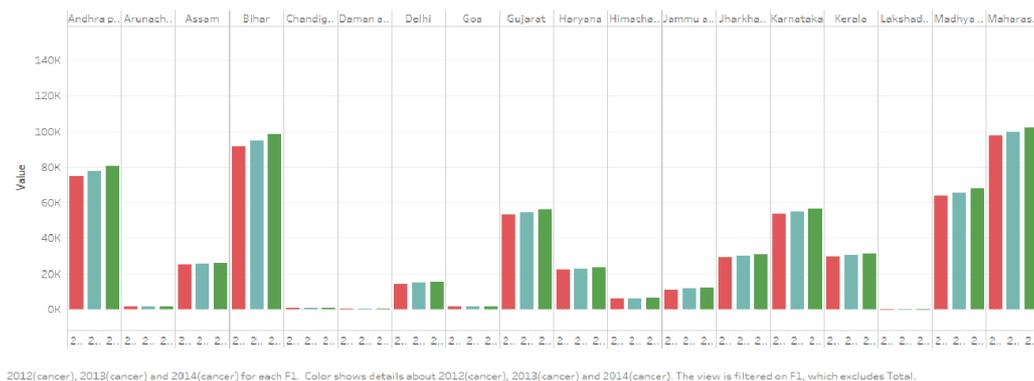
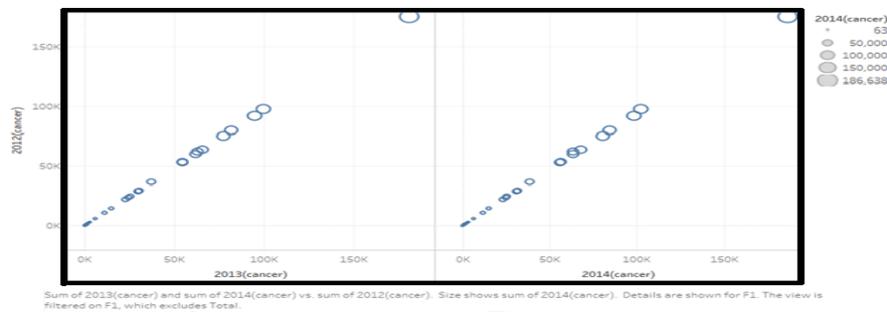


Figure 1: Deaths caused by Cancer (State wise) in 2012-2014.

The second illustration of the data regarding cancer deaths is the form of Scatterplot graph. It is viewed and portrayed by comparing the Y-axis with year 2012 and X-axis with year 2013 and 2014 respectively.

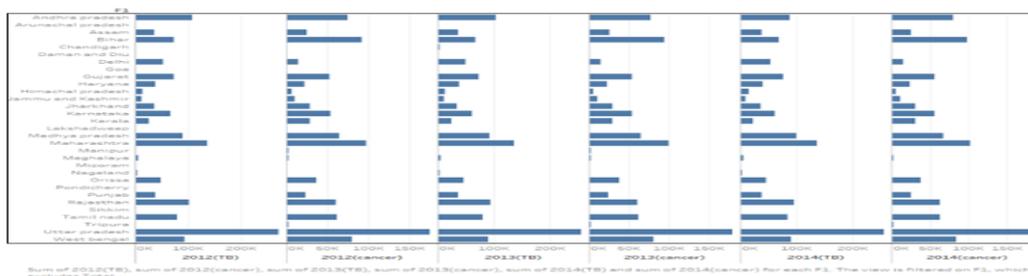
A similar graph unlike the bar format this graph is in the form of Scatter plot graph which is drawn upon states vs value and shown. The values are divided into 5 classes as per the available data. The least class starts from value 63 deaths which was recorded from the UT Lakshadweep and the highest class is recorded for the value 186,638 deaths by the state Uttar Pradesh. The median class was 100,000 deaths which was recorded by Maharashtra. There is a huge difference between the median class and the highest valued class which can be visible from the Scatter plot graph[3]. There is no state with decline every consecutive year on the death rate but apparently the rate has raised every year for all the states and UT's. We can mainly find grouping of a number of states that are divided class wise and are mostly correlated by overlapping of the values in the same region as per the available data.



**Figure 2: Deaths caused by Cancer in 2012 against 2013 and 2014 by Scatter plot graph.**

Figure 1 outlined the data in the form of vertical bar graph, this graph is in the form of horizontal bar graph. The addition in this graph is the disease Tuberculosis is added to compare against Cancer; the death rates between these two diseases are drawn among year wise from 2012-2014 are displayed in the graph. Similar to the death rate case regarding cancer , Uttar Pradesh tops the death rate in Tuberculosis too followed by Maharashtra in the same manner. Sikkim (1638) and Goa (1664) record the least number of death rate states for Tuberculosis. The death rate is less in eastern states than compared to 2013 which can be visible through the graph.

The year wise distribution of values for both the diseases is drawn against each state which can be shown in the graph below.



**Figure 3: Deaths caused by Tuberculosis and Cancer year wise.**

This consists of a table which details the data caused by the next category of diseases that include Cholera, Typhoid, Diarrhea and Viral Hepatitis. Comparison of death rate is done among these 4 diseases and is later visualized[4].

State wise the data is depicted in the form of table graph which includes the category 2 diseases which include Cholera, Typhoid, Diarrhea and Viral Hepatitis.

State / U.T. ...	2012 -	2012 -	2012 -	2012 -	2013 -	2013 -	2013 -	2013 -	2014 -	2014 -	2014 -	2014 -	2015 -	2015 -	2015 -	2015 -
A & N Islands	0.0	3.0	1.0	9.0	11.0	0.0	0.0	3.0	0.0	2.0	0.0	9.0	0.0	0.0	0.0	1.0
Andhra Prad.	0.0	100.0	37.0	84.0	34.0	0.0	110.0	6.0	0.0	10.0	5.0	1.0	0.0	2.0	0.0	6.0
Arun. Prad.	0.0	7.0	10.0	4.0	0.0	0.0	3.0	4.0	0.0	5.0	3.0	6.0	0.0	3.0	6.0	1.0
Assam	0.0	147.0	10.0	0.0	0.0	0.0	147.0	0.0	0.0	64.0	29.0	13.0	0.0	82.0	2.0	7.0
Bihar	0.0	8.0	3.0	2.0	2.0	0.0	24.0	2.0	0.0	24.0	4.0	3.0	0.0	21.0	1.0	2.0
Chandigarh	0.0	0.0	0.0	0.0	1.0	0.0	2.0	0.0	1.0	29.0	0.0	21.0	0.0	72.0	27.0	24.0
Chhattisgarh	0.0	26.0	6.0	5.0	5.0	0.0	37.0	2.0	0.0	32.0	0.0	4.0	0.0	12.0	1.0	6.0
D & N Haveli	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	4.0	0.0	4.0
Daman & Diu	0.0	0.0	0.0	4.0	7.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0
Delhi	0.0	98.0	71.0	66.0	131.0	0.0	62.0	29.0	0.0	73.0	14.0	95.0	0.0	59.0	11.0	48.0
Goa	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	1.0	1.0	0.0
Gujarat	0.0	7.0	0.0	6.0	6.0	1.0	8.0	1.0	0.0	3.0	0.0	7.0	0.0	3.0	1.0	0.0
Haryana	0.0	27.0	1.0	1.0	1.0	0.0	27.0	0.0	0.0	8.0	1.0	4.0	0.0	4.0	0.0	2.0
J & K	0.0	2.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	2.0	0.0	0.0
Jharkhand	0.0	6.0	39.0	0.0	65.0	0.0	4.0	7.0	0.0	28.0	7.0	20.0	0.0	0.0	3.0	1.0
Karnataka	0.0	84.0	1.0	26.0	16.0	0.0	81.0	9.0	0.0	23.0	1.0	9.0	2.0	13.0	1.0	16.0
Kerala	1.0	10.0	1.0	18.0	8.0	0.0	12.0	3.0	0.0	9.0	0.0	4.0	0.0	0.0	0.0	2.0
Lakshadwe...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Madhya Pra...	0.0	91.0	29.0	4.0	11.0	0.0	89.0	28.0	0.0	112.0	31.0	18.0	0.0	65.0	4.0	14.0
Maharashtra	0.0	1.0	2.0	21.0	13.0	1.0	0.0	1.0	2.0	4.0	0.0	13.0	0.0	13.0	0.0	6.0
Manipur	0.0	56.0	5.0	4.0	0.0	0.0	37.0	17.0	0.0	32.0	10.0	0.0	0.0	11.0	0.0	0.0
Meghalaya	0.0	19.0	10.0	1.0	0.0	0.0	12.0	1.0	1.0	27.0	8.0	0.0	0.0	20.0	0.0	0.0
Mizoram	0.0	7.0	1.0	15.0	10.0	0.0	12.0	3.0	0.0	10.0	4.0	0.0	0.0	8.0	0.0	3.0
Nagaland	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Puducherry	0.0	21.0	0.0	19.0	10.0	0.0	28.0	1.0	0.0	11.0	3.0	0.0	0.0	0.0	0.0	2.0
Punjab	0.0	27.0	4.0	0.0	6.0	0.0	13.0	5.0	0.0	22.0	1.0	4.0	0.0	32.0	1.0	7.0
Rajasthan	0.0	12.0	4.0	7.0	10.0	0.0	18.0	5.0	0.0	17.0	4.0	3.0	0.0	6.0	0.0	0.0
Sikkim	0.0	0.0	0.0	6.0	1.0	0.0	1.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
Tamil Nadu	0.0	17.0	0.0	0.0	0.0	3.0	23.0	1.0	0.0	14.0	0.0	0.0	0.0	8.0	0.0	0.0
Tripura	0.0	22.0	3.0	2.0	1.0	0.0	17.0	1.0	0.0	22.0	0.0	1.0	0.0	2.0	1.0	1.0
Uttar Prade...	0.0	254.0	65.0	12.0	17.0	0.0	272.0	161.0	0.0	301.0	203.0	50.0	0.0	241.0	183.0	47.0
Uttarakhand	0.0	21.0	4.0	16.0	11.0	0.0	12.0	0.0	0.0	14.0	14.0	9.0	0.0	4.0	10.0	12.0
West Bengal	0.0	280.0	29.0	102.0	91.0	0.0	302.0	39.0	0.0	200.0	42.0	63.0	0.0	129.0	18.0	54.0

Figure 4: Tables for the deaths caused by Cholera, Typhoid, Diarrhea and Viral Hepatitis.

The category 2 diseases which include Cholera, Typhoid, Diarrhea and Viral Hepatitis are now compared year wise first in the form of regular Horizontal bar graph. Disease wise Cholera stands as the least death causing disease as the first 4 columns describe regarding Cholera followed by Diarrhea, Typhoid and Viral Hepatitis. Uttar Pradesh (241) and West Bengal (129) recorded the highest death rate for Diarrhea. This disease is mostly declined in major effected states as the year increases[5]. This can be shown from the graph below.

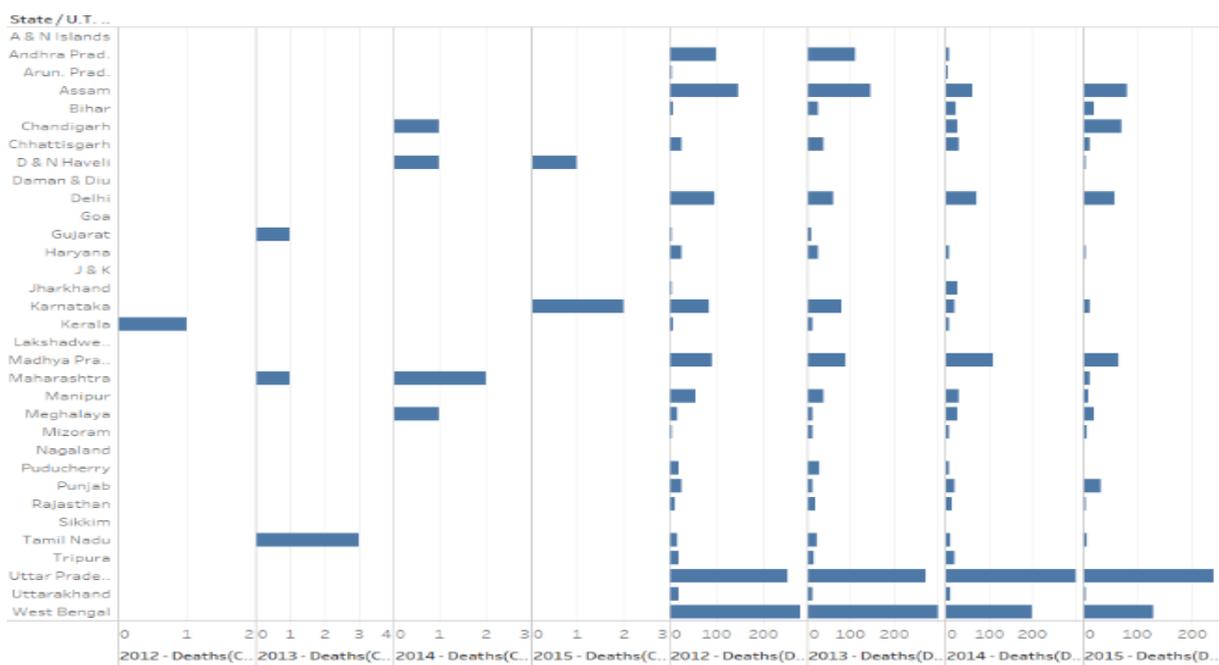
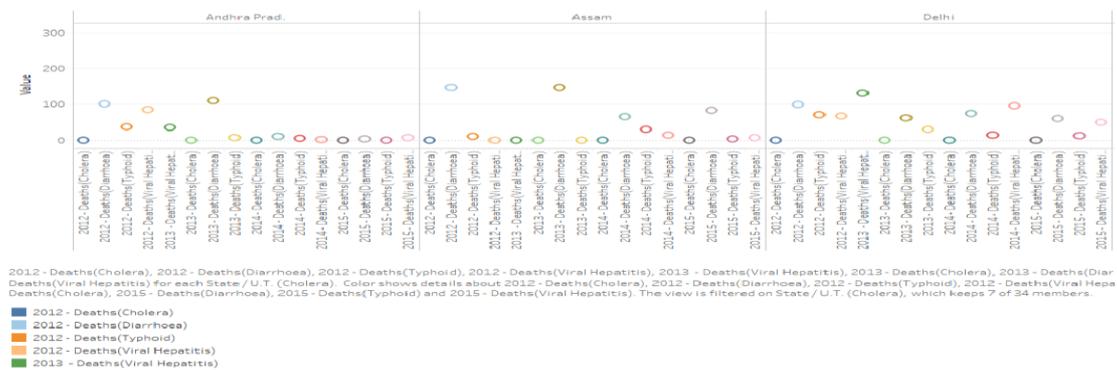


Figure 5: Graph for the deaths caused by Cholera, Typhoid, Diarrhea and Viral Hepatitis.

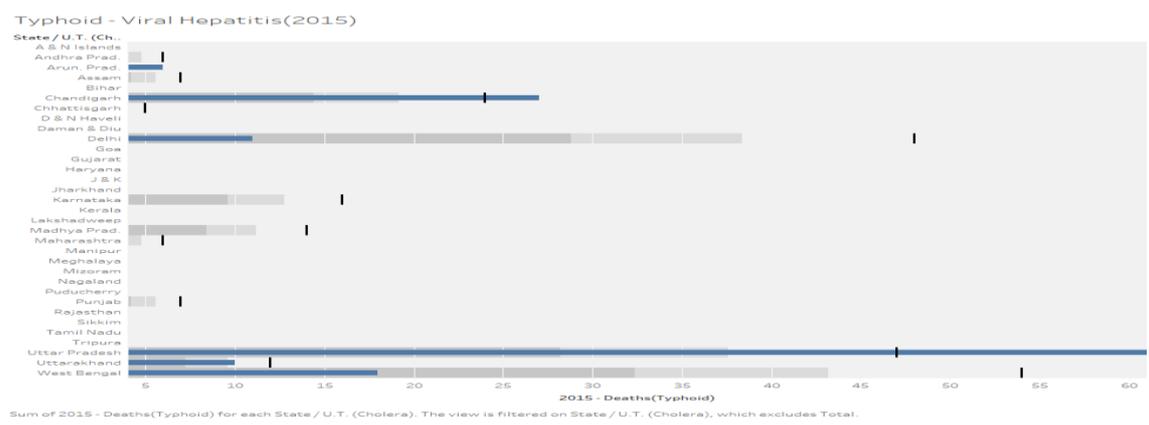
This graph is the Side by Side Circle graph which portrays data in the form of circles for all the available factors and attributes that include the state and disease. Inclusive of all the diseases the top 7 states have been filtered as the top effected states from the available 34 states regarding all the 4 diseases and also for the 4 years. Uttar Pradesh, West Bengal, Assam, Delhi, Madhya Pradesh, Karnataka and Andhra Pradesh are the leading states considering all the other factors. Cholera is the minimal effected disease that cause death rate to be very low with Karnataka registering 3 deaths and Tamil Nadu registering 2 deaths.



**Figure 6: Circle data graph for the deaths caused by Cholera, Typhoid, Diarrhoea and Viral Hepatitis.**

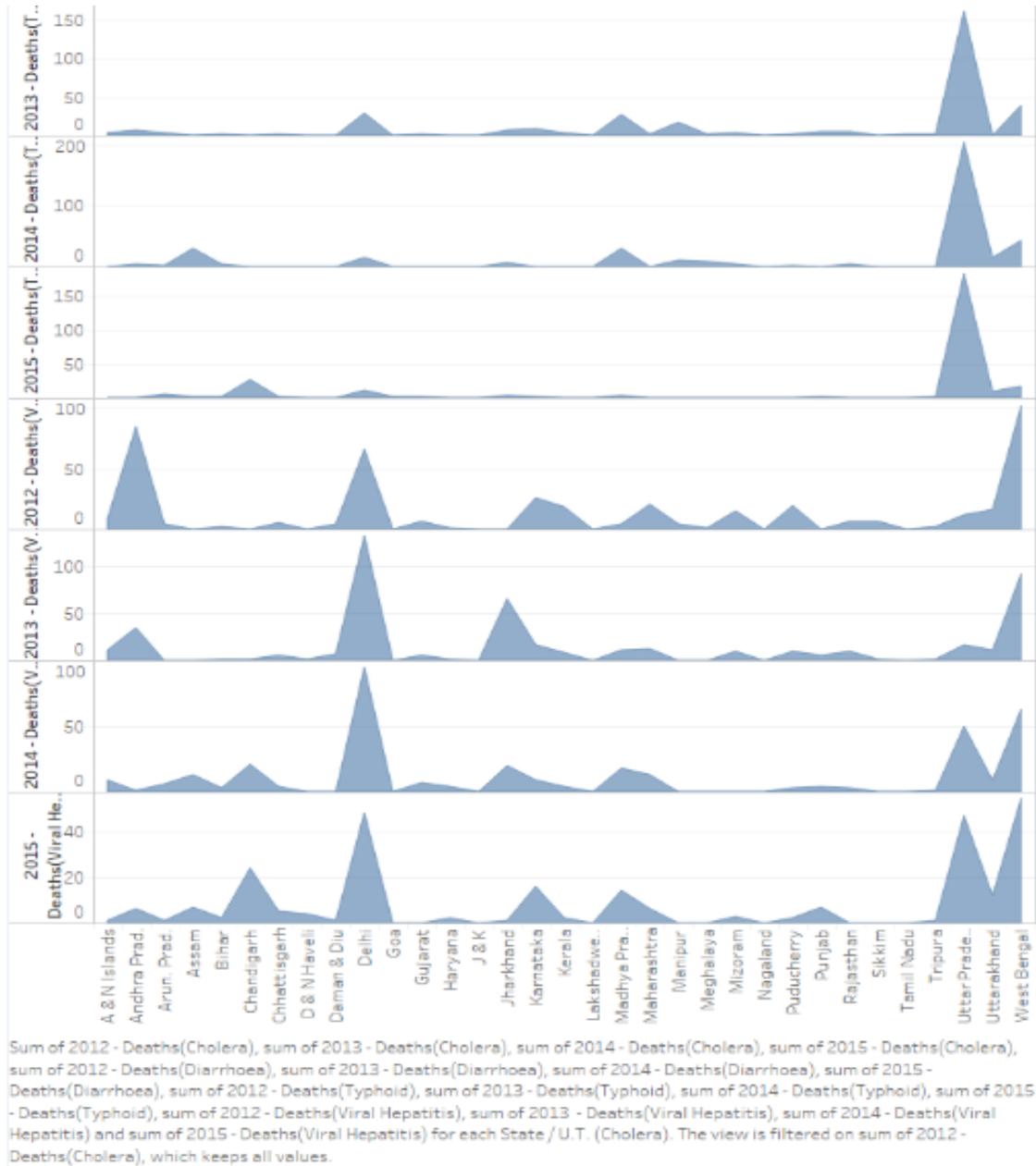
A bullet graph is a variety of a structured presentation. Bullet diagrams were created to commonly show too little data, require an excess of space, and are messed with pointless and diverting enhancement[6]. It mainly displays the response from the graph like satisfactory, excellent and very good. Here we have used Typhoid and Viral Hepatitis as the parameters with year 2015 as the main frame target and have created the graph.

The black line indicates the average for both the diseases combined and it used as a reference mark. The blue bar indicates for Typhoid and the grey bar for Viral Hepatitis. Considering in the case of Delhi which records for the highest account for Viral Hepatitis has an average mark of 48 deaths and the blue bar (Typhoid) shows 11 deaths whereas the grey bar (Viral Hepatitis) shows 42 deaths. Uttar Pradesh accounts high for Typhoid whereas considering the average mark Delhi and West Bengal stay at the highest point.



**Figure 7: Bullet graph for the deaths caused by Typhoid and Viral Hepatitis.**

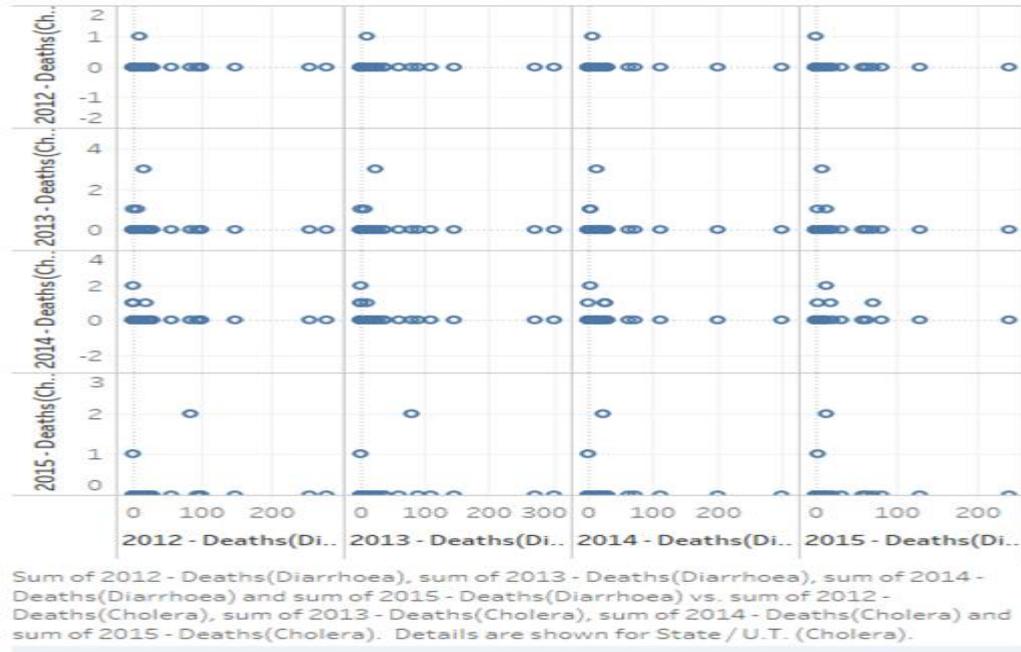
Area Chart graph is another graph which provides the data in the form of curve that contains a peak point and minimal point for every state. The view is filtered which keeps all the values showing the graph for state which is at maximum value at the highest peak point for the particular disease state wise and year wise. Consider Viral Hepatitis in Delhi during 2015 which caused the highest amount of deaths so there is a peak point which reaches its value. Uttar Pradesh stands at a high point in many of the diseases.



**Figure 8: Area chart graph for the deaths caused by Cholera, Typhoid, Diarrhoea and Viral Hepatitis.**

Circle View graph is used in comparing the values for Diarrhoea and Cholera with the available factors in the form of state and year. The deaths value of Cholera is highly negligible and is in the range between 0 to 5. The Cholera values are plotted on the Y-axis against the Diarrhoea values on the X-axis. Uttar Pradesh records the values - (254,272,301,241) deaths during the year 2012-2015 respectively. The values are plotted with taking any scale and are visualized as per the available data.

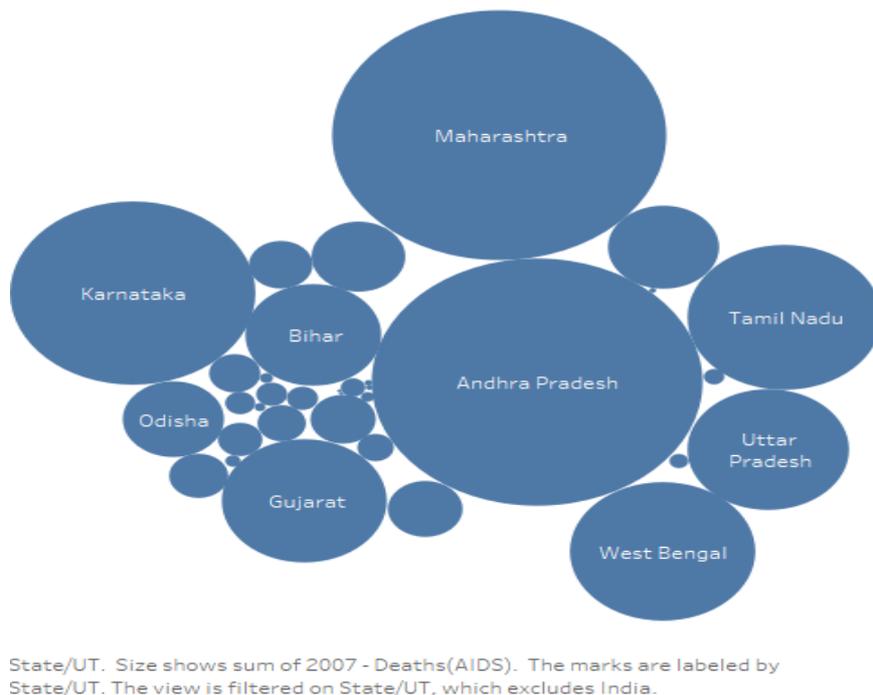
We can find some overlapping for the Cholera values against the Diarrhoea values which indicate negligible Cholera values matching with the same Diarrhoea values.



**Figure 9: Circle view graph for the deaths caused by Diarrhoea and Cholera.**

AIDS is one dangerous disease that is highly present throughout the world. Comparing it state wise in India we come across some factors to visualize which are nothing but the year and the state itself.

Packed Bubbles[7] graph is the graph which visualizes the data in the form highly shaped bubbles depending on the effected number of people. Andhra Pradesh comprises of the highest tally regarding the death rate as it is visible from the bubble size. Maharashtra is the next state that follows up the death toll.



**Figure 10: Packed Bubbles graph for the deaths caused by AIDS.**

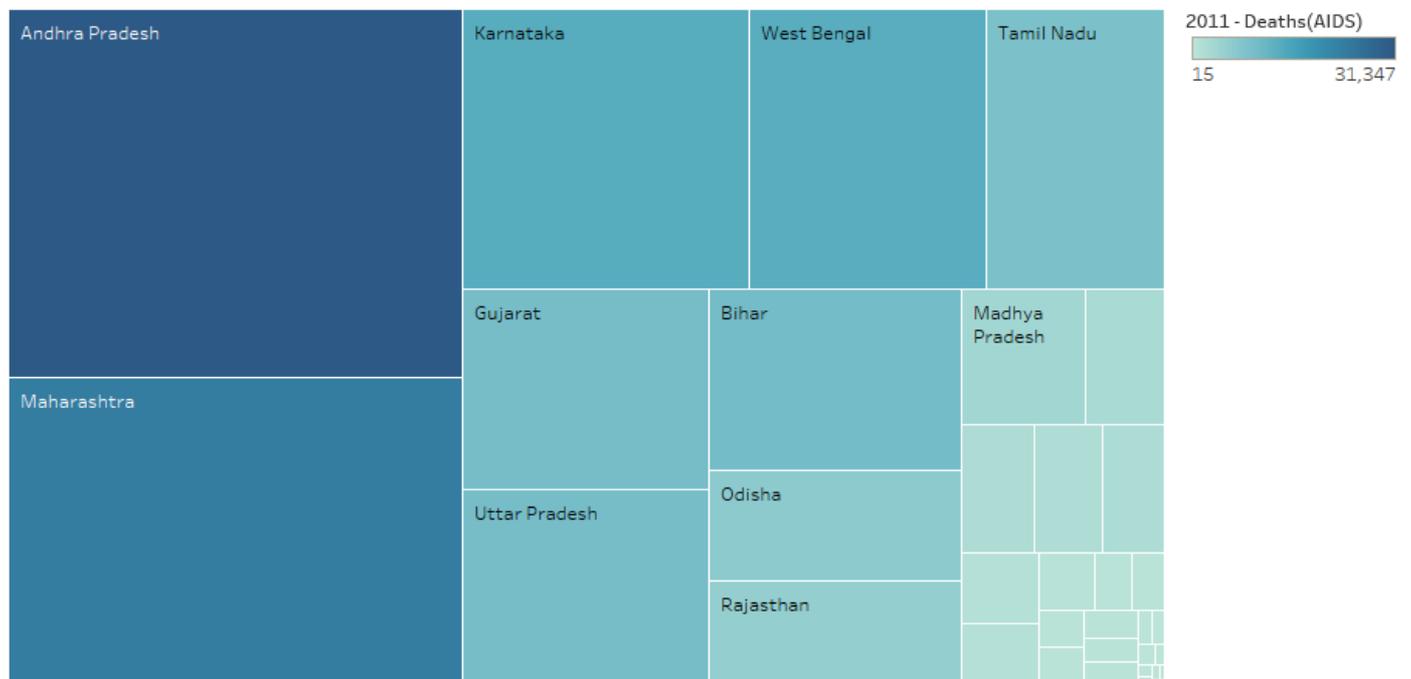
The tree maps graph is the graph which is based on the intensity of the corresponding value. It is clearly visible from the tree map that which state leads among the rankings. The unnamed boxes are very negligible in value and hence are not named.

The top 6 states effected by AIDS data wise are shown in this small table in the form State vs Year :

**Table 1: Top 6 states effected by AIDS in India**

State/UT	2007	2008	2009	2010	2011
Andhra Pradesh	46,427	42,337	38,919	35,276	31,347
Maharashtra	47,343	40,734	35,522	29,350	23,764
Karnataka	25,510	23,136	20,686	16,927	13,514
West Bengal	14,578	14,629	14,514	14,027	13,310
Tamil Nadu	15,946	13,616	12,286	10,508	8,582
Gujarat	11,540	11,426	11,284	10,489	9,510

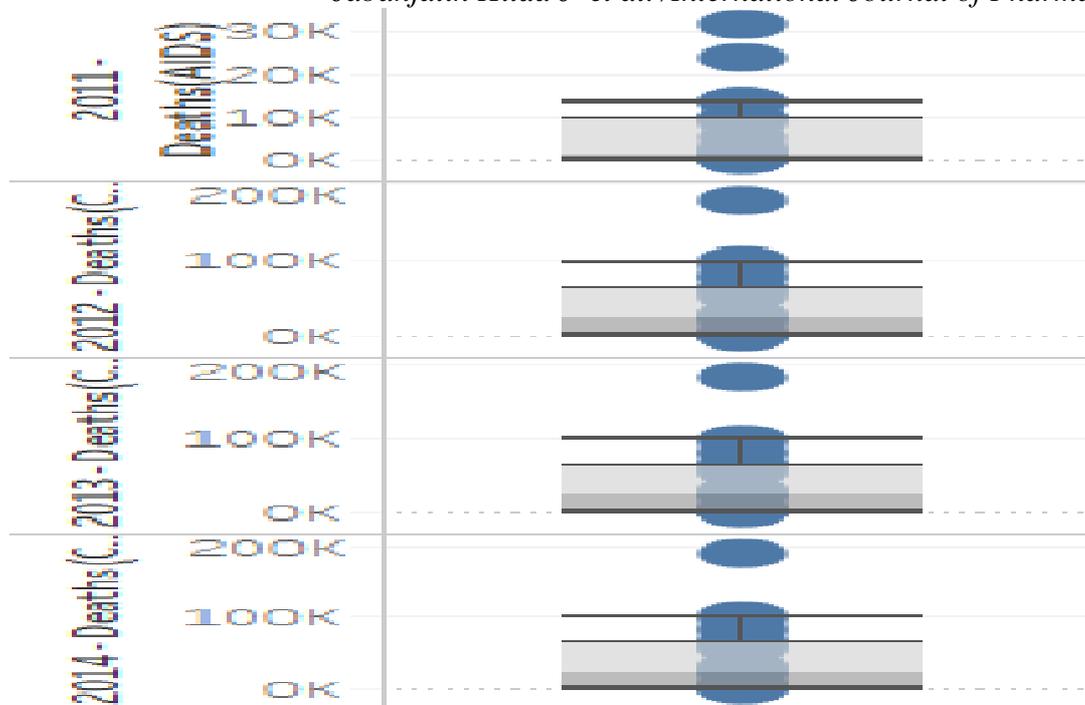
2010-11 Deaths(AIDS)



State/UT. Color shows sum of 2011 - Deaths(AIDS). Size shows sum of 2010 - Deaths(AIDS). The marks are labeled by State/UT. The view is filtered on State/UT, which excludes India.

**Figure 11: Tree Maps graph for the deaths caused by AIDS.**

The next graph which we are going to deal with would be Box and Whisker Plot graph. We are plotting this by comparing and merging two mainly death causing diseases Cancer and AIDS. Here the box plot contains in the form of least value, highest value, mean, median and some quartiles values which illustrate the values in a Box Plot graph.



Sum of 2007 - Deaths(AIDS), sum of 2008 - Deaths(AIDS), sum of 2009 - Deaths(AIDS), sum of 2010 - Deaths(AIDS), sum of 2011 - Deaths(AIDS), sum of 2012 - Deaths(Cancer), sum of 2013 - Deaths(Cancer) and sum of 2014 - Deaths(Cancer). Details are shown for State/UT.

**Figure 12: Box and Whisker Plot graph for the deaths caused by AIDS.**

**4. Conclusion:**

As we have seen various comparisons made and the various visualizations of the death rates in various states in India, by the various comparisons made we can evaluate multiple understandings and insight of the data with these visualizations, if not for the visualization we couldn't have an insight of the data in a much better way leading us to take uninformed decisions which will adversely lead us to un-efficient decisions. Hence whenever a person has a problem relating to these diseases, they can refer to these visualizations and come up with a decision as in where should he be heading to for the best treatment possible. In general wherever the disease is prominent the facilities or quality of treatment is efficient over there.

End of the day these visualizations are helping the general public to take informed and knowledgeable decisions regarding their health ensuring safety and good health.

**5. Acknowledgements:**

We would like to acknowledge Professor Jabanjalin Hilda J, without whom we would not have completed this paper. Her guidance and support was truly instrumental in the successful execution of our paper. We would also like to thank VIT University, Vellore for providing us with necessary resources.

**6. References:**

1. Baker DW, Einstadter D, Thomas CL, Husak SS, Gordon NH, Cebul RD. Mortality trends during a program that publicly reported hospital performance. *Medical care*. 2002 Oct 1;40(10):879-90.
2. Sacks JA, Zehe E, Redick C, Bah A, Cowger K, Camara M, Diallo A, Gigo AN, Dhillon RS, Liu A. Introduction of mobile health tools to support Ebola surveillance and contact tracing in Guinea. *Global Health: Science and Practice*. 2015 Dec 1;3(4):646-59.
3. Ellsworth MA, Lang TR, Pickering BW, Herasevich V. Clinical data needs in the neonatal intensive care unit electronic medical record. *BMC medical informatics and decision making*. 2014 Oct 24;14(1):1.
4. Ritt E, Taylor NP. The identification of post-acute patients at risk for hospital readmission: Clinical implications of the LACE index. *Clinical Nursing Studies*. 2016 Jun 7;4(3):p32.
5. Hilda JJ, Srimathi C, Bonthu B. A review on the development of big data analytics and effective data visualization techniques in the context of massive and multidimensional data. *Indian Journal of Science and Technology*. 2016 Jul 28;9(27).
6. Wattenberg M. Visual exploration of multivariate graphs. In *Proceedings of the SIGCHI conference on Human Factors in computing systems 2006 Apr 22 (pp. 811-819)*. ACM.
7. Mates P, Santos E, Freire J, Silva CT. Crowdlabs: Social analysis and visualization for the sciences. In *International Conference on Scientific and Statistical Database Management 2011 Jul 20 (pp. 555-564)*. Springer Berlin Heidelberg.