BIGDATA VISUALIZATION USING GOLANG SERVICES

A Project Report Submitted

in Partial Fulfilment of the Requirements

for the Degree of

B.tech(Hon)

in

Computer Science

by

S.Sai Siddardha

(Roll No. 2015BCS0032)

to

DEPARTMENT OF COMPUTER SCIENCE

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

KOTTAYAM - 695017, INDIA

April 2018

DECLARATION

I, **S.Sai Siddardha** (**Roll No: 2015BCS0032**), hereby declare that, this report entitled **"BIGDATA VISUALIZATION USING GOLANG SERVICES"** submitted to Indian Institute of Information Technology Kottayam towards partial requirement of **Bachelor of Technology(Hon)** in **Computer Science** is an original work carried out by me under the supervision of **Dr.Shajulin Benedict** and has not formed the basis for the award of any degree or diploma, in this or any other institution or university. I have sincerely tried to uphold the academic ethics and honesty. Whenever an external information or statement or result is used then, that have been duly acknowledged and cited.

Thiruvananthapuram - 695 017 **S.Sai Siddardha**

April 2018

CERTIFICATE

This is to certify that the work contained in this project report entitled **"BIGDATA VISUALIZATION USING GOLANG SERVICES"** submitted by **S.sai Siddardha** (**Roll No: 2015BCS0032**) to Indian Institute of Information Technology Kottayam towards partial requirement of **Bachelor of Technology(Hon)** in **Indian Institute Of Information Technology,Kottayam** has been carried out by him under my supervision and that it has not been submitted elsewhere for the award of any degree.

Thiruvananthapuram - 695
April 2018

(Dr. Shajulin Benedict)
Project Supervisor

ABSTRACT

Almost all research sectors (business sector, education sector and other sectors) might lead to a bigdata problem due to huge data. Data visualization helps users to quickly identify interesting and significant events/patterns from data that are otherwise too detailed or complex to discern. Users find difficult to interpret such large data. In this report, we discussed about the challenges and opportunities for visualizing bigdata. It discloses the modern techniques and solutions applied for visualizing bigdata in various sectors. It showcases the most important charts used for bigdata visualization like composition charts, comparison charts, relationship charts, distribution charts. And, the experiments were conduct to prove bigdata visualization.

Contents

Chapter 1

Introduction

Bigdata is also a data but with huge size. Bigdata contains both structured and unstructured data exponentially increasing with time. Big data is a term that describes the large volume of data both structured and unstructured that inundates a business on a day-to-day basis. But its not the amount of data thats important. Its what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves.

The main problem with bigdata is with this large amounts of data the user cannot choose data that Is useful to him.so this visualization helps to see the data clearly and he/she can predict data. Bigdata contains both structured ad unstructured data. Structured data can be easily and interpreted in database. But unstructured data cannot be easily analyzed and interpreted as like as structured data.

Today businesses struggle to just store the massive amount of data whereas analyzing, interpreting and presenting it in meaningful ways is a thought for

later. The main challenge of Big Data lies in capturing, storing, analyzing, sharing, searching, and visualizing data. One of the major aspect of Big Data analysis is that we can find interesting pattern in huge data set, but actually the result of the analysis is usually raw numbers and by those numbers it is very difficult to interpret anything. But if those numbers are represented visually then it becomes much easier for our brain to find meaningful patterns and take decision accordingly.

In recent years bigdata is growing very rapidly because of growth of social, cloud and multimedia computing. As we know traditional systems cannot store and process data companies turning into bigdata management solutions to convert unstructured data into structured data useful for the companies.

1.1 Bigdata Benefits

The benefits of the bigdata include: Identifying the root causes of failures and issues in real time Fully understanding the potential of data-driven marketing Generating customer offers based on their buying habits increasing custmor loyalty. Reevaluating risk portfolios quickly Personalizing custmor experience. Adding value to online and offline customer interactions.

1.2 Big Data in Todays Business and Technology Environment)

Definition 1.2.1. 2.7 Zetabytes of data exist in the digital universe today. 235 Terabytes of data has been collected by the U.S. Library of Congress

in April 2011. The Obama administration is investing 200 million in big data research projects. IDC Estimates that by 2020,business transactions on the internet- business-to-business and business-to-consumer will reach 450 billion per day. Facebook stores, accesses, and analyzes 30+ Petabytes of user generated data. Akamai analyzes 75 million events per day to better target advertisements. 94 percent of Hadoop users perform analytics on large volumes of data not possible before; 88percent analyze data in greater detail; while 82percent can now retain more of their data. Walmart handles more than 1 million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes of data.

1.3 The Rapid Growth of Unstructured Data

YouTube users upload 48 hours of new video every minute of the day. 571 new websites are created every minute of the day. Brands and organizations on Facebook receive 34,722 Likes every minute of the day. 100 terabytes of data uploaded daily to Facebook. According to Twitters own research in early 2012, it sees roughly 175 million tweets every day, and has more than 465 million accounts. 30 Billion pieces of content shared on Facebook every month.

1.4 why bigdata visualization:

Big data visualization means implementation of visualization techniques to show contemporary relation between two variables in a dataset. If we want to

know relationship between two variables in a dataset which consists of large data it is very difficult to compare those two variables by checking every row but by using some visualization techniques if we can plot the data in graphs like bar chart, scatter chart, pie chart then we can easily predict the data by seeing in a graph.

Some of the bigdata visualization techniques are like world cloud, symbol maps, connectivity charts, line charts, bar charts, map charts, 3D grpahs. We can visualize our data either in Tables or Graphs. Graphs and tables serve different purposes. Choose the appropriate data display to fit your purpose. Common Display Types Bar plots, Box plots, Heatmaps, Tree plot, Pie chart Scatterplots and in many different ways we can visualize our data. Different kinds of visualization techniques are there for Big data and small data. Heatmap and Scatterplot some of the basic plot examples ..

1.5 objectives of bigdata visualization

The objectives of bigdata visualization are:

1). the main objective of big data visualization is to show the graph elegantly and clearly.

2). The improved custmor visualization of data.

3). Better operational efficiency.

4). Every company uses data in its on way. the more the company uses the data the larger it will grow.

1.6 Challenges Of Bigdata Visualization:

Traditional visualization tools have reached to their limits when encountered with very large data sets and these data are evolving continuously. Though there are some extensions to traditional visualization approaches but they lag behind by miles. The visualization tool should be able to provide us interactive visualization with as low latency as possible. To reduce the latency, we can do the following things: Use the pre-computed data Parallelize Data Processing and Rendering Use a predictive middleware Big Data visualization tool must be able to deal with semi-structured and unstructured data because big data usually have this type of format. It is realized that to cope with such huge amount of data there is need for immense parallelization, which is a challenge in visualization. The challenge in parallelization algorithm is to break down the problem into such independent task that they can run independently.

The main challenges of bigdata visualization are:

1). Data storage and quality.

2). People who understand Bigdata analysis.

3). Good quality analysis.

4). Security and Privacy of data.

5). Various Sources of data.

Chapter 2

LITERATURE SURVEY

Data visualization is the main focusing concept in big data analysis for processing and analyzing multi variate data, because of rapid growth of data size and complexity of data. Basically data visualization may achieve three main problems, i.e.

1. Structured and Unstructured pattern evaluation in big data analysis.
2. Shrink the attributes in data indexed big data analysis.
3. Rearrange of attributes in parallel index based data storage.

Big Data analytics plays a key role through reducing the data size and complexity in Big Data applications. Visualization is an important approach to helping Big Data get a complete view of data and discover data values. Big Data analytics and visualization should be integrated seamlessly so that they work best in Big Data applications.

Data visualization transforms data into images to aid the understanding of data; therefore, it is an invaluable tool for explaining the significance of data to visually inclined people. Given a (big) dataset, the essential task of

visualization is to visualize the data to tell compelling stories by selecting, filtering, and transforming the data, and picking the right visualization type such as bar charts or line charts. Our ultimate goal is to automate this task that currently requires heavy user intervention in the existing visualization systems.

An evolutionized system in the field faces the following three main challenges:

(1) Visualization verification: to determine whether a visualization for a given dataset is interesting, from the viewpoint of human understanding.

(2) Visualization search space: a "boring" dataset may become interesting after an arbitrary combination of operations such as selections, joins, and aggregations, among others.

(3) On-time responses: do not deplete the user's patience.

Our services are different from other services like given a bigdata set ant the essential form is visualizing by selecting right chart(composition chart, comparison chart, relationship charts, distribution charts) and plot it by using some R packages and show it in by golang services.

2.1 Bigdata Visualization Tools

Various tools have emerged to help us out from the above pointed problems. The most important feature that a visualization must have is that it should be interactive, which means that user should be able to interact with the visualization. Visualization must display relevant information when hovered over it, zoom in and out panel should be there, visualization should adapt

itself at runtime if we select subset or superset of data.

some of the important bigdata visualization tools are:

1)Tableau:

Tableau is interactive data visualization tool which is focused on Business Intelligence. Tableau provides very wide range of visualization options. It provides option to create custom visualization. It is fast and flexible. It supports mostly all the data format and connection to various servers right from the Amazon Aurora to Cloudera Hadoop and Salesforce. User interface is intuitive, wide variety of charts are available. For simple calculations and statistics one does not require any coding skills but for heavy analytics we can run models in R and then import the results into Tableau. This requires quite a bit of programming skill based upon the task we need to perform.

2)Microsoft Power BI:

Power BI is a powerful cloud-base business analytics service. Visualization are interactive and rich. Power BI consists of 3 elements, Power BI Desktop,Service(SaaS), Apps. Every service is available to us that is why it makes Power BI flexible and persuasive. With more than 60 types of source integration you can start creating visualization in matter of minutes. Power BI combines the familiar Microsoft tools like Office, SharePoint and SQL Server. The feature that it distinguishes from other tools is that you can use natural language to query the data. You dont require programming skills for this tool but there is option available to run your R script.

3)Plotly:

Plotly is also known as Plot.ly is build using python and Django framework. The actions it can perform are analyzing and visualizing data. It is

free for users but with limited features, for all the features we need to buy the professional membership. It creates charts and dashboards online but can be used as offline service inside Ipython notebook, jupyter notebook and panda. Different variety of charts are available like statistical chart, scientific charts, 3D charts, multiple axes, dashboards etc. Plotly uses a tool called Web Plot Digitizer(WPD) which automatically grabs the data from the static image.

4)Gephi:

Gephi is open-source network analysis tool written in Java and OpenGL. It is used to handle very large and complex datasets. The network analysis includes Social Network Analysis Link Analysis Biological Network Analysis With its dynamic data exploration Gephi stands out rest of its competition for graph analysis. No programming skills are required to run thin tools but a good knowledge in graphs is necessary. It uses GPU 3D render engine to accelerate the performance and give real time analysis.

Chapter 3

PROPOSED ARCHITECTURE AND CHARTS

There are four important ways you can present the bigdata the ways are:

1) Composition charts

2)Comparison charts

3)Relationship charts

4)Distribution charts

To determine which chart is best suited for each of those presentation types, first you must answer a few questions:

How many variables do you want to show in a single chart? One, two, three, many?

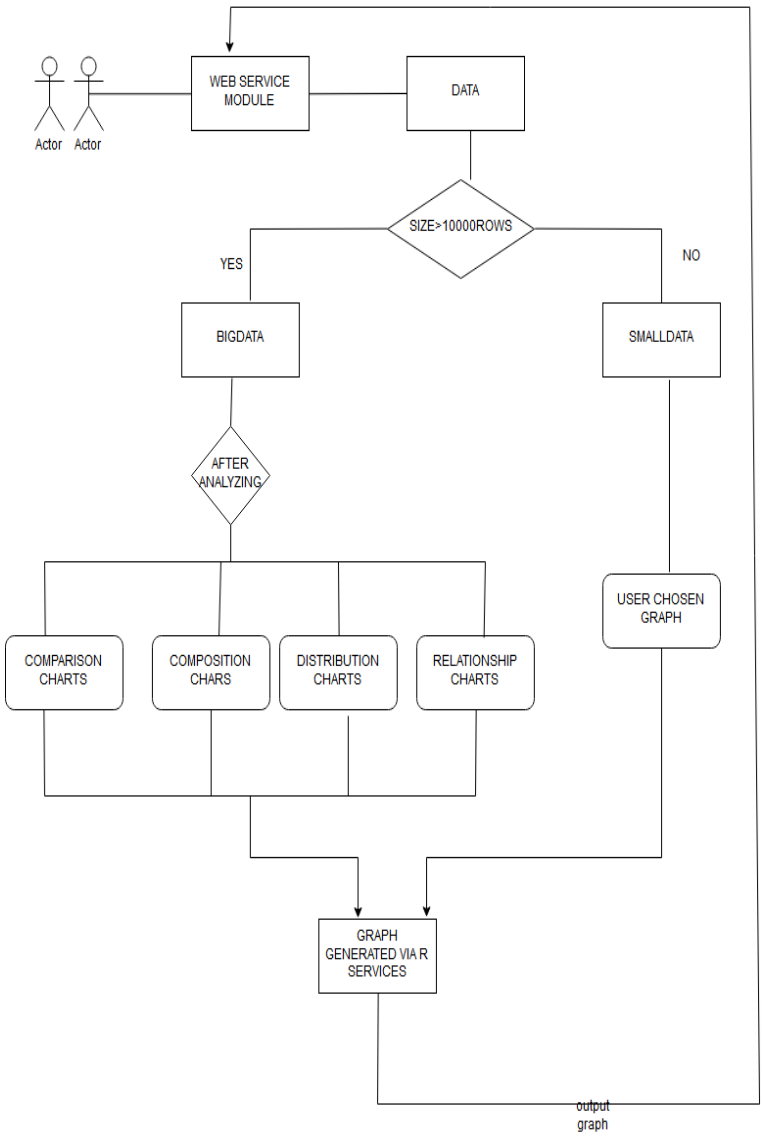
How many items (data points) will you display for each variable? Only a few or many?

Will you display values over a period of time, or among items or groups?

As every chart plots the bigdata some graphs are good for some type. Like Bar charts are good for comparisons, while line charts work better for trends. Scatter plot charts are good for relationships and distributions, but pie charts should be used only for simple compositions never for comparisons or distributions. These are some restriction we need to follow.

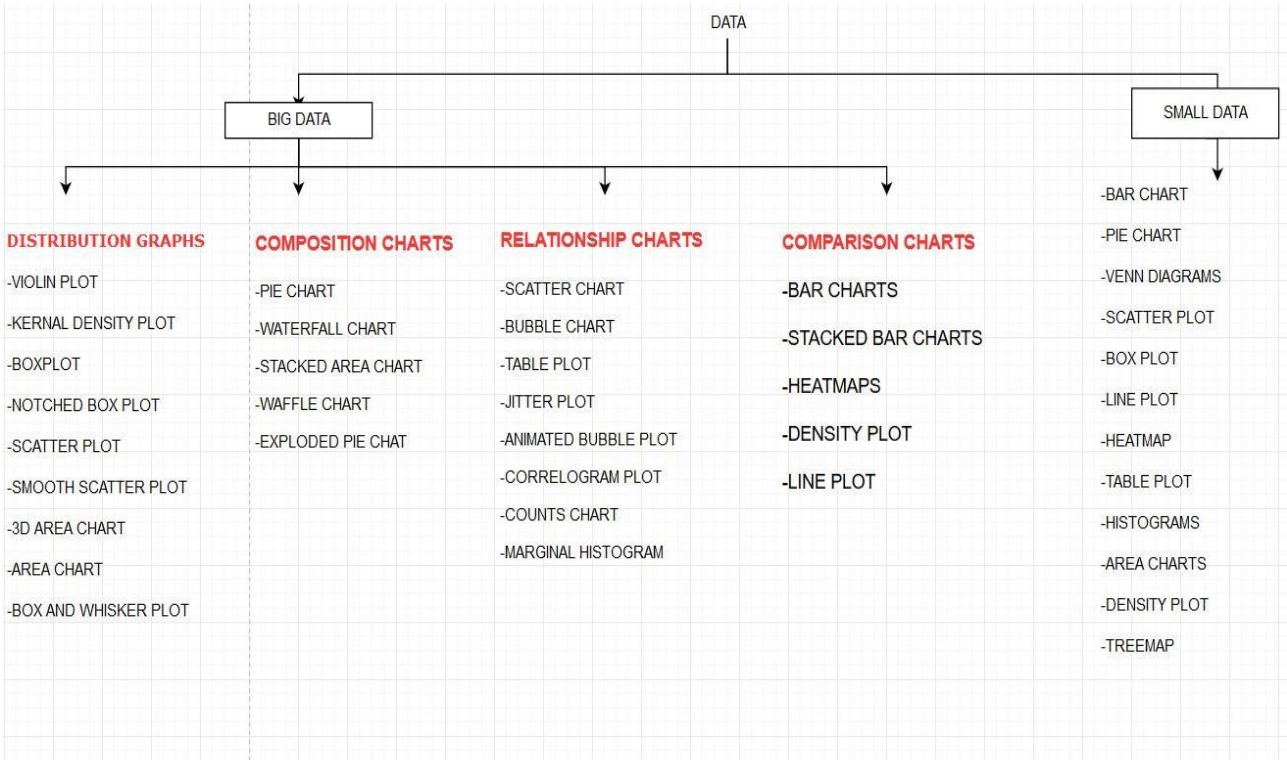
3.1 PROPOSED ARCHITECTURE

Based on above requirements the architecture will be like this. At first the user will input the data in web service module after that based on number of rows in data it will split data either like as small data or bigdata. If the data is bigdata the user will choose any graphs either like composition charts or comparison charts or distribution charts or relationship charts after that it will give a output graph and it will show it in webmodule. if it is small data the user can show it in any graph and it will be given output in wenbservice module.so the proposed architecture will be:


Actor
Actor
WEB SERVICE MODULE
DATA
SIZE>10000ROWS
YES
NO
BIGDATA
SMALLDATA
AFTER ANALYZING
USER CHOSEN GRAPH
COMPARISON CHARTS
COMPOSITION CHARS
DISTRIBUTION CHARTS
RELATIONSHIP CHARTS
GRAPH GENERATED VIA R SERVICES
output graph

3.2 PROPOSED TAXONOMY

In bigdata every chart contains some graphs.This taxonomy contains the graphs that every charts contains:



3.3 COMPOSITION CHARTS:

Composition of data is probably the most misused method in data representation endeavors. The idea is to show how individual parts make up the whole by combining them together and displaying them as a sum. Composition can also be used to show how a total value can be divided into parts or to highlight the significance of each part relative to the total value. Use data composition charts to show Company market share and a few key players in the market Total country population by TOP religions, languages, or ethnical groups Total revenue, by TOP product lines, divisions, or regions The more important composition charts include:

Area chart:

Area chart is used to show continuity across a variable or data set. It is very much same as line chart and is commonly used for time series plots.

Bar chart:

A bar chart represents data in rectangular bars with length of the bar proportional to the value of the variable.

Pie Chart:

A pie-chart is a representation of values as slices of a circle with different colors.

Stacked area chart:

An area chart displays the development of quantitative values over an interval. It resembles a line chart as it uses lines connecting data points with each other.

Waffle Chart:

Waffle charts is a nice way of showing the categorical composition of the

total population.

Scatter chart:

Scatter Plot is used to see the relationship between two continuous variables.

3.4 COMPARISON CHARTS:

Comparison of data points is probably the most common and easy-to-understand method for data analysis. As the name suggests, we use comparison to evaluate and compare values between two or more data points. With comparison you can also easily find the lowest and highest values in the chart.

Usually comparisons are made to accomplish one of the following goals: To list key values to quickly find and read them (i.e., revenue per month) To rank several data categories from best to worst or the other way around To show pattern recognition by visually highlighting gaps, spikes, outliers, or trends

Some of the important comparison charts are:

Bar Plot :

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent.

Density Plot : Density plots are usually a much more effective way to view the distribution of a variable.

Heatmap :

A heat map is a graphical representation of data where the individual

values contained in a matrix are represented as colors. It is a bit like looking a data table from above.

Line Charts :

A line chart is a graph that connects a series of points by drawing line segments between them. These points are ordered in one of their coordinate (usually the x-coordinate) value. Line charts are usually used in identifying the trends in data.

3.5 RELATIONSHIP CHARTS

These types of charts show the relationship, correlation, or connection of two or more variables and their properties. A good use of relationship graphs would be to demonstrate how something does or does not affect another variable positively or negatively.

Use a relationship chart to: Spot flaws in effectiveness by evaluating expenses vs. income by store or region. pot flaws in effectiveness by evaluating expenses vs. income by store or region.

Some of the important relationship charts are:

Scatter Plot with Encircling :

Scatter Plot is used to see the relationship between two continuous variables. It is type of scatter plot that encircles set of points.

Jitter Chart :

It is a type in scatter plot with overlapping values.jo jitter chart uses the same function as scatter plot.

Counts Chart :

To overcome the problem of data points overlap is to use what is called a counts chart.

Bubble Plot :

While scatterplot lets you compare the relationship between 2 continuous variables, bubble chart serves well if you want to understand relationship within the underlying groups based on: 1. A Categorical variable (by changing the color)

2. Another continuous variable (by changing the size of points).

Animated Bubble chart :

It is same as the bubble chart, but, you have to show how the values change over a fifth dimension (typically time).

Correlogram :

Correlogram is used to test the level of co-relation among the variable available in the data set. The cells of the matrix can be shaded or colored to show the co-relation value.

Marginal Histogram :

If you want to show the relationship as well as the distribution in the same chart, use the marginal histogram. It has a histogram of the X and Y variables at the margins of the scatterplot.

Table Plot :

The table plot is a visualization method that is used to explore and analyse large datasets. Table plots are used to explore the relationships between the variables, to discover strange data patterns, and to check the occurrence and selectivity of missing values.

3.6 DISTRIBUTION CHARTS:

Area Chart :

Area chart is used to show continuity across a variable or data set. It is very much same as line chart and is commonly used for time series plots. Alternatively, it is also used to plot continuous variables and analyze the underlying trends.

Box Plot :

Box plot is an excellent tool to study the distribution. It can also show the distributions within multiple groups, along with the median, range and outliers if any.

Dot+Box Plot :

On top of the information provided by a box plot, the dot plot can provide more clear information in the form of summary statistics by each group. The dots are staggered such that each dot represents one observation. So, in below chart, the number of dots for a given manufacturer will match the number of rows of that manufacturer in source data.

Candle Stick Plot :

It is like a combination of line-chart and a bar-chart: each bar represents all four important pieces of information for that day: the open, the close, the high and the low. Being densely packed with information, they tend to represent trading patterns over short periods of time, often a few days or a few trading sessions.

Kernal Density Plot :

Kernal density plots are usually a much more effective way to view the distribution of a variable.

Notched Box Plot :

The boxplot compactly displays the distribution of a continuous variable. It visualizes five summary statistics (the median, two hinges and two whiskers), and all "outlying" points individually .for a notched box plot.

Stripe Plot :

A strip plot is a graphical data analysis technique for summarizing a univariate data set. The strip plot consists of: Horizontal axis = the value of the response variable;

Vertical axis = all values are set to 1.

Tufte Box Plot :

Tuftes Box plot is just a box plot made minimal and visually appealing.

Violin Plot :

A violin plot is a method of plotting numeric data. It is similar to box plot with a rotated kernel density plot on each side. The violin plot is similar to box plots.

Now after understanding everything about graphs we can plot every graph using some R packages. Now using golang as service every plot mentioned above are plotted in localhost. As most graphs use ggplot2 package the packages that this project include:

Ggplot2

Hexbin

Ggthemes

Tabplot

Ggcorrplot

ggExtra

ggalt

hexbin

RColorBrewer

Magrittr

Leaflet

Chapter 4

EXPERIMENTAL RESULTS

Data visualization is a widely studied field; however,data-driven automatic data visualization has not yet been extensively studied. Big Data analytics plays a key role through reducing the data size and complexity in Big Data applications. Visualization is an important approach to helping Big Data get a complete view of data and discover data values.

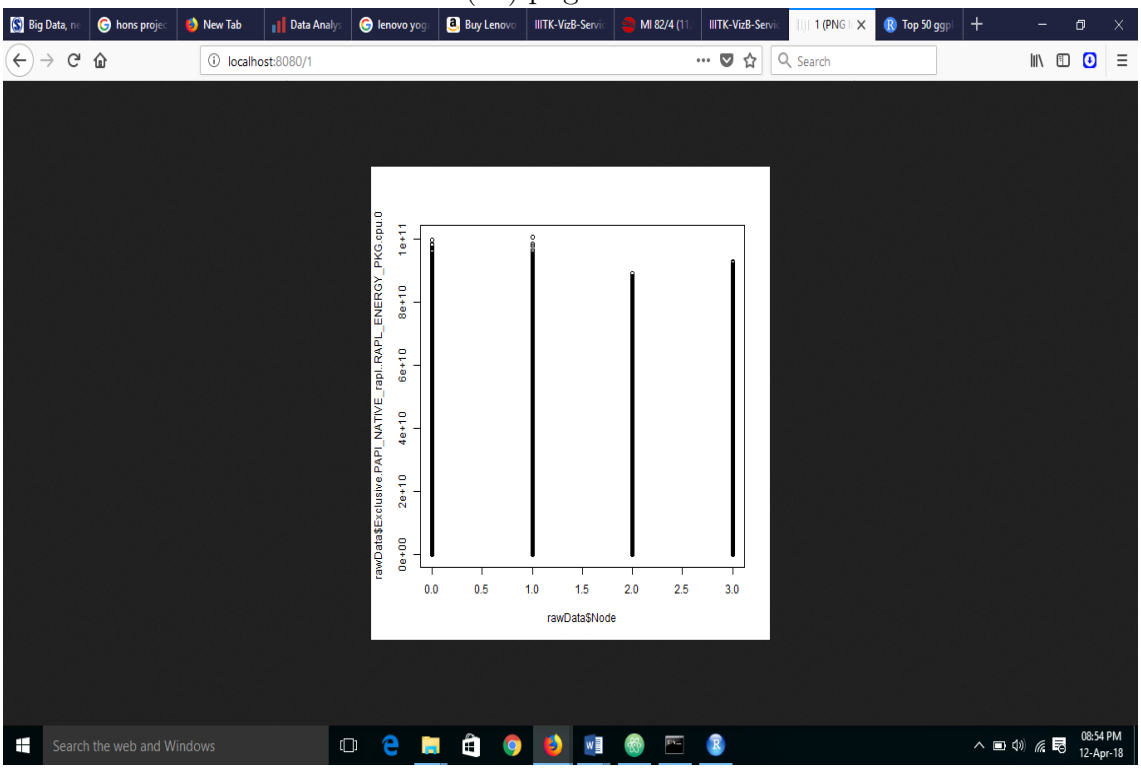
Interactive visualizations often lead to discovery and do a better job than static data tools. Do interactive visualization for more analyzing of data. Interactive brushing and linking between visualization approaches and networks or Web-based tools can facilitate the scientific process. Web-based visualization helps get dynamic data timely and keep visualizations up to date.Here are the some of the visualizations we have practiced:

4.1 Scatter Chart

The most frequently used plot for data analysis is undoubtedly the scatterplot. Whenever you want to understand the nature of relationship between two variables, invariably the first choice is the scatterplot.

It can be drawn using geompoint()

(30).png



4.1 Scatter chart for bigdata

4.2 BOX Chart

Box plot is an excellent tool to study the distribution. It can also show the distributions within multiple groups, along with the median, range and outliers if any.

(31).png

4.2 BOX chart for bigdata

4.3 Dot+BOX Chart

a box plot, the dot plot can provide more clear information in the form of summary statistics by each group. The dots are staggered such that each dot represents one observation.

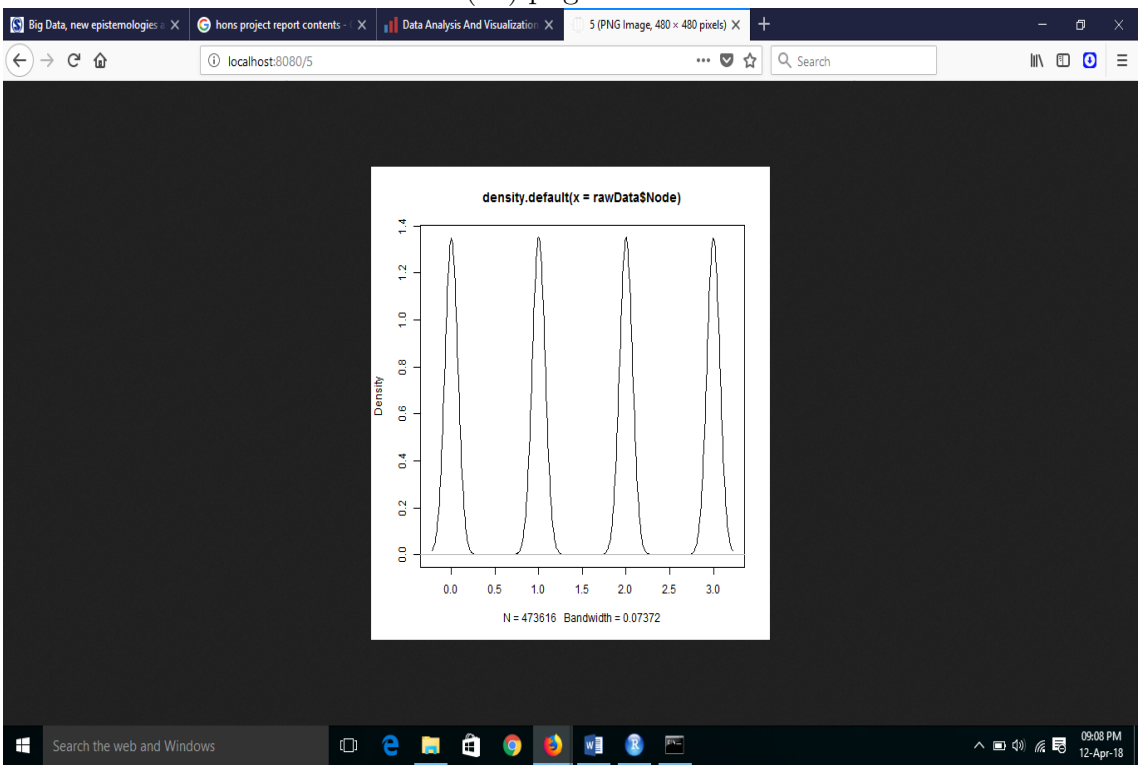
(32).png

4.3 DOT+BOXr chart for bigdata

4.4 KERNALDENSITY Chart

kernal density plot an excellent tool to study the distribution. It can also show the distributions within multiple groups, along with the median, range and outliers if any.

(33).png

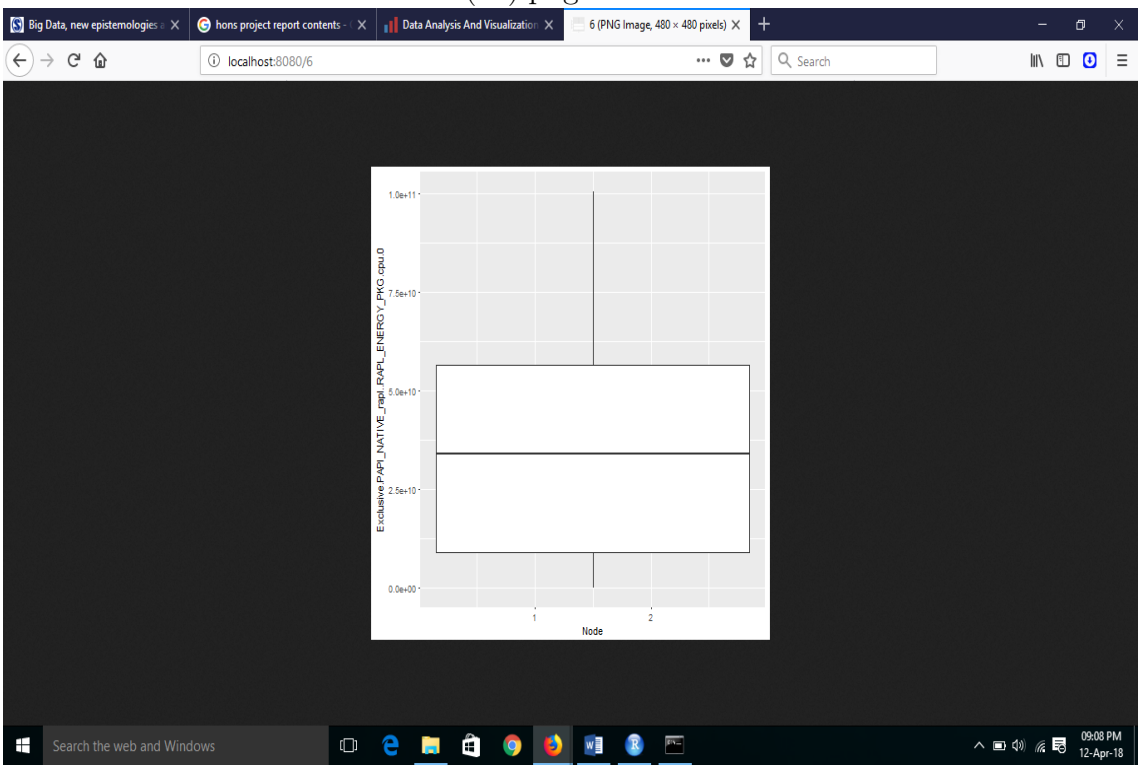


4.4 KERNALDENSITY chart for bigdata

4.5 NOTCHEDBOX Chart

As notched areas do not overlap across box plots, we may expect that the three medians are different from one another. ... Box (vertical) thickness represents the InterQuartile Range (IQR) and allows to visualize the dispersion of 50.

(34).png

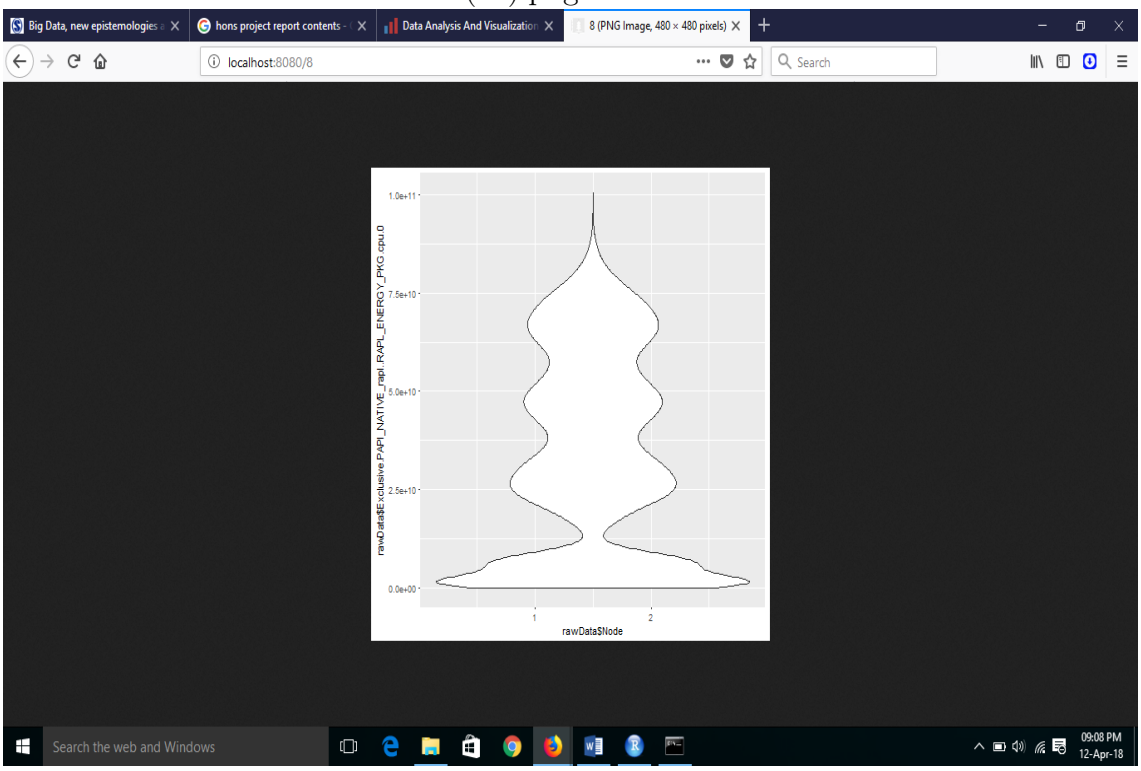


4.5 NOTCHEDBOX chart for bigdata

4.6 Violin Chart

A violin plot is similar to box plot but shows the density within groups. Not much info provided as in boxplots.

(36).png



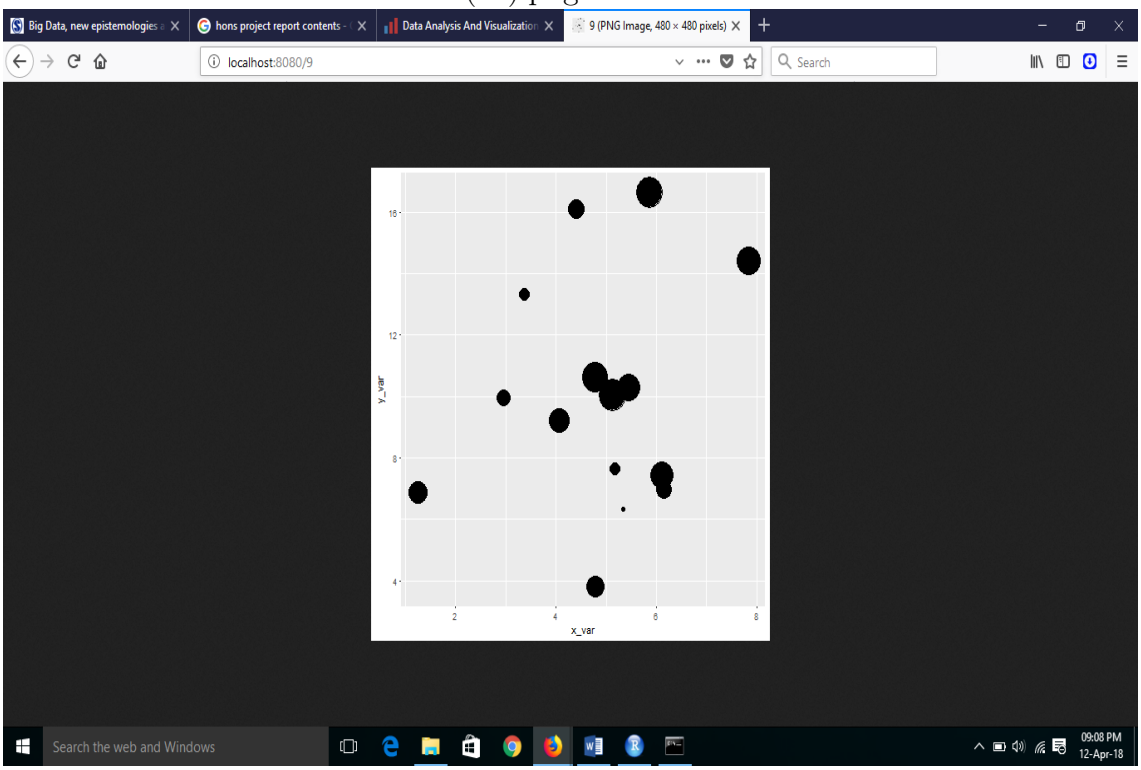
4.6 VIOLIN chart for bigdata

4.7 BUBBLE Chart

While scatterplot lets you compare the relationship between 2 continuous variables, bubble chart serves well if you want to understand relationship within the underlying groups based on:

1). A Categorical variable (by changing the color) and 2). Another continuous variable (by changing the size of points).

(37).png

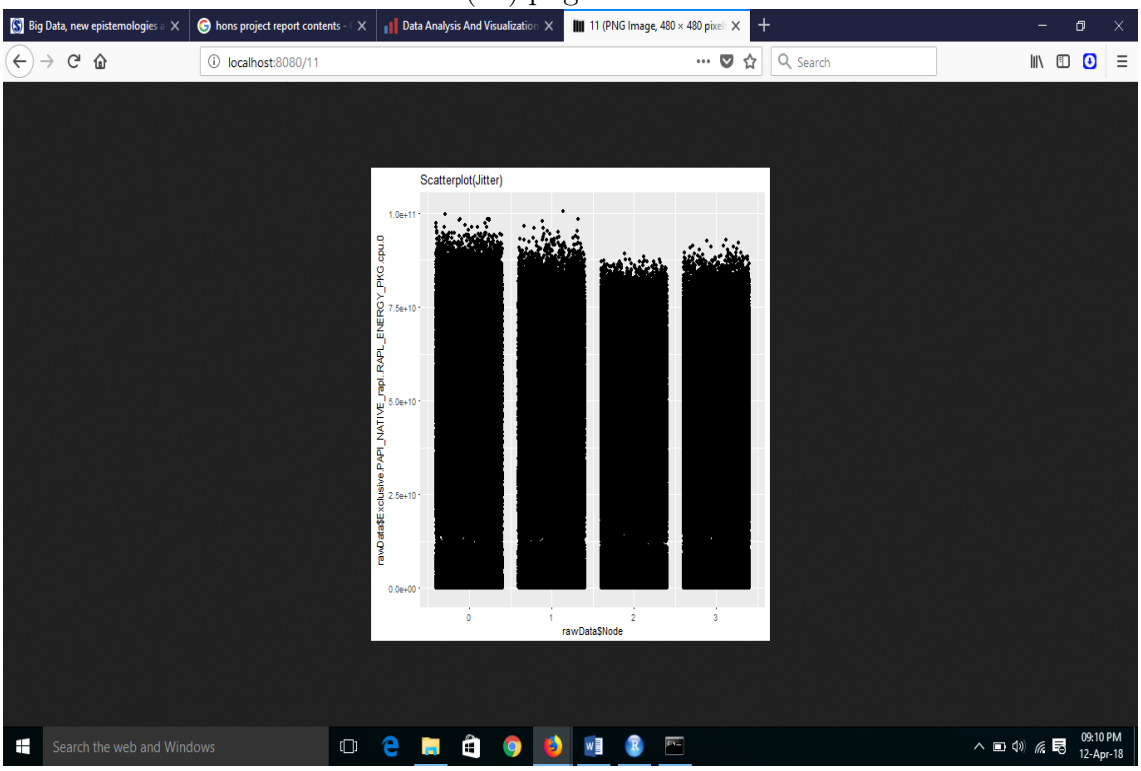


4.7 BUBBLE chart for bigdata

4.8 JITTER Chart

It is a type in scatter plot with overlapping values.jo jitter chart uses the same function as scatter plot.

(39).png

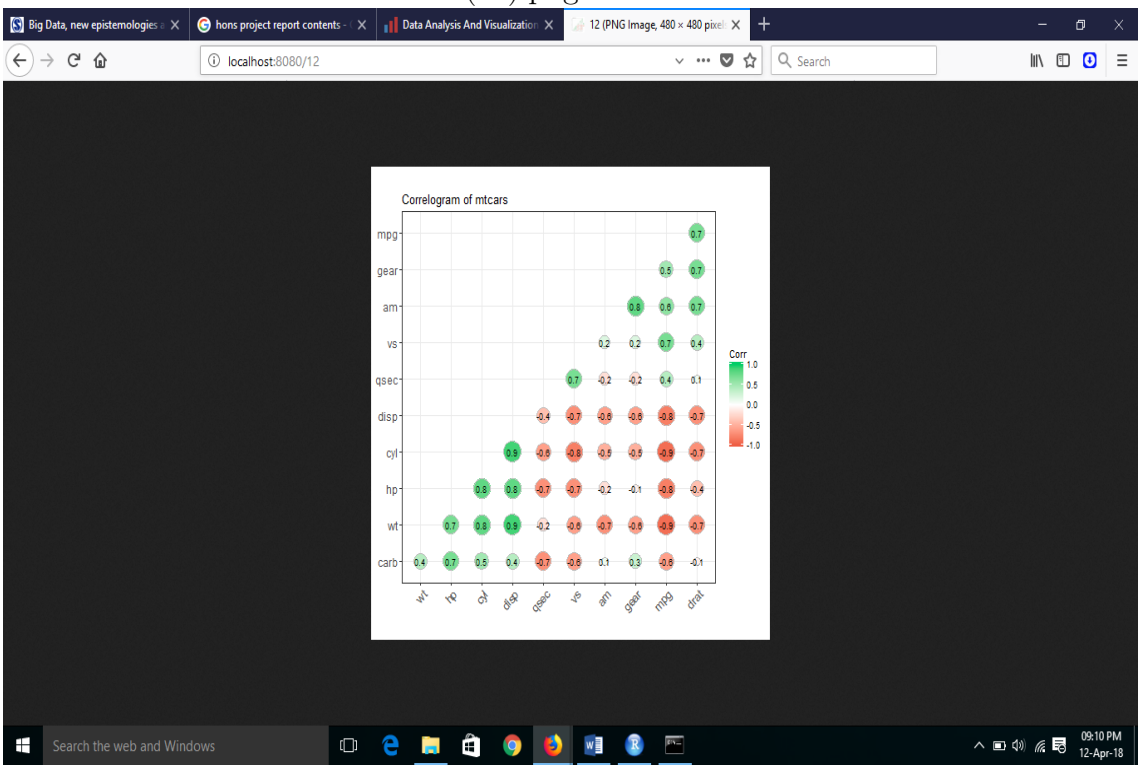


4.8 JITTER chart for bigdata

4.9 Correlogram Chart

Correlogram lets you examine the corellation of multiple continuous variables present in the same dataframe. This is conveniently implemented using the ggcorrplot package.

(40).png



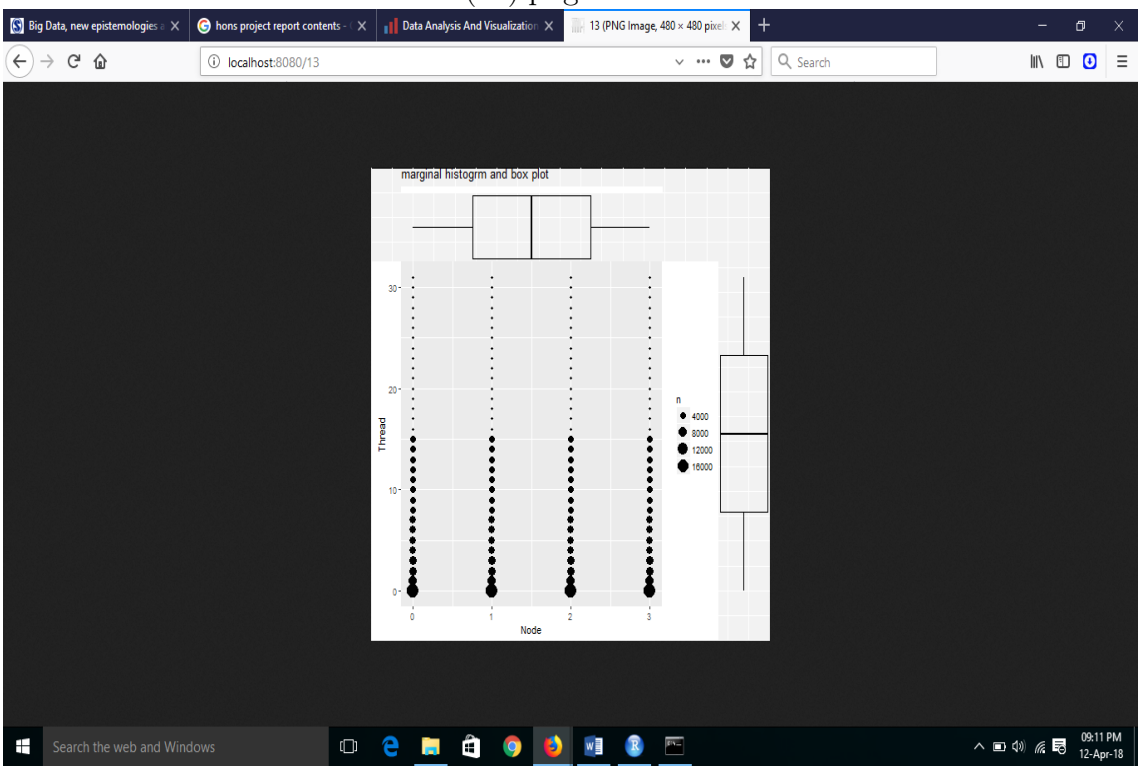
4.9 CORRELOGRAM chart for bigdata

4.10 Marginal histogram+box Chart

If you want to show the relationship as well as the distribution in the same chart, use the marginal histogram. It has a histogram of the X and Y variables at the margins of the scatterplot.

This can be implemented using the ggMarginal() function from the ggExtra package.

(41).png

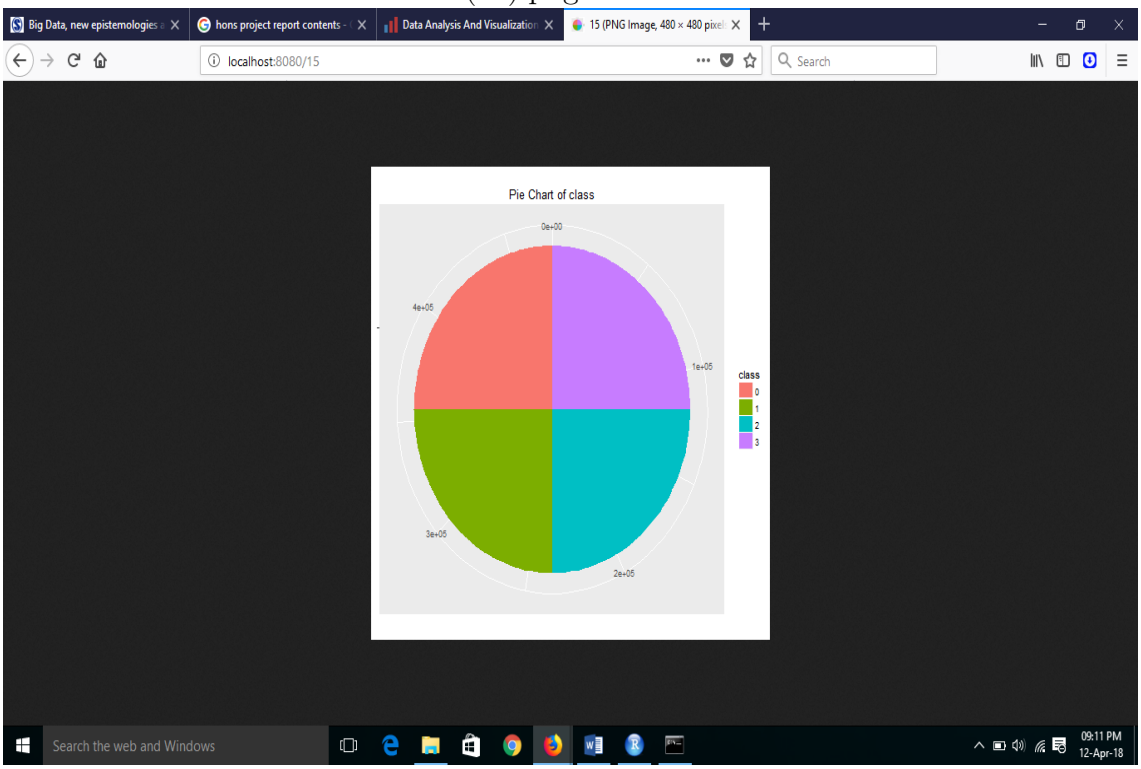


4.10 MARGINALHISTOGRAM+BOX CHART for bigdata

4.11 PIE Chart

Pie chart, a classic way of showing the compositions is equivalent to the waffle chart in terms of the information conveyed. But is a slightly tricky to implement in ggplot2 using the coordpolar()

(43).png

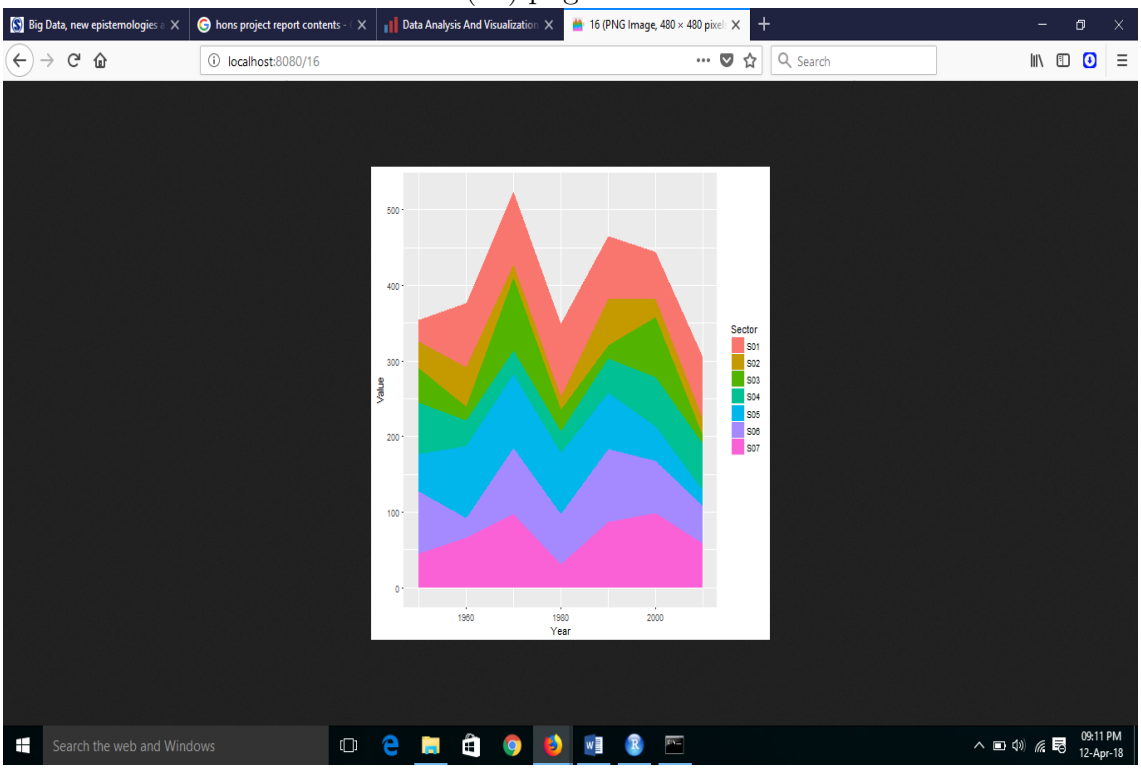


4.11 PIE chart for bigdata

4.12 STACKED AREA Chart

if you want to represent this evolution for several groups in the same time, you are probably interested by stacked area chart, where every groups are displayed one of top of each other.

(44).png



4.12 STACKED AREA chart for bigdata

4.13 WAFFLE Chart

Waffle charts is a nice way of showing the categorical composition of the total population. Though there is no direct function, it can be articulated by smartly maneuvering the ggplot2.

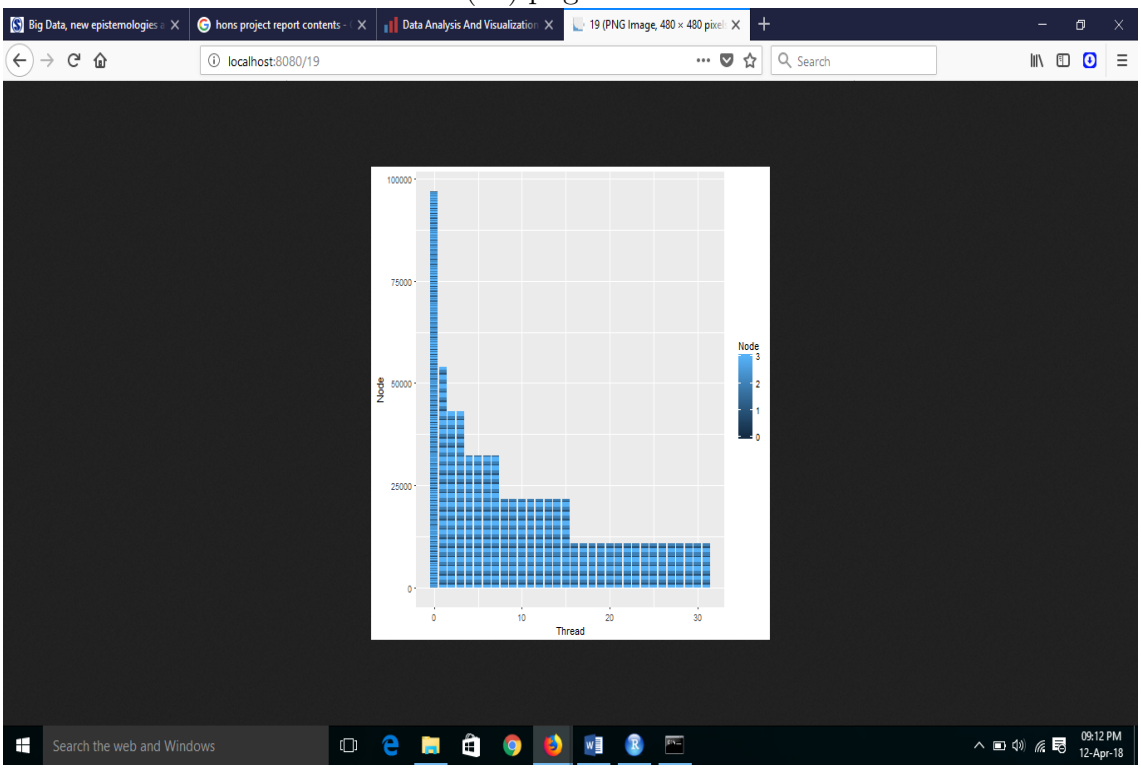
(45).png

4.13MWAFFLE chart for bigdata

4.14 BAR Chart

A bar chart can be drawn from a categorical column variable or from a separate frequency table. By adjusting width, you can adjust the thickness of the bars.

(47).png



4.14 BAR chart for bigdata

4.15 HISTOGRAM Chart

Histogram on a continuous variable can be accomplished using either geom bar() or geom histogram(). When using geom histogram(), you can control the number of bars using the bins option.

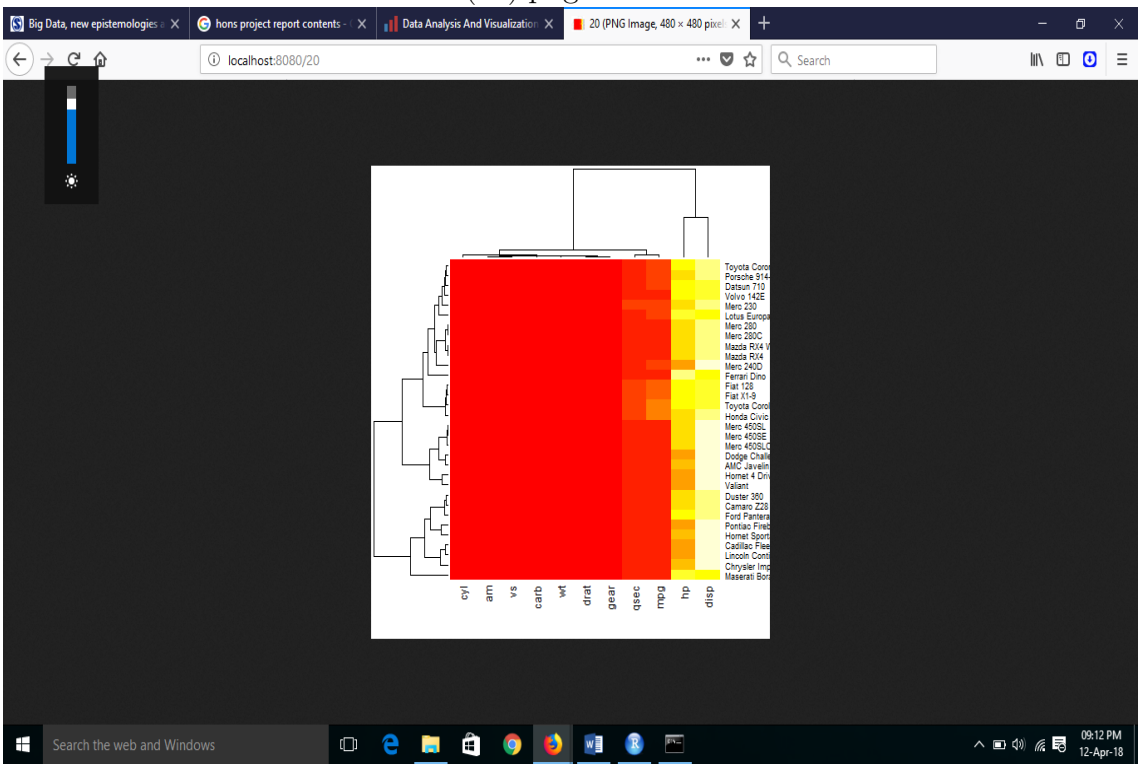
(46).png

4.15 HISTOGRAM chart for bigdata

4.16 HEAT Chart

A heat map (or heatmap) is a graphical representation of data where the individual values contained in a matrix are represented as colors.

(48).png



4.16 HEAT chart for bigdata

Chapter 5

conclusion and future work

From this study we have found that Visualizations can be static or dynamic. Interactive visualizations often lead to discovery and do a better job than static data tools. Do interactive visualization for more analyzing of data. Interactive brushing and linking between visualization approaches and networks or Web-based tools can facilitate the scientific process. Web-based visualization helps get dynamic data timely and keep visualizations up to date. More new methods and tools of Big Data visualization should be developed for different Big Data applications.

Big Data analytics and visualization can be integrated tightly to work best for Big Data applications. More new graphs and More new visualizations could try in the future for better enhancement of visualization. A lot of technologies are developing for better visualization of bigdata and in future this can be developed by asking the user to enter data and ask like to which points data should be plotted.

Bibliography

[1] https://www.rdocumentation.org/packages/ggalt/versions/0.1.1/topics/ggalt. *bigdata packages*.

[2] https://www.rdocumentation.org/packages/gganimate/versions/0.1.0.9000/topics/gganimate. *bigdata packages*.

[3] https://www.rdocumentation.org/packages/ggextra/versions/0.7/topics/ggextra. *bigdata packages*.

[4] https://www.statmethods.net/graphs/index.html. *bigdata packages*.

[5] Dai X Agrawal R, Kadadi A and Andres F. Challenges and opportunities with big data visualization. pages 169–173. ACM, 2015.

[6] Elizabeth Lawler Stacy Yeh Ashish Mahabal ; Anwell Wang, Jerry Zhang. Immersive and collaborative data visualization using virtual reality platforms. *Big Data (Big Data), 2014 IEEE International Conference*, 2014.

[7] Chi P Chen L Cai L, Guan X and Luo J. Big data visualization collaborative filtering algorithm based on rhadoop. International Journal of Distributed Sensor Networks, 2015.

[8] Intel IT Center. Big data visualization: Turning big data into big insights. *Bigdata visualization*, pages 1–14, March 2013.

[9] E.Y. Gorodov and V.V. Gubarev. Analytical review of data visualization methods in application to big data. Embry-Riddle Aeronautical University, Daytona Beach, USA, 2013. ELSEIVER Computer Society.

[10] H. Hauser H. Doleisch, M. Gasser. Interactive feature specification for focus+context visualization of complex simulation data. *IEEE Symposium on Visualization*, pages 239–248, 2015.

[11] Cheng X Jin X, Wah BW and Wang Y. Significance and challenges of big data research. *Bigdata research*, pages 1–6, 2015 june 30.

[12] Noah Iliinsky Julie Steele. *Beautiful Visualization, Looking at Data Through the Eyes of Experts*. march 2018.

[13] Remzi Sekerb Radu F. Babiceanua. Big data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. page 3, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA, 2016. ELSEIVER Computer Society.

[14] scrott berinato. *Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations*. march 2017.

[15] Duygu Sinanc Seref Sagiroglu. Bigdata analysis with interactive visualization using r packages. *Collaboration Technologies and Systems (CTS), 2013 International Conference*, 2013.

[16] stephen few. *Information Dashboard Design: Displaying Data for At-a-glance Monitoring*. 2016.

[17] J. Alberto Espinosa William Money Stephen Kaisler, Frank Armour. Big data: Issues and challenges moving forward.

[18] Nooper Gupta Syed mohd ali. Big data visualization: Tools and challenges. *Bigdata visualization*, pages 1–6, 2017.

[19] Edward R Tufte. *The Visual Display of Quantitative Information by Edward R.Tufte*. march 2018.

[20] Mohan Varma Wonhee Cho. Bigdata analysis with interactive visualization using r packages. *Bigdata visualization*, pages 1–7, 2015.

[21] NanTang GuoliangLi xeudiqin, YuyuLuo. *DeepEye: An automatic big data visualization framework*. IEEE Computer society, march 2018.