

DATA VISUALIZATION

Tableau and QlikView

Rajiya Begam Shaik

Manusha Oruganti

MISI- 799- Capstone

Ferris State University, 2017

Advisor:

Dr. Greg Goglin

Acknowledgement

We would like to thank our advisor, Dr. Greg Goglin and Ferris State University Writing Center for their continuous support and supervision for our Capstone Project.

Index

Chapter 1

Overview.....	7
Scientific Visualization.....	7
Information Visualization.....	7
Multivariate Data Visualization.....	8
Text Data Visualization.....	9
Temporal Data Visualization.....	9
Hierarchical Data Visualization.....	11
Network Data Visualization.....	12
Purpose Statement.....	13
Research Questions.....	13
Limitations.....	13

Chapter 2

Background.....	14
Data Science Cycle.....	15
Visualizing data using charts.....	15
• Line Chart.....	16

• Bar chart.....	17
• Stacked bar chart.....	18
• Scatterplot.....	19
• Area chart.....	20
• Box whisker plot.....	21
• Pie chart.....	22
• Donut chart.....	23
• Radar chart.....	24
Tools used.....	24
Environment Tools.....	25
Necessary Resources.....	25

Chapter 3

Summary.....	26
Primary Purpose.....	26
Why is data visualization important?	26
Statement of problems.....	27
Problem 1	27
Problem 2.....	27
Restating of Research questions.....	28

Question 1.....28

Question 2.....28

Question 3.....28

Question 4.....28

Chapter 4

Visualizing data using Tableau.....30

What is Tableau30

Why Tableau.....30

Maps in Tableau.....32

Symbol Maps.....32

Filled Maps.....33

Custom SQL Queries Tableau.....34

Sample SQL query.....34

Steps to visualize using SQL in Tableau.....35

Live data implementation Tableau.....38

Steps to implement live map using Tableau.....38

Visualizing Tree Maps using Tableau.....42

Steps to visualize Tree map in Tableau.....	42
Geographical map visualization using Tableau.....	44
Visualization of Combo and Gauge Charts using QlickView.....	45
Research questions.....	52
Question 1.....	52
Question 2.....	54
Question 3.....	56
Question 4.....	56

Chapter 5

Recommendations/Conclusion.....	57
---------------------------------	----

References.....	59
------------------------	-----------

Chapter 1

Overview

Data Visualization is the representation of data in the form of graphs, i.e. graphical representation of the information that is available which is used for communication and sense-making. There exist several stories which can be discovered from the sources which need to be presented to others. There is a saying that a picture is worth a thousand words. Thus, representation in the form of diagrams can be used to draw many results. The result must be readable and recognizable. Data visualization transforms raw data into information.

Data visualization has two major subfields:

- Scientific visualization
- Information visualization

Scientific visualization:

Scientific visualization is the representation of data graphically to gain understanding and insight into the data. This allows the researcher to gain insight into the system that studied information in ways previously impossible. It is also referred to as visual data analysis. The purpose is to convey the scientific data accurately, reveal underlying structures in data and encourage the exploration of the data.

Information Visualization:

Information visualization is the communication of abstract data through interactive visual interfaces.

This produces the visual representation of abstract data to reinforce human cognition, thus enabling the viewer to gain knowledge about the internal structure of the data and casual relationships in the data. Abstract data is the data which has no inherent spatial structure. The purpose is to convey the abstract data accurately, reveal the underlying structure in data, encourage the exploration of data and also to display the data aesthetically.

Types of data visualization include:

- **Multivariate Data Visualization**

Multivariate data visualization is used to gain insight and knowledge from high-dimensional data.

Multivariate data: Data that arises from more than two variables

Multidimensional data: Elements those with two or more dimensions

Examples: Bubble chart, Heatmap etc.

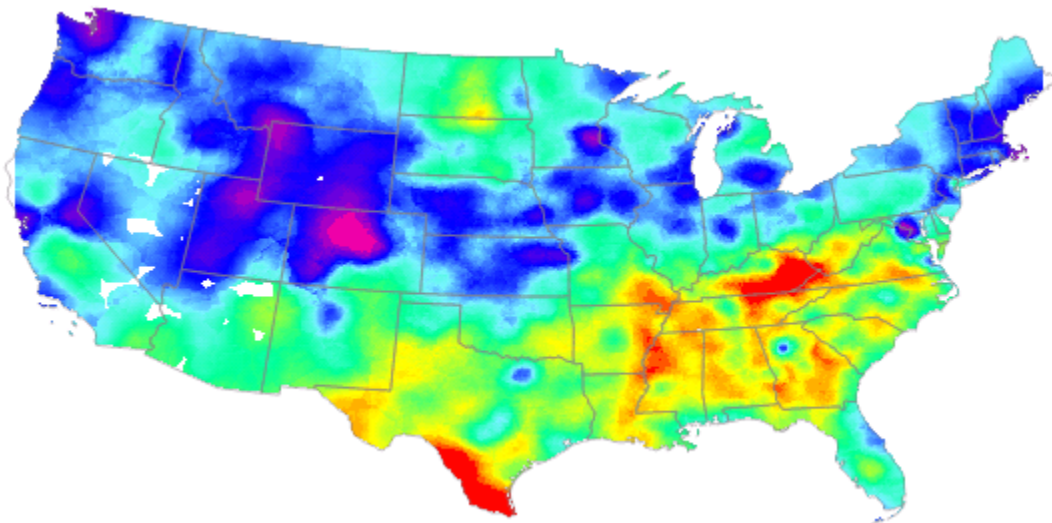


Fig. Heat Map

Examples: Connected Scatter Plot, Time Series, Polar Area Diagram

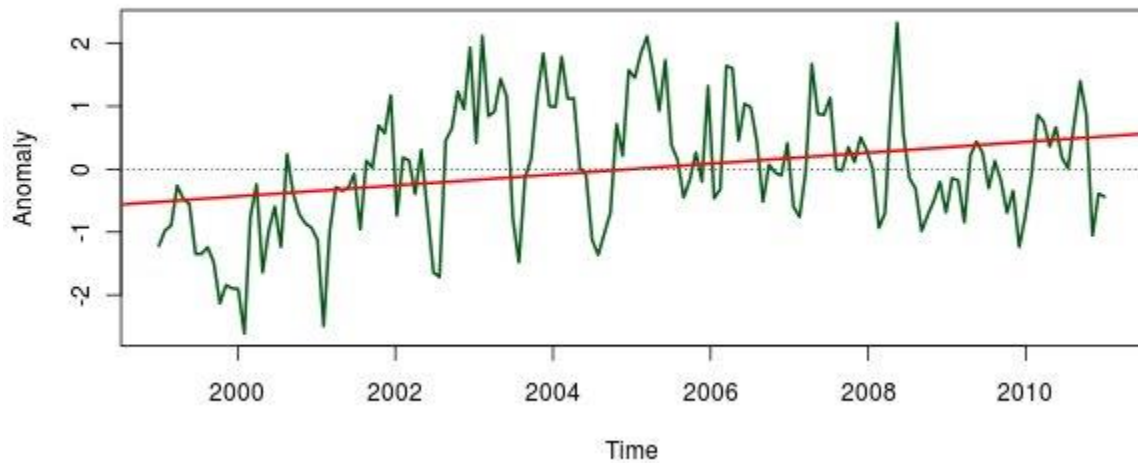


Fig. Time Series

- **Geospatial Data Visualization**

Information about a physical object can be represented by numerical values in a geographic coordinate system. These data relate to the relative position of objects on the Earth's surface. Tools available for this visualization i

nclude Google Fusion Tables, Google Map API, and Tableau.

Examples: Dot Distribution Map, Choropleth, Cartogram

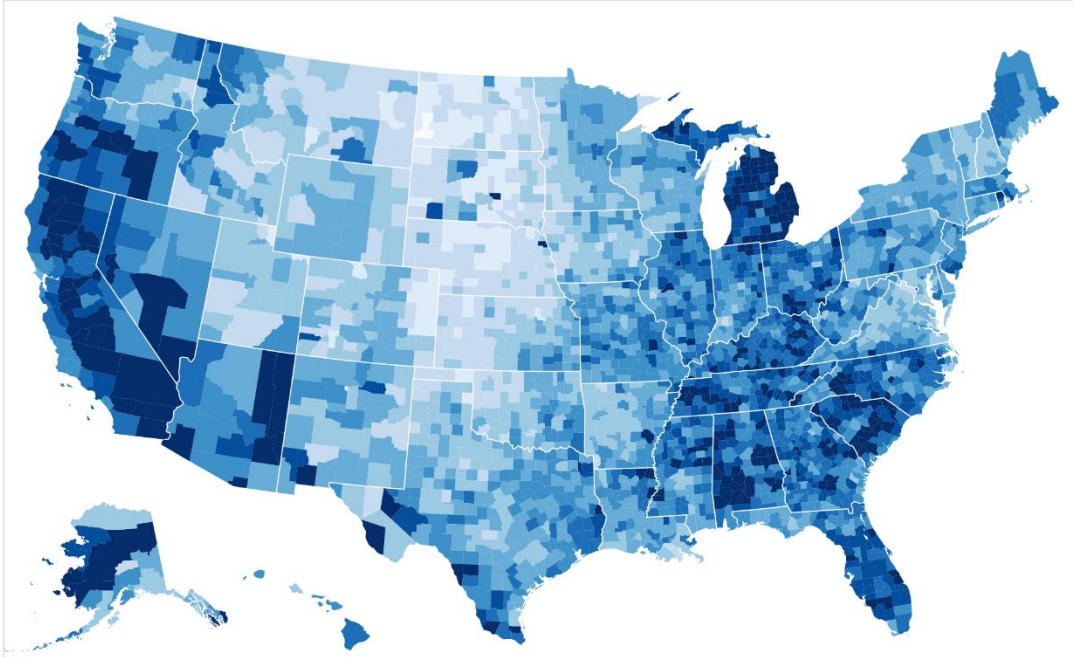


Fig. Choropleth Map

- **Hierarchical Data Visualization**

Groups of objects can be ranked so that everyone but the topmost is subordinate to one above it. Hierarchical data emphasize the subordinate or membership relations between data items. Tools available include Google Charts and d3.

Examples: Dendrogram, tree diagram

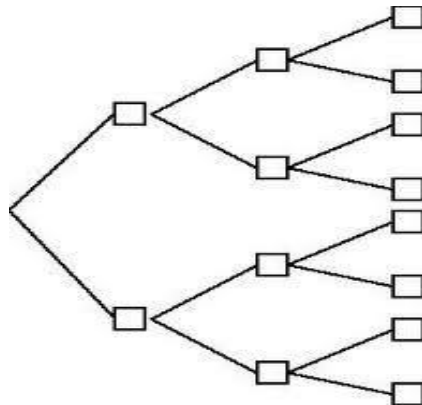


Fig. Tree diagram

- **Network Data Visualization**

Network data visualization is used to show data which exhibit behavior like nodes and links. Network data visualization show how data sets are related to one another within a network. Tools available for this include Alluvial Generator and Google Fusion Tables.

Examples: Alluvial Diagram, Node-Link Diagram, Matrix

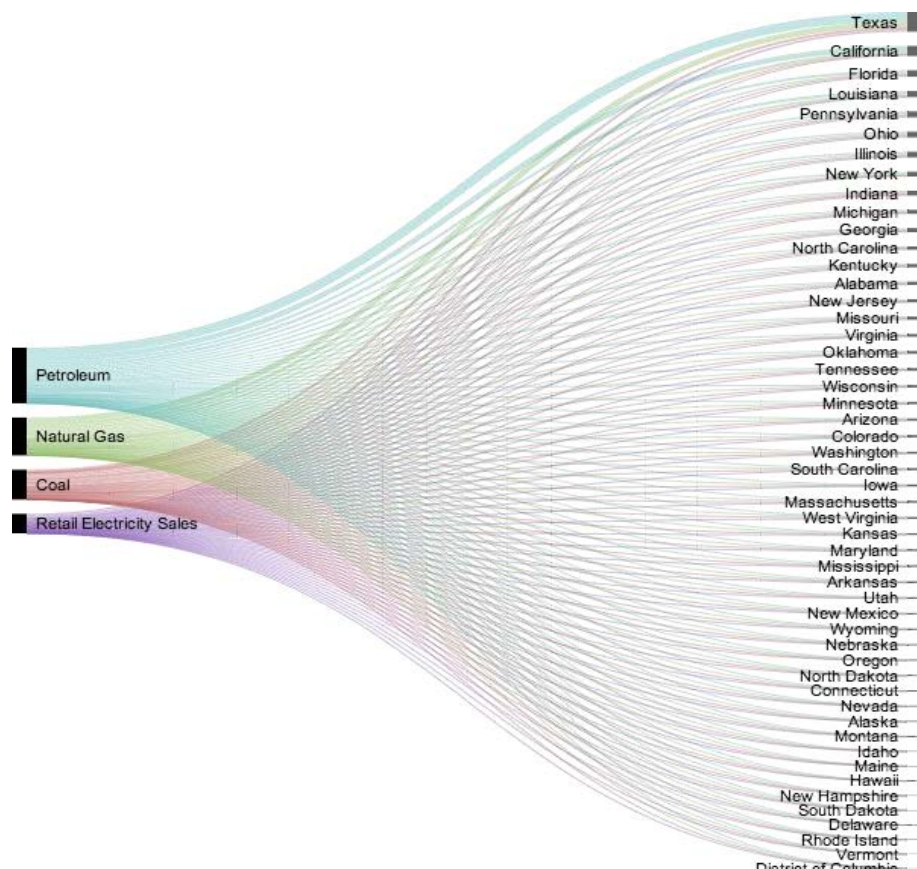


Fig. Alluvial diagram

Purpose of Statement

This paper will discuss the data visualization in real-time data integration, analysis and graphical representation of Tableau. This article explains Custom SQL queries and creating dashboards and map representations like Tree maps in the Tableau tool. Currently, many tools are used for data visualization, and we are considering two popular tools: QlikView and Tableau. This research mainly focuses on data implementation and compares the data visualization tools QlikView and Tableau so that we can analyze and understand the interaction of the data.

Research questions

1. What are the customization capabilities when compare in Excel and SQL?
2. How does the visualization compare and contrast with the compatibility in Tableau and QlikView?
3. How was a dynamic chart developed to represent the live data?
4. How can geographical data represented in color?

Limitations

- Connecting through Custom SQL causes slow performance in Tableau
- Tableau cannot connect across data sources and servers using Custom SQL.
- In order to provide live a data demonstration, we need to have access permission for live data servers.
- These servers provide a layer of data which is visualized using Tableau Maps.

Chapter 2

Background:

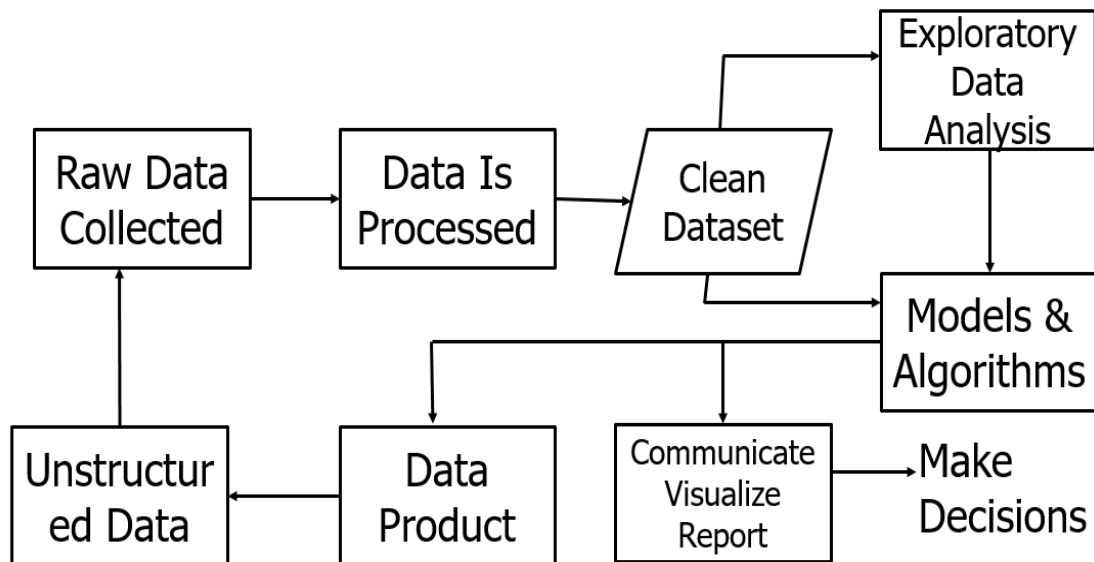
Data visualization is identified as 21st-century research skills, which represent the information in a graphical format so that we can see the patterns and connections of the data. The visual information can make it easy to grasp difficult perceptions and develop the design of the information. The data is mainly displayed visually for two purposes: communication and comprehensibility. The main advantage of the data visualization is not making the data more beautiful, but providing brainstorming methods for displaying complex data through visual communication. It allows communication with a large amount of complex data by using charts and spreadsheets, which make the user focus on the information.

As a picture is worth a thousand words, information can be narrated more effectively in graphs than verbally. Visualization comes from Data Science. Data Science means exploring the available structured and unstructured data and performing quantitative analysis of the data to extract and understand knowledge and get actionable results. Data Science involves several steps: find and acquire the data, clean and transform the data, understand relationships from the data and deliver some value. In the industry, organizations want to know how they are performing and ways to improve their performance as well. It is not always easy to understand the distinctions of the data or identify the solutions for complex information. The large companies look at the dashboards to solve the problem.

Data visualization recognizes what dashboards are, their value, the offers which are provided by the merchants, and information that companies should consider while looking at the solutions.

Data visualization is very important in the business industry because it identifies the area that needs improvement and the factors which influence the customers' requirements, provides a clear understanding of the complex data regarding product marketing, and predicts product sales. Businesses can improve their analytical and operational performances through data visualization which includes tools like, Infogram, QlikView, Exhibit, Qlik Sense, Mode, Gephi, Plotly, Tableau software, and visual mining.

Data Science Cycle



Visualizing data using charts

There are many charts to visualize the data. Some of them are

- Line chart
- Histogram

- Bar chart
- Stacked bar chart
- Scatterplot
- Area chart
- Box whisker plot
- Pie chart
- Donut chart
- Radar chart

Line Chart: Displays information as a series of data points connected by straight line segments. It represents one variable. It is used to show the distribution of the data.

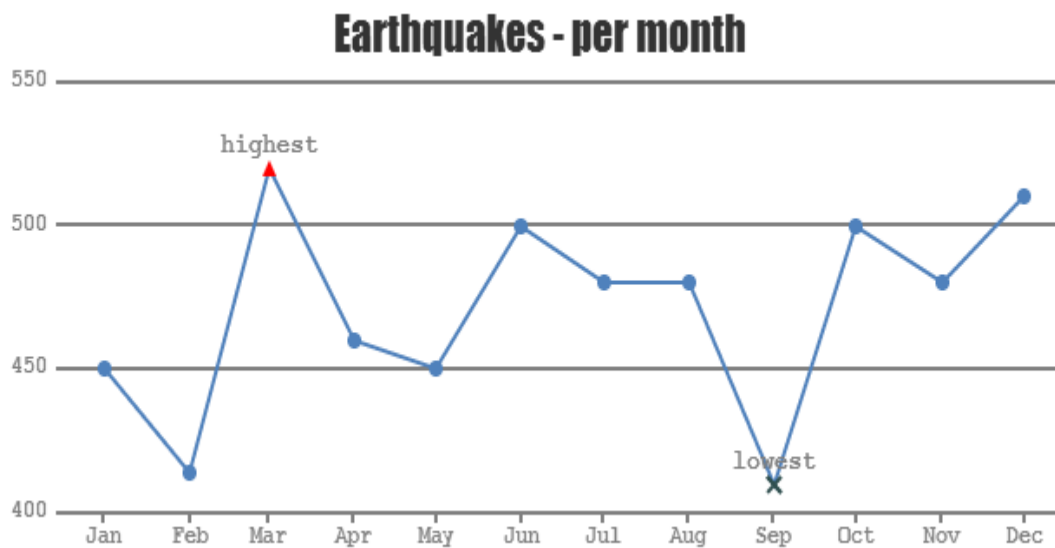


Fig. Line Chart

Histogram: It is useful to illustrate the frequency distribution of data within equal intervals. It consists of rectangles whose area is proportional to the frequency of a variable and width is equal to class interval. It is used to show the distribution of the data.

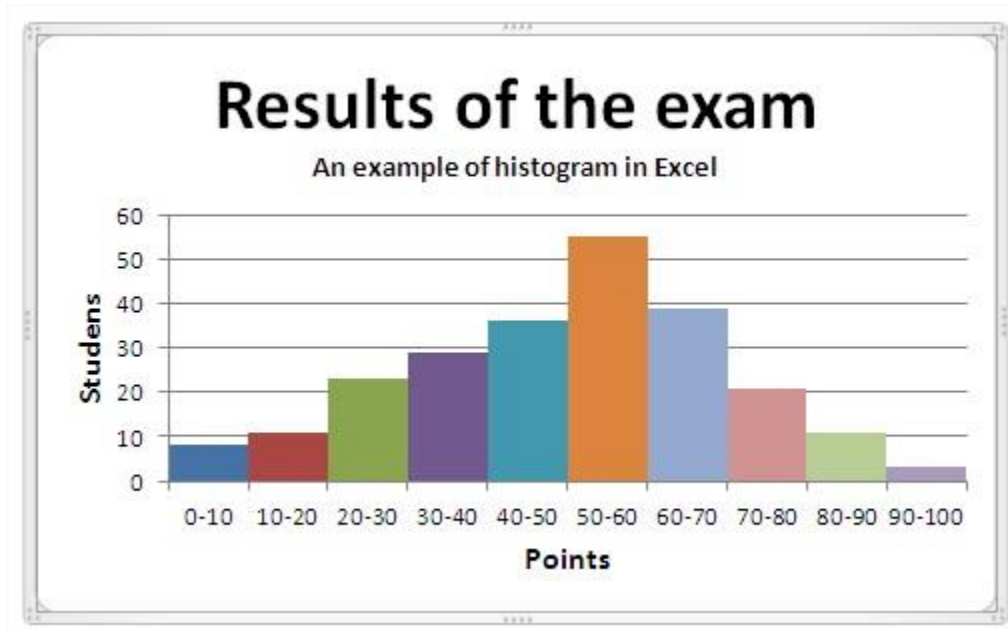


Fig. Histogram

Bar chart: Bar chart is used to display and compare the data. The scale should be divided into equal intervals. It looks like histogram, but differs as this is used for comparison between more than one variables. It is used to show the comparison of data points.

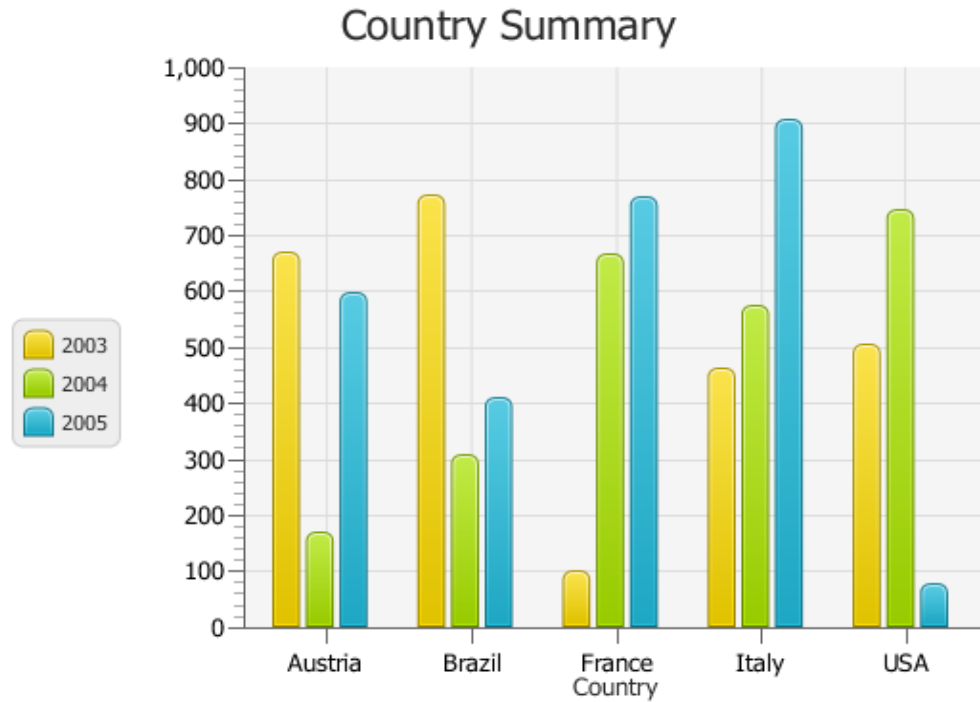


Fig. Bar Chart

Stacked bar chart: Stacked bar chart is used to compare the parts to the whole. The bars are divided into categories. Each category is represented with a different color/pattern. Often each bar represents a total. It is used to show comparison of data points and composition of data.

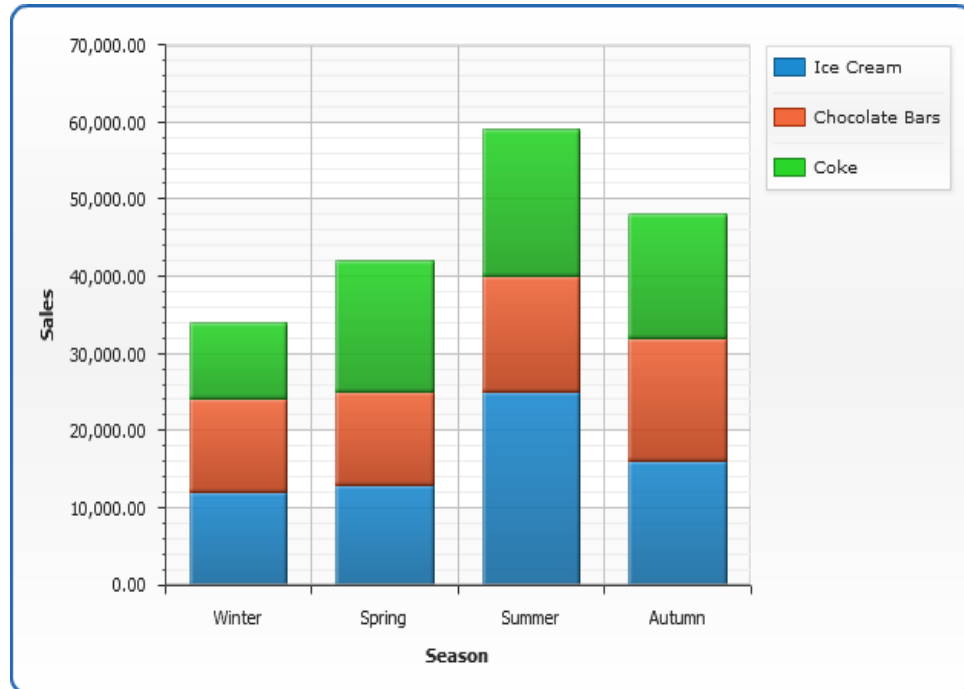


Fig. Stacked Bar Chart

Scatterplot: This displays values for typically one or more variables for a set of data.

The pattern of the resulting points reveals the correlation present. It is often used to know, how much one variable is affected by another. To read a scatter plot, look for the overall pattern tells you something about the direction, form and strength of the relationship. This chart is used to show relationship between data points and distribution of data.

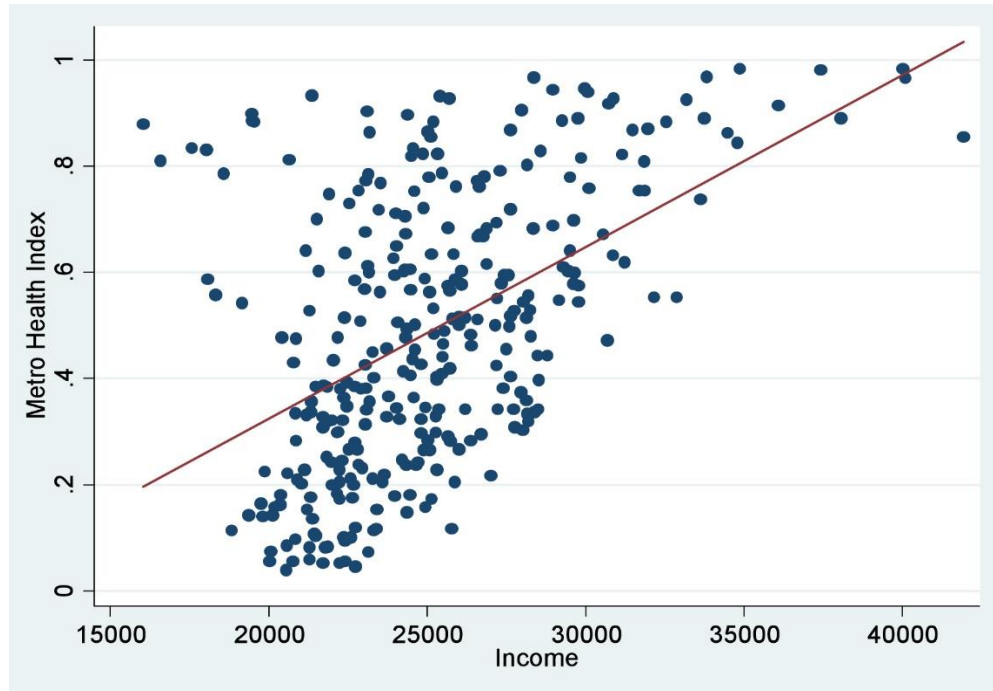


Fig. Scatter plot

Area chart: Area chart displays graphically quantitative data. It is used to represent cumulated totals of more than one variables. Line charts and area charts are very closely related. The main difference is that the area chart is ideal for part to whole relationships. It is used to show the comparison of data points, composition of data and distribution of data.

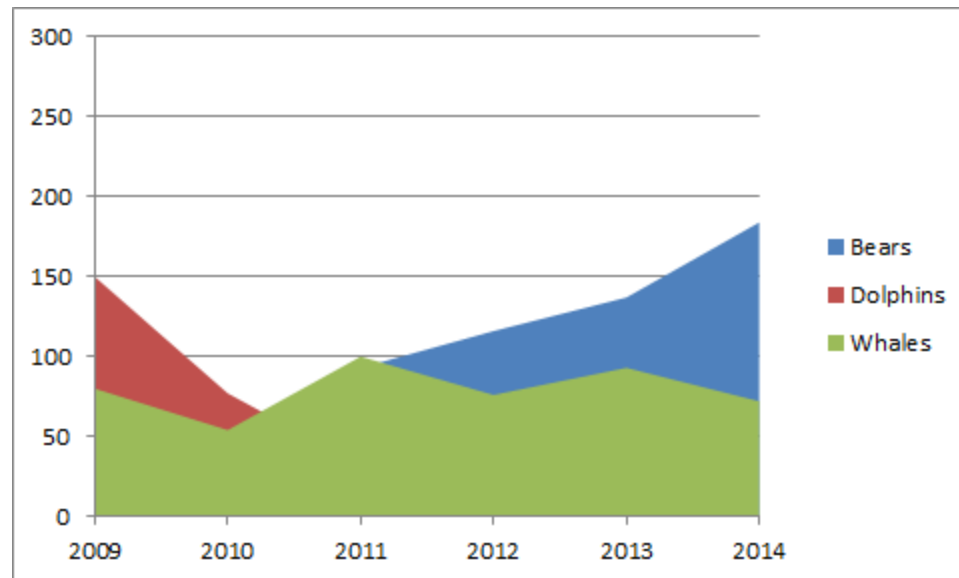


Fig. Area Chart

Box whisker plot: It is the convenient way of graphically depicting groups of numerical data through their quartiles.

- Max value
- Min Value
- First quartile (Q1)
- Median
- Third quartile (Q3)

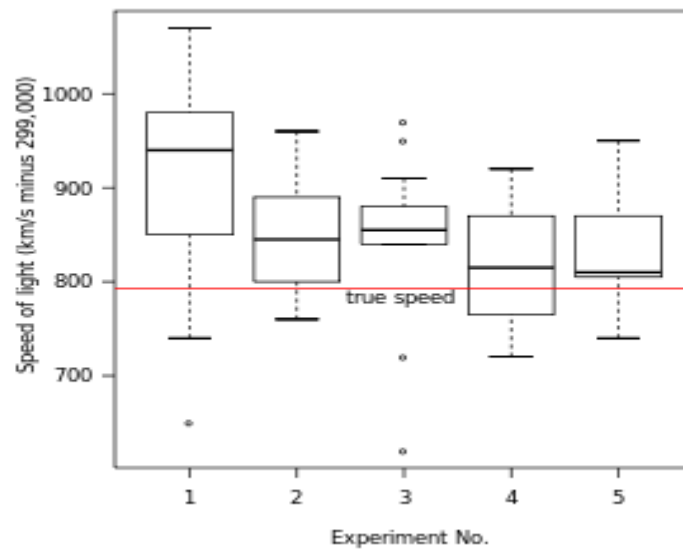
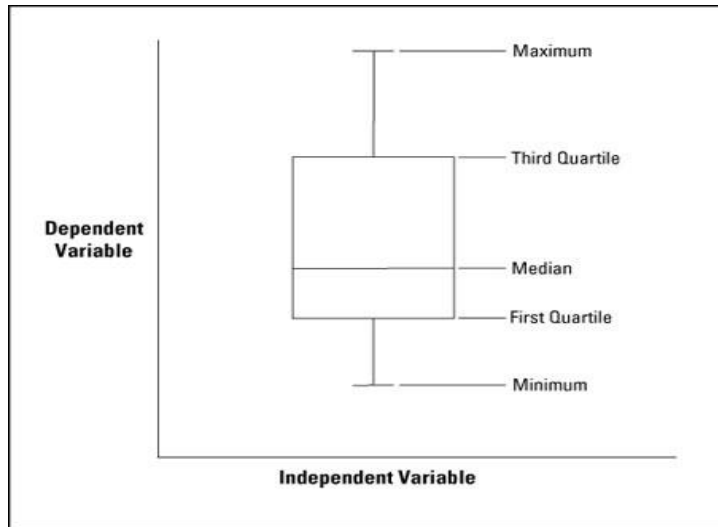


Fig. Box whisker plot

Pie chart: In this circle is divided into sectors that each represent a proportion of the whole. Each arc represents size of each component. Used to show Composition of data.

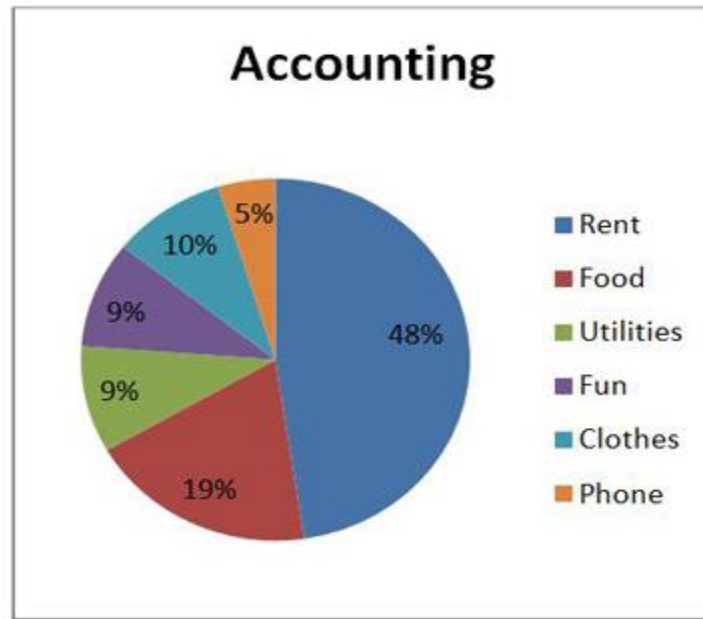


Fig. Pie Chart

Donut chart: It displays value data as percentages of the whole. Length of each slice is proportional to its quantity. It is used to show Composition of data.

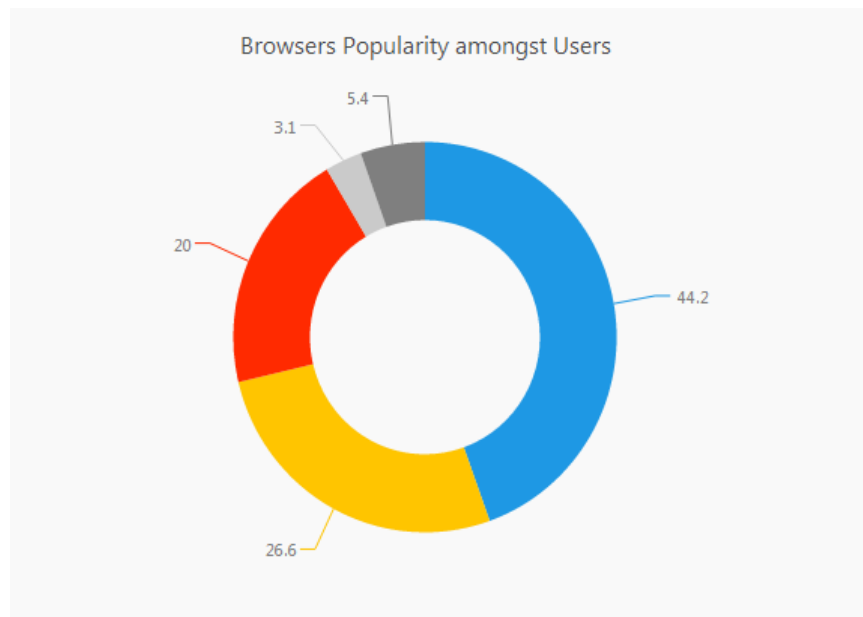


Fig. Donut Chart

Radar chart: It is used to display multivariate data on axes starting from same point. It contains three or more axes. It is used to represent one or more variables and used in comparison of data points and distribution of data.

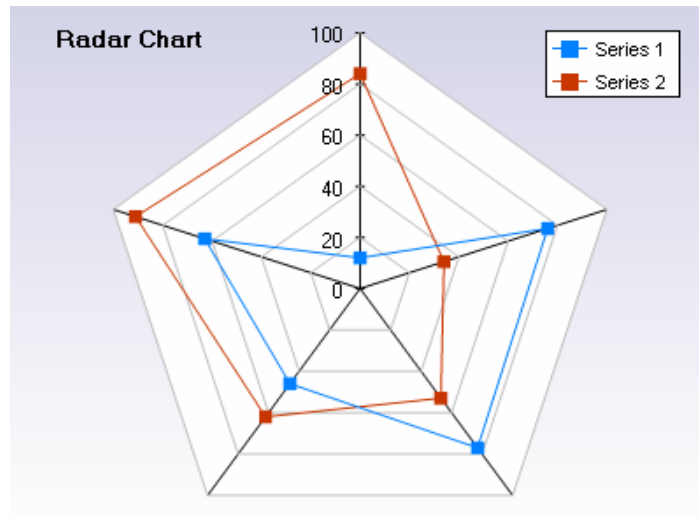


Fig. Radar Chart

Tools used: The major tools used for the visualization are Tableau, QlikView, Qlik Sense, Alteryx etc. Below is the Gartner Quadrant survey as of February 2017. We can see Tableau tops the survey as most used data visualization tool. Tableau, Microsoft and other tools have been widely used from the past few years.



Fig. Gartner Survey for Visualization 2017.

Environment Tools:

We have stored lots of data of the states, sales of the products in Excel sheets. To represent the real time-data, custom SQL queries, and creating tree maps, we used tableau tool. We applied QlikView tool for create Combo chart and Gauge chart.

Necessary Resources:

Ferris State Library, Google search Engine, and YouTube training videos

Chapter 3

Methodology

Summary

- Data Visualization is the representation of data in the form of graphs, i.e., graphical representation of the available information which is used for communication and clarity.
- The presentation of non-visual data visually: The data must come from something that is not immediately visible.
- The production of an image: Visuals must be the primary means of communication.
- The result should be readable and recognizable: There must be some relevant aspects of data that can be read.
- Data visualization transforms raw data into information.

Primary Purpose

- Aims to convey scientific data accurately
- Aims to reveal underlying structure in data
- Aims to encourage exploration of data

Why is data visualization important?

- Visualizations help people see things that were not obvious to them before
- When data volumes are very large, patterns can be spotted quickly and easily
- Visualizations convey information in a universal manner and make it simple to share ideas with others
- Spreadsheets are hard to visualize

- Data visualization makes interpretation easier

Statement of problems:**Problem 1:**

The representation of static data is easy as data is readily available and is not changed as time changes. But representation of live or real-time data is complex and a challenging task to do as data changes constantly as time changes.

Approach:

Web Map Service in Tableau is the approach used to represent live or real-time data. Real-time or live data means that the data is delivered immediately after it is collected without any delay. This is often useful in navigation, tracking, weather etc. Web Map Service is a standard protocol which serves the georeferenced map images using the data from the GIS database. To provide the live data demonstration, we need to have access permission for the live data servers. These servers provide a layer of data which is visualized using Tableau maps.

Problem 2:

In the representation of large data, it is often difficult to visualize the exact type of data from the data source, so the problem is how to customize the data to get accurate output.

Approach:

Custom SQL Queries is used to find a solution to this problem. Tableau queries each data source using SQL that's specific to the data type.

Tableau allows the SQL to query a data source to be customized to manipulate the joins, filters, and field lengths and types to produce a more accurate output.

Restating of Research questions:

1. What are the customization capabilities when compared in Excel and SQL?

Approach:

EXCEL is a program and SQL is a standard language used to create, design and manage data. Microsoft Excel has spreadsheets which contain any type of data in rows and columns. It is easy to learn and share the file copy. Database has Tables where it stores structured data and it is easily scalable. Database allows modifications, inputs, and display of data at a time and with spreadsheets, it supports duplication of the data. Excel sheets allow the user to enter and edit the data manually whereas database uses queries. Database mainly focuses on the data, using a preliminary scrutiny, exploring, and data cleaning. Database is designed in such a way to plug into other things. SQL standards defines functions e.g. Arithmetic. A selective query can give multiple results from different tables.

2. How does the visualization compare and contrast with the compatibility in Tableau and QlikView?

Approach:

In the business world, managing a lot of data and finding the fastest way to access the data are the biggest issues. Both Tableau and QlikView tools combine the data into text, graphics, animated charts, and dashboards where the interaction of the information is easy. Tableau visualization capabilities are many and different. Word clouds, Bubble maps, Tree maps are great tools to represent the comprehensive data. Tree maps provide a feature called add context to the graphics and helps in comparing the multiple categories of the data.

QlickView is the most popular BI tool because it offers a diversity of the features and is highly insightful. It is easily, strategically configurable and produces excellent reports within minutes. As it is a parent of the associative technology, it confirms the user drive intelligence on the demand. Users used both tools, but Tableau was preferred because there was no need to think about defaults in designing and formatting. It has many features like instant connections with more than 30 data types, excellent in mapping, and integration with Relational database.

3. How was a dynamic chart developed to represent the live data?

Approach:

Web Map Service in Tableau is the approach used to represent live or real-time data. Real-time or live data means that the data is delivered immediately after it is collected without any delay. This is often useful in navigation, tracking, weather updates etc. Web Map Service is a standard protocol which serves the georeferenced map images using the data from the GIS database. To provide the live data demonstration, access permission for the live data servers is required. These servers provide a layer of data which is visualized using Tableau maps.

4. How can geographical data represented in color?

Filled maps in Tableau are used to create a geographical representation of the data. Filled maps with pie charts are used, and that might be effective where the total percentage of product profits can be represented by geographic area. By marking color options, geographical representation of the data can be visualized in colors.

Chapter 4

Visualizing data using Tableau

What is Tableau?

Tableau is a visual analysis solution that allows people to explore and analyze data with simple drag and drop operations. It has a user-friendly interface that creates reports that look great right at the beginning. More than 10,000 organizations used Tableau in the office and on-the-go and tens of thousands of people use Tableau to share data in their blogs and websites.

Tableau is a rapid BI Software

Great Visualizations: Allows to connect to the data, visualize and creative interactive, sharable dashboards in a few clicks.

Ease of use: It's easy enough that any excel user can learn it, but powerful enough to satisfy even the most complex analytical problems

Fast: we can create interactive dashboards, quick filters and calculations.

Why Tableau?

As per Gartner survey "Tableau was chosen more often for functionality than any other vendor in the survey, with the highest product functionality scores". "Tableau's products often fill an unmet need in organizations that already have a BI standard "and more frequently deployed than other interactive visualization vendors as a complementary capability to an existing BI platform.

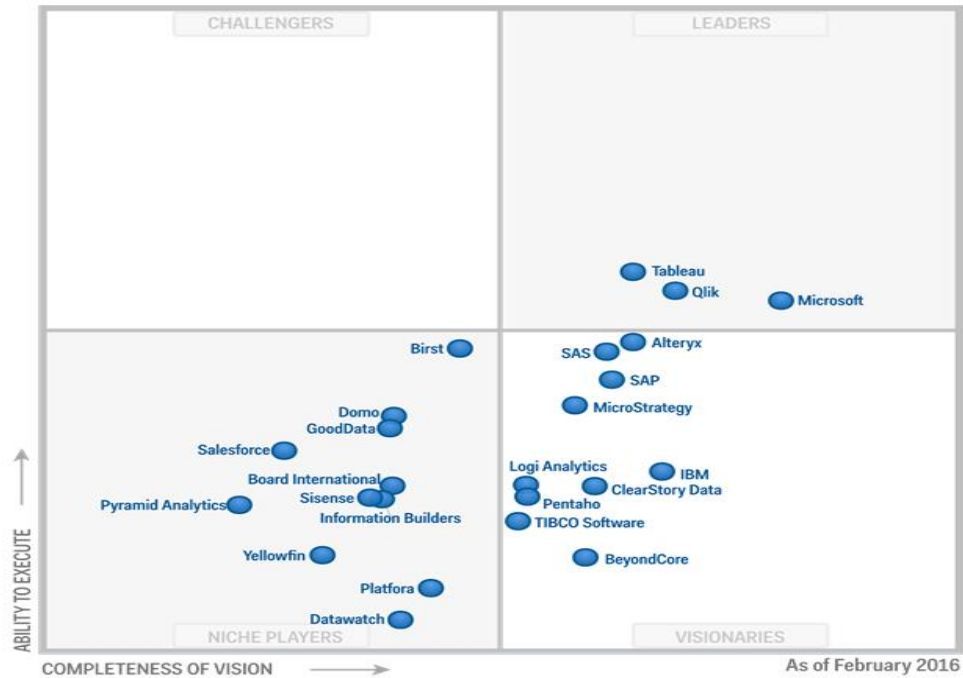
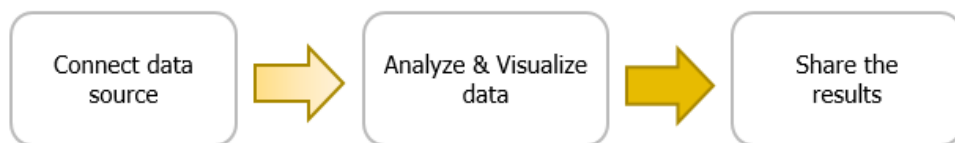


Fig. Gartner Chart 2016

Gartner positions tableau as “Leader”. Companies frequently cite that breadth and ease of use along with high business benefits as the primary reasons that they choose Tableau.

Three main stages in Tableau



Connect data source: Connect Tableau to any data source like MS-Excel, MySQL, and Oracle.

Tableau connects data in two ways Live connect and Extract.

For Extract way of connection, Tableau access data from Excel, text file, tsv and csv. For live data visualization, it almost supports every database like Oracle, MySQL, Microsoft SQL Server, Google Analytics and DB2.

Analyze & Visualize: Analyze the data by filtering, sorting and Visualize the data using the relevant chart provided. Tableau automatically analyzes the data as follows:

Dimensions: Tableau treats any field containing qualitative, categorical information as dimension. All dimensions are indicated by “Blue” color.

Measures: Tableau treats any field containing numeric information as a measure. All measures are indicated by “Green” color

Tableau suggests the recommended charts based on the dimensions and measures.

Share: Tableau Dashboards can be shared as word documents, pdf files, images.

Maps in Tableau

Tableau automatically assigns geographic roles with common geographically names, such as Country, State/Province, City etc. Fields with geographic role will automatically generates longitude and latitude coordinates on a map view. Fields with assigned geographic roles will have a globe icon next to them. In Tableau, there are two types of maps to choose from when creating a visualization with geographic data: Symbol Maps and Filled Maps.

Symbol Maps: These are simple maps that use a type of mark to represent a data point, such as a filled circle.

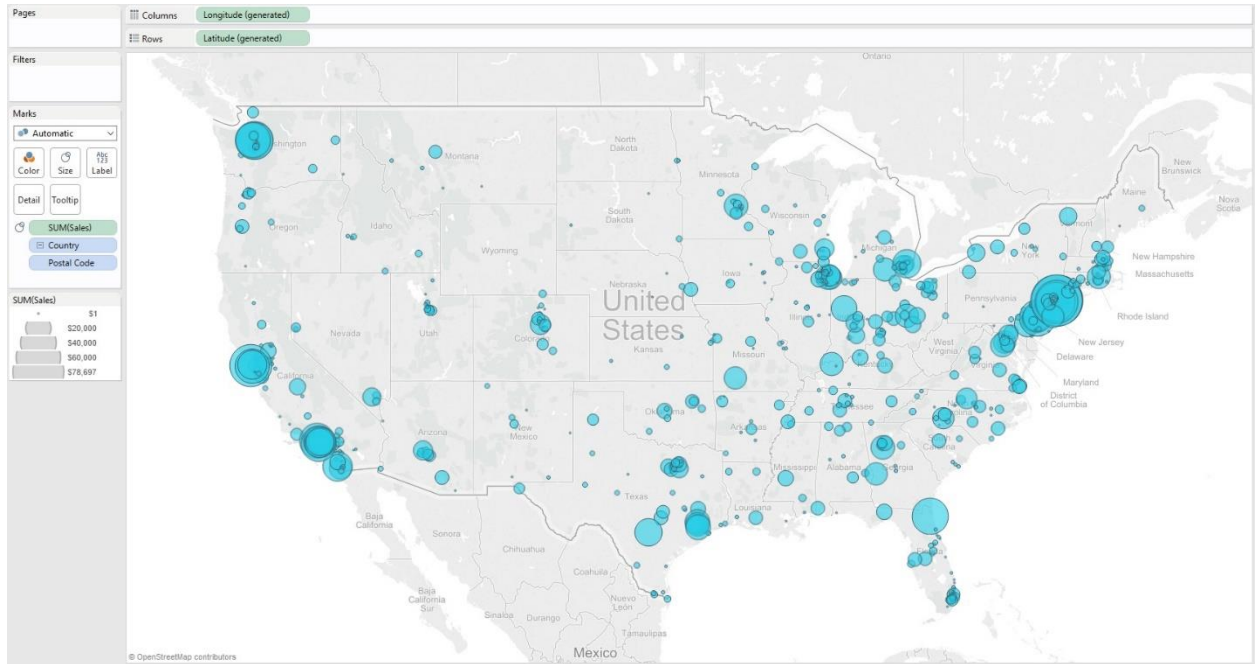


Fig. Symbol Map

Filled Maps: Instead of displaying data points, filled maps uses shading on a country or state basis to indicate relationships.

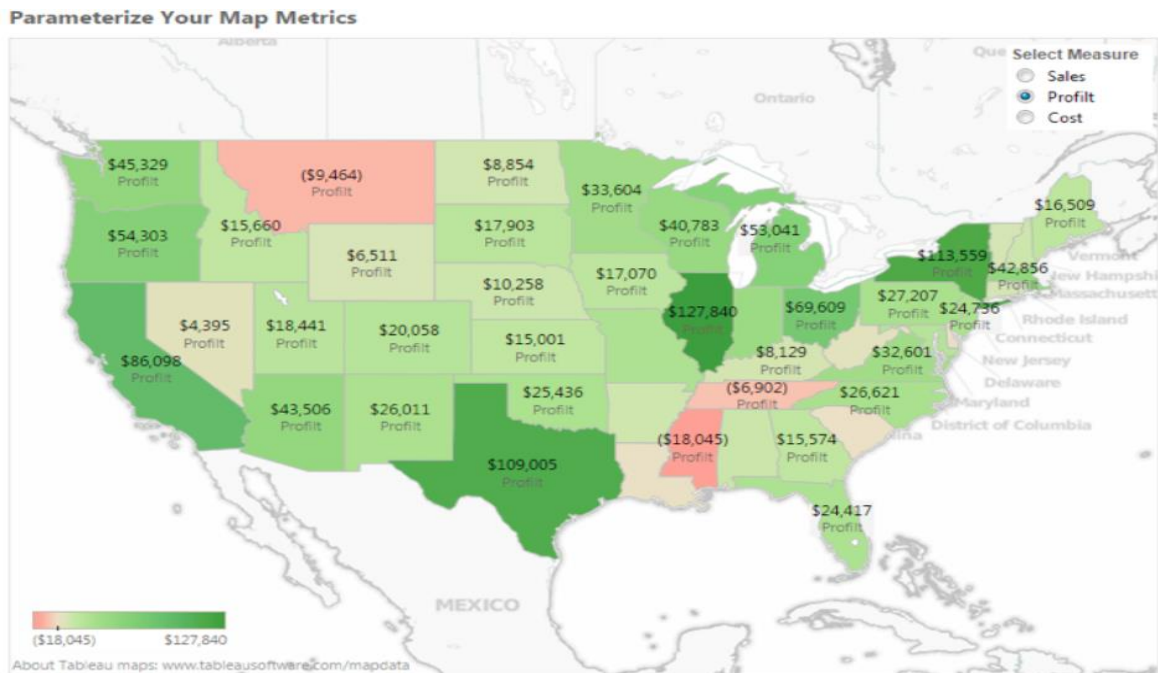


Fig. Filled Map

Custom SQL Queries Tableau

What is Custom SQL Query?

Tableau queries each data source using SQL that's specific to the data type. Tableau allows the SQL used to query a data source to be customized (relatively easily) in order to manipulate the joins, filters, and field lengths and types produce a more accurate output.

Sample SQL query:

```
SELECT ['us states data 1$'].[State] AS [State],  
  
       ['us states data 1$'].[Population] AS [Population],  
  
       ['us states data 1$'].[Region] AS [Region]  
  
FROM ['us states data 1$']  
  
UNION  
  
SELECT ['us states data 2$'].[State] AS [State],  
  
       ['us states data 2$'].[Population] AS [Population],  
  
       ['us states data 2$'].[Region] AS [Region]  
  
FROM ['us states data 2$']  
  
UNION  
  
SELECT ['us states data 3$'].[State] AS [State],  
  
       ['us states data 3$'].[Population] AS [Population],
```

```
        ['us states data 3$'].[Region] AS [Region]

FROM ['us states data 3$']

UNION

SELECT ['us states data 4$'].[State] AS [State],

        ['us states data 4$'].[Population] AS [Population],

        ['us states data 4$'].[Region] AS [Region]

FROM ['us states data 4$']
```

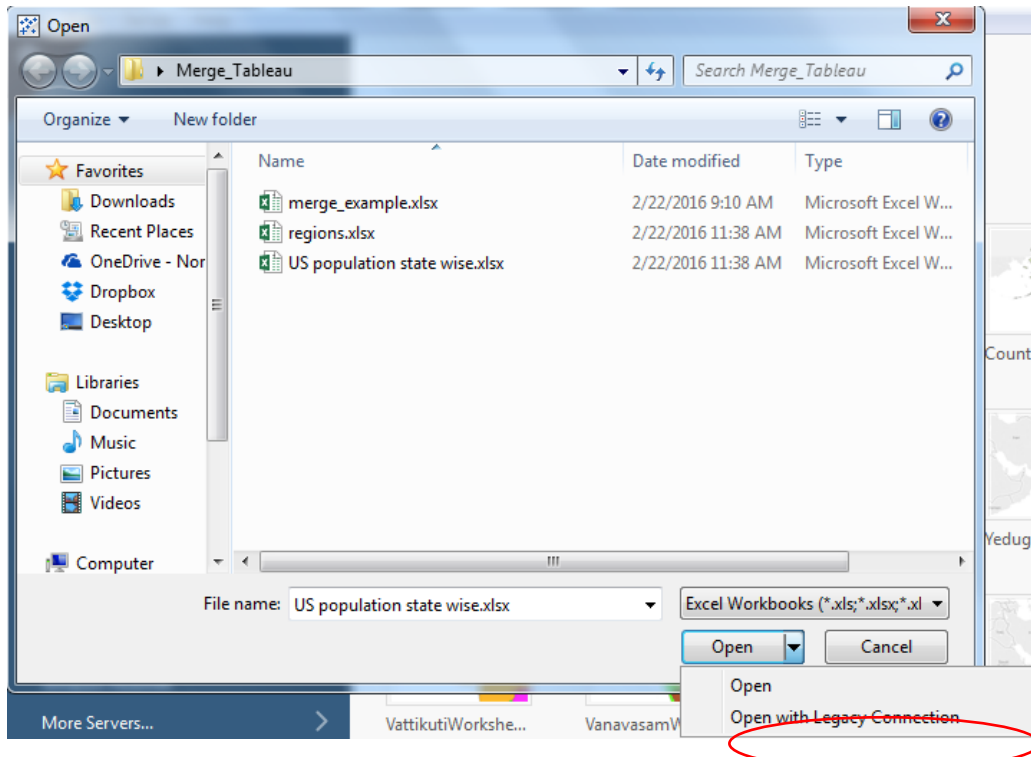
Steps to visualize using SQL in Tableau:

1. The following excel workbook consists of 4 sheets consisting of US population state wise.

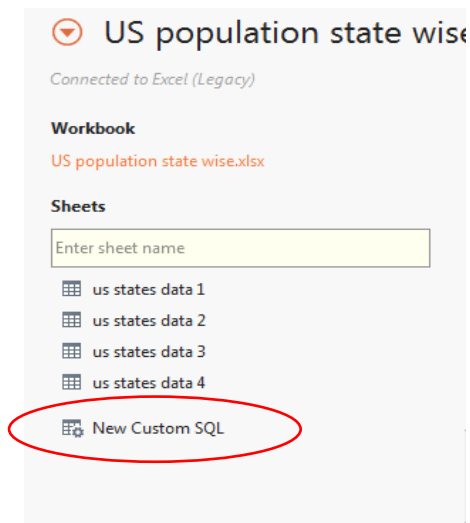


US population state
wise (4).xlsx

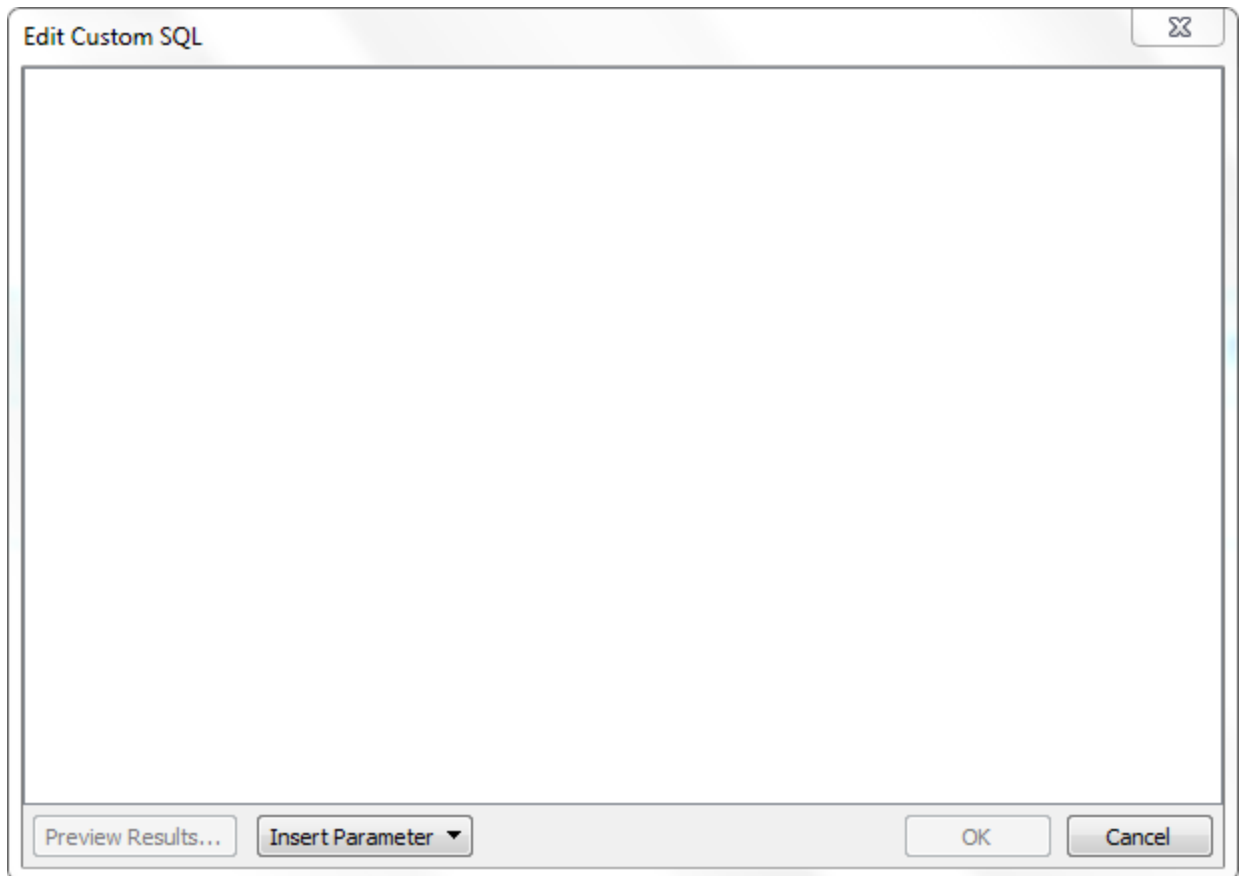
2. Open *Tableau* and connect the Excel data as Open with Legacy Connection.



3. Now your Tableau window looks like this with an additional sheet named as New Custom SQL.



4. Click on New Custom SQL then you will be prompted to Edit Custom SQL as shown.



5. Now we need to write SQL query to perform union operation on the excel sheets and generate a combined sheet.

```
Ex: SELECT [Sheet1$].[ID] AS [ID],  
       [Sheet1$].[Type] AS [Type],  
       [Sheet1$].[Value] AS [Value]  
FROM [Sheet1$]  
  
UNION ALL  
  
SELECT [Sheet2$].[ID] AS [ID],  
       [Sheet2$].[Type] AS [Type],  
       [Sheet2$].[Value] AS [Value]  
FROM [Sheet2$]
```

The above SQL query performs union operation on two sheets namely Sheet1 and Sheet2 with ID, Type, and Value as their columns.

6. After writing the query click on Preview Results on the Edit Custom SQL window pane to check whether the query worked fine or not.
7. The resulting data should contain all the states, population and region name to which it belong to.
8. Click on the Tableau sheet to visualize the data.
9. Now you will see State, Region as Dimensions and Population as Measures.
10. Now visualize the data derived.

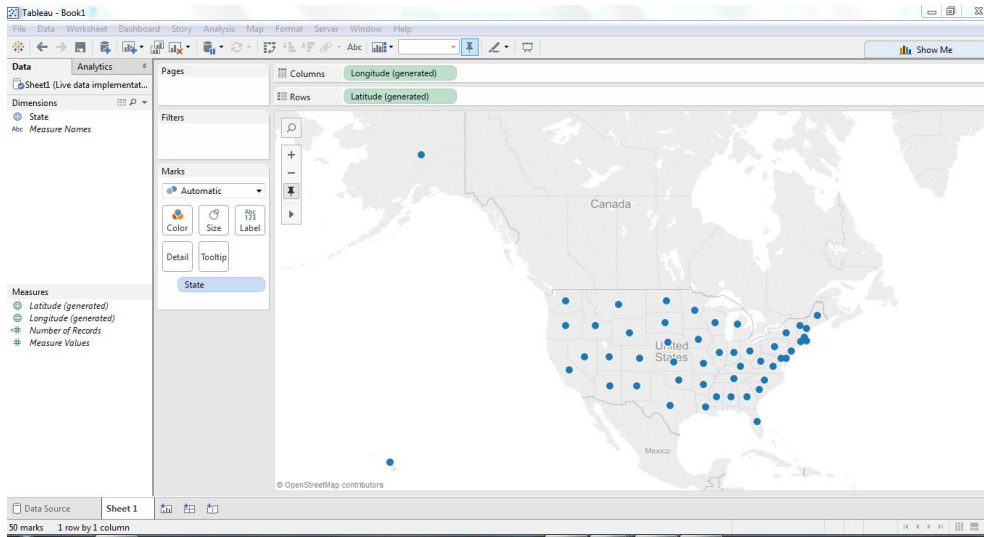
Live data implementation Tableau

Live Data or Real-time data (RTD) denotes information that is delivered immediately after collection. There is no delay in the timeliness of the information provided. Real-time data is often used for navigation or tracking. A Web Map Service (WMS) is a standard protocol for serving (over the Internet) georeferenced map images which a map server generates using data from a GIS database. The Open Geospatial Consortium developed the specification and first published it in 1999. Mapbox is a large provider of custom online maps for websites such as Foursquare, Pinterest, Evernote, the Financial Times, The Weather Channel and Uber Technologies. The Open Geospatial Consortium (OGC) has defined Web Mapping Service (WMS) specification protocols for the transfer of geospatial data from servers to client applications. A Web Map Service (WMS) produces an image (e.g. GIF, JPG) of geospatial data.

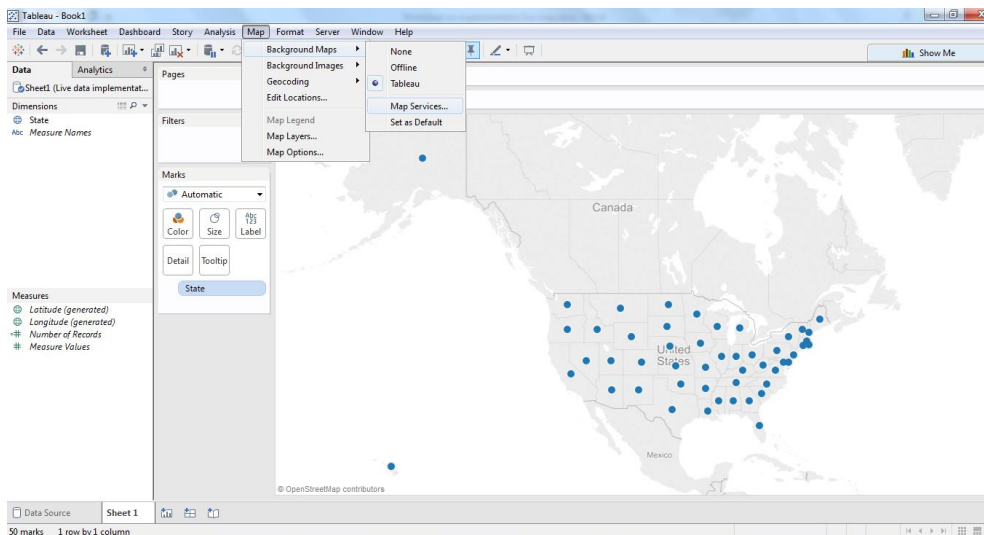
Steps to implement live map using Tableau:

1. Open *Tableau* and connect the Excel data.

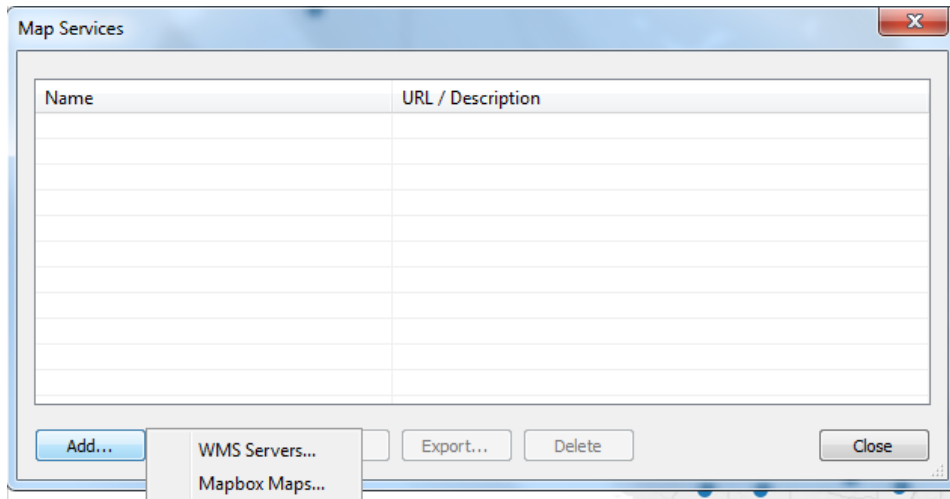
2. Drag and drop state Dimension in Detail menu provided in Marks window pane.
3. Now you can see Symbol map representing the states of United States. It looks as shown below



4. In this we will be using OGCWMS server to get the live geospatial data.
5. Click on Map item in the menu bar, then select Background Maps and select WMS Server as shown

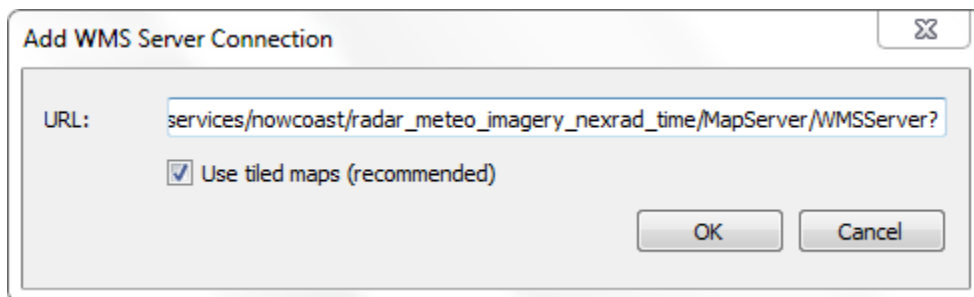


6. After clicking on Map Services you will be prompted a window as shown.

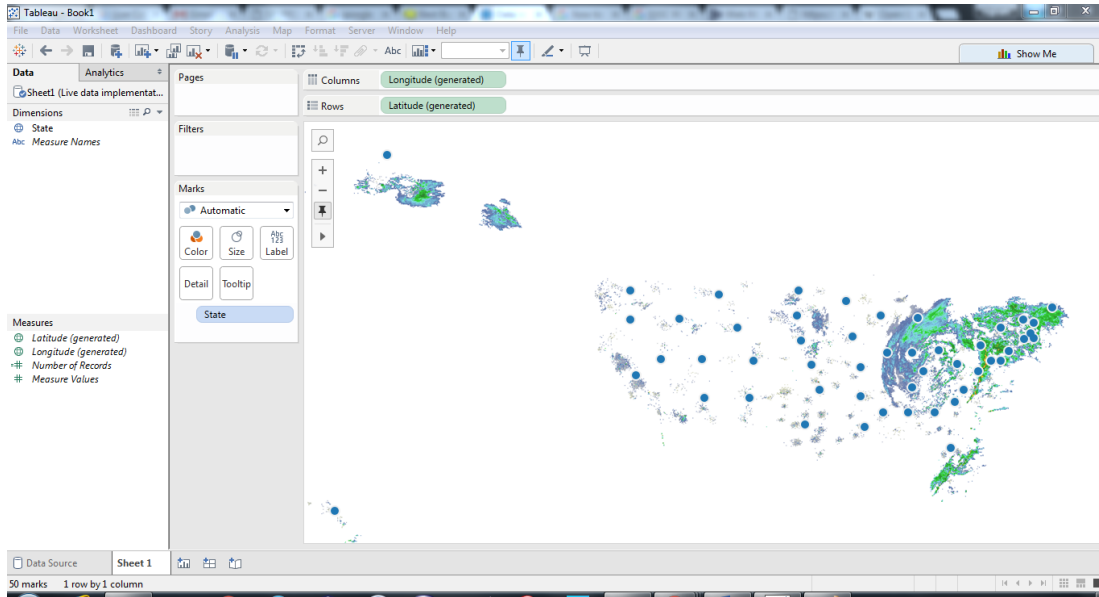


7. Click on Add and select WMS Servers. After clicking on WMS Servers a window pane as shown below will be prompted to enter the URL of the server.

http://nowcoast.noaa.gov/arcgis/services/nowcoast/radar_meteo_imagery_nexrad_time/MapServer/WMSServer?



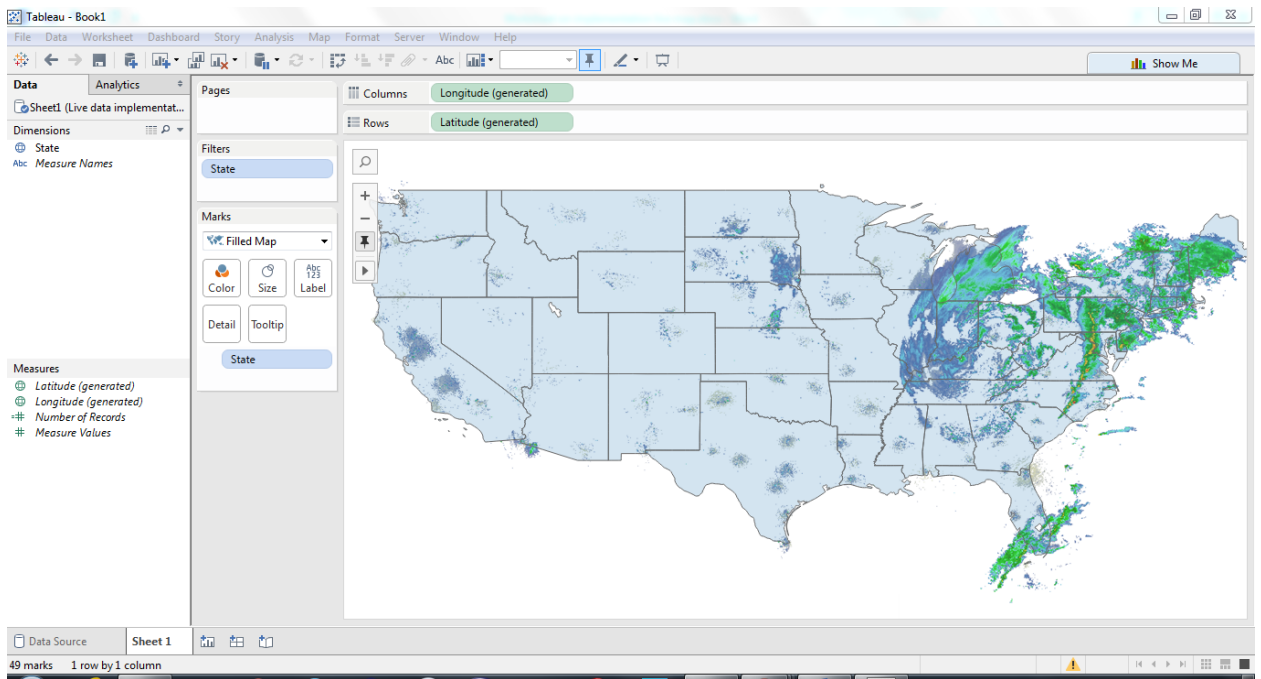
8. Now click on OK and then close the WMS server window pane and a map can be seen as shown



9. Change the map type to Filled map.

10. Change the color transparency to less than 20% so that the required map is visible.

11. Then required map looks like this



Visualizing TreeMaps using Tableau

Tree map displays hierarchical data by using nested rectangles which together represent a whole. They were invented by Ben Shneiderman in 1992 as a way to visualize tree structures in a space-constrained layout. In a treemaps, each branch of a tree is a rectangle, which is tiled with smaller rectangles to represent sub-branches. This allows the viewer to easily see patterns that would be hard to spot in bar charts or area charts. The main benefits of tree maps are, they make efficient use of compact space. Treemaps are becoming popular in infographics and presentations, especially for data analysis that requires detailed views of many items. With this feature one can explore the data hierarchy effortlessly and simultaneously decent level of estimation is also possible for quantitative aspects of information.

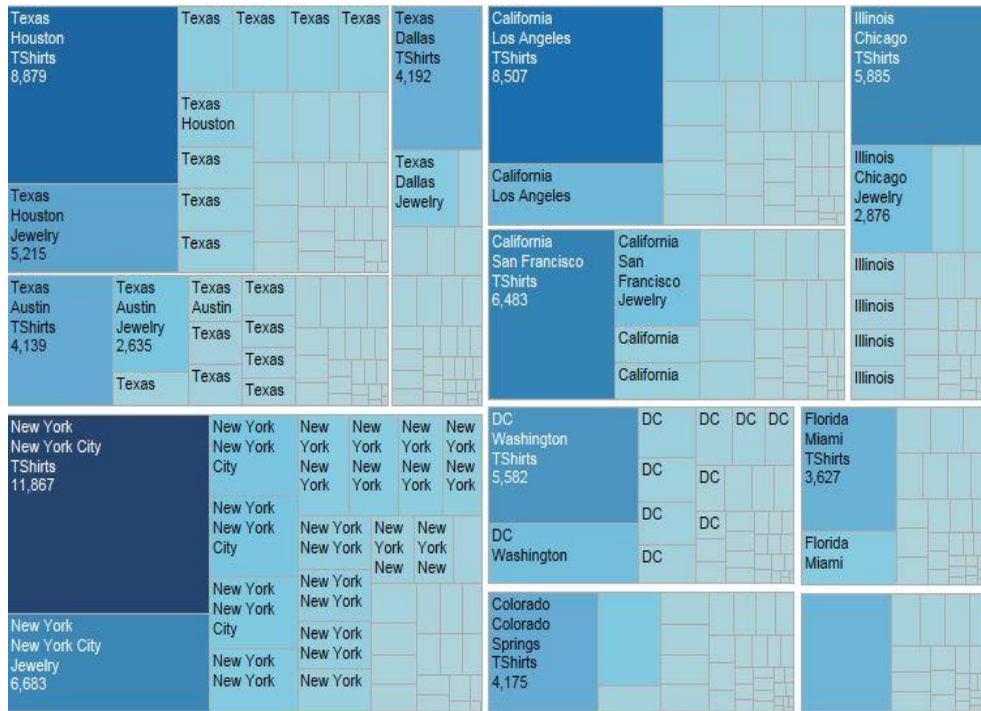
Steps to visualize Treemaps in Tableau:

1. The data is about the various products sold at a fashion store in the cities of different states of the united states
2. Start with opening tableau and connect the data, which is of type excel to Tableau
3. Click on sheet1 which is at the bottom of the page to visualize the data
4. Now to the left of the screen you can see a small window showing Dimensions and Measures. The dimensions of the data are state, city, store name, product, SKU number, year, quarter, month and week. The measures of the data are Quantity sold and sales revenue
5. You can also see rows and columns at the top of the page, now drag sales revenue from the measures and drop it in the rows section. You can see a bar chart with single bar showing sales revenues

6. Now drag states from the dimension to the column section, you can see a bar chart showing the sales revenues of different states
7. Now click on Show Me button on the top right corner of the screen, you can see various maps suggested by the Tableau tool to visualize the data. The maps which are bright in color shows us that those maps are suitable to visualize the given data
8. Select the Tree Map by clicking it, you will see a Tree Map representing various states in green color with different shades
9. The rectangle which is large and bright represents the highest quantity according to sales revenue and similarly the smaller and lighter shade of the rectangles show less sales revenue
10. Now select the city tab from the dimension section and drop it on the Label in the Marks section. You can observe that the rectangles are split into even smaller rectangles according to the cities in each state
11. Now select the product from the dimension section and drop it on the Label in the Marks section. You can observe that the rectangles are split into even smaller rectangles according to the products
12. Now select the Quantity sold tab from the dimension section and drop it on the Label in the Marks section. You can see that the actual quantity sold of each product.



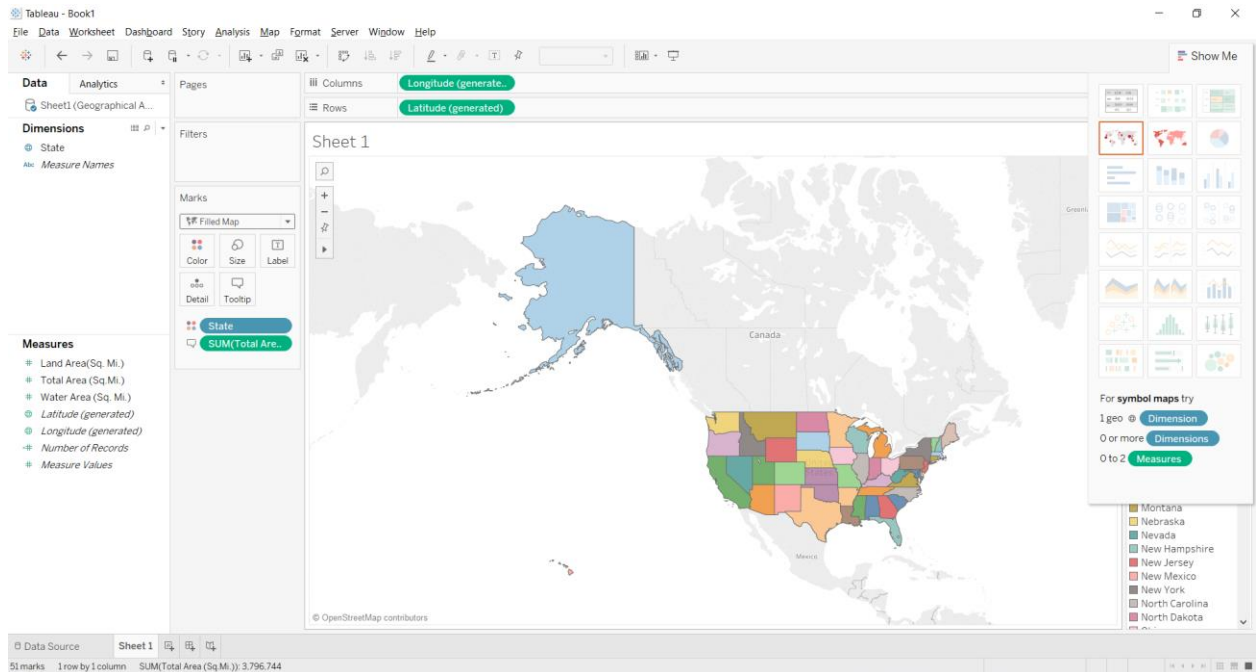
Datatreemap.xlsx



Geographical map visualization using Tableau:

Steps to visualize Filled Maps in Tableau:

1. Open Tableau and connect excel sheet.
2. The data with states and their geographical data is shown in Tableau.
3. Drag and drop states onto columns and Total Area onto rows.
4. Drag and drop state onto color and total area onto tool tip.
5. Now select the filled map. Then the map is represented with the geographical data of each state as shown below.



Visualization of Combo and Gauge Charts using QlikView

1. For this we have taken two different datasets.

- Dataset -1 (Cost_Student). Contains data which representing amount spend by the countries, on a student per year for primary to secondary education and post-secondary education.
- Dataset - 2 (Sales_Rep). Contains data, which represents the sales made by the representatives of a stationary chain around USA over particular dates.

OrderDa	Regio	Rep	Item	Sal	Unit C	Revent	Target
1/6/14	East	Jones	Pencil	95	1.99	189.05	115
1/23/14	Central	Kivell	Binder	50	19.99	999.50	115
2/9/14	Central	Jardine	Pencil	36	4.99	179.64	115
2/26/14	Central	Gill	Pen	27	19.99	539.73	115
3/15/14	West	Sorvino	Pencil	56	2.99	167.44	115
4/1/14	East	Jones	Binder	60	4.99	299.40	115
4/18/14	Central	Andrews	Pencil	75	1.99	149.25	115
5/5/14	Central	Jardine	Pencil	90	4.99	449.10	115
5/22/14	West	Thompson	Pencil	32	1.99	63.68	115
6/8/14	East	Jones	Binder	60	8.99	539.40	115
6/25/14	Central	Morgan	Pencil	90	4.99	449.10	115
7/12/14	East	Howard	Binder	29	1.99	57.71	115
7/29/14	East	Parent	Binder	81	19.99	1,619.19	115
8/15/14	East	Jones	Pencil	35	4.99	174.65	115
9/1/14	Central	Smith	Desk	2	125.00	250.00	15
9/18/14	East	Jones	Pen Set	16	15.99	255.84	115
10/5/14	Central	Morgan	Binder	28	8.99	251.72	115
10/22/14	East	Jones	Pen	64	8.99	575.36	115
11/8/14	East	Parent	Pen	15	19.99	299.85	115
11/25/14	Central	Kivell	Pen Set	96	4.99	479.04	115
12/12/14	Central	Smith	Pencil	67	1.29	86.43	115
12/29/14	East	Parent	Pen Set	74	15.99	1,183.26	115
1/15/15	Central	Gill	Binder	46	8.99	413.54	115
2/1/15	Central	Smith	Binder	87	15.00	1,305.00	115
2/18/15	East	Jones	Binder	4	4.99	19.96	115
3/7/15	West	Sorvino	Binder	7	19.99	139.93	115
3/24/15	Central	Jardine	Pen Set	50	4.99	249.50	115
4/10/15	Central	Andrews	Pencil	66	1.99	131.34	115
4/27/15	East	Howard	Pen	96	4.99	479.04	115
5/14/15	Central	Gill	Pencil	53	1.29	68.37	115
5/31/15	Central	Gill	Binder	80	8.99	719.20	115
6/17/15	Central	Kivell	Desk	5	125.00	625.00	15
7/4/15	East	Jones	Pen Set	62	4.99	309.38	115
7/21/15	Central	Morgan	Pen Set	55	12.49	686.95	115
8/7/15	Central	Kivell	Pen Set	42	23.95	1,005.90	115
8/24/15	West	Sorvino	Desk	3	275.00	825.00	15
9/10/15	Central	Gill	Pencil	7	1.29	9.03	115
9/27/15	West	Sorvino	Pen	76	1.99	151.24	115
10/14/15	West	Thompson	Binder	57	19.99	1,139.43	115
10/31/15	Central	Andrews	Pencil	14	1.29	18.06	115
11/17/15	Central	Jardine	Binder	11	4.99	54.89	115
12/4/15	Central	Jardine	Binder	94	19.99	1,879.06	115
12/21/15	Central	Andrews	Binder	28	4.99	139.72	115

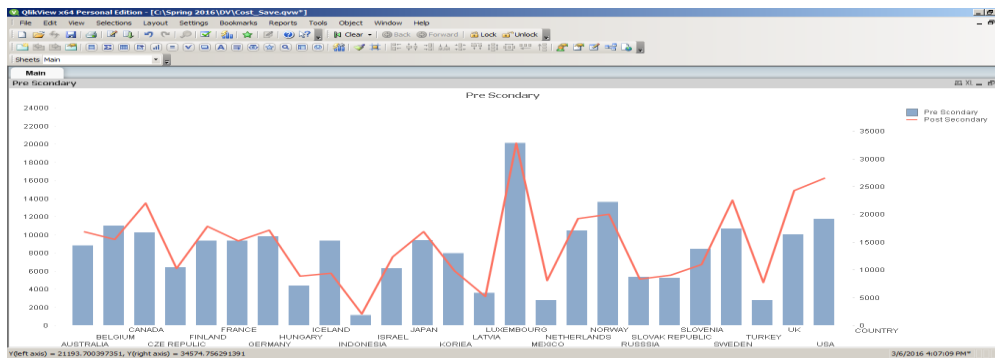
2. Open QlikView. A QlikView getting started screen opens, now click on the New Document button on right below corner of the Wizard.
3. A new Window opens up (Kind of pop up) close the pop up window. (Since it is Auto chart generator, we do not want that wizard).
4. On closing the pop up, user will be able to see a Main blank sheet. Now have quick look on to the toolbar.
5. Over the tool bar below to the selections label, click on the Edit Script icon (Icon looks like pen on a paper).
6. A new Edit Script window opens up (We use this window to load an Excel file), bottom of the window under the Data from files Click Table Files button, and choose the saved Excel file, then click open.
7. A new File Wizard Type window opens up. Now select the Tables drop down and choose Cost_Student if it has not defaulted and click the Labels drop down and choose Embedded Labels. Click next until, next button disables (If it disables you are on final screen). Select check box load all and click finish.
8. On Finish you will be able to see Edit Script screen , with script added to load excel file (Script shown below) On Edit Script window tool bar below file , click the reload icon. It asks you save the file.

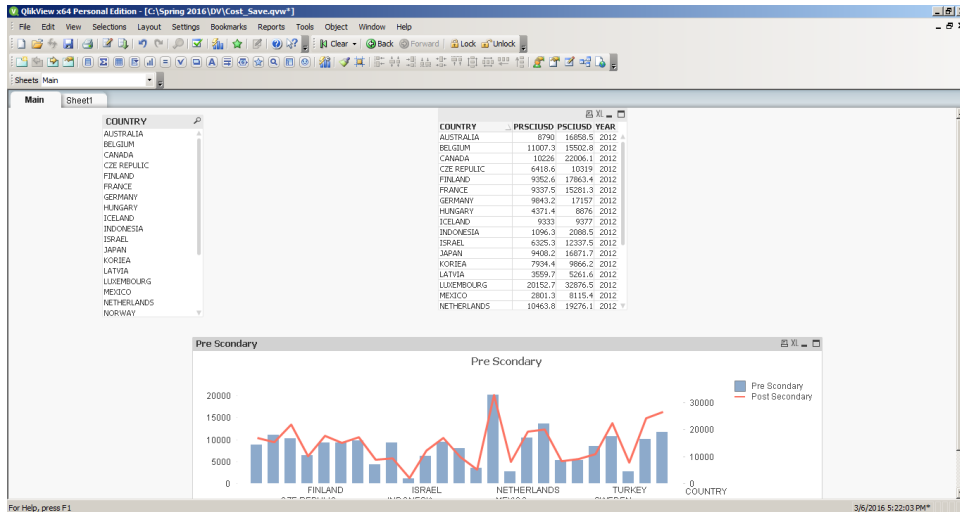
LOAD * FROM

[C:\Manusha\DataViz\DataViz-Qlik\QlikView_Dataset.xls]

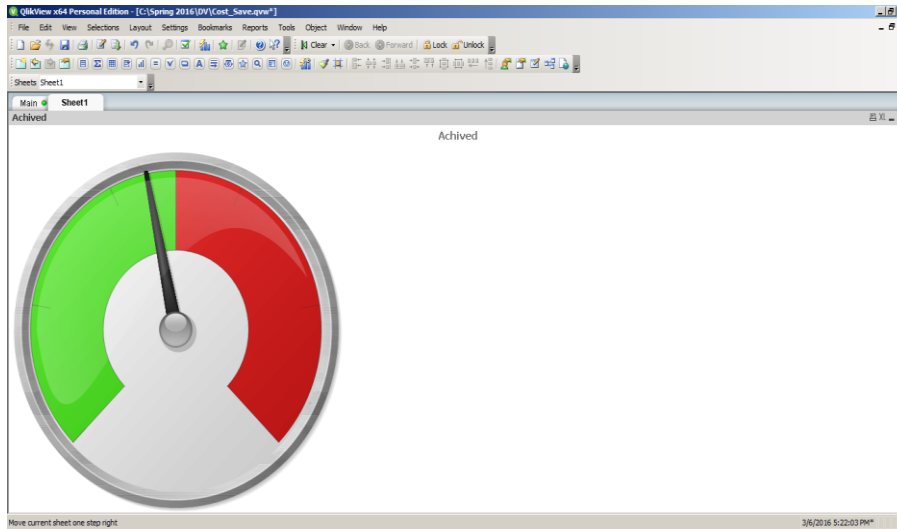
(biff, embedded labels, table is Cost_Per_Student\$);

9. Now a sheet property window opens, please choose Country (Which works as a filter).
10. Right Click anywhere on the sheet, a window opens choose New Sheet Object and select Table box. New window open again, add all the available fields and click apply. You will be able to see a table box with content same as to excel sheet.
11. Right Click anywhere on the sheet, a window opens choose New Sheet Object and select Chart. New window open again, Select Combo Chart from chart type and click next.
12. Choose Country as dimension and click next now choose the expressions.
 - Expression 1 for bar chart Sum (PRSCIUSD), now click Add button to add one more expression Sum (PSCIUSD) and click finish.
13. Now we have seen the Combo Chart.
14. If you observe both pre-Secondary and post-secondary are on same axis. Let us split them. Right click on the chart and choose the properties, and navigate to the Axes tab, Select Post-Secondary Expression and choose the position Right (Top) click apply and ok. You will see charts similar to below.

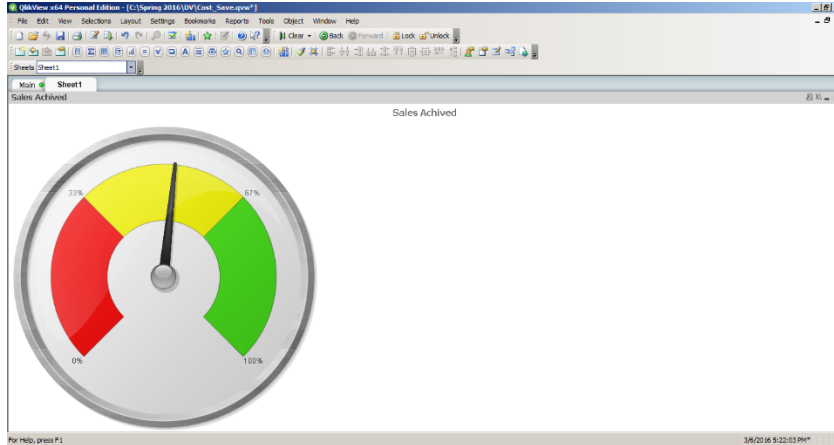




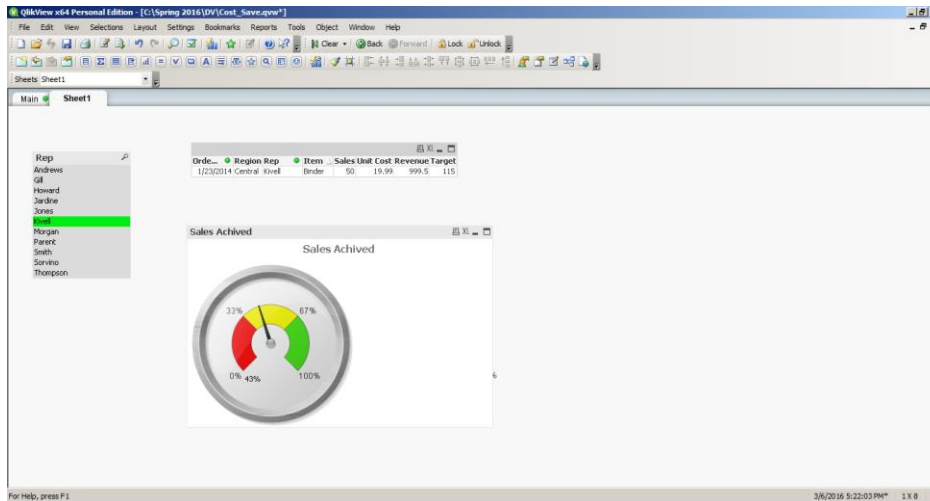
15. Now click on layout shown in the tool bar and add new sheet. A new sheet will open.
16. Repeat step 7 to step 12, in step 9 choose sales_Rep sheet and in step 11 Choose Rep to the filed display in list boxes.
17. Right Click anywhere on the sheet, a window opens choose New Sheet Object and select Chart. New window open again, Select Gauge from chart type and click next.
18. Gauge charts have no dimensions, so we do not select any dimension, will be moving directly to expression.
19. Since we are calculating the performance, we add a division between sales and targets (Sum (sales)/sum (targets)), add label for expression (Sales Achieved) finish you will see a chart like below.



20. Let us make it more readable, right click, choose the properties, navigate to presentation add a segment in segment setup (we can go with 2 segments as well for clear user readability we choose 3) and change the colors of each segment by choosing color band (Red, Yellow, Green).
21. Now look for Show scale section left –below to segment setup add margin units a 4 and show labels on every major entry margin units 1 (Select the check box Show scale if it is not selected).
22. Now Switch to Number tab since we are representing it in percentage select the radio button Fixed to and select the check box Show in percentage and click apply. Will see something similar to below.



23. In order to see the needle value (Speedometer current value). Right click on the chart and select properties move to presentation tab and select Add button next to Text in Chart (located below right corner of the window) and add this formula($=\text{num}(\text{Sum}(\text{Sales})/\text{sum}(\text{Target}),\text{'###\%'})$), and click apply, value appears on the top left corner of the chart. Press CTL +SHIFT to drag the text anywhere in the chart. Final Gauge chart with data along with chart looks like below.



Research questions

1. What are the customization capabilities when compared between Excel and SQL?

Excel is a program and SQL is a standard language used to create, design and manage data. Microsoft Excel has spreadsheets which contain any type of data in rows and columns. It is easy to learn and share the file copy. Databases have tables where they store structured data and are easily scalable. Databases allow users to modify, give inputs for, and display multiple data at once and if users follow the same process in spreadsheets, databases support duplication of the data. Excel sheets allow users to enter and edit data manually whereas databases use queries. Databases mainly focus on the preliminary scrutiny of data and the exploration and cleaning of the data. Databases are designed in such a way as to allow users to retrieve data from other sources. SQL standards define functions, e.g. Arithmetic. A selective query can give multiple results from different tables.

Advantages of Spreadsheets:

Microsoft Excel has a flexible structure and is easy to use. It is associated with other popular office software and analytic services like SQL server. Excel makes it easy for the user to create, store and modify the data, perform mathematical calculations, and share the reports with others. It has other features like charts, graphs, pivot tables, pivot charts and 3D maps, and provides various templates which achieve multiple tasks at once.

Disadvantages of Spreadsheets:

Unfortunately, Excel has some limits. Every cell is a small fragment where data is unreliable. It creates data integrity issues and more complexity when users want to change or manage multiple spreadsheets. As the amount of data increases, performance of the data will become more difficult and take more time to give results. Spreadsheets are vulnerable to fraudulent activities. They are difficult to test or troubleshoot and do not support collaboration.

Advantages of the SQL:

SQL stores the data logically and systematically by uses normalization through various tables. It not only supports scalability, but also formulates queries help to retrieve the bulk amount of data from the database efficiently and quickly. Databases have well-defined standards which are adopted by ANSI & ISO.

Access database offers many tools and these are responsible for maintenance of the data quality Database integrated with emerging technologies like cloud technology, Google BigQuery etc.

Disadvantages of SQL:

It is complicated than the spreadsheets. SQL has stability issues when it handles some functions like transactions and auditing. As SQL database defines by the ANSI & ISO standards, it heavily depends on the extensions that assures to the merchant lock-in. Another disadvantage is interfacing an SQL database.

2. How does the visualization compare and contrast with the compatibility in Tableau and QlikView?

The team doing this study worked on both Tableau and QlikView tools and feel both are fabulous data visualization tools which provides different styles of data reporting. Both tools have their own pros and cons.

QlikView: The users of QlikView believe that it is easy to learn and relatively free compared to the Tableau. QlikView (QV) can be defined as an advanced BI reporting tool and allows for continuous advancements. It is a powerful tool for data discovery and navigations between the data by using key performance indicators. From the data analyst view, QV can produce better reports; consequently, it takes more time to develop the reports than Tableau. If users working in an organized environment with related strategies are in a good supported team, QV would be the best tool to produce detailed reports. QV users are comfortable in associating with Excel. It stores the data efficiently, and it has the capability to integrate with the data warehousing tool. Importing data with QV is the one of the most difficult tasks in the QlikView. QV is Enterprise ready with multi-server configuration. Data integration and Data warehousing processes like Extract, Transform, and load (ETL) are very strong in QV. Story implementation is usual in QV.

Tableau: This tool can link to multiple data sources like Excel, text files, and database. It is also easy to learn, and it can store and explore billion bytes of data.

There is no limitations of speed, performance, and size of the data in Tableau. It is intuitive and transparent. Advance versions such as Tableau 8.1 support the Big Data. The beauty of the Tableau is creating immediate presentation of the data. As it adapts the on-screen visualization, it creates instant and real-time dashboards. Tableau allows the user to quickly move through data, and it is very easy to move data in and out of the sheet. Tableau enables map servers to share the live information.

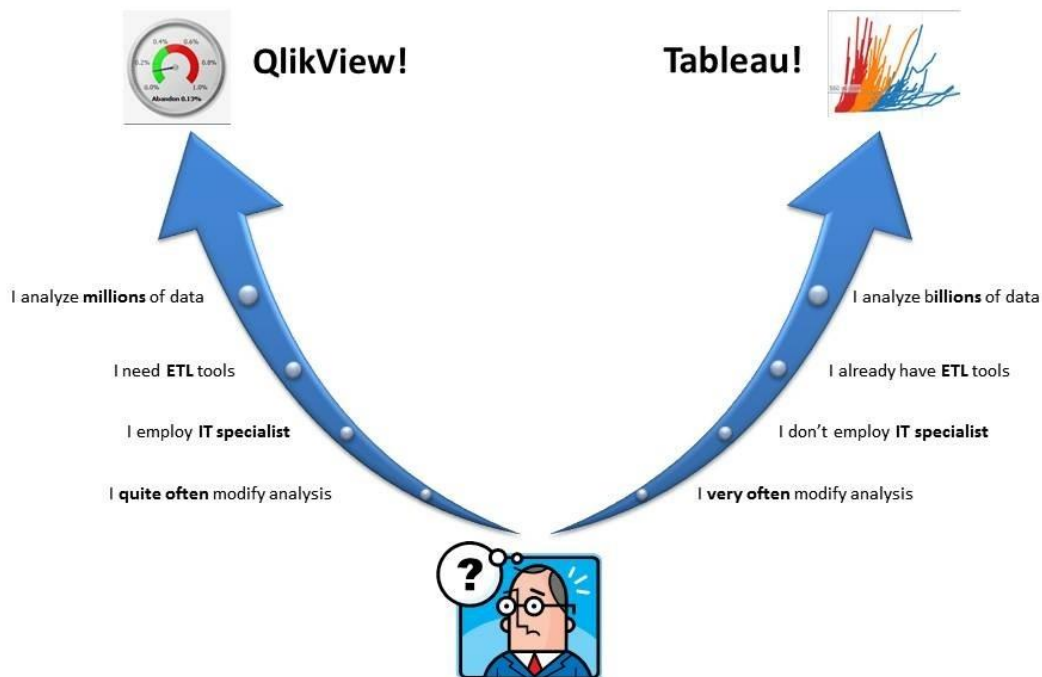


Tableau provides in-built table calculations to add with complex analysis. It has a clean user interface and a drag-and-drop functionality. It can also be integrated with programming languages like Python and R. Tableau is for everybody that push you gently

in the correct path by leaving out worst conditions. Unlike QV, it doesn't have ETL capability, for these users depends on another tool.

3. How was a dynamic chart developed to represent the live data?

Web Map Service in Tableau is the approach used to represent the live or real-time data. Real-time or live data means that the data is delivered immediately after it is collected without any delay. This is often useful in navigation, tracking, weather updates etc. Web Map Service is a standard protocol which serves the georeferenced map images using the data from the GIS database. To provide the live data demonstration, we need to have access permission for the live data servers. These servers provide a layer of data which is visualized using Tableau maps.

4. How can geographical data represented in color?

Geographical data can be represented in Tableau using Filled Maps. After considering the data, take the column of data which you want to represent in the color and place that column on the color mark which is available in the Tableau. Then you can visualize the geographical data in color using Tableau.

Chapter 5

Recommendations/Conclusion

In the present world, a lot of data is being generated and analyzed; this data has been difficult. Thus, Data visualization came into existence, which means creating pictorial or graphical representation of the data so that users can predict and make decisions and also to identify new patterns.

There are several data visualization tools that are available in the market and using these tools, users can represent the data in the form of infographics, geographical maps, heat maps, pie charts, detailed bar charts, etc. Tableau is the most popular data visualization tool which is used even at the corporate level. Data presentation is the means of summarizing, organizing and communicating information with diagrams, charts, graphs etc.

The visualization of data is important because huge amounts of data can be easily visualized. Well-designed graphics from the data are more powerful and also more simple. Data visualization can identify the areas that need improvement or attention, clarify the factors that influence the behavior of the customer, predict the volume of the sales and help one to understand which products to be placed where. Data visualization can represent the quantities accurately, make it easy to compare the quantities, make it obvious how people use the information, and clearly indicate the relationship nature.

The benefit of data visualization is that organizations are likely to find the information which they need when they need it. Data visualization helps users to see the connections between business performance and operating conditions. Finding these connections and correlations is important so that it is easy to identify the cause for any issues and resolve them.

With data visualizations, organizations can easily spot customer behaviors and also the market conditions over various data sets and take advantage of the changing market quickly. The greatest strength comes from the way in which it brings the insights to the surface. Visualization tools help users to interact with the datasets directly and manipulate them. This interaction helps to ask questions about the data sets. It has ability to substitute new business language.

Some of the data visualization tools like Tableau, QlikView, Qlik Sense, power BI help users to get the visualizations quickly and in an easy manner. Each of the tools has unique features and some features in common. These tools help users to visualize any data sets, which absorb information in more constructive ways. Users can also identify the emerging trends faster and in an easy manner. Though data visualizations are easy to understand and also look attractive, it is important to achieve a perfect balance between the functionality and the visual appeal so that we can improve the efficiency.

The future of data visualization is changing gradually. It has been integrating with the geo-spatial and network displays and also some other forms like seamless interaction and simultaneous use. A few areas, like technological support, have the advantage of multiple brains together for collaborative data sense making and for the integration of data visualization with the data mining algorithms to find the meaningful patterns so as to provide better reviews and explore this patterns. Human-computer interface devices are being improved for interacting with the data visualization in a seamless and rapid manner.

Thus, data visualization is playing a vital role in the everyday life of human beings by analyzing various types and also huge amounts of data for predictive analysis and understanding the nature of the data easily.

References

Love, %. (2014, December 03). Geospatial Heat Maps in Tableau via Alteryx. Retrieved April 22, 2017, from <https://www.theinformationlab.co.uk/2014/12/03/geospatial-heat-maps-tableau-via-alteryx/>

Scientific Visualization Tutorial. (n.d.). Retrieved April 22, 2017, from <http://www.cc.gatech.edu/scivis/tutorial/linked/whatis-scivis.html>

(2016, August 1). Retrieved April 22, 2017, from <https://www.seoclerk.com/faq/18531/Does-using-a-wordcloud-in-website-help-you-rank-better>

Kosara, R. (n.d.). Visualization Criticism – The Missing Link Between Information Visualization and Art. Retrieved April 22, 2017, from http://kosara.net/papers/2007/Kosara_IV_2007.pdf

Information Visualization. (n.d.). Retrieved April 22, 2017, from http://www.infovis-wiki.net/index.php?title=Information_Visualization

Santos, S. D., & Brodlie, K. (2004, June). Retrieved April 22, 2017, from <http://www.sciencedirect.com/science/article/pii/S0097849304000251>

(n.d.). Retrieved April 22, 2017, from <http://hci.stanford.edu/courses/cs448b/fl1/lectures/CS448B-20111117-Text.pdf>

Hierarchical Data Visualization. (n.d.). Retrieved April 22, 2017, from <http://cs.iupui.edu/~sfang/cs552/cs552-hierarchy.pdf>

Becker, R. A., Eick, S. G., & Wilks, A. R. (1995, March). Visualizing Network Data. Retrieved April 22, 2017, from <http://igva2013.wikispaces.asu.edu/file/view/visualizingNetworkData.pdf>

Briggs, W. B. (2012, January 25). How To Cheat, Or Fool Yourself, With Time Series: Climate Example. Retrieved April 22, 2017, from <http://wmbriggs.com/post/5107/>

BUTMAN, D. (n.d.). Big Ideas Made Simple. Retrieved April 22, 2017, from <http://www.ib5k.com/blog/posts/choropleth-mapping-with-react-and-d3>

Fregia. (2013, April 02). Fundamental Counting Principle. Retrieved April 22, 2017, from <https://www.pinterest.com/pin/97108935687650076/>

Visualizing Categorical Data as Flows with Alluvial Diagrams. (2014, June 3). Retrieved April 22, 2017, from <http://digitalsplashmedia.com/2014/06/visualizing-categorical-data-as-flows-with-alluvial-diagrams/>

M. (2012, May). What is data visualization? - Definition from WhatIs.com. Retrieved April 22, 2017, from <http://searchbusinessanalytics.techtarget.com/definition/data-visualization>

Heer , J. (n.d.). Research topics in Data Visualization. Retrieved April 22, 2017, from <https://cs.stanford.edu/degrees/phd/cs300/jeffrey-heer.pdf>

Data Visualization: What it is and why it matters. (n.d.). Retrieved April 22, 2017, from https://www.sas.com/en_us/insights/big-data/data-visualization.html

JavaScript Line Charts. (n.d.). Retrieved April 22, 2017, from <http://canvasjs.com/html5-javascript-line-chart/>

Kędzia, %. (n.d.). Retrieved April 22, 2017, from <http://best-excel-tutorial.com/56-charts/76-histogram>

Redko, A. (2014, January). Release: JavaFX 2.2. Retrieved April 22, 2017, from <http://docs.oracle.com/javafx/2/charts/bar-chart.htm>

(n.d.). Retrieved April 22, 2017, from <http://6.anychart.com/products/anychart/docs/users-guide/Stacked-Bar-Column-Chart.html?fromtree>

Public Health & QI Toolbox. (n.d.). Retrieved April 22, 2017, from <http://www.health.state.mn.us/divs/opi/qi/toolbox/scatter.html>

How to Make an Area Chart in Excel. (n.d.). Retrieved April 22, 2017, from <https://www.qimacros.com/excel-charts-qimacros/excel-area-chart/>

Pasi, %. (2013, May 7). Posts about Box-and-whisker plot on Intro to Statistical Methods. Retrieved April 22, 2017, from <https://statsmethods.wordpress.com/tag/box-and-whisker-plot/>

(n.d.). Retrieved April 22, 2017, from <http://study.com/academy/lesson/what-is-a-pie-chart-definition-examples-quiz.html>

Donut Chart. (n.d.). Retrieved April 22, 2017, from <https://www.shieldui.com/documentation/asp.net.chart/chart.types/donut>

Spider (Radar) chart for iOS. (n.d.). Retrieved April 22, 2017, from <http://stackoverflow.com/questions/12720462/spider-radar-chart-for-ios>

Turley, P. (2017, February 17). Gartner Magic Quadrant for BI 2017 Top 3 Rankings. Retrieved April 22, 2017, from <https://sqlserverbi.blog/2017/02/17/gartner-magic-quadrant-for-bi-2017-top-3-rankings/>

Bajaj, C., & Wiley, J. (1998). Retrieved April 22, 2017, from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.6660>

Veltman, N. (2013, November 07). SQL: The Prequel (Excel vs. Databases). Retrieved April 22, 2017, from <https://schoolofdata.org/2013/11/07/sql-databases-vs-excel/>

QlikView vs. Tableau: A Review of Two Visualization Giants. (2014, May 20). Retrieved April 22, 2017, from <https://www.experfy.com/blog/qlikview-vs-tableau-review-two-visualization-giants/>

Make your data make an impact. (n.d.). Retrieved April 22, 2017, from <https://www.tableau.com/>

Tableau Tutorial. (n.d.). Retrieved April 22, 2017, from <https://www.tutorialspoint.com/tableau/>

J. U. (2016, February 9). Gartner 2016 BI and Analytics Magic Quadrant. Retrieved April 22, 2017, from <http://www.jenunderwood.com/2016/02/09/big-changes-in-gartners-2016-magic-quadrant-for-bi-and-analytics/>

Sleeper, R. (2016, March 16). Tableau 201: How to Make a Symbol Map with Mapbox. Retrieved April 22, 2017, from <http://www.evolytics.com/blog/tableau-201-make-symbol-map-mapbox-integration/>

Chandramouly, A. (2015, February 24). Gartner Positions Tableau as. Retrieved April 22, 2017, from <http://www.tableau.com/about/blog/2015/2/gartner-positions-tableau-leader-2015-magic-quadrant>

Measure Values and Measure Names. (n.d.). Retrieved April 22, 2017, from http://onlinehelp.tableau.com/current/pro/desktop/en-us/datafields_understanddatawindow_meavalues.html

Hardin, %, Hom, D., Perez, R., & Williams, L. (n.d.). Which chart or graph is right for you?

Retrieved April 22, 2017, from

http://www.tableau.com/sites/default/files/media/which_chart_v6_final_0.pdf

(n.d.). Retrieved April 22, 2017, from <http://powertoolsfortableau.com/tableau-performance-series-custom-sql-connections>

(2013, October 7). Retrieved April 22, 2017, from

<http://fonsecadatascience.com/2013/10/07/using-mapfluence-to-customize-your-viz-urban-mapping-blog/>

Business Objects. (n.d.). Retrieved April 22, 2017, from <http://webiworx.com/book-sample-files.html>

Branch, G. P. (2012, December 01). State Area Measurements and Internal Point Coordinates.

Retrieved April 22, 2017, from <https://www.census.gov/geo/reference/state-area.html>

Chafik, A. (2016, August 2). Retrieved April 22, 2017, from <https://www.quora.com/What-data-visualization-software-is-better-Tableau-or-QlikView>