BIGDATA

Unit-IV-Data visualization
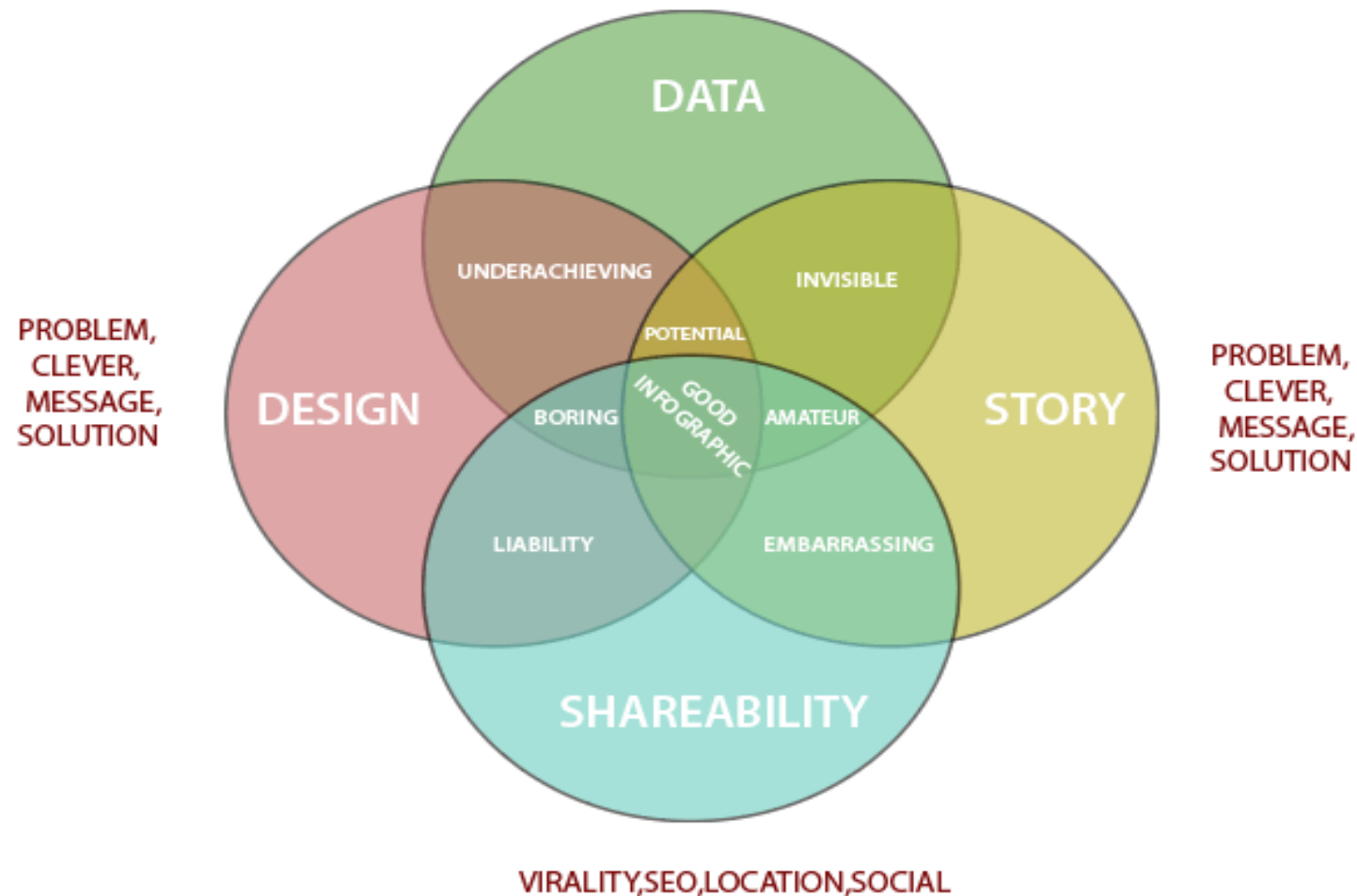
Data visualization

- Data visualizations are common in your everyday life, but they always appear in the form of graphs and charts. The combination of multiple visualizations and bits of information are still referred to as Infographics.

- Data visualizations are used to discover unknown facts and trends. We can see visualizations in the form of line charts to display change over time.

- Bar and column charts are useful for observing relationships and making comparisons. A pie chart is a great way to show parts-of-a-whole,and maps are the best way to share geographical data visually.

- Today's data visualization tools go beyond the charts and graphs used in the Microsoft Excel spreadsheet, which displays the data in more sophisticated ways such as dials and gauges, geographic maps, heat maps, pie chart, and fever chart.

What makes Data Visualization Effective?

- Effective data visualization are created by communication, data science, and design collide. Data visualizations did right key insights into complicated data sets into meaningful and natural.

- American statistician and Yale professor **Edward Tufte** believe useful data visualizations consist of ?complex ideas communicated with clarity, precision, and efficiency.

DATA VISUALIZATION

Importance of Data Visualization

- Data visualization is important because of the processing of information in human brains. Using graphs and charts to visualize a large amount of the complex data sets is more comfortable in comparison to studying the spreadsheet and reports.

- Data visualization is an easy and quick way to convey concepts universally. You can experiment with a different outline by making a slight adjustment.

- **Data visualization have some more specialties such as:**
  - Data visualization can identify areas that need improvement or modifications.
  - Data visualization can clarify which factor influence customer behavior.
  - Data visualization helps you to understand which products to place where.
  - Data visualization can predict sales volumes.

- Data visualization tools have been necessary for democratizing data, analytics, and making data-driven perception available to workers throughout an organization. They are easy to operate in comparison to earlier versions of BI software or traditional statistical analysis software. This guide to a rise in lines of business implementing data visualization tools on their own, without support from IT.

Why Use Data Visualization?

1. To make easier in understand and remember.

2. To discover unknown facts, outliers, and trends.

3. To visualize relationships and patterns quickly.

4. To ask a better question and make better decisions.

5. To competitive analyze.

6. To improve insights.

What Is Big Data Visualization?

- Big data Visualization is , as the name suggests, a visual representation of big data. Visualization techniques vary depending on the goal of the illustration. It could be as simple as line charts, histograms and pie charts or a bit complex like scatter plot, heat maps, tree maps, etc. Visualization of big data can also be done in 3-Dimensional graphs, based on the use case.

- When they say, "Data is the new oil", the data set can be imagined as crude oil and big data visualization as refined oil. In the world of big data, it is often the case that the breadth, let alone the length of a data set even fit on a single screen. We need big data visualization software to make any sense out of such mind-boggling amount of data.

- It allows you to make observations and gather insights. Representing data sets in the form of charts and graphs makes a huge difference in the usability of the data set. Even people who are good with SQL queries prefer a visual format to make observations instead of a tabular format.

# Why Is Data Visualization Important in Big Data?

- Generally, when big data analytics and algorithms are applied to data sets, the results are meant for the decision makers. Other colleagues or top executives would appreciate it if one can reduce time or effort that goes into decision making. It makes their life a lot easier when they are provided with insights in easily consumable format.

- The best part of big data visualization tools is that they are capable of capturing data sets in the visual format without loss of accuracy. One can control the factors like accuracy, precision, level of aggregation that is required to serve the purpose.

- Another major benefit of visualization is the ability to show all the information in a single place. It enables you to create dashboards and reports, which are packed with insights, that can be shared across the organization.

- We need visualization of data in every industry - airline, IoT, energy, media and entertainment, automotive, sports, manufacturing, the list is endless.

- Given the huge scale of data, big data visualization comes with its own set of challenges. Online Big Data Courses provide great resources for building strong big data fundamentals. It will help you make informed choices while choosing appropriate big data visualization techniques.

Types of Big Data Visualization?

1. Line charts

Line chart, also called a line graph or line plot is a common chart. It is used to represent changes in one variable against another, typically the time. The data points are connected by lines. It is used for identifying trends and relationships between two variables. For example, the below chart depicts the sales numbers of three employees.
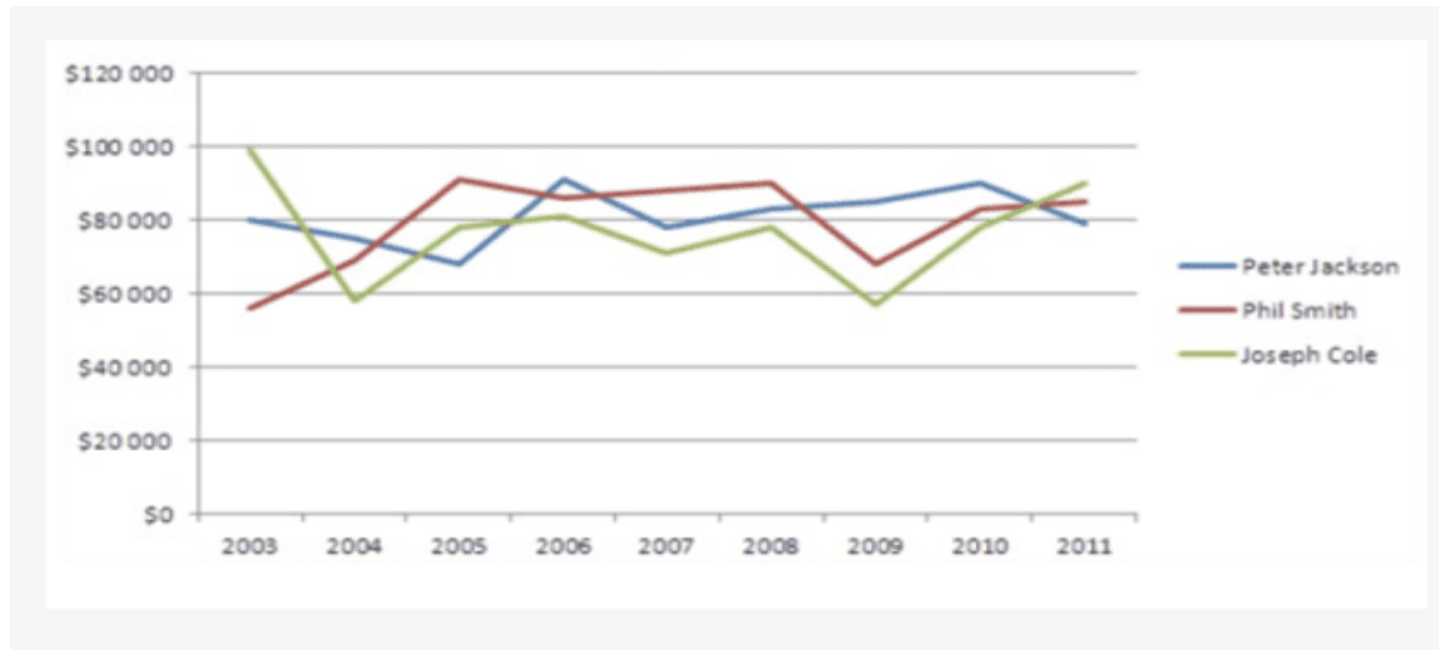
2. Histograms

- A histogram is used to represent the frequency distribution of data. It groups data into logical ranges and depicts the count of how many data points fall into each of those ranges.

- It allows one to understand the nature of frequency distributions. The distribution may be categorized as symmetric, right-skewed and left skewed. For example, how many people are between each range of ages is shown in the following histogram.
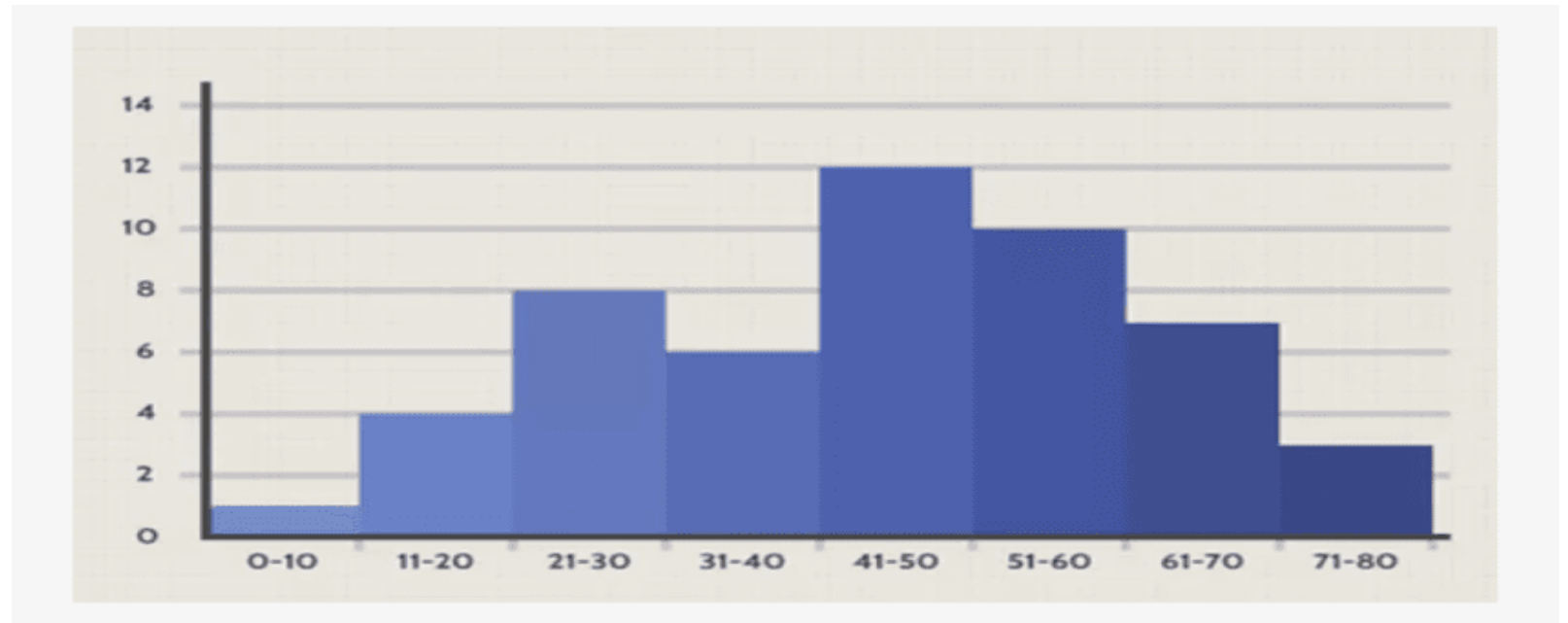
3. Bar chart

Bar chart, also called a bar graph, is used for depicting categorical data with rectangular strips/bars. The length of the bars shows the value or quantity of a variable. The bars might be vertical or horizontal. For example, the below shows how many people like which kind of movies.
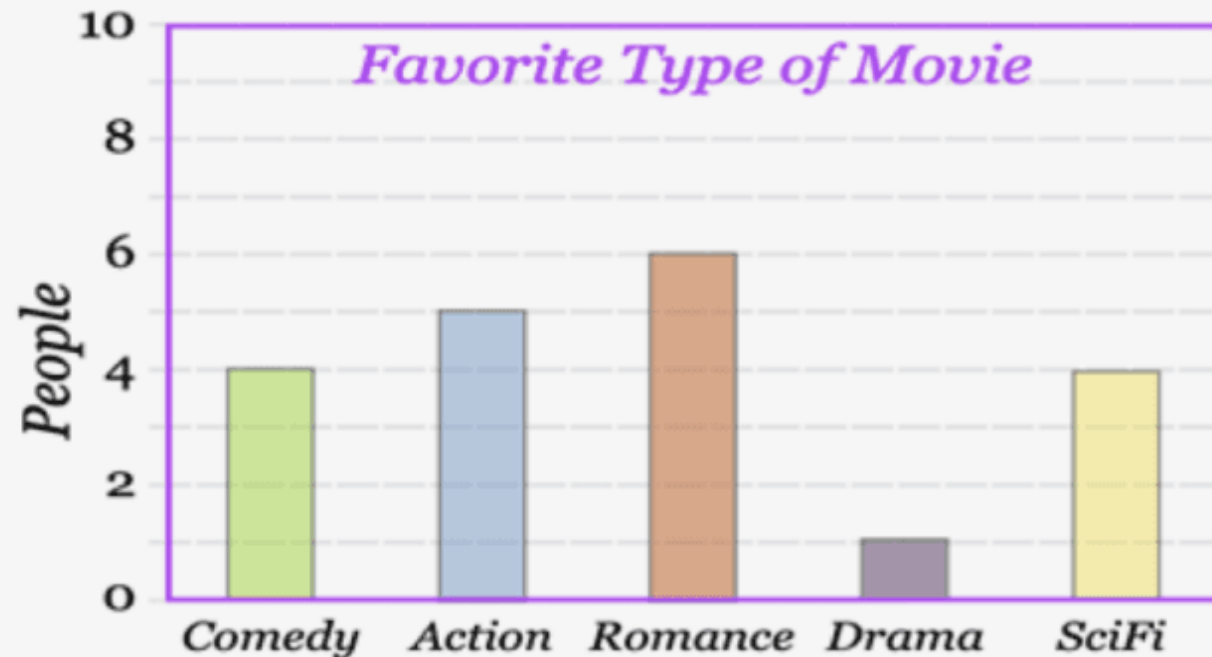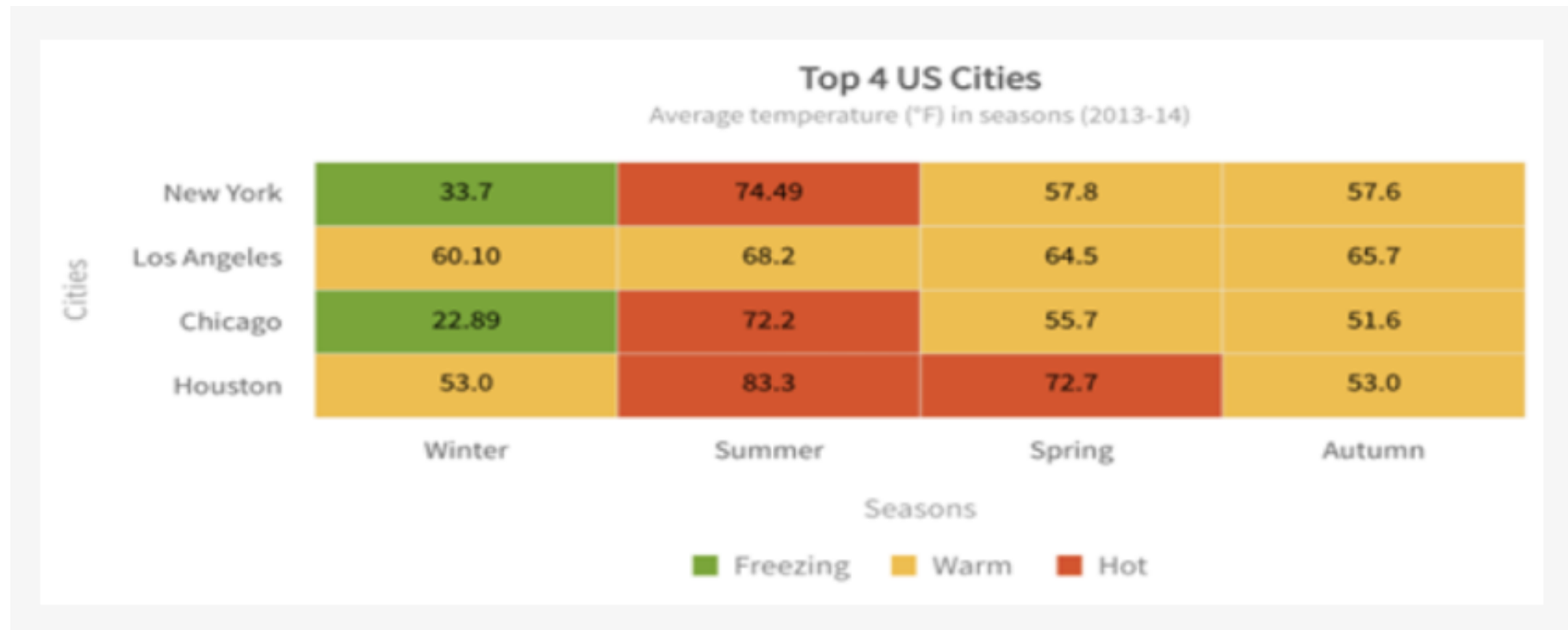
4. Pie charts

Pie chart depicts the information in the form of "pie slices". The "slices" are in proportion to the relative sizes of data. Above example can be represented in the pie chart in the following form.

Favorite Type of Movie

SciFi: 4 (20%)
Comedy: 4 (20%)
Drama: 1 (5%)
Action: 5 (25%)
Romance: 6 (30%)
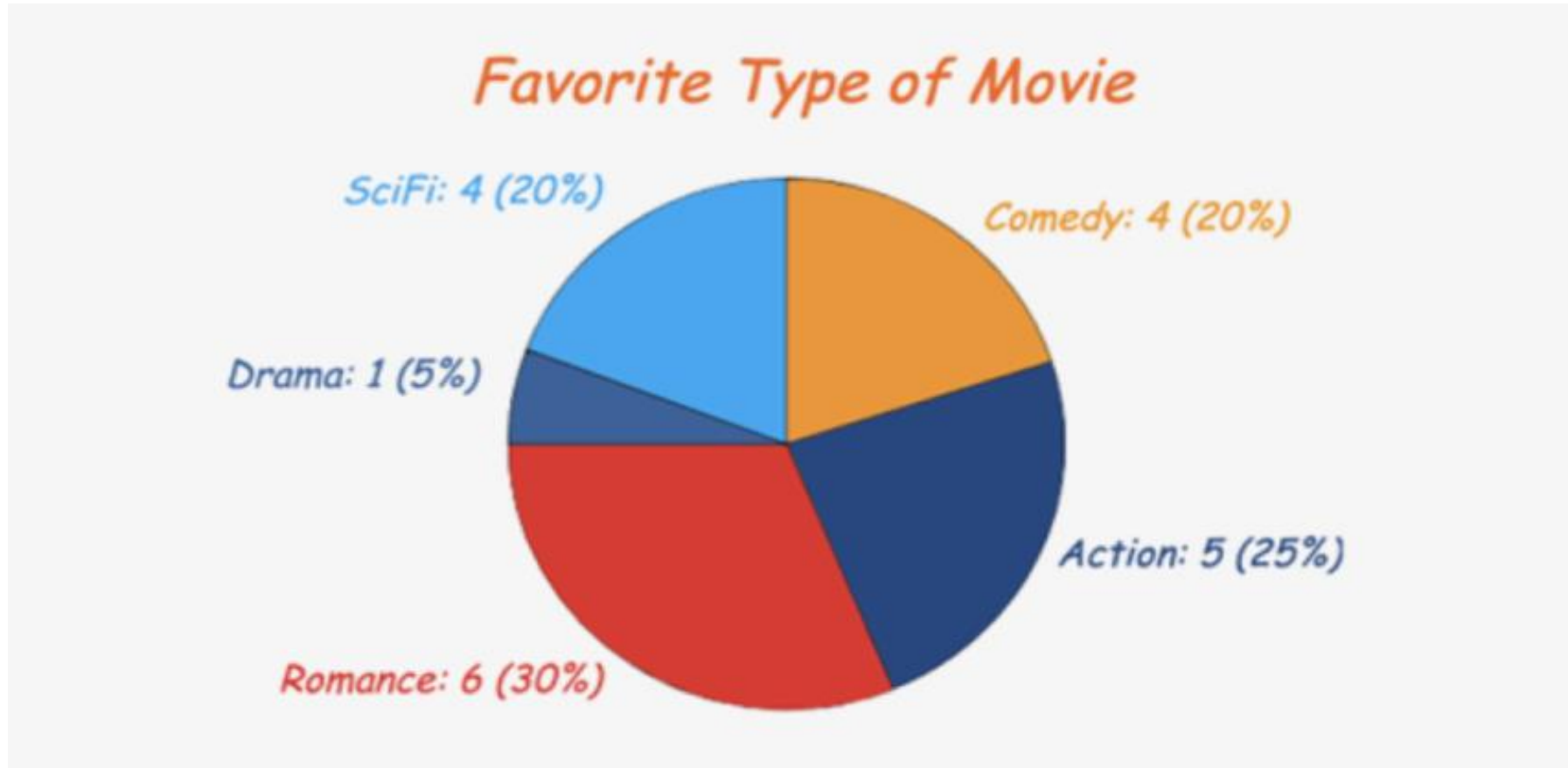
5. Heat Maps

- A heat map uses two-dimensional representation of data in which colors represent the values or ranges. It provides a quick visual summary of information.

- Below is an example of heatmap of temperature-variation data across an year in 4 US cities.

Top 4 US Cities
Average temperature (°F) in seasons (2013-14)

Cities	Winter	Summer	Spring	Autumn
New York	33.7	74.49	57.8	57.6
Los Angeles	60.10	68.2	64.5	65.7
Chicago	22.89	72.2	55.7	51.6
Houston	53.0	83.3	72.7	53.0

Seasons

■ Freezing ■ Warm ■ Hot

6. Scatter plot

- It uses dots/points to show values for numeric variables. The position of the dots against both the axes indicates the value of that particular data point.

- Below is an example of tree height plotted against the girth of the stem.
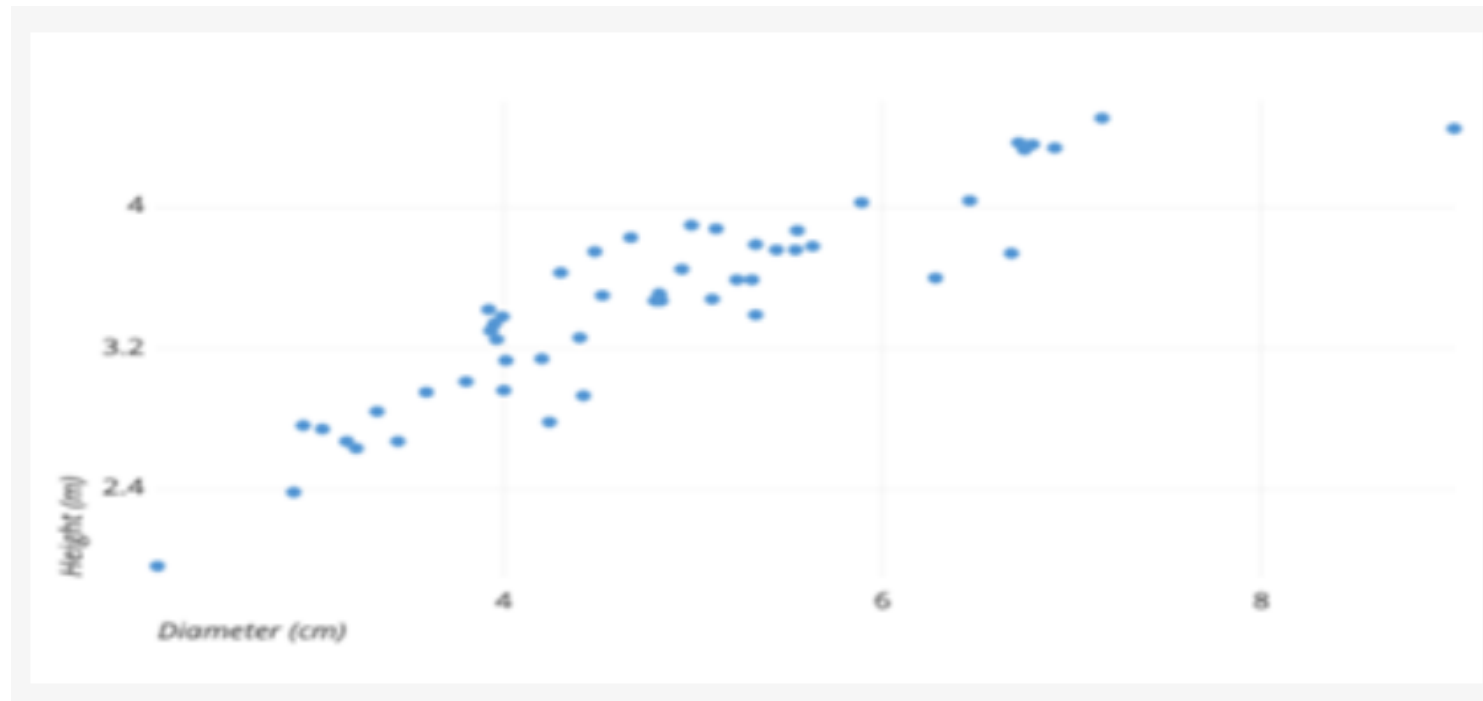
7. Tree map

- This type of chart represents hierarchical data in the format of nested rectangles.

- The size and color of a rectangle represents the value of that category or variable.

- It helps to depict part-to-whole relationships in a complex data set.

Lunch menu

Meals		
Entrees	Deserts	Appetizers
Burger ($8) — Pasta ($8) — Pizza ($8)	Cake ($6) — Pie ($6)	Soup ($6) — Cheese ($4) — Drinks — Water ($2) — Tea ($3) — Coffee ($3)

8. Word cloud

- Word cloud or tag cloud is a representation of word frequency in a data set. The larger the word appears, the higher the frequency of that word. This is used for textual data analysis and summarization. Below is a sample word cloud of jargons commonly used in the big data industry.

- One can also define big data visualization categories in the following manner.

- **Temporal**

- It is a representation of data against time period. For example, gantt charts, timelines, etc.

- **Hierarchical**

- It represents data in tree format. One root node at the top and branches originating from the root. For example, tree map, flow charts.

- **Network**

- It is used when one wants to show connections between various unrelated data sets. Word cloud and matrix charts are examples of network type of visualization.

- **Geospatial**

- Geospatial is a special category in which location data is one of the variables. The variables are plotted against the location variable. Demographic charts, density maps are examples of this category.

Big Data Visualization Tips

- Start with defining the audience and the requirement of the report. Big Data visualization can capture several insights, but what makes the most impact on the audience needs to be identified at the beginning.

- Data can be represented in multiple formats. Deciding the right kind of visualization helps the decision maker to arrive at a decision faster. The visualization should capture all the key insights yet be simple to absorb.

- Providing labels in the report gives meaning to the visual. Some points to keep in mind are - check the alignment of labels, check all required labels are present, add titles that capture the essence of the chart and add legend when needed.

- Make use of multiple colors in the chart, if it makes it better for the viewer to grasp. Colors can be used to encode information without explicitly adding any more elements to the graph. The color scheme should be such that the viewer need not struggle to find the patterns depicted in the graph.

Tools for Big Datass Visualization

- There are many tools available in the market for data visualization in the big data world. Along with these tools, there are languages like R and python that data professionals use for complex analytical functions and plotting. These functions can also be embedded into big data visualization tools for applying to large data sets. Many tools can be categorized as big data analytics visualization tools rather than just visualization tools.

- **Google**, **Apple**, **Facebook**, and **Twitter** are asks , a better question of their data and make a better business decision by using data visualization.

1. Tableau

- Tableau is a well-known visualization tool. Tableau Desktop can collect data from multiple data sources, which can be either on-premises or in the cloud. Users can get started easily without much initial effort. It is built for data professionals as well as professionals from other fields.

- It supports drag-and-drop functionality, interactive dashboards/charts and the ability to connect to multiple data sources simultaneously. Natural language queries are a major plus point for people with limited DBMS knowledge.

- An interactive dashboard helps immensely with story-telling, say, in front of stakeholders.

- It provides support from data preparation, doing analysis and creating reports for sharing amongst colleagues for better decision making.

- Tableau provides SAAS model as well as on-prem model. It has a mobile app too for analysis and viewing.

- By using Tableau we can create graphs, charts, maps, and many other graphics.

- A tableau desktop app is available for visual analytics. If you don't want to install tableau software on your desktop, then a server solution allows you to visualize your reports online and on mobile.

- A cloud-hosted service also is an option for those who want the server solution but don't want to set up manually. The customers of Tableau include Barclays, Pandora, and Citrix.
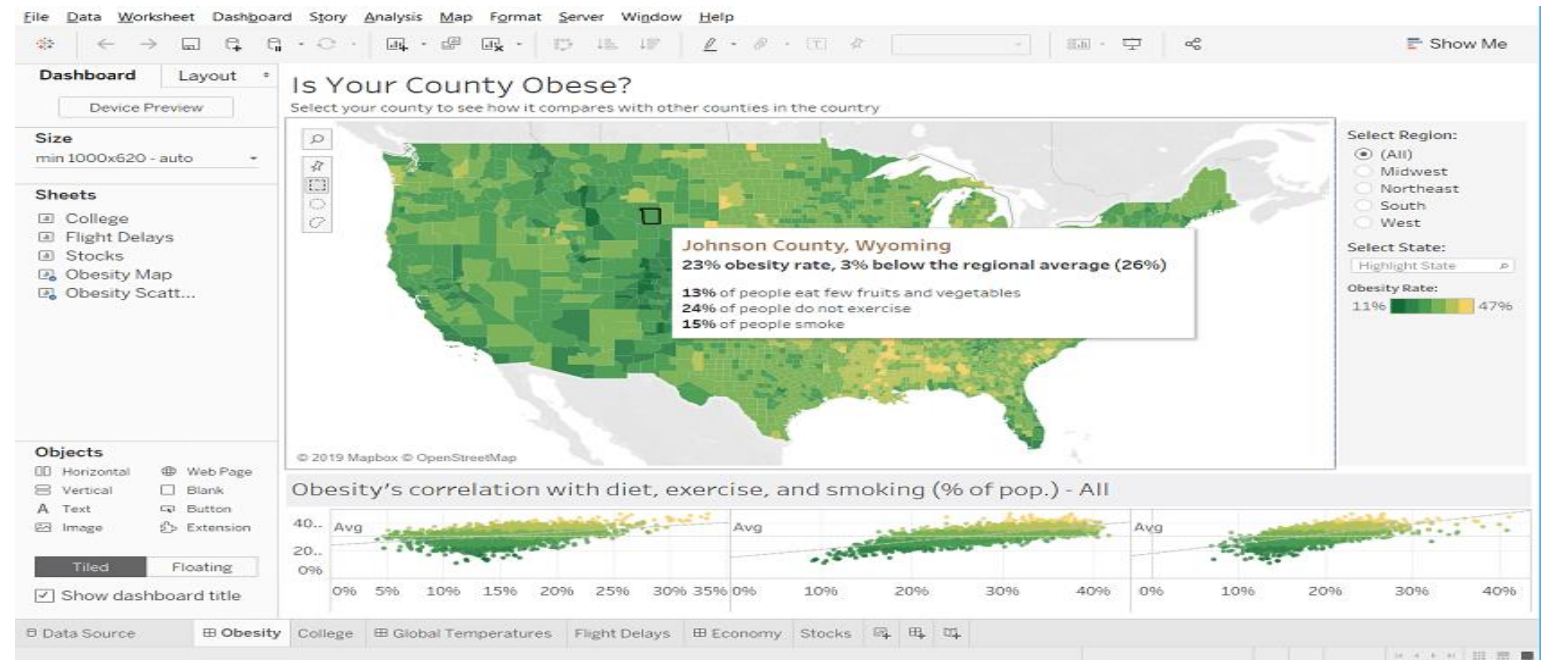
2. Info gram

Info gram is a data visualization tool. It has some simple steps to process that:

1. First, you choose among many templates, personalize them with additional visualizations like maps, charts, videos, and images.

2. Then you are ready to share your visualization.

3. Info gram supports team accounts for journalists and media publishers, branded designs of classroom accounts for educational projects, companies, and enterprises.

- An info gram is a representation of information in a graphic format designed to make the data easily understandable in a view. Infogram is used to quickly communicate a message, to simplify the presentation of large amounts of the dataset, to see data patterns and relationships, and to monitor changes in variables over time.

- Infogram abounds in almost any public environment such as traffic signs, subway maps, tag clouds, musical scores, and weather charts, among a huge number of possibilities.

Infogram
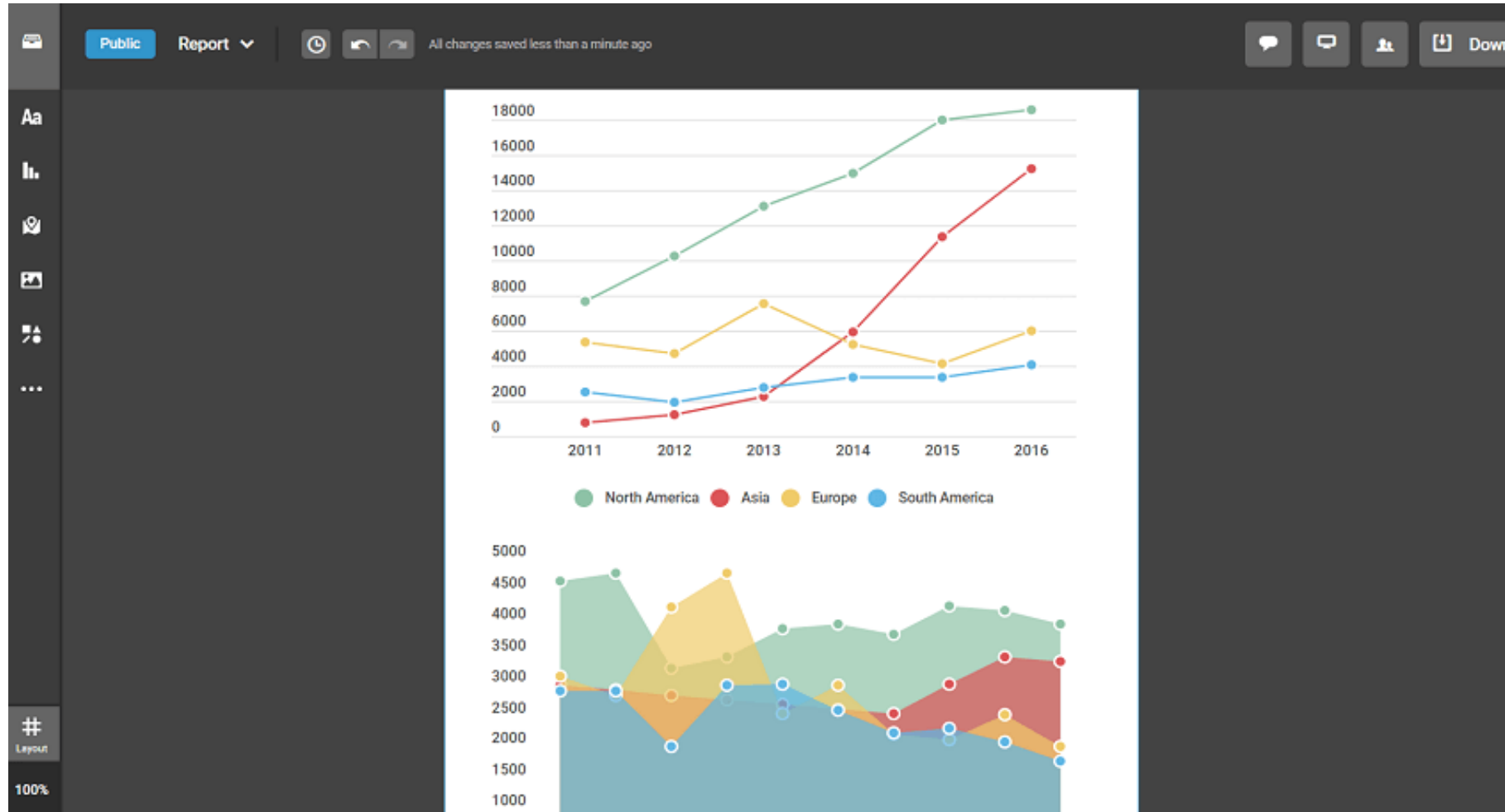
3. Chart blocks

- Chart blocks is an easy way to use online tool which required no coding and builds visualization from databases, spreadsheets, and live feeds.

- Your chart is created under the hood in **html5** by using the powerful JavaScript library **D3.js**. Your visualizations is responsive and compatible with any screen size and device. Also, you will be able to embed your charts on any web page, and you can share it on Facebook and Twitter.
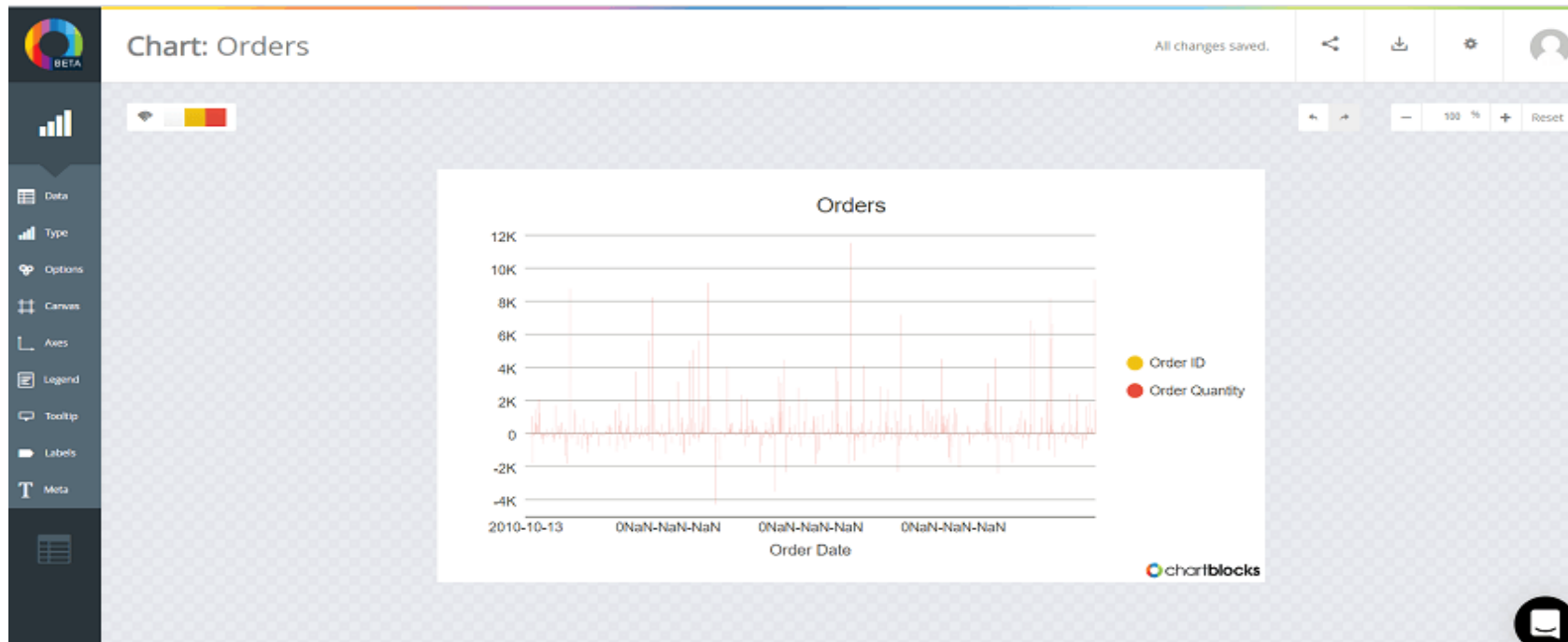
4. Data wrapper

- Datawrapper is an aimed squarely at publisher and journalist. **The Washington Post, VOX, The Guardian, BuzzFeed, The Wall Street Journal** and **Twitter** adopts it.

- Datawrapper is easy visualization tool, and it requires zero codings. We can upload your data and easily create and publish a map or a chart.

- The custom layouts to integrate your visualizations perfectly on your site and access to local area maps are also available.

.
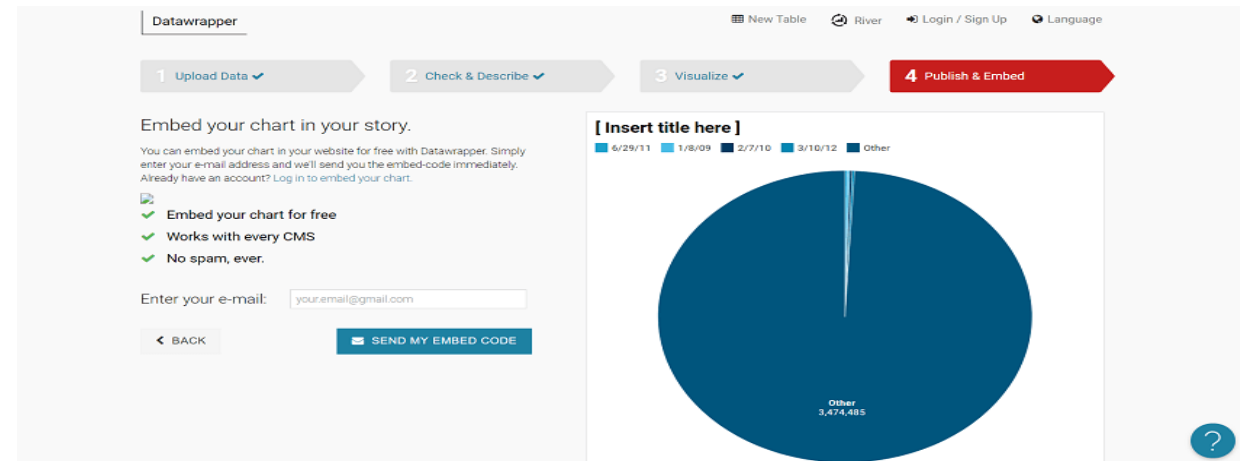
5. Plotly

- Plotly will help you to create a slick and sharp chart in just a few minutes or in a very short time. It also starts from a simple spreadsheet.

- We use Plotly at Google and also by the US Air Force, Goji and The New York University.

- Plotly is very user-friendly visualization tool which is quickly started within a few minutes. If you are a part of a team of developers that wants to have a crack, an API is available for JavaScript and Python languages.
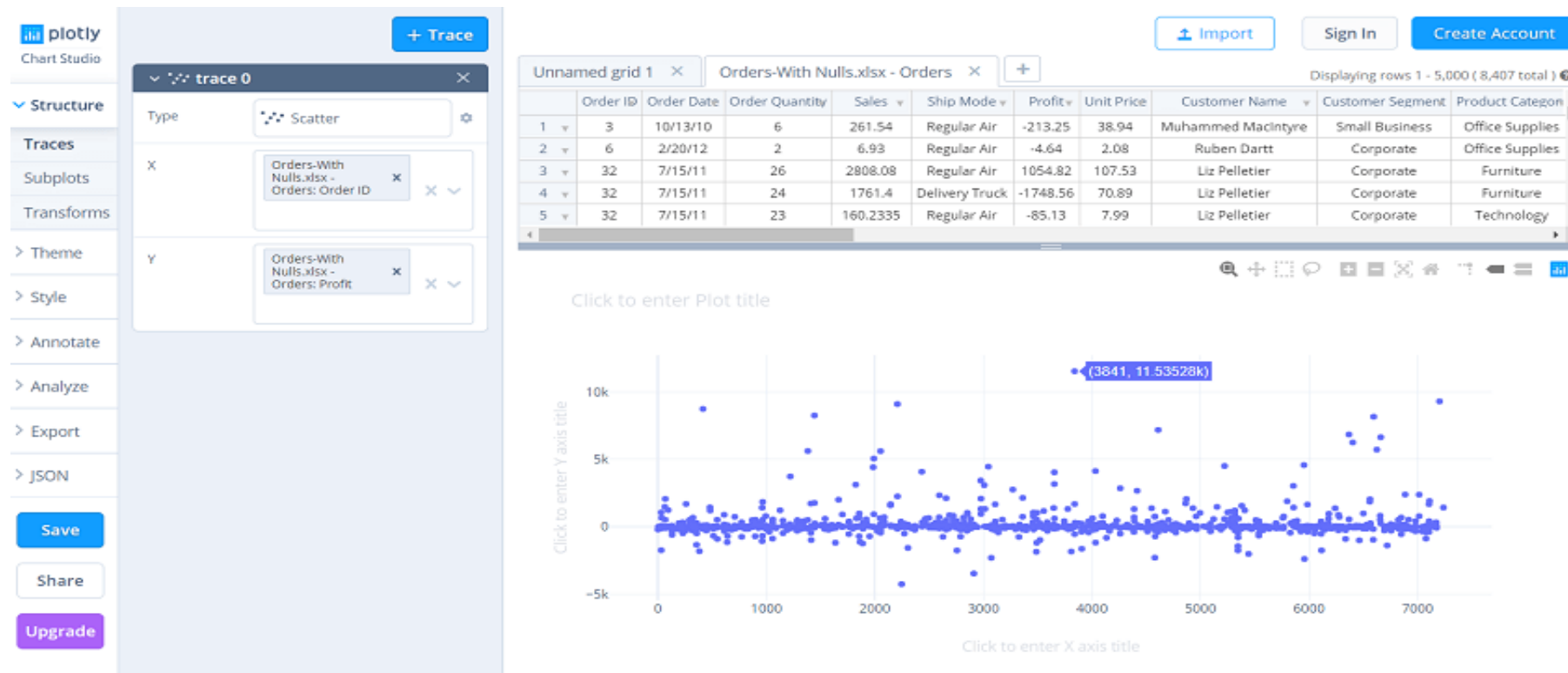
6. RAW

- RAW creates the missing link between spreadsheets and vector graphics on its home page.

- Your Data can come from **Google Docs, Microsoft Excel, Apple Numbers**, or a simple comma-separated list.

- Here the kicker is that you can export your visualization easily and have a designer to make it look sharp. RAW is compatible with Inkscape, Adobe Illustrator, and Sketch. RAW is very easy to use and get quick results.
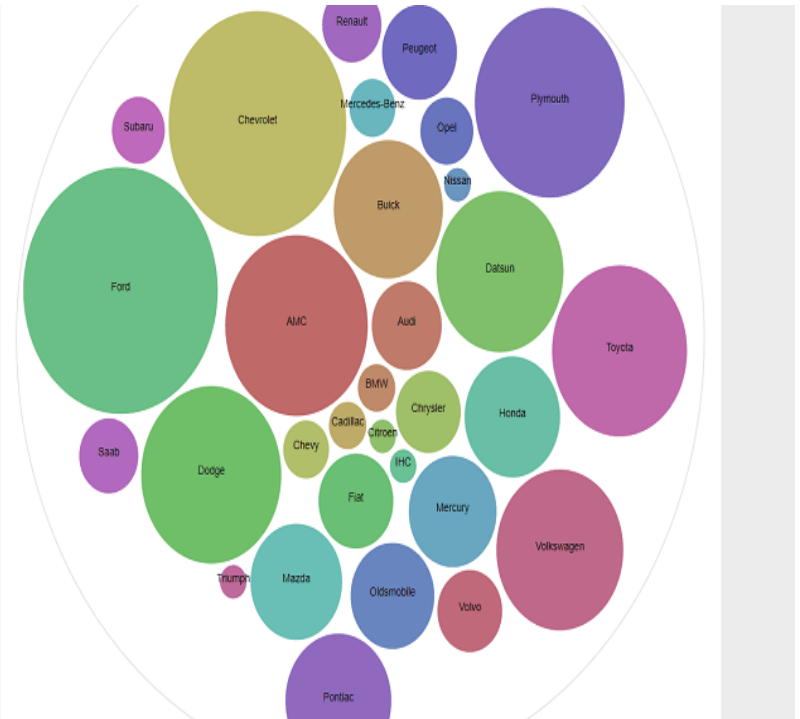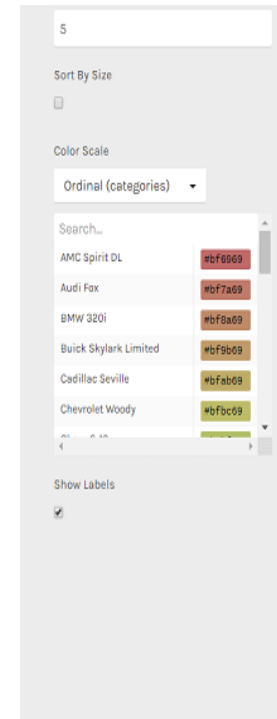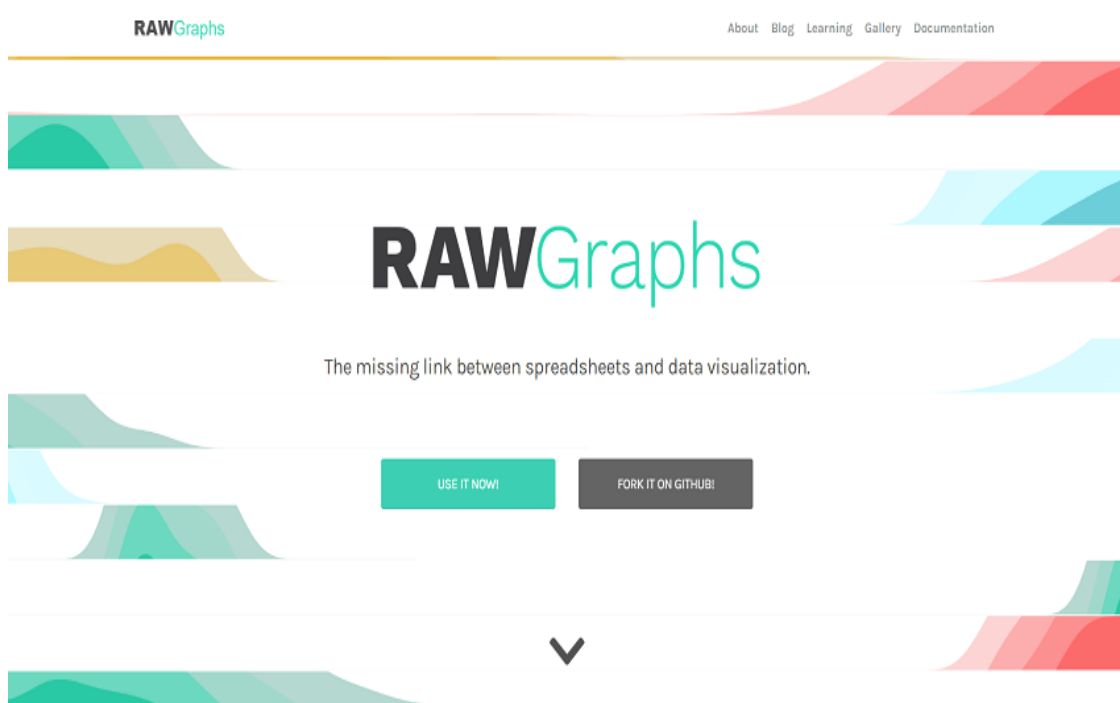
7. Visually

- Visually is a visual content service. It has a dedicated data visualization service and their impressive portfolio that includes work for Nike, VISA, Twitter, Ford, The Huffington post, and the national geographic.

- By a streamlined online process, you can find entire outsource your visualizations to a third-party where you describe your project and connected with a creative team that will stay with you for the entire duration of the project.

- Visually sends you an email notification for all the event you are hitting, and also it will give you constant feedback to your creative team. Visually offer their distribution network for showcasing your project after it's completed.

8. D3.js

- D3.js is a best data visualization library for manipulating documents. D3.js runs on JavaScript, and it uses **CSS**, **html**, and **SVG**. D3.js is an open-source and applies a data-driven transformation to a webpage. It's only applied when data is in **JSON** and **XML** file.

- D3.js emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a single framework and combining powerful visualization components.

- D3.js is as powerful as it is a cutting-edge library, so it comes with no pre-built charts and only IE9+ supports this library.

9. Ember Charts

- Ember charts are based on the ember.js and D3.js framework, and it uses the D3.js under the hood. It also applied when the data is in **JSON** and **XML** file.

- It includes a bar, time series, pie, and scatter charts which are easy to extend and modify. These chart components represent our thoughts on best practices in chart presentation and interactivity.

- The team behind Ember Charts is also the same that created Ember.js. It puts a lot of focus on best practices and interactivity. Error handling is very graceful, and your app will not crash after finding irrelevant data or corrupt data.
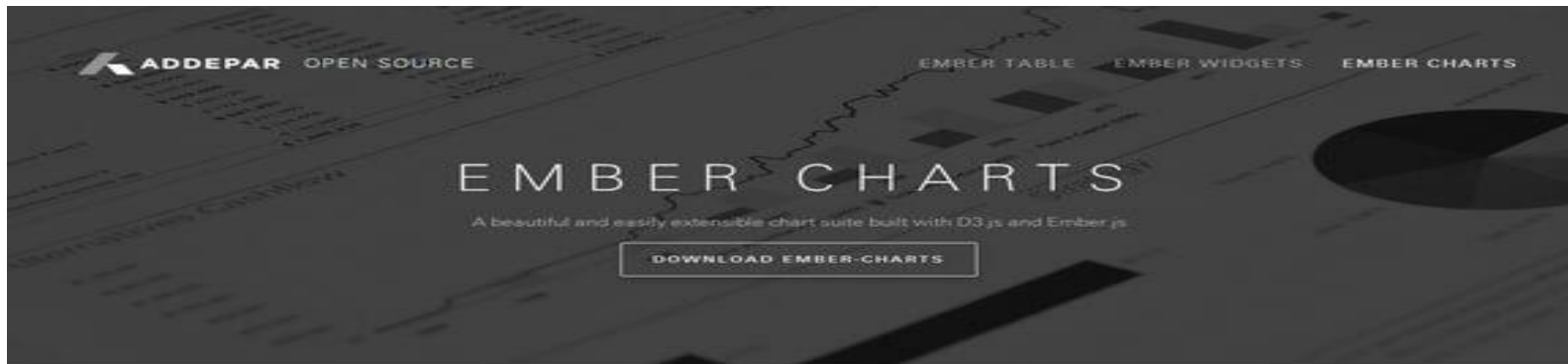
10. NVD3

- NVD3 is a project that attempts to build reusable charts and components. This project is to keeps all your charts neat and consumable

- NDV3 is a simpler interface on the top of the D3.js and keeps all of its powerful features under the hood.

- The front end engineers develop NDV3, and they use their insight into charting technology. This charting technology is used to provide powerful analytics to clients in the financial industry.
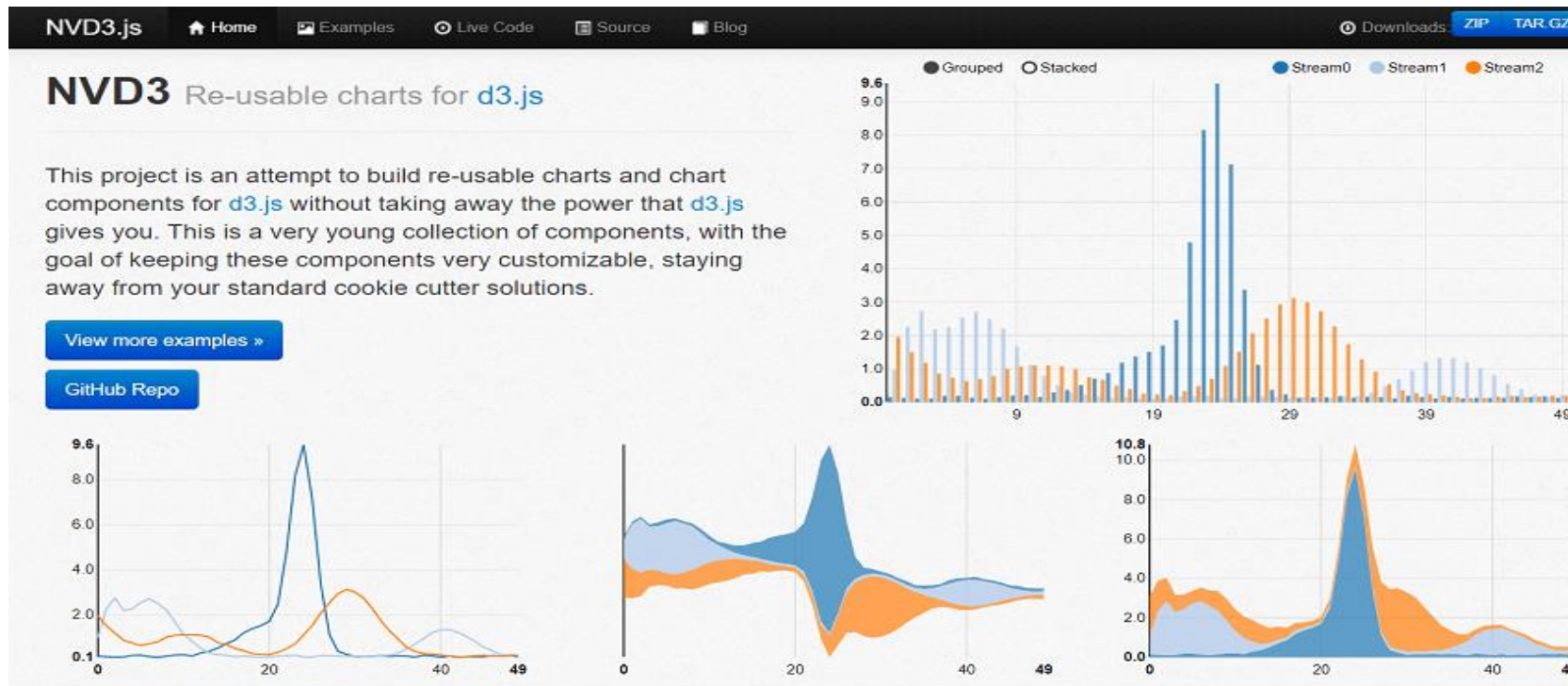
1. What is Data Visualization and explain importance of visualization?'
2. Briefly discuss Visualization Design Principles and Dash Boards ?
3. Explain types of Big Data Visualization?
4. Explain types of Big Data visualization Tools?
5. Discuss Data Exploration?