

# Data Visualization

116N54C301

Module 1

# Introduction data visualization

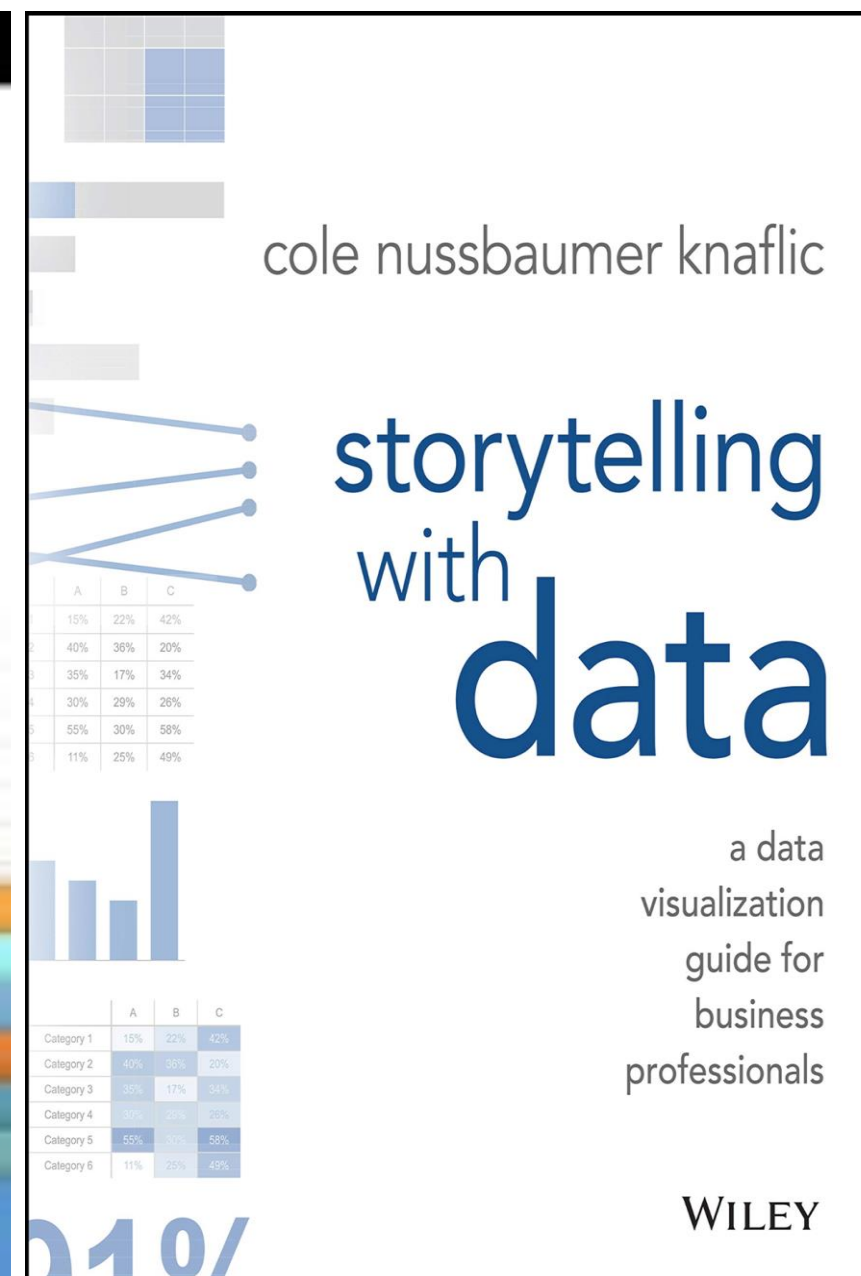
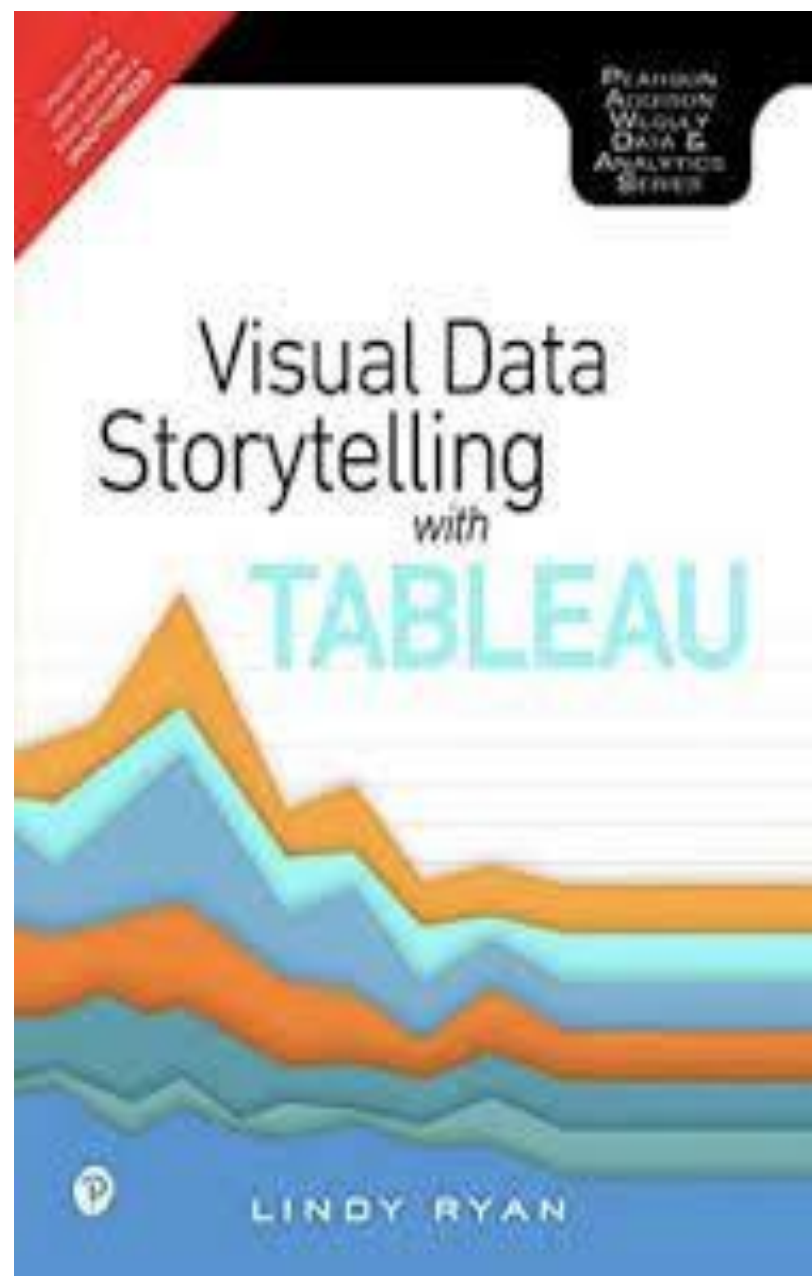
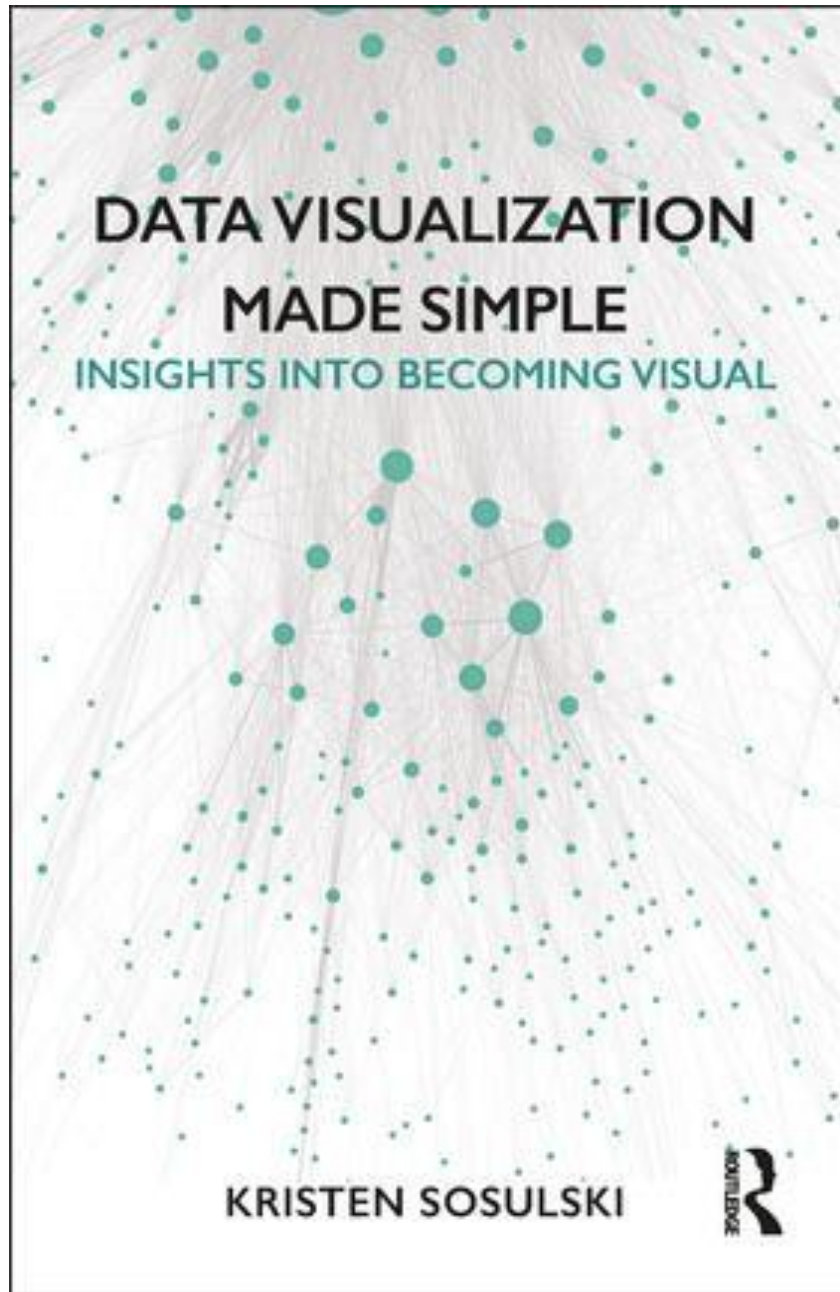
- Introduction to data visualization and it's needs.
- Data analysis lifecycle.
- A Visual Revolution, Various types of visualization with its best practices.
- From Visualization to Visual Data Storytelling: An Evolution, From Visual to Story: Bridging the Gap
- Data Fundamentals, Collecting data, Preparing Data
- Introduction to basic Visualization and it's need.
  - The Bar Chart
  - The Line Chart
  - The Pie and Donut Charts
  - The Scatter Plot
  - The Packed Bubble Chart

# Books

**Sosulski, K.** ,”Data Visualization Made Simple: Insights into Becoming Visual”, New York: Routledge. First edition, 2018

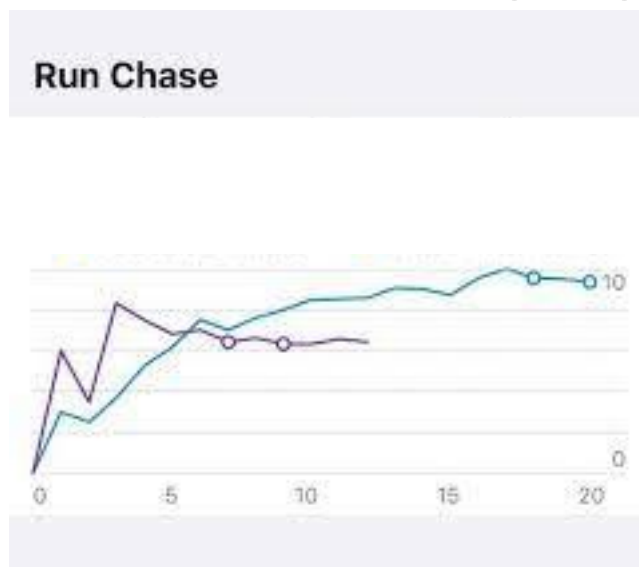
**Lindy Ryan**, “Visual Data Storytelling with Tableau”, Pearson Education First edition, 2018

**Cole Nussbaumer Knaflic**, “Storytelling with Data”, Wiley First edition, 2015

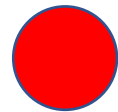


# Data Visualization

- A picture speaks more than 1000 words
- Colors attracts more attention
- Non-verbal communication can influence human mind subtly
- Universal language has no bar



Vegetarian



Non- Vegetarian



# Introduction to data visualization and it's need

- **Data visualization:**

- A graphic representation that expresses the significance of data
- Reveals insights and patterns that are not immediately visible in the raw data
- An art through which information, numbers, and measurements can be made easy to understand and manipulate making the information useful
- **Primary goal is to communicate information clearly and effectively through graphical means**
- **Pictorial format representing some form of collected data which would help in making decision or navigation**

# Introduction to data visualization and it's need

- **Data visualization:**
  - 65% of the audience is visual learner
  - Easy to retain
    - 80% of what we see compared to 20% of what we read 0% of what we hear
  - It takes just 1/10 seconds for brain to process as an image
  - Pictures would help in
    - Decision making
    - Uncovering patterns and trends
    - Presenting arguments or telling story

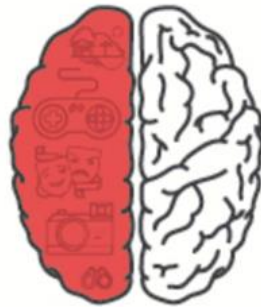
# Introduction to data visualization and it's need

- **Data visualization:**
  - **65% of the audience is visual learner**
  - **Tools for Visual Learners**
    - Use mind pictures or mind maps
    - Take notes
    - Use "clue" words for recalling
    - Use colored highlighters to color code texts and notes
    - Use maps, charts, diagrams, and lists
    - Watch audiovisuals
    - Take photographs
    - Use study cards or flashcards
    - Use notebooks
    - Watch instructor's mouth and face
    - Use visual chains or mnemonics
    - Watch TV
    - See parts of words
    - Write down directions



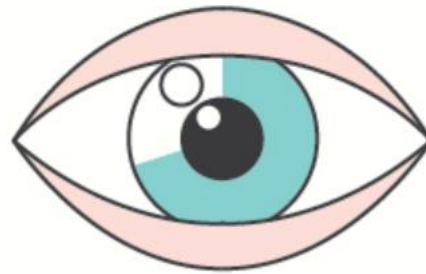
# Introduction to data visualization and it's need

- **Data visualization:**
  - **Easy to retain**
    - **80% of what we see compared to 20% of what we read 10% of what we hear**



ALMOST  
**50%**

of your brain is involved  
in **visual processing**



**70%**

of all your **sensory  
receptors** are in your eyes



**1/10**  
OF A SECOND

is all it takes you to  
understand a **visual scene**

# Introduction to data visualization and it's need

- Data visualization:
  - Easy to retain
    - 80% of what we see compared to 20% of what we read 10% of what we hear

Actually, it only takes us



**EXAMPLE:**



**VS**

While you are travelling down this road there is a chance that one or more rocks of a varying size may fall from the slopes on one or both sides of you. You should be aware of this before you travel this way so that you are cautious of this particular type of hazard.

# Introduction to data visualization and it's need

- Data visualization:
  - Easy to retain
    - 80% of what we see compared to 20% of what we read 10% of what we hear

A study on comprehension rates of medicine labels found a:



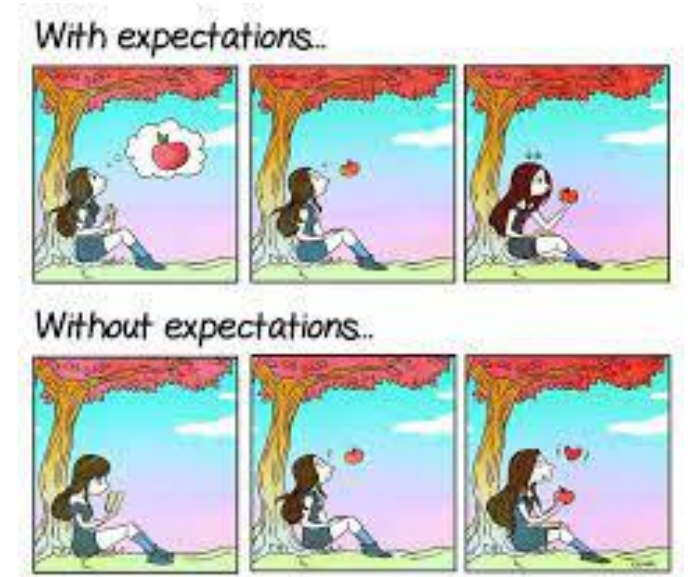
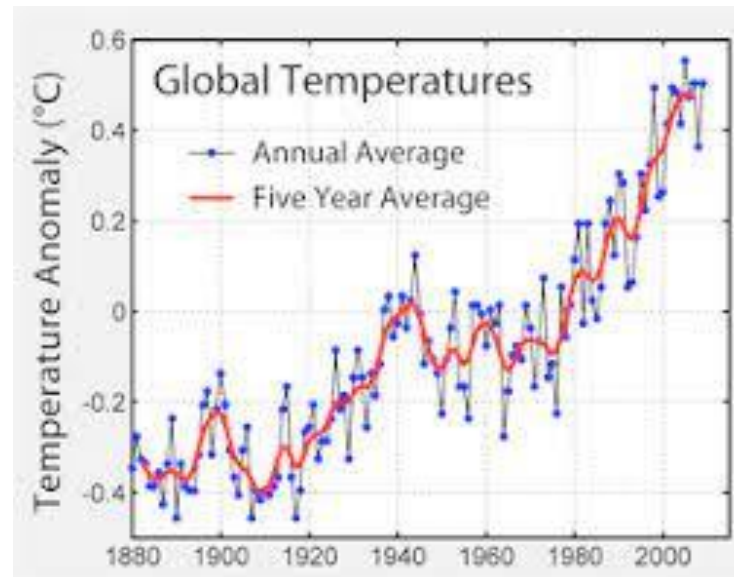
rate of understanding for labels with **text only**



rate of understanding for labels with **text and pictures**

# Introduction to data visualization and it's need

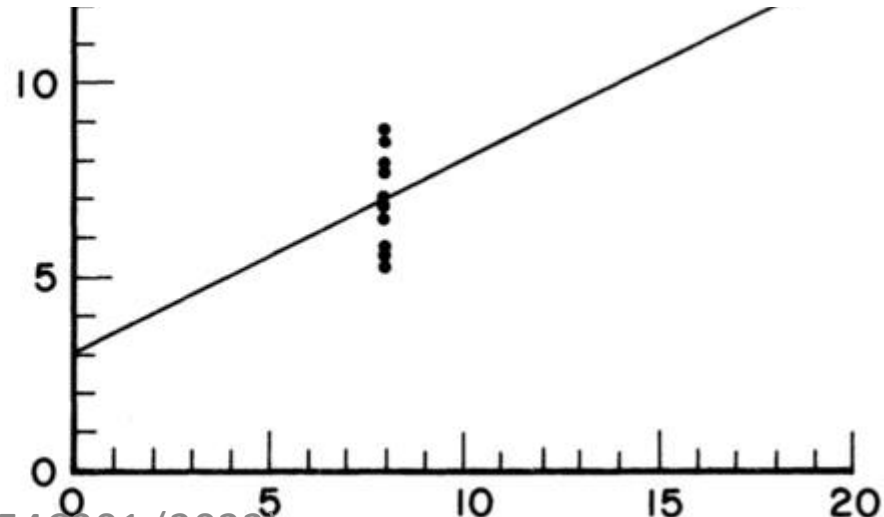
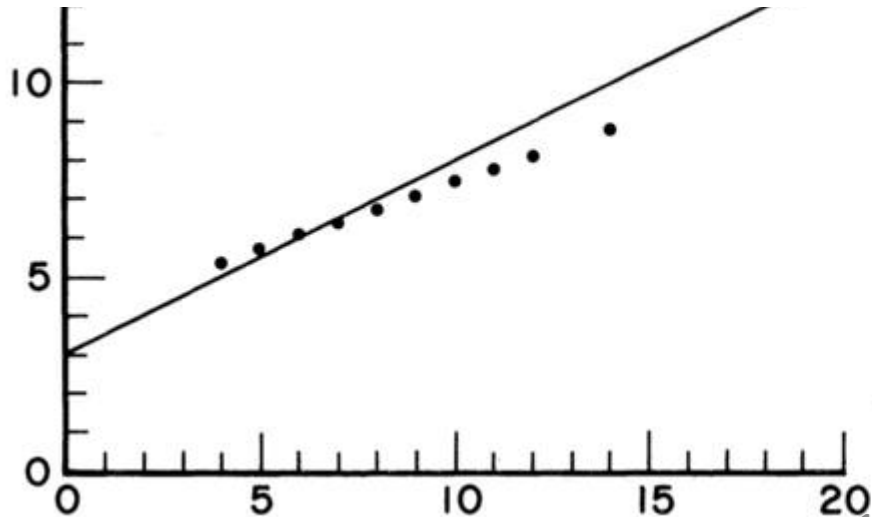
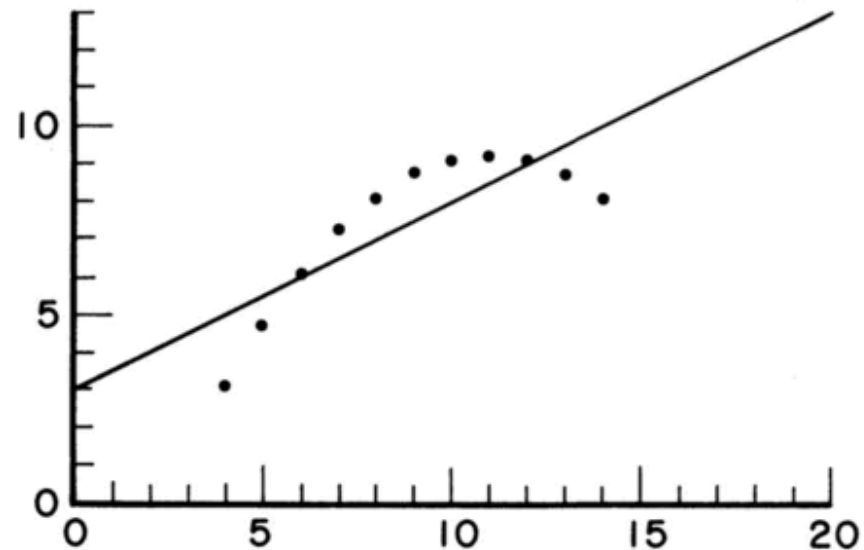
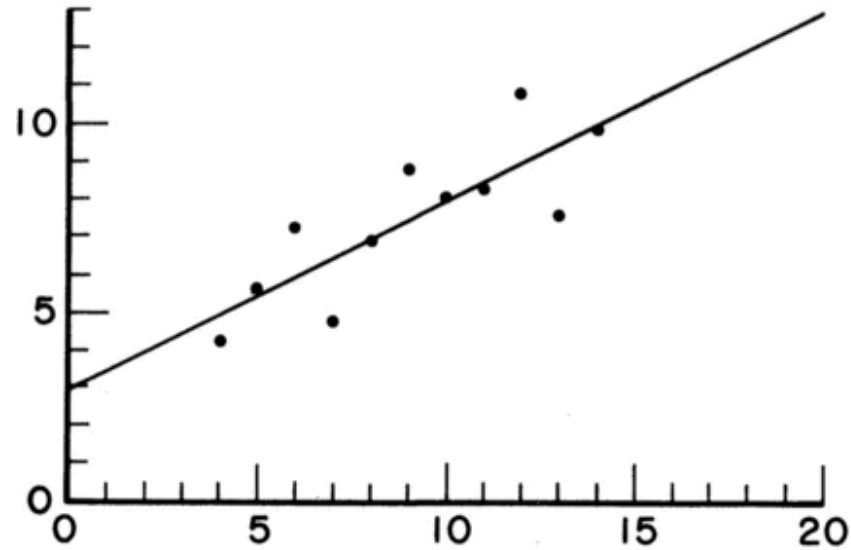
- **Data visualization:**
  - It takes just 1/10 seconds for brain to process as an image
  - Pictures would help in
    - Decision making
    - Uncovering patterns and trends
    - Presenting arguments or telling story



# Importance of data visualisation

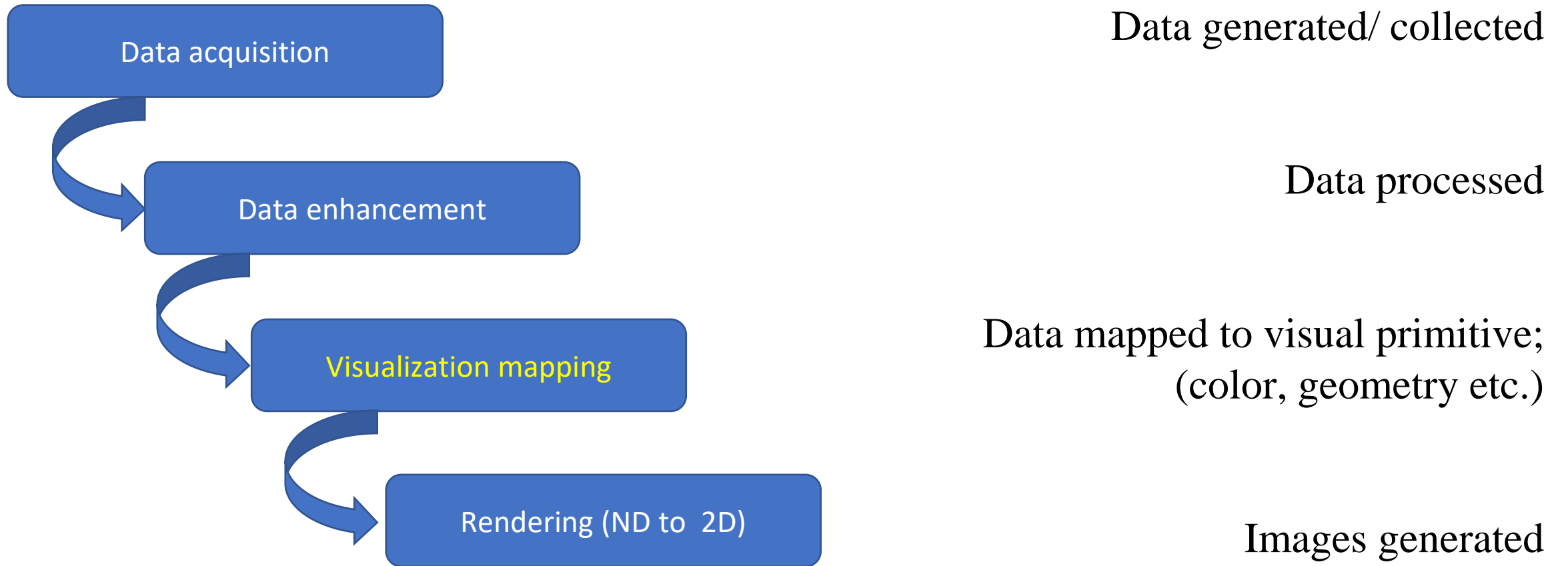
Data set	1-3	1	2	3	4	4
Variable	x	y	y	y	x	y
Obs. no. 1 :	10.0	8.04	9.14	7.46	8.0	6.58
2 :	8.0	6.95	8.14	6.77	8.0	5.76
3 :	13.0	7.58	8.74	12.74	8.0	7.71
4 :	9.0	8.81	8.77	7.11	8.0	8.84
5 :	11.0	8.33	9.26	7.81	8.0	8.47
6 :	14.0	9.96	8.10	8.84	8.0	7.04
7 :	6.0	7.24	6.13	6.08	8.0	5.25
8 :	4.0	4.26	3.10	5.39	19.0	12.50
9 :	12.0	10.84	9.13	8.15	8.0	5.56
10 :	7.0	4.82	7.26	6.42	8.0	7.91
11 :	5.0	5.68	4.74	5.73	8.0	6.89

# Same data presented in graphical form





# Visualization steps



# Types of Visualisation

## **Scientific Visualisation SciVis:**

- used for the clarification of well-known phenomena

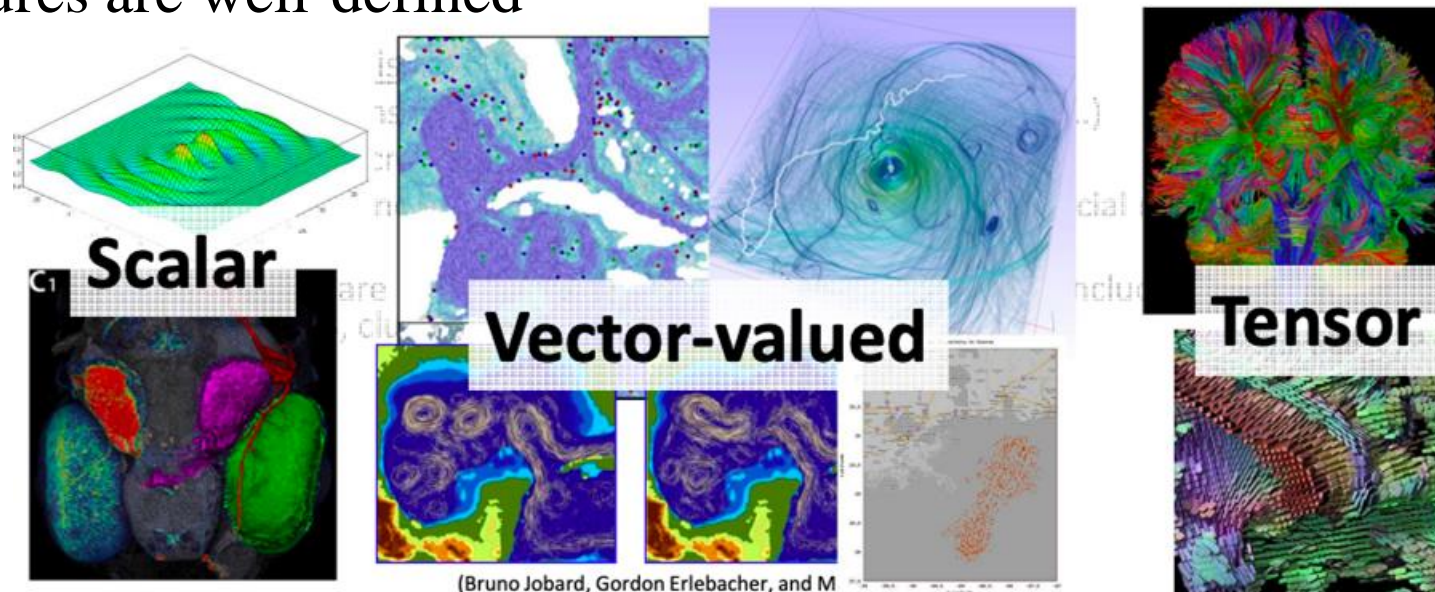
## **Information Visualisation InfoVis:**

- used for searching for interesting phenomena



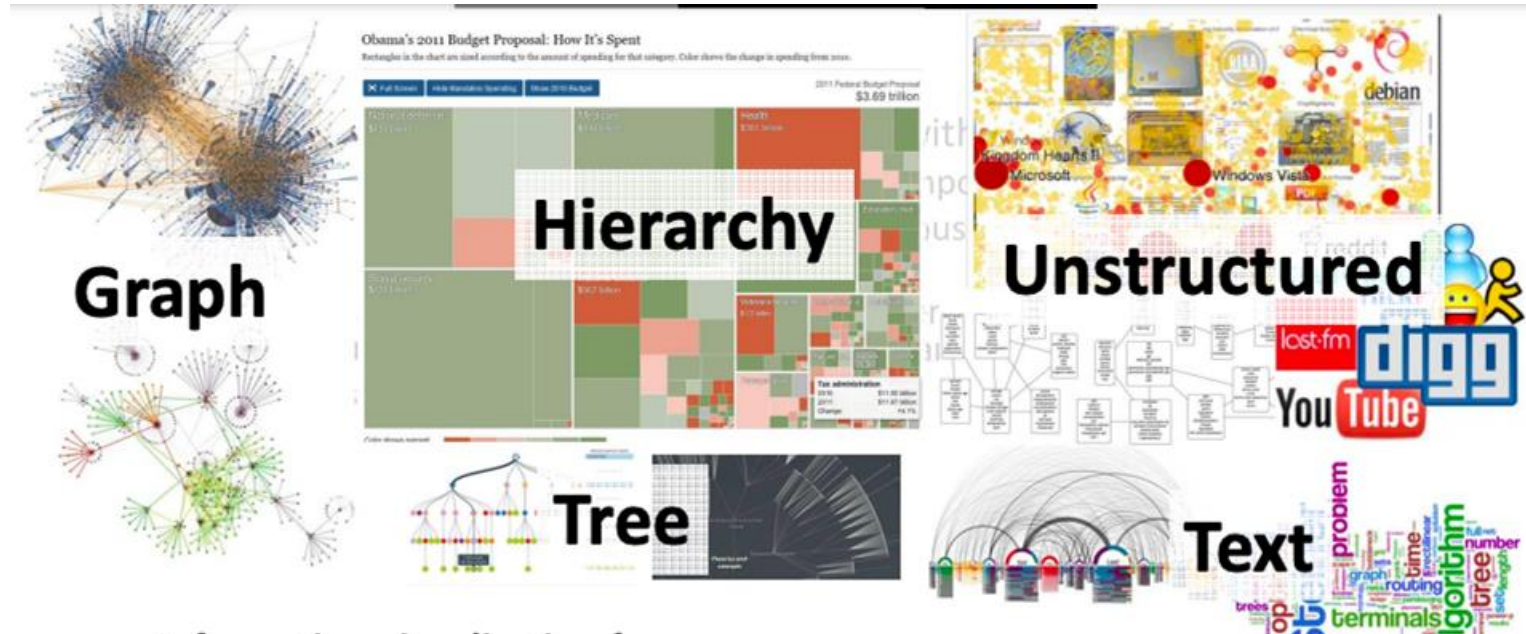
# SciVis vs. InfoVis

- Scientific visualization is mostly concerned with: –
  - Data defined in physical space, i.e. spatio-temporal data (2~4 dimensions)
  - Data describes continuous events in continuous space, however, the representation is discrete (i.e. sampled data)
  - Examples include simulation and measurement data from physics, chemistry, geo-science, medical-biological, climate, oceanography, energy, .....
  - Features are well-defined



# SciVis vs. InfoVis

- Information visualization focuses on:
  - high-dimensional ( $\gg 4$ ), abstract data (i.e. tree, graphs, hierarchy, ...)
  - Data is discrete in the nature
  - Examples include financial, marketing, HR, statistical, social media, political, .....
  - Feature are not well-defined, the typical analysis tasks including finding patterns, clusters, voids, outliers



# Need of Data Visualization

- Data visualization **helps to tell stories by curating data into a form easier to understand, highlighting the trends and outliers.**
- **Analysing the Data in a Better way**
- Data visualization can help by delivering data in the most efficient way possible.
- Data visualization takes the raw data, models it, and delivers the data so that conclusions can be reached.
- **Faster Decision Making**
- Data visualization uses visual data to communicate information in a manner that is universal, fast, and effective.
- **Making Sense of Complicated Data**
- Data visualization positively affects an organization's decision-making process with interactive visual representations of data.

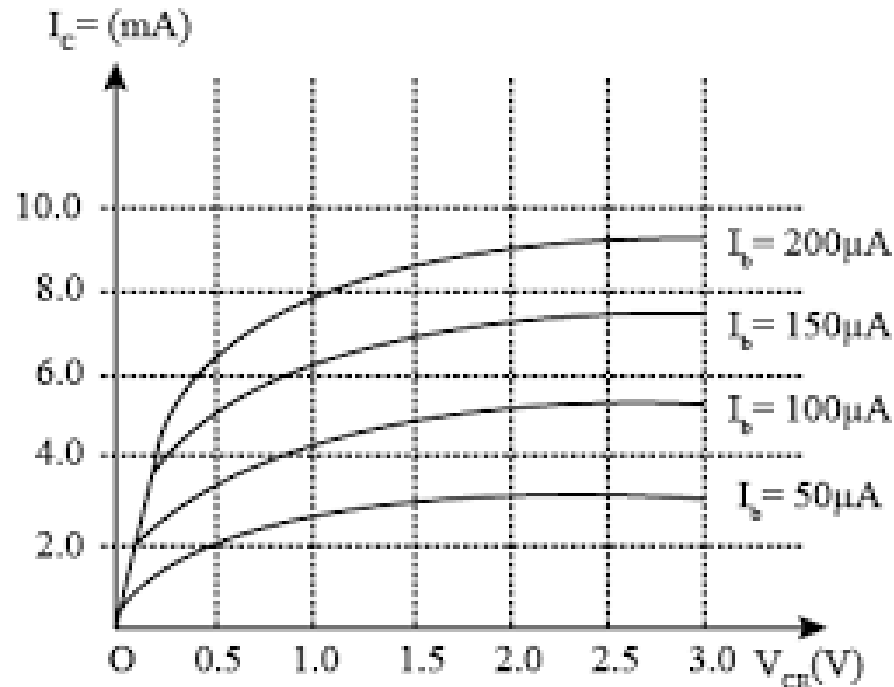
Google form to collect Data

Sample Responses (Excel format)

Sample Responses Graphical form

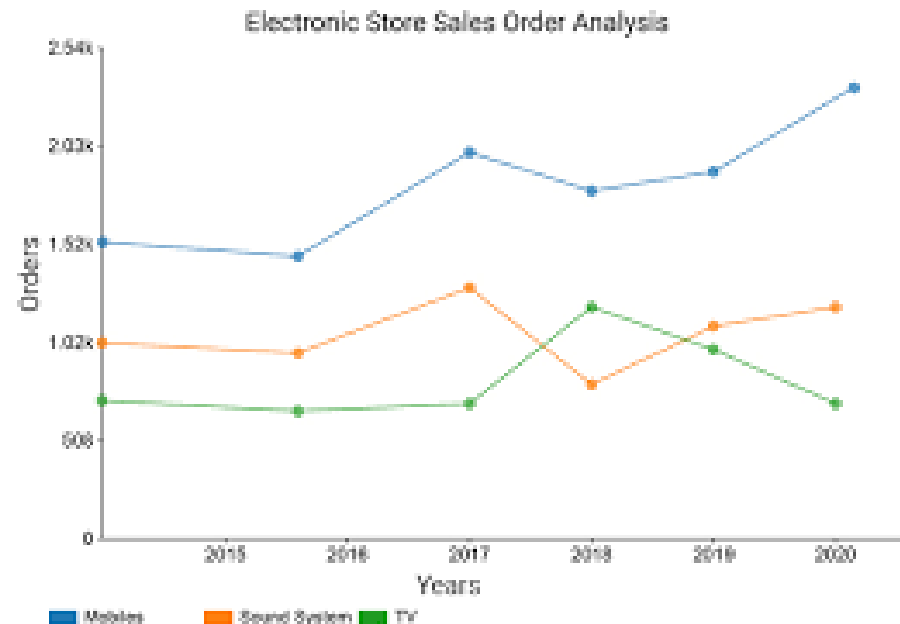
# Need of Data Visualization

- Businesses can now recognize patterns more quickly because they can interpret data in graphical or pictorial forms.
- **Correlations in Relationships:** Without data visualization, it is challenging to identify the correlations between the relationship of independent variables. Make better business decisions by making sense of those independent variables.



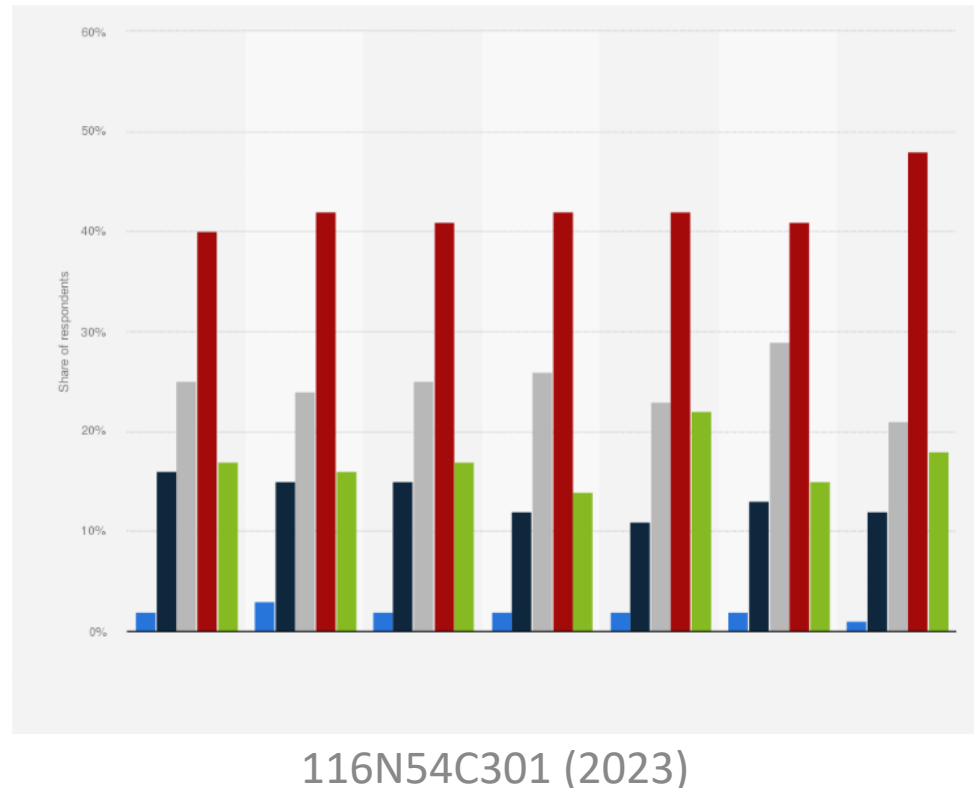
# Need of Data Visualization

- Businesses can now recognize patterns more quickly because they can interpret data in graphical or pictorial forms.
- **Trends Over Time:** an obvious and one of the most valuable applications use of data visualization It's impossible to make predictions without having the necessary information from the past and present. Trends over time tell where we were and where we can potentially go.



# Need of Data Visualization

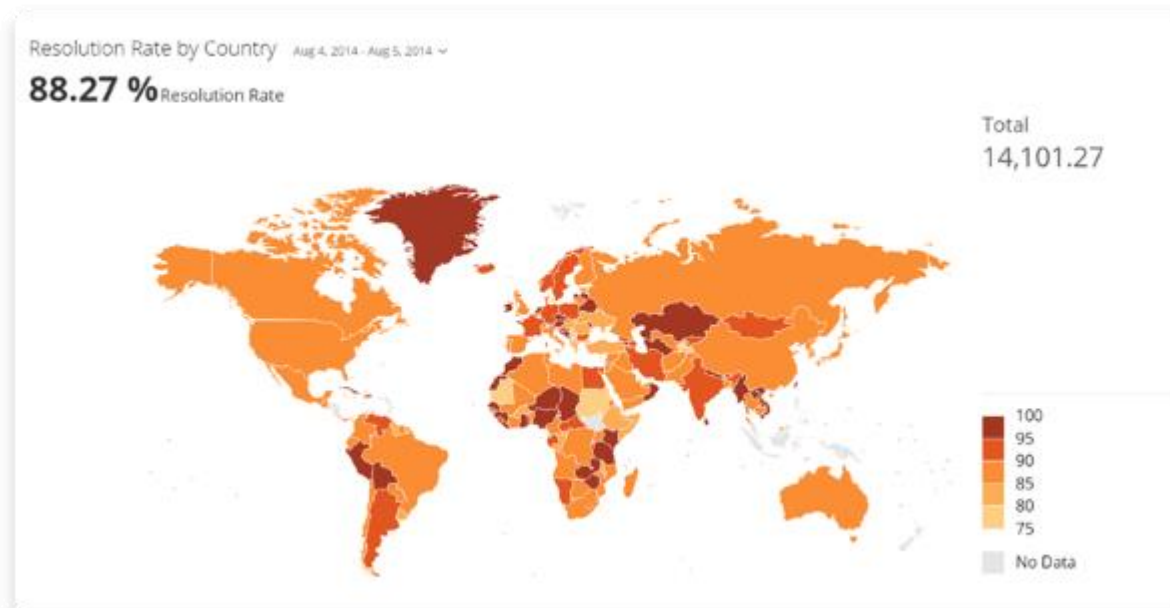
- Businesses can now recognize patterns more quickly because they can interpret data in graphical or pictorial forms.
- **Frequency:** Closely related to trends over time is frequency. By examining the rate, or how often, customers purchase and when they buy gives a better feel for how potential new customers might act and react to different marketing and customer acquisition strategies.





# Need of Data Visualization

- Businesses can now recognize patterns more quickly because they can interpret data in graphical or pictorial forms.
- **Examining the Market:** Data visualization takes the information from different markets to give you insights into which audiences to focus your attention on and which ones to stay away from. We get a clearer picture of the opportunities within those markets by displaying this data on various charts and graphs.





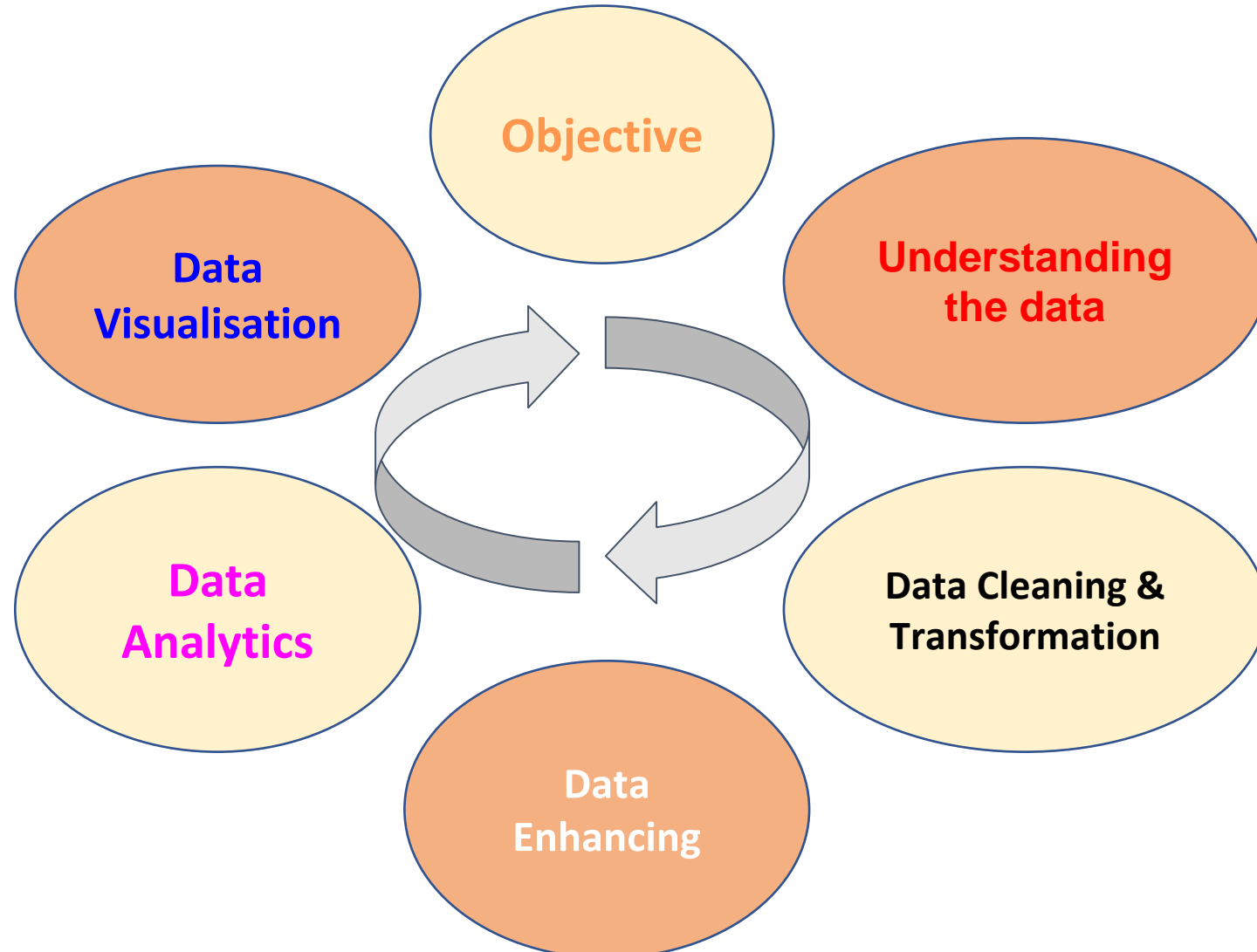
# Need of Data Visualization

- Businesses can now recognize patterns more quickly because they can interpret data in graphical or pictorial forms.
- **Correlations in Relationships:**
- **Trends Over Time:**
- **Frequency:**
- **Examining the Market:**

# Need of Data Visualization

- **Risk and Reward:** Looking at value and risk metrics requires expertise because, without data visualization, we must interpret complicated spreadsheets and numbers. Once information is visualized, we can then pinpoint areas that may or may not require action.
- **Reacting to the Market:** The ability to obtain information quickly and easily with data displayed clearly on a functional dashboard allows businesses to act and respond to findings swiftly and helps to avoid making mistakes.
- Data visualization helps to tell stories by curating data into a form easier to understand, highlighting the trends and outliers
- A good visualization tells a story, removing the noise from data and highlighting the useful information
- Effective data visualization is a delicate balancing act between form and function
- The most stunning visualization could utterly fail at conveying the right message or it could speak volumes
- The data and the visuals need to work together

# Data Analysis Lifecycle



# Data Analysis Lifecycle

## 1. Objective

- Identify **WHY** is the data collected.
- The purpose of data collection decides **WHAT** to collect and **HOW** and from **WHOM** and to **WHOM** will that be presented

## 2. Understanding the Data

- Know **WHAT** is collected

## 3. Data Cleaning & Data Transformation

- Remove the UNWANTED data (noise) which may be irrelevant
- Convert data from one form to another if needed

## 4. Data Enhancing

- Identify the distinguishing features of the data and highlight them

## 5. Data Analytics

- Model the data for presentation
- Interpret the data

## 6. Data Visualisation

- Present the data to the end user in the way (s)he will use to gain

# Data Analytics Lifecycle

## 1. Data Discovery and Formation

- Identify data's purpose and how to achieve it
- Consists of mapping out the potential use and requirement of data,
  - Source of information; Where the information is coming from
  - What story you want your data to convey
  - How your organization benefits from the incoming data
  - **Focus on enterprise requirements related to data, rather than data itself**
  - **Assessing the tools and systems that are necessary to read, organize, and process all the incoming data.**
  - The data science team learn and investigate the problem.
  - Develop context and understanding.
  - Come to know about data sources needed and available for the project.
  - The team formulates initial hypothesis that can be later tested with data

# Data Analytics Lifecycle

## 1. Data Discovery and Formation

## 2. Data Preparation

- Steps to explore, preprocess, and condition data prior to modeling and analysis.
- Execute, load, and transform the data
- Tasks could be performed multiple times and not in predefined order
- Several tools commonly used for this phase are – Hadoop, Alpine Miner, Open Refine, etc.

## 3. Model Planning

- Data is explored to learn about relationships between variables and subsequently, selects key variables and the most suitable models.
- Data science team develop data sets for training, testing, and production purposes
- Team builds and executes models based on the work done in the model planning phase.
- Several tools commonly used for this phase are – Matlab, STASTICA.

# Data Analytics Lifecycle

## 1. Data Discovery and Formation

## 2. Data Preparation

## 3. Model Planning

## 4. Model Building

- Team develops datasets for testing, training, and production purposes.
- Team also considers whether its existing tools will suffice for running the models or if they need more robust environment for executing models.
- Free or open-source tools – R and PL/R, Octave, WEKA.
- Commercial tools – Matlab , STASTICA.

## 5. Communication of Results

- After executing model team need to compare outcomes of modeling to criteria established for success and failure.
- Team considers how best to articulate findings and outcomes to various team members and stakeholders, taking into account warning, assumptions.
- Team should identify key findings, quantify business value, and develop narrative to summarize and convey findings to stakeholders.

# Data Analytics Lifecycle

## 1. Data Discovery and Formation

## 2. Data Preparation

## 3. Model Planning

## 4. Model Building

## 5. Communication of Results

## 6. Optimization

- The team communicates benefits of project more broadly and sets up pilot project to deploy work in controlled way before broadening the work to full enterprise of users.
- This approach enables team to learn about performance and related constraints of the model in production environment on small scale , and make adjustments before full deployment.
- The team delivers final reports, briefings, codes.
- Free or open source tools – Octave, WEKA, SQL, MADlib.



# Visual Revolution, Various types of visualization with its best practices.

**A 'Visual Revolution' is changing the way we communicate**

- Use all kind of visuals - images, graphics, videos and animations to create emotional connections
- Make sure your content is unique so that people will remember for longer duration, creating impact
- Create visuals that are useful to customers, or other supporters
- Invite other people to share their images to encourage interaction and engagement
- Include short captions with your visuals to give context or include a call to action that asks people to take the next step

# Visualization best practices

**With explosion in volume and complexity of data generated, use of data visualization becomes essential**

- Human processes visual information better than tables, numbers, and text
- Converting into graphs and charts data is transformed into easy-to-understand form

# From Visualization to Visual Data Storytelling: An Evolution

Data visualization will depend on statistical understanding of the data, its context and to measure it to avoid false analysis and skewed decision making

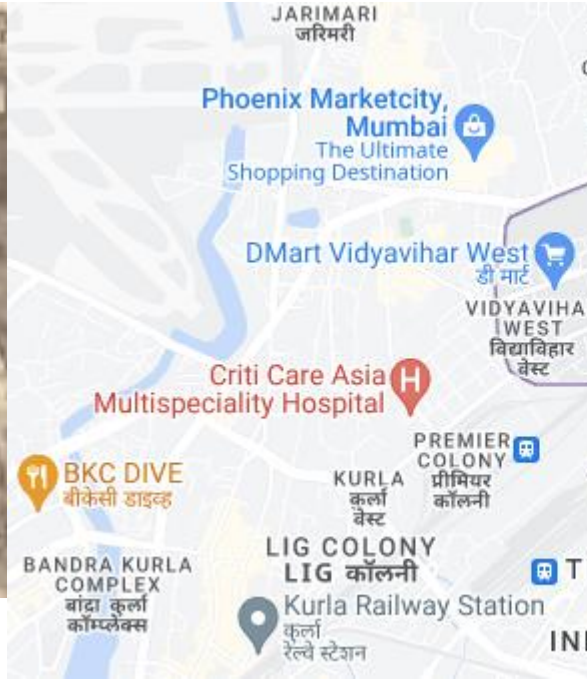
Human visual cognition system demands a way to encode numbers with meaning by relying on colours and shapes to support interpretation

Data visualization is the best way to see and understand huge, most diverse data

# From Visualization to Visual Data Storytelling An Evolution

## Data visualization

- Cave drawings      Early Maps      Modern Maps      Statistical graphs



# From Visualization to Visual Data Storytelling

## An Evolution

Data visualization aided by:

- Advancement in visual design
- Cognitive science
- Technology and
- Business intelligence

An emerging new approach to storytelling attempts to combine data with graphics and tell the world's stories through the power of information visualization.

# From Visualization to Visual Data Storytelling: An Evolution

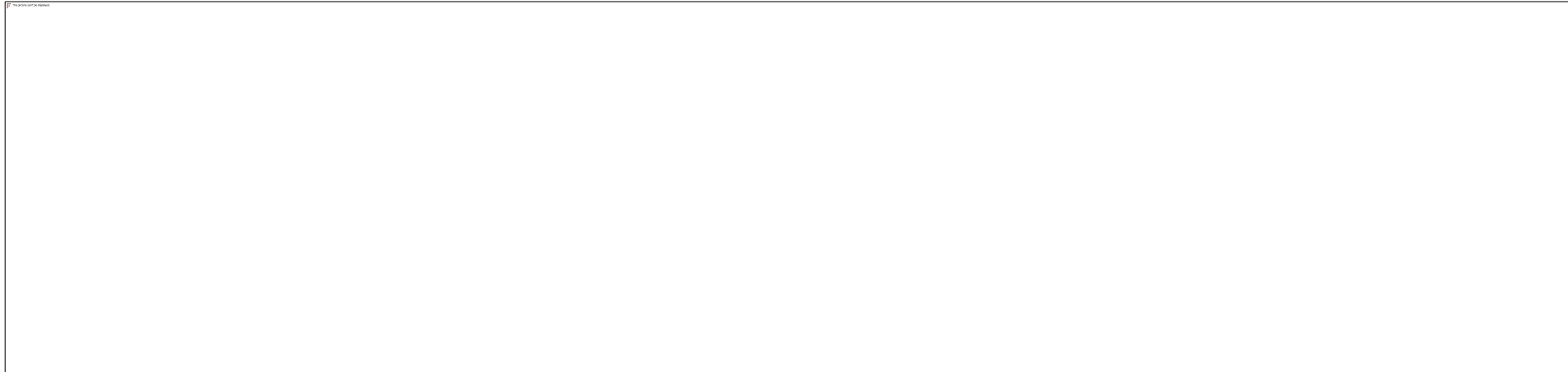
Evidence of the cognitive effects of storytelling in neurology is it's a central way that we learn, remember, and communicate information.

Important goal of visual story is to prepare business decision makers to leave a data presentation with a story in their head that helps them both to **remember** your message and **take action** on it.

Data stories provide tool to communicate information in ways the facts and figures alone can't.

# From Visualization to Visual Data Storytelling: An Evolution

## Storytelling Process



# From Visual to Story: Bridging the Gap

- Data Storytelling is needed as amount of data is growing day-by-day
- Steps
  1. Create Data
  2. Collect Data
  3. Clean/ Prepare Data
  4. Present Data (Visualization)
  5. Storytelling (Communication)



# Data Visualization V/s Data Storytelling

- **Data visualization:** the practice of graphically representing data to help people see and understand patterns, insights, and other discoveries hidden inside information
- **Data storytelling:** translate seeing into meaning by weaving a narrative around the data to answer questions and support decision making

# Digital Storytelling

- **Process of using Digital Technology to communicate to aid taking decision making**

## **Process of Storytelling:**

Use of media

Text,

Visuals

Music

Video

Voice

**To reach to wider audience in more convincing way**

# Digital Storytelling

## Data:

How to prepare pasta :

Ingredients

Process/ Procedure



## Visualisation:

Use images to  
express



## Presentation:

Use presentation to create long lasting impression

# Data Fundamentals, Collecting data, Preparing Data

## Data Fundamental:

### Primary Data:

- Collected afresh
- Need to validate the process and collected information

### Secondary Data:

- Collected by some one
- Validity/ relevance may be checked

# Data Fundamentals, Collecting data, Preparing Data

## Collecting data:

### Primary Data:

- Survey
- Descriptive research
  - Observation: systematic viewing
  - Direct Communication
  - Personal Interviews

### Secondary Data:

- Download from reliable source

# Data Fundamentals, Collecting data, Preparing Data

**Collecting data:**

**Primary Data:**

## **Advantages**

- Unbiased data
- Fresh Information
- Relevant information could be captured

## **Limitations:**

- Time consuming
- Costly
- Limited scope

# Data Fundamentals, Collecting data, Preparing Data

## Collecting Primary data:

- Types of observation:
  - Structured:
    - Standard conditions
    - Well prepared
    - Forms/ formats
    - Checklist
  - Unstructured:
- Without much preparation

# Introduction data visualization

- **Introduction to basic Visualization and it's need.**
  - **The Bar Chart**
  - **The Line Chart**
  - **The Pie and Donut Charts**
  - **The Scatter Plot**
  - **The Packed Bubble Chart**



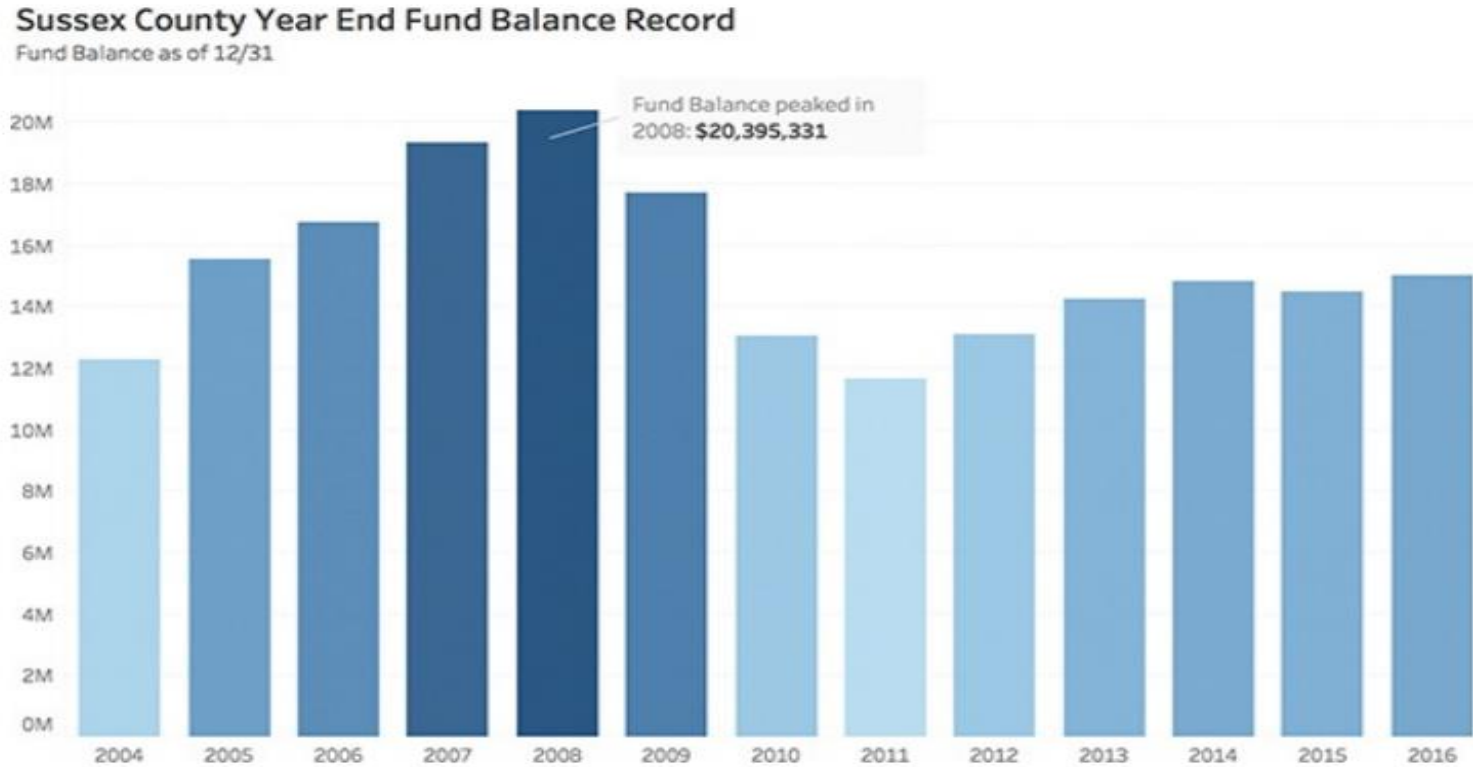
# Charts

- The Bar Chart
- The Line Chart
- The Pie and Donut Charts
- The Scatter Plot
- The Packed Bubble Chart

# Charts: The Bar Chart

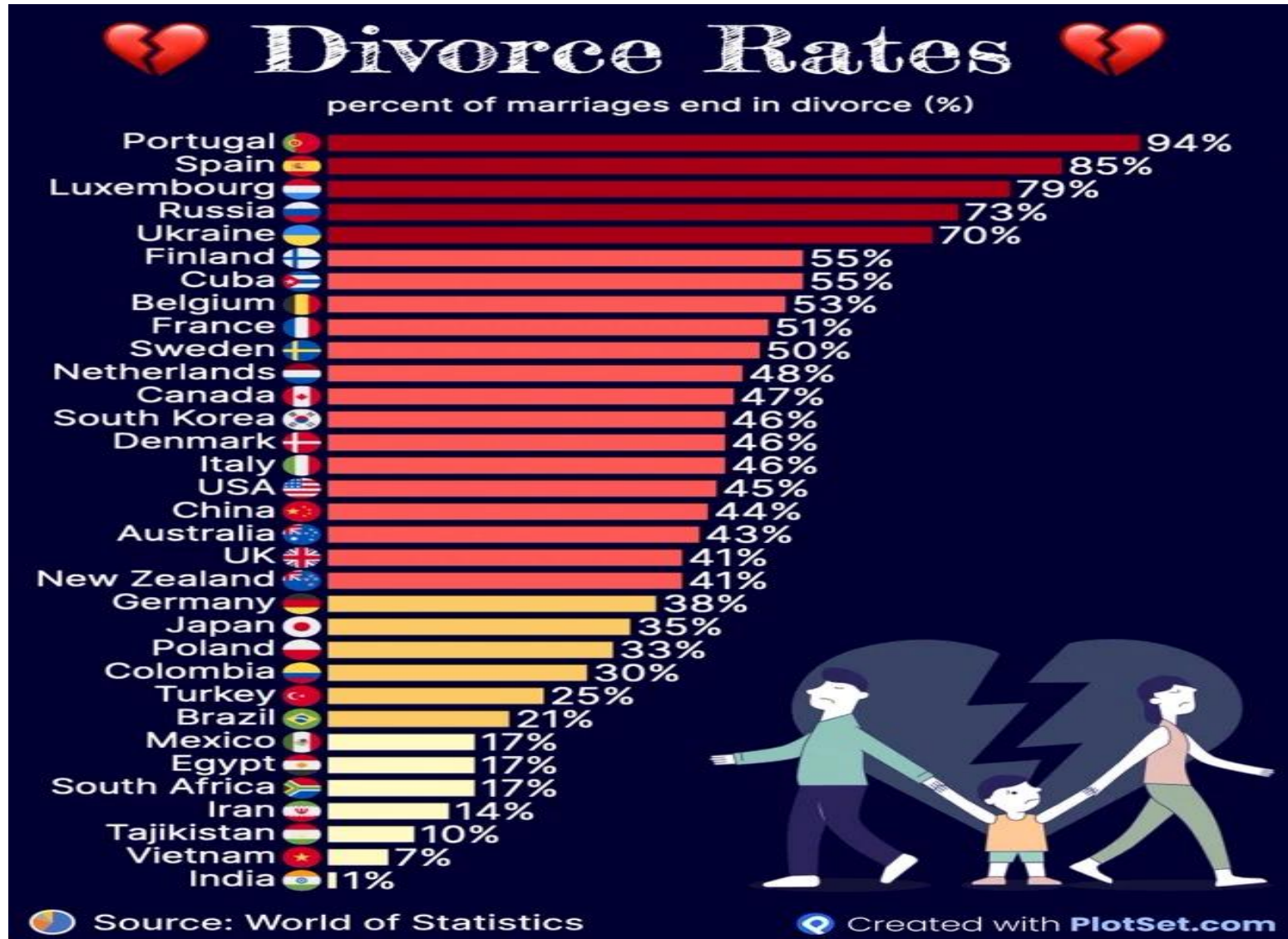
- The bar chart is one of the most common ways to visualize data.
- **Used to compare numerical values using the length of the bar.**
- It is best suited for numerical data that can be divided into distinct categories to compare information and reveal trends at a glance.
- Bars can be oriented on the vertical or horizontal axis, which can be helpful for spotting trends.
- Additional layers of information can be added using clustered bars or by stacking related data.
- Color can be added for more impact or to overlay for immediate insight.
- Trend lines and other annotations can be added to highlight important data points.

# Charts

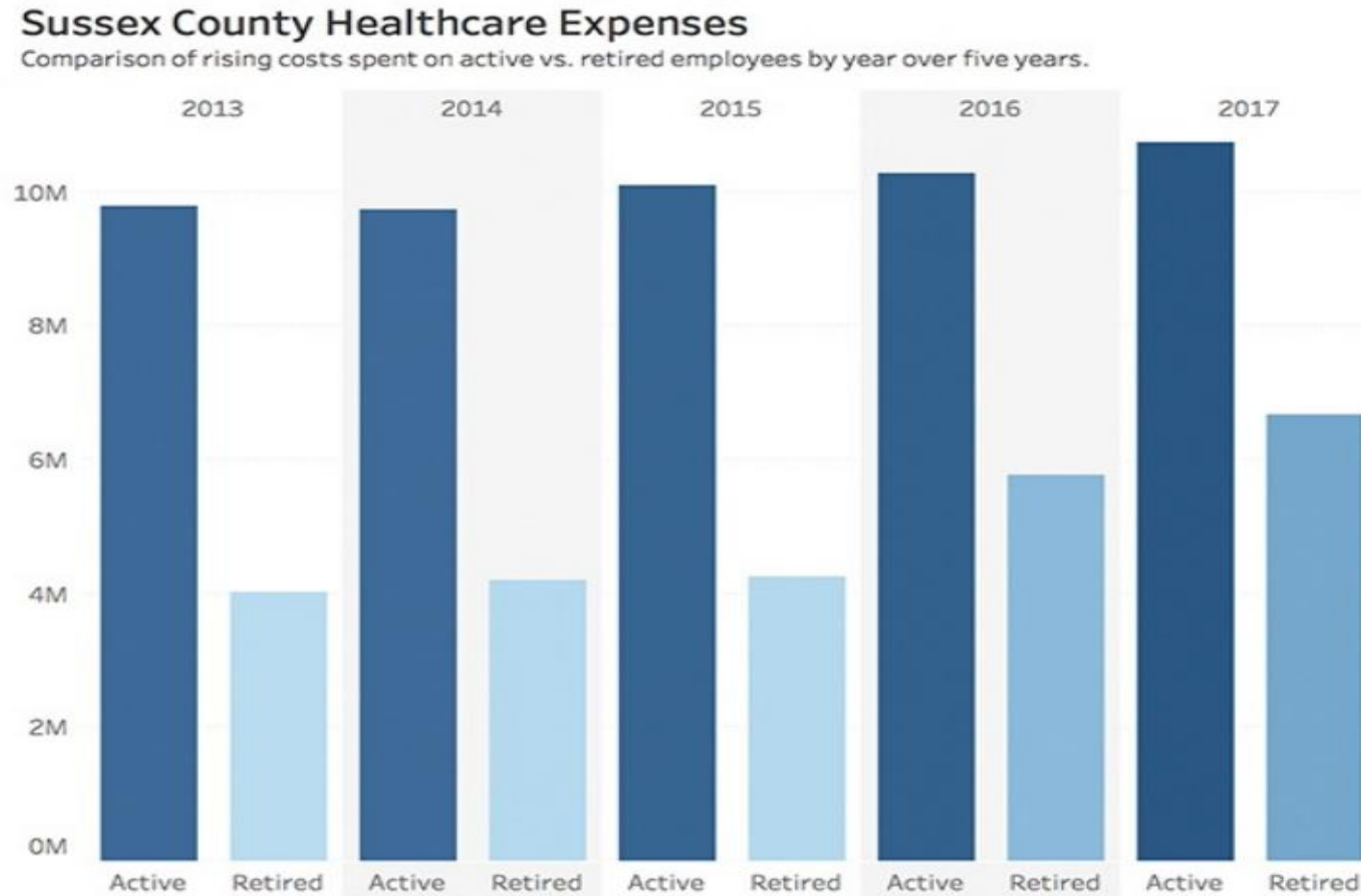


**Figure 5.1** This simple, classic bar chart with color gradient shading and a point annotation compares the year end balance for Sussex County, NJ over a period of 13 years.

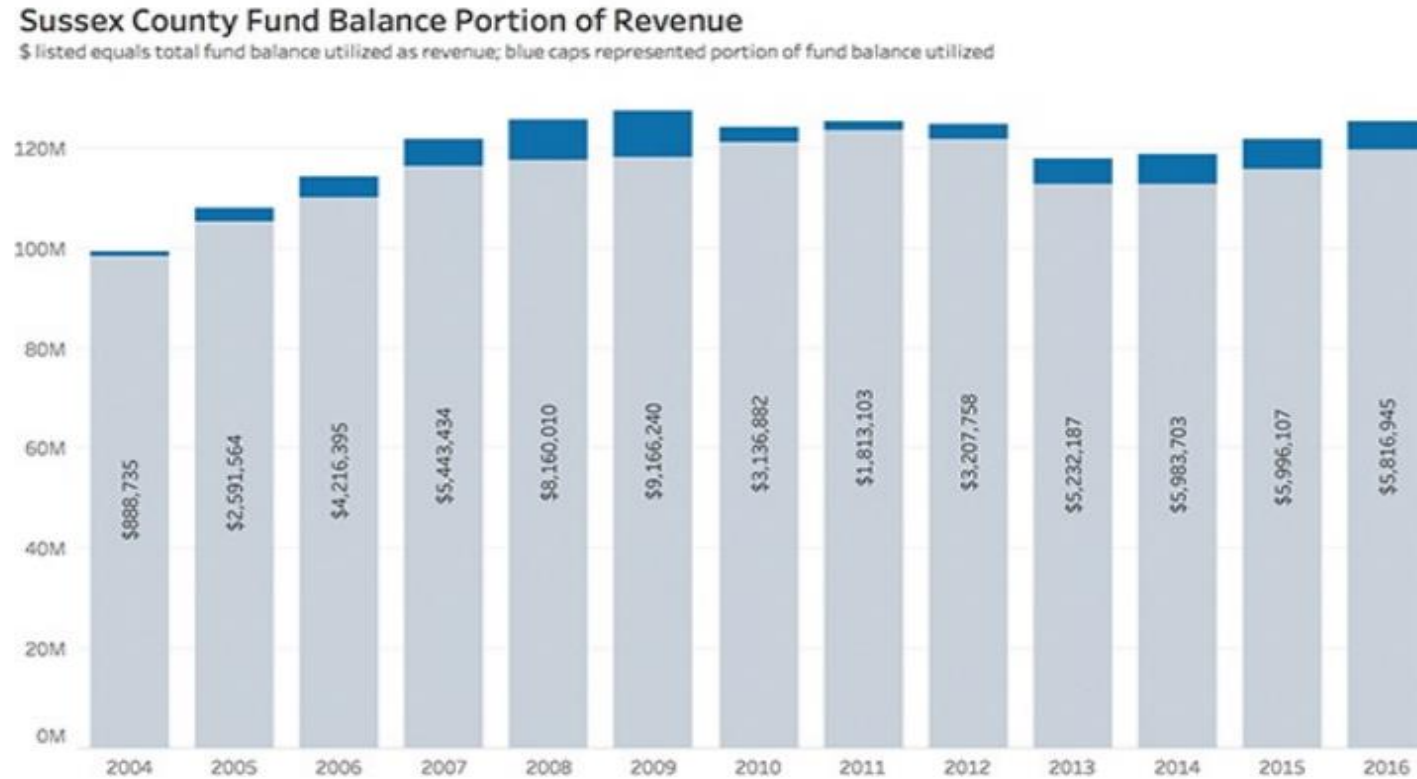
# Horizontal Bars



# Comparison using bars



# Stacked bar chart



**Figure 5.2** Alternative bar charts: a side-by-side bar chart with color gradient shading and a stacked bar chart with labeled and banded columns.

# Charts

## The Bar Chart

- **Advantages:**
- show each data category in a frequency distribution.
- display relative numbers or proportions of multiple categories.
- summarize a large data set in visual form.
- clarify trends better than do tables.
- estimate key values at a glance.
- permit a visual check of the accuracy and reasonableness of calculations.

# Charts

## The Bar Chart

- **Limitations**
- Requires additional written or verbal explanation
- By cleverly choosing the scale of the bar graph some scammers can make sure that one of the bars seems higher relative to the other
- Fails to reveal assumptions, causes, and effects
- Not Suitable if there are large number of categories



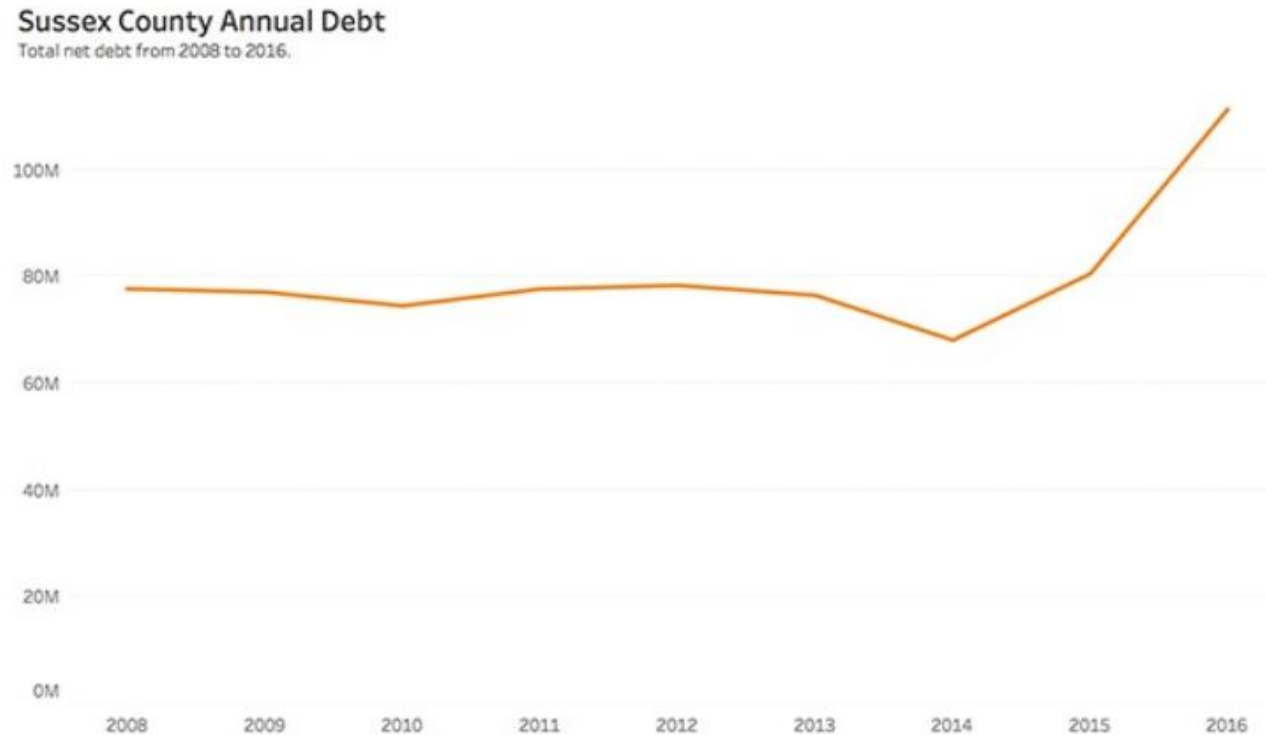
# Charts

## The Line Chart

- A line chart is a type of chart that displays information as a series of data points connected by straight line segments.
- They are most commonly used when an element of time is present.

# The Line Chart

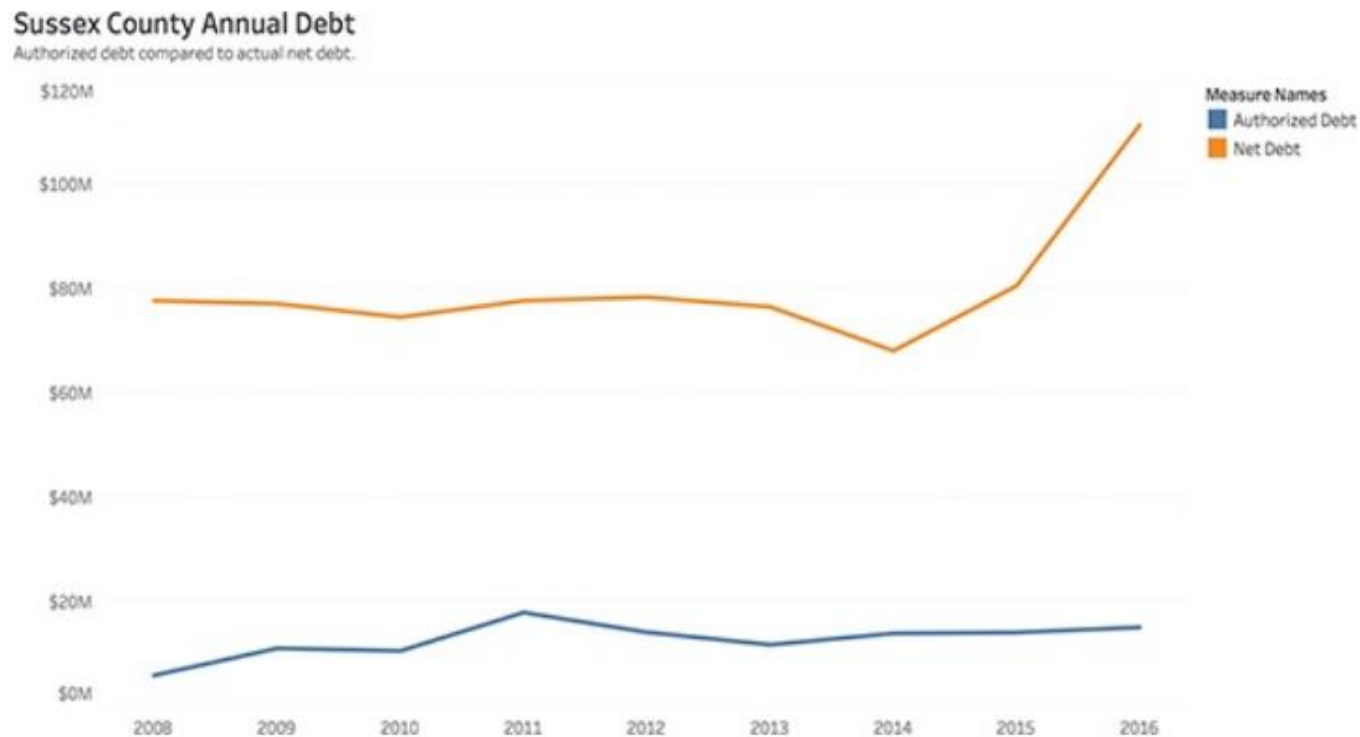
- The best use case for line charts involves displaying trends over a period of time , when your data are ordered, or when interpolation makes sense.



**Figure 5.5** This line chart shows the audited annual net debt for Sussex County over a period of nearly ten years.

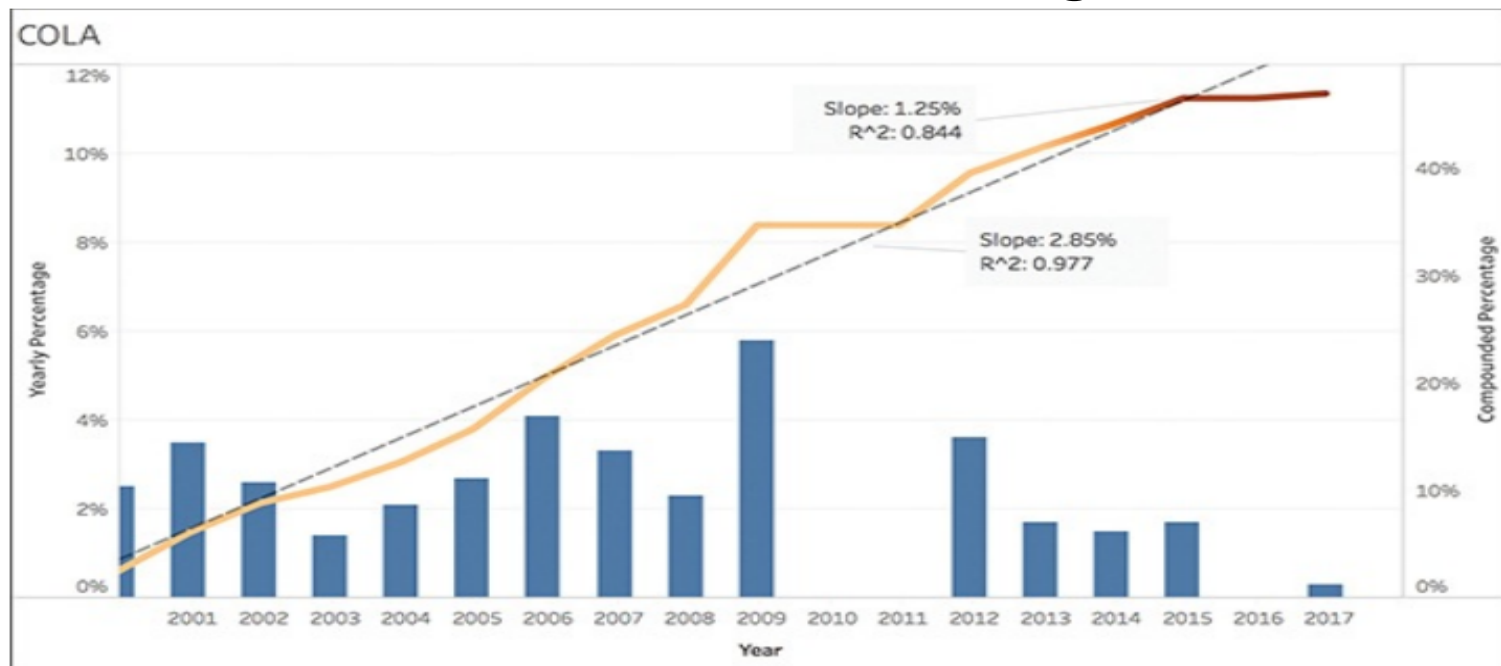
# The Line Chart

- Dual-axis line charts can be created by bringing two measures to the rows shelf, and then right-clicking on the second measure and selecting Dual-axis from the drop-down menu.



# The Line Chart

- Additionally, when two or more lines are present, you can transform line charts by adding additional chart types to deepen insight. For example, a line chart can be combined with a bar chart to provide visual cues for further investigation.



**Figure 5.7** Adjust the Marks card to help you combine chart types. This work-in-progress line chart has been combined with a bar chart. It also includes annotations, trend lines, and a color gradient shade element on the line to enhance insight.

# Charts

## The Line Chart

### Advantages:

- Easy to show the data changes over time over a line graph.
- Helps in showing small changes that are difficult to measure in other graphs.
- A relationship between 2 or more variables get identified.
- It presents a good impression of trends and changes.
- Both negative, as well as positive values, are indicated.

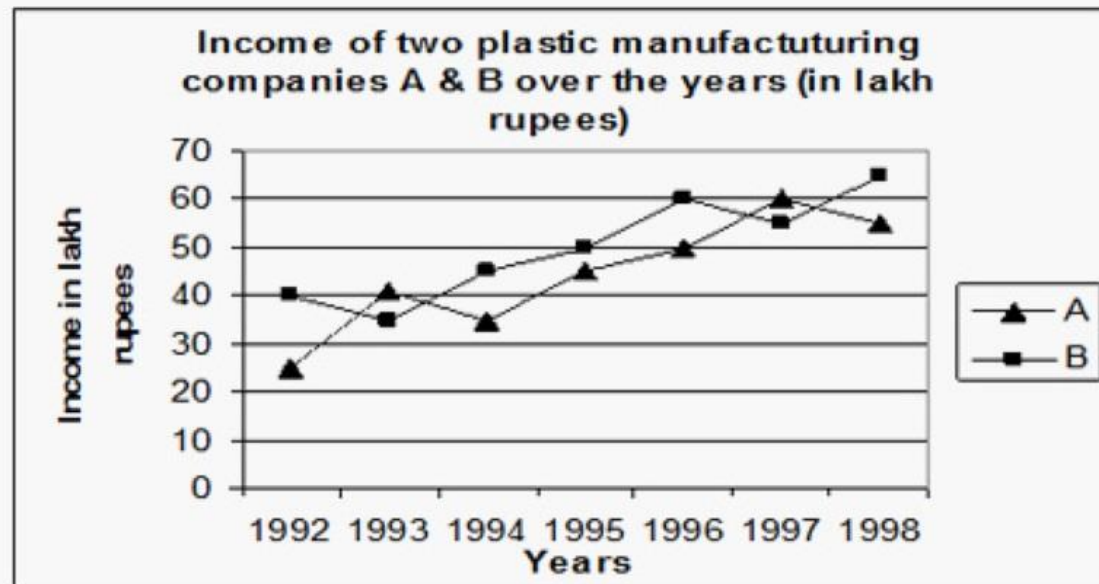
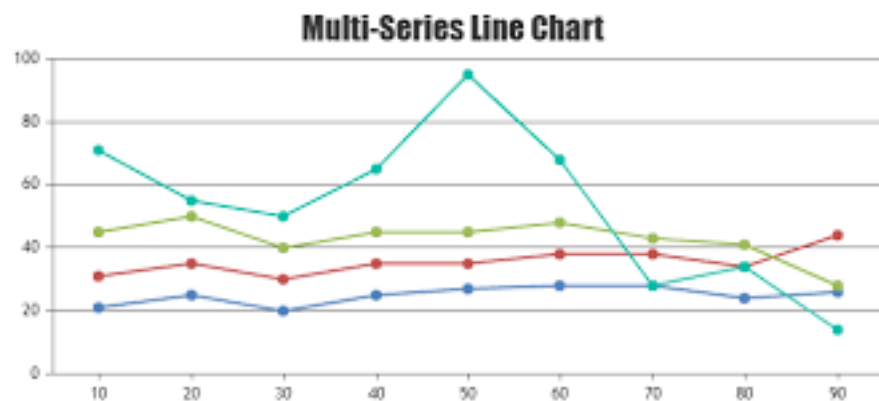
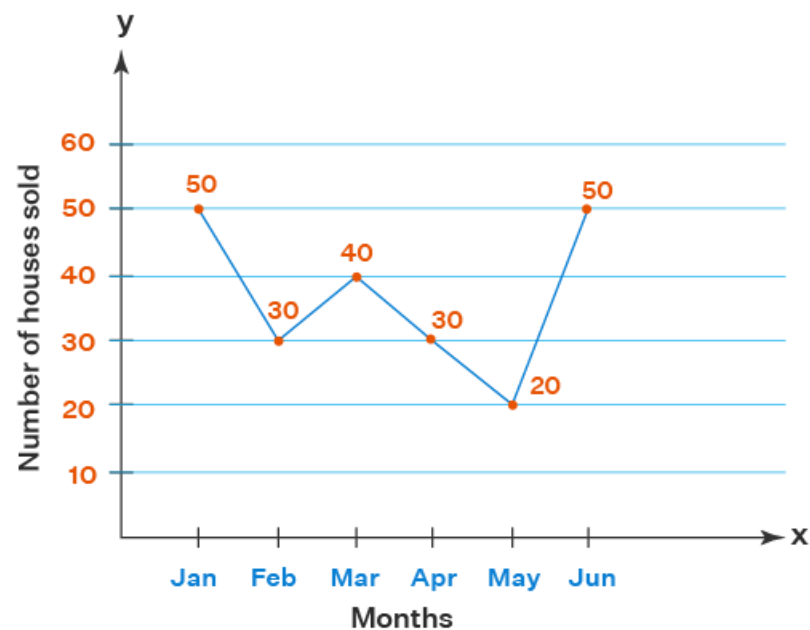
# Charts

## The Line Chart

### Limitations:

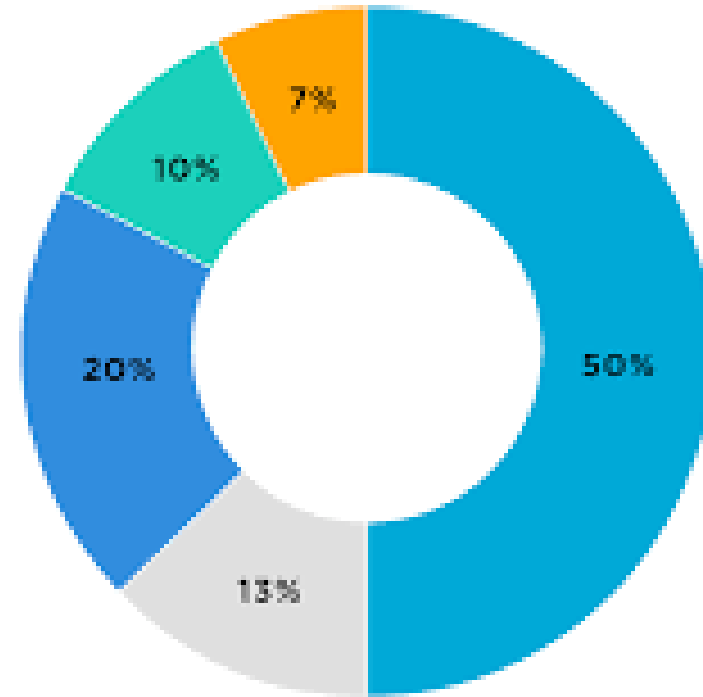
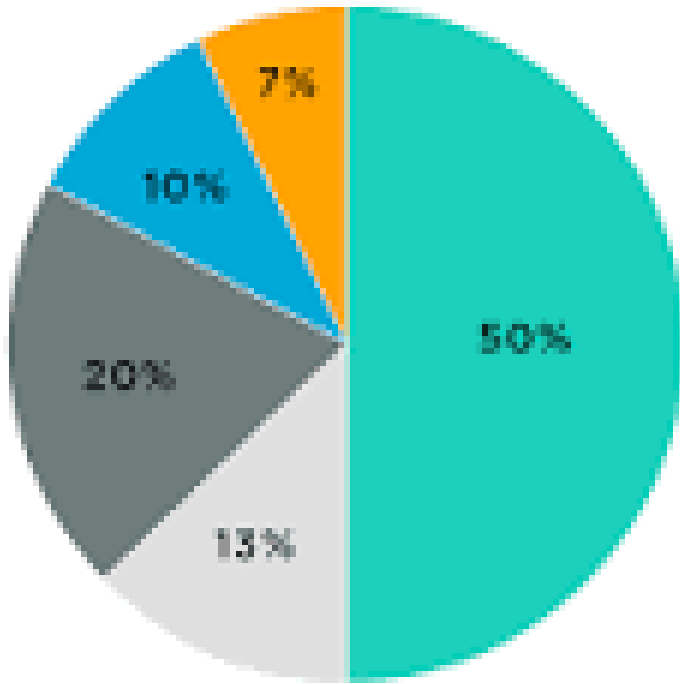
- Plotting too many lines over the graph makes it cluttered and confusing to read.
- A wide range of data is challenging to plot over a line graph.
- They are only ideal for representing data made of total figures such as values of total rainfall in a month.

## Line Chart



# The Pie and Donut Charts:

- A pie chart (or a circle chart) is a circular statistical graphic, which is divided into slices to illustrate numerical proportion.
- Donut charts can help clarify your data story by including a key takeaway in the center white





# Charts

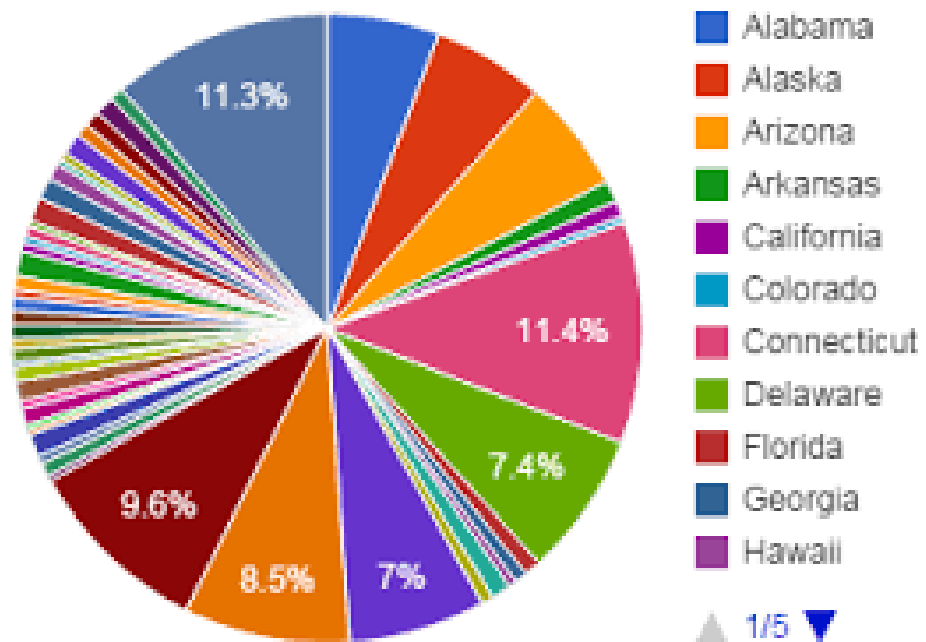
## The Pie and Donut Charts:

- **Advantages :**
- A simple and easy-to-understand picture.
- It represents data visually as a fractional part of a whole, which can be an effective communication tool for the even uninformed audience.
- It enables the audience to see a data comparison at a glance to make an immediate analysis or to understand information quickly.
- The need for readers to examine or measure underlying numbers themselves can be removed by using this chart.
- To emphasize points you want to make, you can manipulate pieces of data in the pie chart.

# Charts

## The Pie and Donut Charts:

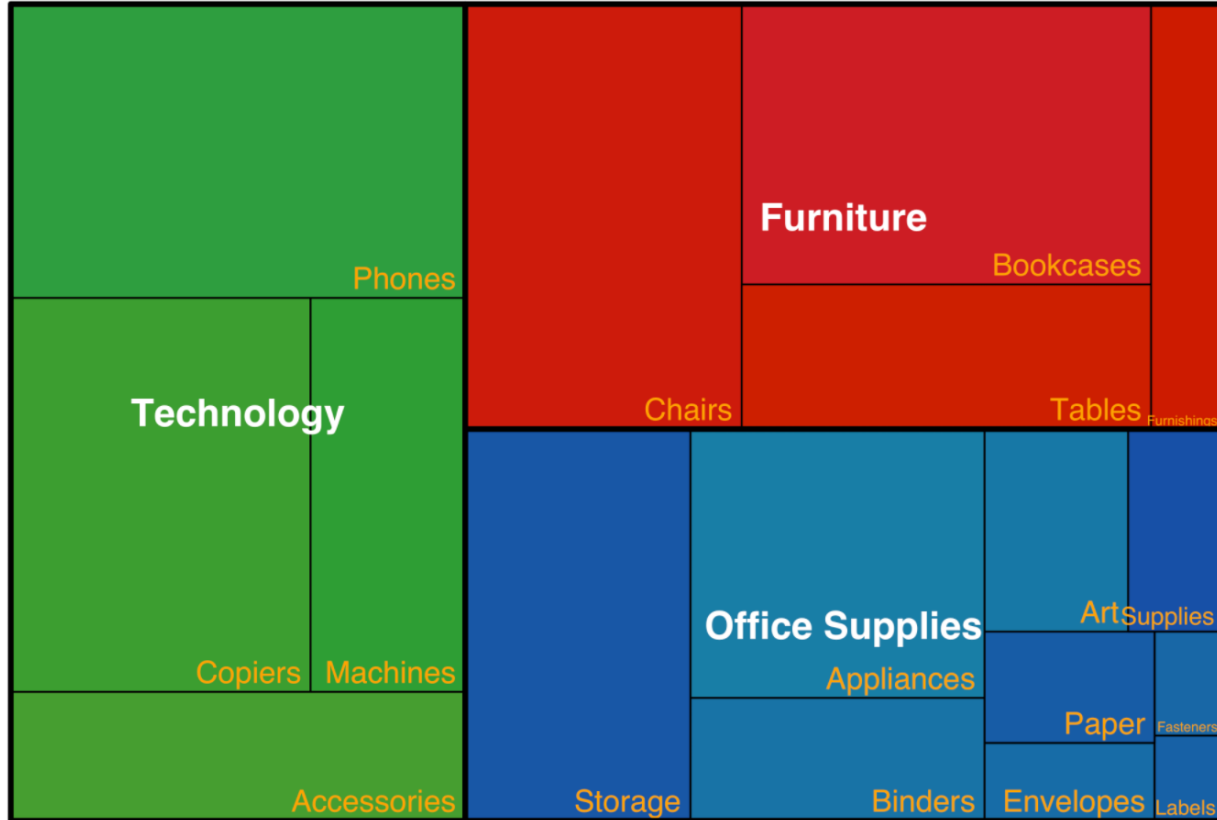
- **Limitations:**
- They cannot show more than a few values, because as the number of values shown increases, the size of each segment/slice becomes smaller. This makes them unsuitable for large amounts of data.
- They take up more space than their alternatives, like a 100% Stacked Bar Chart for example. Mainly due to their size and for the usual need for a legend.
- They are not great for making accurate comparisons between groups of Pie Charts. This being that it is harder to distinguish the size of items via area when it is for length.



TEST MATCHES / BATTING AVERAGES (CURRENT PLAYERS) - 2019



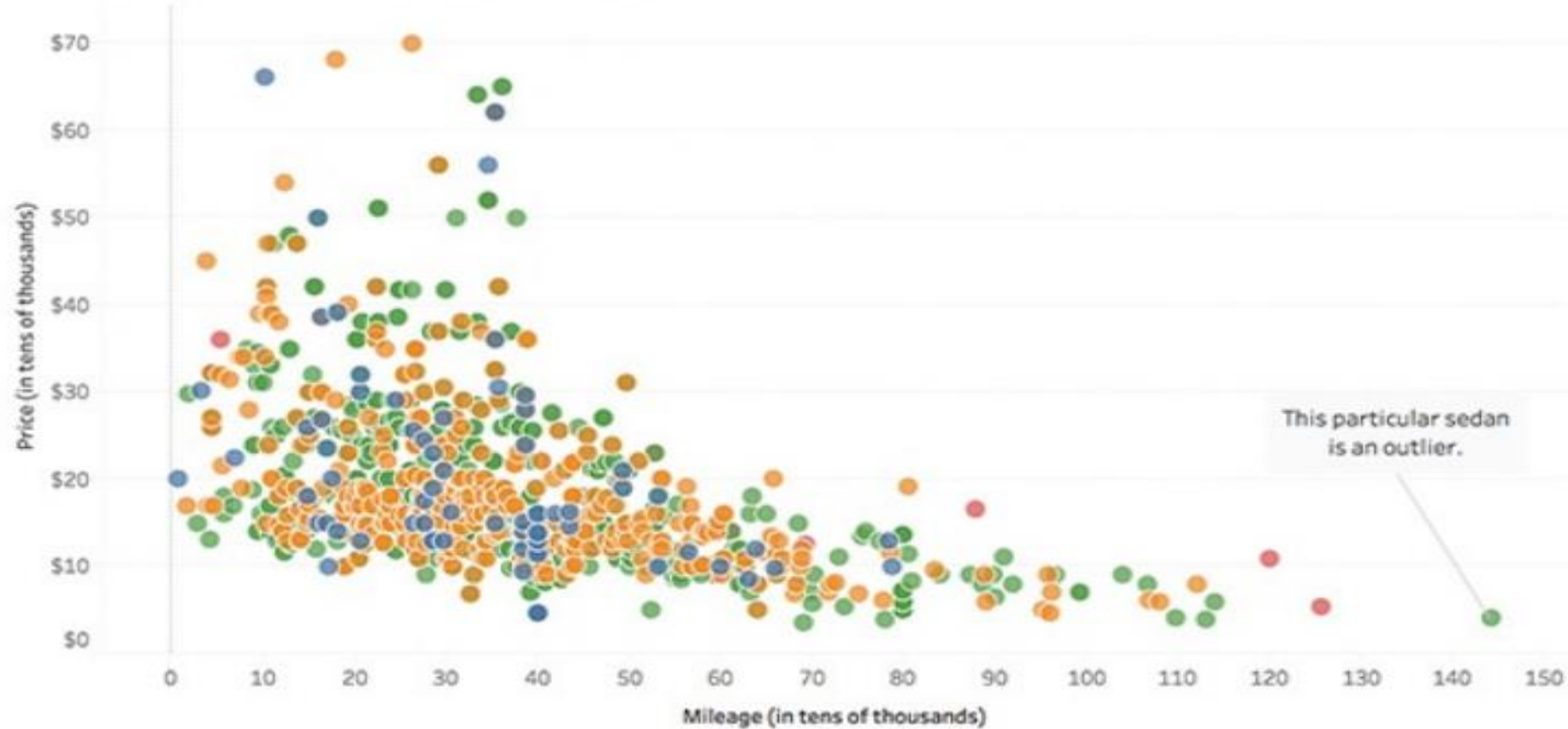
# Treemap an alternative to pie



# The Scatter Plot

What Effect Does Mileage of a Used Car Have On Price?

A regression scatterplot of used coupes, SUVs, sedans, and trucks.



**Figure 5.14** Scatter plot example.

# Charts

## The Scatter Plot

- Scatter plots are an effective way to visualize numerical variables to compare measures and quickly identify patterns, trends, concentrations (clusters), and outliers.
- These charts can give viewers a sense of where to focus discovery efforts further and are best used to investigate relationships between variables.
- Scatter plots are particularly useful when exploring statistical relationships such as linear regression.
- Use a scatter plot to determine whether or not two variables have a relationship or correlation.

# Charts

## The Scatter Plot

### Advantages:

- It shows the relationship between two variables.
- It is the best method to show you a non-linear pattern.
- The range of data flow, like the maximum and minimum value, can be determined.
- Patterns are easy to observe.
- Plotting the diagram is simple.

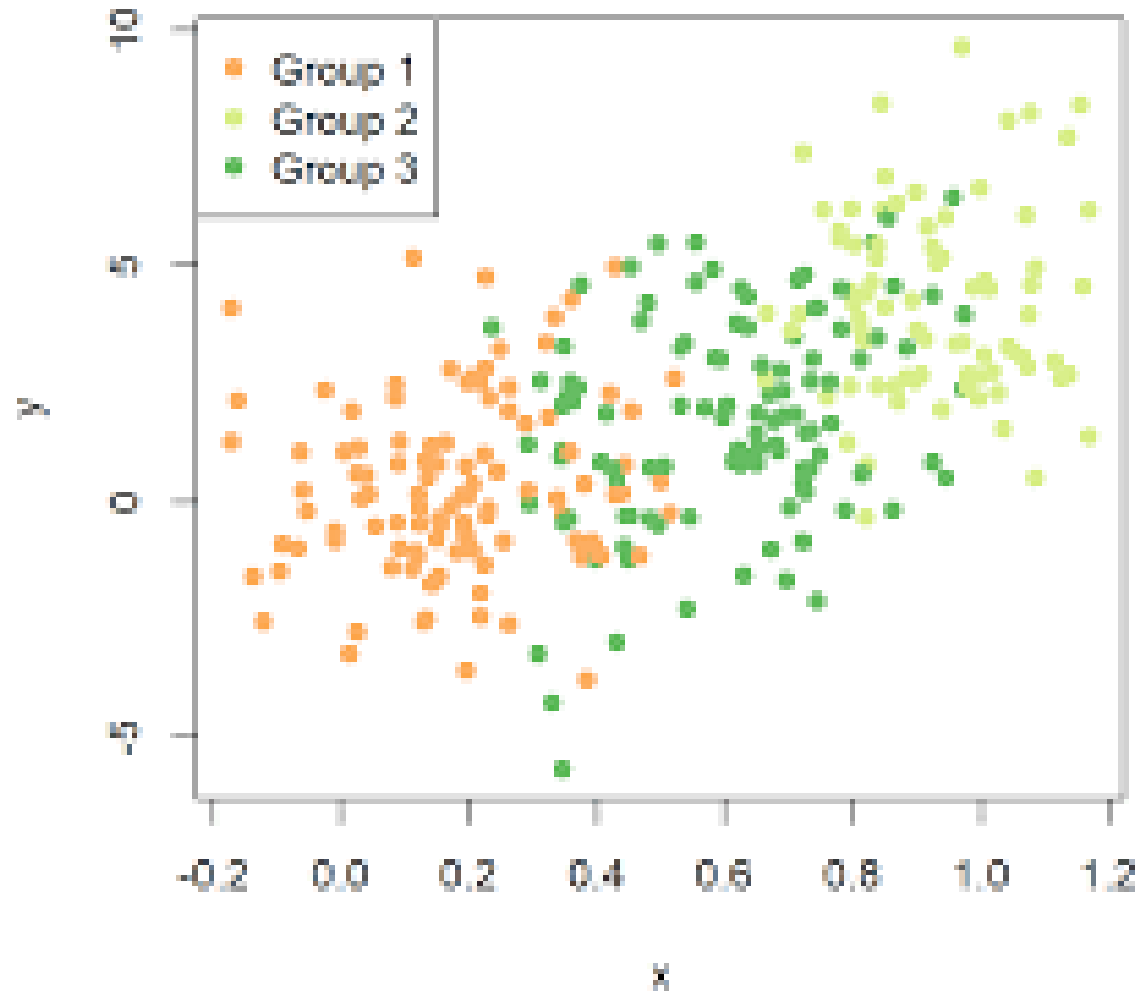
# Charts

## The Scatter Plot

### Limitations:

- Scatter plot does not show the relationship for more than two variables.
- Scatter plots are unable to give the exact extent of correlation.
- Scatter plot doesn't show the quantitative measure of the relationship between the two variables.





# The Packed Bubble Chart

- Packed bubble charts are **visualizations where the size and optionally the color of the bubbles are used to visualize the data.**

America's Favorite Pie Flavor  
Year over year, apple takes the cake...erm, pie.



**Figure 5.21** A packed bubble chart displays data in a cluster of circles, using size and color to encode the bubbles with

# The Packed Bubble Chart

- The bubble chart is a variation of the scatter plot that replaces data points with a cluster of circles (or bubbles), a technique that further emphasizes data that would be rendered on a pie chart, scatter plot, or map.
- This method shows relational values without regard to axes and is used to display three dimensions of data: two through the bubble's location and another through size.
- These charts allow for the comparison of entities in terms of their relative positions with respect to each numeric axis and size.
- The sizes of the bubbles provide details about the data, and colors can be used as an additional encoding cue to answer many questions about the data at once.

# Charts

## The Packed Bubble Chart

- **Advantages:**
- Explaining a complex data set easily.
- Analyzing data sets with multiple inputs more easily.
- Helping to visualize patterns and uncover trends using data analysis.
- Helping to correlate data as part of a data series.
- Being dynamic enough to analyze finance, sales and marketing scenarios beyond time-series or whole-to-whole comparisons using line and bar charts.
- Effectively displaying a relatively large amount of information.
- Depicting the relationship between three or even four variables that can change over time without having to use 3D graphs.

# Charts

## The Packed Bubble Chart

- **Limitations**
- A bubble chart can become difficult to understand.
- The overlapping of bubbles can make it difficult to distinguish them.
- Users may find it hard to depict zero or negative values.
- It can be hard to ascertain exact values using circle sizes.

# Charts

