Big Data Visualization

Content

- Introduction
- The 3v's of Big Data
- Big Data Life Cycle
- Role of Visualization in Big Data
- Importance of Data Visualization
- Design Principle
- Steps for Interactive Data Visualization
- Visualization Techniques
- Make Visualization more interactive
- Visualization Challenges
- References

Introduction

- Data: Any piece of Information formatted in a special Way

Different Forms of Data …

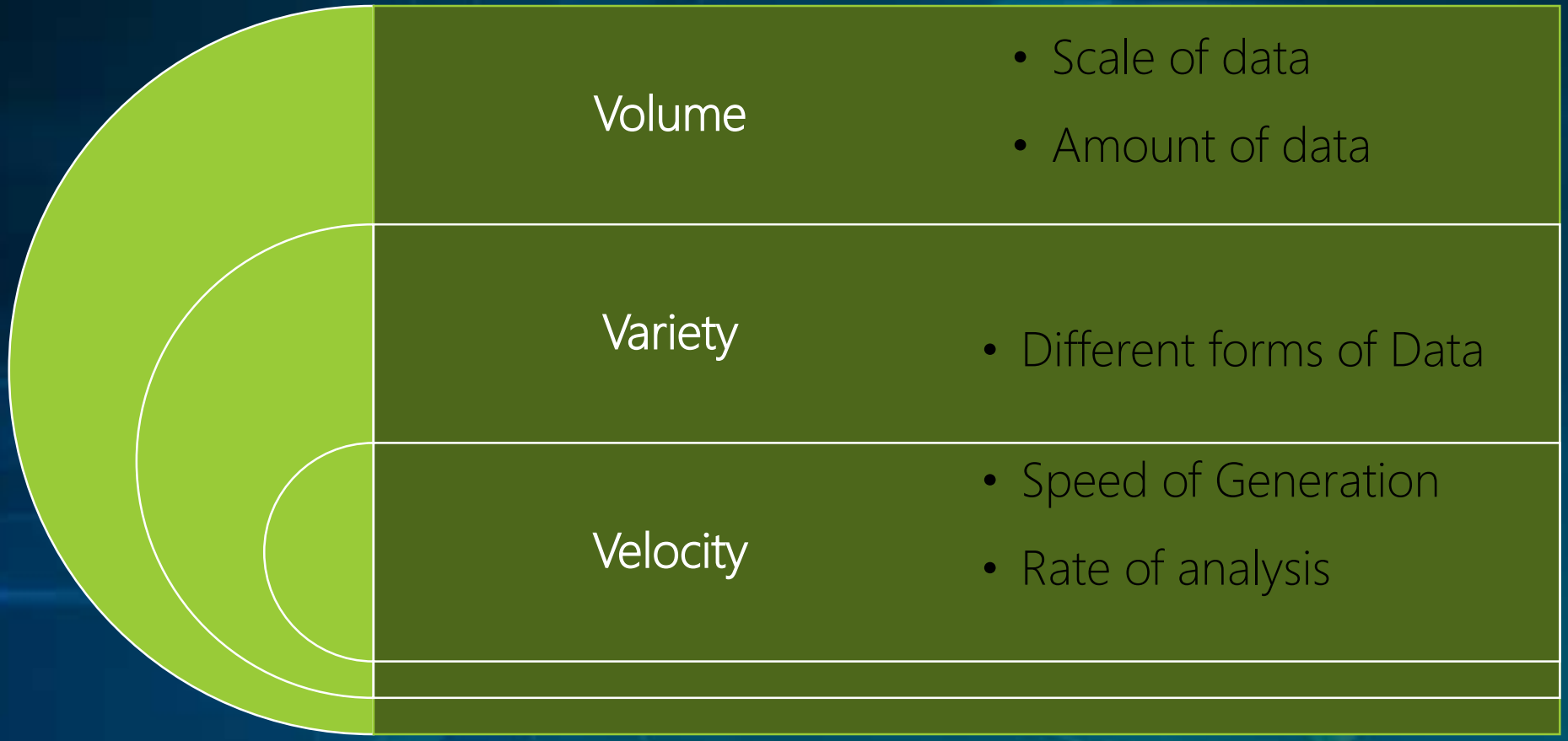
Tabular Data…

Table 1: Salt Concentration and Light Transmittance				
Salt Concentration (%)	Transmittance (%T)			
	Trial #1	Trial #2	Trial #3	Trial #4
0	77.23	74.50	64.88	75.27
3	85.23	92.82	78.91	60.71
6	88.39	100.05	73.66	66.51
9	80.71	100.05	68.29	64.91
12	82.66	117.18	71.01	56.91
15	72.55	115.40	65.72	66.03

Stats…

Natural moisture content	13.4%
Specific gravity	2.40
Liquid limit	45.5%
Plastic limit	31.0%
Plasticity index	14.5%
AASHTO classification	A-7-5
Soil type	Silt-Clay

Three V's of Data



What is Big Data

- "Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" (Gartner 2012)

- Complicated (intelligent) analysis of data may make a small data "appear" to be "big"

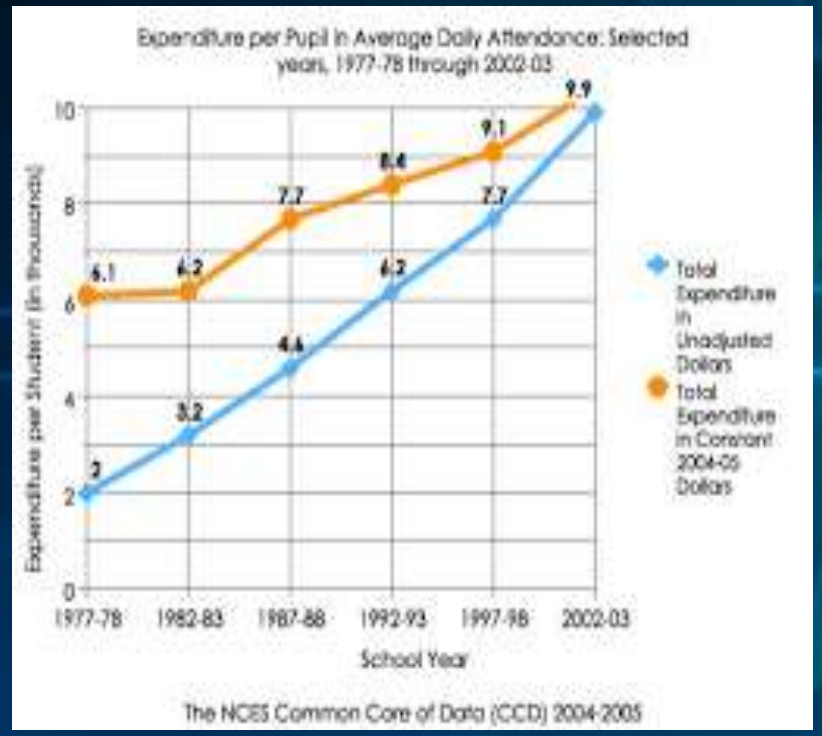
- Any data that exceeds our current capability of processing can be regarded as "big"

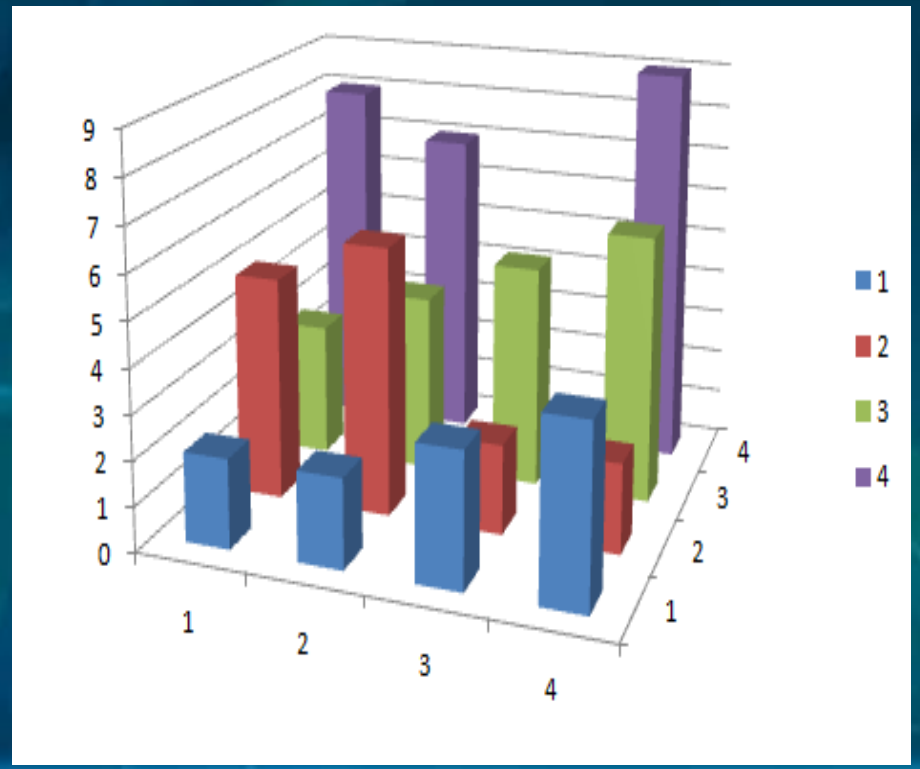
Why Big Data a "big Deal"

- Private Sector
  - Walmart handles more than 1 million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes of data
  - Facebook handles 40 billion photos from its user base.
  - Falcon Credit Card Fraud Detection System protects 2.1 billion active accounts world-wide
- Science
  - Large Synoptic Survey Telescope will generate 140 Terabyte of data every 5 days.
  - Biomedical computation like decoding human Genome & personalized medicine
  - Social science revolution

Visualization

visualization is the process of displaying **data**/information in graphical charts, figures and bars.





What is Big Data Visualization ??

- Big Data visualization is representing data in some systematic form including attributes and variables for the unit of information

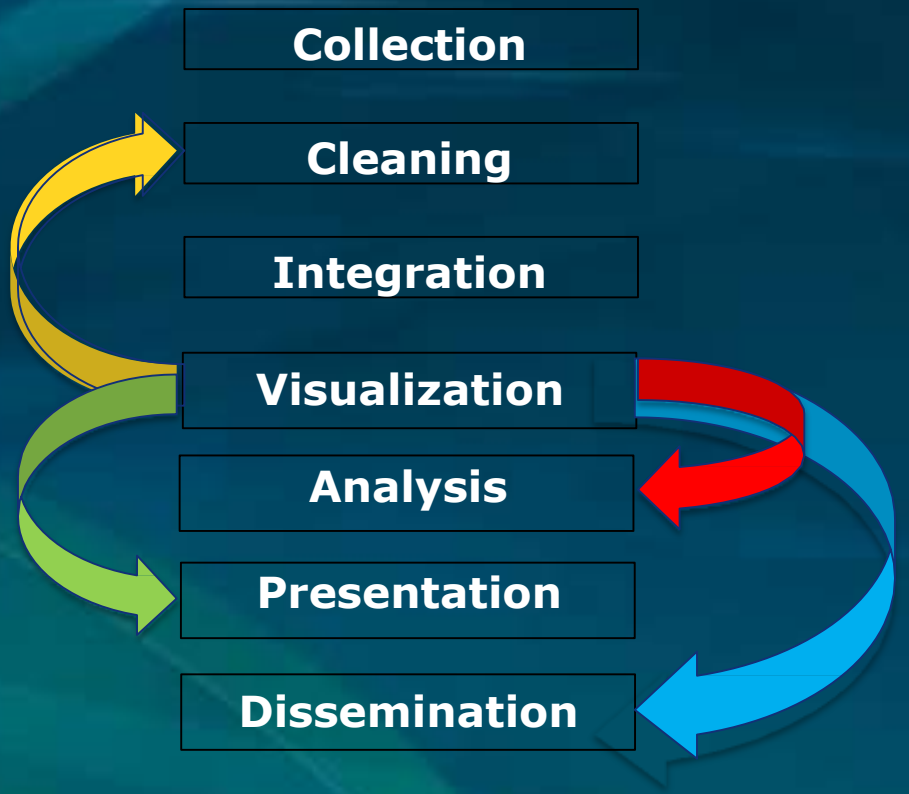
- It uses more interactive, graphical illustrations - including personalization and animation - to display figures and establish connections among pieces of information

- It refers to the implementation of more contemporary visualization techniques to illustrate the relationships within data

Big Data Life Cycle

- Generic process model, Big data analytics processes based on building blocks

- Some building blocks can be skipped, depending on the operating contexts and to go back (two-way street) is admitted



Role of Visualization in Big Data Life Cycle

- Data visualization can play a specific role in several phases of the Big Data Life Cycle
- Data types can affect visualization design
- Visualization methods can informs data cleaning and the choice of analysis algorithms

Along the Big Data life cycle, visualization methods can be properly incorporated in three phases:

- Pre-processing, staging, handling
- Exploratory data analysis
- Presentation of analytical results

Why Data Visualization Important

- The human brain processes information much easily , using charts or graphs to visualize large amounts of complex data.
- It is a quick, easy way to convey concepts in a universal manner.
- We can experiment with different scenarios by making slight adjustments.
- It become easy to predict the future possibilities.

Design Principle

- Objective
  - Think about the content
- Data
  - Numerical : Values measure Something
    - Continuous : Continuity of values
    - Discrete : Discrete set of values
  - Categorical : Values encode a classification
    - Ordinal : Category naturally ordered
    - Nominal : Categories unordered
- Audience
  - Get to know the audience

Steps to Interactive Data Visualization

- *Step 1:* Identify Desired Goals
- *Step 2:* Understand Data Constraints
- *Step 3:* Design Conceptual Model
- *Step 4:* Source & Model Data
- *Step 5:* Design the User Interface
- *Step 6:* Build Core Technology
- *Step 7:* User Test and Refine
- *Step 8:* Launch to Targeted Audience
- *Step 9:* Stay Updated

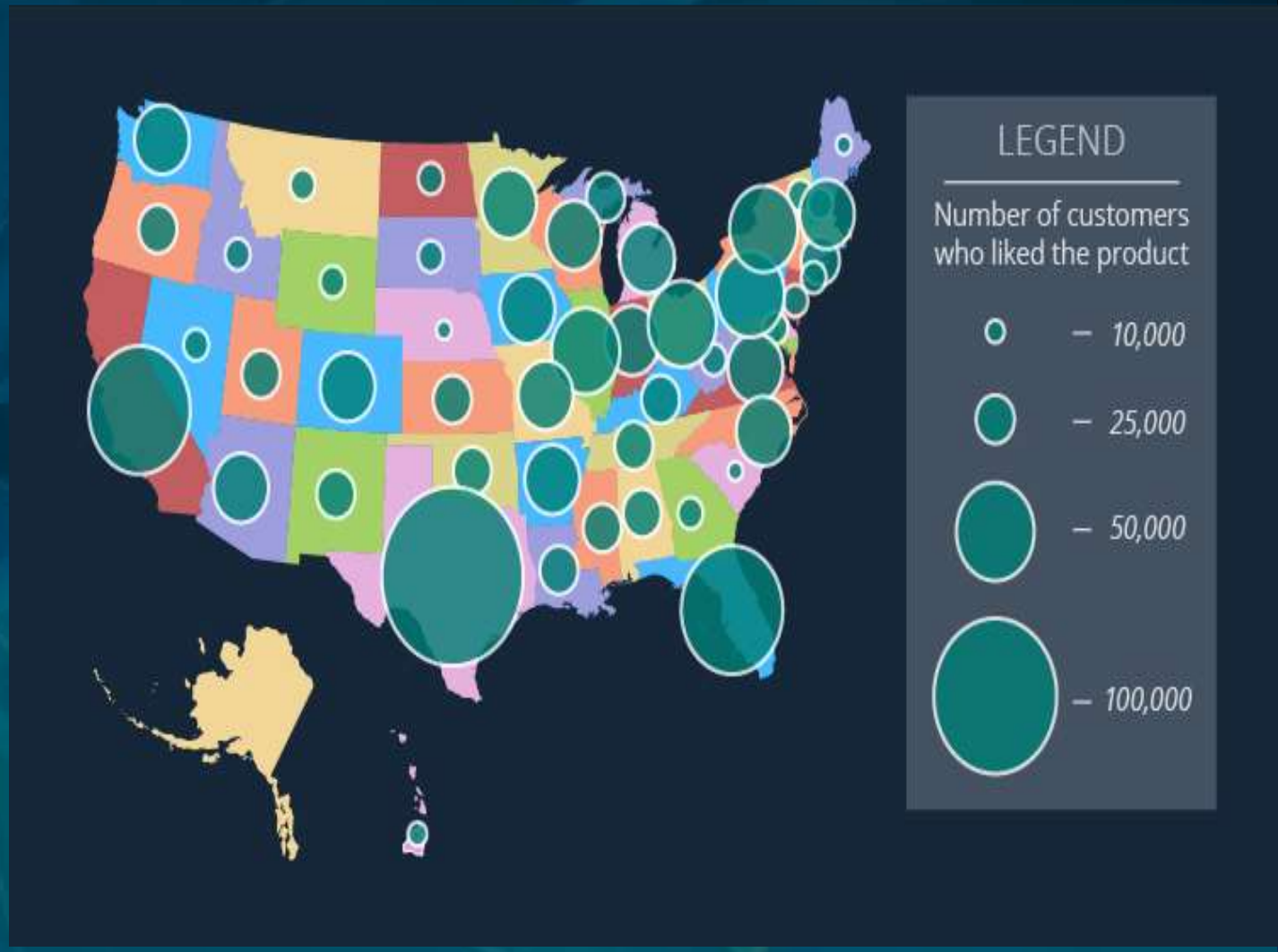
Dedicated big data visualization techniques

Word Cloud

- Displays how frequently words appear in a given body of text
- Words in cloud are of different types
- More the size- higher the frequency
- Used for sentiment analysis of customer's social media posts

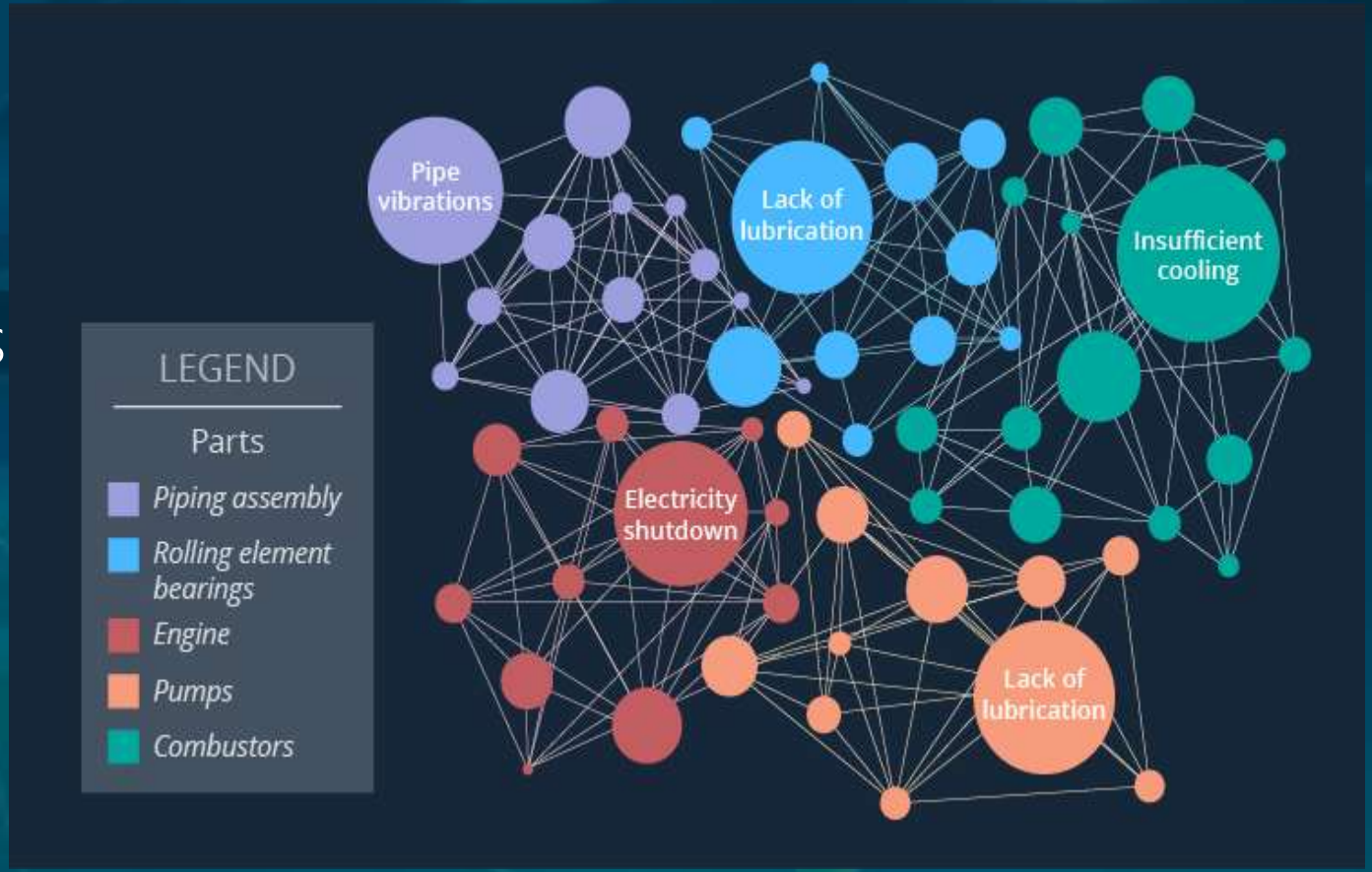
Symbol Maps

- Maps with symbol
- Symbol differ in size, easy to compare
- Used by companies to know the popularity of their product in different areas



Connectivity Charts

- Shows the links b/w phenomena or events
- Based on Connected Graphs theory
- Fig shows the connections between machinery failures and their triggers



Visualization techniques that work for both traditional and big data

Line Charts

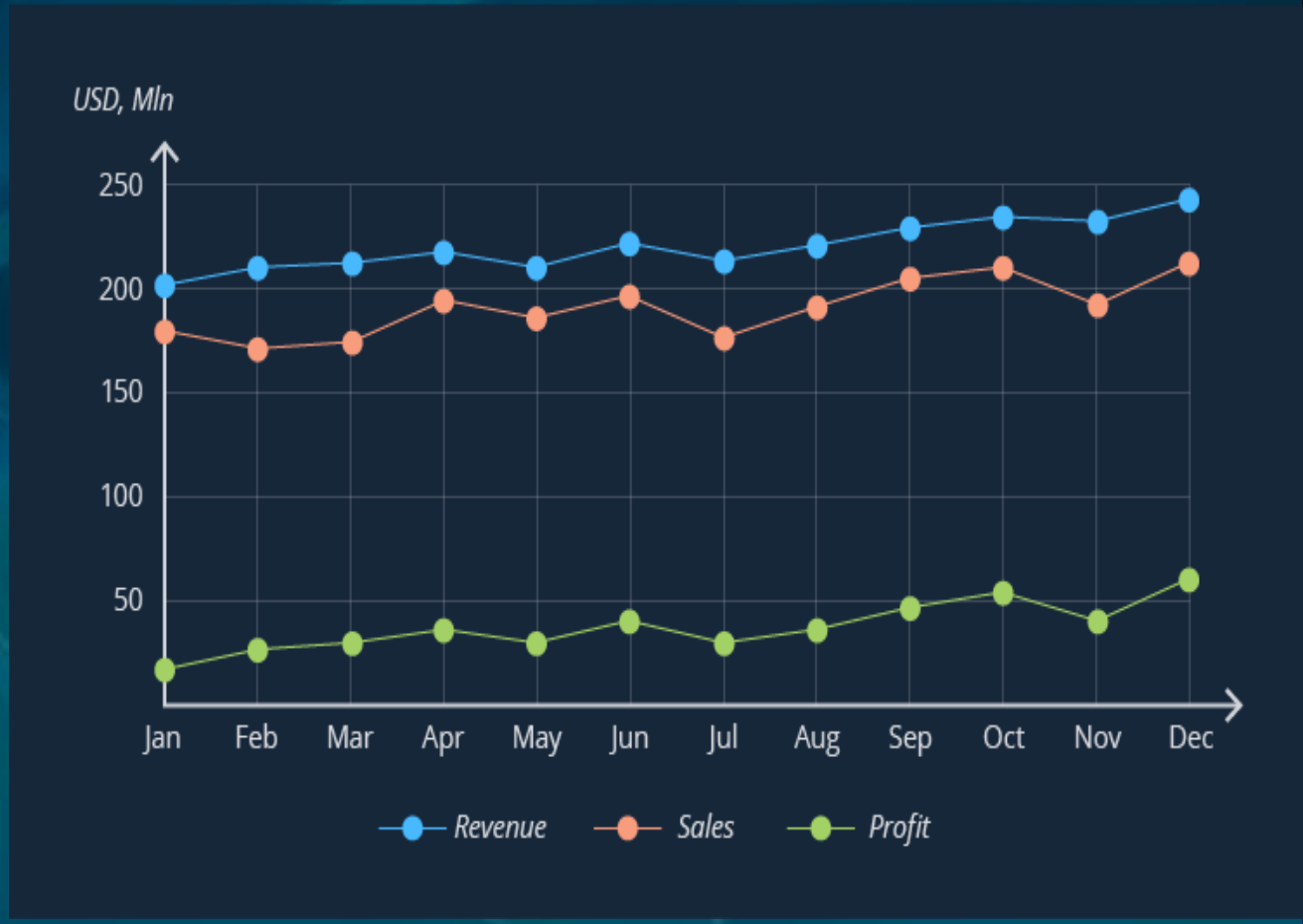
- It looks behavior of one or several variable over time
- It identify the trends between variables.

For traditional

- Shows sales, profit, revenue of last 12 months

For Big Data

- Tracks avg. no. of complaints to call center.
- Total application click by weeks



Heat Maps

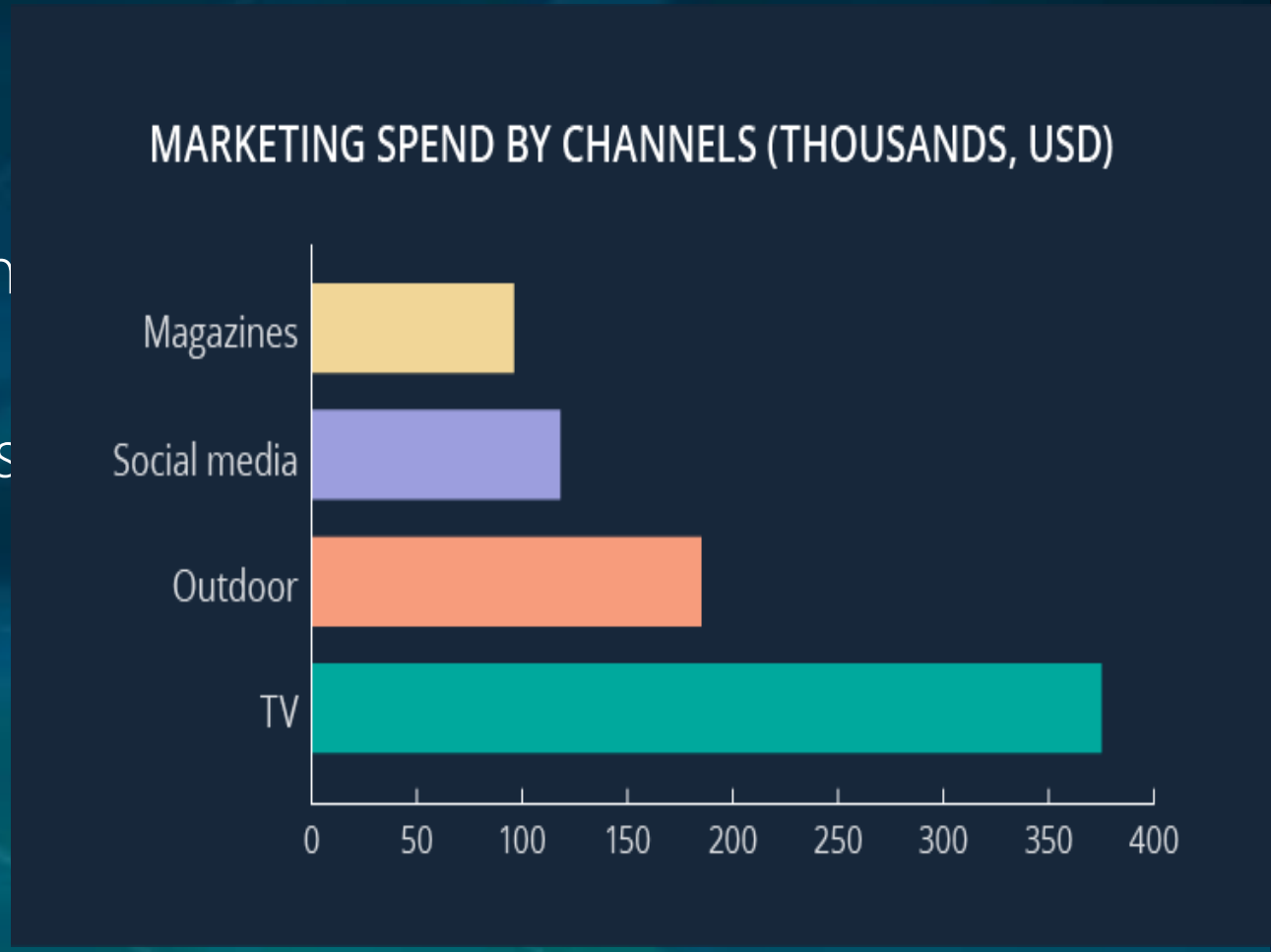
- Two-dimensional representation of data
- Use Color to represent Data
- provides an immediate visual summary of information
- More elaborate heat maps allow the viewer to understand complex data sets

SALES BY QUARTER (BILLIONS, USD)

	Q1	Q2	Q3	Q4
Store 1	121	154	152	185
Store 2	114	156	159	192
Store 3	101	123	138	175
Store 4	131	132	147	164

Bar Charts

- It allow comparing the values of differen variables.
- Graph represents categories on one axis and a discrete value in the other.
- The goal is to show the relationship between the two axes.
- can also show big changes in data over time.



Pie Charts

- It is a circular statistical graphic.
- It is divided into slices to illustrate numerical proportion
- Arc length proportional to quantity it represents.

Making Visualization more Interactive

Zooming	• Zoom In & Zoom Out
Selecting	• Interactive selection of data entities
Linking	• useful for relating information among multiple views
Filtering	• helps users adjust the amount of information for display
Rearranging	• very effective in producing different insights.

Visualization Challenges

- *Visual noise:* Most of the objects in dataset are too relative to each other. Users cannot divide them as separate objects on the screen.
- *Information loss:* Reduction of visible data sets can be used, but leads to information loss.
- *Large image perception:* Data visualization methods are not only limited by aspect ratio and resolution of device, but also by physical perception limits.
- *High rate of image change: Users* observe data and cannot react to the number of data change or its intensity on display.
- *High performance requirements:* It can be hardly noticed in static visualization because of lower visualization speed requirements--high performance requirement.

Benefits : Data Visualization

- Improved Decision-making
- Better ad-hoc data analysis
- Improved collaboration/information sharing
- Time savings
- Increased return of investment (ROI)
- Time savings
- Reduced burden on IT

References

- https://www.promptcloud.com/blog/design-principles-for-effective-data-visualization
- https://www.idashboards.com/blog/2017/07/26/data-visualization-and-the-9-fundamental-design-principles/
- https://www.irjet.net/archives/V4/i1/IRJET-V4I182.pdf
- http://pubs.sciepub.com/dt/1/1/7/

Thank You …