

# Data Visualisation

## CSE3020

# INTRODUCTION

## COURSE OBJECTIVE

1. To understand the various types of data, apply and evaluate the principles of data visualization
2. Acquire skills to apply visualization techniques to a problem and its associated dataset
3. To apply structured approach to create effective visualizations
4. To learn how to bring valuable insight from the massive dataset using visualization
5. To learn how to build visualization dashboard to support decision making
6. To create interactive visualization for better insight using various visualization tools

# INTRODUCTION

## EXPECTED OUTCOMES

On Completion of the course, the students will be able to

1. Identify the different data types, visualization types to bring out the insight.
2. Relate the visualization towards the problem based on the dataset to analyze and bring out valuable insight on large dataset
3. Design visualization dashboard to support the decision making on large scale data
4. Demonstrate the analysis of large dataset using various visualization techniques and tools.

# INTRODUCTION

- **Syllabus:**
- **UNIT - I Introduction to Data Visualization**
  - Introduction to Data Visualization
  - Data Abstraction
  - Task Abstraction
  - Analysis: Four Levels for Validation

# INTRODUCTION

- **Syllabus:**
- **UNIT - II Visualization Techniques**

- Scalar and point techniques
- vector visualization techniques
- multidimensional techniques
- visualizing cluster analysis
- matrix visualization in Bayesian data analysis

# INTRODUCTION

- **Syllabus:**
- **UNIT - III Visual Analytics**
  - Networks and Trees
  - Heat Map
  - Map Color and Other Channels
  - Manipulate View
  - Visual Attributes

# INTRODUCTION

- **Syllabus:**
- **UNIT - IV Visualization Tools and Techniques**
  - Introduction to various data visualization tools
  - Visualization using R

# INTRODUCTION

- **Syllabus:**
- **UNIT - V Diverse Types Of Visual Analysis**
  - Time- Series data visualization
  - Text data visualization
  - Multivariate data visualization and case studies

# INTRODUCTION

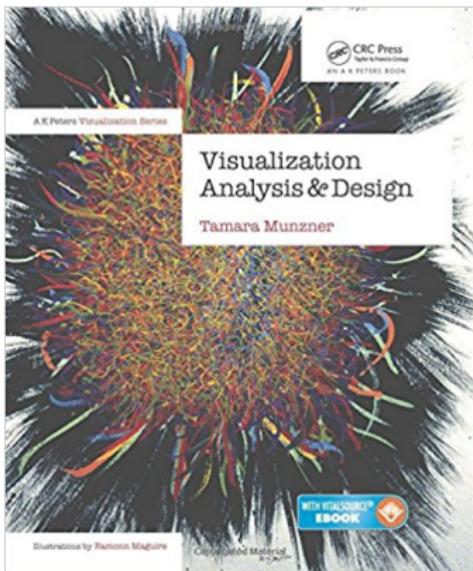
- **Syllabus:**
- **UNIT - V Integration of Data Visualization with Hadoop**
  - Integration of visualization tools with Hadoop
  - Dashboard creation using visualization tools
  - Dashboard creation for Finance
  - Dashboard creation for Marketing
  - Dashboard creation for Healthcare applications

# UNIT - VIII: RECENT TRENDS IN DATA VISUALIZATION

- Guest Lecture from Industry experts

# INTRODUCTION

- **Visualization Analysis and Design** by Tamara Munzer



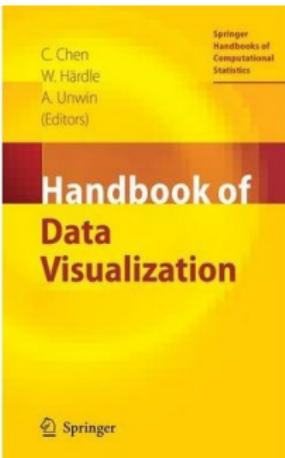
# INTRODUCTION

- **Now You See It** by Stephen Few



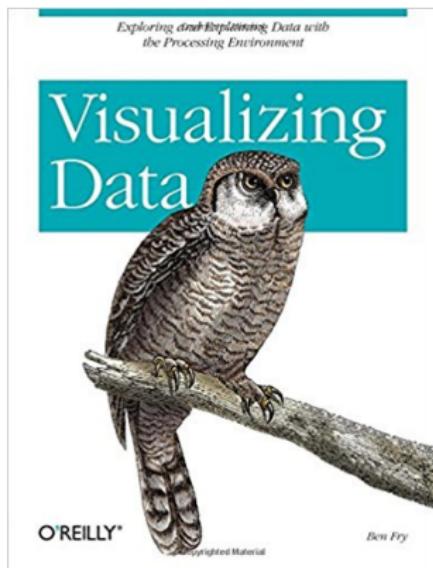
# INTRODUCTION

- **Handbook of Data Visualization** by Dr.Chun-hou Chen



# INTRODUCTION

- **Visualizing Data** by Ben Fry

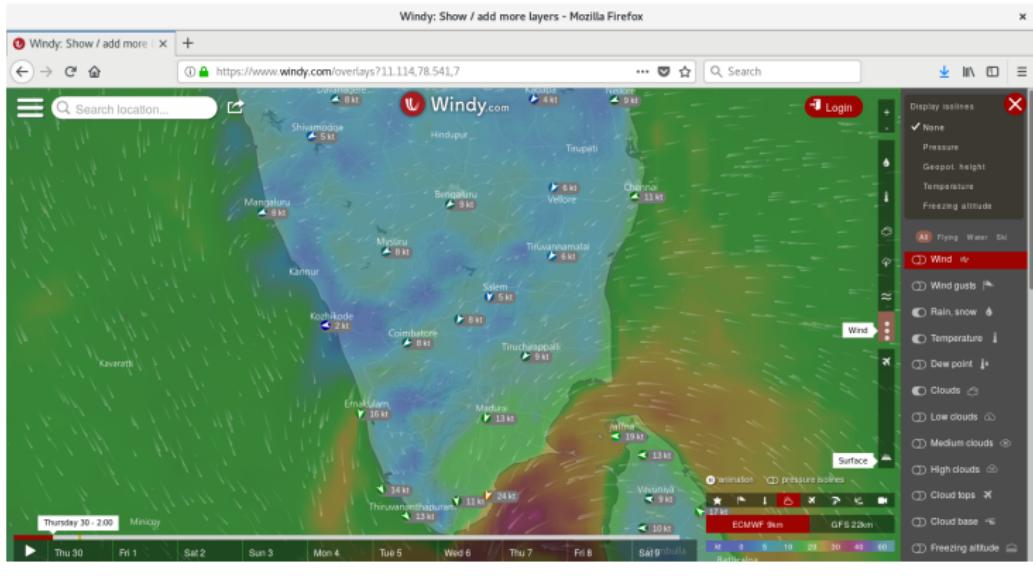


# WHY DO LEARN DATA VISUALIZATION

## • Examples:

- Weather Forecasting

- URL is <https://www.windy.com/>

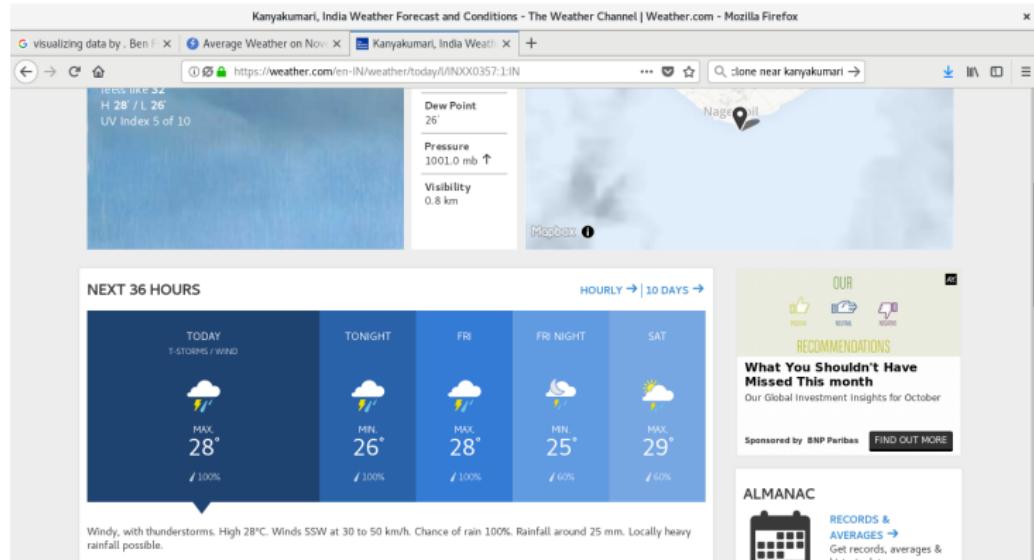


# WHY DO LEARN DATA VISUALIZATION

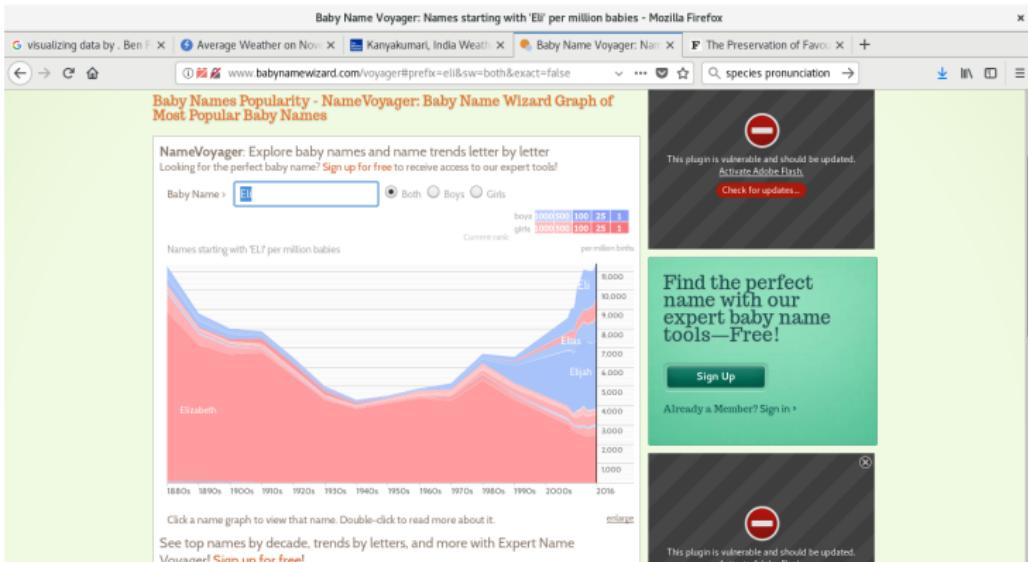
## • Examples:

### • Weather Forecasting

- The URL is *https://weatherspark.com/*



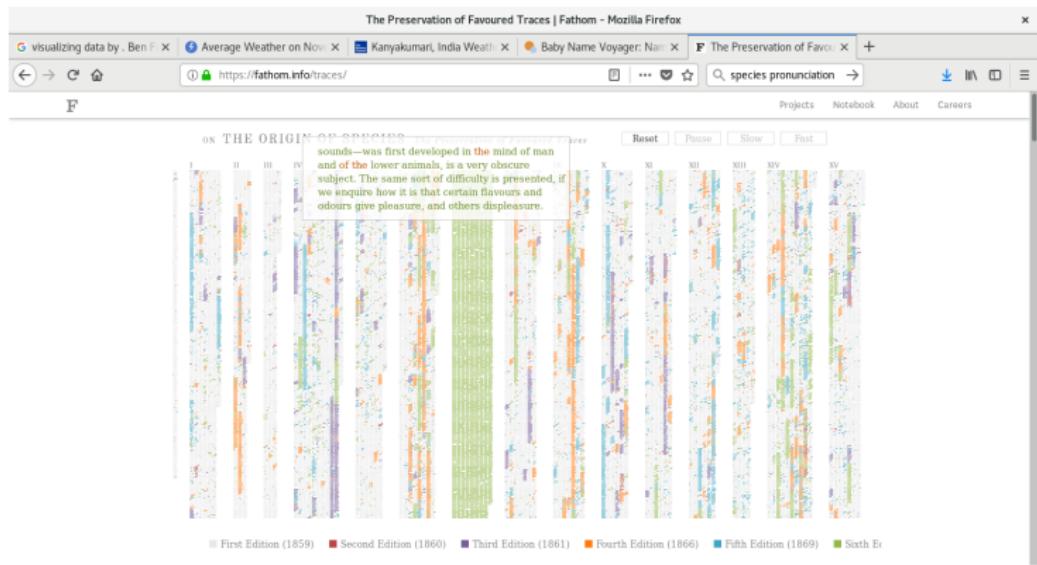
- Baby names Wizard Graph most popular baby names
  - <http://www.babynamewizard.com/>



- Origin of Species Edits
  - <http://benfry.com/traces/>

# • Origin of Species Edits

- <http://benfry.com/traces/>



# GOALS OF THIS COURSE

- Understand what makes a visualization effective through the study of core principles
- Critically evaluate a visual representation of data by looking at various examples in media
- Gain hands-on experience with visualization tools (Tableau, R, D3.js etc)
- Incorporate visualization principles to build an interactive visualization of your own data

# DATA SCIENTIST

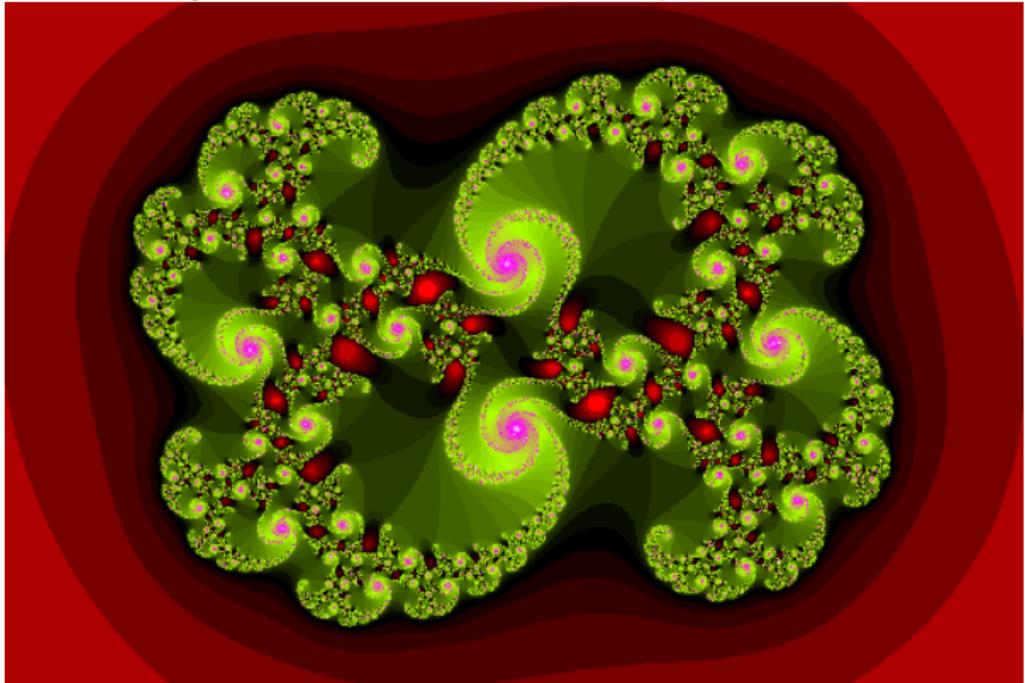
- A Data Scientist is an individual, organization or application that performs statistical analysis, data mining and retrieval processes on a large amount of data to identify trends, figures and other relevant information.
- A data scientist is a job title for an employee or business intelligence (BI) consultant who excels at analyzing data, particularly large amounts of data, to help a business gain a competitive edge.

# VISUALIZATION

- **A picture is worth more than a thousand words** -editor Tess Flanders
- **A picture is worth more than a thousand numbers**
- **A good sketch is better than a long speech** - Napoleon

# VISUALIZATION

- What did you understand from the picture:

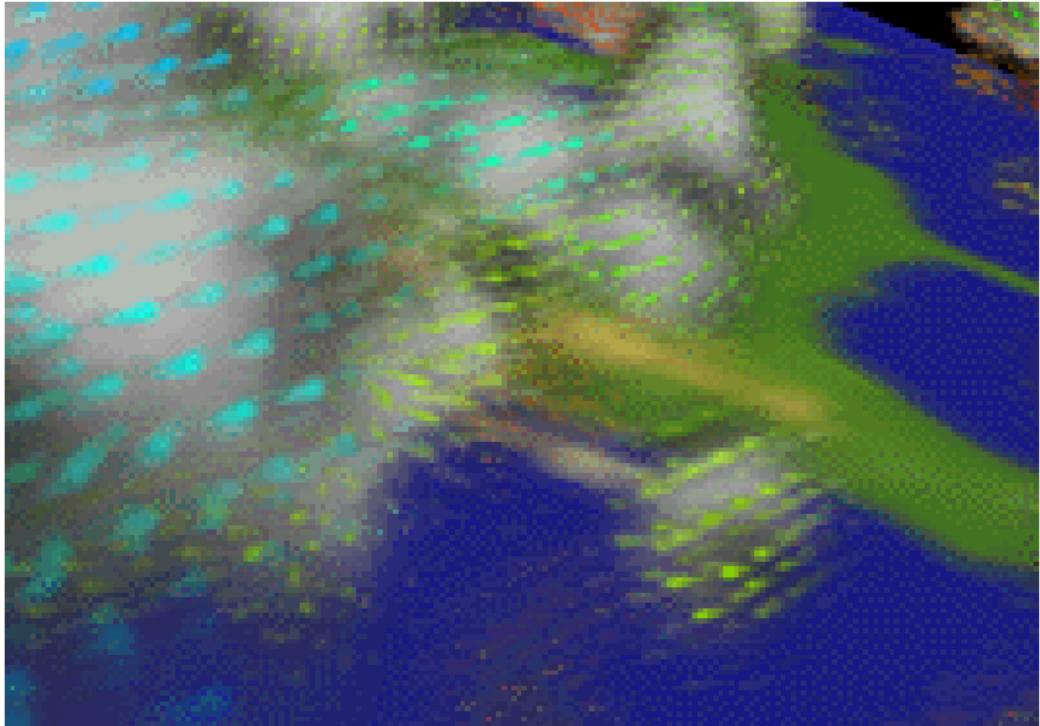


# VISUALIZATION

- Looks like a swirl.
- Small swirls at the edges
- It has different shades, red @ outside , , green @ inside
- The smaller swirls have purple highlights.
- The green has also different shades.
- Each small swirl is composed of even smaller ones.
- The swirls go clockwise.
- Inside the object, there are also red highlights. Those have different shades of red also.
- The green shades vary in a fan, while the purple ones are more uni-color.

# VISUALIZATION

- What will be the scientific information of the following figure:



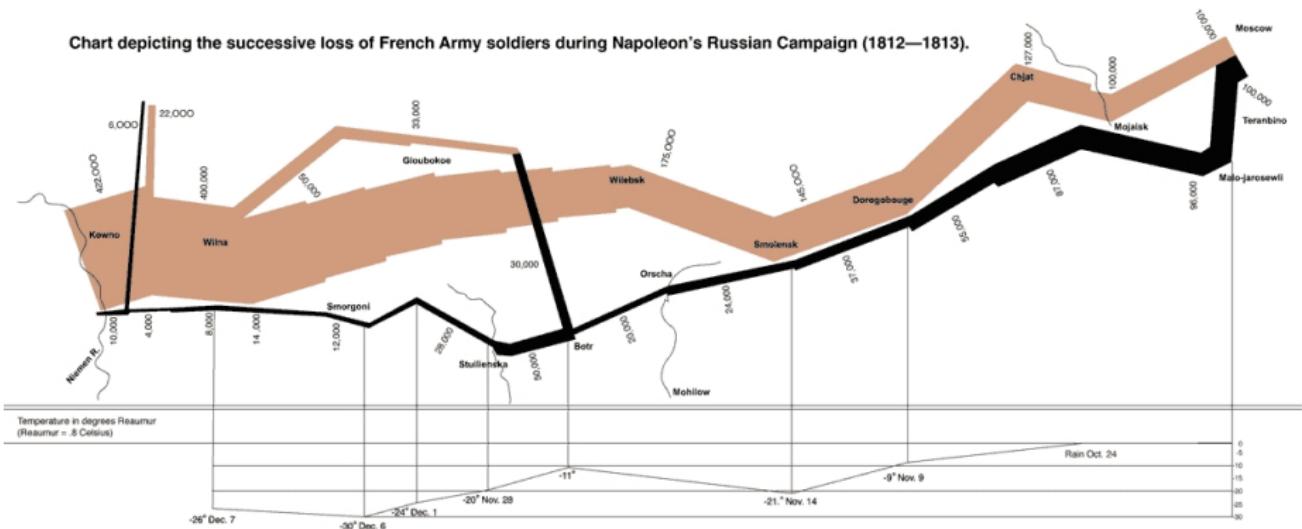
# VISUALIZATION

- Terrain Geometry: (10,20,21), (12,13,14), (13,32,12), (10,20,21) ....
- Terrain Texture: Terrain Geometry: (1,2,3), (2,4,5), (3,5,6),....
- Time 0:
- Volumetric cloud cover:
- 0,0,12,14,15,15,17,12,23,45,....
- Wind Vector:
- (0.2,0.3,0.93,5), (0.4,0.5,0.76,12)....

# VISUALIZATION

- What will be the story, from the figure:

Chart depicting the successive loss of French Army soldiers during Napoleon's Russian Campaign (1812—1813).

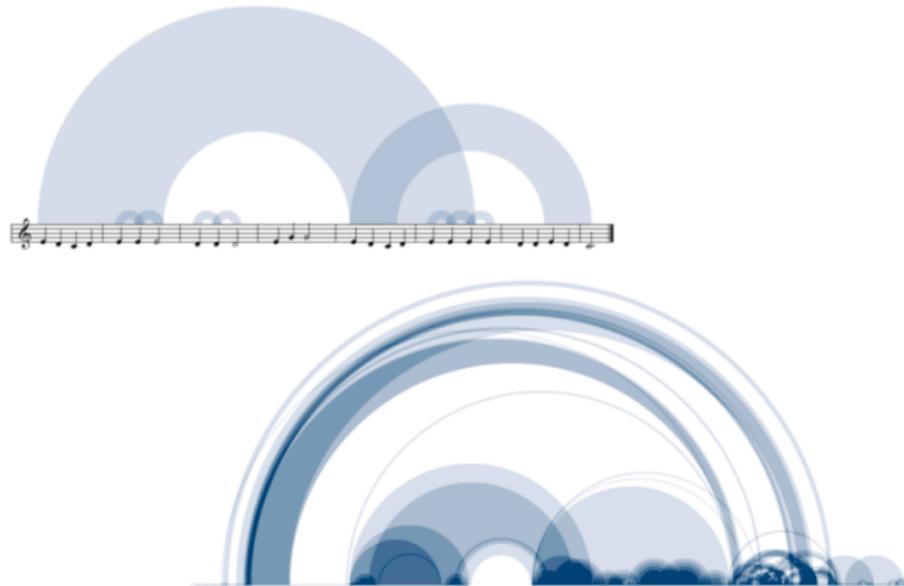


# VISUALIZATION

- Map and line chart showing Napoleon's retreat from Moscow Click for full size Famous visualisation showing Napoleon's advance on Moscow (in brown), and subsequent retreat(in black).
- The X and Y coordinates shows the armies position over time.
- The Width of the line is showing the Napoleon's army size.
- The line-chart at the bottom shows the temperature during the retreat.
- The crossing of river Stultienska on November 28th, with the temperature at -20 'Reamur' (-16 Celsius).

# VISUALIZATION

- Visualization in audio system



Shape of Songs: "Like a Prayer" (Madonna)  
Martin Wattenberg

# VISUALIZATION

- Fried asked help make sense from simulated Data:

| Set A |       | Set B |      | Set C |       | Set D |      |
|-------|-------|-------|------|-------|-------|-------|------|
| X     | Y     | X     | Y    | X     | Y     | X     | Y    |
| 10    | 8.04  | 10    | 9.14 | 10    | 7.46  | 8     | 6.58 |
| 8     | 6.95  | 8     | 8.14 | 8     | 6.77  | 8     | 5.76 |
| 13    | 7.58  | 13    | 8.74 | 13    | 12.74 | 8     | 7.71 |
| 9     | 8.81  | 9     | 8.77 | 9     | 7.11  | 8     | 8.84 |
| 11    | 8.33  | 11    | 9.26 | 11    | 7.81  | 8     | 8.47 |
| 14    | 9.96  | 14    | 8.1  | 14    | 8.84  | 8     | 7.04 |
| 6     | 7.24  | 6     | 6.13 | 6     | 6.08  | 8     | 5.25 |
| 4     | 4.26  | 4     | 3.1  | 4     | 5.39  | 19    | 12.5 |
| 12    | 10.84 | 12    | 9.11 | 12    | 8.15  | 8     | 5.56 |
| 7     | 4.82  | 7     | 7.26 | 7     | 6.42  | 8     | 7.91 |
| 5     | 5.68  | 5     | 4.74 | 5     | 5.73  | 8     | 6.89 |

Summary Statistics Linear Regression

$$\mu_X = 9.0 \quad \sigma_X = 3.317 \quad Y^2 = 3 + 0.5 X$$

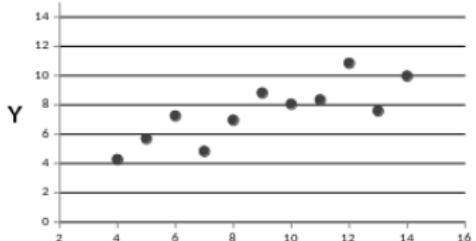
$$\mu_Y = 7.5 \quad \sigma_Y = 2.03 \quad R^2 = 0.67$$

[Anscombe 73]

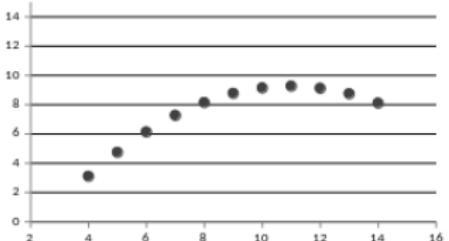
# INTRODUCTION

- Plots for each data set:

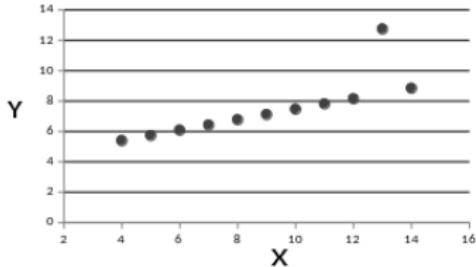
Set A



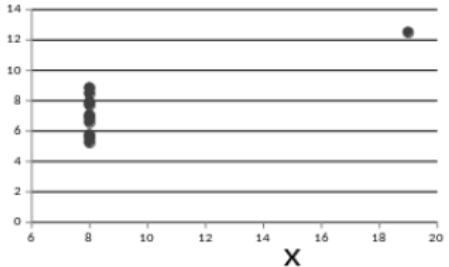
Set B



Set C



Set D



- **"Seeing is believing"** → We have to observe and draw conclusions
- **Seeing is also understanding** → beware of illusion(magicians)
- **What is going do:**
  - Transformation of data or information into pictures
  - It engages primary human sensory apparatus - vision
- **It's like a Tool**
  - Learning or Understanding
  - Compact Representation of Information
  - "Carrier" of Information

# WHAT IS DATA VISUALIZATION

- "Visualization is the process of exploring, transform and represent data as images (or other sensorial forms) to gain insight into phenomena"
- " Computer-based Visualization Systems providing visual representations of datasets intended to help people carry out some task more effectively." -T Munzner
- Visualization links the human eye and computer, helping people to identify patterns and to extract insights from large amounts of data or information

# DATA VISUALIZATION

- **II<sup>nd</sup> Definition:**

- Visual Representation: → Perception Vs Cognition
- Data Sets to help People: → Human in the loop needs the details
- Some task: → Intended Task
- More Efficiently: → Measurable definitions of effectiveness

- **Cognition Vs Perception**

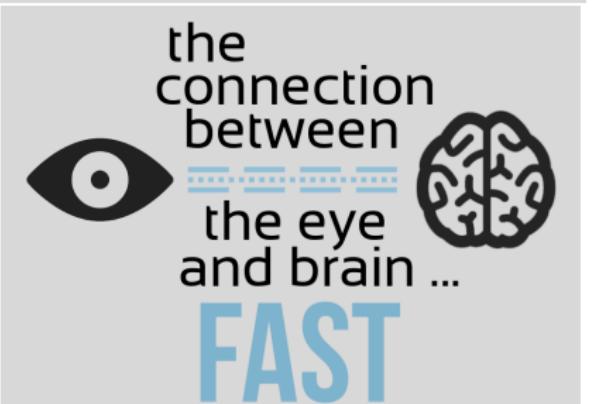
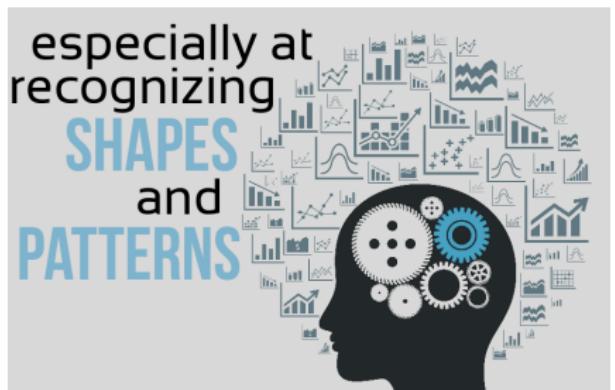
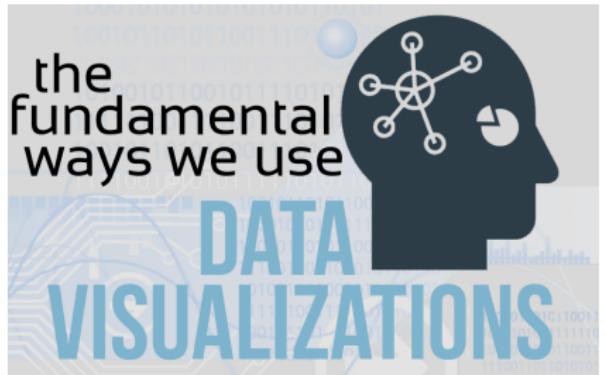
- **Cognitive Task Vs Perceptive Task**

- **eye beats memory**

# DATA VISUALIZATION

- **Data Visualization:**
- To Convey the information through visual representations
- **"produces (interactive) visual representations of abstract data to reinforce human cognition; thus enabling the viewer to gain knowledge about the internal structure of the data and causal relationships in it"**

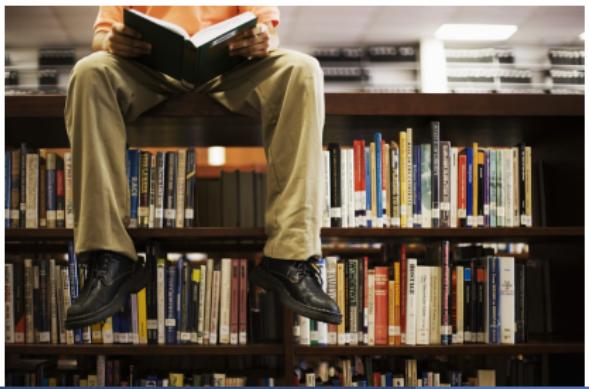
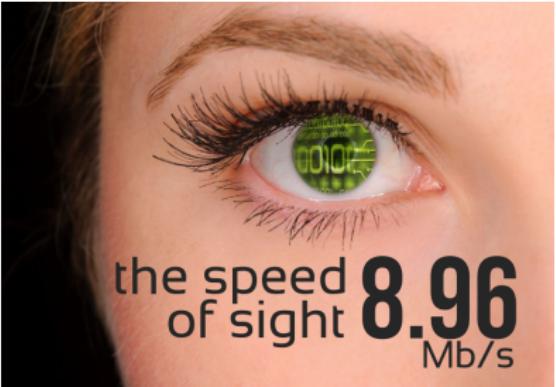
# IMPORTANCE OF VISUALIZATION



# IMPORTANCE OF VISUALIZATION

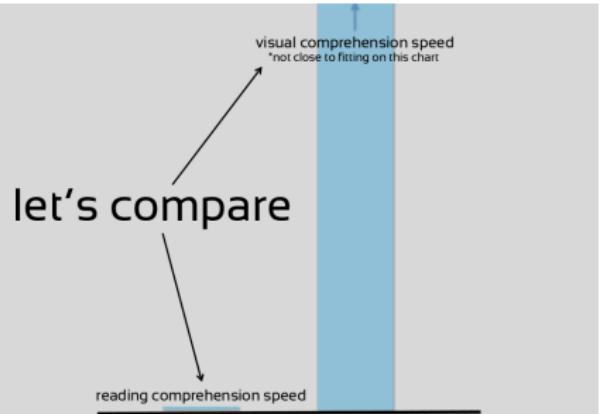
**8.96**  
Mb/s

the brain receives 8.96 megabits of data from the eye every second



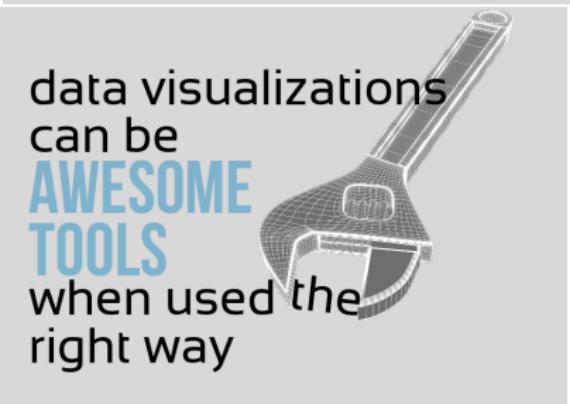
the average person comprehends **120 WORDS** per minute reading

# IMPORTANCE OF VISUALIZATION



A speedometer icon is positioned next to the text 'we are wired to visualize fast'.

we are wired  
to visualize fast



A wrench icon is positioned next to the text 'data visualizations can be AWESOME TOOLS when used the right way'.

data visualizations  
can be  
**AWESOME  
TOOLS**  
when used the  
right way

# TYPES OF DATA VISUALIZATION

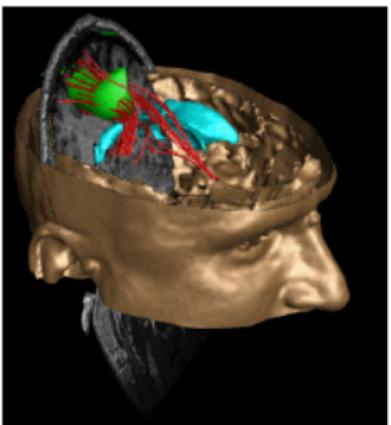
- **Scientific Visualization**
- **Information Visualization**
- **Visual Analytics**

- **Scientific Visualization**

- The science or methodology of quickly and effectively displaying scientific data.
- **The use of interactive visual representations of scientific data, typically physically based, to amplify cognition.**
- It is used for clarification of well-known phenomena.
- Data defined in physical space, i.e. spatio-temporal data (2 to 4 dimensions)
- features of data are well - defined
- Structural Data: Seismic, Medical, data from simulation, etc
- Scientific Visualization and Confirmatory Analysis: Example TeraShake 2.1 Earthquake simulation

# TYPES OF DATA VISUALIZATION

- **Scientific Visualizations:**

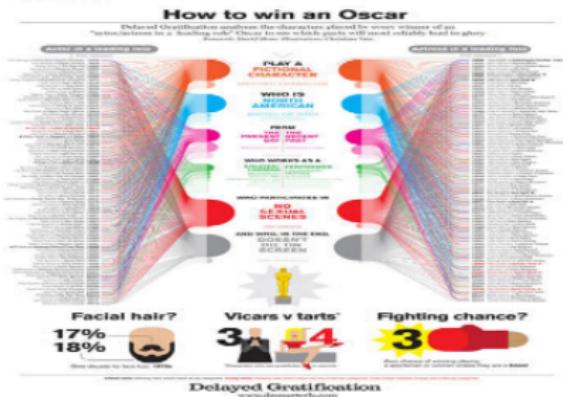


- **Information Visualization**

- **The use of interactive visual representations of abstract, non-physically based data to amplify cognition.**
- it focuses on High Dimensional Data ( $\geq 4$ ) and data is in discrete nature
- It is used for searching for interesting phenomena.
- No inherent structure News, stock market, top grossing movies, facebook connections
- Information Visualization and Exploratory Analysis: Example Grocery Store Analysis

# TYPES OF DATA VISUALIZATION

### • **Information Visualization:**



# TYPES OF DATA VISUALIZATIONS

## • Visual Analytics

- Visual analytics is an outgrowth of the fields of information visualization and scientific visualization that focuses on analytical reasoning facilitated by interactive visual interfaces.
- Use visualization to understand and synthesize large amounts of multimodal data audio, video, text, images, networks of people.

