



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Data Science

Data Visualization

Dr. Ramakrishna Dantu

Associate Professor, BITS Pilani

Introduction to Data Science



Disclaimer and Acknowledgement



Disclaimer

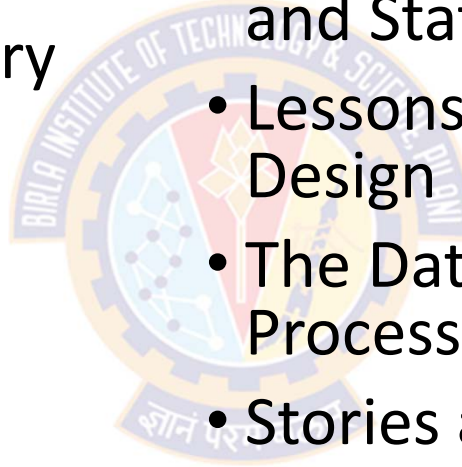
- The content for these slides has been obtained from books and various other source on the Internet
- I here by acknowledge all the contributors for their material and inputs.
- I have provided source information wherever necessary
- I have added and modified the content to suit the requirements of the course

Introduction to Data Science



Data Visualizations

- Data Need for visualization
- Exploratory vs Explanatory Analysis
- Tables , Axis based Visualization and Statistical Plots
- Lessons in Data Visualization Design
- The Data Visualization Design Process
- Stories and Dashboards





Need for Visualization

Need for Visualization



Introduction

- What?
 - Graphical / Visual representation of data
- Why?
 - Way to identify patterns, trends and outliers in data
 - Helps in making data-driven decisions
- "The purpose of visualization is insight, not just the picture."
 - Data visualization pioneer, Ben Shneiderman

Need for Visualization



Why do we visualize quantitative data?

- We visualize quantitative data to perform three fundamental tasks in an effort to achieve three essential goals:

Task	Goal
Exploration: Searching for significant facts	Discovery
Sensemaking: Examining and making sense of data	Understanding
Communication: Conveying information to others	Informed decisions

- Defining data visualization based on these tasks & goals:
 - "Data visualization is the use of visual representations to explore, make sense of, and communicate information"

Need for Visualization



Myths about data visualization

- Questions
 - Why should we use graphical displays (rather than other forms such as tables and audio) to communicate and convey the information to others?
- Some features can be best perceived and understood when quantitative data displayed graphically
- This is because of the ways our brains work
- Vision is by far our dominant sense
- Humans have evolved to perform many data sensing and processing tasks visually
- Data exploration, sensemaking, and communication should always involve an intimate collaboration between seeing and thinking (i.e., visual thinking).

Need for Visualization



Myths about data visualization

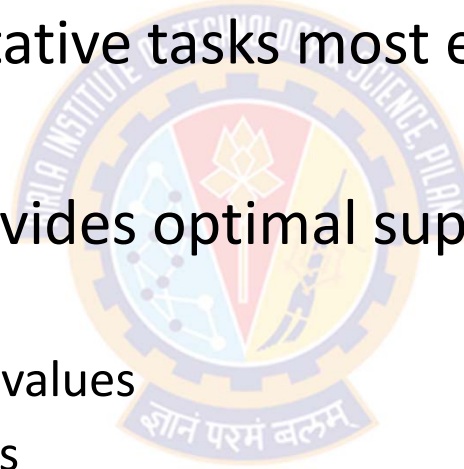
- **Myth #1: We visualize data because some people are visual learners.**
 - True, but everyone benefits from data visualization, whether they consider themselves visual learners or not, including those who prefer numbers.
- **Myth #2: We visualize data for people who have difficulty understanding numbers.**
 - True, but even the brightest mathematicians benefit from seeing quantitative information displayed visually. Data visualization is not a dumbed-down expression of quantitative concepts.
- **Myth #3: We visualize data to grab people's attention with eye-catching but inevitably less informative displays.**
 - It isn't necessary to sacrifice content in lieu of appearance. Data can always be displayed in ways that are optimally informative, pleasing to the eye, and engaging
 - To engage with a data display without being well informed of something useful is a waste.
- **Myth #4: The best data visualizers are those who have been trained in graphic arts.**
 - While training in graphic arts can be useful, it is much more important to understand the data and be trained in visual thinking and communication.
- **Myth #5: Graphics provide the best means of telling stories contained in data.**
 - While it is true that graphics are often useful and sometimes even essential for data-based storytelling, it isn't storytelling itself that demands graphics.
 - Much of storytelling is best expressed in words and numbers rather than images
 - Graphics are useful for storytelling because some features of data are best understood by our brains when they're presented visually.

Need for Visualization



Why do we really visualize data?

- Because humans can perceive particular quantitative features and perform particular quantitative tasks most effectively when the data is expressed graphically
- Visual data processing provides optimal support for the following:
 - Seeing the big picture
 - Easily and rapidly comparing values
 - Seeing patterns among values
 - Comparing patterns



Need for Visualization

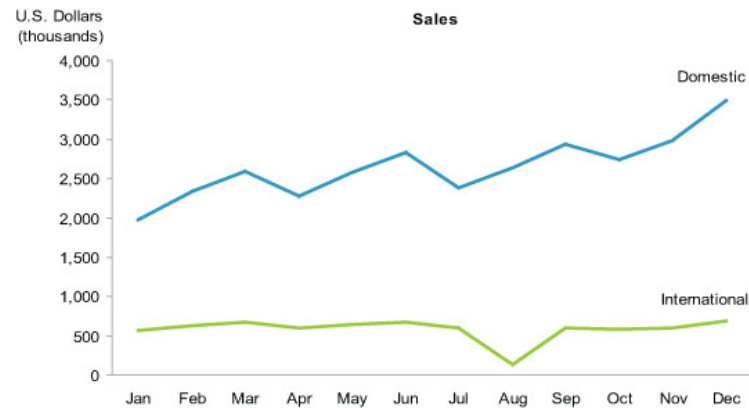


Example

- Seeing patterns among values

Sales

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Domestic	1,983	2,343	2,593	2,283	2,574	2,838	2,382	2,634	2,938	2,739	2,983	3,493
International	574	636	673	593	644	679	593	139	599	583	602	690
	\$2,557	\$2,979	\$3,266	\$2,876	\$3,218	\$3,517	\$2,975	\$2,773	\$3,537	\$3,322	\$3,585	\$4,183



Need for Visualization



Types of Visualizations

- Static Visualizations
- Interactive/Dynamic Visualizations



Need for Visualization



Types of Visualizations

- Static Visualizations

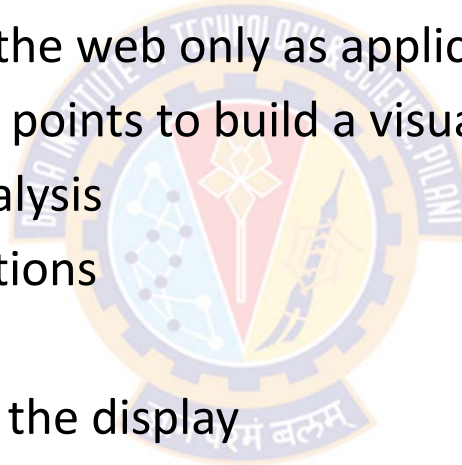
- They are commonly seen as infographics posted on the web or printed as handouts
- Used in presentations, documents, etc.
- Requires careful design as its meant for offline viewing
- User can not adjust the views
- Usually focused on a specific data story, users can't go beyond a single view to explore additional stories beyond what's in front of them
- The story is specifically captured in an engaging single page layout.

Need for Visualization



Types of Visualizations

- Interactive/Dynamic Visualizations
 - They are commonly seen on the web only as applications
 - Users can select specific data points to build a visualized story of their choosing
 - Used for exploratory data analysis
 - Meant for live/online interactions
 - Has more viewing options
 - User gains more control over the display
 - These visualizations allow the user to be part of the data visualization process by building a story of their choosing



Need for Visualization



Types of Visualizations

- Common/General Types of Data Visualization

- Charts
- Tables
- Graphs
- Maps
- Infographics
- Dashboards

- Specific Examples of Methods to Visualize Data

- | | | |
|---|--|--|
| <ul style="list-style-type: none">• Area Chart• Bar Chart• Box-and-whisker Plots• Bubble Cloud• Bullet Graph• Cartogram• Circle View• Dot Distribution Map | <ul style="list-style-type: none">• Gantt Chart• Heat Map• Highlight Table• Histogram• Matrix• Network• Polar Area• Radial Tree• Scatter Plot (2D or 3D) | <ul style="list-style-type: none">• Streamgraph• Text Tables• Timeline• Treemap• Wedge Stack Graph• Word Cloud• Any combination of the above |
|---|--|--|

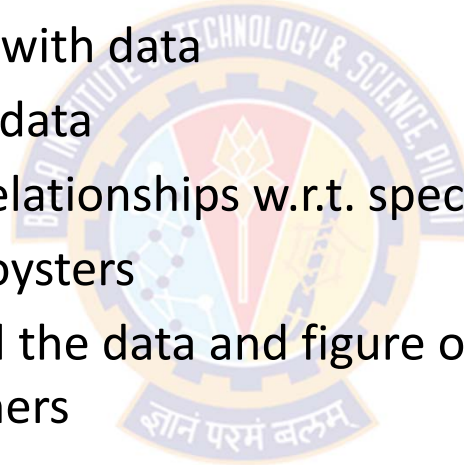
Need for Visualization



Exploratory Vs. Explanatory Analysis

- Exploratory Analysis

- Required to become familiar with data
- Involves digging through the data
- Helps in finding trends and relationships w.r.t. specific goals
- It's like hunting for pearls in oysters
- The purpose is to understand the data and figure out what might be noteworthy or interesting to highlight to others
 - E.g., hunting for pearls in oysters
- We might have to test 100 hypotheses or look at the data in 100 different ways
 - E.g., we might have to open 100 oysters to find perhaps two pearls



Need for Visualization



Exploratory Vs. Explanatory Analysis

- Explanatory Analysis

- It's about communicating our findings
- It's about telling a story with data
 - It's about explaining why it is so or how it is so
- We need to be in the explanatory space when we want to communicate the results of our analysis to our audience
- At this stage, we should have a specific story to tell or a specific thing we want to explain
 - E.g., probably about those two pearls.



Data Visualization - Context

Data Visualization – Context



Agenda

- Under this topic, we will focus on the following aspects:
 - Understanding the context
 - Effective storytelling strategies

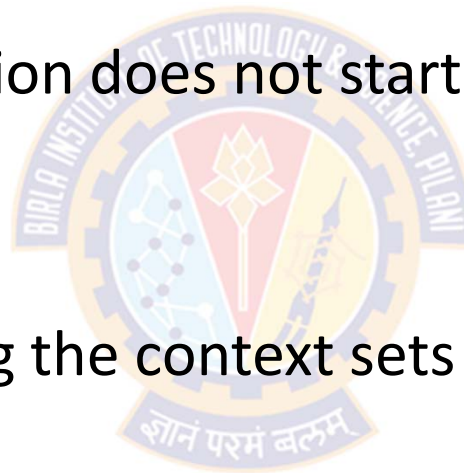


Data Visualization – Context



Success criteria

- "Success in data visualization does not start with data visualization itself." -- Cole Knaflc
- ...understanding the context sets a solid foundation for data visualization creation



Data Visualization – Context



Who, What, & How

- Consider three things in explanatory analysis and data visualization:
 - Who
 - First, to whom are we communicating?
 - Who is our audience and how they perceive us?
 - It is important to identify common ground so that we can ensure they hear our message
 - What
 - Second, what do we want our audience to know or do? – Action items or takeaways.
 - We should be clear about how our audience should act and make use of our communication
 - How
 - Next, how can we use data to help make your point?

Data Visualization – Context



Who – Our Audience

- Our audience
 - Being clear about our audience helps us communicate better
 - Knowing them places us in a better position for communication
 - We need to be specific while identifying our audience
 - Avoid general audiences, such as:
 - "internal and external stakeholders" or
 - "anyone who might be interested"
 - If we have audience with disparate needs, we cannot communicate effectively to anyone
 - We may need to create different communications for different audiences

Data Visualization – Context



Who - We

- We, the presenters
 - Key considerations to be given when determining the structure of our communication:
 - What is our relationship with our audience?
 - How do they perceive us?
 - Are we encountering with our audience for the first time through this communication or do we already have an established relationship?
 - Do they trust us as an expert, or do we need to work to establish credibility?
 - Whether and when to use data, and may impact the order and flow of the overall story we aim to tell

Data Visualization – Context



What

- What do we want our audience to know or to do?
 - Action
 - Mechanism
 - Tone



Data Visualization – Context



What – Action

- Make what we communicate relevant for our audience
- It helps us in ensuring our audience care about what we have to say
- Our objective should be such that we always want our audience to know or do something
- Sometimes, we may not concisely articulate what we need to communicate
- This could be because we believe our audience knows better than us
 - This assumption is false
- We are the subject matter experts here because we are the one analyzing the data
- This puts us in a unique position to interpret the data and help lead people to understanding and take action
- If no action recommendation possible / feasible, then encourage discussion towards one
- In general, the presenter should take a more confident stance when it comes to making specific observations and recommendations based on their analysis