DATA VISUALIZATION PRINCIPLES AND PRACTICE
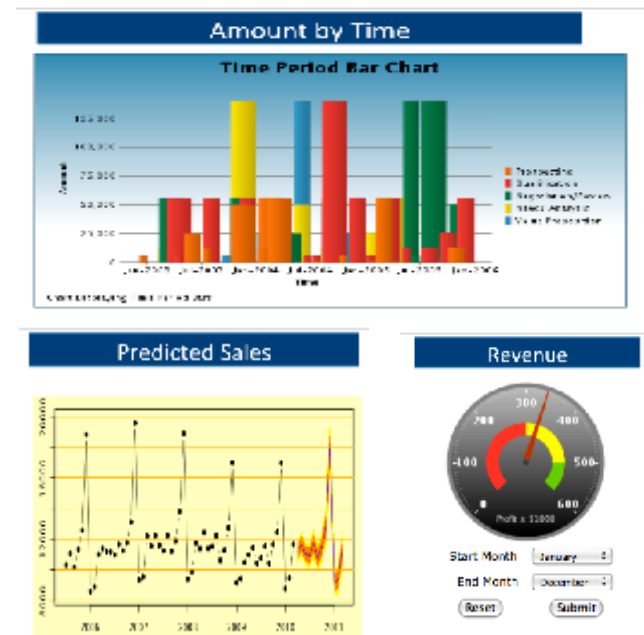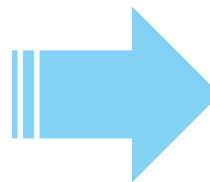
SACHIN KUMAR MANJHI

Agenda

➢ Data Visualization

    ➢ What is it?

    ➢ Why is it important?

➢ Rules of Visual Perception

➢ Edward Tufte principle of chart design

➢ Gestalt Principles

➢ Grice's Maxims

➢ Effective Dashboard Design

Data Visualization – What is it?

- General understanding amongst analysts/business users:

  - Visualization is the <u>graphical presentation</u> of <u>information</u>, with the goal of providing the viewer with a qualitative understanding of the data.

  - Graphical presentation may entail manipulation of graphical entities (points, lines, shapes, images, text) and attributes (color, size, position, shape).

DATA

PICTURES

Data Visualization: Why is it important?

- Vision stands out as the primary and most powerful channel of input to brain(~90% perception).

- Vision is also most intimately connected to cognition.

  - We remember things better when we see them.

- A visual can communicate more information than a table in a much smaller space.

Data Visualization: Why is it important?

- Trends, patterns, exceptions are all easier to spot if we look at pictures rather than tables of data.

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Each set has a completely different pattern.

Can we spot the difference between the 4 sets of data? What about outliers?
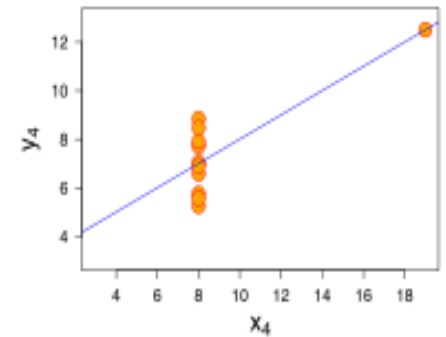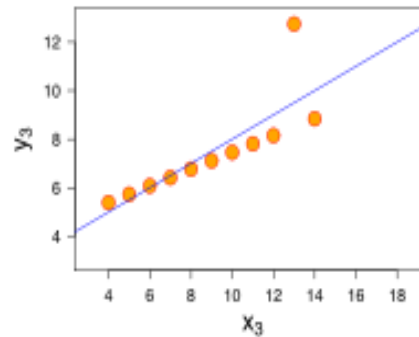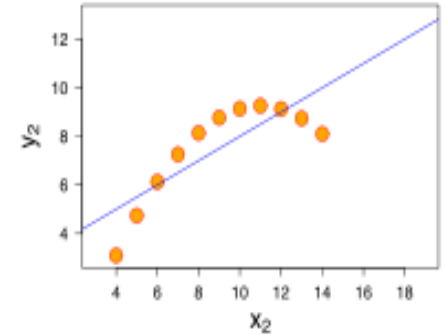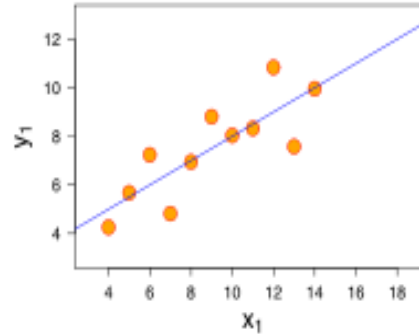
Now can we do it in 2 sec or less?

Computers are good with data. Humans are very good with pictures.

Data Visualization – What is it used for in Business context?

Data visualization is the graphical representation of information for <u>Analysis</u> and <u>Communication</u>.

Visual Analysis	- Use our innate visual advantage for processing information - Spot pattern in large amount of data. See things that are not obvious otherwise. - Interact with data and determine actionable insights - What happens if we make adjustment to that area?
Visual Communication	- Presentation and communication of data in an intuitive manner - Convey information in simple manner, so we all see and understand the same things. - Guide the user to focus on the important aspects of data - Effectively represent large amounts of data in a small area

Any Data Visualization needs to start with an audience in mind.
The "What would you like to know?" needs to be followed up with
"What would you do if you knew this information?"

Data Visualization: Rules of Visual Perception

- Visual perception follows certain rules.
  - Our brain doesn't process all the light the enters the eye. It does this selectively.
  - Our eyes are drawn to familiar patter. We "see" what we expect to see.
  - We see things in relation to its surrounding.

- If we can utilize the rules of visual perception, we can be more effective in
  - analyzing data to find meaning
  - putting the "story" across to the consumer of data

- Three primary set of rules:
  - Tufte's Rules
  - Gestalt Principle of Perception
  - Grice's Maxim of Communication

- If we disobey the rules, our data will be misleading or incomprehensible.

Data Visualization: Why is it important?

We see what we expect to see.

Can you find the second image?

Did anyone expect a Dolphin on a rose?

And now, after seeing this, can we see the Dolphin in the first image too?

Courtesy: Coolbubble.com

Tufte Principles

Edward Tufte – considered the father of Data Visualization - invented the concept of Chart Junk.

His most famous work – *The Visual Display of Quantitative Information was* published in 1983.

It is probably the most important book ever written on data visualization.

- **Principles of Data Visualization:**

    1. **Graphical Integrity:** Visual representations of data must tell the truth.

    2. **Data-Ink Ratio:** Data Ink is the ink on a graph that represents data. Good graphical representations maximize data-ink and erase as much non-data-ink as possible.

    3. **Avoid ChartJunk:** Chart junk is the excessive and unnecessary use of graphical effects in graphs.

    4. **Data Density:** Proportion of the total size of the graph that is dedicated displaying data. Maximize data density and data matrix within reason.

    5. **Small Multiples: S**eries of the same small graph repeated in one visual. Small multiples are a great tool to visualize large quantities of data and with a high number of dimensions.

Gestalt Principle of Perception

A group of researchers in early 20th century, known as Gestalt psychologists formulated the regularities according to which the perceptual input is organized into unitary forms.

- **Proximity**: Objects that are close together or connected are perceived as a group

- **Similarity**: Objects that share similar attributes, color or shape, are perceived as a group

- **Enclosure and Continuity**: Objects that appear to have a boundary or a continuation around them are perceived as a group

- **Closure**: Open structures can be easily perceived as closed, complete
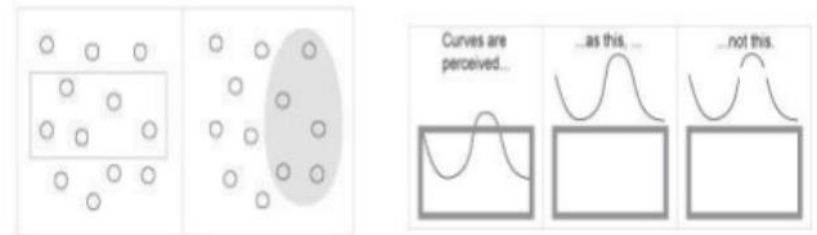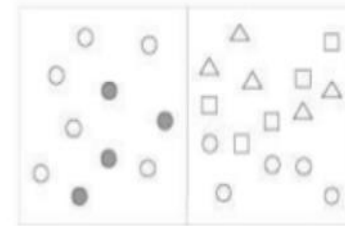
Grice's Maxims of Polite Conversation

- Paul Grice , a 20th century philosopher is well known for his conversational maxims, which attempt the describe the characteristics of polite conversation.

- Grice's conversational maxims: 4 categories:

Quantity	• provide as much information as necessary • do not provide more information than necessary
Quality	• be truthful – do not give information that is false • do not give information that is not supported by evidence.
Relevance	• says things that are pertinent to the discussion
Manner	• be orderly • be brief • avoid ambiguity

- These maxims of conversation apply equally well to **visual communication** of quantitative information in the workplace.

Effective Dashboard Design Principles

1. Consider Your Audience

- Ask how a dashboard will be used ?
- What information does the user need to be successful?
- How much detail does the user need?
- What action can be taken and how?
- How are exceptions or insights that need action highlighted?
- What do colors mean and can they be visually interpreted?
- Which icons are familiar?

2. Use Best Practice Dashboard Designs

- Good design should tell a story. Avoid way too much information, clutter or noise. Limit content to fit entirely on one screen.
- Starting with the highest level of detail in the upper corner of the screen and show more detail you move down in the direction the audience is used to reading.
- Keep your dashboard simple with only a 3 to 5 key values, charts, or tables.
- Remember to provide adequate context and keep related items near each other.
- Avoid displaying "singular numbers" without any other context. Show degrees of change for quick comparisons.
- Avoid data visualization variety for the sake of variety.
- If detail tables are needed, place them on the bottom of the dashboard

Effective Dashboard Design Principles

3. Avoid Common Data Visualization Issues

- Choose appropriate data visualizations. Don't use charts that distort reality i.e. 3-D charts. Keep in mind that it is difficult for the human brain to interpret circular shapes.

- Be consistent with chart scales on axes, chart dimension ordering and also the colors used for dimension values within charts.

- Be sure to encode quantitative data nicely.

- Don't exceed three or four numerals when displaying numbers.

- Display measures to one or two numerals left of the decimal point and scale for thousands or millions i.e. 3.4 million not 3,400,000.

- Make sure that time frames are well understood.

- Don't have one chart that has last month next to filtered charts from a specific month of the year.

- Don't mix big and small measures on the same scale, such as on a line or bar chart.

- Don't clutter your charts with data labels that are not needed.

Presenting Data : Charting Best Practice

Following are examples of some of the Chart Types available for various Scenarios:

Scenario	Chart Types
Comparison	• Line, Bar, Stacked, Bubble, Table, Tag Cloud chart , Word Cloud , Radar chart, Bullet , Gauge
Distribution	• Bar, Column, Scatter, Area
Composition	• Bar, Stacked, Table, Pie, Doughnut, Waterfall, Treemap
Trend	• Line, Bar, Stacked, Combo, Sparkline, Motion Chart
Relationship	• Scatterplot, Bubble, Motion Chart, Table, Network diagram
Performance	• Gauge , HeatMap, Bullet
Hierarchy	• Tree Map, Network diagram

Hans Rosling's Data Visualization

https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen