

# DES433 Lecture Summary

Data Visualization (Final Exam)

SIIT DE-DS Y4T1/2022 – By Paphana Yiwsiw (@waterthatfrozen)

## Contents

<b>01</b> – The Big Idea	2
<b>02</b> – Effective Visual	2
<b>03</b> – Data Representation	4
<b>04</b> – Visual Coding	6
<b>05</b> – Storytelling with Data	7
<b>06</b> – Interactive Data Visualization	7
<b>07</b> – Dashboard Design	7

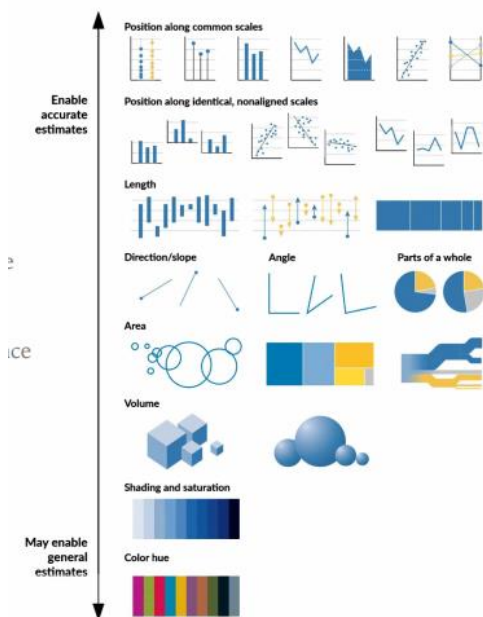
## 01 – The Big Idea

- Data visualization = art of communication + science of data analytics
- You need to bring that story, both visually and contextually, to life with your data and tools.
- Art of communication
  - o You need to know who you want to talk, what you want them to do, how to talk to them.
  - o Don't forget about mood and tone of your message and delivery.
- Big Idea: Articulate your POV, convey what is at stake, complete in one sentence.

## 02 – Effective Visual

- Focused on delivery mechanisms, how you convey your message to the audience.

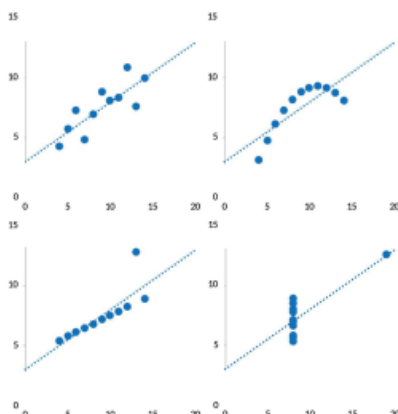
### Perceptual Ranking



- How our brain perceives the visual stimuli.
- What kind of visualizations you choose to create will depend on your goal, audience need, experience, and expertise.
- Order of perceptual ranking from worst to best.
  - o Color Hue
  - o Shading and Saturation
  - o Volume (3D is worse than 2D on flat screen)
  - o Area
  - o Direction/Shape, Angle, Parts to whole
  - o Length
  - o Position along identical, nonaligned scale
  - o Position along aligned scale.

- Standard graphs like bar and line are common because they are more accurately perceived, familiar to people, and easy to create.
- Nonstandard graphs like one used curves and circles may not allow for most accurate perception of the exact data values.
- However, many times that accuracy is not the goal, but reader engagements is also important.

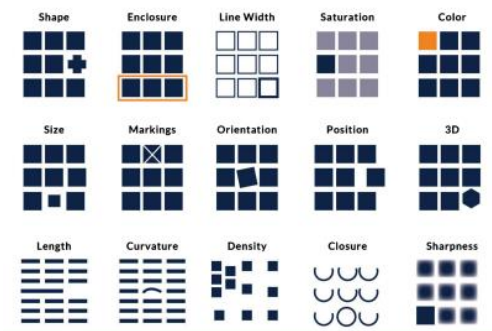
### Cognitive science part



- Anscombe's Quartet
- It is easier to see the relationship in the graph and what's happening in the data such as outlier and curvature even the data is belong to the same regression line, correlation coefficient, mean, and variance.

## Gestalt Theory

- Show how we tend to organize visual element into groups.
- **Proximity:** the closer, more likely that we perceived as a group.
- **Similarity:** we tend to group object that share similar attributes such as color, shape, direction together, even if location is closer to another group, **the color gives stronger sense** more than proximity.
- **Enclosure:** Bounded objects are perceived as a group, enclosure is stronger than color. If the color is not the same as bounded, it will be perceived as an error
- **Closure:** We tend to ignore gaps and try to close it between two segments with a straight line even the data might be a different pattern.
- **Continuity:** We seek smooth path when following a sequence of shape.
- **Connection:** We connect objects as members of the same group.
- **PREATTENTIVE PROCESSING**
  - o The reader will grab attention of what we put as important in the picture first.
  - o To emphasize part of the picture, stress it out. Just grab reader attention.
  - o Example, different color and font weight.



- o
  - Different shape, enclosure, line width, saturation, color, size, markings, orientation, positions, dimensions, length, curvature, density, closure, sharpness.

## Combination of Science and Art

- Science: use visualization that allow readers to perceive the absolute value more accurately.
- Art: create visuals that engage and excite the reader, even graphs are not accurate.
- Perform between the tradeoff between engagement and accuracy. Even if the reader may not interested in the topics as we hope or doesn't have the expertise to grasp the content.
- Don't forget to think about different audience types and their role. → Different reader, different interest, or engagement with the visual itself.

## Visualization Balance



- It is not the same as see picture but rather read a paragraph when reading a graph. You also need a story, accuracy, and attention.
- You ideally want accurate and engaging graphs but need a trade-off. Depends on the message and type of the audience you planned.
- Boring graphs are forgettable. We should not operate from the assumption that reader will pay attention to more familiar graphs.
- Note: we see the curve, think as more creative.

## 03 – Data Representation

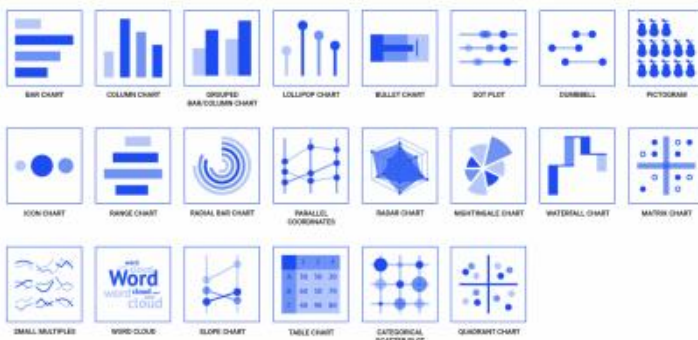
### Five Guidelines for Data Visualization

1. **Show the data:** Don't show something that will shift the reader focus, but not show all data, only the part that convey the story to the reader.
2. **Reduce the clutter:** reduce the confusing element in the graph such as label and number, unrelated graphics such as 3D, background, and pattern.
3. **Integrate graphic and text:** text around the graphics is just as important.
  - a. *Remove legends when possible:* legends force a lot more work and energy to your reader, better approach is to put the text on data directly if possible.
  - b. *Write active title:* It must be powerful, grab attention, and convey the message. Make the legend to graph so it can make graphs more understandable. **YOU NEED TEXT!**
  - c. *Add explainers:* add detail to the graph, don't have to explain, just include necessary fun fact about the data related to your message.
4. **Avoid spaghetti chart:** Avoid graphs with so much information inside a small area that make it difficult to see.
5. **Start with gray:** Start with gray first, so you know what important element in the graph is and know what part needed to stress out.

### Chart Types

1. **Categorical Comparison:** key message is comparison. Most common, easy to read, and clear.

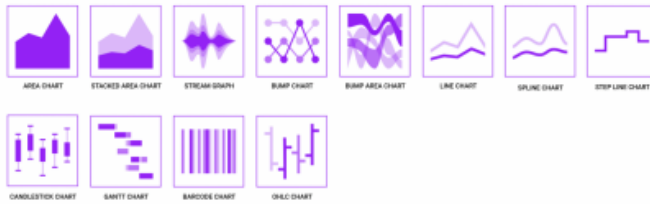
#### COMPARISON



- a. Bar Charts: width or height of bars are the same.
  - i. Column/Bar charts: vertical(column), horizontal (bar). Good with one category.
  - ii. Paired Bar charts: For more category, create a grouped bar chart.
  - iii. Stacked bar charts helps when datasets are based on the same references. Valuable where the relative proportion is as important as the change.
  - iv. Diverging bar charts: normally show data with positive or negative side.
  - v. Nightingale charts: most famous, could be confused for donut or pie.
  - vi. Radial bar charts: less accuracy but attract more audiences.
- b. Mosaic Charts: width or height of bars are not equal. Use area to present data.
- c. Heat Map Charts: use color and saturation to present data.
- d. Gauge Charts: useful with dashboard, not useful on its own.
- e. Bubble Comparison: same with mosaic chart but with circle, area, color, to present the data. It looks cooler because of circles and curves.

## 2. Temporal Data: data is collected as time-series. Show continuity and relationship.

### DATA OVER TIME (TEMPORAL)



- Line chart is the easiest to read, clear in presentation, easily drawn, connected by line to show value over continuous period, tracking trend or patterns. Not worry about sheer volume of your data, focus on the purpose and reader's attention.
- Small multiple: same frame of reference of all multiple graphs used.
- Circular Line chart: wrap in circle, show cycle of the graph that has pattern in a year.
- Slope chart: display data on one time to another, normally compare two points in time. Paired bar chart can also use in this case, but the slope is the better and easier to see.
- Stripes chart: feature heat map where bars present a data point in colors, not accurate.
- Bump chart: "power-up slope chart". Width of ribbon are scaled to data values, more flowing and organic looks.
- Flow chart: steps, command, information over time.

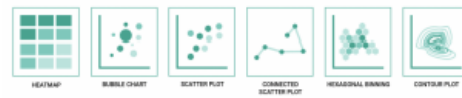
## 3. Distribution & Relation: data has no pattern or without explicit split, a.k.a. random.

### DISTRIBUTION



*When the data has no regular pattern or spread out on an axis with no explicit*

### CORRELATION



*chart type is very appropriate*

- Histogram: most basic for distribution graph. Samples are divided into bins, height show number of sample within these intervals. Bins too wide, hide pattern in distribution. Bins too narrow, may obscure the shape of distribution.
- Scatter plot: show correlations between numeric values and find patterns. Usually use first for data exploration. Give more insight to the data. Similar types are categorical scatter plot, bubble plot, quadrant plot, connected scatter plot, etc.
- Pyramid chart: show the population-related data or similar to that such as income-based data, look like diverging bar chart. May confused with positive/negative data, use lollipop or dot plot instead.

## 4. Geospatial: Show data on a map or spatial area.

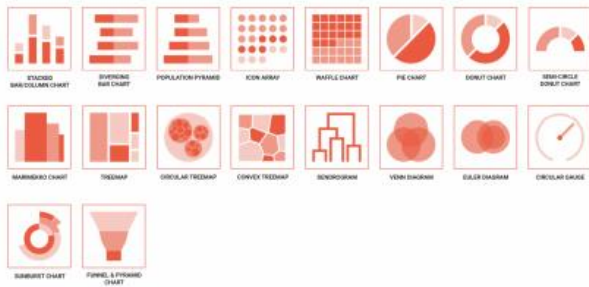
### GEOSPATIAL & OTHER



- Advantages to plot geographic data on a map or any data that related with the location.
- However, the map isn't the best way to communicate the message. Some map is simply created due to the data has the geographic data, but the map is not the best medium for that content. Map doesn't need to be use at all the time. It may consume the brain power and energy of reader too much.

## 5. Part-To-Whole: key message is about how data is being a part of whole picture

### PART-TO-WHOLE & HIERARCHICAL



- Pie chart: Simplest, hard to perceive the difference among slices. Please avoid. More accurate by labeling the data. Donut chart is just pie with holes.
- Tree map: divide rectangles into groups to show hierarchy or part-to-whole relationship of the data. Same perception ranking with pie chart.
- Nightingale chart: stacked column wrapped around a circle. Get more attention while show some part-to-whole relationship of the data.

## 6. Uncertainty Data

- Confidence Interval Chart: use line or shaded area to show range or amount of uncertainty. Usually have central estimate, upper, and lower confidence interval. Shaded/filled area is the more preferred one since it is easier to see.
- Fan chart: show different confidence interval with color or saturation of shaded area.
- Candlestick chart: similar variation, can show error, use in stock market.
- Ridgeline plot: similar to histogram, show distribution of each data, Series of histograms or density plot shown for different group aligned along the same horizontal axis and presented with a slight overlap along vertical axis.
- Strip plot: similar to ridgeline, you plot in separate axis, you have more information, plot on to parallel axis, separate them is better.
- Bee swarm plot: jitter data a little between line, make it easier to reader. off axis a little.

## 7. Relationship Data

- Parallel coordinate plot: shows correlations between two or more variables across multiple vertical axes.
- Radar/Spider/Star chart: line wraps around a circle, each corner present a category. More professional and easier to read. However, you need to label each corner with categories.
- Chord diagram: show relationship in a circle, famous way to show relationship, but you don't exactly know how many relationship it is. You can also focus on some part.
- Arc chart: chord chart but on a single line. Often show thickness as volume of the data.

## 04 – Visual Coding

- Decide how to encode data and expand dimensions of the data. Each dot on the XY graph can only display 2 data.
- Example of how we can encode data to the points
  - o X and Y axes can represent the main message of the graphs.
  - o Angle is an x-axis of the circular form.
  - o Size can give visual comparison. → Can represent the value of the dot.
  - o Area is good indicator
  - o Distance/Radius can be used to represent quantitative properties.
  - o Color is easy for the eyes to distinguish data at a glimpse.



- Stroke color can be used but not popular
- Shape/Icon is for categorical data.
- Label is the simplest way.
- **BE CAREFUL!** When you are using more than one variable to convey data that might cause confusion to the reader. It may overload their memory. It is better to limit to 8 parameters.
- When you use visual coding, it is important to have LABEL because it is not a standard graph.

## 05 – Storytelling with Data

- *There is a story within the data that want to be told.* Story is strong in terms of history and keep telling story down to generation by time. When you want to tell something, make the fact into the story, so that it will get to the listener head and stays in their head for a very long time. (Ex. you remember the movie better than anything else.)
- **Constructing the story**
  - Simple story structure: beginning, middle (whole story), ending.
  - **The beginning:** You want to tell the environment, data, axis, collection, etc. Just to get everyone on the common ground. Answer: Why should I pay attention? What is it for me.
    - **Main questions:** settings, main characters, imbalance, balance, solution
    - Another way to think about balance/imbalance is problem/solution
  - **The middle:** “what could be” with the goal of convincing audience of the need for action.
    - Further develop situations by covering relevant background information such as incorporate context, give examples, include problem-demonstrated data, discuss potential options of the problem, show benefits of solutions, make it clear to your audience why they are in the unique position.
  - **The ending:** end with call for action, want audience to act at the end of the story.
- **Power of repetition:** more information is repeated, more likely to ended up in LTM or retained.
- **Horizontal Logic:** It is important to have action titles.
  - Having executive summary slides with each bullet correspond to subsequent slides. So audience knows what to expect and taken through detail.
- **Vertical Logic:** information slides is self-reinforcing. Content reinforces title, vice versa.
- **5 Steps of storytelling with data**
  - Understand context → Choose appropriate display → Eliminate clutter → Draw attention where you want it → Think like a designer → Tell a story.

## 06 – Interactive Data Visualization

- In scholar purposes, it is recommended to use standard graph since it requires most accurate perceptions of the data.
- Dynamic data visualization: Use when data always change, lots of data over time, stimulating user engagement, or excite audiences. Static graphs may not be enough to convey all information inside it. Even single image cannot deliver full contents because static only captures a moment in time of those data.

## 07 – Dashboard Design

- “One view to rule them all”. Mostly it requires a simple view to present all information, show trends and risky area, updates users what happened. A dashboard is **a glance preview** of the most crucial information for users. Easy to navigate directly.
- We can have dashboard show on multiple screens but not ideal, usually it shown on one screen.

- War room refers to the room with many screens and show a lot of dashboards and monitor data in the area. Usually used in crisis response center or war time.
- Dashboard used to describe to non-tech and tech users. It should be simple and easy enough to the simple user.

## Great Dashboard

- **Clear, Intuitive, Customizable**
  - o Communicate information quickly, display information clearly and efficiently, show trends and changes over time, easily customizable, widgets and components are presented effectively in a limited space.
- **Everything in One Click Away**
  - o All information is immediately accessible, prioritized, displayed clearly in the visual hierarchy on the screen, coherent design.
  - o Improve the usability by using filters by labels, categories, and KPIs.

## 10 Steps of Designing Good Dashboards

1. Define the purpose of the dashboard: analytical & strategic or operational & tactical
  - a. Operational dashboard: impart critical information quickly. Present data deviations to the user quickly, show current resources and their status. Time-sensitive.
  - b. Analytical dashboard: provide at glance information for analysis and decision making. Less time-sensitive and not immediate action. Help data make the best sense, analyze trends and drive decision making.
2. Focus on your audience : know who use the dashboard to create best tool for them.
  - a. User roles: different roles need different dashboard. Each one should get a unique one.
  - b. Personalization vs customization:
    - i. Personalization: done by the system, identify user, deliver context and functionality that matches their roles.
    - ii. Customization: done by the user, make change to meet their specific needs.
3. Choose relevant Key Performance Indicators/KPIs: Help to shape the dashboard direction.
  - a. KPI measure and evaluate the business activities' performance. Can be high-level with global performance or low-level with individual or process performance.
4. Choose the right representation for the data: right chart types or data misinterpretation.
5. Define the layout and flow: alignment and consistency to create a structure of the design.
  - a. Keep this in mind: top left corner will get more attention, try to position key information from left to right, and top to bottom.
  - b. Create a continuous flow layout for easier scanning across the dashboard if there are dependencies that will affect the decision making.
6. Use building blocks with consistent structure: work with widgets. Use cards, they are infinitely manipulatable, good choice for responsive design.
7. Don't hide information or rely on interactions too much: surface information at glance, don't rely on scrolling or many interactions. It may overwhelm users.
8. Don't try to place all the information on the same page: don't demand too much from users and don't overwhelm with data. Maximum of 5-7 different widgets per view.
9. Prioritize simplicity: chart options are useful only when there is a reason to apply them.
10. Learn from good examples: look at good example about how they present the data such as medical or stock dashboard.