

HORIZONS-25-A3-HW

ASHOKA HORIZONS : APPLIED DATA SCIENCE WITH ML AND AI

Name - Tushar Joshi / Week#3 / Assignment-3

Part 1: Fundamentals of Visualization (20 Points)

1. Professor Edward Tufte emphasizes that useful data visualizations consist of "complex ideas communicated with clarity, precision, and efficiency."

(a) Explain what each of these three terms (clarity, precision, efficiency) means in the context of data visualization. (6 points)

In data visualization: -

- **Clarity** means the visualization should be easy to understand without extra effort.
- **Precision** means the information shown should be accurate, with labels and values correctly represented.
- **Efficiency** means the graph should deliver its message quickly without unnecessary decoration.

(b) Provide a hypothetical example of a visualization that might lack one of these qualities and explain how it could be improved. (4 points)

- A bad example might be a 3D pie chart showing election results. It may look cool, but it can distort angles and make the proportions unclear. Changing it to a simple 2D bar chart would improve clarity and precision.

2. Hal Varian, Google's Chief Economist, stated that the ability to "take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill."

(a) Why is the skill of data visualization particularly important in an era of "essentially free and ubiquitous data"? (5 points)

- Since data is everywhere and often free now, the real challenge is making sense of it. Visualization helps us understand and communicate insights quickly, making it a vital skill.

(b) How does effective visualization help in "extracting value" from data? (5 points)

- Since data is everywhere and often free now, the real challenge is making sense of it. Visualization helps us understand and communicate insights quickly, making it a vital skill.

Part 2: Visual Encoding and Perception (25 Points)

3. Explain the difference between "Exploratory Visualization" and "Explanatory Visualization."

(a) Describe a scenario where you would primarily use exploratory visualization. What would be your goals? (4 points)

- Exploratory visualization is used when you're still trying to understand your data. For example, I might use a scatter plot while analyzing survey responses to look for any unexpected patterns.

(b) Describe a scenario where you would primarily use explanatory visualization. What would be your goals? (4 points)

- Explanatory visualization is used when you already have a message and want to tell it clearly. For example, a chart showing that women scored higher than men in a test is made to convince or inform others.

4. Visual channels (like position, length, color, shape) are used to encode data.

(a) According to Tamara Munzner's effectiveness ranking (discussed in the lecture slides, slide 32), which visual channels are generally most effective for encoding quantitative (magnitude) data? Why do you think this is the case? (5 points)

- According to Tamara Munzner, position and length are the most effective for showing magnitude because our brains can judge spatial relationships more accurately than color or area.

(b) What are "identity channels" and when might they be more appropriate than "magnitude channels"? Give an example. (4 points)

- Identity channels like shape or color hue are better for distinguishing categories, like different departments in a company. They're better than bar heights when exact quantities aren't important.

5. Define "expressiveness" and "effectiveness" in the context of visual encoding. Why are both important for creating good data visualizations? (8 points)

- **Expressiveness** means the chart should show all the information in the data, and nothing extra.
- **Effectiveness** means it should show that information in a way that's easy for the viewer to understand. - Both matter because a good chart needs to be both correct (expressive) and clear (effective).

Part 3: Narrative, Color, and Design (30 Points)

6. Choose either William Playfair's charts or Florence Nightingale's "Diagram of the Causes of Mortality."

(a) Briefly describe the visualization. (3 points)

- It was a circular chart that showed causes of death during the Crimean War.

(b) What was the primary message or story this visualization was trying to tell? (4 points)

- The main message was that most soldiers were dying from disease, not battle wounds.

(c) How did its visual design contribute to effectively conveying this narrative? (3 points)

- Its design used colored wedges to show the number of deaths per cause, making the disease impact visually obvious and persuasive.

7. Explain the difference between qualitative, sequential, and diverging color palettes.

(a) For each type of color palette, describe a type of data it would be most suitable for visualizing and provide a hypothetical example. (6 points)

- **Qualitative:** Used for categories, like different fruits in a pie chart.
- **Sequential:** Used for numerical scales, like population density heatmaps.
- **Diverging:** Used to show deviation from a midpoint, like temperatures above and below average.

(b) Why is it important to consider color-blindness when choosing colors for data visualization? What is one strategy to make your color choices more accessible? (4 points)

- Color-blindness matters because many people can't distinguish red from green. One way to fix this is to use colorblind-friendly palettes (like ColorBrewer) or include patterns/symbols.

8. Before creating any data visualization, it's crucial to plan. List and briefly explain three key questions you should ask yourself during the planning phase of a data visualization project, as discussed on slide 34 ("Asking the Right Question"). (10 points)

- 1. Who is your audience? Are they experts or general readers? This changes how much detail or explanation you include.
- 2. What are you trying to show? Is it a comparison, trend, or distribution? This helps choose the right chart.
- 3. How will you show it? Pick the right chart type and visual encoding (e.g., bar chart for comparison, line chart for trends).

Part 4: Dynamic Visualization and Creative Coding (25 Points)

9. The lecture distinguished between static and dynamic visualizations.

(a) What are some potential advantages of using dynamic or interactive visualizations (e.g., HTML+JS based) over static ones (e.g., .png, .pdf)? (5 points)


- Dynamic visualizations let users interact with the data—hover, filter, or zoom. They work better for large or complex datasets that can't fit into one static image.

(b) Describe one specific type of interactivity (e.g., tooltips, filtering, zooming) and how it can enhance a user's understanding of complex data. (5 points)

- **Filtering** is useful when you want to let users focus on a category, like showing sales data by region. It simplifies the view without losing the underlying dataset.

10. The lecture introduced "Creative Coding."

(a) In your own words, what do you understand "creative coding" to mean, especially in relation to data visualization? (5 points)

- 
- Creative coding means using programming not just to solve problems, but to create visuals or art—like painting with code. In data viz, it's using code to make charts that are more interactive, artistic, or meaningful.

(b) Search online for an example of a data visualization that you would consider an example of "creative coding" (it doesn't have to be made with P5.js). Provide a link to it (if possible) or describe it, and briefly explain why you find it compelling or creative. (10 points)

- One great example of **creative coding** is "[People of the Pandemic](https://datastori.es/161-people-of-the-pandemic-with-shirley-wu/)" by Shirley Wu and Moritz Stefaner — a data-driven interactive simulation of COVID-19 spread:

<https://datastori.es/161-people-of-the-pandemic-with-shirley-wu/>

- I find it compelling because rather than just showing static charts, it lets users interact with and explore the pandemic's impact through a simulation platform built with code. This pushes beyond basic visualization: it blends storytelling, exploration, and interactivity.

