

# WS #3 - Tufte

Wednesday, September 4, 2024

DS 002R - Jo Hardin

Your Name: \_\_\_\_\_

Names of people you worked with: \_\_\_\_\_

Discuss with your partner(s) the different moving pieces of this class. Do you have any questions about what assignment to do, what platform to pay attention to, which emails to read, where to submit the assignments, what to do to be prepared for each class, or how to participate when you have no clue what is going on in class? Write down something logistical that still seems daunting, or report that you are feeling comfortable with the logistics.

**Task:** Consider the two examples from Tufte (1997) “Visual and Statistical Thinking: Displays of Evidence for Making Decisions.” For each of the two examples, use the graphical theory we’ve covered to assess the plots (see hints on back, and feel free to write on the back).

## 1. Cholera

- Provide one strength of the cholera visualization that helped scientists get to the root of the problem? Explain and refer to some aspect of theory.
- Provide one aspect of the visualization that clouded the story or made the solution harder to understand / communicate? Explain and refer to some aspect of theory.

## 2. Challenger

- Provide one aspect of the Challenger visualizations that made the story harder to understand / communicate? Explain and refer to some aspect of theory.
- Can you come up with one aspect of the visualization that were creative or helpful? Explain and refer to some aspect of theory.

**Nolan and Perrett (2016)**

1. Make the data stand out
2. Facilitate comparison
3. Add information

**Yau (2013)**

- Visual cues
- Coordinate system
- Scale
- Context

**Class notes give advice on:**

- Think carefully about the basic plot
- Eliminate superfluous material
- Facilitate comparisons
- Choose the scale well
- Make the plot information rich
- Create good captions
- Follow good plot making practice

## Solution:

### 1. Cholera

- What are some of the strengths of the cholera visualization that helped scientists get to the root of the problem? Explain.
  - Snow thought carefully about how to present the data. Instead of simply looking at the data as counts or frequencies, he looked at the death *spatially* - on a map of the area. [Coordinate system]
  - Snow was able to **compare** different groups (brewers and employees at a work-house). [Facilitate comparisons]
- Were there aspects of the visualization that clouded the story or made the solution harder to understand / communicate? Explain.
  - Some aspects of the graph can be scrutinized in terms of choices: Did the individuals die at the place on the map? Live at the place on the map? Which (types of) individuals were missing from the graph? Missing at random? What decisions did he make in creating the graph (axes, binning of histogram bars, time over which data are plotted, etc.) that change the story needing to be told? [Lack of context]
  - No population density or time was included in the graphic. [Poor facilitate comparison with respect to time]
  - Would be visually appealing to highlight the Broad St. pump. [Didn't make the data stand out]

### 2. Challenger

- What are some of the aspects of the Challenger visualizations that made the story harder to understand / communicate? Explain.
  - Time and temperature were never clearly displayed in comparison to one another. [Poor facilitate comparison]
  - The engineers who understood the problem created tables and engineering graphs which were
    - \* Not visually appealing. [Poor thinking about the basic plot]
    - \* Not decipherable to the layman (e.g., “At about 50° F blow-by could be experienced in case joints”) [Poor captions]
    - \* There was also no authorship (reproducibility!). **Figures should always have both accountability and reproducibility.** [Poor plot making practice]
  - The information provided included very relevant points (about temperature) and superfluous information unrelated to temperature. The univariate analysis was insufficient because the story the data were trying to tell was about the **bivariate relationship between temperature and o-ring failure.** [Did not eliminate superfluous material]

- Missing data created an illusion of *lack of evidence*, when in fact, the true story was quite strong given the full set of information. (92% of the temperature data was missing from some of the most vital tables.) [Did not facilitate comparisons]
  - Anecdotal evidence was misconstrued: SRM-15 at 57F had the most damage, but SRM-22 at 75F had the second most damage. [Poor context]
  - In the end, the shuttle launched on a day which was an extrapolation from the model suggested by the data. They had never launched a shuttle at temperatures of 26° – 29°F.
  - Tufte goes on to describe many ways which the final presentation by the engineers to the administrators was inadequate: disappearing legend (labels), chartjunk, lack of clarity depicting cause and effect, and wrong order. [Poor basic plotting, superfluous material, not information rich, poor captions, poor context, poor plot making practice]
- Were there aspects of the visualization that were creative or helpful? Explain.
    - Not really. One thing that the engineers *tried* to do was to communicate as complete a picture as possible. They provided too much information which made it impossible for lay people to understand. [Tried to communicate context]