Data Visualization: Design Principles and Processes

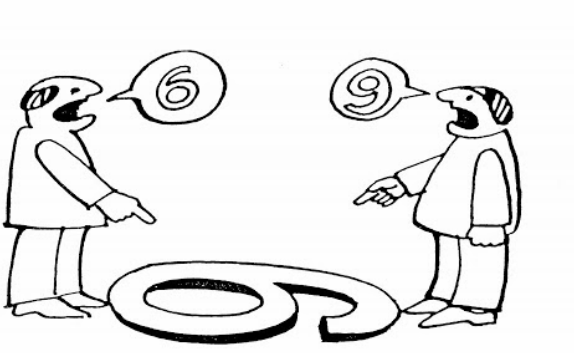
SMM635 - Week 1

Prof. Simone Santoni

Bayes Business School

What is Good Data Visualization?

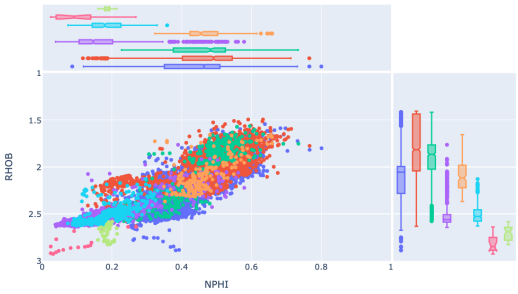
The fundamental question every data analyst must ask



"Excellence in statistical graphs consists of complex ideas communicated with clarity, precision, and efficiency." - Edward Tufte

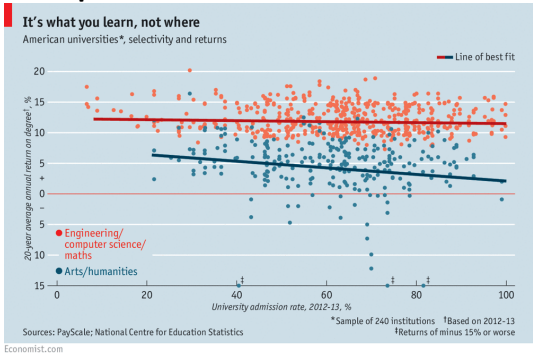
Tale of Two Visualizations

Example A: Technical Plot



Shows data relationships
Cluttered interface
Distracting elements

Example B: The Economist



Clean, focused design
Clear narrative
Professional aesthetics

Tufte's Principles of Graphical Excellence

Excellent visualizations should:

- **Show the data** clearly and accurately

Figure 1: Edward Tufte

Tufte's Principles of Graphical Excellence

Excellent visualizations should:

- **Show the data** clearly and accurately
- **Induce thinking** about substance, not methodology

Figure 1: Edward Tufte

Tufte's Principles of Graphical Excellence

Excellent visualizations should:

- **Show the data** clearly and accurately
- **Induce thinking** about substance, not methodology
- **Avoid distortion** of what the data reveal

Figure 1: Edward Tufte

Tufte's Principles of Graphical Excellence

Excellent visualizations should:

- **Show the data** clearly and accurately
- **Induce thinking** about substance, not methodology
- **Avoid distortion** of what the data reveal
- **Present many numbers** in a small space

Figure 1: Edward Tufte

Tufte's Principles of Graphical Excellence

Excellent visualizations should:

- **Show the data** clearly and accurately
- **Induce thinking** about substance, not methodology
- **Avoid distortion** of what the data reveal
- **Present many numbers** in a small space
- **Make large datasets coherent**

Figure 1: Edward Tufte

Tufte's Principles of Graphical Excellence

Excellent visualizations should:

- **Show the data** clearly and accurately
- **Induce thinking** about substance, not methodology
- **Avoid distortion** of what the data reveal
- **Present many numbers** in a small space
- **Make large datasets coherent**
- **Encourage comparison** between data elements

Figure 1: Edward Tufte

Tufte's Principles of Graphical Excellence

Excellent visualizations should:

- **Show the data** clearly and accurately
- **Induce thinking** about substance, not methodology
- **Avoid distortion** of what the data reveal
- **Present many numbers** in a small space
- **Make large datasets coherent**
- **Encourage comparison** between data elements
- **Reveal data** at multiple levels of detail

Figure 1: Edward Tufte

Tufte's Principles of Graphical Excellence

Excellent visualizations should:

- **Show the data** clearly and accurately
- **Induce thinking** about substance, not methodology
- **Avoid distortion** of what the data reveal
- **Present many numbers** in a small space
- **Make large datasets coherent**
- **Encourage comparison** between data elements
- **Reveal data** at multiple levels of detail
- **Serve a clear purpose**: description, exploration, or decoration

Figure 1: Edward Tufte

Tufte's Principles of Graphical Excellence

Excellent visualizations should:

- **Show the data** clearly and accurately
- **Induce thinking** about substance, not methodology
- **Avoid distortion** of what the data reveal
- **Present many numbers** in a small space
- **Make large datasets coherent**
- **Encourage comparison** between data elements
- **Reveal data** at multiple levels of detail
- **Serve a clear purpose**: description, exploration, or decoration
- **Integrate** with statistical and verbal descriptions

Figure 1: Edward Tufte

The Power of "Show the Data" - Anscombe's Quartet

Four datasets with identical summary statistics

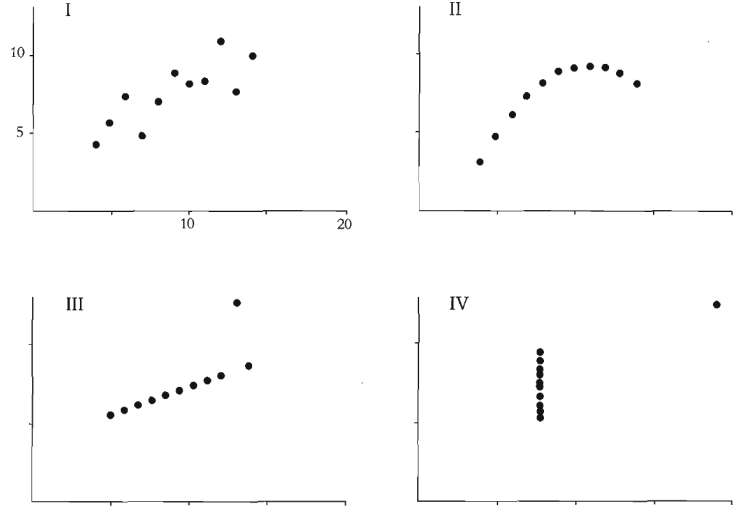
I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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

$N = 11$
mean of X's $= 9.0$
mean of Y's $= 7.5$
equation of regression line: $Y = 3 + 0.5X$
standard error of estimate of slope $= 0.118$
$t = 4.24$
sum of squares $X - \overline{X} = 110.0$
regression sum of squares $= 27.50$
residual sum of squares of $Y = 13.75$
correlation coefficient $= .82$
$r^2 = .67$

Same means, same correlations, same regression lines...

Anscombe's Quartet Revealed

...but completely different data patterns!



The Design Process Framework

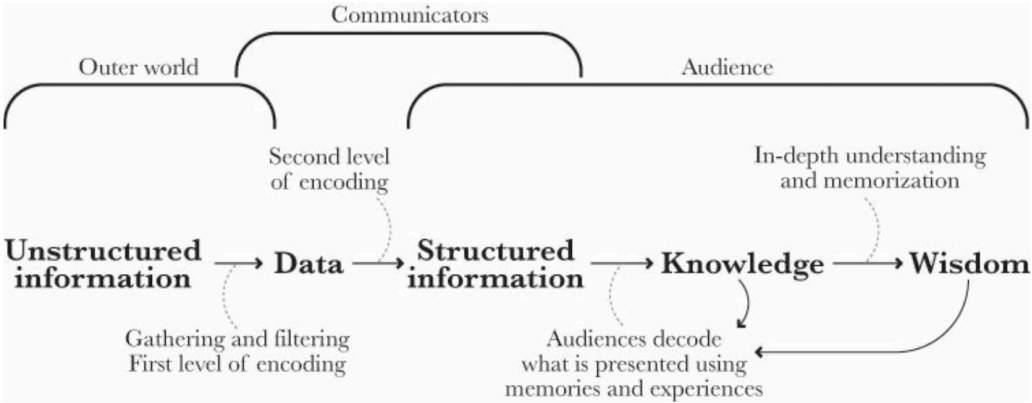
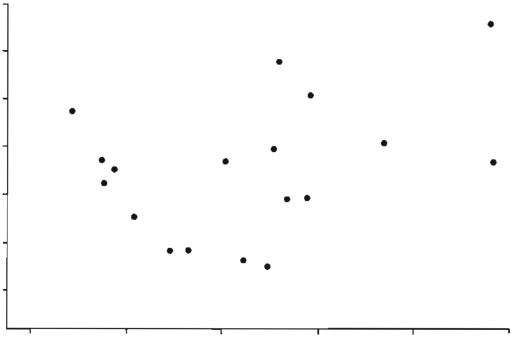


Figure 1.8. From reality to people's brains.

Source: Cairo, A. (2012). The Functional Art

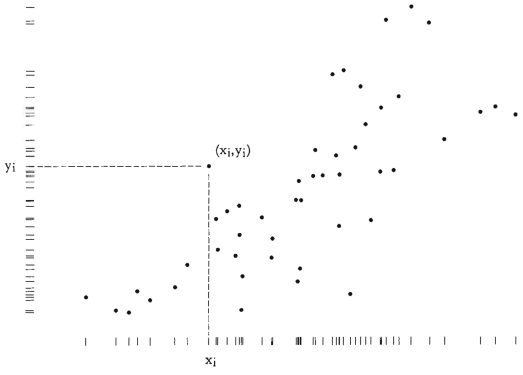
Design Principles in Action

Traditional Approach



Heavy gridlines, excessive decoration

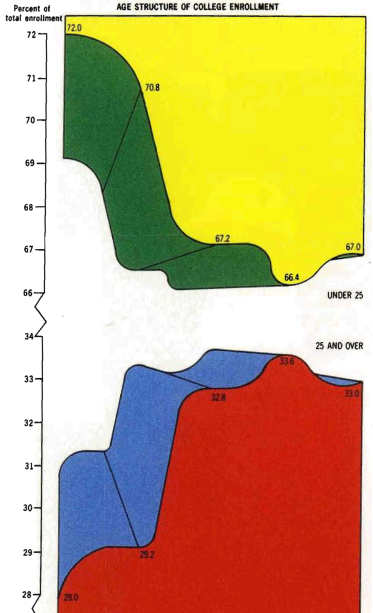
Tufte's Approach



Minimalist, data-focused design

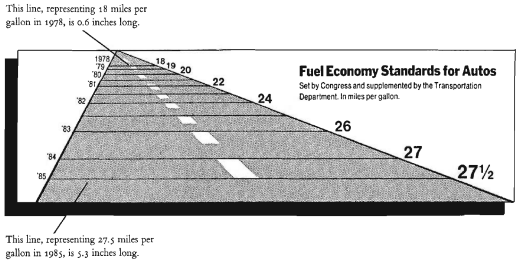
Principle: Maximize the **data-ink ratio** - every mark should represent data

Chart Junk - What Not to Do



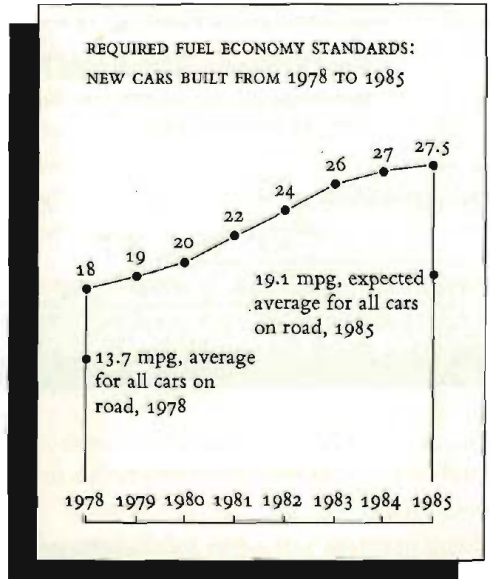
Before and After - Redesign Example

Before: Cluttered Design



Issues: 3D effects, poor labeling, distracting elements

After: Clean Redesign



Key Takeaways for Week 1

Your visualization design checklist

Purpose: Does your chart serve a clear analytical goal?

Data: Does your visualization accurately represent the data?

Clarity: Can viewers understand the message quickly?

Simplicity: Have you removed unnecessary elements?

Aesthetics: Is the design professional and appropriate?

Iteration: Have you tested and refined your design?

Remember: *Good visualization design is both art and science - it requires understanding your data, your audience, and your design principles.*

Next Steps

For next week: Read Tufte Chapter 1 and Cairo Introduction and Chapter 1

Practice: Complete the Data visualization and communication excercise

Resources

- Course GitHub: github.com/simoneSantoni/data-viz-smm635
- Design principles checklist in Moodle
- Office hours: Wednesdays 15:00-17:00