Agenda

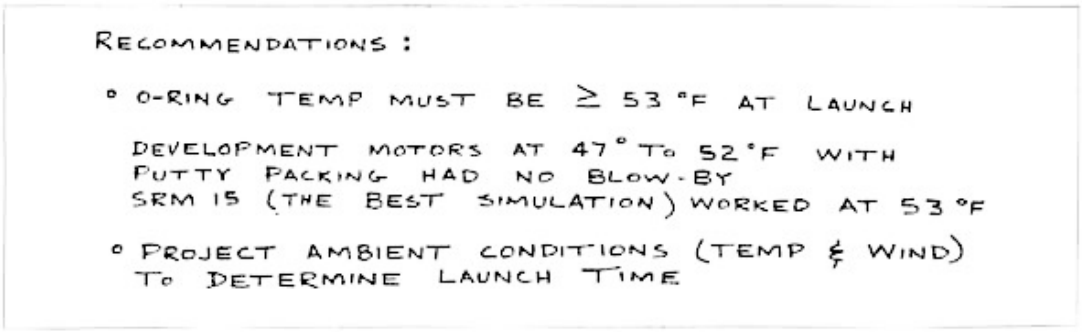
1. Challenger episode
2. Basic data graphics in R
3. Some inspiring and not so inspiring visualizations
4. Syllabus
5. Assignment #1 due Thursday (2/8) by noon.

Information Presentation & the Challenger episode On January 27th, 1986, engineers at Morton Thiokol, who supply Solid Rocket Motors (SRMs) to NASA for the space shuttles, recommended that NASA delay the launch of *Challenger*, due to concerns that the cold weather forecast for the next day's launch would jeopardize the stability of the rubber O-rings that held the rockets together. These engineers provided 13 charts that were reviewed over a two-hour conference call involving the engineers, their managers, and NASA. The engineers' recommendation was overruled due to a lack of persuasive evidence, and the launch proceeded on schedule. The O-rings failed in exactly the manner the engineers had feared 73 seconds after launch, the *Challenger* exploded, and all seven astronauts died.

- A selection of tables/graphics sent to NASA the night before

HISTORY OF O-RING TEMPERATURES
(DEGREES - F)

MOTOR	MBT	AMB	O-RING	WIND
QM-1	68	36	47	10 MPH
QM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH



Activity: Visual Explanations Identify three weaknesses in the "History of O-Ring Damage" data graphic. How could you improve this?

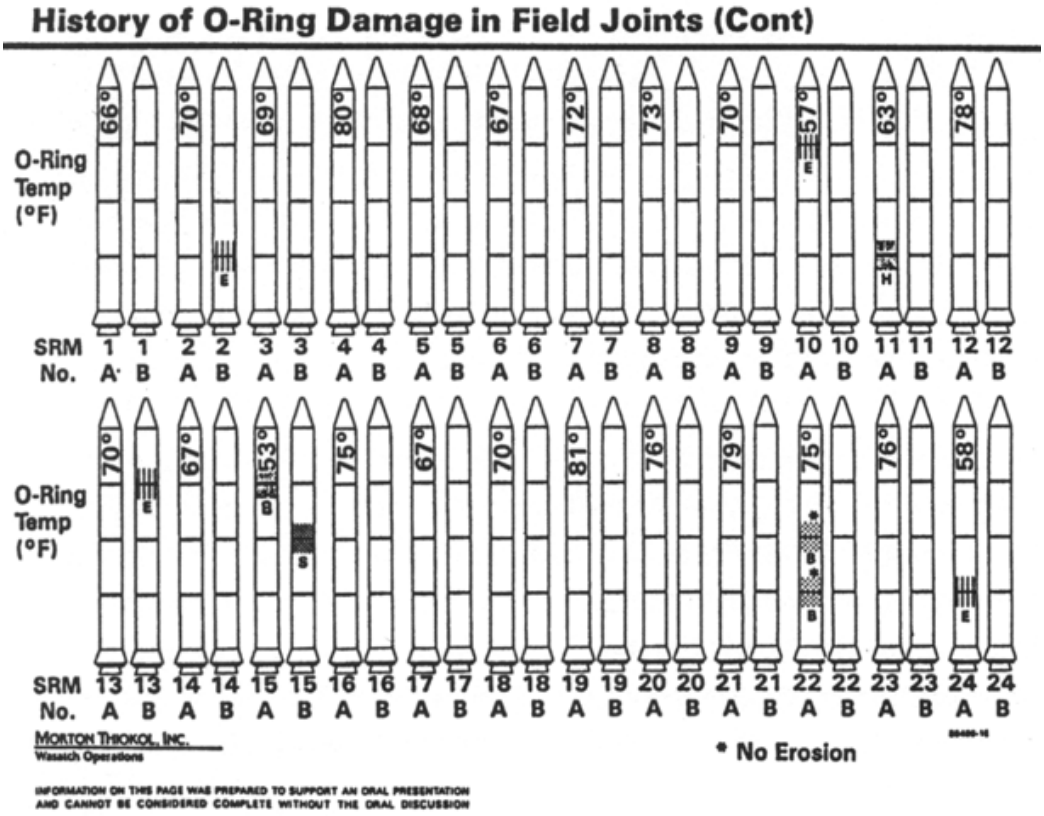
1.

2.

3.

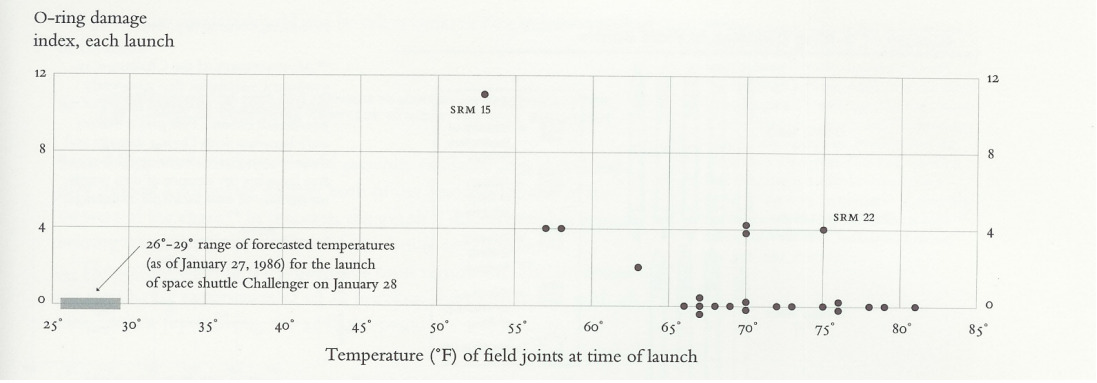
Graphical Practice Tufte provides a systematic study of statistical *data graphics*. This provides a bedrock for *data visualization*.

- Data graphics: instruments for reasoning about quantitative information. Focus on substance – not methodology & technique
- Quantifying visual information in data graphics:
  - Show the data: "A picture is worth 1000 words," so how many numbers are embedded in a data graphic?
  - Tables may be preferrable for $n \leq 20$
  - Efficiency & minimalism: the greatest number of ideas conveyed in the shortest time with the least ink in the smallest space
  - Avoid distortion to promote honesty:
    - visual representation should be proportional to numerical representation
    - Lie Factor $= \frac{\text{size of effect shown}}{\text{size of effect}}$
    - 2D image for a 1D quantity: number of variable dimensions $\leq$ number of data dimensions
- Clarity:
  - Labels should be informative, clear, honest, consistent: deception results from incorrect extrapolation of visual expectations
  - Annotate important elements right on the graphic
  - Decoration is pointless – chartjunk
- Show meaningful comparisons
  - Use multivariate displays



  - Add important visual signposts – *compared to what?*
  - Avoid selective reporting

- Timeline of developments in graphical practice:
  - 1100s: High-quality, accurate maps appear in China
  - 1850s: Abstract quantitative axes do not appear until mid-to-late 18th century!
  - 1854: Snow's cholera map (p. 24): uses clusters to inform location of diseased pump handle. Data science exercise with a happy ending.
  - 1869: Minard's March of Napoleon (p. 31): data map & time series with six variables
  - 1885: Marey's Paris train schedule (p. 41): time series that shows velocities & intersections. Creative, insightful, original visualization that reveals structure not obvious in data.
  - 1960s: Tukey codified statistical graphics
  - Data visualization is young! Contrast with mathematical graphing


O-ring damage
index, each launch
12
8
4
0
SRM 15
SRM 22
26°–29° range of forecasted temperatures
(as of January 27, 1986) for the launch
of space shuttle Challenger on January 28
25°
30°
35°
40°
45°
50°
55°
60°
65°
70°
75°
80°
85°
Temperature (°F) of field joints at time of launch