Introduction to data display and visualization
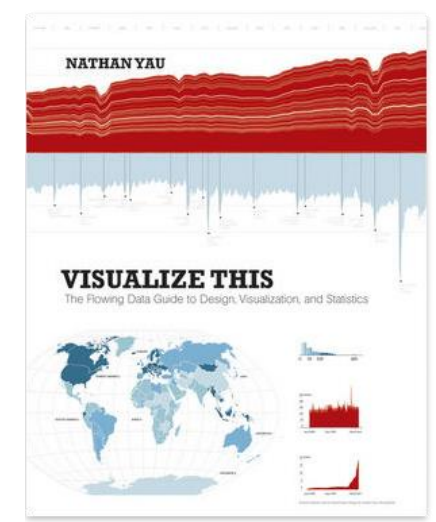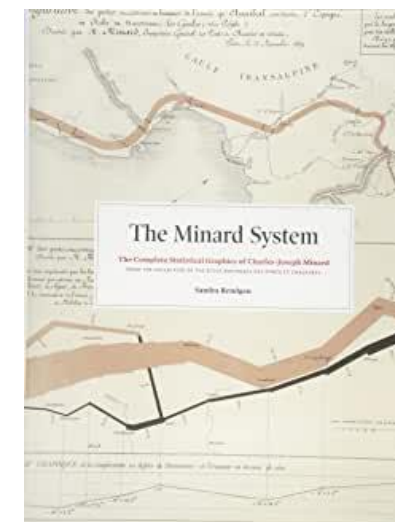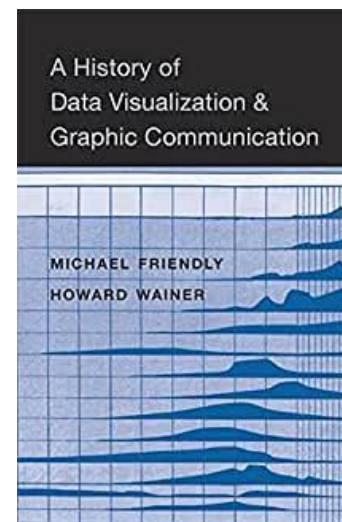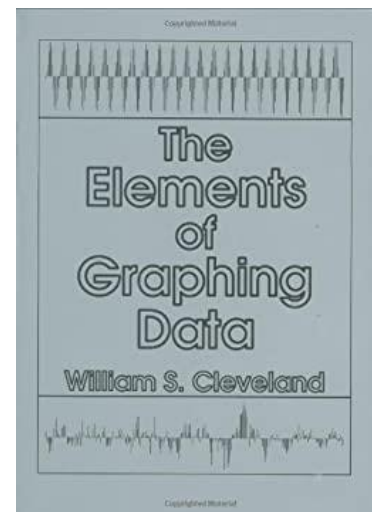
Omar Alonso

6-April-2023

Agenda

- High level introduction
- Main concepts
- Why this is important
- Grad level class at NEU
- Many books and online resources

Books
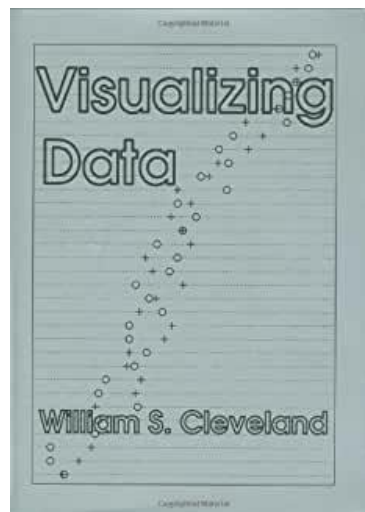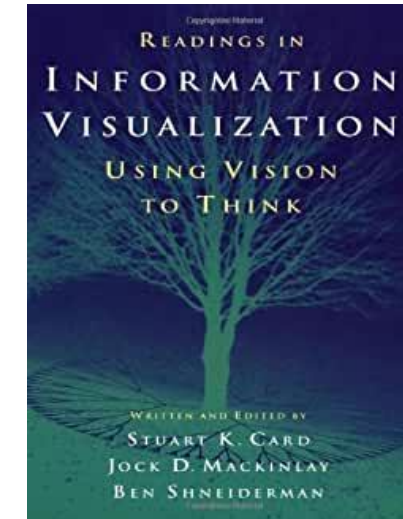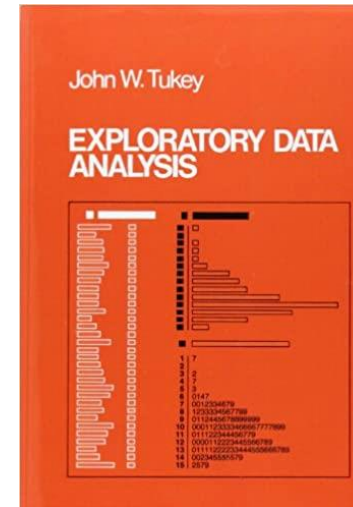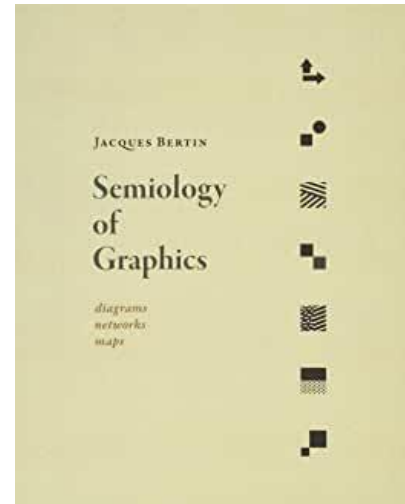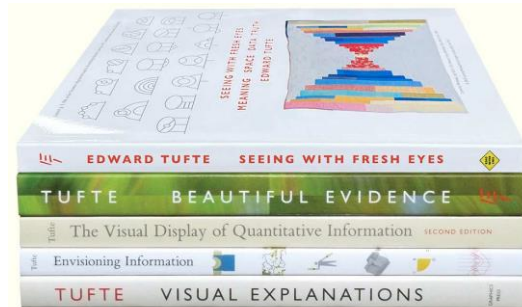
Why data viz is important?

- All technical documentation contains
  - Tables
  - Figures
  - Diagrams
- You need to communicate main findings and results
- Data presentation helps business decision
- Avoid visualization overload
- Lots of technologies
- Design for an audience

Graphical excellence

- Well-designed presentation of interesting data: a matter of substance, of statistics, and of design

- Complex ideas communicated with clarity, precision, efficiency

- Requires telling the truth about the data

Data

- Show the data
- Help the reader to think instead of methodology, graphic design, viz
- Facilitate comparison and to illustrate causality
- Serve a purpose
- Data collection
- Learn all you can about your data before any viz

Annotations

- Add labeled arrows
- Circle anomalies
- Explanations near the data
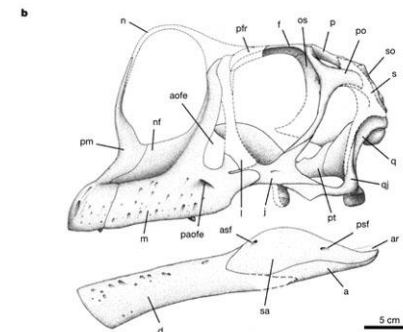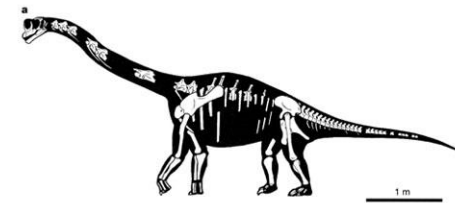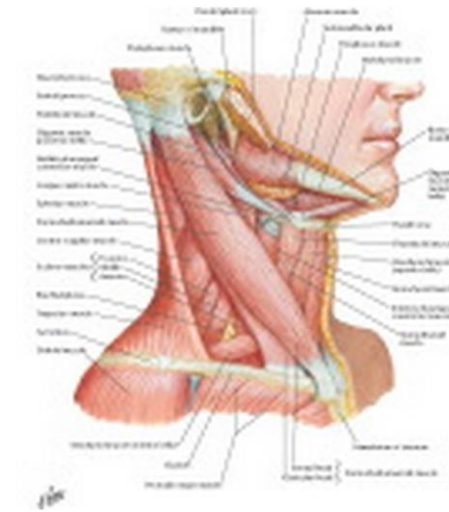
Figure 1 | *Europasaurus holgeri* gen. et sp. nov. a, Reconstruction of the skeleton (known parts shown in white) and outline (black) scaled to the largest individual (body length 6.2 m). b, Skull reconstruction based on several individuals but scaled to the holotype. Abbreviations: a, angular; aofe, antorbital fenestra; ar, articular; asf, anterior surangular foramen; d, dentary; f, frontal; j, jugal; l, lacrimal; m, maxilla; n, nasal; nf, narial fossa; os, orbitosphenoid; p, parietal; paofe, preantorbital fenestra; pfr, prefrontal; pm, premaxillary; po, postorbital; psf, posterior surangular foramen; pt, pterygoid; q, quadrate; qj, quadratojugal; s, squamosal; sa, surangular; so, supraoccipital.

Avoid clutter

- Present information without creating confusion
- Text needs to be easy to read
  - ADVANCED TOPICS IN DATA SCIENCE
  - Advanced Topics in Data Science
- Use colors carefully
- Don't obfuscate the data display

Graphical integrity

- Author names
- Dates
- Quote data in full context
- Show data variation, not design variation
- Data-ink ratio: proportion of a graphic's ink devoted to the non-redundant display of data info
  - Data-ink / total ink used to print the graphic
- Be honest

Minard's map of Napoleon's March
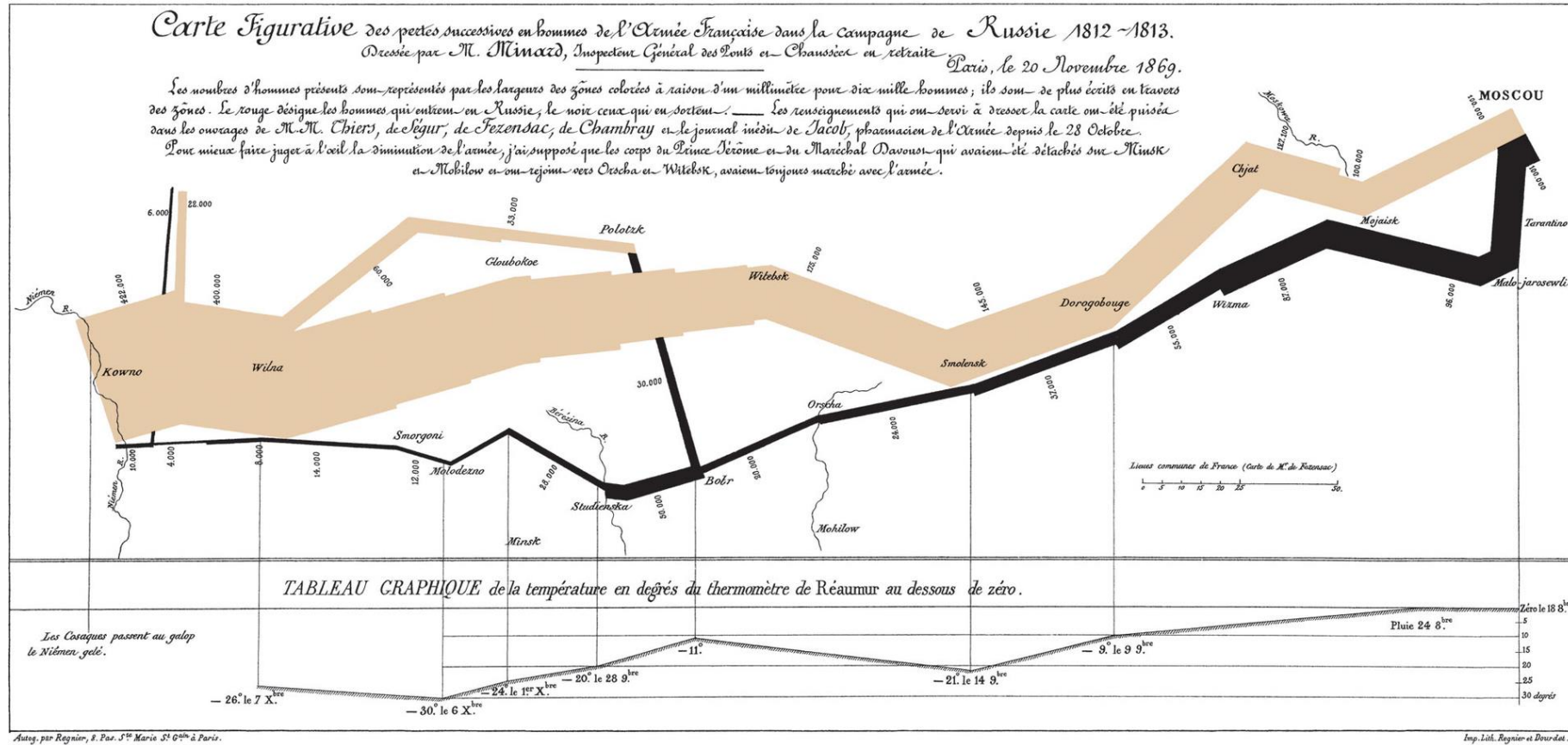
Size of the army
Location of 2-D surface by date (latitude & longitude)
Direction of movement
Date
Temperature

Charles Minard

- Many other maps
- Carte figurative
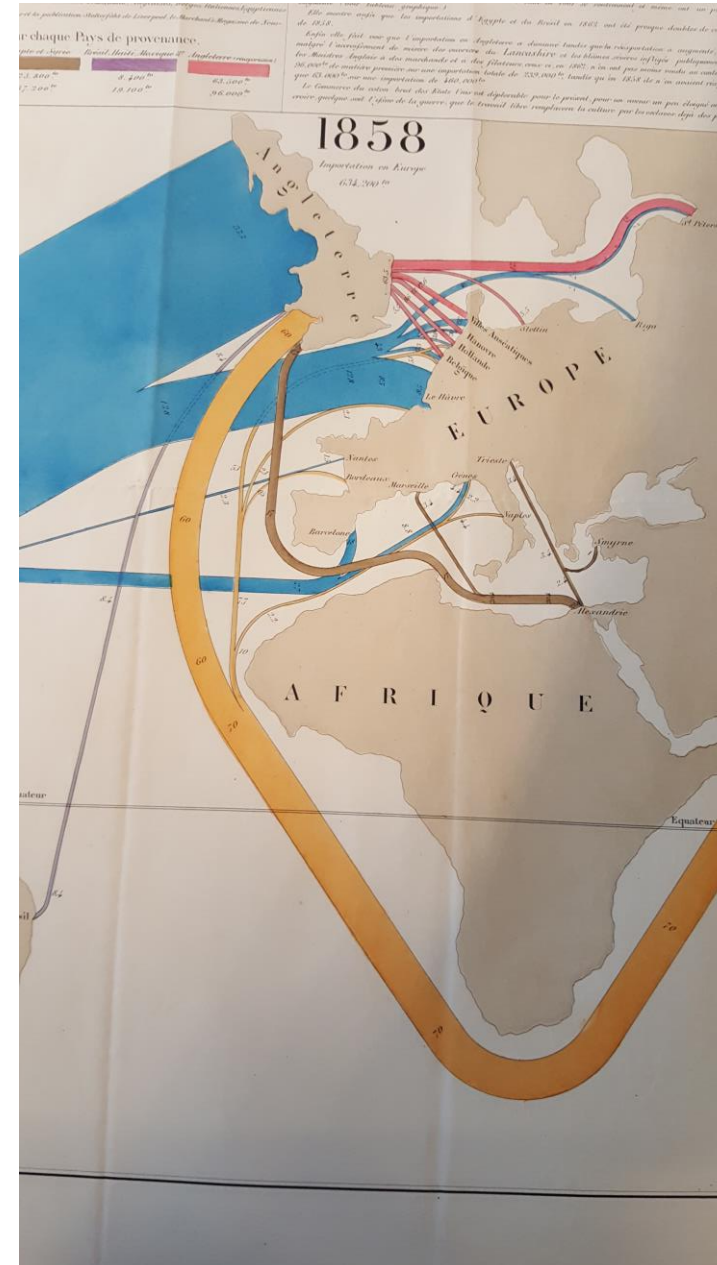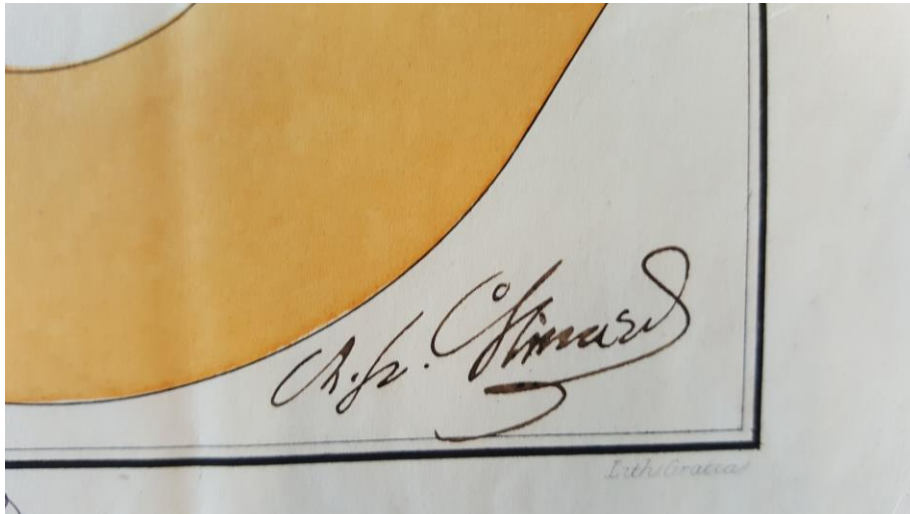
The decision to launch the Shuttle Challenger

- In 1986 the shuttle Challenger exploded because two rubber O-rings leaked

- The engineers who designed the rocket opposed launching

- Evidence presented was inconclusive

1.4 STATISTICAL REASONING. These data were graphed by space shuttle engineers the evening before the Challenger accident to determine the dependence of O-ring failure on temperature. Data for no failures was not graphed in the mistaken belief that it was irrelevant to the issue of dependence. The engineers concluded from the graph that there is no dependence.

1.5 STATISTICAL REASONING. The complete set of O-ring data is now graphed, including the observations with no failures. A dependence of failure on temperature is revealed.

Information Visualization

- Explore, present, describe meaning

- Communicating and clarifying ideas

- Components
  - Visual cues
  - Coordinate system
  - Scale
  - Context

Principles of graph construction

- Scale-line
- Data rectangle encloses data
- Data labels
- Make the data stand out
- Make captions informative and comprehensive
    - Describe everything that is graphed
    - Draw attention to the important features of the data
    - Describe the conclusions that are drawn from the data on the graph
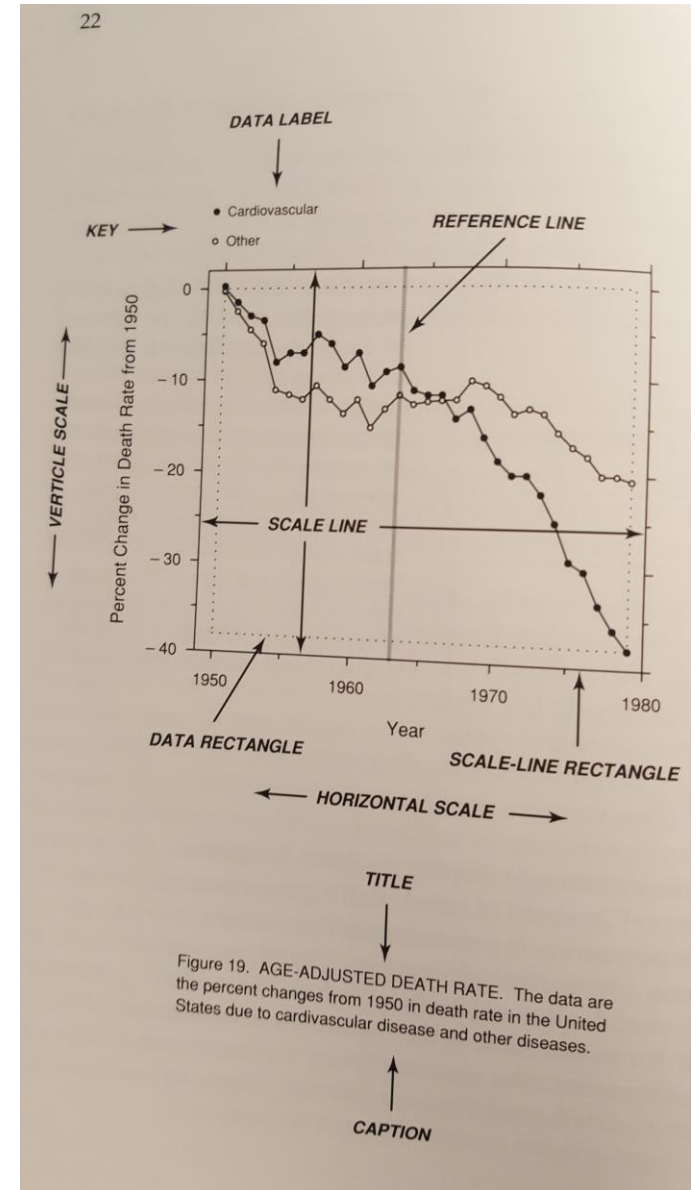
PowerPoint and presentations

- Tufte's critique
- PPT is presenter-oriented, not content-oriented, not audience oriented
- Replace PPT with word-processing software
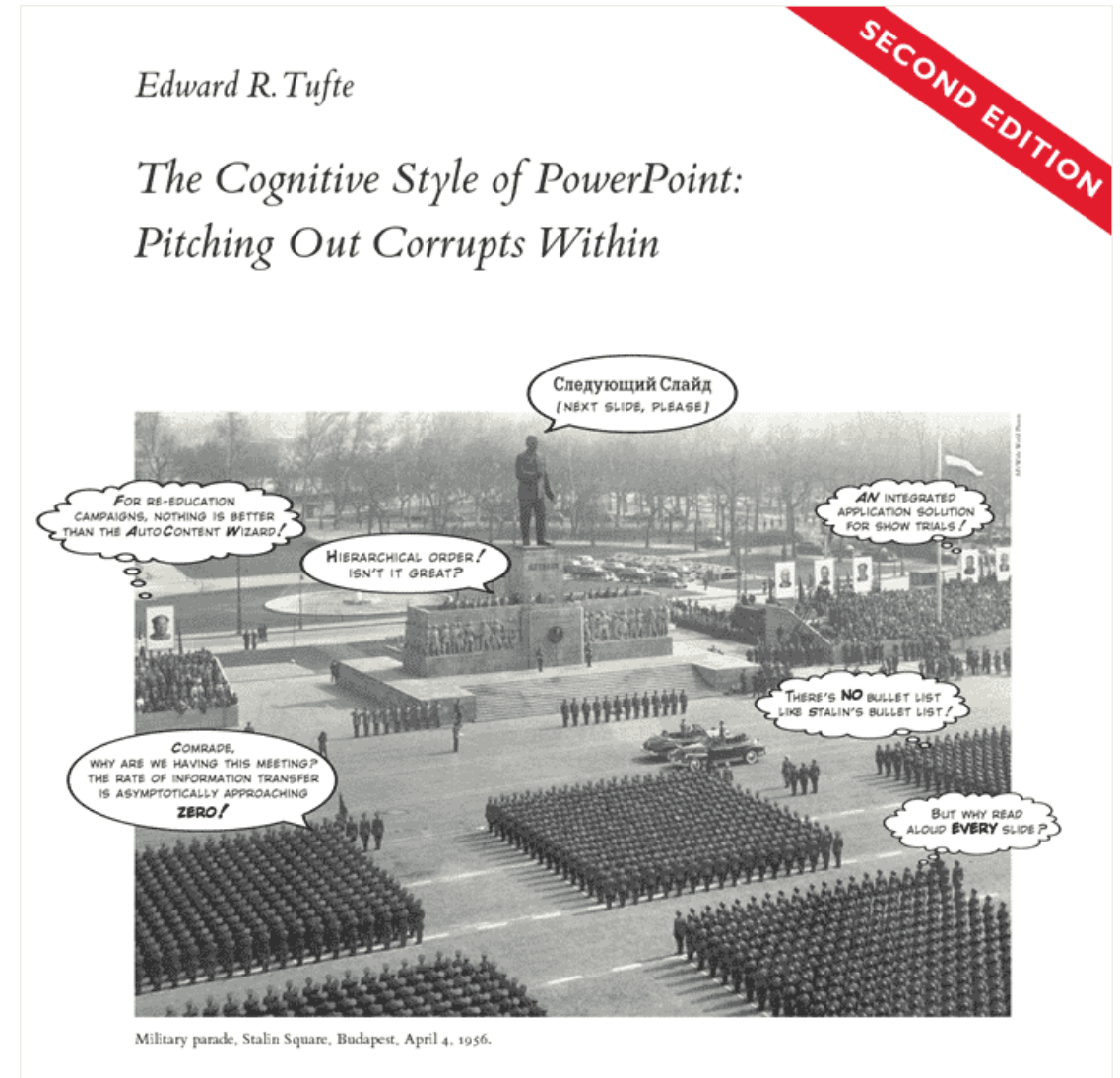
Assessing the quality & credibility of tech reports

- Does PPT cognitive style affect the quality of engineering analysis?
- Columbia shuttle
  - 82 seconds after liftoff, a 1.67 pound piece of foam broke and damaged the wing's thermal protection
  - Burned up during re-entry
- Some tools are better than others for engineering
- Tech reports > PPT

Review of Test Data Indicates Conservatism for Tile Penetration

- The existing SOFI on tile test data used to create Crater was reviewed along with STS-87 Southwest Research data
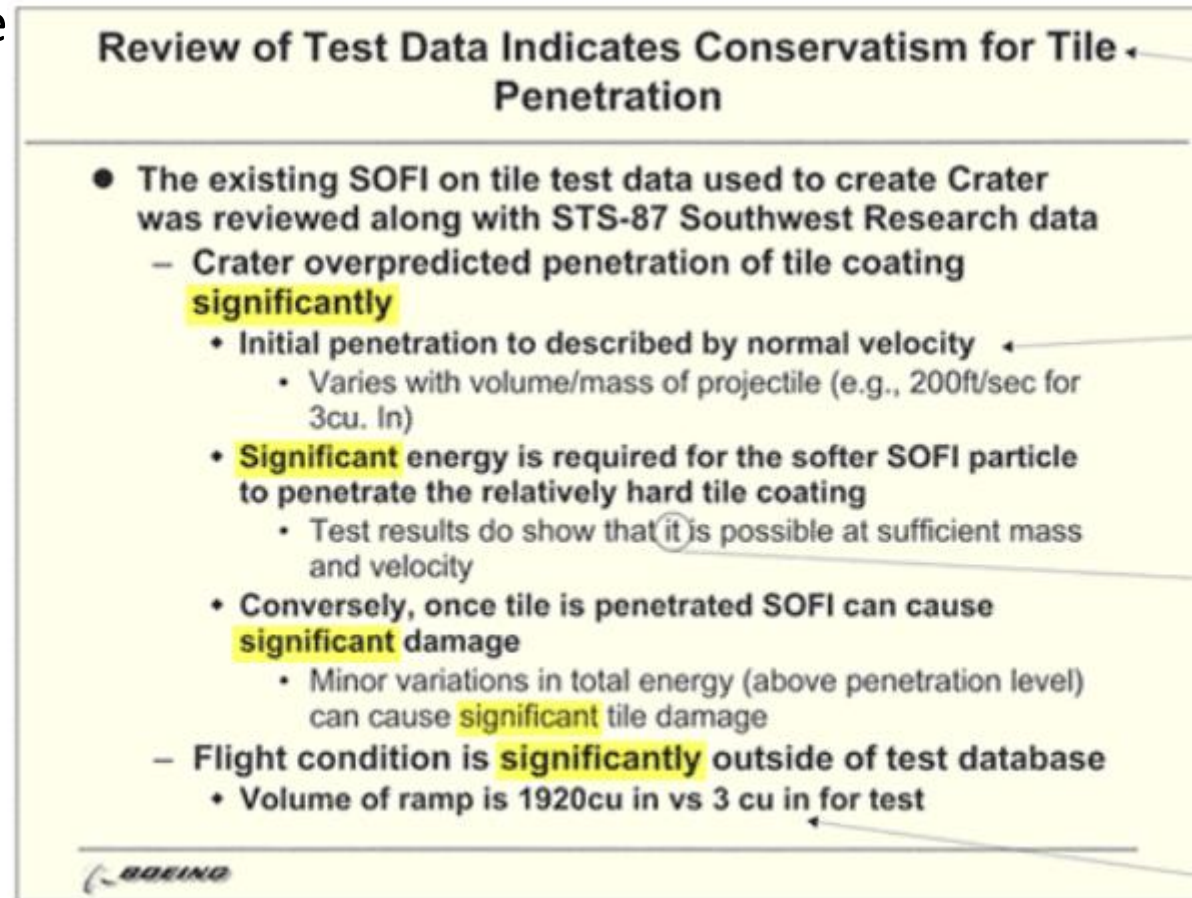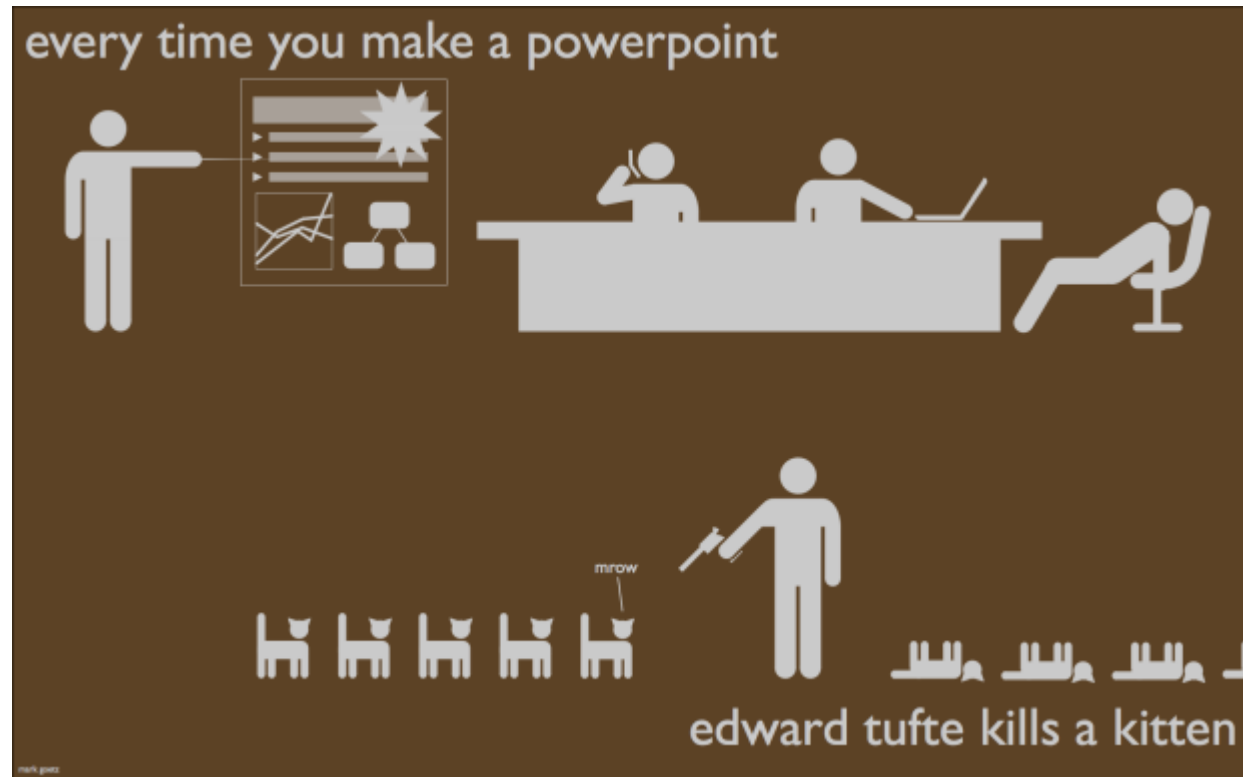 - Crater overpredicted penetration of tile coating **significantly**
    - Initial penetration to described by normal velocity
      - Varies with volume/mass of projectile (e.g., 200ft/sec for 3cu. In)
  - **Significant** energy is required for the softer SOFI particle to penetrate the relatively hard tile coating
    - Test results do show that (it) is possible at sufficient mass and velocity
  - Conversely, once tile is penetrated SOFI can cause **significant** damage
    - Minor variations in total energy (above penetration level) can cause **significant** tile damage
  - Flight condition is **significantly** outside of test database
    - Volume of ramp is 1920cu in vs 3 cu in for test

BOEING

None of the 5 "significants" refer to statistical significance

Vague reference to "it" -> damage to the left wing

Unit of measurement is shown in different ways

https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001yB

We're all guilty

https://flowingdata.com/2010/03/20/powerpoint-and-dying-kittens/
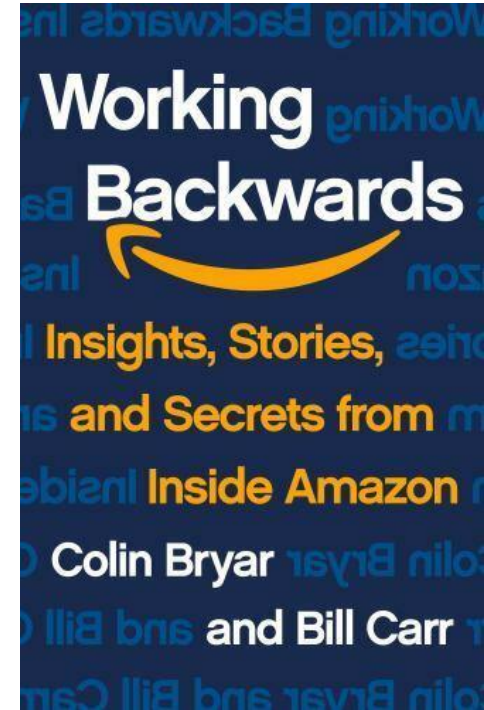
https://markandrewgoetz.com/

Narratives

- The first 20min of many meetings
- Six-pager (~ 3min per page)
- Narratives produce clear thinking
- Stimulate discussion
- Ideas matter most
- A document is portable and easy to circulate

AMAZON STAFF MEETINGS: "NO POWERPOINT"

November 30, 2012 · by Conor Neill · in Decision Making, Great Presentations, Organisation Behaviour, Powerpoint, Public Speaking · 80 Comments

"We have study hall at the beginning of our meetings." says Jeff Bezos.

Staff meetings at Amazon begin with 30 minutes of silent reading.

https://conorneill.com/2012/11/30/amazon-staff-meetings-no-powerpoint/

Links

- Edward Tufte (https://www.edwardtufte.com/tufte/)
- Marti Hearst "Search User Interfaces" (https://searchuserinterfaces.com/)
- J. Heer et al. "A Tour through the Visualization Zoo" (https://dl.acm.org/doi/10.1145/1794514.1805128)
- Processing (https://processing.org/)
- D3 (https://d3js.org/)
- The Gettysburg PPT Presentation (https://norvig.com/Gettysburg/)