Visualization Design

ID 413: Information Graphics and Data Visualization
Spring 2016

Venkatesh Rajamanickam (@venkatrajam)
venkatra@iitb.ac.in
http://info-design-lab.github.io/ID413-DataViz/

Tufte's design principles for graphical excellence

- Show the data
- Induce the viewer to think about the substance, rather than about methodology, graphic design, [or] the technology of graphic production
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail
- Serve a reasonably clear purpose
- Be closely integrated with the statistical and verbal descriptions

1. Organise

// PROGRAMM

MITTWOCH 21.04.2004

12:00 13:00 14:00 15:00 16:00 17:00

LAGERHALLE

KUNSTHALLE DOMINIKANERKIRCHE

DONNERSTAG 22.04.2004

12:00 13:00 14:00 15:00 16:00 17:00

LAGERHALLE
- BRAVE NEW WORLD — 12:00 // 58 min
- ELSEWHERE — 13:30 Le Gout du Koumiz // 66 min
- MEMORIES ARE MADE OF THIS — 15:00 // 64 min
- JONAS AT THE OCEAN — 16:00 // Peter Sempel / 16mm / 2003 / 55 min

ARTHOUSE 4
- DIE PERLE IN DER KACKE — 15:30 // Dirk Bill / 82 min
- EWIG… — 17:30 / »Atlik…

ARTHOUSE 5
- NIWATORI WA HADASHI DA — 15:30 // Atsuma Morisaki / Japan 2003 / 114 min / OmU
- ARAG… — 17:30 / / Farb…

HAUS DER JUGEND (HDJ)
- STUDENT FORUM: MEDIA ACADEMIES / PART I — 12:00 IAMAS Kyoto, Christa Sommerer (A/JP); 13:00 Academy of Fine Arts, Prague, Anetta Mona Chiesa (SK/CZ)
- STUDENT FORUM: MEDIA ACADEMIES / PART II — 15:00 HfG Offenbach, Retrand Pape (D); 16:00 AKI Enschede, Bill Sylnhoven (NL); 17:00 KHM Köln, Katrin Peters (D)

FREITAG 23.04.2004

12:00 13:00 14:00 15:00 16:00 17:00

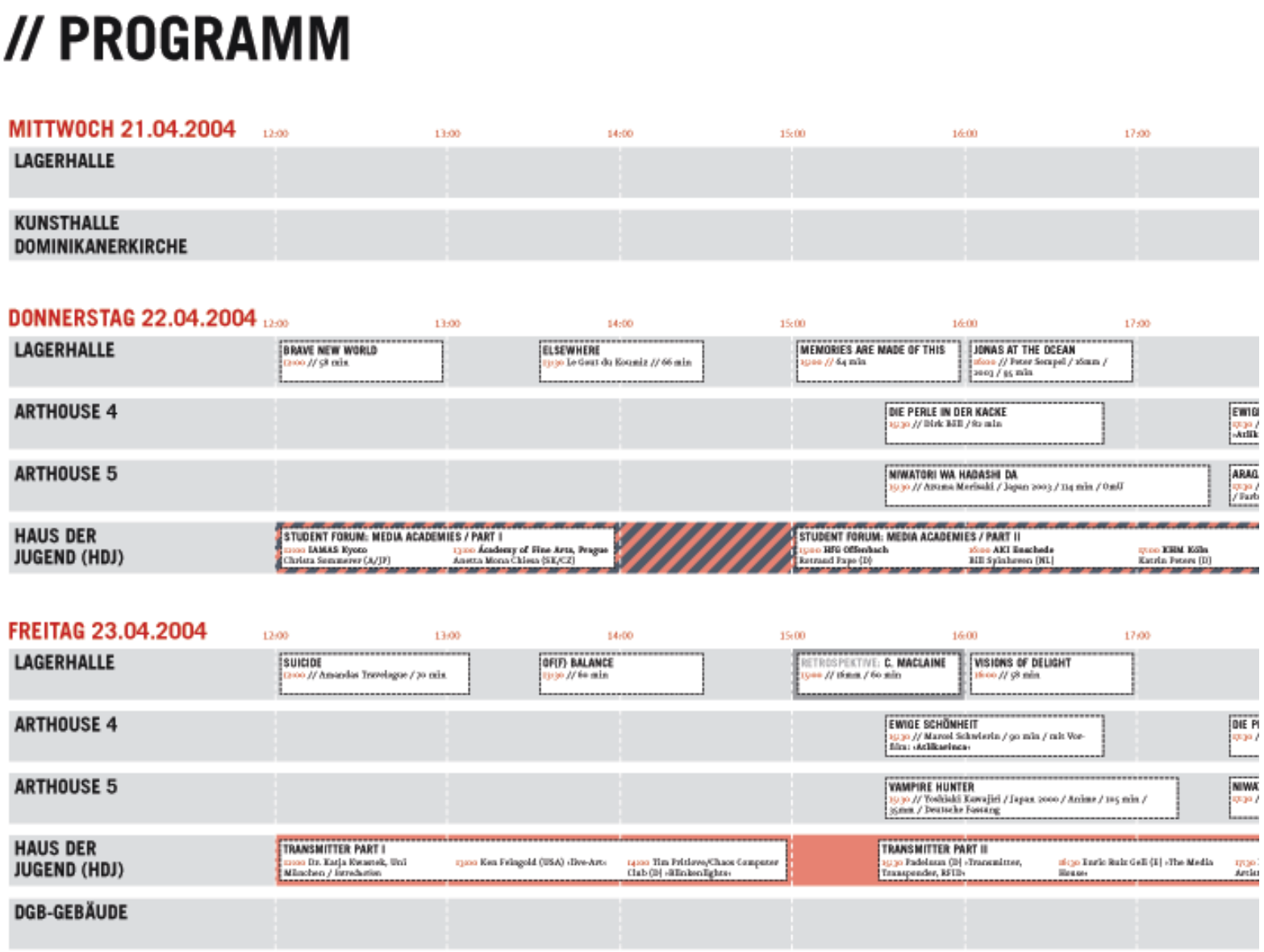
LAGERHALLE
- SUICIDE — 12:00 // Amandas Travelogue / 70 min
- OF(F) BALANCE — 13:30 // 60 min
- RETROSPEKTIVE: C. MACLAINE — 15:00 // 16mm / 60 min
- VISIONS OF DELIGHT — 16:00 // 58 min

ARTHOUSE 4
- EWIGE SCHÖNHEIT — 15:30 // Marcel Schwierin / 90 min / mit Vorfilm: »Atlikarinca«
- DIE P… — 17:30 /

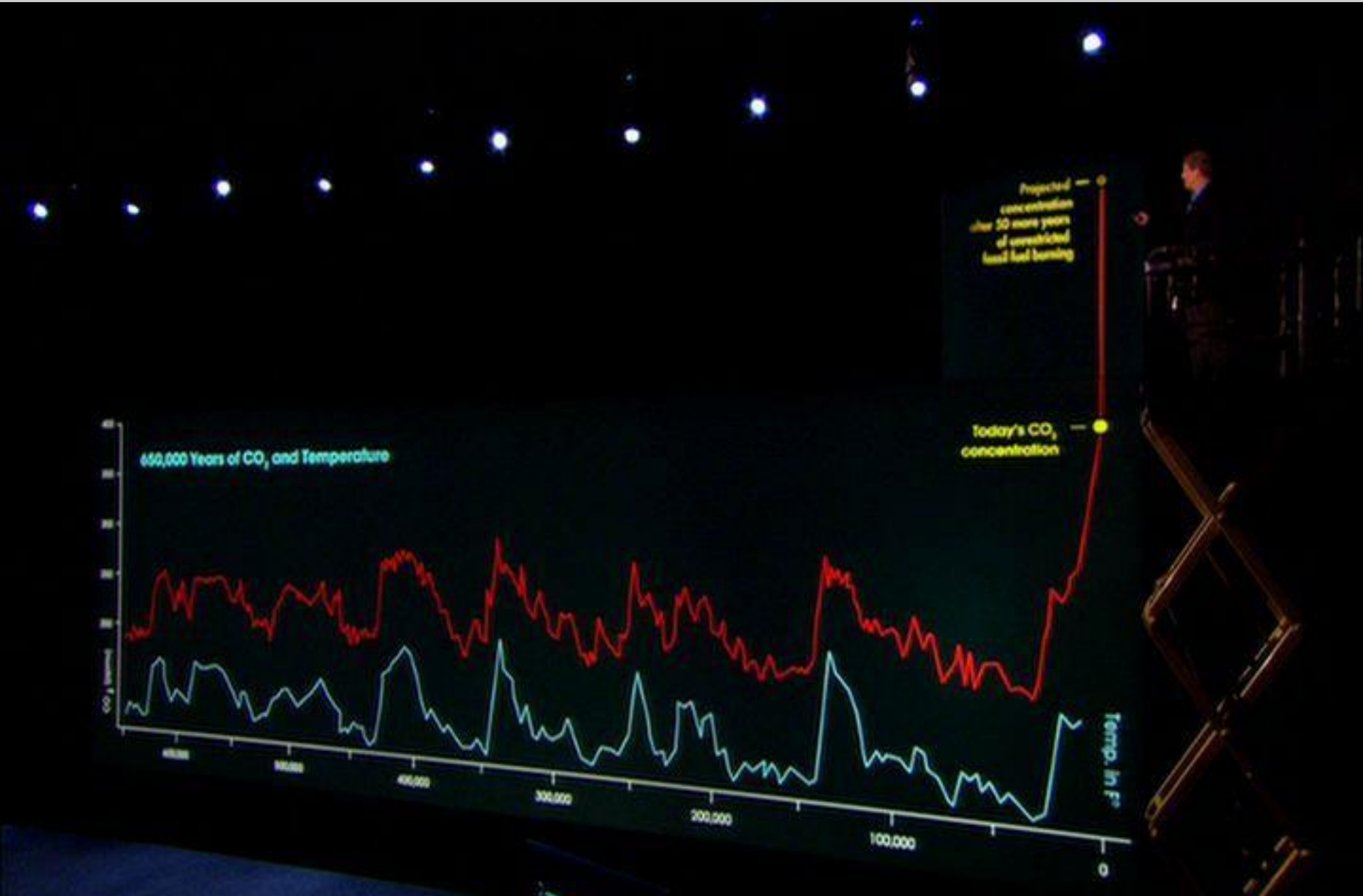
ARTHOUSE 5
- VAMPIRE HUNTER — 15:30 // Yoshiaki Kawajiri / Japan 2000 / Anime / 105 min / 35mm / Deutsche Fassung
- NIWA… — 17:30 /

HAUS DER JUGEND (HDJ)
- TRANSMITTER PART I — 12:00 Dr. Katja Kwastek, Uni München / Introduction; 13:00 Ken Feingold (USA) »Live-Art«; 14:00 Tim Pritlove/Chaos Computer Club (D) »Blinkenlights«
- TRANSMITTER PART II — 15:30 Padelman (D) »Transmitter, Transponder, RFID«; 16:30 Enric Ruiz Gell (E) »The Media House«; 17:30 … Artist…

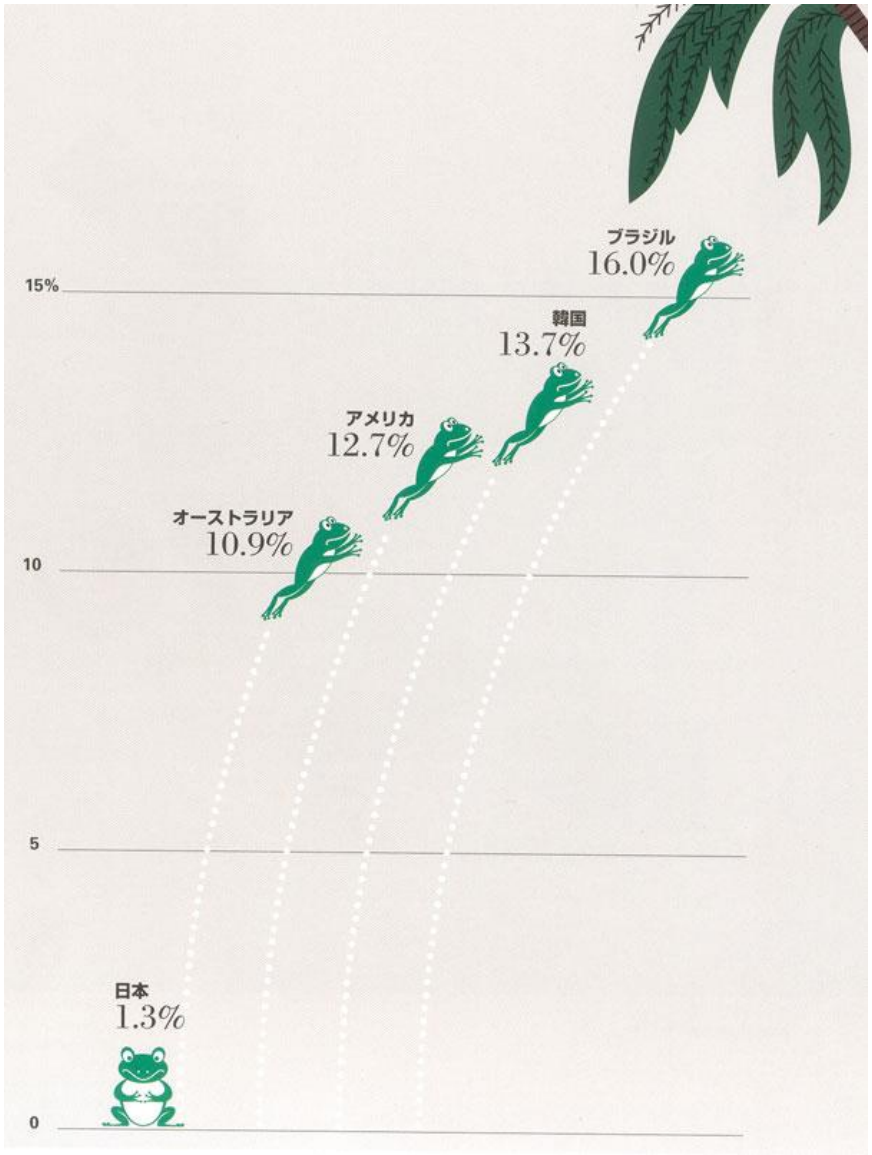
DGB-GEBÄUDE



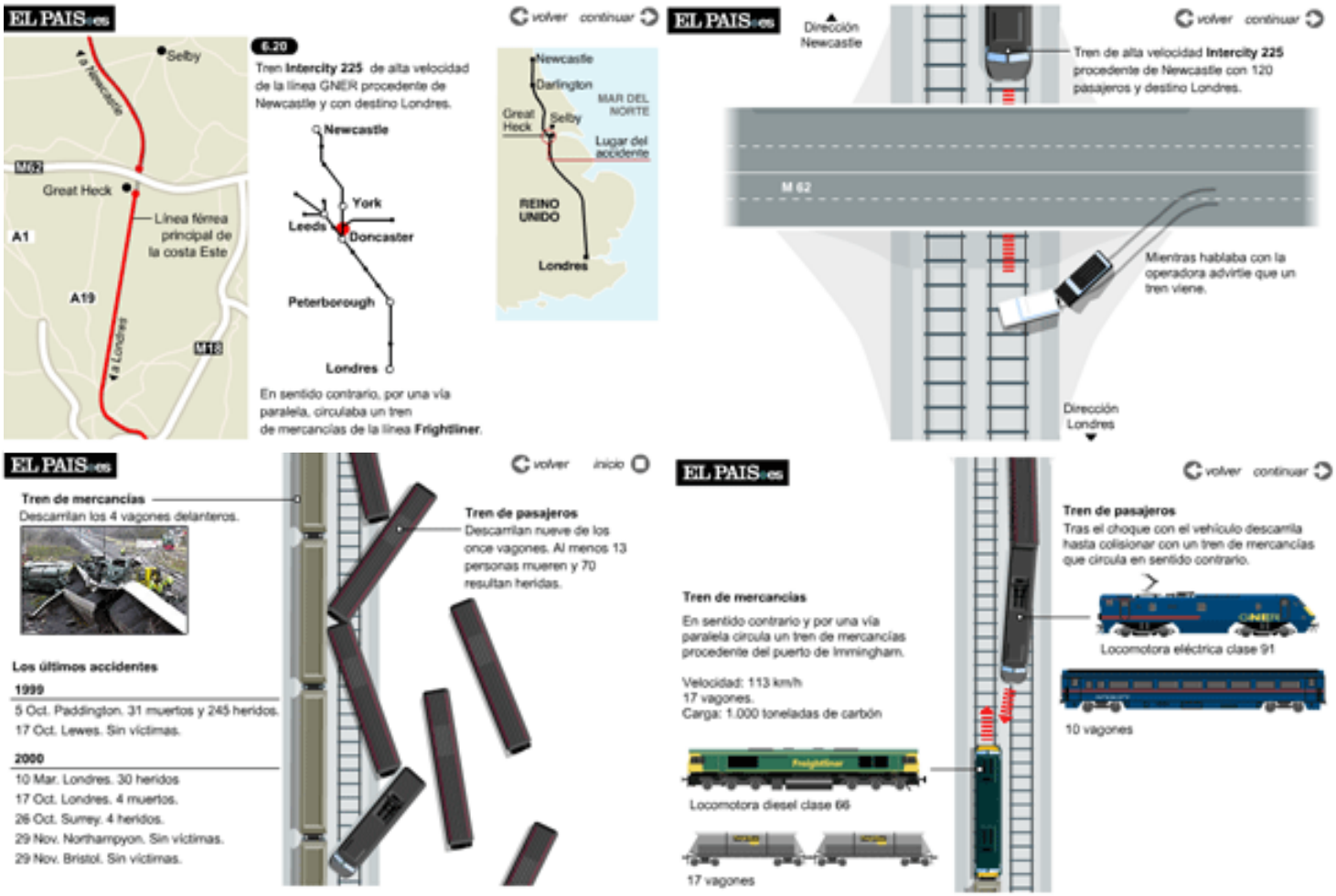
2. Make Visible



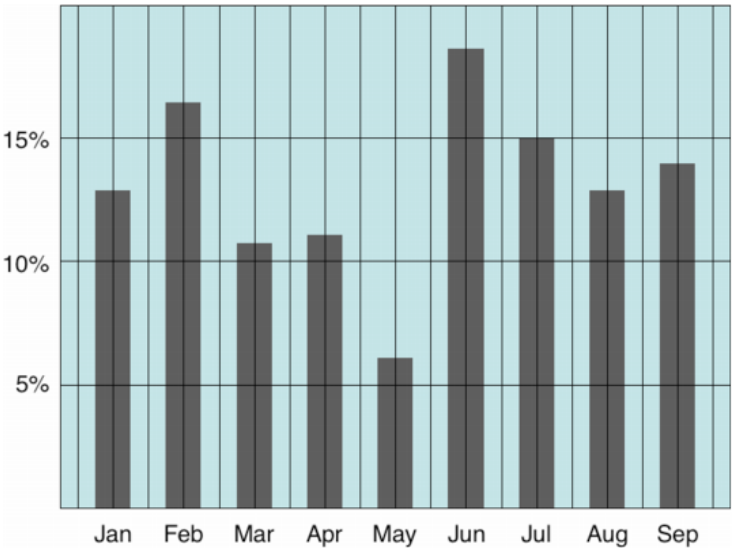
2. Make Visible



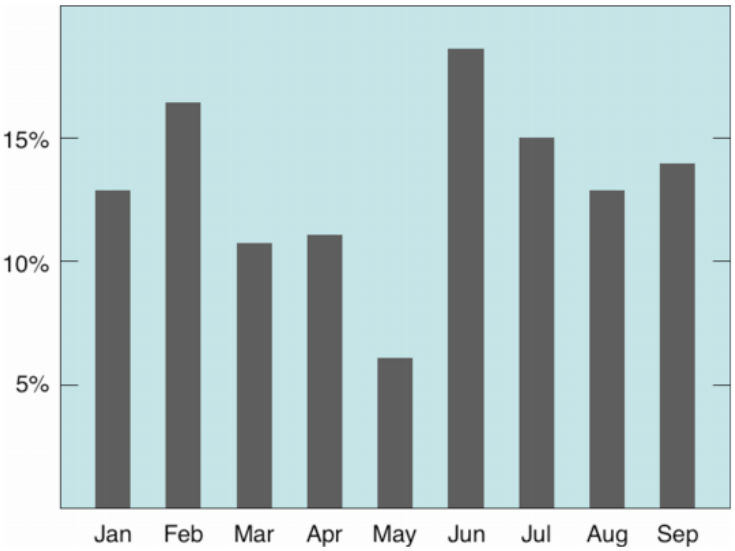
3. Establish Context



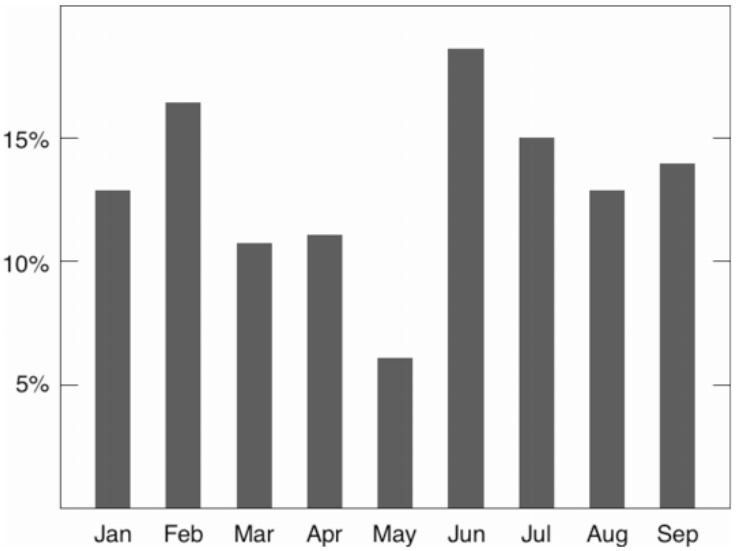
4. Simplify



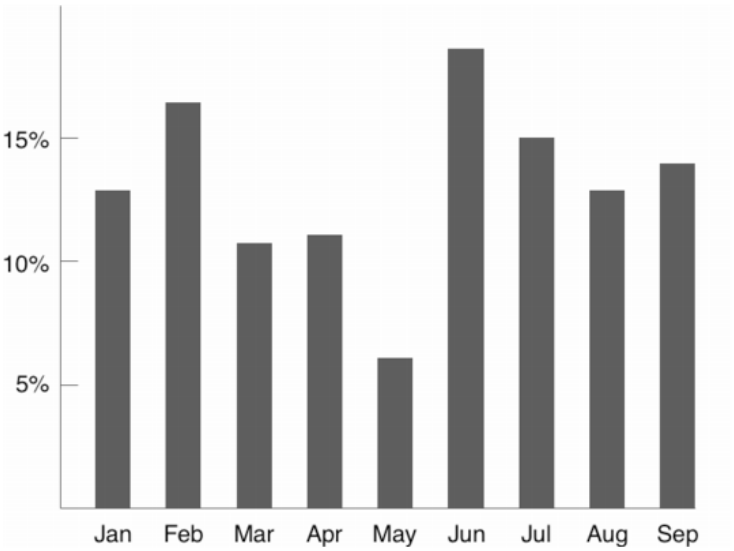
4. Simplify



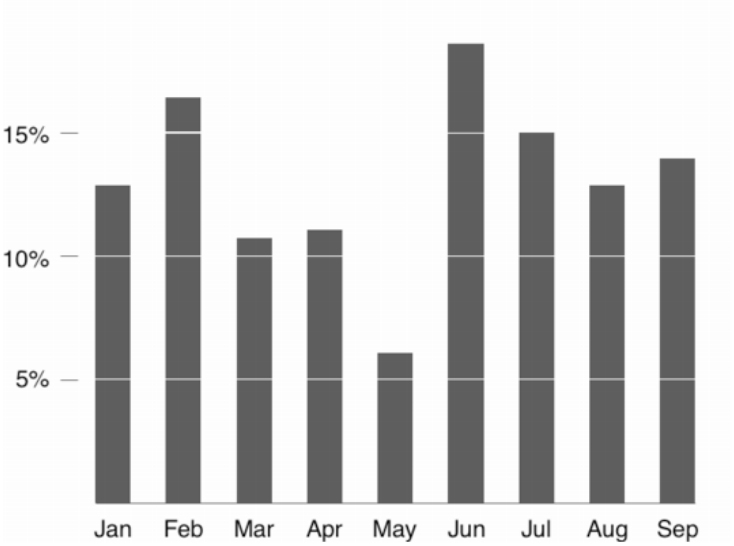
4. Simplify



4. Simplify

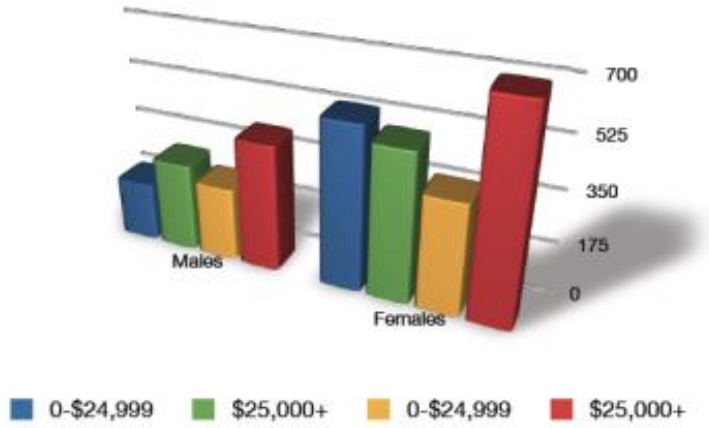


4. Simplify



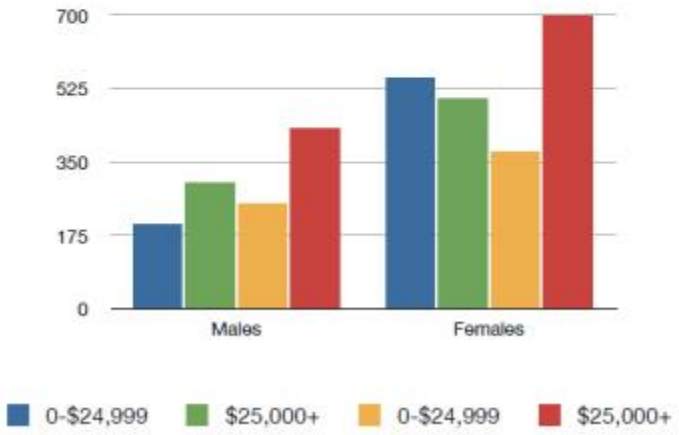
5. Maximize Data-Ink Ratio

$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$

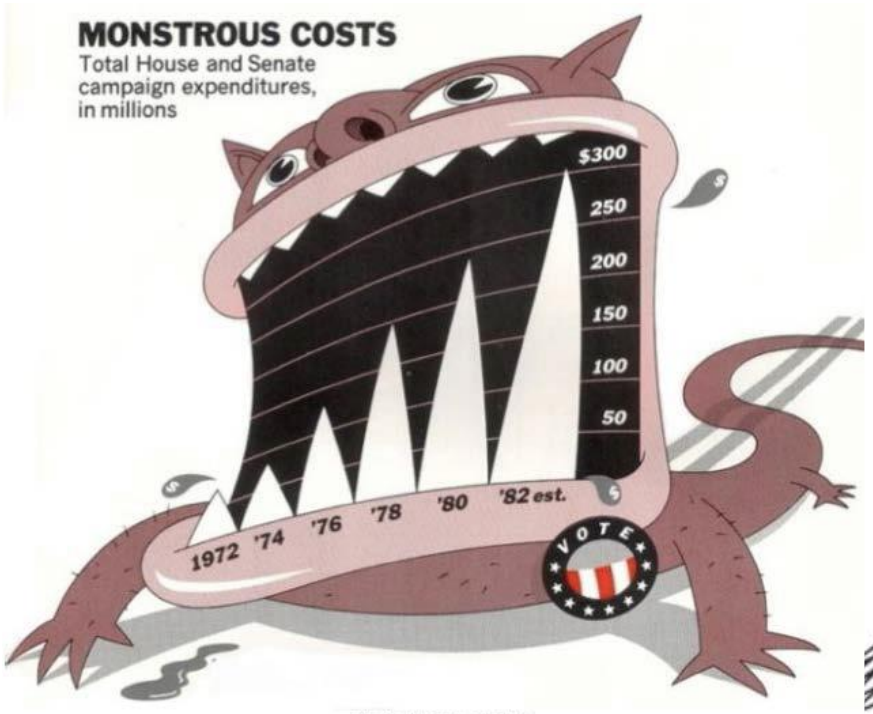


5. Maximize Data-Ink Ratio

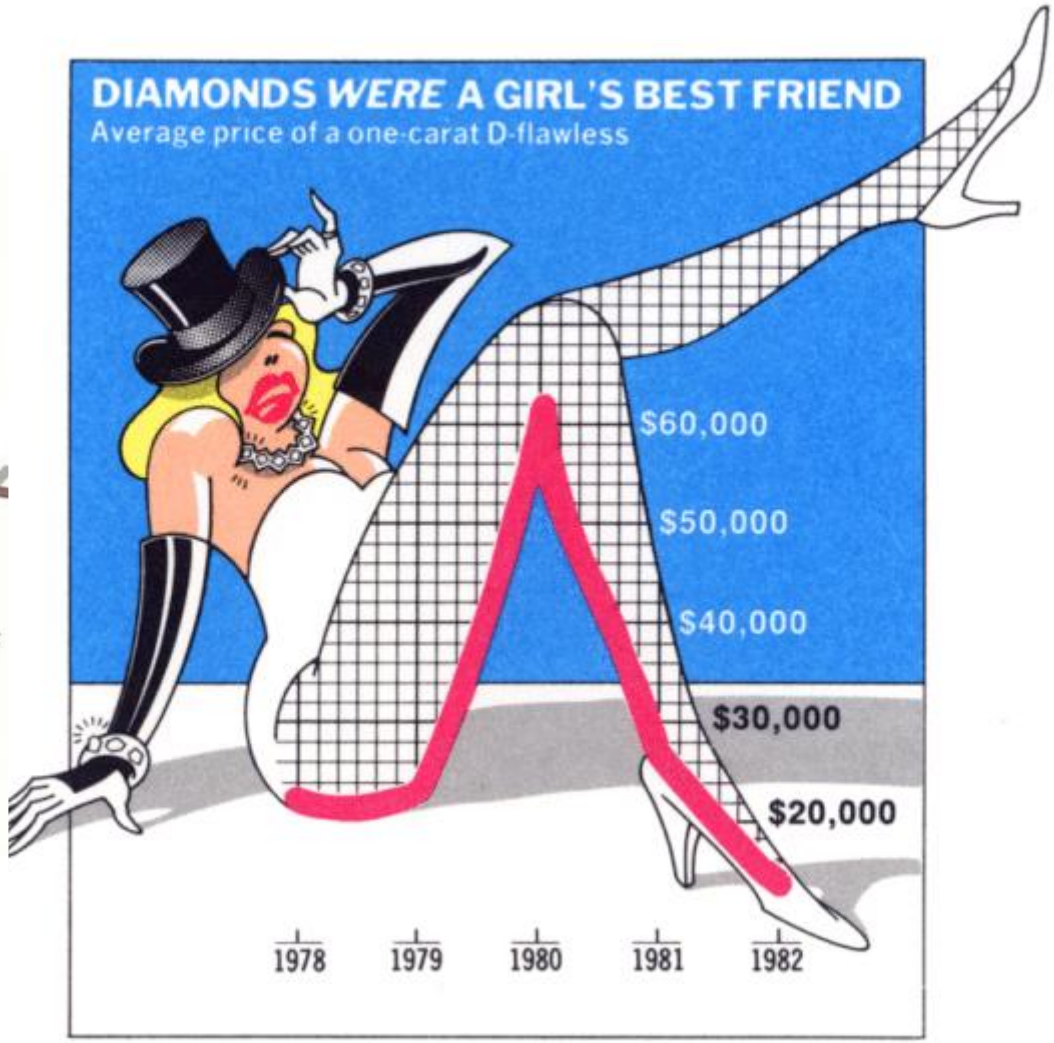
$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



5. Maximize Data-Ink Ratio



MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions



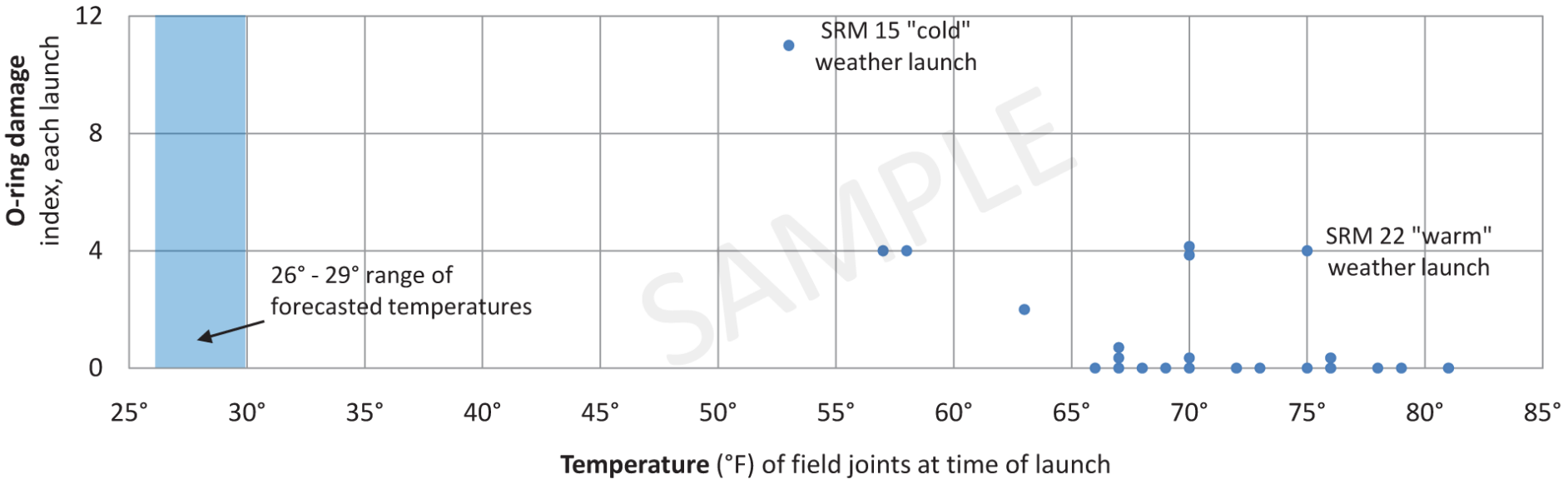
Nigel Holmes

6. Show Cause and Effect

Space Shuttle History of Temperature and O-ring Damage

For All 24 Launches Prior to Challenger on January 28, 1986

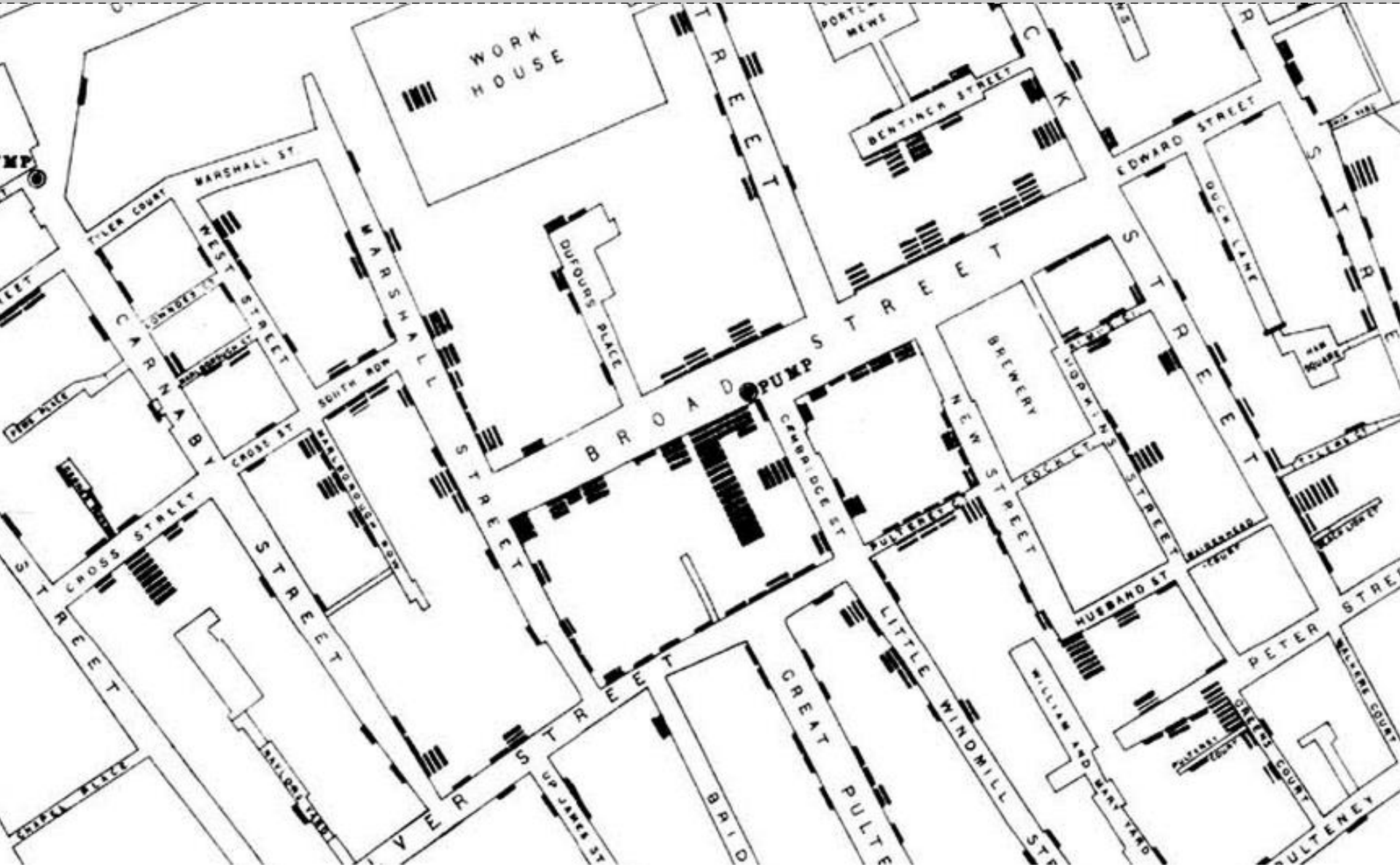
Solid Rocket Motor (SRM) 15 and SRM 22 were the only prior launches discussed in relation to temperature on the eve of the launch.



Sources: Presidential Commission on the Space Shuttle Challenger Accident (PCSSCA) and Post-Challenger Evaluation of Space Shuttle Risk Assessment and Management as quoted in **Visual and Statistical Thinking** by Edward Tufte.

© Joe Bobcat

6. Show Cause and Effect

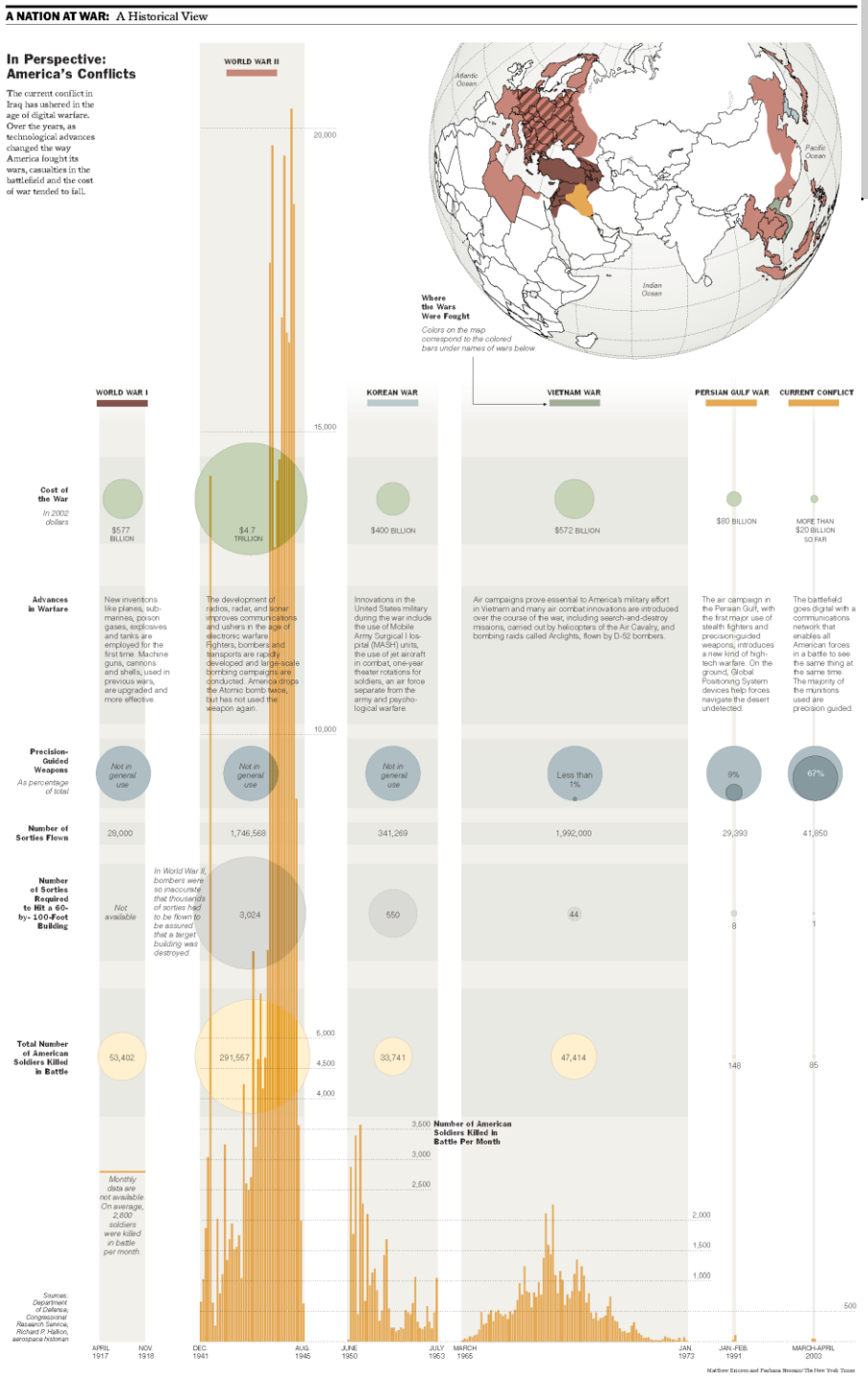


6. Show Cause and Effect

Magician Teller's definition of magic:

"The theatrical linking of a cause with an effect that has no basis in physical reality, but that — in our hearts — ought to."

7. Compare and Contrast



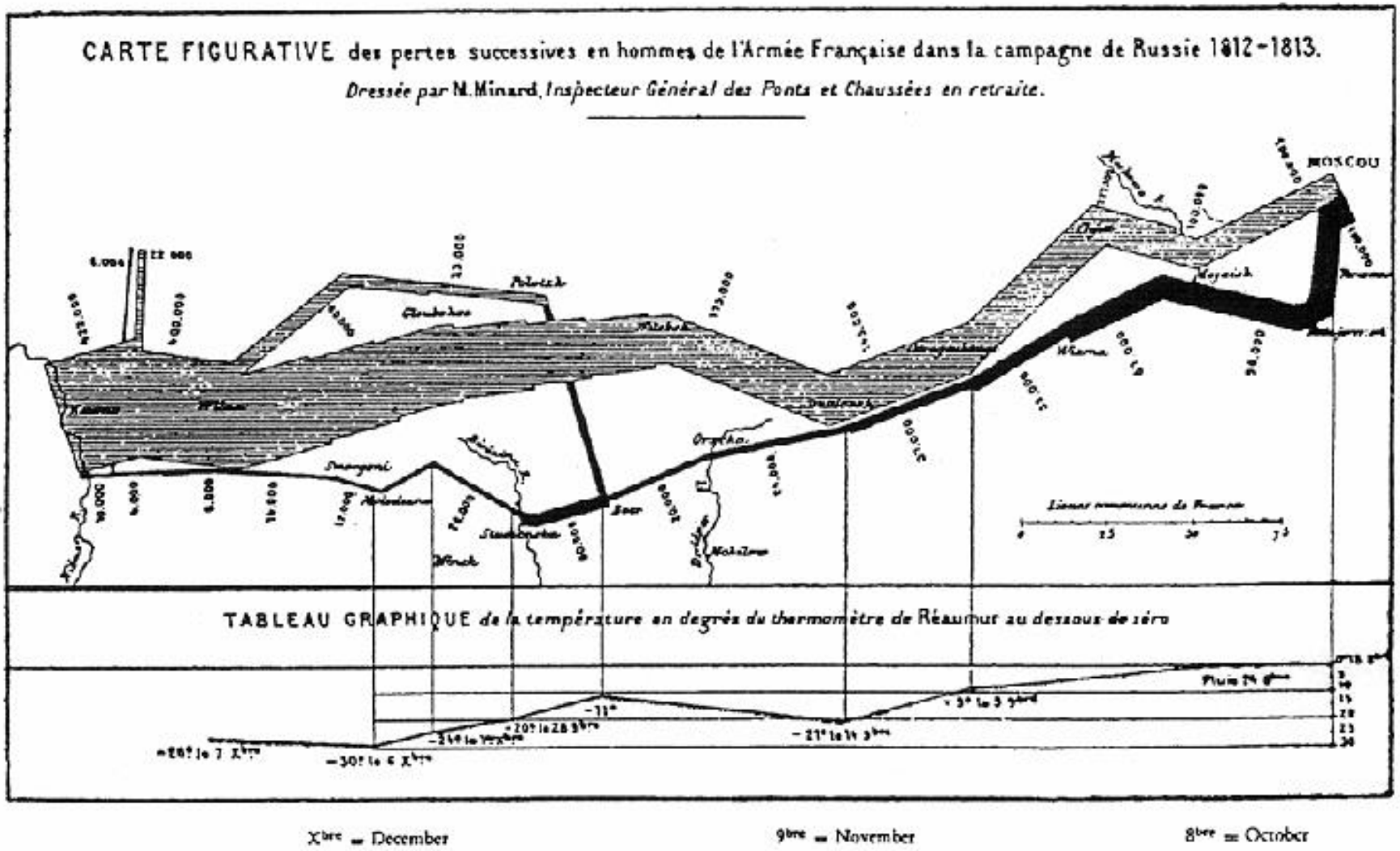
7. Compare and Contrast

The Fallen of World War II is an interactive documentary that examines the human cost of the second World War and the decline in battle deaths in the years since the war. The 15-minute data visualization uses cinematic storytelling techniques to provide viewers with a fresh and dramatic perspective of a pivotal moment in history.

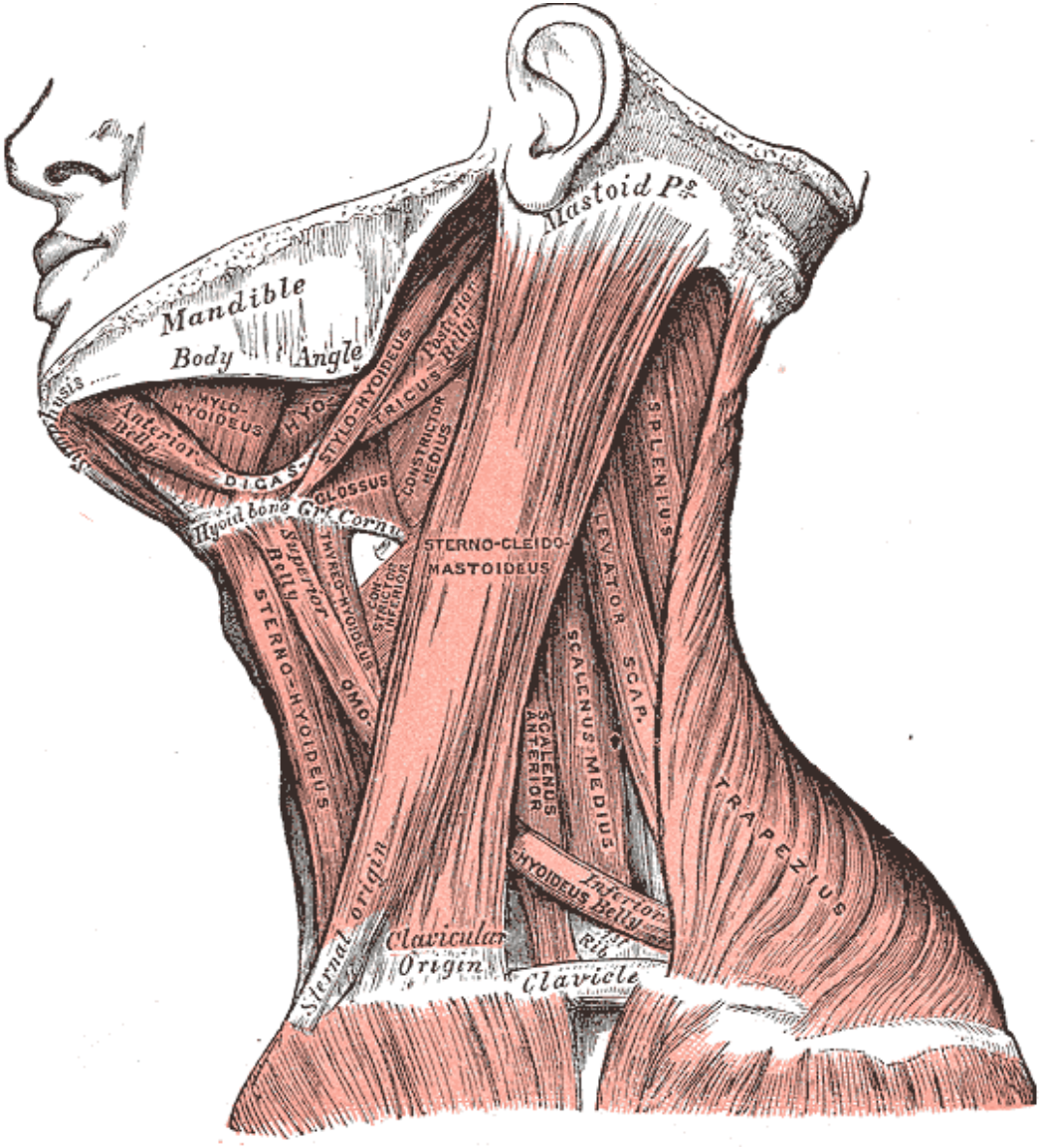
The film follows a linear narration, but it allows viewers to pause during key moments to interact with the charts and dig deeper into the numbers.

http://www.fallen.io/ww2/

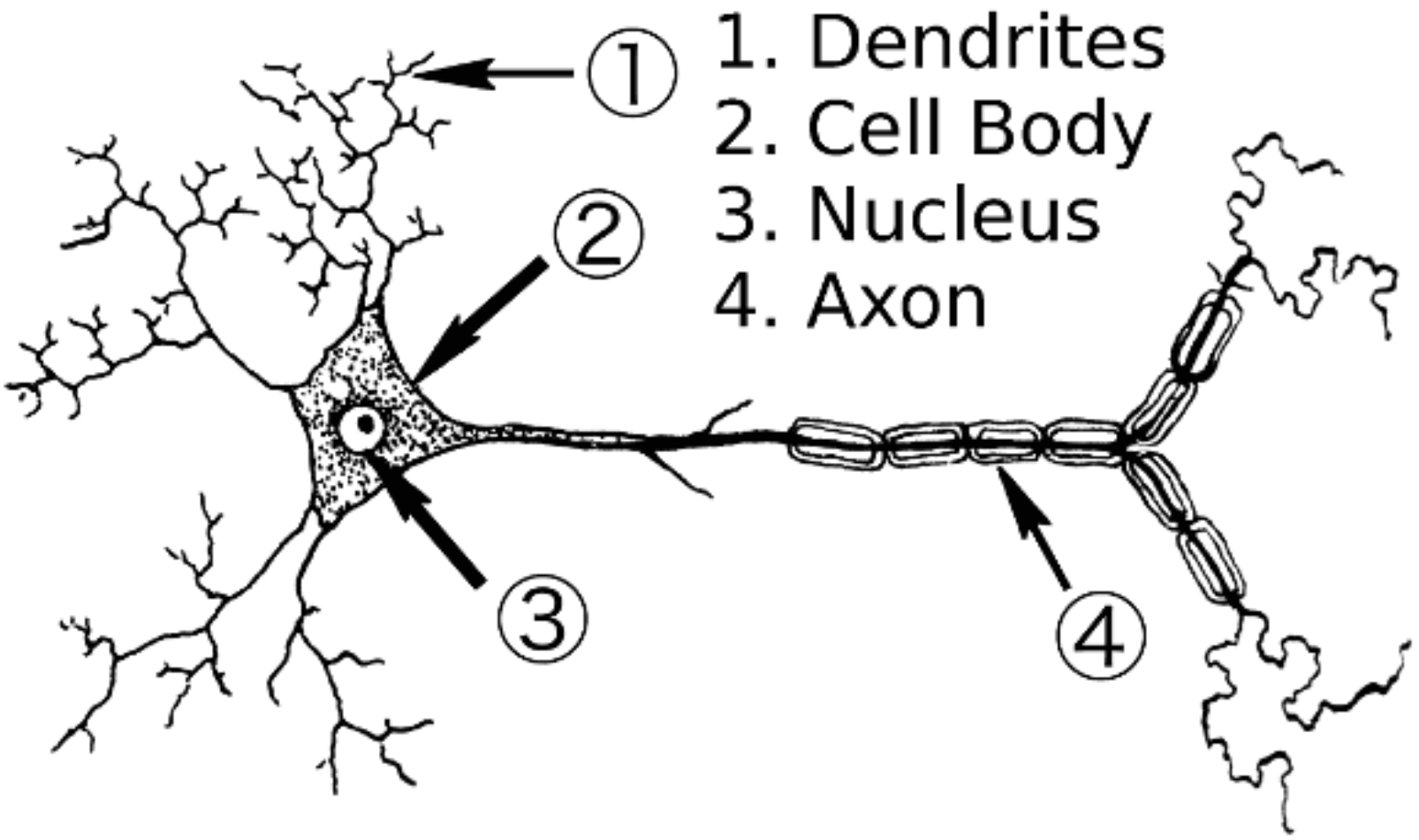
8. Show Multiple Dimensions



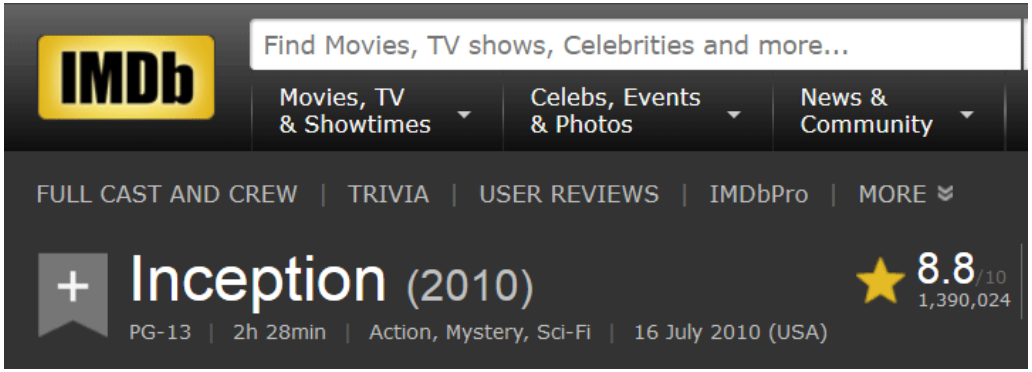
9. Integrate


Mastoid Pr.
Mandible
Body
Angle
Anterior Belly
HYLO-HYOIDEUS
HYO-
STYLO-HYOIDEUS
TRICUS
Posterior Belly
DIGAS-
GLOSSUS
CONSTRICTOR MEDIUS
Hyoid bone
Gr. Cornu
STERNO-CLEIDO-MASTOIDEUS
SPLENIUS
LEVATOR SCAP.
Superior Belly
THYREO-HYOIDEUS
CONSTRICTOR INFERIOR
STERNO-HYOIDEUS
OMO-
SCALENUS ANTERIOR
SCALENUS MEDIUS
TRAPEZIUS
HYOIDEUS
Inferior Belly
Sternal origin
Clavicular Origin
Clavicle

9. Integrate

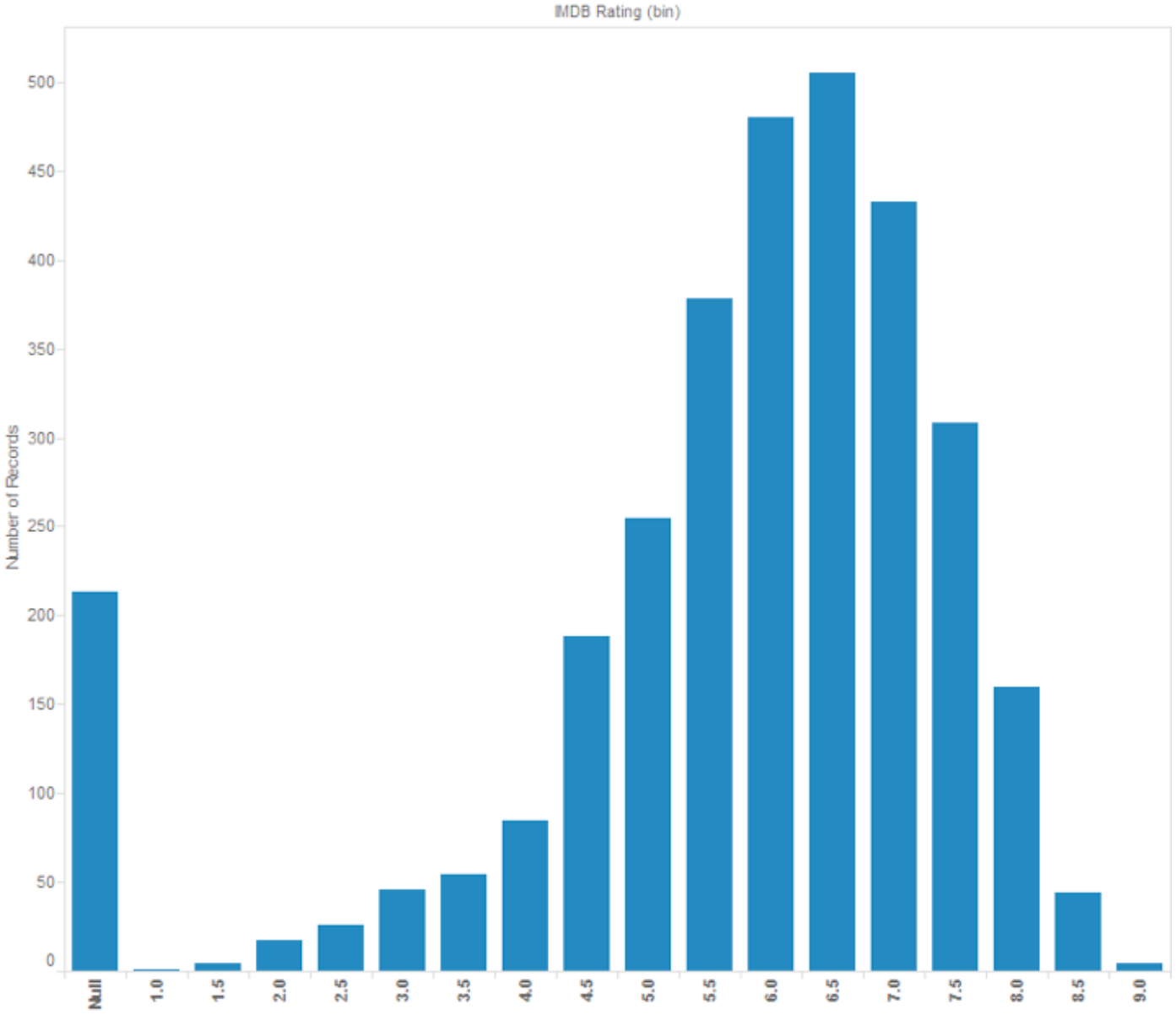


Analysis Example: Motion Pictures Data

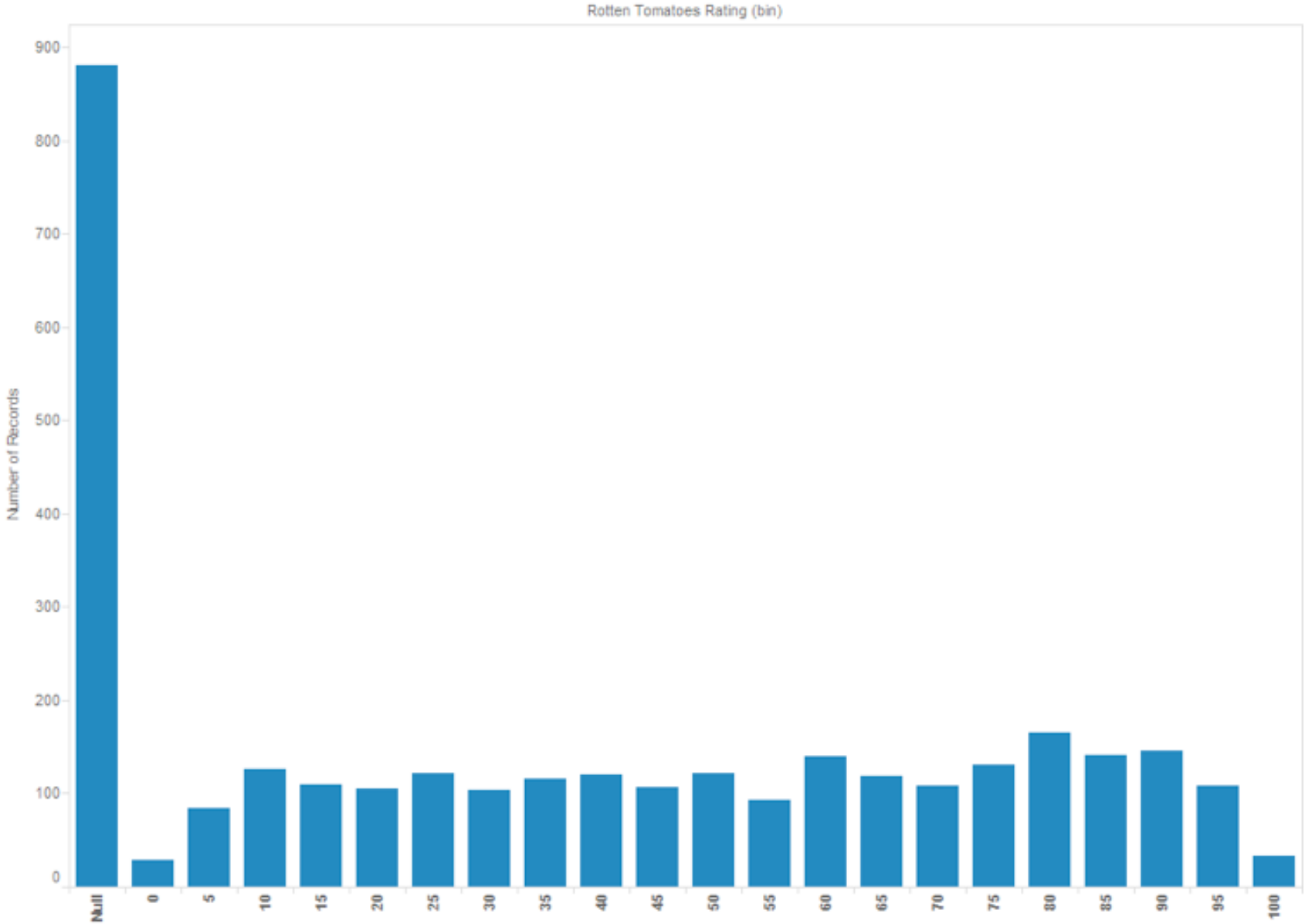


Title	String (N)
IMDB Rating	Number (Q)
Rotten Tomatoes	Rating Number (Q)
Release Date	Date (T)

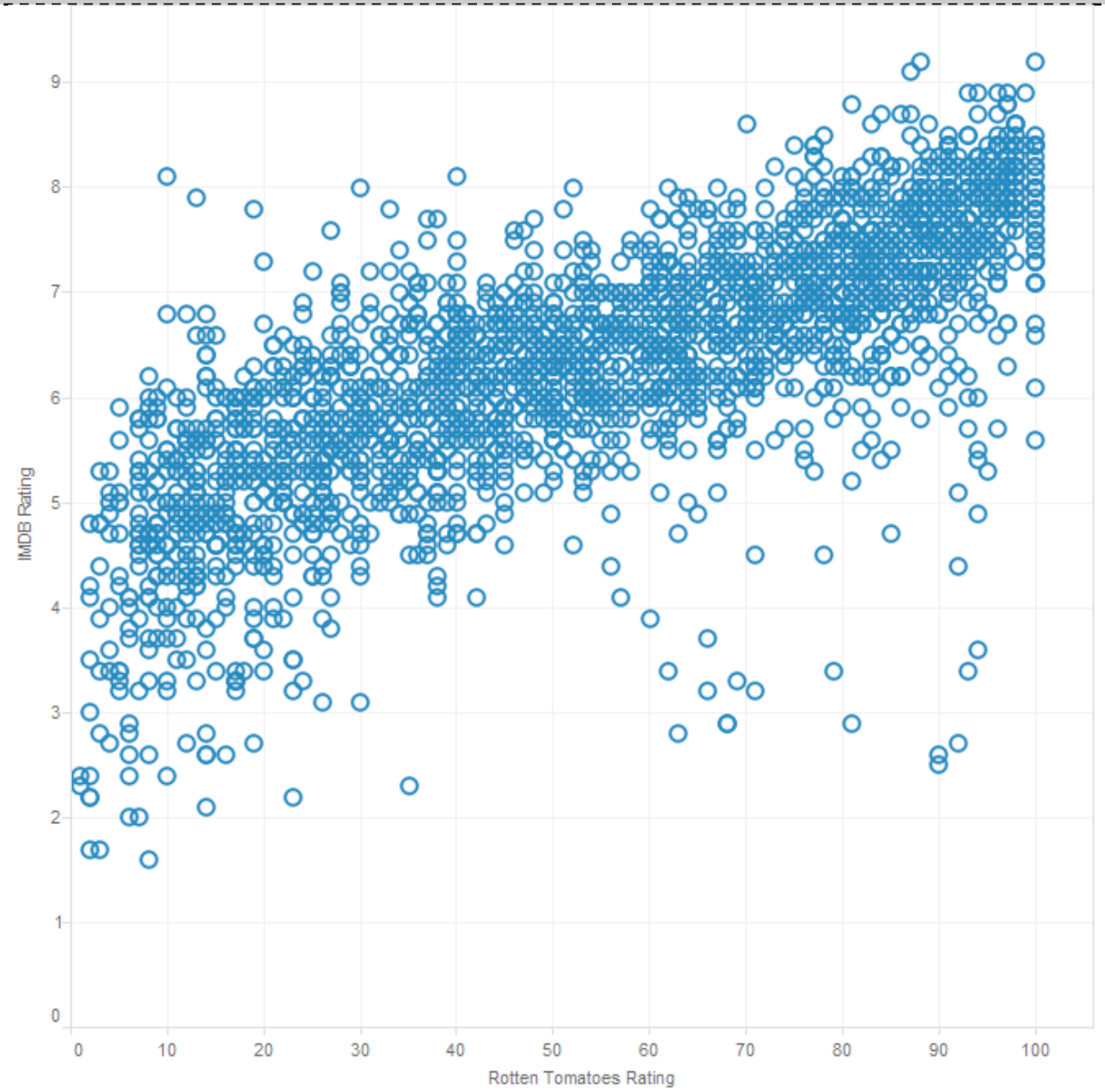
Analysis Example: Motion Pictures Data



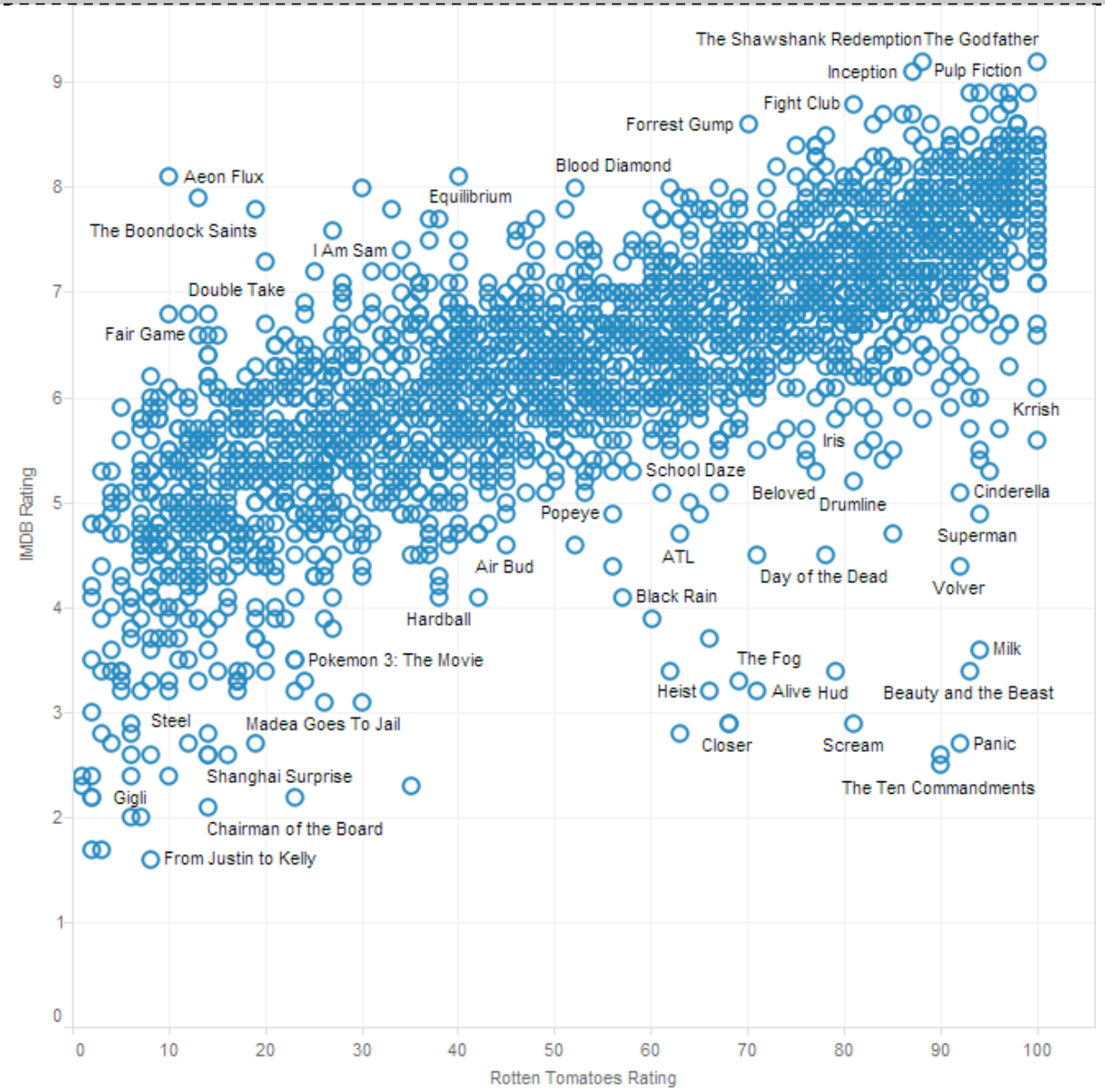
Analysis Example: Motion Pictures Data



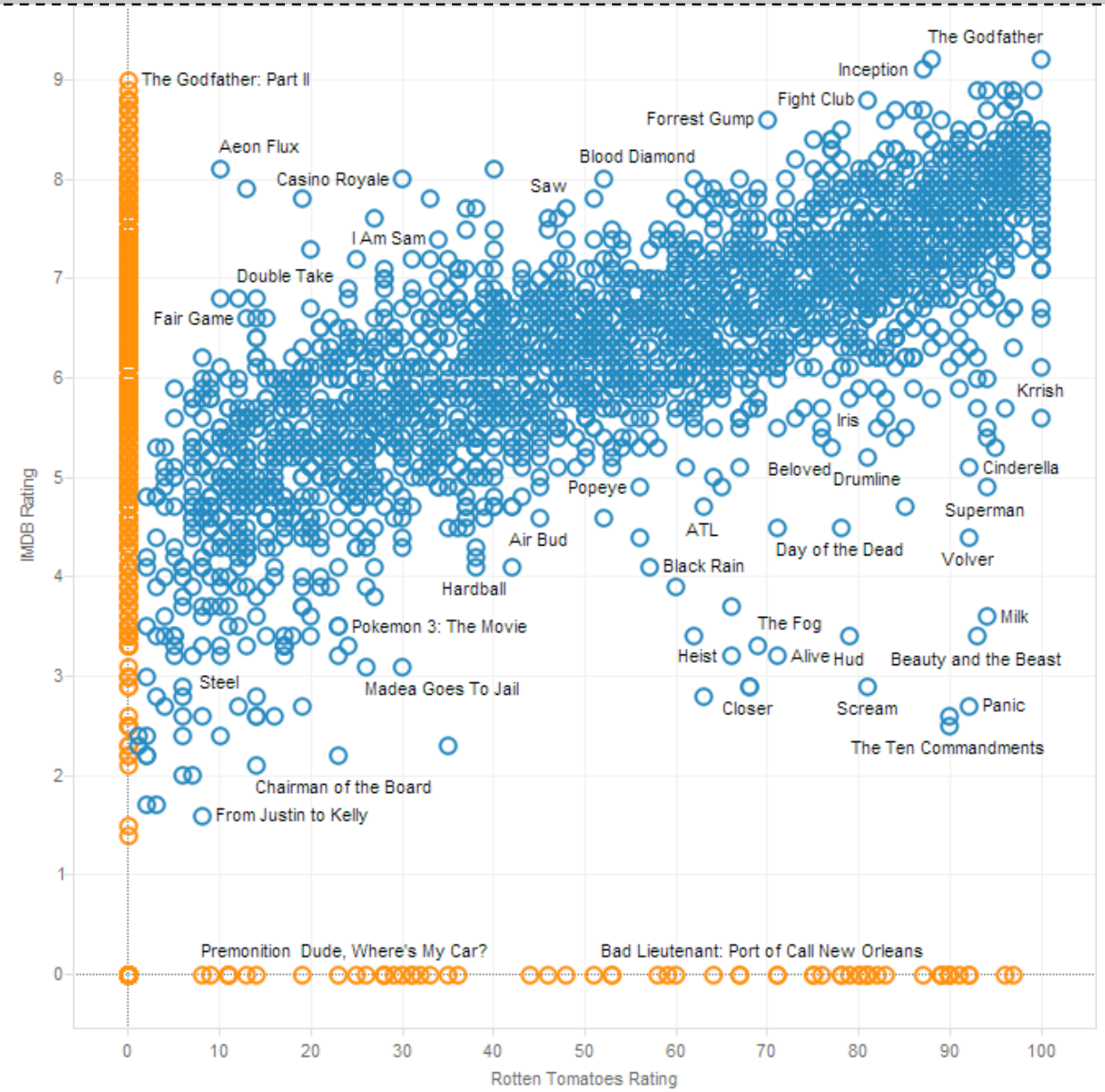
Analysis Example: Motion Pictures Data



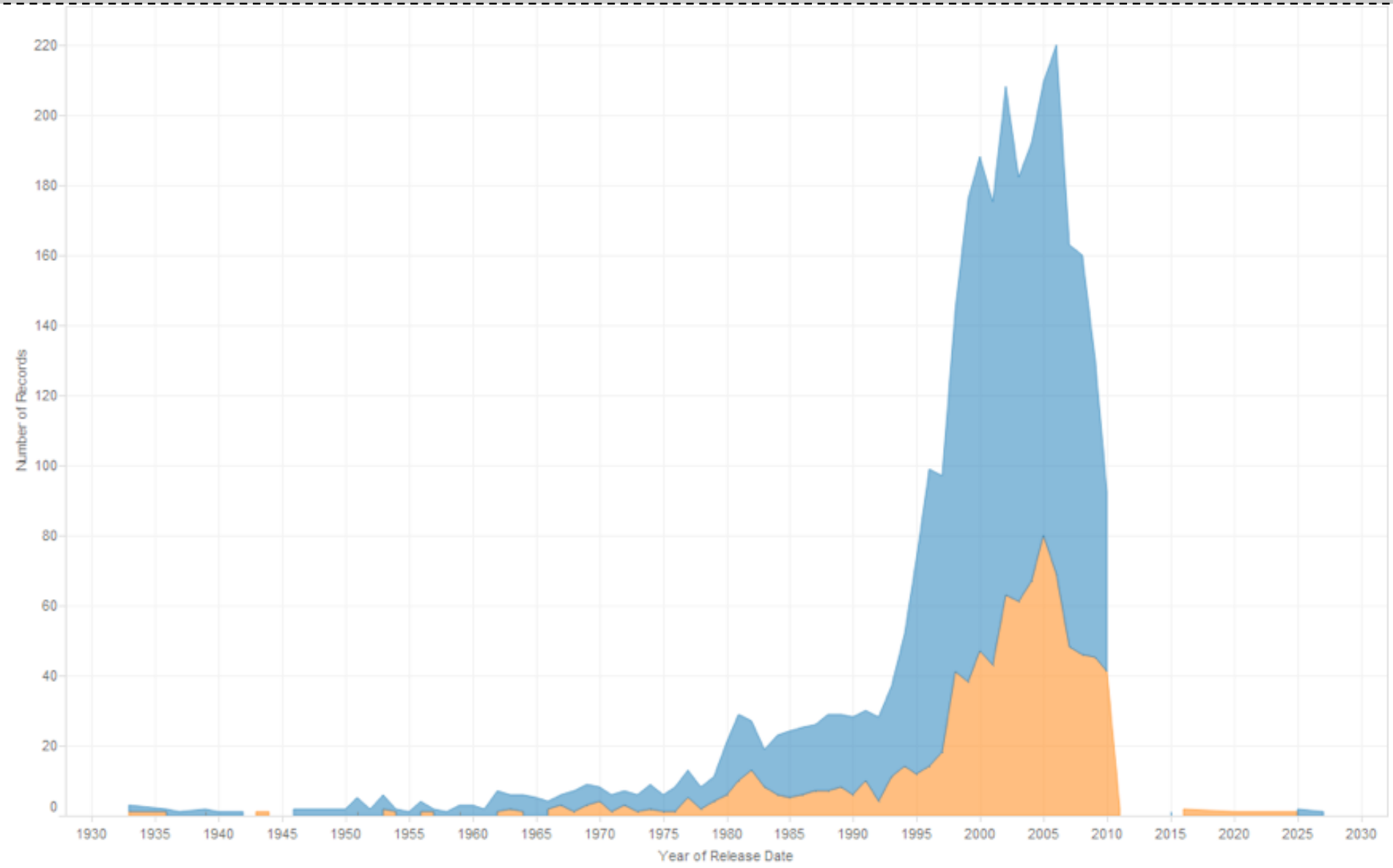
Analysis Example: Motion Pictures Data



Analysis Example: Motion Pictures Data



Analysis Example: Motion Pictures Data



Analysis Example: Motion Pictures Data -- Lessons

- Check **data quality** and your assumptions
- Start with **univariate summaries**, then start to consider **relationships among variables**
- **Avoid premature fixation!**
- Even for “simple” data, a variety of graphics might provide insight. Tailor the choice of graphic to the questions being asked, but be open to surprises
- Graphics can be used to guide and help assess the quality of statistical models

Case Study: Antibiotic Effectiveness

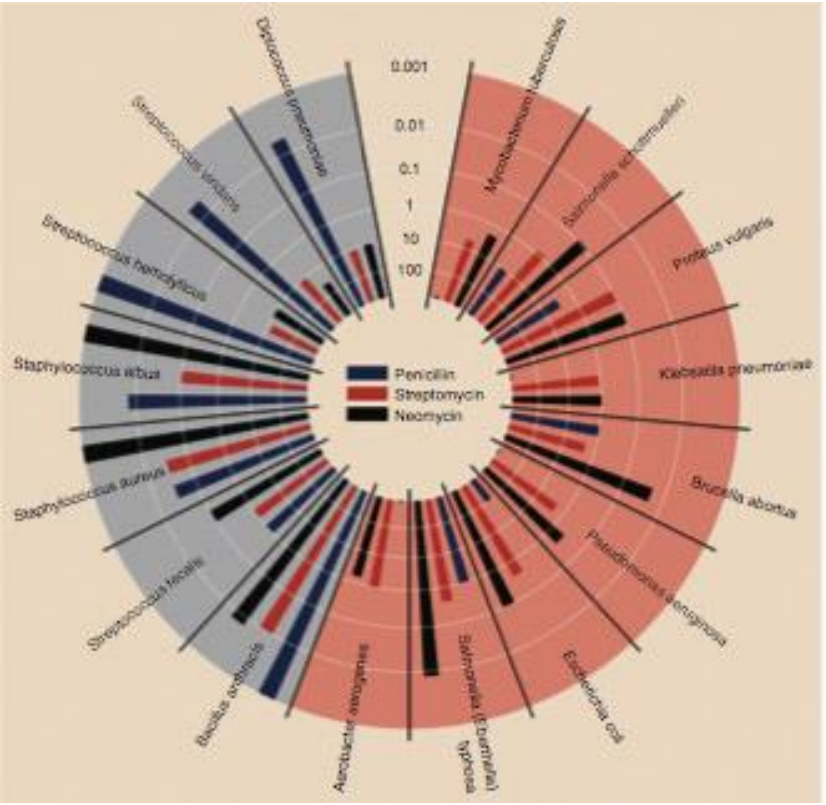
- In 1951, Will Burtin published a graphic display that was admired for the clarity and economy with which it showed the efficacy of three antibiotics on 16 different kinds of bacteria
- The dependent variable was the minimum concentration of the drug required to prevent the growth of the bacteria in vitro—the minimum inhibitory concentration (MIC)
- The three drugs were penicillin, neomycin and streptomycin, and their efficacy varied over six orders of magnitude
- The scale varies from 1,000 micrograms per milliliter to .001 micrograms per millilitre
- Lower is better, indicating less antibiotic is needed to treat the bacteria

Burtin's dataset: What questions might we ask?

Table 1: Burtin's data.

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter *aerogenes*	870	1	1.6	negative
Brucella *abortus*	1	2	0.02	negative
Brucella *anthracis*	0.001	0.01	0.007	positive
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

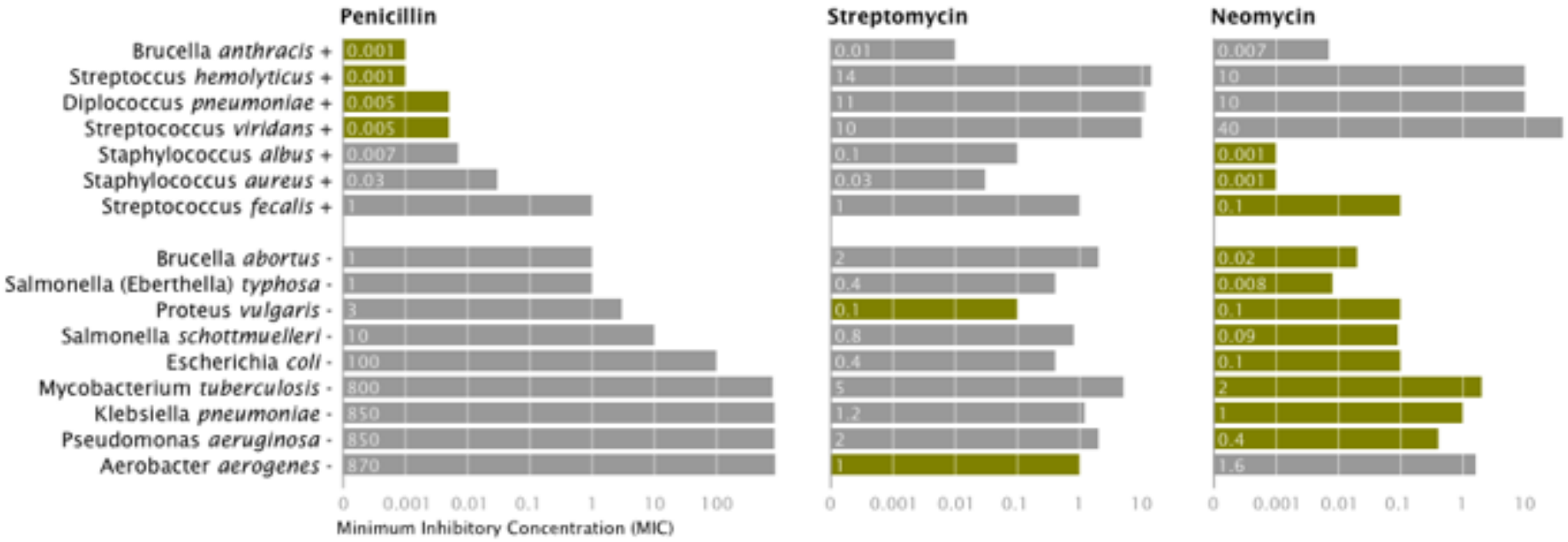
Burtin's dataset: How do the drugs compare?



Bacteria	Penicillin	Antibiotic Streptomycin	Neomycin	Gram stain
Aerobacter aerogenes	870	1	1.6	–
Brucella abortus	1	2	0.02	–
Bacillus anthracis	0.001	0.01	0.007	+
Diplococcus pneumoniae	0.005	11	10	+
Escherichia coli	100	0.4	0.1	–
Klebsiella pneumoniae	850	1.2	1	–
Mycobacterium tuberculosis	800	5	2	–
Proteus vulgaris	3	0.1	0.1	–
Pseudomonas aeruginosa	850	2	0.4	–
Salmonella (Eberthella) typhosa	1	0.4	0.008	–
Salmonella schottmuelleri	10	0.8	0.09	–
Staphylococcus albus	0.007	0.1	0.001	+
Staphylococcus aureus	0.03	0.03	0.001	+
Streptococcus fecalis	1	1	0.1	+
Streptococcus hemolyticus	0.001	14	10	+
Streptococcus viridans	0.005	10	40	+

Radius: 1 / log(MIC)
Bar Colour: Antibiotic
Background Colour: Gram Staining

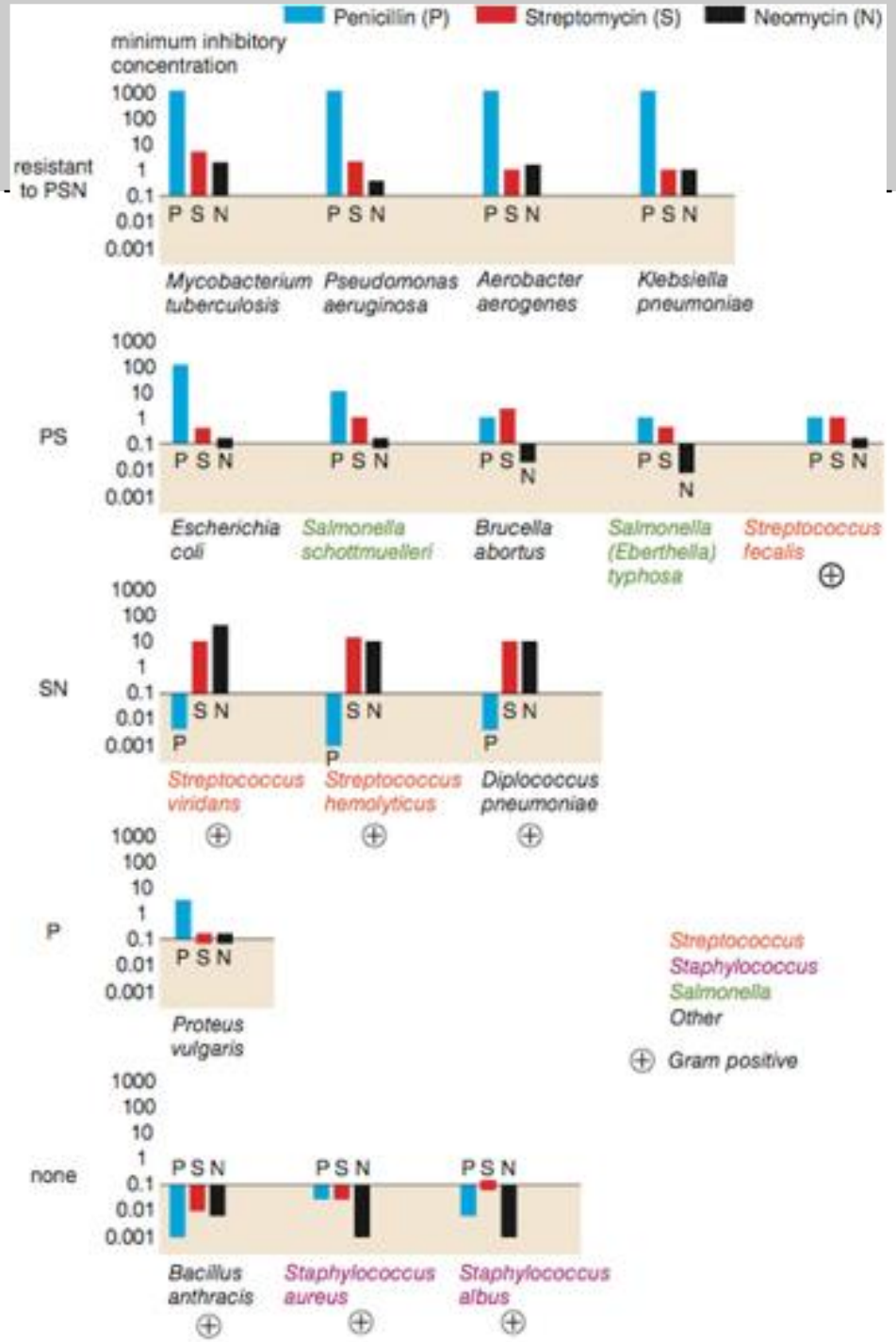
Burtin's dataset: How do the drugs compare?



X-axis: Antibiotic | log(MIC)
Y-axis: Gram-Staining | Species
Colour: Most-Effective

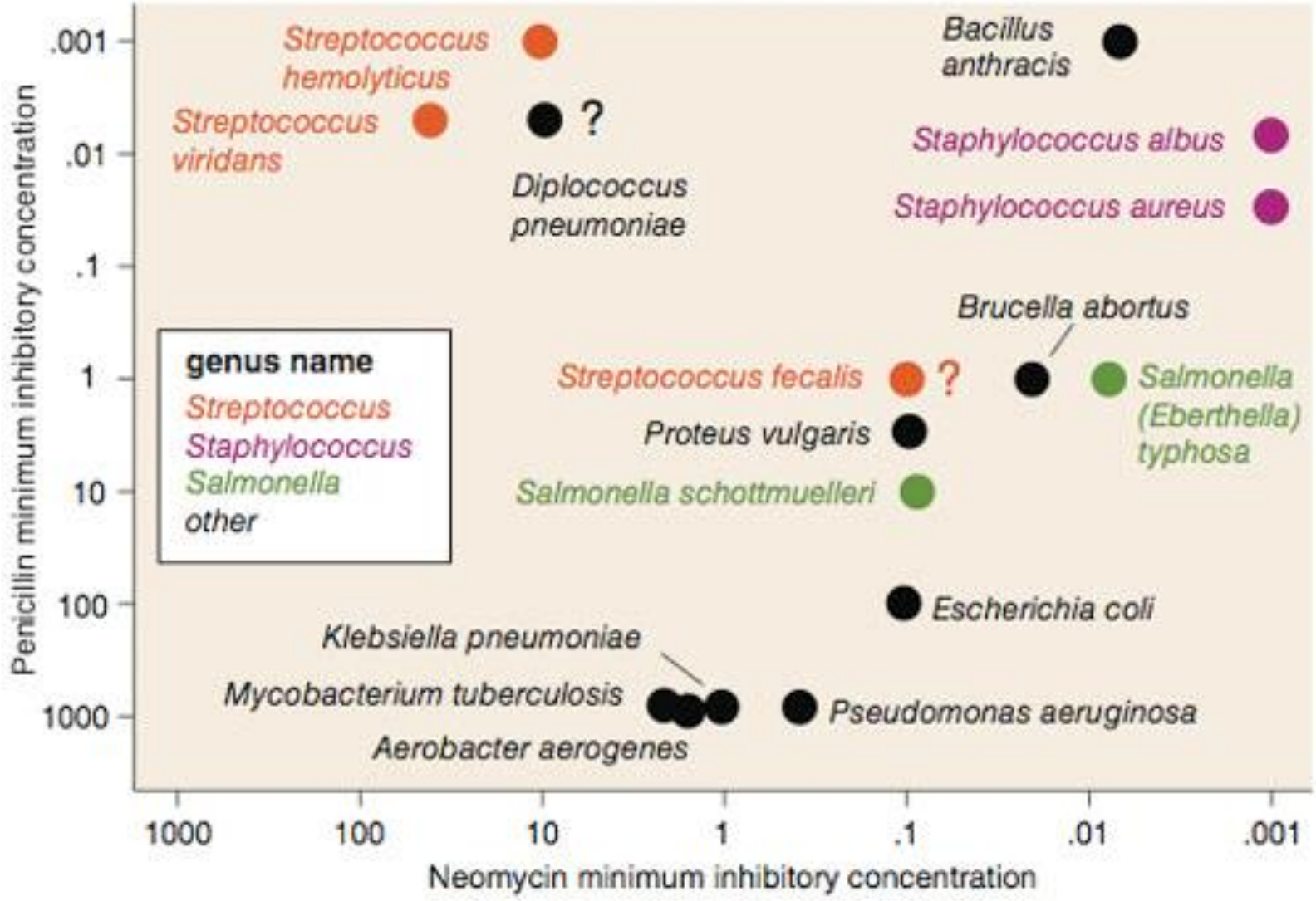
Mike Bostock, 2009

Do bacteria group by antibiotic resistance?



Wainer & Lysen
American Scientist, 2009

Do different antibiotics correlate?



Wainer & Lysen
American Scientist, 2009

Lesson: Iterative Exploration

Exploratory Process:

1. Construct graphics to address questions
2. Inspect “answer” and assess new questions
3. Repeat…

Transform data appropriately (e.g., invert, log)

“Show data variation, not design variation” -Tufte

Visualization Taxonomy

Comparison
Proportion
Distribution
Correlation

Data Visualization Process & Graphs
Hanspeter Pfister's slides on visualization taxonomy

Chart Suggestions—A Thought-Starter

What would you like to show?

Comparison

- Among Items
  - Two Variables per Item
    - Variable Width Column Chart
  - Many Categories
    - Table or Table with Embedded Charts
  - One Variable per Item
    - Few Categories
      - Many Items
        - Bar Chart
      - Few Items
        - Column Chart
- Over Time
  - Many Periods
    - Cyclical Data
      - Circular Area Chart
    - Non-Cyclical Data
      - Line Chart
  - Few Periods
    - Single or Few Categories
      - Column Chart
    - Many Categories
      - Line Chart

Relationship

- Two Variables
  - Scatter Chart
- Three Variables
  - Bubble Chart

Distribution

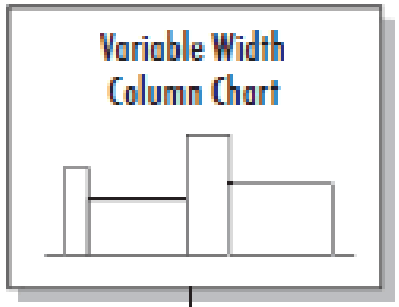
- Single Variable
  - Few Data Points
    - Column Histogram
  - Many Data Points
    - Line Histogram
- Two Variables
  - Scatter Chart
- Three Variables
  - 3D Area Chart

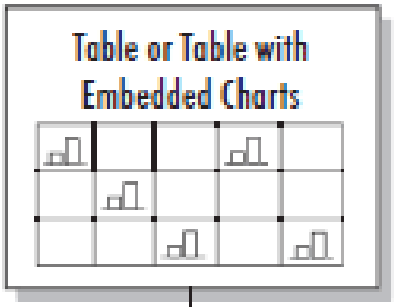
Composition

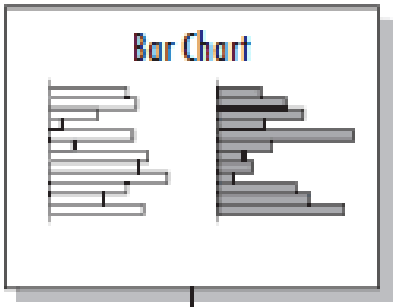
- Changing Over Time
  - Few Periods
    - Only Relative Differences Matter
      - Stacked 100% Column Chart
    - Relative and Absolute Differences Matter
      - Stacked Column Chart
  - Many Periods
    - Only Relative Differences Matter
      - Stacked 100% Area Chart
    - Relative and Absolute Differences Matter
      - Stacked Area Chart
- Static
  - Simple Share of Total
    - Pie Chart
  - Accumulation or Subtraction to Total
    - Waterfall Chart
  - Components of Components
    - Stacked 100% Column Chart with Subcomponents

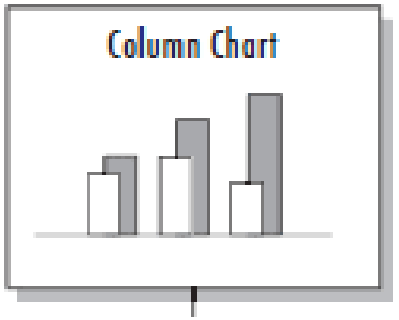
www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com

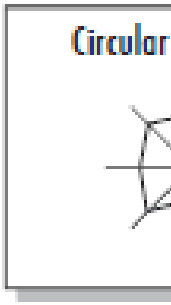
Chart Suggestions—A T











Two Variables per Item

Many Categories

Many Items

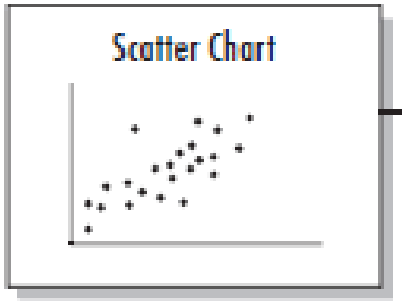
Few Items

Few Categories

One Variable per Item

Among Items

Comparison

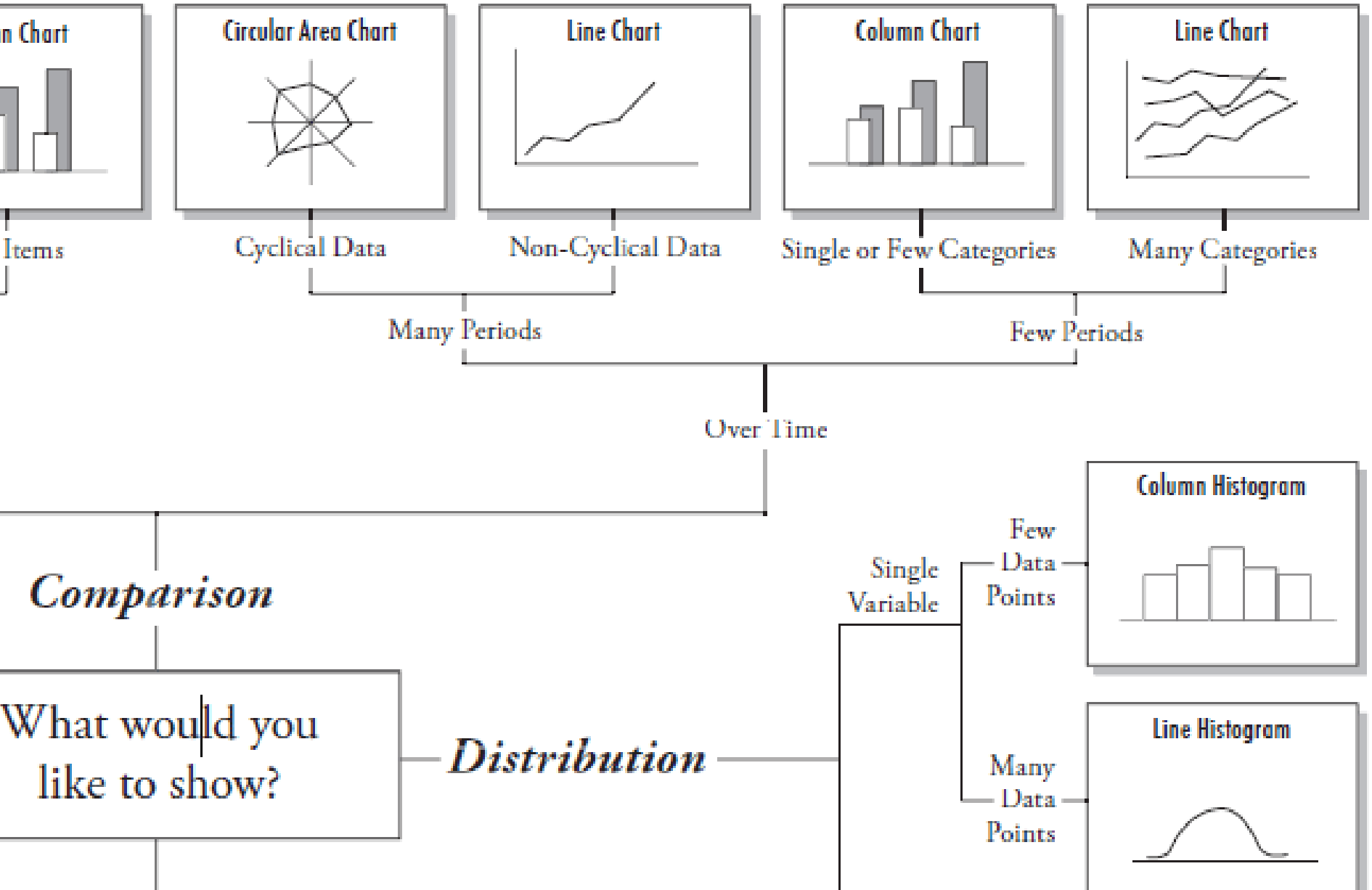


Two Variables

Relationship

What would yo
like to show?

ions—A Thought-Starter



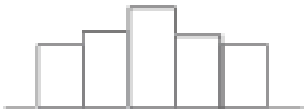
Over Time

Distribution

Single Variable

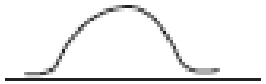
Few Data Points

Column Histogram



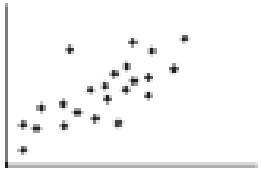
Many Data Points

Line Histogram



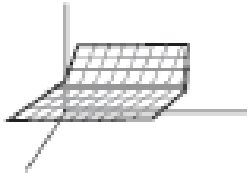
Two Variables

Scatter Chart



Three Variables

3D Area Chart



Static

hare

Accumulation or

Components

Composition

Two Variables

Scatter Cha

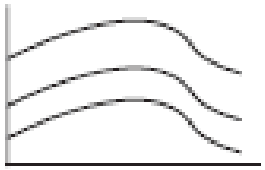
Three Variables

3D Area Cha

Static

Relative and Absolute Differences Matter

Stacked Area Chart

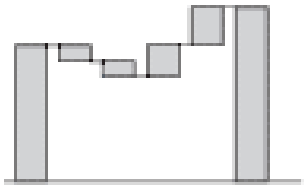


Simple Share of Total

Pie Chart

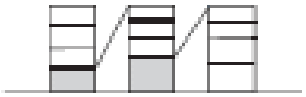
Accumulation or Subtraction to Total

Waterfall Chart



Components of Components

Stacked 100% Column Chart with Subcomponents

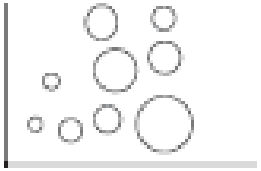


www.Extren
© 2009 A. Abela — a.

Relationship

like to show?

Bubble Chart



Three Variables

Composition

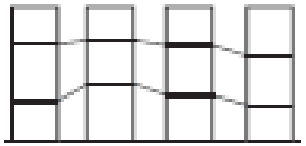
Changing Over Time

Few Periods

Many Periods

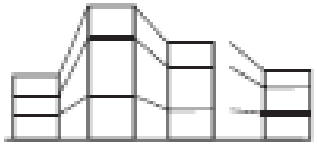
Only Relative Differences Matter

Stacked 100% Column Chart



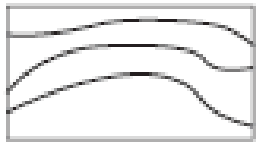
Relative and Absolute Differences Matter

Stacked Column Chart



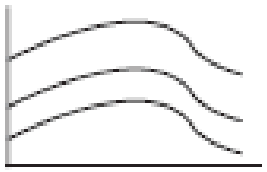
Only Relative Differences Matter

Stacked 100% Area Chart



Relative and Absolute Differences Matter

Stacked Area Chart



Simple
of To

Pie Ch

per Item

Categories

Few Categories

One Variable per Item

Among Items

Comp

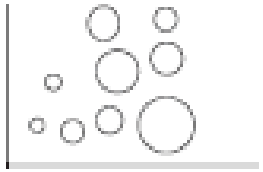
Scatter Chart

Two Variables

Relationship

What v
like t

Bubble Chart



Three Variables

Comp

Changing Over Time

Few Periods

Many Periods

Assignment 2

In this assignment, you will design a visualization for a small data set and provide a rationale for your design choices. The choices you make will demonstrate your understanding of the data, visual and encoding principles you have learned so far.

The data set is a collection of measurements related to the IITB's Million Solar Lamp project -- demographics of beneficiaries, and the assembly, distribution & repairs of solar lamps in the Jhauba Block, Jhauba District of Madhya Pradesh state.

The data are summarised in multiples tables in given report. Your challenge is to combine these data in one single visualization that can fit in a A3 size paper. Submit a short write-up (1 page), providing a rigorous rationale for your design decisions. Explain the visual encodings you used and why they are appropriate for the data.

The best visualization will be incorporated into the final reports and duly credited. Assignment Due on 7 Mar 2016, 11:59 pm.