

Data Visualization

Software Engineering Programme
David Johnson

1

Contents

- What is data visualization?
- Why bother?
- Design critiques
- Importance of design
- Visual Variables to Charts to Networks to High Dimensional Data
- Labs – Python charting using Bokeh

2

Visualization

noun vi-su-al-i-za-tion \,vi-zhə-wə-lə-'zā-shən, ,vi-zhə-lə-, ,vīzh-wə-lə-

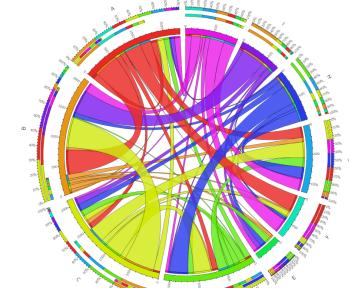
1. formation of mental visual images
2. the act or process of interpreting in visual terms or of putting into visible form
3. the process of making an internal organ or part visible by the introduction (as by swallowing) of a radiopaque substance followed by radiography

Source: Merriam-Webster dictionary

3

(Data) Visualization

To convey information through graphical representations of data

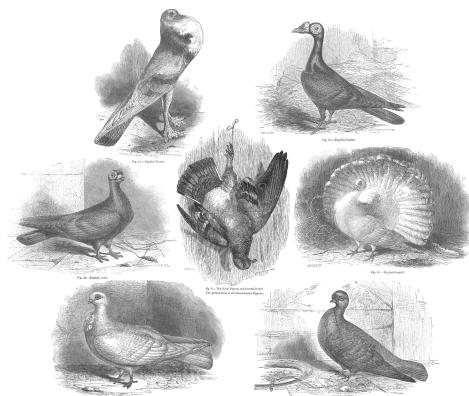


4

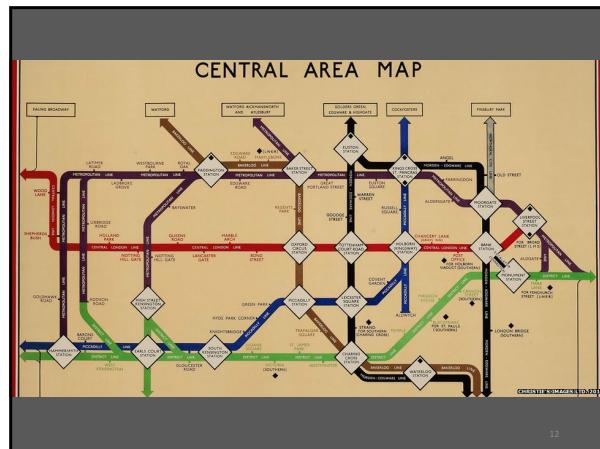
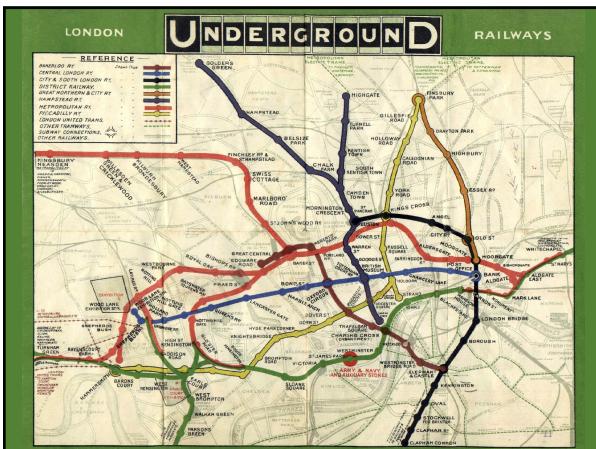
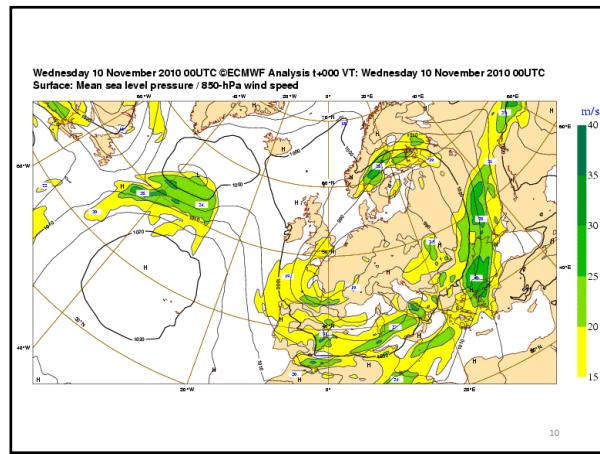
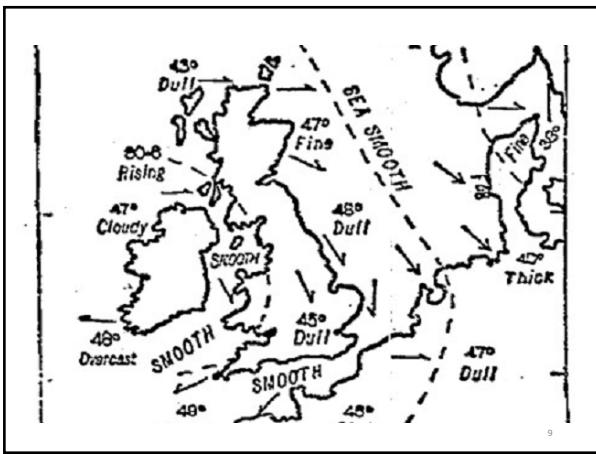
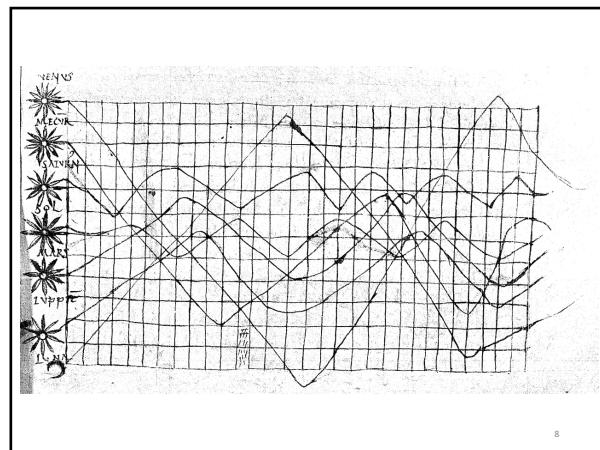
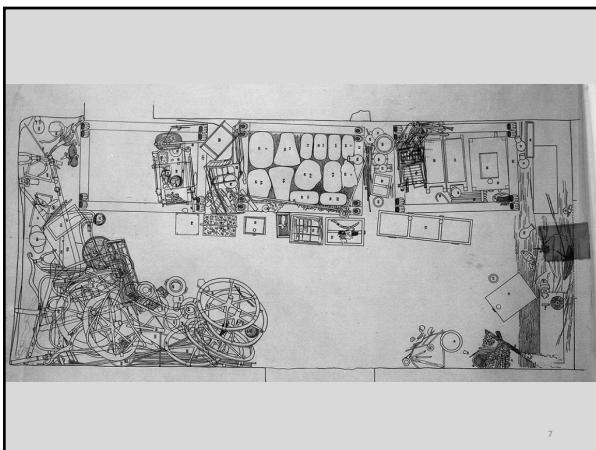
Goals of visualization

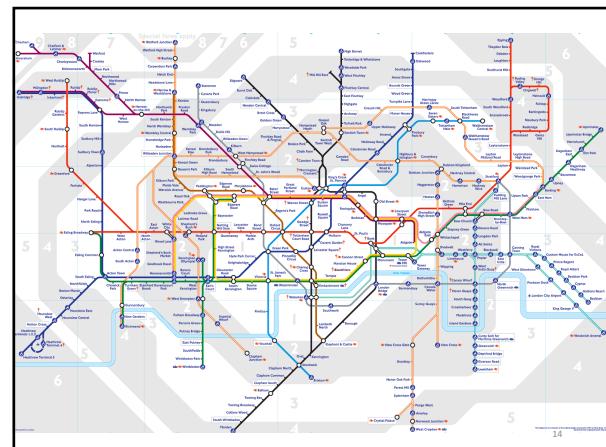
- **Record** information
- **Analyze** data to support decision making
- **Communicate** information to others

5



6

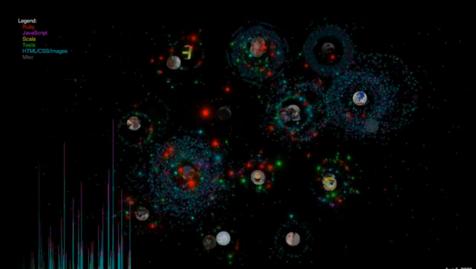




The Best Stats You Have Ever Seen



Storytelling

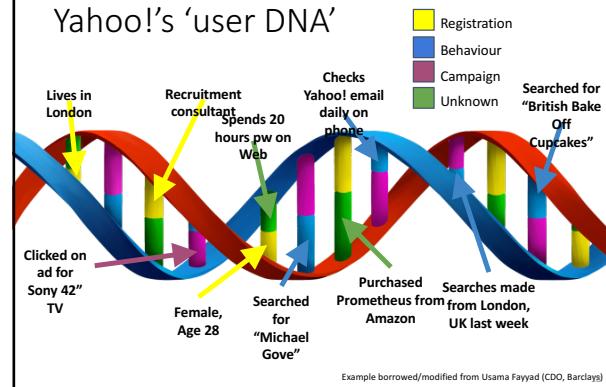


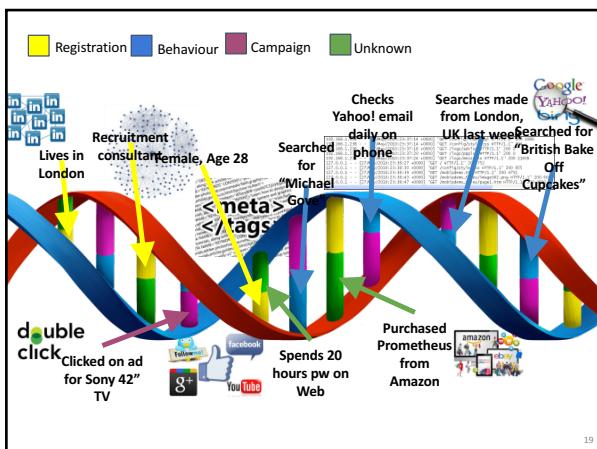
Why data visualization?

- Big data
- Limits to cognition/understanding
- Visual queries are easier to understand (even with summarized statistics, abstract visual cues are fast)

17

Yahoo!'s 'user DNA'





19

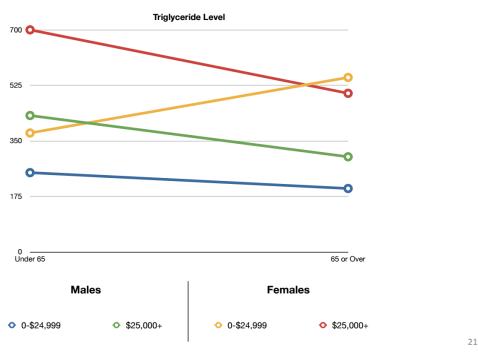
Limits to cognition

- Which gender or income level group shows different effects of age on cholesterol levels?

Income Group	Males		Females	
	Under 65	65 or Over	Under 65	65 or Over
0-\$24,999	250	200	375	550
\$25,000+	430	300	700	500

20

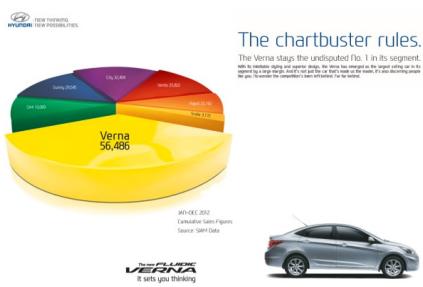
Limits to cognition



21

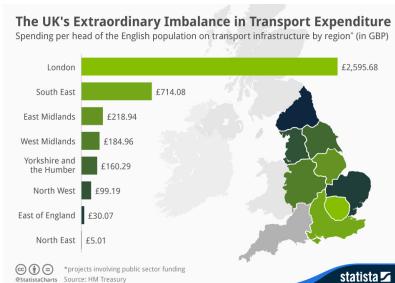
Critique

Critique



23

Critique



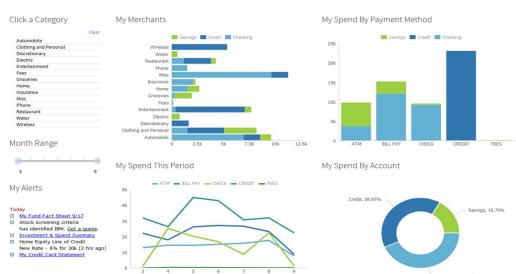
24

Critique



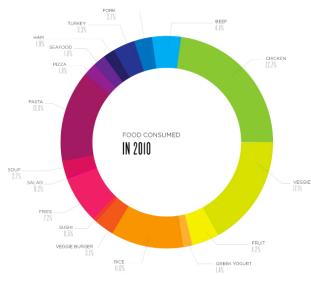
25

Critique



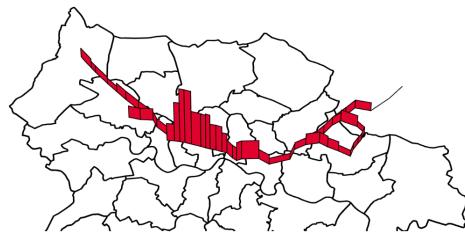
26

Critique



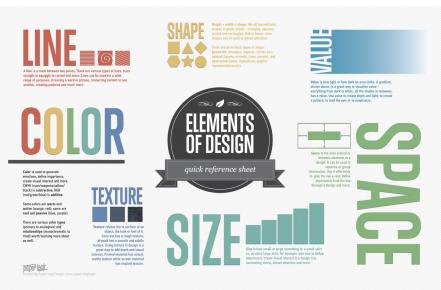
27

Critique



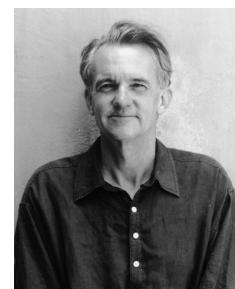
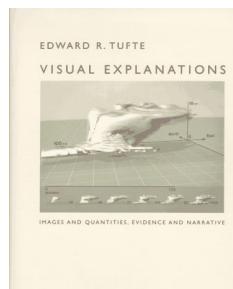
28

Importance of Design



29

Edward Tufte



30

Design Principles

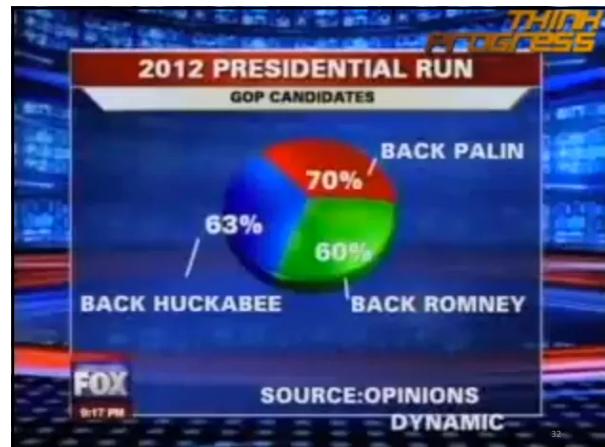
- Graphical integrity

- Always tell the truth about the data
- Use clear, detailed labels to defeat distortion and ambiguity
- Show data variation, not design variation

- Design aesthetics

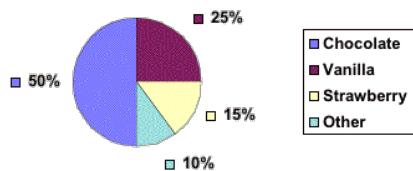
- Be clear and precise
- Maximize data-ink ratio (i.e. reduce redundant display of information)
- Avoid chartjunk
- Maximize data density

31

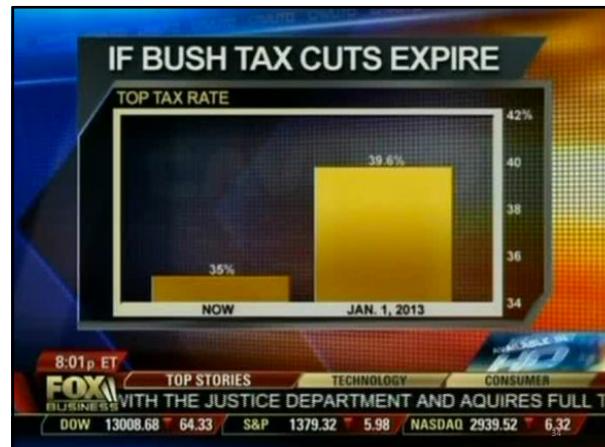


Graphical integrity

Favorite Ice Cream

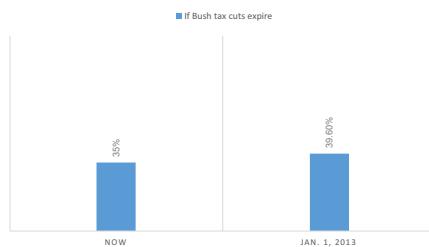


33



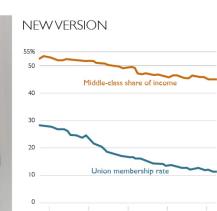
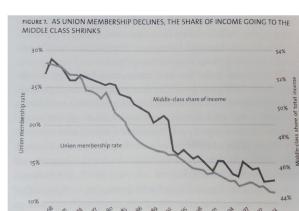
Graphical integrity

IF BUSH TAX CUTS EXPIRE

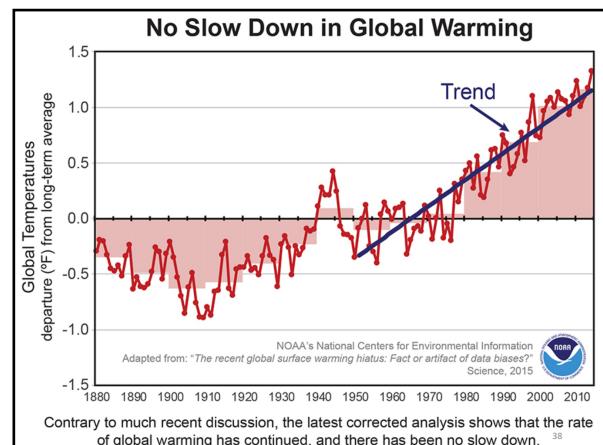


35

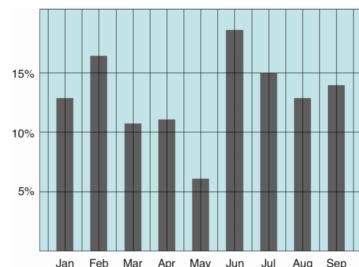
Scale distortions



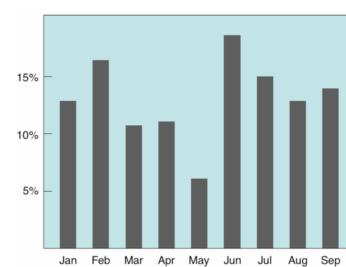
36



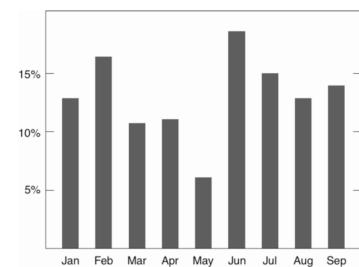
Avoid chartjunk



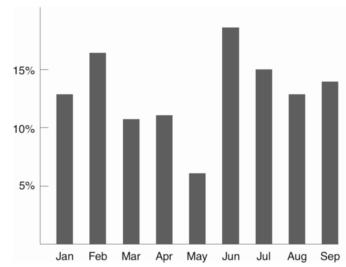
Avoid chartjunk



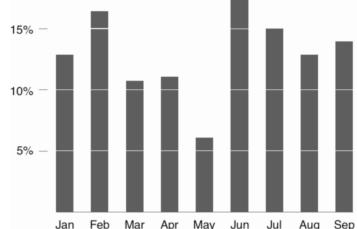
Avoid chartjunk



Avoid chartjunk

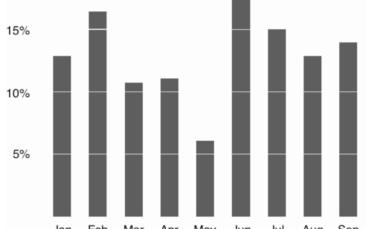


Avoid chartjunk



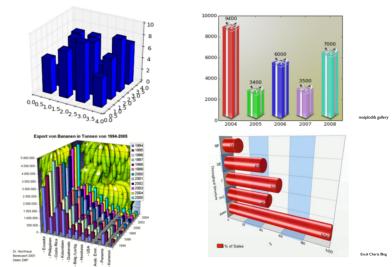
43

Avoid chartjunk



44

Examples of bad charts



45

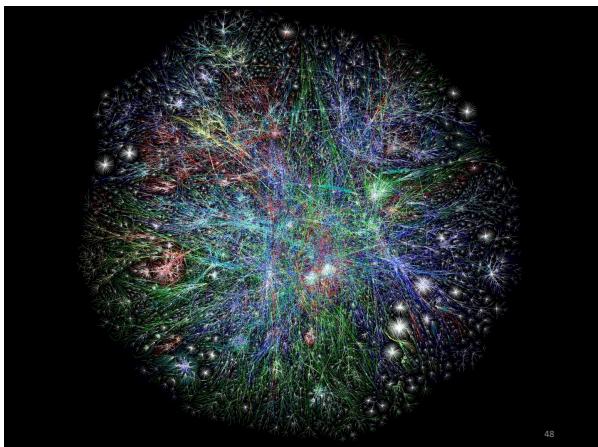
Visual Variables

- Basic model for stuff we measure (data):
 - From Science paper "On The Theory of Scales and Measurements", basically:
 - Nominal scale – categorization
 - Ordinal scale – continuous, comparable
 - Interval scale – kind of like ordinal but no true zero
 - Ratios – can measure ratios and proportions
- Also, entities, relationships and attributes
=> Jacques Bertin's Visual Variables

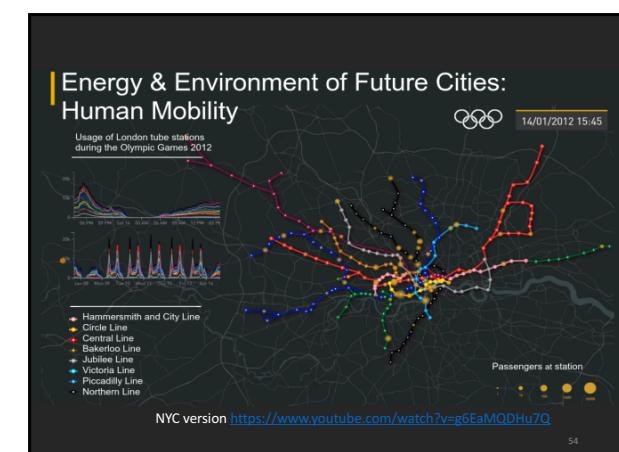
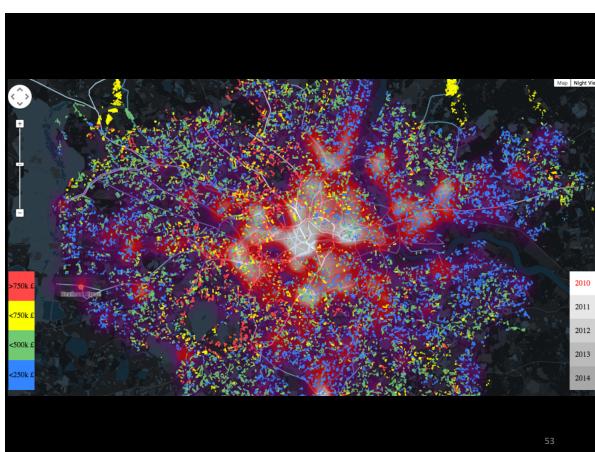
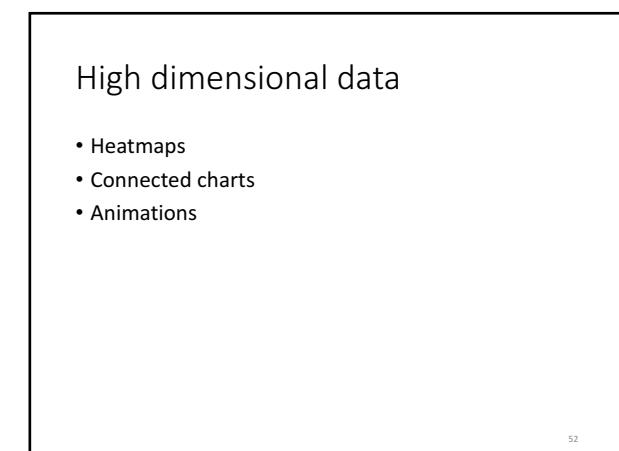
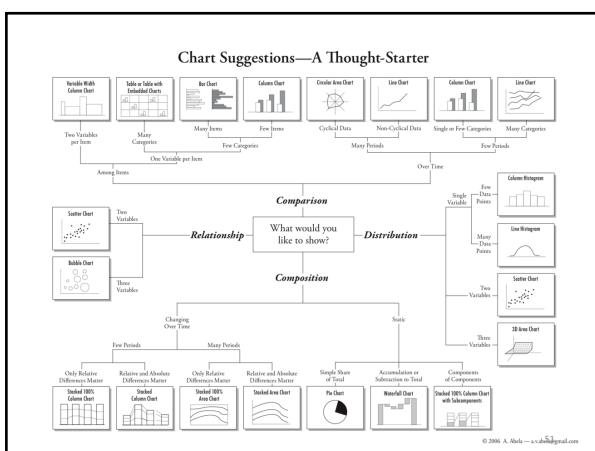
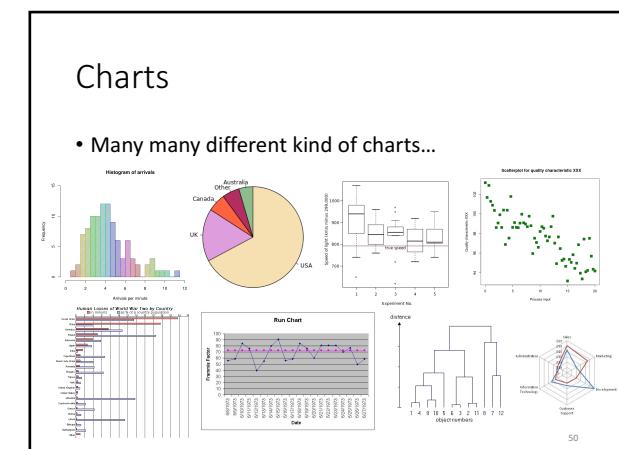
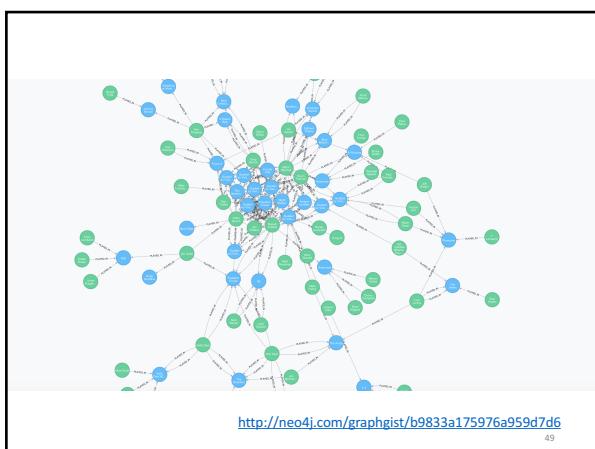
46

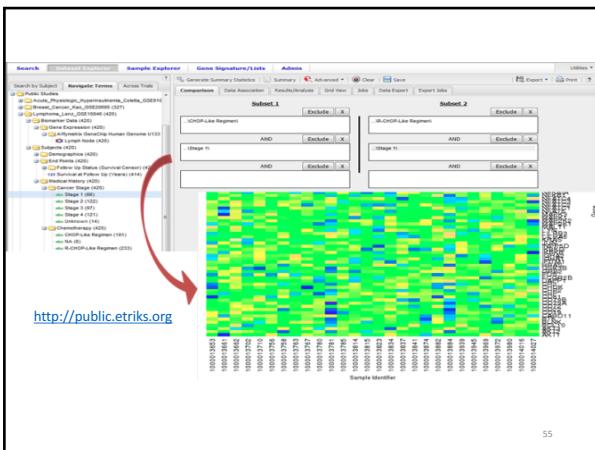
	Points	Lines	Areas	Best to show
Shape	● ▲ □	possible, but too weird to show	cartogram	qualitative differences
Size	● ● ●	cartogram	cartogram	quantitative differences
Color Hue	● ● ●	cartogram	cartogram	qualitative differences
Color Value	● ● ●	cartogram	cartogram	quantitative differences
Color Intensity	● ● ●	cartogram	cartogram	qualitative differences
Texture	● ● ●	cartogram	cartogram	qualitative & quantitative differences

47



48

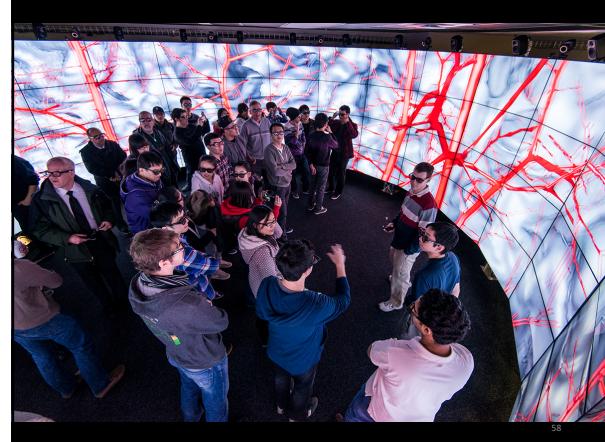


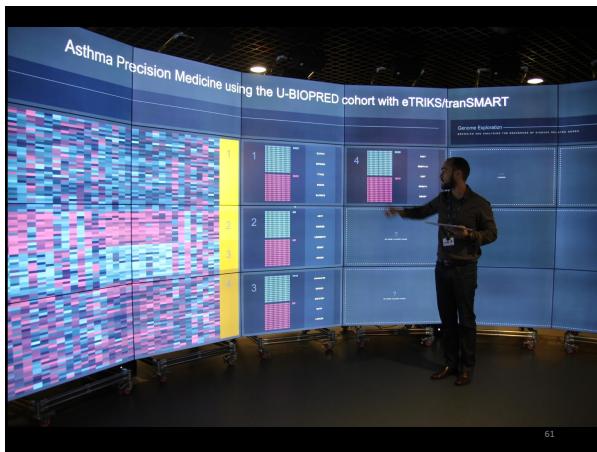


Visualization Systems

- Much visualization tooling is currently centered around how to present information on standard displays
 - Web
 - Desktop/laptop
 - Tablets/mobile
- What about on big devices/systems?

56





How does this fit in with the rest of CLO?

- Visualization could plug into different stages of your processing pipeline in data science/experimental exercises.
- Important for how to turn your big data analyses into summarized and presentable artefacts (typically for human consumption)
- Externalizing data representation into visual artefacts to use our visual cognition compliments our mind's cognitive abilities to extract meaning from data.

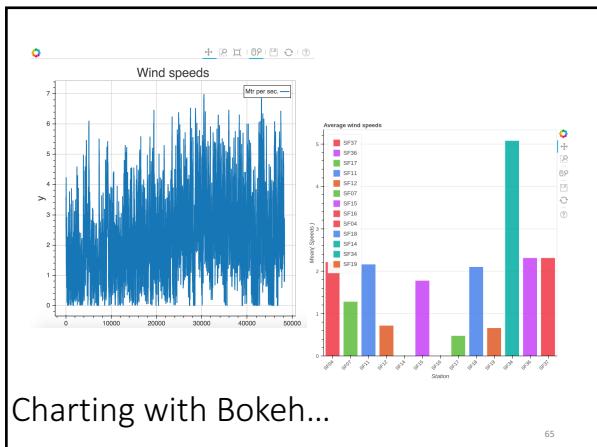
63

Visualization tools and frameworks

- Code/scripting frameworks
 - D3 (JS, very very widely used), bokeh, matplotlib (Python), R, and loads loads more....
- Tools
 - Tableau, Circos, MATLAB, Octave and more...

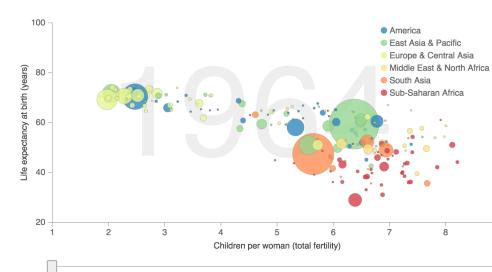
...lab time!

64



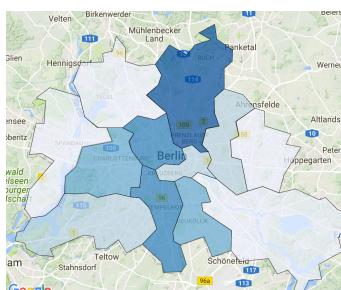
65

Bokeh can do advanced viz too...



66

Bokeh can even do maps...



67