Visualization Design

Maneesh Agrawala

CS 448B: Visualization
Fall 2017

Last Time: Data and Image Models

The big picture

task

data
- physical type
  - int, float, etc.
- abstract type
  - nominal, ordinal, etc.

domain
- metadata
- semantics
- conceptual model

processing algorithms

mapping
- visual encoding
- visual metaphor

image
- visual channel
- retinal variables

[based on slide from Munzner]

Nominal, ordinal and quantitative

On the theory of scales of measurements
S. S. Stevens, 1946

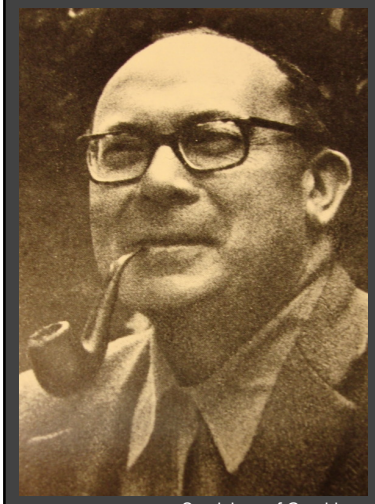
N - Nominal (labels)
Fruits: Apples, oranges, …
Operations: =, ≠

O - Ordered
Quality of meat: Grade A, AA, AAA
Operations: **=, ≠,** <, >, ≤, ≥

Q - Interval (location of zero arbitrary)
Dates: Jan, 19, 2006; Loc.: (LAT 33.98, LON -118.45)
Like a geometric point. Cannot compare directly
Only differences (i.e. intervals) may be compared
Operations: **=, ≠, <, >, ≤, ≥,** -

Q - Ratio (location of zero fixed)
Physical measurement: Length, Mass, Temp, …
Counts and amounts
Like a geometric vector, origin is meaningful
Operations: **=, ≠, <, >, ≤, ≥, -,** +

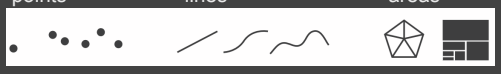
Marks and Visual Variables



Semiology of Graphics
J. Bertin, 1967

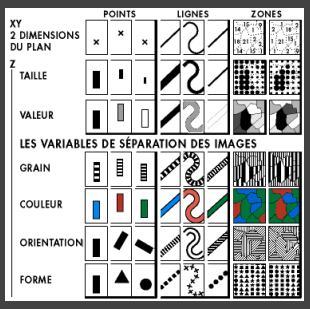
Marks: geometric primitives

points lines areas

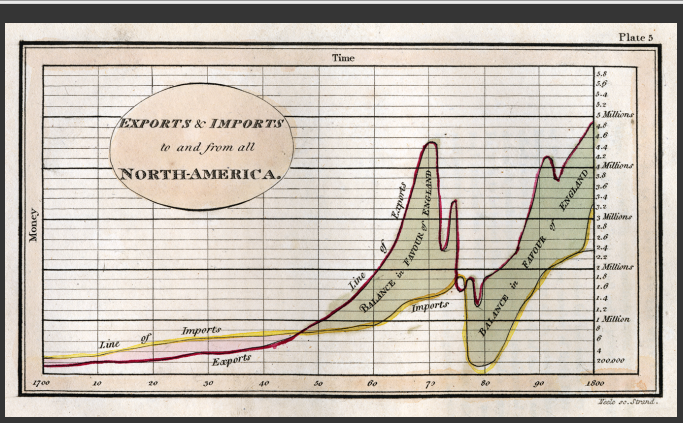


Visual Variables: control mark appearance

- Position (2x)
- Size
- Value
- Texture
- Color
- Orientation
- Shape



Playfair 1786/1801



- Time → x-position (Q, linear)
- Exports/Imports Values → y-position (Q, linear)
- Exports/Imports → color (N, O, nominal)
- Balance for/against → area (maybe length??) (Q, linear)
- Balance for/against → color (N, O, nominal)

Bertins' "Levels of Organization"

Position	N	O	Q
Size	N	O	Q
Value	N	O	Q
Texture	N	O	
Color	N		
Orientation	N		
Shape	N		

N Nominal
O Ordered
Q Quantitative

Note: Q < O < N

Note: Bertin actually breaks visual variables down into differentiating (≠) and associating (≡)

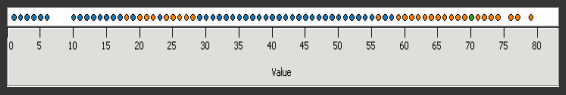
Mackinlay's expressiveness criteria

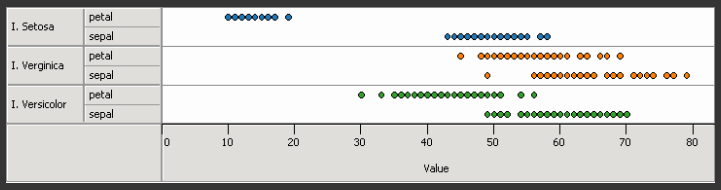
Expressiveness

A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express *all* the facts in the set of data, and *only* the facts in the data.

Cannot express the facts

A one-to-many (1 → N) relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position





Expresses facts not in the data

A length is interpreted as a quantitative value;

∴ Length of bar says something untrue about N data

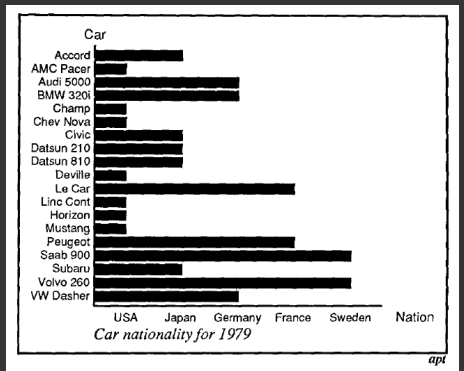


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[Mackinlay, APT, 1986]

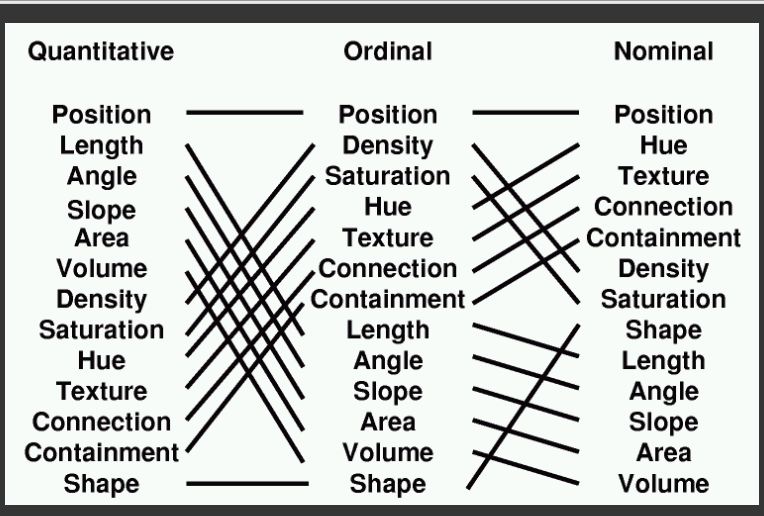
Mackinlay's effectiveness criteria

Effectiveness

A visualization is more effective than another visualization if the information conveyed by one visualization is more readily *perceived* than the information in the other visualization.

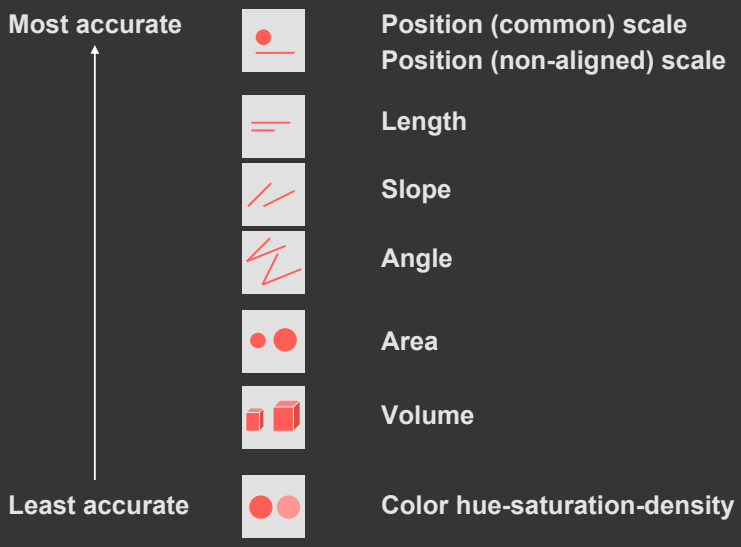
Subject of perception lecture

Mackinlay's ranking

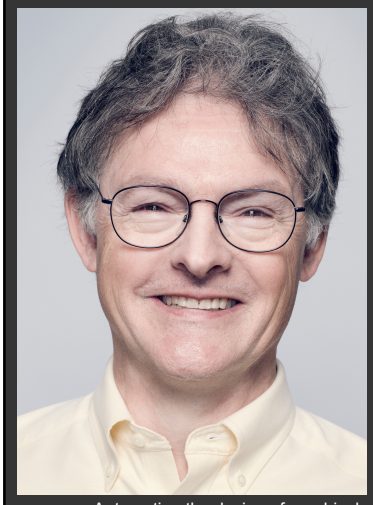


Conjectured *effectiveness* of the encoding

Graphical Perception



APT: Automatic Chart Construction



Automating the design of graphical presentation of relational information

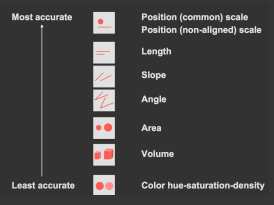
J. Mackinlay, 1986

Encode most important data using highest ranking visual variable for the data type

Year	Exports	Imports
1700	170,000	300,000
1701	171,000	302,000
1702	176,000	303,000
…	…	…

→

1. Year (Q)
2. Exports (Q)
3. Imports (Q)



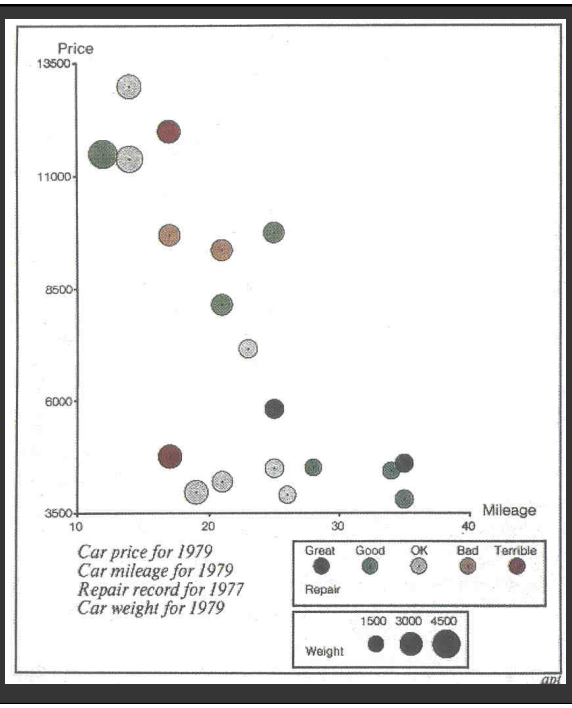
→

mark: lines

Year → x-pos (Q)

Exports → y-pos (Q)

Imports → y-pos (Q)



[Mackinlay, APT, 1986]

Limitations

Does not cover many visualization techniques

- Bertin and others discuss networks, maps, diagrams
- They do not consider 3D, animation, illustration, photography, …

Does not model interaction

Announcements

Announcements

Class participation requirements

- Complete readings before class
- In-class discussion
- Post at least 1 discussion substantive comment/question by noon the day after lecture (short paragraph)

Office hours on website

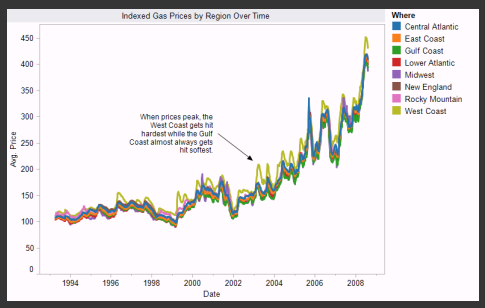
Class wiki
https://magrawala.github.io/cs448b-fa17

Assignment 2: Exploratory Data Analysis

Use Tableau to formulate & answer questions

First steps

- Step 1: Pick a domain
- Step 2: Pose questions
- Step 3: Find data
- Iterate



Create visualizations

- Interact with data
- Question will evolve
- Tableau

Make wiki notebook

- Keep record of all steps you took to answer the questions

Due before class on Oct 16, 2017

Assignment 1: Visualization Design

Barley Yield Data

In 1931 and 1932 Minnesota collected data on the yield in bushels per acre of 10 varieties of barley grown in 1/40 acre plots at University Farm, St. Paul, and at the five branch experiment stations located at Waseca, Morris, Crookston, Grand Rapids, and Duluth (all in Minnesota). The varieties were grown in three randomized blocks at each of the six stations during 1931 and 1932, different land being used each year of the test.

Number of records: 120
Variable Names:
Site: Crookston, Duluth, Grand Rapids, Morris, University Farm, Waseca
Variety: Glabron, Manchuria, No 457, No 462, No 475, Peatland, Svansota, Trebi, Velvet, Wisc. No 38
Yield: bushels/acre
Year: 1931, 1932

We've cleaned up this dataset and posted in csv format: barley2.csv

Barley Yields

Due by noon on Mon Oct 2

Submissions of PDF via Canvas, bring printout to class

Design Considerations

Expressiveness

- Do the mappings show the facts and only the facts?
- Are visual mappings consistent? (e.g., respect color mappings)

Effectiveness

- Are perceptually effective encodings used?
- Are the most important data mapped to the most effective visual variables?

Cognitive Load (Efficiency)

- Are there extraneous visual elements?

Data Transformation

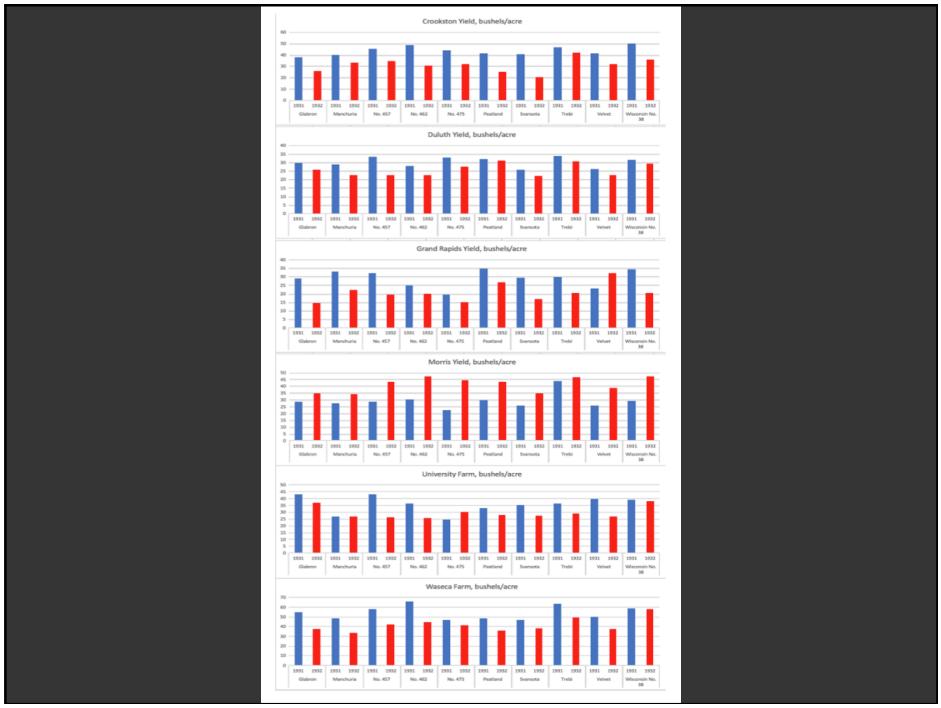
- Are transformations (filter, sort, derive, aggregate) appropriate?

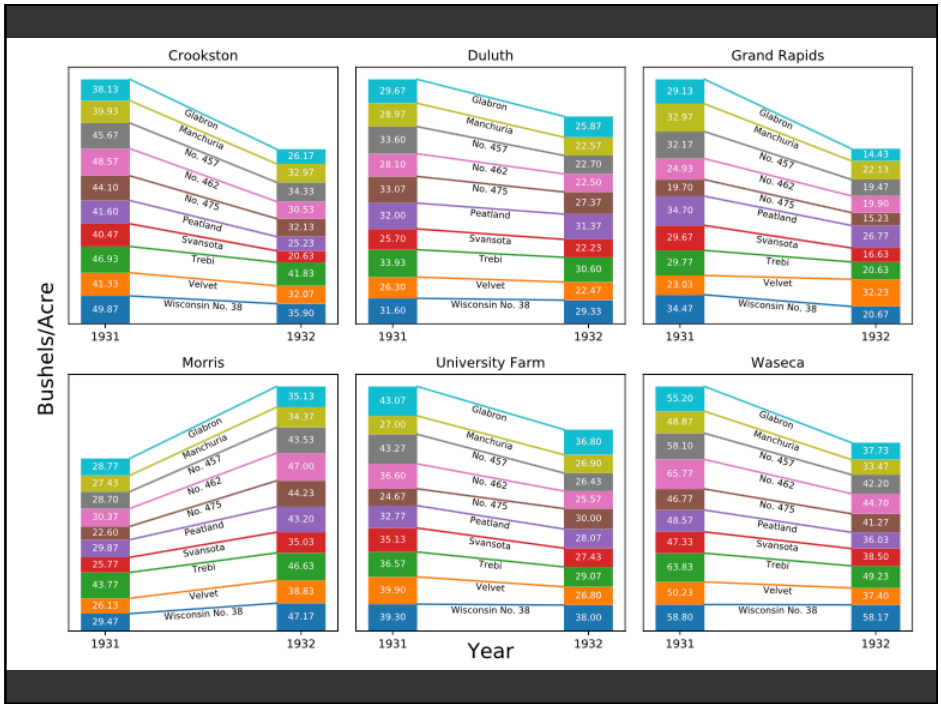
Guides (Non-Data Elements)

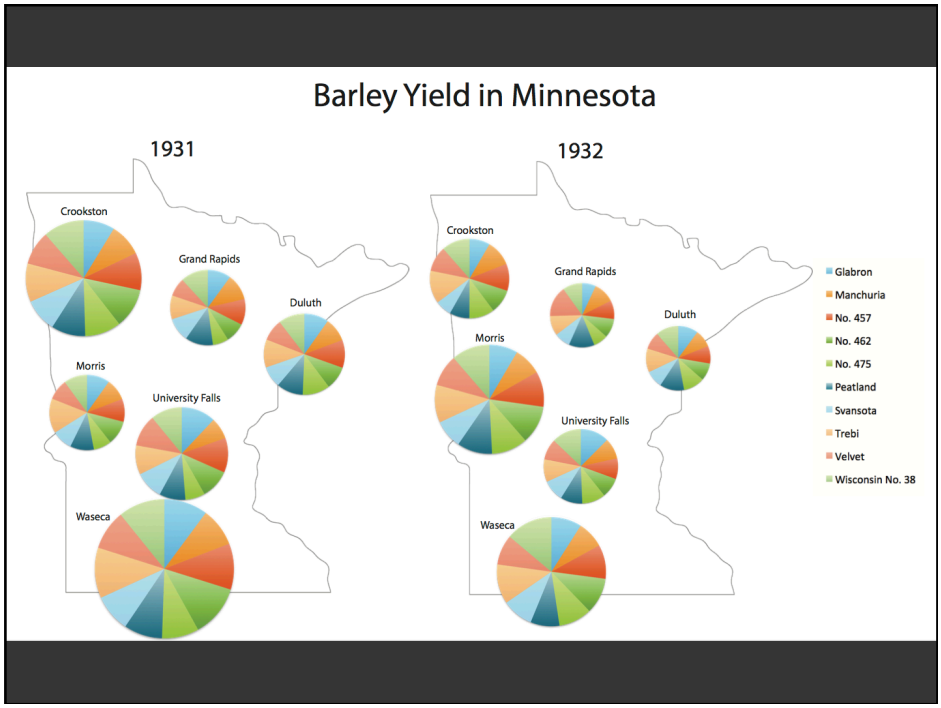
- Descriptive, consistent: Title, Label, Caption, Source, Annotations
- Meaningful references: Gridlines, Legend

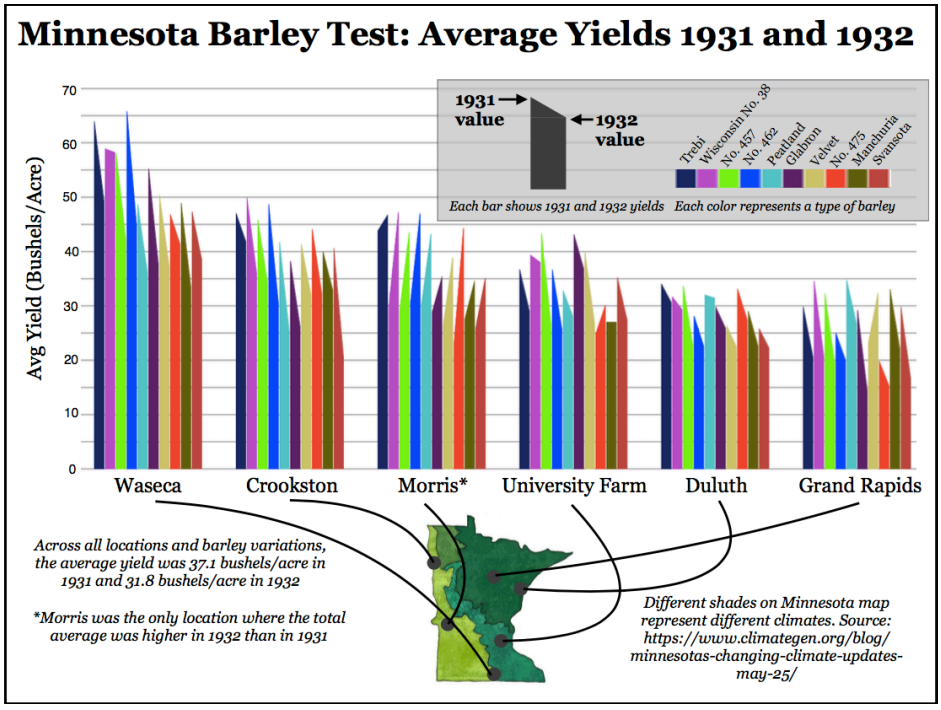
Design Space of A1 Submissions

Spatial Encoding	**Bar charts, Maps, Scatterplot, Pie**
Color Encoding	**Mostly ordered or nominal (year, loc.), Quantitative (dual encoding)**
Data Transformation	**Aggregation (avg. yield across variety)**
Labeling	**Title, Caption, Axis labels, Legends Not many annotations**


Crookston Yield, bushels/acre
Duluth Yield, bushels/acre
Grand Rapids Yield, bushels/acre
Morris Yield, bushels/acre
University Farm, bushels/acre
Waseca Farm, bushels/acre


Crookston
Duluth
Grand Rapids
Morris
University Farm
Waseca
Bushels/Acre
Year
1931
1932
Glabron
Manchuria
No. 457
No. 462
No. 475
Peatland
Svansota
Trebi
Velvet
Wisconsin No. 38


Barley Yield in Minnesota
1931
1932
Crookston
Grand Rapids
Duluth
Morris
University Falls
Waseca
Glabron
Manchuria
No. 457
No. 462
No. 475
Peatland
Svansota
Trebi
Velvet
Wisconsin No. 38


Minnesota Barley Test: Average Yields 1931 and 1932
Avg Yield (Bushels/Acre)
0
10
20
30
40
50
60
70
1931 value
1932 value
Each bar shows 1931 and 1932 yields
Each color represents a type of barley
Trebi
Wisconsin No. 38
No. 457
No. 462
Peatland
Glabron
Velvet
No. 475
Manchuria
Svansota
Waseca
Crookston
Morris*
University Farm
Duluth
Grand Rapids
Across all locations and barley variations, the average yield was 37.1 bushels/acre in 1931 and 31.8 bushels/acre in 1932
*Morris was the only location where the total average was higher in 1932 than in 1931
Different shades on Minnesota map represent different climates. Source: https://www.climategen.org/blog/minnesotas-changing-climate-updates-may-25/

In-Class Review

Procedure

Break into groups of 3

Present your visualization – in order by last name – 3 min each to describe what your visualization shows, and design choices.

Others should write down critique on sheet

We will keep time and tell you to switch

Critique in order by last name – rubric on next slide (~5 min each)

- Tell author your critique.
- Give critiques to author

Author take photos of critiques and add to A1 along with a short response (1 paragraph) to the feedback.

In-Class Review

Expressiveness

- Do the mappings show the facts and only the facts?
- Are visual mappings consistent? (e.g., respect color mappings)

Effectiveness

- Are perceptually effective encodings used?
- Are the most important data mapped to the most effective visual variables?

Cognitive Load (Efficiency)

- Are there extraneous visual elements?

Data Transformation

- Are transformations (filter, sort, derive, aggregate) appropriate?

Guides (Non-Data Elements)

- Descriptive, consistent: Title, Label, Caption, Source, Annotations
- Meaningful references: Gridlines, Legend