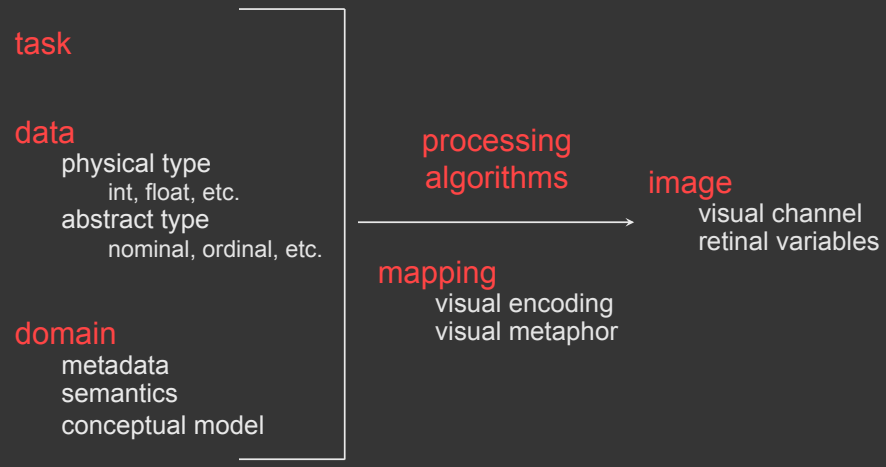
Visualization Designs

Maneesh Agrawala
Jessica Hullman

CS 294-10: Visualization
Fall 2014

Last Time: Data and Image Models

The big picture



[based on slide from Munzner]

Nominal, ordinal and quantitative

N - Nominal (labels)
- Operations: =, ≠

O - Ordered
- Operations: =, ≠, <, >, ≤, ≥

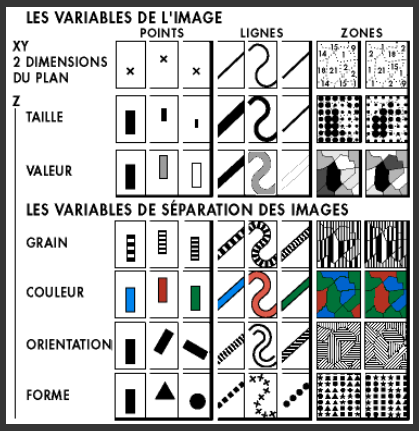
Q - Interval (Location of zero arbitrary)
- Operations: =, ≠, <, >, ≤, ≥, -
- Can measure distances or spans

Q - Ratio (zero fixed)
- Operations: =, ≠, <, >, ≤, ≥, -, ÷
- Can measure ratios or proportions

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

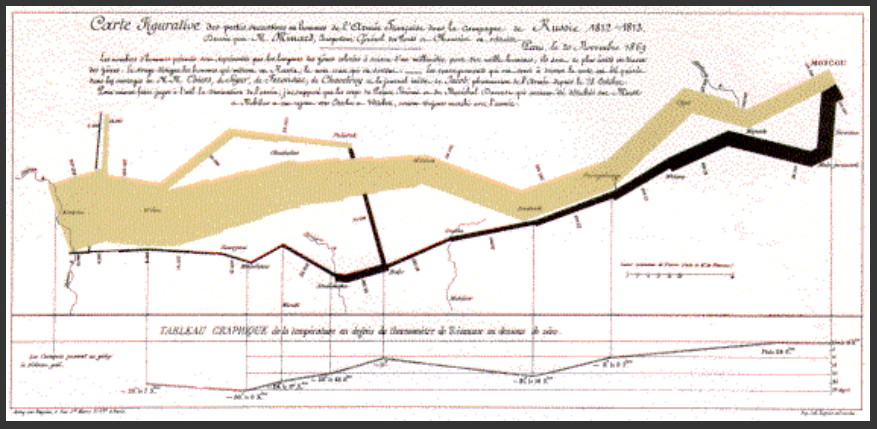
Visual variables

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

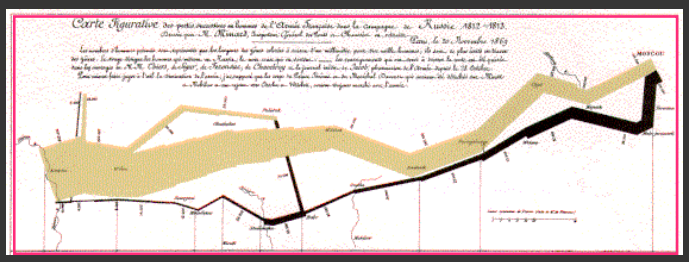


Note: Bertin does not consider 3D or time
Note: Card and Mackinlay extend the number of vars.

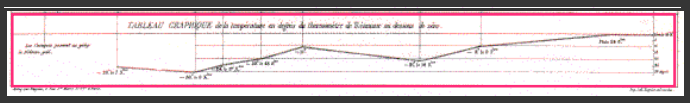
Minard 1869: Napoleon's march



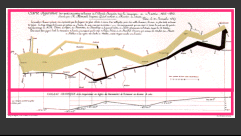
Single axis composition



+



=



[based on slide from Mackinlay]

Mark composition

temperature → y-position (Q, linear)

\+ longitude → x-position (Q, linear)

=

temp over longitude (Q x Q)

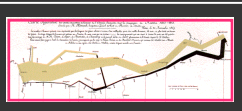
[based on slide from Mackinlay]

Mark composition

latitude → y-position (Q, linear)

\+ longitude → x-position (Q, linear)

\+ army size → width (Q, linear)

= 

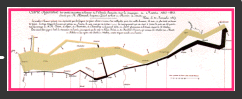
army position (Q x Q) and army size (Q)

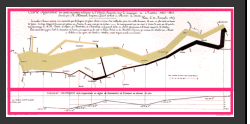
[based on slide from Mackinlay]

latitude (Q, lin)

longitude (Q, lin)

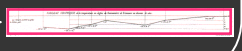
army size (Q, lin)





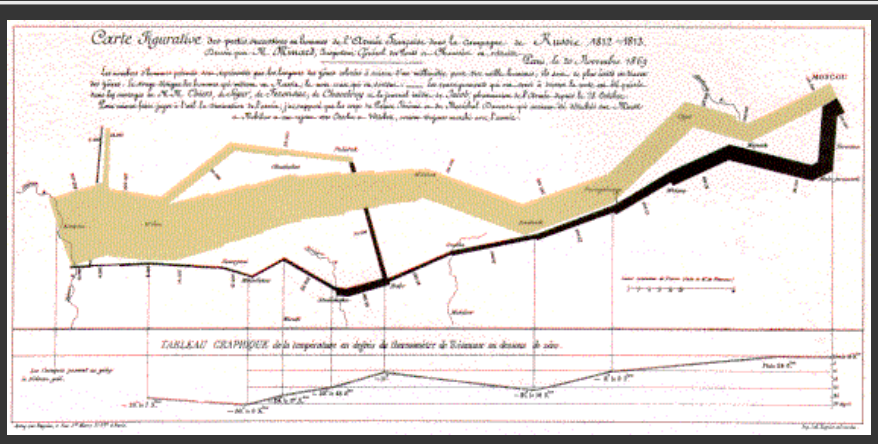
temperature (Q, lin)

longitude (Q, lin)



[based on slide from Mackinlay]

Minard 1869: Napoleon's march

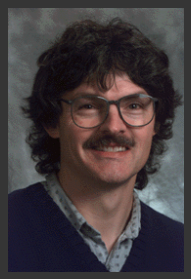


Depicts at least 4 quantitative variables

Any others?

Automated design

Jock Mackinlay's APT 86



Combinatorics of encodings

Challenge:

Assume 8 visual encodings and n data attributes

Pick the best encoding from the exponential number of possibilities $(n+1)^8$

Principle of Consistency:

The properties of the image (visual variables) should match the properties of the data

Principle of Importance Ordering:

Encode the most important information in the most effective way

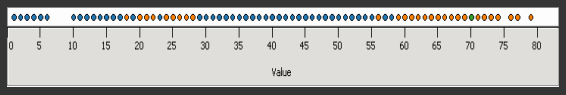
Mackinlay's expressiveness criteria

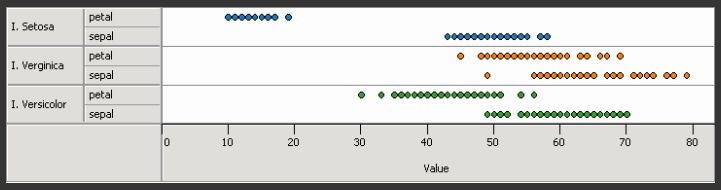
Expressiveness

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

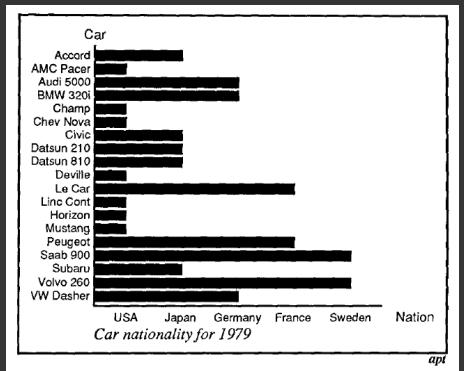


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[Mackinlay, APT, 1986]

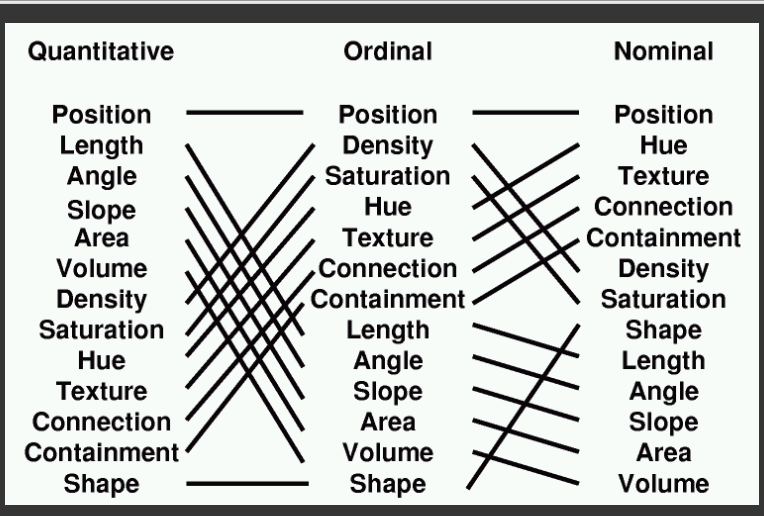
Mackinlay's effectiveness criteria

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

Announcements

Announcements

Auditors, *please* enroll in the class (1 unit, P/NP)

- Requirements: Come to class and participate (online as well)
- Requirements: Assignment 1

Class participation requirements

- Complete readings before class
- In-class discussion
- Post at least 1 discussion substantive comment/question by 11am on day of lecture

All, add yourself to participants page on the wiki

Class wiki
http://vis.berkeley.edu/courses/cs294-10-fa14/wiki/

Assignment 2: Exploratory Data Analysis

Use existing software to formulate & answer questions

First steps

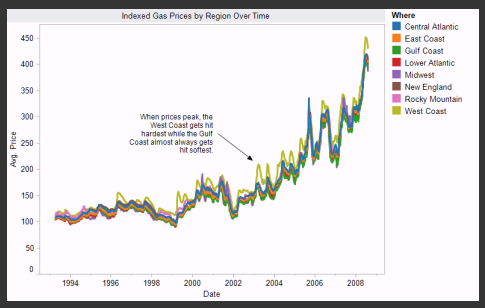
- Step 1: Pick a domain
- Step 2: Pose questions
- Step 3: Find data
- Iterate

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 Sep 29, 2014

Assignment 1: Visualization Design

Worldwide Disasters 1900 - 2008

	Full_date_start	Year_start	Month_start	Day_start	Full_date_end	Year_end	Month_end	Day_end	Country	Location	Type	Sub_Type	Name	Killed	Cost_dollars	Affected	Id
45	31051970	1970	5	31	31051970	1970	5	31	Peru	Chimbote	Earthquake (seismic	Earthquake (ground shaking)		66794	3216240	530	1970-0032
46	101942	1942	10		101942	1942	10		Bangladesh	West Sundar	Storm	Tropical cyclone		61000			1942-0008
47	1910	1910			1910	1910			China P Rep	Manchuria	Epidemic	Bacterial Infectio	Pneumonic	60000			1910-0001
48	1935	1935			1935	1935			India		Storm	Tropical cyclone		60000			1935-0015
49	31051935	1935	5	31	31051935	1935	5	31	Pakistan	Quetta (Balu	Earthquake (seismic	Earthquake (ground shaking)		60000			1935-0005
50	71949	1949	7		71949	1949	7		China P Rep		Flood			57000			1949-0025
51	11918	1918	1		11918	1918	1		Canada	Natiuonwide	Epidemic	Viral Infectious D	Spanish Influ	50000	2000000		1918-0015
52	81912	1912	8		81912	1912	8		China P Rep	Wenchou	Storm	Tropical cyclone		50000			1912-0005
53	101949	1949	10		101949	1949	10		Guatemala	East	Flood			40000		15	1949-0012
54	14101942	1942	10	14	14101942	1942	10	14	India	Orissa, West	Storm	Tropical cyclone		40000			1942-0009
55	21061990	1990	6	21	21061990	1990	6	21	Iran Islam Rep	Rasht, Astara	Earthquake (seismic	Earthquake (ground shaking)		40000	710000	8000	1990-0034
56	11051965	1965	5	11	11051965	1965	5	11	Bangladesh	Barisal distri	Storm	Tropical cyclone		36000	15600000	57.7	1965-0028
57	26122004	2004	12	26	26122004	2004	12	26	Sri Lanka		Earthquake (seismic	Tsunami		35399	1019306	1316.5	2004-0659
58	26121939	1939	12	26	26121939	1939	12	26	Turkey	Erzincan (Ani	Earthquake (seismic	Earthquake (ground shaking)		32962		20	1939-0010
59	1946	1946			1946	1946			Cape Verde Is	Countrywide	Drought	Drought		30000			1946-9004
60	24011939	1939	1	24	24011939	1939	1	24	Chile	Chillan regio	Earthquake (seismic	Earthquake (ground shaking)		30000	58500	920	1939-0001
61	81954	1954	8		81954	1954	8		China P Rep	Hopeh, Wuh	Flood			30000			1954-0014
62	8051902	1902	5	80	8051902	1902	5	80	Martinique		Volcano	Volcanic eruptior	Mount Pelee	30000			1902-0006
63	15121999	1999	12	15	20121999	1999	12	20	Venezuela	Federal distr	Flood	Flash flood		30000	483635	3160	1999-0547
64	13011915	1915	1	13	13011915	1915	1	13	Italy	Avezzano	Earthquake (seismic	Earthquake (ground shaking)		29980		60	1915-0002
65	71974	1974	7		71974	1974	7		Bangladesh		Flood			28700	38000000	579.2	1974-0034
66	26122003	2003	12	26	26122003	2003	12	26	Iran Islam Rep	Bam (Kerran	Earthquake (seismic	Earthquake (ground shaking)		26796	267628	500	2003-0630
67	16091978	1978	9	16	16091978	1978	9	16	Iran Islam Rep	Tabas (Khora	Earthquake (seismic	Earthquake (ground shaking)		25000	40000	50	1978-0115
68	7121988	1988	12	71	7121988	1988	12	71	Soviet Union	Armenia, Ler	Earthquake (seismic	Earthquake (ground shaking)		25000	1642000	14000	1988-0528
69	1920	1920			1920	1920			Cape Verde Is	Countrywide	Drought	Drought		24000			1920-9004

Design Considerations

Guides: Title, labels, legend, captions, source!

Expressiveness and Effectiveness

- Use encodings (visual variables) that match properties of the data
- Express the facts and only the facts
- Use perceptually effective encodings
- Choose encodings based on the importance of the data

Support comparison and pattern perception

- Between elements, to a reference line, or to counts

Design Considerations

Group / sort data by meaningful dimensions

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

- Are model choices (regression lines) appropriate?

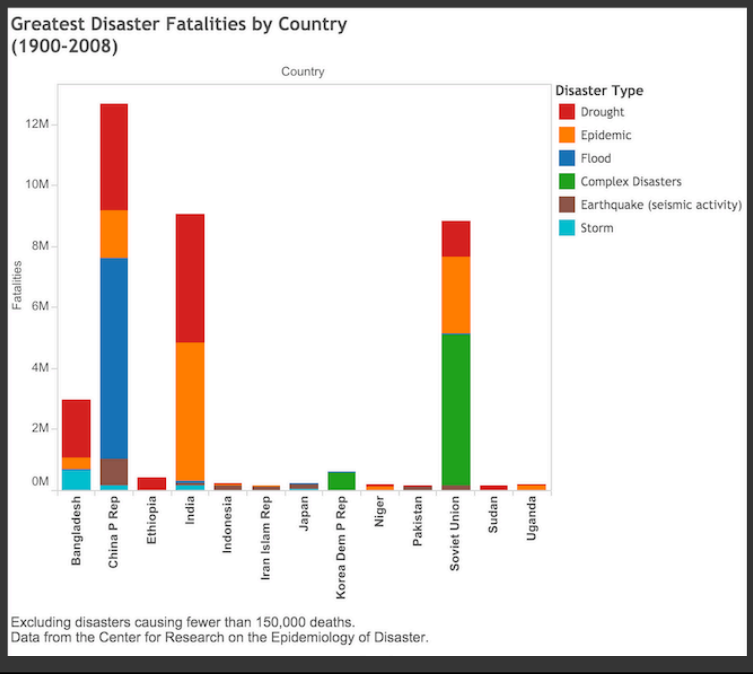
Reduce cognitive overhead

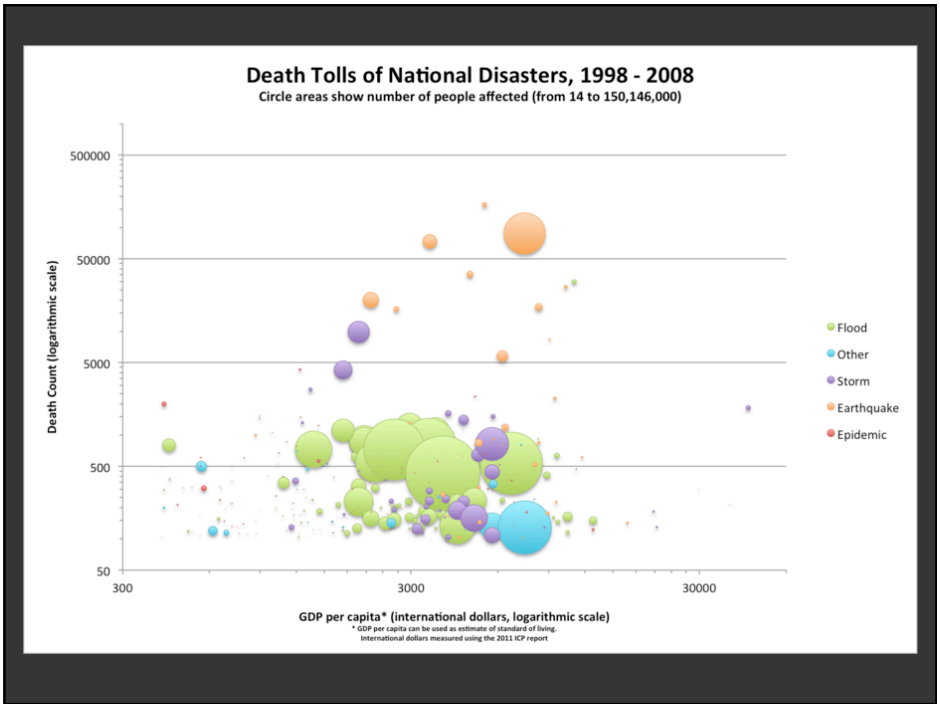
- Don't distract: faint gridlines, pastel highlights/fills
- Start minimal
- Minimize visual search, minimize ambiguity
  - Avoid legend lookups if direct labeling works
  - Avoid color mappings with indiscernible colors

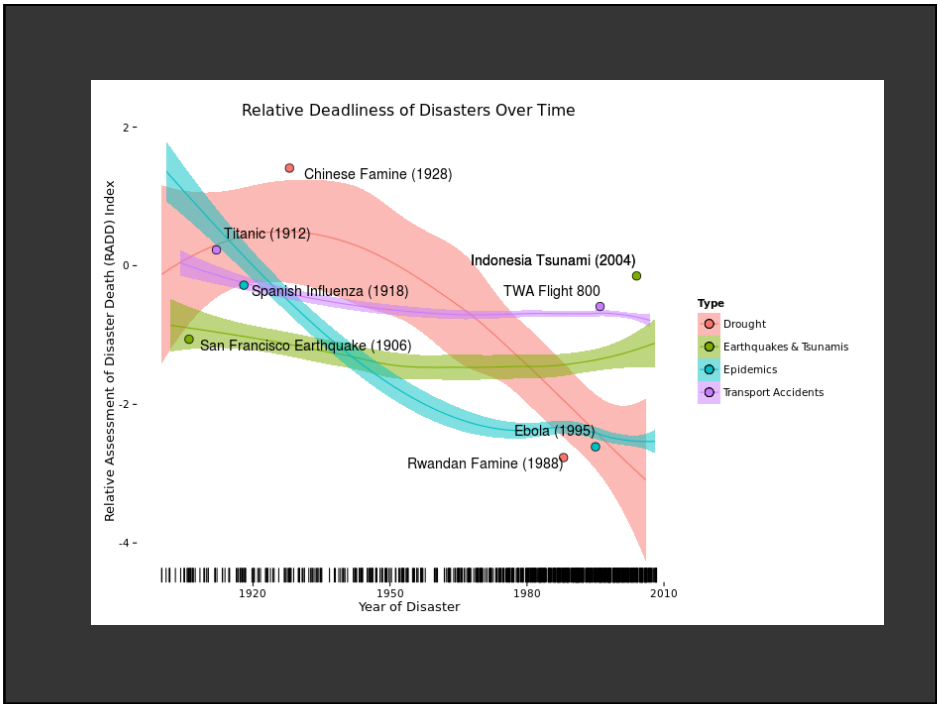
Be consistent! Visual inferences should consistently support data inferences

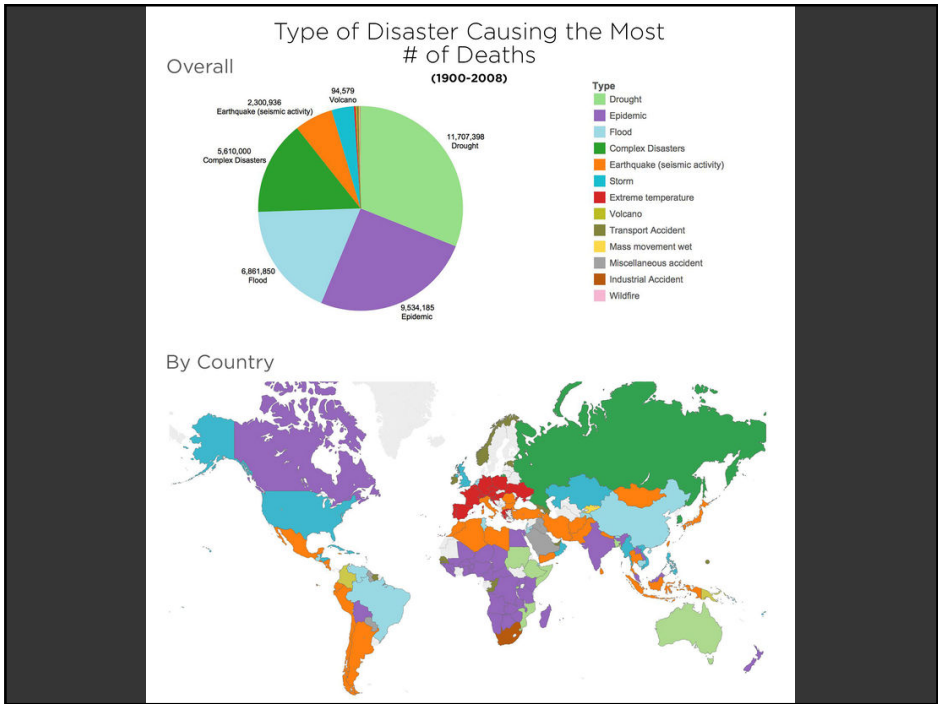
Design Space of A1 Submissions

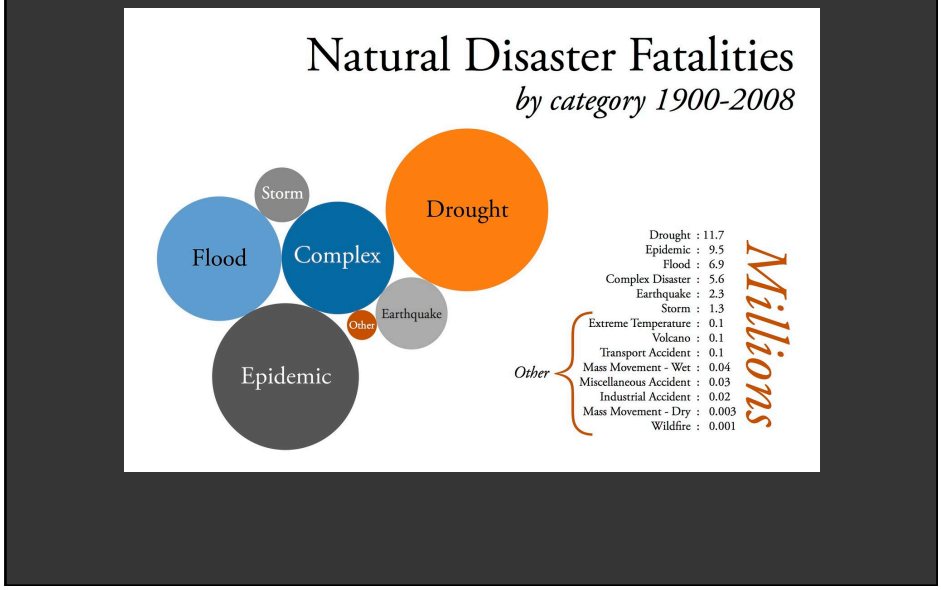
Spatial Encoding	Bar charts, Line charts, Area charts Scatterplots, Bubble charts, Maps
Color Encoding	Mostly nominal, Quantitative (maps)
Data Transformation	Often raw death counts or log scaled death counts
Labeling	Title, Caption, Axis labels Annotations, Many legends




Death Tolls of National Disasters, 1998 - 2008
Circle areas show number of people affected (from 14 to 150,146,000)
Death Count (logarithmic scale)
500000
50000
5000
500
50
300
3000
30000
GDP per capita* (international dollars, logarithmic scale)
* GDP per capita can be used as estimate of standard of living.
International dollars measured using the 2011 ICP report
Flood
Other
Storm
Earthquake
Epidemic


Relative Deadliness of Disasters Over Time
Relative Assessment of Disaster Death (RADD) Index
2
0
-2
-4
Chinese Famine (1928)
Titanic (1912)
Indonesia Tsunami (2004)
Spanish Influenza (1918)
TWA Flight 800
San Francisco Earthquake (1906)
Ebola (1995)
Rwandan Famine (1988)
1920
1950
1980
2010
Year of Disaster
Type
Drought
Earthquakes & Tsunamis
Epidemics
Transport Accidents


Type of Disaster Causing the Most # of Deaths
(1900-2008)
Overall
94,579 Volcano
2,300,936 Earthquake (seismic activity)
11,707,398 Drought
5,610,000 Complex Disasters
6,861,850 Flood
9,534,185 Epidemic
Type
Drought
Epidemic
Flood
Complex Disasters
Earthquake (seismic activity)
Storm
Extreme temperature
Volcano
Transport Accident
Mass movement wet
Miscellaneous accident
Industrial Accident
Wildfire
By Country


Natural Disaster Fatalities
by category 1900-2008
Storm
Drought
Flood
Complex
Earthquake
Other
Epidemic
Millions
Drought : 11.7
Epidemic : 9.5
Flood : 6.9
Complex Disaster : 5.6
Earthquake : 2.3
Storm : 1.3
Other
Extreme Temperature : 0.1
Volcano : 0.1
Transport Accident : 0.1
Mass Movement - Wet : 0.04
Miscellaneous Accident : 0.03
Industrial Accident : 0.02
Mass Movement - Dry : 0.003
Wildfire : 0.001

In-Class Review

Procedure

Break into groups of 3 (assigned by me)
Appoint a time keeper
Take turns showing your visualization – present findings (~3 min each)
Then critique – rubric on next slide (~5 min each)

- Get feedback from everyone in group
- Author must take notes (post critique notes/feedback to wiki after class)

Write-up of critique will be used in grading

In-Class Review Rubric

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 (unmapped) 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