Information Visualization Aggregate & Filter 2

Tamara Munzner

Department of Computer Science University of British Columbia

Lect 19, 17 Mar 2020

https://www.cs.ubc.ca/~tmm/courses/436V-20

News

- Midterm marks and solutions released
- -Gradescope has detailed breakdown, note stats are wrt total of 75 - Canvas has percentages, mean was 79%
- -solutions have detailed rubric w/ answer alternatives & explanations
- MI marks released
- -we specifically suggest meet to discuss during labs or office hrs to several teams
- P3 marks released
- -bimodal distribution

Spatial aggregation

- MAUP: Modifiable Areal Unit Problem
- -changing boundaries of cartographic regions can yield dramatically different results -zone effects

[http://www.e-education.ţ	u/edu/geog486/14_p7.html, Fig 4.cg.6]	





Clustering

- classification of items into similar bins
- -based on similiarity measure
- Euclidean distance, Pearson correlation
- -partitioning algorithms
- divide data into set of bins
- # bins (k) set manually or automatically
- -hierarchical algorithms
- produce "similarity tree" (dendrograms): cluster hierarchy
- agglomerative clustering: start w/ each node as own cluster, then iteratively merge
- cluster hierarchy: derived data used w/ many dynamic aggregation idioms
- -cluster more homogeneous than whole dataset • statistical measures & distribution more meaningful



- Online lectures and office hours start today, using Zoom: https://zoom.us/j/9016202871
- Lecture mode
- -Plan: I livestream with video + audio + screenshare, will also try recording.
- You'll be able to just join the session
- -Please connect audio-only, no video, to avoid congestion
- You'll be auto-muted. If you have a question use the Show Hand (click on Participants, button is at the bottom of the popup window), I'll unmute you myself
- Office hours mode

PI-P3 marks

All Programming Assignments

increasingly bimodal

-Please do connect with video if possible, in addition to audio

Gerrymandering: MAUP for political gain

2. Compact.

5 blue districts.

0 red districts

BLUE WINS

-I'll use the Waiting Room feature, where I will individually allow you in • If I'm already talking to somebody else I'll briefly let you know, then put you back in WR until it's your turn.





https://www.washingtonpost.com/news/wonk/wp/2015/03/01/this-is-the-best-explanation-ofgerrymandering-you-will-ever-see/

2 blue districts

3 red district

RED WINS

Neither compact

E

Idiom: GrouseFlocks

OST.COM/WONKBLOG

Gerrymandering, explained

60% blue, 40% red 1. Perfect

3 blue districts, 2 red districts

BLUE WINS

- data: compound graphs network -cluster hierarchy atop it
- derived or interactively chosen
- visual encoding
- connection marks for network links
- containment marks for hierarchy
- -point marks for nodes
- dynamic interaction
- -select individual metanodes in hierarchy to expand/ contract



A real district in Pennsylvania

18 house seats





IEEE Visualization Conference (Vis '99), pp. 43-50, 1999.]

ts	Dimensionality reduction & visualization
Out Labels for clusters Why?	 why do people do DR? -improve performance of downstream algorithm avoid curse of dimensionality -data analysis -data analysis if look at the output: visual data analysis abstract tasks when visualizing DR data dimension-oriented tasks naming synthesized dims, mapping synthesized dims to original dims cluster-oriented tasks verifying clusters, naming clusters, matching clusters and classes
19	Sequences. Brenimer, Sealmair, Ingram, and Munzner. Proc. BELIV 2014.J
each axis	 Nonlinear dimensionality reduction pro: can handle curved rather than linear structure cons: lose all ties to original dims/attribs new dimensions often cannot be easily related to originals mapping synthesized dims to original dims task is difficult many techniques proposed many literatures: visualization, machine learning, optimization, psychology, techniques: t-SNE, MDS (multidimensional scaling), charting, isomap, LLE, t-SNE: excellent for clusters but some trickiness remains: http://distill.pub/2016/misread-tsne/ minimize stress or strain metrics early formulations equivalent to PCA
	Linear DR
rials (spheres)	 first try: PCA (linear) result: error falls off sharply after ~45 dimensions orgection problem: physically impossible intermediate points when simulating new materials specular highlights cannot have holes!

Embed

