### Data Visualization (CSCI 627/490)

Aggregation & Focus+Context

Dr. David Koop





### Overview: Reducing Items & Attributes



D. Koop, CSCI 627/490, Fall 2022



# → Aggregate

### → Items







[Munzner (ill. Maguire), 2014]



Northern Illinois University



## Item Filtering on Maps



#### D. Koop, CSCI 627/490, Fall 2022

Northern Illinois University







## Attribute Filtering on Star Plots



(c)

#### D. Koop, CSCI 627/490, Fall 2022

(d)









## Aggregation: Histograms

- Very similar to bar charts

- Often shown without space between (continuity)
- Choice of number of bins
  - Important!
  - Viewers may infer different trends based on the layout







# Binning

- Hexbin advantages:
  - Bins are more circular so distance to the edge is not as variable
  - More efficient aggregation around the center of the bin



### • 2D Histogram is a histogram in 2D encoded using color instead of height



Hexagonal Bin









## Spatial Aggregation







[Penn State, GEOG 486]

# Modifiable Areal Unit Problem

- of aggregation you get
- Similar to bins in histograms
- Gerrymandering



D. Koop, CSCI 627/490, Fall 2022

[Wonkblog, Washington Post, Adapted from S. Nass]



Northern Illinois University







# Drawing Different Maps: Compactness

#### borders

specifically the 2018 midterms – based on historical patterns since 2006













# Drawing Different Maps











### Boxplots

- Show **distribution**
- Single value (e.g. mean, max, min, quartiles) doesn't convey everything
- Created by John Tukey
- Show spread and skew of data
- Best for **unimodal** data
- Variations like vase plot for multimodal data
- Aggregation here involves many different marks

#### D. Koop, CSCI 627/490, Fall 2022





## Aggregation: Boxplots



D. Koop, CSCI 627/490, Fall 2022



Northern Illinois University



# Four Distributions, Same Boxplot...



D. Koop, CSCI 627/490, Fall 2022

Northern Illinois University





# Project Design

- Feedback:
  - Data Manipulation?
  - Questions lead, not technique!
  - Be creative! (interaction too) <u>https://xeno.graphics</u>
- Work on turning your visualization ideas into designs
- Turn in:
  - Two Design Sketches (like sheets 2-4 from 5 Sheet Design)
  - One Bad Design Sketch (like sheets 2-4: here, justify why bad)
  - Progress on Implementation
- Due next Wednesday

#### D. Koop, CSCI 627/490, Fall 2022

### ://xeno.graphics deas into designs

### 2-4 from 5 Sheet Design) ts 2-4: here, justify why bad)





### Assignment 5

- Focus on Multiple Views and Interaction
- Soon...





### No Class Tuesday

#### Vote!





### K-Means



#### D. Koop, CSCI 627/490, Fall 2022

### <u>Run</u>







### K-Means Issues



#### D. Koop, CSCI 627/490, Fall 2022



Number of Clusters





# Dimensionality Reduction

- individual attribute
- Example: Understanding the language in a collection of books
  - Count the occurrence of each non-common word in each book
  - (e.g. "western")
  - Don't want to have to manually determine such rules
- techniques

#### D. Koop, CSCI 627/490, Fall 2022

• Attribute Aggregation: Use fewer attributes (dimensions) to represent items • Combine attributes in a way that is more instructive than examining each

- Huge set of features (attributes), want to represent each with an aggregate feature (e.g. high use of "cowboy", lower use of "city") that allows clustering

Techniques: Principle Component Analysis, Multidimensional Scaling family of









# Principle Component Analysis (PCA)

original data space



D. Koop, CSCI 627/490, Fall 2022



PC 1













### PCA



[Principle Component Analysis Explained, Explained Visually, V. Powell & L. Lehe, 2015]









# 17 dimensions to 2

Alcoholic drinks Beverages Carcase meat Cereals Cheese Confectionery Fats and oils Fish Fresh fruit Fresh potatoes Fresh Veg Other meat Other Veg Processed potatoes Processed Veg Soft drinks Sugars

England	N Ireland	Scotland	
375	135	458	
57	47	53	
245	267	242	
1472	1494	1462	
105	66	103	
54	41	62	
193	209	184	
147	93	122	
1102	674	957	
720	1033	566	
253	143	171	
685	586	750	
488	355	418	
198	187	220	
360	334	337	
1374	1506	1572	
156	139	147	

Wales

England

Scotland

D. Koop, CSCI 627/490, Fall 2022



[Principle Component Analysis Explained, Explained Visually, V. Powell & L. Lehe, 2015]





Northern Illinois University







## Non-linear Dimensionality Reduction



### original data space $\mathcal{X}$

D. Koop, CSCI 627/490, Fall 2022



### component space Z











## Dimensionality Reduction in Visualization



#### D. Koop, CSCI 627/490, Fall 2022

[Glimmer, Ingram et al., 2009]



Northern Illinois University









# Tasks in Understanding High-Dim. Data



#### D. Koop, CSCI 627/490, Fall 2022

[Munzner (ill. Maguire), 2014]











## Probing Projections











# Probing Projection Goals

- Examining the Projection
- Exploring the Data
- Design Goals:
  - Show and correct approximation errors
  - Allow for multi-level comparisons
  - Spatial orientation
  - Consistent design
- Allow grouping of samples
  - Selections
  - Classes
  - Clusters













# Tooltips with statistics

Austria • United States

United Kingdom

Israel

Luxembourg

#### Portugal

- Educational attainment 35  $-2.4 \sigma_{\text{Slove}}$
- Employees working ve... 9.31  $-0.034 \sigma$ 
  - Life expectancy 80.8 +0.39  $\sigma$
  - Life satisfaction 5.2
  - Self-reported health 50
- Student skills 488 -0.20 σ
- Time devoted to leisur... 14.95 +0.13  $\sigma$ 
  - Years in education 17.8  $+0.31 \sigma$

correct distances













# Comparing Two Groups

South America 3 samples Northern Europe 9 samples

Educational attainment 50 77 Employees working ve... 18 6.2 Life expectancy 75 81 Life satisfaction 7.1 7.4 Self-reported health 65 77 Student skills 420 500 Time devoted to leisur... 14 Years in education 16 19











### Heatmap from Dimension Hover



	P R O J E C Edit proje	ECTION	🗌 Display dend	rogram 🗌 Display error	rs 🔽 Display labe	ls
	SELECT	IONS				
	+					
	new sele select sar	ection nples				
	CLUSTE	RING 🔘	•			
	Clusters:		O−−−−→ 5 c	clusters		
			(w)		$\langle v_{i}\rangle$	74
	Cluster 1 10 sample	1 es	<b>Cluster 2</b> 9 samples	Cluster 3 4 samples	Cluster 4 4 samples	Cluster 5 9 samples
	DIMENS	IONS				
		Educatio	onal attainmei	nt	32	94
	2.4	Employe	ees working ve	ery long hours	0.17	43.29
sian Federati		Life expe	ectancy		69	82.8
		Life satis	sfaction		4.7	7.8
		Self-repo	orted health		30	90
		Student	skills		402	542
	-	Time dev	voted to leisur	re and personal care	13.42	16.06
	1	Years in	education		14.1	19.7











## Showing Error via Sample-centric Halos













## Showing Projection Errors



D. Koop, CSCI 627/490, Fall 2<mark>022</mark>















# User Study & Results

- Types of Questions:
  - How would you try to characterize the type X?
  - In what way are X and Y different in their properties?
  - Are the projections of X and Y correct or do they deviate? How do you interpret this?
  - Can you discover which parts of the cluster combinations are A, B, and C?
- Discussion:
  - Learnability: need more effective mechanisms for grasping the concepts behind dimensionality reduction
  - Manipulation: What happens with results? \_
  - Large data: What about text corpora?













### Focus+Context

- Show everything at once but compress regions that are not the current focus - User shouldn't lose sight of the overall picture

  - May involve some aggregation in non-focused regions
  - "Nonliteral navigation" like semantic zooming
- Elision
- Superimposition: more directly tied than with layers
- Distortion









### Focus+Content Overview

#### Embed $( \rightarrow)$

### → Elide Data



### → Superimpose Layer



#### → Distort Geometry



#### D. Koop, CSCI 627/490, Fall 2022





[Munzner (ill. Maguire), 2014]



Northern Illinois University







### Elision

- data
- In visualization, usually correlated with focus regions

• There are a number of examples of elision including in text, DOITrees, ... Includes both filtering and aggregation but goal is to give overall view of the







### Degree of Interest Function

- DOI = I(x) D(x,y)
  - I: interest function
  - D: distance (semantic or spatial)
  - x: location of item
  - y: current focus point (could be more than one)
- Interactive: y changes





## Elision: DOITrees

- Example: 600,000 node tree
  - Multiple foci (from search results or via user selection)
  - Distance computed topologically (levels, not geometric)











## Superimposition

- Different from layers because this is restricted to a particular region
  - For Focus+Context, superimposition is **not global**
  - More like overloading
- Lens may occlude the layer below







### Superimposition with Interactive Lenses



#### (a) Alteration

D. Koop, CSCI 627/490, Fall 2022

#### (b) Suppression [ChronoLenses and Sampling Lens in Tominski et al., 2014]



Northern Illinois University



## Superimposition with Interactive



D. Koop, CSCI 627/490, Fall 2022

Enrichment

[Extended Lens in Tominski et al., 2014]





### Distortion











# **Distortion Choices**

- How many focus regions? One or Multiple
- Shape of the focus?
  - Radial
  - Rectangular
  - Other
- Extent of the focus
  - Constrained similar to magic lenses
  - Entire view changes
- Type of interaction: Geometric, moveable lenses, rubber sheet





## Overplotting











### Cartesian Distortion











### Cartesian Distortion











# Stretch and Squish Navigation

🛃 LiveRAC			
File Edit Focus Groups Arrange	Screen shot Reports		
Manual	CPU used (Totals)	Load	# Procs
swamp	80 40 0 00:00 04:00 08:00 12:00 16:00 20:00 00:00 — CPU Used (All) [%] — CPU User (All) [%]		
sobriety	90- 60- 30- 0- 00:00 04:00 08:00 12:00 16:00 20:00 00:00 — CPU Used (All) [%] — CPU User (All) [%]		
spire	100 80 60 40 20 00:00 04:00 08:00 12:00 16:00 20:00 00:00 — CPU Used (All) [%] — CPU User (All) [%] — CPU I/O Wait (All) [%]		1
joint			
tang haversack		-M-	
blowout			
port mortality tior		tern.	
potpourri liberty			









## Fisheye Distortion in Programming

12	FastDate	eFormat.java 🛪
	66pu	blic class FastDateFormat extends
	(1423	printer minister fundationen printerier in einen einen einen einen einen im eine eine
	571	protected List parsePattern() {
	575	String[] ERAs - symbols.getEras();
	576	String[] months - symbols.getMonths();
	585	for (int i = 0: i < length:
	590	int tokenLen = token le
	595	Rule rule:
	596	char c = token charAt(0):
	597	000000000000000000000000000000000000000
	598	switch (c) (
	500	are ICL: // are design
	555	case G . // era design
	600	rule - new lextriei
	601	break;
	602	case 'y': // year (numb
	603	if (tokenLen >= 4)
	604	rule = selectNu
	605	} else {
	606	rule = TwoDigit
	607	}
	608	break;
	609	case 'M': // month in y
	610	if (tokenLen >= 4)
	611	rule = new Text
	612	<pre>} else if (tokenLen</pre>
	613	rule = new Text
	614	} else if (tokenLen
	620	case 'd': // day in mon
	623	case 'h': // hour in am
	626	case 'H': // hour in da
	629	case 'm': // minute in hour
	632	case 's': // second in minu
	635	case 'S': // millisecond (number)
	638	case 'D': // day in week (text)
	504 507	<pre>name 'T': If Gay all result is search (scatters) name 's': If result is year (scatters) result 'T' '' and is much index</pre>
	760	protected NumberRule selectNumb
	1	

#### D. Koop, CSCI 627/490, Fall 2022



#### [Jakobsen and Hornbaek, 2011]





### Distortion vs. Hide

🕖 Defa	aultGalleryItemRenderer.java 🔀
12	<pre>package org.eclipse.n</pre>
37	public class DefaultG
41	boolean dropShado
<mark>≏</mark> 78	<pre>public void draw(</pre>
95	<b>if</b> (itemImage
100	eize = ce
101	SIZE - ge
101	vShift =
102	wShift -
103	ysniic -
104	i fi dan a a C
105	II (urops
100	Color
107	IOT (
108	C
109	g
110	
111	g
113	}
114	}
115	}
152	}
154	public void setDr

D. Koop, CSCI 627/490, Fall 2022



NIU



### Research Questions

- Is a priori importance useful (and for what)?
- What does the user focus on?
  - predictability of view changes when focus changes
  - how direct user control is
  - task & context
- What interesting information should be displayed
  - degree of interest function may produce varied result sizes
- Do fisheye views integrate or disintegrate?
  - interference with other interactions; allow on-demand use?
- Are fisheye views suitable for large displays?

#### D. Koop, CSCI 627/490, Fall 2022







# **Distortion Concerns**

- Distance and length judgments are harder
  - Example: Mac OS X Dock with Magnification
  - Spatial position of items changes as the focus changes
- Node-link diagrams not an issue... why?
- Users have to be made aware of distortion
  - Back to scatterplot with distortion example
  - Lenses or shading give clues to users
- Object constancy: understanding when two views show the same object
  - What happens under distortion?
- 3D Perspective is distortion... but we are well-trained for that • Think about **what** is being shown (filtering) and method (fisheye)







