Data Visualization (CSCI 627/490)

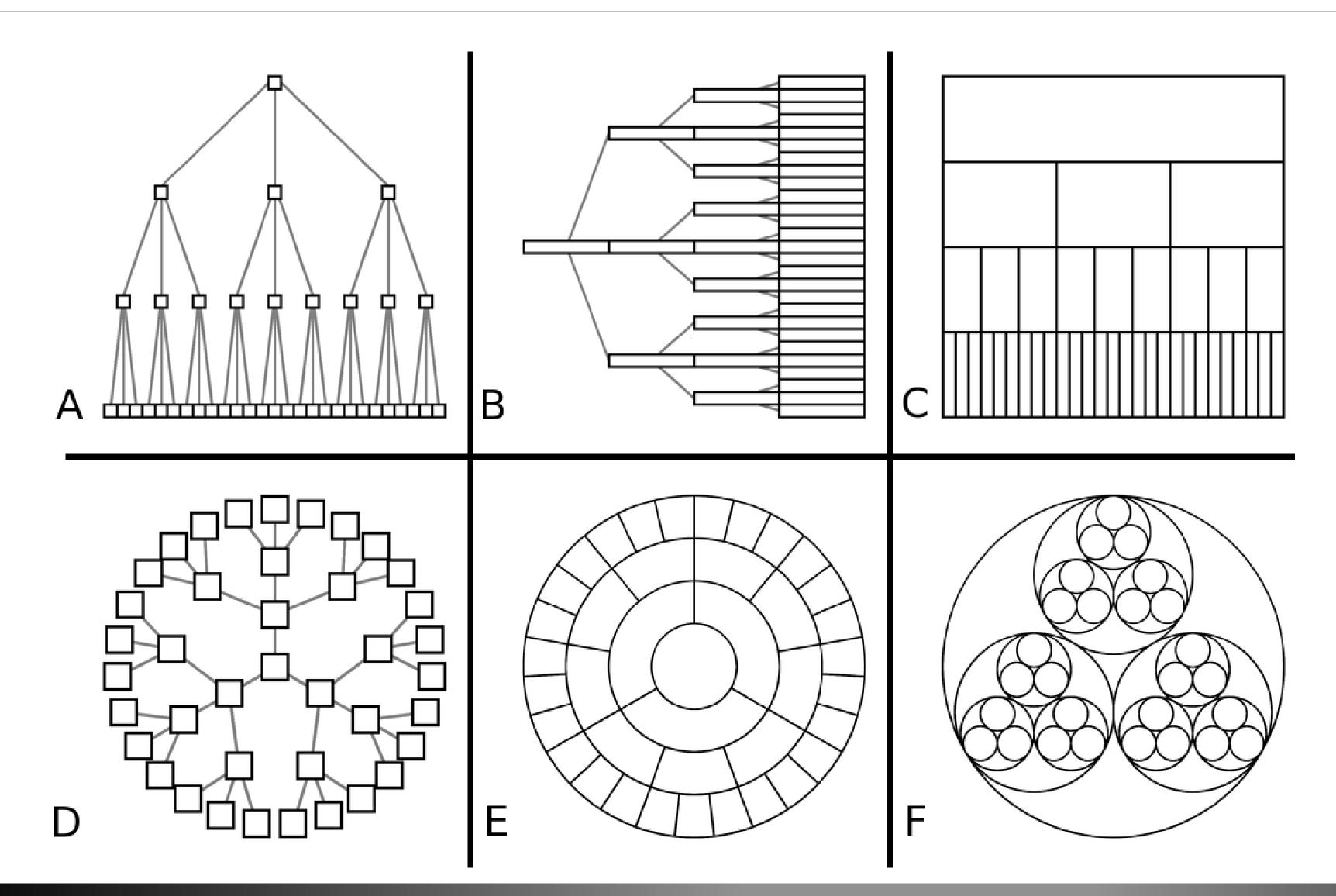
Design

Dr. David Koop

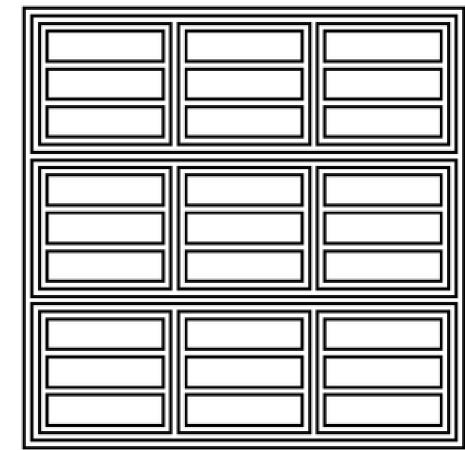


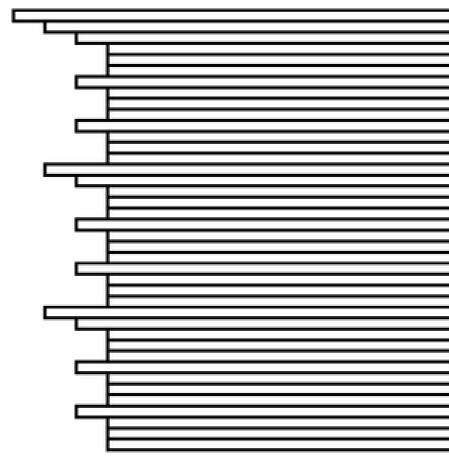


Tree Visualizations



D. Koop, CSCI 627/490, Fall 2024





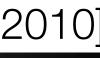




Η

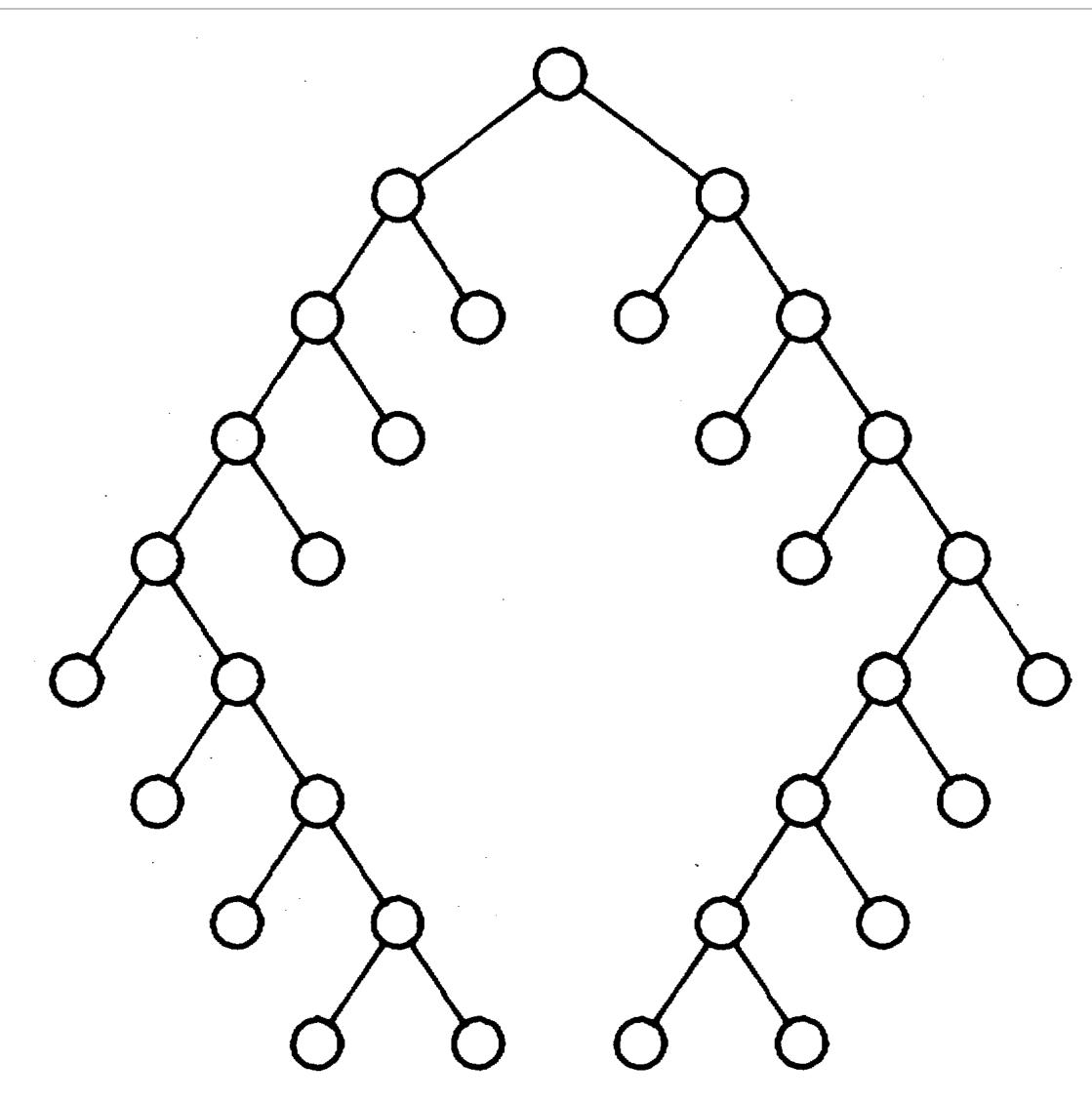
G







Reingold-Tilford Algorithm



- Recurse on left and right subtrees
- Shift subtree over as long as it doesn't overlap
- Place parent centered above the subtrees
- Originally, only binary trees, extended by Walker











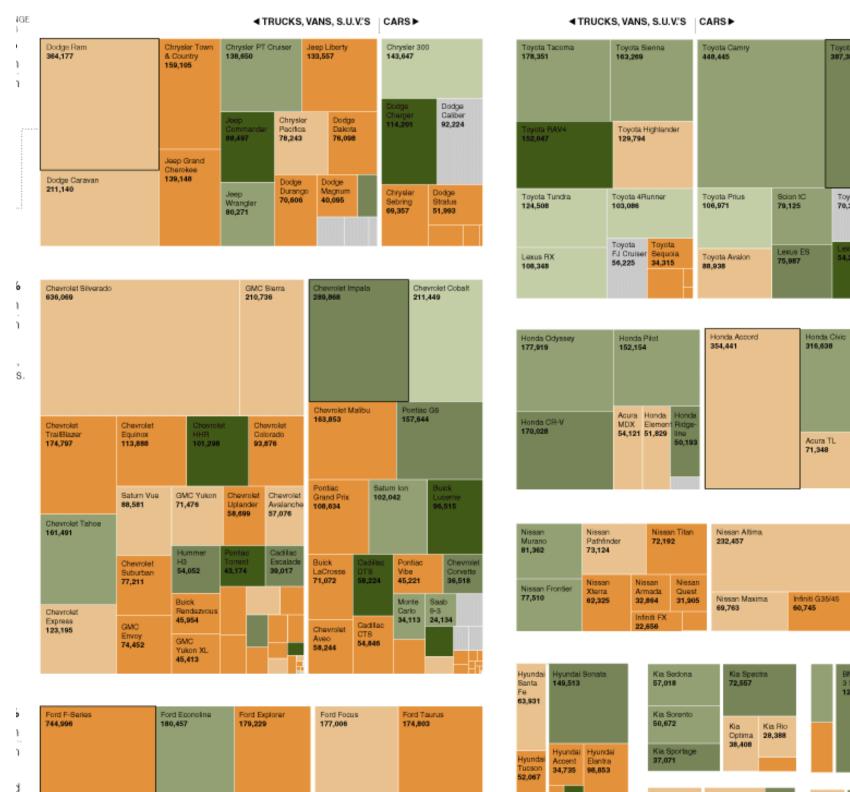




Treemap

Truck Sales Slip, Tripping Up Chrysler

Over the past few years, Chrysler executives said they were following the lead of Toyota and Honda, focusing on vehicles that met the needs of their customers. But as American consumers turned away from large trucks and S.U.V.'s in 2006, Chrysler continued to churn out big vehicles, which are now sitting unsold at dealerships across the country.



READING THE CHART

100.00

Boxes are scaled

proportionally

according to

sold in 2006

number of cars

Change in sales from

-10% -2.6

Aoura TSX 27,934 38,035

Below Above

D. Koop, CSCI 627/490, Fall 2024

- Containment marks instead of connection marks—show hierarchy
- Encodes some quantitative attribute of the items as the size of the rectangles
- Not as easy to see the intermediate rectangles (hierarchy)
 - Scalability: millions of leaf nodes and links possible

[A. Cox and H. Fairfield, NYTimes, 2012]





Treemap Layouts: Slice & Dice

- Split at each level into strips
- At each step, orientation of division (horizontal/vertical) changes
- Better, but some rectangles still have bad aspect ratio

Node Link Tree				Labeler 9,956	Prope Encod 4,138	Distor 6,314	r Fish Tree Filte	OCC LiS 5,42	S<mark>Strings</mark> Q<u>222</u>,026r 21,0236				Interpo N C F / F C I I 8,746 II II II II II I I 2 2 2 1 1 1	ragiyaadilaaqiiddalibbal 7.76235859958801888	Misel I E FiPD(Missi Ci	E Time Scale 5,833	Gr <u>I</u> Cc 9 9,{ (JSNDN CBody CEorce 1,10,49	rtes erte(
Layout 12.870 Radial Tree							5,21								7,		2	1,00,49	3
Layout 12,348 Circle																			
Packing Layout 12,003					Encod 4,060				Shapes			_							
Circle Layout 9,317									19,118					Query 13,896					
Tree Map Layout						Bifoca Distor	Visit Filte							13,690		Quan Scale 4,839			
Stacked Area Layout					Color Encod	4,461	3,50						Transitioner 19,975			4,000			
Force Directed Layout				Radial Labeler 3,899					Maths 17,705				10,070						
Layout 7,881							Grap							Expression					
Axis Layout Icicle				Stacked Area		Fisher Distor 3,444		r	Disch					5,130		Scale 4,268		Simu	
Dendrogram Lavout Bundled				Labeler 3,202	Shape Encod				Displays 12,555	i				Comparison 5,103		1,200		9,983	
Indented Pie Data	Data	Node Sprite	Sc	cale Da	1,690 Ita Tr prite B	ree uilder	Edge Rend	Tree 6 7,147	E				Easing	Date					Dir Sp
	List 19,788	Sprite 19,382	Bi	nding Sp ,275 10	,349 9,	uilder ,930	Rend 5,569	e 7,147	Sr Color Palette 6,367	Ĭ	Size Palett 2,291 2,05	t 1.2	17,010	Date Util 4,141					8,8
									Geomet	Geometry				String Util 4,130		Ordin Scale 3,770			
									10,993	,				Arithmetic 3,891	Hip A C	Morgo			
									Fibonac Heap	сі		Hea		Match 3,748	Hie A C Clu C S 6,7 3, 3	Edge 743			
							Shap Rend 2,247		9,354			1,2	Transition 9,201	Composite Expression		Log Scale 3,151	, Data Sour	c	
							Arrow	1	Colors 10,001					3,677 Expression Iterator		3,131	3,33		
Tooltip Control	Selection Control	Pan Zoom Control	Hover Control	Control List	Click Contr		1	g Anc C t Cor ₁		_			Tween	3,617 Fn				Parti	
8,435	7,862	Control 5,222	4,896	4,665	3,824	1 2,83	32 2,64	8 2,1(Sparse Matrix 3,366	Dense Matrix 3,165	I Matrix 2,815	x	6,006	3,240 Binary Expression		Quan Scale 2,435	Data Util	2,822	2
									Arrays 8,258				Function Sequence	2 893 lf			3,322	2	
Legend 20,859				Legend Range			Le	egend em	Dates				5,842	2,732 IsA 2,030		l Scale Map)	Sprin 2,210	n Re 3 Sp 3,6
				10,530			4,	614	8,217				Scheduler 5,593	Variance Aggregate		2,105	Data		
Axis					Ca Ax	artesian	ן ו	Ax 1,5	Sort 6,887				Sequence	Range Not	Aspect Ratio Banker	Scale Type 1,821	2,165	Sprin Force 1,68	
24,593					Ax 6,7			1,5	6 Stats 6,557				5,534	Literal Variable Xor	7,074	Root Scale	Data Field		
Visualization									Property 5,559				Parallel 5,176	And Or Distinct		1,756	1,75	1,336	6 Lir Sp
16,540									Filter					Maximum		Linea Scal	Data	Drag	, /



Northern Illinois University





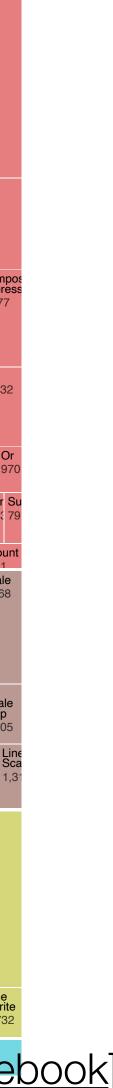


Treemap Layouts: Squarify

- Slice & Dice and Strip can lead to bad aspect ratios
- Solution: Strip only uses rows, allow columns to be used, too
- Choose divisions (x/y) based on the width/height of region in order to maintain good aspect ratios
 - Use left and right side
 - Process large rectangles first
- Ordering not preserved which may cause issues if the data is updated

Node Link Tree Layout 12,870		Circle Layou 9,317	ut	Tree Map Layo 9,19	out	A L	Stacked Area ayout ,121		Data 20,544			Data List 19,78			Interpolator 8,746 Matrix Color R	1	ransitio 9,975	ner	gt mi 60; 60 div e	748 625 61 sut ne It 60(59 55)
Radial Tree Layout 12,348		Force Direc Layou 8,411	ut	La	kis ayout 725	lcic Tre Lay 4,8	e ا رout	Dendrog Layout 4,853	1						Array 1,983 1,629	Num			fn o not u stdc s Expres		n 2 2 nparison	Date
Circle Packing Layout 12,003		Layo 7,881		Ec	undled dge outer 727	Inder Tree Layo	ut	R Li 8'	Node Sprite 19,382				Scale Binding 11,275		Point Interpola 1,675 Easing 17,010		Tran 9,20	isition 1	5,130 String Util	5,1 Arithm 3,891	03 et Match 3,748	Util 4,141 Compo Expres
Labeler 9,956	Radial Labele 3,899 Stacke Area Labele	d	Distortio 6,314 Bifocal Distortio		Di	istor Lis	e Perator	Operat Sequer 4,190	Data Sprite 10,349			Edge Rend 5,56		⊤ree ′,147	Tween			Parallel	Iterato	ssic Fn	Binar	3,677 y If 355 2,732
Property Encoder 4,138 Encoder 4,060	3,202 Color Encoder 3,179 Shape Encoder	1,83	4,461 Fisheye Tree Filter 5,219 Visibility Filter		Dis Fil	ter 2,5 165 Sc	ort Derator	Operato 2,490 I Opera	Builder 9,930			Shap Reno 2,24 Edge Sprit	 Por		6,006 Function Sequence 5,842		5,534	5,176	IsA 2,039 Varian 1,876		1,10	And Or 1,02 97 inct Mir S 84: 7
Tooltip Control 8,435	Encoder Selection Control 7,862	P Z C		Legend 20,859		2,0	023	1,286	Legend Range 10,530		Lege Item 4,614	nd Vis	ualization 540	1	Scheduler 5,593 Max Flow Min Cut 7,840	Lir	Transit Event 1 116 hk stance 731	I Pa Sched 44 1 041 Betwe Centra 3,534	Aggre Expres	Literal gate 1.214 ssic Variat 1 124 Time Scale	Aver 891 Ile Max	
Hover Control 4,896 Control	Click Control 3,824	Expar Contro 2,832	Drag Contrc 2,649	Axis 24,593						Cartes Axes 6,703	ian	Da	ta	Tooltip Event	Shortest Paths 5,914	Tr 3,₄	416			Ordinal Scale 3,770	Quant Scale 2,435	ile I Scale Map 2,105
List 4,665 Strings 22,026	Anchor Control 2,138	1,3	htro I 53 Cor 763 Aaths 7,705				Geom			Axes 1,302 Spars Matrix	Axis Axis Dense Matrix	2,3 Se Ev	lection ent	1,701 Visualiz Event tes	Hierarchical Cluster 6,714	3,938	nunity ture	re NE 7		Log Scale 3,151	Scale Type Root Scale	Lir Sc 1,3
										3,366	3,165				GraphML Converter 9,800		JSON Conve 2,220	Source	Data Util 3,322	2 10,0	e 566 5	Dirty Sprite 8,833
Shapes 19,118			Displays 2,555				Fibon Heap 9,354		H N 1,	Matrix 2,815 Sort 6,887	(Property 5,559	/	Delimited Text Converter N Body Force		I Data Conve 701	Data Field	ticle Sp	Data Table 779 Data Set Dring 213		
		F	Color Palette 5,367			Shape Palette 2,059		5 1		Stats 6,557			Filter 2,324	Pro	10,498 Simulation 9,983			Spr For	ce	Drac Forc 1.08 Flare	e 3	Line Sprite 1,732
					Palette	9							Orientat 1,486	ion I I Pre				Gra For	ce	1,08 Flare Vis I 4.11	<u>N</u>	ote

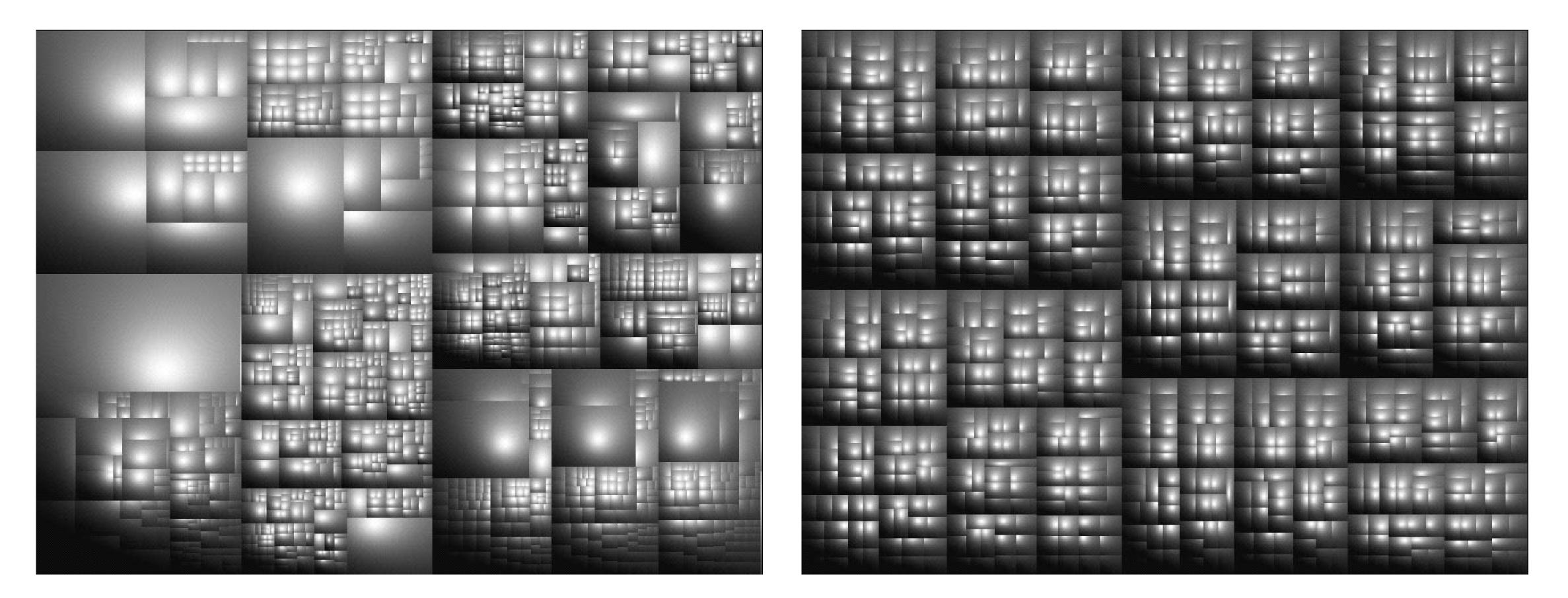








Squarified + Cushioned Treemaps



(a) File system

D. Koop, CSCI 627/490, Fall 2024

(b) Organization



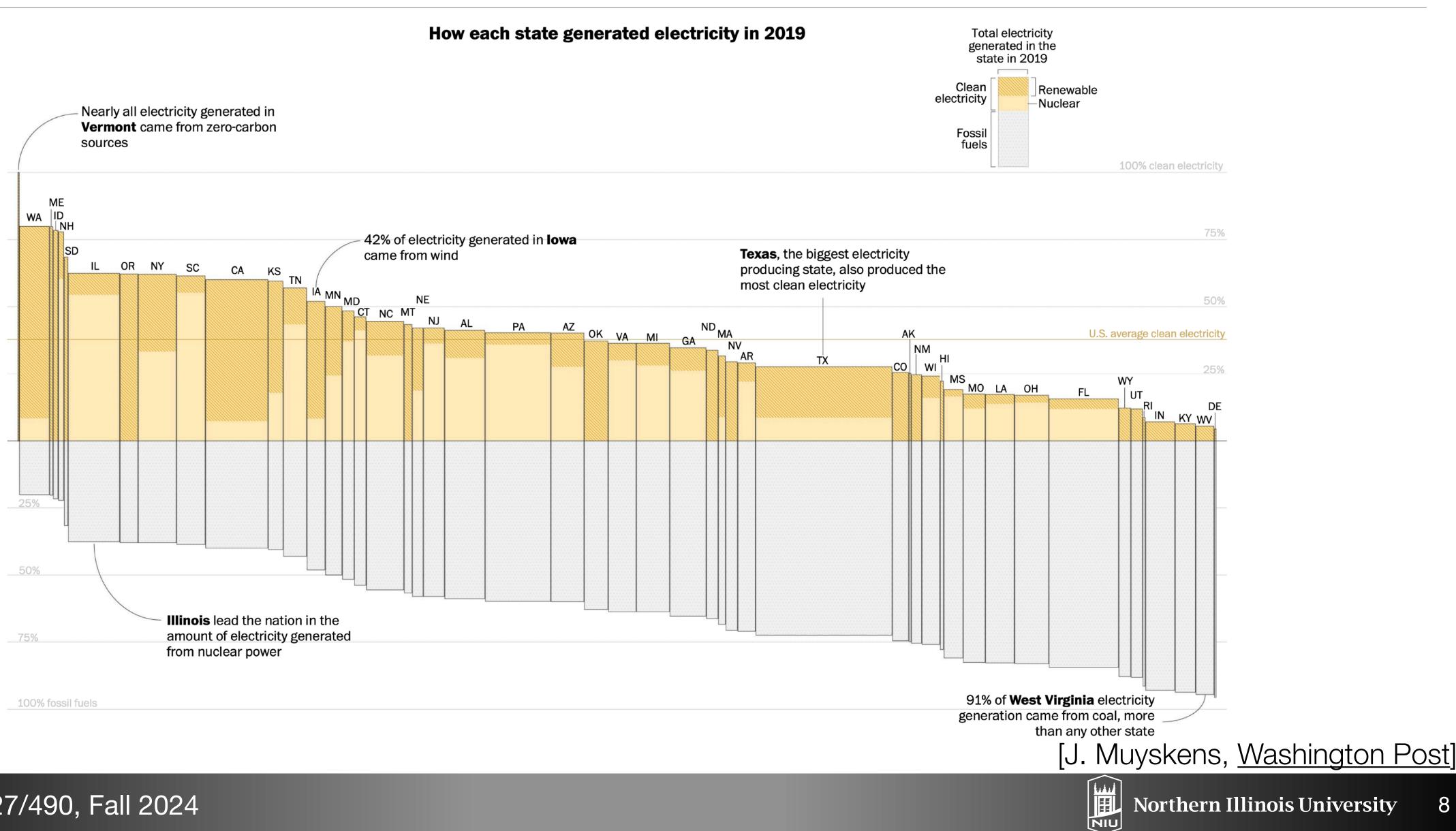








Variations: Marimekko Chart



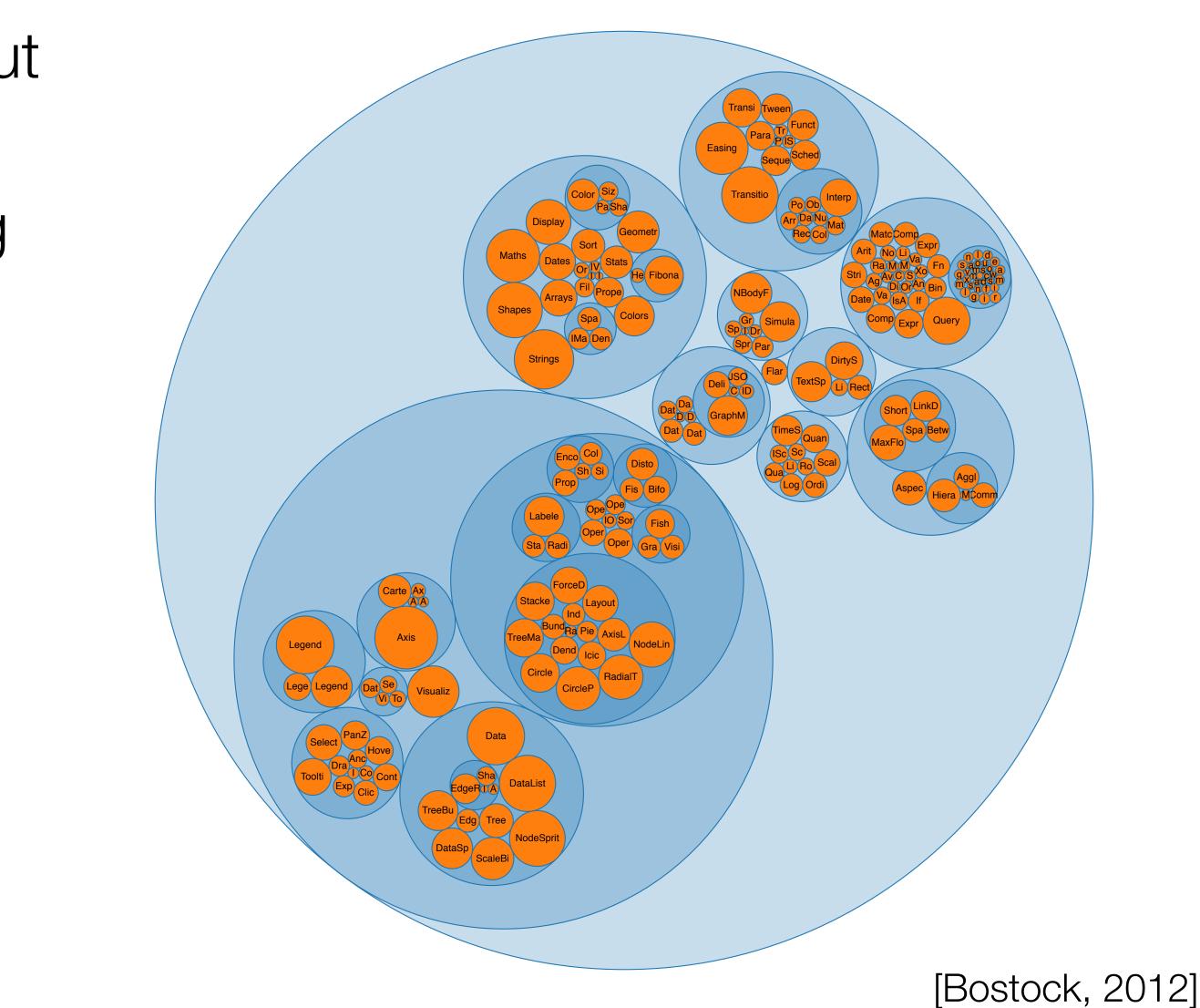






Nested Circles

- Looks more like cluster diagram, but shows hierarchy
- Containment shown by the layering of semi-transparent circles
- Labeling becomes more difficult





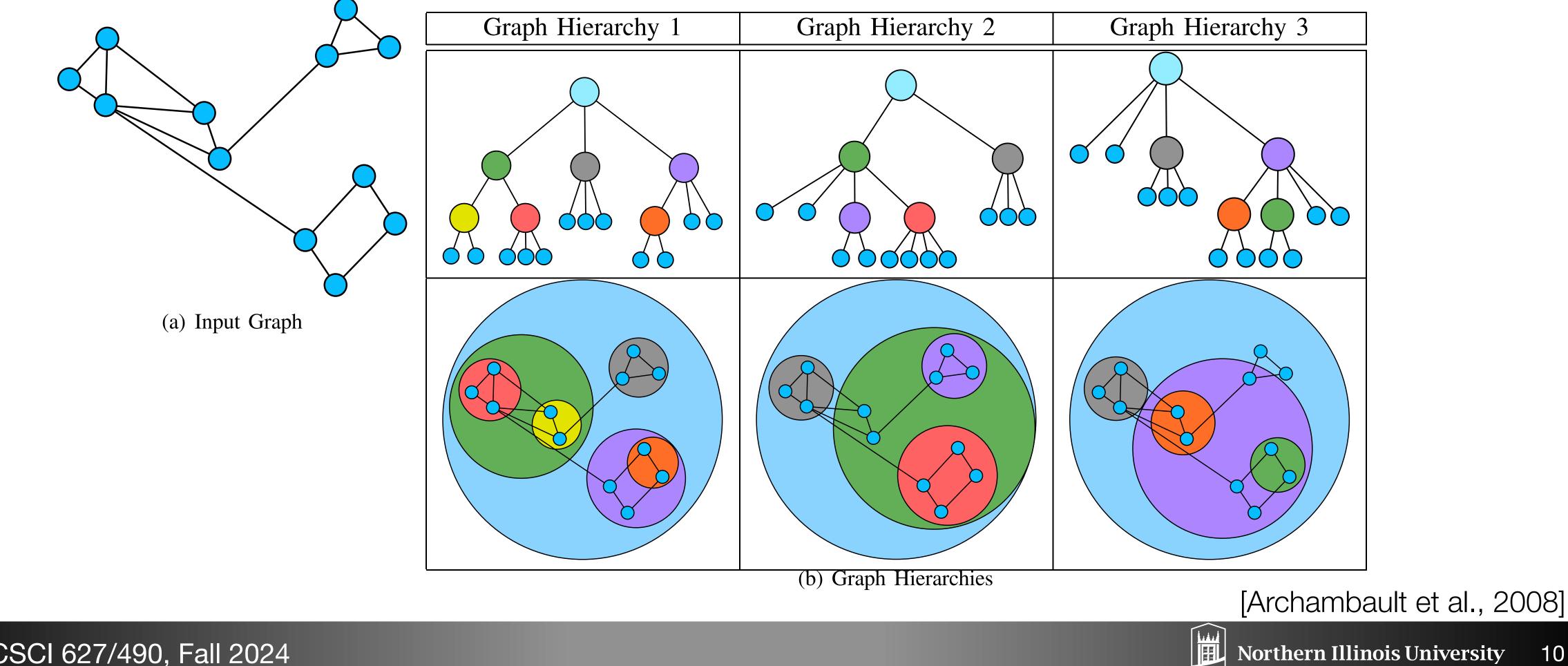






Compound Networks

- Add a hierarchy to the network (e.g. from clustering)
- GrouseFlocks: uses nested circles with colors



D. Koop, CSCI 627/490, Fall 2024

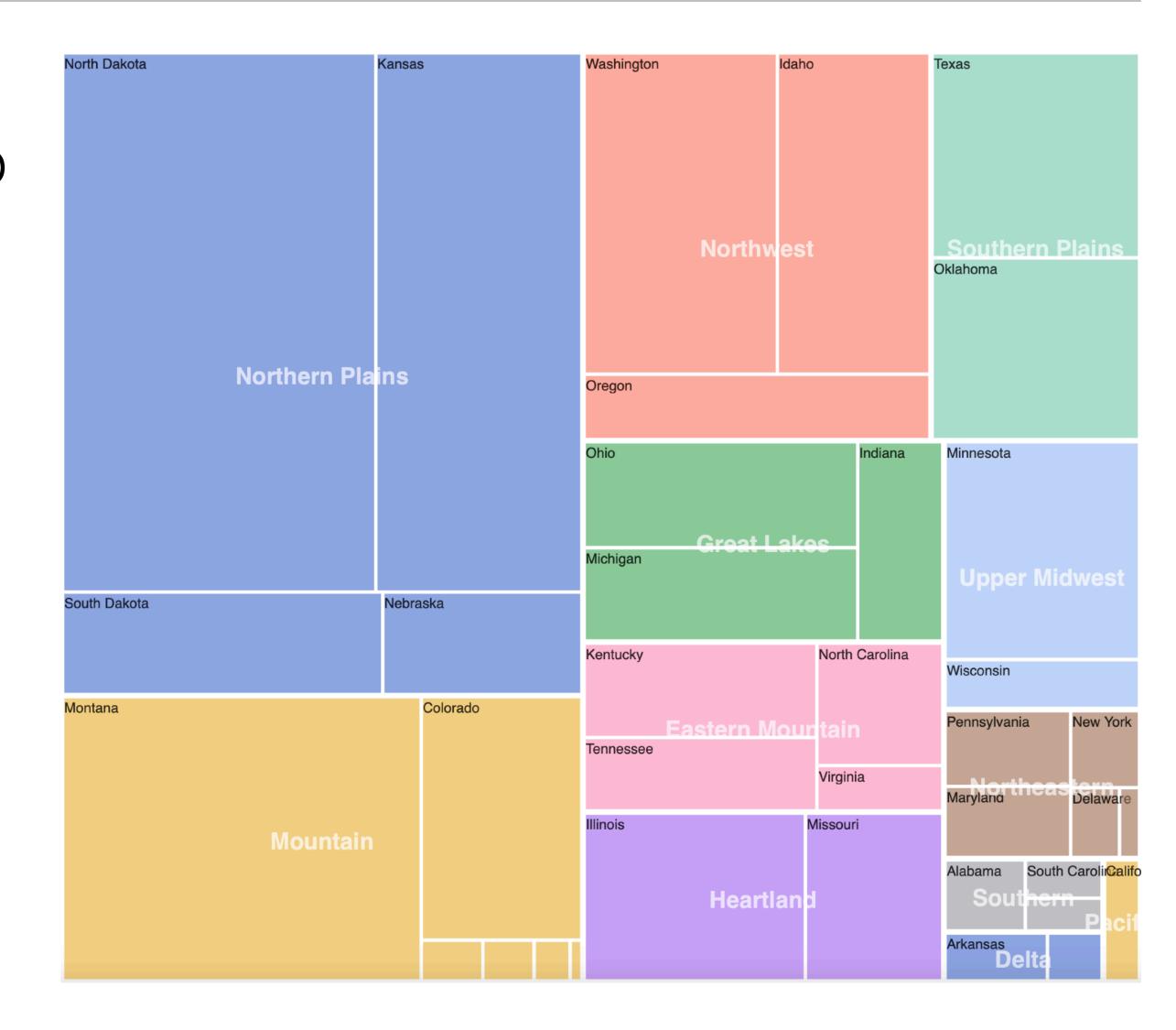




NIU

<u>Assignment 4</u>

- Crop Production in the US
- Geospatial Visualizations & Treemap
 - Choose colormaps carefully
 - Add legend
- You may use D3 or Observable Plot
 - Part 1a: D3
 - Part 3 will require some D3 for treemap layout







- Next steps:
 - Start thinking about the designs that help answer the questions
 - Tasks should drive your design
 - Different designs are great
 - Multiple views
 - Single view with details on demand
 - Interaction design (linked highlighting, navigation)
 - In general, don't force the user to make choices without first seeing an overview





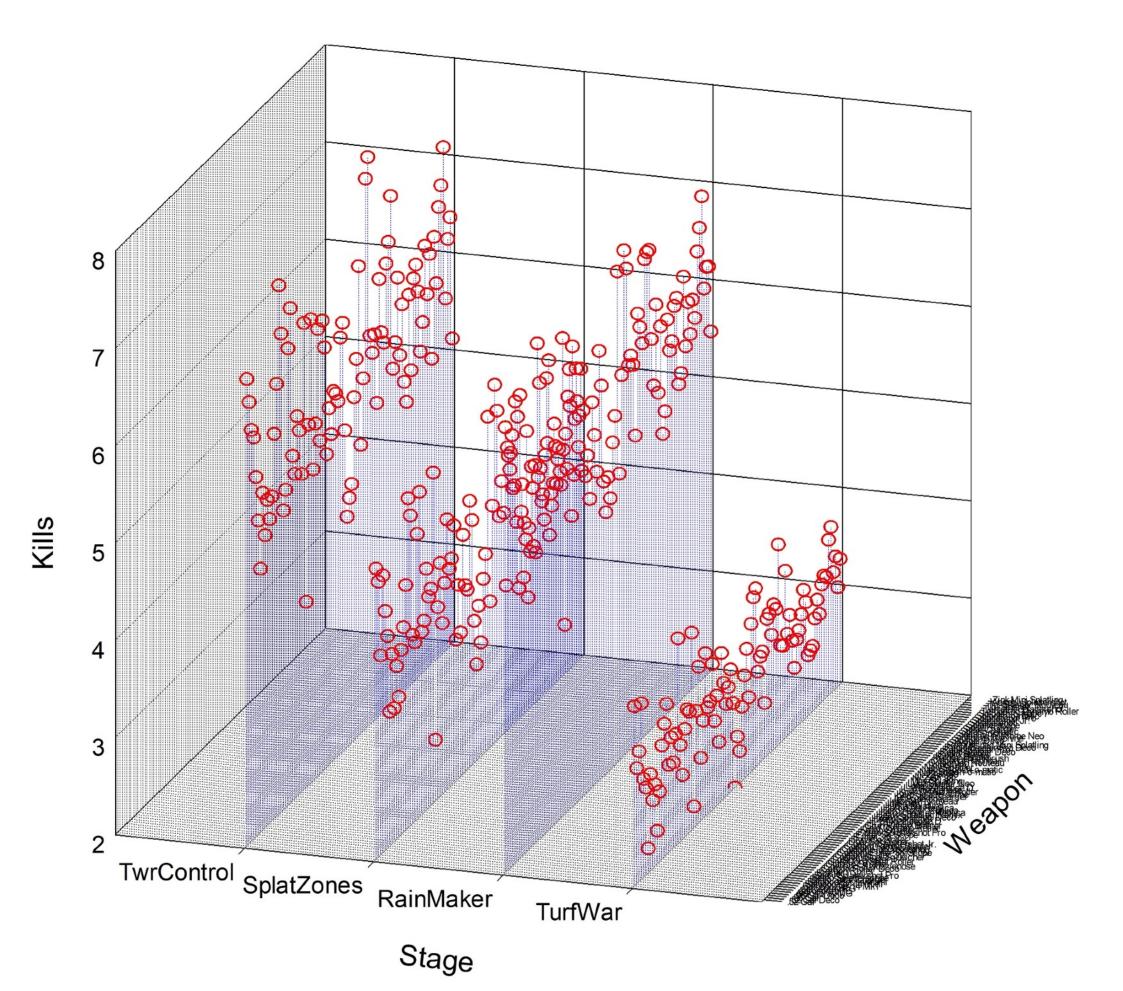
Guidelines for Visualization Design





WTF Visualizations (<u>wtfviz.net</u>)

3D Category Scatter



D. Koop, CSCI 627/490, Fall 2024







Northern Illinois University



Tufte: "The da Vinci of Data" —<u>NYTimes</u>



D. Koop, CSCI 627/490, Fall 2024

BEAUTIFUL EVIDENCE

The Visual Display of Quantitative Information SECOND EDITION

VISUAL EXPLANATIONS

[https://www.edwardtufte.com/tufte/, 2017]

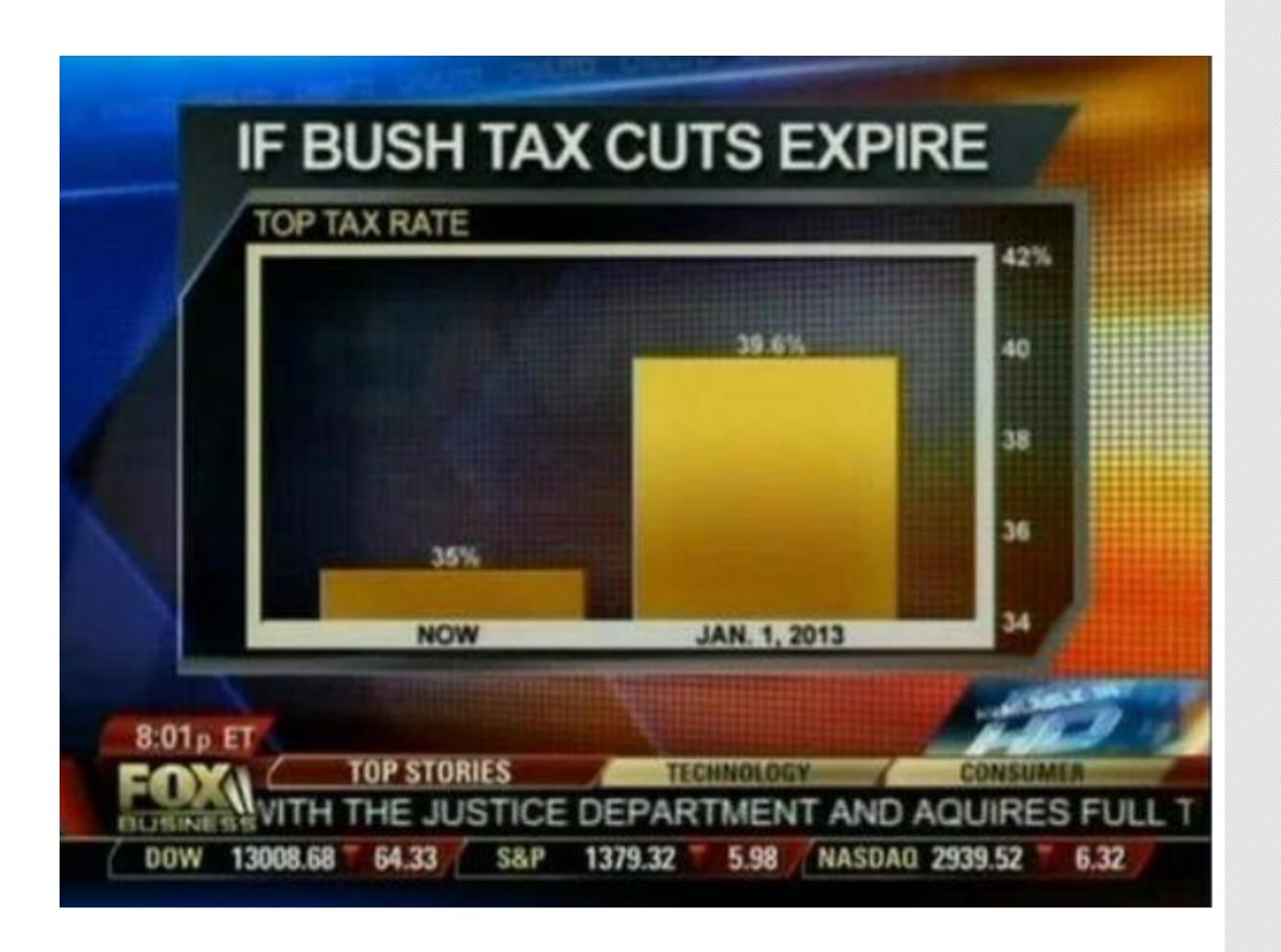








Bad: Data magnitude $\langle \neq \rangle$ Mark magnitude









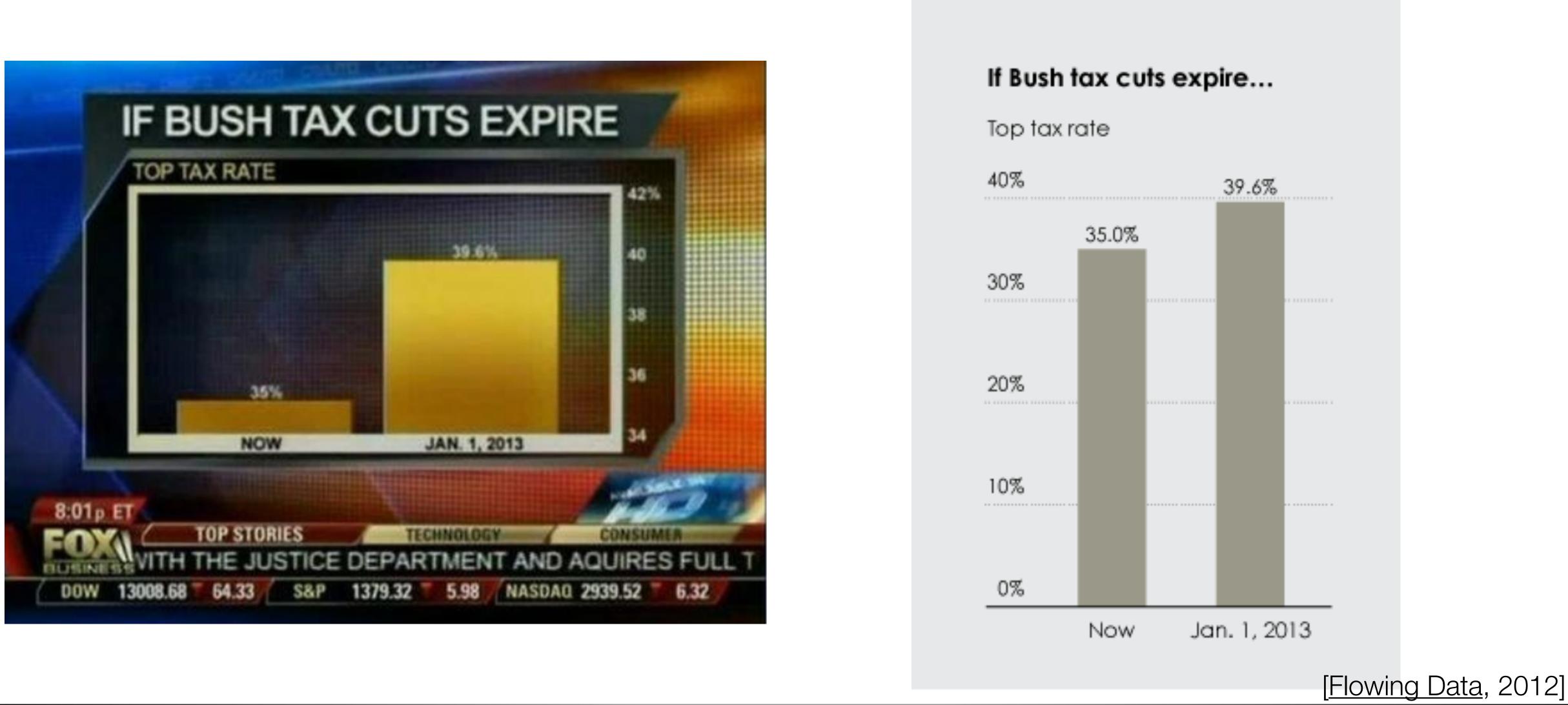








Good: Data magnitude <=> Mark magnitude



D. Koop, CSCI 627/490, Fall 2024

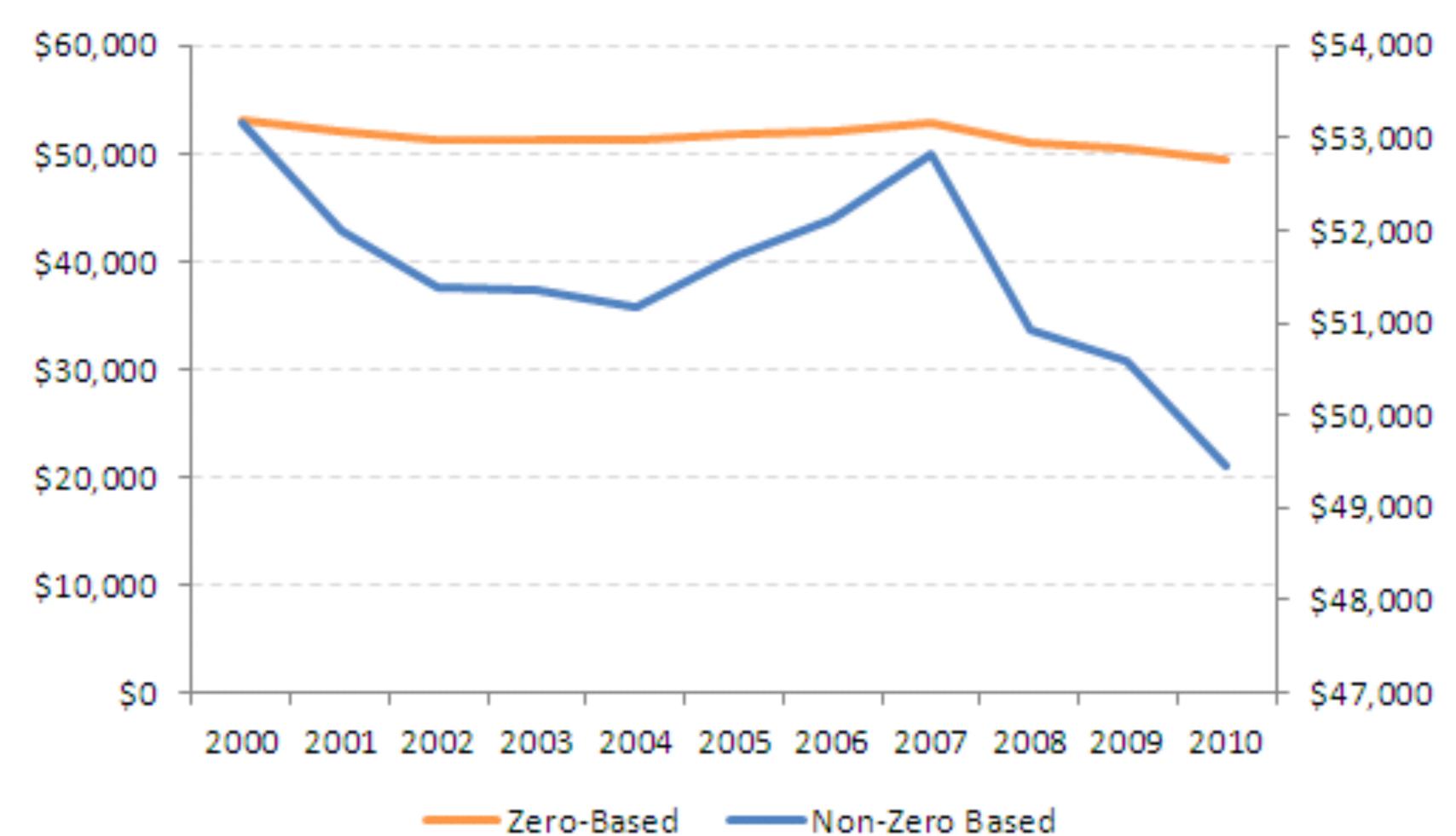




17

Starting Scales at Zero?

Median household income in 2010 inflation adjusted dollars



D. Koop, CSCI 627/490, Fall 2024

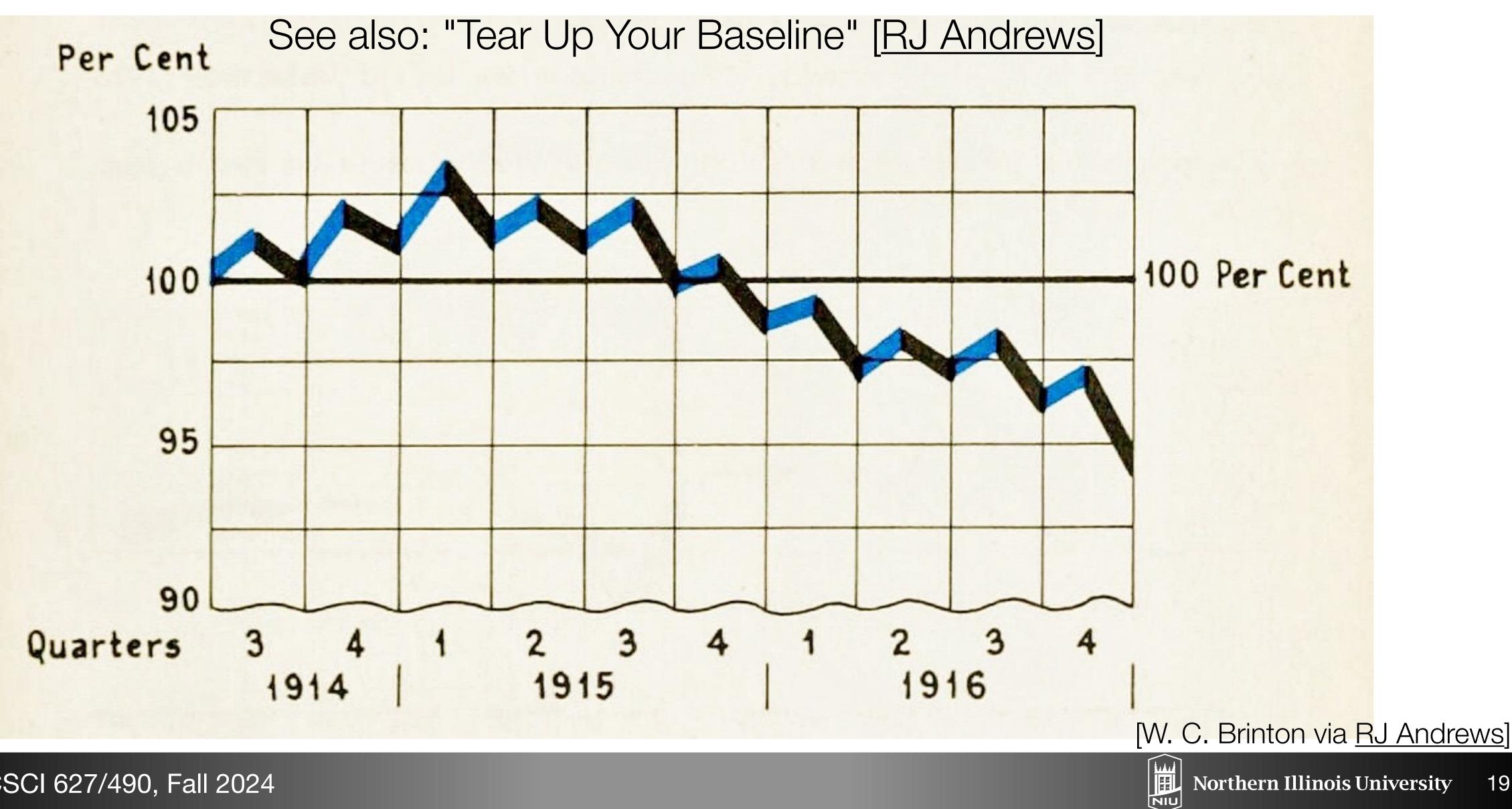
——Non-Zero Based





18

Wavy baselines for non-zero starts

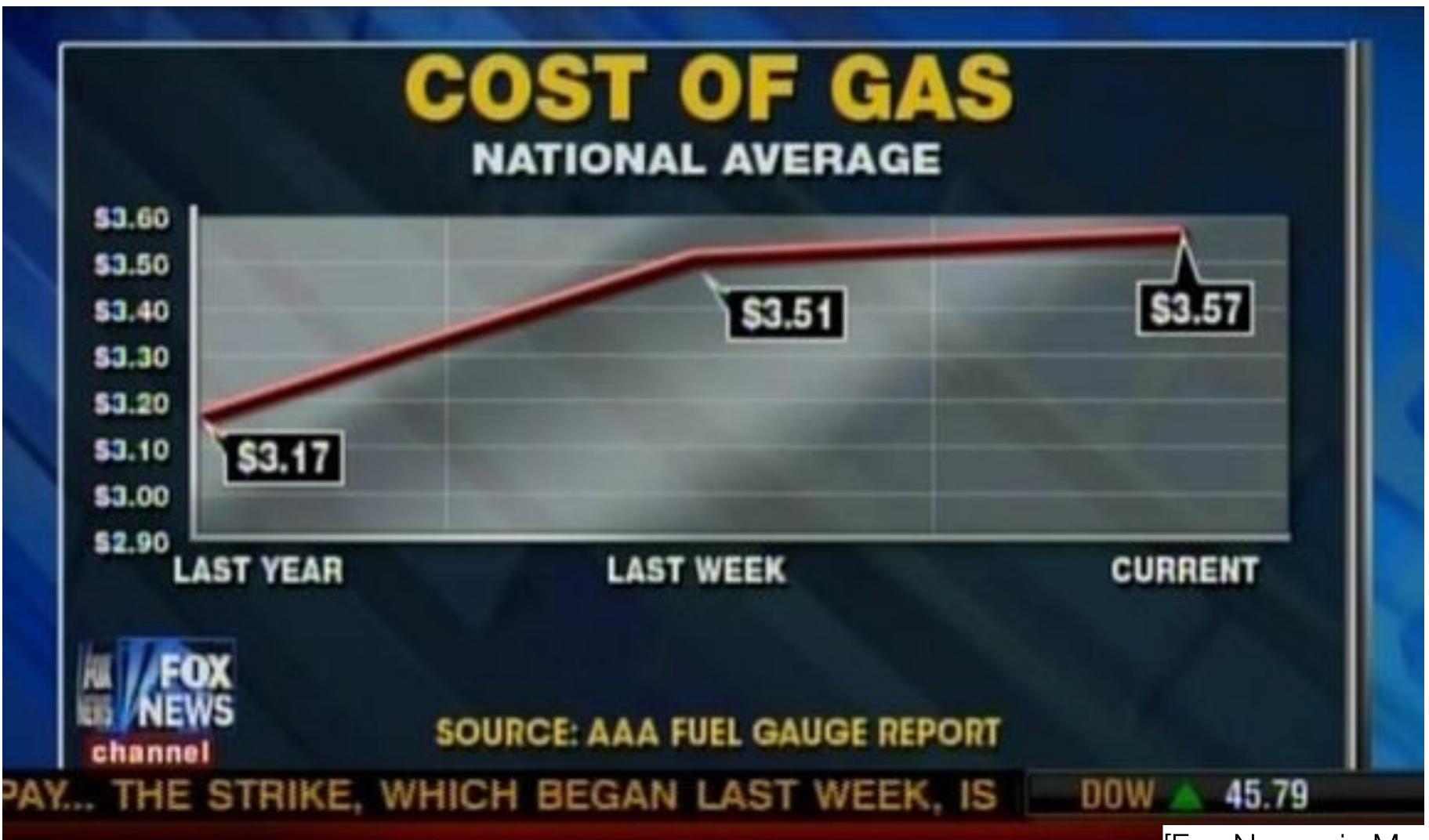








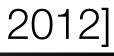
Cherry-picking data



D. Koop, CSCI 627/490, Fall 2024

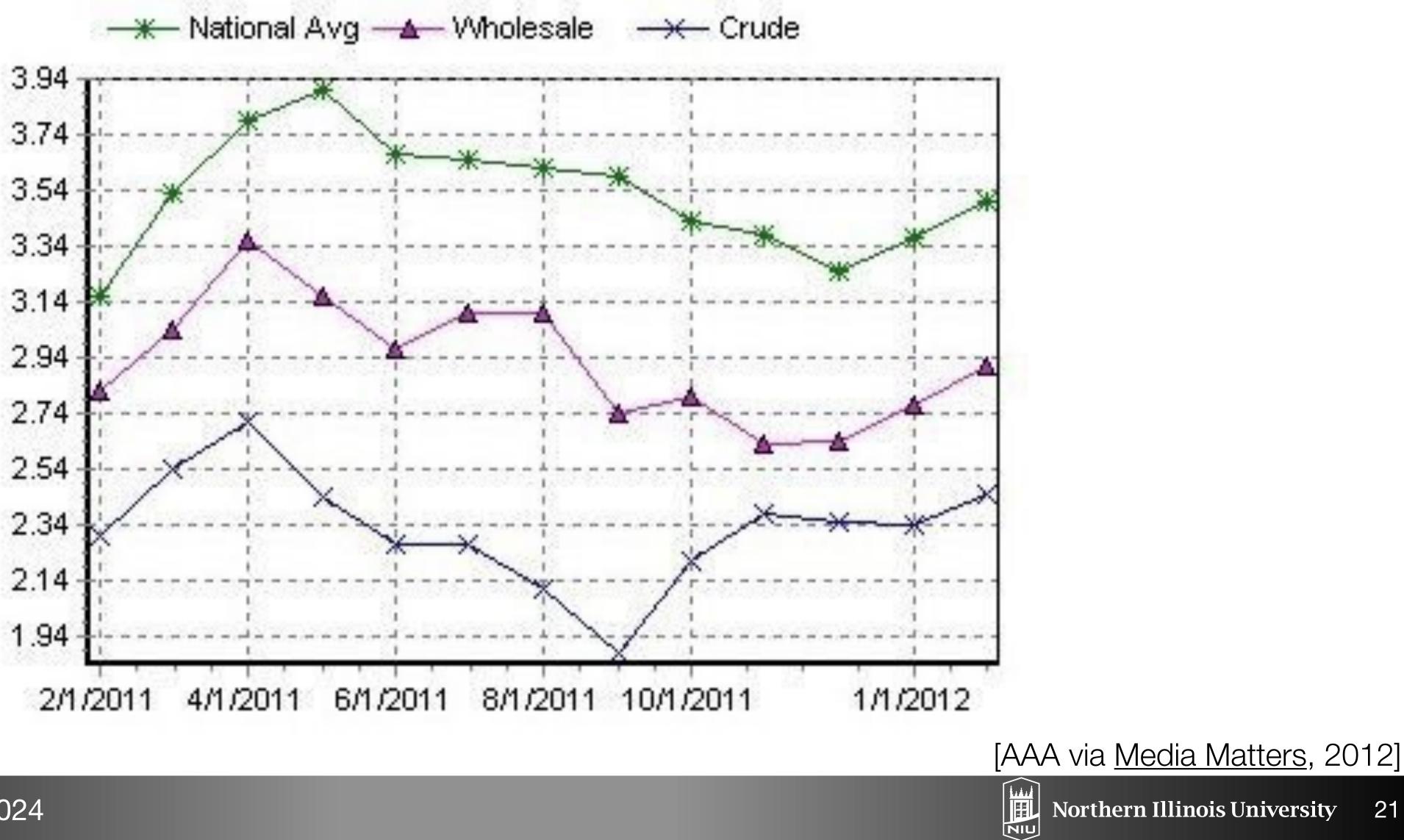
[Fox News via Media Matters, 2012]







Show all the data



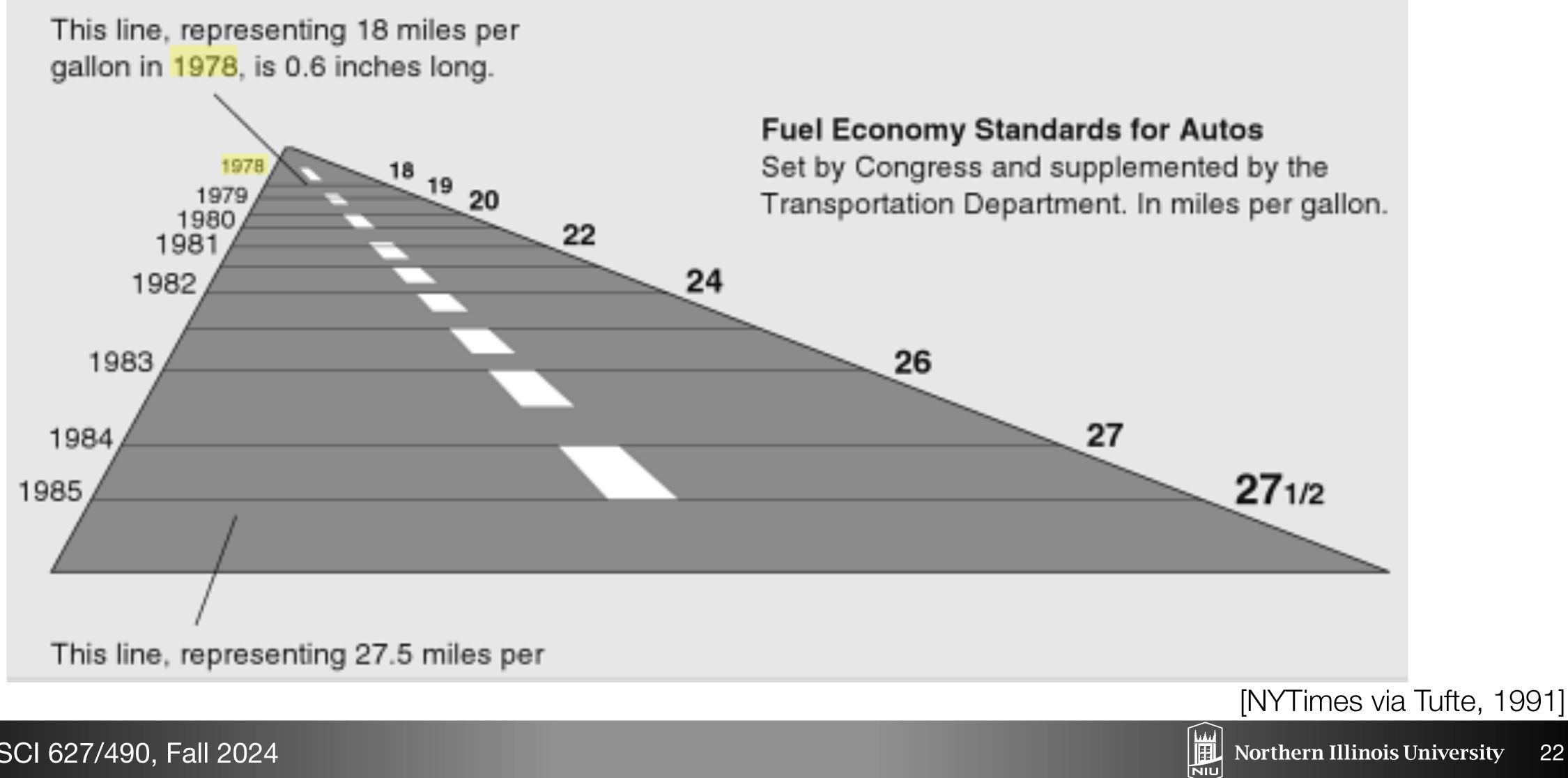
D. Koop, CSCI 627/490, Fall 2024

Northern Illinois University





Tufte's Lie Factor









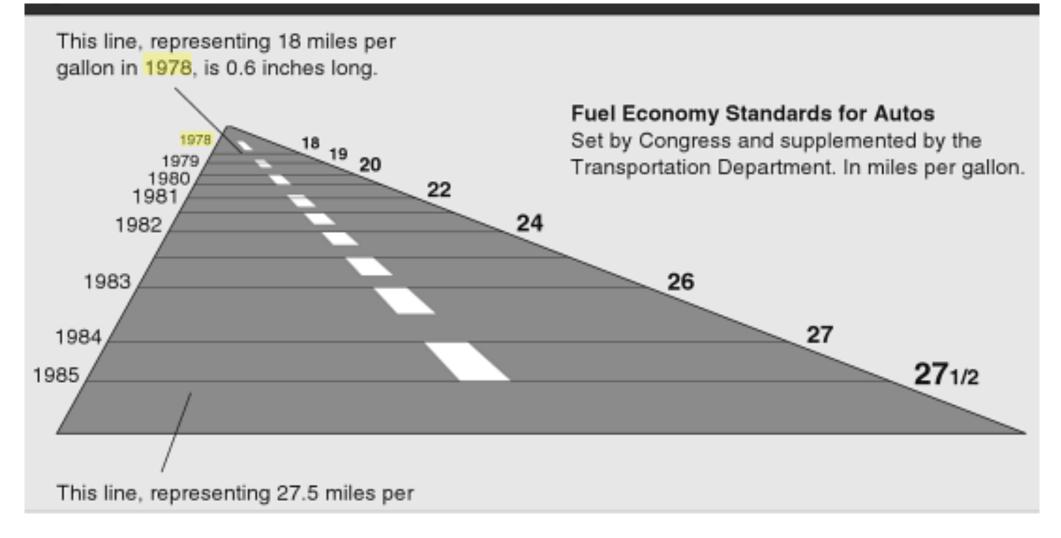


Tufte's Lie Factor

- Size of effect = (2nd value 1st value) / (1st value)
- Lie factor = (size of effect in graphic) / (size of effect in data)
- In the graphic:

Lie Factor =





D. Koop, CSCI 627/490, Fall 2024

5.3 - 0.6 0.6 14.8 27.5 - 18









(Some of) Tufte's Integrity Principles

- Show data variation, not design variation
- Clear, detailed, and thorough labeling and appropriate scales
- Size of the graphic effect should be directly proportional to the numerical quantities ("lie factor")

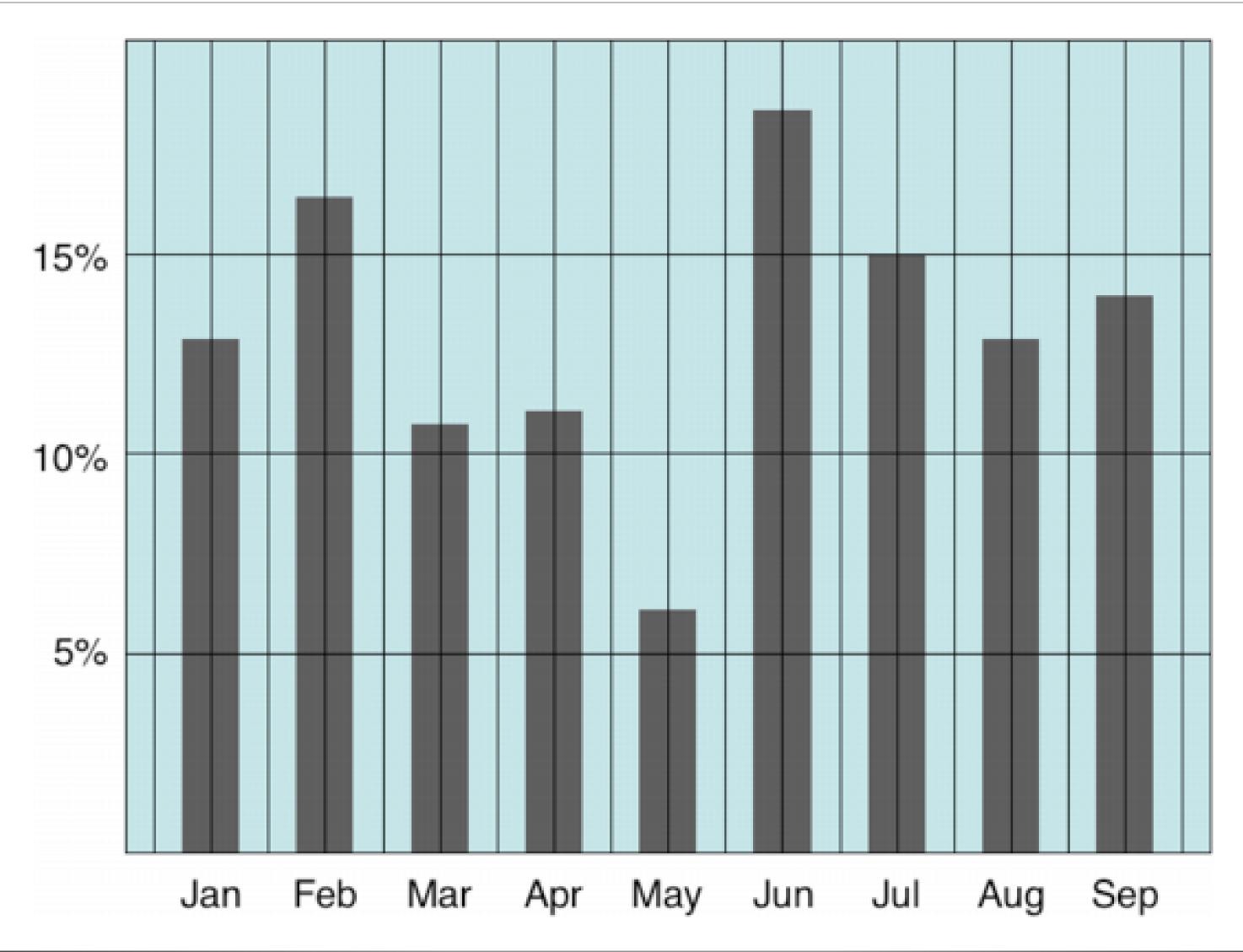








Avoid Chartjunk





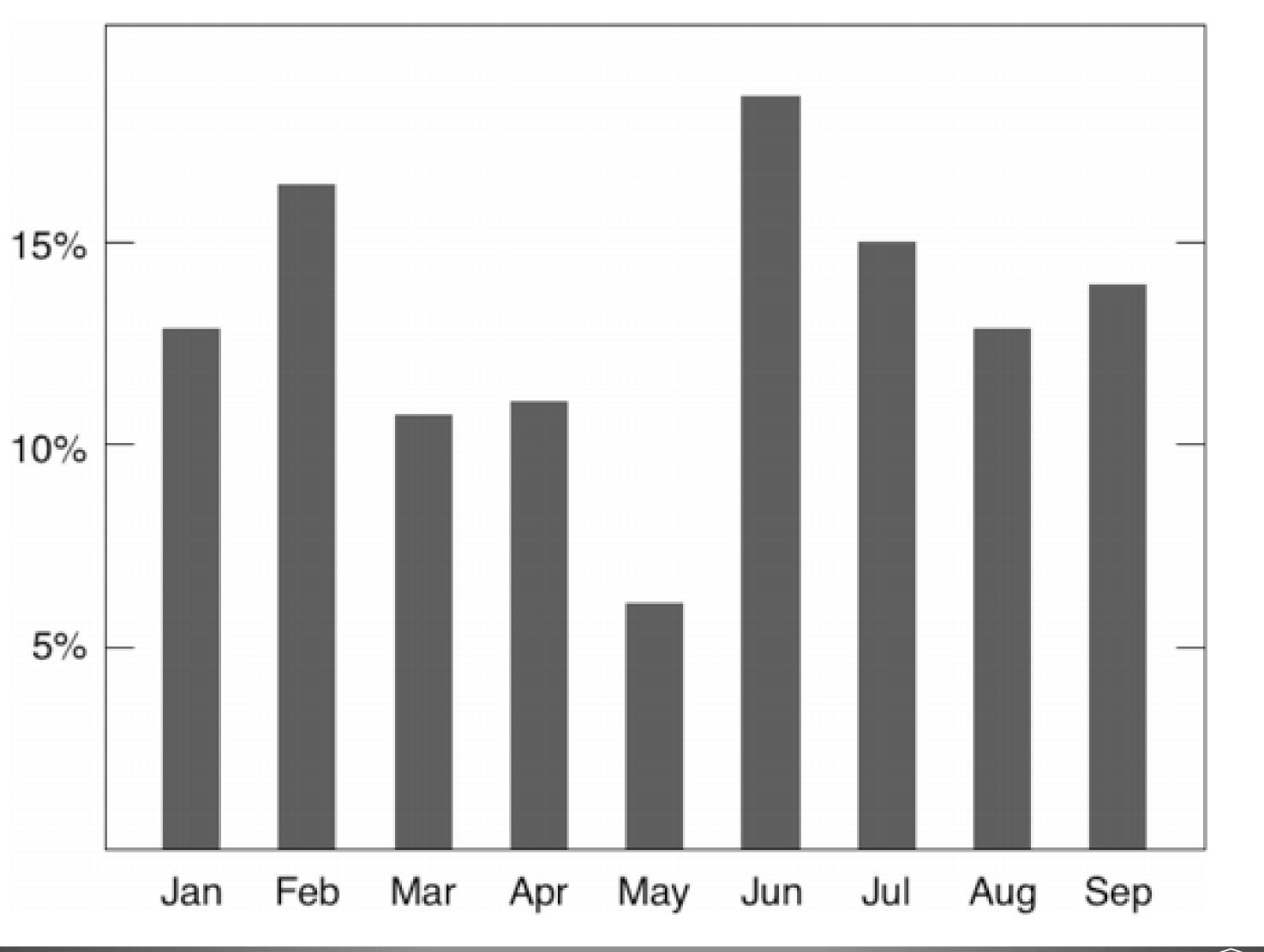








Avoid Chartjunk



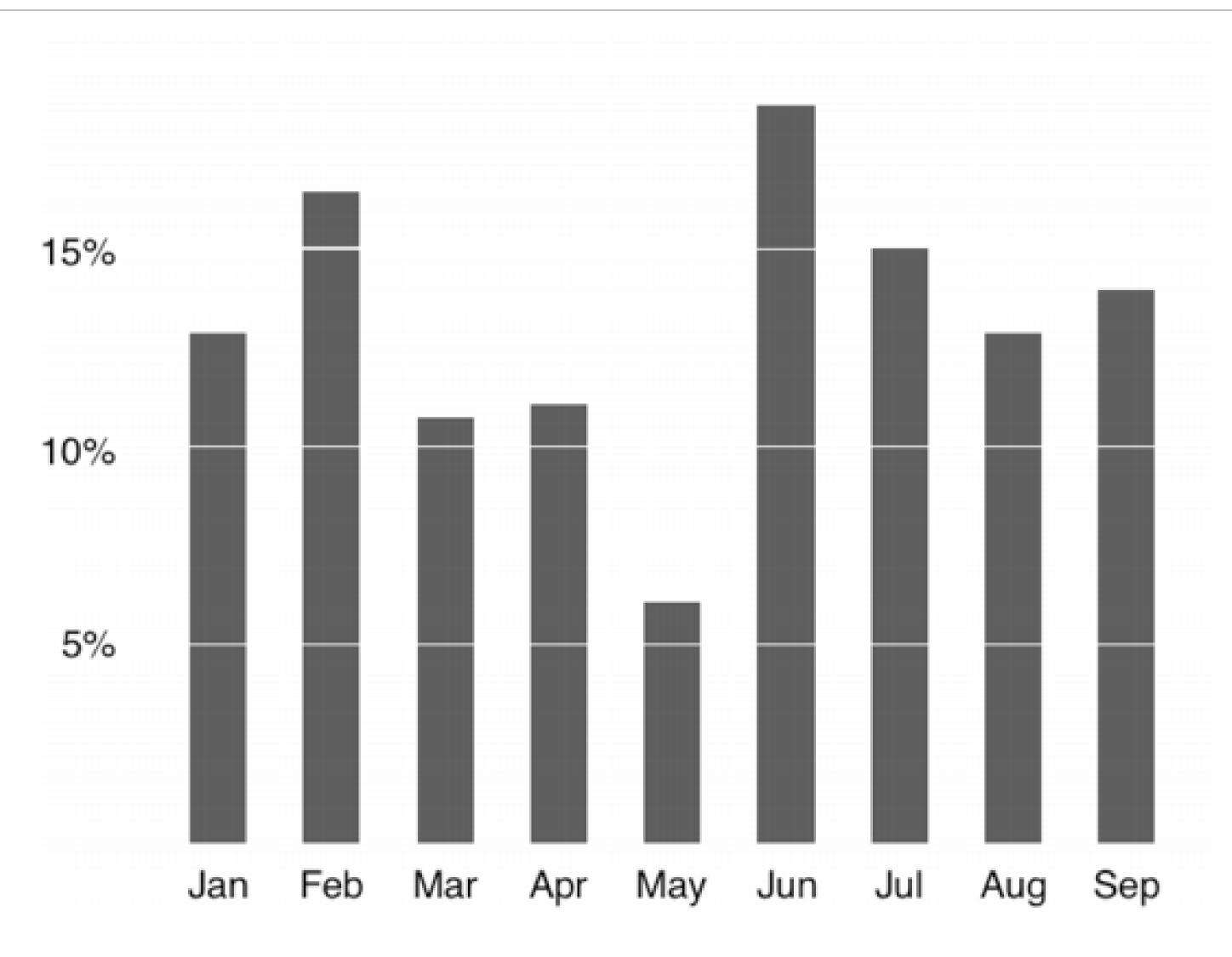








Avoid Chartjunk





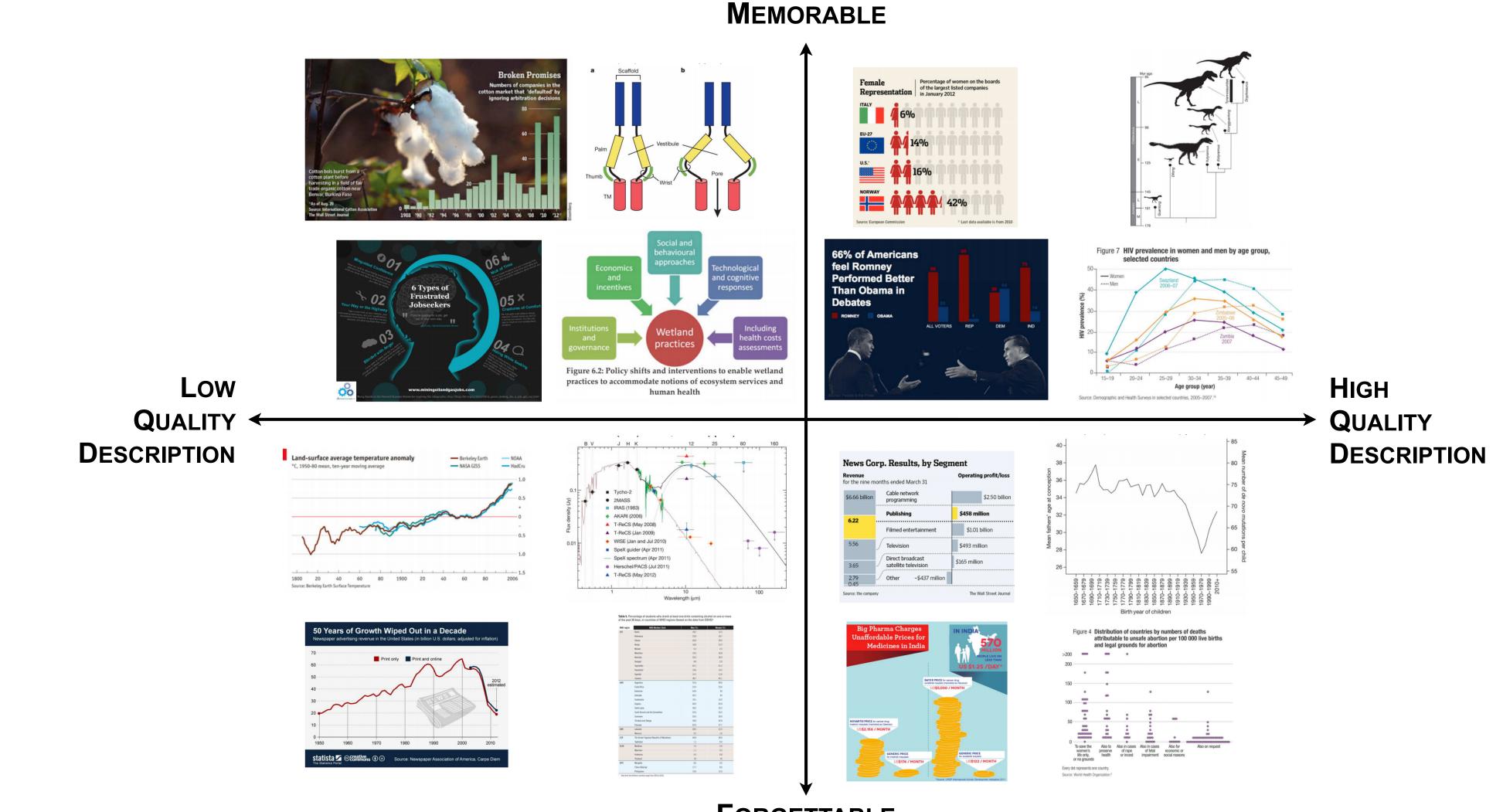








Avoid Chartjunk?



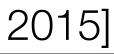
D. Koop, CSCI 627/490, Fall 2024

FORGETTABLE





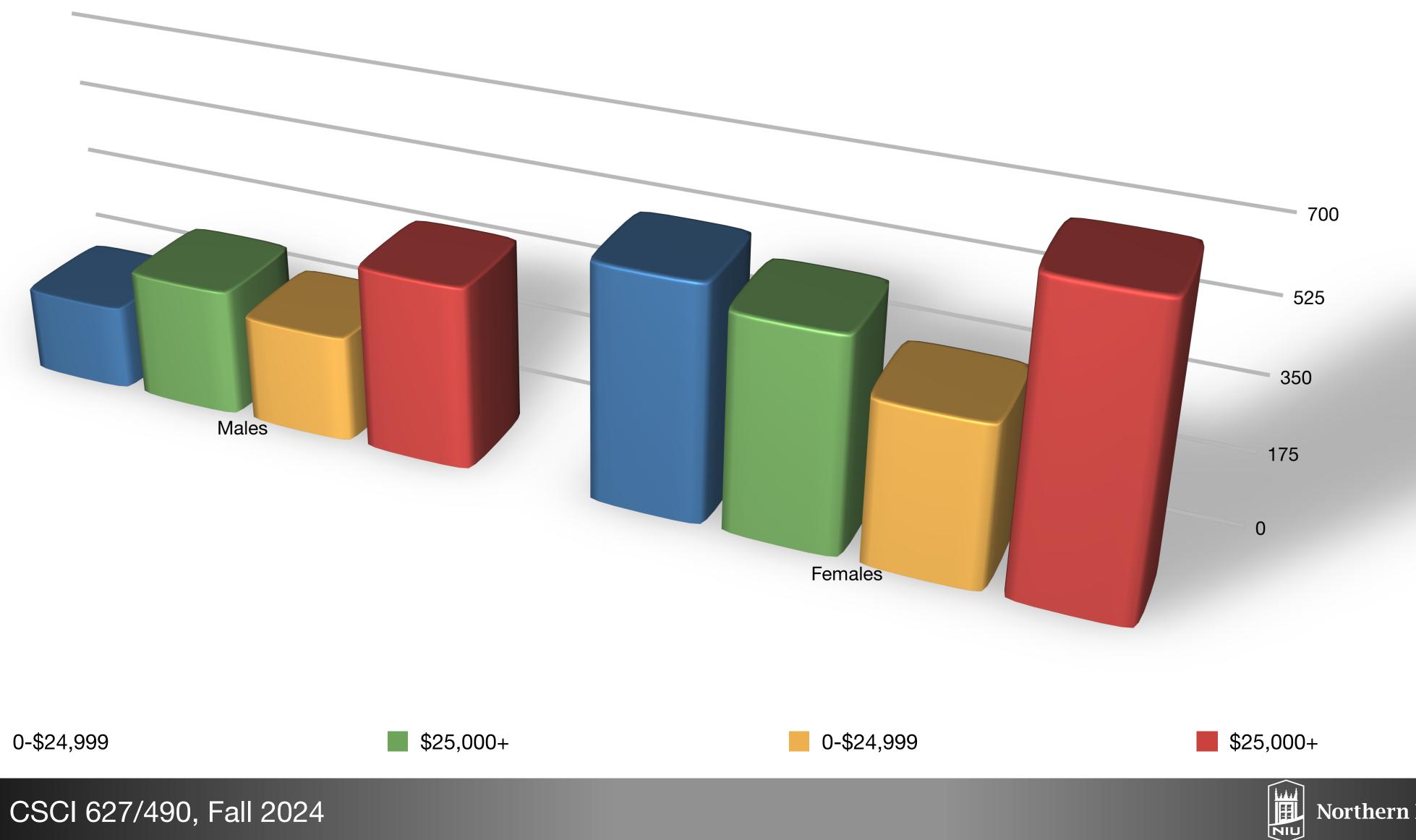
Northern Illinois University

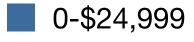






Data-to-Ink Ratio (Also Unjustified 3D)







D. Koop, CSCI 627/490, Fall 2024

Northern Illinois University



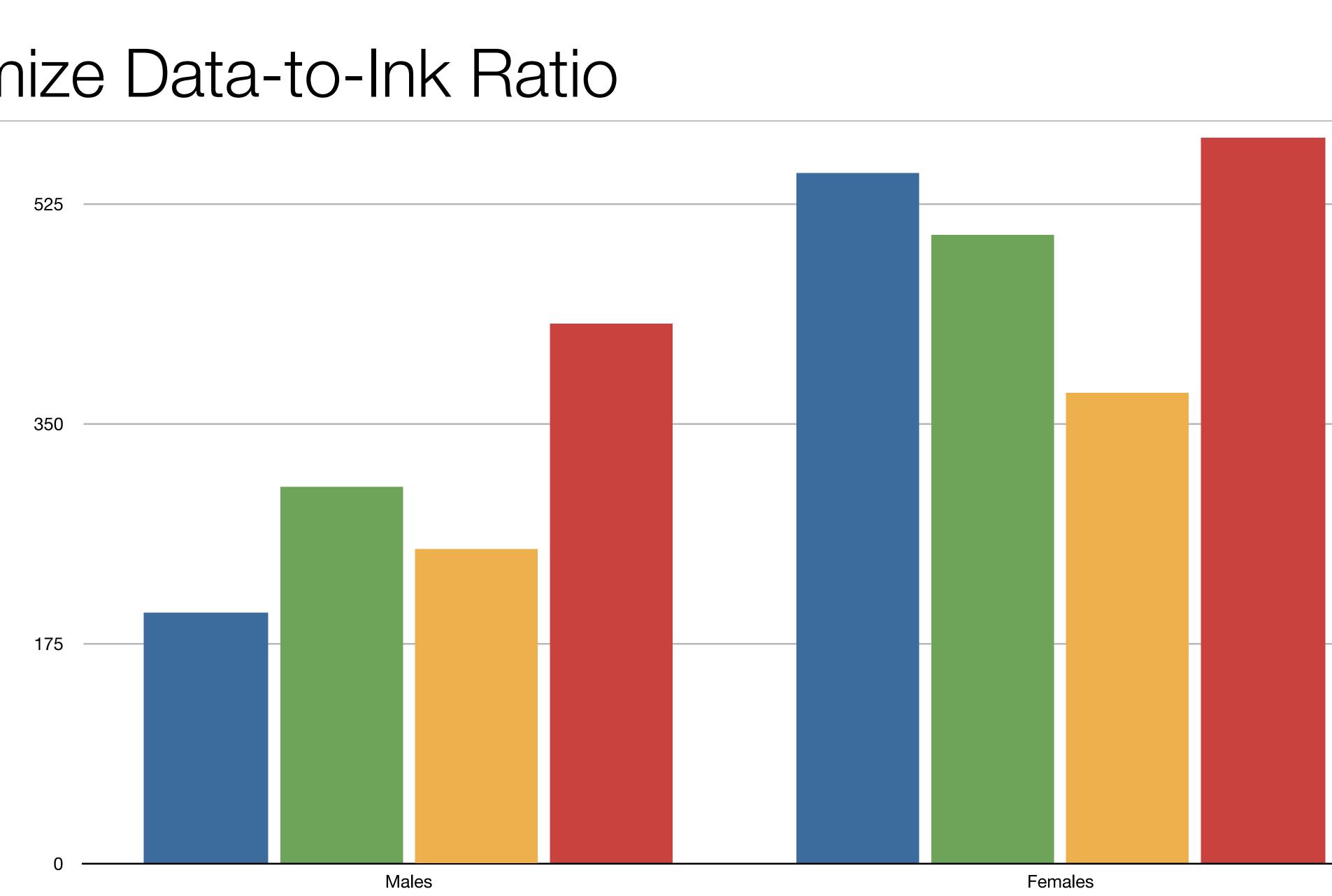






700

Maximize Data-to-Ink Ratio





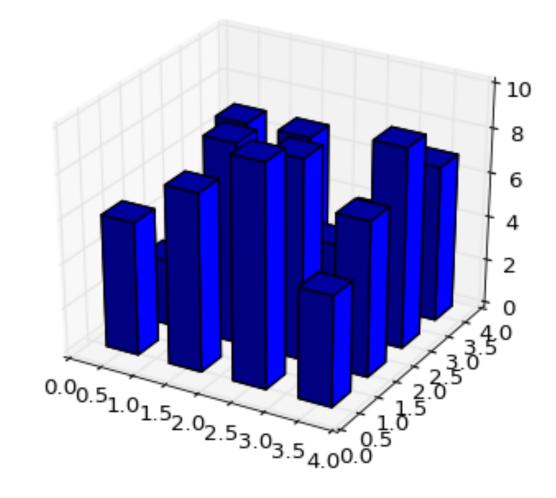


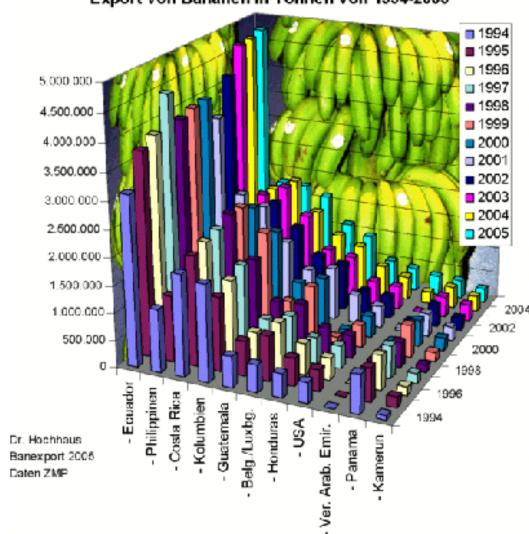






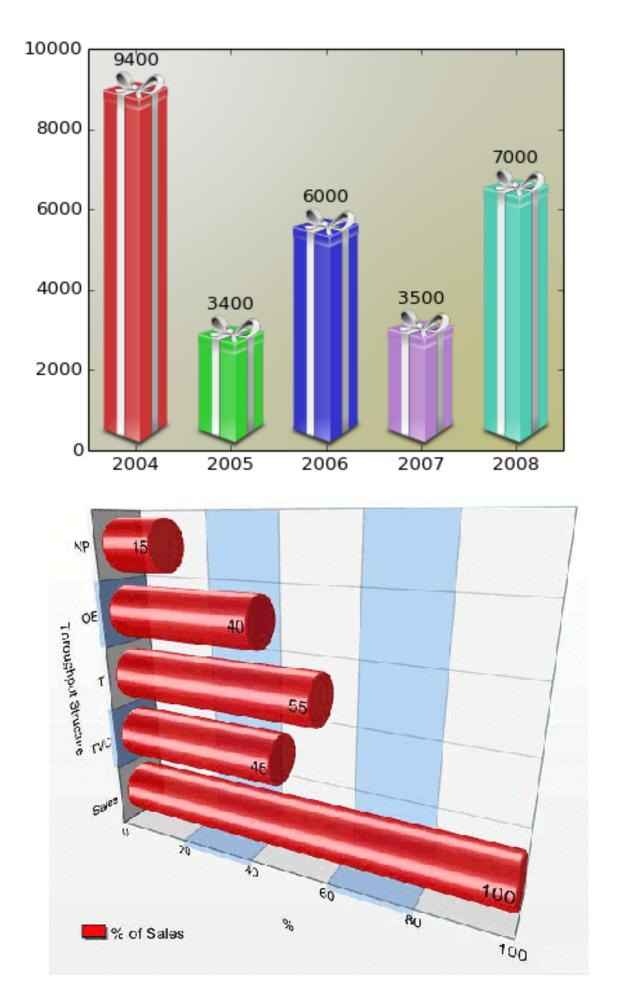
No Unjustified 3D





Export von Bananen in Tonnen von 1994-2005

D. Koop, CSCI 627/490, Fall 2024



matplotlib gallery





Excel Charts Blog

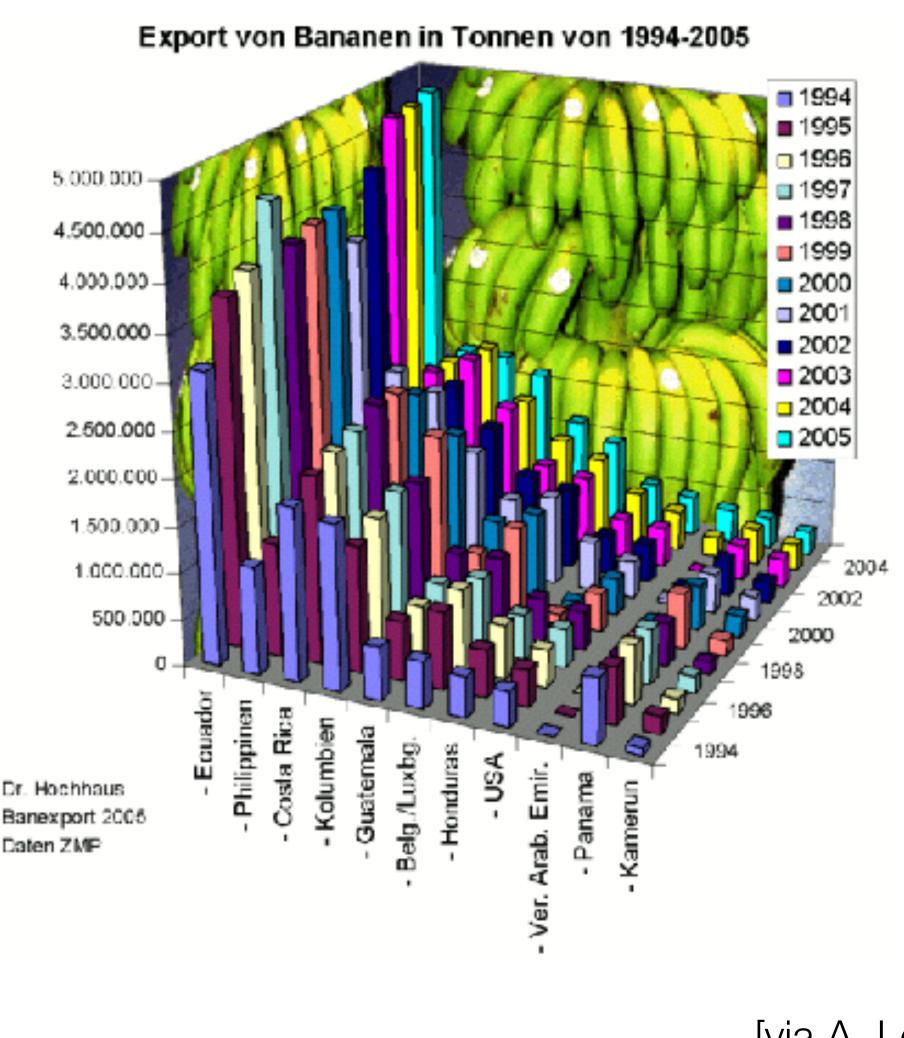






No Unjustified 3D

- Occlusion hides information
- Perspective distortion dangers
- Tilted text isn't legible
- Can help with shape perception











Eyes Beat Memory

- Reduce cognitive load (using up working memory)
- Animation versus side-by-side views
- Change blindness







"Computer-based visualization systems provide visual tasks more effectively."

D. Koop, CSCI 627/490, Fall 2024



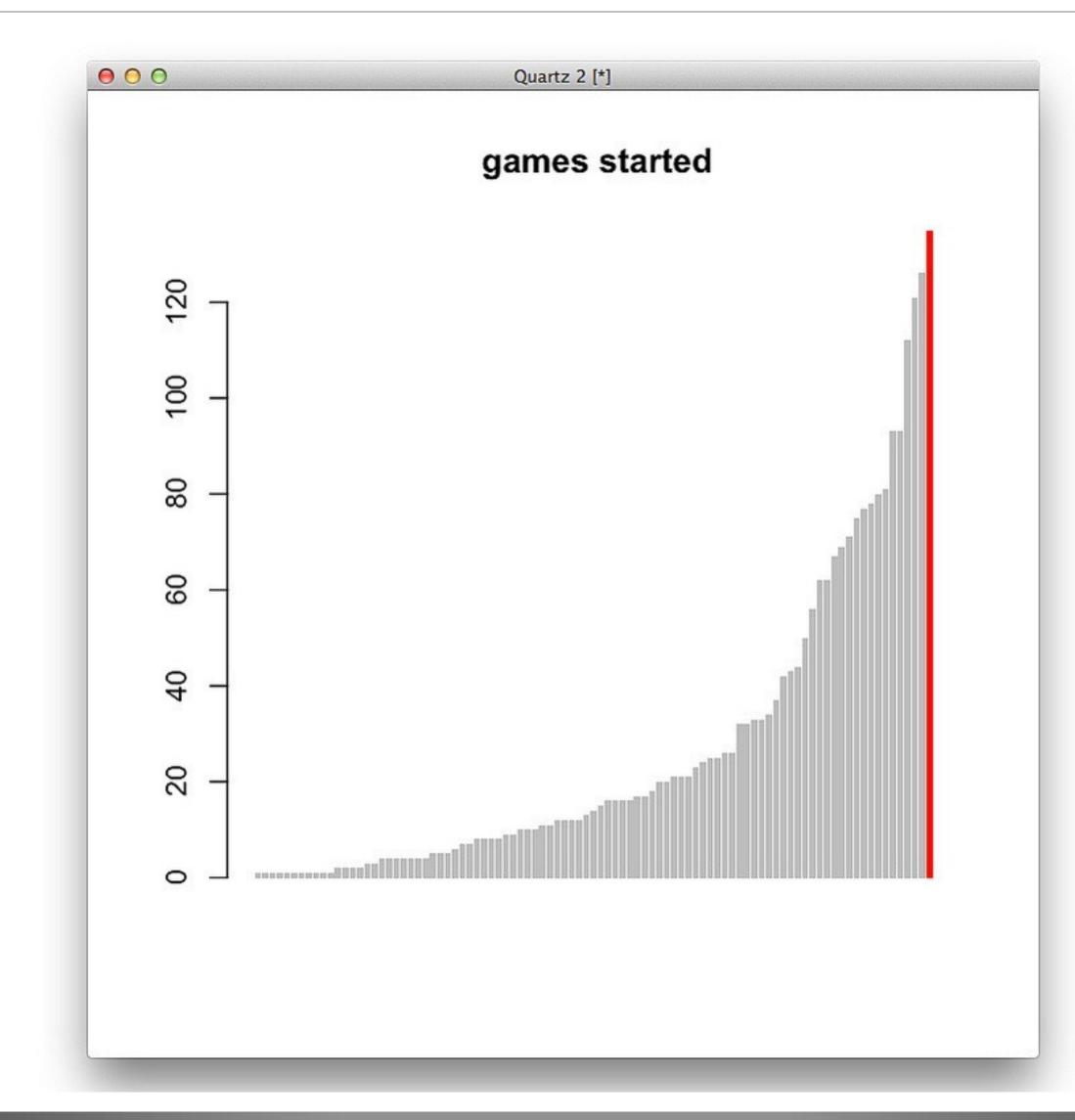
representations of datasets **designed** to help people carry out

– T. Munzner



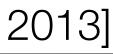


Design Iteration











Design Iteration

New York Giants Indianapolis Colts San Diego Chargers **Baltimore Ravens New England Patriots Green Bay Packers New Orleans Saints Atlanta Falcons New York Jets** Cincinnati Bengals **Houston Texans Carolina Panthers Denver Broncos Arizona Cardinals Jacksonville Jaguars Detroit Lions** ampa Bay Buccaneers **Dallas Cowboys**

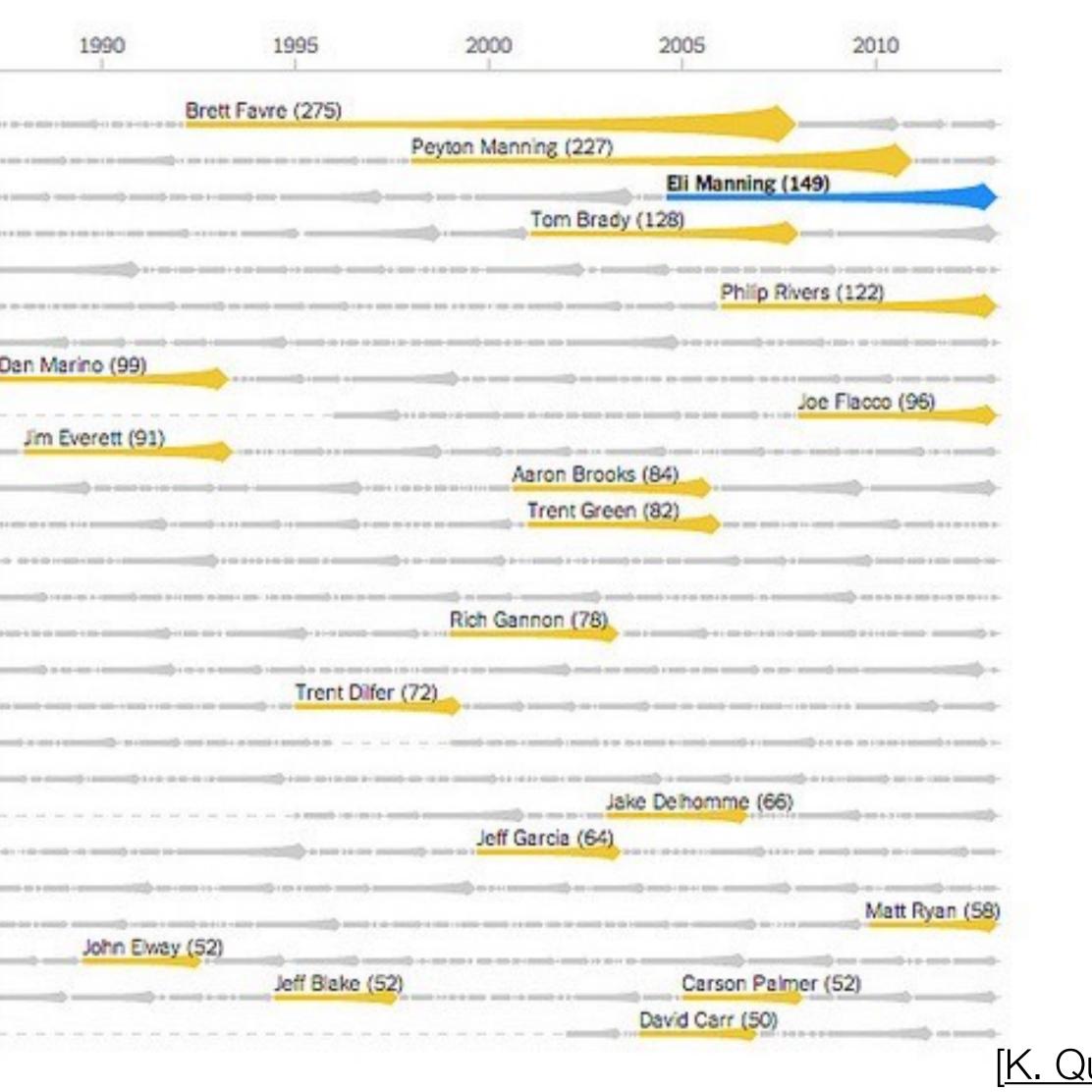
Peyton Manning					Andrew Luck
					Pullinew Eder
	Philip Rivers				
Kyle Boller S	Steve McNair	Joe Flacco			
Tom Brady		Matt Cassel	Tom Brady		
Brett Favre		Aaron Rodgers		Aaron Rodgers	Aaron Rodgers
Aaron Brooks D	rew Brees		Drew Bree	s	
Michael Vick Michael V	/ick	Matt Ryan	Matt Rya	n	
	Chad Penningtor	Brett Favre	Mark Sanch : Mark Sa	nchez	
Carson Palm Carson Palmer		Ryan Fitzp Cars	on Palmer	Andy Dalton	
David Carr		Matt Sch	aub		Matt Schaub
Jake Delhomme		Jake Delhomme		Cam Newto	on
Jake Plummer	Jay Cutler		Kyle Orton	Tim Tebow	Peyton Manning
	Matt Leina · Kur	t Warner			
Byron Leftwich	David Garrard	David Garrard		Blaine Gab	bert
Joey Harrington	Jon Kitna			Matthew Staf	ford
Chris Sim	ms Bruce Gra (Jeff Garci	a	Josh Freema	n	Josh Freeman
Drew Bledsoe	Tony Romo	Tony Romo)	Tony Romo	[<u></u>





Design Iteration

	1970	1975	1980	1985		
PACKERS	100-00-000					
COLTS						
GIANTS						
PATRIOTS						
EAGLES	10-100 (0-1-100) - 1	Real Property and a second sec	on Jaworski (123)			
CHARGERS						
BILLS		ol	e Ferguson (110)		-	
DOLPHINS			and an event on a second second		P	
RAVENS						
RAMS						
SAINTS						
CHIEFS	*******					
REDSKINS			Joe The	eismann (79)		
CARDINALS		Jim Hart (79)				
RAIDERS						
JETS			Richard Todd (76)	-	
BUCCANEERS				()		
BROWNS	Brian Sipe (71)					
SEAHAWKS		Jim Zorn (69)				
PANTHERS				a ai ai ai ai ai ai ai ai		
49ERS						
STEELERS	ana a-sa amatek-mana	Ter	ry Bradshaw (63)	nd-mail a stramind-s s-m res		
FALCONS	ADDID 110-10-10000		Steve Bartk	owski (58)		
BRONCOS					D-1	
BENGALS	a minimum para mana mana				0	
TEXANS		Roger	Staubach (50)			









Design

- Unlike a math problem, there are many different approaches for the visualization of some data
- Need to have some way to discuss how to determine whether a visualization is doing what we want
- Validation: Understand why a design is effective
 - What problems can be effective
 - Do this at different levels

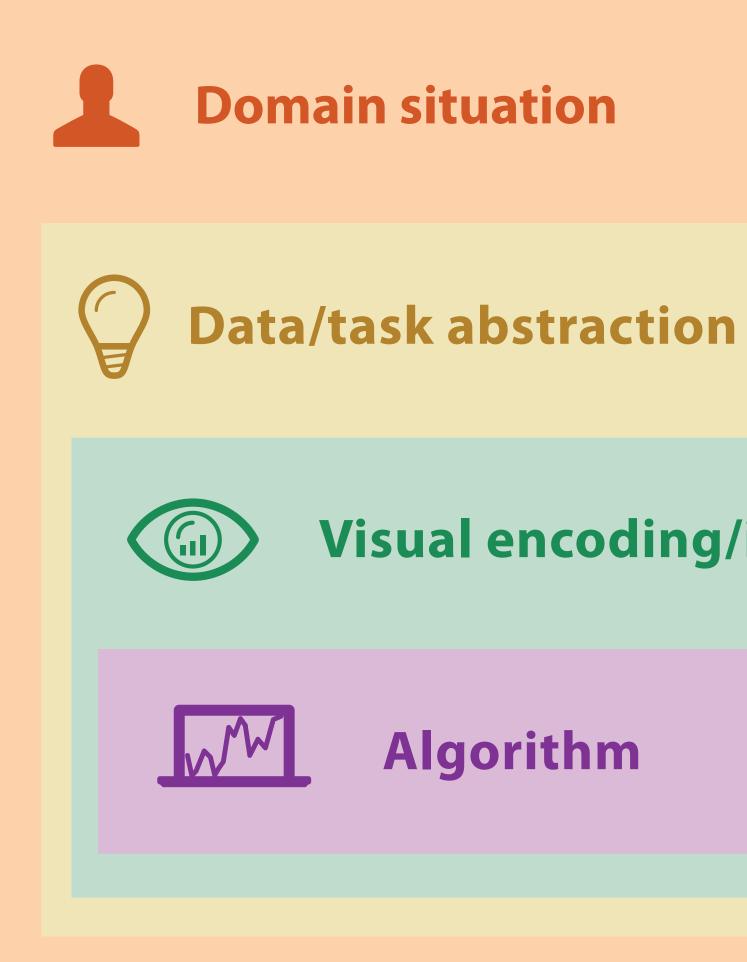
D. Koop, CSCI 627/490, Fall 2024





38

Four Nested Levels of Design



D. Koop, CSCI 627/490, Fall 2024

Visual encoding/interaction idiom











Potential problems at each level

Domain situation You misunderstood their needs

Data/task abstraction You're showing them the wrong thing



Wisual encoding/interaction idiom The way you show it doesn't work

Algorithm Your code is too slow

D. Koop, CSCI 627/490, Fall 2024







40

Validation at each level

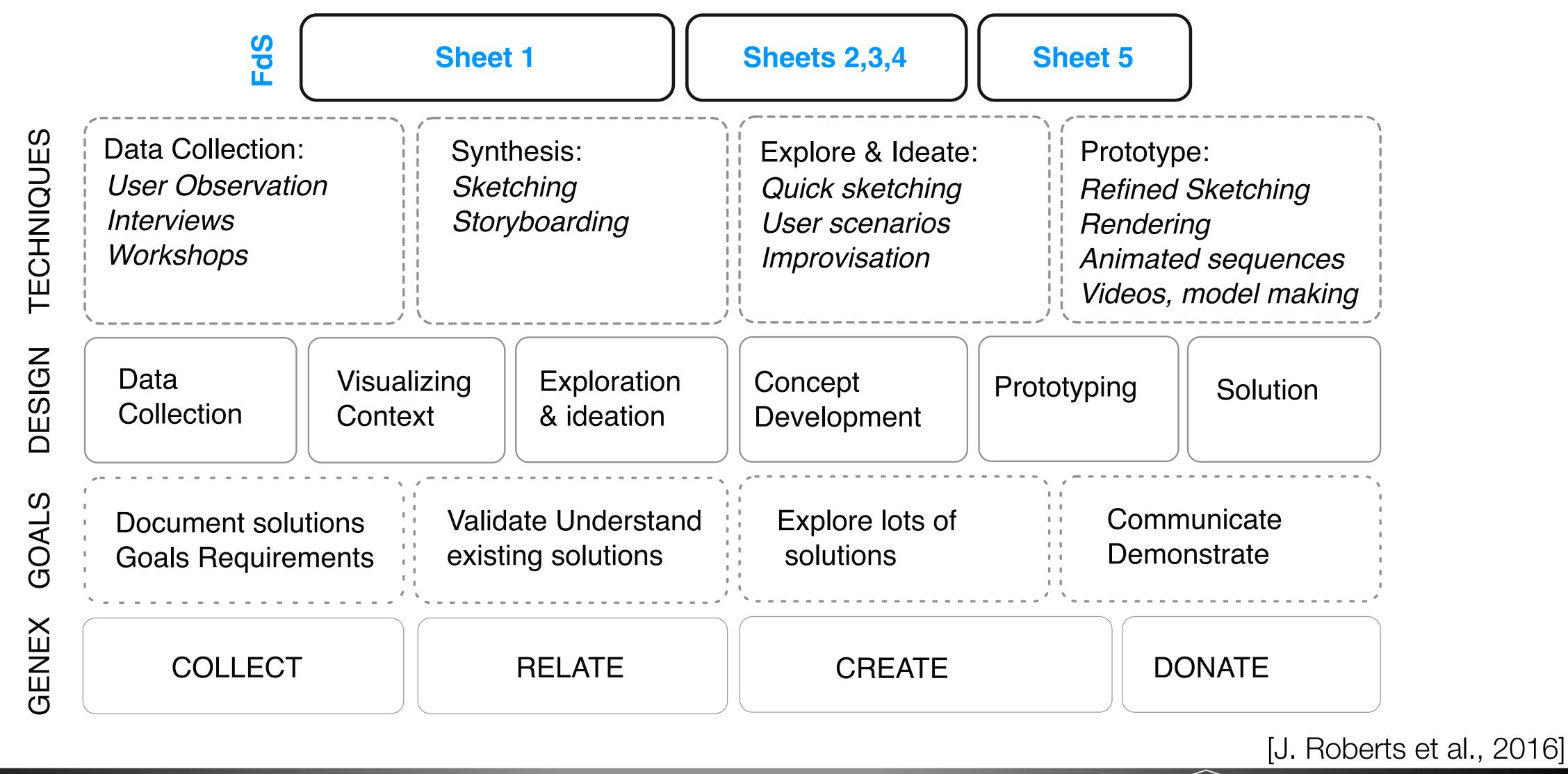


- Ineffective encoding/interaction idiom
- Validate Test on target users, collect anecdotal evidence of utility Validate Field study, document human usage of deployed system





Five Design-Sheet Methodology









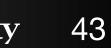
Five Stages

- 1. Meet with client and consider task; or contemplate task on own.
- 2. Ideate and sketch small ideas.
- 3. Sketch and plan three alternative designs.
- 4. Consider solutions with client; or deliberate on own.
- 5. Generate realization sheet, and implement prototype. Discuss with client and re-iterate if necessary.

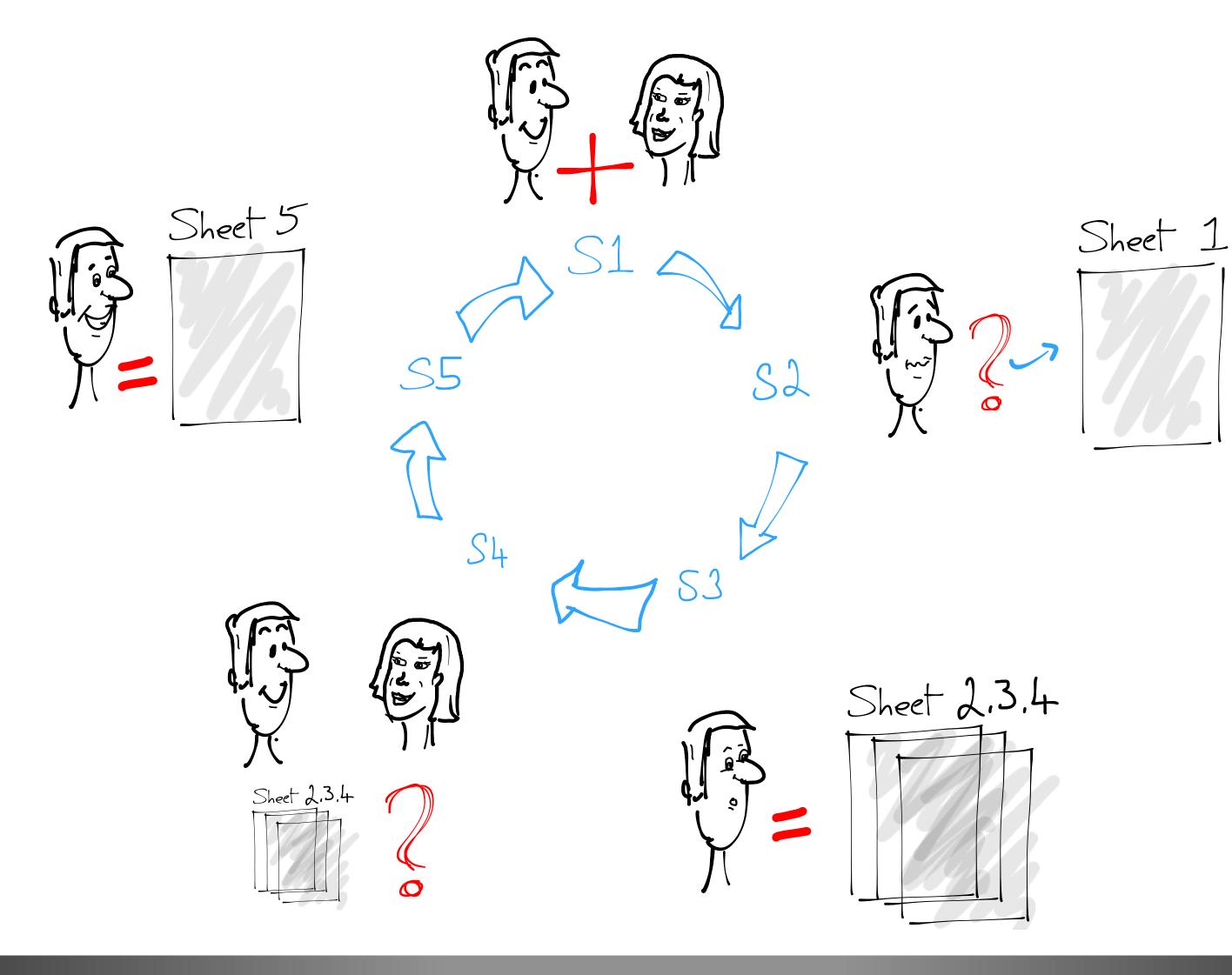








Five Stages



D. Koop, CSCI 627/490, Fall 2024



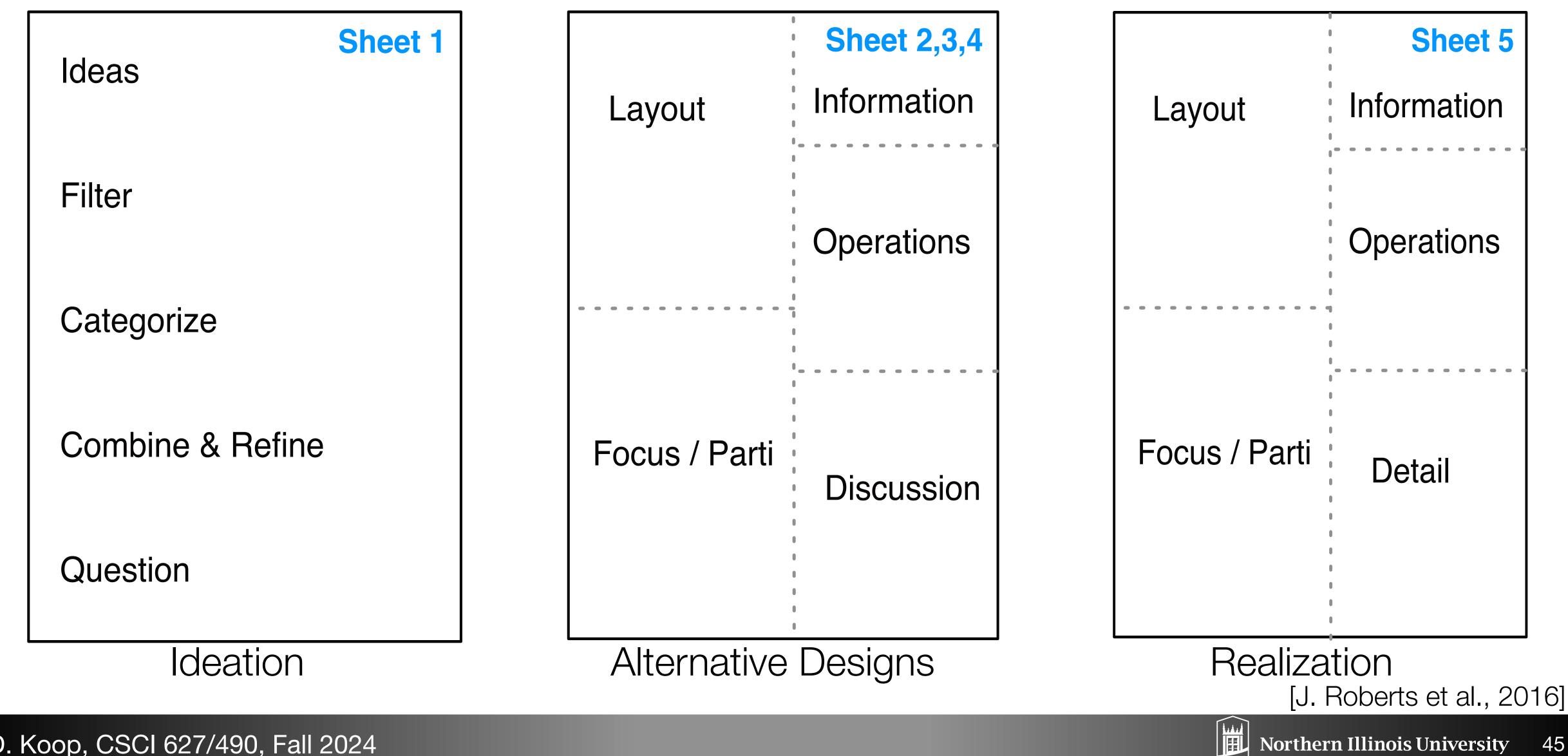


Northern Illinois University





The Five Sheets



D. Koop, CSCI 627/490, Fall 2024



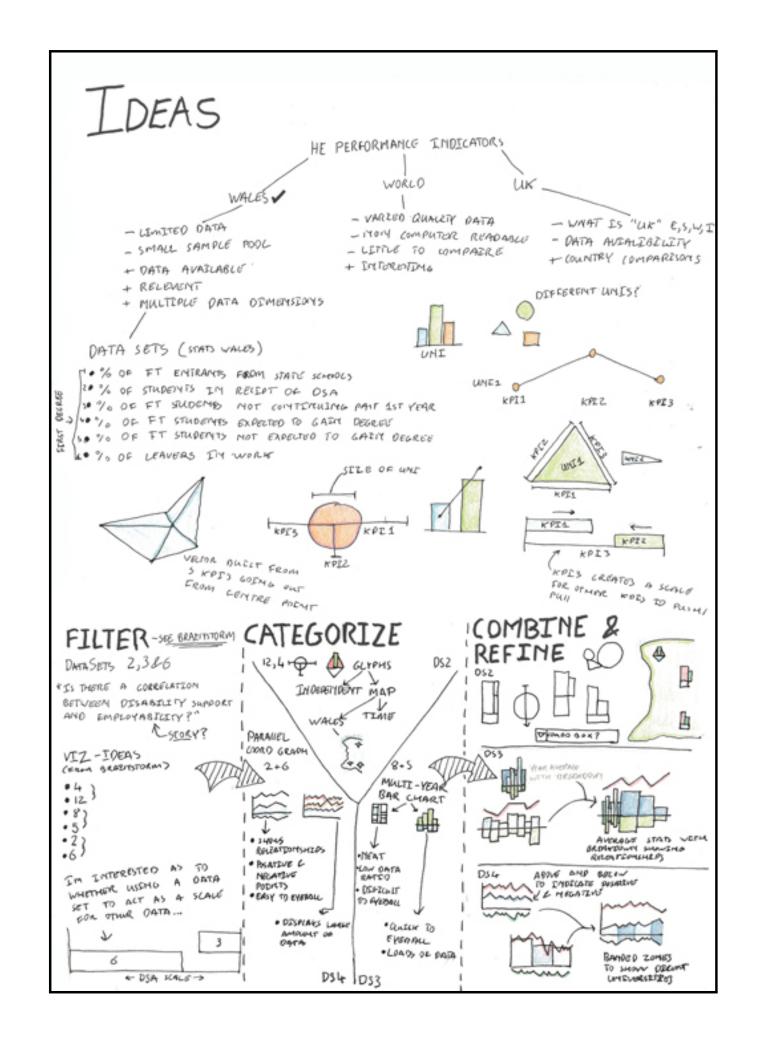




NIU



Example: University Access for Disabled Students





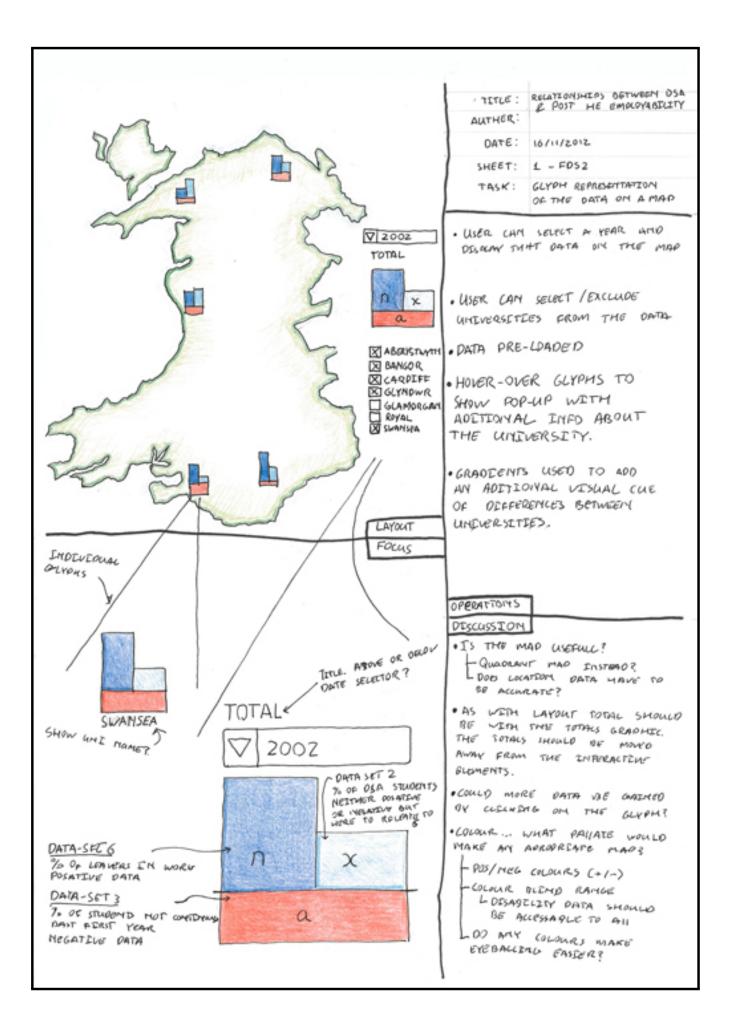


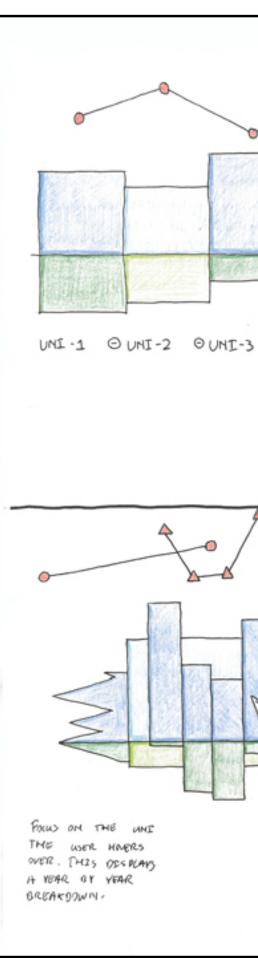




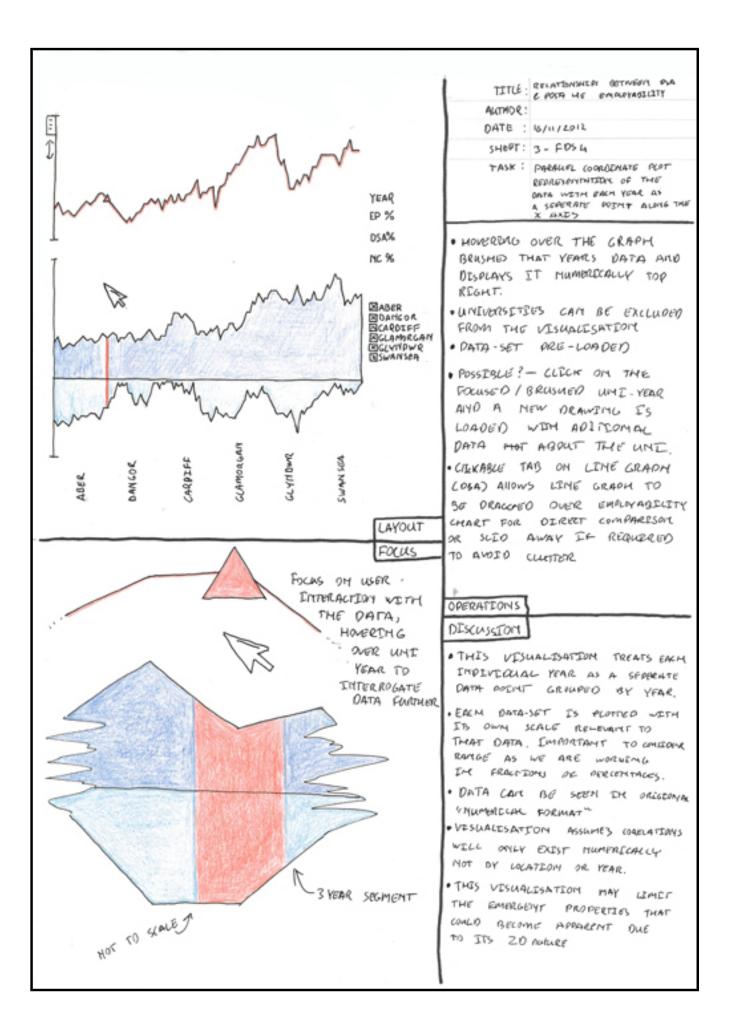


Sheets 2-4





	TITLE : AGLATIONSHEDS OF HEAD OSA & DOST WE BHOLOYADILITY AUTHER: DATE : 16/11/2012 SHEET : 2 - FOS 3 TASK : BAR-CHART REPRESENTATION PF THE OSA . EMPLOYABILITY & LEANERS DATA
OUNI-4 O	 HOLERING OVER EACH AVERAGED UNTEVERSIFY BAR DRAWS THE YBAR BY YEAR BREAKDOWN. (LICKENG ON AN INDEVEDUAL YEAR WOULD SET ALL DARAGRAUMS "MAIN" BARS TO THE FIGURES OF THAT YEAR ALONE FOR (OMPARISON. CLECKING ON THE MINUS SYMBOL WOULD MENIMISE THAT UNIL WITH THE REMARKED UNILS EXAGNOING TO FILL THE SPACE. DATA DRE-LOADED ADDICTY TO CHANGE DISPLAY HOME COLDURS FOR DEFERENT USERS PREFERENCES?
Mor TO Scale	OPERATIONS DISCUSSION • DOES MENTEMISDIG' UNIS ADD ANY USEFULI DITTERACTIONS • DUADNG THE HOVER - OVER BREAK DOWN OF DIATA SHALLO THE VERA OF DISPLAYED IN TOXT POR FASE OF DUTEPRIDATION • SHOW ALL BUTTON COULD BE WARCHI BLAT IT MAY CLUTTER THE DATA. • VISUALISATION ASSUMES THE ONLY CORELATIONS WILL BE NUMERICUL. • COULD OTHER DATA BE ODSPLAYED ON THE BARS? PERCENTAGES DE DEGREE CLASSEFICATEON BY TYPE?



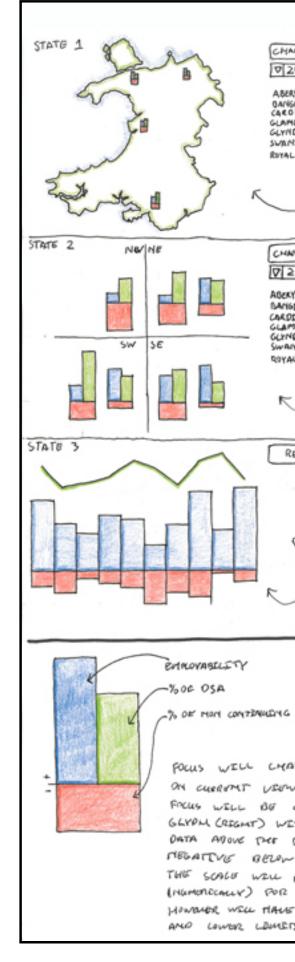








Sheet 5



D. Koop, CSCI 627/490, Fall 2024

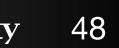
	TITLE : ARCHTENSHERS BETWEEN DA
HANGS VEDW	AUTHOR:
2004	DATE : 17/11/2012
REATSTWATH 23 Moore 23	SHEET : 4 - FPS 5
KOIFF BANDRGAN D	TASK : FENAL DESTAN CONCEPT WARAS STATUS, MAR, RUASEDTT MAR
TAL D PAP SSUAL	ONELOTMOT OFFICEPO'D RREAT FOS2, FOS3 & CLEVINT COMMYECATEON
LOCATEON COL DEATEON	· All DATA DEG LOADED
LOCATEON COL STAT	· SWETCH VEEN" BUTTON AROWS USER TO SWETCH DETWEEN STATE & STATE Z
2004	· IN STATE 2 CLECKENG ON A LINE GLYDH THANSETERYS TO STATE 3,
ERISTWITH B	A FUIL BREAKDOWN OF THE UNIS DATA BY YEAR.
ngdr Bi Kolff Bi Amorgan Bi Knowr Bi	· EN STATE 3 CLECKENG ON THE "RETURN" OLUTION WOULD RETURN THE WARE TO STATE 2.
YAL SEMPLEFICE EASLER	WHE TO STATE 2. • WHES CAM BE EXCUDED / INCLUDED IN THE VEW (STATES 1 62) BY
MAP ADE ATTOM MAP ADE ATTOM MAP ADE ATTOM INTRACEATON INTRACTION	SELECTERS AND DESELECTERS THEM.
ENTRACE ATTA	· YEAR SELECTABLE BY USER
0° 0°	
RETURN ONE - CUBRT REPRESENTATION ONE ALL THE UNCUBLICK	
TALIAL	OPERATIONS
Othe - CHART REPRESENT	DETAIL
CLART OWTO WATE	. TEME TO BUELD ESTEMATED AT
BAR ALL LUP WALTER	16 HOURS
a sen phund p	· DATA-SETS AQUIRED FROM SCATS WALES
Ora ALL THE UNCOME B STREET BACK OF A SCORE FALLY OF A SCORE FOR STREET	SCALE FOR EACH ITEM SET TO
yene us	THE RANGE AS THE DIFFERMUE
LAYOUT	13 FRACTSONS OF %3.
FOCUS	and a start of the old
	SURVEY DATA, HOWEVER WOULD NOT BE SUSTABLY FOR COLOUR BLETD. MAY CHOSE YOULD, PURPLO & PLUE
G AFTER 1ST YEAR	TO SUET ORITERANDMALY & PROTANDPY COLOMR BLENDNESS.
	* D85AAG - 0146.
CANGE DEPENDING	жеве544 - Yirlow, Ф. В33552 - РИХРИГ,
AV STATE . MARN	· DATA-SET WILL MEED SAME COUNTING
ON THE BI-POLAR	AS NOT ALL UNDY HAVE ALL VEARS
PETH POSATEVE	Oata.
DATION AND	· MAY NOT DICLUDE STATE THREE L STATE THREE BASED ON FDS3.
BE REFERENT	DIME ON FOSS
2 MAR PATA - STRM	
IS THE SAME UPDOR	
ers cosciaccy.	1





Northern Illinois University





Prototype

