Data Visualization (CSCI 627/490)

Introduction

Dr. David Koop





What is Data Visualization?









How is it different from Computer Graphics?







The purpose of computing is about insight, not numbers

D. Koop, CSCI 627/490, Fall 2022

- R. W. Hamming





4

The purpose of visualization is about insight, not pictures

- Card, Mackinlay, Shneiderman









Why do we visualize data? (vs. looking at tables?)







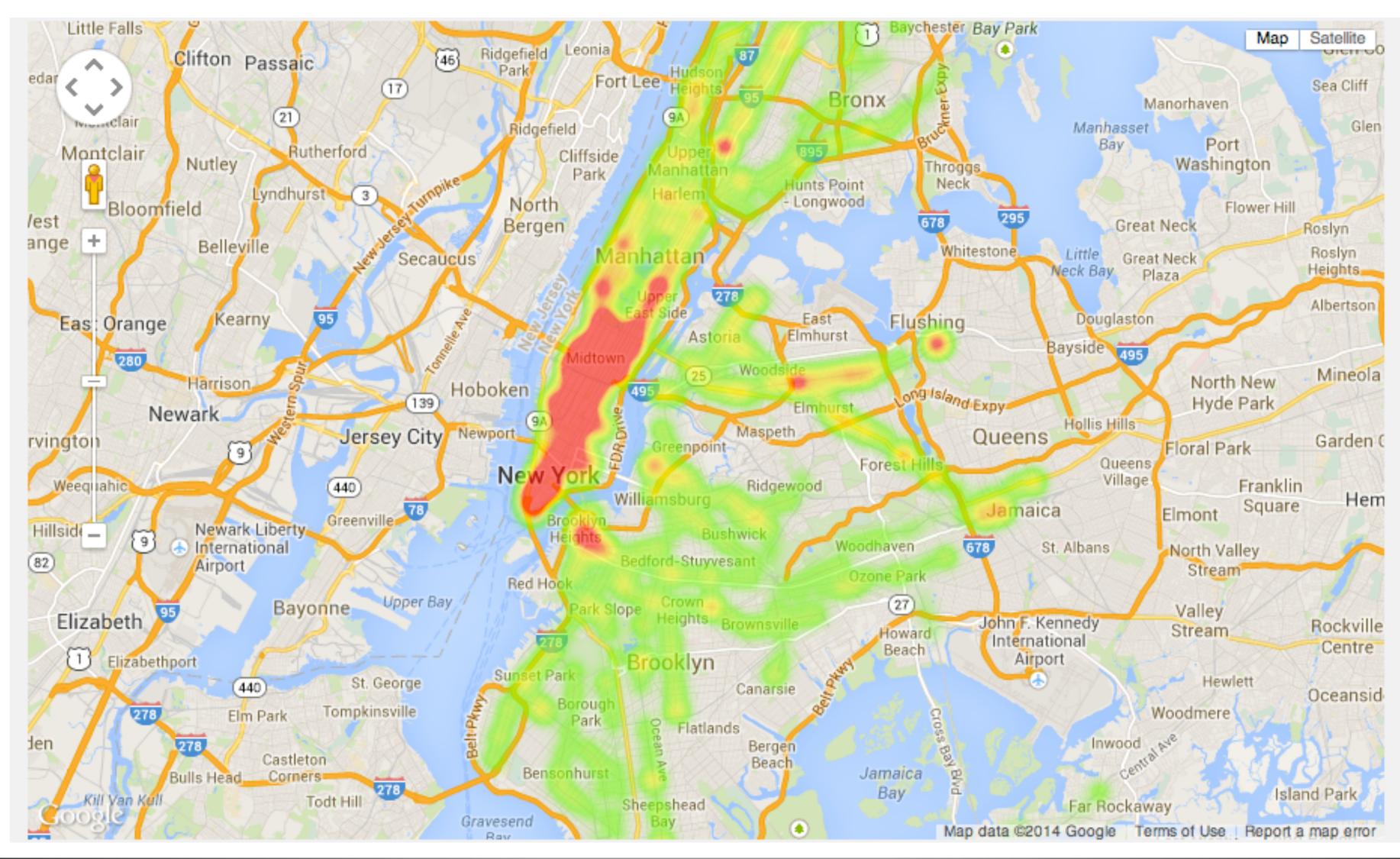
MTA Fare Data Table

	REMOTE	STATION	FF V	SEN/DIS	7-D AFAS UNL	D AFAS/RMF I	JOINT RR TKT	7-D UNL	30-D UNL
1 F	R011	42ND STREET & 8TH AVENUE	00228985	00008471	00000441	00001455	00000134	00033341	0007125
2 F	R170	14TH STREET-UNION SQUARE	00224603	00011051	00000827	00003026	00000660	00089367	00199843
3 F	R046	42ND STREET & GRAND CENTRAL	00207758	00007908	00000323	00001183	00003001	00040759	00096613
4 I	R012	34TH STREET & 8TH AVENUE	00188311	00006490	00000498	00001279	00003622	00035527	00067483
5 F	R293	34TH STREET - PENN STATION	00168768	00006155	00000523	00001065	00005031	00030645	00054376
6 I	R033	42ND STREET/TIMES SQUARE	00159382	00005945	00000378	00001205	00000690	00058931	00078644
7 F	R022	34TH STREET & 6TH AVENUE	00156008	00006276	00000487	00001543	00000712	00058910	00110466
8 F	R084	59TH STREET/COLUMBUS CIRCLE	00155262	00009484	00000589	00002071	00000542	00053397	00113966
9 I	R020	47-50 STREETS/ROCKEFELLER	00143500	00006402	00000384	00001159	00000723	00037978	0009074
10 F	R179	86TH STREET-LEXINGTON AVE	00142169	00010367	00000470	00001839	00000271	00050328	00125250
11 F	R023	34TH STREET & 6TH AVENUE	00134052	00005005	00000348	00001112	00000649	00031531	00075040
12 F	R029	PARK PLACE	00121614	00004311	00000287	00000931	00000792	00025404	00065362
13 F	R047	42ND STREET & GRAND CENTRAL	00100742	00004273	00000185	00000704	00001241	00022808	0006821
14 F	R031	34TH STREET & 7TH AVENUE	00095076	00003990	00000232	00000727	00001459	00024284	0003867
15 F	R017	LEXINGTON AVENUE	00094655	00004688	00000190	00000833	00000754	00020018	0005506
16 F	R175	8TH AVENUE-14TH STREET	00094313	00003907	00000286	00001144	00000256	00038272	0007466
17 F	R057	BARCLAYS CENTER	00093804	00004204	00000454	00001386	00001491	00039113	0006811
18 F	R138	WEST 4TH ST-WASHINGTON SO	00093562	00004677	00000251	00000965	00000127	00031628	0007445





MTA Fare Data Visualization



D. Koop, CSCI 627/490, Fall 2022



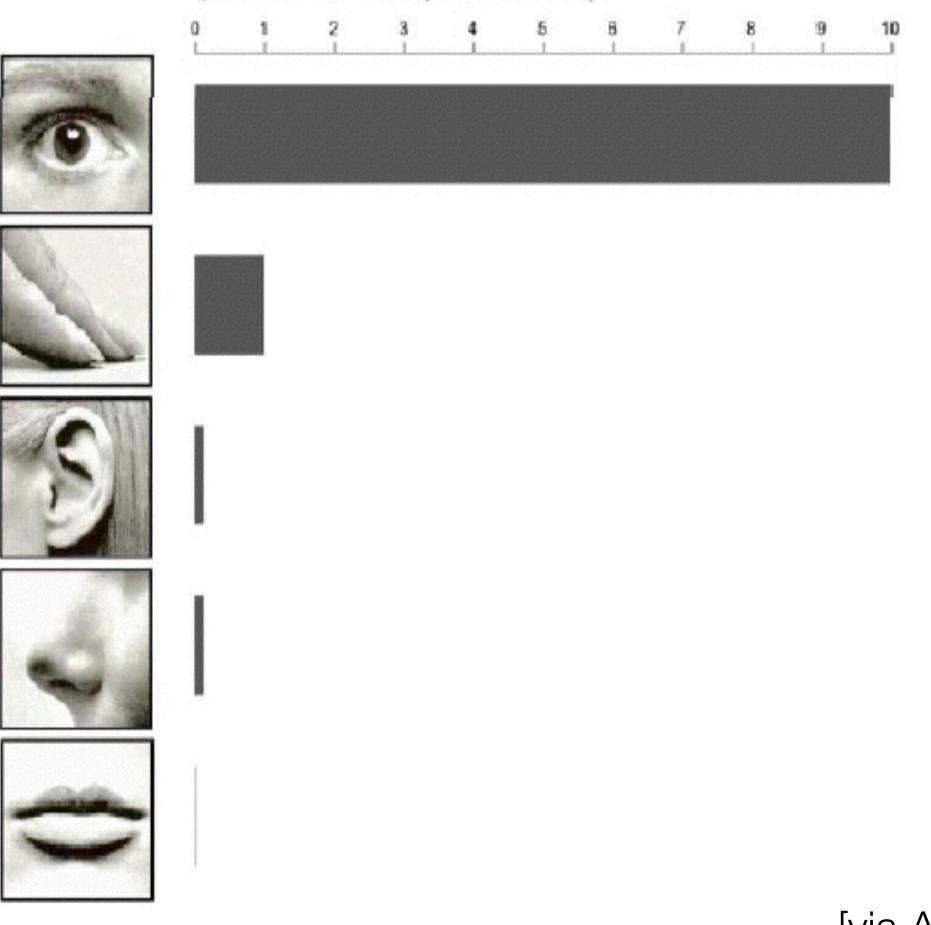


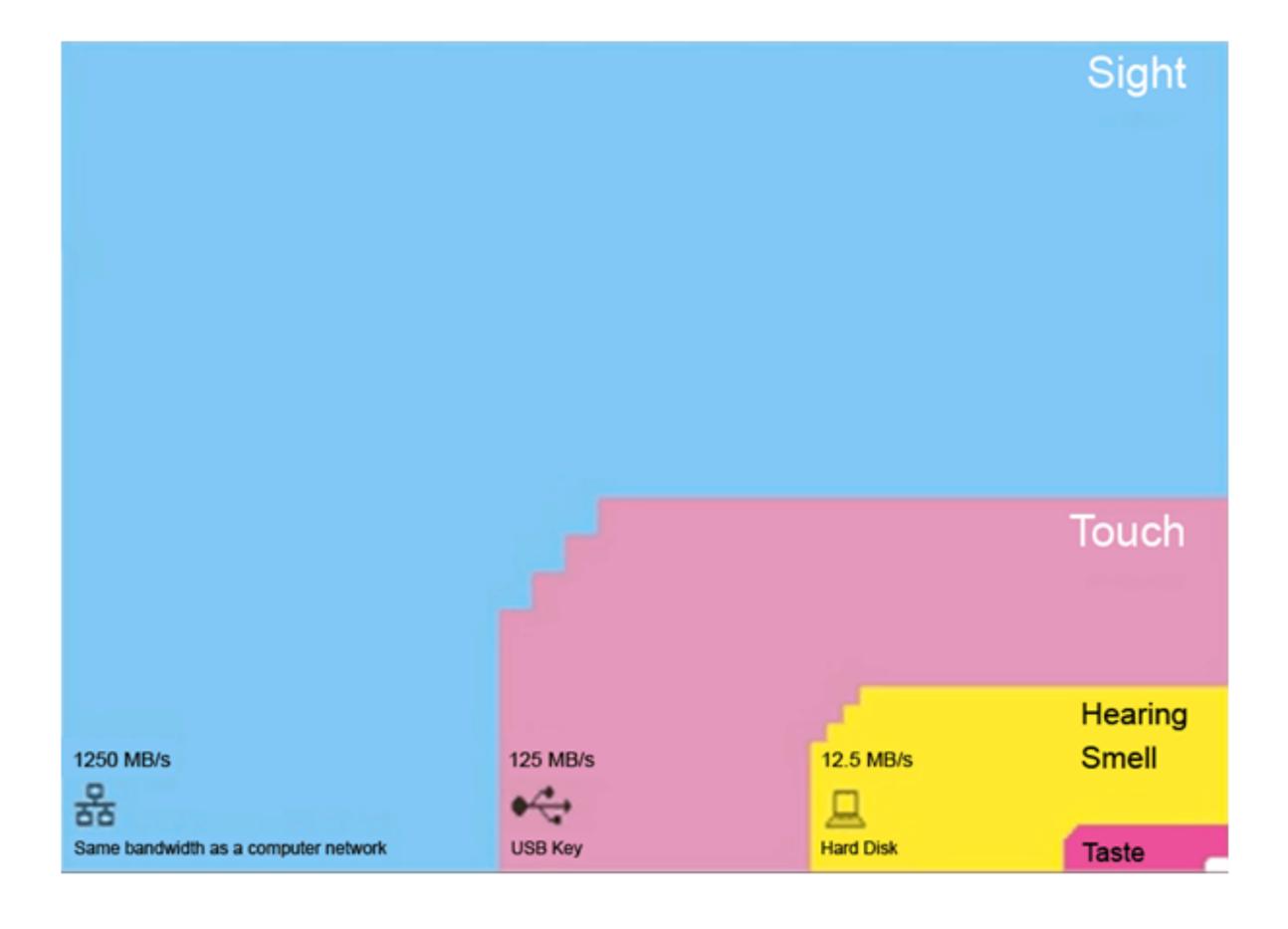
8

Why do we visualize data?

Total Bandwidth

(millions of bits per second)













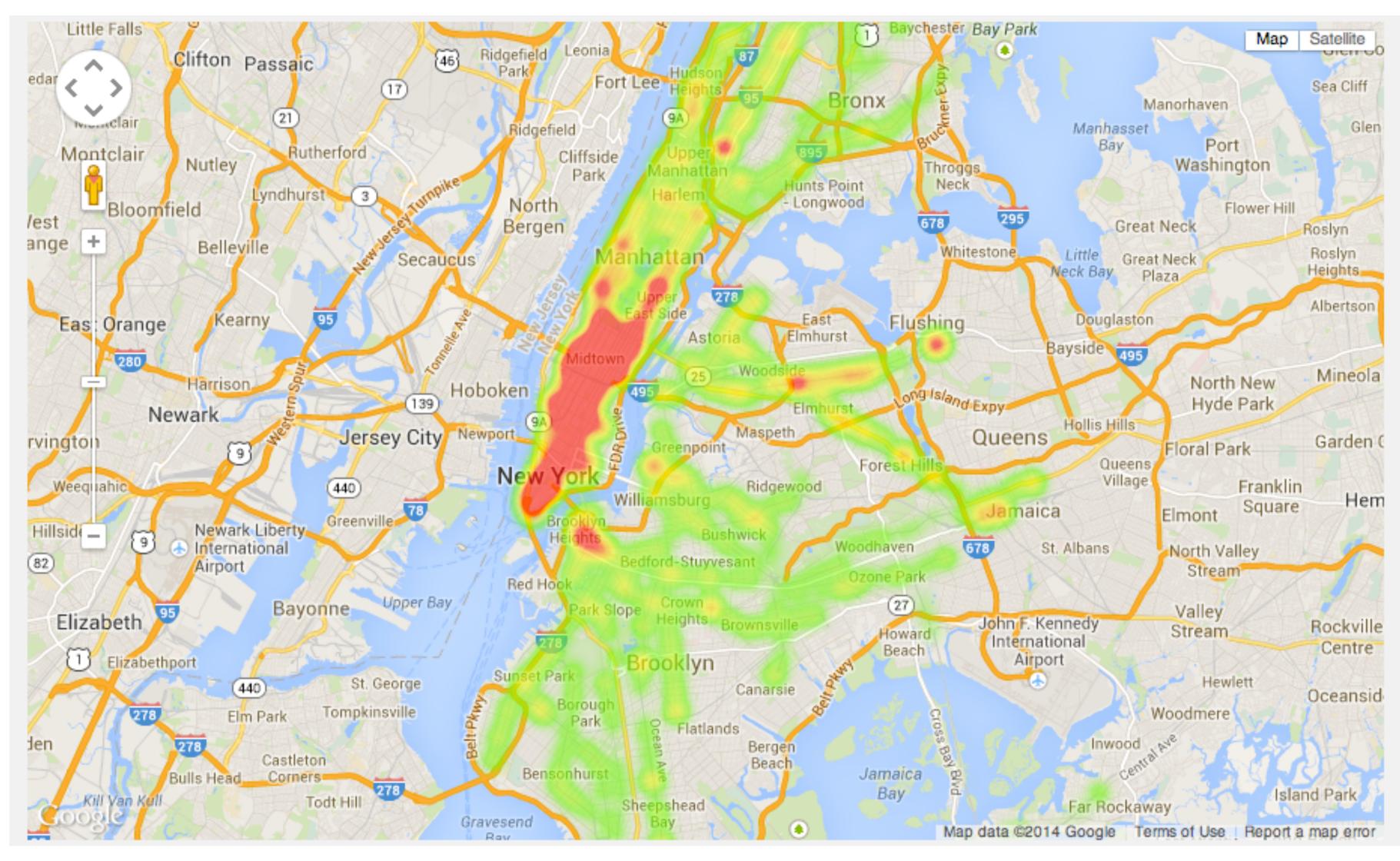


What are the purposes for visualization?





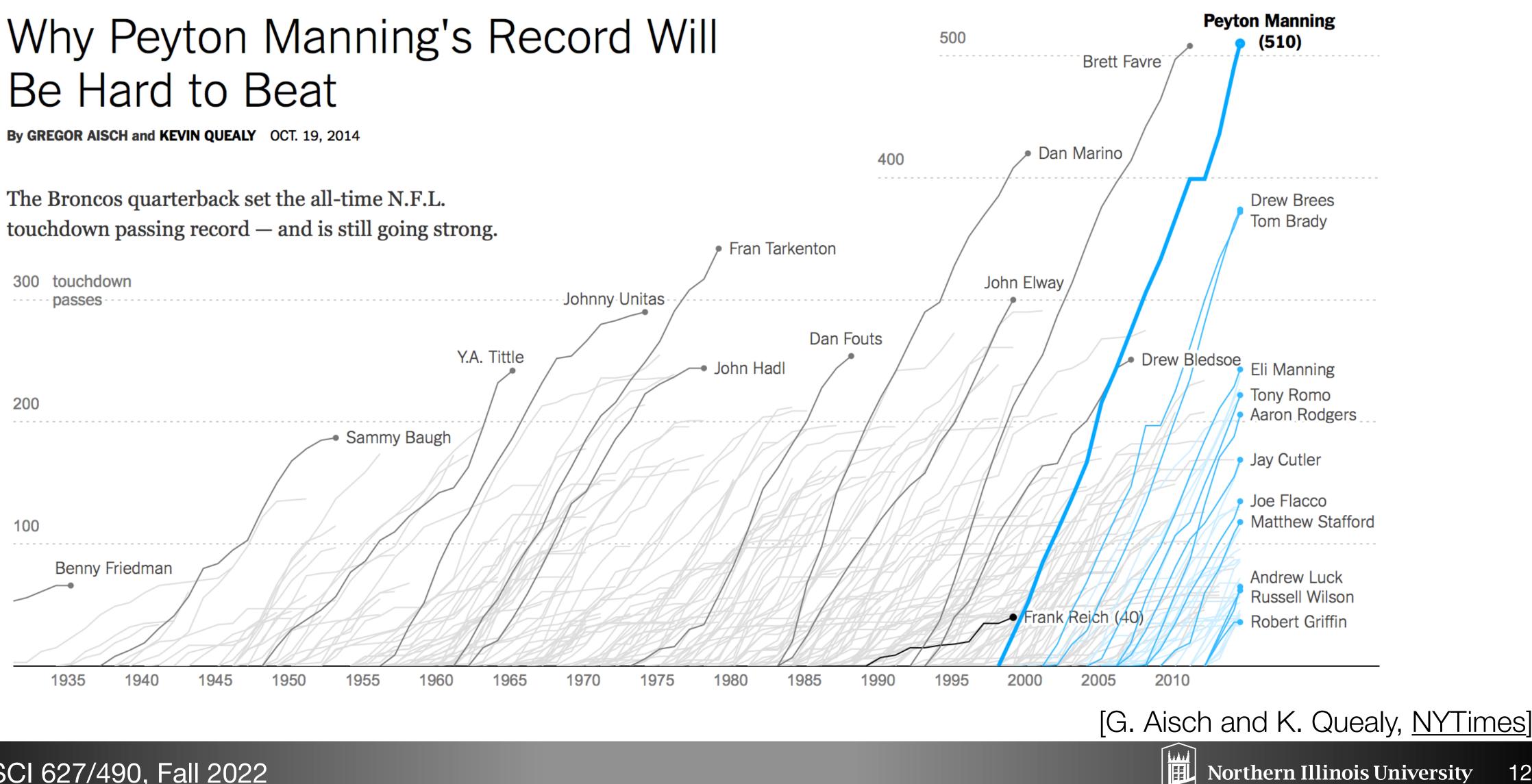
Analysis: Subway Ridership Density







Communication: Peyton's Records



D. Koop, CSCI 627/490, Fall 2022



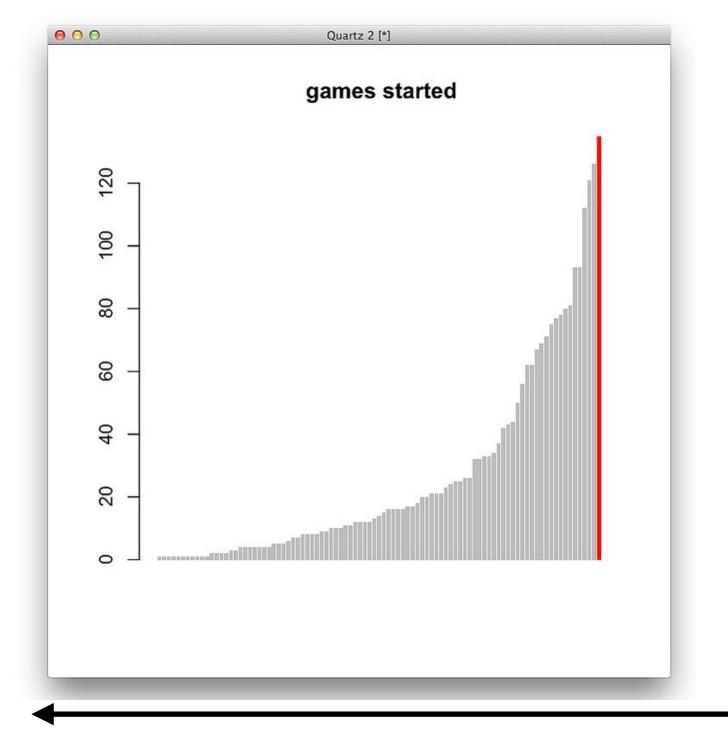




Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team

Each streak ------ shows

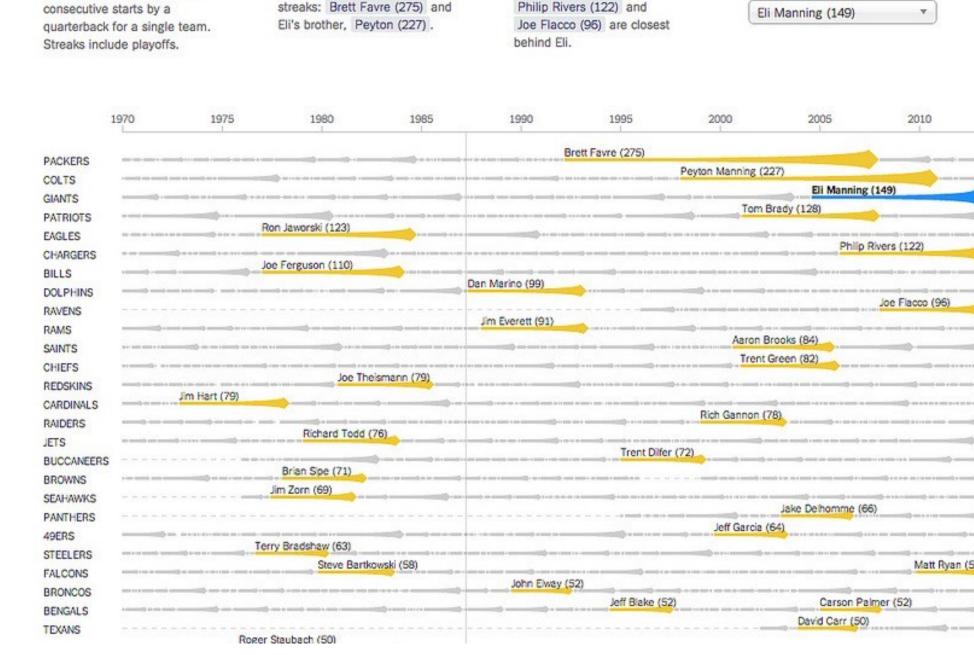


Exploration



Questions

D. Koop, CSCI 627/490, Fall 2022



Only two players have longer

streaks: Brett Favre (275) and

Among active players,

Philip Rivers (122) and

Confirmation

Communication

Answers/Persuasion





Find a quarterback



13



What types of data can we visualize?





Types of Data

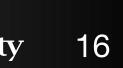
- Tables
- Networks (Graphs)
- Spatial Data
 - Geography
 - Physical (e.g. Scientific, Medical)
- Text
- Sets



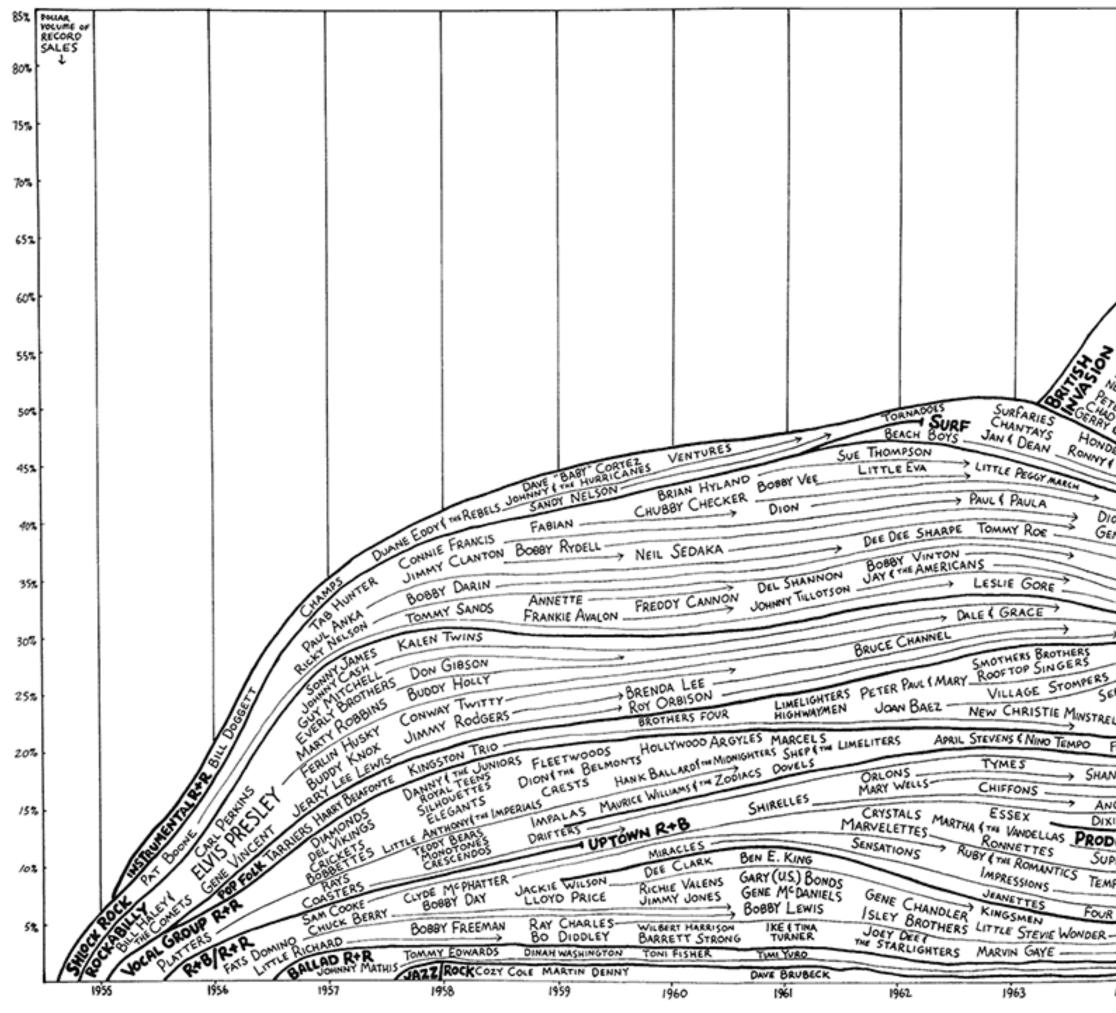


Where have you seen visualizations?

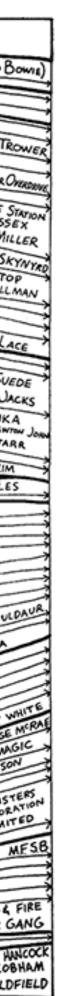




Books/Posters



		-																		
													METAL BI			Aure Coo	EDG	AR WINTER	(David	Barres
														KK SABB	MILER-	ALICE CO.	Dee	P PURPLE -		Comit)
												LIEAWY	METAL D	ACK		ARGENT	Pin	and the second se		
											- RAN	LROAD	EMERSON	N, LAKE, 4	PALMER		F	k FLoyp -	Rogins	
										GRAN	TRAFFI	CATH-			• Y	ES	 URIA 	H HEEP-	ROBINT	ROWER
									BLIND FA		O ANG	and the second s						н Неер → Васни → Васни	MANTURNER	Over
			VERE (THE RA	MARD	ROCK W	HO	CREAM			ODELIN	ANNTA		-> JETHR GEORGE	ARRISON	+					
			VERE STNE RA	NOERS	VAN MOR	RISON	IRON BUTT	ERFLY_	LED ZE	PPEC	MOUTA	NA		WART -	> Ma	LO	J. GE		DAVID ES	SEX
	/	PAUL ME	KINGSME		PROCUL H	ARUM	STEPPEND	VOLF-	GUESS	WHO De		AVER .			LEON	RUSSELL	and the second se	VALSH	LYNYDD	ILLER
	UNG STONE	ICCOYS MBIES	ASSOCIATI			es — E	LUE CHEE	R Cass	C 44	RWATER S	JOE CO		JANIS PLASTIC	DOPLIN -	ALLMAN	BROTHE	Doons	BROTHERS	Z.Z.T	KYNYKI TOP
80	URKS ZO	MBIES	2 (THE MY	STERIAN	S Door		the second se	the second second	C. C.	CAT -	CUNCKIN	NG BLUE	JANIS PLASTIC MOODY	ONO BAN	ND	« ()	STEE	LY DAN	GREGG AL	LMAN
DAVE	LS YAP	OBIROS	STANDE		VANU	BIG BS	OTHER A	\rightarrow	BLOOD, JW		JOHN L	and the second s	M0001	OPACO.		CS GANS	L006/	NS & MESSIN	4	
HONEY	COMBS YAP	DBENDERS	TROGG	S ACID	ROCKOW	HENDRIN		HOLDING	S COMPAN THREE DO	~	-	_	ALLAN .	Javial (GILBERT C	SULLIVAN	BETTE	Midler	PAPER L	ACE
S DEDT	IFS MO	ODY BLUES	HOLLIES	5	VEFFER	SON AIRO		_			NORMAN	GREENE	ELTON					MOLEN		
SEARCHE	SIR DOUG	LAS QUINTE	T VOGUE S DUSTY SP CYRCLI S LEFT BA	SPOPRO	KBOX TO	PS	ANE	\rightarrow	SPIRAL ST ETER, AUGN	URCASE	Nº OR	LANDON	OSMO	NDS		_		\rightarrow	BLUE S	VEDE
PETE BEAT	FREDOY	F DREAMER	CYRCL	RINGFIELI E	BEE GE				SPIRAL ST	AARY	ANNE N	NURRA	0.48	NELSON	CHUCK	C BERRY PRESLEY	MAUREEN	MGOVERN	PAUL AN	the second se
RY S JERE	- HERMAN	I'S UCOMER	LEFT BA	WK	HAPPENII BUCKING	VGS	Cui	TT D	TER, PAULST		CARPE	OITION	HELE	N REDOY -	ELVIS			\rightarrow	CLIVIA NEI RINGO ST	KA WTOW JOSE
NDEL PACE	SEE	KERS	THEIC OW	-0110			CLASSICS GRASS RO	OTS -		310 1	1	AV DEB		ER-		NGS	DONNY	OSMOND		
TI THE DAYTA	AKERS PET	ULA CLARK	NEIL DIAN	INATRA	CowsiL	LS->	UNION GA	P	LACKIE DE	es	B.J. B.J. B. B.J. B.	THUNK	RICK HELE	×		PBERRIES	STOR	66	ANDY K	
					BOBBY C	DENT.	LEMON PIL	PERS	JACKIE UNT	APBELL E	PAUL MC	AMILY	PARTRID	E FAMIL	Y RAS	SUTHRIE	NITTY GR	TTY DIAT BANK	ENGL	
DIONNELL		Lorent T.	B		SCOTT MC	Ken	O.C. SM	TH	JOHNNY GLEN	<u></u>	800	AVR 16	PARI	ONEZ.	ARLO	VauNG -				
GENE PITNE	WICK HERB	ALPERT & TU	BOB LIND	2	FIFTH D	SAN ZIE	MARY HO	PKIN	JOHN		POPPY F	THOUSE	JOAN D	on.	NEIL	YOUNG -)			
		Curry To	Same James ;	THE SHO	LULU	MENSION		-	OLIVE	63	DISONUES IN	and -	116	HTFOOT -	MAC	BROWN	6	\rightarrow		
	BBLEGV	CHER	BOB LING BOB LING MONKEES PLAYBOYS	STR	Love		ossy Gold	SBORO	Bomy Sh BROOKLY BROOKLY	ASH +	DI FU	LOF	JOAN JORDON LIG CAT STE CAROLE	VENS	JACKSO					
	GARY	EWIS & THE	PLAYBOYS -		WBERRY A	ARM	1910 FRUIT	SUM CO.	BORNY	880	GRATE	/	CAT STE	KING	PAUL SI	MON		SI CROFTS	HARIA M	ULDAUR
\rightarrow	WE FU	VE		\rightarrow			OHIO EN	PRESS	BROOS	N-	- NN	16		ENNER -	JIM CE	000	ART GAS	CROFTS	Abb	
		N N	JEW VAUDVII	LE Du		Вов	DYLAN	NYR0	BOBDY	ON DOS	BREAD	\rightarrow	JOHN D CARLY KRIS KRI	SIMON_	ON DON M	200	OS SEAL	<u>ste</u>	Au	
SERENDIPITY	FOLBOB	SPOONFUL -	BARRY SA	DLER	UANIS	JAN >	COUNT	BOC	K NILSS	BLOODS	BADFIN	GER TAVIOR	KRIS KRI JONI MIT	STOPPEN -	JONATH	N EDWAR		ROTHERS	\sim	_
SERENGERS	MAR	SPOONFOC UNNE FAITHFU MCGUIRE	DONOVAN SIMON & G			SPRINGE	SOF	COLL	INS	NA	SH JAMES	YOUNG	JONI MI	icities y	O'JAY	rs	ISLEY DOBIE	GRAY	\sim	WHITE
RELS-	→ BARR BYRT	1100	MAMAS (PAP	AS	¢				CROSBY, ST	nus, tree	SPINNE	ERS-	A1 6955		CHI LIT	65		\sim	BARRY	E MCRA
FOUR SEA							INTRUDE	T SOUL	L JERRY	SUTLER	BROOK B	ENTON	AL GREE		5171.151		BILLY	AUL	GEORG BLUE N	AGIC
ANGRI-LAS			<u>></u>	\rightarrow	- T. HAY	TERRELL	INTRUD	ERS	MARVIN	GATE	STAIRS	TEPS	DONNY H	ATHAWAY					AL WILL	
		OUS BROTH	ERS	RVIN GAY	EGIAMONI		DELPHON	SOUL	60.00	OS OF	MOMEN EDWIN	STARR	ROBERTA	FLACK		INGRAM		KENDRICK	INTER SI	STERS
ANGELS DIXIE CUPS	RIGHTE	OUS DROT	s						FOUN	DATIONS	JACKSON	FINE -	JEAN K	NIGHT	LUTHER	TTY WRIG	HT E00	HV HV	INTER CORPORES UNLIN	ITED
DUCEDS	BARR	ARA LEWI	\sim			$\rightarrow SL$	Y THE FAM				FREDA PA	DSS G	LA DYS KNI	GHTS	PIPS	1468	KON		NE	MESE
UPREMES -		~		-	ETHA FR	RANKLIN-	PRO	GRECO	IVE SOU			C	HONEY CO	ONIE	14.00	JAEL UNT		R OF POWER	2	
EMPTATIONS		\square	CARLA THOM	MAS A	ARTHUR C	ONNELLY	ARTHUR BA	Rower	IVE Sour				DRAMAT	ICS A	MAIN ING	REDIENT	TOWE			
WR TOPS			CARLA DERCY SLE	DGE	JOE TEX SAMS DA	AVE (DTIS RED	DING OR.	I SAAC H	AVE S	TEWAR		0.000		BILLY P	DECTON '		WEIELD E	ARTH, WIND	GANG
8			MES & BOSE	BY PURIFY		TI CU	ARENCE CAR	ER, ISL	EY BROTHL	RS	TEVIE WO	NDER -			-		CURIT			
	YOUNG	RASCALS	CAPITOL		OUL SURVIV	and the second se	IN MASE KI	ELA JR	WALKER (THE ALLSTA	RS					CINICER	S DR J	OHN	HERBIE BILLY C	OBHAM
-		LEWIS	MITCH RYDE	RE	ESQUIRE	s	ARCHIE BE	ill (WATTS 10340 RAYTHM B	AND IK	E (TINA TI	URNER			STAPL	E SINGER	CARLOS	SANTANA MELAUGHKI	N MIKE OL	DFIELD
1964	196		1966		1967		1968		1969		1970		1971		197	2	197	3	1974	ł
			ΓĽ	$2 \wedge c$	וי אר	NI' F		ie	Hor	$r \Delta t r$	ר D	<u>a</u> \/	R	G_{2}	rnf		1 C)77	hio	Т
			Lı	IUL	רז√ I			5				ay,	Ι \.	Ja		aiO,	ΙČ	, , ,	lvia	
														Jont	hor	n Ill	noid	s Univ	TORO	ity.
														NOLT.	ner			5 UIII	vers	ILY_

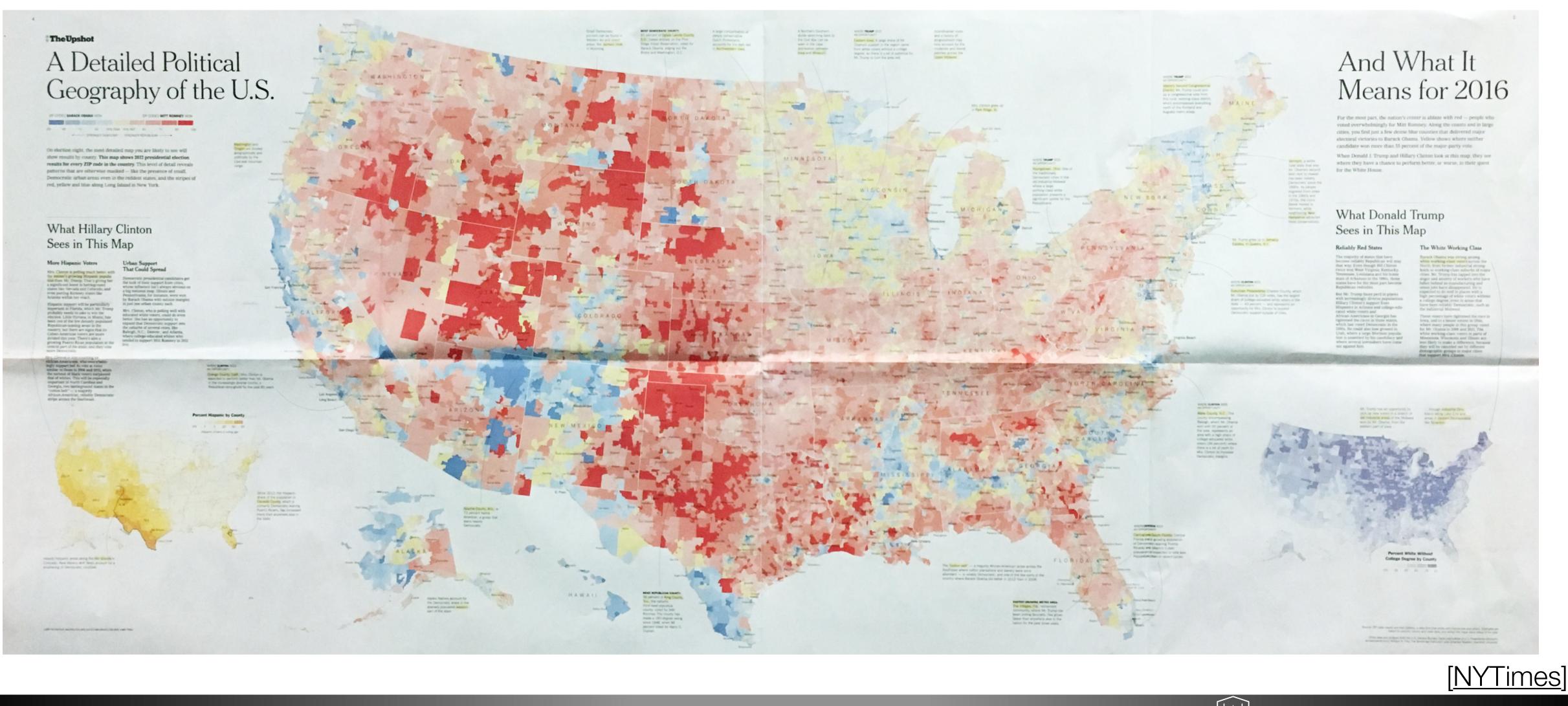








Newspapers

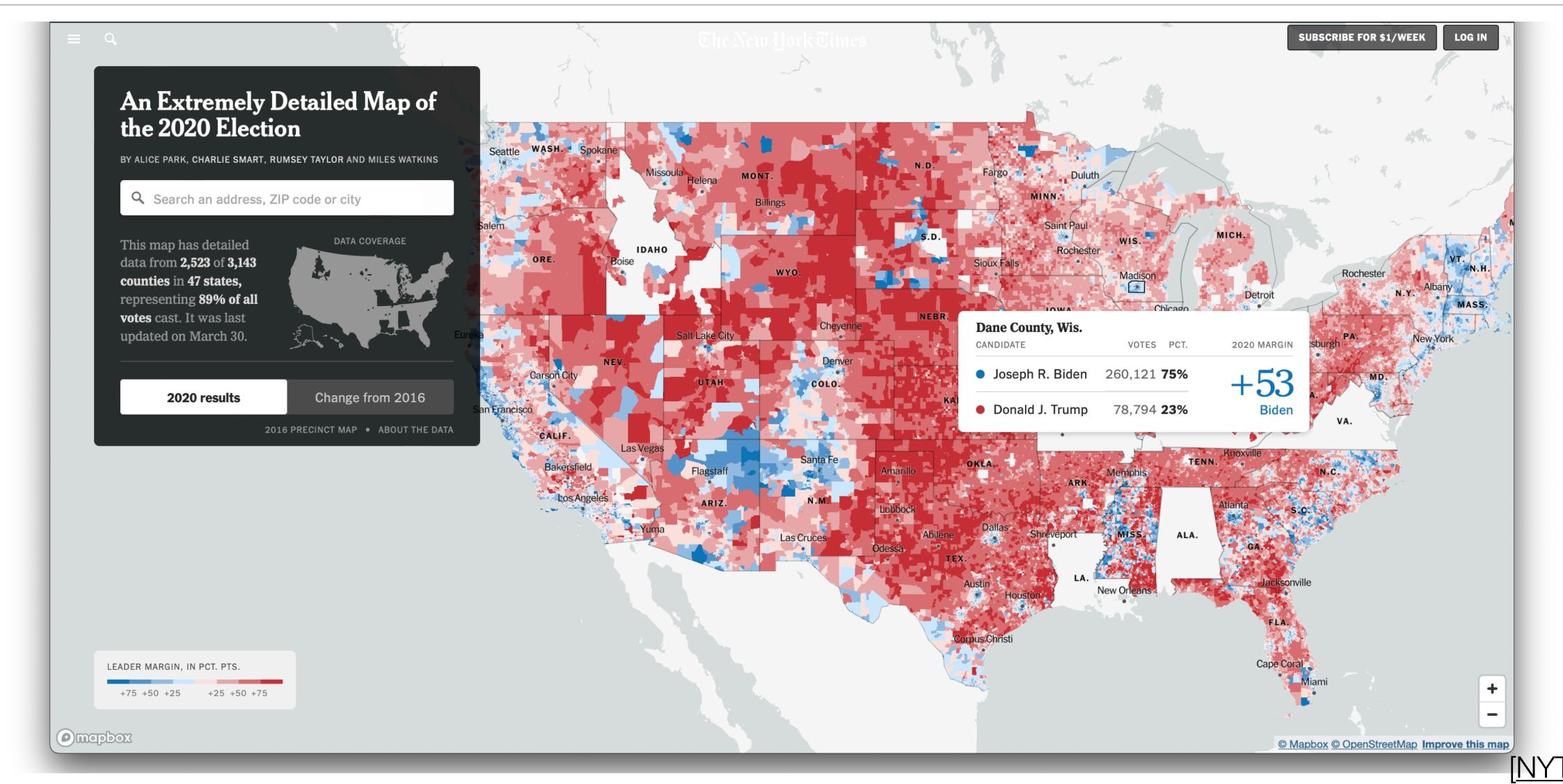








Web



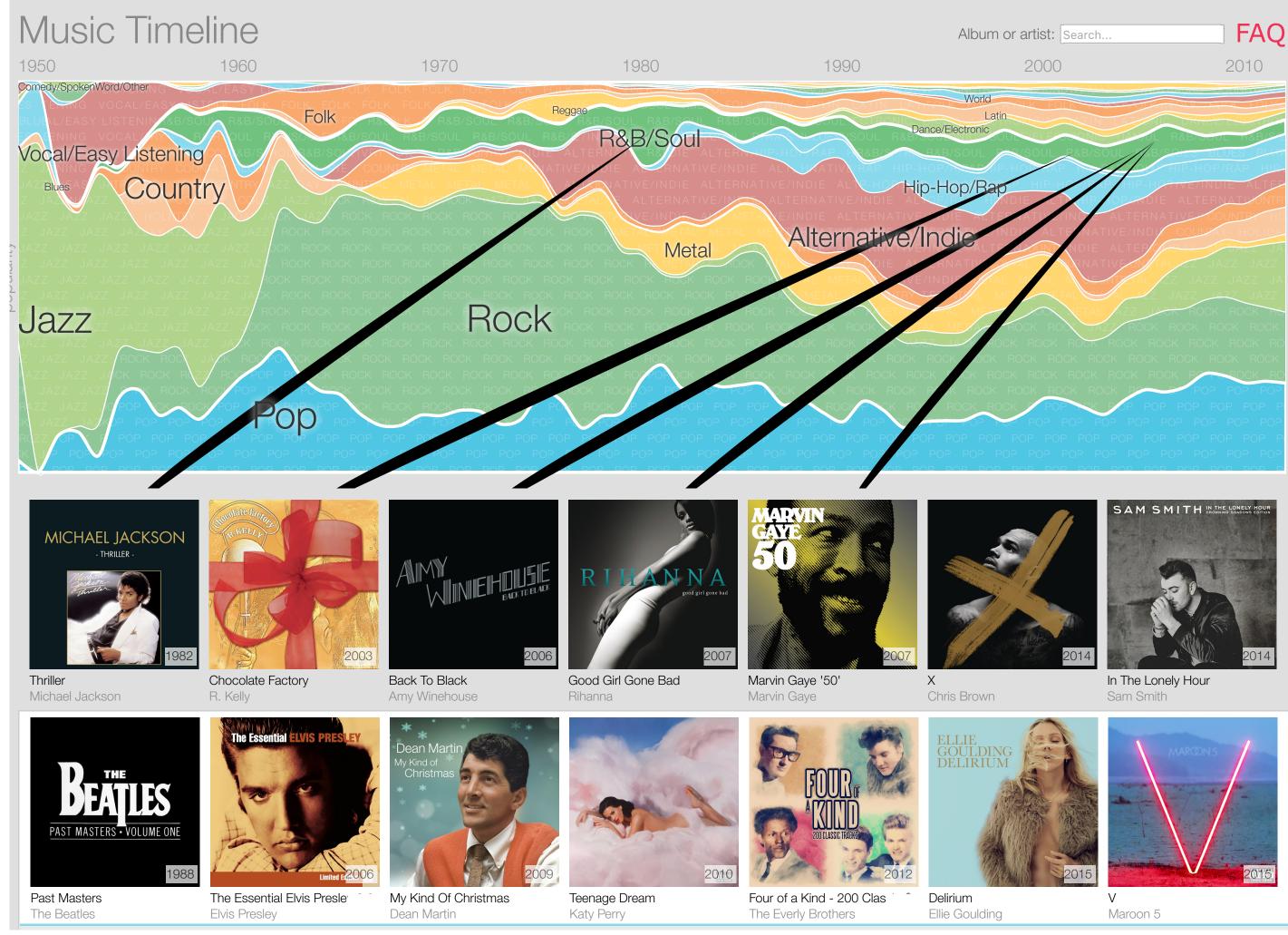








Web



D. Koop, CSCI 627/490, Fall 2022

[Music Timeline, Google Research (No Longer Working)]





What is the advantage of the second version?







Interaction







How do we create modern visualizations?







Tools

- ...

- Desktop Applications:
 - Excel (see <u>excelcharts.com</u>)
 - Tableau
- Grammars:
 - Vega-Lite
- Programming Frameworks
 - d3.js

- Observable Plot, plot.ly, deck.gl

D. Koop, CSCI 627/490, Fall 2022

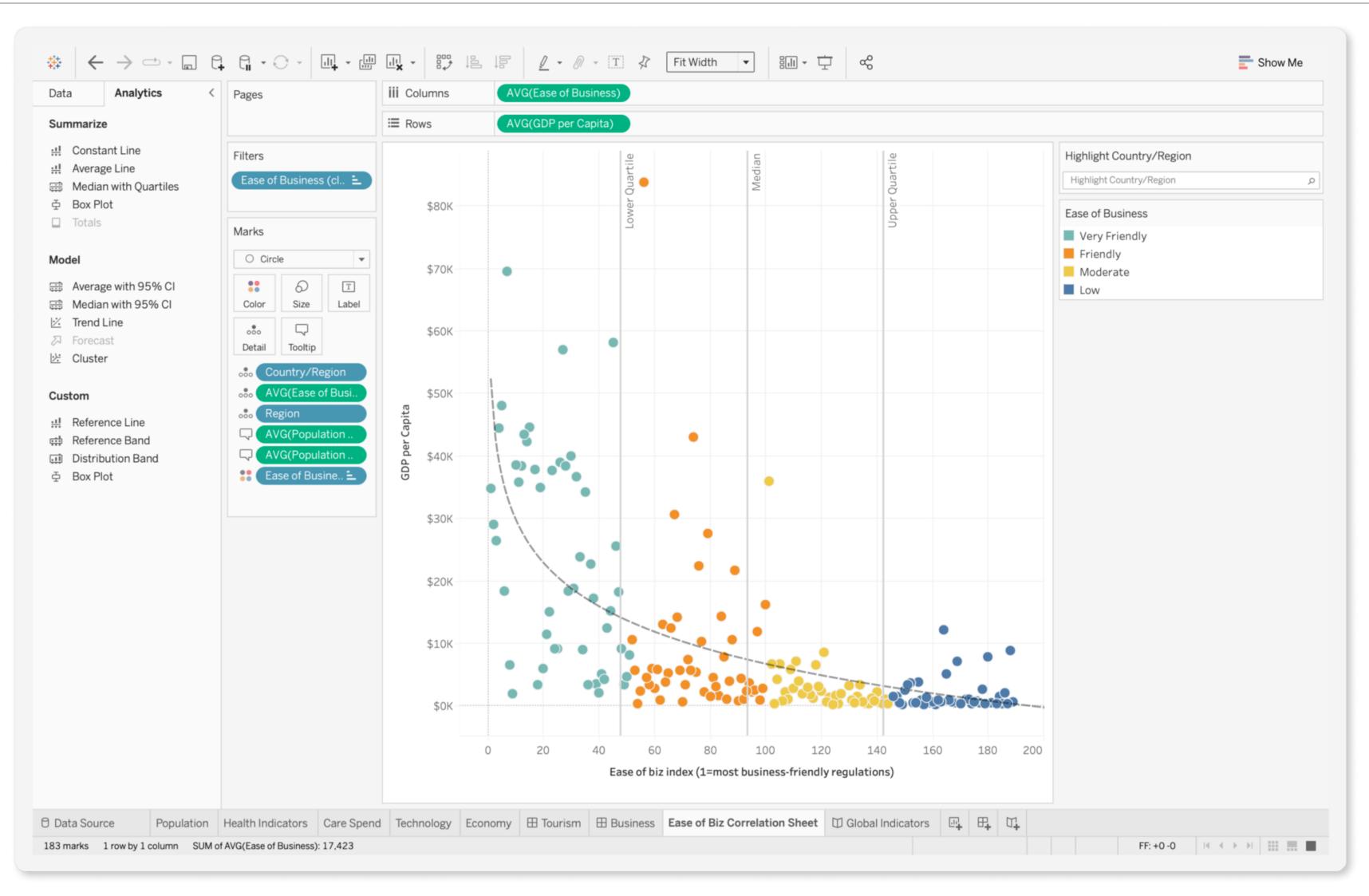
- Tradeoffs
 - Speed
 - Customization
 - Understanding
 - Dissemination





24

Tableau







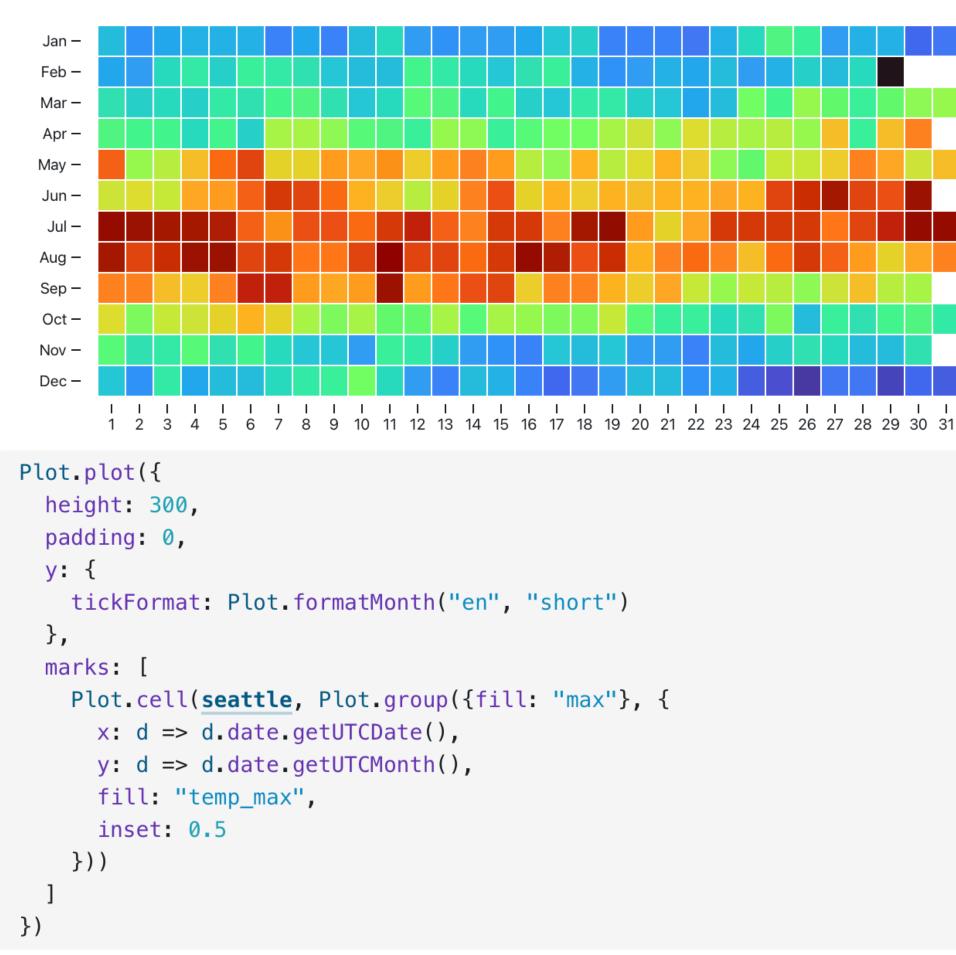






Observable Plot

seattle = > Array(1461) [Object, Object, Objec seattle = FileAttachment("seattle-weather.csv").csv({typed: true})













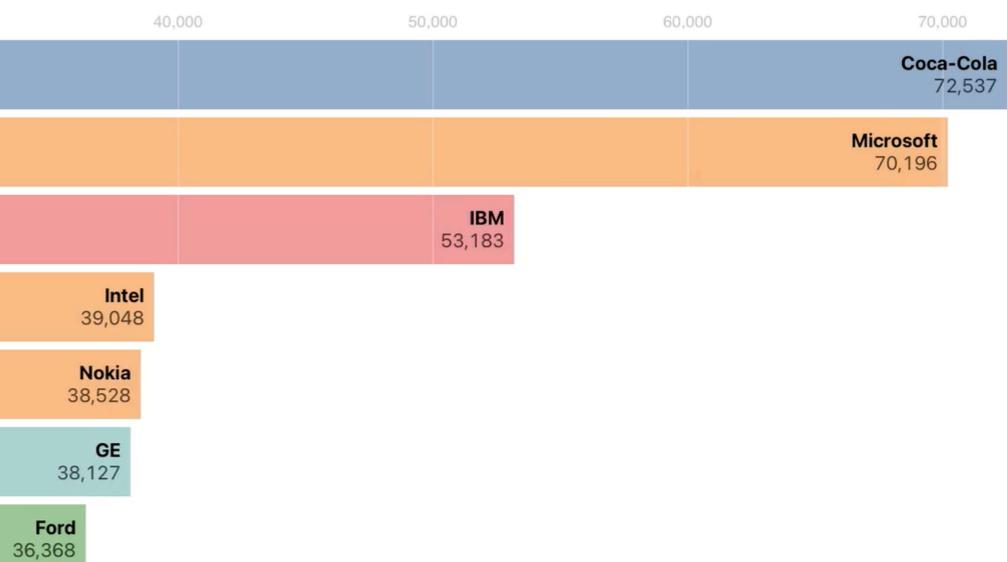


<u>d3.js</u>

Best Global Brands

Value in \$M; color indicates sector. Data: Interbrand

10	,000 20,0	00	30,000
			Disney 33,553
		McDona 27,8	Id's 359
		AT&T 25,548	
	Marlı 22,	boro ,110	
	Mercedes-Ben 21,10	z 4	













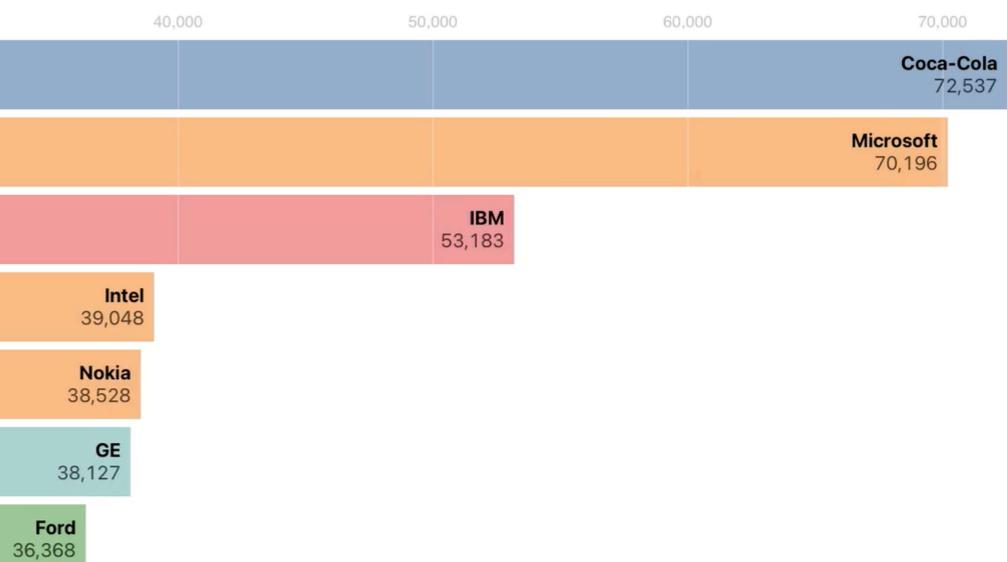


<u>d3.js</u>

Best Global Brands

Value in \$M; color indicates sector. Data: Interbrand

10	,000 20,0	00	30,000
			Disney 33,553
		McDona 27,8	Id's 359
		AT&T 25,548	
	Marlı 22,	boro ,110	
	Mercedes-Ben 21,10	z 4	















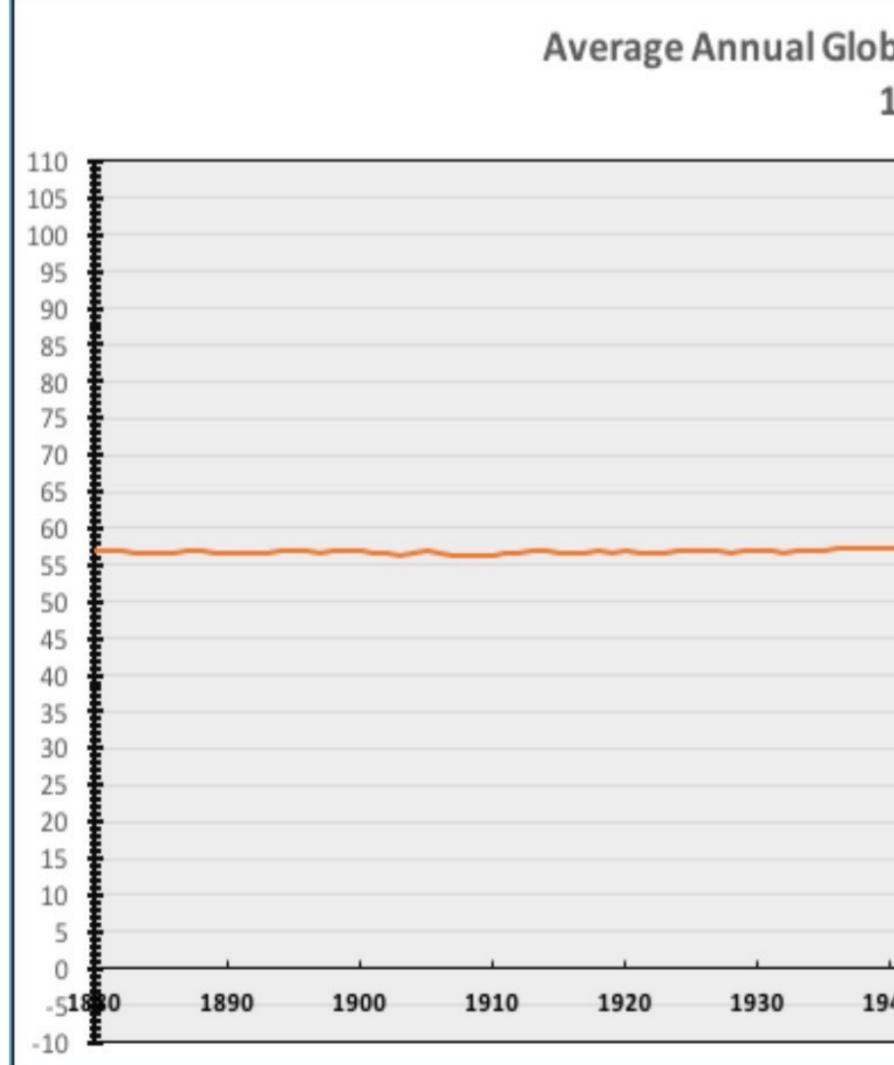
Why do we care about the design of visualizations?







Design: Focus on only the y-axis



D. Koop, CSCI 627/490, Fall 2022

Average Annual Global Temperature in Fahrenheit 1880-2015

								, I		
								1		
								1		
								1		
				~~~~						
								1		
								1		
3	1									
40	1950	1960	1970	1980	1990	2000	2010			
							[	5	laywa	

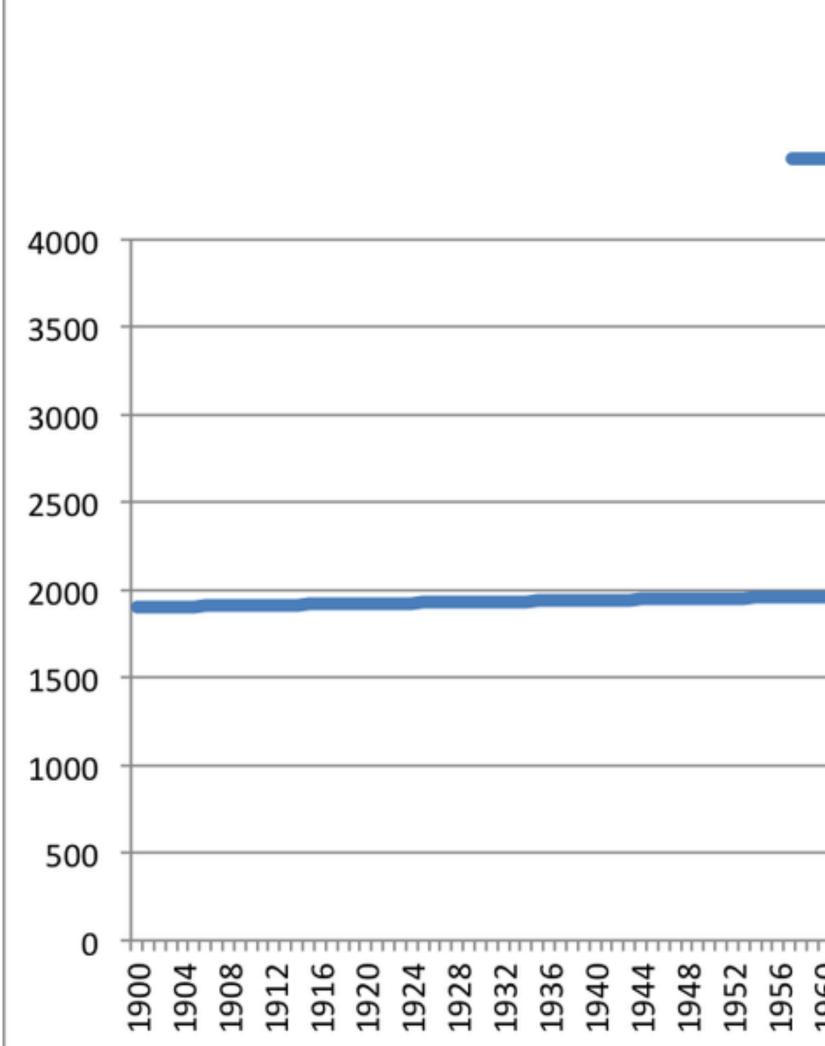








### Design: Year on the y-axis



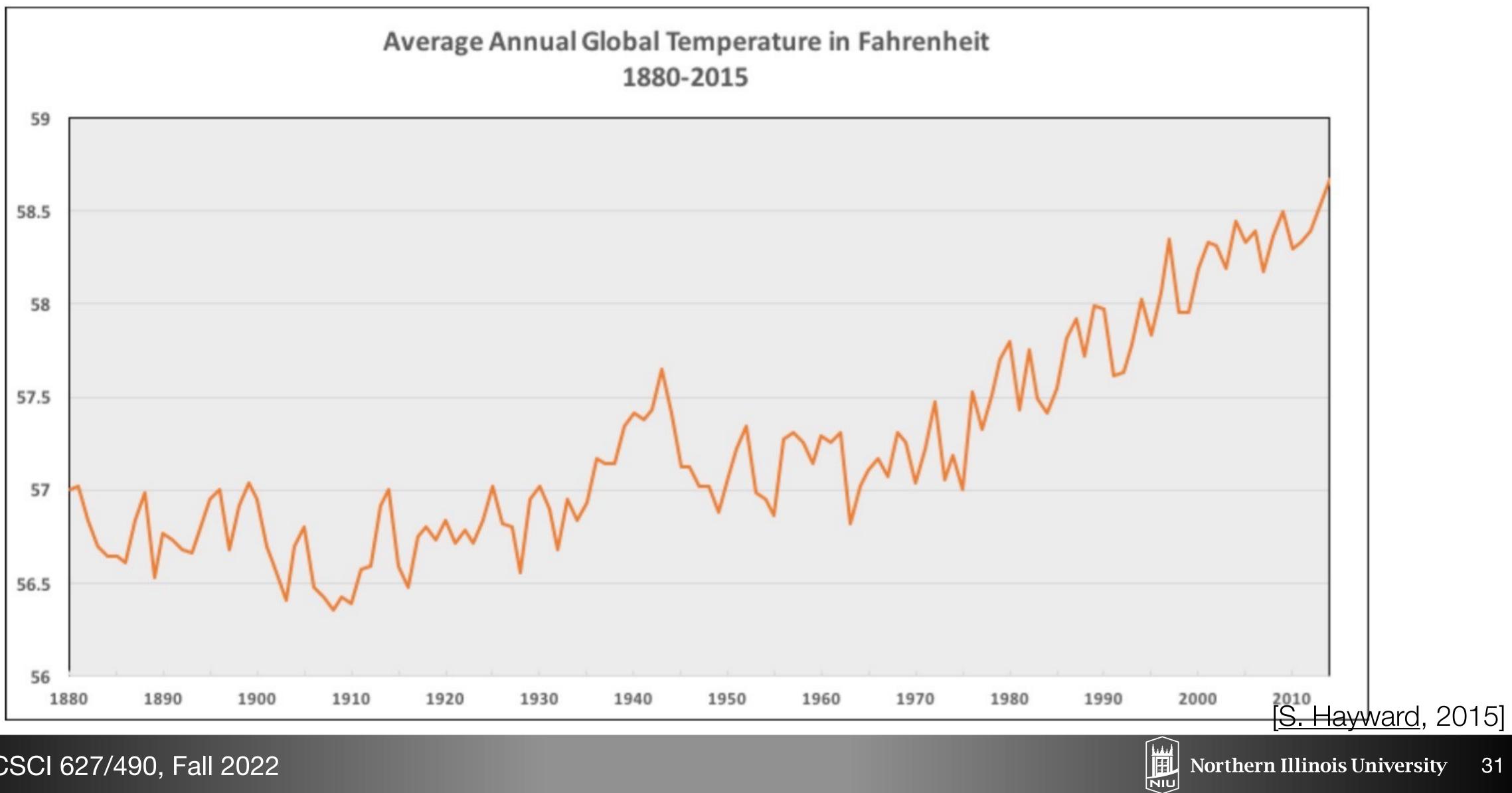
Year	
Year (A.D.)	
$\frown$ $\rightarrow$ $\frown$	
1960 1964 1968 1976 1976 1976 1976 1976 1976 1976 1976	ଁ କୁଁ <u>[@bizweekgraph</u>
	Iorthern Illinois University







### Design: Different y-axis



D. Koop, CSCI 627/490, Fall 2022

Northern Illinois University

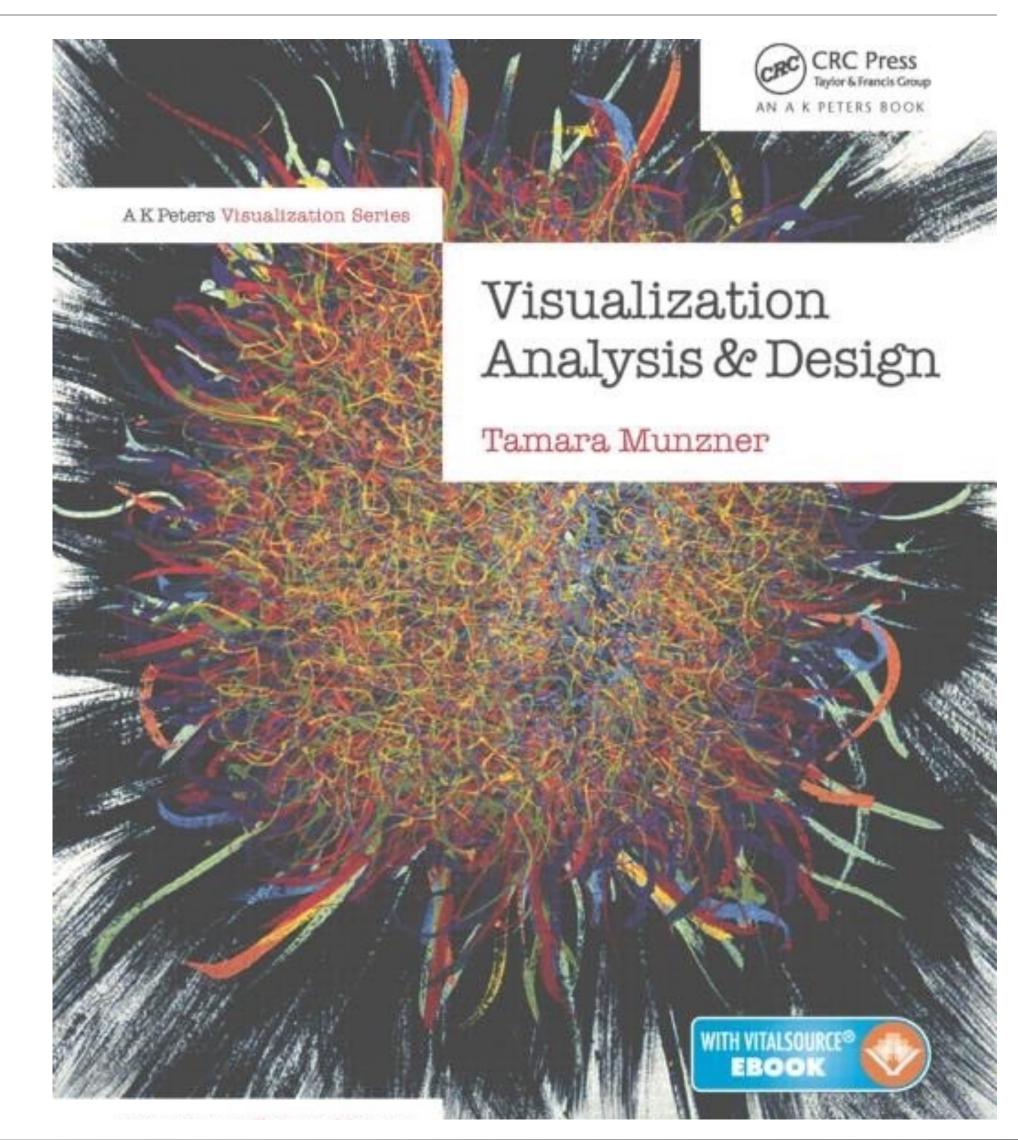






## Administrivia

- Course Web Site
- Syllabus
  - Plagiarism
  - Accommodations
- Textbook:
  - Required: <u>Munzner</u> (VAD)
- Assignments
- Exams: Midterm (Oct. 13) and Final (Dec. 8)
- Registration







## Administrivia

- Undergraduate (CSCI 490) and Graduate (CSCI 627)
  - Graduate: Extra reading, exam questions, project emphasis
- Research Topics:
  - Also investigate some topics in depth
  - Research papers as assigned reading (CSCI 627)
- Project: Create an interactive visualization (or vis research)
  - Design
  - Data analysis
  - Insight
  - Presentations: Last week of class









### Office Hours & Email

- Office hours will be held in person
- Scheduled office hours are open to all students - Tu: 1:45-3pm, Th: 10:45am-12pm, or by appointment
- You do not need an appointment to stop in during scheduled office hours
- If you need an appointment outside of those times, please email me with **details** about what you wish to discuss
- Many questions can be answered via email. Please consider writing an email before scheduling a meeting.







# Do not cheat!







## Do not plagiarize

- It is **Academic Misconduct**
- Do your own work, do not copy anyone else's work, text, sentences, ...
  - Anyone = another student, an internet source, book, blog, ...
- Never quote text unless there is a specific need.
  - Usually, only famous quotes or very specific definitions
  - "I think there is a world market for maybe five computers." -Thomas Watson (1874-1956), Chairman of IBM, 1943)
- Cite sources that back up your claims or reflect the origin of an idea
  - Vertex cover is an NP-Complete problem [1]. ... [1] Garey, M. R., and Johnson, D. S., "Computers and intractability: a guide to NP-completeness." 1979.







### Do not cheat

- Cheating on assignments, projects, and exams is not allowed
- You will receive a **zero** on the assignment/project/exam
- It will be reported to the department and university
- If it repeats, you will fail the course
- You can be kicked out of the university





# Do ask questions!







### Do ask questions

- If you are stuck on a specific issue with an assignment:
  - Do email me with **specific** questions
  - Do consult books, online documentation, tutorials
  - Do discuss that specific issue with a classmate
- If you are asked about a question:
  - Do not share your code

  - If the questioner is trying to cheat, walk away - If you see an obvious mistake, kindly point it out - Suggest a specific function or library that may be useful







#### Questions?



