

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

— R. W. Hamming

The purpose of **visualization** is about **insight**, not pictures

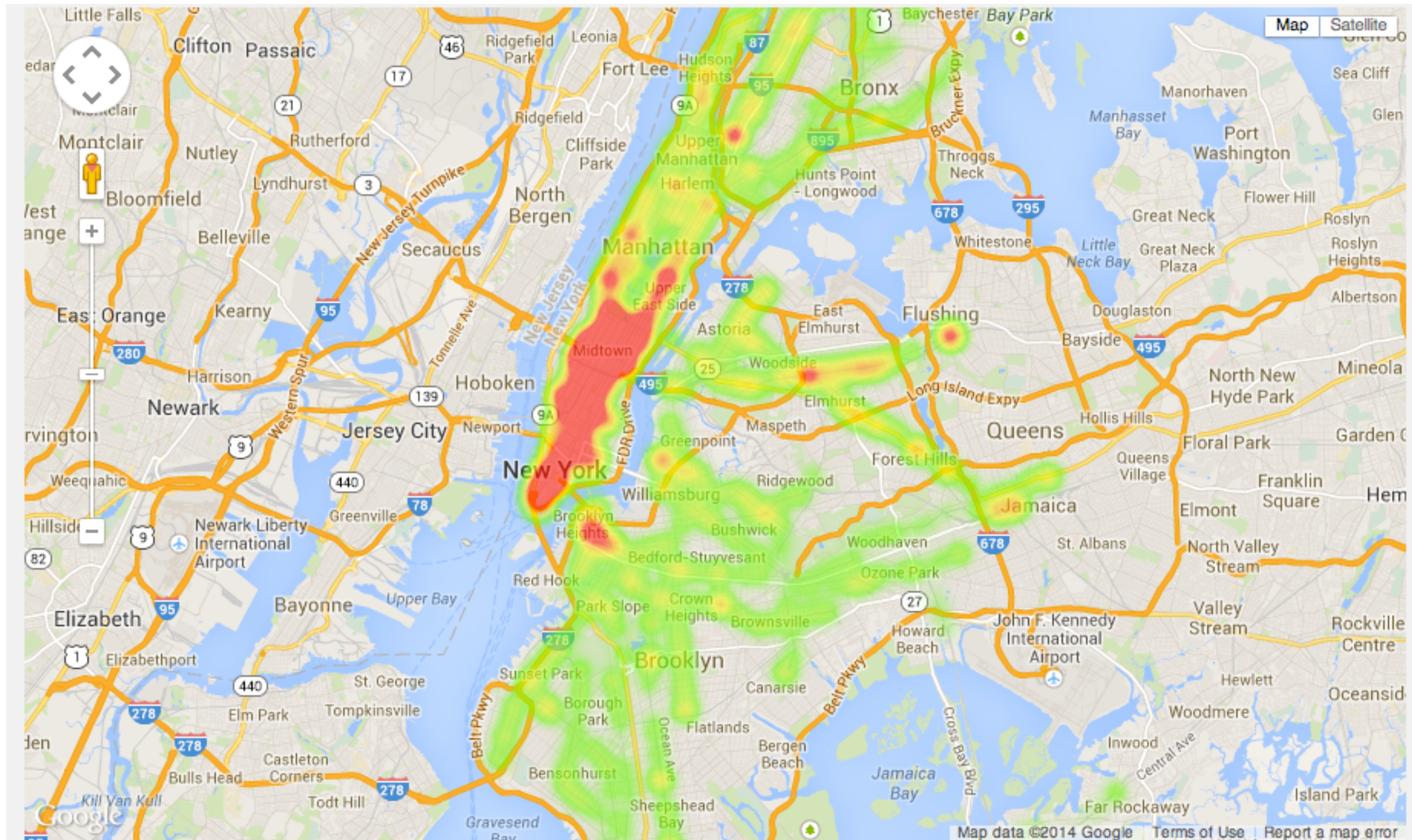
— B. Shneiderman

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

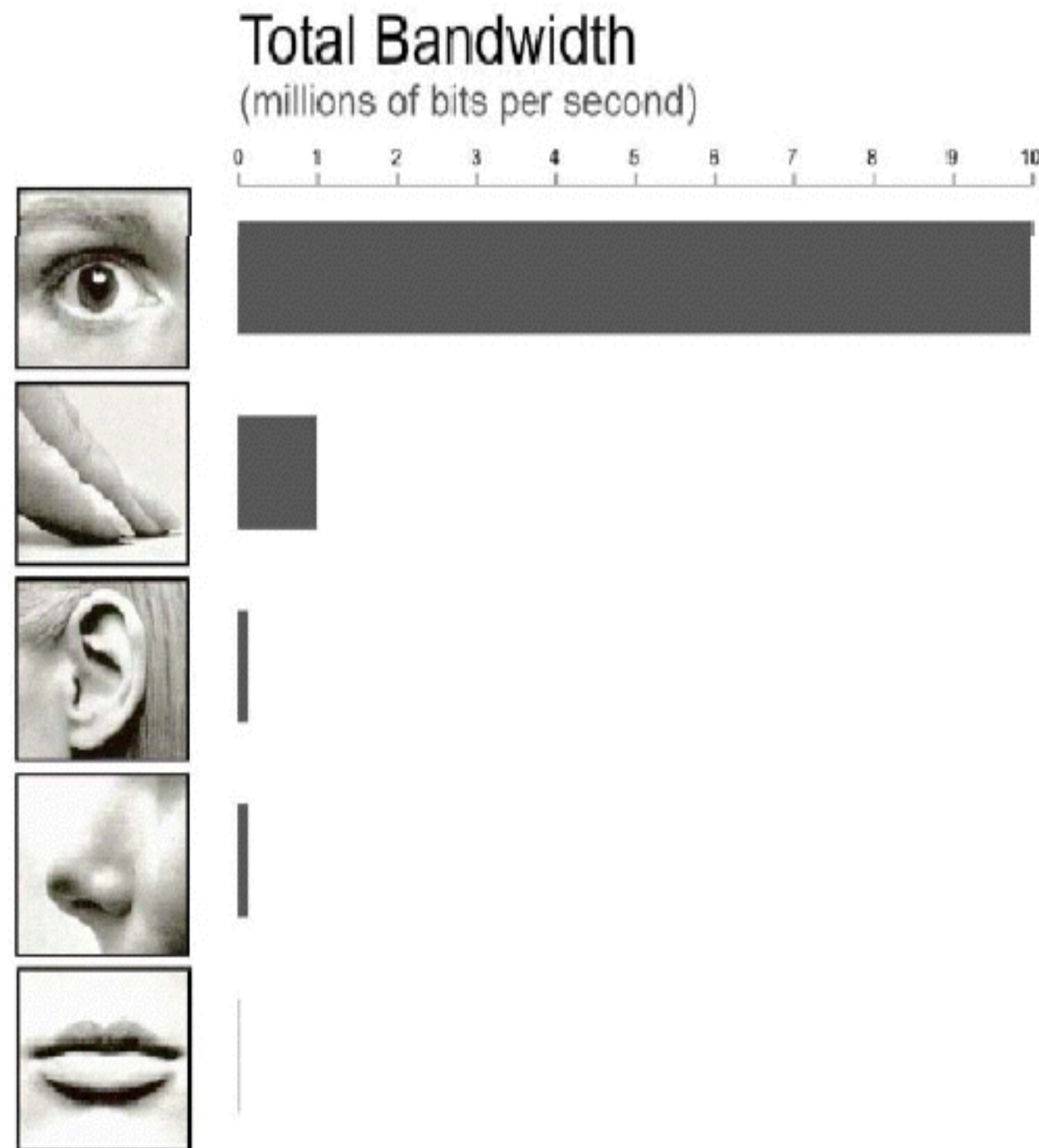
MTA Fare Data Table

	REMOTE	STATION	FF ▼	SEN/DIS	7-D AFAS UNL	D AFAS/RMF I	JOINT RR TKT	7-D UNL	30-D UNL
1	R011	42ND STREET & 8TH AVENUE	00228985	00008471	00000441	00001455	00000134	00033341	00071255
2	R170	14TH STREET-UNION SQUARE	00224603	00011051	00000827	00003026	00000660	00089367	00199841
3	R046	42ND STREET & GRAND CENTRAL	00207758	00007908	00000323	00001183	00003001	00040759	00096613
4	R012	34TH STREET & 8TH AVENUE	00188311	00006490	00000498	00001279	00003622	00035527	00067483
5	R293	34TH STREET - PENN STATION	00168768	00006155	00000523	00001065	00005031	00030645	00054376
6	R033	42ND STREET/TIMES SQUARE	00159382	00005945	00000378	00001205	00000690	00058931	00078644
7	R022	34TH STREET & 6TH AVENUE	00156008	00006276	00000487	00001543	00000712	00058910	00110466
8	R084	59TH STREET/COLUMBUS CIRCLE	00155262	00009484	00000589	00002071	00000542	00053397	00113966
9	R020	47-50 STREETS/ROCKEFELLER	00143500	00006402	00000384	00001159	00000723	00037978	00090745
10	R179	86TH STREET-LEXINGTON AVE	00142169	00010367	00000470	00001839	00000271	00050328	00125250
11	R023	34TH STREET & 6TH AVENUE	00134052	00005005	00000348	00001112	00000649	00031531	00075040
12	R029	PARK PLACE	00121614	00004311	00000287	00000931	00000792	00025404	00065362
13	R047	42ND STREET & GRAND CENTRAL	00100742	00004273	00000185	00000704	00001241	00022808	00068216
14	R031	34TH STREET & 7TH AVENUE	00095076	00003990	00000232	00000727	00001459	00024284	00038671
15	R017	LEXINGTON AVENUE	00094655	00004688	00000190	00000833	00000754	00020018	00055066
16	R175	8TH AVENUE-14TH STREET	00094313	00003907	00000286	00001144	00000256	00038272	00074661
17	R057	BARCLAYS CENTER	00093804	00004204	00000454	00001386	00001491	00039113	00068119
18	R138	WEST 4TH ST-WASHINGTON SO	00093562	00004677	00000251	00000965	00000127	00031628	00074458

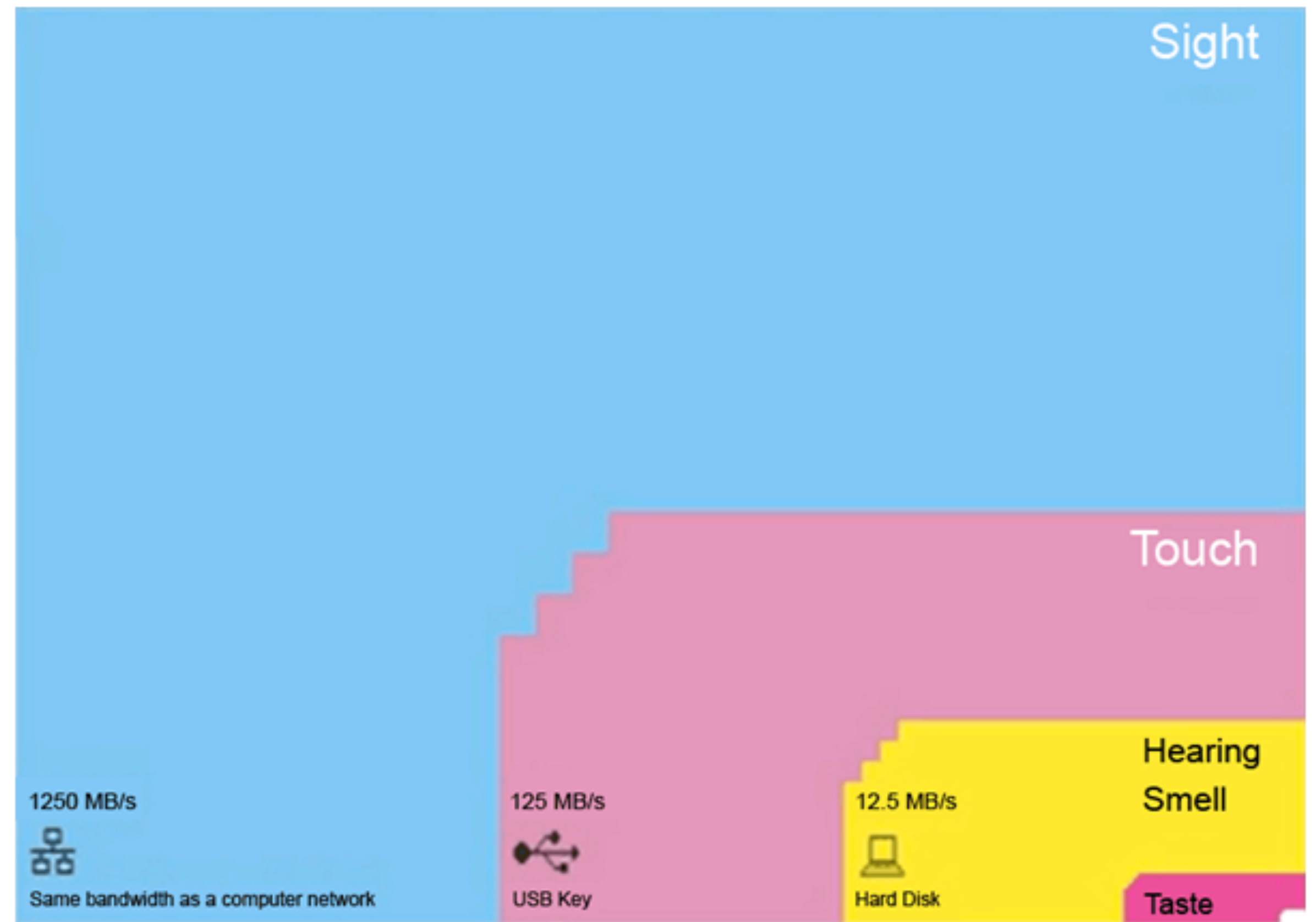
MTA Fare Data Visualization



Why do we visualize data?



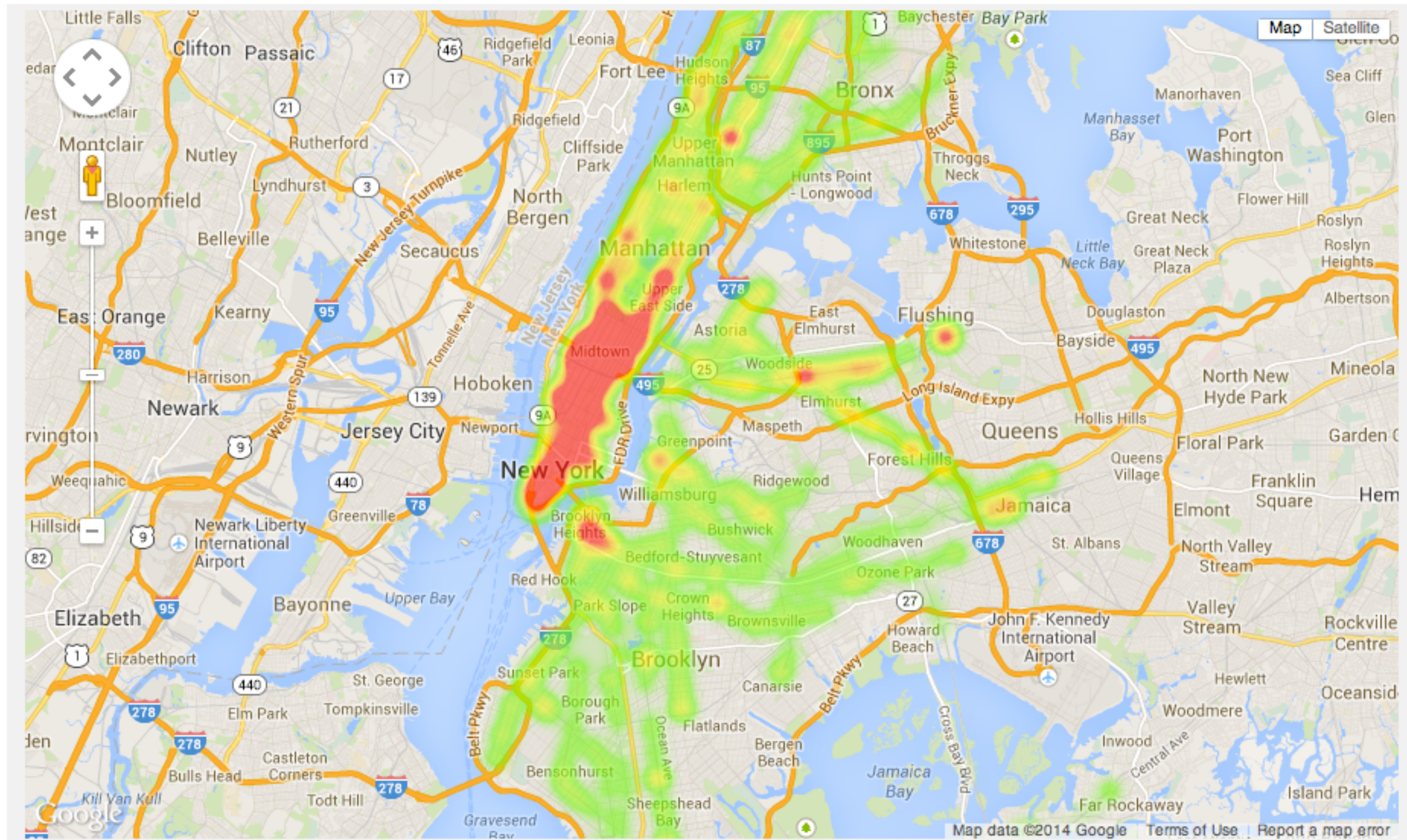
[via A. Lex]



[T. Nørretranders]

What are the goals of visualization?

Analysis: Subway Ridership Density

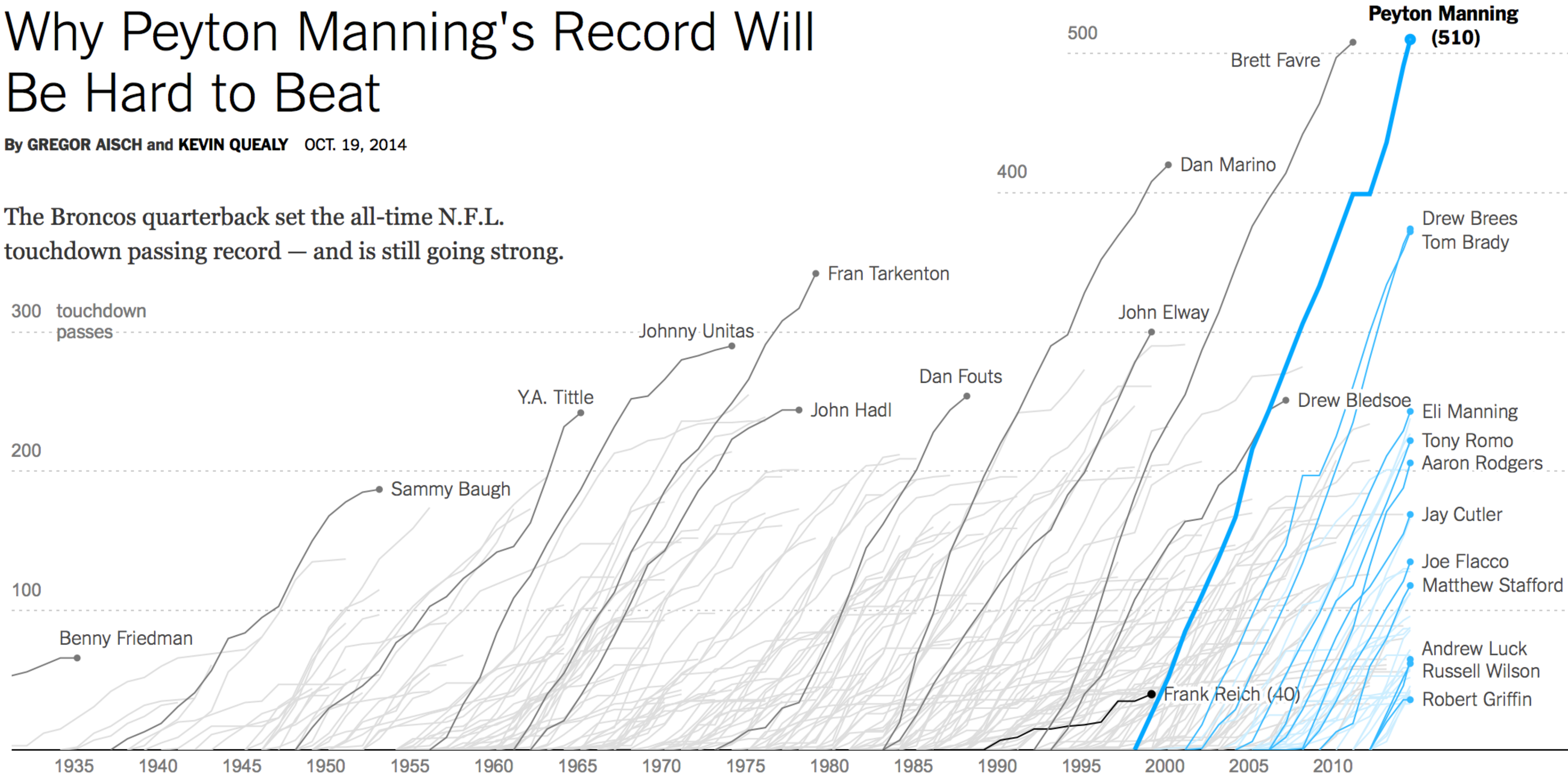


Communication: Peyton's Records

Why Peyton Manning's Record Will Be Hard to Beat

By GREGOR AISCH and KEVIN QUEALY OCT. 19, 2014

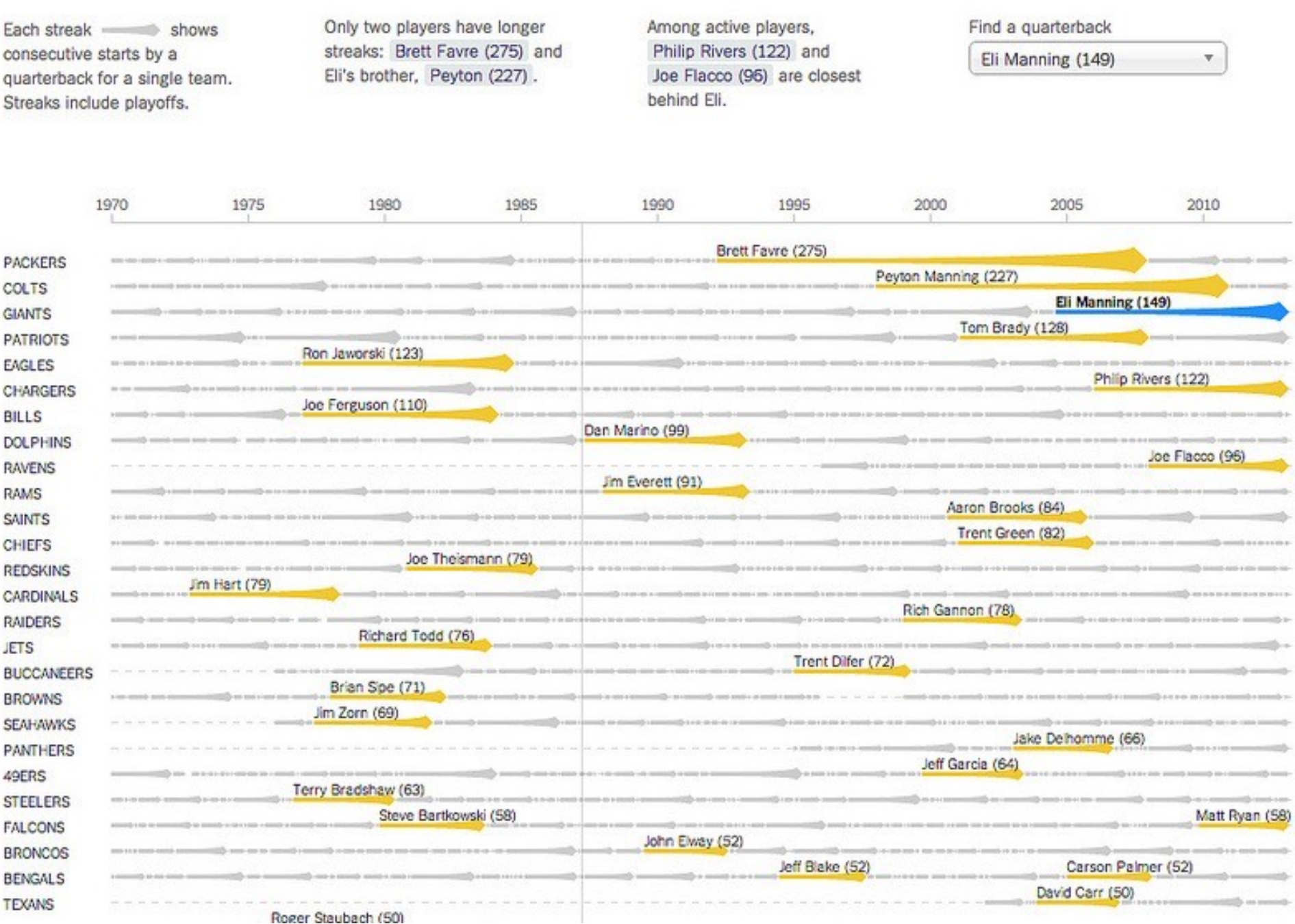
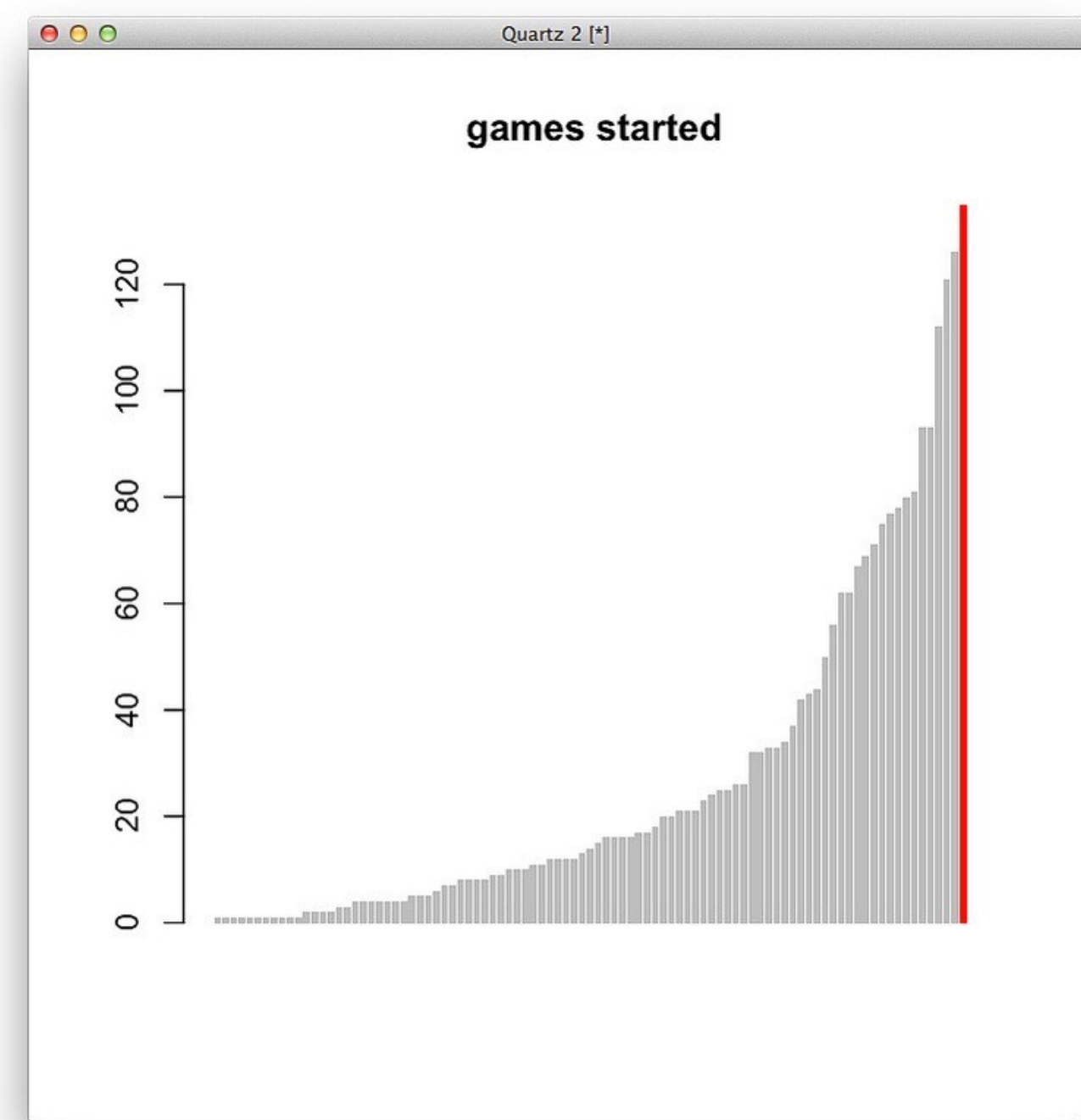
The Broncos quarterback set the all-time N.F.L. touchdown passing record — and is still going strong.



[G. Aisch and K. Quealy, [NYTimes](#)]

Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team



Exploration

Confirmation

Communication

Questions

Answers/Persuasion

[K. Quealy, 2013]

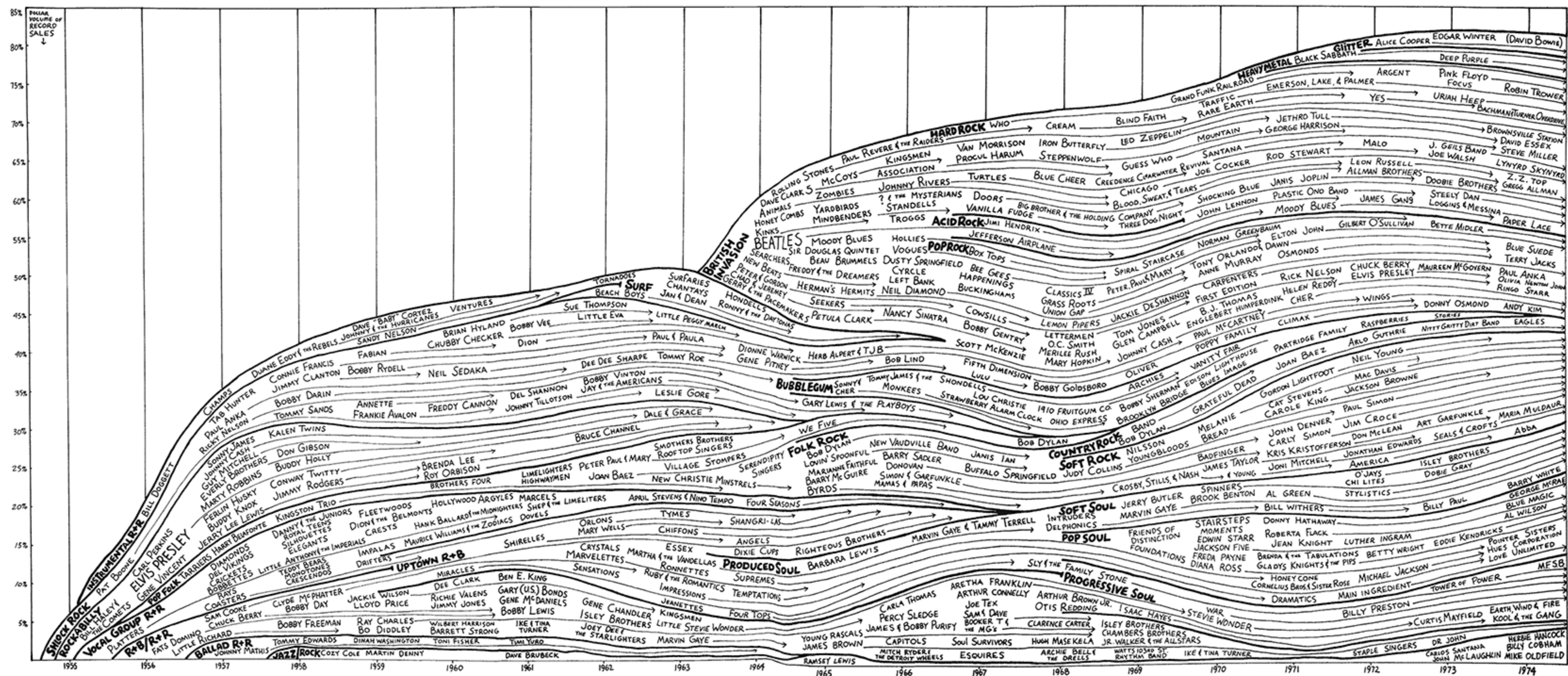
What types of data can we visualize?

Types of Data

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

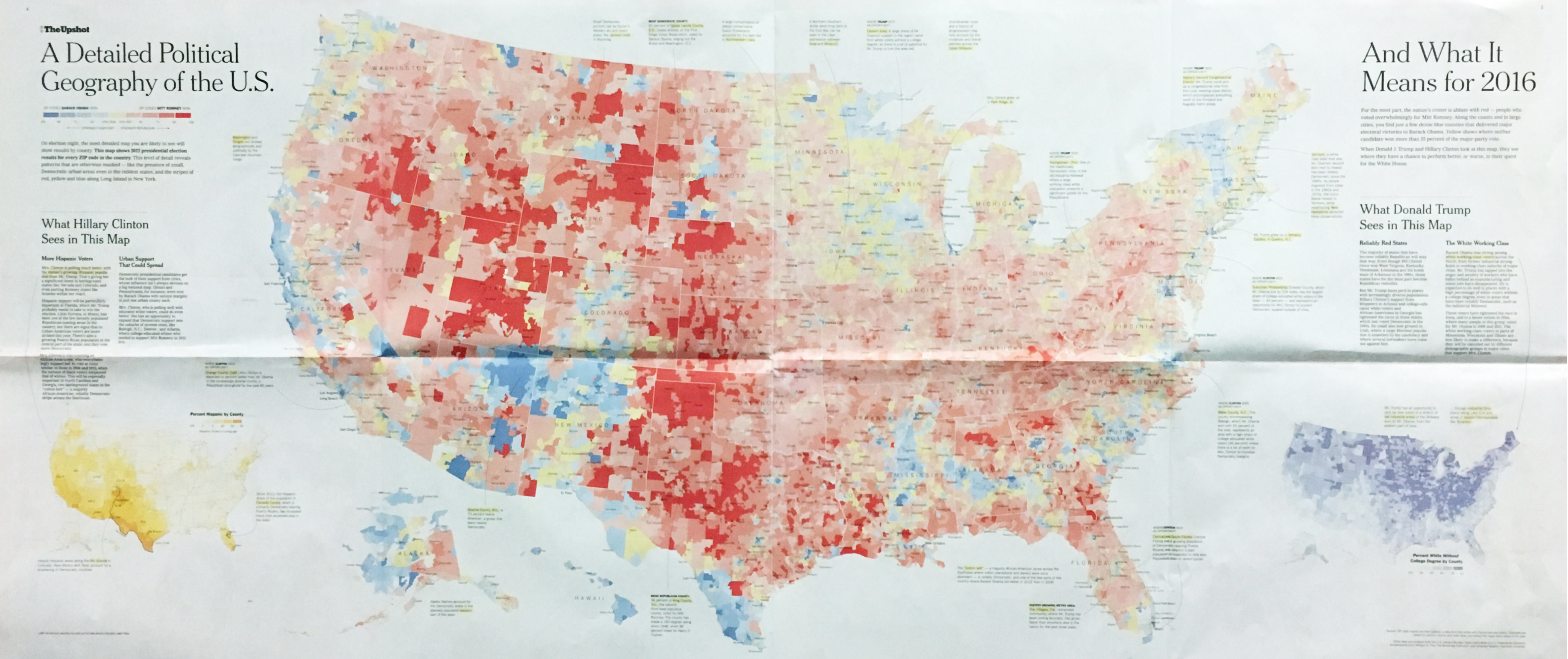
Where have you seen visualizations?

Books/Posters

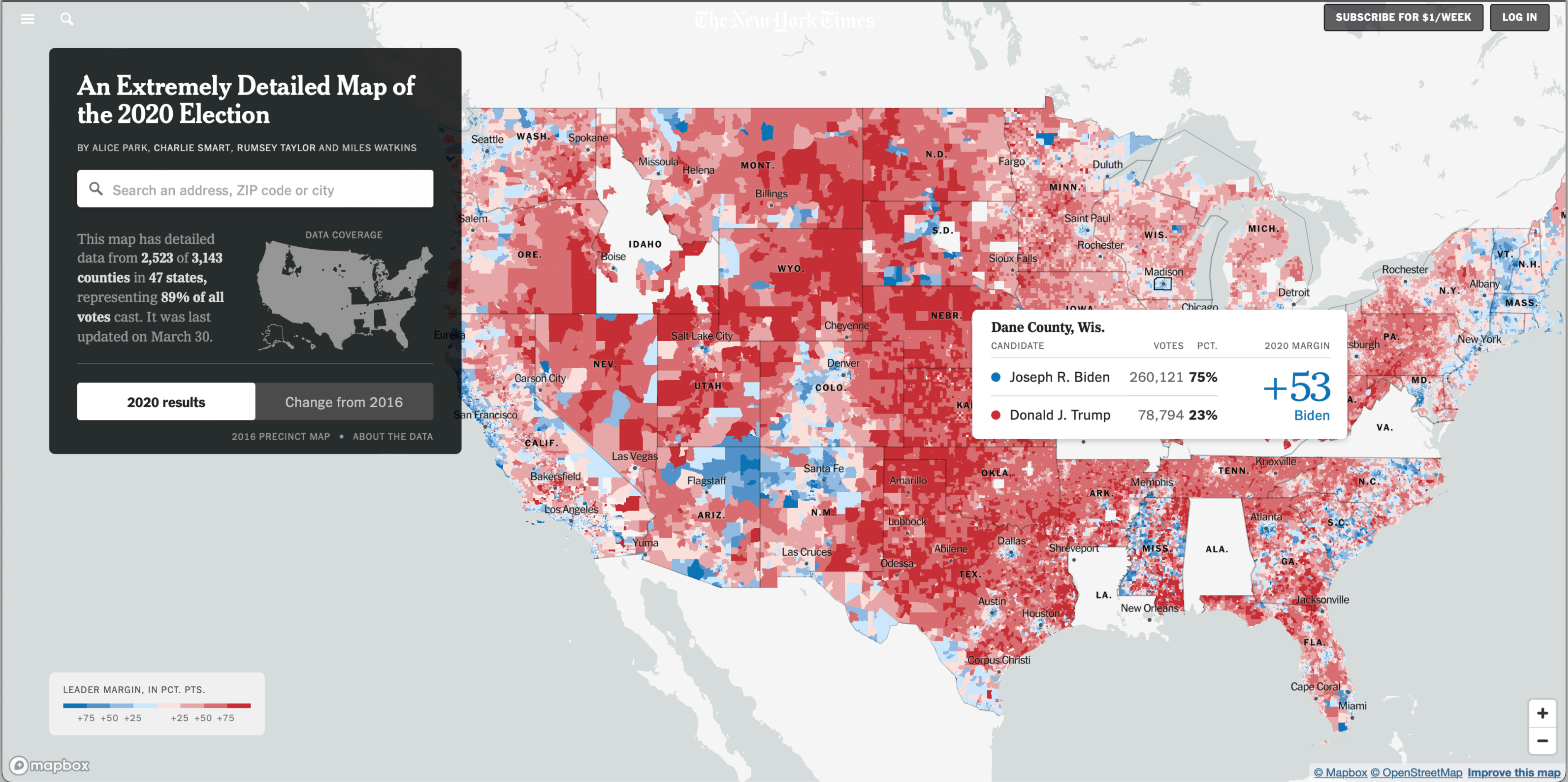


[Rock 'N' Roll is Here to Pay, R. Garofalo, 1977 (via Tufte)]

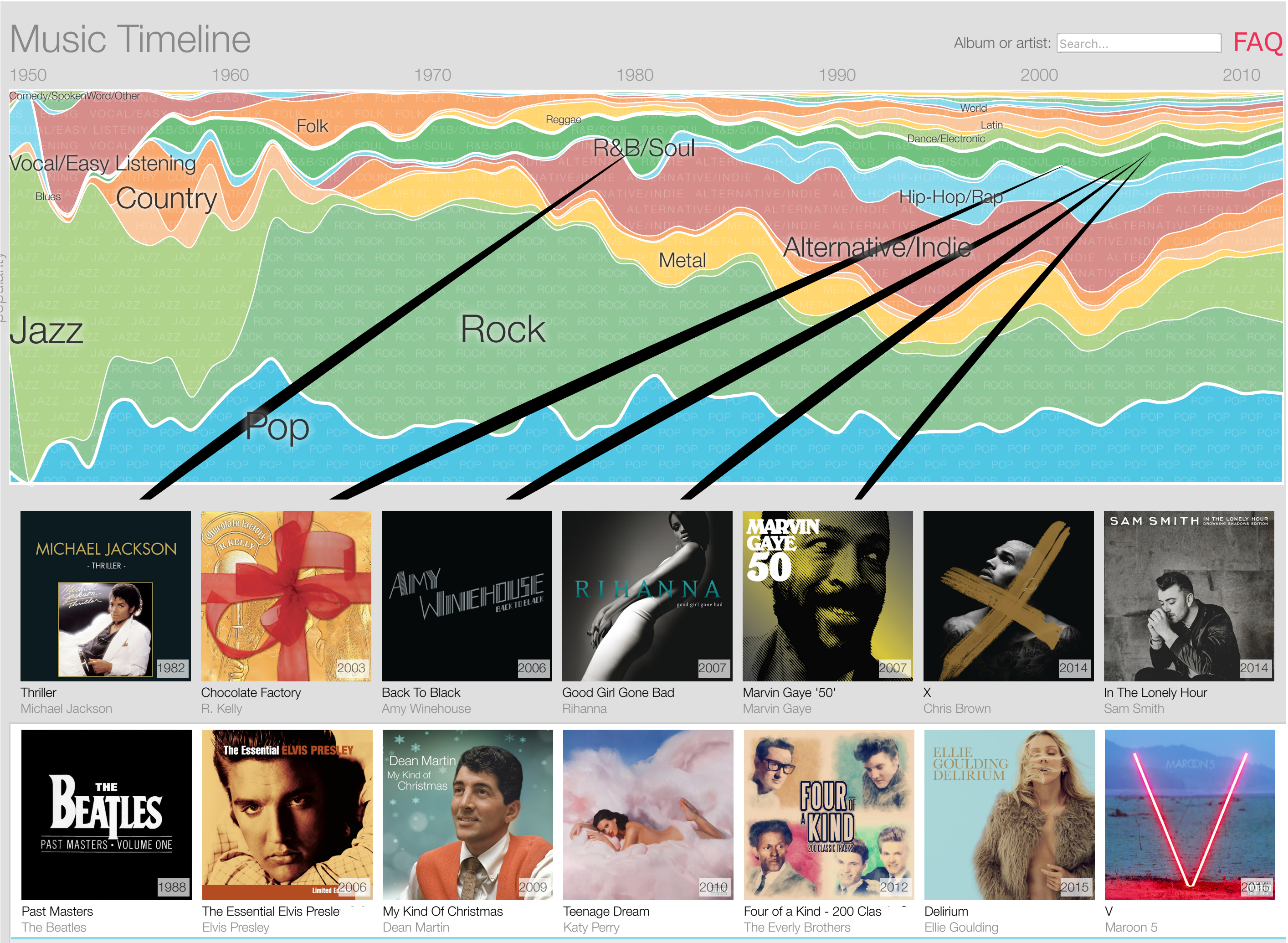
Newspapers



[NYTimes]



[NYTimes]



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

What is the advantage of the second version?

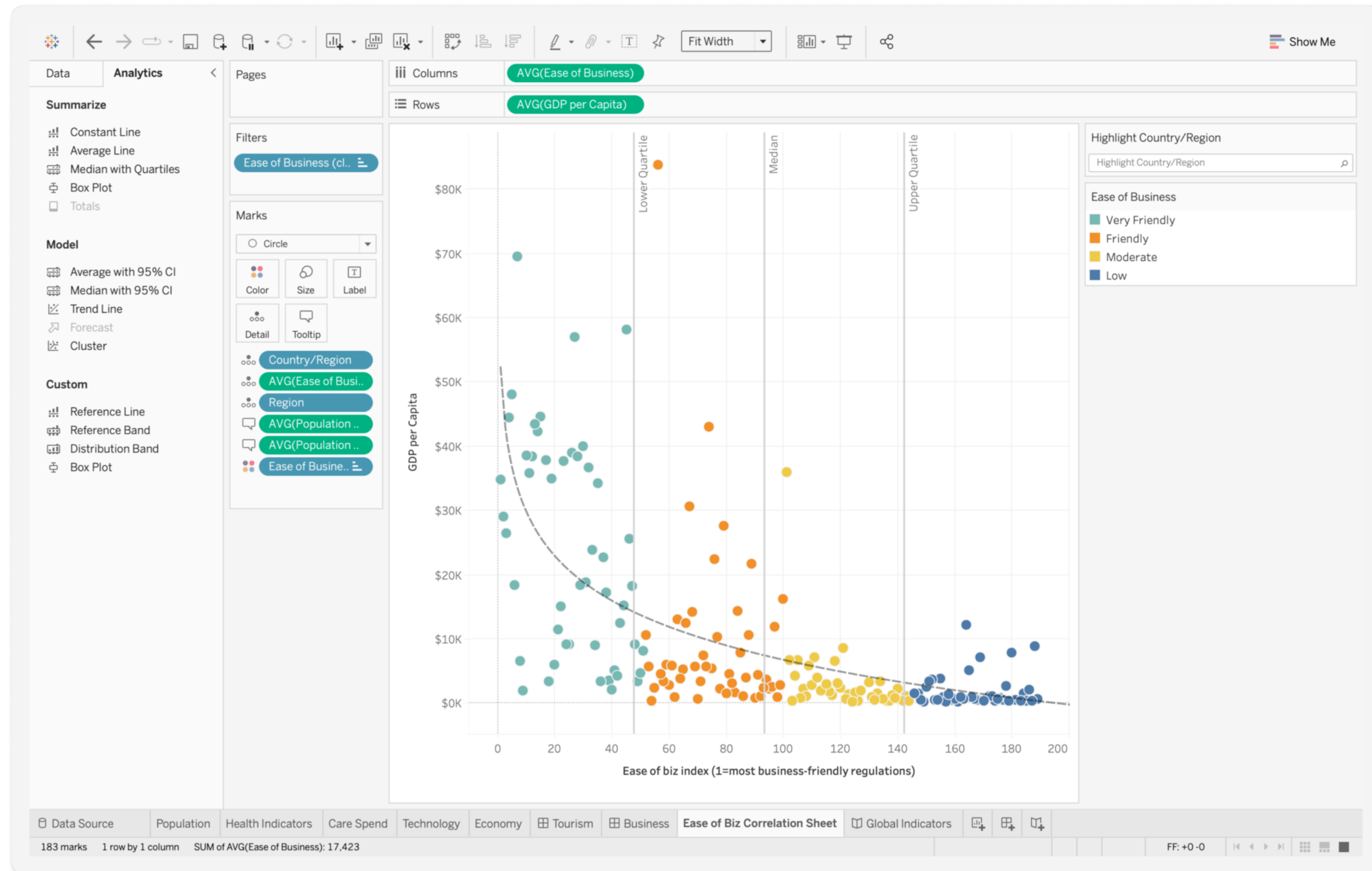
Interaction

How do we create modern visualizations?

Tools

- Desktop Applications:
 - Excel (see [excelcharts.com](https://www.excelspider.com/excelcharts.com))
 - Tableau
 - ...
- Grammars:
 - Vega-Lite
- Programming Frameworks
 - d3.js
 - Observable Plot, plot.ly, deck.gl
 - ...
- Tradeoffs
 - Speed
 - Customization
 - Understanding
 - Dissemination

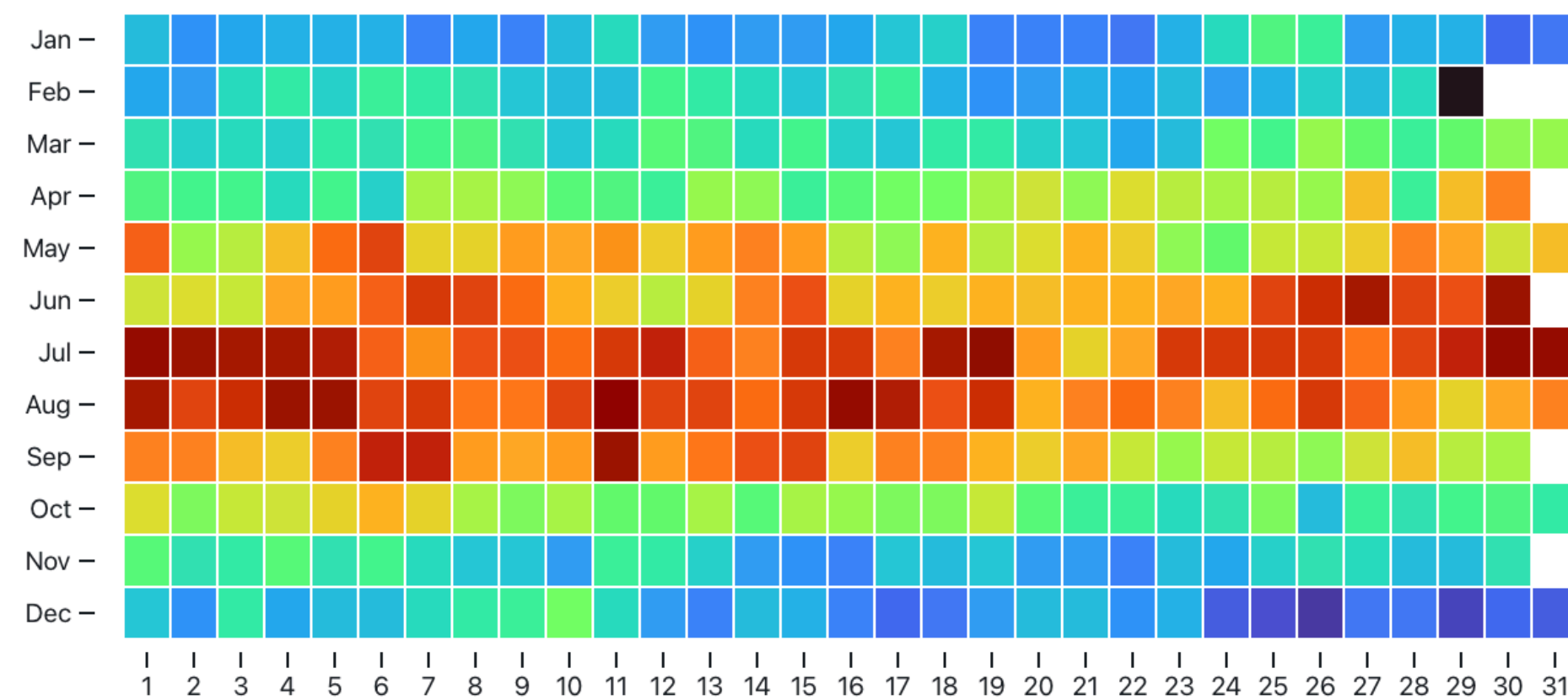
Tableau



[Tableau Desktop]

Observable Plot

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

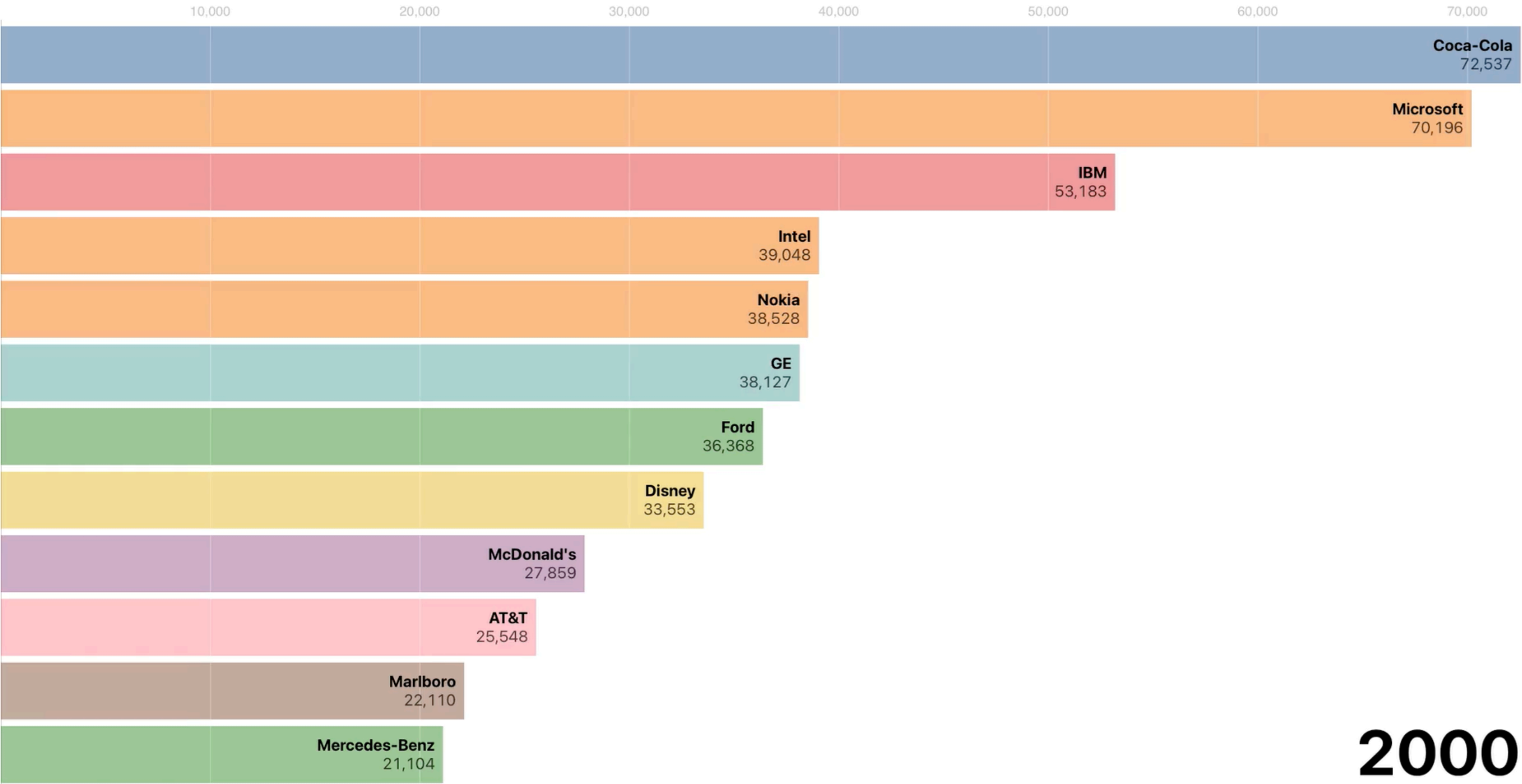


```
Plot.plot({
  height: 300,
  padding: 0,
  y: {
    tickFormat: Plot.formatMonth("en", "short")
  },
  marks: [
    Plot.cell(seattle, Plot.group({fill: "max"}, {
      x: d => d.date.getUTCDate(),
      y: d => d.date.getUTCMonth(),
      fill: "temp_max",
      inset: 0.5
    })))
  ]
})
```

[Observable]

Best Global Brands

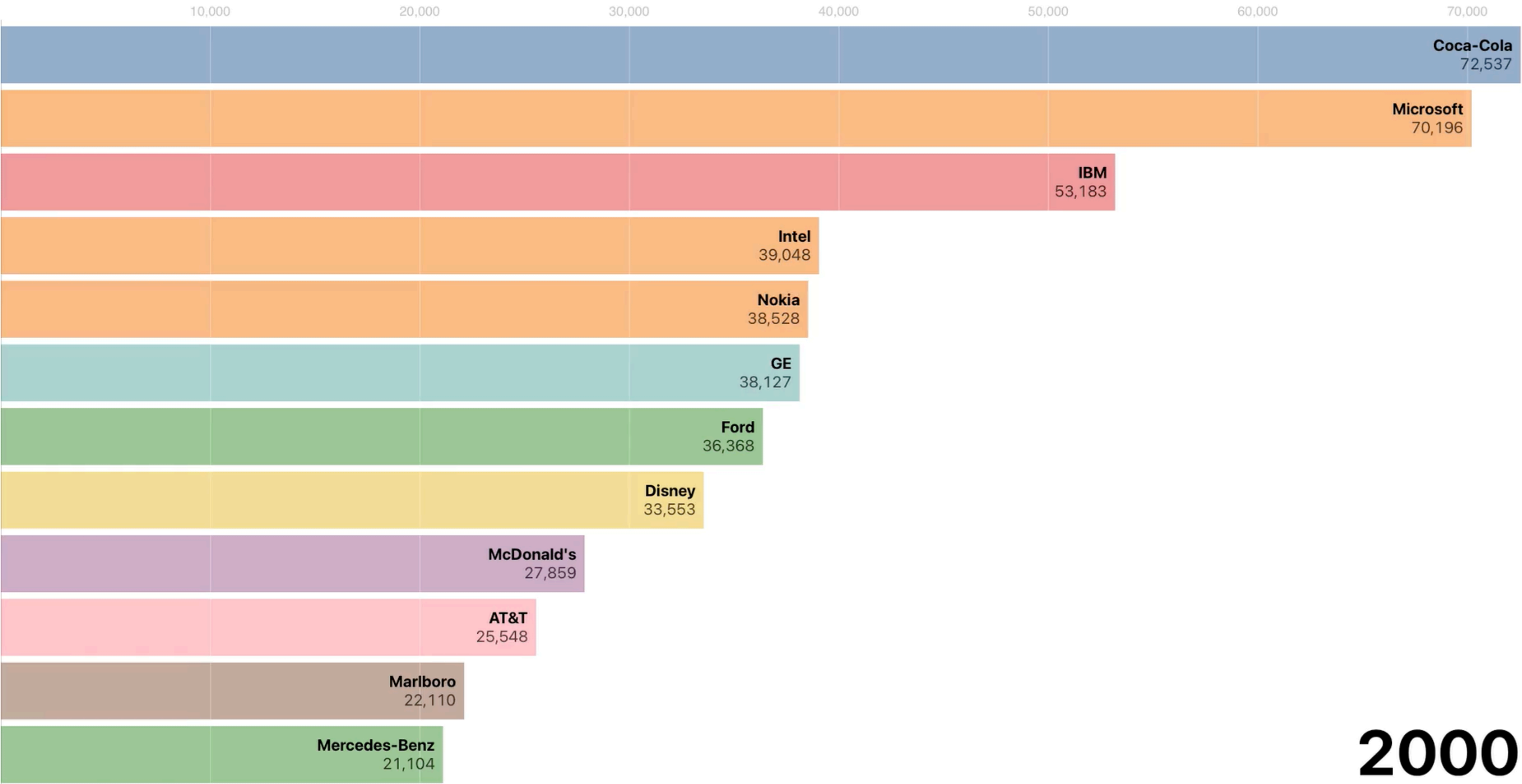
Value in \$M; color indicates sector. Data: [Interbrand](#)



2000 [M. Bostock]

Best Global Brands

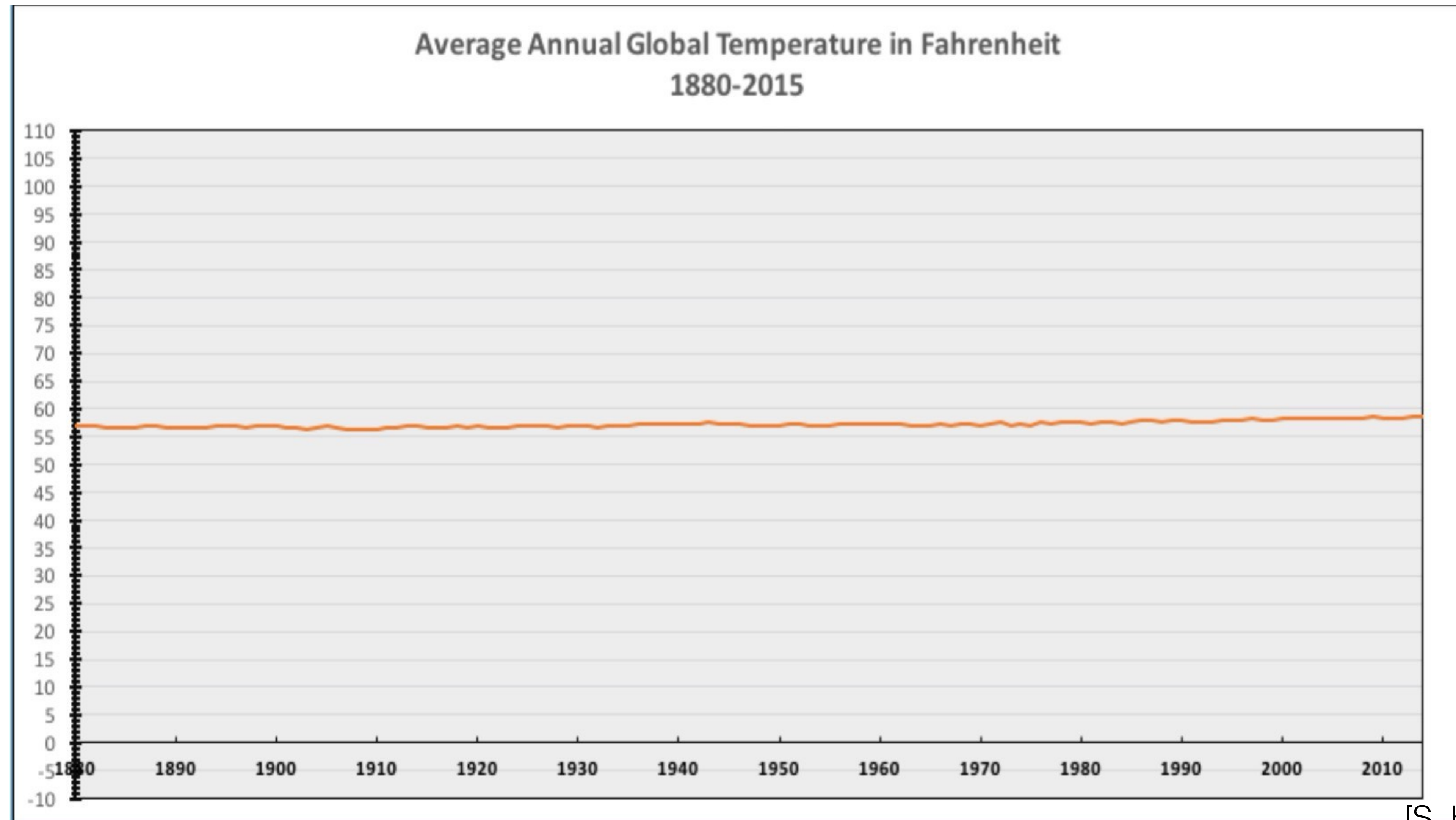
Value in \$M; color indicates sector. Data: [Interbrand](#)



2000 [M. Bostock]

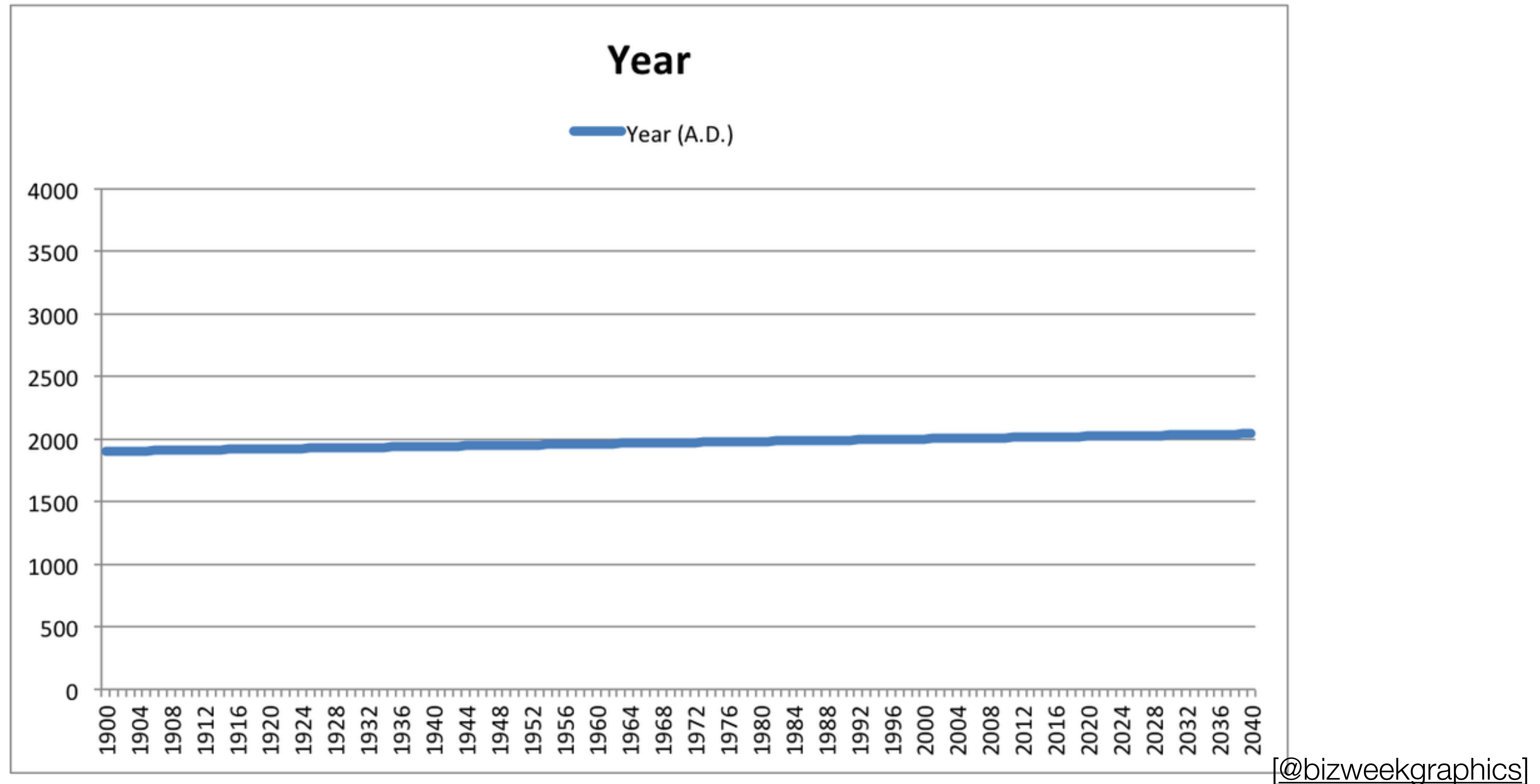
Why do we care about the design of visualizations?

Design: Focus on only the y-axis

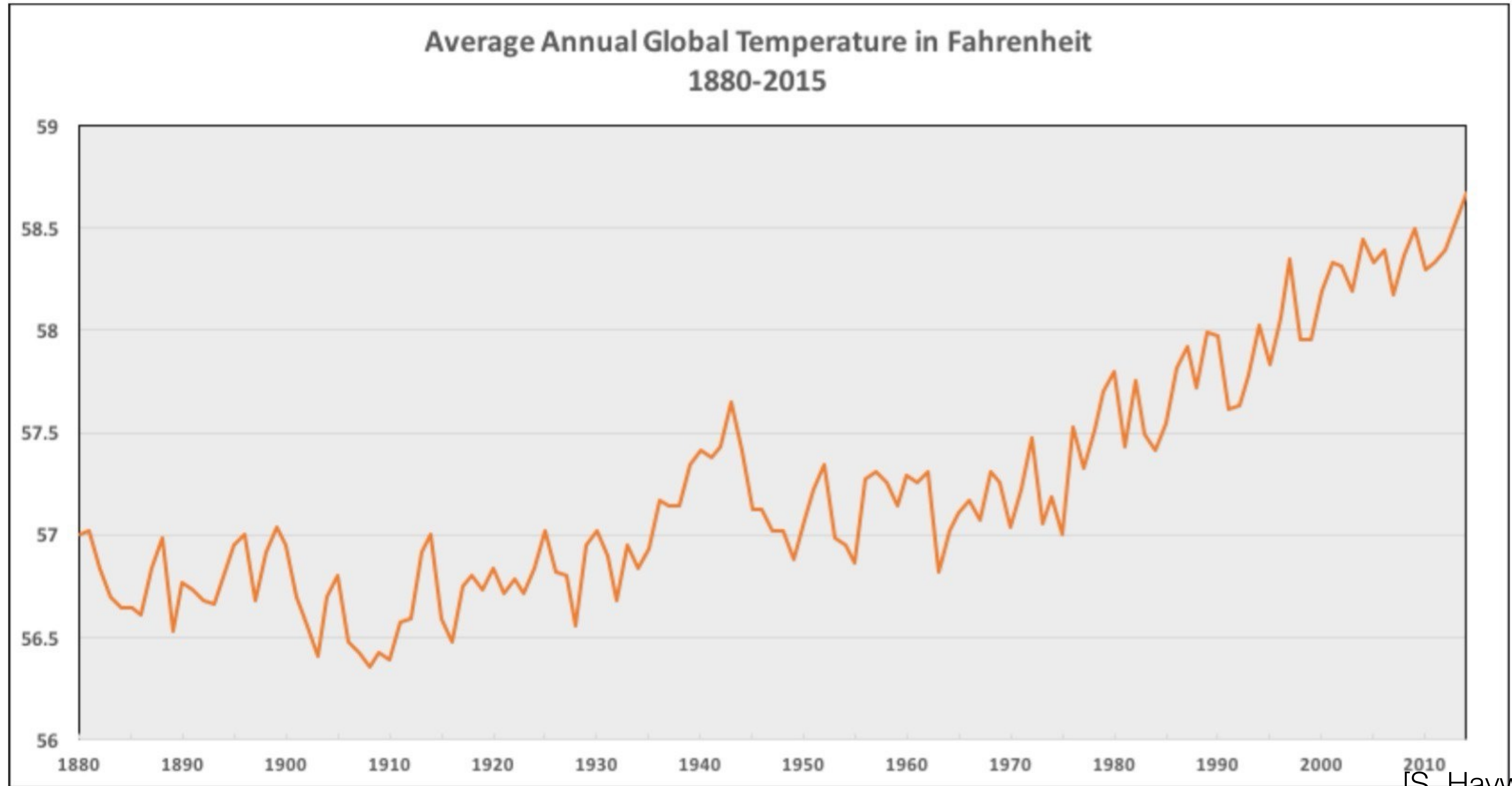


[S. Hayward, 2015]

Design: Year on the y-axis



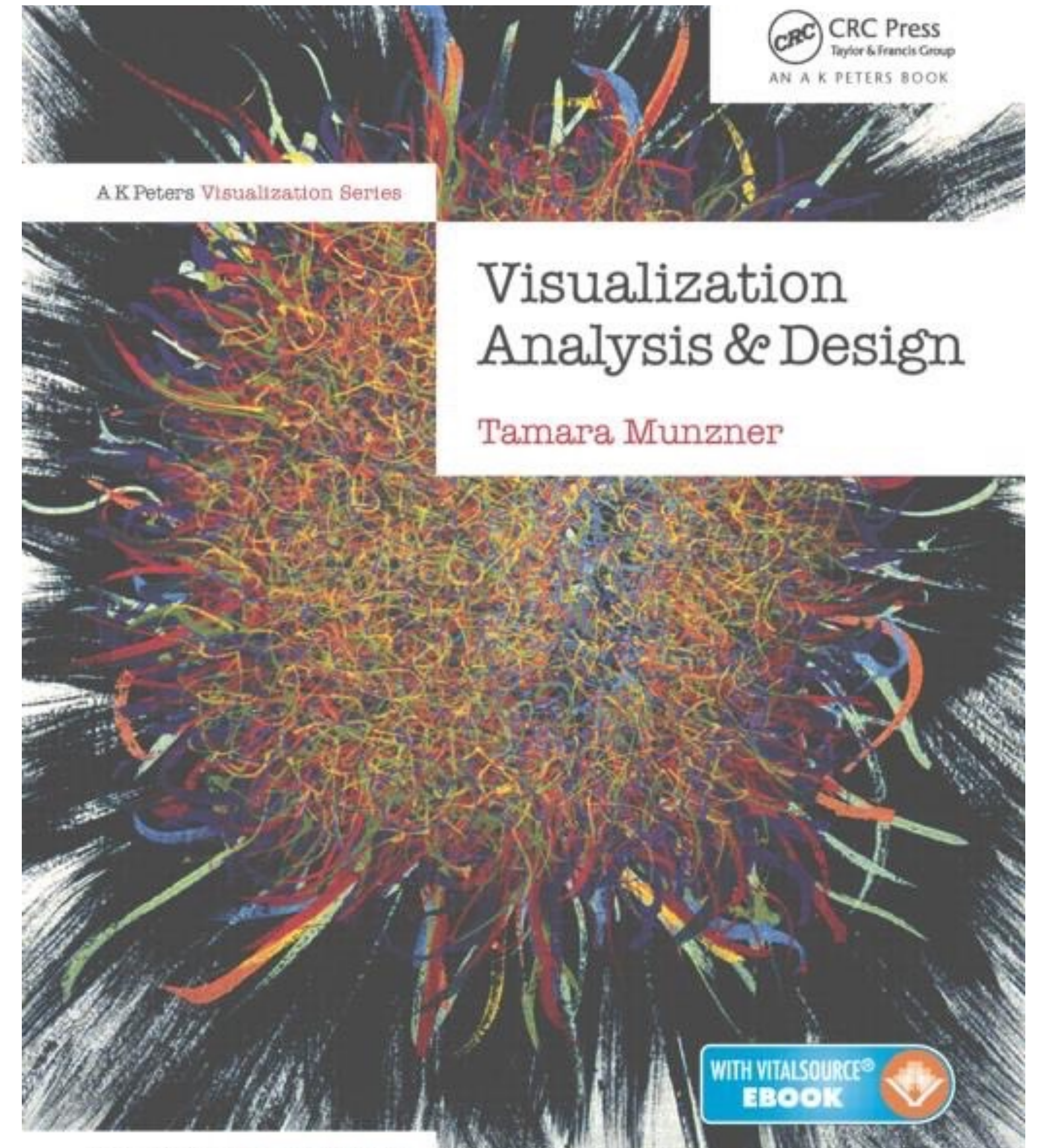
Design: Different y-axis



[S. Hayward, 2015]

Administrivia

- Course Web Site
- Syllabus
 - Plagiarism
 - Accommodations
- Textbook:
 - Required: Munzner (VAD)
- Assignments
- Exams: Midterm (Mar. 5) and Final (May 7)



Administrivia

- Undergraduate (CSCI 490) and Graduate (CSCI 627) Course
 - 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

Exams, Quizzes, and Missed Classes

- All quizzes and exams are **in-person**, **closed-book**, and **closed-notes**
- **No makeup** quizzes or exams will be given*
- You are responsible for all material presented in class and assigned readings. If you miss a lecture, be sure to obtain class notes from a classmate before the next class meeting

Office Hours & Communication

- Office hours will be held **in person**
- Scheduled office hours are open to all students
 - M: 1:45-3pm, W: 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 send me screenshots of code!** (send code or Observable links)

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, ...
- Do not share your work with anyone else
- Never quote text unless there is a specific need.
 - "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** for any assignment/exam/etc. where cheating has occurred
- You will **fail** the course if you cheat more than once
- Misconduct is reported through the university's system
- You **may** discuss problems and approaches with other students
- You **may not** copy or transcribe code from another source (includes generative AI)

Do ask questions!

Do ask questions

- If you are stuck on a specific issue with an assignment:
 - Do consult books, online documentation, tutorials
 - Do discuss that specific issue with a classmate
 - Do email me with **specific** questions
- 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
 - Discuss approaches or techniques

Questions?