Data Visualization (CSCI 627/490)

Definition & Web Programming

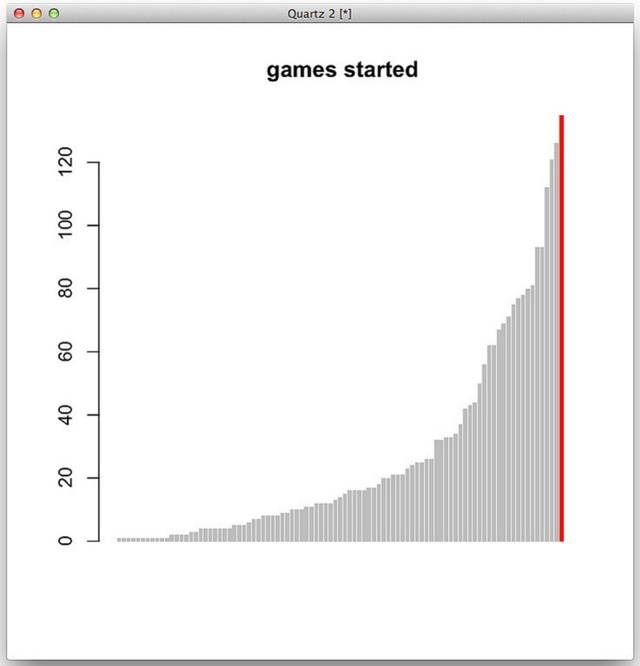
Dr. David Koop

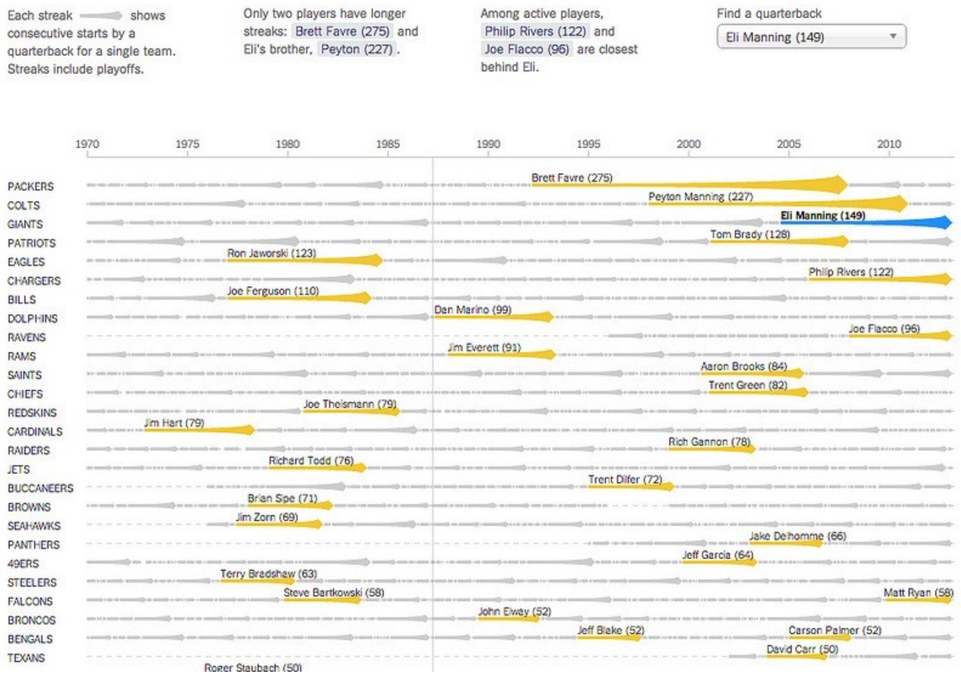
The purpose of **visualization** is about **insight**, not pictures

—B. Shneiderman

Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team



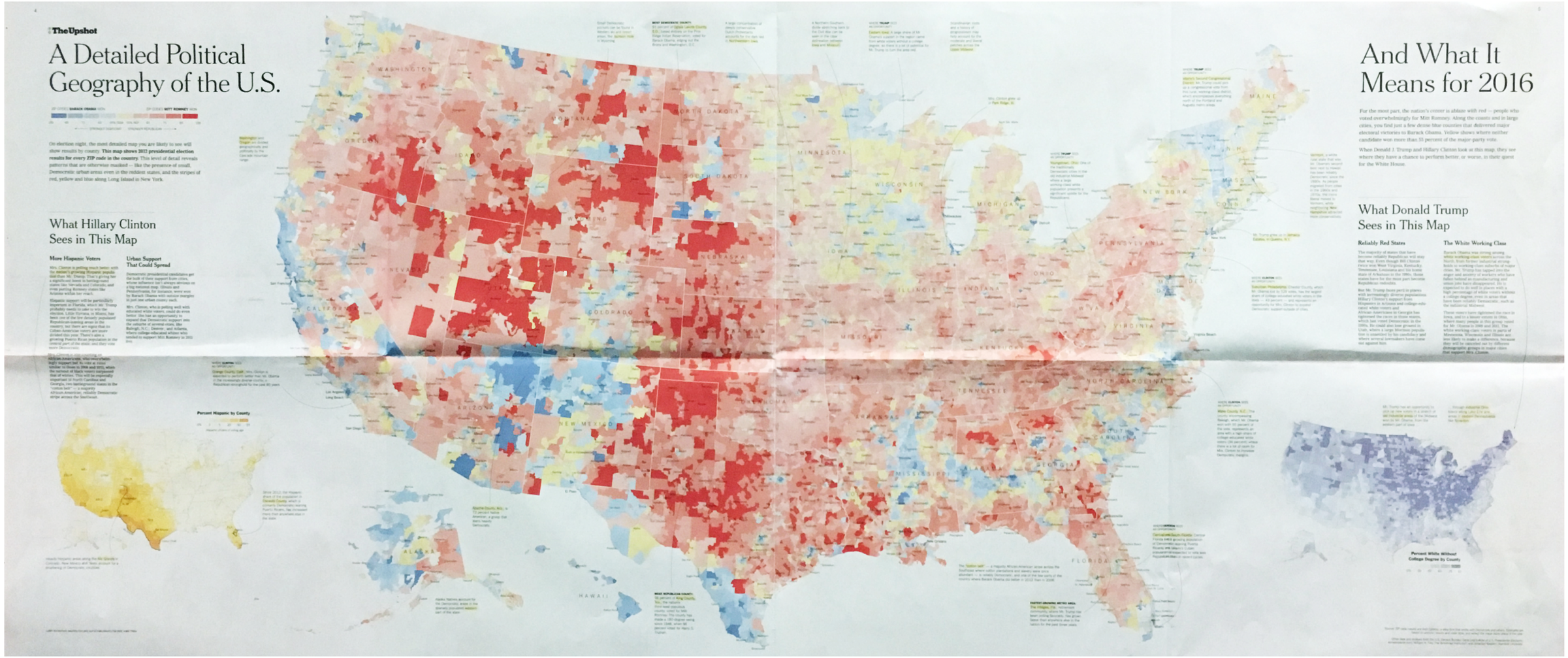


Exploration — Confirmation — Communication

Questions — **Answers/Persuasion**

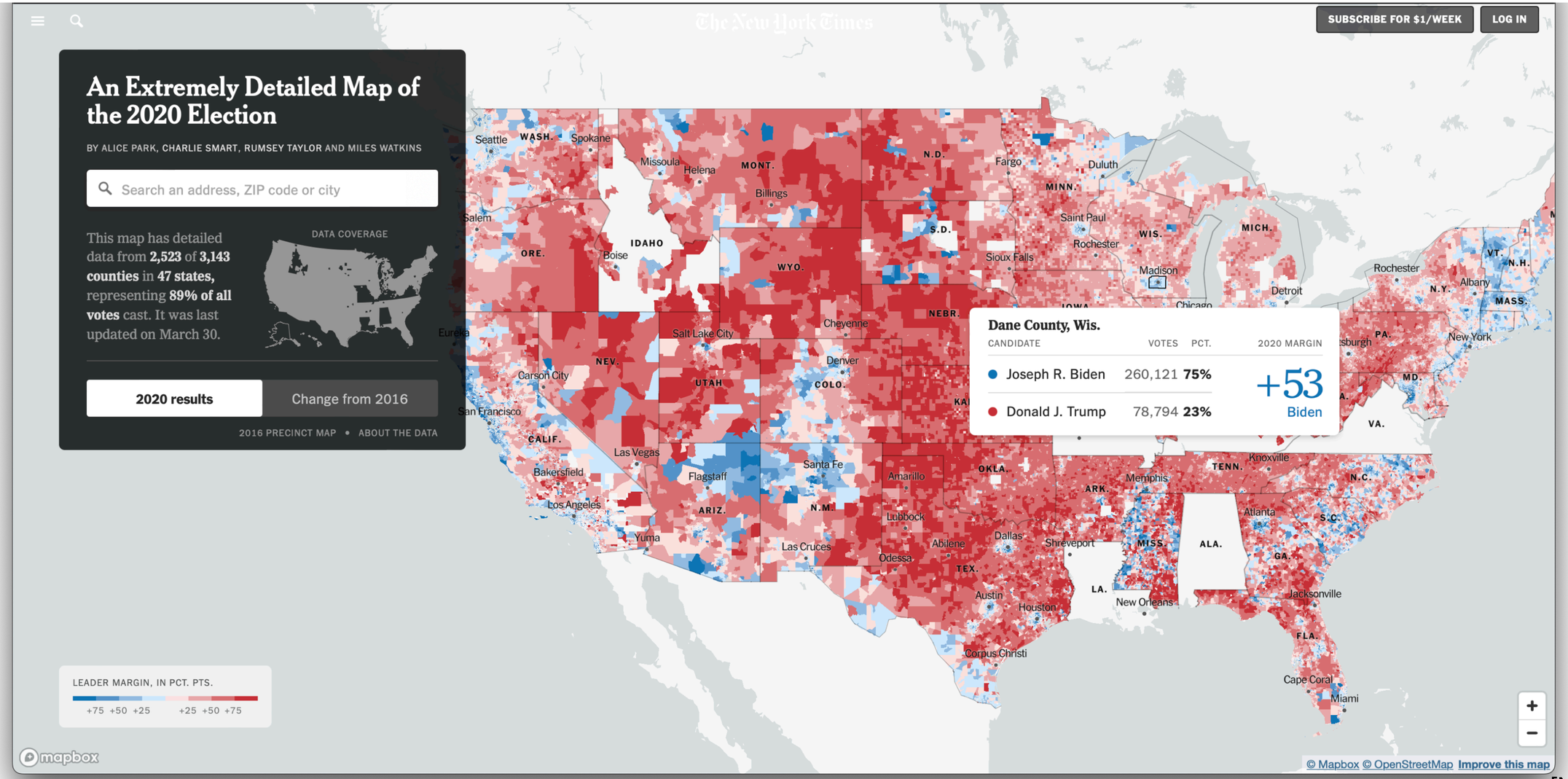
[K. Quealy, 2013]

Static Visualization



[NYTimes]

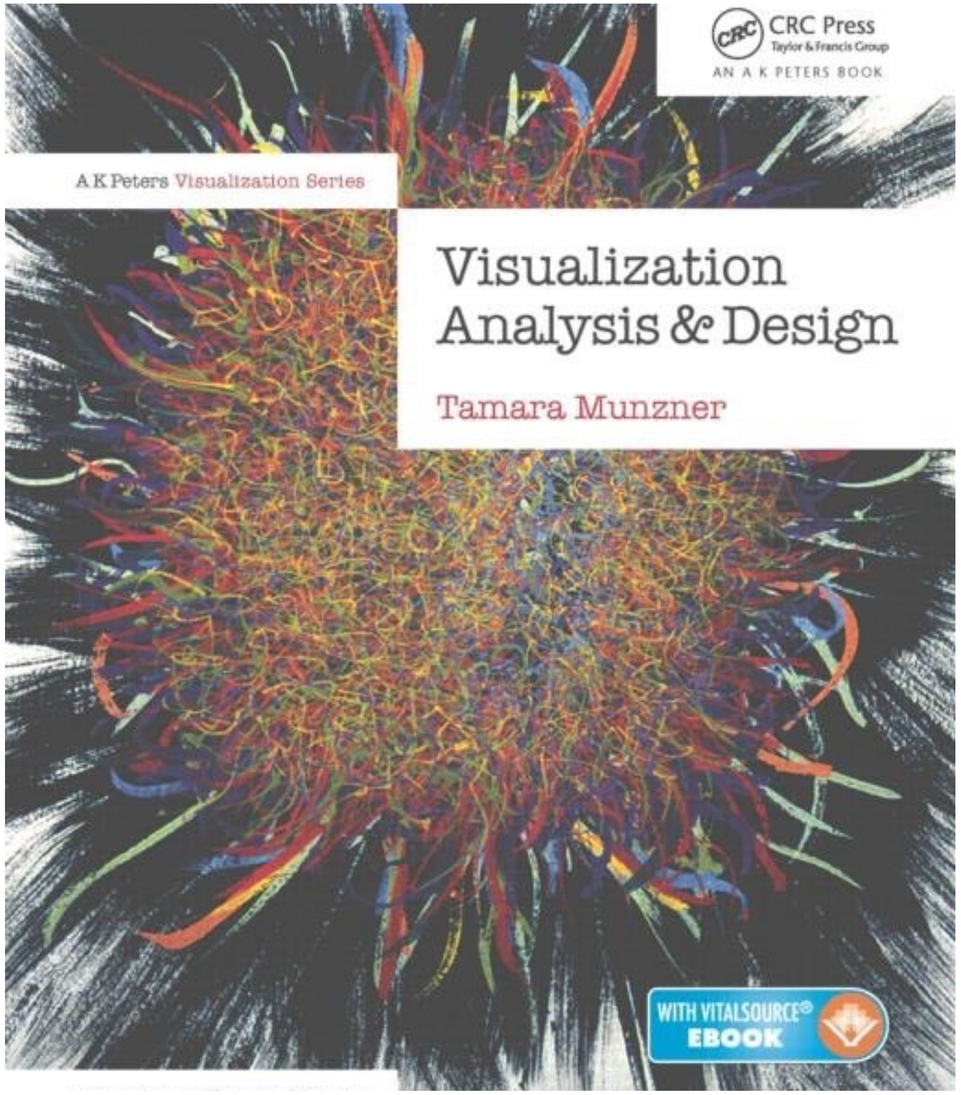
Interactive Visualization



[NYTimes]

Administrivia

- Course Web Site
- Syllabus
  - Plagiarism
  - Accommodations
- Textbook:
  - Required: Munzner (VAD)
- Assignments
- Exams: Midterm (Oct. 14) and Final (Dec. 9)



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 & Communication

- Scheduled in-person 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 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 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?

Assignment 1

- To be released soon…
- Write HTML, CSS, and SVG
- Text markup and styling
- Drawing markup and styling
- Data Visualization using Observable Plot

Definition of Visualization

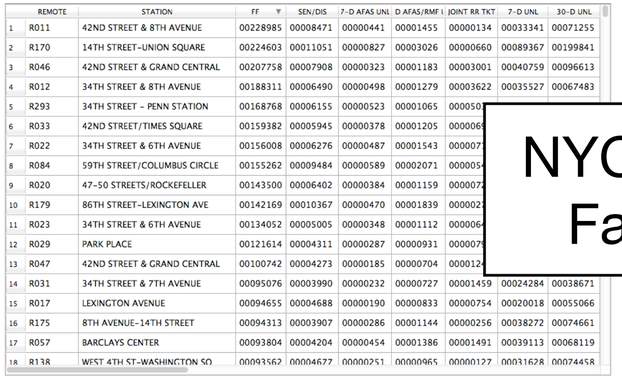
"Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively."

— T. Munzner

Definition

"Computer-based visualization systems provide visual representations of **datasets** designed to help people carry out **tasks** more effectively."

Definition



	REMOTE	STATION	FF	SEN/DIS	7-D AFAS UNL	D AFAS/RMF U	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	0000050		
6	R033	42ND STREET/TIMES SQUARE	00159382	00005945	00000378	00001205	0000006		
7	R022	34TH STREET & 6TH AVENUE	00156008	00006276	00000487	00001543	0000007		
8	R084	59TH STREET/COLUMBUS CIRCLE	00155262	00009484	00000589	00002071	0000005		
9	R020	47-50 STREETS/ROCKEFELLER	00143500	00006402	00000384	00001159	0000007		
10	R179	86TH STREET-LEXINGTON AVE	00142169	00010367	00000470	00001839	0000002		
11	R023	34TH STREET & 6TH AVENUE	00134052	00005005	00000348	00001112	0000006		
12	R029	PARK PLACE	00121614	00004311	00000287	00000931	0000007		
13	R047	42ND STREET & GRAND CENTRAL	00100742	00004273	00000185	00000704	0000012		
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 SQ	00093562	00004672	00000251	00000965	00000127	00031628	00074458

NYC Subway Fare Data

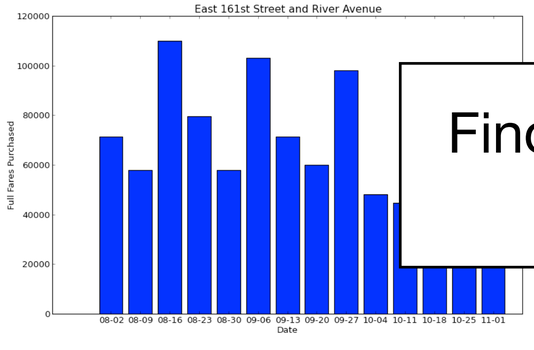
"Computer-based visualization systems provide visual representations of **datasets** designed to help people carry out **tasks** more effectively."

Definition

	REMOTE	STATION	FF	SEN/DIS	7-D AFAS UNL	D AFAS/RMF U	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	0000050		
6	R033	42ND STREET/TIMES SQUARE	00159382	00005945	00000378	00001205	0000006		
7	R022	34TH STREET & 6TH AVENUE	00156008	00006276	00000487	00001543	0000007		
8	R084	59TH STREET/COLUMBUS CIRCLE	00155262	00009484	00000589	00002071	0000005		
9	R020	47-50 STREETS/ROCKEFELLER	00143500	00006402	00000384	00001159	0000007		
10	R179	86TH STREET-LEXINGTON AVE	00142169	00010367	00000470	00001839	0000002		
11	R023	34TH STREET & 6TH AVENUE	00134052	00005005	00000348	00001112	0000006		
12	R029	PARK PLACE	00121614	00004311	00000287	00000931	0000007		
13	R047	42ND STREET & GRAND CENTRAL	00100742	00004273	00000185	00000704	0000012		
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 SQ	00093562	00004672	00000251	00000965	00000127	00031628	00074458

NYC Subway Fare Data

"Computer-based visualization systems provide visual representations of **datasets** designed to help people carry out **tasks** more effectively."



Find Interesting NYC Subway Ridership Patterns

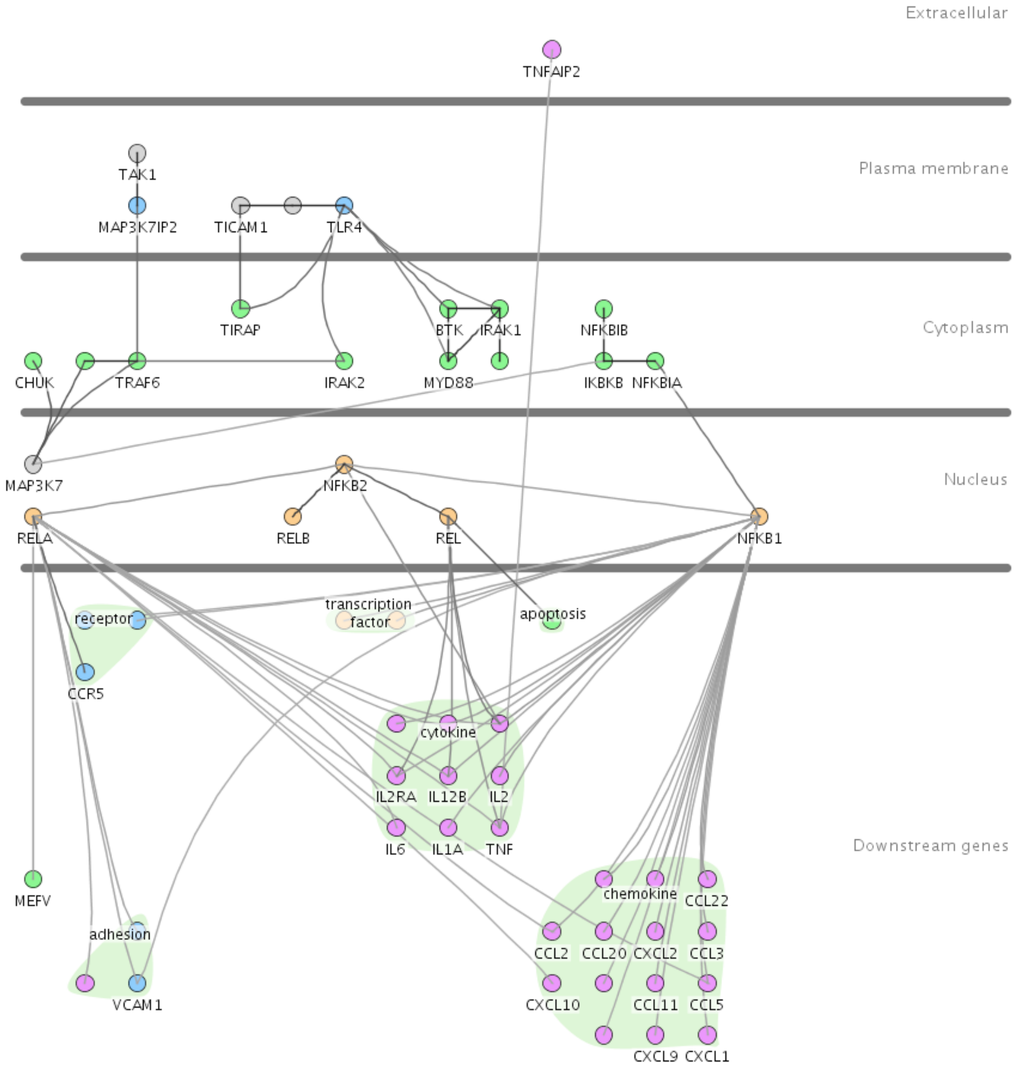
Definition

"**Computer-based** visualization systems provide visual representations of datasets designed to help **people** carry out tasks more effectively."

Why People?

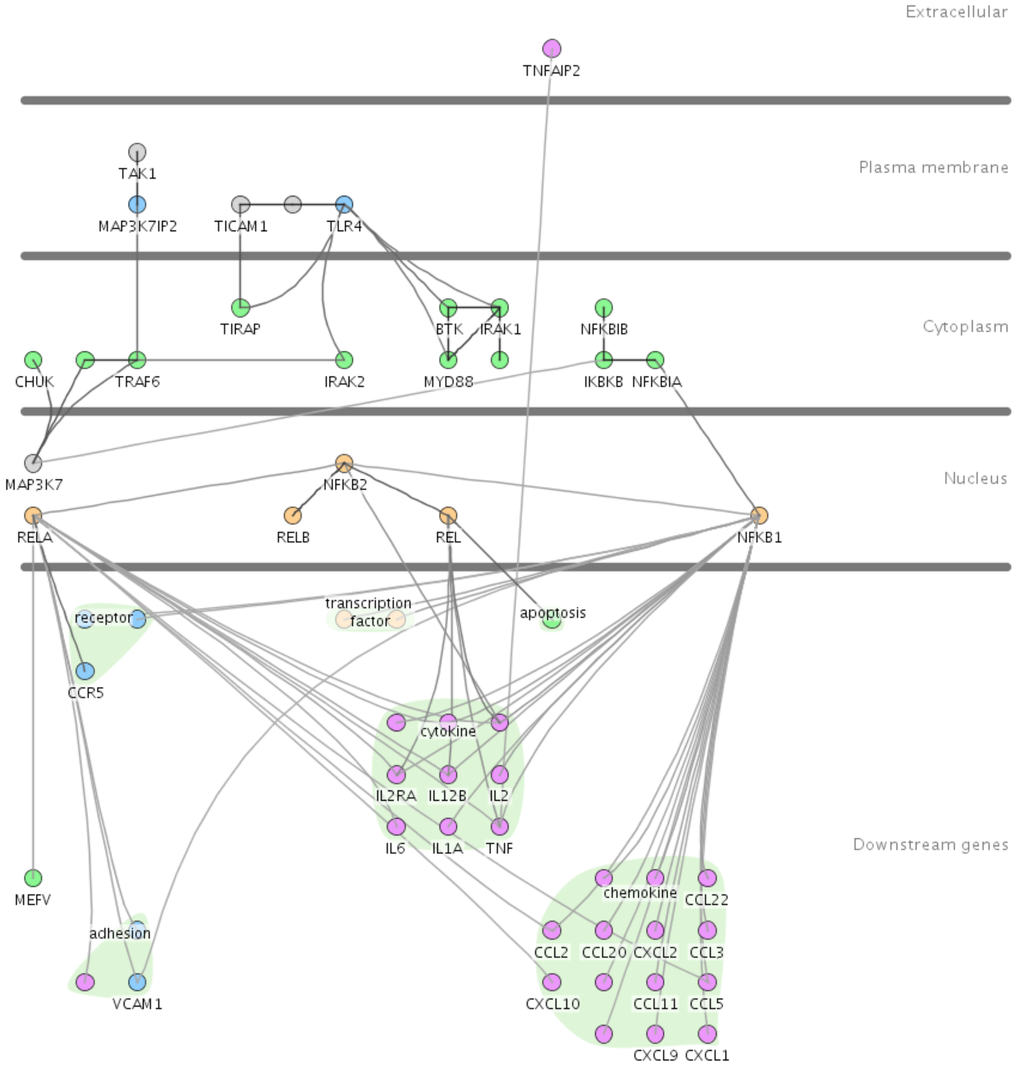
- Certain tasks can be totally **automated**
  - Statistical computations
  - Machine learning algorithms
  - Don’t need visualization for these tasks (maybe for debugging them…)
- Analysis problems are often **ill-specified**
  - What is the correct question?
  - Exploit human visual system, pattern detection capabilities
  - Goal may be an automated solution or a visual analysis system
- Presentation
  - It is often easier to show someone something than to tell them a bunch of facts about the data (and let them explore it)

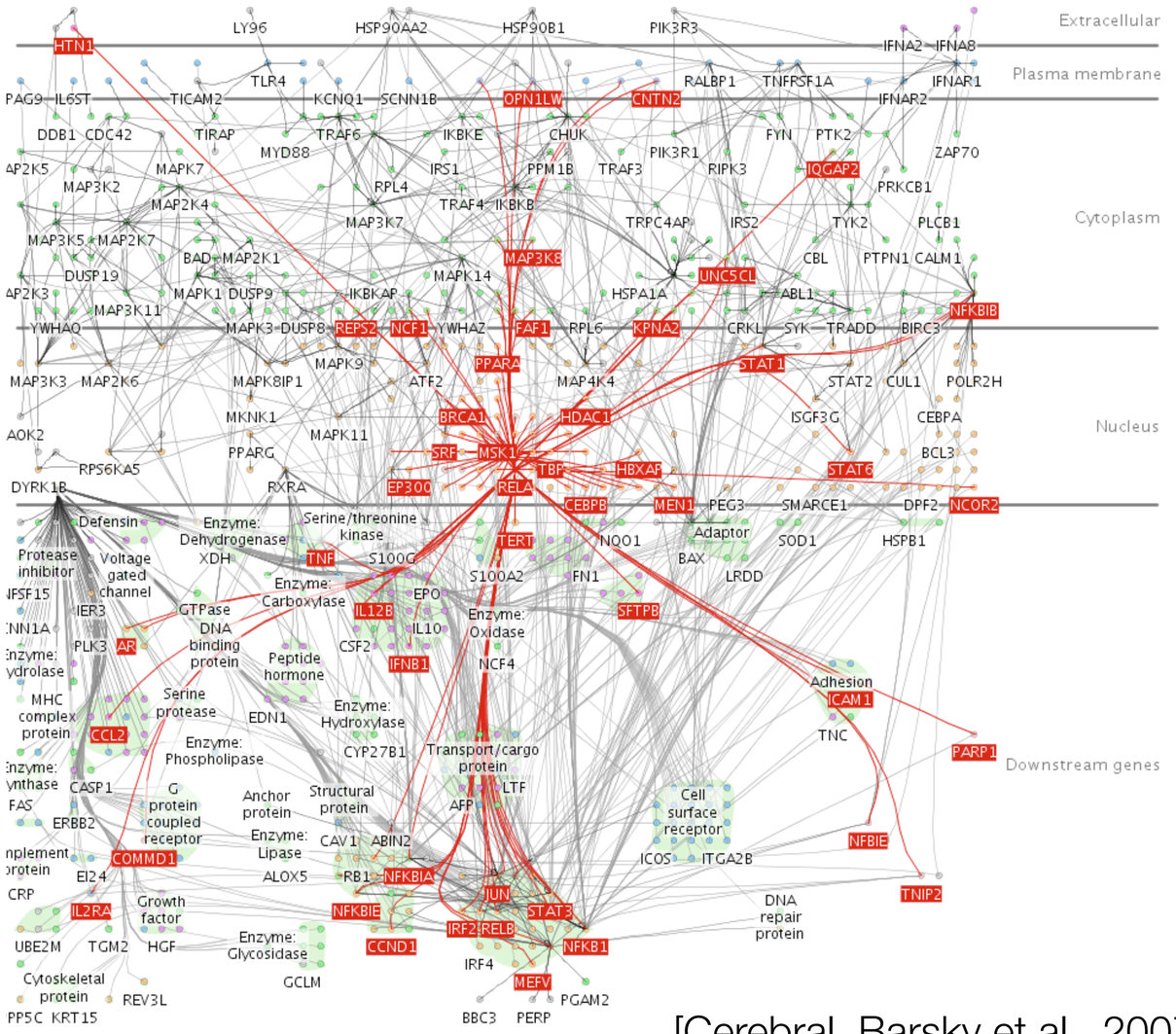
Why Computers?



[Cerebral, Barsky et al., 2007]

Why Computers?

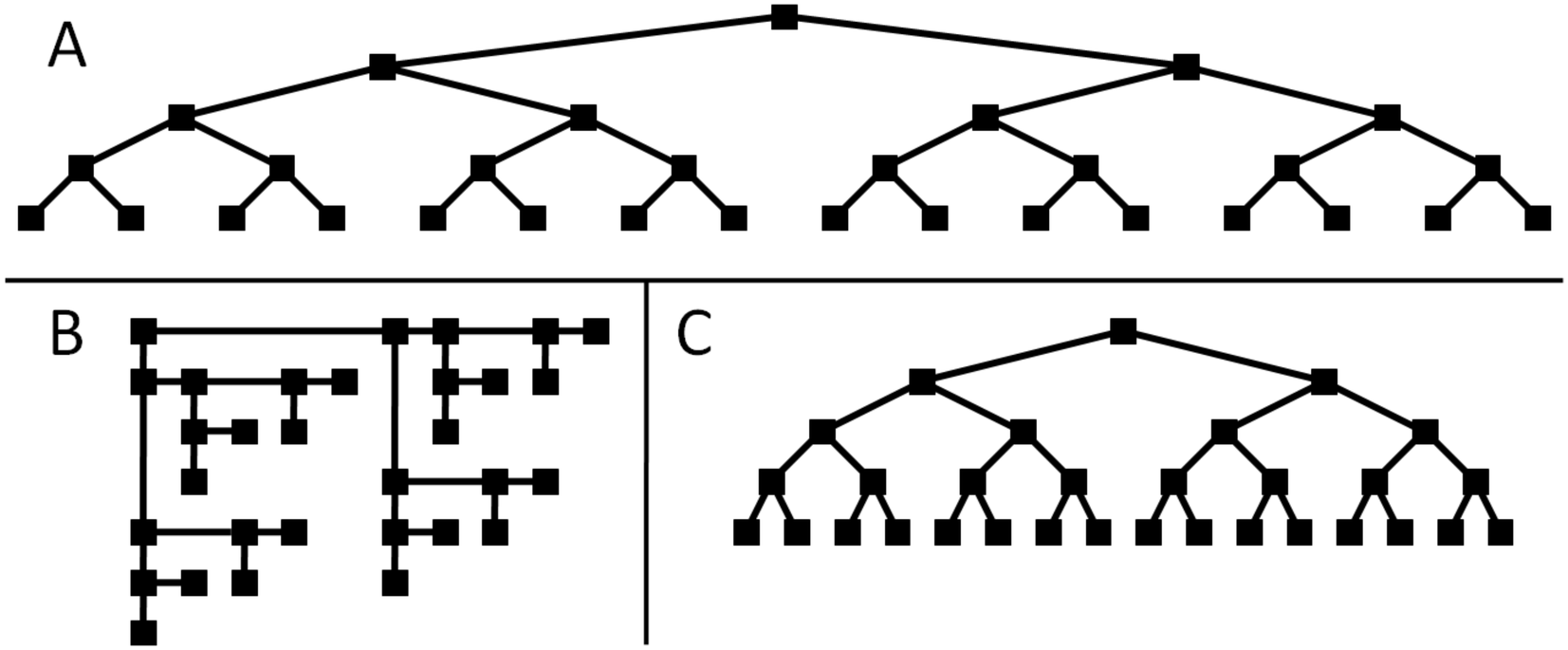




[Cerebral, Barsky et al., 2007]

Resource Limitations

- Memory and space constraints
- How many pixels do I have?
- Information Density

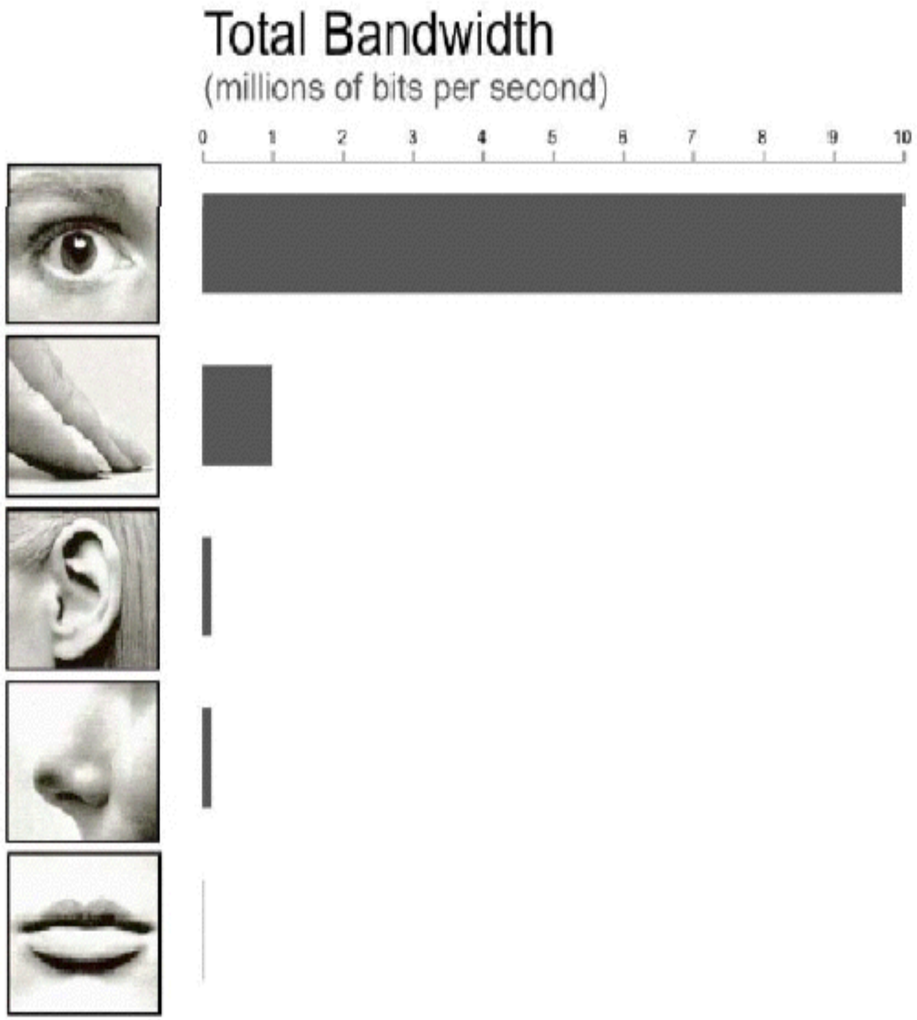


[McGuffin & Robert, 2010]

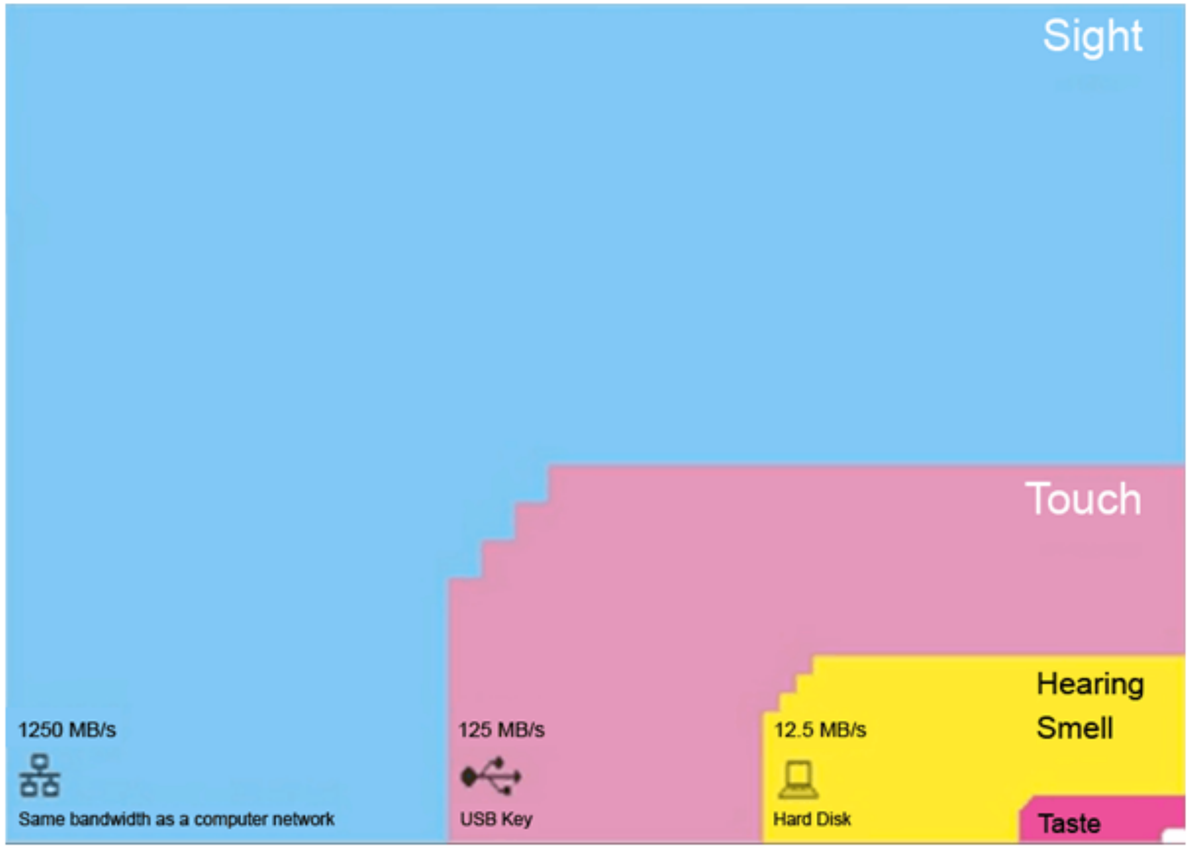
Definition

"Computer-based visualization systems provide **visual** representations of datasets designed to help people carry out tasks more effectively"

Why do we visualize data?



[via A. Lex]



[T. Nørretranders]

Why Visual?

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

[F. J. Anscombe]

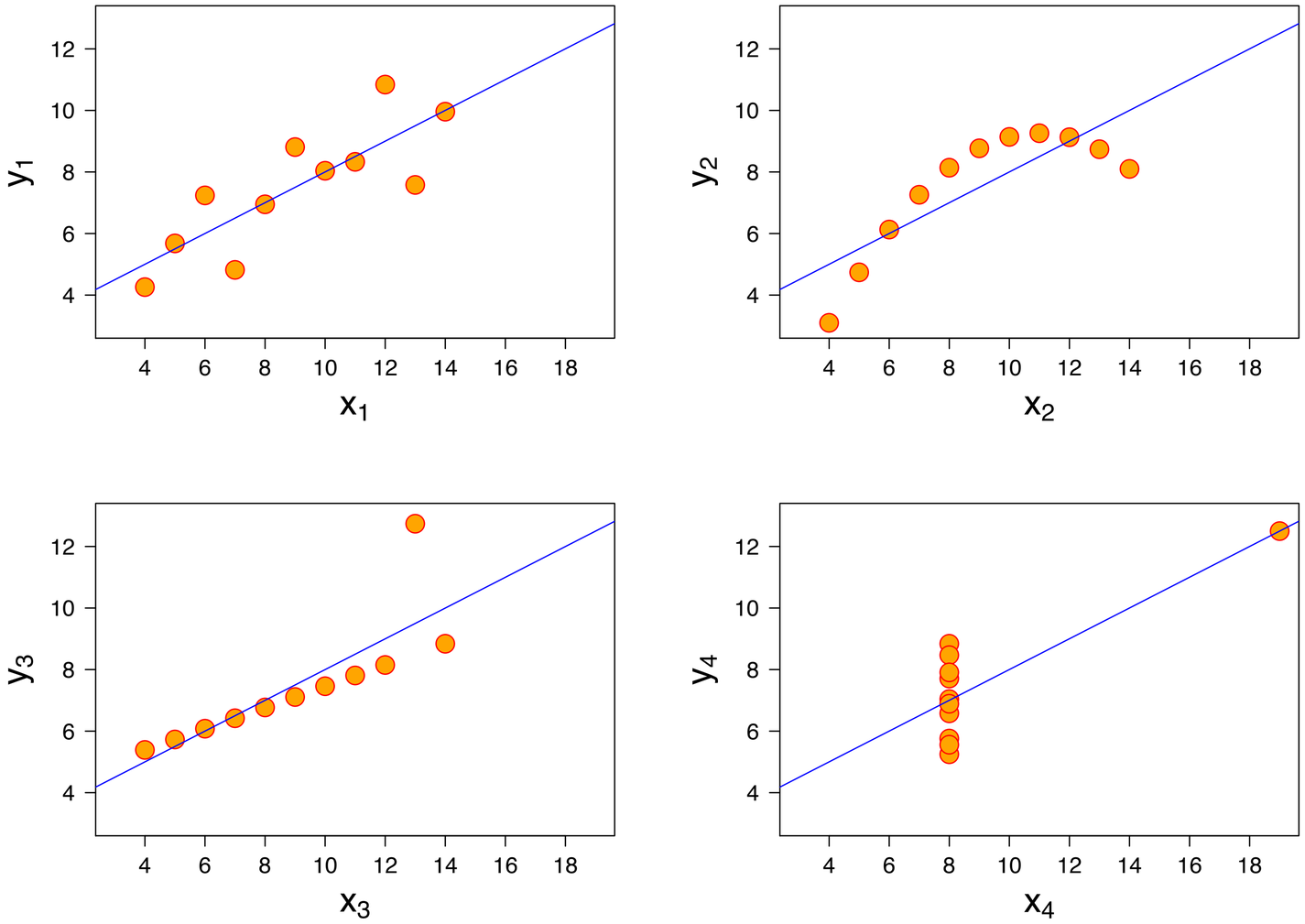
Why Visual?

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Mean of x	9
Variance of x	11
Mean of y	7.50
Variance of y	4.122
Correlation	0.816

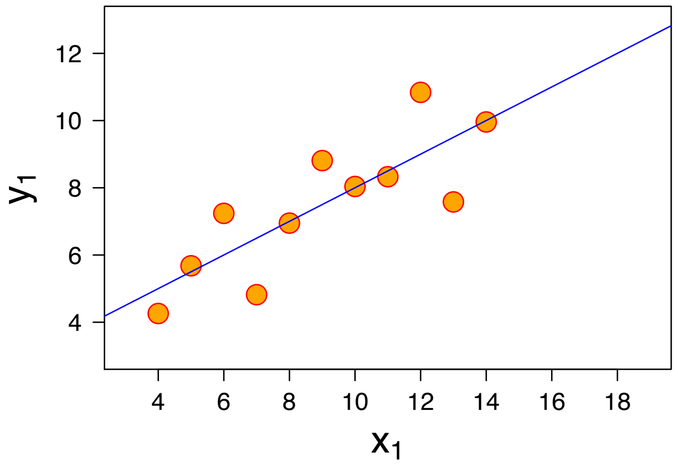
[F. J. Anscombe]

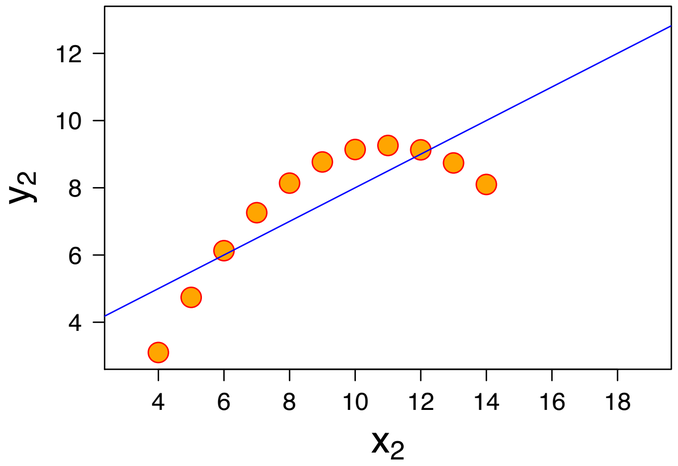
Why Visual?

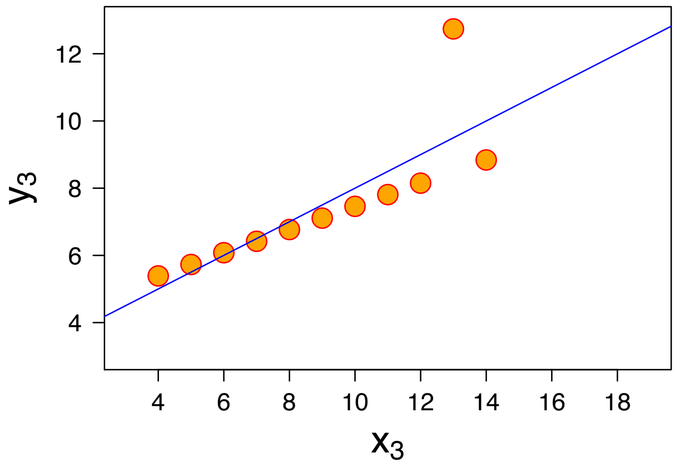


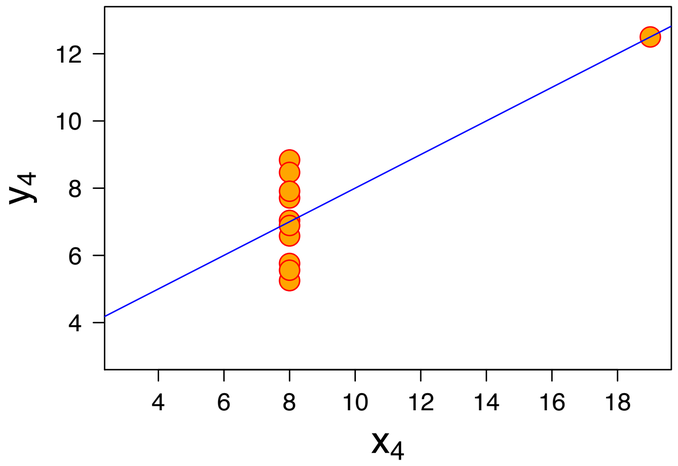
[F. J. Anscombe]

Why Visual?





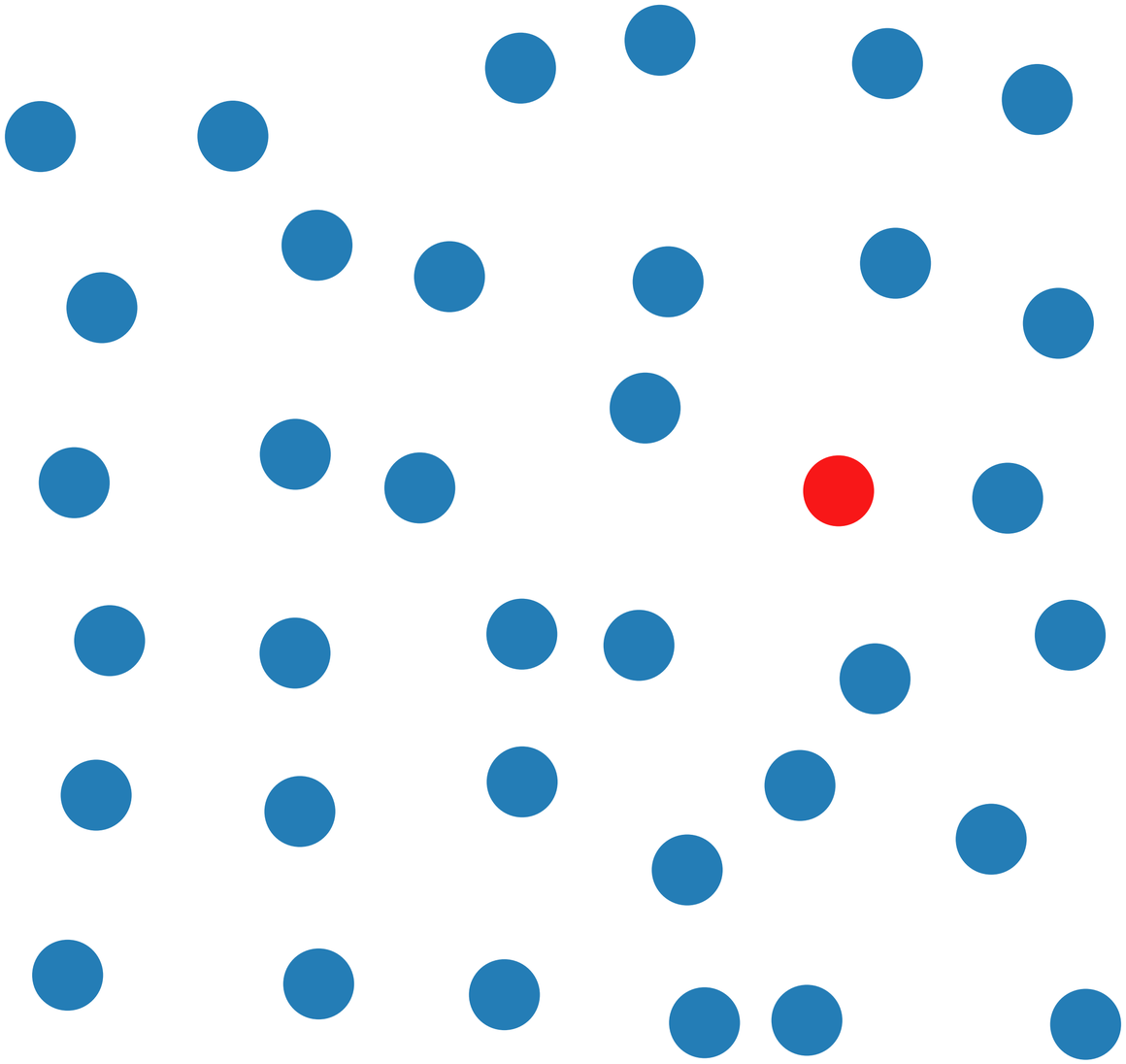




Mean of x	9
Variance of x	11
Mean of y	7.50
Variance of y	4.122
Correlation	0.816

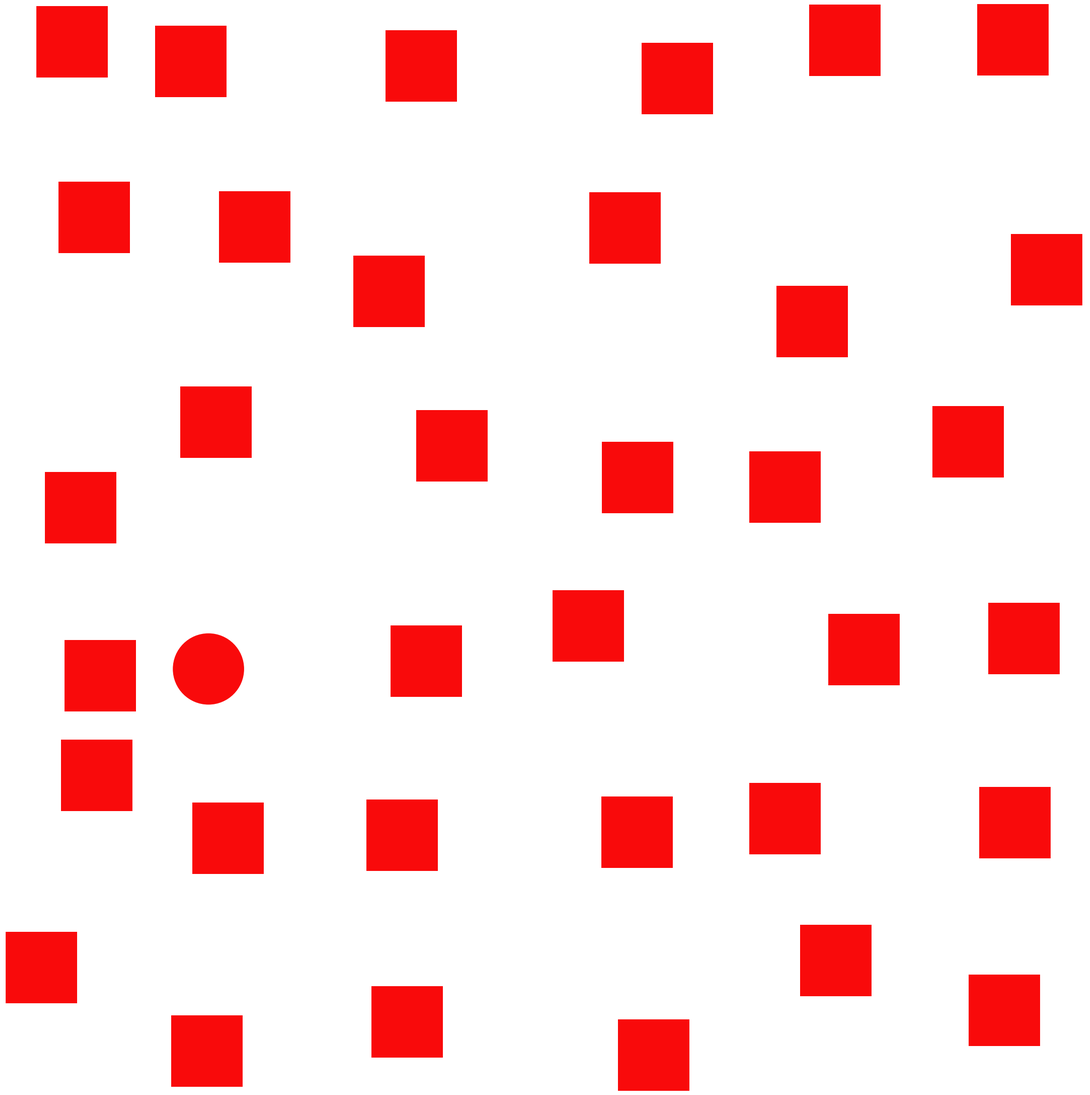
[F. J. Anscombe]

Visual Pop-out



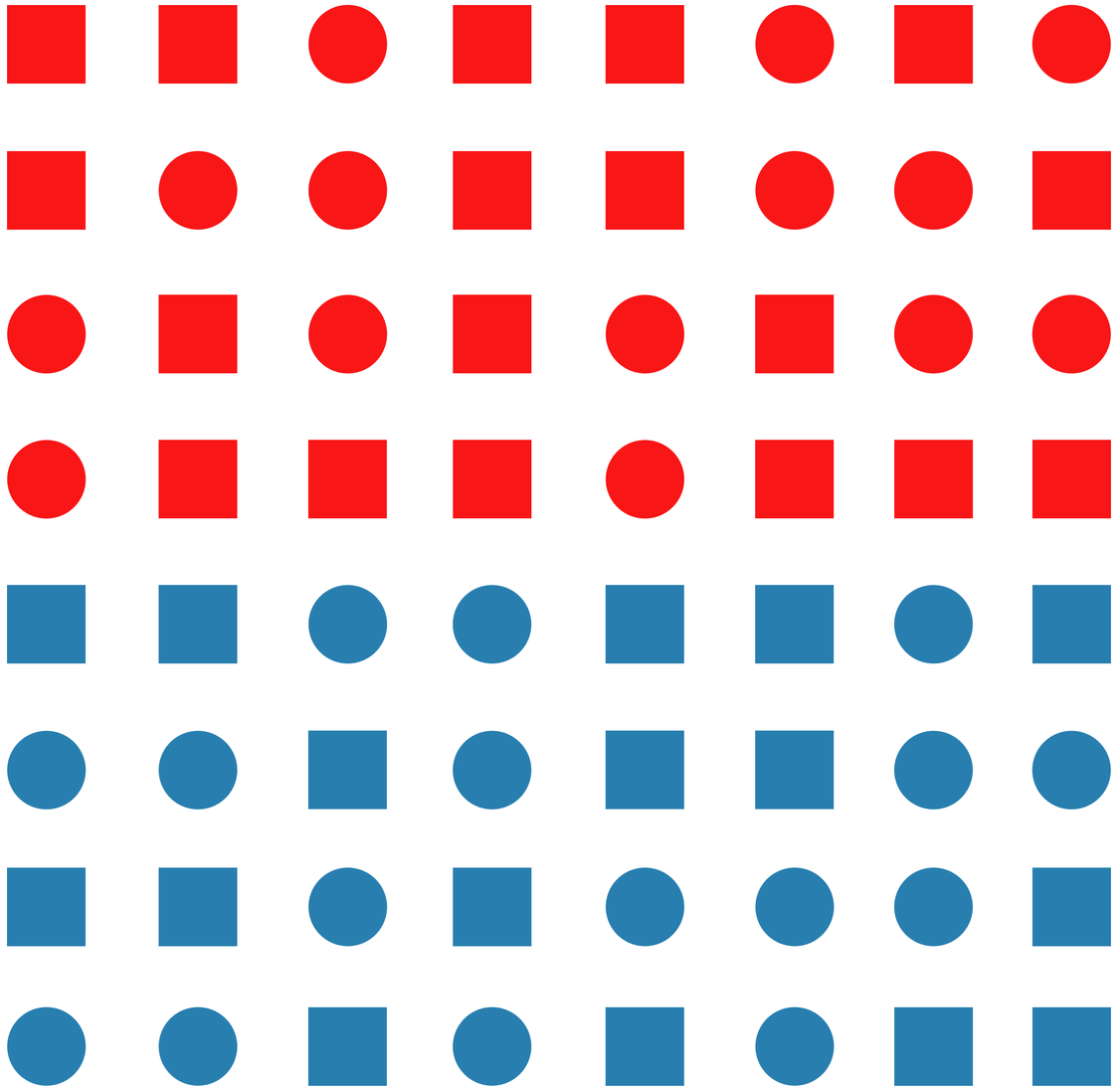
[C. G. Healey]

Visual Pop-out



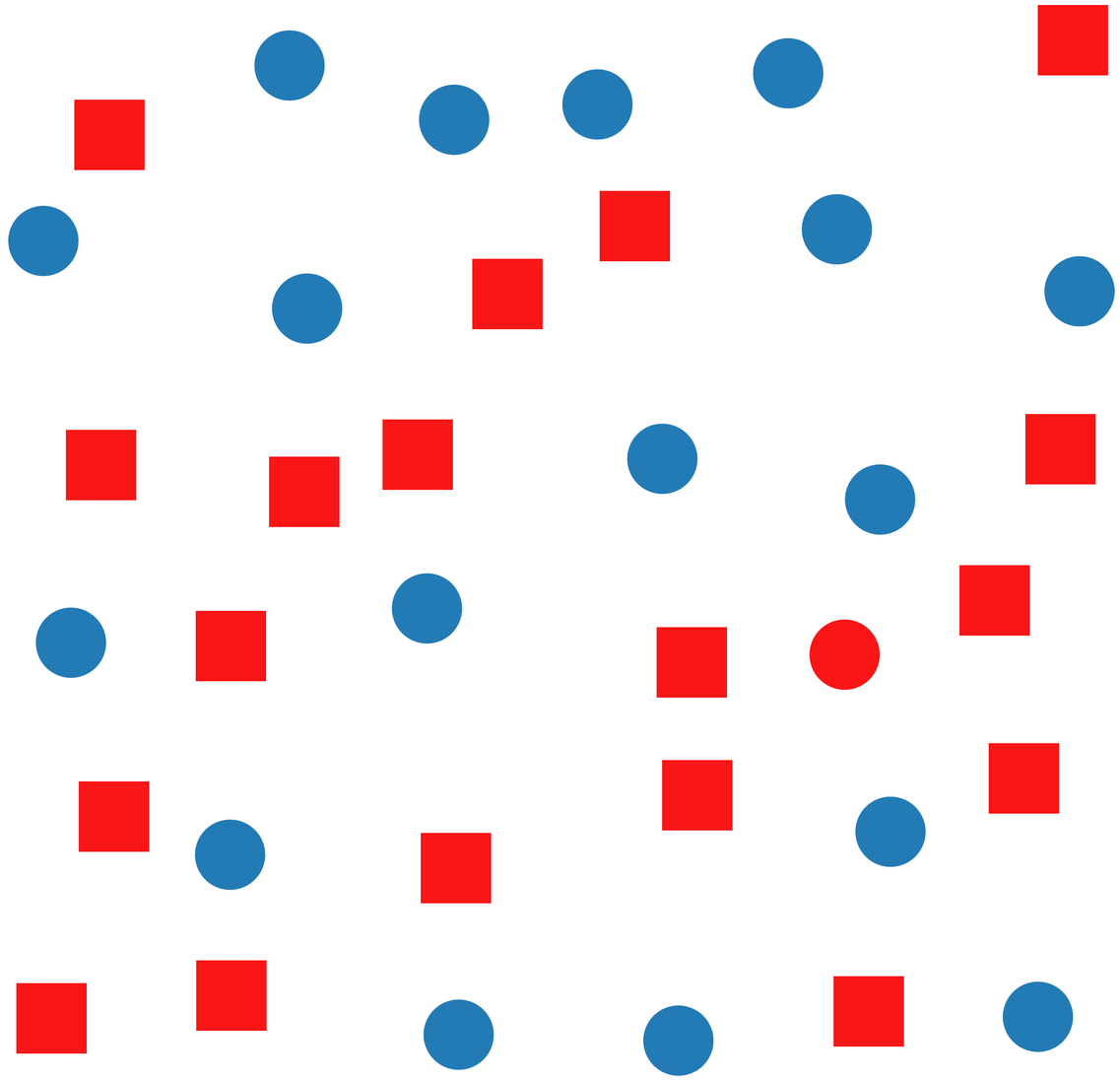
[C. G. Healey]

Visual Pop-out



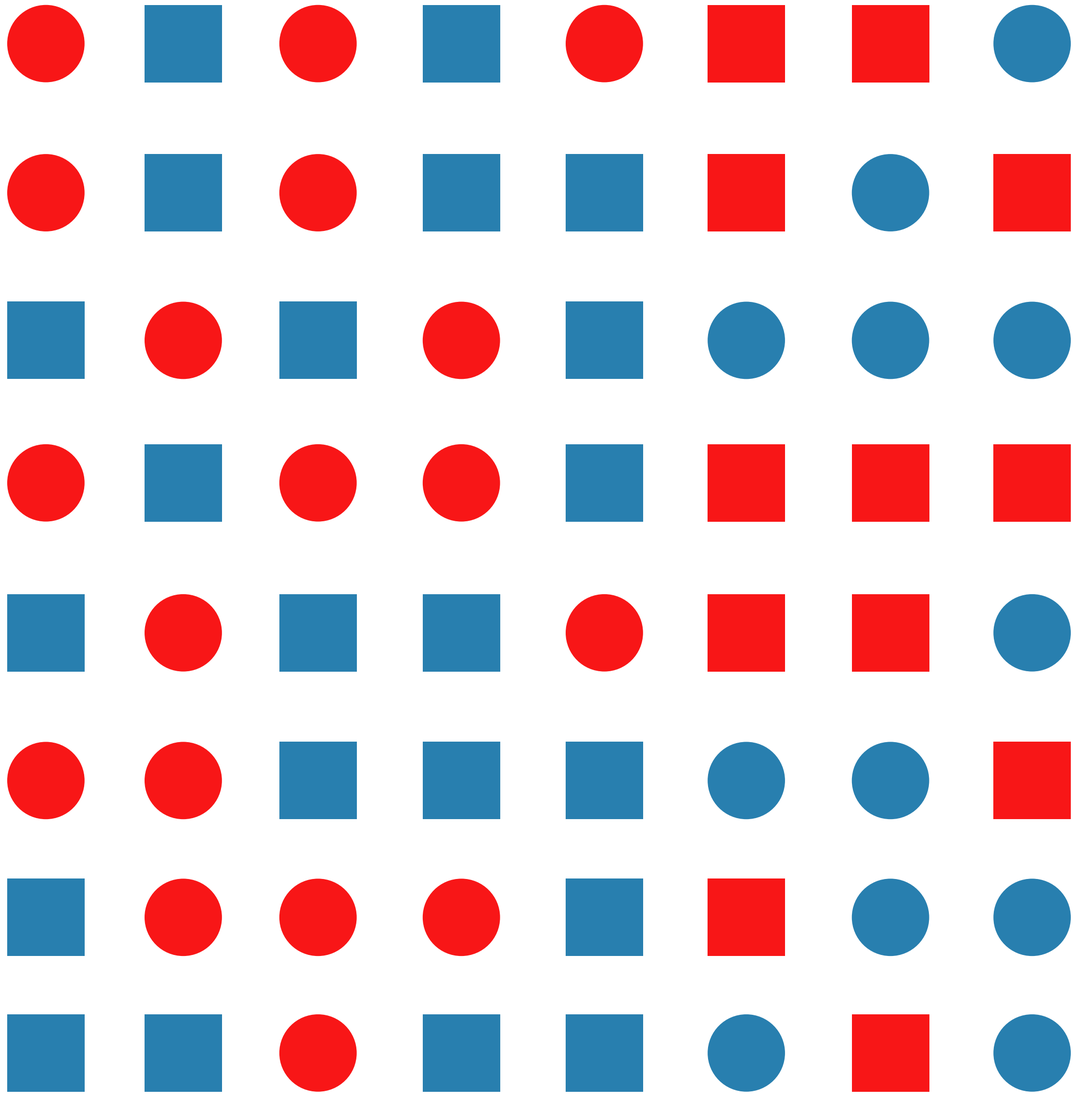
[C. G. Healey]

Visual Perception Limitations



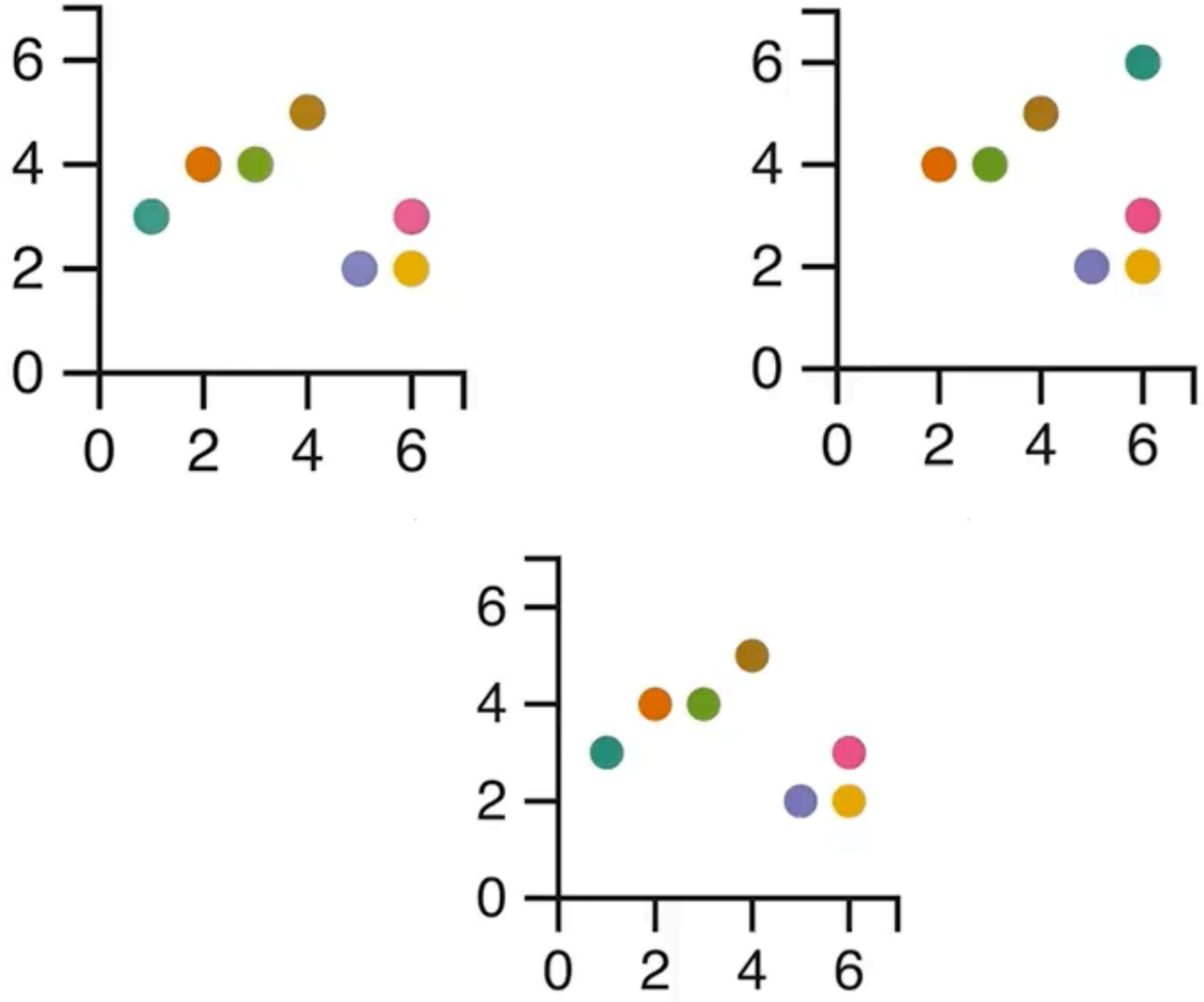
[C. G. Healey]

Visual Perception Limitations



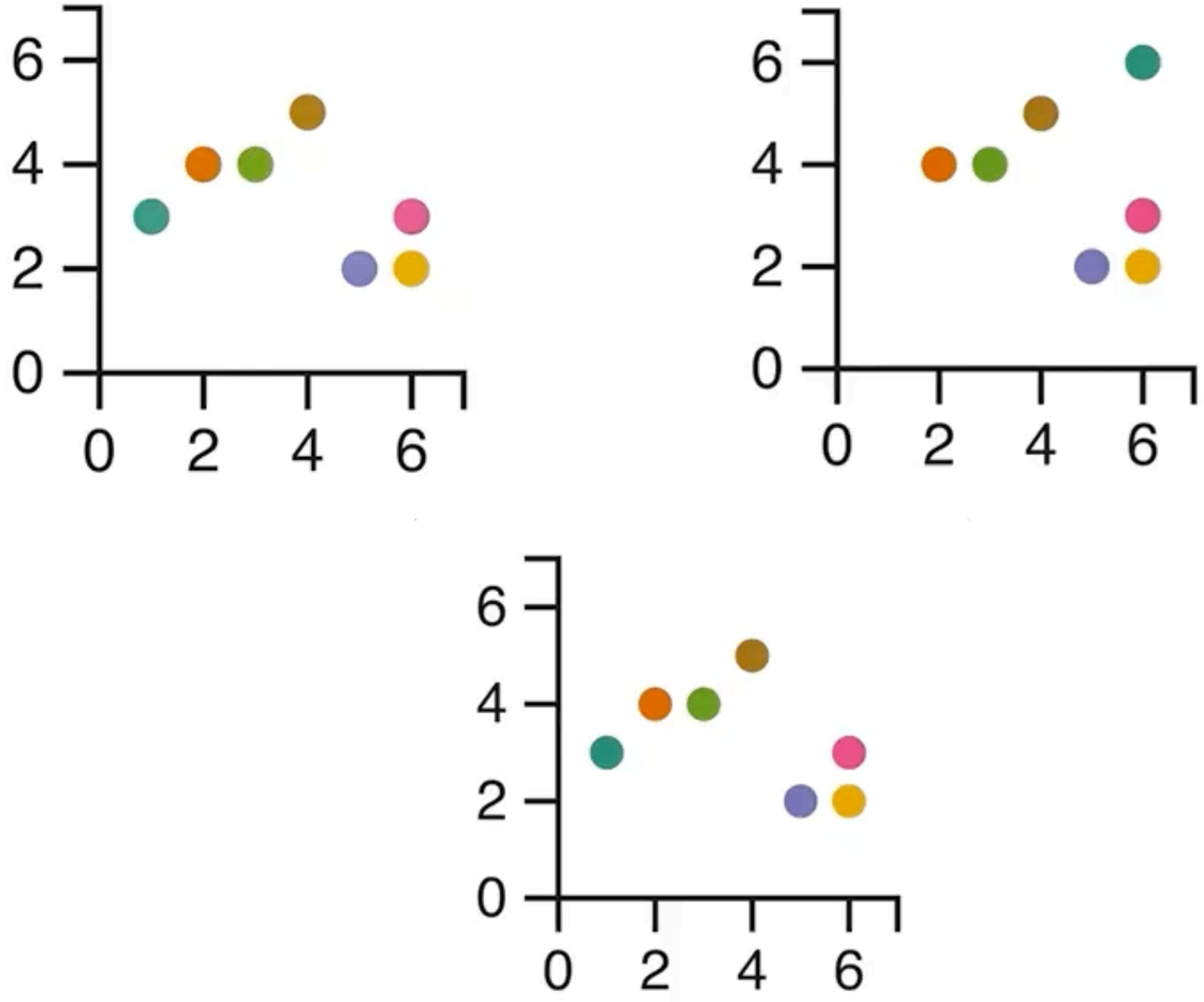
[C. G. Healey]

Animation Can Help



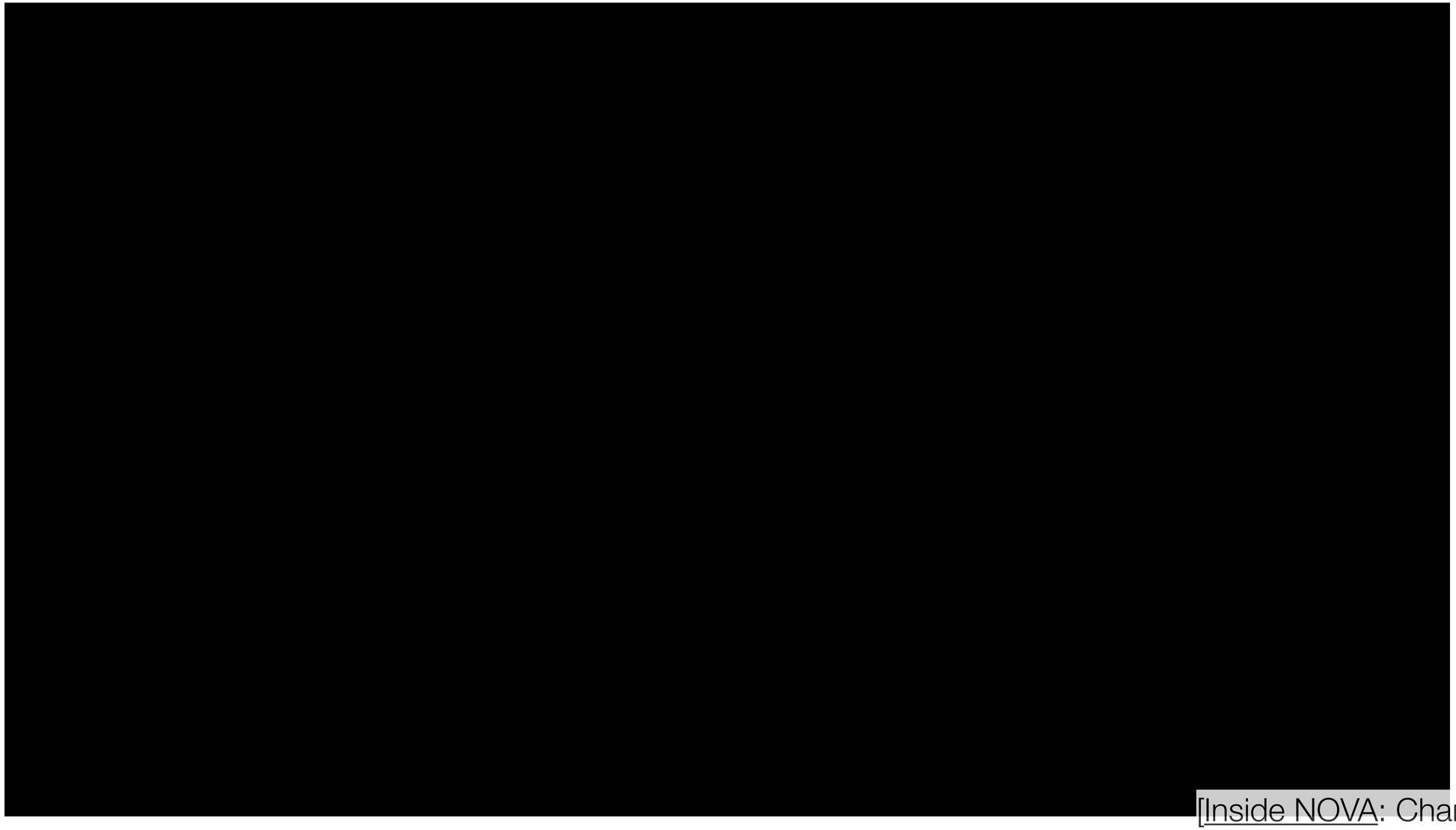
[A. Barrett via J. Cherdarchuk]

Animation Can Help



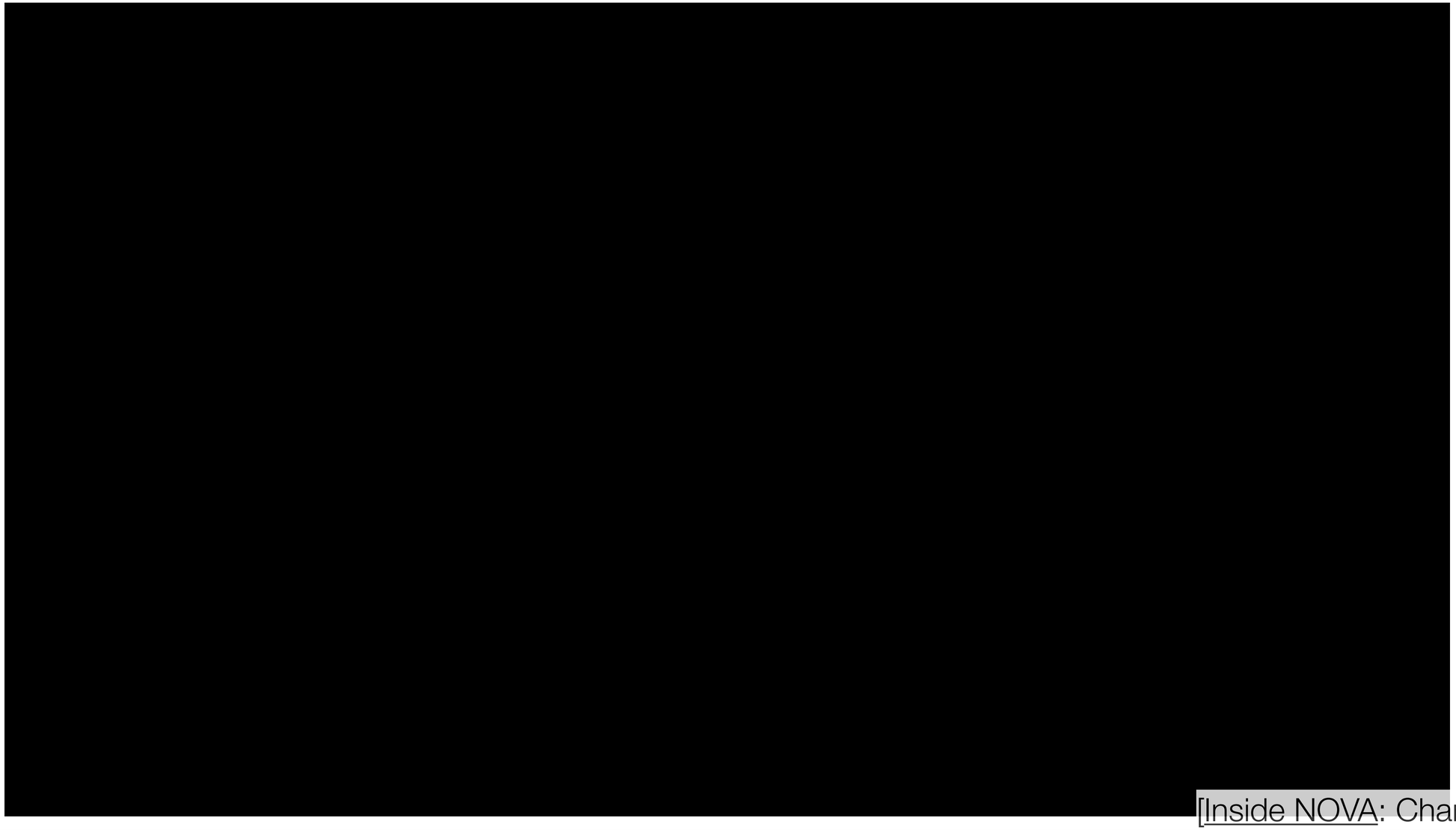
[A. Barrett via J. Cherdarchuk]

But…



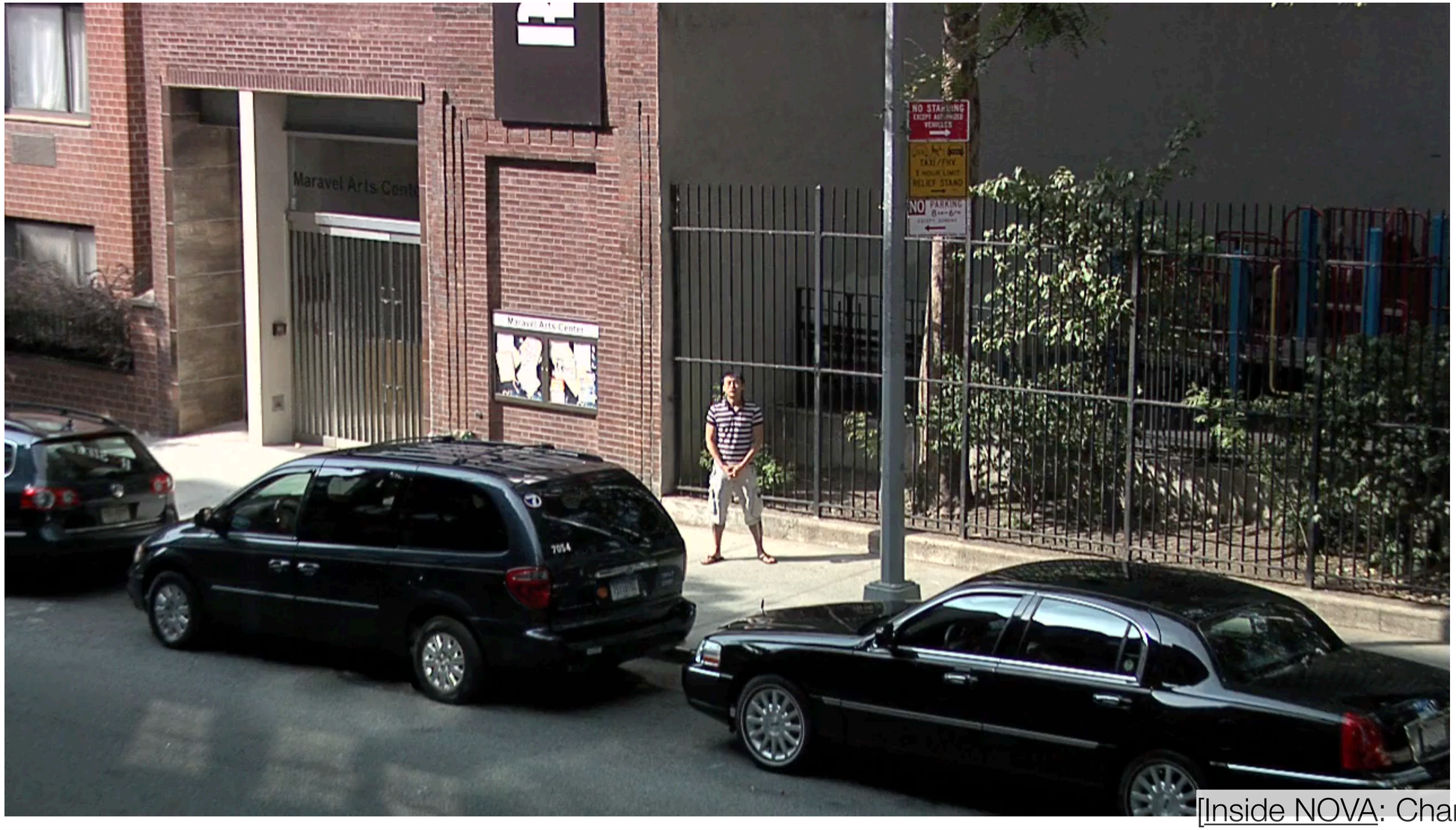
[Inside NOVA: Change Blindness]

But…



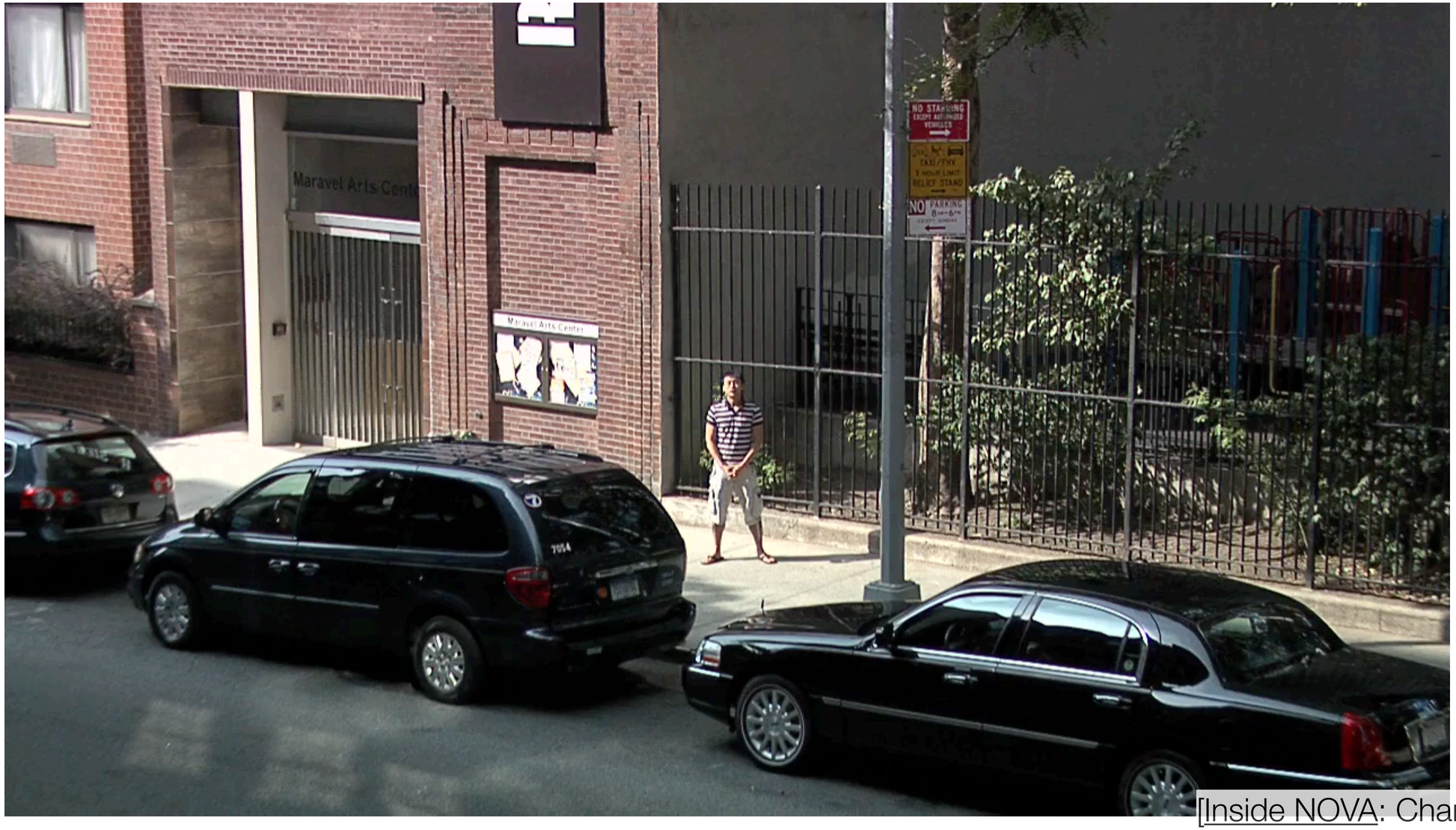
[Inside NOVA: Change Blindness]

Change & Inattentional Blindness Not Uncommon



[Inside NOVA: Change Blindness]

Change & Inattentional Blindness Not Uncommon



[Inside NOVA: Change Blindness]

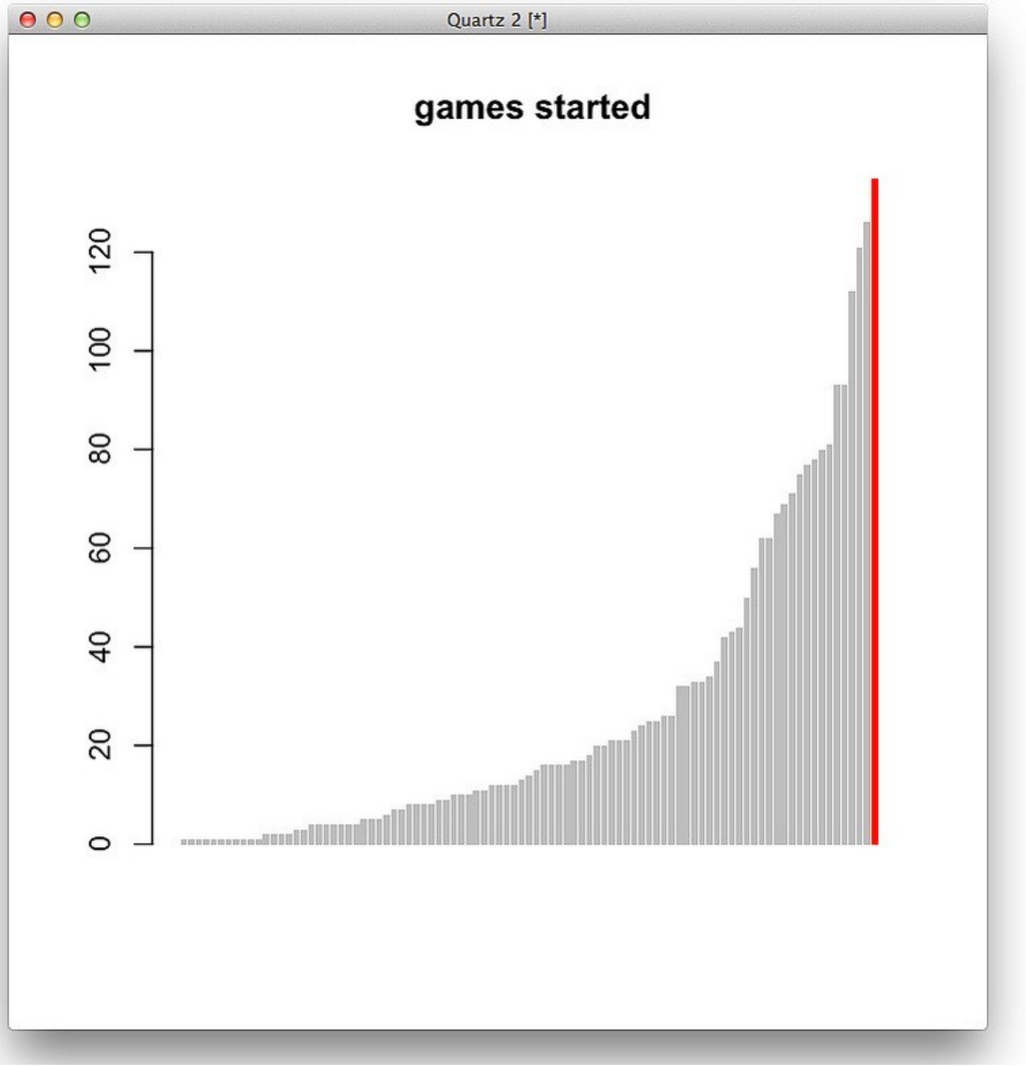
Other Human Limitations

- Visual working memory is **small**
- **Change blindness:** A failure to notice a change in our view
- **Inattentional blindness**: A failure to notice something else going on in our view while focusing on a particular task
- "The goal of vision is not to build a complete photograph or model of the world in your mind. The goal of vision is to make sense of the meaning of the world around you." - D. Simons

Definition

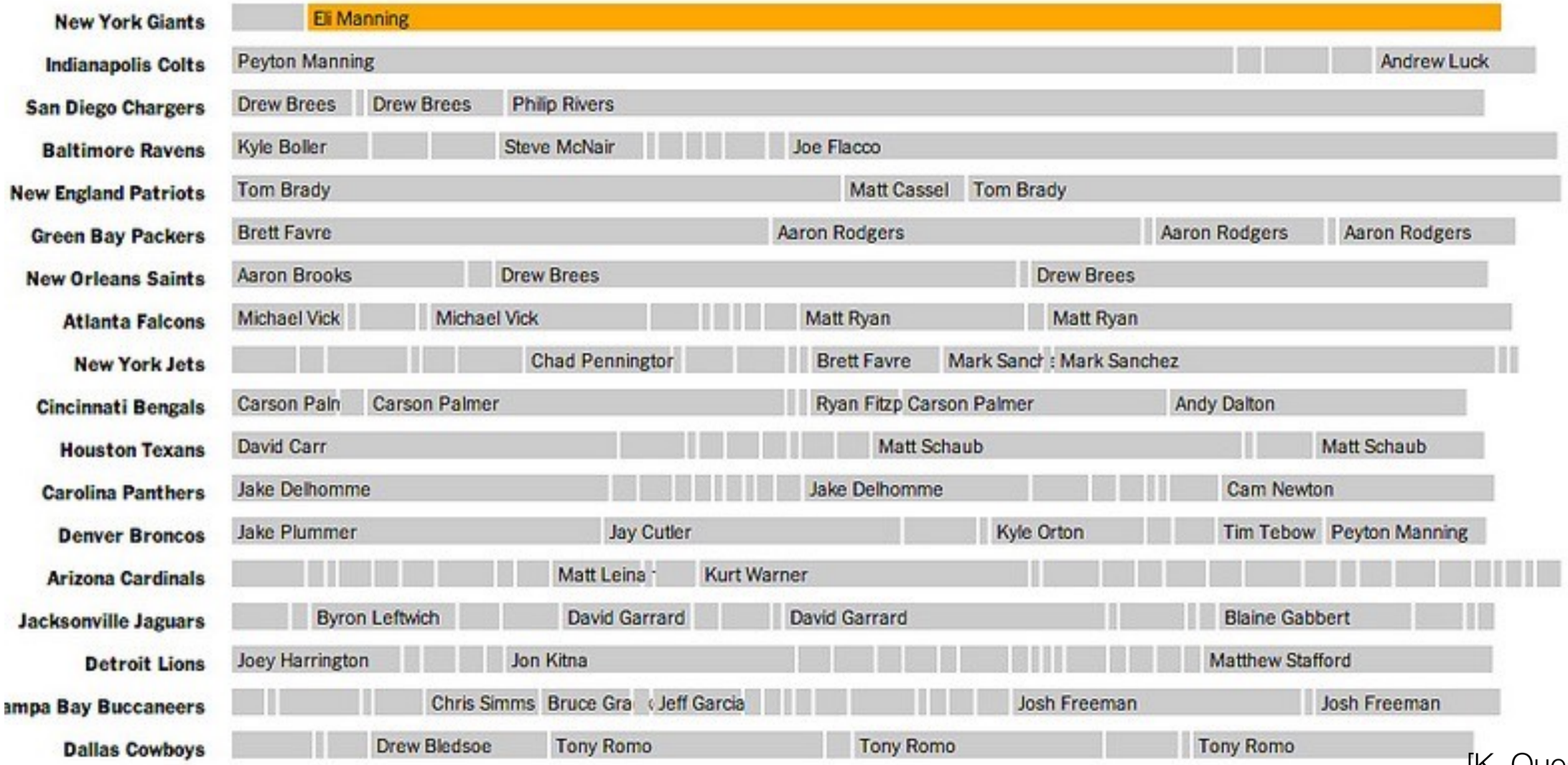
"Computer-based visualization systems provide visual representations of datasets **designed** to help people carry out tasks more effectively"

Design Process: Iteration



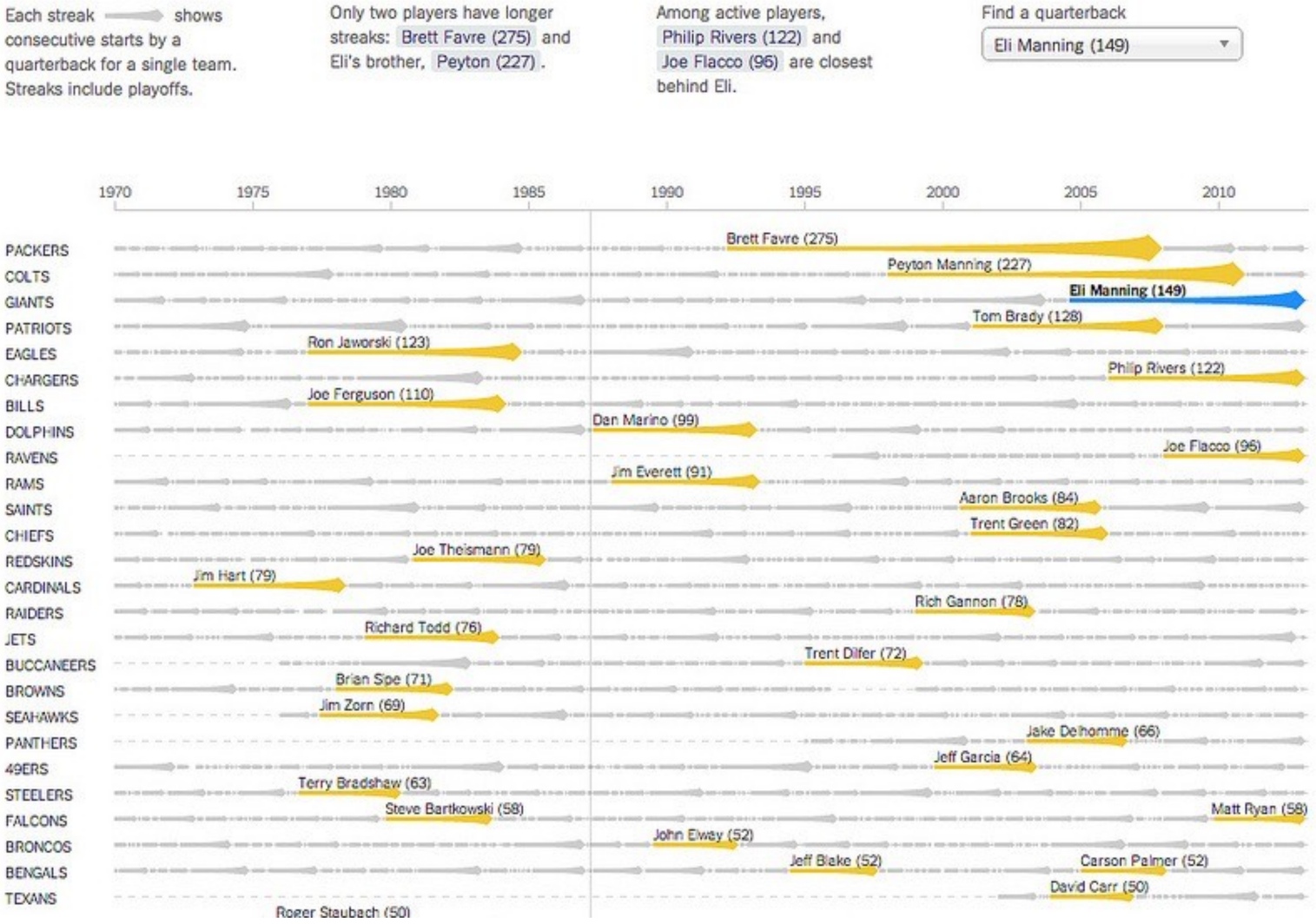
[K. Quealy, 2013]

Design Process: Iteration



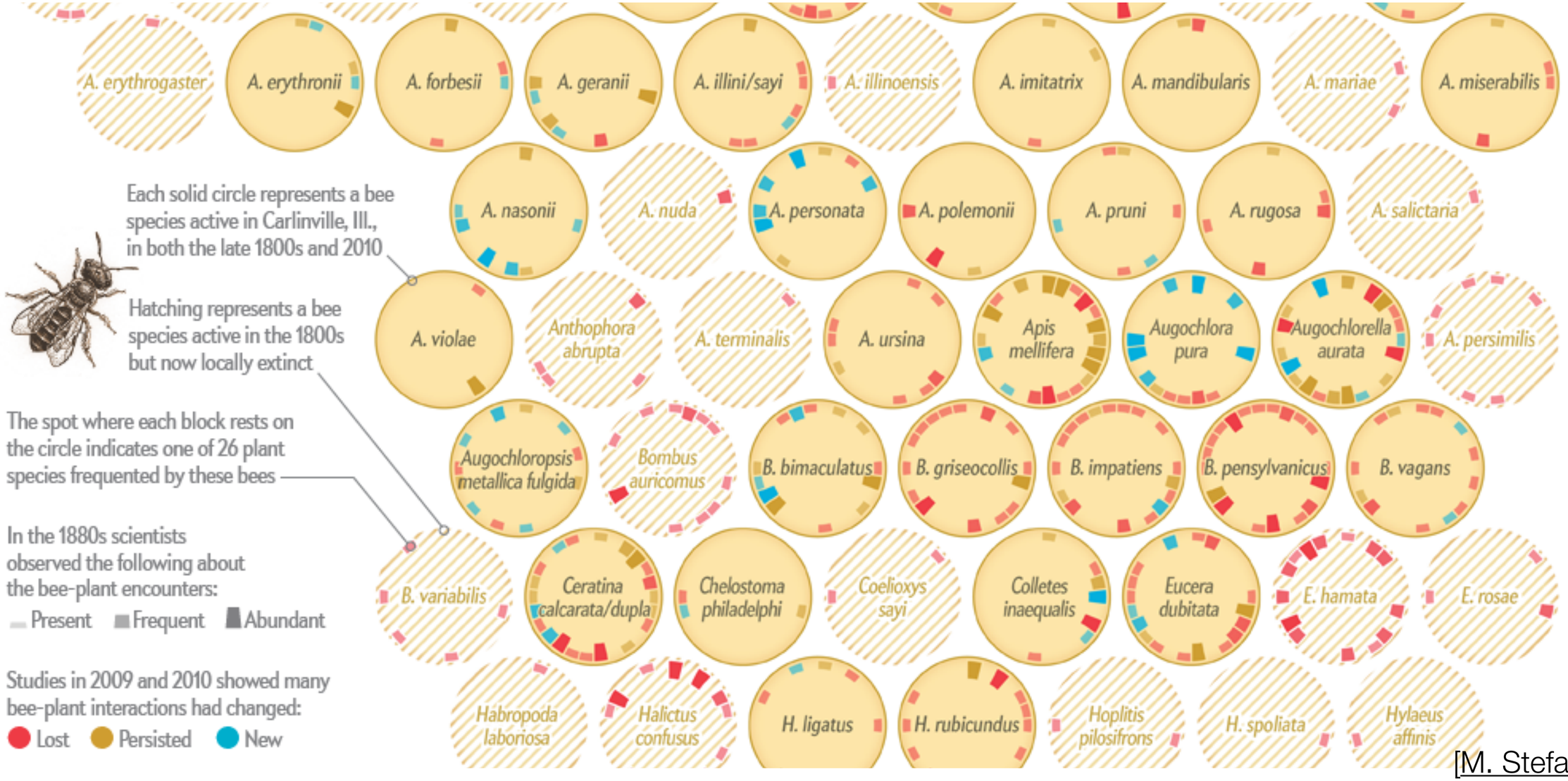
[K. Quealy, 2013]

Design Process: Iteration



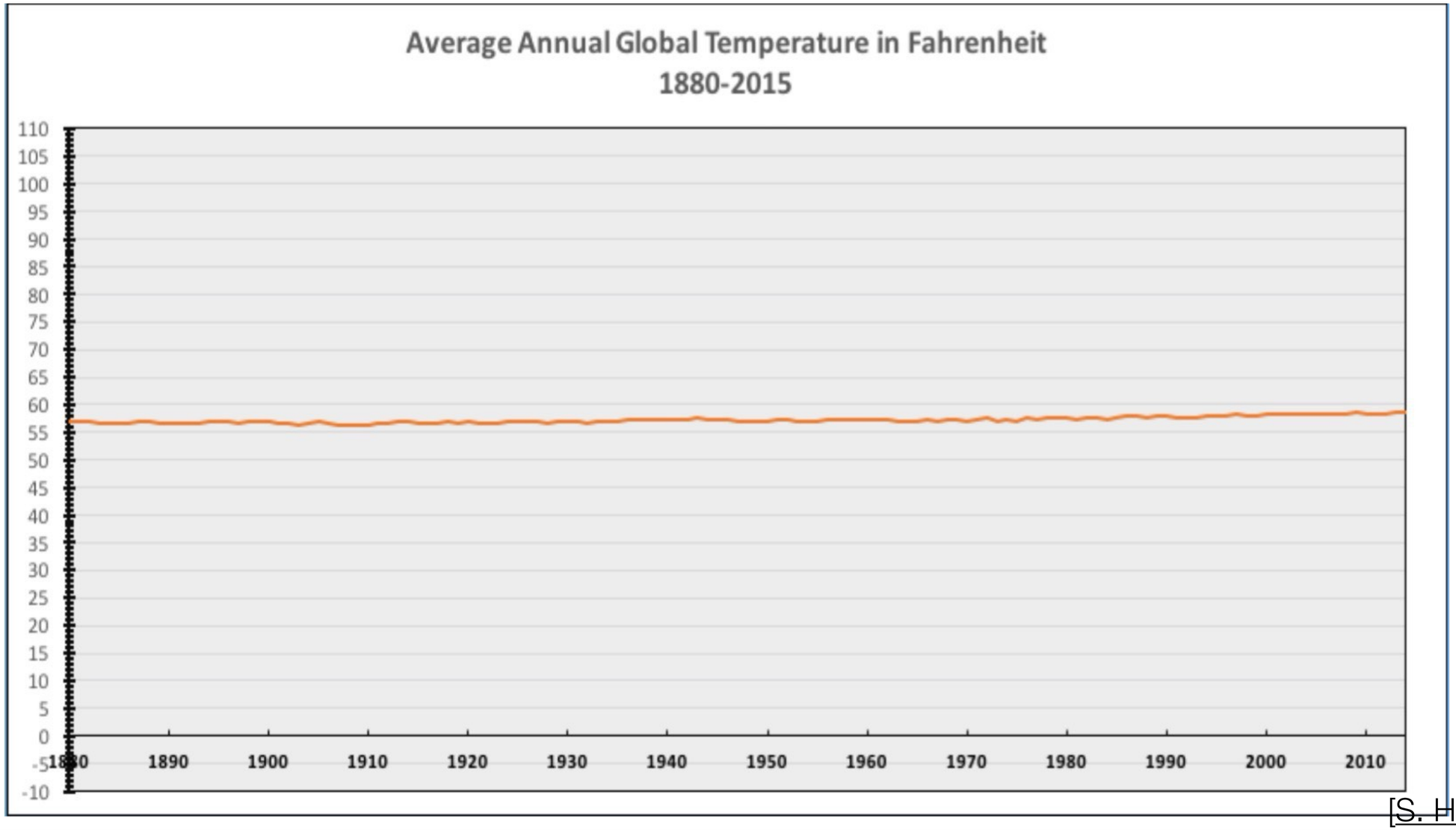
[K. Quealy, 2013]

Design Example



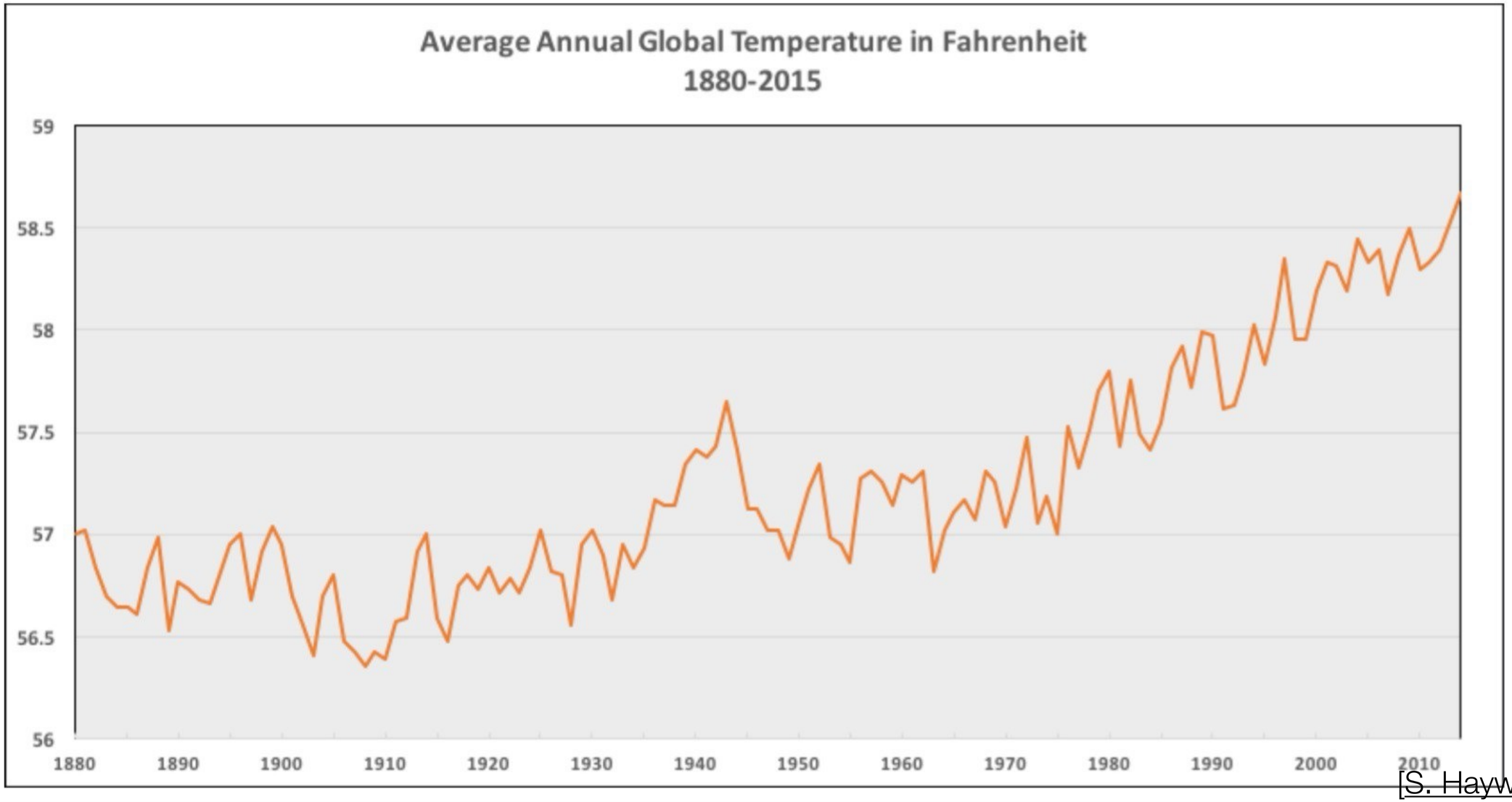
[M. Stefaner, 2013]

Impact of Design Choices: y-axis scale



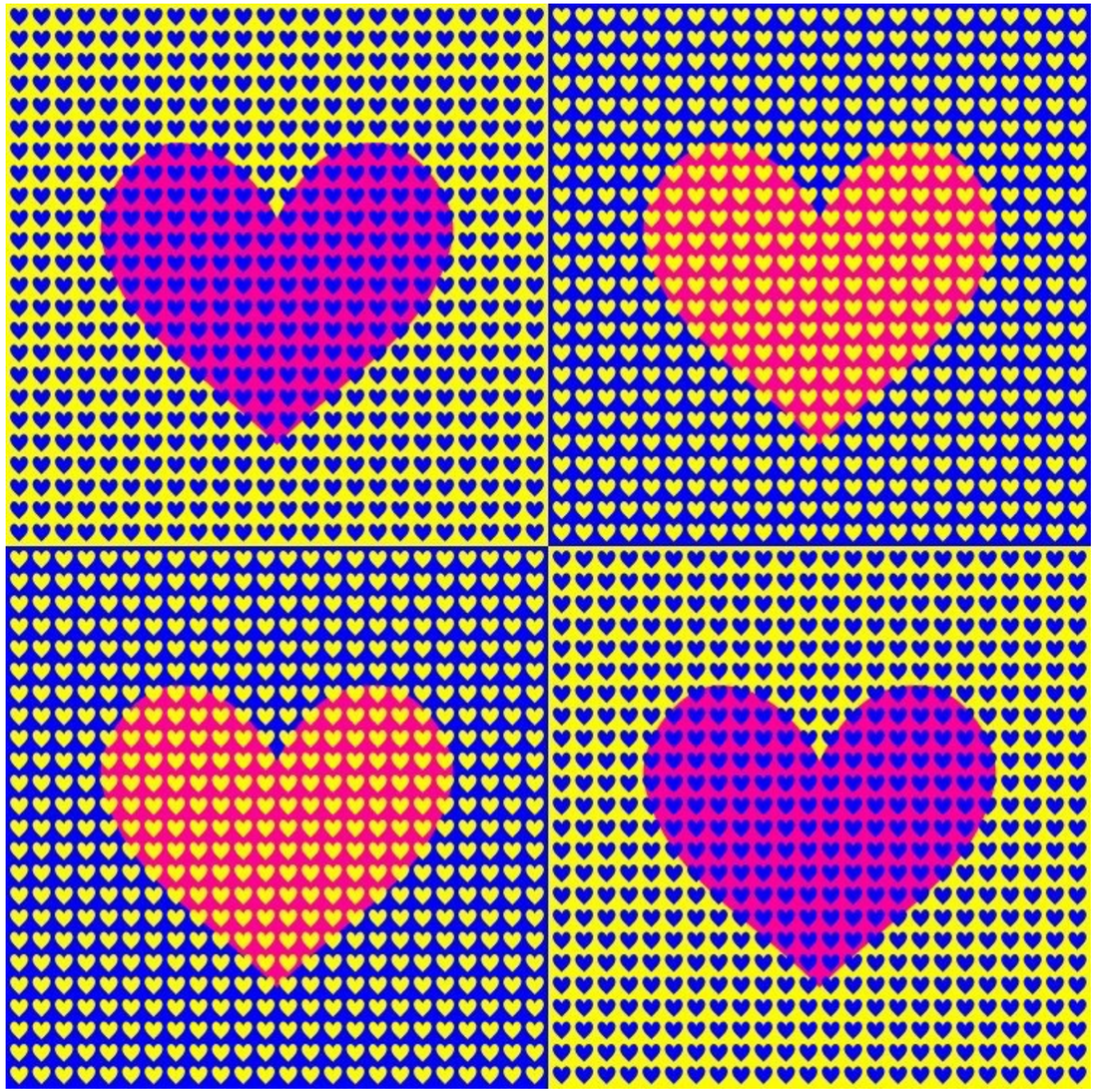
[S. Hayward, 2015]

Impact of Design Choices: y-axis scale



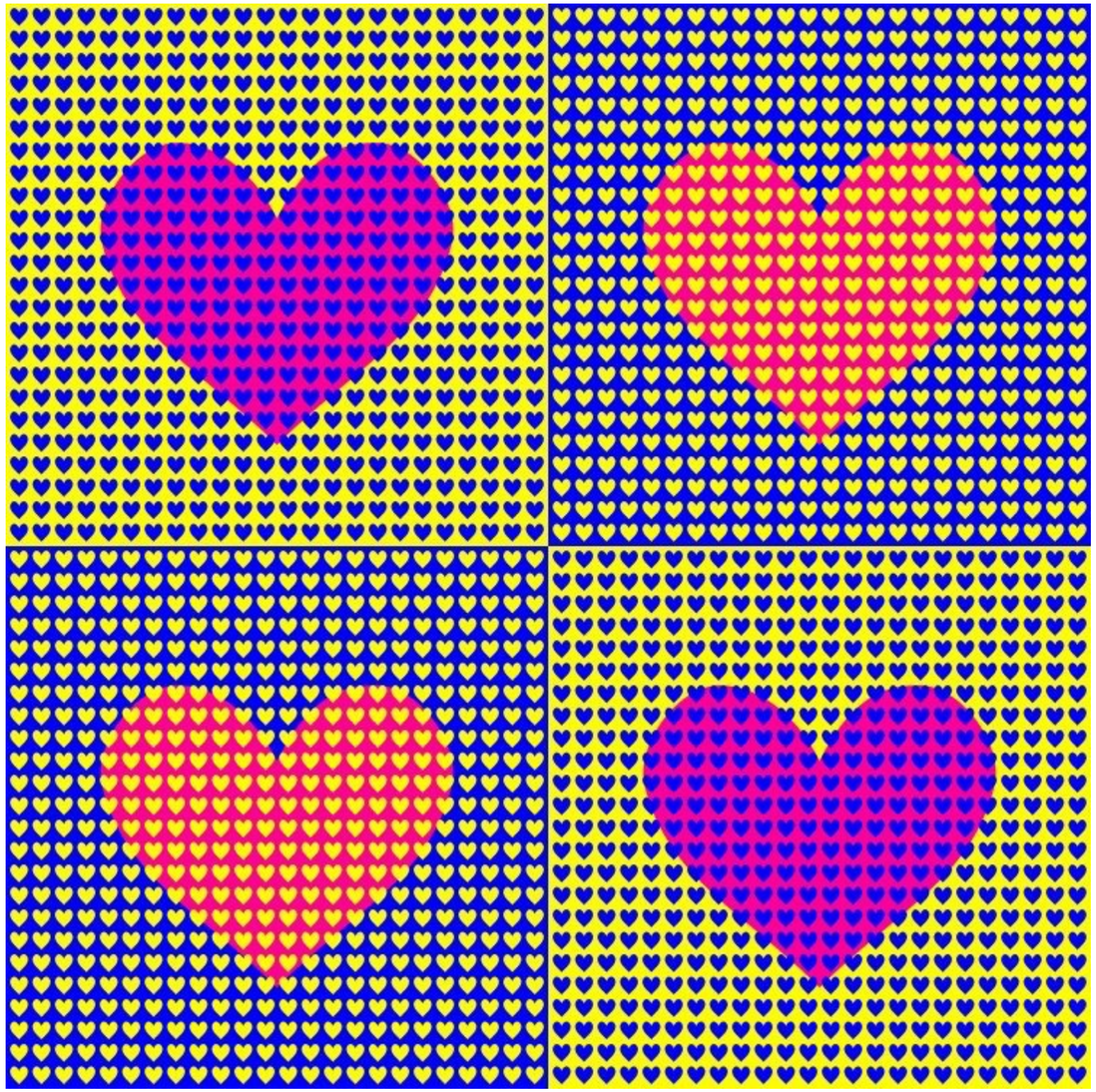
[S. Hayward, 2015]

Impact of Design Choices: color



[A. Kitaoka]

Impact of Design Choices: color

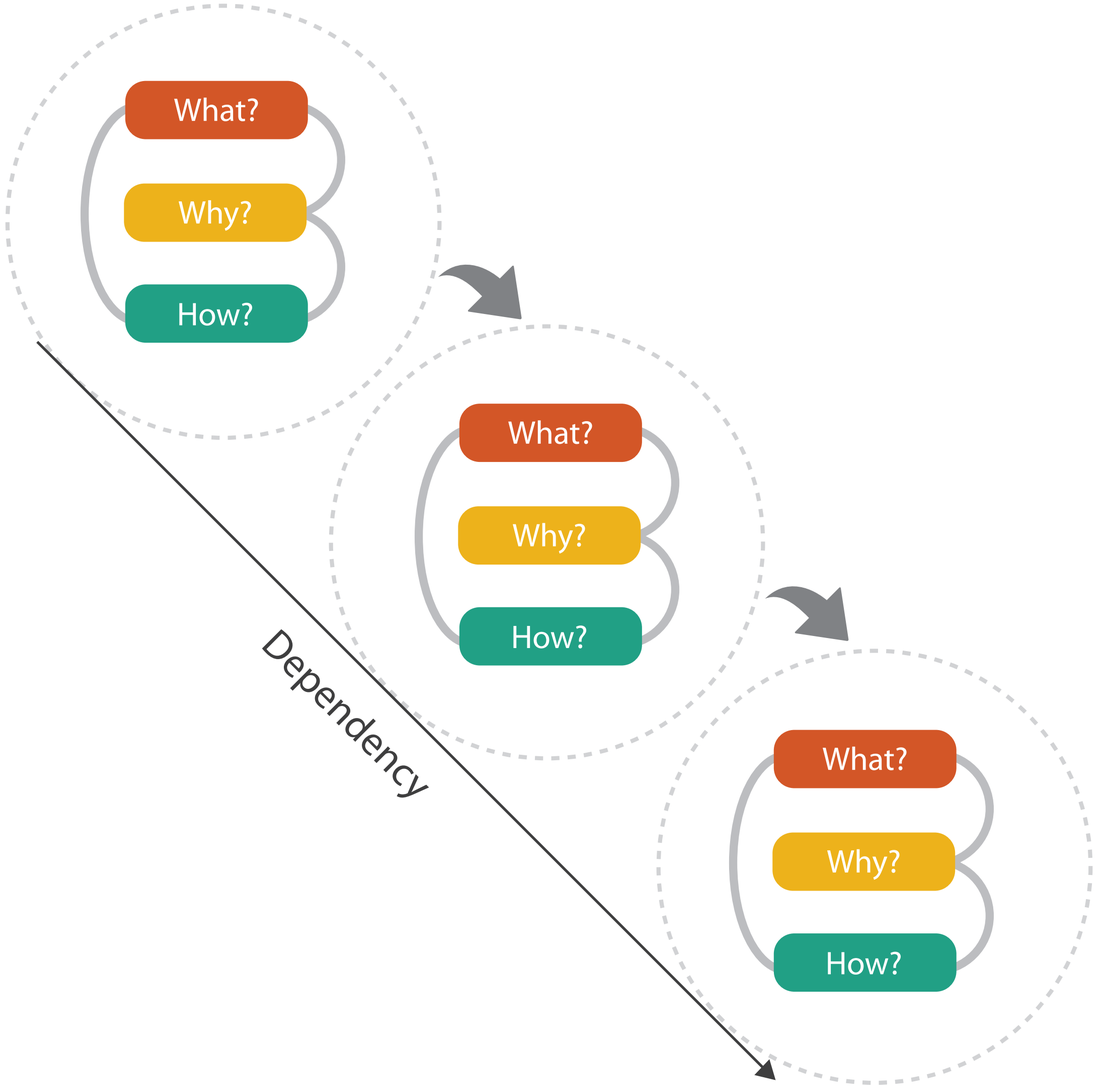


Red, yellow, blue

Purple, orange do not exist!

[A. Kitaoka]

Analyzing Visualizations



[Munzner (ill. Maguire), 2014]

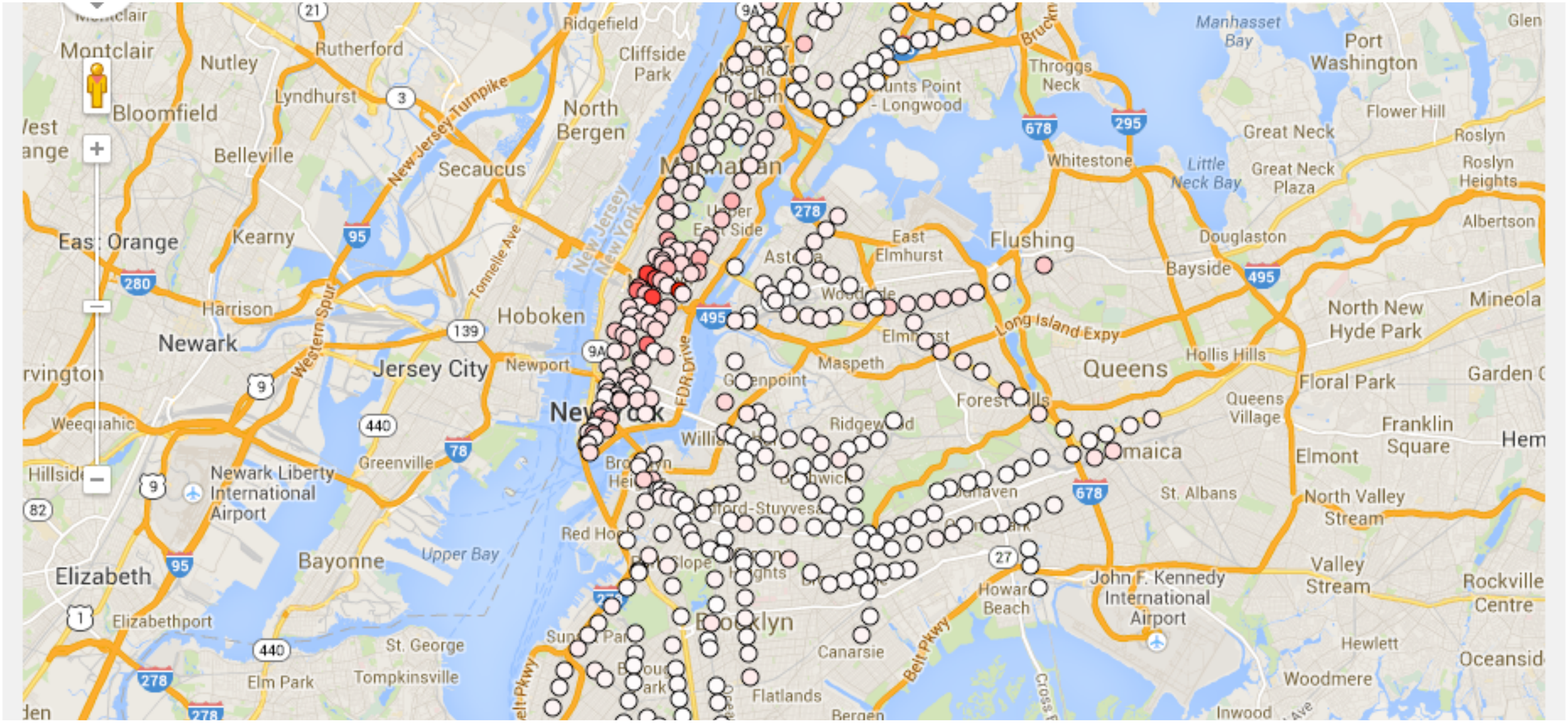
Data Exploration through Visualization

Transportation Data - NYC MTA

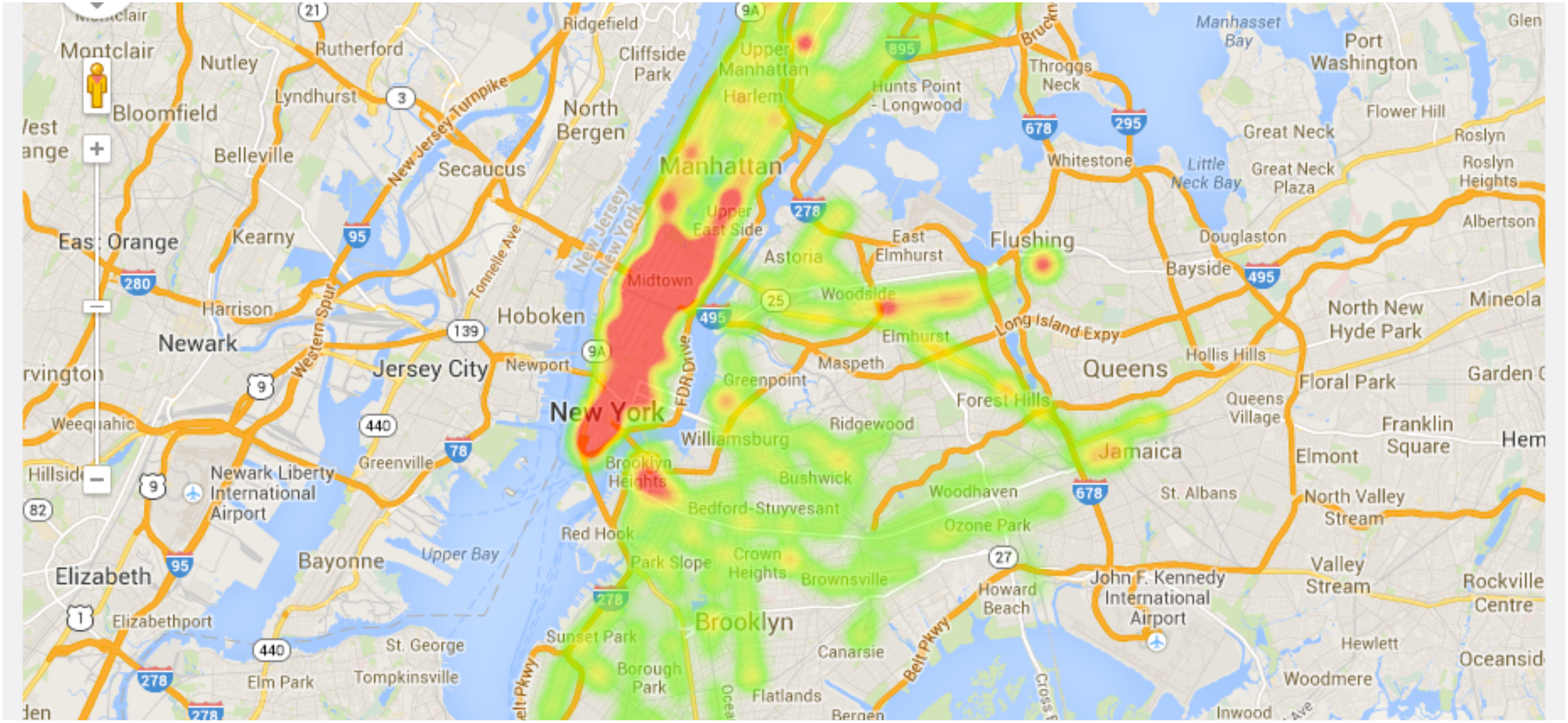
MTA Fare Data Exploration

	REMOTE	STATION	FF	SEN/DIS	7-D AFAS UNL	D AFAS/RMF U	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

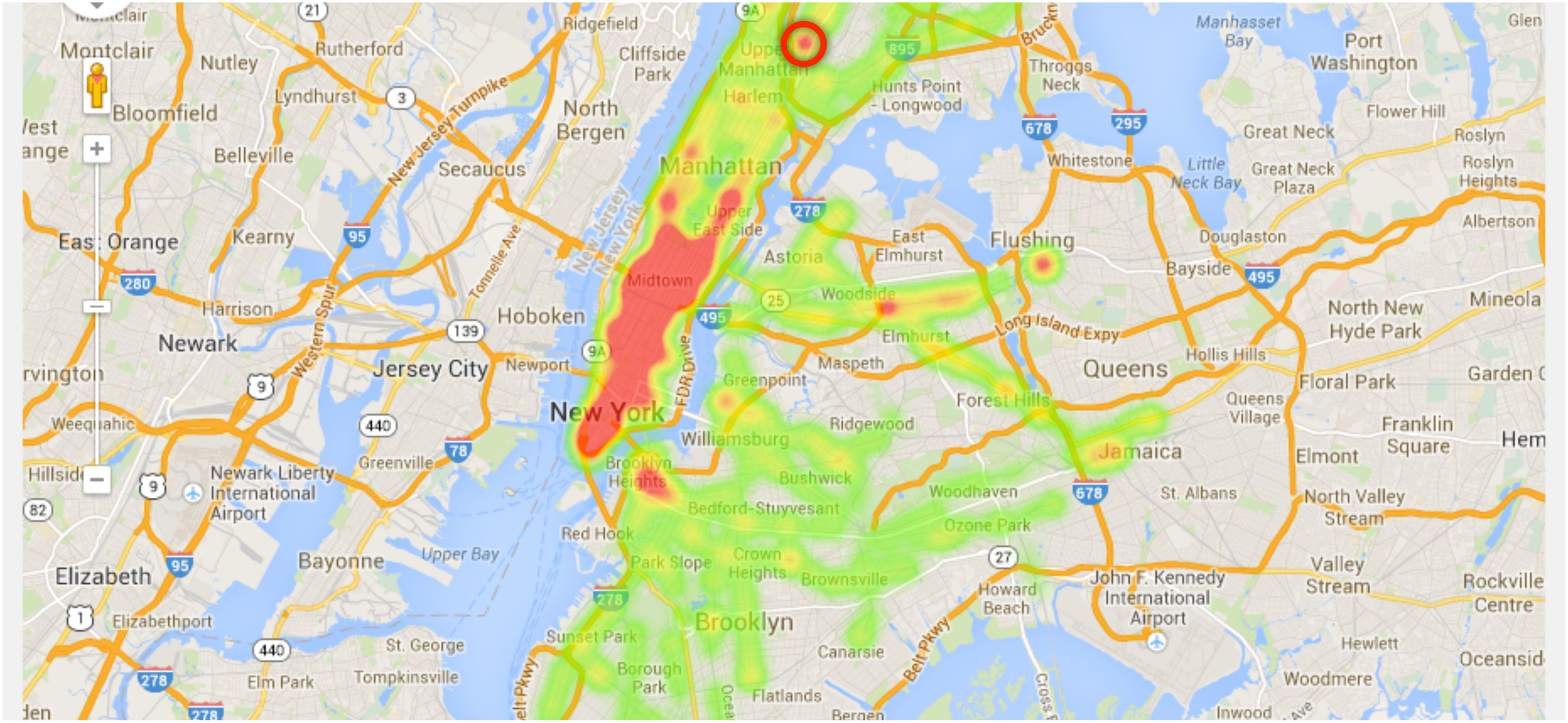
MTA Fare Data Exploration



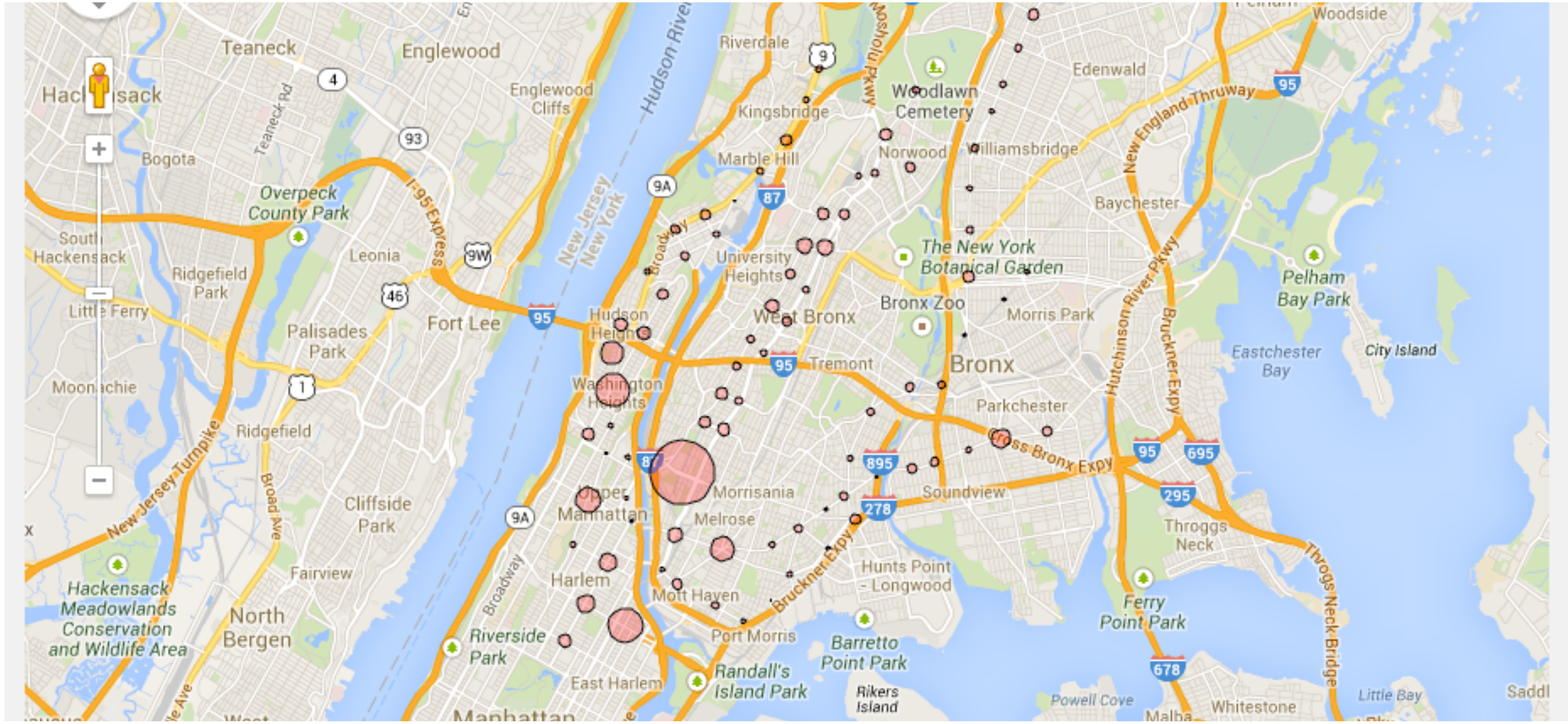
MTA Fare Data Exploration



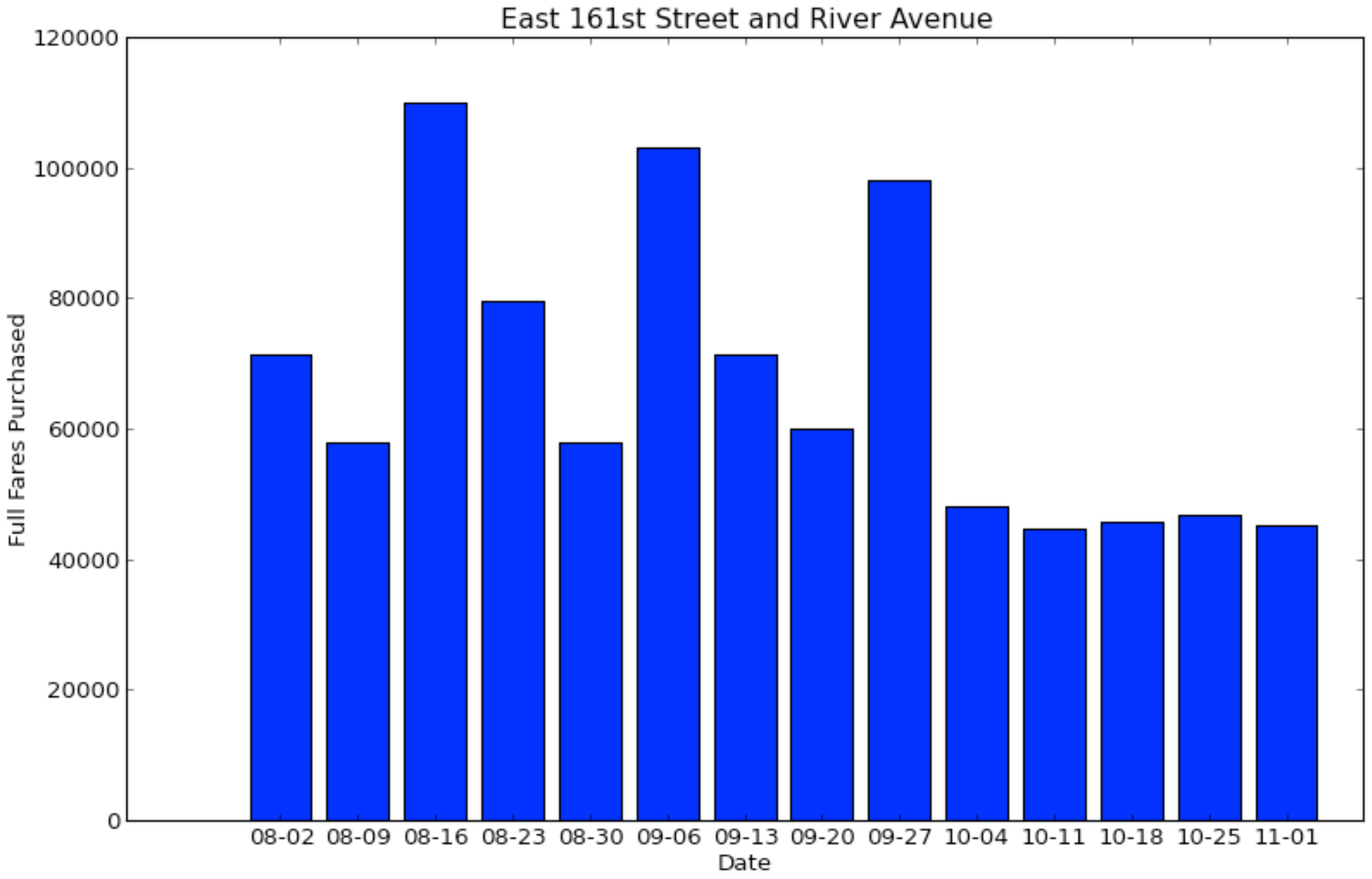
MTA Fare Data Exploration



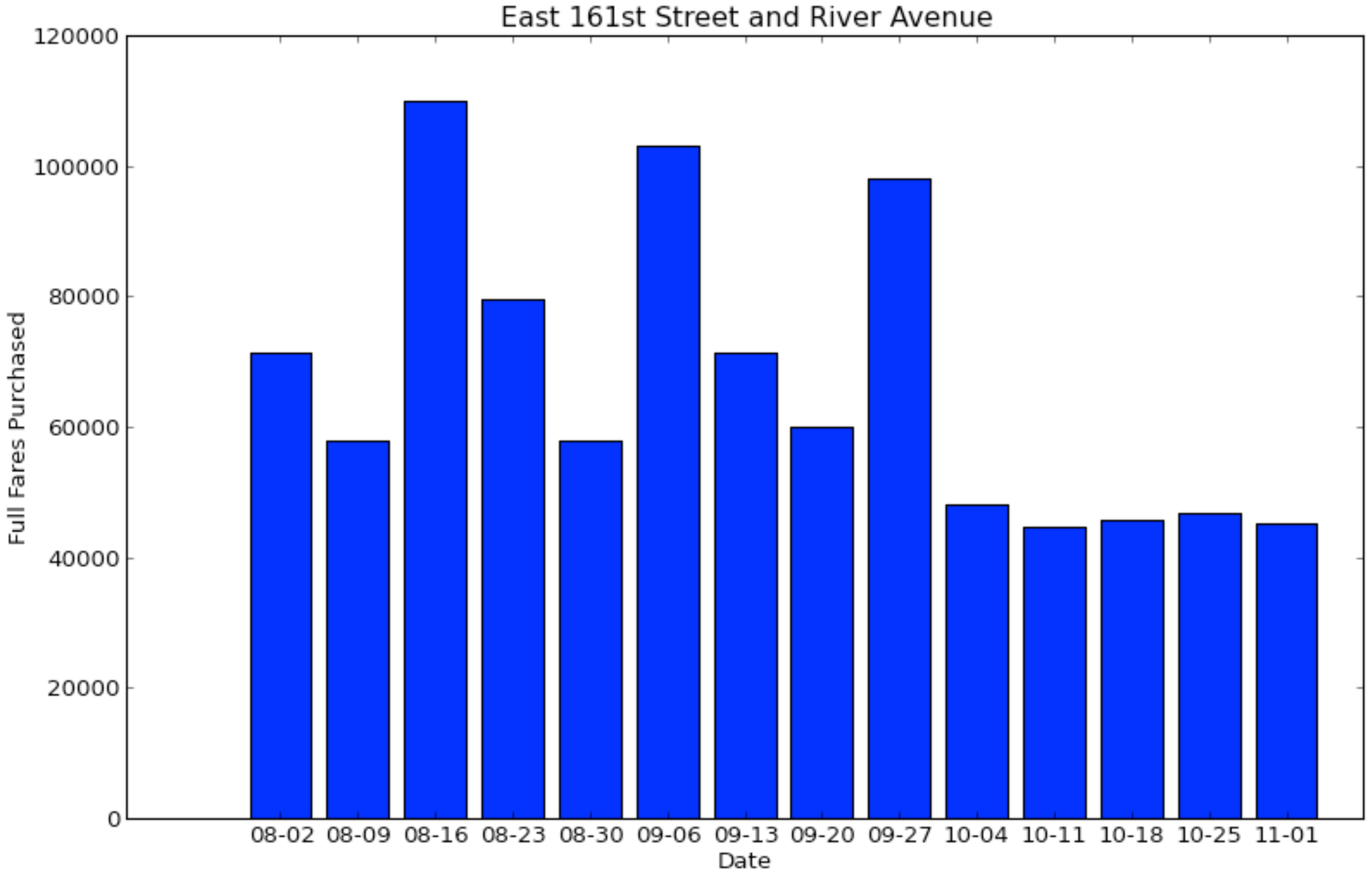
MTA Fare Data Exploration



MTA Fare Data Exploration



MTA Fare Data Exploration



Tools

- Desktop Applications:
  - Excel (see excelcharts.com)
  - Tableau
  - …
- Grammars:
  - Vega-Lite
- Programming Frameworks
  - d3.js
  - Observable Plot, plot.ly, deck.gl
  - …

- Tradeoffs
  - Speed
  - Customization
  - Understanding
  - Dissemination

Programming

- "Programming is blindly manipulating symbols." - B. Victor
- "Code is often the best tool we have because it is the most **general** tool we have; code has almost unlimited expressiveness" - M. Bostock
- You will write code in this class
  - Your assignments will involve code
  - Your project will involve code
- JavaScript is the language of the Web
  - Somewhat forgiving, not always the easiest to debug
  - Lots of references out there
  - A quickly-changing environment of frameworks

What languages do we use on the Web?

Languages of the Web

- HTML
- CSS
- SVG
- JavaScript
  - Versions of Javascript: ES6, ES2015, ES2020…
  - Specific frameworks: react, jQuery, bootstrap, D3

Web Programming Tools

- Basic: Text editor and Modern Browser
- Developer Tools: Built in to browsers (e.g. Chrome Developer Tools)
- Web Environments: Observable, CodePen, JSFiddle, etc.
- IDEs: WebStorm, VSCode

Observable

- Observable is a platform that allows users to create notebooks using JavaScript, Markdown, and HTML
- Strong support of data visualization (company and community interests)
- Introduction: A Taste of Observable
- Type markup/code,"execute" the cell, and results appear **above** the code
- Pin the cell to keep the code visible
- Can choose the type of cell (JavaScript, Markdown, or HTML)
- Can create an output (variable) in each cell that can be used in other cells
- Content is all **global scope**!

Other Platforms

- CodePen
- deno.land