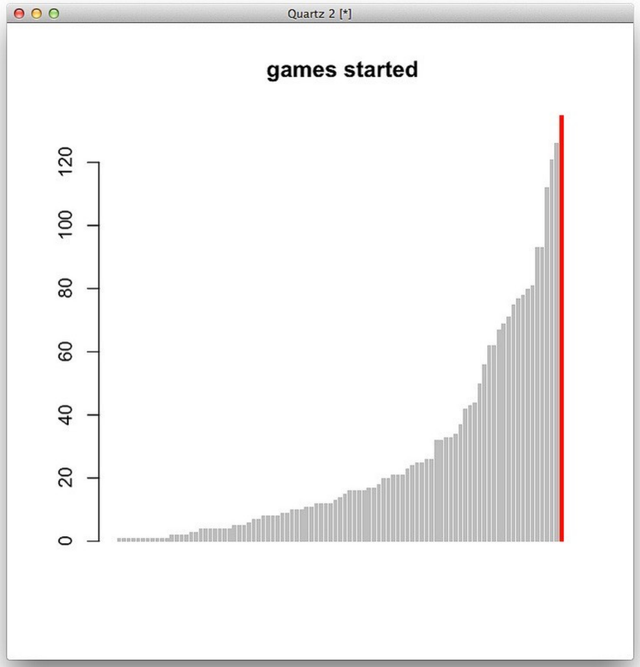
Data Visualization (CIS 490/680)

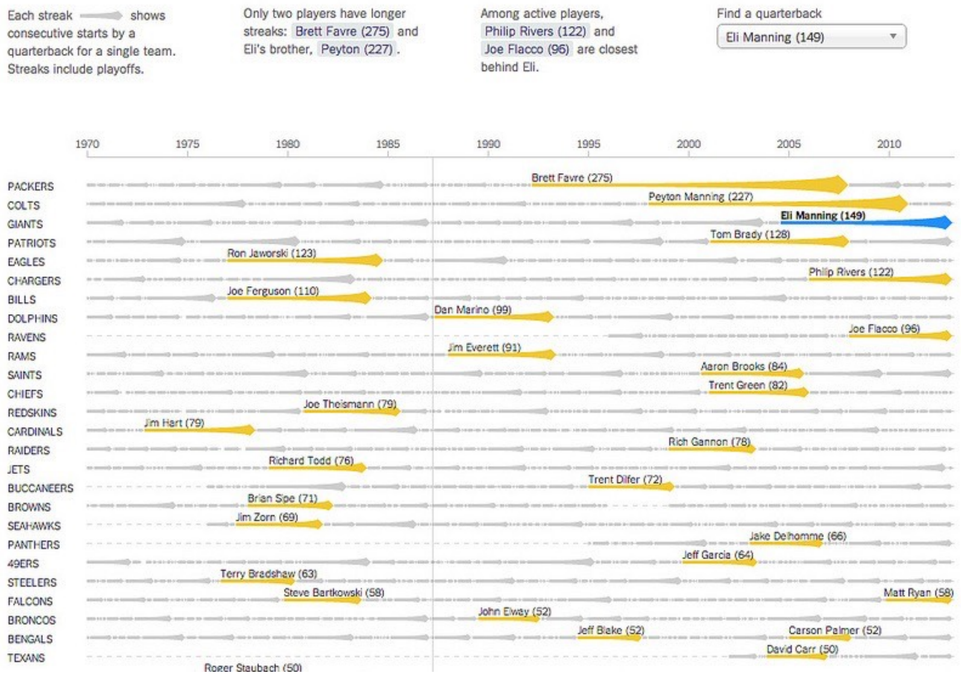
Definition & Web Programming

Dr. David Koop

Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team



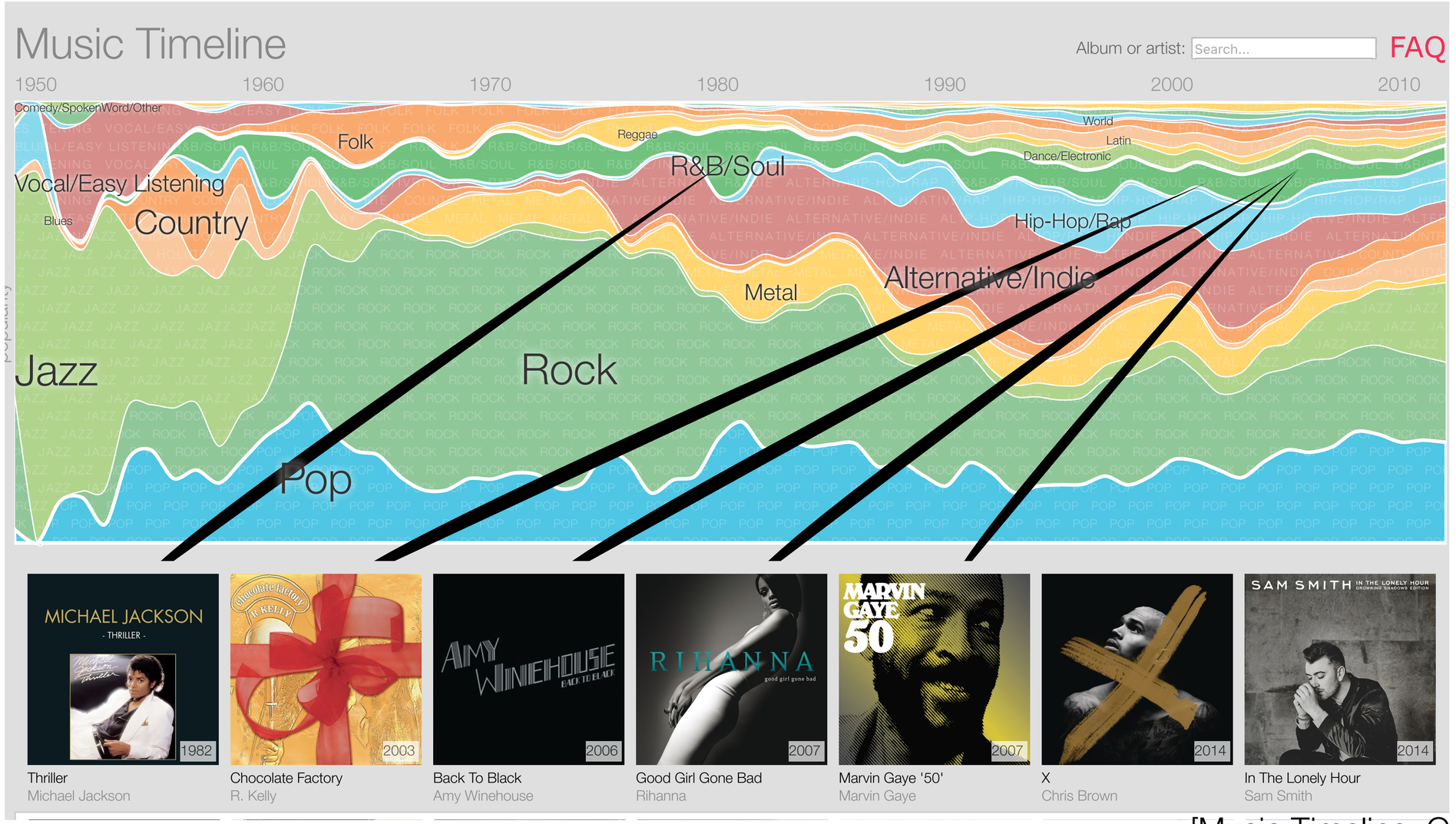


Exploration — Confirmation — Communication

Questions — **Answers/Persuasion**

[K. Quealy, 2013]

The Power of Interactive Visualization



[Music Timeline, Google Research]

Definition of Visualization

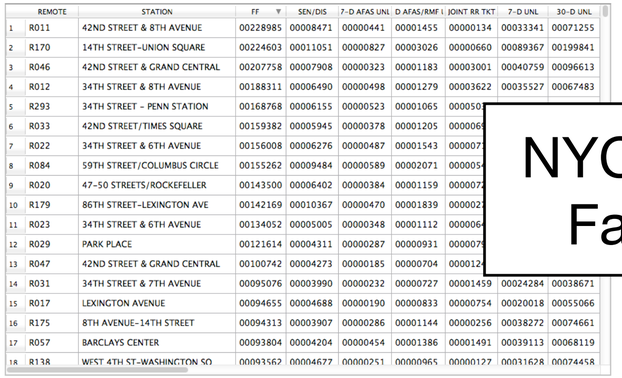
"Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively."

— 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

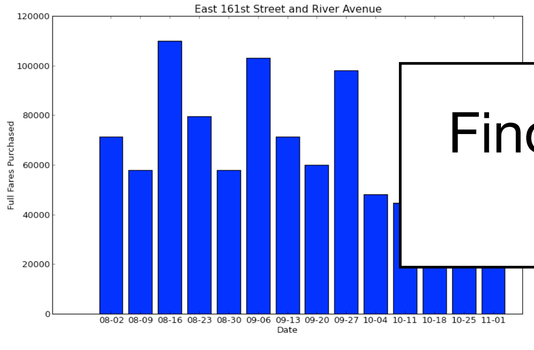
"Computer-based visualization systems provide visual representations of **datasets** designed to help people carry out **tasks** more effectively."

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	0000069		
7	R022	34TH STREET & 6TH AVENUE	00156008	00006276	00000487	00001543	0000071		
8	R084	59TH STREET/COLUMBUS CIRCLE	00155262	00009484	00000589	00002071	0000054		
9	R020	47-50 STREETS/ROCKEFELLER	00143500	00006402	00000384	00001159	0000072		
10	R179	86TH STREET-LEXINGTON AVE	00142169	00010367	00000470	00001839	0000022		
11	R023	34TH STREET & 6TH AVENUE	00134052	00005005	00000348	00001112	0000064		
12	R029	PARK PLACE	00121614	00004311	00000287	00000931	0000079		
13	R047	42ND STREET & GRAND CENTRAL	00100742	00004273	00000185	00000704	0000124		
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

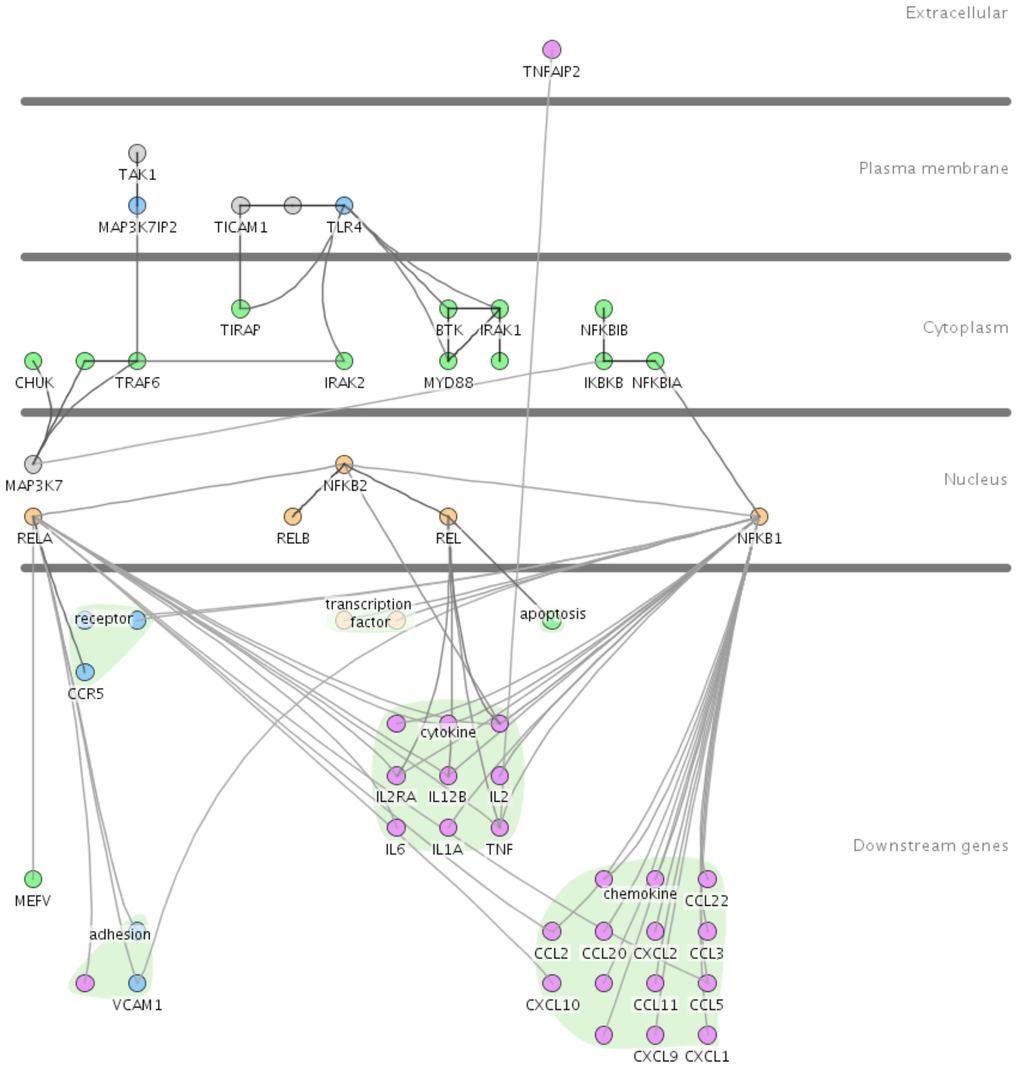
Definition

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

Why People?

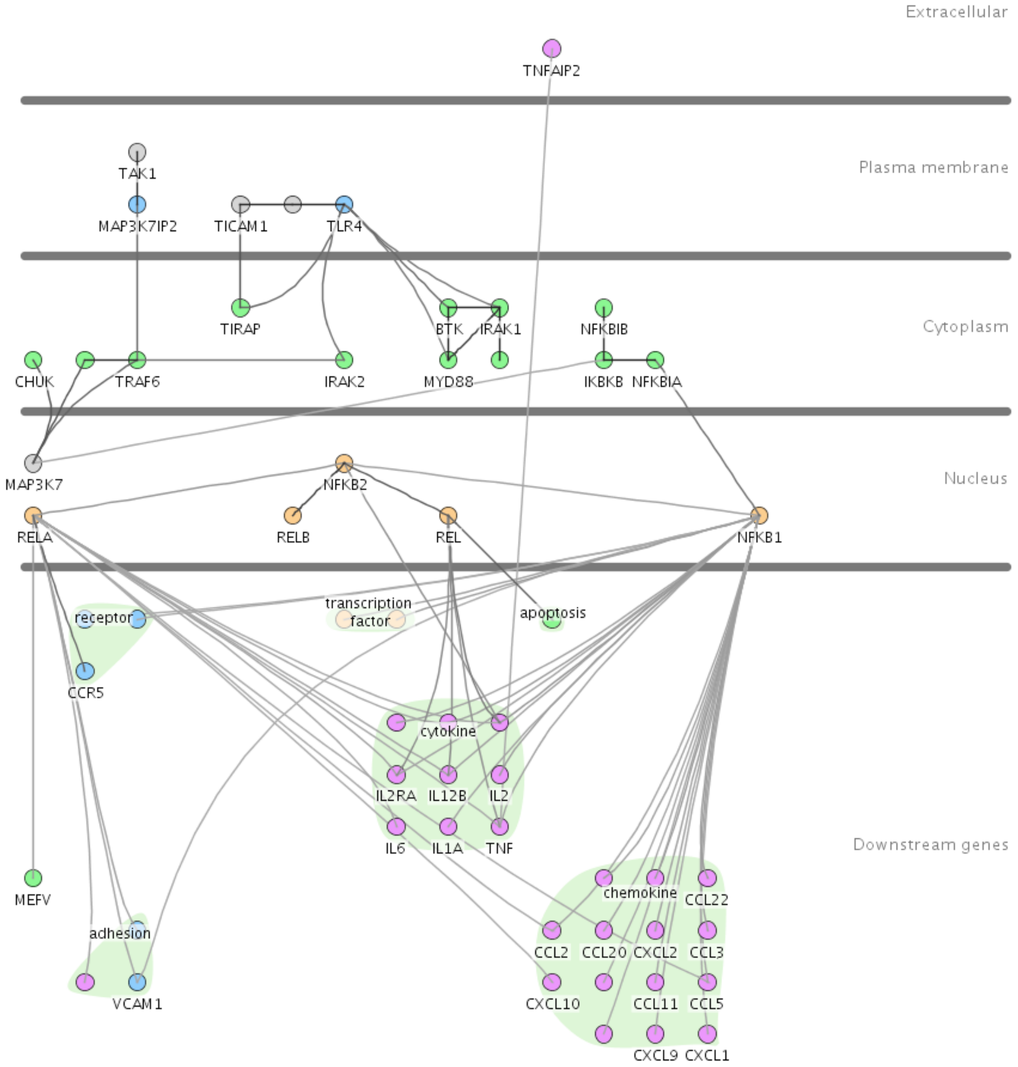
- Certain tasks can be totally **automated**
  - Statistical computations
  - Machine learning algorithms
  - We don’t need visualization for these tasks (although perhaps for debugging them…)
- Analysis problems are often **ill-specified**
  - What is the correct question?
  - Exploit human visual system, pattern detection capabilities
  - Goal may be an automated solution or a visual analysis system
- Presentation

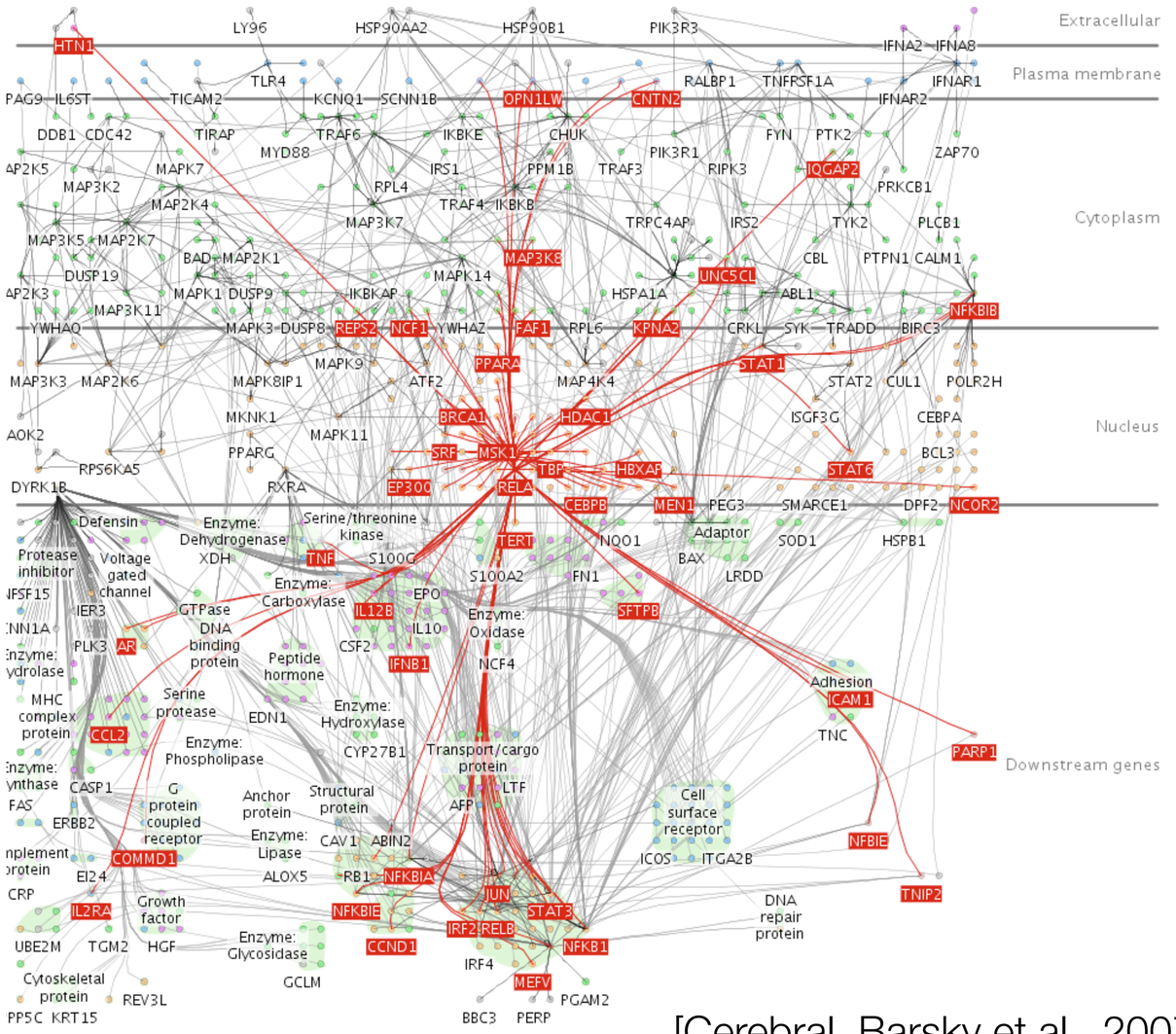
Why Computers?



[Cerebral, Barsky et al., 2007]

Why Computers?

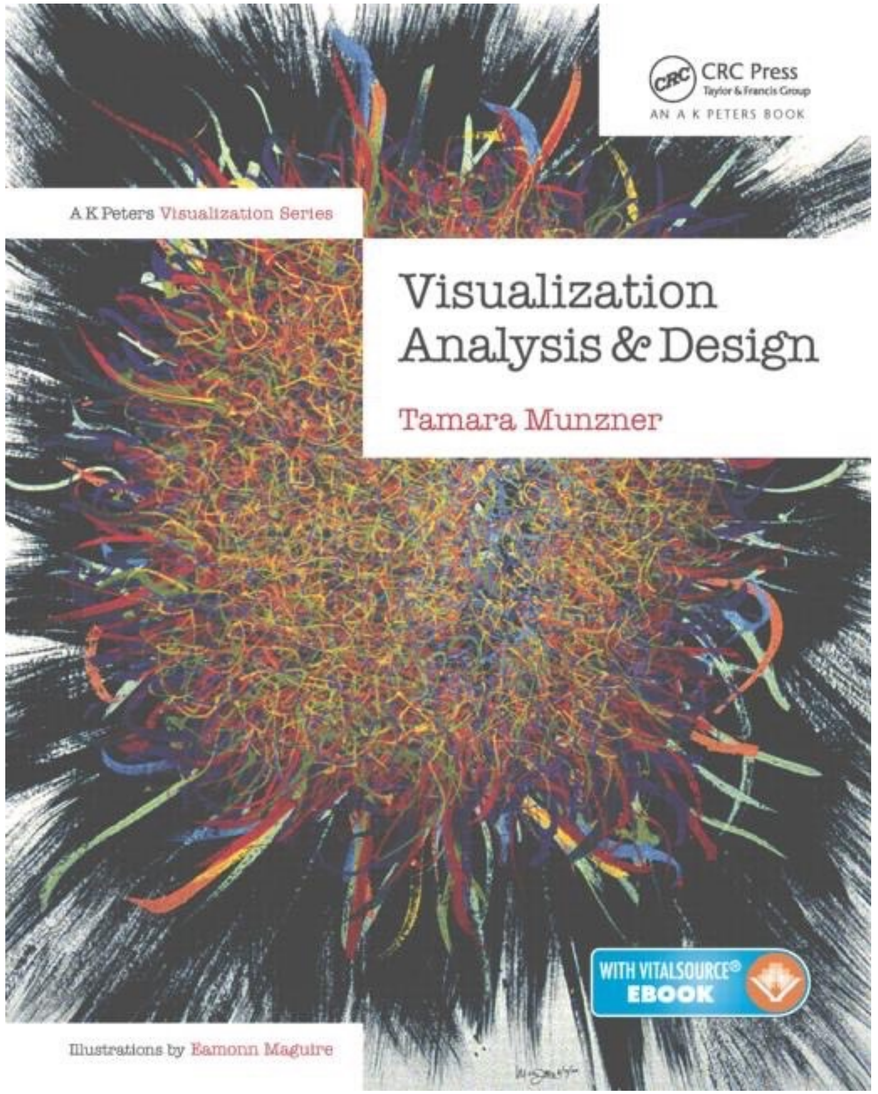




[Cerebral, Barsky et al., 2007]

Administrivia

- Course Web Site
- Syllabus
  - Plagiarism
  - Accommodations
- Textbook:
  - Required: Munzner (VAD)
  - Recommend: Murray, 2nd ed. (IDV)
- Assignments
- Exams: Midterm (Oct. 17) and Final (Dec. 11)
- Registration



Assignment 1

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

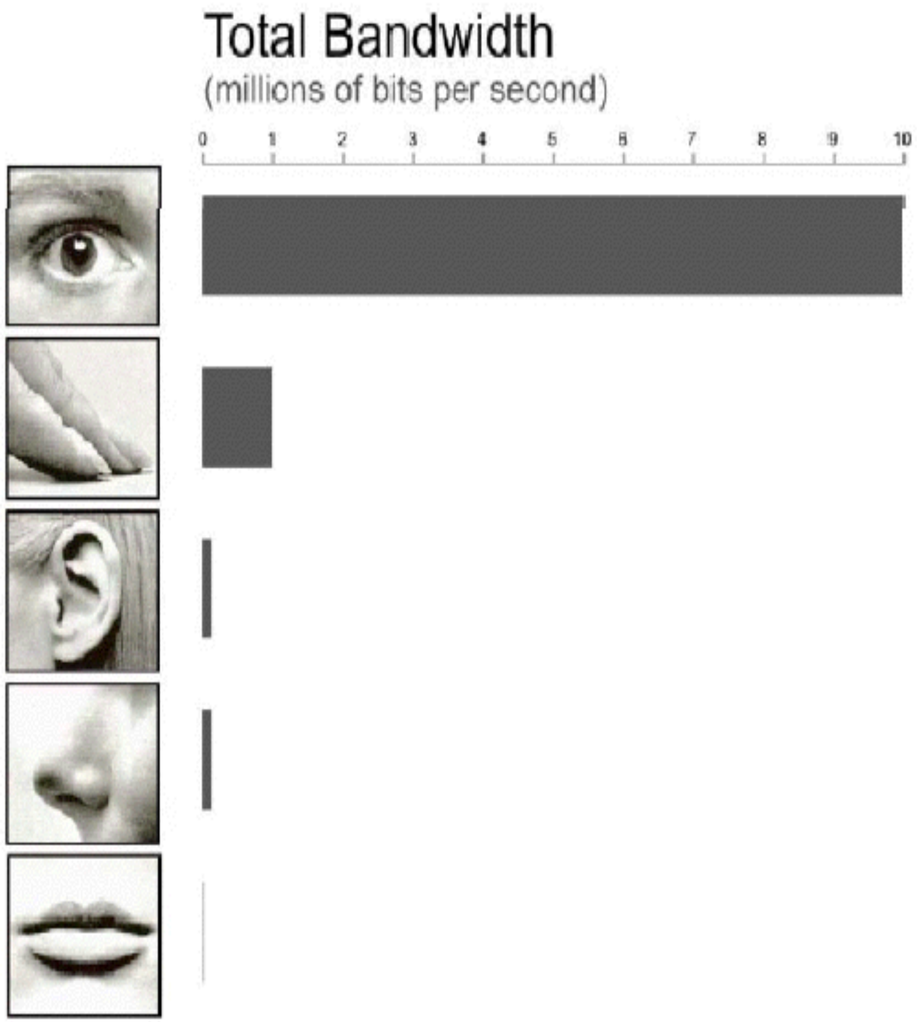
Do not cheat!

Questions?

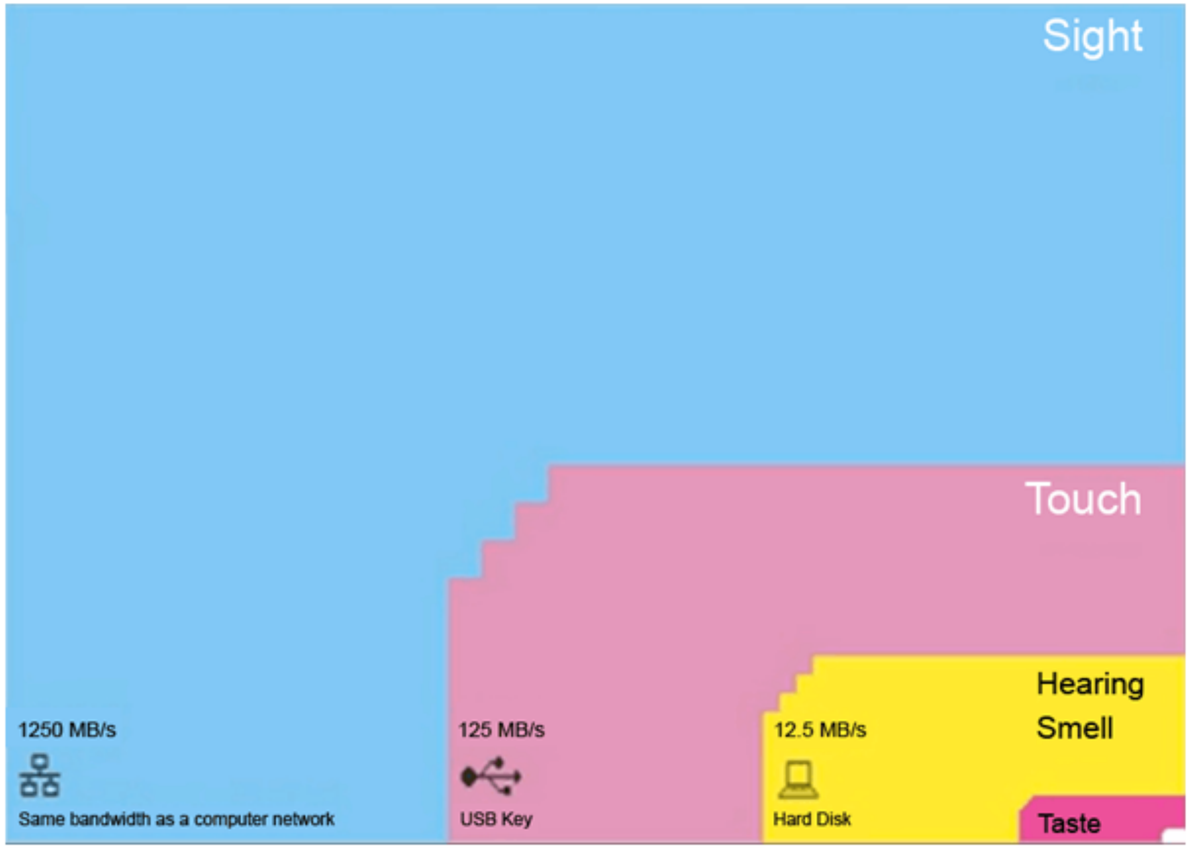
Definition

"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]

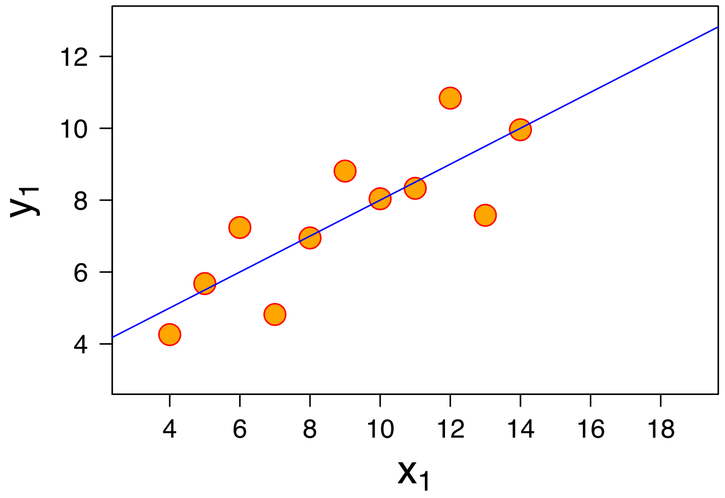
Why Visual?

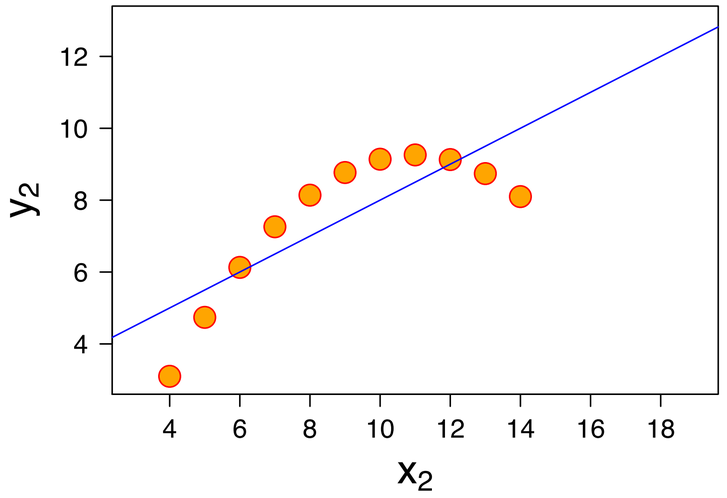
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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

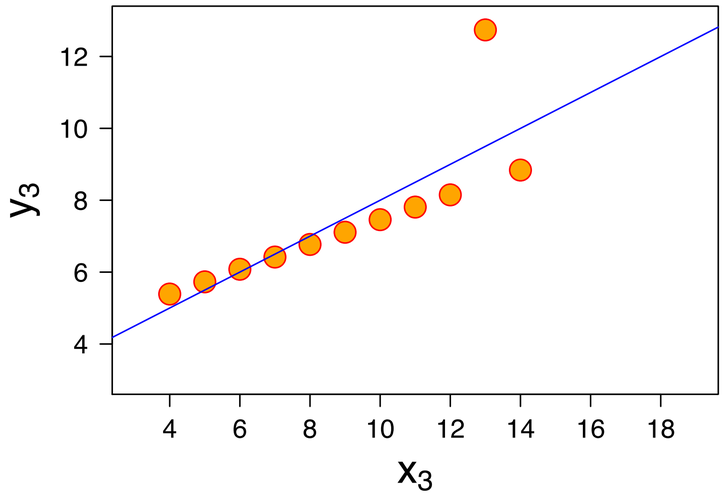
Mean of x	9
Variance of x	11
Mean of y	7.50
Variance of y	4.122
Correlation	0.816

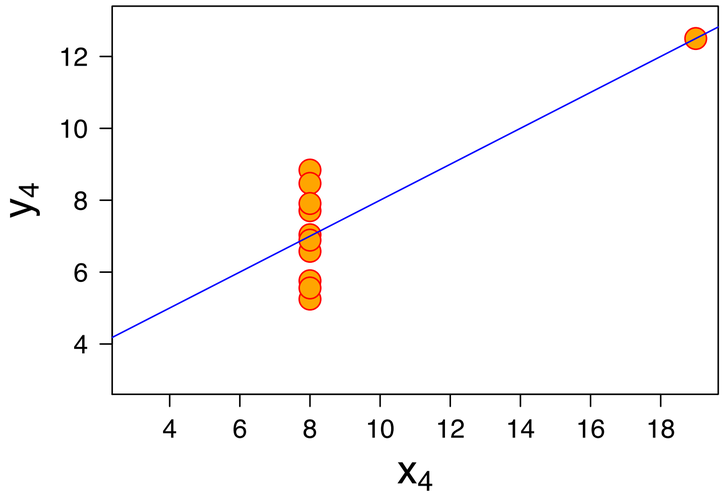
[F. J. Anscombe]

Why Visual?



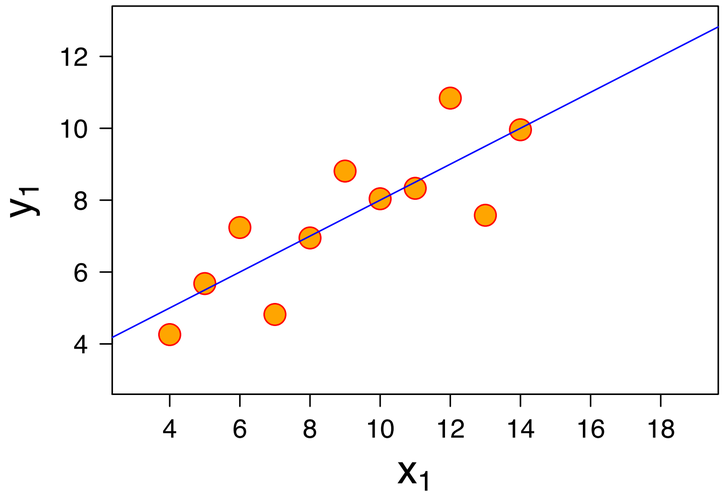


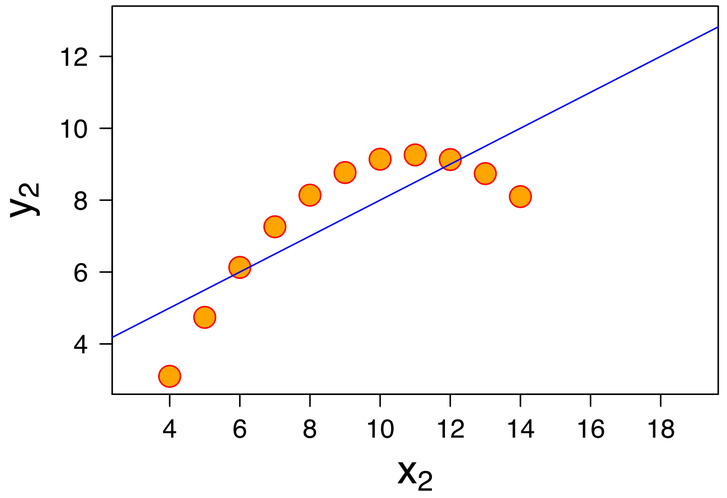


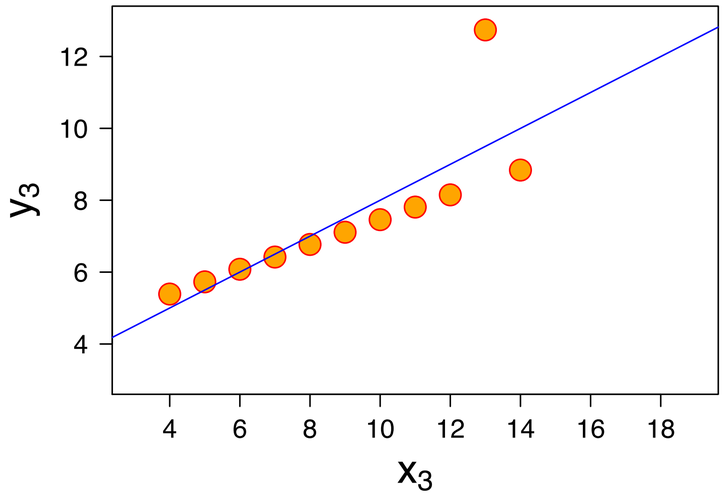


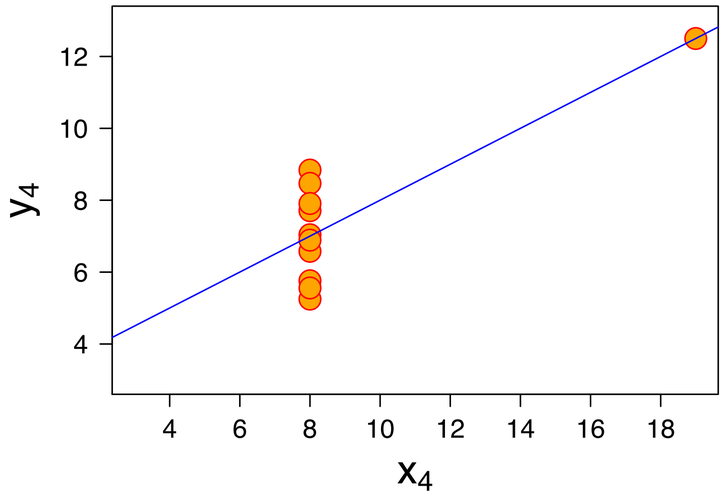
[F. J. Anscombe]

Why Visual?





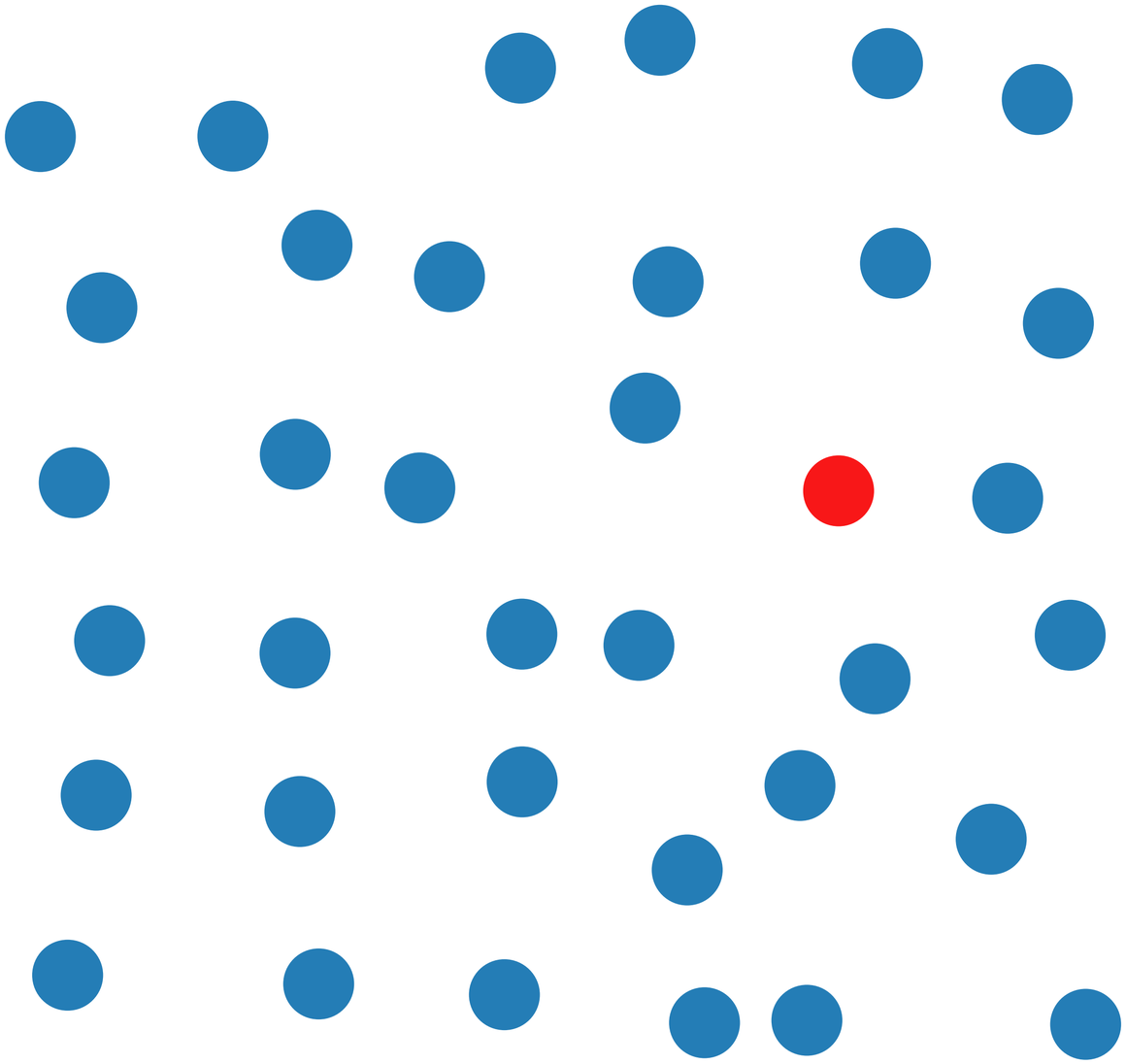




Mean of x	9
Variance of x	11
Mean of y	7.50
Variance of y	4.122
Correlation	0.816

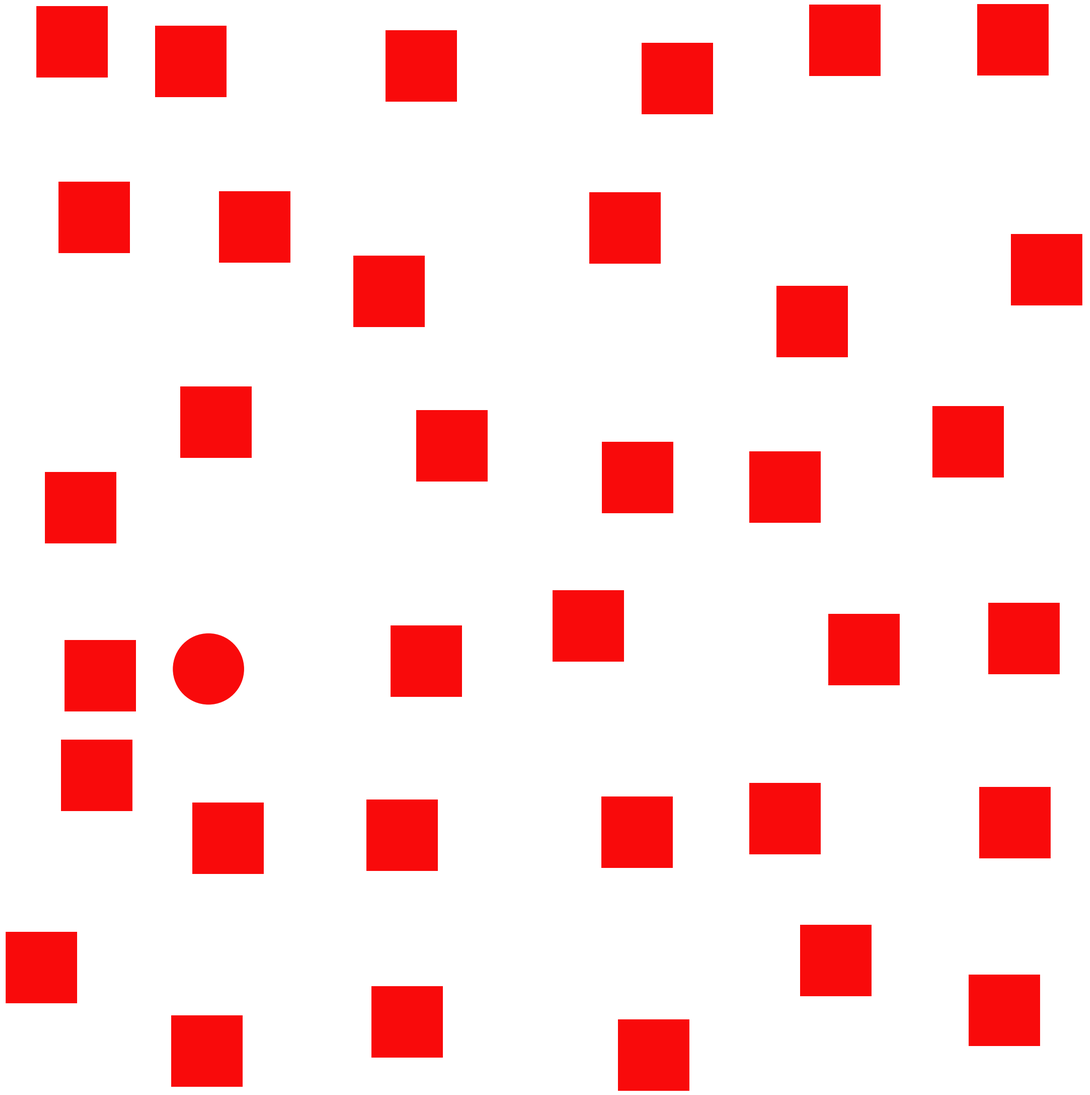
[F. J. Anscombe]

Visual Pop-out



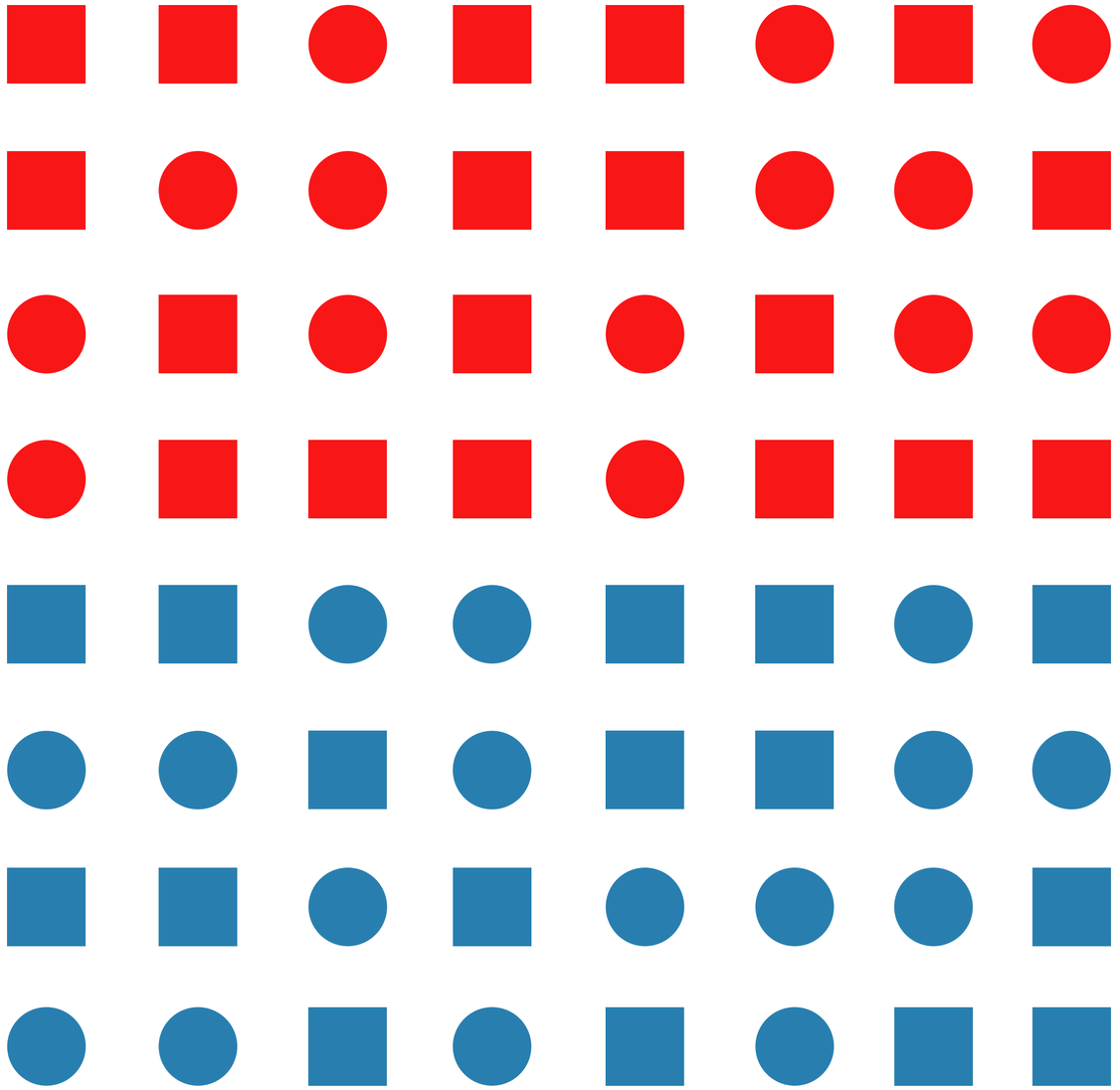
[C. G. Healey]

Visual Pop-out



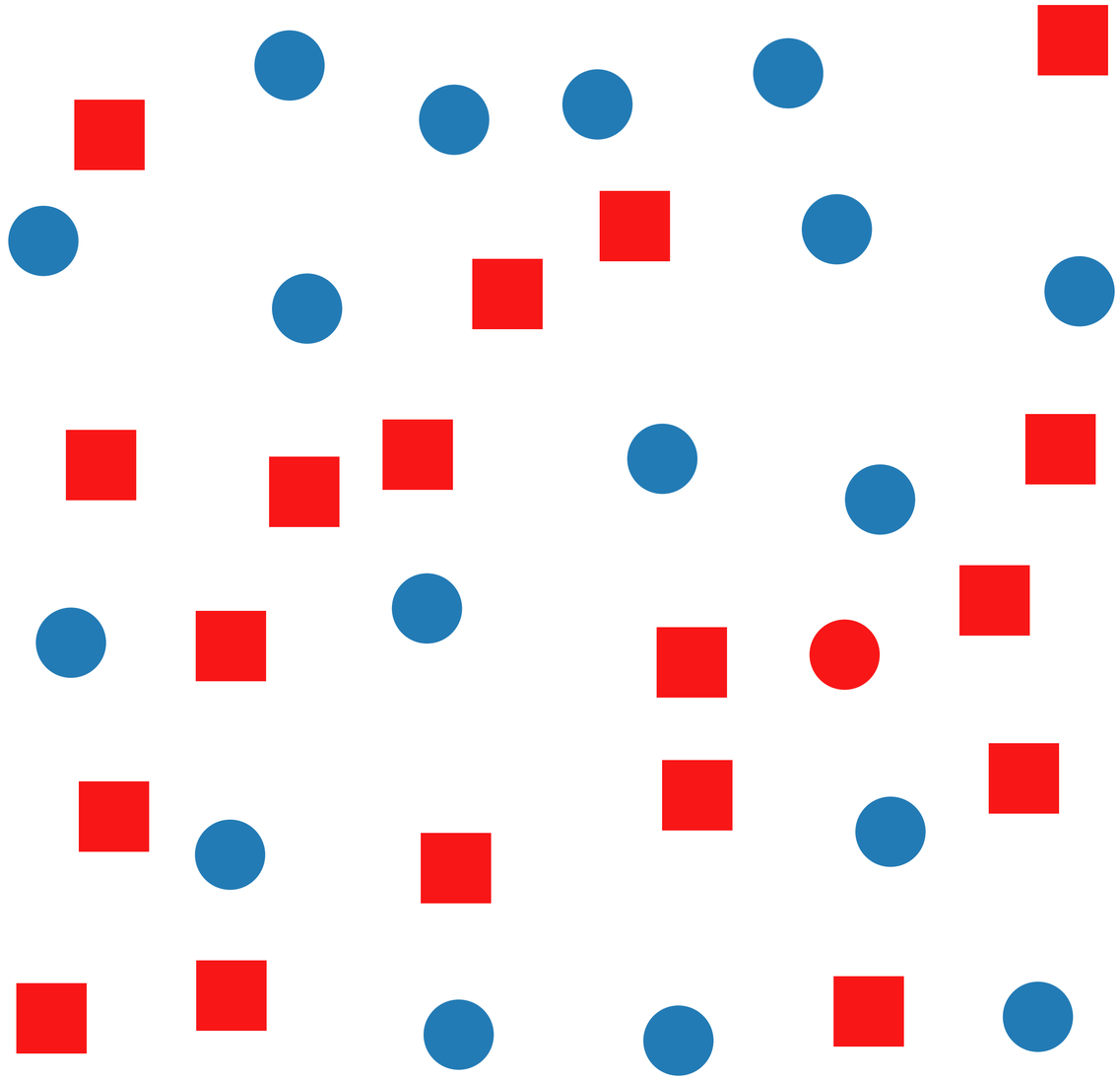
[C. G. Healey]

Visual Pop-out



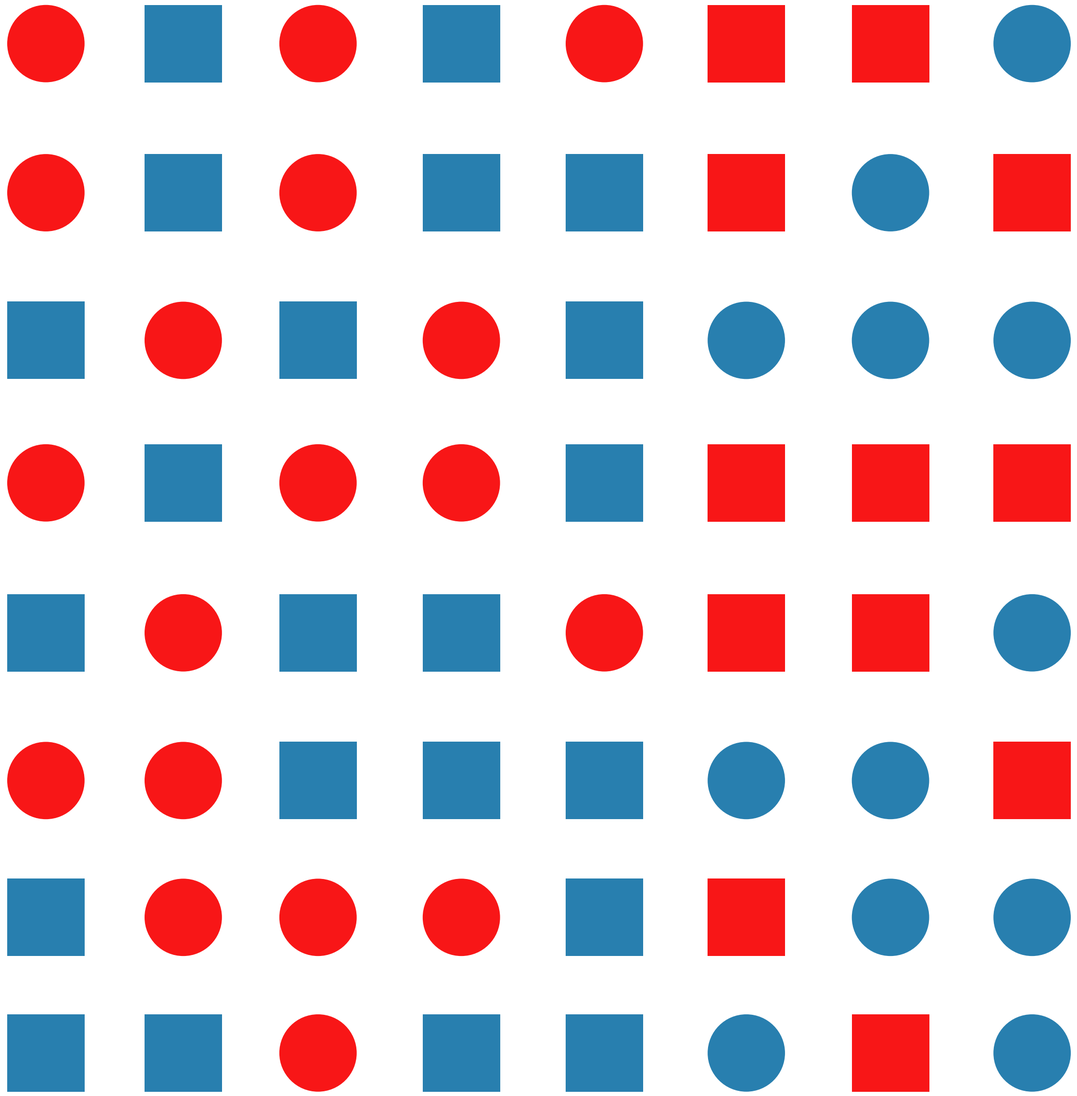
[C. G. Healey]

Visual Perception Limitations



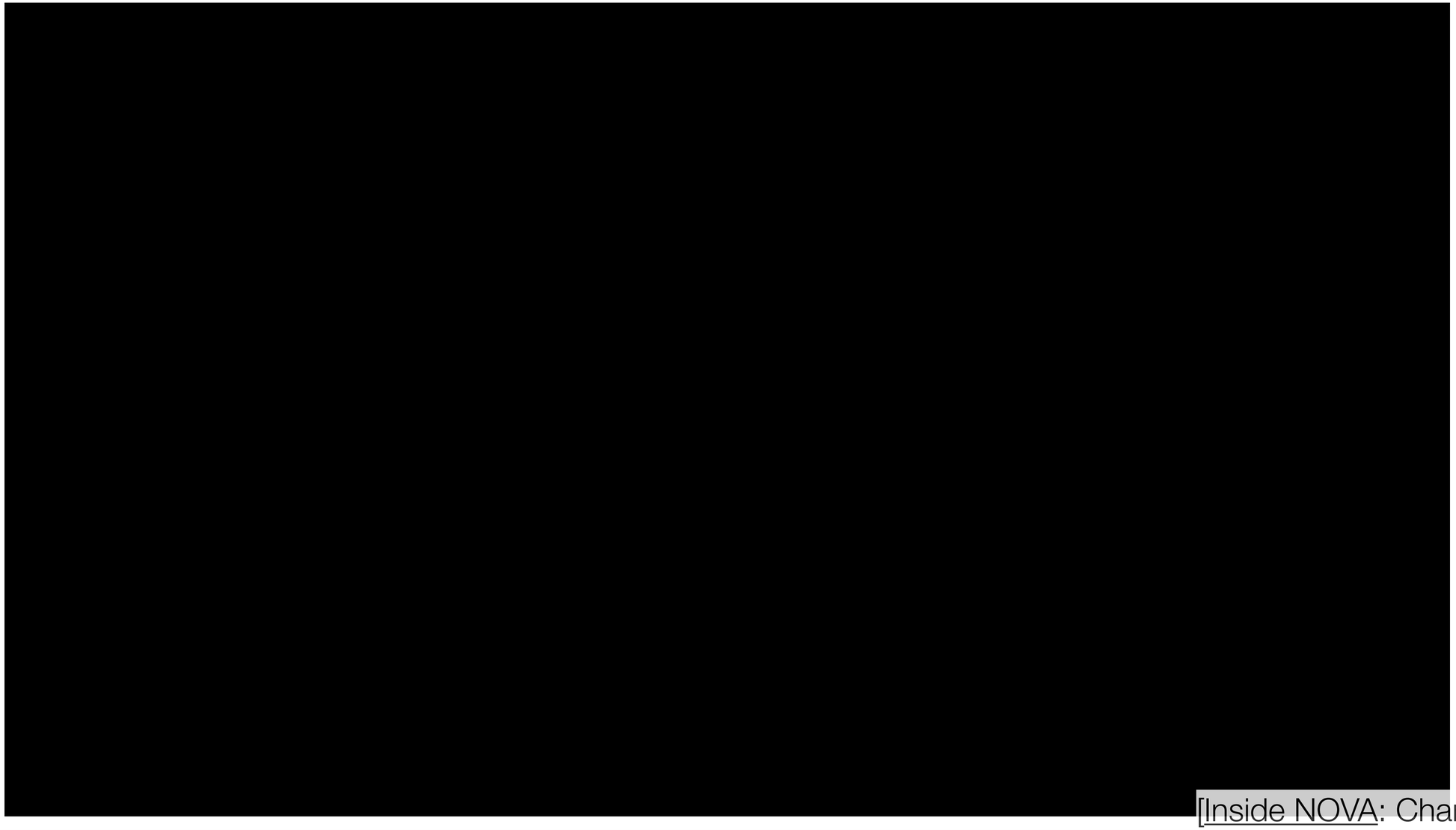
[C. G. Healey]

Visual Perception Limitations



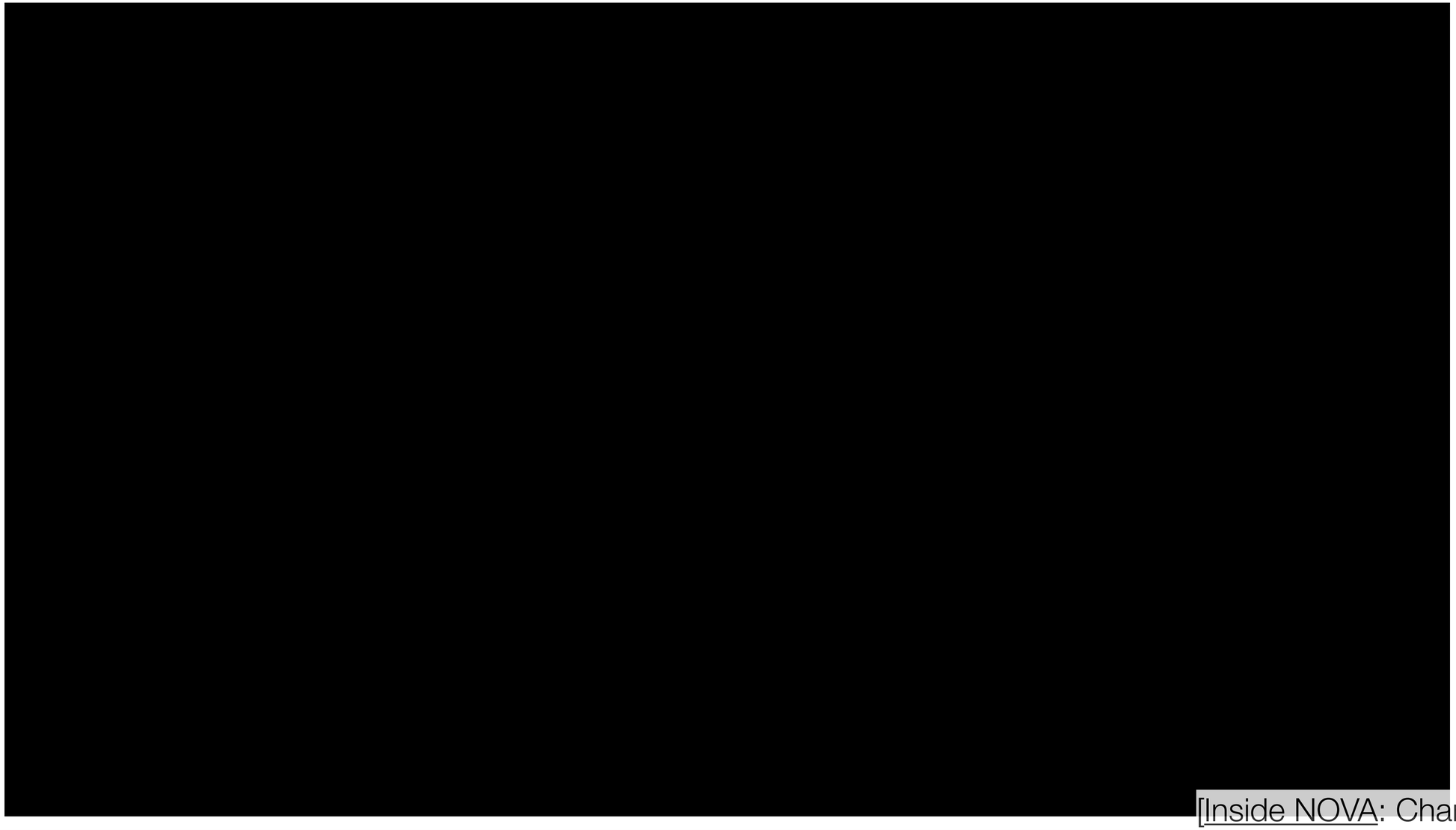
[C. G. Healey]

Human Perception



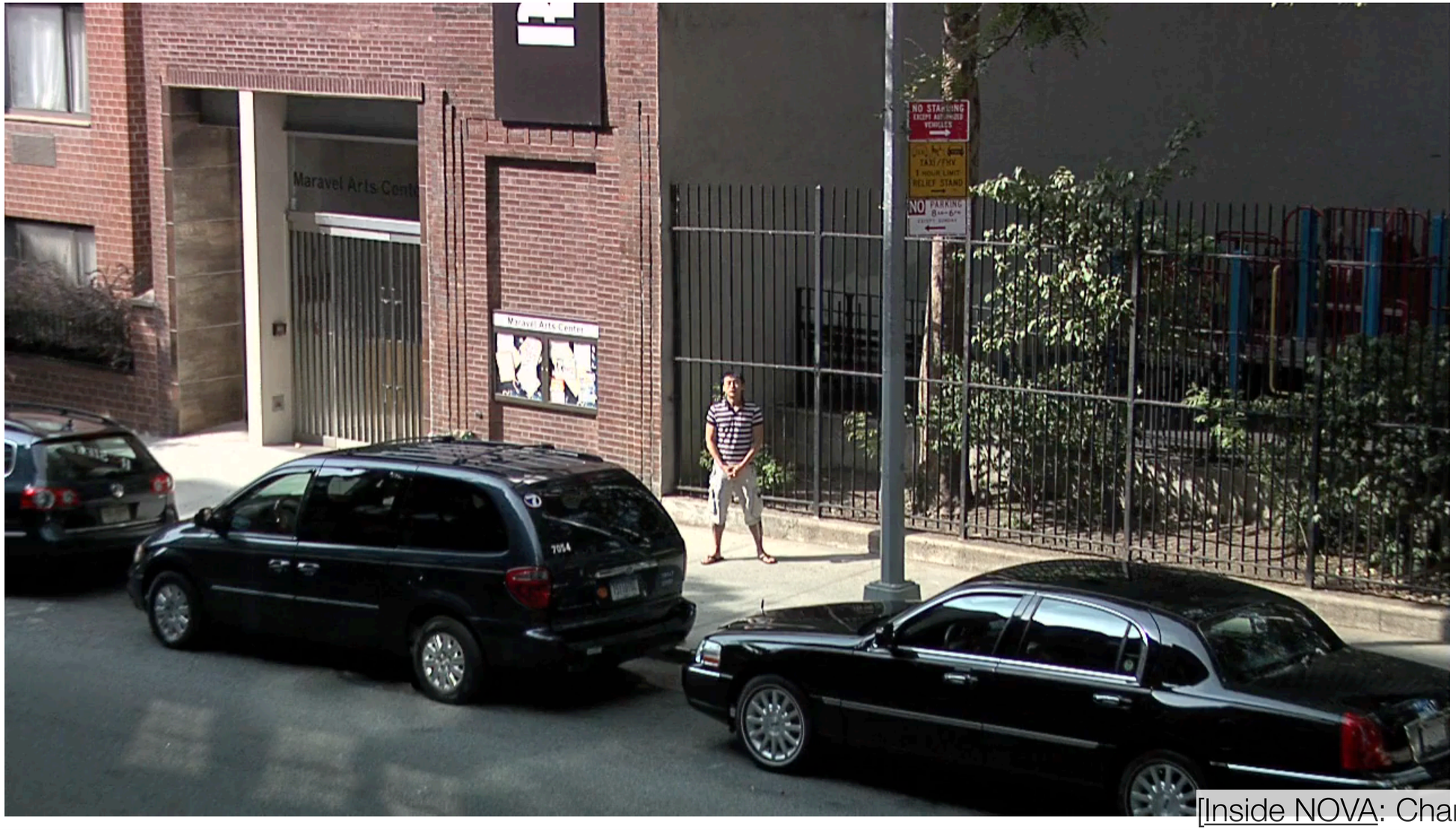
[Inside NOVA: Change Blindness]

Human Perception



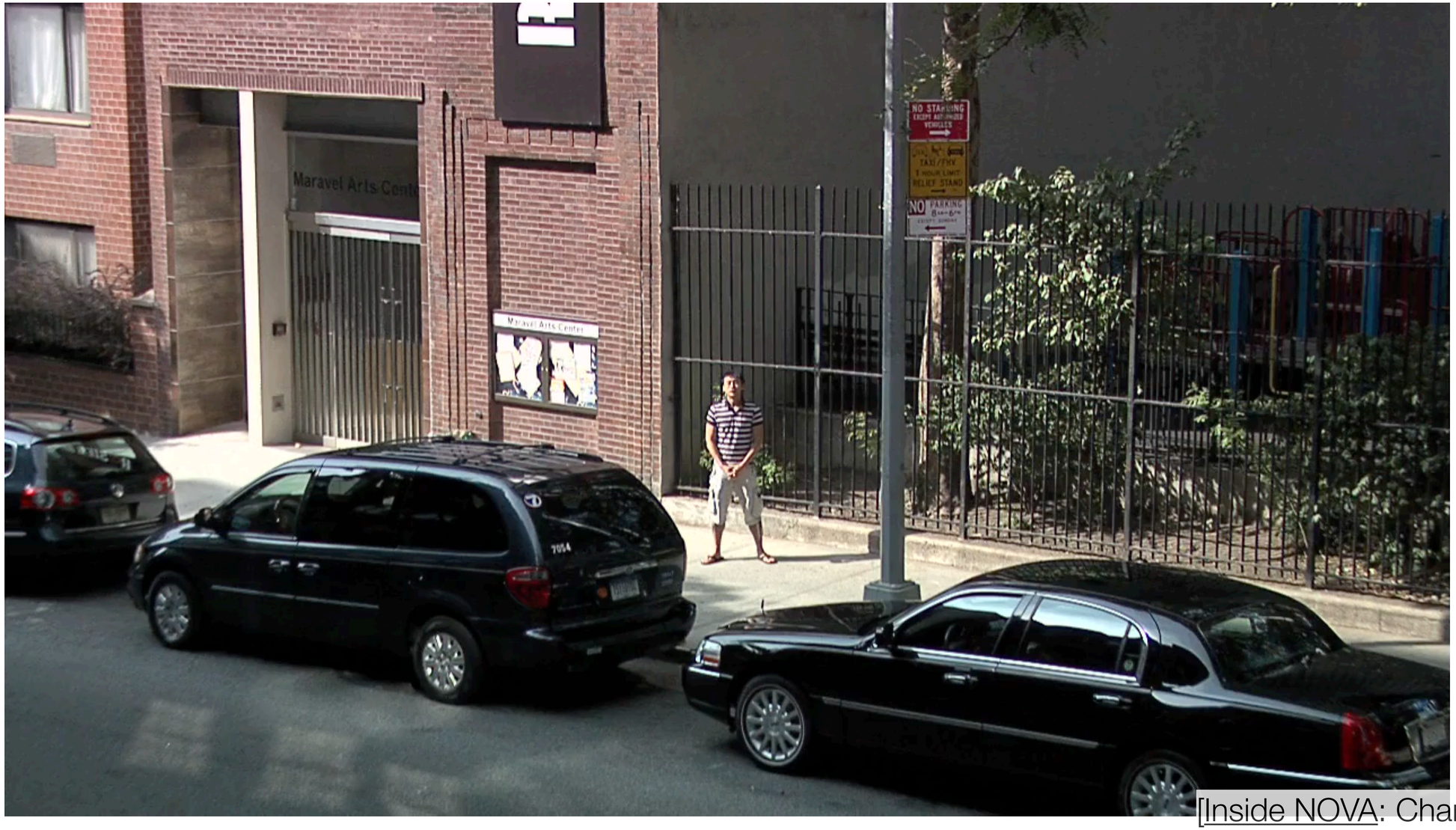
[Inside NOVA: Change Blindness]

Not Uncommon



[Inside NOVA: Change Blindness]

Not Uncommon



[Inside NOVA: Change Blindness]

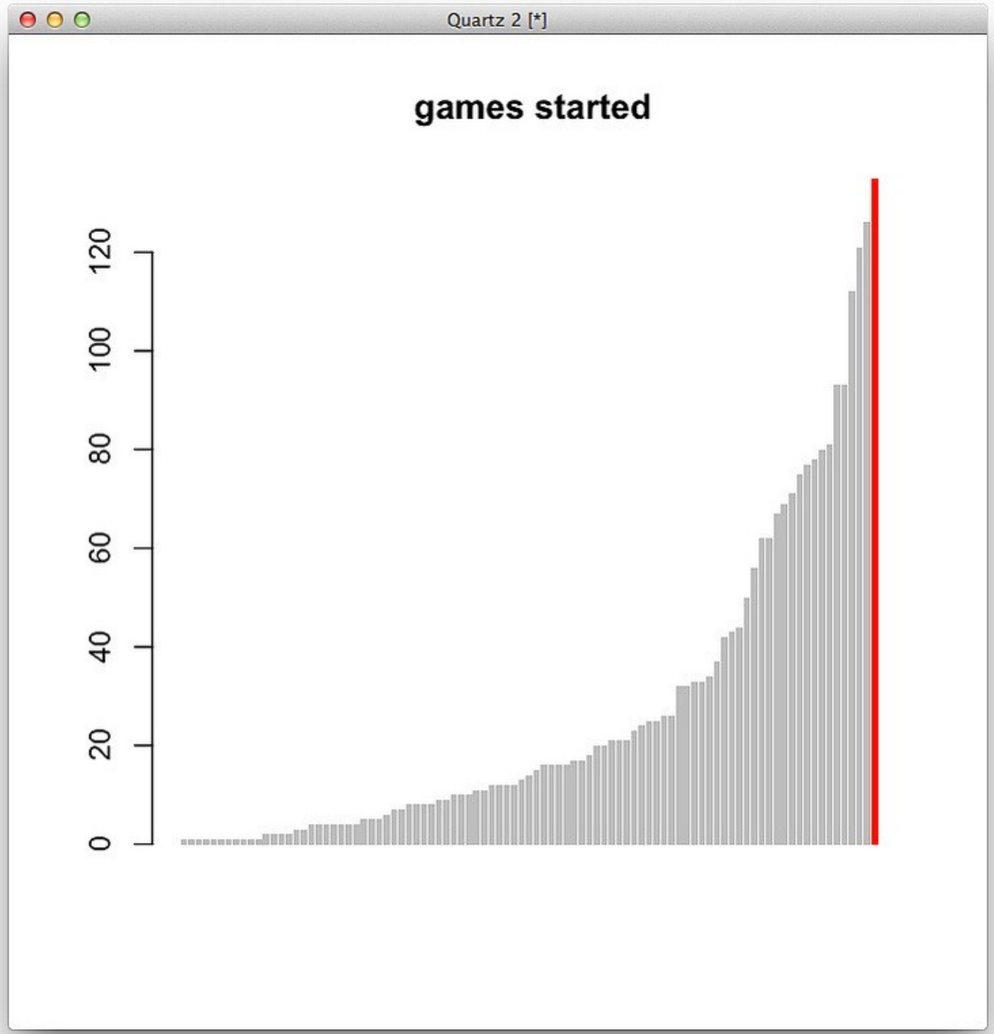
Other Human Limitations

- 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

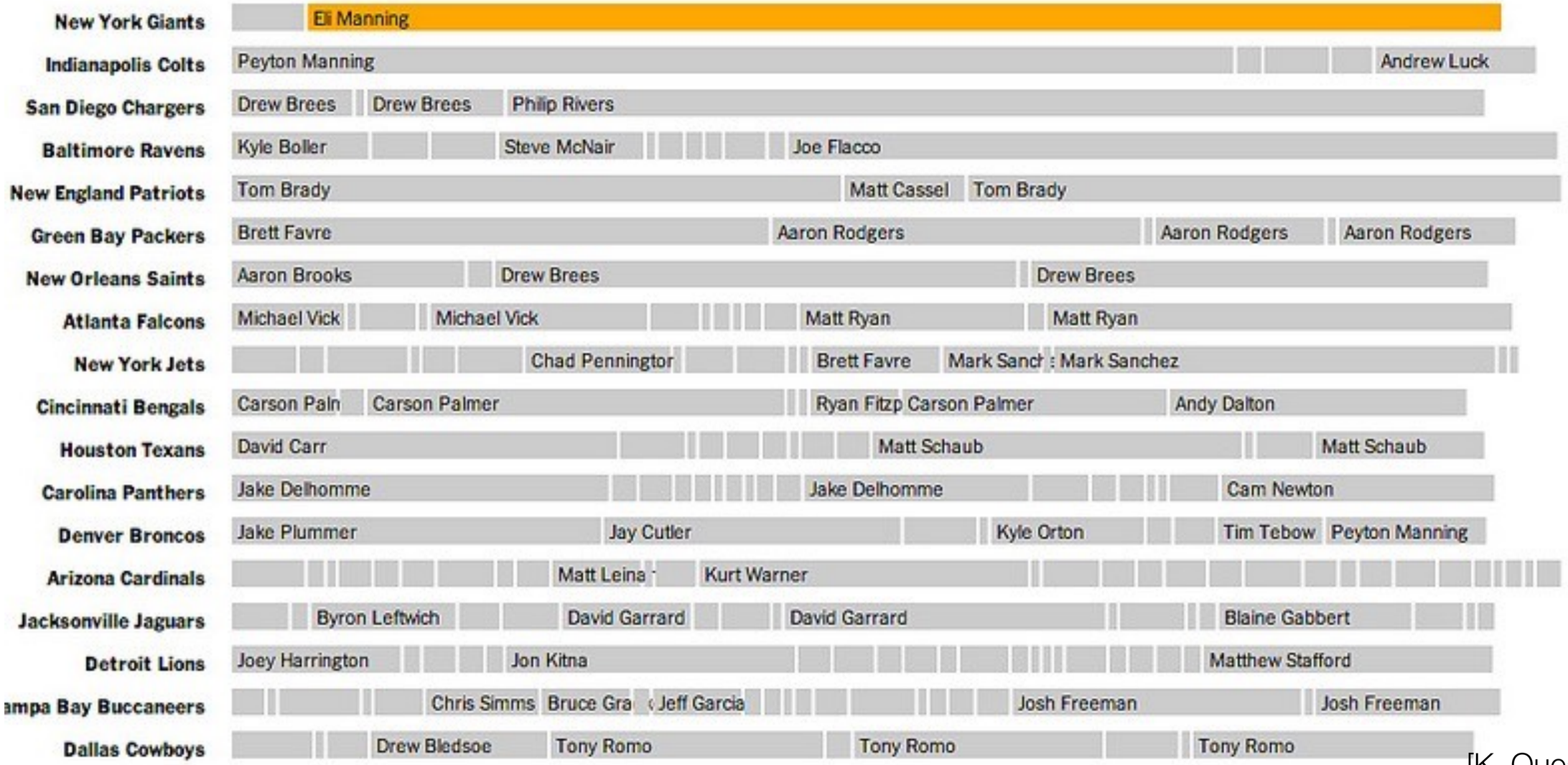
"Computer-based visualization systems provide visual representations of datasets **designed** to help people carry out tasks more effectively"

Design Iteration



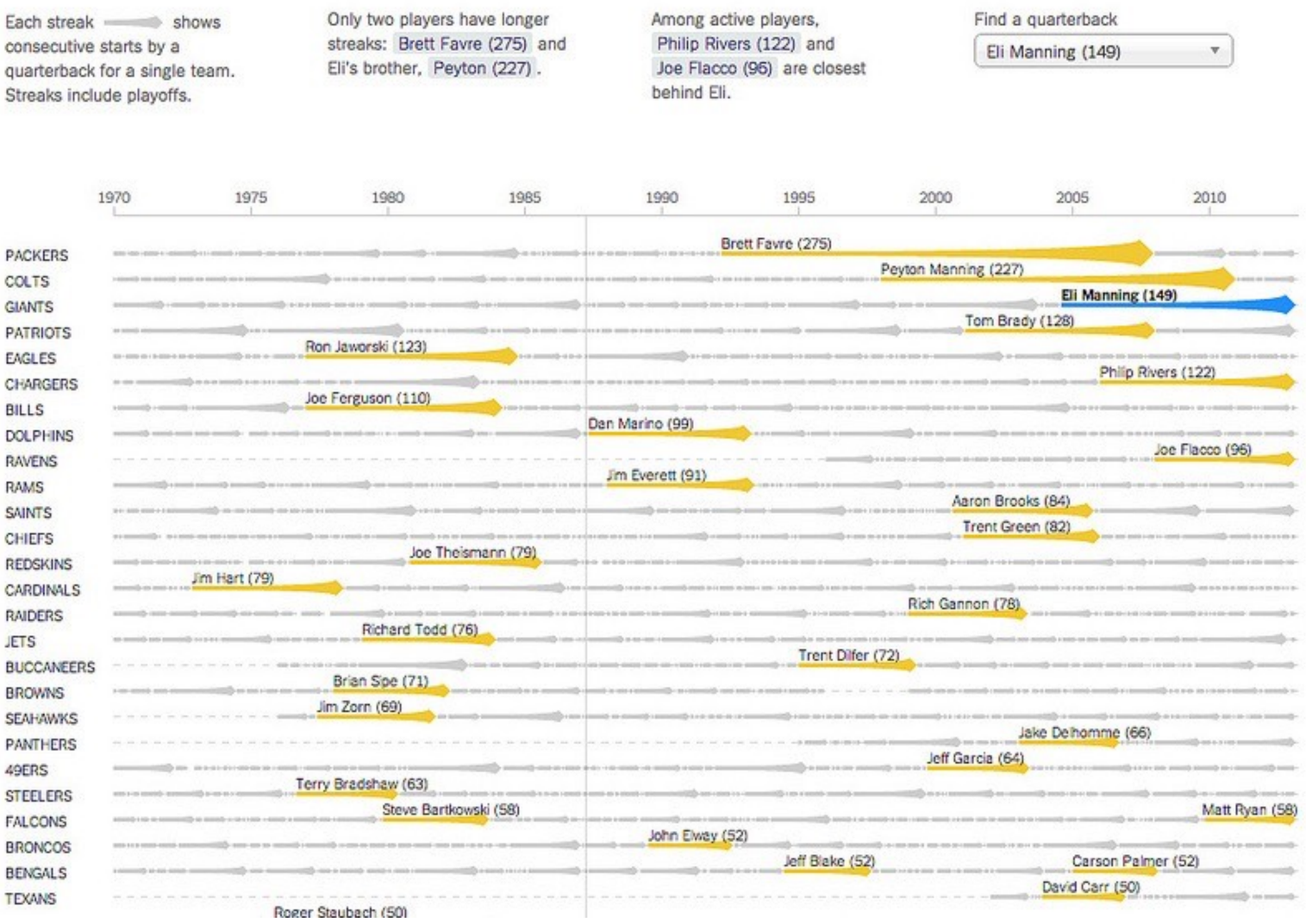
[K. Quealy, 2013]

Design Iteration



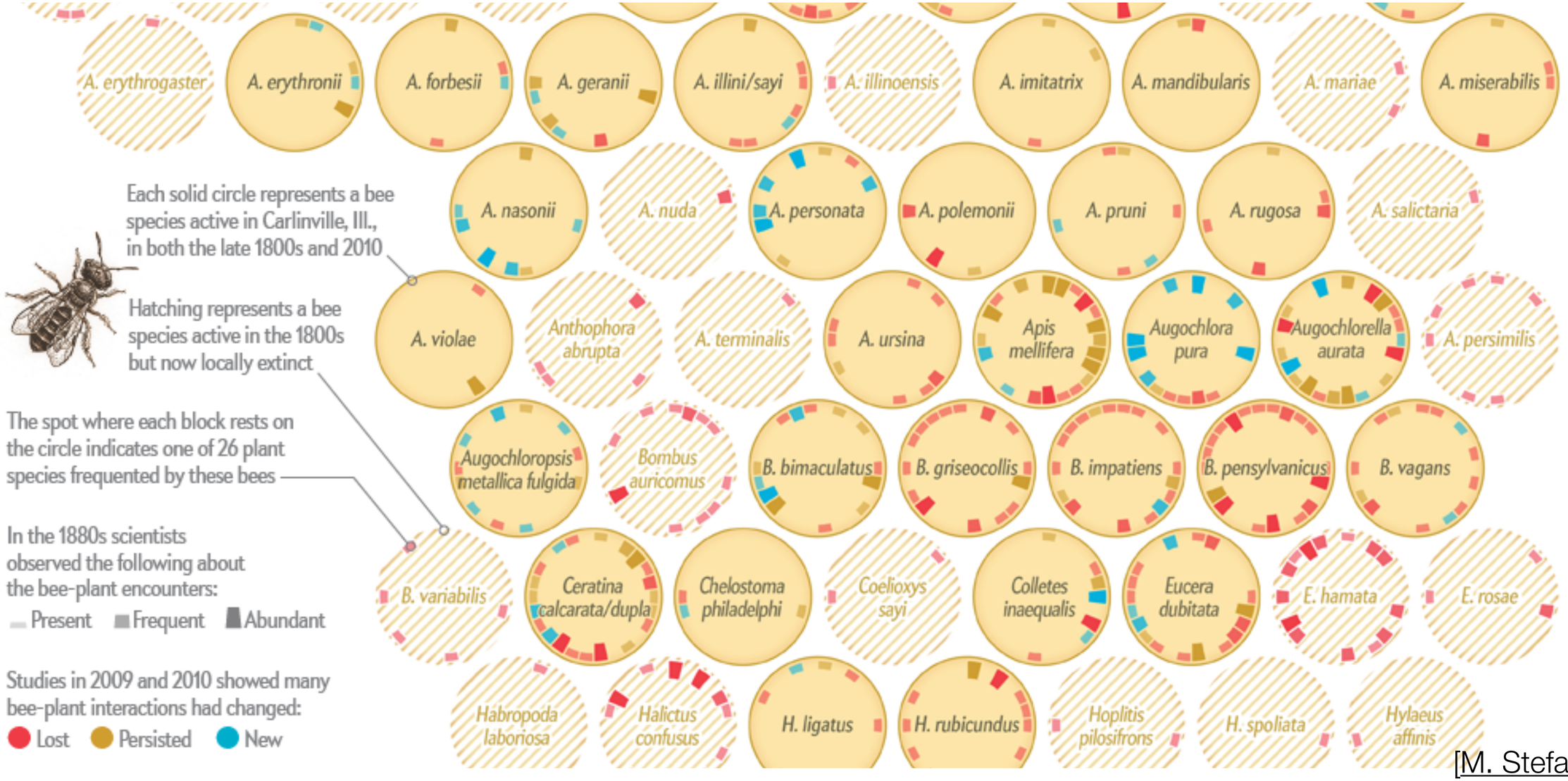
[K. Quealy, 2013]

Design Iteration



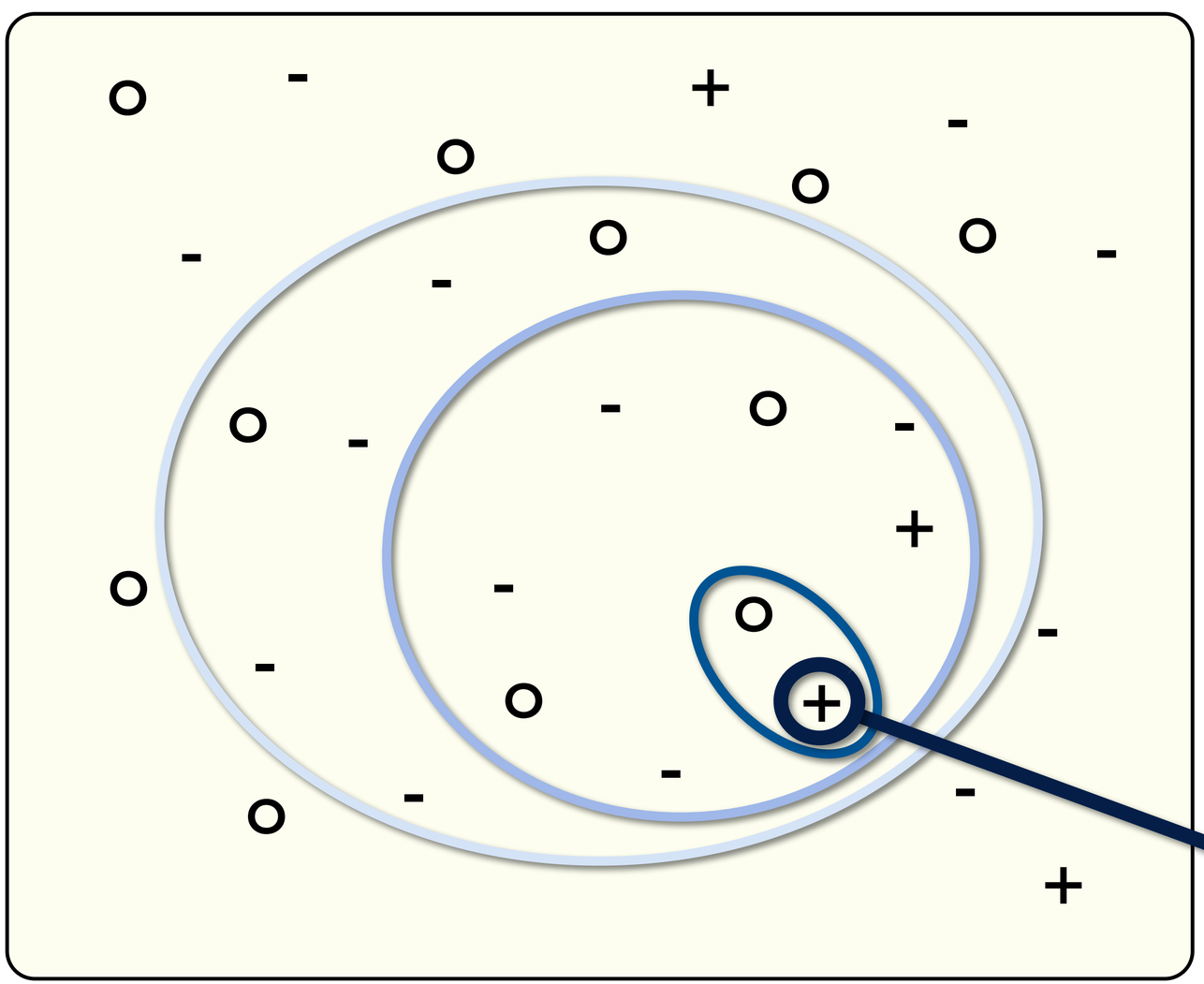
[K. Quealy, 2013]

Another Design Example



[M. Stefaner, 2013]

Design Space: Think Broad



know

consider

propose

select

[Design Study Methodology, Sedlmair et al., 2012]

Definition

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

Why Effectiveness?

- “It’s not just about pretty pictures”
- Any depiction of data requires the designer to make choices about how that data is visually represented
  - Analogy to photography
  - Lots of possibilities (see quarterback study)
- Effectiveness measures how well the visualization helps a person with their tasks
  - How? insight, engagement, efficiency?
  - Benchmarks and user studies

What languages do we use on the Web?

Languages of the Web

- HTML
- CSS
- SVG
- JavaScript
  - Versions of Javascript: ES6, ES2015, ES2017…
  - 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: CodePen, JSFiddle, Liveweave, CodeSandbox, etc.
- IDEs: WebStorm, etc.

Hyper Text Markup Language (HTML)

- Markup languages allow users to encode the **semantics** of text
- Tags define the boundaries of the structures of the content
  - Tags are enclosed in angle brackets (e.g. `<html>`)
  - Most of the time, you have a start and end tag
  - End tags are just like start tags except that they have forward slash after the open bracket (e.g. `</html>`)
  - Tags may be nested but not mismatched
    - `<p>A <strong><em>very</em></strong> cool example</p>`
    - `<p>A <strong>very <em>cool</strong></em> example</p>`
  - What about `<img src="mypicture.png" alt="My Image">`?

HTML Elements and Attributes

- Tags denote **elements** of the content (e.g. sections, paragraphs, images)
- Each element may have **attributes** which define other information about the element
  - An attribute has a **key** and **value** (`key="value"`)
  - e.g. `<img src="mypicture.png" alt="My Image">`
- Many different elements available
  - Common: headers (`h1`, …, `h6`), paragraph (`p`), lists (`ul`, `ol`, `li`), emphasis (`em`, `strong`), link (`a`), spans & divisions (`span`, `div`)
  - Lots more (e.g. `abbr`): see MDN Documentation
- Many different attributes available
  - See MDN Documentation: note that some are legacy due to CSS

HTML Element Structure & Naming

- Elements structure a document
  - Document Object Model (DOM)
  - We can visualize this information
  - More importantly, we can **navigate** this tree
- Identifying and Classifying elements: `id` and `class` attributes
  - `id` identifies a **single** element—use for a unique case
  - `class` may identify **multiple** elements—use for common cases
  - Each element may have multiple classes, separate by spaces
  - Use normal identifiers: don't start the name with a number

Other HTML Trivia

- `<`, `>`, `&`, and `"` are special characters, escape with `<`, `>`, `&`, and `&quot;` (note the semi-colon)
- Comments are enclosed by `<!--` and `-->`

```
- <!-- This is a comment -->
```

- HTML Documents begin with a `DOCTYPE` declaration
  - For HTML5, this is easier `<!DOCTYPE html>`
- meta tag: `<meta charset="UTF-8"/>`
- HTML has audio and video tags, math equation support, and more

Basic HTML File

```
<!DOCTYPE html>
<html>
  <head>
  <title>A Basic Web Page</title>
</head>
<body>
  <h1>My Wicked Awesome Web Page</h1>
  <p><em>This is <strong>cool</strong>. What about
<u><strong>this?</strong></u></em></p>
  <img src="https://upload.wikimedia.org/wikipedia/commons/thumb/9/9f/
Tom_Brady_2011.JPG/409px-Tom_Brady_2011.JPG"/>
</body>
</html>
```

- https://codepen.io/dakoop/pen/PdRKEL

What is CSS?

Cascading Style Sheets (CSS)

- Separate from content, just specifies how to style the content
- Style information can appear in three places:
  - External file
  - In a style element at the beginning of the HTML file
  - In a specific element in the body of a document (least preferable)
- Why Cascading?
  - Don’t want to have to specify everything over and over
  - Often want to use the same characteristics in a region of the DOM
  - Use inheritance: properties that apply to children cascade down

CSS Selectors

- How do we specify what part(s) of the page we want to style?
- The **element types** themselves (the HTML tag)
  - `strong { color: red; }`
- **Classes** of elements (ties to HTML `class` attribute)
  - `.cool { color: blue; }`
- A **specific** element (ties to HTML `id` attribute)
  - `#main-section { color: green; }`
- Relationships
  - Descendant: `p em { color: yellow; }`
  - Child: `p > em { color: orange; }`
- Pseudo-classes: `a:hover { color: purple; }`

Other CSS Bits

- Comments: `/* This is a comment in CSS */`
- Grouping Selectors: `p, li { font-size: 12pt; }`
- Multiple Classes: `.cool.temp { color: blue; }`
- Colors:
  - Names (Level 1, 2, & 3): `red, orange, antiquewhite`
  - Dash notation (3- & 6-character): `#fff, #00ff00`
  - Integer or % RGB and HSL Functions: `rgb( 255, 0, 0), rgb(50%, 50%, 0%), hsl(120, 100%, 50%)`
  - Also `background-color`
- Watch out for multiple rules (look at how a web browser parses)
- Again, much more documentation at MDN

Example CSS

```
body {
  font-face: sans-serif;
  font-size: 12pt;
}
em { color: green; }
em u { color: red; }
em > strong { color: blue; }
img { border: 4px solid red; }
```

- What colors are displayed for this HTML (with the above stylesheet)?
  - `<em>This is <strong>cool</strong>. What about <u><strong>this?</strong></u></em>`
- https://codepen.io/dakoop/pen/ErNJvJ

CSS Specificity

- Example:
  - CSS:

```
p.exciting { color: red; }
p { color: blue; }
```

  - What is the color of the paragraph

```
<p class="exciting">Cool</p>
```
?

- Generally, last rule listed overrides previous rules
- …but anytime a selector is **more specific**, it has precedence
- `p.exciting` is a more specific selector
- When in doubt, **test it** in a browser
- https://codepen.io/dakoop/pen/MLbRQz

How to add CSS to HTML

- External: a separate file via a link element (in the `<head>` section):
  - `<link rel="stylesheet" href="styles.css">`
- Embedded: in the header:
  - `<style type="text/css"> … </style>`
- Inline: for a specific element: (**Discouraged!**)
  - `<p style="font-weight: bold;">Some text</p>`