

Data Visualization (CIS 490/680)

Data & Tasks

Dr. David Koop

SVG Manipulation Example

- Draw a horizontal bar chart
 - `var a = [6, 2, 6, 10, 7, 18, 0, 17, 20, 6];`
- Steps:
 - Programmatically create SVG
 - Create individual rectangle for each item
- Link:
 - <https://codepen.io/dakoop/pen/mdbxQKe>

Data

- What is this data?

R011	42ND STREET & 8TH AVENUE	00228985	00008471	00000441	00001455	00000134	00033341	00071255
R170	14TH STREET-UNION SQUARE	00224603	00011051	00000827	00003026	00000660	00089367	00199841
R046	42ND STREET & GRAND CENTRAL	00207758	00007908	00000323	00001183	00003001	00040759	00096613

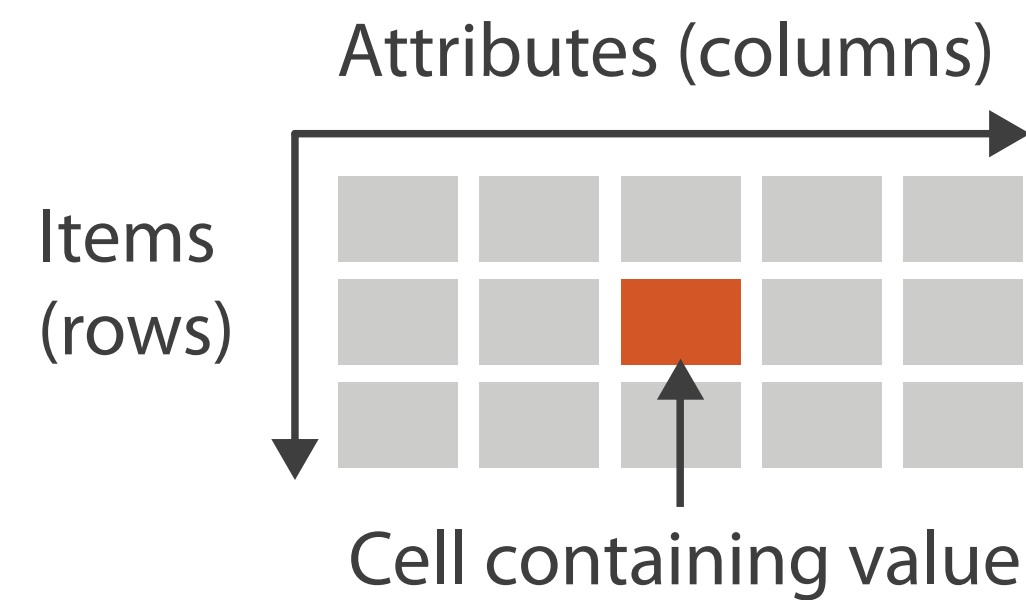
- **Semantics**: real-world meaning of the data
- **Type**: structural or mathematical interpretation
- Both often require **metadata**
 - Sometimes we can infer some of this information
 - Line between data and metadata isn't always clear

Data Terminology

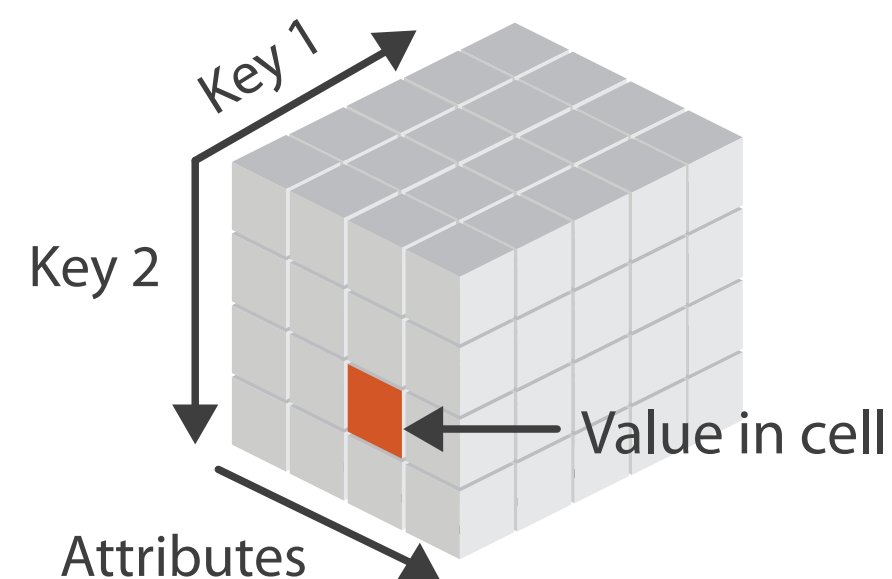
- Item (also Nodes): an entity
- Link: relationship between two items
- Attribute: property of an item
- Position: location in space
- Grid: how data is sampled

Dataset Types

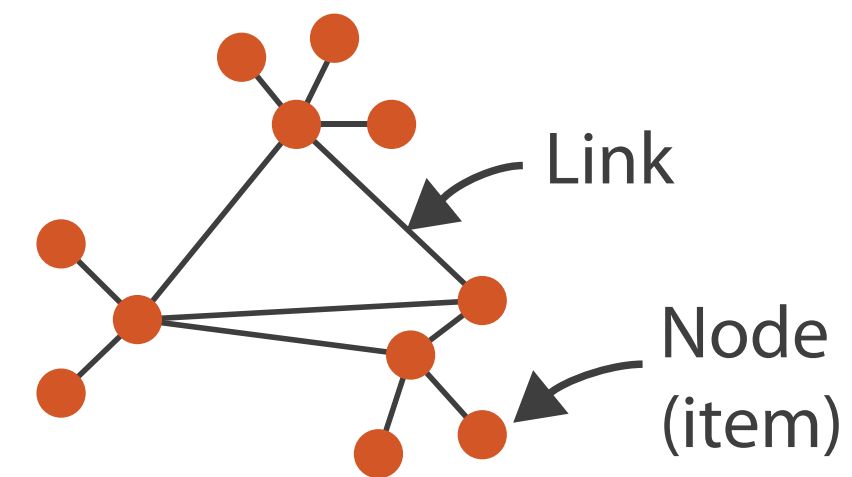
→ Tables



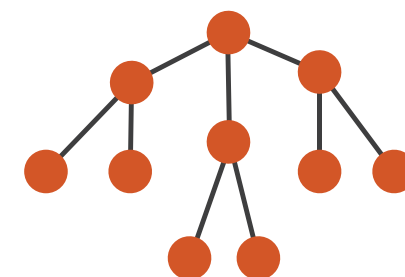
→ *Multidimensional Table*



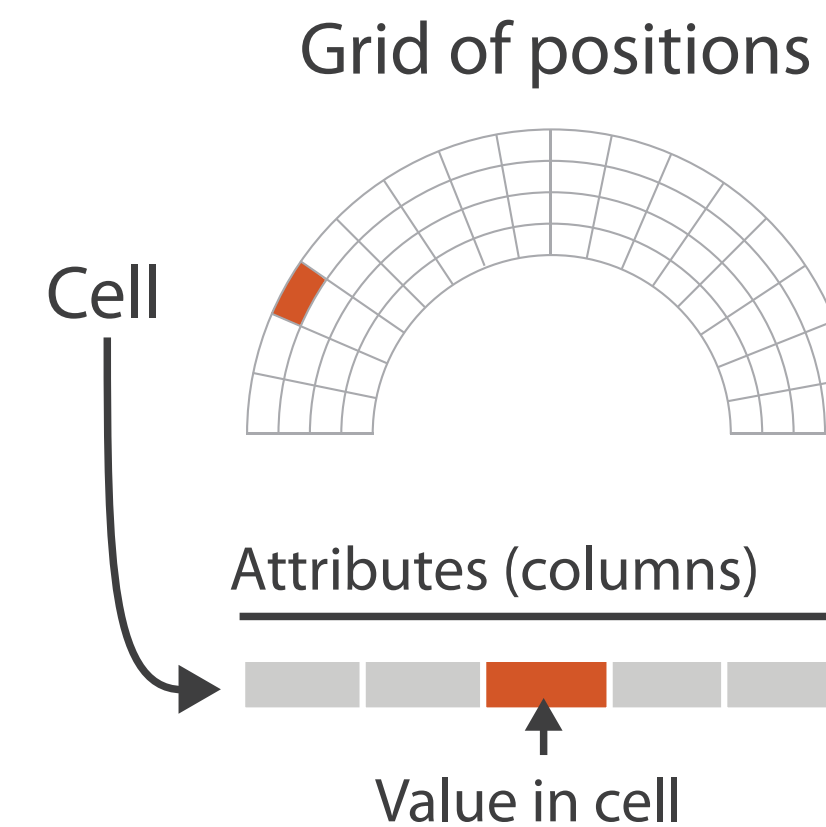
→ Networks



→ Trees



→ Fields (Continuous)

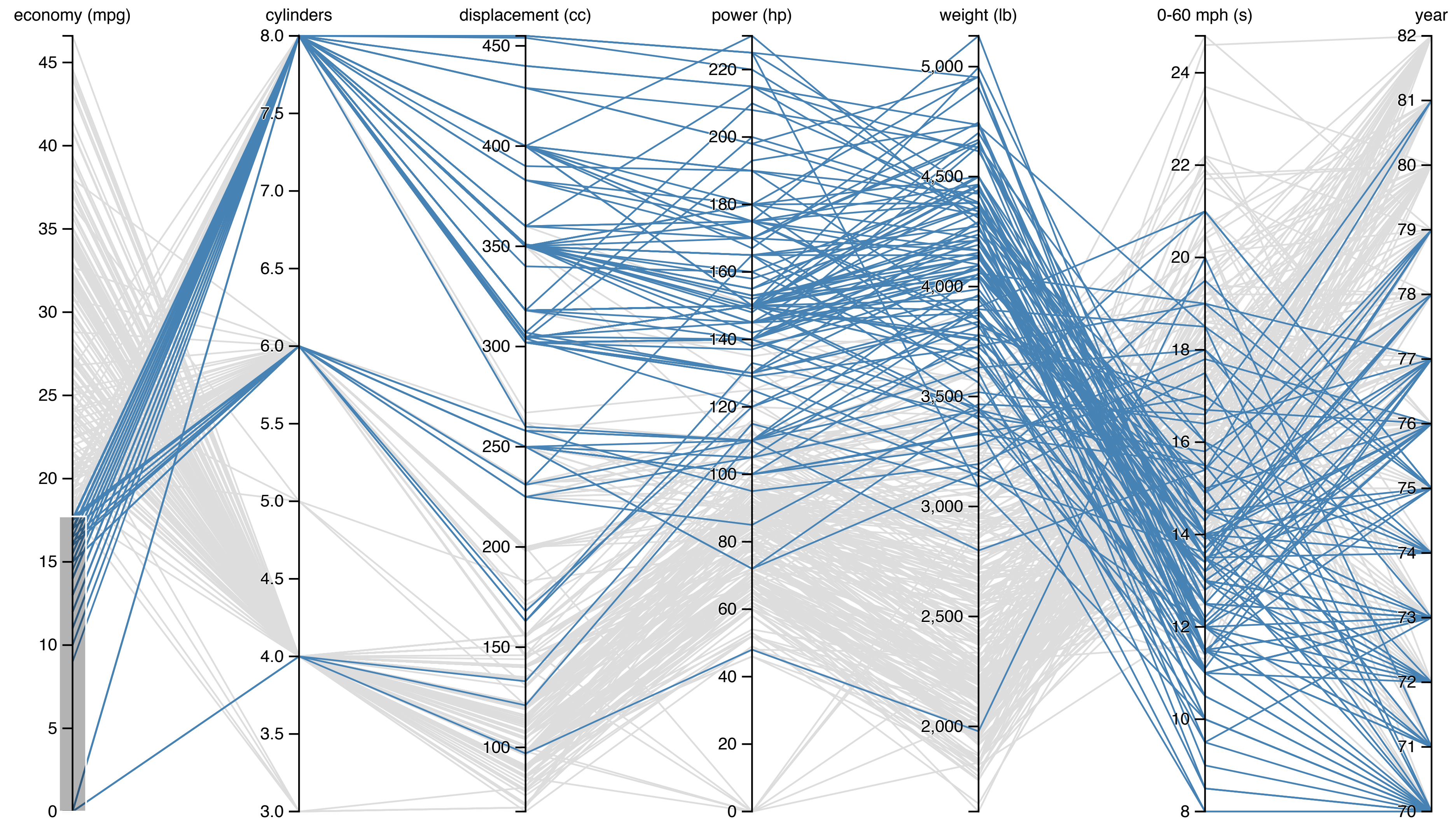


→ Geometry (Spatial)



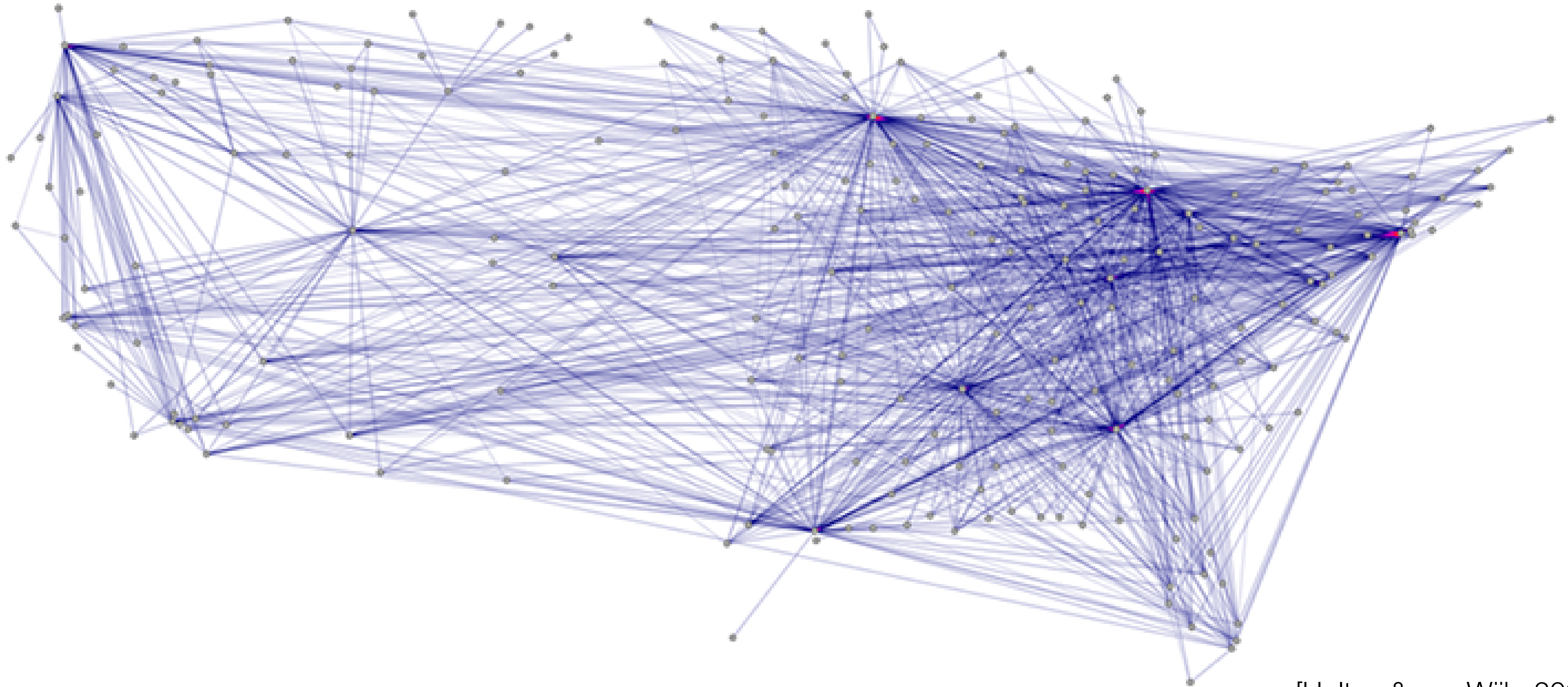
[Munzner (ill. Maguire), 2014]

Table Visualizations



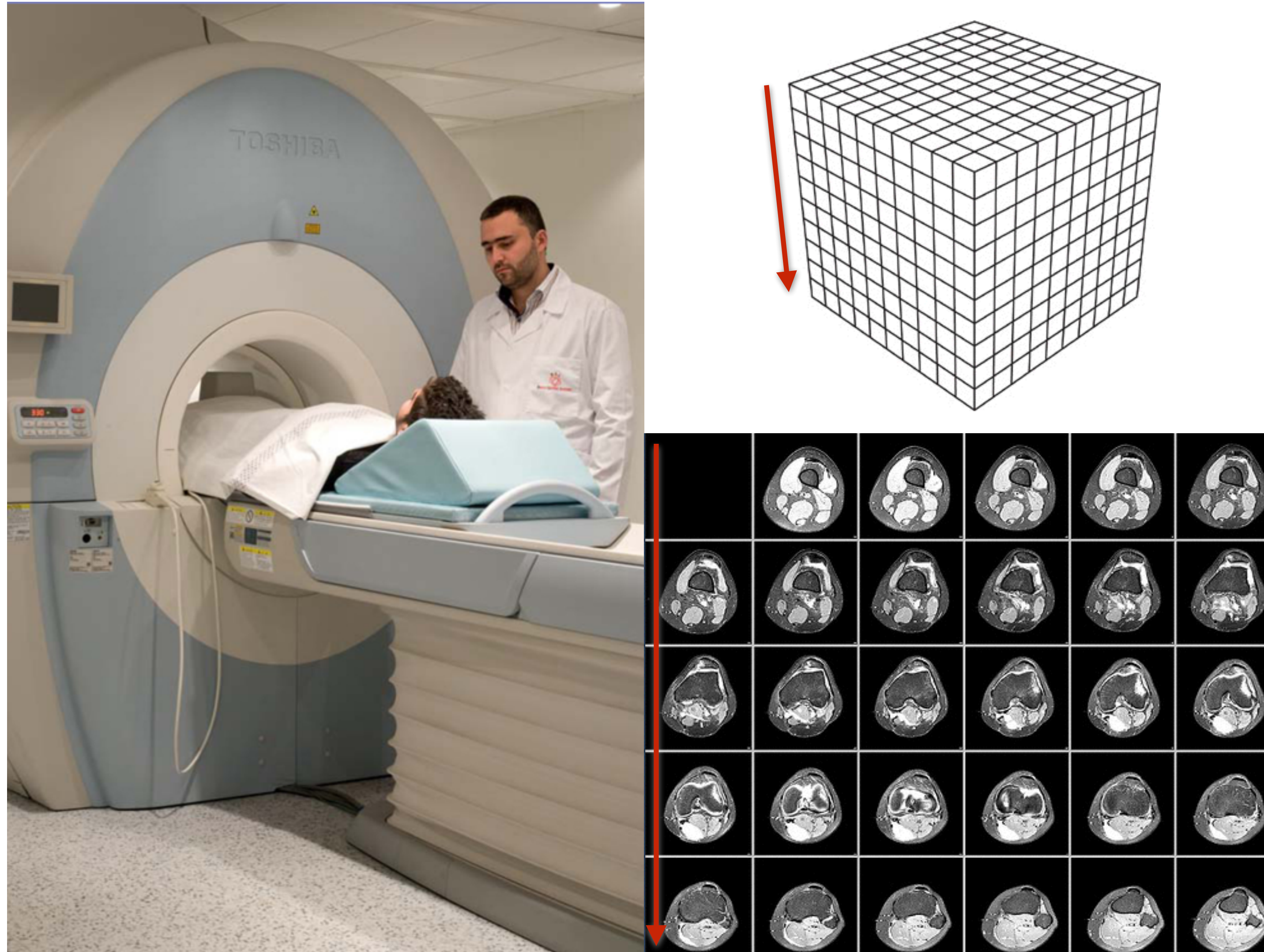
[M. Bostock, 2011]

Networks



[Holten & van Wijk, 2009]

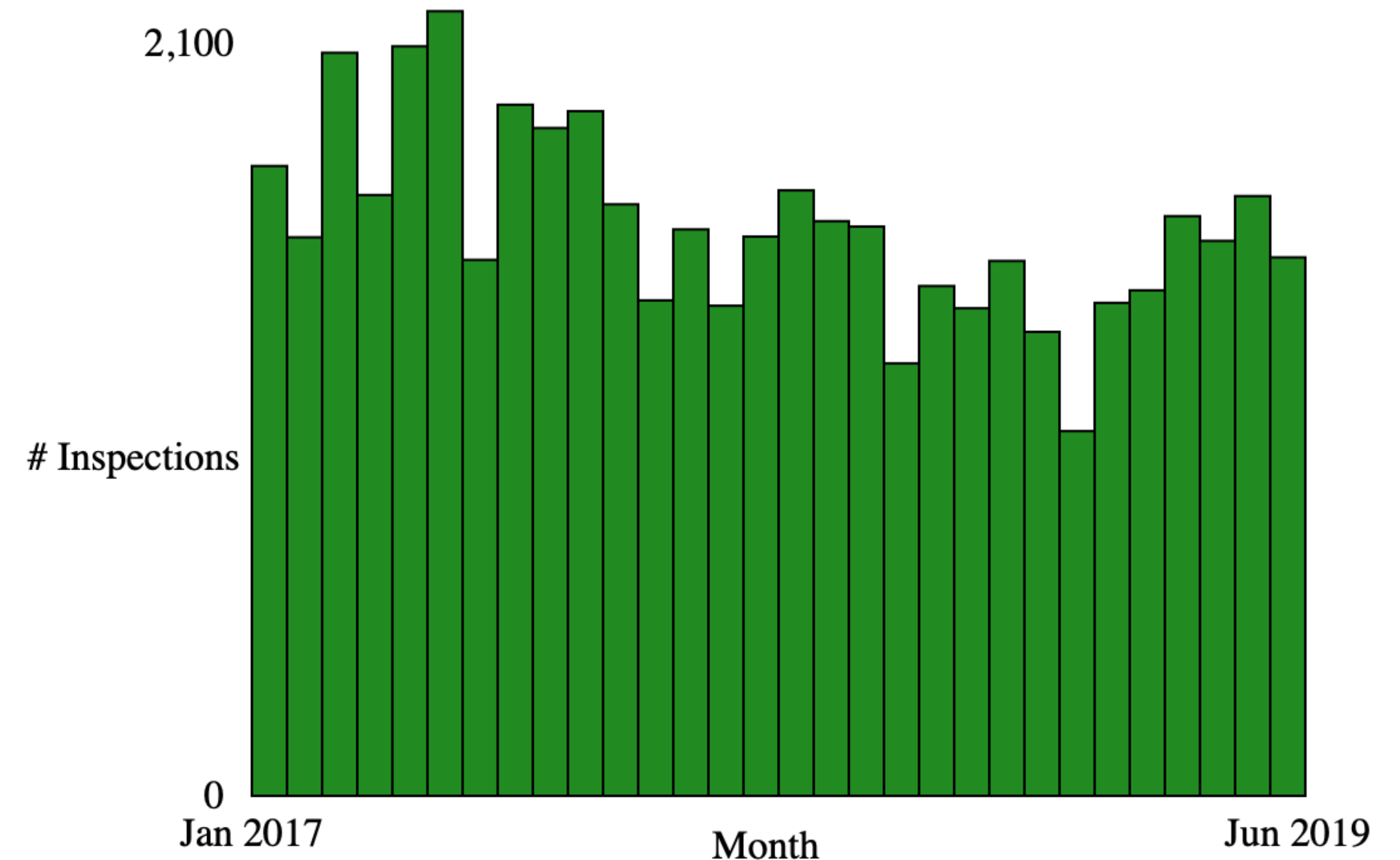
Fields



[via Levine, 2014]

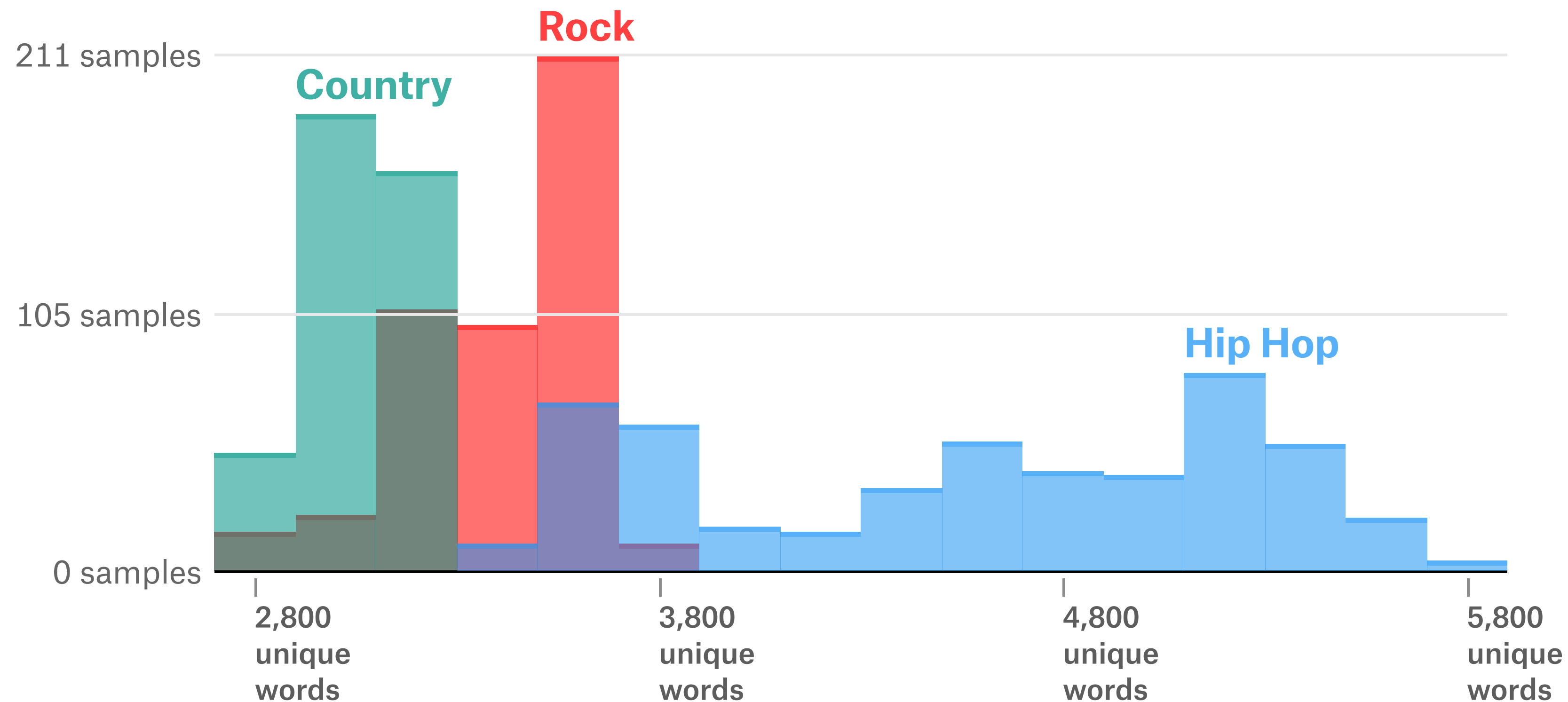
Assignment 2

- Link
- Three parts: table, horizontal bar chart, vertical bar chart
 - data processing
 - highlighting (CS 680)
- Vertical chart can be tricky
- Start early!
- Questions?



Sets & Lists

of Unique Words Used in 500 Random Samples of 35,000 Lyrics from Country, Rock, Hip Hop

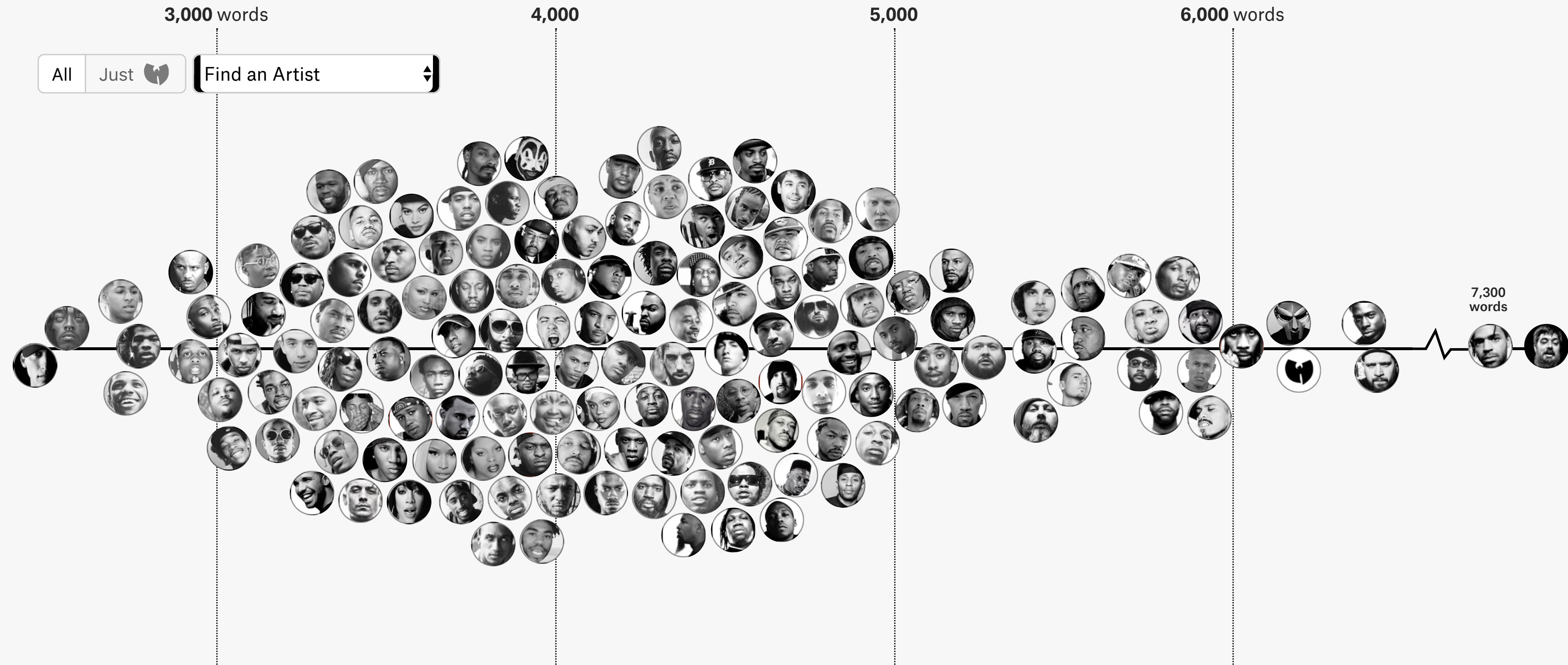


Raw Lyrics Data via John W. Miller

[M. Daniels, 2019]

Sets & Lists

of Unique Words Used Within Artist's First 35,000 Lyrics



Notes/sources:

All lyrics are via [Genius](#).

[M. Daniels, 2019]

# of Unique Words Used Within Artist's First 35,000 lyrics											
BY ERA ¹											
1980s 1990s 2000s 2010s											
Lil Uzi Vert NF	DMX 21 Savage A Boogie wit... Lil Baby Lil Durk Wiz Khalifa YG YoungBoy Nev...	Bone Thugs-n-... 50 Cent Juicy J Drake Future Kid Cudi Kid Ink Kodak Black Lil Yachty Logic Migos Travis Scott Young Thug	Foxy Brown Juvenile Master P Salt-n-Pepa Snoop Dogg Eve Gucci Mane Kanye West Lil Wayne Missy Elliot Trick Daddy Trina Young Jeezy Big Sean BoB Childish Gam... G-Eazy J Cole Machine Gun ... Meek Mill Nicki Minaj Russ	Run-D.M.C. 2Pac Big L Insane Clown... MC Lyte Scarface Three 6 Mafia UGK Dizzee Rascal Jadakiss Kano Lil' Kim Nelly Rick Ross T.I. 2 Chainz A\$AP Ferg Big KRIT Brockhampton Cupcakke Hopsin Jay Rock Kendrick Lamar Mac Miller Schoolboy Q Tyga Vince Staples	Biz Markie Ice T Rakim Brand Nubian Geto Boys Ice Cube Jay-Z Mobb Deep Outkast Public Enemy Cam'ron Eminem The Game Joe Budden Kevin Gates Royce da 5'9 Tech n9ne Twista Ab-Soul A\$AP Rocky Death Grips Denzel Curry \$uicideboy\$ Tyler the Cr... Wale	Beastie Boys Big Daddy Kane LL Cool J Busta Rhymes Cypress Hill De La Soul Fat Joe Gang Starr KRS-One Method Man A Tribe Call... Atmosphere Ludacris Lupe Fiasco Mos Def Murs Talib Kweli Xzibit Flatbush Zom... Joey BadA\$\$ Rittz	Common Das EFX E-40 Goodie Mob Nas Redman Brother Ali Action Bronson KAAN	Kool G Rap Kool Keith Raekwon CunninLynguists Sage Francis Watsky	Del the Funk... The Roots Blackalicious Canibus Ghostface Ki... Immortal Tec... Jean Grae Killah Priest RZA	GZA Wu-Tang Clan Jedi Mind Tr... MF DOOM	Aesop Rock Busdriver
<2,675 unique words	2,675-3,050 unique words	3,050-3,425 unique words	3,425-3,800 unique words	3,800-4,175 unique words	4,175-4,550 unique words	4,550-4,925 unique words	4,925-5,300 unique words	5,300-5,675 unique words	5,675-6,050 unique words	6,050-6,425 unique words	6,425+ unique words

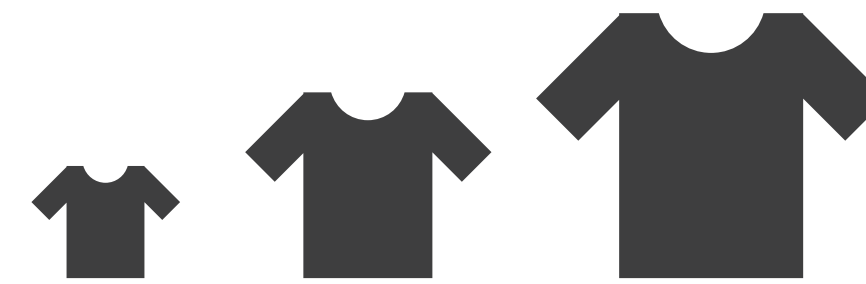
Attribute Types

→ Categorical

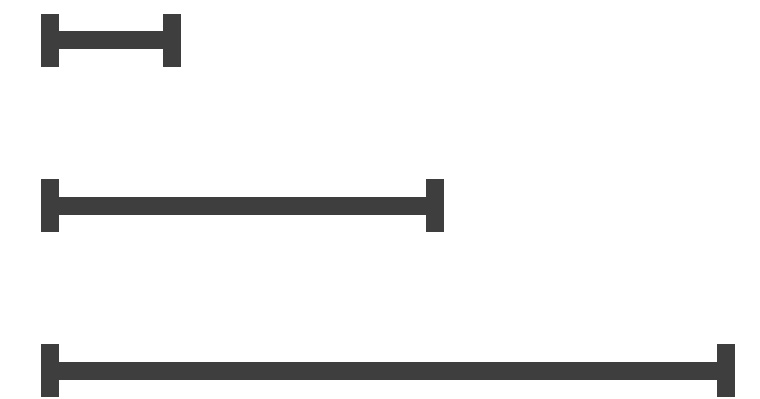


→ Ordered

→ *Ordinal*



→ *Quantitative*



Categorical, Ordinal, and Quantitative

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low		0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

quantitative
ordinal
categorical

Categorical, Ordinal, and Quantitative

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low		0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

quantitative
ordinal
categorical

Data Model vs. Conceptual Model

- Data Model: raw data that has a specific data type (e.g. floats):
 - Temperature Example: [32.5, 54.0, -17.3] (floats)
- Conceptual Model: how we think about the data
 - Includes semantics, reasoning
 - Temperature Example:
 - Quantitative: [32.50, 54.00, -17.30]

[via A. Lex, 2015]

Data Model vs. Conceptual Model

- Data Model: raw data that has a specific data type (e.g. floats):
 - Temperature Example: [32.5, 54.0, -17.3] (floats)
- Conceptual Model: how we think about the data
 - Includes semantics, reasoning
 - Temperature Example:
 - Quantitative: [32.50, 54.00, -17.30]
 - Ordered: [warm, hot, cold]

[via A. Lex, 2015]

Data Model vs. Conceptual Model

- Data Model: raw data that has a specific data type (e.g. floats):
 - Temperature Example: [32.5, 54.0, -17.3] (floats)
- Conceptual Model: how we think about the data
 - Includes semantics, reasoning
 - Temperature Example:
 - Quantitative: [32.50, 54.00, -17.30]
 - Ordered: [warm, hot, cold]
 - Categorical: [not burned, burned, not burned]

[via A. Lex, 2015]

Ordering Direction

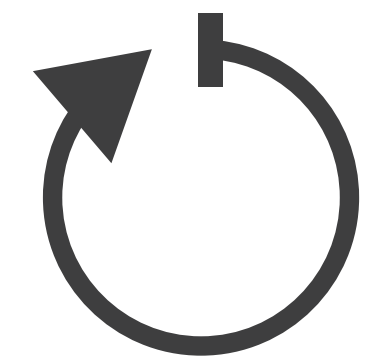
→ Sequential



→ Diverging

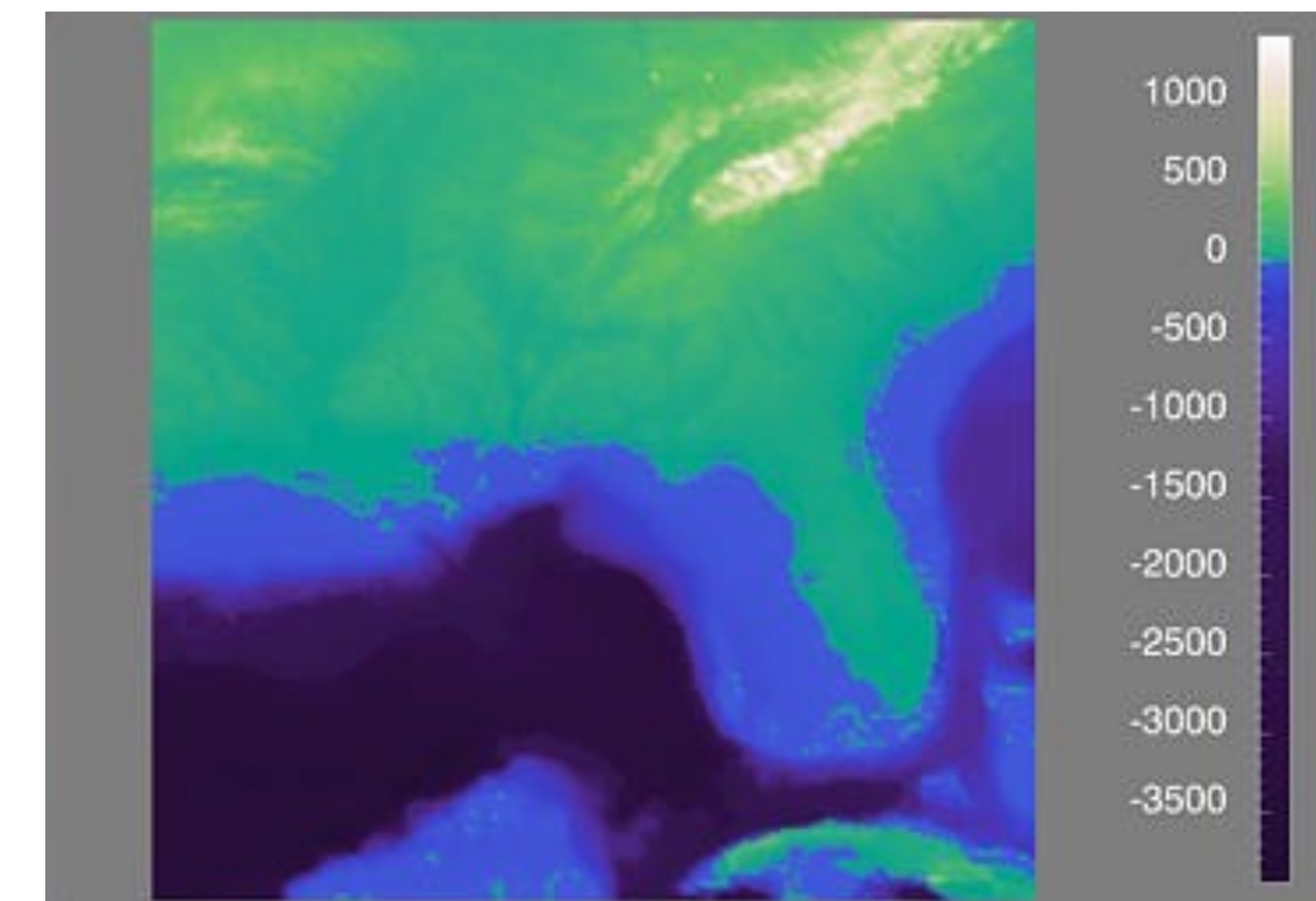
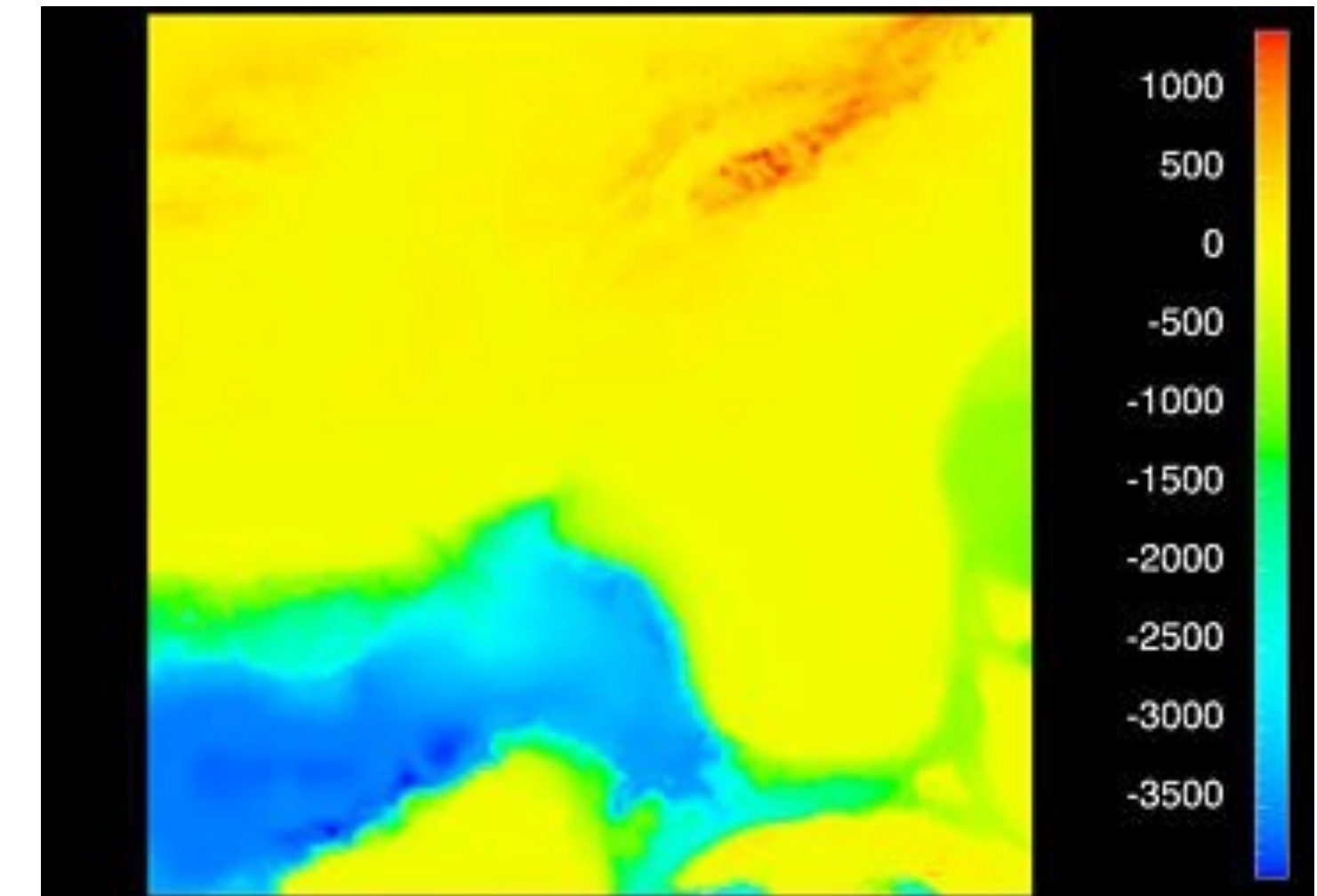


→ Cyclic



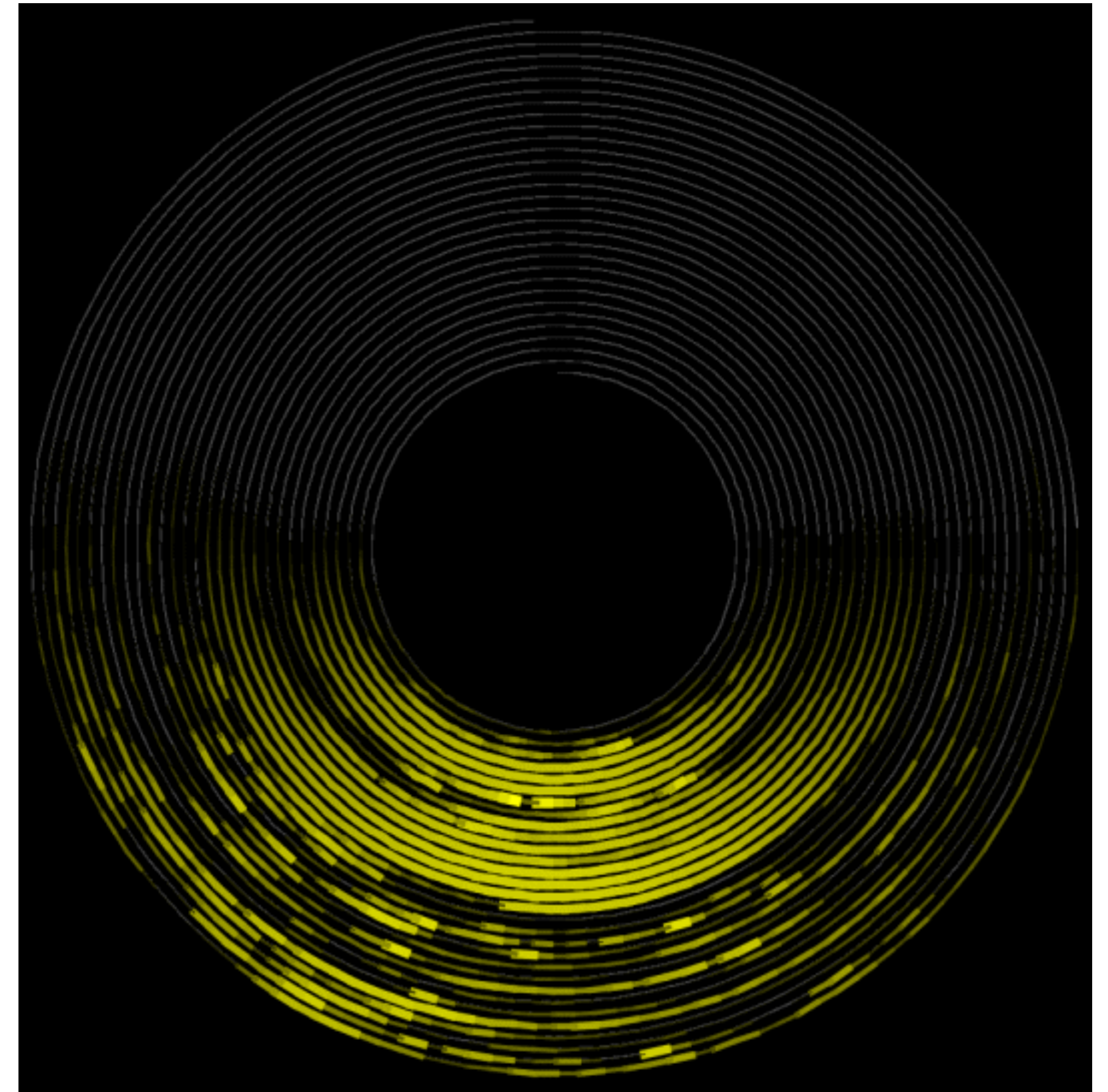
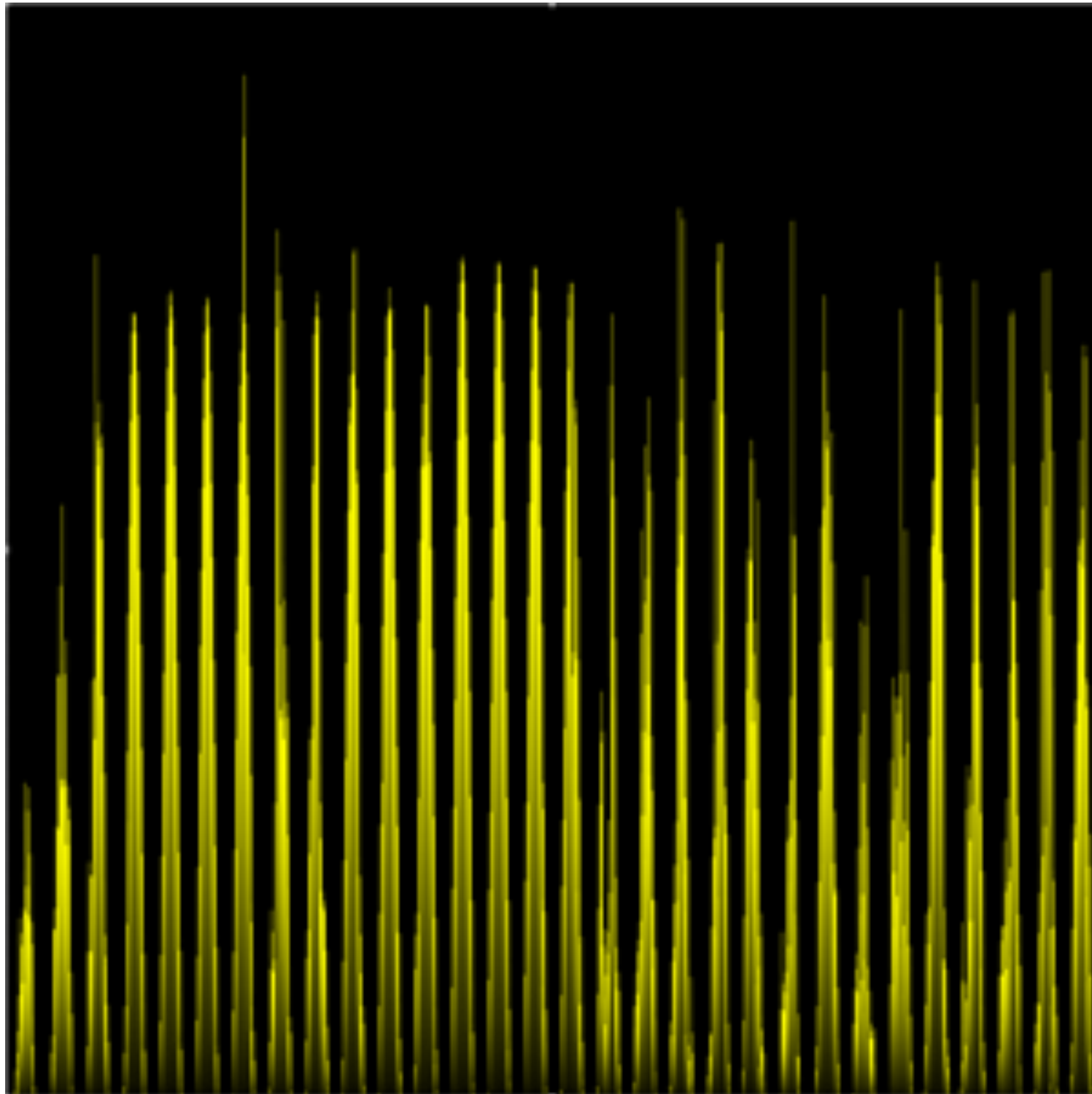
Sequential and Diverging Data

- Sequential: homogenous range from a minimum to a maximum
 - Examples: Land elevations, ocean depths
- Diverging: can be deconstructed into two sequences pointing in opposite directions
 - Has a **zero point** (not necessary 0)
 - Example: Map of both land elevation and ocean depth



[Rogowitz & Treinish, 1998]

Cyclic Data



[Sunlight intensity, Weber et al., 2001]

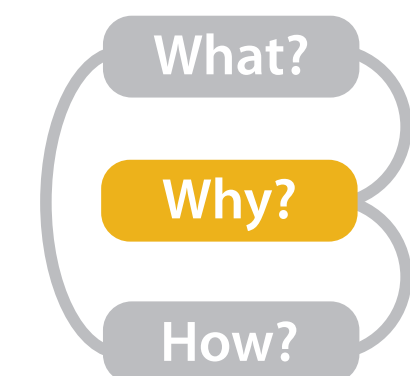
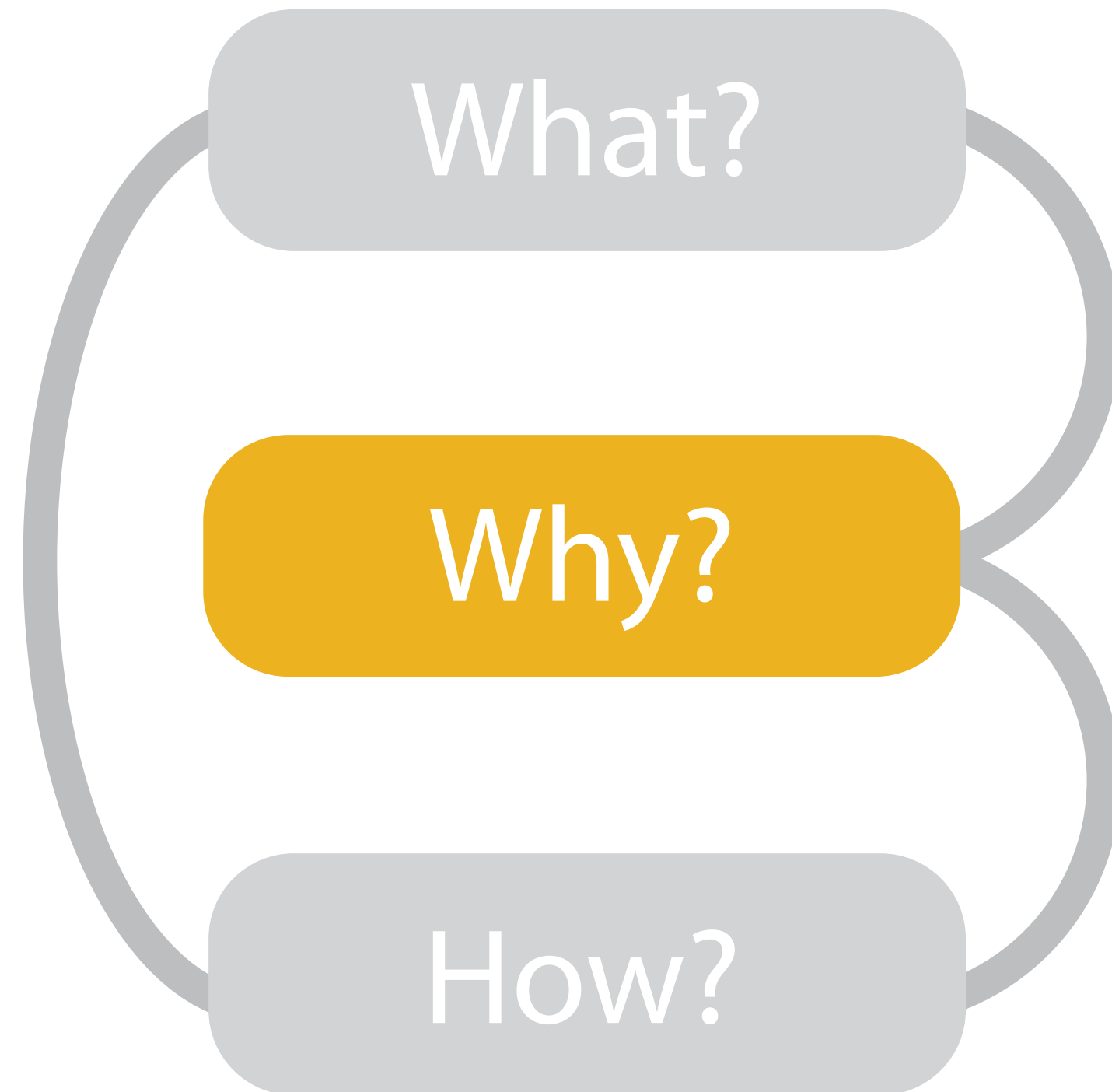
“Computer-based visualization systems provide visual representations of datasets designed to help people carry out **tasks** more effectively.”

— T. Munzner

Tasks

- Why? Understand data, but what do I want to do with it?
- Levels: High (Produce/Consume), Mid (Search), Low (Queries)
- Another key concern: Who?
 - Designer <-> User (A spectrum)
 - Complex <-> Easy to Use
 - General <-> Context-Specific
 - Flexible <-> Constrained
 - Varied Data <-> Specific Data

Tasks



[Munzner (ill. Maguire), 2014]

Actions: Analyze

→ Consume

→ *Discover*



→ *Present*

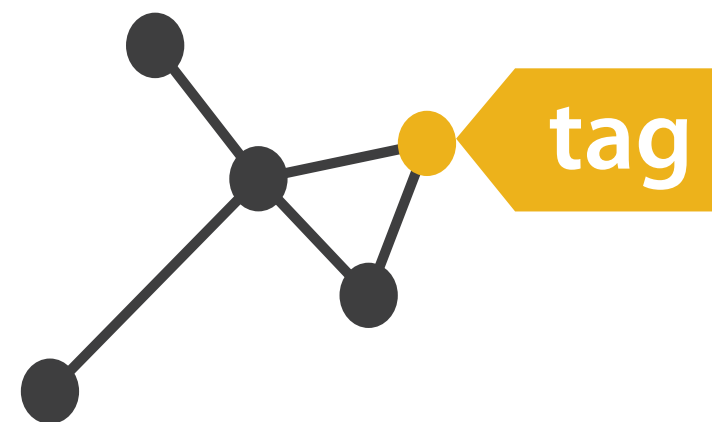


→ *Enjoy*

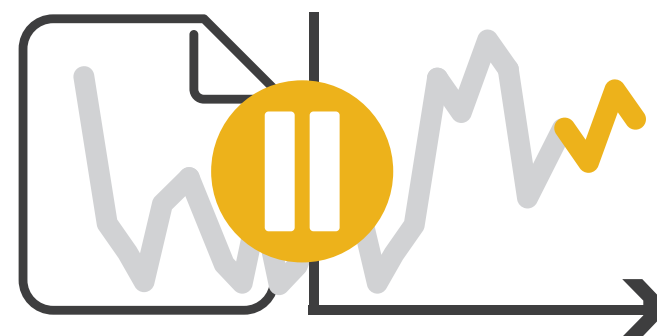


→ Produce

→ *Annotate*



→ *Record*



→ *Derive*



[Munzner (ill. Maguire), 2014]

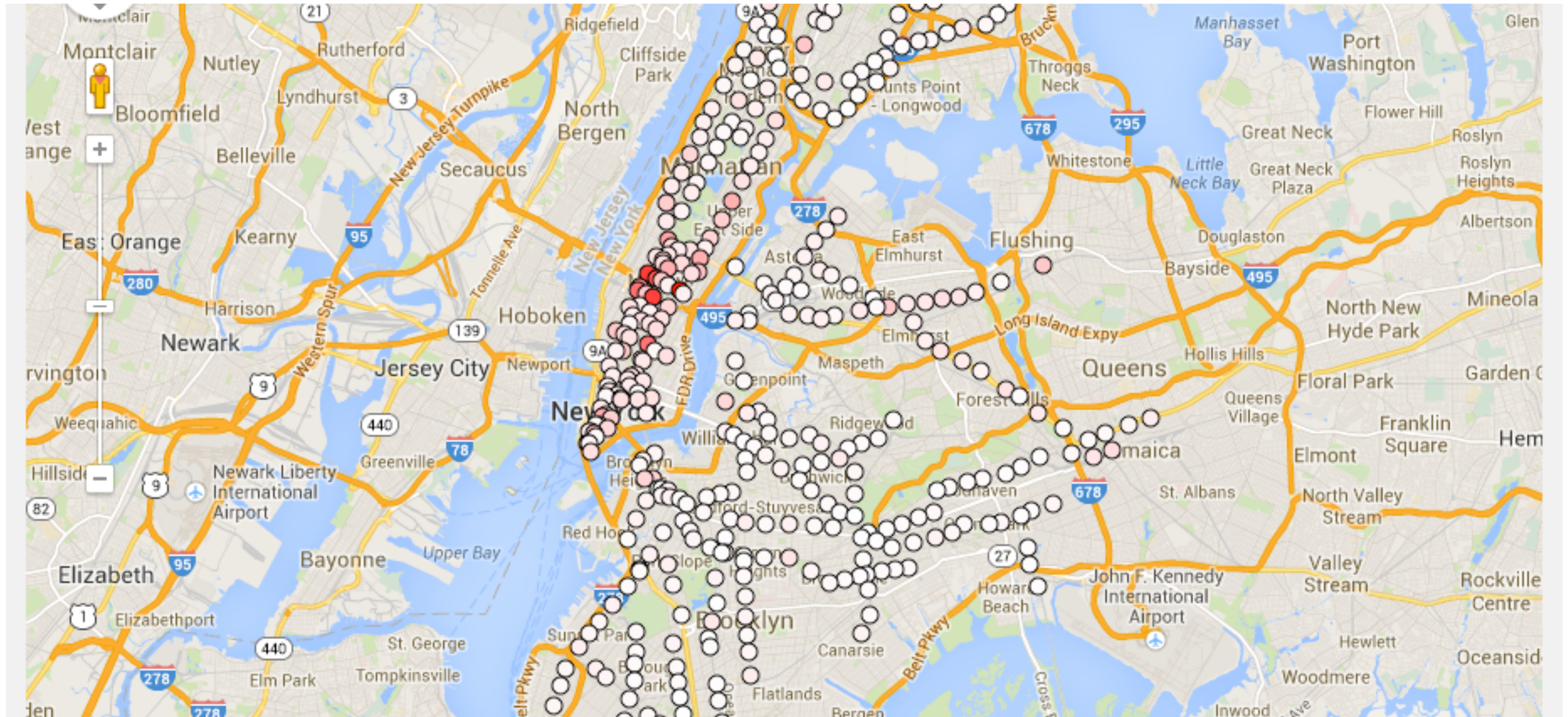
Visualization for Consumption

- Discover new knowledge
 - Generate new hypothesis or verify existing one
 - Designer doesn't know what users need to see
 - "why doesn't dictate how"
- Present known information
 - Presenter already knows what the data says
 - Wants to communicate this to an audience
 - May be static but not limited to that
- Enjoy
 - Similar to discover, but without concrete goals
 - May be enjoyed differently than the original purpose

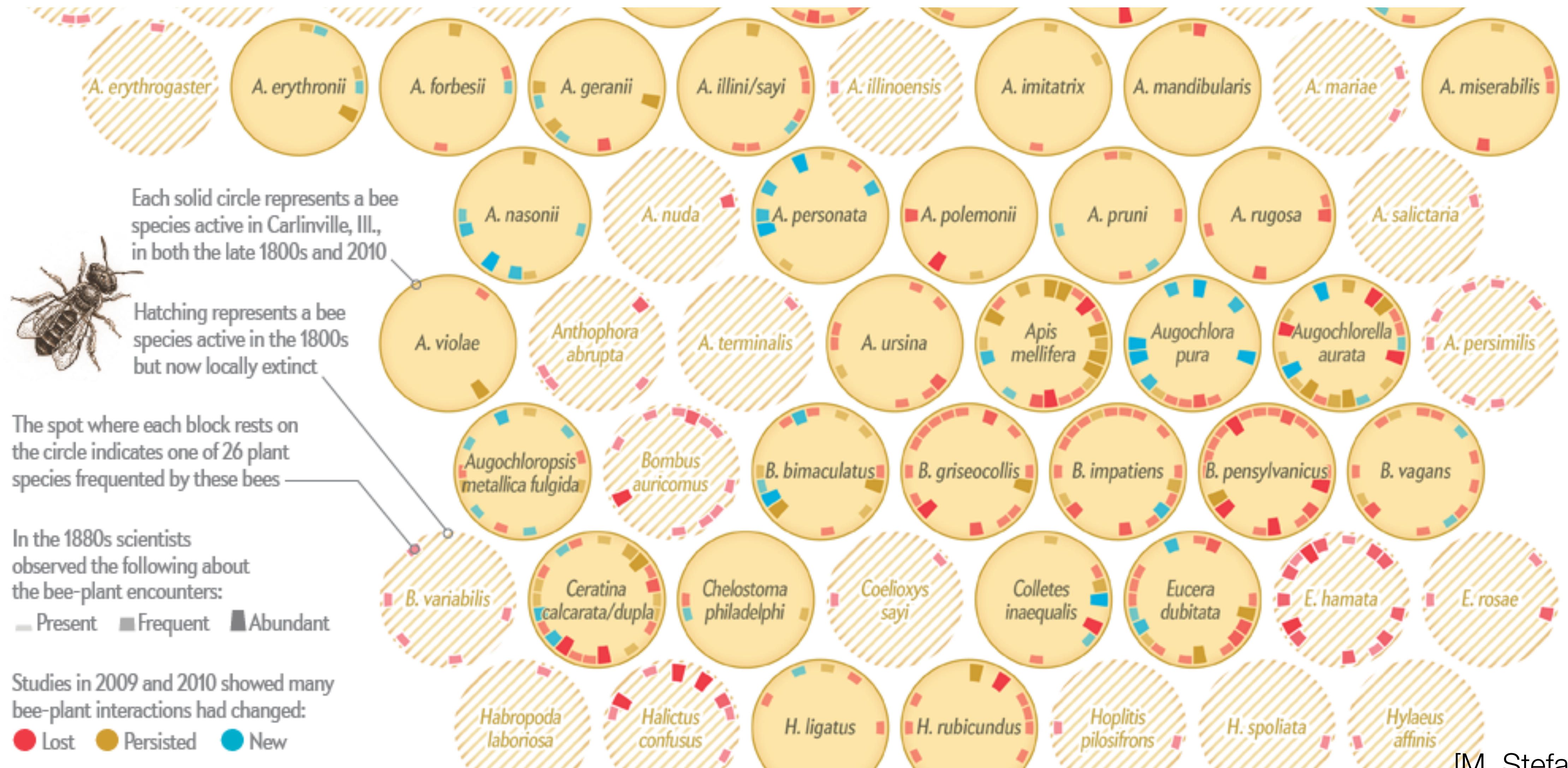
Asking good **questions** is very important

Answers often lead to **more** questions

Explore MTA Fare Data

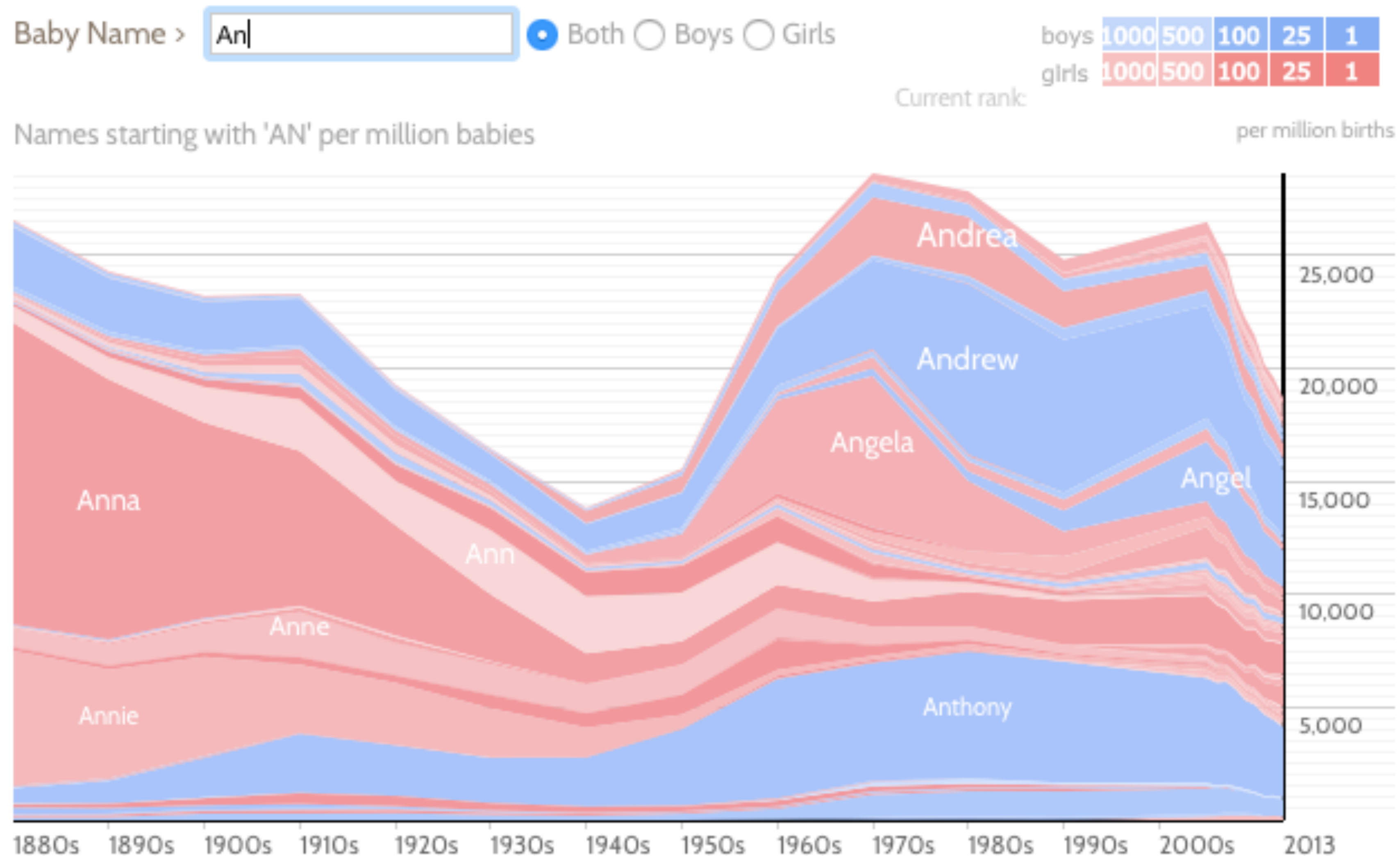


Present Known Information



[M. Stefaner, 2013]

Enjoy Visualizations of Names



[Wattenberg, 2005, www.babynamewizard.com]

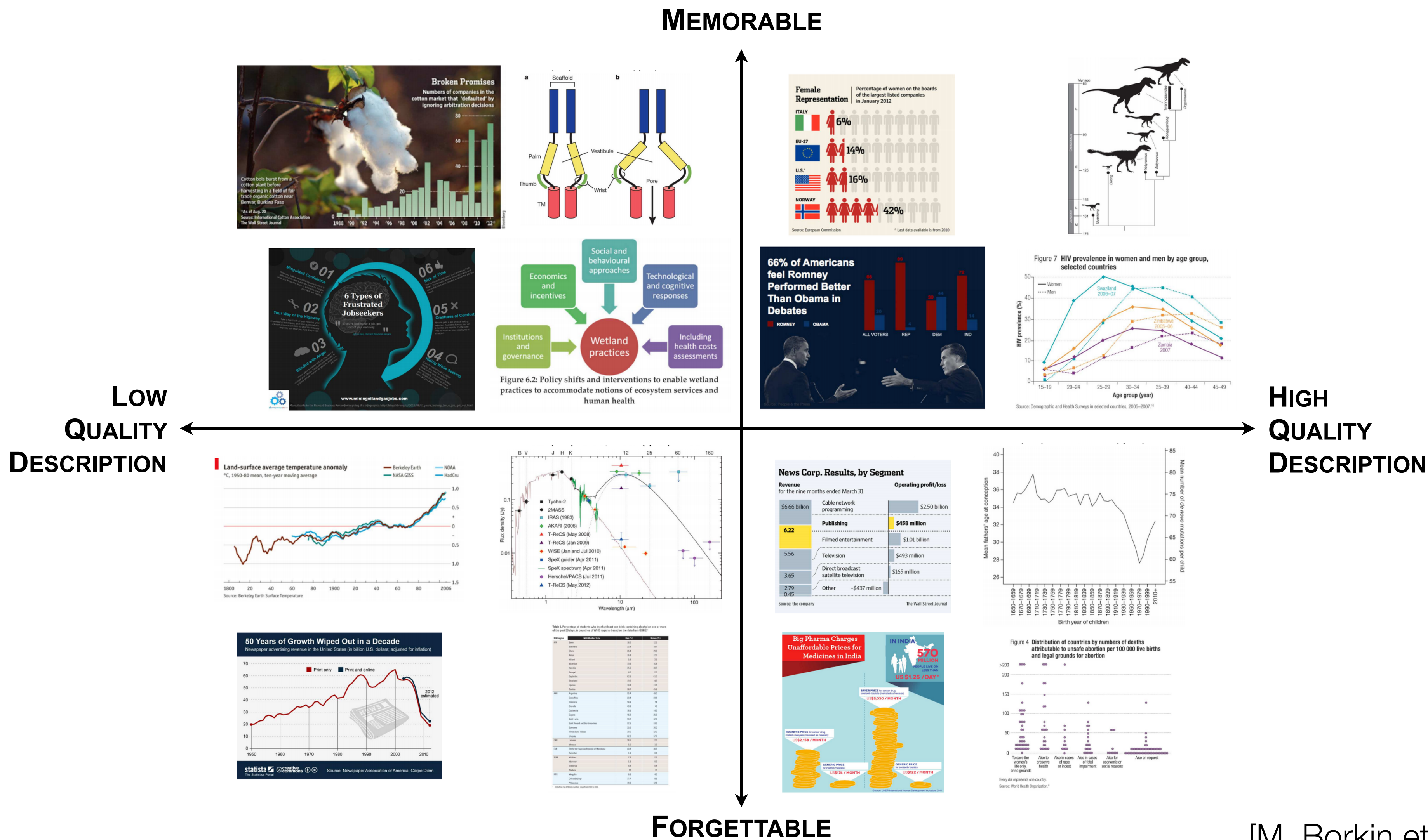
“[W]e scientists now understand how important emotion is to everyday life, how valuable. Sure, utility and usability are important, but without fun and pleasure, joy and excitement, and yes, anxiety and anger, fear and rage, our lives would be incomplete.”
—D. Norman (Emotional Design)

Measuring User Experience in Visualization

- Memorability: Capability of maintaining and retrieving information [J. Brown et al., 1977]
- Engagement: Emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource. [S. Attfield et al., 2011]
- Enjoyment: Feeling that causes a person to experience pleasure. Pleasure is recognized with occurrent happiness and excitement, which can be explained in terms of belief, desire, and thought. [W. A. Davis, 1982]

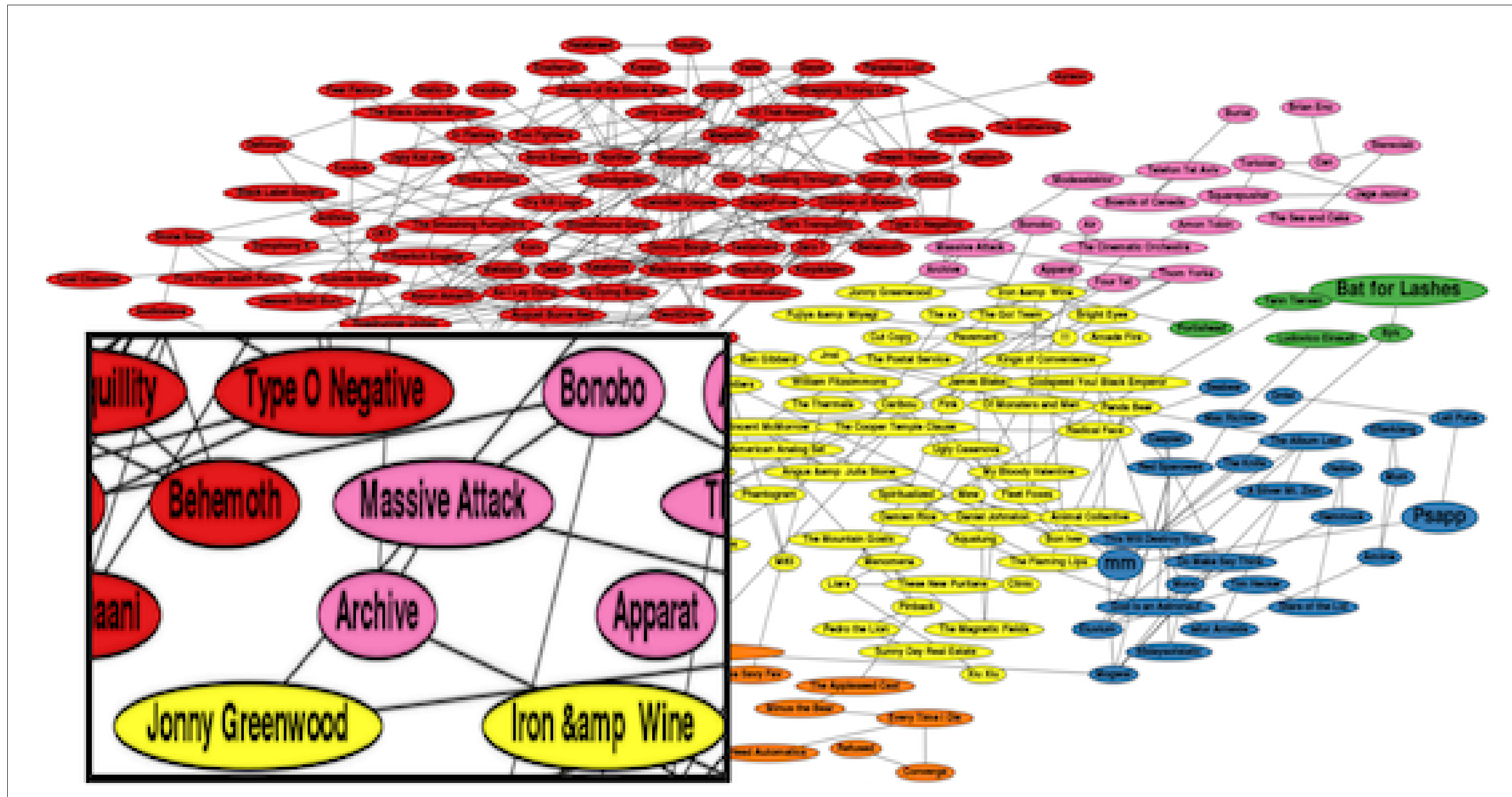
[B. Saket et al., BELIV 2016]

Memorability



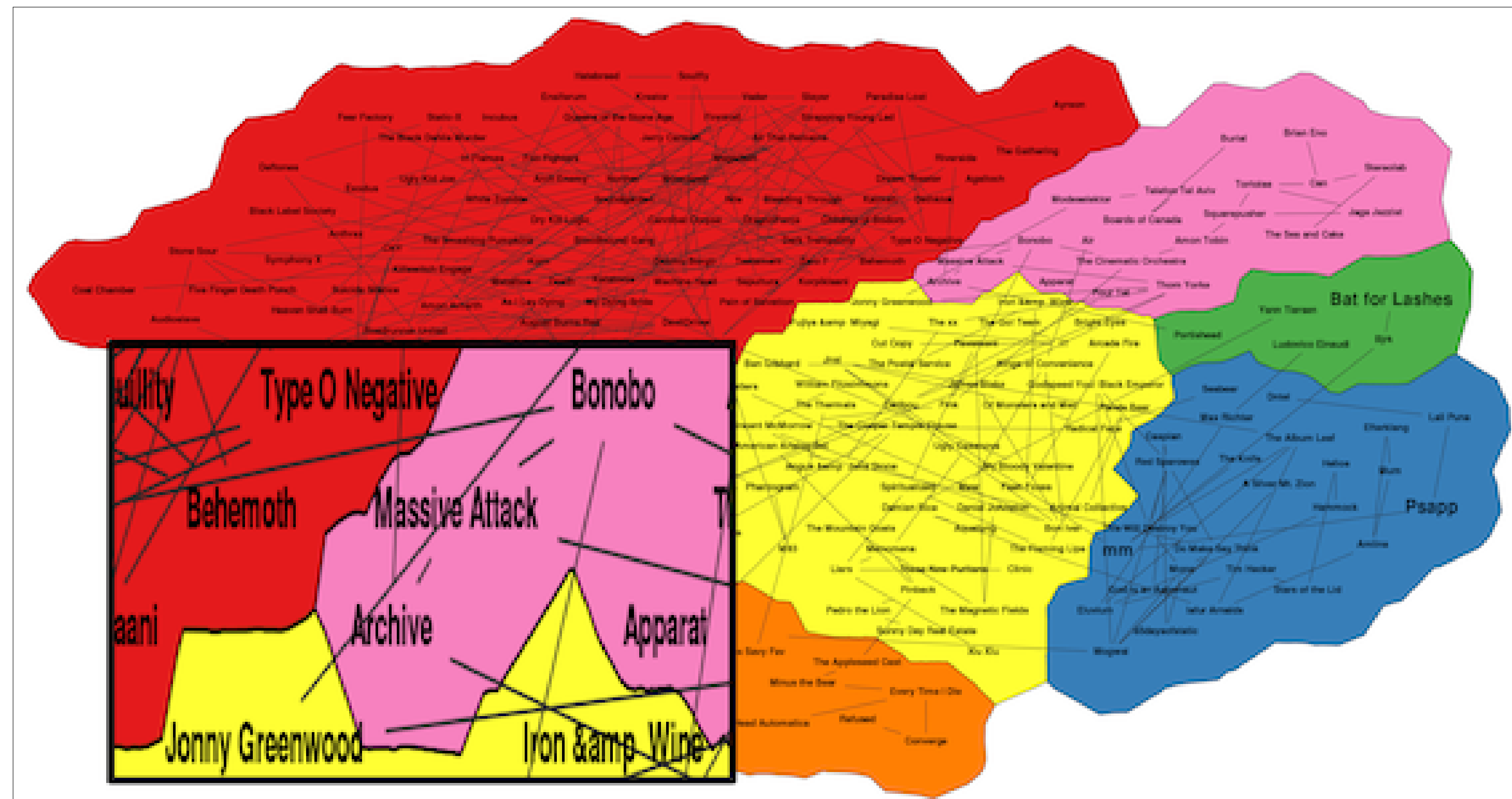
[M. Borkin et al., InfoVis 2015]

Memorability: Maps instead of Networks



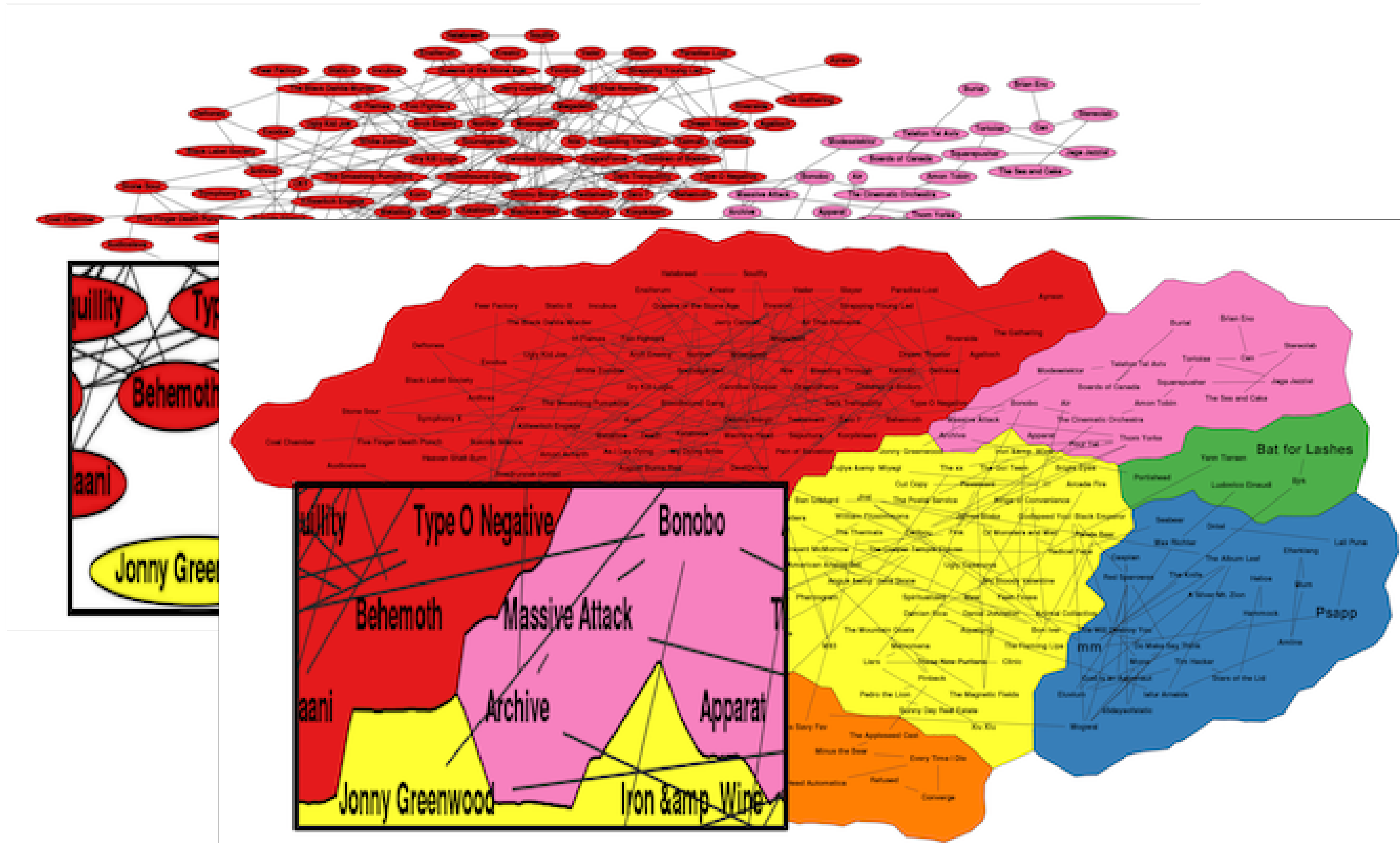
[B. Saket et al., EuroVis 2015]

Memorability: Maps instead of Networks



[B. Saket et al., EuroVis 2015]

Memorability: Maps instead of Networks

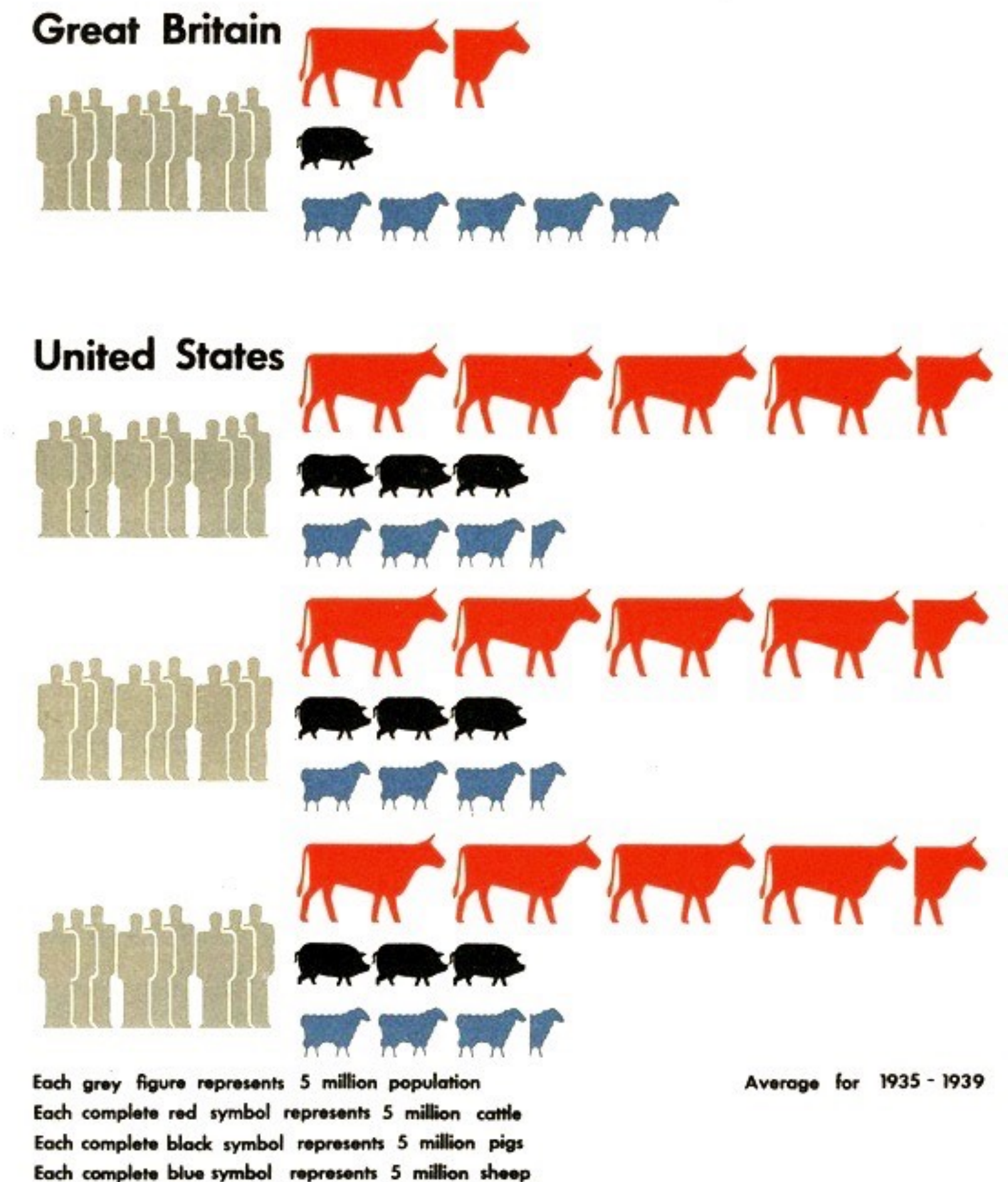


[B. Saket et al., EuroVis 2015]

ISOTYPE Visualizations

- Study [Haroz et al., 2015]
 - Want quick understanding and ease of remembering
 - Does ISOTYPE help?
- Results:
 - Stacked icons allow both length and quantity encoding
 - Icons are more memorable
 - Images that aren't used to show data are distracting

Population and Live Stock



[Image by O. and M. Neurath, Study by S. Haroz et al., 2015]

Memorability

- Capability of maintaining and retrieving information
[J. Brown et al., 1977]
- How to measure?
 - test users
- How long?
 - short-term, intermediate, or long-term?
- What types of visualizations?
 - bar/line/pie, networks, graphs, etc.

[B. Saket et al., BELIV 2016]

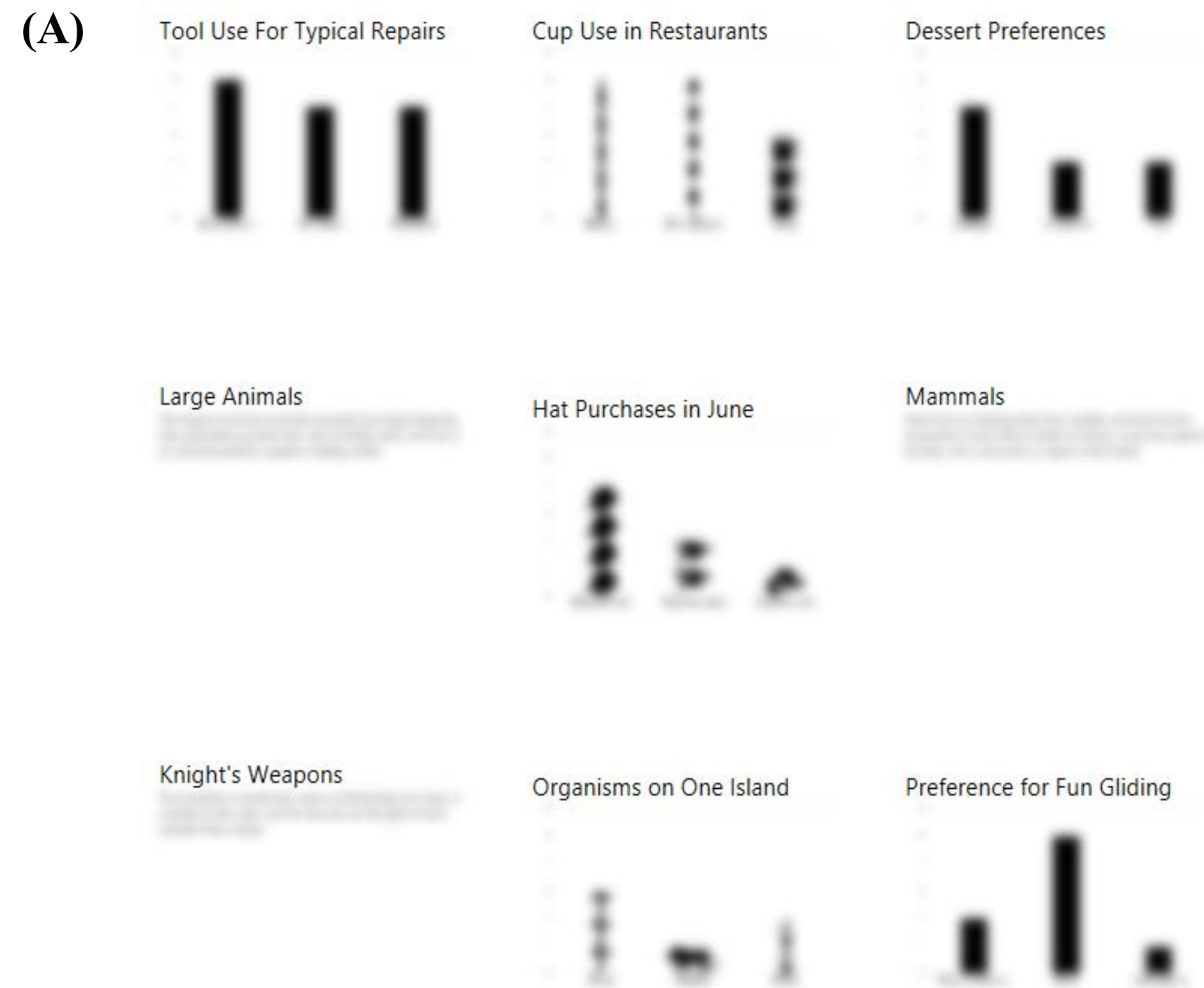
Engagement

- "Emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource." [S. Attfield et al., 2011]
- How to measure? total time spent looking at a chart

[B. Saket et al., BELIV 2016]

Measuring Engagement

Grid is blurred,
click for detail

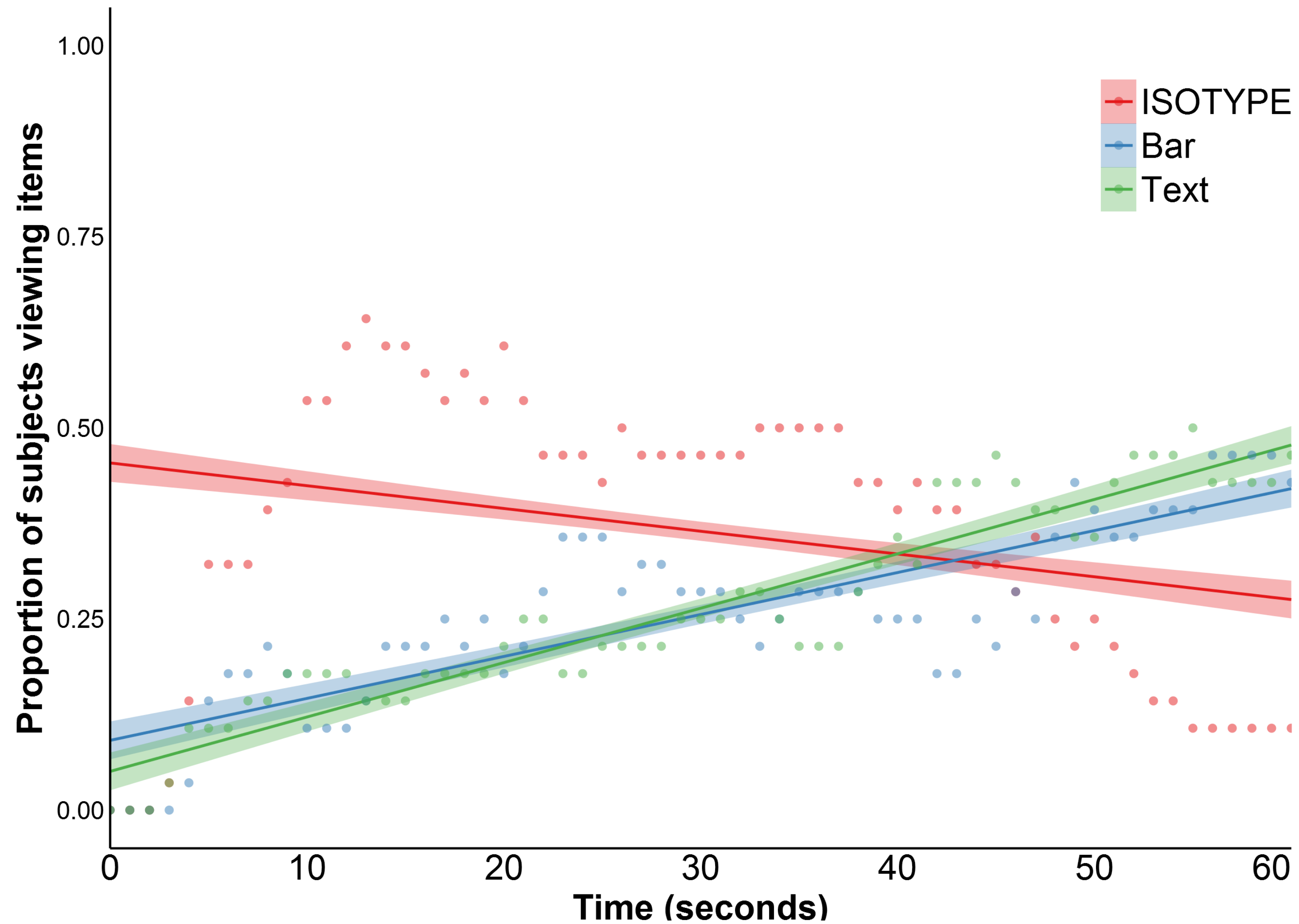


(B) **Mammals**

Mammals are distinguished from reptiles and birds by the possession of hair, three middle ear bones, mammary glands in females, and a neocortex (a region of the brain).

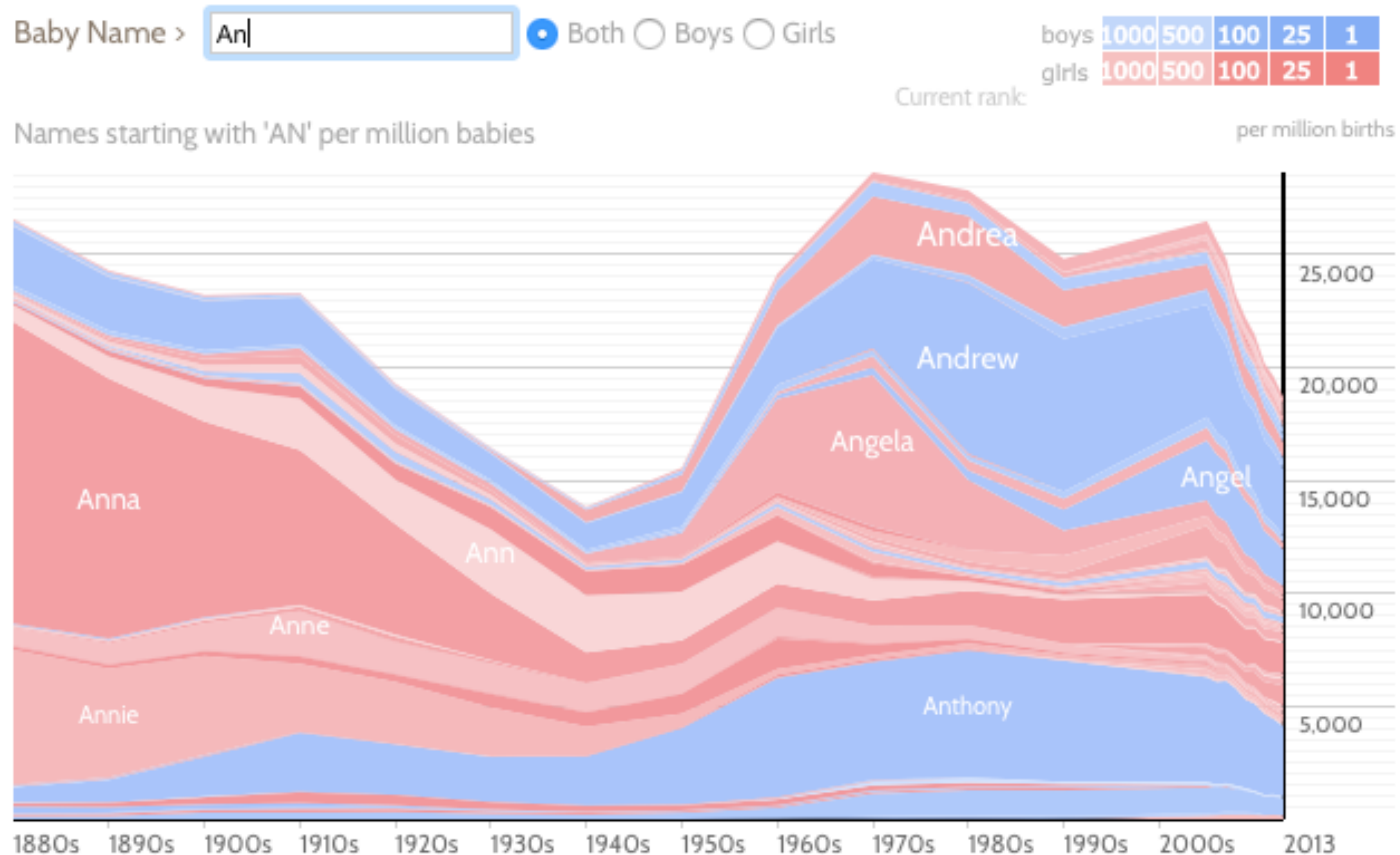
[S. Haroz et al., 2015]

Measuring Engagement



[S. Haroz et al., 2015]

Enjoyment: Name Voyager



[Wattenberg, 2005, www.babynamewizard.com]

Measuring Enjoyment

- Difference from engagement (e.g. may be for a job)
- Self-reporting (e.g. comparison between different charts)
- Measure why someone enjoys a visualization:
 - Challenge
 - Focus
 - Clarity
 - Feedback
 - Control
 - Immersion

[B. Saket et al., BELIV 2016]

“Visualizations don’t need to be designed for memorability – they need to be designed for comprehension. For most visualizations, the comprehension that they provide need only last until the decision that it informs is made. Usually, that is only a matter of seconds.”
— S. Few

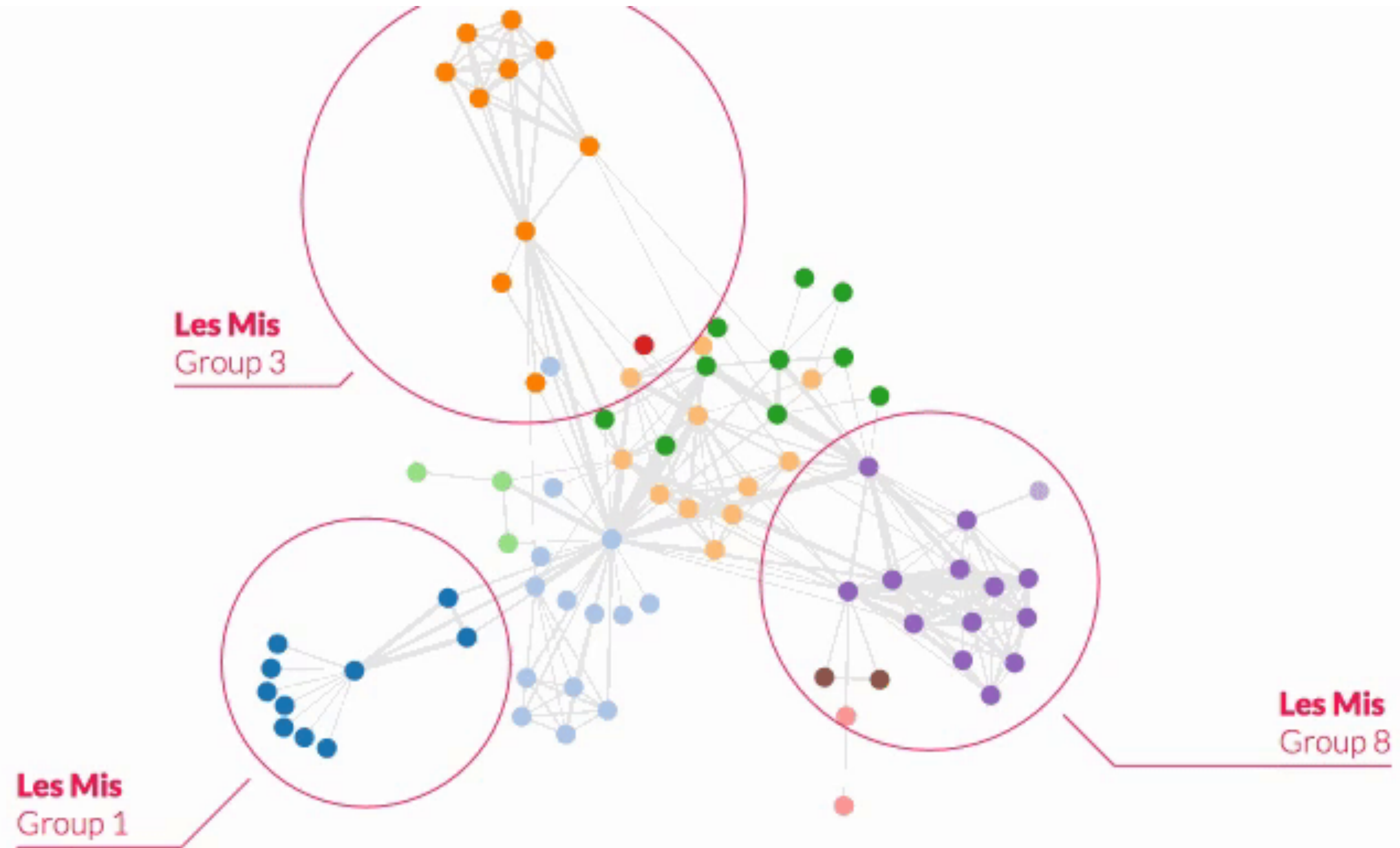
Reaction

- B. Jones (paraphrased): People make decisions using visualizations but this isn't instantaneous like robots or algorithms; they often chew on a decision for a while
- R. Kosara: there are cases where people benefit from remembering a visualization (e.g. health-related visualization)
- Are there tradeoffs between the characteristics?

Visualization for Production

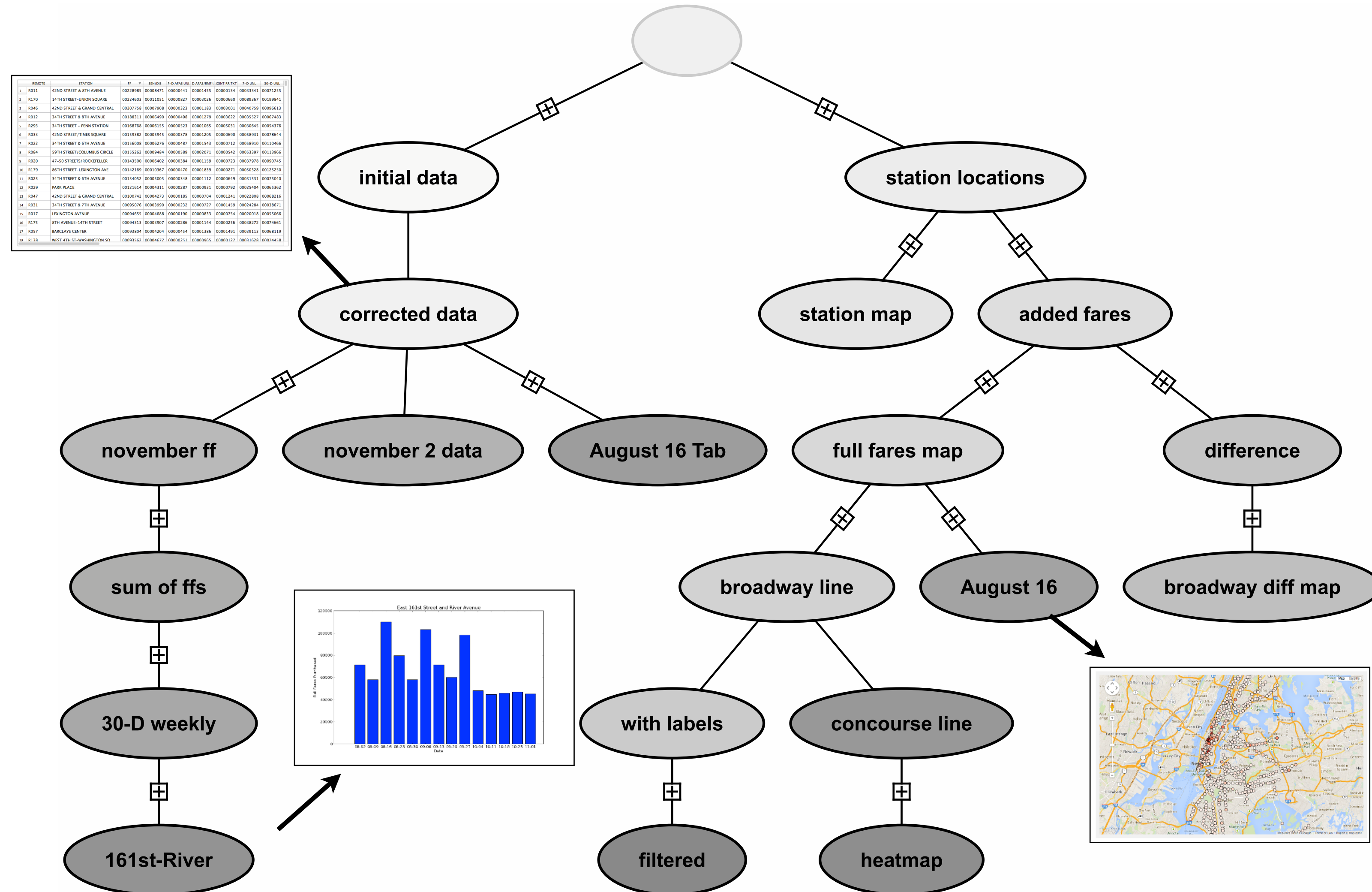
- Generate new material
- Annotate
- Record
- Derive (Transform)

Annotation: Circle Annotations

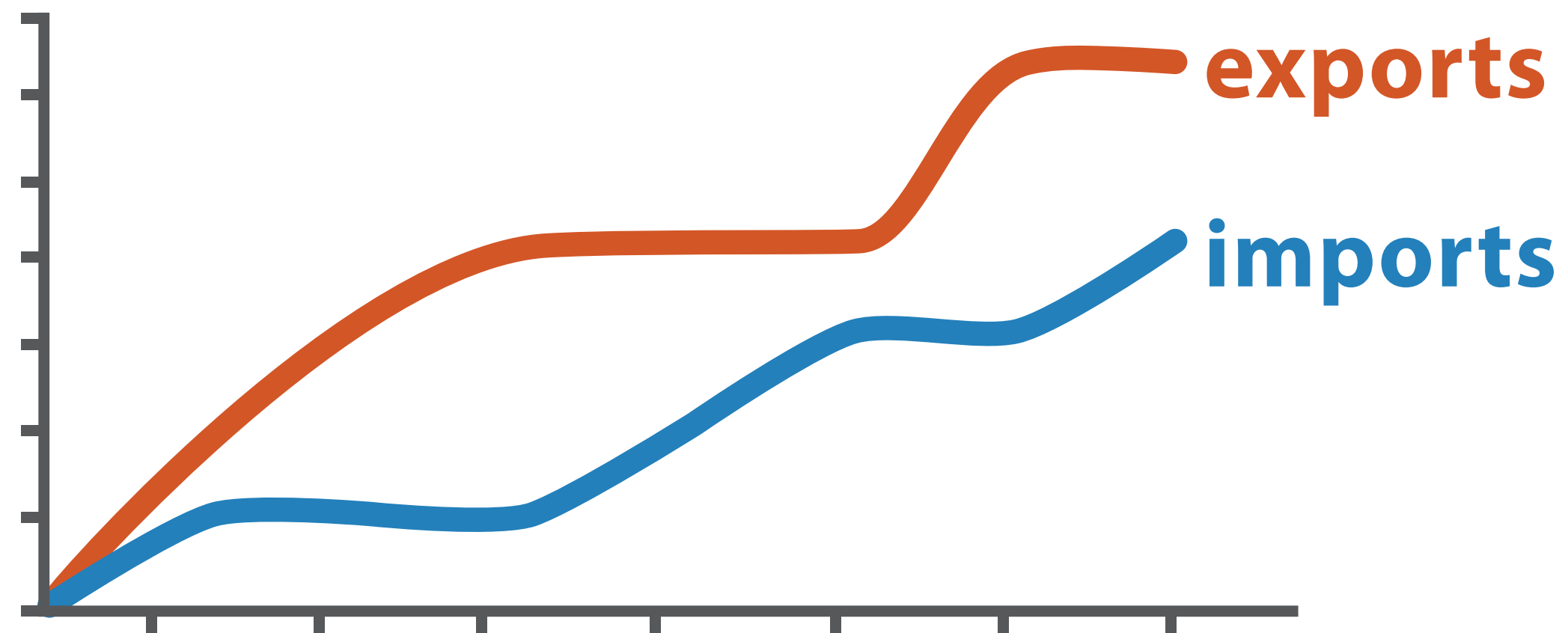


[S. Lu, 2017]

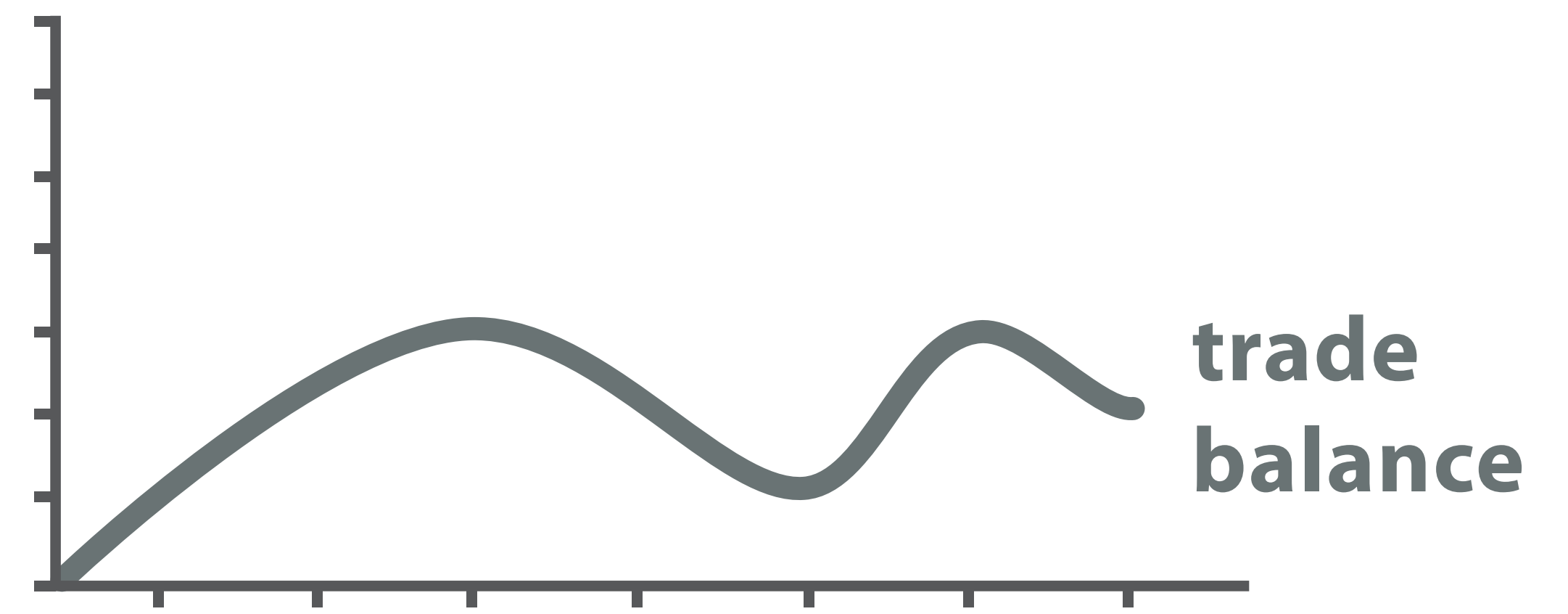
Record: Provenance of MTA Data Exploration



Derived Data



Original Data



$$\text{trade balance} = \text{exports} - \text{imports}$$

Derived Data

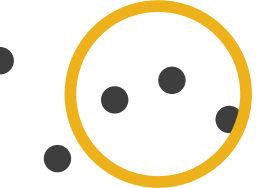
[Munzner (ill. Maguire), 2014]

Visualization for Production

- Generate new material
- Annotate:
 - Add more to a visualization
 - Usually associated with text, but can be graphical
- Record:
 - Persist visualizations for historical record
 - Provenance (graphical histories): how did I get here?
- Derive (Transform):
 - Create new data
 - Create derived attributes (e.g. mathematical operations, aggregation)

Actions: Search

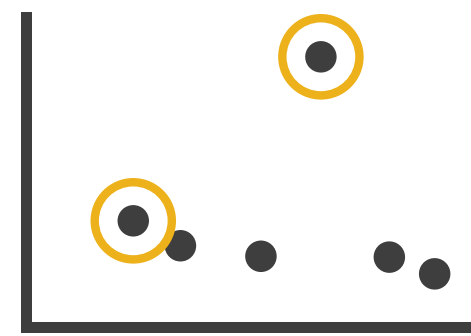
- What does a user know?
 - Lookup: check bearings
 - Locate: find on a map
 - Browse: what's nearby
 - Explore: where to go
 - Patterns

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

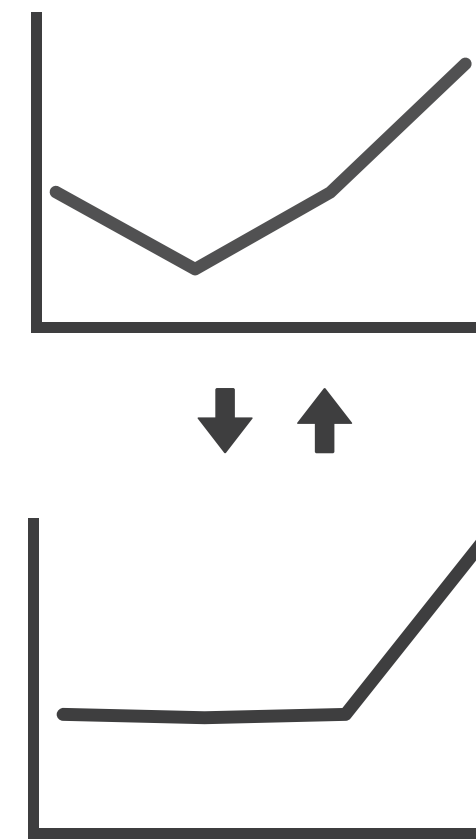
[Munzner (ill. Maguire), 2014]

Query

→ Identify



→ Compare



→ Summarize



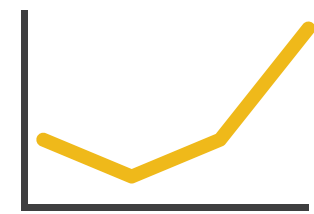
- Number of targets: One, Some (Often 2), or All
- Identify: characteristics or references
- Compare: similarities and differences
- Summarize: overview of everything

[Munzner (ill. Maguire), 2014]

Targets

➔ ALL DATA

➔ Trends



➔ Outliers



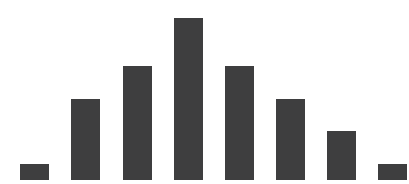
➔ Features



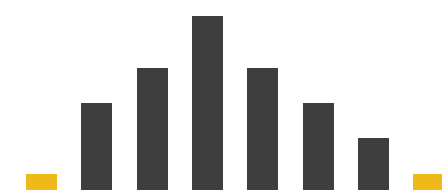
➔ ATTRIBUTES

➔ One

➔ *Distribution*



↓ *Extremes*

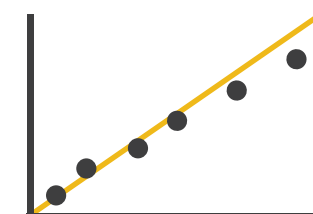


➔ Many

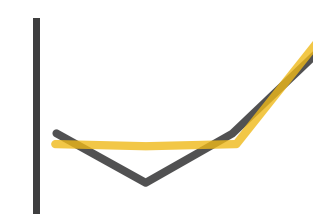
➔ *Dependency*



➔ *Correlation*

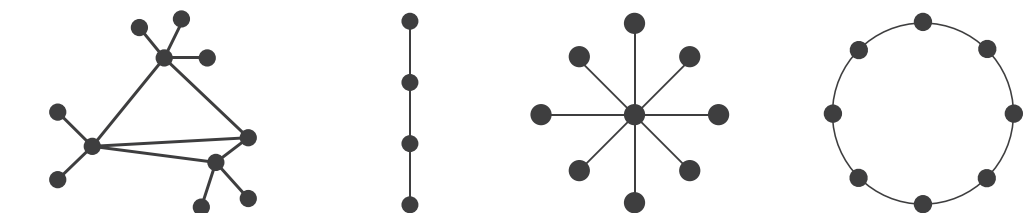


➔ *Similarity*

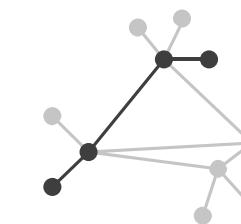


➔ NETWORK DATA

➔ Topology

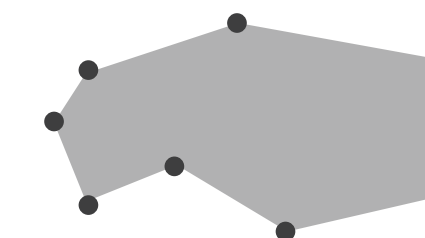


➔ *Paths*



➔ SPATIAL DATA

➔ Shape



[Munzner (ill. Maguire), 2014]