

MassMutual DSDP 2018: INTERACTION

June 19, 2018

R. Jordan Crouser

Assistant Professor of Computer Science

Smith College

Outline

- **Mini-lab: #badviz**
- **Interaction and Analysis**
 - Definition
 - Interaction with data and problem space
 - Interaction with visual interfaces
- **BREAK**
- **Introduction to plotly**

Mini-lab: #badviz

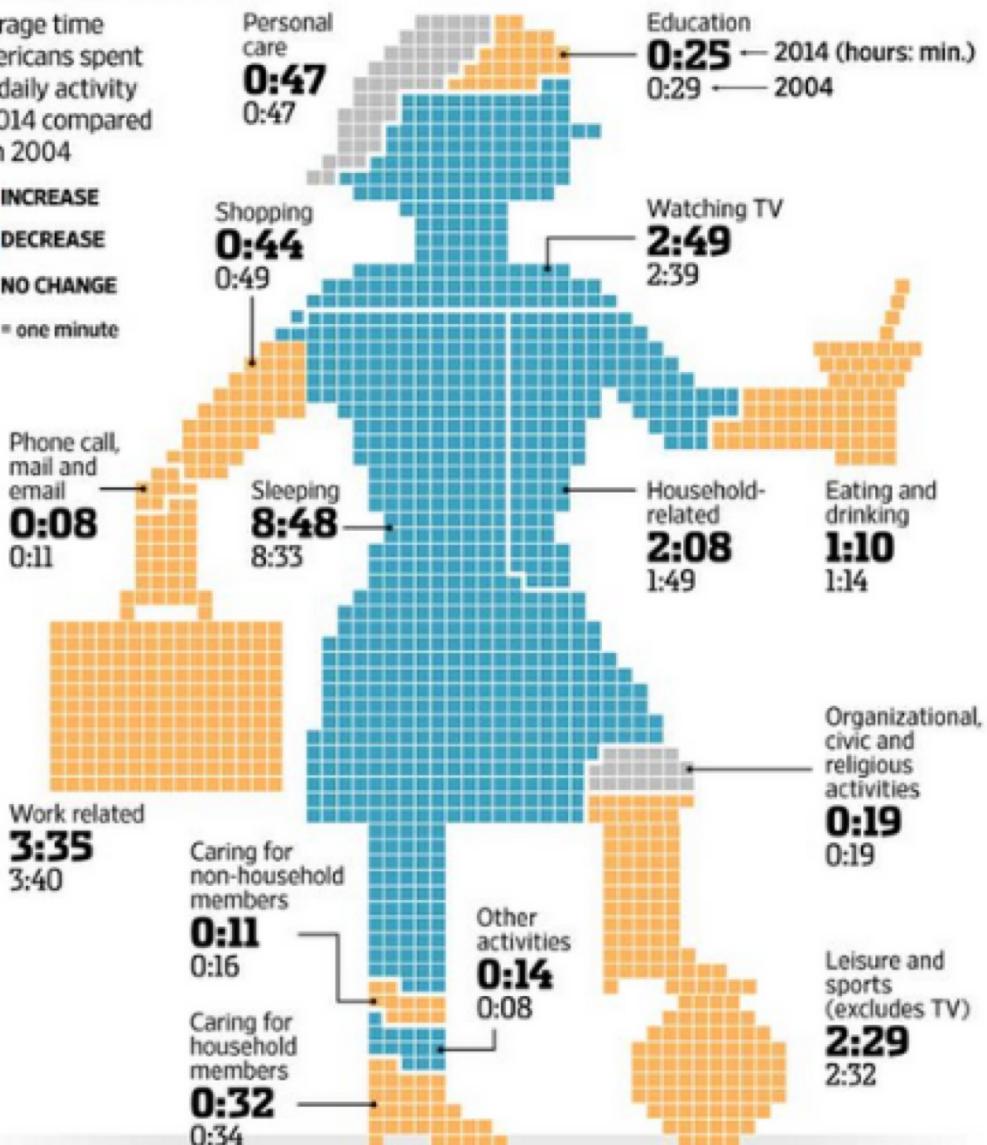
- Break into groups of 2-3 people, and go to:
<http://www.science.smith.edu/~jcrouser/design>
- Goal: **critique** some professionally-made visualizations
- Try to think about the following questions:
 - What is the **first thing you notice** about this visualization?
 - What **point** is this visualization trying to make?
 - Who is the **intended audience**?
 - What is the visualization **doing well**?
 - What **problems** do you see with the visualization design?
 - **Why** do you think the designer made those choices?

Study shows Americans are working more, sleeping more and watching more TV
on.wsj.com/1QRHRBF

A Day in the Life

Average time
Americans spent
per daily activity
in 2014 compared
with 2004

- INCREASE
- DECREASE
- NO CHANGE
- = one minute



Note: Time may not total 24 hours due to rounding.

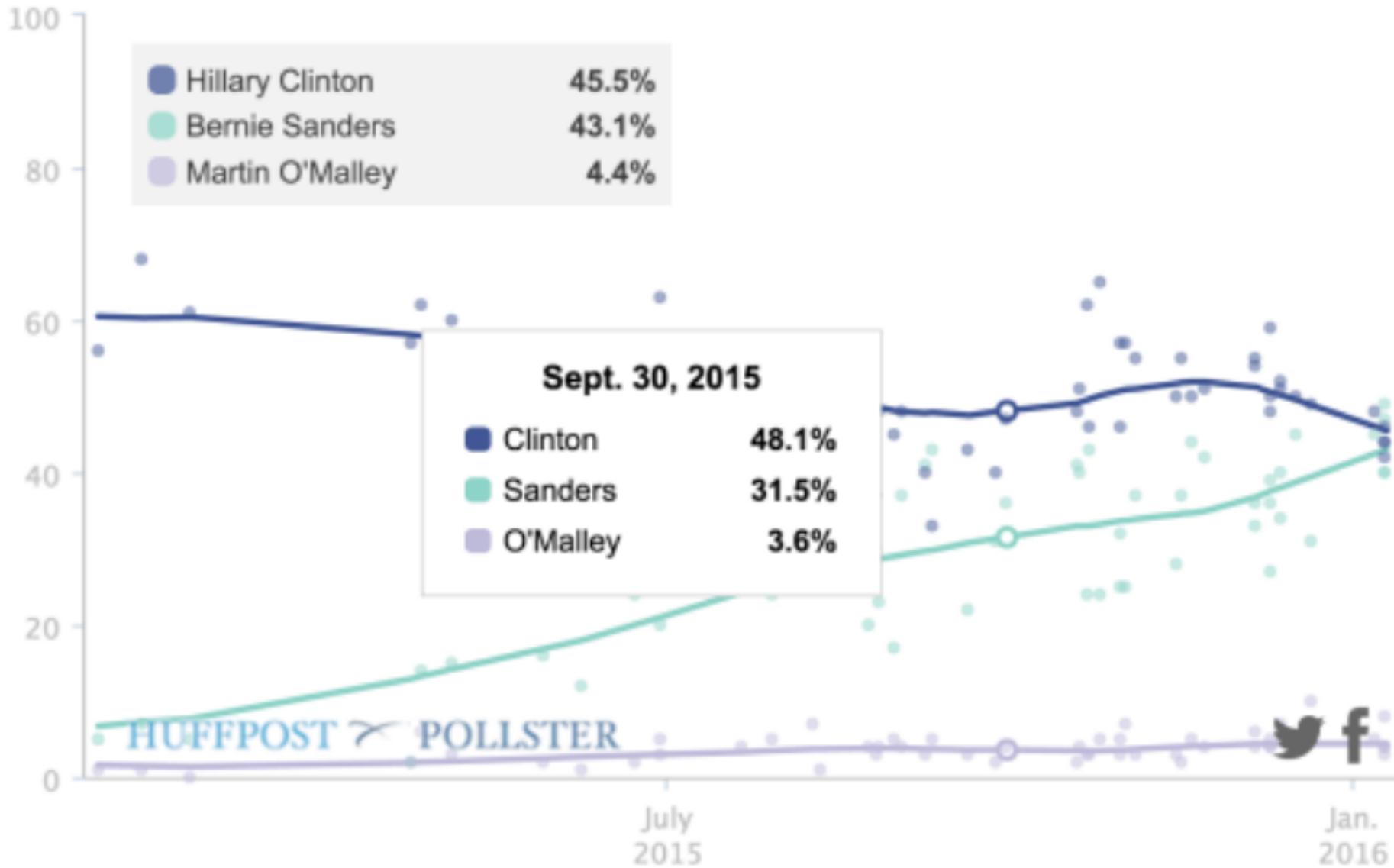
Source: Labor Department

Christopher Kaeser/THE WALL STREET JOURNAL.

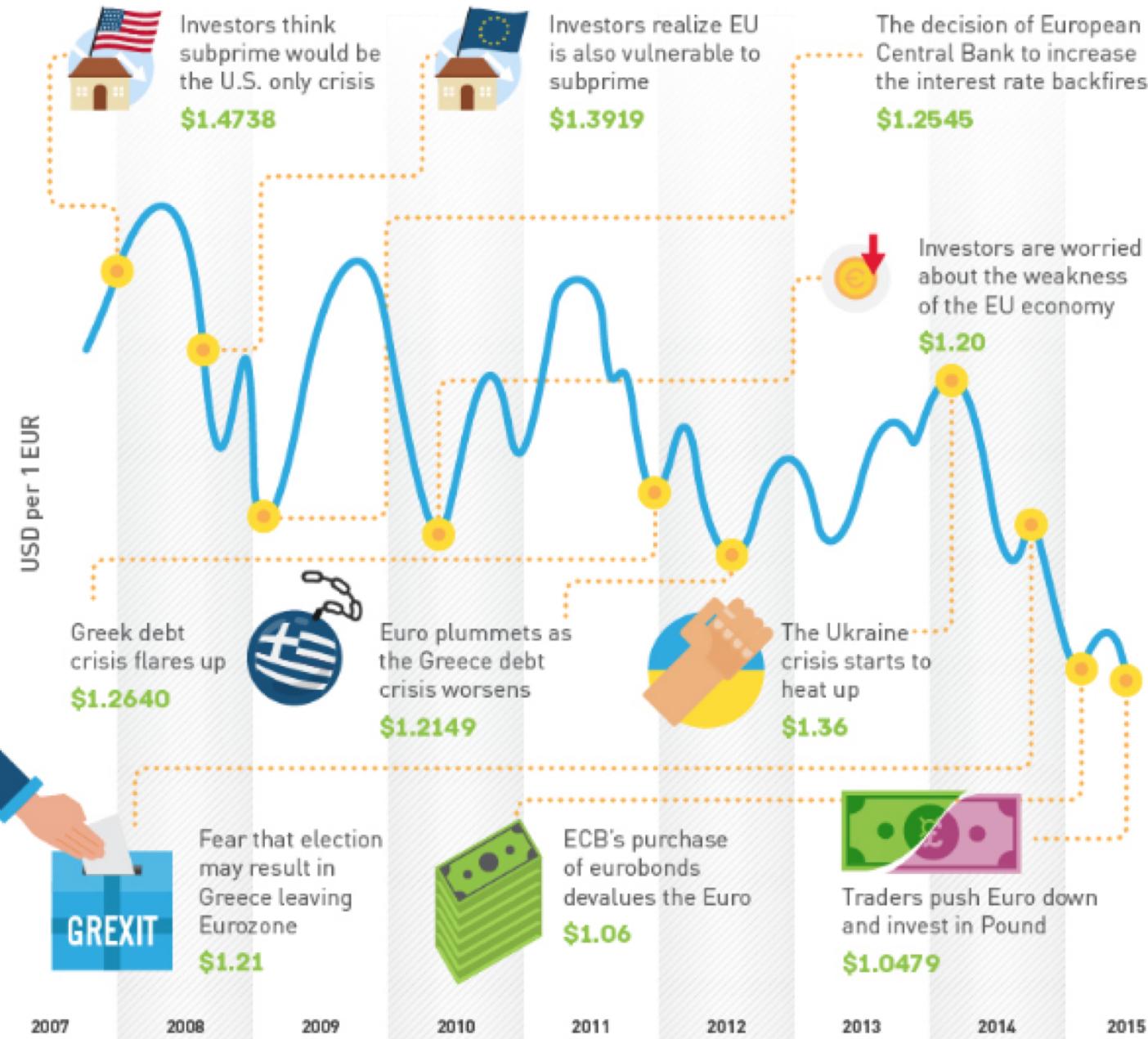
What your
BRAND COLOR
SAYS ABOUT YOUR BUSINESS



2016 Iowa Democratic Presidential Caucus



EVENTS CONTRIBUTING TO DROP OF EURO



UNDER PRESIDENT OBAMA,
**MORE STUDENTS ARE EARNING THEIR HIGH
SCHOOL DIPLOMAS THAN EVER BEFORE**



#LeadOnEducation

SOURCE: U.S. DEPARTMENT OF EDUCATION,
NATIONAL CENTER FOR EDUCATION STATISTICS



57%

of Europeans are
worried their
**personal
information**
is not safe.



Symantec.

Illinois

Variable: Net Job Creation (Per 100)
Employees, Same Sex and Age Group
Year: 2000 Quarter:1
Sex: All and Age Group: Ages 19–21

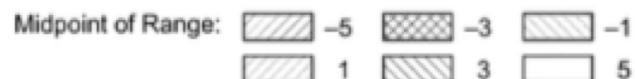


Fig. 5.7 Job creation for young workers, by county, Illinois

Who do Nike sponsor?

International sports and events sponsor

The American based company is the largest sports supplier in the world, supplying equipment, shoes and apparel.

1,016
athletes sponsored worldwide

\$230m

spent on athlete sponsors

1 Figure Skater



1 Martial Artist



2 Cyclists



2 Skiers



2 Squash Players



29 American Footballers

25 Male Golfers

21 Cricketers

10 Swimmers

4 Female Golfers

15 Female Footballers

8 Rugby Players

8 Boxers

64 Ice Hockey Players

93 Basketball Players

485 Male Footballers

120 Male Track and Field Athletes

72 Female Track and Field Athletes

34 Baseball Players

31 Male Tennis Players

21 Female Tennis Players

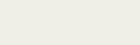
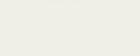
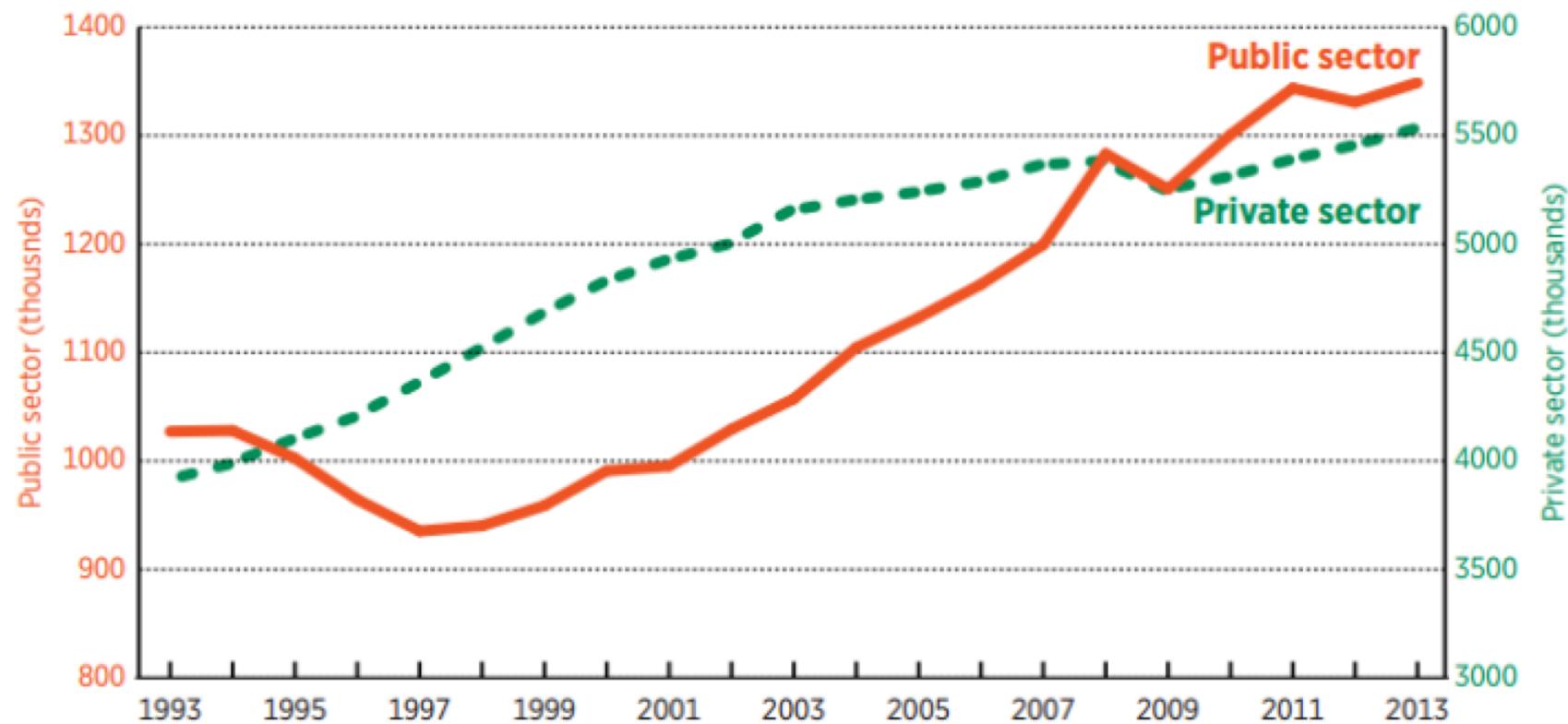


Figure 10: Public- and private-sector jobs (000s) in Ontario, 1993–2013



Source: Statistics Canada, CANSIM Table 282-0089: *employment by class of worker and sex, seasonally adjusted and unadjusted; Ontario; Public sector and private sector employees; Both sexes; Seasonally adjusted (x 1,000)*.



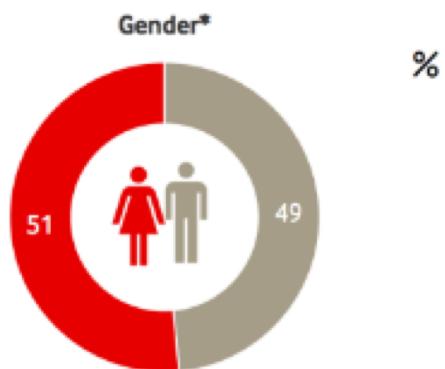
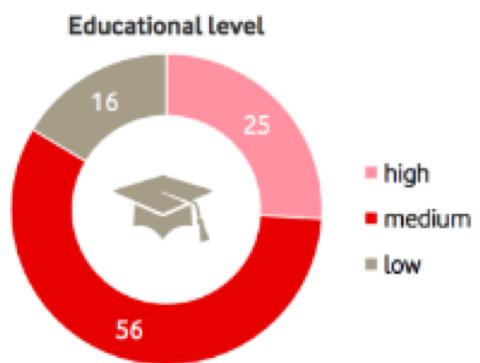
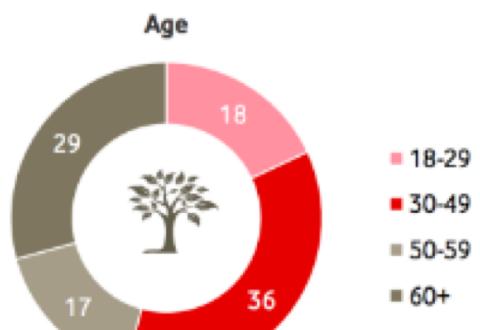
2011
193,600

2015
117,161

**Despite the hysteria, the
number of refugees in the
UK has actually fallen by
76,439 since 2011.**



Sample Description



* Due to rounding, numbers presented in this document may not add up precisely to 100%



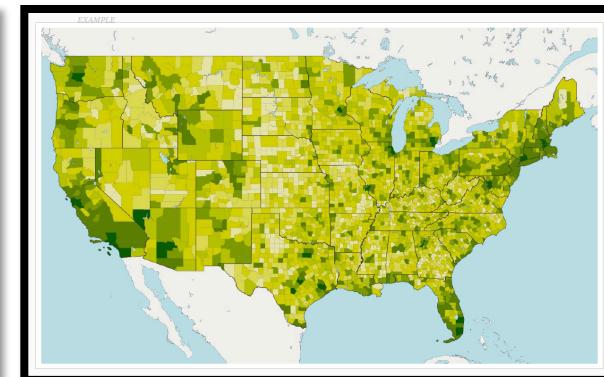
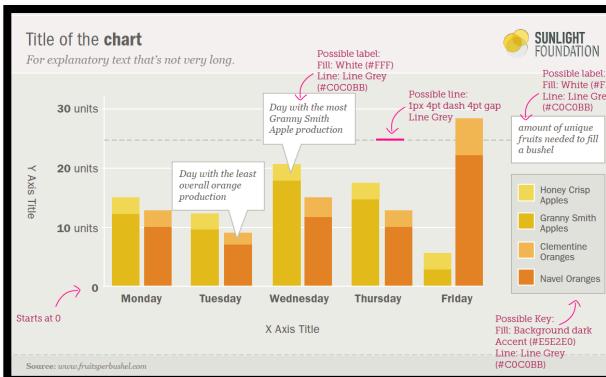
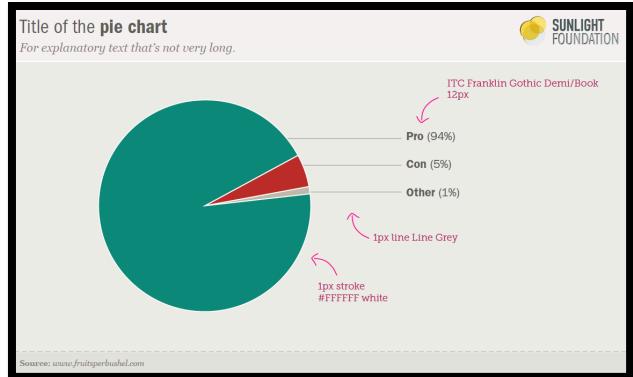
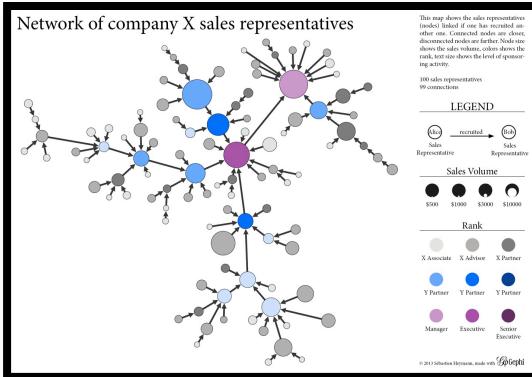
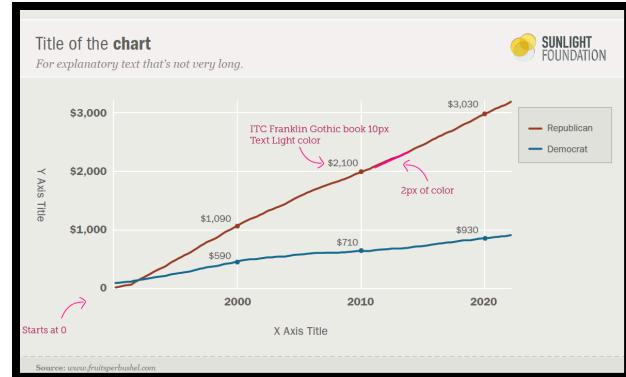
LIKES PER TEAM (AVG)



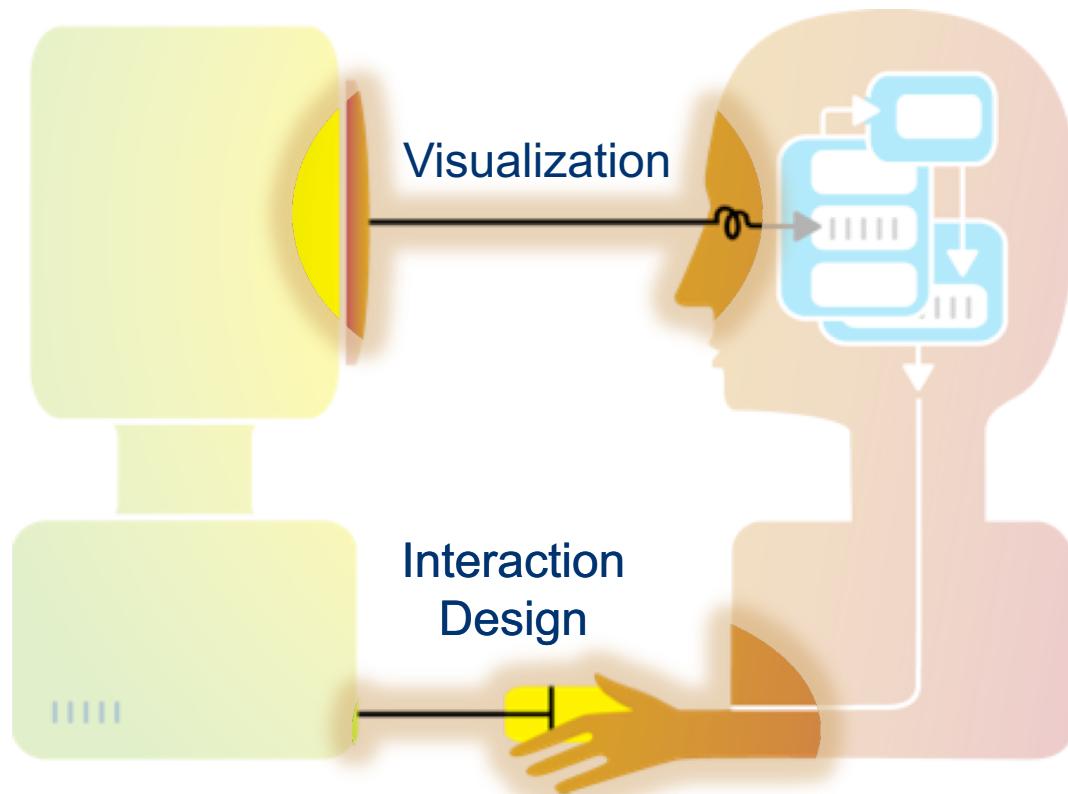
FOLLOWERS PER TEAM (AVG)



So far: Information Visualization



Flashback: about this workshop



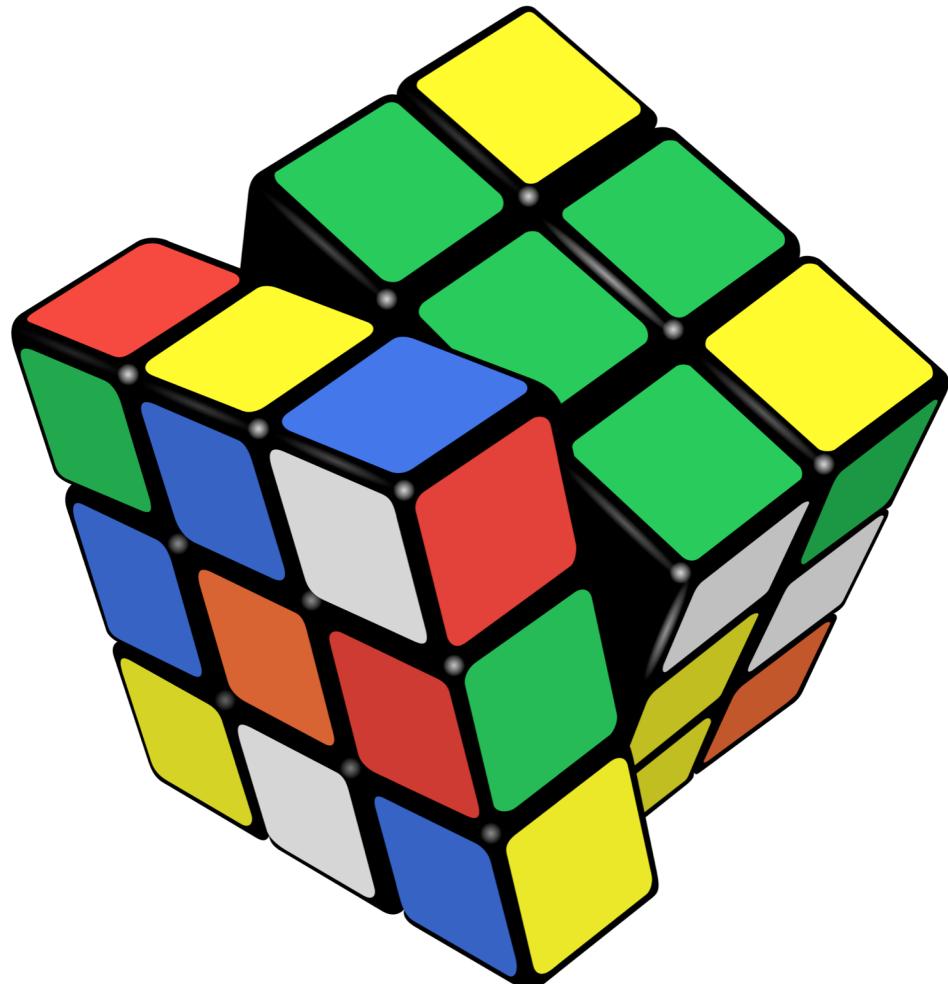
Interaction (def.)

- Methods by which humans **create knowledge** through the manipulation of an interface
- Low level: **between human and interface**
 - the set of operations available
 - the relationship between the human and the visualization
- High level: **between human and problem space**
 - a cognitive act *enabled* by the tool
 - does not need to take place exclusively within them
 - might be distributed across multiple tools

Example: Rubik's Cube

What **low-level**
interactions can you
have?

What **high-level**
interactions can you
have?



Interaction and Analysis

- Interaction is the observable result of a cognitive process: “**an externalization of thought**”
- In visual analytics, there is a growing belief that interaction and analysis are one in the same
- **Analytic discourse:** the idea that knowledge is constructed, tested, refined, and shared through the interactive manipulation of a visual interface¹

¹Pike, W. A., Stasko, J., Chang, R., & O'Connell, T. A. (2009). The science of interaction. *Information Visualization*, 8(4), 263-274.

Part I: High Level

Interaction with
Data / Problem Space

Interaction as a Reasoning Aid

- Interaction is situated in the **context** of a problem or goal-directed activity
- This context helps the human identify relevant concepts and **link** them into appropriate structures
- Interaction brings together background contexts and current observations
- This is known as “**situated cognition**”

Reality Check

Hypothesis: the more ways a user can ‘**touch**’ their data (by changing their form or exploring them from different perspectives), the more **insight** will accumulate.

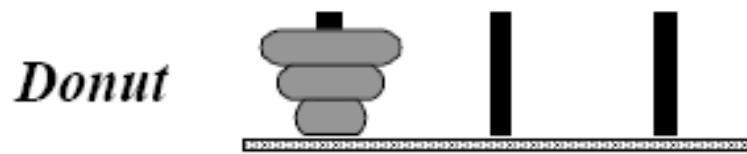
Interaction as Distributed Cognition

- Visualization is often used as a tool to “**offload**” storage or computation from the human’s brain
- In order for this to be useful, we have to be able to “**reload**” parts of the data and operate on it
- Internal (in your head) vs. external (on the screen) representations
- Consider the impact of the **affordances** of the interface



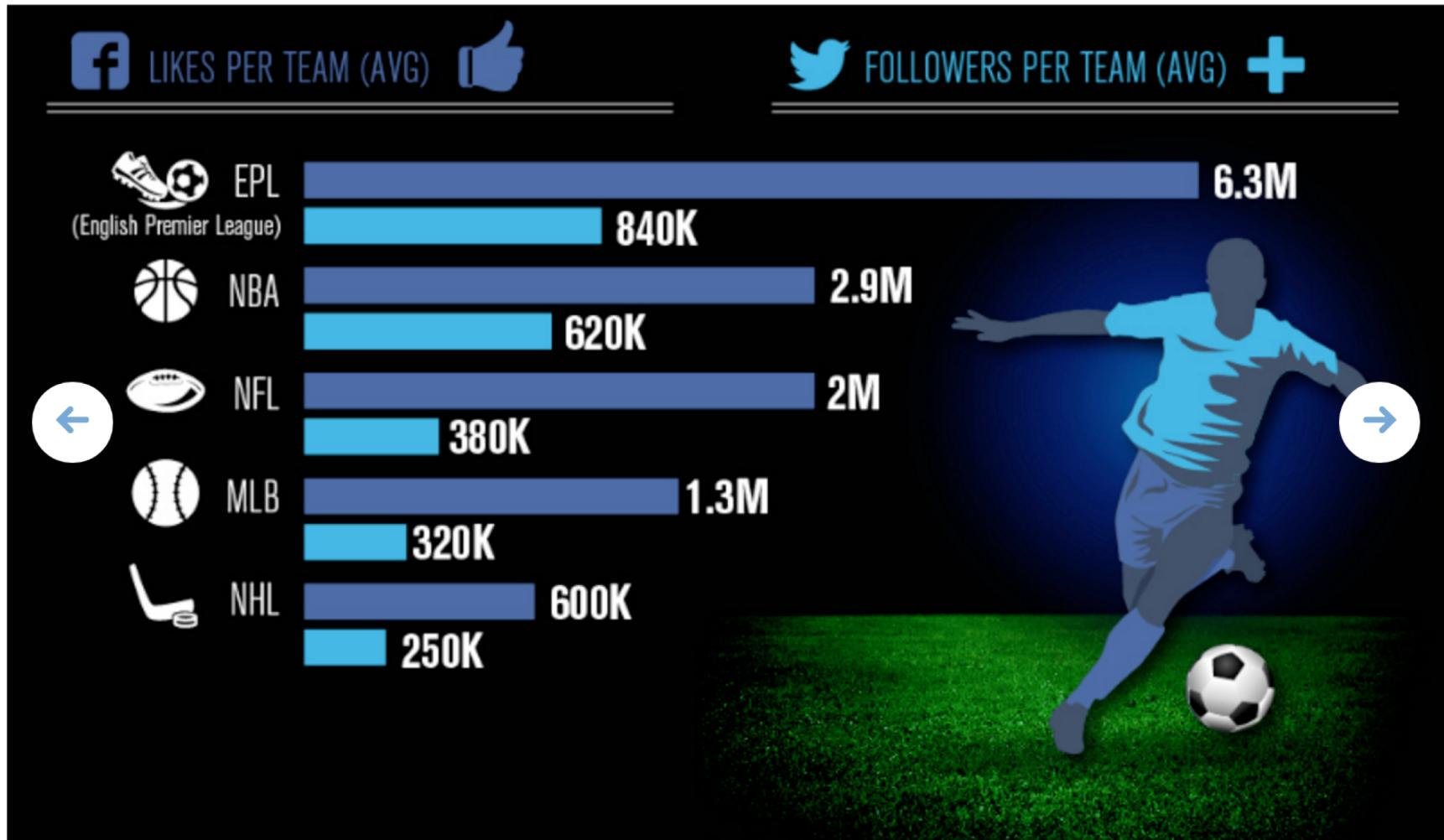
Example: Tower of Hanoi

Rule1 Rule2 Rule3



1. Only one disk can be transferred at a time.
2. A disk can only be transferred to a pole on which it will be the largest.
3. Only the largest disk on a pole can be transferred to another pole.

Flashback: #badvis

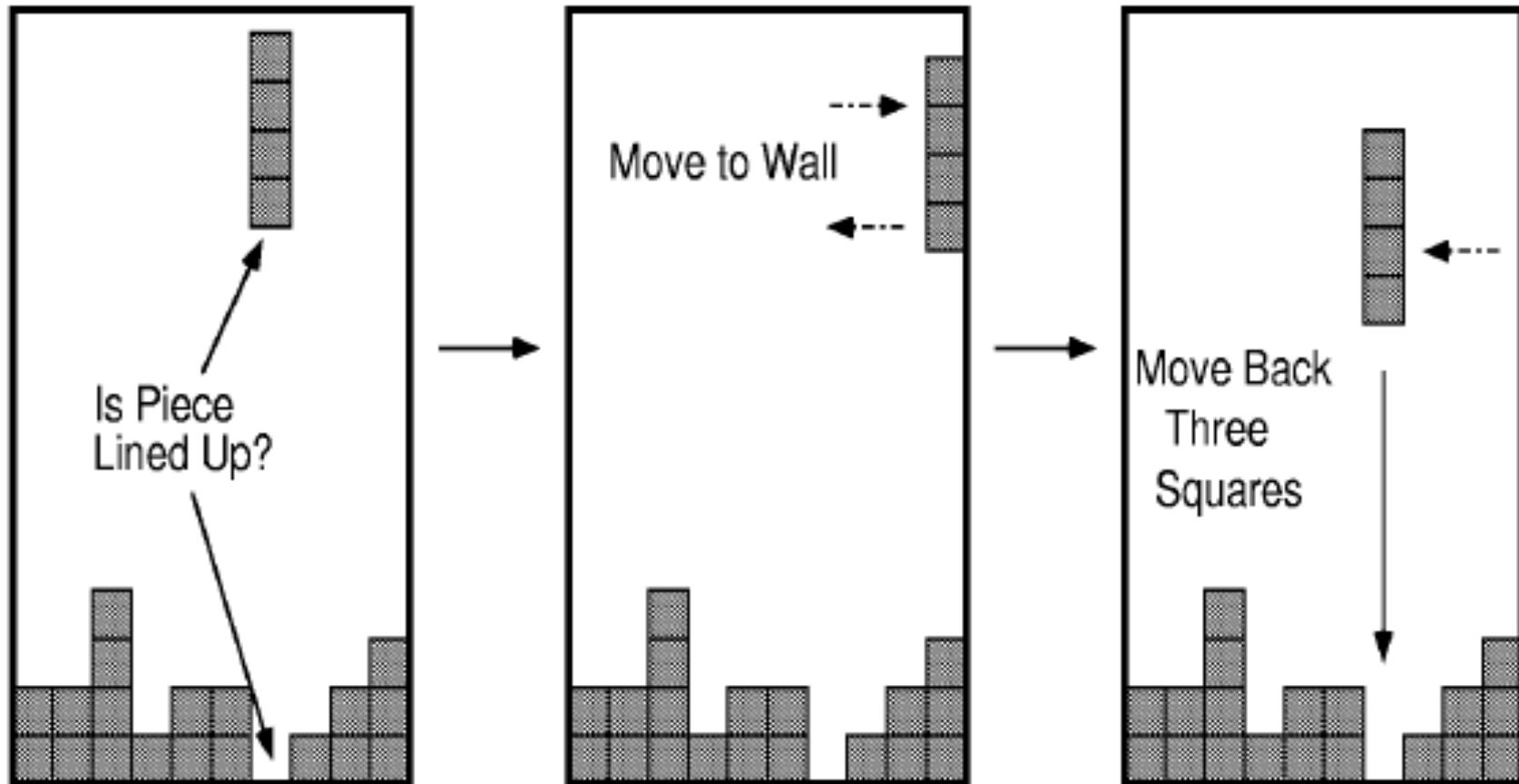


What representations are internal vs. external?

Pragmatic vs. Epistemic Action

- **Pragmatic actions** move a person and their analysis closer to the desired destination.
- **Epistemic actions** enable humans to leverage environmental structures to **link internal structures**.
- The purpose of some actions is not for the effect they have on the environment but **for the effect they have on the humans**.

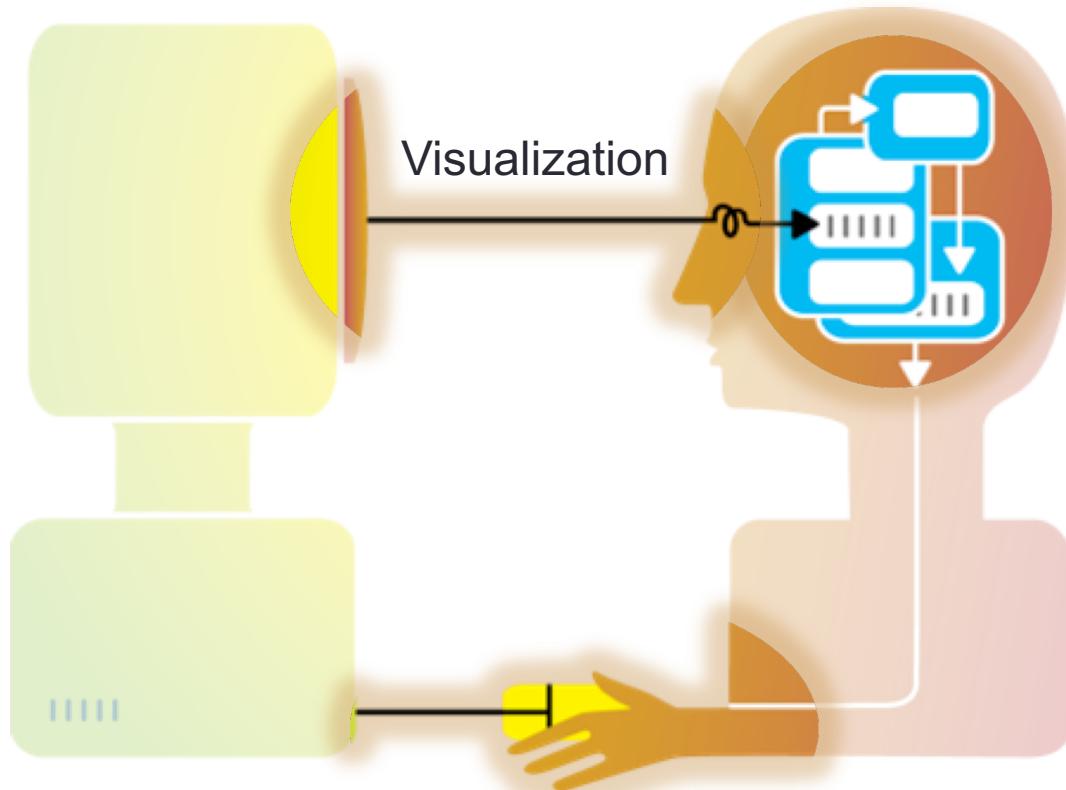
Example: Tetris



Part II

Interaction with a Visual Interface

So far...



Flashback: interaction as a reasoning aid

- Interaction is situated in the **context** of a problem or goal-directed activity
- **Question:** what kinds of things might someone want to do using a visualization?



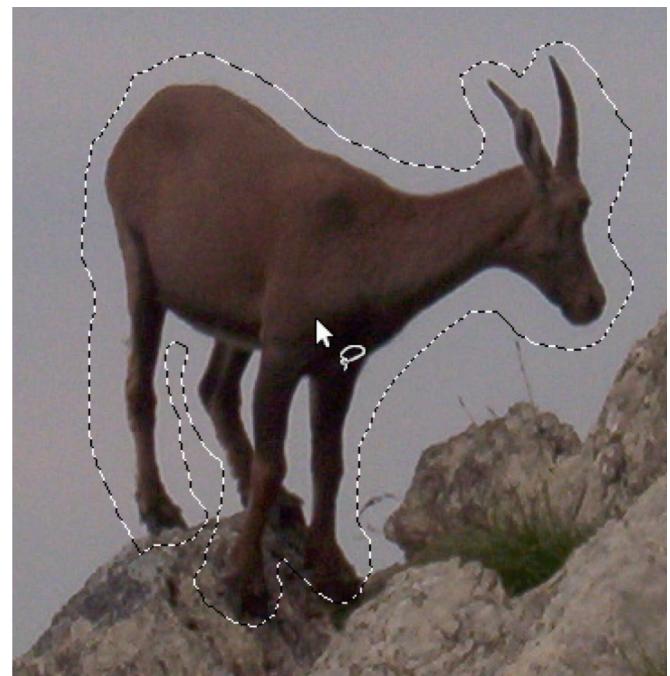
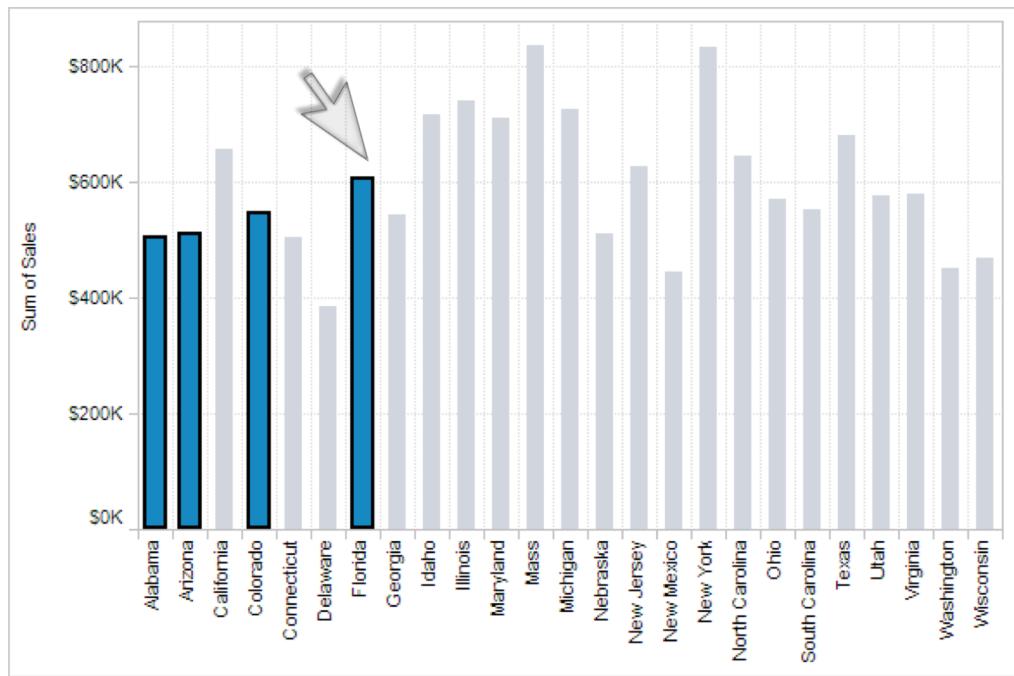
Yi, Kang, Stasko and Jacko (2007)

1. Select: mark something as interesting
2. Explore: show me something else
3. Reconfigure: show me a different arrangement
4. Encode: show me a different representation
5. Abstract/Elaborate: show me more or less detail
6. Filter: show me something conditionally
7. Connect: show me related items

Yi, J. S., ah Kang, Y., Stasko, J. T., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6), 1224-1231.

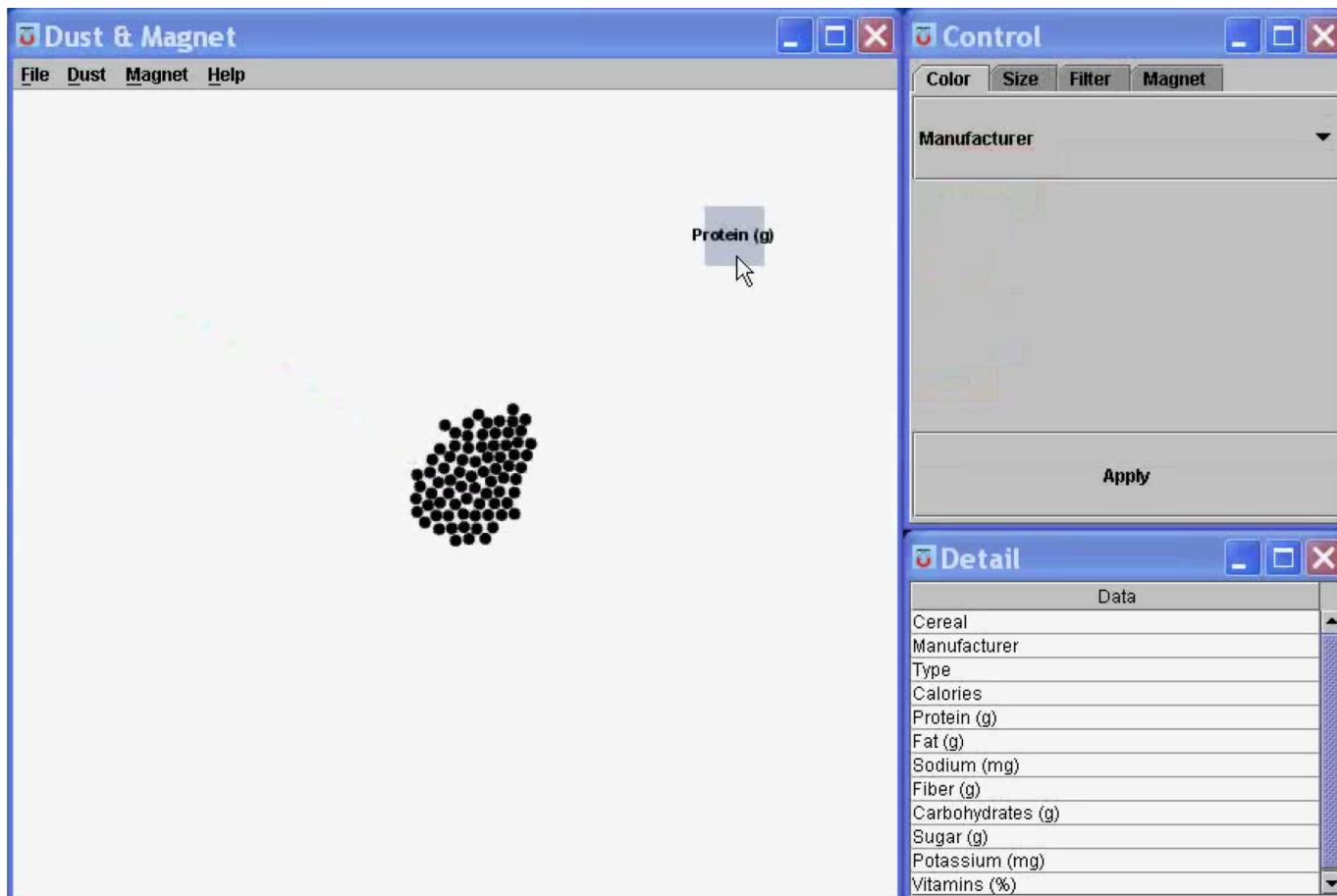
1. Select

Mark something as interesting: direct



1. Select

Mark something as interesting: indirect



2. Explore

Show me something else

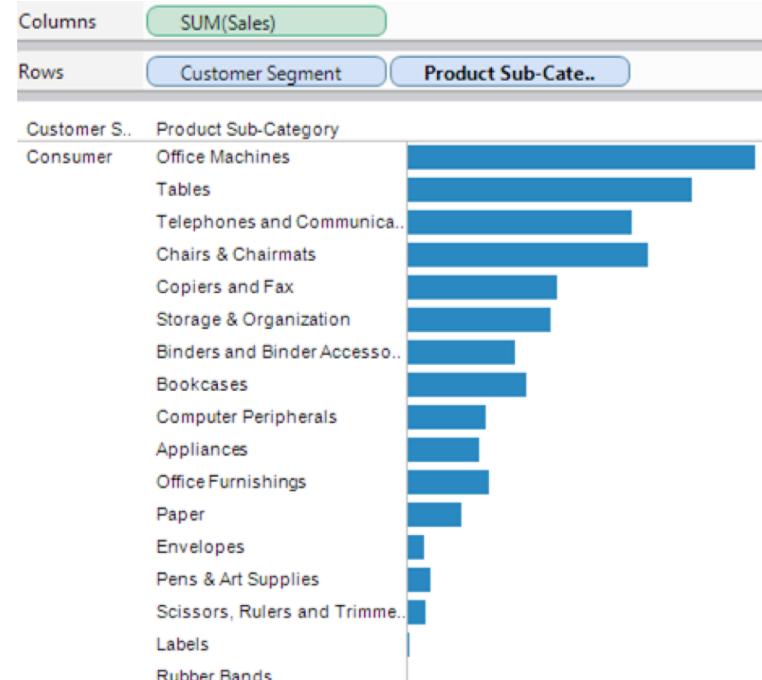
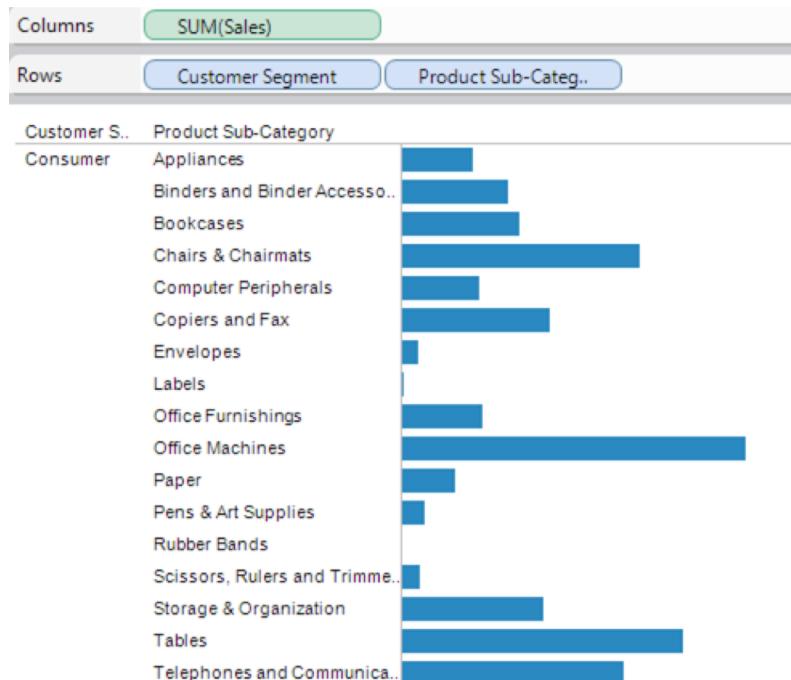
- Scroll bars
- Panning
- Direct-Walk (e.g. hyperlink traversal)

The image displays two side-by-side examples of user interface design for exploration:

- Left Side (Map View):** A map of Pittsburgh, Pennsylvania, showing a dense network of streets and major highways (I-65, I-79, I-376). The map includes labels for various neighborhoods and landmarks such as the Andy Warhol Museum and the Hill District. It features a zoom control (+/-), a compass icon, and a 'Map/Satellite' switch.
- Right Side (Wikipedia Homepage):** The homepage of Wikipedia, featuring a large globe graphic and the text "WIKIPEDIA The Free Encyclopedia". Below the title is a "navigation" sidebar with links to "Main page", "Contents", "Featured content", "Current events", and "Random article".

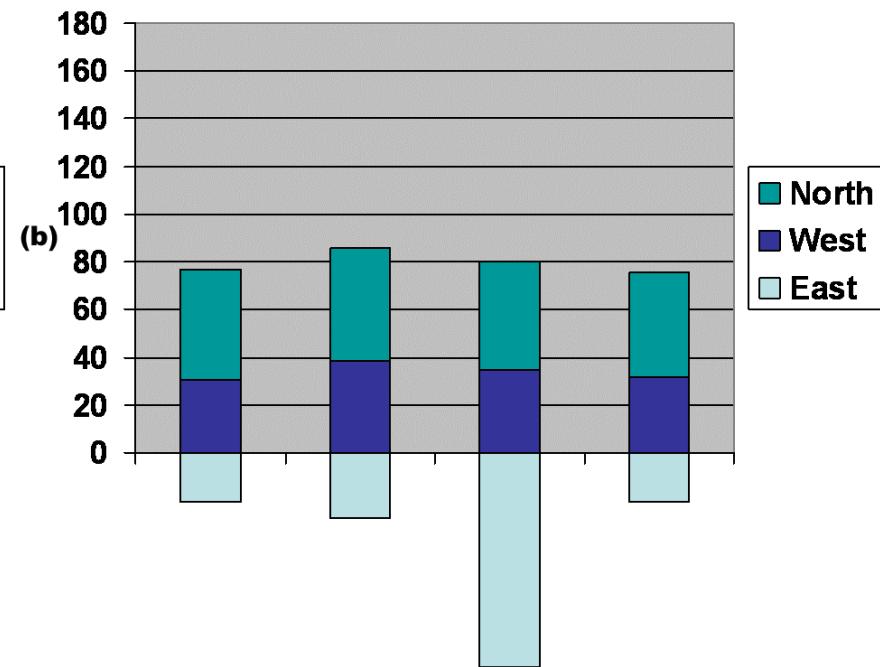
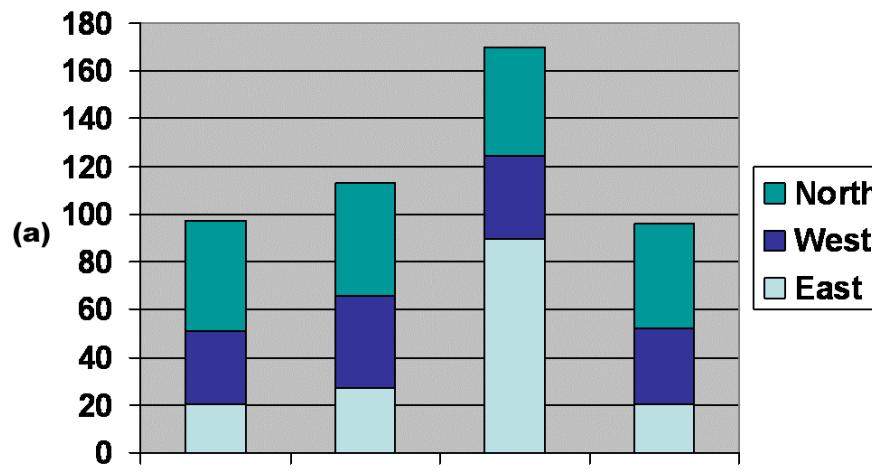
3. Reconfigure

Show me a different arrangement: sorting



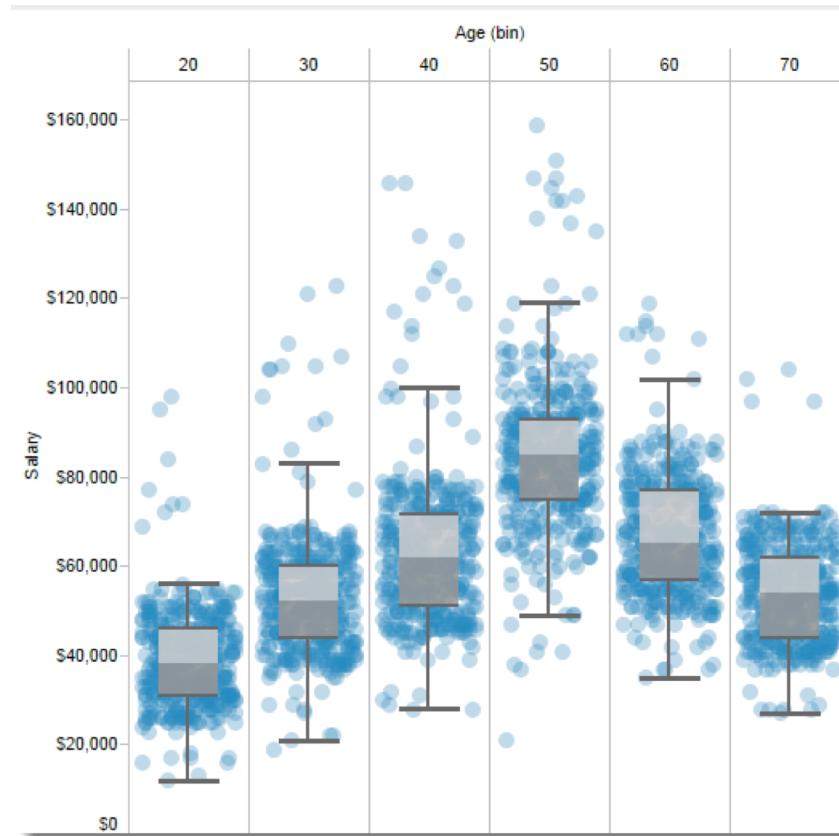
3. Reconfigure

Show me a different arrangement: baseline adjustment



3. Reconfigure

Show me a different arrangement: reduce occlusion (jitter)



4. Encode

Show me a different representation: visualization type, color, size, orientation, etc.



5. Abstract / Elaborate

Show me more or less detail: drill up/down



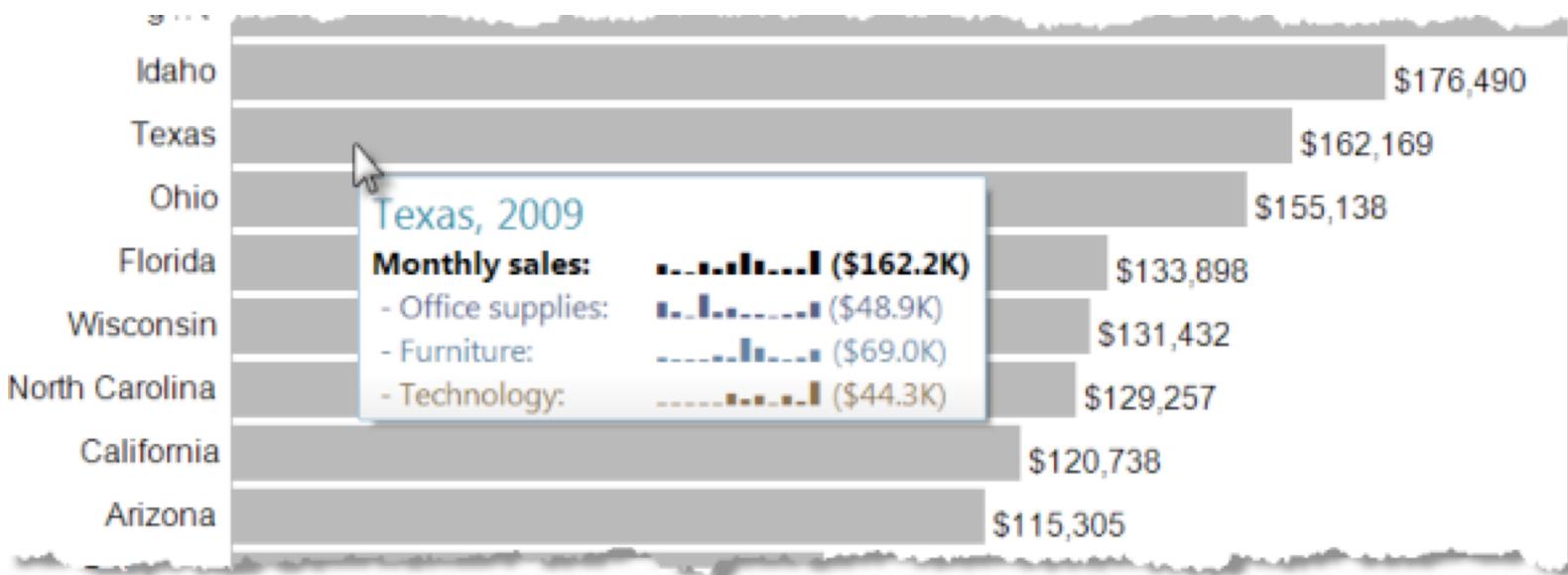
5. Abstract / Elaborate

Show me more or less detail: zooming



5. Abstract / Elaborate

Show me more or less detail: tooltips



6. Filter

Show me something conditionally

NameVoyager: Explore baby names and name trends letter by letter

Looking for the perfect baby name? [Sign up for free](#) to receive access to our expert tools!

Baby Name >

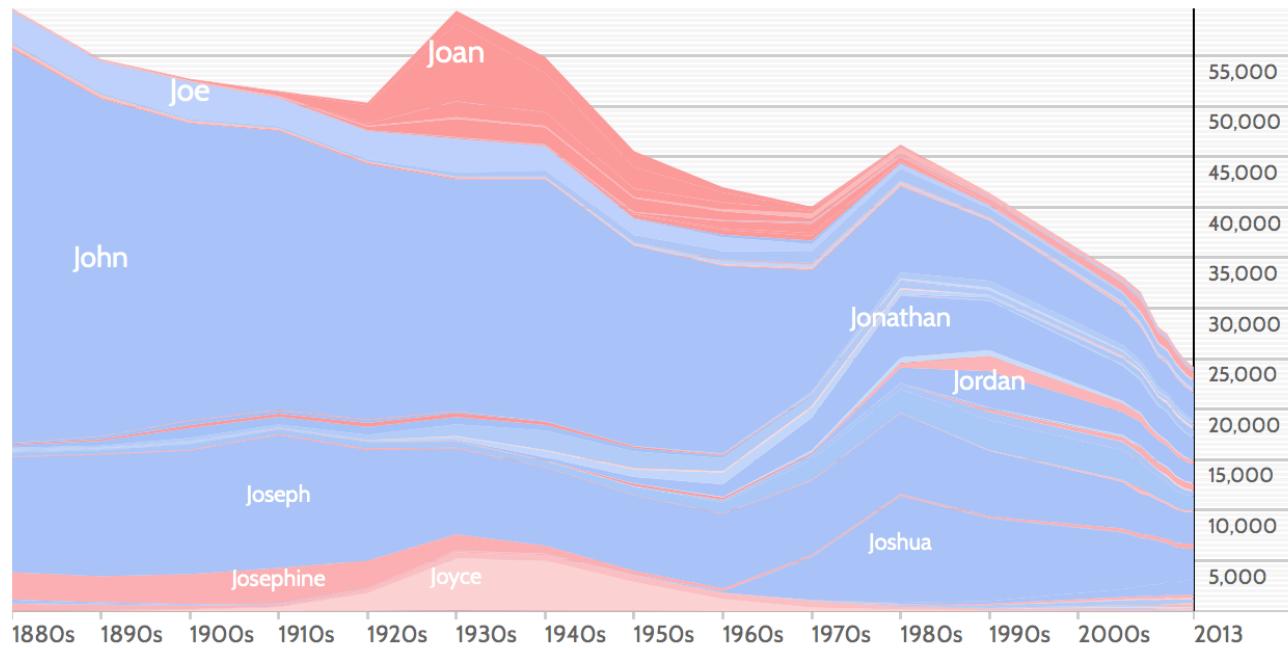
Both Boys Girls

boys	1000	500	100	25	1
girls	1000	500	100	25	1

Current rank:

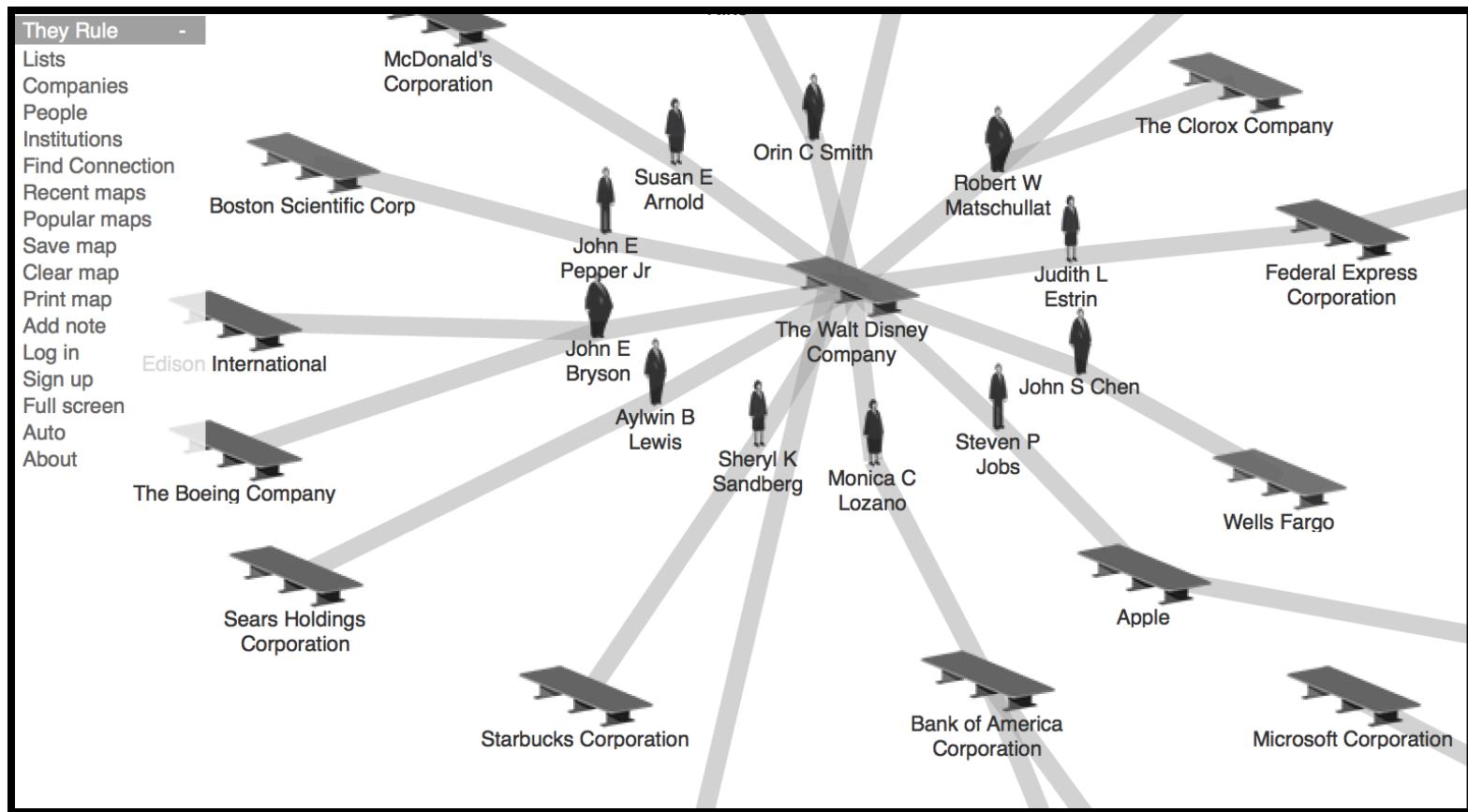
per million births

Names starting with 'JO' per million babies



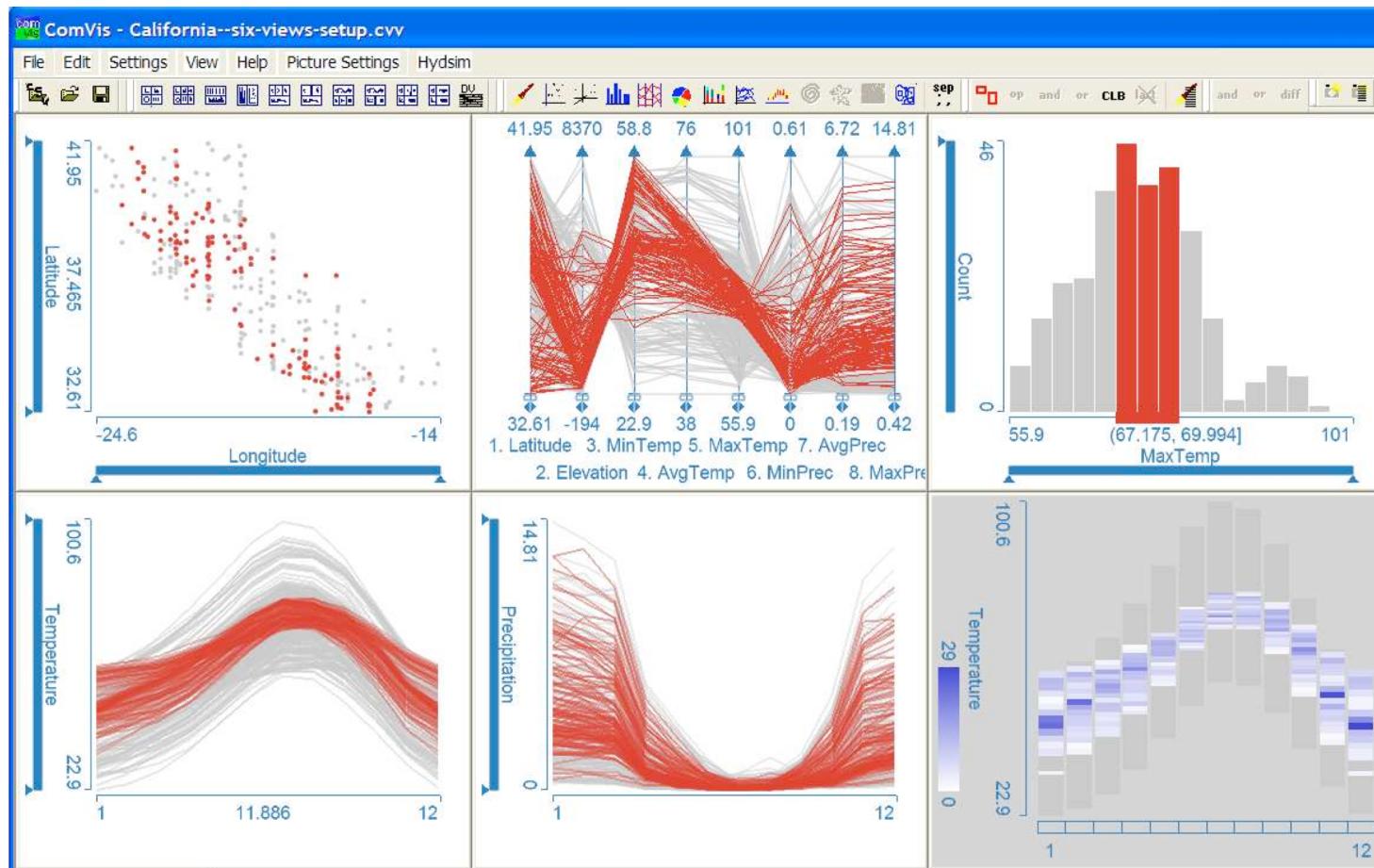
7. Connect

Show me related items: build-out

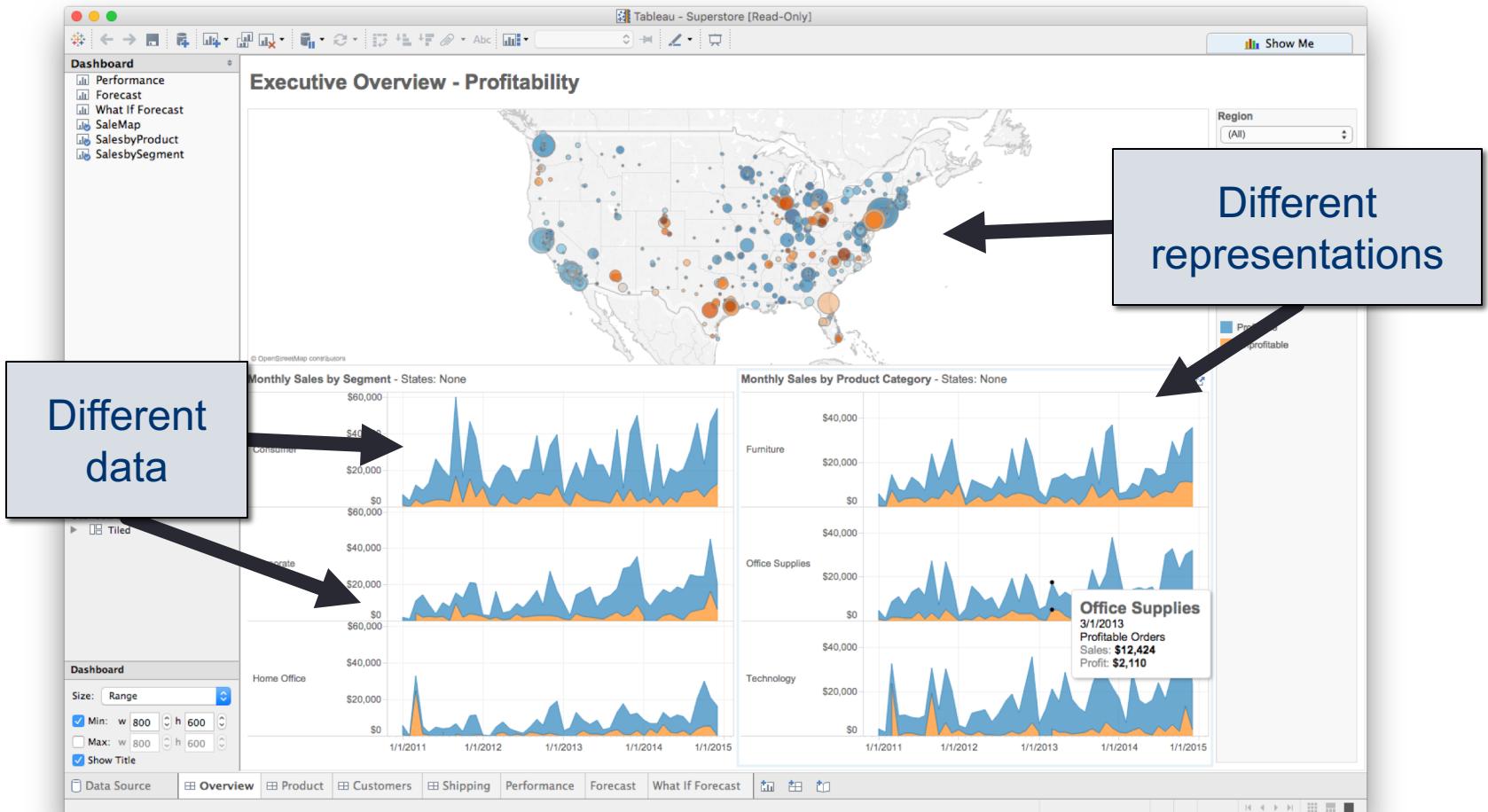


7. Connect

Show me related items: coordinated multiple views (CMV)



Multiple views



Systems that use **two or more distinct views** to support the exploration of a single concept or domain

When to use multiple views?

- The data is too big
 - Lots of attributes
 - Lots of observations
- The data is too complicated
 - Lots of data sources
 - Lots of data types
- The data has several interesting parts, but no one visualization highlights them all

Need to think about: resource optimization

SiSense

Sheet Document3

Visitors - Map Distribution

A choropleth map of the United States where states are colored based on visitor distribution. Darker shades of blue represent higher visitor counts, while lighter shades represent lower counts.

U.S Visitors and Income By State (source)

State	Stock Size	Satisfaction Index	Visitors	Income (Sum)	\$ / visitor
New York	High	A+	3,313,508	\$2,071,008.60	\$0.625
California	High	A+	2,008,366	\$1,824,459.42	\$0.908
Pennsylvania	Medium-High	A+	503,898	\$424,459.38	\$0.842
Vermont	Medium	A	318,332	\$324,459.02	\$1.019
West Virginia	Low	A	218,560	\$202,327.00	\$0.925
Massachusetts	Low	B+	202,271	\$58,035.27	\$0.286
Oklahoma	Low	B+	102,503	\$82,539.38	\$0.805
Maine	Low	B	98,689	\$63,935.80	\$0.647
Alabama	Very Low	B-	54,201	\$62,502.02	\$1.153
North Carolina	Very Low	B-	46,185	\$53,297.32	\$0.647

10 ▾

Satisfaction Index - at least A

A stacked bar chart showing the average sales for locations with a satisfaction index of at least A. The y-axis represents Average Sale (\$), ranging from \$49 to \$249. The x-axis lists locations: NY 8th Ave., LA Hollywood West, Queens #1, NY Webster Ave., NC #1 - Asheville, Queens #2, LA East, Brooklyn, LA Downtown #2, and Portland Ctr. Products A, B, C, and D are represented by different colors in the stacked bars.

Satisfaction Index - B+ or lower

A stacked bar chart showing the average sales for locations with a satisfaction index of B+ or lower. The y-axis represents Average Sale (\$), ranging from \$49 to \$249. The x-axis lists locations: Queens #1, Queens #2, LA East, Brooklyn, LA Downtown #2, and Portland Ctr. Products A, B, C, and D are represented by different colors in the stacked bars.

Income Per Visitor

A gauge chart showing the maximum income per visitor. The scale ranges from < \$1 to \$1.9. The current value is \$1.156.

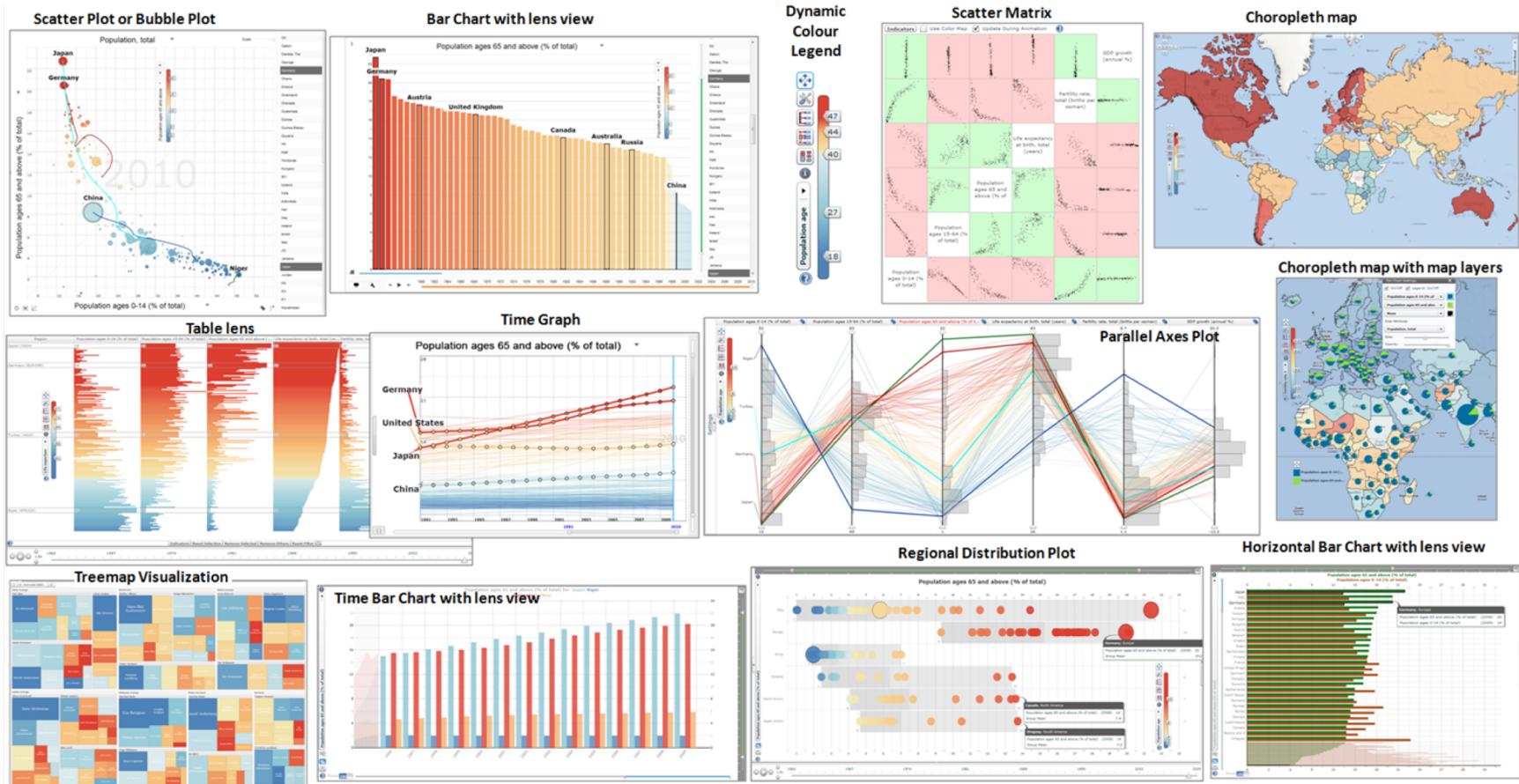
Max income per visitor: \$1.156

Total Income

A gauge chart showing the total income. The scale ranges from 1M to 10M. The current value is \$5,419,143.

Total Income: \$5,419,143

Need to think about: resource optimization



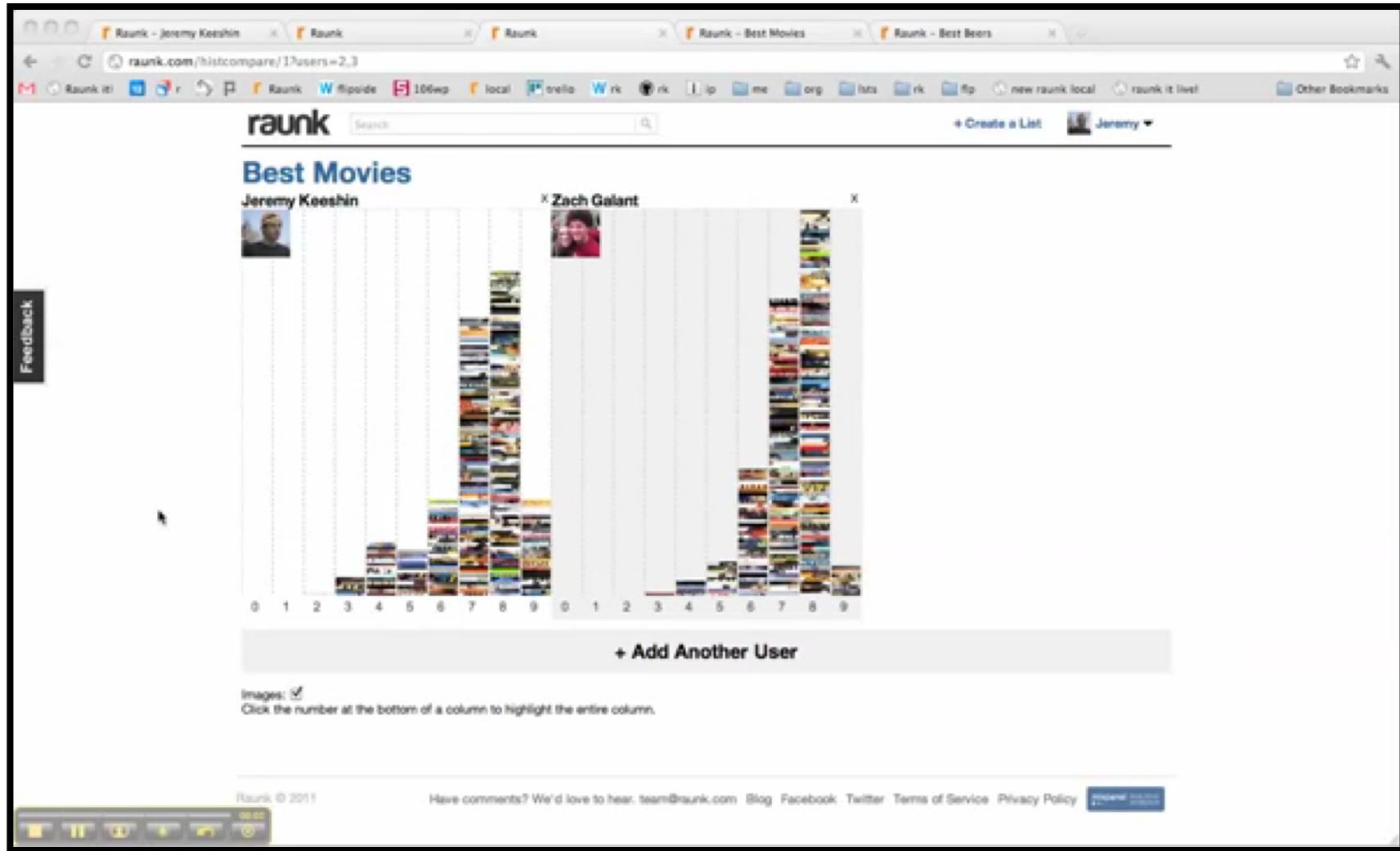
Guidelines for multiple views

- Balance the **costs** of presenting multiple views with the **benefits** of using the views
- Split complex data into multiple views to create **manageable chunks**
- Use views that are complimentary, bringing out **correlations and/or disparities**
- Use **perceptual cues** to:
 - make relationships more apparent to the reader
 - focus the reader's attention on the right view at the right time

Coordination: brushing and linking

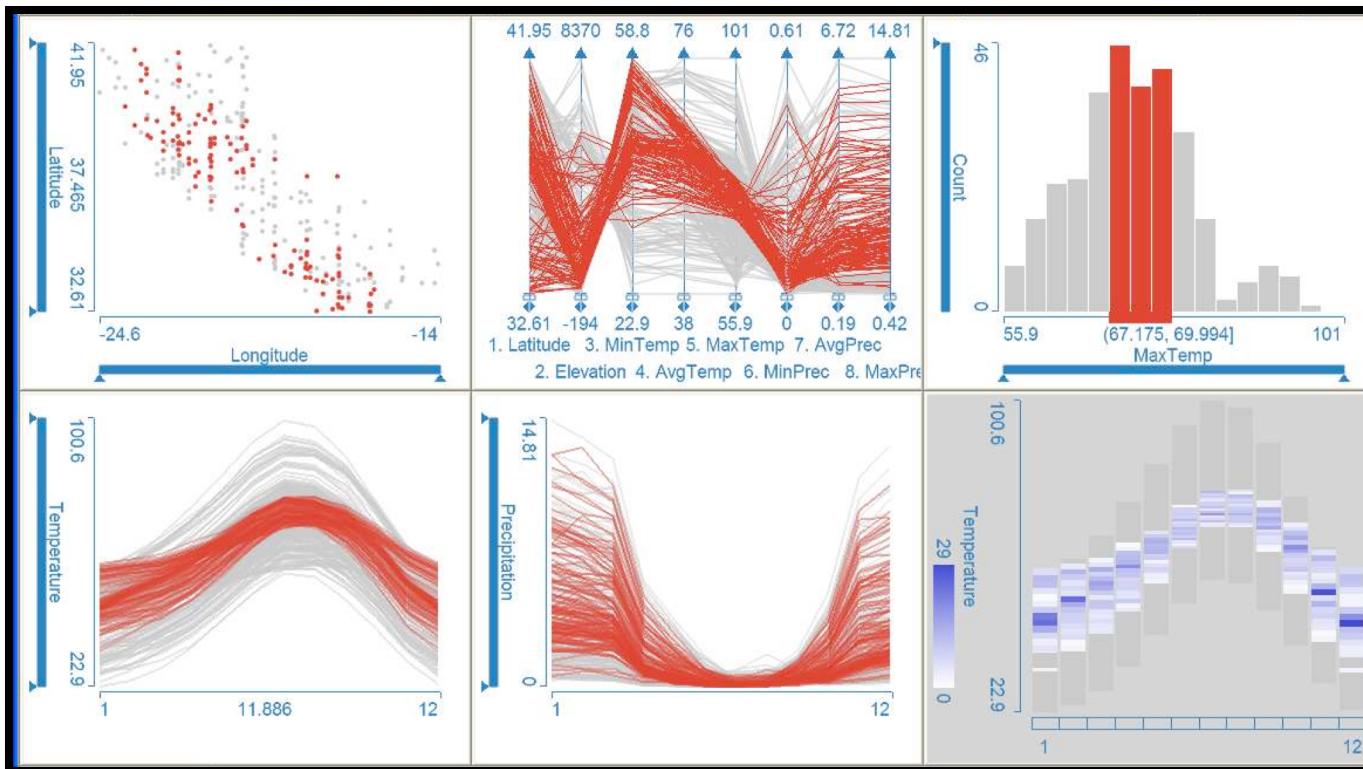
- **Big idea:** actions in the real world have ripple effects; actions on a visualization should too
- **Brushing:** the visualization responds (usually through highlighting) as a person “brushes past” data points
- **Linking:** the visualization connects related data points across multiple views

Coordination: brushing and linking



Coordinated multiple views

- Multiple views + brushing and linking
- **Big idea:** propagate interaction from one view to all others, respond as appropriate



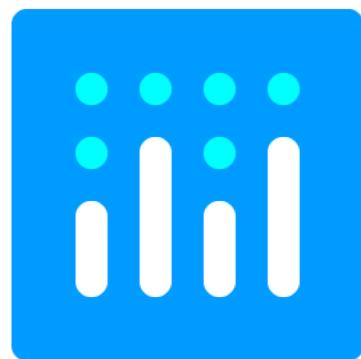
Discussion

What would we need
to make this work?



Suggested questions to ask when designing

1. What is the goal of the analysis?
 - Decision-making
 - Better understand a domain or a problem
 - Identify the trends of a phenomenon
 - Forecast the future
 - ...etc.
2. What kinds of operations do we need to enable?
3. How can the visualization support those operations?



plotly

Outline

- Introduction to `plot_ly()` and `ggplotly()`
- Case Study – Housing Sales in Texas
 - Using `plot_ly()`
 - Using `ggplotly()`
- Combining views
 - with `htmlwidgets`
 - with `shiny`
 - with `subplot(...)`

What you need to know

- 2 main ways to initiate a `plotly` object in R:
 - `plot_ly()` function transforms *data* into a `plotly` object
 - `ggplotly()` function transforms a *ggplot object* into a `plotly` object ([Wickham 2009](#)); ([Sievert et al. 2016](#))
- Both result in an interactive web-based visualization with tooltips, zooming, and panning enabled by default
- Let's play...