

# DeCART Summer School 2018

## Data Visualization

*Day 1 - Afternoon*

*16 July 2018*

Nils Gehlenborg, PhD - Harvard Medical School

# Syllabus

- **Day 1**

- Morning: Introduction to Data Visualization
- Afternoon: Design Process, Evaluation, and Interaction

- **Day 2**

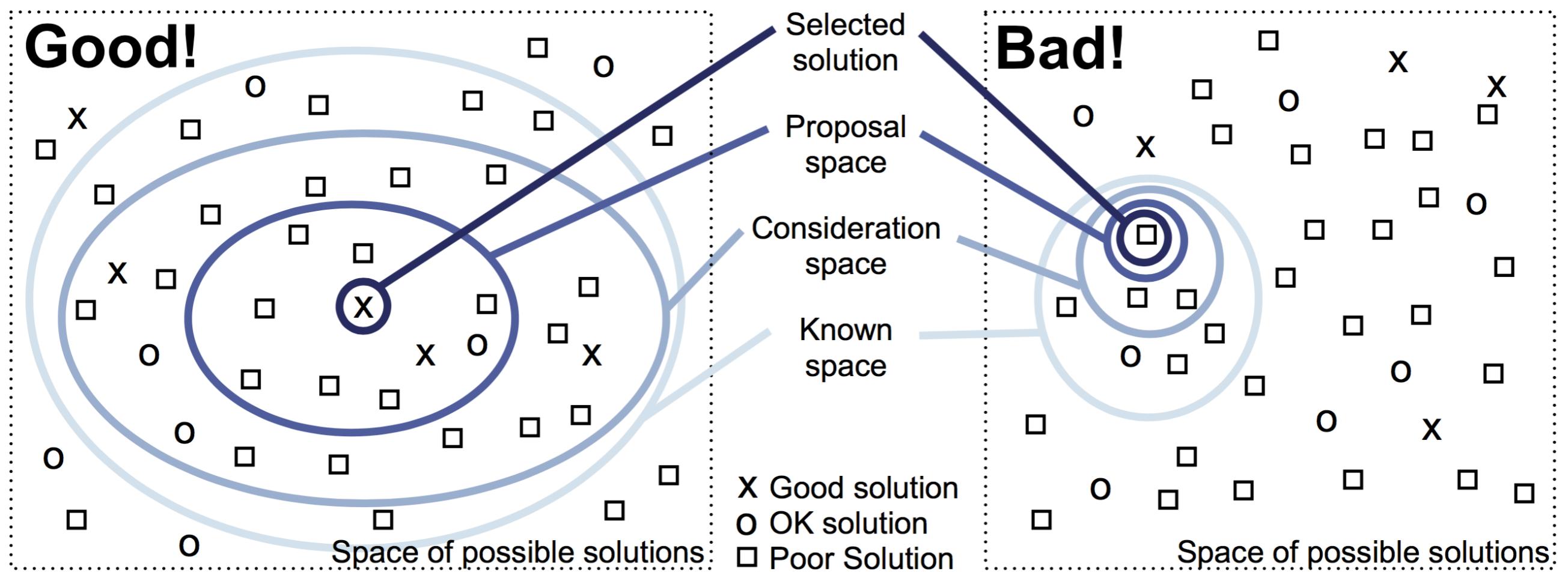
- Morning: Introduction to Altair and visualization of high-dimensional and temporal data
- Afternoon: Advanced Altair and visualization of genomic and network data

# Syllabus - Day 1 (Afternoon)

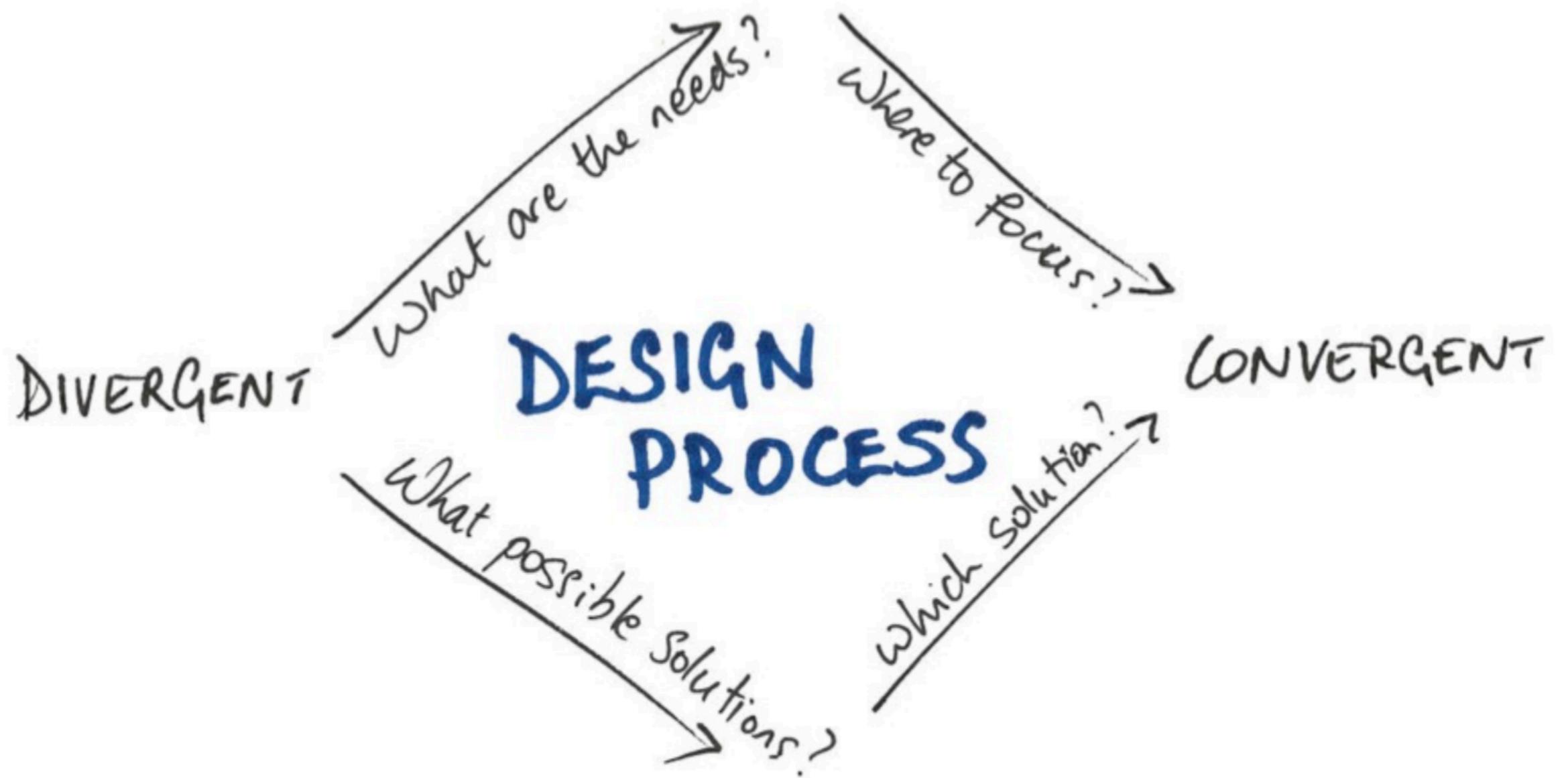
- 1:30 - 2:00 | Design Process & Five Design Sheet Method
- 2:00 - 2:45 | Design Exercise 2: Flu Dashboard
- 2:45 - 3:00 | Break
- 3:00 - 3:30 | Design Exercise 2: Reporting Back
- 3:30 - 4:00 | Evaluation
- 4:00 - 4:30 | Interaction
- 4:30 - 5:00 | Questions and Preview

# Design Process

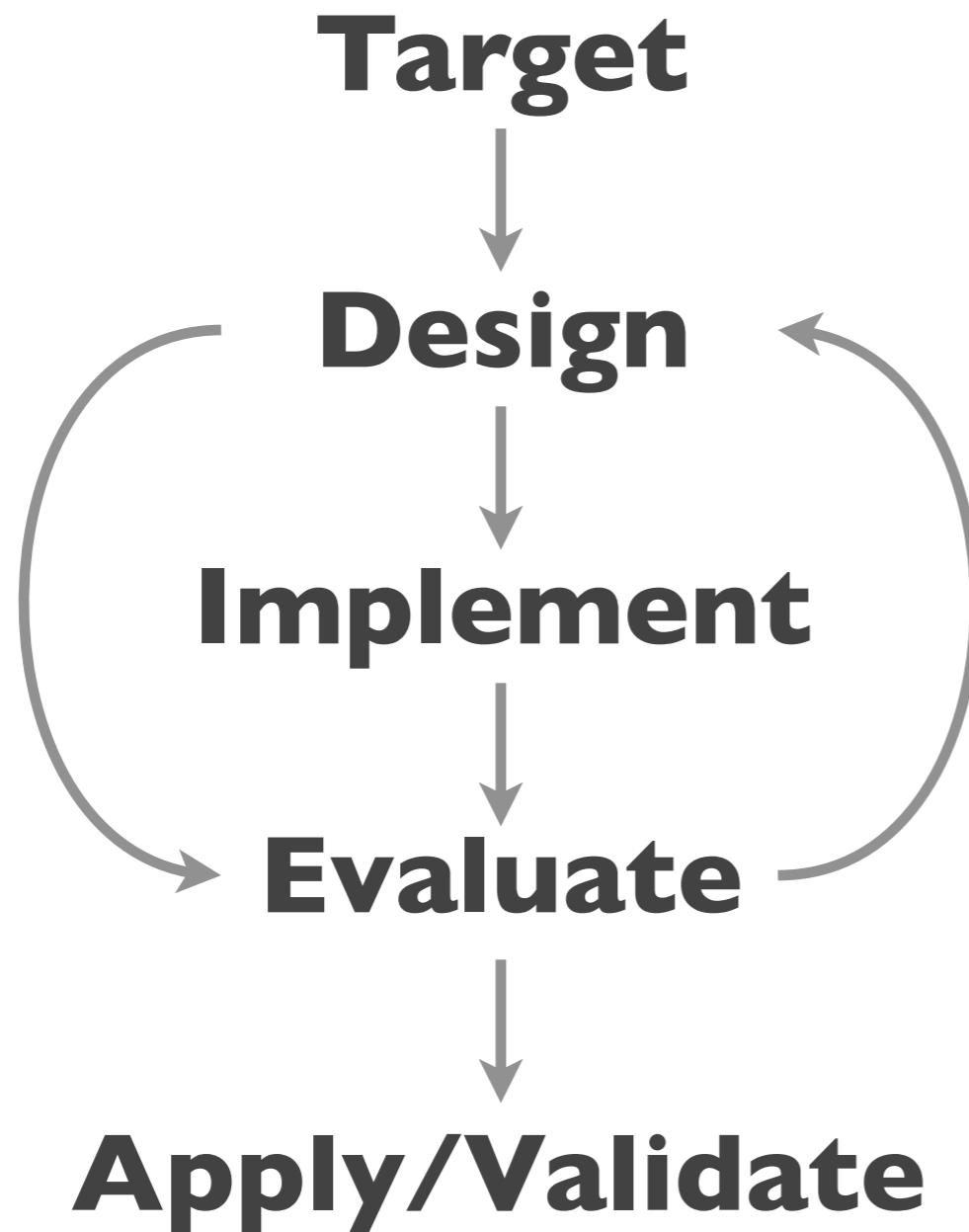
# Goal of the Design Process



# Five Design-Sheet Methodology



# User-Centered Participatory Design



user-centered design  
usability engineering  
participatory design

# Five Design-Sheet Methodology

Five sheets

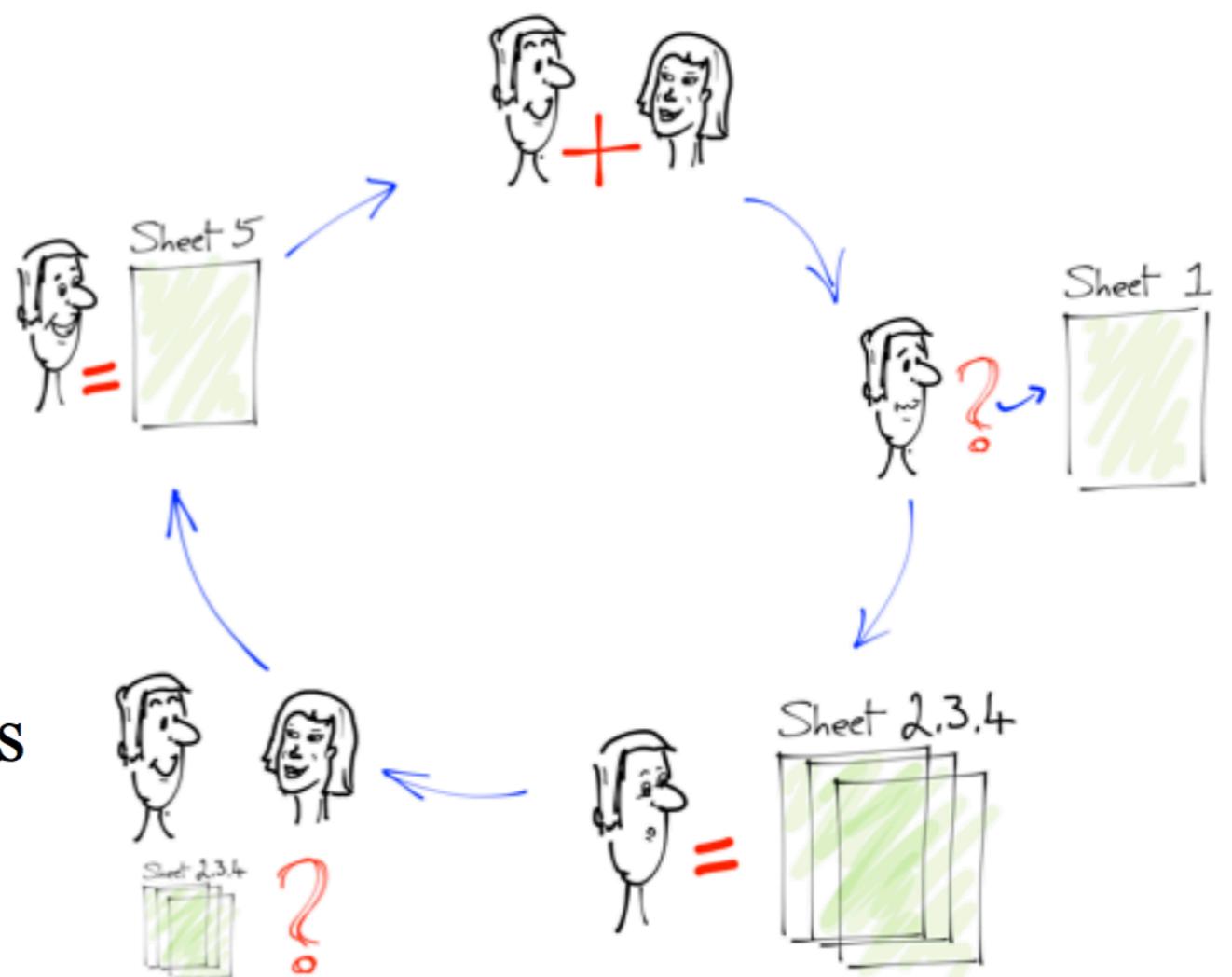
Five parts per sheet

Five parts to the process

Sheet 1 = Ideas

Sheets 2,3,4 = Main designs

Sheet 5 = Realization



# Five Design Sheet Exercise

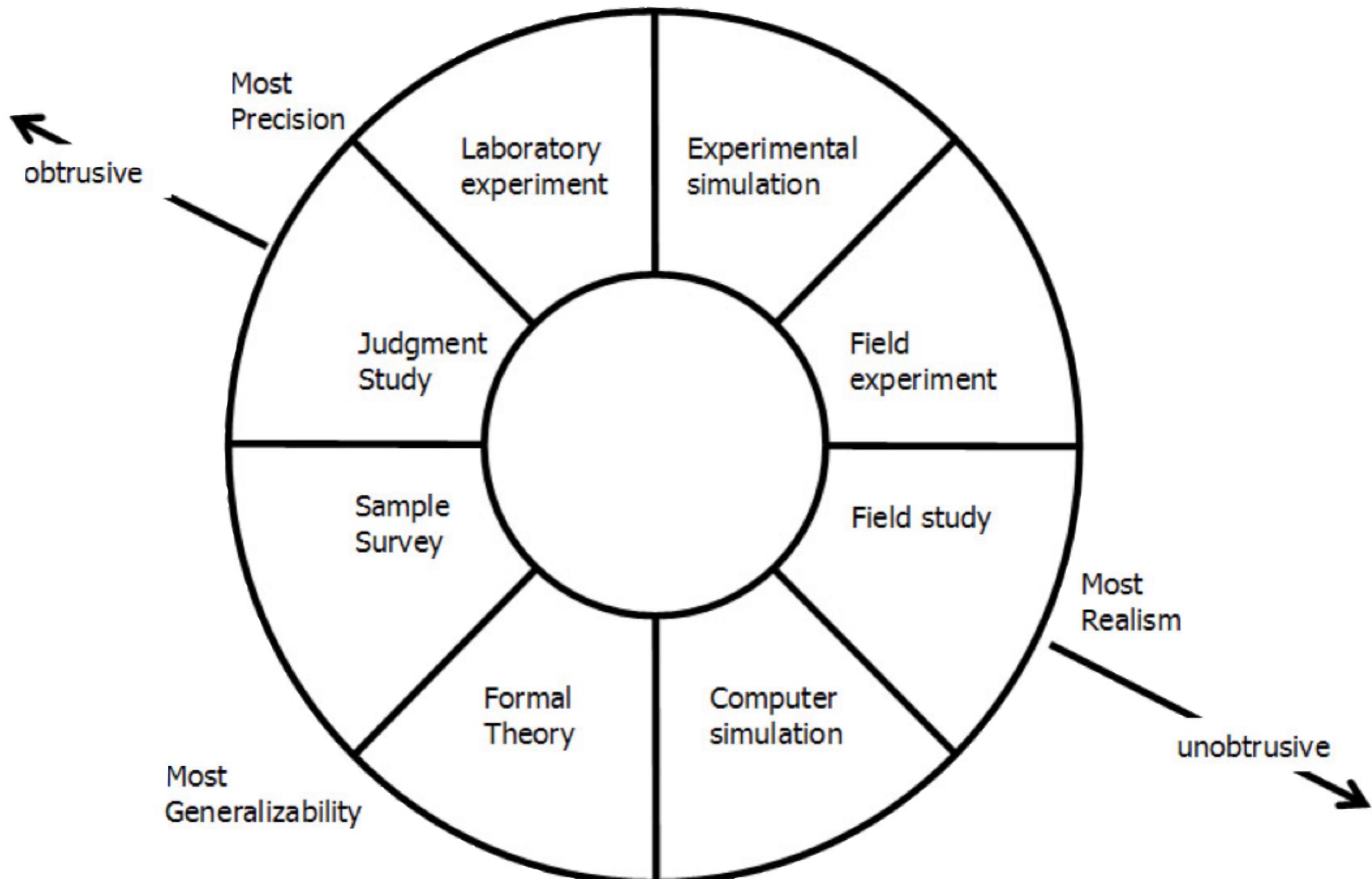
- **Develop a Flu Dashboard for General Public**
  - Geographic distribution of flu cases
  - Temporal change in flu cases per week
  - Comparison between countries and year
  - Drill down to time windows or geographic locations
- **Data**
  - Same as this morning but for all countries and for 5 years (2010 - 2015) with 52 weeks per year

# Five Design Sheet Exercise

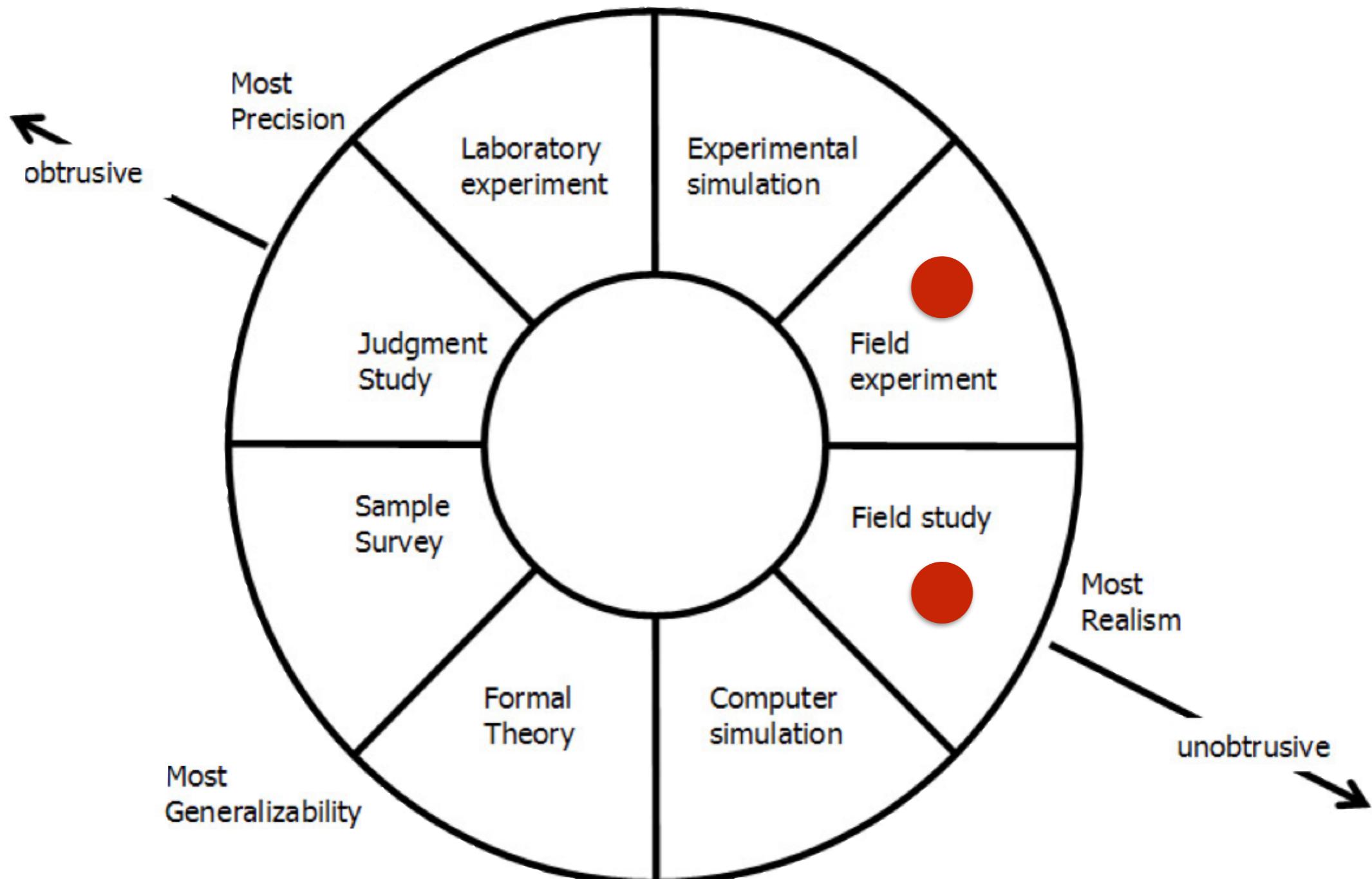
- **Sheet 1**
  - everyone sketches separately
  - filter and group together (merge)
- **Sheet 2 - 5**
  - group exercise
- **Website:** <http://fds.design>

# Evaluation

# Validation Techniques



# Validation Techniques



# Qualitative Usability Testing

- **Field study**
  - Intended to be unobtrusive
  - Conducted in the actual situation (a user with their own analysis tasks and data)
  - Watching someone use your tool can be very informative; you may be surprised!

# Qualitative Usability Testing

- **Field experiment**
  - Design user tasks to simulate real analyses
  - Recruit group of users and arrange one-on-one sessions
  - Encourage thinking aloud and note top usability issues

# Qualitative Usability Testing

- Learn through observation
- Learn about behavior, not just opinions
- Not about proof, but about insight and context
- Test early and test often
- **Important that your users are assured that you are testing the software, not them**

# Quantitative User Testing

- Error Rate
- Time (see LineUp paper next week)
- Insights
  - **Oh boy!**

*See required reading for Week 5*

**Gremlin: an interactive visualization model for analyzing genomic rearrangements**

O'Brien TM, Ritz AM, Raphael BJ, Laidlaw DH

IEEE Trans. Vis. Comput. Graph., 2010

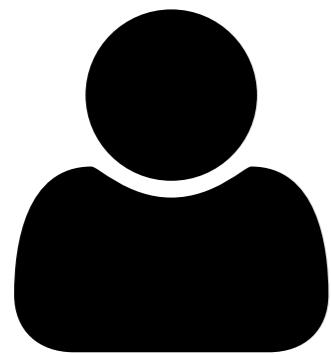
# Evaluation

**What is an effective visualization and what is not?**

**Considerations** - e.g. measuring effectiveness  
quantitatively is extremely problematic

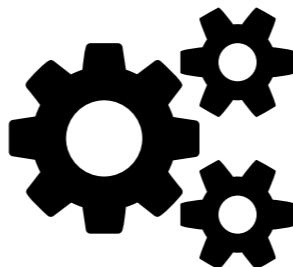
# Evaluation

1



human aspect

2



system complexity

3

? vs !

outcome is question,  
not answer

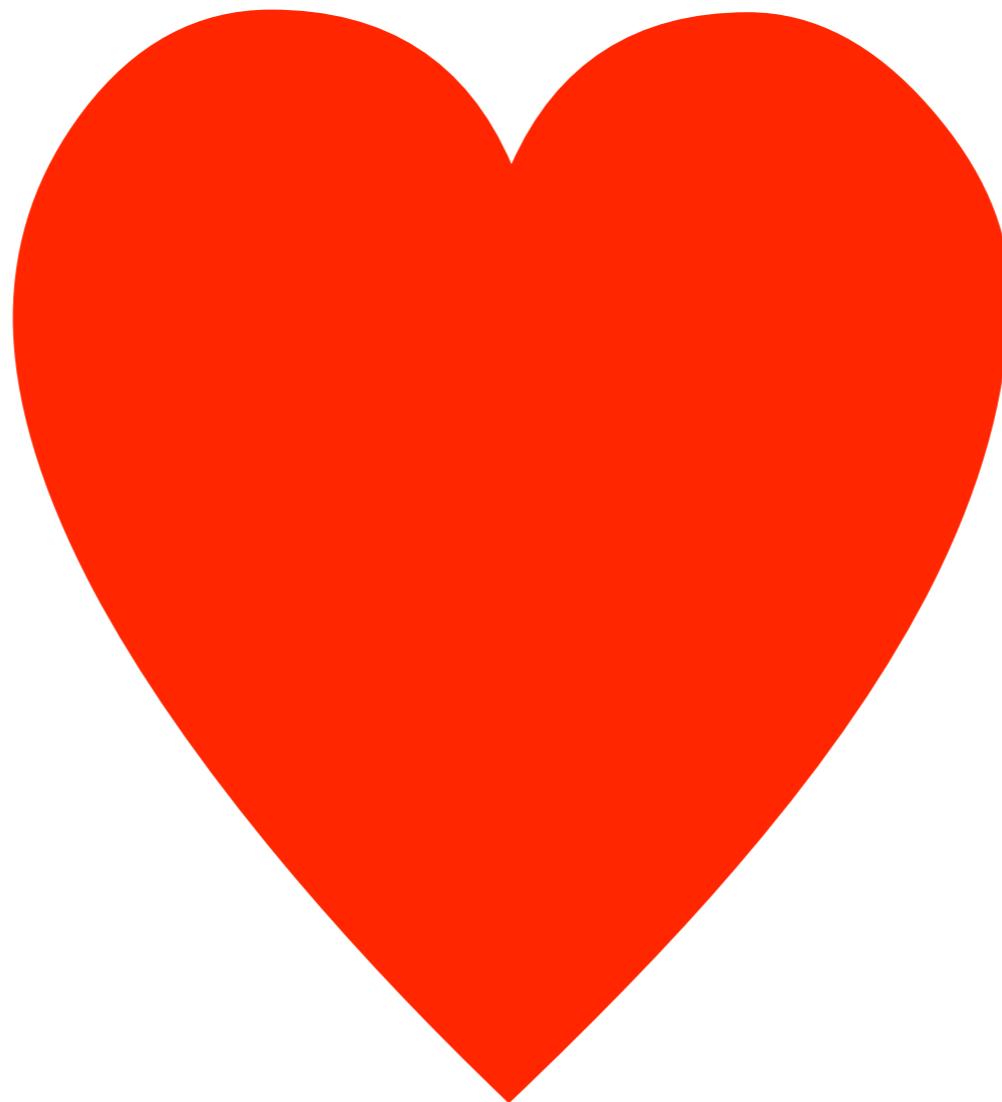
# Evaluation

**What is an effective visualization and what is not?**

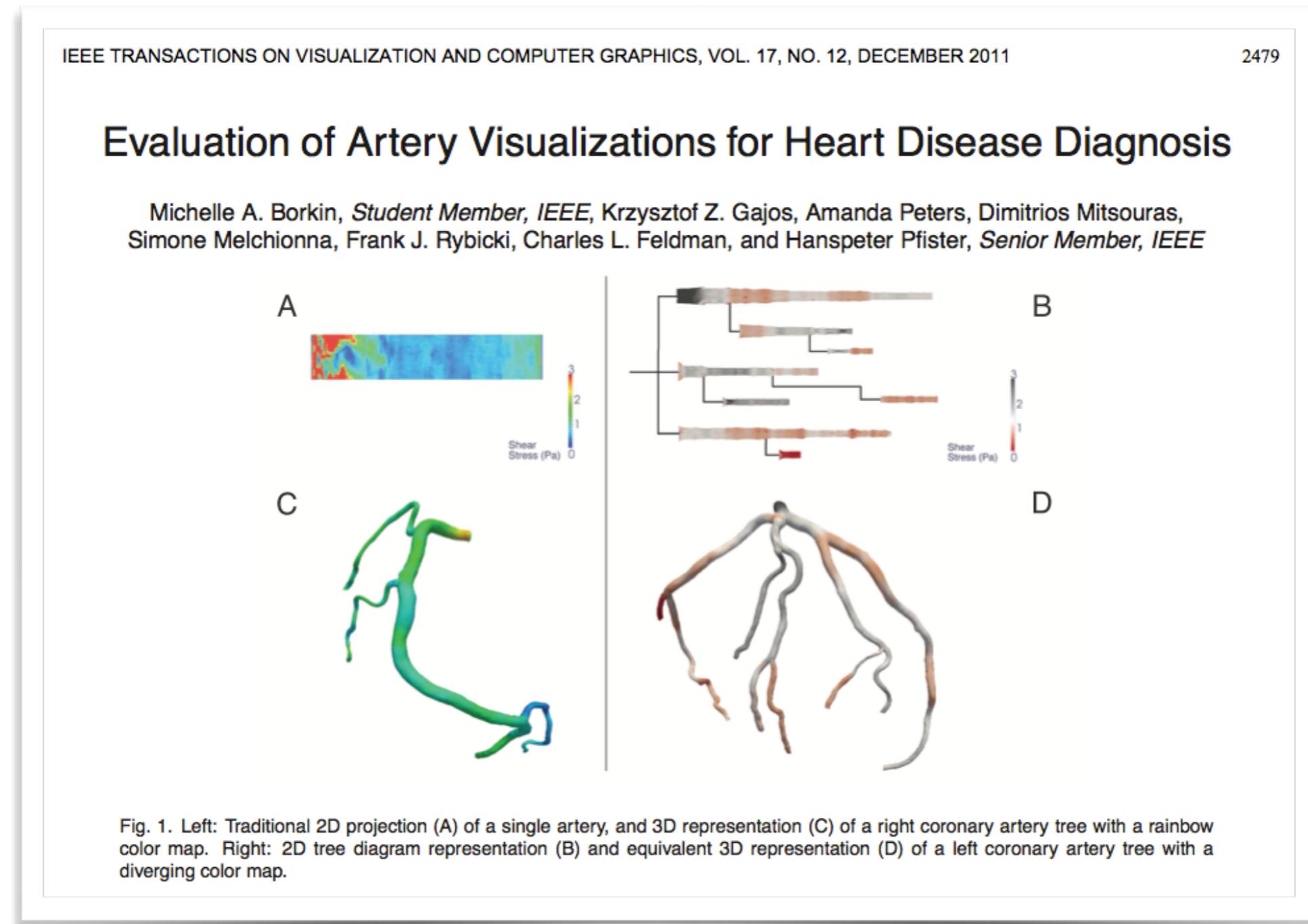
**Considerations** - e.g. measuring effectiveness quantitatively is extremely problematic

**Impact** - e.g. on developers (are we doing good work?), users (is this a good tool?), and reviewers (should this be published/funded?), e.g. in evaluating visual against analytical approaches

# When good visualization really matters

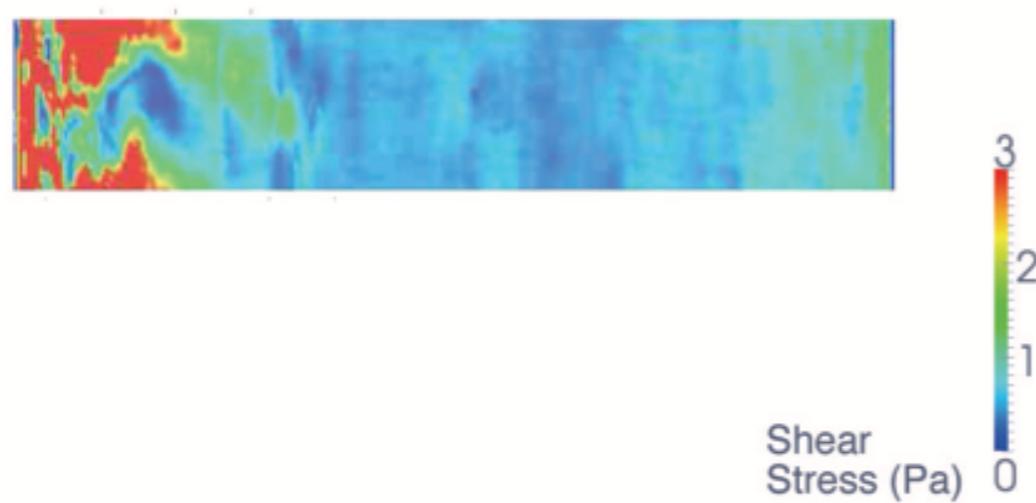


# Why we should ❤️ good visualizations

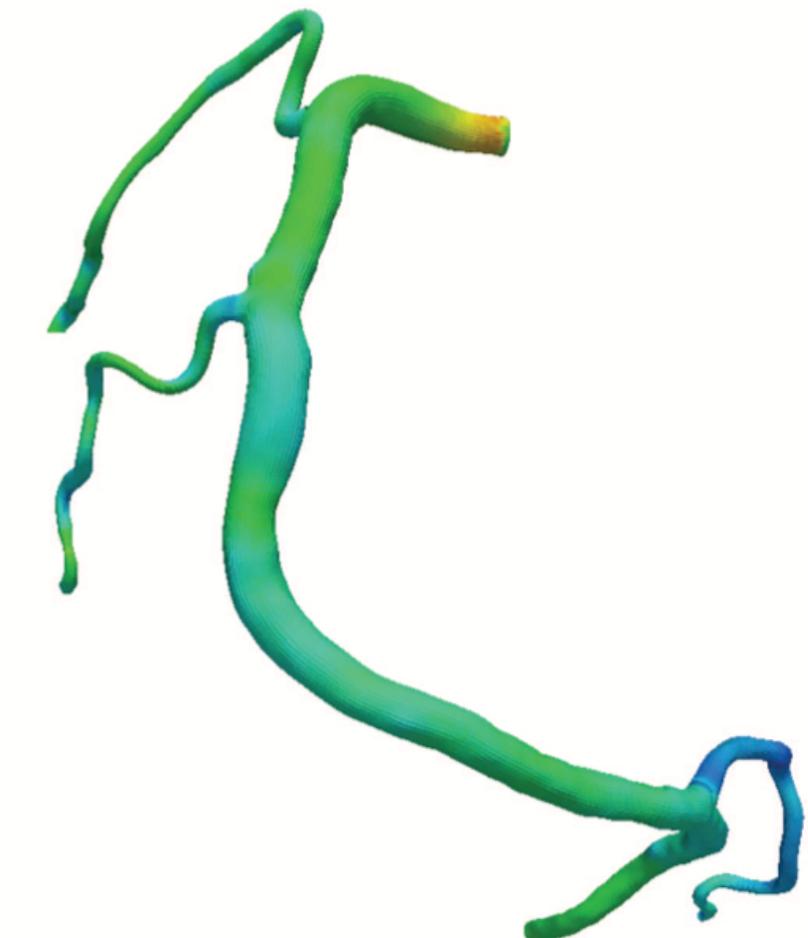


# Why we should ❤️ good visualizations

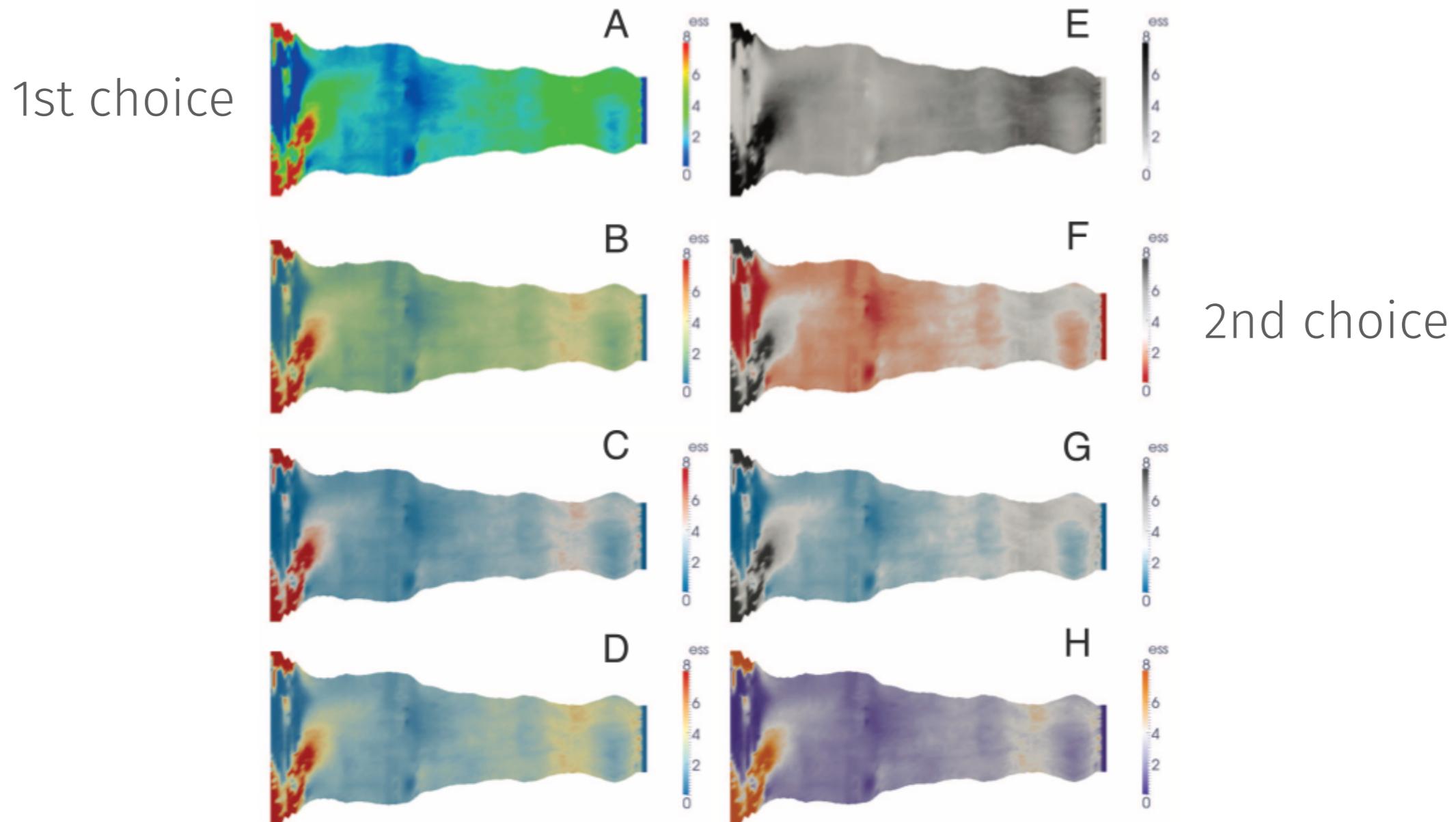
Traditional 2D projection  
of a single artery



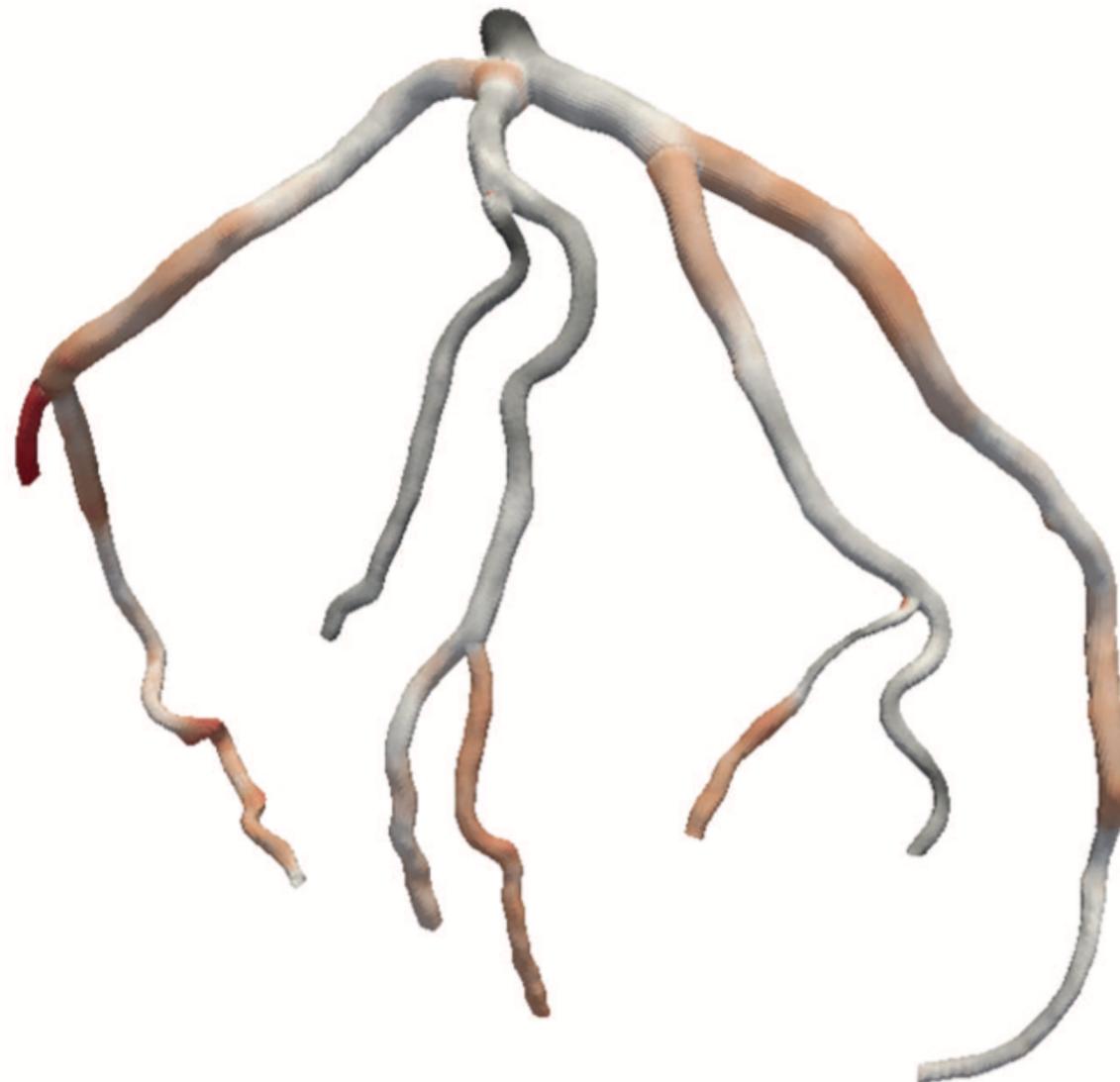
Traditional 3D projection  
of right artery tree



# Why we should ❤️ good visualizations



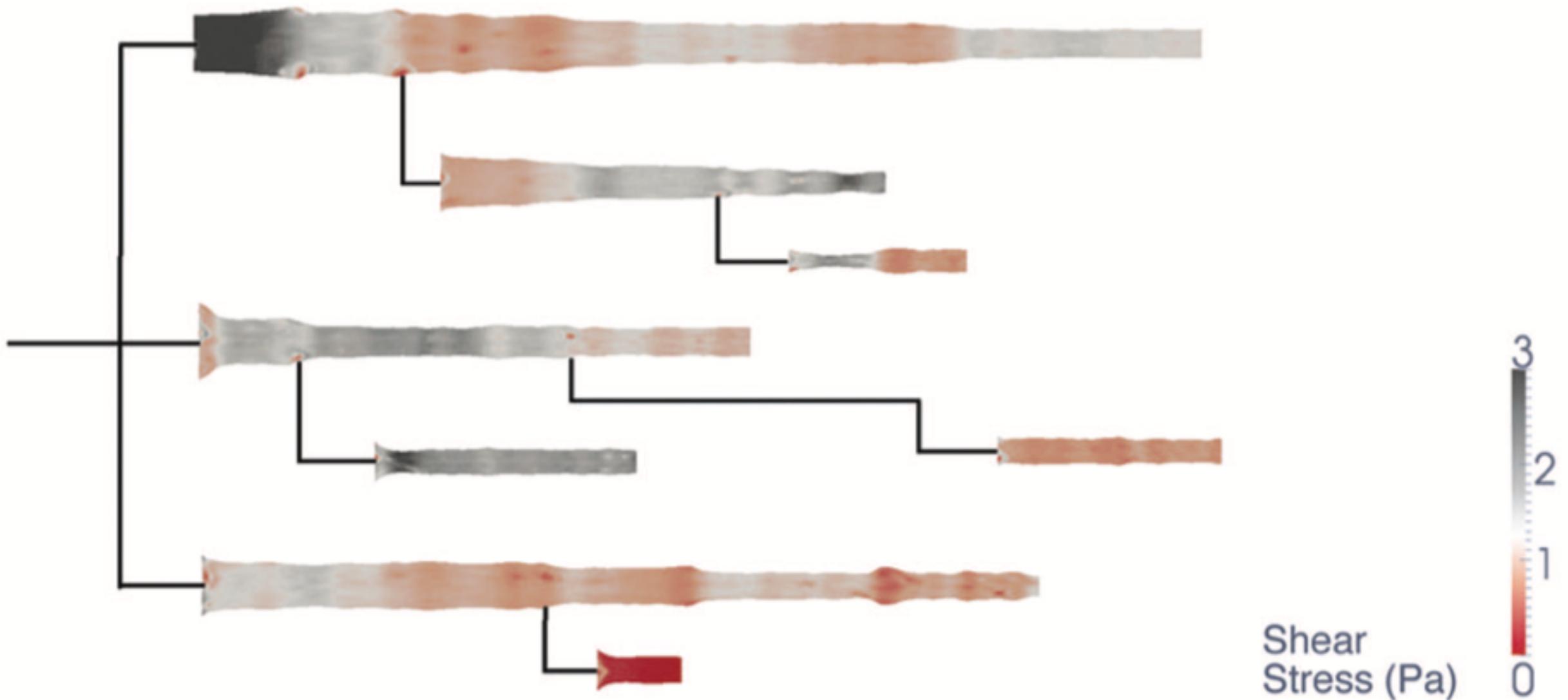
# Why we should ❤️ good visualizations



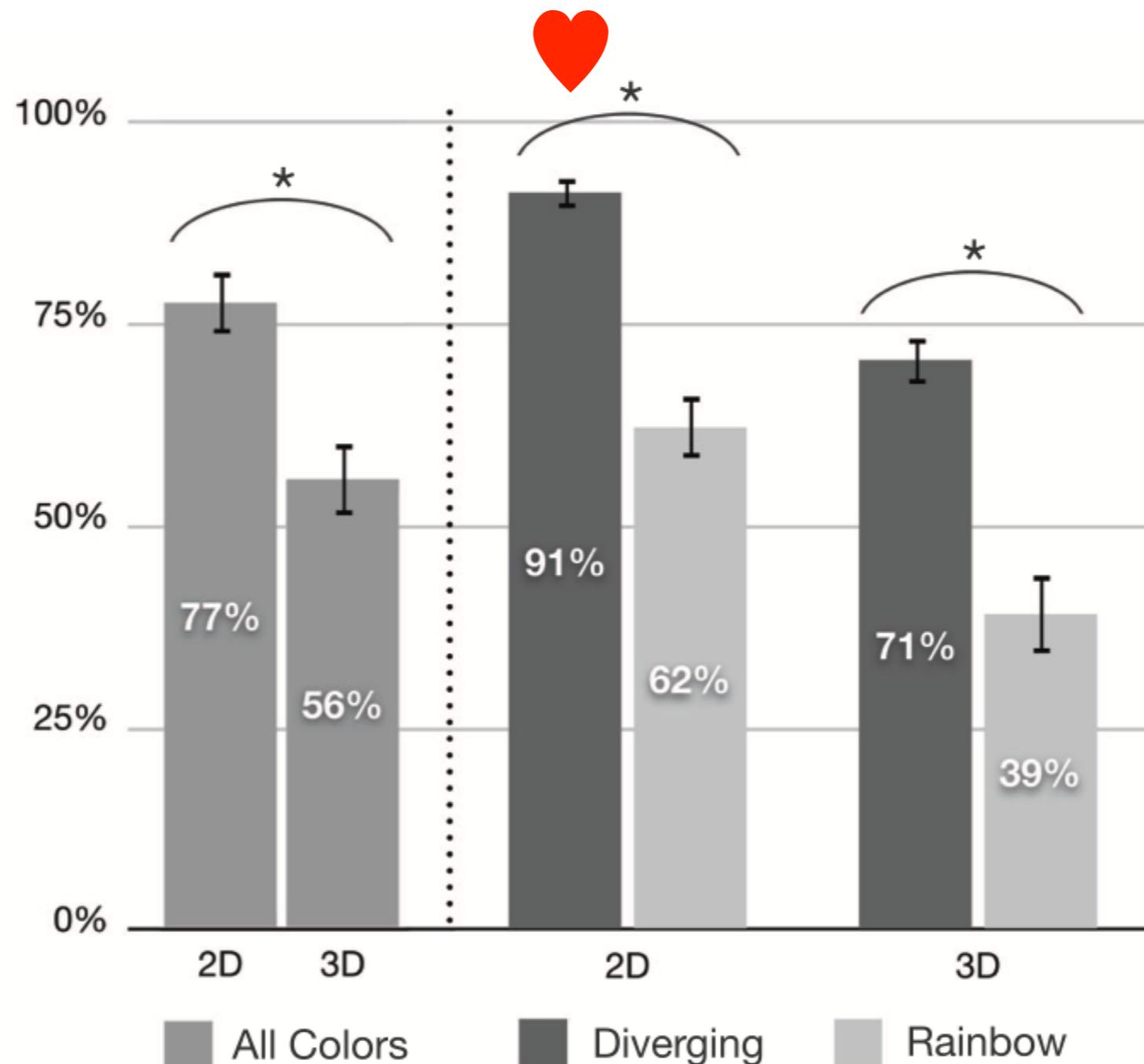
3D projection of left  
artery tree with  
diverging color map

# Why we should ❤️ good visualizations

Novel 2D projection of left artery tree



# Result: % of low ESS areas spotted



# Interactions

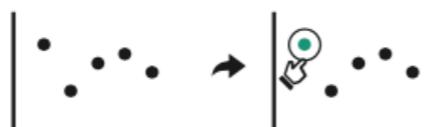
# Single View Interactions

## Manipulate

### ⌚ Change over Time



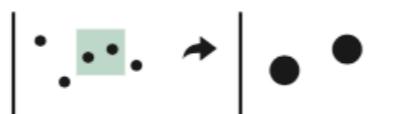
### 🔍 Select



### ניווט (Navigate)

#### → Item Reduction

→ Zoom  
Geometric or Semantic



#### → Pan/Translate



#### → Constrained



#### → Attribute Reduction

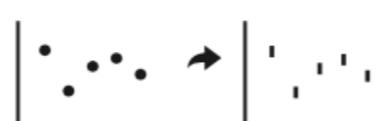
##### → Slice



##### → Cut



##### → Project



# Exploration: Hypothesis Generation

- Visualization for exploration is an “Exploratory Data Analysis” technique (Tukey 1977). Statistical graphics such as box plots and scatter plots are early examples.
- When there is a specific question that can easily be determined algorithmically (“What is the highest value?”), then visualization is usually not the right tool.
- When it is not clear what should be asked or when the answer can not be summarized easily (“What is the distribution of the values?”), then visualization is an excellent choice.
- **Visualization for exploration is challenging because the data sets are getting bigger and more heterogeneous.**

# Exploration: Information Seeking Mantra

- In explorative settings the user is normally dealing with large amounts of data.
- Impossible to grasp everything at once.
- Solution: Make visualizations **interactive** to support the user in exploring subsets of the data at different resolutions.
- Ben Shneiderman's **Information Seeking Mantra**:
  - Overview first
  - Zoom and filter
  - Then details on demand.

# Exploration: Visual Analytics Mantra

- Keim-Mansmann-Thomas **Visual Analytics Mantra:**
  - Analyze first
  - Show the important
  - Zoom, filter and analyze further
  - Then Details on demand

# Interaction Techniques: Basic Methods

## - Selection

- click on elements
- lasso/drag over elements
- hover over elements

## - Manipulation

- move elements
- sort elements
- add or delete elements (or filter elements)

# Interaction Techniques: Advanced Methods

- Focus + Context (*single view*)
- Overview + Detail (*multiple views*)
- Brushing + Linking (*multiple views*)
- Zooming + Panning
  - geometric zooming (modifies graphical representation)
  - semantic zooming (modifies selection of data)
  - often combined (Google Maps)

## Pan and Zoom in D3

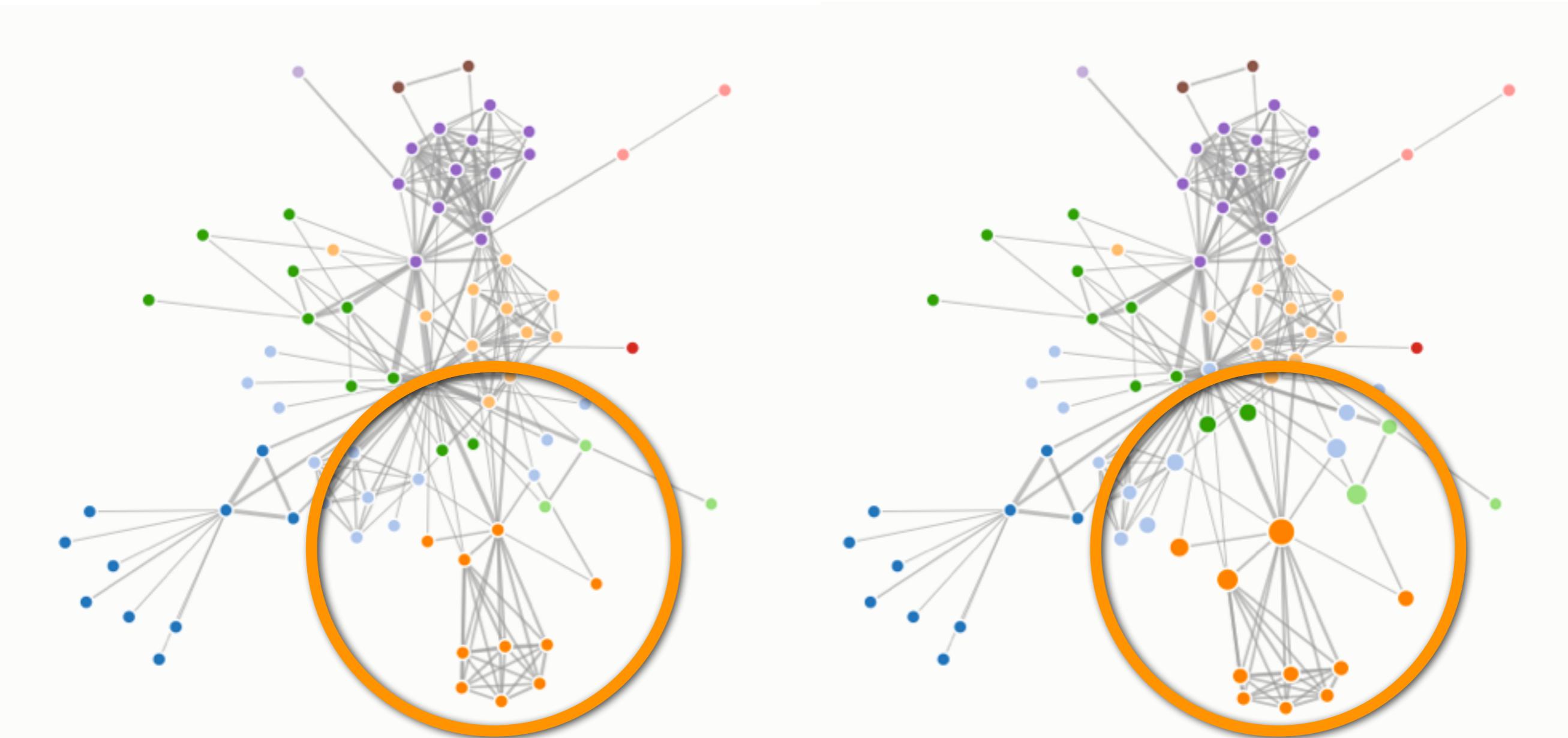
<https://bl.ocks.org/mbostock/db6b4335bf1662b413e7968910104f0f>

## Semantic Zoom

<https://bl.ocks.org/mbostock/3680957>

# Interaction Techniques: Focus + Context

Fisheye Distortion



Fisheye and Rectilinear Cartesian Distortion in D3

<https://bostocks.org/mike/fisheye/>

Mike Bostock, <http://www.d3.js>

# Multi-View Interactions

## Facet

### ④ Juxtapose and Coordinate Multiple Side-by-Side Views

→ Share Encoding: Same/Different

→ *Linked Highlighting*



→ Share Data: All/Subset/None



→ Share Navigation



		Data		
		All	Subset	None
Encoding	Same	Redundant	Overview/Detail	Small Multiples
	Different	Multiform	Multiform, Overview/Detail	No Linkage

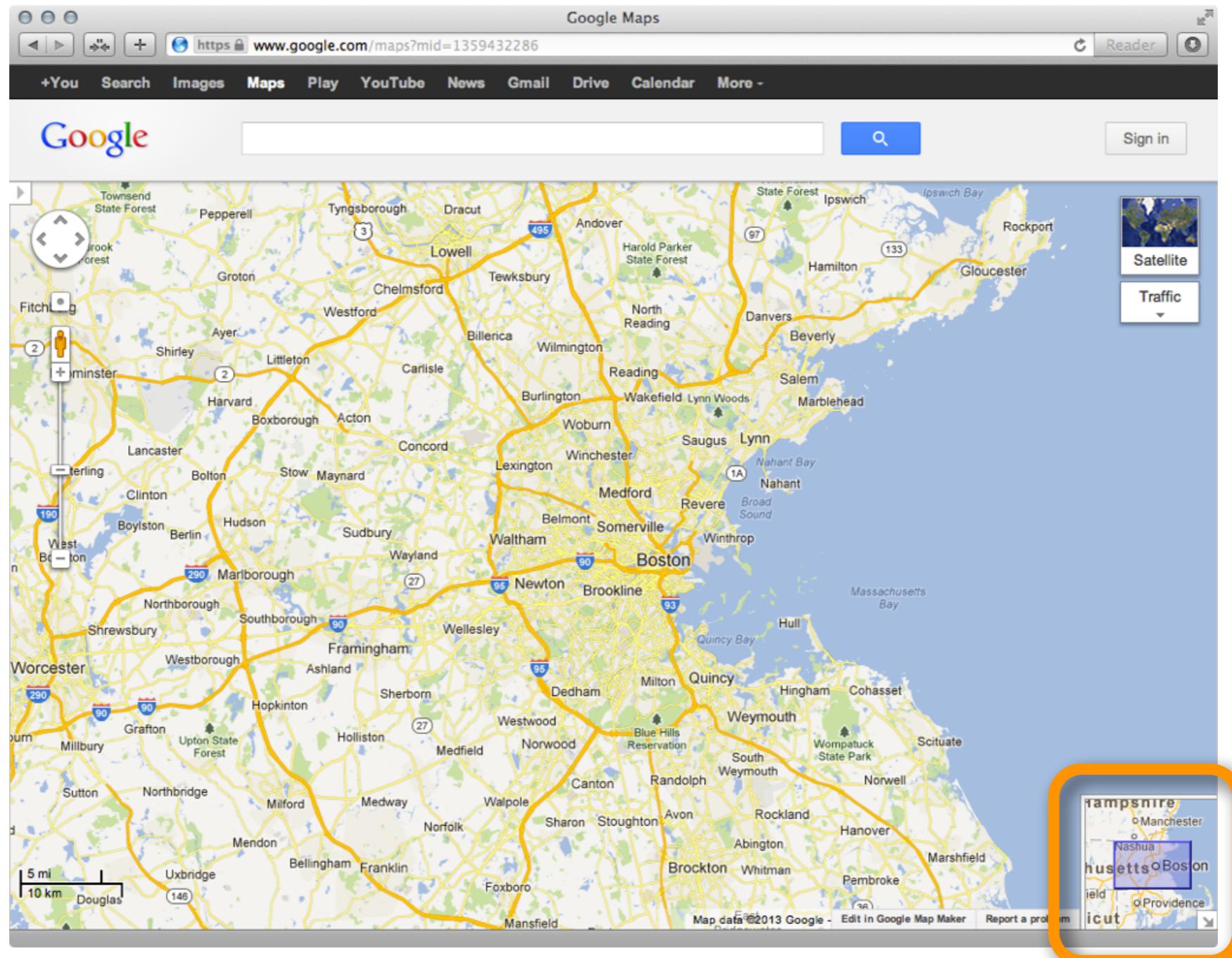
### ⑤ Partition into Side-by-Side Views



### ⑥ Superimpose Layers



# Interaction Techniques: Overview + Detail



Then ...

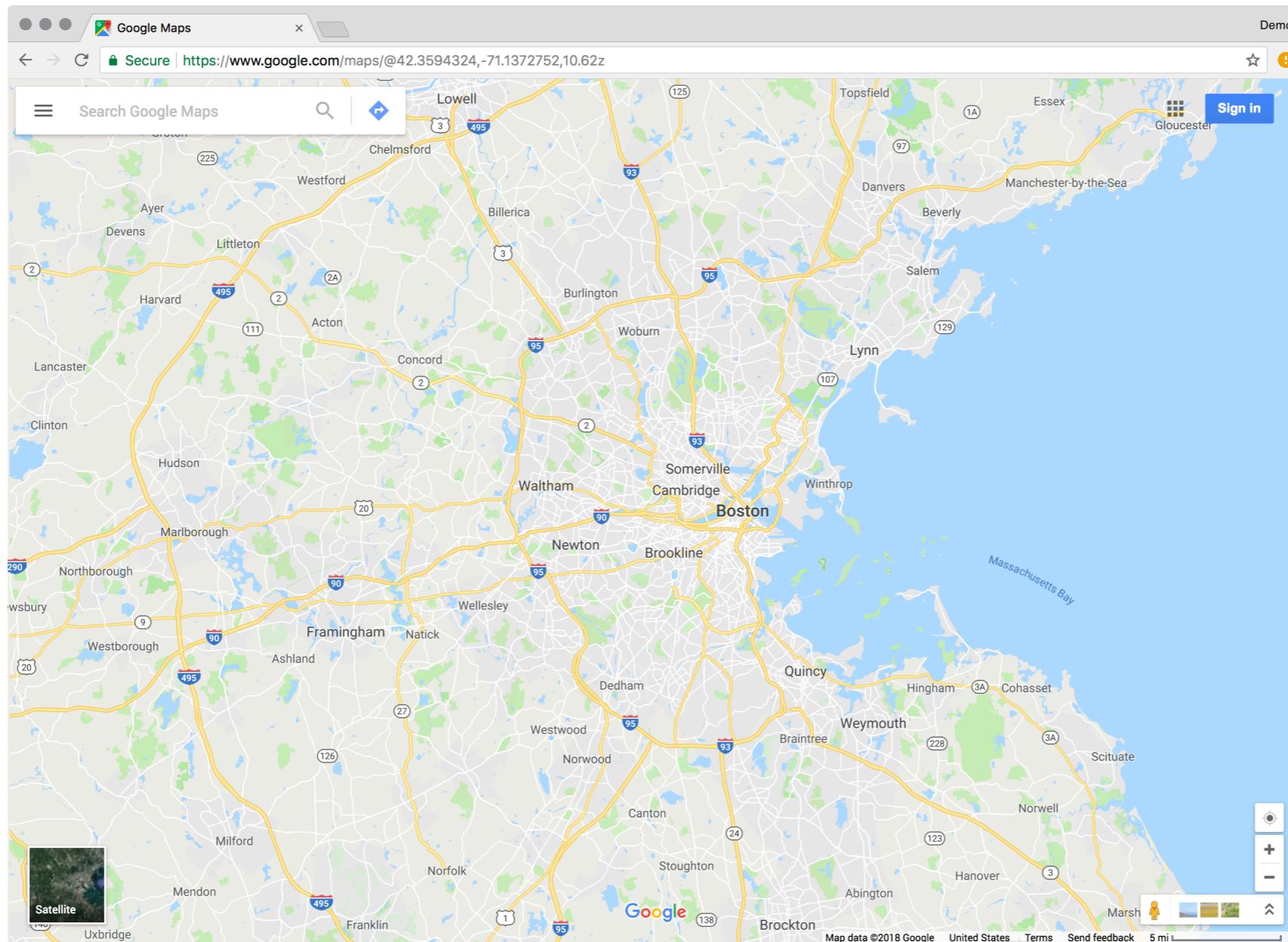
## Brush and zoom in D3

<https://bl.ocks.org/mbostock/f48fcdb929a620ed97877e4678ab15e6>

<https://bl.ocks.org/mbostock/34f08d5e11952a80609169b7917d4172> (also overview + detail)

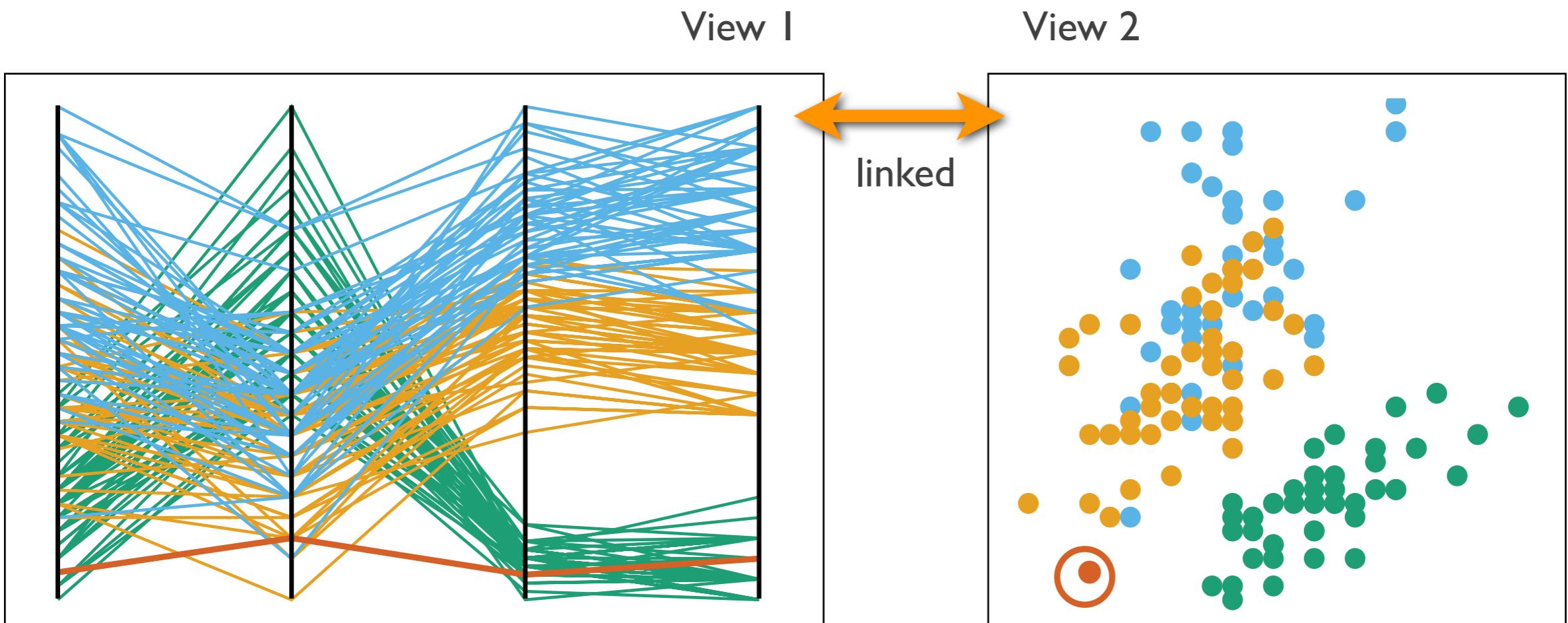
<https://maps.google.com>

# Interaction Techniques: Overview + Detail



... & today!

# Interaction Techniques: Brushing + Linking



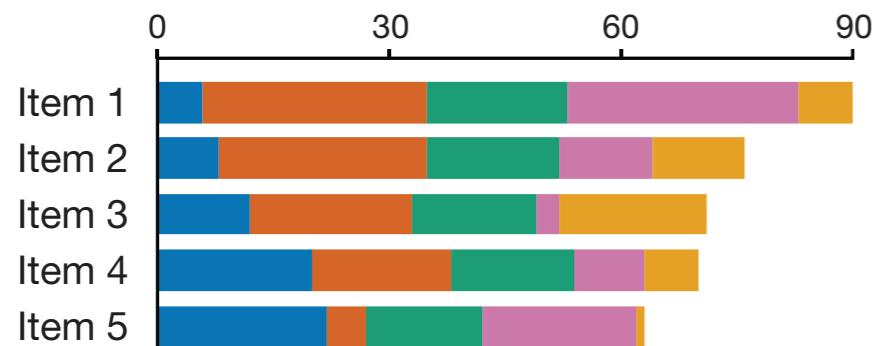
## Brushing and Linking

<https://bl.ocks.org/mbostock/4063663>

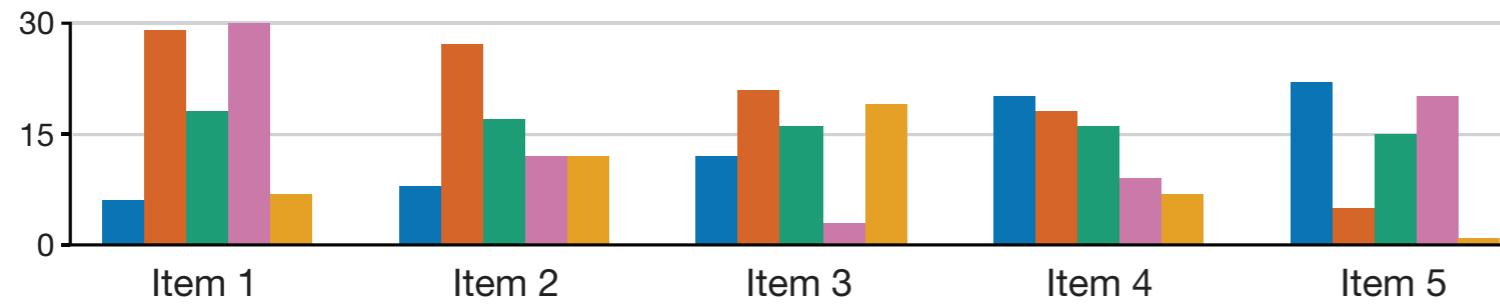
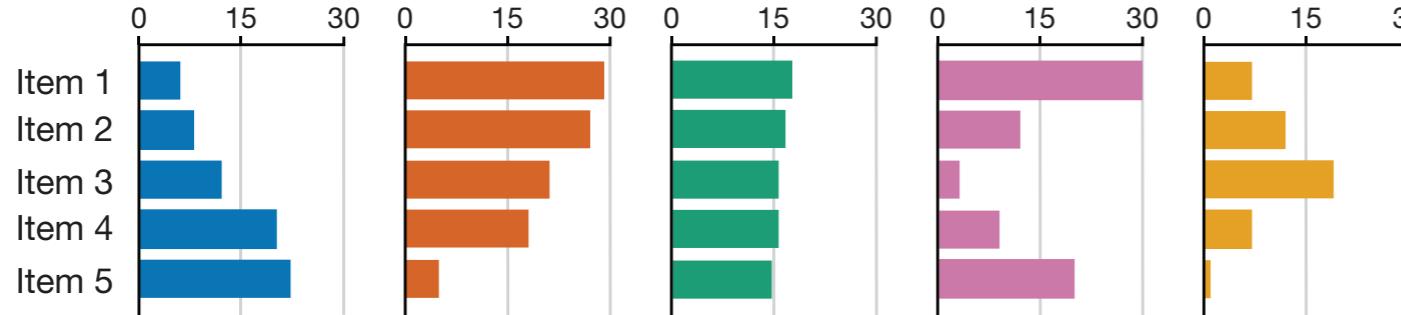
<https://stateofobesity.org/adult-obesity/>

# Optional Homework

# Re-Create these Plots in Altair



	1	2	3	4	5
Item	6	29	18	30	7
Item	8	27	17	12	12
Item	12	21	16	3	19
Item	20	18	16	9	7
Item	22	5	15	20	1

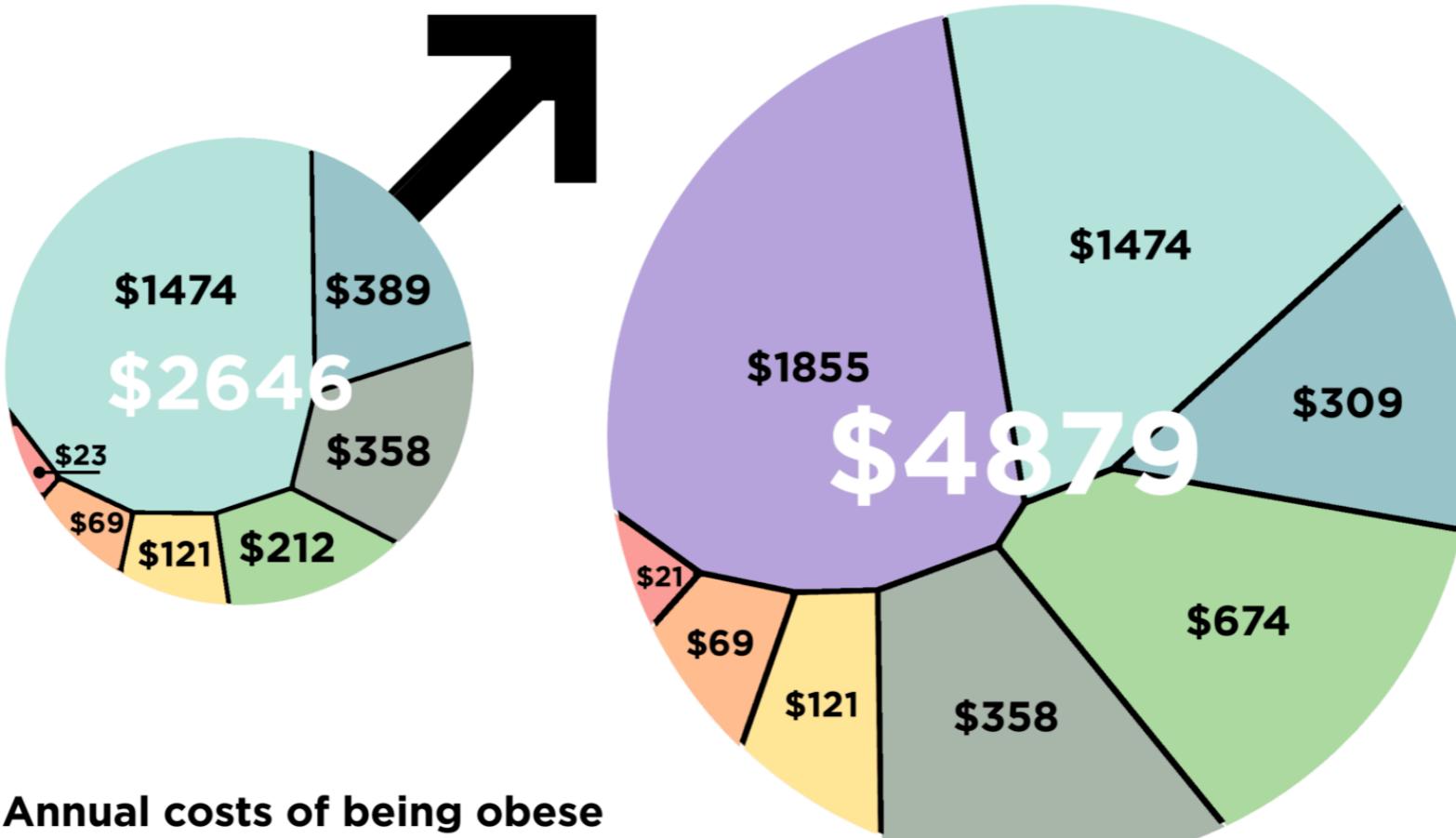


<https://altair-viz.github.io>

# Redesign Exercise

# Redesign Exercise

## The Individual Costs of Being Obese in the United States



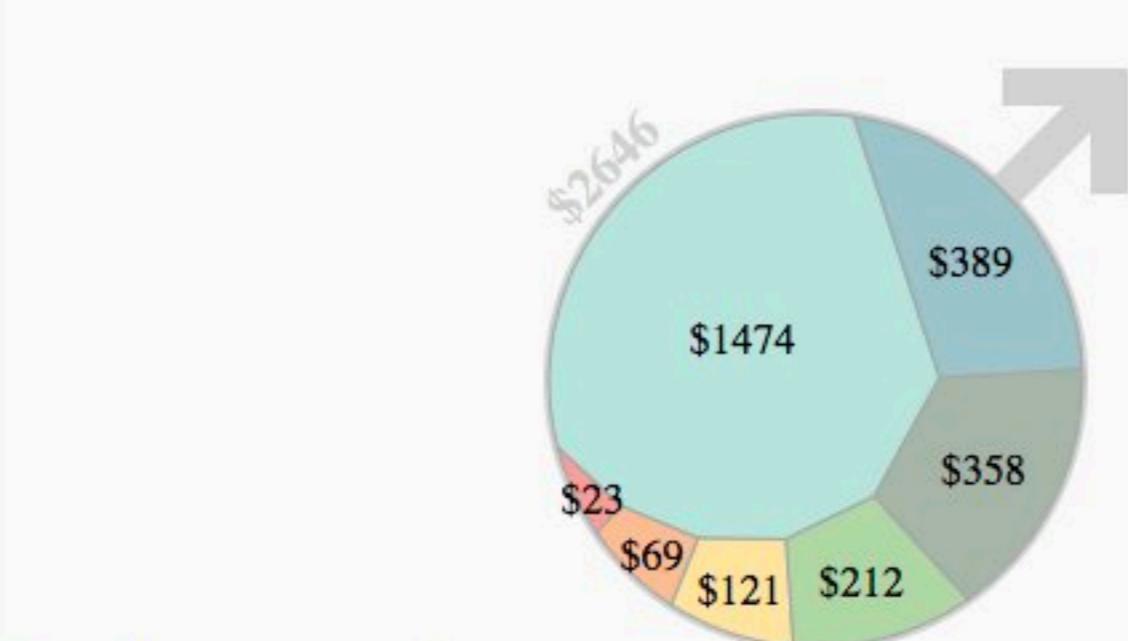
### Annual costs of being obese

- Wage Discrimination
- Direct Medical
- Short-term Disability
- Productivity (Presenteeism)
- Sick Leave (Absenteeism)
- Life Insurance
- Disability Pension Insurance
- Gasoline for Cars

**howmuch**.net

# Redesign Exercise

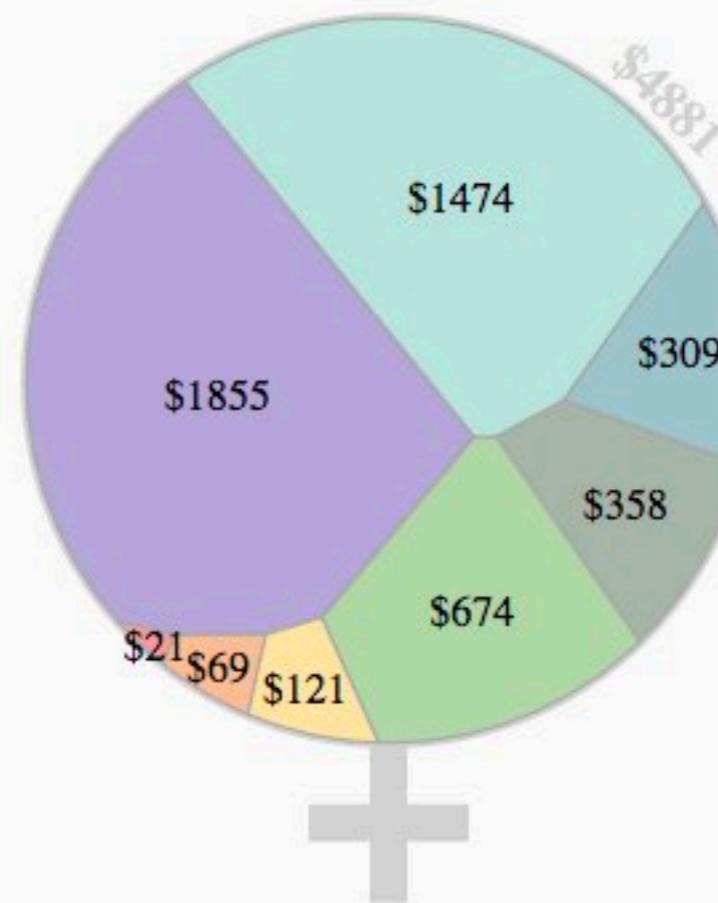
## The Individual Costs of Being Obese in the U.S. (2010)



Annual costs of being obese

- Wage Discrimination
- Direct Medical
- Short-term Disability
- Productivity (Presenteeism)
- Sick Leave (Absenteeism)
- Life Insurance
- Disability Pension Insurance
- Gasoline for cars

Remake of HowMuch.net's post 'The Costs of Being Fat, in Actual Dollars'

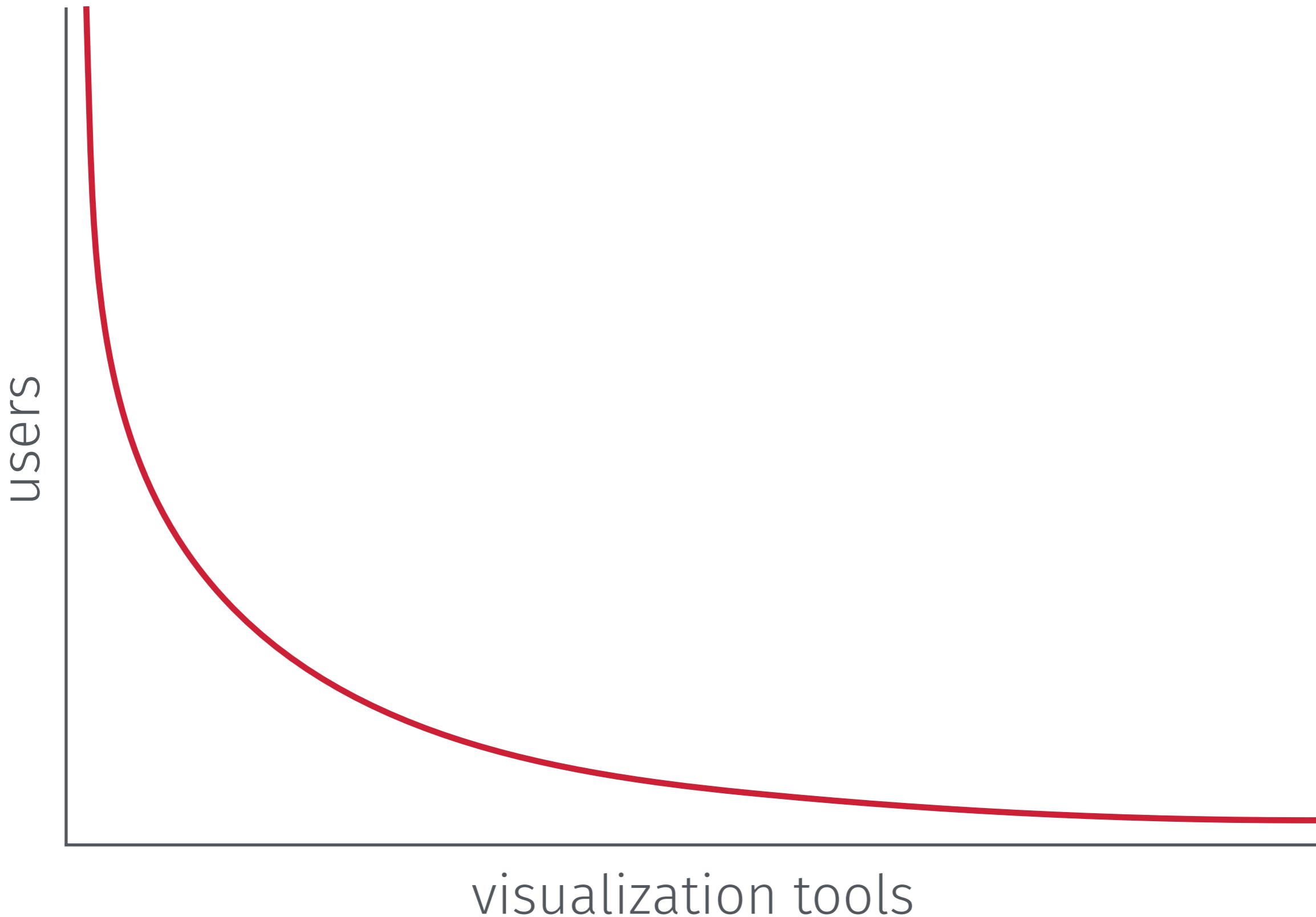


by @\_Kcnarf

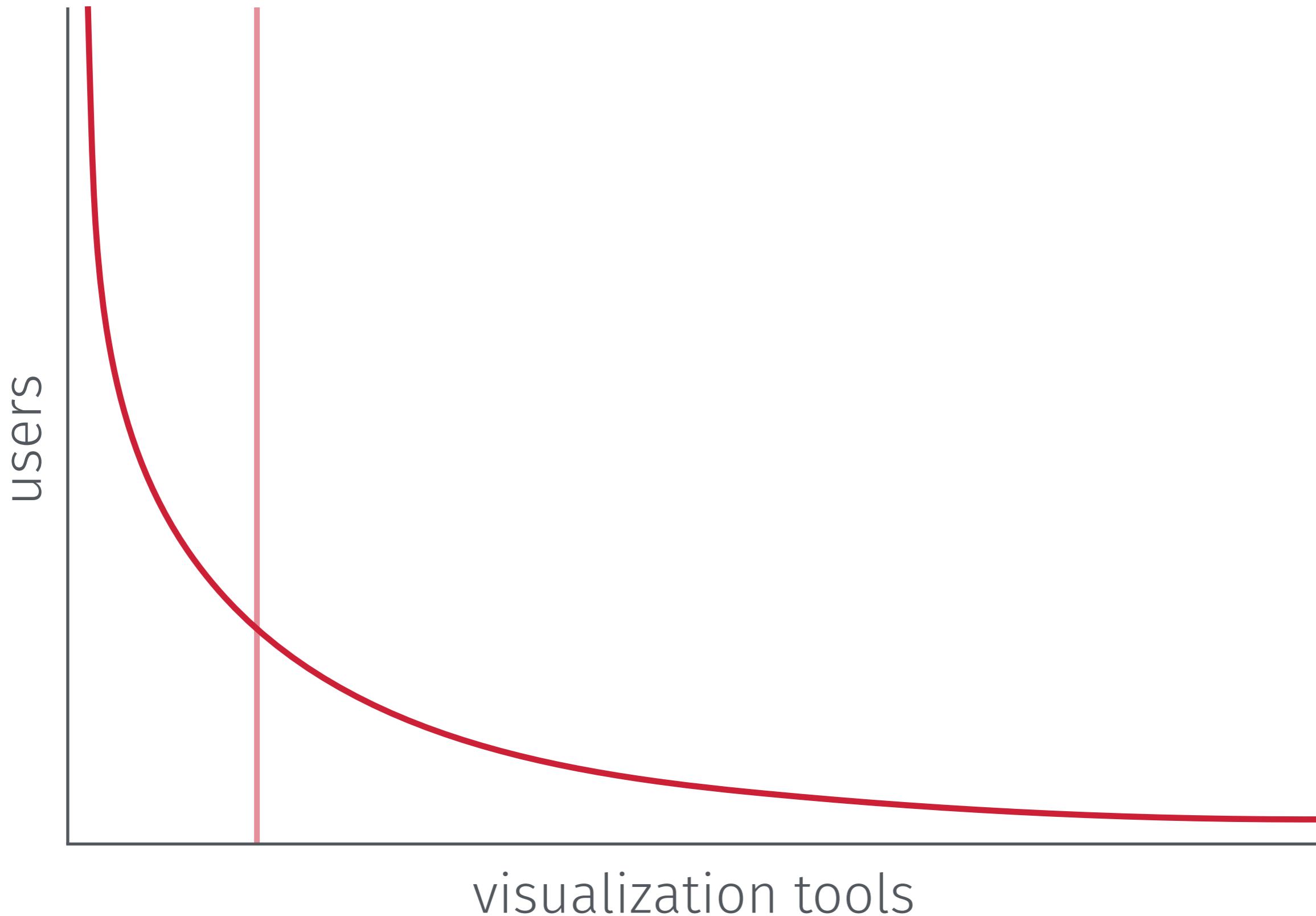
[bl.ocks.org/Kcnarf/e649c8723eff3fd64a23f75901910930](https://bl.ocks.org/Kcnarf/e649c8723eff3fd64a23f75901910930)

# The Curse of the Long Tail

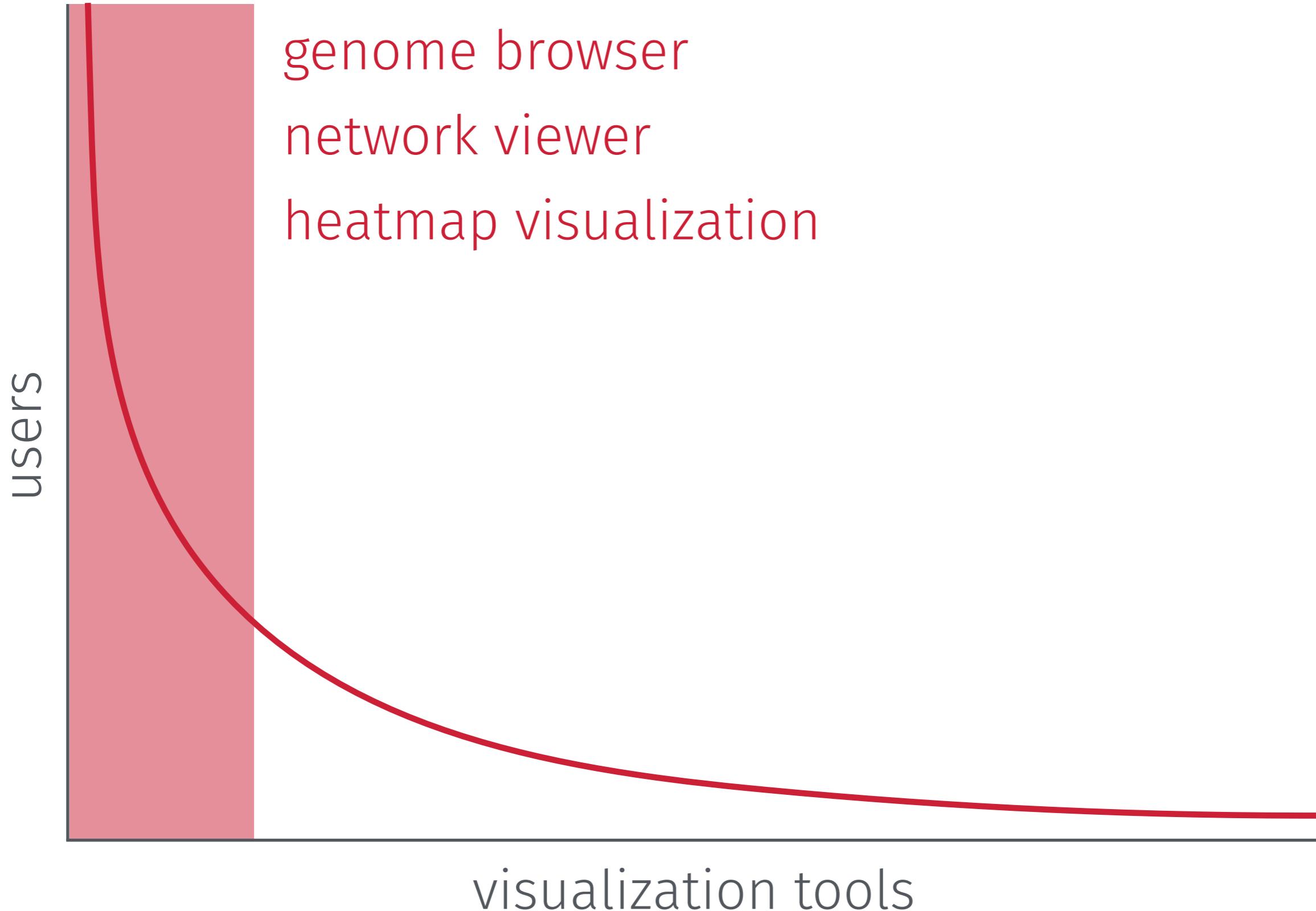
# Data Visualization: The Long Tail



# Data Visualization: The Long Tail



# Data Visualization: The Long Tail



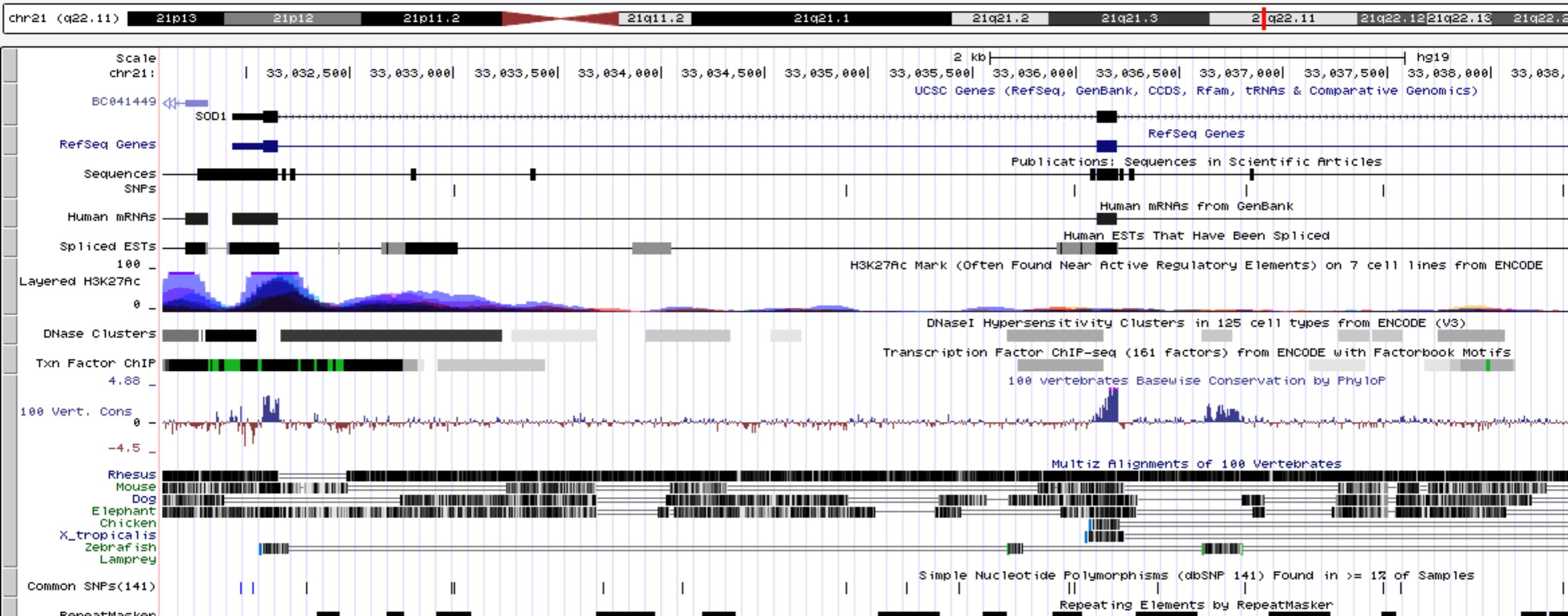

[Genomes](#)
[Genome Browser](#)
[Tools](#)
[Mirrors](#)
[Downloads](#)
[My Data](#)
[View](#)
[Help](#)
[About Us](#)

## UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr21:33,031,597-33,041,570 9,974 bp. enter position, gene symbol or search terms

go



move start

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to new position.

< 2.0 >

[track search](#) [default tracks](#) [default order](#) [hide all](#) [add custom tracks](#) [track hubs](#) [configure](#) [reverse](#) [resize](#) [refresh](#)

[collapse all](#)

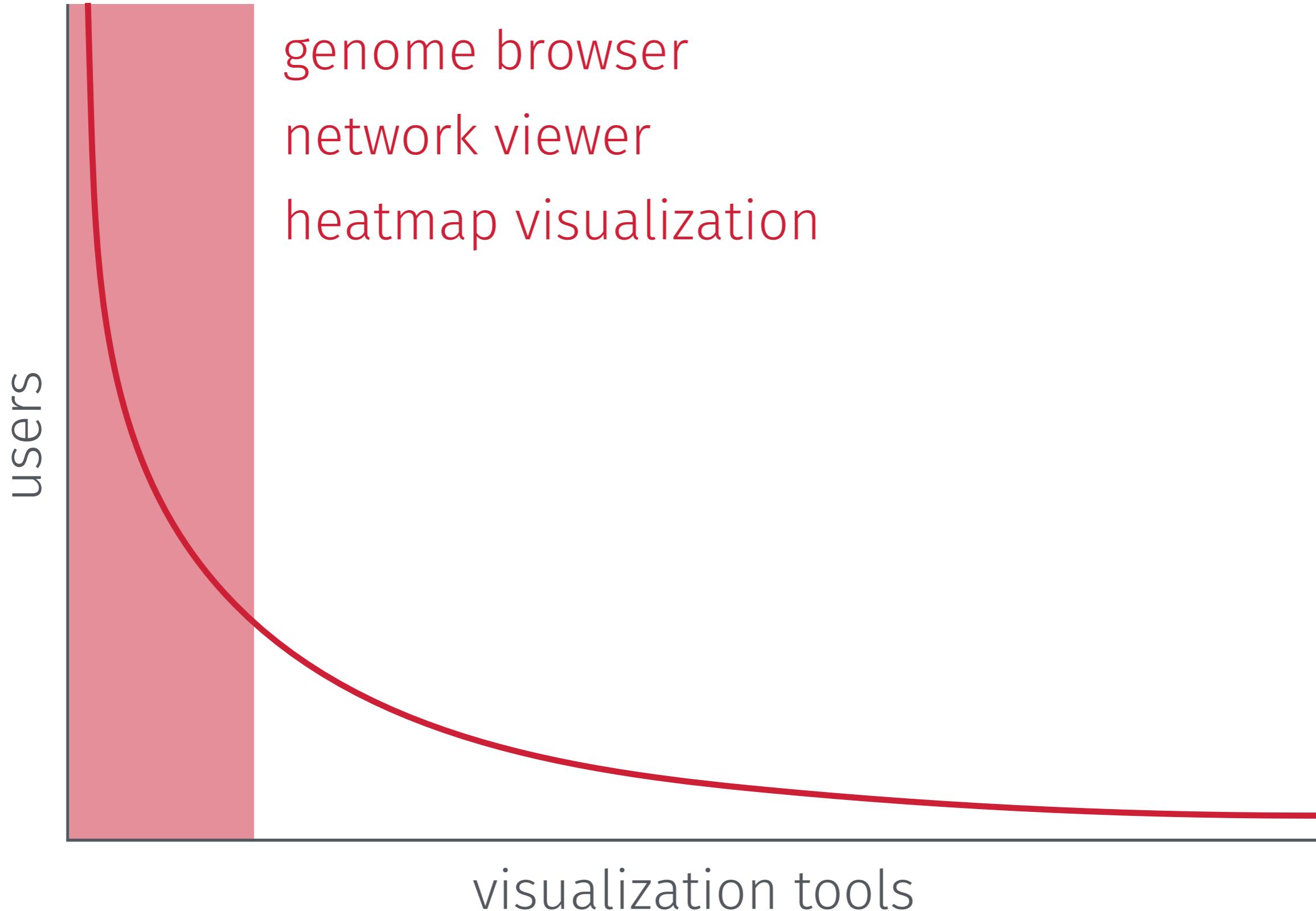
Use drop-down controls below and press refresh to alter tracks displayed.  
Tracks with lots of items will automatically be displayed in more compact modes.

[expand all](#)

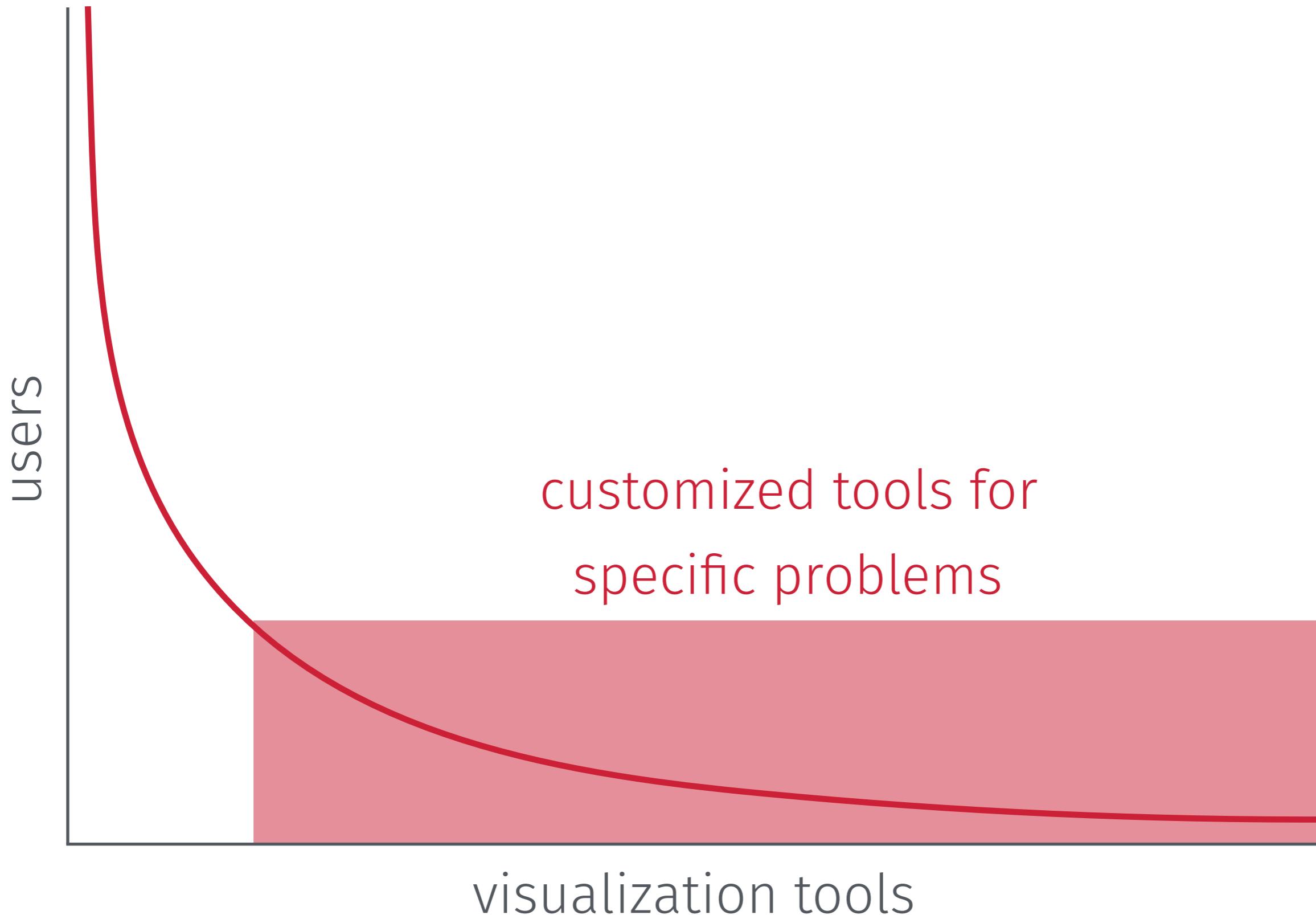
Mapping and Sequencing

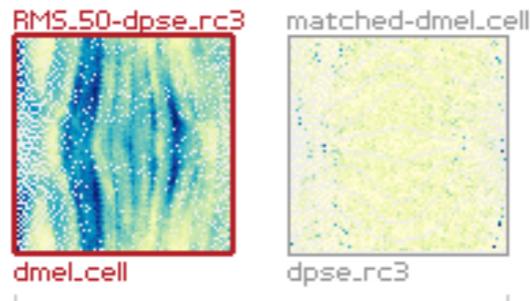
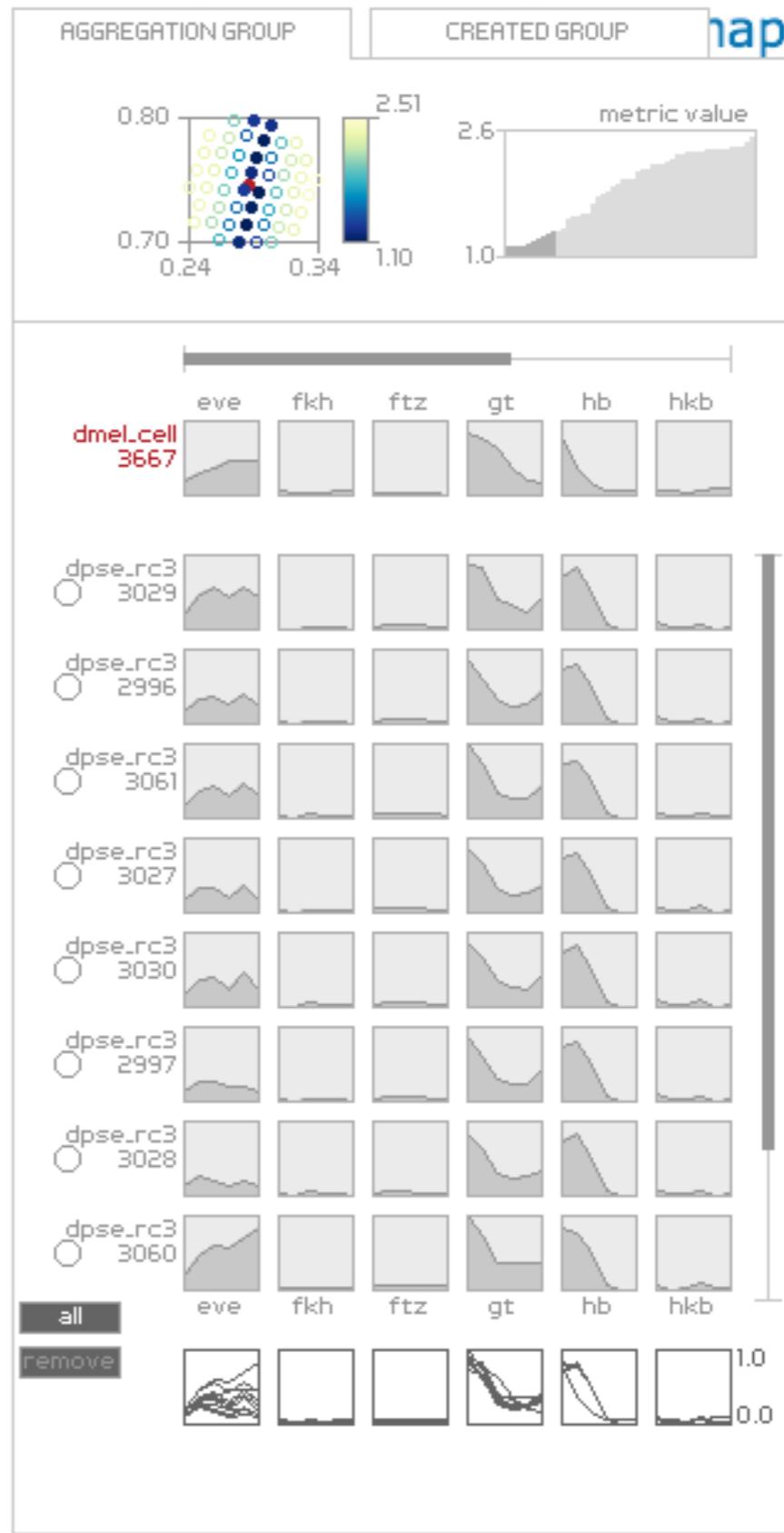
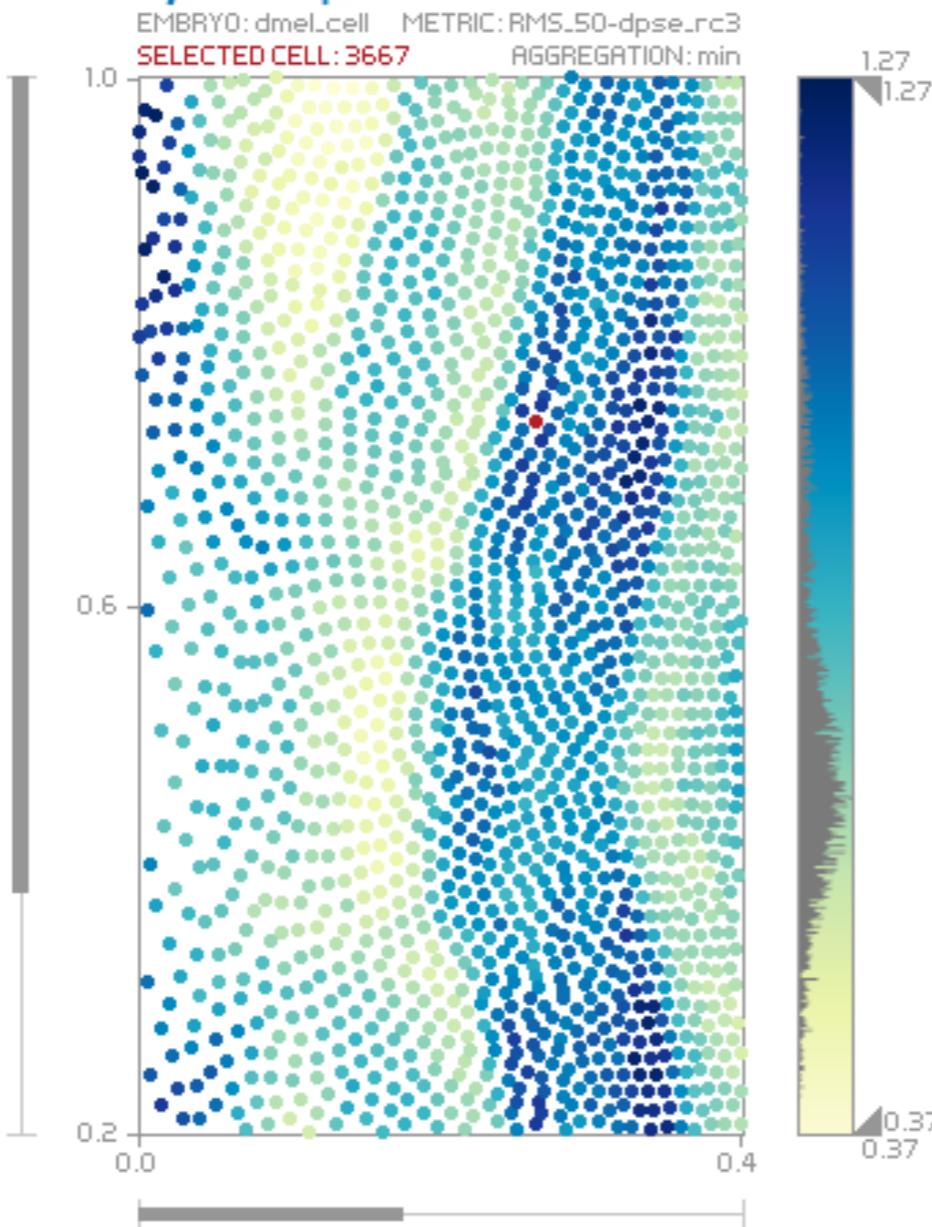
Display a menu for "http://genome-euro.ucsc.edu/cgi-bin/hgc?hgsid=200830848\_Pr7vSRBAuNDdCe9FN6cl81oDcNRD&c=chr21&o=33031596&t=33041570&g=phyloP100wayAll&i=phyloP100wayAll"

# Data Visualization: The Long Tail



# Data Visualization: The Long Tail



**Summaries****Embryo Map**

# Data Visualization: The Long Tail

