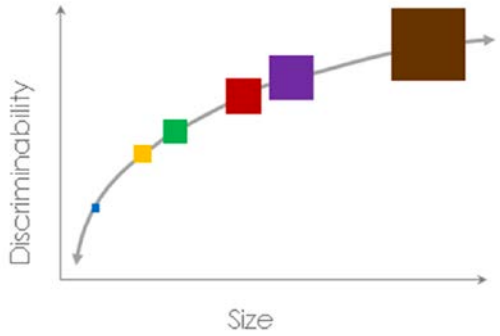
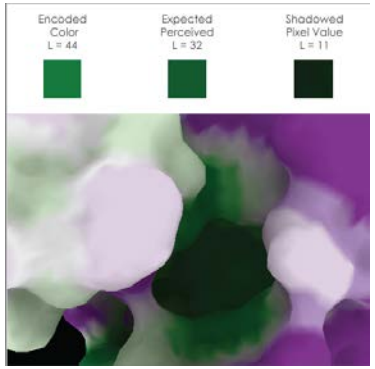
Visualization & Perception Across Scales

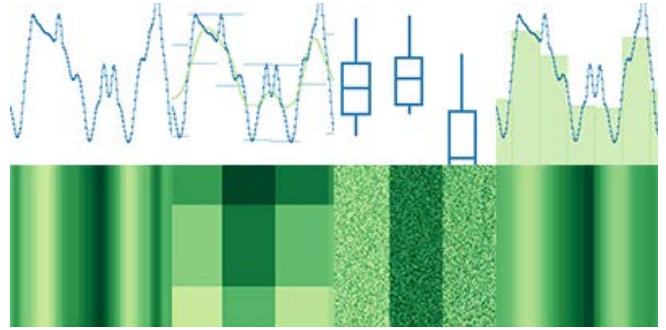
Danielle Albers Szafir
Department of Information Science
danielle.szafir@colorado.edu

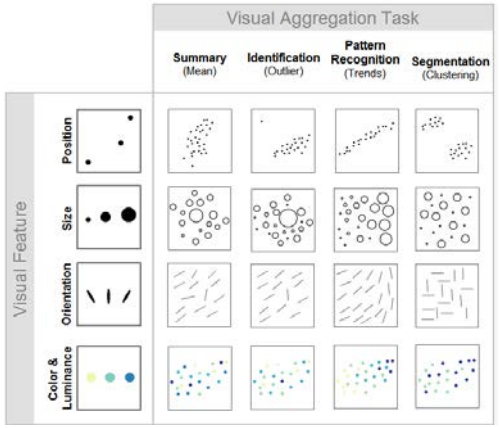
cmci

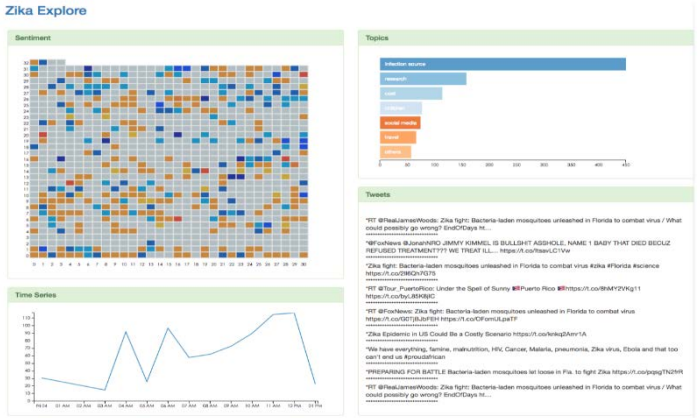
VisualLab

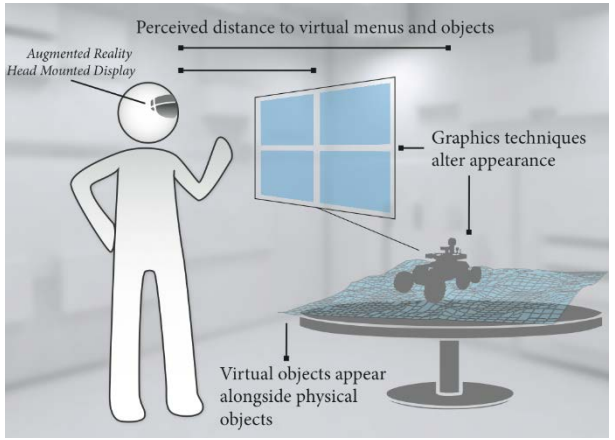

Discriminability
Size

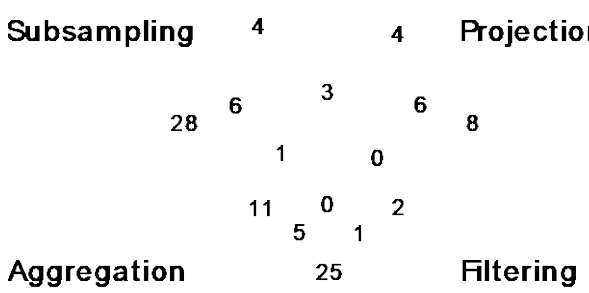

Encoded Color L = 44
Expected Perceived L = 32
Shadowed Pixel Value L = 11

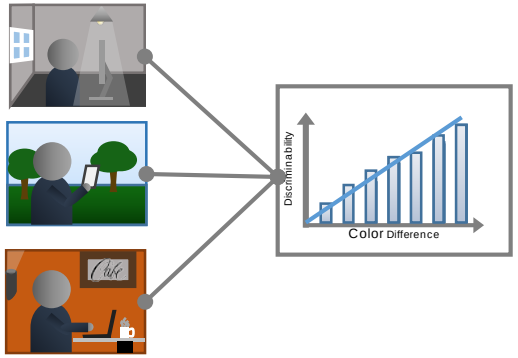


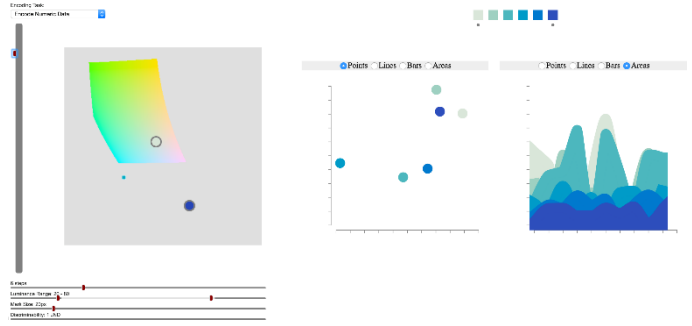

Visual Aggregation Task
Summary (Mean)
Identification (Outlier)
Pattern Recognition (Trends)
Segmentation (Clustering)
Visual Feature
Position
Size
Orientation
Color & Luminance

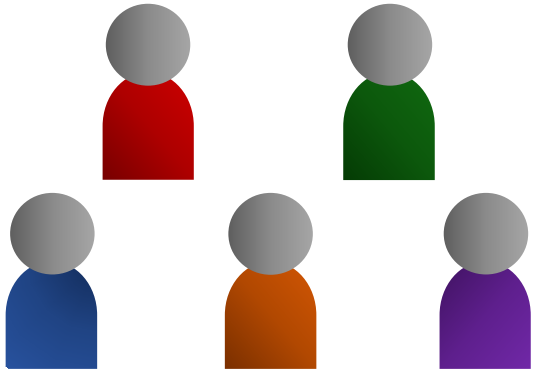

Zika Explore
Sentiment
Topics
Time Series
Tweets


Perceived distance to virtual menus and objects
Augmented Reality Head Mounted Display
Graphics techniques alter appearance
Virtual objects appear alongside physical objects

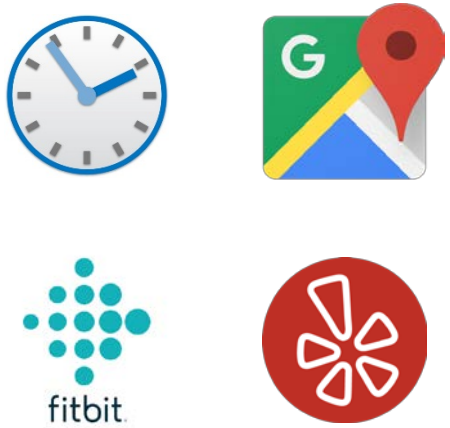

Subsampling
Projection
Aggregation
Filtering
4
4
6
3
6
28
8
1
0
11
0
2
5
1
25


Discriminability
Color Difference





More Analysts
(& Fewer Experts)



More (& Different) Data

?

More Questions

Why don’t we just compute the answer?

Data Sample 1:

Mean(x) = 9

Variance(x) = 11

Correlation(x, y) = 0.816

Regression: y = 3 + 0.5x

Data Sample 2:

Mean(x) = 9

Variance(x) = 11

Correlation(x, y) = 0.816

Regression: y = 3 + 0.5x

Data Sample 3:

Mean(x) = 9

Variance(x) = 11

Correlation(x, y) = 0.816

Regression: y = 3 + 0.5x

Data Sample 4:

Mean(x) = 9

Variance(x) = 11

Correlation(x, y) = 0.816

Regression: y = 3 + 0.5x

Anscombe, American Statistician, 1973

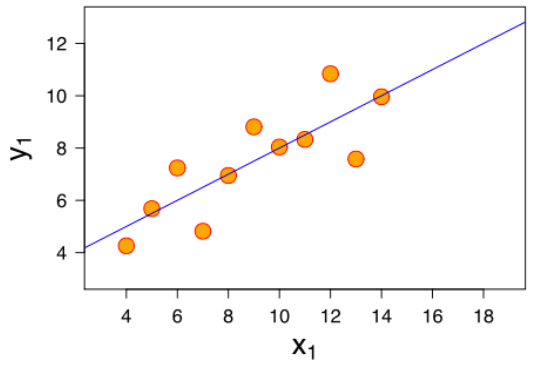
Data Sample 1:

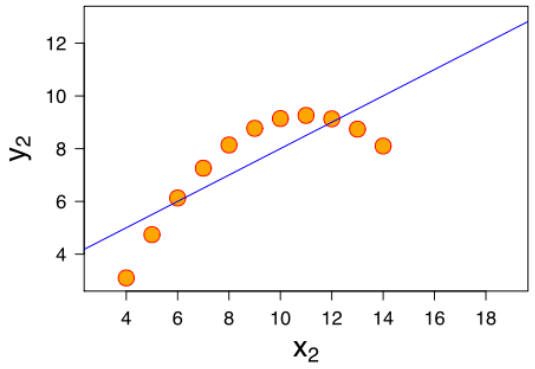
Mean(x) = 9

Variance(x) = 11

Correlation(x, y) = 0.816

Regression: y = 3 + 0.5x





Data Sample 2:

Mean(x) = 9

Variance(x) = 11

Correlation(x, y) = 0.816

Regression: y = 3 + 0.5x

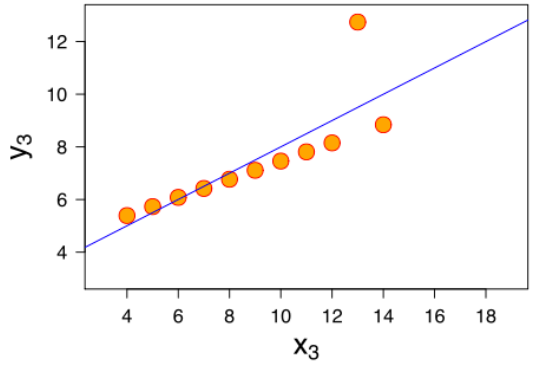
Data Sample 3:

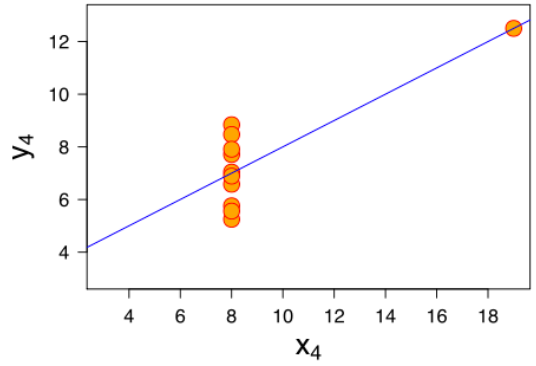
Mean(x) = 9

Variance(x) = 11

Correlation(x, y) = 0.816

Regression: y = 3 + 0.5x





Data Sample 4:

Mean(x) = 9

Variance(x) = 11

Correlation(x, y) = 0.816

Regression: y = 3 + 0.5x

Statistical tools are powerful,

but people understand patterns

How we represent data changes the questions we can answer

These questions shift as the available data grows

We can manage scales using two strategies:

1. **Harness Human Vision**
2. **Collaborate with Computation**

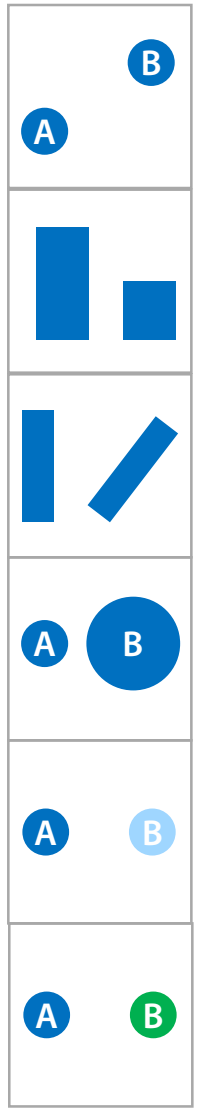
How we represent data changes the questions we can answer

These questions shift as the available data grows

We can manage scales using two strategies:

1. Harness Human Vision
2. Collaborate with Computation

Cleveland & McGill, 1984



Position

Length

Orientation

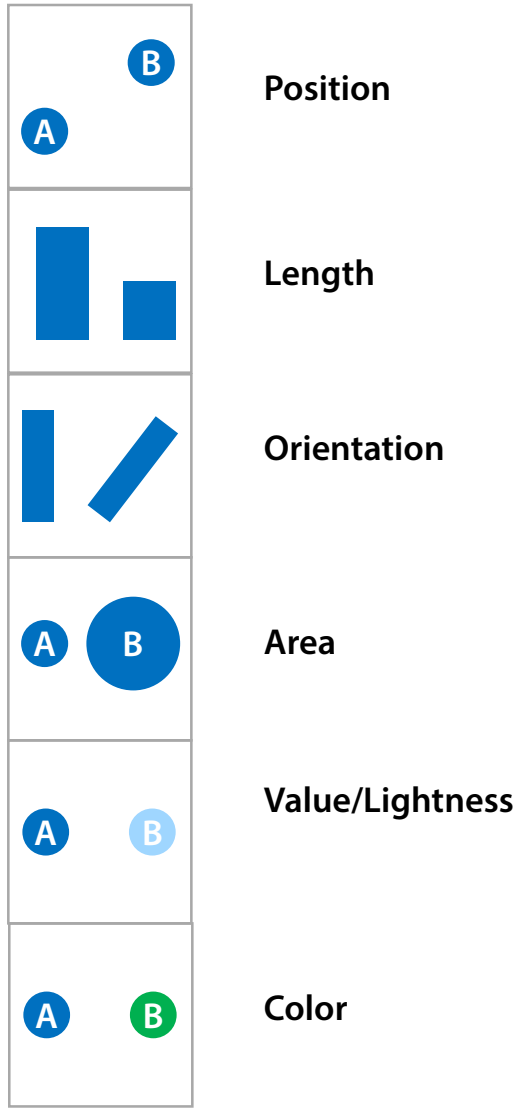
Area

Value/Lightness

Color

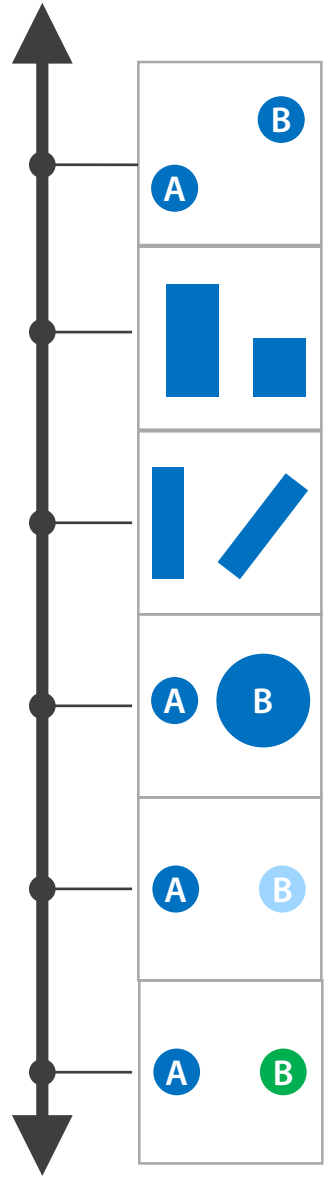
Cleveland & McGill, 1984

How much bigger is A than B?



Cleveland & McGill, 1984

More Effective



Position

Length

Orientation

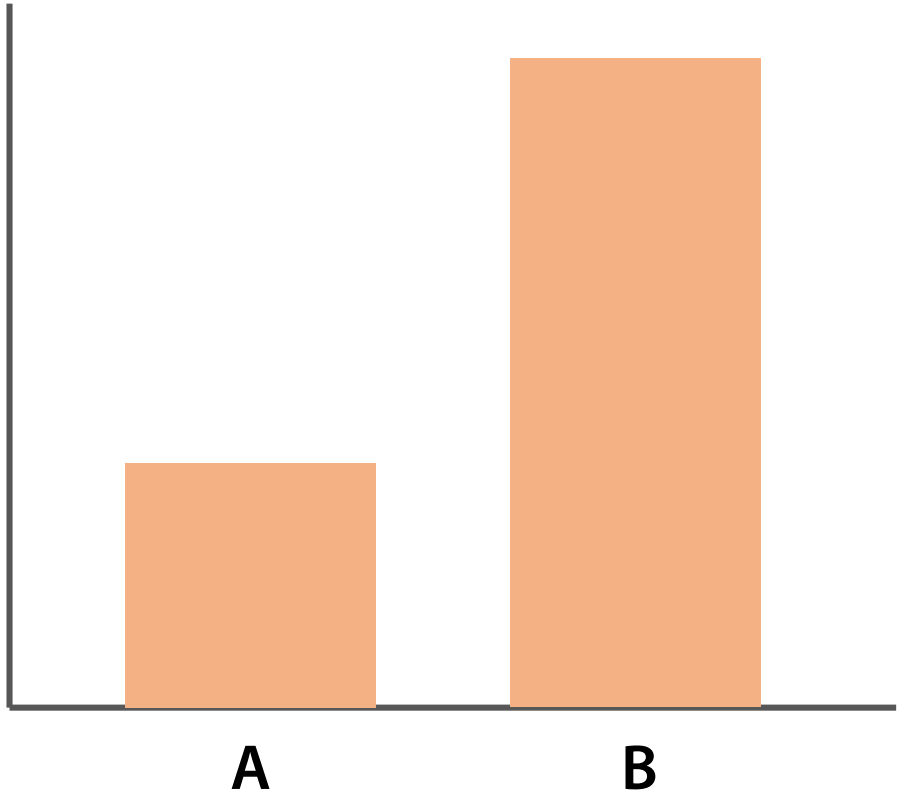
Area

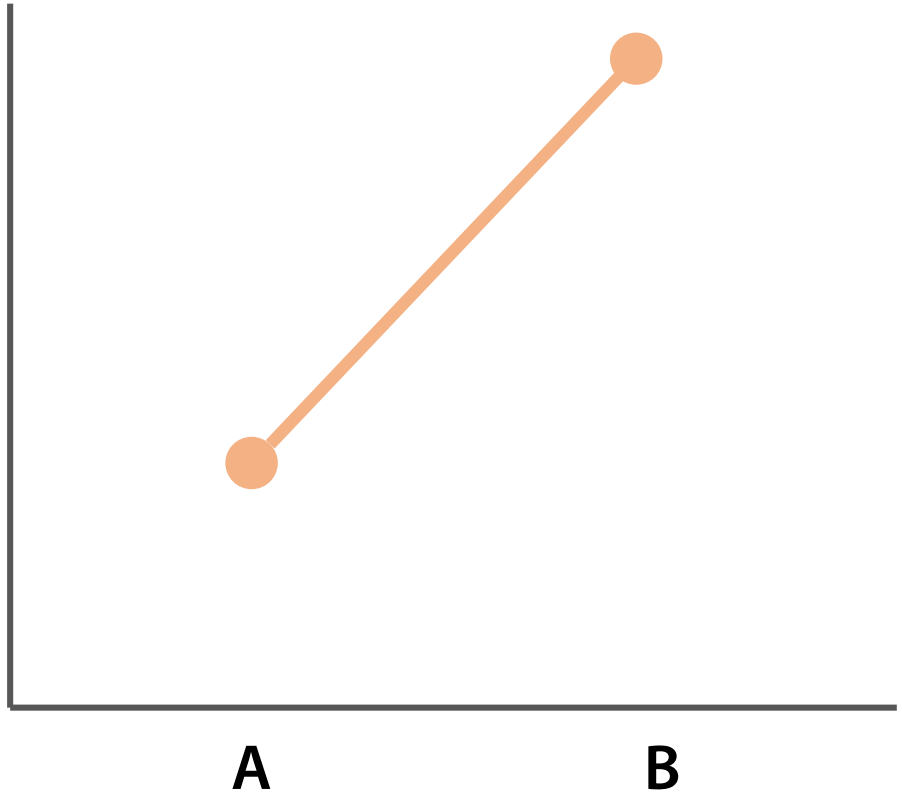
Value/Lightness

Color

Less Effective

Zacks & Tversky, 1999





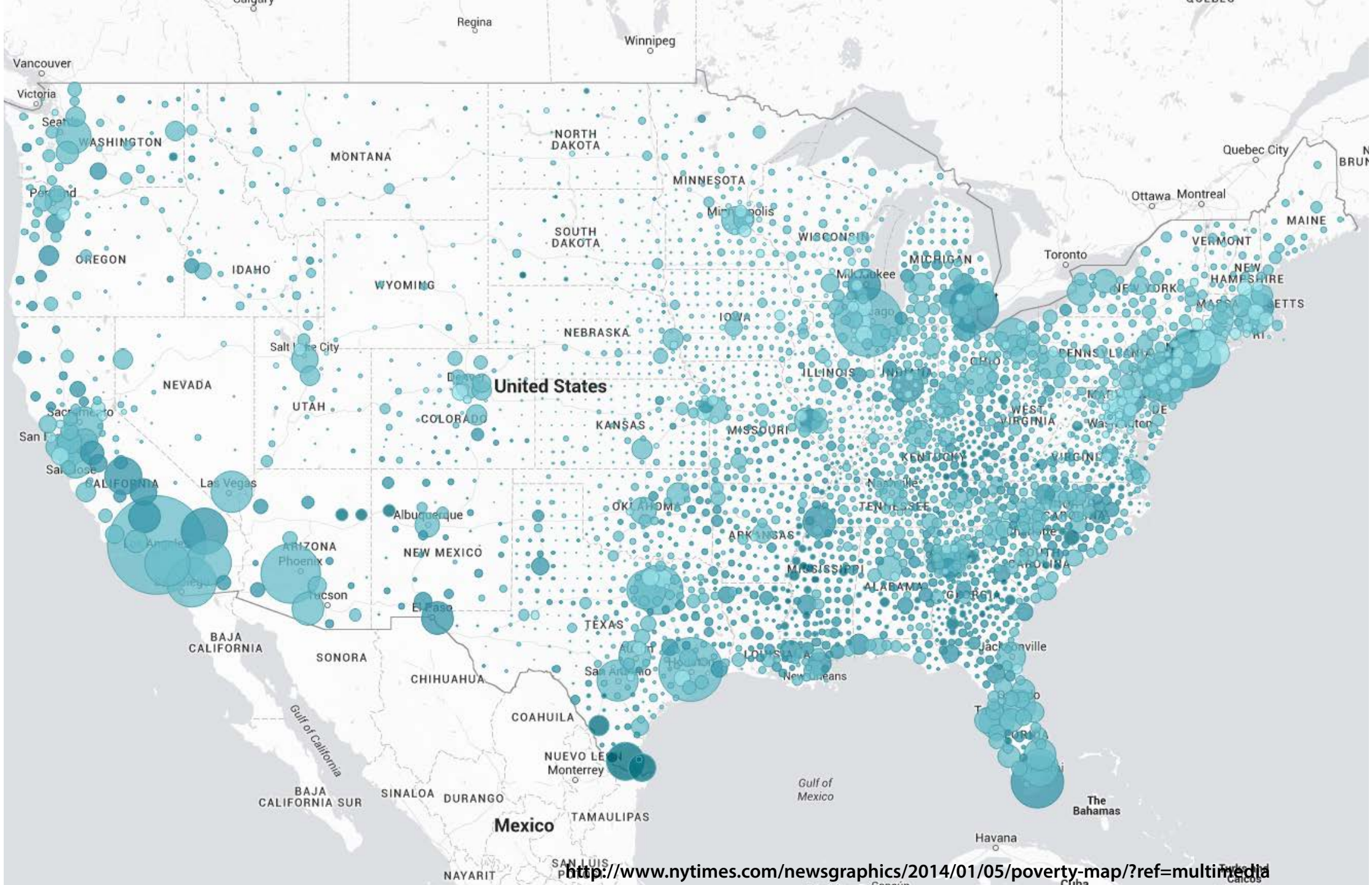
How we represent data affects the patterns people see

How we represent data changes the questions we can answer

These questions shift as the available data grows

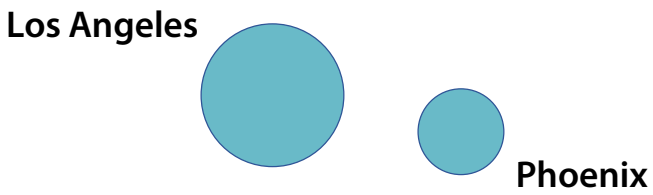
We can manage scales using two strategies:

1. Harness Human Vision
2. Collaborate with Computation



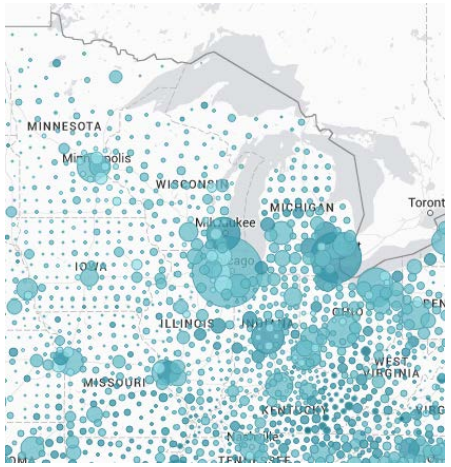
http://www.nytimes.com/newsgraphics/2014/01/05/poverty-map/?ref=multimedia

Low-Level Tasks → Individual Values

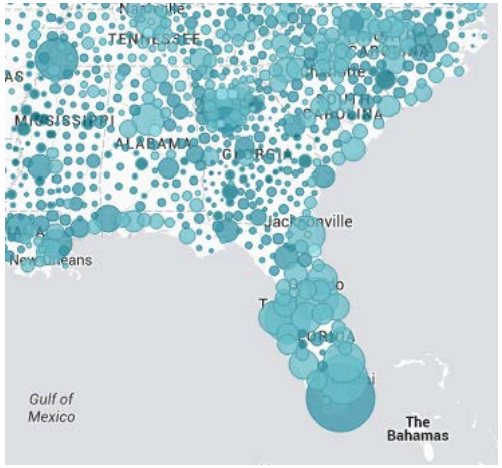


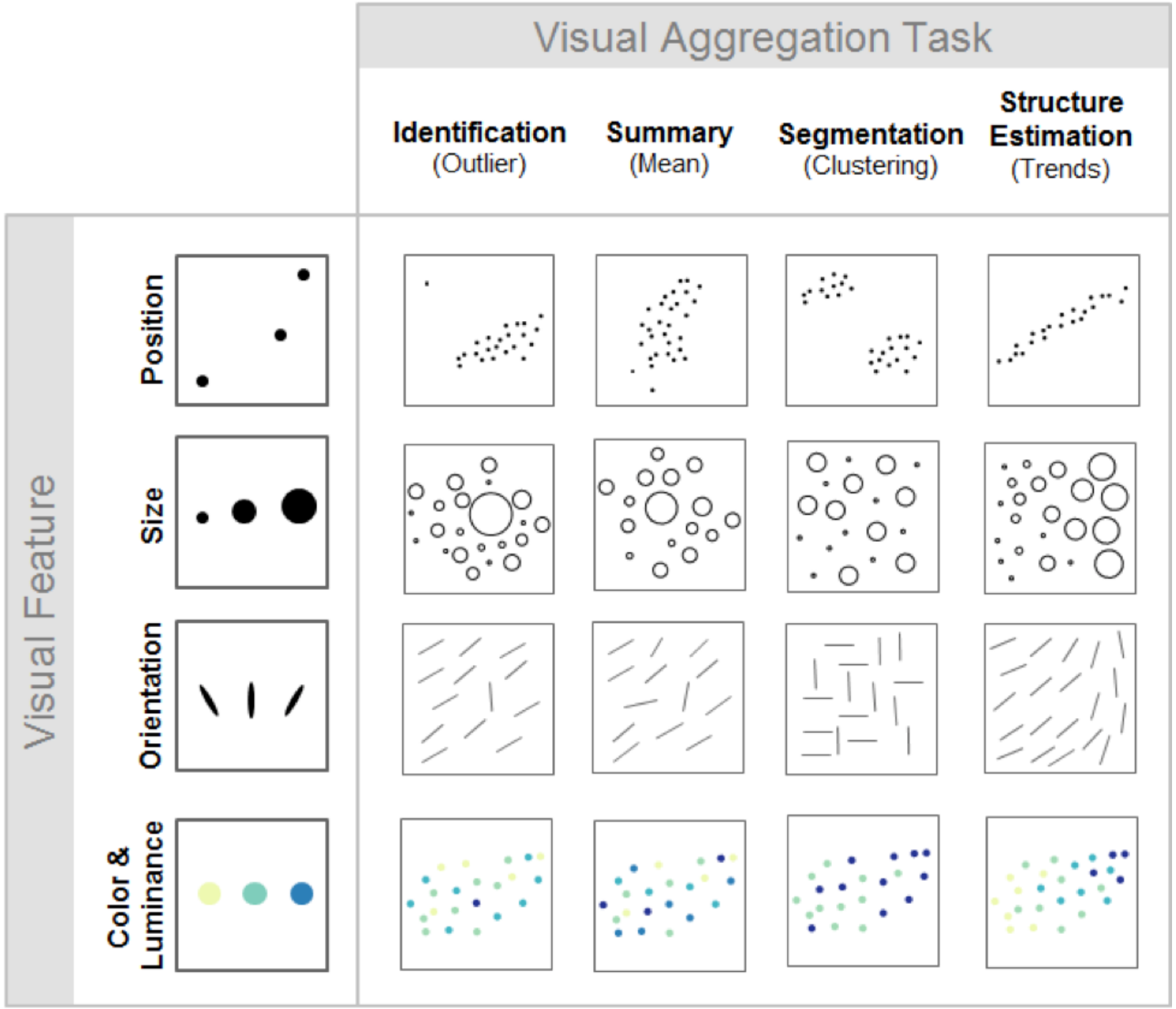
High-Level Tasks → Combine Many Values

Midwest



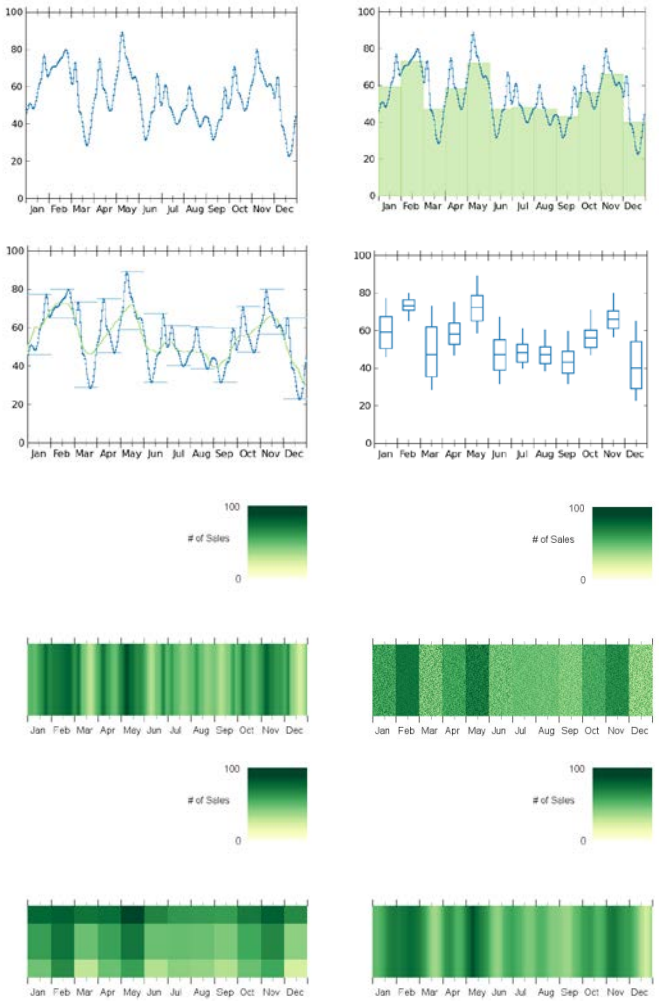
Southeast





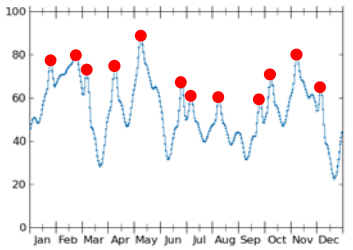
Szafir, Haroz, Gleicher & Franconeri, 2016

Encodings

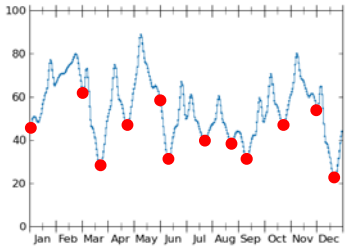


X

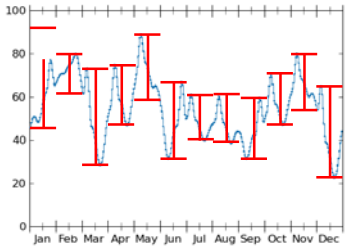
Tasks



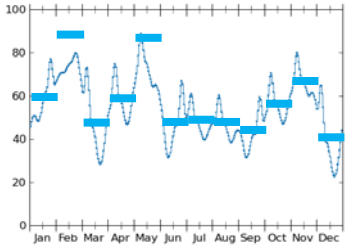
Maxima



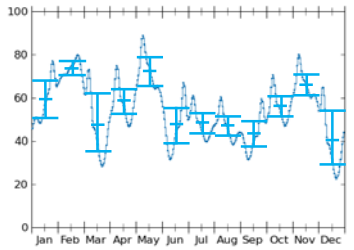
Minima



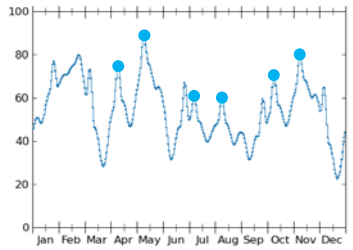
Range



Average

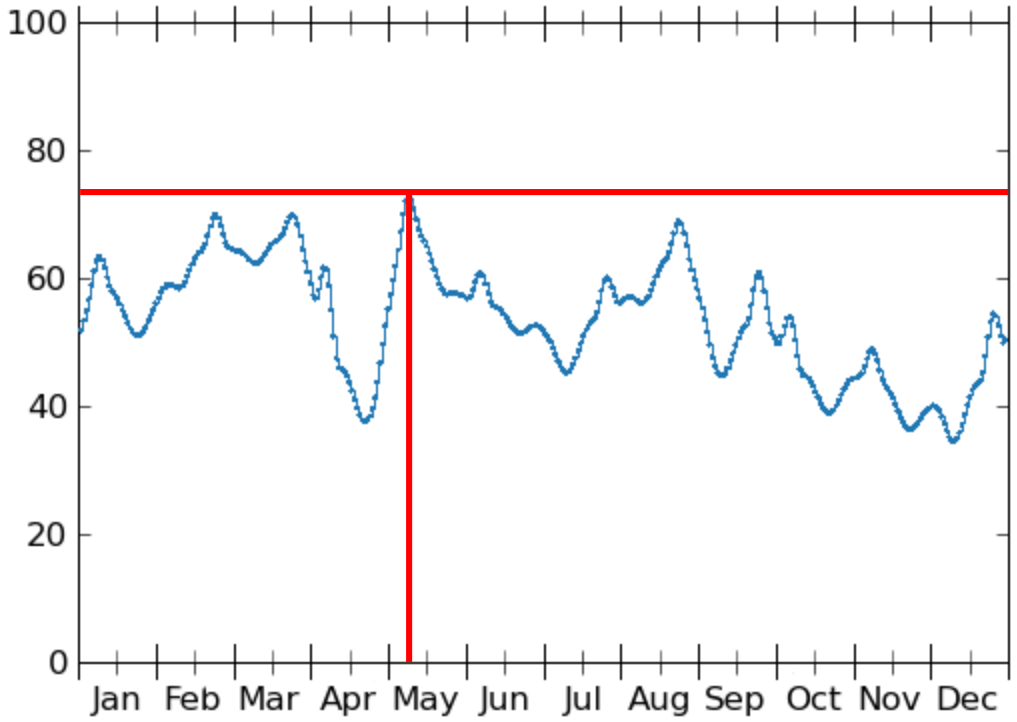


Variance



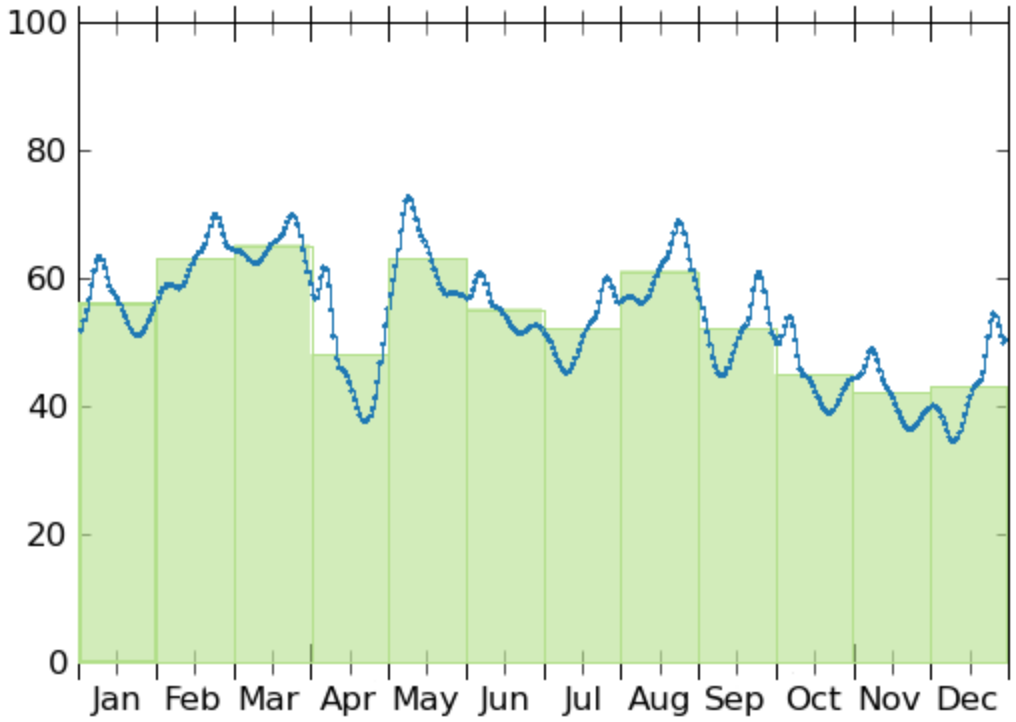
Outliers

Albers, Correll, Franconeri & Gleicher, 2014



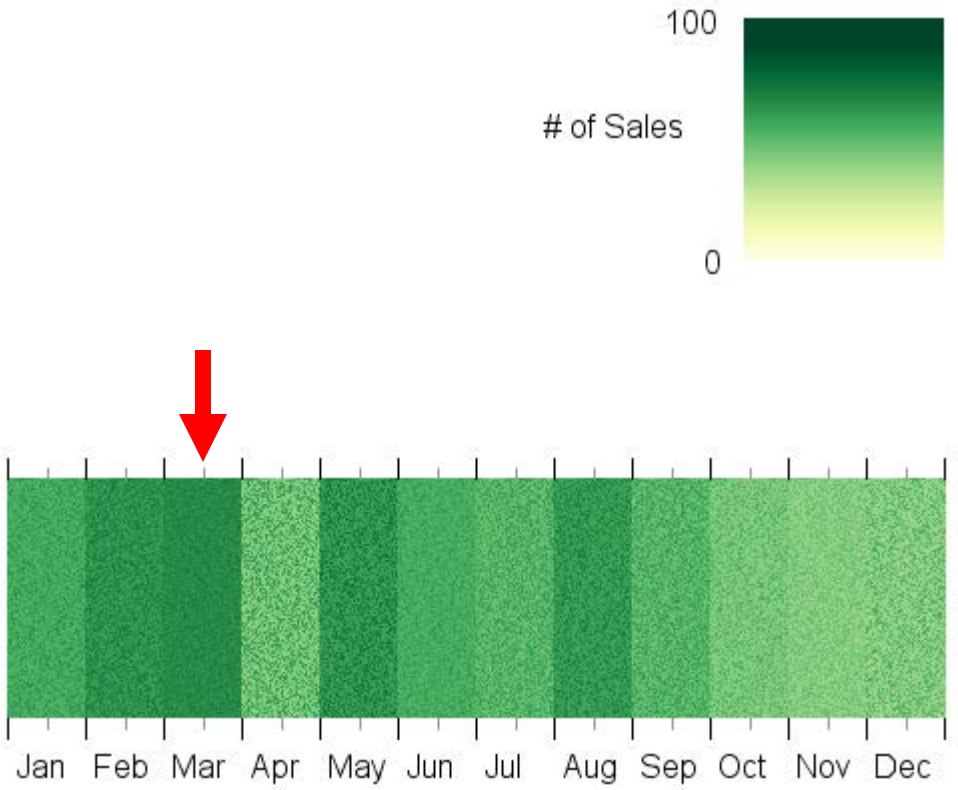
What month has the highest sales day?

Albers, Correll, Franconeri & Gleicher, 2014



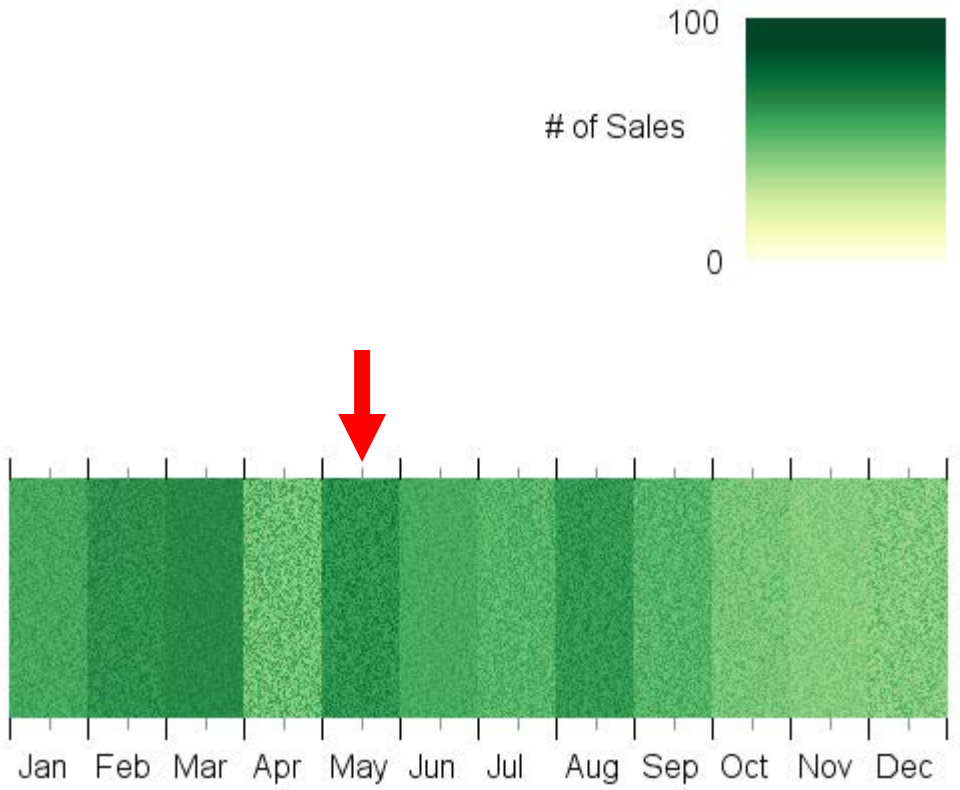
What month has the highest sales on average?

Albers, Correll, Franconeri & Gleicher, 2014



What month has the highest sales on average?

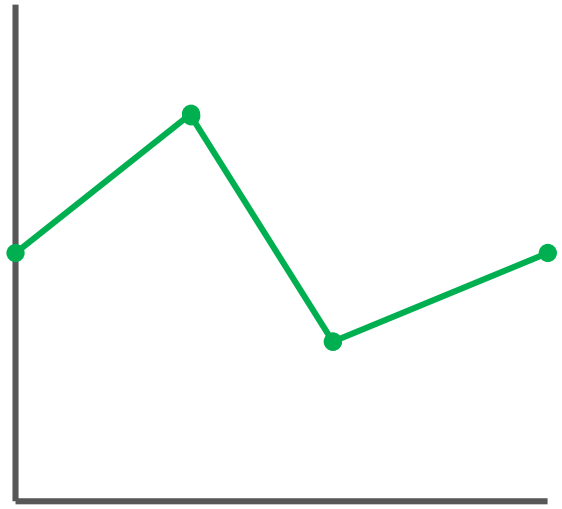
Albers, Correll, Franconeri & Gleicher, 2014



What month has the highest sales day?

Albers, Correll, Franconeri & Gleicher, 2014

Albers, Correll, Franconeri & Gleicher, 2014



Position for
Point Tasks

Color for
Summary Tasks

How you map the data impacts what statistics people see

How we represent data changes the questions we can answer

These questions shift as the available data grows

We can manage scales using two strategies:

1. **Harness Human Vision**
2. **Collaborate with Computation**

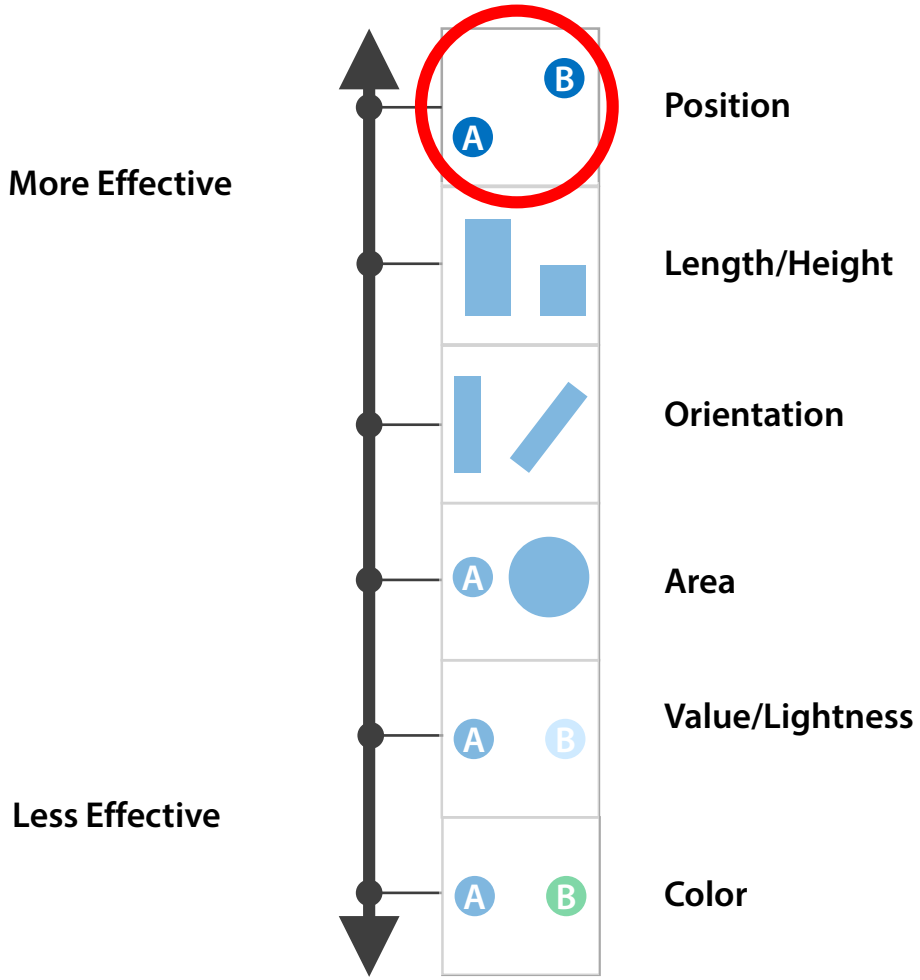
All the world's a stage,
And all the men and women merely players:
They have their exits and their entrances;
And one man in his time plays many parts,
His acts being seven ages. At first the infant,
Mewling and puking in the nurse's arms.
And then the whining school-boy, with his satchel
And shining morning face, creeping like snail
Unwillingly to school. And then the lover,
Sighing like furnace, with a woeful ballad
Made to his mistress' eyebrow. Then a soldier,
Full of strange oaths and bearded like the pard,
Jealous in honour, sudden and quick in quarrel,
Seeking the bubble reputation
Even in the cannon's mouth.

Important Texts

She: A History of Adventure: 15 chapters

Large Digitized Collections

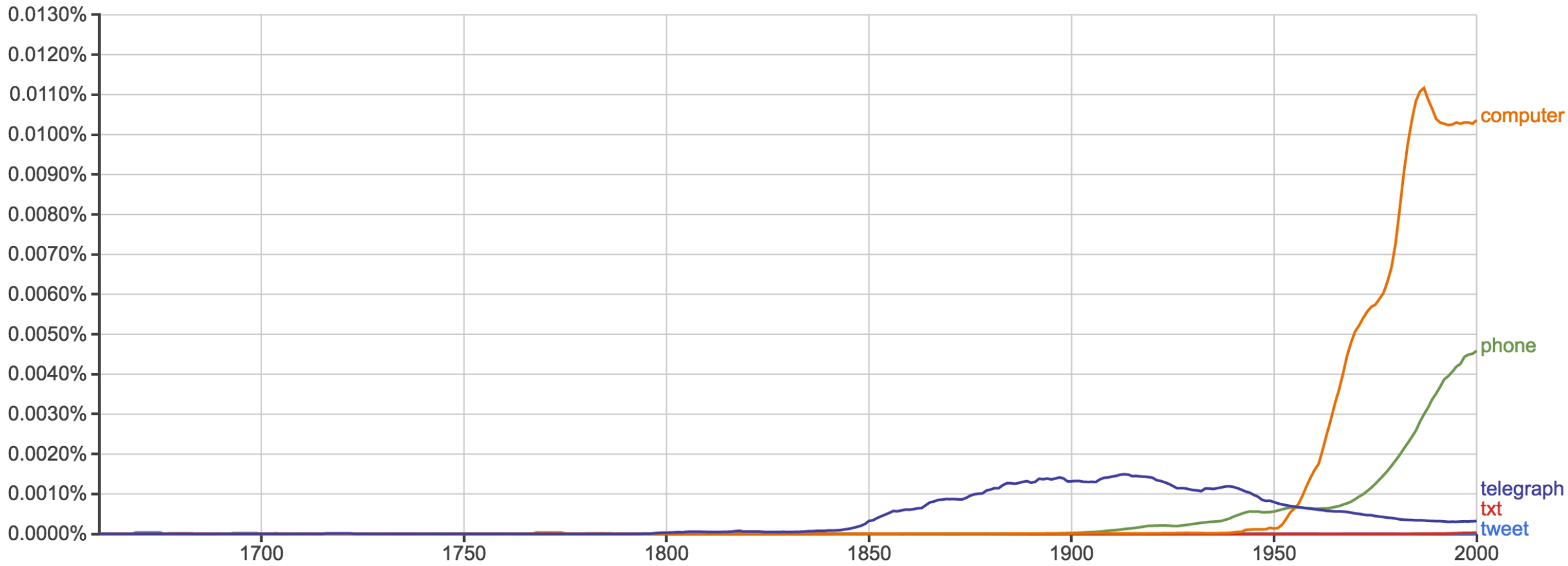
Google N-Grams: 5,195,769 books

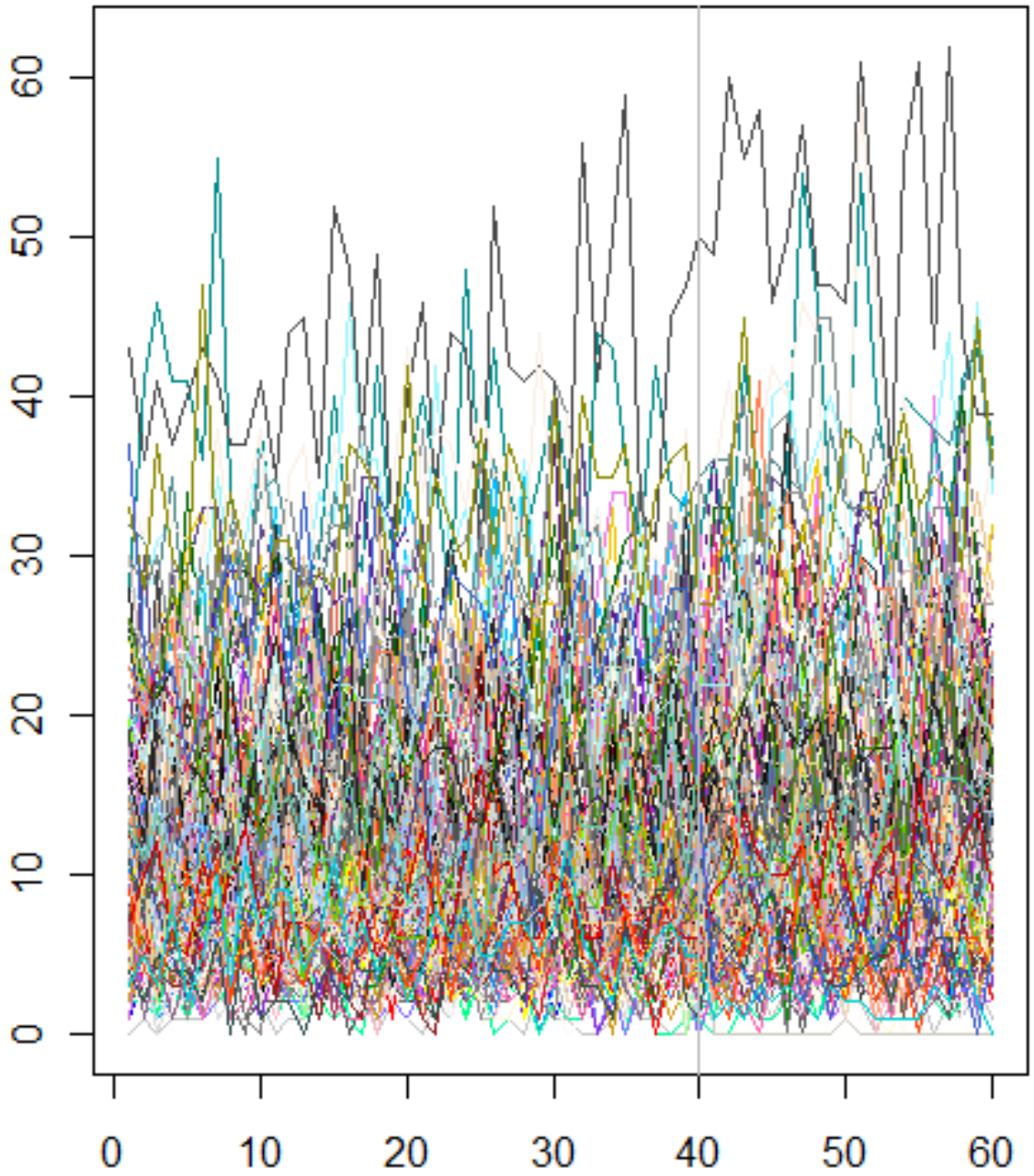

More Effective
Less Effective
A
B
Position
Length/Height
Orientation
A
Area
A
B
Value/Lightness
A
B
Color

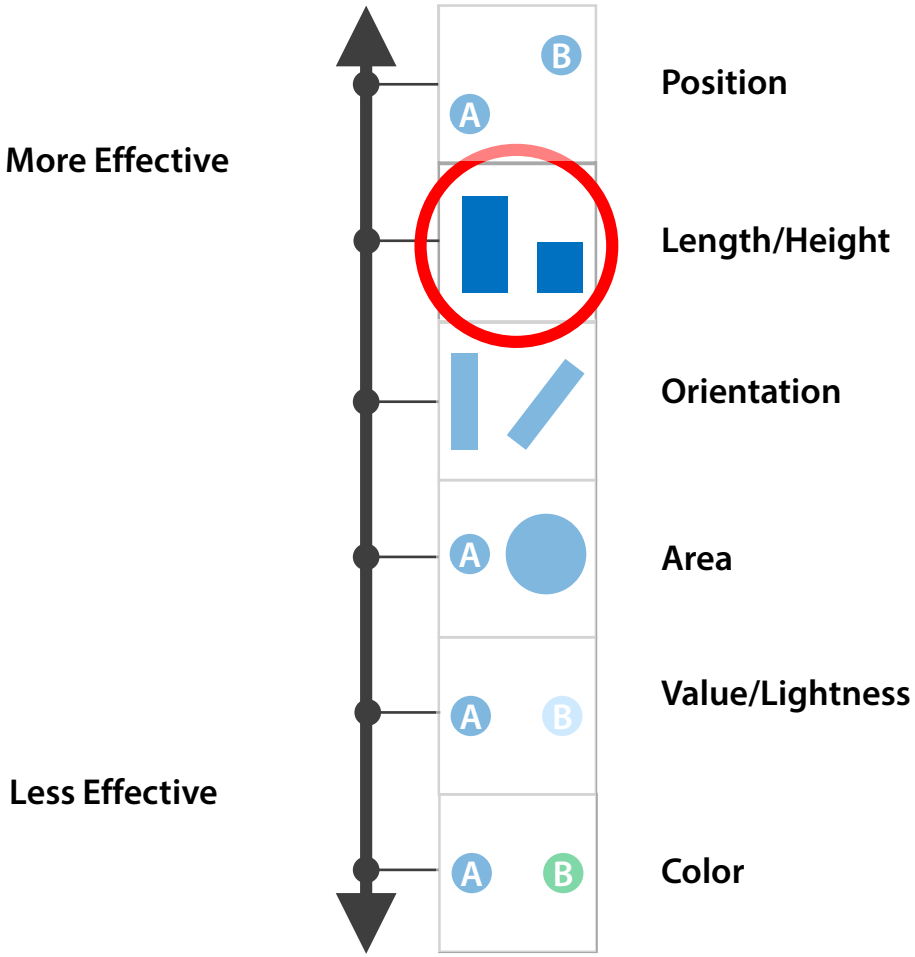
Google Books Ngram Viewer

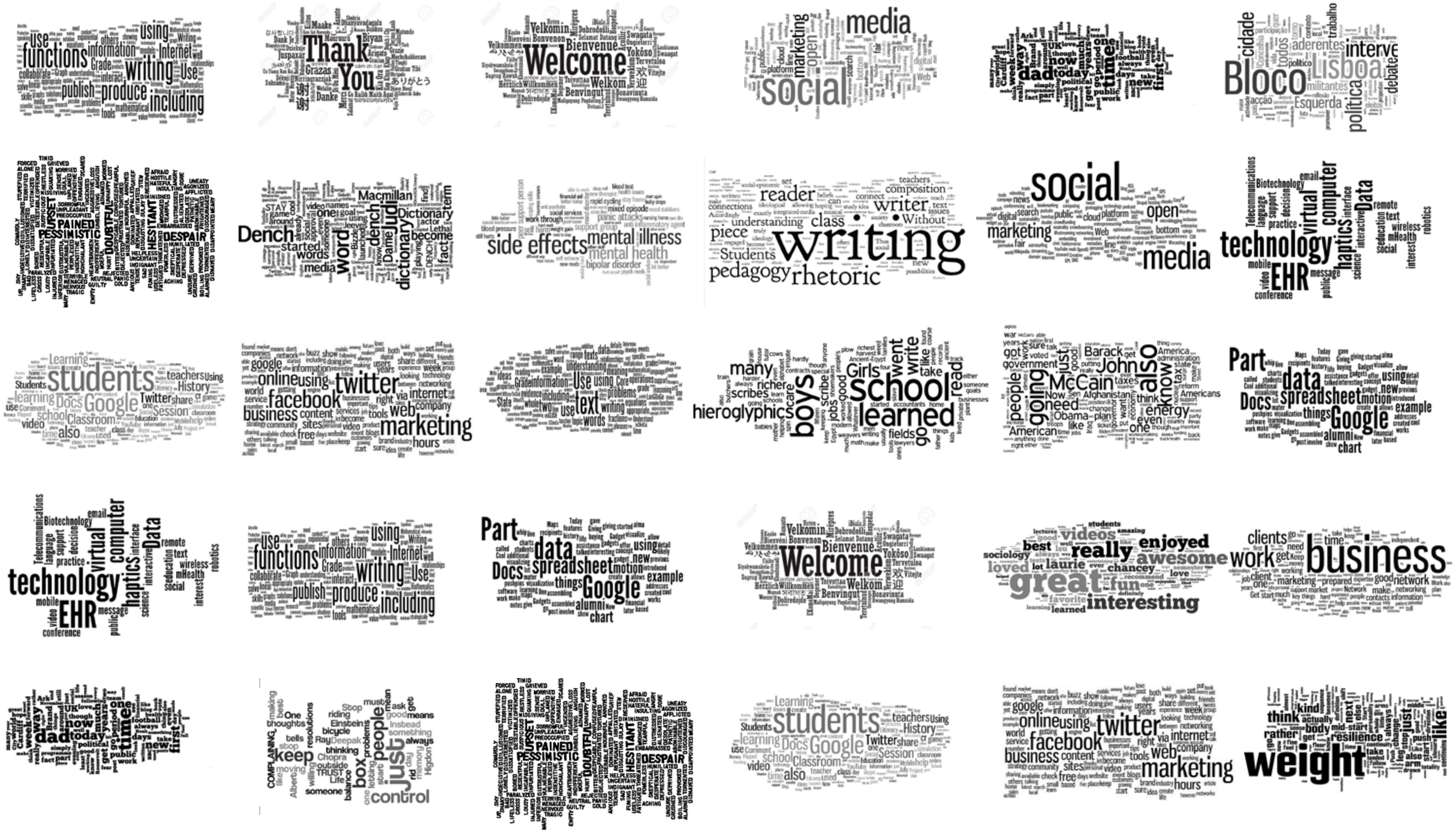
Graph these comma-separated phrases: tweet,txt,phone,computer,telegraph case-insensitive

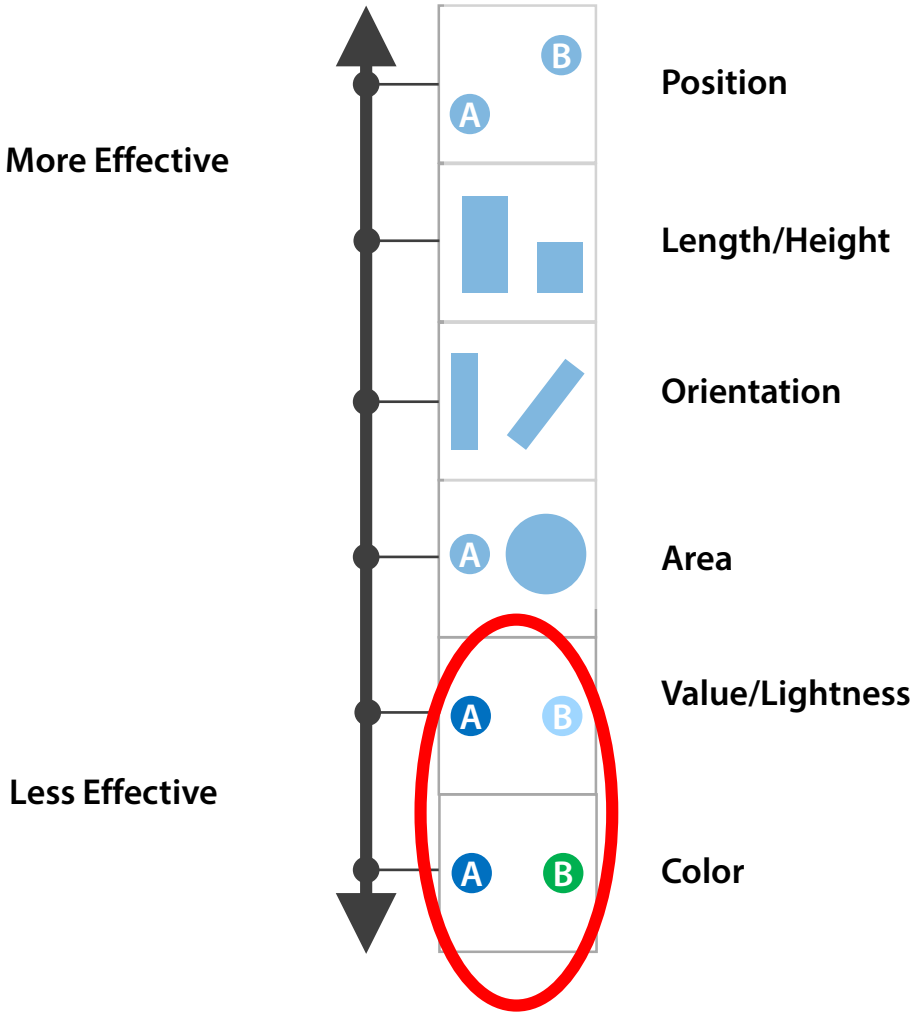
between 1660 and 2000 from the corpus English with smoothing of 3. Search lots of books




0
10
20
30
40
50
60
0
10
20
30
40
50
60


More Effective
Less Effective
B
A
Position
Length/Height
Orientation
A
Area
A
B
Value/Lightness
A
B
Color

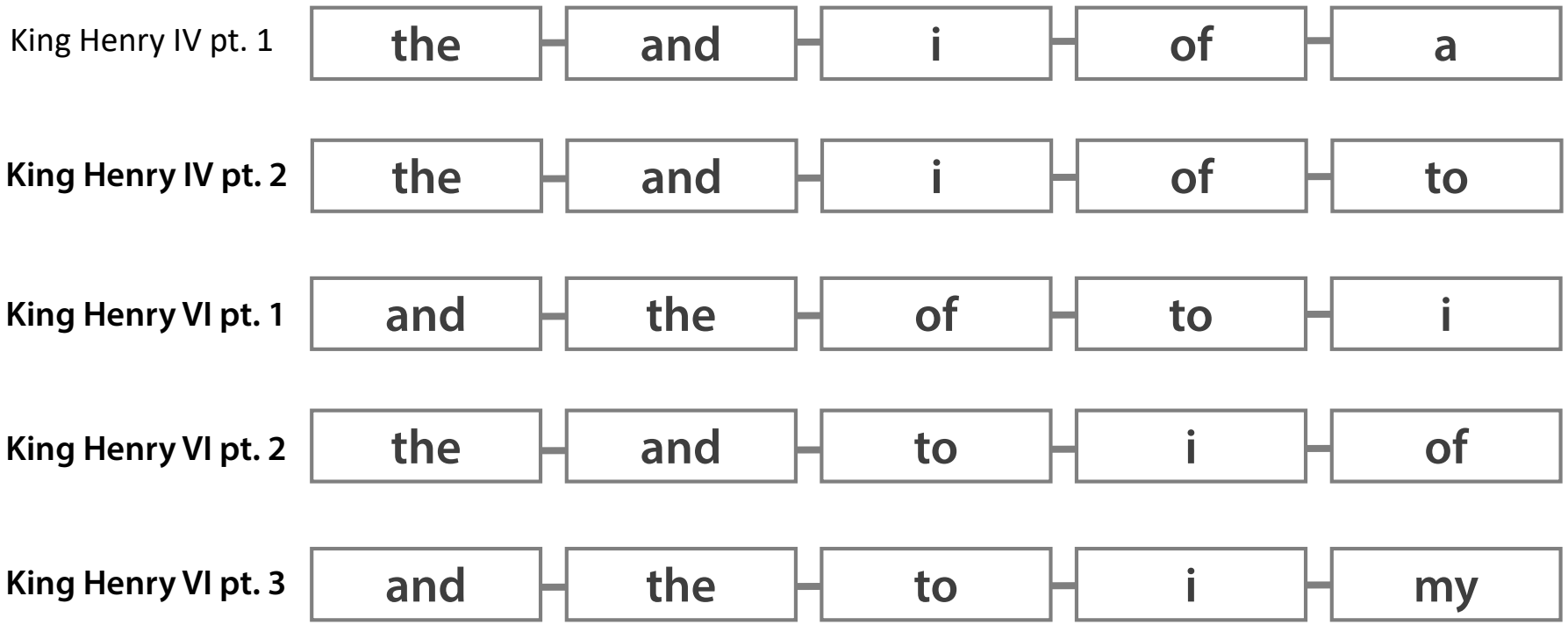

functions
writing
Thank You
Welcome
social
media
Bloco
Lisboa
PESSIMISTIC
DESPAIR
Dench
dictionary
side effects
mental illness
writing
rhetoric
pedagogy
social
media
technology
EHR
students
Google
twitter
facebook
marketing
school
learned
McCain
spreadsheet
Google
technology
functions
writing
Google
spreadsheet
Welcome
great
awesome
business
dad
time
control
just
PESSIMISTIC
students
twitter
marketing
weight


More Effective
Less Effective
B
A
Position
Length/Height
Orientation
A
Area
A
B
Value/Lightness
A
B
Color

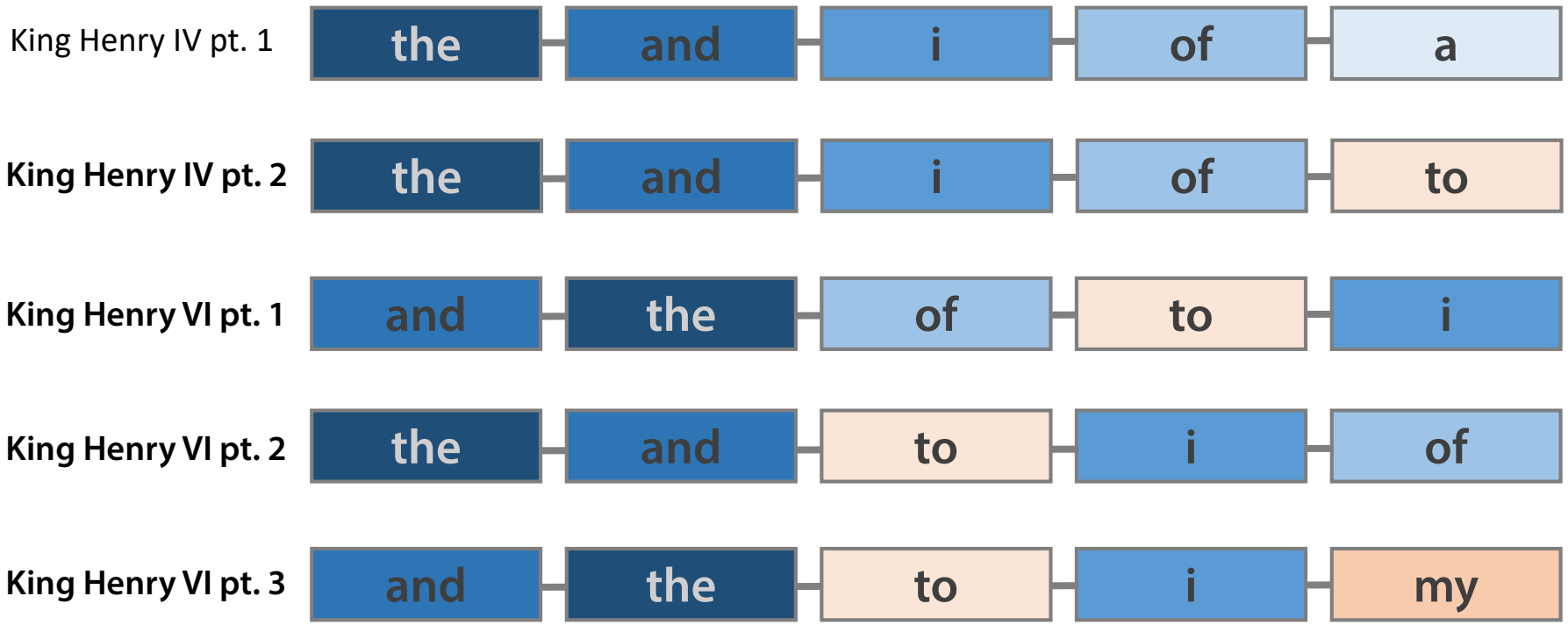
Turning texts into sequences

```
All the world's a stage,
And all the men and women merely players:
They have their exits and their entrances,
```

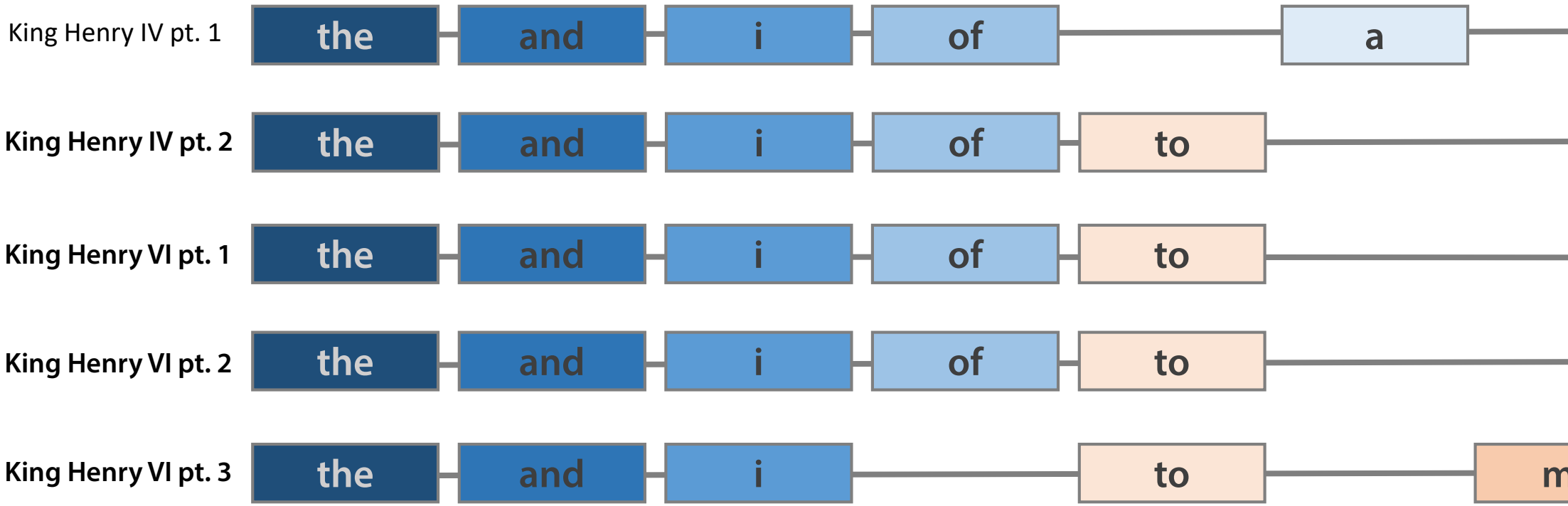
```
all the world a stage
and all the men and women merely players
they have their exits and their entrances
```



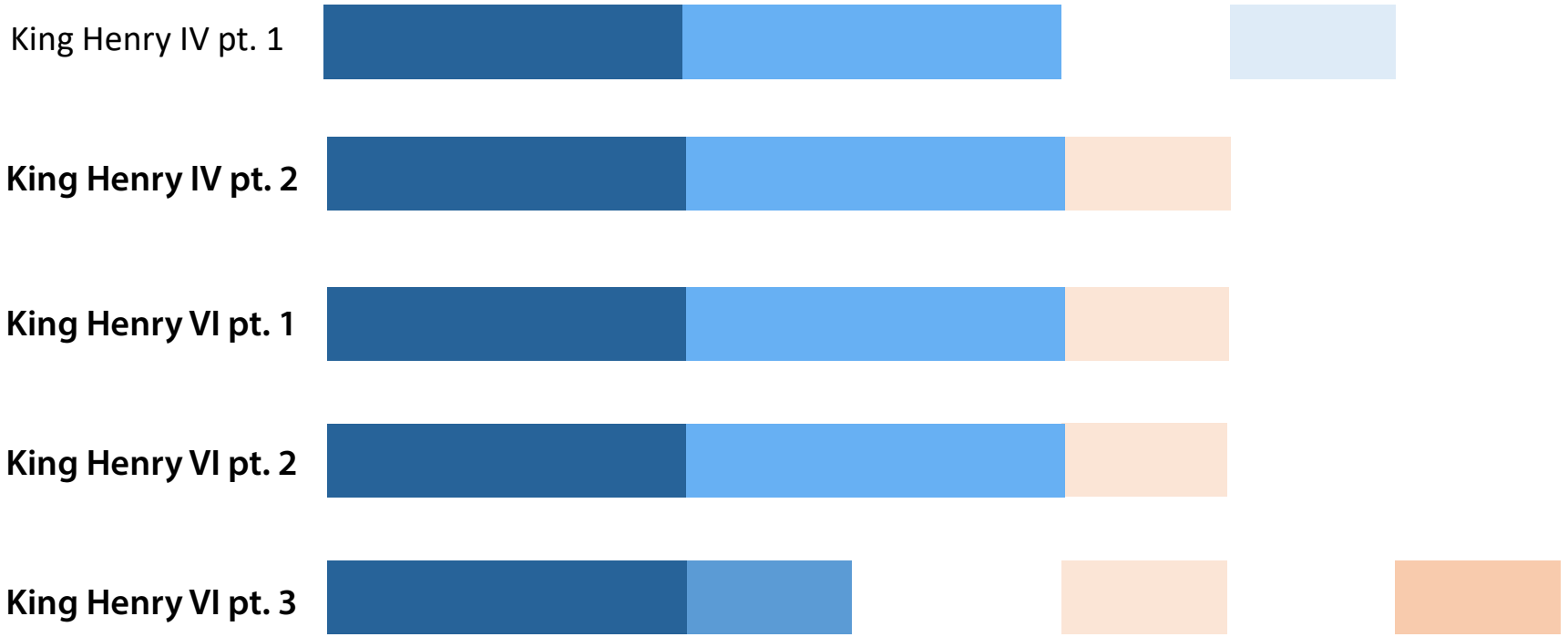
Szafir, Stuffer, Sohail, & Gleicher 2016



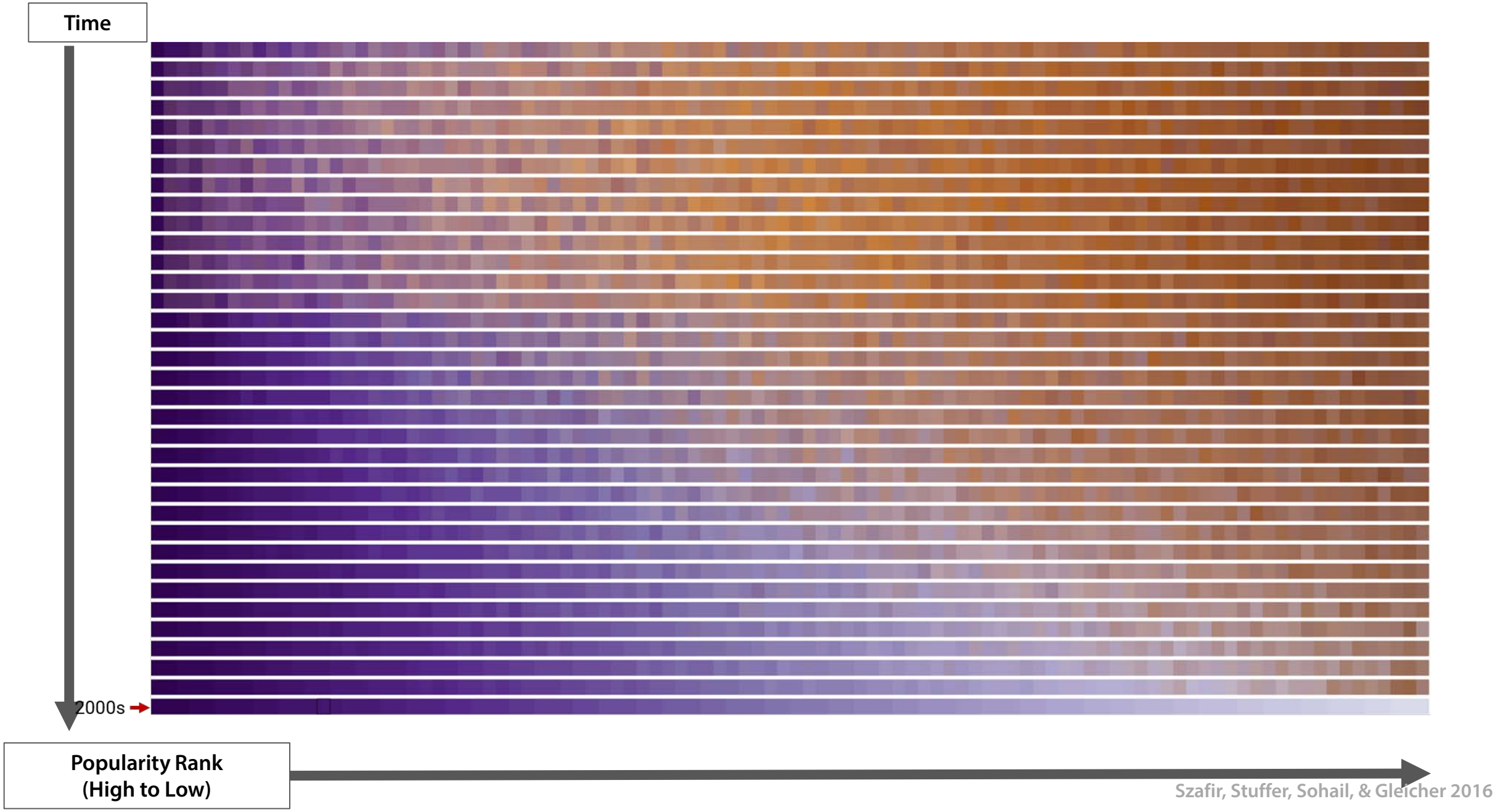
Szafir, Stuffer, Sohail, & Gleicher 2016



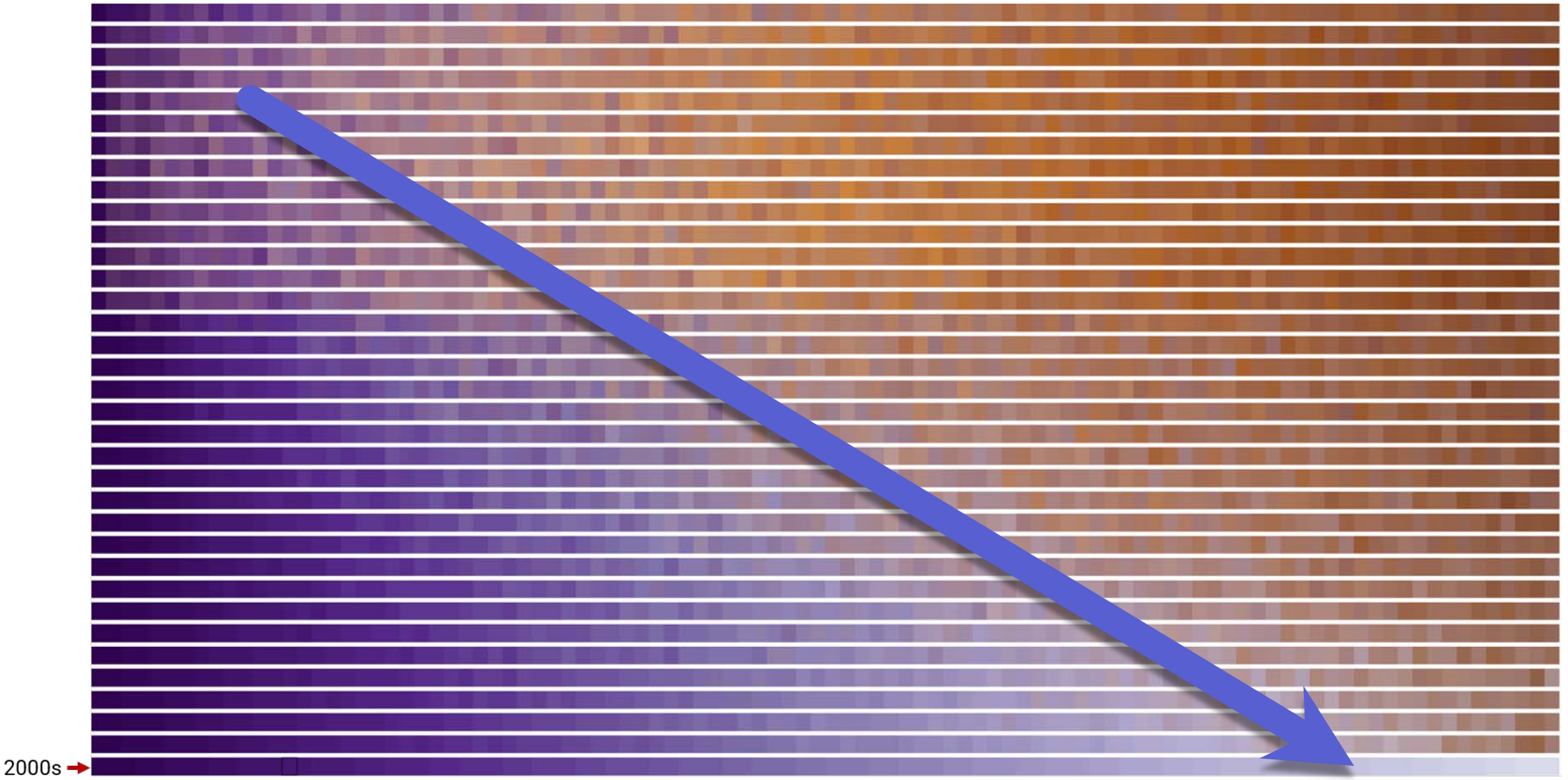
Szafir, Stuffer, Sohail, & Gleicher 2016



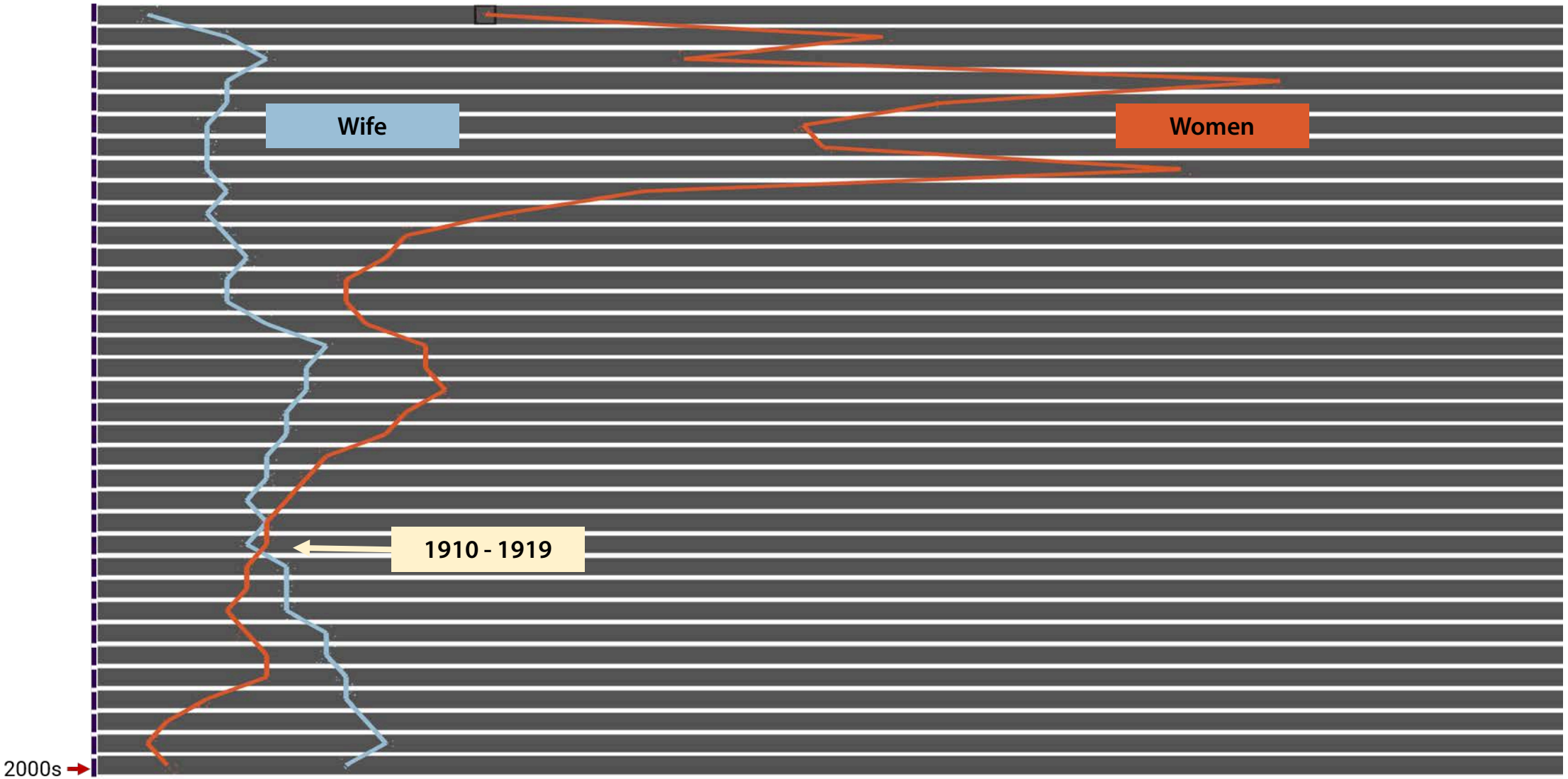
Szafir, Stuffer, Sohail, & Gleicher 2016



Szafir, Stuffer, Sohail, & Gleicher 2016



Szafir, Stuffer, Sohail, & Gleicher 2016



Szafir, Stuffer, Sohail, & Gleicher 2016

Scales matter because

Scales change questions we ask

Scales change representations that work

Scales require multiple perspectives on data

How we represent data changes the questions we can answer

These questions shift as the available data grows

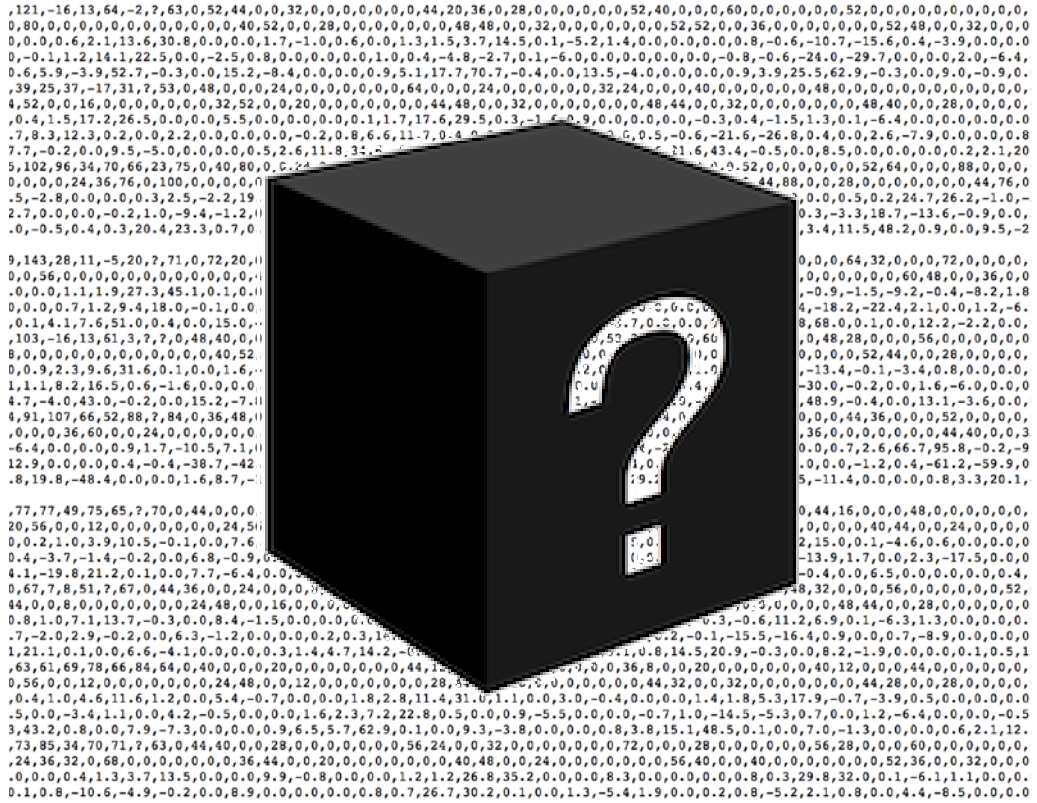
We can manage scales using two strategies:

1. **Harness Human Vision**
2. **Collaborate with Computation**

Statistical methods scale well, but their processes are often opaque

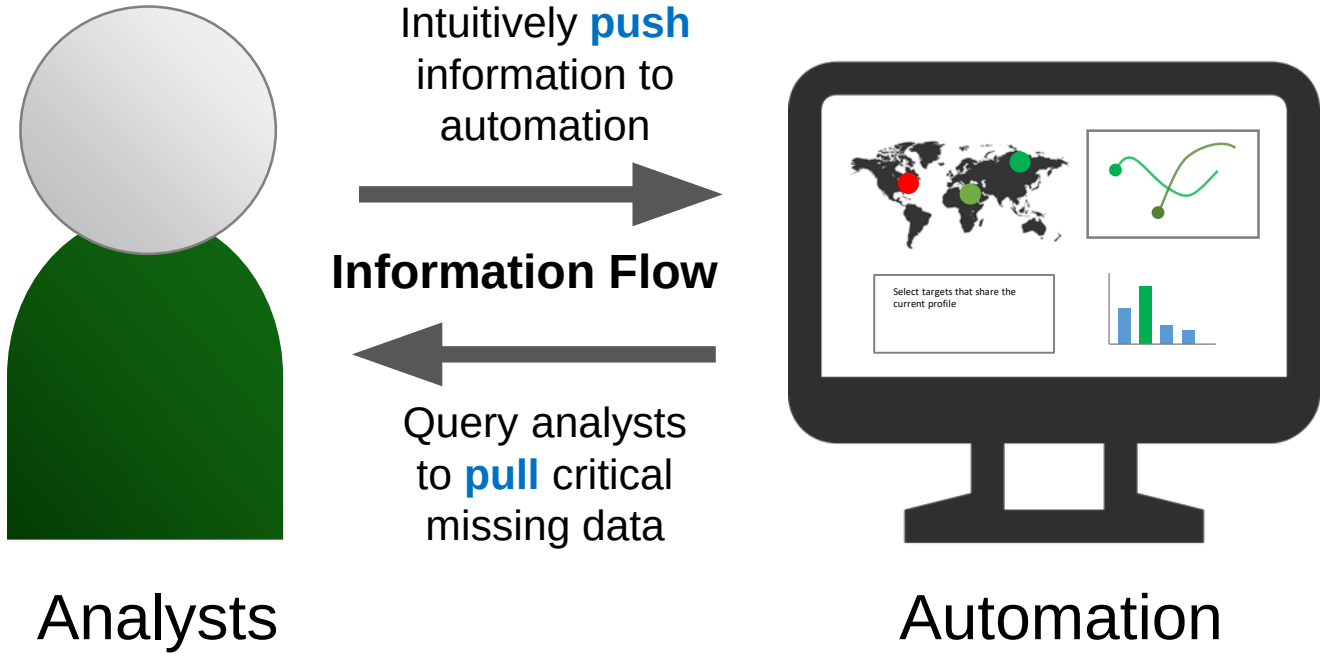
People bring context and expertise, but are slow

How can we combine statistical scalability with contexts from domain expertise?



Develop new data fusion and machine learning methods for collaborative human-machine perception in remote sensing

Apply methods from unmanned autonomous vehicles domain to SBIRS/OPIR domain



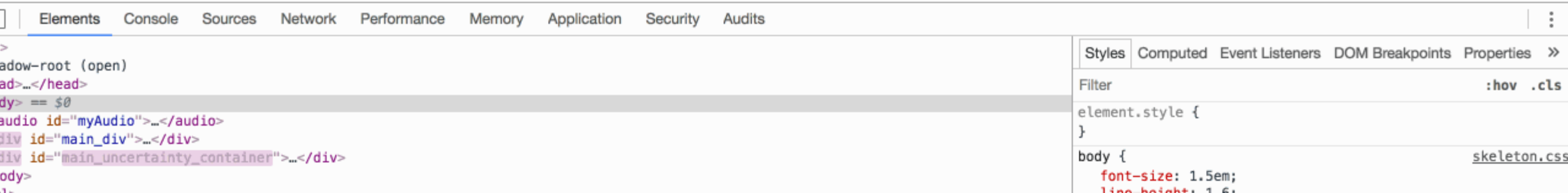
CAMP-SMC

File Edit View Help

For 10-15 second frame updates,

human + automation can accurately

classify targets ~3 minutes earlier



Open Questions

How do we quantify cognitive and perceptual elements of data analysis? How do we make that data actionable?

What do we do with imperfect data?

What factors of models allow people to collaboratively interact with automated models?

Open Questions

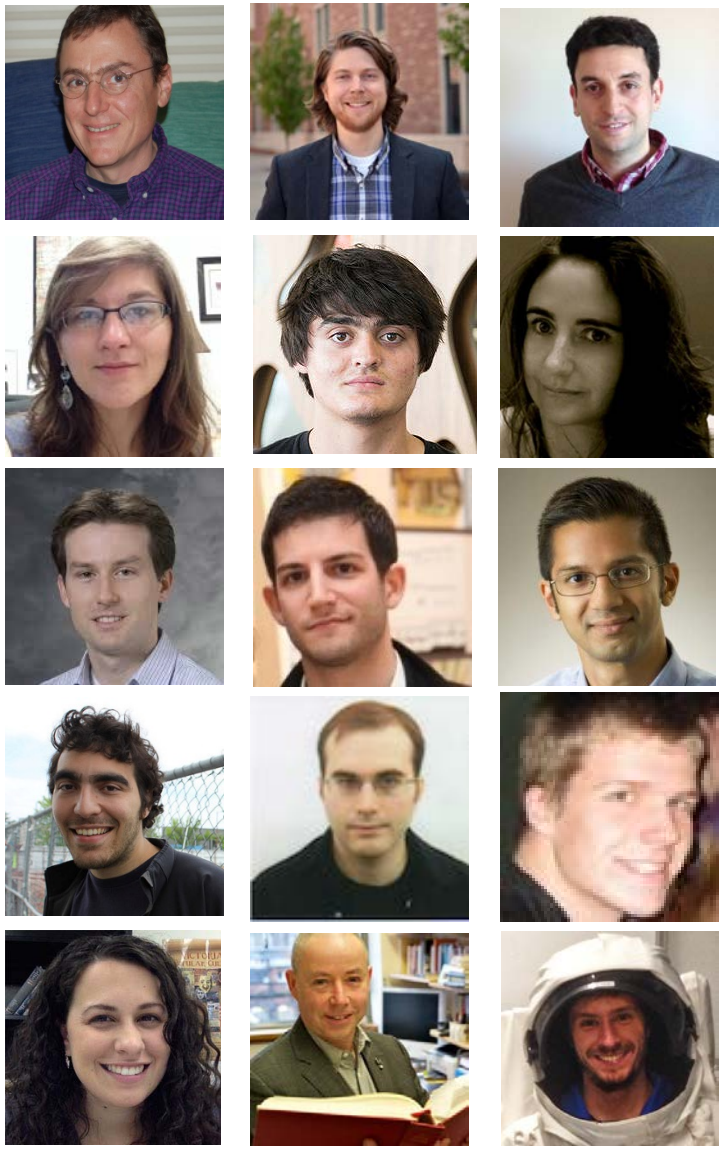
How do we quantify cognitive and perceptual elements of data analysis? How do we make that data actionable?

What do we do with imperfect data?

What factors of models allow people to collaboratively interact with automated models?

Thank You!

VisualLab



Danielle Albers Szafir
danielle.szafir@colorado.edu
@dalbersszafir

Funding:
National Science Foundation
US Air Force SMC
Mellon Foundation

Demos & Papers:
http://danielleszafir.com
http://cmci.colorado.edu/visualab