Visualization & Perception Across Scales

Danielle Albers Szafir
Department of Information Science
danielle.szafir@colorado.edu
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
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
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
<table>
<thead>
<tr>
<th>Position</th>
<th>Area</th>
<th>Length</th>
<th>Value/Lightness</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value/Lightness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How much bigger is A than B?
More Effective

Position

Length

Orientation

Area

Value/Lightness

Color

Less Effective

Cleveland & McGill, 1984
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
Low-Level Tasks $\rightarrow$ Individual Values

Los Angeles

Phoenix

High-Level Tasks $\rightarrow$ Combine Many Values

Midwest

Southeast
## Visual Aggregation Task

<table>
<thead>
<tr>
<th>Visual Feature</th>
<th>Identification (Outlier)</th>
<th>Summary (Mean)</th>
<th>Segmentation (Clustering)</th>
<th>Structure Estimation (Trends)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
<td><img src="image4" alt="Diagram" /></td>
</tr>
<tr>
<td>Size</td>
<td><img src="image5" alt="Diagram" /></td>
<td><img src="image6" alt="Diagram" /></td>
<td><img src="image7" alt="Diagram" /></td>
<td><img src="image8" alt="Diagram" /></td>
</tr>
<tr>
<td>Orientation</td>
<td><img src="image9" alt="Diagram" /></td>
<td><img src="image10" alt="Diagram" /></td>
<td><img src="image11" alt="Diagram" /></td>
<td><img src="image12" alt="Diagram" /></td>
</tr>
<tr>
<td>Color &amp; Luminance</td>
<td><img src="image13" alt="Diagram" /></td>
<td><img src="image14" alt="Diagram" /></td>
<td><img src="image15" alt="Diagram" /></td>
<td><img src="image16" alt="Diagram" /></td>
</tr>
</tbody>
</table>

Szafir, Haroz, Gleicher & Franconeri, 2016
Encodings

Tasks

Maxima

Minima

Range

Average

Variance

Outliers

Albers, Correll, Franconeri & Gleicher, 2014
What month has the **highest sales day**?
What month has the highest sales on average?
What month has the **highest sales on average**?

Albers, Correll, Franconeri & Gleicher, 2014
What month has the **highest sales day**?
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
- Position
- Length/Height
- Orientation
- Area
- Value/Lightness
- Color

More Effective vs. Less Effective
More Effective

Position
Length/Height
Orientation
Area
Value/Lightness
Color

Less Effective
Position
Length/Height
Orientation
Area
Value/Lightness
Color

More Effective
Less Effective
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
King Henry IV pt. 1: the and i of a
King Henry IV pt. 2: the and i of to
King Henry VI pt. 1: and the of to i
King Henry VI pt. 2: the and to i of
King Henry VI pt. 3: and the to i my

Szafir, Stuffer, Sohail, & Gleicher 2016
King Henry IV pt. 1
King Henry IV pt. 2
King Henry VI pt. 1
King Henry VI pt. 2
King Henry VI pt. 3

Szafir, Stuffer, Sohail, & Gleicher 2016
Women

1910 - 1919

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.

Intuitively push information to automation.

Query analysts to pull critical missing data.

Information Flow:
- Analysts
- Automation
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!

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