Paul Rosen

paul.rosen@utah.edu
@paulrosenphd
https://cspaul.com

www.sci.utah.edu

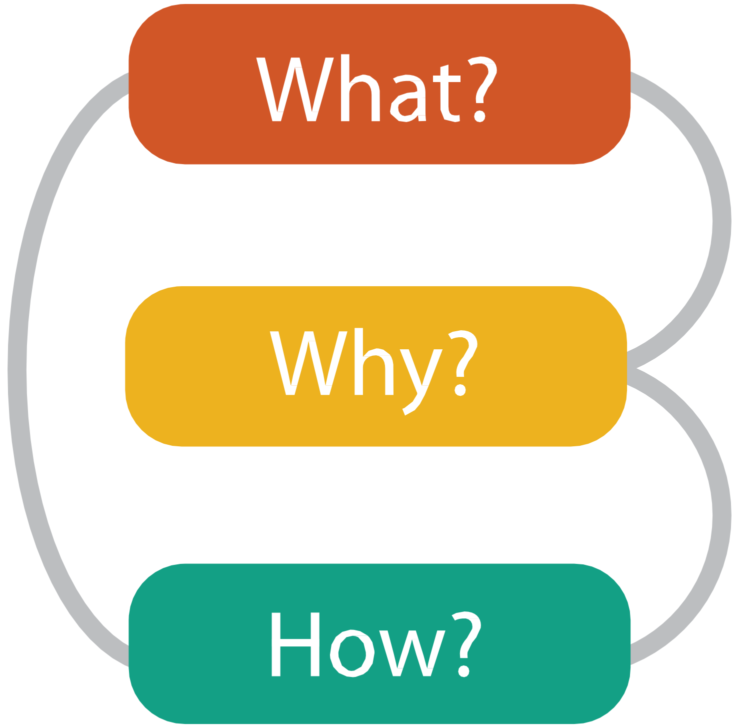
THE UNIVERSITY OF UTAH

Visualization for Data Science DS-4630 / CS-5630 / CS-6630

TASKS & INTERACTION

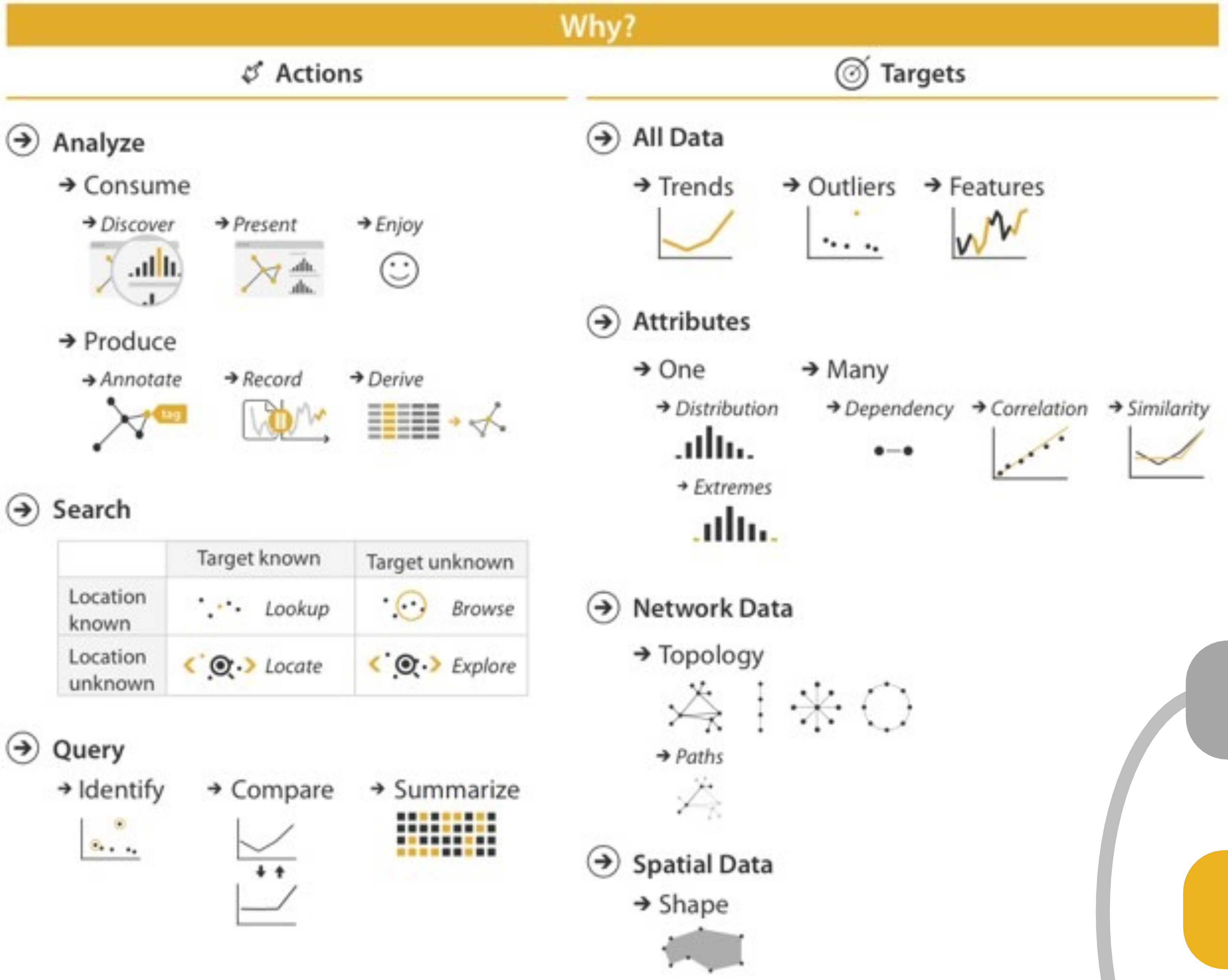
High-level Task Taxonomy: what, why, and how

- what is shown?
- why is the user looking at it?
- how is it shown?

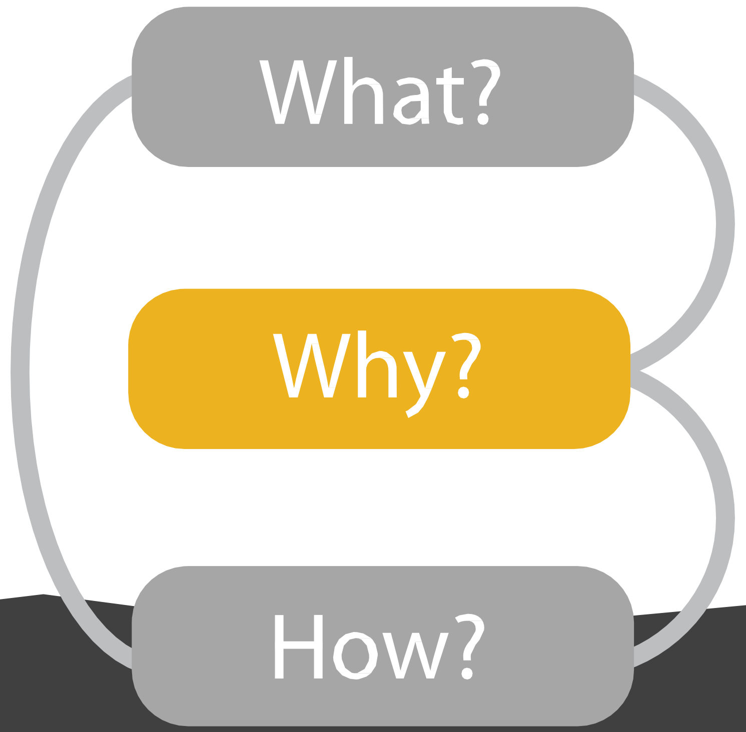


  - abstract vocabulary avoids domain-specific terms
  - what-why-how analysis framework as scaffold to think systematically about design space

task abstraction



	Target known	Target unknown
Location known	Lookup	Browse
Location unknown	Locate	Explore



What?

Why?

How?

{action, *target*} pairs

- discover distribution
- compare trends
- locate outliers
- browse topology

{action, target}

➔ Analyze

➔ Search

➔ Query

{action, target}

➔ **Analyze**

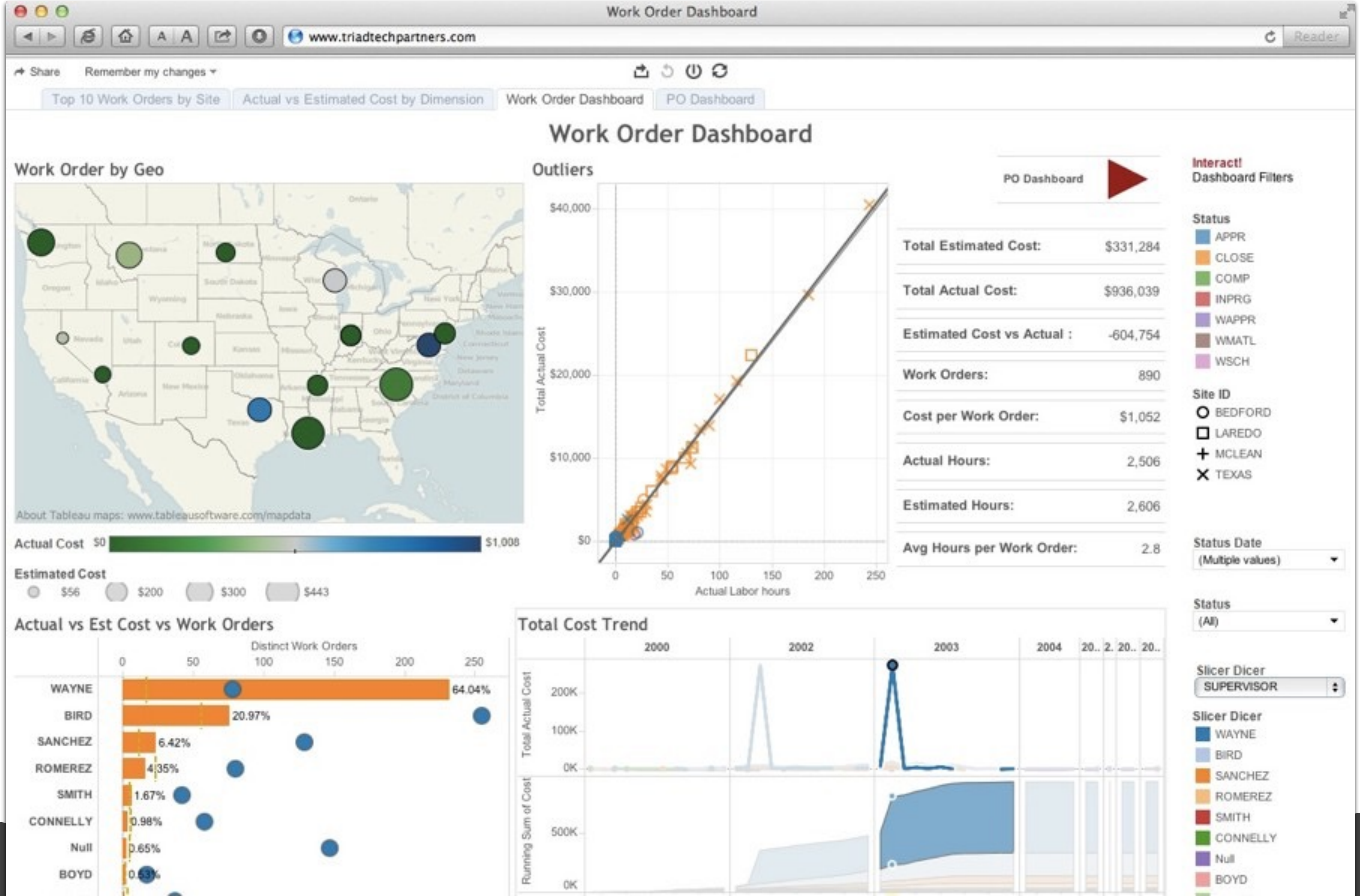
➔ Consume

➔ *Discover* ➔ *Present* ➔ *Enjoy*

➔ **Search**

➔ **Query**

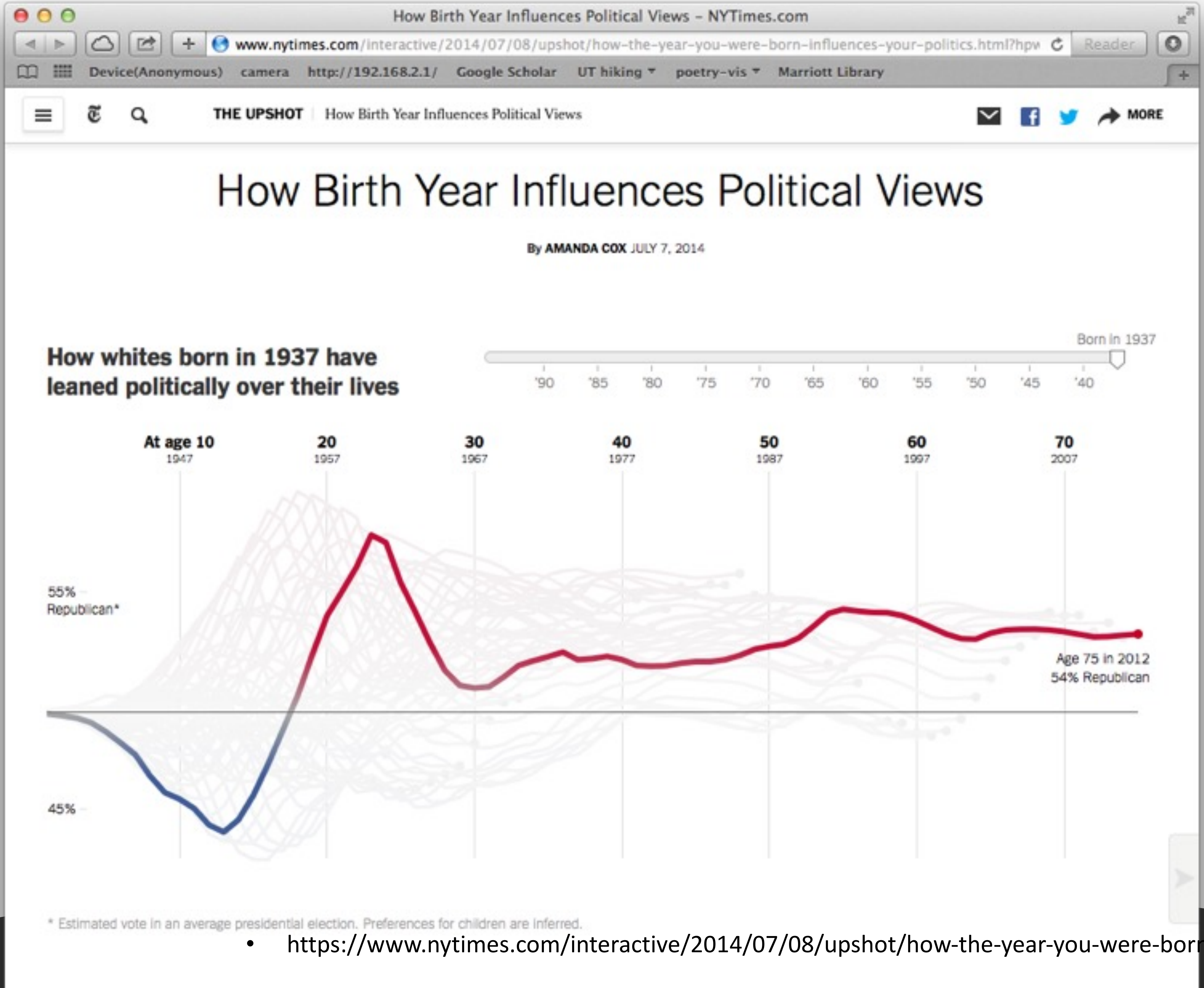
discover



Work Order Dashboard

Total Estimated Cost:	$331,284
Total Actual Cost:	$936,039
Estimated Cost vs Actual :	-604,754
Work Orders:	890
Cost per Work Order:	$1,052
Actual Hours:	2,506
Estimated Hours:	2,606
Avg Hours per Work Order:	2.8

present


How Birth Year Influences Political Views – NYTimes.com
www.nytimes.com/interactive/2014/07/08/upshot/how-the-year-you-were-born-influences-your-politics.html?hpw
Reader
Device(Anonymous)
camera
http://192.168.2.1/
Google Scholar
UT hiking
poetry-vis
Marriott Library
THE UPSHOT
How Birth Year Influences Political Views
MORE

How Birth Year Influences Political Views

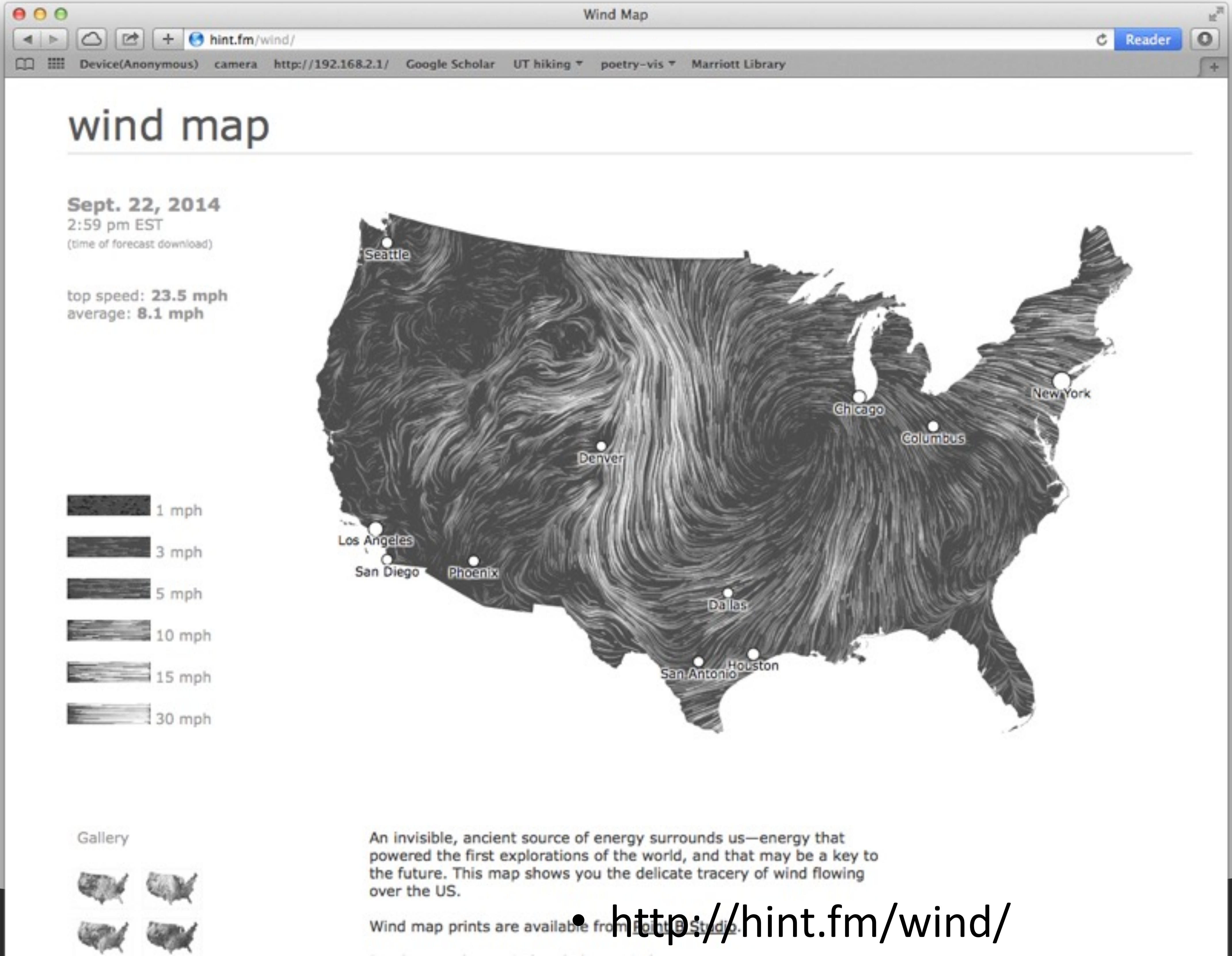
By AMANDA COX JULY 7, 2014

How whites born in 1937 have leaned politically over their lives

* Estimated vote in an average presidential election. Preferences for children are inferred.

- https://www.nytimes.com/interactive/2014/07/08/upshot/how-the-year-you-were-born-influences-your-politics.html

enjoy



wind map

Sept. 22, 2014
2:59 pm EST
(time of forecast download)

top speed: **23.5 mph**
average: **8.1 mph**

▨	1 mph
▨	3 mph
▨	5 mph
▨	10 mph
▨	15 mph
▨	30 mph

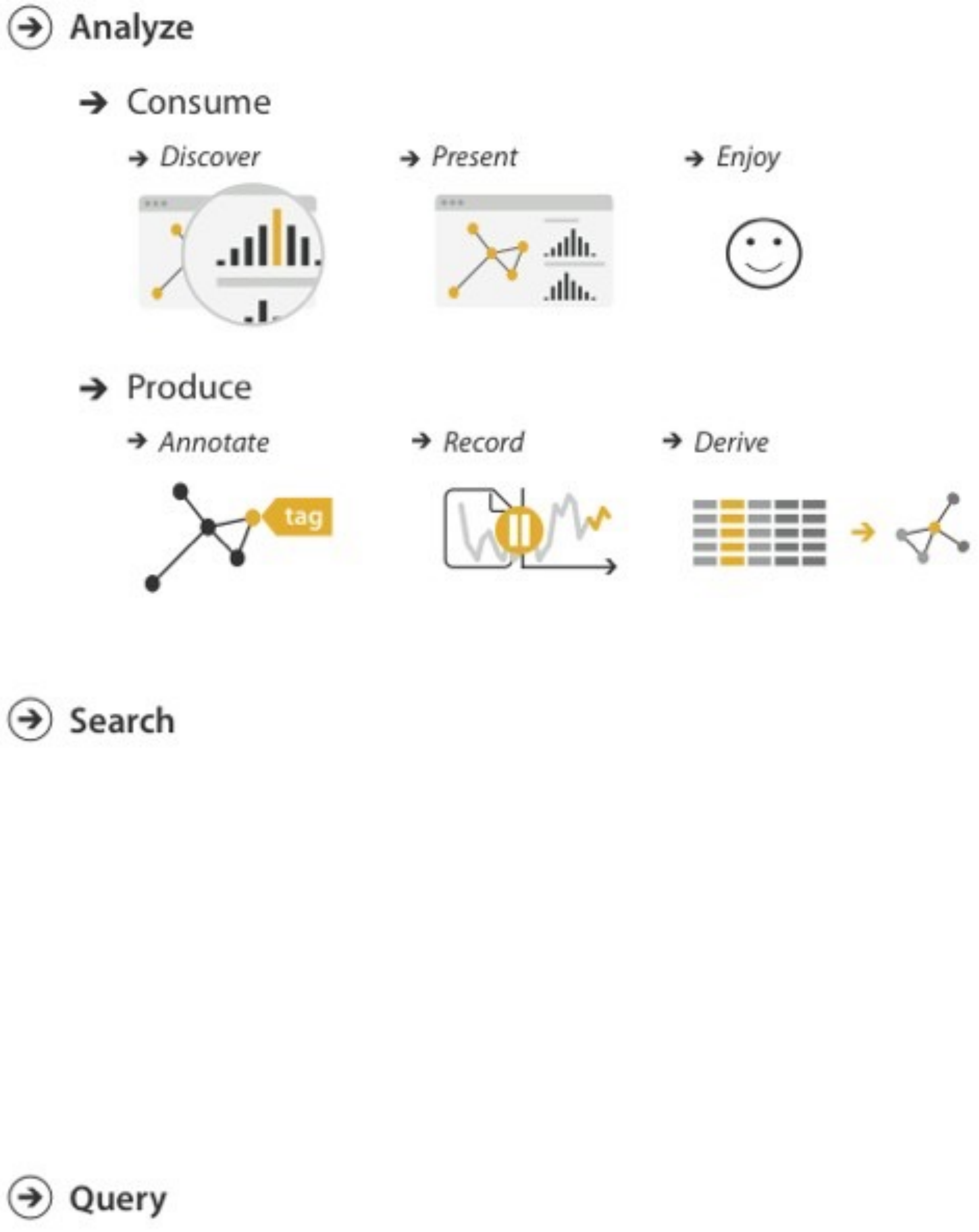
Gallery

An invisible, ancient source of energy surrounds us—energy that powered the first explorations of the world, and that may be a key to the future. This map shows you the delicate tracery of wind flowing over the US.

Wind map prints are available fr

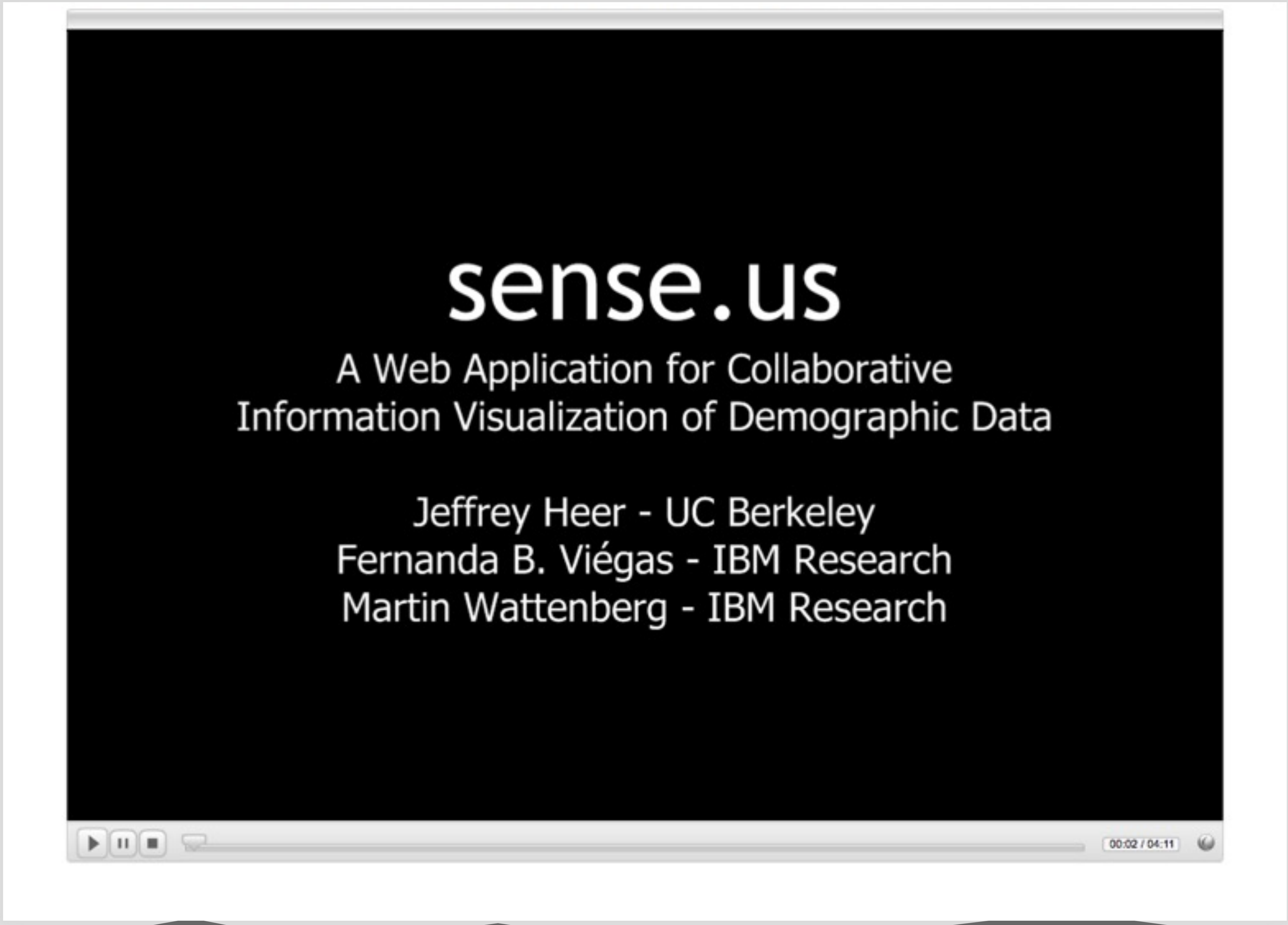
- http://hint.fm/wind/

{action, target}



- Analyze
  - Consume
    - *Discover*
    - *Present*
    - *Enjoy*
  - Produce
    - *Annotate*
    - *Record*
    - *Derive*
- Search
- Query

annotate & record



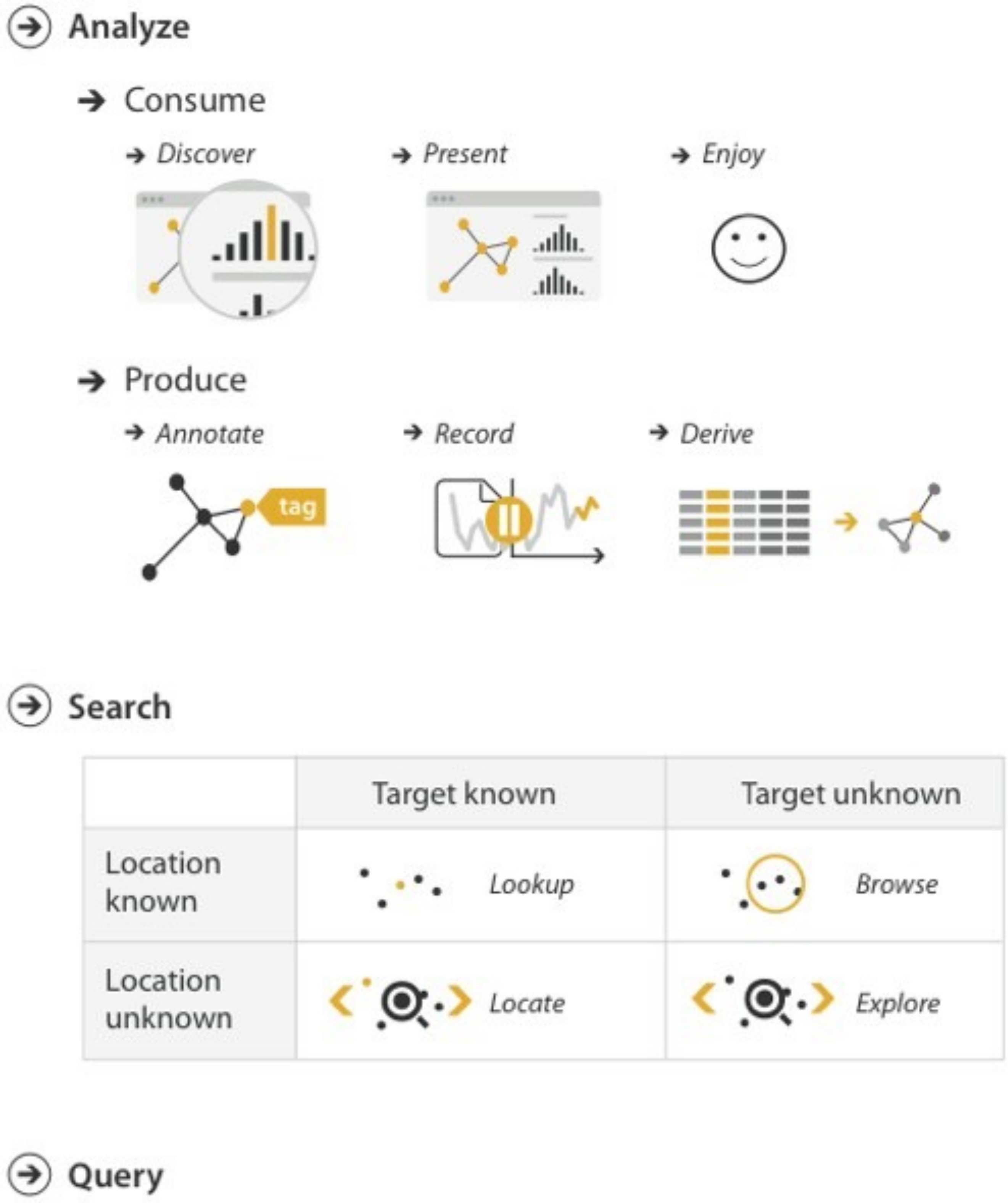
sense.us

A Web Application for Collaborative
Information Visualization of Demographic Data

Jeffrey Heer - UC Berkeley
Fernanda B. Viégas - IBM Research
Martin Wattenberg - IBM Research

http://vis.stanford.edu/files/sense.us/

{action, target}



Analyze

Consume

- Discover
- Present
- Enjoy

Produce

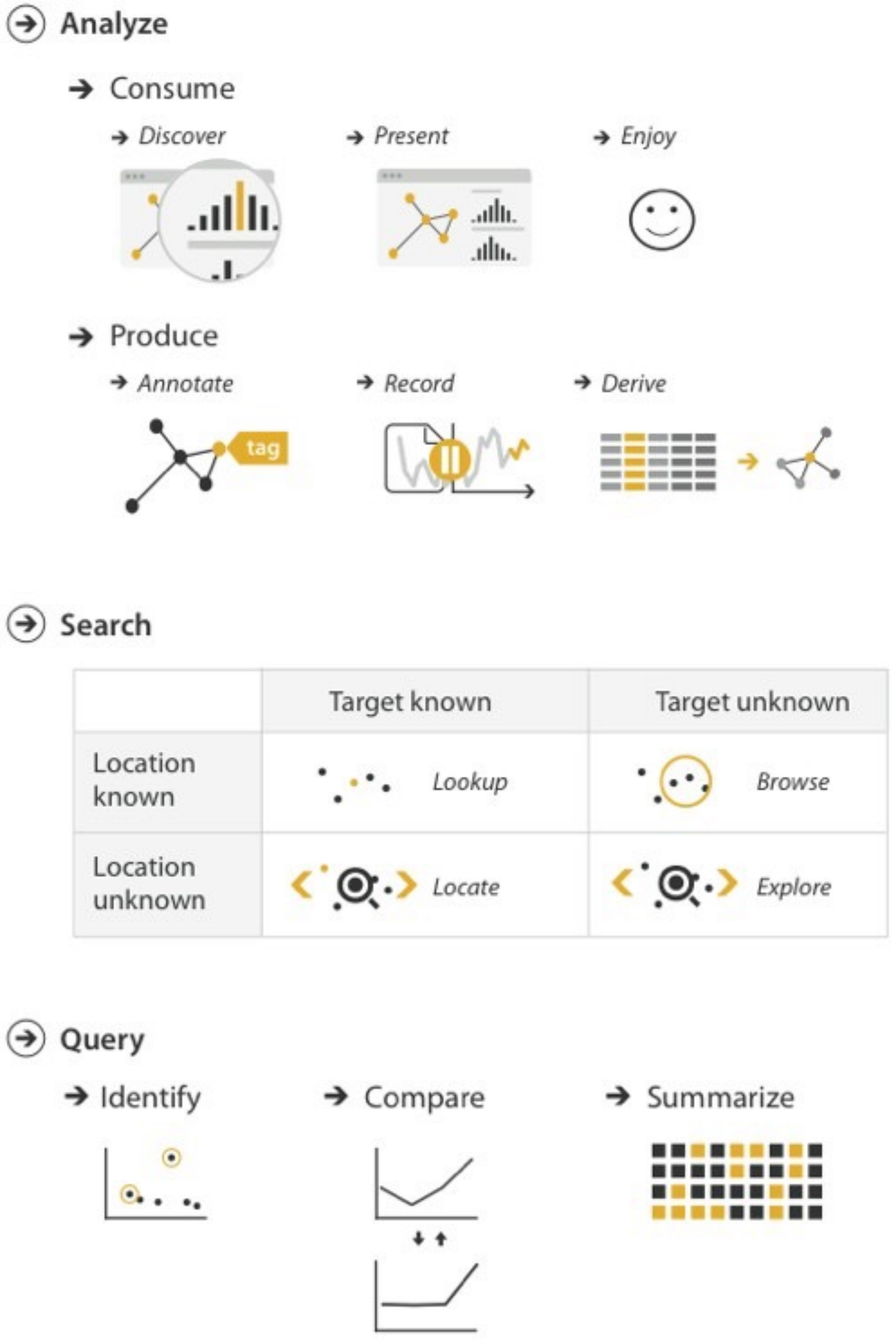
- Annotate
- Record
- Derive

Search

	Target known	Target unknown
Location known	Lookup	Browse
Location unknown	Locate	Explore

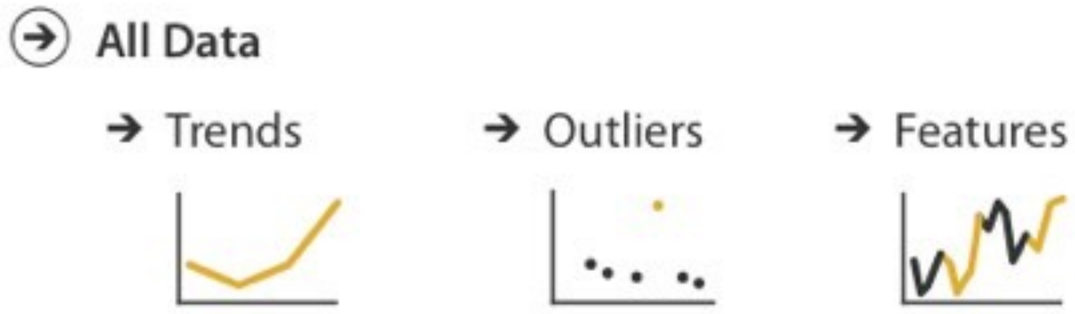
Query

{action, target}



	Target known	Target unknown
Location known	Lookup	Browse
Location unknown	Locate	Explore

{action, **target**}



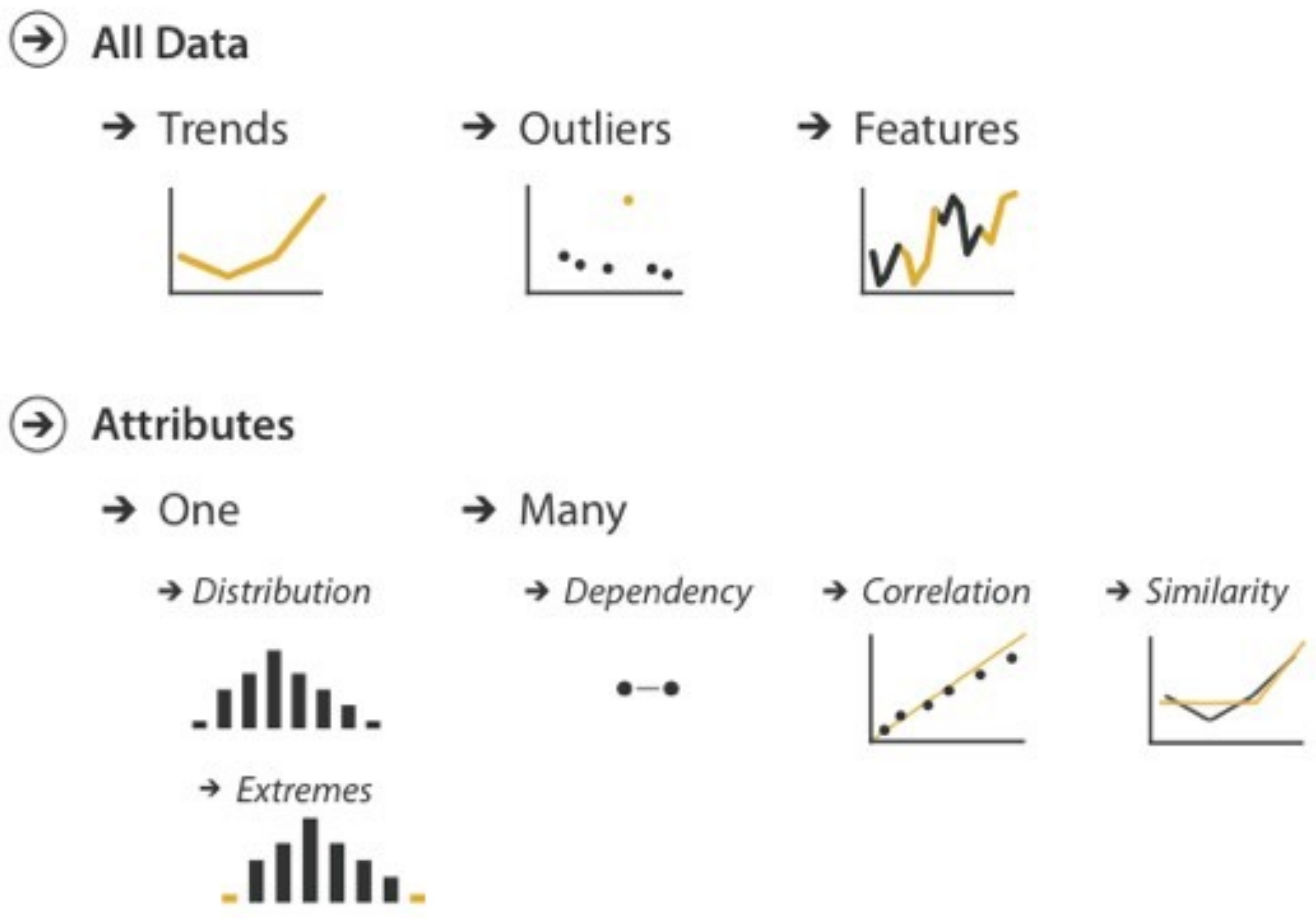
All Data

Trends

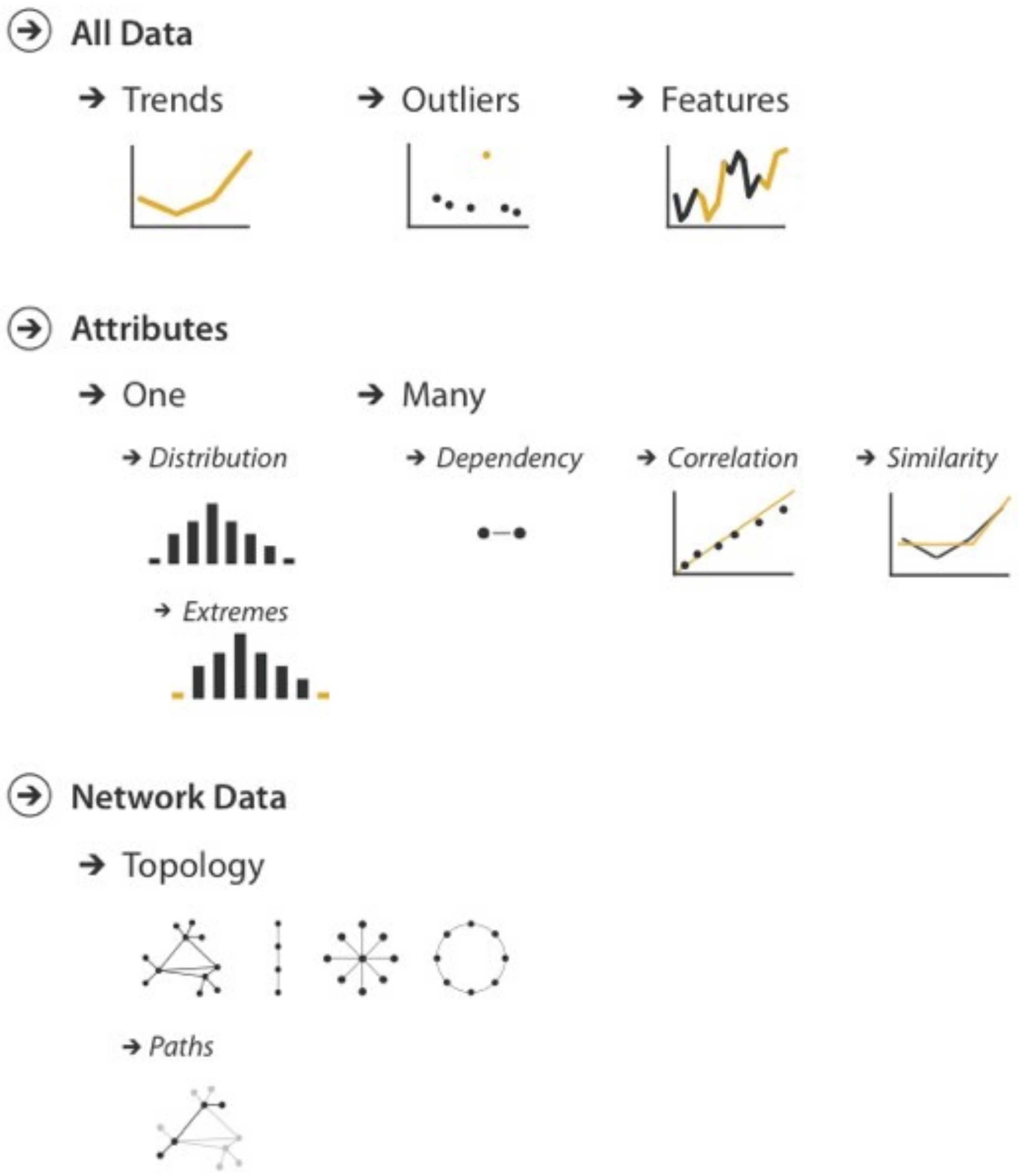
Outliers

Features

{action, **target**}



{action, **target**}



All Data

- Trends
- Outliers
- Features

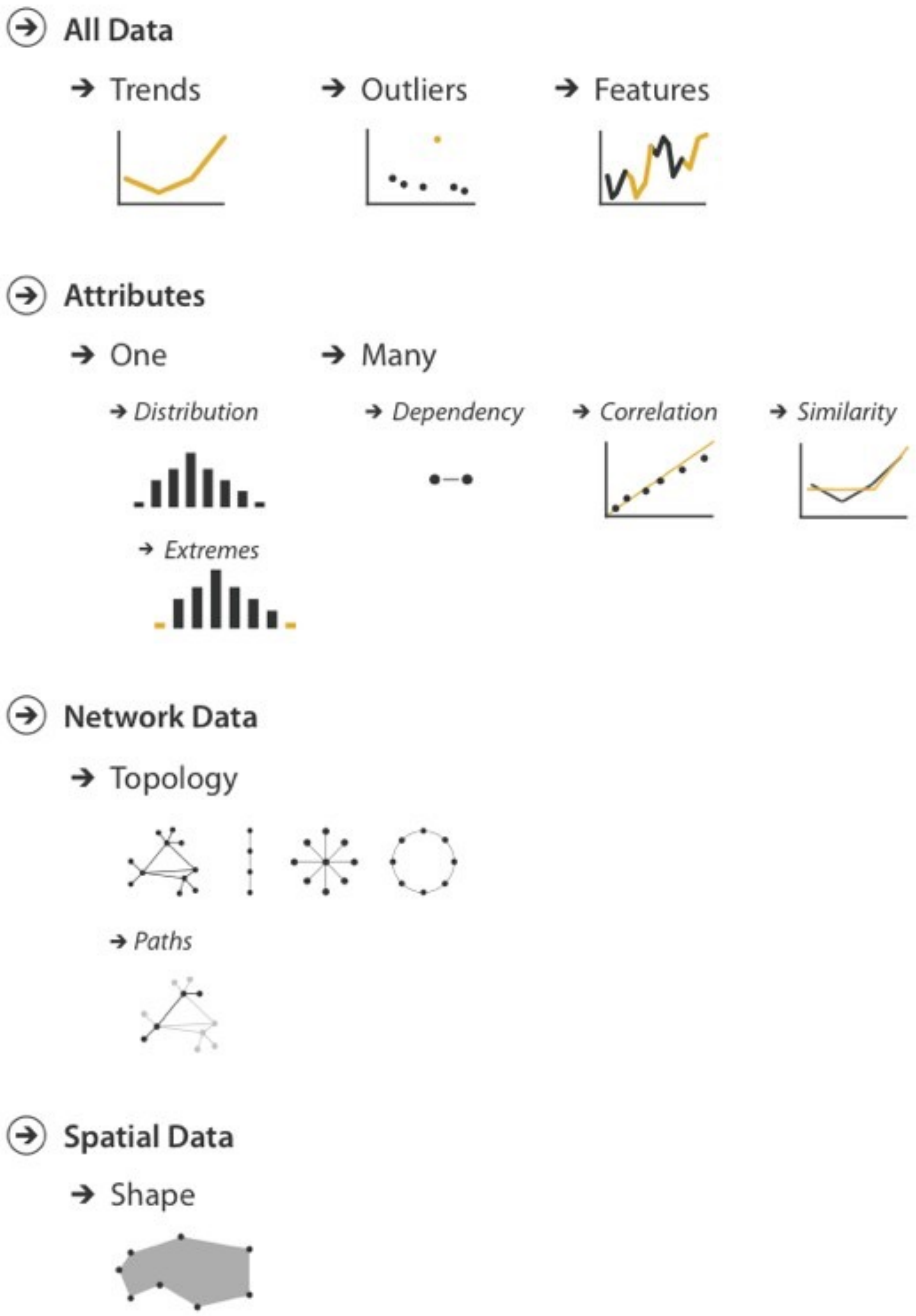
Attributes

- One
  - *Distribution*
  - *Extremes*
- Many
  - *Dependency*
  - *Correlation*
  - *Similarity*

Network Data

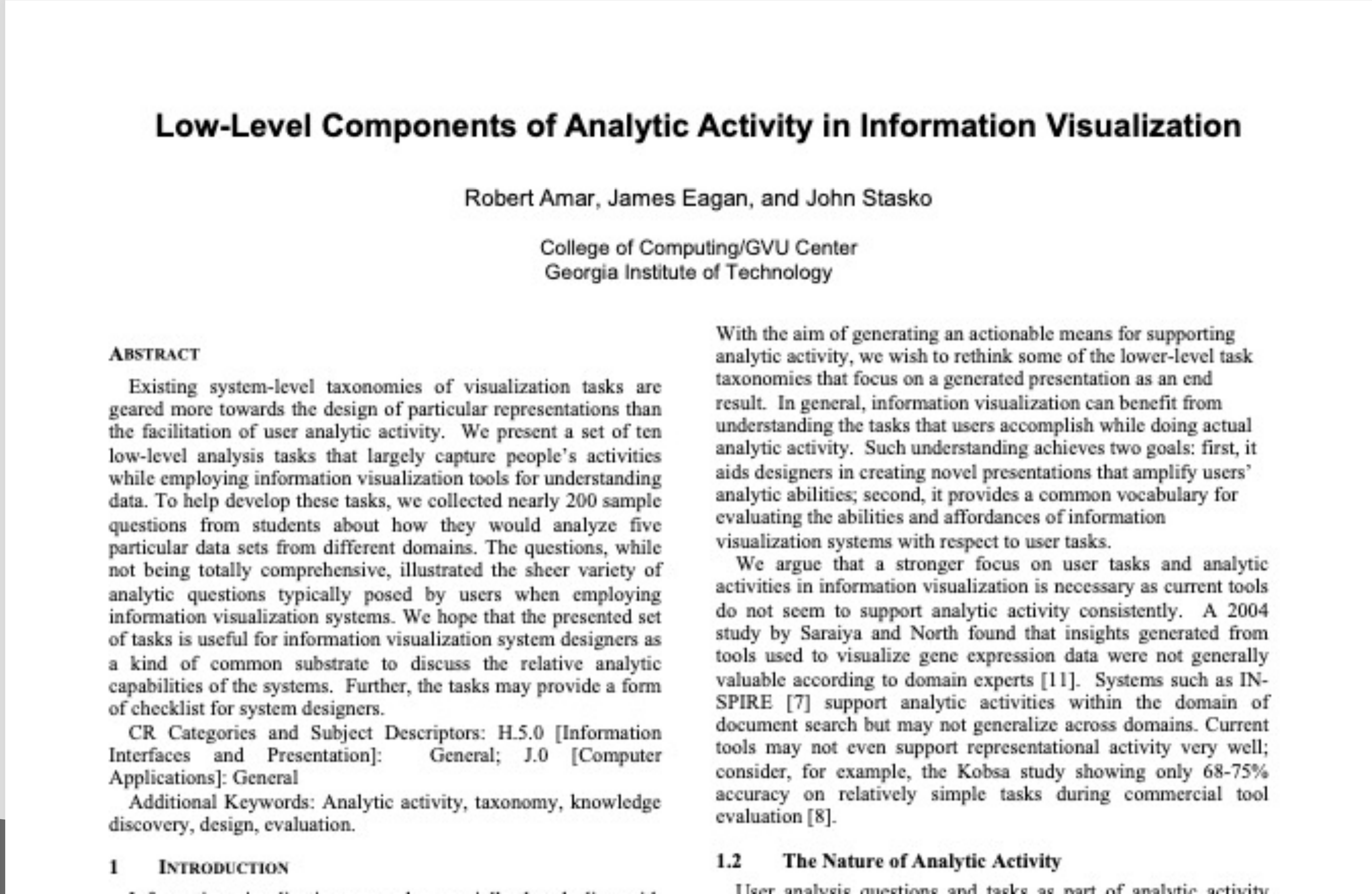
- Topology
  - *Paths*

{action, **target**}



- All Data
  - Trends
  - Outliers
  - Features
- Attributes
  - One
    - *Distribution*
    - *Extremes*
  - Many
    - *Dependency*
    - *Correlation*
    - *Similarity*
- Network Data
  - Topology
    - *Paths*
- Spatial Data
  - Shape

Low-level task taxonomy



Low-Level Components of Analytic Activity in Information Visualization

Robert Amar, James Eagan, and John Stasko

College of Computing/GVU Center
Georgia Institute of Technology

Abstract

Existing system-level taxonomies of visualization tasks are geared more towards the design of particular representations than the facilitation of user analytic activity. We present a set of ten low-level analysis tasks that largely capture people's activities while employing information visualization tools for understanding data. To help develop these tasks, we collected nearly 200 sample questions from students about how they would analyze five particular data sets from different domains. The questions, while not being totally comprehensive, illustrated the sheer variety of analytic questions typically posed by users when employing information visualization systems. We hope that the presented set of tasks is useful for information visualization system designers as a kind of common substrate to discuss the relative analytic capabilities of the systems. Further, the tasks may provide a form of checklist for system designers.

CR Categories and Subject Descriptors: H.5.0 [Information Interfaces and Presentation]: General; J.0 [Computer Applications]: General

Additional Keywords: Analytic activity, taxonomy, knowledge discovery, design, evaluation.

1 Introduction

With the aim of generating an actionable means for supporting analytic activity, we wish to rethink some of the lower-level task taxonomies that focus on a generated presentation as an end result. In general, information visualization can benefit from understanding the tasks that users accomplish while doing actual analytic activity. Such understanding achieves two goals: first, it aids designers in creating novel presentations that amplify users' analytic abilities; second, it provides a common vocabulary for evaluating the abilities and affordances of information visualization systems with respect to user tasks.

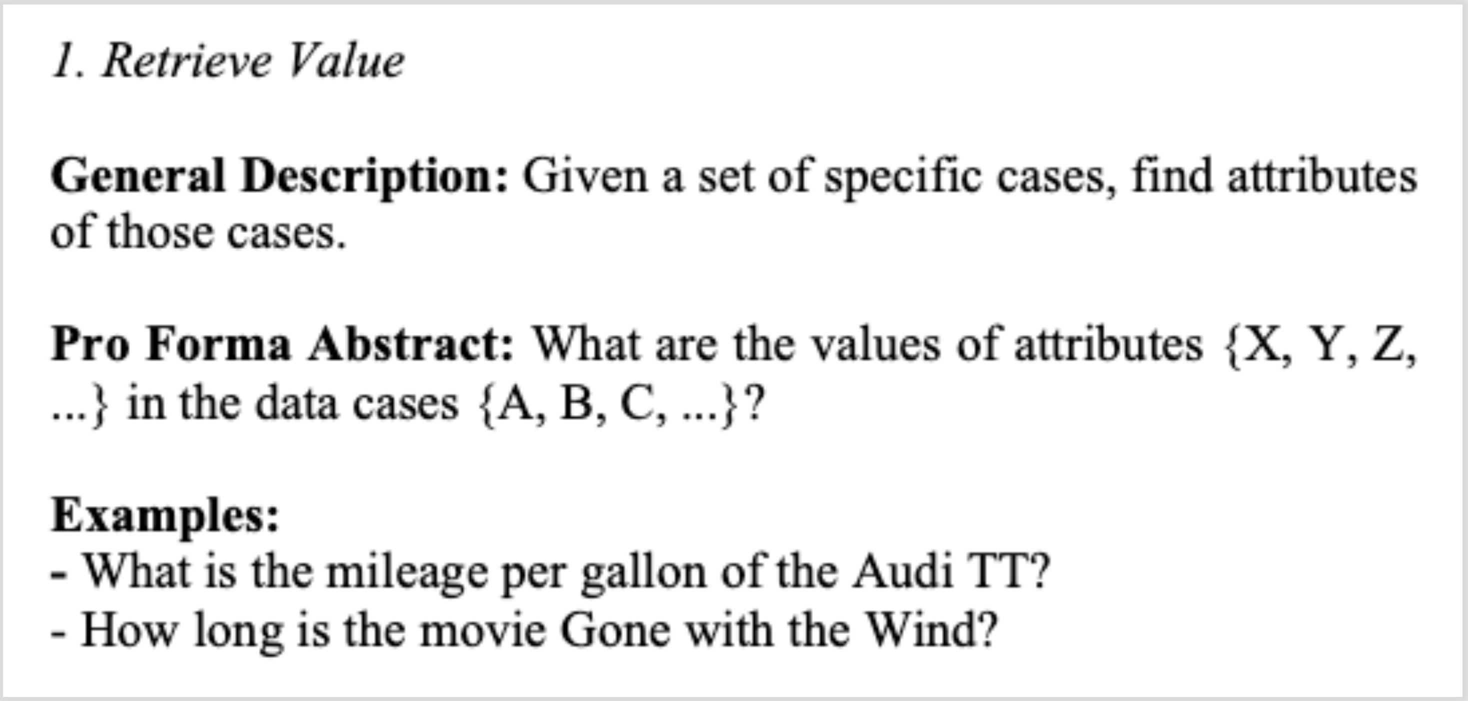
We argue that a stronger focus on user tasks and analytic activities in information visualization is necessary as current tools do not seem to support analytic activity consistently. A 2004 study by Saraiya and North found that insights generated from tools used to visualize gene expression data were not generally valuable according to domain experts [11]. Systems such as IN-SPIRE [7] support analytic activities within the domain of document search but may not generalize across domains. Current tools may not even support representational activity very well; consider, for example, the Kobsa study showing only 68-75% accuracy on relatively simple tasks during commercial tool evaluation [8].

1.2 The Nature of Analytic Activity

User analysis questions and tasks as part of analytic activity

Visualization, InfoVis 2005

Tasks

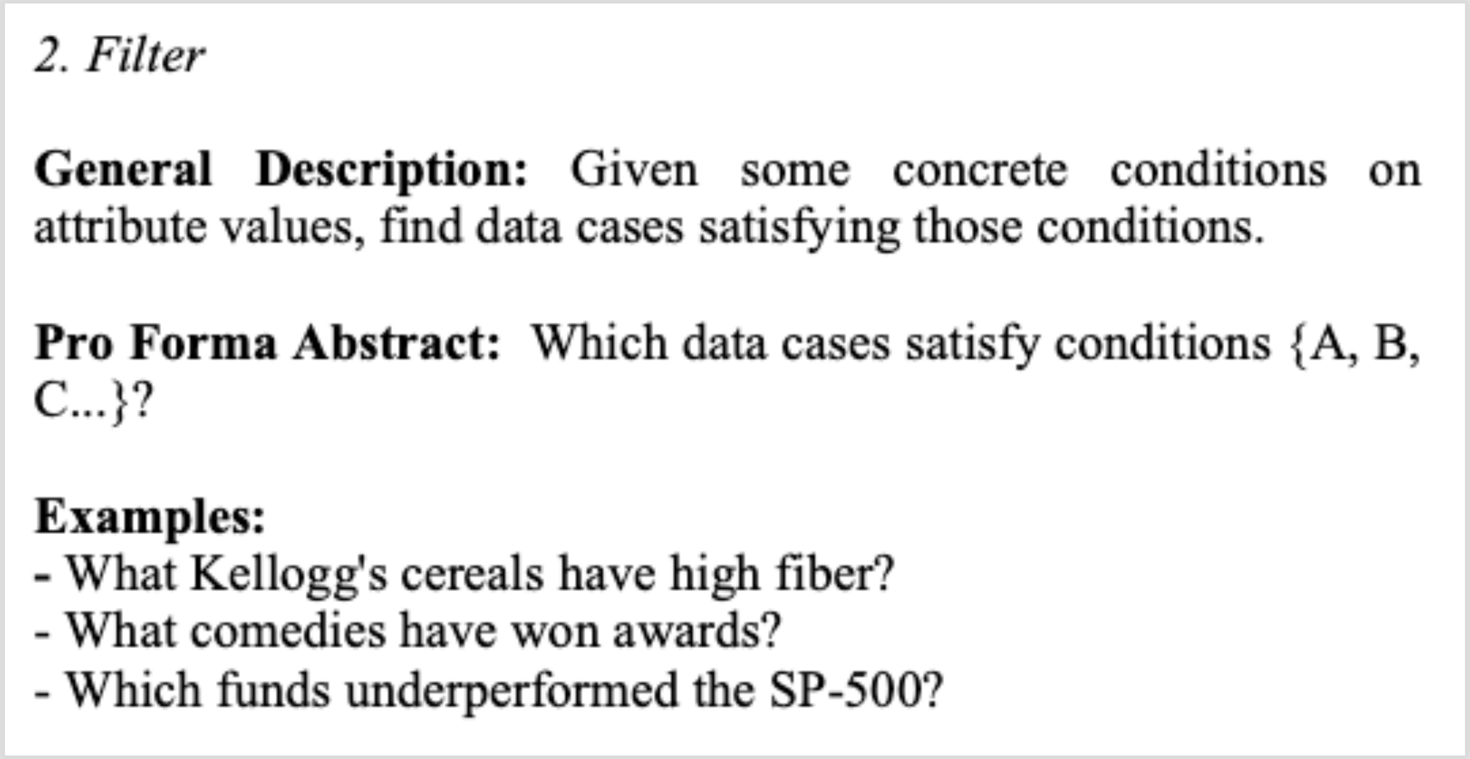


1. Retrieve Value

General Description: Given a set of specific cases, find attributes of those cases.

Pro Forma Abstract: What are the values of attributes {X, Y, Z, ...} in the data cases {A, B, C, ...}?

Examples:
- What is the mileage per gallon of the Audi TT?
- How long is the movie Gone with the Wind?



2. Filter

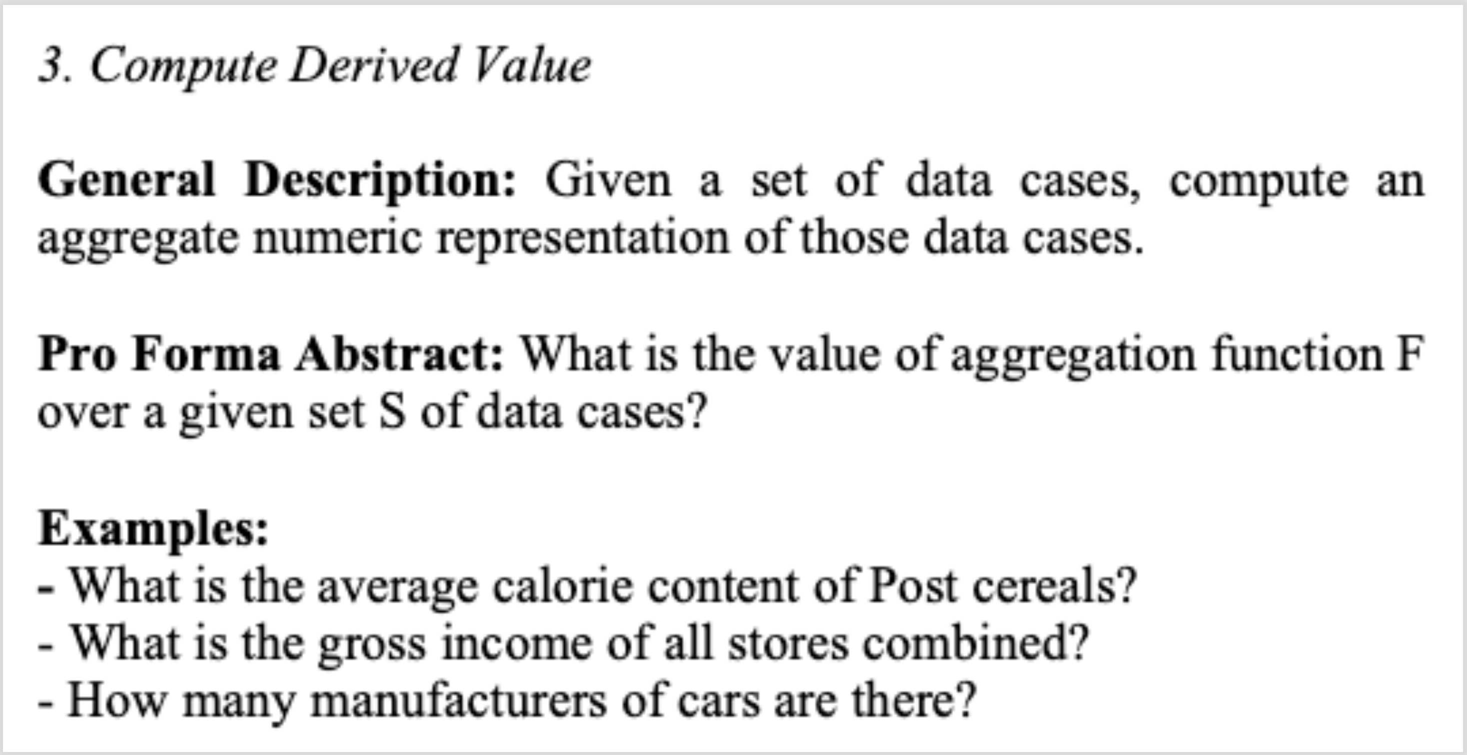
General Description: Given some concrete conditions on attribute values, find data cases satisfying those conditions.

Pro Forma Abstract: Which data cases satisfy conditions {A, B, C...}?

Examples:
- What Kellogg's cereals have high fiber?
- What comedies have won awards?
- Which funds underperformed the SP-500?

- Amar et al, ow-Level Components of Analytic Activity in Information Visualization, InfoVis 2005

Tasks

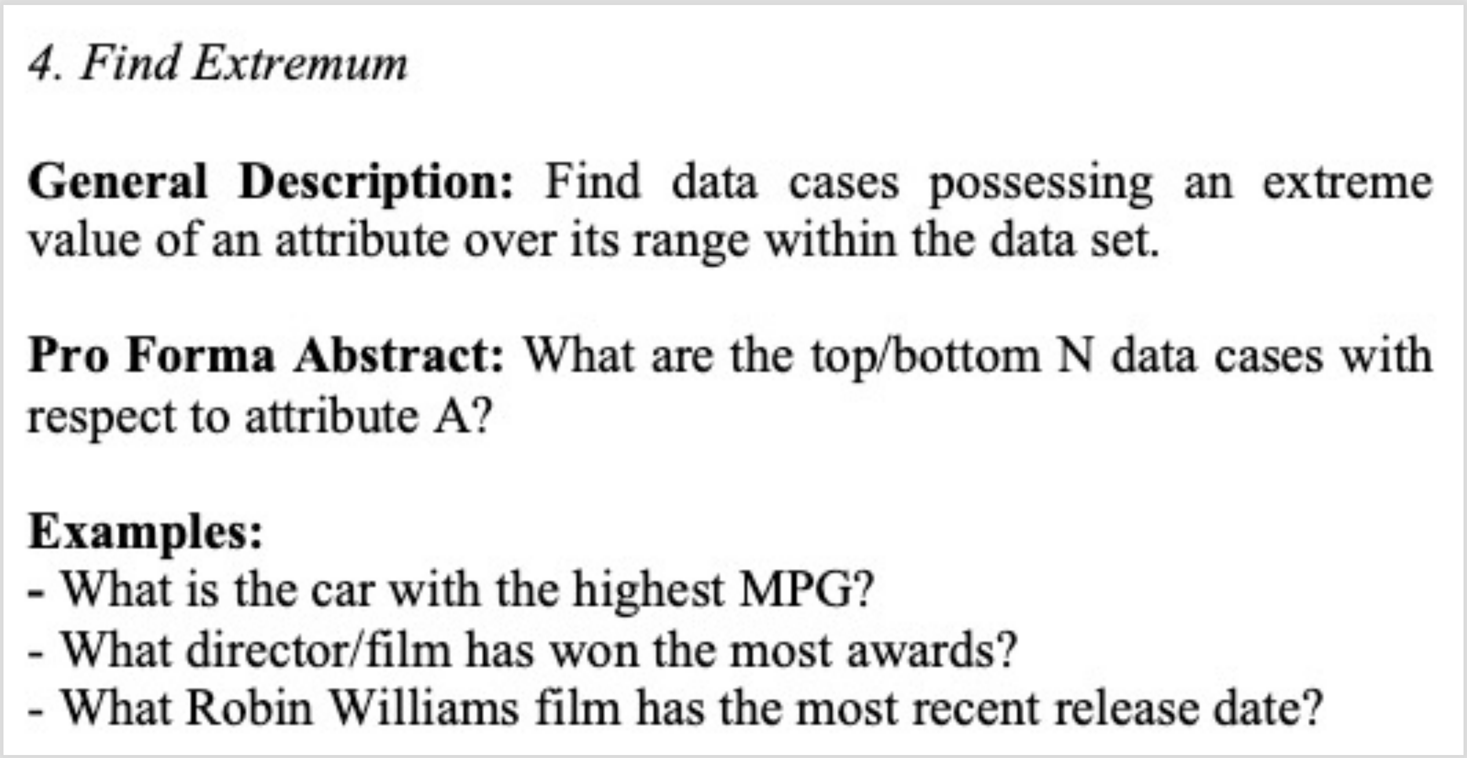


3. Compute Derived Value

General Description: Given a set of data cases, compute an aggregate numeric representation of those data cases.

Pro Forma Abstract: What is the value of aggregation function F over a given set S of data cases?

Examples:
- What is the average calorie content of Post cereals?
- What is the gross income of all stores combined?
- How many manufacturers of cars are there?



4. Find Extremum

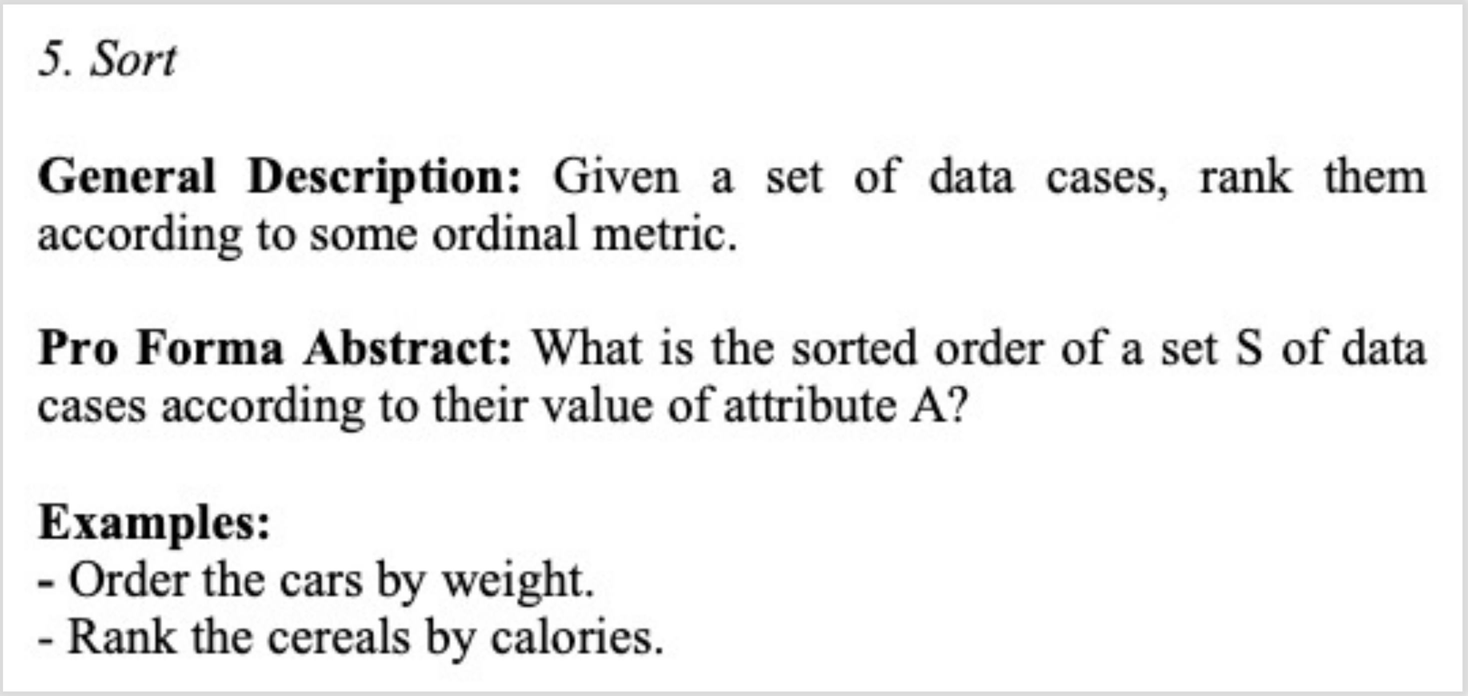
General Description: Find data cases possessing an extreme value of an attribute over its range within the data set.

Pro Forma Abstract: What are the top/bottom N data cases with respect to attribute A?

Examples:
- What is the car with the highest MPG?
- What director/film has won the most awards?
- What Robin Williams film has the most recent release date?

- Amar et al, ow-Level Components of Analytic Activity in Information Visualization, InfoVis 2005

Tasks

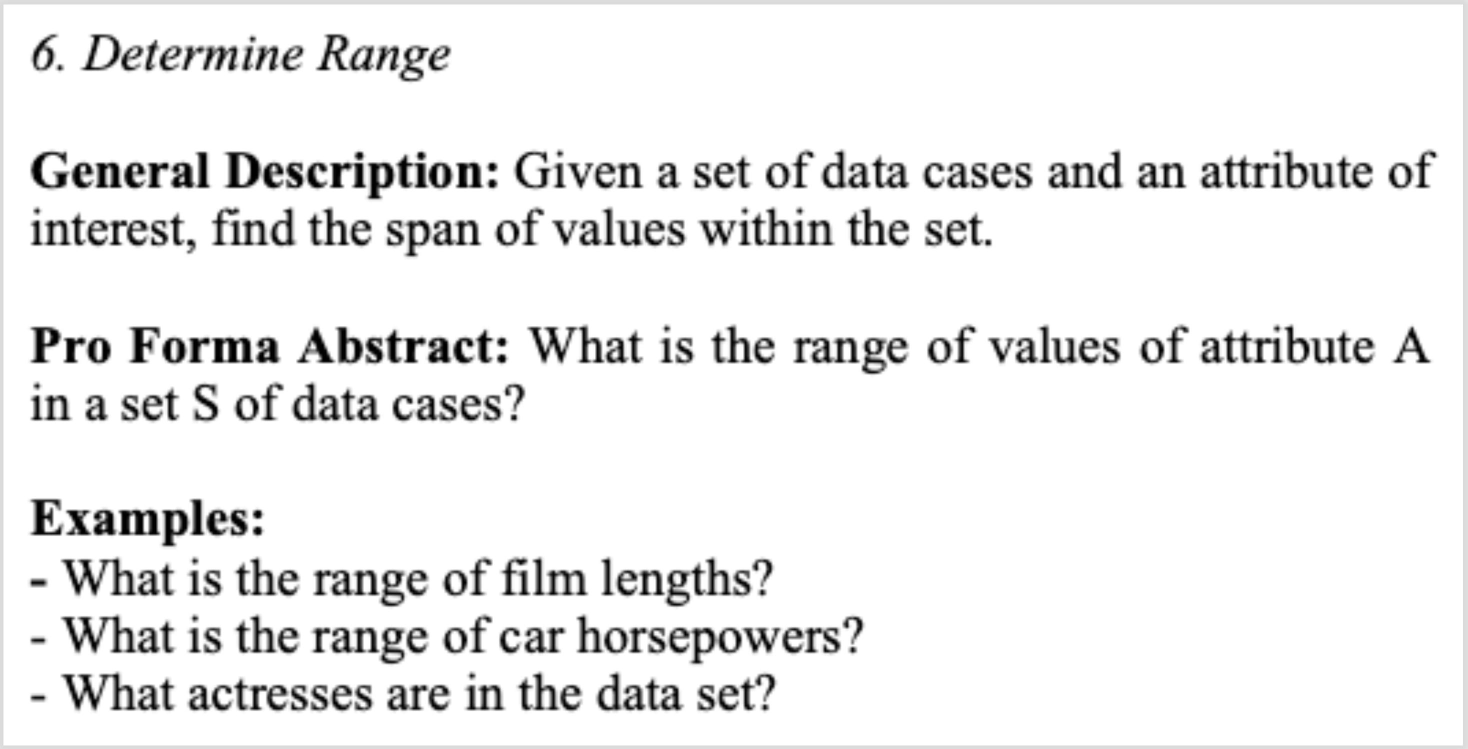


5. Sort

General Description: Given a set of data cases, rank them according to some ordinal metric.

Pro Forma Abstract: What is the sorted order of a set S of data cases according to their value of attribute A?

Examples:
- Order the cars by weight.
- Rank the cereals by calories.



6. Determine Range

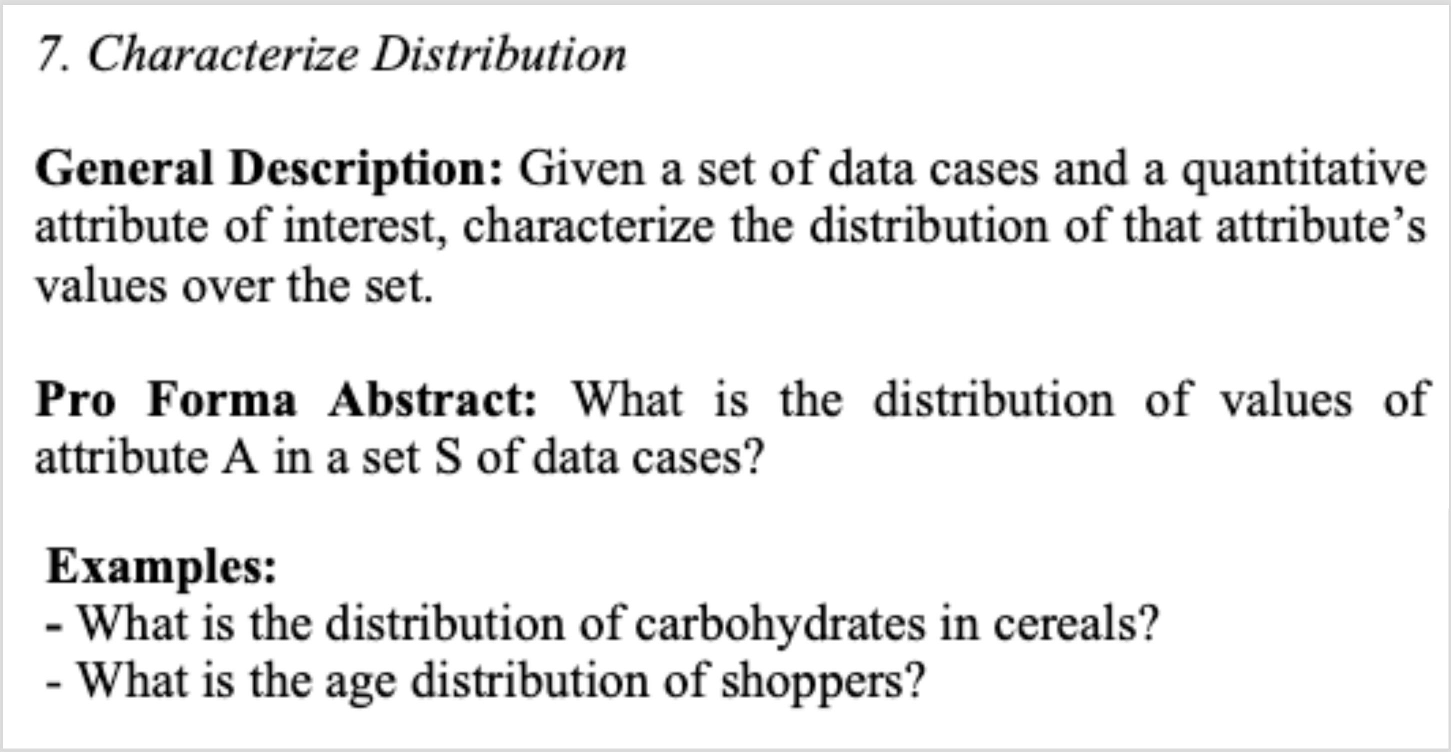
General Description: Given a set of data cases and an attribute of interest, find the span of values within the set.

Pro Forma Abstract: What is the range of values of attribute A in a set S of data cases?

Examples:
- What is the range of film lengths?
- What is the range of car horsepowers?
- What actresses are in the data set?

- Amar et al, ow-Level Components of Analytic Activity in Information Visualization, InfoVis 2005

Tasks

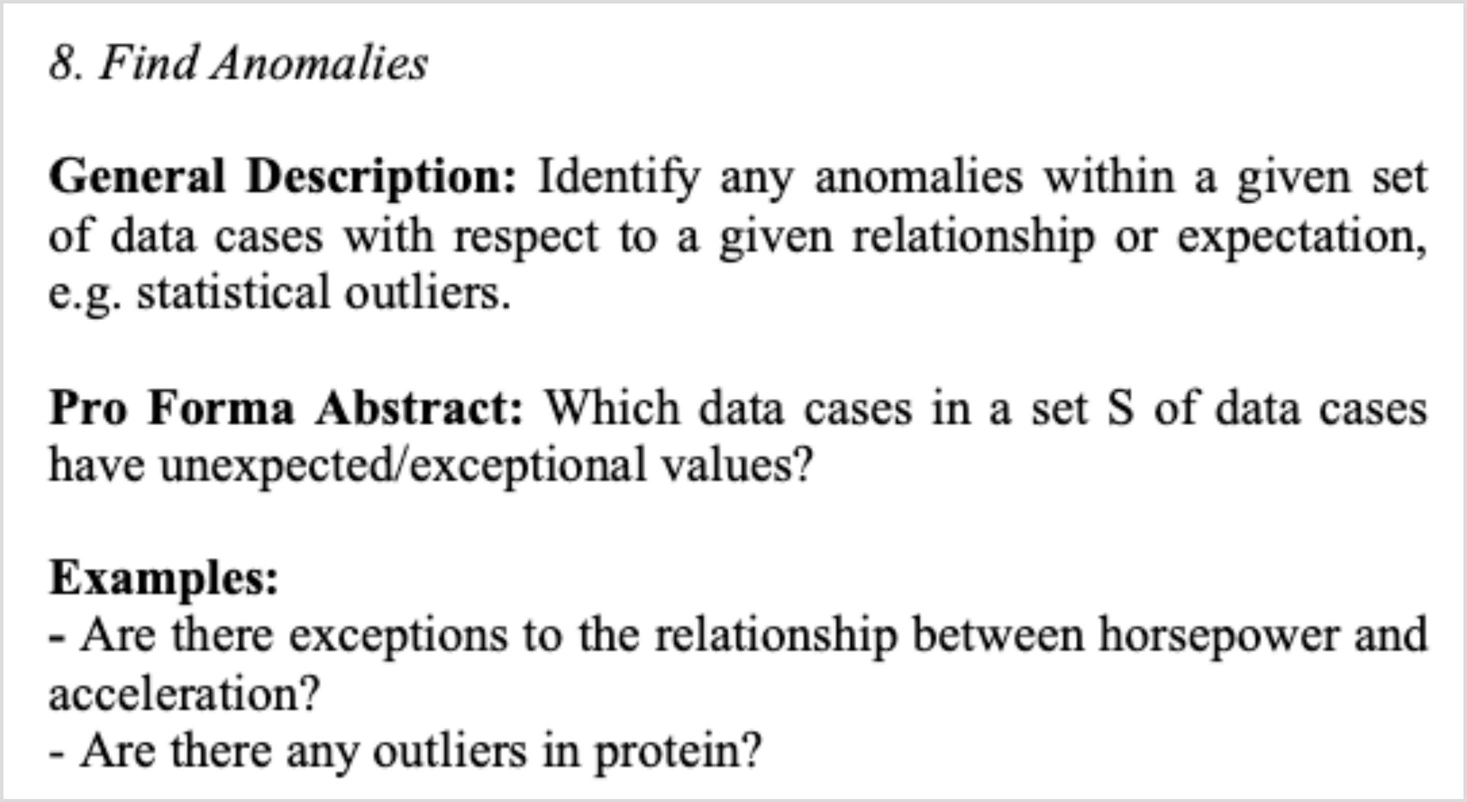


7. Characterize Distribution

General Description: Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute's values over the set.

Pro Forma Abstract: What is the distribution of values of attribute A in a set S of data cases?

Examples:
- What is the distribution of carbohydrates in cereals?
- What is the age distribution of shoppers?



8. Find Anomalies

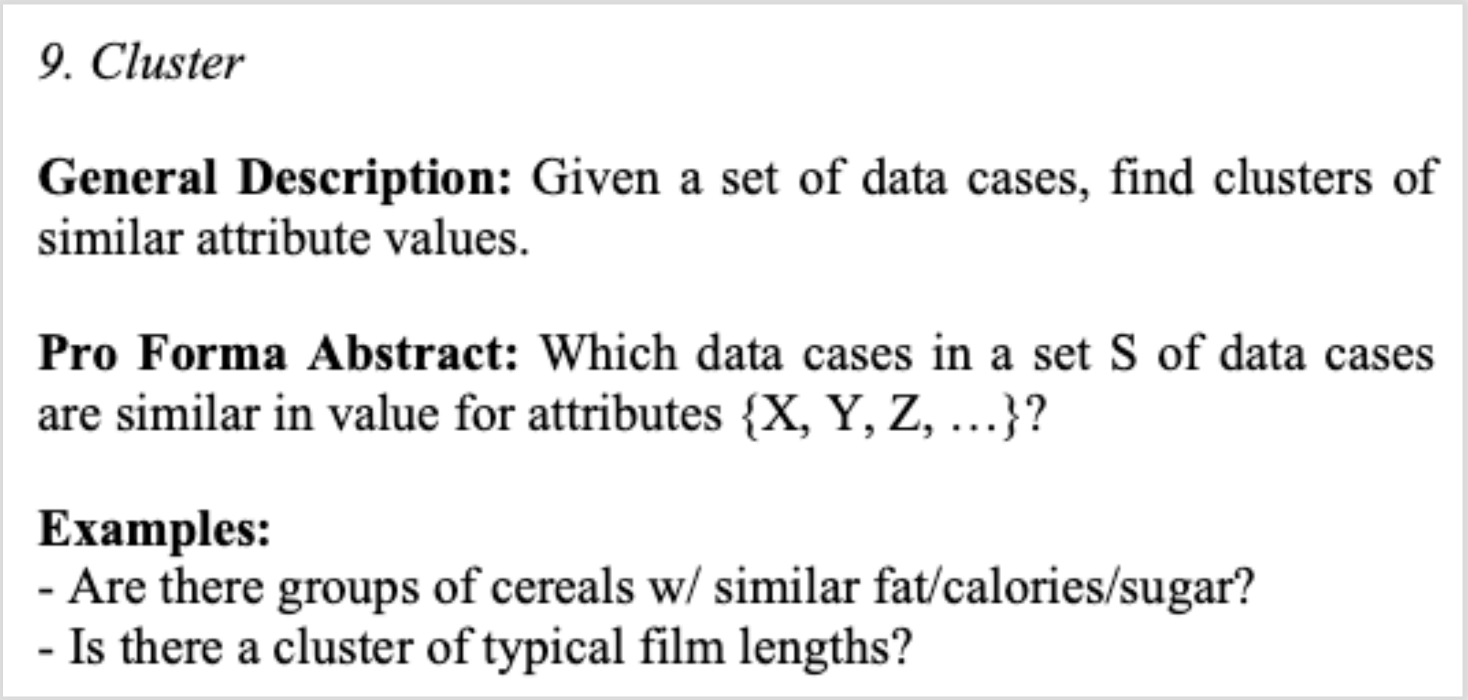
General Description: Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g. statistical outliers.

Pro Forma Abstract: Which data cases in a set S of data cases have unexpected/exceptional values?

Examples:
- Are there exceptions to the relationship between horsepower and acceleration?
- Are there any outliers in protein?

- Amar et al, ow-Level Components of Analytic Activity in Information Visualization, InfoVis 2005

Tasks

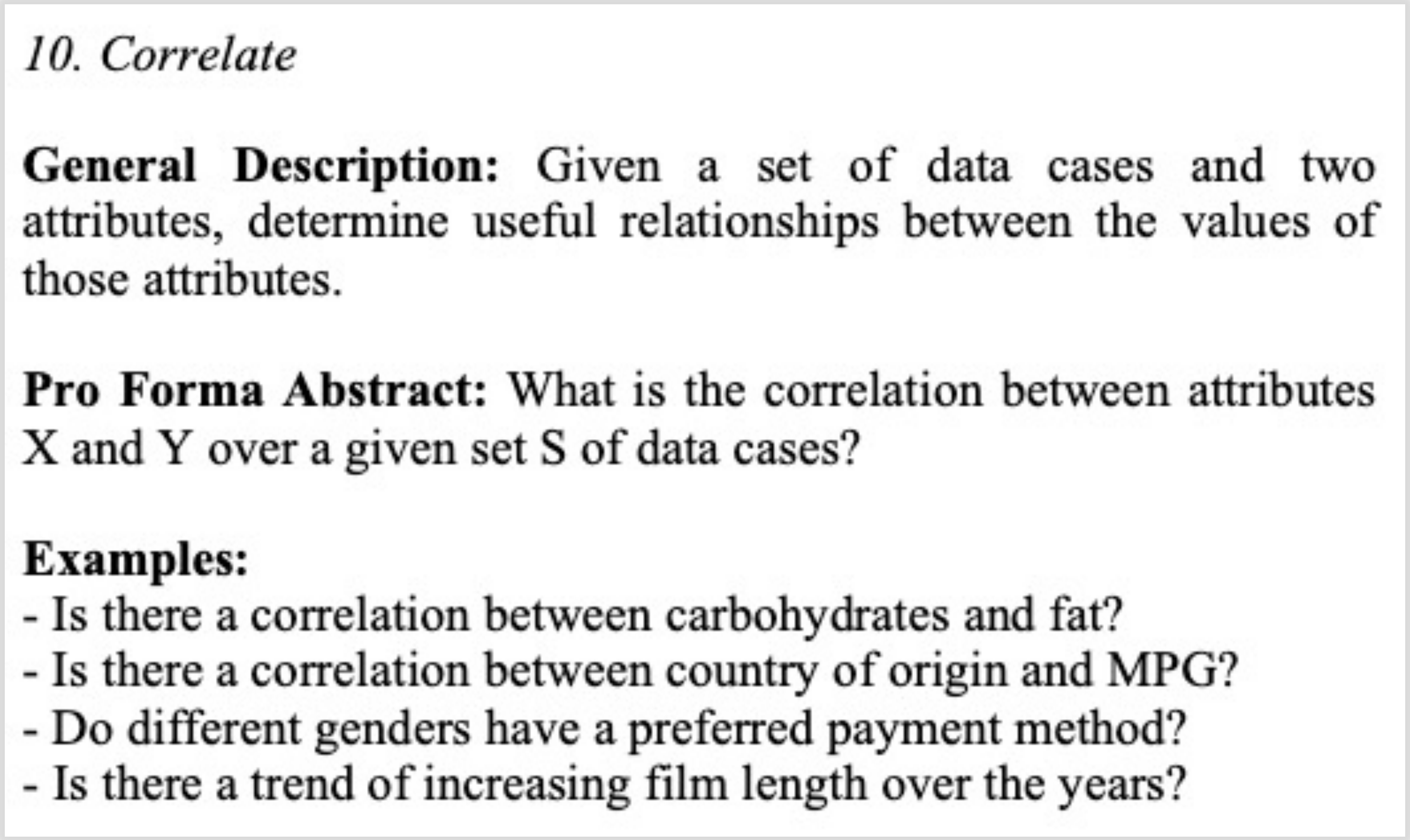


9. Cluster

General Description: Given a set of data cases, find clusters of similar attribute values.

Pro Forma Abstract: Which data cases in a set S of data cases are similar in value for attributes {X, Y, Z, …}?

Examples:
- Are there groups of cereals w/ similar fat/calories/sugar?
- Is there a cluster of typical film lengths?



10. Correlate

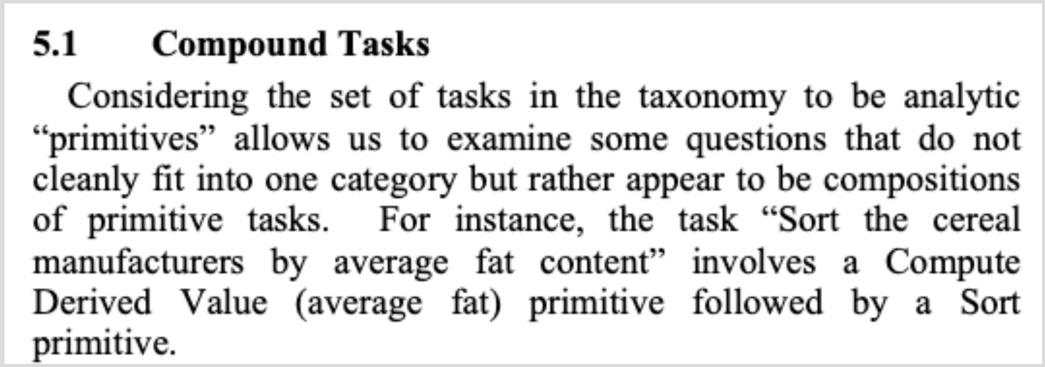
General Description: Given a set of data cases and two attributes, determine useful relationships between the values of those attributes.

Pro Forma Abstract: What is the correlation between attributes X and Y over a given set S of data cases?

Examples:
- Is there a correlation between carbohydrates and fat?
- Is there a correlation between country of origin and MPG?
- Do different genders have a preferred payment method?
- Is there a trend of increasing film length over the years?

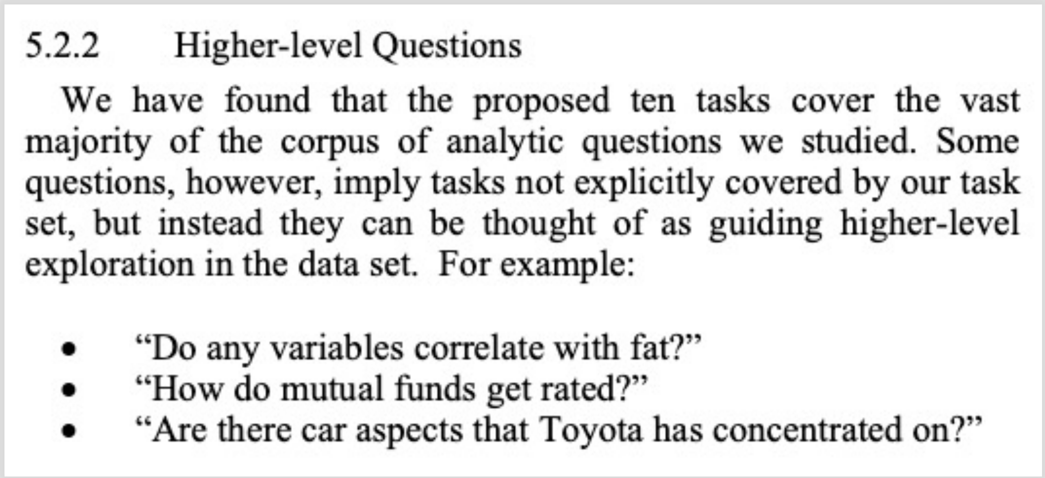
- Amar et al, ow-Level Components of Analytic Activity in Information Visualization, InfoVis 2005

Tasks Not Considered



5.1 Compound Tasks

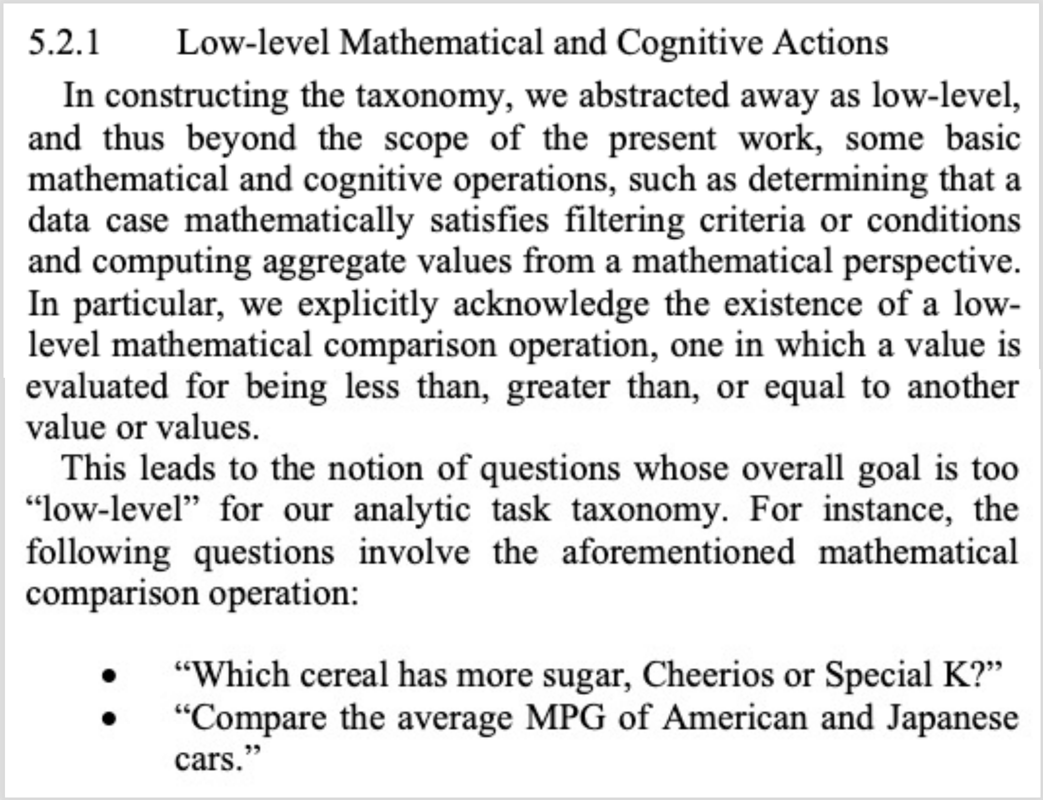
Considering the set of tasks in the taxonomy to be analytic “primitives” allows us to examine some questions that do not cleanly fit into one category but rather appear to be compositions of primitive tasks. For instance, the task “Sort the cereal manufacturers by average fat content” involves a Compute Derived Value (average fat) primitive followed by a Sort primitive.



5.2.2 Higher-level Questions

We have found that the proposed ten tasks cover the vast majority of the corpus of analytic questions we studied. Some questions, however, imply tasks not explicitly covered by our task set, but instead they can be thought of as guiding higher-level exploration in the data set. For example:

- “Do any variables correlate with fat?”
- “How do mutual funds get rated?”
- “Are there car aspects that Toyota has concentrated on?”

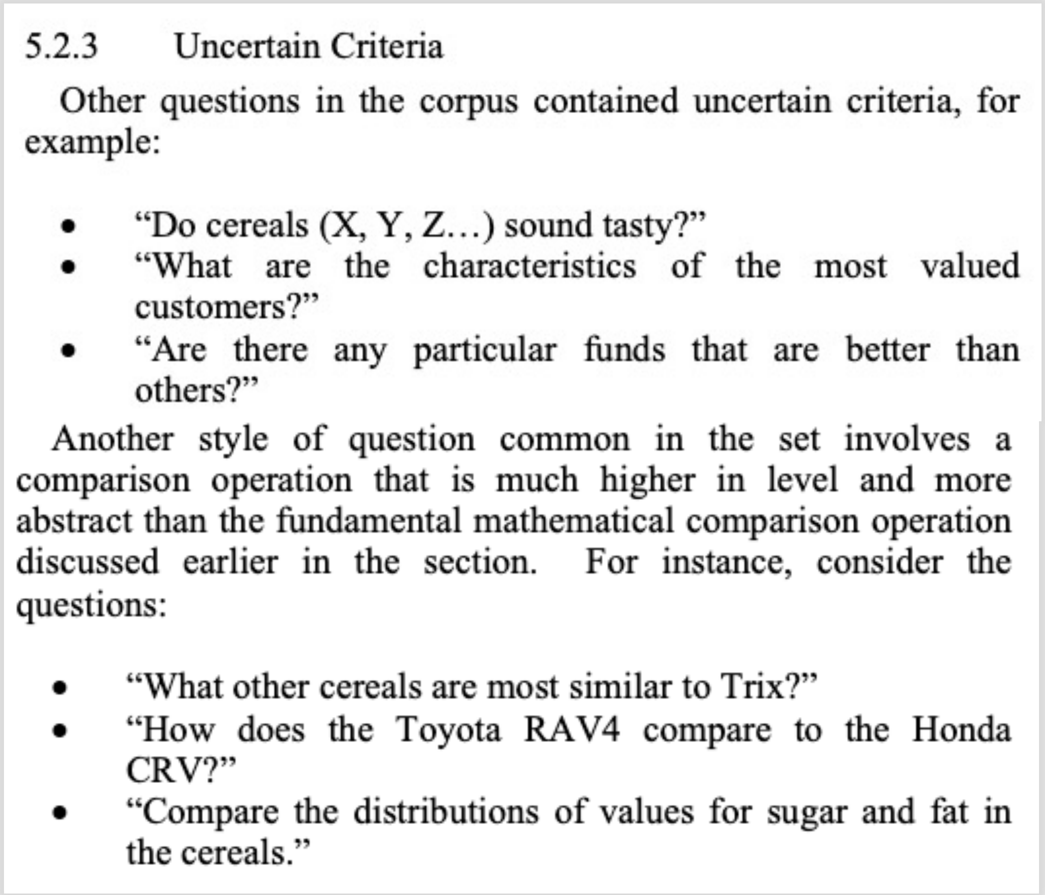


5.2.1 Low-level Mathematical and Cognitive Actions

In constructing the taxonomy, we abstracted away as low-level, and thus beyond the scope of the present work, some basic mathematical and cognitive operations, such as determining that a data case mathematically satisfies filtering criteria or conditions and computing aggregate values from a mathematical perspective. In particular, we explicitly acknowledge the existence of a low-level mathematical comparison operation, one in which a value is evaluated for being less than, greater than, or equal to another value or values.

This leads to the notion of questions whose overall goal is too “low-level” for our analytic task taxonomy. For instance, the following questions involve the aforementioned mathematical comparison operation:

- “Which cereal has more sugar, Cheerios or Special K?”
- “Compare the average MPG of American and Japanese cars.”



5.2.3 Uncertain Criteria

Other questions in the corpus contained uncertain criteria, for example:

- “Do cereals (X, Y, Z…) sound tasty?”
- “What are the characteristics of the most valued customers?”
- “Are there any particular funds that are better than others?”

Another style of question common in the set involves a comparison operation that is much higher in level and more abstract than the fundamental mathematical comparison operation discussed earlier in the section. For instance, consider the questions:

- “What other cereals are most similar to Trix?”
- “How does the Toyota RAV4 compare to the Honda CRV?”
- “Compare the distributions of values for sugar and fat in the cereals.”

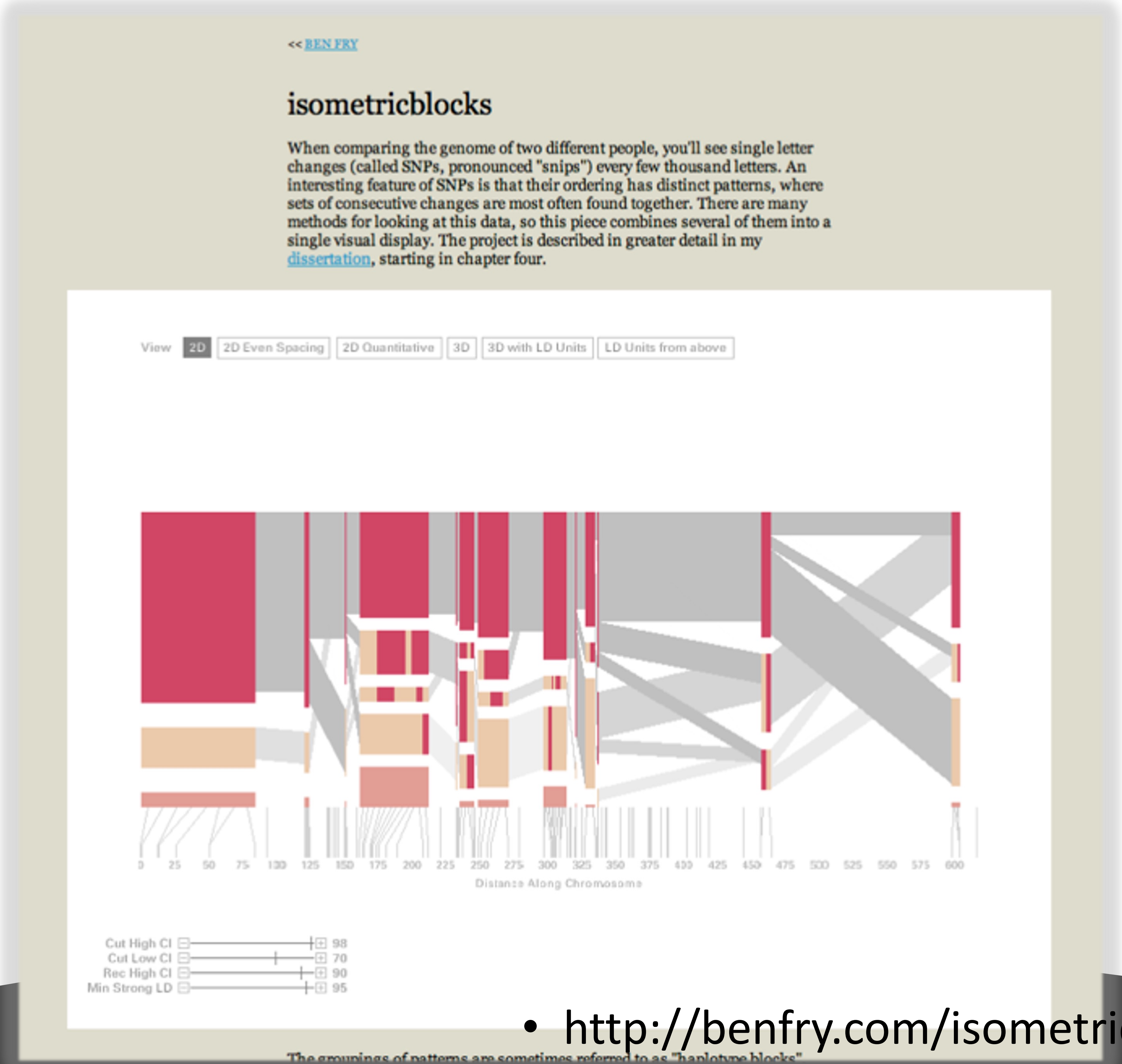
- Amar et al, ow-Level Components of Analytic Activity in Information Visualization, InfoVis 2005

Using Interaction

- Change Over Time
- Rearranging
- Selection & Highlighting
- Linking
- Filtering
- Navigation

CHANGE OVER TIME

animated transitions



<< BEN FRY

isometricblocks

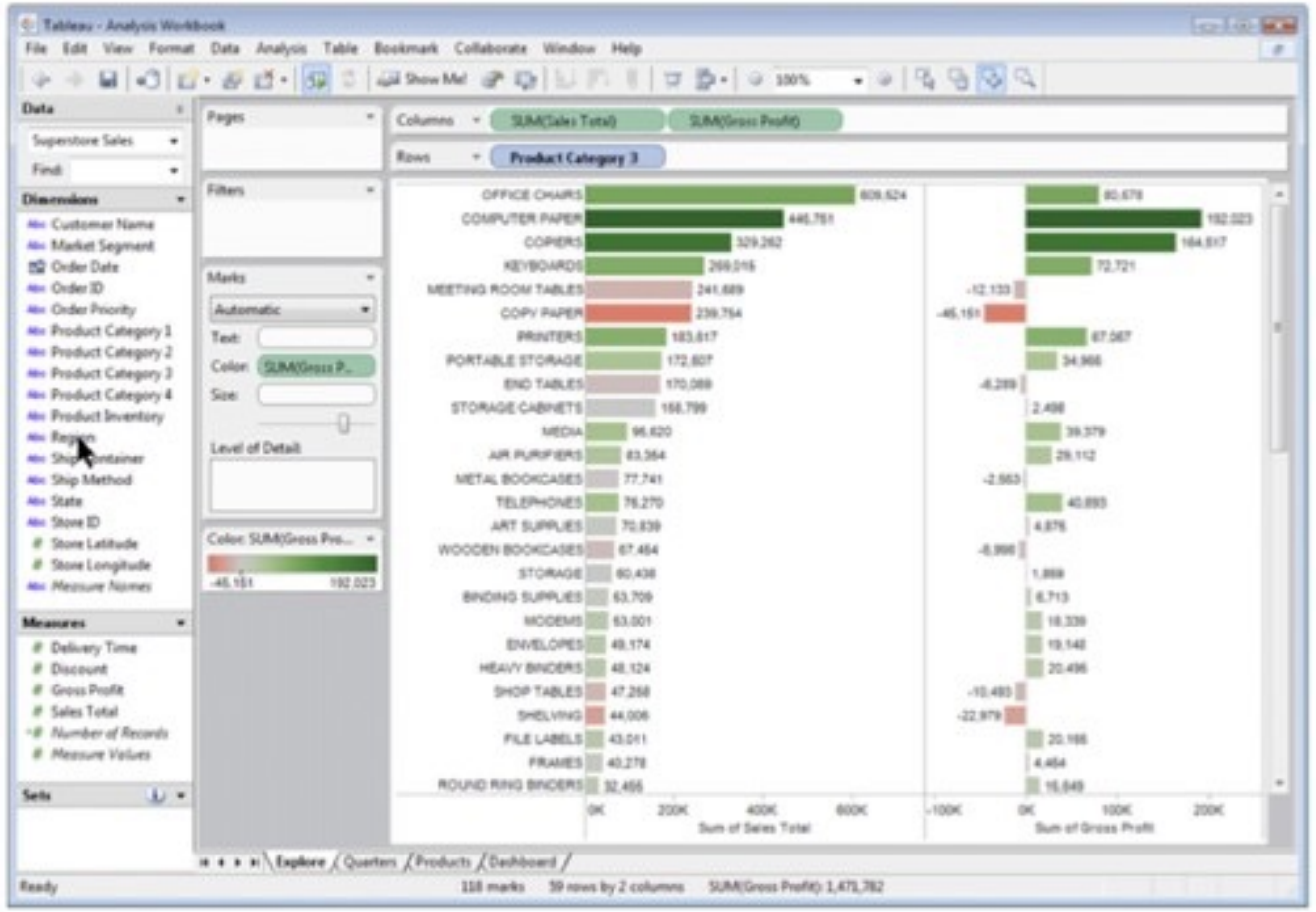
When comparing the genome of two different people, you'll see single letter changes (called SNPs, pronounced "snips") every few thousand letters. An interesting feature of SNPs is that their ordering has distinct patterns, where sets of consecutive changes are most often found together. There are many methods for looking at this data, so this piece combines several of them into a single visual display. The project is described in greater detail in my dissertation, starting in chapter four.

Cut High CI 98
Cut Low CI 70
Rec High CI 90
Min Strong LD 95

The groupings of patterns are sometimes referred to as "haplotype blocks".

- http://benfry.com/isometricblocks/

Rearranging



Columns: SUM(Sales Total) SUM(Gross Profit)

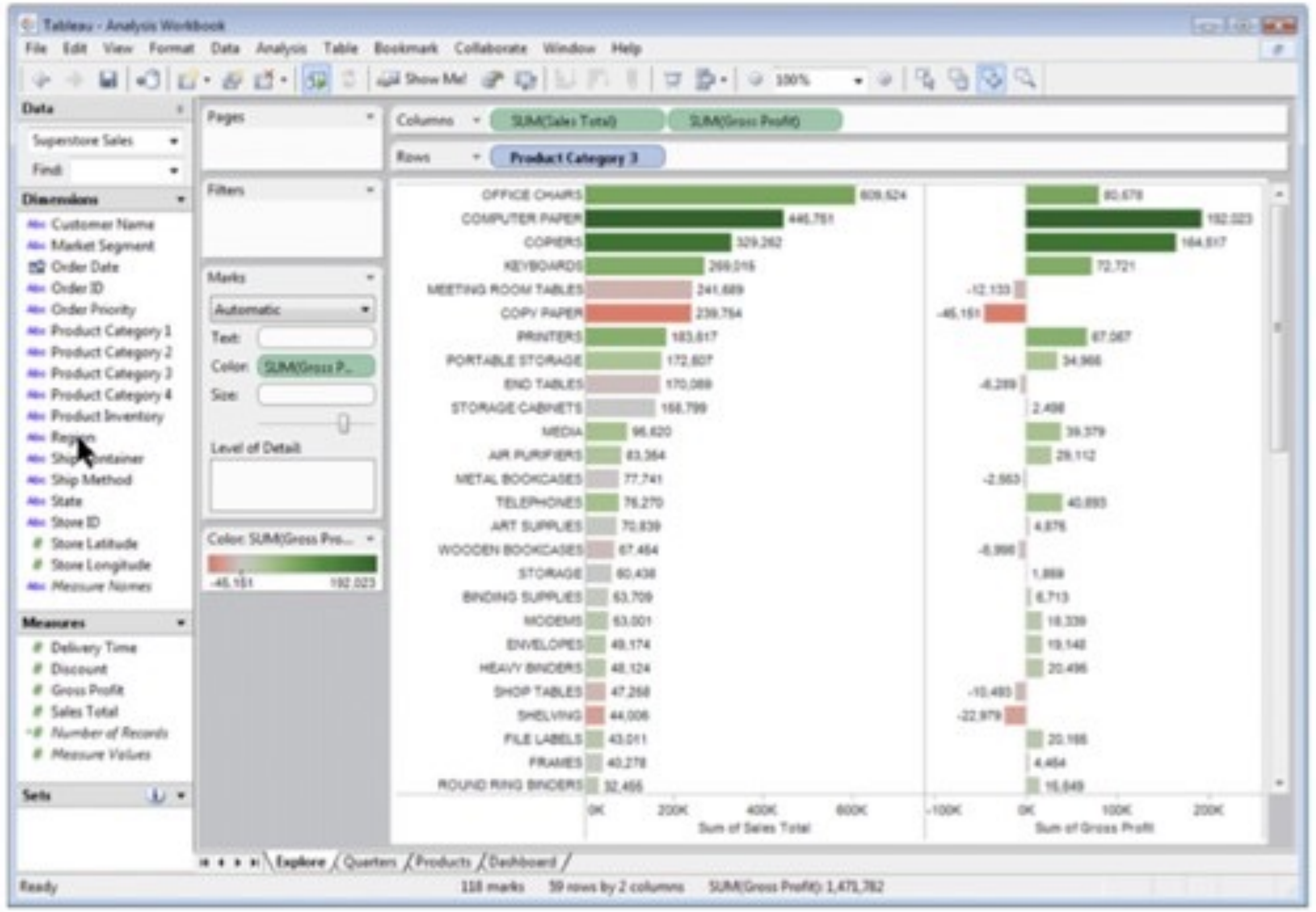
Rows: Product Category 3

Marks: Automatic

Color: SUM(Gross P...

Ready 118 marks 39 rows by 2 columns SUM(Gross Profit): 1,471,782

change encoding

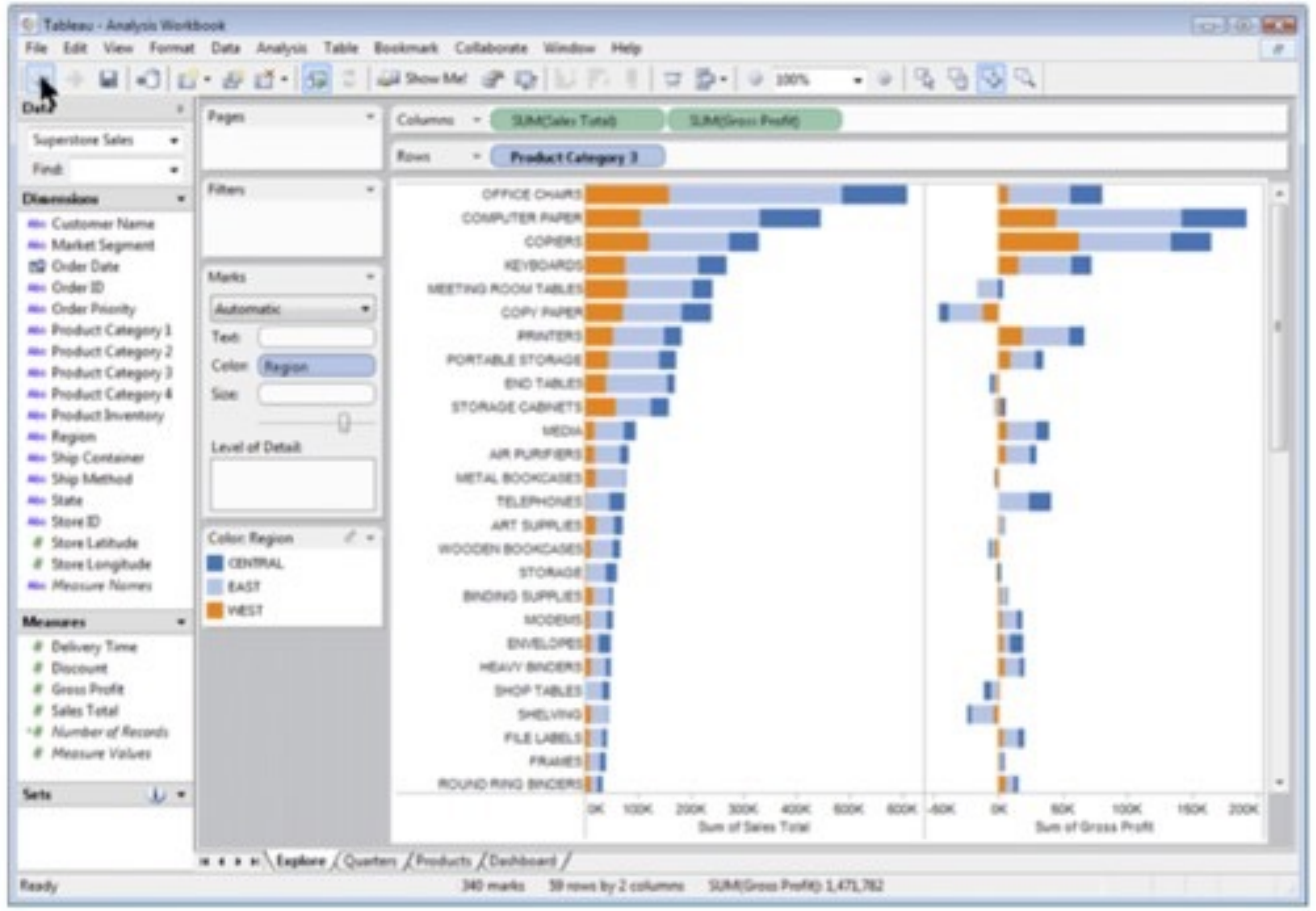


Columns: SUM(Sales Total) SUM(Gross Profit)

Rows: Product Category 3

Color: SUM(Gross P...

Ready 118 marks 59 rows by 2 columns SUM(Gross Profit): 1,471,782

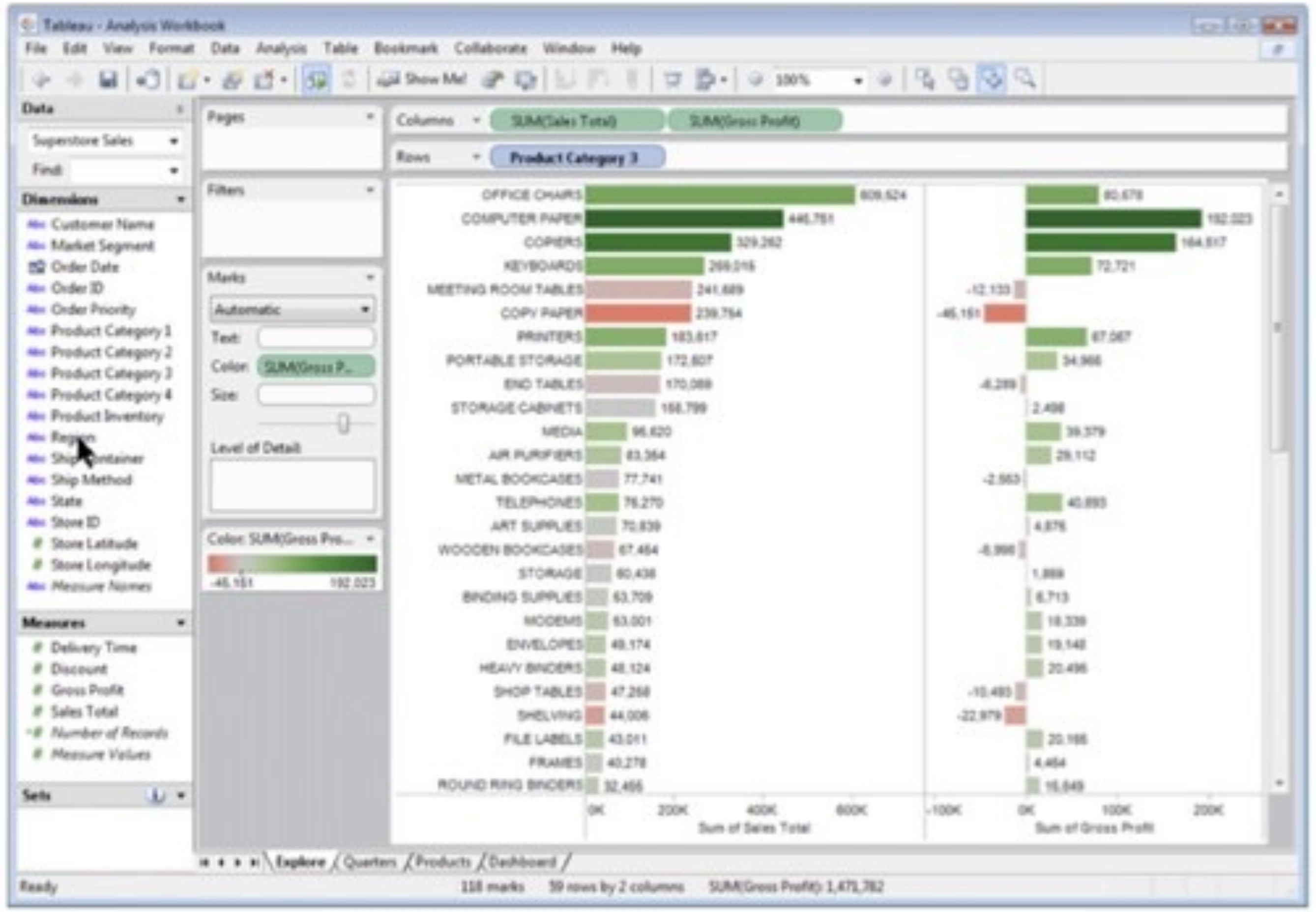


Columns: SUM(Sales Total) SUM(Gross Profit)

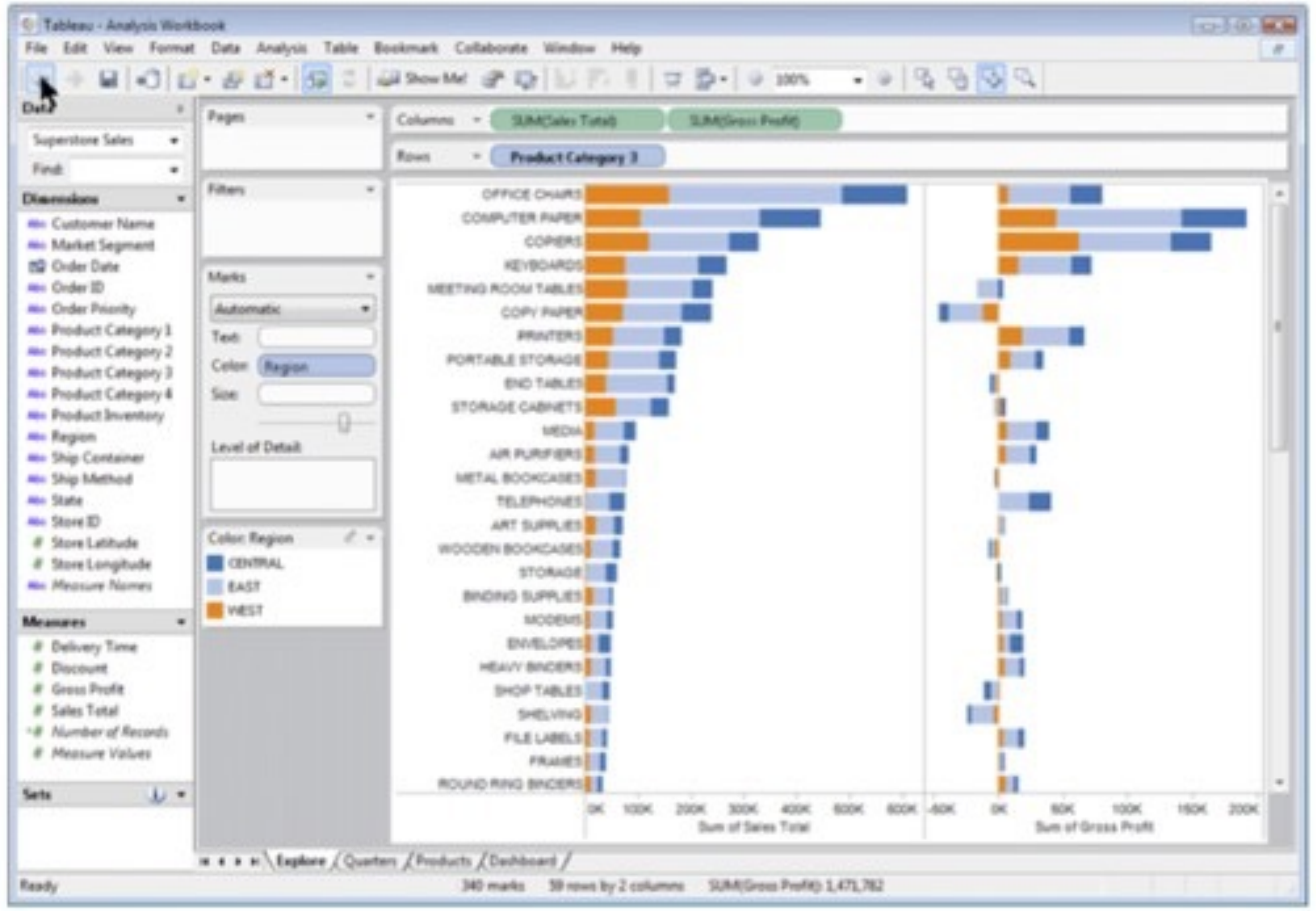
Rows: Product Category 3

Color: Region

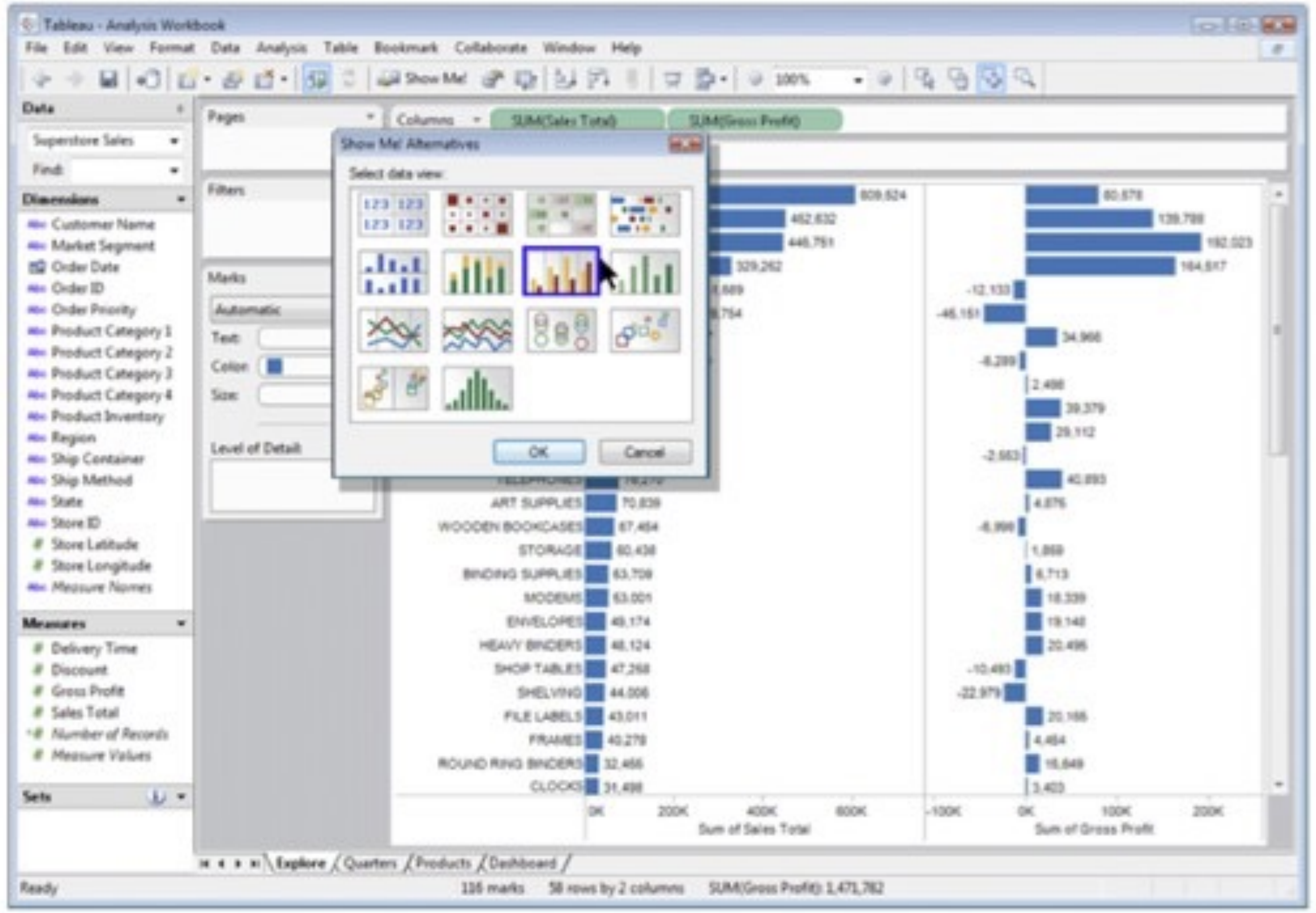
Ready 340 marks 59 rows by 2 columns SUM(Gross Profit): 1,471,782



118 marks 59 rows by 2 columns SUM(Gross Profit): 1,471,782



340 marks 59 rows by 2 columns SUM(Gross Profit): 1,471,782



118 marks 58 rows by 2 columns SUM(Gross Profit): 1,471,782

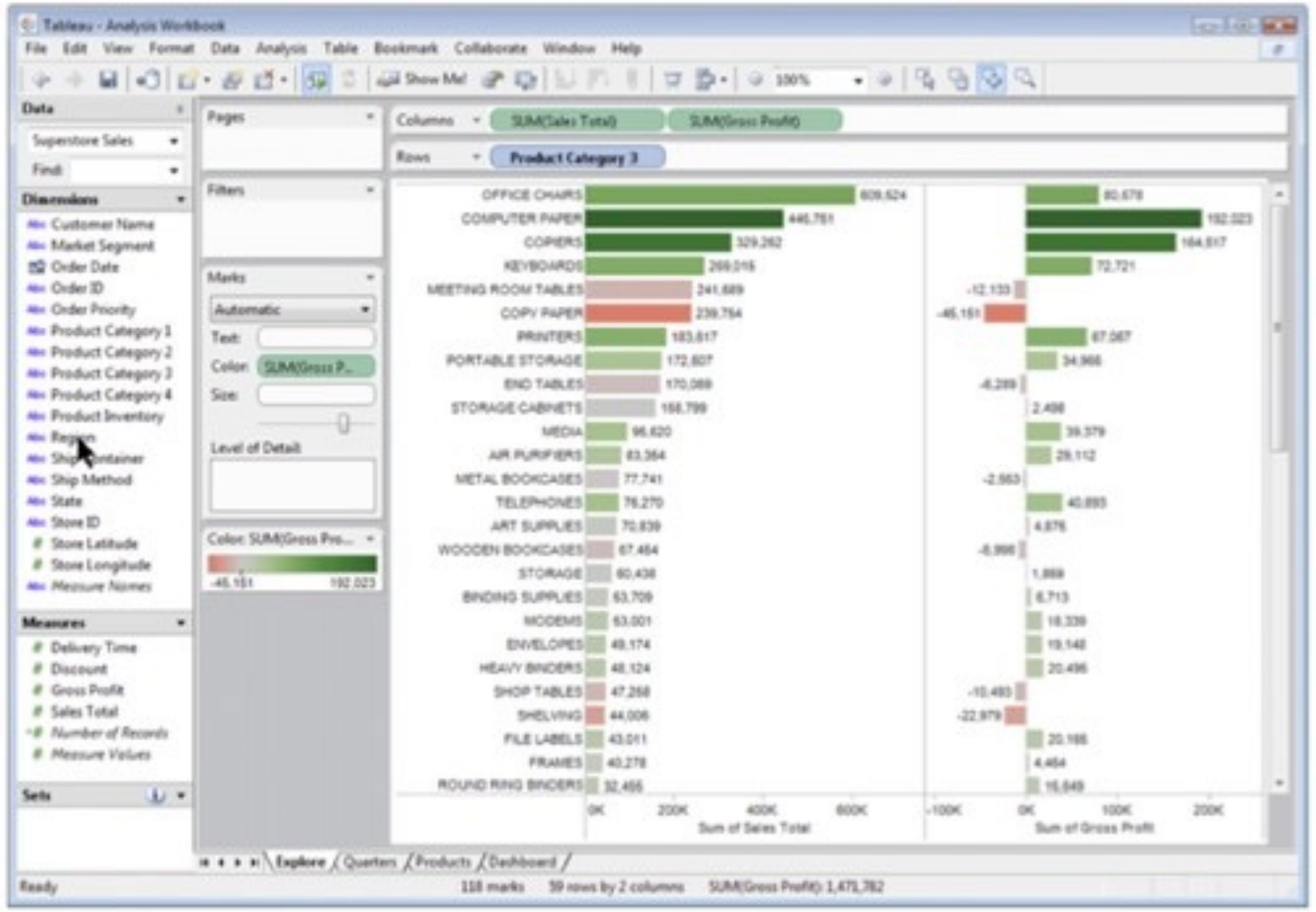


Tableau - Analysis Workbook
File Edit View Format Data Analysis Table Bookmark Collaborate Window Help
Columns SUM(Sales Total) SUM(Gross Profit)
Rows Product Category 3
Superstore Sales
Dimensions
Measures
Marks Automatic
Color SUM(Gross P...
Color: SUM(Gross Pro...
-46,151 192,023
OFFICE CHAIRS
COMPUTER PAPER
COPIERS
KEYBOARDS
MEETING ROOM TABLES
COPY PAPER
PRINTERS
PORTABLE STORAGE
END TABLES
STORAGE CABINETS
MEDIA
AIR PURIFIERS
METAL BOOKCASES
TELEPHONES
ART SUPPLIES
WOODEN BOOKCASES
STORAGE
BINDING SUPPLIES
MODEMS
ENVELOPES
HEAVY BINDERS
SHOP TABLES
SHELVING
FILE LABELS
FRAMES
ROUND RING BINDERS
Sum of Sales Total
Sum of Gross Profit
Explore Quarters Products Dashboard
Ready 118 marks 59 rows by 2 columns SUM(Gross Profit): 1,471,782

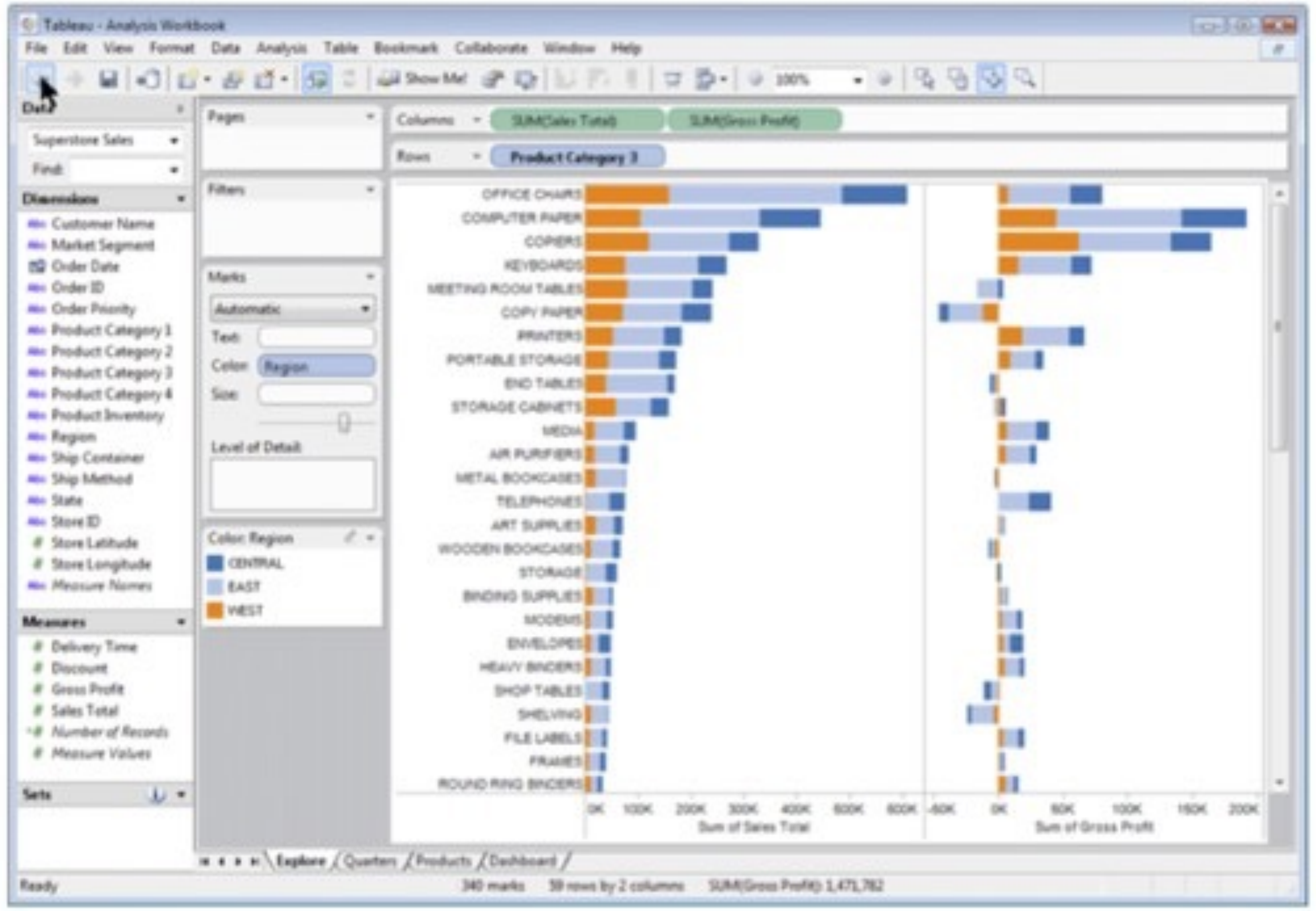


Tableau - Analysis Workbook
Columns SUM(Sales Total) SUM(Gross Profit)
Rows Product Category 3
Color Region
Color: Region
CENTRAL
EAST
WEST
Sum of Sales Total
Sum of Gross Profit
Explore Quarters Products Dashboard
Ready 340 marks 59 rows by 2 columns SUM(Gross Profit): 1,471,782

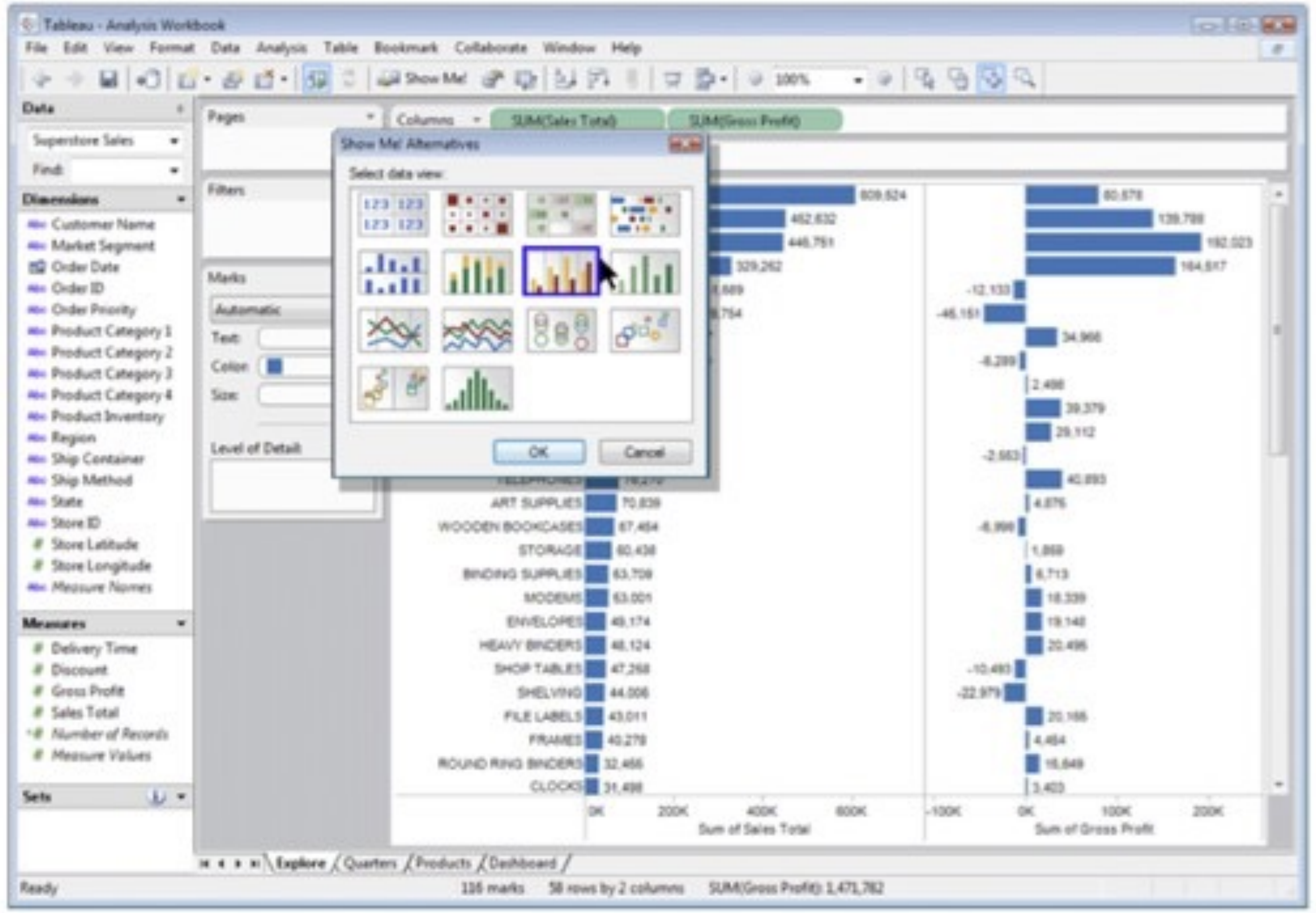


Tableau - Analysis Workbook
Show Me! Alternatives
Select data view:
OK
Cancel
Columns SUM(Sales Total) SUM(Gross Profit)
Sum of Sales Total
Sum of Gross Profit
Explore Quarters Products Dashboard
Ready 118 marks 58 rows by 2 columns SUM(Gross Profit): 1,471,782

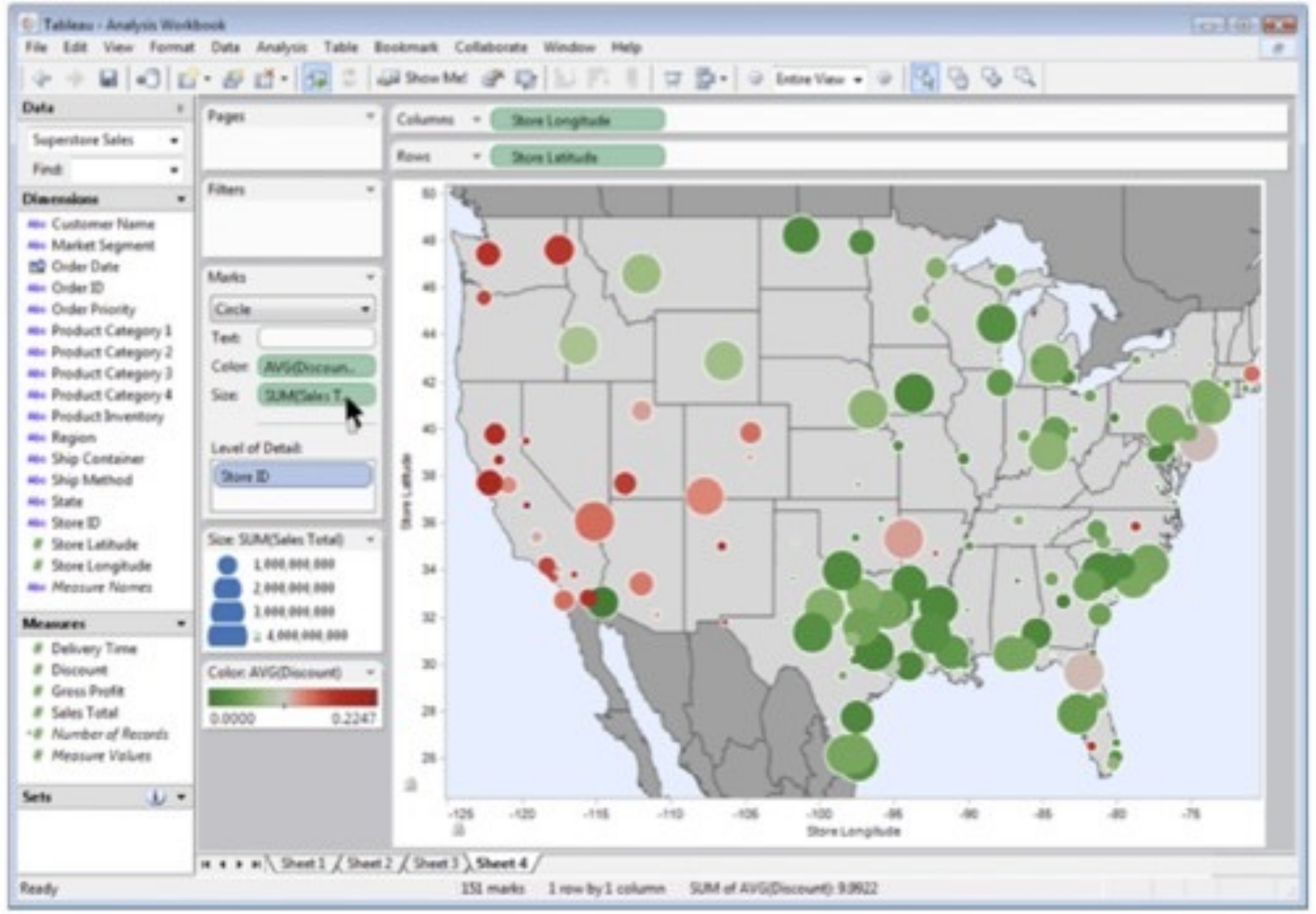
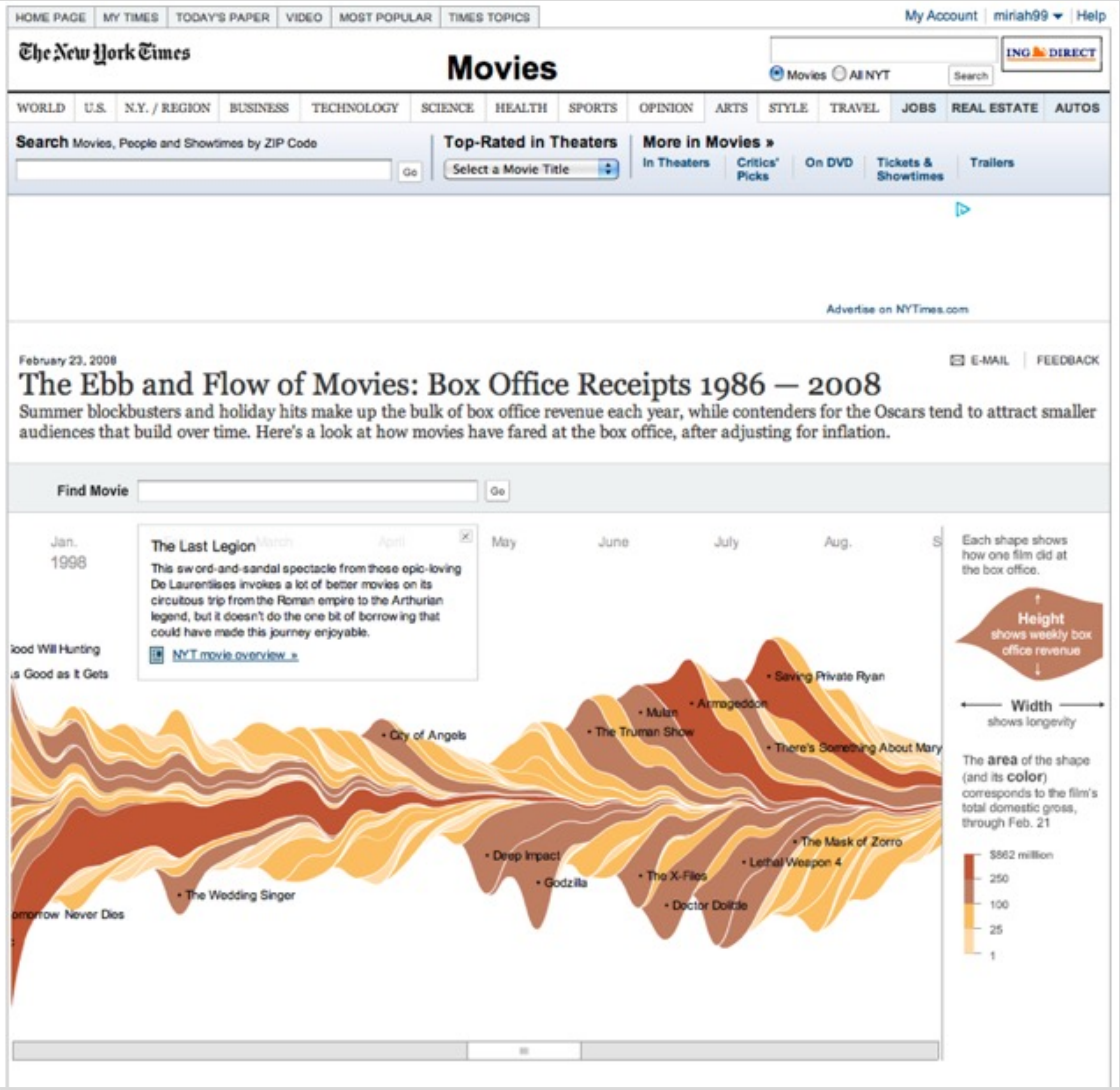


Tableau - Analysis Workbook
Columns Store Longitude
Rows Store Latitude
Marks Circle
Color AVG(Discoun...
Size SUM(Sales T...
Level of Detail
Store ID
Size: SUM(Sales Total)
Color: AVG(Discount)
0.0000 0.2247
Store Latitude
Store Longitude
Sheet 1 Sheet 2 Sheet 3 Sheet 4
Ready 151 marks 1 row by 1 column SUM of AVG(Discount): 9.9922

SELECTION & HIGHLIGHTING



February 23, 2008

The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008

Summer blockbusters and holiday hits make up the bulk of box office revenue each year, while contenders for the Oscars tend to attract smaller audiences that build over time. Here's a look at how movies have fared at the box office, after adjusting for inflation.

The Last Legion

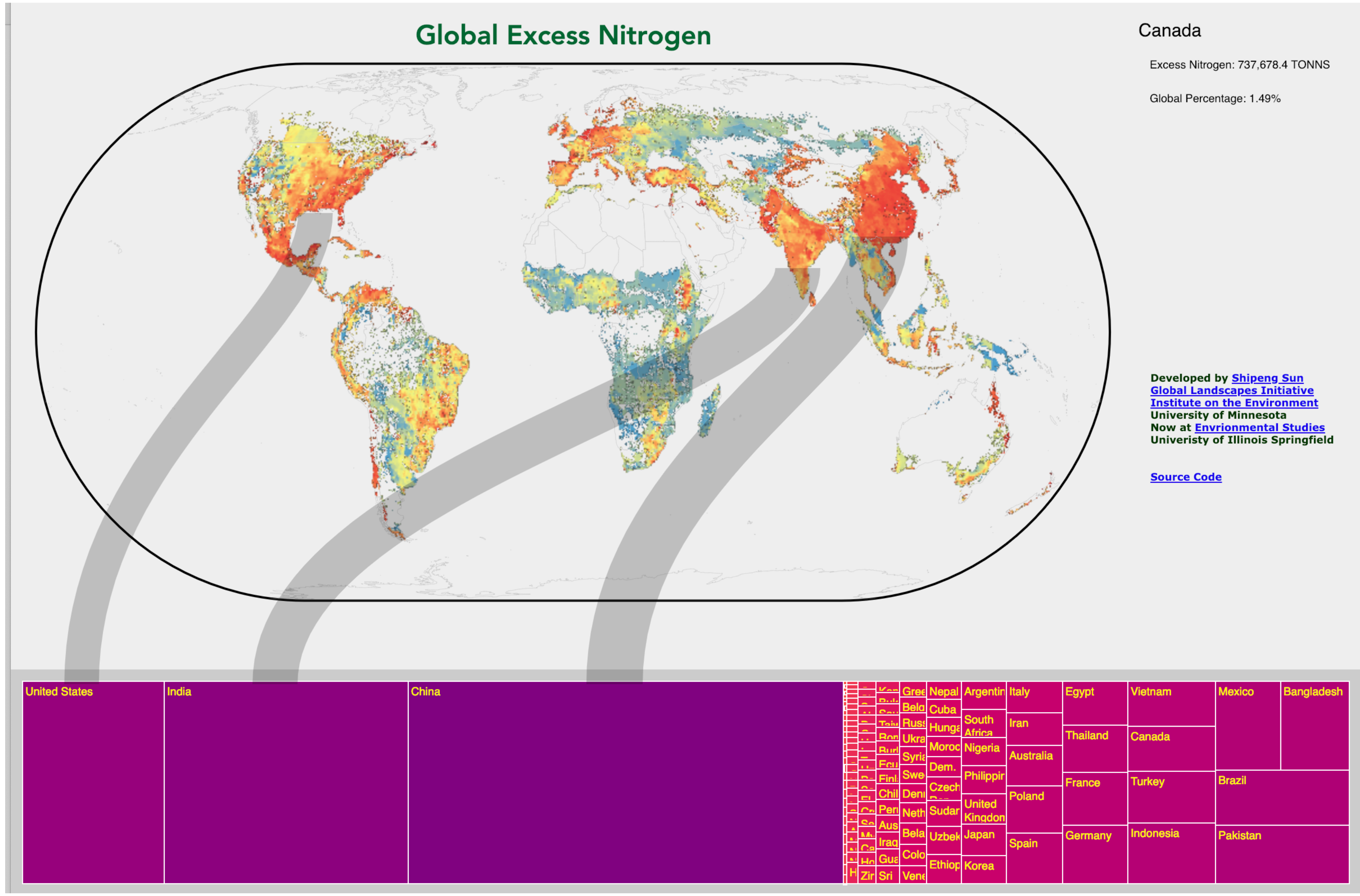
This sword-and-sandal spectacle from those epic-loving De Laurentiises invokes a lot of better movies on its circuitous trip from the Roman empire to the Arthurian legend, but it doesn't do the one bit of borrowing that could have made this journey enjoyable.

Each shape shows how one film did at the box office.

The **area** of the shape (and its **color**) corresponds to the film's total domestic gross, through Feb. 21

http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html

Linking



Global Excess Nitrogen

Canada

Excess Nitrogen: 737,678.4 TONNS

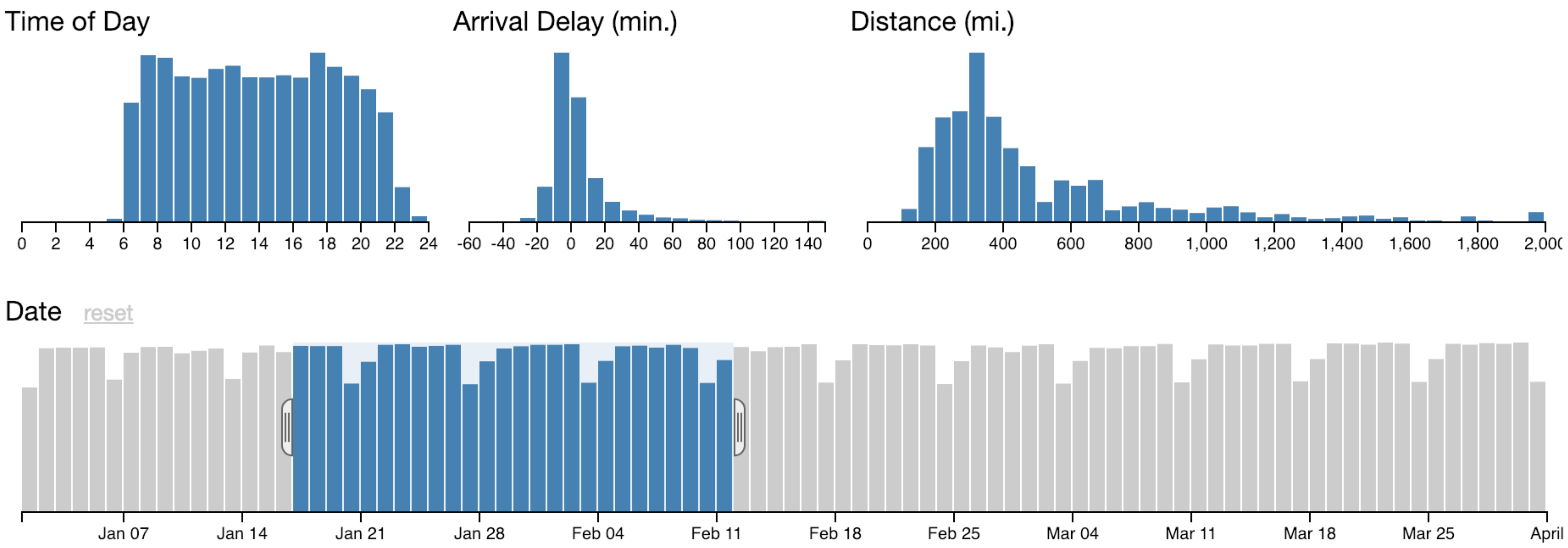
Global Percentage: 1.49%

Developed by Shipeng Sun
Global Landscapes Initiative
Institute on the Environment
University of Minnesota
Now at Envrionmental Studies
Univeristy of Illinois Springfield

Source Code

http://sunsp.net/demo/GeogTreeMaps/

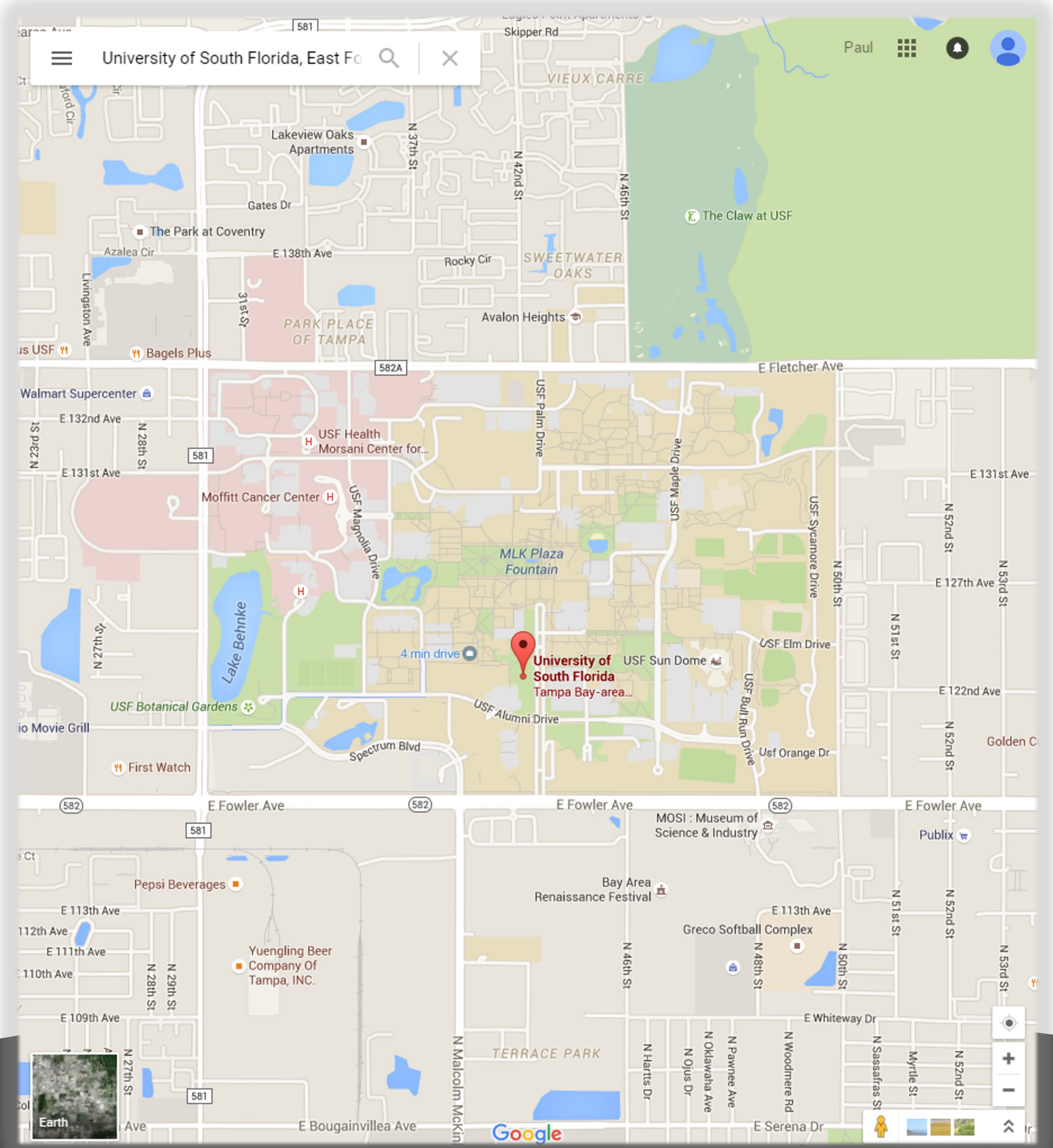
Filtering



http://square.github.io/crossfilter/

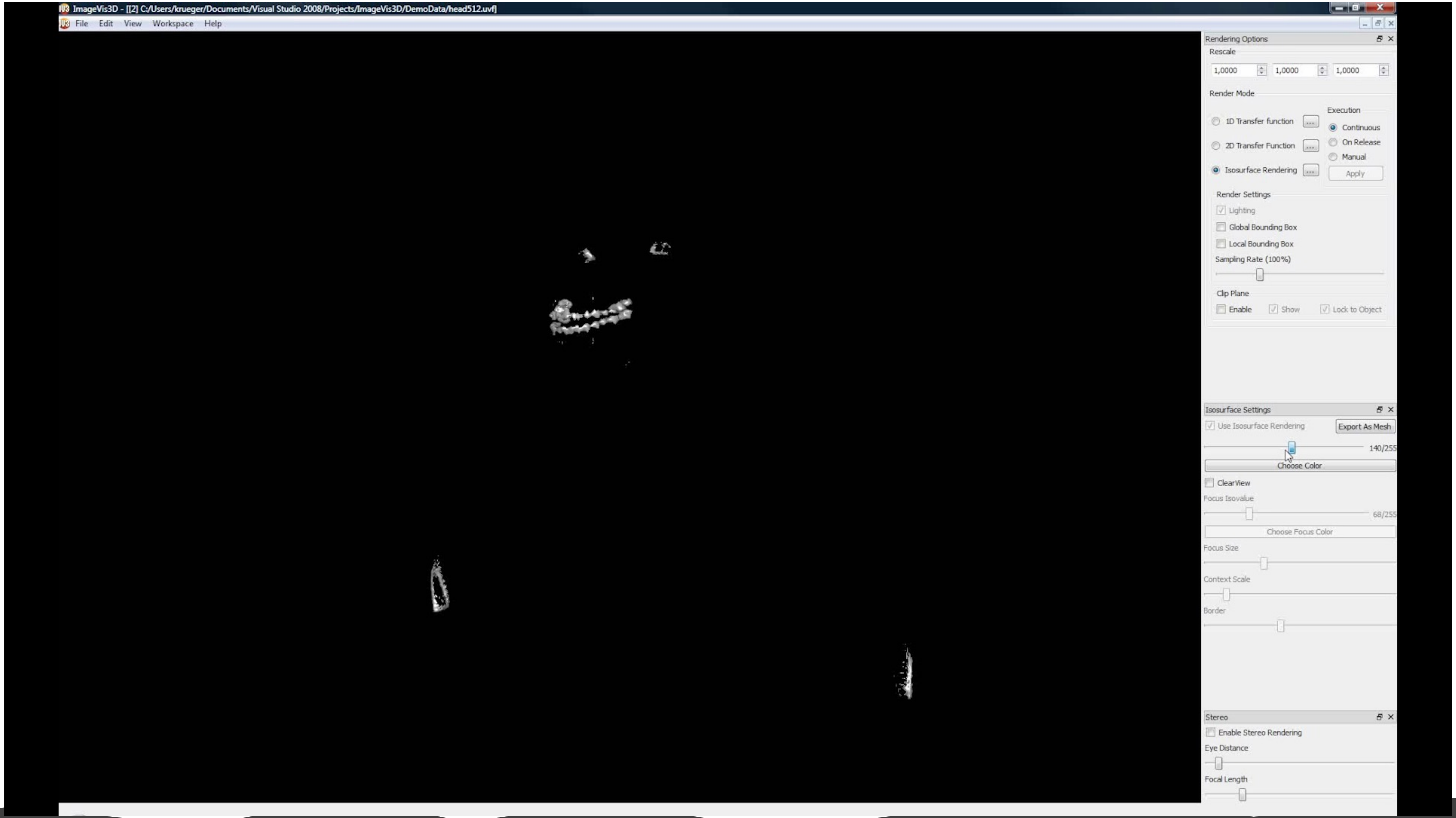
NAVIGATION

pan (and translate)



University of South Florida, East Fo

rotate



ImageVis3D - [[2] C:/Users/krueger/Documents/Visual Studio 2008/Projects/ImageVis3D/DemoData/head512.uvf]

Rendering Options

Rescale

1,0000 1,0000 1,0000

Render Mode

○ 1D Transfer function

○ 2D Transfer Function

● Isosurface Rendering

Execution

● Continuous

○ On Release

○ Manual

Render Settings

☑ Lighting

□ Global Bounding Box

□ Local Bounding Box

Sampling Rate (100%)

Clip Plane

□ Enable ☑ Show ☑ Lock to Object

Isosurface Settings

☑ Use Isosurface Rendering

140/255

□ ClearView

Focus Isovalue

68/255

Focus Size

Context Scale

Border

Stereo

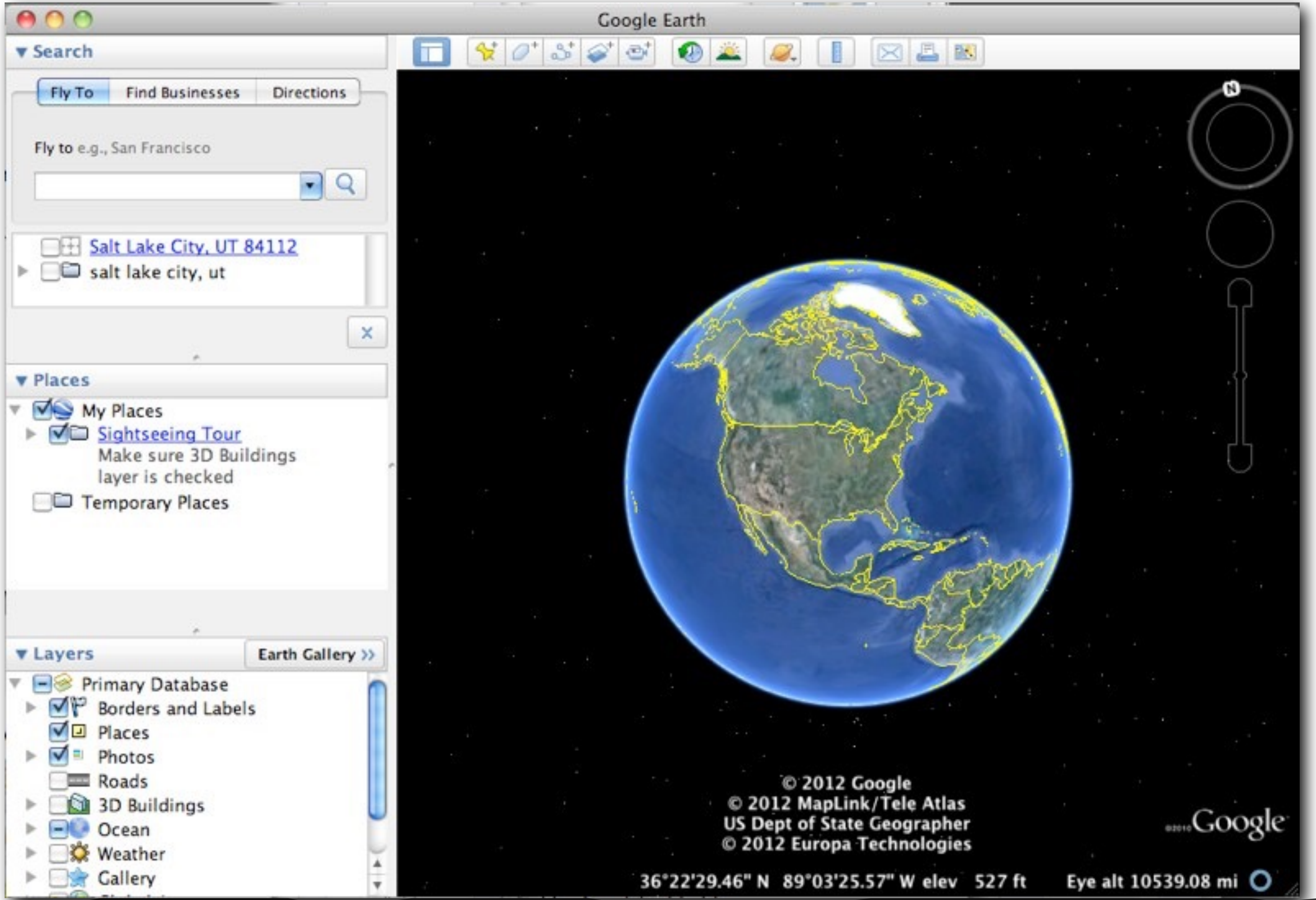
□ Enable Stereo Rendering

Eye Distance

Focal Length

GEOMETRIC vs SEMANTIC ZOOMING

geometric



□ Salt Lake City, UT 84112
□ salt lake city, ut

☑ My Places
☑ Sightseeing Tour
Make sure 3D Buildings layer is checked
□ Temporary Places

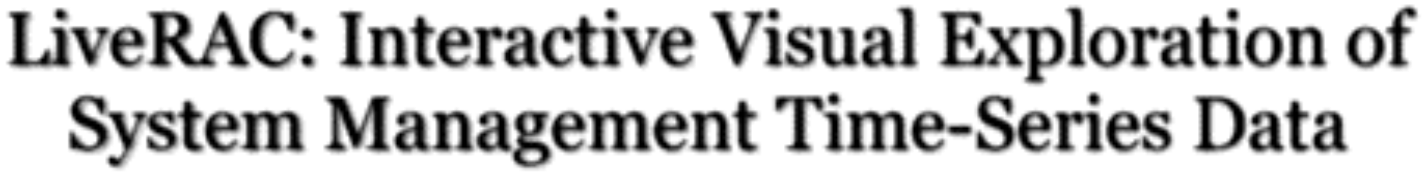
⊟ Primary Database
☑ Borders and Labels
☑ Places
☑ Photos
□ Roads
□ 3D Buildings
⊟ Ocean
□ Weather
□ Gallery

© 2012 Google
© 2012 MapLink/Tele Atlas
US Dept of State Geographer
© 2012 Europa Technologies

36°22'29.46" N 89°03'25.57" W elev 527 ft Eye alt 10539.08 mi

semantic

semantic



LiveRAC: Interactive Visual Exploration of
System Management Time-Series Data

SCI