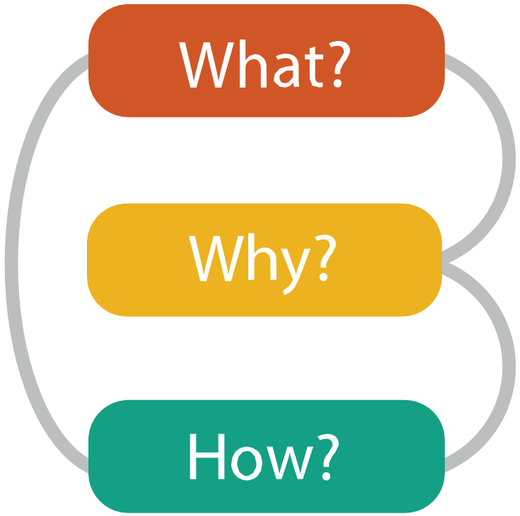
Paul Rosen
paul.rosen@utah.edu
@paulrosenphd
https://cspaul.com

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

TASKS & INTERACTION

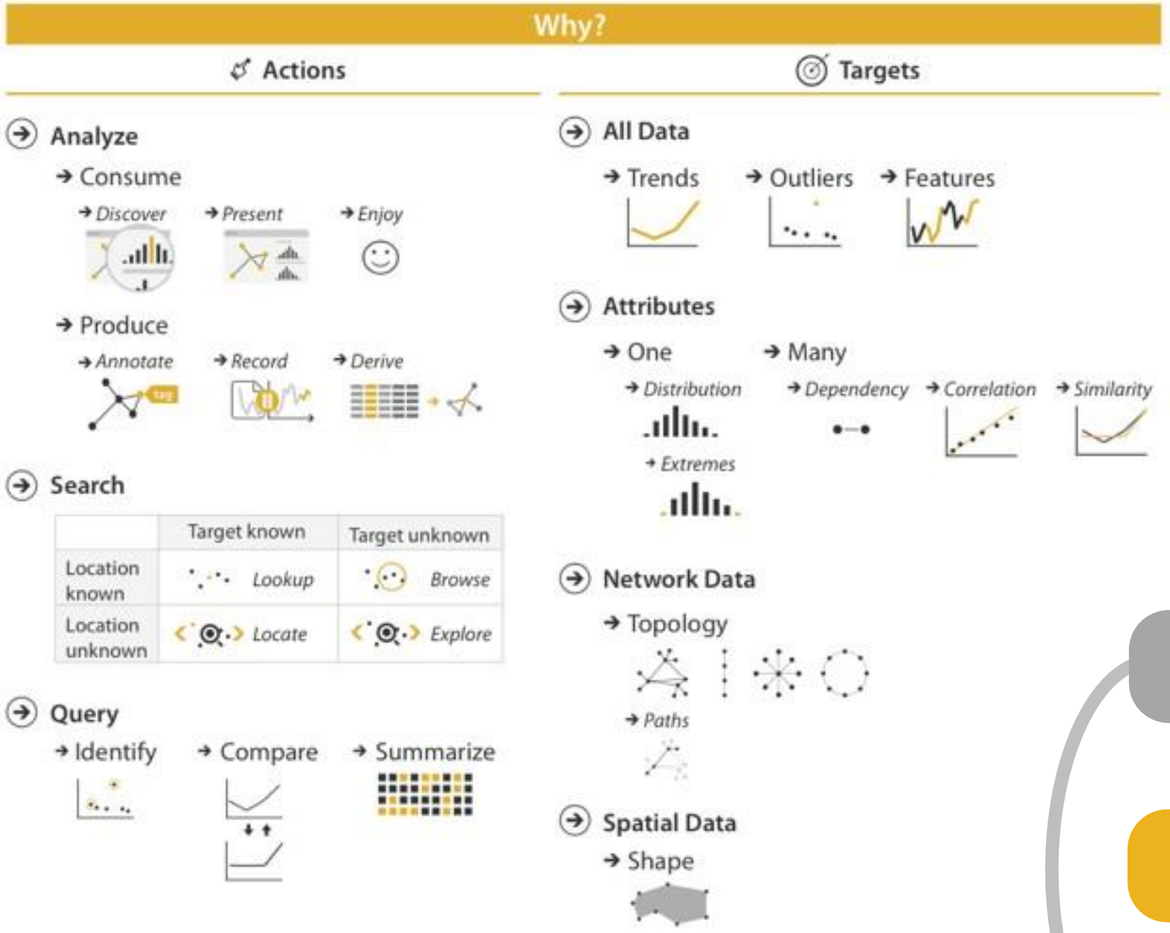
High-level Task Taxonomy: what, why, and how

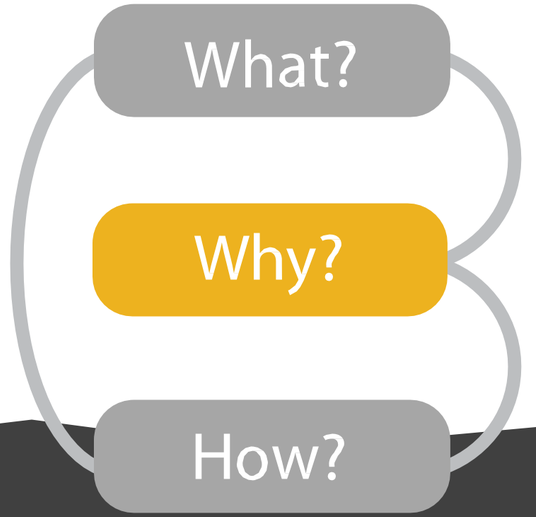
- what is shown?
- why is the user looking at it?
- how is it shown?



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

task abstraction





{action, *target*} pairs

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

Analyze

{**action**, target}

Search

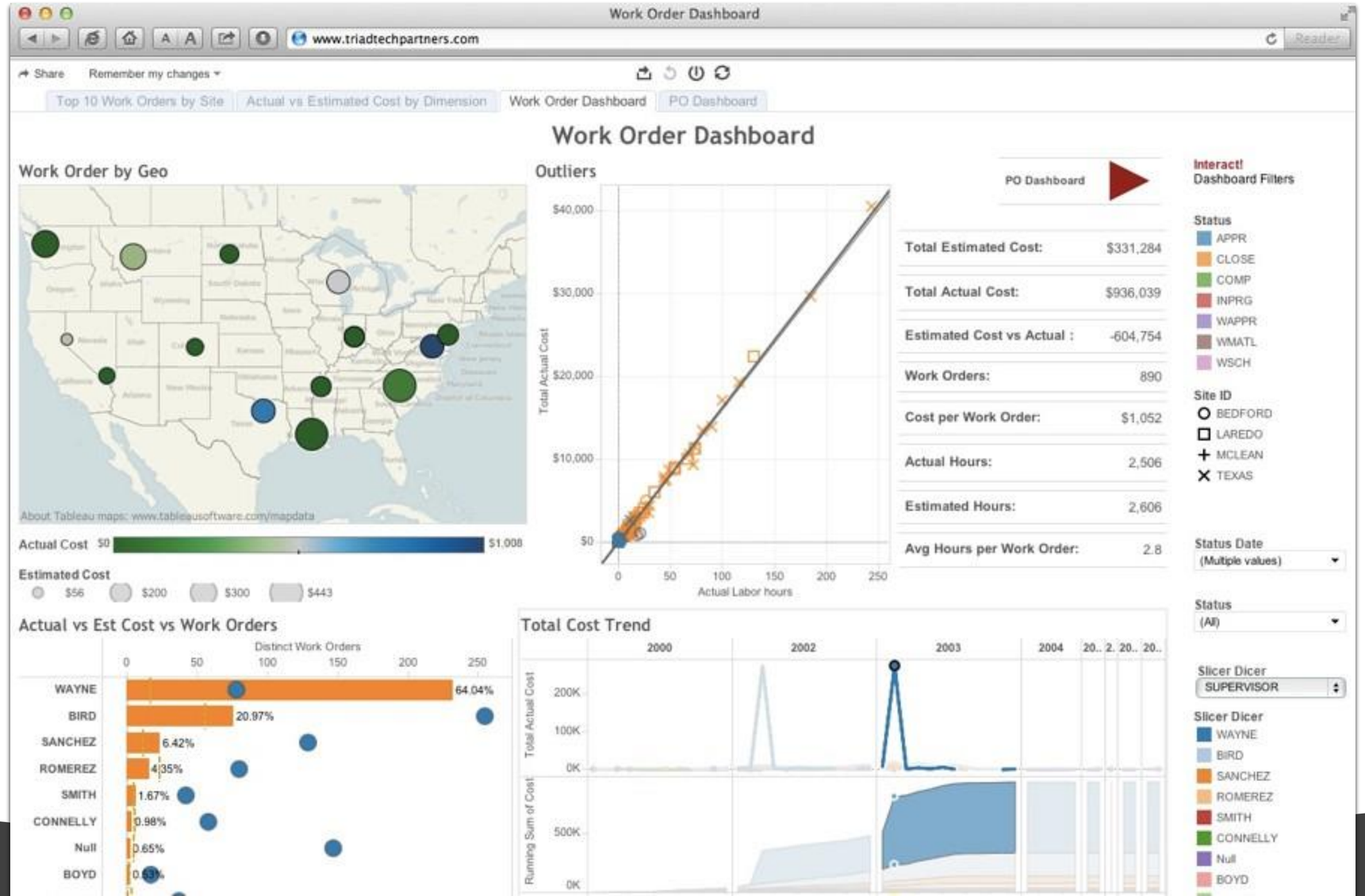
Query

{**action**, target}

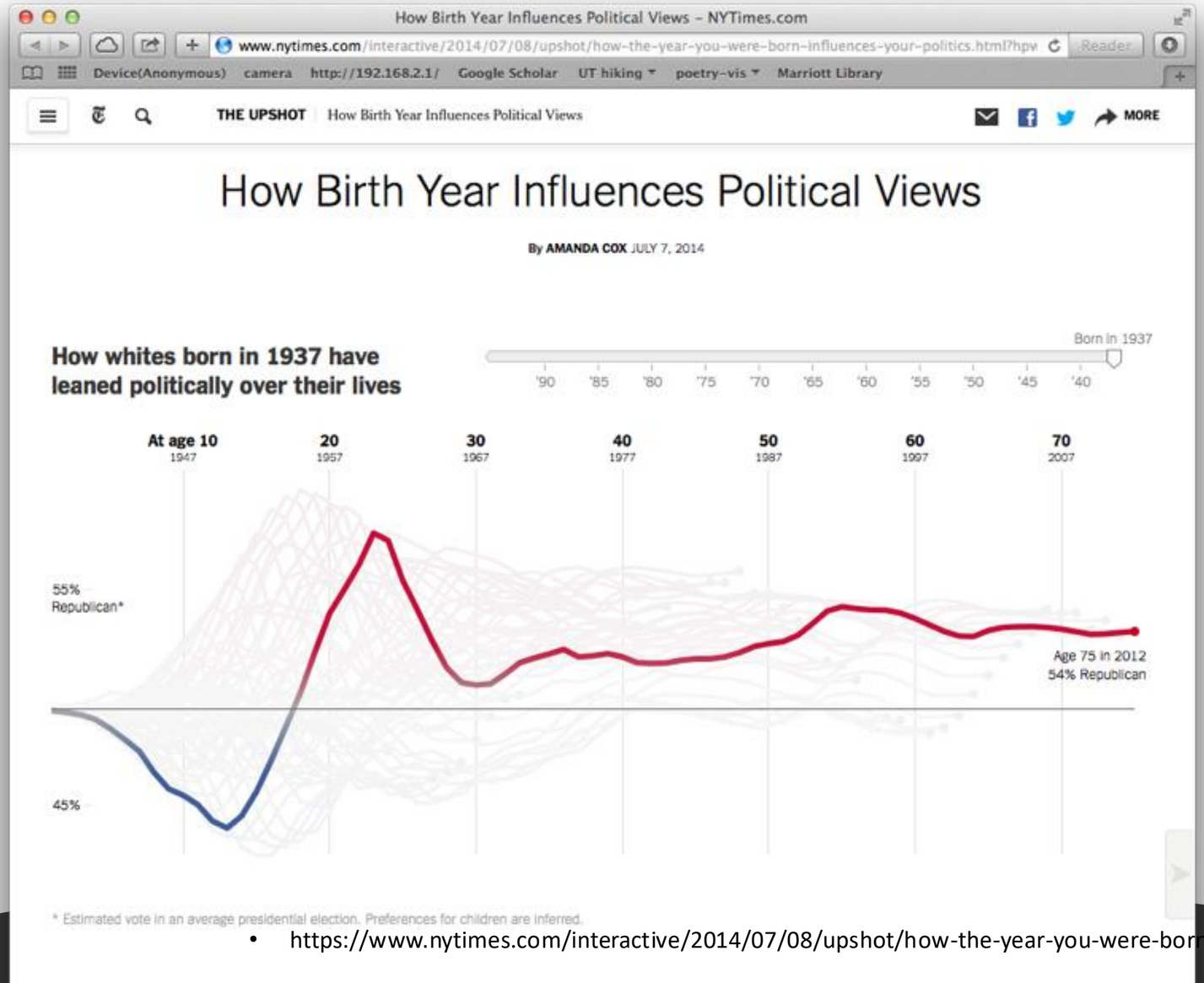
- **Analyze**
  - Consume
    - *Discover*
    - *Present*
    - *Enjoy*

- **Search**
- **Query**

discover


Work Order Dashboard
www.triadtechpartners.com
Share
Remember my changes
Top 10 Work Orders by Site
Actual vs Estimated Cost by Dimension
Work Order Dashboard
PO Dashboard
Work Order Dashboard
Work Order by Geo
About Tableau maps: www.tableausoftware.com/mapdata
Actual Cost $0 $1,008
Estimated Cost
$56 $200 $300 $443
Outliers
$40,000
$30,000
$20,000
$10,000
$0
Total Actual Cost
0 50 100 150 200 250
Actual Labor hours
PO 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
Interact!
Dashboard Filters
Status
APPR
CLOSE
COMP
INPRG
WAPPR
WMATL
WSCH
Site ID
BEDFORD
LAREDO
MCLEAN
TEXAS
Status Date
(Multiple values)
Status
(All)
Slicer Dicer
SUPERVISOR
Slicer Dicer
WAYNE
BIRD
SANCHEZ
ROMEREZ
SMITH
CONNELLY
Null
BOYD
Actual vs Est Cost vs Work Orders
Distinct Work Orders
0 50 100 150 200 250
WAYNE 64.04%
BIRD 20.97%
SANCHEZ 6.42%
ROMEREZ 4.35%
SMITH 1.67%
CONNELLY 0.98%
Null 0.65%
BOYD
Total Cost Trend
2000 2002 2003 2004 20.. 2.. 20.. 20..
Total Actual Cost
200K
100K
0K
Running Sum of Cost
500K
0K

present



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

enjoy

- http://hint.fm/wind/

{**action**, target}

Analyze

- Consume
  - *Discover*
  - *Present*
  - *Enjoy*

- Produce
  - *Annotate*
  - *Record*
  - *Derive*

Search

Query

annotate & record

{**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}

Analyze

Consume

- *Discover*
- *Present*
- *Enjoy*

Produce

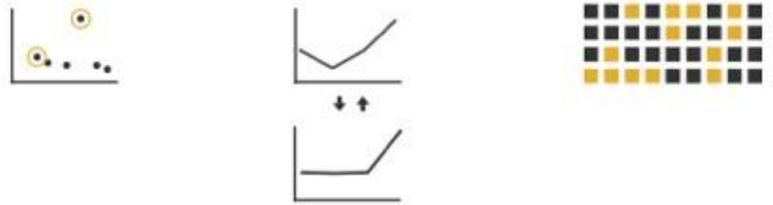
- *Annotate*
- *Record*
- *Derive*

Search

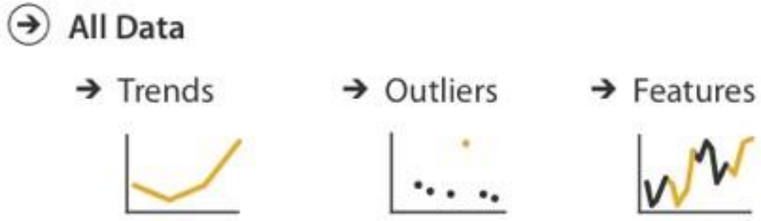
	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

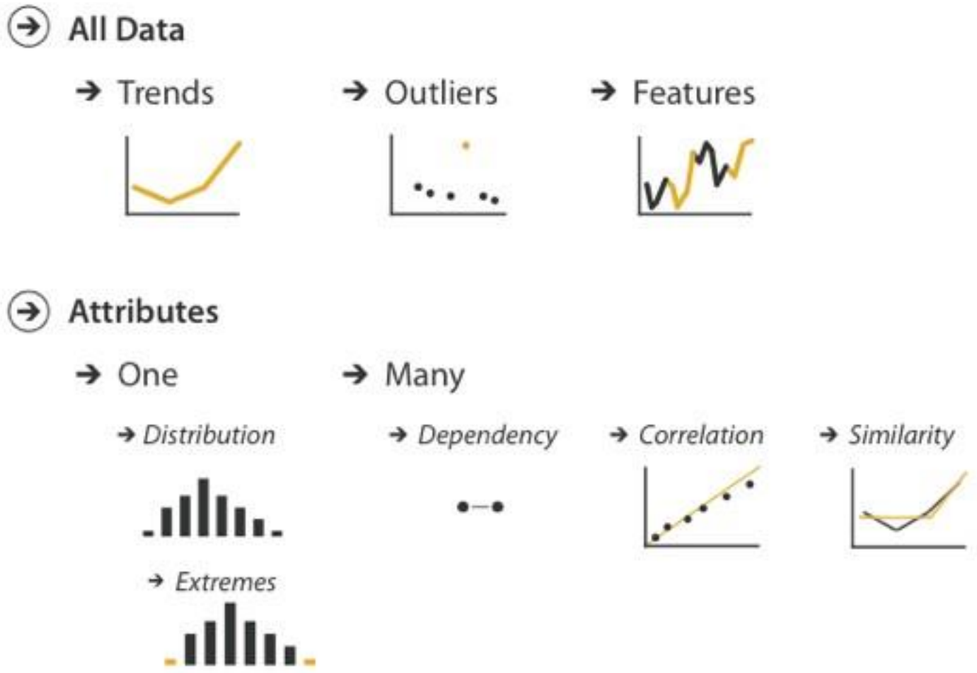
- Identify
- Compare
- Summarize



{action, **target**}



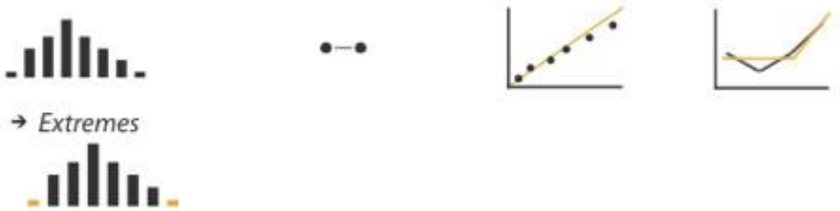
{action, **target**}



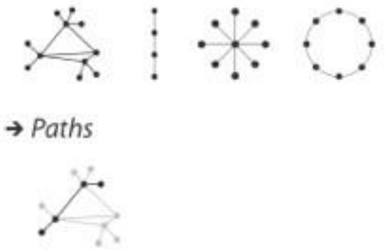
{action, **target**}

- All Data
  - Trends
  - Outliers
  - Features

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



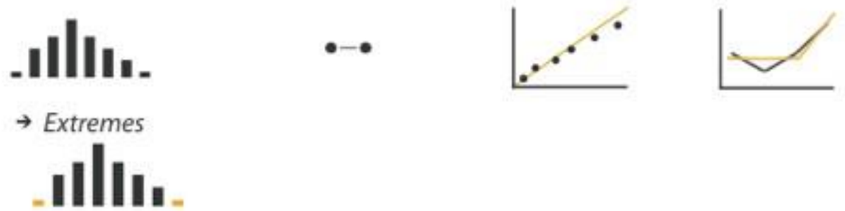
- Network Data
  - Topology
  - *Paths*



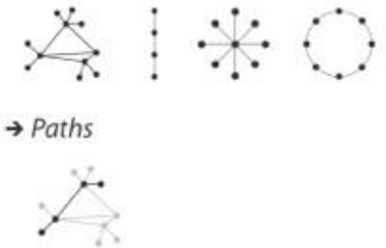
{action, **target**}

- **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

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?

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?

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?

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?

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?

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."

Using Interaction

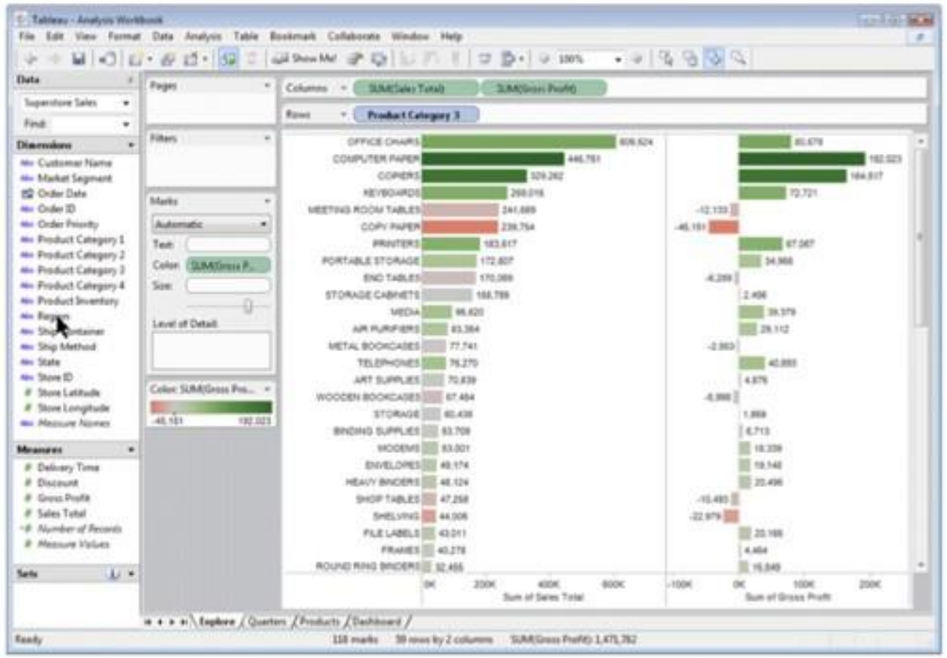
- Change Over Time
- Rearranging
- Selection & Highlighting
- Linking
- Filtering
- Navigation

CHANGE OVER TIME

animated transitions

- http://benfry.com/isometricblocks/

Rearranging



change encoding

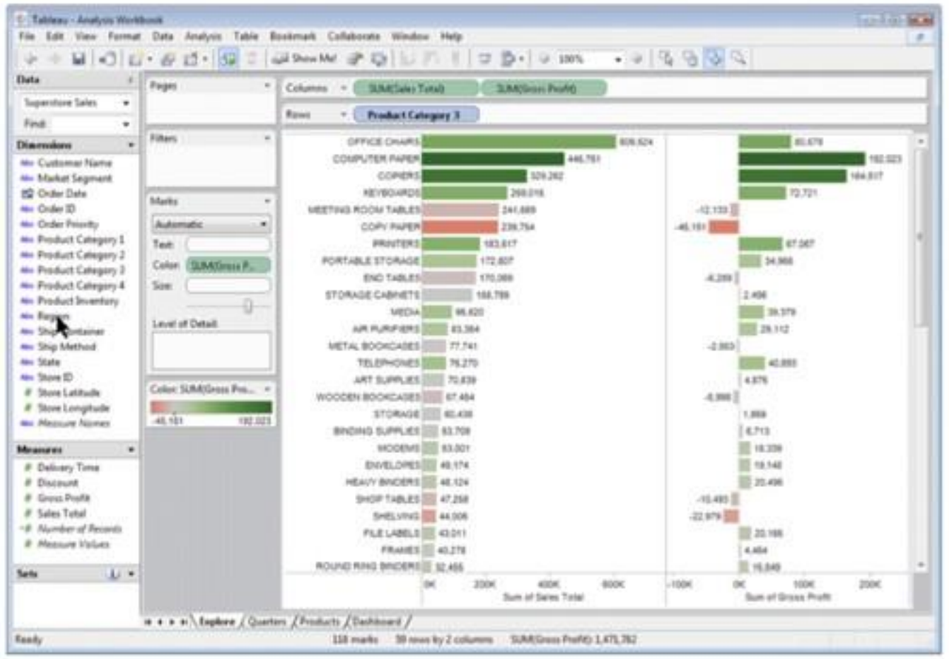


Tableau - Analysis Workbook
File Edit View Format Data Analysis Table Bookmark Collaborate Window Help
Data
Superstore Sales
Dimensions
Customer Name
Market Segment
Order Date
Order ID
Order Priority
Product Category 1
Product Category 2
Product Category 3
Product Category 4
Product Inventory
Region
Ship Container
Ship Method
State
Store ID
Store Latitude
Store Longitude
Measure Names
Measures
Delivery Time
Discount
Gross Profit
Sales Total
Number of Records
Measure Values
Sets
Pages
Filters
Marks
Automatic
Text
Color SUM(Gross P...
Size
Level of Detail
Color: SUM(Gross Pro...
-46,151
192,023
Columns SUM(Sales Total) SUM(Gross Profit)
Rows Product Category 3
OFFICE CHAIRS 608,624 80,679
COMPUTER PAPER 446,751 192,023
COPIERS 329,282 164,817
KEYBOARDS 268,016 72,721
MEETING ROOM TABLES 241,689 -12,133
COPY PAPER 239,754 -46,151
PRINTERS 183,617 67,267
PORTABLE STORAGE 172,607 34,966
END TABLES 170,069 -4,299
STORAGE CABINETS 168,799 2,498
MEDIA 98,620 39,579
AIR PURIFIERS 83,384 29,112
METAL BOOKCASES 77,741 -2,883
TELEPHONES 76,270 40,893
ART SUPPLIES 70,839 4,875
WOODEN BOOKCASES 67,484 -8,998
STORAGE 60,439 1,869
BINDING SUPPLIES 53,708 6,713
MODEMS 53,001 18,339
ENVELOPES 49,174 19,148
HEAVY BINDERS 48,124 20,496
SHOP TABLES 47,258 -15,485
SHELVING 44,006 -22,979
FILE LABELS 43,011 20,166
FRAMES 40,278 4,464
ROUND RING BINDERS 32,455 15,849
0K 200K 400K 600K
Sum of Sales Total
-100K 0K 100K 200K
Sum of Gross Profit
Explore Quarters Products Dashboard
Ready
118 marks 39 rows by 2 columns SUM(Gross Profit): 1,471,782

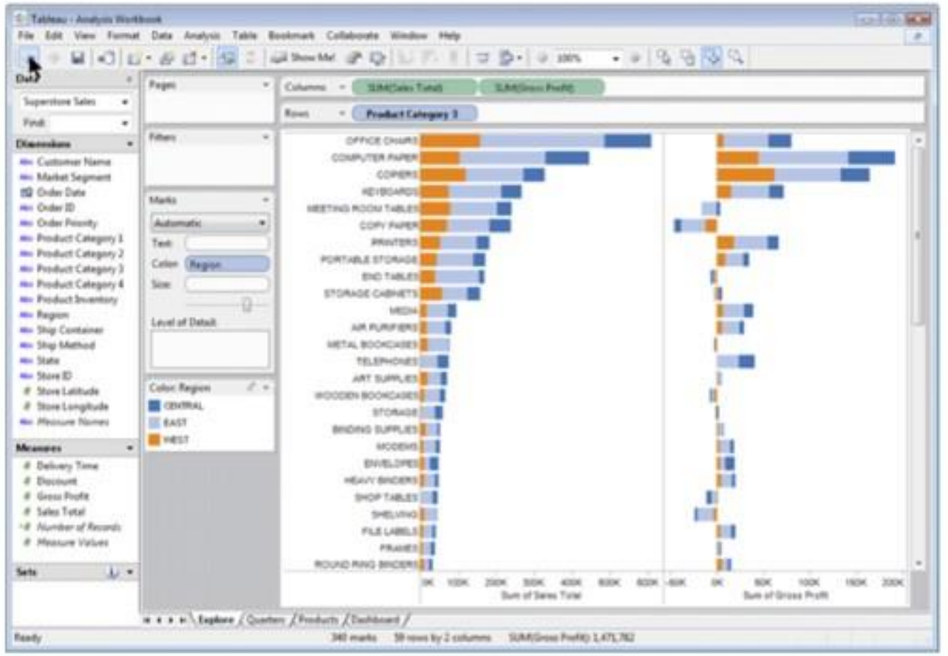
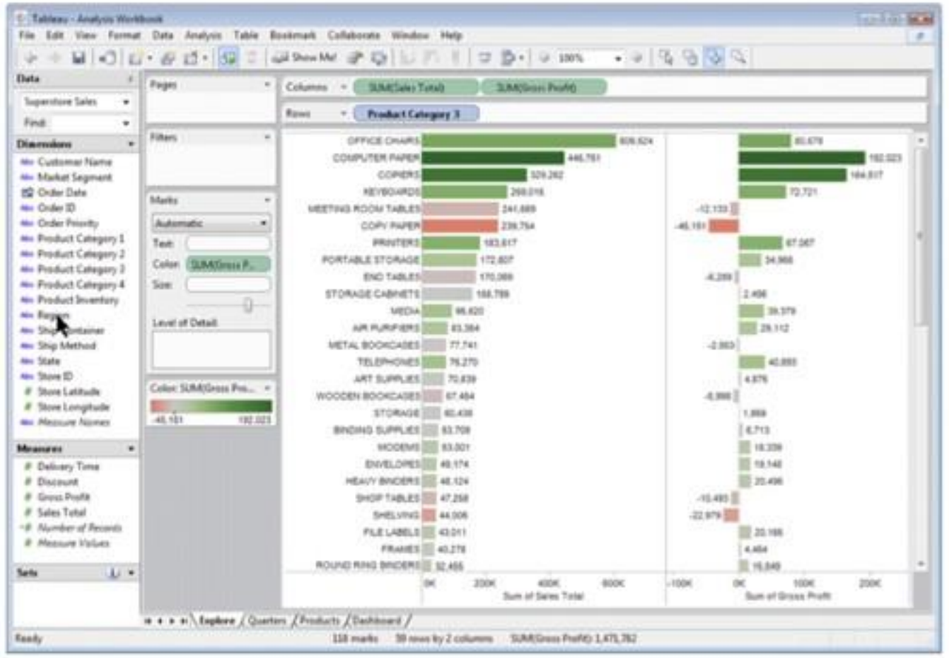
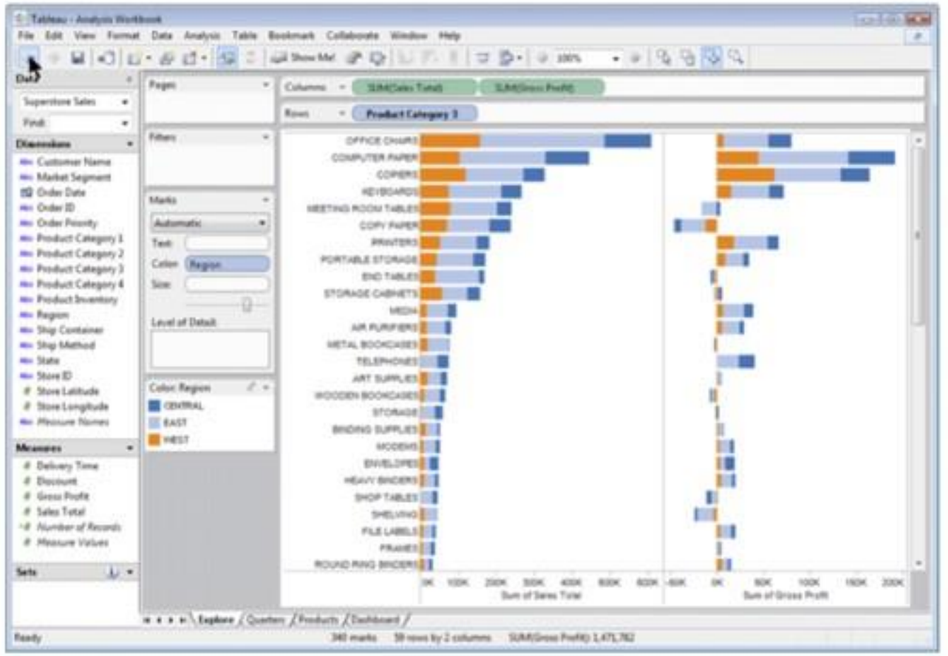
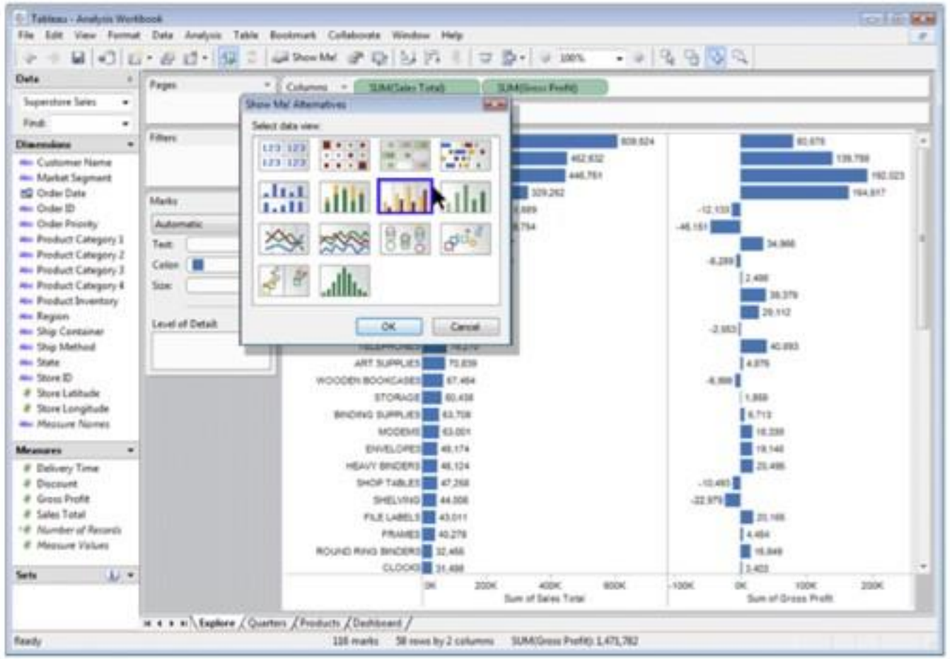
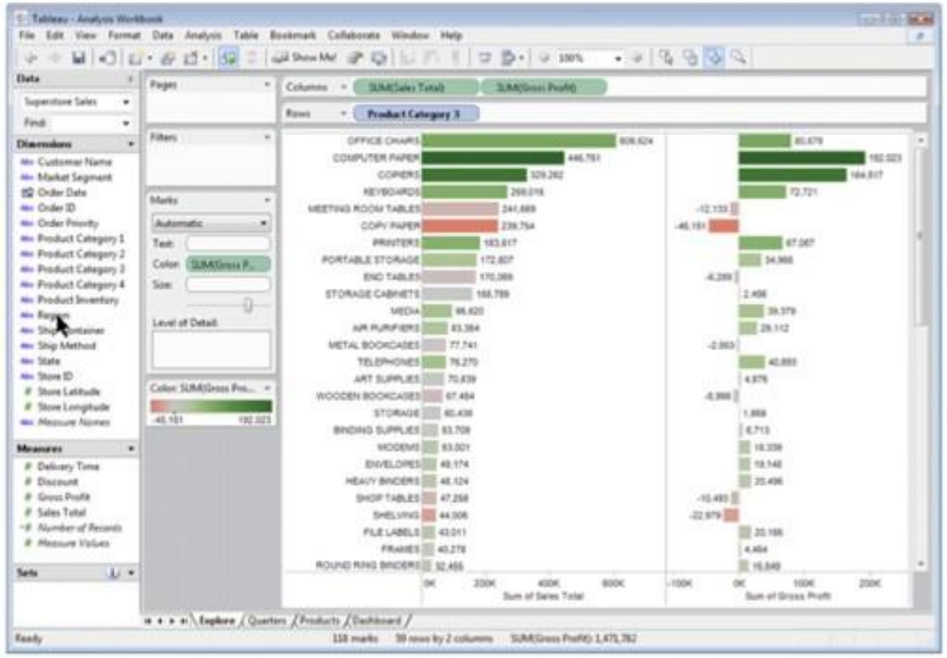


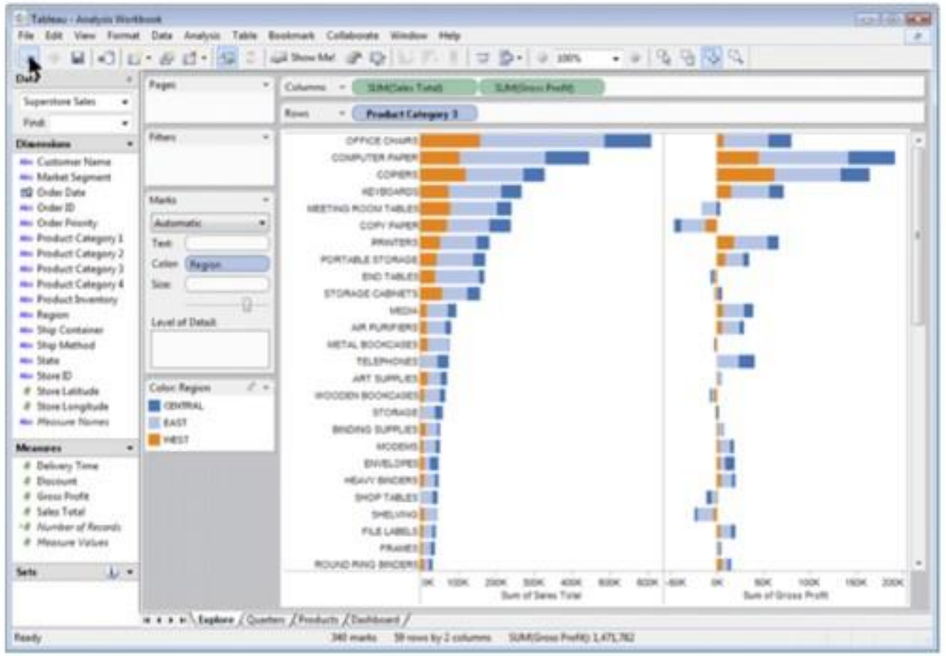
Tableau - Analysis Workbook
File Edit View Format Data Analysis Table Bookmark Collaborate Window Help
Data
Superstore Sales
Dimensions
Customer Name
Market Segment
Order Date
Order ID
Order Priority
Product Category 1
Product Category 2
Product Category 3
Product Category 4
Product Inventory
Region
Ship Container
Ship Method
State
Store ID
Store Latitude
Store Longitude
Measure Names
Measures
Delivery Time
Discount
Gross Profit
Sales Total
Number of Records
Measure Values
Sets
Pages
Filters
Marks
Automatic
Text
Color Region
Size
Level of Detail
Color: Region
CENTRAL
EAST
WEST
Columns SUM(Sales Total) SUM(Gross Profit)
Rows Product Category 3
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
0K 100K 200K 300K 400K 500K 600K
Sum of Sales Total
-60K 0K 60K 100K 150K 200K
Sum of Gross Profit
Explore Quarters Products Dashboard
Ready
340 marks 39 rows by 2 columns SUM(Gross Profit): 1,471,782

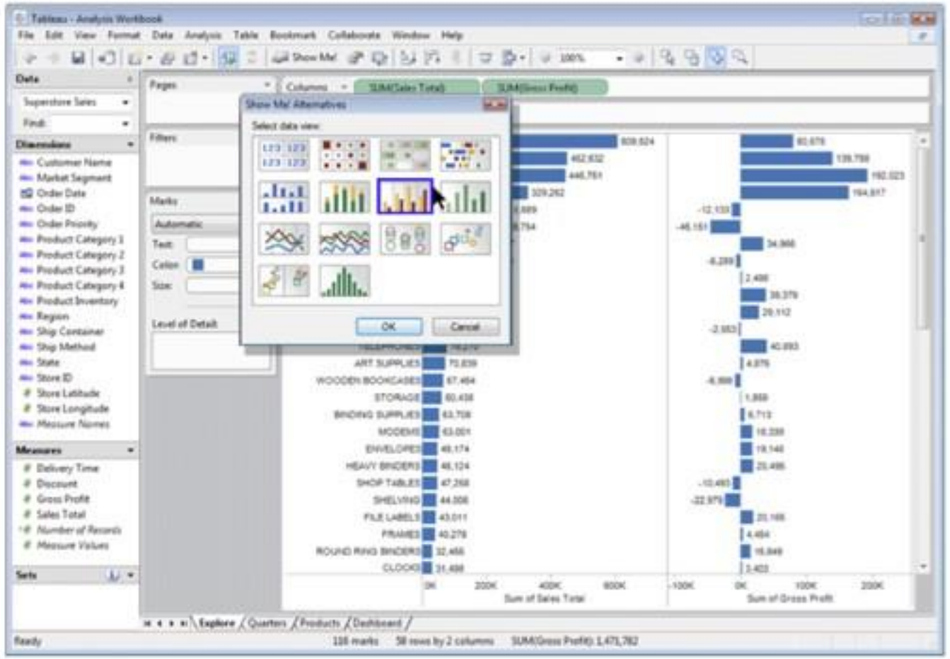




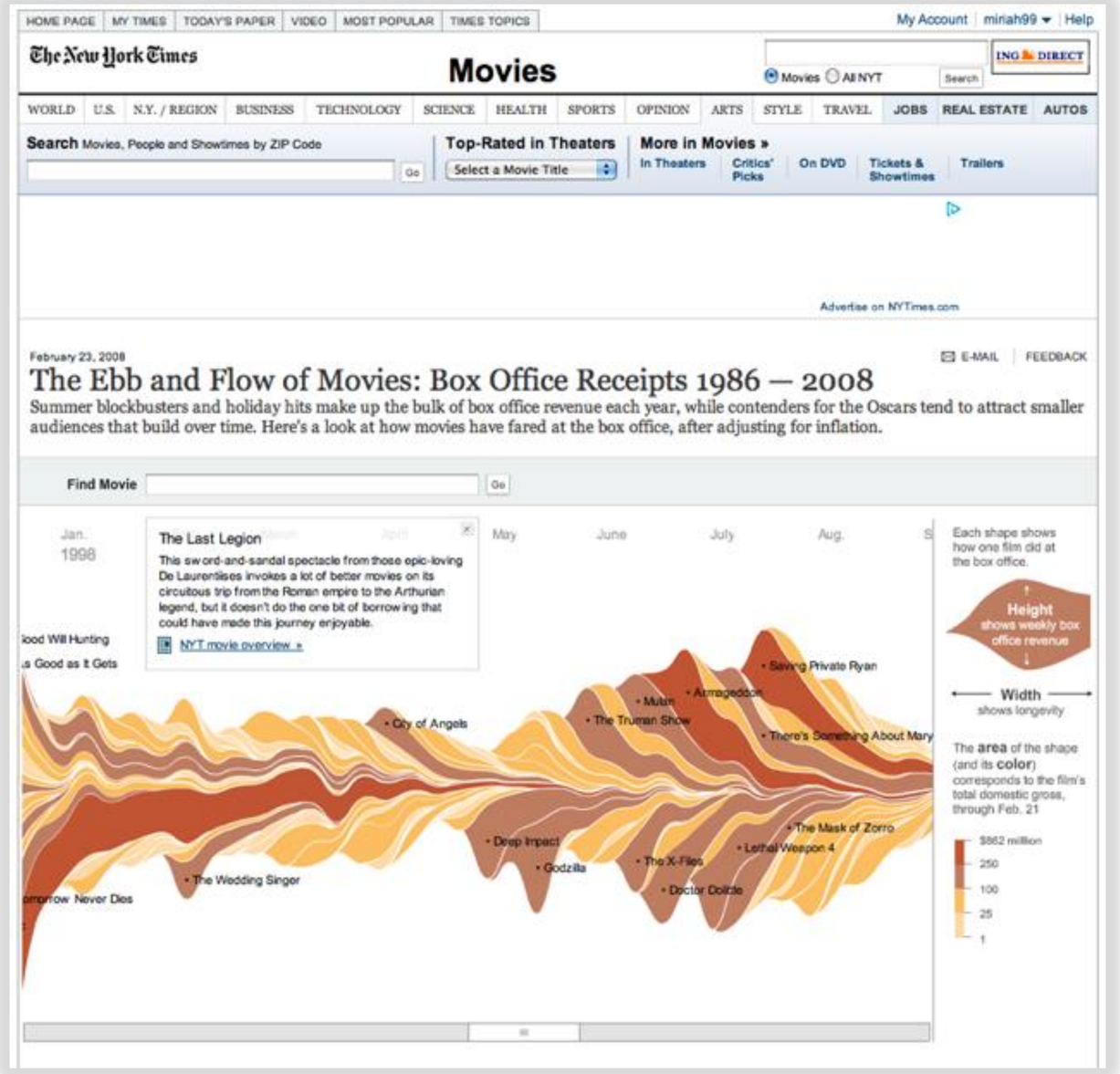




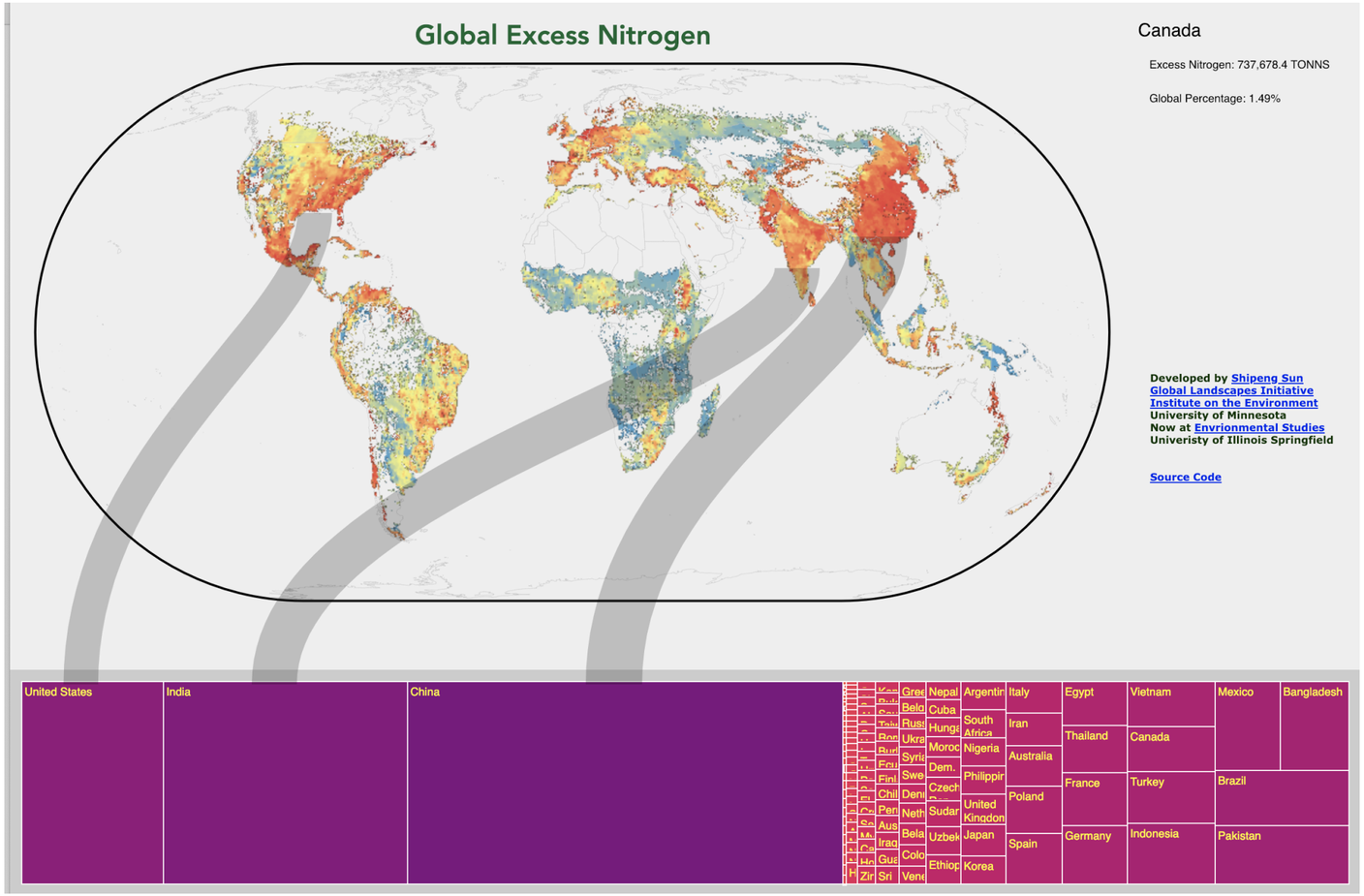




SELECTION & HIGHLIGHTING

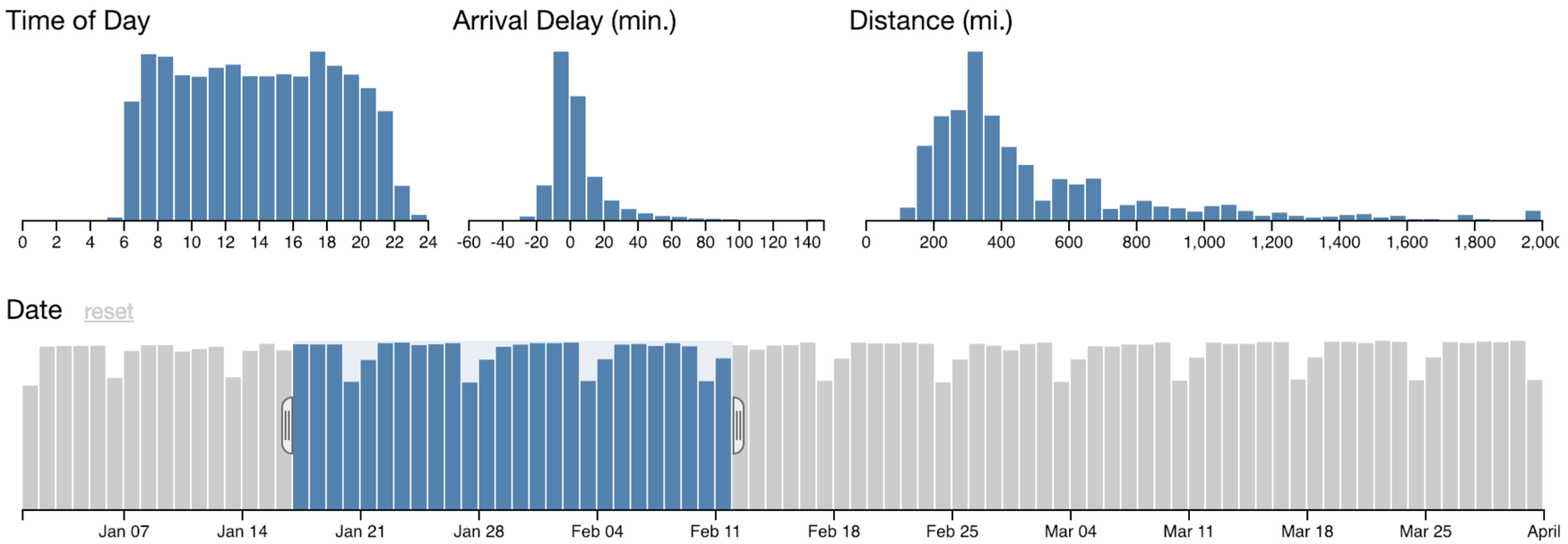

HOME PAGE | MY TIMES | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS
My Account | miriah99 | Help
The New York Times
Movies
Movies | All NYT | Search
WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL | JOBS | REAL ESTATE | AUTOS
Search Movies, People and Showtimes by ZIP Code
Top-Rated in Theaters
Select a Movie Title
More in Movies »
In Theaters | Critics' Picks | On DVD | Tickets & Showtimes | Trailers
Advertise on NYTimes.com
February 23, 2008
E-MAIL | FEEDBACK
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.
Find Movie
Go
Jan. 1998
May
June
July
Aug.
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.
NYT movie overview »
ood Will Hunting
s Good as It Gets
City of Angels
Saving Private Ryan
Armageddon
Mulan
The Truman Show
There's Something About Mary
The Wedding Singer
Deep Impact
Godzilla
The X-Files
Lethal Weapon 4
The Mask of Zorro
Doctor Dolittle
Each shape shows how one film did at the box office.
Height shows weekly box office revenue
Width shows longevity
The area of the shape (and its color) corresponds to the film's total domestic gross, through Feb. 21
$862 million
250
100
25
1

Linking



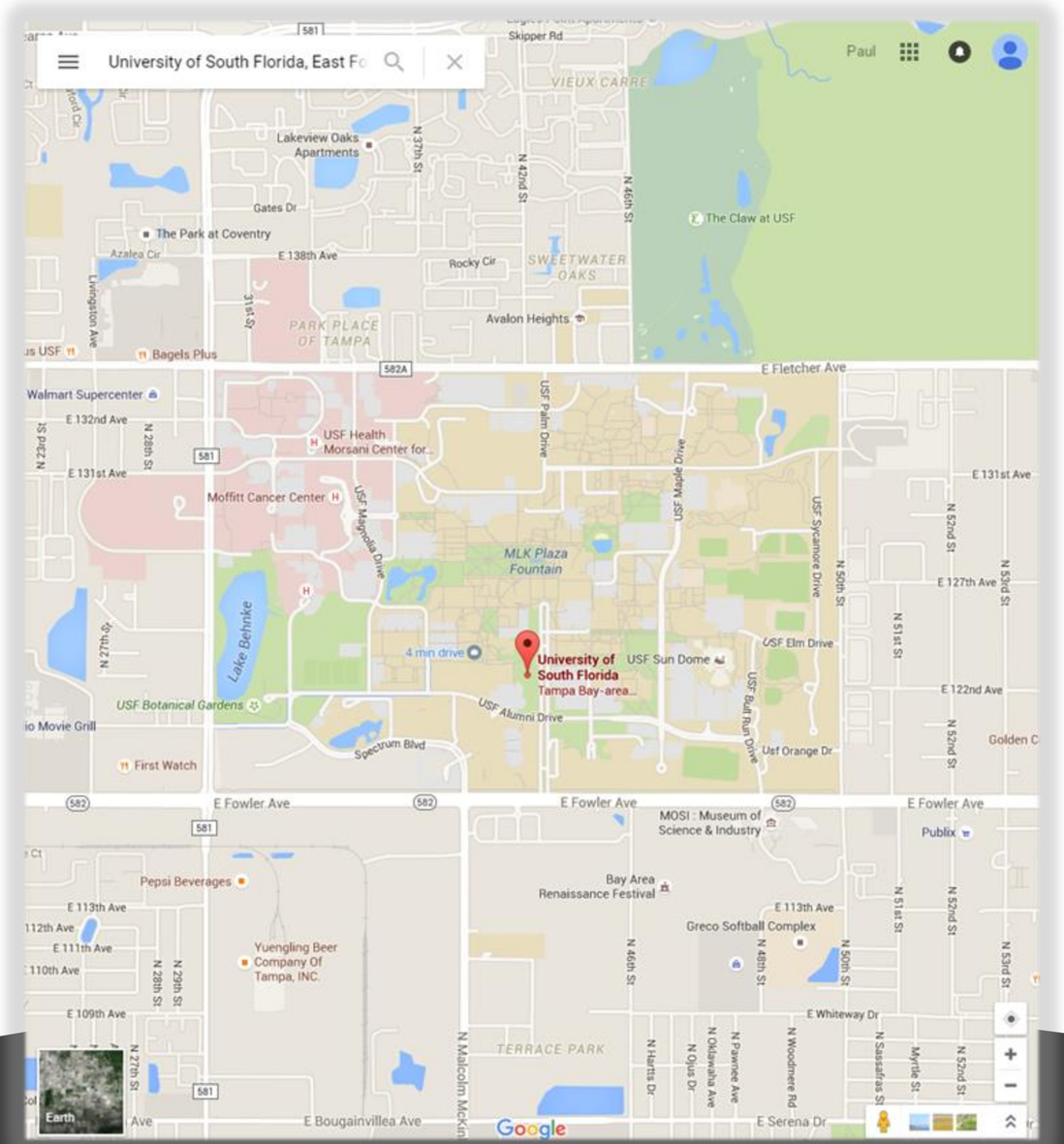
http://sunsp.net/demo/GeogTreeMaps/

Filtering

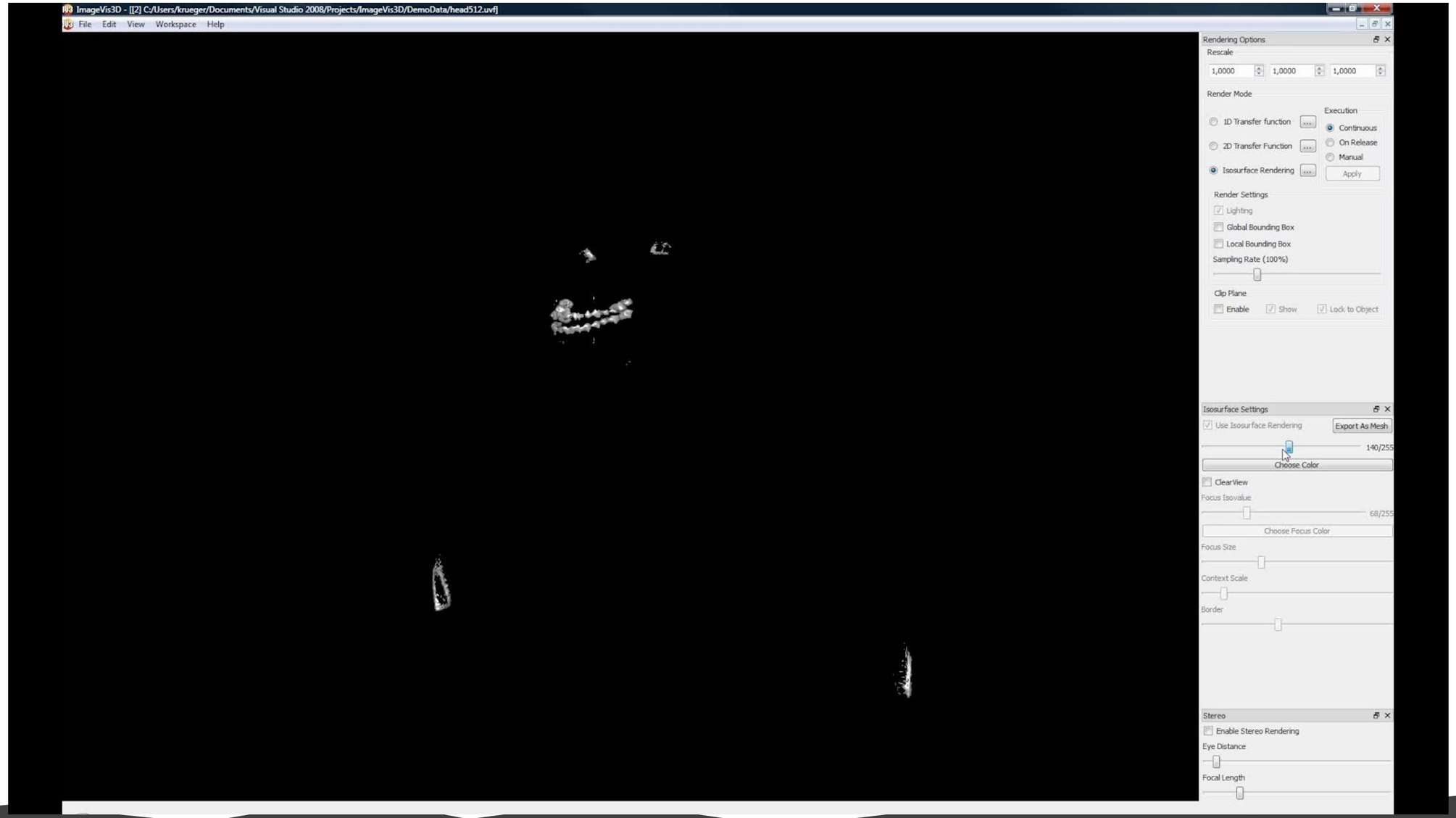

Time of Day
0 2 4 6 8 10 12 14 16 18 20 22 24
Arrival Delay (min.)
-60 -40 -20 0 20 40 60 80 100 120 140
Distance (mi.)
0 200 400 600 800 1,000 1,200 1,400 1,600 1,800 2,000
Date
reset
Jan 07 Jan 14 Jan 21 Jan 28 Feb 04 Feb 11 Feb 18 Feb 25 Mar 04 Mar 11 Mar 18 Mar 25 April

NAVIGATION

pan (and translate)

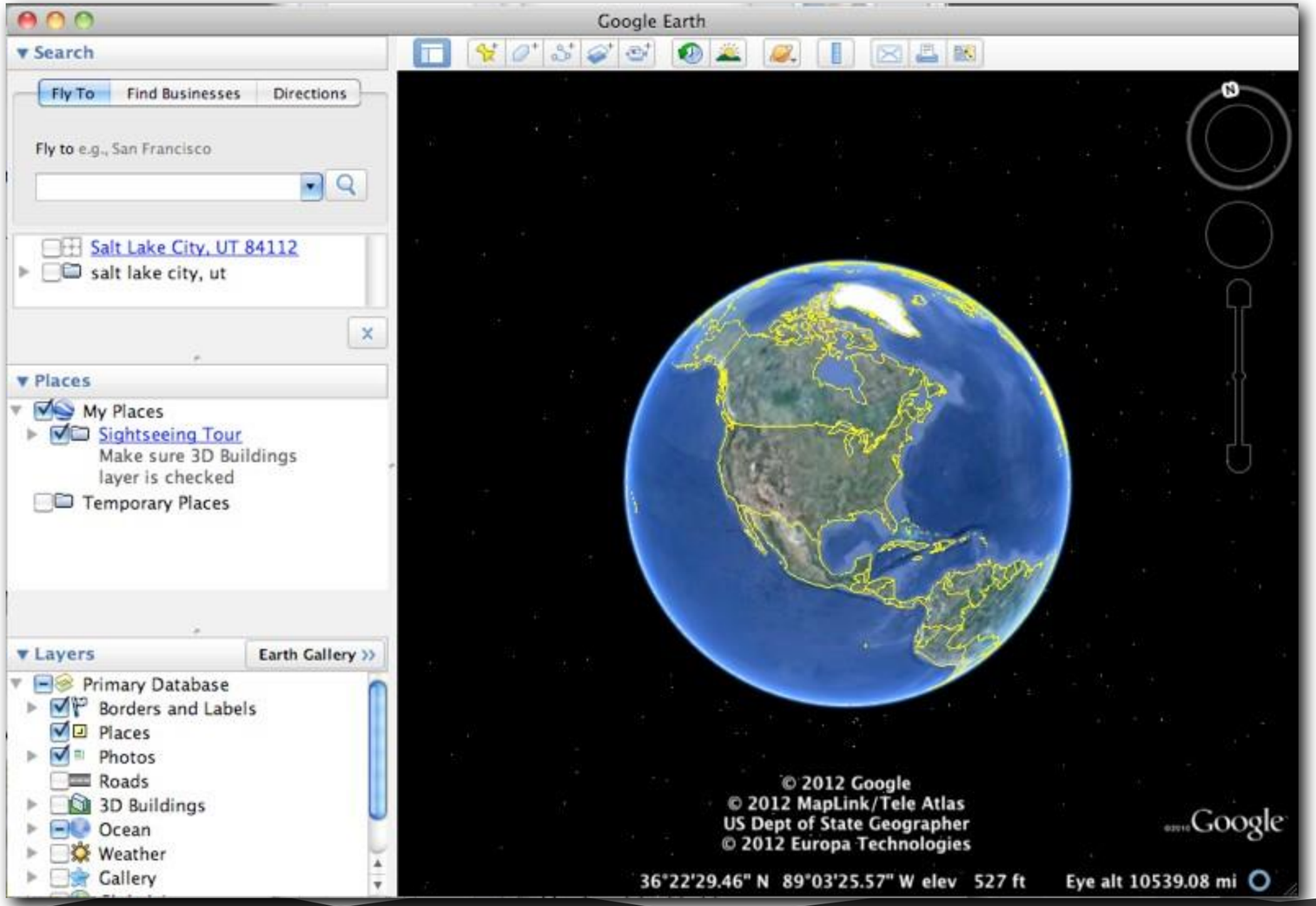


rotate


ImageVis3D - [[2] C:/Users/krueger/Documents/Visual Studio 2008/Projects/ImageVis3D/DemoData/head512.uvf]
File Edit View Workspace Help
Rendering Options
Rescale
1,0000 1,0000 1,0000
Render Mode
1D Transfer function
2D Transfer Function
Isosurface Rendering
Execution
Continuous
On Release
Manual
Apply
Render Settings
Lighting
Global Bounding Box
Local Bounding Box
Sampling Rate (100%)
Clip Plane
Enable
Show
Lock to Object
Isosurface Settings
Use Isosurface Rendering
Export As Mesh
140/255
Choose Color
ClearView
Focus Isovalue
68/255
Choose Focus Color
Focus Size
Context Scale
Border
Stereo
Enable Stereo Rendering
Eye Distance
Focal Length

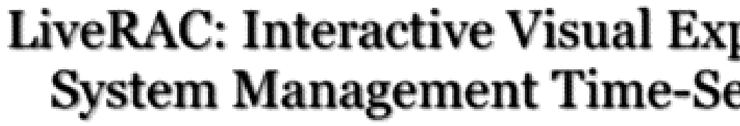
GEOMETRIC vs SEMANTIC ZOOMING

geometric


Google Earth
Search
Fly To
Find Businesses
Directions
Fly to e.g., San Francisco
Salt Lake City, UT 84112
salt lake city, ut
Places
My Places
Sightseeing Tour
Make sure 3D Buildings layer is checked
Temporary Places
Layers
Earth Gallery
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
Google
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