Data Visualizations

Dr. Bo (Beth) Sun

Posted on CAVAS

acmqueue

A Tour through the Visualization Zoo

A survey of powerful visualization techniques, from the obvious to the obscure

Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky, Stanford University

Thanks to advances in sensing, networking, and data management, our society is producing digital information at an astonishing rate. According to one estimate, in 2010 alone we will generate 1,200 exabytes—60 million times the content of the Library of Congress. Within this deluge of data lies a wealth of valuable information on how we conduct our businesses, governments, and personal lives. To put the information to good use, we must find ways to explore, relate, and communicate the data meaningfully.

https://rowan.instructure.com/courses/2381249/files/129900344/download


Data
Insights

How to do that?

COMPUTATION
+
HUMAN INTUITION

How to do that?

COMPUTATION	INTERACTIVE VIS
Automatic	User-driven; iterative
Summarization, clustering, classification	Interaction, visualization
>Millions of nodes	Thousands of nodes

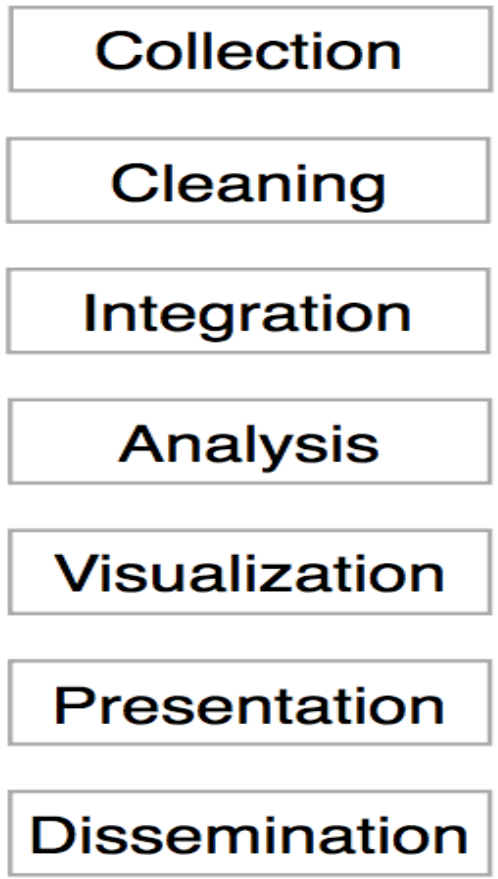
Both develop methods for making sense of network data

Our Approach for Big Data Analytics

DATA MINING	HCI Human-Computer Interaction
Automatic	User-driven; iterative
Summarization, clustering, classification	Interaction, visualization
>Millions of items	Thousands of items

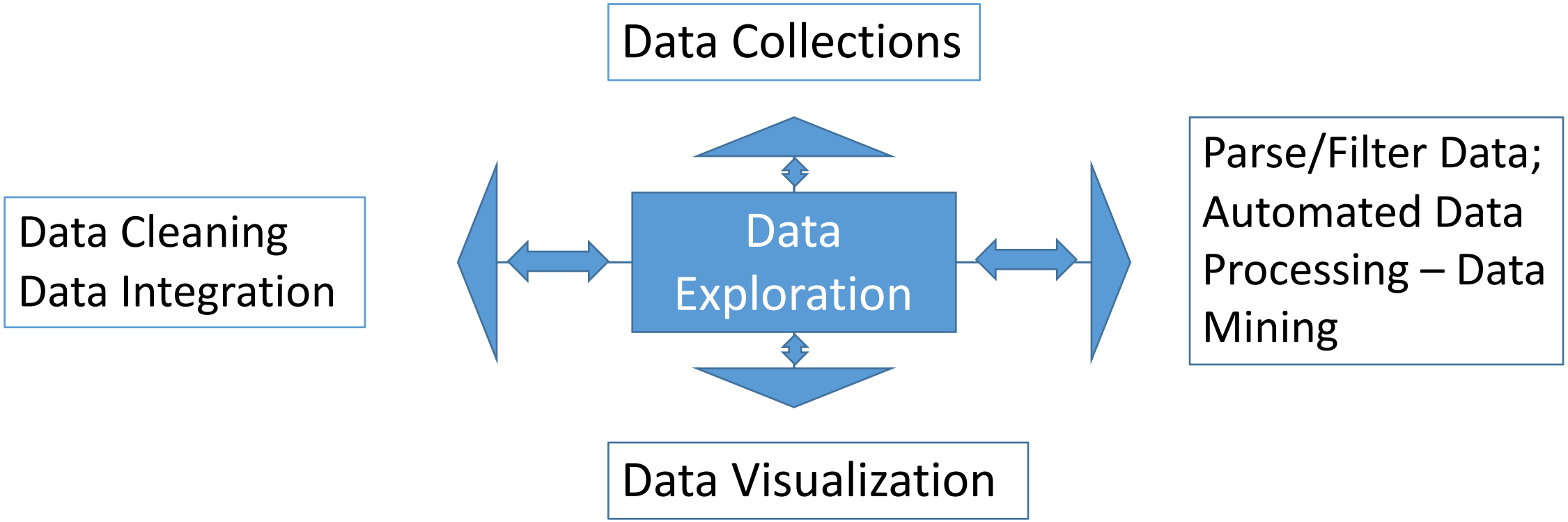
Our research combines the **Best of Both Worlds**

Building blocks, not “steps”



- **Can skip some**
- **Can go back (two-way street)**
- Examples
  - Data types inform visualization design
  - Data informs choice of algorithms
  - Visualization informs data cleaning (dirty data)
  - Visualization informs algorithm design (user finds that results don’t make sense)

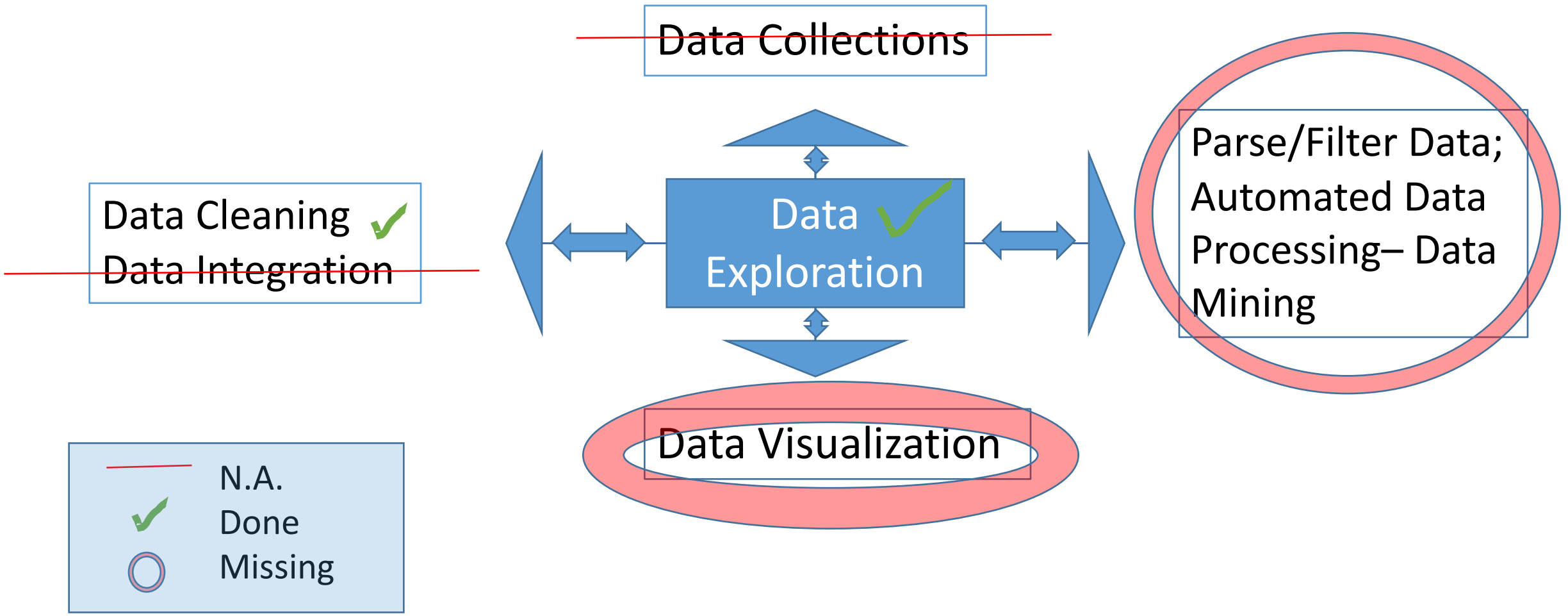
Dr. Sun's Visual Analytics Process



Data Exploration---Use Graphs and Human Intuition/HCI to explore data set to find out data trends and patterns;

Data Visualization: design an interactive graph to best present data insights

Regarding Final Projects



Automated Data Processing – Data Mining

1. Classification

(or Probability Estimation)

Predict which of a (small) set of classes an entity belong to.

1. Classification

(or Probability Estimation)

Predict which of a (small) set of classes an entity belong to.

- email spam (y, n)
- sentiment analysis (+, -, neutral)
- news (politics, sports, …)
- medical diagnosis (cancer or not)
- shirt size (s, m, l)
- cat detection
- face detection (baby, middle-aged, etc.)
- buy /not buy - commerce

2. Regression ("value estimation")

Predict the **numerical value** of some variable for an entity.

2. Regression ("value estimation")

Predict the **numerical value** of some variable for an entity.

- point value of wine (50-100)
- credit score
- stock prices
- relationship between price and sales
- weather
- sports and game scores

3. Similarity Matching

Find similar entities (from a large dataset) based on what we know about them.

- find similar gene sequences (that may be repeating, or does similar things)
- online dating
- patent search
- carpool matching (find people to carpool)

4. Clustering (unsupervised learning)

Group entities together by their similarity.

(For most algorithms, user provides # of clusters)

4. Clustering (unsupervised learning)

Group entities together by their similarity.

- groupings of similar bugs in code
- topical analysis (tweets?)
- land cover: tree/road/…
- for advertising: grouping users for marketing purposes
- cluster people by accents (y'all, you all)

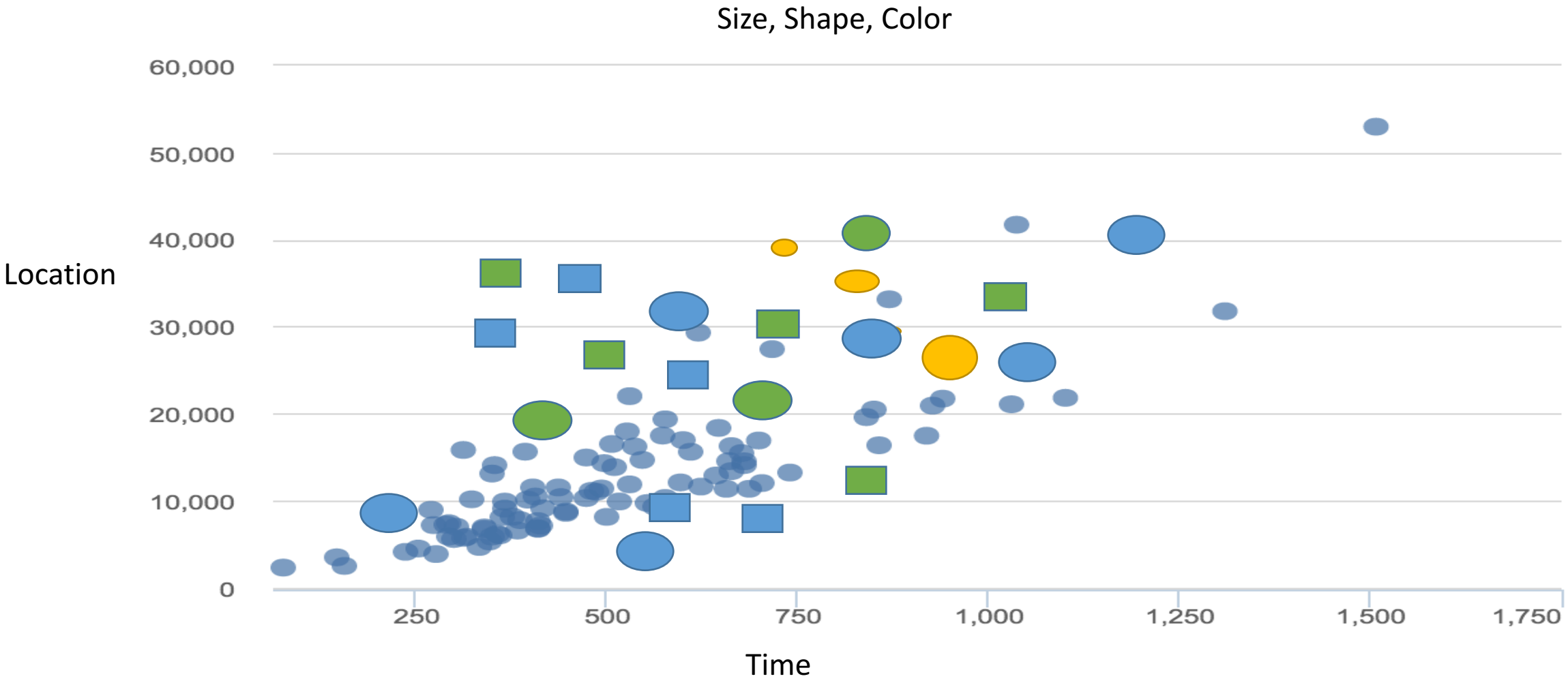
5. Data reduction ("dimensionality reduction")

Shrink a large dataset into smaller one, with as little loss of information as possible

1. if you want to visualize the data (in 2D/3D)
2. faster computation/less storage
3. reduce noise

Data Visualization--design an interactive graph to best present data insights

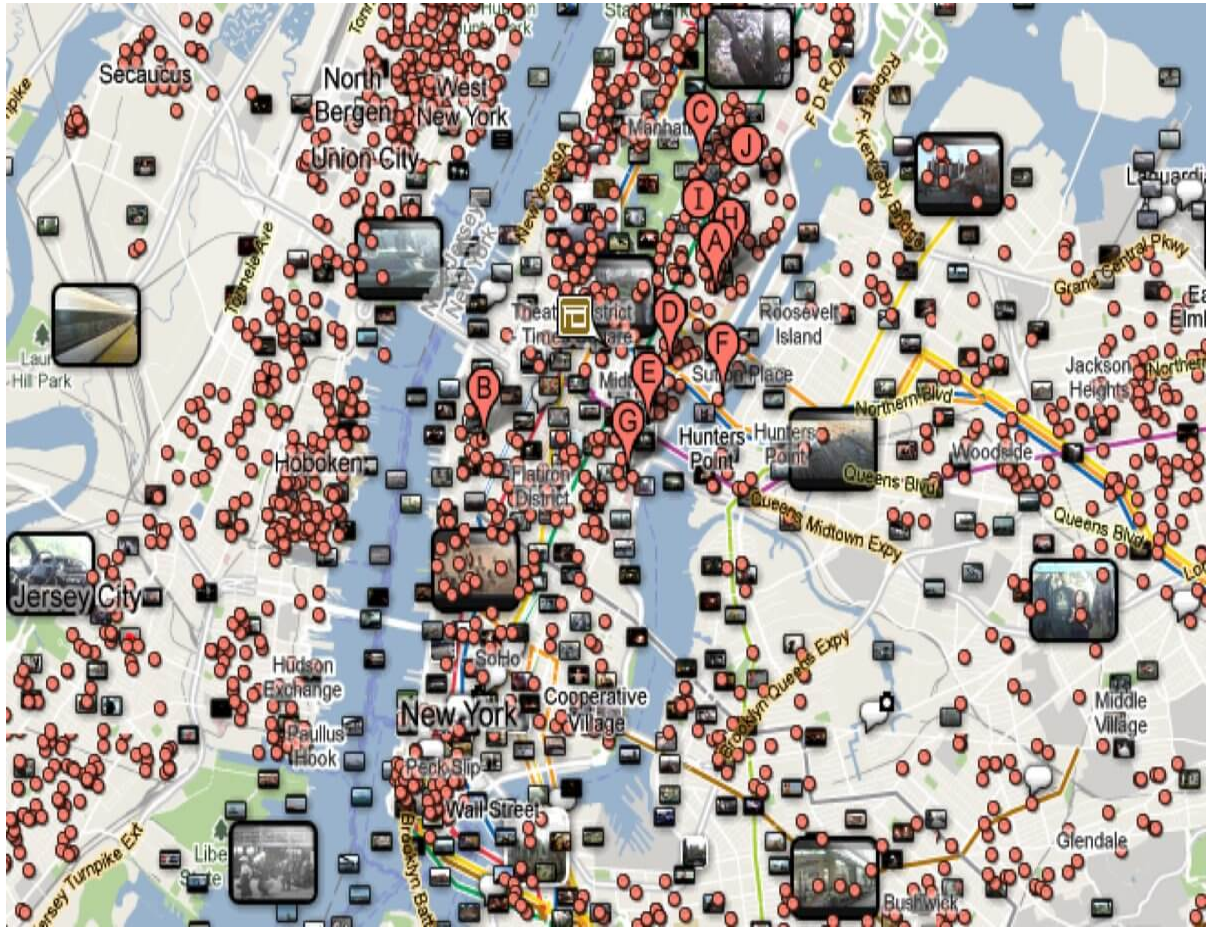
Scatter Plots/Bubble Charts



Scatter Plots/Bubble Charts

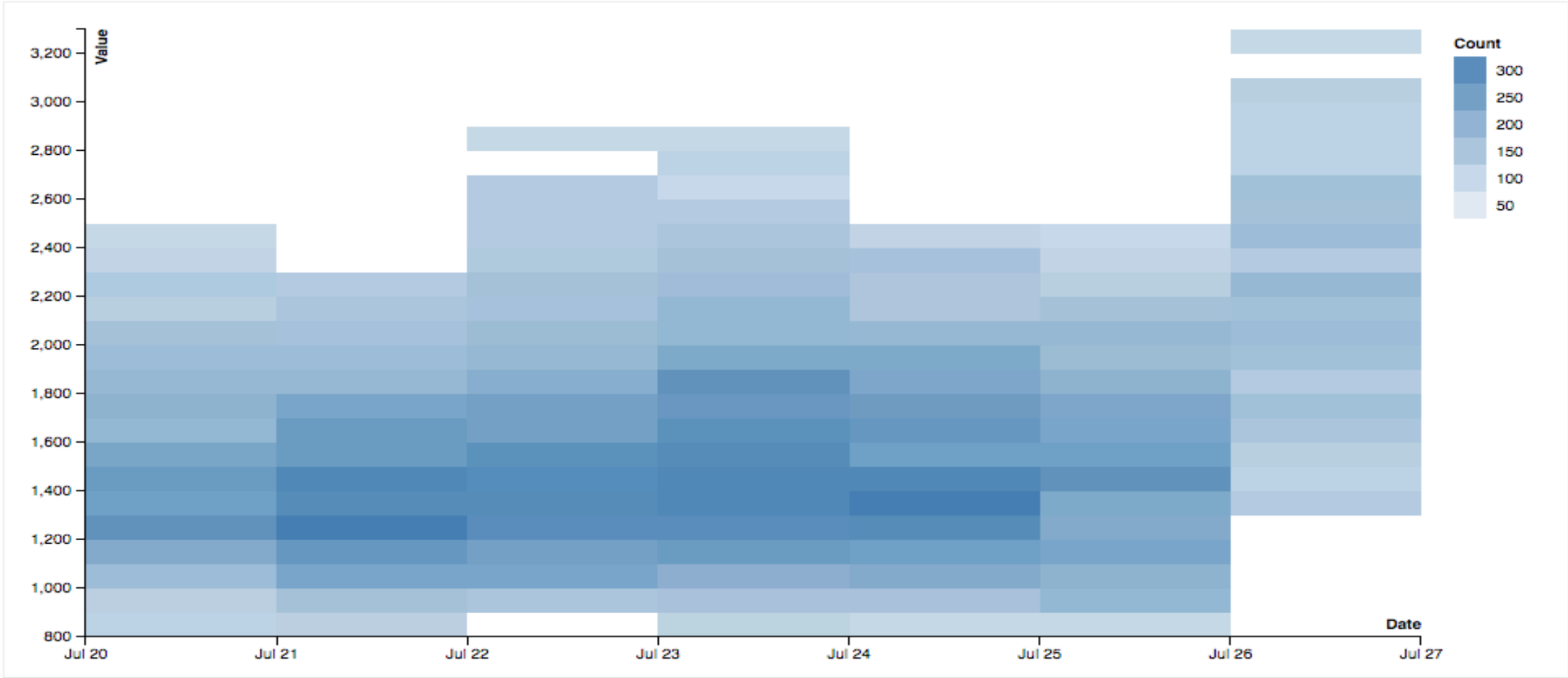
- Present relationships between two (scatter) or three (bubble) numerical variables,
- Plot two or three sets of variables on one x-y coordinate plane,
- Present patterns in large sets of data, linear or non-linear trends, correlations, clusters, or outliers.
- Compare large number of data points. The more data you include in a scatter chart, the better comparisons you can make.
- Present relationships, but not exact values for comparisons.

Map Chart

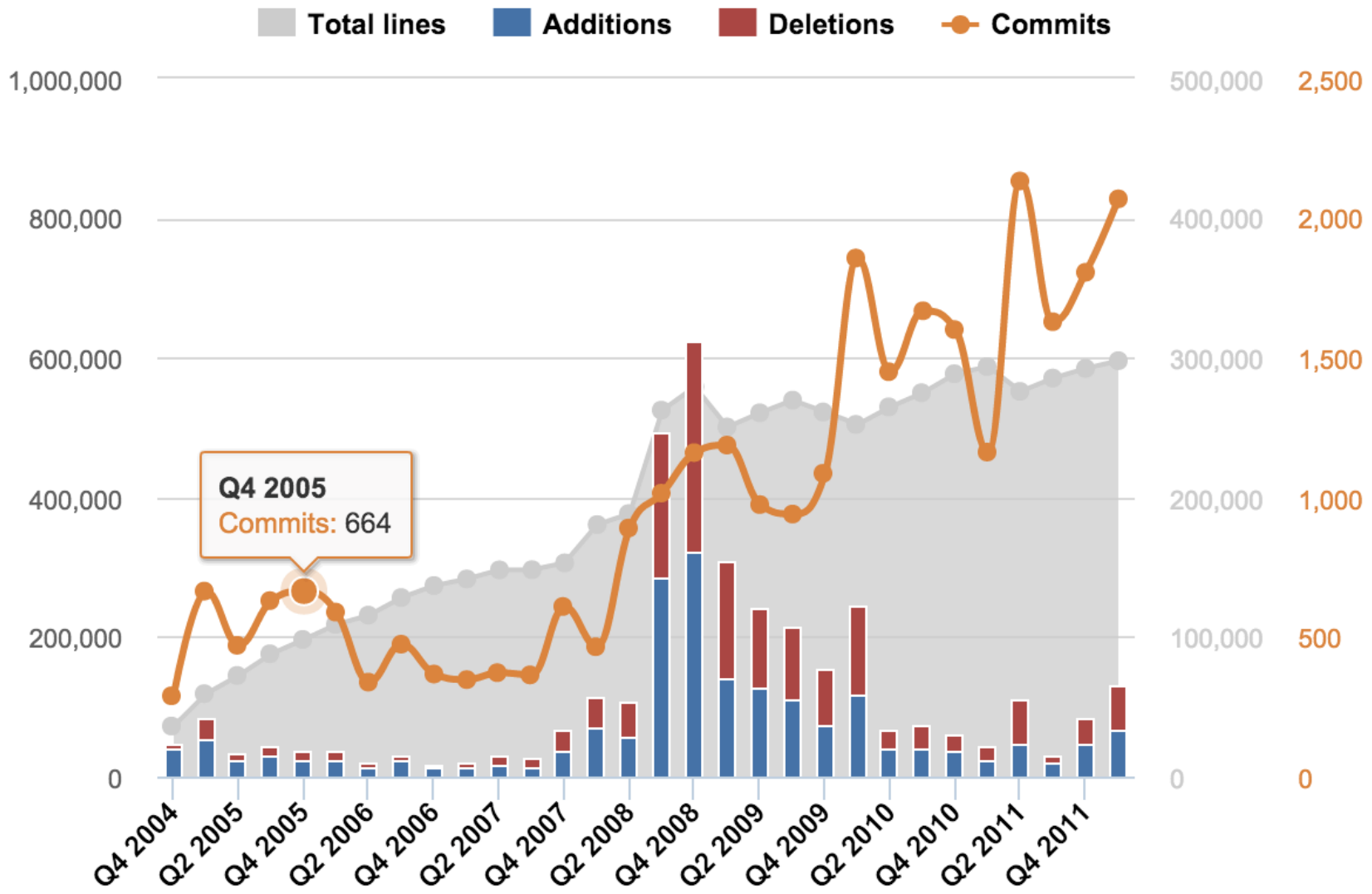


- If you want to display quantitative information on a map.
- To present spatial relationships and patterns.
- When a regional context for your data is important.
- To get an overview of the distribution across geographic locations.
- Only if your data is standardized (that is, it has the same data format and scale for the whole set).

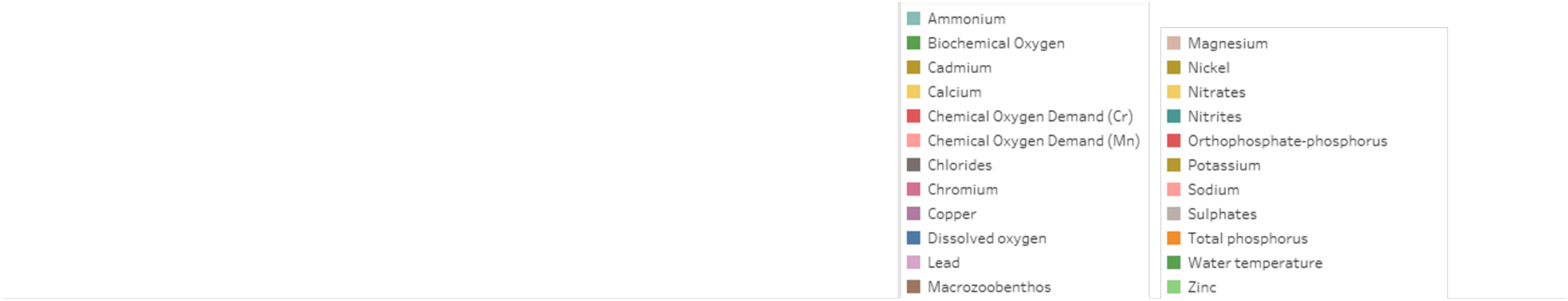
Heatmap (2D Histogram, CSV)



Multi Axes Charts



- Display more charts with the same X-axis.
- Compare multiple measures with different value ranges.
- Illustrate the relationships, correlation, or the lack thereof between two or more measures in one visualization.
- Save canvas space.



Calendar View By Year

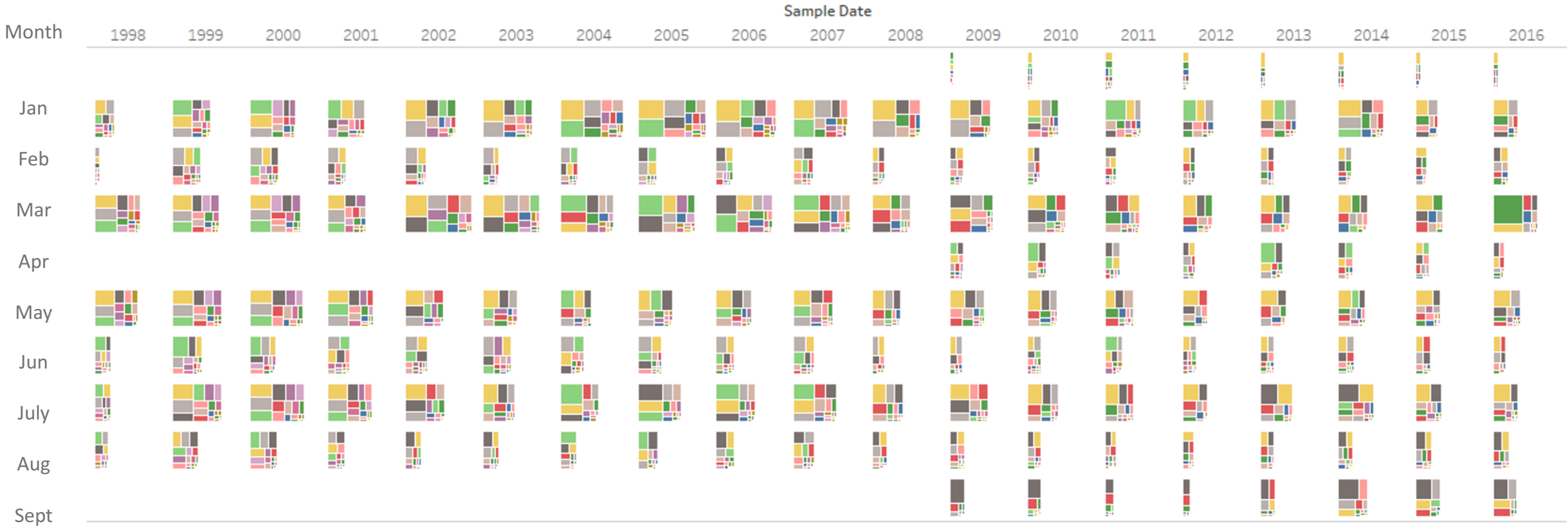
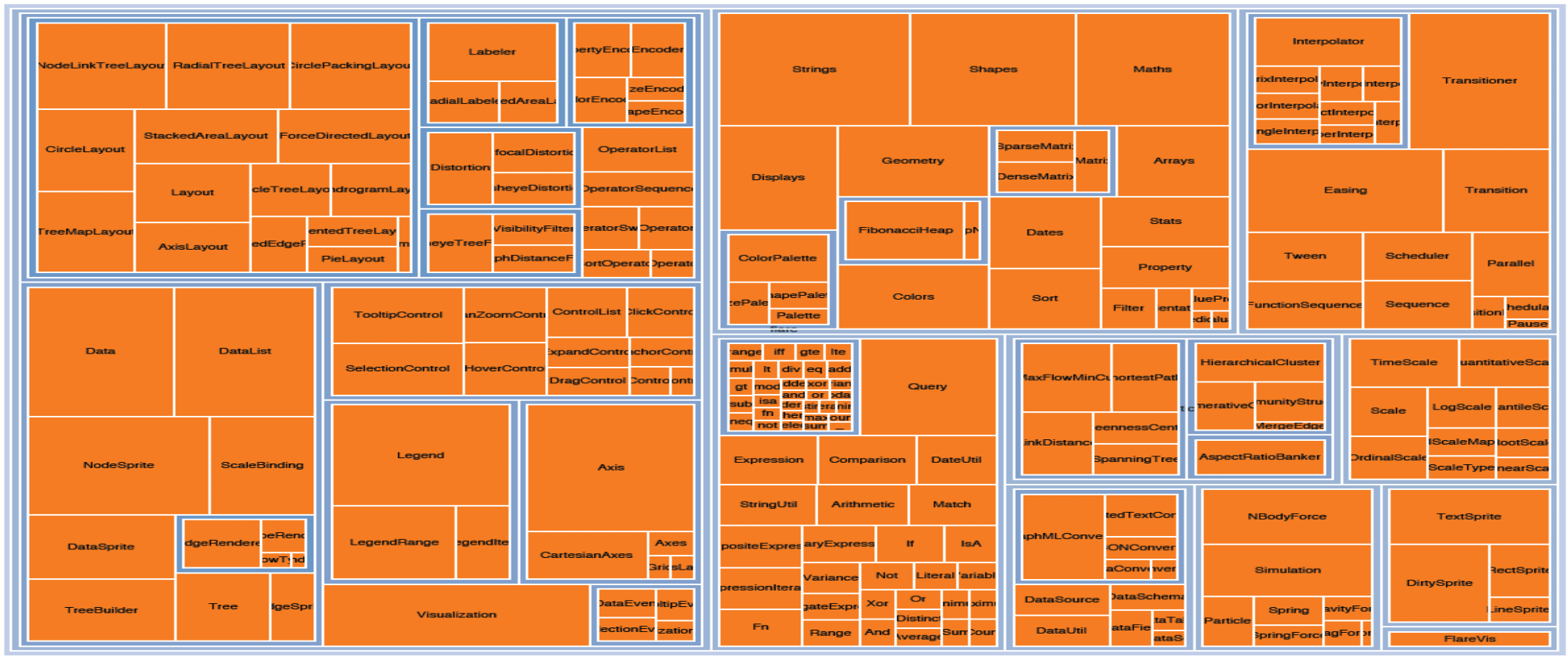


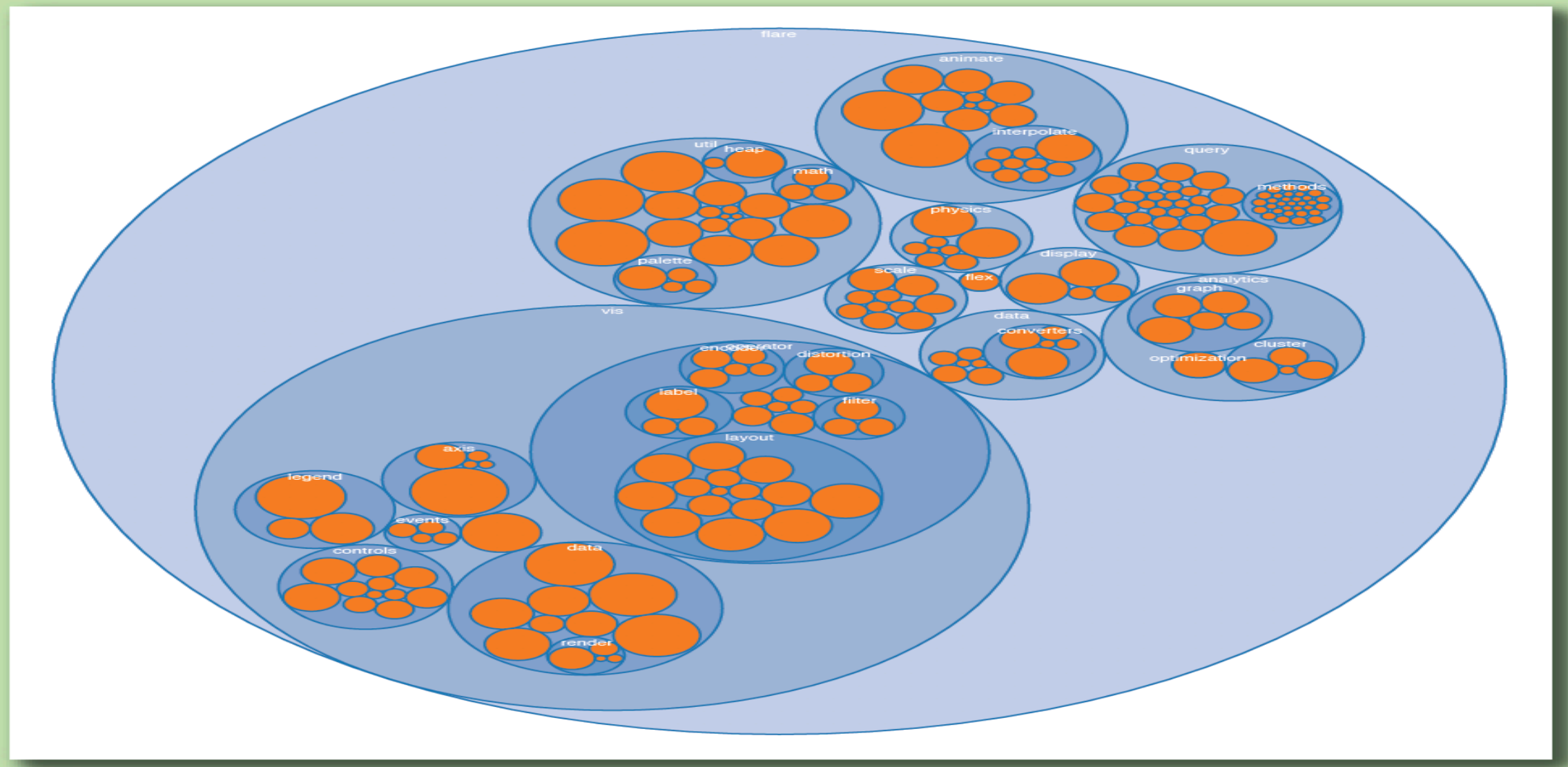
FIGURE 4F

Treemap Layout of the Flare Package Hierarchy



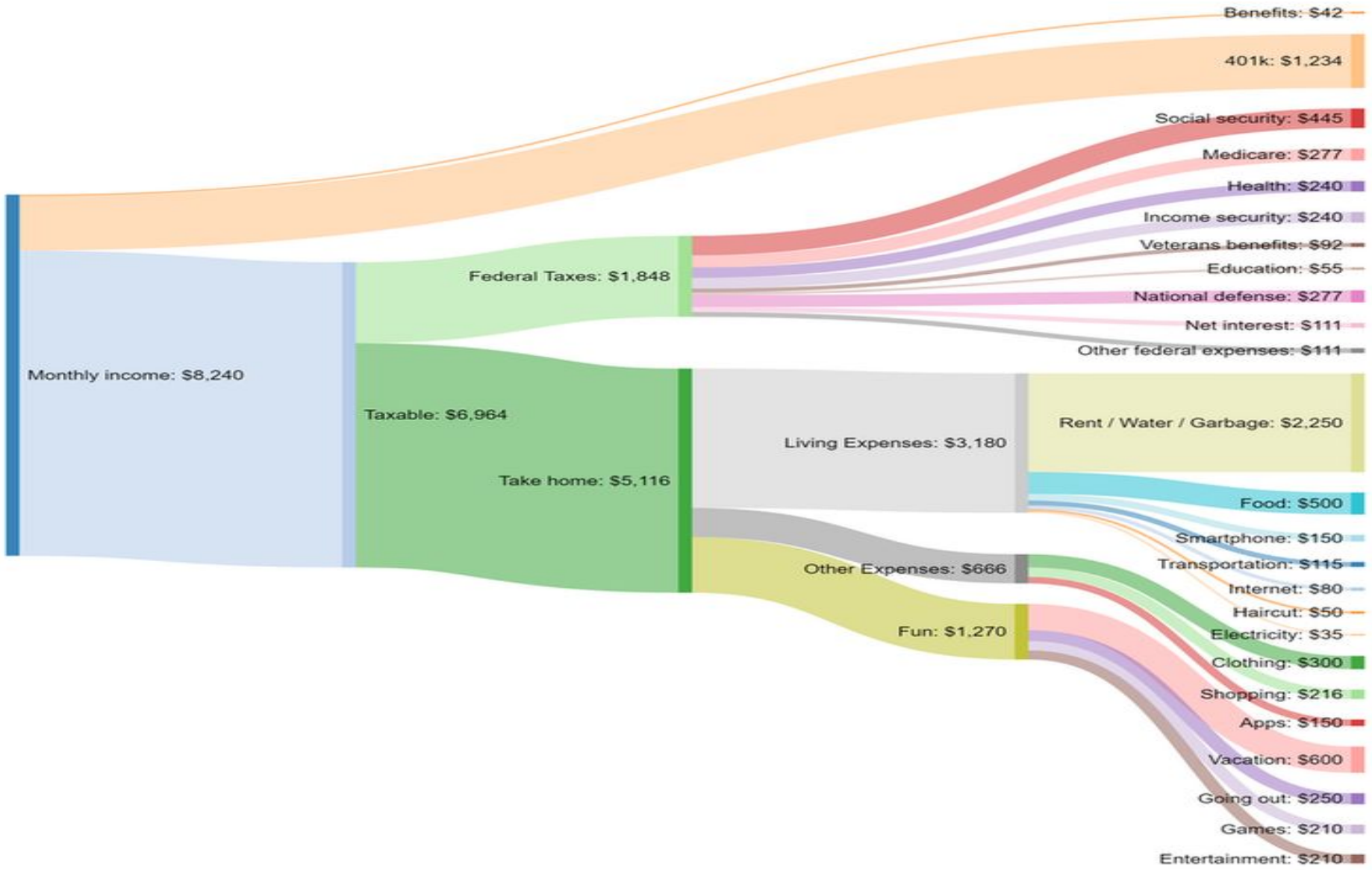
Source: Flare Visualization Toolkit (http://flare.prefuse.org)
http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/treemap.html

Nested Circles Layout of the Flare Package Hierarchy



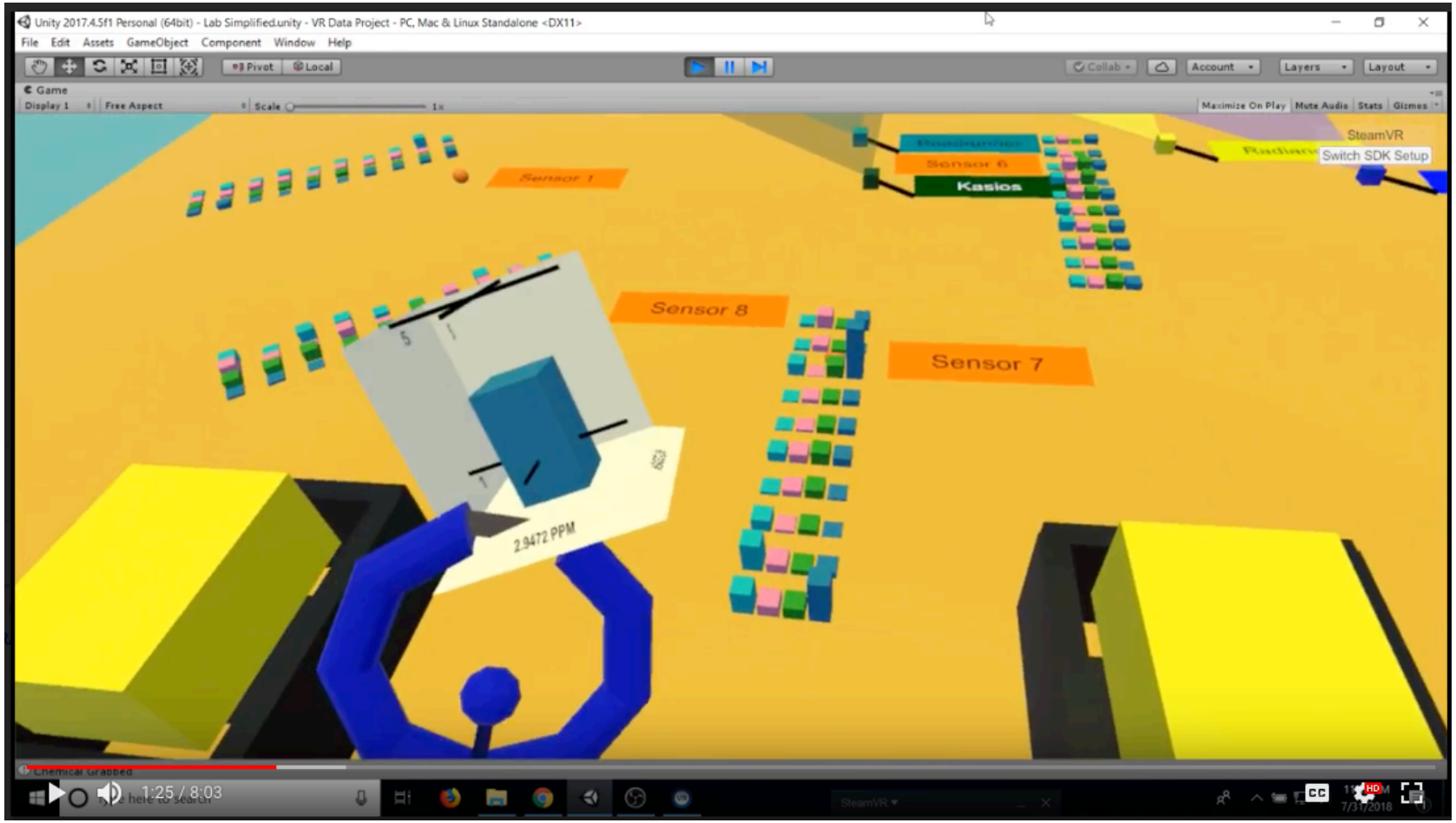
Source: Flare Visualization Toolkit (http://flare.prefuse.org)
http://hci.stanford.edu/jheer/files/zoo/ex/hierarchies/pack.html

Sankey Diagrams



- Sankey Diagrams display flows and their quantities in proportion to one another. The width of the arrows or lines are used to show their magnitudes;
- Typically, Sankey Diagrams are used to visually show the transfer of energy, money or materials, but they can be used to show the flow of any isolated system process

Immersive Data Visualization- VAST Challenge 2018 MC2



D3 Examples

- Fisheye Distortion: https://bost.ocks.org/mike/fisheye/
- Geographic Bounding Boxes: https://www.jasondavies.com/maps/bounds/
- At the National Conventions, the Words They Used: https://archive.nytimes.com/www.nytimes.com/interactive/2012/09/06/us/politics/convention-word-counts.html#Jobs
- Airports: http://mbostock.github.io/d3/talk/20111116/airports.html
- Baseball: https://archive.nytimes.com/www.nytimes.com/interactive/2013/03/29/sports/baseball/Strikeouts-Are-Still-Soaring.html?ref=baseball

* From https://d3js.org/