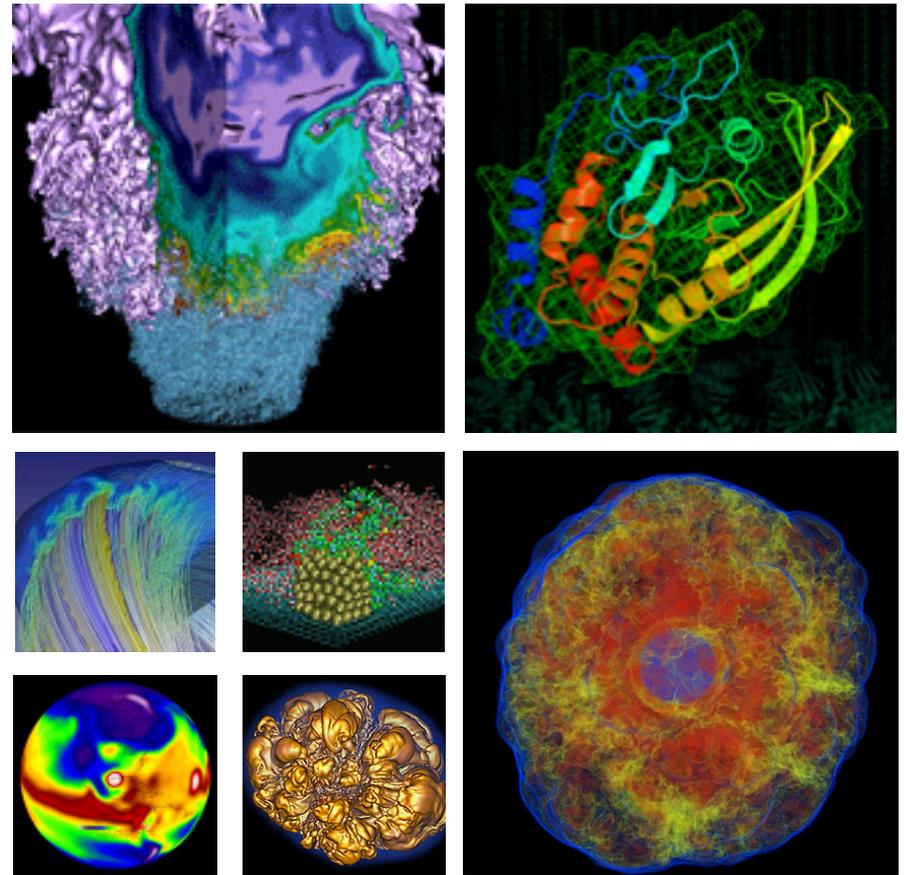


# Tools for Data Visualization



**Annette Greiner**  
**NERSC Data and Analytics Services**

Data Day, August 22, 2016

# Why Visualize?

# Anscombe's Quartet



X	Y
10.0	8.04
8.0	6.95
13.0	7.58
9.0	8.81
11.0	8.33
14.0	9.96
6.0	7.24
4.0	4.26
12.0	10.84
7.0	4.82
5.0	5.68

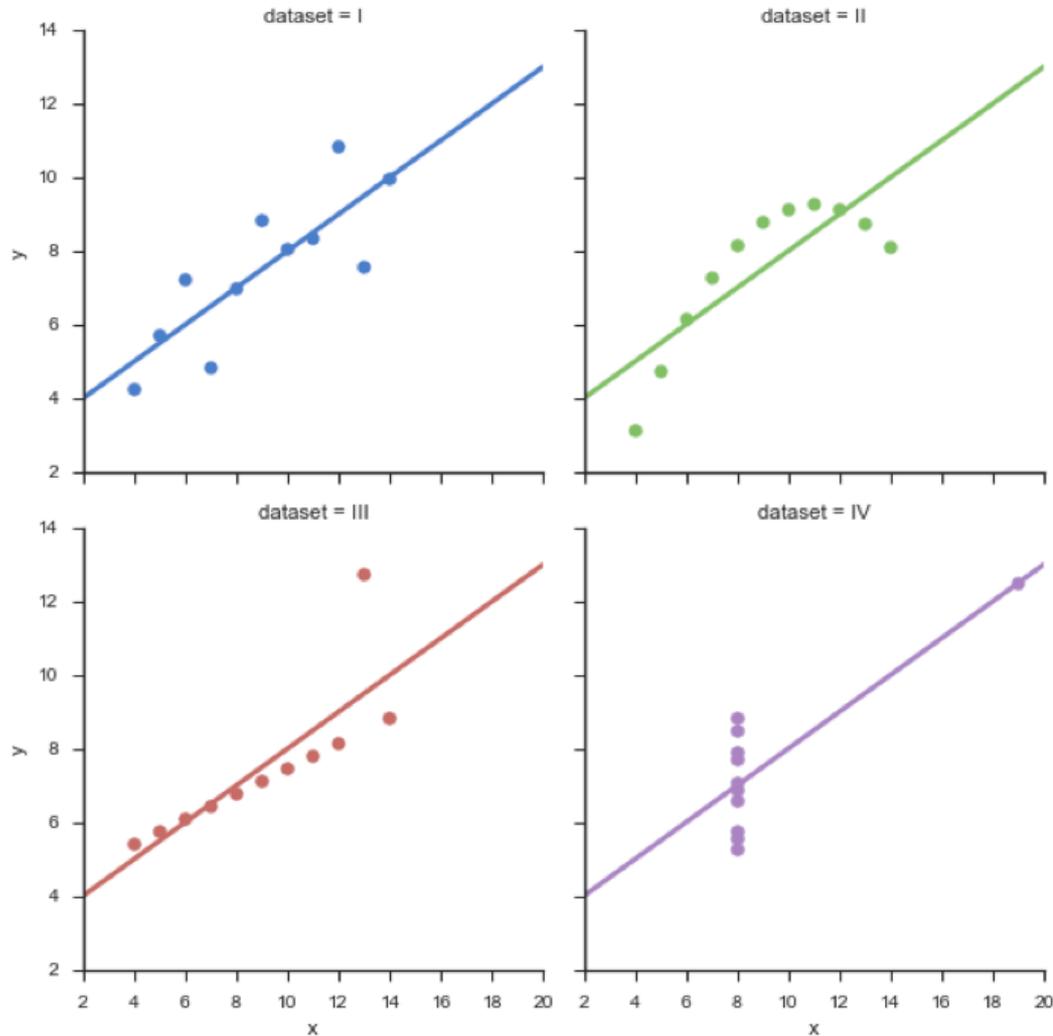
X	Y
10.0	9.14
8.0	8.14
13.0	8.74
9.0	8.77
11.0	9.26
14.0	8.10
6.0	6.13
4.0	3.10
12.0	9.13
7.0	7.26
5.0	4.74

X	Y
10.0	7.46
8.0	6.77
13.0	12.74
9.0	7.11
11.0	7.81
14.0	8.84
6.0	6.08
4.0	5.39
12.0	8.15
7.0	6.42
5.0	5.73

X	Y
8.0	6.58
8.0	5.76
8.0	7.71
8.0	8.84
8.0	8.47
8.0	7.04
8.0	5.25
19.0	12.50
8.0	5.56
8.0	7.91
8.0	6.89

Same:  
 N  
 Mean X  
 Mean Y  
 Variance X  
 Variance Y  
 Regression  
 Correlation

# Anscombe's Quartet



# Choosing a Tool

# Good News/Bad News



+ DATAVISUALIZATION.CH SELECTED TOOLS

Search...

All

Maps

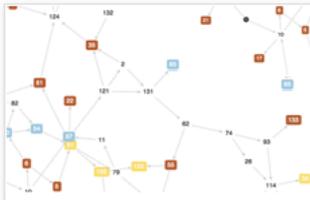
Charts

Data

Color

Code?

✓ ✕



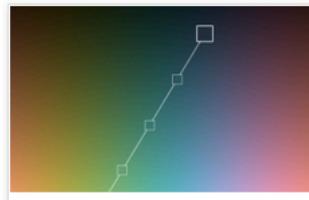
## Arbor.js

A library of force-directed layout algorithms plus abstractions for graph organization and refresh handling.



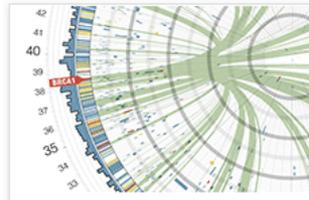
## CartoDB

A web service for mapping, analyzing and building applications with data.



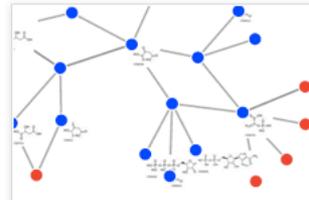
## Chroma.js

Interactive color space explorer that allows to preview a set of linear interpolated equidistant colors.



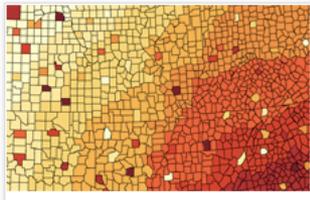
## Circos

A software package for visualizing data in a circular layout.



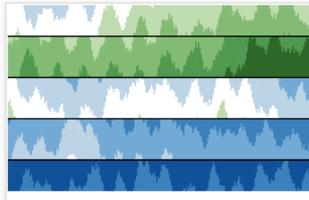
## Cola.js

A library for arranging networks using constraint-based optimization techniques.



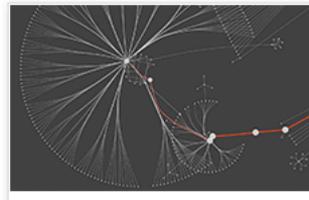
## ColorBrewer

A web tool for selecting colors for maps.



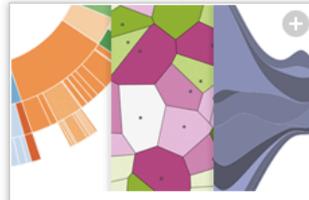
## Cubism.js

A library for creating interactive time series and horizon graphs based on D3.js



## Cytoscape

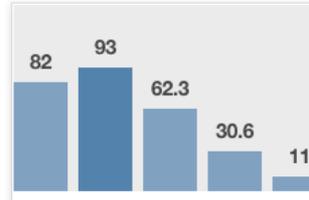
An application for visualizing complex networks and integrating these with any type of attribute data.



## D3.js

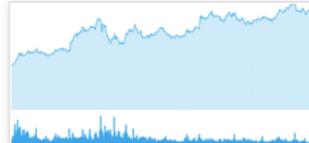
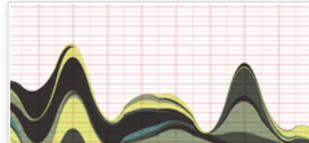
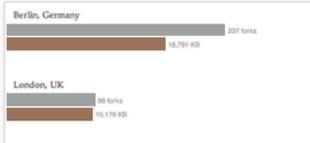
An small, flexible and efficient library to create and manipulate interactive documents based on data.

Ⓜ <http://mbostock.github.com/d3>



## Dance.js

A simple data-driven visualization framework based on Data.js and Underscore.js



## Is your goal explanation or exploration?

### Explanatory:

You already know what you want to say.

### Exploratory:

You want to find out what the data means.

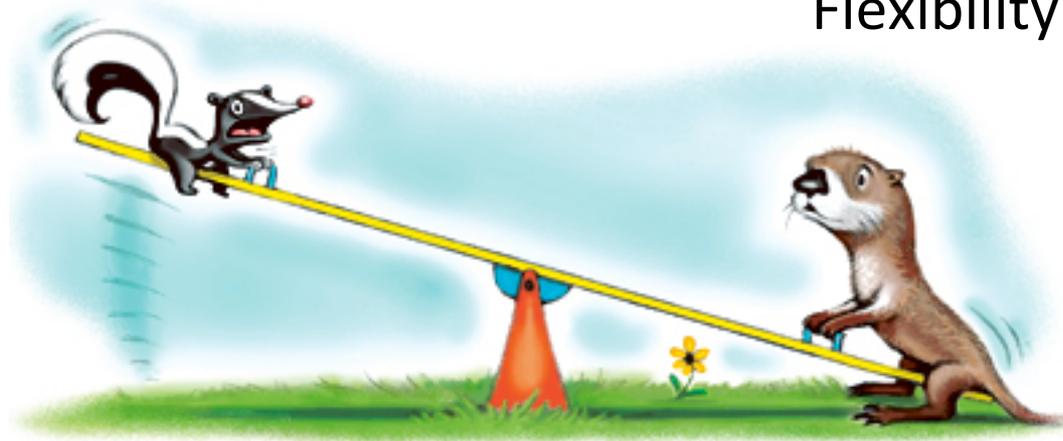
# Speed/Flexibility Tradeoff



Do you have lots of time? Do you need customization?

Speed, Ease

Flexibility



What is compatible with your working habits?



Python

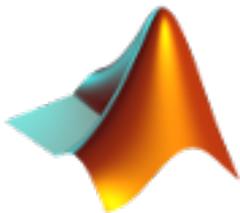


R

Analytics app



Web



Libraries

Versions



## Do you need to share with others?

- **Image files (PNG, TIFF, PDF, etc.)**
- **Interactivity (click, brush, rollover, zoom, pan, etc.)**
- **Code sharing (notebook interface)**
- **Privacy (authentication, authorization)**

Is there a tool aimed at your problem in particular?

yt project About Docs Community Develop Gallery Project Members Quick Links

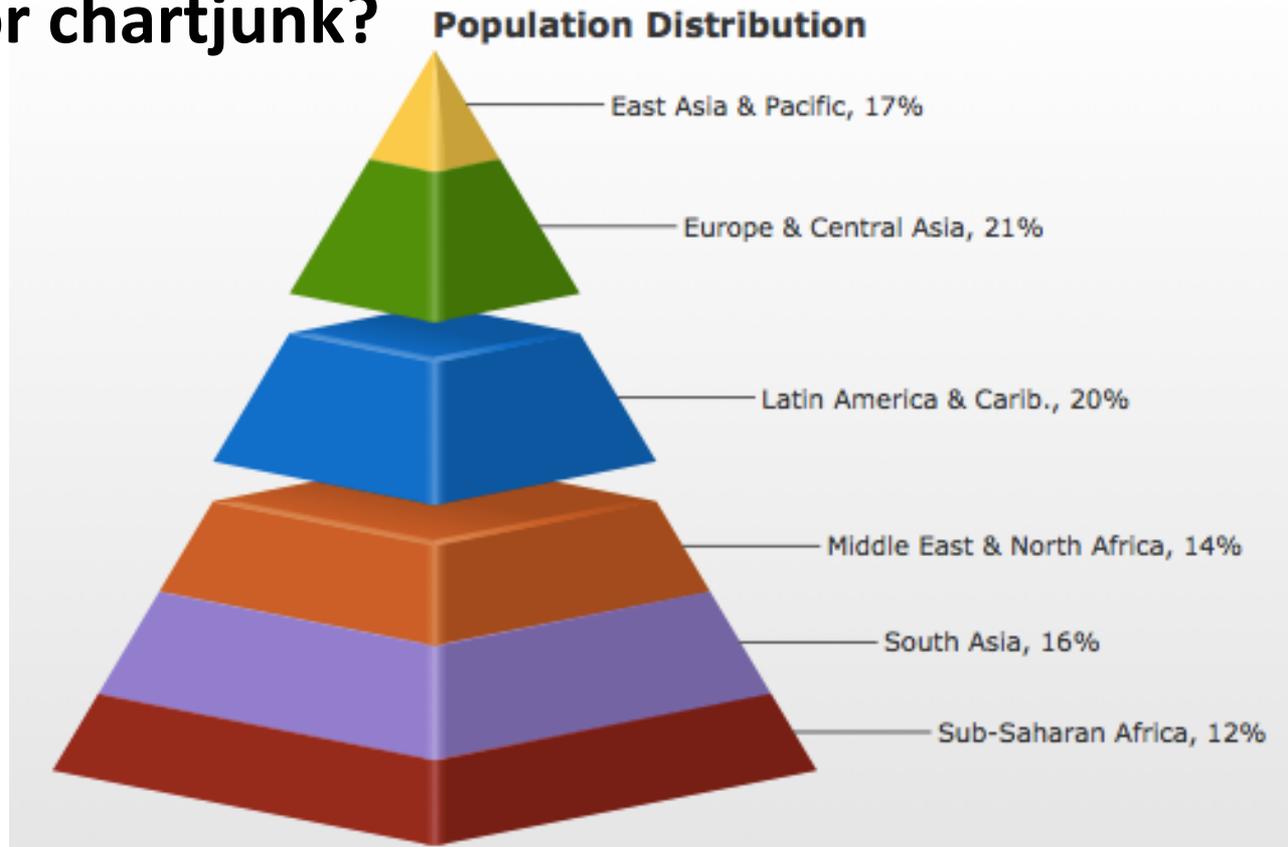
## Volumetric Data Analysis

yt is a python package for analyzing and visualizing volumetric, multi-resolution data from astrophysical simulations, radio telescopes, and a burgeoning interdisciplinary community.

Get yt



What tool will let you show your data without distortion or chartjunk?



## What are the costs?

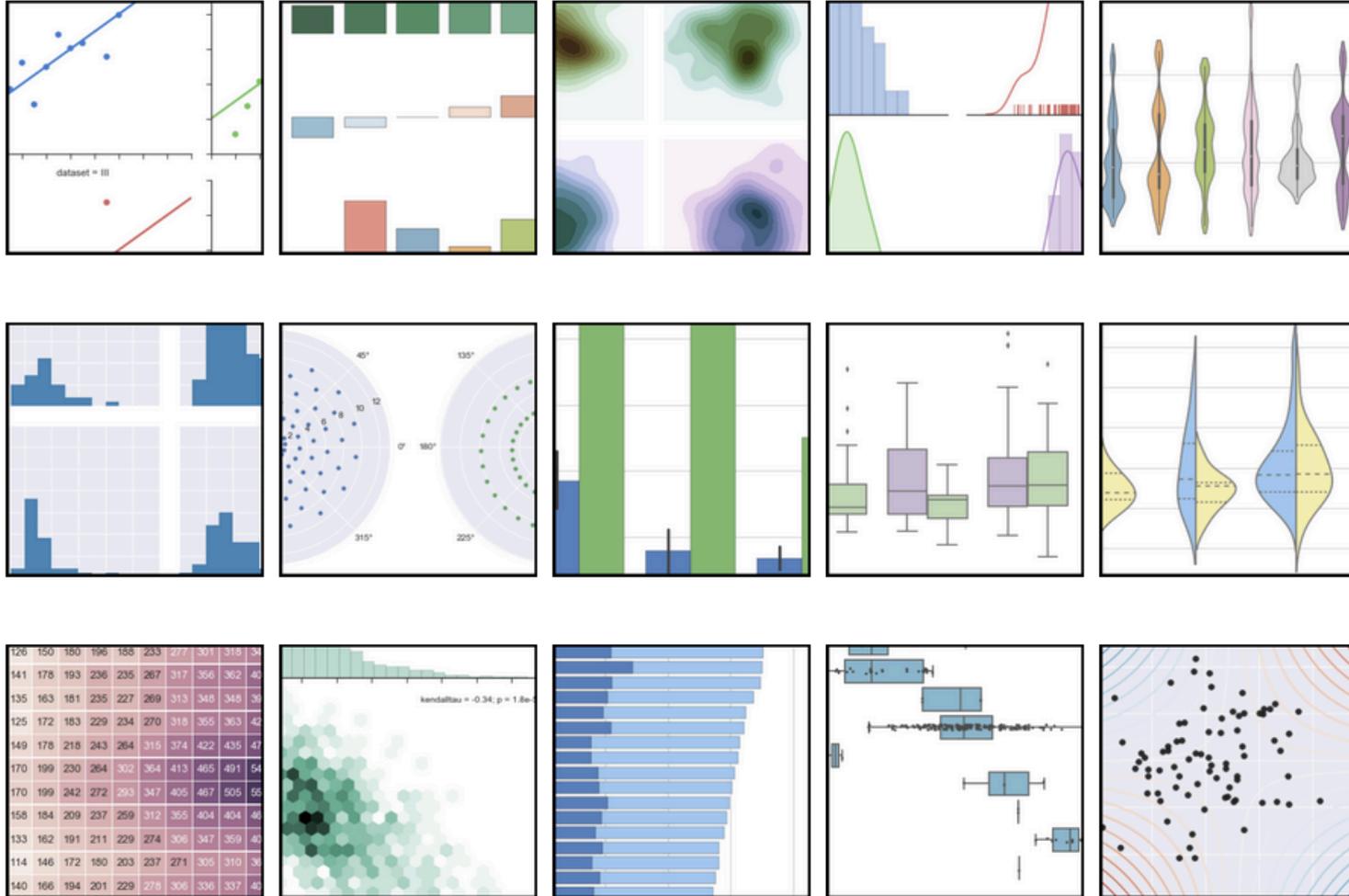
- Purchase price
- License over time
- License constraints (e.g., attribution, constraints on derivative works)





# New Tools of Note

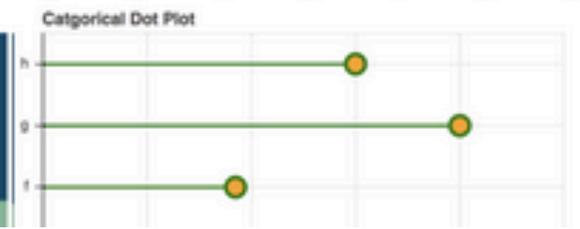
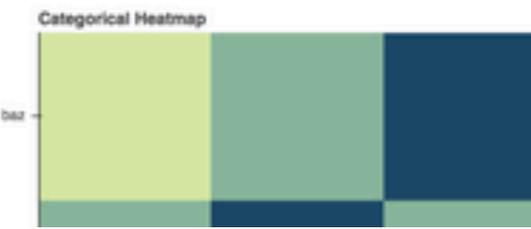
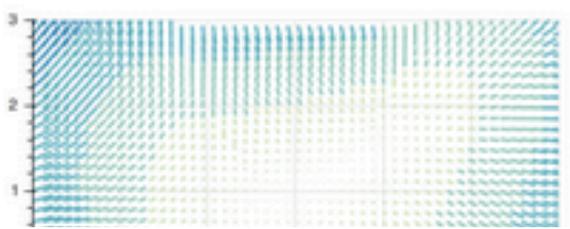
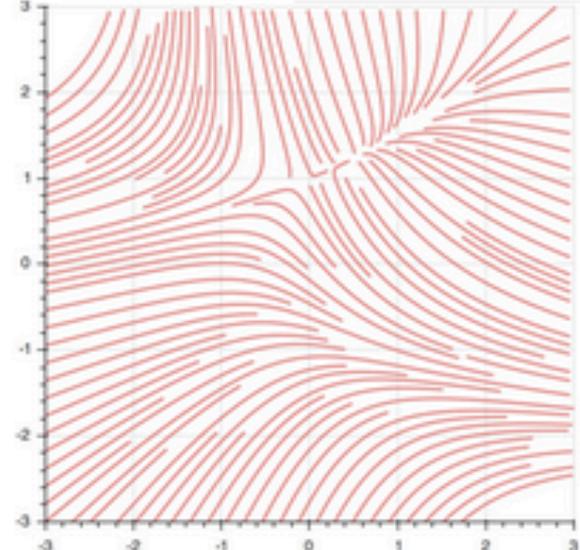
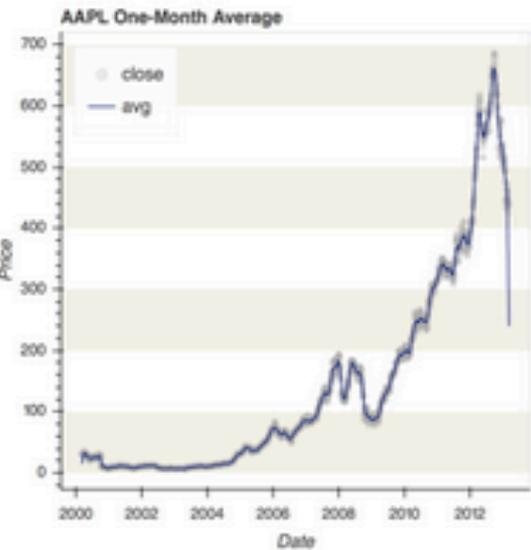
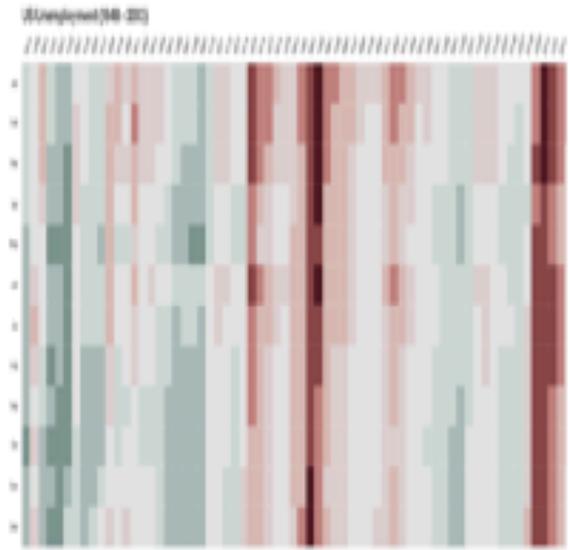
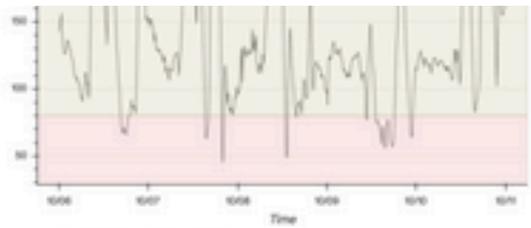
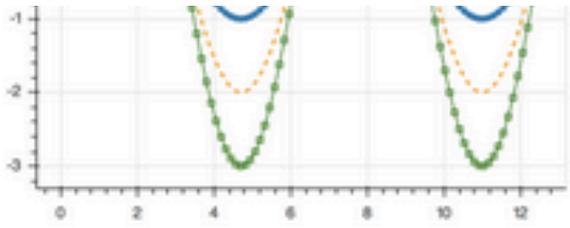
# Seaborn



126	150	150	196	155	233	277	301	318	34
141	178	193	236	235	267	317	356	362	40
135	163	181	235	227	269	313	348	345	38
125	172	163	229	234	270	318	355	363	42
149	178	218	243	264	315	374	422	435	47
170	199	230	264	302	364	413	465	491	54
170	199	242	272	293	347	405	467	505	55
158	184	209	237	259	312	355	404	404	46
133	162	191	211	229	274	306	347	359	40
114	146	172	180	203	237	271	305	310	36
140	166	194	201	229	278	306	336	337	46

- **Goal: exploratory**
- **Speed/Flexibility: Speed**
- **Dependencies: python 2.7 or 3.3+, numpy, scipy, matplotlib, pandas**
- **Sharing: export image files**
- **Specific Uses: statistical graphs**
- **Graphical Quality: good quality graphs, ability to tweak colors, axes, etc.**
- **Costs: open source**

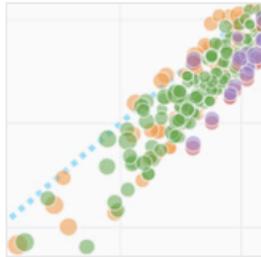
# Bokeh



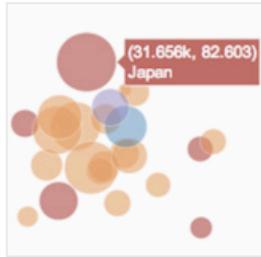
- **Goal: exploratory or explanatory**
- **Speed/Flexibility: low-, intermediate-, or high-level**
- **Dependencies: python**
- **Sharing: Bokeh server, embedding in web pages or notebooks**
- **Specific Uses: novel interactive visualizations in the browser**
- **Graphical Quality: very good quality**
- **Costs: open source**



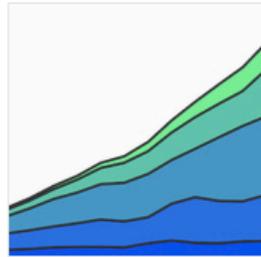
Dashboards



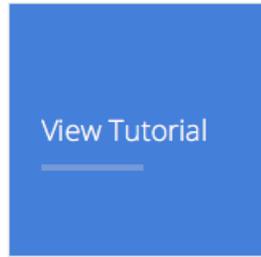
Line and Scatter Plots



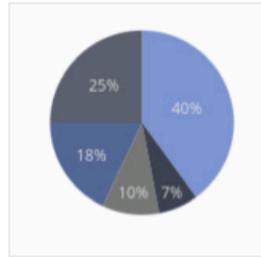
Bubble Charts



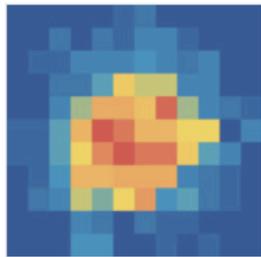
Filled Area Plots



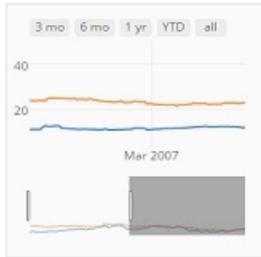
[Bar Charts](#)



Pie Charts



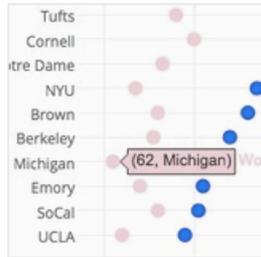
2D Histograms



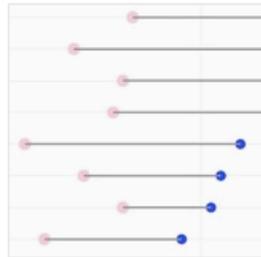
Range Sliders and Selectors



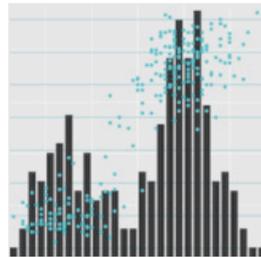
Gauge Chart



Dot plots



Dumbbell plots



Graphing Multiple Chart Types

**SCATTER PLOT**

Click the column headers to choose x and y columns to graph. Use the different colors to match x columns with y columns.

Scatter plot ▾

**OPTIONS**

Error Bars

Asymmetric Errors

Group By

Text

**INSERT INTO**

Make a new plot ▾

	Environment name ▾	Organism name ▾	method ▾	Project id ▾	name ▾	chemical formula ▾	neutral mas
x	choose as x	choose as x	choose as x	choose as x	choose as x	choose as x	choose as x
y	choose as y	choose as y	choose as y	choose as y	choose as y	choose as y	choose as y
G	choose as G	choose as G	choose as G	choose as G	choose as G	choose as G	choose as G
1	M94	The Environment	analysis	ENIGMA	tyrosine	C9H11NO3	181.073893
2	M94	The Environment	analysis	ENIGMA	betaine	C5H11NO2	117.078978
3	M94	The Environment	analysis	ENIGMA	monosacchari... (fructose, mannose, glucose, galactose, inositols)	C6H12O6	180.063388
4	M94	The Environment	analysis	ENIGMA	disaccharide 1 (sucrose, maltose, trehalose, cellobiose, lactose, etc)	C12H22O11	342.116211
5	M94	The Environment	analysis	ENIGMA	disaccharide 2 (sucrose, maltose, trehalose, cellobiose, lactose, etc)	C12H22O11	342.116211
6	M94	The Environment	analysis	ENIGMA	lactate	C3H6O3	90.0316940

- **Goal: exploratory**
- **Speed/Flexibility: Speed**
- **Dependencies: library for python, R, Matlab, Excel, or JS, or use web app**
- **Sharing: Plot.ly, Plotly Server**
- **Specific Uses: collaborative analysis**
- **Graphical Quality: good quality**
- **Costs: plotly libs are open source, Plotly Server is licensed**



**MULTIDIMENSIONAL SCALING**

Similarity analysis tool.



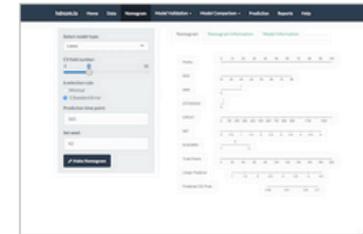
**INTERACTIVE REPORTING**

Interactive report with Shiny and R Markdown



**GEOMETRY OF CLASSIFIERS**

Comparison of machine learning algorithms.



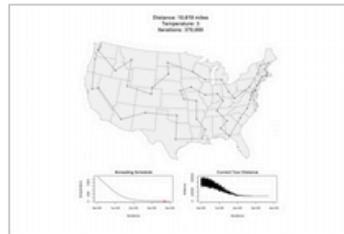
**NOMOGRAM GENERATOR**

Tune analysis then export as pdf, HTML, or Word file.



**RADIANT**

Extensive app for teaching business analytics. ([documentation](#))



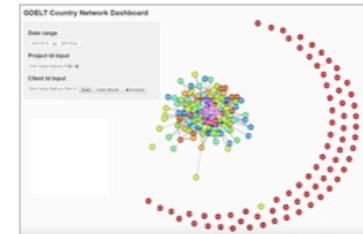
**TRAVELING SALESMAN**

Optimization fun.



**FILE CONVERTER**

Upload a data file, then download in various formats.



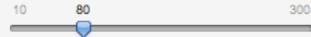
**GDELT & BIGQUERY**

Dashboard to [GDELT](#) database with quarter-billion records

## Movie explorer

### Filter

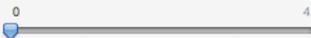
Minimum number of reviews on Rotten Tomatoes



Year released



Minimum number of Oscar wins (all categories)



Dollars at Box Office (millions)

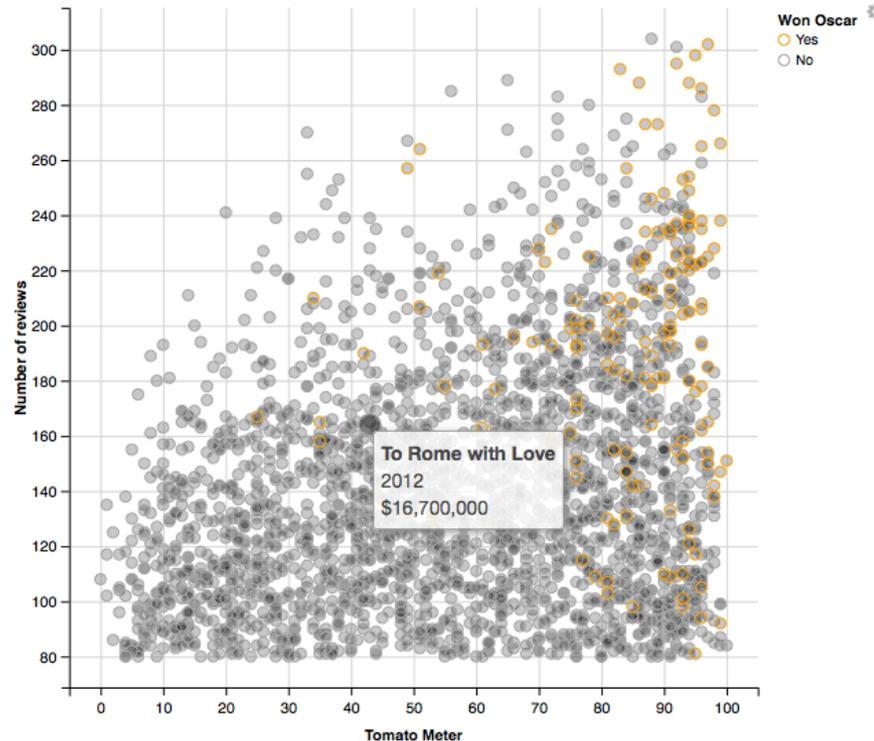


Genre (a movie can have multiple genres)

All

Director name contains (e.g., Miyazaki)

Cast names contains (e.g., Tom Hanks)



Number of movies selected:  
2557

- **Goal: exploratory or explanatory**
- **Speed/Flexibility: Speed**
- **Dependencies: R or R Studio**
- **Sharing: shinyapps.io, Shiny Server**
- **Specific Uses: interactive web applications**
- **Graphical Quality: good quality**
- **Costs: open source, freemium model**

# Questions?

---



I have some for you . . .

# Questions?

---



- **What tools do you guys use (on or off HPC)?**

# Questions?

---



- **What tools do you guys use (on or off HPC)?**
- **What would you like to see available at NERSC?**

# Questions?

---



- **What tools do you guys use (on or off HPC)?**
- **What would you like to see available at NERSC?**
- **Do you currently do data vis on HPC systems?**

# Questions?

---



- **What tools do you guys use (on or off HPC)?**
- **What would you like to see available at NERSC?**
- **Do you currently do data vis on HPC systems?**
- **If no, why not? Are there roadblocks we can remove?**

# Questions?

---



- **What tools do you guys use (on or off HPC)?**
- **What would you like to see available at NERSC?**
- **Do you currently do data vis on HPC systems?**
- **If no, why not? Are there roadblocks we can remove?**
- **Are there other issues with currently available software for data vis?**



**National Energy Research Scientific Computing Center**