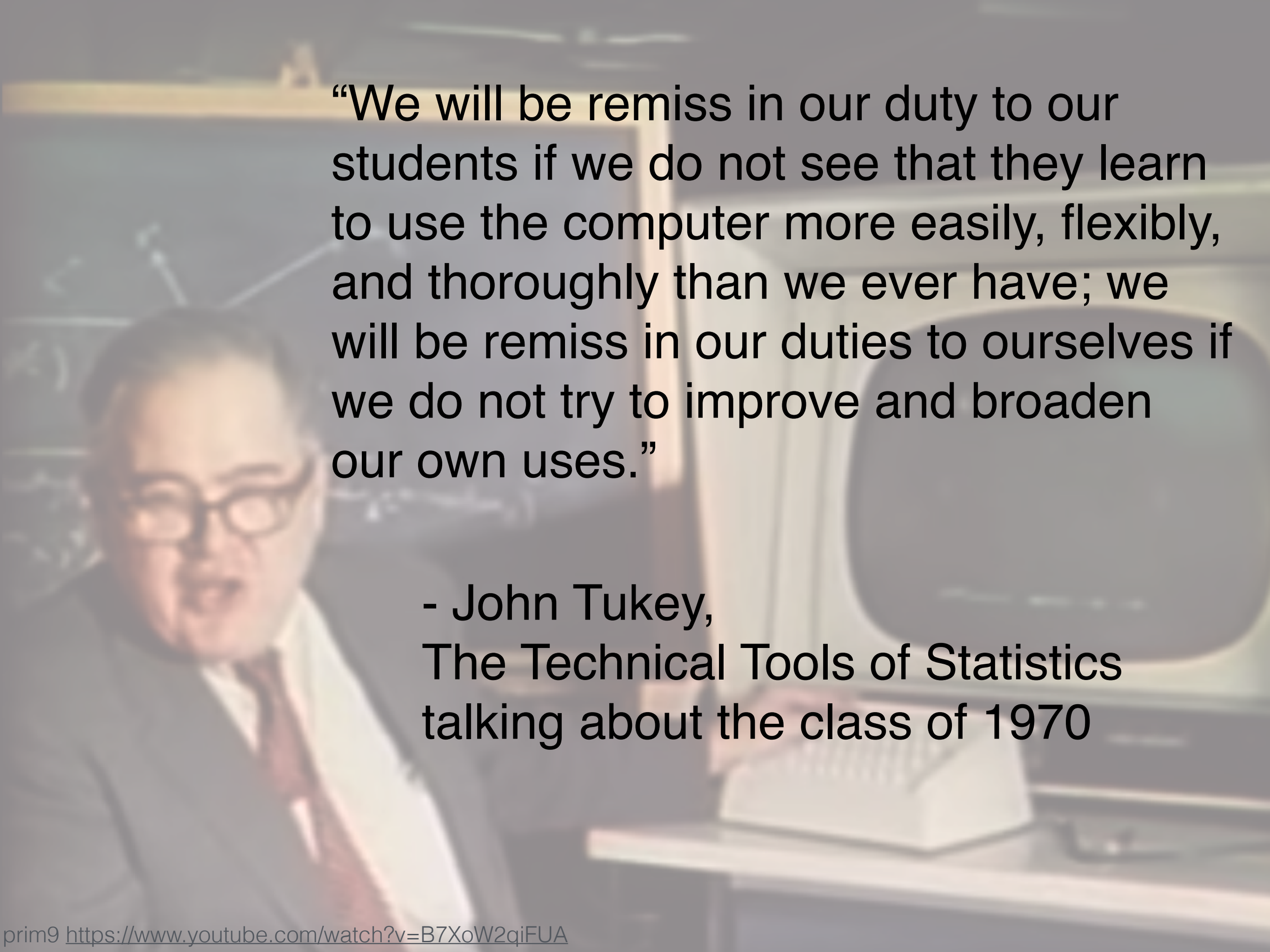
The gap between tools for learning and for doing statistics

Amelia McNamara
Visiting Assistant Professor of Statistical and Data Sciences
Smith College
@AmeliaMN

flickr: eulothg

- Tools for teaching and learning statistics
- Tools for doing statistics
- Beginning to bridge the gap
- The future



"We will be remiss in our duty to our students if we do not see that they learn to use the computer more easily, flexibly, and thoroughly than we ever have; we will be remiss in our duties to ourselves if we do not try to improve and broaden our own uses."

- John Tukey,
The Technical Tools of Statistics
talking about the class of 1970

prim9 https://www.youtube.com/watch?v=B7XoW2qiFUA

Tools for teaching and learning statistics

TinkerPlots™
DYNAMIC DATA EXPLORATION

Fathom®
Dynamic Data Software

StatCrunch

Rossman/Chance Applet Collection

Data Analysis

- Descriptive Statistics (js)
- Guess the Correlation (js)
- Least Squares Regression (js)

Sampling Distribution Simulations

- Reeses Pieces (js)
- Sampling Words (js)
- Sampling from a Finite Population (js)
- Sampling from a Probability Model (j)
- Sampling Regression Lines - Population Model (j)
- Simulating Confidence Intervals for Population Parameter (js)
- Improved Batting Averages (Power) (js)
- ANOVA simulation (js)

Classics (j)

- Histogram Bin Width
- Dotplot Summaries
- Sampling Pennies
- Sampling Change
- Sampling 2005 Senators
- Friendly Observers
- Dolphin Study applet
- Yawning Study applet
- Two-way Table simulation applet
- Randomization Test for quantitative response (two groups) (f)
- Simulating Confidence Intervals for Population Parameter
- Simulating t-intervals for different population shapes
- Random Babies

Probability

- Random Babies (js)
- Secretary Problem (j)
- Normal Probability Calculator (js)
- Randomizing Subjects (js)
- Random number generator (js)

Statistical Inference

- One proportion inference (js)
- Analyzing Two-way Tables (js)
- Matched Pairs (js)
- Randomization test for quantitative response (multiple groups) (js)
  - two means
- Randomization test for categorical response (multiple groups) (js)
  - Dolphin Study applet
- Analyzing Two Quantitative Variables (js)
- Theory-based Inference (js)

Click here to access old applets page

j = java applet (click here for help on running java on macs, pc)
js = javascript
f = flash

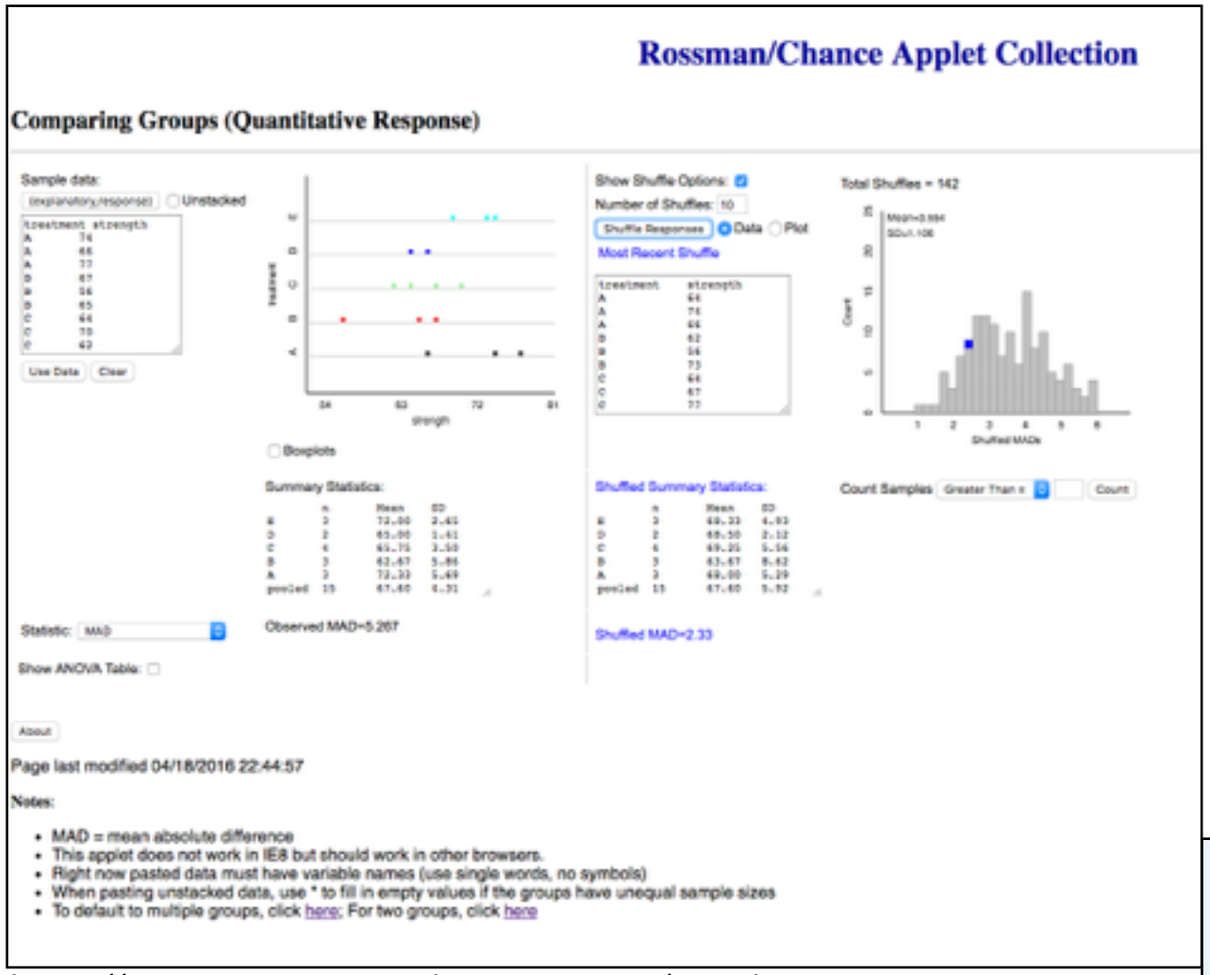
StatKey to accompany *Statistics: Unlocking the Power of Data* by Lock, Lock, Lock, Lock, and Lock

Descriptive Statistics and Graphs	Bootstrap Confidence Intervals	Randomization Hypothesis Tests
One Quantitative Variable	CI for Single Mean, Median, St.Dev.	Test for Single Mean
One Categorical Variable	CI for Single Proportion	Test for Single Proportion
One Quantitative and One Categorical Variable	CI for Difference in Means	Test for Difference in Means
Two Categorical Variables	CI for Difference in Proportions	Test for Difference in Proportions
Two Quantitative Variables	CI for Slope, Correlation	Test for Slope, Correlation

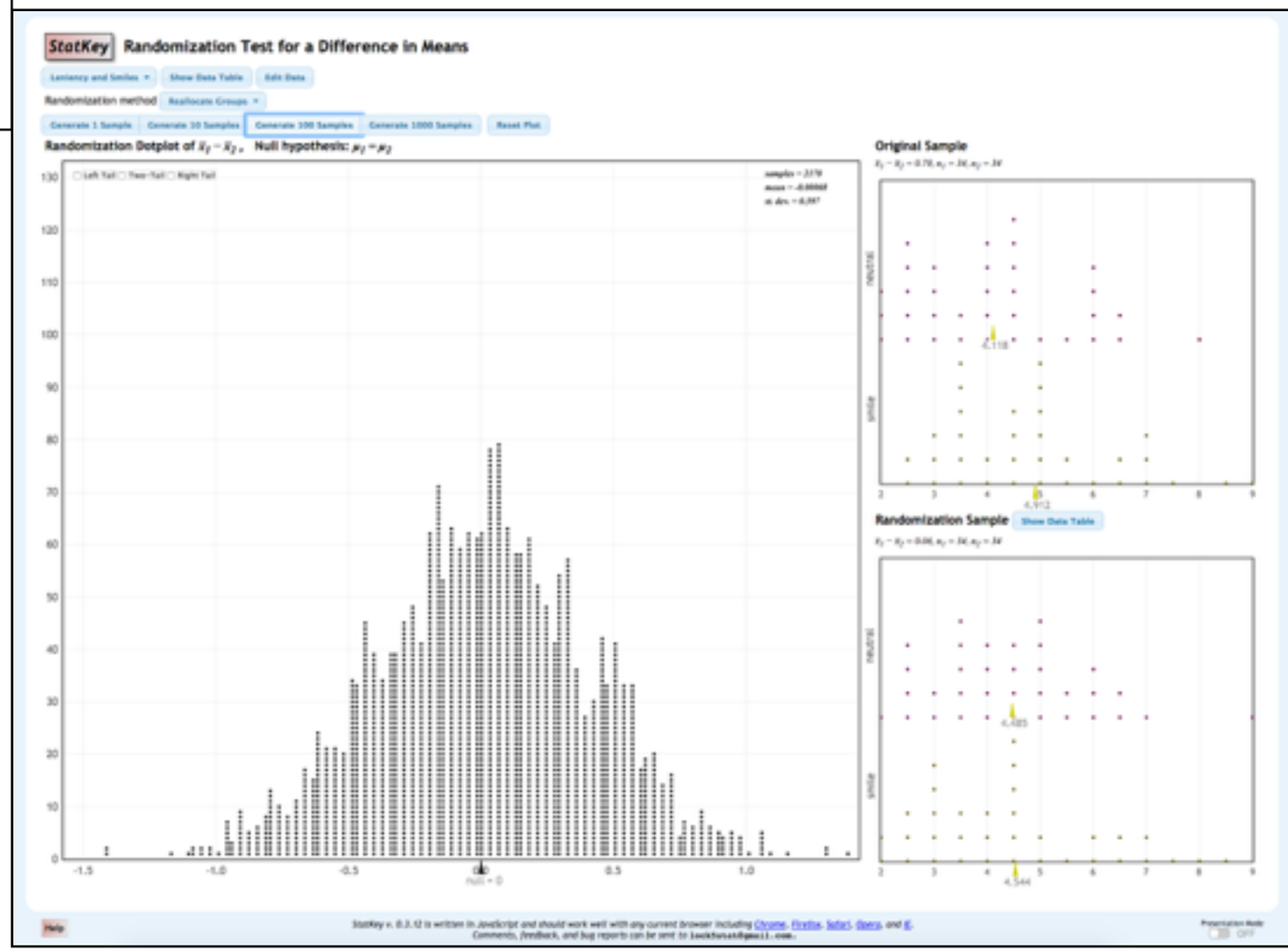
Sampling Distributions	Mean	Proportion

Theoretical Distributions	Normal	t	χ^2	F

More Advanced Randomization Tests	χ^2 Goodness-of-Fit	χ^2 Test for Association	ANOVA for Difference in Means	ANOVA for Regression



http://www.rossmanchance.com/applets



http://lock5stat.com

Result Properties

Thumbnail:

Owner: websterwest

Created: Sep 21, 2015

Size: 122KB

Share: yes

Views: 1950

Data set for this result:

Metropolitan Statistical Areas in the U.S. - Population, Location

By statcrunchhelp
On Oct 14, 2014

Recently shared results for this data set:

The changing population of U.S. Metropolitan Areas

By scsurvey
On Aug 28, 2015

Percent Change in population from 2010 to 2013 for U.S. Metro Areas

By statcrunchhelp
On Oct 14, 2014

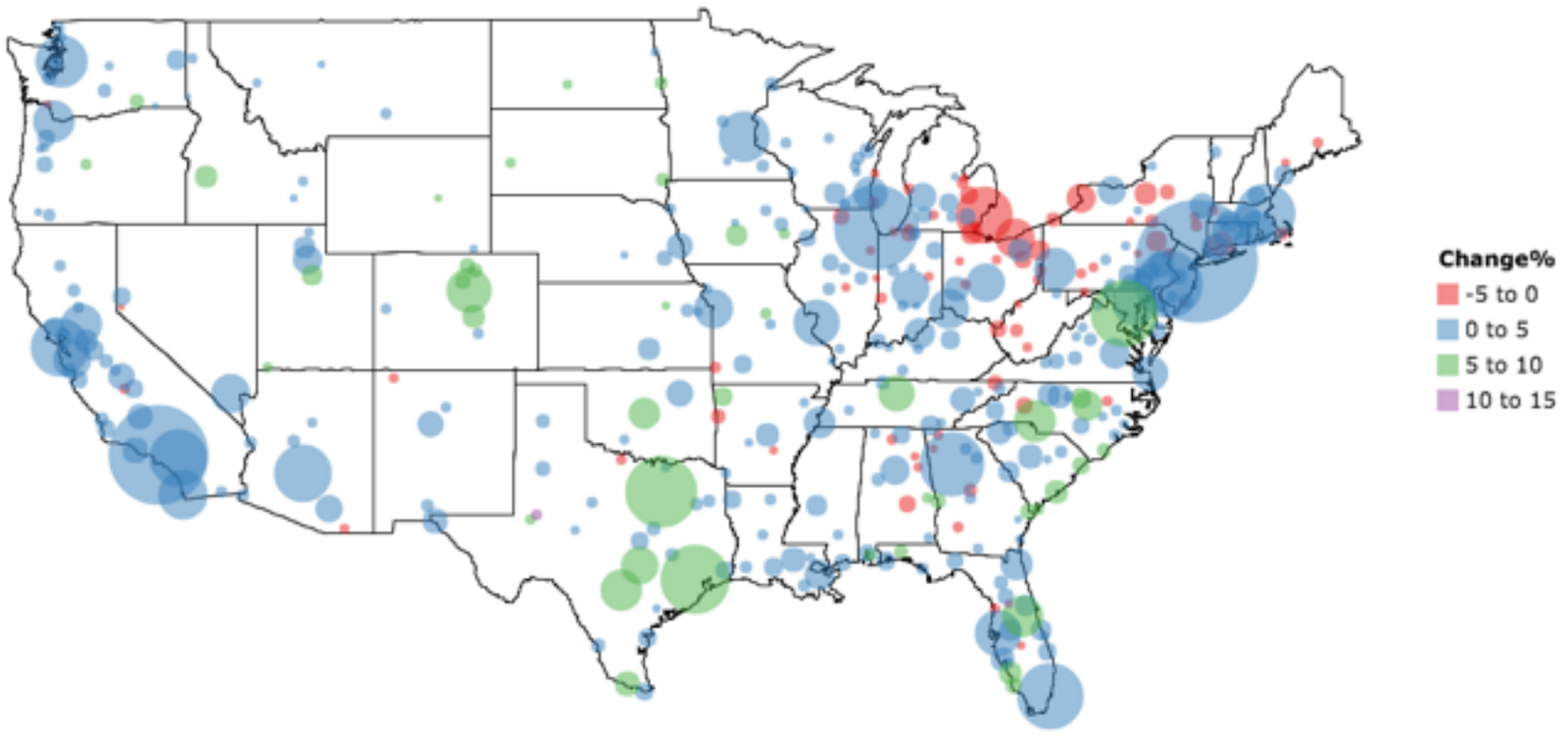
View all shared >

Reports with this result:

None

Visualizing the population change in US Metro areas

Copy Print Mail Link Embed Twitter Facebook



Tags:
map

HTML link:
<A href="http://www.statcrunch.com/5.0/viewresult.php?resid=1802998">Visualizing the population change in US Metro areas</A>

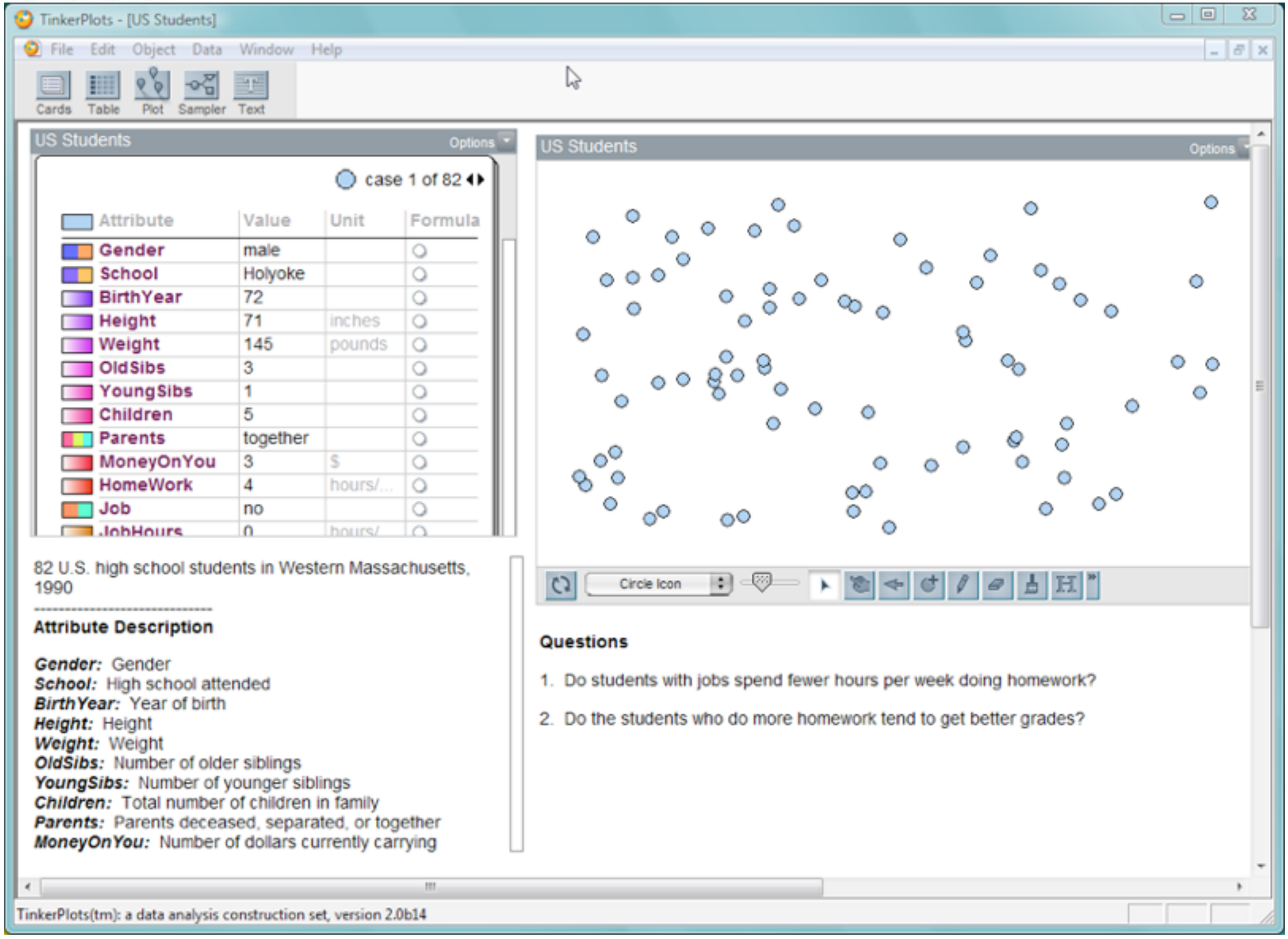
Comments

Want to comment? Subscribe
Already a member? Sign in.

http://www.statcrunch.com

TinkerPlots™

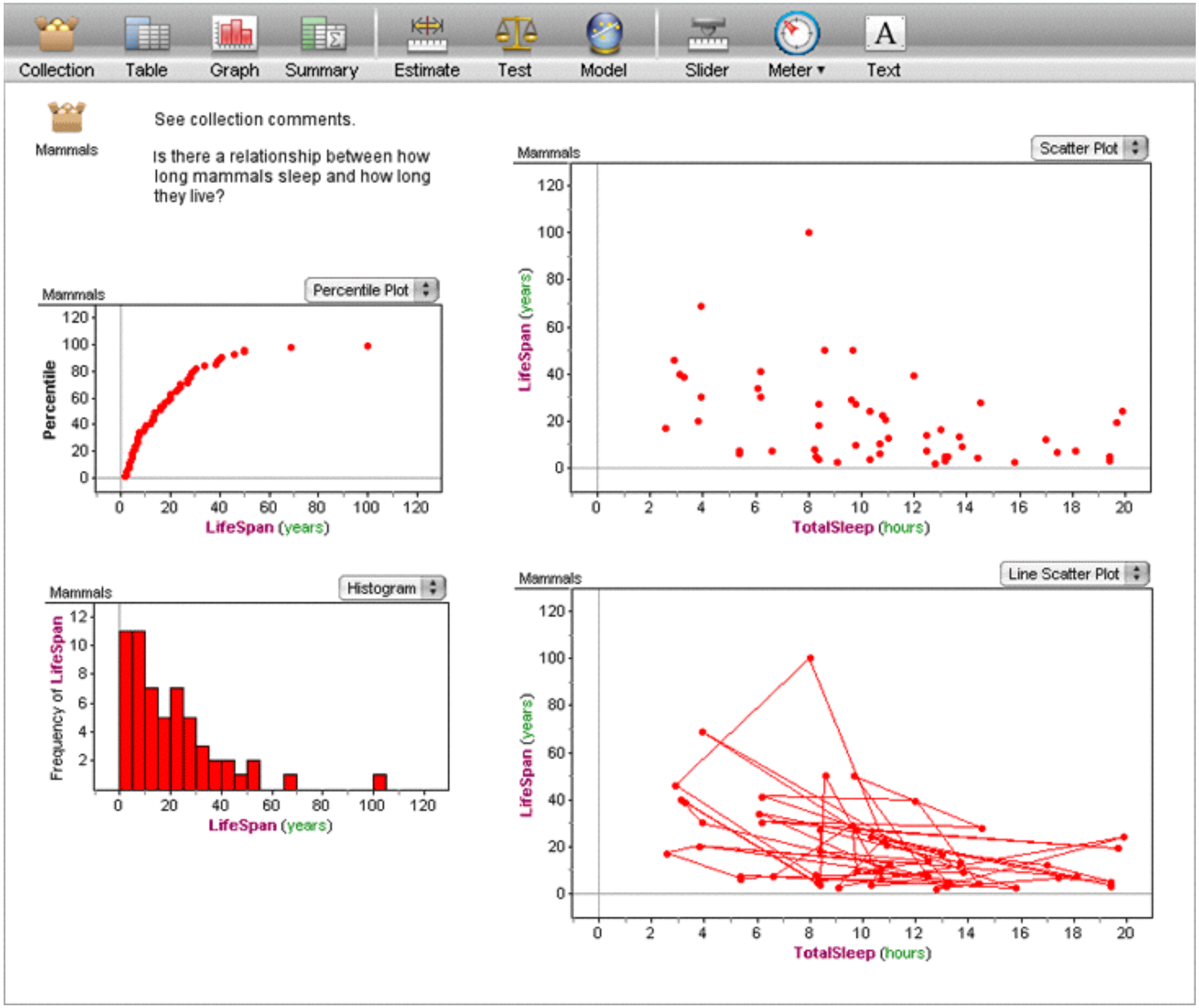
DYNAMIC DATA EXPLORATION



https://www.tinkerplots.com/

Fathom®

Dynamic Data Software



http://fathom.concord.org/

Tools for teaching and learning statistics are typically

- Inexpensive
- Accessible
- Interactive
- Curated

Tools for doing statistics

STATA®

R

SPSS®

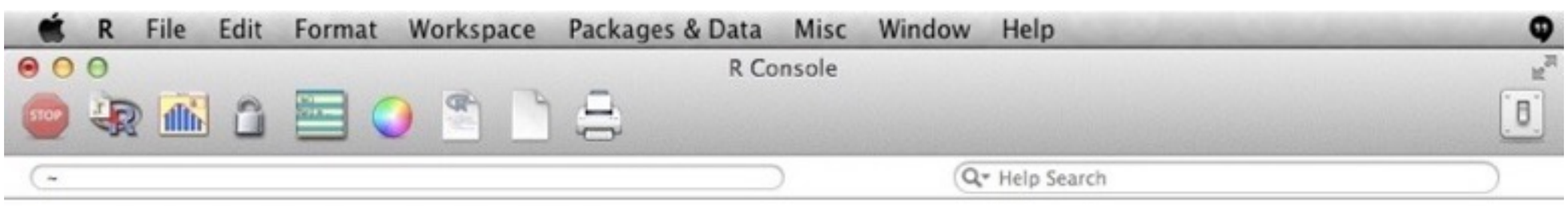
sas

MATLAB®

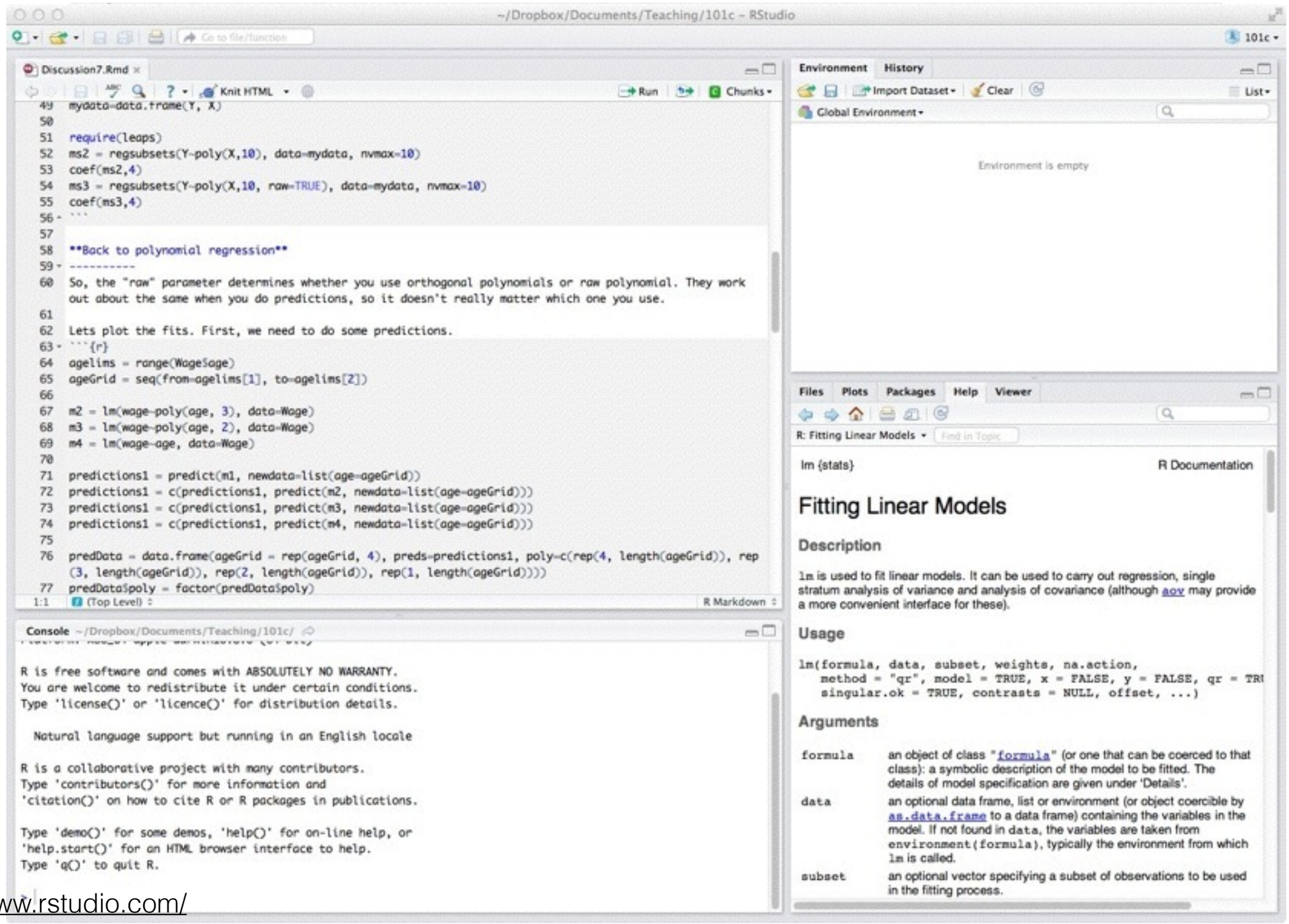
Excel

```
Terminal  Shell  Edit  View  Window  Help
amelia — R — 80×24
Last login: Wed Apr 16 15:39:29 on ttys000
Amelias-MacBook-Air:~ amelia$ R

R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin10.8.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

https://cran.r-project.org/



https://cran.r-project.org/



https://www.rstudio.com/

Tools for doing statistics are often

- Expensive
- Static
- Flexible
- Extensible
- Reproducible


flickr: josef42marbach
What's in between?

Shiny

by RStudio

Movie explorer

Filter

Minimum number of reviews on Rotten Tomatoes

10 80 300

Year released

1,940 1,970 2,014

Minimum number of Oscar wins (all categories)

0 4

Dollars at Box Office (millions)

0 800

Genre (a movie can have multiple genres)

All

Director name contains (e.g., Miyazaki)

Cast names contains (e.g. Tom Hanks)

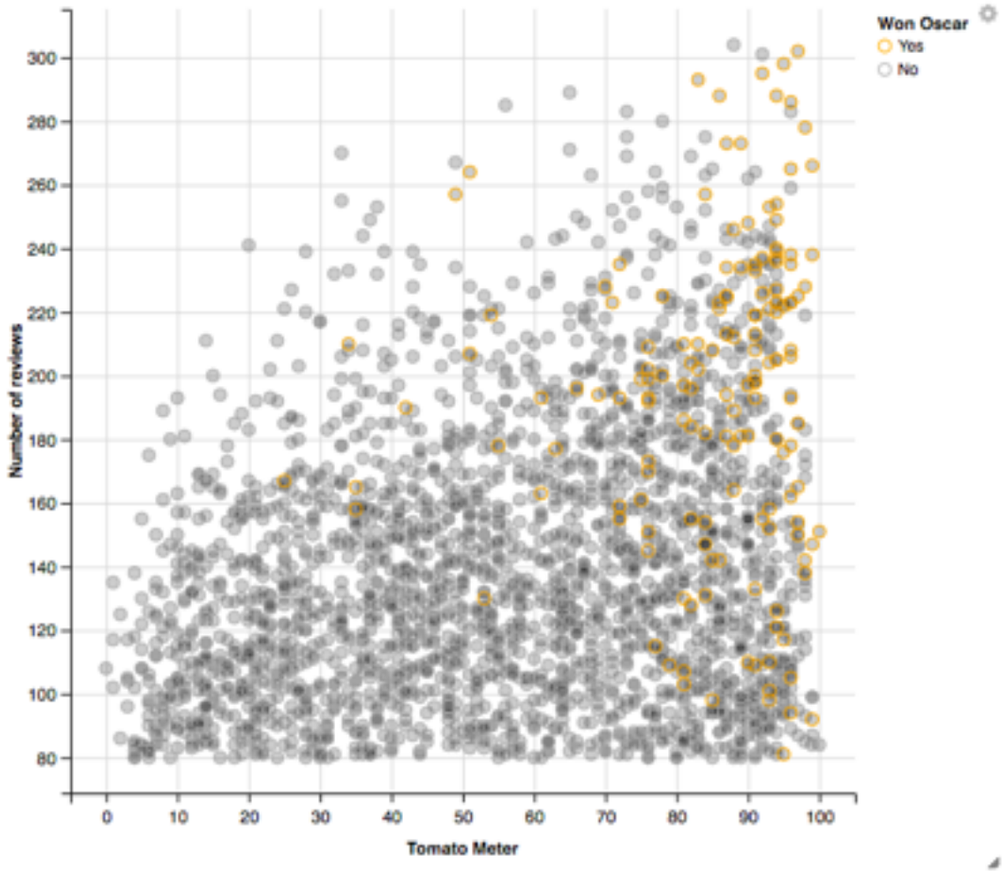
X-axis variable

Tomato Meter

Y-axis variable

Number of reviews

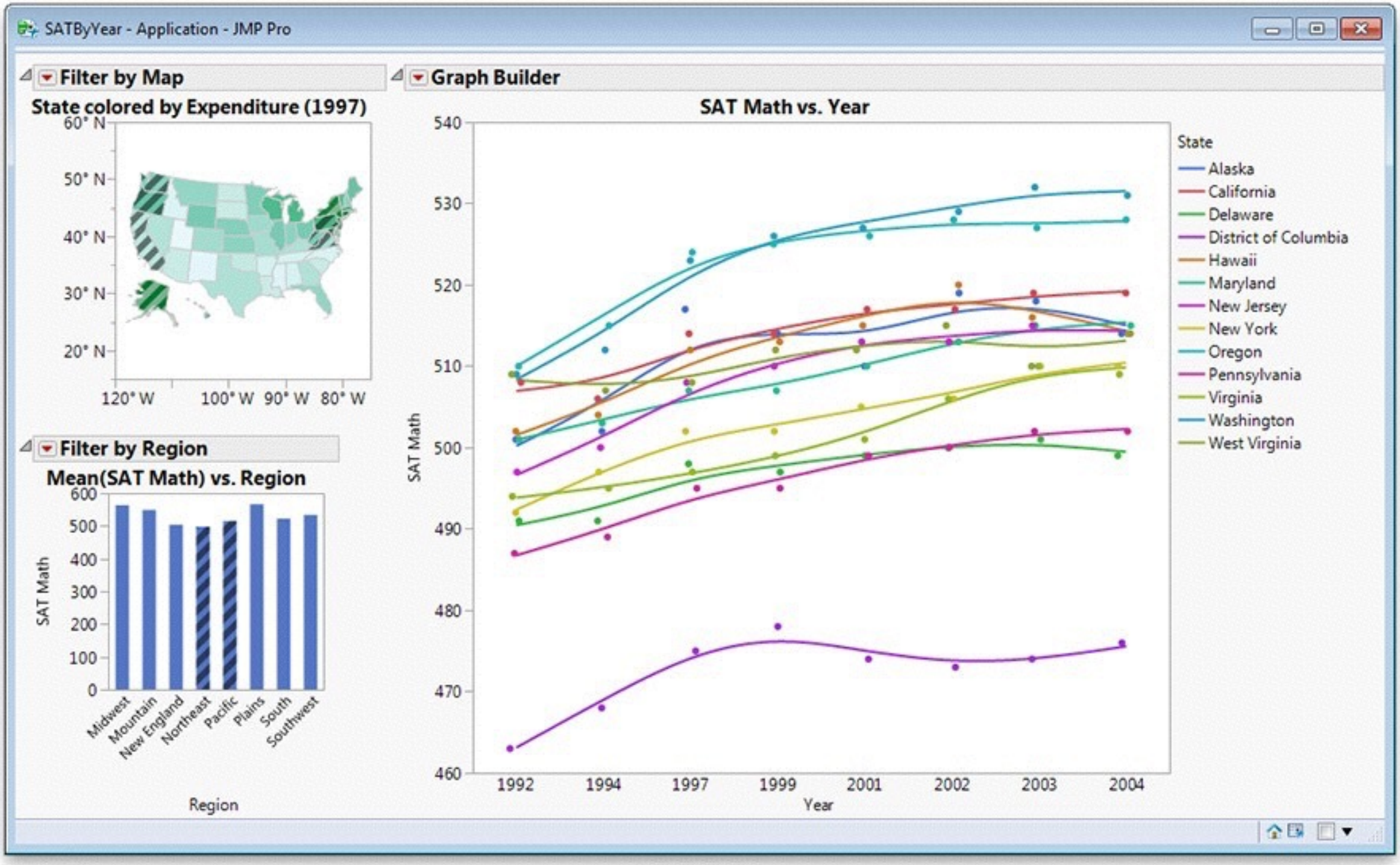
Note: The Tomato Meter is the proportion of positive reviews (as judged by the Rotten Tomatoes staff), and the Numeric rating is a normalized 1-10 score of those reviews which have star ratings (for example, 3 out of 4 stars).



Number of movies selected:
2557

http://shiny.rstudio.com/

Statistical Discovery.™ From SAS.

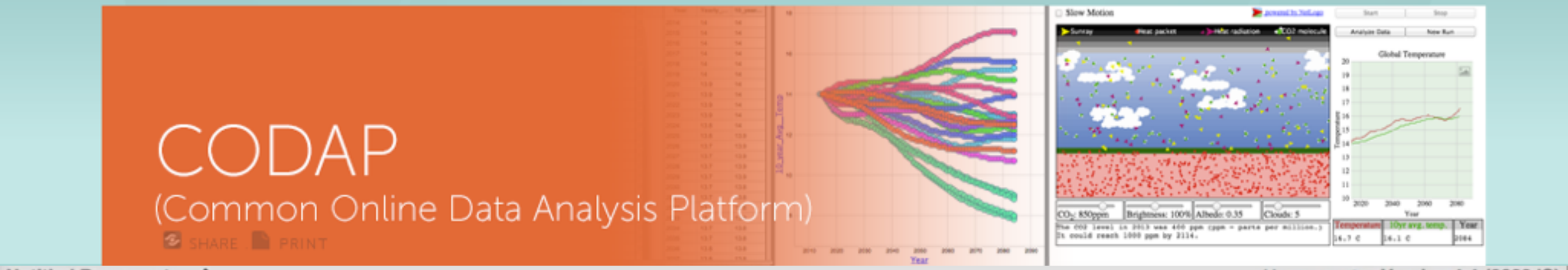


http://www.jmp.com/

CODAP

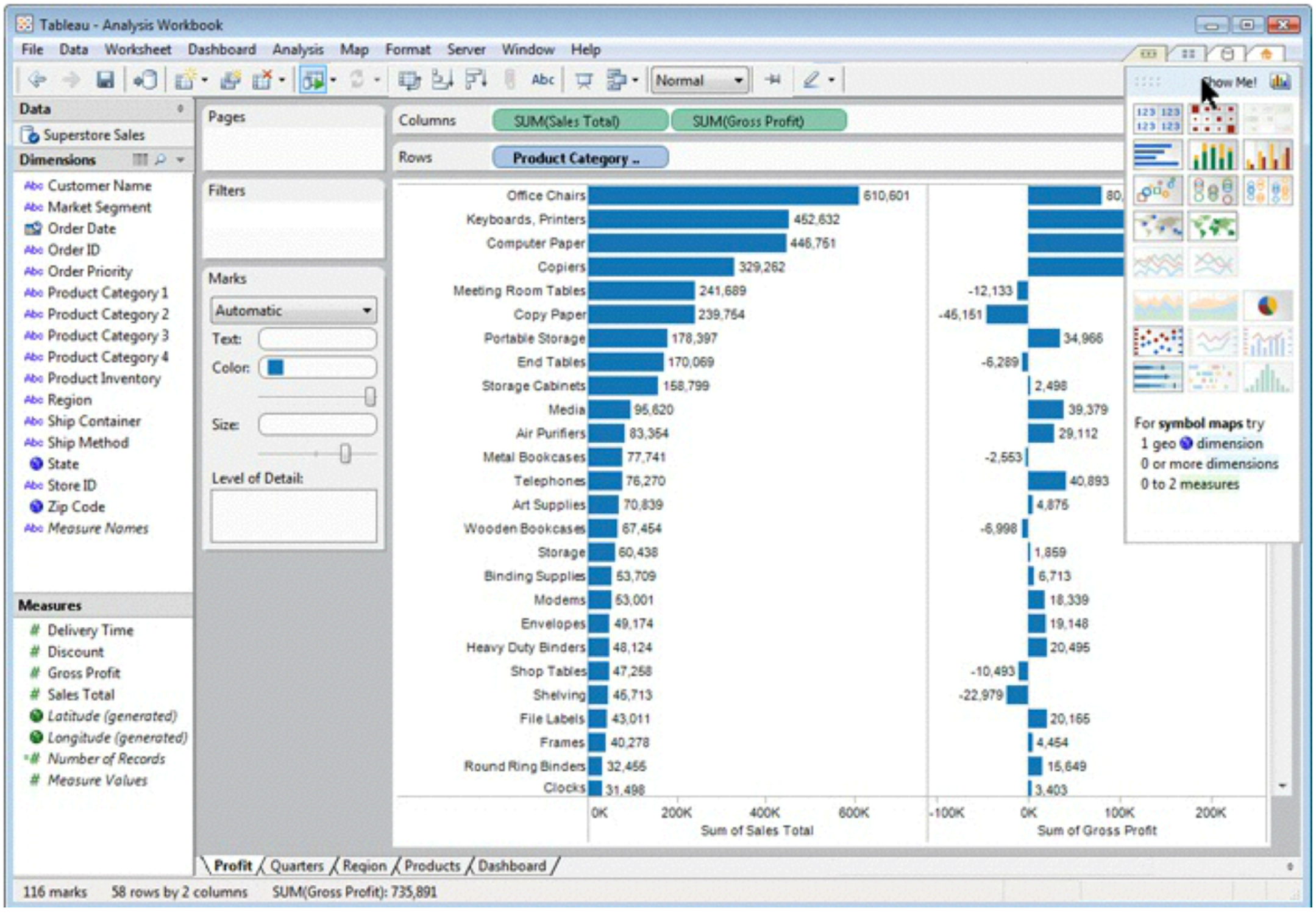
(Common Online Data Analysis Platform)

SHARE PRINT



http://concord.org/projects/codap

+ a b | e a u®



http://www.tableau.com/

DATA PIPELINES

Characters

From Source: characters

Data

name	Myriel	Napoleon	Mlle.Baptistine
group	1	1	1
index	0	1	2

1-20 of 77

+ NEW TRANSFORM

Scales

Fill Color | Size | Font Size

+ NEW PIPELINE

VISUALIZATION

Layer 1

AXES

MARKS

Labels

Nodes

Type

Pipeline

Characters + VISUAL LAYOUTS

Force-Directed Layout

LINKS DATA

Data connections

Source: source

Target: target

LINKS STROKE

Line Type: line

Color: #ccc

Width: 0.5

CONNECTIONS

Distance	Strength	Tension
80	1	0

NODES

Charge	Friction	Gravity
-220	0.9	0.1

Output: x y weight

Properties

POSITION

X: x

Y: y

GEOMETRY

Shape: circle

Size: Size weight

FILL

Color: Fill Color group

Opacity: 0.3

STROKE

Color: Fill Color group

Width: 0.75

THE LYRA VDE beta

http://idl.cs.washington.edu/projects/lyra/

TRIFACTA

Generate

24 Columns 345 Rows 2 Data Types Grid Filter in grid

	#.# AZ_Phoenix	#.# CA_Los_Angeles	#.# CA_San_Francisco	#.#
	65 - 228	59 - 273	47 - 219	47 - 16
	PHXR-SA	LXXR-SA	SFXR-SA	DNXR-
		59.43	46.96	50.3
		59.89	47.3	50.1
		60.4	47.84	50.3
		61.32	47.98	50.7
		62.03	48.31	50.7
		62.78	48.61	50.4
		63.46	49.08	50.2
		64.13	49.54	50.1

The Transformer

1 of 5

This is **the Transformer**, where you can transform your messy data into clean data.

The Transformer is populated with a sample of your dataset. Let's discover what's in your data!

Learn more

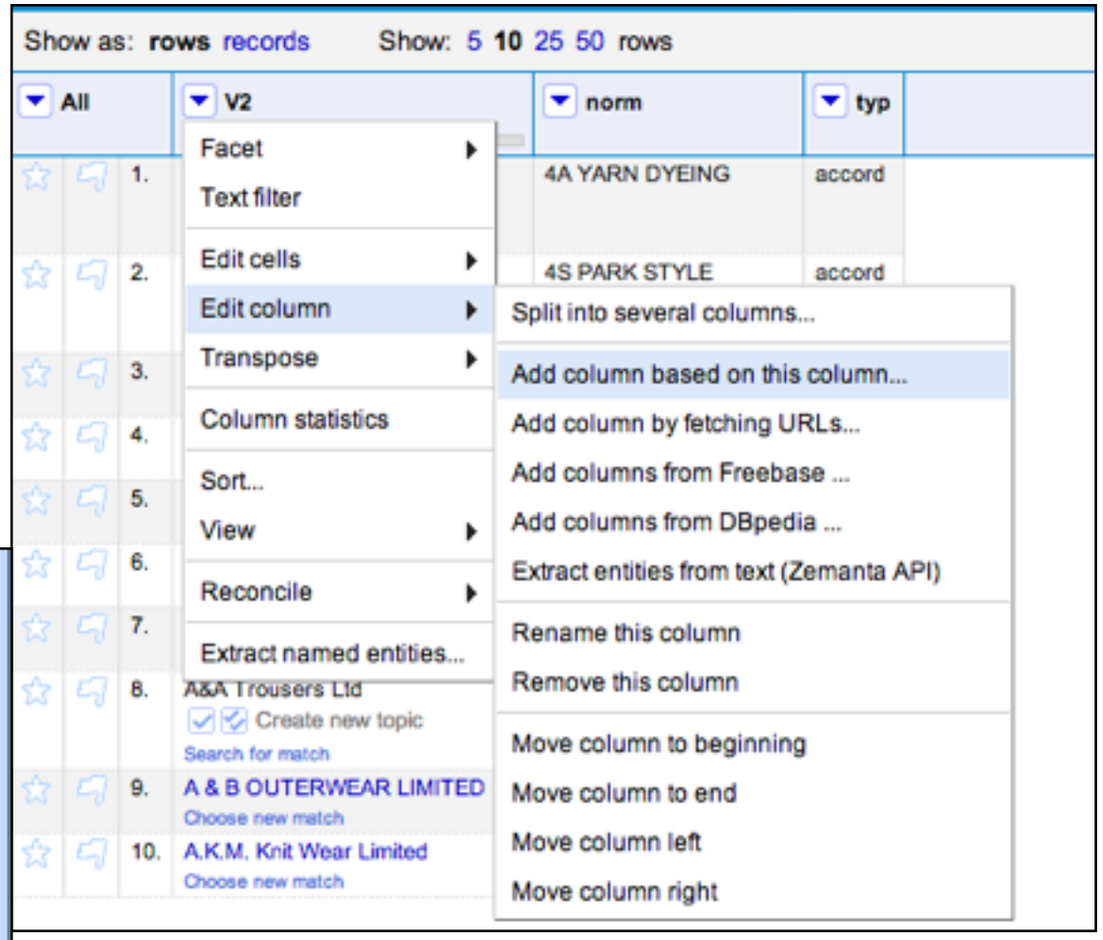
Don't show me any helpers

Next

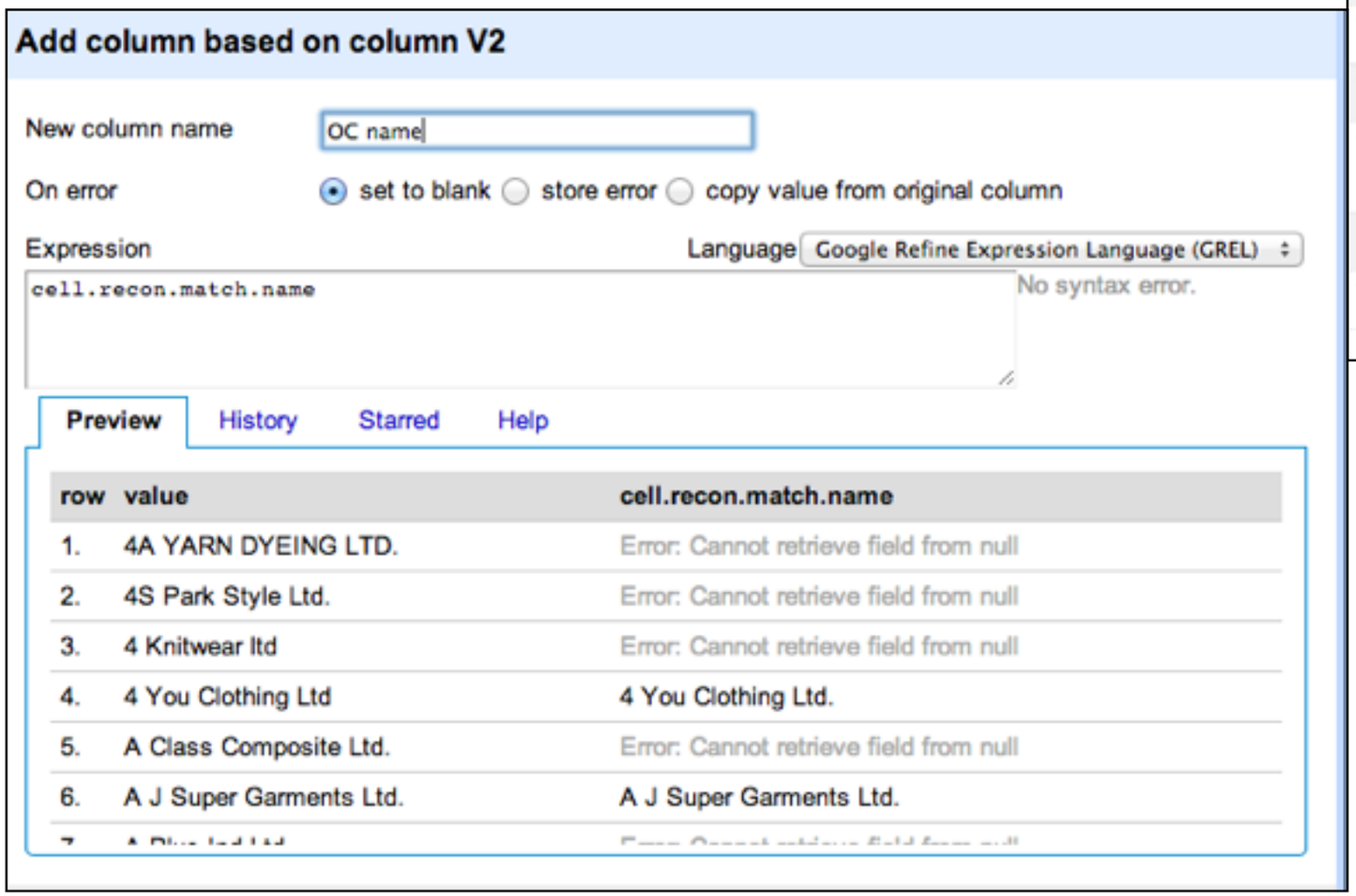
R

expression

https://www.trifacta.com/



flickr: psychemedia



http://openrefine.org/

Reaching across the gap

Tools for learning should aim for

- Reproducibility
- Shareability

Tools for doing should aim for

- Interactivity
- Accessibility

wikipedia: Sydney Harbour Bridge

Selected References

- Biehler, Rolf. (1997). Software for Learning and for Doing Statistics. *International Statistical Review*,65(2). http://onlinelibrary.wiley.com/doi/10.1111/j.1751-5823.1997.tb00399.x/abstract

- Biehler, Rolf., Dani Ben-Zvi, Arthur Bakker and Katie Makar. (2012). “Technology for enhancing statistical reasoning at the school level." *Third International Handbook of Mathematics Education*. http://link.springer.com/chapter/10.1007/978-1-4614-4684-2_21

- McNamara, Amelia. (2015). Bridging the Gap Between Tools for Learning and For Doing Statistics. http://escholarship.org/uc/item/1mm9303x

- Tukey, John. (1965). The Technical Tools of Statistics. *The American Statistician*, 19(2). http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1965.10479711#.VzOMAxUrJBw

Questions? Comments? Thoughts?

amcnamara@smith.edu

@AmeliaMN

flickr: hugo90