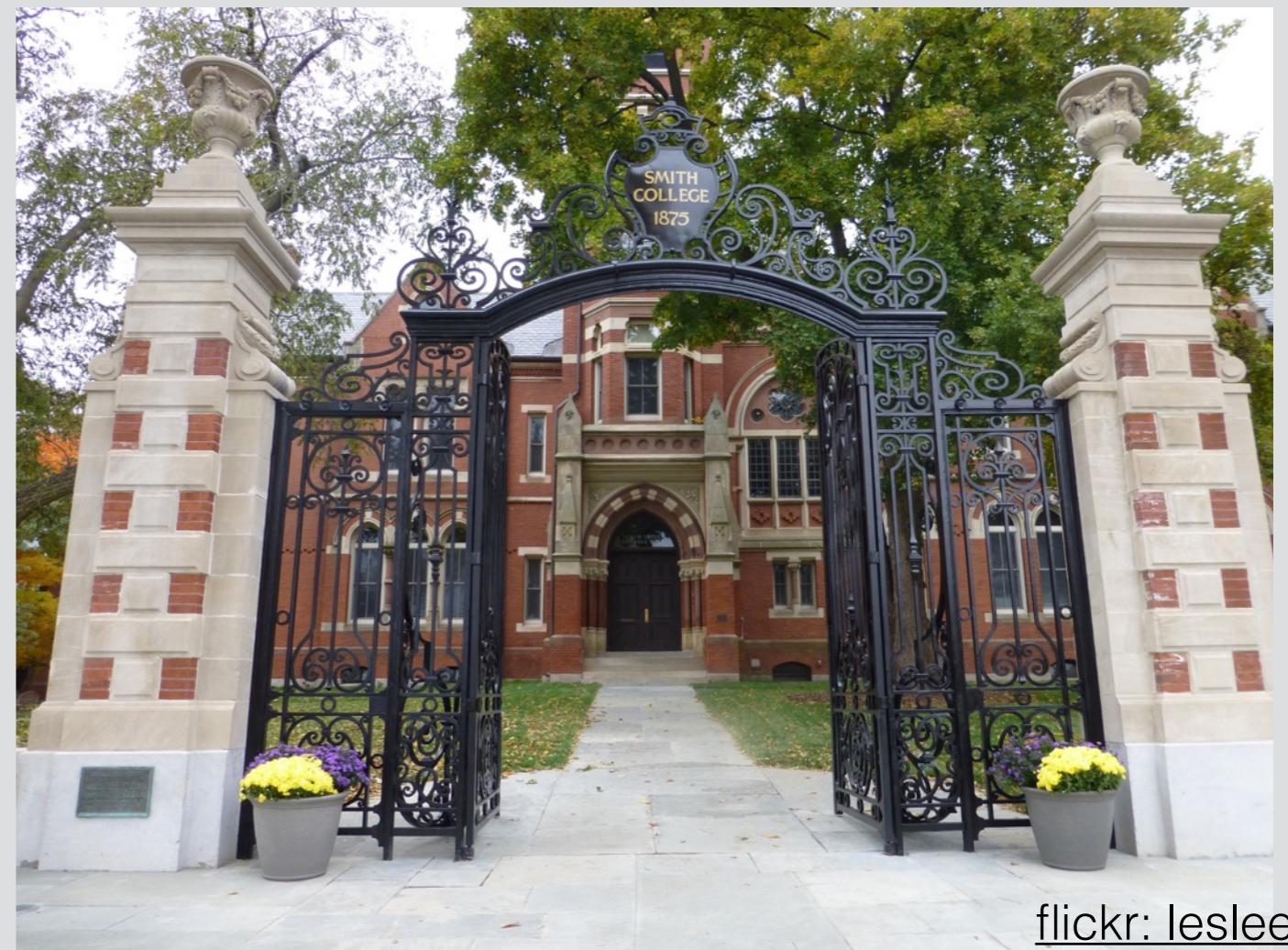


The importance of parameter manipulation for statistics

Amelia McNamara [@AmeliaMN](https://twitter.com/AmeliaMN)
Program in Statistical & Data Sciences
Smith College

About me

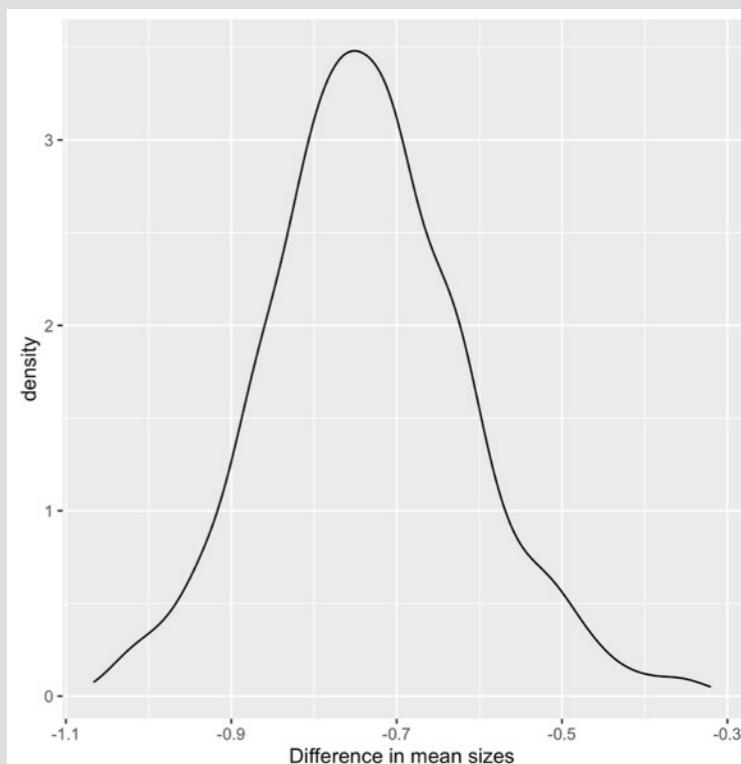
- BA in Mathematics and English from Macalester College
- PhD in Statistics from UCLA
- Visiting Assistant Professor of Statistical and Data Sciences at Smith College
- Research at the intersection of statistical computing, statistics education, data visualization



[flickr: leslee](#)

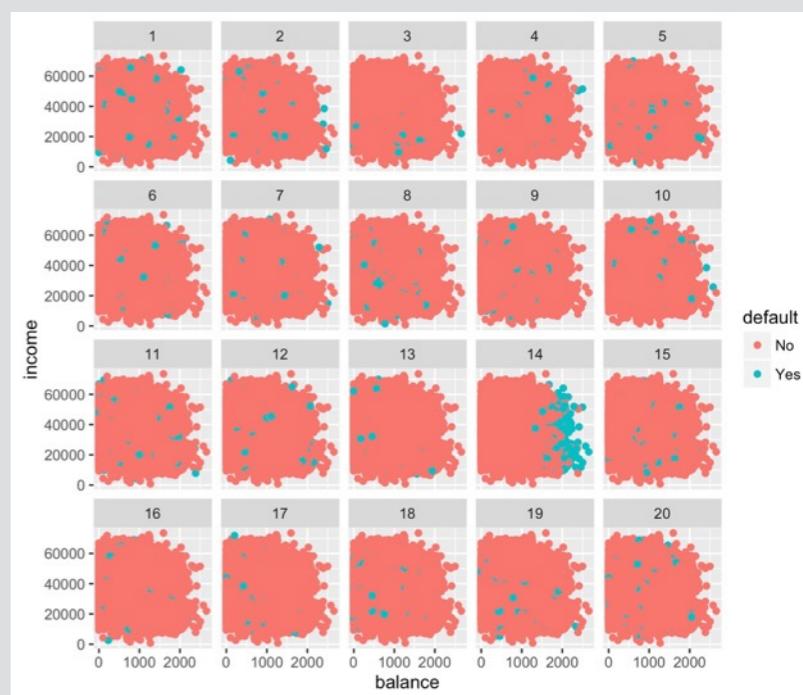
Statistics seeks to answer general questions:

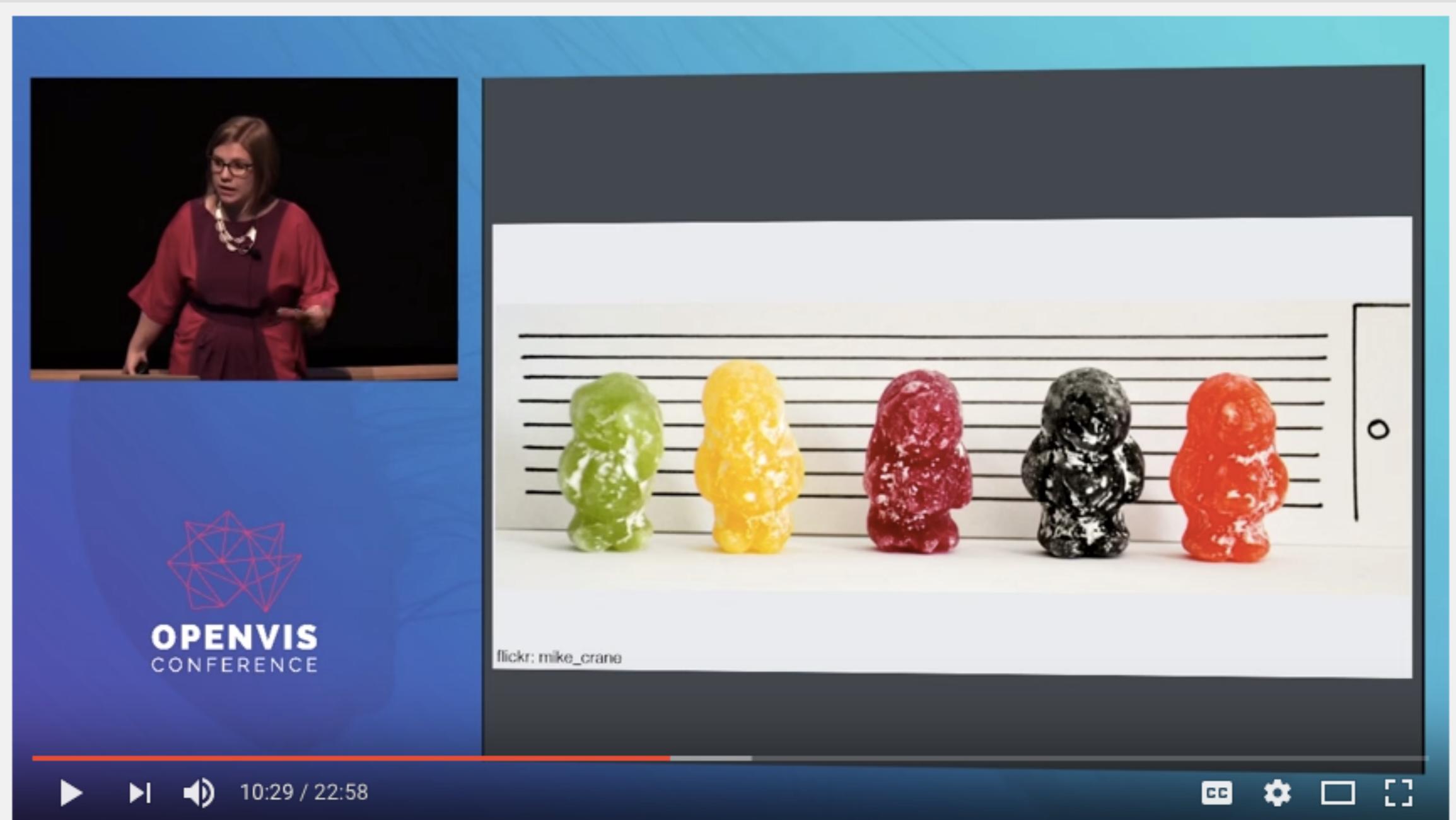
- Is this number different than zero?
- How sure am I about this number?



Similar questions in visualization:

- Is this visualization different than “nothing”?
- How sure am I about the story I see?





Do you know Nothing when you see it – Amelia McNamara



Bocoup LLC

Subscribe

3,279

279 views

The video player displays a presentation slide with the title "SAD: the folk music of uncertainty vis". Below the title is a black and white photograph of three people: a man with a beard and a mustache, a woman with long blonde hair, and another person partially visible. The man is holding a microphone. In front of the video frame, there is a blue graphic for the "OPENVIS CONFERENCE" featuring a red geometric logo.

SAD: the folk music of uncertainty vis

Complicated and abstract

OPENVIS CONFERENCE

◀ ▶ ⏸ 🔍 12:59 / 30:59 CC ⚙️

The Visual Uncertainty Experience - Jessica Hullman



Bocoup LLC

[Subscribe](#) 3,279

317 views

The future of statistical programming tools

- Provide easy entry for novice users
- Data as first-order persistent object
- Support a cycle of exploratory and confirmatory analysis
- Allow users to develop new visual representations
- Support randomization throughout
- Provide interactivity at every level
- Provide inherent documentation
- Encourage documentation, publishing and reproducibility
- Be modular and extensible

McNamara, Amelia Ahlers.
(2015). Bridging the Gap
Between Tools for Learning and
for Doing Statistics.

<http://bit.ly/BridgingTheToolGap>

UNIVERSITY OF CALIFORNIA
Los Angeles

Bridging the Gap Between Tools for Learning
and for Doing Statistics

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Statistics

by

Amelia Ahlers McNamara

2015

Parameter manipulation in existing tools

R/ggplot2

lab-intro.Rmd x R data sets x R data sets x

ncbirths North Carolina births
oscars Oscar winners, 1929 to 2012
poker Poker winnings during 50 sessions
possum possum
prRace08 Election results for the 2008 U.S.
president Presidential race
run10 United States Presidential History
run10Samp Cherry Blossom 10 mile run data, 2009
run10Samp Cherry Blossom 10 mile run data, 2009
run10_09 Cherry Blossom 10 mile run data, 2009
satGPA SAT and GPA data
senateRace10 Election results for the 2010 U.S. Senate
races
smoking UK Smoking Data
textbooks Textbook data for UCLA Bookstore and Amazon
tgSpending Thanksgiving spending, simulated based on
Gallup poll.
tips Tip data
unempl Annual unemployment since 1890

Use 'data(package = .packages(all.available = TRUE))'
to list the data sets in all *available* packages.

Console ~/ Attaching package: 'openintro'

The following object is masked from 'package:mosaic':
dotPlot

The following object is masked from 'package:datasets':
cars

```
> ggplot(tgSpending) + geom_histogram(aes(x=spending))
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
> |
```

Environment History Import Dataset Global Environment Data tgSpending 436 obs. of 1 variable

Files Plots Packages Help Viewer Publish

A histogram titled 'tgSpending' showing the distribution of spending. The x-axis is labeled 'spending' and ranges from 0 to 300. The y-axis is labeled 'count' and ranges from 0 to 50. The distribution is right-skewed, with the highest frequency occurring between 50 and 75.

Fathom

untitled

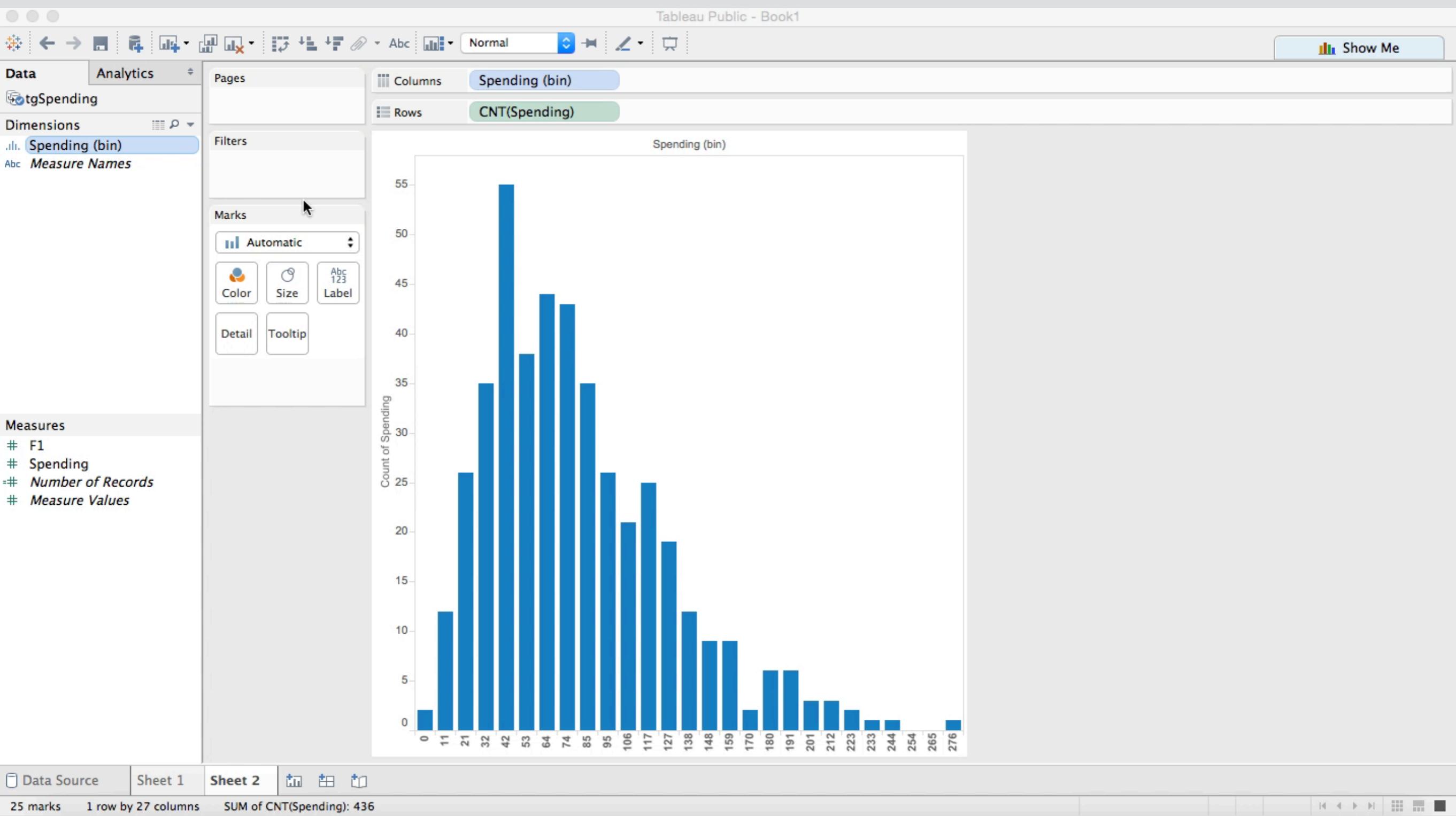
Collection Table Graph Summary Estimate Test Model Slider Meter Text

tgSpending.csv

	Attr1	Attr2	<new>	
429	429	50.1943		
430	430	41.7233		
431	431	203.553		
432	432	92.1924		
433	433	52.1701		
434	434	21.2527		
435	435	66.7804		
436	436	149.905		

Fathom Dynamic Data Software

Tableau



manipulate

RStudio

Project: (None)

lab-intro.Rmd R data sets R data sets Untitled1* ×

Source on Save Run Source

```
1 manipulate(
2   ggplot(tgSpending) + geom_histogram(aes(x=spending), binwidth=x),
3   x = slider(0,100, initial=10)
4 )
5
6
7 manipulate(
8   histogram(tgSpending$spending, breaks=slider(0,20, initial=10))
9 )
10
11
12 manipulate(
13   histogram(tgSpending$spending, breaks=x),
14   x = slider(4,20)
15 )
```

12:12 (Top Level) R Script

Console ~/

```
> manipulate(
+   histogram(tgSpending$spending, breaks=x),
+   x = slider(4,20)
+ )
> manipulate(
+   ggplot(tgSpending) + geom_histogram(aes(x=spending), binwidth=x),
+   x = slider(0,100, initial=10)
+ )
> manipulate(
+   ggplot(tgSpending) + geom_histogram(aes(x=spending), binwidth=x),
+   x = slider(0,100, initial=10)
+ )
> |
```

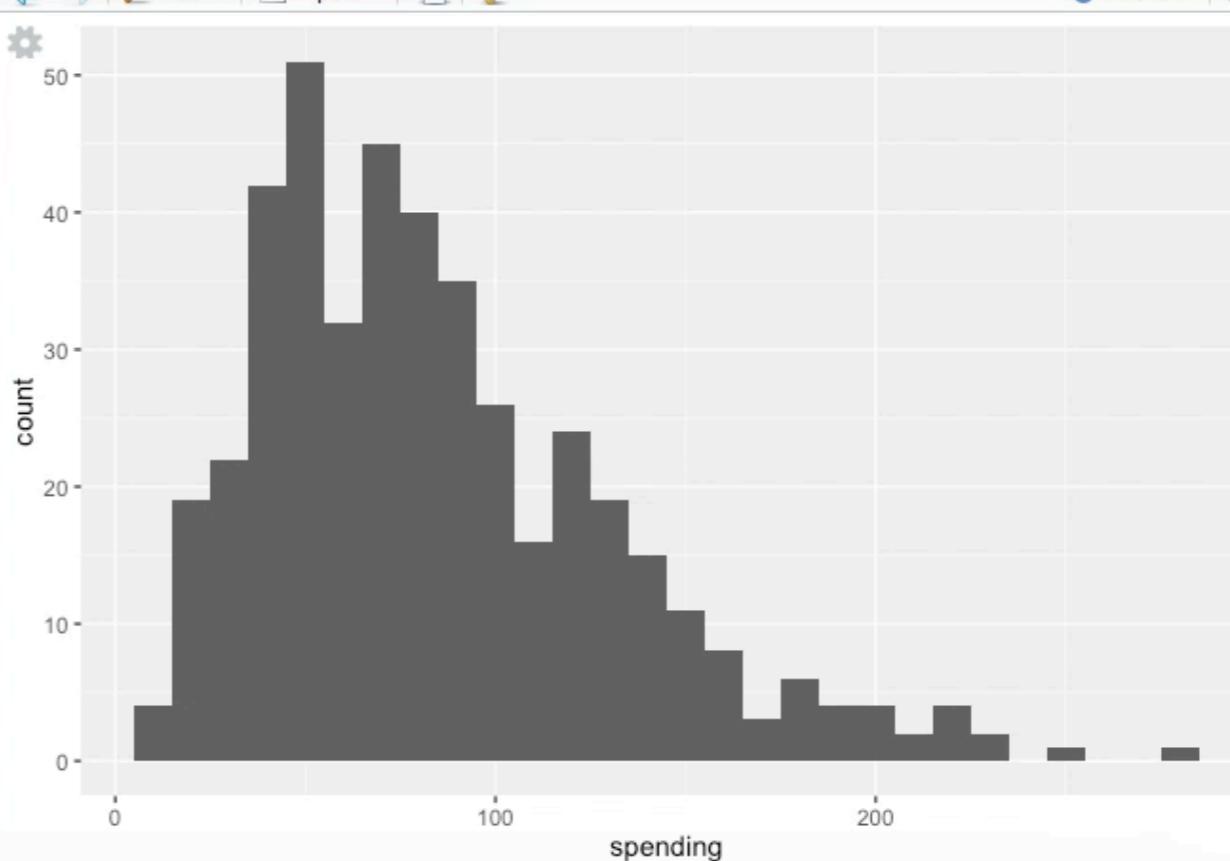
Environment History

Import Dataset Global Environment

Data tgSpending 436 obs. of 1 variable

Files Plots Packages Help Viewer

Zoom Export Publish



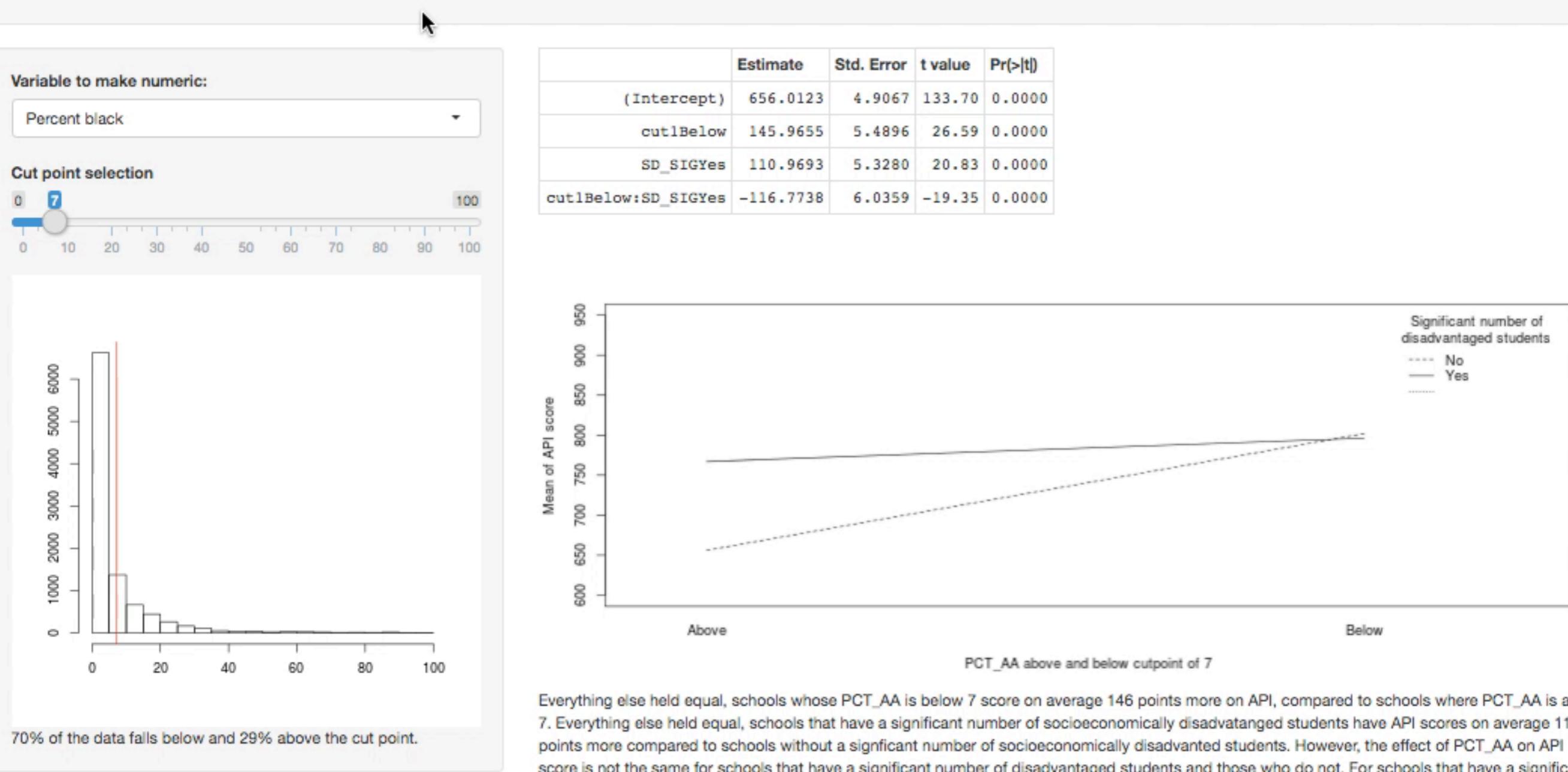
A histogram showing the distribution of spending. The x-axis is labeled 'spending' and ranges from 0 to 250. The y-axis is labeled 'count' and ranges from 0 to 50. The distribution is right-skewed, with the highest frequency occurring between 50 and 75.

Bin Range (spending)	Count (count)
0 - 25	~5
25 - 50	~18
50 - 75	~52
75 - 100	~42
100 - 125	~35
125 - 150	~45
150 - 175	~40
175 - 200	~35
200 - 225	~25
225 - 250	~15
250 - 275	~10
275 - 300	~5
300 - 325	~2
325 - 350	~1

Shiny

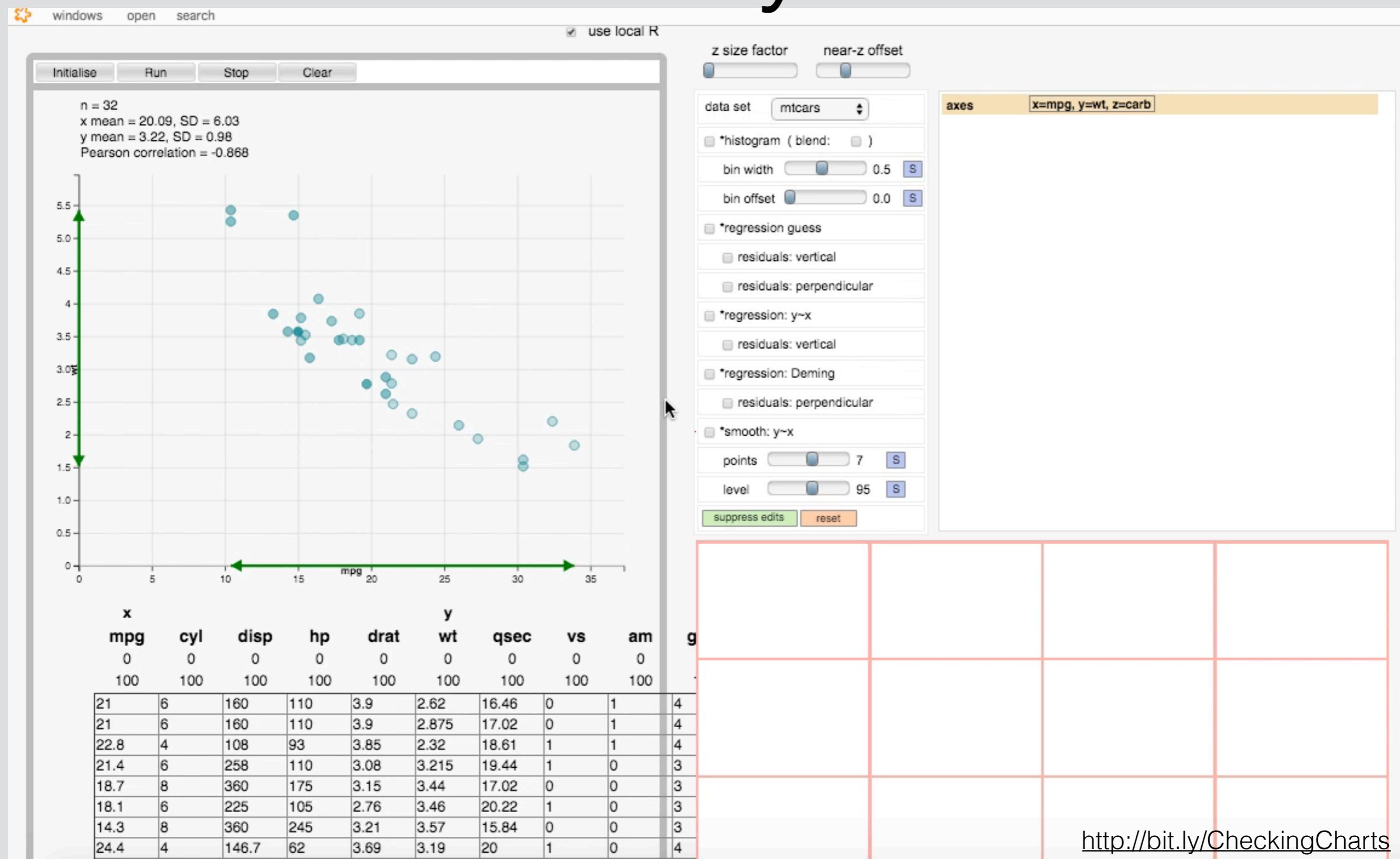
California education data

For more exposition, see <https://ameliamn.shinyapps.io/InteractionPlotExplanation/InteractionPlot.Rmd>

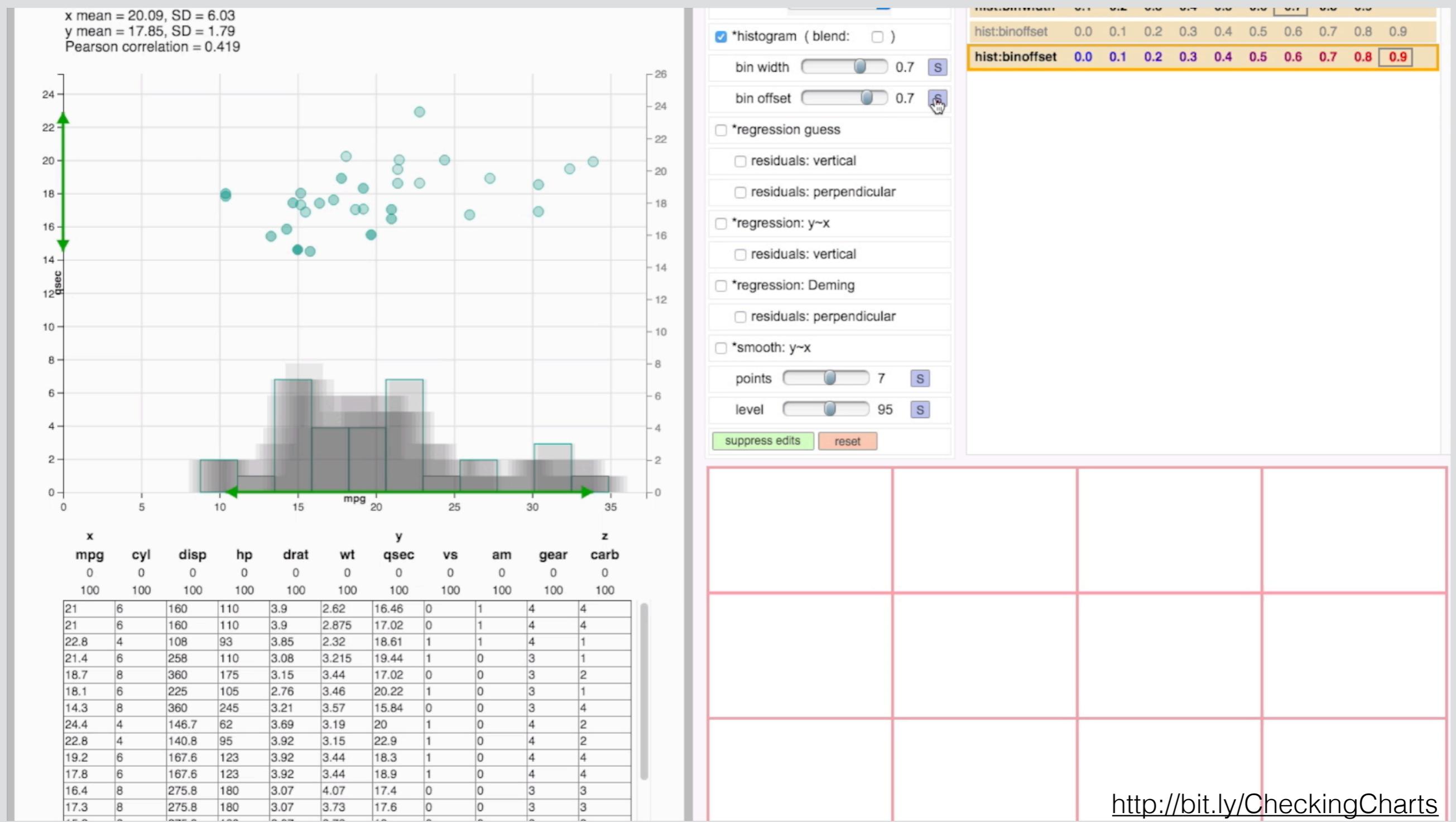


Parameter manipulation in experimental interfaces

LivelyR

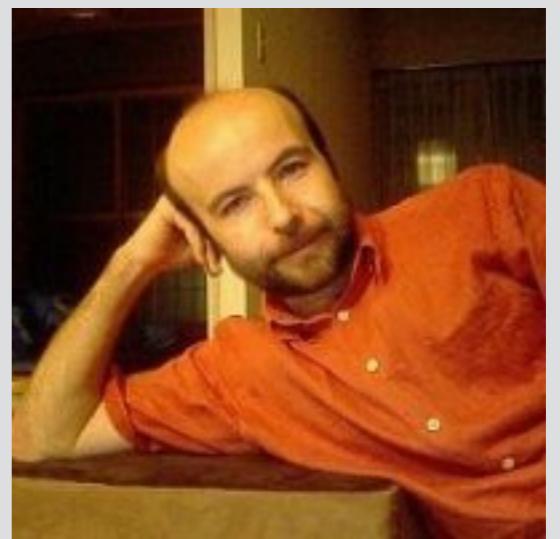


LivelyR

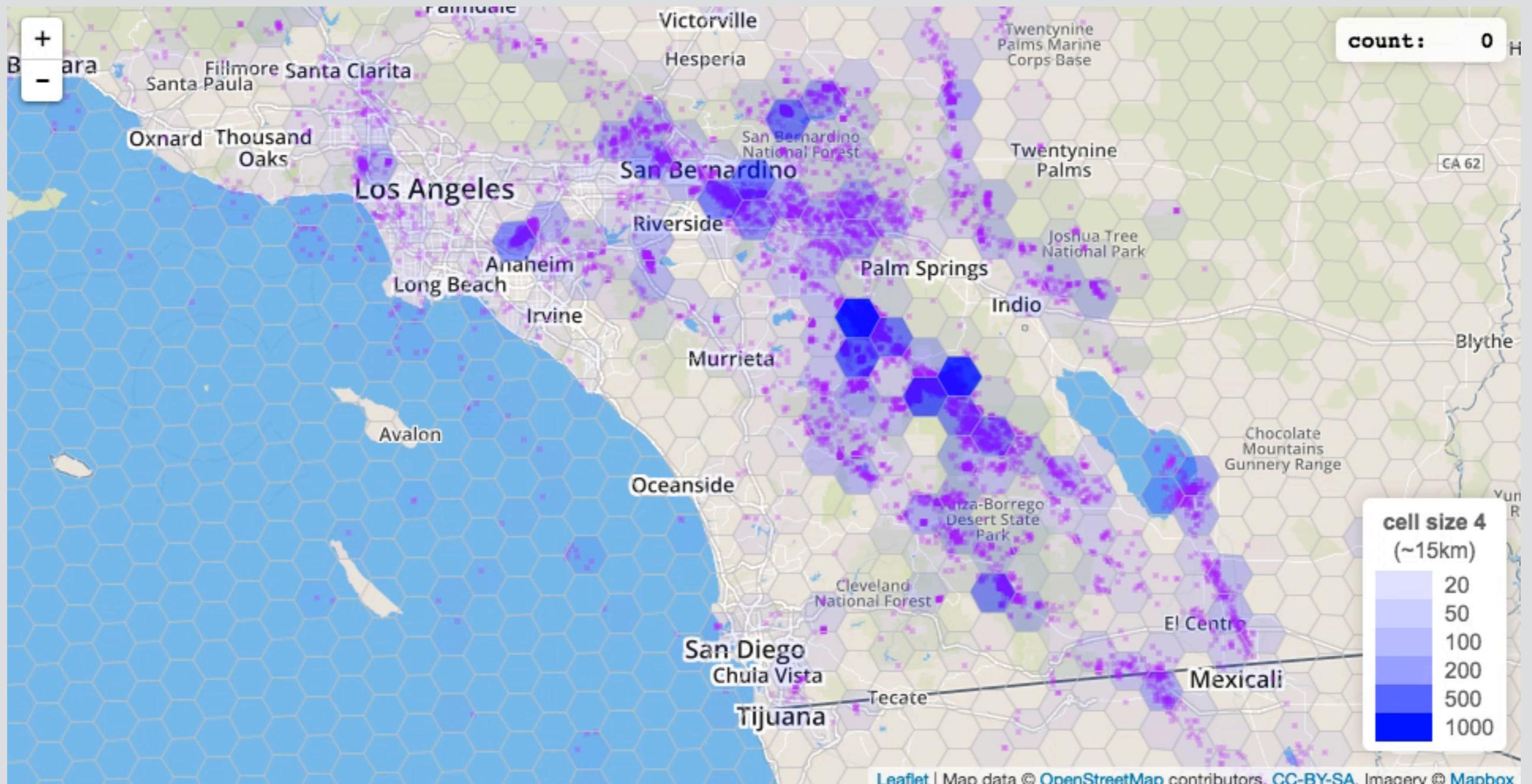


Aran Lunzer and Amelia McNamara. (2014). It ain't necessarily so: Checking charts for robustness. In *IEEE Vis 2014*.

bit.ly/CheckingCharts



Spatial aggregation toy



http://bit.ly/spatial_agg

What would be possible if
we made interactive
parameter manipulation
available everywhere?

Thank you