

# Teaching Data Science to Teenagers

Amelia McNamara

University of California, Los Angeles

[@AmeliaMN](#)

[bit.ly/AmeliaMN](http://bit.ly/AmeliaMN)

Mark Hansen

Columbia University

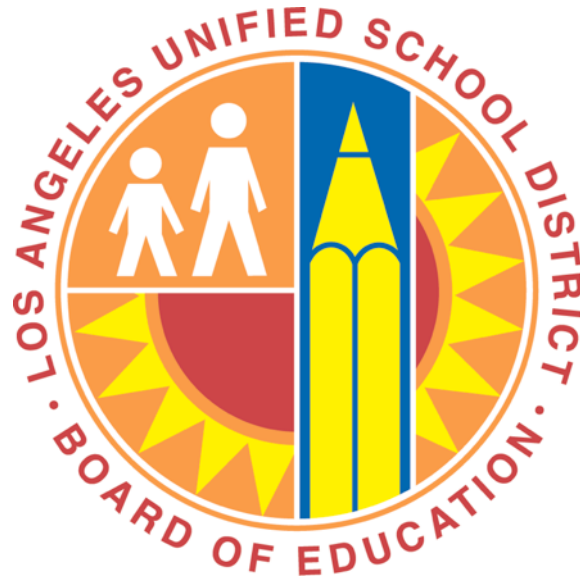
[@cocteau](#)

[www.stat.ucla.edu/~cocteau/](http://www.stat.ucla.edu/~cocteau/)

**m**  **obilize**

<http://www.mobilizingcs.org/>

# mobilizePartners



# mobilize Principal Investigators

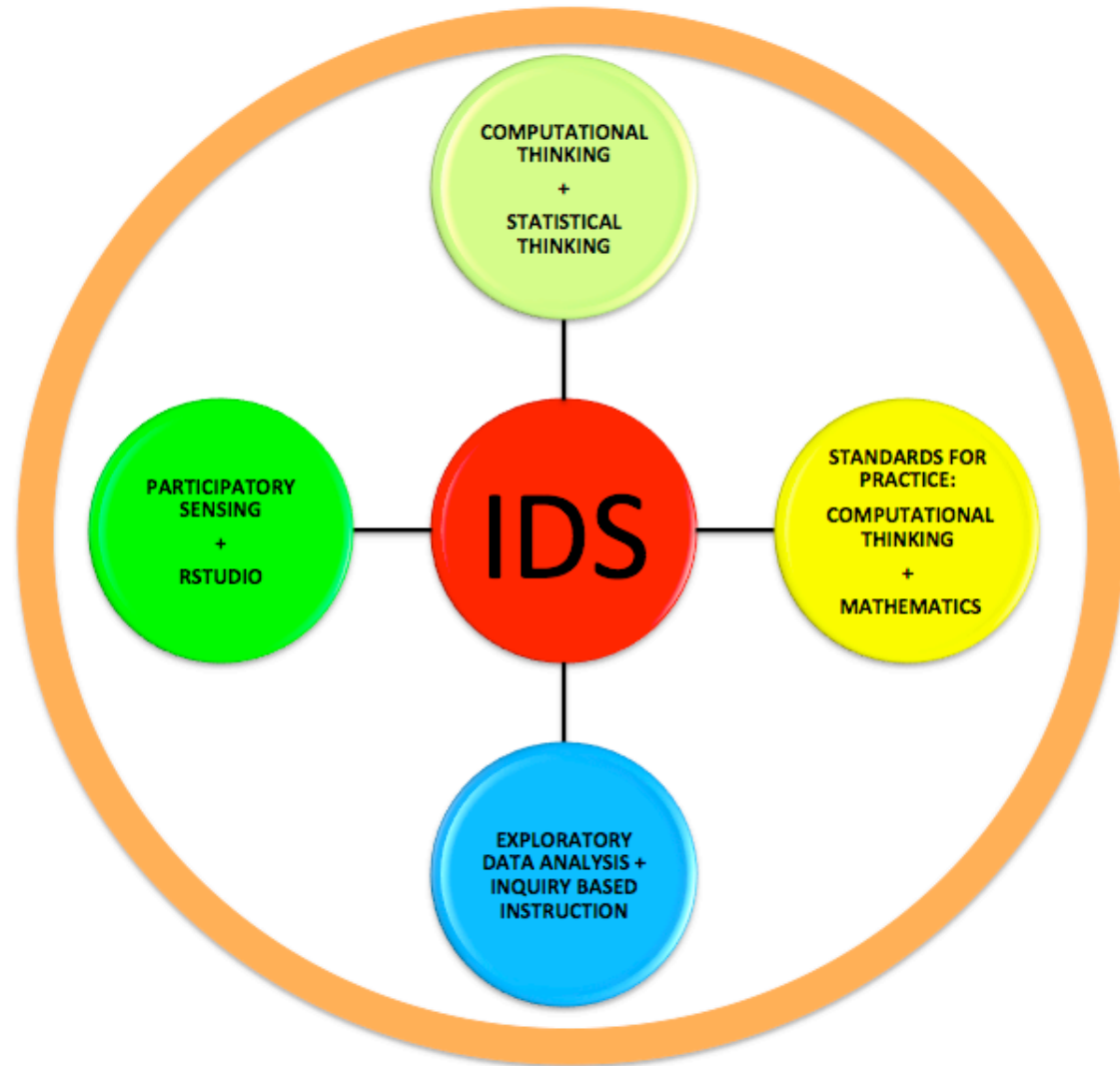
- Rob Gould (UCLA-- Statistics)
- Deborah Estrin (Cornell Tech-- Computer Science)
- Joanna Goode (University of Oregon-- Education)
- Mark Hansen (Columbia University-- Journalism)
- Jane Margolis (UCLA-- Education)
- Thomas Philip (UCLA-- Education)
- Jody Priselac (UCLA-- Education)
- Todd Ullah (LAUSD)

**Mobilize has many facets**



# Introduction to Data Science

- High school course
- Year-long
- Piloting in Fall 2014
- Computation-based
- “C” designation, from A-G requirements
- Using R
- Participatory sensing





Campaigns



Surveys



Response  
History



Upload  
Queue



Profile



Help

What is “Participatory sensing”?



You are logged in as mobipd.amelia

[Dashboard](#)[Campaigns](#)[Responses](#)[Explore Data](#)[Documents](#)[Classes](#)[Browse Responses](#) | [Edit Responses](#)**Campaign:**

PD\_2011\_Advertiseme

**Survey:**

All

**Participant:**

All

**Privacy:**

All

**From date:****To date:** Only photo responses**Showing 187 responses from all users**

Campaign participants may share or delete their responses while the campaign is still running. Once a campaign has been stopped, only supervisors may change responses.

[select all](#) | [none](#)    [expand](#) | [collapse](#)

<input type="checkbox"/>	Tue Jul 12 09:41:45 GMT-700 2011	mobipd.baf	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Tue Jul 12 09:37:44 GMT-700 2011	mobipd.baf	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Tue Jul 12 09:35:17 GMT-700 2011	mobipd.baf	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Tue Jul 12 09:32:51 GMT-700 2011	mobipd.baf	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Tue Jul 12 09:31:04 GMT-700 2011	mobipd.baf	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Tue Jul 12 09:29:00 GMT-700 2011	mobipd.baf	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Tue Jul 12 09:26:44 GMT-700 2011	mobipd.baf	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Tue Jul 12 09:25:05 GMT-700 2011	mobipd.baf	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Tue Jul 12 09:22:33 GMT-700 2011	mobipd.baf	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Tue Jul 12 08:00:05 GMT-700 2011	mobipd.bba	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 09:37:21 GMT-700 2011	mobipd.bab	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 09:32:27 GMT-700 2011	mobipd.bab	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 09:28:27 GMT-700 2011	mobipd.bab	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 09:25:41 GMT-700 2011	mobipd.bab	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 08:43:43 GMT-700 2011	mobipd.baj	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 08:42:18 GMT-700 2011	mobipd.baj	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 08:41:12 GMT-700 2011	mobipd.baj	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 08:39:46 GMT-700 2011	mobipd.baj	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 08:26:05 GMT-700 2011	mobipd.baj	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 08:09:17 GMT-700 2011	mobipd.baj	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 00:01:36 GMT-700 2011	mobipd.baa	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>
<input type="checkbox"/>	Mon Jul 11 00:00:04 GMT-700 2011	mobipd.baa	PD_2011_Advertisement	Advertising Media	<input type="button" value="Shared"/>

[select all](#) | [none](#)    [expand](#) | [collapse](#)

Teaching Data Science  
to Teenagers  
...and their teachers

# Tools

What is a “tool”?

What qualities should a tool have?

Authentic

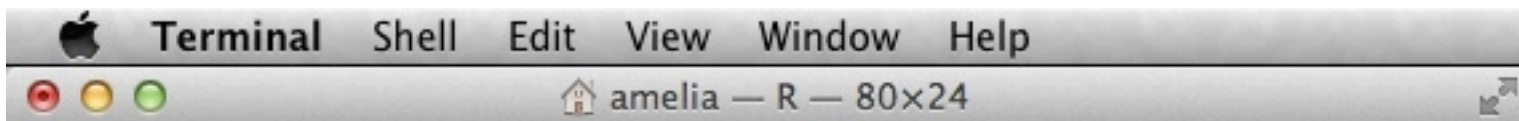
Low threshold

High Ceiling

Free/open source



Andi Ioannidou, et al. Computational Thinking Patterns.  
In *American Educational Research Association*, April 2011.



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.

> █

Apple R File Edit Format Workspace Packages & Data Misc Window Help

R Console



Help Search

```
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.
```

```
[R.app GUI 1.62 (6558) x86_64-apple-darwin10.8.0]
```

```
[History restored from /Users/amelia/.Rapp.history]
```

```
> |
```



```
Loading required package: foreign

Attaching package: 'Deducer'

The following object(s) are masked from
'package:stats':

  summary.lm

Loading required package: tm
Loading required package: wordcloud
Loading required package: Rcpp
Loading required package: Snowball
Loading required package: RColorBrewer
Loading required package: DeducerSpatial
Loading required package: sp
Loading required package: maptools
Checking rgeos availability: FALSE
  Note: when rgeos is not available, polygon
  geometry computations in maptools depend on gpclib,
  which has a restricted licence. It is disabled by
  default;
  to enable gpclib, type gpclibPermit()

>
```



cdc x 14,041 observations of 26 variables

	age	grade	gender	hisp_latino	race	height	w
1	16 years old	10th grade	Male	No	E	1.70	84
2	17 years old	10th grade	Male	No	A	1.75	95
3	16 years old	10th grade	Male	No	E	1.80	66
4	17 years old	10th grade	Female	No	E	1.47	53
5	15 years old	10th grade	Female	No	E	1.83	16
6	16 years old	10th grade	Female	No	E	1.68	56
7	16 years old	10th grade	Female	No	E	1.63	75
8	15 years old	10th grade	Female	No	A	1.65	73
9	15 years old	10th grade	Female	No	E	1.55	68
10	16 years old	10th grade	Female	No	E	1.57	45
11	16 years old	10th grade	Female	No	E	1.57	76
12	16 years old	10th grade	Female	No	E	1.68	73
13	16 years old	10th grade	Female	Yes	E	1.65	63

Displayed 1000 rows of 14,041 (13,041 omitted)

Environment History

Import Dataset Clear

Global Environment

Data

- cdc 14041 obs. of 26 variables

Values

- lab\_titles chr [1:11] "Unit 1 - Lab 1: Data, Code & Rstudio" "Unit 1 ..."

Functions

- load\_lab function ()

```

Console -/
> load_labs()
Error: could not find function "load_labs"
> load_lab()

1: Unit 1 - Lab 1: Data, Code & Rstudio 2: Unit 1 - Lab 2: Lab 2
3: Unit 1 - Lab 3: Lab 3                4: Unit 1 - Lab 4: Lab 4
5: Unit 1 - Lab 5: Lab 5                6: Unit 1 - Lab 6: Lab 6
7: Unit 1 - Lab 7a: Lab 7a             8: Unit 1 - Lab 7b: Lab 7b
9: Unit 1 - Lab 8: Lab 8                10: Unit 1 - Lab 9: Lab 9
11: Unit 1 - Lab 10: Lab 10

Selection: 1
> data(cdc)
> View(cdc)
> |
  
```

Files Plots Packages Help Viewer

# Data, Code & RStudio

## Unit 1 - Lab 1

Directions: Follow along with the slides and answer the questions in **red** font in your journal.



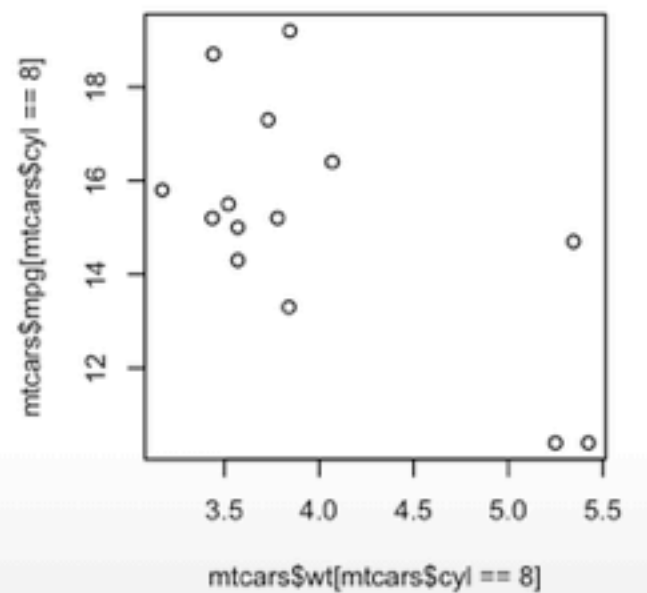
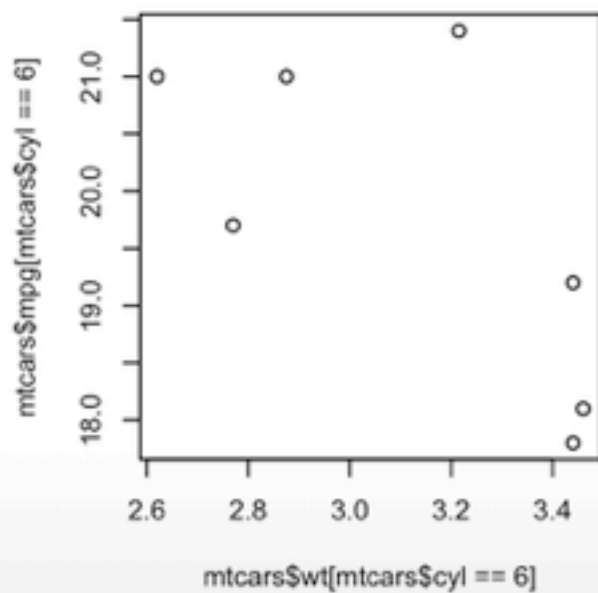
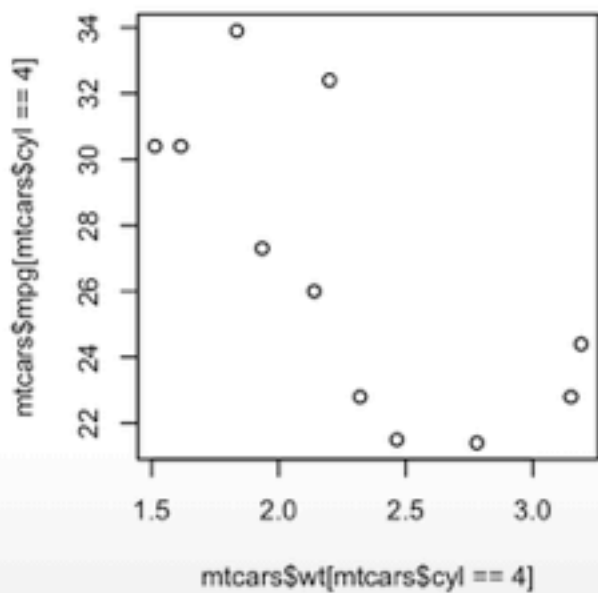
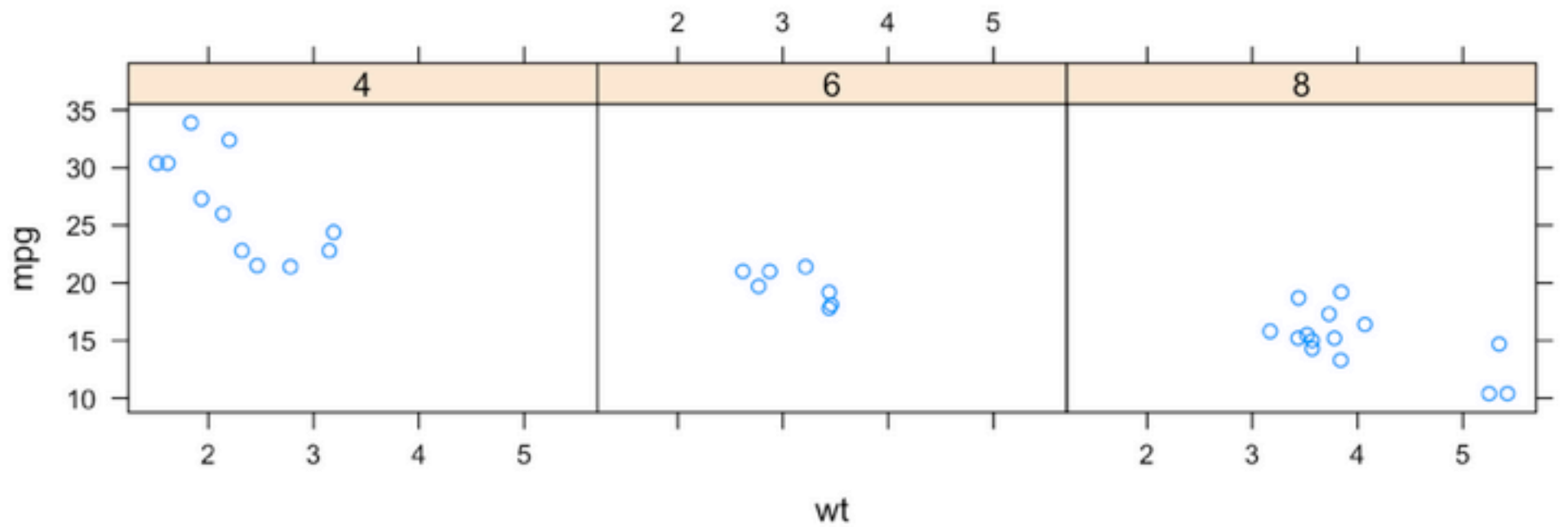
# Syntax

```
xyplot(mpg ~ wt | as.factor(cyl), data = mtcars)
```

VS.

```
par(mfrow = c(1,3))  
plot(mtcars$wt[mtcars$cyl == 4], mtcars$mpg[mtcars$cyl == 4])  
plot(mtcars$wt[mtcars$cyl == 6], mtcars$mpg[mtcars$cyl == 6])  
plot(mtcars$wt[mtcars$cyl == 8], mtcars$mpg[mtcars$cyl == 8])
```







This repository ▾

Search or type a command



Explore Gist Blog Help



AmeliaMN + - ✕



PUBLIC



**mobilizingcs / MobilizeSimple**

Unwatch ▾ 6

★ Star 3

Fork 3

branch: master ▾

**MobilizeSimple / R / +**



Fix typo



**Jimmylovestea** authored 25 days ago

latest commit [c40dee9332](#)

..

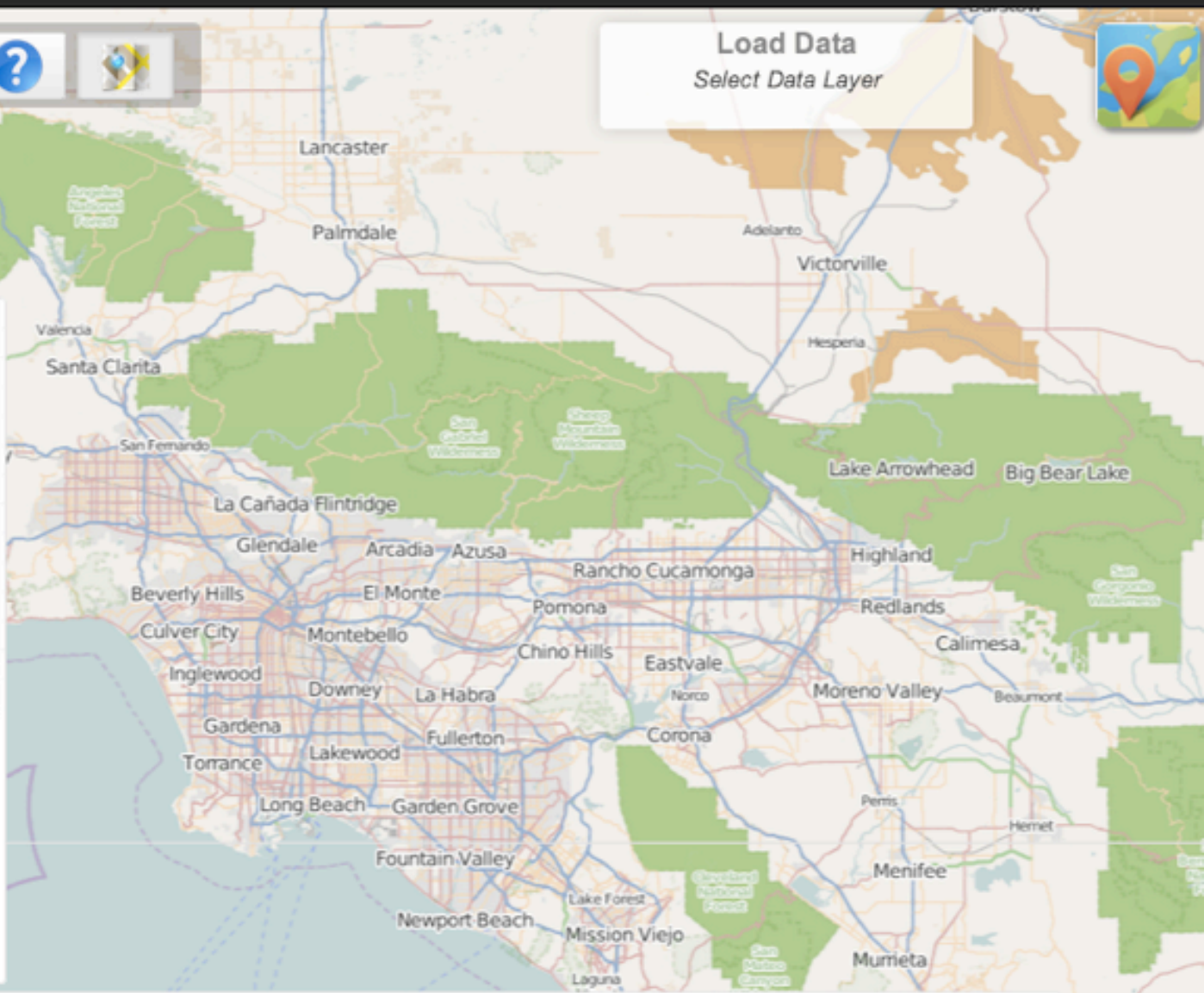
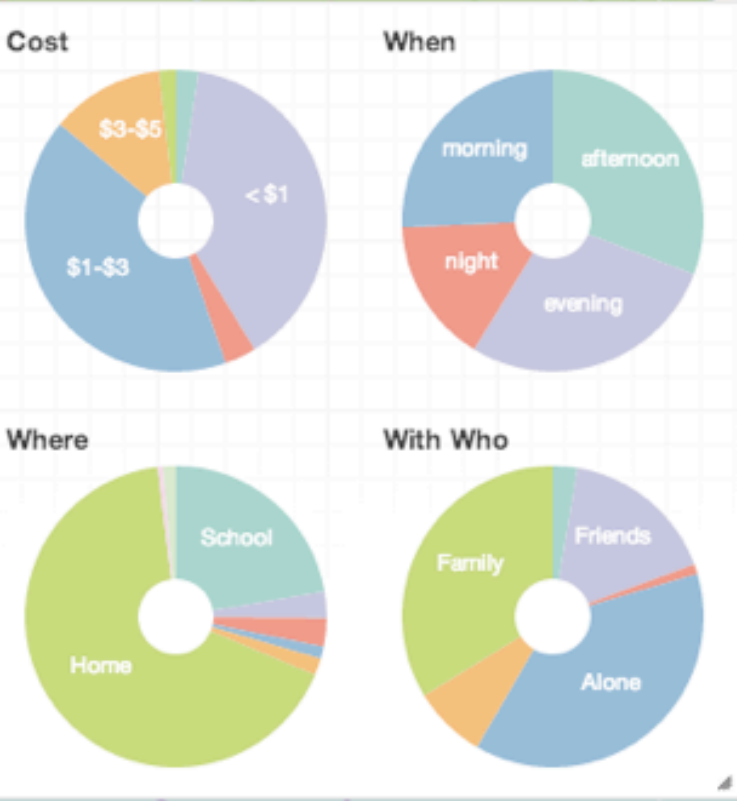
<a href="#">InitializeText.R</a>	Documentation changes, splitting out of spatial subsetting	a year ago
<a href="#">MakeMap.R</a>	Fixed some syntax from the merge	5 months ago
<a href="#">MakeWordBar.R</a>	Fixed merge conflicts in MakeMap	5 months ago
<a href="#">MakeWordCloud.R</a>	Fixed merge conflicts in MakeMap	5 months ago
<a href="#">MobilizeSimple-internal.R</a>	second commit	a year ago
<a href="#">ProcessText.R</a>	Fixed merge conflicts in MakeMap	5 months ago
<a href="#">PullNumbers.R</a>	Some new functions	a year ago
<a href="#">SpatialSubset.R</a>	Respecified the function to work specifically with dynamically-genera...	a year ago
<a href="#">UpdateStopwords.R</a>	Fixed merge conflicts in MakeMap	5 months ago
<a href="#">WordFreqTable.R</a>	Fixed merge conflicts in MakeMap	5 months ago
<a href="#">atus_format.R</a>	Fix typo	25 days ago
<a href="#">bargraph.R</a>	Added IDS functions and updated dependencies	a month ago
<a href="#">get_IDS_labs.R</a>	Added get_IDS_labs function	28 days ago
<a href="#">histogram.R</a>	Added IDS functions and updated dependencies	a month ago
<a href="#">mad.R</a>	Added IDS functions and updated dependencies	a month ago
<a href="#">xyplot.R</a>	Added IDS functions and updated dependencies	a month ago

<http://github.com/mobilizingcs/MobilizeSimple>

**Load Data**  
Select Data Layer



1,709 selected out of 1,709 records | [\(reset all\)](#)





# **mobilize** Web Tools select a tool to get started

## Web Frontend



Data management and viz

## Survey Taking



Browser-based data collection tool

## Class Setup



LAUSD Teacher class creation tool

## Authoring Tool



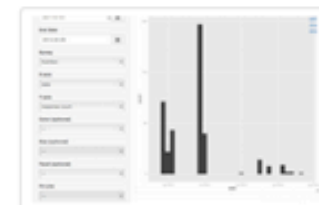
Interactive campaign creation

## Dashboard



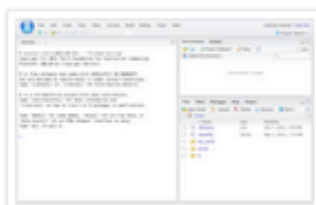
Interactive data exploration.  
Also available: [Public board](#)

## Plot App



R-based data exploration

## RStudio



User-friendly interface for R

## Campaign Monitor



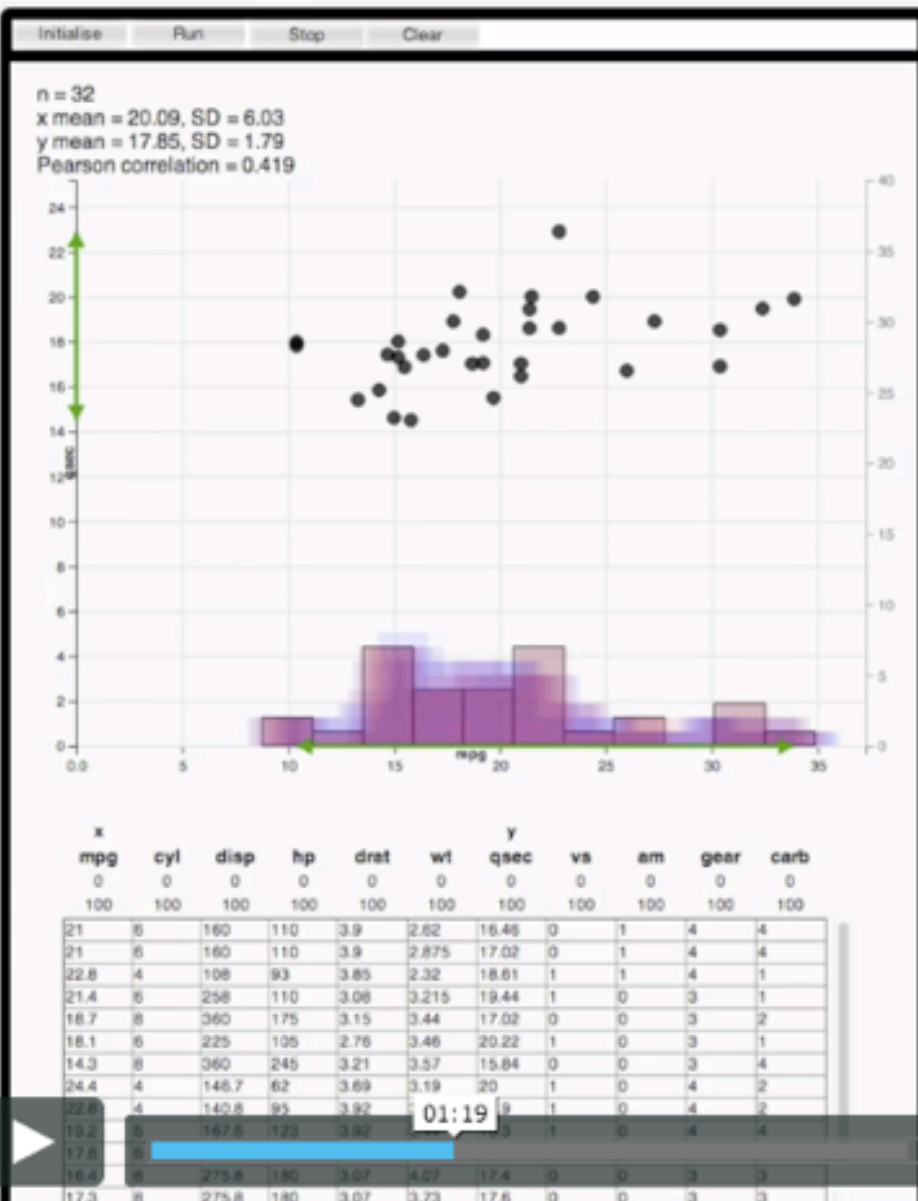
Campaign progress tracking

## Wiki



All materials: curricula, how-to guides, video tutorials, etc.

So, the change of software has helped.  
What is still making this hard?



data set: mtcars

\*histogram (blend: )

bin width: 0.7

bin offset: 0.7

\*regression guess

residuals: vertical

residuals: perpendicular

\*regression: y~x

residuals: vertical

\*regression: Deming

residuals: perpendicular

\*smooth: y~x

points: 7

x SD lines

suppress edits reset

axes: x=drat, y=mpg

axes: x=hp, y=mpg

axes: x=disp, y=mpg

axes: x=cyl, y=mpg

axes: x=mpg, y=mpg

axes: x=mpg, y=qsec

hist binwidth: 0.5 0.6 0.7

hist binoffset: 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7

hist binoffset: 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

```
Workspace
LivelyR is...
...a mashup of
• R, with packages from RStudio:
  ggvis, shiny, evaluate, httpuv
• Javascript, with libraries:
  Vega, d3, Lively Web
...a collaborative work-in-progress
• direction: Amelia McNamara
• code: Aran Lunzer, Robert Krohn
• support: Bret Victor, Dan Ingalls
```

# LivelyR introductory demo

[bit.ly/LivelyR\\_demo1](http://bit.ly/LivelyR_demo1)

**Thank you**