

Integrated R labs for high school students

Amelia McNamara, James Molyneux, and Terri Johnson

@AmeliaMN, @jimmylovestea, @terriannabana



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



flickr: leslee

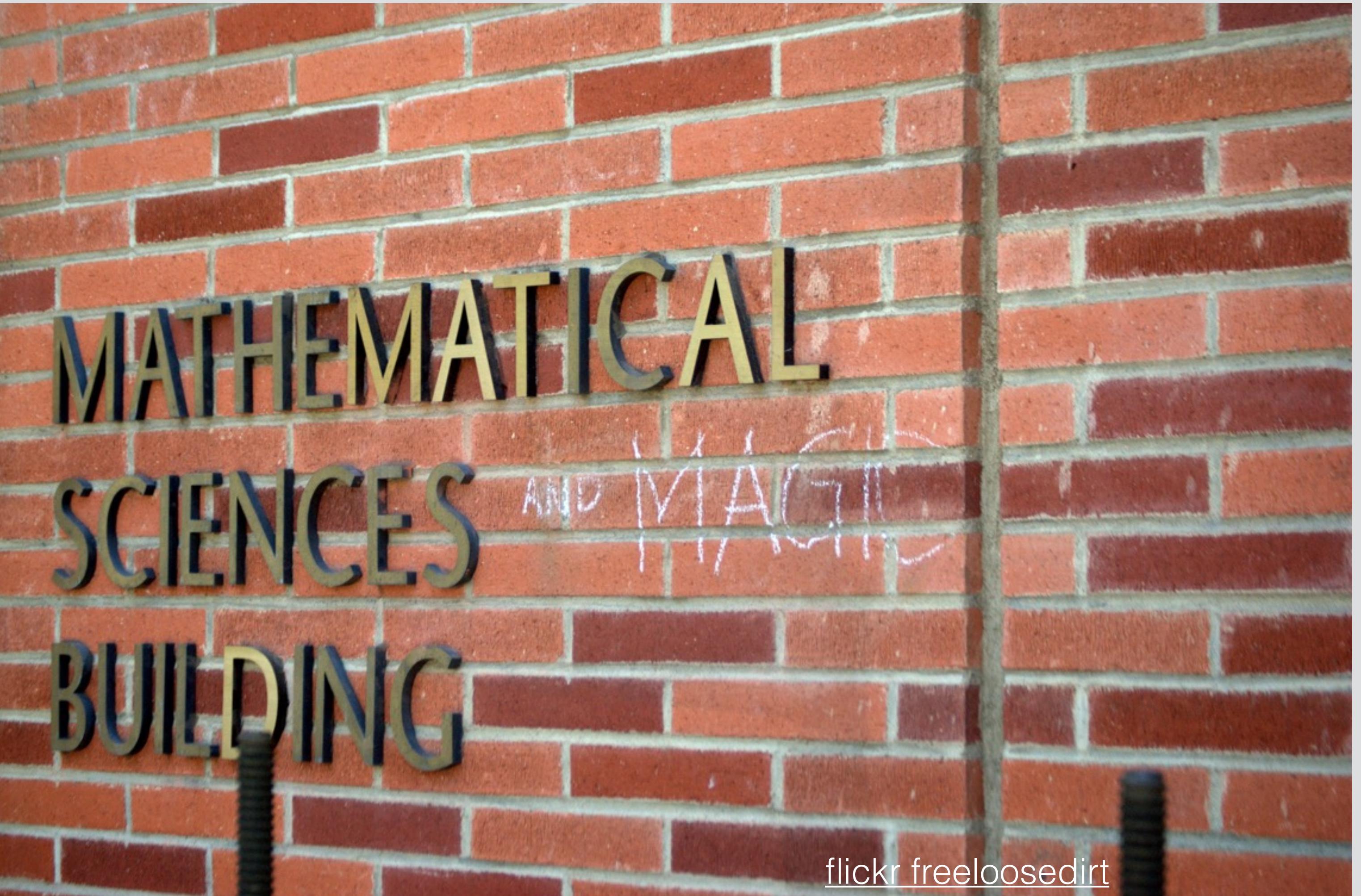


James Molyneux



Terri Johnson

My coauthors





Our data. Our lives.



Curriculum



Technology



Professional
Development



Publications &
Presentations



Mobilize News

Participatory sensing

Briefly: using devices with sensors (smartphones) to collect data from a group of people to form a body of knowledge

MANY moving parts (pt 1)

MANY moving parts (pt 1)

- Lead PI:
 - Rob Gould, Director of the Center for Teaching Statistics/Vice-Chair of Undergraduate Studies, UCLA Statistics



MANY moving parts (pt 1)

- Lead PI:
 - Rob Gould, Director of the Center for Teaching Statistics/Vice-Chair of Undergraduate Studies, UCLA Statistics
- Co-PIs:
 - Derek Chau, Director of Secondary Instruction, LAUSD Office of Curriculum and Instruction
 - Joanna Goode, Associate Professor, University of Oregon College of Education
 - Jane Margolis, Senior Researcher, UCLA Center X
 - Thomas Philip, Associate Professor, UCLA Graduate School of Education & Information Studies
 - Jody Priselac, Associate Dean for Community Programs, UCLA Graduate School of Education & Information Studies



MANY moving parts (pt 1)

- Lead PI:
 - Rob Gould, Director of the Center for Teaching Statistics/Vice-Chair of Undergraduate Studies, UCLA Statistics
- Co-PIs:
 - Derek Chau, Director of Secondary Instruction, LAUSD Office of Curriculum and Instruction
 - Joanna Goode, Associate Professor, University of Oregon College of Education
 - Jane Margolis, Senior Researcher, UCLA Center X
 - Thomas Philip, Associate Professor, UCLA Graduate School of Education & Information Studies
 - Jody Priselac, Associate Dean for Community Programs, UCLA Graduate School of Education & Information Studies
- Faculty advisors:
 - Anna Bargagliotti, Associate Professor, Mathematics, College of Science & Engineering, Loyola Marymount University
 - Deborah Estrin, Professor of Computer Science, Cornell Tech
 - Mark Hansen, Director, Brown Institute for Media Innovation & Professor of Journalism, Columbia University
 - Tim Jacobbe, Associate Professor, College of Education, University of Florida



MANY moving parts (pt 2)

- Goals:
 - Design a High School Curriculum
 - Develop Participatory Sensing Software
 - Build Multidisciplinary STEM Teacher Teams and Professional Development
 - Enhance Student Learning and Engagement
 - Support LAUSD Initiatives
 - Campaign for New Educational Policy
 - Build National Dissemination and Impact
 - Conduct Research, Evaluation, and Assessment of Student Learning



MANY moving parts (pt 3)

- Curriculum:
 - add-on units:
 - Exploring Computer Science (ECS)
 - Algebra I
 - Biology
 - Year-long course:
 - Introduction to Data Science (IDS)



- Year-long course
- Validates Algebra II requirement
- “Data science”
- Taught in R within RStudio server
- Participatory sensing
- Content includes:
 - Exploratory data analysis
 - Randomization, simulation, bootstrapping
 - Simple linear regression, multiple regression
 - Decision trees, clustering, k-means



Introduction to Data Science

Introduction to Data Science



Robert Gould

Suyen Moncada-Machado

Terri Anna Johnson

James Molyneux

RStudio Lab Codes and Functions

Contents

Loading, saving and viewing data	1
Scraping web data	3
Data manipulation functions	3
Numerical summaries and frequency tables	5
Plotting functions	7
Maps	11
Statistical modeling	12
Sampling and permutation functions	13
Probability functions	14
Symbols and operators	15
Creating functions	15

Loading, saving and viewing data

`data()`: Loads and displays a pre-loaded data file from RStudio.

Example:

```
data(cdc)
```

`read.csv()`: Imports data from a `.csv` formatted file into R.

Example:

```
read.csv("Time Use.csv")
```

`View()`: Displays the data as a spreadsheet in a new tab.

Example:

```
View(cdc)
```

`head()`: Prints the first 6 values or rows of data in the console.

Examples:

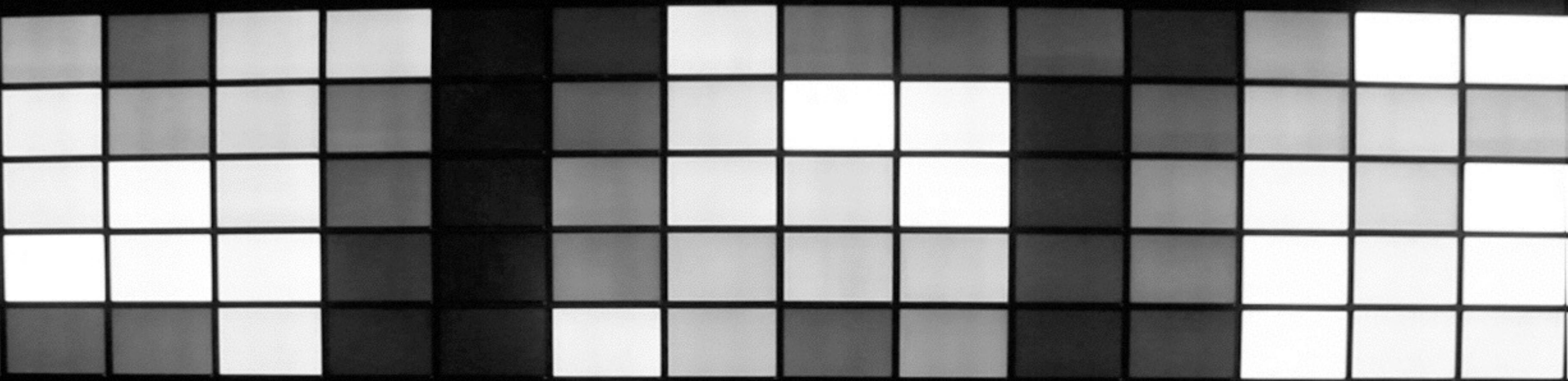
```
# Observations of a dataset  
head(cdc)
```

```
# Observations of a variable  
head(~gender, data = cdc)
```

`tail()`: Prints the last 6 values or rows of data in the console.

Examples:

```
# Observations of a dataset  
tail(cdc)
```



Labs



DataCamp

OpenIntro[®]



{swirl}

Learn R, in R.



Introduction to Data Science

Lab1.Rmd x cdc x

Filter

	age	grade	gender	hisp_latino	race	height	weight	general_health	hours_sleep
1	16 years old	10th grade	Male	No	E	1.70	84.37	Good	8 hours
2	17 years old	10th grade	Male	No	A	1.75	99.79	Good	7 hours
3	16 years old	10th grade	Male	No	E	1.80	68.04	Very good	7 hours
4	17 years old	10th grade	Female	No	E	1.47	51.71	Fair	5 hours
5	15 years old	10th grade	Female	No	E	1.83	102.06	Very good	4 or less hours
6	16 years old	10th grade	Female	No	E	1.68	56.70	Very good	7 hours
7	16 years old	10th grade	Female	No	E	1.63	75.75	Good	7 hours
8	15 years old	10th grade	Female	No	A	1.65	73.48	Good	6 hours
9	15 years old	10th grade	Female	No	E	1.55	68.04	Fair	8 hours
10	16 years old	10th grade	Female	No	E	1.57	49.44	Very good	8 hours
11	16 years old	10th grade	Female	No	E	1.57	76.66	Good	8 hours
12	16 years old	10th grade	Female	No	E	1.68	71.67	Fair	5 hours
13	16 years old	10th grade	Female	Yes	E	1.65	61.69	Fair	4 or less hours
14	15 years old	10th grade	Male	No	E	1.80	73.94	Excellent	6 hours
15	15 years old	10th grade	Male	No	E	1.80	72.58	Very good	6 hours
16	16 years old	10th grade	Male	No	C	1.80	72.58	Fair	5 hours
17	16 years old	10th grade	Female	No	E	1.69	57.16	Excellent	8 hours

Showing 1 to 17 of 14,041 entries

Console ~/Dropbox/user2016/ ↵

```
> library(mobilizr)
> load_labs()
```

Enter the number next to the lab you would like to load:

1: Lab 1A	2: Lab 1B	3: Lab 1C	4: Lab 1D	5: Lab 1E
6: Lab 1F	7: Lab 1G	8: Lab 1H	9: Lab 2A	10: Lab 2B
11: Lab 2C	12: Lab 2D	13: Lab 2E	14: Lab 2F	15: Lab 2G
16: Lab 2H	17: Lab 2I	18: Lab 3A	19: Lab 3B	20: Lab 3C
21: Lab 3D	22: Lab 3E	23: Lab 3F	24: Lab 4A	25: Lab 4B
26: Lab 4C	27: Lab 4D	28: Lab 4E	29: Lab 4F	30: Lab 4G
31: Lab 4H	32: PD Lab 1D	33: PD Lab 1E		

Selection: 1

```
> data(cdc)
> View(cdc)
>
```

Environment History

Import Dataset

Global Environment

Data

cdc 14041 obs. of 26 variables

Files Plots Packages Help Viewer

So let's get started!

- The data for our first few labs comes from the Centers for Disease Control (CDC)
 - The CDC is a federal institution that studies public health.
- Type these two commands into your console:
 - `data(cdc)`
 - `View(cdc)`

- Describe the data that appeared after running `View(cdc)`:
 - Who is the information about?
 - What sorts of information about them was collected?



This repository

Search

Pull requests Issues Gist



mobilizingcs / ids_labs

Unwatch 5

Star 0

Fork 0

Code

Issues 2

Pull requests 0

Wiki

Pulse

Graphs

Branch: gh-pages

ids_labs / unit_1 / lab1a / lab1a.Rpres

Find file Copy path

 jimmylovestea Updating the newest version of the labs since it's been so long since... 02e907d 14 days ago

1 contributor

154 lines (116 sloc) | 4.71 KB

Raw

Blame

History

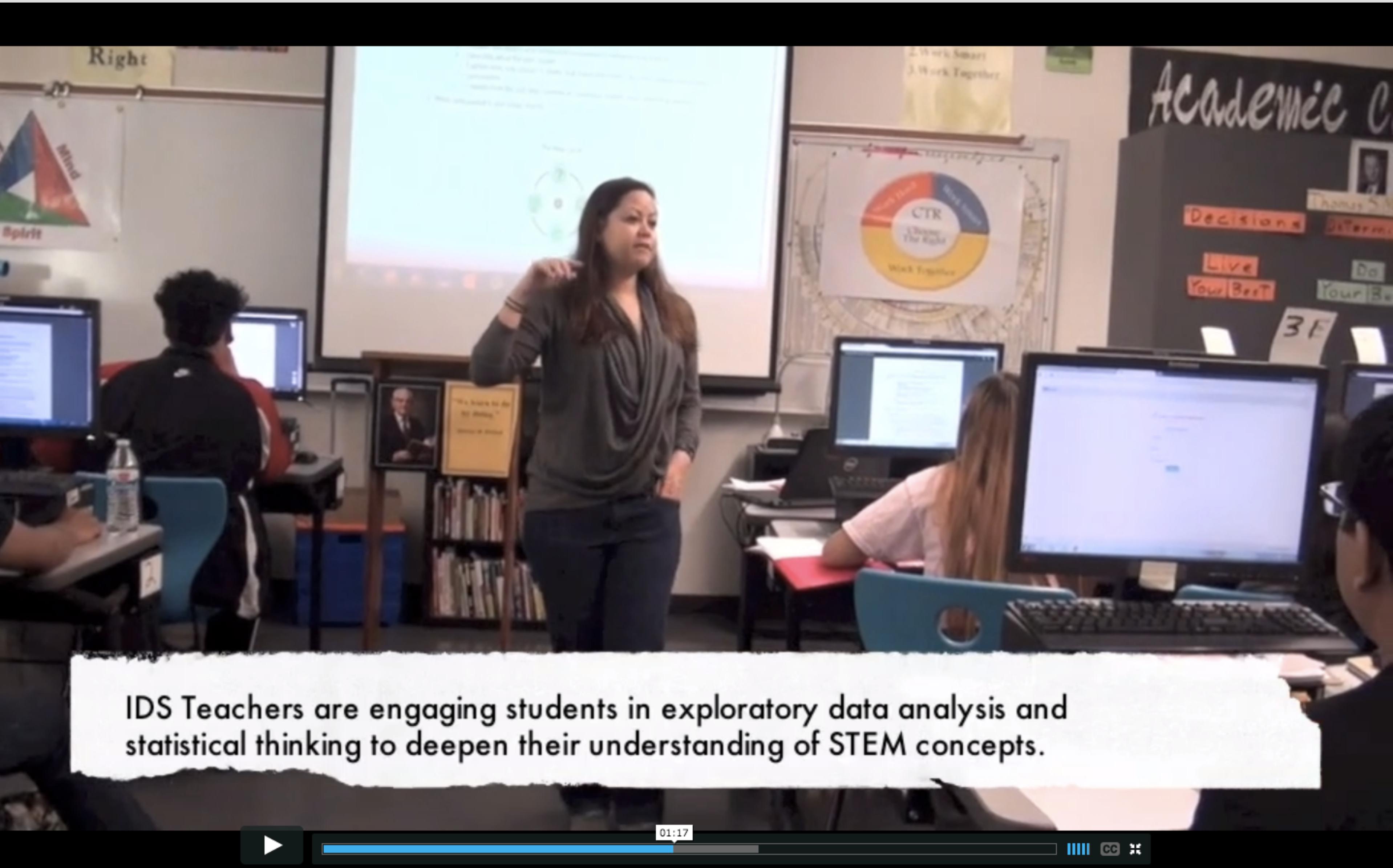


```
1 Data, Code & RStudio
2 =====
3 transition: none
4 css: ../../IDSLabCSS.css
5
6 Lab 1A
7
8 Directions: Follow along with the slides and answer the questions in **red** font in your journal.
9
10 ```{r, eval=TRUE, echo=FALSE, results='hide', message=FALSE, warning=FALSE}
11 library(mobilizr)
12 data(cdc)
13 ```
14
15 Welcome to the labs!
16 =====
17
18 Throughout the year, you'll be putting your data science skills to work by completing the labs.
```

Package: mobilizr
Type: Package
Title: Functions for the Mobilize Introduction to Data Science course.
Description: This package is intended to simplify the code writing process for Introduction to Data Science, a Mobilize grant high school course.
Authors@R: as.person(c(
 "James Molyneux <jmolyneux@ucla.edu> [aut, cre]",
 "Amelia McNamara <amcnamara@smith.edu> [aut]",
 "Terri Johnson <terri.johnson@stat.ucla.edu> [aut]",
 "Steve Nolen <technolengy@gmail.com> [aut]"
))
Depends: R (>= 3.0.2), grid, stats, MASS, plyr, dplyr, mosaic,
ggplot2, XML, leaflet, RColorBrewer, network, stringr, rpart
Imports: curl, rstudioapi, log4r
Suggests: testthat

Want to check out the labs?

```
library(devtools)
install_github("mobilizingcs/mobilizr")
library(mobilizr)
load_labs()
```



Resources

- Project website: <http://www.mobilizingcs.org/>
- Curriculum page: <http://www.mobilizingcs.org/curriculum>
- GitHub organization: <https://github.com/mobilizingcs>
- mobilizr package: <https://github.com/mobilizingcs/mobilizr>
- Lab code: https://github.com/mobilizingcs/ids_labs
- Wiki: <https://wiki.mobilizingcs.org/>

A large, bright classroom with rows of desks and chairs, looking out onto a snowy landscape through large windows.

Thank you