IMMERSE Pre-Training Day 5

June 2, 2023

UC SANTA BARBARA

Overview

Review of Days 3 and 4

Break

Review of Day 2

Review of Day 1

Questions about in-person training?

Review of Days 3 and 4

Day 3: Data wrangling, exploration, and visualization

Day 4: Collaborative, reproducible workflows with R, R Studio, git, and Github

What is your scariest/most embarrassing data moment?

Take a couple minutes to think about it, write some notes here. Next, you'll have a opportunity to share and hear from others in small groups

What is your scariest/most embarrassing data moment?

- Example: I have to triple check that I've transferred the results from my output to my table in Word to make sure I didn't make any mistakes (which I have done before!)

Individual: Write your thoughts here (2 minutes)

Small groups (10 minutes)

Nominations for the best (worst?) data story?

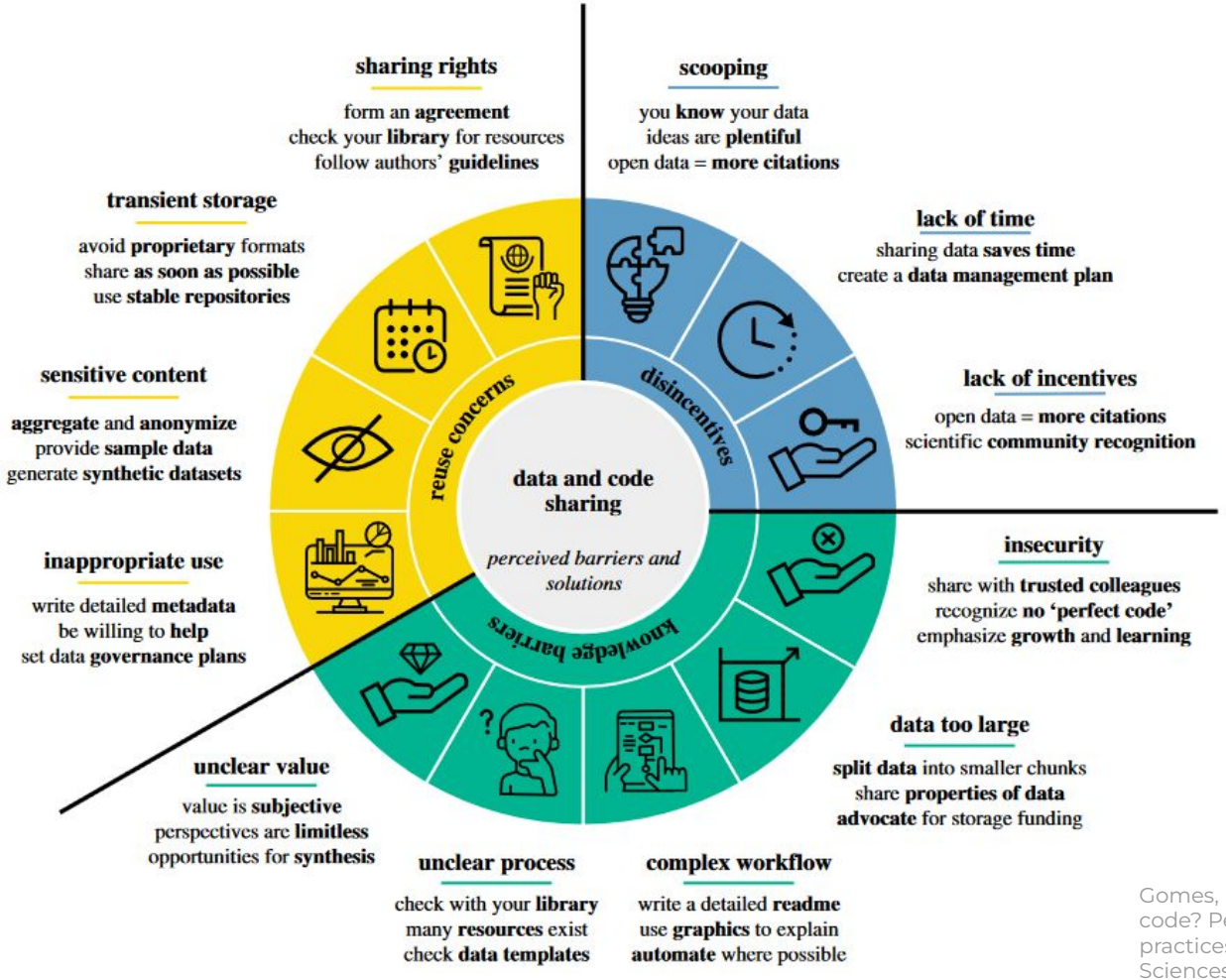


Figure 1. Perceived barriers and solutions to sharing data and code.

We highlight 12 distinct barriers (see icons and corresponding underlined titles) to researchers publicly sharing data and code, which can be broken into three larger groups (knowledge barriers, reuse concerns and disincentives; innermost circle).

Underneath the section titles, we list a few suggestions for overcoming these barriers (see main text for more details).

Gomes, D. G. E., et al. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. Proceedings of the Royal Society B: Biological Sciences, 289(1987), 20221113

“Good enough” practices: data management

From *Good enough practices in scientific computing*:

- Save the raw data
- Ensure that raw data are backed up in more than one location
- Create the data you wish to see in the world
- Create analysis-friendly data
- Record all the steps used to process data

(partial list - see article for more)

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. PLOS Computational Biology, 13(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

“Good enough” practices: software/coding

From *Good enough practices in scientific computing*:

- Place a brief explanatory comment at the start of every program
- Decompose programs into functions
- Be ruthless about eliminating duplication
- Give functions and variables meaningful names
- Search for well-maintained software libraries that do what you need
- Test libraries before relying on them

(partial list - see article for more)

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. PLOS Computational Biology, 13(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

“Good enough” practices: collaboration/organization

From *Good enough practices in scientific computing*:

- Collaboration
  - Decide on communication strategies
- Project organization
  - Put each project in its own directory, which is named after the project
  - Put raw data and metadata in a `data` directory
  - Put files generated during cleanup and analysis in a `results` directory
  - Put project source code in the `src` directory
  - Name all files to reflect their content or function
- Keeping track of changes
  - Keep changes small; share changes frequently
  - Use a version control system

(partial list - see article for more)

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. PLOS Computational Biology, 13(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

More “good enough” practices: Casey’s pro tips

- Raw data stays raw
- The script is the key to everything
- Document for your audience, collaborators, and for future you
- Absolute vs relative file paths: `.Rproj` and `here::here()` is your friend
- Version control
- R Markdown as a reproducibility check!

Thinking about your moment or reflecting on someone's experience that you just heard about, how might data science practices or habits been helpful in that situation?

Take a couple minutes to think about it, then we will put you in breakout rooms to share with each other in pairs.

How could data science have helped you in that moment?

- Example: I have to triple check that I've transferred the results from my output to my table in Word to make sure I didn't make any mistakes (which I have done before!)
    - Solution: MplusAutomation can extract the values from your output into a nicely formatted table! No copy/paste needed.

Write your notes here (5 minutes)

Pairs (7 minutes)

BREAK (5 minutes)

Review of Day 2

Introduction to Mplus and MplusAutomation in R

WHAT is MplusAutomation & WHY should we use it?

WHAT?

- MplusAutomation is an R package
- It "wraps around" the Mplus program
- Requires both R & Mplus software
- Requires learning some basics of 2 programming languages
- Car metaphor: R/Rstudio is the steering wheel or dashboard & Mplus is the engine

WHY?

- MplusAutomation can provide clearly organized work procedures in which every research decision can be documented in a single place
- Increase reproducibility, organization, efficiency, and transparency

HOW?

- The interface for MplusAutomation is entirely within R-Studio. You do not need to open Mplus
- The code presented will be very repetitive by design

IMMERSE Project– IES funded Training Grant (R305B220021)

BREAKOUT ROOM

What are your lingering questions about MplusAutomation/R/Rprojects?

Take a couple minutes to think about it, then we will put you in breakout rooms to share with each other. You can write notes in your groups here

What are your lingering questions about Mplus/ MplusAutomation/R?

Example: In what cases would we use Mplus over MplusAutomation?

- Small groups: Write your notes here start on slide 16 (10 minutes)
- Individual (2 minutes): Take a moment to look at questions from other groups
- Whole group (10 minutes)

DISCUSSION

Take a look at what other groups wrote and add comments to the slides.

Review of Day 1

Training goals and how to apply mixture modeling to address those goals

Goals for Training

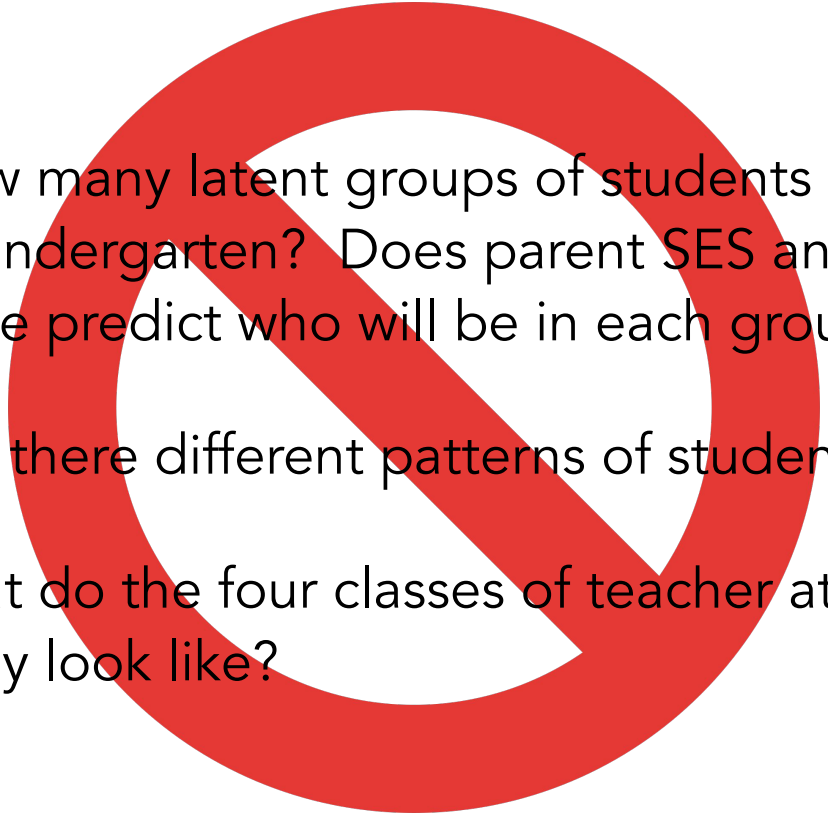
Our goal for this training program is support your learning and appropriate use of about mixture modeling approach so that you can conduct rigorous and relevant educational research

- Pretraining
- In-person training
- Ongoing support

Outcomes:

- formative and summative feedback from you about your learning experiences
- your application of mixture modeling to research objectives that you are interested in that by the end of this year that results in a conference presentation, peer reviewed publication, grant proposal, etc.

Typical Research Questions



RQ1: How many latent groups of students learners are there in Kindergarten? Does parent SES and preschool attendance predict who will be in each group?

RQ2: Are there different patterns of student motivation?

RQ3: What do the four classes of teacher attitudes towards technology look like?

Research Goals

Many applications of mixture modeling is exploratory

- Assume that there are subgroups
- No specific hypothesis about what emergent classes will be or how many there are

Building an argument for the presence of classes

- Previous work discuss differences among students?
- Literature review and find papers with mixtures/clusters?
- Theoretically ground your hypothesis that is it reasonable and meaningful to identify classes.

The argument of the existence of classes is theoretical, not statistical

Research Goals

Research goals/objectives:

- Explore what groups exist
  - Note: we don't ask "do classes exist"
- What are the characteristics of latent subgroups?
- What relations are there with the emergent groups
  - Covariates/distal outcomes

Example Research Goals

- The primary goal of this study was to characterize population heterogeneity in the person-specific stability and change of depression symptomatology spanning across portions of the first two decades of life— from ages 4 through 16.5 years (linked)

- The goals of this study were (1) to examine the diverse ways in which body image concerns and behaviors are manifested in males from middle to late adolescence and (2) to examine how sexual orientation modifies risk for problematic patterns of concerns and behaviors across adolescence (linked)
- Using data from a large nationally representative youth health survey, this paper examines indicators of socioeconomic deprivation and how these indicators vary by demographic characteristics of adolescents. We identify adolescents experiencing household poverty, using latent class analysis, and examine the relationship with a well established measure of neighbourhood deprivation in New Zealand (linked)

IMMERSE Project– IES funded Training Grant (R305B220021)

Let's continue to work on your research goals

We added comments to those who provided draft research goals after the day 1 pre-training. Check out comments. Please don't delete the comments. We want to keep those there for others to view.

- Helpful to allude to the analysis that you will run and type of data
- Mixtures on demographics that are not malleable are generally discouraged-- though certain circumstances, is done

Let's continue to work on your research goals

Here is a link to the day 5 pre-training workspace for you to continue to work on your research goals (10 minutes)

We are available for breakout rooms if you want to talk with us individually now. Just send us a message in the chat and we will open up a breakout room

We will provide written feedback on your slides before Monday.

All pre-training information is housed below.

https://immerse-ucsb.github.io/pre-training

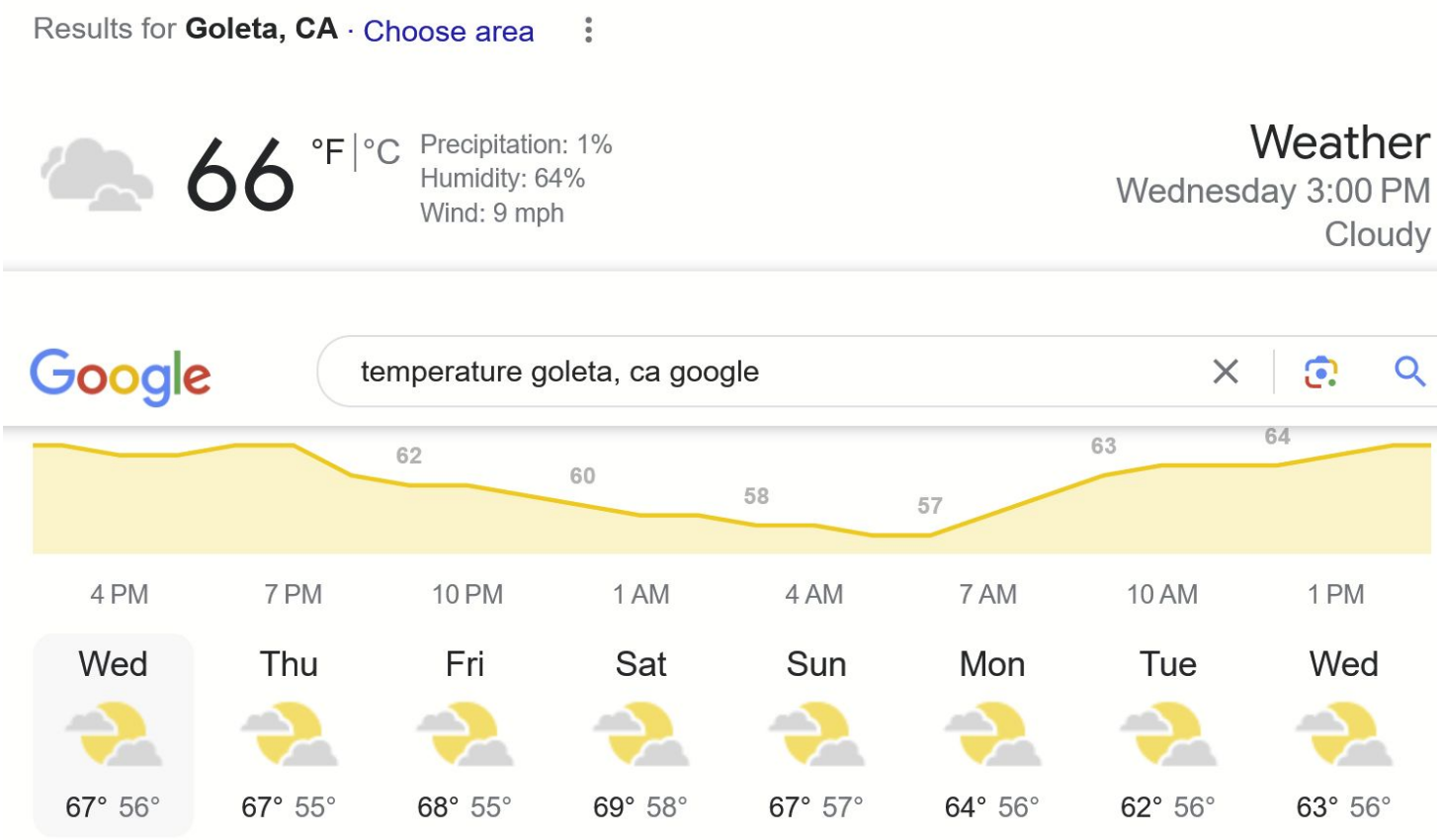
All in-person training information is housed below.
We will be using this link next week.

https://immerse-ucsb.github.io/in-person-training

Your quick, anonymous feedback is appreciated.
Here is a link

Any questions?

It might be colder here than you think…

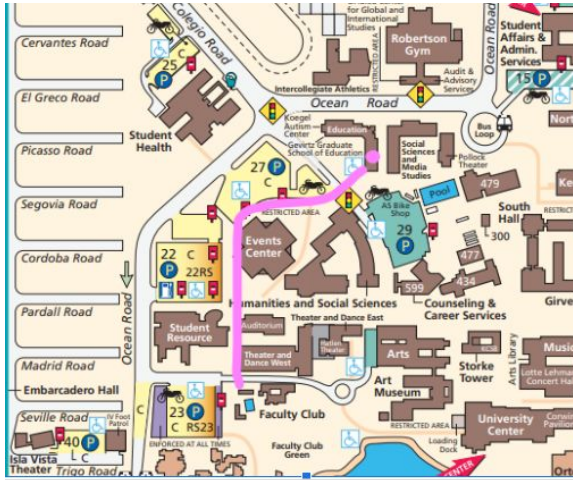


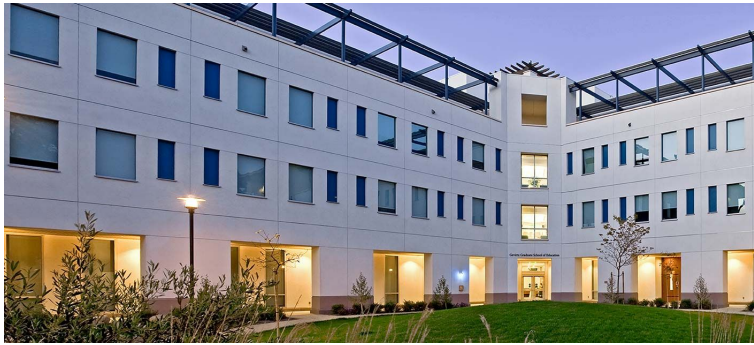
IMMERSE Project– IES funded Training Grant (R305B220021)

UC SANTA BARBARA

UCSB Visit: General Schedule

- Here is a draft
- Training will be be 9-5:00 with breaks and one hour lunch
- Location: Education 2209 and 4211
- 10 minute walk from hotel
- Lunch will be provided Monday and Tuesday, optional organized group order
- Dinners on own
- Materials will be share on GitHub account
- Optional activities;
  - Yoga at the beach (Monday)
  - Dinner in Downtown Santa Barbara (Tuesday)
  - Santa Barbara wine happy hour (Wednesday)





UCSB Visit

We plan to take pictures during the in-person training. The pictures will be used on twitter and project website.

If you do not want your picture to be used, please contact immerse@education.ucsb.edu

If we don't hear from you by Monday, we'll assume that it's ok to include your picture.

You might be on the same flight with another fellow (in case you want to coordinate travel from airport to hotel)

Here is a link to a google doc with travel information from some of the fellows

For those flying into Santa Barbara, it is about a 7 minute ride from airport to Club and Guest House at UCSB.

See you in Santa Barbara!

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B220021 to The Regents of the University of California, Santa Barbara. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.