R Notebook

# Pitt CSD R Workshop - Meeting notes

## Friday, Jan 28, 2022

Mike Dickey, Bharath Chandrasekaran, Emily Goldberg, Alex Swiderski, Sarah Wallace, Rob Cavanaugh

## Action Items

* **Start putting a survey that we pass along through secret CSD channels (or broadly advertise on twitter)**
* **Run it by Elaine Mormer to stakeholder perspective**
* **Meet again in 2 weeks**

## Notes from meeting

Mike:

* data management and visualization and simple descirptive stats using R providing a quick introduction to different kinds of data and basic data principles.
* A quick introduction for doing simple inferential statistics on those data. Ideally here’s glmer and lmer. Not thinking about mastery in any of these but giving people some basic tools that they can build on in the context of their home labs.

Alex:

* Why someone learns a computer language (to do data science). What’s challenging about learning about data science and a programming language is that you’re learning a new language and statistics at the same time.
* students need some basic skills. Load a data frame, get summary statistics, basic wrangling. The key being that you can focus on the stats because you have some basic programming skills. Perhaps you can do data visualization after that.

Emily:

* Agrees with Alex with respect to the basic data management/wrangling skills

Mike:

* goal is data management and simple wrangling in a programming language (like R) that will facilitate being able to expand these skills in subsequent classes.

Rob:

* troubleshooting, data management, best practices with R projects, how to google stuff, how to create a markdown document to share scripts.

Discussion:

* when to do this? Embed in an expanded PhD student orientation (week before classes start)
* A lot of effort for targeting two students. What do you get out of it?
* Incorporate the uniqueness of the CSD perspective/psychology perspective
* Go big. lacking in PhD programs across the US. Make it a bootcamp that Pitt PhD students could host every year for CSD PhD students throughout the country.
* Create a survey and share it widely with CSD programs in CSD.
* Enrollment of 30 CSD PhD students that could pilot the workshop.
* Also research SLPs who are working in research labs
* Limit to US students? No reason to do that now. But a more homogenous group to start with will be helpful.
* What is the unique contribution to the CSD specific field.
* Work data into a T32/ collect outcomes to iterate/disseminate on the outcomes
* Ask people to bring datasets to the workshop. better curation of what experiments look like.
* Could collect datasets for use and dissemination for learning.
* How many days should the workshop be…1, 2, 3?
* Stakeholder engagement is necessary - need to figure out what people want….do now
* If planning for the first time, just do it virtual
* Mike - mornings of that orientation week in August.
* Question about when universities start classes.
* Sustainability
  + is there a way to monetize so that there’s a financial incentive for PhD students to teach this in the future
    - and return funds back to the budget for the program.
  + there is a good pipeline of students that could step up
  + Opportunity to incorporate a younger PhD student to ‘co teach’ so that they could lead the following year
  + What is the cost you would charge.
  + Bharath can help through the administrative hassles
* Needs to be a ‘team’ putting on the course. presenter cant troubleshoot and teach. need multiple people.
* Inform Charity/Lauren. Possible in the future years could involve the SHRS data center.
* Key ingredients:
  + starting with something a little bigger reaching out across CSD PhD programs
  + CSD specific - datasets are specific to the field.
  + where between 2 and 30 feels like - can maintain the scale of personal/hands on introduction

#### Pre-meeting thoughts on Workshops

Define goals & scope of workshop(s)

* Target audience(s)
* Content areas
  + Some overarching context that should inform content:
    - Relevance to different career trajectories (academia, industry)
    - Relevance to different research areas within CSD
    - Potential for a basic course (with fall PhD orientation) and a more advanced course (?spring/summer)
  + Knowledge/skills
    - Beginner:
      * Basics of R/Rstudio/Jupyter software setup and use
      * Best practices for organization (e.g., use of Rprojects)
      * Basic R programming skills
      * Basic Data wrangling skills
      * Basic Plotting/Visualization in R
      * Trouble shooting & reprexes
      * Use of RMarkdown reports for sharing analyses
      * Basic knowledge of git/github
      * Making analyses reproducible for ASHA journal’s badges
    - Advanced:
      * Advanced wrangling, plotting
      * Generating tables and results text from statistical model objects for manuscripts
      * Creating shiny web-apps & interactive plots
      * Simulating data
  + Concrete deliverable(s)
    - Single analysis document/project
    - Reproducible ‘long-form’ abstract
    - Website creation
  + Practical considerations
    - Rstudio cloud vs. personal computers vs. jupyter/anaconda
    - Length of workshop (hours/days)
    - Single or multiple sessions
    - Ongoing support after learning?
* Resources for generating workshop content (don’t reinvent the wheel)
  + <https://psyteachr.github.io/>
  + <https://rstudio-education.github.io/hopr/>
  + <https://education.rstudio.com/learn/>
  + <https://r4ds.had.co.nz/>
  + <https://bookdown.org/yihui/rmarkdown-cookbook/>
  + <https://github.com/ujjwalkarn/DataScienceR>
* Plan for sustainability
* Action items moving forward