

# Welcome to the Workshop!

- Find a seat with a nametag, and plug in!
  - WiFi instructions printed on your table.
- Make sure you've completed the “**Setup**” steps:

<https://uw-madison-aci.github.io/2018-06-04-uwmadison-dc/>

- Get to know the people at your table.
- *Restrooms are out to the left, then a quick right.*

Welcome to



<https://github.com/UW-Madison-ACI/2018-06-04-uwmadison-dc>

# Introductions

# Workshop Impetus and Content

# Data Problems\*

- "Where did this data file come from?!"
- "Can you repeat the analysis, omitting subject X?"
- "You used the wrong data- the full dataset has more subjects."
- "The color-coding in this spreadsheet was off."

\*modified from Karl Broman's intro to [Tools for Reproducible Research](#)

# Our Goals:

## Reproducibility

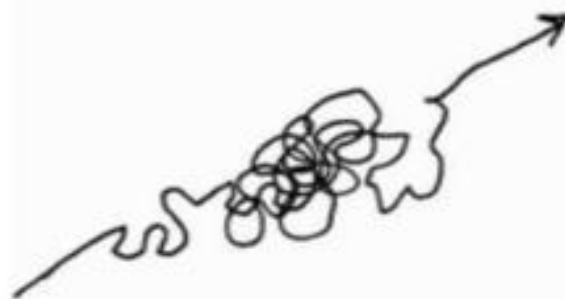
- Learn data analysis tools and best practices for organizing your data “studio”
- Two themes:
  - Make yourself a road map: take time to organize and document
  - Avoid doing things “by hand”: use tools that parse and analyze data, rather than editing data directly
- Take away something immediately useful

Success



what people think  
it looks like

Success



what it really  
looks like

Reproducibility vs. Replicability?



# What We'll Learn

- Data Organization and Formatting in Spreadsheets
- Data Cleaning/Reformatting
- *(Meta: Tips for Data Entry/Creation)*
- Data Management using a Database
- Automated Data Analysis
- Automated Data Visualization and Reporting

# Tools We'll Use

- Spreadsheet Program (Excel, or other)
- OpenRefine
- SQL
- R w/packages:
  - dplyr
  - ggplot2
  - knitr

# Workshop Logistics

# Workshop Format

- Hands-on, working with real data.
- Be prepared to:
  - (almost) always, follow along with the instructor on your own computer
  - work on exercises by yourself or with people near you
- Ask questions!
- Share notes in the etherpad:
  - <http://pad.software-carpentry.org/2018-06-04-uwmadison-dc>

# Nametags

- Your nametags are a visual feedback tool for instructors
- If you are lost, confused, or need help, put the red side up.
- If all is well, put up the green side up.

# Code of Conduct

- In order to facilitate the best learning for everyone, we expect everyone to abide by the Data Carpentry Code of Conduct:  
<http://www.datacarpentry.org/code-of-conduct/>

If you have any concerns, email ACI (Contact [ACI@lists.wisc.edu](mailto:ACI@lists.wisc.edu)), or contact the Data Carpentry staff listed on the CoC above.

# Workshop Logistics

- We will have breaks during the morning, afternoon and for lunch.
- Coffee, tea, and a morning snack are provided. Lunch is on your own.
  - We can try to find a fridge or microwave if you REALLY need it..
- **Restrooms** (and drinking fountains) are across the hall (right, then quick left).

# Meet your neighbors!

- **Talk to 1-2 people around you:**
  - summarize your research/work in three words (and then briefly describe it)
  - share one thing you hope to learn at this workshop



# Important Links

- **Workshop Website** (schedule, materials):  
<https://uw-madison-aci.github.io/2018-06-04-uwmadison-dc/>
- **Etherpad** (links, online notes + live chat):  
<http://pad.software-carpentry.org/2018-06-04-uwmadison-dc>
- **Twitter:**  
@datacarpentry, #dcuwm
- **All-Data Shortcut:** (most useful after the workshop)  
kbroman.org/datacarp