Here is what we want you to get out of this bootcamp:

- Think expansively about data -- data formats, data types, data sources - how can we create data from existing sources and analyze the data using innovative tools? In this, I will show you about web scraping and automation using examples from my own work, with a disclaimer that you will not be able to do this by the end of the week, but rather a longer project I can help you with in the future.
- What software is available that can facilitate the interpretation, analysis and visualization of data beyond the typical tools (e.g., SPSS) and what are the benefits and drawbacks
- Data wrangling refers to the process cleaning, merging, recoding and otherwise preparing data for analysis. This is the most important part of any analysis and the part nobody likes to do.
- Show me the data! There is tons of data available to download on the internet from reliable sources. I am going to review the most important ones that we can use to analyze and map. We will download and merge this data for analysis. Much of it can provide rich contextual information for your research, including your dissertation.

2. DAY 1

A. Intro to open source software

1. There are a wide variety of different open source and proprietary software tools that are simple to learn and make data management and analysis easier, more effective and fun. Here is my list of the most important tools to be familiar with. The items in bold are the software that we will focus on for the boot camp.

2. software

a. open source software

1. **R**: IMHO the best statistical analysis tool ever (yes, better than python for statistics)

2. **jamovi**: the next best option, built on an R framework, nobody will know you are NOT using R :)

3. **JASP**: another 'next best' option, in some ways better than jamovi, but it seems less stable

4. **QGIS**: do geospatial analysis like a (ArcGIS) pro

b. proprietary software that OSU is licensed for

1. **SPSS**: its like a necessary evil - everyone must know how to use SPSS. There are some things that are actually easier in SPSS

2. **SAS**: ugh!

3. **EXCEL**: can be great to clean your data, particularly if you use the built-in functions.

4. **ArcGIS Map, Pro**

5. **jmp**: this is a very cool data science program from the makers of SAS (the archaic and soon to be extinct software program)

c. Add-ons

1. **R data integration for SPSS**: use R functionality from SPSS

2. **NodeXL**: a software addon for excel that performs social network analyses

B. Intro to data & data types:

Before starting an analysis it is important to know the kind of data that you have. This requires understanding not only the distinction between qualitative and quantitative data, but also the difference between structured and unstructured data, time series data, spatial data, etc. We wills start by briefly reviewing the different types of data that are available and the multiple ways in which data can be measured. This is of upmost importance because the analysis you choose is driven by the type of data you have. In other words, you cannot make your data fit the analysis you know, but rather you must learn how to analyze the data you have.

. Qualitative v Quantitative

2. Structured v unstructured data

1. data types

3. Data innovation - being creating with data: here we address the question: what really constitutes data? Over the years, I have learned that even advanced students do not really know how to properly identify data that *can be used* in an analysis. How can we identify, and more importantly utilize, different data sources in our analysis.

1. [Los Angeles Homicide Report]([Victims in the last 12 months - The Homicide Report - Los Angeles Times (latimes.com)](latimes.com))

2. [Mapping Attacks on LGBTQ Rights in U.S. State Legislatures | American Civil Liberties Union (aclu.org)](aclu.org)

3. [Child Maltreatment Data and Statistics]([Child Maltreatment and Data Statistics – PCAKY](PCAKY)) by County in Kentucky

4. Other examples

B. Show me the data!

1. Finding and downloading data on the Internet

1. Major datasets for Social Work

3. Day 2: The Boring Stuff

A. Data wrangling

- entering data

- important data

- computing variables

- reshaping data

- Ex_1 Hospice Utilization - Patient Demographics - Hospice_Demographics_18-21_Dataset - California Health and Human Services Open Data Portal

- recoding & defining variables

- filtering cases

- merging data

Day 3: It's getting real! Today we will start analyzing the data we downloaded and cleaned in Days 1 & 2. The goal is not to learn/review inferential statistics. The goal for this section is to demonstrate how to be effective at describing your data -- which is the first step of all analyses -- Know thy data. Therefore, we learn how to present what should always be (imho) the first two tables in any publication -- the table of descriptive statistics (i.e., Table 1) and some representation of the relationships in your data (Table 2, Figure 1, etc.)

4. Analyzing data
    1. Describing data
        1. Using jamovi and JASP to display APA formatted results
            1. Table 1: Describing the variables
            2. Table 2: Describing relationships
    2. Innovative data visualization tools
        1. Scatterplots
        2. Correlation matrix
        3. Box plots & violin plots
        4. Dot charts
        5. Alluvial diagrams
        6. Age pyramids
        7. Venn diagrams
        8. Variable tree plots
5. Interpreting data
    1. telling a story with data
    2. when insignificant findings are significant
    3. publishing an article using data
    Day 4: Where in the world?
    1. Using open source GIS
    2. Data types
    3. Census data
    4. OpenStreetMap
    5. Merging public datasets
    6. Making Maps
6. Resources
    1. [learning statistics with jamovi (davidfoxcroft.github.io)](davidfoxcroft.github.io)