**Math Camp R Topics**

1. **Getting started**

* What is R? Why R? (Talk briefly about alternative software)
* Installing R
* What is RStudio?
* Why use RStudio to interface with R?
* Install RStudio
* TAKE AS MUCH TIME TO MAKE SURE EVERYONE HAS RSTUDIO INSTALLED AND RUNNING
* Getting to know the RStudio panels, show briefly how things operate in the console
* Create a .qmd file. Why do we work on markdown files?
* Objects (and big picture of object-oriented programming)
* Different types of objects: Vectors, functions on vectors
* Data frames and lists
* R packages: Install tidyverse
* Intersperse short exercises as you introduce new things

1. **Tidyverse I**

* Remind that last time we ended with R packages
* What is the tidyverse? Why do we use it?
* Load a data set (whatever you like, ideally on something that already exists in a package, like the penguins data set, we will talk about different data sources later)
* Selecting columns (this is the first time they would see a pipe operator, %>%, so take time to explain what is going on)
* Create columns with mutate
* Filter rows, an excuse to learn about logical operators
* Order rows with arrange
* Summarize data
* Summarize data by groups
* Visualizing data with ggplot
* Univariate, categorical and numerical
* Different kinds of bivariate plots

1. **Tidyverse II**

* Loading data from different formats
* RData vs rds files
* CSV
* Excel (readxl package)
* Stata, SPSS (haven package)
* Have an open ended conversation about where to get “real” data
* Grab a “real” data set that you like, nice opportunity to plug in your research
* Tabulate with count function
* Recoding variables (practice mutate again, now with if\_else and case\_when)
* Missing values
* Pivoting wider and longer (try to find a good motivation for this, maybe something like data with year observations in the columns, need to move it to the rows)
* Merging data sets, bind\_rows, merge, left\_join, right\_join, all kinds of joins!
* If enough time: More about ggplot, especially customizing.
* Practice as you introduce new things

1. **Sampling and simulation**

* Different kind of random sampling from distributions (e.g. runif, rnorm, rbinom, etc…)
* Different kind of distribution functions (e.g. rnorm vs. dnorm vs. pnorm)
* Sampling from existing data
* Why would we want to do sampling from distributions/data? Maybe have exercises that illustrate why these things are useful
* Custom functions (very small example, mostly to get used to the syntax)
* Loops (for, while, repeat)
* Generalizing loops (apply family of functions, map function on tidyverse
* When to use a loop? When to use a function? Talk about minimizing mistakes from copy pasting or extensive editing
* Maybe leave like 20 minutes for lingering questions/wrap up