ENSC 445/545 – Watershed analysis

Unit 2 – Data wrangling

In this unit we will begin doing some programming and data wrangling in R/RStudio. Data wrangling is a term used for “cleaning” up data sets, making calculations, and shaping data frames into the shapes we want (e.g., long or wide formats) for a particular analysis or visualization we will do. Generally speaking, in R once you have the data frame in the correct “shape” that analysis is relatively easy. We will generally be using tidyverse for our programming in the course and the data wrangling component of tidyverse is dplyr. When we wrangle data we are often making it “tidy”. When we tidy data we make the ensuing analysis much easier and often (but not always) we are shaping it into a “long” format.

Reading for this unit

Tidy data: <https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html>

Wide and long data formats: <https://www.stephenskalicky.com/pivoting-data.html#:~:text=Pivoting%20data%20using%20tidyverse%20functions&text=Wide%20data%20is%20where%20each,an%20individual%20observation%20per%20subject>.

To begin this unit follow this link to the GitHub repository:

<https://github.com/tpcovino/unit_2_data_wrangling>