WELCOME TO NDACAN MONTHLY OFFICE HOURS!

NATIONAL DATA ARCHIVE ON CHILD ABUSE AND NEGLECT
DUKE UNIVERSITY, CORNELL UNIVERSITY, & UNIVERSITY OF CALIFORNIA: SAN FRANCISCO

NDACAN

Children's Bureau
An Office of the Administration for Children & Families

- The session will begin at 11am EST
  - 11:00 - 11:30am – LeaRn with NDACAN (Introduction to R)
  - 11:30 - 12:00pm – Office hours breakout sessions
- Please submit LeaRn questions to the Q&A box
- This session is being recorded.
- See ZOOM Help Center for connection issues: https://support.zoom.us/hc/en-us
  - If issues persist and solutions cannot be found through Zoom, contact Andres Arroyo at aa17@cornell.edu.

LEARN WITH NDACAN

Presented by Frank Edwards

MATERIALS FOR THIS COURSE

- Course Box folder (https://cornell.box.com/v/LeaRn-with-R-NDACAN-2024-2025) contains
  - Data (will be released as used in the lessons)
    - Census state-level data, 2015-2019
    - AFCARS state-aggregate data, 2015-2019
    - AFCARS (FAKE) individual-level data, 2016-2019
    - NYTD (FAKE) individual-level data, 2017 Cohort
  - Documentation/codebooks for the provided datasets
  - Slides used in each week's lesson
  - Exercises as that correspond to each week's lesson
  - An .R file that will have example, usable R code for each lesson – will be updated and appended with code from each lesson

WEEK 2: "TIDYVERSE" FUNCTIONS

October 18, 2024

DATA USED IN THIS WEEK'S EXAMPLE CODE

- Census aggregate data from 2015-2019 (census_2015_2019.csv)
  - Population counts by state, year, sex, race, and ethnicity
  - Publicly available from CDC Wonder:
    - https://wonder.cdc.gov/single-race-population.html
- AFCARS aggregate data from 2015-2019 (afcars_aggreg_suppressed.csv)
  - Counts by state, year, sex, race/ethnicity of children in foster care; number of children removed due to physical or sexual abuse, or neglect; the number of children who entered or exited foster care in that year
  - Can order full data from NDACAN:
    - https://www.ndacan.acf.hhs.gov/datasets/request-dataset.cfm

TIDYVERSE

WHAT IS THE TIDYVERSE

- "tidyverse" is a special R package that contains a collection of R packages designed to streamline data exploration and analysis
- Core tidyverse installation includes packages: `ggplot2, dplyr, tidyr, readr, purr, tibble, stringr, forcats`
- The set of packages share common syntax, and data formatting and types, like tibbles (i.e., data frames)

ADDITIONAL TIDYVERSE RESOURCES

- For more details and background, and additional tidyverse packages see: https://www.tidyverse.org/
- A good text reference is Chapter 5 in *R for Data Science*, available here: https://r4ds.had.co.nz/

ADDITIONAL TIDYVERSE RESOURCES EXAMPLES

And many more!

THE PIPING OPERATOR

- The biggest difference with "tidyverse" programming is using the special syntax called "pipe" operators to make code cleaner, more intuitive, and consolidated

%>%

- NOTE: Newest versions of tidyverse can use the new pipe syntax |> and you may see this in documentation. They are functionally the same!

USING THE PIPING OPERATOR

- The operator %>% is used to sequentially apply functions to a data set
- Avoids having to overwrite or save many intermediate steps as usually happens with analogous base R constructions
- Compares to using $ in base R to reference variables
- DATA %>% select(var1, var2, var3) %>%

```
mutate(var4 = var1+var2/var3) %>%
rename(variable1 = var1,
       variable2 = var2,
       outcome = var3)
```

MOST OFTEN USED TIDYVERSE FUNCTIONS

- `select()`: select variables to keep/drop or to reorder variables
- `mutate()`: creates new variables that are functions of existing variables
- `filter()`: filter/subset data using logical operators over variables
- `summarize()`: summarizes data to 1 row per grouping variable with specified function (e.g. mean, sum, number of unique values)
- `arrange()`: reorder of data based on alphanumeric order of listed variables
- `group_by()`: grouping variables that functions can then be applied over (for example, grouping by race and to get means or totals by race)
- `rename()`: rename variable names

ADVANCED DATA MANIPULATION FUNCTIONS

- `full_join()`, `left_join()`, `right_join()`, `anti_join()`: join/merge data
- `pivot_wider()`, `pivot_longer()`: go from long to wider format, or vice versa
- `mutate_at()`, `mutate_if()`: apply same function to multiple variables at same time
- `summarize_at()`, `summarize_if()` : summarise multiple variables at same time over the same function and group