

WELCOME TO NDACAN MONTHLY OFFICE HOURS!

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



- 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 4: LINKING MULTIPLE YEARS

November 11, 2024

DATA USED IN THIS WEEK'S EXAMPLE CODE

- **AFCARS fake individual-level data (afcars_2016_indv_fake.csv, ...)**
 - Simulated (fake) individual-level data that mimics AFCARS structure. We will use the 2016, 17, and 18 files today
 - Can order full data from NDACAN:
 - <https://www.ndacan.acf.hhs.gov/datasets/request-dataset.cfm>

CONSIDERING TIME

SUBYR IN NCANDS

Variable Name: SubYr

Variable Label: Submission Year

Definition:

The submission year is the Federal Fiscal Year (FFY) 12-month period. The FFY is from October 1 through September 30. The year of data submission is based on the report disposition date. All report disposition dates fall within the submission year. The report date may have occurred in a previous fiscal year. All records have the same year in this field.

RPTDT IN NCANDS

Report Date

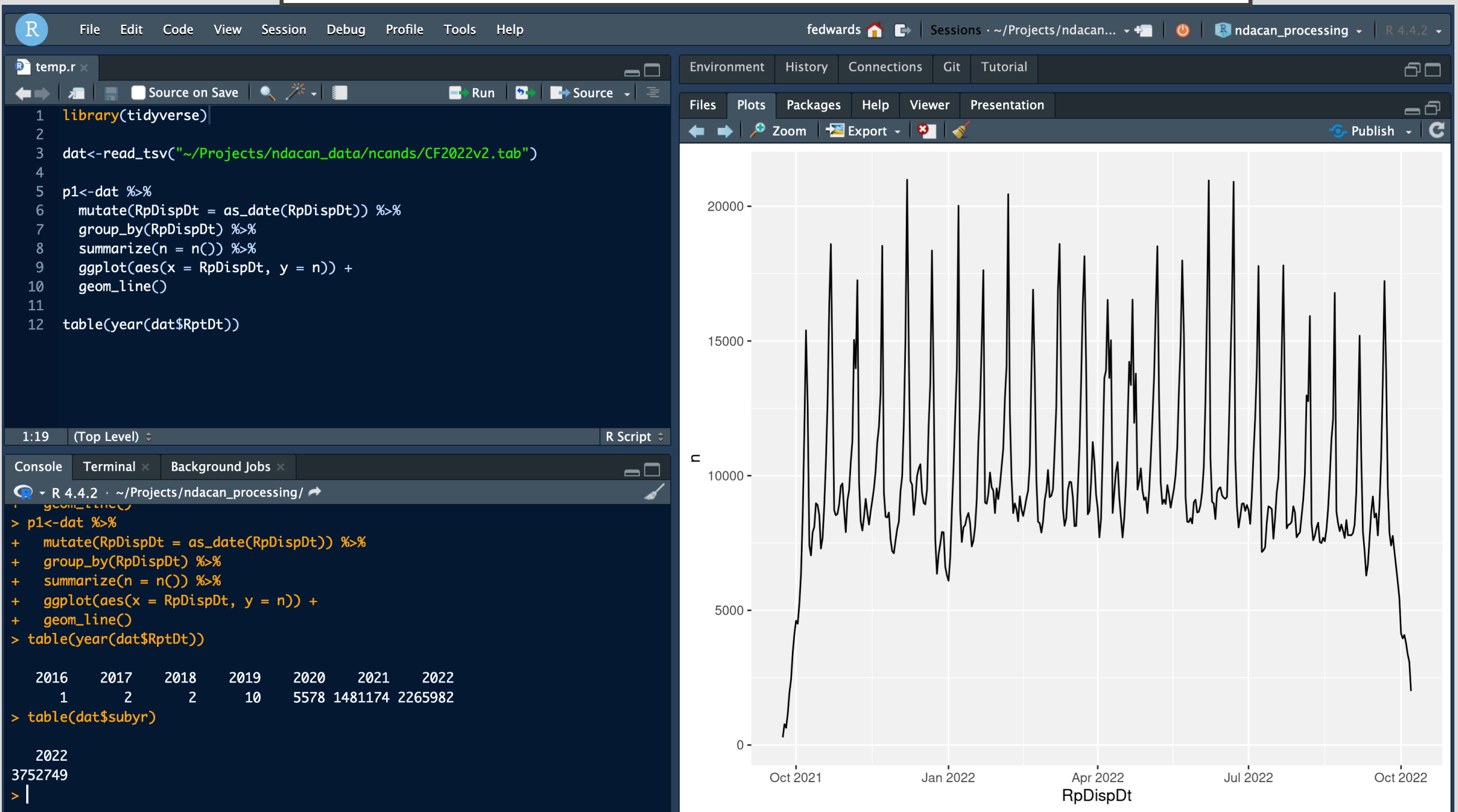
Variable Name: RptDt

Definition:

The month, day, and year that the responsible agency was notified of the suspected child maltreatment referral. This is the date that the report of maltreatment was made. If a state combines several allegations into one report, the date of the report is the date of the initial allegation. The determination as to whether additional allegations are considered part of the original report or whether they are considered a new report is left to each state's procedures.

To help assure confidentiality of the data, this date has been rounded to either the 8th or the 23rd of the month. All other dates in this data are then similarly adjusted by the same number of days to maintain the timespan between all dates.

R CODE AND PLOT



GUIDANCE

- **Carefully consider HOW you want to consider time.**
- **For both NCANDS and AFCARS, many cases that occur in the calendar year 2022 will not be present in the 2022 FY data**
- **For annualized analysis, treating SubYr or FY as equal to the calendar year is OK, if not exactly correct. With a long time series, and for large geographies, the 2 month mismatch isn't likely to cause much error**
- **But for smaller geographies and sub-annual analysis, care is required!**

OVER TO RSTUDIO

```
##### Joining multiple years of AFCARS data LeaRn demo
##### Author: Frank Edwards
##### Email: frank.edwards@rutgers.edu

# load tidyverse packages
library(tidyverse)

# read in data. I am using an RStudio project so that paths are relative
afcars16<-read_csv("./data/afcars_2016_indv_fake.csv")

afcars17<-read_csv("./data/afcars_2017_indv_fake.csv")

afcars18<-read_csv("./data/afcars_2018_indv_fake.csv")

# evaluate data structure
glimpse(afcars16)

# evaluate unique ID coverage across years
# with intersection operator for each pair of years
table(afcars16$id_num %in% afcars17$id_num)
table(afcars16$id_num %in% afcars18$id_num)
table(afcars17$id_num %in% afcars18$id_num)

# simple case: binding rows to combine into single long table
# Note: columns must identically match. harmonize beforehand if needed

afcars_full<-afcars16 %>%
  bind_rows(afcars17) %>%
  bind_rows(afcars18)
```

```
## lets evaluate the proportion of entries involving
```

```
## neglect allegations
```

```
afcars_full %>%  
  filter(entered_f == 1) %>%  
  group_by(fy) %>%  
  summarize(total_entries = n(),  
            neglect_entries = sum(neglect_f == "I")) %>%  
  mutate(entries_prop_neglect = neglect_entries / total_entries) %>%  
  ggplot(aes(x = fy, y = entries_prop_neglect)) +  
  geom_line()
```

```
# slightly more complex case: wide data by subject
```

```
# Maybe we want to know about children's placement setting trajectories
```

```
# start with all unique IDs across years
```

```
# then drop duplicates
```

```
wide_data<-data.frame(id_num =  
  c(afcars16$id_num,  
    afcars17$id_num,  
    afcars18$id_num)) %>%  
  distinct()
```

```
# now join 2016 curplset
```

```
wide_data<-wide_data %>%
```

```
  left_join(afcars16 %>%  
    select(id_num, curplset_f) %>%  
    rename(curplset_f_2016 = curplset_f))
```

and 2017

```
wide_data<-wide_data %>%  
  left_join(afcars17 %>%  
    select(id_num, curplset_f) %>%  
    rename(curplset_f_2017 = curplset_f))
```

and 2018

```
wide_data<-wide_data %>%  
  left_join(afcars18 %>%  
    select(id_num, curplset_f) %>%  
    rename(curplset_f_2018 = curplset_f))
```

here's what we get

```
head(wide_data)
```