Data Cleaning

## Importing IPUMS Data

# NOTE: To load data, you must download both the extract's data and the DDI  
# and also set the working directory to the folder with these files (or change the path below).  
  
if (!require("ipumsr")) stop("Reading IPUMS data into R requires the ipumsr package. It can be installed using the following command: install.packages('ipumsr')")

Loading required package: ipumsr

Warning: package 'ipumsr' was built under R version 4.3.3

ddi <- read\_ipums\_ddi("cps\_00005.xml")  
data <- read\_ipums\_micro(ddi)

Use of data from IPUMS CPS is subject to conditions including that users should cite the data appropriately. Use command `ipums\_conditions()` for more details.

## Loading Packages

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.4  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(vtable)

Loading required package: kableExtra  
  
Attaching package: 'kableExtra'  
  
The following object is masked from 'package:dplyr':  
  
 group\_rows

library(lubridate)

## Selecting Variables + Joining Indnames

dataSelected <- select(data, 'YEAR', 'MONTH', 'STATECENSUS', 'EMPSTAT', 'LABFORCE', 'WHYUNEMP', 'IND','FAMINC')  
  
dataSelected <- dataSelected %>% filter(IND != 0)  
  
indnames <- read.csv('indnames.csv')  
  
dataJoined <- inner\_join(dataSelected, indnames, by = c("IND" = "ind"))

## Retail + Information Data

dataFilteredRI <- dataJoined %>%  
 filter(indname == c('Retail Trade','Information'))

Warning: There was 1 warning in `filter()`.  
ℹ In argument: `indname == c("Retail Trade", "Information")`.  
Caused by warning in `indname == c("Retail Trade", "Information")`:  
! longer object length is not a multiple of shorter object length

dataYearRI <- dataFilteredRI %>% filter(YEAR <= 2022 & YEAR >= 2017)  
  
dataFilteredRI <- dataYearRI %>%  
 filter(YEAR != 2022 | MONTH <= 4)  
  
dataFinalRI <- dataFilteredRI %>%  
 mutate(date = make\_date(year = dataFilteredRI$YEAR, month = dataFilteredRI$MONTH, day =1))  
  
dataFinalRI <- dataFinalRI %>%  
 select(-YEAR, -MONTH)

## Retail + Finance/Insurance/Real Estate/Rental Leasing

dataFilteredFIRR <- dataJoined %>%  
 filter(indname == c('Retail Trade', 'Finance and Insurance, and Real Estate and Rental and Leasing'))

Warning: There was 1 warning in `filter()`.  
ℹ In argument: `indname == c("Retail Trade", "Finance and Insurance, and Real  
 Estate and Rental and Leasing")`.  
Caused by warning in `indname == c("Retail Trade",  
 "Finance and Insurance, and Real Estate and Rental and Leasing")`:  
! longer object length is not a multiple of shorter object length

dataYearFIRR <- dataFilteredFIRR %>% filter(YEAR <= 2022 & YEAR >= 2017)  
  
dataFilteredFIRR <- dataYearFIRR %>%  
 filter(YEAR != 2022 | MONTH <= 4)  
  
dataFinalFIRR <- dataFilteredFIRR %>%  
 mutate(date = make\_date(year = dataFilteredFIRR$YEAR, month = dataFilteredFIRR$MONTH, day =1))  
  
dataFinalFIRR <- dataFinalFIRR %>%  
 select(-YEAR, -MONTH)