Data Translation

## Data Cleaning

Loading Required Packages

library(vtable)

Loading required package: kableExtra

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::group\_rows() masks kableExtra::group\_rows()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(rio)  
library(ipumsr)

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

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')")  
  
ddi <- read\_ipums\_ddi("cps\_00002.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.

Data Cleaning

dataSelected <- select(data, 'YEAR', 'MONTH', 'STATECENSUS', 'EMPSTAT', 'LABFORCE', 'WHYUNEMP', 'IND')  
  
dataSelected <- dataSelected %>% filter(IND != 0)  
  
indnames <- read.csv('indnames.csv')  
  
dataJoined <- inner\_join(dataSelected, indnames, by = c("IND" = "ind"))  
  
datarFiltered <- dataJoined %>%  
 filter(indname == 'Retail Trade')  
  
write.csv(dataJoined, "joined\_data.csv", row.names = FALSE)  
  
write.csv(datarFiltered, "filtered\_and\_joined\_data.csv", row.names = FALSE)