Problem Set 1 - PSID: Labor Market Outcomes

Tate Mason

```r
psid_raw <- read_dta("~/SchoolWork/Y2S1/Macro/Data/PSID/PSID.dta")

cpi_data <- tibble(
  year = c(1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017),
  cpi_2020_ratio = c(1.54, 1.44, 1.37, 1.26, 1.14, 1.18, 1.10, 1.08, 1.07, 1.02)
)
```

```r
create_psid_panel <- function(data) {
  var_mapping <- tribble(
      ~year, ~age_var, ~sex_var, ~earnings_var, ~hours_var, ~wage_var, ~educ_var, ~ind_var, 
      # These are PSID variable codes - you MUST verify these match your data
      1999, "ER13010", "ER13011", "ER13218", "ER13363", "ER13224", "ER15937", "ER13216", "ER
      2001, "ER17013", "ER17014", "ER17229", "ER17393", "ER17235", "ER19998", "ER17227", "ER
      2003, "ER21017", "ER21018", "ER21153", "ER21356", "ER21159", "ER23435", "ER21146", "ER2
      2005, "ER25017", "ER25018", "ER25142", "ER25345", "ER25148", "ER27402", "ER25128", "ER2
      2007, "ER36017", "ER36018", "ER36147", "ER36350", "ER36153", "ER40574", "ER36133", "ER3
      2009, "ER42017", "ER42018", "ER42182", "ER42148", "ER42188", "ER46552", "ER42168", "ER4
      2011, "ER47317", "ER47318", "ER47495", "ER47456", "ER47501", "ER51913", "ER47480", "ER4
      2013, "ER53017", "ER53018", "ER53195", "ER53156", "ER53201", "ER57669", "ER53180", "ER5
      2015, "ER60017", "ER60018", "ER60210", "ER60171", "ER60216", "ER64821", "ER60195", "ER6
      2017, "ER66017", "ER66018", "ER66211", "ER66172", "ER66492", "ER70755", "ER66196", "ER6
  )

  data <- data %>%
    mutate(
      famid = coalesce(ER66009, ER60009, ER53009, ER47309, ER42009,
                       ER36009, ER25009, ER21009, ER17022, ER13019)
    )
    panel_data <- map_dfr(1:nrow(var_mapping), function(i) {
      year_info <- var_mapping[i,]
```

```
      year_data <- data %>%
        select(famid, all_of(unlist(year_info[-1]))) %>%
        setNames(c("famid", "age", "sex", "earning", "hours", "hourly_wage",
                   "education", "industry", 'lfp_status', "wealth")) %>%
      mutate(year = year_info$year) %>%
      filter(!is.na(famid))
    return(year_data)
    })
  return(panel_data)
}

psid_panel <- create_psid_panel(psid_raw)

glimpse(psid_panel)
```

```
Rows: 166,010
Columns: 11
$ famid       <dbl> 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4~
$ age         <dbl> 76, 47, 45, NA, 39, 34, 25, NA, 23, 22, 20, NA, NA, NA, NA~
$ sex         <dbl> 1, 1, 2, NA, 2, 1, 2, NA, 1, 1, 2, NA, NA, NA, NA, 1, 1, N~
$ earning     <dbl> 0, 400, 0, NA, 0, 600, 0, NA, 0, 0, 0, NA, NA, NA, NA, 500~
$ hours       <dbl> 0, 40, 65, NA, 45, 50, 0, NA, 40, 30, 40, NA, NA, NA, NA, ~
$ hourly_wage <dbl> 0.00, 0.00, 0.00, NA, 15.35, 0.00, 0.00, NA, 8.50, 8.00, 5~
$ education   <dbl> 3, 3, 1, NA, 1, 3, 1, NA, 3, 1, 1, NA, NA, NA, NA, 1, 1, N~
$ industry    <dbl> 0, 628, 669, NA, 907, 69, 0, NA, 139, 67, 669, NA, NA, NA,~
$ lfp_status  <dbl> 0, 0, 0, NA, 0, 0, 0, NA, 0, 0, 0, NA, NA, NA, NA, 0, 0, N~
$ wealth      <dbl> 91500, 26000, 413500, NA, 42000, 20000, 12700, NA, 3000, 1~
$ year        <dbl> 1999, 1999, 1999, 1999, 1999, 1999, 1999, 1999, 1999, 1999~
```