

Milestone 01

FirstName LastName

Environment

Update the YAML with your first and last name.

Load the packages from the code-along. (Install first if not already done so.)

```
library(here)
library(tidyverse)
library(haven)
library(gssr)
library(gssrdoc)
library(summarytools)
```

Load all years of the GSS data.

Then, use the `gss_get_yr()` function to load the data only for the 2024 survey respondents.
(*Hint:* Use the code from code-along-01.qmd)

```
# Load the data (will appear in your Global Environment pane)
data(gss_all)

# load the gss 2024 data (add your code below)
gss24 <- gss_get_yr(2024)
```

Fetching: https://gss.norc.org/documents/stata/2024_stata.zip

Load the GSS codebook

```
# Load the codebook
data(gss_dict)
```

Code, Output, Meaning

Run the code and add **narrative** (outside code chunks) or **comments** (inside the code chunks) describing what each of the below R chunks do.

```
here()
```

```
[1] "C:/Users/Joanna/Documents/GitHub/Stats for Sociologists"
```

```
list.files(path = here())
```

```
[1] "_extensions"          "_quarto-speaker.yml"
[3] "_quarto.yml"          "code-alongs"
[5] "data"                 "docs"
[7] "lectures"             "memos"
[9] "milestones"           "project"
[11] "slides"               "SOC6302_data.qmd"
[13] "SOC6302_punch-list.qmd" "SOC6302_readings.qmd"
[15] "SOC6302_syllabus.qmd" "Stats for Sociologists.Rproj"
[17] "tutorials"
```

```
list.files(path = here("project"))
```

```
[1] "data"      "outputs" "scripts"
```

```
Sys.time()
```

```
[1] "2025-08-19 19:05:43 EDT"
```

Use R to complete the checkpoints below. Show your work (e.g., R code chunks) where appropriate. Add narrative (text outside code chunks) or comments (text inside the code chunks) throughout.

Reference specific statistics (where appropriate) from your output to justify your answers. Explain what the values tell you about the data; interpret their meaning in relation to the question.

Use the `gss_all` dataset for checkpoints 01 - 07.

Use the `gss24` dataset for checkpoints 08 - 10.

Checkpoint 01

Report the survey question that corresponds to the `fefam` variable. How did you find out?

Checkpoint 02

Is `fefam` a dichotomous, nominal, ordinal, or interval-ratio variable? Justify your answer.

Checkpoint 03

Use `gss_which_years()` below to show the survey years with the variable `fefam`.

```
gss_which_years(gss_all, fefam)
```

```
# A tibble: 35 x 2
  year      fefam
  <dbl> <lgl>
1 1972    FALSE
2 1973    FALSE
3 1974    FALSE
4 1975    FALSE
5 1976    FALSE
6 1977     TRUE
7 1978    FALSE
8 1980    FALSE
9 1982    FALSE
10 1983    FALSE
# i 25 more rows
```

Which years of the survey did the `fefam` variable appear in the GSS?

Checkpoint 04

Use `print_labels()` to see a list of the values and labels for the `fefam` variable.

```
print_labels(gss_all$fefam)
```

```

Labels:
  value          label
    1      strongly agree
    2          agree
    3      disagree
    4  strongly disagree
NA(d)          don't know
NA(i)          iap
NA(j)  I don't have a job
NA(m)          dk, na, iap
NA(n)          no answer
NA(p)      not imputable
NA(r)          refused
NA(s)      skipped on web
NA(u)          uncodeable
NA(x) not available in this release
NA(y)  not available in this year
NA(z)          see codebook

```

If a respondent has a value of 4 for the `fefam` variable, what does that mean?

Checkpoint 05

Convert the `fefam` variable to a factor variable.

1. Use `zap_missing()` to get rid of all the 'missing' (NA) levels.
2. Use `as_factor()` to replace the values with labels.
3. Use `drop_levels()` to get rid of the empty levels (if any).

```

gss_all$fefam <- zap_missing(gss_all$fefam)
gss_all$fefam <- as_factor(gss_all$fefam)
gss_all$fefam <- drop_levels(gss_all$fefam)

```

Checkpoint 06

Use `table()` to show the frequencies of the `fefam` variable.

```

table(gss_all$fefam)

```

strongly agree	agree	disagree	strongly disagree
2977	10331	16097	7888

Checkpoint 07

Use `freq()` to create a (pretty) relative frequency table to show for the `fefam` variable using the 2024 data.

```
freq(gss_all$fefam, report.nas = FALSE, headings = FALSE)
```

	Freq	%	% Cum.
strongly agree	2977	7.98	7.98
agree	10331	27.70	35.68
disagree	16097	43.16	78.85
strongly disagree	7888	21.15	100.00
Total	37293	100.00	100.00

How many respondents “strongly agree” with the `fefam` statement in 2024? What % of respondents either “strongly agree” or “agree”?

Checkpoint 08

Create a (pretty) relative frequency table for the `happy` variable using the 2024 data.

```
freq(gss24$happy, report.nas = FALSE, headings = FALSE)
```

Tagged NA values were detected and will be reported as regular NA; use `haven::as_factor()` to

	Freq	%	% Cum.
very happy [1]	684	20.85	20.85
pretty happy [2]	1892	57.67	78.51
not too happy [3]	705	21.49	100.00
Total	3281	100.00	100.00

Were most respondents “not too happy,” “pretty happy,” or “very happy” in 2024?

Checkpoint 09

Create a (pretty) cross-tab of the **happy** and **sex** variables using only the 2024 survey data. Make sure your values are labelled for easy interpretation.

Use `zap_missing()`, `as_factor()`, `droplevels()`, and `ctable()` as appropriate.

```
# Fix the sex variable
gss24$sex <- zap_missing(gss24$sex)
gss24$sex <- as_factor(gss24$sex)
gss24$sex <- droplevels(gss24$sex)

# Fix the happy variable
gss24$happy <- zap_missing(gss24$happy)
gss24$happy <- as_factor(gss24$happy)
gss24$happy <- droplevels(gss24$happy)

# Create a cross-tab
ctable(gss24$happy, gss24$sex,
  prop = "c",
  format = "p",
  useNA = "no"
)
```

Cross-Tabulation, Column Proportions

happy * sex

Data Frame: gss24

	sex	male	female	Total
happy				
very happy		293 (20.2%)	386 (21.3%)	679 (20.8%)
pretty happy		830 (57.1%)	1054 (58.2%)	1884 (57.7%)
not too happy		330 (22.7%)	370 (20.4%)	700 (21.5%)
Total		1453 (100.0%)	1810 (100.0%)	3263 (100.0%)

Were women or men more likely to report they were very happy?

Checkpoint 10

According to the table above, were a greater percentage of women or men not too happy in 2024?

Meet the IPUMS Data

To keep your Research Brief progress on track, you'll complete short exercises that correspond with the new course material using your own dataset as part of your milestones.

Consider a research topic you are interested in and explore the documentation for some of the available [IPUMS \(microdata\) datasets](#) that could be relevant for the subject. Start broadly, such as looking at the [Topical Availability Summary](#). Then, look at a couple of individual datasets more in depth. Explore the sidebar of available documentation for 2-3 datasets: About, Data, Documentation (including User Guides and Questionnaires, FAQs).

- Write 1-2 sentences about your research interests.
- Then, write 2-3 sentences about which IPUMS dataset you are thinking of using, explaining the sample population and time frame of the dataset, and why it is a good fit for your research interests.
- Describe 2-3 key variables available in the dataset that are a good fit for your research topic. Report the exact survey questions and response options for these variables.
- Finally, create an IPUMS account for the dataset of interest. See this [help video](#) for details.

NOTE: For your Research Brief, you must choose an [IPUMS \(microdata\) dataset](#) to use.