

SOC 4015/5050: Lab-08 - Factors

Christopher Prener, Ph.D.

Fall 2018

Directions

Please complete all steps below. All work should be uploaded to your GitHub assignment repository by 4:15pm on Monday, October 29th, 2018.

Analysis Development

Using RStudio and your operating system's file manager, create an R Project in the *existing* directory in your assignments repository named Lab-08. Add a README.md file, notebook, and all necessary folders before beginning.¹

¹ This initial section follows the project workflow that is available in the lecture-03 repo!

Part 1: Auto Data

1. Edit a copy of the driveStr variable so that it is more descriptive. It should, however, remain a string/character variable. Use the following new value labels:

- TWD, Rear when driveStr is R
- AWD when driveStr is A
- TWD, Front when driveStr is F
- FWD when driveStr is 4
- FWD, Part-time when driveStr is P

Your code should include the creation of a frequency table with rounding to the thousandths place for the percentages.

2. Create a copy of the variable constructed in the previous question, but reformat it so that it is a factor variable. The levels should be in the following order: TWD, Front, TWD, Rear, AWD, FWD, Part-time, and AWD. Once again, your code should include the creation of a frequency table with rounding to the thousandths place for the percentages.
3. Now, write the frequency table you've already displayed within your R Markdown notebook to a csv file so that you can easily

move it into another document (like a presentation or a paper).
You can pipe the table directly into `readr::write_csv()`:

```
df %>%
  tabyl(var) %>%
  adorn_pct_formatting(digits = 3) %>%
  write_csv(here("results", "var_freq.csv"))
```

Part 2: GSS Data

5. Using the variable `HRS1`, recode the value -1 to missing (i.e. NA).
6. Create two new variables - `SEX` (which is a factor version of the current `SEX` variable and `FEMALE` (which is a logical variable that is `TRUE` when the respondent is female). Display frequency tables for `SEX` and `FEMALE` in your notebook.
7. Using the new factor variable `SEX`, create a bar plot and save the plot to your `results/` directory.
8. Create a new factor version of `MARITAL` that does not include observations when `MARITAL == 9` (i.e. the are missing). The valid categories should be:
 - Married (1) when `MARITAL` is 1
 - Widowed (2) when `MARITAL` is 2
 - Divorced (3) when `MARITAL` is 3
 - Separated (4) when `MARITAL` is 4
 - Never married (5) when `MARITAL` is 5
9. Finally, export the frequency tables for both `SEX` and `MARITAL` to `.csv` files, saved in `results/`. You should also save the final analytic data set you've created to the `data/` folder as a `.csv` file.