

# SOCI 385 - Social Statistics

Fall 2021

## Assignment 3

*This assignment is due via Canvas by 10:00 AM on Monday, October 4, 2021.*

*Please show all your code and write your responses in complete sentences as appropriate. Submit your knitted pdf and the R Markdown (.Rmd) file for your notebook. Remember to include your name in the header.*

This week you will continue practicing with descriptive statistics. This assignment requires the `assignment_03.csv` file which includes data from the 2008-2018 waves of the General Social Survey. Your dataset should have 7,946 observations and 5 variables.

The variables are:

`id`: a unique identifier for each respondent;

`hrs1`: the number of hours respondents reported working in the previous week (restricted to those who had part- or full-time jobs), from 1 to 89;

`class`: respondent's self-reported social class, categorized as Lower, Working, Middle, or Upper;

`racehisp`: respondent's self-reported race and Hispanic identification, categorized as Black, Hispanic, Other, or White;

`region`: respondent's geographic location, categorized as Middle Atlantic, Midwest, New England, South, or West

## Questions

1. Create a histogram showing the distribution of hours worked. Remember to label your axes and provide a title and subtitle. Use your histogram to briefly summarize the center and shape of this distribution. Note: don't waste time adjusting the binwidth; you can use the default of 30 here.
2. Create a figure with overlapping density plots of hours worked for each social class. Use this plot to roughly compare the probabilities that respondents from each class will work 40 hours a week (it's the mode for each class).
3. Create a new factor variable for hours worked with the following categories: less than 20, 20-39, 40, more than 40. The levels should be in order from least to most hours.
4. Using the new variable from #3, create a table showing the proportion of respondents in each region who are in each category of hours worked. Which region has the highest proportion of respondents who work more than forty hours a week?

5. Use the `group_by()` and `summarize()` functions to find the standard deviation of hours worked by race/ethnicity. In a sentence, describe any similarities or differences you notice. You do not have to create a figure for this question.
6. Approximately how long did it take you to complete this assignment?