

Assignment 3

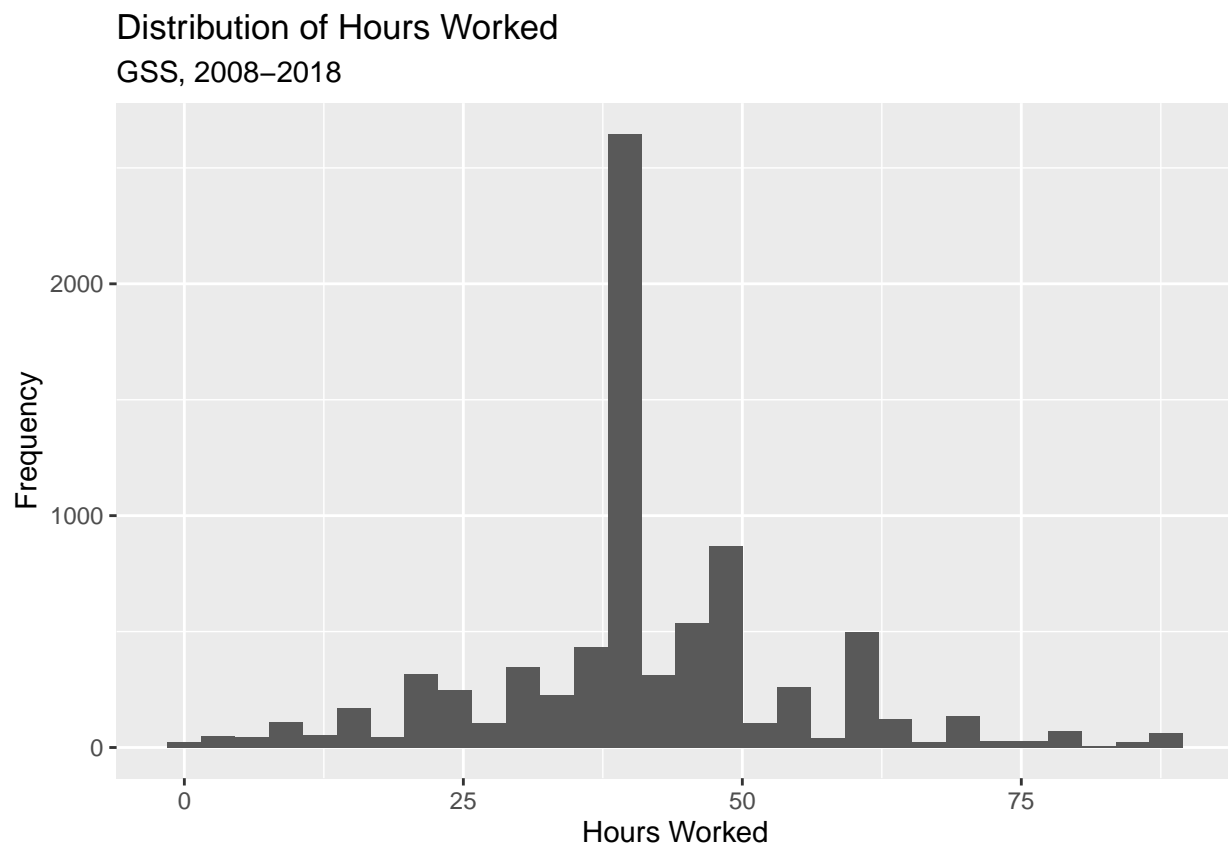
Matt Lawrence

October 4, 2021

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.

```
hours_plot <- ggplot(assignment_3, aes(x = hrs1))  
  
hours_plot + geom_histogram() +  
  labs(x = "Hours Worked", y = "Frequency",  
        title = "Distribution of Hours Worked",  
        subtitle = "GSS, 2008-2018")
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

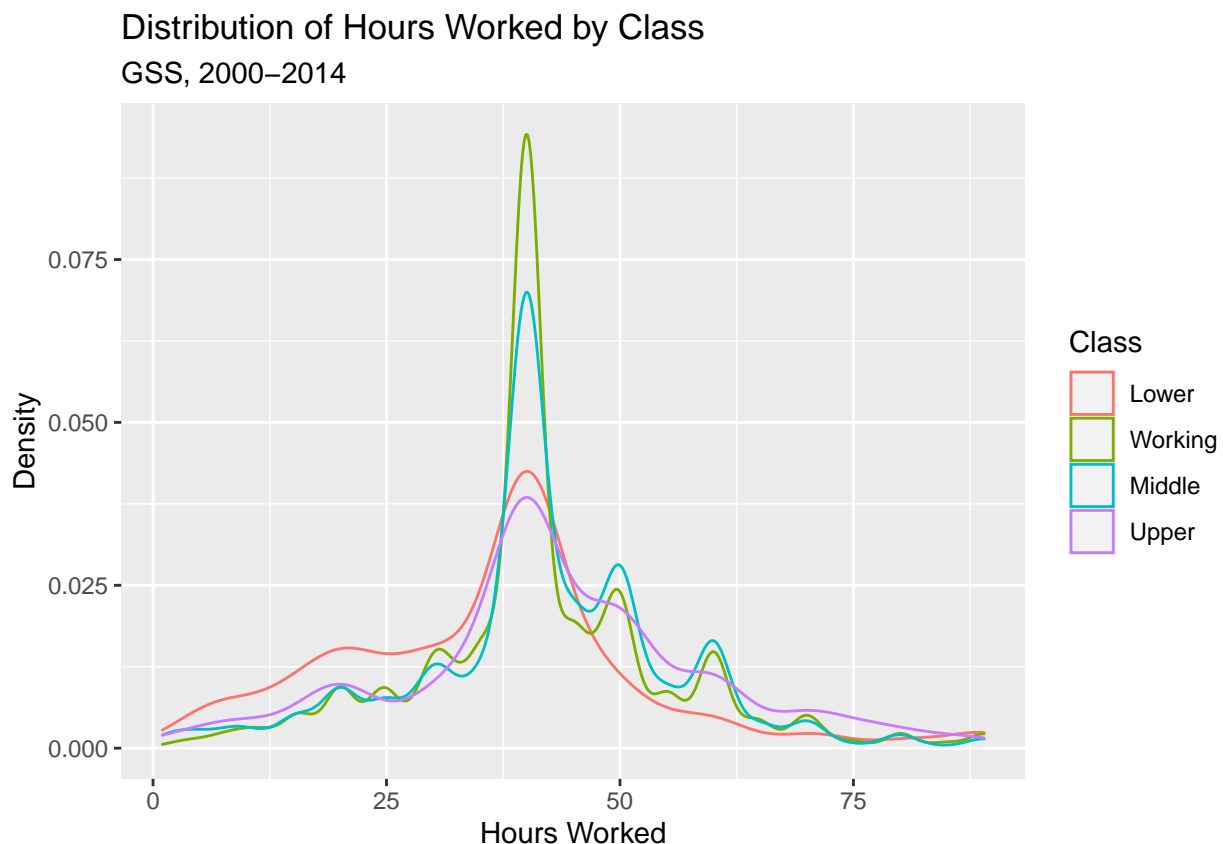


2. Create a figure with overlapping density plots of hours worked for each social class. Use this plot to roughly compare the probabilities that respondents from each class will work approximately 40 hours a week (it's the mode).

```
assignment_3 <- assignment_3 %>%
  mutate(class = factor(class,
    levels = c("Lower", "Working", "Middle", "Upper"))) # Remember to put the class categories in

hours_class_plot <- ggplot(assignment_3, aes(x = hrs1, color = class))

hours_class_plot + geom_density() +
  labs(x = "Hours Worked", y = "Density",
    title = "Distribution of Hours Worked by Class",
    subtitle = "GSS, 2000-2014",
    color = "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.

```
assignment_3 <- assignment_3 %>%
  mutate(hours_cat = ifelse(hrs1 < 20, "Less Than 20",
    ifelse(hrs1 %in% 20:39, "20-39",
      ifelse(hrs1 == 40, "40",
        "More Than 40"))),
    hours_cat = factor(hours_cat,
      levels = c("Less Than 20", "20-39", "40", "More Than 40")))
```

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 smallest proportion of respondents who work more than forty hours a week?

```
round(prop.table(
  table(assignment_3$region, assignment_3$hours_cat),1),3)
```

```
##
##           Less Than 20 20-39    40 More Than 40
## Middle Atlantic      0.070 0.239 0.320      0.371
## Midwest              0.061 0.206 0.299      0.434
## New England          0.065 0.270 0.288      0.378
## South                0.054 0.234 0.325      0.388
## West                 0.075 0.234 0.315      0.376
```

```
# Region should be rows, hours should be columns
```

5. Use the `group_by()` and `summarize()` functions in `dplyr` 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.

```
hours_race <- assignment_3 %>%
  group_by(racehisp) %>%
  summarize(sd_hours = round(sd(hrs1),3))
```

Optional Figure

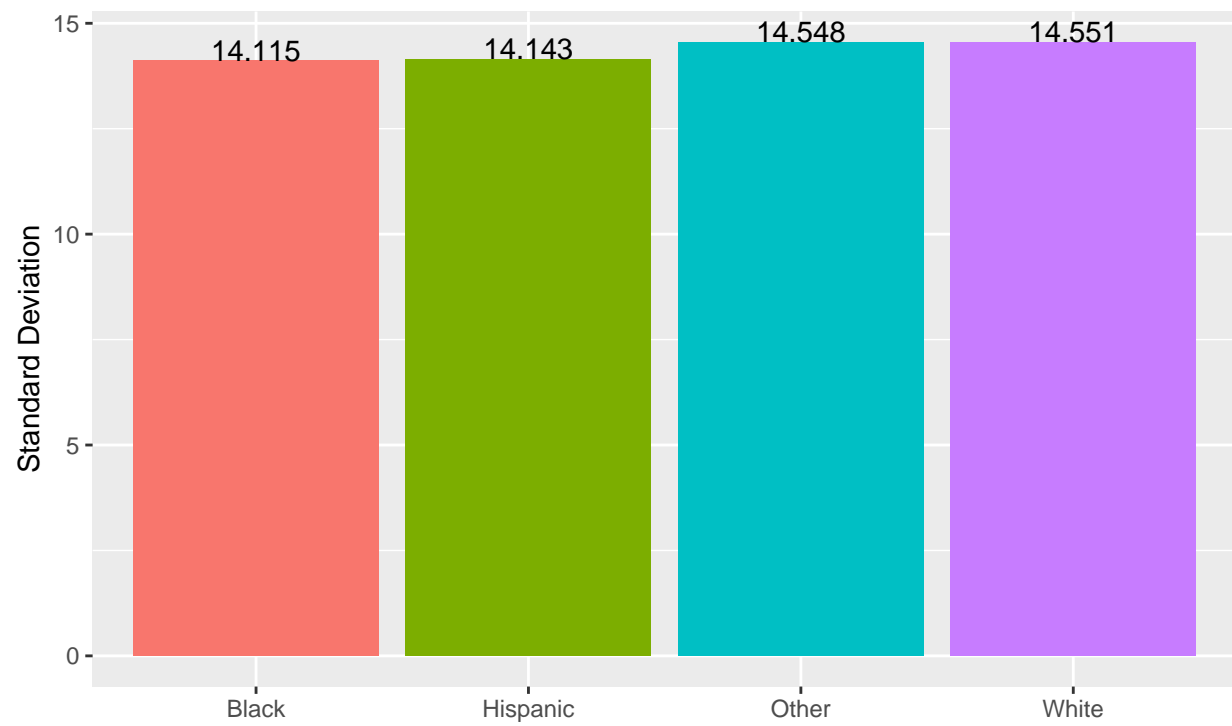
```
hours_race_plot <- ggplot(hours_race,
  aes(x = racehisp, y = sd_hours,
    fill = racehisp))

hours_race_plot + geom_col() +
  geom_text(aes(label = sd_hours, vjust="outward")) +
  labs(x = "", y = "Standard Deviation",
    title = "Standard Deviation in Hours Worked",
    subtitle = "General Social Survey, 2000-2014") +
  guides(fill = FALSE)
```

```
## Warning: 'guides(<scale> = FALSE)' is deprecated. Please use 'guides(<scale> =
## "none")' instead.
```

Standard Deviation in Hours Worked

General Social Survey, 2000–2014



Knitting A Table Without Kable

```
##
##           Less Than 20 20-39    40 More Than 40
## Middle Atlantic      0.070 0.239 0.320      0.371
## Midwest              0.061 0.206 0.299      0.434
## New England          0.065 0.270 0.288      0.378
## South                0.054 0.234 0.325      0.388
## West                 0.075 0.234 0.315      0.376
```

Knitting A Table With Kable

	Less Than 20	20-39	40	More Than 40
Middle Atlantic	0.070	0.239	0.320	0.371
Midwest	0.061	0.206	0.299	0.434
New England	0.065	0.270	0.288	0.378
South	0.054	0.234	0.325	0.388
West	0.075	0.234	0.315	0.376