STS 112: Lab 5

Code

Selamawit Berekat

This lab will require you to make column graphs and line graphs to explore women's role in households and labor markets. Be sure to make all graphs publication quality (informative titles, appropriate axis labels, legend titles, attractive theme, appropriate color scheme).

Getting started

Load the tidyverse package and read in the data for Week 5 as a data frame called `ipums`. Remove records for Alaska and Hawaii before 1960.

Hide

```
library(tidyverse)
ipums = read.csv("wk5.csv", header = T)

ipums <- ipums %>% filter(YEAR > 1950 | !STATEFIP %in% c(2, 15)) #removes
ipums %>% filter(YEAR < 1960 & STATEFIP %in% c(2, 15)) #test
```

0 rows | 1-10 of 11 columns

Women's occupations

You are going to make a column chart showing occupational change for U.S. women aged 15-64 from 1900 to 1990 as you did in Notebook 5, but faceting by four race categories (White, Black, Asian/Pacific, Other). Start by creating a new data frame called `women` that includes only women aged 15-64. Add two new variables: `OCCUP` (the same occupational categories as in Notebook 5) and `RACEF` (a factor variable for race with the four categories listed above).

Hide

```
women <- ipums %>% filter(SEX == 2 & AGE %in% 15:64) %>%
           mutate(OCCUP = OCC1950 %/% 100 + 1) %>%     #Integer division/floor removes r
emainder
           mutate(OCCUP = ifelse(OCCUP == 9, 2,
                         ifelse(OCC1950 %in% 910:979, 9, OCCUP))) %>%
           mutate(OCCUP = factor(OCCUP, labels = c("Prof/Tech", "Farming", "Managers",
"Clerical",
                                                   "Sales", "Crafts", "Operatives", "Ser
vice", "Laborers", "None"))) %>%
  mutate(RACEF = ifelse(RACE < 3, RACE,
                 ifelse(RACE %in% 4:6, 3, 4))) %>%
  mutate(RACEF = factor(RACEF, labels = c("White", "Black", "Asian/Pacific", "Other")))


#Test:
sort(unique(women$OCCUP)); sort(unique(women$RACEF))
```

```
 [1] Prof/Tech  Farming    Managers   Clerical   Sales      Crafts     Operatives Servic
e
 [9] Laborers   None
10 Levels: Prof/Tech Farming Managers Clerical Sales Crafts Operatives Service ... None
[1] White         Black         Asian/Pacific Other
Levels: White Black Asian/Pacific Other
```

Hide

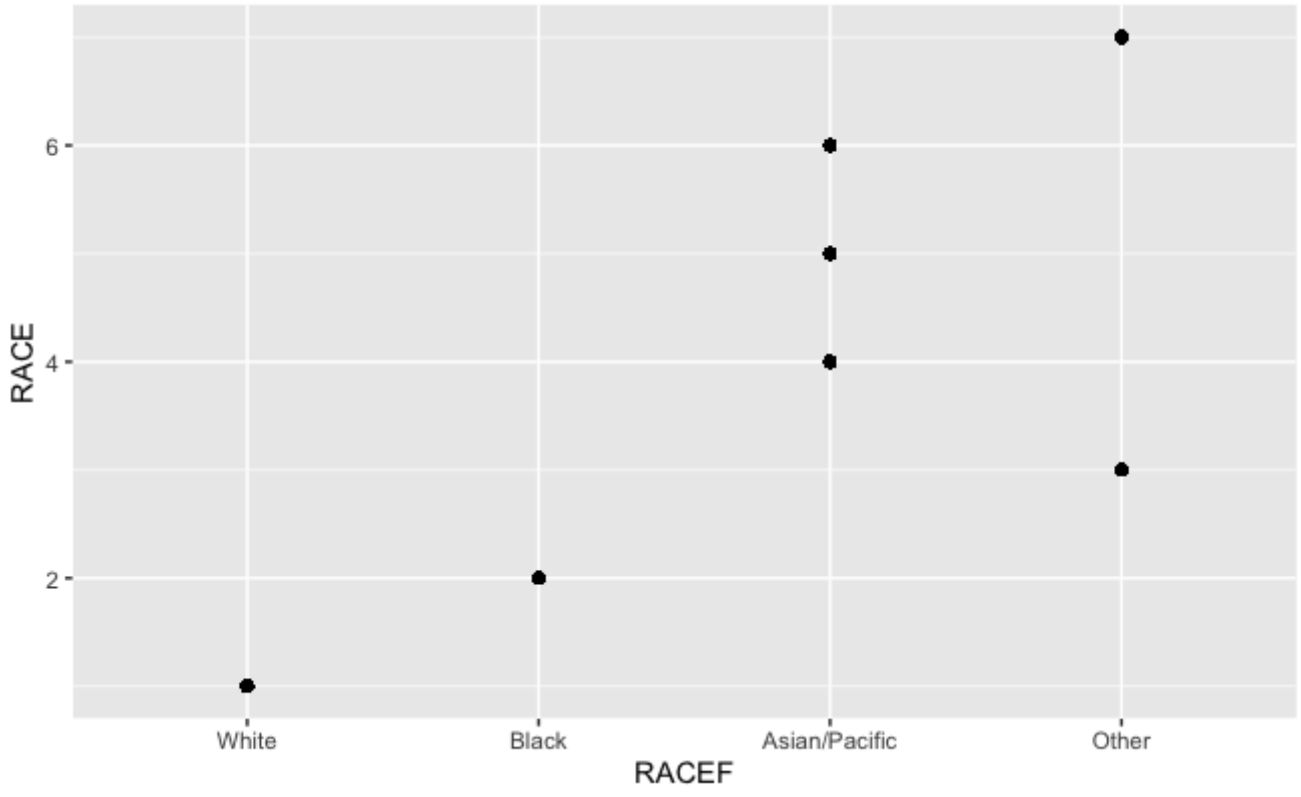
```
table(women$RACEF, women$OCCUP)
```

```
               Prof/Tech Farming Managers Clerical   Sales  Crafts Operatives Service
  White           357800   43683   120078   706934  187976   42237      323578  405435
  Black            29966   34624     6856    50614    7765    4491       40315  135345
  Asian/Pacific     7702     486     2272     9999    2135     699        4750    6256
  Other             3838    1290     1659     9383    1579     881        6092    9260

               Laborers    None
  White           28683 2172374
  Black            6710  227387
  Asian/Pacific     418   13707
  Other             953   19432
```

Hide

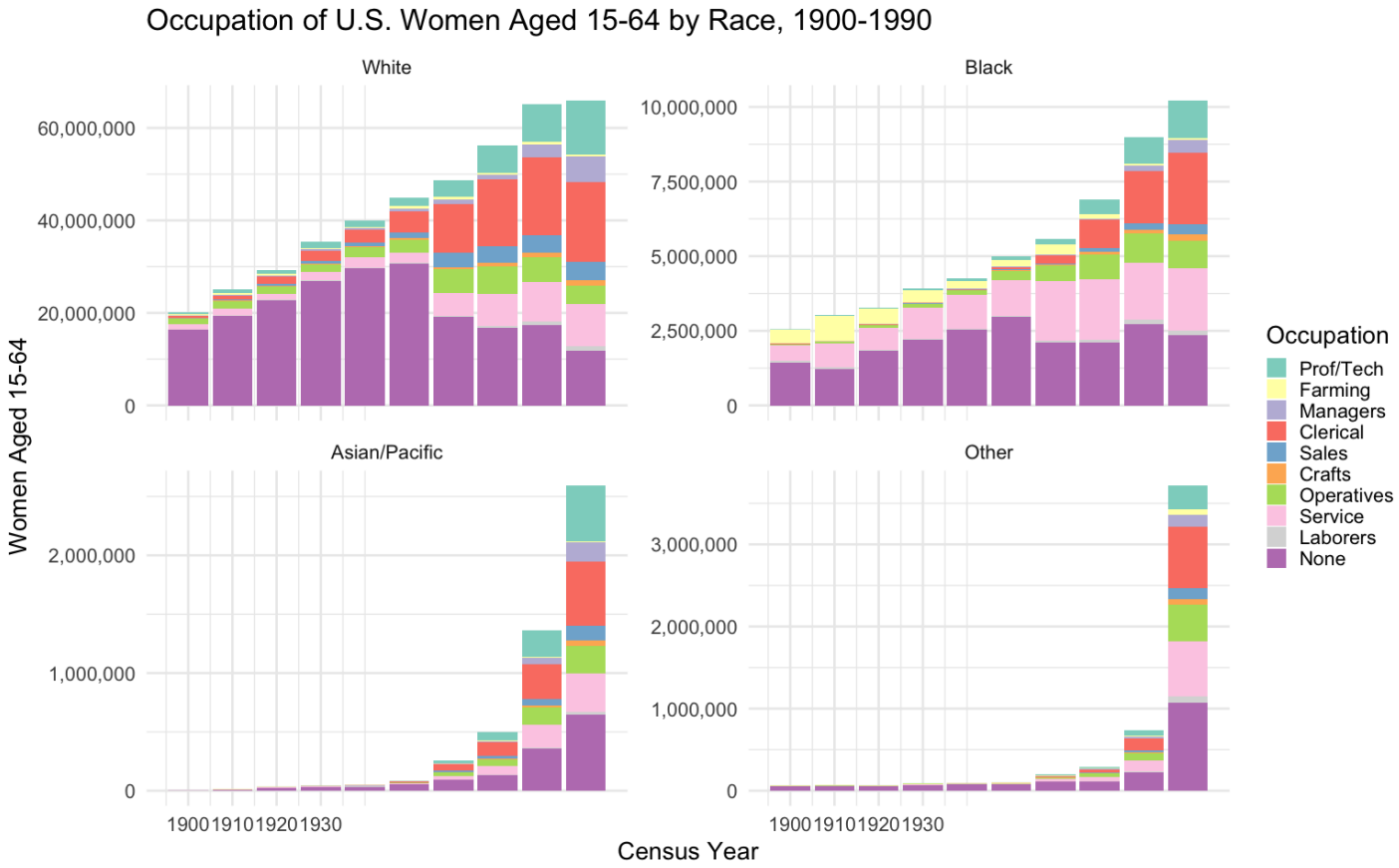
```
qplot(RACEF, RACE, data = women)
```



Now make a publication-quality graph of `OCCUP` by `YEAR`, faceting by `RACEF`. Do whatever you need to do to make all facets visible and all text readable. Remember to add an informative title.

Hide

```
women %>% group_by(YEAR, RACEF, OCCUP) %>% count(wt = PERWT) %>%
  ggplot(aes(YEAR, n, fill = OCCUP)) + geom_col() +
  scale_x_continuous(breaks = c(1870, 1880, 1900, 1910, 1920, 1930)) +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal(base_size = 20) + scale_fill_brewer(palette = "Set3") +
  facet_wrap(vars(RACEF), scales = "free_y") +
  labs(x = "Census Year", y = "Women Aged 15-64", fill = "Occupation",
       title = "Occupation of U.S. Women Aged 15-64 by Race, 1900-1990")
```



Hide

```
women
```

YEAR <int>	STATEFIP <int>	PERWT <dbl>	RELATE <int>	RELATED <int>	SEX <int>	AGE <int>	MARST <int>	RACE <int>	RACED <int>
1900	1	101	2	201	2	41	1	1	100
1900	1	101	2	201	2	27	1	1	100
1900	1	101	12	1215	2	17	6	2	200
1900	1	101	12	1214	2	21	2	2	200

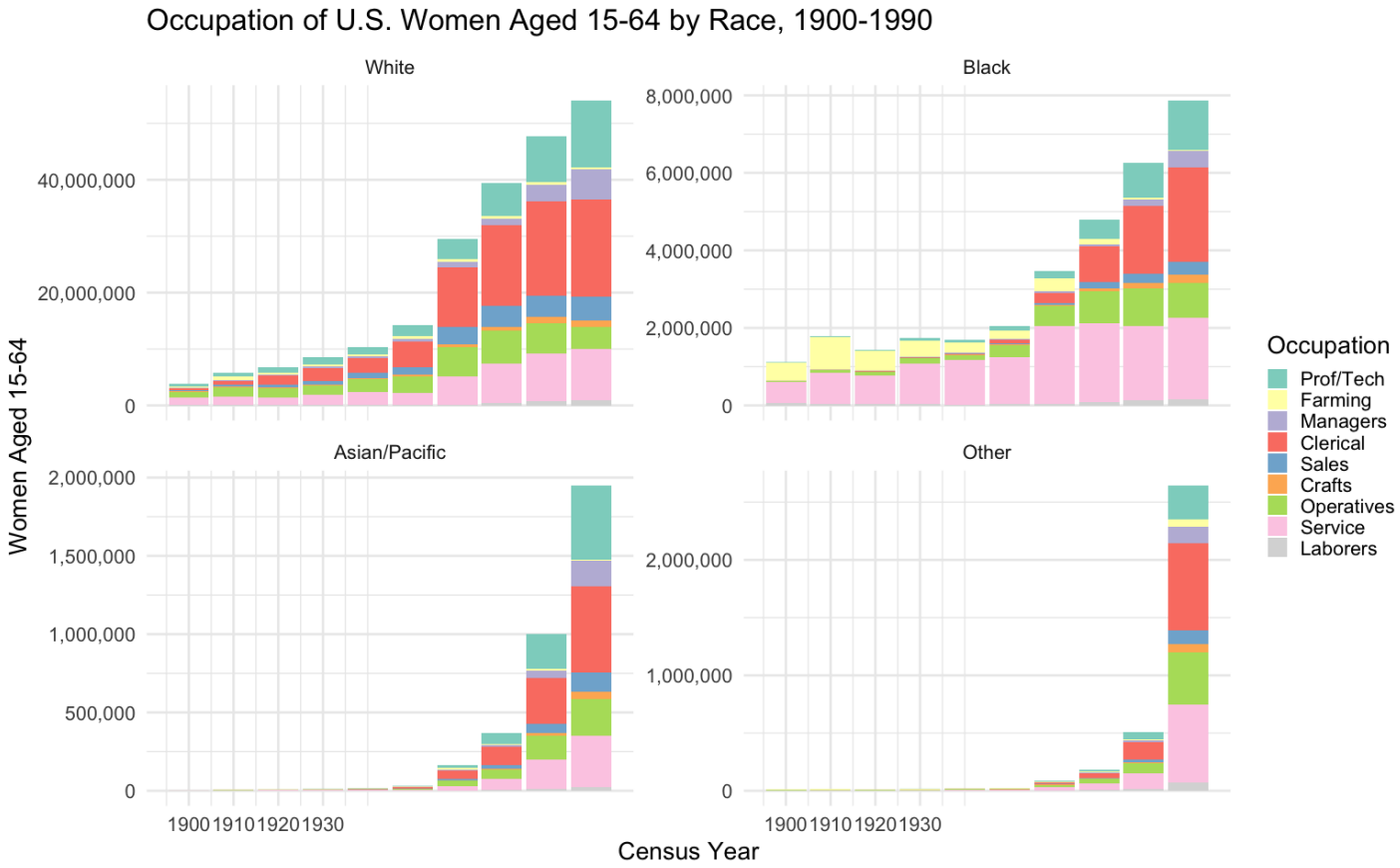
YEAR <int>	STATEFIP <int>	PERWT <dbl>	RELATE <int>	RELATED <int>	SEX <int>	AGE <int>	MARST <int>	RACE <int>	RACED <int>
1900	1	100	2	201	2	34	1	1	100
1900	1	100	5	501	2	55	5	1	100
1900	1	100	6	601	2	60	5	1	100
1900	1	100	2	201	2	22	1	2	200
1900	1	100	2	201	2	34	1	2	200
1900	1	100	5	501	2	60	5	2	200

1-10 of 5,035,642 rows | 1-10 of 13 columns Previous **1** 2 3 4 5 6 … 100 Next

Given how many women were not listed as having an occupation in most of the censuses, it can be hard to see what the women with occupations were doing. Make the same graph again, but this time remove women who were not listed as having an occupation. Do this without changing the `women` data frame.

Hide

```
women %>% filter(!OCCUP %in% c("None")) %>%
  group_by(YEAR, RACEF, OCCUP) %>% count(wt = PERWT) %>%
  ggplot(aes(YEAR, n, fill = OCCUP)) + geom_col() +
  scale_x_continuous(breaks = c(1870, 1880, 1900, 1910, 1920, 1930)) +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal(base_size = 20) + scale_fill_brewer(palette = "Set3") +
  facet_wrap(vars(RACEF), scales = "free_y") +
  labs(x = "Census Year", y = "Women Aged 15-64", fill = "Occupation",
       title = "Occupation of U.S. Women Aged 15-64 by Race, 1900-1990")
```



Hide

```
women
```

YEAR <int>	STATEFIP <int>	PERWT <dbl>	RELATE <int>	RELATED <int>	SEX <int>	AGE <int>	MARST <int>	RACE <int>	RACED <int>
1900	1	101	2	201	2	41	1	1	100
1900	1	101	2	201	2	27	1	1	100
1900	1	101	12	1215	2	17	6	2	200
1900	1	101	12	1214	2	21	2	2	200
1900	1	100	2	201	2	34	1	1	100
1900	1	100	5	501	2	55	5	1	100
1900	1	100	6	601	2	60	5	1	100
1900	1	100	2	201	2	22	1	2	200
1900	1	100	2	201	2	34	1	2	200
1900	1	100	5	501	2	60	5	2	200

1-10 of 5,035,642 rows | 1-10 of 13 columns Previous **1** 2 3 4 5 6 … 100 Next

Women's labor force participation

If we eliminate women without an occupational classification from our graph, then we need another graph to indicate what proportion of working-aged women are included in this graph so that readers can also get a sense of the number of women who did not have an occupation listed and how that changed over time. We can do this with a line graph.

What we want is a line graph with `YEAR` on the x-axis and the proportion of women who were listed as having an occupation on the y-axis, with race indicated by line color.

First create a new variable in `women` called `JOB` indicating (true or false) whether each woman had an occupation listed. Test your new variable.

Hide

```
women <- women %>% mutate(JOB = OCCUP != "None")
summary(women$JOB)
```

```
   Mode   FALSE    TRUE
logical 2432900 2602742
```

Now create a new data frame called `working` that includes one row for each race category in each year year, with a column indicating the number of employed women (this will be `n`) and the total number of women (`TOTAL`) in each race category in each year. The code below already includes all of the functions you will need, so you just need to fill in the parentheses.

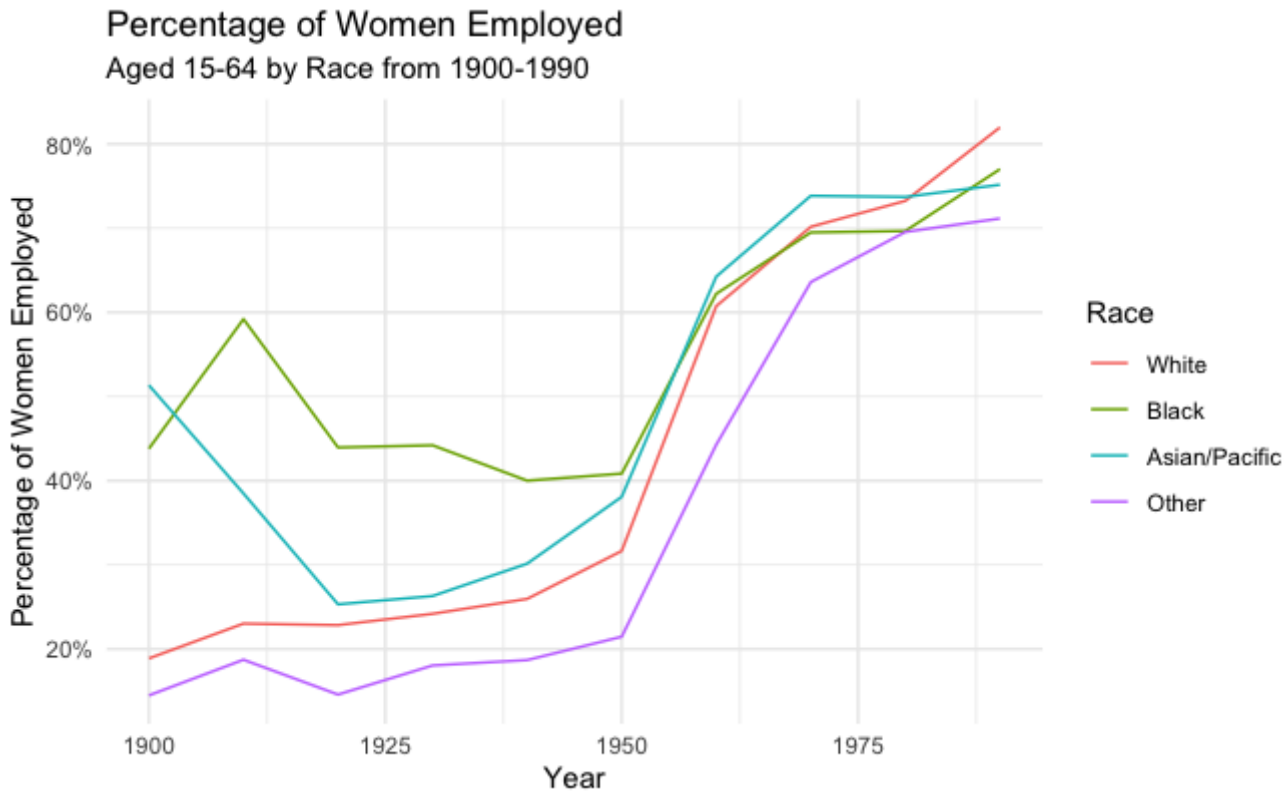
Hide

```
working <- women %>% group_by(RACEF, YEAR) %>%
                     count(JOB, wt = PERWT) %>%
                     mutate(TOTAL = sum(n)) %>% filter(JOB == TRUE)
```

Now make a line graph showing the percentage of women in each race category who were listed as being employed in each year. The year will go on the x-axis, percentage of women will go on the y-axis, and line color will indicate race classification. Make sure the graph is publication quality.

Hide

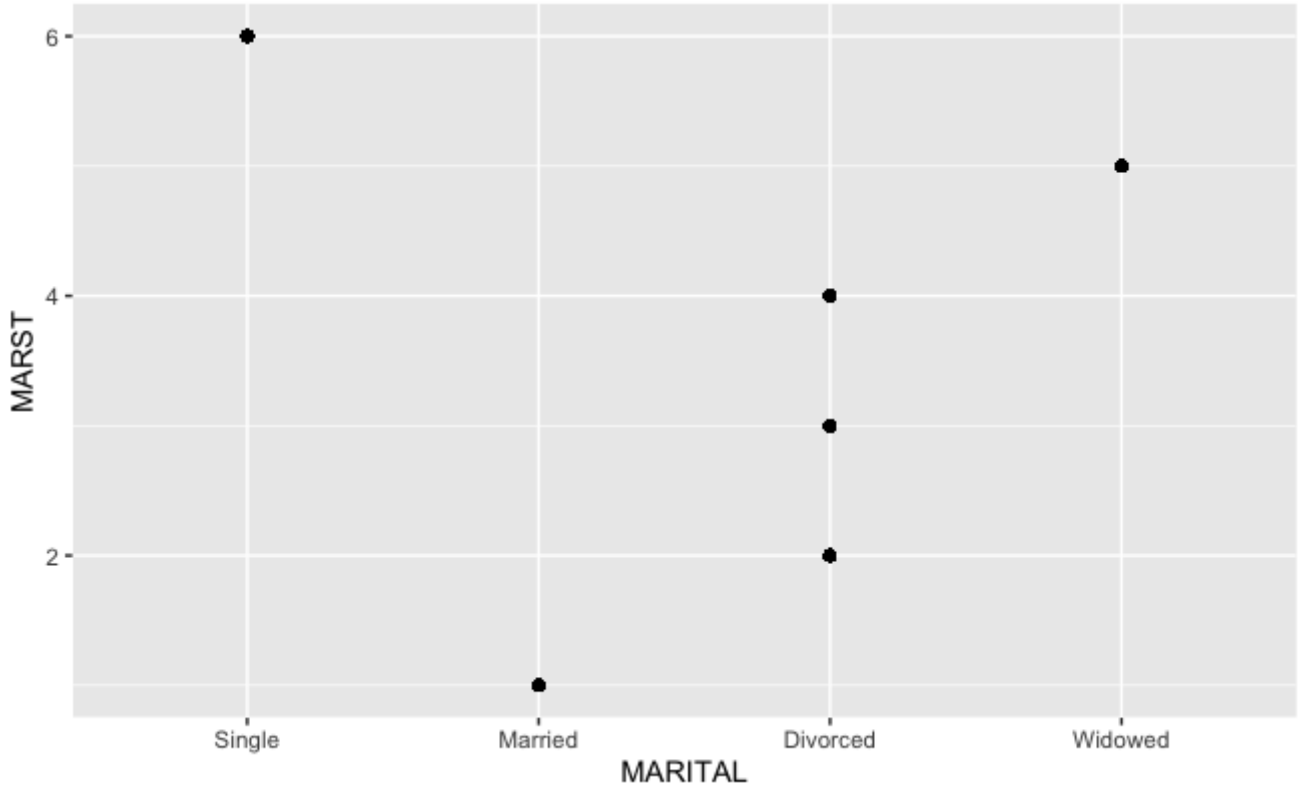
```
ggplot(working, aes(YEAR, n/TOTAL, color = RACEF)) + geom_line() +
  scale_y_continuous(labels = scales::percent) +
  theme_minimal() + scale_fill_brewer(palette = "Paired") +
  labs(x = "Year", y = "Percentage of Women Employed", color = "Race",
       title = "Percentage of Women Employed", subtitle = "Aged 15-64 by Race from 1900-
1990")
```



Going back to the `women` data frame, add a column for marital status (`MARITAL`), using the same categories we used in Notebook 5.

Hide

```
women <- women %>% mutate(MARITAL = ifelse(MARST == 6, 1,
                                   ifelse(MARST == 1, 2,
                                   ifelse(MARST == 5, 4,3)))) %>%
                   mutate(MARITAL = factor(MARITAL, labels = c("Single", "Married", "Div
orced", "Widowed")))
#Test
qplot(MARITAL, MARST, data = women)
```

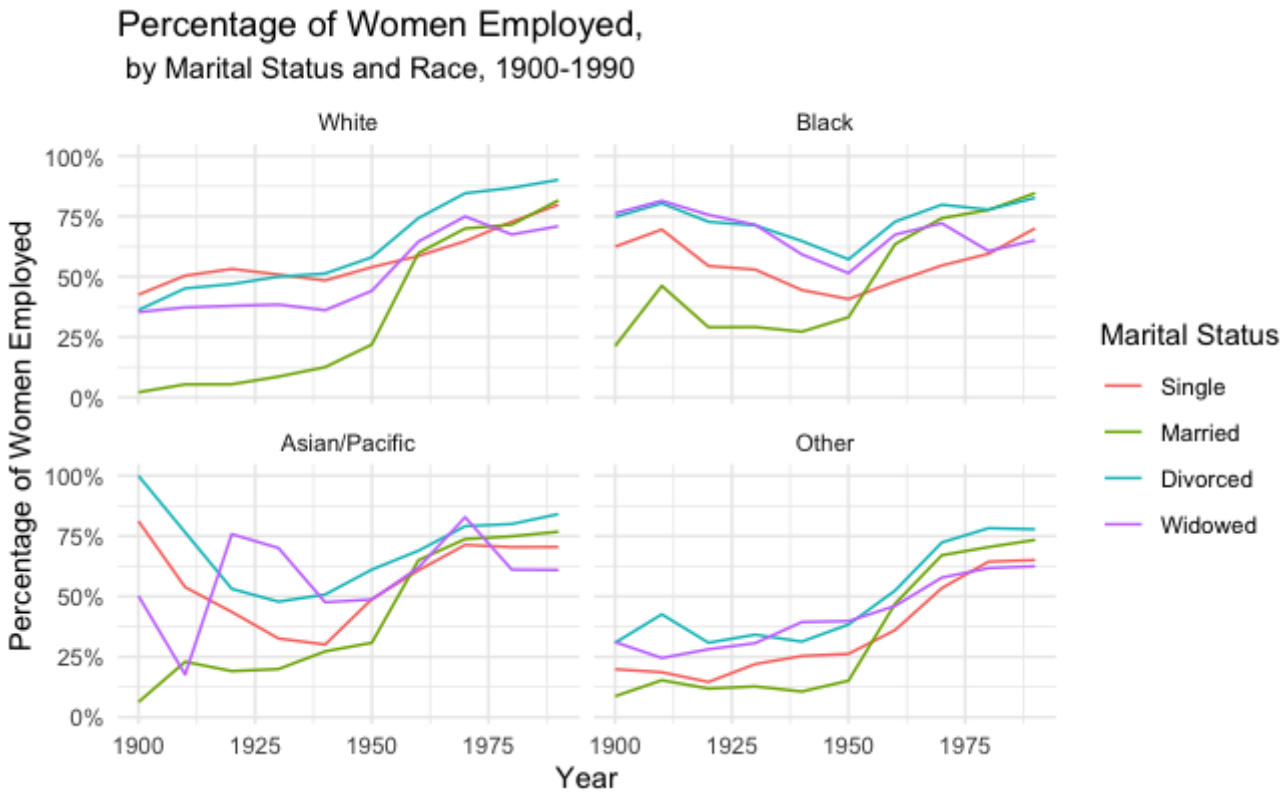


Now go through all of the necessary steps to make a line graph showing the percentage of women in each year who were listed as having an occupation, just as you did before, but this time use color to indicate marital status and facet by race.

Hide

```
working <- women %>% group_by(MARITAL, RACEF, YEAR) %>%
                     count(JOB, wt = PERWT) %>%
                     mutate(TOTAL = sum(n)) %>% filter(JOB == TRUE)

ggplot(working, aes(YEAR, n/TOTAL, color = MARITAL)) + geom_line() +
  scale_y_continuous(labels = scales::percent) +
  facet_wrap(vars(RACEF)) +
  theme_minimal() + scale_fill_brewer(palette = "Paired") +
  labs(x = "Year", y = "Percentage of Women Employed", color = "Marital Status",
       title = "Percentage of Women Employed,", subtitle = " by Marital Status and Race,
1900-1990")
```



Women's households

Now make a graph identical to the previous one, but instead of showing the percent of women aged 15-64 who are listed as having an occupation, show the percent of women aged 15-64 who are listed as household head. Remember that you will need to start by making a variable that indicates whether or not each woman was a head of household.

Hide

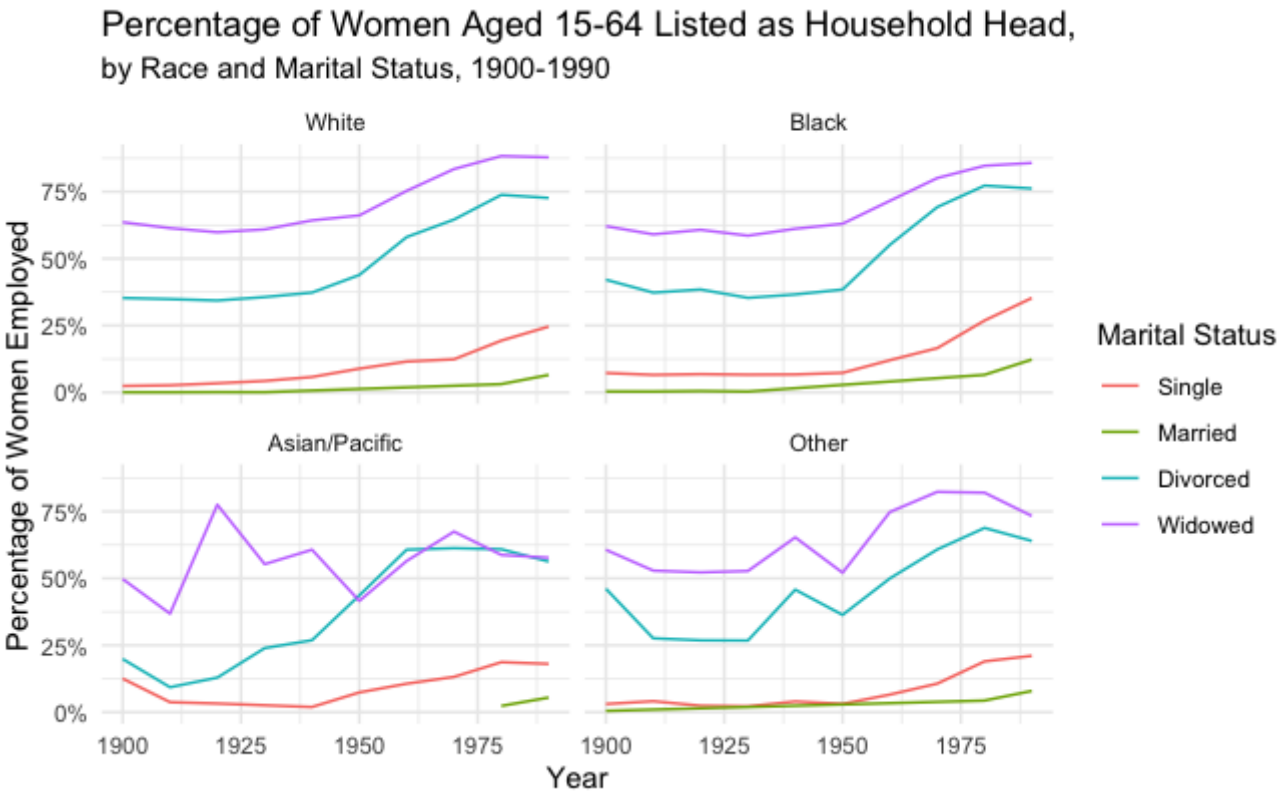
```
women <- women %>% mutate(HOUSE = RELATE == 1)
summary(women$HOUSE)
```

```
   Mode   FALSE    TRUE
logical 4394680  640962
```

Hide

```
women %>% group_by(MARITAL,RACEF, YEAR) %>%
          count(HOUSE, wt = PERWT) %>%
          mutate(TOTAL = sum(n)) %>% filter(HOUSE == TRUE) %>%

ggplot(aes(YEAR, n/TOTAL, color = MARITAL)) + geom_line() +
  scale_y_continuous(labels = scales::percent) +
  facet_wrap(vars(RACEF)) +
  theme_minimal() + scale_fill_brewer(palette = "Paired") +
  labs(x = "Year", y = "Percentage of Women Employed", color = "Marital Status",
       title = "Percentage of Women Aged 15-64 Listed as Household Head,", subtitle = "b
y Race and Marital Status, 1900-1990")
```



Submit the lab

Save this file, preview it to make sure everything is showing up (if it isn't, try running all the code again), and upload the .nb.html file to Canvas.