

Homework # 4 -

Riyesh Nath

2025-09-25

GROUP MEMBERS: Riyesh Nath

Summary:

We will continue data exploration to see if Race has an effect on marriage rate. We will divide the ratio of marriage with race based on each region. We will divide marriage type to two category, married or not married. We filter out all the NA since that is a none answer.

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2    3.5.1      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(viridis)
```

```
## Loading required package: viridisLite
```

```
library(modelsummary)
```

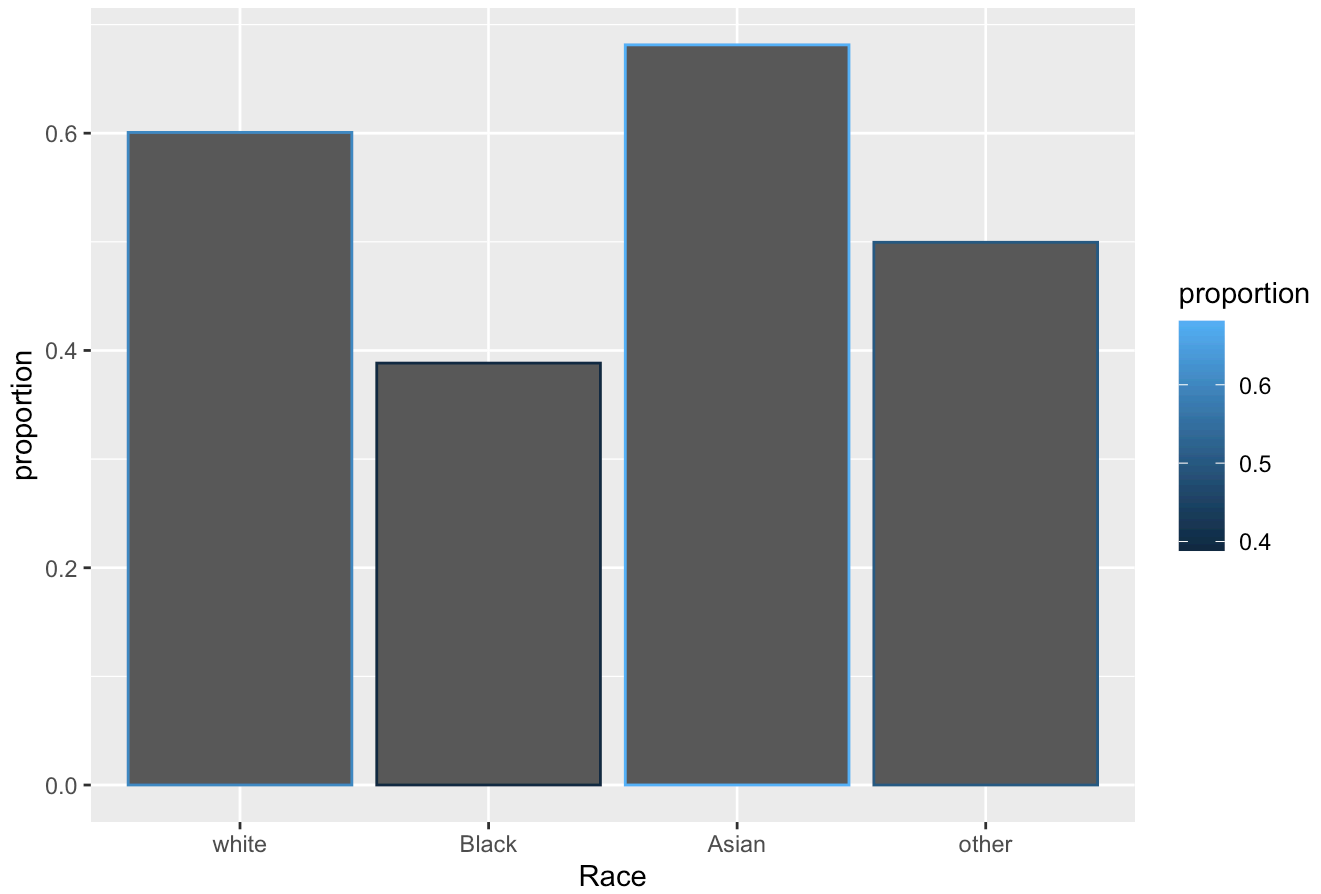
```
setwd("~/Desktop/GIT/Econometrics/Homework#3")
load("data/d_HHP2020_24.Rdata")
```

#FUNCTIONS

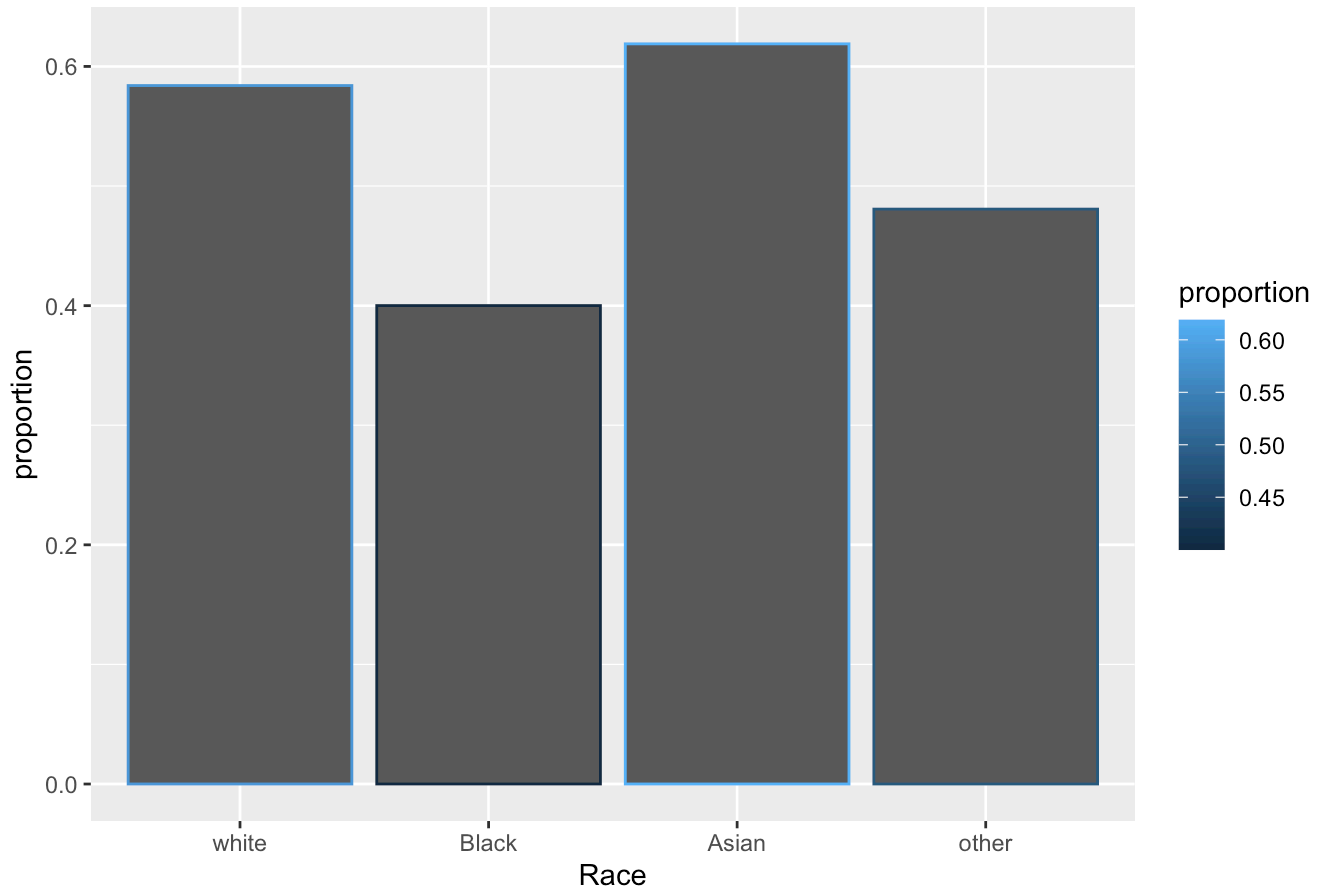
```
get_ratio_by_race_per_region <- function(dataframe, region) {  
  dataframe_race_proportion_region <- dataframe_marriage_rate_based_on_race %>%  
    filter(Region == region) %>%  
    group_by(Race) %>%  
    mutate(proportion = n/sum(n)) %>%  
    arrange(Race) %>%  
    ungroup()  
  
  dataframe_race_proportion_region <- dataframe_race_proportion_region %>%  
    filter(married)  
  
  plot <- ggplot(dataframe_race_proportion_region, aes(x=Race, y=proportion, color=proportion)) +  
    geom_col() +  
    labs(title = paste("Proportion of married population in", region))  
  
  print(plot)  
}
```

```
dataframe_marriage_rate_based_on_race <- d_HHP2020_24 %>%  
  filter(!is.na(Mar_Stat)) %>%  
  mutate(married = if_else(Mar_Stat == "Married", TRUE, FALSE)) %>%  
  count(Race, married, Region)  
  
list_of_region <- unique(dataframe_marriage_rate_based_on_race$Region)  
  
for(region in list_of_region){  
  get_ratio_by_race_per_region(dataframe_marriage_rate_based_on_race, region)  
}
```

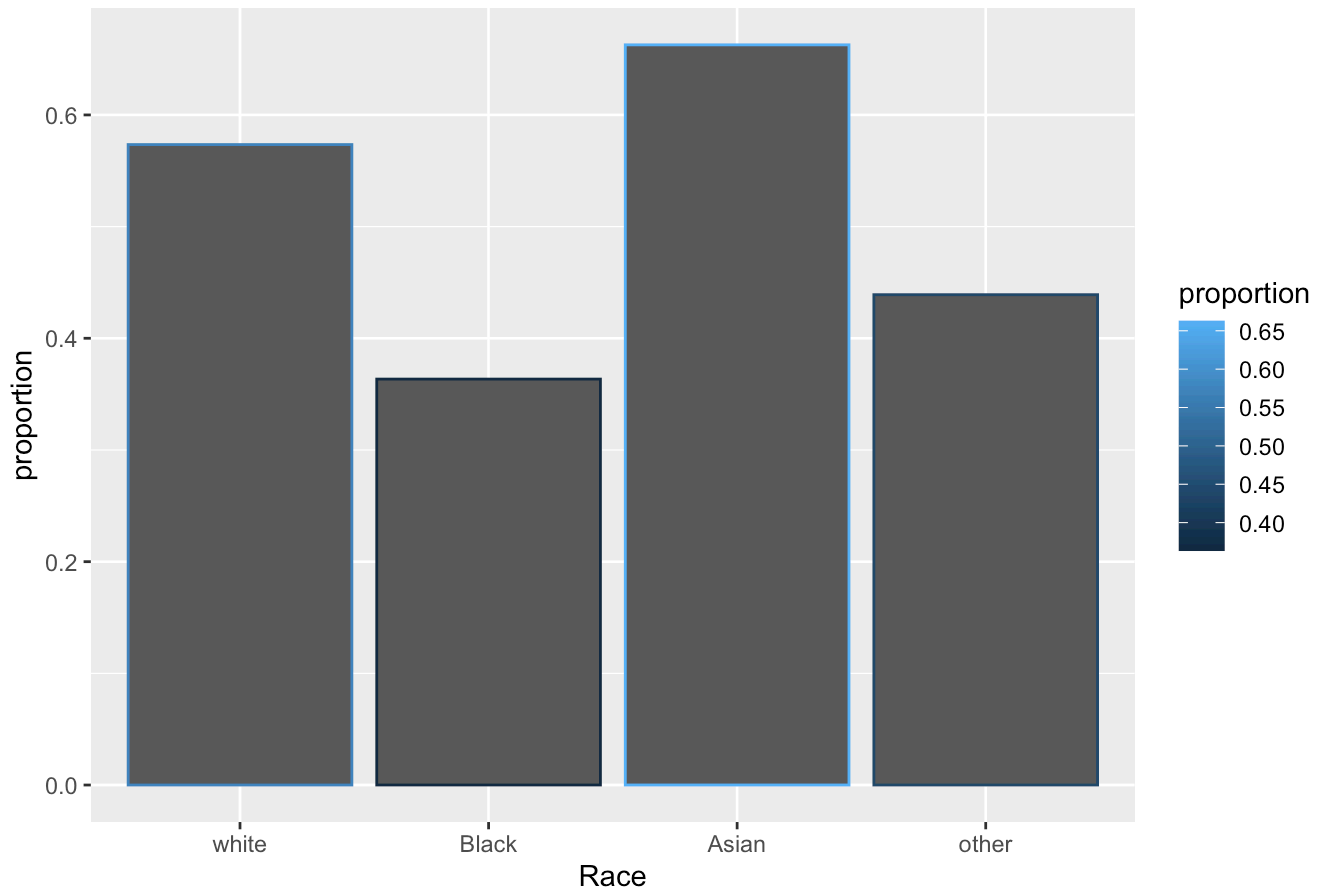
Porportion of married population in South



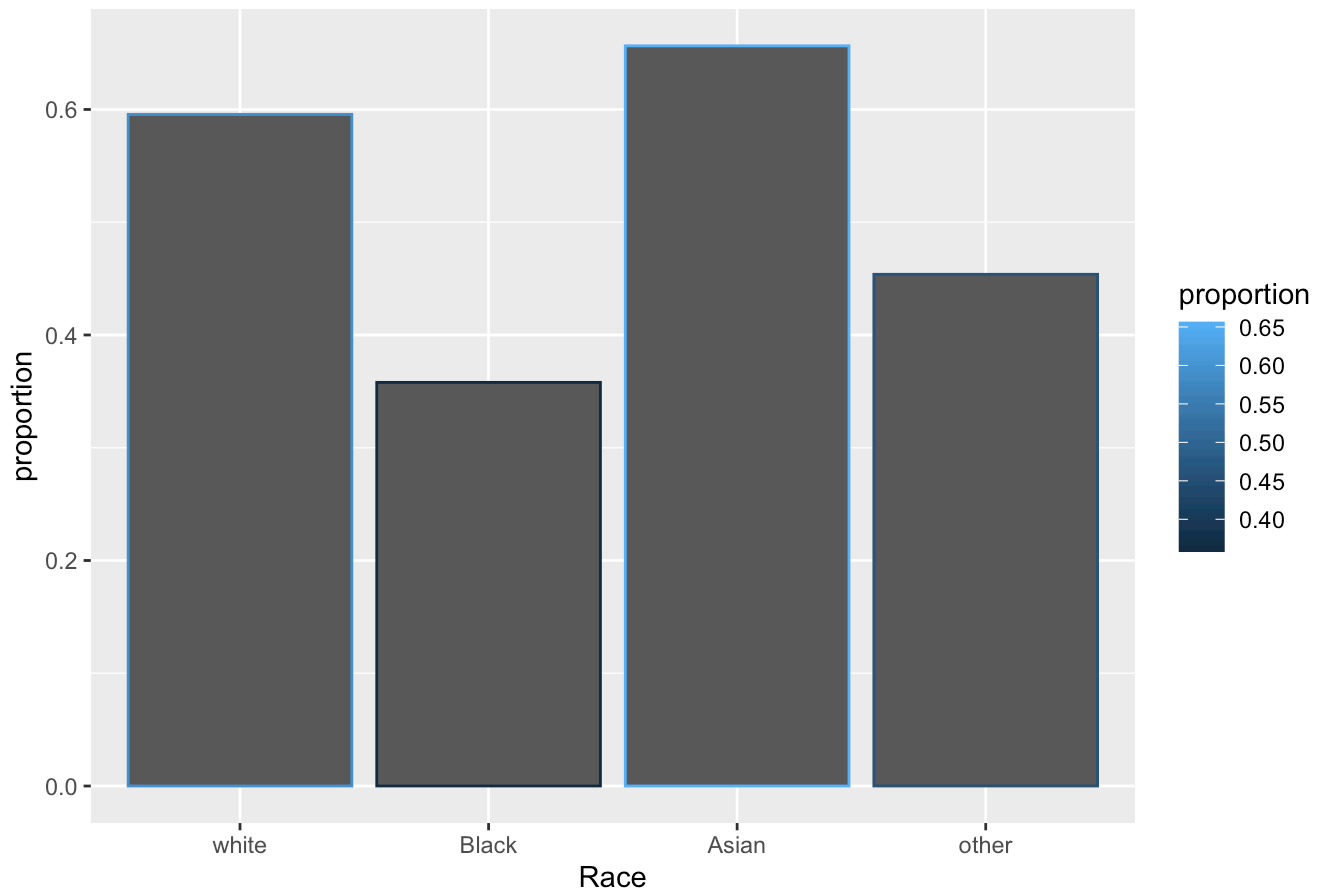
Porportion of married population in West



Porportion of married population in Northeast



Porportion of married population in Midwest



From a quick look, I see that despite the region, we have have the same marriage rate proportion for different race.