HW0 - Xinan Wang

Xinan Wang

2023-01-26

```
getwd()
```

```
## [1] "/Users/nannmemeda/Desktop/IE 6600"
```

Section A

```
# Reference: https://data.census.gov/table?q=United+States

# Import the dataset into R studio

library(readxl)
library(dplyr)
library(ggplot2)
library(tidyverse)
library(ggrepel)
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
race <- read_excel("DECENNIALPL2020.P1-2023-01-26T201630.xlsx")
head(race)
```

```
## # A tibble: 6 x 2
##   Label                                  Population
##   <chr>                                       <dbl>
## 1 Total:                                  331449281
## 2 Population of one race                  297600338
## 3 White alone                             204277273
## 4 Black or African American alone          41104200
## 5 American Indian and Alaska Native alone   3727135
## 6 Asian alone                              19886049
```

Plot 1: The Percentage Of Different Races Count Population

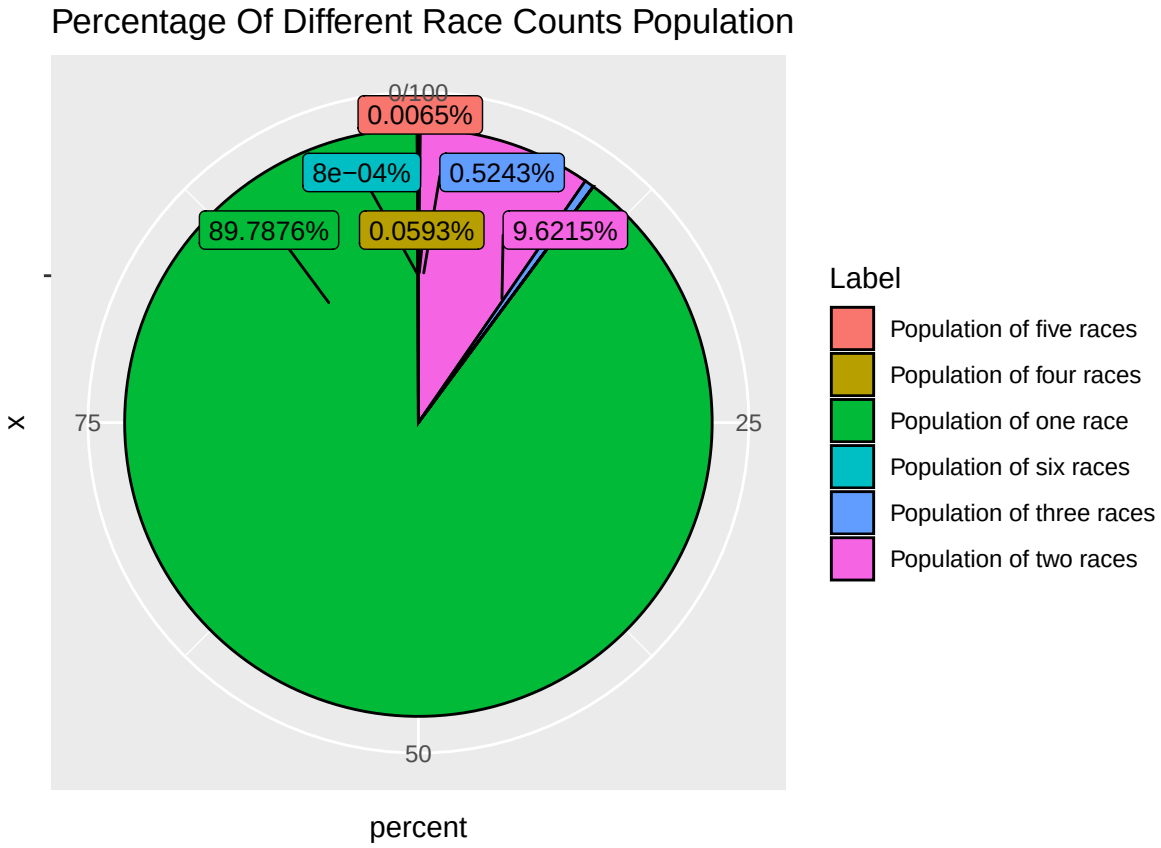
```
race_num <- race %>% filter(
  Label %in% c("Population of one race", "Population of two races", "Population of three races", "Popula
)

total <- race %>% filter(Label == "Total:")
total$Population
```

```
## [1] 331449281
```

```
race_num <- race_num %>% mutate(percent = round(100 * Population / total$Population, digit = 4))
race_num
```

```
## # A tibble: 6 x 3
##   Label                      Population percent
##   <chr>                           <dbl>   <dbl>
## 1 Population of one race     297600338 89.8
## 2 Population of two races     31890339  9.62
## 3 Population of three races    1737832  0.524
## 4 Population of four races      196582  0.0593
## 5 Population of five races       21685  0.0065
## 6 Population of six races         2505  0.0008
```

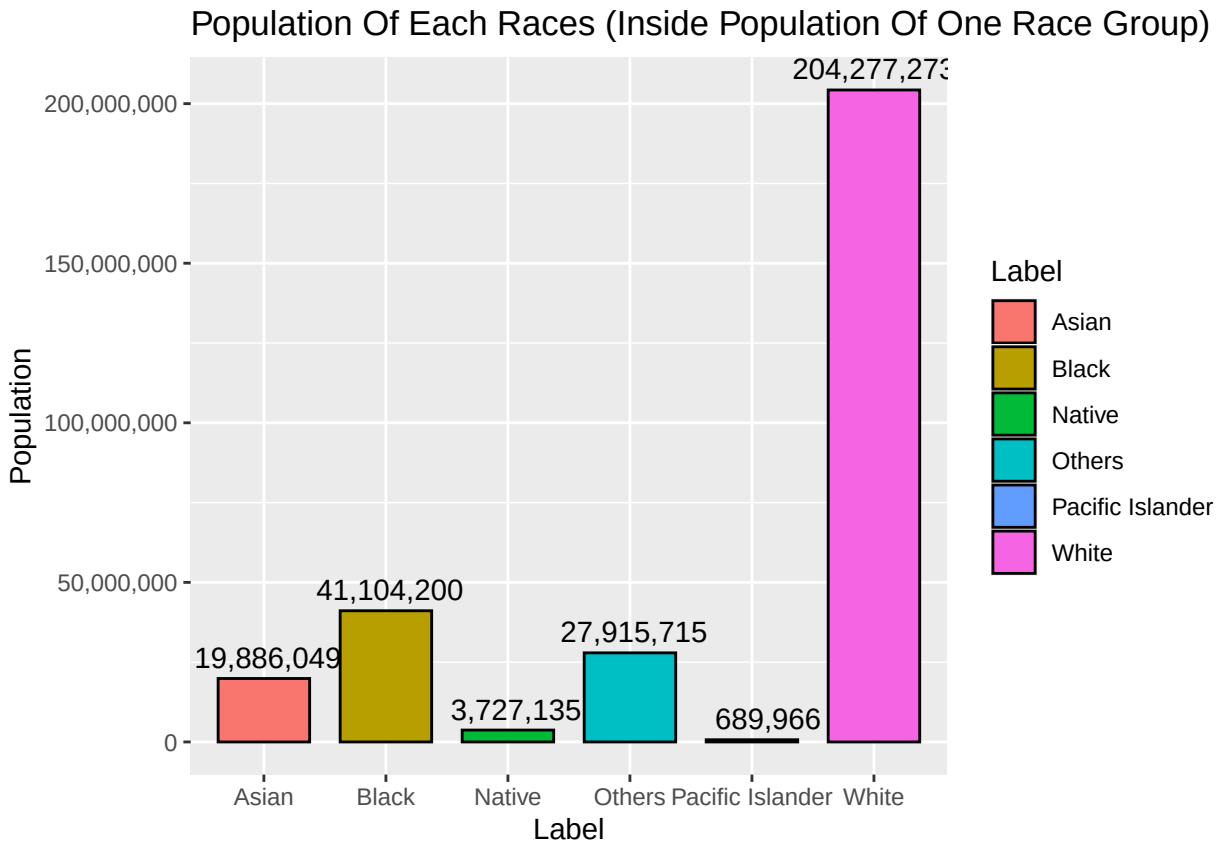
```
ggplot(race_num, aes(x = "", y = percent, fill = Label)) +
  geom_bar(stat="identity",width = 3, color = 1) +
  coord_polar("y", start = 0) +
  geom_label_repel(data = race_num,
                   aes(label = paste0(percent, "%")), size = 3.5, nudge_x = 1, show.legend = FALSE) +
  ggtitle("Percentage Of Different Race Counts Population")
```



Plot 2: United States Race Population (Inside Population Of One Race)

```
race_alone <- race %>% filter(Label %in% c("White alone", "Black or African American alone", "American I

race_alone$Label[race_alone$Label == "White alone"] <- "White"
race_alone$Label[race_alone$Label == "Black or African American alone"] <- "Black"
race_alone$Label[race_alone$Label == "American Indian and Alaska Native alone"] <- "Native"
race_alone$Label[race_alone$Label == "Asian alone"] <- "Asian"
race_alone$Label[race_alone$Label == "Native Hawaiian and Other Pacific Islander alone"] <- "Pacific Isl
race_alone$Label[race_alone$Label == "Some Other Race alone"] <- "Others"




ggplot(race_alone, aes(x = Label, y = Population, fill = Label)) +
  geom_line() +
  scale_color_manual(values = c("pink","brown", "red", "yellow", "green", "cyan")) +
  geom_bar(stat = "identity", width = 0.75, color = "black") +
  scale_y_continuous(labels = scales::comma) +
  geom_text(aes(label = format(Population, big.mark = ",")), position = position_dodge(width = 0.4), vju
  ggtitle("Population Of Each Races (Inside Population Of One Race Group)")
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```


Population Of Each Races (Inside Population Of One Race Group)
Population
200,000,000
150,000,000
100,000,000
50,000,000
0
19,886,049
41,104,200
3,727,135
27,915,715
689,966
204,277,273
Asian
Black
Native
Others
Pacific Islander
White
Label
Label
Asian
Black
Native
Others
Pacific Islander
White