

Final Project - Visualization

Hana Yerin Lim

3/10/2021

R Markdown

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggribes)
```

```
## Warning: package 'ggribes' was built under R version 4.0.4
```

```
library(viridis)
```

```
## Warning: package 'viridis' was built under R version 4.0.4
```

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

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 4.0.4
```

```
library(tidyr)
library(RColorBrewer)
library(wesanderson)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.4
```

```
data <- readr::read_csv("ca_div.csv")
```

```
##
## -- Column specification -----
## cols(
##   .default = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
colnames(data) <- tolower(colnames(data))
datacopy <- data
datacopy <- datacopy[, -c(1, 2, 12,14,18)]
```

```
cols <- c("sex", "marst", "race", "citizen", "speakeng", "educ", "nchild")
datacopy[cols] <- lapply(datacopy[cols], as.factor)
```

```
levels(datacopy$race) <- c("White", "African\n American", "American\n Indian", "Asian/Pacific\n Islander", "Hispanic", "Other")
levels(datacopy$educ) <- c("NA", "Elementary", "Middle", "High School", "High School", "High School", "High School", "College", "College", "College", "College", "Graduate")
levels(datacopy$sex) <- c("male", "female")
levels(datacopy$marst) <- c("Married", "Married", "Separated", "Divorced", "Widowed", "Single")
levels(datacopy$speakeng) <- c("NA", "No", "Well", "Well", "Well", "Well", "No")
levels(datacopy$nchild) <- c("0", "1", "2", "3+", "3+", "3+", "3+", "3+", "3+", "3+")
```

```
# datacopy$race[which(datacopy$race == "White" & datacopy$hispan != 0)] <- "Hispanic"
# datacopy$race[which(datacopy$race == "Other" & datacopy$hispan != 0)] <- "Hispanic"
datacopy$race[which(datacopy$hispan != 0)] <- "Hispanic"
```

```
## Warning in `[<-factor`(`*tmp*`, which(datacopy$hispan != 0), value =
## structure(c(2L, : invalid factor level, NA generated
```

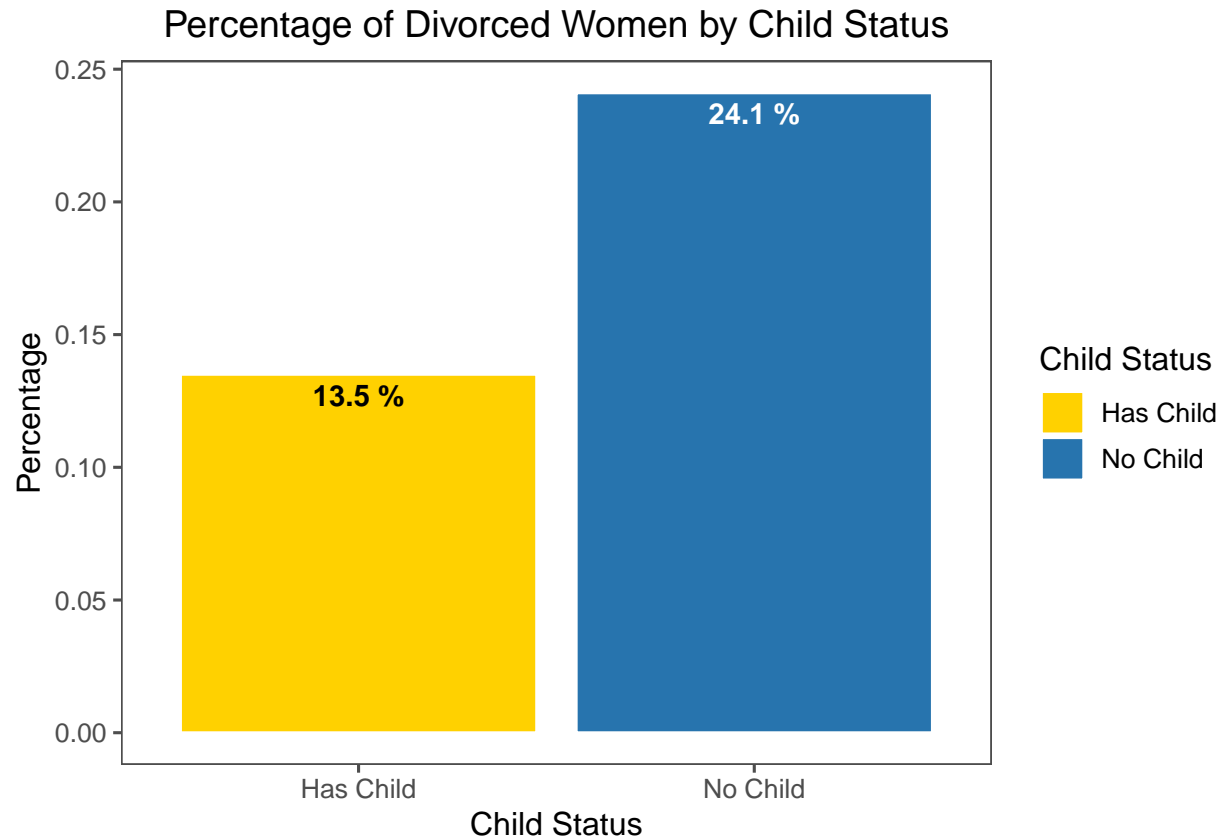
```
datacopy$race <- addNA(datacopy$race)
levels(datacopy$race) <- c(levels(datacopy$race), "Hispanic")
datacopy$race[is.na(datacopy$race)] <- "Hispanic"
```

```
datacopy <- datacopy %>% mutate(child_binary = as.factor(ifelse(nchild == 0, "No Child", "Has Child")))
datacopy$age_interval <- cut(datacopy$age, c(0, 19, 29, 39, 49, 59, 69, 79, 89, 99), labels = c("10s", "20s", "30s", "40s", "50s", "60s", "70s", "80s", "90s"))
```

```
female <- datacopy %>% filter(sex == "female")
```

```
female %>% filter(marst == "Married" | marst == "Divorced") %>%
  group_by(child_binary, marst) %>% summarise(ct = n()) %>% mutate(prop = ct / sum(ct)) %>% filter(marst == "Divorced")
ggplot(aes(x = child_binary, y = prop, fill = child_binary)) + geom_bar(stat = "identity", color = "white", fill = "#FFD100") +
  scale_fill_manual("Child Status", values = c("#FFD100", "#2774AE")) +
  geom_text(aes(label = paste(round(prop, 3)*100, "%"), vjust = 1.5, color = c("Black", "White"), fontface = "bold")) +
  ggtitle("Percentage of Divorced Women by Child Status") +
  theme_few() +
  theme(plot.title = element_text(hjust = 0.5)) +
  ylab("Percentage") + xlab("Child Status")
```

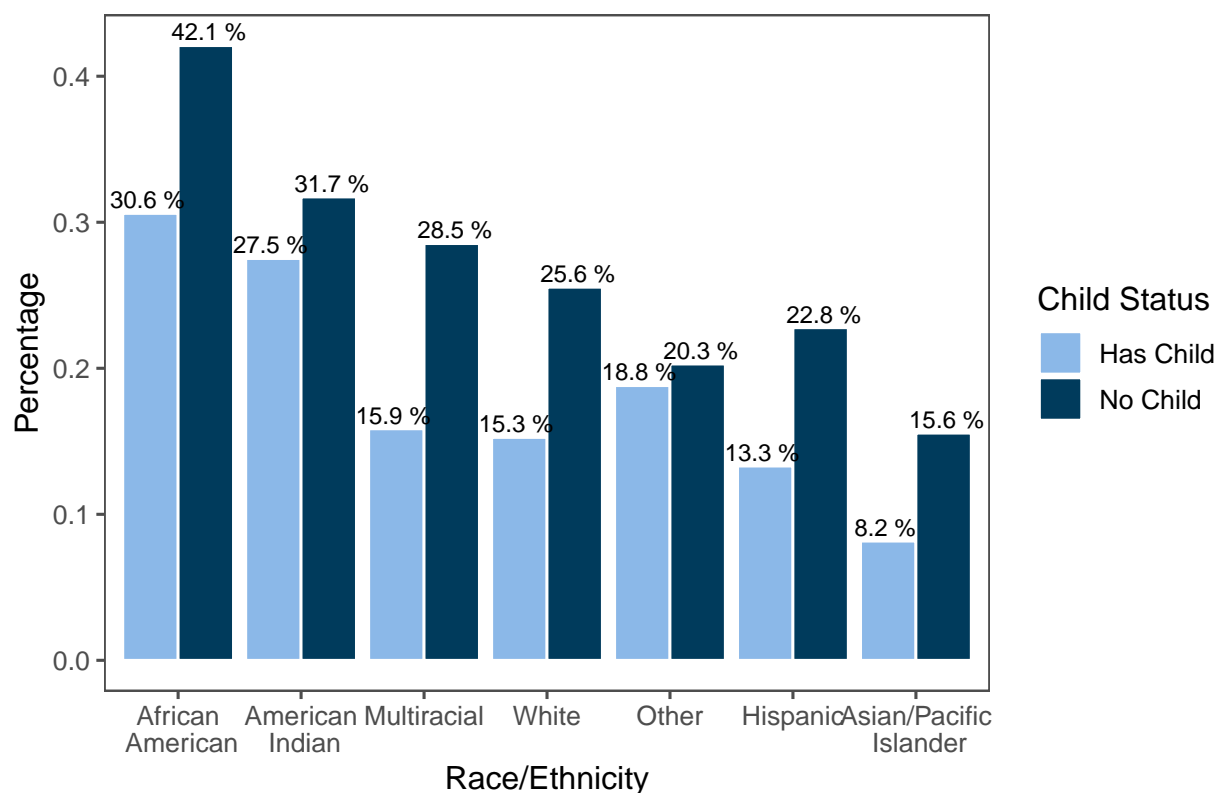
```
## `summarise()` regrouping output by 'child_binary' (override with `.groups` argument)
```



```
female %>% filter(marst == "Married" | marst == "Divorced") %>%
  group_by(child_binary, race, marst) %>% summarise(ct = n()) %>% mutate(prop = ct / sum(ct)) %>% filter(marst == "Divorced")
ggplot(aes(x = reorder(race, -prop), y = prop, fill = child_binary)) + geom_bar(stat = "identity", color = "black") +
  geom_text(aes(y = prop, label = paste(round(prop, 3)*100, "%"), size=3.3, position = position_dodge(0.5))) +
  ggtitle("Percentage of Divorced Women by Demographic and Child Status") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_fill_manual("Child Status", values = c("#8BB8E8", "#003B5C")) + ylab("Percentage") + xlab("Race")
```

```
## `summarise()` regrouping output by 'child_binary', 'race' (override with `.groups` argument)
```

Percentage of Divorced Women by Demographic and Child Status

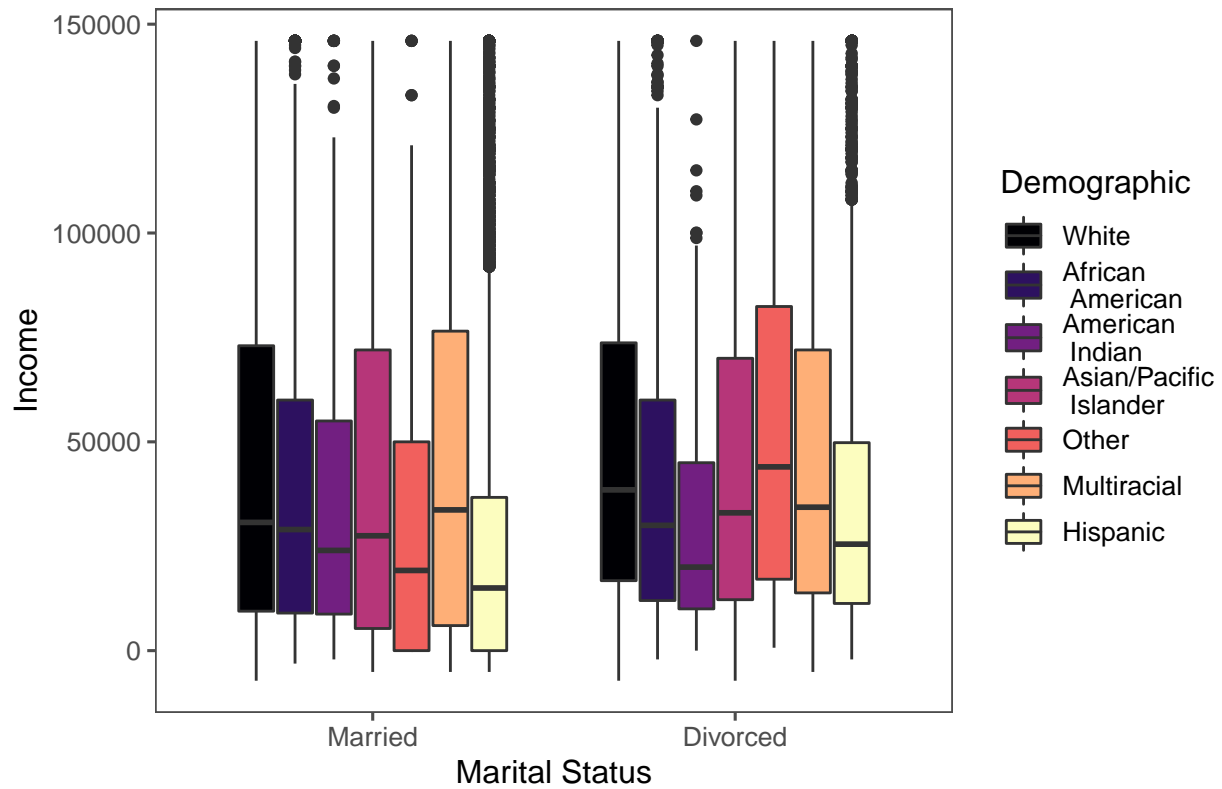


income (if needed)

```
filter_income <- female %>% filter(inctot != 9999999)
lowerq <- quantile(filter_income$inctot)[2]
upperq <- quantile(filter_income$inctot)[4]
iqr = upperq - lowerq
upper_outlier <- (iqr * 2) + upperq
lower_outlier <- lowerq - (iqr * 2)
filter_income$inctot[filter_income$inctot > upper_outlier] <- upper_outlier
filter_income$inctot[filter_income$inctot < lower_outlier] <- lower_outlier

# removed outliers
filter_income %>% filter(inctot != 9999999, marst == "Married" | marst == "Divorced") %>%
  ggplot(aes(x = marst, y = inctot, fill = race)) + geom_boxplot() +
  scale_fill_viridis("Demographic", option = "A", discrete = TRUE) +
  scale_color_viridis(option = "A", discrete = TRUE) +
  ggtitle("Income comparison between married vs divorced women by race") +
  theme_few() +
  theme(plot.title = element_text(hjust = 0.5)) +
  xlab("Marital Status") + ylab("Income")
```

Income comparison between married vs divorced women by race



```
filter_income %>% filter(inctot != 9999999, marst == "Married" | marst == "Divorced") %>%
  ggplot(aes(x = race, y = inctot, fill = marst)) + geom_boxplot() +
  ggtitle("Income comparison between married vs divorced women by race") +
  scale_fill_manual("Marital Status", values = c("#f7cb44ff", "#7e4e90ff")) +
  coord_flip() +
  theme_few() +
  theme(plot.title = element_text(hjust = 0.5)) +
  ylab("Income Total") + xlab("Race/Ethnicity")
```

Income comparison between married vs divorced women by race

