Paper Assignment #2

Clayton Perrin

2023-10-16

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(descr)  
library(forcats)  
library(ggplot2)

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

## create a table of gss\_cat  
table(gss\_cat$race)

##   
## Other Black White Not applicable   
## 1959 3129 16395 0

table(gss\_cat$rincome)

##   
## No answer Don't know Refused $25000 or more $20000 - 24999   
## 183 267 975 7363 1283   
## $15000 - 19999 $10000 - 14999 $8000 to 9999 $7000 to 7999 $6000 to 6999   
## 1048 1168 340 188 215   
## $5000 to 5999 $4000 to 4999 $3000 to 3999 $1000 to 2999 Lt $1000   
## 227 226 276 395 286   
## Not applicable   
## 7043

## Including Plots

You can also embed plots, for example:

## create data frame & remove missing values  
df1 <- gss\_cat %>%   
 na.omit() %>%   
 filter(year == 2010) %>%   
 select(year, race, rincome)

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

## making variable dichotomous, eliminating categories from data  
df1 %>%   
 filter(year == 2010) %>%   
 filter(race != "Other")

## # A tibble: 1,296 × 3  
## year race rincome   
## <int> <fct> <fct>   
## 1 2010 White $7000 to 7999   
## 2 2010 Black Not applicable  
## 3 2010 White Not applicable  
## 4 2010 Black $10000 - 14999  
## 5 2010 Black Not applicable  
## 6 2010 Black $20000 - 24999  
## 7 2010 Black $25000 or more  
## 8 2010 White Not applicable  
## 9 2010 White Not applicable  
## 10 2010 White Not applicable  
## # ℹ 1,286 more rows

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

df1 %>%   
 filter(year == 2010) %>%   
 filter(rincome != " Don't know"   
 & rincome != "Refused"   
 & rincome != "Not applicable" )

## # A tibble: 834 × 3  
## year race rincome   
## <int> <fct> <fct>   
## 1 2010 Other $25000 or more  
## 2 2010 White $7000 to 7999   
## 3 2010 Black $10000 - 14999  
## 4 2010 Black $20000 - 24999  
## 5 2010 Black $25000 or more  
## 6 2010 White $25000 or more  
## 7 2010 White $25000 or more  
## 8 2010 White $25000 or more  
## 9 2010 White $25000 or more  
## 10 2010 Other $1000 to 2999   
## # ℹ 824 more rows

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

## get a quick count of the sample size  
count(df1)

## # A tibble: 1 × 1  
## n  
## <int>  
## 1 1423

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

## get the relative frequency for the race variable  
df1 %>%   
 filter(year == 2010) %>%   
 filter(race != "Other") %>%   
 count(race) %>%   
 mutate(prop = prop.table(n))

## # A tibble: 2 × 3  
## race n prop  
## <fct> <int> <dbl>  
## 1 Black 226 0.174  
## 2 White 1070 0.826

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

## get relative frquency for the response variable (income)  
df1 %>%   
 filter(year == 2010) %>%   
 filter(rincome != "No answer"  
 & rincome != "Refused"  
 & rincome != "Not applicable"  
 & rincome != "Don't Know" ) %>%   
 mutate(rincome = fct\_recode(rincome,   
 "More than 20000" = "$25000 or more",  
 "More than 20000" = "$20000 - 24999",  
 "Less than 20000" = "$15000 - 19999",  
 "Less than 20000" = "$10000 - 14999",  
 "Less than 20000" = "$8000 to 9999",  
 "Less than 20000" = "$7000 to 7999",  
 "Less than 20000" = "$6000 to 6999",  
 "Less than 20000" = "$5000 to 5999",  
 "Less than 20000" = "$4000 to 4999",  
 "Less than 20000" = "$3000 to 3999",  
 "Less than 20000" = "$1000 to 2999",  
 "Less than 20000" = "Lt $1000")) %>%   
 count(rincome) %>%   
 mutate(prop = prop.table(n))

## # A tibble: 3 × 3  
## rincome n prop  
## <fct> <int> <dbl>  
## 1 Don't know 10 0.0120  
## 2 More than 20000 513 0.615   
## 3 Less than 20000 311 0.373

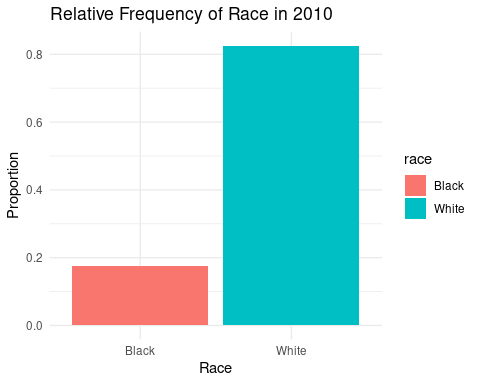
Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

# Filter and prepare the data for visualization  
data\_to\_plot <- df1 %>%  
 filter(year == 2010) %>%  
 filter(race != "Other" &  
 rincome != "No answer" &  
 rincome != "Refused" &  
 rincome != "Not applicable" &  
 rincome != "Don't Know")

## Including Plots

You can also embed plots, for example:

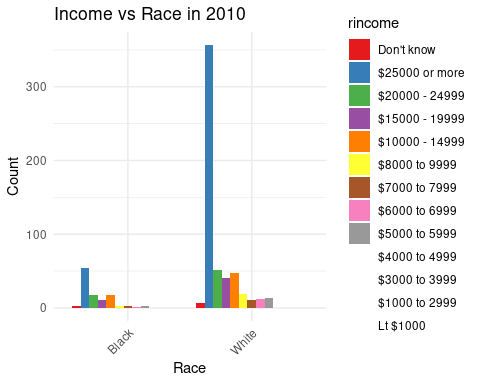
# Create a bar plot for race  
ggplot(df1 %>%   
 filter(year == 2010) %>%   
 filter(race != "Other") %>%   
 count(race) %>%   
 mutate(prop = prop.table(n)),   
 aes(x = race, y = prop, fill = race)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Relative Frequency of Race in 2010",  
 x = "Race",  
 y = "Proportion") +  
 theme\_minimal()



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

# Create a grouped bar plot  
ggplot(data\_to\_plot, aes(x = race, fill = rincome)) +  
 geom\_bar(position = "dodge") +  
 labs(title = "Income vs Race in 2010",  
 x = "Race",  
 y = "Count") +  
 scale\_fill\_brewer(palette = "Set1") + # Adjust the color palette  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels for better readability

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum for palette Set1 is 9  
## Returning the palette you asked for with that many colors



## Including Plots

You can also embed plots, for example:

# Create a bar plot for reported income  
ggplot(df1 %>%   
 filter(year == 2010) %>%   
 filter(rincome != "No answer" &  
 rincome != "Refused" &  
 rincome != "Not applicable" &  
 rincome != "Don't Know"),   
 aes(x = rincome, fill = rincome)) +  
 geom\_bar() +  
 labs(title = "Relative Frequency of Reported Income in 2010",  
 x = "Reported Income",  
 y = "Count") +  
 theme\_minimal()

