Homework 3

Reyna Orellana

4/12/2021

```
#Installing packages that I will need to use markdown
install.packages ("tidyverse")
install.packages ("tinytex")
install.packages ("dplyr")
install.packages("rmarkdown")
#loading tidyverse
library(tidyverse, tinytex, rmarkdown, "dplyr")
```

#data from GSS #My dataset has 2348 rows and 1065 columns (variables) #my unit of analysis is people. I am using the General Social Survey

```
> GSS2018 <- read_dta("GSS2018.dta")
> View(GSS2018)

#data descriptions/exploration
head(GSS2018)
dim(GSS2018)
colnames(GSS2018)
```

#making my dataset smaller, keeping only variables that will be used for my analysis

```
ab18<- select(GSS2018, "abany":"abstate1","age","ballot","bible","born", "class",
              "cohort", "coldeg1", "degree", "depress", "occ10", "partyid", "race",
              "sex", "hispanic")

summarize(ab18, mean.abany = mean (`abany`, na.rm = TRUE))

summary(gss18$abany)
table(gss18,"abany")
length(unique(GSS2018$abany))
table(GSS2018$abany)

#remove data that I will not use
rm(Remitdata,column_names,ab,gss18)
```

#For this assignment, you must select at least two variables of interest from your dataset, provide a #basic description of the variables, clean and recode as needed, and present descriptive statistics and #any visualizations. Your R Markdown document should include:

#1 The remaining NAs were missing data and I decided to leave them as NAs #2. My two variables will be age and race # I ended up with a smaller dataset of 35 variales with 2348 observations

#race

```
summarise(ab18, race)
table(select (ab18, race))
```

```
head(select(ab18, race))
#recode race variable
#changing first to make it a nominal variable since it is a numerical value
#Race variable had 3 values that I changed
#Age variable had 18 as the lowest number and 89 as its highest. I created 4
#categories for the new variable.
#two ways that work to recode race

#does not work, why? what is haven_labelled mean?
#ab18<- ab18 %>%
#  mutate(race3 = recode(race, `1` = "White", `2` = "Black", `3` = "Other"))


ab18<- ab18 %>%
  mutate(race_3cat = case_when(
    race == 1 ~ "White",
    race == 2 ~ "Black",
    race == 3 ~ "Other"
  ))

#checking
table(select(ab18, race, race_3cat ))

#age
summary(ab18$age)
#recoding age variable
ab18 <- ab18%>%
  mutate(age_4cat = case_when( age >= 18 & age <= 29 ~ "18-29",
                               age >= 30 & age <= 39 ~ "30-39",
                               age >= 40 & age <= 49 ~ "40-49",
                               age >= 50 & age <= 89 ~ "50+"))
table(select(ab18, age_4cat))
#checking if recoded correctly
table(select(ab18, age_4cat, age))
```