```
## Name: Robert James
## Assignment: Lab 3
## Date: 11/08/2023
## Purpose: Explore the Relationship between income at 16 and years spend in school
# install packages
install.packages("remotes", repos = "http://cran.us.r-project.org")
install.packages("tidyverse", repos = "http://cran.us.r-project.org")
install.packages("tidyr", repos = "http://cran.us.r-project.org")

# load gssr package
remotes::install_github("kjhealy/gssr")

# load libraries
library(gssr)
library(dplyr)
library(tidyr)

# load the master documentation files
data(gss_all) # note that this is a large file of all GSS data
data(gss_doc) # this is the documentation for the GSS data
# use the dictionary to get information in a different format
data(gss_dict)
gss_dict
df_2018 <- gss_all %>% #filter for only the year 2018
  filter(year == 2018)
df_2018

df <- df_2018 %>% #income at 16 and years in school, and wtssall
  select(educ, incom16, wtssall) %>%
  drop_na() #remove missing values

sapply(df, function(x) sum(is.na(x))) #count missing values
df

# run your model and diagnostic plots
model <- lm(df$incom16 ~ df$educ)
summary(model)

# Plot the regression line
plot(df$incom16, df$educ)
abline(model, col="blue")
cor(df$incom16, df$educ)
df
```

```
# Check residuals
resids2 <- residuals(model)
plot(df$incom16, resids2)

par(mfrow=c(2,2))
plot(model)
```