HW5_ahcooper

Andrew Cooper

11/2/2020

```
library(tidyverse)
library(readr)
library(janitor)
library(MASS)
```

Problem 3

My r-session crashed everytime I attempted to load the file "EdStatsData.csv" in its entirety. TO avoid this issue I only read in the first 1000 rows, which makes my compuations of the number of rows in the data before and after munging it innacurate.

```
EdStatsCountry <- read_csv("~/STAT5014/EdStatsCountry.csv")
EdStatsCountry_Series <- read.csv("~/STAT5014/EdStatsCountry-Series.csv", col.names = c("Country Code",
EdStatsData <- read_csv("~/STAT5014/EdStatsData.csv", n_max = 1000)
EdStatsFootNote <- read_csv("~/STAT5014/EdStatsFootNote.csv")
EdStatsSeries <- read_csv("~/STAT5014/EdStatsSeries.csv")
```

```
df1 <- left_join(EdStatsCountry, EdStatsCountry_Series, by = c("Country Code" = "Country.Code")) %>%
  left_join(EdStatsData, by = "Country Code") %>%
  left_join(EdStatsFootNote, by = c("Country Code" = "CountryCode")) %>%
  left_join(EdStatsSeries, by = c("SeriesCode" = "Series Code"))
```

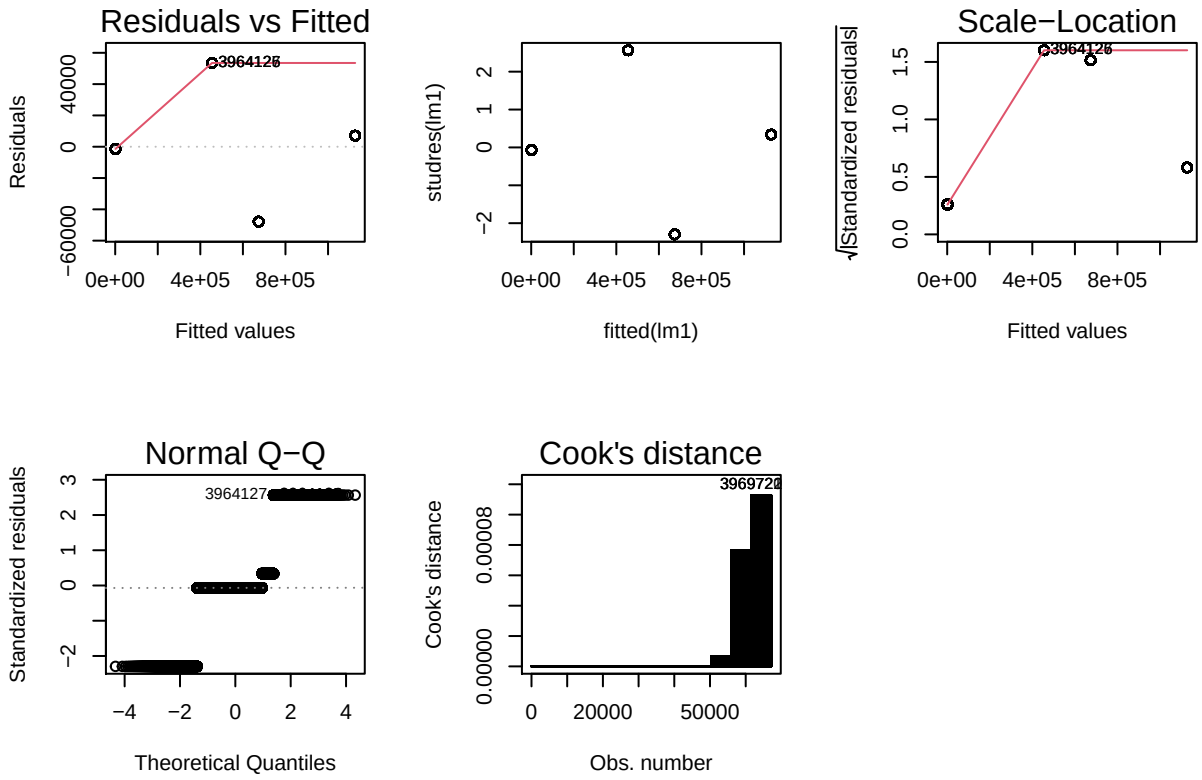
```
df1_clean <- df1 %>%
  janitor::clean_names()
```

There were 6.49157×10^5 rows in the original data. There are now 7.317784×10^6 rows in the cleaned data.

Problem 4

```
lm1 <- lm(x1971 ~ x1970, df1_clean)
```

```
par(mfrow = c(2, 3))
plot(lm1, which = 1)
plot(fitted(lm1), studres(lm1))
plot(lm1, which = 3)
plot(lm1, which = 2)
plot(lm1, which = 4)
```


Residuals vs Fitted
Residuals
Fitted values
studres(lm1)
fitted(lm1)
Scale−Location
√|Standardized residuals|
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Cook's distance
Cook's distance
Obs. number

Problem 5