Final Report

due November 16, 2021 by 11:59 PM

Your names and team name here

Load Packages

```
library(tidyverse)
library(ggplot2)
library(sf)
install.packages("rnaturalearth")
library(rnaturalearth)
install.packages("rnaturalearthdata")
library(rnaturalearthdata)
library(rgeos)
library(scales)
```

Load Data

```
library(readr)
h_spend <- read_csv("../data/Health-Spending.CSV")
wb_tot_lifeexp <- read_csv("../data/World Bank Life Exp.csv")
wb_f_lifeexp <- read_csv("../data/World Bank Female Life Exp.csv")
wb_m_lifeexp <- read_csv("../data/World Bank Male Life Exp.csv")
econ <- read_csv("../data/EconMetrics.csv")
econ <- select(econ, -4)
```

#Combine Datasets

```
wb_tot_lifeexp_long <- pivot_longer(wb_tot_lifeexp, cols = "1990":"2020", names_to = "Year", values_to =

wb_f_lifeexp_long <- wb_f_lifeexp %>%
  select(c(1:2,34:64)) %>%
  pivot_longer(cols = "1990":"2020", names_to = "Year", values_to = "Female Life Expectancy")

wb_m_lifeexp_long <- wb_m_lifeexp %>%
  select(c(1:2,34:64)) %>%
  pivot_longer(cols = "1990":"2020", names_to = "Year", values_to = "Male Life Expectancy")

lifeexp <- left_join(wb_tot_lifeexp_long, wb_f_lifeexp_long, by = c("Country Name", "Country Code", "Yea

lifeexp <- left_join(lifeexp, wb_m_lifeexp_long, by = c("Country Name", "Country Code", "Year"))
```

```
yrs <- c(2011:2020)

colnames(econ) <- c("Country Name", "Country Code", "Series Name", yrs)
```

```r
econ_temp <- econ %>%
  pivot_longer(cols = 4:13, names_to = "Year", values_to = "Value")
econ_temp2 <- econ_temp %>%
  group_by("Country Name") %>%
  pivot_wider(names_from = "Series Name", values_from = "Value") %>%
  ungroup()
```

```
## Warning: Values are not uniquely identified; output will contain list-cols.
## * Use `values_fn = list` to suppress this warning.
## * Use `values_fn = length` to identify where the duplicates arise
## * Use `values_fn = {summary_fun}` to summarise duplicates
```

```r
econ_temp3 <- select(econ_temp2, -4)
econ2 <- filter(econ_temp3, !is.na(econ_temp3$'Country Code'))
j <- c(4:58)
econ2[j] <- lapply(econ2[j], unlist)
econ2[j] <- lapply(econ2[j], as.numeric)
```

```
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
```

```
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
```

```
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
```

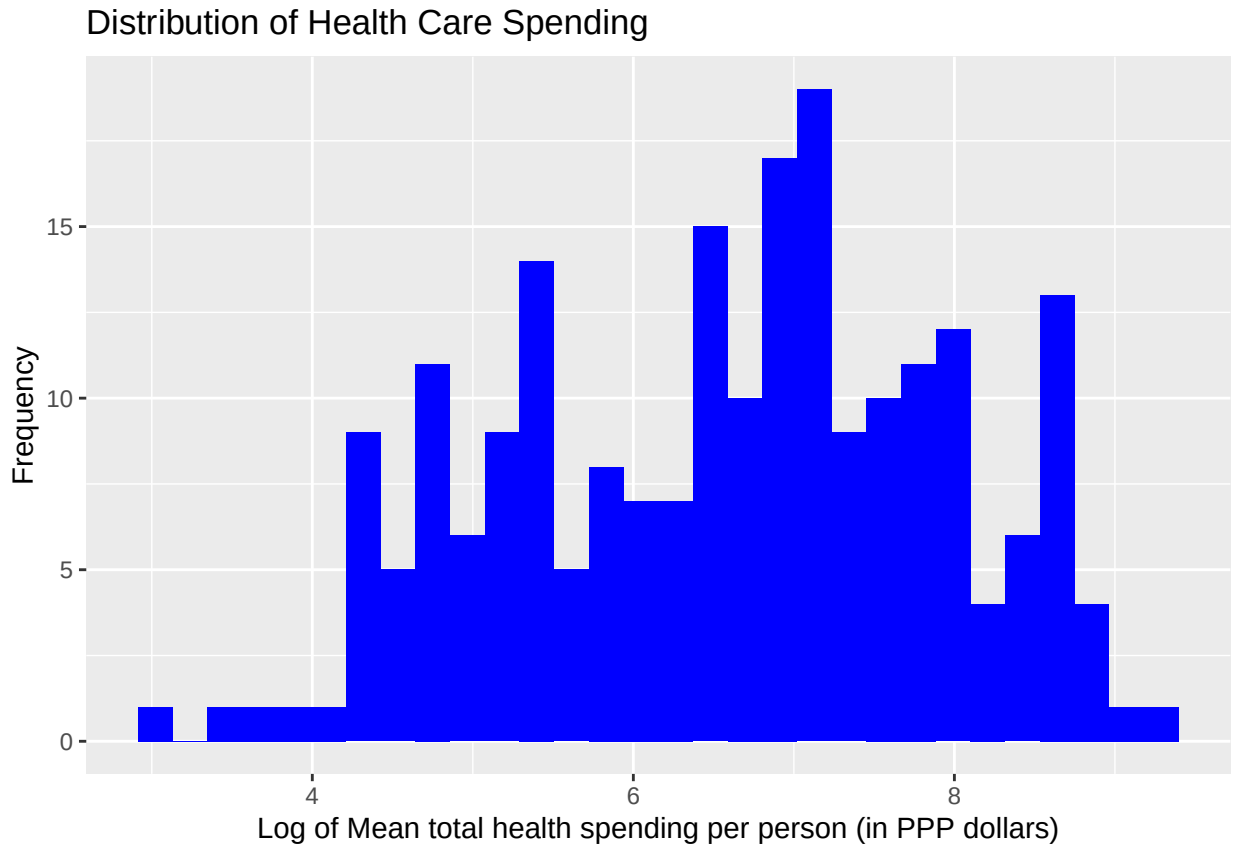
```
## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion

## Warning in lapply(econ2[j], as.numeric): NAs introduced by coercion
```

```
h_spend2 <- select(h_spend, -1, -4)
colnames(h_spend2)[1:3] <- c("Country Name", "Country Code", "Year")
```

```
full_data <- full_join(lifeexp, econ2, by = c("Country Name", "Country Code", "Year"))
full_data$Year = as.double(full_data$Year)
full_data <- full_join(full_data, h_spend2, by = c("Country Name", "Country Code", "Year"))
```

```
#Below is the distribution of log of Mean total health spending per person (in PPP dollars)
full_data %>%
  filter(Year == 2018) %>%
  filter(is.na(the_per_cap_ppp_mean) == FALSE) %>%
ggplot(aes(x = log(the_per_cap_ppp_mean))) +
  geom_histogram(fill = "blue") +
  labs(
    x = "Log of Mean total health spending per person (in PPP dollars)",
    y = "Frequency",
    title = "Distribution of Health Care Spending"
  )
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

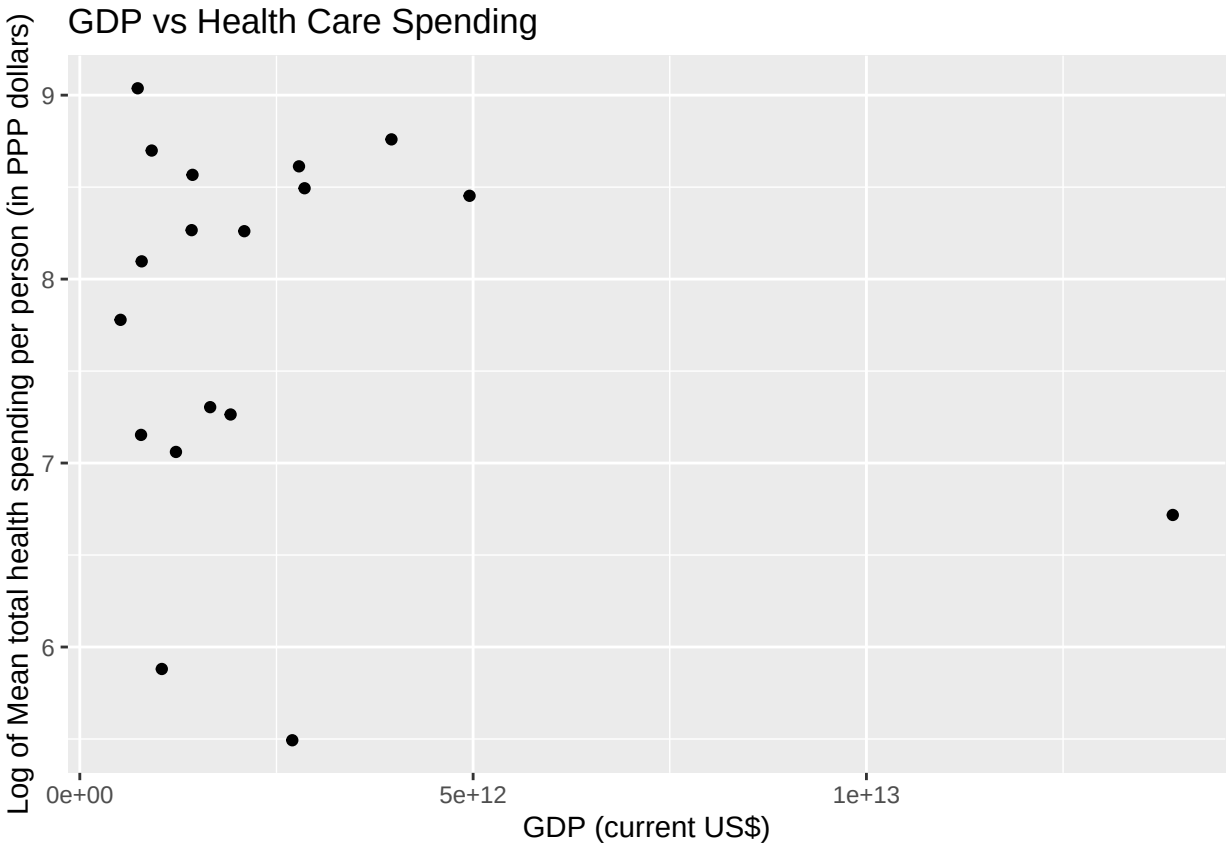



```
#NOTE: TO THE PLOT ABOVE, ADD NORMAL CURVE TO THE ABOVE DISTRIBUTION
```

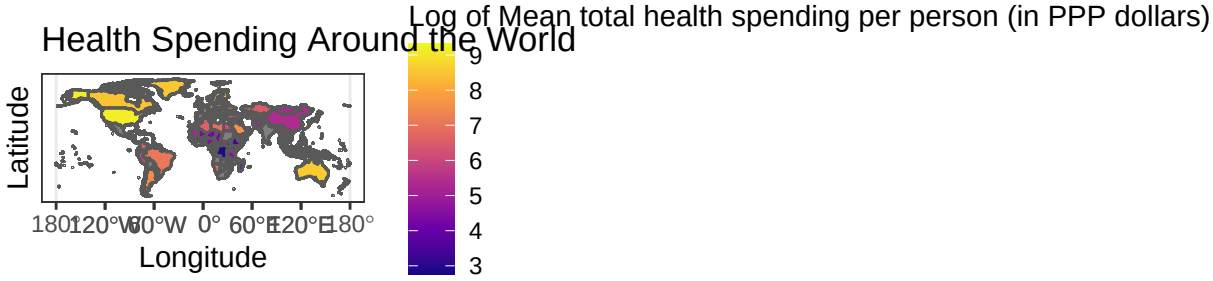
```r
#Below are a few quick summary stats for health spending per person
full_data %>%
  select(the_per_cap_ppp_mean) %>%
   filter(is.na(the_per_cap_ppp_mean) == FALSE) %>%
summary(the_per_cap_ppp_mean)
```

```
##  the_per_cap_ppp_mean
##  Min.   :   16.0
##  1st Qu.:  177.0
##  Median :  602.5
##  Mean   : 1202.1
##  3rd Qu.: 1477.8
##  Max.   :11027.0
```

```r
full_data %>%
 filter(Year == 2018) %>%
  filter(is.na(the_per_cap_ppp_mean) == FALSE) %>%
    filter(is.na(`GDP (current US$)`) == FALSE) %>%
ggplot(aes(x = `GDP (current US$)`, y = log(the_per_cap_ppp_mean))) +
geom_point(fill = "blue") +
 labs(
   x = "GDP (current US$)",
    y = "Log of Mean total health spending per person (in PPP dollars)",
    title = "GDP vs Health Care Spending"
 )
```



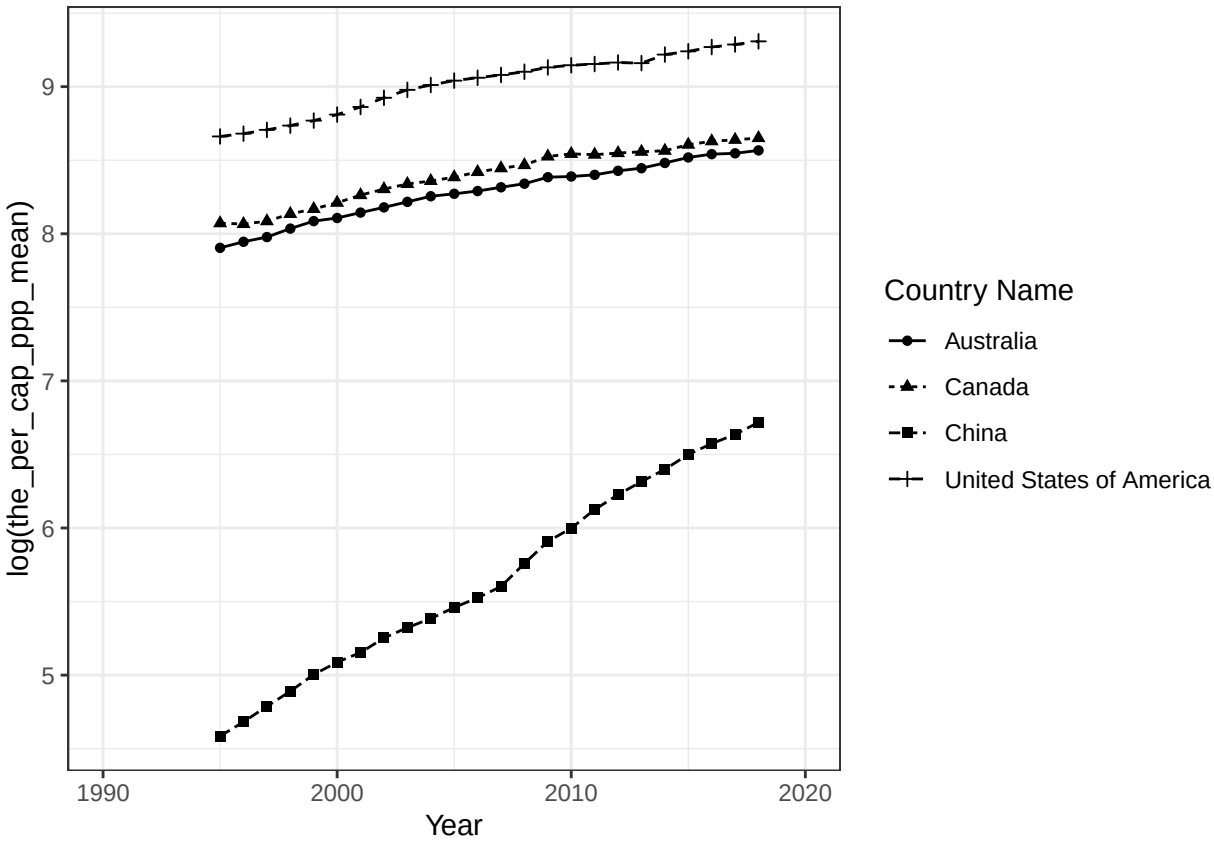
```
#Below, we're plotting a map of health spending around the world
# the classic dark-on-light theme for ggplot2 is nice for maps
theme_set(theme_bw())
# world contains the country information for plotting in addition to a lot of other information about t
world <- ne_countries(scale = "medium", returnclass = "sf")

healthspending.world <- merge(world, full_data, by.x="admin", by.y="Country Name")

ggplot(data = healthspending.world) +
    geom_sf(aes(fill = log(the_per_cap_ppp_mean))) +
    scale_fill_viridis_c(option = "plasma") +
    labs(x = "Longitude",
       y = "Latitude",
       title = "Health Spending Around the World",
       fill = "Log of Mean total health spending per person (in PPP dollars)")
```




```
#Now we're going to choose 5 countries and plot line graphs for trends in health spending over time
full_data %>%
  filter(`Country Name` == "Australia" | `Country Name` == "United States of America" |  `Country Name`
ggplot(aes(x=Year, y=log(the_per_cap_ppp_mean), group = `Country Name`)) +
  geom_line(aes(linetype= `Country Name`))+
  geom_point(aes(shape=`Country Name`))
```

```
## Warning: Removed 21 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```


log(the_per_cap_ppp_mean)
9
8
7
6
5
1990
2000
2010
2020
Year
Country Name
Australia
Canada
China
United States of America