# Problem Set #3

EH6127 - Quantitative Methods

Steven V. Miller

This homework makes use of data available in `{stevedata}` and implies the use of `{tidyverse}` to at least graph the data as prompted. `{tidyverse}` is not necessary to answer these questions though it will assuredly make the process easier. Load these two libraries to get started answering these questions.

```
library(tidyverse)
library(stevedata)
```

## Democracy and Economic Development (Around) 1949-50

This homework will refer to the `Lipset59` data set that is available in `{stevedata}`. This data set roughly approximates what Lipset (1959) did in his pioneering *American Political Science Review* paper on the "modernization" thesis.[1] The data include observations of 48 countries from around 1949-1950 based on an estimate of their democracy and their per capita income. You can find out more information about the data by visiting this part of the package's website, or with the following command.

```
?Lipset59
```

This codebook has a discussion of what the democracy variables are communicating and you are responsible for understanding it. Here's a little preview of these data.

```
Lipset59
```

```
## # A tibble: 48 x 11
##    country     cat    iso3c wbgdp2011est wbpopest unpop uninc unincpc xm_qudsest
##    <chr>       <chr>  <chr>        <dbl>    <dbl> <dbl> <dbl>   <dbl>      <dbl>
##  1 Australia   EE: S~ AUS        1.12e11   8.07e6  7912  5374     679       1.87
##  2 Belgium     EE: S~ BEL        7.75e10   8.29e6  8614  5015     582       1.85
##  3 Canada      EE: S~ CAN        1.79e11   1.33e7 13549 11797     870       1.52
##  4 Denmark     EE: S~ DNK        4.66e10   4.09e6  4230  2908     689       1.58
##  5 Ireland     EE: S~ IRL        1.77e10   2.86e6  2991  1260     420       1.20
##  6 Luxembourg  EE: S~ LUX        4.16e 9   2.85e5   295   162     553       1.75
##  7 Netherlands EE: S~ NLD        9.28e10   9.64e6  9956  5000     502       1.71
##  8 New Zealand EE: S~ NZL        2.48e10   1.82e6  1881  1610     856       1.85
```

[1] You can read more about the exact argument at stake here: http://svmiller.com/blog/2023/09/democracy-income-correlation-analysis/. I offer my blog synopsis in lieu of the article itself, though successful completion of this problem set implies an understanding of what the debate is. Implicitly, the exercise here assumes my interpretation of the relationship that Lipset (1959) first described (i.e. in the abstract, the causal sequence Lipset (1959) first described runs in reverse).

```
##  9 Norway      EE: S~ NOR        3.44e10   3.15e6  3233  1898     587       1.71
## 10 Sweden      EE: S~ SWE        7.87e10   6.70e6  6956  5426     780       1.79
## # i 38 more rows
## # i 2 more variables: v2x_polyarchy <dbl>, polity2 <dbl>
```

Answer these questions. A successful answer of these question must include the R code you used to help you answer the question.

1. It is not a formal assumption of OLS that the dependent variable is normally distributed, but the normality assumption for the conditional distribution of errors wants to imply the marginal distribution of the dependent variable is also normal. We need to create a per capita income variable (called: `gdppc`) that is equal to the estimate of GDP (`wbgdp2011est`) divided over the estimate of population size (`wbpopest`). Create this variable and show me how you might look at the distribution of this variable for these observations. Describe it to me in a sentence.
2. Let's assume our primary independent variable for democracy is Xavier Marquez' "Quick UDS" extensions (`xm_qudsest`). Show me how you might look at the distribution of this variable for these observations. Describe it to me.
3. (2 POINTS) Let's run a naive bivariate OLS regression that regresses `gdppc` (the variable you created in the first prompt) on `xm_qudsest`. Describe the results to me.
4. You can do a proto-bivariate OLS with `{ggplot2}` and a combination of `geom_point()` and `geom_smooth(method = "lm")`. Create a scatter plot with linear trend for our independent variable and dependent variable.
5. Change the smoother on this scatter plot to a LOESS smoother (i.e. `geom_smooth(method = "loess")`). What do you see and what do you think this implies for the effect of democracy on economic development?
6. It's not uncommon you'll be asked to do "robustness tests" for an independent variable of interest. We have a few other democracy indicators in the data frame as well. Regress `gdppc` on `polity2` and describe the results to me. Make sure you understand the distribution of the independent variable.
7. Regress `gdppc` on `v2x_polyarchy` and describe the results to me. Make sure you understand the distribution of the independent variable.