

Problem Set #2

IRIII.2 - Quantitative Methods in the Study of International Relations

Steven V. Miller

This homework is designed to get students making plots in `{ggplot2}`. For convenience's sake, it assumes the use of `{tidyverse}` when really its call is to use `{ggplot2}`. The data used will come by way of `{stevethemes}`, which students should have also installed. Students can also just manually load the data from Athena for this purpose, though the student should be explicit about this if this is how they load the data into R. Failure to do this will result in a one-point deduction.

Kilowatt Hours per Capita and GDP per Capita, 2010

This homework assignment will use data included in `{stevethemes}` as the `kwh_gdp` data set. This is a very simple data set that includes GDP per capita data and kilowatt hours consumed per capita data of over 130 sovereign states in 2010.

```
?kwh_gdp
```

Here's a little preview of these data.

```
kwh_gdp
```

```
## # A tibble: 136 x 7
##   country    iso3c  year  kwhpc  gdppc  ln_kwhpc  ln_gdppc
##   <chr>      <chr> <int> <dbl> <dbl>   <dbl>   <dbl>
## 1 Albania    ALB    2010  1943.  4094.   7.57    8.32
## 2 Algeria    DZA    2010  1020.  4496.   6.93    8.41
## 3 Angola     AGO    2010   206.  3497.   5.33    8.16
## 4 Argentina  ARG    2010  2878. 10386.   7.96    9.25
## 5 Armenia    ARM    2010  1686.  3143.   7.43    8.05
## 6 Australia  AUS    2010 10727. 52134.   9.28   10.9
## 7 Austria    AUT    2010  8383. 46904.   9.03   10.8
## 8 Azerbaijan AZE    2010  1603.  5844.   7.38    8.67
## 9 Bahrain    BHR    2010 18443. 21187.   9.82    9.96
## 10 Bangladesh BGD    2010   246.   777.   5.50    6.66
## # i 126 more rows
```

Answer these questions/do these commands. A successful answer of these question must include the R code you used to help you answer the question. Each question is worth a point.

1. Create a scatterplot where GDP per capita is on the x -axis and kilowatt hours per capita is on the y -axis.

2. Take the same plot you created above and add a “line of best fit” through the scatterplot. Your preferred method can be the LOESS smoother or the OLS (“lm”) line of best fit, but you should be explicit about which one you choose. Describe what that line tells you in two complete sentences.
3. Take the same template for the plot you created above and change the x -axis and y -axis variables to their log-transformed equivalents (i.e. the two variables with prefixes of `ln_`). In two complete sentences, describe what you see in comparison to the previous plot.
4. Take the same basic code for the plot you created in the third prompt and add a theme of your choice.¹ Choose whichever one you like (hint: they all typically begin with `theme_`).
5. Take the scatterplot where GDP per capita is on the x -axis and kilowatt hours per capita is on the y -axis (i.e. the one you created in the first two prompts). Notice that the x -axis is in dollars but that both values are large nominal numbers. Change the labeling on the x -axis and y -axis where the scale of x is in dollars and the scale of the y -axis has separators (like a comma) after every three digits.²

Extra Credit (1 Point Each)

6. Assuming the use of `{ggplot2}` to this point, notice that defaults on the x -axis that has been supplied to you have been placing breaks after every 30,000 USD for the GDP per capita variable. This is sensible, given the distribution of the data. However, people are more inclined to think in bases of 10 rather than bases of 30. Change that to where there are breaks after every 10,000 USD for the scatterplot of GDP per capita and kilowatt hours per capita.
7. Add a caption to this plot that tells the reader the data are sourced from the World Bank and OECD.

¹This assumes you’ve been using `{ggplot2}` code to this point, though it’s effectively mandatory to answering this question.

²Note that the default for this in many R packages is the American use of a comma (e.g. “60,000”) whereas Europeans favor decimal points (e.g. “60.000”) and Swedes use space separators (e.g. “60 000”). I’m ambivalent about which of these you choose and any one of them counts for credit.