

Analyzing GDP data

For Assignment 3, you will analyze GDP data in different ways. Part 1 will ask you to analyze time trends of GDP data. For part 1, you will use the dataset called *WorldGDP6697.csv*. You will create line graphs illustrating the growth trajectories of three different countries.

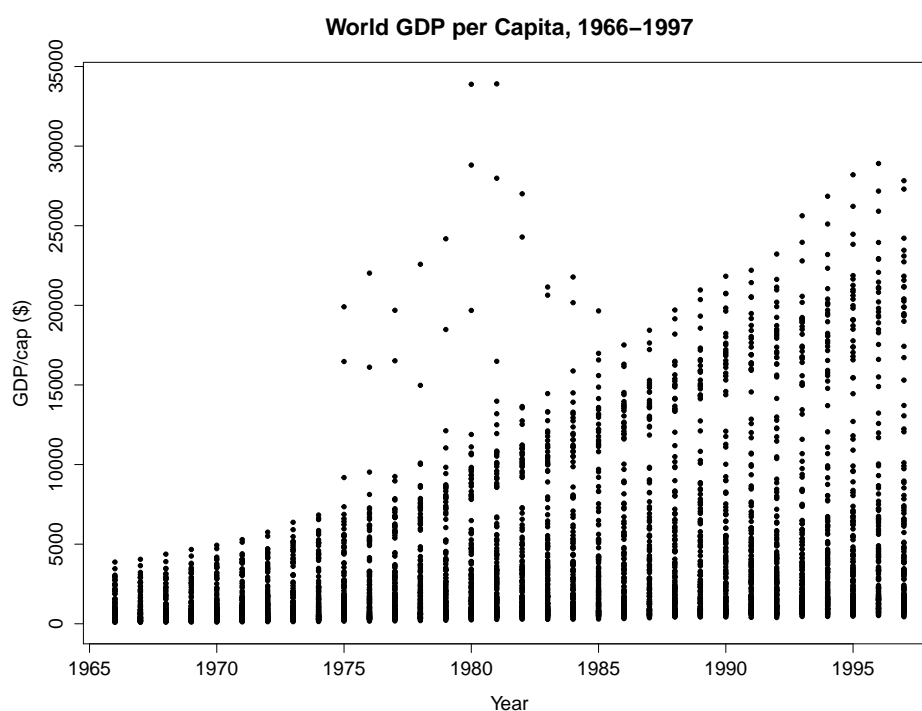
Part 2 will ask that you investigate a possible relationship between GDP per capita and a political instability. Here you will use the data set, *gdppc_instability.csv*, to create a scatterplot and run your first regression model.

This assignment will demand that you do some research into some countries. This is a very rare case where using Wikipedia will be tolerated. Be sure to use parenthetical citations.

1 Part 1: Time trends in wealth across countries

1.1 Motivation

As discussed in lecture, we find that many countries have grown wealthier through time while many have remained in the lower-end of income. Are there countries that best represent this growth of GDP from the 1960s to the 1990s? What about those that could not grow as much?



1.2 Exercises

Exercise 1: Run the code for the scatterplot but add the following argument to create a blank plot: `plot(x-variable, y-variable, type='n')`. Once you create this blank plot, add the following plot lines using the requested line types below. Please add a legend and the appropriate labels.

- Qatar OR Kuwait: (`lwd=2, lty=1, col='black'`)
- Singapore OR Japan: (`lwd=2, lty=3, col='red'`)
- Brazil OR Mexico: (`lwd=2, lty=5, col='darkgreen'`)

Exercise 2: Answer the following questions:

- Describe the trajectories for each of the countries you chose
- Provide one factor for each country to explain its trajectory
- Are there regional differences between these countries that explains the different growth trajectories?

1.3 Pointers

The best way to add country lines to the plot is to subset the data using the country of interest. Below I use the United States as an example.

```
#Subset specific country
usa = subset(gdpdat, gdpdat$country=="United States")

#Create plot
plot(gdpdat$year, gdpdat$GDPcapImp, xlab='xlabel',
     ylab='ylabel', main='title', type='n')
# Add lines (make sure you already created the subset)
lines(usa$year, usa$GDPcapImp, lwd=2, lty=3, col='blue')
# Add legend
legend("topleft", c("USA"), lwd=c(2), lty=c(3), col='blue', cex=0.8)
```

Notice that syntax in lines. Instead of using *gdpdat\$*, I make the line using *usa\$*, which is the subsetted data. The X and Y variables are the same, the main difference is what precedes the dollar sign. Remember that R will consistently think of this structure: *dataset\$variable*.

You can continue adding the second and third country. Remember to make your subsets first and add the lines and legend after opening up the blank plot.

As for the legend, you should have three items within each argument. For example, say I add a second country line, my legend should look something like this:

```
legend("topleft", c("USA", "Country2"), lwd=c(2, 2), lty=c(3, 2) col=c('blue', 'red'), cex=0.8)
```

The line types should correspond with the country so the ordering is important.

2 Part 2: Bivariate relationship between political instability and GDP per capita

2.1 Motivation

Does accumulating more wealth lead to political stability? In other words, are poorer countries likely to be more politically *unstable* than wealthier ones? You are expected to analyze and describe the relationship between economic development and political instability in the world. You will also examine whether regional differences exist or not. **Instability is your dependent variable.**

The data for this part of the assignment is called *gdppc_instability.csv*. There are seven variables in the dataset. The first is **country** and contains the names of 156 countries. The second variable is **gdppc05**. The variable measures the GDP per capita of each country in 2005. The third variable is called **instability09**. This variable is an index that describes the social and political situation in each country in 2009. Bigger values reflect higher instability. The least unstable country has a score of -17.11 and the most unstable country has a score of 44.19. The next three variables are dummy variables indicating whether a country is in Africa (**africa**), Western Europe (**weurope**), or other (**other**).¹ You can refer to the units for instability as points or units.

2.2 Exercises

Exercise 3: Create a scatterplot based on the relationship between political instability and log GDP per capita. Use *type='n'* in the plot. Add the regression line to the plot and then color code the points by region:

- Africa (*col='darkgreen', pch=20*)
- W. Europe (*col='blue', pch=20*)
- Other (*col='red', pch=20*)

¹Note: USA, Canada, Australia and New Zealand are classified into W.Euro. So W.Euro can be looked at as a group of wealthy, developed countries.

Exercise 4: Run a bivariate regression in R using log GDP per capita as one of your variables. Keep in mind what the unit is. Round to 2-3 decimal places and present the results like in the table below. For the R-Squared, use the *Multiple R-squared* from the R output. Please label accordingly.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)				
independent variable				
R-Squared:				
Number of Observations:				

Exercise 5: Answer EACH question below in 1-2 sentences:

- What is the slope and how is it interpreted?
- Is it statistically significant? How do we know?
- Is the result substantively significant and how so?
- What percent of the variation in Y is explained by X?
- Provide one reason why this relationship is plausible
- Provide one reason why this relationship is questionable

Exercise 6: You will focus on the expected values and the residual values for Costa Rica and Argentina. For **each country**, please answer the following questions.

- Given the level of Log GDP, what is the expected value for instability? Please indicate the observed Log GDP value.
- What is the residual? Is it positive or negative?
- Provide one reason why this country has a positive or negative residual.

You can find the observed values by using this code:

```
data[data$country=='Argentina',]
data[data$country=='Costa Rica',]
```

2.3 Pointers

Adding points to a blank scatterplot is very similar to adding lines. Both require subsets of the data, and have a similar X and Y format.

```
plot(data$Xvariable, data$Yvariable, type='n')
abline(lm(data$Yvariable ~ data$Xvariable), lwd=2)
```

```
asia = subset(data, data$asia==1)
points(asia$Xvariable, asia$Yvariable, col='orange', pch=20)
```

Notice the x and y variables are consistent with the syntax for *plot*. What changes in the code is the subset preceding the dollar sign.

To get number of observations, use:

```
nobs(model)
```

3 Administrative Details

- This assignment is due at 11:59 PM, 11/13/2020
- Must submit in PDF format
- You may work with one partner in regards to the R aspect of the assignment. Remember that working with a partner does not mean copy from each other.
- You **do not** need to submit your R code.