Economics 312 Spring 2019
Project #1 Assignment Due: 11:59pm, Mon., February 4

Project #1 aims to give you some experience locating data, downloading them into an Excel spreadsheet, moving the data into Stata, and doing some basic descriptive analysis. The data in question are annual "harmonized" unemployment rates for OECD countries from 1990 to the present. The source is the OECD statistical library, which can be accessed at http://www.oecd-ilibrary.org/statistics. You must access this from a Reed campus computer (or use the Reed proxy server) because the data set is available only under a subscription held by the Reed Library. (To help with general questions about using datasets, the Reed Library and Computer User Services have created a team of data-savvy staff members. Go to www.reed.edu/data-at-reed/ or email data@reed.edu for more information.)

For this project, each student will do the project independently and submit his or her own report. The report is due by midnight on Monday, February 4. You should submit your report in Word or pdf format as an email attachment to the instructor at parker@reed.edu. Please copy and paste your Stata outputs into your document. (That makes it easier for me to track down the source of any anomalies.)

Cautionary note on using "found data"

The Organisation for Economic Cooperation and Development (OECD), whose data you will use for this exercise, is a group of rich economies. These countries have a joint data repository and try to maintain some degree of compatibility in data-collection methods and definitions. OECD data are generally of high quality, but there are still some potential problems that must concern you.

The membership of the OECD has expanded dramatically since its 1961 inception, bringing in a total of 34 countries. Because some countries joined quite recently (Slovenia, Israel, Estonia, and Chile were the most recent in 2010), their data will not go back as far as the original members. In addition, some countries have experienced significant changes since joining, notably the unification of Germany and the fission of Czechoslovakia and Yugoslavia. These changes may make the data before and after the change date incompatible.

Even where no dramatic changes have occurred, countries often change their methods of data collection in ways that may (or may not) compromise the consistency of the time series. You must be attentive to these situations as you collect data, both for this project and for the rest of your careers. Data collecting organizations such as the OECD usually mark discontinuities in data series with some kind of notation such as a superscript or a line drawn

at the point of the change. Be alert to these and always follow links to the explanation of the data change. In many cases it will be a minor change that you can ignore, but in some cases you may need to think carefully (and document in your report, paper, or thesis) the implications of the change for your sample.

More broadly, a huge *caveat emptor* applies to data use in general: *Just because someone typed a column of numbers onto a Web page does not mean that they are consistent, reliable, accurate or even useful.* Always investigate the source of any data whose provenance you do not recognize. Official statistical agencies of advanced countries are *usually* reliable, though recall that for political reasons the previous government of Argentina published bogus inflation statistics during the early part of this decade. Data that have been vetted by major international organizations such as the OECD, IMF, World Bank, United Nations, etc. are likely to be more reliable, but it's always useful to check any notes they may provide about data quality.

A. Locating and downloading the data

The following instructions will sketch the process of finding and downloading the data, but will not (by design) give you all the details. If you get lost, use available resources (peers or the instructor) to help find your way. The object of your search is the OECD data on "harmonized unemployment rates" for the entire population.

- Begin at http://www.oecd-ilibrary.org/statistics.
- Unemployment rates are found in the "Labour market statistics" section (which may itself be inside the "Labour" section) of the "Main Economic Indicators" database.
- The little pink button that says "Data" is helpful in getting the actual data onto the screen.
- Once you get data onto the screen, the "Customize" and "Export" buttons near the top will be useful in getting the data into the form you want and downloading it as an Excel spreadsheet.
- Your download should have only one series, the "Harmonized unemployment rate, Total." These data are available on a monthly, quarterly, or annual basis. To keep this simple, you should only get annual data for the period 1990 to the present.
    - Hint: When you are looking for the series, be sure that you open *all* of the little [+] boxes on the left of the list so that you get down to the finest level of detail. If you check an unopened category rather than individual variables within it, your data may not display.
- Once you have the data on your screen, pay careful attention to several pieces of information in addition to the numbers:
    - Some data are missing, indicated by ".."
    - Some countries have a little blue "i" indicator. This tells you that there is information about potential problems with that variable for that country.

    - In some cases, the "i" is warning you about breaks in the data series, which are indicated in the table by "(B)."
    - Gather as much information about these data problems as you can and include what you find in your report. Based on the nature of the problem, you should decide whether or not the country can reasonably be included in your dataset and justify that decision.
- For convenience in loading into Stata, you will want each country's data in a single column, with each year in a single row. (The default may be the opposite.) You can change this in Customize → Layout.
    - (A different format—a single column with one country on top of the other—would be desirable if we were doing different things with the data in Stata.)
- In addition to the member countries, the OECD data will by default include four aggregated variables for the Euro Area, the EU, the G7, and the total OECD. You don't want these, so either deselect them on the OECD site or delete these columns in Excel after downloading.

B. Copying the data from Excel to Stata

Once you have downloaded the data in Excel format, open the dataset in Excel. (It *may* be possible to use Excel substitutes, but there may be problems. I strongly recommend that you use Excel, which is available on all public Reed computers. Even using Excel, I get a warning message opening the downloaded file, but it opens fine in spite of the message.) If you have done the download correctly, the top-left of your spreadsheet should look something like this:

	A	B	C	D	E	F	G	H
2	Dataset: Short-Term Labour Market Statistics							
3	**Subject**		Harmonised unemployment rate (monthly), Total, All persons					
4	**Measure**		Level, rate or quantity series, s.a.					
5	**Frequency**		Annual					
6	**Unit**		Percentage					
7	**Country**		Australia	Austria	Belgium	Canada	Chile	Czech Republic
8	**Time**	i						
9	1990		6.9	..	6.6	8.2	7.8	..
10	1991		9.6	..	6.4	10.3	8.2	..
11	1992		10.7	..	7.1	11.2	6.7	..
12	1993		10.9	4	8.6	11.4	6.5	4.3
13	1994		9.7	3.9	9.8	10.4	7.8	4.3
14	1995		8.5	3.9	9.7	9.5	7.3	4
15	1996		8.5	4.3	9.5	9.6	6.3	3.9

- We usually need to do some "cleaning" of the data format in Excel before copying into Stata.

- In this case, we want to get rid of the row and column with the little "i" flags (Row 8 and Column B above). After deleting these, your left column will be year numbers but will be headed by the "Country" cell. You should change that cell to "Year" so that Stata will call the year variable "year" instead of "country."
- A second cleaning task is to change the ".." notations for missing observations to empty cells (which Stata will recognize as missing). The Replace function in Excel (on the Find & Select menu at the top) allows you to replace the string .. with an empty string throughout the sheet. This will make those cells empty.
- A third involves the year numbers. Note the little green triangles in the cells of Column A above. This means that there is something anomalous about the cell's value. To change these cells to numbers, first select all of the affected cells. Once you select them, you will see a yellow diamond icon with an exclamation mark in it. Click this icon and it will tell you that these cells have "Number Stored as Text" and offer you options about what to do. Click on "Convert to Number" in order to make Excel (and later Stata) know that you want to treat these as numbers.

- As long as they are not too large, the easiest way to get data sets from Excel to Stata is simply to copy and paste. In Excel, copy the set of cells containing the data, including the row of variable (country) names at the top. Open Stata, then open its data editor. Select the top corner cell of the empty Stata data window and paste, selecting the option to treat the top row as variable names. Your dataset should now be in Stata. Note that the empty cells on the spreadsheet now contain a single dot, which is Stata's indicator for missing data. Make sure that all of your variables are displayed in black (numeric data) rather than red (text strings). Close the data editor and save your dataset.

C. Basic descriptive analysis in Stata

Now you have a Stata dataset with each country's unemployment rate as a single variable. Your dataset is time series in nature; you need to tell Stata about the time dimension in order to take advantage of any of the time-related functions. The appropriate command is "tsset year" where "year" is the name of your variable that counts the years. (Note that all of the variable names have been converted to lower case when Stata brings them in from Excel. Stata variables are case-sensitive, so "year," "Year," and "YEAR" could be three different variables.) You should begin your report with a summary of the data-collection procedure and outcome.

Once you are set up in Stata, perform the following analyses. Describe the results verbally in your report and copy the relevant Stata outputs into the report as tables or figures.

- Use the summarize command to show sample statistics for the unemployment rate for your countries. Describe the key features of the results.
    - Some countries have very short samples (*e. g.*, Switzerland). Are these results comparable with the other countries? What would be an appropriate solution given the limitations of your data?
    - Any Stata command can be restricted to a subset of observations by appending something like if time < 2000 after the command. (You can combine conditions such as if time < 2000 & time >= 1995 to get the period 1995–99.) Look at some 5-year (or so) subsamples and comment on the variations in the level of unemployment over time. Look for sets of countries that seem to have similar movements in unemployment over time.
- Choose a few countries (no more than 5 to keep the graph readable) and use the tsline command to plot their unemployment rates on the same graph. (You may want to use the Graphics menu at the top rather than typing the command in to the command line. The menu will help you determine what options are available. Each variable [country] that you include will be a separate "plot" that must be defined.)
- Choose a subset of countries (choose 7 so that Stata does not have to break the table) and calculate the correlation matrix among them. (I suggest the pwcorr command rather than the correlate command. The former calculates each correlation coefficient using all observations for which each particular pair of variables is not missing; the latter uses only the common set of observations that is available for *all* of the variables in the table.) Describe the results in your report (and copy the table), indicating which countries seem to be strongly positively correlated and which do not. Is there evidence of a common business cycle across these OECD countries? Are there any surprises here?