

Graphing and Data Transformation in Excel

ECON 285

Chris Georges

This is a brief tutorial in Excel and a first data exercise for the course. The tutorial is written for a novice user of Excel and is not meant to be comprehensive. If you are already familiar with the use of spread sheet programs, you should still complete the data exercises in this handout (parts 6-12). This work should not be turned in.

1. Spread Sheets

Spread sheet programs allow users to transform and display data in convenient ways. Excel is a widely used spread sheet program, and learning to use it will give you a fairly generic knowledge of spread sheet applications.

2. Access to Excel and the Data:

I will make the Excel data files that we will use in this course available from my Web page <http://people.hamilton.edu/christophre-georges/>. You should be able to use any recent version of Excel. If you don't have Excel on your own computer, you can access it in the KJ labs and store your work on the SSS. Unfortunately, Excel menus vary across platforms and vintages. This tutorial is written for Excel in Office 2016, with a few comments about other versions.

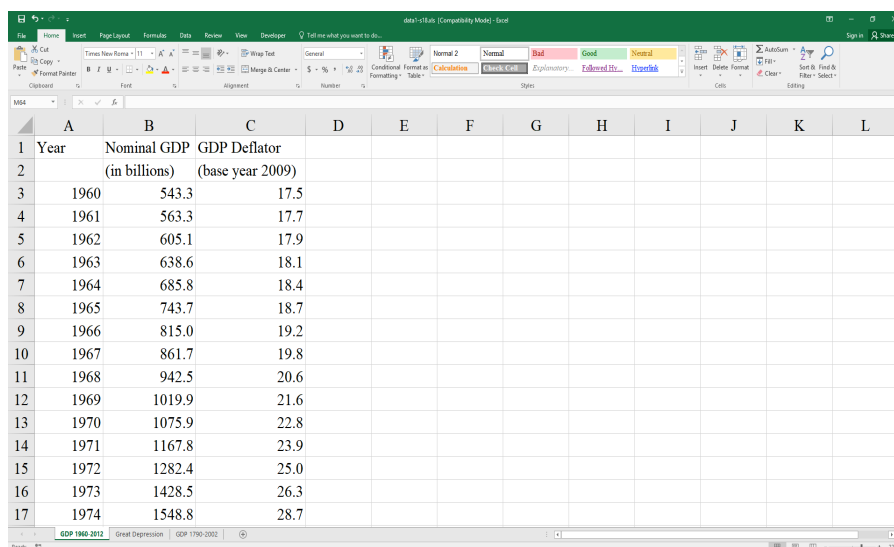
3. Saving Your Work:

You will want to save your work periodically, especially during long exercises.

4. Opening a file (spread sheet) in Excel

Let's open the Excel file (data1.xls) that we will use in this tutorial. First go to my web page using a web browser (e.g., Chrome). From my home page, select my **course work** page, and then the **Econ 285** page. Left click on **data1.xls** and select "open with Microsoft Office Excel." Make sure that it opens in Excel and not your web browser, as the web browser will give you very limited functionality. (Note: If you are not given the option to open in Excel, then right click on the link to data1.xls, save it to your computer or SSS space, and then open it in Excel.) Once you have opened the file in Excel, you can work on it and save changes to your machine or SSS space as you wish.

The file contains data on Nominal GDP and the GDP Deflator for 1960-2016 and should look similar to the picture at the right. Note that, while the Commerce Department calculates GDP quarterly (four times a year) the data in the file are annual (averages of the quarterly data for each year).



Year	Nominal GDP (in billions)	GDP Deflator (base year 2009)
1960	543.3	17.5
1961	563.3	17.7
1962	605.1	17.9
1963	638.6	18.1
1964	685.8	18.4
1965	743.7	18.7
1966	815.0	19.2
1967	861.7	19.8
1968	942.5	20.6
1969	1019.9	21.6
1970	1075.9	22.8
1971	1167.8	23.9
1972	1282.4	25.0
1973	1428.5	26.3
1974	1548.8	28.7

5. Data Manipulation I: It Slices, It Dices, It Does Calculations...

You should now have the spreadsheet data1.xls open in Excel on your screen. Each of the first three columns contains data. Column A contains the dates 1960-2016 Columns B and C contain the values of Nominal GDP and the GDP Deflator respectively for the U.S. for each year (as reported by the Department of Commerce).

The spread sheet is organized in cells. For example, Nominal GDP for 1960 is located in column **B** row **3**, which Excel calls cell **B3**.

Before using the data in columns A-C, we can note that Excel can be used like a hand calculator to do simple calculations. First go to an empty cell (say cell D5) by pointing and clicking (with the left mouse button) on it. The cell should now be highlighted. Now let's divide 100 by 5. To do this, type =100/5 and then press the <Enter> key. The answer, 20, should now appear in the cell.

Note that the equals sign in =100/5 tells Excel that this is a calculation rather than text. Had you entered 100/5 without the equals sign, Excel would have treated your entry as text and not have done the calculation. You can proofread your entry by looking at the white edit box, above the column markers.

You can now clear the cell that you just edited by right clicking on the cell (which brings up a menu) and choosing "Clear Contents."

6. Data Manipulation II: Per Capita GDP in 1960

Now let's transform a piece of data already in the spreadsheet. The population of the U.S. in 1960 was 179.323 million (which is 0.179323 billion). What was the level of Nominal GDP per capita (i.e., the dollar amount of stuff produced per person) in that year? To find out, we need to divide Nominal GDP in 1960 (which according to cell B3 was 543.3 billion) by the population in 1960. To do this, go to any blank cell (e.g., cell D3) by pointing and clicking there. We could type

=543.3/ 0.179323 <Enter>

(both are in billions). But we can also use the cell name for the Nominal GDP in 1960 and type

=B3 /0.179323 <Enter>

Try the latter. The answer, 3029.634, should now appear in the cell. The dollar value of output per person in the U.S. in 1960 was \$3,029.63.

7. Data Manipulation III: Real GDP for 1960-2016

Nominal GDP for 1960 gives us the dollar amount of stuff produced in the economy in 1960, at 1960 prices. To compare the "real" amount of stuff produced in that year to the amount produced in a more recent year, we need to "deflate" Nominal GDP by a price deflator such as the GDP Deflator. Real GDP is defined as follows:

$$\text{Real GDP} = (\text{Nominal GDP} / \text{GDP Deflator}) * 100$$

To calculate Real GDP for 1960, go to cell D3 and type

$=(B3 / C3)*100$ <Enter>

The answer, 3108.738, should now appear in the cell. Real GDP for 1960 was about \$3.1 trillion.

Shortcut: note that in the formula above, rather than typing in “B3” and “C3”, you can click on the cells to insert them into the formula as you are typing it.

Question: Why is Real GDP in 1960 a much larger number than nominal GDP for 1960 (\$543.3billion)?

To see how Real GDP (a measure of the amount of stuff produced in a year) for the U.S. has changed over time, we need to perform this calculation for each year (1960-2016). Now we can really put the spread sheet to work for us. Cell D3 has the calculation for 1960 in it. Excel shows you the numerical answer in the cell, but has also saved the formula by which it was calculated: (B3/C3). We can copy the formula to other years by doing the following:

- point and click on cell D3
- move your cursor over the little square in the lower right corner of the cell (D3), so that the cursor changes to a thin plus sign (from a thick plus sign)
- then either
 - 1) double click on the little square, or
 - 2) click on the little square, and, while holding the mouse button down, drag the cursor down the column until you reach cell D59, release the mouse button, and then click anywhere to get rid of the highlighting.

The values of Real GDP for 1960-2016 should now appear in cells D3 – D59. You can double check this by hand or by noticing that if you click on one of these cells (say D10, which should be Real GDP for 1967) that the correct formula $=(B10/C10)*100$ appears in the edit window (middle top). Whenever you copy a formula from one cell to another in a spread sheet program, the specific cells in the formula are not copied, but rather the relative spatial locations of those cells. I.e., the spreadsheet interprets the formula in D3 as, "divide the number in the cell two to the left by the number in the cell one to the left, and multiply by 100."

8. Data Manipulation IV: Growth of Real GDP for 1961-2016

Let's keep our calculations of Real GDP, in column D, and calculate one more variable, the annual growth rate of Real GDP, for each year. The growth rate of Real GDP for 1961 is the percent change of Real GDP from 1960 to 1961, which is the difference between the 1961 and 1960 values, divided by the 1960 value.

$\text{Growth Rate of Real GDP}_{1961} = ((\text{Real GDP}_{1961} - \text{Real GDP}_{1960}) / \text{Real GDP}_{1960}) * 100$

To calculate this

- point and click on cell E4
- type: $=(D4 - D3)/D3*100$ <Enter>

The answer, 2.55244, should now appear in cell E4. The economy grew by about 2.5 percent from 1960 to 1961.

Now calculate the growth rates for the remaining years just as you did above for real GDP (remember to always let Excel do the work for you).

You will now probably want to label the two new variables that you have created by typing in the names (e.g., "Real GDP" and "Growth Rate Real GDP") in cells D1 and E1 (just type in these names in those cells). If the titles do not fit into the cells of the spreadsheet you can resize the column by moving the cursor between the column headings and then either clicking and dragging horizontally or simply double-clicking.

9. Economic Growth

Question: What is the average growth rate for the U.S. economy since 1960?

To answer this question, we can take the average of the annual growth rates of Real GDP from column E for 1961-2016. These are found in the range E4 – E59. Excel calls this range E4:E59 (or E4..E59). To take the average of these values,

- go to a blank cell (point and click)
- type the following: =average(e4:e59) <Enter>

You should now see the answer, 3.071. The U.S. economy grew at an average rate of about 3.1 percent per year since 1961.

Question: Has growth sped up or slowed down on average over this period?

To answer this question, we can (for example) compare average growth rates for earlier and later years in this time period. Specifically, let's split the time period in half and compare the average growth rates for two periods: 1961-88, 1989-2016. To do this,

- go to a blank cell (point and click)
- type: =average(e4:e31) <Enter>
- go to another blank cell (point and click)
- type: =average(e32:e59) <Enter>

Shortcut: rather than typing “average” and the cell ranges, you can start to type =average, then double click on AVERAGE when it pops up in a list, then highlight the cell range that you want to select.

You should find that annual growth has slowed by about 1.2 percentage points. This result is fairly robust to the choice of cutoff dates. In general, there appears to have been a slowdown in growth over the post WWII period. If you were to look at each decade, you would find that there was very rapid growth on average in the 1960s (about 4.29% per year), with slower average growth in the 1970s and 1980s (about 3.20% and 3.36% per year respectively). Growth rates rebounded in the late 1990s, prompting some economists and politicians to speculate that we had entered a new golden age of growth (a “new economy”). The recession in 2001 ended that burst of growth, and growth was lackluster in the 2000s (average of 2.44% through 2007 followed by another recession in 2008-9), and more lackluster growth since 2009 (2.14%). The prospects for the pace of future growth continue to be hotly debated.

Question: Is a 1.2 percentage point decline in the annual growth rate substantial?

Question: What do you think might cause differences in growth rates over long periods of time?

10. Business Cycles

The U.S. economy has tended to grow over time. As we saw above, the average growth rate since 1961 has been roughly 3.1% per year. However, as we see in column E, growth rates have also varied substantially from year to year. Indeed in many years the growth rate has been negative: i.e., the amount of stuff produced in the economy has fallen rather than risen. We can call a year in which Real GDP falls a "recession year." The semi-regular sequence of recessions and expansions that tend to characterize capitalist economies is called the "business cycle."

Exercise: Look in column E and find all of the recession years since 1960.

Question: What do you think may have caused each recession?

As a caveat, note that, even though real GDP did not fall at an annual frequency in 2001, that year is also considered a recession year (e.g., by the National Bureau of Economic Research (NBER)) because GDP growth slowed markedly, and unemployment rose markedly. The NBER dates the most recent recession of 2008-09 (sometimes called the "great recession") as having started in December of 2007 and ended in June of 2009 (i.e., the recovery as having started in July 2009).

11. Graphing I: Growth and Business Cycles since 1970

Let's look at the pattern of long term economic growth and short term business cycles graphically. First, let's graph Real GDP from 1970-2016. For Excel in Office 2016:

- highlight the series of cells (D13 – D59) that we want to graph (point, click, and drag)
- select the "Insert" tab (upper left region of the Excel window)
- select 2D Line chart by hovering over the chart icons until you find the icon for "line chart", click on line chart, and select 2D line style

You should see a graph "floating" on your spreadsheet. You can resize the graph by clicking and dragging on a corner edge of the graph. You can move the graph on the spreadsheet by clicking and dragging on the center of the graph.

Note: some Excel toolbars include a graph icon that you can click after highlighting the data rather than using the "Insert" menu.

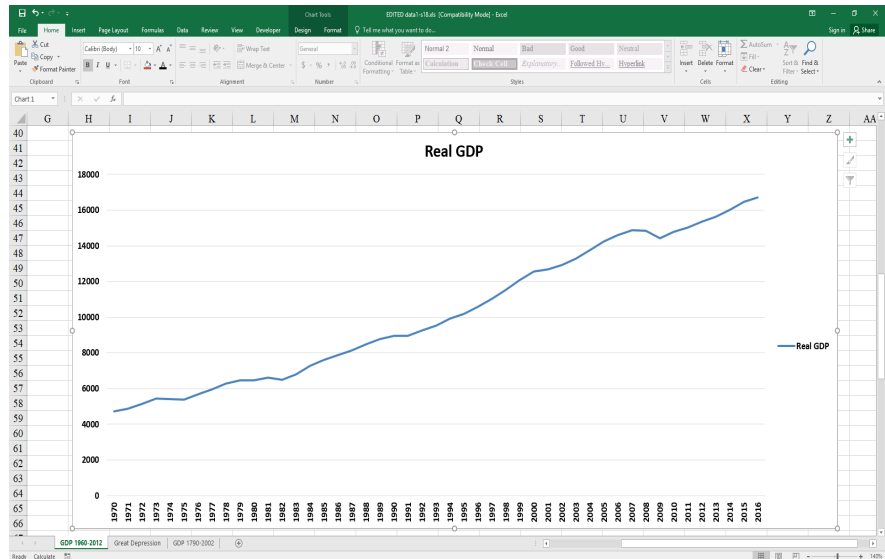
Now let's go back and add dates to the X-axis. For Office 2016:

- left click on an empty part of the graph to select it and select the "Design" tab (above)
- select "Select Data"
- In the Horizontal Axes window select the "Edit" button and then click the box to the right of the "Axis label range" window
- highlight the cells in which the dates 1970-2016 reside (cells A13 – A59) (point, click, and drag)
- hit <Enter>

Let's also add a legend (label what's being graphed)

- go back to "Select Data"
- Under the "Series" window, select Series1 (left click), and then "Edit"
- click on the box to the right of the "Series Name" window
- click on cell D1 (cell D1 is where you placed the name "Real GDP") and hit <Enter>
- select "OK"

You should now have a useful graph that looks similar to the one at the right. You can print the graph, if you wish to do so, by clicking on it once (left mouse button) and selecting "Print" in the "File" menu (top left corner of screen) or the print icon (top left).



Note: in Excel 2016 you can also get into the "Select Data" window by right clicking on an empty part of the graph. In some versions of Excel for Mac, you can select "Source Data" in the "Chart" menu or click on the chart and using the "Formatting Palette."

We will now make a series of additions and changes. First, let's add a second variable, Nominal GDP (from column E), to the graph so that we can compare the patterns of Nominal and Real GDP over time. In Excel 2016:

- Select "Select Data" (under the "Design" tab as above)
- select "Add"
- click on the box to the right of the "series values" window (to add the second data series)
- highlight cells B13 – B59 (point, click, and drag) and hit <Enter>
- click on the "Series name" box and then cell B1 (which is the cell with the text "Nominal GDP" in it) and then <ENTER> (to add the name of this second variable)
- select "OK"

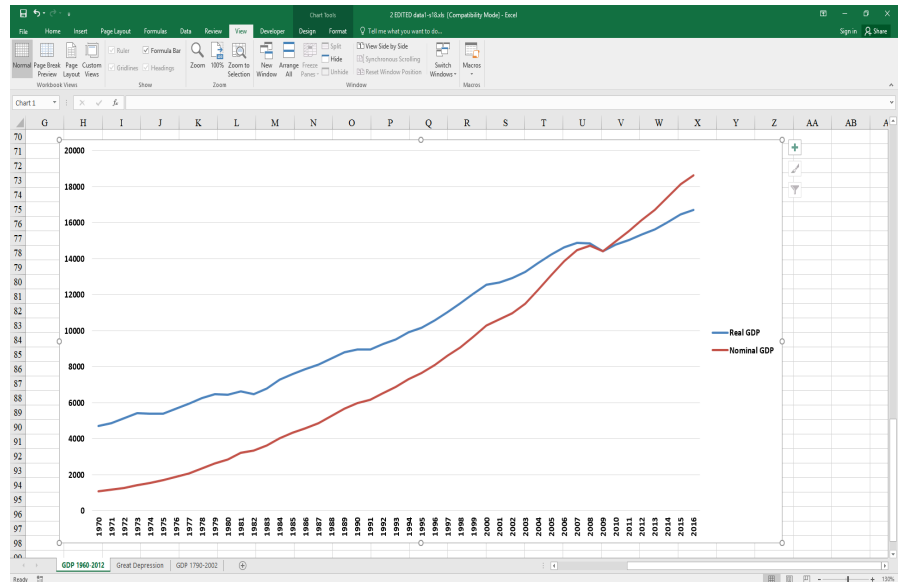
Note: for other versions of Excel, try clicking the right mouse button on the graph or going to the "Chart" menu, select "Source Data" then "Add" etc.

Notice that Real GDP falls in 7 of the years since 1970 while Nominal GDP has fallen only once (2009). The difference is due to inflation, which has, at least in the years that we are looking at, been great enough to drive up Nominal GDP even in most recession years when Real GDP has fallen.

Recall that:

$$\text{Nominal GDP} = (\text{Real GDP} * \text{GDP Deflator})/100$$

If Real GDP falls, but the GDP Deflator rises enough (specifically: by a larger percentage), then Nominal GDP rises.

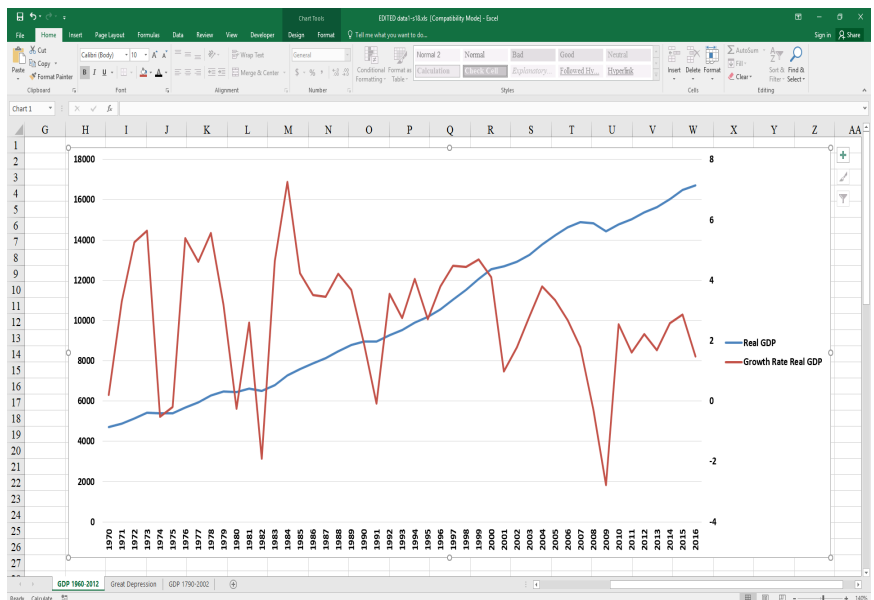


Now let's replace Nominal GDP in the graph with the annual Growth Rate of Real GDP (from column E), to get a better look at the size of business cycle fluctuations.

- go back to “Select Data” and select “Nominal GDP” and then “Edit”
- click on the box to the right of the “Series values” window
- highlight cells E13 – E59 (point, click, and drag) and hit <Enter> (this replaces the series)
- click on the "Series name" box, and then cell E1, and then <Enter> (this corrects the name)
- select “OK”

Now let's give the growth rate a second scale (so that it isn't scrunched into the bottom of the graph)

- click the left mouse button on the actual line-plot of the growth rate (not the plot area) to select that series
- Choose the “Format” tab
- double click on “Format Selection”
- select “secondary axis”



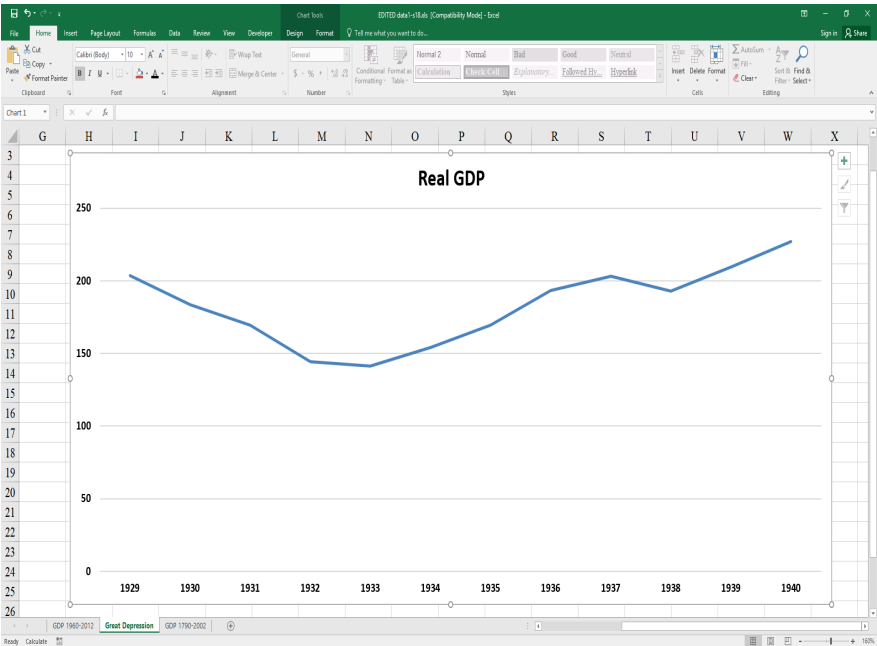
Now you should be able to see clearly both the overall trend and degree of fluctuations of Real GDP, reading the level of Real GDP off of the left hand Y-Axis scale and the Growth Rate off of the right hand Y-Axis scale.

Question: How smooth or volatile has economic growth been since 1970?

12. Graphing II: The Great Depression

Now let's take a look at the Great Depression, which lasted from 1930-1933. Go to sheet B of the spreadsheet (by clicking on the folder tab marked B near the bottom of the screen). This sheet contains data on Nominal GNP and the GNP deflator for 1929-1940. You will first need to construct a Real GDP series in column D as you did (in section 7) above. Once you have done that, let's proceed to make a graph of Real GDP (from column D) for 1929-1940.

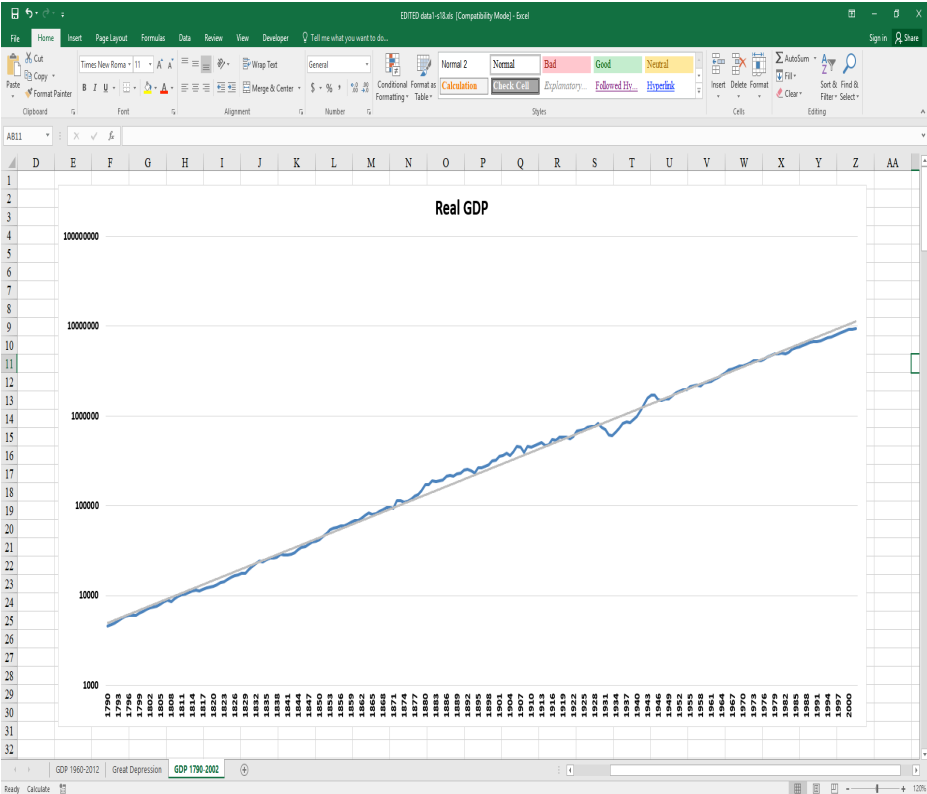
- highlight the series of cells (D3 - D14) that we want to graph (point, click, and drag)
- use "Insert" to create the new "line" graph
- put dates on the X axis by selecting 1929-1940 (i.e., cells A3-A14) as the horizontal axis labels and cell D1 as the corresponding label name.



Question: By how much (what percentage) did Real GDP fall between 1929 and 1933? How long did it take the economy to recover from the Depression?

13. Graphing III: Log plots

Finally, let's look at a longer period of growth. The third sheet of the Excel workbook contains estimates for Real GDP from 1790 to 2002. Please plot this data with time on the X-axis. Now please make the scale of the Y-axis logarithmic (and also make the minimum value on the Y-axis 1000 -- see the Addendum below for how to do this). How does this change the plot? Now also please add an exponential trend line. Note that, if the growth rate of real GDP were constant, the log plot of the data would lie on a straight line (like the trend line).



Addendum

Other Useful Operators

=sum()	(sum of elements in a cell range)
=median()	(median of elements in a cell range)
=average()	(mean of elements in a cell range)
=min()	(minimum of elements in a cell range)
=max()	(maximum of elements in a cell range)
=ln()	(natural log of number in a cell)
=x^y	(raise x to the power y)

More on Graphing:

You can change a number of aspects of a graph in Excel. For example you can put in titles, gridlines, data markers, etc. (select the graph and then use the Layout tab) You can also change the appearance of the plots in various ways, such as changing plot-line colors, markers, etc. , and change the appearance (e.g., fonts) of legends, axis labels, etc. You can also see different types of graphs (such as 3-D, bar, pie, etc.) with “Change Chart Type” under the Design tab.

Log Plots:

Select the Format tab. Then double click on the labels for the Y-axis of a graph to select the axis. Then select “Format Selection” to get a menu that lets you change the scale to “logarithmic scale.” Note that a linear (straight line) log plot indicates a constant growth rate.

Trend Line:

To add a trend line in a line plot, left click onto the data plot to highlight it, then right click and pick the trend line option. You can select linear or nonlinear options for the trend line. Note that an exponential trend line will be linear in a logarithmic plot.

Scatter Plots

For scatter plots of two variables, chose the scatter chart type rather than the line chart type. Note that you need to treat the two variables as the the X and Y components (values) of a single series (not as two separate series).

Absolute Column and Row References

If you want a column or row reference in an equation to be absolute (so that it does not change if you copy the equation elsewhere), place a \$ before it in the cell reference. E.g., if you want to divide the number in each cell in cell range C5:C40 by the number in cell F3, then you can use the equation =C5/\$F\$3 for the first number and then copy this formula down for the other 35 numbers in the series.

Sorting Data

You can sort data in a spreadsheet by using the Data/Sort menu item. However, be sure to sort the entire data set, not just the one column that you want the data sorted by. E.g., if you have country names in column 1 and GDP for each country in column 2, you can reorder the data by e.g., listing countries alphabetically (sort by column 1) or listing by size of GDP (sort by column 2).

Filters

Recall item **10** above, where you found all the recession years. You could do this fairly easily by visual inspection. A fancier way to do this (and a very useful shortcut when working with larger data sets) is to add a “filter” to the growth rate column (E) by clicking on the column header, then click on “Filter” under the “Data” menu. You will see a little menu tag appear at the top of the E column. You can use this to set a filter – e.g., show only rows for which the value in E (growth rate) is less than zero. Note that you can also use the filter to sort the data, as above, for example. Make sure to clear the filter after you use it.

Cell Formats:

Excel can display data in different ways. For example, if you want a series to be displayed with 3 decimal places showing, highlight the series, then go to “Number” in the tool bar above the spreadsheet and click on the button indicated with a right arrow. This does not affect the precision with which Excel stores and manipulates the data, only its appearance.

More generally, to alter the format of the content of a range of cells, highlight the range, right click on the highlighted area, and select “format cells.” For example, you could have adjusted the number of decimal places displayed above by selecting “format cells,” choosing “number” in the “number” tab, and then changing the number of decimal places displayed.

Note that if Excel displays only pound signs (#####) in a cell after you make a calculation, the cell was not wide enough for the answer to fit. You need to widen the column (double click or click and drag between the columns at the top of the sheet).

Freeze Panes:

You may notice that as you scroll down or across the worksheet you have been using, the dates and variable labels scroll out of sight. You can keep these in sight at all times by using View/Freeze Panes. In the worksheets for this exercise you should first left click on cell C3 and then click on select View/Freeze Panes. Now scroll around in the worksheet and see the effect. Note that I suggested cell C3 as Freeze Panes locks in rows above and columns to the left of the cell selected. To turn off the feature, select View/Freeze Panes again.

Printing:

To print a section of a spreadsheet, use Page Layout/Print Area/Set Print Area after highlighting the area that you want to print. To highlight more than one area on the spreadsheet, highlight the first area in the usual way and then hold down the <Ctrl> key while highlighting additional areas.

You can preview what will print with “print preview” (from the Office button at the top left of the window). You can also squeeze a large print area onto one sheet by selecting Page

Layout/Scale To Fit. Similarly, you can add gridlines, and make other formatting decisions under the Page Layout tab.

Paste Special

With “paste special” you can copy and paste values (rather than formulas), transpose data, etc. Note that “paste” will copy and paste formulas, so if you want to copy and paste just values, use “paste special.”

Headers for Take Home Exercises:

For the take home exercises, you will be asked to print out your work in Excel. Please make the effort to format your work in a way that will make it accessible to the reader. You will also be asked to embed your name in the header of each worksheet. To do this, select Insert/Headers and Footers and type your name in the left header panel (or select Print Layout/Page-Setup/Header-Footer/Custom-Header and type in your name in the “left section”). You will need to do this for each worksheet that you print.

Other Useful Stuff You May Want to Learn About at a Later Date

VLOOKUP – operator for pulling data from one column based on values in another column

Pivot Table – for cross tabulation (under “Data” menu)

IF – operator for conditional if/then/else statements

Reading and writing non-Excel formatted data files (e.g., .csv or .txt files)

Keyboard shortcuts for saving time and effort

Macros – for automation of tasks