**CORE121 Sections 1 and 11   
Fall 2021  
Prof. McDonald  
Excel Lab #1**

This is a beginner exercise for learning some of the most commonly used formulas and functions in Excel. The exercise starts with an Excel spreadsheet called “Excel Lab 1”.

Download the data from here: <https://bit.ly/3iZtNbV>

**Part 1: Presidential Turnout**

The first worksheet called **Presidential turnout**, has one record for each US State (rows), with three columns of numbers. The source for this data covers ten election cycles, and here you will two of them: 2014 and 2018.

Our goal is to look for some patterns and trends in this information.

Let’s start by thinking of some questions we might want to know from this data. Some possibilities:

* Which state had the greatest increase in its turnout?
* Which state had the greatest decrease?
* How has support for Republicans and Democrats varied?
* Which county had the highest total turnout?

Can we answer all of these questions with this particular data? **Briefly discuss in your groups which of the above questions you will be able to answer.**

Answer: This worksheet only shows the total turnout for each state; we don’t have any details on specific parties or counties so we can’t answer the last two questions. To get those we would need data that shows one record for each candidate or county that includes votes received by those units.

Now, we need to translate those questions (the questions we *can* answer) into syntax that Excel will understand.

**Which state had the greatest increase in turnout?**

This can be a tricky one, and the language in the question isn’t precise. Of course, we can take the 2018 total and subtract the 2014 total to see which state had the largest increase in number of votes or raw percentages. But is this a fair way to compare states? Does the change relative to population size matter?  
  
To give you some perspective, in the U.S. one state (California) comprises approximately 12% of the entire population. The state with the fewest people, Wyoming, has about 0.175%. Nearly one in eight Americans lives in California, and only 1 of every 570 Americans lives in Wyoming.

So it would be surprising if California’s turnout increase did not exceed Wyoming, even if there are consistent trends or trends that contradict that observation.

This is a situation where we are trying to compare entities of different sizes and we need to put them on a level playing field first. Instead of using the turnout change, we can use a percentage change – that will tell us the relative size of the change in comparison to the others. We will do this in two ways.

First, we will calculate raw growth rates. Second, we will calculate turnout as a percentage of the population, and then calculate the percentage change from that. Discuss the merits of these two strategies in your group once calculations are complete.

We can calculate these quantities using formulas in Excel. Spreadsheet programs like Excel use two dimensional grids as their visual metaphor. In the early days of computing, all this kind of work involved interaction with a command prompt, which you will see when work with RStudio. Around 1980, someone invented a commercially successful program that evolved into the modern spreadsheet. Along with new word processing programs and database programs, spreadsheets effectively launched the modern personal computer industry.[[1]](#footnote-1)

You enter content (formulas, numbers or text) into one or more of the individual cells on a spreadsheet. Each cell has a row and a column designation, and the organization of cells makes perfect sense the first time you see it.

A formula in Excel starts with an equal sign and references the cell addresses (e.g. A1) of data and text to perform calculations. Notice that a worksheet in Excel use letters to label columns and numbers to label rows. If you click on a cell and it will highlight the letter and number that comprise its address. For example, in this worksheet, the header the word “state” is located in cell “A1”. That label names the content of the cells below it, which you can see are the names of U.S. states.

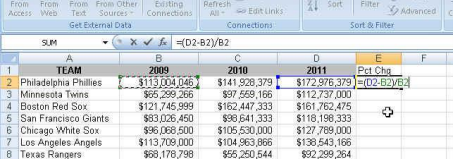
As we have learned, a percentage change is calculated by taking the new number minus the old number and dividing by the old number. If we designate “N” to mean new and “O” to mean old, the formula is:

=(N-O)/O (think of “Noo!” as a way to remember this).

The parentheses are very important, and if you don’t enter them, you’ll get the wrong result. Play with it when you enter the Alabama example. Why are parentheses necessary?

In our worksheet (labeled (Presidential Turnout) we can use D2 (the 2018 total voted number for Alabama), subtract B2 (the 2014 Alabama number) and divide that difference by B2. You type the formula in cell F2, which should be the first row of data below the headers. Add a label to this column called “raw pct change”. [[2]](#footnote-2)

Here is an example of such an analogous operation from a different dataset, which in this case shows major league baseball salary totals by team.





After typing in your spreadsheet, hit Enter. Then click on cross in the lower right corner of cell E2, and the formula will copy down to every row that has data in the adjacent column.

You can improve the readability of these cells by choosing an appropriate format. We’re working with a percentage. Highlight the column, then right-mouse click and choose “Format Cells” to change the formatting to percentage.

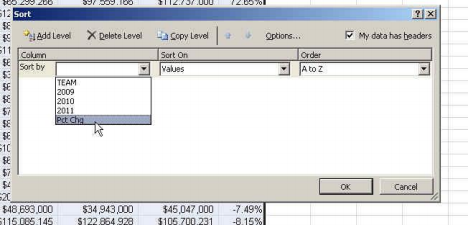


Let’s reorder the rows so we can easily see the state with the biggest percentage that we just computed.

The easiest way[[3]](#footnote-3) is to

* Make sure your cursor in somewhere in the grid with data in it.
* Select the “Data” menu on the menu bar.
* Click on the “Sort” button.

A dialog box pops up. In the list that shows the names of different columns, pick the one we just generated (“raw pct change”), indicate you want to see the order from Largest to Smallest, and that you are sorting on values.  
  
Here’s an example from the baseball dataset. Notice that the labeling might change a bit from version to version of Excel but the basic template remains the same.



Hit OK and your list should change, and you should see one state far ahead of the rest.

Now create two new columns that present the percentage of the population that voted in each state for each year (“voted\_pct\_population”). This variable will occupy column “G” and “H” for 2014 and 2018 respectively.

The results understate the percentage that we usually associate with turnout. Why is that?

Construct another variable in column “I” called “pct pop change” that subtracts the 2014 percentage from the 2018 percentage. Sort this list from largest to smallest. The same state continues to top the list, but the percentage change is much smaller in magnitude. **Discuss in your group the difference between these two percent change calculations. Which do you think should be reported in newspaper coverage?**

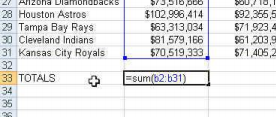
,

The next question we want to ask is “What is the total turnout in the United States for each year?” To do that we need to add up all the states for each year.

Much of the time you can use Excel with simple operators as we have done so far, such as add, subract, multiple or divide. Excel can do much more sophisticated with functions. To add along sequence of data, SUM that allows us to add up all the values that are in a range of cells. In this case, the 2014 turnout is in cells B2 through B52. The 2018 turnout is in the range from D2 to D52, etc.

We’ll put these totals on a row below our table. So go to cell A54 (leaving an empty row between our data table and these new totals) and put the word “Totals” in cell A54. This will be our label to remind us what this is.

Then in cell B54, we’ll put our formula. For example: =SUM(B2:B52)



Hit enter and then get the copy tool and drag the formula across to columns C and D. If you see ############# populate a column, that just means that the columns are not wide enough to display the entire number. To make them wider, go up to the letters at the top. You can either drag the line between the C and the D or double click on it to get it wider. Then do the same for the line between the D and E columns.

To answer the next question --- how has this changed over time – we can calculate percentage changes using the same two methods as before.

An easy way to get the percentage change is to dr ag the 

percentage change formula down from cell F52 so that it lands

in cell F53. You’ll end up with a Divide-By-Zero error message

on the blank row (which you can just delete). Repeat to calculate totals for all the new columns you created earlier.

Next question: What’s the average state turnout? Here we will focus ONLY on turnout as a percentage of the population in 2018.

Again, Excel has a function for this (two, actually). When you’re dealing with dollar amounts – especially pay or housing values or things like that – you will want to look at both the AVERAGE and the MEDIAN. The average is calculated by adding all the values together and dividing by the number of records. The Median is the value that would be in the middle if you sorted all the payrolls and ordered them smallest to largest. An average can be heavily skewed by one or two “outliers” – numbers that are significantly bigger or smaller than all the other values. A median will not be skewed. When there are outliers, the median will more likely represent the “typical” value.

The rule of thumb is to calculate both and if they are close to each other, it’s safe to use either one. But if the average is significantly higher or lower than the median, it’s best to use the median (if you’re attempting to tell your readers/viewers what the “typical” is).

To calculate these in Excel, we’re going to use functions just like we did when summing our data.

Start a new line below the totals and put “Averages” in the A55 cell and “Medians” in the A56 cell. Then put these formulas in the B55 and B56 cell, respectively. Then copy them across to the other columns in your worksheet.

=AVERAGE(B2:B52)

=MEDIAN(B2:B52)

Example from other data:



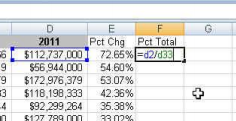
**Are the averages much different than the medians? Discuss this in your group.**

Let’s do one last thing with this data. Let’s figure out just how much bigger California is than the `other states. There are a couple things we can do.

The first is a “percent of total.” We can calculate what percentage of the total turnout (what we calculated on row 54) each state accounts for.

To do this, let’s go to column J and put a header – “Pct Total” and then start a new formula in cell F2.

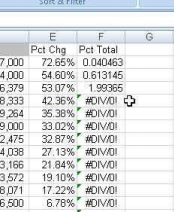
A percent of total is calculated by taking the

state number and dividing it by the grand 

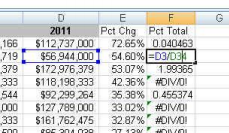
Total (use 2018 raw data, column D).

After typing in this formula, hit enter and

copy it down for the other states.

Uh oh! We’ve got a problem. 

We can trouble-shoot by clicking on the answer in J3. If you double-click on it, it will highlight the cells that it is using to do the math and it will display the formula is the formula bar just above your data.



What’s wrong? We can see that the first cell address in the formula is correct. But what about the second address? It’s grabbing the average we calculated for 2018 – not the total. If you go down to the ones that are showing a divide-by-zero error, you’ll see that it’s trying to get data from a cell that is empty.

We need to adjust our formula so that the divisor (the second cell address) is ALWAYS on the total that is located in D54 (the Total).

Excel allows us to do that with what’s called an “anchor”

To anchor a cell, we put a dollar sign ($) in front of the letter and/or the number of the cell address. To lock an address completely (regardless of whether you copy your formula down or to the right) you would use two anchors, like this:

=D2/$D$54

So go back to the state on line 2 – and edit your formula (easiest way is to go up to the formula bar and edit there), adding in two dollar signs. Then copy the formula down again. This time you should have answers! (You may need to highlight the column, right-mouse click and choose “Format Cells” to change it to a percentage). What percent of total turnout comes from California?

Finally, another question we could answer is “How does each state’s voter turnout rates compare to the national average?” Here we will take the difference between each state’s 2018 turnout relative to the national average.

Start a new column – in K – and call it “Pct Diff US Ave”. Take the 2018 Total US turnout as the percent of Total US population-- don’t forget to use anchors -- and subtract the state 2018 turnout as a percentage of the population (do this in row 2 of your worksheet).

This time, to copy down this formula cells below, let's use keyboard shortcuts instead of the dragging method. Click on your formula and then hold down the shift key while simultaneously clicking on the downward arrow button on your keyboard. Repeat until all relevant cells are highlighted. You can then execute the shortcut Command-d in order to “fill down” to all those highlighted cells.

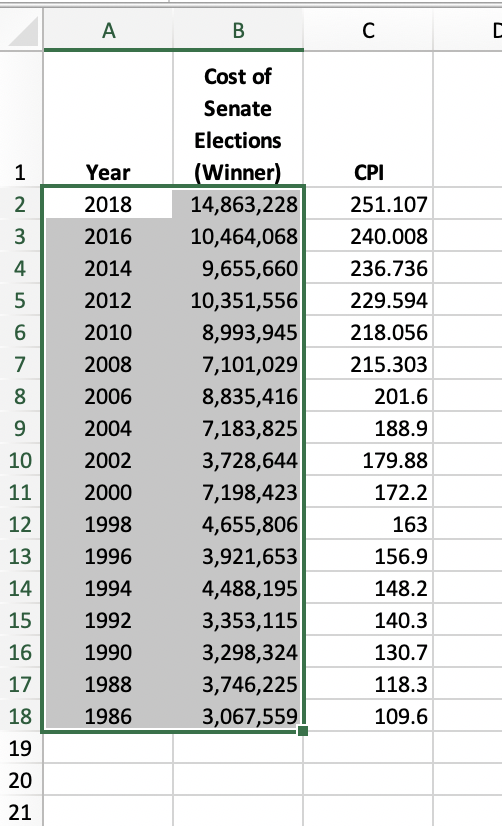
Another quick shortcut to practice is a quick way to highlight all cells below or all cells above. Click on any cell in your workbook and execute the short-cut “command-shift-arrow up”. You will see all the cells above your cell automatically highlight. Now try command-shift-arrow down”. All the cells below highlight, though sometimes it highlights down to the very bottom of the workbook, encompassing blank cells. For this reason if you want to execute this shortcut is it better to go to the last cell and highlight up (and then “fill-down”).

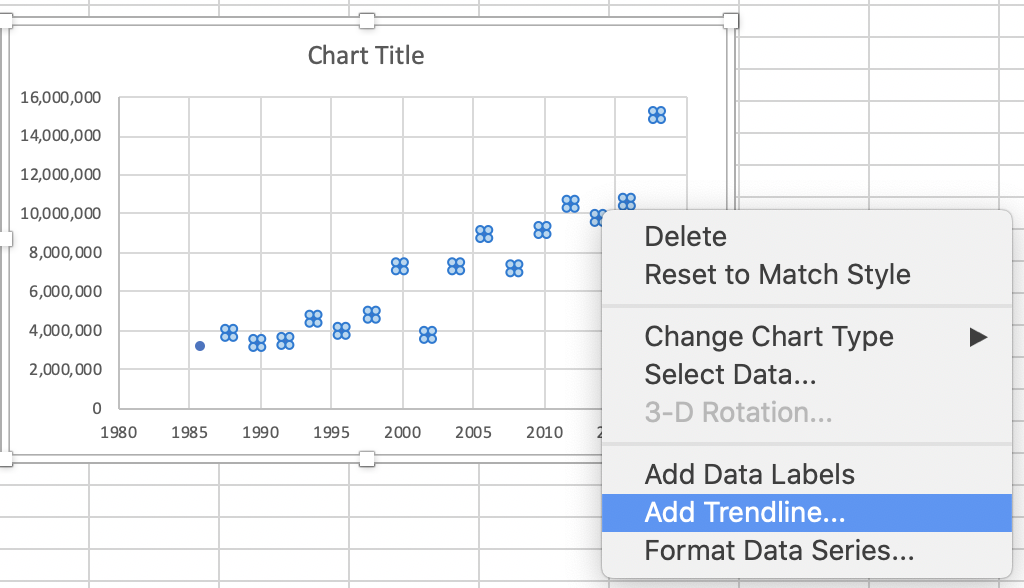
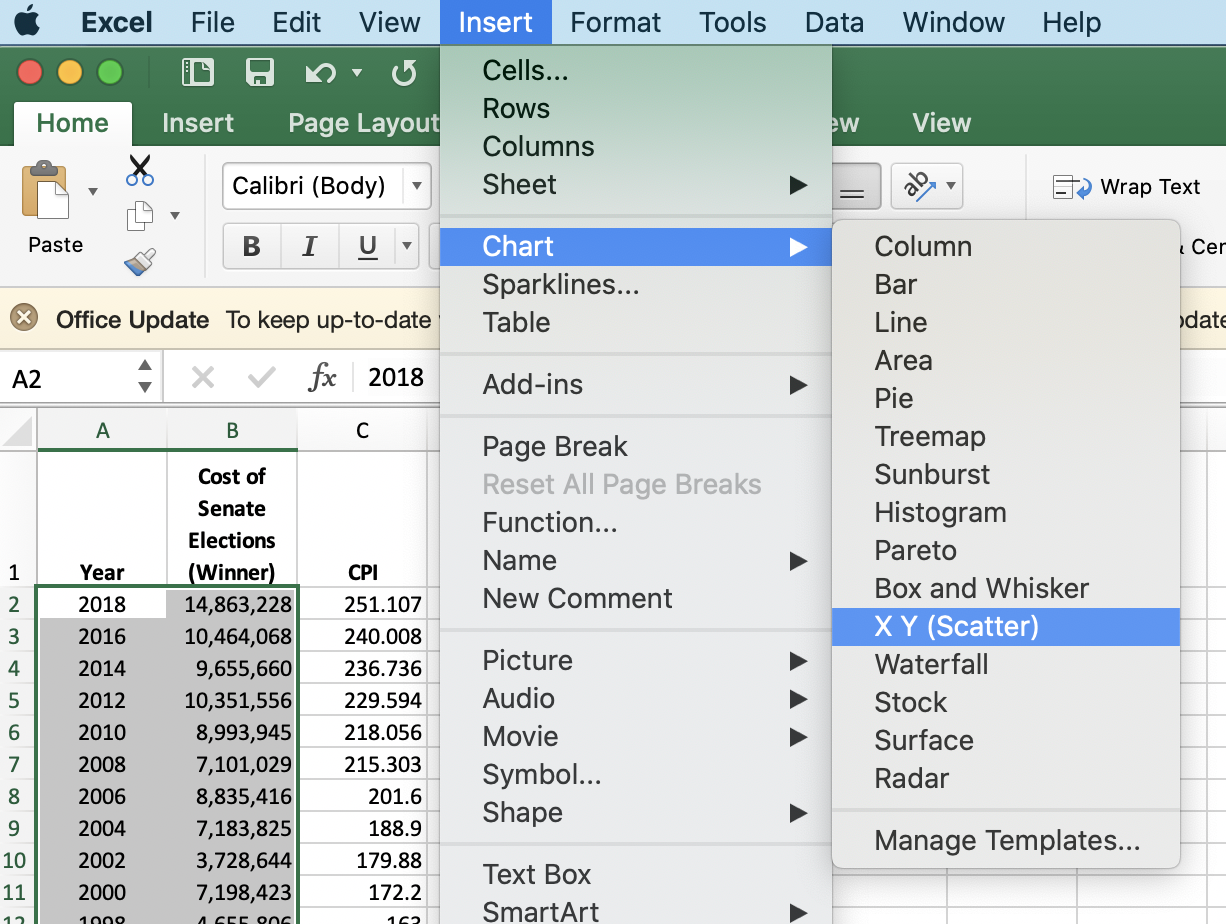
[Slack me with a time-check now so I can make sure it is reasonable to have both Part 1 and 2 due Monday, which is the current plan]

Save you worksheet and upload it here: <https://classroom.google.com/c/MjU4MTgxMDAyODk5/a/MjU4MTgxMDAyOTU3/details>

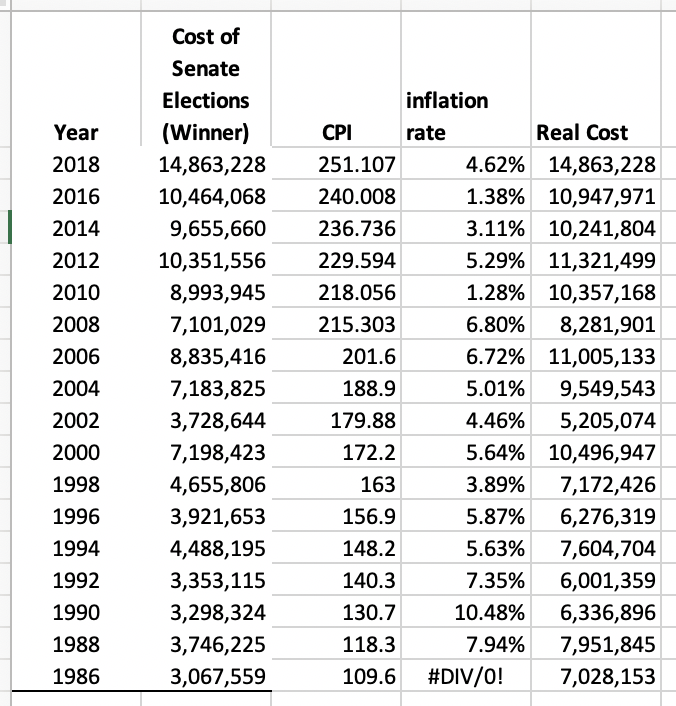
**PART 2: Campaign Finance Inflation**

Worksheet 2 is called “Campaign Finance”. This worksheet contains information over time, with each row representing a general election year. The second column represents the average cost of winning a Senate election, in nominal dollars (NOT real). The third column contains CPI.

1. Highlight the data that appears in Column A and Column B (Drag cursor over data). 
2. Go to the Insert menu at the top of the application and go to Insert > Chart. Choose the “X Y Scatterplot option”. After the graph appears, right-click on a data point (a dot). Choose the option to “Add Trendline…”. Play around with the options and discuss in your group whether the data is better fit by an exponential or linear growth model. You can delete this chart when you are done.



1. In Column D, create a variable called “inflation rate” that captures the percent change in CPI. Just like in Part 1 of this lab, this task is completed by creating a percent change (N-O)/O and filling down the column. Your first cell will be empty because there is no previous data to subtract from. Highlight the column and right-click, selecting the option “Format Cells”. Format to percentages.
2. In Column E, create a variable called “Real Cost” that uses the CPI column to recalculate the cost of winning a campaign in 2018 dollars. To do this, first divide by the CPI in each year [e.g. each row will refer to a different year], and then multiply by the 2018 CPI [anchor this value]. Fill down the entire column. Format this column to include separators for 1,000s (option available under number). At this point your data should look like this:



1. Highlight ONLY the data in column A and Column E by dragging your cursor over that data while holding down the command button. Again, go to Insert > Chart and create a new scatter plot with the Real data. Add a trendline. **What model do you think best fits the data now?**
2. Within the trendline option menu, click on the “display equation” option. **Use this equation to calculate the expected cost of winning a Senate election in 2024 (in 2018 dollars).**
3. **If inflation is 2%, how much will the candidate have to raise in 2024 dollars?**

Save your Excel worksheet and upload it here:

<https://classroom.google.com/c/MjU4MTgxMDAyODk5/a/MjYyOTk4ODIwMTQx/details>

For the remainder of class-time we will practice inflation adjustments. The answers for the questions asked are provided at the end of this document and you are not required to turn in your work for this practice session.

Use the CPI table below for your calculations in questions 1-5.

When working with CPI, all of the following formulas are applicable -- indeed they are the exact same thing, just rearranged terms!!! I find that different formulas “click” for different students (and some students don’t like thinking in terms of formulas at all, which is FINE -- you can also just think about CPI as “Ratios” to use for conversion!!

* Value in X Dollars = Value in Y Dollars \* (CPI-X / CPI-Y)
* Past dollars in terms of recent dollars = Dollar amount ÷ Beginning-period CPI × Ending-period CPI.
* More recent dollars in terms of past dollars = Dollar Amount ÷ Ending-period CPI.× Beginning-period CPI

1. What was the inflation rate between 2019 and 2020? [Calculate Percent Change]
2. What is $1,000 today worth in 1982 dollars?
3. How much was $1,000 from 1982 (nominal) worth in today’s money?
4. How much was $1,000 from 2000 (nominal) worth in today’s money?
5. Contributions to federal candidates to political office are limited by federal campaign finance laws. Specifically, in 2020 a candidate for federal office could only receive $2,800 from an individual. These limits have been indexed by inflation. Using the CPI table, estimate what the individual contribution limit was in 2002 (the year when these limits were last altered by law).
6. A friend says he will either give you $100 now, or $125 in 5 years. If inflation the inflation rate over those five years is 3%, which is the better deal? [Note: Since inflation here is constant, the easiest way to do this problem is the explicit exponential growth formula!]

*Note: This table uses 1982 as its baseline [1982 CPI = 100].*

|  |  |
| --- | --- |
| **Year** | **CPI** |
| 2020 | 259 |
| 2019 | 256 |
| 2018 | 251 |
| 2017 | 245 |
| 2016 | 240 |
| 2015 | 237 |
| 2014 | 237 |
| 2013 | 233 |
| 2012 | 230 |
| 2011 | 225 |
| 2010 | 218 |
| 2009 | 215 |
| 2008 | 215 |
| 2007 | 207 |
| 2006 | 202 |
| 2005 | 195 |
| 2004 | 189 |
| 2003 | 184 |
| 2002 | 180 |
| 2001 | 177 |
| 2000 | 172 |

*Practice Answers*

1. What was the inflation rate between 2019 and 2020? [Calculate Percent Change]
   1. (259-256)/256 = 1.17%
2. What is $1,000 today worth in 1982 dollars?
   1. 1000/259\*100 = $336.10
3. How much was $1,000 from 1982 (nominal) worth in today’s money?
   1. Past dollars in terms of recent dollars = Dollar amount ÷ Beginning-period CPI × Ending-period CPI.
   2. 1000/100\*259 = $2,590
4. How much was $1,000 from 2000 (nominal) worth in today’s money?
   1. 1000/172\*259 = $1505.81
5. Contributions to federal candidates to political office are limited by federal campaign finance laws. Specifically, in 2020 a candidate for federal office could only receive $2,800 from an individual. These limits have been indexed by inflation. Using the CPI table, estimate what the individual contribution limit was in 2002 (the year when these limits were last altered by law).
   1. 2800/259\*180 = $1945 [In practice, contribution limits are rounded to nearest 100 dollars, the limit put in place in 2002 was for $2,000]
   2. More recent dollars in terms of past dollars = Dollar Amount ÷ Ending-period CPI.× Beginning-period CPI
6. A friend says he will either give you $100 now, or $125 in 5 years. If inflation the inflation rate over those five years is 3%, which is the better deal? [Note: Since inflation here is constant, the easiest way to do this problem is the explicit exponential growth formula!]
   1. 100\*1.03^5 = $115.92. It is better to take the $125 (unless of course you are investing your money wisely!)

1. In other words, spreadsheets are the reason large numbers of businesses and consumers bought their first personal computer starting in the 1980’s. Semiconductors and especially the design of a tiny central processing unit on a silicon chip, were necessary pieces of this revolution, which preceded widespread presence of the internet by 15-20 years. [↑](#footnote-ref-1)
2. There are conventions and style guides for naming columns like this, but for now, just make it succinct and descriptive. [↑](#footnote-ref-2)
3. But not the only way. Get used to fact that almost everything you can do with a computer (or an application like Excel) is something you can do in many different ways. Often the right way is the first try that happens to work. [↑](#footnote-ref-3)