Lab 1: Introduction to Excel

The goal of today's class is to get more comfortable working with Excel, manipulating data, and getting familiar with housing and demographic data from the American Community Survey. There are other proprietary statistical packages (e.g. SPSS, STATA) and Python libraries (e.g. Pandas) that can do the same things (and sometimes more quickly); still, Excel remains a staple of most planning offices. Excel is also often used for Accounting and Budgeting, so becoming familiar with looking at and manipulating data in Excel is an important professional planning skill. In addition, it can be useful to provide or present information in Excel if that better matches your audiences'/coworkers' skill level.

We will cover the following topics to help you get from the spreadsheet to these outputs:

1. Navigating the Environment
2. Basic Functions
3. Summary statistics
4. Charts
5. Formatting issues

Please download the file labeled "Excel_Lab_1" from bCourses and open the file in Excel. Note that the data in the first two sheets have been sourced from the U.S. Census Bureau website. We'll learn more about how to use census.data.gov later in our lab next week. The data in the first sheet is from Summary Tables *DP04: SELECTED HOUSING CHARACTERISTICS* and *DP05: ACS DEMOGRAPHIC AND HOUSING ESTIMATES*, drawing from 2014-2018 American Community Survey 5-year Estimates. The data in the second sheet comes from Summary Tables *B03002: POPULATION DEMOGRAPHICS*, drawing from the 2017 American Community Survey 5-year Estimates.

Part 1: Understanding Excel's Environment (15 mins)

Let's start by playing around with the "Lab_1_Data" worksheet.

- By positioning the cursor on the division between the header of columns A and B you can left-click and drag the cursor to the right to make column A wider.
- Now wrap the text within a cell to be able to see the whole variable description.
- Let's now sort the data so you can see which census tract in San Francisco has the largest population
- Scroll to the bottom of the worksheet. Darn – now you can't see the variable labels along the top of the worksheet! Select cell A2, go to View (Mac: Window) – "Freeze Panes" and "Freeze" the top row.
- You can also freeze the left-most column by selecting cell B1 and pressing "Freeze panes". To select both your first row and your left-most column, so that you can always see the labels associated with your data, select cell B2 and "Freeze panes." This will work by freezing the row above your cell selection (Row 1, in this case) and the column to the left of your cell selection (Column A, in this case).

Just as with Word, you can cut, copy, paste individual data, or a whole column or row (or multiple columns and rows at once). If you have the hang of it, think about using a shortcut for selecting columns and rows (PC: ctrl + arrow direction; Mac: shift + cmd + arrow direction).

Part 2. Basic Functions (15 mins)

Let's start analyzing our data. Make sure you are on the sheet that is labeled "Lab_1_Data." Your first goal is to calculate the total population, the total number of housing units, and the total number of rental and owner units in San Francisco.

First, start by scrolling to the bottom of the worksheet, and adding a new row labeled *County Totals* on cell B200. Freeze the top panes as needed. Now let's input a formula that calculates the total population for values in a selected column. Go to the "Formulas" menu and select "AutoSum". Press Enter. Voilá!

Another alternative is to write the formula yourself. Write in cell C200 the following expression *=SUM(B3:B199)* and press enter. It tells Excel to Sum all the cells in column B, starting with the cell in row 3 and going all the way to row 199. Now let's delete cell C200 to have a cleaner table.

Second, calculate the totals for the other columns. Try dragging the little black square in the first "total" cell over to the right.

See that Excel tries to guess what you want to copy or duplicate, which is a great shortcut, but you have to be careful to check that its guess is correct. If you ever want to "freeze" either a column or row reference, you need to put a dollar sign in front. Try typing *=SUM($B$3:$B$199)* in the cell and dragging that over – it should stay exactly the same.

Third, let's calculate the percent renters in San Francisco. To do so type the following expression: =(F199/C199)*100. See that we are dividing Total Rental Units by Total Occupied Units and then multiply that by 100.

Please click your cell, which should note the cell type as "General." If you change the type to "Percentage," what happens? To fix that, simply delete the *100 from the expression =(F199/C199)*100.

Part 3.1: Summary Statistics Table: Mean, Maximum, and Minimum Values (15 mins)

Let's create a new worksheet called "Summary_Table_A" and then cut and paste the totals calculated in Part 2 into the new worksheet to start building a summary table. What went wrong?

It is important to remember the difference between actual number values and "instructions" that you give Excel on what to put in a cell. When you copy the instructions, the "reference" disappears, giving you the dreaded "#REF!".

One solution is to copy and paste the number as "values" ("Paste Special" > "Text"). But perhaps a better alternative is to copy the cells and then select Paste Special and then Paste Link. With this alternative,

you are referencing the cells in the Lab_1_Data worksheet that has the information.

Now use the AVERAGE() formula as we did with the SUM() formula to obtain the average for each Column in our Lab_1_Data worksheet. Try also using the MAX() and MIN() formulas to obtain the maximum and minimum values for each variable.

Part 3.2: Summary Statistics Table – Proportions and Format (15 mins)

Now your task is to generate a table with summary statistics of demographic data in Alameda using the data provided in the sheet "ACS_B03002".

Start by creating a new worksheet called "Summary_Table_B". Then, calculate the total population by each race/ethnicity group using the =SUM() formula. Once you have the totals, calculate the percentages for each population group as the total per group divided by the total population in Alameda.

Now we need to format the Table, so it is easy to digest and has all the information needed

1. Let's change the font and add number format
2. Now insert a row above the column header and add a table title
3. **Important!** You need to carefully note the source of your data and make clear processes. Failing to do so is an easy way to lose track of your work and lose credibility as a researcher.

Figure 1: Alameda County 2017 Population Demographics

	White (NH)	Black (NH)	Native American (NH)	Asian (NH)	Native Hawaiian or Pacific Islander (NH)	Other (NH)	Two or More Races (NH)	Hispanic or Latino
Population	524,881	175,063	5,008	468,356	13,000	4,489	71,777	367,041
Percent of Population	32%	11%	0%	29%	1%	0%	4%	23%

Source: 2017 ACS Five Year Estimates, Table B03002
Notes: Total popualation of Alameda County, n = 1,629,615.
Percentages are rounded to the nearest whole number, resulting in some rates being reported as 0 percent.

Part 4.1 Chart Generation: Column/Bar Chart *(15 min)*

Now that we have our table, let's put it in a chart. There are many chart options – we will talk about them later in the course. For now, we are going to create two charts showing the demographics of Alameda County using "Recommended Chart". Treat it like an assistant in training. It's still learning, so it's good to help it out by putting your data into a simple format.

- Let's start by generating a column chart showing population statistics. Highlight your table, click Insert, and choose a column chart.
- Notice: there are a bunch of options for charts - types, types of types, etc. Which one do you think works the best? Should the percentage of the population be included on the same table?
- You can remove data by right-clicking on the chart and selecting "select data" and removing the percentage data. Note, once you get more familiar with Excel it will be easier to create effective charts using "Select Data" Play around a bit in this window.
- Excel will recommend colors and you can try on your own to select good colors by selecting the 'Change Colors" icon and then choosing the color you want. Alternatively, you can click on one of the bars on your chart and then go to 'Home' and select a color fill. That will change the color of all bars at once. You can also change the color of only one bar by double-clicking on any bar and choosing the desired color. Not sure what's the best combination? Go to http://colorbrewer2.org/ for some suggestions

Part 4.1 Chart Generation: Pie Chart (10 min)

- Now that we have walked through creating a bar chart can you create a pie chart? It is a very similar process.

Part 5: Two Common Issues (10 minutes)

Click on the Common Issues worksheet.

1. *Summarizing numbers* – Example 1 has two sets of numbers and a calculated total. Uh-oh, the totals are not the same. Why do you think that is? When you are adding numbers, make sure you understand what you are really adding.

2. *Format Issues* – In Excel, you will sometimes run into problems where your functions don't work. This issue very frequently is related to how your data is formatted. If a cell is recorded as "Text" it will not be included when you conduct functions. What cell is the problem in Example 2?

DIY: VLOOKUP -- aka your future best friend in Excel (30 minutes)

Let's now say that your boss has notified you that she is going to go to a series of meetings around the county and needs the same demographic data about the neighborhoods where she is presenting. She gave you an old list of four neighborhoods in Oakland and Berkeley with the Census tracts that make up the neighborhoods. You can either manually identify the neighborhoods or... you could use the VLOOKUP function and be done super fast.

1. Go to the "Neighborhood_List" worksheet.
2. Insert a column between columns A and B in the B03002_Clean_Data sheet (by right-clicking column B and selecting "Insert") and give the column the title, "Neighborhood."

3. Now we use the function! Click in the first cell below the header and type in “=VLOOKUP” The function asks for four pieces of information: lookup_value, table_array, index_column, and whether you need an exact match.

    - lookup_value is the reference information we will use to add the new information;
    - table_array is the table of information that we will be using to pull in new information;
    - index_column tells excel what column the new information is in (Note: the left-most column in your table of new information is column 1, then 2, etc. moving to the right); and
    - exact match (FALSE) allows us to confirm whether we need an exact match.

Now that you have added the neighborhood information, develop a table that provides demographic information by neighborhood and for the County as a whole. Remember that each time you use these functions the more natural they become.