

Lab 3: Excel Basics—Formulas and Generating Charts

Objectives

There are three basic goals for today's lab. We will be covering:

- Table Generation;
- Picking Colors; and
- Using the VLOOKUP function

Table Generation

Let's start by using our American FactFinder skills to download some data. We want to look at some demographic information for all of the census tracts in Alameda County - specifically, a racial breakdown of the population, categorized by whether respondents identify as Hispanic or Latino. So we want to find a table that summarizes data on Hispanic and Latino origin, and further, racial identification. First, then, we should refine our Advanced Search in American FactFinder by selecting our geography (Census tracts within Alameda County). Then, we should probably proceed by topic (Which one? Were you paying attention last week?). Then find a table that summarizes the data we're looking for: a report on people in Alameda County's Hispanic or Latino origin status and their race. Which table did you find for this?

Once you identify the correct table to use, examine the table in Table Viewer. Does it look how you want? Does it look a little too "wide" when you might want it "long"? This can be corrected in American FactFinder. Open the table in "Table Viewer." Select "Modify Table," and then "Transpose Rows/Columns." Now download that and open the .csv. I suggest you start by saving the workbook as an Excel Workbook (.xls), so that you can work with multiple sheets (which we'll want to do in a bit).

Now we want to get working with the data, so we'll probably want to clean it up for more efficient use. Right now, we just need a table including our Geography (e.g. "Census Tract 4001, Alameda County, California"), FIPS code (e.g. 6001400100), and estimates of Total Population, the non-Hispanic population that reported their race as "White Alone" (call this, "White alone (NH)"), "Black or African American alone" ("Black (NH)"), "Asian alone" ("Asian (NH)"), "Two or more races" ("Two or More Races (NH)"), "American Indian and Alaska Native alone" ("Native American (NH)"), "Native Hawaiian and Other Pacific Islander alone" ("Native Hawaiian or Pacific Islander (NH)" or "Some other race alone" ("Other (NH)"). We also want a total estimate of the population that reported being of Hispanic or Latino origin for each Census Tract ("Hispanic or Latino").

Since cleaning takes a while (but is an essential part of data analysis), we've cleaned the data for you. Open up the "Lab3data.xls" file. You'll see that the first tab is our raw data, transposed. The second tab, "Clean_Data", looks...well...cleaner. It always good practice to keep your raw

Census data as it appears when downloaded as a tab in your workbook. You never know when you'll need to draw from it again. This allows you to have a back-up, or sorts, if you run into any issues.

Your Job – Let's pretend that you are an entry level planner who has been asked to provide a demographic description for your boss with both total population and percent of total population. She is going to a county meeting and asked that you provide the information in a good visual representation. She would like an easy to read table for her records and either a bar chart or a pie chart that she can present at the meeting.

First – Excel generates visualization using data from tables. As a result, we need to start by creating a table with the information requested. What statistics do you need to generate? (Hint: we need to generate summary statistics for the county).

Now let's calculate the amounts together. Do your summary statistics match your neighbors? How about the numbers projected at the front of the class?

Second – Now that we have the necessary information, we need to put it into a good table. You can think about this as prettying up the table, but in reality what you are doing is making the information easily digestible. Your table should have all the information your boss needs, but nothing more. A great way to check if your tables are good is to ask a neighbor to explain what your table is showing in one sentence.

1. Create a new sheet. Then Paste in the column heading.
2. Copy the summary statistics from "Clean Data"
3. Format the table to make it easy to understand. We will go into more detail in class about best practices.
4. **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.

What does your table look like? Does it look something like this?

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 population of Alameda County, n = 1,629,615.

Percentages are rounded to the nearest whole number, resulting in some rates being reported as 0 percent.

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

- **Column/Bar Chart:** 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 percentage of population be included on the same table? You can remove data by right clicking on the table 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 Source.” Play around a bit in this window.
- **Pie Chart:** Now that we have walked through creating a bar chart can you create a pie chart? It is a very similar process.
- **TIP:** When making charts, it might be easier to make your data “long,” rather than “wide” - for similar reasons that we discussed earlier in the lab. To make your summary statistics long, you need to transpose them. You can “copy” your table, and the “Paste Special.” Check “Transpose” in the “Paste Special” window.

Selecting Colors

Right click on “Clean Data” and choose unhide. The tab shows the visuals that Abby created when she was practicing doing the lab. What issues do you see with her pie chart (hint: what is the title of this mini-section)?

Excel will recommend colors and you can try on your own to select good colors; however, I would strongly recommend using the website Color Brewer (<http://colorbrewer2.org/>) if you are thinking of presenting your tables (and later maps) to a wider audience. The website provides a

number of options for whatever kind of visual you are providing. You can then use the information to manually adjust the colors for your visual. It can take a little time, so I would recommend finalizing your visual first and then doing color adjustments as the last step.

VLOOKUP (aka your future best friend in Excel)

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 superfast.

1. Open the file "Neighborhood_List.xls"
2. Copy the tab "Future_Meeting_Sites" into the workbook we have been using throughout the lab. This can be done by dragging the tab into the workbook.
3. Insert a column between columns A and B in the Clean_Data sheet (by right clicking column B, and selecting "Insert") and give the column the title, "Neighborhood."
4. 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 allows us to confirm whether we need an exact match.

Let's fill this out together:

- lookup_value - _____
 - table_array - _____
 - index_column - _____
 - exact match - FALSE
5. Now that you have added the neighborhood information, can you work with your neighbor to 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.