

Lab 1: Profiling Boulder

Introduction

Is the People's Republic of Boulder really that different from the rest of Colorado? Aside from the abundance of vegan muffins, is the population, housing stock, and economy of Boulder different from the rest of the state? In this lab we will profile the City of Boulder by comparing it to the rest of Colorado. We will be working with data from the American Community Survey (ACS) that describes selected characteristics of all of the towns in Colorado that had a population greater than 20,000 people between 2006 and 2008.

Reading Survey Data

The American Community Survey is an annual survey of households conducted by the US Census Bureau. The ACS is a survey not a census, meaning that each year a small portion of the population is selected to participate. From this small sample, through the magic of statistics, the census bureau estimates the characteristics of the entire population in each town.

In this lab we are going to use data from the ACS. Our goal is to compare Boulder to the rest of Colorado by using the estimates of income, ethnicity, education, and housing characteristics and their associated margins of error (MOE). In the data we will use for this exercise each variable is described using two columns, an estimated value (EST) and a MOE. When you compare Boulder to the rest of Colorado be sure to look at the EST and the MOE.

The Data

The Excel File contains info on the population, income, commuting patterns, education, and housing for each city with a population greater than 20,000 people in the State of Colorado. The data is stored on tabs, if you look at the lower left corner of the excel page you'll see the tabs, click on them to change between each type of data (tabs are indicated by the red box in figure 1). This is a lot of information to synthesize, to effectively communicate the difference between Boulder and the rest of Colorado you are going to have to use the skills including functions, cell references, and filling. In addition, you are going to have to create graphs.

This is "real data" from the census bureau's most recent release of ACS data (2015). Sometimes, based on the title of the column it is difficult to figure out what the data are describing. I have included a document, called lab1variables.txt, this file should answer any questions you have about the data.

Some Excel Tips

Working with large tables can be cumbersome. One trick that I like to use is splitting a window into panes and/or freezing columns. On the view tab you can freeze the first column, the one that contains the city names so that when you scroll across the data you can still see the city

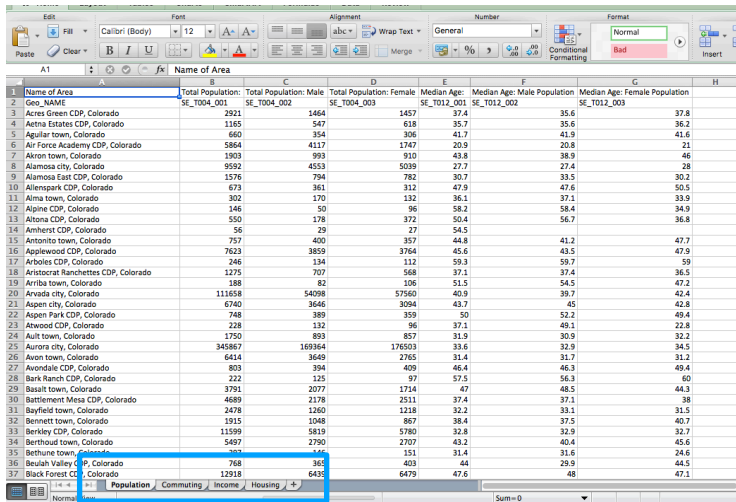


FIGURE 1: EXCEL TABS

names. Click the freeze panes buttons to access the freezing options. You can also split the window vertically or horizontally using a button on the upper right above the vertical scroll bar and the lower right below the scroll bar (indicated by the red arrows in Figure 2). The horizontal and vertical split buttons are not labeled, to use them click on the buttons and drag to create a new pane. You'll know you are on the right button when your cursor changes into a black symbol with two arrows pointing in opposite directions.

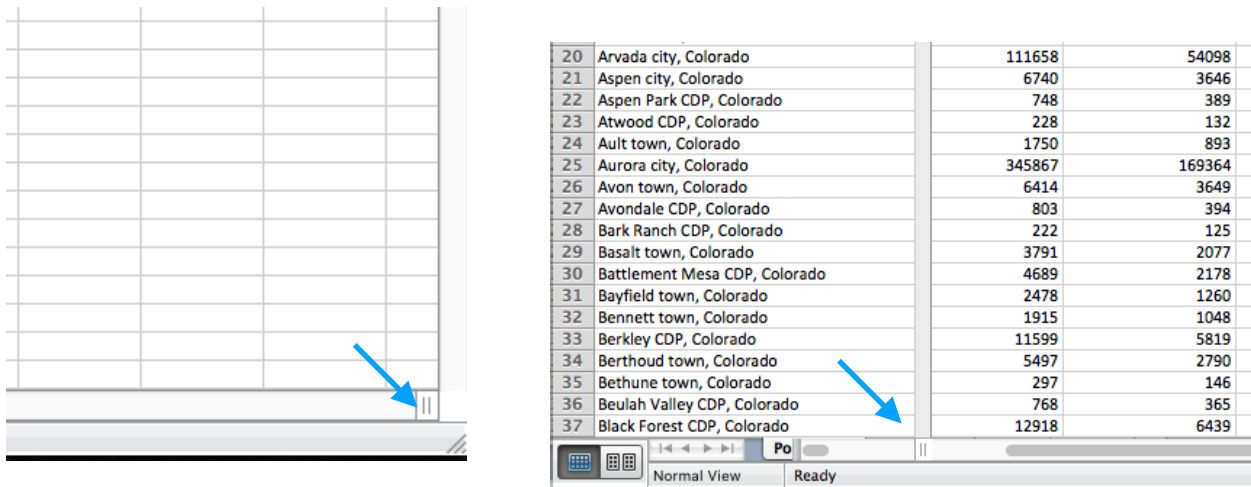


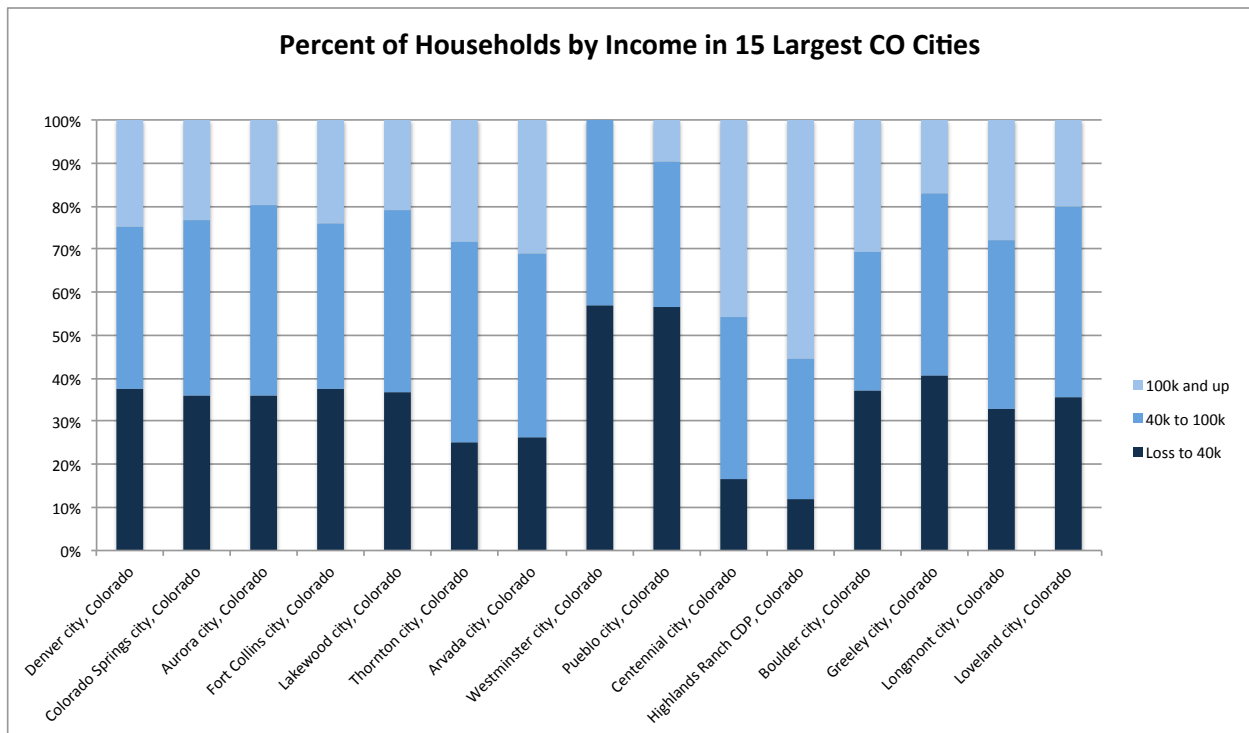
FIGURE 2: WINDOW SPLITTING BUTTONS

Questions

1. Use the data on the population tab to answer the following questions:
 - A. What is the largest city in Colorado and how does Boulder compare in size? How many Boulder's would it take to equal the population of the largest city?
 - B. What is median age in Boulder? How does Boulder compare to the rest of the state? Is there a gender difference? In the United States the median age of men is 35.5 years and of women 38.1 years (2010 est.). In general why is there a difference in the median ages of men and women?

2. Use the data on the commuting tab to answer the following questions:
 - A. What percent of commuters in Boulder travel by taxicab, motorcycle, bicycle, walked, or other means. How does this compare to the national average (4.6%)? Are most places in CO above or below the national average?
 - B. If you think Boulder is different from the rest of the state, can you think of any reason why commuters in Boulder might behave differently from the rest of Colorado? Consider variables such as the rates of driving alone, public transit use, and use of "taxi, bikes, and other."

FIGURE 3: GRAPHING INCOME CATEGORIES



3. Use the data in the income tab to answer the following questions:

- A. Create a graph like Figure 3 showing the income distribution in each of the 15 most populous cities in Colorado. All of the data you'll need to answer this question is on the income tab. Once you're happy with how it looks, copy your graph and save it.

To make this graph you'll have to create 3 or 4 income categories from the 8 original income categories. You can reduce the number of income categories by adding two columns together (using cell references and filling). Remember you are working with columns that tell you the estimated number of people in each income group not the percent of all workers in each income group. Compute the percent of workers in each income category for each city by dividing your new columns by the total number of workers. The Universe column is the one you should use as a denominator when computing percents, each column has its Universe identified in the first row of the table. Make sure the percents for each city sum to 100.

In figure 3 I grouped all of the households earning less than \$40,000, those that earned between \$40 and \$100,000, and those that earned more than \$100,000. To reiterate, you can create new categories by creating a new column that is a combination of existing columns. If columns D, E, F contained the categories you wanted to combine create a new column, with the expression in a new empty column, so for arguments sake lets say you want to use column T to hold the data. In cell T1 write = D1 + E1 + F1 and then use fill to populate the rest of column T with data.

To create a graph like figure 3 go to the insert tab. Select the new income columns you created, these should hold information about the 5 of workers in each city in each income category. In the spot where it says charts choose 100% stacked column from the list of column types. Your graph doesn't have to look exactly like mine, I customized my graph by playing with the graph layout and options. Generally, you change elements of the chart by right clicking on them and choosing properties. You don't have to mess with the formatting, but feel free to explore. You will be graded, in part, on the appearance of your graph. One easy way to change the appearance of your graph is to select the graph, click the design tab, and choose a new layout for the graph.

In excel, if you want to display data visually you just select the data that you want to graph and choose the graph type. It is important to make sure that your graph has appropriate labels on the axes and in the title, otherwise it will be hard for people to read your graph.

- B. The Median Household Income tells us the amount of money that half of the households earn less than and half of the households earn more than. Let's compare the Median Household Incomes to another measure of "typical" income, by looking at the difference between the median and the mean. Create a column that computes this difference. What would cause this difference to be positive, and what would cause it to be negative? Does the difference you observe relate to the patterns on the graph you just completed?

C. Next, let's look at the distribution of income within Boulder. What share of all earnings to the top 20% of households make in Boulder? How does this compare to the rest of the state? What share of all earnings do the lowest 20% of households make? How would you interpret this finding? What do you think causes this?

1. What is the universe for these columns? Do you think the top 20% of households and the bottom 20% of households represent the same number of people? What data would you need to answer this question? Explain your opinion.

4. Use the data on the housing tab to answer the following question:

A. In each city what percent of ALL housing units are vacant and are not currently for rent, for sale only?

B. Which city has most expensive mobile homes? What is standard error telling you about mobile home prices?

C. Consider both rented and owner occupied housing. Which city in Colorado has highest housing costs? Use at least three pieces of data from the housing page to support your argument. Does the picture become more or less extreme if you consider housing costs relative to that cities median household income?

One more question below!

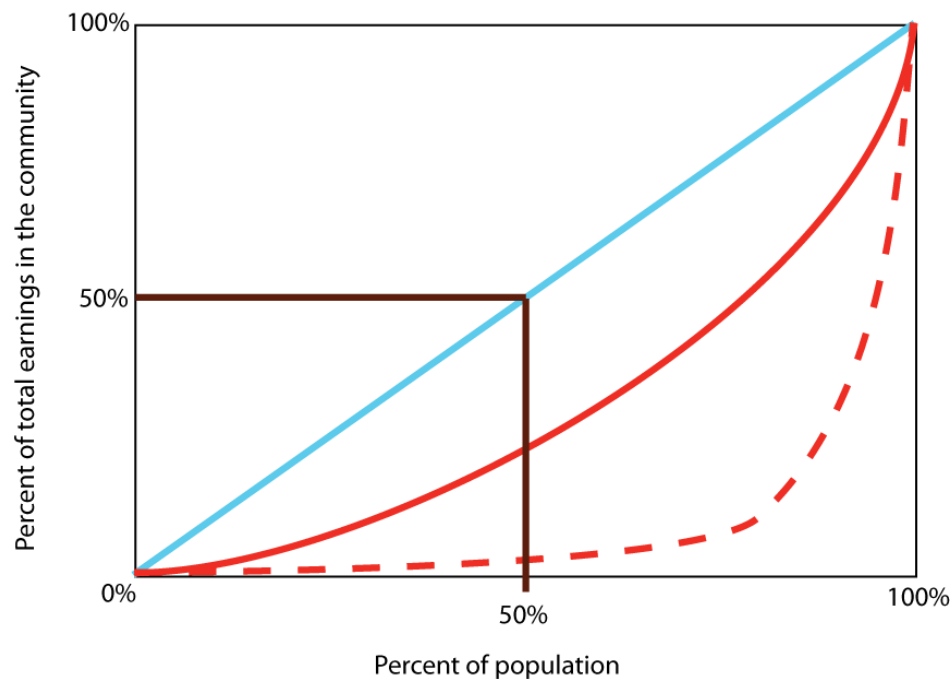
Looking more at measures of income distribution

Most of the variables in the data table describe things that don't need much explanation, income, commuting behavior, years of education. There is one column (on the income tab) that contains something called a gini index of income inequality or a gini coefficient.

The gini coefficient summarizes the dispersion of incomes, it easiest way to describe it is with a graph called a Lorenz Curve. Figure 4 is a Lorenz curve showing the total earnings in a community on the vertical axis. Imagine adding up the annual earning for every person in a town- this annual amount is the 100% mark on the vertical axis. On the horizontal axis we have the percent of the total population. 100% of the population always earns 100% of the total earnings however, 50% of the population does not always earn 50% of the total earnings. The blue line on the Lorenz curve (Figure 4) represents a totally egalitarian society, like the town where everyone worked at the same factory, in this every society each person earns the same amount of money so that 50% of the people control 50% of the wealth. However few places are totally egalitarian, the red line represents a place where the wealth is moderately concentrated, half of the population controls 25% of the wealth. This means that the wealthiest half of the population controls 75% of the wealth. The dashed red line represents a society with extreme inequality, 50% of the wealth is controlled by less than 10% of the population.

Conceptually, the gini coefficient measures the area between the blue diagonal line and the red line(s). If there is a lot of inequality the distance between the blue and red lines will be large and the area covered by the space between the red and blue lines will be a large percent of the lower portion of the graph- the gini coefficient measures this area. South Africa, a society with a legacy of enormous inequality, according the UN's Human Development report (2007/2008) has a gini index of around .57, the gini coefficient for Canada is .32, the US has more inequality than Canada but less than South Africa (Gini coefficient is around .40).

FIGURE 4: THE GINI COEFFICIENT



5. Now that you know what the “gini index” says, look back at your graph from question 3. How do the gini coefficients of the top 15 most populous Colorado towns relate, if at all, to the patterns in income distribution shown in your graph?

Deliverables

This week, submit a word document file (YOUR_NAME_LAB_#) with the answers to the questions in section 3.

After looking at all this data what do you think: Is Boulder different from the rest of Colorado? Submit a 1 page essay arguing your position. Essays will be graded on how well use data to support your position. The more data you use the better but I'm looking for random facts. Make sure you are using the data to support your position. You may use graphs and charts as you see necessary but they don't count toward the length.

Note: You may work on this assignment in groups of up to 3 people. If you work together, please write the names of your group members at the top of the lab and summarize each of your contributions at the end of the write-up.