Hands-On Lab

Collect and Compare Data

Lab version: 1.1.0

Last updated: 5/20/2017

Contents

[Overview 3](#_Toc422004603)

[Exercise 1: Collecting and Formatting State Data 3](#_Toc422004604)

[Task 1 – Collecting the Data 3](#_Toc422004605)

[Task 2 – Format the Data Into a Single Sheet 6](#_Toc422004606)

[Exercise 2: Basic Data Analysis 7](#_Toc422004607)

[Task 1 – Calculate Raw Job Changes 7](#_Toc422004608)

[Task 2 – View the data trend 8](#_Toc422004609)

[Exercise 3: Absolute and Relative Change 11](#_Toc422004610)

[Task 1 – Create a sheet for raw job change 11](#_Toc422004611)

[Task 2 – Create a sheet for relative job change 12](#_Toc422004612)

Overview

* 1. In this lab, we will collect jobs data from multiples sources, organize it and compare it visually.

# Prerequisites

* 1. The following prerequisites are required for this hands-on lab:
  + Microsoft Excel
  + Internet Connection

# Estimated completion time

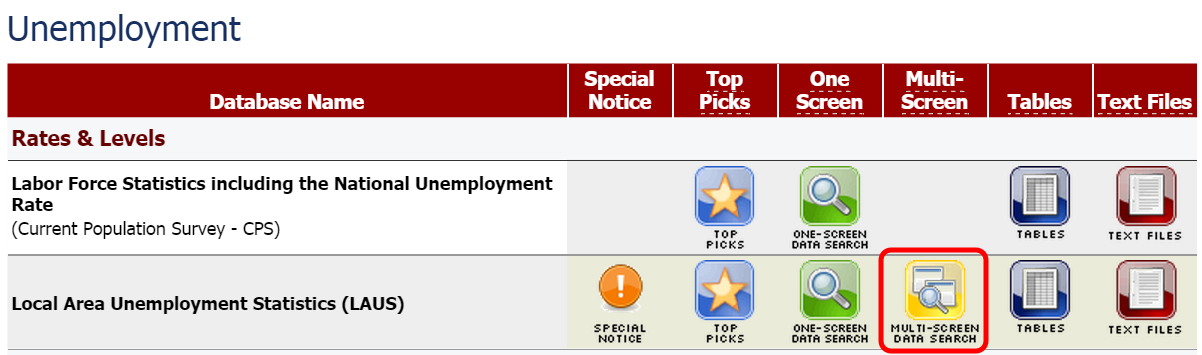
* 1. Completing this lab should take at least 60 minutes.

Exercise 1: Collecting and Formatting State Data

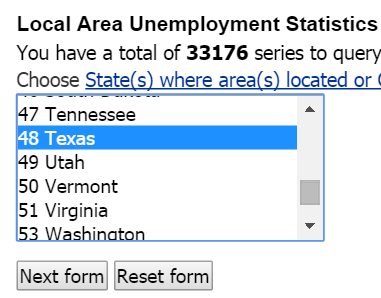
* 1. In this exercise we will collect data on state jobs from the Bureau of Labor Statistics and format it in such a way that we can graph it in Excel. We will start by collecting data on California and Texas and comparing the two.

Task 1 – Collecting the Data

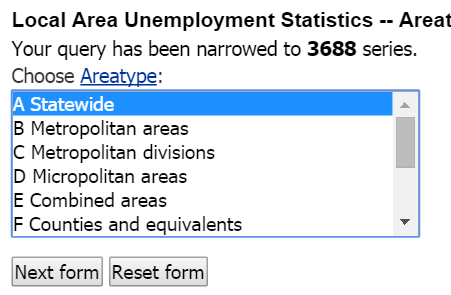
* 1. First, we need to go to
  2. Open up <http://www.bls.gov/data/> in your favorite browser that isn’t Safari.
  3. Scroll down to the “**Unemployment**” section and in the **Local Area Unemployment Statistics** row, click on **Multi-Screen Data Search.** Note: you can use the One Screen Data Search but I find it harder to use and it involves installing Java and I wouldn’t do that to you.



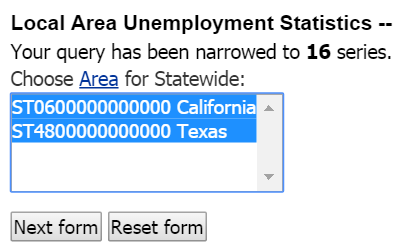
* 1. Select the following items:
     1. Screen 1 of 6 – Select “California” and scroll down and, pressing the CTRL key, select “Texas”



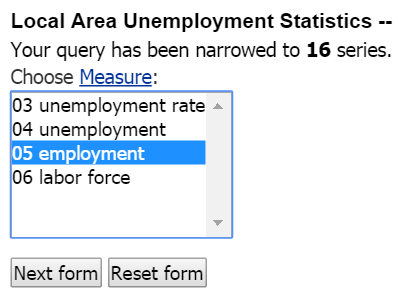
* + 1. Screen 2 of 6 – Select “A Statewide”



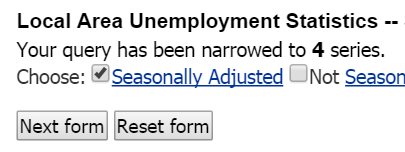
* + 1. Screen 3 of 6 – Use Shift or CTRL (Command on Mac) to select both options



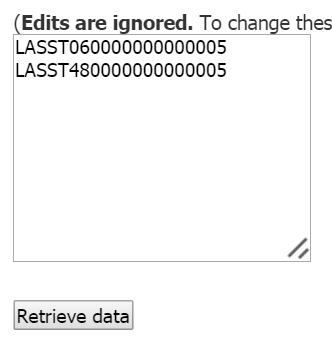
* + 1. Screen 4 of 6 – Select “05 employment”



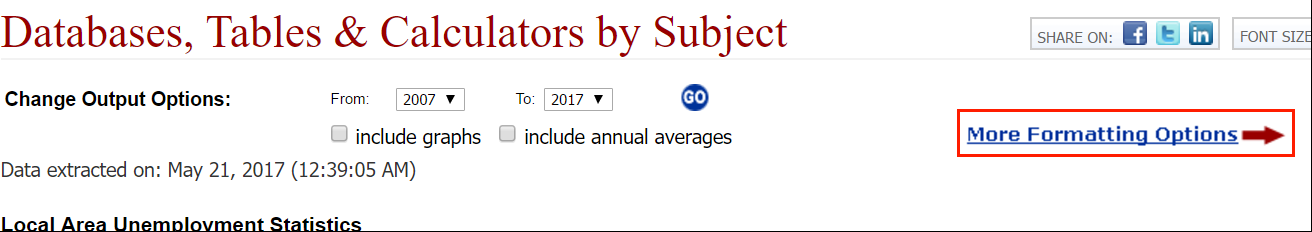
* + 1. Screen 5 of 6 – Select Seasonally Adjusted



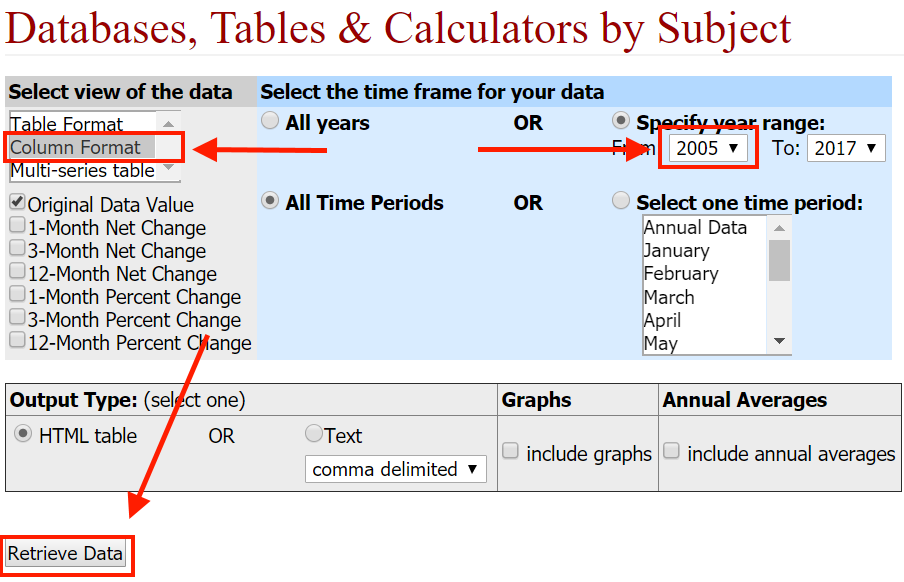
* + 1. Screen 6 of 6 – Click “Retrieve data”



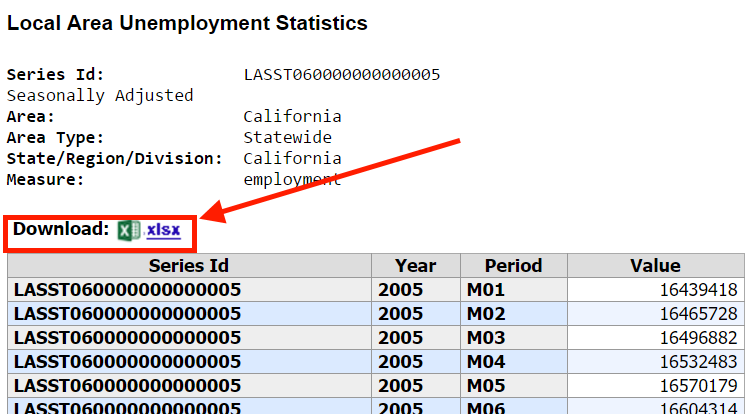
* 1. This will take you to a data retrieval page where you can download the data. For our purposes, we want the data in a very specific format, so click on the **More Formatting Options** button in the top right of the page.



* 1. Change the view of the data from Table Format to **Column Format**. Select **2005** for start year. The click **Retrieve Data**



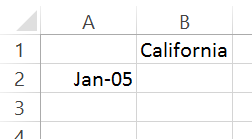
* 1. This will take you to the data as we want it to be formatted. Click the Download link for both California and Texas.



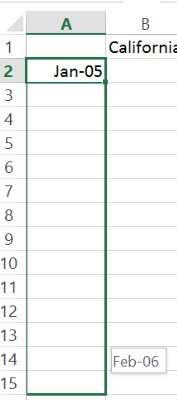
* 1. These files will have a coded name (something like “SeriesReport-20170521004001\_abcde.xlsx”). Rename them to “California Employment Data” and “Texas Employment Data”

Task 2 – Format the Data Into a Single Sheet

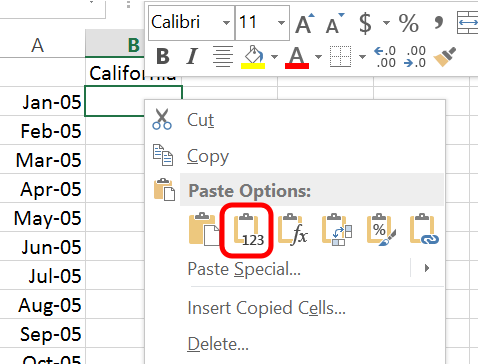
1. Open up a new Excel document and name it “**State Employment Data**”.
2. At the bottom of your sheet, you should see that it is named “Sheet “. Rename this “**Jobs by State**” by double clicking into the sheet name field.
3. We want to copy all this information but do it in a way that gets rid of this formatting and adds a column to identify the month. Start by typing “California” into cell B1 and “Jan 2005” into cell A2.



1. Click on the handle on cell B1 and drag it down until you get to “Apr-17”



1. Copy the column of data from your California document (column D). Go to your Jobs document and right click on cell B2. Select “Paste Values” from the pop-up menu.



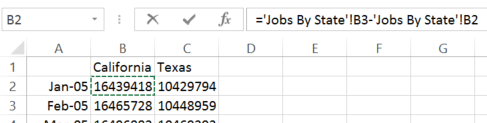
1. Do the same for the Texas document, opening it and copying the values from column D and pasting them into a “Texas” column in your jobs doc
2. If you’re running out of time, you can open “Lab 2 – midpoint” in the **Lab 2 Assets** folder.

Exercise 2: Basic Data Analysis

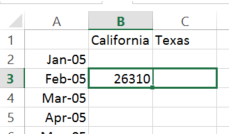
* 1. In this exercise we will look at how jobs in these two states change from a single point in time. We’ll calculate this as both raw job changes and as a percentage of the original number.

Task 1 – Calculate Raw Job Changes

* 1. We will ultimately want to calculate all the job changes from state to state in a way that is extensible (will be reusable as we add more states to our data set). This method will let us do that.
  2. Create a new sheet using the method we used earlier and name it “Monthly Change”.
  3. Copy and paste column A from the **Jobs by State** and copy-paste the header cells (California/Texas) into the **Monthly Change** sheet
  4. In **Monthly Change**, click on cell B3 and type “=”. This will kick Excel into “formula” mode. Navigate back to **Jobs by State** and click cell B3. Type “-“ and select B2. Hit enter.



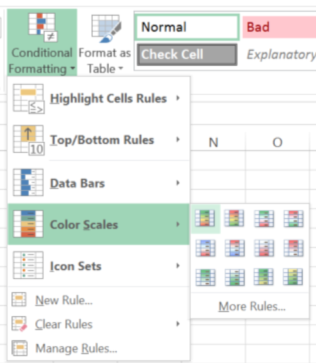
* 1. Our B3 cell in Monthly Change now has the monthly employment change in California from January 2005 to February 2005. We can extend this to Texas by dragging the formula across to the C column.



* 1. Drag the formula down to row 130 (Sept-15) and we now have the month-by-month job change for both states for the last decade.

Task 2 – View the data trend

* 1. It’s never too early to start looking at our data. Seeing data from many different angles gives us a comprehensive understanding of the data set. In this task, we’ll look at the monthly changes data two ways: first with Excel’s built-in color range tool and then by creating a simple graph.
  2. Select all the data cells in Monthly Change (B2 – C130). In the “Home” tab, select “Conditional Formatting”, then “Color Scales”, then the green-to-red option. This will auto-analyze our data and color the largest value green, the smallest value red, and everything in between an appropriate gradient shade.

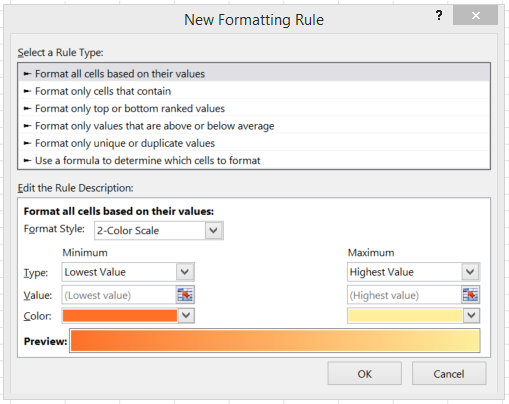


* 1. There is a zoom slider on the bottom right corner of Excel. Zoom out until you can see a good 10,000 foot view of the data. We can see some cycles of growing employment (green) and job losses (red).

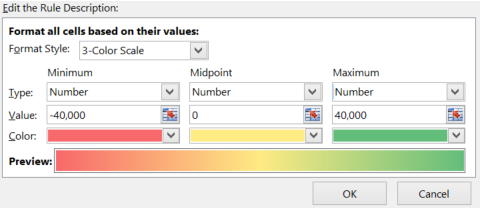




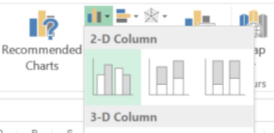
* 1. The problem we have right now is that California has such incredible sways in data that the midpoint for our color scale isn’t at “0 jobs lost”… it’s actually at -17,930. Let’s fix that and get a better view of our data.
  2. Select all the data cells in Monthly Change (B2 – C149). In the “Home” tab, select “Conditional Formatting”, then “Color Scales”, but select “more rules” (Mac: select “New Rules”).
  3. PC: You’ll see a pop-up allowing us to create a new formatting rule.



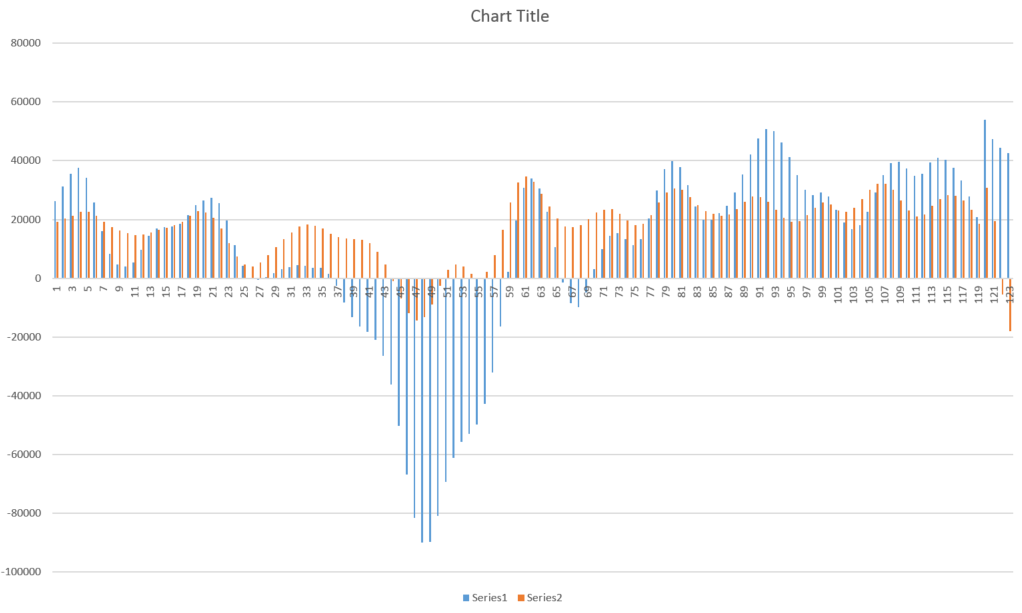
* 1. In “Format Style”, change the selection from “2-Color Style” to “3 Color Style”. Change the Types for Minimum, Midpoint, and Maximum all to “Number”. Set the minimum number to -40,000, the midpoint number to 0, and the maximum number to 40,000. Click “OK”.



* 1. Zoom in (you can also zoom quickly by holding the CTRL key and scrolling your mouse) and look at the strongest red area. We notice that the big decline starts around the beginning of 2008. Let’s investigate that with a graph.
  2. Select all the data cells in Monthly Change (B2 – C130). Select the “Insert” tab. You should see some chart icons in the ribbon. Click on the vertical bar icon and select the “clustered column” chart option.



* 1. You’ll get a simple graph. Resize it so that we can get a better view of this graph.



We notice that California was hit much harder by the recession than Texas but also that they have gained more jobs in much of the last few years. This is partly because California is substantially larger than Texas. This would be an even bigger problem if we were trying to compare job growth in California to a much smaller state like Georgia or Connecticut. So how do we compare these two states in a way that reflects their relative size?

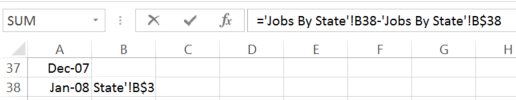
* 1. Shifting our thought back to our other observation of how the economy started to lose jobs at the beginning of 2008, let’s pose a question: How have jobs in these two states grown since the high point before the recession. We’ll spend the rest of this lab analyzing our data to answer that question.

Exercise 3: Absolute and Relative Change

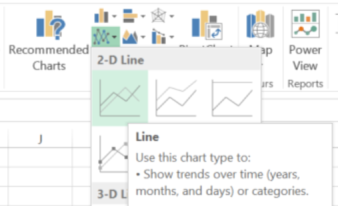
* 1. In this exercise, we will look at the change in raw jobs in both states using an “anchor” date. We’ll then look at the relative job change, which will give us job growth as a percentage of that anchor data point.

Task 1 – Create a sheet for raw job change

* 1. First, we will create a sheet that will track the absolute job change in each state since January 2008.
  2. Create a new sheet using the method we used earlier and name it “Change since 08”. Copy and paste the A column (our dates) and the 1 Row (our data headers).
  3. Click on cell B38 and type “=”, then click on the Jobs By State sheet and select B38, type “-“ and select B38 again. Hit enter.
  4. Select B38 again and in the formula field (above the spreadsheet) change the second B38 to B$38. This will anchor the row (but not the column) for that part of the formula as we drag it across the spreadsheet.



* 1. Drag the formula over to the C column (Texas) and down to row 149 (Apr-17).
  2. Now we have a set of data showing how both states were hit by the recession and how they recovered. With the data selected, click on Insert tab and select the Line Chart to create a line graph of this data.

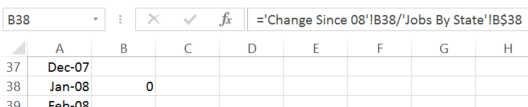


* 1. The resulting graph gives us a better view of this story of how the respective states have fared since the recession. But it still doesn’t account for the fact that California has a much larger population than Texas.

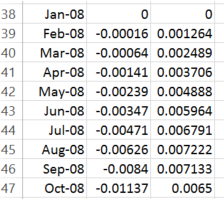
Task 2 – Create a sheet for relative job change

Let’s create a sheet that accounts for the differences in state size.

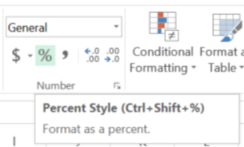
* 1. Create a new sheet using the method we used earlier and name it “% Change since 08”. Copy and paste the A column (our dates) and the 1 Row (our data headers).
  2. Click on cell B38 and type “=”, then click on the **Change Since 2008** sheet and select B38, type “/“ and click on the **Jobs By State** sheet and select B38 again. Hit enter.
  3. Select B38 again and in the formula field (above the spreadsheet) change the second B38 to B$38.



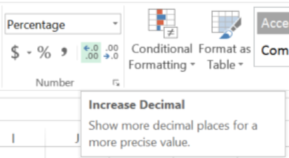
* 1. Drag the formula over to the C column (Texas) and down to row 149 (Apr-17). We have a raw calculation that has a lot of data but not a lot of context.



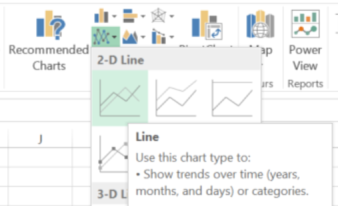
* 1. Select all the calculated cells (from B38 – C149) and press the “Percent Style” button in the Home tab.



* 1. Now press the “increase decimal” button to give another decimal point to the data.



* 1. Create a line chart like we did before. With the data selected, click on Insert tab and select the Line Chart to create a line graph of this data.



* + 1. The resulting graph gives us the best view of the hit of the recession and how these two states recovered.
    2. 