**MAKING MAPS WITH GOOGLE FUSION TABLES**

Robert Gebeloff

[rgebeloff@nytimes.com](mailto:rgebeloff@nytimes.com)

Even if you’re highly skilled at parsing data and writing formulas, sometimes you need to visualize your data to get a better understanding of any patterns that might emerge.

Journalists working with data organized by geography like to use mapping – connecting the numbers in your spreadsheet to locations on a map.

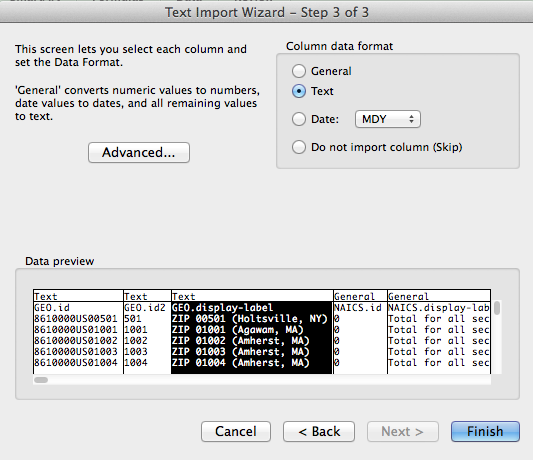
There are several ways of doing this. For intensive mapping jobs, many journalists use a “GIS” program like ArcView from Esri, or the open-source GIS app Qgis. You can also map data within more robust data frameworks such as Python and R.

But the easiest way to get started is with a Google plugin called Fusion Tables. A Fusion table file can read your data right out of a spreadsheet or CSV file and then connect it to the Google mapping system.

Let’s do a quick example to see how this works. We’ll use a file from the Census Bureau’s Zip Code Business Patterns survey (see <http://www.census.gov/econ/cbp/> for more information).

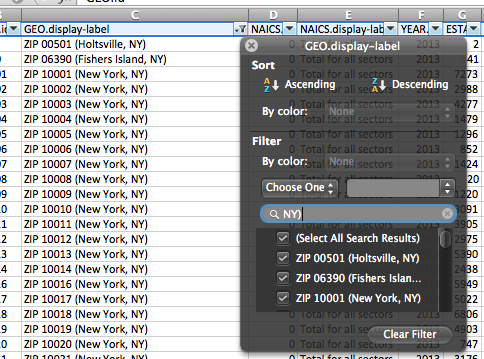
Our goal is to look at the average pay for jobs located in the largest zip codes in New York – those with at least 1,000 establishments. So we’ll need to clean the file in Excel first.

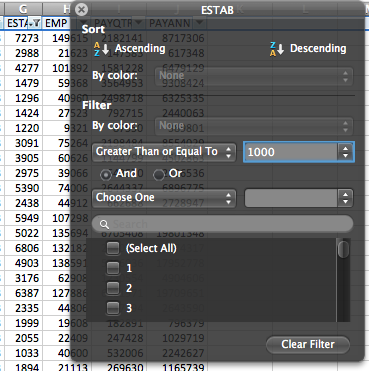
1. Open up the Excel and import the file biz.csv. You can download it from the class directory. REMEMBER: It’s always better to “import” data than to just open it by double clicking, because this will allow us to use the import wizard to check our column data types.



Note here how I’m declaring my first three columns as “text” format. This is because I want to make sure that zip codes that begin with 0 don’t lose the “leading” zero. Once imported, save the file as a spreadsheet with name like “allzips.”

1. Let’s first use a filter to pull out NY Zip codes with at least 1,000 establishments. Highlight the top row, click on Data/Filter. For column C, let’s filter on records that contain the string “NY)”, with the closing parens – this gives us every zip code where the label ends in NY but ignores zip codes where the letters “ny” might be in the middle of the postal name.

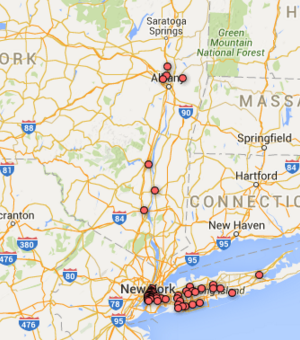


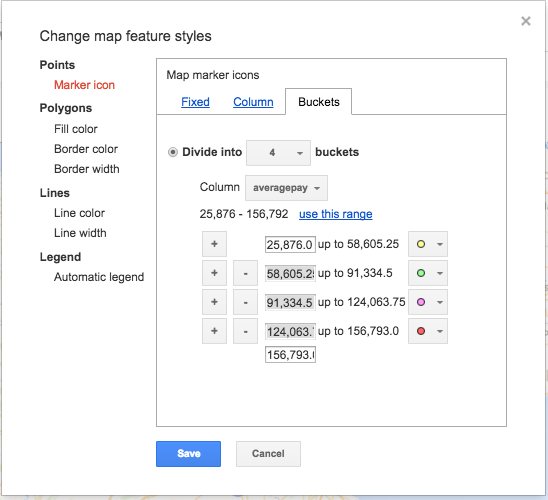
3) Now on column G, let’s add another filter to pull out only zip codes where there are 1,000 or more establishments. Once you have this filter on you can close the filter dialog box, and copy/paste the filtered data into a new file, which you can save as “bigny”. The idea is to preserve our source data but keep separate our work files.

1. We now can get to work on “bigny”. We have two jobs now – to make sure we have a column with valid zip codes, and to calculate the average pay. Look at the data as it exists now – you can see the zip code exists in the first three columns, but column B seems to have what we want. (If we were dealing with a state where the zip code begins with 0, and we lost it, here’s where we’d have to fix it.) So let’s rename column B “zipcode”.
2. We now want to calculate the average annual pay for workers in each zipcode. Let’s add a column called averagepay, and use a formula for the calculation:

=(j2/h2)\*1000

As you can see, there is quite a disparity among the different zip codes. But unless you’re a mail clerk, you probably cannot associate any of the numbers with the actual neighborhood. So let’s ask Google Fusion Tables to help.

1. From within a Google account, click on Google Apps and Google Drive. Then under My Drive/New File, create a new Fusion Table. If you don’t see Fusion Tables as one of the sources, click on connect more apps and search for Fusion Tables in the app library.
2. Create a new Fusion Table by importing your bigny spreadsheet. Once you see that it recognizes your rows and columns, click through to finish.
3. Let’s make the map. We need to tell Fusion Tables that our zipcode column is, in fact, a zip code. Click on the column header and then click “change” – in the dialog box change the “type” to a location – this is a data type we haven’t seen before in Excel.
4. Once you save the change to “location”, the data will appear under a yellow highlight in your table. You can now hit the red plus sign and click “Make a map”. It will take a minute or two, but you’ll end up with a dot that represents each zip code.
5. 
6. This is interesting to see that most of the largest employment centers are in the city and on Long Island, but what we really want to examine is the pay difference. Under “configure map” on the left, let’s change our “feature styles”. Let’s choose four buckets based on averagepay.



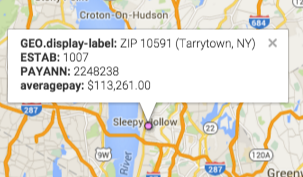
Let’s also turn on the automatic legend. You’ll now see a geographic pattern:



There are other features to explore. Let’s say you only wanted to see zip codes where the average annual pay topped $100,000. Click on the “Filter” button on the left.

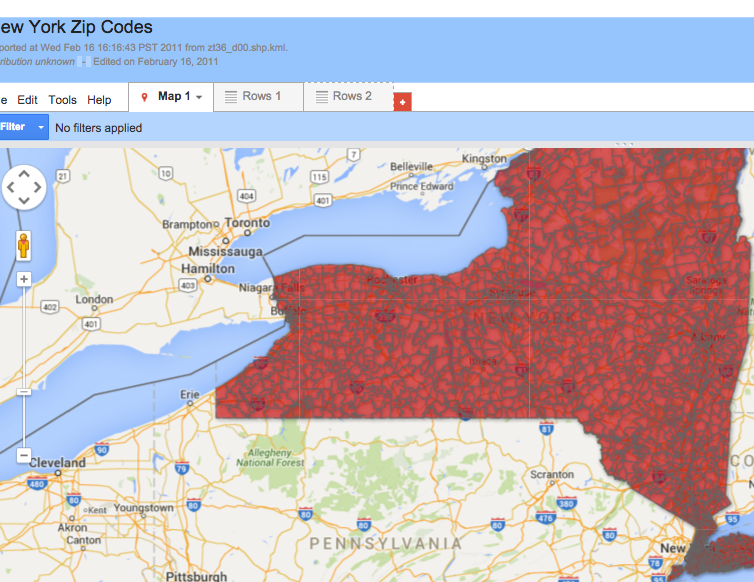
You’re left with a handful of zip codes in Manhattan and one in Tarrytown.

Lastly, let’s clean up the info window that pops up when you click on an icon – all we want to see is the zip code’s name, the number of establishments, and the average pay. Click on “Change Info Window” and revise your choices.

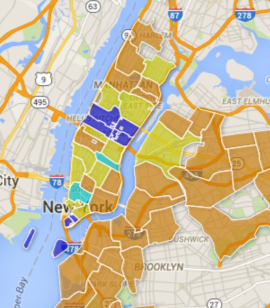


Ok, one more cool Google Fusion trick. Google maps can read polygons in a format known as KML, and for something like New York zip codes, it’s possible that there is already a zip code map in that format. Let’s do a regular google search for “new york city zip code kml”.

Voila, found something:



1. Generally in journalism, we don’t take data we find online without a lot of vetting, but in this case, it’s a map, and it looks correct. Click on the map and take note that in the underlying datafile, the zip code is called ex\_ZCTA – we can use that field to merge the zip code polygons to our map. So switch to the rows1 tab, copy the URL, and go back to our map.
2. Click File/Merge and paste the Web address of the polygon map into the form on the bottom. A dialog box appears asking how we want to make the connect. In our file, we have something called zipcode – let’s pick that and ex\_ZCTA for the new file. We can now “change feature styles” and switch to a polygon presentation.



This is just the tip of the iceberg when it comes to mapping. Google Fusion Tables are fast and easy once you get the hang of them, but they can be slow, so if you have a project where you need to do big-time mappy, I’d suggest getting to know QGIS or ArcView.