In this tutorial you’ll learn how to take an Excel file and make a map out of it, like this:

<https://postmediamontrealgazette2.files.wordpress.com/2014/10/most-common-industries-boroughs2.png>

Specifically, we’ll take a spreadsheet with job numbers for different industries for each borough in Montreal, find the largest industry in each, and join that to a map of Montreal in QGIS.

**Requirements:**

* Microsoft Excel (or LibreOffice or Google Spreadsheet)
* QGIS
* This Excel file: <http://ville.montreal.qc.ca/portal/page?_pageid=6897,102849654&_dad=portal&_schema=PORTAL>
* This shapefile of Montreal’s borough and demerged suburbs:  
  <https://drive.google.com/file/d/0B2u6nj7h8baHbDgzbnF0QkJFRTQ/view?usp=sharing>

**Excel formulas used in this tutorial:**

Arguments in [brackets] are optional.

**=MAX(number, [number…])**   
Find the highest number in a sequence or a cell range

**=MATCH(lookup\_value, lookup\_array, [match\_type])**  
Returns the row number of a value in a cell range.   
Match type: 0 for exact match, 1 for close match

**=INDEX(array, row\_number, [column\_number])**  
Given a cell range, it returns the value in the same row of a column you specify.

**=MID(text, start\_num, num\_chars)**  
Copy a text starting at start\_num up to num\_chars.  
If you want to start at the beginning best to use the LEFT function.

**=FIND(find\_text, within\_text, [start\_num])**  
Returns the position of a text you specify within a larger text. For example:

=FIND(“ab”, “Alabama”) returns 3

**=LEN(text)**  
Returns the number of characters of a text or cell.

**1. Open the Excel file and get comfortable with it**

Look at the top row. Look at the first column. What is this data bout?

What do the numbers represent? What about those number codes before each industry name in column A? Do your research first. Google is your friend.

**2. Cover your ass**

Whenever manipulating any kind of data, always keep a copy of your original. Copy the raw data, and paste it into a new worksheet.

**3. Filter the values you want**

We just want the numbers for the top-most level of industry classification, those that start with two digits or a range of two-digit numbers, like 11, 21, 31-33, etc.

Select the top left cell. On your toolbar, click the little funnel button. That’s your filter. It will add a little drop-down menu on your column headers.

Click ‘Select all’ or ‘Clear’ to unselect everything. Now go down the list and check off the values you want. That is, all values that start with two numbers.

You should have 18 rows of data after filtering.

**4. Find the highest job number in each area**

Copy your selected data and paste it into a new worksheet.

Name a new row on Column A “Highest industry”. In Column B, write in this formula:

**=INDEX($A$2:B18,MATCH(MAX(B2:B18),B2:B18,0),1)**

That’s a big formula, so let’s break it down, from innermost to outermost:

**MAX(B2:B18)** - this will find the highest number in column B.

**MATCH(MAX(B2:B18),B2:B18,0)** - This returns the row number of the max value of Column B.

**INDEX($A$2:B18,MATCH(MAX(B2:B18),B2:B18,0),1)** - This returns the value in the first column that’s on the same row as the max value in column B. That’s the industry that has the highest number of jobs.

Click on the little black box on the bottom right and drag it across your 32 columns. It will copy the formula across.

**5. Clean your data**

Again, copy your data to a new sheet.

Let’s get rid of those number codes. They’re useless to us now. We want to just keep all the text after the space following the numbers.

Right click on the column B label (the letter B) and choose ‘Insert’. It will add a blank column. Write in this formula on B2:

**=MID(A2, FIND(“ “,A2)+1, LEN(A2))**

Again, let’s break this down:

**FIND(“ “,A2)** - this will return the position of the first space in A2. If the text in A2 is “11 Farms”, it will return 3.

**=MID(A2, FIND(“ “,A2)+1 …** - this will start copying the text in A2 starting at the first space. But we don’t want that space. The **+1** makes sure we don’t capture it. That is, we want to start at one position after the space. Using the example above, this would make it 4.

**=MID(A2, FIND(“ “,A2)+1, LEN(A2)) -** The LEN function makes sure we copy everything until the end, no matter how long the text is.

Copy that formula down the column (remember the little black box on the bottom right?). You should have the values in Column A without the number codes. Copy that column you just populated. Select A1, right-click and choose ‘Paste Special’ and ‘Values only’. This ensures you only paste the values of the cells and not the formulas you wrote.

Notice how the industry names on the bottom row also changed to the cleaner version.

Now let’s get rid of the (ville) and (arrondissement) qualifiers next to each region name in the top row. Select all of Row 1 and bring up the ‘Find’ dialog. It’s in the Edit menu or Ctrl-F (Command-F on Mac).

Choose the ‘Find and replace’ option. In the ‘Find’ box, type in “ (\*)”. This means: find a space, an open parenthesis, anything that follows it until a close parenthesis. In the ‘Replace’ box, put nothing. Choose ‘Replace all’. All clean.

**6. Prepare the data for mapping**

Again, copy all the data and paste it into a new sheet. But make sure you paste values only. We don’t want any more formulas.

Delete all the rows that aren’t your max row and your largest industry row. We no longer need that data. You should be left with three rows: Area, max jobs, and highest industry.

We need to reshape this data from 3 rows x 32 columns to 32 rows x 3 columns. Select your three rows and copy them. On a blank cell (say, A5), hit ‘Paste special’, then select ‘Transpose’. Your data is reshaped to “long” form. Delete the “wide” version in your top three rows and any blank rows. We’re done.

If you want, you can translate the industry names to English.

**7. Map that data**

Open QGIS. Click the ‘Add Vector Layer’ button. The Source Type should be File, and the Encoding should be latin1. Click the ‘Browse’ button and find your Montreal shapefile. Make sure you unzipped it first! Select the file with the .shp extension. Click ‘Open’.

Hello, Montreal island. You’ll see that the shapefile has been added to your layers panel on the left. Right click on it and choose ‘Open Attribute Table’. That’s the data behind that shapefile. We only have one column, with the name of each borough and demerged suburb. We’ll use that to join our Excel data to it.

Make sure the names on the Excel file and the shapefile are spelled EXACTLY the same. The tiniest difference will make the join fail for that area. Once you’re satisfied, save your final Excel sheet as a CSV (comma-separated values) file.

Click the ‘Add Vector Layer’ button again. Add the CSV you just created. It will appear on your layers list. Right click it and look at the attribute table. If you’re data looks good, the way you made it in Excel, you’re ready to join. If there are weird symbols in place of letters, you’ll have to choose another encoding when you open the file. Try UTF-8 or macintosh.

Double click the Montreal layer in the layer list. This is the main control panel for that layer. Choose ‘Joins’ on the left menu. Click the green + button. Your join layer is the CSV file you added. The join field is the column in your CSV with the names of your areas. And the target field is the column on the shapefile with the same area names. Click OK. Then OK again.

Nothing happened right? Wrong. Look at the attribute table of the Montreal shapefile again. It has the data you added! If there are any rows without data, it means the area names between the files didn’t match. Go back and fix it in Excel, and add the corrected CSV file again.

Now you can style the map.

Double-click the Montreal layer again. Choose the ‘Style’ menu. Right now all areas are styled the same as a single symbol. In the top menu, choose ‘Categorized’. Now you need a column to categorize it by. Choose your largest industry column. Click on ‘Classify’ to add colours to values in the column. Click OK.

Doesn’t that look great? Notice there’s also a colour legend under the Montreal layer. You can customize the colours for each value by double-clicking on it in the Style menu.