# **Mapping with Carto**

## This guide will require a:

- **Gmail** account
- <u>Carto</u> account

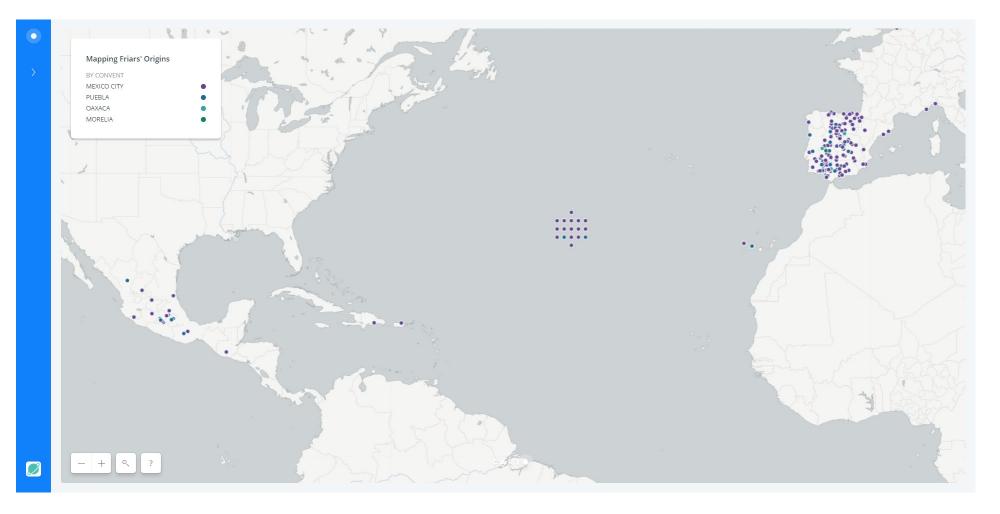
# This guide uses the following sources:

• Primary Source:

"Book of professions from the Order of St. Augustine Convent of Mexico City, 1537-1653," Benson Latin American Collection, Genaro García Collection, G21

• Secondary Source:

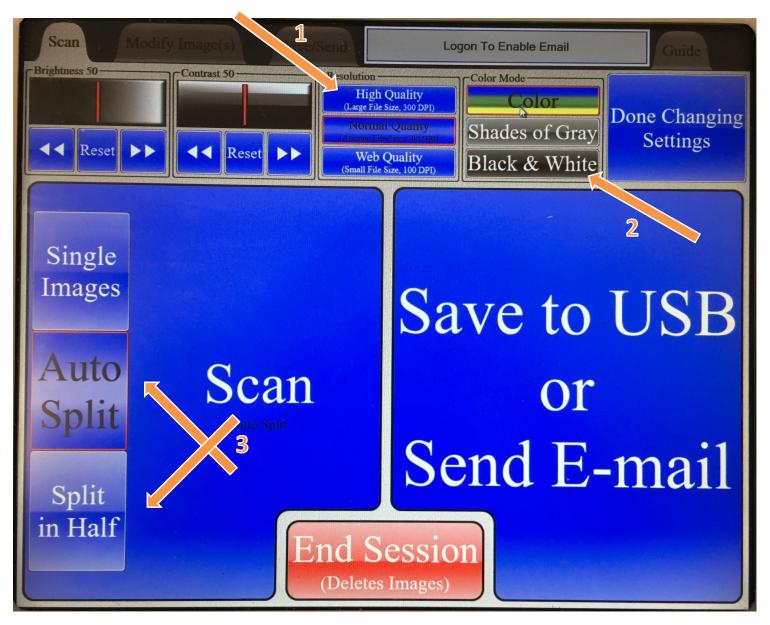
Ennis, Arthur J. Augustinian religious professions in sixteenth century Mexico (Villanova, Pa.: Augustinian Historical Institute, 1986).



## Scanning Published Data for OCR (Optical Character Recognition) Software

- For UT Libraries' KIC Scanning Stations, select:
  - o Resolution: High Quality
  - o Color Mode: Black & White (for high contrast)
  - o Scan: either Auto Split or Split in Half

- Smart phone scanning apps with OCR capabilities:
  - Scanner Pro (iPhone, \$4-5)
  - Text Fairy (Android)
  - o Google Keep (Android)

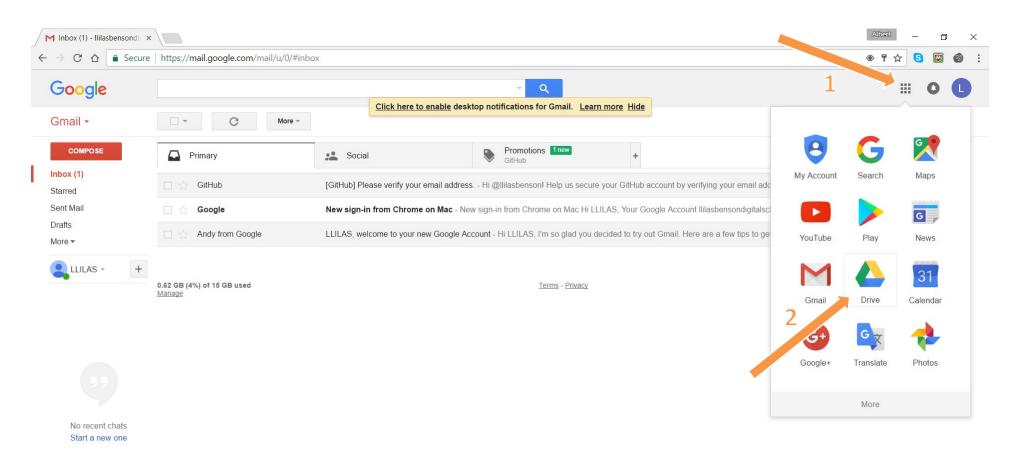


# Applying OCR (Optical Character Recognition) Software on PDFs using Google

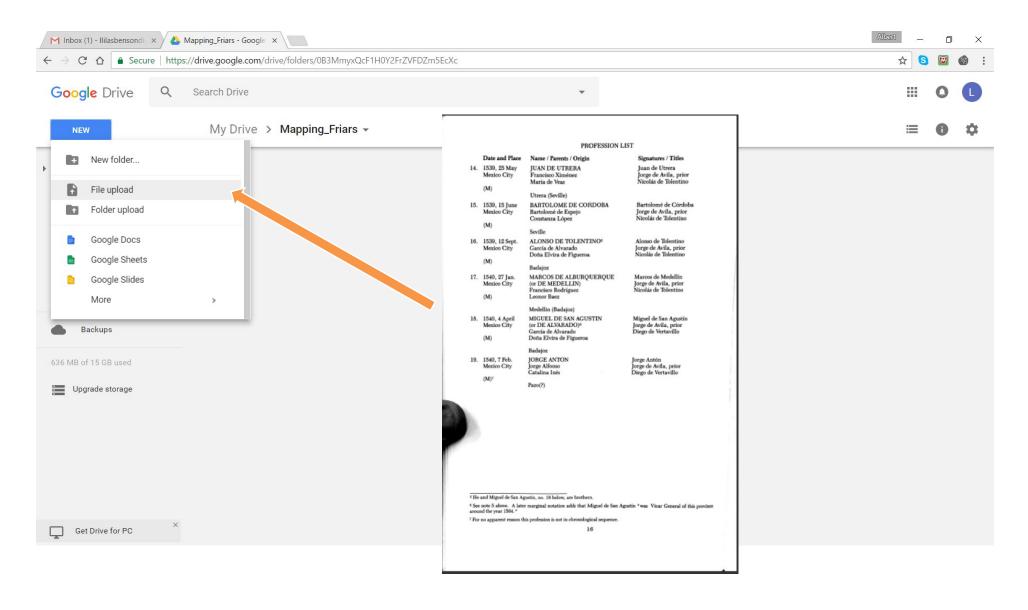
- **Note**: Google Drive will only run its OCR software on 25-35 page PDFs (it depends on file size). If your file is larger than this, you might consider breaking up the PDF into 20-page chunks using <u>PDFtk</u> or <u>Smallpdf</u>.
- Log into your Gmail account

https://drive.google.com/?tab=mo&authuser=0

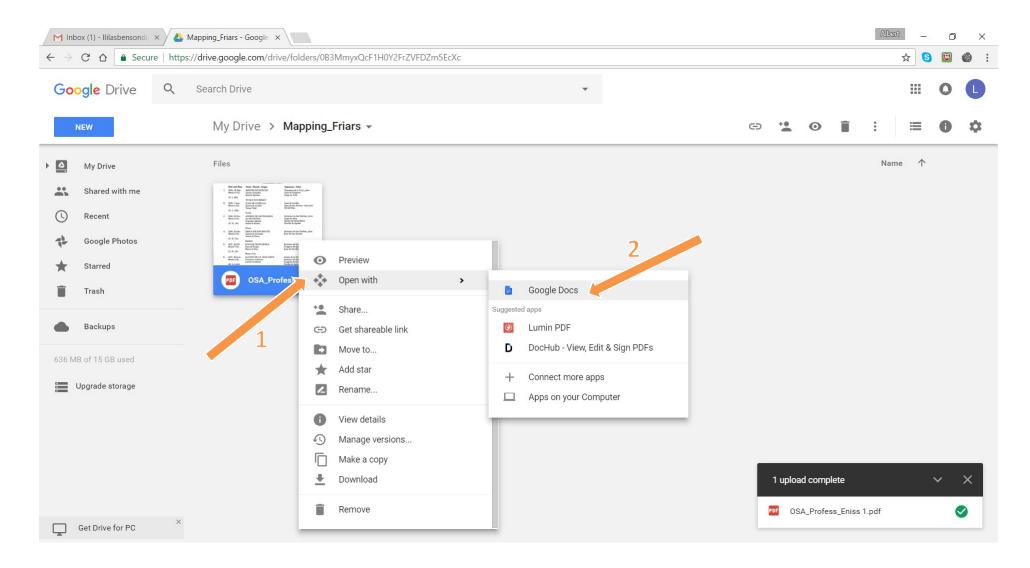
• Click on the **Google apps** icon and select **Google Drive** (it will open in a new tab)



• Select New, then File upload to upload the black & white PDF scan of the document you want to OCR

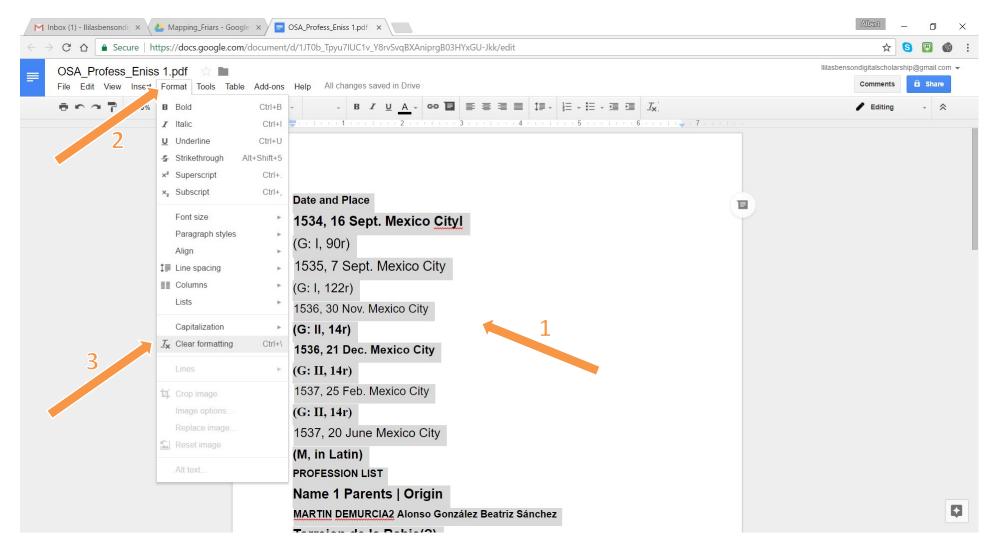


- When the PDF file uploads, right-click on the file thumbnail, select **Open with**, then **Google Docs**.
- This will initiate Google's OCR software. The PDF will open with Google Docs in a new tab.

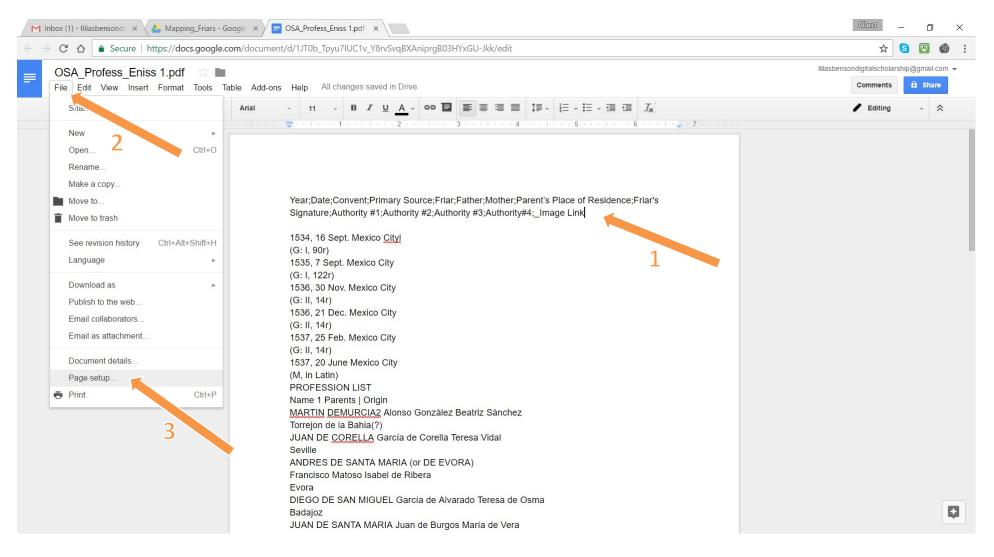


### **Cleaning Up Extracted Text**

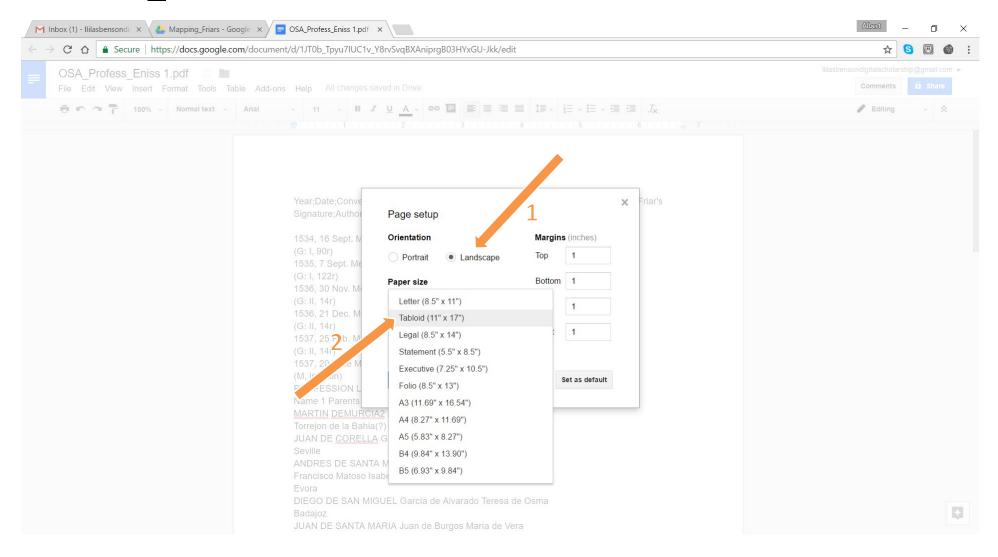
- In this section, you will be structuring the extracted/OCRed text to convert it into a table.
- To simplify the data visually, select all the text, click on <u>Format</u> in the main menu, then click on <u>Clear formatting</u>.



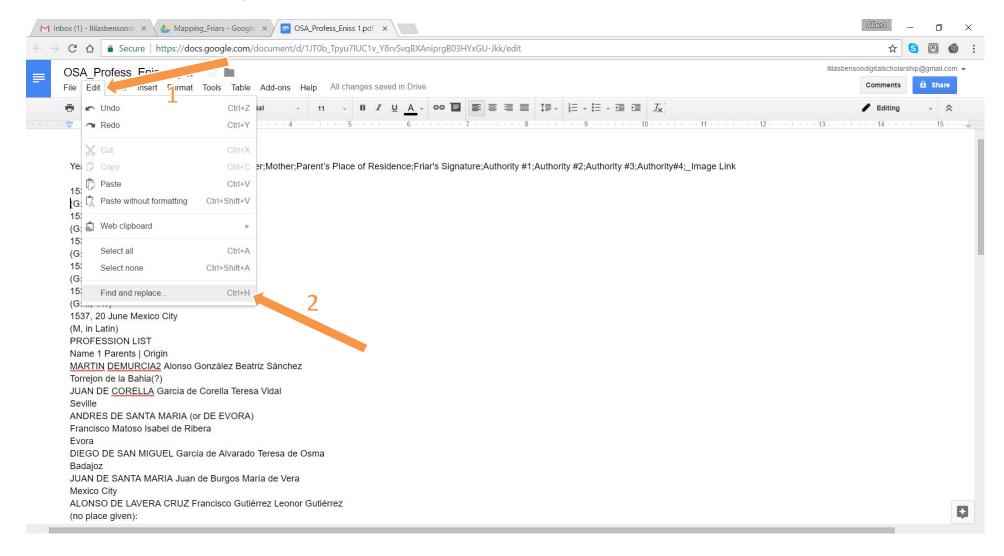
- At the top of the document, list the headings of the table columns you want to create and separate them with either a semicolon or | (no spaces in between text and ;/|). Think of the ; or | as the vertical gridline separating out columns.
- Re-orient and resize the document to help visualize the table rows. In other words, we want to limit the data that will go in one table row on one text line in Google Docs. Click on <u>File</u>, then select <u>Page setup...</u>.



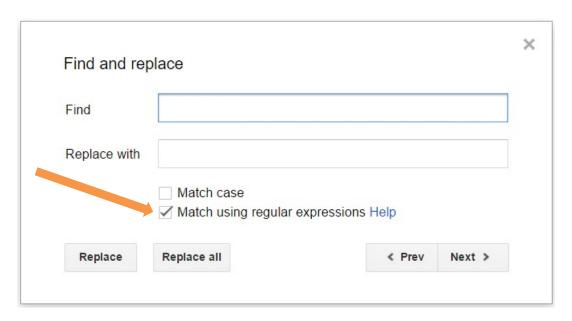
- In the Page setup window, change:
  - o Orientation: Landscape
  - Paper size: Tabloid (11" x 17")
  - o Click **OK**



- We will use Google Docs' "Find and replace" functionality to start cleaning up the data.
- Select <u>Edit</u>, then <u>Find and replace...</u>.



• In the Find and replace window, check the Match using regular expressions box.



### **Functions**

Replace line breaks [\n] with semicolons:

Find: City[\n]
Replace with: City;

1534, 16 Sept. Mexico City (G: I, 90r)
1535, 7 Sept. Mexico City (G: I, 122r)
1536, 30 Nov. Mexico City (G: II, 14r)
1536, 21 Dec. Mexico City (G: II, 14r)
1537, 25 Feb. Mexico City (G: II, 14r)
1537, 20 June Mexico City (M, in Latin)

Replace common abbreviations and insert semicolons:

Find: Sept. M

Replace with: September; M

1534, 16 Sept. Mexico City;(G: I, 90r) 1535, 7 Sept. Mexico City;(G: I, 122r) 1536, 30 Nov. Mexico City;(G: II, 14r) 1536, 21 Dec. Mexico City;(G: II, 14r) 1537, 25 Feb. Mexico City;(G: II, 14r) 1537, 20 June Mexico City;(M, in Latin) Replace parentheses with semicolons:

Find: City;\(

Replace with: City;

1534, 16 September; Mexico City; (G: I, 90r) 1535, 7 September; Mexico City; (G: I, 122r) 1536, 30 November; Mexico City; (G: II, 14r) 1536, 21 December; Mexico City; (G: II, 14r) 1537, 25 February; Mexico City; (G: II, 14r) 1537, 20 June; Mexico City; (M, in Latin)

# <u>Find</u>: \)

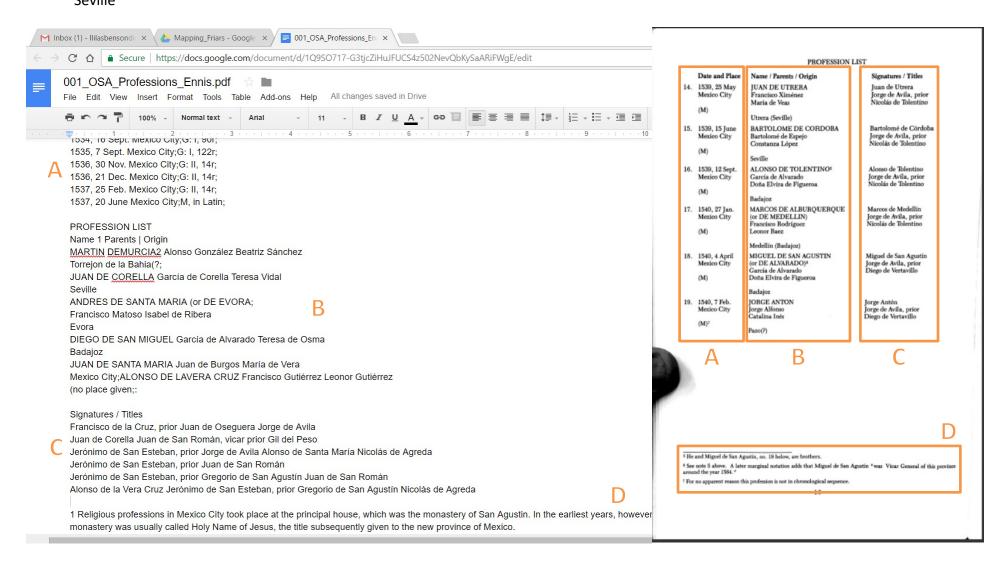
### Replace with: ;

1534, 16 September; Mexico City; G: I, 90r) 1535, 7 September; Mexico City; G: I, 122r) 1536, 30 November; Mexico City; G: II, 14r) 1536, 21 December; Mexico City; G: II, 14r) 1537, 25 February; Mexico City; G: II, 14r) 1537, 20 June; Mexico City; M, in Latin

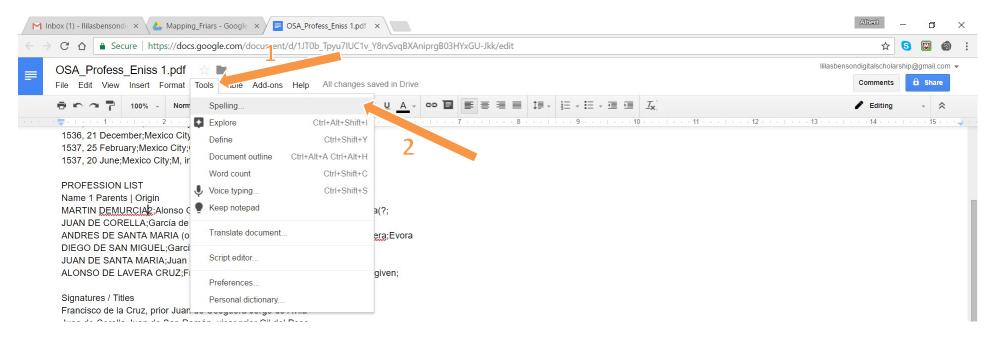
- In this example, Google extracted the data from the PDF column by column, instead of row/profession by row/profession.
- Our goal is to have all the data for one record (in this case, a friar's profession) in one line of text so that it will transform into a table row.
- In this example, the data from column A, C, & D corresponding to each record/profession is in a single line of text. Data from column B is not. We will consolidate the data related to each friar's profession in a line of text. Remember to separate the data values with a semicolon or |.
- Example:

MARTIN DEMURCIA2 Alonso González Beatriz Sánchez Torrejon de la Bahia(?; JUAN DE CORELLA García de Corella Teresa Vidal Seville MARTIN DEMURCIA2; Alonso González; Beatriz Sánchez; Torrejon de la Bahia (?;

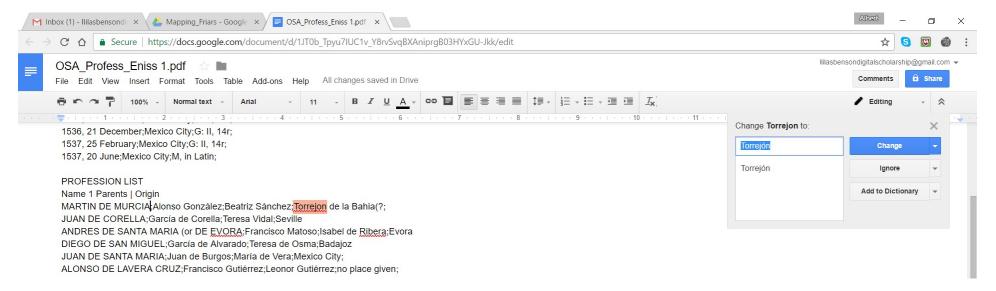
JUAN DE CORELLA: García de Corella: Teresa Vidal: Seville



• Check spelling by selecting **Tools**, then **Spelling...**.



Google autodetects language in your document to help spellcheck.



Insert semicolons between data values to signal table column changes.

Note: Do not enter spaces between data values and semicolons.

PROFESSION LIST

Name 1 Parents | Origin

MARTÍN DE MURCIA; Alonso González; Beatriz Sánchez; Torrejón de la Bahía(?);

JUAN DE CORELLA; García de Corella; Teresa Vidal; Seville

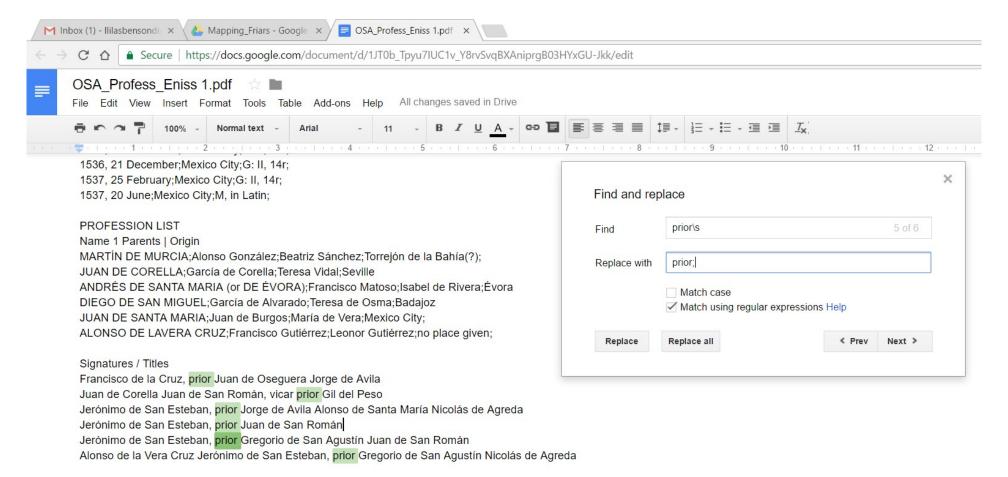
ANDRÉS DE SANTA MARIA (or DE ÉVORA); Francisco Matoso; Isabel de Rivera; Évora

DIEGO DE SAN MIGUEL;García de Alvarado;Teresa de Osma;Badajoz

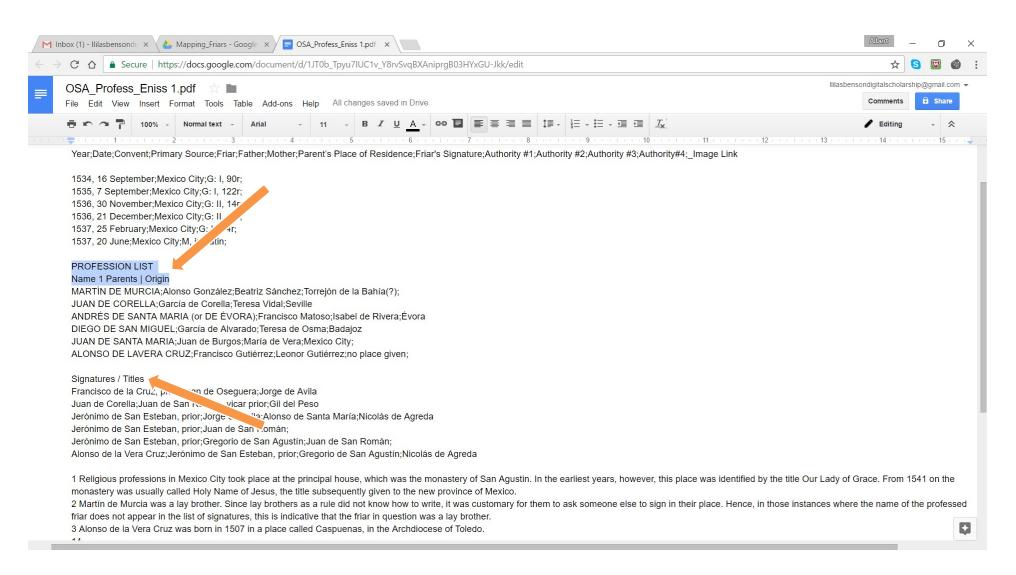
JUAN DE SANTA MARIA; Juan de Burgos; María de Vera; Mexico City;

ALONSO DE LAVERA CRUZ; Francisco Gutiérrez; Leonor Gutiérrez; no place given;

• Use Google Docs' <u>Find and replace</u> functionality to automate these semicolon insertions. In this example, "prior" is commonly the last word in a data value.

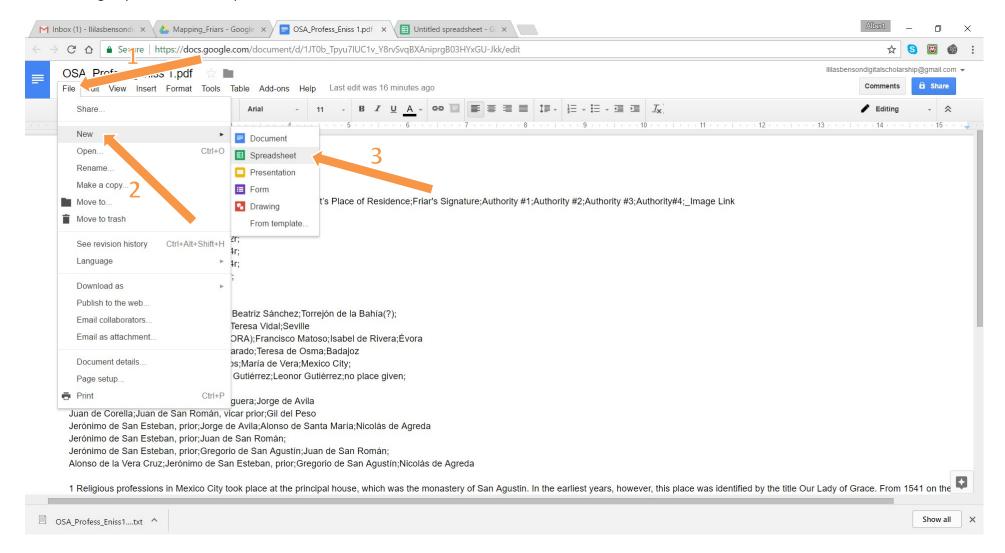


• Delete the section titles (ex. "Date and Place", "PROFESSION LIST Name 1 Parents | Origin", and "Signatures / Titles") to complete the data clean up.

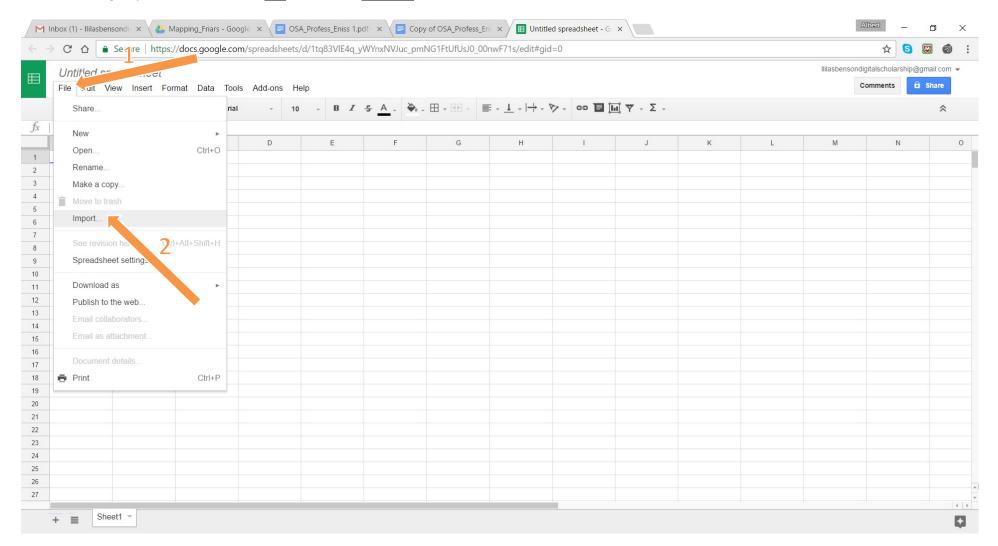


## Transforming the Google Doc into a Google Sheet

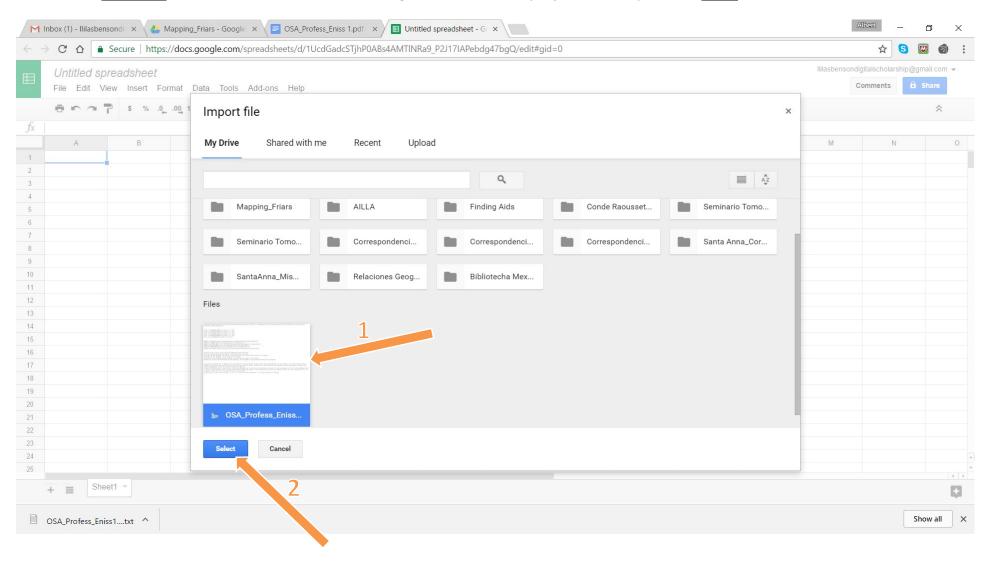
- Click on File, hover over New, and select Spreadsheet.
- Google Spreadsheets will open in a new tab.



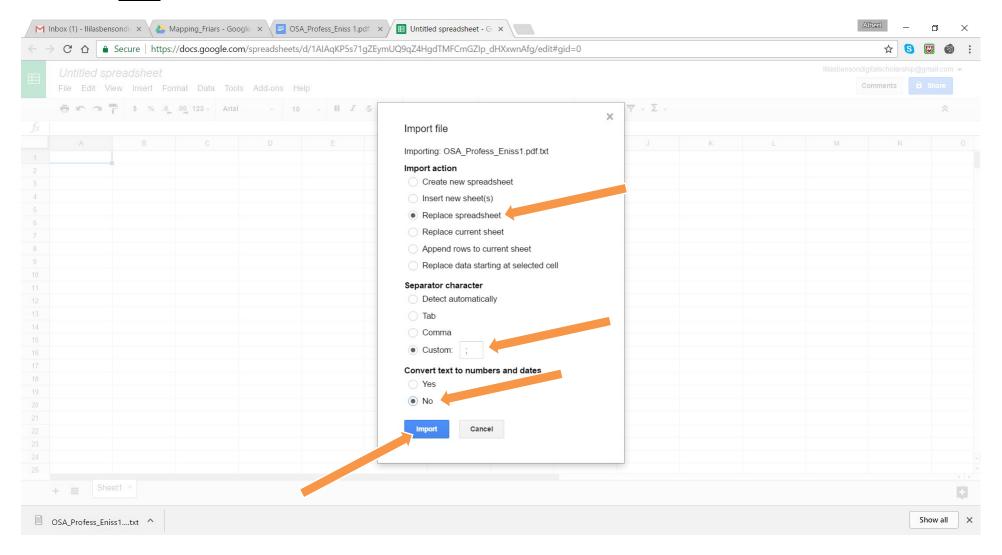
• In Google Spreadsheets, click on File, then select Import....



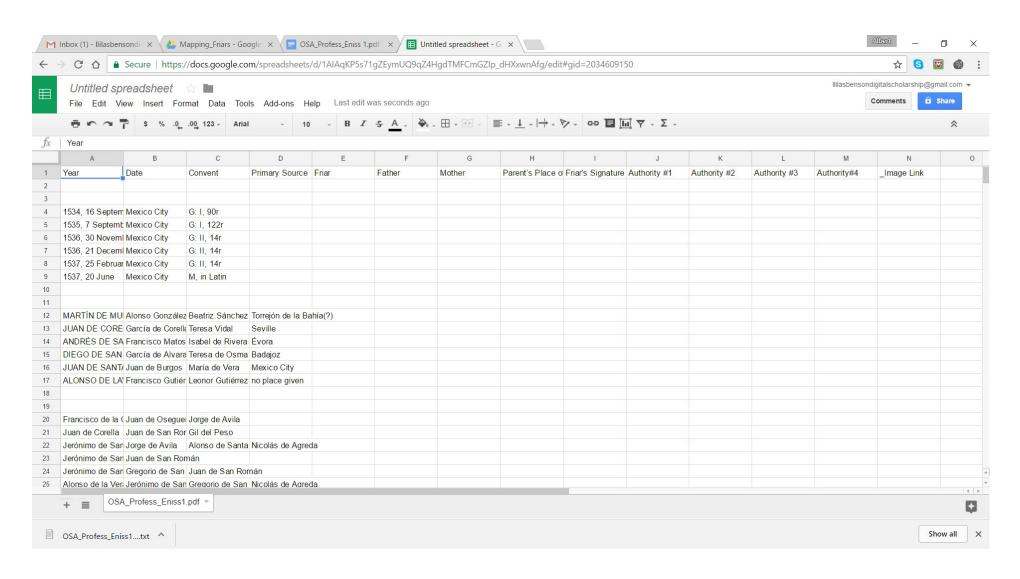
• In the **Import file** window, search for and select the Google Doc with the data you just cleaned up. Click on **Select**.



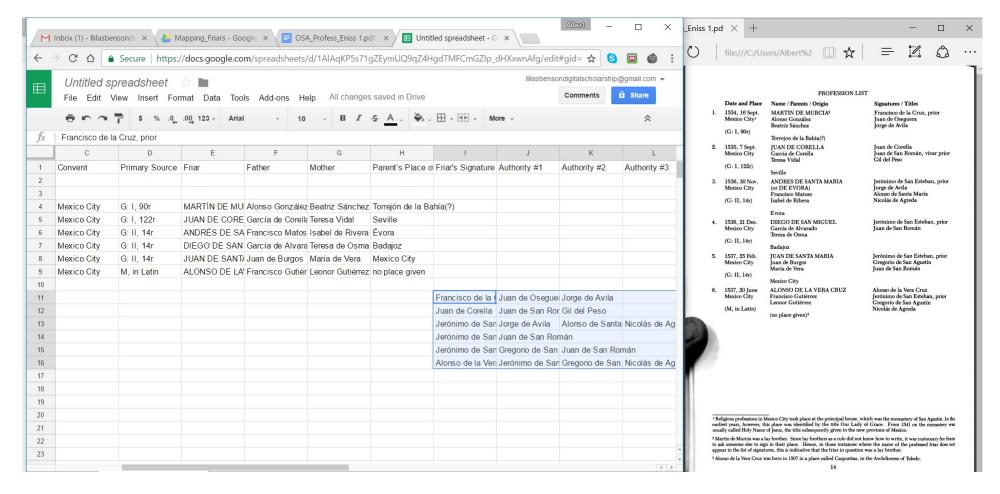
- In the **Import file** window, select:
  - o Import action: Replace spreadsheet
  - o <u>Separator character</u>: Custom: [depending on which symbol you chose to separate your column values, enter a semicolon or | ]
  - o Convert text to numbers and dates: No
- Click on Import.



• Your imported data should look like this:



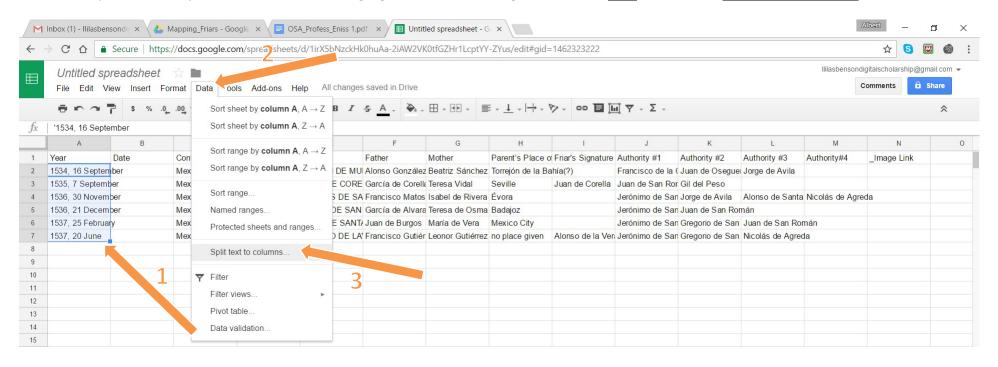
- Select and move the sections of cell values so that it aligns with the corresponding column.
- Remember, the goal is to have all the data relating to a record (in this example, a friar's profession) in a single table row.
- For this example, we will delete the footnotes at the bottom of the spreadsheet.



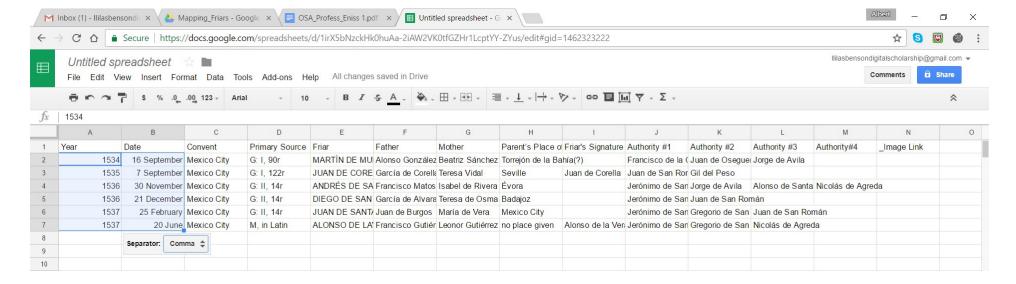
#### Result

|   | А                  | В    | С           | D              | E                         | F                 | G                | Н                | 1                      | J               | K                 | L                              | M             |
|---|--------------------|------|-------------|----------------|---------------------------|-------------------|------------------|------------------|------------------------|-----------------|-------------------|--------------------------------|---------------|
| 1 | Year               | Date | Convent     | Primary Source | Friar                     | Father            | Mother           | Parent's Place   | Friar's Signature      | Authority #1    | Authority #2      | Authority #3                   | Authority#    |
| 2 | 1534, 16 September |      | Mexico City | G: I, 90r      | MARTÍN DE MURCIA          | Alonso González   | Beatriz Sánchez  | Torrejón de la B | ahía(?)                | Francisco de la | Juan de Osegue    | Jorge de Avila                 | 11-20-11-20   |
| 3 | 1535, 7 Septem     | ber  | Mexico City | G: I, 122r     | JUAN DE CORELLA           | García de Corella | Teresa Vidal     | Seville          | Juan de Corella        | Juan de San Ro  | r Gil del Peso    |                                |               |
| 4 | 1536, 30 Noven     | nber | Mexico City | G: II, 14r     | ANDRÉS DE SANTA MARIA (or | Francisco Matos   | Isabel de Rivera | Évora            |                        | Jerónimo de Sar | Jorge de Avila    | de Avila Alonso de Santa María |               |
| 5 | 1536, 21 Decen     | nber | Mexico City | G: II, 14r     | DIEGO DE SAN MIGUEL       | García de Alvara  | Teresa de Osma   | Badajoz          |                        | Jerónimo de Sar | Juan de San Román |                                |               |
| 6 | 1537, 25 Februa    | ary  | Mexico City | G: II, 14r     | JUAN DE SANTA MARIA       | Juan de Burgos    | María de Vera    | Mexico City      |                        | Jerónimo de Sar | Gregorio de San   | Juan de San Ro                 | or Nicolás de |
| 7 | 1537, 20 June      |      | Mexico City | M, in Latin    | ALONSO DE LAVERA CRUZ     | Francisco Gutiér  | Leonor Gutiérrez | no place given   | Alonso de la Vera Cruz | Jerónimo de Sar | Gregorio de San   | Nicolás de Agre                | eda           |
|   |                    |      |             |                |                           |                   |                  |                  |                        |                 |                   |                                |               |

• To separate out the year from the date, select/highlight the cells containing dates, click on **Data**, then click on **Split text to columns...**.

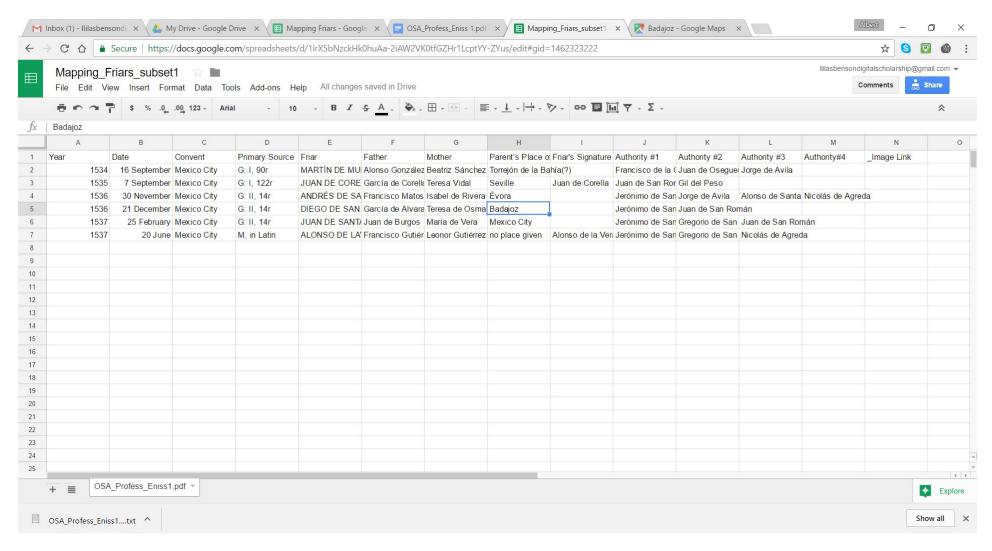


• Google Spreadsheets should detect the comma as the separating character and move the day and month over to the next column. If not, set **Separator** to **Comma**.

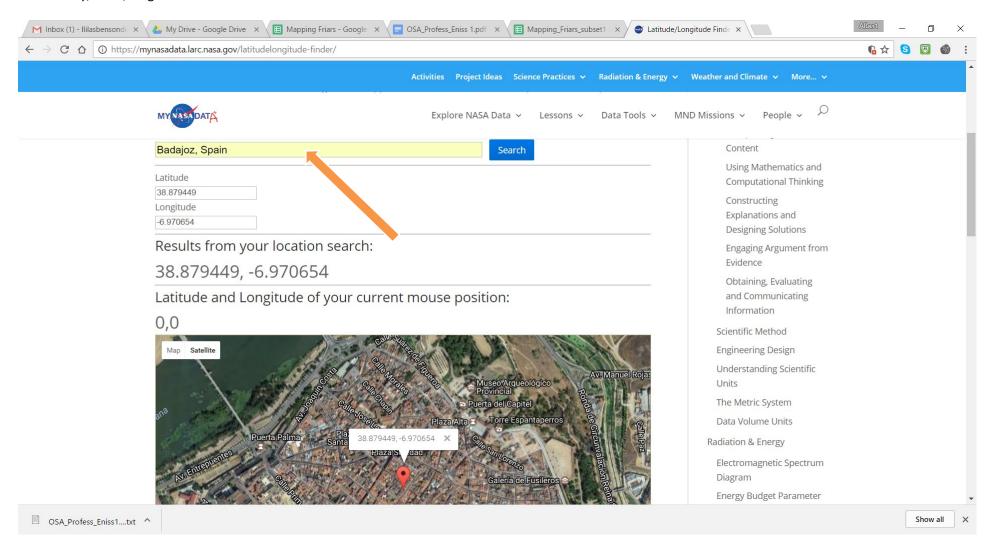


### **Adding GIS Coordinates to Map Records**

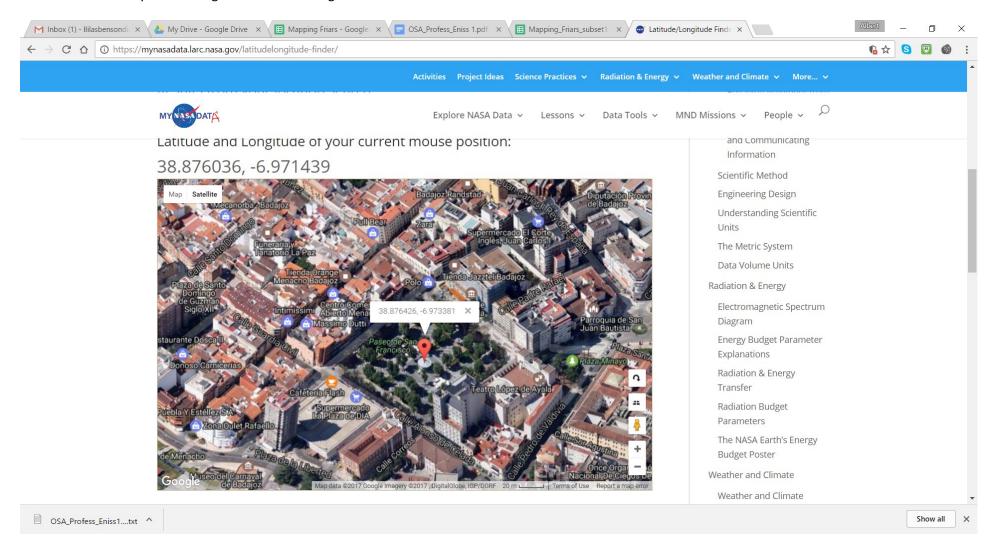
- GIS coordinates (latitude and longitude values) of the places in your profession records/rows are needed to map the information.
- For this example, we will be mapping the "Parent's Place of Origin" values.



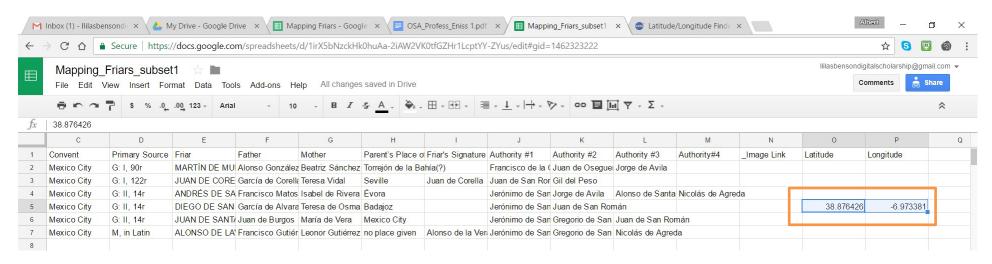
- Navigate to https://mynasadata.larc.nasa.gov/latitudelongitude-finder/
- Enter the place name you wish to find GIS coordinates for in the **Search** entry box. This will provide you the general coordinates for this city/town/neighborhood.



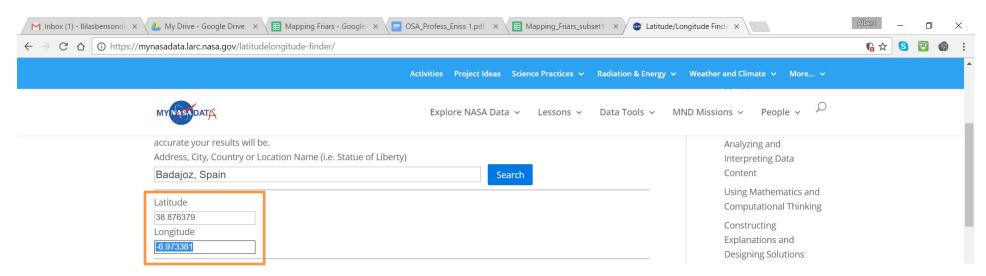
• If you wish to find the coordinates of a much more specific place, zoom in and move the map/satellite image under the coordinate information and click on the area you wish to get latitude and longitude values.



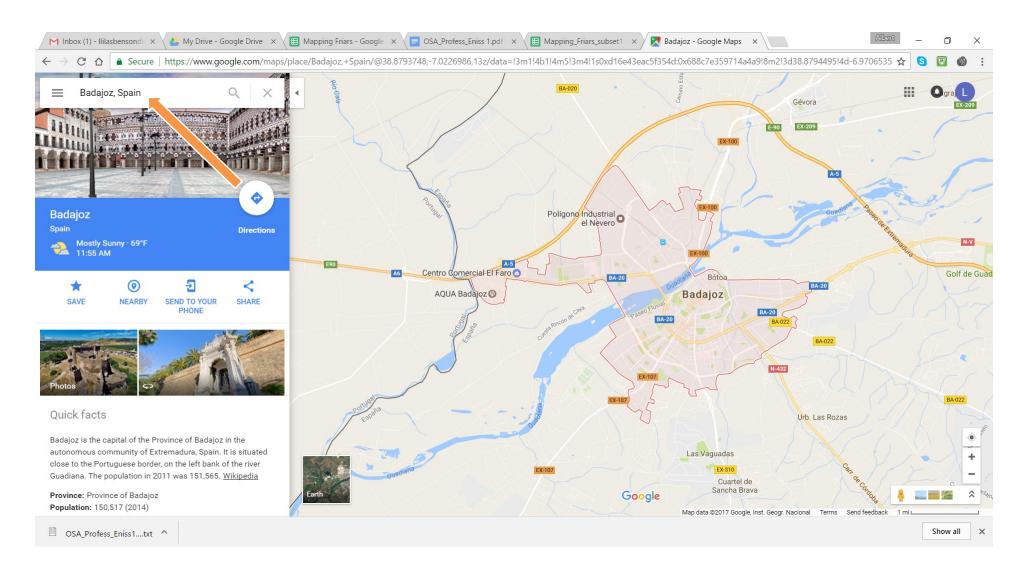
• Create a Latitude and Longitude column in your Google Spreadsheet.



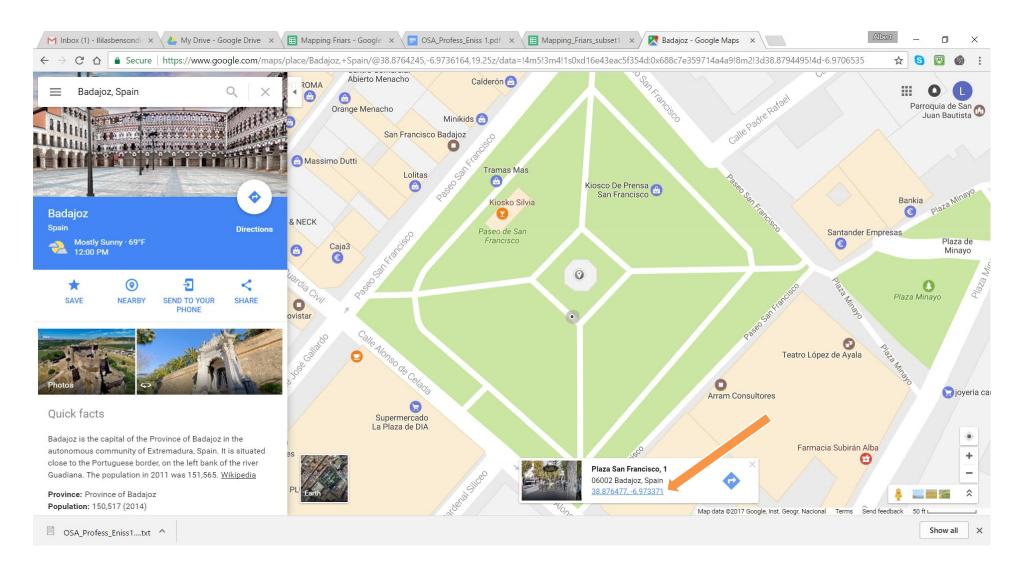
• Copy the latitude and longitude values from *MyNASAData+* and paste them into the corresponding row in your Google Spreadsheet.



• You can also find GIS coordinates using Google Maps. Search for the city/town/place.

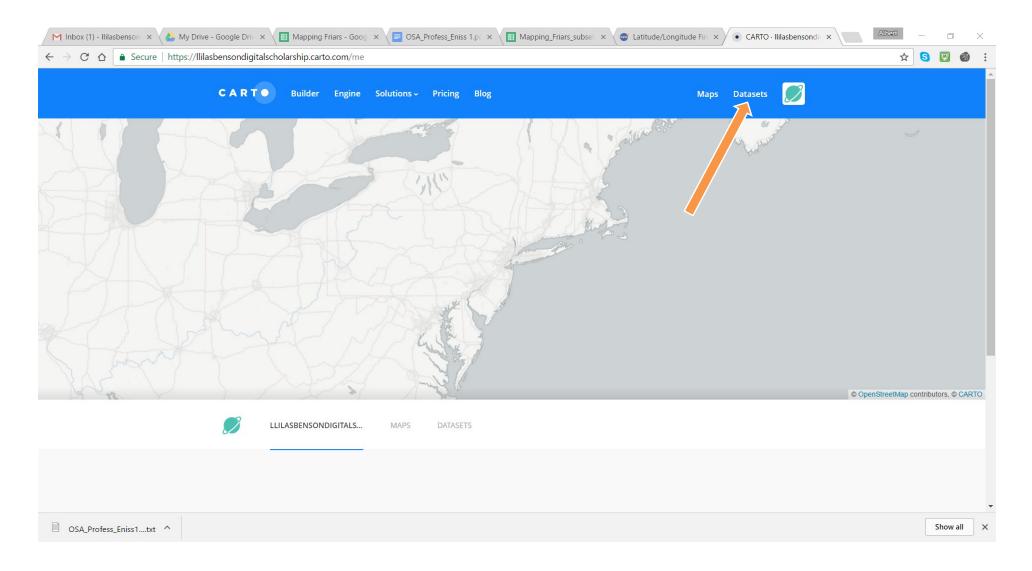


• If you click on a specific place, latitude and longitude coordinates will appear at the bottom of the screen.

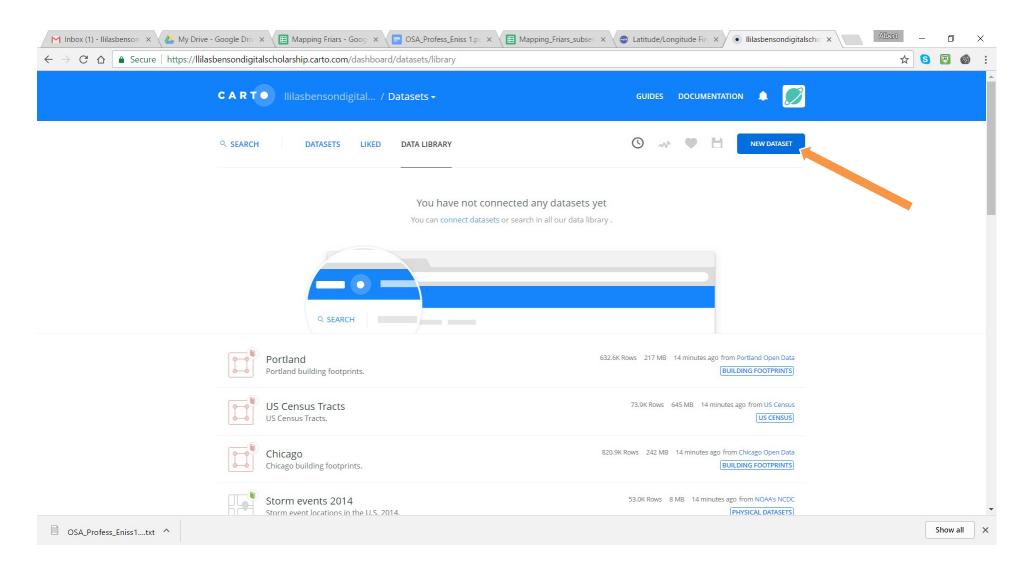


## **Importing Data Set into Carto**

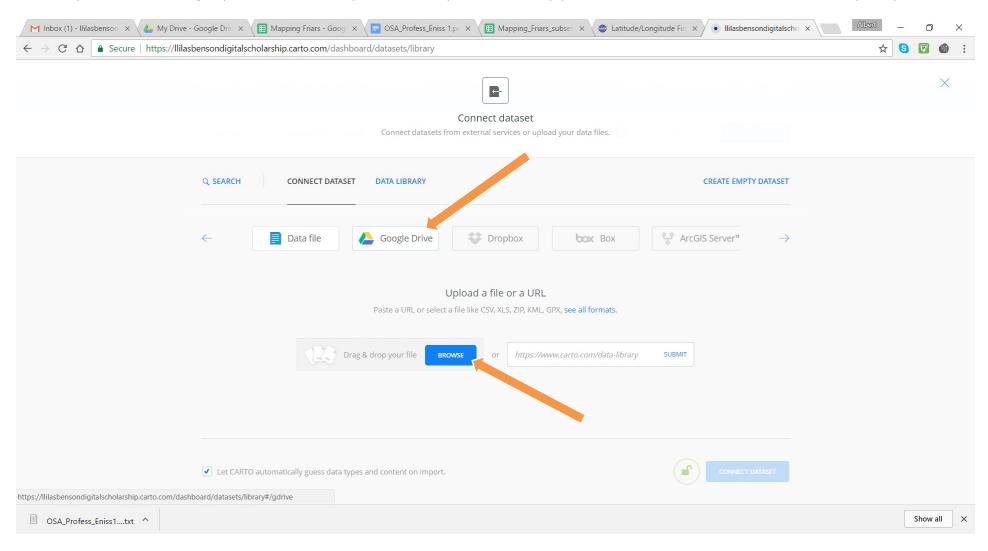
- Log into Carto. It is recommended that you login using your Gmail account so that Carto can directly access your Google Spreadsheet.
- On your dashboard, click on **Datasets**.



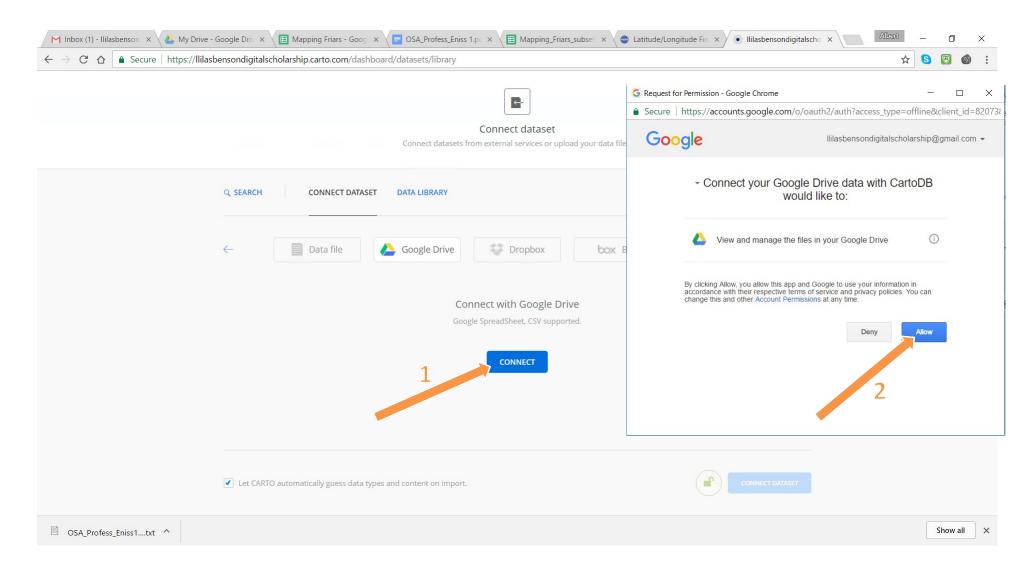
## • Click on New Dataset.



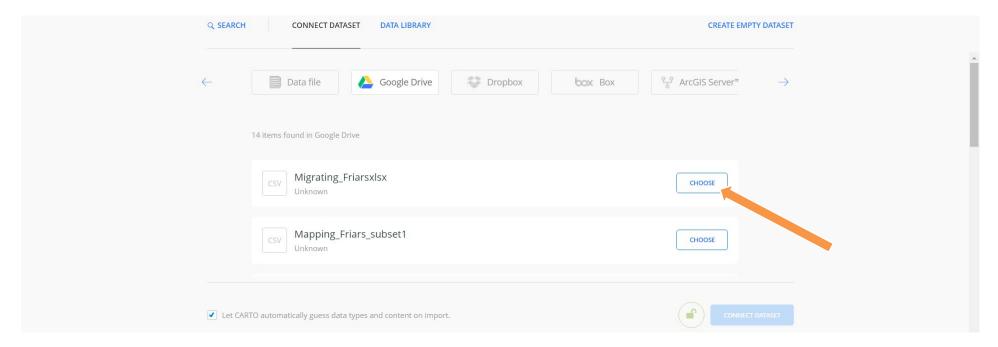
- Click on Google Drive.
- If you did not use Google Spreadsheet to create your dataset or you did not link up your Gmail account to Carto, click on Browse to upload your dataset.



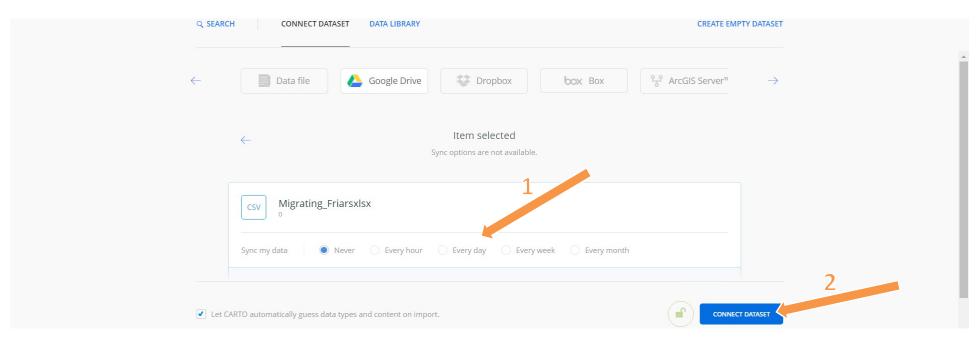
- Click on Connect.
- Grant Carto permission to access your Google Drive files by clicking on **Allow**.



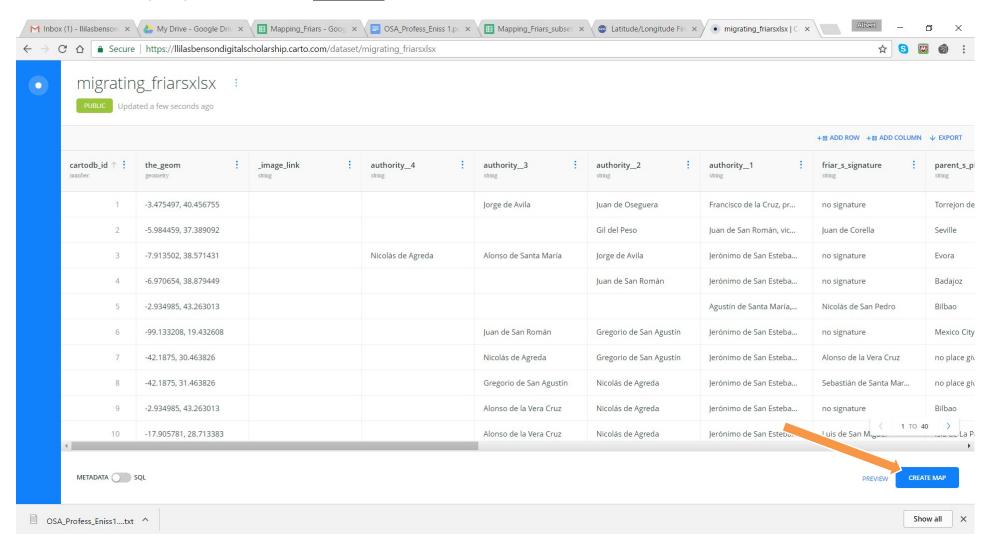
• Carto will automatically identify file types it can import. Click on **Choose** to select the one you wish to import.



• Select how often you want Carto to update the dataset from Google Drive, then click on **Connect Dataset**.

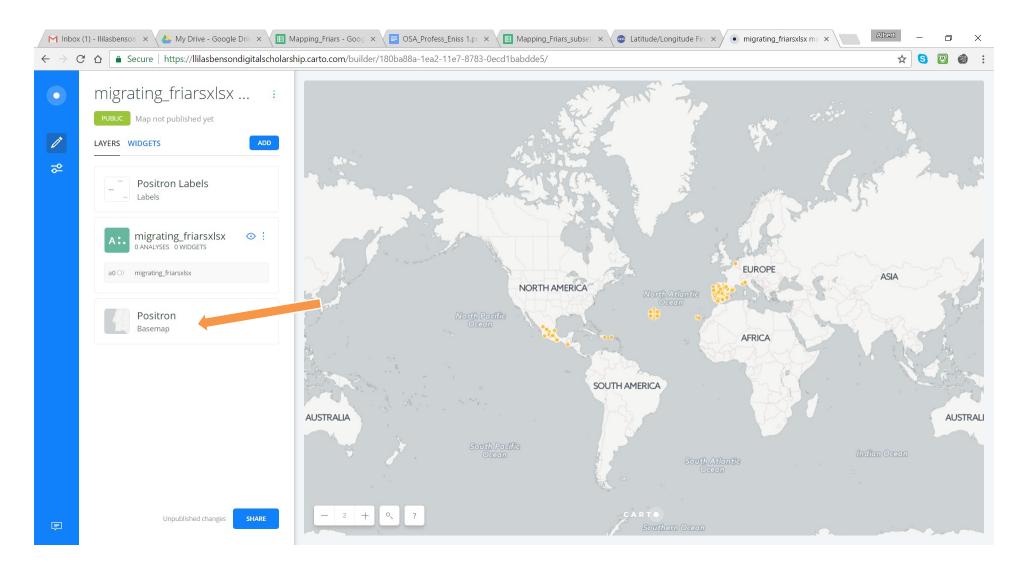


• After Carto imports your dataset, click on Create Map.

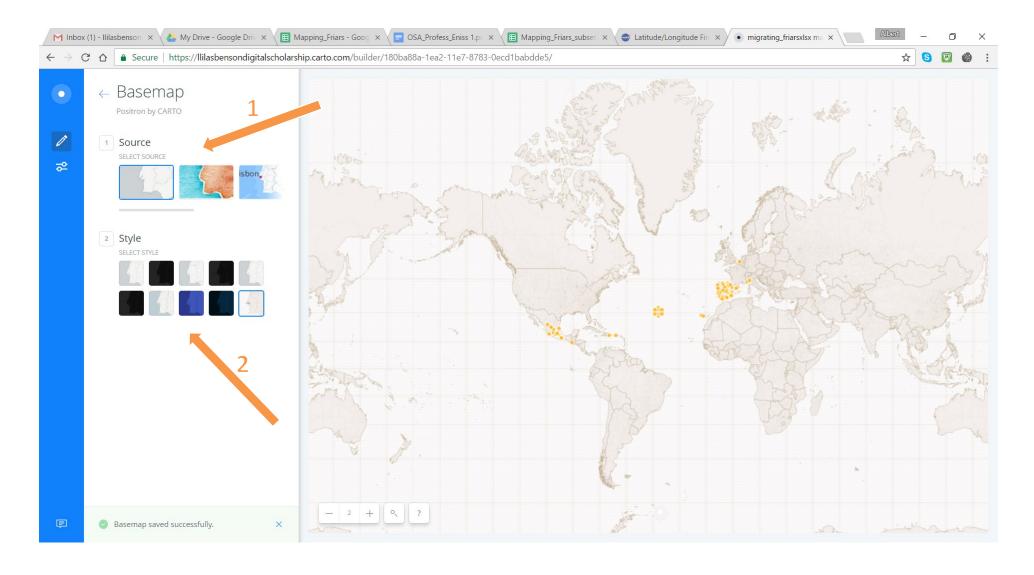


## **Customizing Carto Map**

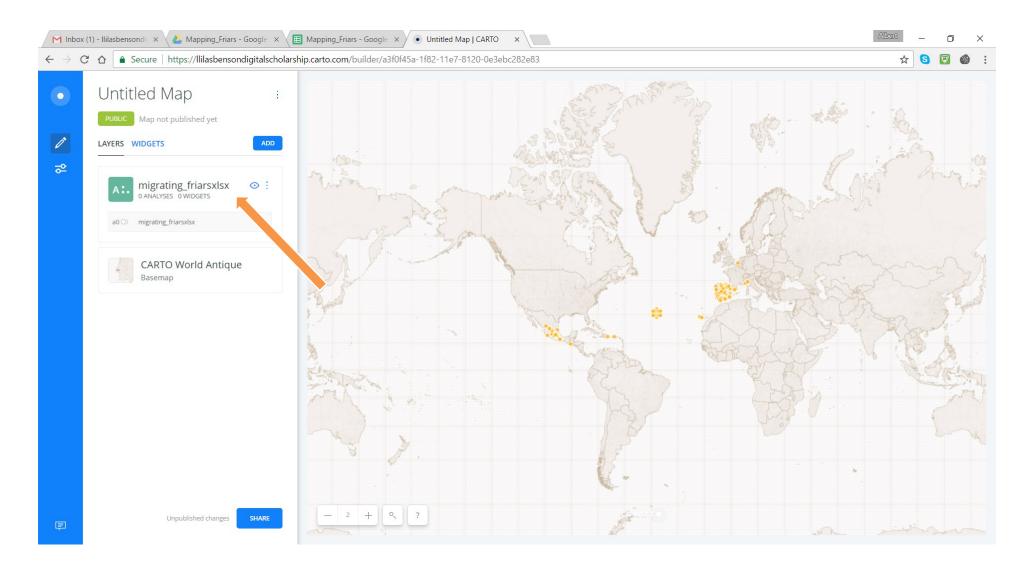
- Note: The remainder of this guide will be using a full dataset derived from the Ennis publication. You can find this dataset here.
- To change the style of the background map, click on <u>Positron: Basemap</u>.



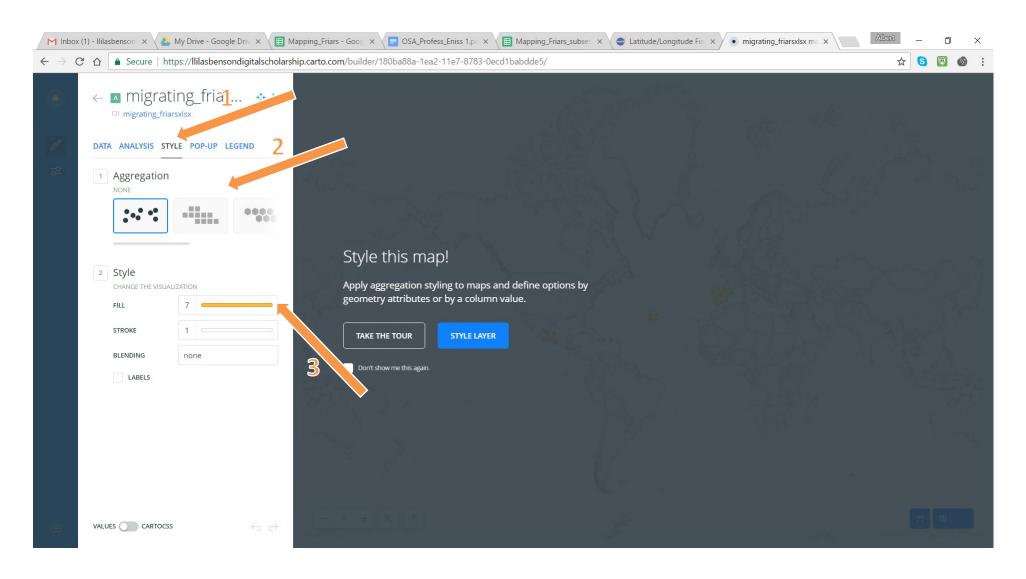
- Here you can select the **Source** and **Style**.
- Click on the  $\underline{\leftarrow}$  to apply the style and return to the previous menu.



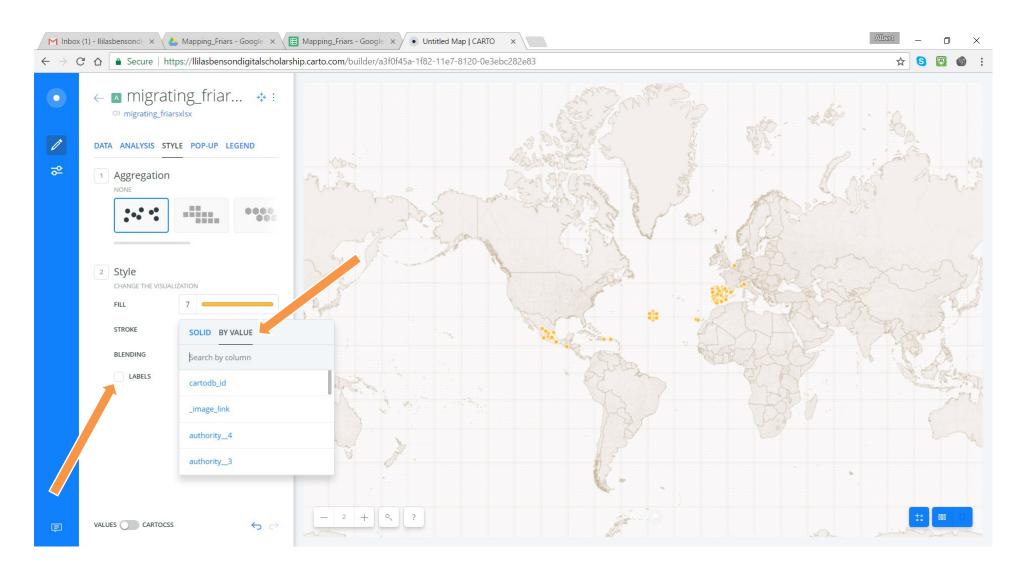
• Click on the layer with the dataset you imported to start manipulating the data.



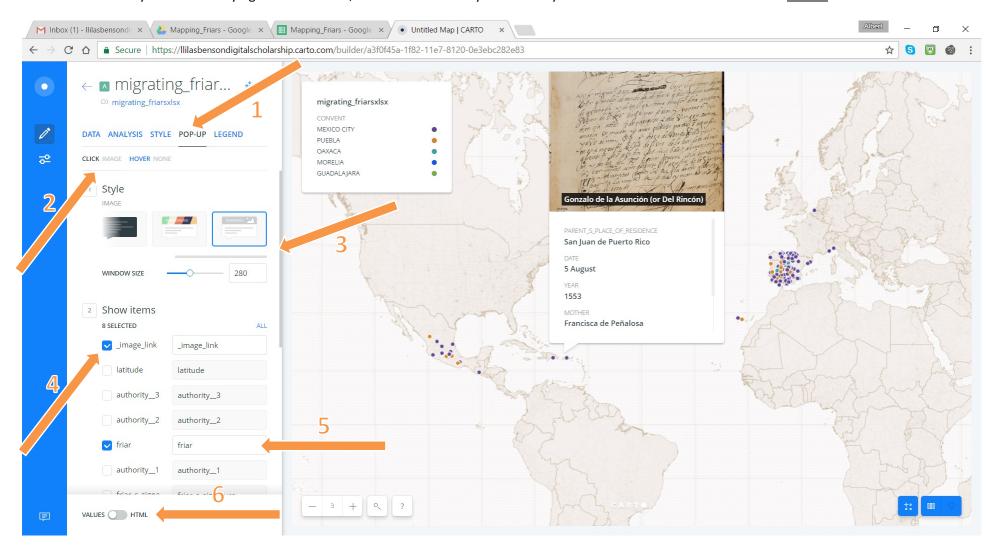
- The layer opens with the <u>Style</u> tab selected.
- Under **Aggregation**, you can choose to aggregate the data points and how you would like the aggregations rendered.
- For this example, we will not aggregate the data so that each point on the map represents records with that specific GIS coordinate.
- To differentiate the points on the map with color, click on the color bar next to Fill.



- If you click on **By Value**, you can differentiate the points with color based on the column values.
- For this example, we will select the "Convent" column. Once you select a column name, Carto will give you the option to change the colors for each unique data value under the "Convent" column.
- Once you have selected a column and applied your color scheme, click outside the menu box to apply it.
- You also have the option to add labels and customize them by checking the **Label** box.



- To enable the "pop-up" feature when a point is hovered over or clicked on by the user, click on the **Pop-Up** tab.
- Under the main tabs, you can select if you want a pop-up to come up when the user hovers over or clicks on the data point.
- Select the **Style** of the pop-up window and the **Window Size**.
- Select the data values you would like to show in the pop-up by checking the boxes under **Show Items**. You can rearrange the order of the displayed data by clicking and dragging the column names up or down. If you selected **Image** under **Style**, the "\_image\_link" column heading should come first. You can also edit the data descriptor in the entry text boxes next to each column name.
- We will modify how this information is displayed a bit more by switching from the **Values** to the **HTML** screen.
- Note: once you start modifying the HTML code, Carto will not allow you to modify the data values and their order in the <u>Values</u> view.

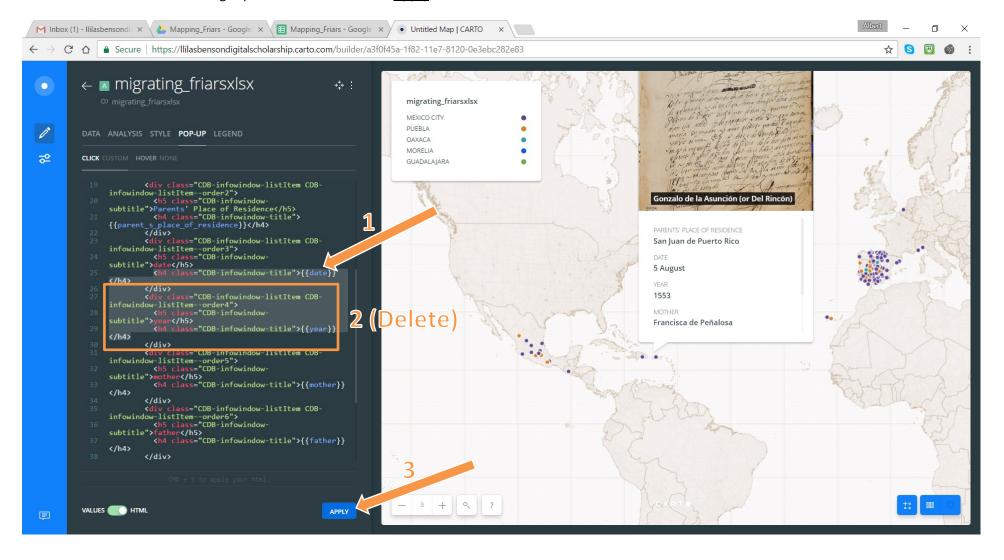


- First, we will bring the "Date" and "Year" data together to tidy up the information in the pop-up.
- Scroll down on the HTML code until you find the piece that brings up the "Date" and "Year" data. This is the relevant code for the "Date" field (apply the highlighted changes):

### **Before**

Delete the code that brings up the "Year" and click on Apply.

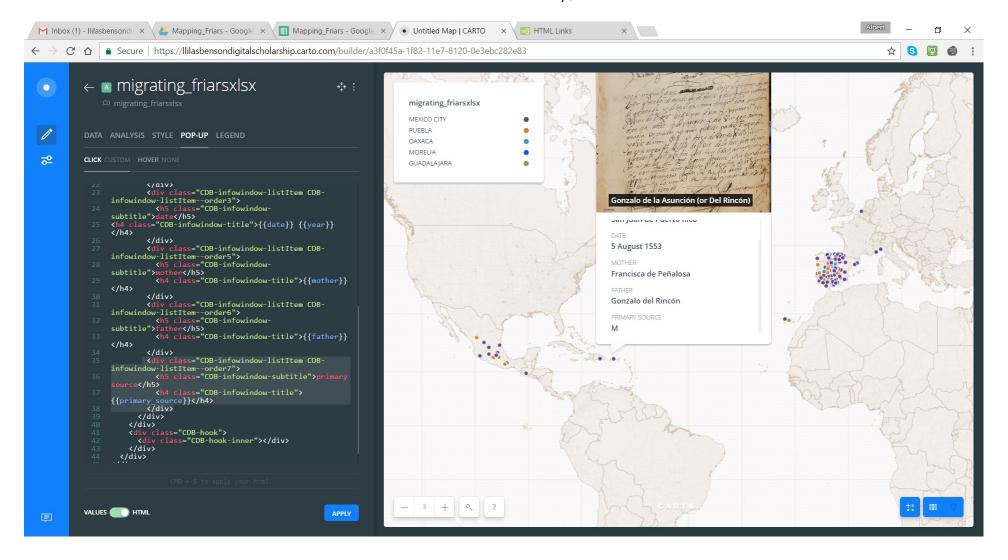
## After



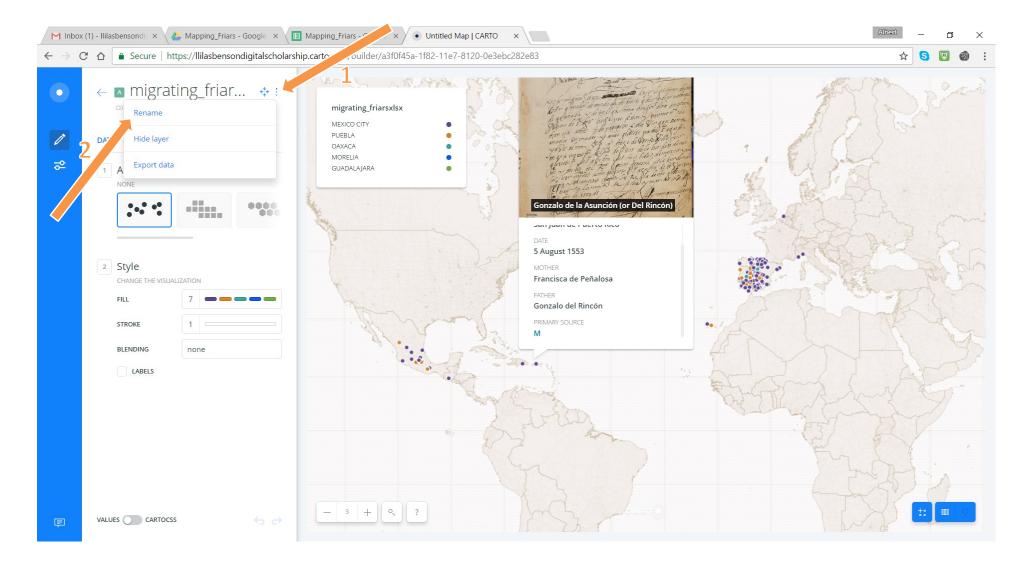
• To hyperlink the "Primary Source" data to a full image of the document (also being used as the pop-up header image), scroll down to the relevant code for the "Primary Source" data.

### **Before**

### After

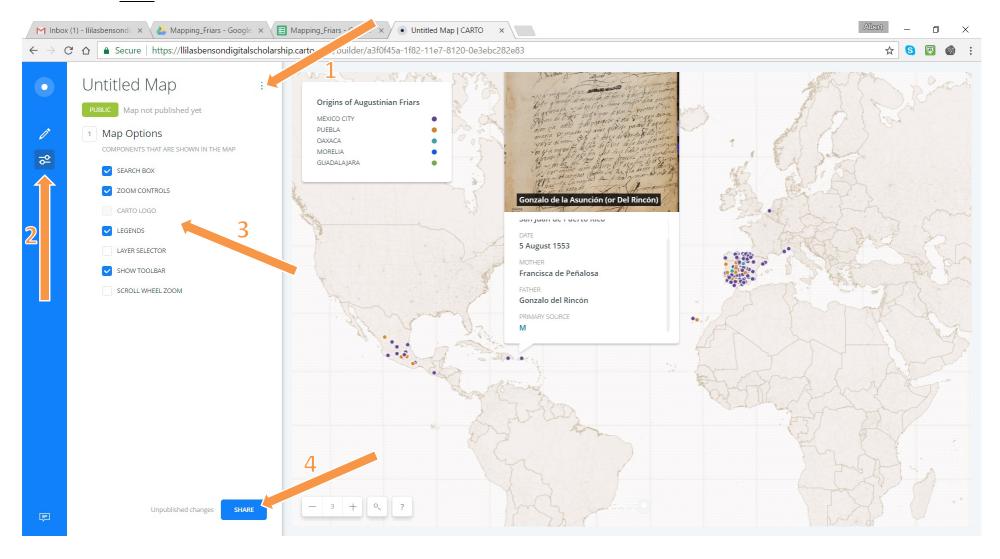


- To change the name of the layer (which is reflected on the Legend), click on the vertical ellipsis icon next to the name and select **Rename**.
- Enter the name in the entry field that comes up and hit the **Enter** key.



## **Publishing the Map**

- Click on the icon below the pencil icon to the far left.
- To name your map, click on the vertical ellipsis icon next to "Untitled Map" and select **Rename**. Enter the name in the entry field that comes up and hit the **Enter** key.
- Select the Map Options you would like for your audience.
- Click on <u>Share</u>.



• Click on **Publish**, and share the link Carto provide

