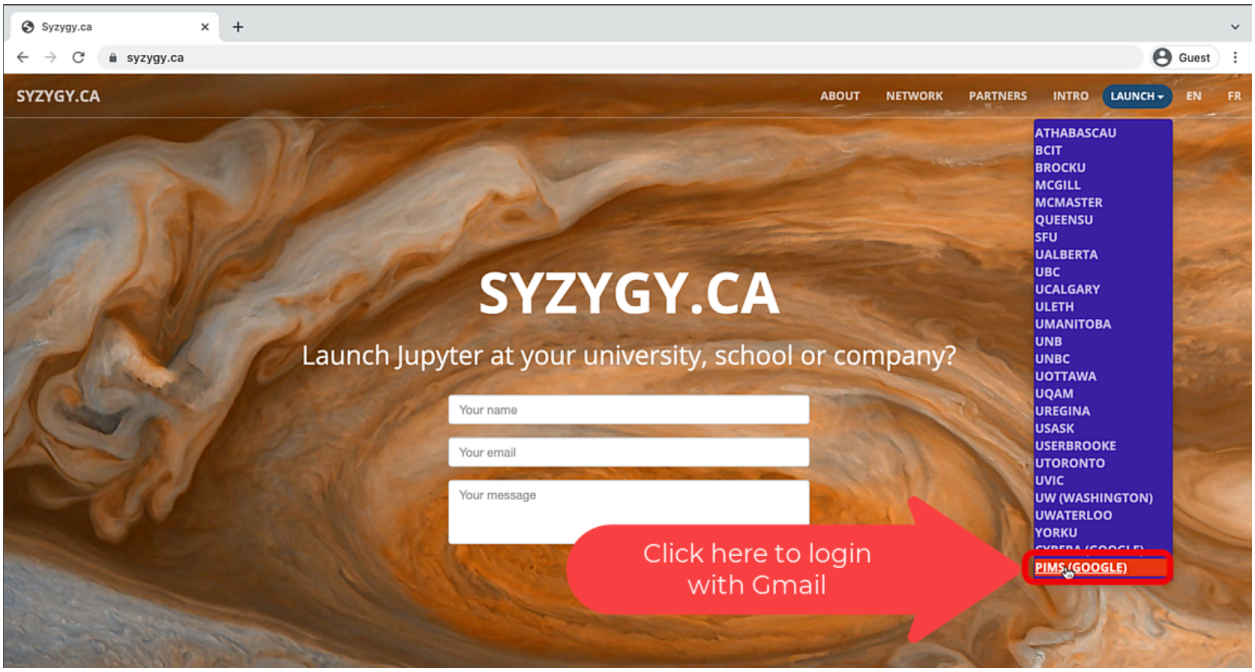
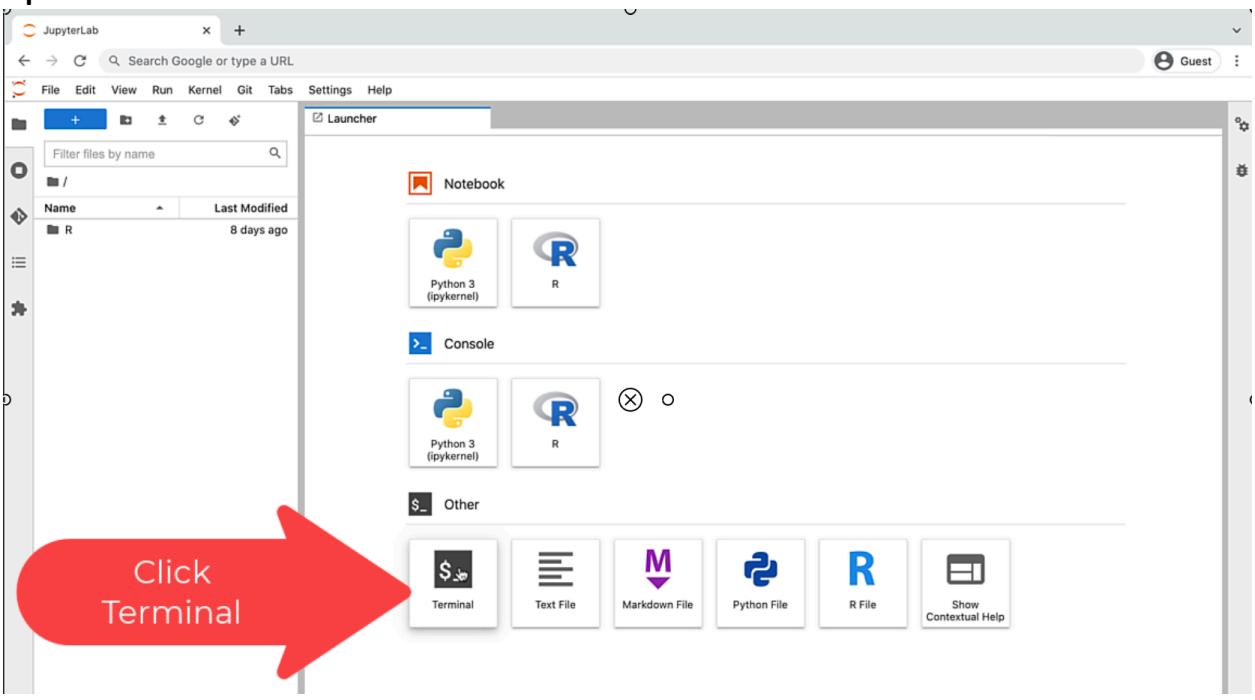
Scholar Metrics Scraper Instructions

Step 0: Setup

Go to syzygy.ca, click Launch, and either sign in with your Gmail or institutional account.

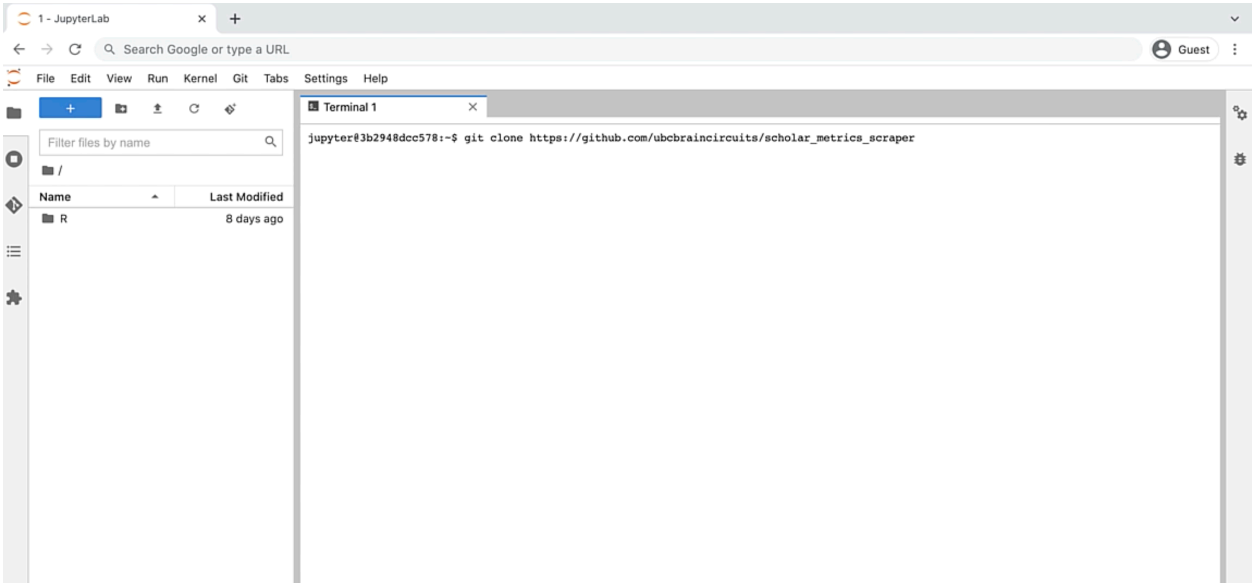


Open Terminal

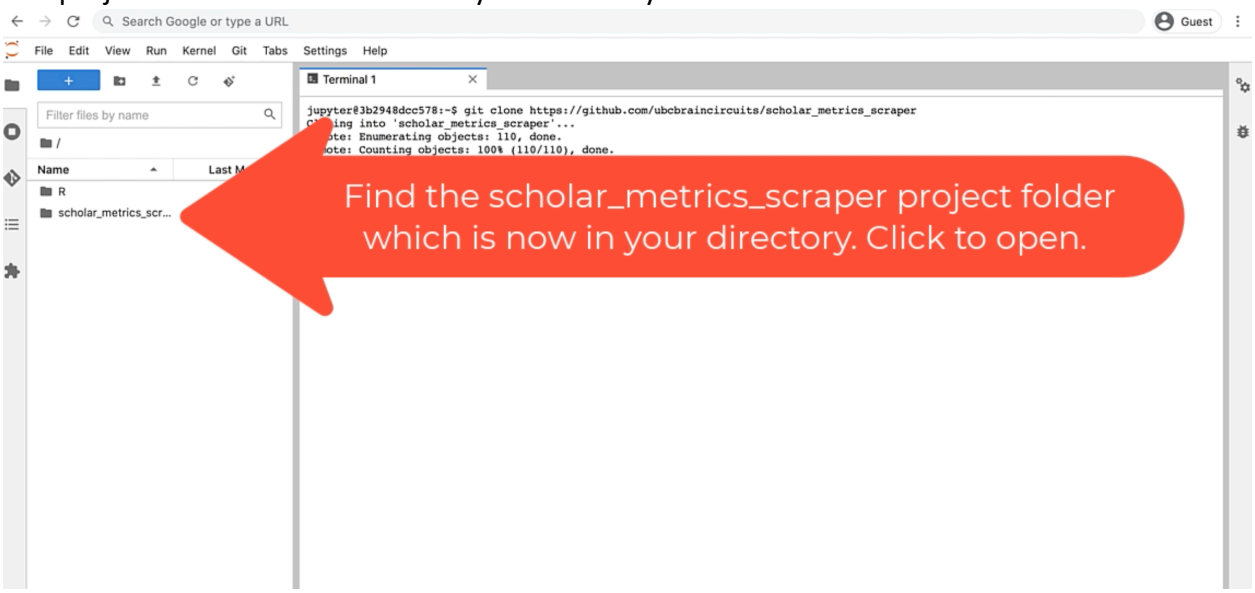


Step 1: Clone project from GitHub

Clone the project from Git. Type "git clone https://github.com/ubcbraincircuits/scholar_metrics_scraper" and press enter.

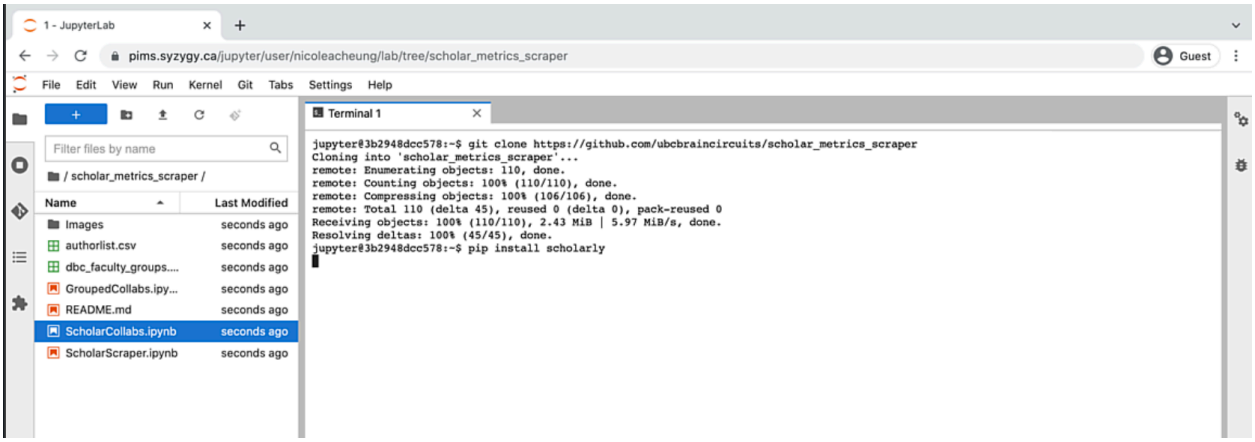


The project folder should now be in your directory

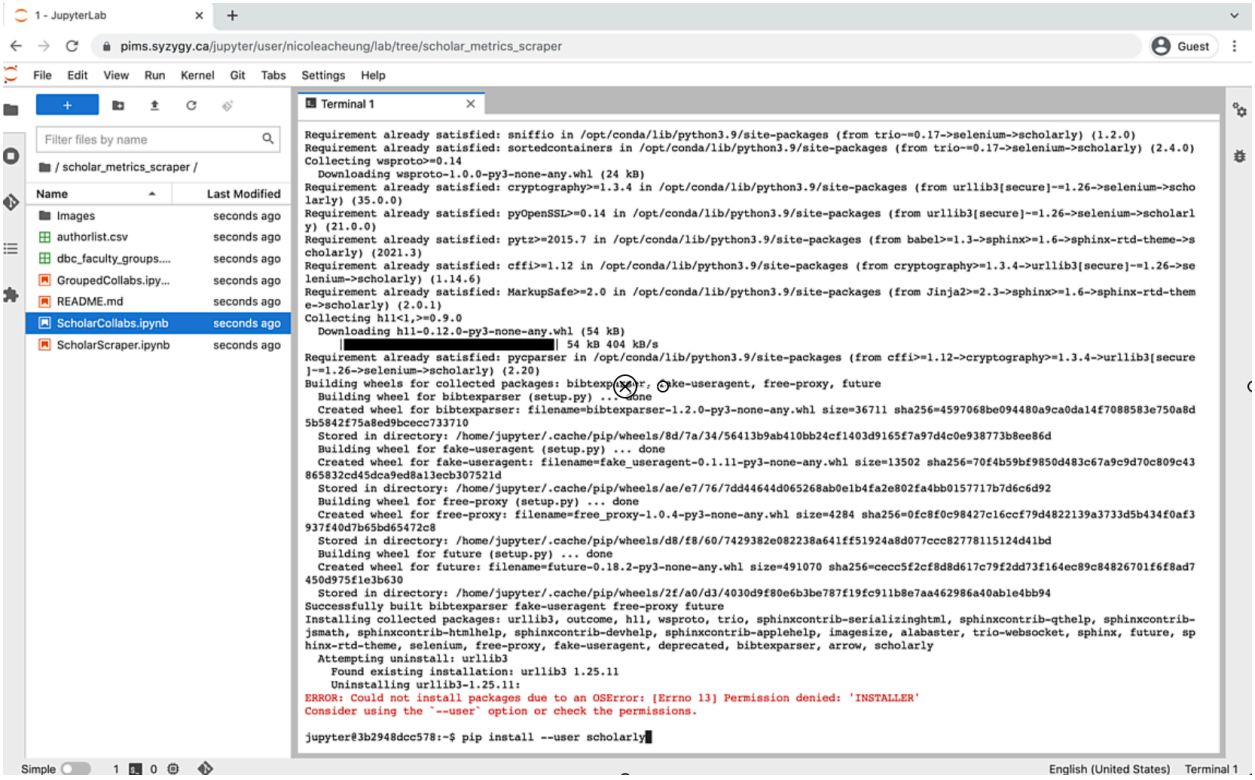


Step 2: Install scholarly

Install scholarly. Type “pip install scholarly” and press enter.



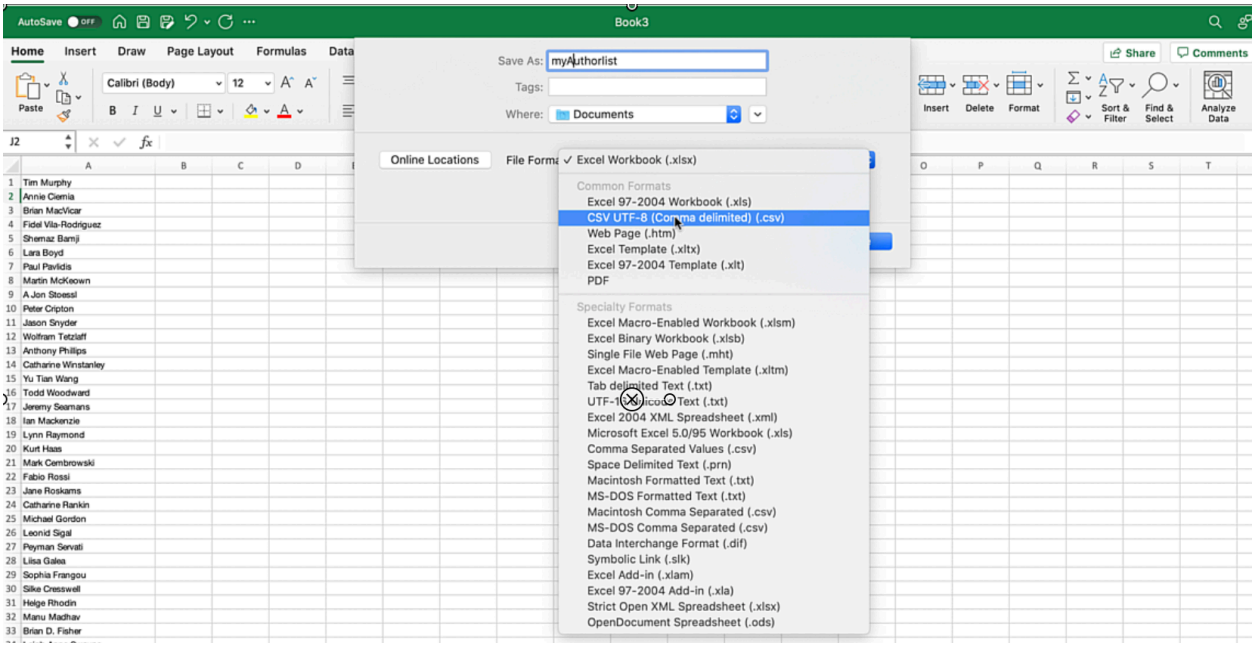
- If you receive an error, type “pip install --user scholarly” and press enter.



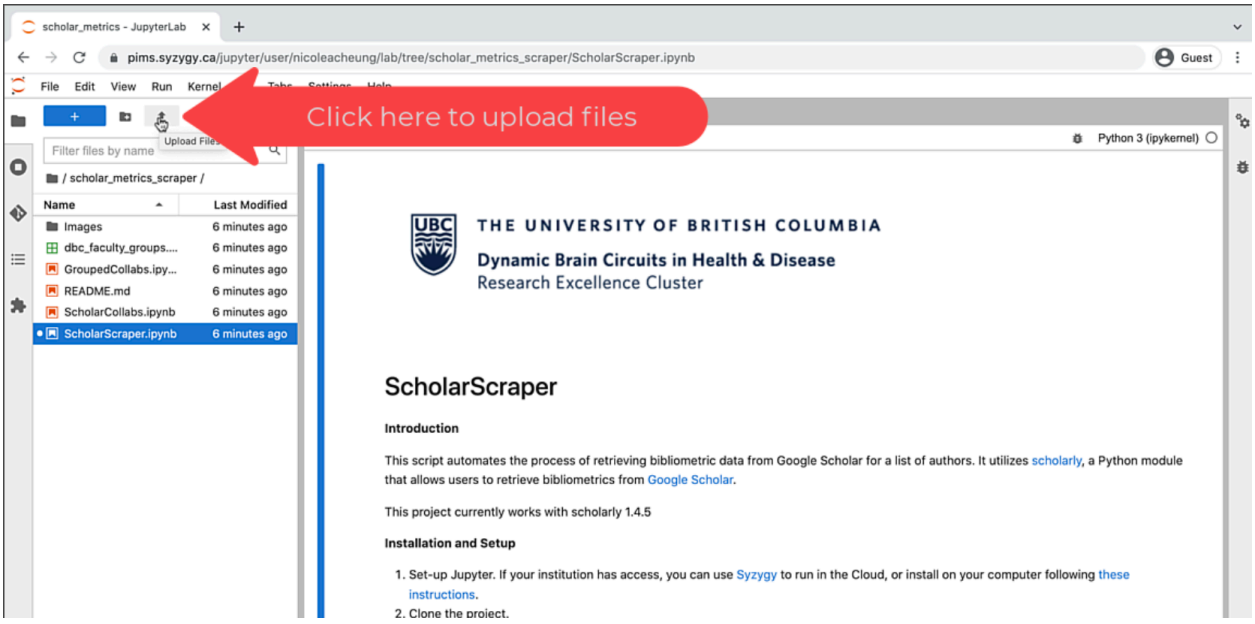
Step 3: Upload author list

Create a CSV file with a list of author names and/or Scholar IDs in a single column. Upload this to your project directory.

- Save the author list as a CSV file

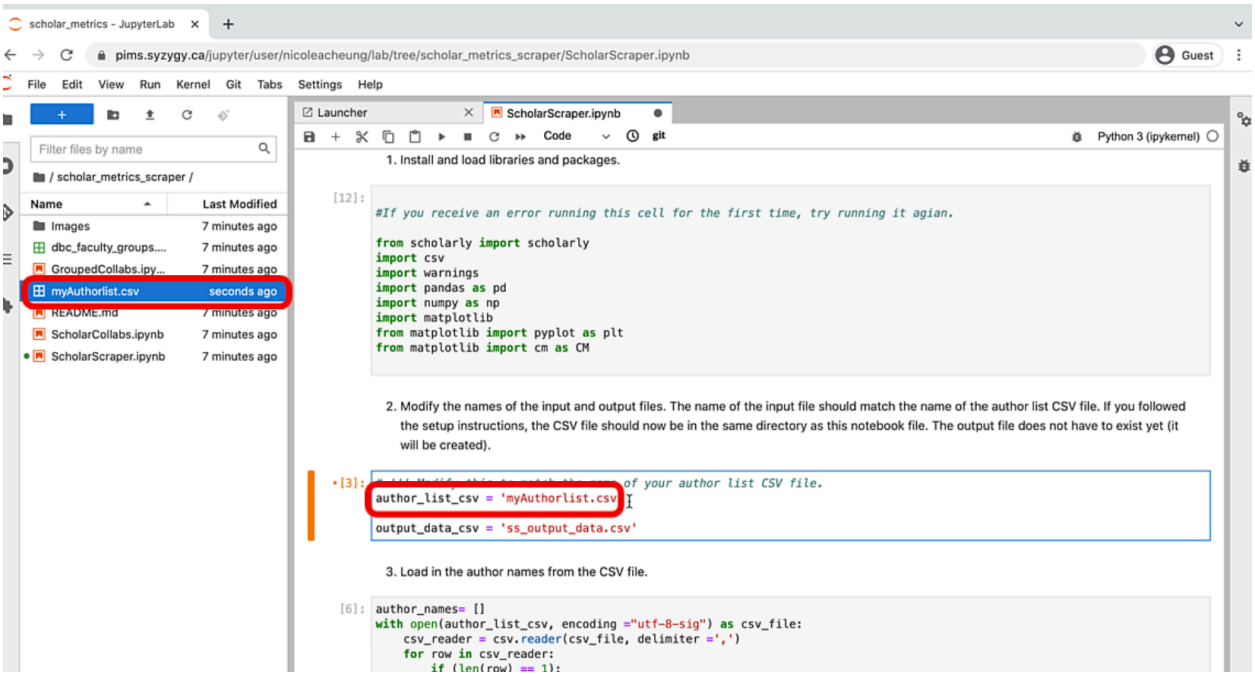


- Upload the author list CSV to the project directory

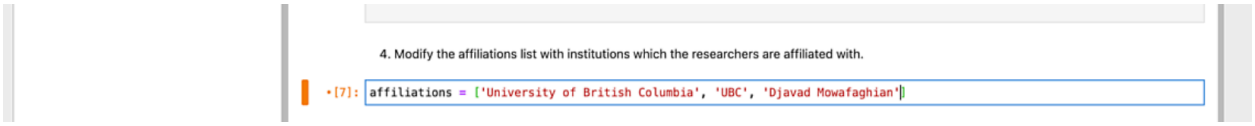


Make some modifications in the ScholarScraper notebook.

- Modify author_list_csv to match the author list CSV filename which you just uploaded.



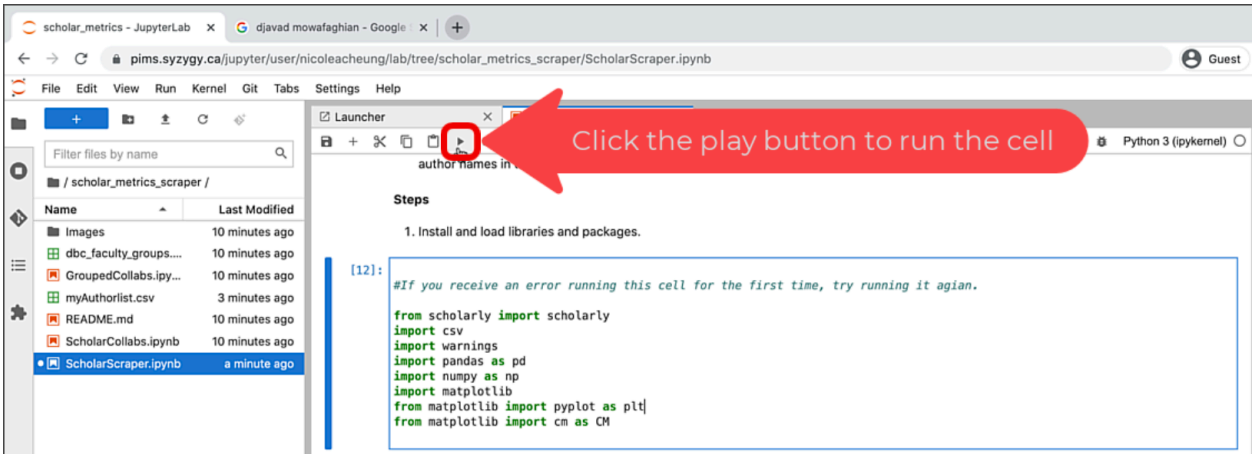
- Modify the affiliations. Add institution names which you expect the authors to be affiliated with on Google Scholar. Authors who don't have any of these affiliations will be flagged.



Step 4: Run ScholarScraper (this step continues after step 5)

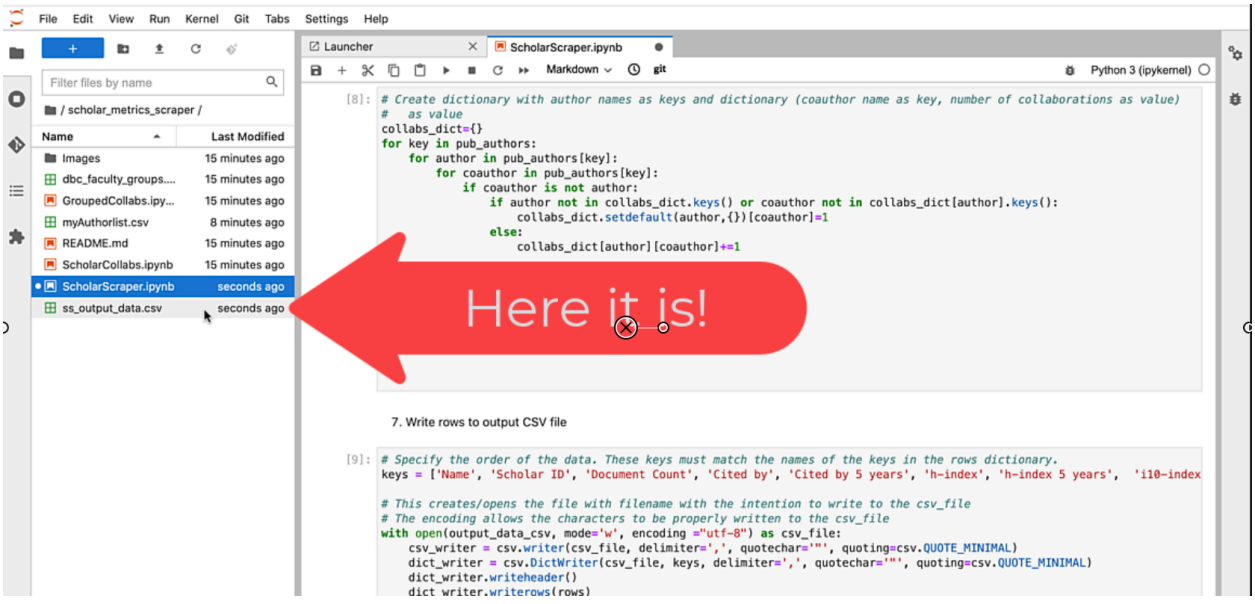
Run the ScholarScraper notebook code cells 1-7.

- You can run a cell by clicking anywhere inside the cell and either clicking the play button or pressing shift + enter.



- Cell number 5 (which retrieves the data for each author from Google Scholar) may take several minutes.

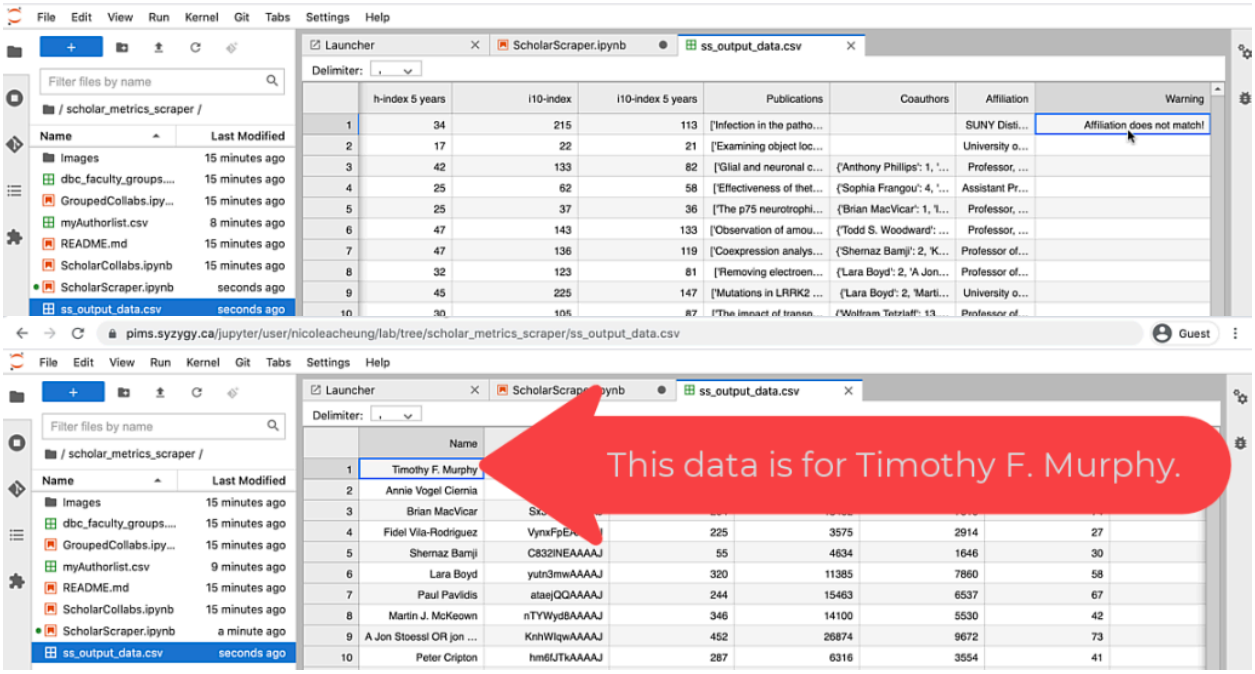
- After running cell number 7, you should find a CSV saved to your directory with the data for each author. You can click it to view it or right-click to download.



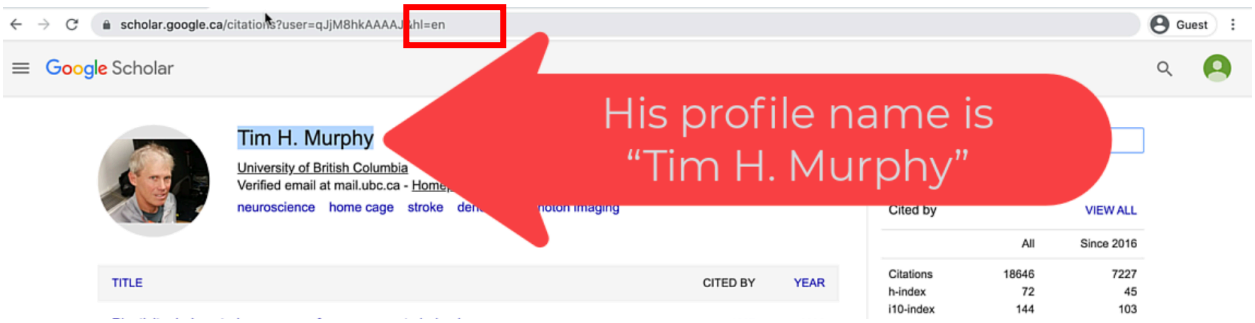
Step 5: Check for warnings

Check the last column of the table for warnings and if necessary, modify your author list CSV.

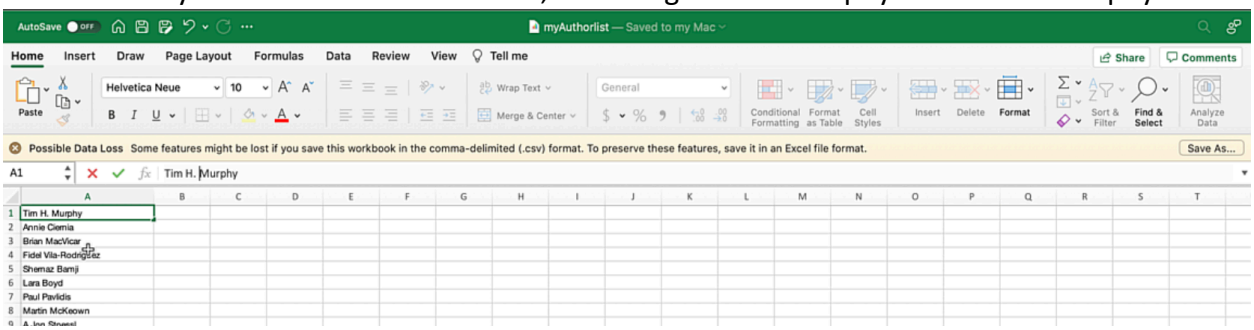
- Check the last column for warnings - the row for Dr. Tim Murphy has been flagged with a warning. The author data retrieved contains an affiliation that does not match one from the affiliations list.



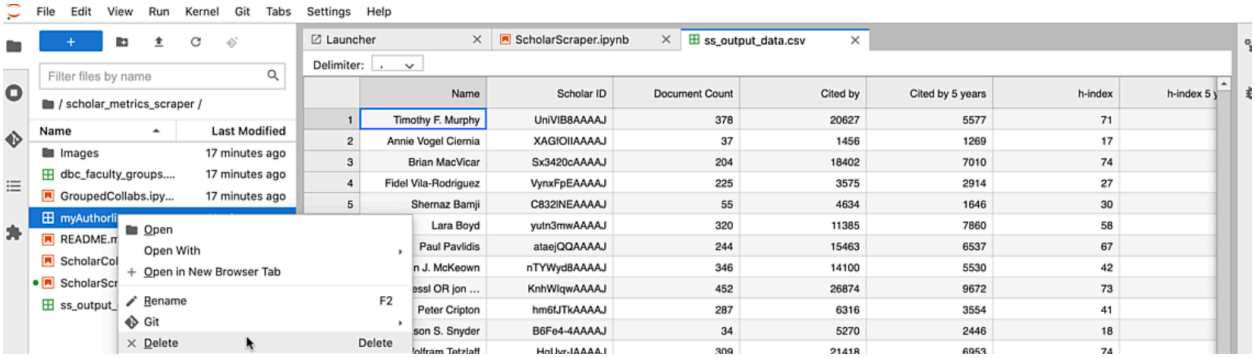
- Find the author's Google Scholar profile and make sure you have their correct information. If you don't, you will need to modify your author list. You can either correct their name or replace their name with their Google Scholar ID which can be found in their profile URL (see the red box below).
    - In our case, we attempted to retrieve data for "Tim Murphy", but the retrieval resulted in data from another author named "Timothy F. Murphy". We either need to change the name in the author list to "Tim H. Murphy" or replace it with the Scholar ID (qJjM8hkAAAAJ).



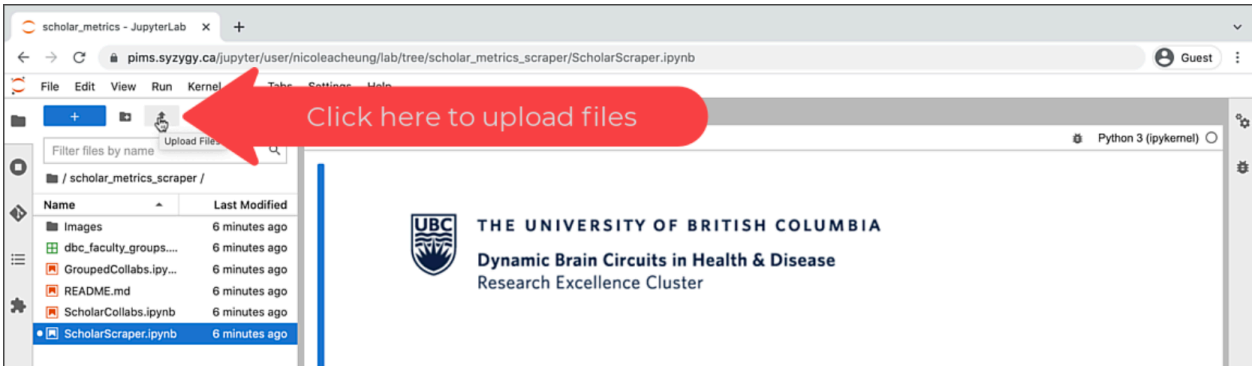
- Modify the author list. In this case, we changed “Tim Murphy” to “Tim H. Murphy”



- Right click to delete the old author list

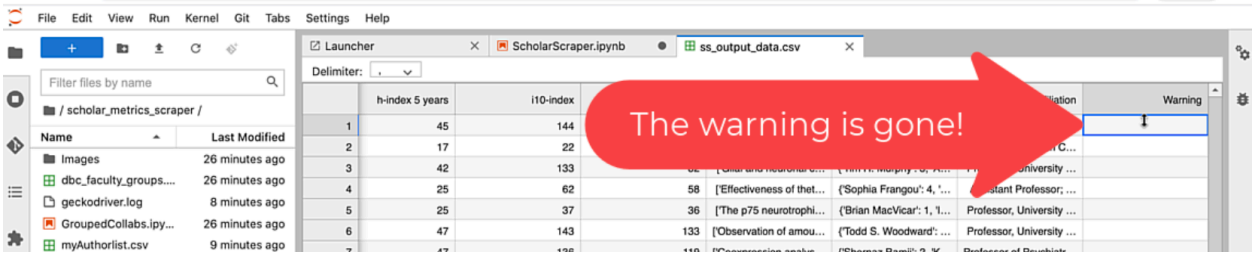


- Upload the corrected author list



- Run code cells 1-7 again

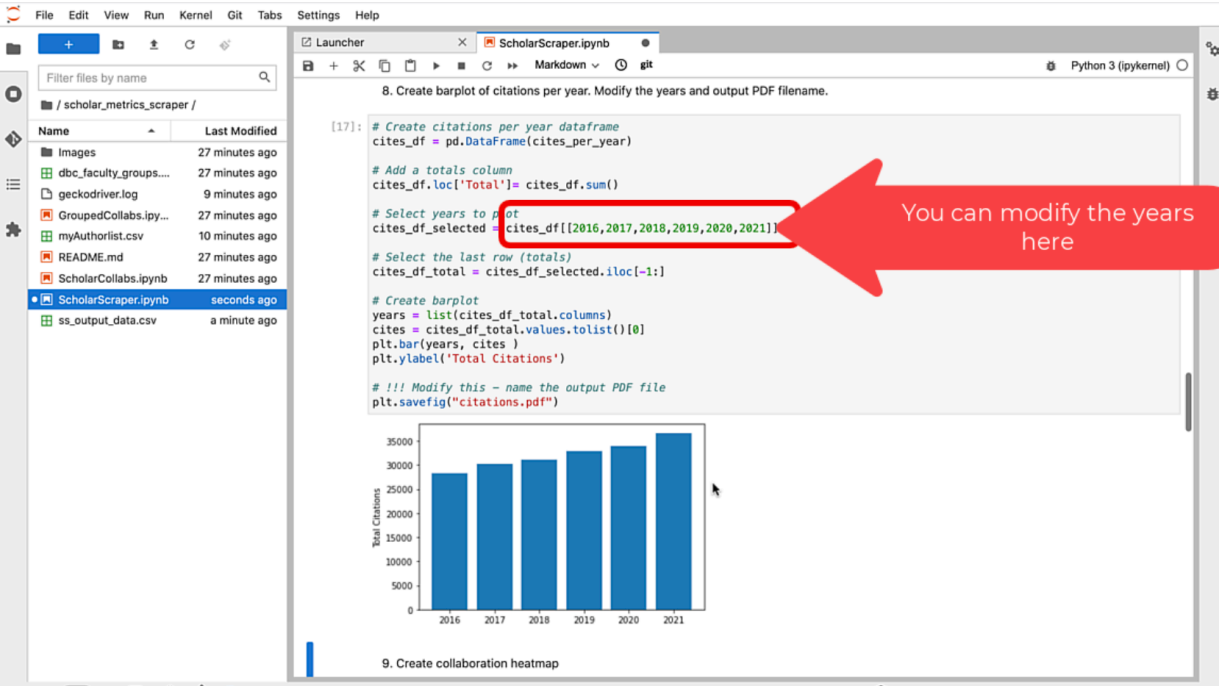
- Check the warnings column again to ensure the author’s data has been corrected.



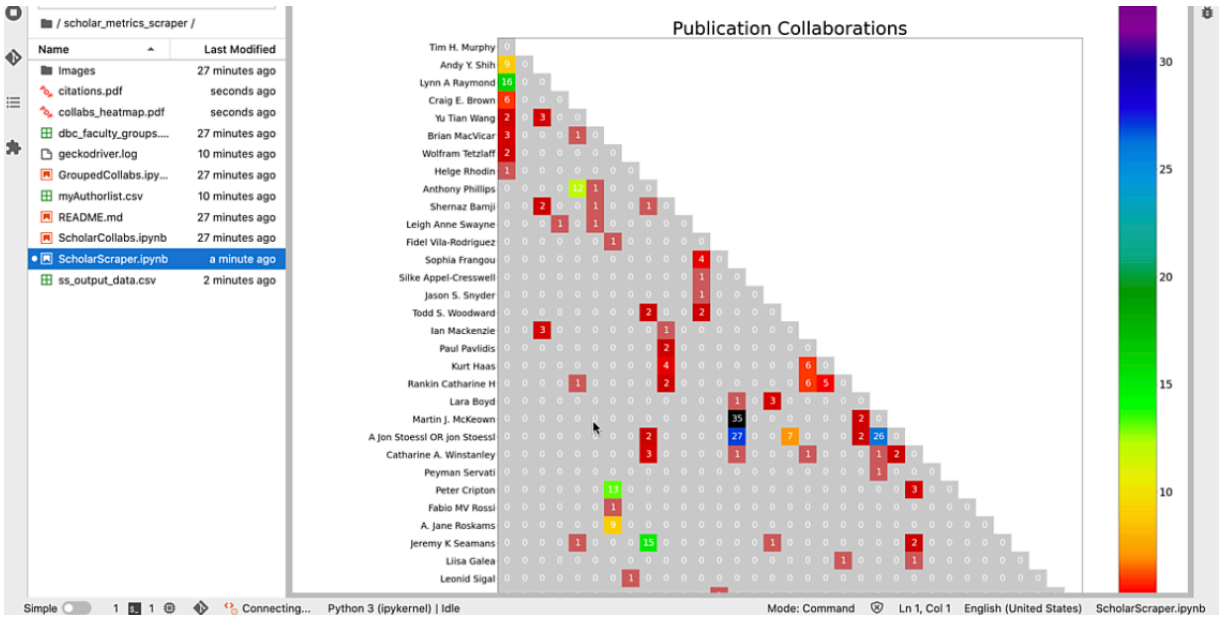
Step 4 (continued)

Run the ScholarScraper notebook code cells 8 and 9.

- Code cell 8 creates a bar chart of total citations/year for the group. The range of years can be modified as shown below.



- Code cell 9 creates a collaboration heatmap, visualizing the number of collaborations between two coauthors

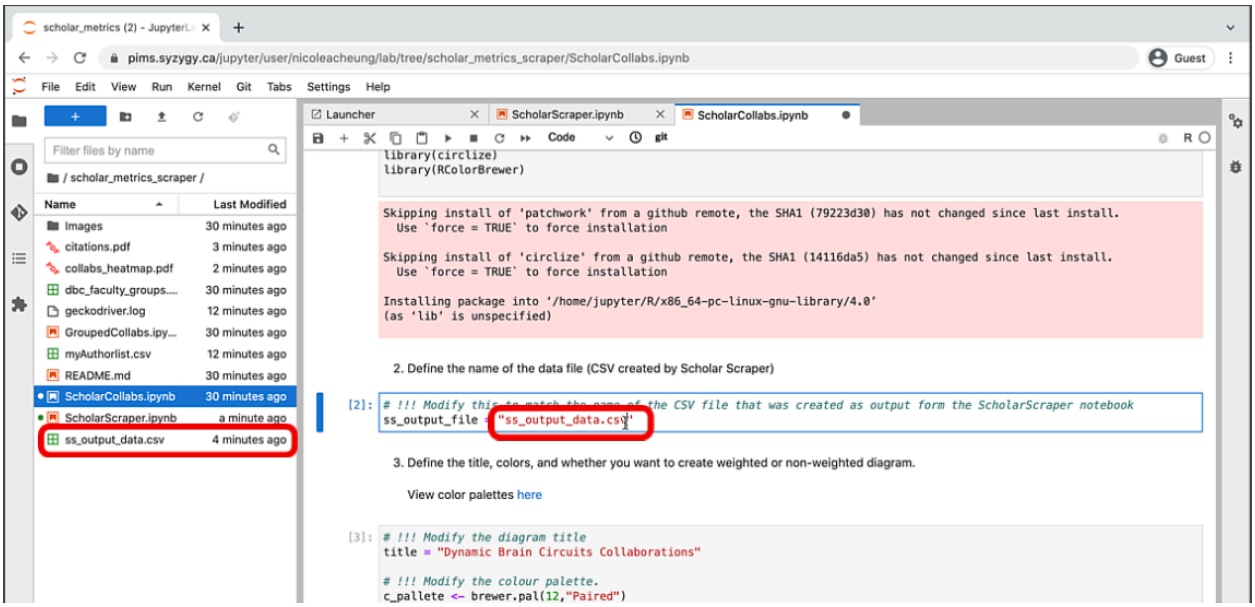


- Both diagrams are saved to the project directory as PDFs for viewing and downloading.

Step 6: Run ScholarCollabs

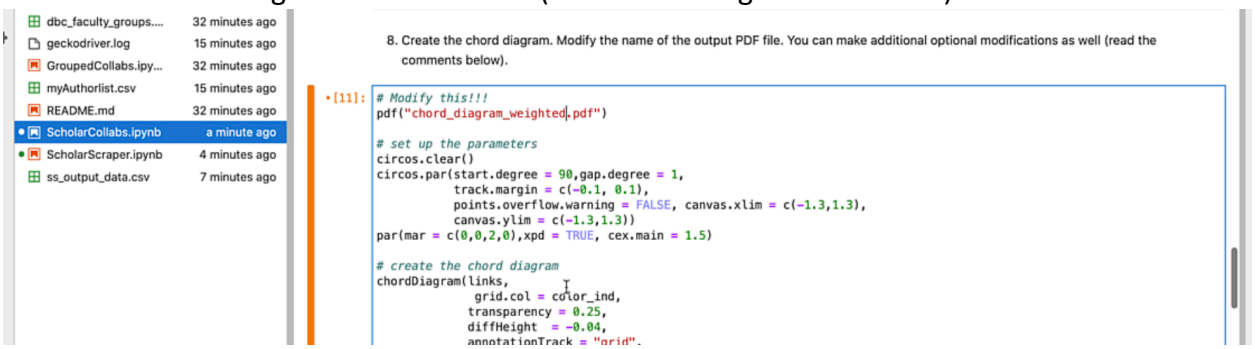
Open the ScholarCollabs notebook. Make some minor modifications to the notebook before running.

- Make sure the output data name matches the CSV filename.



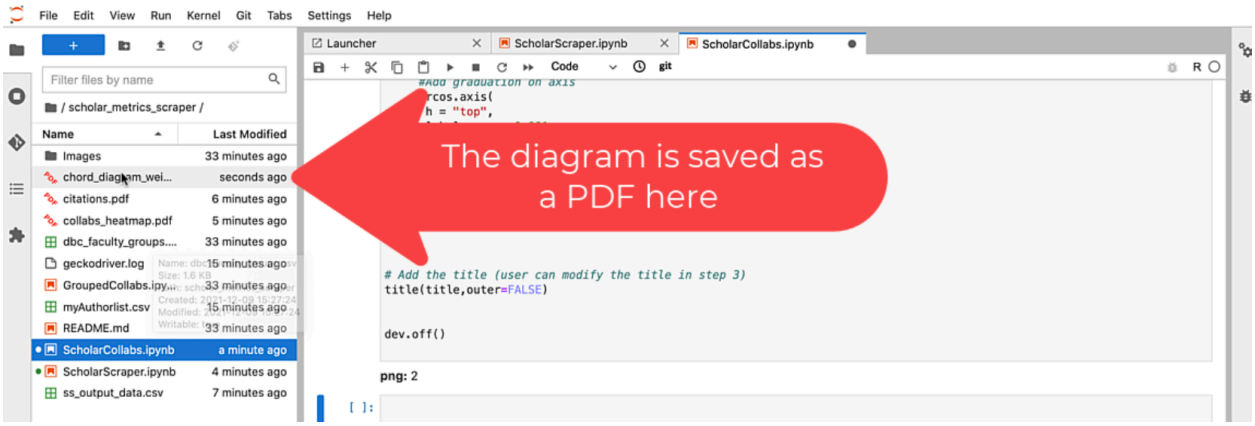
- Modify the title, colour palette, and set weighted to TRUE if you want a weighted diagram (which weights links by the number of collaborations) or FALSE if you want an unweighted diagram.

- You can change the PDF filename (where the image will be saved) here.

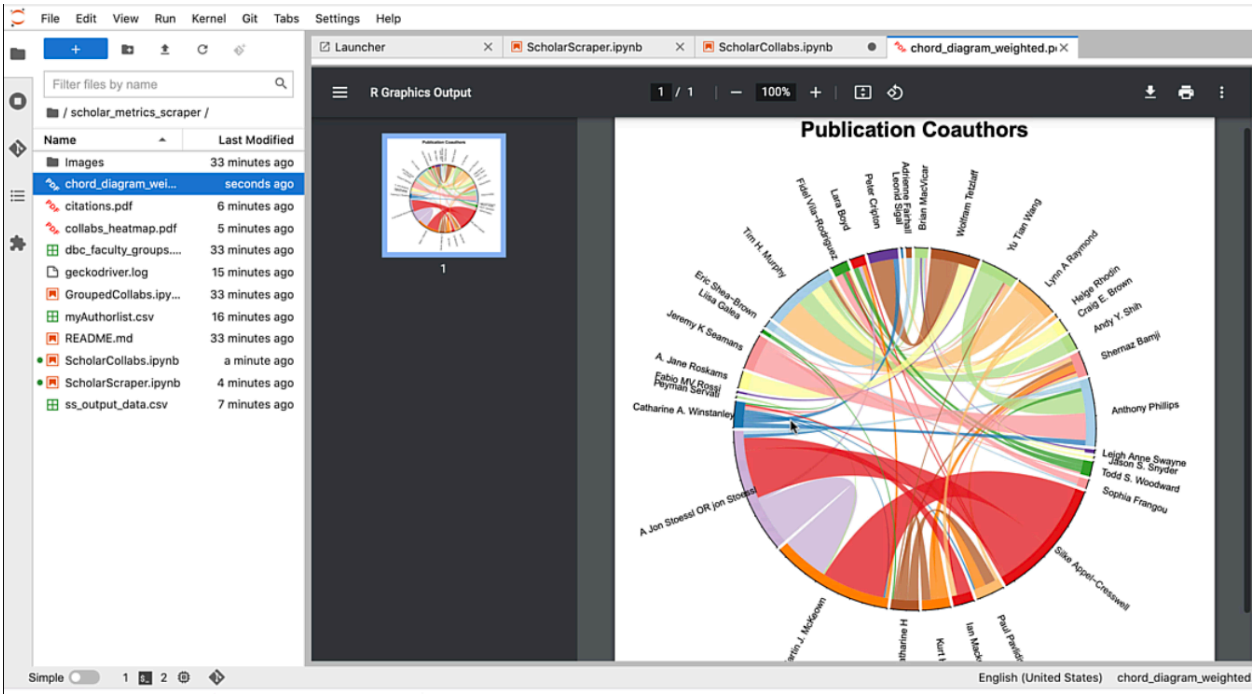


Run the ScholarCollabs notebook code cells.

- After running the last code cell (8), the diagram will be saved in your project directory.



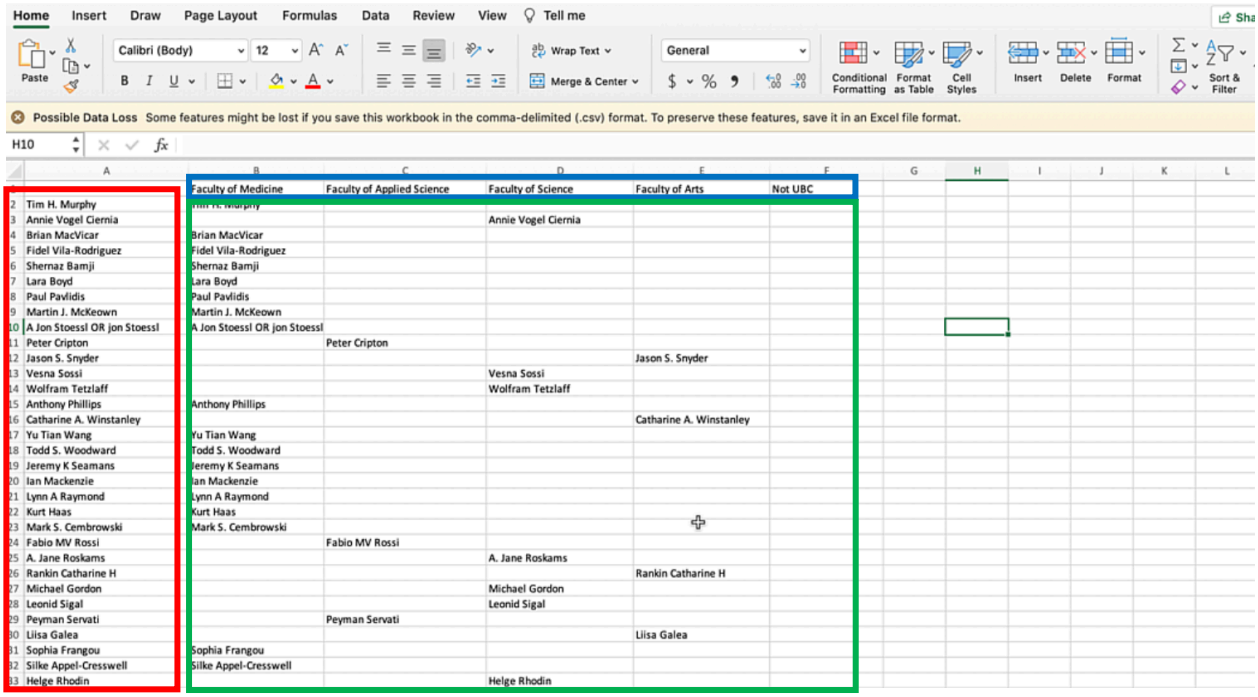
- Click it to view the diagram.



Step 7: Run GroupedCollabs

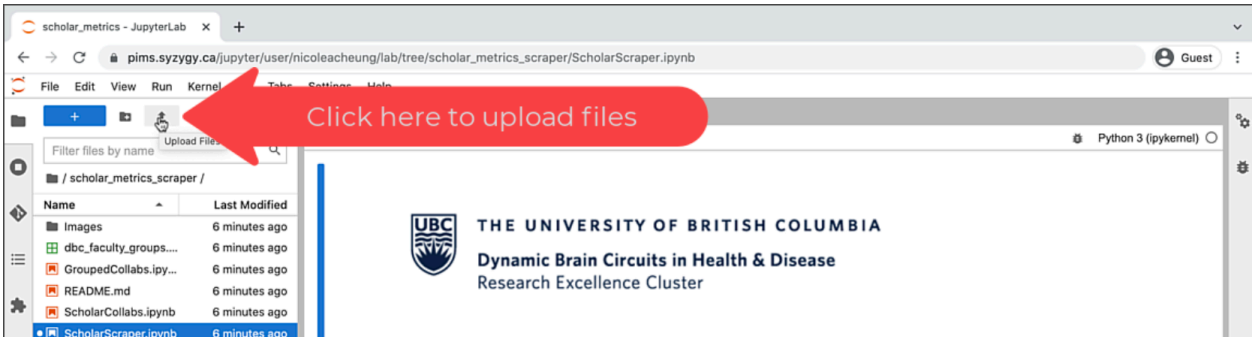
Open the GroupedCollabs notebook. You will need an additional CSV file with each author's group to run this notebook.

- Create a CSV file with author names in the first column (red box), group names in the first row (blue box), and author names in their respective rows and under their group's column (green box). Author names will need to match the names in the output data CSV file (and therefore, match their Google Scholar profile names).



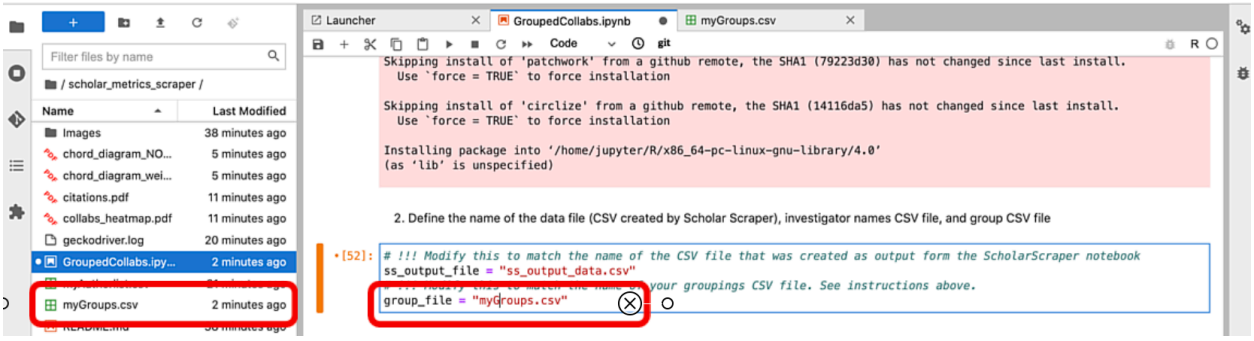
	Faculty of Medicine	Faculty of Applied Science	Faculty of Science	Faculty of Arts	Not UBC
Tim H. Murphy					
Annie Vogel Ciernia			Annie Vogel Ciernia		
Brian MacVicar	Brian MacVicar				
Fidel Vila-Rodriguez	Fidel Vila-Rodriguez				
Shernaz Bamji	Shernaz Bamji				
Lara Boyd	Lara Boyd				
Paul Pavlidis	Paul Pavlidis				
Martin J. McKeown	Martin J. McKeown				
A Jon Stoessl OR jon Stoessl	A Jon Stoessl OR jon Stoessl				
Peter Cripton		Peter Cripton			
Jason S. Snyder				Jason S. Snyder	
Vesna Sossi			Vesna Sossi		
Wolfram Tetzlaff			Wolfram Tetzlaff		
Anthony Phillips	Anthony Phillips				
Catharine A. Winstanley				Catharine A. Winstanley	
Yu Tian Wang	Yu Tian Wang				
Todd S. Woodward	Todd S. Woodward				
Jeremy K Seamans	Jeremy K Seamans				
Ian Mackenzie	Ian Mackenzie				
Lynn A Raymond	Lynn A Raymond				
Kurt Haas	Kurt Haas				
Mark S. Cembrowski	Mark S. Cembrowski				
Fabio MV Rossi		Fabio MV Rossi			
A. Jane Roskams			A. Jane Roskams		
Rankin Catharine H				Rankin Catharine H	
Michael Gordon			Michael Gordon		
Leonid Sigal			Leonid Sigal		
Peyman Servati		Peyman Servati			
Liisa Galea				Liisa Galea	
Sophia Frangou	Sophia Frangou				
Silke Appel-Cresswell	Silke Appel-Cresswell				
Helge Rhodin			Helge Rhodin		

- Save as CSV and upload to the project directory.

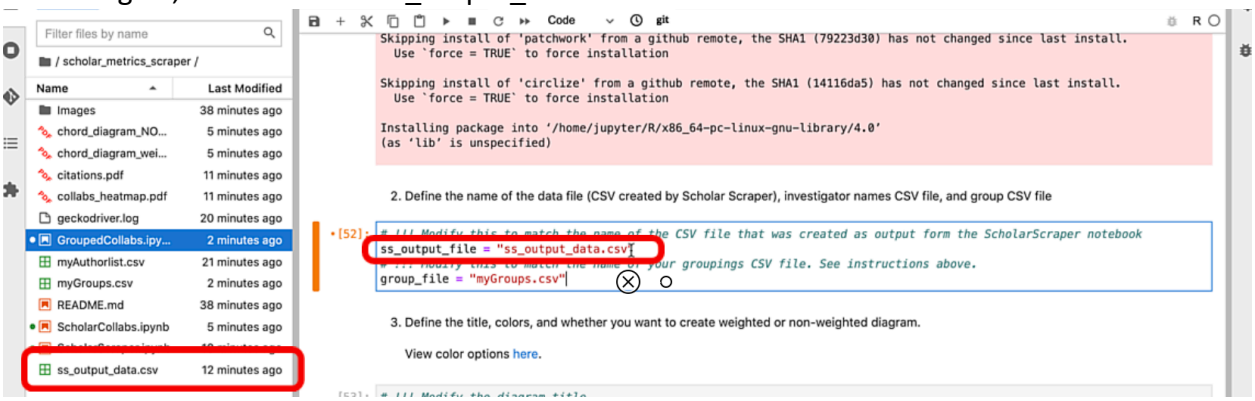


Make some modifications to the GroupedCollabs notebook.

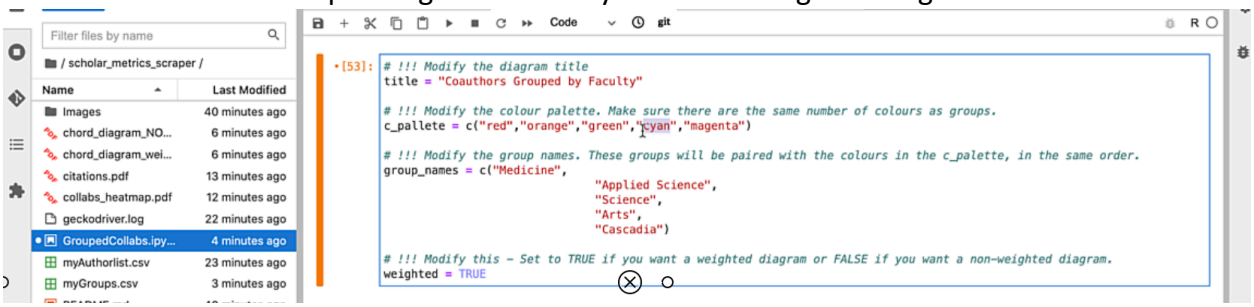
- Modify the group_file name to match the CSV filename.



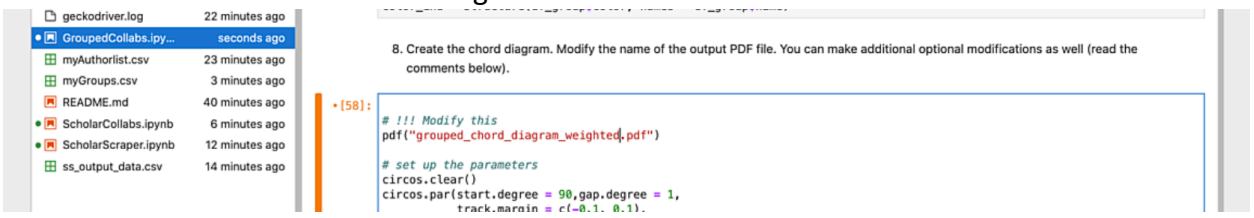
- Again, make sure the ss_output_data name matches the CSV file.



- Modify the title, colours (one for each group), and group names. Again, set weighted to TRUE or FALSE depending on whether you want a weighted diagram.

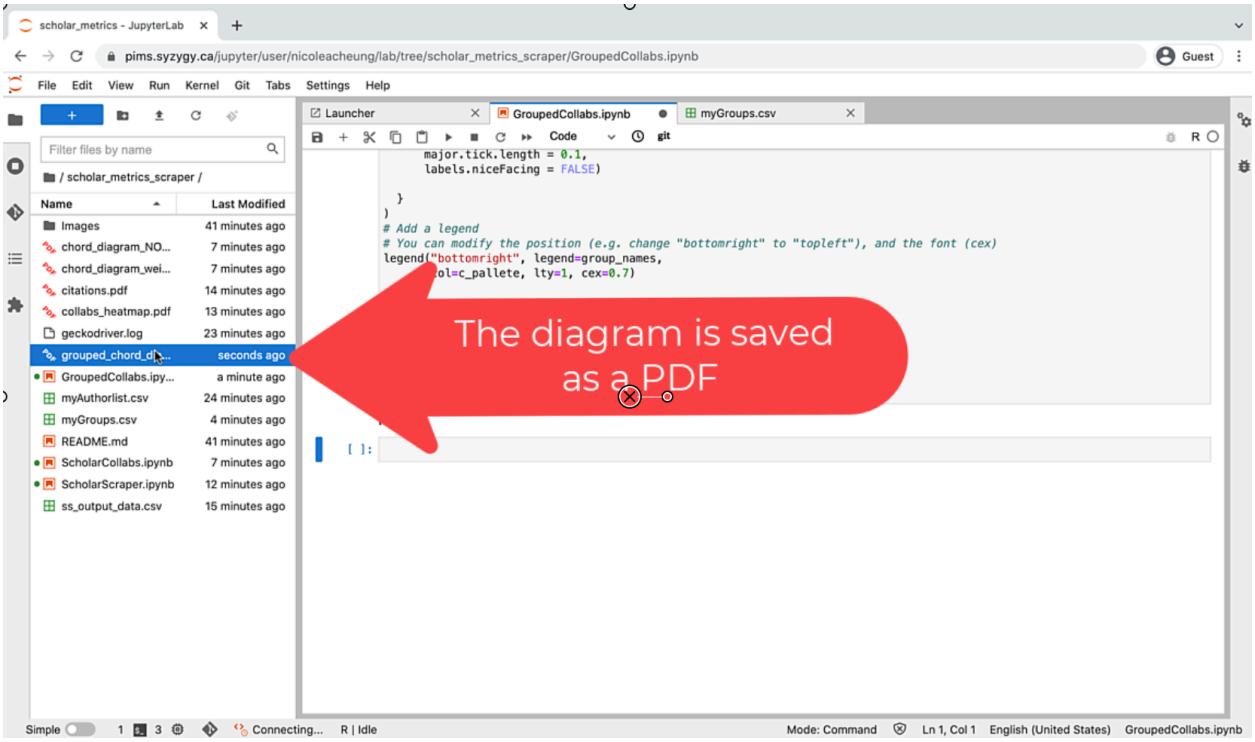


- Name the PDF where the diagram will be saved to.



Run the GroupedCollabs notebook code cells.

- After running the last code cell (8), the diagram will be saved to your project directory.



- Click it to view the diagram.