

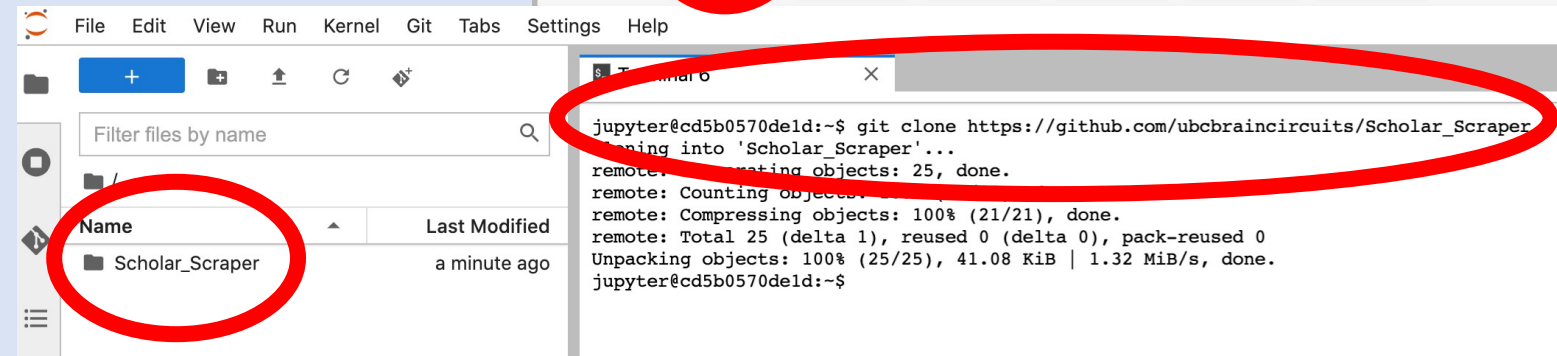
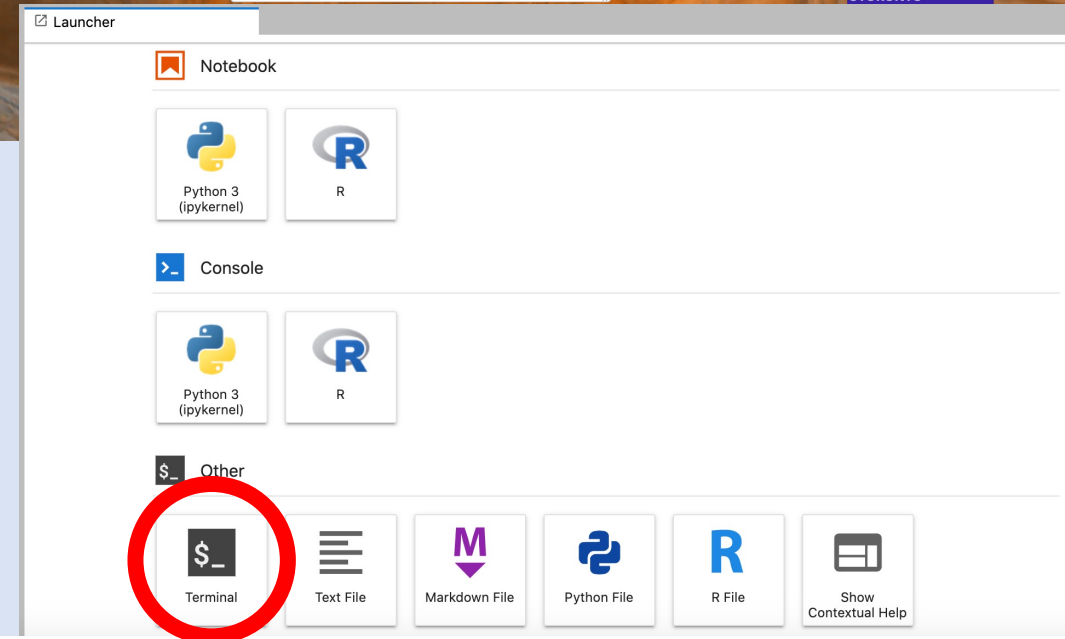
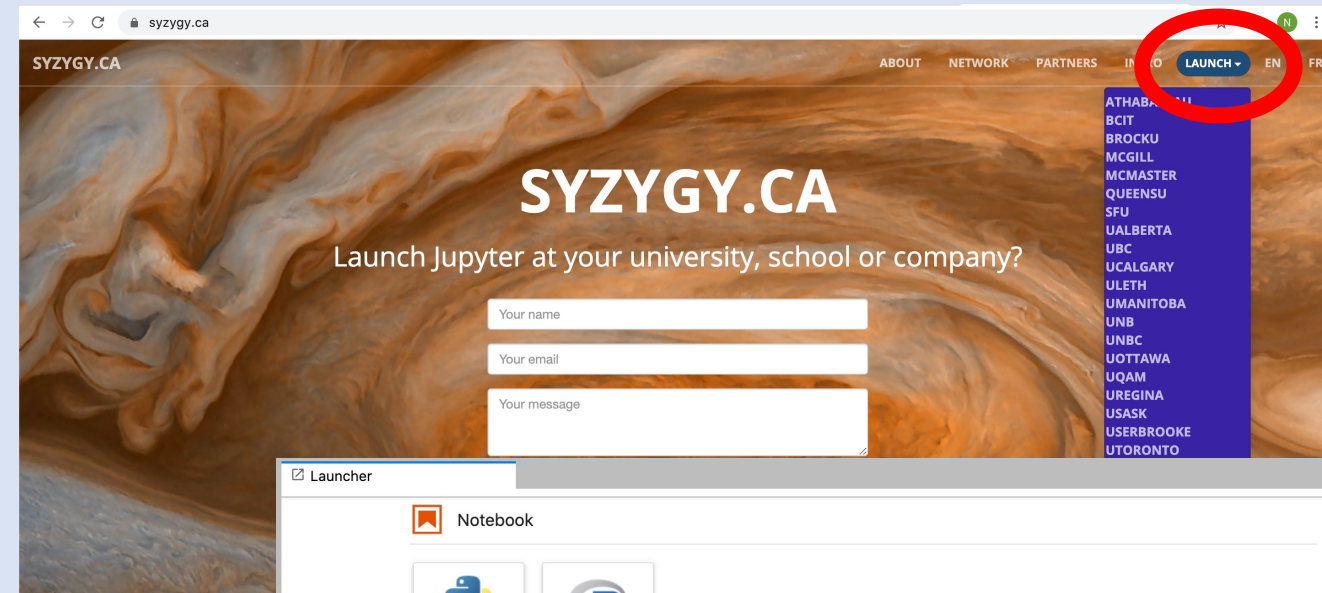
1. Clone Notebook from GitHub

Go to syzygy.ca and click "Launch" and select your institution. Sign in.

Jupyter should open. In the Launcher, select Terminal

Type in: `git clone https://github.com/ubcbraincircuits/Scholar_Scraper`
Press enter

The Scholar_Scraper folder should populate in your directory



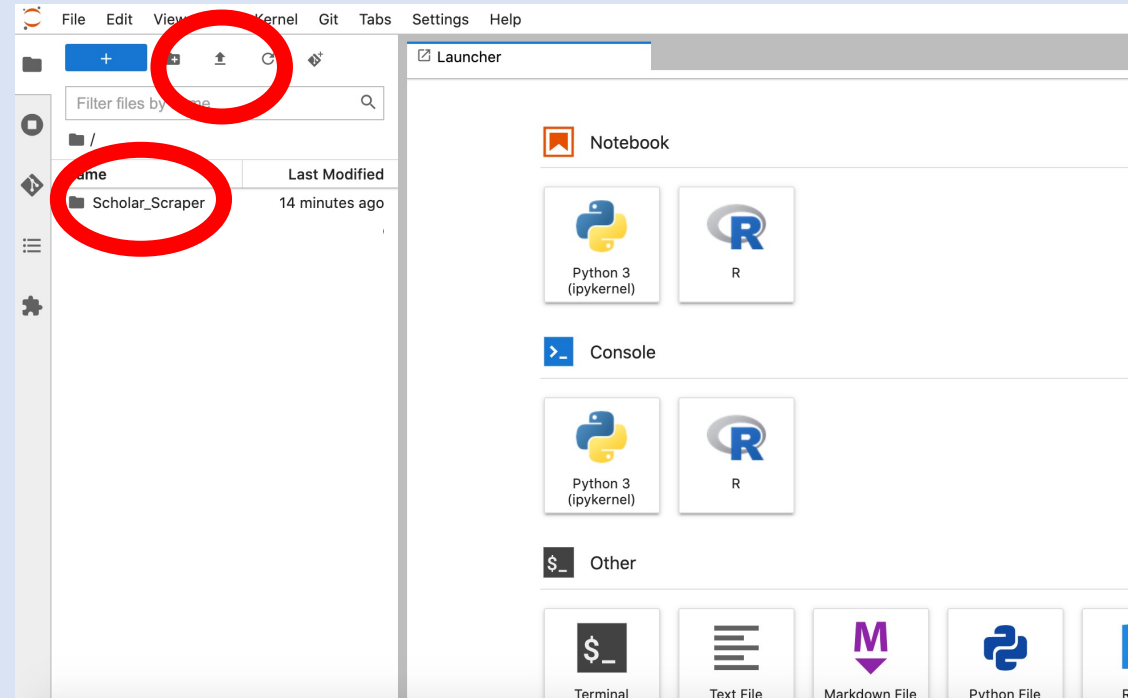
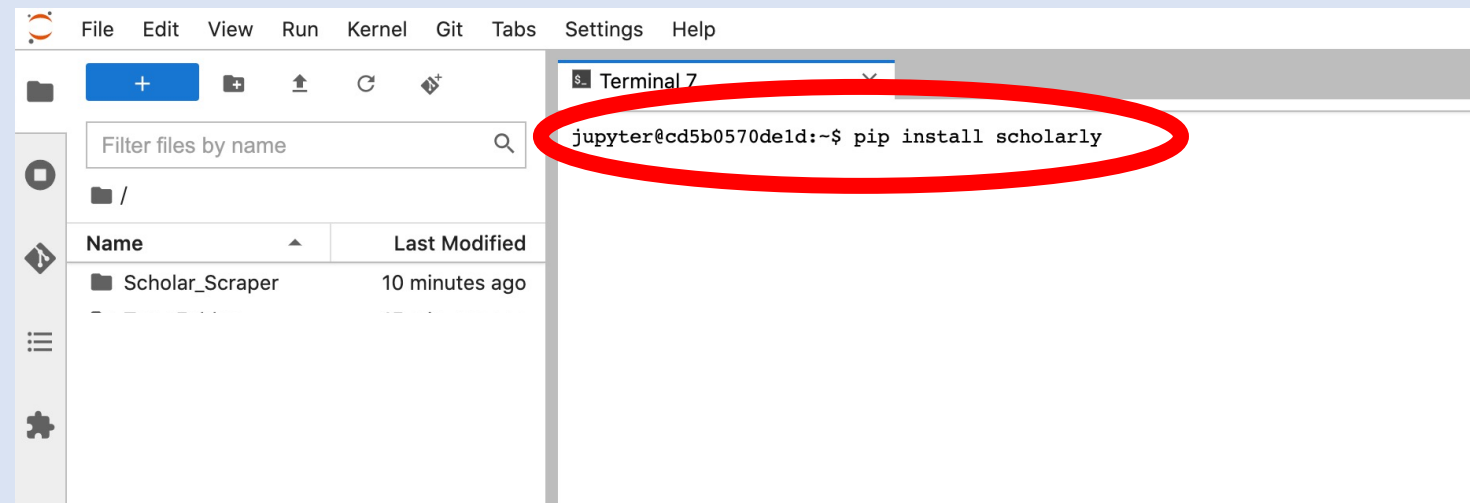
2. Install scholarly

In the terminal again (can be same as before), type:
pip install scholarly
Press enter

3. Import input CSV file with list of author names

Click on the Scholar_Scraper folder.

Then click on the up arrow and upload your csv file
with author names. Author names should be in a
single column, one author per row.



4. Modify the names of the input and output files

Click on scholar_scraper.ipynb

Scroll down to step 1.

Modify the input_authors variable to match the name of the csv file you just imported.

You can also modify the output_data file name. This will be the name of the csv file that is outputted.

Make sure the file names have quotations around them.

The top screenshot shows the JupyterLab file browser for the 'Scholar_Scraper' directory. The file 'scholar_scraper.ipynb' is highlighted with a red circle. The bottom screenshot shows the notebook editor for 'scholar_scraper.ipynb'. The code cell containing the following code is highlighted with a red circle:

```
[ ]: from scholarly import scholarly
import csv
import warnings

1. Modify the names of the input and output files. The input file should be in this directory. The output file does not have to exist yet (it will be created).

[ ]: input_authors = 'DBC Investigators.csv'
output_data = 'output_dbc.csv'

2. Load author list as .csv. Note: author names should match their name on Google Scholar.

[ ]: j = 0
with open(input_authors, encoding="utf-8-sig") as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',')
    for row in csv_reader:
        j = j + 1
```

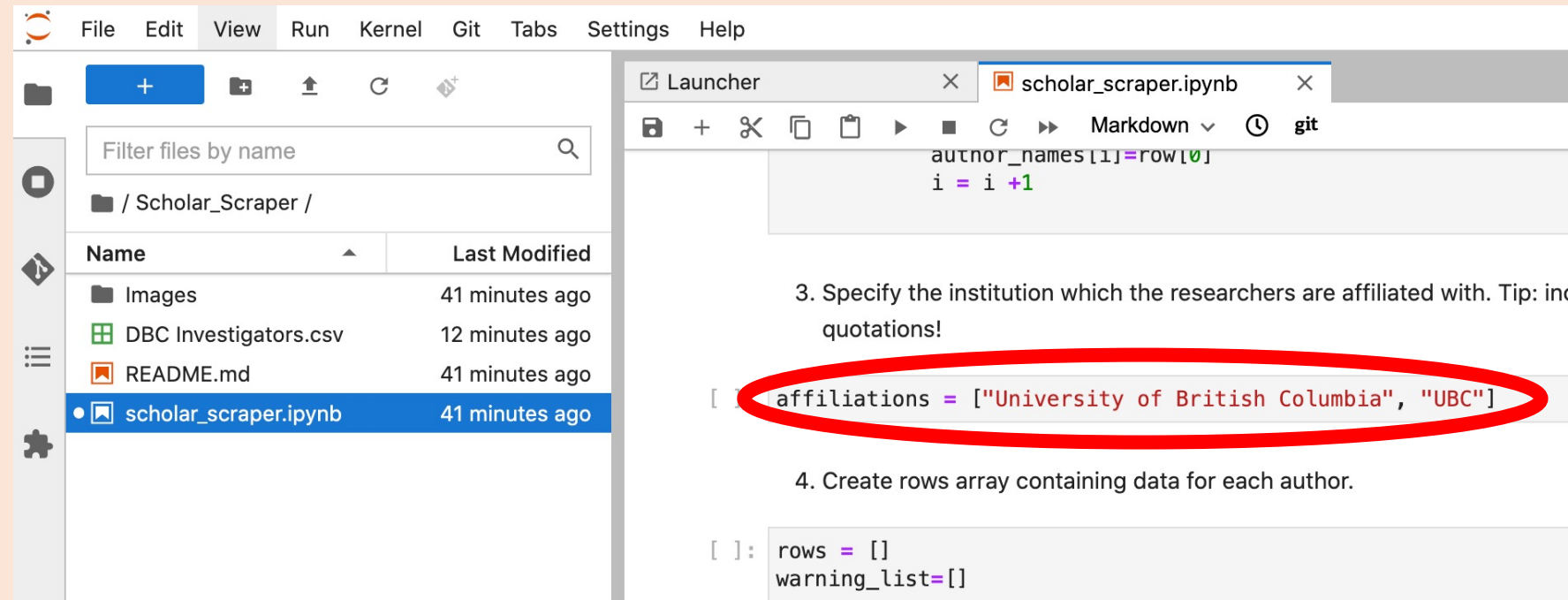
5. Modify the "affiliations" variable

Scroll down to step 3.

Modify the affiliations variable as a list of institution names you expect your author names to be affiliated with.

Add both abbreviated and long forms.

Make sure the list is surrounded by square brackets, each name has quotations, and there is a comma between each name.



The screenshot shows the JupyterLab interface with the Scholar_Scraper directory selected. The file scholar_scraper.ipynb is open, and the code cell is visible. The code cell contains the following code:

```
author_names[i]=row[0]  
i = i +1
```

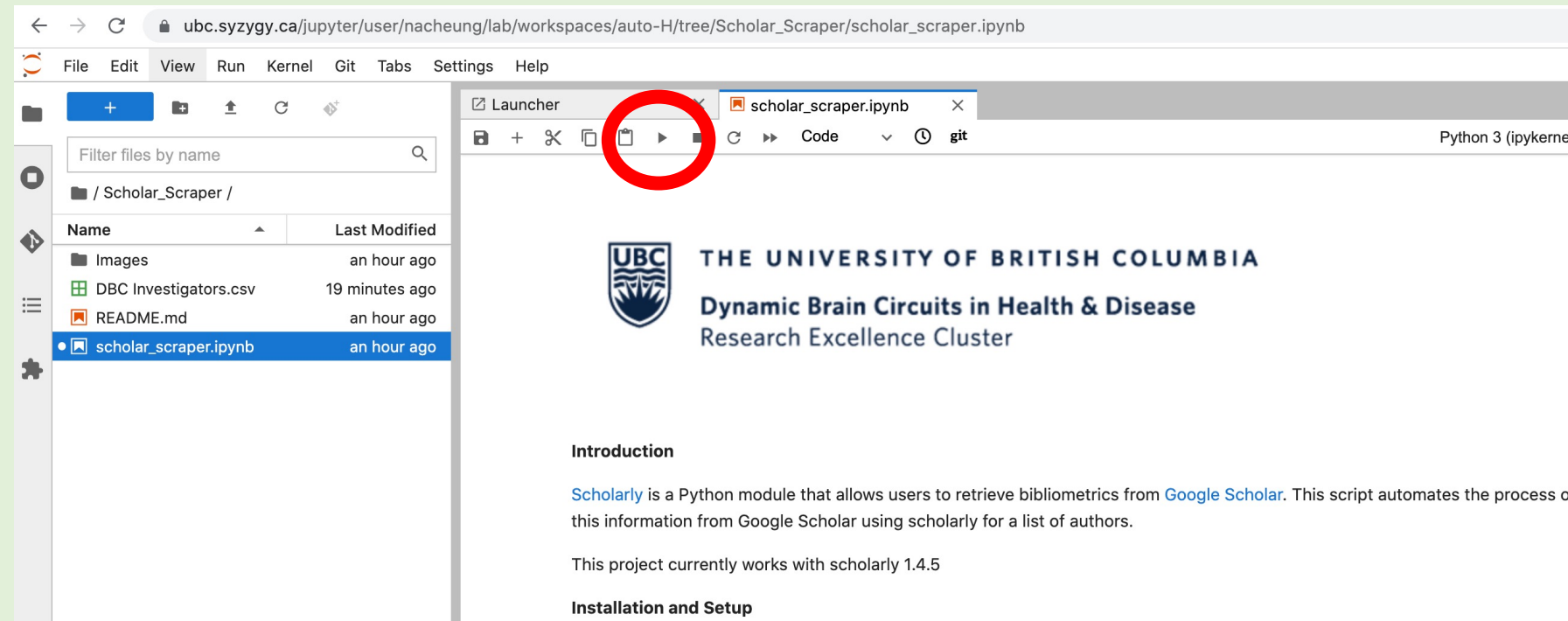
Below the code cell, there is a text prompt: "3. Specify the institution which the researchers are affiliated with. Tip: include both abbreviated and long forms of the institution name in your list. Use square brackets and quotations for each institution name." Below this prompt, the code cell shows the following code:

```
[ ]: affiliations = ["University of British Columbia", "UBC"]
```

The code cell is highlighted with a red circle, indicating the step to modify the affiliations variable.

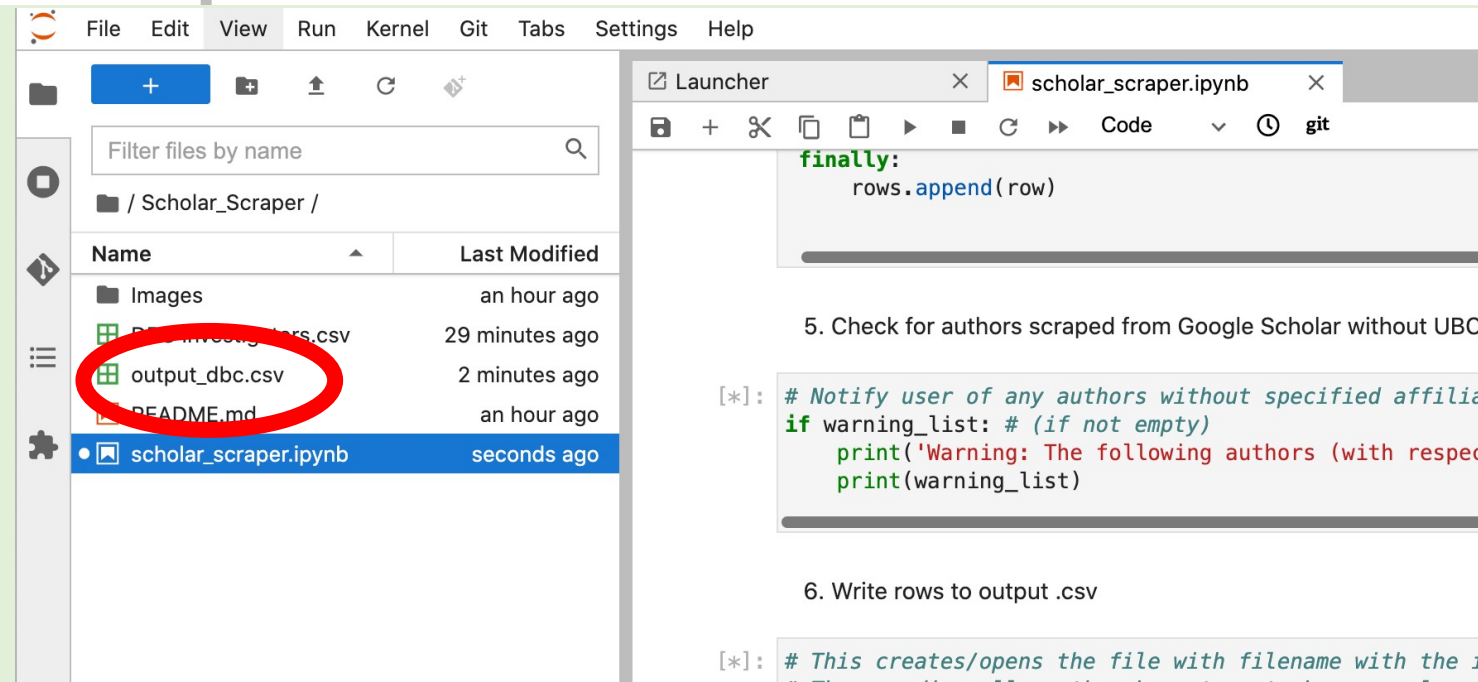
6. Run notebook

You can run each cell by pressing shift enter or clicking the play button.



7. Creates Output CSV

After running step 6, you should find an output csv file in the directory. You can click it to preview it or right click -> download to download it.



8. Check for warnings, make adjustments to input CSV and reimport to Syzygy

You can check for warnings in step 5.

This will list any authors who did not have the specified institutions (from the affiliations variable) as their affiliation on Google Scholar.

You can also check the Warnings column in the output csv file.

This column will also specify if the author was not found on Google Scholar (no information found).

Check these warnings and confirm that you have typed in the correct name in the input CSV file and successfully scraped the correct author's information.

If needed, edit the input csv, reupload it to Syzygy, and run again.

5. Check for authors scraped from Google Scholar without UBC listed as their affiliation. Make sure y

```
[11]: # Notify user of any authors without specified affiliation in case the wrong author
      if warning_list: # (if not empty)
          print('Warning: The following authors (with respective row numbers) do not have
          print(warning_list)
```

Warning: The following authors (with respective row numbers) do not have any of the their affiliation:
[[6, 'Cheryl Rivers'], [13, 'Jason S. Snyder'], [34, 'Sophia Frangou'], [38, 'Brian yne'], [41, 'Adrienne Fairhall'], [42, 'Eric Shea-Brown'], [43, 'Emil Sylwestrak'],

output_dbc

| | | |
|--|---|--|
| | I | J |
| | | |
| _dbc | | |
| tion | | Warning |
| sity of British Columbia | | |
| sity of British Columbia, Vancouver | | |
| | | no information found |
| ssor, University of British Columbia | | |
| ant Professor; University of British Columbia | | |
| r Lecturer in International Business, Victoria University of Wellington | | Specified institutions not found in affiliation! |
| isor, University of British Columbia | | |
| isor, University of British Columbia | | |
| isor of Psychiatry, University of British Columbia | | |
| isor of Medicine (Neurology), University of British Columbia | | |
| sity of British Columbia | | |
| isor of Biomedical Engineering, University of British Columbia | | |
| iate Professor, Department of Psychology, Djavad Mowafaghian Centre for Brain Health | | Specified institutions not found in affiliation! no information found |
| isor and Director ICORD, University of British Columbia | | |