

Oregon Bee Project Data Retrieval and Formatting Script

Summary

These scripts retrieve observation data from iNaturalist.org and reformat it for printing bee specimen labels.

There are two main processes:

1. Pulling and formatting data from iNaturalist.org
2. Merging and indexing formatted data

These processes can be run by executing (double-clicking) iNaturalist_Process.bat or Merge_Process.bat, respectively.

Pulling and Formatting Data from iNaturalist.org

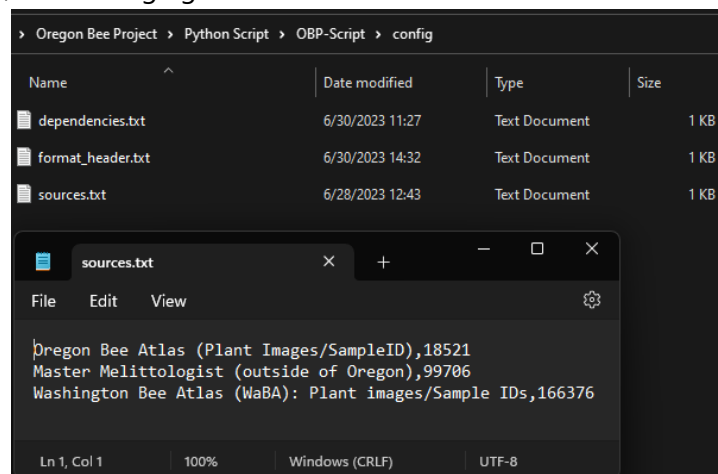
This script (iNaturalist_Process.bat) will query iNaturalist.org for observation data from a given year and from a given list of iNaturalist projects. When a user runs the script, they will be prompted to type a year to query. The list of iNaturalist projects to query is stored in OBP-Script/config/sources.txt (see "Changing Sources" below).

As of July 2023, the script pulls from the following projects:

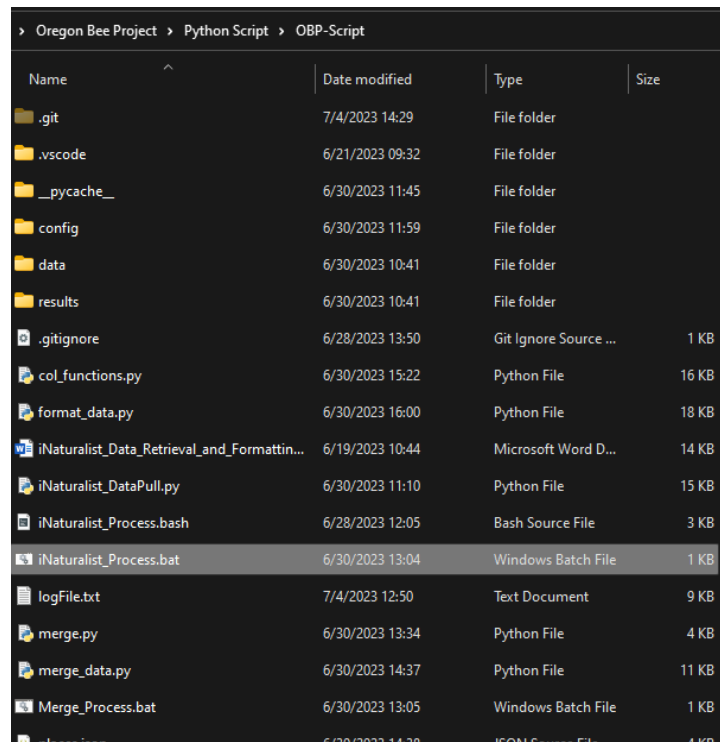
- Oregon Bee Atlas (OBA)
- Master Melittologist (MM)
- Washington Bee Atlas (WaBA)

Running the Process

0. Check that the script is set to pull data from the desired projects (OBP-Script/config/sources.txt).
 - If not, see "Changing Sources" below.

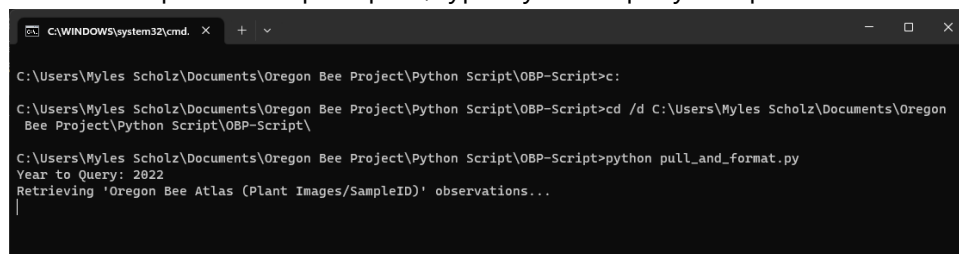


1. Execute (double-click) iNaturalist_Process.bat.



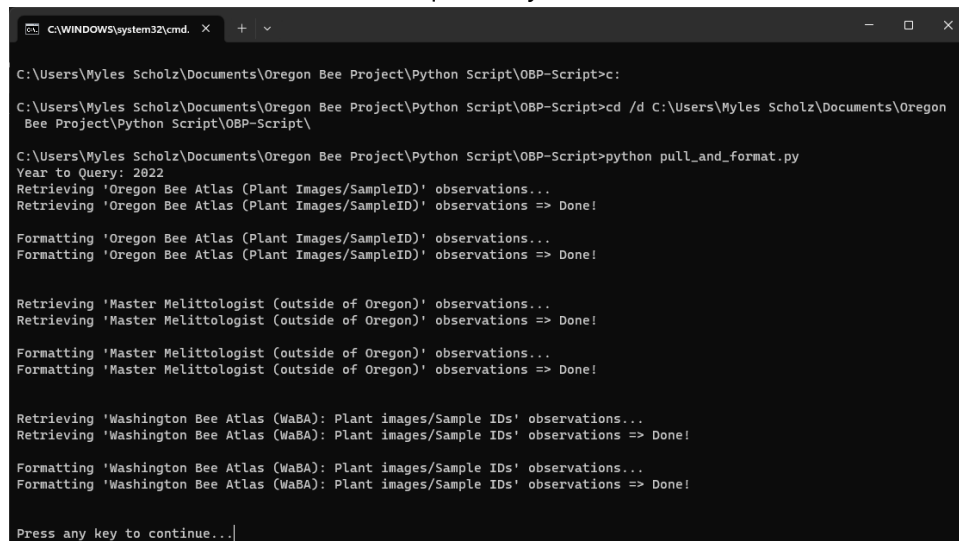
Oregon Bee Project > Python Script > OBP-Script			
Name	Date modified	Type	Size
.git	7/4/2023 14:29	File folder	
.vscode	6/21/2023 09:32	File folder	
__pycache__	6/30/2023 11:45	File folder	
config	6/30/2023 11:59	File folder	
data	6/30/2023 10:41	File folder	
results	6/30/2023 10:41	File folder	
.gitignore	6/28/2023 13:50	Git Ignore Source ...	1 KB
col_functions.py	6/30/2023 15:22	Python File	16 KB
format_data.py	6/30/2023 16:00	Python File	18 KB
iNaturalist_Data_Retrieval_and_Formattin...	6/19/2023 10:44	Microsoft Word D...	14 KB
iNaturalist_DataPull.py	6/30/2023 11:10	Python File	15 KB
iNaturalist_Process.bat	6/28/2023 12:05	Bash Source File	3 KB
iNaturalist_Process.bat	6/30/2023 13:04	Windows Batch File	1 KB
logFile.txt	7/4/2023 12:50	Text Document	9 KB
merge.py	6/30/2023 13:34	Python File	4 KB
merge_data.py	6/30/2023 14:37	Python File	11 KB
Merge_Process.bat	6/30/2023 13:05	Windows Batch File	1 KB
observers	6/30/2023 14:38	JSON Source File	4 KB

2. A terminal will open. When prompted, type a year to query and press Enter.



```
C:\WINDOWS\system32\cmd. x + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python pull_and_format.py
Year to Query: 2022
Retrieving 'Oregon Bee Atlas (Plant Images/SampleID)' observations...
```

3. Wait while the program fetches the data (this may take several minutes). The program will pull and format each source's data sequentially.



```
C:\WINDOWS\system32\cmd. x + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python pull_and_format.py
Year to Query: 2022
Retrieving 'Oregon Bee Atlas (Plant Images/SampleID)' observations...
Retrieving 'Oregon Bee Atlas (Plant Images/SampleID)' observations => Done!
Formatting 'Oregon Bee Atlas (Plant Images/SampleID)' observations...
Formatting 'Oregon Bee Atlas (Plant Images/SampleID)' observations => Done!

Retrieving 'Master Melittologist (outside of Oregon)' observations...
Retrieving 'Master Melittologist (outside of Oregon)' observations => Done!
Formatting 'Master Melittologist (outside of Oregon)' observations...
Formatting 'Master Melittologist (outside of Oregon)' observations => Done!

Retrieving 'Washington Bee Atlas (WaBA): Plant images/Sample IDs' observations...
Retrieving 'Washington Bee Atlas (WaBA): Plant images/Sample IDs' observations => Done!
Formatting 'Washington Bee Atlas (WaBA): Plant images/Sample IDs' observations...
Formatting 'Washington Bee Atlas (WaBA): Plant images/Sample IDs' observations => Done!

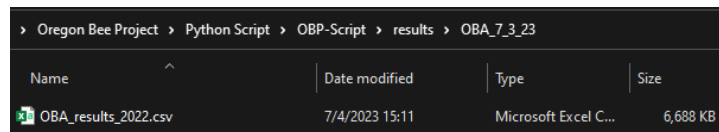
Press any key to continue...|
```

4. The resulting formatted CSV files will appear in source-specific folders under OBP-Script/results/.

Each source-specific folder will be named according to the format Abbr_M_D_YY, where Abbr is an abbreviation of the source and M_D_YY is the date when the program ran. For example, the results of fetching data from the Oregon Bee Atlas on July 3rd, 2023 would appear in OBP-Script/results/OBA_7_3_23/.

The resulting CSV file will be named according to the format Abbr_results_YYYY.csv, where Abbr is an abbreviation of the source and YYYY is the year that was queried. For example, querying Oregon Bee Atlas data from 2022 would produce a file named OBA_results_2022.csv.

If the program is run a second time on the same day, the results file will be entirely overwritten with new data.



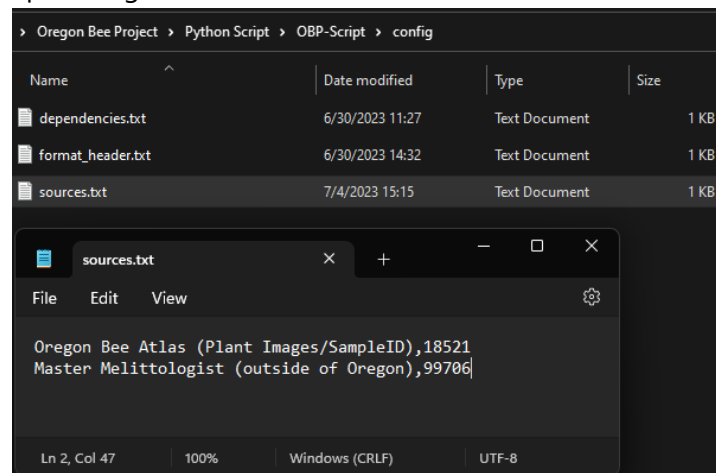
Name	Date modified	Type	Size
OBA_results_2022.csv	7/4/2023 15:11	Microsoft Excel C...	6,688 KB

Changing Sources

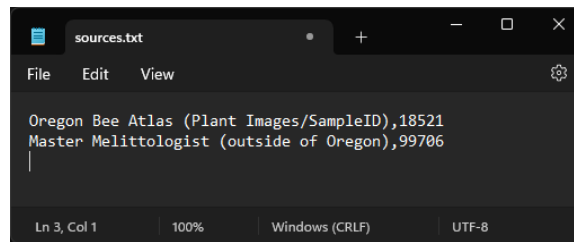
The script pulls data from a list of iNaturalist projects in OBP-Script/config/sources.txt. This file has a specific format that must be maintained for the script to run properly. Each source has its own line, and each line has two comma-separated fields: the name of the source and the iNaturalist project ID. The source name is not essential for the script to function; it is only used for improving the readability of the program's output. The iNaturalist project ID is essential for pulling data from the correct sources.

To add a source, do the following:

1. Open OBP-Script/config/sources.txt.

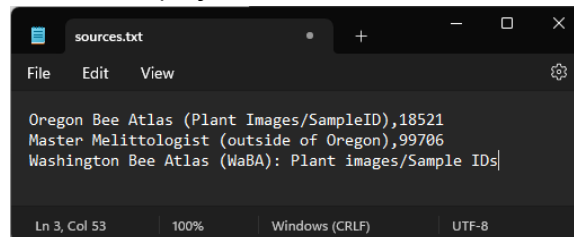


2. Add a line at the end of the file.



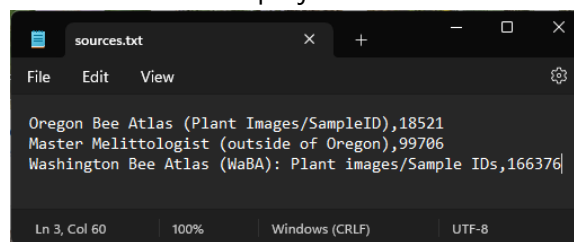
```
sources.txt
File Edit View
Oregon Bee Atlas (Plant Images/SampleID),18521
Master Melittologist (outside of Oregon),99706
Ln 3, Col 1 100% Windows (CRLF) UTF-8
```

3. Type a name for the source.
 - The name cannot contain commas, but all other keyboard characters are allowed.
 - New sources will not have an abbreviation unless the code is modified. Output files and folders will use the project ID instead.



```
sources.txt
File Edit View
Oregon Bee Atlas (Plant Images/SampleID),18521
Master Melittologist (outside of Oregon),99706
Washington Bee Atlas (WaBA): Plant images/Sample IDs
Ln 3, Col 53 100% Windows (CRLF) UTF-8
```

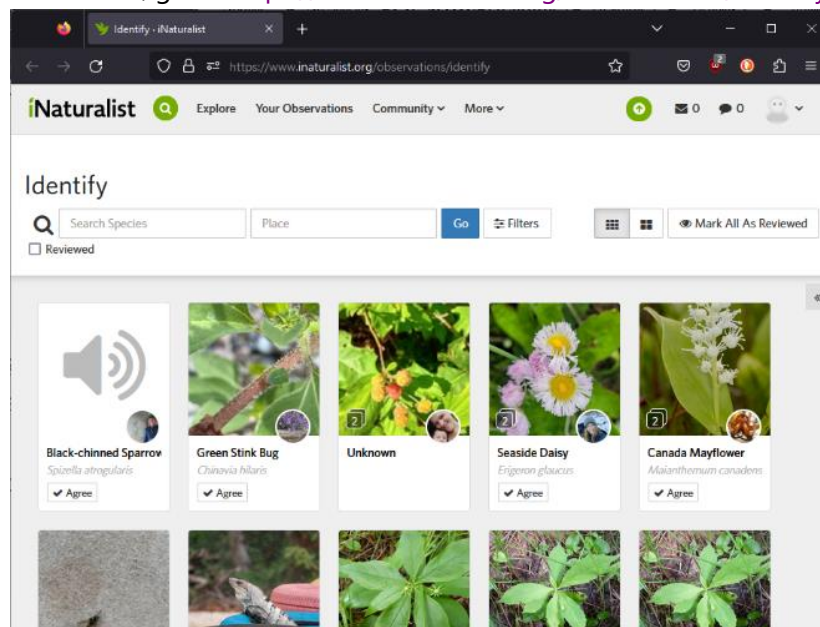
4. Type a comma and then the iNaturalist project ID.



```
sources.txt
File Edit View
Oregon Bee Atlas (Plant Images/SampleID),18521
Master Melittologist (outside of Oregon),99706
Washington Bee Atlas (WaBA): Plant images/Sample IDs,166376
Ln 3, Col 60 100% Windows (CRLF) UTF-8
```

The iNaturalist project ID can be found by doing the following:

- In a browser, go to <https://www.inaturalist.org/observations/identify>.



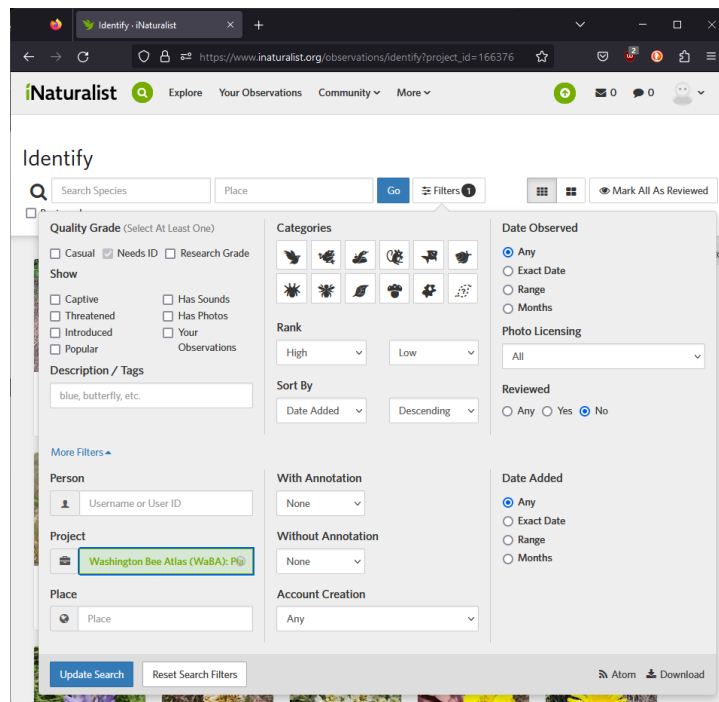
- Click the "Filters" button.

The screenshot shows the iNaturalist 'Identify' page. At the top, there's a navigation bar with 'iNaturalist' logo and links for 'Explore', 'Your Observations', 'Community', and 'More'. Below this, the 'Identify' section has a search bar for 'Search Species' and a 'Place' dropdown. To the right of these are 'Go' and 'Filters' buttons. The 'Filters' button is highlighted with a blue box. Below the search bar, there are several filter categories: 'Quality Grade' (with checkboxes for 'Casual', 'Needs ID', and 'Research Grade'), 'Show' (with checkboxes for 'Captive', 'Threatened', 'Introduced', 'Popular', 'Has Sounds', 'Has Photos', and 'Your Observations'), 'Description / Tags' (with a text input field containing 'blue, butterfly, etc.'), 'Categories' (with icons for various organisms), 'Rank' (with 'High' and 'Low' dropdowns), 'Sort By' (with 'Date Added' and 'Descending' dropdowns), 'Date Observed' (with radio buttons for 'Any', 'Exact Date', 'Range', and 'Months'), 'Photo Licensing' (with a dropdown set to 'All'), and 'Reviewed' (with radio buttons for 'Any', 'Yes', and 'No'). At the bottom of the filter section, there are 'Update Search' and 'Reset Search Filters' buttons, and a 'More Filters' link.

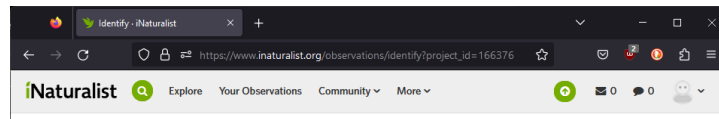
- Click the "More Filters" button.

This screenshot shows the iNaturalist 'Identify' page with the 'More Filters' button highlighted. The 'More Filters' button is located at the bottom left of the filter section. The filter section is expanded to show additional filters: 'Person' (with a dropdown for 'Username or User ID'), 'Project' (with a text input field for 'Name or URL slug, e.g. my-project'), 'Place' (with a dropdown for 'Place'), 'With Annotation' (with a dropdown for 'None'), 'Without Annotation' (with a dropdown for 'None'), 'Account Creation' (with a dropdown for 'Any'), and 'Date Added' (with radio buttons for 'Any', 'Exact Date', 'Range', and 'Months'). The 'Update Search' and 'Reset Search Filters' buttons are still visible at the bottom of the filter section.

- Type the name of the project in the "Project" field.
 - Be sure to select the project from the results. The full project name should appear in green.



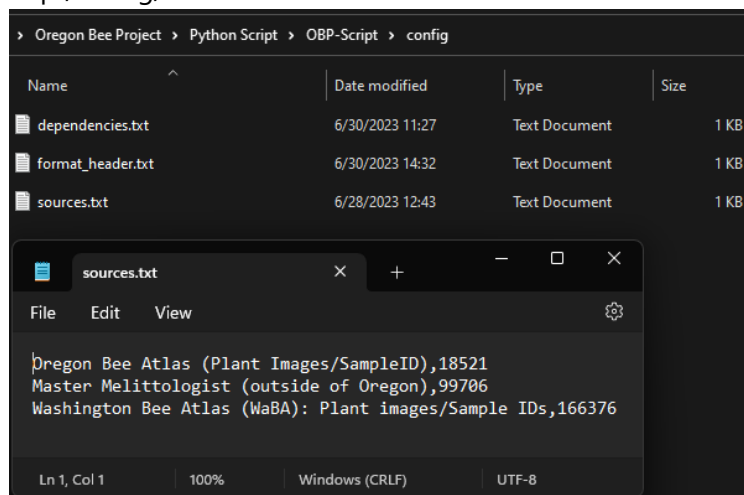
- The project ID will appear as a string of about five numbers at the end of the browser's URL bar.



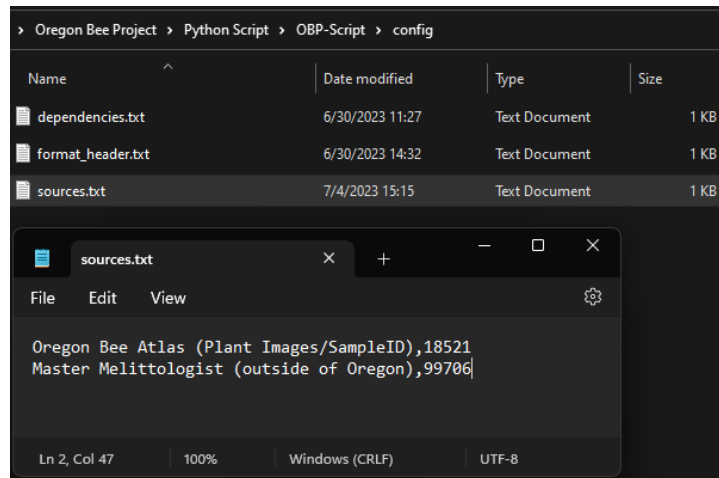
5. Save and close sources.txt.

To remove a source, do the following:

1. Open OBP-Script/config/sources.txt.



2. Select and delete the line containing the source to be removed.
 - Do not leave the line empty. Press backspace on the empty line to remove it.



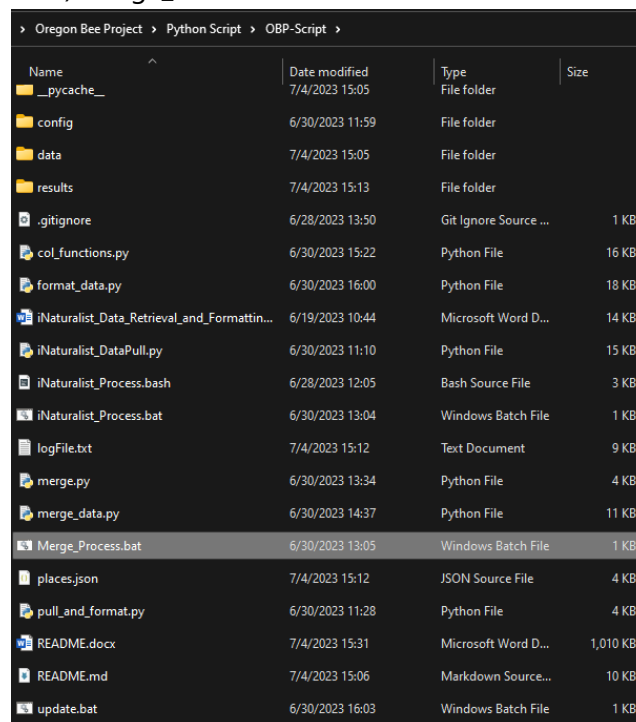
3. Save and close sources.txt.

Merging and Indexing Formatted Data

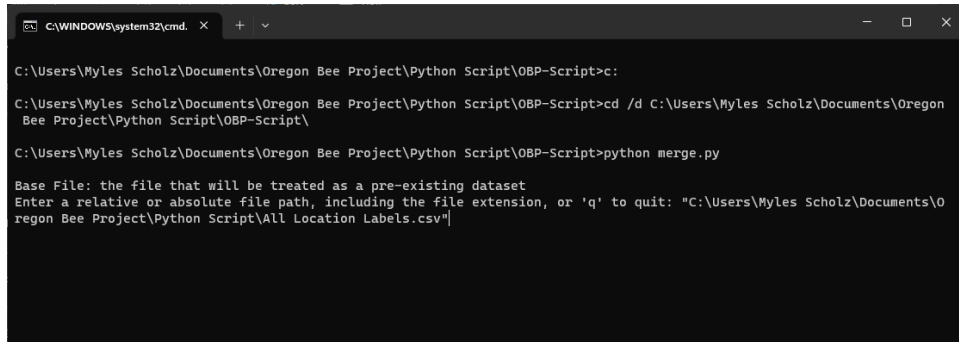
This script (Merge_Process.bat) will combine, sort, and index two formatted data files, such as the "results" files that are output from the data pulling process. When the user runs the script, they will be prompted for three file paths: the base file, the file to append to the base file, and the output file. These files must be CSV files and must have the exact header specified in OBP-Script/config/format_header.txt (see "Help and Tips" below).

Running the Process

1. Execute (double-click) Merge_Process.bat.

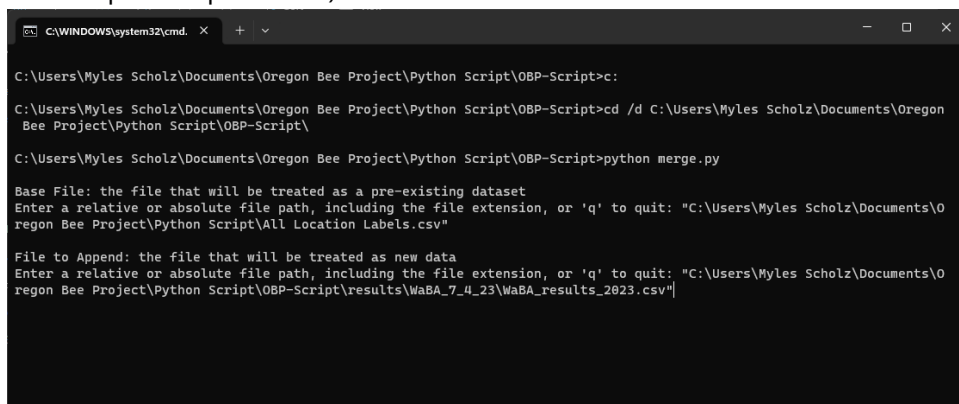


2. When prompted, type or paste the file path of the base file.
 - The program automatically trims quotation marks from the ends of the file path, so users may paste file paths copied directly from the File Explorer on Windows (see "Help and Tips" below).



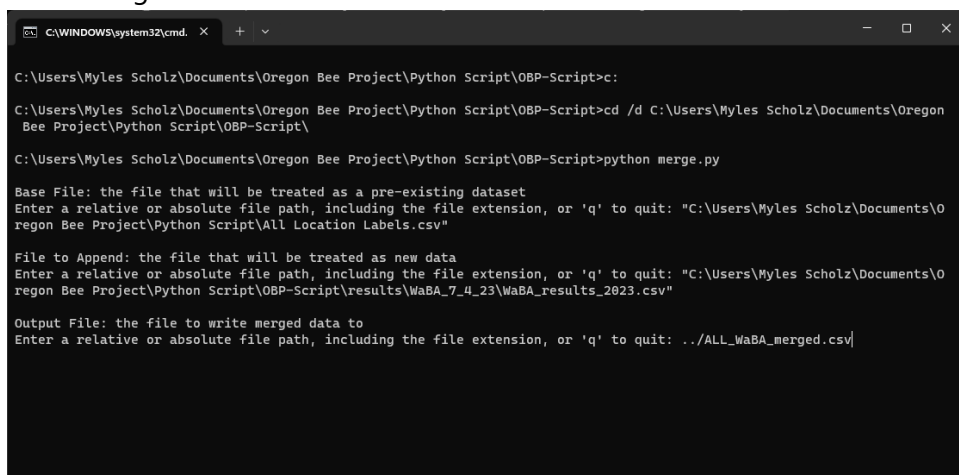
```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py
Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"
```

3. When prompted, type or paste the file path of the file to append to the base file.
 - The program automatically trims quotation marks from the ends of the file path, so users may paste file paths copied directly from the File Explorer on Windows (see "Help and Tips" below).



```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py
Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"
File to Append: the file that will be treated as new data
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\OBP-Script\results\WaBA_7_4_23\WaBA_results_2023.csv"
```

4. When prompted, type or paste a file path to output the merged data to.
This may be a new file or an existing file but be careful because the program will overwrite all data in the target file.



```
C:\WINDOWS\system32\cmd. X + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py
Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"
File to Append: the file that will be treated as new data
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\OBP-Script\results\WaBA_7_4_23\WaBA_results_2023.csv"
Output File: the file to write merged data to
Enter a relative or absolute file path, including the file extension, or 'q' to quit: ..\ALL_WaBA_merged.csv
```


- Wait while the program merges the data (this may take a minute, depending on the size of the input files).

```
C:\WINDOWS\system32\cmd. x + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py

Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"

File to Append: the file that will be treated as new data
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\OBP-Script\results\WaBA_7_4_23\WaBA_results_2023.csv"

Output File: the file to write merged data to
Enter a relative or absolute file path, including the file extension, or 'q' to quit: ../ALL_WaBA_merged.csv
Merging '..\All Location Labels.csv' and 'results\WaBA_7_4_23\WaBA_results_2023.csv' into '..\ALL_WaBA_merged.csv'...
```

- The resulting merged data will appear in the specified output file path.

```
C:\WINDOWS\system32\cmd. x + v
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py

Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"

File to Append: the file that will be treated as new data
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\OBP-Script\results\WaBA_7_4_23\WaBA_results_2023.csv"

Output File: the file to write merged data to
Enter a relative or absolute file path, including the file extension, or 'q' to quit: ../ALL_WaBA_merged.csv
Merging '..\All Location Labels.csv' and 'results\WaBA_7_4_23\WaBA_results_2023.csv' into '..\ALL_WaBA_merged.csv'...
Merging '..\All Location Labels.csv' and 'results\WaBA_7_4_23\WaBA_results_2023.csv' into '..\ALL_WaBA_merged.csv' => Do
ne!

Press any key to continue...|
```

The merged data will not have duplicate entries, given that the combination of the following fields uniquely identify an entry:

- o "iNaturalist Alias"
- o "Sample ID"
- o "Specimen ID"
- o "Collection Day 1"
- o "Month 1"
- o "Year 1"

The script will index (assign a unique number to) the data in the "Observation No." field using the format YY#####, where YY is the two-digit abbreviation of the year when the script ran and ##### are five sequentially assigned digits. The script assigns these numbers by sorting the data by

- "Observation No."
- Then by, "Collector - Last Name"
- Then by, "Collector - First Name"
- Then by, "Month 1"
- Then by, "Collection Day 1"
- Then by, "Sample ID"
- Then by, "Specimen ID"

in ascending order, with blank values being put at the end. If all "Observation No." fields are blank in the merged data, the indexing will start at YY00000. Otherwise, the indexing will start at the previous largest "Observation No." plus one.

Help and Tips

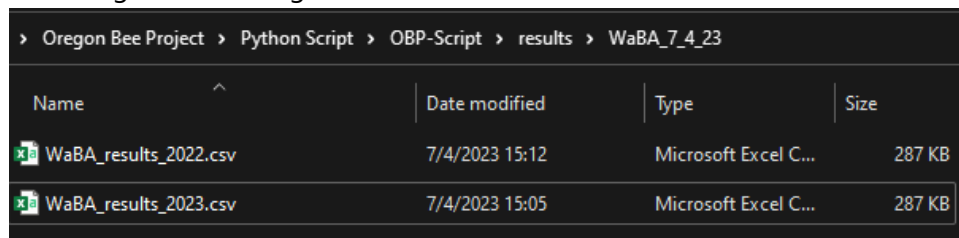
- **File Paths**

A file path is a string of characters representing the unique location of a file on a computer. It contains a sequence of slash-separated folders followed by a file name and its extension. This program accepts absolute and relative file paths and will display given file paths relative to the directory in which the program is running.

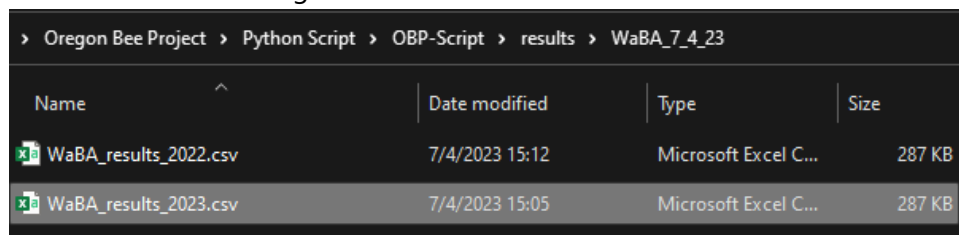
This script accepts file paths pasted directly from File Explorer. To copy a file path in File Explorer, do the following:

On Windows 11 and some versions of Windows 10,

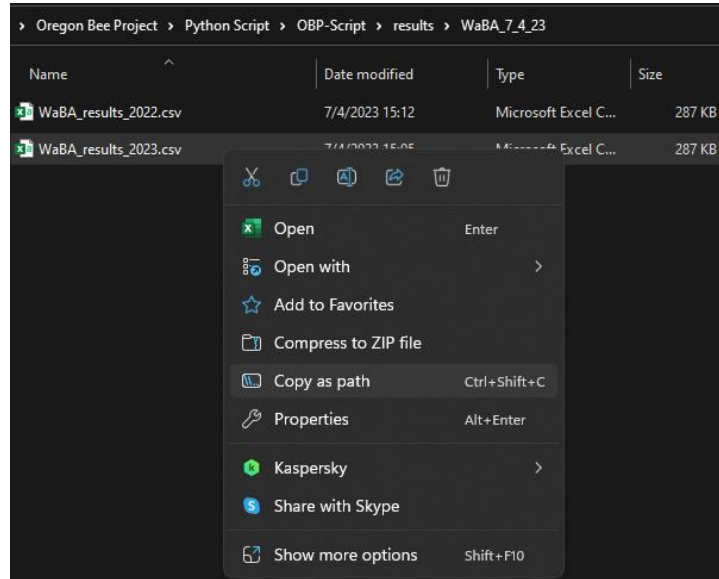
1. Navigate to the target file.



2. Left click on the target file to select it.



3. Right click on the target file and click "Copy as Path", or press Ctrl + Shift + C, to copy the file path.



4. Click the terminal window in which Merge_Process.bat is running.
5. When prompted for a file path, right click, or press Ctrl + V, to paste the file path.

```
C:\WINDOWS\system32\cmd. x + v

C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:

C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\

C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py

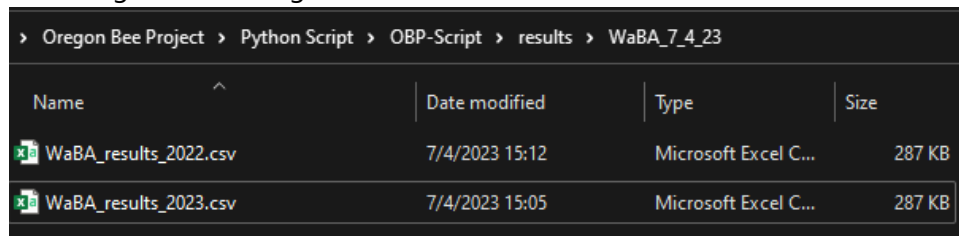
Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"

File to Append: the file that will be treated as new data
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\OBP-Script\results\WaBA_7_4_23\WaBA_results_2023.csv"
```

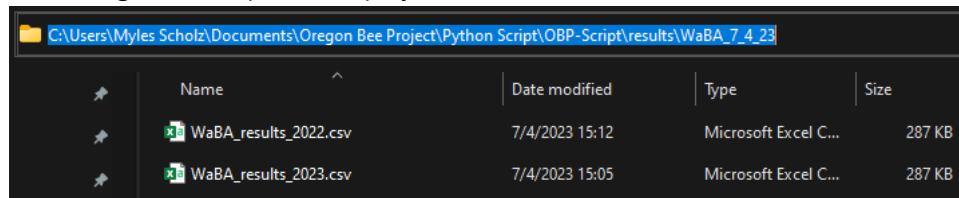
6. Press Enter to submit the file path.

On Windows 10,

7. Navigate to the target file's folder.



8. Click on the blank space in the file path bar near the top of the window. This should change the file path's display to text, which should be selected.



9. Press Ctrl + C, or right click and click "Copy", to copy the file path. This will only copy up to target file's parent folder.
10. Click the terminal window in which Merge_Process.bat is running.
11. When prompted for a file path, right click, or press Ctrl + V, to paste the folder path.

```
C:\WINDOWS\system32\cmd. X + -
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py

Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"

File to Append: the file that will be treated as new data
Enter a relative or absolute file path, including the file extension, or 'q' to quit: C:\Users\Myles Scholz\Documents\Or
egon Bee Project\Python Script\OBP-Script\results\WaBA_7_4_23\
```

12. Type slash and the target file's name and extension.

```
C:\WINDOWS\system32\cmd. X + -
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>c:
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>cd /d C:\Users\Myles Scholz\Documents\Oregon
Bee Project\Python Script\OBP-Script\
C:\Users\Myles Scholz\Documents\Oregon Bee Project\Python Script\OBP-Script>python merge.py

Base File: the file that will be treated as a pre-existing dataset
Enter a relative or absolute file path, including the file extension, or 'q' to quit: "C:\Users\Myles Scholz\Documents\O
regon Bee Project\Python Script\All Location Labels.csv"

File to Append: the file that will be treated as new data
Enter a relative or absolute file path, including the file extension, or 'q' to quit: C:\Users\Myles Scholz\Documents\Or
egon Bee Project\Python Script\OBP-Script\results\WaBA_7_4_23\WaBA_results_2023.csv\
```

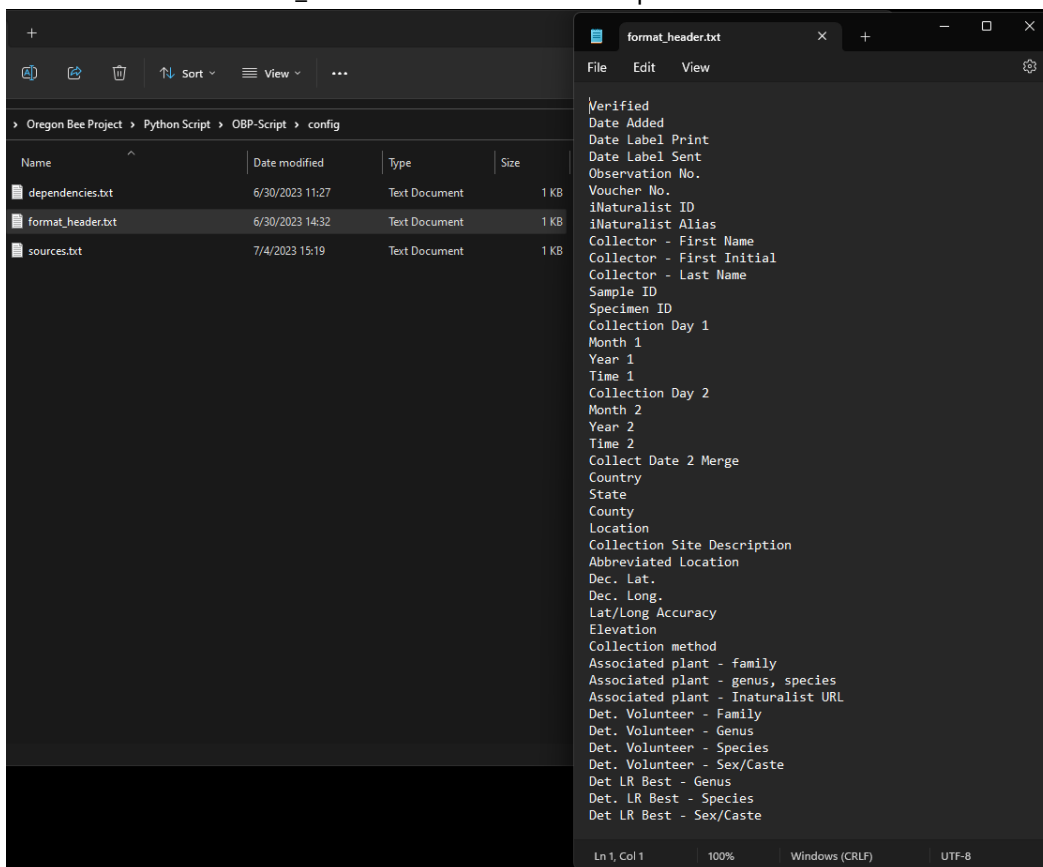
13. Press Enter to submit the file path.

- **Resolving Non-matching Header Errors (format_header.txt)**

The "base and append file headers do not match" error is easy to trigger with this script because the program requires that the headers of its input files match *exactly*. Resolving this issue depends on which files are being used as input and whether their headers can be modified.

If this error occurs when merging with "results" files from the data pulling and formatting process, the fix may be relatively simple. The "results" files use the header specified in OBP-Script/config/format_header.txt. This file lists the column names for the data, one on each line. To resolve the issue,

1. Check whether the number of columns and the names of the columns match between format_header.txt and the other input file.



2. If the headers have the same number of columns and the columns roughly correspond but they do not match in spelling, capitalization, or spacing,
 1. Modify the column names in either format_header.txt or the other input file so the two match exactly.
 2. If format_header.txt was changed, save it, and run the data pulling and formatting process again.
 3. Run the merging process again (with the newly pulled data, if applicable).

3. If the number of columns does not match or the columns do not correspond in order, there are a couple options:
 1. Manually modify the columns of either input file to match exactly. If reordering or renaming the columns, make sure that the data is reordered or rewritten in the same way.
 2. Modify the code (format_data.py) to change how data is formatted. Run the data pulling and formatting process again.
 3. After either, run the merging process again with the reformatted data.