

Software Environment Setup

Workshop on Delta Flow-Salinity Modeling Using Machine Learning
Pre-Workshop Setup Meeting, January 25, 2023

Raymond Hoang, Peyman Namadi, and Brad Tom
DWR Delta Modeling Section



CALIFORNIA DEPARTMENT OF
WATER RESOURCES



UCDAVIS
UNIVERSITY OF CALIFORNIA

Part B: Using Google Colaboratory (Colab)

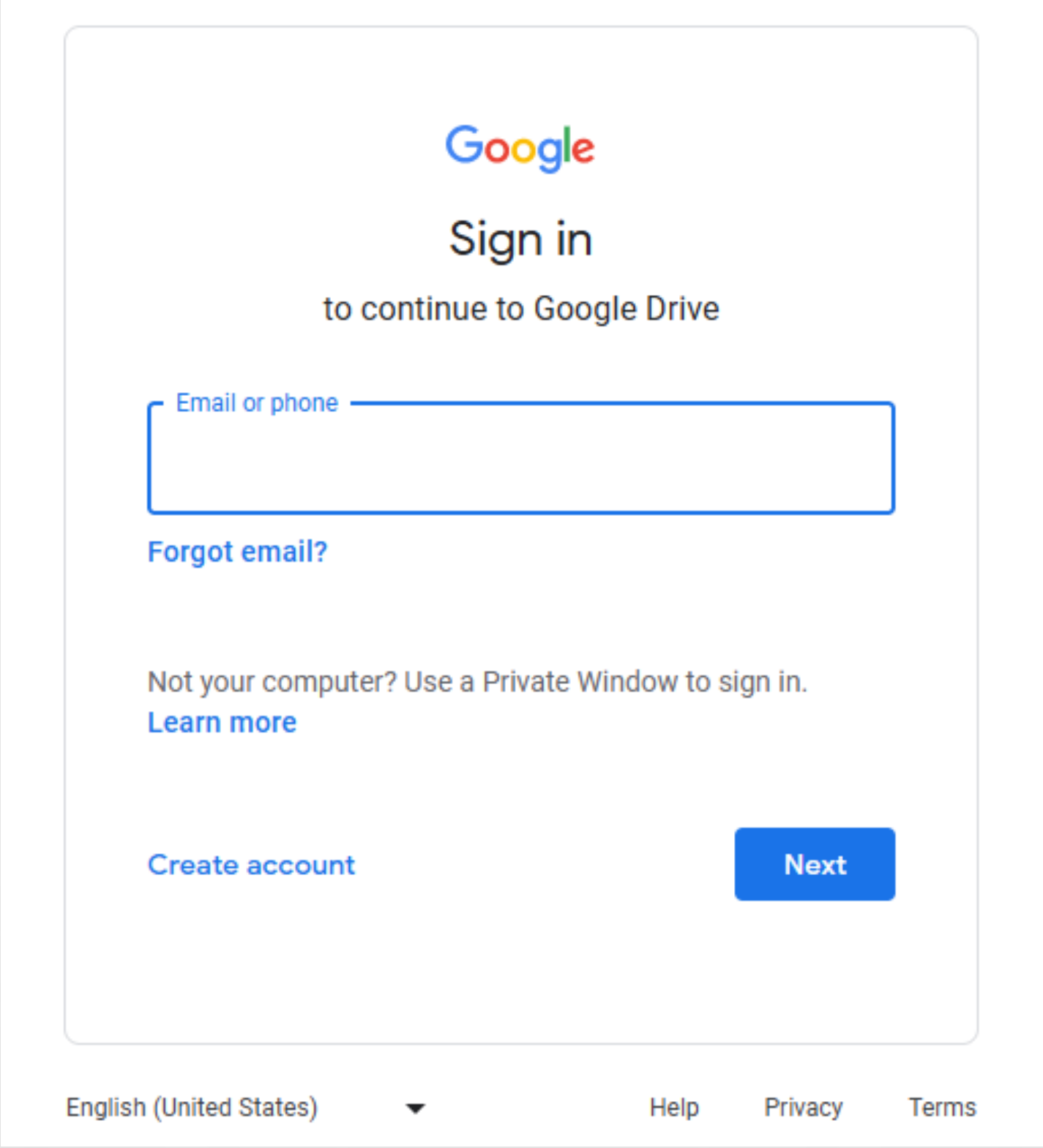


CALIFORNIA DEPARTMENT OF
WATER RESOURCES



UC DAVIS
UNIVERSITY OF CALIFORNIA

Step 1: Create a Google account or use an existing Google account and log in to your Google Drive at <https://drive.google.com>



The image shows the Google sign-in page for Google Drive. At the top is the Google logo, followed by the text "Sign in to continue to Google Drive". Below this is a text input field labeled "Email or phone". Under the input field is a link "Forgot email?". Further down is the text "Not your computer? Use a Private Window to sign in." with a link "Learn more". At the bottom left is a link "Create account", and at the bottom right is a blue button labeled "Next". The footer contains the text "English (United States)" with a dropdown arrow, and links for "Help", "Privacy", and "Terms".



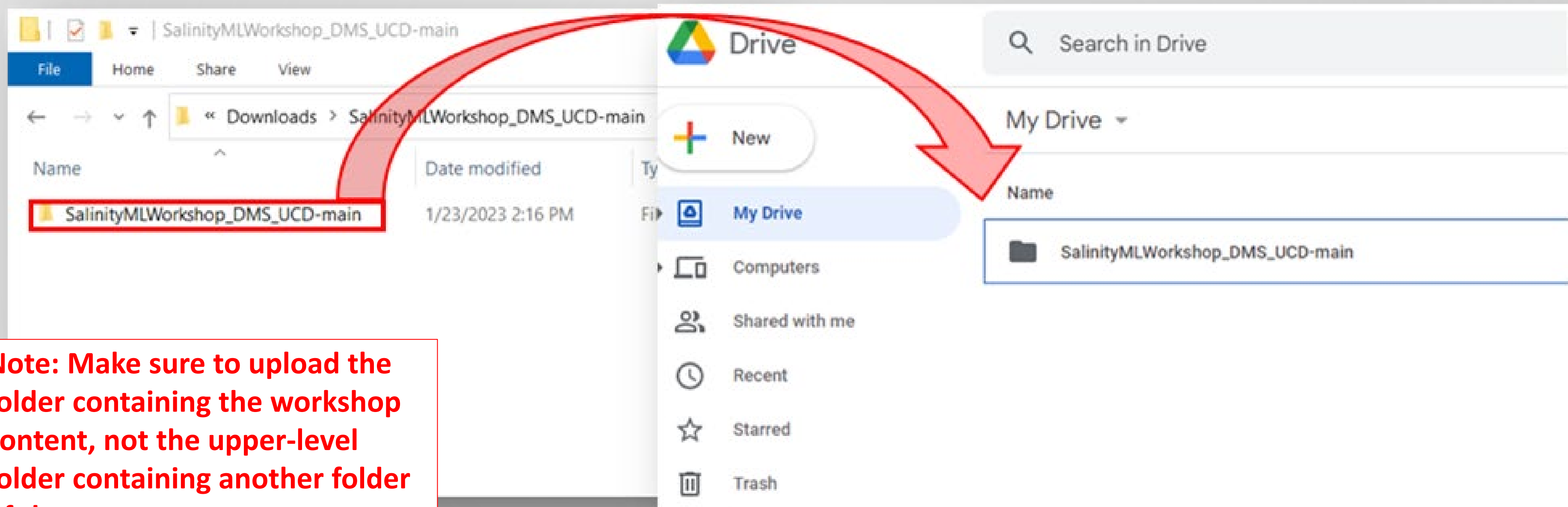
Step 2: Use this link to download the folder containing data and code:
https://github.com/CADWRDeltaModeling/SalinityMLWorkshop_DMS_UCD

The screenshot shows the GitHub repository page for 'CADWRDeltaModeling / SalinityMLWorkshop_DMS_UCD'. The repository is public and has 1 branch and 0 tags. The 'Code' button is highlighted with a red arrow labeled '1'. The dropdown menu for the 'Code' button is open, showing options for cloning the repository (HTTPS, SSH, GitHub CLI) and downloading the ZIP file. The 'Download ZIP' option is highlighted with a red arrow labeled '2'.

| File | Commit |
|-------------------------------------|------------------------------|
| Colab_Train_ANN_on_Augmented_Da... | make folder name consisted w |
| Local_Train_ANN_on_Augmented_Dat... | Rename Train_ANN_on_Augme |
| README.md | Update README.md |
| Salinity_DWR.yml | first commit |
| annutils.py | First commit |
| dsm2_ann_inputs_base.xlsx | First commit |
| dsm2_ann_inputs_dcc0.xlsx | First commit |
| dsm2_ann_inputs_dcc1.xlsx | First commit |
| dsm2_ann_inputs_rsacminus15day.xlsx | First commit |

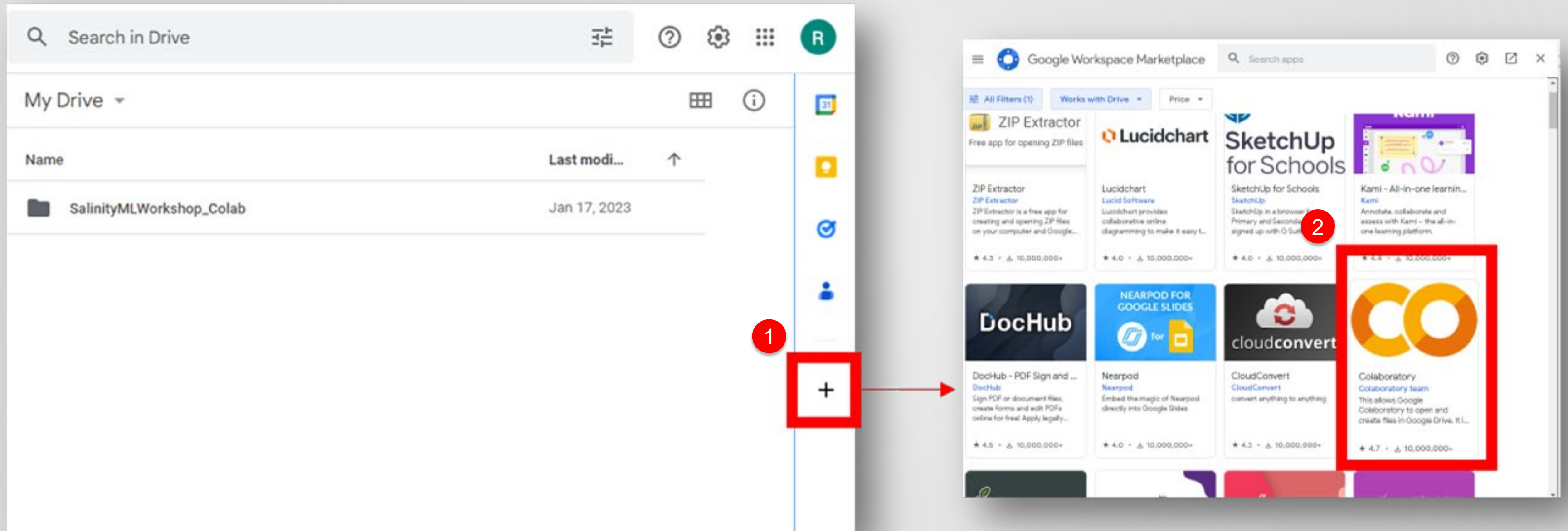
Step 3: Navigate to your download location and unzip the folder.

Step 4: Drag-and-drop the unzipped folder with the workshop content to your Google Drive's top-level directory (i.e., My Drive/).



Note: Make sure to upload the folder containing the workshop content, not the upper-level folder containing another folder of the same name.

Step 5: Link the Google Colab Add-On to your Google Drive. See screenshot below. When complete, refresh your browser for the changes to take effect.



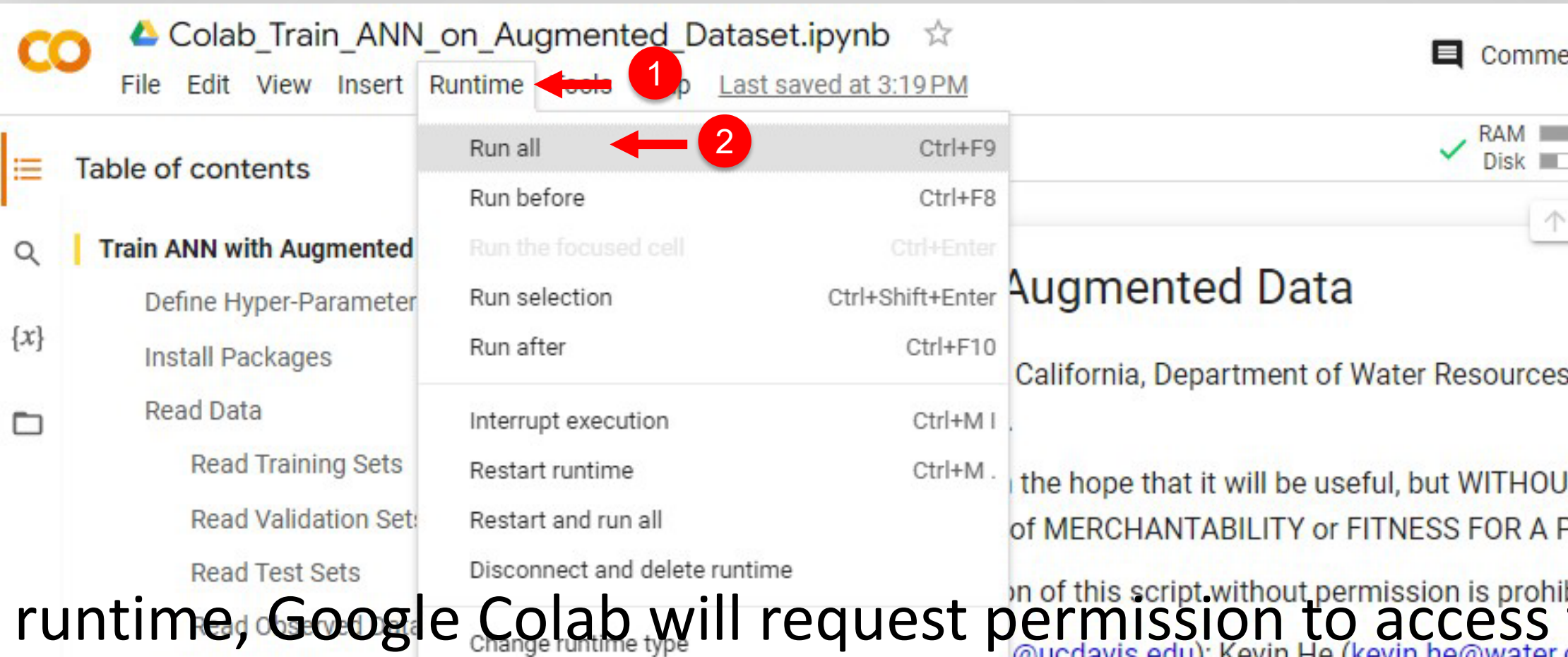
Step 6: Go into the “*SalinityMLWorkshop_DMS_UCD-main*” folder you uploaded and open ***Colab_Train_ANN_on_Augmented_Dataset.ipynb***. Right-click and choose “Open With”, and then select “Google Colaboratory”.

The screenshot shows a Google Drive interface with a list of files. A right-click context menu is open over the file 'Colab_Train_ANN_on_Augmented_Dataset.ipynb'. The 'Open with' option is selected, and a sub-menu is displayed showing 'Google Colaboratory' as a suggested app. Red annotations highlight the steps: 1. Right-click on the file. 2. Select 'Open with'. 3. Select 'Google Colaboratory'.

| File Name | Date | Size |
|--|--------------|--------|
| dsm2_ann_inputs_rsacminus20pct.xlsx | Jan 20, 2023 | 6.2 MB |
| dsm2_ann_inputs_rsacplus15day.xlsx | Jan 20, 2023 | 6.3 MB |
| dsm2_ann_inputs_rsanminus20pct.xlsx | Jan 20, 2023 | 6.3 MB |
| dsm2_ann_inputs_rsanminus15day.xlsx | Jan 20, 2023 | 6.3 MB |
| dsm2_ann_inputs_rsanplus15day.xlsx | Jan 20, 2023 | 6.3 MB |
| dsm2_ann_inputs_rsanplus20pct.xlsx | Jan 20, 2023 | 6.3 MB |
| martinez_stage1hr.csv | Jan 20, 2023 | 7.2 MB |
| dsm2_ann_inputs_smscg1.xlsx | Jan 20, 2023 | 6.3 MB |
| dsm2_ann_inputs_smscg0.xlsx | Jan 20, 2023 | 6.2 MB |
| observed_data_daily.xlsx | Jan 20, 2023 | 3 MB |
| Colab_Train_ANN_on_Augmented_Dataset.ipynb | Jan 23, 2023 | 1.6 MB |

Note: If Google Colaboratory does not show up, you may need to refresh your browser.

Step 7: With the script now open in Colab, run the script by clicking *Runtime: Run All*.



Note: during code runtime, Google Colab will request permission to access the folder – follow the prompts to allow access.

Permit this notebook to access your Google Drive files?

This notebook is requesting access to your Google Drive files. Granting access to Google Drive will permit code executed in the notebook to modify files in your Google Drive. Make sure to review notebook code prior to allowing this access.

[Connect to Google Drive](#)

Choose an account

to continue to [Google Drive for desktop](#)

Your Account

Use another account

view the activity record of files in your Google Drive

See, edit, create, and delete any of your Google Drive documents

Make sure you trust Google Drive for desktop

You may be sharing sensitive info with this site or app. You can always see or remove access in your [Google Account](#).

Learn how Google helps you [share data safely](#).

See Google Drive for desktop's [Privacy Policy](#) and [Terms of Service](#).

[Cancel](#) [Allow](#)

CALIFORNIA DEPARTMENT OF WATER RESOURCES

UC DAVIS UNIVERSITY OF CALIFORNIA

Each cell of the script should now commence running sequentially. The script will take about 10 minutes to run, depending on Colab resources.

This icon indicates the cell is currently running



```
observed_stations_ordered_by_median = ['RSMKL008', 'RSAN032', 'RSAN037', 'RSAC092', 'SLTRM004', 'RO  
CHVCT000', 'RSAN018', 'CHSWP003', 'CHDMC006', 'SLDUT007  
OLD_MID', 'RSAN058', 'ROLD059', 'RSAN007', 'RSAC081', '  
'RSAC075', 'SLMZU011', 'SLSUS012', 'SLCBN002', 'RSAC064'  
  
num_sheets = 9  
  
output_stations = None  
  
xscaler = None  
yscaler = None  
  
for data_file in train_data + [val_data,] + list(test_data.values()) + list(extra_data.values()):  
    data_path = os.path.join(gdrive_root_path, data_file)
```

This icon indicates the cell has completed running














```
[4] observed_stations_ordered_by_median = ['RSMKL008', 'RSAN032', 'RSAN037', 'RSAC092', 'SLTRM004', 'RO  
CHVCT000', 'RSAN018', 'CHSWP003', 'CHDMC006', 'SLDUT007  
OLD_MID', 'RSAN058', 'ROLD059', 'RSAN007', 'RSAC081', '  
'RSAC075', 'SLMZU011', 'SLSUS012', 'SLCBN002', 'RSAC064'  
  
num_sheets = 9  
  
output_stations = None  
  
xscaler = None  
yscaler = None  
  
for data_file in train_data + [val_data,] + list(test_data.values()) + list(extra_data.values()):  
    data_path = os.path.join(gdrive_root_path, data_file)
```

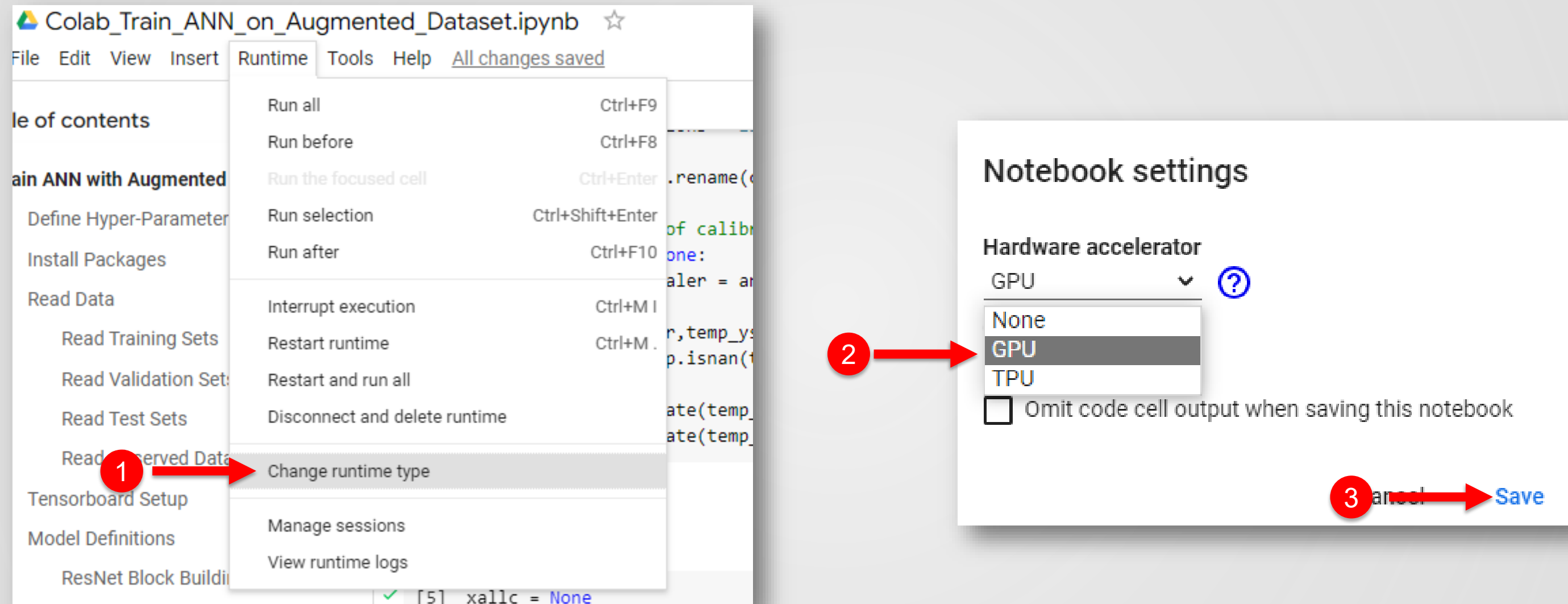
Post-Run:

Upon completion, the outputs will be in these three folders

My Drive > SalinityMLWorkshop_DMS_UC... ▾

| Name | |
|---|--|
|  | __pycache__ |
|  | models |
|  | results |
|  | images |
|  | Local_train_ann_on_augmented_dataset.py |
|  | Salinity_DWR.yml |
|  | Local_Train_ANN_on_Augmented_Dataset.ipynb |
|  | annutils.py |
|  | dsm2_ann_inputs_base.xlsx |
|  | dsm2_ann_inputs_dcc0.xlsx |
|  | dsm2_ann_inputs_rsacminus15day.xlsx |

Optional Step: You may opt to use free GPU acceleration offered by Colab to decrease the run time by clicking Runtime: Change runtime type and selecting GPU. This can decrease runtimes by approximately 40%.



Repeat Step 7 after making these changes.

Questions?

Kevin.He@water.ca.gov



CALIFORNIA DEPARTMENT OF
WATER RESOURCES



UC DAVIS
UNIVERSITY OF CALIFORNIA