

Statistics for Data Science

Introduction



Introductions

- Your name
- The place you call home
- A number that has some significance to you



Classroom guidelines

- Ask lots of questions
 - You are welcome to unmute and ask your question at any time
 - Raise hand button in Zoom
 - You can also use the chat
- Help each other; learn from each other
- Get comfortable with discomfort. Making mistakes, figuring them out, and then correcting them is part of the learning process



Class Format

- New concepts introduced through slides or interactive demonstrations
- Jupyter Notebook to put these concepts into code
- Exercises to practice + build off of what was learned in class
 - Exercises will be reviewed at the beginning of class on Tuesday (with one exception at the end)
- Some review questions at the beginning of class on Thursdays
- For students who have or are planning to do the data science bootcamp, we'll have a post-course assessment



Topic Outline

Week 1: Introduction to Statistics, Descriptive Statistics, and Single-Variable EDA

Week 2: Multi-variable EDA, Introduction to Probability

Week 3: Probability and Random Variables (Binomial and Normal Distributions)

Week 4: Estimation, Sampling Distributions, Confidence Intervals, and the Bootstrap

Week 5: Hypothesis Testing

Week 6: Linear Regression

Week 7: Logistic Regression



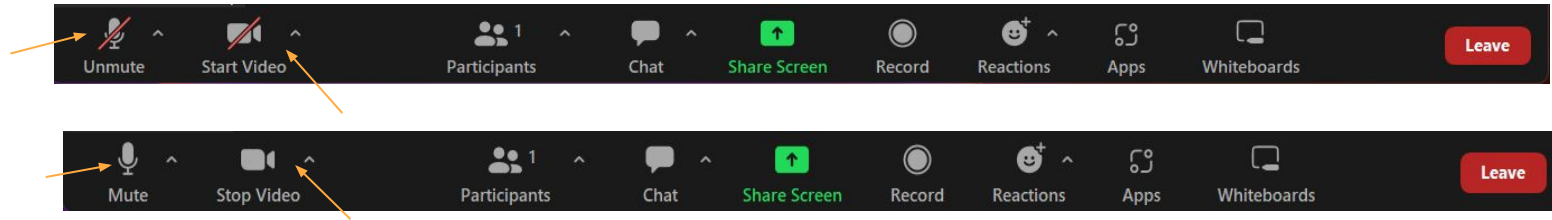
Checklist

- Joined Slack
- Installed Anaconda



Zoom Features - Audio and Camera Settings

The buttons to mute/unmute and turn your camera off/on are located along the bottom of the Zoom window. Simply click the relevant icon to switch yourself from muted to unmuted or your camera from off to on. For more advanced audio and video settings, click on the carrot next to the relevant icon.

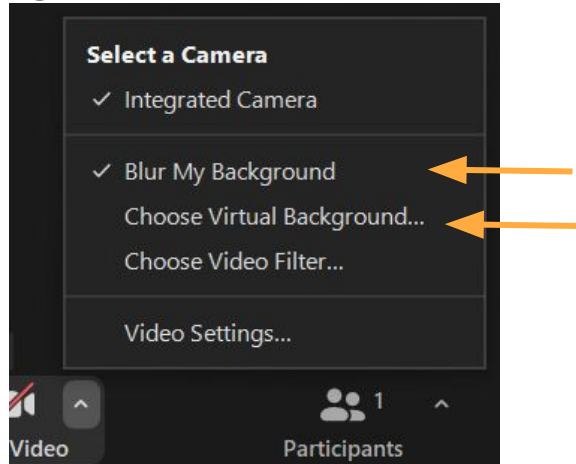


Icon appearances when unmuted and camera on:



Zoom Features - Backgrounds

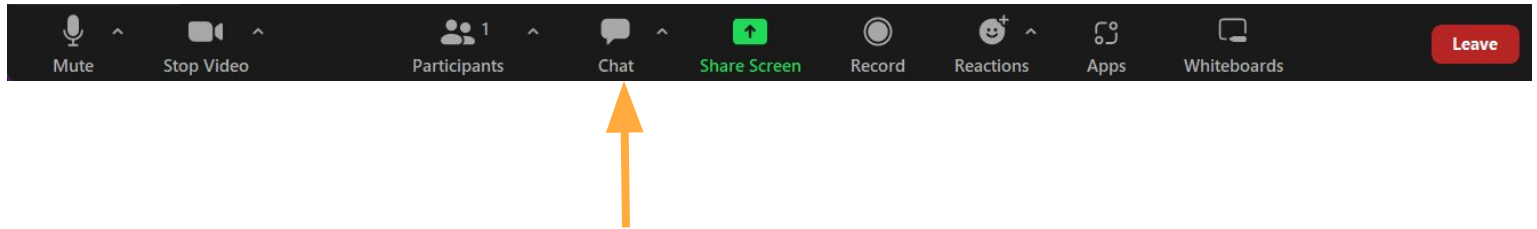
You may at times be working in a space that has a distracting appearance. You can make use of Zoom's built in features to maintain a professional background appearance. By clicking the carrot next to the video icon, you can blur your background or open the Virtual Background options to choose a digital background.



Zoom Features - Chat

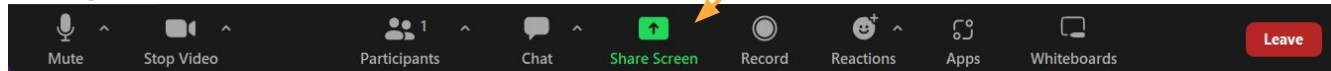
During lectures, additional information/links will often be shared in the session's chat, so keep an eye out for activity here.

This can also be a good place to add your observations and/or questions if you don't want to interrupt the person who is speaking.

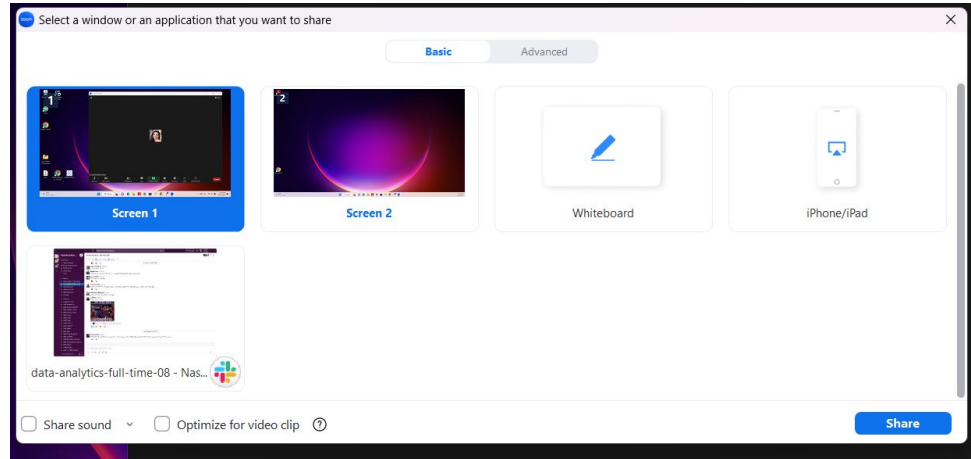


Zoom Features - Share Screen

If you are having difficulty with something, we will often ask you to share your screen so that we can help you troubleshoot. To initiate screen sharing, click on the green “Screen Share” button:

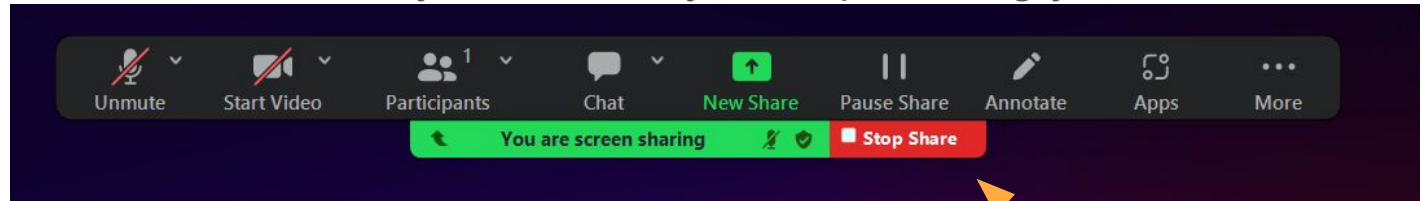


Then select the relevant screen or specific window to share:



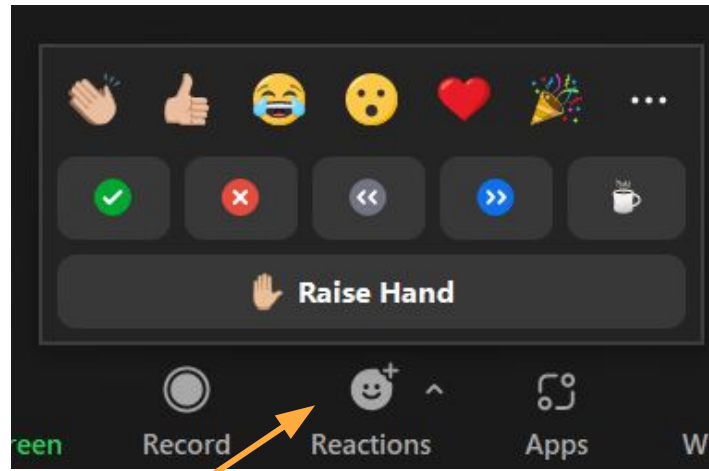
Zoom Features - Screen Share

Once you are sharing your screen, the Zoom controls tend to move to the top of the screen, though if you have multiple screens, they may move to a screen you are NOT sharing. Use the red “Stop Share” button when you are ready to stop sharing your screen.



Zoom Features - Reactions

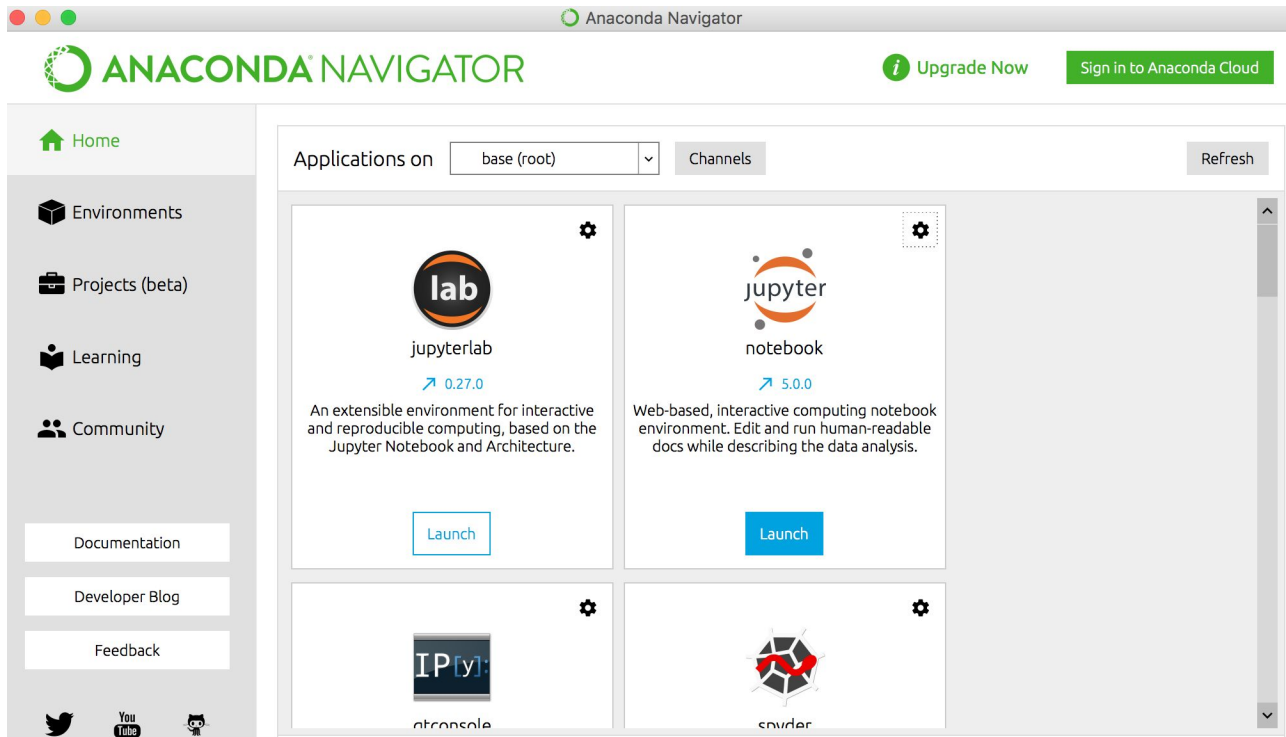
You can click the “Reactions” button to see available reactions. Options like the green check or “Raise Hand” will remain in place until you or the host remove them. Other reactions like clapping will only remain visible for a short time.



Orientation to Jupyter Notebook

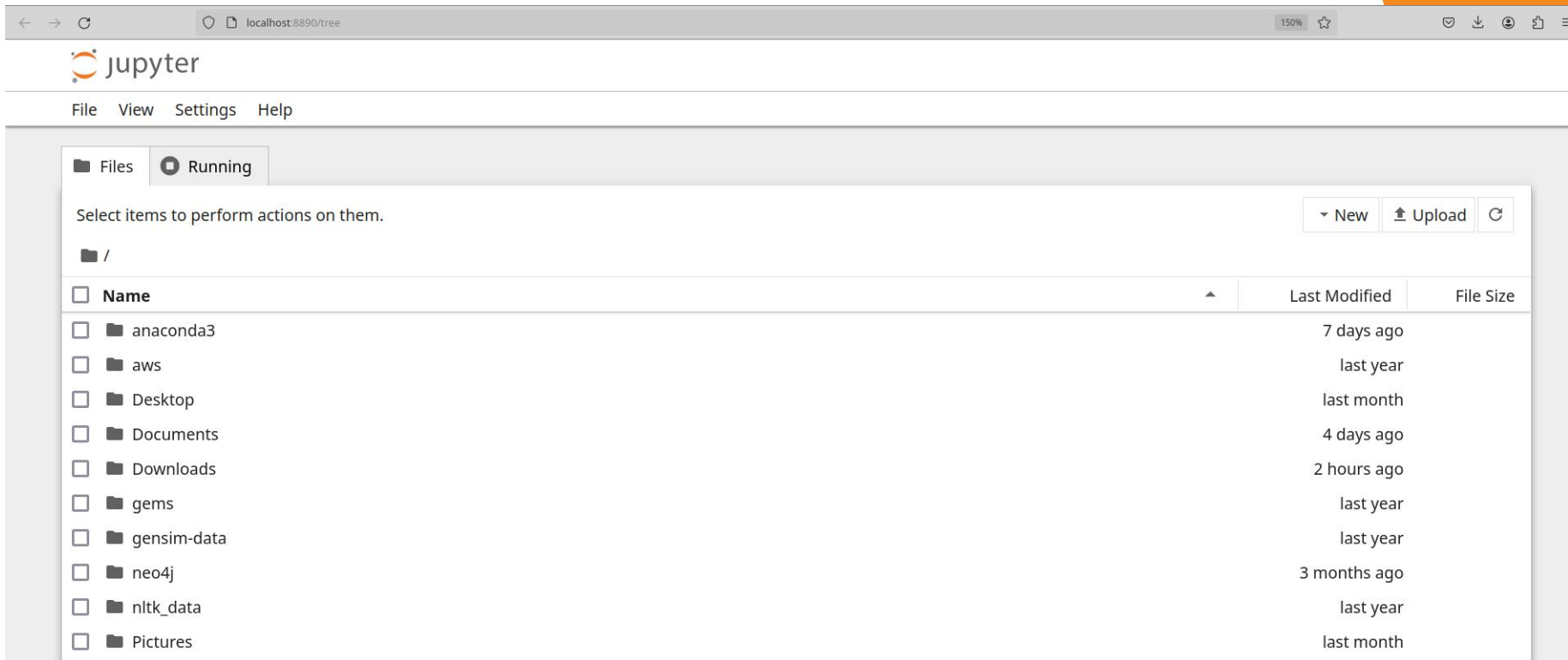
NASHVILLE  SOFTWARE SCHOOL

Open Anaconda Navigator, install, and launch Jupyter Notebook



A new tab will open in your default browser. It's not actually connecting to the internet; it's just running on your machine

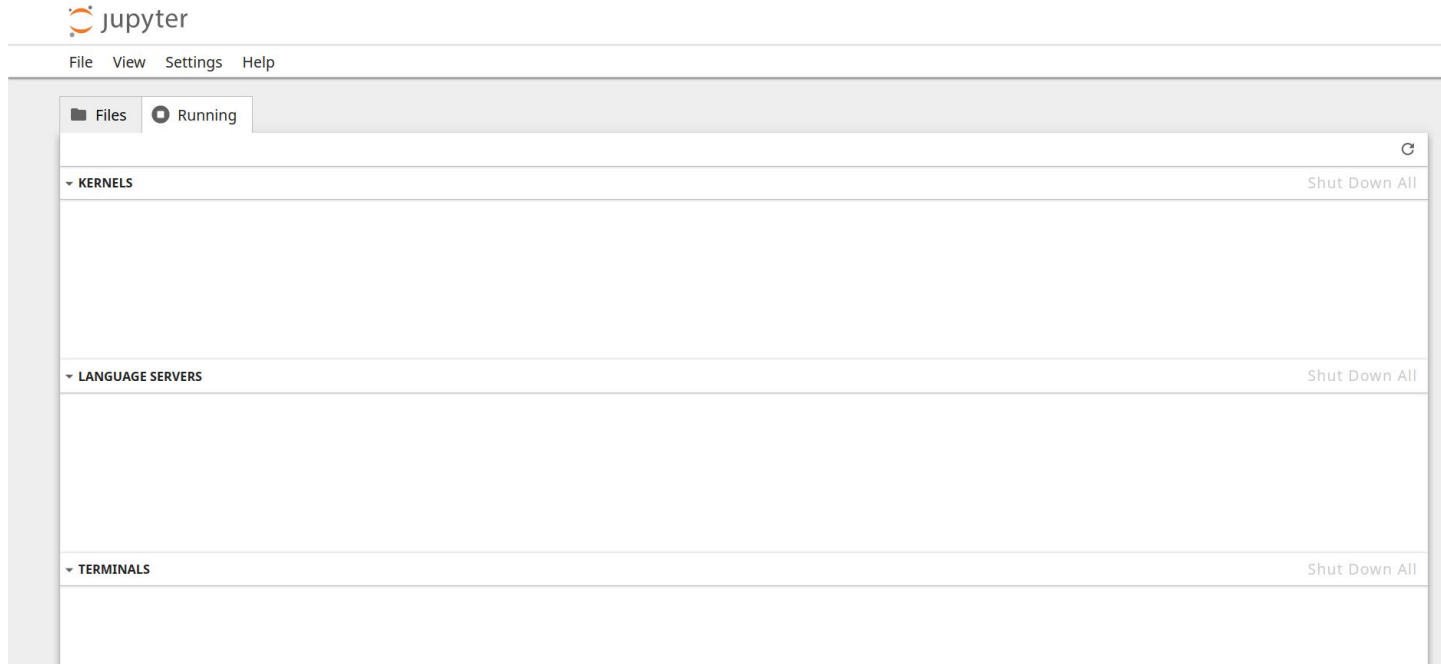
You will see the file structure from your computer, and you can navigate as normal



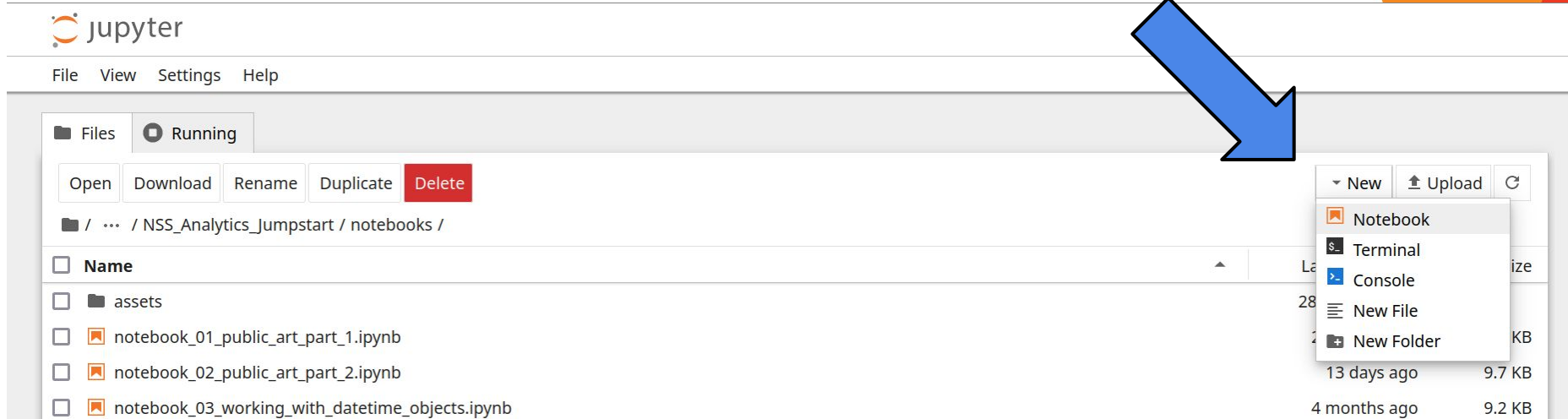
The screenshot shows the JupyterLab web interface in a browser window. The address bar shows 'localhost:8890/tree'. The JupyterLab logo and 'jupyter' text are at the top left. Below the logo is a menu bar with 'File', 'View', 'Settings', and 'Help'. The main interface has two tabs: 'Files' (selected) and 'Running'. Below the tabs is a message 'Select items to perform actions on them.' and three buttons: 'New', 'Upload', and a refresh icon. Below this is a file browser table showing the local file system structure.

<input type="checkbox"/> Name	Last Modified	File Size
<input type="checkbox"/> /		
<input type="checkbox"/> anaconda3	7 days ago	
<input type="checkbox"/> aws	last year	
<input type="checkbox"/> Desktop	last month	
<input type="checkbox"/> Documents	4 days ago	
<input type="checkbox"/> Downloads	2 hours ago	
<input type="checkbox"/> gems	last year	
<input type="checkbox"/> gensim-data	last year	
<input type="checkbox"/> neo4j	3 months ago	
<input type="checkbox"/> nltk_data	last year	
<input type="checkbox"/> Pictures	last month	

See what notebooks are already running (should be empty if just opening Jupyter)



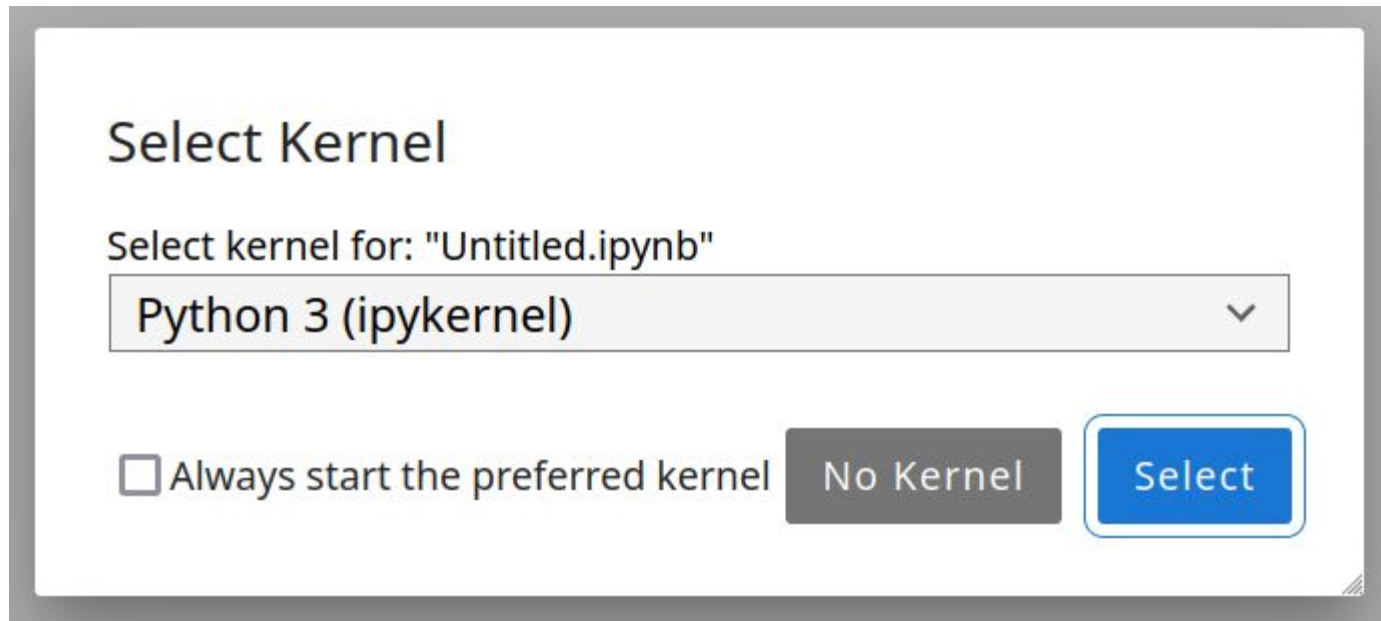
Navigate to `analytics_jumpstart/notebooks` and create a new notebook.



The screenshot displays the JupyterLab web interface. At the top, the 'jupyter' logo is visible, followed by a menu bar with 'File', 'View', 'Settings', and 'Help'. Below the menu bar, there are two tabs: 'Files' and 'Running'. The 'Files' tab is active, showing a file browser. The current path is '/ NSS_Analytics_Jumpstart / notebooks /'. A large blue arrow points from the top right towards the 'New' button in the file browser. The file browser shows a list of files and folders. The 'New' button is open, showing a dropdown menu with options: 'Notebook', 'Terminal', 'Console', 'New File', and 'New Folder'. The 'Notebook' option is selected.

Name	Last Modified	Size
assets		
notebook_01_public_art_part_1.ipynb	28	
notebook_02_public_art_part_2.ipynb	2	
notebook_03_working_with_datetime_objects.ipynb	13 days ago	9.7 KB
	4 months ago	9.2 KB

When prompted, select the Python 3 kernel.



The image shows a 'Select Kernel' dialog box from a Jupyter Notebook interface. The title 'Select Kernel' is at the top. Below it, the text 'Select kernel for: "Untitled.ipynb"' is displayed. A dropdown menu shows 'Python 3 (ipykernel)' as the selected option. At the bottom, there is a checkbox labeled 'Always start the preferred kernel' which is currently unchecked. To the right of the checkbox are two buttons: 'No Kernel' and 'Select'. The 'Select' button is highlighted with a blue border.

Select Kernel

Select kernel for: "Untitled.ipynb"

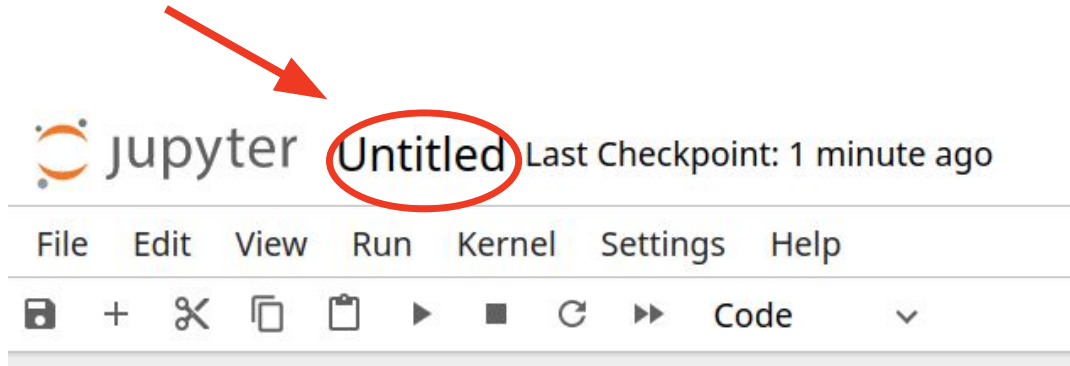
Python 3 (ipykernel)

☐ Always start the preferred kernel

No Kernel

Select

Give your notebook a title



Rename File

File Path
Untitled.ipynb

New Name

****The notebook name should be meaningful and contain no spaces**

Useful buttons (and shortcuts) for running code and moving cells around

The image shows the JupyterLab interface with several annotations pointing to specific buttons and shortcuts:

- save notebook**: Points to the save icon (floppy disk) in the toolbar.
- run code in selected cell (shift + enter)**: Points to the run button (play icon) in the toolbar.
- move cell (up/down)**: Points to the up and down arrow buttons in the bottom right corner of the cell toolbar.
- toggle between markdown and code**: Points to the 'Code' button in the toolbar.
- add cell**: Points to the '+' button in the toolbar.
- cut/copy/paste cell**: Points to the scissors, clipboard, and document with plus icons in the toolbar.

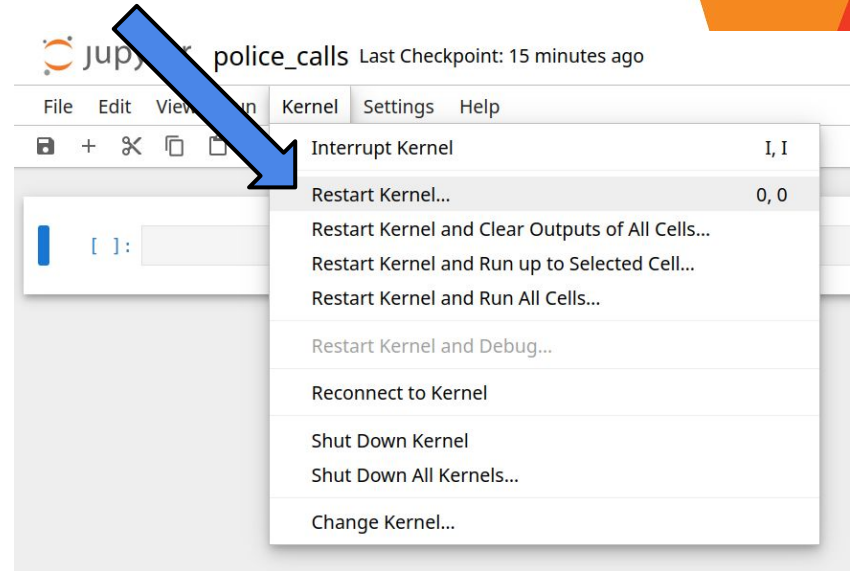
The interface shows the JupyterLab logo, the notebook name 'police_calls', and the last checkpoint time 'Last Checkpoint: 3 minutes ago'. The menu bar includes File, Edit, View, Run, Kernel, Settings, and Help. The toolbar includes icons for saving, adding, cutting, copying, pasting, running, and toggling between code and markdown. The bottom right corner of the cell toolbar includes icons for moving the cell up and down, adding a new cell, and deleting the cell.

If your code is taking too long or is giving unexpected results, try restarting the kernel

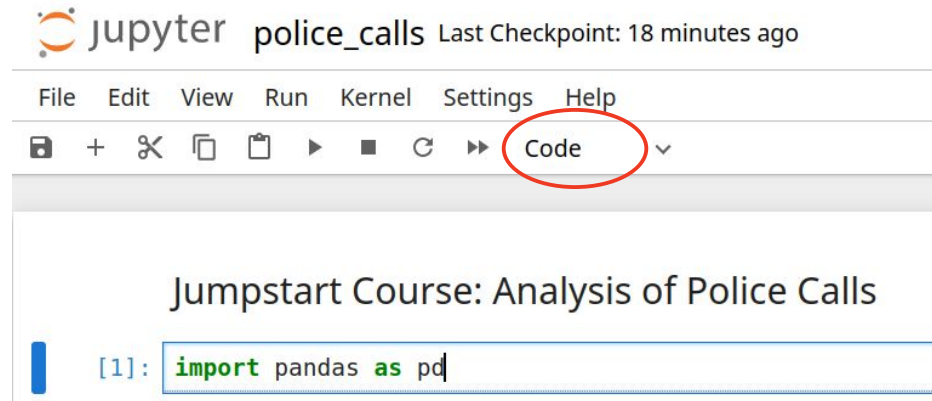
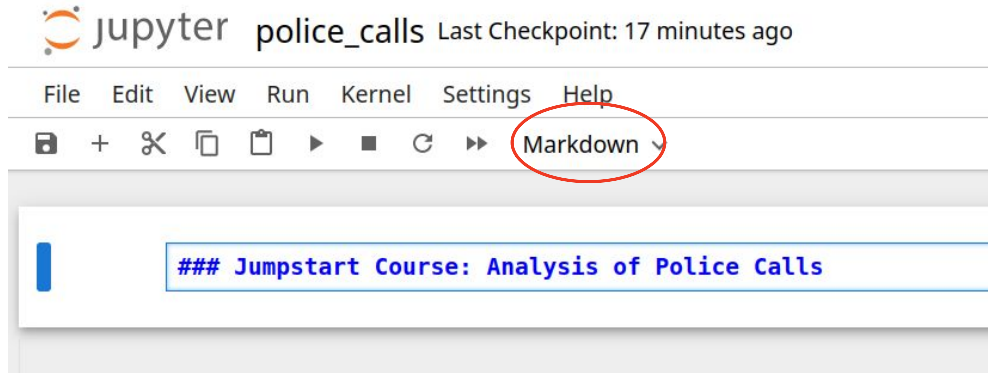
Each time you run a piece of code in a Jupyter Notebook, that process is saved to a kernel. All the inputs, outputs, variables, etc. are saved.

Even if you modify or delete a cell, the earlier instances of it being run were saved. This can sometimes lead to strange results. Restarting the kernel will clear out the memory so you can start fresh.

Closing and opening the notebook will also do this, but remember to rerun all your code after you restart the kernel!



Change the format of the cell to add notes or run code



Questions?