

Data Science – intro 2

Course text, Jupyter Notebooks, arithmetic (the computer way),
variables

Course TAs

- Arpan Nayak
- Thomas Roberts
- Poppy Aves

Keep learning

- I have learned and used lots of languages in my data science journey
 - assembly, Basic, FORTRAN, COBOL, Perl, C, C++, R, Python
 - At the same time, I am relatively new to Python
 - I can make mistakes. Making mistakes is part of coding.
 - If it looks like I am doing something wrong, ask. I may be and we will correct it together.
 - Seeing the process of making and correcting mistakes is important (how to find and correct mistakes is one of the main things coders need to learn)
 - The goal is not to make no mistakes. That is impossible.
 - The goal is to learn to test for, find and fix mistakes
- No prior knowledge is assumed for this course

Outline – What we will cover – Getting set up for the course

- Course resources
 - VLab and jupyter notebooks
 - Piazza (Q & A)
 - CANVAS page
 - e-textbook
 - Student Whatsapp group?
- Surviving the computer – some advice
- Our first exercise – Dracula

Course organization – course map

- The course runs on two tracks in parallel:

- 1) Python programming concepts
- 2) Data science concepts

These don't always look connected (and sometimes they aren't), but you need both.

More detailed map for each week

Week	Data concepts	Python	Examples
1-2	Sampling and distributions	Python basics: Variables, operations, functions	
3-5	Simulations	Arrays, loops, data frames	Births of boys/girls, Ethnicity of the jury
7	Permutations	Random shuffle, 'if' statement, compare means (t-tests)	McDonalds/Starbucks, Mosquitos and beer
8	Linear relationship	Handling missing data, correlation	Fertility and GDP
9	Linear relationship	Optimization, regression	Cholesterol levels and heart disease
10	Count data	chi-square test	Titanic

Guide to course tools

- CANVAS page – has links to tools and announcements
- Lecture notes and exercises – Vlab - notes have runnable sections of python if you want to experiment with them
- Exercises – Vlab – Make sure you do the exercises
 - This is the one most important tip for success
 - The computer provides feedback to guide you during the exercises
 - If you don't understand something in an exercise, please post to piazza and/or ask in class – Questions are essential and you do everyone else a service by asking them
- Online textbook – Information to support lectures and activities – Make sure you read the weekly sections.
- Piazza discussion board – Q&A

Assessment

- Homework 1 – formative
- Homework 2 – summative (counts for your mark)
- Group project
- Individual reflective statement covering your project work

A note about AI

- AI is NOT always a positive force when learning to program. It depends how you use it.
- I suggest you just use google search at the beginning and not AI
- Why?
- You might have the impression that ChatGPT is a better programmer than you (now)
- You are probably right
- Guess what? There are lots of humans who are also better programmers than you.
- Do you want to learn to program or just learn to copy things?

A note about AI

- There is a level of programming and data analysis that AI can't do (and has no immediate prospect of doing)
 - This is the thinking part
 - This is a valuable skill
-
- You can't learn it without going through the beginning phase where ChatGPT (and other humans) are better than you
 - No change – it has always been this way
 - There is no substitute for learning it yourself and that means doing it yourself

A note about AI

- I cannot determine how you use the course
- If you want a valuable skill, you will do things yourself
- You might be able to get by by doing very little work and copying ChatGPT. Many have tried. Some have failed. You won't know until it is too late if your strategy has worked
- If you adopt the AI-dependency route you will have a 'skill' that is **guaranteed** to have ZERO value
- Who will hire you when they can hire ChatGPT

A note about AI

- There is lots of good information for programmers on the internet
- Stackoverflow is one site you will frequently see
- Use these sources, but use them to learn, not to copy
- Our assessments (e.g. projects) are much more weighted to understanding now, and less to coding. Coding is just a basic level on the way to understanding. We will ask you to explain code and also to explain what data mean.

Guide to the CANVAS page:

<https://canvas.bham.ac.uk/courses/81492>

The screenshot shows the Canvas LMS interface for the course 'LI/LH/LM Data Science for Brain and Behaviour'. The browser address bar shows the URL 'canvas.bham.ac.uk/courses/81492'. The course title is displayed at the top. A search bar and 'View as Student' button are present. A left sidebar contains a navigation menu with items: Home, Announcements, Modules, Assignments, Grades, Discussions, Quizzes, Panopto, Rubrics, Zoom, Pages, Outcomes, Syllabus, and Collaborations. The main content area is titled 'Recent announcements' and shows the course ID '39530/39531/40426 LI/LH/LM Data Science for Brain and Behaviour'. Below the title are buttons for 'Assign to', 'Edit', and a menu icon. A collage of historical and modern figures is displayed below the buttons. A blue circular help icon with a question mark is located to the right of the announcement section.

Module Overview and timetable information	Week by Week	Assessment & Feedback	Project information	Project marking rubric	Reflective statement information (and link to rubric)	EDI information
Discussion board (Piazza)	How to download Homework and submit to CANVAS	Spreadsheet for listing projects/ members/ data sources	How to submit your project	How to share your project with me	How to set up a local data science environment on Windows (so you can do work locally instead of on deepnote)	How to set up a local data science environment on Mac

[Online course text \(click to access\)](#)

[Panopto folder for the course](#)

[Using AI/ChatGPT and other internet sources](#)

[An interesting perspective on AI bots](#)

Accessing VLab

- You should be able to sign in to Vlab with your normal login. Try it.
 - <https://vlab.oc1.aws.cs.bham.ac.uk/>
 - This gives you a 'personal space' to:
 - Do exercises
 - Do your homework
 - Do your projects
 - View lectures (and run examples from lecture)
 - You need to copy exercises from a course repository to your personal space in order to work on them
 - We will set that up together today.
- If you can't log in using your UoB login, don't panic
 - There are always some people who are not on the course lists for one reason or another.
 - Let me know your email address and I will get you put on
 - You will not be falling far behind.

Type the link in your browser (or click on it)
You will see something like this...

Select the environment

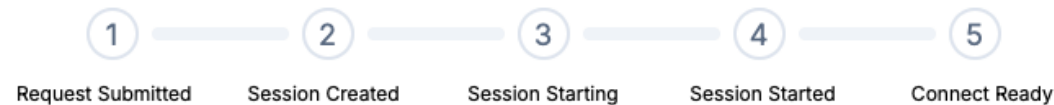
Environment Selection

AO Andrew Olson

Support

NAME	AVAILABLE FROM	EXPIRES ON	SESSION SESSION LENGTH	SUBMIT LEFT	USAGE
<input type="radio"/> Jupyter - Psychology (Autumn 2025)	2025-09-25 00:00	2025-11-09 00:00	2h 00m	148h	1.3%

Session Status:

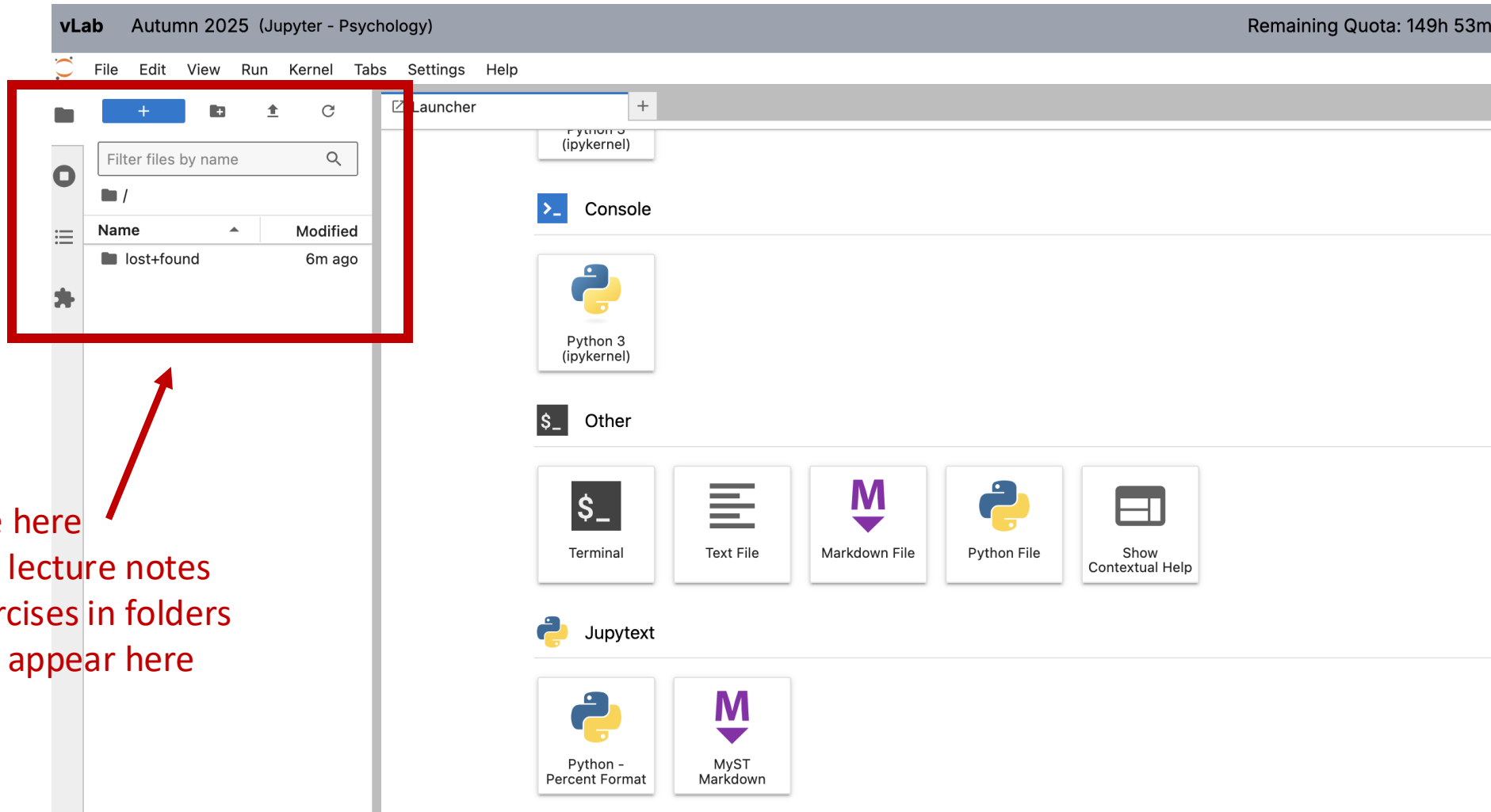


Action:

Start Session

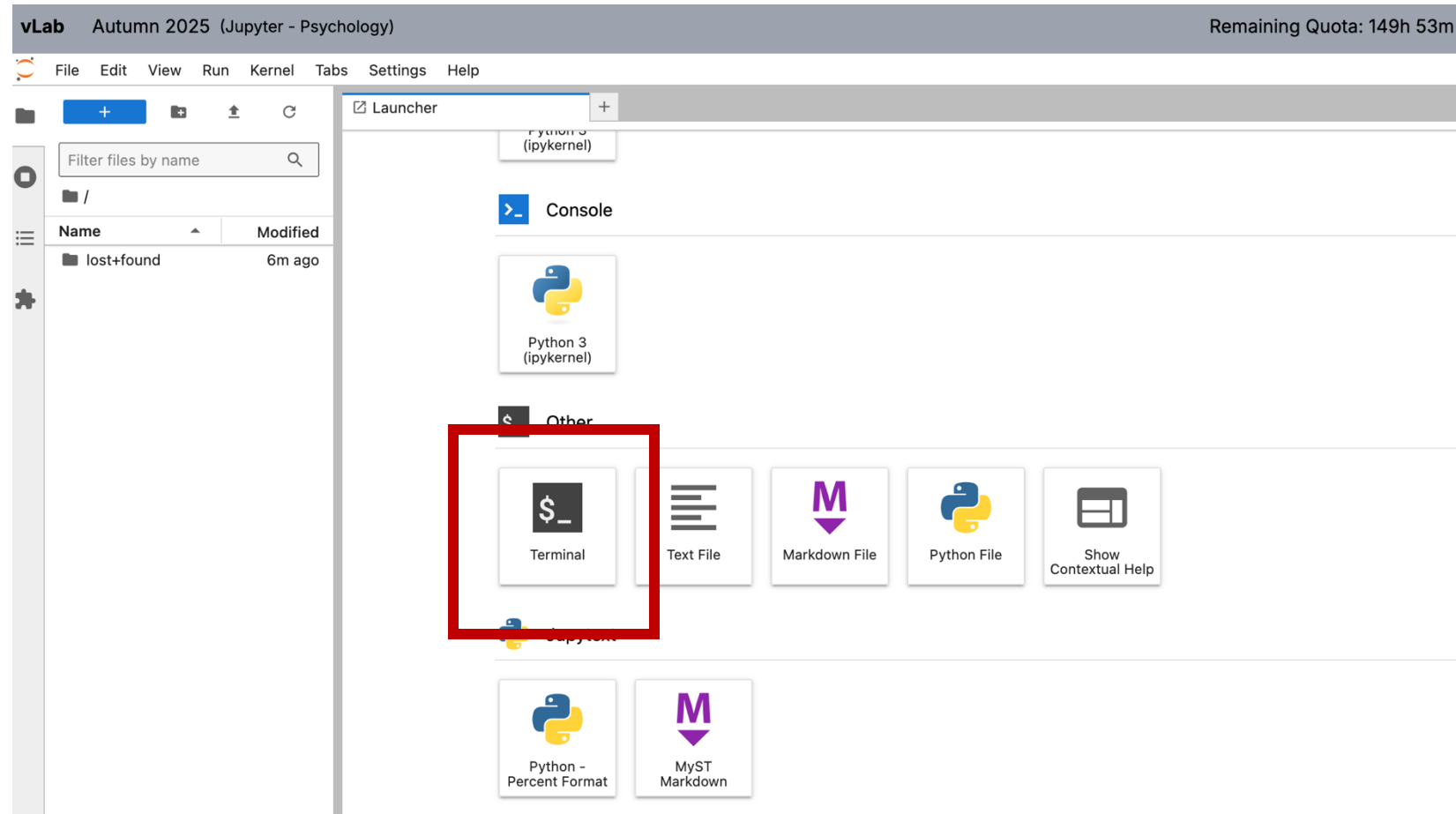
click on "start session"

When the environment starts you will see:



Setting up for getting exercises and notes

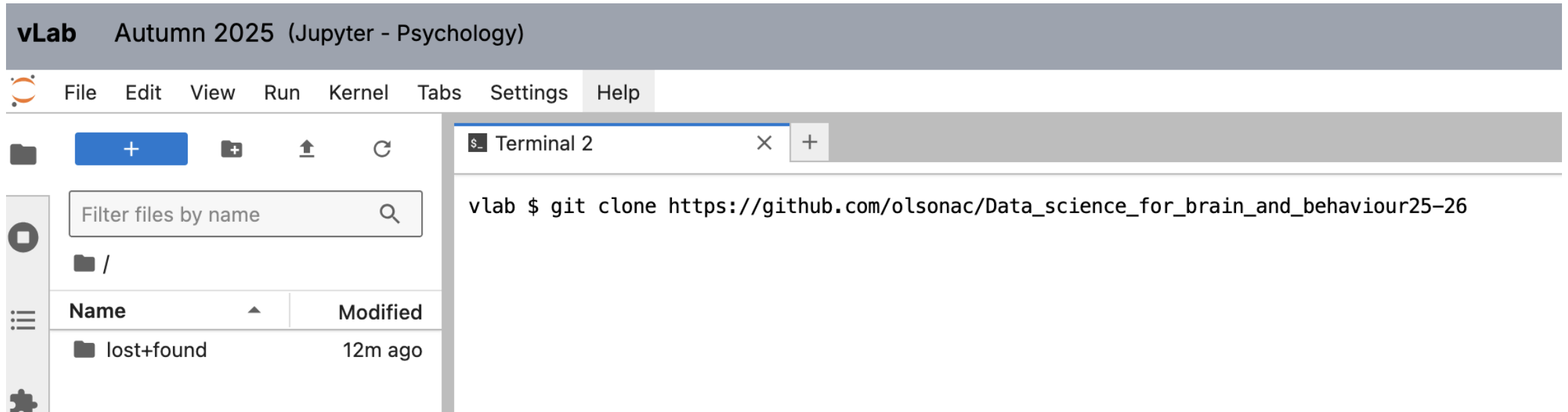
- We use a popular system for sharing code – ‘git’ and ‘github’
- We only need to set up once
- Start a terminal – Click on the ‘terminal’ icon



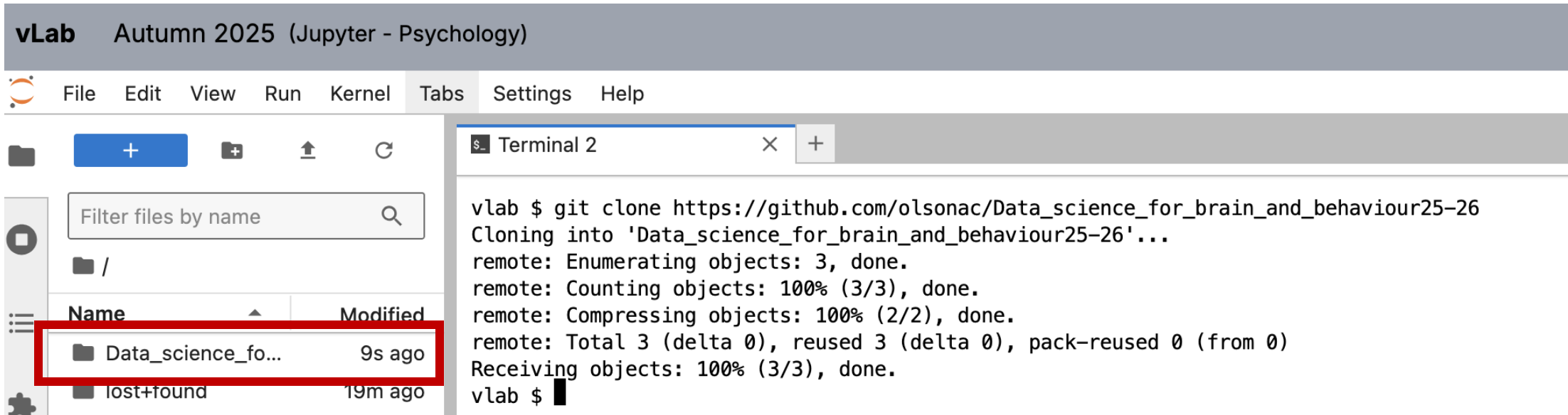
Type the 'git clone' command to connect to the 'repository' where files are stored on the cloud

- Copy the command below and paste into your terminal

```
git clone https://github.com/olsonac/Data_science_for_brain_and_behaviour25-26.git
```



The result should match this:



Notice that you now have a folder for the course. This is where exercises and lecture notes will appear

We only do that setup step once – from now on, each week...

- Start a terminal
- type 'git pull'
- new material will be added to your course folders
- e.g.

```
vlab $ git pull
remote: Enumerating objects: 13, done.
remote: Counting objects: 100% (13/13), done.
remote: Compressing objects: 100% (6/6), done.
remote: Total 12 (delta 4), reused 12 (delta 4), pack-reused 0 (from 0)
Unpacking objects: 100% (12/12), 53.18 KiB | 6.65 MiB/s, done.
From https://github.com/olsonac/Data_science_for_brain_and_behaviour25-26
   ba4ebee..82598af  main    -> origin/main
Updating ba4ebee..82598af
Fast-forward
 Exercises/dracula/.Rhistory                | 0
 Exercises/dracula/.ipynb_checkpoints/dracula-checkpoint.ipynb | 365 ++++++
 Exercises/dracula/.ipynb_checkpoints/dracula_template-checkpoint.Rmd | 269 ++++++
 Exercises/dracula/.ipynb_checkpoints/dracula_template-checkpoint.ipynb | 521 ++++++
 Exercises/dracula/dracula.Rmd              | 258 ++++++
 Exercises/dracula/dracula.ipynb            | 379 ++++++
 Exercises/dracula/play_arrow.png           | Bin 0 -> 13663 bytes
7 files changed, 1792 insertions(+)
create mode 100644 Exercises/dracula/.Rhistory
create mode 100644 Exercises/dracula/.ipynb_checkpoints/dracula-checkpoint.ipynb
create mode 100644 Exercises/dracula/.ipynb_checkpoints/dracula_template-checkpoint.Rmd
create mode 100644 Exercises/dracula/.ipynb_checkpoints/dracula_template-checkpoint.ipynb
create mode 100644 Exercises/dracula/dracula.Rmd
create mode 100644 Exercises/dracula/dracula.ipynb
create mode 100644 Exercises/dracula/play_arrow.png
```

Our tools: Do I need to install python on my own machine? Jupyter notebooks?

- Most of our work is done through vlab.
- Still, you might want to have a local copy of python and jupyter notebooks (e.g. for working offline)
- If you are interested I can help with that – contact me

Jupyter notebook – how does it work?

- VLab is run through a web browser on your computer
- When you run a cell, it doesn't do the computation on your machine.
- When you click on “run” or <cmd><enter>
 - The browser sends the text you have typed in to the vlab machine somewhere in the networld (maybe California). The vlab machine is running your “kernel”.
 - The “kernel” is what executes your commands
 - the kernel does the arithmetic and sends back the answer
 - The answer comes over the web to your browser.
 - Your browser gets the message and shows it to you

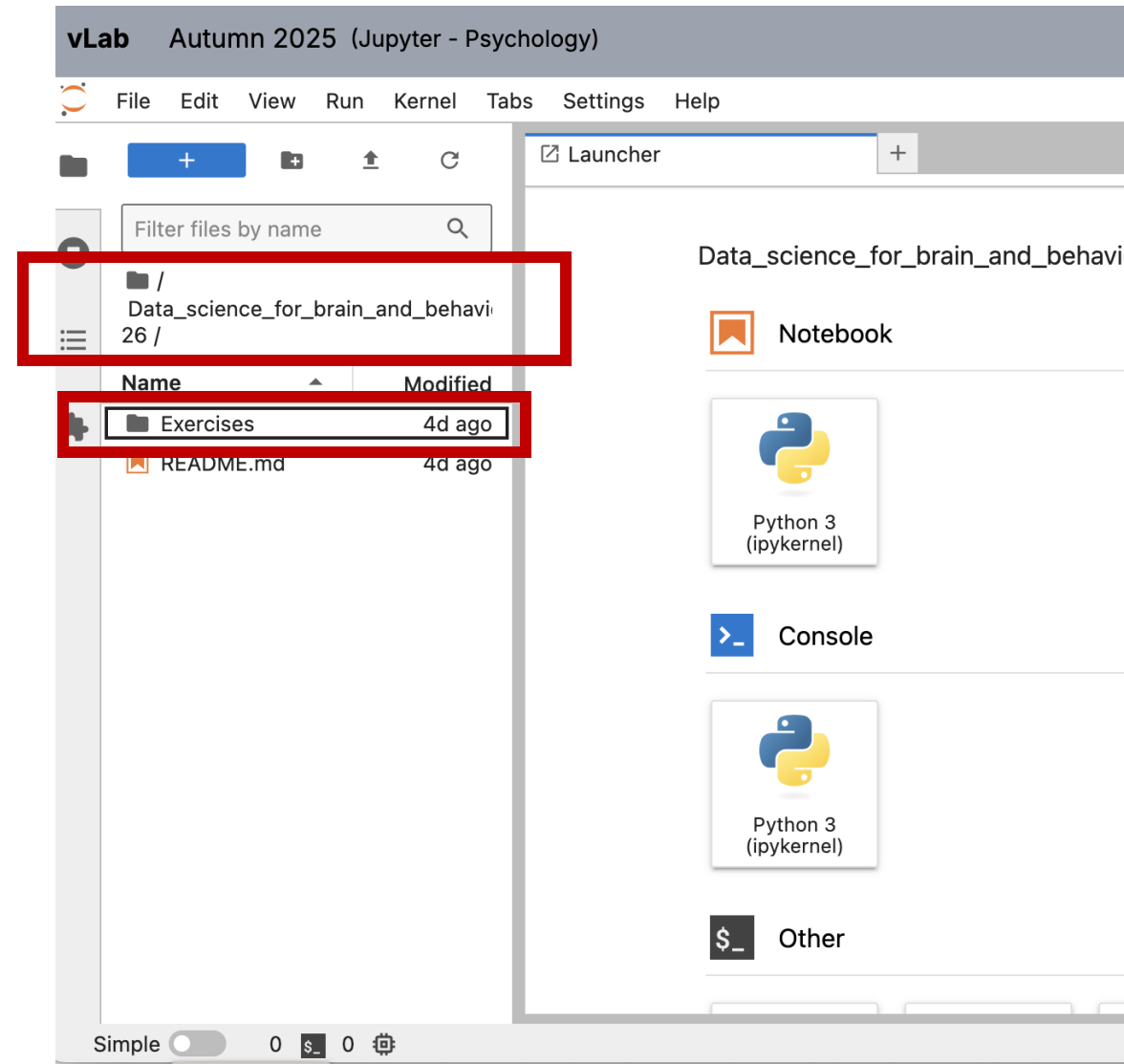


Why do I need to know that?

- Two reasons
- 1) VLab doesn't work without a web connection
 - That may have been obvious already
- 2) There are two processes looking after you.
 - One on your machine (your web browser).
 - One out in the cloud where your kernel is running

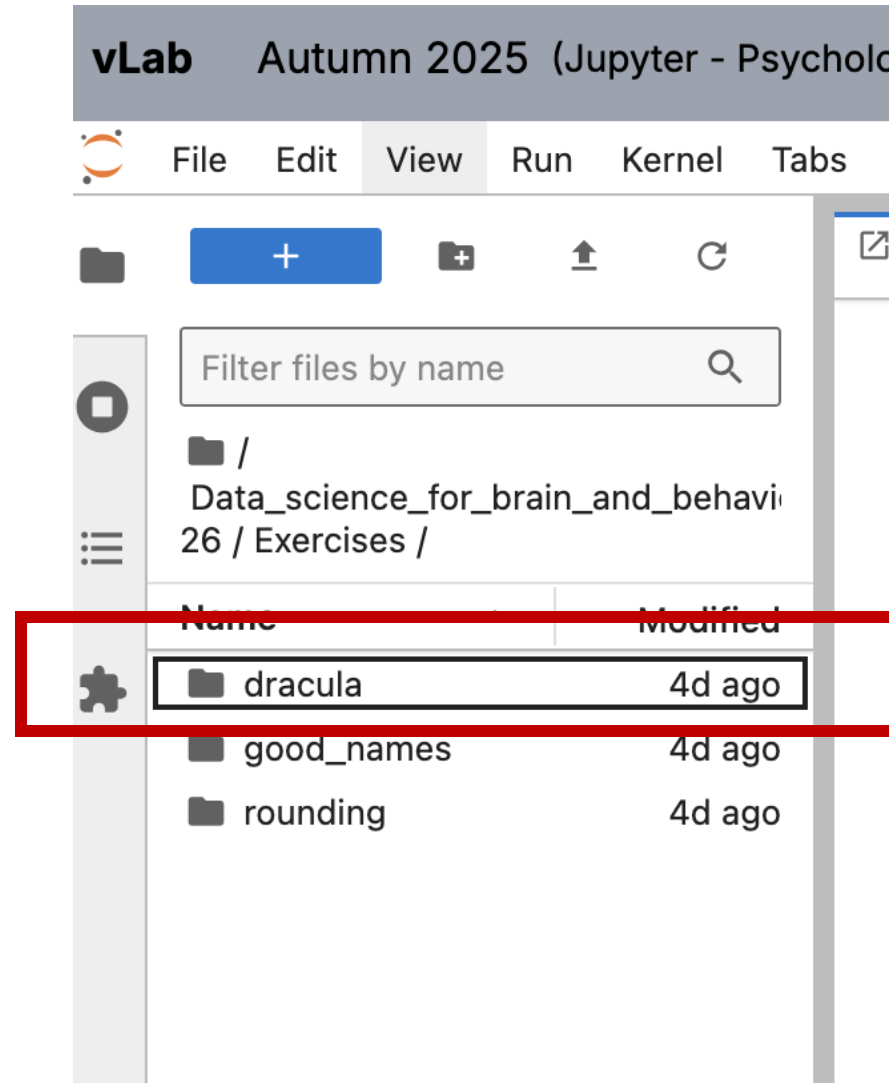
Your first exercise: Dracula

- Click on the Data_science_for_brain_and_behaviour_25-26 folder
- Click on the exercises folder underneath this



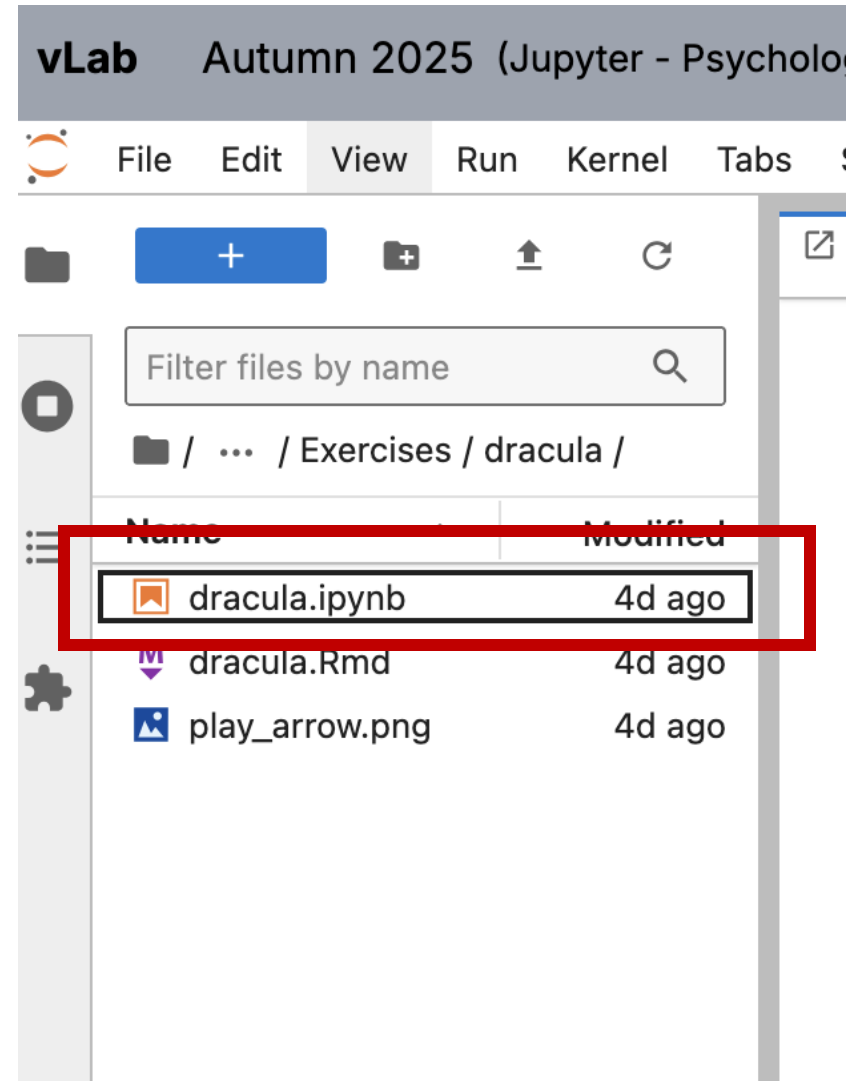
click on the 'Dracula' folder

- Dracula is our first exercise



Double click on the 'dracula.ipynb' file

- .ipynb files are jupyter notebooks
- We will be using this type of file for our work
- Double clicking on the file will open it



You should now see the notebook:

The screenshot displays the JupyterLab interface. At the top, a header bar shows 'vLab Autumn 2025 (Jupyter - Psychology)' on the left, 'Remaining Quota: 146h 46m Session Time: 1h 13m / 2h 0m' in the center, and a red 'Stop Session' button on the right. Below the header is a menu bar with 'File', 'Edit', 'View', 'Run', 'Kernel', 'Tabs', 'Settings', and 'Help'. The left sidebar contains a file browser with a search bar 'Filter files by name' and a list of files in the directory '/ ... / Exercises / dracula /': 'dracula.ipynb' (4d ago), 'dracula.Rmd' (4d ago), and 'play_arrow.png' (4d ago). The main area shows a notebook titled 'dracula.ipynb' with a 'Python 3 (ipykernel)' kernel. The notebook content includes a section header 'Pride, Prejudice and Vampires', followed by a paragraph: 'If you squint your eyes and look sideways, this exercise is a little like the classic [Pride and Prejudice and Zombies](#). OK, it's not as classic as the original book, but still, a classic in its own way.' Below this is another paragraph: 'For this exercise, you are going to adapt the original [Literary Characters](#) notebook so that, instead of analyzing the text of "Pride and Prejudice", it analyzes the text of [Dracula](#) by Bram Stoker.' This is followed by a link: 'You can find the full text of Bram Stoker's Dracula at <http://www.gutenberg.org/ebooks/345.txt.utf-8>.' The next section is 'Don't worry about the details', with the text: 'The exercise asks you to edit the code in the original notebook, but, like the original notebook, we do not expect you to fully understand this code. You will learn many aspects of how this code works over the rest of the course.' The following section is 'The task', with the text: 'Below, you will see the code cells from the original Literary Characters notebook. Your task is to edit these code cells so the'. At the bottom, a status bar shows 'Simple' mode, a cursor at 'Ln 1, Col 1', and the file name 'dracula.ipynb'.

End of vlab setup

- Now you are set up to use vlab to get exercises and notes
- New material will appear each week
- Remember to do 'git pull' before starting your work
- Then click on folders to find the notebook you need
- We will introduce notebooks in a separate step

Piazza – For questions/discussions

- Ask questions!
- There are no silly questions
- sign up to Piazza using this link:
- <https://piazza.com/bham.ac.uk/fall2025/395303953140426/home>

There are no silly questions about DS

- If you have a question, several other people have it too
 - Guaranteed
- Questions help me
 - Sometimes I don't know how people are thinking about an issue
 - “Ooooh, I see how you're thinking about it”
 - That can be very helpful

Signing up for piazza – enter your UoB email address

- Signup page:

The screenshot shows the Piazza website for the University of Birmingham. The header includes the Piazza logo, a link for 'Looking for Piazza Careers', and a 'Log In' button. The main heading is 'University of Birmingham' with a '(change school)' link. A sidebar on the left contains a 'Welcome to Piazza!' message and information about the platform. The main content area shows a 'Selected Term' dropdown set to 'Fall 2025'. Below this, a list of classes is displayed. Class 1 is highlighted in yellow and contains the text 'Class 1: 39530/39531/40426: Data Science for Brain and Behaviour (edit)', 'Instructors: Andrew Olson · 12 Enrolled', and 'Join as: ☒ Student'. A red box is drawn around the 'Join as' section, and a red arrow points from the text 'Select "Student"' to the 'Student' radio button. Below Class 1 are four empty class input fields labeled 'Class 2:', 'Class 3:', 'Class 4:', and 'Class 5:'. At the bottom, there is a link 'Add Another Class' and a blue 'Join Classes' button.

Looking for Piazza Careers Log In

University of Birmingham

(change school)

Welcome to Piazza!

Piazza is an intuitive platform for instructors to efficiently manage class Q&A. Students can post questions and collaborate to edit responses to these questions. Instructors can also answer questions, endorse student answers, and edit or delete any posted content.

Piazza is designed to simulate real class discussion. It aims to get high quality answers to difficult questions, fast!

The name Piazza comes from the Italian word for plaza--a common city square where people can come together to share knowledge and ideas. We strive to recreate that communal atmosphere among students and instructors.

Are you a professor?
Click here to create & join classes

Selected Term: Fall 2025

Fall 2025

Class 1: 39530/39531/40426: Data Science for Brain and Behaviour (edit)
Instructors: Andrew Olson · 12 Enrolled

Join as: ☒ Student *Instructor self-enrollment has been disabled for this class.*

Class 2: ×

Class 3: ×

Class 4: ×

Class 5: ×

[Add Another Class](#)

[Join Classes](#)

Select "Student"

sign up with UoB email (needs to be UoB)

The screenshot shows the Piazza website interface for the University of Birmingham. At the top, there is a blue header with the 'piazza' logo on the left, and two buttons on the right: 'Looking for Piazza Careers' and 'Log In'. Below the header, the page title 'University of Birmingham' is centered, with a '(change school)' link underneath. A 'Selected Term' dropdown menu is set to 'Fall 2025', with a '(go back & edit classes)' link to its right. Below this, the text 'Fall 2025' is displayed. A list of classes follows, with the first item being '1. 39530/39531/40426: Data Science for Brain and Behaviour'. Below the class name, it says 'Instructors: Andrew Olson · 12 Enrolled' and '✓ Joining as Student'. At the bottom of the page, there is a white box with the heading 'Please enter your school email address'. Inside this box, it says 'Please enter the bham.ac.uk email address to which you would like to add your classes.' There are two input fields: 'Email:' and 'Confirm Email:'. A blue 'Submit Email' button is located at the bottom of the box.

piazza

Looking for Piazza Careers Log In

University of Birmingham

(change school)

Selected Term: Fall 2025 (go back & edit classes)

Fall 2025

1. 39530/39531/40426: Data Science for Brain and Behaviour
Instructors: Andrew Olson · 12 Enrolled
✓ Joining as Student

Please enter your school email address

Please enter the **bham.ac.uk** email address to which you would like to add your classes.

Email:

Confirm Email:

Submit Email

Piazza course page

folders with categories of questions

questions listed here

The screenshot shows the Piazza course page for ID 39530/39531/40426. The top navigation bar includes 'Q & A', 'Resources', and 'Statistics'. A blue box highlights the 'Class at a Glance' section, which contains status indicators: 'Needs Attention' (1 unread post), 'All caught up' (No unanswered questions), and 'All caught up' (No unanswered followups). Below this is a grid of statistics: Total Posts (5), Total Contributions (5), Students Enrolled (9), License Status (pending instructor license, 21 days left), Instructor Engagement (0 instructor responses), and Student Participation (0 student responses). A red box on the left highlights the 'All Posts' section, showing a 'Welcome to Piazza!' post from yesterday at 05:23 PM. Another red box at the top highlights the category folders: 'hw1', 'hw2', 'project', 'other', 'week1', and 'week2'. A red arrow points from the text 'folders with categories of questions' to the 'project' folder.

piAZZA 39530/39531/40426 Q & A Resources Statistics Andrew Olson

Students, we've launched the new Piazza user interface! For more information, and How-To resources, [Click Here](#)

+ New Post Search posts...

hw1 hw2 project other week1 week2

All Posts

Yesterday

Welcome to Piazza! 05:23 PM
Piazza is a Q&A platform designed to get you great answers from classmates and instructors fast. We've put together

Class at a Glance

Needs Attention
1 unread posts

All caught up
No unanswered questions

All caught up
No unanswered followups

Total Posts 5

Total Contributions 5

Students Enrolled 9

License Status
pending instructor license (21 days left)

Instructor Engagement
0 instructor responses

Student Participation
0 student responses

Ask a question

PIAZZA 39530/39531/40426 ▾ Q & A Resources Statistics Andrew Olson

Students, we've launched the new Piazza user interface! For more information, and How-To resources, [Click Here](#) x

+ New Post Search posts...

hw1 hw2 project other week1 week2

other x

⋮ □ ← + New Post

Post Type*

☒ **Question**
if you need an answer

☐ **Note**
if you **don't** need an answer

☐ **Poll/In-Class Response**
if you need a vote

Post To* ☒ Entire Class ☐ Instructor(s)

Select Folder(s)* hw1 hw2 project **other** week1 week2

Summary* Enter a one line summary, 100 characters or less

Details ☐ Rich text editor ☐ Plain text editor ☒ Markdown editor

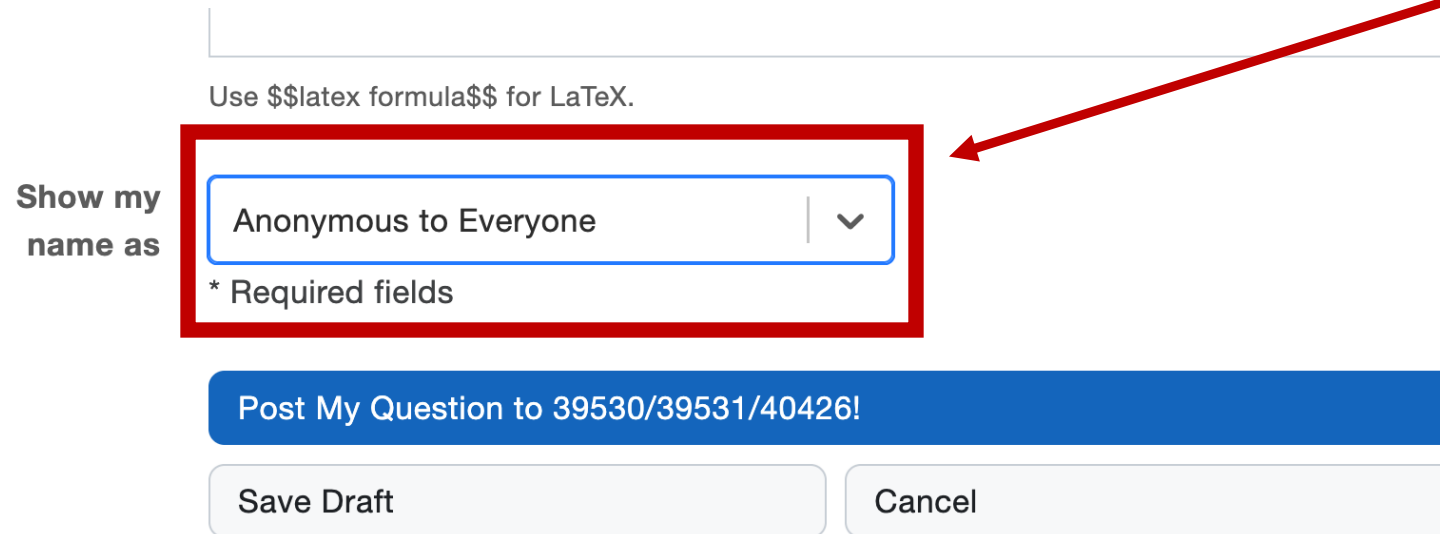
☒ preview

Choose a
folder
that fits
your
question

To make it anonymous

- See the bottom of the post section:

Change this to be anonymous



A screenshot of a web form for creating a post. At the top is a large text input field. Below it is a small text hint: "Use \$\$latex formula\$\$ for LaTeX." Below the hint is a dropdown menu labeled "Show my name as" with the option "Anonymous to Everyone" selected. This dropdown menu is highlighted with a red rectangular box. A red arrow points from the text "Change this to be anonymous" to the dropdown menu. Below the dropdown is the text "* Required fields". At the bottom of the form are three buttons: a blue button labeled "Post My Question to 39530/39531/40426!", a light gray button labeled "Save Draft", and another light gray button labeled "Cancel".

Use \$\$latex formula\$\$ for LaTeX.

Show my name as

Anonymous to Everyone

* Required fields

Post My Question to 39530/39531/40426!

Save Draft

Cancel

Piazza allows you to have code and text blocks.

- This can make your question easier to read

1) Use the rich text editor

2) Click here to create a code block

3) the grey box is code

The white part is text

The screenshot displays the Piazza Q&A interface. On the left, a sidebar shows 'Drafts' with a 'test question' containing a code block and plain text, and 'Yesterday' with a 'Welcome to Piazza!' post. The main area shows a 'New Post' form. A red box highlights the 'Rich text editor' option in the 'Details' section. A red arrow points from the 'Rich text editor' label to the 'test question' draft. Another red arrow points from the 'Rich text editor' label to the 'Rich text editor' option. A third red arrow points from the 'Rich text editor' label to the 'code block' in the 'test question' draft. A fourth red arrow points from the 'Rich text editor' label to the 'code block' in the 'test question' draft. A fifth red arrow points from the 'Rich text editor' label to the 'code block' in the 'test question' draft. A sixth red arrow points from the 'Rich text editor' label to the 'code block' in the 'test question' draft. A seventh red arrow points from the 'Rich text editor' label to the 'code block' in the 'test question' draft. A eighth red arrow points from the 'Rich text editor' label to the 'code block' in the 'test question' draft. A ninth red arrow points from the 'Rich text editor' label to the 'code block' in the 'test question' draft. A tenth red arrow points from the 'Rich text editor' label to the 'code block' in the 'test question' draft.

Piazza 39530/39531/40426 Q & A Resources Statistics Andrew Olson

+ New Post Search posts...

hw1 hw2 project other week1 week2

All Posts

Drafts

test question

```
<pre> # This is a code block my_variable = 10
print(my_variable)</pre> <p>Hi Andrew -</p>
<p>This is plain text - Is this the correct way to...
```

The CANVAS page and link to the online course text

- Link on the CANVAS page – also here:
<https://olsonac.github.io/textbook/intro.html>
- Assessment and feedback information
- Link to the e-textbook
- Link to the piazza discussion board
- How to submit homework (TBA)

2024/25

College of Life and Environmental Sciences | School of Psychology

First tasks: Make sure to join the deepnote site for the course and the piazza discussion board (see links below)

Module Overview and timetable information	Week by Week	Assessment & Feedback	Project information	Reflective statement information (and link to rubric)	EDI information	Discussion board (Piazza)
---	------------------------------	---	-------------------------------------	---	---------------------------------	---

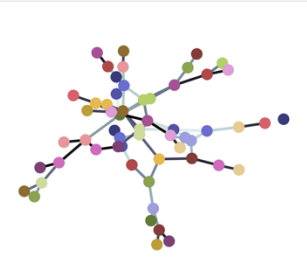
How to download Homework and submit to CANVAS	How to submit your project	How to share your project with me
---	--	---

[Online course text \(click to access\)](#)

Join the deepnote space for the course by clicking the link below

Course text page

- Read the entries in the first section: “Coding for data”



Coding for data

This is the textbook for a course teaching *data science* – coding for data.

We go into some detail about what we mean by “Data science” in the [next section](#), but here is the one-line summary:

Data science is an approach to data analysis with a foundation in code and algorithms.

The textbook aims to teach you this approach.

We designed the textbook so you can learn from it without doing the course. If you read this textbook carefully, and do the exercises, you will have a solid foundation for learning more about data science.

The background you need

You do not need any previous experience in programming to use this book. We aim to teach you the programming you need as we go.

The book deliberately uses very little formal mathematics. Instead, we show how the procedures work with code. We hope you agree that this can be much easier to understand, especially for those without much background in mathematics.

Contents

- The background you need
- A summary of the book
- Based on the Berkeley textbook
- Many sections are interactive

Coding for data

- What is data science?
- Why Data Science?
- Tools and techniques
- Text is data
- Surviving the computer
- Our tools
- Using Jupyter notebooks
- More on the Jupyter notebook

On code

- Ode to code
- A sampling problem

Student organized Whatsapp group

- Students have organized a Whatsapp group in the past
- If you want to do that I will distribute the link
 - Send me an email and I will put it on the CANVAS page
- I will not participate in the whatsapp group
 - no lurking

Surviving the computer

- Good advice

Have a piece of paper and a pen next to you

When *working* on the computer, have a piece of paper and a pen next to you on your desk. When you notice that you don't fully understand an error or a task on the computer, stop, move the laptop out of the way, and write down your problem on a piece of paper. Give yourself some time to reflect on what the problem is, and how to solve it. Only then, you can return to the computer and try your solutions, or do more research. You will find that disengaging from the computer is a) difficult and b) very productive in releasing you from the various mental traps that it is easy to fall into when you get stuck on the computer.

- Take a break. Get away from your machine. Think about something else. Then come back
- Writing things down can also help you to think in a different way (e.g. diagrams)

The first exercise: Dracula

Plotting the classics

- You don't have to understand the code (yet)
- Checks that you can open a notebook and run the sections
- Has some examples of what you can do...
 - Get text from the internet
 - break it up into chunks (words, chapters)
 - Count how often the principle characters are mentioned
 - Make a summary table by chapter (of cumulative numbers - important for interpretation [not numbers for each chapter])
 - Plot the summary
 - See what you can learn from the summary plot
 - e.g. who is the principle character
 - who is the "love interest"
 - when are secondard characters important

Summary

- We've looked at VLab
- We've seen the canvas page
- We've seen how to access the course text
- We've seen what jupyter notebooks look like and how to open them
- We've seen piazza – Ask questions!
- Dracula – Our first exercise
 - Plotting the classics (you don't have to understand the code parts – we don't expect that yet)
- Just see what is possible, getting text from the internet, breaking it up, plotting the results
- See what you can understand about a book from even a few lines of python code
- Practice understanding what a graph is showing you (first identify the parts clearly)