

Introduction to *Urban* Data Science

Introduction

(EPA1316A)

Lecture 1

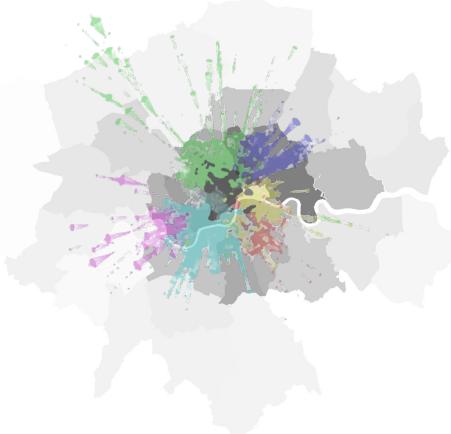
Trivik Verma



Adapted from the work of Sean Perez



Work &
Education



Data
+
Cities
+
Justice



Startup

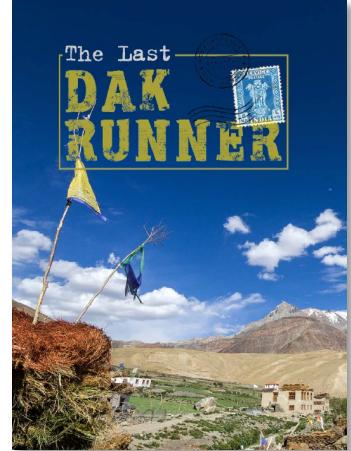


Photo +
Teaching





Centre for Urban Science & Policy

We are a transdisciplinary research group working to advance urban research, planning and policy in a way that strives for just and equitable outcomes for communities. We use a mix of computational spatial science and qualitative participatory methods to investigate how social, economic, environmental and political processes shape cities. Our goal is to develop a body of computational approaches and curate evidence that facilitates an integrated systems-based approach for urban planning.

[EXPLORE OUR WORK](#)

Space and Place

Cities & Social Justice

Understand the complex nature of urban spaces in transformation

Understand how inequalities intersect with space

Propose guidelines and develop tools to support public participation and citizen action

Identify inequalities associated with lack of recognition and legitimacy

Intersectionality

Democratisation of urban design, planning & policy

Teaching Support



Auriane Técourt
2nd year EPA



Ludovica Bindi
2nd year EPA



Dorukhan Yeşilli
2nd year EPA



Nicolò Canal
2nd year EPA



Vaibhavi Srivastava
2nd year EPA



Ruth Nelson
PhD @ CUSP lab

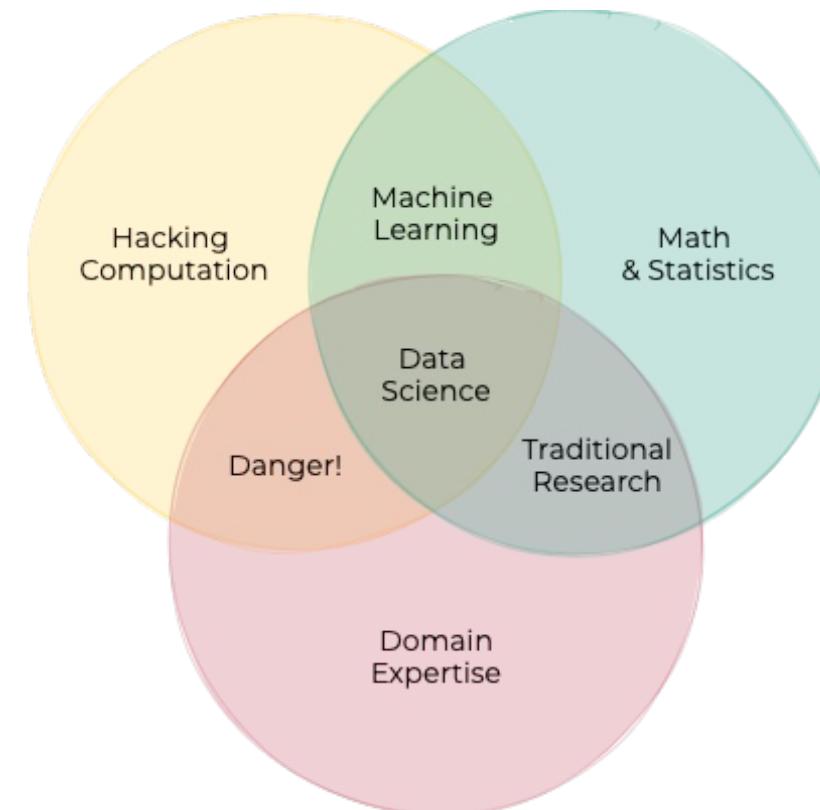
Today

- Introduction to the Course
- Tools - Python and Conda
- Post break, Intro part II

Introduction to the Course

More stats than a GIS course... more GIS than a stats course

With a few twists!



After this course

You will be able to...

- **Obtain**: Obtaining data from multiple **open** data sources.
- **Scrub**: Data cleaning, munging, sampling to consolidate all information into a dataset that is manageable, informative and relates to your problem.
- **Explore**: Exploratory data analysis to make sense of what your data is trying to say.
- **Model**: Estimation and modelling based on statistical tools such as regression and clustering.
- **Interpret**: Communicating results and reflections through visualisation, storytelling and interpretable summaries.

DATA



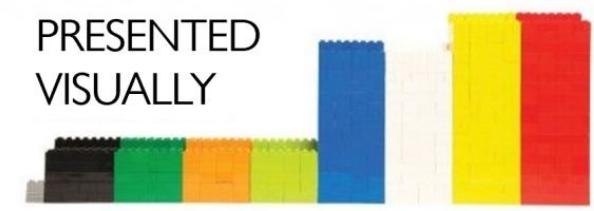
SORTED



ARRANGED



PRESENTED VISUALLY



EXPLAINED WITH A STORY



Can you find the source for me?

Philosophy

- (Lots of) **methods** and techniques
 - General overview
 - Intuition
 - Very little math
 - Lots of ways to continue learning more
- Emphasis on the **application** and **use**
- Close connection to “**real world**” applications

Format – Ways of Working

Seven weeks of:

- **Prep. Materials:** videos, podcasts, articles... 1h. approx. (most recommended!)
- **2x 1h. Lecture:** interaction, concepts, methods, examples
- **2h. Computer labs:** hands-on, application of concepts, Python (*highly employable*)
- **1h. Paper Discussions:** reading a paper/article and deliberating on a set of questions with your peers in small groups. (extremely important if you are interested in applying concepts to real-world problems).
- **Extra material:** how to go beyond the minimum*

*(not necessary for the course but useful in life)

Content

- **Weeks 1-4:** “big picture” lectures + introduction to computational tools (learning curve) + lots and lots of data + lots of visualisation
- **Weeks 5-7:** lots of spatial, network and machine learning concepts
- **Weeks 8-10:** wrap up + prepare an awesome final project in groups (opportunities to follow up with internships in the ***CUSP*** lab)

An overview of the course

- **Weeks 1-4:** “big picture” lectures + introduction to computational tools (learning curve) + lots and lots of data + lots of visualisation
- **Weeks 5-7:** lots of spatial, network and machine learning concepts
- **Weeks 8-10:** wrap up + prepare an awesome final project in groups (opportunities to follow up with internships in the ***CUSP*** lab)

Week	Lecture + Discussion	Topic	Learning Goals	Python Libraries	Labs and Homework	Assessment
W1	L1	Introduction to Urban Data Science		Anaconda and Jupyter, Numpy	Lab 0 + 1	
	L2	Spatial and Urban Data				
W2	L3 + D1	Data Grammar	Obtain, Discuss	Pandas, Seaborn	Lab 2	
	L4	Data Engineering	Manipulate, consolidate	Pandas		Assignment 1
W3	Ommen Trip					
W4	L5 + D2	EDA and Visualisation	Discuss, manipulate and Consolidate	Geopandas, Matplotlib, Rasterio	Lab 3	
	L6	Geo-Visualisation	Interpret			
W5	L7 + D3	Networks and Spatial Weights	Describe, Analyse	Networkx, Osmnx, Pysal	Lab 4	
	L8	Exploratory Spatial Data Analysis	Describe, Analyze			Assignment 2
W6	L9 + D4	Machine Learning for Everyone	Apply	Sklearn, Scipy, Statsmodels	Lab 5	
	L10	Anatomy of a Learning Algorithm	Infer			
W7	L11 + D5	Clustering	Apply	Pysal, Sklearn-Cluster	Lab 6	
	L12	Dimensionality Reduction	Apply			Assignment 3
W8	L13 + D6	Spatial Density Estimation	Infer	More Sklearn	Lab 7	
	L14	Responsible Data Science	Create			
W9	Project Preparation					
W10						Final Project

Schedule

	Week 1	2	3	4	5	6	7	8	9	10
	L1: Introduction	L3: Data Grammar		L5: EDA and Visualisation	L7: Networks and Spatial Weights	L9: Machine Learning for Everyone	L11: Clustering	L13: Spatial Density Estimation		
	L2: Spatial and Urban Data	L4: Data Engineering		L6: Geo-Visualisation	L8: Exploratory Spatial Data Analysis	L10: Anatomy of a Learning Algorithm	L12: Dimensionality Reduction	L14: Responsible Data Science		
	5-9 Sep	12-16 Sep	19-23 Sep	26-30 Sep	3-7 Oct	10-14 Oct	17-21 Oct	24-28 Oct	31 Oct - 4 Nov	7-11 Nov
Mo	Assignment 1 Release		Final project release	Assignment 2 Release		Assignment 3 Release				
	At: 8:00		At: 8:00	At: 8:00		At: 8:00				
	Lecture 1	Lecture 3	Ommen	Lecture 5	Lecture 7	Lecture 9	Lecture 11	Lecture 13		
	13:15 - 15:00	11:15 - 12:30		13:15 - 14:30	13:15 - 14:30	13:15 - 14:30	11:15 - 12:30	13:15 - 14:30		
	Trivik	Trivik		Trivik	Trivik	Trivik	Trivik	Trivik		
	Lab 0	Paper Discussion		Paper Discussion	Paper Discussion	Paper Discussion	Paper Discussion	Paper Discussion		
	15:15 - 17:00	12:45 - 14:30		14:45 - 16:30	14:45 - 16:30	14:45 - 16:30	12:45 - 14:30	14:45 - 16:30		
	TAs	In-person groups with TAs		In-person groups with TAs	In-person groups with TAs	In-person groups with TAs	In-person groups with TAs	In-person groups with TAs		
		Formative Feedback		Formative Feedback	Formative Feedback	Formative Feedback	Formative Feedback	Formative Feedback		
		14:30 - 14:45		16:30 - 16:45	16:30 - 16:45	16:30 - 16:45	14:30 - 14:45	16:30 - 16:45		
Tu										
We	Lecture 2	Lecture 4	Ommen	Lecture 6	Lecture 8	Lecture 10	Lecture 12	Lecture 14		
	9:15 - 10:30	9:15 - 10:30		9:15 - 10:30	13:15 - 14:30	9:15 - 10:30	9:15 - 10:30	9:15 - 10:30		
	Trivik	Trivik		Trivik	Trivik	Trivik	Trivik	Trivik		
	Lab 1	Lab 2		Lab 3	Lab 4	Lab 5	Lab 6	Lab 7		
	10:45 - 11:45	10:45 - 11:45		10:45 - 11:45	14:45 - 16:30	10:45 - 11:45	10:45 - 11:45	10:45 - 11:45		
	12:00 - 13:00	12:00 - 13:00		12:00 - 13:00	TAs	12:00 - 13:00	12:00 - 13:00	12:00 - 13:00		
	TAs	TAs		TAs	TAs	TAs	TAs	TAs		
					Assignment 1 summative feedback		Assignment 2 summative feedback		Assignment 3 summative feedback	Final report submission
					Released by: 18:00		Released by: 18:00		Released by: 18:00	Deadline: 23:30 TBD
Th										
Fr	Assignment 1			Group formation	Assignment 2	Selection of final project	Assignment 3	Problem formulation and Data checks		
	Deadline: 23:30			Deadline: 23:30	Deadline: 23:30	Consult and check in with TAs	Deadline: 23:30	Consult and check in with TAs		

Logistics - Website

Course Material

- <https://cusp.tbm.tudelft.nl/courses/epa1316>

Lectures, Labs and Discussions

- Physical Space: Check Timetable every week

Communication Channels

- With instructors and TA: Email
- Announcements on Brightspace
- Discussion Forums on Brightspace

Submissions, Groups, Grades and Feedback

- Brightspace

The screenshot shows the homepage of the CUSP website for the course EPA 1316 A. The top navigation bar includes links for About, Research, Teaching, People, Opportunities, and Philosophy, along with the CUSP logo and "Centre for Urban Science & Policy". The main content area has a sidebar on the left with a search bar and links for Courses (EPA 1316 A), Lectures, Software, Resources, and FAQ. The main content area features a section titled "Introduction" with an "IMPORTANT" note stating that the course is freely available online and hybrid meetings are not held. It also welcomes users to Introduction to Urban Data Science (EPA 1316 A) at Delft University of Technology, taught by Dr Trivik Verma. The "Instructor Details" section provides contact information for Dr Trivik Verma, including his title as Assistant Professor in Urban Science & Policy, office location B2.390, Building 31, Faculty of Technology, Policy and Management, Jaffalaan 5, 2628 BX Delft, The Netherlands, and email tverma [at] tudelft.nl. The "Schedule" section links to a detailed schedule, and the "Locations" section states that all lectures, labs, discussions, and office hours will be hosted in person.

Centre for Urban
Science & Policy **CUSP** About Research Teaching People Opportunities Philosophy

Search... /

← Courses
EPA 1316 A
Introduction
To Begin With
Overview
Syllabus
Schedule
Lectures
Labs
Discussions
Assessment
Assignments
Software
1. Standard Installation
2. Minimalist Installation
3. Comprehensive Installation
Virtual Environments
Resources
FAQ

Introduction

IMPORTANT:

The course is freely available on this website. There are no online or hybrid meetings.
If you just want all of the course materials and not participate in the course any longer,
please unenroll from Brightspace. Thank you.

Welcome to Introduction to Urban Data Science (EPA 1316 A) at Delft University of Technology.
The course is taught by Dr Trivik Verma with the support of a fantastic team of teaching
assistants.

Instructor Details

Dr Trivik Verma
Assistant Professor in Urban Science & Policy
B2.390, Building 31
Faculty of Technology, Policy and Management
Jaffalaan 5
2628 BX Delft
The Netherlands
E: tverma [at] tudelft.nl

Schedule

A detailed schedule of the course is provided [here](#).

Locations

All lectures, labs, discussions, and office hours will be hosted in person.

Self-directed learning

This course is much more about “**learning to learn**” and problem solving rather than acquiring specific programming tricks or stats wizardry

- **Prepare** for the labs
- **I won't** be leading/lecturing at the computer labs. TAs will be present for abundant help and feedback.
- **Go over the notebooks** before the lecture and the computer lab
- If the first time you see a notebook is at the lab, you may find it difficult to follow on. The best thing to do is to prepare a set of questions to ask us.
- **Bring** questions, comments, feedback, (informed) rants to class/labs. The more you bring, the more we all learn.
- **Collaborate** (it's **NOT** a zero-sum win!)

Python

- **General purpose** programming language
- “Sweet spot” between “*proof-of-concept*” and “*production-ready*”
- Industry standard: **GIS** (Esri, QGIS) and **Data Science** (510, World Bank, OECD, The Atlantic, Gemeente Den Haag...)

How many of you have written a line of computer code before?

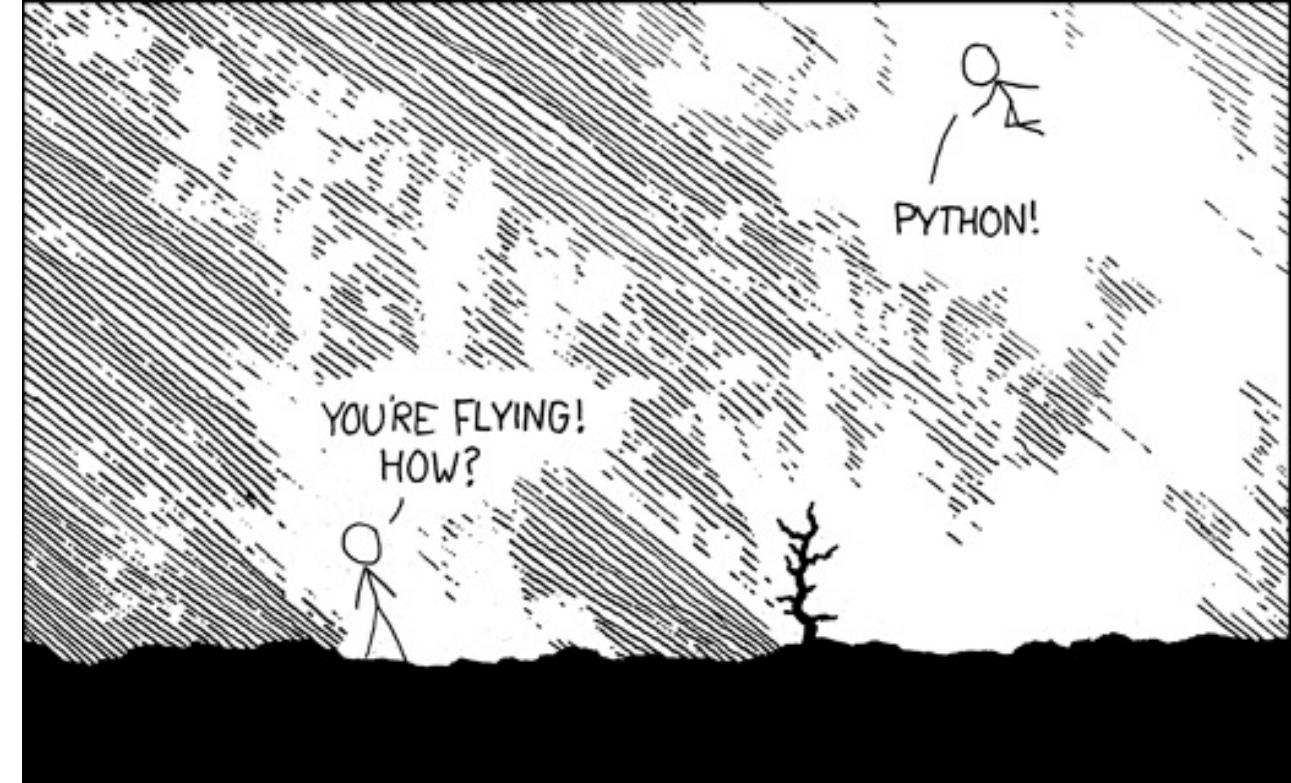


Figure under Creative Commons Attribution Non-Commercial 2.5 License



Assessments

Formative : These are ungraded

- 7 In-Class Labs
- 7 Homework Labs

Discuss the completed homework with your peers using a list of Do's and Don'ts to evaluate each other's work

Assessments

Summative : There are 4 graded components that contribute to the final mark for the course as follows:

50% Individual assessment include

- Assignment 1: Data Collection and Wrangling (15%)
- Assignment 2: Geographic Visualisation (15%):
- Assignment 3: Prediction/Inference (20%)

50% Group assessment include

- Final Project

Rubric for Assignments

- Assignments are graded based on **four criteria**
- The criteria have different weights that add up to a **total of 10**
- The score for each criterion range from **0 to 2 points**
- Only English

Criteria	Indicative weight	Needs Improvement (0.5-0.9)	Meets Some Expectations (1.0-1.9)	Exceeds Expectation (2)
Output	0.5			
Formatting	0.5			
Methods (e.g. Tidy Data, EDA, Graphical Excellence, Spatial Autocorrelation, Network Weights, Regression, etc.)	2			
Documentation (Markdown/Comments)	2			

Rubric for Assignments

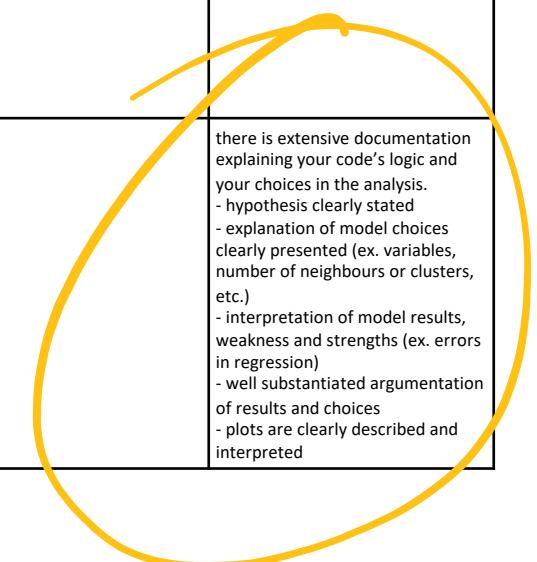
Output

Formatting

Methods

Documentation

Assignments are about
Methods and Documentation,
not your ability to code

Criteria	Indicative weight	Needs Improvement (0.5-0.9)	Meets Some Expectations (1.0-1.9)	Exceeds Expectation (2)
Output	0.5			
Formatting	0.5			
Methods (e.g. Tidy Data, EDA, Graphical Excellence, Spatial Autocorrelation, Network Weights, Regression, etc.) 	2			
Documentation (Markdown/Comments) 	2			<p>there is extensive documentation explaining your code's logic and your choices in the analysis.</p> <ul style="list-style-type: none">- hypothesis clearly stated- explanation of model choices clearly presented (ex. variables, number of neighbours or clusters, etc.)- interpretation of model results, weakness and strengths (ex. errors in regression)- well substantiated argumentation of results and choices- plots are clearly described and interpreted 

Do's

- Finish the corresponding labs before starting an assignment
- Think about the objective of the assignment: data exploration vs data analysis
- Think about the method and its limitations
- Make sure that your code runs and produces the expected output
- Use clear and interpretable variable names: '**SP.DYN.TFRT.IN**' vs '**average_fertility_rate**'
- Think about the data quality: are there missing values? How will you deal with them? What are the implications?
- Use headers to structure your notebook
- Use markdown to explain the code and interpret the findings
- Please do not use any language other than **Python and English**

Don'ts

- Don't copy and paste the code from the labs without understanding what the added value is
- Don't print huge matrices or use enormous font. Make sure the notebook can be read as a report
- Don't print empty cells, import useless packages or import packages multiple times
- Don't print figures without labels or titles
- Don't use colours that are difficult for colour-blind people to distinguish
- Don't hardcode, try instead to use functions to make the code easily reusable later
- Don't overwork, focus on what is asked from you in the assignment

Rubric for Final Project

- Projects are graded based on seven criteria
- The criteria have different weights and the score for each criterion range from 1 to 10 points in increments of .5

Criteria (increments of .5)	Indicative %	(1-4) Unacceptable	(4.5-5.5) Insufficient	(6-6.5) Minimally acceptable	(7-8.5) Meets Expectations	(9-10) Exceeds Expectations
		Defines work that is generally incomplete or substandard	An incomplete attempt to address the task.	An adequate accomplishment of the task.	An above average accomplishment of the task.	An exemplary accomplishment of the task.
Formatting and Legibility (use Grammarly.com for grammar and spelling if unsure)	5%					
Problem Structuring	15%					
Related Work	5%					
Data Collection and Processing	20%					
Methodology (reproducibility means that I understand what you did and how you did it and can repeat the analysis if I had access to the same data)	20%					
Results and Visualisation	20%					
Discussion/Conclusion	15%					

Support!!!

 Search... [← Courses](#)[EPA 1316 A](#) [Introduction](#)[To Begin With](#)[Overview](#)[Syllabus](#)[Schedule](#) [Lectures](#) [Labs](#) [Discussions](#) [Assessment](#)[Assignments](#) [Software](#)[1. Standard Installation](#)[2. Minimalist Installation](#)[3. Comprehensive Installation](#)[Virtual Environments](#) [Resources](#) [FAQ](#)

Helping Material

Our teaching support team has prepared and updated two wonderful resources for the students.

There is one resource on programmin support. You can find links to library documentation, data science practices, analysis and much more. Another resources is on debugging. Everytime your program fails to do what you expected of it, go to this resource first.

Programming Help Sheet

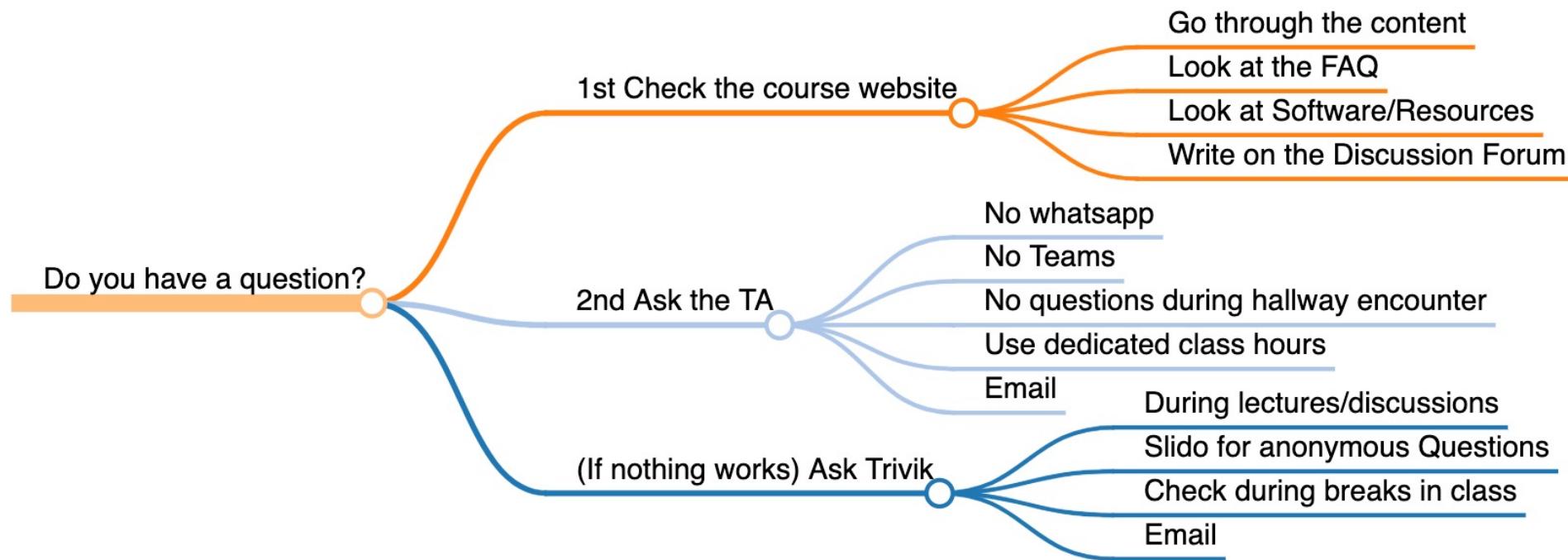
Analysis: explanations and tutorial	<ul style="list-style-type: none">◦ https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e09c (with links to other PCA content)• PCA in Python using scikit-learn (sklearn) with a good step-by-step guide◦ https://www.datacamp.com/tutorial/principal-component-analysis-pca-in-python
Interactive visualizations (extra stuff, not in the course)	<ul style="list-style-type: none">• For the usual charts:<ul style="list-style-type: none">◦ https://realpython.com/python-data-visualization-basics-introduction/◦ https://www.geeksforgeeks.org/using-plotly-for-interactive-data-visualizations/• For maps:<ul style="list-style-type: none">◦ https://www.mapbox.com/
And everything else	<ul style="list-style-type: none">• General help websites:<ul style="list-style-type: none">◦ https://stackoverflow.com/• Website to check difference between texts (e.g., use it to check if two files are different)<ul style="list-style-type: none">◦ https://www.diffchecker.com/• General stuff about data science:<ul style="list-style-type: none">◦ https://towardsdatascience.com/◦ https://medium.com/• Interesting notebooks about anything:

More Support!!!

- Ask questions
- Help others as much as you can (the best way to learn is to share perspectives)
- Search heavily on Google + Stack Overflow (learn to troubleshoot)
- Bring questions, comments, feedback, (informed) rants to class
- Collaborate with each other
- Use the Discussion Forum **heavily**

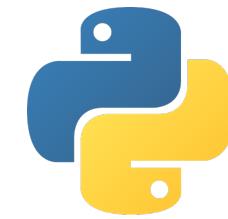


General Support Structure!!!



ex. you send me an email that your code just wouldn't run.

Installing Python



Search... /[← Courses](#)[EPA 1316 A](#)[!\[\]\(1d44e689db7887f5f7d7a4ea2fb82e45_img.jpg\) Introduction](#)[To Begin With](#)[Overview](#)[Syllabus](#)[Schedule](#)[!\[\]\(e7b151aec1eb227a716fbba8e2a59b0e_img.jpg\) Lectures](#)[!\[\]\(d1714a7609c81149f0a84903636f0f81_img.jpg\) Labs](#)[!\[\]\(847172ace9f417f0ef2d71cc34021152_img.jpg\) Discussions](#)[!\[\]\(1fe35e27b64f2cbe9d91692407739e94_img.jpg\) Assessment](#)[Assignments](#)[!\[\]\(d0348104d649dce8ac4d39dfb5bdc990_img.jpg\) Software](#)[1. Standard Installation](#)[2. Minimalist Installation](#)[3. Comprehensive Installation](#)[Virtual Environments](#)[!\[\]\(7ac9149c311be44e3668c36226e60d19_img.jpg\) Resources](#)[!\[\]\(4fc72ea608e9fd40fa4cc9b4cf961884_img.jpg\) FAQ](#)

Software

This course is best followed if you can reproduce the examples and tutorials provided with it. To do so, you will need to install in your machine a series of software packages. These are all open-source and available for free to download.

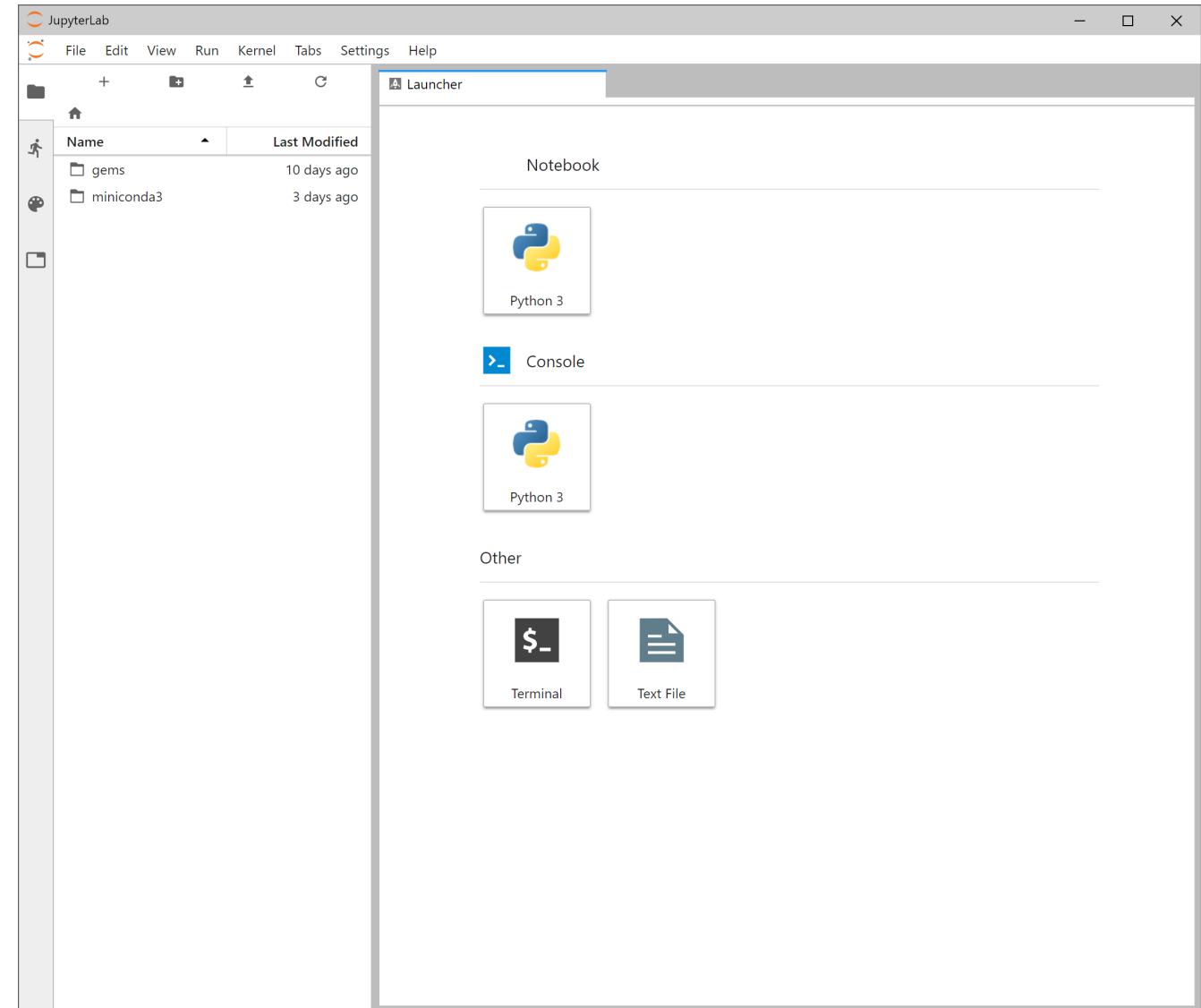
There are three main pathways to install required Python libraries on your machine.

- A **(1) standard** one is installing the software on your operating system without using the command line interface.
- A **(2) minimalist** one will provide basic Python resources and the ability to expand them.
- A **(3) comprehensive** one will install not only a Python stack but also several useful libraries (including some from the programming language, R).

If you want to learn to explore Python and its capabilities, while going beyond this course, I **recommend option 2**. If all else fails, option 3 is the last resort for this course. It is guaranteed to work and very powerful, so you will not be limited in any way. But it does not allow you to install new libraries, which means you are limited by what it offers.

 The difference in these options can be explained through the illustration of a living place. If you own a house, you might be able to expand it, paint the walls, add new furniture, even keep a dog. This is akin to the **minimalist approach** which gives you everything you need and the freedom to build upon it. Instead, if you rent a house, in most cases you will not be allowed to make any changes. A **comprehensive approach** gives you everything too, but no freedom to experiment with new python libraries. The **standard option** is like visiting a hotel where others service you for a bit without you having to do the heavy lifting. You choose what works for you!

Installing Python



You?

Join at
slido.com
#3960 125





Meet your peers

Turn to your neighbors and chat for three minutes about,

- Name
- What did you study before coming here?
- What do you expect from this course?
- How do you see this course helping you with your future ambitions?

Community

EPA + Others

- We are from different parts of the world
- We have different educational backgrounds
- We have different experiences in life



Adapted from the work of Sean Perez

Discussions

Only by listening to others can I become aware of the conceptual shackles imposed by my own identity and experiences.

- David Takacs



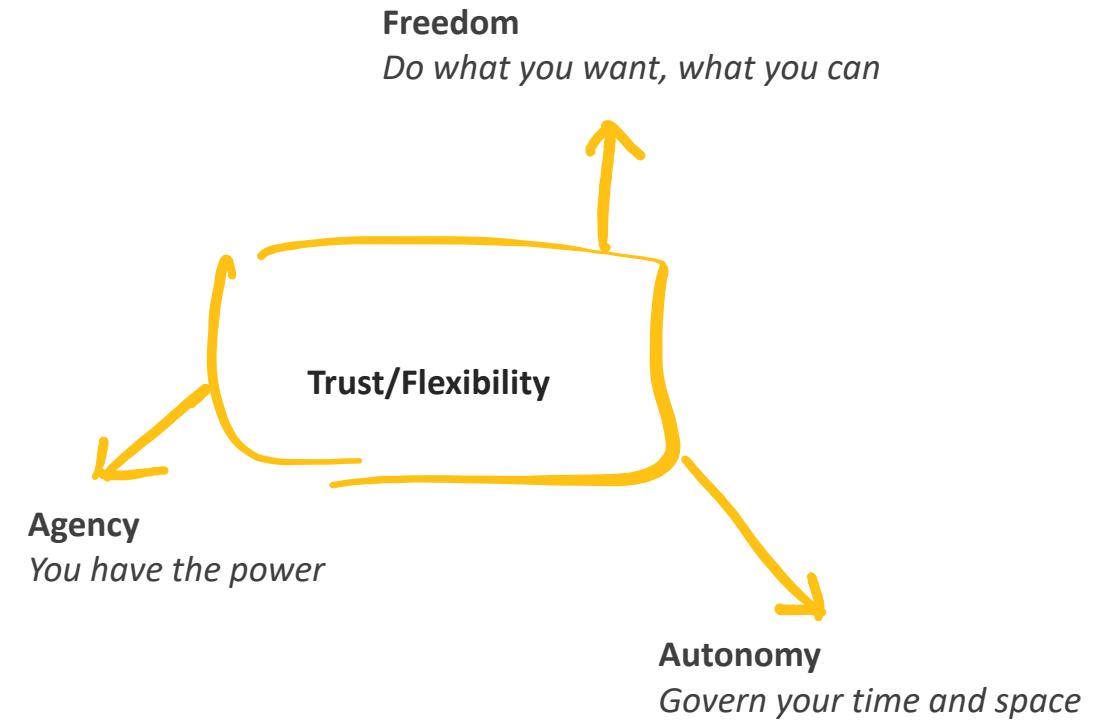
*Experience
Identity
Theories and Evidence*

Flexibility in learning



Flexibility

in everything



You do what fits your lives!

Flexibility

- You do not owe me “productivity” or “efficiency”. I just want your participation, so we can all learn something new.
- You do not owe me any information concerning your personal situation or mental or physical health condition.
- If you want to, you are welcome to talk to me about anything you are going through. **Just drop me an email**, and we will figure it out from there.
- This course is just one small part of life. If you have to work around it to figure life out, go ahead and do it. I trust you will reach out to me if you need support, and I will be here to offer it.
- Exams only focus on your ability to regurgitate knowledge in a 3-hour window. There is no exam in this course. The final project will provide you an opportunity to learn from each other and create something awesome.

Deadline Philosophy

Deadlines are there for a reason!

But if you need more time because... well life, you need to send me an **email** explaining **why** you need more time, and send me that email in good time.

DEADLINES!



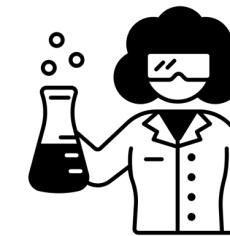
Break



CHILL



WALK



COFFEE OR TEA



MAKE FRIENDS