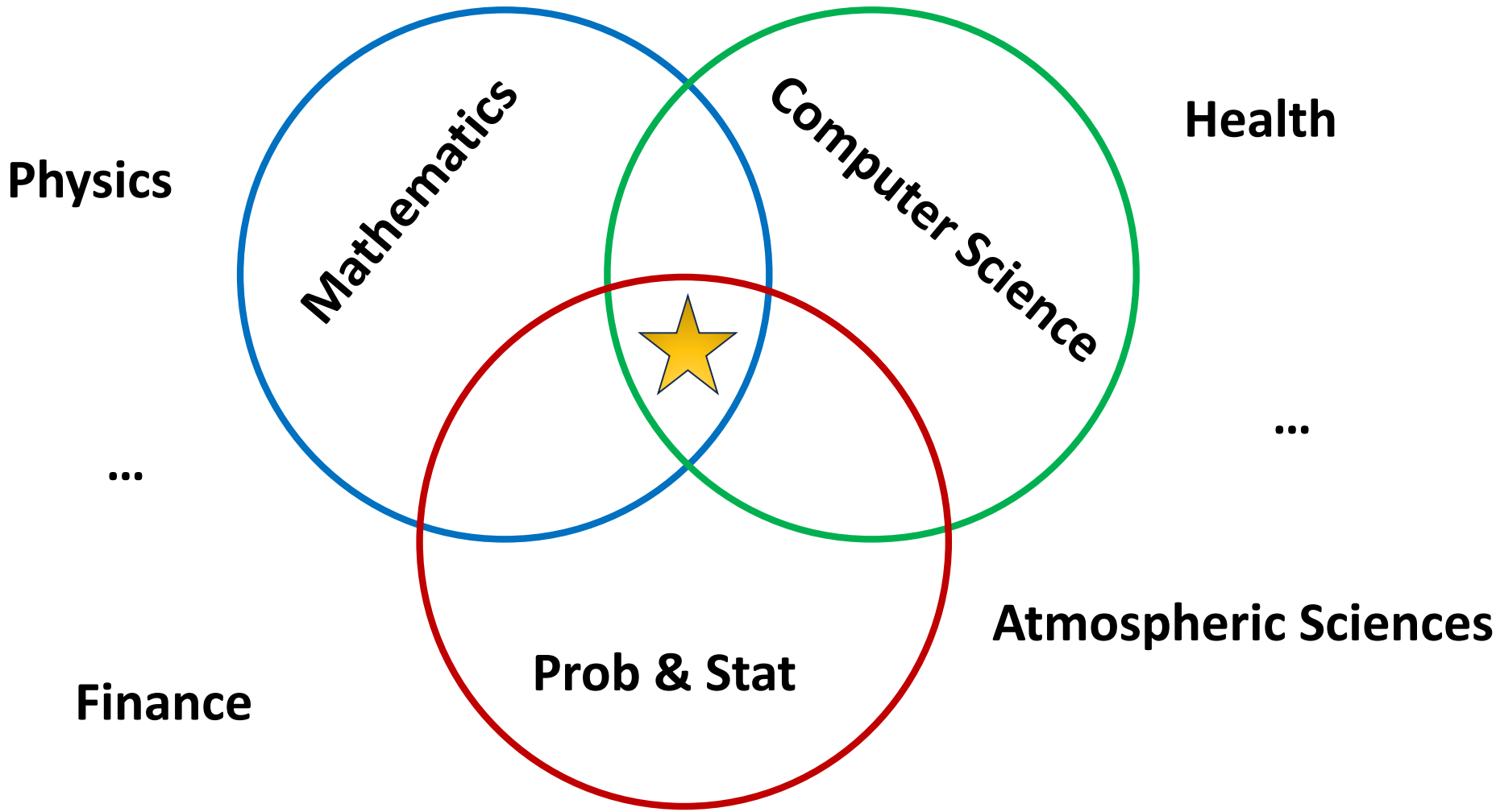
Setting up an Integrated Development Environments (IDE), Virtual Environment, Overleaf, LaTeX, and GitHub

DSCI 222: Overview

What skills do I need to be an excellent data scientist?

Is Data Science in High Demand?

The Bureau of Labor Statistics projects 35.8% employment growth for data scientists between 2021 and 2031. In that period, an estimated 40,500 jobs should open up.


Physics
Mathematics
Computer Science
Health
...
...
Finance
Prob & Stat
Atmospheric Sciences


Soft Skills
Communication Skills
Presentation Skills
Job Interview Skills
Team Skills
Time Management Skills
Leadership Skills
Customer Service Skills

To keep in mind…

Although our main focus is towards data manipulation, analysis and visualization through Python, during this course It is expected to further develop all other necessary skills to be an excellent data scientist.

In-person **Recordings** **Projects** **Assignments**

Presentation **Guests** **Passion**

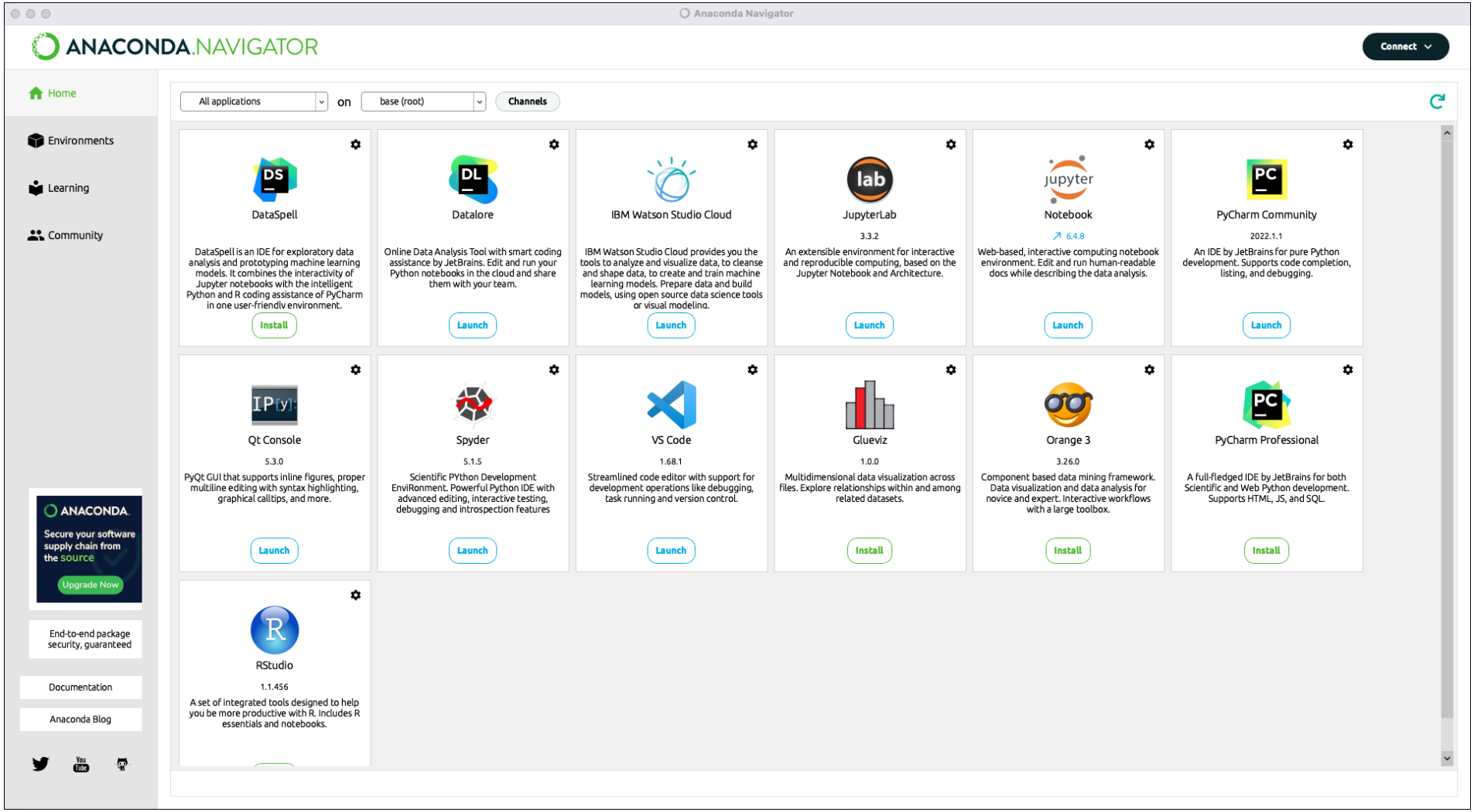
Tools to succeed

https://forms.gle/ERuDtSETousxazyk7

- Anaconda.Navigate
- Visual Studio Code
- PyCharm
- Conda
- Google Colab
- Jupyter Notebooks
- LateX
- Overleaf
- Github
- AI Assistants

Anaconda

VS Code and PyCharm	• Python IDE • Multiple applications and IDEs can be run through Anaconda.Navigator
Environment manager	• Conda (alternative to pip) • Easy Package Management • Many pre-installed packages generally used in machine learning and data science
Jupyter Notebook (Google Colab alternative)	• Combination of Julia, Python, and R, highlighting its support for multiple languages through the use of different kernels


Anaconda Navigator
ANACONDA.NAVIGATOR
Connect
Home
Environments
Learning
Community
All applications
on
base (root)
Channels
DataSpell
DataSpell is an IDE for exploratory data analysis and prototyping machine learning models. It combines the interactivity of Jupyter notebooks with the intelligent Python and R coding assistance of PyCharm in one user-friendly environment.
Install
Datalore
Online Data Analysis Tool with smart coding assistance by JetBrains. Edit and run your Python notebooks in the cloud and share them with your team.
Launch
IBM Watson Studio Cloud
IBM Watson Studio Cloud provides you the tools to analyze and visualize data, to cleanse and shape data, to create and train machine learning models. Prepare data and build models, using open source data science tools or visual modeling.
Launch
JupyterLab
3.3.2
An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.
Launch
Notebook
6.4.8
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
Launch
PyCharm Community
2022.1.1
An IDE by JetBrains for pure Python development. Supports code completion, listing, and debugging.
Launch
Qt Console
5.3.0
PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.
Launch
Spyder
5.1.5
Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features
Launch
VS Code
1.68.1
Streamlined code editor with support for development operations like debugging, task running and version control.
Launch
Glueviz
1.0.0
Multidimensional data visualization across files. Explore relationships within and among related datasets.
Install
Orange 3
3.26.0
Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.
Install
PyCharm Professional
A full-fledged IDE by JetBrains for both Scientific and Web Python development. Supports HTML, JS, and SQL.
Install
RStudio
1.1.456
A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.
ANACONDA.
Secure your software supply chain from the source
Upgrade Now
End-to-end package security, guaranteed
Documentation
Anaconda Blog

VS Code or PyCharm

- Lightweight Code Editor vs Full Featured IDE
- Jupyter (requires extension) vs PRO feature
- Limited AI Support vs PRO feature
- Python support (requires extension for VS Code)
- Supports multiple languages vs primarily just Python

Virtual Environments: venv and conda

A **virtual environment** is an isolated workspace that contains its own Python interpreter and libraries — separate from the system-wide Python installation.

venv	Built-in Python module for creating lightweight environments	• Simple projects • Python-only • Uses pip
Conda*	Environment manager from Anaconda; supports non-Python packages too	• Data science • Machine Learning • Scientific computing

*I will be using a DSCI 222 virtual environment, created in Anaconda

Python isn't the only language we will be using

Markdown

LaTeX

- Used to create Jupyter Notebooks
- Uses Markdown for notebook cells
- Made for collaboration
- Natively supports many Data Science packages
- No installation needed, only a Gmail account
- Cloud-based
- Access to datasets through Google Drive

Collaboration while writing Python code

- One of the most important features of Google Colab is the ability to share your work with others in real-time. You can collaborate with your team members, colleagues, or friends by sharing your Google Colab notebook with them.

- Sometimes you may need to share a Google Drive with all the members of Google Colab, especially if you are working on a group project where data needs to be shared among team members.

LaTex: Overleaf

Menu | Upgrade | Project | Review | Share | Submit | History | Layout | Chat

Source | Source (legacy) | Rich Text | Recompile

frog.jpg
main.tex
sample.bib

File outline
- Introduction
- Some examples to get st...
  - How to create Sectio...
  - How to include Figur...
  - How to add Tables
  - How to add Comme...
  - How to add Lists
  - How to write Mathe...
  - How to change the ...
  - How to change the d...
  - How to add Citation...
  - Good luck!

```
1  \documentclass{article}
2
3  % Language setting
4  % Replace `english' with e.g. `spanish' to change the
   document language
5  \usepackage[english]{babel}
6
7  % Set page size and margins
8  % Replace `letterpaper' with `a4paper' for UK/EU
   standard size
9  \usepackage[letterpaper,top=2cm,bottom=2cm,left=3cm,right
   =3cm,marginparwidth=1.75cm]{geometry}
10
11 % Useful packages
12 \usepackage{amsmath}
13 \usepackage{graphicx}
14 \usepackage[colorlinks=true, allcolors=blue]{hyperref}
15
16 \title{Your Paper}
17 \author{You}
18
19 \begin{document}
20 \maketitle
21
22 \begin{abstract}
23 Your abstract.
24 \end{abstract}
25
26 \section{Introduction}
27
28 Your introduction goes here! Simply start writing your
   document and use the Recompile button to view the
```

Your Paper

You

January 14, 2023

Abstract

Your abstract.

1 Introduction

Your introduction goes here! Simply start writing your document and use the Recompile button to view the updated PDF preview. Examples of commonly used commands and features are listed below, to help you get started.

Once you're familiar with the editor, you can find various project settings in the Overleaf menu, accessed via the button in the very top left of the editor. To view tutorials, user guides, and further documentation, please visit our help library, or head to our plans page to choose your plan.

2 Some examples to get started

2.1 How to create Sections and Subsections

Simply use the section and subsection commands, as in this example document! With Overleaf, all the formatting and numbering is handled automatically according to the template you've chosen. If you're using Rich Text mode, you can also create new section and subsections via the buttons in the editor toolbar.

2.2 How to include Figures

First you have to upload the image file from your computer using the upload link in the file-tree menu. Then use the includegraphics command to include it in your document. Use the figure environment and the caption command to add a number and a caption to your figure. See the code for Figure 1 in this section for an example.

Note that your figure will automatically be placed in the most appropriate place for it, given the surrounding text and taking into account other figures or tables that may be close by. You can find out more about adding images to your documents in this help article on including images on Overleaf.

Figure 1: This frog was uploaded via the file-tree menu.

1

Important !!!

DSCI 221 provided you with basic programming skills very useful for this course.

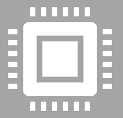
Review Chapters 2 and 3 (from pages 11 to 46) of the book “Introduction to Python for Computational Science and Engineering” by Hans Fangohr. It is an online freely available book, there is a link in the syllabus.

Read and understand the material before next class.

We will use your new skills to solve multiple problems during class.

Summary

There are multiple tools we can use to get important insights from the data.



Python provides ways to manipulate and analyse multiple datasets in a short period of time.

During this course we will learn techniques to extract significant information from data.