NYU Langone Health

NYU Grossman School of Medicine

MACHINE LEARNING: TECHNICAL INTRODUCTION

Lecture 2: Technical Introduction

- 2.1 Integrated Development Environment
- 2.2 Linux command and package managers
- 2.3 R markdown, Jupyter notebook and Google Colab

Michelle Hollenberg
2nd year PhD Student
Fenyo Lab

NYU Langone Health

Course Logistics

Please form your groups, enter your project info, and pick your project dates in the spreadsheet by September 15:
https://docs.google.com/spreadsheets/d/1xodPYg-vJQCmf1nt1gekcAyXXD6-_r0qz_DYgKK_Zbo/edit?usp=sharing
Reminder that the first presentation will be on September 27.
2-3 People per group.

If you have issues with Brightspace, please contact Sandra (Sandra.Squarcia@nyulangone.org).

Please join our course Slack to reach out to other students and TAs, and ask/discuss questions (https://ml2022-nyu.slack.com).

NYU Langone Health

Grading

After class quizzes (starting from lecture 2): **10%**

- due 2 weeks from lecture through Brightspace
- each quiz has 2 attempts and 2 hours per attempt
- attempts with highest scores will be recorded

Homeworks: **30%**

- HW1 [post Sep 13; due Sep 22 11:59pm EST]
- HW2 [post Sep 22; due Sep 29 11:59pm EST]
- HW3 [post Sep 29; due Oct 13 11:59pm EST]
- HW4 [post Oct 13; due Oct 25 11:59pm EST]
- HW5 [post Oct 25; due Nov 3 11:59pm EST]
- HW6 [post Nov 3; due Nov 17 11:59pm EST]

Project: **60%**

- Planning 10%
- Exploratory 20%
- Final report 30%

You will be able to view your grades in Brightspace.

HWs will be posted and due in Brightspace. Old HW answers from previous years' class does not help (we update the HWs every year). Every 1 hour late of HW submission results in 1% jeopardy to the HW grade (eg. 3-hour late submission can only get at most 97 points of the total grade for that 100-point HW). We will round time to the next closest hour (eg. 1hr and 2min late count as 2 hour late). This implies that late HW submission will not be accepted after 99 hours past due time, roughly 4 days. Exceptions could only be granted by Dr. Fenyo for valid reasons.

Auditors please feel free to join any project groups if you want and please on the form label yourself as auditor so that your group mates could know. We will not grade your HW/quizzes but will be happy to share answers if you want them.

Please put all **questions** in the respective HW's channel. Do not email TA's individually unless you have a question about your grade. Also, please do not share your answers of HWs or quizzes on public Slack channels.

Instructors:
David Fenyö:
David@FenyoLab.org
Wenke Liu:
Wenke.Liu@nyulangone.org

TAs:
Joshua Wang:
Joshua.Wang@nyulangone.org
Michelle Hollenberg:
Michelle.Hollenberg@nyulangone.org
Jimin Tan:
Jimin.Tan@nyulangone.org
Tayyibah Khanam:
Tayyibah.Khanam@nyulangone.org

Guest lecturers' contacts will be posted during the classes.

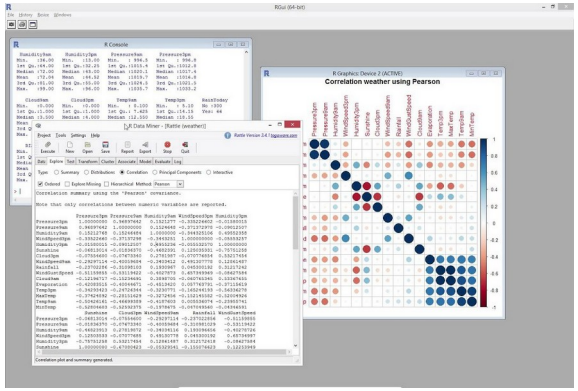
NYU Langone Health

Zoom Poll Questions

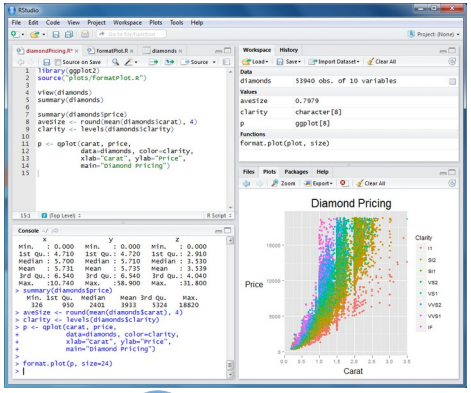
1. What is your primary programming language?
2. What is your Python experience level

Common IDEs (integrated Development Environment) for R

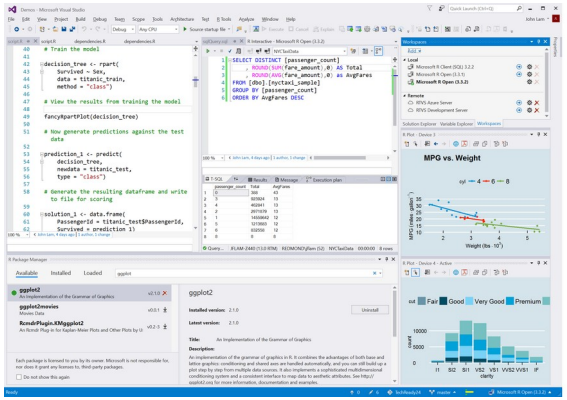
- Rstudio (the most specialized R IDE)
- R Tools for Visual Studio (RTVS)
- Rattle GUI
- RKWard



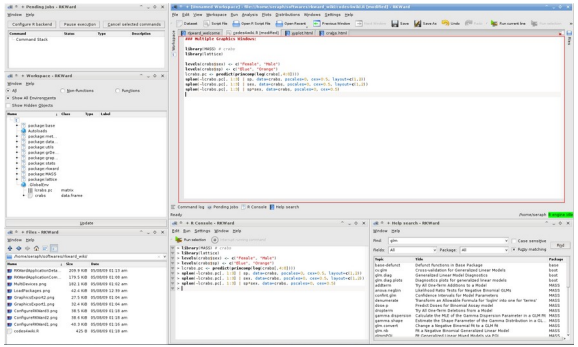
Rattle GUI



RStudio



RTVS

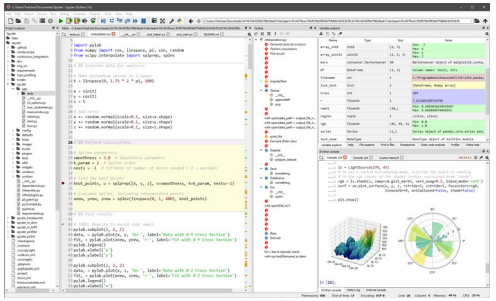


RKWard

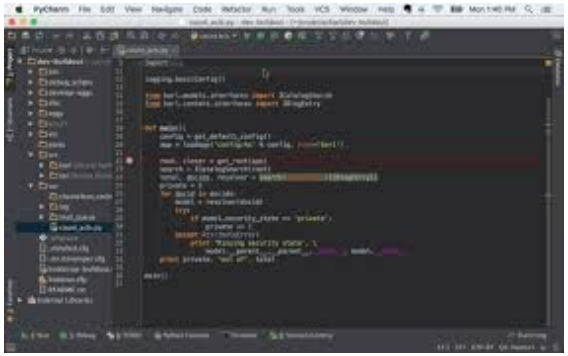
NYU Langone Health

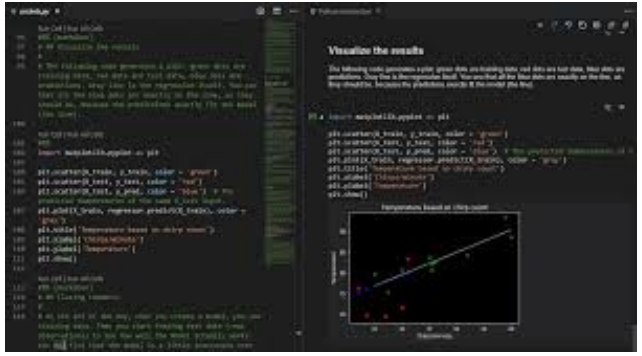
Common IDEs for Python

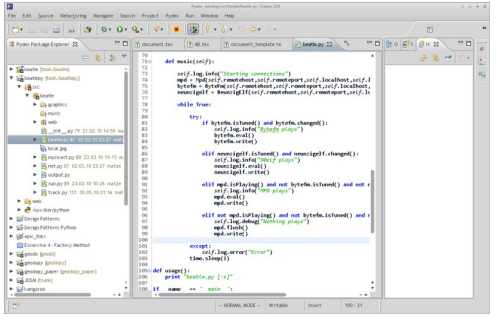
- Pycharm
- Python
- Spyder (scientific)
- PyDev (for Eclipse)



SPYDER







PyDev

PC

Python for VS Code

NYU Langone Health

Basic Linux Command

- **pwd**: print current working directory
- **ls**: list items in directory
- **cd**: change directory
- **mkdir**: make directory
- **rm**: remove file/directory
- **cp**: copy
- **mv**: move
- **cat**: display content in file
- **head**: display first X rows in file
- **chmod**: change permissions
- **ssh**: secure access
- **scp**: secure copy files
- **man**: manual of command
- **emacs/vim/nano**: text editor
- **sudo**: command with admin privilege
- **df**: available space on disk
- **du**: disk usage
- **ping**: check connection to server
- **grep**: search for regular expression matches
- **find**: find files

https://itsubuntu.com/linux-command-cheat-sheet-download-for-free/

conda

- Package, dependency and environment management for any languages
- Install a package included in Anaconda
  - conda install PACKAGENAME
- Create a new environment named py35, install Python 3.5
  - conda create --name py35 python=3.5
- Activate the new environment to use it
  - WINDOWS: activate py35
  - LINUX, macOS: conda activate py35
- List installed packages in environment
  - conda list

https://docs.conda.io/projects/conda/en/latest/user-guide/getting-started.html
https://docs.anaconda.com/anaconda/user-guide/tasks/using-r-language/

pip

- A *de facto* python package management system
- Install packages
  - pip install some-package-name
- Uninstall packages
  - pip uninstall some-package-name
- Python Package Index (PyPI)
  - Search, install, and publish python packages
    - pypi.org
    - test.pypi.org
  - pip's default source; similar to cran in R

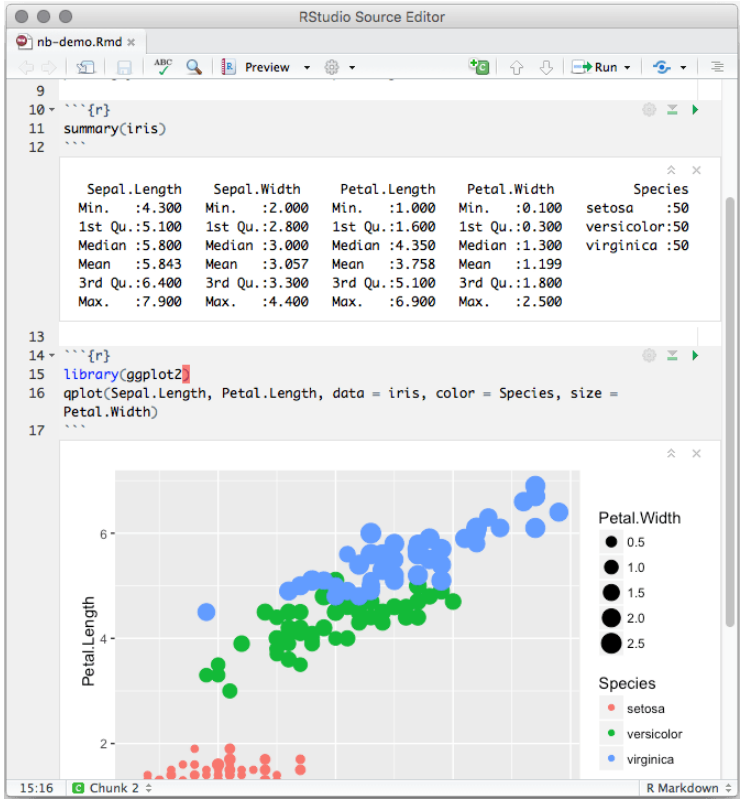
https://pip.pypa.io/en/stable/

Other Package Managers

- Homebrew (mac os and linux)
  - https://docs.brew.sh/Homebrew-and-Python
- Poetry
  - https://python-poetry.org

R Markdown

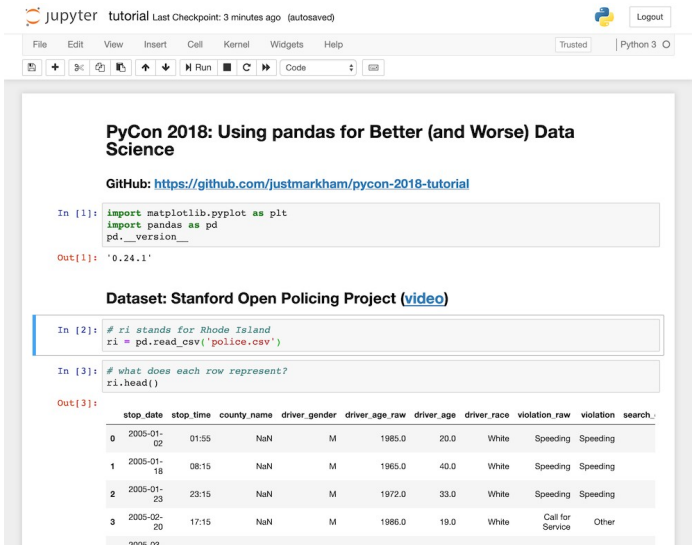
- Reproducible notebook interface for R
- Installation
  - install.packages("rmarkdown")
- File extension
  - .Rmd



https://bookdown.org/yihui/rmarkdown/notebook.html
https://rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf?_ga=2.85326297.1332753122.1593809579-2144530356.1593809579

Jupyter Notebook

- Interactive notebook interface
  - mainly for python, but also available for R and other languages
- Installation
  - pip install notebook
  - conda install -c conda-forge notebook
- Launch
  - jupyter notebook
- File extension
  - .ipynb



https://jupyter.org/index.html

Google Colab

- “Online version” of Jupyter notebook
- **We will use Colab for most of the demos and homework in this course**
  - Make sure to use your **nyu.edu** account as your Google account for Colab
  - Always save a copy to your own Google drive
  - Do not make any changes to the shared Colab files
  - Use either R or python for your homework
- Practical tutorials

Homework 1 question #1

Edit the Colab notebook:

1. Go to Brightspace Content; open “lecture 2 demo” colab notebook in lecture 2 folder.
2. Save a copy of this colab notebook to your nyu.edu Google Drive.
3. Open this copy. Rename it as “MyDemo”.
4. Add a text cell with your name before the “Getting started with Colab” cell.
5. Add another cell below your name and install “sklearn” package from pip.
6. Add another cell below and type some code to check what version of sklearn you have installed.
7. Import metrics module from sklearn.
8. Download “MyDemo.ipynb” to your computer.
9. Submit “MyDemo.ipynb” to Brightspace.

Homework 1 question #2

Download data to google drive, mount drive, and plot

1. Open HW1_googlecolab.ipynb from the assignments tab in brightspace
2. Save a copy to your nyu.edu google drive folder and make sure to do the rest of the questions in there. Do NOT edit the original HW1_googlecolab!
3. Go to Brightspace Assignments and download diabetes.prev.csv to your local computer.
4. Upload the diabetes.prev.csv file to google drive.
5. In a code cell below, mount your nyu.edu google drive to this google colab notebook.
6. In another code cell below, read in the diabetes.prev.csv file from the mounted google drive using pandas and print out the first 5 rows.
7. Make a scatter plot comparing 2 columns of your choice

Homework 1 question #3

Using the text cell below, in a few sentences, tell me a little about your computational background and why you decided to take this class!

Download this notebook as HW1_KerberosID.ipynb and upload to brightspace.

Final Remarks

- Please let us know...
  - if you have difficulties accessing NYU Google services (Colab, Google Slides, Google Drive)
- Please use Slack to
  - ask/discuss questions
  - reach out to and collaborate with your classmates

  https://ml2022-nyu.slack.com/
- After class quiz on Brightspace due Sep 27 11:59pm EST
- HW1 due Sep 22 11:59pm EST at Brightspace