DS105 - Data for Data Science

Syllabus

Dr. Jon Cardoso-Silva

31 August 2022

Syllabus

A list of what will happen every week.

Click on each Week's link for more information (slides, lab instructions, recommended resources, etc.).

Note about **formative assessment**: besides the in-person lab exercises, we might give take-home assignments on certain weeks. Even though we do not grade these problem sets, you will get written feedback on these formative assignments.

This course will help you become familiarised with the most fundamental practical tools needed to gather (Weeks 02-04) and pre-process data (Weeks 05-09) and will give you some inspiration for some fundamental analysis (Weeks 10-11) to perform in your selected datasets.

These skills are **cumulative**, so practice what you learned each week and make the most of lectures, labs and our Slack group. We believe these points of contact and integration will create a fertile environment of ideas for your project.

Remember: collaboration is key to the success of a data science project!

Intro

Week 01

Lecture

Introduction and the Data Science Toolbox

Lab

No class this week. *(Use this time to revisit basic R or python programming)*

Theme: Behind the scenes

Week 02

Lecture

Operating Systems, Files & The Terminal

Lab

Navigating the command line in your own computer

Week 03

Lecture

The Cloud: accessing and getting data in and out.

Lab

Connecting to the cloud via the command line

Summative

Worth: 10% of final marks

Problem set involving the upload of data to the cloud.

Deadline: day before next week's lab

Week 04

Lecture

The Internet: protocols + scrapping + APIs.

Lab

Web scraping exercise

Summative

Worth: 15% of final marks

Problem set involving web scraping.

Deadline: day before next week's lab

Group project

after class, students pitch their ideas of preferred APIs/datasets

there will be a designated channel on our Slack for this

Theme: Working with data

Week 05

Lecture

(Re)shaping data.

Lab

Github & Markdown

Group project

students form groups of 3

students must submit a team contract to Moodle (not graded)

Week 06

Reading Week

Week 07

Lecture

Dataframes, databases & data normalisation.

Lab

Computational Notebooks and DataFrames

Week 08

Lecture

Unstructured data (text, audio & image).

Lab

We will have group presentations instead of a structured class this week.

Summative

Worth: 15% of final marks

each group will present about their selected data.

see instructions for marking criteria

Week 09

Lecture

Managing your data science workflow.

Lab

Setting up Github for your group project & GitFlow

Theme: Applications

Week 10

Lecture

Data viz with the grammar of graphics

Lab

  Super tech-support (get help with your project)

  Week 11

Lecture

Sentiment analysis, topic modelling and social networks

Lab

We will have group presentations instead of a structured class this week

Summative

Worth: 20% of final marks

each group will present about their selected data

see instructions for marking criteria