Inf2 – Foundations of Data Science
S2 Week 7: Project Q&A

THE UNIVERSITY *of* EDINBURGH
informatics

FOUNDATIONS
OF
DATA
SCIENCE

Agenda

- Project
- Any other questions

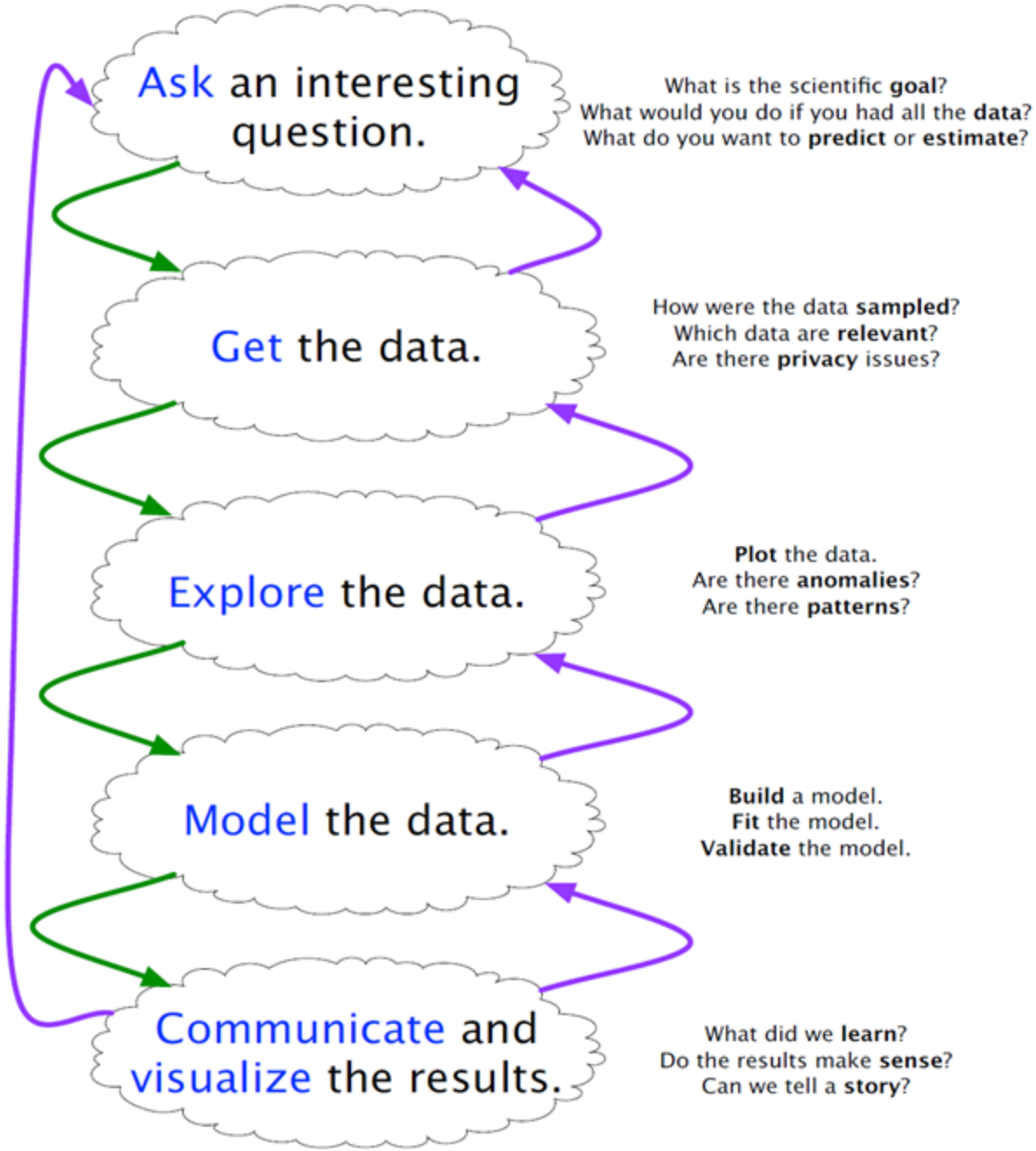


Image taken from Harvard CS109

Project overview

- Project description available in **Assessment->Coursework 3: Project** in Learn
- Can be done in groups of 2 or 3 (recommended) or individually
- This is a **marked** assignment which will count towards **40%** of your final grade for **Inf2-FDS**
- **Submission deadline:** Tuesday 2 April at 12:00 UK time
- This coursework uses the Informatics Late Submission of Coursework Rule 1: Extensions are permitted (3 days) and Extra Time Adjustments (ETA) are permitted and can be combined*.
  - See details about groups with students with ETAs

https://web.inf.ed.ac.uk/infweb/student-services/taught-students/information-for-students/information-for-all-students/your-studies/late-coursework-extension-requests

Project description

- The goal of the project is to go through the complete data science process to answer a question. You will:
  - Acquire the data, explore and visualise it
  - Apply one or more basic techniques from descriptive and inferential statistics and machine learning
  - Interpret and describe the output from your analysis
  - Communicate the results so that there is a clear story.

Dataset options

- Historical and world-wide trends in ultramarathon running
- The cancellations of planned operations in the Scottish NHS
- National university student satisfaction survey data

Individual vs. Groups

- If you are working individually, you should address the main question we have supplied.
- If you are working in a pair, you should address the main question we have supplied, and propose and address an extra question.
- If you are working in a group of three, you should address the main question we have supplied and propose and address two extra questions.
- Page limit increases with number of questions – see later.

Three requirements for submission

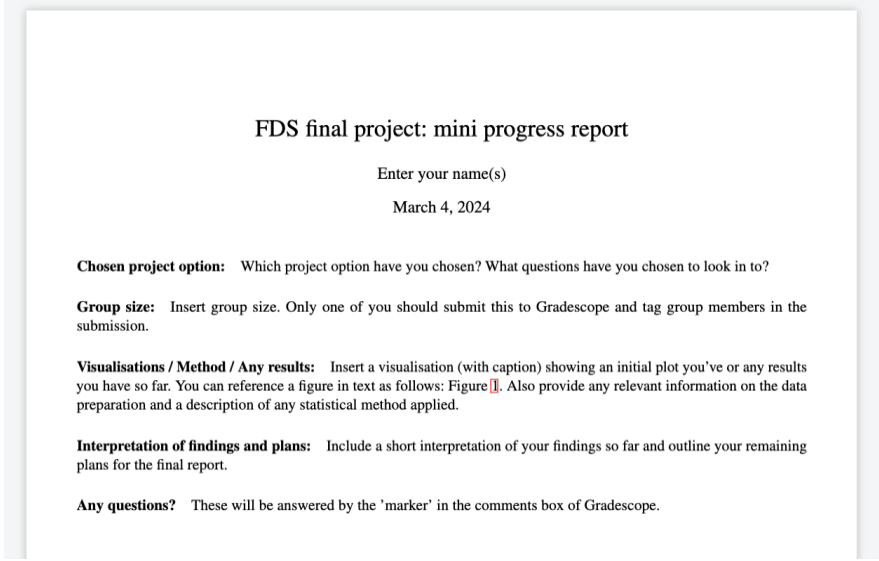
1. A short report of your project written in LaTeX
   - Submitted using Gradescope and marked using rubric on Learn
2. Jupyter notebooks and/or python files containing the code
   - Submitted as zip file, not marked, but used in cases of doubt
3. Project survey form
   - If working in pairs or threes, write a short individual statement how you divided the work, e.g. “X & Y designed the analysis, Y implemented the analysis, X did the visualisations, X & Y wrote the report”.
   - All: A few questions on the project

Feedback via written update or presentations (not for credit)

At week 8 (Monday), required to let us know whether you will either:

- Be attending a week 9 or 10 workshop to present an update on your project (e.g. at least one visualisation)
- Or submitting a mini one-page document of your update to receive some written feedback on.

Please use given latex template: https://www.overleaf.com/read/rcjrrftvqmkj#ab096



FDS final project: mini progress report

Enter your name(s)

March 4, 2024

Chosen project option: Which project option have you chosen? What questions have you chosen to look in to?

Group size: Insert group size. Only one of you should submit this to Gradescope and tag group members in the submission.

Visualisations / Method / Any results: Insert a visualisation (with caption) showing an initial plot you've or any results you have so far. You can reference a figure in text as follows: Figure 1. Also provide any relevant information on the data preparation and a description of any statistical method applied.

Interpretation of findings and plans: Include a short interpretation of your findings so far and outline your remaining plans for the final report.

Any questions? These will be answered by the 'marker' in the comments box of Gradescope.

If opting to present (not for credit)

- You or your group will sign up to a presentation slot in week 9 or 10 that happen during usual workshop slots
  - Details to follow
- We'll split workshop into two rooms (AT5.04 and AT5.01)
- You'll give a *short* presentation to your group and tutor
- The presentation is intended:
  - to be low-stress
  - to help you reflect on your progress
  - to get feedback from your tutor and peers.

Report

- Overview
- Introduction
- Context and Motivation
- Objectives
- Data Description
- Exploration and Analysis
- Discussion and Conclusion
- LaTeX template provided in Overleaf – take a copy of this template https://www.overleaf.com/read/brpnfsptvxnp

Page limit

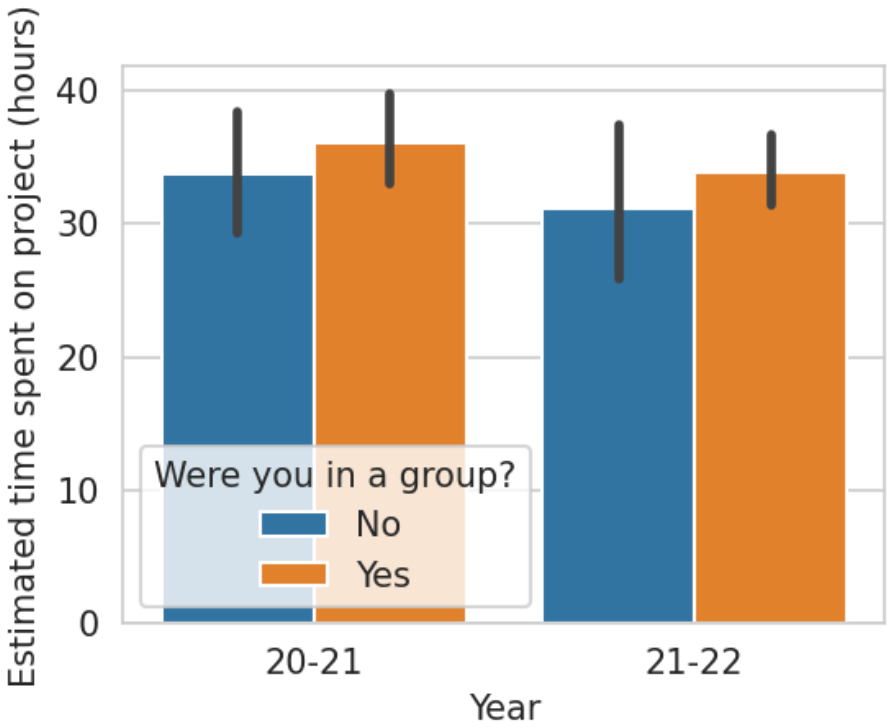
- Individual: 6 pages, excluding references
- Pairs: 8 Pages, excluding references
- Threes: 10 Pages, excluding references

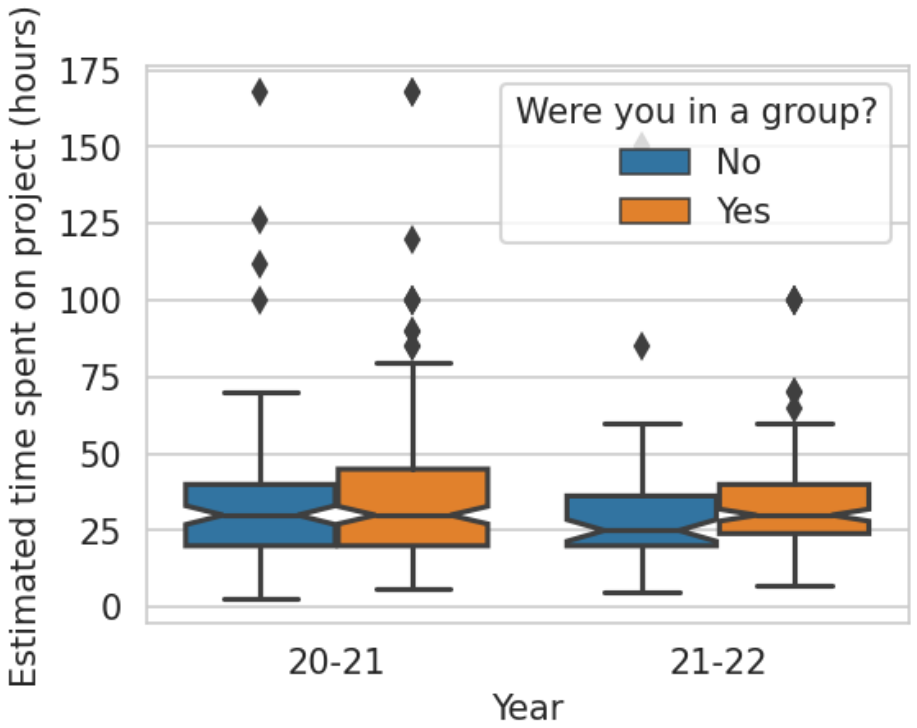
Support

- Exemplars from previous years in Learn
  - Feedback to be supplied
- InfPALS have done LaTeX tutorials - and more info later
- Writing Q&A session in Week 8 – details to be supplied
- Feel free to ask questions on Piazza
  - If in doubt make them private
  - Anna will be keeping track of them
- Feedback via presentations (last year's students appreciated them) or project update (new this year)
- We hope to offer "surgeries" later on in the project, depending on staff time.

How long should you work on this?

- 7 hours a week for 4.2 weeks => about 30 hours, the median time students estimated they took In 20-21





Some FAQs

- "Extension questions" for groups – how different do they have to be?
  - Should be distinct, but ideally would follow on from 1st question so that there is a coherent story
- Do I have to answer the main question exactly as given on the project description sheet?
  - The main question can be addressed in a number of ways – it might help to reframe as a more precise question
- Do I need to use one technique from each of descriptive stats, inferential stats and ML?
  - You **do not,** but **do** use techniques that make sense make a convincing argument

More FAQs

- Yes, please upload CSVs along with your code, but ideally compressed (e.g. .gz)
- Yes, please feel free to look for extra data
  - but question how much it adds
- Yes, feel free to use extra libraries or techniques
  - But be careful about doing complicated things before simple things

FAQ: If I get an **extension**, does it apply to my group?

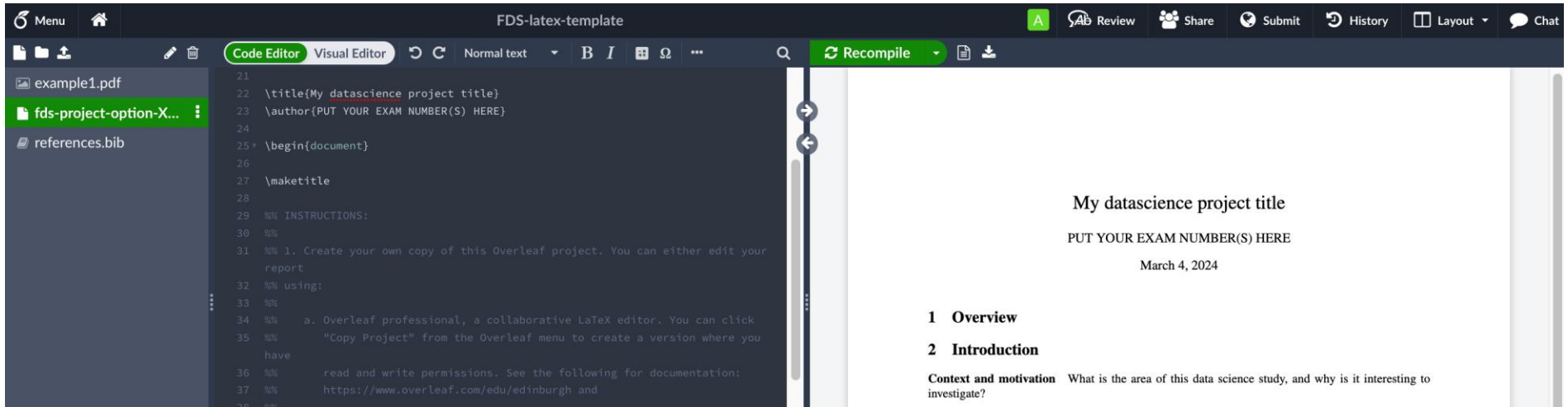
- If one student in a group has circumstances that affect their ability to complete your coursework on time, the affected student should:
  - apply for an extension
  - inform the other members of their group about the new deadline
  - inform the ITO <ito@inf.ed.ac.uk> that their group has an extension, and about the names/IDs of students in their group.
- Once the extension for the affected student is approved, the rest of the group can also apply. Applications from the rest of the group should contain the following information:
  - the affected student's name or application number
  - a short explanation of the circumstances.
- Each group member must apply separately before the original assessment deadline.

FAQ: If I have applied my **Extra Time Adjustment (ETA)**, does it apply to my group?

- Students with Extra Time Adjustments can apply for an ETA on behalf of a group.
- The student with an ETA must inform the ITO <ito@inf.ed.ac.uk> of their ETA date as well as all other students in their group, as ETAs are not normally extended to all group members.
- **The student with an ETA must make the submission.**

Why LaTeX?

- Consistent format, especially with margins and font sizes
- Used for academic papers in computer science
- Excellent for typesetting maths
- Used for the UG4 project
- Git can be used to track LaTeX code (also using Overleaf)



LaTeX resources

- *"Something that I would suggest for future years is to add a bit of course content about how to use Latex, since at least for me I had no experience whatsoever and I would have appreciated having some background on it."*
- InfPALS tutorials
- https://uoe.sharepoint.com/:f:/r/sites/digitalskillsandtraining/Shared%20Documents/LaTeX/LaTeX%20for%20Beginners%20using%20Overleaf?csf=1&web=1&e=LeBV5i
- https://www.overleaf.com/events/webinars

Advice from students from previous years

- *"Do work regularly, submit early, write a draft for report early."*
- *"Work was not easy to divide into tasks but that also forced us to meet and make progress together which was more efficient."*

Writing your project report.

- **Don't worry!** Writing is hard, but the more you do, the better you get.
- **Don’t wait.** Write. Poor writing is the enemy of great ideas!
- Papers require analysis, not just description.
- Break down big descriptions into concrete, measurable parts.
- Support key assertions with evidence: concrete examples, sources of information, footnotes, etc.

Writing Recommendations

- Develop only one important idea per paragraph.
- Writeup needs to convey: what is the problem, why it is hard, what others have done, what is your approach, why it is interesting, what are the results, what is the implication of the work.
- Introduction should be a self contained description of the above!
- A figure is worth a thousand words.
- Who is your audience?

Any more questions on the project?