Data Science Toolbox Formative Coursework 0: Data Review and group work primer

Deadline: Wednesday 12:00 Week 3

Group Project description

You will **choose an application domain** that your group will work with during Assessment 0-1.

This first challenge has the following purposes:

1. To give you experience with the assessment used,
2. To enable you to explore **collaborative working**,
3. To give you additional **experience with programming**.

Remember that this is a **formative assessment**, that is, it does not contribute to your final mark. You should however engage with it as if it were assessed.

Your challenge is to perform a literature review of the resources for helping you do data science in **your chosen application domain**. Think about the following topics:

- What are the broad types of data?
- What are the main types of resource?
- What type of problems can the resources solve?
- Are there any generic data science resources that might be applicable? In what sense are they applicable?
- How might the approach be compared to other approaches, and/or applied across different datasets?
- How is the experience of sharing code via GitHub limiting, and/or enabling?

You group should:

- Find a range of books, websites and other resources that **provide code**;
- Download and **run the examples** you find;
- Make **modifications** to the code to create **your groups' own** visualisations of the data they are designed to analyse;
- Combine them into a **report** that structures selected content logically.

It may be natural to use multiple programming languages. It is likely that your team has some prior experience with coding, so consider continuing an analysis to obtain additional insight.

Individual reflection description

- Discuss the limitations of your literature search.
- Discuss the mathematical content of at least one resource.
- Reflect on where your own comfort zone is.

- Reflect on what additional knowledge is needed to utilise these resources more effectively.
- Reflect on collaborative coding practice, and what you might change for a smoother experience next time.

Coursework guidance

This section is the same for every coursework.

Submission

Every group member must submit something to Blackboard, by the deadline. You must submit an **Individual Reflection** and a **Group Report**.

It is recommended to upload the Reflection directly to blackboard, and in the notes add a link to your Report repository.

You should submit a Report Repository containing:

1. **README.md**: An explanation of:
   - **Project Group**: List who was in the group.
     - It can be helpful to describe briefly what their contribution was, here or in the reading order. Any **Equity** variation should be clearly noted.
   - **Reading order**: the order that your files should be read in, which should explain:
     - Preparation: how to install any packages or software etc that should be installed.
     - Your report content reading order (if there are multiple files), with any additional info about the file you feel appropriate (e.g. if only some of the team were an author; the purpose of the file such as "data downloading", etc)
2. **report/**: a FOLDER, containing all of the files that will be read as part of assessing your project.
   - Label these alphanumerically as "<number>-<name>.<file ending>" for the reading order.
3. **Documentation folders**:
   - Each member of the group should be using the repository to work on the project. Make **one folder** per group member, name and merge your content into the report as possible. This will act as documentation that you have contributed to the project.

There is an Example for you to emulate, with the structure:

- README.md
- report/
  - 01-Data.Rmd
  - 02-R_analysis.Rmd
  - 03-Python_Analysis.ipynb

  - 04-Wrapup.Rmd
- RachelR/test.Rmd
- PeterP/work.ipynb

Assessment

- 75% of your mark will be for the group project itself. All students in a project should submit the same project; only one project will be run. The individual marks may be moderated away from the group project mark.
- 25% of your mark will be for an individual reflection, which should be written by you. It should be approx 500-800 words (not strict) which should be individually written.

Report

All coursework for this unit is based on group work in teams of around 3. Your team will address a single data science challenge. You will have choice about the topic, within the remit of the project description. It is always the intention that you each learn from, and teach, your teammates any skills you can bring to bear on your chosen problem. Your team will submit a single project report, which is a script that can be run to a) obtain data, b) analyse data, and c) produce any figures and tables that you feel are illuminating.

Your project script would **typically** take the form of an **Rstudio markdown** project or a **Jupyter Notebook**. It should be annotated with factual statements describing what you have done and why in basic terms. Unless otherwise stated, you may choose the programming language but we recommend sticking with python or R since all students are expected to become familiar with these. The results of computations including plots should be displayed and labelled (e.g. with numbers) and if you have not used a seamless method then you must provide a zip file containing both a script, and a pdf or similar document that also contains the output of your script. Your script is expected to run, and if at any stage some manual step is required (for example, to wait for a bluecrystal job submission to finish, or data must be downloaded) this should be carefully noted. You may lose marks if your script needs debugging.

There is no word, page or other limit. Credit will be awarded for making your arguments thoroughly but without repetition or meandering off-topic. Only include material that you feel makes a contribution to the overall project scope. If some research led to a dead end, work it into the results.

Remember to **reference** websites and other resources for content and ideas, in addition to the usual academic referencing. This will assist you in your future projects.

Report Assessment Criteria

Your project will be assessed against the following criteria:

- Fit and Success.
  - *Choice of project is very important for learning. This section includes finding good datasets and matching them to questions; making progress on hard problems; fulfilling the learning outcomes.*
- Innovation.
  - *Thinking outside of the box and finding resources that are not presented in the course. Innovation can come in the form of data, methods, and mathematical ideas brought in from elsewhere.*
- Citations, Referencing, Literature.
  - *Cite your sources, build up a repertoire of useful content. Link your results to those on analogous problems. Note that many resources are not published papers.*
- Structure and Description.
  - *Your project should be well introduced, and easy to read and undestand. Make good figures and explain them. Structure your project well, stick to the point and note what your results mean.*

Equity

Your team should try to agree an **equity** or proportional contribution to the group project, accounting for both practical (implementation) and conceptual (theory, methods choice, etc) contributions. If you cannot agree, you should approach the tutor to try to agree equity before submitting divergent opinions. Try to agree any non-even equity before the project gets underway.

Contributions will be taken into account when assigning individual marks from group reports. Small deviations are unlikely to be given divergent grades.

Individual grades can be moderated up and down based on equity but are unlikely to be increased as much as they are decreased, and the final decision takes into account documentation.

Additional notes:

- It is expected that all group members understand the group submission.
- It is also the intention that they put in equal effort.
- It is not expected that the final script contains content proportional to equity. There are many good reasons that work does not make the final report.
- If you put in lower effort and agree a lower equity, you may receive a proportionally lower group mark.
- If you put in extra effort and agree a higher equity, you may receive a higher mark but the reward is not linear. It is better to have an equal share of a good project, than a high share of a poor project.
- Mathematical contributions and programming contributions can be considered. All contributions should be documented.
- If some people choose lower equity because they could not contribute fully, make this clear in all reflections. The lower manpower may mitigate a low

grade.

Documentation

All students are expected to contribute to programming. You should each submit your own scripts, session history or similar, that demonstrate that you made some independent effort, even if these did not make it to the final report. If you cannot demonstrate an amount of effort commensurate with your claimed equity, then your mark may be reduced.

Your documentation is likely to take the form of an Rstudio markdown or Jupyter Notebook. It can be long and contain dead ends. It does not need to be documented, nor be able to run from top-to-bottom. It should be unique to you, though is likely to contain content from others' notebooks. You may refer to it in your individual reflection, but if there is excessive material that should have been shared with the group then you will not receive credit for it. You should not try to boost your individual grade by doing extra work here. It may not be carefully read and you may not receive feedback on it. It should be no additional effort to produce this as it should consist of files that you already have.

Individual Reflection

The purpose of your reflection is to encourage changes in your practice that improve your understanding of Data Science, as well as improve your ability to work in a team on Data Science projects.

You are being assessed on your progression and understanding of the content of the project. It is better to note deficiencies with what you have done, than to try to post-hoc justify something. It is understood that you are under time pressure and may make a poor irreversible decision for the project performance, but that will not strongly affect your mark if the reason for the failure is clear. You must write your writeup independently of the other students, though using the shared understanding gained from working with them.

Reflection Assessment Criteria

Your reflection should always adress the following areas in addition to what is asked in the specific project description.

- Fit.
  - *Introduce the area and explain the overall goals.*
  - *Justify the decisions made in the project.*
- Depth.
  - *Explain the results and discuss the conclusions. This can focus on your contributions but should also include the project as a whole.*
  - *Briefly explain some aspect of the mathematical model(s) that has been used.*

        * *It is expected that your group will discuss this in detail, and that contribution of understanding is included in the project contributions.*
        * *Each student still must write something in their own words.*
    - *Reflect on the strengths and weaknesses of your approach, and how you might do it differently next time.*
    - *Reflect on any aspects of the project that could be improved, paying particular attention to* **group working**, *technological barriers and solutions.*
- Structure and Description.
    - *Write clearly, use referencing if appropriate. This is a reflection, not a report, so use appropriate language.*
- Evidence.
    - *Your documentation and reflection together are assessed to evidence your individual contribution. Explain this and how it fits into the whole.*

Learning outcomes

You are reminded that:

- Teamwork is a learning outcome.
- Progress, teaching and sharing is more important than individual technical ability or project success.
- The difficulty of these assessments is beyond what would be expected of an average student alone.
- Most groups will contain a mixture of expertise which should be exploited.
- In the event that your entire group is inexperienced at programming, you still need to meet a minimum standard. However, you can still score well if you focus on a mathematically interesting question.

Working practice

You should work on this project together. This may mean all group members trying different things and coalescing on a final approach. Trying things that fail is still a contribution. Failure can be included in the report if something meaningful was learnt.

To work physically separately, you should:

a) arrange a suitable **discussion forum** for your group such as a WhatsApp group, slack, etc.
b) arrange a suitable **file sharing location** such as github, OneDrive, Dropbox, or GoogleDrive.
c) Collaboratively decide the final content, merging all versions of the analysis.

You should finalise the project content at least 48 hours before the deadline, so that individual writeups can be written.

Marking Criteria

The mark ranges and descriptions in normal type below are the University of Bristol Generic Marking criteria that apply to any assessment at the University - these can be found at www.bristol.ac.uk/esu/assessment/codeonline.html. The descriptions in bold type are additional maths-specific criteria introduced primarily to clarify the descriptors in the case of marking maths examinations.

0-100 scale	**Criteria to be satisfied** University generic marking criteria in normal type, **Maths-specific marking criteria in bold**
100 94 89	• Work would be worthy of dissemination under appropriate conditions • Mastery of advanced methods and techniques at a level beyond that explicitly taught • Ability to synthesise and employ in an original way ideas from across the subject • In group work, there is evidence of an outstanding individual contribution • Excellent presentation • Outstanding command of critical analysis and judgement **and** • **Work develops concepts not directly presented in course material or uses known concepts to answer hard, unfamiliar questions that require calculations/methods not similar to any course material** • **An elegance of mathematical work beyond that expected for the level of the course** • **Of a quality that could be distributed to fellow students as an example of exceptional work**
83 78 72	• Excellent range and depth of attainment of intended learning outcomes • Mastery of a wide range of methods and techniques • Evidence of study and originality clearly beyond the bounds of what has been taught • In group work, there is evidence of an excellent individual contribution • Excellent presentation **and** • **On standard but unfamiliar problems, carrying out calculations with no errors of understanding** • **Demonstrates a high level of technical competence with very few mistakes of any kind** • **Great clarity in mathematical arguments**
68 65 62	• Attained all the intended learning outcomes • Able to use well a range of methods and techniques to come to conclusions • Evidence of study, comprehension and synthesis beyond the bounds of what has been explicitly taught • Very good presentation of material • Able to employ critical analysis and judgement • Where group work is involved there is evidence of a productive individual contribution **and** • **Able to make a good attempt at standard but unfamiliar problems, with some minor errors** • **Demonstrates technical competence, perhaps with some shortcomings** • **Clear mathematical arguments**

0-100 scale	Criteria to be satisfied University generic marking criteria in normal type, **Maths-specific marking criteria in bold**
58 55 52	• Some limitations in attainment of learning objectives, but has managed to grasp most of them • Able to use most of the methods and techniques taught • Evidence of study and comprehension of what has been taught • Adequate presentation of material • Some grasp of issues and concepts underlying the techniques and material taught • Where group work is involved there is evidence of a positive individual contribution **and** • **Able to start standard but unfamiliar problems but with significant errors** • **Able to complete competently "bookwork" questions that have been seen in the course material**
48 45 42	• Limited attainment of intended learning outcomes • Able to use a proportion of the basic methods and techniques taught • Evidence of study and comprehension of what has been taught, but grasp insecure • Poorly presented • Some grasp of the issues and concepts underlying the techniques and material taught, but weak and incomplete **and** • **Able to complete "bookwork" questions that have been seen in course material with few errors** • **Gaps or inconsistencies in the mathematical argument**
35	• Attainment of only a minority of the learning outcomes • Able to demonstrate a clear but limited use of some of the basic methods and techniques taught • Weak and incomplete grasp of what has been taught • Deficient understanding of the issues and concepts underlying the techniques and material taught **and** • **Able to reproduce work seen in course material, but with some errors**
7-29	• Attainment of nearly all the intended learning outcomes deficient • Lack of ability to use at all or the right methods and techniques taught • Inadequately and incoherently presented • Wholly deficient grasp of what has been taught • Lack of understanding of the issues and concepts underlying the techniques and material taught **and** • **Unable to reproduce satisfactorily even "bookwork" questions that have been seen in course material**
0	• No significant assessable material, absent or assessment missing a "must pass" component