

# Midterm project instructions

STATS/DATASCI 531, Winter 2022

---

**Midterm project outline.** Find a time series dataset of your choice. Carry out a time series analysis, taking advantage of what we have learned so far in this course. Write a report, submitted on Canvas as a zip file by the deadline, 11:59pm on Monday 2/21. The zip file should contain the following:

1. A file called `unblinded.Rmd` and its compiled version `unblinded.html` which are the full report file. This version should have the names of the group and should include a section explaining the contributions of group members. These are the files that the grader is expected to spend most time with, though other files may also be inspected.
  2. A file called `blinded.Rmd` and its compiled version `blinded.html` in which all identifying text is removed. This version will be used for anonymous peer review and posted on the course website.
  3. Data files and any other files needed to compile the Rmd on a standard R or Rstudio environment. You can assume that the grader will install any required libraries.
- 

**Groups.** There will be randomly assigned groups of approximately 3 people. As pointed out in the syllabus, in special situations you can request to write an individual midterm project. This may be appropriate if you have a particular dataset or scientific question that has motivated your interest in learning time series analysis. You must request this before groups are assigned. Once you are in a group you have to come to an agreement with your group on what data to analyze. You will not be able to request your own group partners - in order to treat everyone fairly, groups will be randomized.

---

**Choice of data.** The time series should hopefully have at least 100 time points. You can have less, if your interests demand it. Shorter data needs additional care, since model diagnostics and asymptotic approximations become more delicate on small datasets. If your data are longer than, say, 1000 time points, you can subsample if you start having problems working with too much data. Come ask the instructor or GSI if you have questions or concerns about your choice of data.


---

**Data privacy and project anonymity.** The projects, together with their data and source code, will be posted anonymously on the class website unless you have particular reasons why this should not be done. For example, you may have access to data with privacy concerns. The projects will be posted anonymously. After the semester is finished, you can request for your name to be added to your project if you want to.

---

**Expectations for the report.** The report will be graded on the following categories.

- Communicating your data analysis. [10 points]
  - Raising a question. You should explain some background to the data you chose, and give motivation for the reader to appreciate the purpose of your data analysis.
  - Reaching a conclusion. You should say what you have concluded about your question(s).

- Communication style. Material should be presented in a way that helps the reader to appreciate the contribution of the project. Code and computer output are included **only if they are pertinent** to the discussion. Usually, code remains in the source file, and numerical results are presented in tables or graphs or text, rather than raw computer output. 
- You will submit your source code, but you should not expect the reader to study it. If the reader has to study the source code, your report probably has not explained well enough what you were doing.
- Statistical methodology. [10 points]
  - Justify your **choices** for the statistical methodology.
  - The models and methods you use should be **fully explained**, either by references or within your report.
  - Focus on a few, carefully explained and justified, figures, tables, statistics and **hypothesis tests**. You may want to try many things, but only write up evidence supporting how the data help you to get from your question to your conclusions. Value the reader's time: you may lose points for including material that is of borderline relevance, or that is not adequately explained and motivated.
  - Correctness. Obviously, we aim to avoid errors in the math we present, our code, or the reasoning used to draw conclusions from results. Being self-critical and paying attention to detail can help here.
- Scholarship. [10 points]
  - Your report should make references where appropriate. For a well-written report the citations should be clearly linked to the material. The reader should not have to do detective work to figure out what assertion is linked to what reference.
  - You should properly acknowledge any sources (people or documents or internet sites) that contributed to your project.
  - You are welcome, and encouraged, to look at previous projects, linked from the course website ([../index.html](#)). If you address a question related to a previous project, you should put your contribution in the context of the previous work and explain how your approach varies or extends the previous work. It is especially important that this is clearly explained: substantial points will be lost if the reader has to carry out detective work to figure out clearly the relationship to a previous project.
  - When using a reference to point the reader to descriptions elsewhere, you should provide a brief summary in your own report to make it self-contained.
  - Credit between group members. Your report should explain clearly **how work was divided among** members of the subgroup. Usually, there should be **a short section explaining** how the group operated.

---

**Plagiarism.** If material is taken directly from another source, that source must be cited and the copied material clearly attributed to the source, for example by the use of quotation marks. Failing to do this is plagiarism (<https://en.wikipedia.org/wiki/Plagiarism>) and will, at a minimum, result in zero credit for the scholarship category and the section of the report in which the plagiarism occurs. Further discussion of plagiarism can be found in *On Being a Scientist: A Guide to Responsible Conduct in Research: Third edition* (2009) ([http://www.nap.edu/catalog.php?record\\_id=12192](http://www.nap.edu/catalog.php?record_id=12192)), by *The National Academies Press*. Here is how the Rackham Academic and Professional Integrity Policy (<http://www.rackham.umich.edu/current-students/policies/academic-policies/section11#1132>) describes plagiarism:

### 11.2.2 Plagiarism

Includes:

Representing the words, ideas, or work of others as one's own in writing or presentations, and failing to give full and proper credit to the original source.

Failing to properly acknowledge and cite language from another source, including paraphrased text.

Failing to properly cite any ideas, images, technical work, creative content, or other material taken from published or unpublished sources in any medium, including online material or oral presentations, and including the author's own previous work.