

10% off points possible for each day late. No late projects will be accepted after December 16th.

There are 3 different tracks for this project, depending on your interests and what you registered for.

Option I) STAT 436: DATA ANALYSIS VERSION – REPLACES LOWEST HW GRADE

Take a data set (not one that I used) and perform a complete time series analysis of the data using an ARIMA model. The main focus is on the analysis of the data set, with the write-up focused on what the time series measured, the conclusions based on the model(s) fit and justifying the model(s) selected. This project would lead to an article-style report that is focused on the results with minimal discussion of the methods. It will contain just enough detail on the methods so that a somewhat sophisticated reader could replicate your results if they had your data set. It will have a data set motivating the study and requires some explanation of where/how it was collected. It will focus mainly on conclusions and interpretations of results. You can use the “report” format you used in STAT 411/511 to structure your results.

Option II) STAT 536: DATA ANALYSIS VERSION

Take a data set and analyze it using methods not discussed in class or with a combination of methods discussed in class and new methods. It must contain a notable element of methods not discussed in class. For example, you could find an R package that includes a time series method that we did not discuss and try to use it on a data set and write up the results. The main focus is on the analysis of the data set, with the write-up focused on conclusions and justifying the models selected. You would need to contain enough detail on the methods so that a somewhat sophisticated reader could replicate your results if they had your data set. It will have a data set motivating the study and requires some explanation of where/how it was collected. It will focus mainly on conclusions and interpretations of results. Try to identify and read a few articles in an area related to your application that use similar methods – use the papers for ideas of how to structure your paper and “sell” the analysis that you did.

Option III) STAT 536: REVIEW OF METHODS

You could review a set of methods related to time series that we did not cover in class, trying to summarize them for an audience that might not be familiar with those methods and might find them interesting. To motivate the methods, it could be useful to provide an example of their use on a real data set, but this is not required. You should think carefully about the audience you are writing for and make the level of review and explanation match your audience. The emphasis for this project will be on the quality of your explanation of the methods and how much interest in the methods you create. A reader should want to consider your methods in their next application paper in this area. It should contain an application of the methods to a data set but the data set could be simulated or not of primary interest.

Option IV) STAT 536: NEW METHODS

You can explore or compare different methods that have not been compared previously or take a set of methods and move them into a slightly different area of application and assess their performance. This type of project begins to resemble original research so would be looking for “a hook” that could potentially lead to this being publishable research. It could be a case-study that compares methods or discusses a new area of application of a method that has been previously proposed. This should look like a stat research article but also possibly include an application to motivate the methods if possible. You should have more stat references in this sort of paper to lay out how this fits with other existing stat research. It could contain a simulation study even if it is just to illustrate a weakness in a particular method.

Not all journal articles fit into these options and yours might not as well. It might combine a real data set of interest and a comparison of methods and conclusions that would be made using the different methods. In this case, you would want to contrast the assumptions of the methods and highlight how the inferences using the different methods are similar or different for your application. Your paper can share aspects of different types of papers, but try to find a simple story to tell.

For STAT 536 only:

I will be evaluating your selection of and suitability of writing for your selected audience. **To aid in this evaluation, you must clearly identify the journal that you are writing this for.** When you submit journal articles, you typically write a short letter to the journal editor briefly describing your submission and why it is of interest for the journal and include any co-authors and affiliations. **You must attach this as a cover sheet for your manuscript.**

You also need to include a title and an abstract!

Depending which path you take, I will grade accordingly. Don't go overboard but also don't leave out important details or stop before you should be done. **Conciseness and completeness** will both be valued. In thinking about the journal article format, note that you need to introduce the problem, summarize your explorations, and make some conclusions. You need to be efficient with your words but include enough detail and work that a reader knows what you did and doesn't have too many openings to criticize what you included or did not include.

Journal articles except those in the *R Journal* typically do not include R code and they **do not include R output**. They do include references to R packages used. See `citation()` in R for more on finding this information. You can include edited R code as an Appendix but the appendix shouldn't contain any R output, just the code used to produce the needed results. Move any useful R output into nicely formatted Tables and do not include extraneous aspects of results anywhere in the document.

Use graphs as much as possible to tell the story of your data set. **All graphs and tables need to be labeled, captioned, and referenced in the text.** If you don't reference it directly in the text of the document, don't include it. Diagnostics are important but might be considered for inclusion in an appendix to the main document. You should at least acknowledge checking diagnostics if you are applying typical time series models.

You do not need to follow exact conventions of the journal you pick as a target, but you do need to follow some version of regular citation system and labeling.

- For all 536 projects, I will expect a well-written report in a **journal article style format**. It needs to be double spaced. No R code should show up in the report but can be included in the appendix in a "clean" format. It must contain at least 2 references that are cited in the document.
- There is no minimum or maximum length – both including too much and including too little will be a reason for penalty. Think about getting to the point but not being vague or incomplete on details.

Grading rubric:

Choice of project (interesting, challenging, and it was of interest to you): _/5
Title and abstract _/5

If you have a data set (depending on how hard this, it can be a significant portion of the work or some of you just loaded data which already existed so this might be worth more or less on your projects): _/10

Description of ALL variables (full detail can be provided in Appendix)

Appropriate visual description of data (maybe numerical?)

Data processing (any observations removed (missing data?) or variables transformed, reason?)

Description of the story of the data/ method and why the analysis you are doing is of interest: _/5

Figure labels _/5

References _/5

"Work" _/10

Details of methods used (explain them briefly but focus on why they are used and how they work)

Reason for any choices made in considering techniques or making them work

Presentation of results (include test stat's, df(s), p-values(?), other important stuff)

Tying the results into the story that you were trying to tell
Detection of any unusual observation(s) or unusual features in data set or performance of the methods

General criteria:

Clarity	_/5
Accuracy/ Correctness	_/5
Conciseness	_/5
Ability to find and report a clear story from the data/method	_/5
Completing the project (not stopping before you found the end)	_/5
Quality	_/5

Total = /75

Some possible ideas for projects involving methods I will not cover:

Time series models for discrete data (Count or Binary)

Multiple testing adjustments:

for ACF/PACF exploration

for trend tests

Comparison of differencing and spline-based derivative estimation techniques

Model averaging for ARIMA models

Bootstrapping the AIC

Bootstrapping GLS models

Comparing AIC, AICc, BIC or other model selection criteria (simulation study?)

Nonlinear regression models with correlated errors (gnls function in nlme)

Effects of scale/ filtering on applications of regression to time series

State space models (multiple packages in R available)

Varying coefficient models (different fitting methods?)

Dynamic linear models (sometimes via state space models)

Threshold autoregressive models

Bayesian versions of everything in the books

Multivariate time series:

VAR (R packages VARS or MARSS?)

Clustering techniques (clustering functional data or time series)

State space models

Transfer function models

Frequency domain methods:

Spectral estimation and multiple testing issues

Lomb-Scargle (Spectral estimation for time series with missing or unevenly spaced data)

Spectral regression

Wavelets (R packages available)

Estimation

Filtering applications

Regression

....

This is just a quick list of the vast possibilities for methods to explore. Please feel free to identify a project that is not in this list.