Final project time! This is an opportunity to showcase all that you've learned this semester. This project is rather open-ended, which means it is better to start thinking about what kind of project you'd like to do sooner rather than later.

1 Basic description

For the final project, you are to complete a data analysis using Bayesian methods that we have learned from this class. Ideally, you would find some data that interests you (more on that later), develop a research question, and answer that question using a Bayesian statistical model. The final project consists of five deliverables on three different dates:

1. Wednesday, 12/10 at 11:59pm: a rough draft of the Introduction and Model description of your report (see Section 2.1) in PDF form, submitted to Canvas.
2. Friday, 12/12 at 9am-12pm: presentation
3. Wednesday, 12/17 at 11:59pm:
   (a) A final draft of the written report in PDF form, submitted to Canvas.
   (b) A final draft of the written report in the original `.Rmd` or `.qmd` file, submitted to Canvas.
   (c) A short reflection on the project, submitted to Canvas.

As a zero-point deliverable: please email Becky by Tuesday 12/2 at 12:00pm a list of four models from Section 3.1 you would be interested in exploring for your final project, or if you have a dataset/other model in mind. Then by later that afternoon, Becky will assign everyone their final project topic.

The main difference between the final project and the case studies is that you will be working alone and you may be teaching yourself something new. The reports should have a similar structure to the case studies. However, the final report for this project should have a more robust description of the data and research question, as well as a more lengthy results and interpretation section. My hope is that you are invested and interested in the data/model you choose, so you cannot help but want to describe everything you find!

2 Project requirements

Once you have selected a method or data set, you should identify how you are going to use the methods you have learned in this class to answer some research questions. My expectation is that you use, at a minimum, some sort of MCMC algorithm (e.g. hand-coded Gibbs or Metropolis, or JAGS), instead of performing all analyses using the basic conjugate models that we learned in the first third of the semester. You are required to fully specify and defend your choice of priors. Then, you should provide results – beautiful plots and informative tables – that are related to the research question of interest, along with a healthy amount of interpretation, once again in the context of the research question.

Each project should teach you something new, and you will in turn share that new topic/idea with everyone else during the presentations.

2.1 Report requirements

Your report must be completed in an `.Rmd` or `.qmd` file, which will be submitted alongside your final report. At a minimum, the report must have the following components as clearly marked sections. Each section must contain the information as described below.

1. **Introduction**: this section should describe the data and any interesting nuances/features that would need to be addressed by a model. This may include some exploratory data analysis (i.e. plots/tables demonstrating interesting features to the reader who have never seen your data before). If you found your own data, you should also clearly describe from where you acquired the data. The last paragraph/few sentences of the Introduction should clearly describe your research question(s)/analysis goal(s).

2. **Model description**: you should clearly (ideally with Latex) write out the full model(s) that you will implement. It should be clearly stated which parameters are unknown and need to be estimated during model fitting. You must also specify all the priors you will use. For *both* the model specification and priors, you should provide a few sentences defending or justifying why you made the choices you did. Though not required, you may find it a good idea to include some plots that demonstrate your prior beliefs.

   You should also clearly describe how this model will address your research question(s). What parameters/derived quantities will be of interest to you?

3. **Model implementation**: clearly describe how you will implement your model. If you will run a Gibbs sampler, derive the full conditionals here. If you code up the Metropolis algorithm, discuss proposal distributions. If you need to do change of variable, provide the new density. Also be sure to include information such as length of burn-in, number of samples stored after burn-in, and number of chains run (if applicable). *When possible, I would prefer you code up your sampler by hand than use JAGS. You may not know what "when possible" means, so feel free to ask Becky if she thinks JAGS is actually appropriate for your method.*

   The model description should be clear enough such that someone else could implement the model without any doubts.

4. **Model fitting**: this section should contain only the `R` code used to *fit* the model, assess convergence and, if appropriate perform a posterior predictive check. No results related to the research question should be provided here. The code you use to implement the model should be made visible to the reader. Code for assessing convergence does not.

5. **Results**: results related to your research question(s). Everything provided in this section should be related to the research question. Do not present plots/summary statistics/etc. just for the sake of doing so. All outputs should be followed with thorough interpretation. Each plot/table should illuminate new information for the research question; try to avoid presenting the same information multiple times.

6. **Conclusion**: a brief conclusion summarizing your findings in the context of the research question, limitations, and possible future directions.

Additionally, your code should be as reproducible as possible!

2.1.1 Rough draft

The rough draft should include a relatively complete Introduction section and Model Description section, and ideally some preliminary code + results. The rough draft you submit need not be fully polished, but it should still have all these components. If you don't have all the results yet, that's okay! But you should have at least a few preliminary results. The purpose of the rough draft is for Becky to give you feedback before you submit your final report.

2.1.2 Final draft

The final draft of your report should be polished and professional. In addition to the specific content detailed above, I am expecting the following:

- All the code presented in your knitted/rendered report should serve a purpose. Be sure to include the code for fitting your model. Code for creating figures/tables is usually not needed. However, if there is some figure/table whose code you think is really worth showing the reader, do feel free to include it! *Please please please: hide the code where you load in packages, and any accompanying messages that are printed when you load them in. Come ask Becky if you don't know how to do this.*
- Reasonably-sized figures. This can be achieved by typing this code in your first chunk of code: `knitr::opts_chunk$set(fig.width = 5, fig.height = 3, fig.align = "center")`. Feel free to play around with the width and height.
- Addressing at least two comments or questions from your peers' questions/comments during your presentation.
- Addressing all the feedback from Becky's comments on your rough draft.
- Uploading your data: if you found data elsewhere, please also submit it alongside your final report.

2.2 Project presentation

Everyone will be required to give a presentation on the last day of class. These presentations will be 10 minutes long (tentative), with 1-2 minutes for questions and comments from your peers.

2.2.1 As a presenter

Your presentation should be shared with the help of a brief slide deck. Due to the short time frame, I suggest no more than 10 slides total. Good presentations use slides to add to the content being shared verbally by the presenter; the slides should not be telling the whole story. I also highly encourage you to practice a few times! You're welcome to schedule a time with Becky to practice.

Because the final report is due after the project presentations, you are not required to have all the results by the day of presentations. If this is the case, you should clearly state during your presentation what you intend to explore further so your peers can give quality feedback.

Becky will be taking notes during the presentation, especially with regards to the feedback and questions from the audience. She will then compile this feedback and send it to the presenter. The presenter should clearly address at least two of the comments/questions from this feedback in their final report.

2.2.2 As an audience member

You are expected to give full attention to each presentation, and to ask questions and provide constructive feedback at the end of the presentation. If you do not ask questions or provide suggestions, then whoever is presenting will not be able to incorporate feedback into their final report and will therefore automatically not receive points!

2.3 Reflection

One the same day you submit your final report, you will also be asked to submit a brief reflection (no more than one page, single-spaced) on the final project and the course as a whole. Please address the following in the reflection:

- What feedback from your peers did you incorporate into your final report?
- What aspect(s) of your project (written and/or presentation) are you most proud of?
- What aspect(s) of your project (written and/or presentation) do you wish you could have improved on?
- How do you think your final project demonstrates your learning from the course?
- How do you think you grew/developed as a statistician through this project and this class as a whole?
- What suggestions/feedback do you have for the project if Becky teaches this class again?
- And anything else that you'd like to reflect on!

3 Parting thoughts

3.1 Project ideas

Here is a brief list, in no particular order, of new models or topics you might learn for this final project. Each one will be accompanied by a dataset already cleaned and ready-to-go:

1. Latent class modeling: these models identify unobserved (latent) groups within a population based on patterns in observed responses. They assume individuals within each group behave similarly, allowing researchers to infer hidden structure
   (a) **Option 1**: extending the disputed authorship model: we will extend the work from Case Study 2 to estimate/predict who wrote the disputed articles.
   (b) **Option 2**: occupancy modeling (ecology): these models estimate the probability that a species occupies a site while accounting for imperfect detection. They separate the ecological and observation process
   (c) **Option 3**: Cormack-Jolly-Seber (CJS) model for survival and detection (ecology): following a mark-recapture sampling scheme, this model estimates survival and detection probabilities for marked individuals over time
2. **Zero-inflated modeling**: these models handle data with more zeros than expected under classic discrete distributions (e.g., Poisson). Particularly useful when zeros can arise for multiple reasons.
3. **Change point detection**: these models identify points in a data sequence where model parameters (such as mean, variance, or trend) shift. These methods help detect regime changes, anomalies, or structural breaks in time series or spatial data.
4. **Binomial data, unknown n and p**: what happens when only aggregated successes are observed but the total number of opportunities is not directly known?
5. **Simple variable selection in regression**: attempts to identify the subset of predictors that best explain or predict the outcome while avoiding unnecessary complexity. Particularly useful when there are tens-hundreds of predictors available to you
6. **Item response theory (IRT) modeling**: these models estimate the relationship between a test-taker's latent ability and their probability of answering items correctly. They characterize item properties (difficulty, discrimination) and individual ability simultaneously
7. **Survival analysis**: methods that model the time until a certain event occurs (e.g. patient mortality, machine failure). In this project, you will work specifically with the Weibull to model survival times.
8. **Model comparison of frequentist versus Bayesian regression models**: let's compare these two approaches by evaluating model fit and predictive performance.

You are also welcome to do a regression analysis using some data of your choosing (e.g. from another class) or teach yourself a different model than one provided here. If using your own data or a model not provided above, you will have to come up with the research question(s) yourself.

3.2 Meet with Becky

You are encouraged to meet with Becky to discuss your project and to practice your presentation if you think that would be helpful to you. There will be a Calendly link where you can sign up for 1:1 meetings.

4 Rubric

Component	**Description**	**Points**
Rough draft	Submitted on time with good-faith effort.	4
Presentation	Clearly describes project data, research question, model, and results to the audience.	10
	Has an accompanying slide deck that is error-free.	5
	Within 30 seconds of time limit.	1
	Is an active listener, asks questions, and provides constructive feedback to presenters.	4
Final report I	Introduction (see Section 2.1)	10
	Model description (see Section 2.1)	15
	Model implementation (see Section 2.1)	12
	Model fitting (see Section 2.1)	10
	Results: quality + interpretation (see Section 2.1)	13
	Conclusion (see Section 2.1)	3
	Addresses at least two comments from peers	3
	Well-written report written free of grammatical errors and has reasonably-sized figures.	5
	Only necessary code included in report.	1
Final report II	Original `.Rmd` or `.qmd` file with reproducible code	5
Reflection	See Section 2.3	4
Total		105