Select a research problem and accompanying dataset that interest you. In selecting a dataset, use publicly available data or data that you own (a few websites providing publicly available data are listed below). If you do not personally own the data, be sure to respect copyright laws and do not distribute the data without permission of the owner. If the data are owned by your employer or are otherwise sensitive (e.g., contain protected information), be sure that none of your analyses violate non-disclosure agreements or otherwise compromise anonymity. Responsible, ethical data handling is critical in science, public service, and business.

Identify a few specific questions that you believe the data will help you answer. With these clear objectives, proceed through the statistical modeling process and specifically address the following:

**1. Understand the problem**

- Clearly identify your problem and the specific question you wish to answer.

**2. Plan and properly collect relevant data**

- Justify why the data chosen provide insight to answering your question.
- Describe how the data were collected.
- Describe any challenges relating to data acquisition/preparation (such as missing values, errors, etc.).

**3. Explore data**

- Graphically explore the data using plots that could potentially reveal insight relating to your question.

**4. Postulate a model**

- Justify why the model you chose is appropriate for the type of data you have.
- Describe how the model is well suited to answer your question.
- Identify how inference for parameters in the model will provide evidence relating to your question.
- Write the full hierarchical specification of the model.
- Justify your choice of prior distributions.

**5. Fit the model**

- Fit the model using JAGS and R.

**6. Check the model**

- Assess MCMC convergence. It is not necessary to include trace plots or other diagnostics in the report. Commenting on the results of your diagnostics is sufficient.
- Check that modeling assumptions are met (e.g., residual analyses, predictive performance, etc.).

**7. Iterate if necessary**

- Decide if your model is adequate. Postulate and fit at least one alternative model and assess which is best for answering your question. If neither is adequate, report that and move on.

**8. Use the model**

- Provide relevant posterior summaries.
- Interpret the model results in the context of the problem.
- Use the results to reach a conclusion.
- Acknowledge shortcomings of the model or caveats for your results.

Write a report that does not exceed four pages (including figures and tables). It will be challenging to address all of the items above in so short a space. It is up to you to decide which parts should be emphasized and discussed in detail and which parts should merely be summarized in a sentence or two. Remember that a peer reviewer, after completing this course, should be able to re-create your results if you were to provide the data with your report (which you will not).

In your report, include the following sections.

**Executive summary/abstract**: This should be a paragraph or two summarizing the problem, how you addressed it, your results, and your conclusions. It is called an executive summary because someone who doesn't have enough time or technical expertise to read through the report should be able to read the summary and understand what you did and what conclusions you reached.

**Introduction**: Describe the problem and how you will address it (process step 1).

**Data**: Describe the data and summarize your work for process steps 2 and 3.

**Model**: Completely specify your statistical model and summarize your work for process steps 4 through 7.

**Results**: Summarize the results from the model's fit to the data (process step 8).

**Conclusions**: Summarize your conclusions based on the results (process step 8). This section may be combined with the results section.

Submit your report in PDF format using the file upload below. To receive credit for this assignment, serve as a peer reviewer for two classmates' assignments.

**A few websites with publicly available data:**

https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html

https://www.r-bloggers.com/datasets-to-practice-your-data-mining/

http://www.public.iastate.edu/~hofmann/data_in_r_sortable.html

http://www.stat.ufl.edu/~winner/datasets.html

http://www.math.uah.edu/stat/data/