# Stat 21 Final Project

Proposal due April 1

Project due April 29

The goal of the final project is to apply what you’ve learned in this course to conduct a statistical analysis. It should be an in-depth regression analysis of a question that interests your group. This question may come from one of your other courses, your research interests, your future career interests, etc.

The project has two main deadlines:

* Project proposal: due Fri, April 1 at 11:59pm
* Project write up and presentation: due Fri, April 29 at class time

## Data

It is best to **start with the question of interest** and finding the data second. As you’re looking for data, keep in mind your regression analysis must be done in R Studio. Once you find a data set, you should make sure you are able to load it into R Studio, especially if it is in a format we haven’t used in class before. If you’re having trouble loading your data set into R Studio, ask for help as soon as possible, so you can make any necessary adjustments before the project proposal is due.

In order for you to have the greatest chance of success with this project it is important that you choose a manageable dataset. This means that the data should be readily accessible and large enough that multiple main effects and interactions can be explored for your model. As such, your dataset must have at least 50 observations and at least 5 variables (exceptions can be made but you must speak with me first). The data set should include both numeric and categorical variables.

**Do not reuse datasets used in examples/homework/labs in class.**

## Final Project Components

Your final project grade is broken into 5 parts corresponding to your

* Proposal
* Report
* Presentation
* Poster
* Team peer evaluation

### Proposal

*Due Fri, April 1 by 11:59pm*

This represents a draft of the introduction section of your project report and a draft of the regression analysis plan for your dataset. Your write up and all typesetting must be done using R Markdown. You may use the final project template provided on Moodle if you wish. Make sure the final proposal document is saved as a PDF. Each group member must submit your group's proposal to Gradescope by the deadline.

There are two main purposes of the project proposal:

* To help you think about the project early, so you can get a head start on finding data, reading relevant literature, thinking about the questions you wish to answer, etc.
* To ensure that the data you wish to analyze, methods you plan to use, and the scope of your analysis are feasible and will help you be successful for this project.

#### Section 1: Who?

List all group members and the responsibilities of each group member. Verify that each group member has read these instructions carefully.

#### Section 2: What?

Introduce the research question you wish to explore. Include the motivation behind your research question (citing any relevant literature) and state your hypothesis/hypotheses regarding this question.

In this section, you will describe your data set, including the data source. This section will include

* Description of the observational units
* Description of the response variable and its variable type
* Description of the predictor variables and the population coefficients you wish to understand using statistical inference
* Description of any variables relevant to the analysis

#### Section 3: How?

Outline your analysis steps in this section. (Outline does not need to be in complete sentences.) Plan which steps you will take to answer your research question and make sure these steps include an exploratory analysis of the data and relate to the four stages of model development (choose, fit, assess, use).

#### Section 4 - References

Include the appropriate references for any outside literature or data sources.

#### Proposal Grading

Total 20 pts

Appropriate data 5 pts

Reasonable analysis plan 5 pts

Group rolls clearly considered 5 pts

Document organization and writing 5 pts

#### Late penalty

**Late, but within 24 hours of due date/time:** -20% (only applies to proposal portion)

**Any later:** no credit

### Report

*Due Fri, April 29 at class time*

The goal of the written report is to demonstrate your ability to ask meaningful questions and answer them with the results from regression analysis, that you are proficient in using R, and that you are proficient at interpreting and presenting the results. Focus on methods that help you begin to answer your research questions. **You do not have to apply every statistical procedure we learned.** Also pay attention to your presentation. Neatness, coherency, and clarity will count. Your report will be checked for plagiarism with TurnItIn. If you include any quotes or re-word any points made in other sources, you must cite these sources and reference them in your text.

Make sure the final report is saved as a PDF. Each group member must submit your group's proposal to Gradescope by the deadline. You can add sections as you see fit to the template. At a minimum, your report should have the following sections:

#### Section 1: Introduction

This is basically a revised version of what is in the project proposal. This should include your research question, hypotheses, and a description of the data. It should also include a summary of an exploratory data analysis.

#### Section 2: Regression Analysis

This section includes the results of your final regression model. In addition to displaying the model output, you should include a brief description of why you chose that type of model and any interpretations/ interesting findings from the coefficients. You should also include a discussion of the model assumptions and model fit analysis.

#### Section 3: Discussion & Limitations

This section should include any relevant predictions and/or conclusions drawn from the model. Also critique your own methods and provide suggestions for improving your analysis. Issues pertaining to the reliability and validity of your data and appropriateness of the regression analysis should also be discussed here. A paragraph on what you would do differently if you were able to start over with the project or what you would do next if you were going to continue work on the project should also be included.

#### Section 4: Conclusion

In this section, you should summarize your project and highlight any final points you wish the reader to get from the project.

#### Section 5: Additional Work

This section should include any other models you tried, a check of the assumptions, and a brief explanation of why you didn’t select the model.

Before you finalize your write up, make sure your chunks are turned off by including echo = FALSE in the header of each code chunk. This will hide the R code in the .md file of your final write up.

The main part of the write up (sections 1 - 4) should be no more than 3 pages. The Additional Work section may be up to 5 pages. (Note: These page limits exclude relevant plots and tables.)

#### Late penalty

**Late, but within 24 hours of due date/time:** -20% (only applies to written report portion)

**Any later:** no credit

### Presentation

*Due Fri, April 29 at class time*

All groups will present during class on the last day of the semester. Your presentation must be no longer than **7 minutes** and **each team member should say something substantial**.

Your presentation should not just be an account of everything you tried (“then we did this, then we did this, etc.”), instead it should convey what choices you made, and why, and what you found. You won't be able to describe all the work you did in this presentation, that detail will be available in your written report. In your presentation, highlight the most interesting, big-picture points from your analysis.

There is no late option for your presentation.

### Poster

*Due Fri, April 29 at class time*

You can use any software you want to create your poster. Posters will be presented virtually. Each group member must submit your poster to Gradescope as a PDF file before the start of class.

Your poster is a visual representation of the main messages from your analysis. The writing on your poster is meant to supplement the images and figures. A poster is NOT just a smaller version of your written report. As such, your poster should be visually attractive and readable and must include

* A catchy title;
* Names of each group member;
* The date;
* Font (besides the works cited) no smaller than 24pt;
* A descriptive summary of your data;
* At least 3 images or figures;
* Key ideas and main take-aways.

You may use/adapt any of these free [poster presentation templates](https://www.posterpresentations.com/free-poster-templates.html#postertemplate-40x30) for your final poster.

Read this [short paper](https://drive.google.com/file/d/1FZBV5XonRhM0NjUqkvKE7G9oBVRpzUZy/view?usp=sharing) summarizing how to create and present an academic poster. You may also find this [guide to creating a quality research poster](https://guides.nyu.edu/posters) useful.

There is no late option for submitting your poster.

### Group peer evaluation

*Due Fri, April 29 before 11:59pm*

You will be asked to complete this [evaluation form](https://docs.google.com/document/d/1vA6qlG0MgjDabv3WLz2i82ISkTJs6KI94Zd28LVle4Y/edit?usp=sharing) for each team member's contribution to the project. Completing this evaluation is a prerequisite for getting credit for the team peer evaluation component of your individual project grade. The purpose of this component of your project is to hold group members accountable for a fair share of the work and to provide you with an opportunity to critically reflect on the role everyone plays in producing your final project.

Complete the evaluation based on this [rubric](https://docs.google.com/document/d/1YzYoOSW4TbK7QFgfNAKMpZsaxq1GFl2hVxrF8FDplKc/edit?usp=sharing) and total up the point for each group member. (The self-evaluation component will not impact your evaluation grade, it is merely an exercise to help you self-reflect while you evaluate others.) If you indicate that any individual did less than 4 points worth of work in total, please provide some explanation. If any individual gets an average peer total score less than 2, this person will receive half the grade of the rest of the group. Save the document as a PDF before submitting it to Gradescope.

Provided you submit a group evaluation form, the grade you receive as an individual for this component of your project is a mixture of two, equally weighted factors:

* A grade from me regarding the thoughtfulness and attention put into this evaluation and
* A grade representing the average scores you received from your group mates on their evaluations.

Please note that you will not receive any of the feedback or scores from your group mates directly. I **strongly recommend** agreeing with your group mates to not discuss these evaluations with one another at all and to instead set dates to check in with one another and ensure roles are clearly communicated and mutually agreed upon.

There is no late option for submitting your group peer evaluation.

## Deliverables

Your final project submission must include a zipped folder containing

* An RMarkdown file of your final report (formatted to clearly present all of your code and results);
* Your dataset (in csv or txt format); and
* A PDF version of your poster.

Style and format **does** **count** for this assignment, so please take the time to make sure everything looks good and your data and code are properly formatted.

## Grading

Total 110 pts

Proposal 20 pts

Report 35 pts

Presentation 20 pts

Poster 25 pts

Team peer evaluation 10 pts

The project will be graded based on the following criteria:

* Consistency: Did you clearly answer the question of interest?
* Clarity: Can the audience easily understand your analysis process and any sort of conclusions/arguments you make?
* Relevancy: Did you use the appropriate statistical techniques to address your question? Was your analysis thorough (e.g. did you consider interactions in addition to main effects?)?
* Interest: Did you attempt to answer a challenging and interesting question rather than just calculating a lot of descriptive statistics and simple linear regression models?
* Organization: Is your write up and presentation organized in a way that is neat and clear for the audience to understand?

A general breakdown of scoring is as follows:

* 90%-100%: Outstanding effort. Student understands how to apply all statistical concepts, can put the results into a cogent argument, can identify weaknesses in the argument, and can clearly communicate the results to others.
* 80%-89%: Good effort. Student understands most of the concepts, puts together an adequate argument, identifies some weaknesses of their argument, and communicates most results clearly to others.
* 70%-79%: Passing effort. Student has misunderstanding of concepts in several areas, has some trouble putting results together in a cogent argument, and communication of results is sometimes unclear.
* 60%-69%: Struggling effort. Student is making some effort, but has misunderstanding of many concepts and is unable to put together a cogent argument. Communication of results is unclear.
* Below 60%: Student is not making a sufficient effort.

## Tips

* Review the grading guidelines and ask questions if any of the expectations are unclear.
* Make sure each team member is contributing, both in terms of quality and quantity of contribution (we will be reviewing commits from different team members).
* Set aside time to work together both in the same location and remotely.
* When you’re done, review the .md document on GitHub to make sure you’re happy with the final state of your work.

### Code

In your write up your code should be hidden (echo = FALSE) so that your document is neat and easy to read. However your document should include all your code such that if I re-knit your Rmd file I should be able to obtain the results you presented. Exception: If you want to highlight something specific about a piece of code, you’re welcome to show that portion.

### Teamwork

You are to complete the assignment as a team. All group members are expected to contribute equally to the completion of this assignment and group assessments will be given at its completion - anyone judged to not have sufficiently contributed to the final product will have their grade penalized. While different group members may have different backgrounds and abilities, it is the responsibility of every group member to understand how and why all code and approaches in the assignment works.