NOTE: It is expected that work will be divided fairly among group members. If this is not the case, grades may be reduced for some members.   You should be able to work out minor differences in contributions, but if there is serious inequity, please let me or the TA know, as early as possible, and we can help you work out a working solution. After handing in the project, you will fill out a group evaluation, describing your collaborative process.

Proposal, due on Fri, Oct 8:

I will provide a number of data sets. If you would like to use other data, we can meet to decide if it is appropriate. Once you have your data, decide on some research questions.

Type a short proposal (approximately one full page), explaining your research question, which data you will be using, why your study is interesting, and some general ideas that you would like to analyze. That is, describe some relationships that would like to investigate, and graphs that you plan to make. Suggest some of your hypotheses. For example, you may plan to look at data on films to see if newer films make more money as a percent of budget, than older films.

Upload a (word or pdf) copy of the proposal to the File Exchange of your Project Group site on Blackboard by 11:59pm **on Friday, October 8.**   Sooner is even better, as I can give you feedback, and there is more time to refine your choice.

Draft, due on Wed, Oct 27:

After your proposal has been approved, begin to explore your data set, using R. Following the instructions for Data Analysis and Writeup below, prepare a draft with **all of the parts** noted. The more finished your draft, the better feedback I can give you before you submit the final version.

The draft should be a pdf or word file, and you should upload it to the File Exchange for you group, calling the file DRAFT.

Peer Evaluation, on Tues, Nov 16 and Thur, Nov 18:

Each group will be assigned another group with which to do peer evaluations. By class time on Tues, Nov 16, make your latest draft available to your partner group (you'll probably want to update it after I send you comments). We will use a Microsoft Teams channel for this purpose – the channel called Discussion group x_x. Upload the project to the channel by class time on Tuesday. By Friday at 11:59pm, each member should evaluate the paper of the other group, and upload the evaluation to the same channel, so members of the other group can see them. The evaluation form will be provided, and will contain relevant components of the Project Rubric (see below).

The Data Analysis and Writeup, due on Wed, Dec 1:

Your project should be in a document that you will present to the class. It should be written using complete sentences with correct spelling and grammar. Your presentation will have four parts: An Introduction, the Data Analysis, Conclusions, and Limitations/Recommendations:

I.   **Introduction:** At the beginning of your paper, you must describe the data, in a paragraph. Note the following, and any similar information:

- What is the source of the data?   Where and when was it created?
- If it is a sample, from what population was it drawn, and how was the sample selected?
- Do you suspect any sampling bias?
- Was it an experiment or an observational study?
- How were measurements taken, or questions asked?
- Do you suspect any bias in the questions or measurements?
- Why is this data of interest to you, and why should the class find it interesting?
- What kind of data cleaning was necessary (R code for this must show…)

II.   **Data Analysis:** Write R code to create some relevant graphs, using techniques that we've used in class (ggplot, maybe dplyr). About **4 or 5 graphs** should be plenty, depending on complexity. Include some numerical summaries as well. If possible and appropriate, include **a bootstrap confidence interval.**

For each graph and numerical summary, write a paragraph or two summarizing what you see, and suggesting some implications. For example, describe patterns that you observe in a graph, and suggest why they make sense, given what you know about the subject, or if they are unexpected. Do you think there is a cause-effect relationship between any variables?   Explain your reasoning.

III.   **Conclusions:** Write some overall conclusions – an overall summary of what you learned from your analysis. Summarize in one paragraph.

IV.   **Limitations / Recommendations:** Write a paragraph describing some of the limitations that are inherent in your study. Also discuss ideas for future research that might build on the work you did in this project. Summarize in one paragraph.

When you have everything assembled, put your code and text in an Rmd script that will help you present your results. The R code in your Rmd script **must have comments** that explain parts of the R code. The **presentation created by your Rmd script must show the R code** (e.g., do not use include=FALSE).

Submit your Rmd script by uploading it to your File Exchange on your Project Group site.   Be sure that your data is also on the File Exchange site. **Also upload a pdf or word version** of your project, produced by Rmd.   The project must be uploaded by **11:59pm on Wednesday, December 1.**

Group Evaluation, due by 11:59pm on Friday, Dec 10.

This is a brief summary of how your group worked. It will be completed on Blackboard, after you have submitted and presented your project to the class. A form will show up during the last week in class, in the Project Description tab on Blackboard.

Please read the rubric on the next pages which will tell you the criteria that I will use to grade this project. Feel free to contact me if you have questions.

Rubric for Grading Project:

Criteria	0	1	2	3
Introduction	Data background is not stated.	Data background is not clearly stated and/or is not accurate.	Data background description is reasonably clear.	Data background description is very clear and thorough. In addition, the student gave a compelling reason why the question is interesting.
Data Analysis -- Coding	No R code is included in the presentation.	R code for graphs and data summary has errors, or is not completely present.	R code is present and largely correct. Students made use of dplyr, ggplot, techniques learned in class.	R code is correct, uses techniques learned in class, and is well documented with comments in the Rmd script.
Data Analysis – Graphs	No graphs are included in the presentation.	Some graphs are presented, but there is an insufficient number, or some are incorrect or inappropriate for the particular data.	A sufficient number of graphs are presented, appropriate to the data.	A sufficient number of appropriate graphs are presented. Graphs are interesting, attractive, and easy for the audience to interpret.
Data Analysis – Summary Statistics	No summary statistics are included in the presentation.	Some summary stats are presented, but there is an insufficient number, or some are incorrect or inappropriate for the particular data.	A sufficient number of summary stats are presented, appropriate to the data.	A sufficient number of appropriate summary stats are presented. Stats are interesting, useful, and easy for the audience to interpret.
Data Analysis (Graphs and Numerical Summaries)-- Written Description	There is no written description of the data analysis.	The written description of the data analysis is incorrect or not relevant to answering the research question	The written description of the data analysis is accurate but not complete	The written description of the data analysis is accurate and completely describes the important features of the distribution
Conclusions	There is no written interpretation of the overall project.	There is a written interpretation, but it is incorrect.	Written interpretation is correct, but not clear and/or not in context	Written interpretation is clear and correct and in the context of the research question.
Limitations and Recommendations	There is no discussion of limitations of the project or ideas for future work	There is some discussion of limitations and ideas for future work, but the ideas	Limitations of the study and ideas for future work are described and are generally sensible but are lackluster	Student sensibly describes limitations of the study and has strong suggestions for future work.

		are unclear and/or don't make sense		
Class Presentation	Not there for presentation.	Presentation is completed, but it is incomplete and/or hard to follow.	Presentation is adequate, but not totally clear, somewhat lackluster.	Presentation is thorough, clear, and engaging.
Quality of Writing	Work is not submitted.	Write up does not use complete sentences and/or uses poor spelling and grammar	Write up uses complete sentences but has quite a few spelling and/or grammatical errors	Write up uses complete sentences and has almost no spelling and/or grammatical errors
Quality of Rmd script	Rmd is not submitted.	Rmd is submitted, but does not work, or html format is not correct, or code does not show up.	Rmd script is mostly correct, but there are a few issues with completeness and/or format.	Rmd script is correct, and produces a good-looking html document with all the necessary parts.
Total Score	**Points above:** /30 + **Proposal and Draft:** /4 + **Peer Evaluation of other group's project:** /4 + **Completing a group evaluation afterwards:** /2 = /40 points total			