

STA 571 Course Project Timeline

This document presents a rough outline to the projects conducted in STA 570: Statistical Methods II. The outline gives when certain topics should be discussed, when decisions should be made on moving forward with projects, and what students need to be focused on. Proposal and final project rubrics will be provided in additional PDFs. As always, there is flexibility based on the way you run your course.

Weeks 1 – 4 (Project Feasibility)

Leading up to Exam 1, decisions will be made if the projects can be conducted effectively within the course. This requires students to determine what data set they want to analyze and what methods they think they would be interested in using. For non Math/Stat students, the idea is to allow them to use their own data (likely data they have collected) to advance the project. Students should be engaged in looking into a source of data and determining what type of analysis can be completed. This can be motivated through **wouldn't it be great if ...** statements, or WIBGIs.

The first four weeks should include students trying to create as many WIBGIs as possible. At the end of each homework assignment, I added an additional *optional* section entitled "Project Development" where I reminded the students they need to produce at minimum 3 WIBGI statements. From my assignment sheet:

Write out 3 - 5 ideas for a project to be done this semester. Think about what data you may have available or what data you may be interested in analyzing. If possible try to write a statement of this form:

Wouldn't it be great if ***what you want to analyze*** could be used to ***solve a problem using statistics/data science***.

Example: Wouldn't it be great if plasma thermogram data could be used to classify lupus using logistic regression.

The main concept here is to encourage exploration and broad thinking. It is okay if students write statements that are not feasible. It is also okay if they write statements that don't conform to the example statement. The goal here is creativity. The idea is to begin generating ideas and determining sources of data; we are trying to encourage data exploration and thinking outside the box from their core studies. I would also suggest that ALL students check www.kaggle.com for additional ideas. It must be made very clear to students that the data needs to be available, and while I did encourage students to try and get their data together, by week 5, if they don't have the data, they should use Kaggle.

I would have the students start a Word/PDF document where they write the WIBGI statements. They should write as many as they can possibly think of. I did not do a great job of collecting the WIBGIs, but maybe at the end of week 4, have them all submit the Word/PDF with their WIBGIs on it, so that we can evaluate what types of projects ideas are produced contrasting Fall/Spring semesters.

Week 5 (Decision on Project)

Around the time of the first exam, I had students choose one WIBGI to move forward with as a project. You could have them formally expand a WIBGI into a workable idea. The students were then required to give a 30s – 1min presentation, in class, informally, about their project idea. If they did not have a WIBGI they were interested in, they were asked to present a topic of interest, which they should then follow up with from Kaggle. The presentations allowed me to gauge enthusiasm for the project as well as feasibility. When my students presented, it was clear that there was overwhelming interest and really great project ideas. During Exam 1, I announced we would move forward with project. This would also be the time you could cancel the project, and announce that a Final Exam will be given instead.

Weeks 6 – 9 (Data Cleaning and Visualization)

This period of time should be use begin understanding their data sets, cleaning, visualizing, and ensuring they have a data set that can be used to move forward with. This was also the first period where I began having out-of-class meetings with students discussing project ideas. Begin helping them understand what they will need to use, if the data set is useful, or if further changes need to be made.

At this stage, I also had students submit images/graphics/visualization of their data so they could show me they were making progress. Many presented boxplots or other simple graphics. A VERY IMPORTANT step was making them discuss every variable within the data set. I had most prepare tables giving a description of each variable (name, description, variable type, summary, ...). I believe pressuring them to understand their variables is important at this stage, and will give you a better understanding of what they are trying to accomplish.

Weeks 10 – 11 (Proposals)

Data should be ready, cleaned, and visualized. I gave them 2 weeks to write a proposal. Proposal rubric can be sent in an additional PDF. The main ideas for the proposals: 1) Introduction – what are we analyzing and why is it important (not what math/stats to use, but what problem to be analyzed; let them feel like experts in their field of study, or make a deeper dive into the kaggle project chosen), 2) Data/Materials – what data set is being used, present a clean version of data, present visualization, present variable discussion, 3) Methods – what do they PROPOSE to use? No analysis necessary, maybe something preliminary 4) Project Goals – what do they hope to accomplish and can they do it within the required timeline.

I have a rubric for this step. I gave about 2 weeks for submission. A limit of 2 pages with additional appendices allowed was set. This should be brief but informative and when done correctly, you will see they have a solid grasp of where they are going.

Week 12 (Graded Proposals and Feedback)

Grading them in a timely fashion was critical for me. You could shorten the period they are visualizing data to give yourself more time to grade. Big here was making decisions on if the projects are going well. I had to redirect maybe 2 – 3 students to entirely different projects. Several just needed slightly new directions but the ideas were there. This is the last big step before they are going to begin working on models and analysis. I did hand-written notes on the submitted proposals and required students to bring my notes to any office hour/meetings about the projects. Be prepared for a lot of office hour visits. If the proposals are strong, they are moving in the right direction. Be VERY CRITICAL of the data presented, and if it doesn't seem strong, knock scores down. They have to have data ready and useable at this stage. They have had 12 weeks to work on it!!!

Weeks 13 – 16 (Final Projects)

With graded proposals in hand, I gave them about 1 month to write the project. I have a formal rubric for this as well. I met with students frequently. Due the day of the final exam. Page limit was 10 pages, but I made them produce graphics and submit all R code as well, most of which could be placed in appendices. Most students used all 10 pages + another 10 pages of appendices. This is also a good learning experience for them in formal writing. Most of the final projects looked like dissertations, short of a table of contents. Everything should be labeled, neat, clear, and easy to follow. All graphics must have legends and explanations. There are points for clarity and topic, as well as if they meet the goals of the course. This is a big challenge for some, but they need it!

The work I saw during this period was remarkable, and students really felt like they were learning how to apply statistical methods to their own laboratory experiments! I would say 80% or more of the projects successfully answered a question of interest. If I had pushed more on having data available and the use of Kaggle, I think we will see even more really interesting projects that are helping students feel they can use these tools outside the classroom.