

STAT 6130 Applied Multivariate Statistics Spring 2020

Class Project Proposal

Purpose of Proposal and Evaluation Criteria:

The major purpose of this proposal is simple: to make sure that you start thinking about the final project at the early stage of this semester.

Specifically, at this stage, you need to: a) Choose a suitable data set for the final project (you are free to switch to another data set if you change your idea later); b) Conduct some basic exploratory analysis on the data set.

The proposals will be graded solely based on whether you have followed the basic requirements (listed below). Moreover, I will also try to give you some feedbacks based on the proposals.

Submission Guideline:

The proposal is due on **March 3, 17:00pm**. You are expected to submit a pdf version of proposal to **my email address** as well as a hard copy to **my personal mailbox** (not the class mailbox).

For the email submission, please stick to the following rules:

- a) The subject of your email should be the following :

[Stat 6130] Project Proposal, Name 1, Name 2 (if applicable).

- b) The name of pdf file should be named based on the following format:

S6130_Proposal_Name1_Name2.pdf

Make sure to put your first name before your last name. For instance, if I worked with Mr. John Smith, I would name my proposal as
“S6130_Proposal_ChaoDu_JohnSmith.pdf”

- c) If you form a team with another classmate, only one of you need to send the pdf version to me. **Make sure to cc another team member in your email.**

- d) Upon receiving your email, I will reply to you to confirm your submission. If you did not receive my confirmation within 24 hours, please let me know.

Proposal Requirement:

Your proposal should include three major parts: **main text, tables and figures, appendix.**

Main Text:

- 1) Format Requirement: **2 pages limit (single-spaced, 12 pt, 1 inch margins)**, you should not include any tables and figures in the main text (which should be placed after the main text). Also, make sure to include your names and computing ids at the beginning of the main text.

- 2) Content Requirement: The main text should contain at least three sections:

- a) **Background.** You are expected to provide some background information on your data. For instance, you may try to answer the following questions:

Where do you obtain your data? Who collected this data in the first place? How was this data collected (i.e. from sampling survey or from experiments)? What was the main purpose of collecting this data? Does this data provide us opportunity to study some major social or scientific questions? Is it possible to situate this data within a big picture?

Make sure to cite the source of your data.

- b) **Summary of Data.** You are expected to present a detailed summary of the key features of the data in this section. For instance, you may try to address the following questions:

How many observations/variables are there in your data? Are there any missing or unreliable values? What are the meanings of the key variables in the data set? Are there any patterns/trends that can be revealed by simple exploratory analysis? Can you draw some interesting conclusions based on your observation?

If possible, support your summary with appropriate tables and/or figures.

Keep in mind that you do not need and should not attempt to summarize every aspects of the data. Stick to the major features of the data and highlight the points that you believe to be important/interesting.

- c) **Potential Problem.** You should try to think about the potential problems you plan to investigate using this data. You do not need to discuss the detailed method you plan to use to address your problems.

Tables and figures

- 1) Format Requirement: All the tables/figures should be numbered in the order they appear in the main text. Proper captions should be used so that the reader can easily understand the information contained in the tables/figures.
- 2) Content Requirement: You may include any number of tables or figures in this section. However, **every table/figure must be relevant to the main text**. You will lose points if you include unnecessary tables/figures.

Appendix:

- 1) Format Requirement: You should make sure that this part is formatted so that it is easy for your reader to read. Otherwise there is no special requirement.
- 2) Content Requirement: You are expected to explain how you produce the figures/tables used in your proposal. **The information you provide in this section should be sufficient enough so that other people can easily reproduce your work if they have your data.** If you are using R/Python, it is sufficient to include the codes you use to produce all the figures/tables. If you are using other software package, you need to provide sufficient description on how your figures/tables can be reproduced.