FINAL PROJECT PROPOSAL
DUE WEDNESDAY NOVEMBER 17TH, 2021

JUSTIN M. CURRY

1. Instructions for Turning in the Project Proposal

- Before class on **Monday, November 15th, 2021**, you must have discussed with your team a potential data set along with ideas of what methods or questions you want to address.
- Before class on **Wednesday, November 17th, 2021**, a representative from your team must email `jmcurry@albany.edu` and `rlcardona@albany.edu` and CC your entire team. You must also turn in a **PDF** LaTeX-ed[1] project proposal that addresses the questions below. The project proposal should be one to three paragraphs and between a half-page to one-page in length. ***Don't just answer the questions line-by-line below.*** Try to get the reader's attention as if this were the start of a Medium data science article, or a grant proposal for federal funding or an attempt to attract venture capital funding.
- Final presentations will be on **Monday December 6th**, the last day of classes.
- Writeups will be due on **Wednesday December 15th**, with no exceptions.

2. Project Proposal Questions (10 HW points)

Your proposal will be graded on the following rubric, which counts as homework.

Everyone on your team will receive the same HW grade grade for the proposal.

(1) (2 points) List the complete names of your data science team.
(2) (4 points) Describe a data set that you want to investigate. Note that you **cannot use any of the following three data sets**:
   - the iris plants data set,
   - the wine data set,
   - the MNIST data set.

   Your data set should have **four or more *independent* features** and **at least 20 samples**. Independence in this case means not 100% correlated. Make sure you address the following questions:
   - Why do you think this data set is interesting? What features does it have?
   - What do you hope to find or achieve?

(3) (4 points) Every group must apply at least three to five (one per person) different data science methods from at least two of the following data science method categories:
   - Classification

Date: November 8, 2021.

[1]It MUST BE IN LaTeX OR IT WILL BE RETURNED WITHOUT REVIEW.

- Clustering
- Regression
- Dimensionality Reduction

Pick four methods (or one per team member) now and identify which categories above they fall into. Your methods can change before submitting your final project at the end of the semester.

3. More Details on the Final Project and Presentation

Recall that 30% of your individual overall course grade depends on this final project and presentation (FP&P grade). However, since this final portion requires that you work as a team, we have devised the following rubric that reflects individual and group effort.

- **(60% of FP&P grade)** You will turn in only one 4 to 8+ page LaTeX-ed written report, including figures and references, as a team. **Whatever grade the written report get, each person in the group gets that same 60% of their FP&P grade.** Your report **must include** a final section that indicates what each team member contributed to the work in the paper. Sometimes this includes literature review (looking up published articles that considered your data set or a similar problem), ideas, coding, plotting, writing and editing, and so on.
- **(40% of FP&P grade)** Your group will give a $\sim$ 20 minute presentation as a team during the last day or two of classes. **Every person in the group must have some speaking role during their presentation and will get an individual score for how well they did as during their turn.** I encourage you all to rehearse several times before the final presentation.