Data Science 2016 - Choose Your Own Adventure!
Project Proposal and Learning Goals

1. **What dataset you are planning to work with**
   San Francisco Crime Classification

2. **Who your partner is**
   Emily Wang, Filippos Lymperopoulos

3. **For each group member: what are you hoping to learn from this project? These goals could be specific data manipulation skills, software engineering skills, or communication skills.**
   The list below came from a collaborative goal-setting discussion. We aren't committed to doing ALL of the suggestions listed below, and may devote relatively more time to some ideas more than others.
   We will reassess our goal progress and scoping in about a week (Tuesday February 9th), and update our goals/plans accordingly.

   1. Visualizations
      a. map-based; spatial data
      b. emphasis on journalism/communication with external audience
      c. additional D3 generated graphs for documentation
   2. Predicting location instead of crime type??
   3. Pattern Detection over time
      a. Time-series predictions
   4. Algorithm design and implementation
      a. Building algorithms from the ground up, design informed by our exploration/context
      b. Creative ways of ensembling existing models
   5. Publish our results somewhere
   6. Sharing of resources found and draw inspiration from external sources (something we do at the start of each meeting)
   7. Mastery of tools:
      a. pandas,
      b. seaborn,
      c. Are there any map-based vis libraries? We're comfortable in both python and javascript.
      d. scikit learn + our own ensembling
         i. or vice versa, our own ML algorithms + scikit learn's ensembling
      e. MAYBE:
         i. D3,
         ii. SQL databases if we need the "heavy artillery"

4. **How do you foresee this project fulfilling the learning goals identified above. Be specific. If there are adjustments that could be made to bring things into better alignment, let me know (I won't guarantee that I will allow it, but it's worth a discussion).**
   As mentioned in #3, we will reassess our goal progress and scoping in about a week (Tuesday February 9th), and update our goals/plans accordingly.

    1. During team meetings, Emily and Filippos will critique each other's data visualizations in their current ipython notebooks. We'll put this critique on the agenda and write up the takeaways in a concise list that we can use for future reference. This will motivate us to be creative with data visualizations and be diligent about making these figures easy to understand and well-labeled.
    2. As a stretch goal, we are interested to see whether we can develop a model that will predict location instead of just crime type. We will try to see the feasibility of such a goal in the early stages of the project.
    3. Studying the dataset and exploring trends and patterns as they have evolved through time is something we want to look into and believe that [these visualizations] can be easily read by an external audience, indirectly contributing to goal #5.
    4. Opening the black box of ML algorithms? Only as desired; it's not a major focus of this process.
    5. We are very interested in publishing our work, struggles and successes throughout the project. We will revisit this idea midway through our work and aim for an ecosystem where this story can be posted and discussed upon. Examples include Kaggle Scripts, Medium, dataisbeautiful reddit, and so on.
    6. Before every scheduled meeting, we will have read some external blog post, paper, or script for inspiration and team discussion during the meeting. Depending on the content, we may or may not implement these external ideas into our work.
    7. The project is going to help us master tools alluded to at (7). Exploration of D3 and SQL databases are optional and will depend on the progress of the project.