# Project: Exploring and Visualizing Data with R

## Objective:

The objective of this assignment is for you to practice your data manipulation, analysis, and visualization skills using R by working with a dataset of your choice. This project should be completed in teams of one, two, or three individuals (Submit teams [here](https://forms.office.com/r/htEgfFU2Lv)).

Description:

In this assignment, you will find a dataset and answer a question using the dataset. You will need to perform various data manipulation and analysis techniques using R to answer question and create at least two data visualizations to communicate your findings. You will also be required to provide a R markdown notebook that contains your code and visualizations.

Dataset:

You will need to find and choose your own dataset(s) to work with. You can find datasets from various sources, including Kaggle, UCI Machine Learning Repository, data.gov, or any other reputable data sources. You should choose a dataset that is interesting to you and that has enough data to enable you to answer your question(s) effectively.

## Deliverables:

The students will be required to submit a public GitHub repository that includes the following:

1. A README file with a brief introduction to the dataset, description of the primary question being explored, citations of all data and other resources you used or referenced, and any instructions required to reproduce your analysis.
2. All data, code, and other files required to reproduce the results of your analysis.
3. An R markdown notebook file that:
   * Introduces the dataset and the primary question you’re interested in answering.
   * Explains the steps you took to clean and manipulate your data to answer your primary question. Be sure to provide justifications if you’ve removed or adjusted data in any way.
   * Answers the question you were exploring. Be sure to include at least two data visualizations that support your answer.
   * Discusses the implications of your work (e.g., additional questions worth exploring, validity of your analysis).
   * Provides citations to your data and other resources you used.
4. Link to a two-minute (max) presentation video (inside README file) summarizing your analysis.
5. Peer review of all other project submissions.

You must submit the link to your GitHub repository to this [form](https://forms.office.com/r/1NiYDN3dGD) by 11:59 AM on Wednesday 5/24. During class on Thursday 5/25, you will review all projects that’s been submitted by your peers.

## Resources:

Albert will be available during Office Hours (i.e., 5/11 and 5/18 sessions) to answer your questions related to the project. In addition, you can book consultation appointments with him through [tiny.ucsf.edu/AlbertLeeDSI](http://tiny.ucsf.edu/AlbertLeeDSI) (one-week advanced notice required).