# Final Project Guideline

1. Summary of the project:

The purpose of the project is to find the significant predictors that explain one's happiness (level), and model their relationship.

2. Data description

We use "General Social Survey 2016" data for the analysis. (http://gss.norc.org) These are the data on contemporary American society in order to monitor and explain trends and constants in attitudes, behaviors, and attributes. The data contains a standard core of demographic, behavioral, and attitudinal questions, plus topics of special interest. Among the topics covered are civil liberties, crime and violence, intergroup tolerance, morality, national spending priorities, psychological well-being, social mobility, and stress and traumatic events.
For the project, 13 variables are selected including the happiness level. Detailed information about the variables is provided in the separate codebook.

3. Submission due and format

- You are allowed to submit the project for two times: one as a draft version by March 16th, and the other as a final version by March 23rd 11:59pm. The first submission is optional, but you will get an opportunity to get feedback for the draft version and reflect the feedback to your final report.
- You need to submit two files: 1) final report (in pdf) and 2) R program file (either in R or Rmd).

4. Grading guideline

- Completeness of the analysis (40%)
- Interpretation (25%)
- R implementation (20%)
- Prediction Power (15%) : There are two data sets from GSS survey; 1) training data ("Happiness.txt : posted on CCLE and will be used to find the best model), and 2) test data (only the professor has it). You will find the "good" model to predict one's happiness level using some or all of the given predictors in the training data set. Then, we will predict the happiness level for the test data and you will get a score for "prediction power" based on how well you have predicted the happiness level relative to their true values, i.e. RSS of the model.

4. Final report formation

- Length : Maximum 6 pages including plots and tables. Single space with 11pt. You may put additional plots in Appendix.
- Structure :
    - Introduction
        - State your general interest and motivation for the data set
        - Discuss the general expectation and plan for the project
    - Methods and procedures
        - Explain and justify the method you choose to model the relationship
        - If transformation, variable selections, case omission have been done throughout the

process, show the reasoning and necessity of the process.

- Results and interpretation
  - Report the R result and interpret the results in the context of the study.
  - Suggest two or three models (candidates) you have tried.
  - Determine the "best" predictive model and justify your choice.
- Discussion
  - Summarize your project.
  - Discuss if your final model makes sense in real world situation. If you can find any literature or articles that can support your result, then that can be plus.
  - Discuss the limitation (if any) of the analysis, and suggest how it can be improved in the future.

5. Additional guideline

- You may consider the happiness level as a continuous variable.
- You are allowed to transform the variables in different units.
- You are allowed to use only some of the variables for the predictors in your model.
- You are allowed to remove some observations with reasonable evidence.
- You are allowed to change the variable type (numerical -> categorical, categorical with 3 levels -> categorical with 2 levels) when it makes your model more valid.
- You may ask any theoretical/technical questions to professor/TAs, but may not ask for reviewing your results before the first due date. Show your independent work.
- You are not allowed to use any libraries or R modeling functions that have not been discussed in class.
- **Note: You might have low R-square, but it is not surprising for the social/behavioral data set.**