- **Task** Choose an existing datasets or create your own data, carry out exploratory data analyses and regression analyses to explain the relationships among the variables involved.
- **Team members**: For this project you may choose to work with 1-2 persons and submit a joint project. If you cannot find a team member, I will assign a teammate to you.
- **Opt out deadline**: 03/08/2017. (If you feel your teammates are being incorporative, you can choose to start your own team as one individual team. However, you CANNOT switch to other teams. You can work on the same project you has been worked on but you are expected to write your own reports and make your own presentation. You must choose to do so voluntarily. Any team CANNOT force a team member out.)
- **Grading policies**: team members will receive the same grade for the project and it is up to you to make sure that the work is shared equitably. The total points of the project is 100 points, which can be divided into three parts:
  1. Initial report (20 pts): due by 03/08/2017;
  2. Presentation (20 pts): each team will give a 30 minutes presentation of the project; (dates to be assigned)
  3. Final report (60 pts): due by the end of the final exam date (TBA);
- **General guidelines of the project**
  1. Identify the problem of interest: choose a data set, describe the data set and identify the problem you are interested in;
  2. Perform preliminary studies of the data: data visualization; check model assumptions, etc
  3. Select most promising predictors: what variables are potentially most useful for your problem;
  4. Choose the best regression model that serves your purpose and justify your choice (e.g. model diagnostics, outliers, normality, model selection criterion, etc…)
  5. Interpret the final regression model;
  6. Discuss your findings: What do the results mean?
  7. Put all your codes in the appendix.
- **Data sets**
  - Find your own data set online (e.g. google “regression data set”), you will find plenty;
  - UCI Machine Learning Repository
  - Kaggle Machine Learning Competition
  - Academic Torrents: a large date sharing website
  - MLcomp
  - MLdata
  - Google Public Datasets
  - A much more complete list