# Final Project Rubric

## PPOL670 – Introduction to Data Science

### Spring 2020

**Student**:_____

**Project Name**: _____

**Total Score**: _____ / 48

**Project Materials**

- Report was posted on Canvas on time? (1 point)
    - `.Rmd` file and all figures were included with the submission?
- Report was generated using `R`? (1 point)
- student's entire project was version controlled on Github:
    - Does the `git` record extend back to when (at least) the proposal was due? (1 point)
    - Is all code for the project contained within? (1 point)
    - Were the materials organized as an `R` Project? (1 point)
- Student provided access to the data (1 point)
    - Via Dropbox link (if greater than a Gigabyte)
    - Via Canvas (if less than a Gigabyte)

**Document Presentation**

- **Student used professional looking visualizations in the report:**
    - Figures were easy to read and understand? (1 point)
    - Figures made sense within the context of the report? (1 point)
    - Student described the purpose and the insight drawn from the figure in the text? (1 point)
    - Color scheme made sense; easy to differentiate between colored items (1 point)
    - Figures were appropriately proportioned (1 point)
- **Student used `R` Markdown for a professional looking report:**
    - Report was rendered without errors or warnings. (1 point)
    - No `R` code was visible in the report. (1 point)
    - Report contained a table of contents. (1 point)
    - Report contained no (or few) gramatical/spelling errors. (1 point)
    - Report reads as a single cohesive document. (1 point)
    - Report is 12 pages in length (double-spaced; 12 pt font). (1 point)
    - Student cited academic, data, and package sources. (1 point)

∗ To cite a package, use `citation("package_name")` to get a the citation information for a package, e.g. `citation("ggplot2")` will yield "*H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.*"

- [*BONUS*] **Student used professional looking tables:** (+ 2 points)
  - When presenting data and/or figures, student formatted the data as a clean table (i.e. latex)?
    ∗ see `stargazer` package or `gt`
  - Table made sense within the context of the report?
  - Table was clear and easy to read.
  - Table was concise and did not contain unnessary information.

**Content**

The student's project sufficiently addressed these general areas.

- **Introduction** (5 point)

  - Student clearly established the aim of the project.

  - Student offered a clear roadmap of the report (i.e what is covered in the report).

- **Problem Statement and Background** (5 point)

  - Student offered a clear and complete statement of the problem and/or aim of their analysis.

  - Student included a brief summary of any related work (i.e. a *light* literature review)

- **Data** (5 point)

  - Student outlined where their data came from.

  - Student clearly specified:

    ∗ the unit of observation;
    ∗ variables of interest;
    ∗ potential issues in the data (e.g. missingness, coverage, etc.)

  - Student articulate the steps they took to wrangle the data.

- **Analysis** (5 point)

  - Student described the methods/tools they explored in their project.
    ∗ Justified the tools/methods that they used.
    ∗ Adequately described what the tools/methods are doing.
    ∗ Note: "Assume the reader is smart but doesn't know `R`/Machine Learning well. That is, be crystal clear about what you're doing and why."

- **Results** (5 point)

  - Student gave a detailed summary of their results.

  - Student presented their results clearly and concisely.

  - Student used visualizations (and tables) whenever possible/appropriate.

- **Discussion** (5 point)

    - Student spoke on the "success" of their project (as defined in their proposal).

        * "Did you achieve what you set out to do? If not why?"

    - Student outlined the tools/methods they considered but ultimately did *not* use in their final analysis.

    - Student articulate how they would expand the analysis if given more time.