MSDA 607 - Project 3

Project 3 is due end of day on Sunday October 25th. This is a project for the entire class to work on together, since being able to work effectively on a virtual team is a key "soft skill" for data scientists. Please note especially the requirement about making a presentation during the class office hours on Thursday October 29th.

●

> Every year at the Academy Awards the most notable prize is for "Best Picture." The media speculate on it for weeks prior to the broadcast, and most viewers stay up well past their bedtimes to see it awarded. There is a far less hyped award on the night: the one for film editing. Let's face it: most viewers flip the channel or go into the kitchen to refill their popcorn bowl when the winner of "Best Film Editing" is announced. Yet what most people don't know is that the two awards are highly correlated: since 1981 not a single film has won Best Picture without at least being nominated for Film Editing. In fact, in about two-thirds of the cases the movie nominated for Film Editing has gone on to win Best Picture.

Source: "Greg McKeown, Essentialism: The Disciplined Pursuit of Less"

W. Edwards Deming said, "In God we trust, all others must bring data." **Please use data to determine whether Best Film Editing is the best predictor of Best Picture.** Consider your work as an exploration; there is not necessarily a "right answer."

Grading rubric:

- You will need to determine what tool(s) you'll use as a group to effectively collaborate, share code and any project documentation (such as motivation, approach, findings).
- You will have to determine what data to collect, where the data can be found, and how to load it.
- The data that you decide to collect should reside in a PostgreSQL database, in a set of normalized tables.
- You should perform any needed tidying, transformations, and exploratory data analysis in R.
- Your deliverable should include all code, results, and documentation of your motivation, approach, and findings.
- As a group, you should appoint (at least) three people to lead parts of the presentation on Tuesday March 31st.
- While you are strongly encouraged (and will hopefully find it fun) to try out statistics and data models to determine the validity of statement that Best Film Editing is the best predictor of Best Picture, ***your grade will not be affected by the statistical analysis and modeling performed*** (since this is a semester one course on Data Acquisition and Management).
- Every student must be prepared to explain how the data was collected, loaded, transformed, tidied, and analyzed for outliers, etc. in the 10/29 office hours. This is the only way I'll have to determine that everyone actively participated in the process, so you need to hold yourself responsible for understanding what your class-size team did! If you are unable to attend the meet up, then you need to either present to me one-on-one *before* the 10/29 office hours, or post a 3 to 5 minute video (e.g. on YouTube) explaining the process. Individual students will not be responsible for explaining any forays into statistical analysis, modeling, data mining, regression, decision trees, etc.

You are encouraged to start early, ask many questions, actively post on the provided Project 3 discussion forum, etc.