Zach Gambill's Oscar Nomination Prediction Algorithm

For my Luther project, I've been working on predicting the number of Oscar nominations a film will receive, based on a variety of factors. For my first pass, I decided to go simple and just run all of the possible variables I collected, which is what I have created for my MVP. With this, I can then go through and select which variables have the most substantial t-statistics (i.e. -2.5+ and 2.5+) and make are the most correlated with Oscar nominations. As I said, this is the first run, so the model is far from perfect, but with an R-squared and Adjusted R-squared of 0.532 and 0.440 respectively, it's a solid start. The features in this model are categorical dummy variables for all genres, production studios, MPAA ratings, and the month of release, as well as the film budget, release year, gross domestic revenue and runtime (in minutes). While the gross domestic revenue is a lifetime statistic, and thus is less accurate for more recent movie releases, it is a placeholder until I have the time (if I have the time) due to its accessibility in my scraping procedure. It's not perfect, but there is definitely some correlation. The chart below is a residual chart for my initial model. While it is visibly linear, there is a promising amount of apparent randomization/scatter. As I said, this is a very early first model, so it is somewhat promising. And no the graph doesn't have labels or a title, but it's an MVP so I'm working on it.