EDA Homework 6

Not due, just for practice.

The file movie_budgets.txt contains data on the budgets of 5,183 movies from 1906 to 2005, along with their lengths in minutes. Read your data into R as a data frame called movie_budgets. We wish to study log10(budget) as the response variable and year and length as explanatories. Note that these movies are not a representative sample of all movies, so we're not trying to generalize, only describe the data we have.

- Using loess or otherwise, fit a model to predict log10(budget) from year and length. For simplicity, do not transform year and length (even though a transformation of length would probably be sensible.) You will have to make a number of modeling choices:
  - Should you fit a linear or curved function for year?
  - Should you fit a linear or curved function for length?
  - Do you need an interaction between year and length?
  - What span should you use in your loess smoother?
  - Should you fit using least squares or a robust fit?

  Some of these choices are clear-cut, while others will be a matter of preference. Either way, you must justify all your choices.
- Draw one set of faceted plots (one coplot) to display the fit – either condition on year or length, whichever seems to you to be more interesting. Choose a sensible number of panels. Briefly describe what this set of plots shows you.