STAT – S670: Exploratory Data Analysis
Assignment 4

Author: Sricharraan Ramaswamy (UID: 2000855651)

Question set 1:

The lines of R Code used to produce loess model:

```
movie_budgets$log_budget = log10(movie_budgets$budget)
movie_budgets_lo = loess((log_budget) ~ length*year, data = movie_budgets, span = 0.35, family = 'symmetric', degree = 2)

prediction_grid = data.frame(expand.grid(year = seq(1906, 2005, 10), length = seq(1, 390, 1), budget =seq(2,10,1)))

pred_movie_budgets <- augment(movie_budgets_lo, newdata = prediction_grid)
```

Should you fit a linear or curved function for year?

After using both a loess and a linear fit to determine the relationship between budget and year. After plotting, it was clear that both plots fit similarly. After plotting the residual plots for both models, we discovered that the points in both residual plots were scattered around zero. For log Budget vs Year, Therefore Loess is the option to go.

Should you fit a linear or curved function for length?

When we plot the coplot of budget over length condition on year we can see that the loess model fits better than the linear model as the model changes according to the year bins.

Do you need an interaction between year and length?

Yes, the interaction between year and length is needed as we plot the graph for budget over length condition on year we can see that the slope of loess changes in every year bin and similarly, when plotted for budget over year condition on length we can see that the slope of loess changes here also in every length bin.

What span should you use in your loess smoother?

After using different spans to plot the loess model the one which was neither overfitting nor underfitting is 0.35 and the results are as expected as one can see in the graph below.

Should you fit using least squares or a robust fit?

As we can clearly see there are many outliers in the plot and with this we can conclude that a robust fit that is **symmetric** fit is the right way to go about it.

Question set 2:

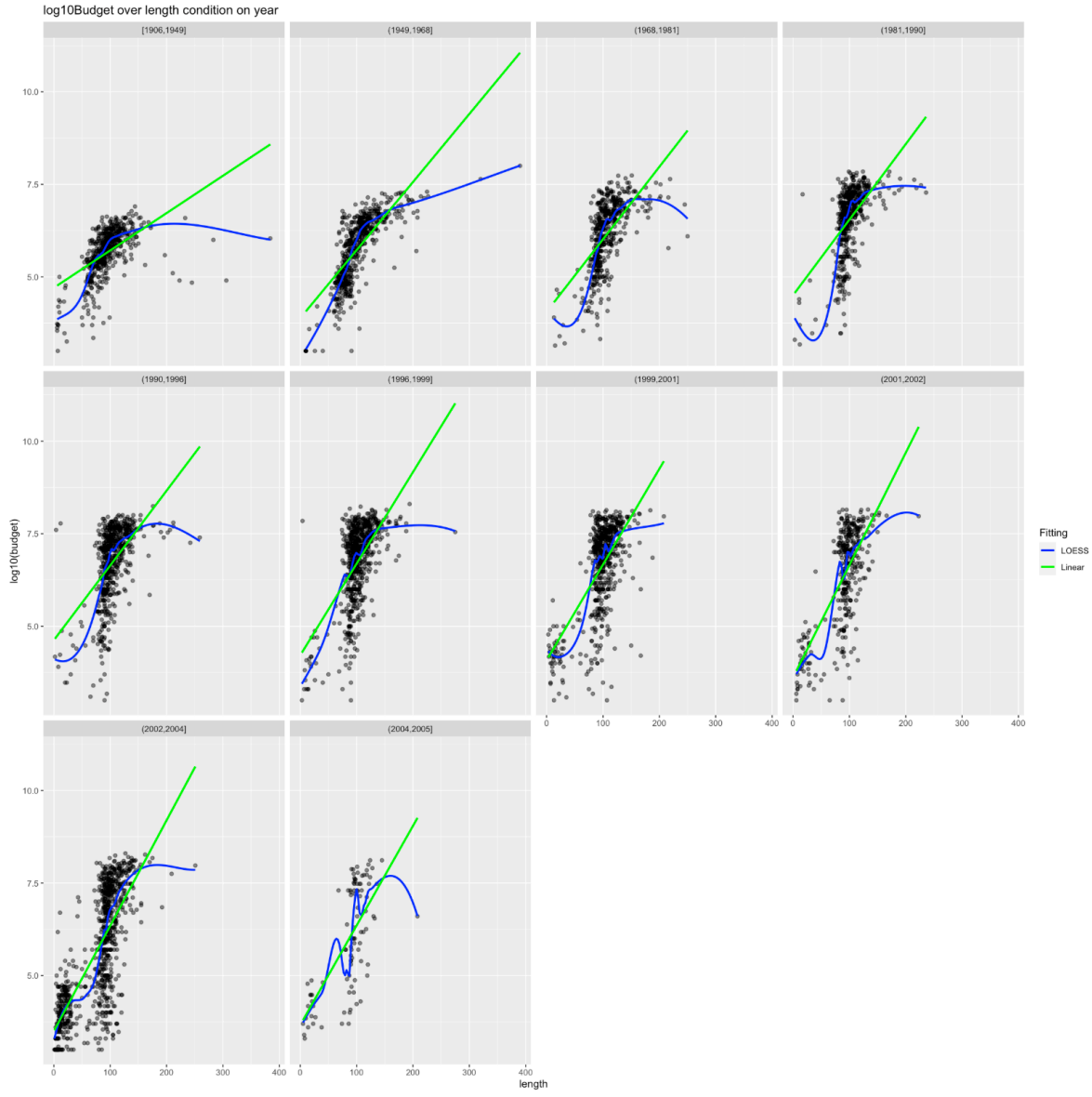

Fig. 1: Log10Budget over length of the movie condition on the year

As seen in the Fig. 1, we can the log10budget over length of the movie condition of the year and as discussed earlier we can see that the loess model gives fit properly compared to the linear model and the model is different for different year bins. The legend clearly depicts which is the linear model and which is the loess model.

Question set 3:

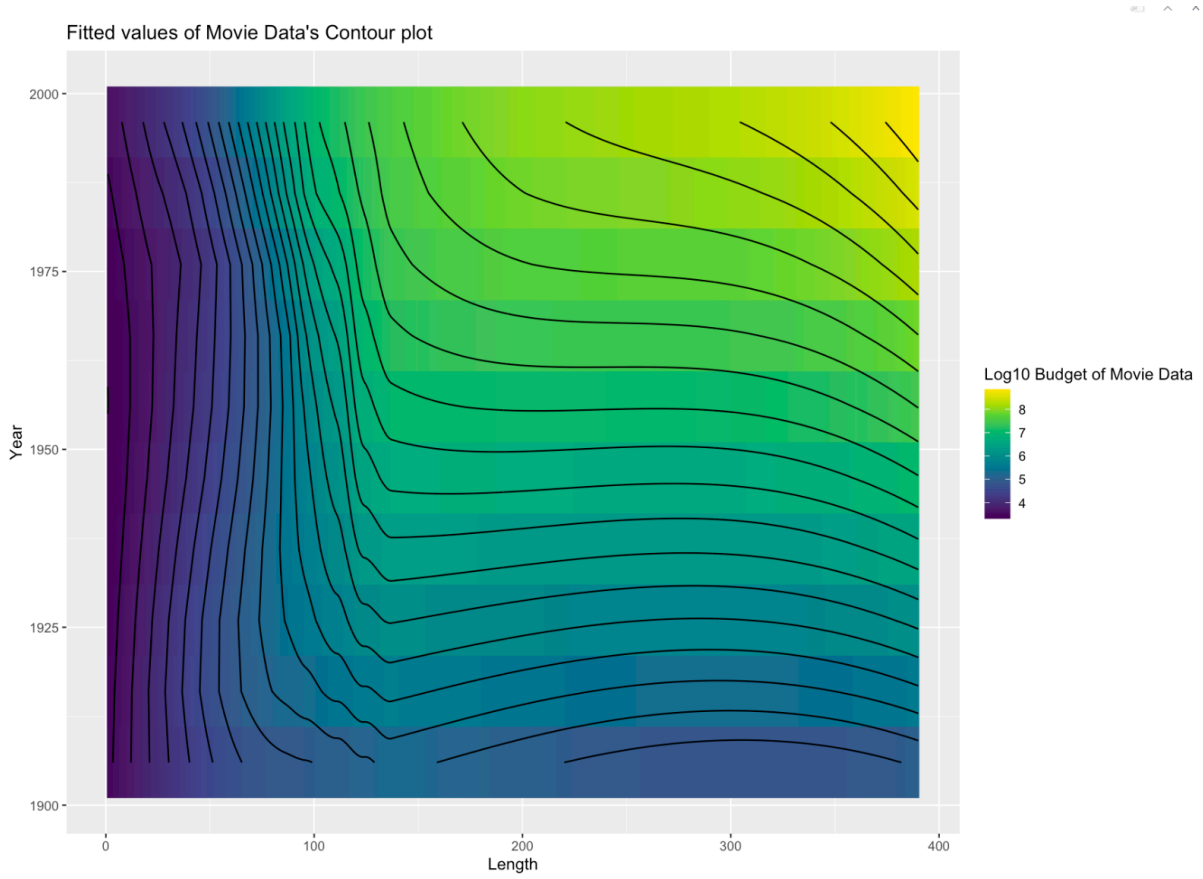

Fig. 2: Contour Plot of Movie Data

As seen in Fig. 2, there are closely spaced contours indicating a large slope around the 100-length and 2000-year marks, while evenly spaced lines indicate a plane, which is essentially what we saw in the coplot above. A film's budget grows in proportion to its length and duration and in Fig. 1, we are not able to see the interaction between time and length.