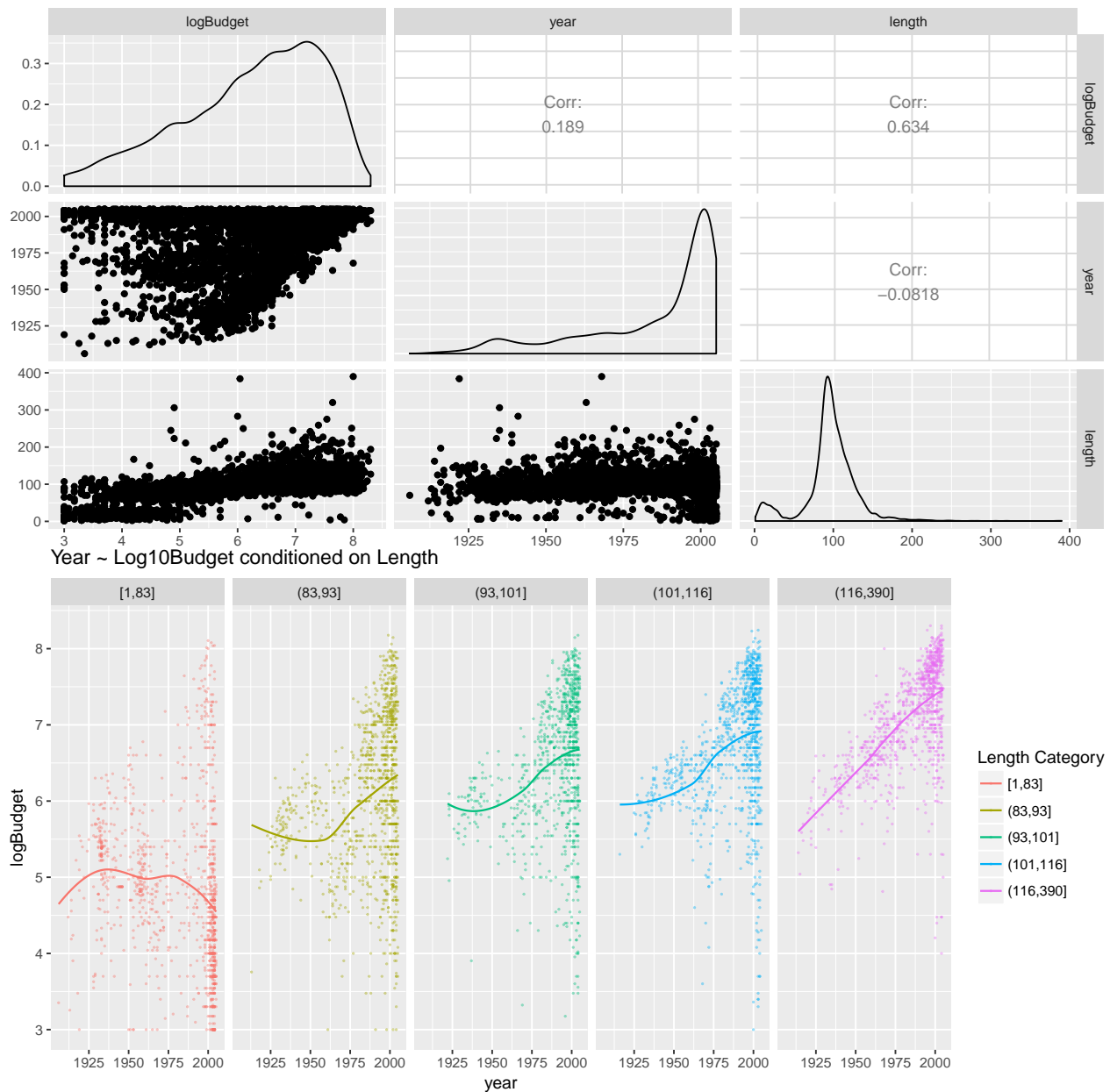


# Assignment - 5

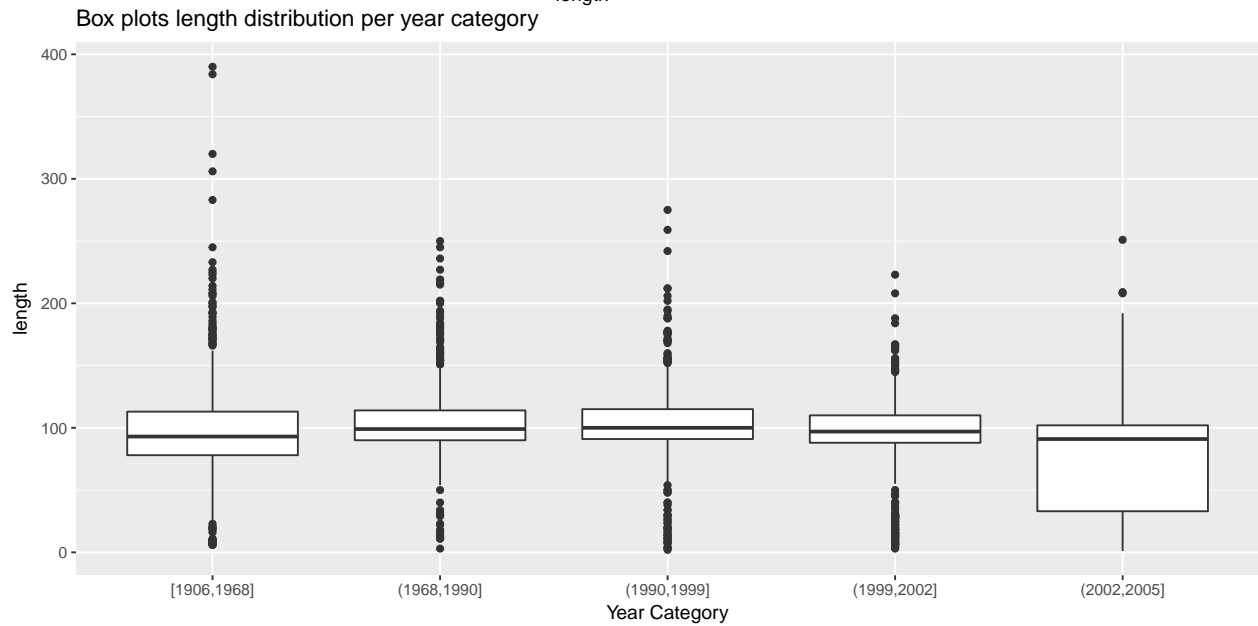
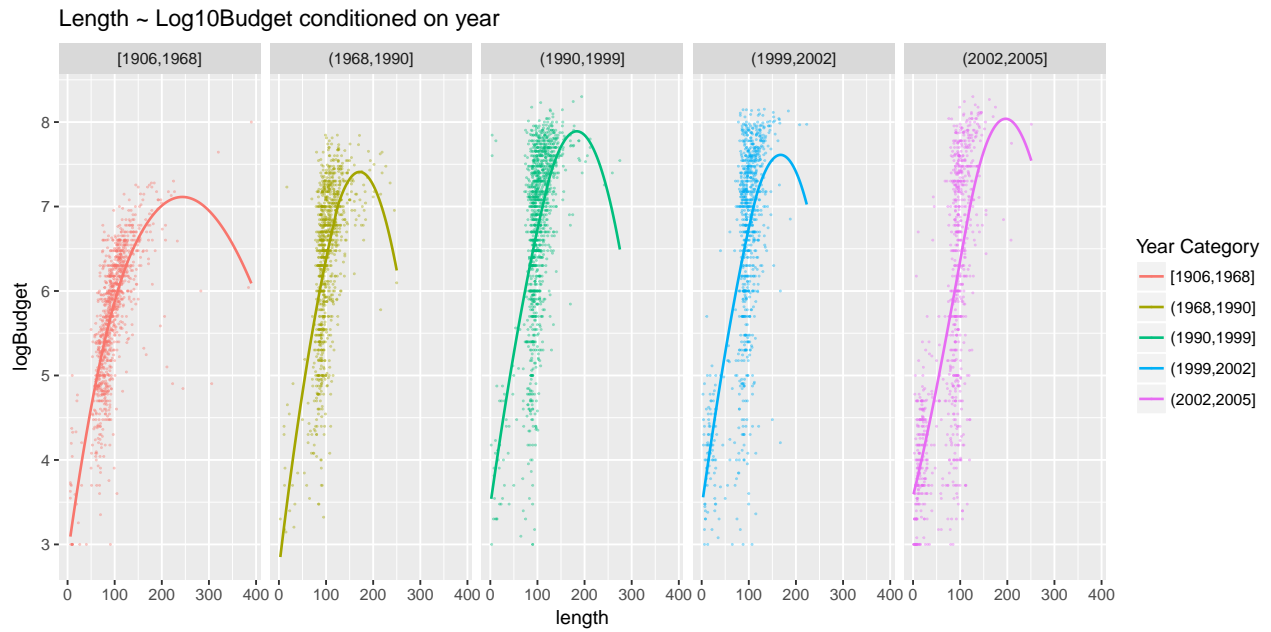
*Dwipam*

3/1/2017

## Question 1:

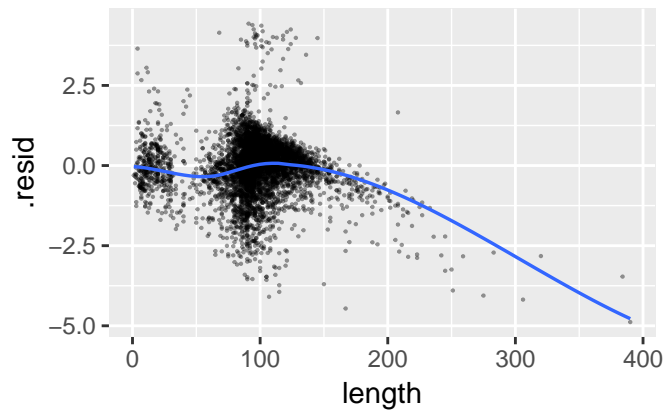


There seems to be some linear increase in log10 Budget over year, however this relationship is still unclear due to non-monotonic non-linear and skew nature of year predictor. But there seems to be linear relation between length and log10 budget. We can assess this from pair plot. Can also check the relationship of length to log10 budget conditioned on year. We can see that highest values for length from 116 - 390 gives highest values of log budget and lower values gives lower values of log10 budget.

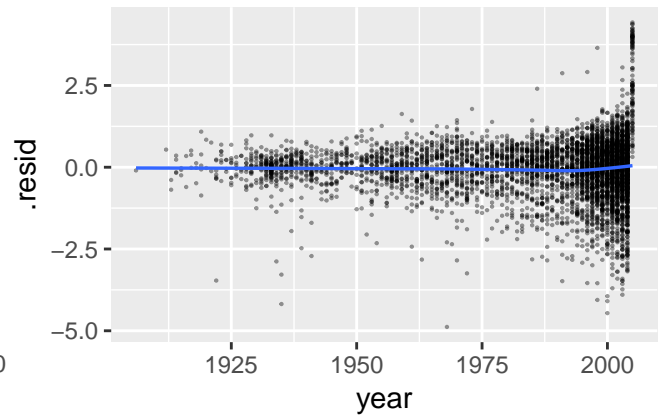


Relationship of log10budget as a function of year seems to be non-monotonic and non-linear. Relationship between year and length is also somewhat complicated. Due to outliers loess function gives a curve to drop down for highest length in each of the categories of year. It would be reasonable to have fit loess model with an interaction and linear term for length with curve term for year.

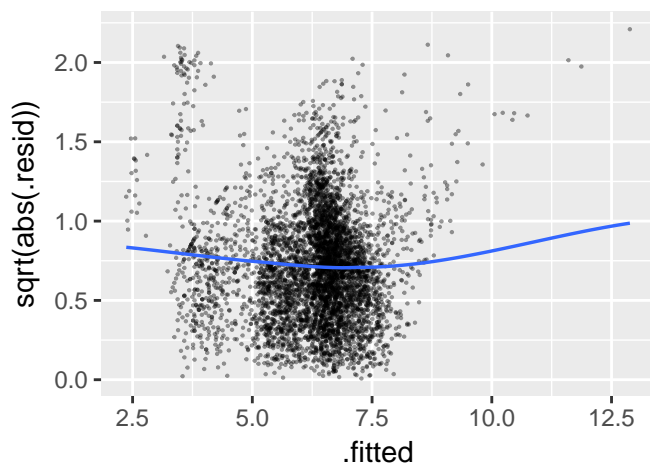
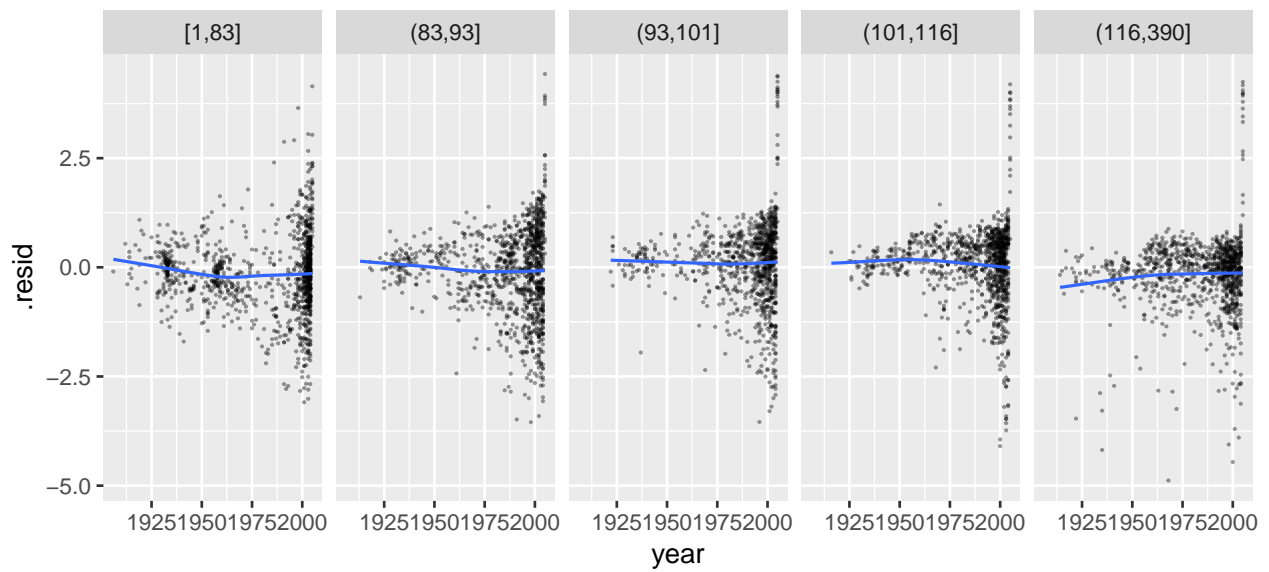
Residuals fit for length



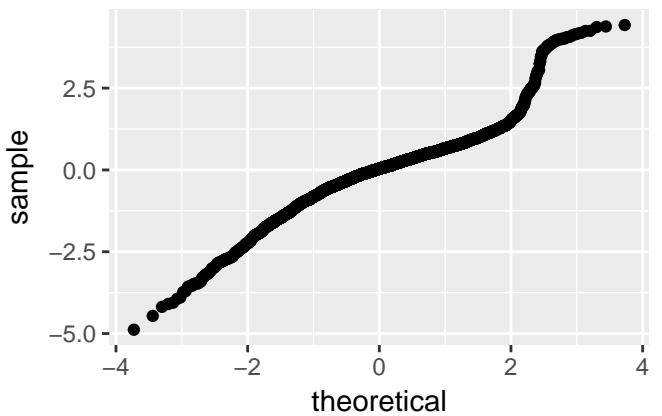
Residuals for Year



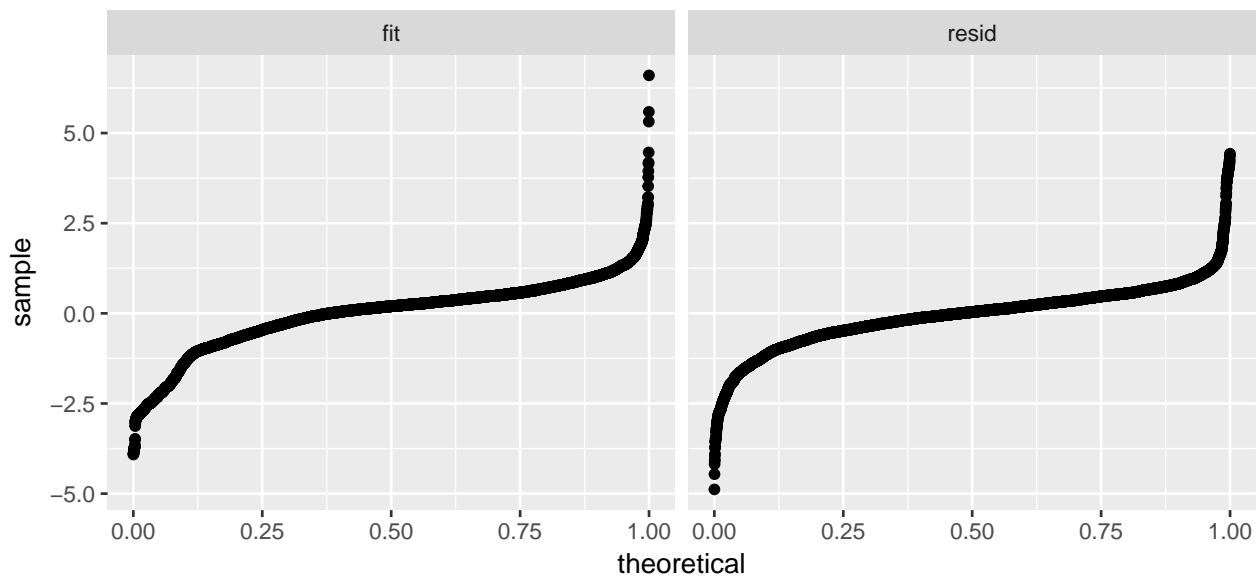
Residuals conditonal on Length



QQplot for residuals



Fitted vs Residual QQplot for uniform distribution



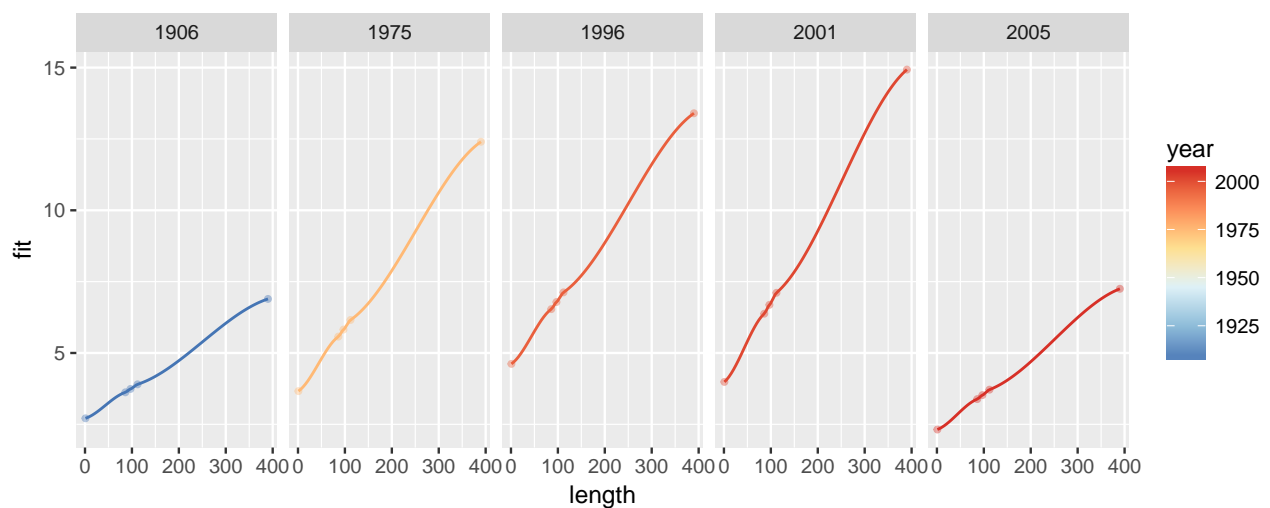
There seems to be some non-monotonic pattern in residual because, year data is skewed and we haven't taken any transformation to reduce the skewness. There also seems to be still lots of variance left within the residuals. QQplot for residuals does not seem to be normal and hence we cannot make any probabilistic judgements. However, there seems to be no heteroskedasticity.

Summarizing for question 1:

1. We do require to fit curved feature of year.
2. We don't require to fit curved feature for length.
3. Yes we do need an interaction between year and length.
4. I got 0.1 as the best span for loess smoother.
5. I won't use Robust fit as Loess with span 0.1 gives me a good linearity.

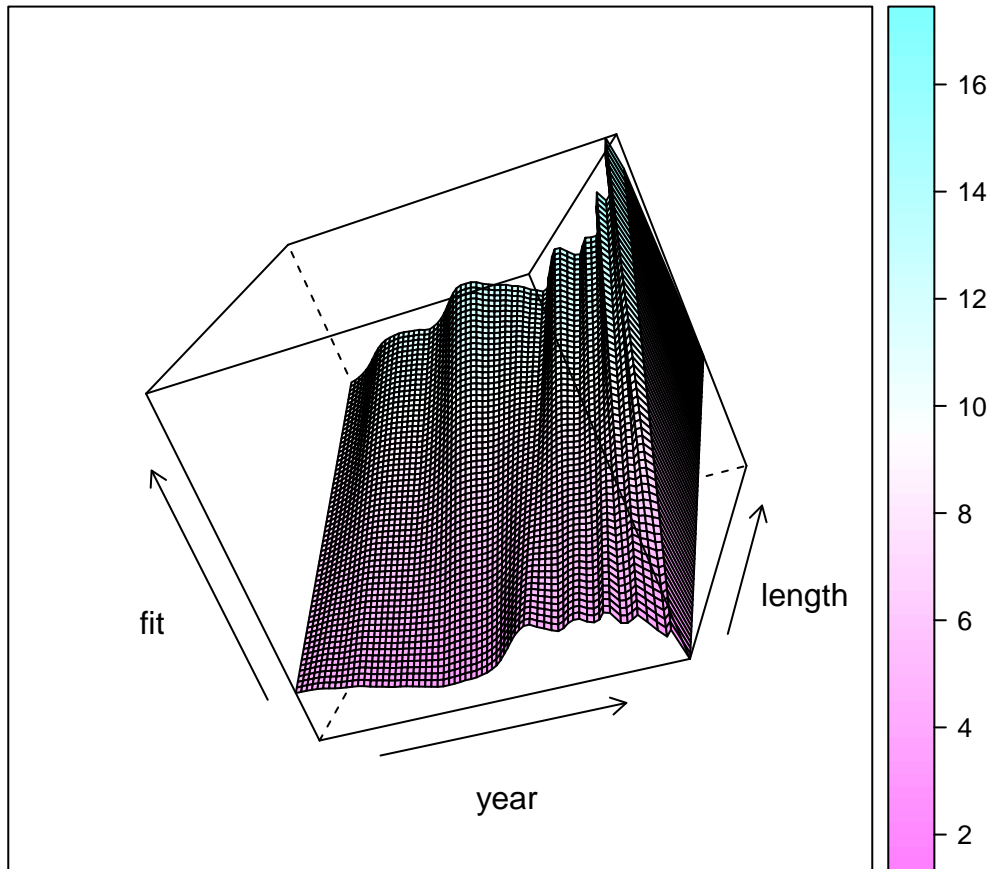
## Question 2:

Log10 budget fit conditional on Year



There is a linear increasing relationship between length and log10 budget. However for Year log10 budget rise until 2001 and then drops of after that. The slope also varies for 1906 and 2005 year and remains constant for 1975 - 2001 year.

### Question 3:



It seems to have linear trend for Length and non linear trend for year. In this plot for high length and high year there is extremely high rise in log 10 budget while for low length and high year there is drop in the log 10 budget. There is major slope in length for high year.