Evaluating the Classical Linear Model Assumptions

Q1.2 No Perfect Collinearity

First the linear regression model was created.

```
model_12 <- lm(log(views) ~ rate + length, data= videos)
```

Identifying **perfect** collinearity is relatively straightforward. If perfect collinearity is present either the regression will not run or a feature will be dropped within the model. In the model created above R does not drop any features and the regression runs successfully. We can validate this by looking at the model coefficients.

```
model_12$coefficients
```

```
##  (Intercept)         rate       length
## 5.4109124371 0.4724853515 0.0004680142
```

Identifying **nearly perfect** collinearity is harder to spot. If nearly perfect collinearity is present the model will have large standard errors on colinear features. In the model created above the standard errors are very small.

```
summary(model_12)$coefficients[,2]
```

```
##  (Intercept)         rate       length
## 4.409977e-02 1.038474e-02 7.779312e-05
```

Additionally, we can determine the correlation between the rate and length variables is relatively low, 0.157, where perfectly collinear correlation would be +/- 1.0.

```
cor(videos$rate, videos$length)
```

```
## [1] 0.1568035
```

Therefore, we conclude that the assumption of No Perfect Collinearity has been met because the regression model ran successfully and did not drop any features. Additionally, the low standard errors and correlation of the rate and length features indicates nearly perfect collinearity is not present.