R Notebook

```
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(message=FALSE)

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(patchwork)

knitr::opts_chunk$set(message=FALSE)
theme_set(theme_minimal())
```

Building the model: lm(log(views) ~ length + rate, data = df)

Correlation of length and rate columns:

```
## [1] 0.1583326
```

Getting model summary:

```
##
## Call:
## lm(formula = log(views) ~ length + rate, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6031 -1.2712 -0.0114  1.2548  6.6880
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.409e+00  4.448e-02 121.587  < 2e-16 ***
## length      4.734e-04  7.892e-05   5.998 2.07e-09 ***
## rate        4.727e-01  1.047e-02  45.132  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.801 on 9477 degrees of freedom
## Multiple R-squared:  0.1894, Adjusted R-squared:  0.1892
## F-statistic:  1107 on 2 and 9477 DF,  p-value: < 2.2e-16
```

Neither feature was dropped, so there must be no perfect collinearity. Also, the p-value assigned to both features is very low, so they're both important to our regression. Furthermore, the correlation between length and rate is only 0.16, so one is not a linear combination of the other.