# Final Project

Navya

2024-03-02

```
data = read.csv("C:/Users/navya/Documents/DATA557-Project/Factors-Influencing-Online-News-Popula
rity/data/OnlineNewsPopularity.csv")
head(data)
```

```
##                                                    url timedelta
## 1   http://mashable.com/2013/01/07/amazon-instant-video-browser/       731
## 2    http://mashable.com/2013/01/07/ap-samsung-sponsored-tweets/       731
## 3 http://mashable.com/2013/01/07/apple-40-billion-app-downloads/       731
## 4       http://mashable.com/2013/01/07/astronaut-notre-dame-bcs/       731
## 5               http://mashable.com/2013/01/07/att-u-verse-apps/       731
## 6               http://mashable.com/2013/01/07/beewi-smart-toys/       731
##   n_tokens_title n_tokens_content n_unique_tokens n_non_stop_words
## 1             12              219       0.6635945                1
## 2              9              255       0.6047431                1
## 3              9              211       0.5751295                1
## 4              9              531       0.5037879                1
## 5             13             1072       0.4156456                1
## 6             10              370       0.5598886                1
##   n_non_stop_unique_tokens num_hrefs num_self_hrefs num_imgs num_videos
## 1                0.8153846         4              2        1          0
## 2                0.7919463         3              1        1          0
## 3                0.6638655         3              1        1          0
## 4                0.6656347         9              0        1          0
## 5                0.5408895        19             19       20          0
## 6                0.6981982         2              2        0          0
##   average_token_length num_keywords data_channel_is_lifestyle
## 1             4.680365            5                         0
## 2             4.913725            4                         0
## 3             4.393365            6                         0
## 4             4.404896            7                         0
## 5             4.682836            7                         0
## 6             4.359459            9                         0
##   data_channel_is_entertainment data_channel_is_bus data_channel_is_socmed
## 1                             1                  0                     0
## 2                             0                  1                     0
## 3                             0                  1                     0
## 4                             1                  0                     0
## 5                             0                  0                     0
## 6                             0                  0                     0
##   data_channel_is_tech data_channel_is_world kw_min_min kw_max_min kw_avg_min
## 1                    0                     0          0          0          0
## 2                    0                     0          0          0          0
## 3                    0                     0          0          0          0
## 4                    0                     0          0          0          0
## 5                    1                     0          0          0          0
## 6                    1                     0          0          0          0
##   kw_min_max kw_max_max kw_avg_max kw_min_avg kw_max_avg kw_avg_avg
## 1          0          0          0          0          0          0
## 2          0          0          0          0          0          0
## 3          0          0          0          0          0          0
## 4          0          0          0          0          0          0
## 5          0          0          0          0          0          0
## 6          0          0          0          0          0          0
##   self_reference_min_shares self_reference_max_shares
## 1                       496                       496
## 2                         0                         0
```

```
## 3                      918                      918
## 4                        0                        0
## 5                      545                    16000
## 6                     8500                     8500
##   self_reference_avg_sharess weekday_is_monday weekday_is_tuesday
## 1                    496.000                1                   0
## 2                      0.000                1                   0
## 3                    918.000                1                   0
## 4                      0.000                1                   0
## 5                   3151.158                1                   0
## 6                   8500.000                1                   0
##   weekday_is_wednesday weekday_is_thursday weekday_is_friday
## 1                    0                   0                 0
## 2                    0                   0                 0
## 3                    0                   0                 0
## 4                    0                   0                 0
## 5                    0                   0                 0
## 6                    0                   0                 0
##   weekday_is_saturday weekday_is_sunday is_weekend     LDA_00     LDA_01
## 1                   0                 0          0 0.50033120 0.37827893
## 2                   0                 0          0 0.79975569 0.05004668
## 3                   0                 0          0 0.21779229 0.03333446
## 4                   0                 0          0 0.02857322 0.41929964
## 5                   0                 0          0 0.02863281 0.02879355
## 6                   0                 0          0 0.02224528 0.30671758
##       LDA_02     LDA_03     LDA_04 global_subjectivity
## 1 0.04000468 0.04126265 0.04012254           0.5216171
## 2 0.05009625 0.05010067 0.05000071           0.3412458
## 3 0.03335142 0.03333354 0.68218829           0.7022222
## 4 0.49465083 0.02890472 0.02857160           0.4298497
## 5 0.02857518 0.02857168 0.88542678           0.5135021
## 6 0.02223128 0.02222429 0.62658158           0.4374086
##   global_sentiment_polarity global_rate_positive_words
## 1                0.09256198                 0.04566210
## 2                0.14894781                 0.04313725
## 3                0.32333333                 0.05687204
## 4                0.10070467                 0.04143126
## 5                0.28100348                 0.07462687
## 6                0.07118419                 0.02972973
##   global_rate_negative_words rate_positive_words rate_negative_words
## 1                0.013698630           0.7692308           0.2307692
## 2                0.015686275           0.7333333           0.2666667
## 3                0.009478673           0.8571429           0.1428571
## 4                0.020715631           0.6666667           0.3333333
## 5                0.012126866           0.8602151           0.1397849
## 6                0.027027027           0.5238095           0.4761905
##   avg_positive_polarity min_positive_polarity max_positive_polarity
## 1             0.3786364            0.10000000                   0.7
## 2             0.2869146            0.03333333                   0.7
## 3             0.4958333            0.10000000                   1.0
## 4             0.3859652            0.13636364                   0.8
## 5             0.4111274            0.03333333                   1.0
```

```
## 6             0.3506100           0.13636364                  0.6
##   avg_negative_polarity min_negative_polarity max_negative_polarity
## 1            -0.3500000               -0.600            -0.2000000
## 2            -0.1187500               -0.125            -0.1000000
## 3            -0.4666667               -0.800            -0.1333333
## 4            -0.3696970               -0.600            -0.1666667
## 5            -0.2201923               -0.500            -0.0500000
## 6            -0.1950000               -0.400            -0.1000000
##   title_subjectivity title_sentiment_polarity abs_title_subjectivity
## 1          0.5000000              -0.1875000             0.00000000
## 2          0.0000000               0.0000000             0.50000000
## 3          0.0000000               0.0000000             0.50000000
## 4          0.0000000               0.0000000             0.50000000
## 5          0.4545455               0.1363636             0.04545455
## 6          0.6428571               0.2142857             0.14285714
##   abs_title_sentiment_polarity shares
## 1                    0.1875000    593
## 2                    0.0000000    711
## 3                    0.0000000   1500
## 4                    0.0000000   1200
## 5                    0.1363636    505
## 6                    0.2142857    855
```

```
data$n_tokens_content = scale(data$n_tokens_content)
data$n_tokens_title = scale(data$n_tokens_title)
```

```
data$is_popular = data$shares>30000
# View the modified dataset
head(data)
```

```
##                                                    url timedelta
## 1   http://mashable.com/2013/01/07/amazon-instant-video-browser/       731
## 2    http://mashable.com/2013/01/07/ap-samsung-sponsored-tweets/       731
## 3 http://mashable.com/2013/01/07/apple-40-billion-app-downloads/       731
## 4       http://mashable.com/2013/01/07/astronaut-notre-dame-bcs/       731
## 5               http://mashable.com/2013/01/07/att-u-verse-apps/       731
## 6               http://mashable.com/2013/01/07/beewi-smart-toys/       731
##   n_tokens_title n_tokens_content n_unique_tokens n_non_stop_words
## 1      0.7574377      -0.69520168       0.6635945               1
## 2     -0.6616483      -0.61878600       0.6047431               1
## 3     -0.6616483      -0.71218294       0.5751295               1
## 4     -0.6616483      -0.03293246       0.5037879               1
## 5      1.2304663       1.11542538       0.4156456               1
## 6     -0.1886196      -0.37468036       0.5598886               1
##   n_non_stop_unique_tokens num_hrefs num_self_hrefs num_imgs num_videos
## 1                0.8153846         4              2        1          0
## 2                0.7919463         3              1        1          0
## 3                0.6638655         3              1        1          0
## 4                0.6656347         9              0        1          0
## 5                0.5408895        19             19       20          0
## 6                0.6981982         2              2        0          0
##   average_token_length num_keywords data_channel_is_lifestyle
## 1             4.680365            5                         0
## 2             4.913725            4                         0
## 3             4.393365            6                         0
## 4             4.404896            7                         0
## 5             4.682836            7                         0
## 6             4.359459            9                         0
##   data_channel_is_entertainment data_channel_is_bus data_channel_is_socmed
## 1                             1                  0                     0
## 2                             0                  1                     0
## 3                             0                  1                     0
## 4                             1                  0                     0
## 5                             0                  0                     0
## 6                             0                  0                     0
##   data_channel_is_tech data_channel_is_world kw_min_min kw_max_min kw_avg_min
## 1                    0                     0          0          0          0
## 2                    0                     0          0          0          0
## 3                    0                     0          0          0          0
## 4                    0                     0          0          0          0
## 5                    1                     0          0          0          0
## 6                    1                     0          0          0          0
##   kw_min_max kw_max_max kw_avg_max kw_min_avg kw_max_avg kw_avg_avg
## 1          0          0          0          0          0          0
## 2          0          0          0          0          0          0
## 3          0          0          0          0          0          0
## 4          0          0          0          0          0          0
## 5          0          0          0          0          0          0
## 6          0          0          0          0          0          0
##   self_reference_min_shares self_reference_max_shares
## 1                       496                       496
## 2                         0                         0
```

```
## 3                      918                      918
## 4                        0                        0
## 5                      545                    16000
## 6                     8500                     8500
##   self_reference_avg_sharess weekday_is_monday weekday_is_tuesday
## 1                    496.000                1                   0
## 2                      0.000                1                   0
## 3                    918.000                1                   0
## 4                      0.000                1                   0
## 5                   3151.158                1                   0
## 6                   8500.000                1                   0
##   weekday_is_wednesday weekday_is_thursday weekday_is_friday
## 1                    0                   0                 0
## 2                    0                   0                 0
## 3                    0                   0                 0
## 4                    0                   0                 0
## 5                    0                   0                 0
## 6                    0                   0                 0
##   weekday_is_saturday weekday_is_sunday is_weekend     LDA_00     LDA_01
## 1                   0                 0          0 0.50033120 0.37827893
## 2                   0                 0          0 0.79975569 0.05004668
## 3                   0                 0          0 0.21779229 0.03333446
## 4                   0                 0          0 0.02857322 0.41929964
## 5                   0                 0          0 0.02863281 0.02879355
## 6                   0                 0          0 0.02224528 0.30671758
##       LDA_02     LDA_03     LDA_04 global_subjectivity
## 1 0.04000468 0.04126265 0.04012254           0.5216171
## 2 0.05009625 0.05010067 0.05000071           0.3412458
## 3 0.03335142 0.03333354 0.68218829           0.7022222
## 4 0.49465083 0.02890472 0.02857160           0.4298497
## 5 0.02857518 0.02857168 0.88542678           0.5135021
## 6 0.02223128 0.02222429 0.62658158           0.4374086
##   global_sentiment_polarity global_rate_positive_words
## 1                0.09256198                 0.04566210
## 2                0.14894781                 0.04313725
## 3                0.32333333                 0.05687204
## 4                0.10070467                 0.04143126
## 5                0.28100348                 0.07462687
## 6                0.07118419                 0.02972973
##   global_rate_negative_words rate_positive_words rate_negative_words
## 1                0.013698630           0.7692308           0.2307692
## 2                0.015686275           0.7333333           0.2666667
## 3                0.009478673           0.8571429           0.1428571
## 4                0.020715631           0.6666667           0.3333333
## 5                0.012126866           0.8602151           0.1397849
## 6                0.027027027           0.5238095           0.4761905
##   avg_positive_polarity min_positive_polarity max_positive_polarity
## 1             0.3786364            0.10000000                   0.7
## 2             0.2869146            0.03333333                   0.7
## 3             0.4958333            0.10000000                   1.0
## 4             0.3859652            0.13636364                   0.8
## 5             0.4111274            0.03333333                   1.0
```

```
## 6             0.3506100           0.13636364                  0.6
##   avg_negative_polarity min_negative_polarity max_negative_polarity
## 1            -0.3500000               -0.600            -0.2000000
## 2            -0.1187500               -0.125            -0.1000000
## 3            -0.4666667               -0.800            -0.1333333
## 4            -0.3696970               -0.600            -0.1666667
## 5            -0.2201923               -0.500            -0.0500000
## 6            -0.1950000               -0.400            -0.1000000
##   title_subjectivity title_sentiment_polarity abs_title_subjectivity
## 1          0.5000000              -0.1875000             0.00000000
## 2          0.0000000               0.0000000             0.50000000
## 3          0.0000000               0.0000000             0.50000000
## 4          0.0000000               0.0000000             0.50000000
## 5          0.4545455               0.1363636             0.04545455
## 6          0.6428571               0.2142857             0.14285714
##   abs_title_sentiment_polarity shares is_popular
## 1                    0.1875000    593      FALSE
## 2                    0.0000000    711      FALSE
## 3                    0.0000000   1500      FALSE
## 4                    0.0000000   1200      FALSE
## 5                    0.1363636    505      FALSE
## 6                    0.2142857    855      FALSE
```

```
# Logistic Model with only title length
model_title = glm(is_popular ~ n_tokens_title, family=binomial, data = data)
summary(model_title)
```

```
##
## Call:
## glm(formula = is_popular ~ n_tokens_title, family = binomial,
##     data = data)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.52774    0.04913 -92.159  < 2e-16 ***
## n_tokens_title  0.15846    0.04795   3.305 0.000951 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4736.9  on 39643  degrees of freedom
## Residual deviance: 4726.0  on 39642  degrees of freedom
## AIC: 4730
##
## Number of Fisher Scoring iterations: 7
```

```
# Logistic Model with only content length
model_content = glm(is_popular ~ n_tokens_content, family=binomial, data = data)
summary(model_content)
```

```
##
## Call:
## glm(formula = is_popular ~ n_tokens_content, family = binomial,
##     data = data)
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -4.51931    0.04875 -92.708   <2e-16 ***
## n_tokens_content -0.09358    0.05456  -1.715   0.0863 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4736.9  on 39643  degrees of freedom
## Residual deviance: 4733.7  on 39642  degrees of freedom
## AIC: 4737.7
##
## Number of Fisher Scoring iterations: 7
```

```
# Logistic Model with title length and content length combined
model_title_content_combined = glm(is_popular ~ n_tokens_content + n_tokens_title, family=binomi
al, data = data)
summary(model_title_content_combined)
```

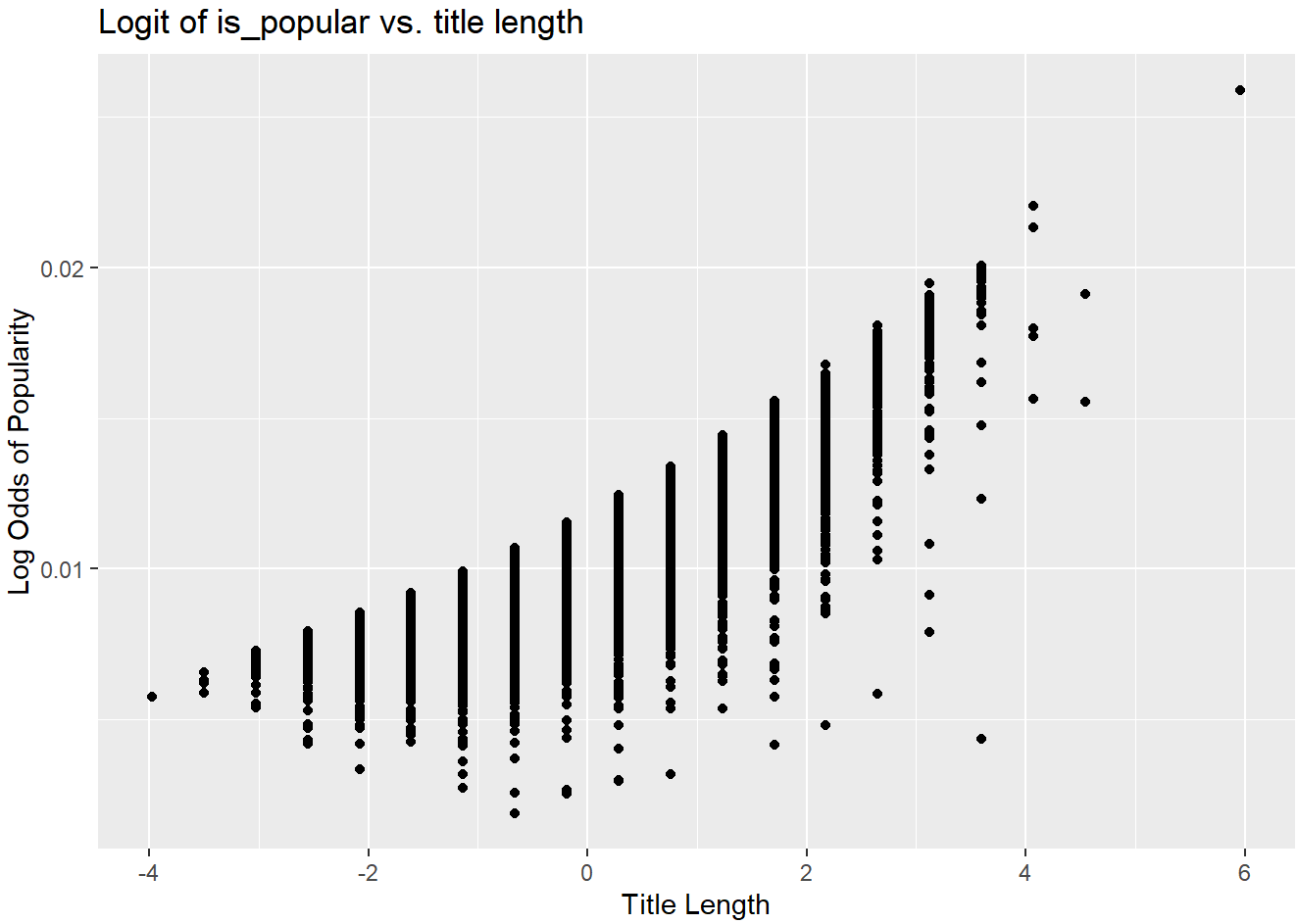
```
##
## Call:
## glm(formula = is_popular ~ n_tokens_content + n_tokens_title,
##     family = binomial, data = data)
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -4.53198    0.04935 -91.841  < 2e-16 ***
## n_tokens_content -0.09657    0.05445  -1.773 0.076157 .
## n_tokens_title    0.16043    0.04805   3.339 0.000841 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4736.9  on 39643  degrees of freedom
## Residual deviance: 4722.6  on 39641  degrees of freedom
## AIC: 4728.6
##
## Number of Fisher Scoring iterations: 7
```

```
# Confidence Interval for Combined Model
(confint(model_title_content_combined, level = 0.95))
```

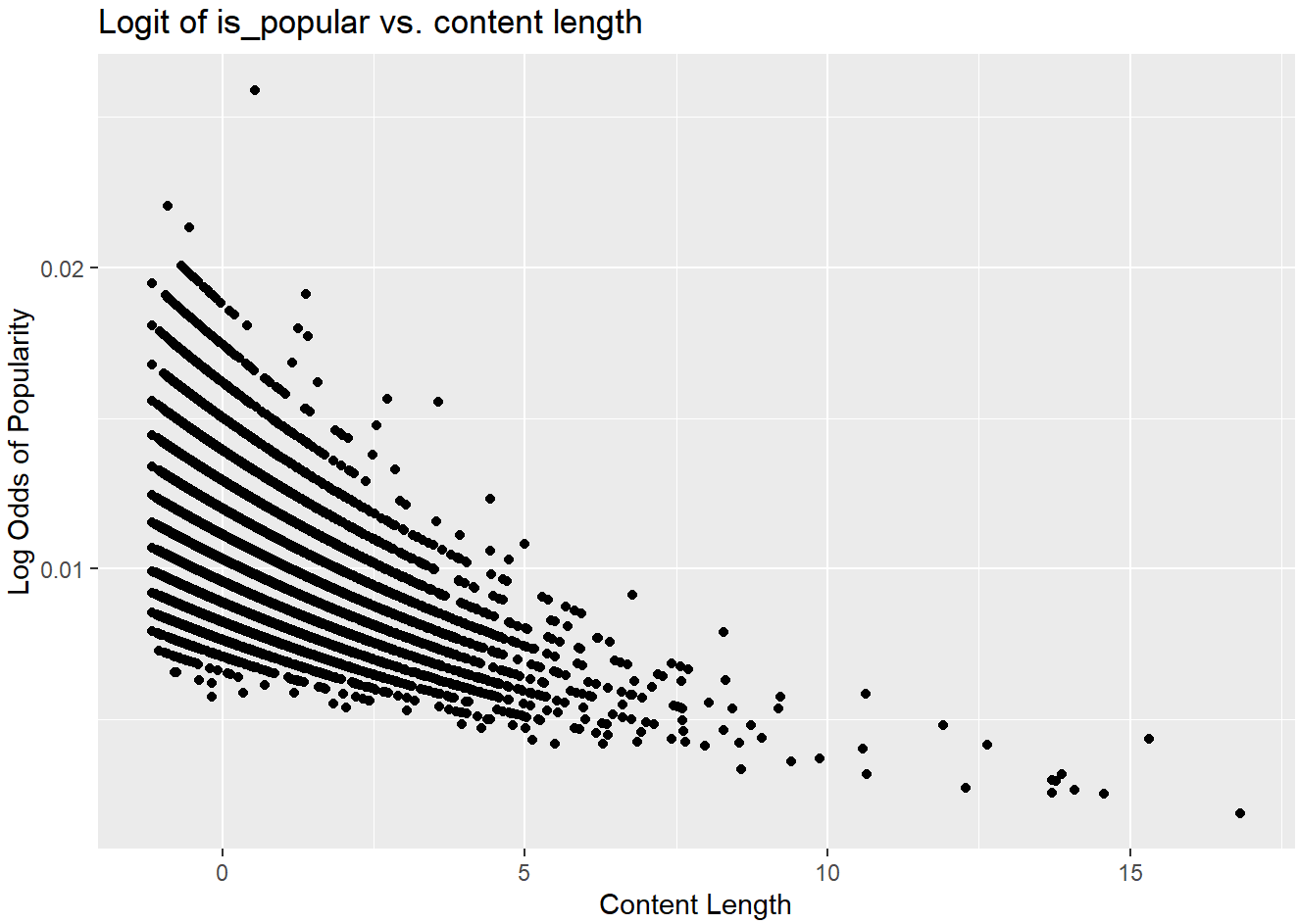
```
## Waiting for profiling to be done...
```

```
##                        2.5 %       97.5 %
## (Intercept)      -4.63030261 -4.436811090
## n_tokens_content -0.20736289  0.005849317
## n_tokens_title    0.06604559  0.254393222
```

```
# Testing logistic regression assumption - Linearity between log odds of probability and title l
ength
library(ggplot2)
ggplot(data, aes(x = n_tokens_title, y = predict(model_title_content_combined, type = "respons
e"))) +
  geom_point() +
  labs(title = "Logit of is_popular vs. title length", x = "Title Length", y = "Log Odds of Popu
larity")
```



```
# Testing logistic regression assumption - Linearity between log odds of probability and content length
library(ggplot2)
ggplot(data, aes(x = n_tokens_content, y = predict(model_title_content_combined, type = "response"))) +
  geom_point() +
  labs(title = "Logit of is_popular vs. content length", x = "Content Length", y = "Log Odds of Popularity")
```




```
# Testing Logistic Regression Assumption of multicollinearity
(correlation <- cor(data[, "n_tokens_title"], data[, "n_tokens_content"]))
```

```
##             [,1]
## [1,] 0.01815965
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

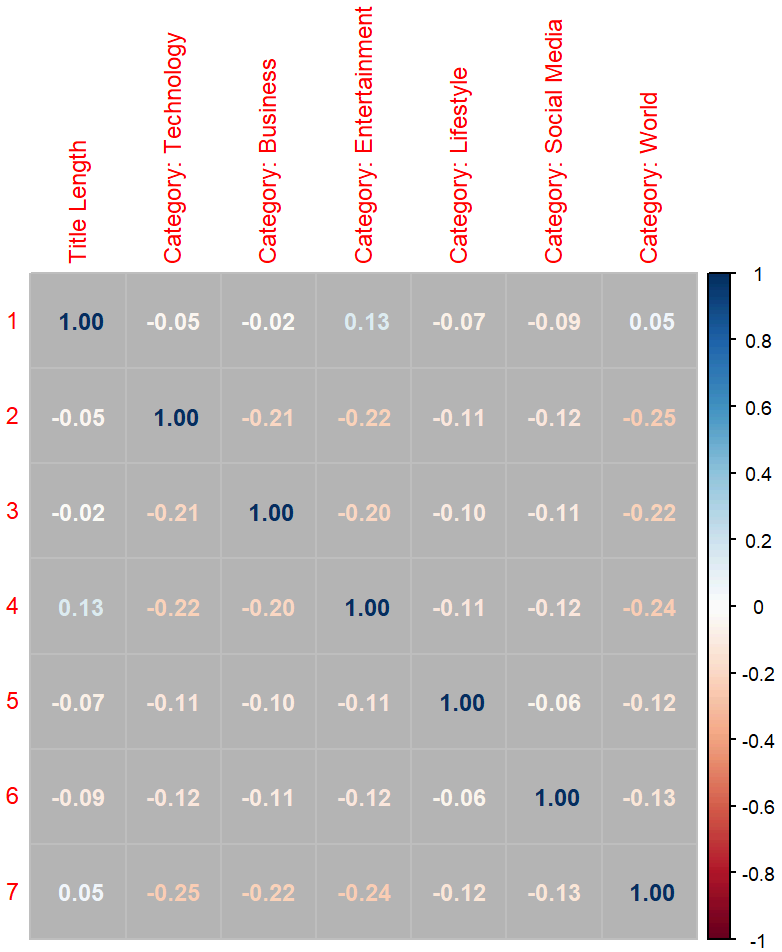
```
par(cex = 0.75)

# Assuming house_data is a data frame
data_list <- list(data$n_tokens_title, data$data_channel_is_tech, data$data_channel_is_bus, data
$data_channel_is_entertainment, data$data_channel_is_lifestyle, data$data_channel_is_socmed, dat
a$data_channel_is_world)

# Convert the list of variables into a matrix and calculate correlations
correlation_values <- cor(do.call(cbind, data_list))

# Add column names to the correlation matrix
colnames(correlation_values) <- c("Title Length", "Category: Technology", "Category: Business",
"Category: Entertainment", "Category: Lifestyle", "Category: Social Media", "Category: World")

# Plot the correlation matrix
corrplot(correlation_values, method = "number", bg = "#B4B4B8", number.digits = 2)
```




```
# Logistic regression model of title length, content length and their combine effect
model_interaction = glm(is_popular ~ n_tokens_content * n_tokens_title, family=binomial, data =
data)
summary(model_interaction)
```

```
##
## Call:
## glm(formula = is_popular ~ n_tokens_content * n_tokens_title,
##     family = binomial, data = data)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -4.53154    0.04932 -91.871  < 2e-16 ***
## n_tokens_content             -0.09029    0.05502  -1.641 0.100765
## n_tokens_title                0.15864    0.04816   3.294 0.000987 ***
## n_tokens_content:n_tokens_title -0.03330    0.05144  -0.647 0.517319
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4736.9  on 39643  degrees of freedom
## Residual deviance: 4722.2  on 39640  degrees of freedom
## AIC: 4730.2
##
## Number of Fisher Scoring iterations: 7
```

logit(is_popular) = $\beta_0$ + $\beta_1$[n_tokens_title] +
$\beta_2$[Category] + $\beta_3$[n_tokens_title * Category]

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data <- data %>%
  mutate(Category = case_when(
    data_channel_is_lifestyle == 1 ~ "Lifestyle",
    data_channel_is_entertainment == 1 ~ "Entertainment",
    data_channel_is_bus == 1 ~ "Business",
    data_channel_is_socmed == 1 ~ "Social Media",
    data_channel_is_tech == 1 ~ "Technology",
    data_channel_is_world == 1 ~ "World",
    TRUE ~ "Other"
  ))
head(data$Category)
```

```
## [1] "Entertainment" "Business"      "Business"      "Entertainment"
## [5] "Technology"    "Technology"
```

```
lm_model_total_before = glm(is_popular ~ n_tokens_title + Category, family=binomial, data = dat
a)
summary(lm_model_total_before)
```

```
##
## Call:
## glm(formula = is_popular ~ n_tokens_title + Category, family = binomial,
##     data = data)
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -4.86833    0.14504 -33.566  < 2e-16 ***
## n_tokens_title         0.17048    0.04957   3.439 0.000583 ***
## CategoryEntertainment  0.32479    0.18457   1.760 0.078460 .
## CategoryLifestyle      0.31292    0.26326   1.189 0.234576
## CategoryOther          1.39312    0.16303   8.545  < 2e-16 ***
## CategorySocial Media   0.06328    0.27793   0.228 0.819891
## CategoryTechnology    -0.50040    0.22485  -2.225 0.026048 *
## CategoryWorld         -0.41173    0.20952  -1.965 0.049399 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4736.9  on 39643  degrees of freedom
## Residual deviance: 4507.8  on 39636  degrees of freedom
## AIC: 4523.8
##
## Number of Fisher Scoring iterations: 8
```

```
lm_model_total = glm(is_popular ~ n_tokens_title * Category, data= data)
summary(lm_model_total)
```

```
##
## Call:
## glm(formula = is_popular ~ n_tokens_title * Category, data = data)
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         0.0078959  0.0013051   6.050 1.46e-09 ***
## n_tokens_title                      0.0040882  0.0012754   3.205  0.00135 **
## CategoryEntertainment               0.0031840  0.0018262   1.744  0.08125 .
## CategoryLifestyle                   0.0017713  0.0027061   0.655  0.51275
## CategoryOther                       0.0224274  0.0018546  12.093  < 2e-16 ***
## CategorySocial Media               -0.0001018  0.0026240  -0.039  0.96906
## CategoryTechnology                 -0.0033340  0.0017787  -1.874  0.06088 .
## CategoryWorld                      -0.0029736  0.0017251  -1.724  0.08476 .
## n_tokens_title:CategoryEntertainment -0.0036863  0.0017810  -2.070  0.03848 *
## n_tokens_title:CategoryLifestyle    -0.0052152  0.0027993  -1.863  0.06247 .
## n_tokens_title:CategoryOther        -0.0004642  0.0018897  -0.246  0.80596
## n_tokens_title:CategorySocial Media  -0.0039626  0.0025010  -1.584  0.11311
## n_tokens_title:CategoryTechnology   -0.0047672  0.0017542  -2.718  0.00658 **
## n_tokens_title:CategoryWorld        -0.0009313  0.0017102  -0.545  0.58606
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.01062735)
##
##     Null deviance: 424.36  on 39643  degrees of freedom
## Residual deviance: 421.16  on 39630  degrees of freedom
## AIC: -67634
##
## Number of Fisher Scoring iterations: 2
```

```
(confint(lm_model_total, level = 0.95))
```

```
## Waiting for profiling to be done...
```

```
##                                             2.5 %        97.5 %
## (Intercept)                            0.0053380233  0.0104537328
## n_tokens_title                         0.0015883820  0.0065879509
## CategoryEntertainment                 -0.0003952369  0.0067632337
## CategoryLifestyle                     -0.0035325971  0.0070752846
## CategoryOther                          0.0187923474  0.0260623596
## CategorySocial Media                  -0.0052447988  0.0050412401
## CategoryTechnology                    -0.0068202057  0.0001521602
## CategoryWorld                         -0.0063547438  0.0004074641
## n_tokens_title:CategoryEntertainment -0.0071769643 -0.0001955612
## n_tokens_title:CategoryLifestyle     -0.0107018071  0.0002714072
## n_tokens_title:CategoryOther         -0.0041680299  0.0032396270
## n_tokens_title:CategorySocial Media  -0.0088644877  0.0009393052
## n_tokens_title:CategoryTechnology    -0.0082054028 -0.0013289836
## n_tokens_title:CategoryWorld         -0.0042831509  0.0024205788
```

```
(confint(lm_model_total_before, level = 0.95))
```

```
## Waiting for profiling to be done...
```

```
##                              2.5 %        97.5 %
## (Intercept)           -5.16647579 -4.5966811800
## n_tokens_title         0.07309991  0.2674126496
## CategoryEntertainment -0.03275729  0.6926683532
## CategoryLifestyle     -0.22303378  0.8147687380
## CategoryOther          1.08263274  1.7231711663
## CategorySocial Media  -0.50710501  0.5893672240
## CategoryTechnology    -0.94837258 -0.0634185836
## CategoryWorld         -0.82500679 -0.0008748532
```

```
(summary(data$n_tokens_title))
```

```
##        V1
##  Min.   :-3.9728
##  1st Qu.:-0.6616
##  Median :-0.1886
##  Mean   : 0.0000
##  3rd Qu.: 0.7574
##  Max.   : 5.9608
```

```
(summary(data$n_tokens_content))
```

```
##        V1
##  Min.   :-1.1601
##  1st Qu.:-0.6379
##  Median :-0.2919
##  Mean   : 0.0000
##  3rd Qu.: 0.3598
##  Max.   :16.8273
```