Stat 432 Data Descriptive Statistics

Group: Score 100

March 21, 2019

Contents

Input Data

```
library(readr)
data <- read.csv("online_news.csv")
dim(data) #There're total 61 variables with 39,644 observations
```

```
## [1] 39644    61
```

First 5 observations

```
head(data)[1:5, ]
```

```
##                                                           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
##   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
##   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
##   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
##   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
##   data_channel_is_tech data_channel_is_world kw_min_min kw_max_min
## 1                    0                     0          0          0
## 2                    0                     0          0          0
## 3                    0                     0          0          0
## 4                    0                     0          0          0
## 5                    1                     0          0          0
##   kw_avg_min kw_min_max kw_max_max kw_avg_max kw_min_avg kw_max_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
##   kw_avg_avg self_reference_min_shares self_reference_max_shares
## 1          0                       496                       496
## 2          0                         0                         0
## 3          0                       918                       918
## 4          0                         0                         0
## 5          0                       545                     16000
##   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
##   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
##   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 0.05004667
## 3                   0                 0          0 0.21779229 0.03333446
## 4                   0                 0          0 0.02857322 0.41929964
## 5                   0                 0          0 0.02863281 0.02879355
##       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
##   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
##   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
##    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
##    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
##    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             0.04545454
##    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
```

Basic Statistics

```
library(pastecs)
options(scipen=100)
options(digits=2)
stat.desc(data, desc=F) #basic statistics
```

```
##           url timedelta n_tokens_title n_tokens_content n_unique_tokens
## nbr.val    NA     39644          39644            39644           39644
## nbr.null   NA         0              0             1181            1181
## nbr.na     NA         0              0                0               0
## min        NA         8              2                0               0
## max        NA       731             23             8474             701
## range      NA       723             21             8474             701
## sum        NA  14055006         412248         21666030           21733
##           n_non_stop_words n_non_stop_unique_tokens num_hrefs
## nbr.val              39644                    39644     39644
## nbr.null              1181                     1181      1299
## nbr.na                   0                        0         0
## min                      0                        0         0
## max                   1042                      650       304
## range                 1042                      650       304
## sum                  39504                    27322    431473
```

```
##          num_self_hrefs num_imgs num_videos average_token_length
## nbr.val           39644    39644      39644                39644
## nbr.null           5350     6987      25026                 1181
## nbr.na                0        0          0                    0
## min                   0        0          0                    0
## max                 116      128         91                    8
## range               116      128         91                    8
## sum              130573   180148      49550               180310
##          num_keywords data_channel_is_lifestyle
## nbr.val         39644                     39644
## nbr.null            0                     37545
## nbr.na              0                         0
## min                 1                         0
## max                10                         1
## range               9                         1
## sum            286379                      2099
##          data_channel_is_entertainment data_channel_is_bus
## nbr.val                          39644               39644
## nbr.null                         32587               33386
## nbr.na                               0                   0
## min                                  0                   0
## max                                  1                   1
## range                                1                   1
## sum                               7057                6258
##          data_channel_is_socmed data_channel_is_tech data_channel_is_world
## nbr.val                   39644                39644                 39644
## nbr.null                  37321                32298                 31217
## nbr.na                        0                    0                     0
## min                           0                    0                     0
## max                           1                    1                     1
## range                         1                    1                     1
## sum                        2323                 7346                  8427
##          kw_min_min kw_max_min kw_avg_min kw_min_max   kw_max_max
## nbr.val       39644      39644      39644      39644        39644
## nbr.null         79        819         81      17108           79
## nbr.na            0          0          0          0            0
## min              -1          0         -1          0            0
## max             377     298400      42828     843300       843300
## range           378     298400      42829     843300       843300
## sum         1034978   45747260   12383476  539648166  29825135300
##          kw_avg_max kw_min_avg kw_max_avg kw_avg_avg
## nbr.val       39644      39644      39644      39644
## nbr.null         79      17102         79         79
## nbr.na            0          0          0          0
## min               0         -1          0          0
## max          843300       3613     298400      43568
## range        843300       3614     298400      43568
## sum     10278973153   44288160  224274479  124317980
##          self_reference_min_shares self_reference_max_shares
## nbr.val                      39644                     39644
## nbr.null                      7174                      7175
## nbr.na                           0                         0
## min                              0                         0
## max                         843300                    843300
```

```
## range                            843300                 843300
## sum                           158526659              409491307
##           self_reference_avg_sharess weekday_is_monday weekday_is_tuesday
## nbr.val                        39644             39644              39644
## nbr.null                        7174             32983              32254
## nbr.na                             0                 0                  0
## min                                0                 0                  0
## max                           843300                 1                  1
## range                         843300                 1                  1
## sum                        253788899              6661               7390
##           weekday_is_wednesday weekday_is_thursday weekday_is_friday
## nbr.val                  39644               39644             39644
## nbr.null                 32209               32377             33943
## nbr.na                       0                   0                 0
## min                          0                   0                 0
## max                          1                   1                 1
## range                        1                   1                 1
## sum                       7435                7267              5701
##           weekday_is_saturday weekday_is_sunday is_weekend    LDA_00
## nbr.val                 39644             39644      39644  39644.00
## nbr.null                37191             36907      34454      1.00
## nbr.na                      0                 0          0      0.00
## min                         0                 0          0      0.00
## max                         1                 1          1      0.93
## range                       1                 1          1      0.93
## sum                      2453              2737       5190   7318.24
##            LDA_01   LDA_02   LDA_03   LDA_04 global_subjectivity
## nbr.val  39644.00 39644.00 39644.00 39644.00               39644
## nbr.null     1.00     1.00     1.00     1.00                1184
## nbr.na       0.00     0.00     0.00     0.00                   0
## min          0.00     0.00     0.00     0.00                   0
## max          0.93     0.92     0.93     0.93                   1
## range        0.93     0.92     0.93     0.93                   1
## sum       5599.94  8575.83  8871.12  9277.86               17577
##           global_sentiment_polarity global_rate_positive_words
## nbr.val                    39644.00                   39644.00
## nbr.null                    1197.00                    1217.00
## nbr.na                         0.00                       0.00
## min                           -0.39                       0.00
## max                            0.73                       0.16
## range                          1.12                       0.16
## sum                         4729.90                    1570.89
##           global_rate_negative_words rate_positive_words
## nbr.val                     39644.00               39644
## nbr.null                     2550.00                1217
## nbr.na                          0.00                   0
## min                             0.00                   0
## max                             0.18                   1
## range                           0.18                   1
## sum                           658.57               27043
##           rate_negative_words avg_positive_polarity min_positive_polarity
## nbr.val                 39644                 39644                 39644
## nbr.null                 2550                  1217                  1217
## nbr.na                      0                     0                     0
```

```
## min                                        0                   0                   0
## max                                        1                   1                   1
## range                                      1                   1                   1
## sum                                    11415               14027                3784
##            max_positive_polarity avg_negative_polarity min_negative_polarity
## nbr.val                    39644                 39644                 39644
## nbr.null                    1217                  2550                  2550
## nbr.na                         0                     0                     0
## min                            0                    -1                    -1
## max                            1                     0                     0
## range                          1                     1                     1
## sum                        30000                -10289                -20692
##            max_negative_polarity title_subjectivity title_sentiment_polarity
## nbr.val                    39644              39644                    39644
## nbr.null                    2550              18013                    19914
## nbr.na                         0                  0                        0
## min                           -1                  0                       -1
## max                            0                  1                        1
## range                          1                  1                        2
## sum                        -4262              11194                     2832
##            abs_title_subjectivity abs_title_sentiment_polarity    shares
## nbr.val                    39644.0                        39644     39644
## nbr.null                    2643.0                        19914         0
## nbr.na                         0.0                            0         0
## min                            0.0                            0         1
## max                            0.5                            1    843300
## range                          0.5                            1    843299
## sum                        13552.0                         6187 134606452
```