# data_stats

*Lushi Chen*

*6 March 2020*

## Data statistics

```
#path = "~/Documents/lucia/QuoteAndDepression/data/"
path = "/Users/lucia/Desktop/quote_local/data/"

setwd(path)

stats  <- read.csv('stats_table.csv')
```

```
#convert table to long format
stats['userid'] <- NULL
stats['X'] <- NULL

stats$pos_l_all <- stats$pos_lyrics/stats$all_post_count
stats$neg_l_all <- stats$neg_lyrics/stats$all_post_count
stats$neu_l_all <- stats$neu_lyrics/stats$all_post_count

#use absolute values for magnitude
stats$neg_mag_lyrics <- abs(stats$neg_mag_lyrics)
stats$neg_mag_quote <- abs(stats$neg_mag_quote)

stats  %>%
  dplyr::select('pos', 'neg', 'neu') -> all_posts

#get users
stats_table <- melt(all_posts)
```

```
## No id variables; using all as measure variables
```

```
p1 <- ggplot(data = stats_table,
       aes(x = variable, y = value, fill=variable)) +
  geom_boxplot() +
  stat_summary(fun.y=mean, geom="point", shape=20, size=5, color="black", fill="black")+
  scale_y_continuous(limits = c(0, 50)) +
  ylab('Post Counts') +
  theme(axis.text.x=element_text(size = 12)) +
  labs(title = 'Post Counts in one year (N = 781)')

#significant difference
t.test(stats$pos_quote,stats$neg_quote)
```

```
##
##  Welch Two Sample t-test
##
## data:  stats$pos_quote and stats$neg_quote
## t = 3.245, df = 1217.2, p-value = 0.001207
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  0.1858008 0.7540199
## sample estimates:
## mean of x mean of y
## 1.1638924 0.6939821
```

```
summary(stats$pos_quote)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   1.164   1.000  55.000
```

```
sd(stats$pos_quote)
```

```
## [1] 3.54043
```

```
#lyric counts
stats  %>%
  dplyr::select('pos_lyrics', 'neg_lyrics', 'neu_lyrics', 'pos_quote', 'neg_quote', 'neu_quote', 'all_c


lyrics <- lyrics[lyrics$all_count > 0, ]
lyrics$all_count <- NULL
lyrics_table <- melt(lyrics)
```

```
## No id variables; using all as measure variables
```

```
p2 <- ggplot(data = lyrics_table,
       aes(x = variable, y = value, fill=variable)) +
  geom_boxplot() +
  scale_y_continuous(limits = c(0, 3)) +
  stat_summary(fun.y=mean, geom="point", shape=20, size=5, color="black", fill="black")+
  ylab('Counts') +
  theme(axis.text.x=element_text(angle=30, hjust=1, size = 12)) +
  labs(title = 'Non-original Content Counts in one year (N = 305)')

#t.test(lyrics$pos,lyrics$neg)
sd(lyrics$nega_counts)
```

```
## [1] NA
```

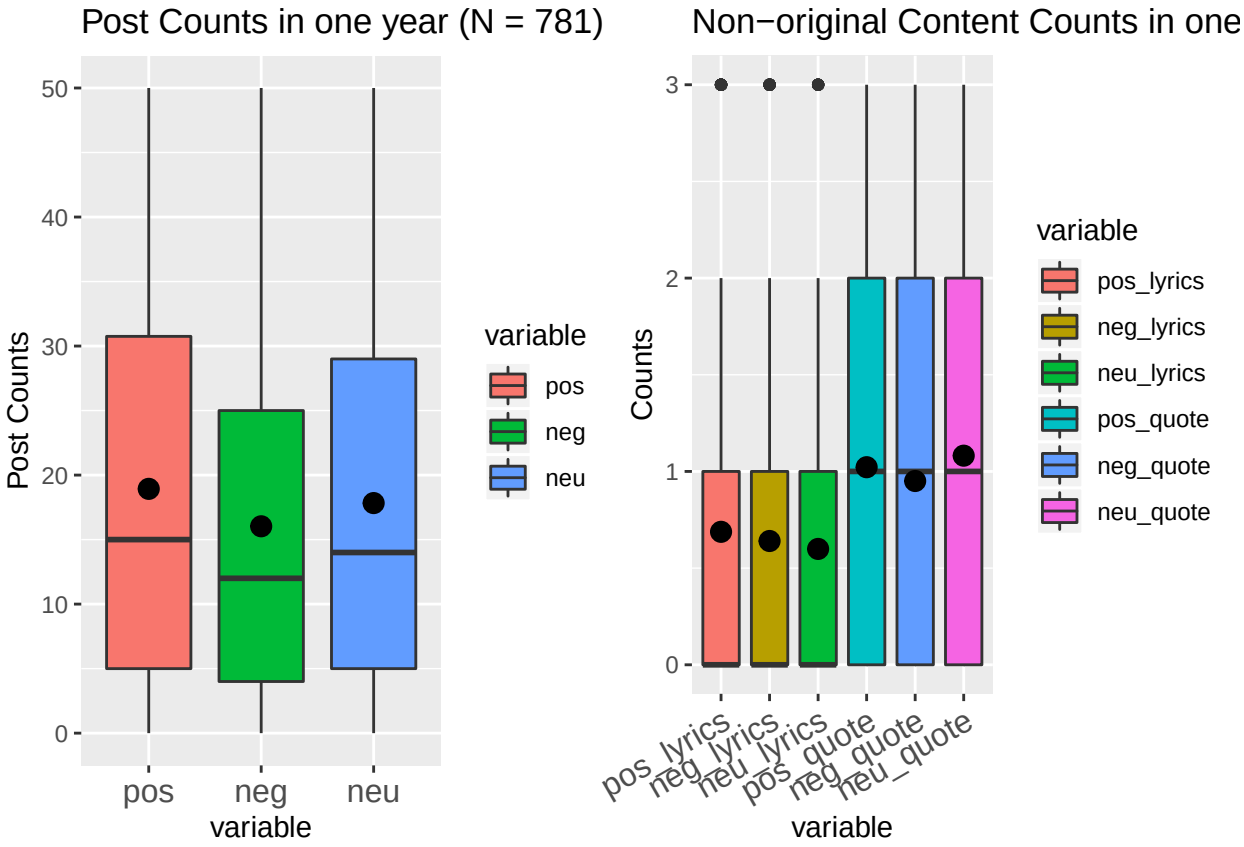
```
grid.arrange(p1, p2, ncol=2)
```

```
## Warning: Removed 614 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 614 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 313 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 313 rows containing non-finite values (stat_summary).
```




```
##
## Call:
## glm(formula = cesd_sum_binary ~ non_origin_ratio + pos_lyr_ratio +
##     neg_lyr_ratio + neu_lyr_ratio + pos_quo_ratio + neg_quo_ratio +
##     neu_quo_ratio + neg_ratio + pos_ratio + lyr_ratio + all_count +
##     pos_mag_lyrics + neg_mag_lyrics + pos_mag_quote + neg_mag_quote,
##     family = "binomial", data = stats)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9123  -1.3834   0.8564   0.9591   1.2994
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        0.95728    0.43515   2.200   0.0278 *
## non_origin_ratio   2.17773    2.35576   0.924   0.3553
## pos_lyr_ratio      0.87229    1.05111   0.830   0.4066
## neg_lyr_ratio     -0.14199    0.88390  -0.161   0.8724
## neu_lyr_ratio      1.99175    0.86775   2.295   0.0217 *
## pos_quo_ratio     -0.46709    0.65491  -0.713   0.4757
## neg_quo_ratio      0.65580    0.78463   0.836   0.4033
## neu_quo_ratio     -0.37438    0.44594  -0.840   0.4012
## neg_ratio         -0.08600    0.78735  -0.109   0.9130
## pos_ratio         -0.97654    0.71126  -1.373   0.1698
## lyr_ratio        -15.37936    6.05687  -2.539   0.0111 *
## all_count          0.02067    0.01191   1.735   0.0827 .
```

```
## pos_mag_lyrics    -0.08890    0.19390  -0.458   0.6466
## neg_mag_lyrics     0.07592    0.18898   0.402   0.6879
## pos_mag_quote     -0.06831    0.19627  -0.348   0.7278
## neg_mag_quote      0.04515    0.18855   0.239   0.8107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1018.13  on 780  degrees of freedom
## Residual deviance:  995.33  on 765  degrees of freedom
## AIC: 1027.3
##
## Number of Fisher Scoring iterations: 4
```

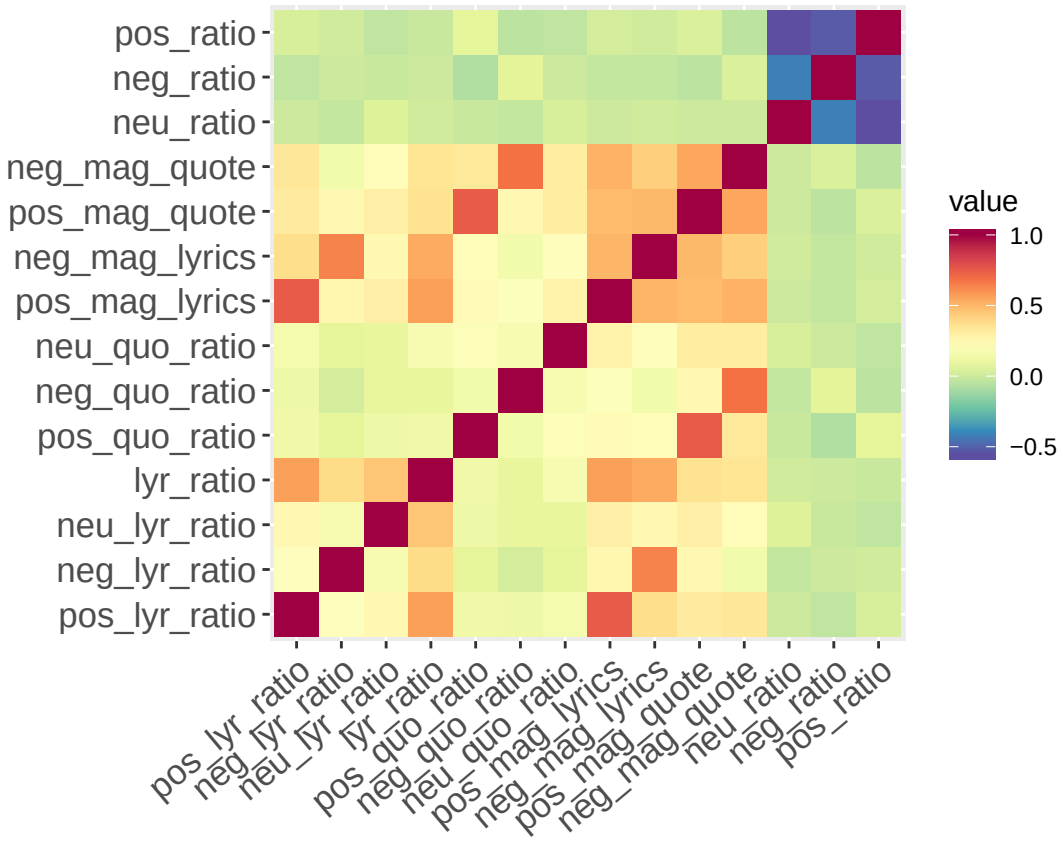
```
##
## Call:
## glm(formula = cesd_sum_binary ~ pos_lyr_ratio + neg_lyr_ratio +
##     neu_lyr_ratio + lyr_ratio, family = "binomial", data = stats)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0805  -1.4235   0.9498   0.9498   1.2130
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.56206    0.08514   6.602 4.06e-11 ***
## pos_lyr_ratio  0.52490    0.78251   0.671   0.5023
## neg_lyr_ratio -0.17324    0.69166  -0.250   0.8022
## neu_lyr_ratio  1.98314    0.84350   2.351   0.0187 *
## lyr_ratio     -7.04003    3.74340  -1.881   0.0600 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1018.1  on 780  degrees of freedom
## Residual deviance: 1010.1  on 776  degrees of freedom
## AIC: 1020.1
##
## Number of Fisher Scoring iterations: 4
```

```
##
## Call:
## glm(formula = cesd_sum_binary ~ pos_lyr_ratio + neg_lyr_ratio +
##     neu_lyr_ratio + lyr_ratio + pos_quo_ratio + neg_quo_ratio +
##     neu_quo_ratio, family = "binomial", data = stats)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0352  -1.4106   0.9085   0.9609   1.1725
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.53330    0.09495   5.617 1.95e-08 ***
```

```
## pos_lyr_ratio  0.40497    0.79056   0.512    0.608
## neg_lyr_ratio -0.13686    0.69208  -0.198    0.843
## neu_lyr_ratio  1.89785    0.84076   2.257    0.024 *
## lyr_ratio     -6.92801    3.76892  -1.838    0.066 .
## pos_quo_ratio -0.23244    0.39598  -0.587    0.557
## neg_quo_ratio  1.15285    0.58302   1.977    0.048 *
## neu_quo_ratio -0.10640    0.40955  -0.260    0.795
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1018.1  on 780  degrees of freedom
## Residual deviance: 1005.5  on 773  degrees of freedom
## AIC: 1021.5
##
## Number of Fisher Scoring iterations: 4
```

```
##
## Call:
## glm(formula = cesd_sum_binary ~ pos_lyr_ratio + neg_lyr_ratio +
##     neu_lyr_ratio + lyr_ratio + pos_quo_ratio + neg_quo_ratio +
##     neu_quo_ratio + pos_mag_lyrics + neg_mag_lyrics + pos_mag_quote +
##     neg_mag_quote + neu_ratio + neg_ratio + pos_ratio, family = "binomial",
##     data = stats)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9689  -1.3848   0.8559   0.9593   1.2391
##
## Coefficients: (1 not defined because of singularities)
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.01454    0.36673  -0.040   0.9684
## pos_lyr_ratio   0.52210    1.02203   0.511   0.6095
## neg_lyr_ratio  -0.59584    0.85417  -0.698   0.4855
## neu_lyr_ratio   1.82390    0.84973   2.146   0.0318 *
## lyr_ratio      -8.75454    4.15208  -2.108   0.0350 *
## pos_quo_ratio  -0.26043    0.60903  -0.428   0.6689
## neg_quo_ratio   0.55871    0.76890   0.727   0.4674
## neu_quo_ratio  -0.22250    0.41991  -0.530   0.5962
## pos_mag_lyrics -0.06257    0.19297  -0.324   0.7457
## neg_mag_lyrics  0.17948    0.18443   0.973   0.3305
## pos_mag_quote  -0.02460    0.19688  -0.125   0.9006
## neg_mag_quote   0.16459    0.18417   0.894   0.3715
## neu_ratio       0.95621    0.70994   1.347   0.1780
## neg_ratio       0.90047    0.73036   1.233   0.2176
## pos_ratio            NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1018.1  on 780  degrees of freedom
## Residual deviance: 1000.7  on 767  degrees of freedom
```

```
## AIC: 1028.7
##
## Number of Fisher Scoring iterations: 4
```




```
## Warning in if (pmat[ut] < 0.001) {: the condition has length > 1 and only
## the first element will be used
```

```
##                row          column           cor if..pmat.ut....0.001...
## 1   pos_lyr_ratio  neg_lyr_ratio  0.2240221680            2.432325e-10
## 2   pos_lyr_ratio  neu_lyr_ratio  0.2587208282            2.069456e-13
## 3   neg_lyr_ratio  neu_lyr_ratio  0.1739909256            9.982261e-07
## 4   pos_lyr_ratio      lyr_ratio  0.5725552955            0.000000e+00
## 5   neg_lyr_ratio      lyr_ratio  0.3901897453            0.000000e+00
## 6   neu_lyr_ratio      lyr_ratio  0.4643928321            0.000000e+00
## 7   pos_lyr_ratio  pos_quo_ratio  0.1402178349            8.433192e-05
## 8   neg_lyr_ratio  pos_quo_ratio  0.0845652181            1.809143e-02
## 9   neu_lyr_ratio  pos_quo_ratio  0.1314745469            2.292811e-04
## 10      lyr_ratio  pos_quo_ratio  0.1356163853            1.438274e-04
## 11  pos_lyr_ratio  neg_quo_ratio  0.1300462695            2.684478e-04
## 12  neg_lyr_ratio  neg_quo_ratio  0.0215304667            5.479702e-01
## 13  neu_lyr_ratio  neg_quo_ratio  0.0901414099            1.172804e-02
## 14      lyr_ratio  neg_quo_ratio  0.0890606902            1.277839e-02
## 15  pos_quo_ratio  neg_quo_ratio  0.1541991255            1.502529e-05
## 16  pos_lyr_ratio  neu_quo_ratio  0.1572510409            1.009846e-05
## 17  neg_lyr_ratio  neu_quo_ratio  0.0835262688            1.956440e-02
## 18  neu_lyr_ratio  neu_quo_ratio  0.0931037190            9.230395e-03
## 19      lyr_ratio  neu_quo_ratio  0.1672189995            2.616907e-06
```

```
## 20  pos_quo_ratio  neu_quo_ratio  0.1987898138           2.112034e-08
## 21  neg_quo_ratio  neu_quo_ratio  0.1745522516           9.200207e-07
## 22  pos_lyr_ratio pos_mag_lyrics  0.7480753411           0.000000e+00
## 23  neg_lyr_ratio pos_mag_lyrics  0.2657792231           4.285461e-14
## 24  neu_lyr_ratio pos_mag_lyrics  0.2954788022           0.000000e+00
## 25      lyr_ratio pos_mag_lyrics  0.5706546262           0.000000e+00
## 26  pos_quo_ratio pos_mag_lyrics  0.2455086381           3.479883e-12
## 27  neg_quo_ratio pos_mag_lyrics  0.2126748296           1.939748e-09
## 28  neu_quo_ratio pos_mag_lyrics  0.2764760964           3.552714e-15
## 29  pos_lyr_ratio neg_mag_lyrics  0.3779910057           0.000000e+00
## 30  neg_lyr_ratio neg_mag_lyrics  0.6373267938           0.000000e+00
## 31  neu_lyr_ratio neg_mag_lyrics  0.2606559548           1.350031e-13
## 32      lyr_ratio neg_mag_lyrics  0.5416678302           0.000000e+00
## 33  pos_quo_ratio neg_mag_lyrics  0.2368618998           2.021494e-11
## 34  neg_quo_ratio neg_mag_lyrics  0.1511712411           2.212158e-05
## 35  neu_quo_ratio neg_mag_lyrics  0.2313491705           5.989653e-11
## 36 pos_mag_lyrics neg_mag_lyrics  0.5062574879           0.000000e+00
## 37  pos_lyr_ratio  pos_mag_quote  0.3180855163           0.000000e+00
## 38  neg_lyr_ratio  pos_mag_quote  0.2625270549           8.926193e-14
## 39  neu_lyr_ratio  pos_mag_quote  0.2964750061           0.000000e+00
## 40      lyr_ratio  pos_mag_quote  0.3557356386           0.000000e+00
## 41  pos_quo_ratio  pos_mag_quote  0.7497658245           0.000000e+00
## 42  neg_quo_ratio  pos_mag_quote  0.2586801157           2.089440e-13
## 43  neu_quo_ratio  pos_mag_quote  0.3106980364           0.000000e+00
## 44 pos_mag_lyrics  pos_mag_quote  0.4898434403           0.000000e+00
## 45 neg_mag_lyrics  pos_mag_quote  0.4953034084           0.000000e+00
## 46  pos_lyr_ratio  neg_mag_quote  0.3401912806           0.000000e+00
## 47  neg_lyr_ratio  neg_mag_quote  0.1473073723           3.585595e-05
## 48  neu_lyr_ratio  neg_mag_quote  0.2367044387           2.085976e-11
## 49      lyr_ratio  neg_mag_quote  0.3525056397           0.000000e+00
## 50  pos_quo_ratio  neg_mag_quote  0.3333412473           0.000000e+00
## 51  neg_quo_ratio  neg_mag_quote  0.6752990415           0.000000e+00
## 52  neu_quo_ratio  neg_mag_quote  0.3107129385           0.000000e+00
## 53 pos_mag_lyrics  neg_mag_quote  0.5220334196           0.000000e+00
## 54 neg_mag_lyrics  neg_mag_quote  0.4306768805           0.000000e+00
## 55  pos_mag_quote  neg_mag_quote  0.5476299879           0.000000e+00
## 56  pos_lyr_ratio      neu_ratio  0.0007916231           9.823781e-01
## 57  neg_lyr_ratio      neu_ratio -0.0207227424           5.630885e-01
## 58  neu_lyr_ratio      neu_ratio  0.0501605307           1.613811e-01
## 59      lyr_ratio      neu_ratio  0.0144852835           6.860774e-01
## 60  pos_quo_ratio      neu_ratio -0.0127555695           7.219039e-01
## 61  neg_quo_ratio      neu_ratio -0.0191845670           5.924219e-01
## 62  neu_quo_ratio      neu_ratio  0.0310862790           3.856314e-01
## 63 pos_mag_lyrics      neu_ratio  0.0006408033           9.857350e-01
## 64 neg_mag_lyrics      neu_ratio  0.0108555458           7.619690e-01
## 65  pos_mag_quote      neu_ratio -0.0013232216           9.705488e-01
## 66  neg_mag_quote      neu_ratio  0.0017272402           9.615626e-01
## 67  pos_lyr_ratio      neg_ratio -0.0315763396           3.781822e-01
## 68  neg_lyr_ratio      neg_ratio  0.0049193022           8.908274e-01
## 69  neu_lyr_ratio      neg_ratio -0.0141368873           6.932402e-01
## 70      lyr_ratio      neg_ratio -0.0042906758           9.047071e-01
## 71  pos_quo_ratio      neg_ratio -0.0687434826           5.481669e-02
## 72  neg_quo_ratio      neg_ratio  0.0658510342           6.586416e-02
## 73  neu_quo_ratio      neg_ratio -0.0001113793           9.975204e-01
```

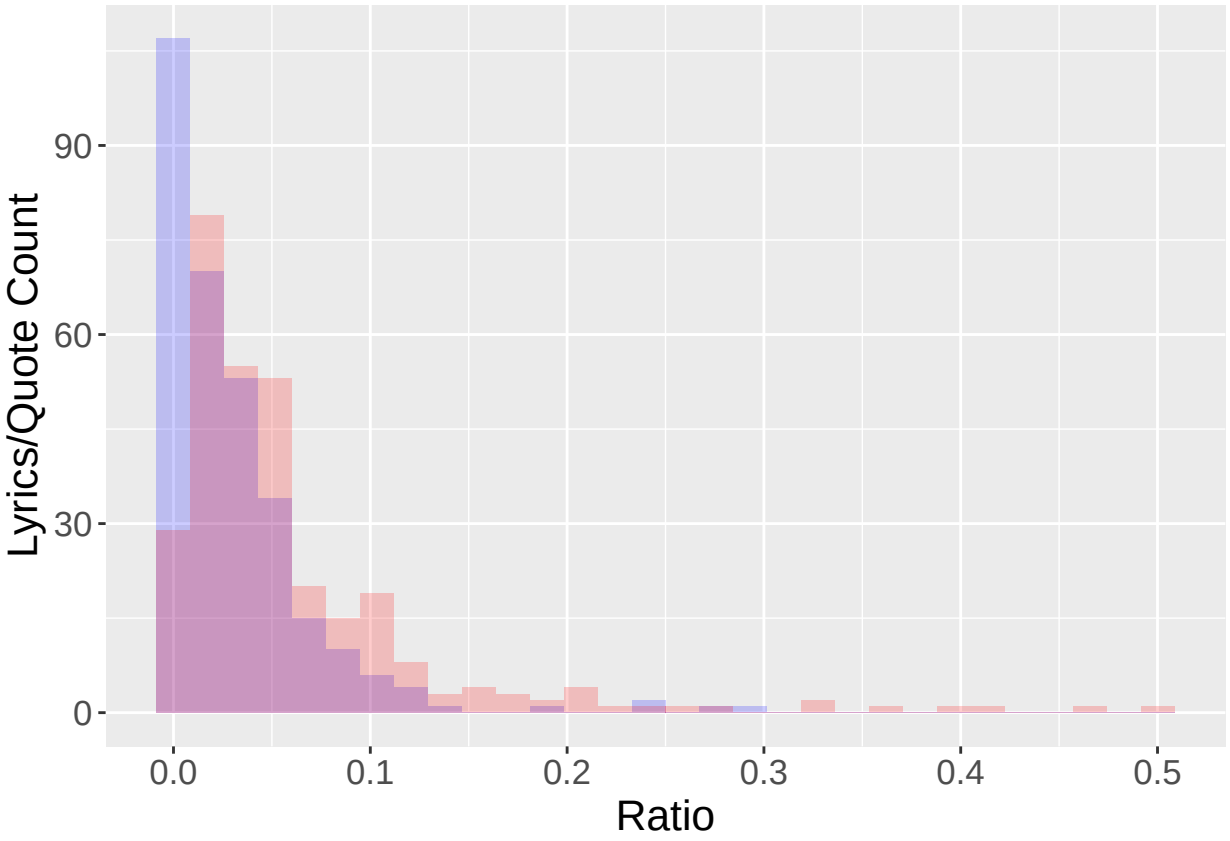
```
## 74 pos_mag_lyrics      neg_ratio -0.0212413995           5.533577e-01
## 75 neg_mag_lyrics      neg_ratio -0.0218747352           5.415879e-01
## 76  pos_mag_quote      neg_ratio -0.0390109304           2.762057e-01
## 77  neg_mag_quote      neg_ratio  0.0374461186           2.959427e-01
## 78      neu_ratio      neg_ratio -0.4245311642           0.000000e+00
## 79  pos_lyr_ratio      pos_ratio  0.0284027948           4.279854e-01
## 80  neg_lyr_ratio      pos_ratio  0.0149581654           6.764002e-01
## 81  neu_lyr_ratio      pos_ratio -0.0341490856           3.405431e-01
## 82      lyr_ratio      pos_ratio -0.0096692928           7.873178e-01
## 83  pos_quo_ratio      pos_ratio  0.0754585176           3.499457e-02
## 84  neg_quo_ratio      pos_ratio -0.0427344068           2.329070e-01
## 85  neu_quo_ratio      pos_ratio -0.0291478808           4.159625e-01
## 86 pos_mag_lyrics      pos_ratio  0.0190046714           5.958978e-01
## 87 neg_mag_lyrics      pos_ratio  0.0099777133           7.807044e-01
## 88  pos_mag_quote      pos_ratio  0.0372555280           2.984094e-01
## 89  neg_mag_quote      pos_ratio -0.0361912344           3.124355e-01
## 90      neu_ratio      pos_ratio -0.5490731717           0.000000e+00
## 91      neg_ratio      pos_ratio -0.5236225004           0.000000e+00
```

You can also embed plots, for example:

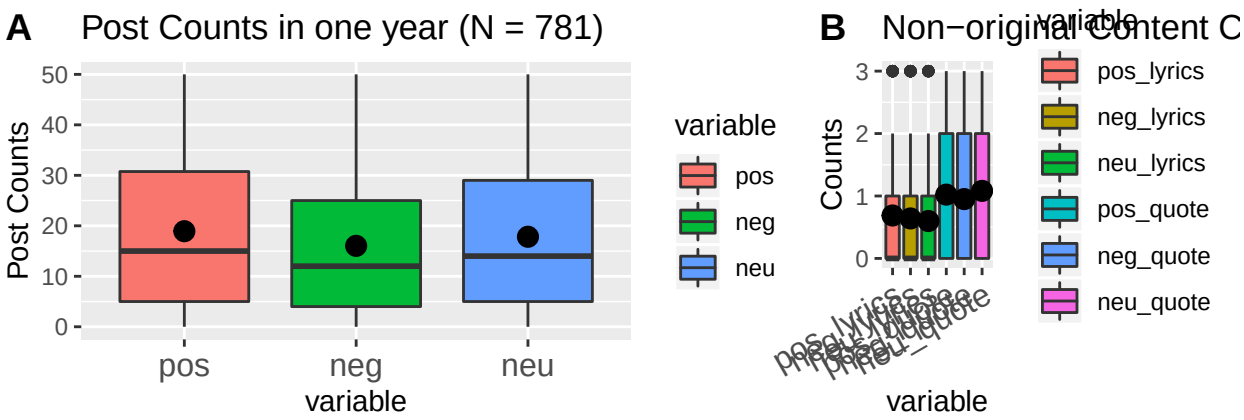
## plot distribution of content originality

```
## No id variables; using all as measure variables
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



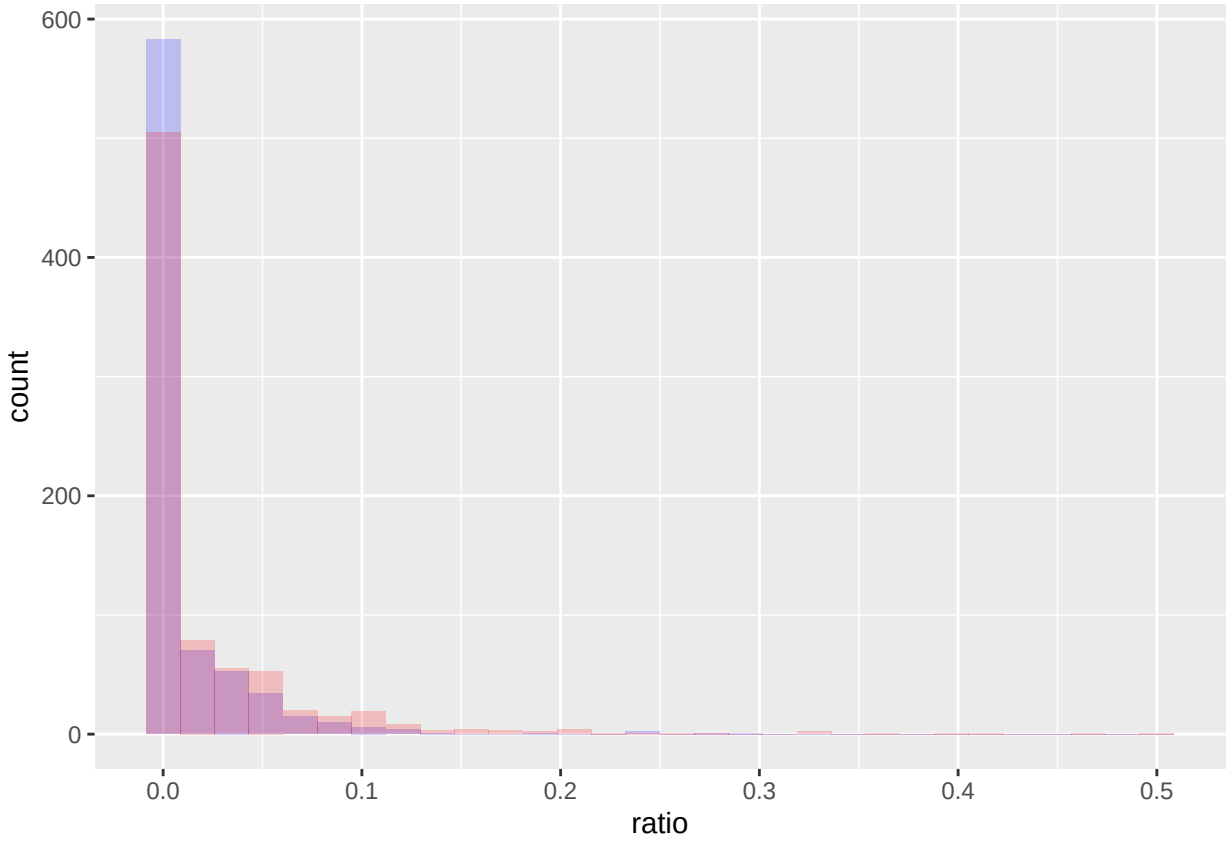

```
## Warning: Removed 614 rows containing non-finite values (stat_boxplot).
## Warning: Removed 614 rows containing non-finite values (stat_summary).
## Warning: Removed 313 rows containing non-finite values (stat_boxplot).
## Warning: Removed 313 rows containing non-finite values (stat_summary).
```




## plot ratio of lyrics, negative quotation ratio and depression score

```
## No id variables; using all as measure variables
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




## sum column value

```
## [1] 909
```

```
## [1] 542
```

```
## [1] 783
```

```
##    pos_counts      nega_counts       neu_counts       all_count
##  Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   :  0.00
##  1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.:  0.00
##  Median : 0.000   Median : 0.000   Median : 0.000   Median :  0.00
##  Mean   : 2.243   Mean   : 1.385   Mean   : 1.923   Mean   :  5.55
##  3rd Qu.: 2.000   3rd Qu.: 1.000   3rd Qu.: 1.000   3rd Qu.:  5.00
##  Max.   :57.000   Max.   :27.000   Max.   :40.000   Max.   :122.00
##
##      lyrics           quote          pos_lyrics        neg_lyrics
##  Min.   : 0.000   Min.   :  0.000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.000   1st Qu.:  0.000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.000   Median :  0.000   Median : 0.0000   Median : 0.0000
##  Mean   : 2.209   Mean   :  3.341   Mean   : 0.8828   Mean   : 0.5397
##  3rd Qu.: 1.000   3rd Qu.:  3.000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :71.000   Max.   :116.000   Max.   :25.0000   Max.   :21.0000
##
##    neu_lyrics        pos_quote       neg_quote         neu_quote
##  Min.   : 0.0000   Min.   : 0.00   Min.   : 0.0000   Min.   : 0.000
```

```
##  1st Qu.: 0.0000   1st Qu.: 0.00   1st Qu.: 0.0000   1st Qu.: 0.000
##  Median : 0.0000   Median : 0.00   Median : 0.0000   Median : 0.000
##  Mean   : 0.7866   Mean   : 1.36   Mean   : 0.8452   Mean   : 1.136
##  3rd Qu.: 0.0000   3rd Qu.: 1.00   3rd Qu.: 1.0000   3rd Qu.: 1.000
##  Max.   :25.0000   Max.   :55.00   Max.   :23.0000   Max.   :38.000
##
##       pos              neg              neu         pos_quo_ratio
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   :0.00000
##  1st Qu.:  9.25   1st Qu.:  6.00   1st Qu.:  9.00   1st Qu.:0.00000
##  Median : 31.00   Median : 18.00   Median : 26.00   Median :0.00000
##  Mean   : 50.56   Mean   : 29.82   Mean   : 44.33   Mean   :0.09796
##  3rd Qu.: 72.00   3rd Qu.: 41.00   3rd Qu.: 64.75   3rd Qu.:0.15385
##  Max.   :365.00   Max.   :244.00   Max.   :360.00   Max.   :1.00000
##
##  neg_quo_ratio     neu_quo_ratio     neg_pos_quo_ratio pos_lyr_ratio
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000    Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000    1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000    Median :0.00000
##  Mean   :0.07233   Mean   :0.09709   Mean   :0.1861    Mean   :0.05277
##  3rd Qu.:0.06957   3rd Qu.:0.14187   3rd Qu.:0.0000    3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :4.0000    Max.   :1.00000
##
##  neg_lyr_ratio     neu_lyr_ratio     neg_pos_lyr_ratio all_post_count
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000    Min.   :  1.0
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000    1st Qu.: 25.0
##  Median :0.00000   Median :0.00000   Median :0.0000    Median : 76.5
##  Mean   :0.03811   Mean   :0.05597   Mean   :0.1189    Mean   :124.7
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000    3rd Qu.:177.8
##  Max.   :1.00000   Max.   :1.00000   Max.   :2.0000    Max.   :938.0
##
##  non_origin_ratio    neg_ratio        pos_ratio        neu_ratio
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.2000   1st Qu.:0.3645   1st Qu.:0.3158
##  Median :0.00000   Median :0.2381   Median :0.4049   Median :0.3542
##  Mean   :0.03639   Mean   :0.2399   Mean   :0.4058   Mean   :0.3543
##  3rd Qu.:0.05155   3rd Qu.:0.2702   3rd Qu.:0.4484   3rd Qu.:0.3915
##  Max.   :0.50000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##  neg_pos_ratio      lyr_ratio      pos_mag_lyrics   neg_mag_lyrics
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.4786   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.5854   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :   Inf   Mean   :0.0113   Mean   :0.3595   Mean   :0.3176
##  3rd Qu.:0.7163   3rd Qu.:0.0108   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :   Inf   Max.   :0.1268   Max.   :4.0000   Max.   :4.0000
##  NA's   :5
##  neu_mag_lyrics pos_mag_quote    neg_mag_quote    neu_mag_quote
##  Min.   :0      Min.   :0.0000   Min.   :0.0000   Min.   :0
##  1st Qu.:0      1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0
##  Median :0      Median :0.0000   Median :0.0000   Median :0
##  Mean   :0      Mean   :0.4675   Mean   :0.4514   Mean   :0
##  3rd Qu.:0      3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0
##  Max.   :0      Max.   :4.0000   Max.   :3.0000   Max.   :0
##
```

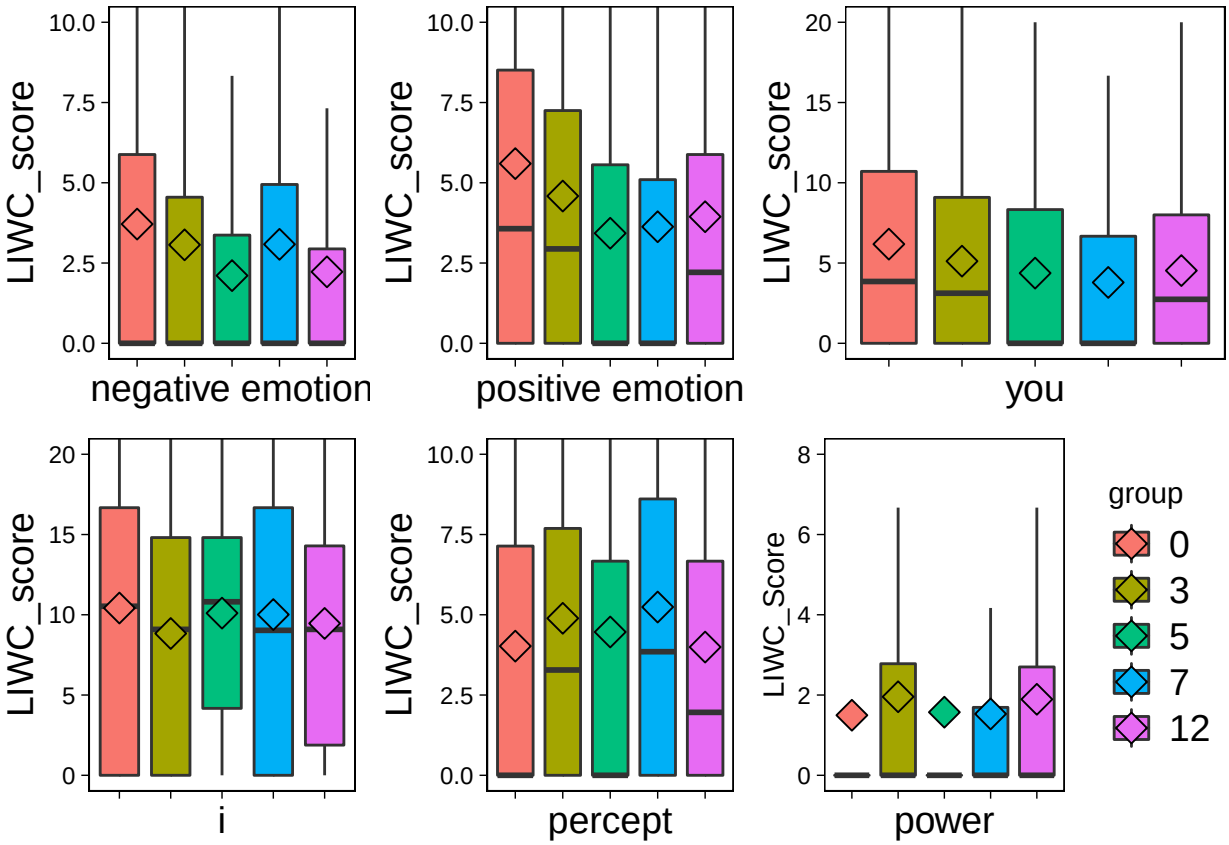
```
##     pos_mag         neg_mag          neu_mag     cesd_sum
##  Min.   :0.000   Min.   :-3.500   Min.   :0   Min.   :23.00
##  1st Qu.:1.478   1st Qu.:-1.804   1st Qu.:0   1st Qu.:27.00
##  Median :1.558   Median :-1.675   Median :0   Median :32.00
##  Mean   :1.538   Mean   :-1.612   Mean   :0   Mean   :31.97
##  3rd Qu.:1.652   3rd Qu.:-1.538   3rd Qu.:0   3rd Qu.:36.00
##  Max.   :4.000   Max.   : 0.000   Max.   :0   Max.   :53.00
##
##    pos_l_all          neg_l_all          neu_l_all        cesd_sum_binary
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :1
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:1
##  Median :0.000000   Median :0.000000   Median :0.000000   Median :1
##  Mean   :0.004343   Mean   :0.002648   Mean   :0.004311   Mean   :1
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:1
##  Max.   :0.062500   Max.   :0.055556   Max.   :0.090909   Max.   :1
##
##   quote_ratio
##  Min.   :0.00000
##  1st Qu.:0.00000
##  Median :0.00000
##  Mean   :0.02509
##  3rd Qu.:0.02915
##  Max.   :0.50000
##

##    pos_counts      nega_counts        neu_counts       all_count
##  Min.   : 0.000   Min.   : 0.0000   Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.000
##  Median : 0.000   Median : 0.0000   Median : 0.000   Median : 0.000
##  Mean   : 1.491   Mean   : 0.8817   Mean   : 1.358   Mean   : 3.731
##  3rd Qu.: 1.000   3rd Qu.: 1.0000   3rd Qu.: 1.000   3rd Qu.: 3.000
##  Max.   :35.000   Max.   :19.0000   Max.   :31.000   Max.   :82.000
##
##      lyrics           quote          pos_lyrics       neg_lyrics
##  Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.0000
##  1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.0000
##  Median : 0.000   Median : 0.000   Median : 0.000   Median : 0.0000
##  Mean   : 1.523   Mean   : 2.208   Mean   : 0.595   Mean   : 0.4194
##  3rd Qu.: 1.000   3rd Qu.: 2.000   3rd Qu.: 0.000   3rd Qu.: 0.0000
##  Max.   :63.000   Max.   :39.000   Max.   :23.000   Max.   :13.0000
##
##    neu_lyrics       pos_quote         neg_quote         neu_quote
##  Min.   : 0.000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.509   Mean   : 0.8961   Mean   : 0.4624   Mean   : 0.8495
##  3rd Qu.: 0.000   3rd Qu.: 1.0000   3rd Qu.: 0.0000   3rd Qu.: 1.0000
##  Max.   :27.000   Max.   :17.0000   Max.   :10.0000   Max.   :15.0000
##
##       pos              neg              neu        pos_quo_ratio
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.0   Min.   :0.00000
##  1st Qu.: 10.00   1st Qu.:  6.00   1st Qu.:  8.0   1st Qu.:0.00000
##  Median : 31.00   Median : 17.00   Median : 25.0   Median :0.00000
##  Mean   : 44.99   Mean   : 24.89   Mean   : 38.2   Mean   :0.09800
```

```
##  3rd Qu.: 63.00   3rd Qu.: 35.50   3rd Qu.: 52.0   3rd Qu.:0.09545
##  Max.   :283.00   Max.   :161.00   Max.   :251.0   Max.   :1.00000
##
##  neg_quo_ratio     neu_quo_ratio     neg_pos_quo_ratio pos_lyr_ratio
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000     Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000     1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.000     Median :0.00000
##  Mean   :0.04819   Mean   :0.09346   Mean   :0.141     Mean   :0.05071
##  3rd Qu.:0.00000   3rd Qu.:0.09348   3rd Qu.:0.000     3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :4.000     Max.   :0.66667
##
##  neg_lyr_ratio     neu_lyr_ratio     neg_pos_lyr_ratio all_post_count
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000    Min.   :  1.0
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000    1st Qu.: 24.5
##  Median :0.00000   Median :0.00000   Median :0.0000    Median : 72.0
##  Mean   :0.04153   Mean   :0.03728   Mean   :0.1194    Mean   :108.1
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000    3rd Qu.:146.0
##  Max.   :1.00000   Max.   :1.00000   Max.   :3.0000    Max.   :695.0
##
##  non_origin_ratio    neg_ratio        pos_ratio        neu_ratio
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.1925   1st Qu.:0.3750   1st Qu.:0.3062
##  Median :0.00000   Median :0.2308   Median :0.4183   Median :0.3505
##  Mean   :0.03385   Mean   :0.2326   Mean   :0.4219   Mean   :0.3456
##  3rd Qu.:0.03756   3rd Qu.:0.2662   3rd Qu.:0.4612   3rd Qu.:0.3937
##  Max.   :0.46154   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##  neg_pos_ratio      lyr_ratio       pos_mag_lyrics   neg_mag_lyrics
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.4381   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.5455   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :   Inf   Mean   :0.01291   Mean   :0.3353   Mean   :0.2851
##  3rd Qu.:0.6796   3rd Qu.:0.00755   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :   Inf   Max.   :0.28899   Max.   :4.0000   Max.   :4.0000
##  NA's   :1
##  neu_mag_lyrics pos_mag_quote    neg_mag_quote    neu_mag_quote
##  Min.   :0      Min.   :0.0000   Min.   :0.0000   Min.   :0
##  1st Qu.:0      1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0
##  Median :0      Median :0.0000   Median :0.0000   Median :0
##  Mean   :0      Mean   :0.4265   Mean   :0.3291   Mean   :0
##  3rd Qu.:0      3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0
##  Max.   :0      Max.   :4.0000   Max.   :4.0000   Max.   :0
##
##     pos_mag         neg_mag          neu_mag     cesd_sum
##  Min.   :0.000   Min.   :-3.000   Min.   :0   Min.   : 0.00
##  1st Qu.:1.444   1st Qu.:-1.817   1st Qu.:0   1st Qu.:14.50
##  Median :1.531   Median :-1.667   Median :0   Median :17.00
##  Mean   :1.502   Mean   :-1.609   Mean   :0   Mean   :16.91
##  3rd Qu.:1.613   3rd Qu.:-1.520   3rd Qu.:0   3rd Qu.:20.00
##  Max.   :2.500   Max.   : 0.000   Max.   :0   Max.   :21.00
##
##    pos_l_all          neg_l_all          neu_l_all        cesd_sum_binary
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0
```

```
##  Median :0.000000   Median :0.000000   Median :0.000000   Median :0
##  Mean   :0.005442   Mean   :0.003914   Mean   :0.003559   Mean   :0
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0
##  Max.   :0.181818   Max.   :0.117647   Max.   :0.125000   Max.   :0
##
##   quote_ratio
##  Min.   :0.00000
##  1st Qu.:0.00000
##  Median :0.00000
##  Mean   :0.02093
##  3rd Qu.:0.01962
##  Max.   :0.46154
##
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.



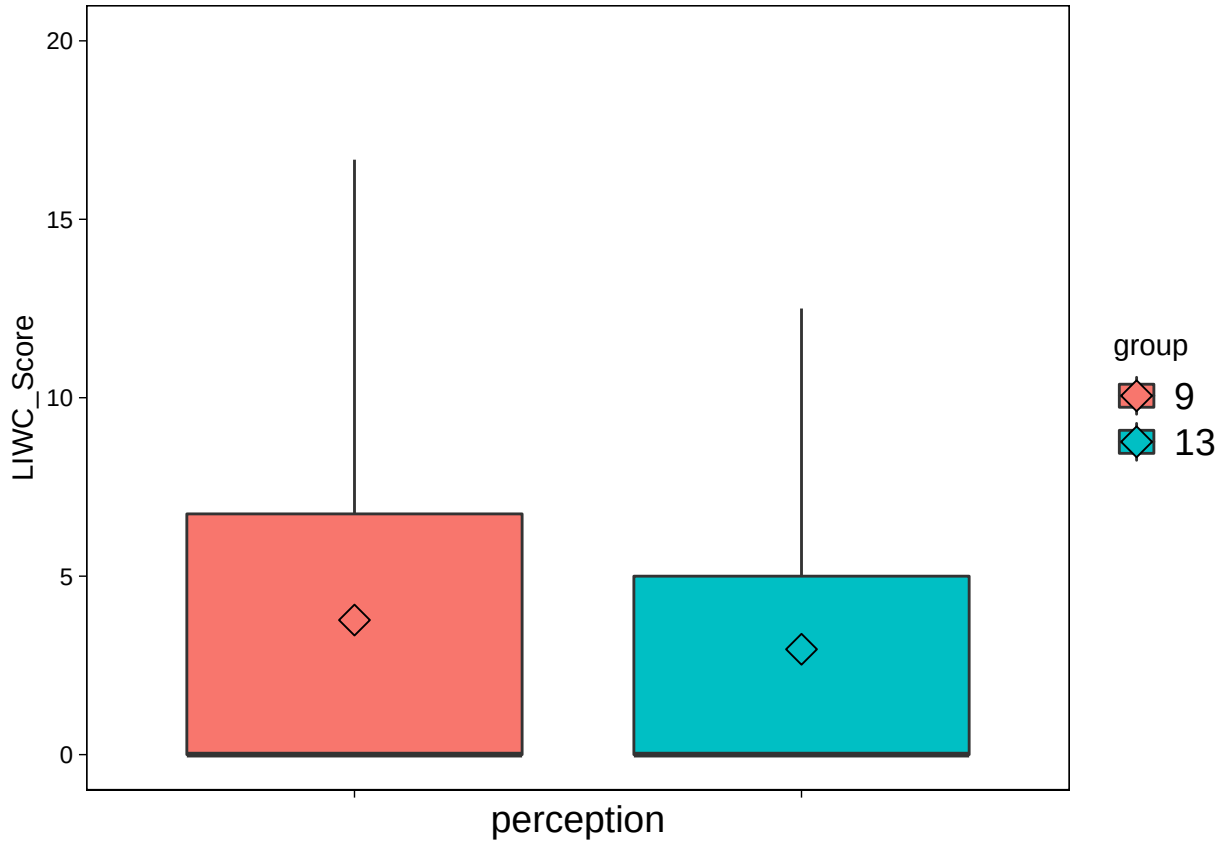

lyrics stats

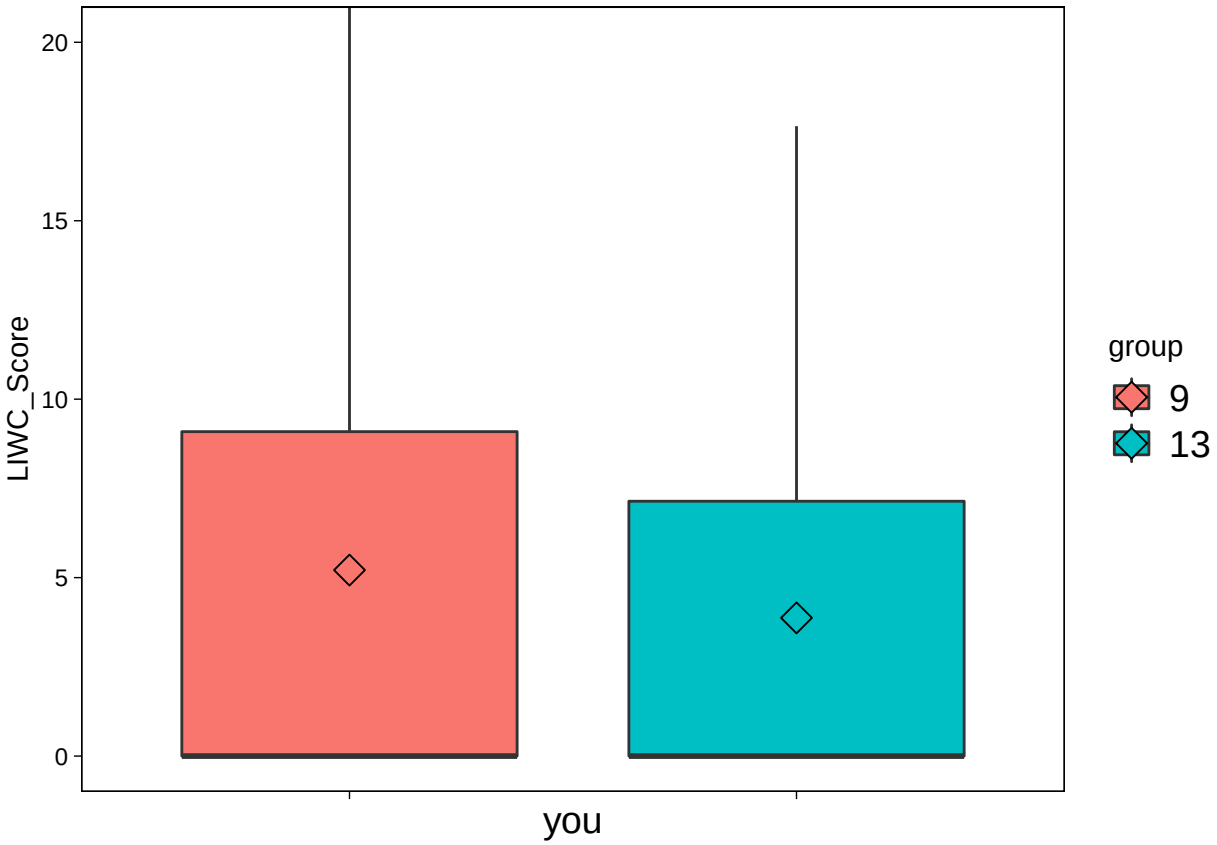
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   2.894   4.350  40.000
```

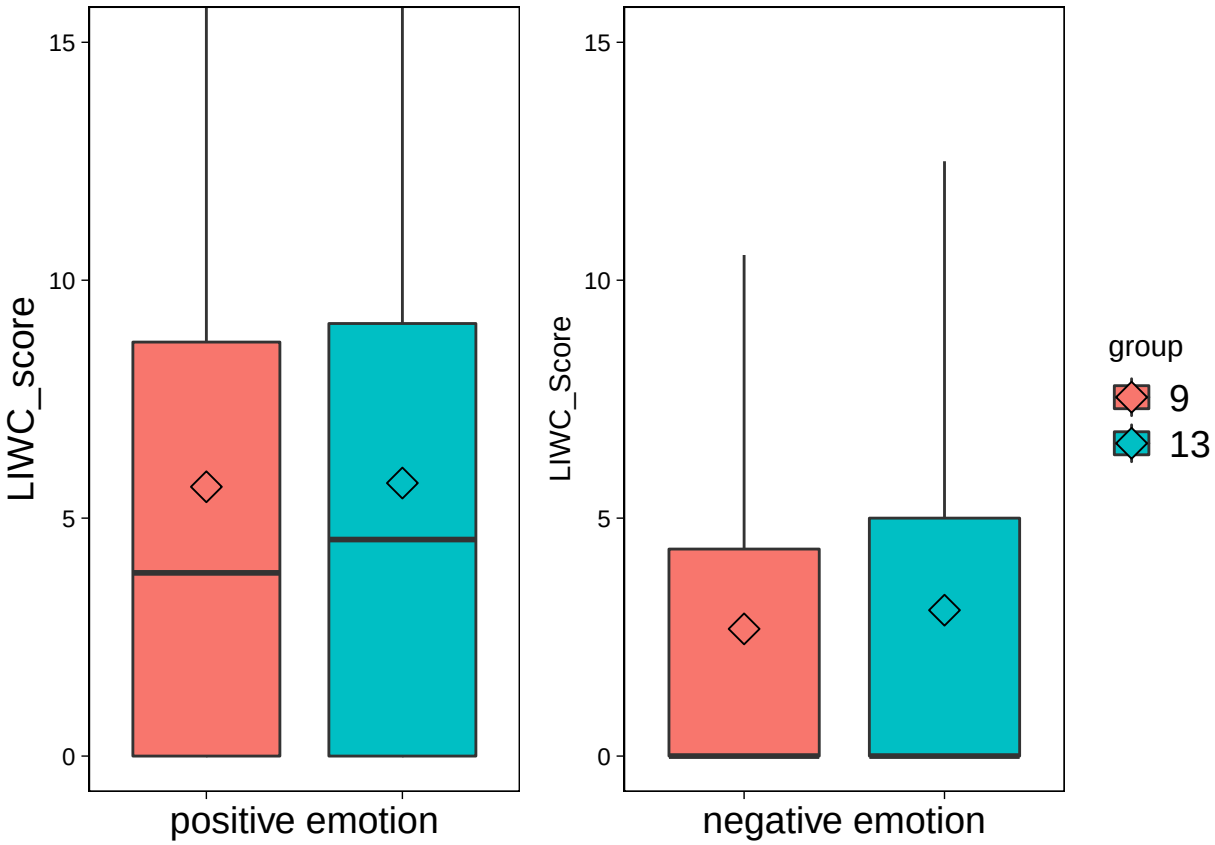
```
##
##  Welch Two Sample t-test
##
## data:  top2$you by top2$dominant_t
## t = 3.3838, df = 467.39, p-value = 0.000775
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.5613954 2.1165030
```

```
## sample estimates:
##  mean in group 9 mean in group 13
##         5.211254         3.872304
```

```
##
##  Welch Two Sample t-test
##
## data:  top2$percept by top2$dominant_t
## t = 2.5086, df = 456.76, p-value = 0.01247
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1772305 1.4590806
## sample estimates:
##  mean in group 9 mean in group 13
##         3.770146         2.951990
```




20
15
10
5
0
LIWC_Score
you
group
9
13


LIWC_score
15
10
5
0
positive emotion
LIWC_Score
15
10
5
0
negative emotion
group
9
13