Buzzfeed Data Pairs Matrix Code

By Max Woolf (http://minimaxir.com)

This notebook is the complement to my blog post Facebook Reactions and the Problem With Quantifying Likes Differently.

```r
options(warn = -1)

# IMPORTANT: This assumes that all packages in "Rstart.R" are installed,
# and the fonts "Source Sans Pro" and "Open Sans Condensed Bold" are installed
# via extrafont. If ggplot2 charts fail to render, you may need to change/remove the theme call.

source("Rstart.R")
library(GGally) # ggpairs

sessionInfo()
```

```
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

Registering fonts with R

Attaching package: ‘’scales

The following objects are masked from ‘package:’readr:

    col_factor, col_numeric


Attaching package: ‘’GGally

The following object is masked from ‘package:’dplyr:

    nasa






R version 3.2.3 (2015-12-10)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.11.3 (El Capitan)
```

```

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] grid      stats     graphics  grDevices utils     datasets  methods  
[8] base     

other attached packages:
[1] GGally_1.0.1       stringr_1.0.0      digest_0.6.8       RColorBrewer_1.1-2
[5] scales_0.3.0       extrafont_0.17     ggplot2_2.0.0      dplyr_0.4.3       
[9] readr_0.1.1       

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.1      Rttf2pt1_1.3.3   magrittr_1.5     munsell_0.4.2   
 [5] uuid_0.1-2       colorspace_1.2-6 R6_2.1.1         plyr_1.8.3      
 [9] tools_3.2.3      parallel_3.2.3   gtable_0.1.2     DBI_0.3.1       
[13] extrafontdb_1.0  assertthat_0.1   IRdisplay_0.3    repr_0.4        
[17] base64enc_0.1-3  IRkernel_0.5     evaluate_0.8     rzmq_0.7.7      
[21] stringi_0.5-5    reshape_0.8.5    jsonlite_0.9.19 
```

```r
df <- read_csv("buzzfeed_data_social_10k.csv")

print(df)
```

```
Source: local data frame [10,388 x 22]

                                                           title
                                                           (chr)
1                        How Well Do You Know Your Banned Books?
2  16 Things F. Scott Fitzgerald Doesn't Want You To Worry About
3           Watch Nick And Amy's Fatal Attraction In "Gone Girl"
4  Alison Bechdel Is The Ultimate Genius "Dyke To Watch Out For"
5                      16 Reasons You'd Probably Die At Hogwarts
6                  19 Banned Books If They Were Made Appropriate
7                               "Zelda's Dreams," By James Franco
8                        How Scandalous Is Your Reading History?
9                     "Gone Girl" Is Now A Sleek But Hollow Movie
10                 17 Things English Majors Are Tired Of Hearing
..                                                           ...
Variables not shown: url (chr), author (chr), date (date), category (chr),
  special (chr), responses (int), num_fb_shares (int), num_tweets (int),
  num_fb_comments (int), love (int), yaaass (int), helpful (int), omg (int),
  lol (int), cute (int), win (int), wtf (int), fail (int), trashy (int), ew
  (int), hate (int)
```

Select only the columns with reaction data, and get spot correlations.

```r
df_reactions <- na.omit(df %>% select(love:hate))

print(df_reactions)

print(cor(df_reactions))
```

```
Source: local data frame [9,883 x 12]
```

```

    love yaaass helpful   omg   lol  cute   win   wtf  fail trashy    ew  hate
   (int)  (int)   (int) (int) (int) (int) (int) (int) (int)  (int) (int) (int)
1     31      0       3     7     1     1     3     5     4      0     0     1
2    110      0       0     2     9    17    18     7     0      1     0     0
3      5      0       0     0     0     0     2     0     0      0     0     0
4     16      0       0     0     0     0     1     0     0      0     0     0
5     72      0       0     2    25     1     4     0     4      0     0     0
6     44      7       0     4    20     1     8     3     7      1     0     0
7     25      0       0     0     0     0     0     7     2      0     0     0
8    139      2       1     5    10     1    20     1     0      2     0     1
9     19      0       0     2     2     0     1     0     0      0     0     0
10   119     23       2     3    22     1    25     0     1      0     0     0
..   ...    ...     ...   ...   ...   ...   ...   ...   ...    ...   ...   ...
              love     yaaass     helpful        omg        lol         cute
love    1.00000000 0.46626799 0.124755232 0.68036925 0.47360895  0.629094452
yaaass  0.46626799 1.00000000 0.175511580 0.35403737 0.26705946  0.096387912
helpful 0.12475523 0.17551158 1.000000000 0.04352179 0.01926325  0.008081270
omg     0.68036925 0.35403737 0.043521787 1.00000000 0.38471634  0.539838706
lol     0.47360895 0.26705946 0.019263247 0.38471634 1.00000000  0.305064425
cute    0.62909445 0.09638791 0.008081270 0.53983871 0.30506443  1.000000000
win     0.83126618 0.45288311 0.114922581 0.59319268 0.43868351  0.523278287
wtf     0.09907593 0.05272187 0.022267654 0.31346725 0.20007750  0.008643063
fail    0.07005368 0.07472599 0.021095192 0.18130431 0.17963674 -0.031200420
trashy  0.03739368 0.09077292 0.014492817 0.13685420 0.09558570 -0.031452816
ew      0.05038921 0.10098157 0.009602044 0.20792642 0.11147785 -0.024514311
hate    0.15831206 0.02737651 0.015482722 0.27294572 0.05448569  0.007341387
               win         wtf        fail      trashy           ew        hate
love    0.83126618 0.099075927  0.07005368  0.03739368  0.050389209 0.158312061
yaaass  0.45288311 0.052721871  0.07472599  0.09077292  0.100981567 0.027376513
helpful 0.11492258 0.022267654  0.02109519  0.01449282  0.009602044 0.015482722
omg     0.59319268 0.313467249  0.18130431  0.13685420  0.207926422 0.272945720
lol     0.43868351 0.200077499  0.17963674  0.09558570  0.111477851 0.054485686
cute    0.52327829 0.008643063 -0.03120042 -0.03145282 -0.024514311 0.007341387
win     1.00000000 0.061382292  0.04877020  0.02347292  0.023725465 0.070561338
wtf     0.06138229 1.000000000  0.63592405  0.50851441  0.566388147 0.332060843
fail    0.04877020 0.635924055  1.00000000  0.51560199  0.505881072 0.348757439
trashy  0.02347292 0.508514410  0.51560199  1.00000000  0.805459962 0.255968387
ew      0.02372546 0.566388147  0.50588107  0.80545996  1.000000000 0.255072265
hate    0.07056134 0.332060843  0.34875744  0.25596839  0.255072265 1.000000000
```

Note that the `helpful` and `trashy` reactions are not used in 2016, so we will not use them.

Use `ggpairs` to plot multidimensional data (lower and diag functions adapted from the GGally package viginette; upper correlation function adopted from Barret Schloerke on GitHub).

```
pairs_theme <- function (x) {
                theme_bw(base_size = 5) +
                theme(panel.grid.minor.x = element_blank()) +
                theme(panel.grid.minor.y = element_blank())
                }


gglower <- function(data, mapping, ..., high = "#c0392b") {
  ggplot(data = data, mapping = mapping) +
```

```r
    geom_bin2d(...) +
    scale_x_log10(limits=c(10^0,10^3), breaks=10^(0:3)) +
    scale_y_log10(limits=c(10^0,10^3), breaks=10^(0:3)) +
    geom_smooth(alpha = 0.5, size = 0.25, color = "#1a1a1a", method = "lm") +
    scale_fill_gradient(low = "#EEEEEE", high = high, trans = "log") +
    pairs_theme()
}

ggdiag <- function(data, mapping, ..., color = "#1a1a1a") {
  ggplot(data = data, mapping = mapping) +
    geom_density(..., color = color) +
    scale_x_log10(limits=c(10^0,10^3), breaks=10^(0:3)) +
    pairs_theme()
}

# From https://github.com/ggobi/ggally/issues/139#issuecomment-176271618

ggupper <- function(data, mapping, color = I("grey50"), sizeRange = c(1, 3), ...) {

  # get the x and y data to use the other code
  x <- eval(mapping$x, data)
  y <- eval(mapping$y, data)

  ct <- cor.test(x,y)
  sig <- symnum(
    ct$p.value, corr = FALSE, na = FALSE,
    cutpoints = c(0, 0.001, 0.01, 0.05, 0.1, 1),
    symbols = c("***", "**", "*", ".", " ")
  )

  r <- unname(ct$estimate)
  rt <- format(r, digits=2)[1]

  # since we can't print it to get the strsize, just use the max size range
  cex <- max(sizeRange)

  # helper function to calculate a useable size
  percent_of_range <- function(percent, range) {
    percent * diff(range) + min(range, na.rm = TRUE)
  }

  # plot the cor value
  ggally_text(
    label = as.character(rt),
    mapping = aes(),
    xP = 0.5, yP = 0.5,
    size = I(percent_of_range(cex * abs(r), sizeRange)),
    color = color,
    ...
  ) +
    # add the sig stars
    geom_text(
      aes_string(
        x = 0.8,
    y = 0.8
   ),
   label = sig,
   size = I(cex),
   color = color,
   ...
  ) +
  pairs_theme() +
  theme(panel.grid.major.x = element_blank()) +
  theme(panel.grid.major.y = element_blank())

}
```

```
pos_color <- "#27ae60"

plot <- ggpairs(df_reactions, columns = c("love", "yaaass", "omg", "lol", "cute", "win"),
    title = sprintf("Pairs Plot of Positive Reaction Counts on %00d BuzzFeed Articles",
      nrow(df_reactions)),
    upper = list(continuous = wrap(ggupper, color = pos_color)),
    lower = list(continuous = wrap(gglower, high = pos_color)),
    diag = list(continuous = wrap(ggdiag, color = pos_color))) +
    theme(title = element_text(size=10))

png("buzzfeed-pos.png", w=1600, h=1600, res=300)
plot
dev.off()
```

pdf: 2

```
neg_color <- "#c0392b"

plot <- ggpairs(df_reactions, columns = c("love", "wtf", "fail", "ew", "hate"),
    title = sprintf("Pairs Plot of Love + Negative Reaction Counts on %00d BuzzFeed
      Articles", nrow(df_reactions)),
    upper = list(continuous = wrap(ggupper, color = neg_color )),
    lower = list(continuous = wrap(gglower, high = neg_color)),
    diag = list(continuous = wrap(ggdiag, color = neg_color))) +
    theme(title = element_text(size=10))

png("buzzfeed-neg.png", w=1600, h=1600, res=300)
plot
dev.off()
```

pdf: 2

The MIT License (MIT)

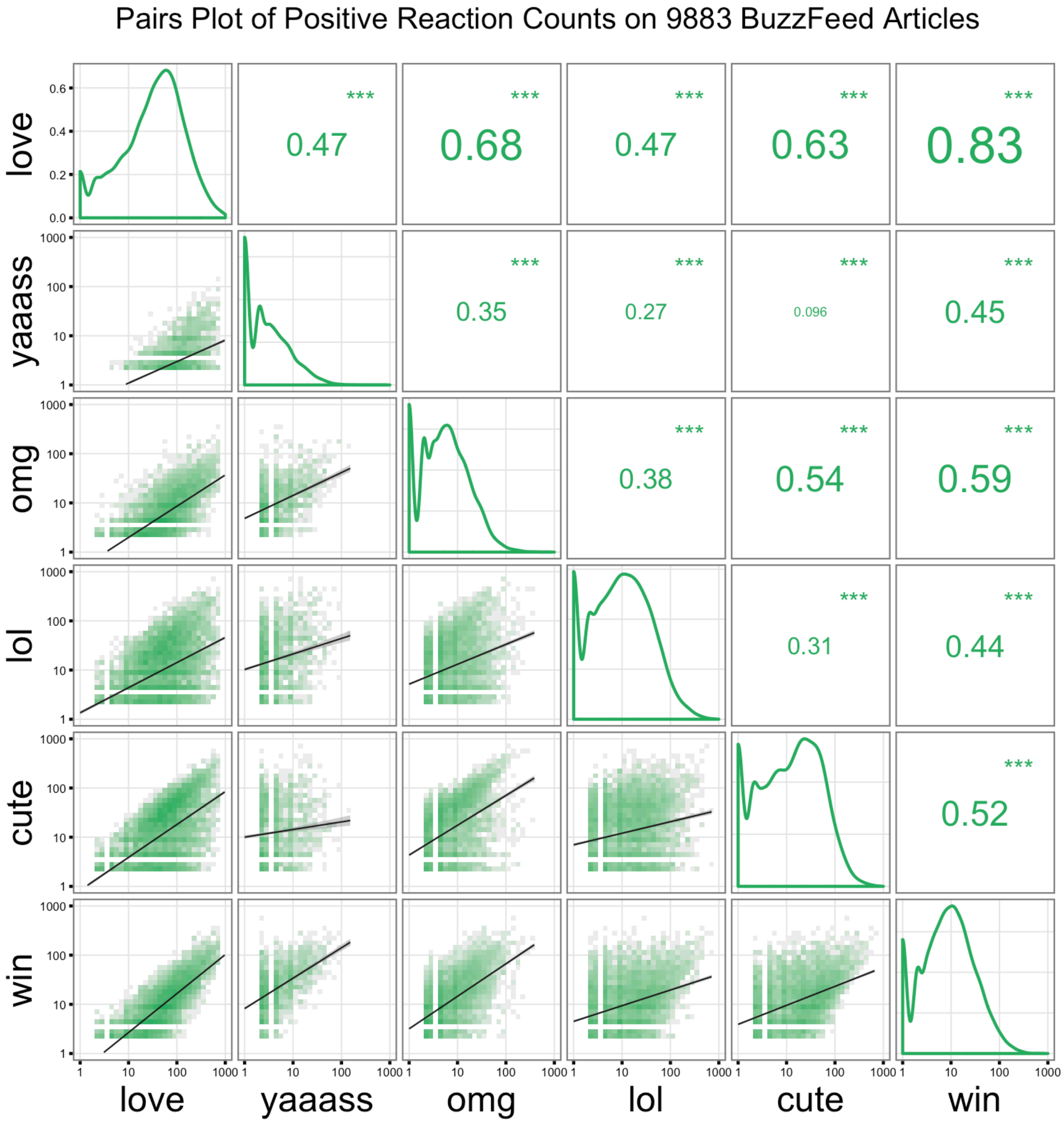
Pairs Plot of Positive Reaction Counts on 9883 BuzzFeed Articles
love
yaaass
omg
lol
cute
win
0.0
0.2
0.4
0.6
1
10
100
1000
*** 0.47
*** 0.68
*** 0.47
*** 0.63
*** 0.83
*** 0.35
*** 0.27
*** 0.096
*** 0.45
*** 0.38
*** 0.54
*** 0.59
*** 0.31
*** 0.44
*** 0.52
love
yaaass
omg
lol
cute
win

Figure 1:

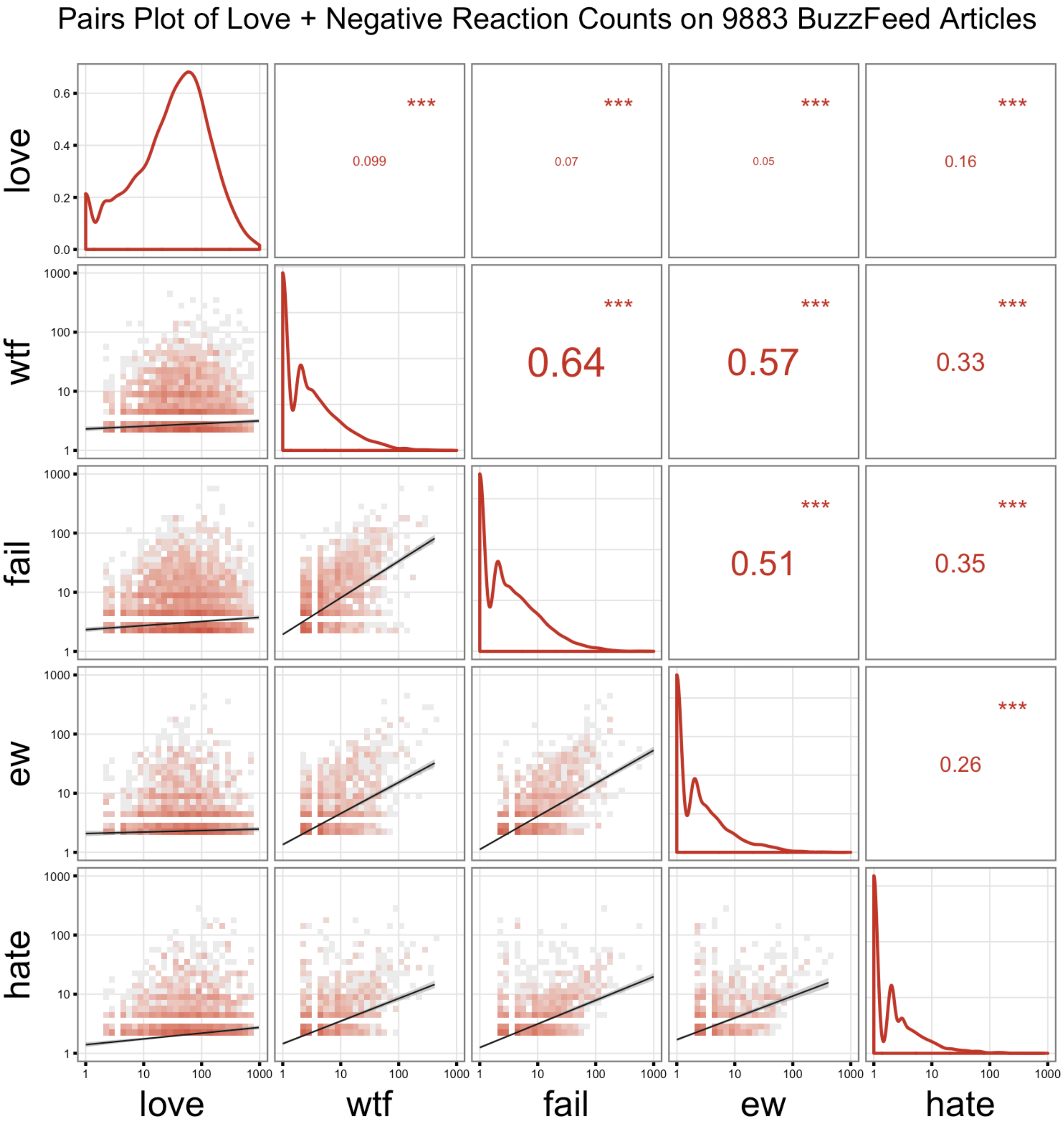
Pairs Plot of Love + Negative Reaction Counts on 9883 BuzzFeed Articles
love
wtf
fail
ew
hate
0.0
0.2
0.4
0.6
1
10
100
1000
***
0.099
0.07
0.05
0.16
0.64
0.57
0.33
0.51
0.35
0.26

Figure 2: