Krysten_241_model

```
d <- read.csv("craigslist.csv")
#d <- read.csv("craigslist.csv", stringsAsFactors = FALSE)
```

```
head(d)
```

```
##   Email.sent.date Email.sent.time sent_timestamp
## 1         3/21/19         8:44 PM  3/21/19 20:44
## 2         3/21/19         8:47 PM  3/21/19 20:47
## 3         3/21/19         8:57 PM  3/21/19 20:57
## 4         3/21/19         9:00 PM  3/21/19 21:00
## 5         3/21/19         9:10 PM  3/21/19 21:10
## 6         3/21/19         9:52 PM  3/21/19 21:52
##                                                                                  Posting.URL
## 1   https://indianapolis.craigslist.org/roo/d/indianapolis-share-house-near-airport/6847331771.html
## 2     https://indianapolis.craigslist.org/roo/d/greenwood-2-bedroom-1-bath-to-share/6847218336.html
## 3 https://indianapolis.craigslist.org/roo/d/indianapolis-furnished-room-in-adorable/6847161772.html
## 4                https://indianapolis.craigslist.org/roo/d/indianapolis-room-mates/6847167359.html
## 5    https://indianapolis.craigslist.org/roo/d/indianapolis-roommate-wanted-in-nice/6847102799.html
## 6               https://indianapolis.craigslist.org/roo/d/hanover-looking-roomates/6847136751.html
##                                                                                    Posting.Title
## 1                        $600 Share house near airport - month to month (West Washington street)
## 2                                     $420 / 820ft2 - 2 bedroom 1 bath to share (Greenwood)
## 3 $425 Furnished Room in Adorable Bungalow, Everything Included! (N Bosart Ave Indianapolis, IN)
## 4                                                          $400 Room mates (Indianapolis)
## 5                                     $360 Roommate wanted in NICE house (Indianapolis, IN)
## 6                                          $500 Looking roomates (Indianapolis,in)
##
## 1
## 2
## 3
## 4
## 5
## 6 Looking to rent a room ... LOCATION MADISON, INDIANA, About us we are a couple with one small chil
##   Post.timestamp Listing.Price Listing.Type..house..apt...etc.
## 1  3/21/19 22:38           600                          house
## 2  3/21/19 19:15           420                      apartment
## 3  3/21/19 18:00           425                          house
## 4  3/21/19 18:07           400                           <NA>
## 5  3/21/19 16:51           360                          house
## 6  3/21/19 17:30           500                          house
##   Poster.Gender              Poster.Age Treatment Reply. Favorable.Reply.
## 1          <NA>                    <NA>  katie_43      Y                Y
## 2        female                    <NA>  katie_43      Y                Y
## 3          <NA> late 20s/mid 30s couple  katie_43      N                N
## 4          <NA>                    <NA>  katie_43      N                N
## 5          <NA>                    <NA>  katie_43      N                N
## 6          male                    <NA>  katie_27      Y                Y
##   reply_date reply_time Gender.in.email Age.in.email
## 1   3/22/19 5:49:00 AM                            NA
## 2   3/23/19 1:53:00 PM                            NA
## 3       <NA>                              NA
```

```
## 4       <NA>                                    NA
## 5       <NA>                                    NA
## 6    3/22/19 8:51:00 AM presumed female         NA
```

```
dim(d)
```

```
## [1] 124  18
```

```
colSums(is.na(d))
```

```
##                 Email.sent.date                 Email.sent.time
##                               0                               0
##                  sent_timestamp                     Posting.URL
##                               0                               0
##                   Posting.Title               Posting.Body.Text
##                               0                               0
##                  Post.timestamp                   Listing.Price
##                               0                               0
## Listing.Type..house..apt...etc.                   Poster.Gender
##                              12                              82
##                      Poster.Age                       Treatment
##                             102                               0
##                          Reply.                 Favorable.Reply.
##                               0                               0
##                      reply_date                      reply_time
##                              62                               0
##                 Gender.in.email                    Age.in.email
##                               0                             123
```

```
#rename columns for clarity, ease of modeling
colnames(d)[1] <- "email_sent_date"
colnames(d)[2] <- "email_sent_time"
colnames(d)[8] <- "list_price"
colnames(d)[9] <- "list_type"
colnames(d)[12] <- "katie_27"
colnames(d)[13] <- "reply"
colnames(d)[14] <- "fave_reply"
```

```
colnames(d)
```

```
##  [1] "email_sent_date"   "email_sent_time"   "sent_timestamp"
##  [4] "Posting.URL"       "Posting.Title"     "Posting.Body.Text"
##  [7] "Post.timestamp"    "list_price"        "list_type"
## [10] "Poster.Gender"     "Poster.Age"        "katie_27"
## [13] "reply"             "fave_reply"        "reply_date"
## [16] "reply_time"        "Gender.in.email"   "Age.in.email"
```

```
colSums(is.na(d))
```

```
##   email_sent_date   email_sent_time    sent_timestamp       Posting.URL
##                 0                 0                 0                 0
##     Posting.Title Posting.Body.Text    Post.timestamp        list_price
##                 0                 0                 0                 0
##         list_type     Poster.Gender        Poster.Age          katie_27
##                12                82               102                 0
##             reply        fave_reply        reply_date        reply_time
##                 0                 0                62                 0
##   Gender.in.email      Age.in.email
```

```
##                 0               123
```

```
#create new subsetted dataframe
d_sub <- d[c(1:2, 8:9, 12:16)]
head(d_sub)
```

```
##   email_sent_date email_sent_time list_price list_type katie_27 reply
## 1         3/21/19         8:44 PM        600     house katie_43     Y
## 2         3/21/19         8:47 PM        420 apartment katie_43     Y
## 3         3/21/19         8:57 PM        425     house katie_43     N
## 4         3/21/19         9:00 PM        400      <NA> katie_43     N
## 5         3/21/19         9:10 PM        360     house katie_43     N
## 6         3/21/19         9:52 PM        500     house katie_27     Y
##   fave_reply reply_date reply_time
## 1          Y    3/22/19 5:49:00 AM
## 2          Y    3/23/19 1:53:00 PM
## 3          N       <NA>
## 4          N       <NA>
## 5          N       <NA>
## 6          Y    3/22/19 8:51:00 AM
```

```
str(d_sub)
```

```
## 'data.frame':    124 obs. of  9 variables:
##  $ email_sent_date: Factor w/ 8 levels "3/21/19","3/24/19",..: 1 1 1 1 1 1 1 1 1 ...
##  $ email_sent_time: Factor w/ 88 levels "1:06 PM","1:20 PM",..: 71 72 78 79 80 86 87 82 88 5 ...
##  $ list_price     : int  600 420 425 400 360 500 450 500 400 600 ...
##  $ list_type      : Factor w/ 4 levels "apartment","condo",..: 3 1 3 NA 3 3 3 1 3 1 ...
##  $ katie_27       : Factor w/ 3 levels "katie_27","katie_27 ",..: 3 3 3 3 3 1 1 3 1 1 ...
##  $ reply          : Factor w/ 2 levels "N","Y": 2 2 1 1 1 2 2 1 1 1 ...
##  $ fave_reply     : Factor w/ 2 levels "N","Y": 2 2 1 1 1 2 2 1 1 1 ...
##  $ reply_date     : Factor w/ 13 levels "3/22/19","3/23/19",..: 1 2 NA NA NA 1 1 NA NA NA ...
##  $ reply_time     : Factor w/ 52 levels "","1:02:00 PM",..: 31 5 1 1 1 48 52 1 1 1 ...
```

```
summary(d_sub$list_type)
```

```
## apartment     condo     house townhouse      NA's
##        34         2        73         3        12
```

```
# There are extra white spaces in 'katie_27', need to remove; returns string w/o leading or trailing wh
trim <- function (x) gsub("^\\s+|\\s+$", "", x)
d_sub$katie_27 <- trim(d_sub$katie_27)
str(d_sub$katie_27)
```

```
##  chr [1:124] "katie_43" "katie_43" "katie_43" "katie_43" "katie_43" ...
```

```
#binarized 'list_type', 'katie_27', 'reply', and 'fave_reply'
d_sub <- within(d_sub, {
  list_type = ifelse(list_type == "house", 1, 0)
  katie_27 = ifelse(katie_27 == "katie_27", 1, 0)
  reply = ifelse(reply == "Y", 1, 0)
  fave_reply = ifelse(fave_reply == "Y", 1, 0)
})

head(d_sub, 25)
```

```
##    email_sent_date email_sent_time list_price list_type katie_27 reply
## 1          3/21/19         8:44 PM        600         1        0     1
```

```
## 2          3/21/19          8:47 PM        420          0        0     1
## 3          3/21/19          8:57 PM        425          1        0     0
## 4          3/21/19          9:00 PM        400         NA        0     0
## 5          3/21/19          9:10 PM        360          1        0     0
## 6          3/21/19          9:52 PM        500          1        1     1
## 7          3/21/19          9:53 PM        450          1        1     1
## 8          3/21/19          9:42 PM        500          0        0     0
## 9          3/21/19          9:56 PM        400          1        1     0
## 10         3/21/19         10:00 PM        600          0        1     0
## 11         3/21/19          9:43 AM        550          1        0     0
## 12         3/21/19         10:00 PM        540          1        1     0
## 13         3/21/19         10:02 PM        390          0        1     0
## 14         3/21/19         10:03 PM        400         NA        1     1
## 15         3/21/19         10:05 PM        550         NA        1     1
## 16         3/21/19         10:06 PM        400          0        1     0
## 17         3/21/19          9:44 PM        525          1        0     0
## 18         3/21/19         10:07 PM        380          0        1     0
## 19         3/21/19          9:35 PM        400          1        0     0
## 20         3/21/19          9:45 PM        550         NA        0     1
## 21         3/21/19          8:44 PM        600          1        0     1
## 22         3/21/19          8:47 PM        420          0        0     1
## 23         3/21/19          8:57 PM        425          1        0     0
## 24         3/21/19          9:00 PM        400         NA        0     0
## 25         3/21/19          9:10 PM        360          1        0     0
##    fave_reply reply_date reply_time
## 1           1    3/22/19 5:49:00 AM
## 2           1    3/23/19 1:53:00 PM
## 3           0       <NA>
## 4           0       <NA>
## 5           0       <NA>
## 6           1    3/22/19 8:51:00 AM
## 7           1    3/22/19 9:51:00 AM
## 8           0       <NA>
## 9           0       <NA>
## 10          0       <NA>
## 11          0       <NA>
## 12          0       <NA>
## 13          0       <NA>
## 14          1    3/23/19 6:32:00 AM
## 15          1    3/23/19 6:17:00 AM
## 16          0       <NA>
## 17          0       <NA>
## 18          0       <NA>
## 19          0       <NA>
## 20          1    3/22/19 5:32:00 AM
## 21          1    3/22/19 5:49:00 AM
## 22          1    3/23/19 1:53:00 PM
## 23          0       <NA>
## 24          0       <NA>
## 25          0       <NA>
```

```r
#changed NA's in 'list_type' to 0
d_sub[c("list_type")][is.na(d_sub[c("list_type")])] <- 0
```

```r
colSums(is.na(d_sub))
```

```
## email_sent_date email_sent_time      list_price       list_type
##               0               0               0               0
##        katie_27           reply      fave_reply      reply_date
##               0               0               0              62
##      reply_time
##               0
```

```r
str(d_sub)
```

```
## 'data.frame':    124 obs. of  9 variables:
##  $ email_sent_date: Factor w/ 8 levels "3/21/19","3/24/19",..: 1 1 1 1 1 1 1 1 1 ...
##  $ email_sent_time: Factor w/ 88 levels "1:06 PM","1:20 PM",..: 71 72 78 79 80 86 87 82 88 5 ...
##  $ list_price     : int  600 420 425 400 360 500 450 500 400 600 ...
##  $ list_type      : num  1 0 1 0 1 1 1 0 1 0 ...
##  $ katie_27       : num  0 0 0 0 0 1 1 0 1 1 ...
##  $ reply          : num  1 1 0 0 0 1 1 0 0 0 ...
##  $ fave_reply     : num  1 1 0 0 0 1 1 0 0 0 ...
##  $ reply_date     : Factor w/ 13 levels "3/22/19","3/23/19",..: 1 2 NA NA NA 1 1 NA NA NA ...
##  $ reply_time     : Factor w/ 52 levels "","1:02:00 PM",..: 31 5 1 1 1 48 52 1 1 1 ...
```

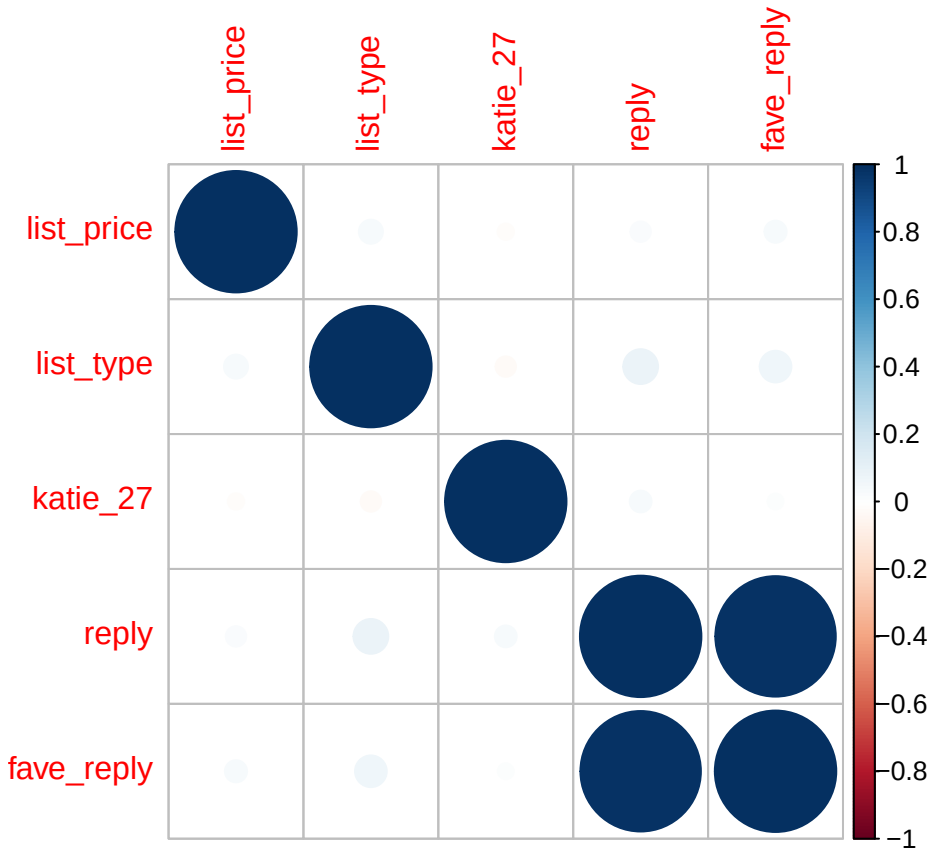
```r
a <- d_sub[, c('list_price', 'list_type', 'katie_27',  'reply', 'fave_reply')]
b <- round(cor(a), 3)
b
```

```
##             list_price list_type katie_27 reply fave_reply
## list_price       1.000     0.039   -0.018 0.028      0.033
## list_type        0.039     1.000   -0.028 0.082      0.068
## katie_27        -0.018    -0.028    1.000 0.032      0.017
## reply            0.028     0.082    0.032 1.000      0.984
## fave_reply       0.033     0.068    0.017 0.984      1.000
```

```r
library(corrplot)
```

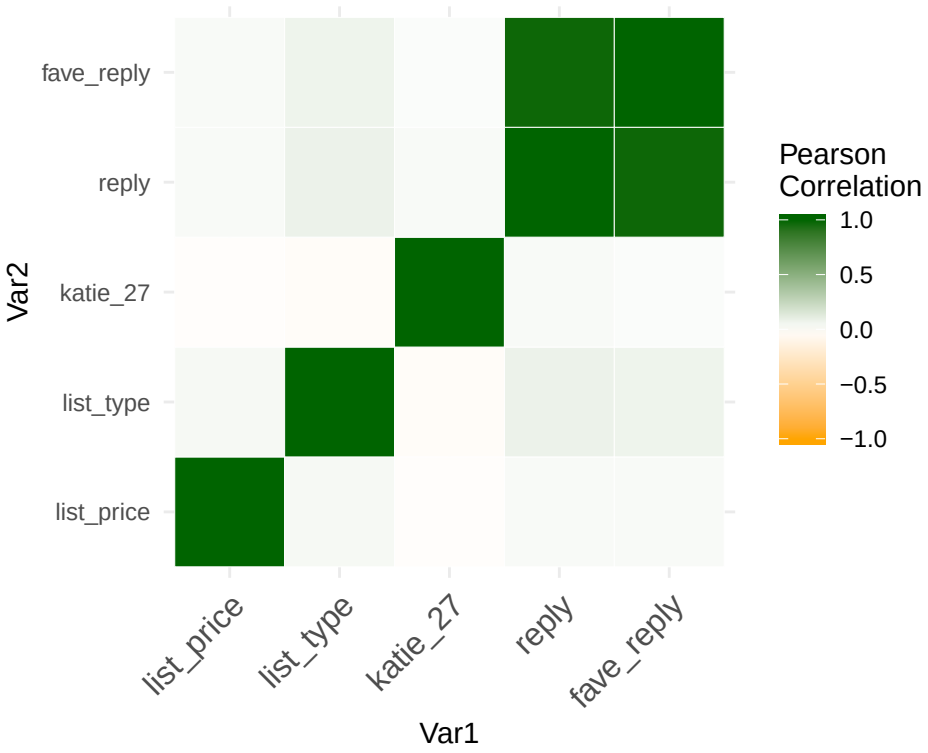
```
## corrplot 0.84 loaded
```

```r
corrplot(b)
```




```
#Correlations

a <- d_sub[, c('list_price', 'list_type', 'katie_27',  'reply', 'fave_reply')]
a.cormat <- round(cor(a), 2)
melt.a <- melt(a.cormat)
#head(melt.speed)
ggplot(data=melt.a, aes(x=Var1, y=Var2, fill=value)) + geom_tile(color="white") +  scale_fill_gradient2
   midpoint = 0, limit = c(-1,1), space = "Lab",
   name="Pearson\nCorrelation") +
  theme_minimal()+
 theme(axis.text.x = element_text(angle = 45, vjust = 1,
    size = 12, hjust = 1))+
 coord_fixed()
```




```
mod.1 <- glm(reply ~ katie_27 + list_price + list_type, family = binomial(link = logit), data=d_sub)
summary(mod.1)
```

```
##
## Call:
## glm(formula = reply ~ katie_27 + list_price + list_type, family = binomial(link = logit),
##     data = d_sub)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.33855  -1.18536   0.00924   1.15068   1.34358
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.4424976  0.6796809  -0.651    0.515
## katie_27     0.1417980  0.3617879   0.392    0.695
## list_price   0.0003426  0.0011872   0.289    0.773
## list_type    0.3344524  0.3671640   0.911    0.362
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 171.90  on 123  degrees of freedom
## Residual deviance: 170.83  on 120  degrees of freedom
## AIC: 178.83
##
## Number of Fisher Scoring iterations: 3
```

```
#model with interactions
mod.2 <- glm(reply ~ katie_27 + list_price + list_type + katie_27:list_price + katie_27: list_type, fami
summary(mod.2)
```

```
##
## Call:
## glm(formula = reply ~ katie_27 + list_price + list_type + katie_27:list_price +
##     katie_27:list_type, family = binomial(link = logit), data = d_sub)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.04168  -1.07000   0.08143   1.08020   1.63894
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           1.007746   0.922958   1.092   0.2749
## katie_27             -3.011077   1.412862  -2.131   0.0331 *
## list_price           -0.001508   0.001655  -0.911   0.3625
## list_type            -0.540733   0.546151  -0.990   0.3221
## katie_27:list_price   0.004373   0.002580   1.695   0.0900 .
## katie_27:list_type    1.672546   0.759717   2.202   0.0277 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 171.9  on 123  degrees of freedom
## Residual deviance: 162.7  on 118  degrees of freedom
## AIC: 174.7
##
## Number of Fisher Scoring iterations: 4
```