

Logistic regression in R

Erin Shellman

May 04, 2015

Contents

Exploratory data analysis	1
Logistic regression	9

For the second project we'll explore user data from Twitter. The data set has a variable called `bot` that denotes whether the user is a bot, and then some features about the user account.

```
library(dplyr)
library(ggplot2)
library(lubridate)
library(GGally)
library(scales)
library(caret)

data = read.delim('bot_or_not.tsv',
                  sep = '\t',
                  header = TRUE)
```

Exploratory data analysis

We've got a brand new dataset here, so let's familiarize ourselves by conducting an exploratory data analysis. First, let's make sure the variables are of the right types, *i.e.* continuous variables are continuous and dummy variables are factors.

```
# tell R which variables are categorical (factors)
data$bot = factor(data$bot)
data$default_profile = factor(data$default_profile)
data$default_profile_image = factor(data$default_profile_image)
data$geo_enabled = factor(data$geo_enabled)
data$profile_background_tile = factor(data$profile_background_tile)
data$verified = factor(data$verified)
```

Now, we can explore the data a bit.

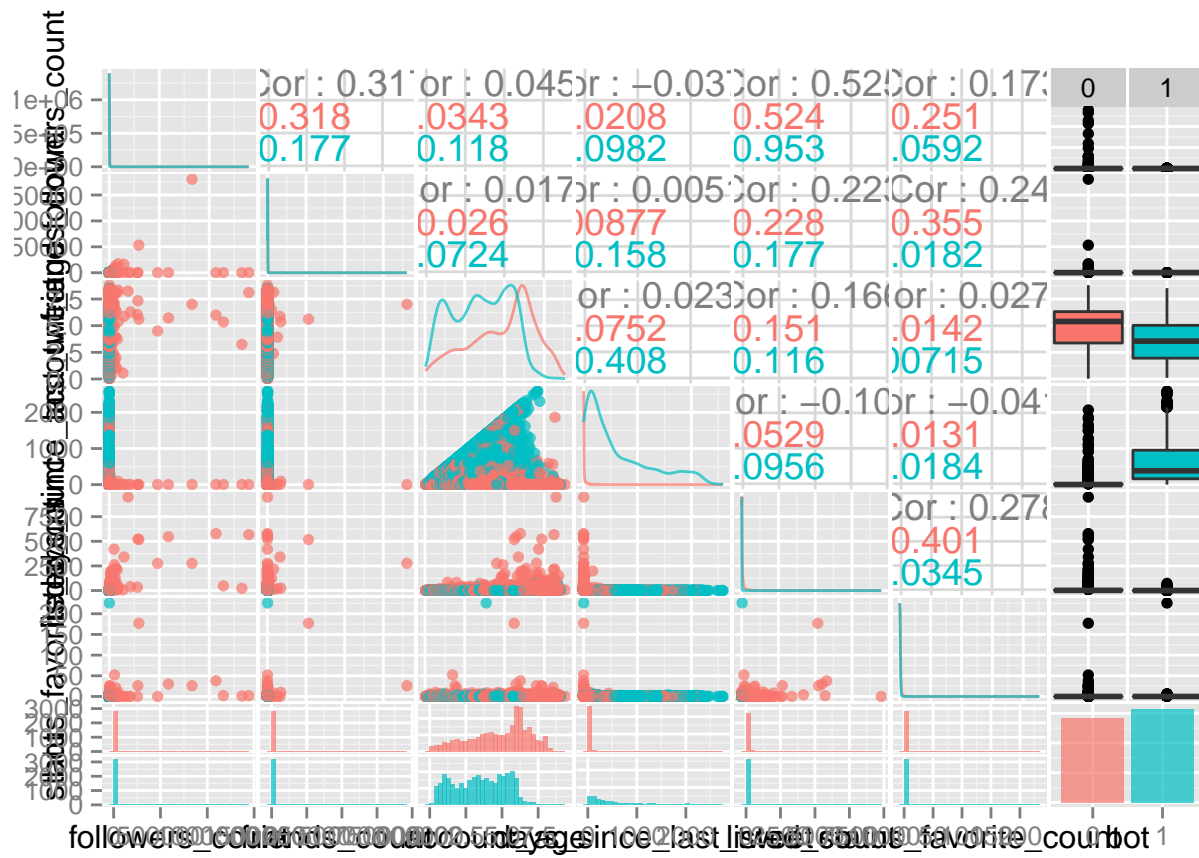
```
summary(data)
```

##	id	bot	followers_count	friends_count
##	Min. :9.630e+02	0:2837	Min. : 0	Min. : 1.0
##	1st Qu.:5.502e+07	1:3146	1st Qu.: 6	1st Qu.: 645.5
##	Median :2.411e+08		Median : 32	Median : 1073.0
##	Mean :6.636e+08		Mean : 2193	Mean : 1321.3
##	3rd Qu.:8.668e+08		3rd Qu.: 324	3rd Qu.: 1448.0
##	Max. :3.166e+09		Max. :1396699	Max. :902454.0

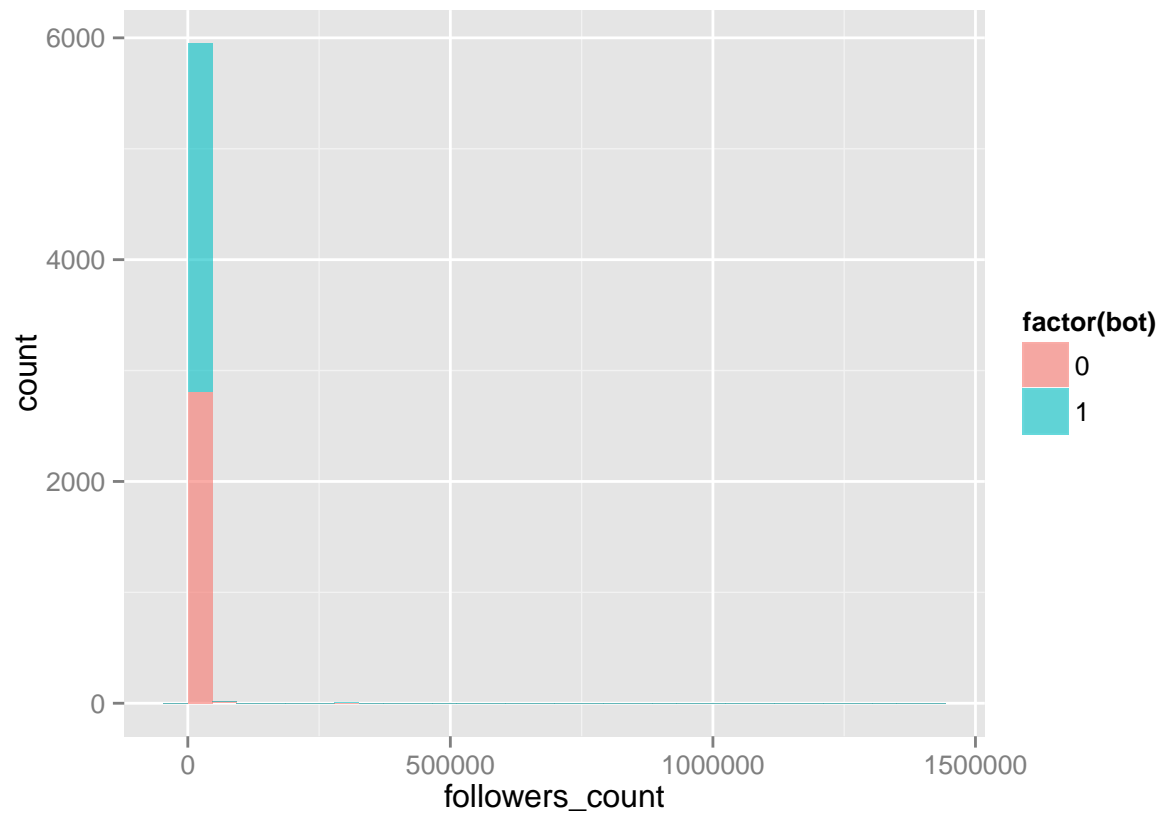
```
## default_profile default_profile_image favourites_count geo_enabled
## 0:3253          0:1566          Min.   :    0.0  0:4341
## 1:2730          1:4417          1st Qu.:    0.0  1:1642
##                                     Median :   15.0
##                                     Mean    :  335.9
##                                     3rd Qu.:  531.0
##                                     Max.    :113640.0
## account_age      days_since_last_tweet listed_count
## Min.   :0.03143   Min.    :  0.7733   Min.    :  0.00
## 1st Qu.:2.55717   1st Qu.:  2.7112   1st Qu.:  0.00
## Median :4.27858   Median : 76.7645   Median :  1.00
## Mean    :4.14829   Mean    :341.5358   Mean    : 44.36
## 3rd Qu.:5.81762   3rd Qu.:433.5421   3rd Qu.: 18.00
## Max.    :8.80368   Max.    :2589.0560   Max.    :9491.00
## profile_background_tile status_favorite_count verified
## 0:5130          Min.    :  0.0000   0:5933
## 1: 853          1st Qu.:  0.0000   1: 50
##                                     Median :  0.0000
##                                     Mean    :  0.3744
##                                     3rd Qu.:  0.0000
##                                     Max.    :226.0000
```

```
# inspect the trends
```

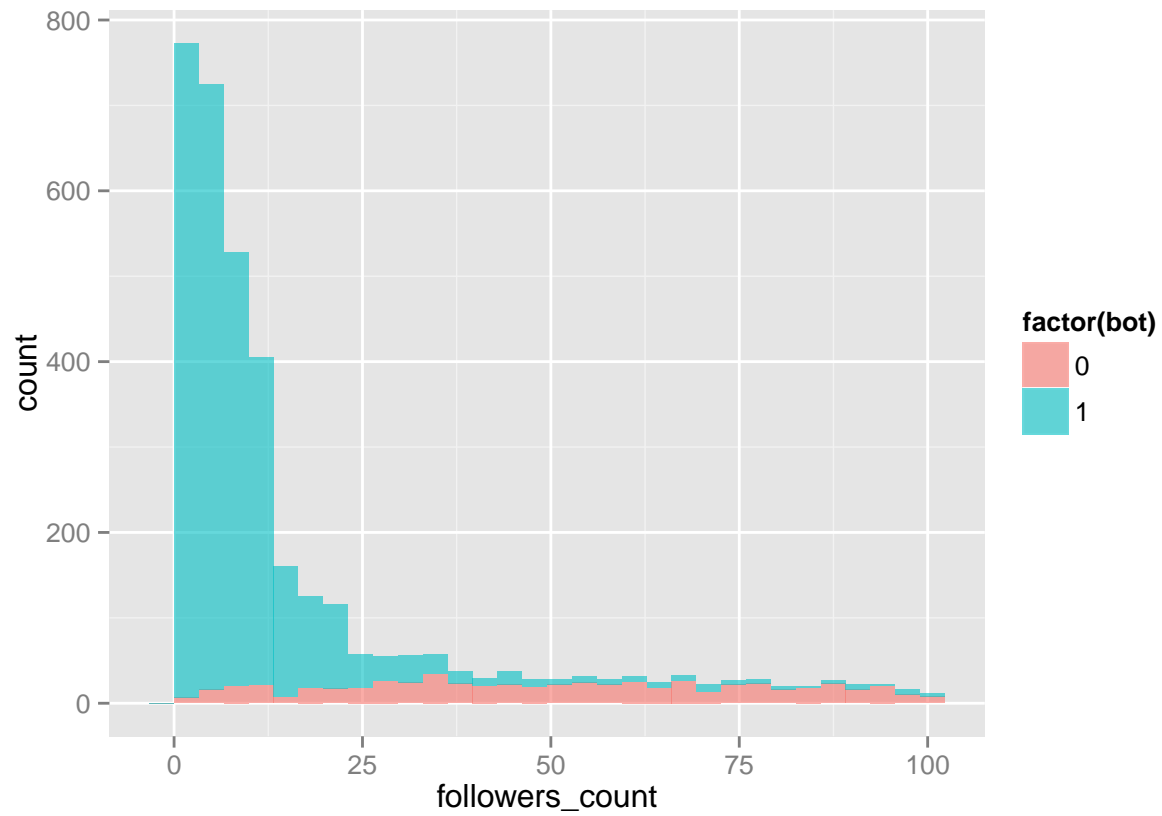
```
ggpairs(data[, c('followers_count', 'friends_count', 'account_age',
                 'days_since_last_tweet', 'listed_count',
                 'status_favorite_count', 'bot')],
        lower = list(continuous = 'points', params = list(alpha = 0.70)),
        diag = list(continuous = 'density', params = list(alpha = 0.70)),
        upper = list(continuous = 'cor'),
        axisLabels = 'show', color = 'bot')
```



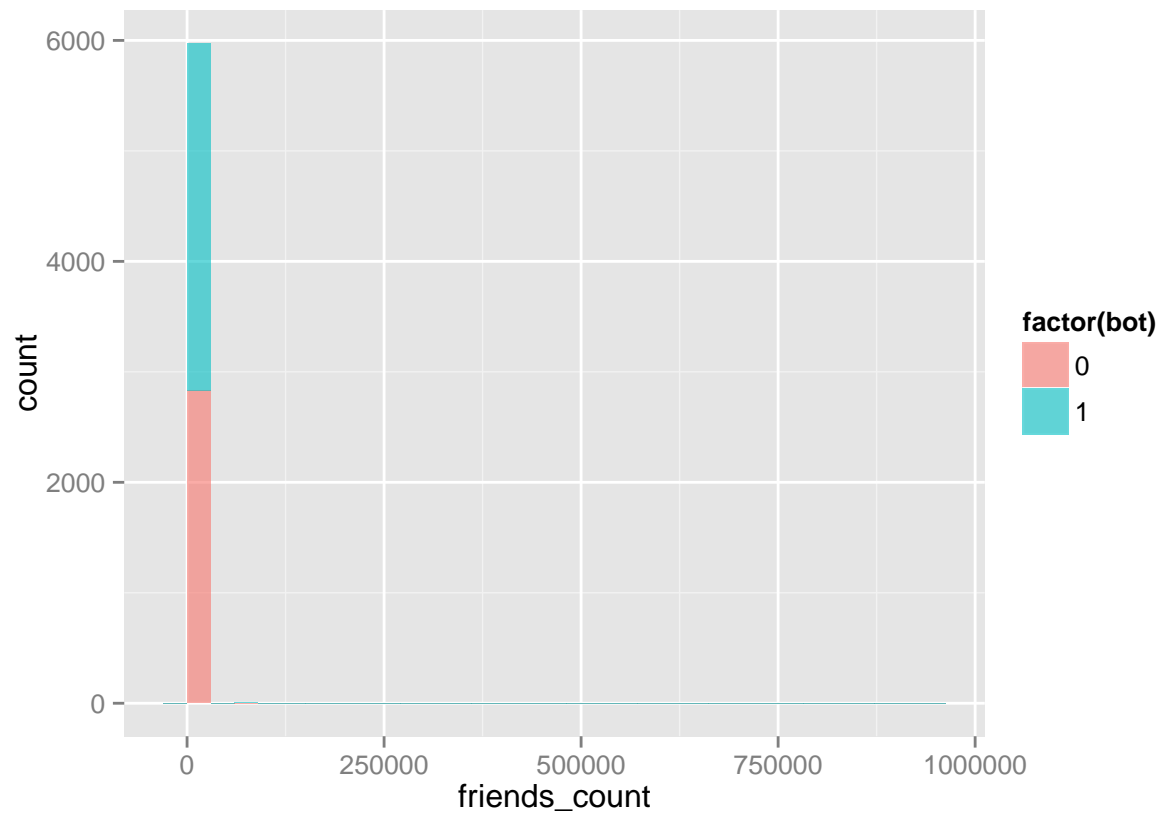
```
ggplot(data, aes(x = followers_count, fill = factor(bot))) +
  geom_histogram(alpha = 0.60)
```



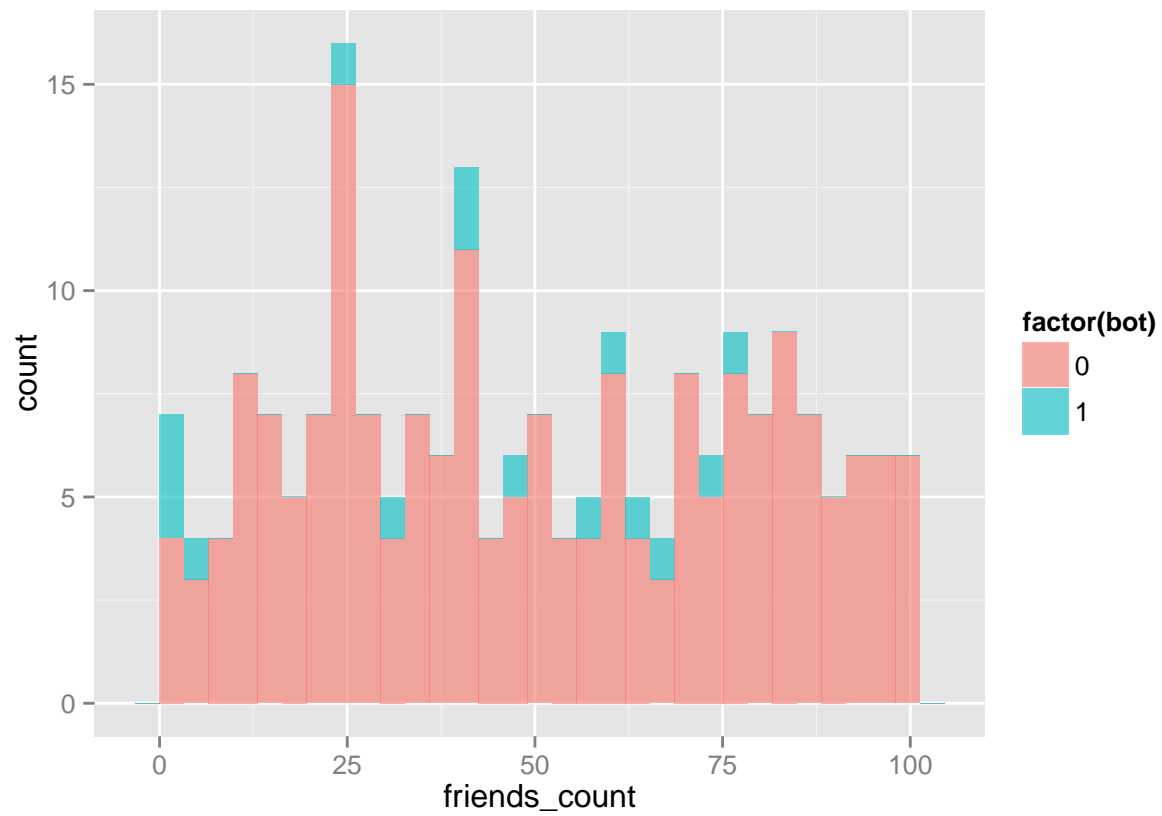
```
# whoa, some people have a lot of followers, but most don't. we need to lob off  
# the long tail so we can see the distribution better  
ggplot(filter(data, followers_count < 100),  
  aes(x = followers_count, fill = factor(bot))) +  
  geom_histogram(alpha = 0.60)
```



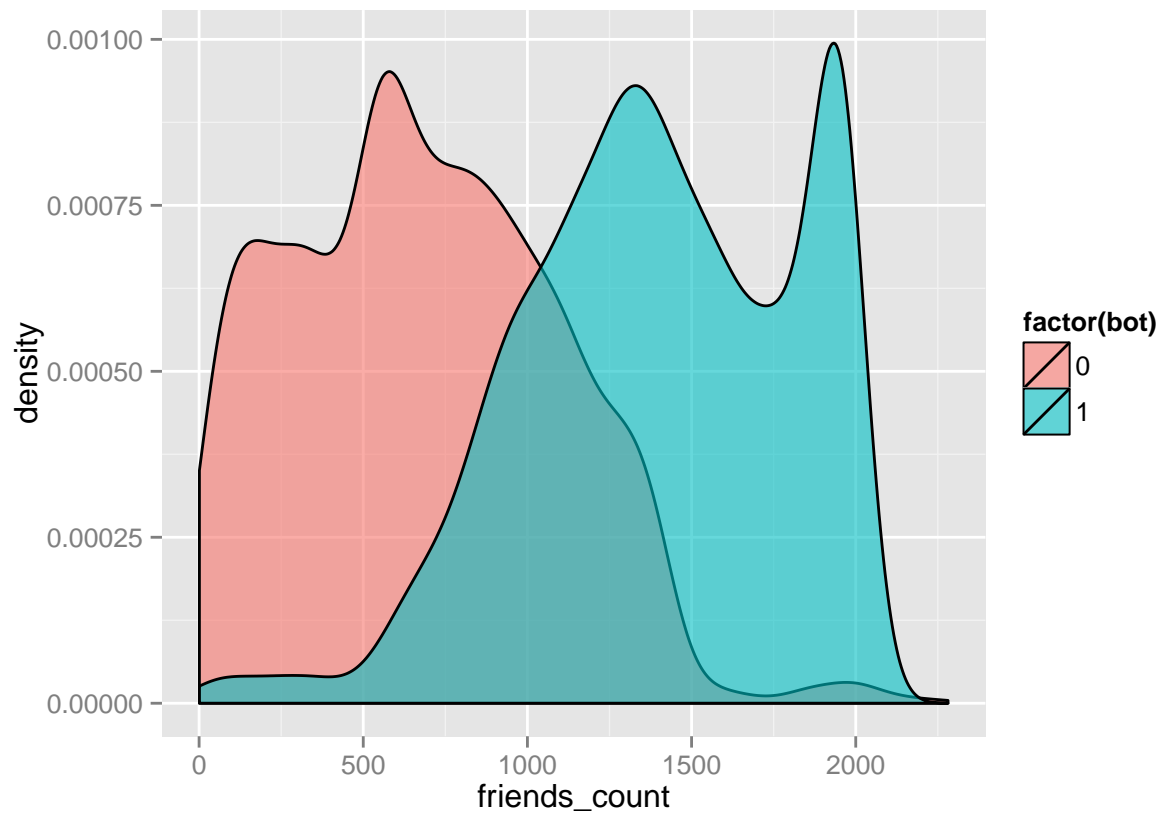
```
# how about the number of people they follow?  
ggplot(data, aes(x = friends_count, fill = factor(bot))) +  
  geom_histogram(alpha = 0.60)
```



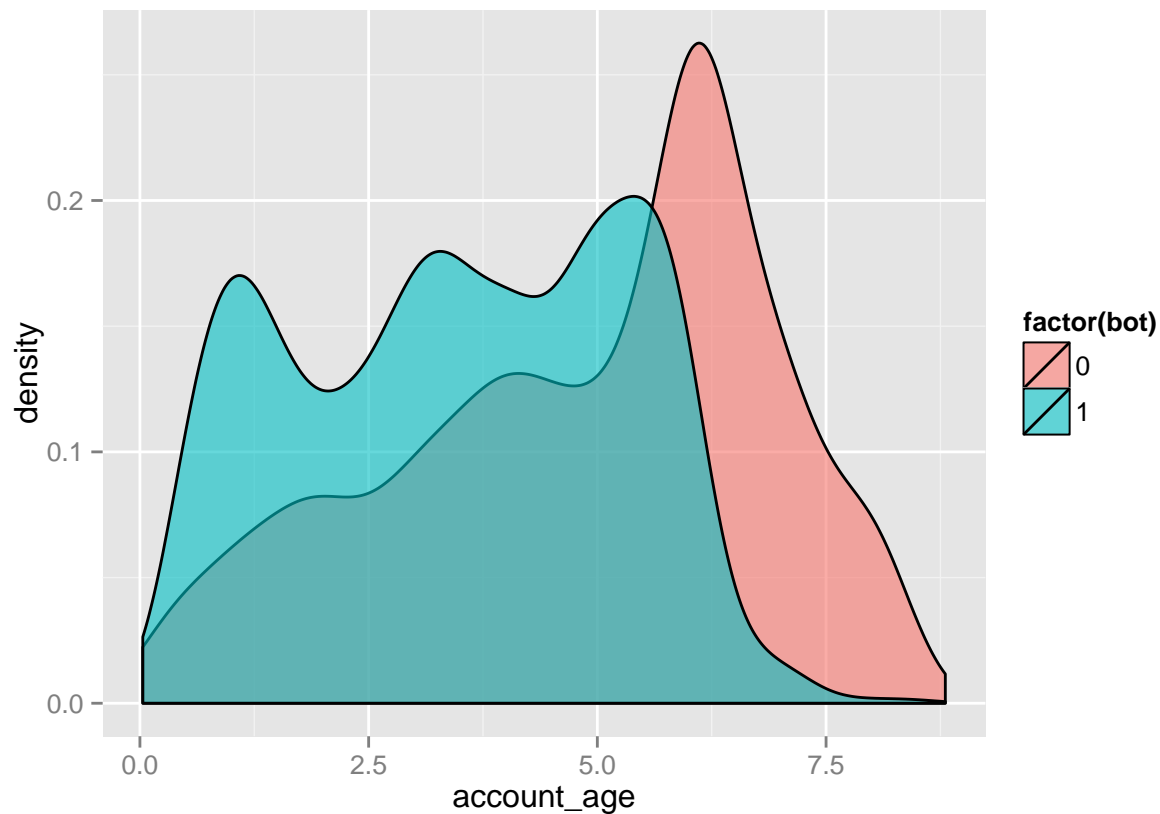
```
# it's a little hard to see  
ggplot(filter(data, friends_count < 100),  
  aes(x = friends_count, fill = factor(bot))) +  
  geom_histogram(alpha = 0.60)
```



```
# that's better  
ggplot(filter(data, friends_count < 2500),  
  aes(x = friends_count, fill = factor(bot))) +  
  geom_density(alpha = 0.60)
```



```
# what about account age?  
ggplot(data, aes(x = account_age, fill = factor(bot))) +  
  geom_density(alpha = 0.60)
```

```
# geo enabled?
xtabs(~bot + geo_enabled, data = data)
```

```
##      geo_enabled
## bot      0      1
##  0 1479 1358
##  1 2862  284
```

Logistic regression

Lucky for us, we can use the handy `train()` function for fitting logistic regressions.

```
set.seed(243)
data = na.omit(data)

# select the training observations
in_train = createDataPartition(y = data$bot,
                                p = 0.75, # 75% in train, 25% in test
                                list = FALSE)

train = data[in_train, ]
test = data[-in_train, ]

# drop the ids
train$id = NULL
test$id = NULL
```

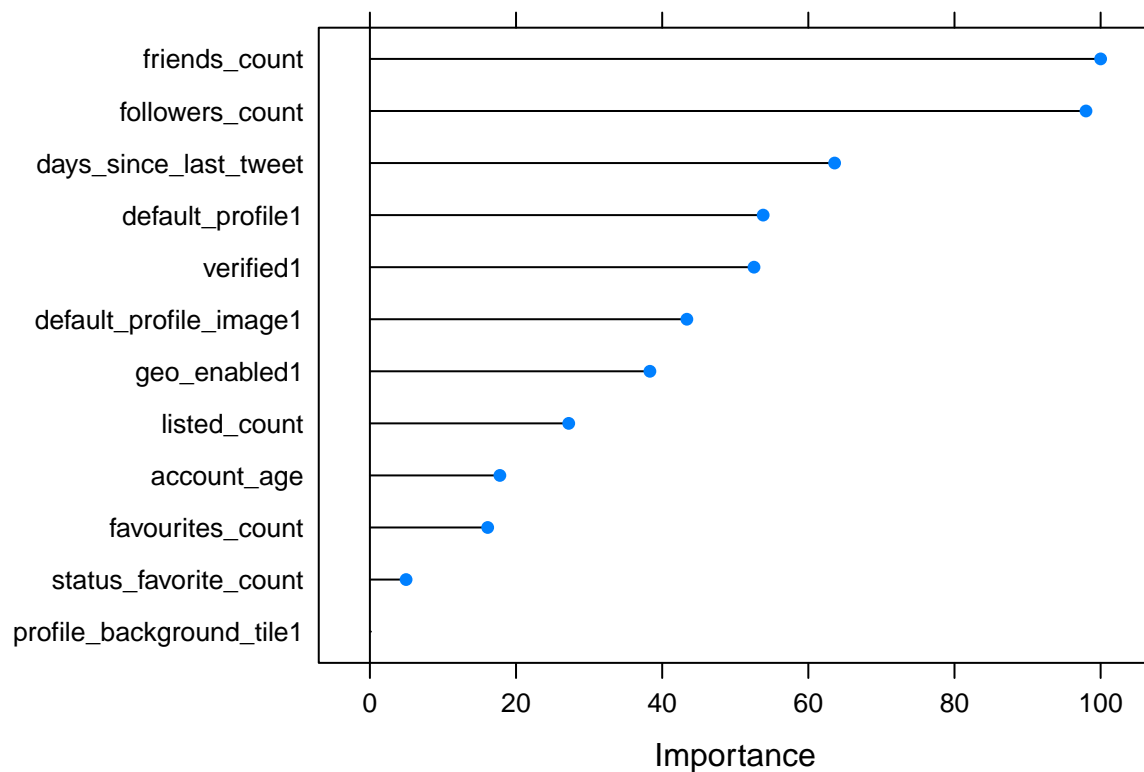
Check out [this page](#) for more types of logistic regression to try out.

```
logistic_model = train(bot ~ .,
                        data = na.omit(train),
                        method = 'glm',
                        family = binomial,
                        preProcess = c('center', 'scale'))

summary(logistic_model)
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5880  -0.0973   0.0001   0.0263   6.5947
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -6.53206     0.42481 -15.377 < 2e-16 ***
## followers_count -196.46069    10.35443 -18.974 < 2e-16 ***
## friends_count   68.81052     3.55725  19.344 < 2e-16 ***
## default_profile1 -0.99847     0.09230 -10.817 < 2e-16 ***
## default_profile_image1 -1.83530     0.20647  -8.889 < 2e-16 ***
## favourites_count -0.79899     0.20715  -3.857 0.000115 ***
## geo_enabled1    -0.73467     0.09234  -7.956 1.78e-15 ***
## account_age     -0.34950     0.08393  -4.164 3.12e-05 ***
## days_since_last_tweet  3.45322     0.27357  12.623 < 2e-16 ***
## listed_count     3.14718     0.53299   5.905 3.53e-09 ***
## profile_background_tile1 0.08489     0.09639   0.881 0.378506
## status_favorite_count  0.13845     0.07713   1.795 0.072668 .
## verified1        2.48665     0.23490  10.586 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 6209.7  on 4487  degrees of freedom
## Residual deviance: 1019.5  on 4475  degrees of freedom
## AIC: 1045.5
##
## Number of Fisher Scoring iterations: 14
```

```
plot(varImp(logistic_model))
```



```
# test predictions
```

```
logistic_predictions = predict(logistic_model, newdata = test)
confusionMatrix(logistic_predictions, test$bot)
```

```
## Confusion Matrix and Statistics
```

```
##
```

```
##           Reference
```

```
## Prediction  0    1
```

```
##           0 689  30
```

```
##           1  20 756
```

```
##
```

```
##           Accuracy : 0.9666
```

```
##           95% CI : (0.9561, 0.9751)
```

```
##           No Information Rate : 0.5258
```

```
##           P-Value [Acc > NIR] : <2e-16
```

```
##
```

```
##           Kappa : 0.933
```

```
##           McNemar's Test P-Value : 0.2031
```

```
##
```

```
##           Sensitivity : 0.9718
```

```
##           Specificity : 0.9618
```

```
##           Pos Pred Value : 0.9583
```

```
##           Neg Pred Value : 0.9742
```

```
##           Prevalence : 0.4742
```

```
##           Detection Rate : 0.4609
```

```
##           Detection Prevalence : 0.4809
```

```
##           Balanced Accuracy : 0.9668
```

```
##
```

```
##           'Positive' Class : 0
```

```
##
```

```
# stepwise logistic regression
```

```
step_model = train(bot ~ .,  
                   data = na.omit(train),  
                   method = 'glmStepAIC',  
                   family = binomial,  
                   preProcess = c('center', 'scale'))
```

```
## Loading required package: MASS
```

```
##
```

```
## Attaching package: 'MASS'
```

```
##
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      select
```

```
## Start:  AIC=16101.47
```

```
## .outcome ~ followers_count + friends_count + default_profile1 +  
##      default_profile_image1 + favourites_count + geo_enabled1 +  
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +  
##      status_favorite_count + verified1
```

```
##
```

	Df	Deviance	AIC
## - followers_count	1	1345	1369
## - friends_count	1	1782	1806
## - status_favorite_count	1	14562	14586
## <none>		16075	16101
## - default_profile_image1	1	16580	16604
## - listed_count	1	17806	17830
## - account_age	1	25303	25327
## - favourites_count	1	29844	29868
## - days_since_last_tweet	1	31574	31598
## - geo_enabled1	1	33953	33977
## - verified1	1	40225	40249
## - default_profile1	1	51110	51134
## - profile_background_tile1	1	51182	51206

```
##
```

```
## Step:  AIC=1369.21
```

```
## .outcome ~ friends_count + default_profile1 + default_profile_image1 +  
##      favourites_count + geo_enabled1 + account_age + days_since_last_tweet +  
##      listed_count + profile_background_tile1 + status_favorite_count +  
##      verified1
```

```
##
```

	Df	Deviance	AIC
## - status_favorite_count	1	1346.7	1368.7
## <none>		1345.2	1369.2
## - profile_background_tile1	1	1348.6	1370.6
## - account_age	1	1354.4	1376.4
## - favourites_count	1	1366.6	1388.6
## - geo_enabled1	1	1408.6	1430.6
## - default_profile_image1	1	1455.3	1477.3
## - default_profile1	1	1489.0	1511.0

```

## - listed_count          1    1630.0 1652.0
## - days_since_last_tweet 1    1663.2 1685.2
## - verified1             1    1711.9 1733.9
## - friends_count         1    1873.8 1895.8
##
## Step:  AIC=1368.66
## .outcome ~ friends_count + default_profile1 + default_profile_image1 +
##     favourites_count + geo_enabled1 + account_age + days_since_last_tweet +
##     listed_count + profile_background_tile1 + verified1
##
##              Df Deviance    AIC
## <none>                1346.7 1368.7
## - profile_background_tile1 1    1350.2 1370.2
## - account_age              1    1355.4 1375.4
## - favourites_count         1    1391.0 1411.0
## - geo_enabled1             1    1409.8 1429.8
## - default_profile_image1   1    1460.4 1480.4
## - default_profile1         1    1489.8 1509.8
## - listed_count             1    1642.7 1662.7
## - days_since_last_tweet    1    1678.1 1698.1
## - verified1                1    1712.0 1732.0
## - friends_count            1    1887.7 1907.7
## Start:  AIC=29365.53
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##              Df Deviance    AIC
## - account_age              1      912   936
## - favourites_count         1      946   970
## - listed_count             1      956   980
## - default_profile_image1   1     1035  1059
## - days_since_last_tweet    1     1161  1185
## - friends_count            1     1634  1658
## - followers_count          1     1775  1799
## <none>                29340 29366
## - profile_background_tile1 1    30132 30156
## - geo_enabled1             1    37413 37437
## - default_profile1         1    41811 41835
## - status_favorite_count    1    50101 50125
## - verified1                1    64014 64038
##
## Step:  AIC=936.27
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##              Df Deviance    AIC
## - profile_background_tile1 1      912   934
## <none>                912   936
## - favourites_count         1      971   993
## - listed_count             1      971   993

```

```

## - geo_enabled1          1      994  1016
## - default_profile1      1     1047  1069
## - default_profile_image1 1     1060  1082
## - days_since_last_tweet 1     1179  1201
## - friends_count         1     1657  1679
## - followers_count       1     1806  1828
## - status_favorite_count  1    44694 44716
## - verified1            1    45055 45077
##
## Step: AIC=934.27
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      days_since_last_tweet + listed_count + status_favorite_count +
##      verified1
##
##              Df Deviance   AIC
## <none>              912   934
## - favourites_count    1    971  991
## - listed_count        1    971  991
## - geo_enabled1        1    994 1014
## - default_profile_image1 1   1060 1080
## - default_profile1    1   1065 1085
## - days_since_last_tweet 1   1179 1199
## - friends_count       1   1657 1677
## - followers_count     1   1806 1826
## - verified1           1  40946 40966
## - status_favorite_count 1  45920 45940
## Start: AIC=33330.34
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      status_favorite_count + verified1
##
##              Df Deviance   AIC
## - followers_count     1   1467  1491
## - friends_count       1   1825  1849
## - listed_count        1  25375 25399
## - profile_background_tile1 1  27465 27489
## <none>              33304 33330
## - account_age         1  35827 35851
## - favourites_count    1  38278 38302
## - verified1           1  42604 42628
## - status_favorite_count 1  45487 45511
## - days_since_last_tweet 1  45848 45872
## - geo_enabled1        1  46785 46809
## - default_profile1    1  59977 60001
## - default_profile_image1 1  62067 62091
##
## Step: AIC=1490.79
## .outcome ~ friends_count + default_profile1 + default_profile_image1 +
##      favourites_count + geo_enabled1 + account_age + days_since_last_tweet +
##      listed_count + profile_background_tile1 + status_favorite_count +
##      verified1
##

```

```

##                                Df Deviance    AIC
## - profile_background_tile1  1   1466.9 1488.9
## <none>                        1466.8 1490.8
## - status_favorite_count      1   1469.2 1491.2
## - account_age                1   1489.5 1511.5
## - favourites_count           1   1516.3 1538.3
## - geo_enabled1              1   1530.5 1552.5
## - default_profile1           1   1558.0 1580.0
## - default_profile_image1     1   1567.2 1589.2
## - listed_count               1   1742.5 1764.5
## - days_since_last_tweet      1   1850.9 1872.9
## - verified1                  1   1851.7 1873.7
## - friends_count              1   1939.0 1961.0
##
## Step: AIC=1488.9
## .outcome ~ friends_count + default_profile1 + default_profile_image1 +
##           favourites_count + geo_enabled1 + account_age + days_since_last_tweet +
##           listed_count + status_favorite_count + verified1
##
##                                Df Deviance    AIC
## <none>                        1466.9 1488.9
## - status_favorite_count      1   1469.3 1489.3
## - account_age                1   1489.7 1509.7
## - favourites_count           1   1519.2 1539.2
## - geo_enabled1              1   1531.2 1551.2
## - default_profile_image1     1   1567.5 1587.5
## - default_profile1           1   1575.9 1595.9
## - listed_count               1   1742.8 1762.8
## - days_since_last_tweet      1   1852.0 1872.0
## - verified1                  1   1852.2 1872.2
## - friends_count              1   1939.5 1959.5
## Start: AIC=37799.75
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##           status_favorite_count + verified1
##
##                                Df Deviance    AIC
## - profile_background_tile1  1     922   946
## - account_age                1     931   955
## - geo_enabled1              1     988  1012
## - days_since_last_tweet      1    1273  1297
## - friends_count              1    1716  1740
## - followers_count            1    1849  1873
## - status_favorite_count      1   37197 37221
## <none>                        37774 37800
## - listed_count               1   44262 44286
## - default_profile1           1   44622 44646
## - favourites_count           1   46064 46088
## - verified1                  1   47938 47962
## - default_profile_image1     1   52696 52720
##
## Step: AIC=945.68
## .outcome ~ followers_count + friends_count + default_profile1 +

```

```

##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + status_favorite_count +
##      verified1
##
##              Df Deviance   AIC
## <none>              922   946
## - account_age      1     932   954
## - geo_enabled1     1     989  1011
## - default_profile1 1    1049  1071
## - days_since_last_tweet 1    1273  1295
## - friends_count    1    1716  1738
## - followers_count  1    1850  1872
## - status_favorite_count 1   35467 35489
## - favourites_count 1   39432 39454
## - listed_count     1   48226 48248
## - verified1        1   49812 49834
## - default_profile_image1 1   52047 52069
## Start:  AIC=1044.32
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      status_favorite_count + verified1
##
##              Df Deviance   AIC
## - status_favorite_count 1    1018  1042
## - profile_background_tile1 1    1018  1042
## <none>              1018  1044
## - favourites_count      1    1028  1052
## - account_age           1    1042  1066
## - geo_enabled1          1    1084  1108
## - default_profile1      1    1149  1173
## - days_since_last_tweet 1    1272  1296
## - followers_count       1    1457  1481
## - verified1             1    1457  1481
## - friends_count         1    1793  1817
## - default_profile_image1 1   36909 36933
## - listed_count          1   40297 40321
##
## Step:  AIC=1042.32
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      verified1
##
##              Df Deviance   AIC
## - profile_background_tile1 1    1018  1040
## <none>              1018  1042
## - account_age           1    1042  1064
## - geo_enabled1          1    1084  1106
## - default_profile1      1    1149  1171
## - default_profile_image1 1    1157  1179
## - followers_count       1    1456  1478
## - verified1             1    1458  1480
## - friends_count         1    1797  1819

```



```

## - favourites_count          1    23284 23306
## - listed_count              1    27537 27559
## - days_since_last_tweet    1    63797 63819
##
## Step: AIC=1040.45
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + listed_count + verified1
##
##               Df Deviance    AIC
## <none>                1018 1040
## - account_age          1    1042 1062
## - geo_enabled1         1    1084 1104
## - default_profile_image1 1    1157 1177
## - default_profile1     1    1171 1191
## - followers_count      1    1456 1476
## - verified1            1    1458 1478
## - friends_count        1    1797 1817
## - favourites_count     1   41090 41110
## - listed_count         1   47217 47237
## - days_since_last_tweet 1   57309 57329
## Start: AIC=965.05
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##           status_favorite_count + verified1
##
##               Df Deviance    AIC
## - profile_background_tile1 1   939.53 963.53
## <none>                939.05 965.05
## - favourites_count        1   949.57 973.57
## - account_age             1   956.67 980.67
## - listed_count            1   957.99 981.99
## - status_favorite_count    1   960.02 984.02
## - geo_enabled1            1  1032.41 1056.41
## - default_profile1        1  1042.67 1066.67
## - default_profile_image1   1  1093.86 1117.86
## - days_since_last_tweet    1  1254.62 1278.62
## - verified1               1  1317.71 1341.71
## - friends_count           1  1764.91 1788.91
## - followers_count         1  1823.21 1847.21
##
## Step: AIC=963.53
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + listed_count + status_favorite_count +
##           verified1
##
##               Df Deviance    AIC
## <none>                939.53 963.53
## - favourites_count        1   950.12 972.12
## - account_age             1   956.96 978.96
## - listed_count            1   958.98 980.98
## - status_favorite_count    1   966.04 988.04

```

```

## - geo_enabled1          1  1032.38 1054.38
## - default_profile1      1  1051.76 1073.76
## - default_profile_image1 1  1093.88 1115.88
## - days_since_last_tweet 1  1258.56 1280.56
## - verified1            1  1318.29 1340.29
## - friends_count         1  1765.05 1787.05
## - followers_count       1  1823.69 1845.69
## Start:  AIC=1156.93
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##              Df Deviance   AIC
## <none>                1130.9 1156.9
## - profile_background_tile1 1  1136.4 1160.4
## - status_favorite_count    1  1140.8 1164.8
## - account_age              1  1145.9 1169.9
## - listed_count             1  1147.1 1171.1
## - favourites_count         1  1152.1 1176.1
## - geo_enabled1            1  1207.5 1231.5
## - default_profile_image1   1  1234.4 1258.4
## - default_profile1        1  1276.6 1300.6
## - days_since_last_tweet    1  1350.4 1374.4
## - verified1               1  1415.9 1439.9
## - friends_count           1  1911.8 1935.8
## - followers_count         1  1918.2 1942.2
## Start:  AIC=43783
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##              Df Deviance   AIC
## - geo_enabled1          1    1062  1086
## - default_profile1      1    1114  1138
## - default_profile_image1 1    1135  1159
## - days_since_last_tweet 1    1230  1254
## - friends_count         1    1815  1839
## - followers_count       1    1819  1843
## - listed_count          1   40513 40537
## - profile_background_tile1 1  41090 41114
## - favourites_count      1   42532 42556
## - verified1            1   43757 43781
## <none>                  43757 43783
## - account_age          1   44550 44574
## - status_favorite_count 1   47722 47746
##
## Step:  AIC=1085.63
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + account_age +
##     days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##

```

```

##                                Df Deviance   AIC
## - profile_background_tile1  1      1062  1084
## <none>                        1062  1086
## - status_favorite_count      1      1075  1097
## - days_since_last_tweet      1      1330  1352
## - friends_count              1      1879  1901
## - followers_count            1      1896  1918
## - default_profile_image1     1     40441 40463
## - account_age                1     41090 41112
## - listed_count               1     44694 44716
## - verified1                  1     46857 46879
## - default_profile1           1     47506 47528
## - favourites_count           1     52768 52790
##
## Step:  AIC=1083.64
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + account_age +
##           days_since_last_tweet + listed_count + status_favorite_count +
##           verified1
##
##                                Df Deviance   AIC
## <none>                        1062  1084
## - status_favorite_count      1      1075  1095
## - default_profile1           1      1176  1196
## - friends_count              1      1879  1899
## - followers_count            1      1896  1916
## - default_profile_image1     1     36332 36352
## - days_since_last_tweet      1     37557 37577
## - account_age                1     37630 37650
## - listed_count               1     42387 42407
## - verified1                  1     42676 42696
## - favourites_count           1     50677 50697
## Start:  AIC=966.84
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##           status_favorite_count + verified1
##
##                                Df Deviance   AIC
## - status_favorite_count      1        941   965
## <none>                        941   967
## - account_age                1        966   990
## - geo_enabled1               1       1022  1046
## - default_profile1           1       1090  1114
## - default_profile_image1     1       1093  1117
## - friends_count              1       1766  1790
## - followers_count            1       1877  1901
## - listed_count               1      38206 38230
## - verified1                  1     47433 47457
## - favourites_count           1     53777 53801
## - profile_background_tile1  1     54138 54162
## - days_since_last_tweet      1     55868 55892
##
## Step:  AIC=965.16

```

```

## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     verified1
##
##           Df Deviance   AIC
## - profile_background_tile1  1      941   963
## <none>                      941   965
## - default_profile1         1     1090  1112
## - default_profile_image1   1     1093  1115
## - followers_count          1     1881  1903
## - verified1                1    33160 33182
## - geo_enabled1             1    36044 36066
## - account_age              1    36260 36282
## - favourites_count         1    42748 42770
## - listed_count              1    48659 48681
## - days_since_last_tweet    1    53993 54015
## - friends_count            1    54065 54087
##
## Step:  AIC=963.17
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + verified1
##
##           Df Deviance   AIC
## <none>                      941   963
## - default_profile_image1   1     1094  1114
## - default_profile1         1     1105  1125
## - followers_count          1     1881  1901
## - account_age              1    35395 35415
## - geo_enabled1             1    38495 38515
## - favourites_count         1    42459 42479
## - friends_count            1    42892 42912
## - verified1                1    45199 45219
## - listed_count              1    46713 46733
## - days_since_last_tweet    1    56012 56032
## Start:  AIC=33041.99
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##           Df Deviance   AIC
## - account_age              1      940   964
## - favourites_count         1      958   982
## - default_profile_image1   1     1053  1077
## - friends_count            1     1651  1675
## - followers_count          1     1737  1761
## <none>                      33016 33042
## - status_favorite_count    1    38999 39023
## - default_profile1         1    40657 40681
## - days_since_last_tweet    1    46568 46592
## - verified1                1    48371 48395
## - geo_enabled1             1    49019 49043

```

```

## - profile_background_tile1 1 49019 49043
## - listed_count 1 68483 68507
##
## Step: AIC=963.57
## .outcome ~ followers_count + friends_count + default_profile1 +
## default_profile_image1 + favourites_count + geo_enabled1 +
## days_since_last_tweet + listed_count + profile_background_tile1 +
## status_favorite_count + verified1
##
## Df Deviance AIC
## - profile_background_tile1 1 940 962
## <none> 940 964
## - favourites_count 1 988 1010
## - status_favorite_count 1 993 1015
## - default_profile_image1 1 1063 1085
## - verified1 1 1230 1252
## - friends_count 1 1657 1679
## - followers_count 1 1747 1769
## - listed_count 1 37557 37579
## - default_profile1 1 43324 43346
## - geo_enabled1 1 43469 43491
## - days_since_last_tweet 1 49091 49113
##
## Step: AIC=961.58
## .outcome ~ followers_count + friends_count + default_profile1 +
## default_profile_image1 + favourites_count + geo_enabled1 +
## days_since_last_tweet + listed_count + status_favorite_count +
## verified1
##
## Df Deviance AIC
## - status_favorite_count 1 940 960
## <none> 940 962
## - favourites_count 1 975 995
## - default_profile_image1 1 1064 1084
## - verified1 1 1231 1251
## - friends_count 1 1657 1677
## - followers_count 1 1747 1767
## - geo_enabled1 1 31935 31955
## - listed_count 1 38783 38803
## - default_profile1 1 44766 44786
## - days_since_last_tweet 1 50173 50193
##
## Step: AIC=959.77
## .outcome ~ followers_count + friends_count + default_profile1 +
## default_profile_image1 + favourites_count + geo_enabled1 +
## days_since_last_tweet + listed_count + verified1
##
## Df Deviance AIC
## <none> 940 960
## - favourites_count 1 975 993
## - geo_enabled1 1 1057 1075
## - default_profile_image1 1 1064 1082
## - verified1 1 1232 1250
## - friends_count 1 1660 1678

```

```

## - followers_count      1      1747  1765
## - listed_count         1      43685 43703
## - default_profile1     1      44982 45000
## - days_since_last_tweet 1      51903 51921
## Start:  AIC=37583.49
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      status_favorite_count + verified1
##
##              Df Deviance   AIC
## - favourites_count      1      924   948
## - listed_count          1      928   952
## - default_profile_image1 1     1006  1030
## - default_profile1      1     1037  1061
## - days_since_last_tweet 1     1237  1261
## - verified1             1     1418  1442
## - friends_count         1     1698  1722
## - followers_count        1     1850  1874
## - account_age           1    35395 35419
## - status_favorite_count  1    36260 36284
## - geo_enabled1          1    37341 37365
## <none>                  37557 37583
## - profile_background_tile1 1    37846 37870
##
## Step:  AIC=948.15
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + geo_enabled1 + account_age + days_since_last_tweet +
##      listed_count + profile_background_tile1 + status_favorite_count +
##      verified1
##
##              Df Deviance   AIC
## - profile_background_tile1 1      924   946
## <none>                  924   948
## - account_age           1      938   960
## - status_favorite_count  1      946   968
## - geo_enabled1          1     1012  1034
## - days_since_last_tweet 1     1345  1367
## - verified1             1     1451  1473
## - friends_count         1     1793  1815
## - followers_count        1     1958  1980
## - default_profile_image1 1    42315 42337
## - default_profile1      1    50317 50339
## - listed_count          1    52191 52213
##
## Step:  AIC=946.24
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + geo_enabled1 + account_age + days_since_last_tweet +
##      listed_count + status_favorite_count + verified1
##
##              Df Deviance   AIC
## <none>                  924   946
## - account_age           1      938   958
## - status_favorite_count  1      946   966

```

```

## - geo_enabled1          1      1013  1033
## - days_since_last_tweet 1      1346  1366
## - verified1             1      1451  1471
## - friends_count         1      1795  1815
## - followers_count       1      1960  1980
## - default_profile_image1 1     41666 41686
## - listed_count          1     46496 46516
## - default_profile1      1     53345 53365
## Start:  AIC=976.83
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      status_favorite_count + verified1
##
##              Df Deviance    AIC
## - profile_background_tile1 1   951.46  975.46
## <none>                      950.83  976.83
## - status_favorite_count    1   961.12  985.12
## - favourites_count         1   977.20 1001.20
## - account_age              1   980.93 1004.93
## - listed_count             1   984.23 1008.23
## - geo_enabled1             1  1036.58 1060.58
## - default_profile1         1  1046.92 1070.92
## - default_profile_image1   1  1124.06 1148.06
## - days_since_last_tweet    1  1178.25 1202.25
## - verified1                1  1379.66 1403.66
## - friends_count            1  1755.30 1779.30
## - followers_count          1  1821.29 1845.29
##
## Step:  AIC=975.46
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + status_favorite_count +
##      verified1
##
##              Df Deviance    AIC
## <none>                      951.46  975.46
## - status_favorite_count    1   961.68  983.68
## - favourites_count         1   978.46 1000.46
## - account_age              1   981.01 1003.01
## - listed_count             1   985.91 1007.91
## - geo_enabled1             1  1037.41 1059.41
## - default_profile1         1  1052.59 1074.59
## - default_profile_image1   1  1124.23 1146.23
## - days_since_last_tweet    1  1181.08 1203.08
## - verified1                1  1384.06 1406.06
## - friends_count            1  1758.66 1780.66
## - followers_count          1  1824.49 1846.49
## Start:  AIC=49477.89
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      status_favorite_count + verified1
##

```

```

##                                Df Deviance   AIC
## - listed_count                1      1021  1045
## - friends_count               1      1796  1820
## - verified1                   1     43685 43709
## - geo_enabled1                1     44117 44141
## - account_age                 1     47866 47890
## - favourites_count            1     48443 48467
## - followers_count             1     49164 49188
## - profile_background_tile1    1     49380 49404
## - status_favorite_count       1     49380 49404
## <none>                        1     49452 49478
## - default_profile1            1     51110 51134
## - default_profile_image1      1     51182 51206
## - days_since_last_tweet       1     62211 62235
##
## Step:  AIC=1045.08
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + profile_background_tile1 +
##           status_favorite_count + verified1
##
##                                Df Deviance   AIC
## <none>                        1      1021  1045
## - favourites_count            1      1031  1053
## - verified1                   1      1500  1522
## - friends_count               1      1806  1828
## - followers_count             1      2006  2028
## - account_age                 1     38999 39021
## - status_favorite_count       1     41955 41977
## - profile_background_tile1    1     42964 42986
## - default_profile_image1      1     44478 44500
## - geo_enabled1                1     46280 46302
## - default_profile1            1     47506 47528
## - days_since_last_tweet       1     65383 65405
## Start:  AIC=40539.07
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##           status_favorite_count + verified1
##
##                                Df Deviance   AIC
## - status_favorite_count       1        902   926
## - account_age                 1        919   943
## - geo_enabled1                1        986  1010
## - default_profile_image1      1       1002  1026
## - default_profile1            1       1036  1060
## - days_since_last_tweet       1       1091  1115
## - verified1                   1       1344  1368
## - friends_count               1       1725  1749
## - followers_count             1       1838  1862
## <none>                        1     40513 40539
## - favourites_count            1     43068 43092
## - profile_background_tile1    1     47073 47097
## - listed_count                1     55507 55531

```



```

##
## Step: AIC=925.84
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     verified1
##
##           Df Deviance    AIC
## <none>           902    926
## - geo_enabled1      1     991  1013
## - default_profile_image1  1    1019  1041
## - default_profile1    1    1049  1071
## - days_since_last_tweet  1    1108  1130
## - verified1          1    1352  1374
## - friends_count      1    1738  1760
## - followers_count    1    1844  1866
## - listed_count       1   44478 44500
## - account_age        1   47145 47167
## - profile_background_tile1  1  48587 48609
## - favourites_count    1   52047 52069
## Start: AIC=1053.54
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##           Df Deviance    AIC
## - profile_background_tile1  1  1027.9 1051.9
## - status_favorite_count     1  1029.2 1053.2
## <none>           1027.5 1053.5
## - listed_count             1  1037.2 1061.2
## - favourites_count         1  1050.5 1074.5
## - account_age              1  1057.8 1081.8
## - geo_enabled1             1  1076.0 1100.0
## - default_profile1         1  1122.7 1146.7
## - default_profile_image1    1  1176.4 1200.4
## - days_since_last_tweet     1  1344.3 1368.3
## - verified1                1  1437.6 1461.6
## - followers_count          1  1716.7 1740.7
## - friends_count            1  1761.6 1785.6
##
## Step: AIC=1051.93
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + status_favorite_count +
##     verified1
##
##           Df Deviance    AIC
## - status_favorite_count     1  1029.7 1051.7
## <none>           1027.9 1051.9
## - listed_count             1  1037.9 1059.9
## - favourites_count         1  1051.2 1073.2
## - account_age              1  1057.8 1079.8
## - geo_enabled1             1  1076.0 1098.0

```

```

## - default_profile1      1    1130.9 1152.9
## - default_profile_image1 1    1176.4 1198.4
## - days_since_last_tweet 1    1345.6 1367.6
## - verified1             1    1438.1 1460.1
## - followers_count       1    1716.7 1738.7
## - friends_count         1    1761.8 1783.8
##
## Step: AIC=1051.7
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + listed_count + verified1
##
##               Df Deviance   AIC
## <none>                1029.7 1051.7
## - listed_count        1    1039.5 1059.5
## - favourites_count    1    1052.6 1072.6
## - account_age         1    1059.6 1079.6
## - geo_enabled1        1    1076.8 1096.8
## - default_profile1    1    1134.3 1154.3
## - default_profile_image1 1    1179.6 1199.6
## - days_since_last_tweet 1    1353.2 1373.2
## - verified1           1    1439.7 1459.7
## - followers_count     1    1730.7 1750.7
## - friends_count       1    1772.7 1792.7
## Start: AIC=1086.7
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##           status_favorite_count + verified1
##
##               Df Deviance   AIC
## - profile_background_tile1 1    1061 1085
## <none>                1061 1087
## - status_favorite_count    1    1070 1094
## - account_age              1    1073 1097
## - favourites_count         1    1084 1108
## - listed_count             1    1093 1117
## - geo_enabled1             1    1118 1142
## - default_profile_image1   1    1209 1233
## - default_profile1         1    1223 1247
## - days_since_last_tweet    1    1326 1350
## - friends_count            1    1884 1908
## - followers_count          1    1975 1999
## - verified1                1   46568 46592
##
## Step: AIC=1084.97
## .outcome ~ followers_count + friends_count + default_profile1 +
##           default_profile_image1 + favourites_count + geo_enabled1 +
##           account_age + days_since_last_tweet + listed_count + status_favorite_count +
##           verified1
##
##               Df Deviance   AIC
## <none>                1061 1085
## - status_favorite_count 1    1070 1092

```

```

## - account_age          1      1073  1095
## - favourites_count     1      1084  1106
## - listed_count         1      1094  1116
## - geo_enabled1        1      1118  1140
## - default_profile_image1 1      1209  1231
## - default_profile1     1      1238  1260
## - days_since_last_tweet 1      1330  1352
## - friends_count        1      1888  1910
## - followers_count      1      1980  2002
## - verified1           1     44982 45004
## Start:  AIC=41764.55
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      status_favorite_count + verified1
##
##              Df Deviance    AIC
## - friends_count      1      1853  1877
## - followers_count    1      1973  1997
## <none>                41739 41765
## - geo_enabled1       1     43108 43132
## - profile_background_tile1 1     43324 43348
## - default_profile_image1 1     43685 43709
## - favourites_count   1     44838 44862
## - status_favorite_count 1     45992 46016
## - days_since_last_tweet 1     46785 46809
## - verified1          1     46785 46809
## - default_profile1   1     47289 47313
## - account_age        1     47866 47890
## - listed_count       1     49668 49692
##
## Step:  AIC=1876.65
## .outcome ~ followers_count + default_profile1 + default_profile_image1 +
##      favourites_count + geo_enabled1 + account_age + days_since_last_tweet +
##      listed_count + profile_background_tile1 + status_favorite_count +
##      verified1
##
##              Df Deviance    AIC
## - profile_background_tile1 1     1852.8 1874.8
## <none>                1852.7 1876.7
## - listed_count          1     1861.4 1883.4
## - account_age           1     1868.1 1890.1
## - status_favorite_count  1     1871.7 1893.7
## - geo_enabled1          1     1903.2 1925.2
## - default_profile1      1     1956.3 1978.3
## - favourites_count       1     1967.2 1989.2
## - default_profile_image1 1     1993.1 2015.1
## - followers_count        1     1997.2 2019.2
## - verified1             1     2250.6 2272.6
## - days_since_last_tweet  1     2342.6 2364.6
##
## Step:  AIC=1874.85
## .outcome ~ followers_count + default_profile1 + default_profile_image1 +
##      favourites_count + geo_enabled1 + account_age + days_since_last_tweet +

```

```

##      listed_count + status_favorite_count + verified1
##
##              Df Deviance    AIC
## <none>              1852.8 1874.8
## - listed_count      1   1861.5 1881.5
## - account_age       1   1868.6 1888.6
## - status_favorite_count 1   1871.9 1891.9
## - geo_enabled1      1   1903.8 1923.8
## - favourites_count  1   1967.2 1987.2
## - default_profile1  1   1973.5 1993.5
## - default_profile_image1 1  1993.7 2013.7
## - followers_count   1   1997.4 2017.4
## - verified1         1   2250.6 2270.6
## - days_since_last_tweet 1  2343.8 2363.8
## Start:  AIC=44720.13
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      status_favorite_count + verified1
##
##              Df Deviance    AIC
## - listed_count      1      774   798
## - favourites_count  1      786   810
## - geo_enabled1      1      864   888
## - default_profile_image1 1     897   921
## - default_profile1  1      944   968
## - days_since_last_tweet 1     997  1021
## - verified1         1     1286  1310
## - friends_count     1     1692  1716
## - followers_count   1     1859  1883
## - status_favorite_count 1    39648 39672
## - account_age       1    41522 41546
## <none>              44694 44720
## - profile_background_tile1 1  47506 47530
##
## Step:  AIC=797.91
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + profile_background_tile1 +
##      status_favorite_count + verified1
##
##              Df Deviance    AIC
## <none>              773.91  797.91
## - profile_background_tile1 1  777.07  799.07
## - account_age       1  781.03  803.03
## - favourites_count  1  784.50  806.50
## - status_favorite_count 1  791.64  813.64
## - geo_enabled1      1  862.82  884.82
## - default_profile_image1 1  900.47  922.47
## - default_profile1  1  946.40  968.40
## - days_since_last_tweet 1  997.48 1019.48
## - verified1         1 1413.28 1435.28
## - friends_count     1 1698.35 1720.35
## - followers_count   1 2034.22 2056.22

```

```

## Start:  AIC=38520.62
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##           Df Deviance   AIC
## - favourites_count      1    1156  1180
## - friends_count         1    1777  1801
## - listed_count           1   36692 36716
## - profile_background_tile1 1   36765 36789
## - default_profile_image1  1   37485 37509
## <none>                   38495 38521
## - geo_enabled1          1   40081 40105
## - status_favorite_count  1   40369 40393
## - default_profile1       1   41306 41330
## - followers_count        1   41811 41835
## - days_since_last_tweet  1   44838 44862
## - account_age            1   47506 47530
## - verified1              1   52984 53008
##
## Step:  AIC=1179.9
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + geo_enabled1 + account_age + days_since_last_tweet +
##     listed_count + profile_background_tile1 + status_favorite_count +
##     verified1
##
##           Df Deviance   AIC
## <none>                   1156  1180
## - profile_background_tile1 1   1161  1183
## - status_favorite_count    1   1164  1186
## - listed_count             1   1188  1210
## - default_profile_image1   1   1388  1410
## - verified1                1   1564  1586
## - days_since_last_tweet    1   1571  1593
## - friends_count            1   1884  1906
## - followers_count          1   1942  1964
## - account_age              1   49380 49402
## - geo_enabled1             1   52047 52069
## - default_profile1         1   70718 70740
## Start:  AIC=1069.77
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##           Df Deviance   AIC
## - profile_background_tile1 1   1044.2 1068.2
## <none>                   1043.8 1069.8
## - status_favorite_count    1   1051.5 1075.5
## - account_age              1   1055.6 1079.6
## - favourites_count         1   1060.0 1084.0
## - listed_count             1   1067.7 1091.7
## - geo_enabled1             1   1135.8 1159.8

```

```

## - default_profile_image1      1   1160.0 1184.0
## - followers_count             1   1172.5 1196.5
## - default_profile1           1   1196.0 1220.0
## - verified1                  1   1306.5 1330.5
## - days_since_last_tweet      1   1390.5 1414.5
## - friends_count              1   1790.5 1814.5
##
## Step: AIC=1068.24
## .outcome ~ followers_count + friends_count + default_profile1 +
##   default_profile_image1 + favourites_count + geo_enabled1 +
##   account_age + days_since_last_tweet + listed_count + status_favorite_count +
##   verified1
##
##               Df Deviance    AIC
## <none>                1044.2 1068.2
## - status_favorite_count  1   1051.8 1073.8
## - account_age           1   1055.9 1077.9
## - favourites_count      1   1060.6 1082.6
## - listed_count          1   1068.5 1090.5
## - geo_enabled1          1   1135.8 1157.8
## - default_profile_image1 1   1160.0 1182.0
## - followers_count       1   1173.5 1195.5
## - default_profile1      1   1210.5 1232.5
## - verified1             1   1307.2 1329.2
## - days_since_last_tweet 1   1393.4 1415.4
## - friends_count         1   1791.5 1813.5
## Start: AIC=932.31
## .outcome ~ followers_count + friends_count + default_profile1 +
##   default_profile_image1 + favourites_count + geo_enabled1 +
##   account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##   status_favorite_count + verified1
##
##               Df Deviance    AIC
## <none>                906.31  932.31
## - listed_count        1   911.64  935.64
## - profile_background_tile1 1   911.85  935.85
## - favourites_count    1   919.33  943.33
## - status_favorite_count 1   919.49  943.49
## - account_age         1   920.32  944.32
## - geo_enabled1        1   992.51 1016.51
## - default_profile1    1  1032.82 1056.82
## - default_profile_image1 1  1035.72 1059.72
## - days_since_last_tweet 1  1137.28 1161.28
## - verified1           1  1538.73 1562.73
## - friends_count       1  1731.78 1755.78
## - followers_count     1  1841.68 1865.68
## Start: AIC=1045.36
## .outcome ~ followers_count + friends_count + default_profile1 +
##   default_profile_image1 + favourites_count + geo_enabled1 +
##   account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##   status_favorite_count + verified1
##
##               Df Deviance    AIC
## - status_favorite_count  1   1019.8 1043.8

```

```

## - profile_background_tile1 1 1020.4 1044.4
## <none> 1019.4 1045.4
## - favourites_count 1 1041.0 1065.0
## - account_age 1 1057.3 1081.3
## - geo_enabled1 1 1099.3 1123.3
## - listed_count 1 1111.9 1135.9
## - default_profile_image1 1 1133.2 1157.2
## - default_profile1 1 1158.2 1182.2
## - days_since_last_tweet 1 1328.4 1352.4
## - verified1 1 1490.6 1514.6
## - friends_count 1 1792.4 1816.4
## - followers_count 1 1867.0 1891.0
##
## Step: AIC=1043.77
## .outcome ~ followers_count + friends_count + default_profile1 +
## default_profile_image1 + favourites_count + geo_enabled1 +
## account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
## verified1
##
## Df Deviance AIC
## - profile_background_tile1 1 1020.9 1042.9
## <none> 1019.8 1043.8
## - favourites_count 1 1041.6 1063.6
## - account_age 1 1057.6 1079.6
## - geo_enabled1 1 1099.5 1121.5
## - listed_count 1 1112.5 1134.5
## - default_profile_image1 1 1134.7 1156.7
## - default_profile1 1 1159.7 1181.7
## - days_since_last_tweet 1 1329.7 1351.7
## - verified1 1 1490.6 1512.6
## - friends_count 1 1802.0 1824.0
## - followers_count 1 1886.0 1908.0
##
## Step: AIC=1042.89
## .outcome ~ followers_count + friends_count + default_profile1 +
## default_profile_image1 + favourites_count + geo_enabled1 +
## account_age + days_since_last_tweet + listed_count + verified1
##
## Df Deviance AIC
## <none> 1020.9 1042.9
## - favourites_count 1 1044.0 1064.0
## - account_age 1 1057.9 1077.9
## - geo_enabled1 1 1099.7 1119.7
## - listed_count 1 1114.4 1134.4
## - default_profile_image1 1 1135.1 1155.1
## - default_profile1 1 1171.8 1191.8
## - days_since_last_tweet 1 1331.4 1351.4
## - verified1 1 1491.7 1511.7
## - friends_count 1 1803.9 1823.9
## - followers_count 1 1887.3 1907.3
## Start: AIC=1107.2
## .outcome ~ followers_count + friends_count + default_profile1 +
## default_profile_image1 + favourites_count + geo_enabled1 +
## account_age + days_since_last_tweet + listed_count + profile_background_tile1 +

```

```

##      status_favorite_count + verified1
##
##              Df Deviance   AIC
## - status_favorite_count      1    1058 1082
## - listed_count                1    1069 1093
## - favourites_count            1    1074 1098
## - account_age                 1    1076 1100
## <none>                        1    1081 1107
## - geo_enabled1                1    1143 1167
## - default_profile_image1      1    1170 1194
## - default_profile1            1    1193 1217
## - days_since_last_tweet       1    1294 1318
## - verified1                   1    1575 1599
## - friends_count               1    1780 1804
## - followers_count             1    1813 1837
## - profile_background_tile1    1   48298 48322
##
## Step:  AIC=1082.23
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      verified1
##
##              Df Deviance   AIC
## <none>                        1    1058 1082
## - listed_count                1    1069 1091
## - account_age                 1    1077 1099
## - favourites_count            1    1093 1115
## - geo_enabled1                1    1143 1165
## - default_profile_image1      1    1170 1192
## - default_profile1            1    1193 1215
## - days_since_last_tweet       1    1295 1317
## - verified1                   1    1575 1597
## - friends_count               1    1785 1807
## - followers_count             1    1828 1850
## - profile_background_tile1    1   37990 38012
## Start:  AIC=962.94
## .outcome ~ followers_count + friends_count + default_profile1 +
##      default_profile_image1 + favourites_count + geo_enabled1 +
##      account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##      status_favorite_count + verified1
##
##              Df Deviance   AIC
## <none>                        1    937 963
## - favourites_count            1    949 973
## - status_favorite_count       1    949 973
## - account_age                 1    953 977
## - geo_enabled1                1   1019 1043
## - default_profile1            1   1065 1089
## - default_profile_image1      1   1104 1128
## - verified1                   1   1283 1307
## - friends_count               1   1718 1742
## - followers_count             1   1799 1823
## - profile_background_tile1    1  32151 32175

```



```

## - listed_count          1    40513 40537
## - days_since_last_tweet 1    48082 48106
## Start:  AIC=54019.39
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##              Df Deviance   AIC
## - default_profile_image1 1    1163 1187
## - followers_count        1    1838 1862
## - friends_count          1    1851 1875
## - status_favorite_count   1   27826 27850
## - geo_enabled1           1   28258 28282
## - verified1              1   29340 29364
## - default_profile1       1   38495 38519
## - profile_background_tile1 1   38711 38735
## - days_since_last_tweet  1   39143 39167
## - favourites_count        1   52624 52648
## <none>                   53993 54019
## - listed_count           1   58319 58343
## - account_age            1   59040 59064
##
## Step:  AIC=1187.29
## .outcome ~ followers_count + friends_count + default_profile1 +
##     favourites_count + geo_enabled1 + account_age + days_since_last_tweet +
##     listed_count + profile_background_tile1 + status_favorite_count +
##     verified1
##
##              Df Deviance   AIC
## <none>                   1163 1187
## - default_profile1       1   1379 1401
## - followers_count        1   1946 1968
## - friends_count          1   1974 1996
## - geo_enabled1           1   25303 25325
## - account_age            1   26384 26406
## - listed_count           1   32944 32966
## - days_since_last_tweet  1   41811 41833
## - profile_background_tile1 1   42315 42337
## - status_favorite_count   1   47145 47167
## - favourites_count        1   48226 48248
## - verified1              1   50894 50916
## Start:  AIC=1045.49
## .outcome ~ followers_count + friends_count + default_profile1 +
##     default_profile_image1 + favourites_count + geo_enabled1 +
##     account_age + days_since_last_tweet + listed_count + profile_background_tile1 +
##     status_favorite_count + verified1
##
##              Df Deviance   AIC
## - profile_background_tile1 1   1020.3 1044.3
## <none>                   1019.5 1045.5
## - status_favorite_count   1   1029.4 1053.4
## - favourites_count        1   1034.7 1058.7
## - account_age            1   1037.1 1061.1

```

```
## - listed_count          1    1043.6 1067.6
## - geo_enabled1          1    1090.9 1114.9
## - default_profile1      1    1152.3 1176.3
## - default_profile_image1 1    1161.2 1185.2
## - days_since_last_tweet 1    1297.5 1321.5
## - verified1             1    1402.6 1426.6
## - friends_count         1    1814.0 1838.0
## - followers_count       1    1901.6 1925.6
##
## Step: AIC=1044.26
## .outcome ~ followers_count + friends_count + default_profile1 +
##   default_profile_image1 + favourites_count + geo_enabled1 +
##   account_age + days_since_last_tweet + listed_count + status_favorite_count +
##   verified1
##
##               Df Deviance    AIC
## <none>                1020.3 1044.3
## - status_favorite_count 1    1030.1 1052.1
## - favourites_count      1    1035.6 1057.6
## - account_age           1    1037.5 1059.5
## - listed_count          1    1044.9 1066.9
## - geo_enabled1          1    1091.1 1113.1
## - default_profile_image1 1    1161.4 1183.4
## - default_profile1      1    1162.7 1184.7
## - days_since_last_tweet 1    1300.3 1322.3
## - verified1             1    1403.8 1425.8
## - friends_count         1    1815.0 1837.0
## - followers_count       1    1903.0 1925.0
```

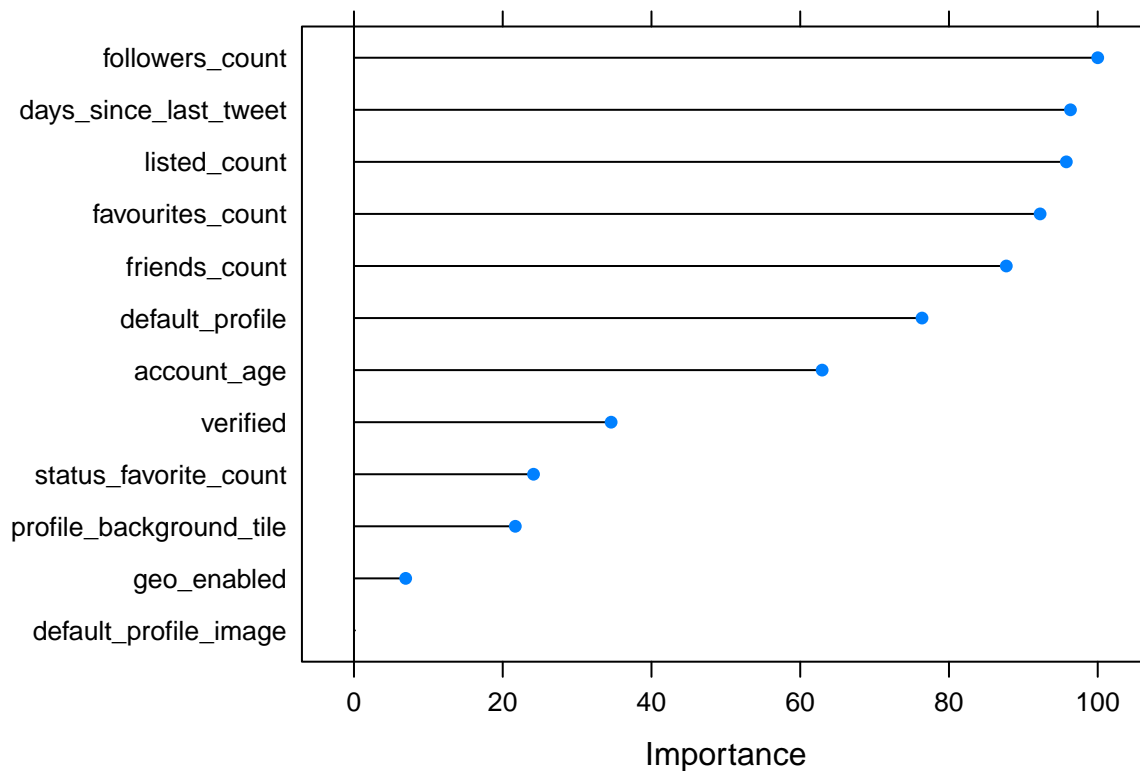
```
summary(step_model)
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5864  -0.0983   0.0001   0.0265   6.6165
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -6.54129     0.42539  -15.377 < 2e-16 ***
## followers_count -196.58360    10.36857  -18.960 < 2e-16 ***
## friends_count   68.84841     3.55951   19.342 < 2e-16 ***
## default_profile1 -0.97051     0.08618  -11.261 < 2e-16 ***
## default_profile_image1 -1.82876     0.20602   -8.877 < 2e-16 ***
## favourites_count -0.80147     0.20696   -3.873 0.000108 ***
## geo_enabled1    -0.73042     0.09217   -7.924 2.29e-15 ***
## account_age     -0.34498     0.08369   -4.122 3.75e-05 ***
## days_since_last_tweet 3.45246     0.27275   12.658 < 2e-16 ***
## listed_count     3.18778     0.53380    5.972 2.35e-09 ***
## status_favorite_count 0.13712     0.07634    1.796 0.072456 .
## verified1       2.48551     0.22903   10.852 < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 6209.7  on 4487  degrees of freedom
## Residual deviance: 1020.3  on 4476  degrees of freedom
## AIC: 1044.3
##
## Number of Fisher Scoring iterations: 14
```

```
plot(varImp(step_model))
```

```
## Loading required package: pROC
## Type 'citation("pROC")' for a citation.
##
## Attaching package: 'pROC'
##
## The following objects are masked from 'package:stats':
##
##    cov, smooth, var
```



```
step_predictions = predict(step_model, newdata = test)
confusionMatrix(step_predictions, test$bot)
```

```
## Confusion Matrix and Statistics
##
##           Reference
```

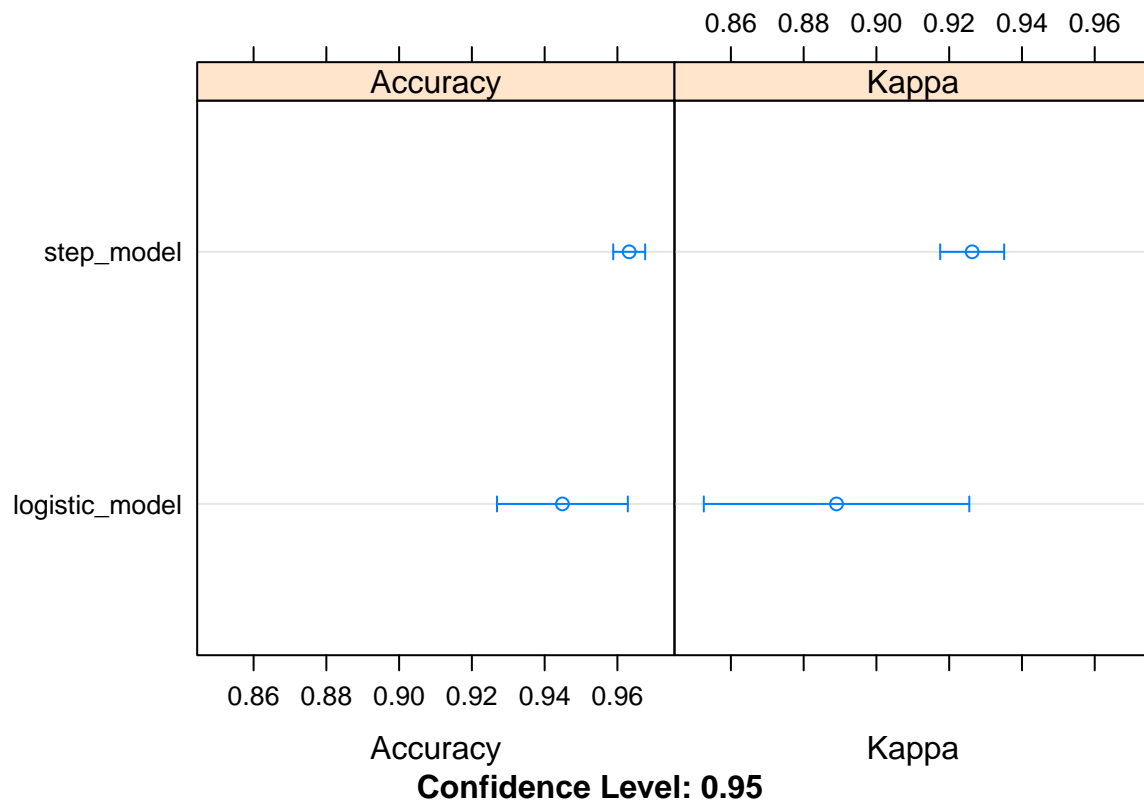
```
## Prediction    0    1
##              0 689  32
##              1  20 754
##
##              Accuracy : 0.9652
##              95% CI : (0.9546, 0.9739)
##      No Information Rate : 0.5258
##      P-Value [Acc > NIR] : <2e-16
##
##              Kappa : 0.9303
##  McNemar's Test P-Value : 0.1272
##
##      Sensitivity : 0.9718
##      Specificity : 0.9593
##      Pos Pred Value : 0.9556
##      Neg Pred Value : 0.9742
##      Prevalence : 0.4742
##      Detection Rate : 0.4609
##      Detection Prevalence : 0.4823
##      Balanced Accuracy : 0.9655
##
##      'Positive' Class : 0
##
```

```
# compare
results = resamples(list(logistic_model = logistic_model,
                        step_model = step_model))

# compare accuracy and kappa
summary(results)
```

```
##
## Call:
## summary.resamples(object = results)
##
## Models: logistic_model, step_model
## Number of resamples: 25
##
## Accuracy
##              Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
## logistic_model 0.8375  0.9409 0.9671 0.9449  0.9698 0.9735    0
## step_model     0.9240  0.9617 0.9648 0.9632  0.9693 0.9767    0
##
## Kappa
##              Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
## logistic_model 0.6706  0.8812 0.9337 0.8890  0.9395 0.9471    0
## step_model     0.8477  0.9232 0.9296 0.9263  0.9385 0.9534    0
```

```
# plot results
dotplot(results)
```



Now that we have a discrete binary outcome we'll use performance metrics called accuracy and kappa instead of R^2 and RMSE.