Exploratory Data Analysis

Lissa Harrop, Katrina Watkins, Ricky Loo and Max Tan

16 August 2022

Contents

1 Exploratory Data Analysis of our overall data **2**

2 Counts_profileVisits, counts_kisses and isVip **3**

3 Country, isVip and Distance **12**

4 UPDATE the path for the data here!!!!! **12**

5 Bivariate **19**

6 Counts_pictures, counts_profiles and counts_kisses **21**

6.1 Summary statistics 21

6.2 Univariate histograms 23

7 Bivariate EDA **23**

7.1 Variance and covariance 23

7.2 Correlation 23

7.3 Mahalanobis distance 23

7.4 Pairs scatter plots and distribution of Mahalanobis classification. 23

7.5 Pairs contour plots 24

8 UPDATE the path for the data here!!!!! **24**

9 Age, country and number of pictures **33**

10 Update file path **33**

11 References **35**

1 Exploratory Data Analysis of our overall data

We (group 10) decided to use the Dating App User Profiles' stats data set. The data set is available on kaggle (@Lovoov3) and the license to use the data ser is available on creativecommons (@License).

After some basic exploration of the variables available in the Lovoo v3 data set we decided to explore the variables age, counts_pictures, counts_profileVisits, counts_kisses, distance, country and isVip.

Age is the users age, **counts_pictures** is the number of pictures on the user's profile, **counts_profileVisits** is the number of clicks on this user (to see his/her full profile) from other user accounts, **counts_kisses** is the number of unique user accounts that "liked" (called "kiss" on the platform) this user account, **distance** is the distance between this user's city/location and the location of the user account that was used to fetch the data of this user, **country** is the user's country, **isVip** is a 1 if the user is VIP. [It was possible to buy a VIP status with real money. This status came with benefits.].

It was discovered that there were 46 missing values in the variable distance. These have been replaced by the mean of the distance column, 207.23.After replacing the 46 missing distance variables to ensure we have a full data set, we have a sample size of 3992 for all seven variables.

The ages of the user's of the lovoo app range from 18 years to 28 years with the median age being 22 year. The minimum number of pictures on a user's profile is 0 with the maximum being 30 pictures and the median being 4.The number of clicks on a user's profile to see his/her full profile (from another users account) ranges from 0 to 164425 clicks, with the median being 1222 clicks. The number of unique user accounts that "liked" a users account ranges from 0 to 9288 likes, with the median being 44 likes. The distance between this user's city/location and the location of the user account that was used to fetch the data of the user ranges from 0 to 6918, with the median being 173. These and other summary statistics can be seen in table 1.

The summary of the countries and their counts can be found in table 2 and a visualisation can be seen in figure 3. There are 32 different countries with varying numbers of users. Table 3 shows that 3901 users are not Vip's while only 91 are Vip's.

There appears to be strong positive correlation between the number of profiles visits and the number of likes that a user receives. There is also positive correlation between the number of pictures a user has and the number of profile visits they receive, as well as the number of likes the user has. There is slight positive correlation between the age of the user and the distance between this user's city/location and the location of the user

account that was used to fetch the data of the user. There appears to be no correlation between age and likes, profile visits and pictures, nor distance and likes, profile visits and pictures. This can been seen in figure 1 and supported by the pairs plots in figure 2.

Table 1: Summary Statistics - Numerical Variables

	age	counts_pictures	counts_profileVisits	counts_kisses	distance
sample size	3992.00	3992.00	3992.00	3992.00	3992.00
minimum	18.00	0.00	0.00	0.00	0.00
first quartile	20.00	2.00	383.00	11.00	85.27
median	22.00	4.00	1222.00	44.00	173.00
third quartile	24.00	6.00	4063.25	141.00	317.00
maximum	28.00	30.00	164425.00	9288.00	6918.00
IQR	4.00	4.00	3680.25	130.00	231.73
standard deviation	1.96	4.42	6845.04	377.65	195.46
mean	21.99	4.79	3705.47	156.60	207.23

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```

2 Counts_profileVisits, counts_kisses and isVip

How does profile visits affect profile likes?

We see that there is a high correlation (0.89) between profile visits and profile likes (figure 4). We see from figure 5 that there is a positive relationship between profile visits and profile likes. We can also see there are several outliers that we should remove as they can be very influential to our data. After removing the outliers from figure 5 we see the same positive relationship but the outlieing profiles are gone (figure 6).

Does having "VIP" mean you get more profile visits and likes?

The distribution of profile visits and profile likes is right skewed and definitely does not follow a normal distribution as seen in figure 7. The density graphs (figure **??**) for both profile visits and profile likes is right skewed. This means that for both profile likes and visits, the mean is greater than the median. The mean of the boxplot for profiles with "VIP" is less than the mean of the boxplot for profiles without "VIP", this means on average profiles with "VIP" get less profile visits and likes than profiles without "VIP" as shown in figure 8

```
##                       counts_profileVisits counts_kisses
## counts_profileVisits                 1.00          0.89
## counts_kisses                        0.89          1.00
```

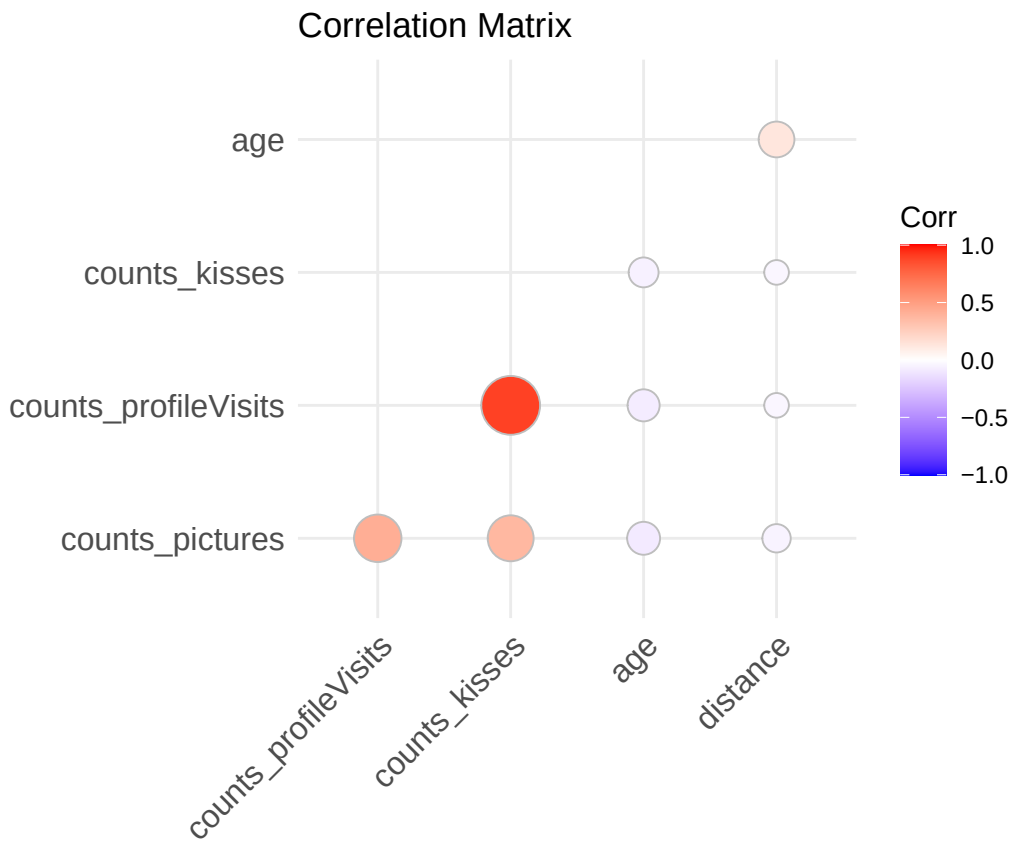


Figure 1: Correlation Matrix

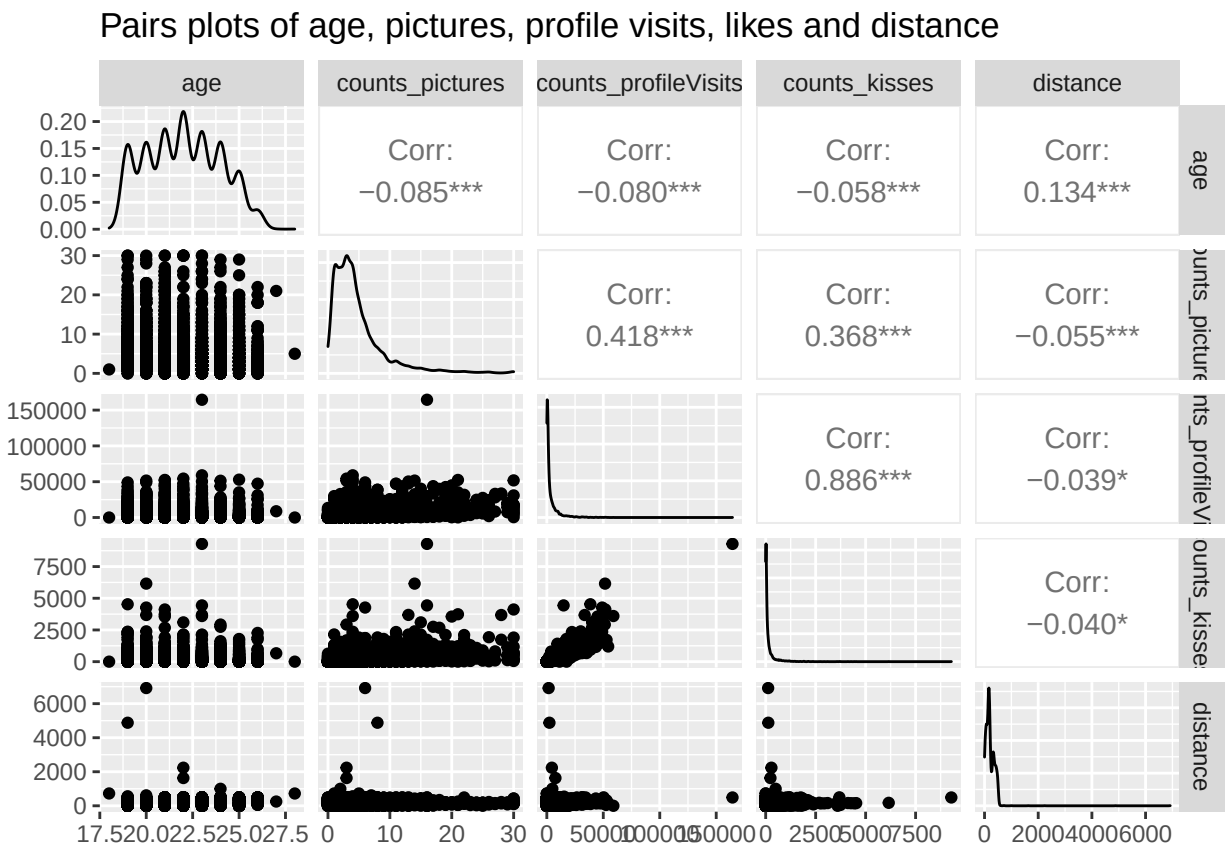


Figure 2: Pairs plot

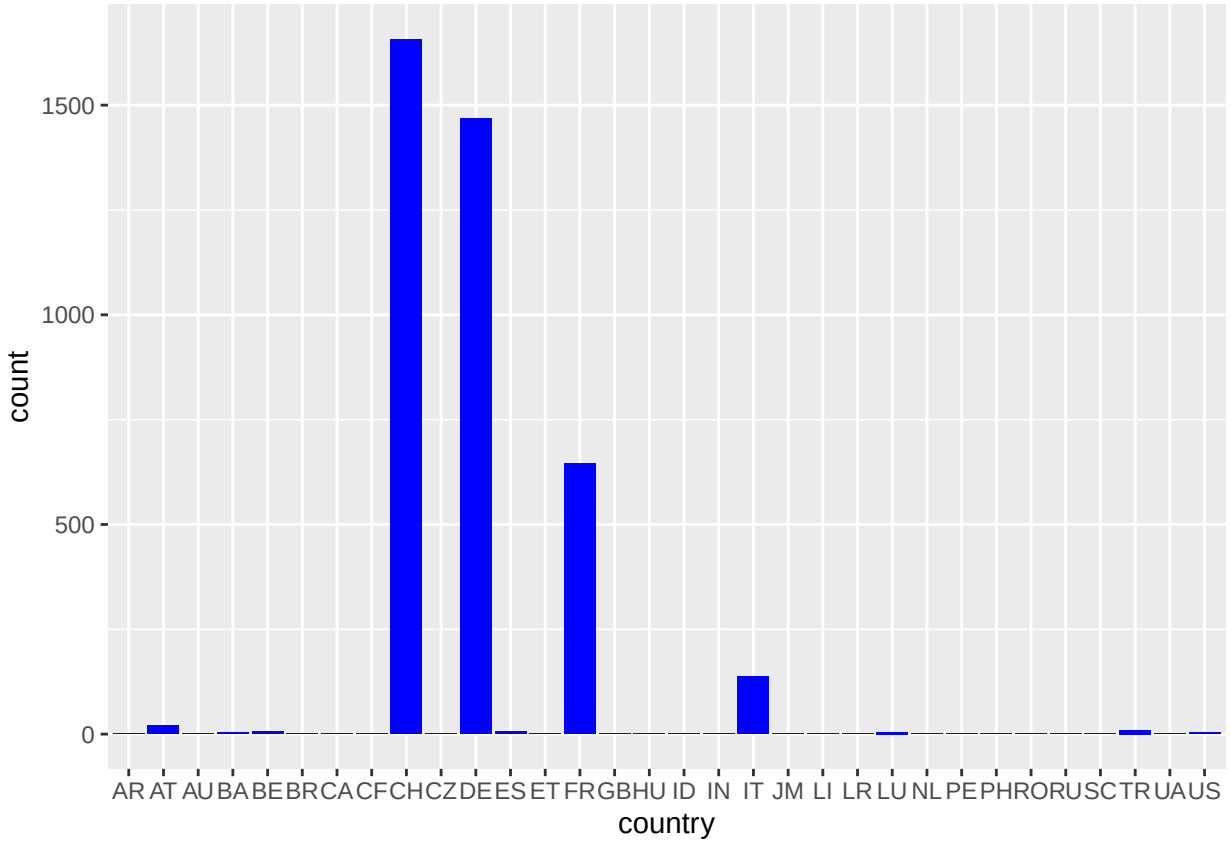


Figure 3: Number of user's by Country

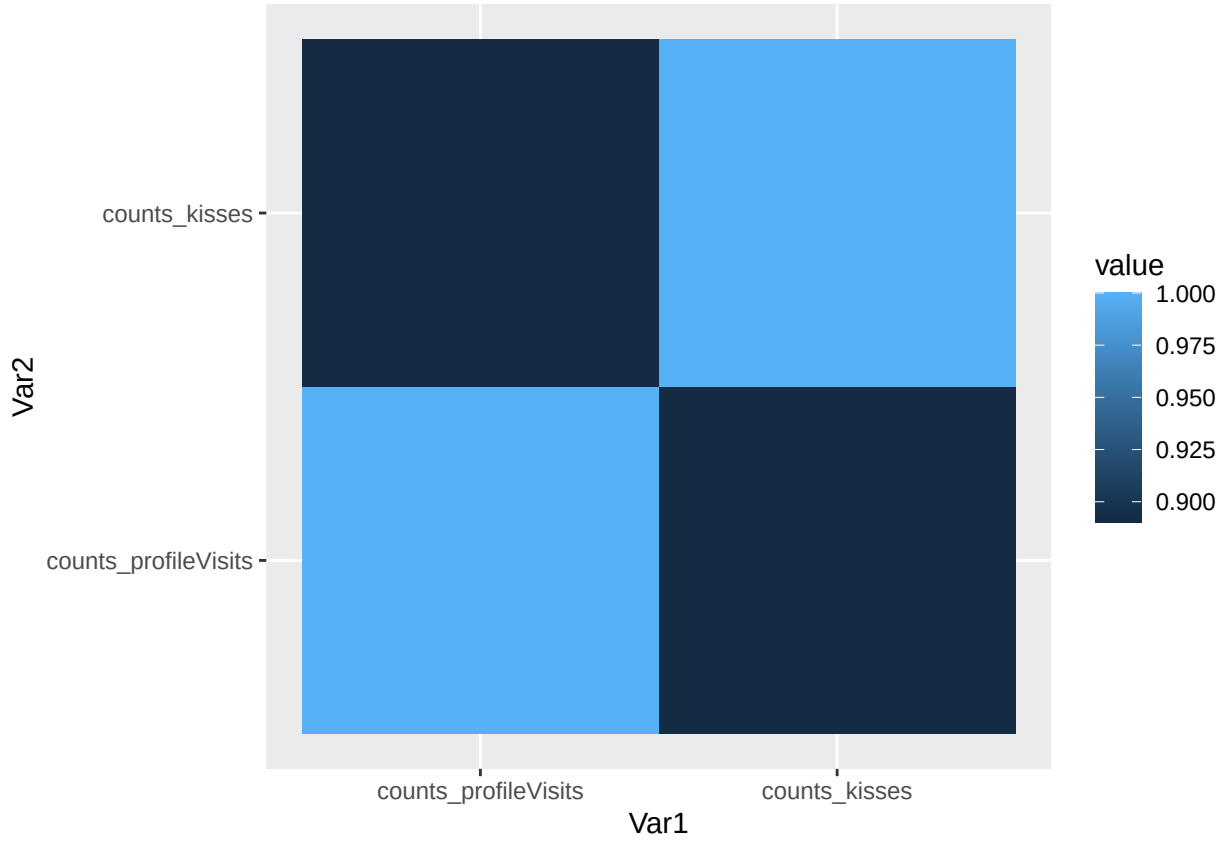


Figure 4: Correlation of profile visits and kisses

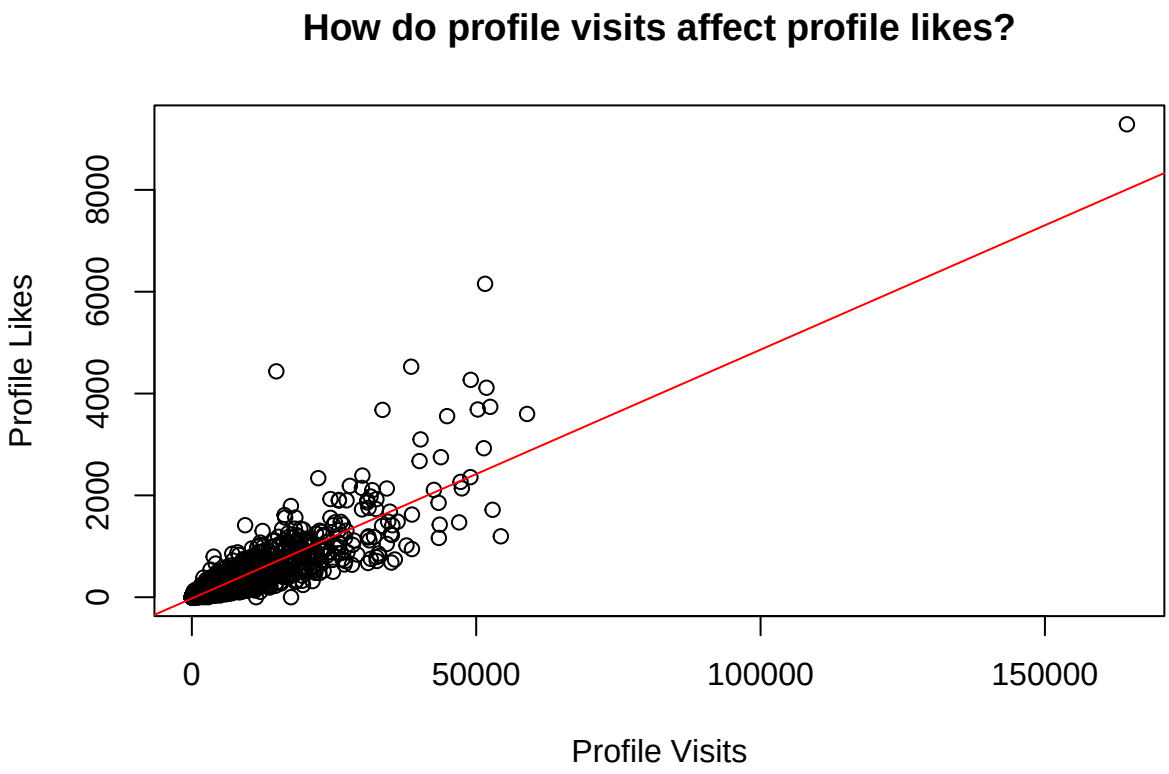


Figure 5: Scatterplot of profile visits vs profile likes

```
plot(lovoo$counts_profileVisits, lovoo$counts_kisses, main="How do profile visits affe
   xlab="Profile Visits", ylab="Profile Likes")
abline(lm(lovoo$counts_kisses ~ lovoo$counts_profileVisits, data = lovoo), col = "red"
```

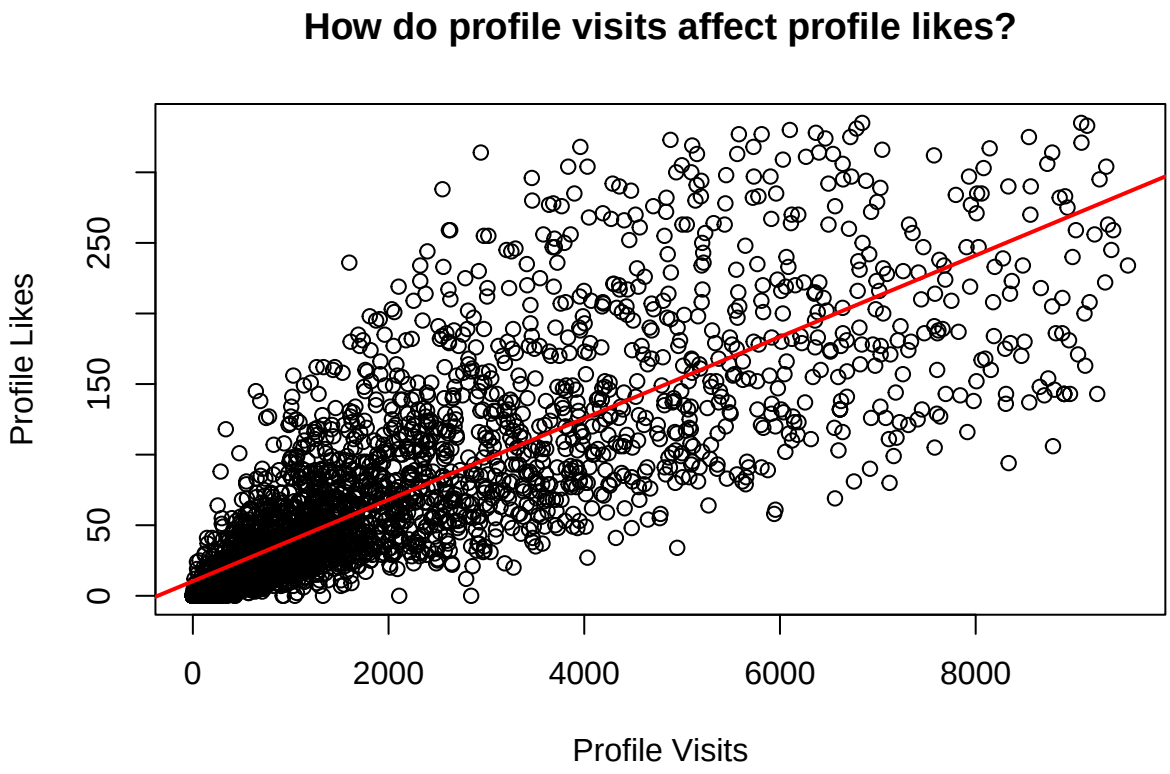


Figure 6: Revised scatterplot of profile visits vs profile likes

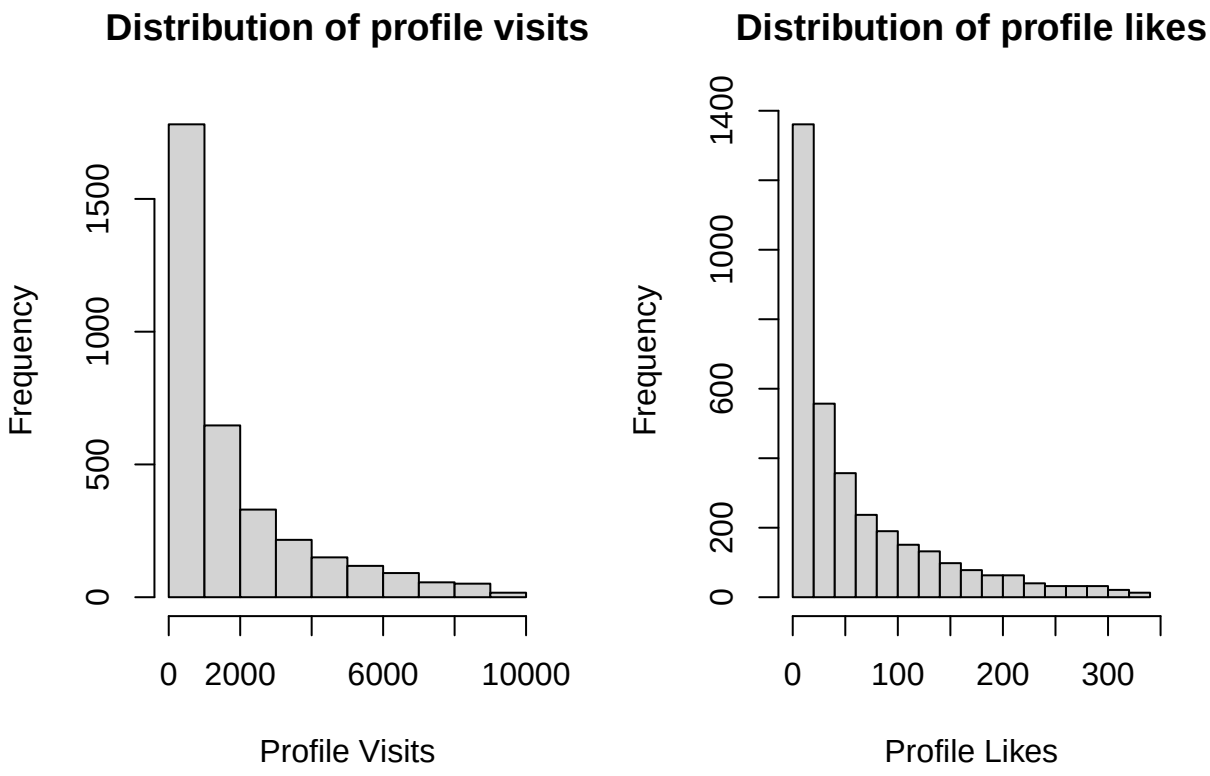
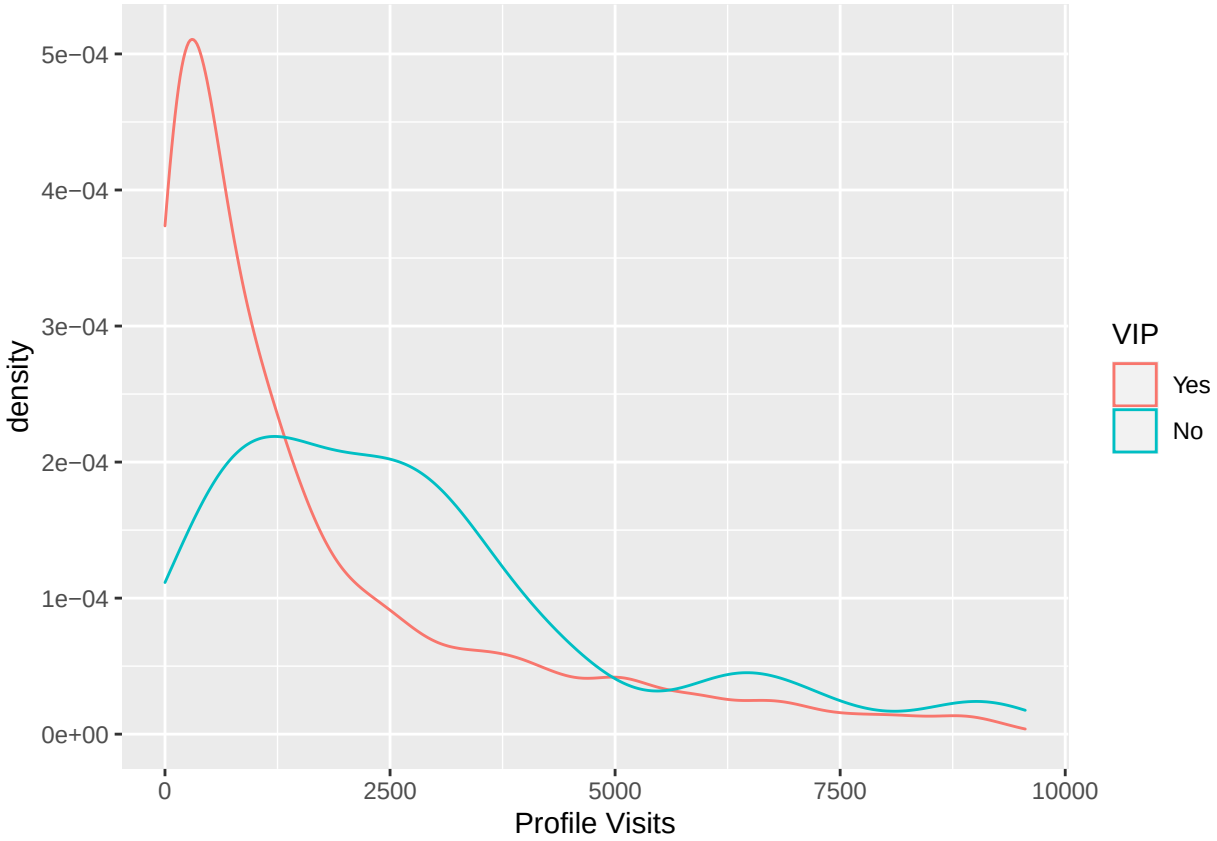
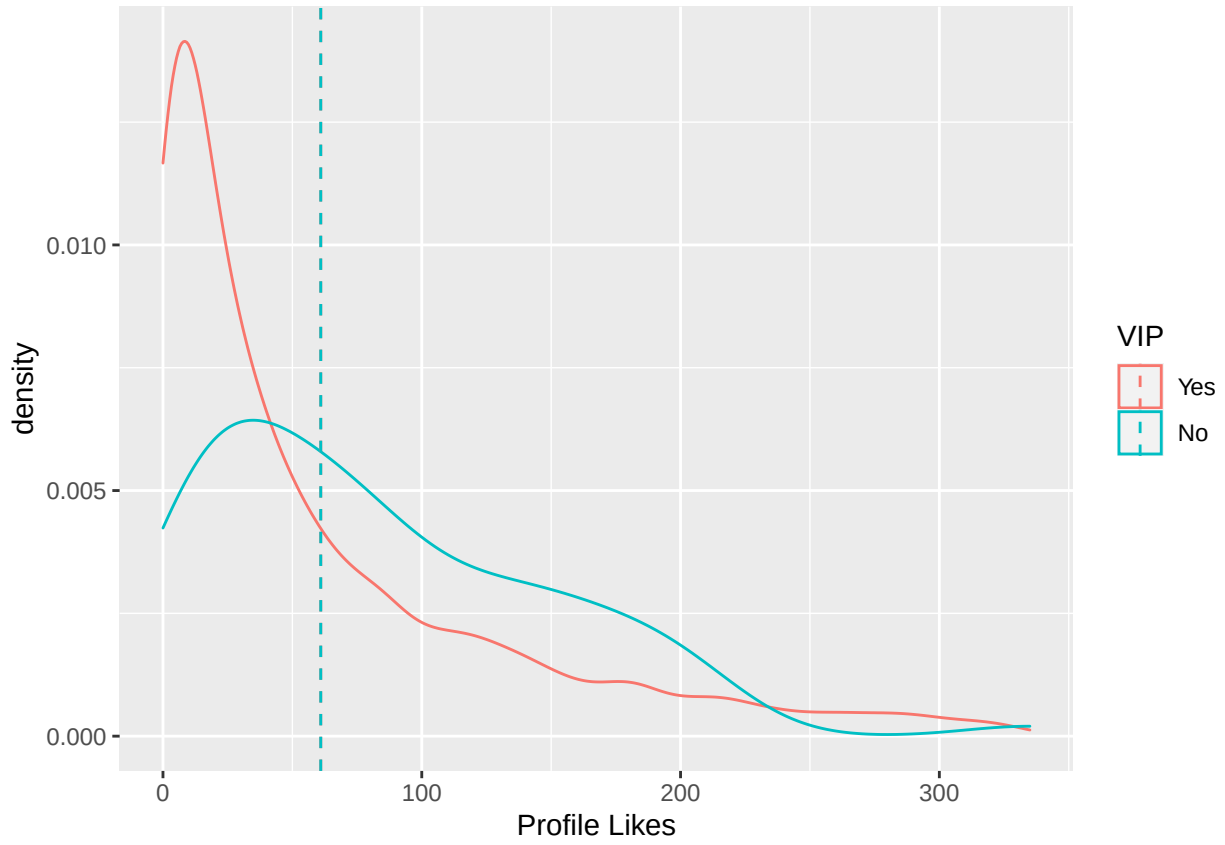


Figure 7: Histograms of Frequency of Profile Visits and Frequency of Profile Likes


5e−04
4e−04
3e−04
2e−04
1e−04
0e+00
density
0
2500
5000
7500
10000
Profile Visits
VIP
Yes
No



3 Country, isVip and Distance

Looking at the variables - country, isVIp and distance individually. It can be observed that the greatest number of users come from: Switzerland (CH), Germany (DE), France (FR) and Italy (IT) (figure 9. From figure 10 we can see that the majority of users haven't purchased VIP status. It can be observed that the distance data is left skewed. The majority of distance between this user's city/location and the location of the user account that was used to fetch the data is higher than the median (figures @fig(fig:histdistance) and @fig(fig:histdistance2)).

?? - need to comment on this 4 - need to comment on this

13 - this is showing the overall count It can be observed from this graph that CH, DE, FR and IT have the greatest number of users who have purchased VIP status.

4 UPDATE the path for the data here!!!!!

```
##
##   AR   AT   AU   BA   BE   BR   CA   CF   CH   CZ   DE   ES   ET   FR   GB   HU
```

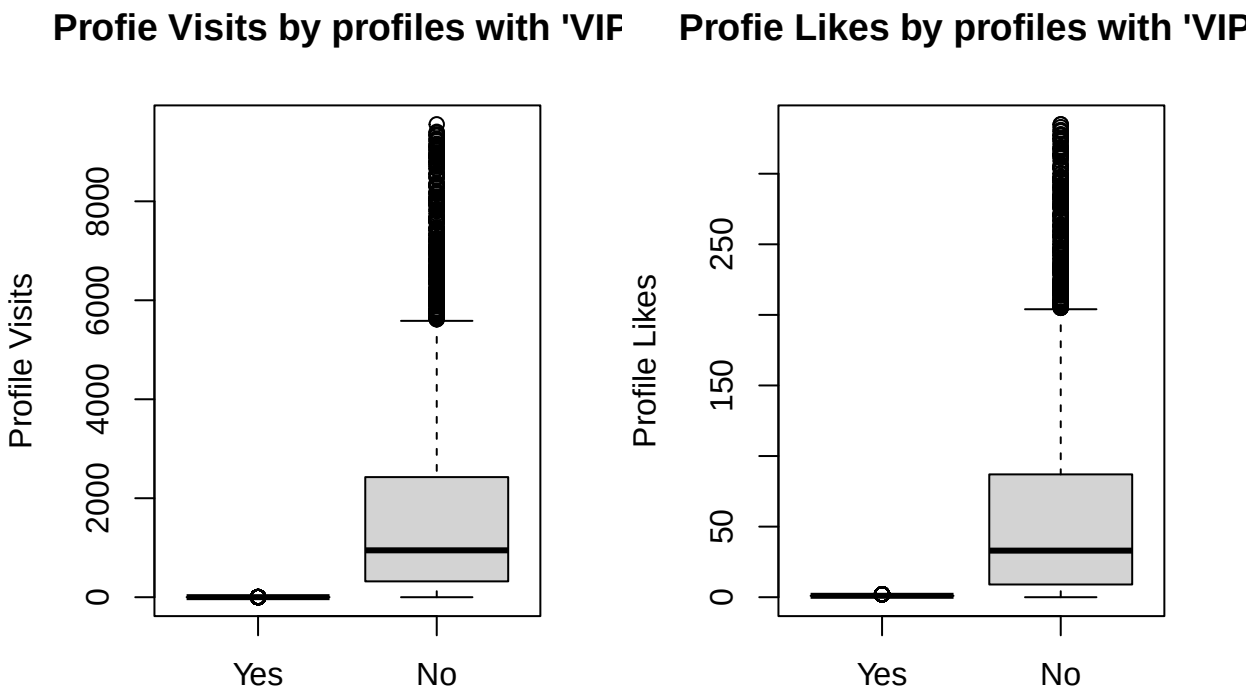


Figure 8: side by side boxplots of Profile Visits and Profile Likes against VIP status

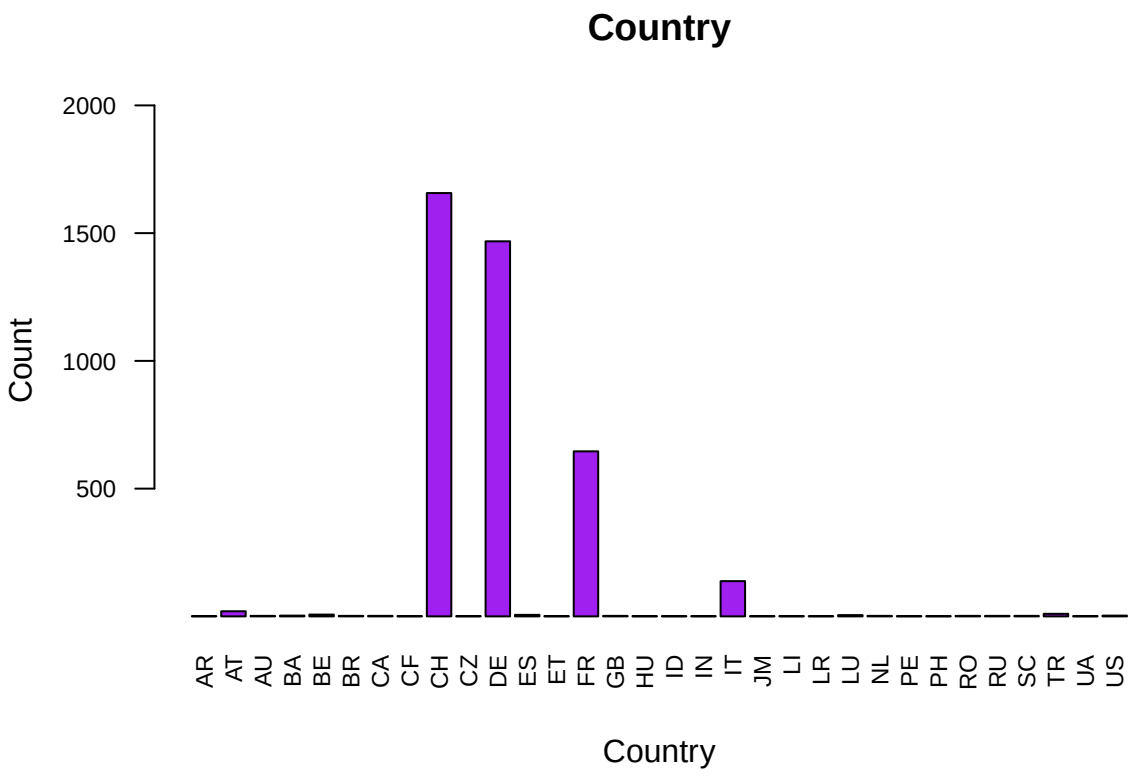


Figure 9: Number of users by country

```
##    1   20    2    3    7    2    2    1 1657    1 1468    6    1  646    2    1
##   ID   IN   IT   JM   LI   LR   LU   NL   PE   PH   RO   RU   SC   TR   UA   US
##    1    1  138    1    1    1    5    2    1    1    2    2    2   10    1    3
```

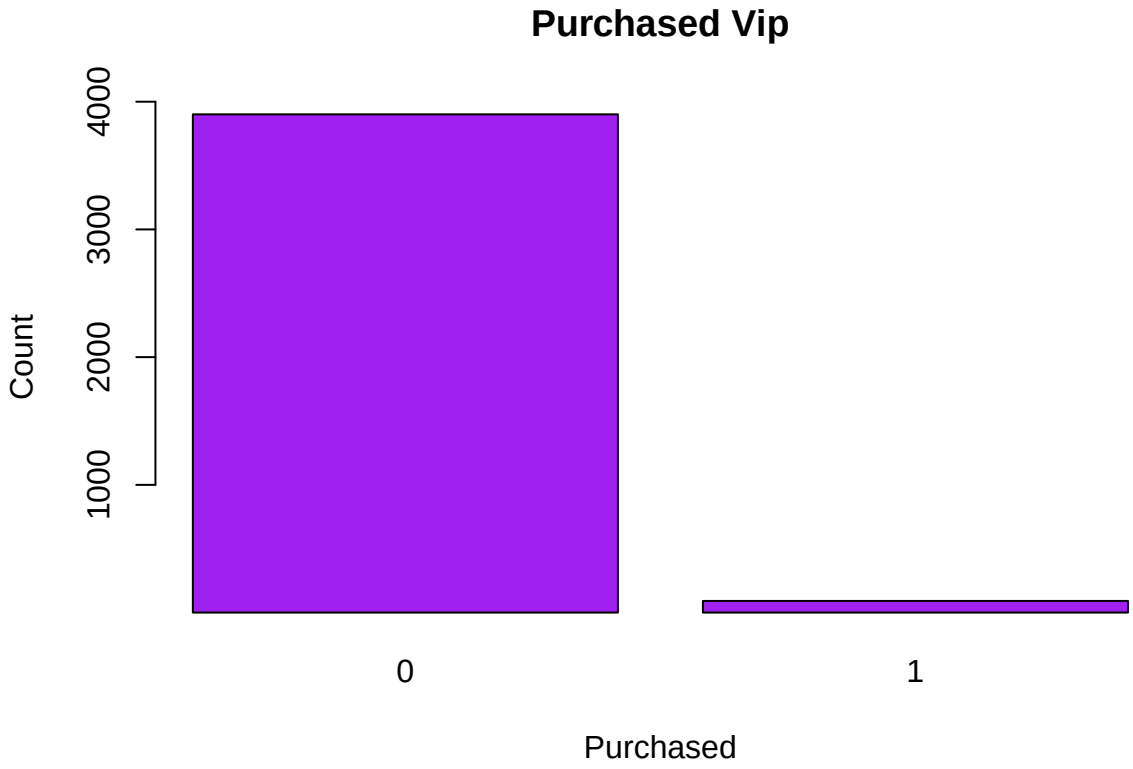


Figure 10: Count of whether a user is a VIP (nor not)

```
##
##    0    1
## 3901   91
```

```
##
##    0  0.1  0.2  0.3  0.4  0.6  0.7  0.8  0.9    1  1.1  1.2  1.3  1.4  1.5  1.6
##  125    7    2    2    2    2    3    6    2    7    4    2    1    4    2    6
##  1.7  1.8    2  2.1  2.3  2.4  2.5  2.6  2.7  2.8    3  3.2  3.3  3.4  3.5  3.6
##    3    2    2    3    1    3    3    1    2    1    2    1    1    1    1    1
##  3.8    4  4.1  4.2  4.3  4.4  4.5  4.6  4.7  4.8  4.9    5  5.2  5.3  5.5  5.6
##    2    2    1    1    1    1    1    1    3    5    1    2    2    3    1    2
##  5.7  6.4  6.8  6.9    7  7.1  7.2  7.3  7.9  8.2  9.4  9.9   10 10.5 10.6 10.7
##    1    2    1    1    1    1    1    1    1    1    1    1    1    2    1    2
```

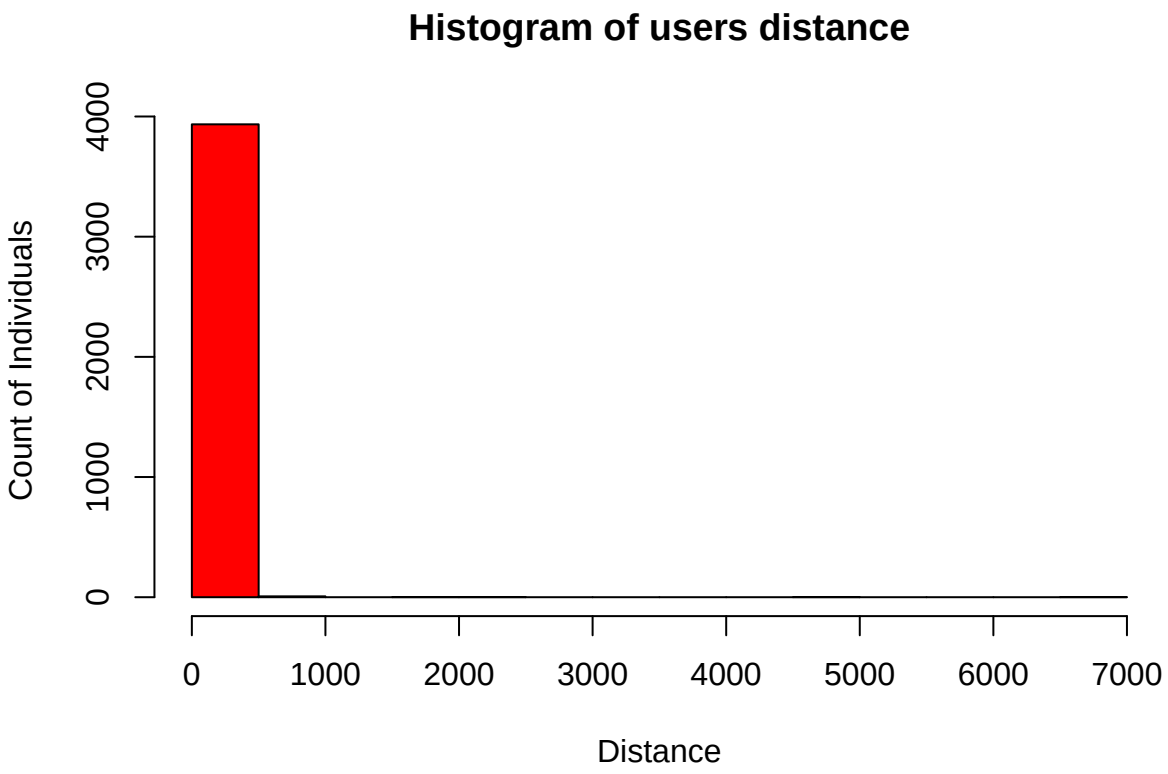


Figure 11: Histogram of individuals distance

```
## 10.8   11 11.4 11.7 12.2 12.3 12.6 12.8   13 13.2 13.4 13.6 13.8 14.3 14.4 14.6
##    2    1    1    1    1    1    1    1    1    2    1    1    1    1    1    1
## 14.7 14.9 15.2 15.4 15.5   16 16.1 16.4   17 17.1 17.2 17.3 17.4 17.5 17.6 17.7
##    2    1    1    1    1    1    1    1    1    1    1    1    1    1    2    1
## 17.8   18 18.1 18.2 18.6   19 19.2 19.4 19.7 19.8 19.9   20 20.1 20.2 20.3 20.4
##    1    1    1    4    1    1    4    1    1    2    1    1    2    2    1    1
## 20.5 20.8 20.9 21.3 21.4 21.5 21.7   22 22.2 22.3 22.7 22.8 22.9 23.1 23.2 23.3
##    2    3    1    1    1    1    4    3    1    1    1    2    1    2    2    1
## 23.4 23.9   24 24.6   25 25.9 26.4 26.8 27.2 27.5 27.6 27.9   28 28.4 28.5 28.7
##    4    1    1    2    1    1    2    2    1    1    1    1    2    3    1    1
## 28.8 28.9   29 29.3 29.4 29.6 29.7 29.9   30 31.5 31.7 31.8 32.2 32.3 32.6 32.7
##    2    1    1    1    2    1    1    1    1    1    1    2    1    1    1    1
## 32.8 33.5 33.7 33.8 33.9   34 34.2 34.3 34.4 34.5 34.9   35 35.1 35.3 35.5 35.9
##    1    3    1    2    1    2    5    3    2    1    1    1    2    1    1    1
## 36.6 36.8 37.2 37.4 37.6 37.9 38.1 38.2 38.3 38.4 38.6 38.7 38.8 39.3 39.4 39.6
##    1    1    1    1    1    1    2    2    1    1    2    1    2    1    2    1
## 39.7 40.2 40.3 40.5   41 41.2 41.3 41.4 41.7 41.8   42 42.1 43.4 43.6   44 44.2
##    1    2    1    1    2    1    2    1    1    1    3    2    1    1    1    1
## 44.5 44.7 44.9   45 45.8 45.9 46.1 46.2 46.3 46.5 46.9   47 47.2 47.3 47.4 47.5
##    1    1    2    1    1    1    2    1    2    1    1    2    2    4    2    2
## 47.6 47.8 47.9   48 48.2 48.4 48.9   49 49.2 49.4 49.5 49.6 49.7 49.8 49.9 50.1
##    1    1    3    3    3    1    3    3    2    3    3    1    3    2    1    3
## 50.2 50.3 50.4 50.5 50.6 50.7 50.9   51 51.1 51.2 51.3 51.4 51.5 51.6 51.7 51.8
##    8    3    3    2    4    5    5    2    1    2    1    3    2    5    3    3
## 51.9   52 52.1 52.2 52.3 52.4 52.5 52.6 52.7 52.8 52.9   53 53.1 53.2 53.3 53.5
##    1    2    1    3    3    4    4    5    3    2    5    2    3    4    2    2
## 53.6 53.9   54 54.3 54.4 54.6 54.7 54.8 54.9   55 55.1 55.3 55.4 55.5 55.6 55.7
##    3    1    1    1    2    2    3    2    2    1    4    1    3    1    2    2
## 55.8 55.9   56 56.1 56.4 56.6 56.7 56.8 56.9   57 57.1 57.2 57.4 57.5 57.6 57.7
##    1    1    2    1    2    2    1    2    1    1    1    2    1    2    1    2
## 57.8 57.9 58.1 58.3 58.8   59 59.2 59.7 59.8   60 60.4 60.5 60.6 60.7 60.8 60.9
##    2    1    2    1    2    2    1    1    1    4    1    2    2    1    1    1
## 61.3 61.9 62.1 62.2 62.5 62.6 62.7   63 63.2   64 64.1 64.2 64.3 64.6 64.9   65
##    2    1    1    1    1    2    2    2    3    1    1    1    2    1    1    1
## 65.6 65.7 65.9 66.1 66.2 66.3 66.4 66.6 66.7 66.9 67.2 67.3 67.5 67.6   68 68.4
##    1    2    1    1    1    1    1    3    1    1    1    1    1    1    1    1
## 69.1 69.5   70 70.2 70.4 70.5 70.7 70.8 70.9 71.1 71.3 71.4 71.6 71.7 72.1 72.6
##    1    2    2    1    1    1    1    1    2    1    1    1    1    1    2    1
## 72.7 72.9   73 73.2 73.4 73.5 73.7 73.8 73.9 74.3 74.4 74.5 74.6 74.7 74.8   75
##    1    1    1    2    1    1    1    4    1    1    2    4    2    2    2    1
## 75.2 75.3 75.4 75.5 75.6 75.7 75.8 76.1 76.2 76.3 76.4 76.5 76.6 76.7 76.8 76.9
##    4    2    1    3    2    1    1    4    2    1    2    4    2    3    1    4
##   77 77.2 77.3 77.4 77.5 77.6 77.7 77.8 77.9 78.1 78.2 78.5 78.7 78.8 78.9   79
##   43    1    4    1    5    1    1    1    2    2    2    2    1    2    3    1
## 79.1 79.2 79.3 79.4 79.5 79.6 79.7 79.9   80 80.1 80.2 80.3 80.4 80.5 80.6 80.7
```

```
##    2    2    3    2    5    1    3    4    4    4    2    2    2    1    2    1
## 80.8 80.9   81 81.2 81.3 81.4 81.5 81.6 81.8 81.9   82 82.1 82.2 82.4 82.5 82.7
##    2    2    4    1    2    1    1    2    1    2    4    4    1    2    1    2
## 82.8 82.9   83 83.1 83.2 83.3 83.4 83.6 83.7 83.8 84.1 84.2 84.4 84.8   85 85.1
##    2    3    2    2    1    2    7    1    1    1    2    1    2    2    1    1
## 85.2 85.3 85.7   86 86.3 86.4 86.5 86.7 86.9 87.2 87.5 88.1 88.5 88.7 89.1 89.5
##    2    1    2    1    1    2    2    2    2    2    2    1    2    2    2    2
## 89.8   90 90.3 90.7 90.8   91 91.3 91.4 91.7 91.8   92 92.4 92.5 92.7 92.8 93.1
##    1    1    1    1    2    1    2    1    1    2    2    2    1    4    2    1
## 93.2 93.3 93.6 93.7 93.8 93.9   94 94.1 94.2 94.3 94.9   95 95.1 95.2 95.4 95.5
##    2    1    1    2    2    1    2    1    1    2    1    4    1    3    1    1
## 95.6 95.7 95.8   96 96.2 96.5 96.7 96.9   97 97.3 97.5 97.7 97.8 98.4 98.7 98.8
##    1    1    2    1    3    1    4    1    2    1    1    1    1    1    2    1
## 98.9 99.1 99.2 99.7  100  101  102  103  104  105  106  107  108  109  110  111
##    1    2    1    1    2    7   16    6    6    2    7    3    6    1    6    1
##  112  113  114  115  116  117  118  119  120  121  122  123  124  125  126  127
##    3    2    4    1    3    8    4    4    7    6    4    4    2    6    4    3
##  128  129  130  131  132  133  134  135  136  137  138  139  140  141  142  143
##   16   14   13   17    5   13    7   22   37   21   21   24   34   18   13    8
##  144  145  146  147  148  149  150  151  152  153  154  155  156  157  158  159
##   21   27   10    8    8   13   11   16   19   13   11    4   18   13    9   15
##  160  161  162  163  164  165  166  167  168  169  170  171  172  173  174  175
##   27   25   30   14   15   13   13   15   12   16   12   31   22   88   28   19
##  176  177  178  179  180  181  182  183  184  185  186  187  188  189  190  191
##   15   16   15   15   20   14   27   22    8   19   18   28   20    6   11    8
##  192  193  194  195  196  197  198  199  200  201  202  203  204  205  206  207
##    9    9    4    4    6    3    9    1    6    4    1    1    5    7    7    6
##  208  209  210  211  212  213  214  215  216  217  218  219  220  221  222  223
##    2    2    3    5    8    5    3    5    5    7    5    2    9    2   13    9
##  224  225  226  227  228  229  230  231  232  233  234  235  237  239  240  241
##    1    3    4    2    3   10    5    7    1    5    2    2    7    3    6    3
##  242  243  244  245  246  247  248  249  250  251  252  253  254  255  256  257
##    2   11    6    4    2    2    2    4    4    3    3    3    2    2    1    1
##  258  260  261  262  263  264  265  267  268  269  270  272  273  274  275  276
##    5    5    2    1    3    1    3    3    3    2    1    2    5    3    1    3
##  277  278  279  280  281  282  283  285  286  287  288  289  290  291  292  293
##    5    2    1    2    3    5    3    1    7    2    3    3    3    3    2    2
##  294  295  296  297  298  299  300  301  302  303  304  305  306  307  308  309
##    4    5    2    2    1    2    2    7    9   16    6    5   23    9    6   13
##  310  311  312  313  314  315  316  317  318  319  320  321  322  323  324  325
##    4    7    9   11   11   35   12   16   10    8   13    4    8    7    5    7
##  326  327  328  329  330  331  332  333  334  335  336  337  338  339  340  341
##   11    8    7   11    2   11    5    9    5    9    7    5    6    6    4    6
##  342  343  344  345  346  347  348  349  350  351  352  353  354  355  356  357
##    4    3    2    6    8    6    9   10    6    5    3    3    7    6    9    7
```

```
##  358  359  360  361  362  363  364  365  366  367  368  369  370  371  372  373
##    8    8   10    4    8    2    2    4    4    6   10    3    8    5    4    3
##  374  375  376  377  378  379  380  381  382  383  384  385  387  388  389  390
##    5    2    4    1    3    3    5    5    2    2    2    2    2    5    3    1
##  391  392  393  394  395  396  397  398  399  400  401  402  403  404  405  406
##    3    6    6    6    3    3    3    3    4    3    5   11   11    4    6    2
##  407  408  409  410  411  412  413  414  416  417  418  421  422  423  424  425
##    5    9    7    2    7    9   10    6    5    6    5    6    4    4    1    3
##  426  427  428  429  430  431  432  433  434  435  436  437  438  439  440  441
##    7   13   45    7    9    6    7    3    5    6    2    3    5    3    4    2
##  442  443  444  445  446  447  448  449  450  451  452  453  454  456  457  458
##    3    2    3    3    6    2    2    1    2    3    3    3    1    1    4    1
##  460  461  462  463  464  465  466  467  468  469  470  471  472  473  474  475
##    2    3    3    3    4    2    3    2    7    1    2    4    5    4    6   12
##  476  477  478  479  480  481  482  483  484  485  486  487  488  489  490  491
##    4    8    8    6    3    1    1    6    4    4    3    4    8    9    6   14
##  492  493  494  495  496  498  499  502  538  558  613  723  724  999 1632 2238
##   47    9    5   11    5    3    1    1    1    1    1    1    1    1    1    1
## 4880 6918
##    1    1
```

5 Bivariate

5.0.1 Correlation between distance and isVip

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```

```
## Warning: Removed 46 rows containing missing values (stat_boxplot).
```

```
## Warning: Removed 125 rows containing non-finite values (stat_boxplot).
```

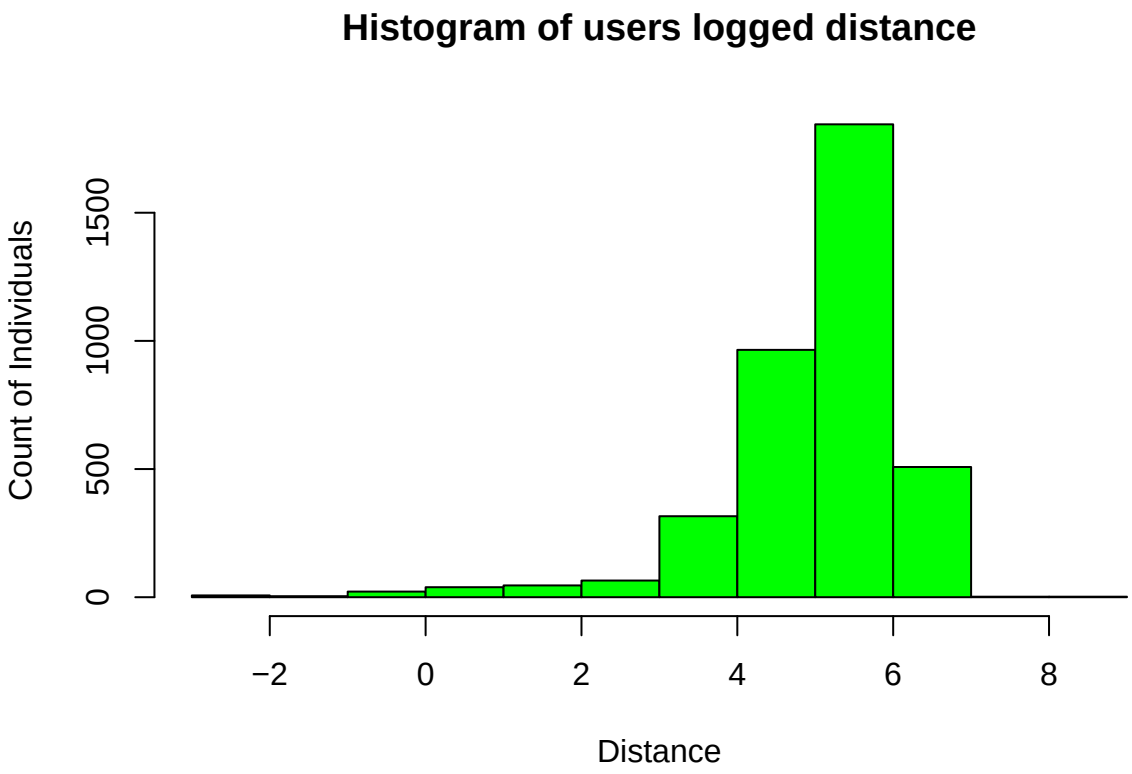
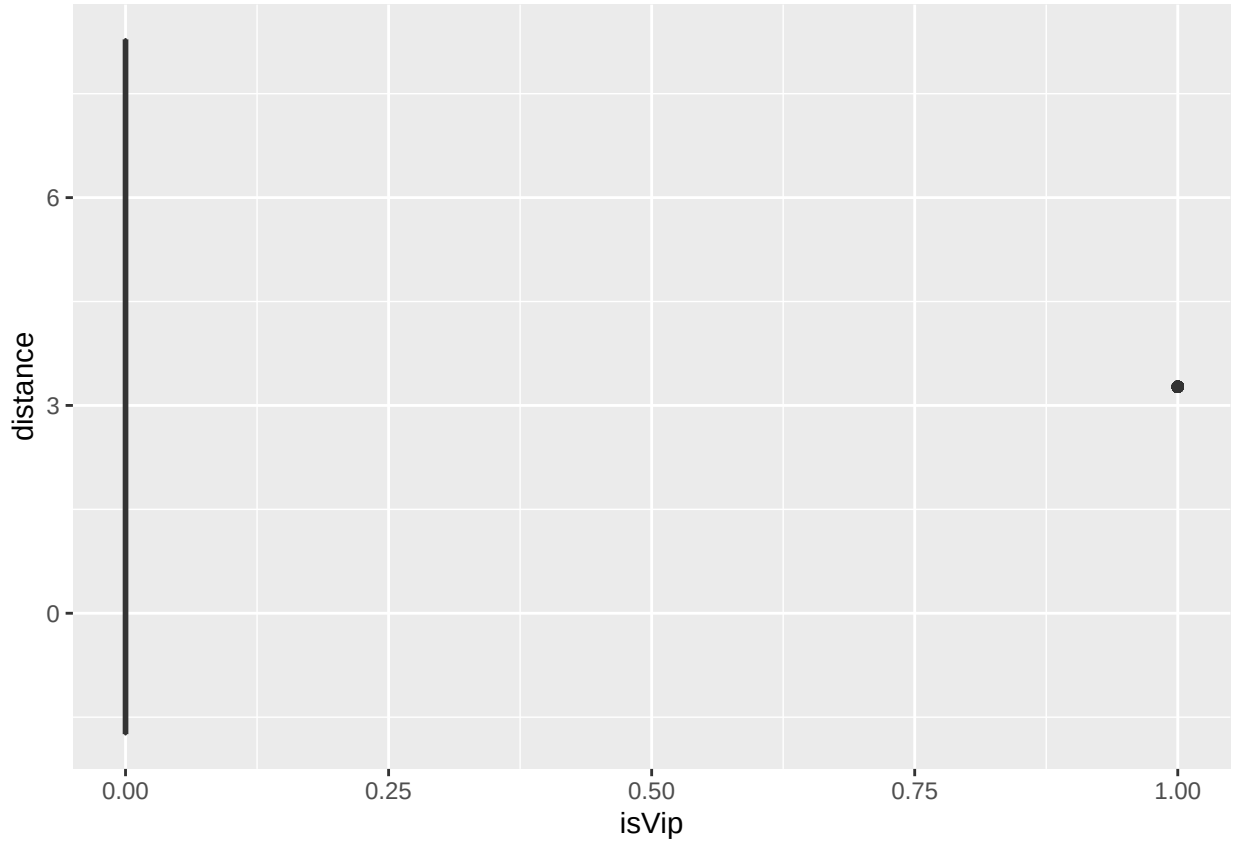


Figure 12: Histogram of users logged distance



6 Counts_pictures, counts_profiles and counts_kisses

6.1 Summary statistics

6.1.1 Pictures

Heavy right tail, showing most women have about 4 photos attached to their accounts and the maximum the can choose to have is 30. All values are positive.

6.1.2 Profile visits

Slightly more symmetric than the pictures distribution but still has a heavy skew with some extreme values and large range. This variable would benefit from the analysis of distances. All values are positive.

6.1.3 Kisses

Another heavy right tail with large range. All values are positive.

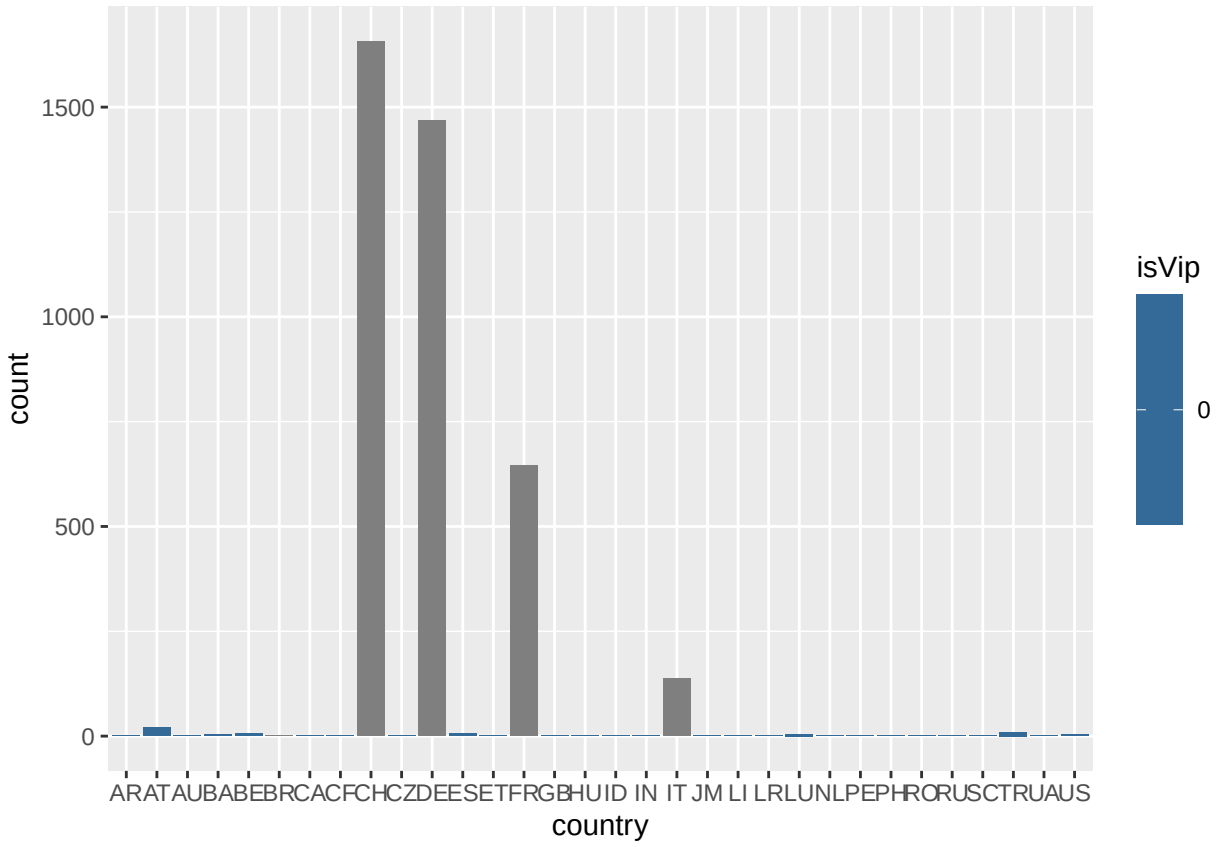


Figure 13: Count of users by country, indicating if

6.2 Univariate histograms

Figure 14 shows the Square-root transformed (for readability) distribution of counts_pictures. As we can see, is it pretty normal to have between 0 & 12 photos as this is where the boxplot spans. It is heavily centered around the lower end of the count. Figure 15 shows the Square-root transformed distribution of counts_profileVisits. There has been an extreme value (100,000 visits counted) removed as it made the graph unreadable. It is heavily centered around the lower end of the count. Figure 16 shows the Square-root transformed distribution of counts_kisses. It is heavily centered around the lower end of the count.

7 Bivariate EDA

7.1 Variance and covariance

All variables are positively related

7.2 Correlation

All variables are positively related, this is what we expect as with more visits there will be more likes and so on. Figure 17 shows that there is a correlation of interest between profile visits and kisses.

7.3 Mahalanobis distance

As we can see there are some very surprising values (108) from the table and Figure 18, it could be worth while examining these and possibly removing them as outliers. These might be famous people that receive a lot of attention online or other cases that are not simply a single woman looking for love online.

7.4 Pairs scatter plots and distribution of Mahalanobis classification.

Figure 19 further shows where the very surprising values are - they are mainly on the fringes of the data. although in the univariate distributions we can see that most very suprising values do make up the lower end of the photos which is what we would assume was normal. There will have to be more discussion around the exclusion of points from the data.

The contour plots are giving a warning message, I think due to scale

7.5 Pairs contour plots

All contour plots (Figure 20, Figure 21, Figure 22) are centered around zero and spread with a positive relationship to each other.

8 UPDATE the path for the data here!!!!!

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

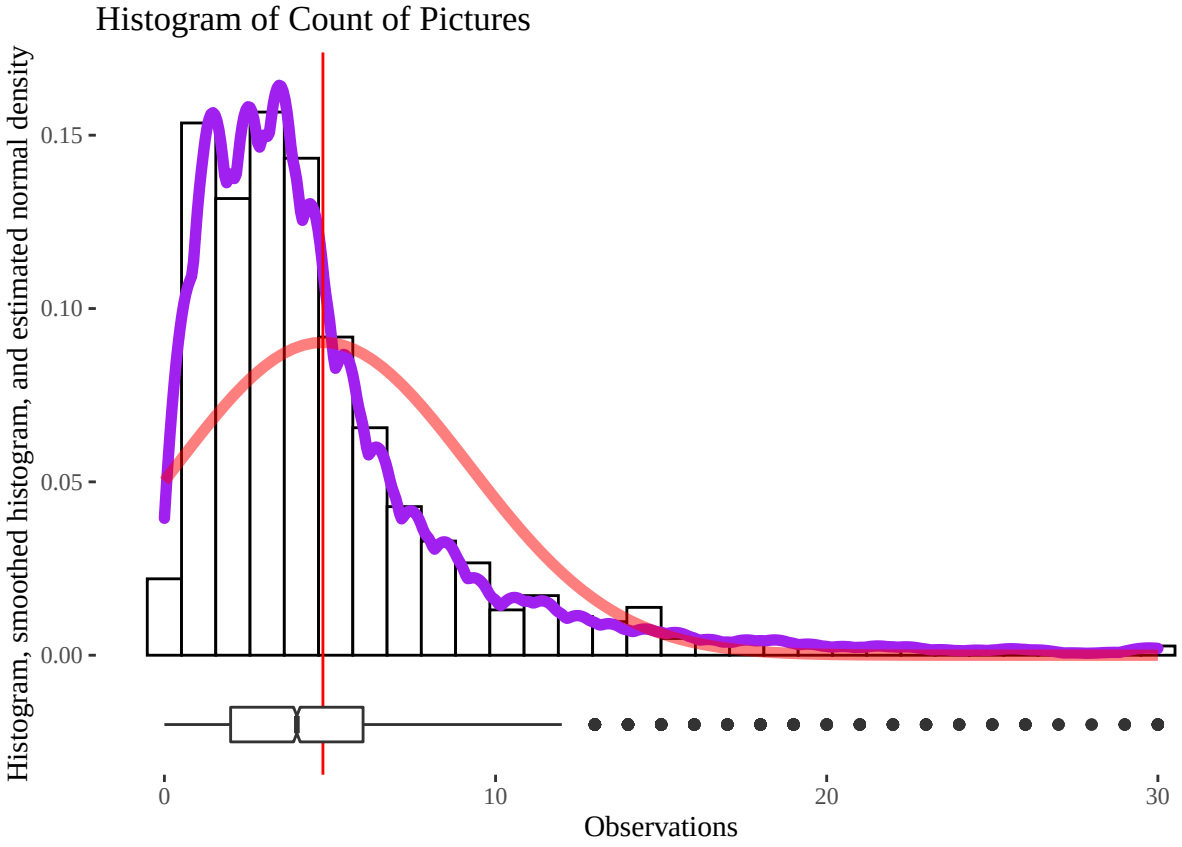


Figure 14: Square-root transformed histogram of counts_pictures

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

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

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```

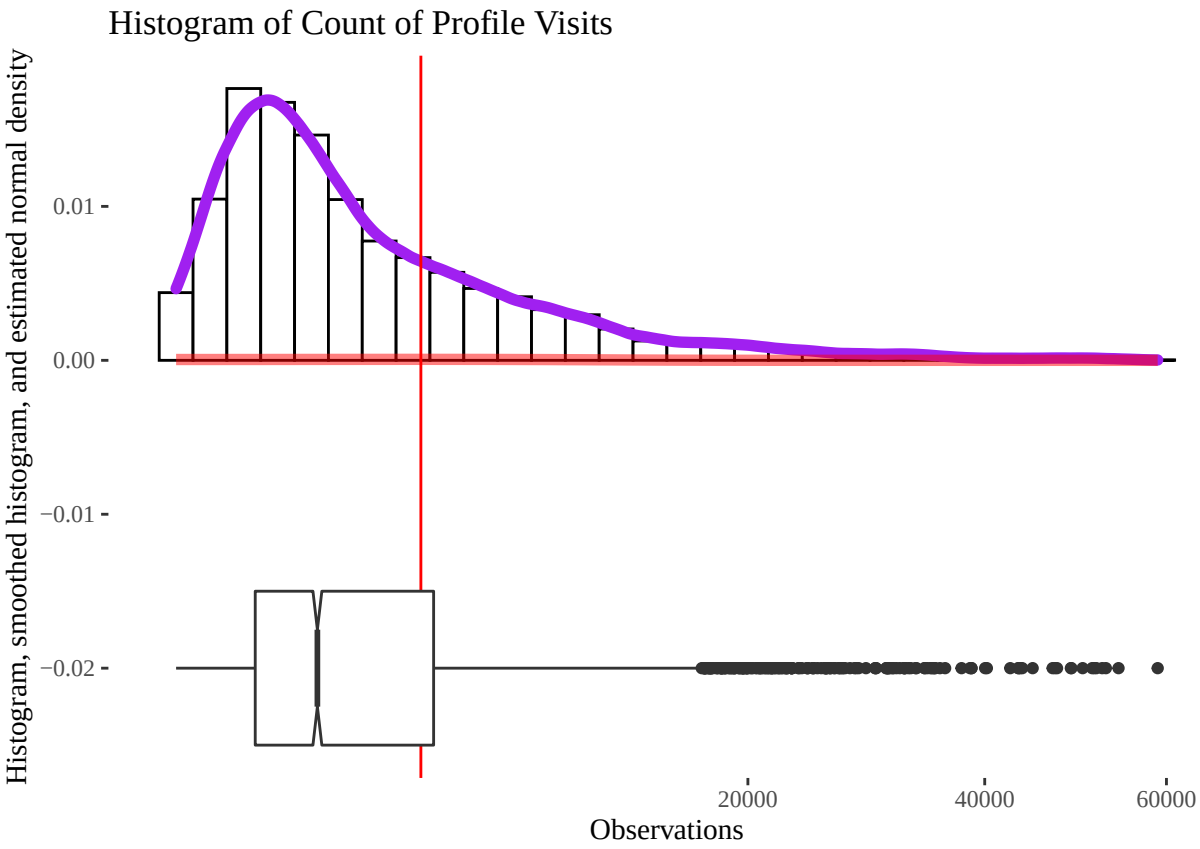


Figure 15: Square-root transformed histogram of counts_profileVisits

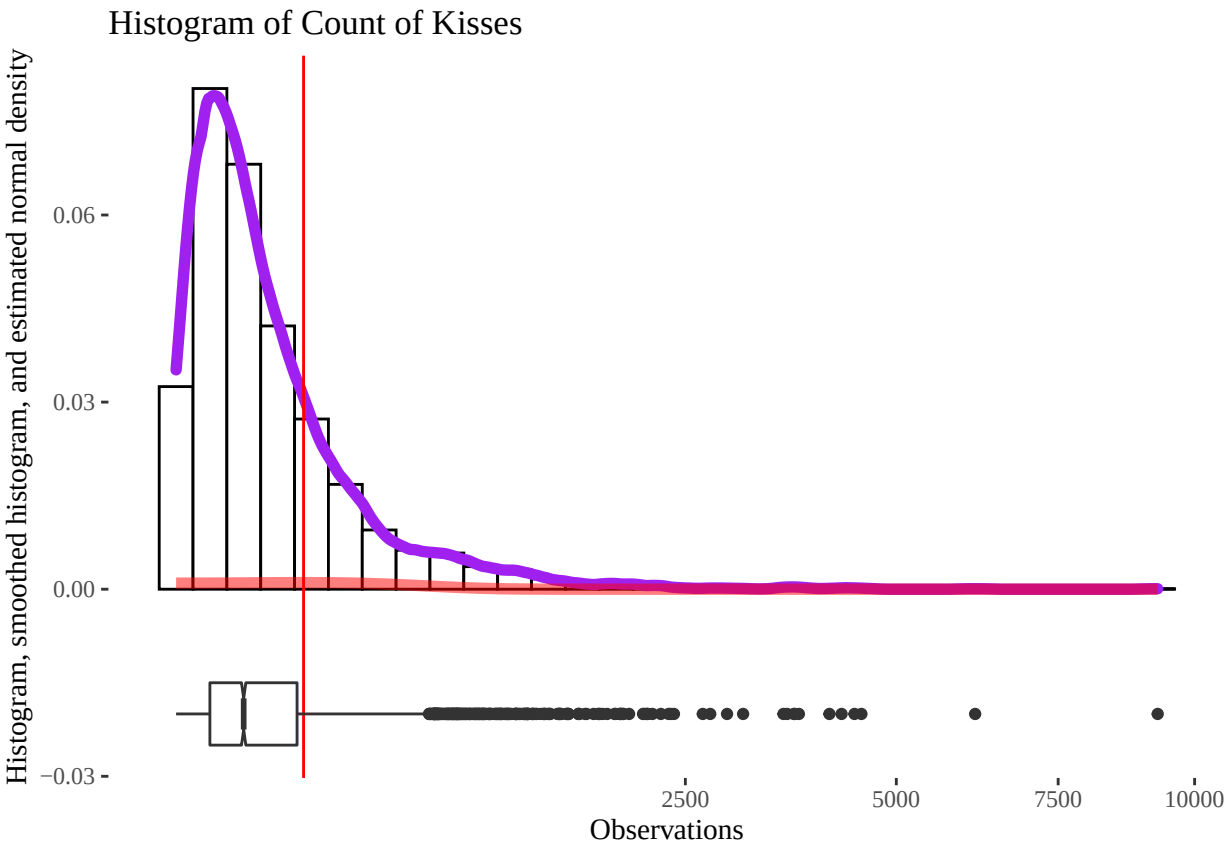


Figure 16: Square-root transformed histogram of counts_kisses

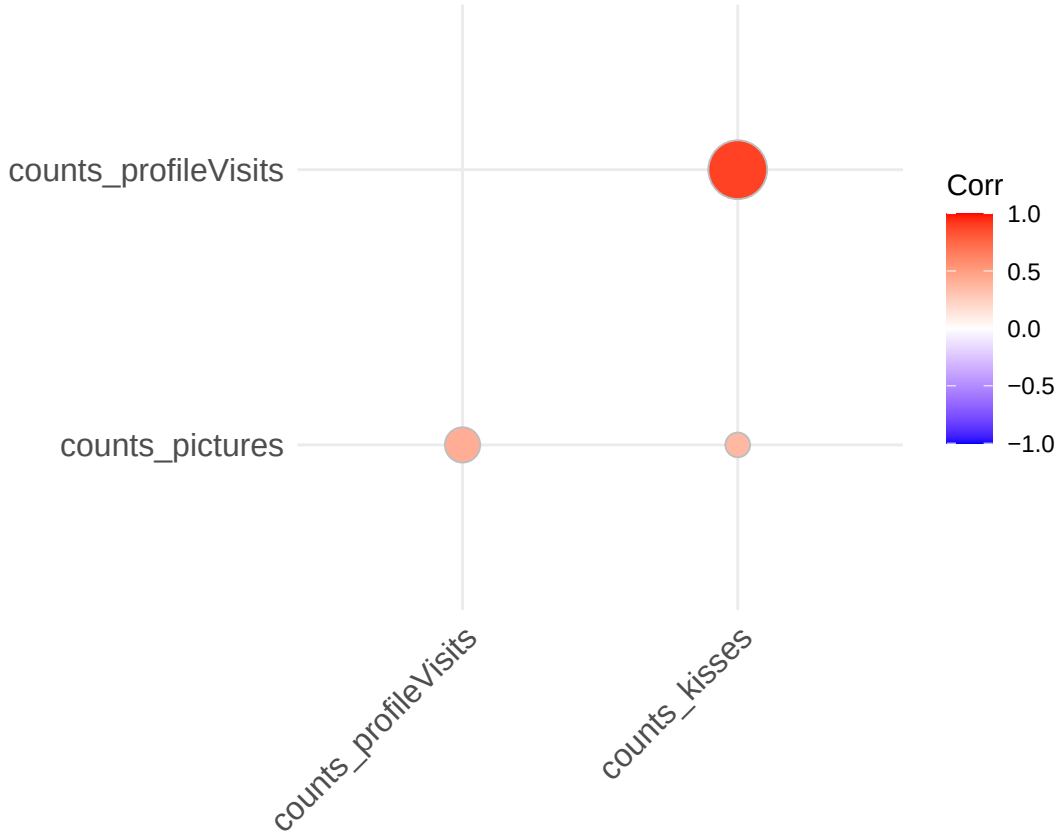


Figure 17: Correlation plot for counts pictures, profile visits, and kisses visualised

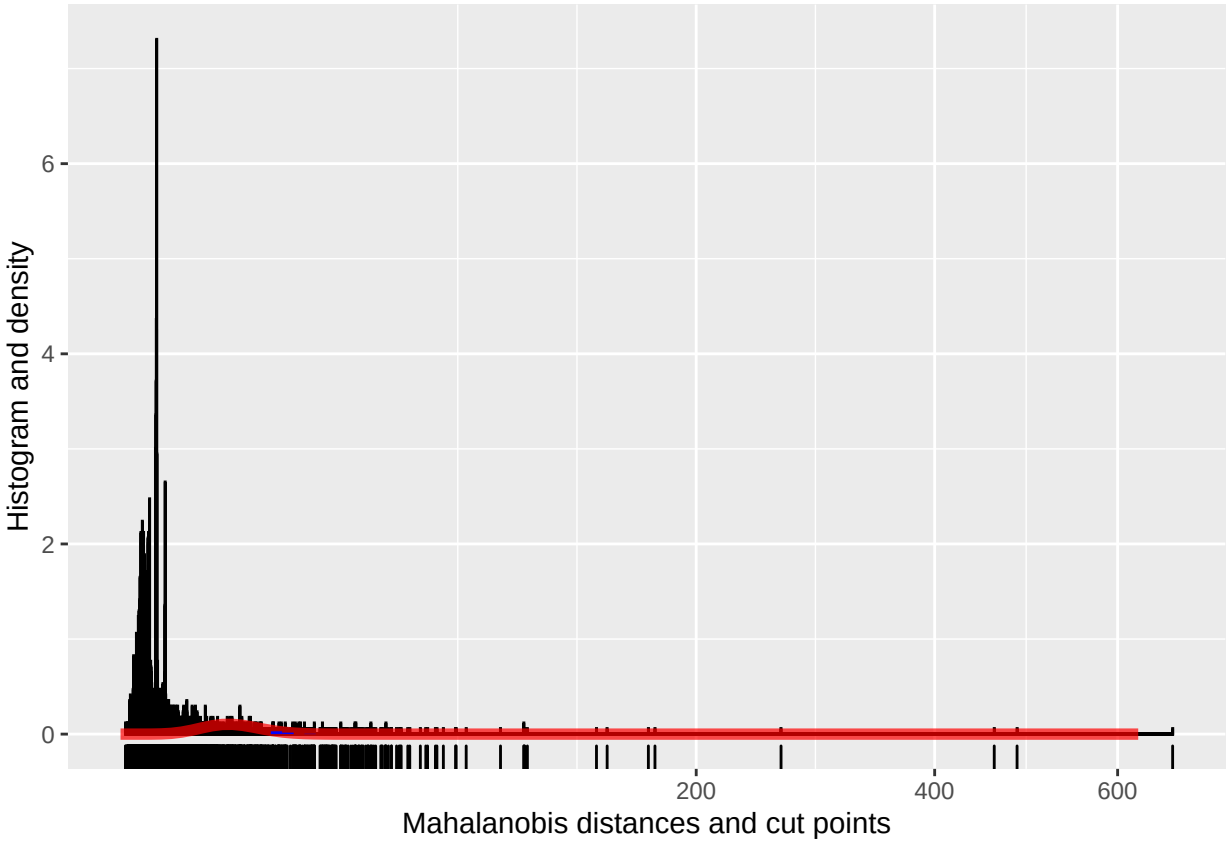


Figure 18: Mahalanobis distince plotted

```
## \begin{table}[ht]
## \centering
## \caption{Surprising table}
## \label{surprise}
## \begin{tabular}{rr}
##   \hline
##  & V1 \\
##   \hline
## Typical & 3832 \\
##   Somewhat &  23 \\
##   Surprising &  29 \\
##   Very & 108 \\
##    \hline
## \end{tabular}
## \end{table}
```

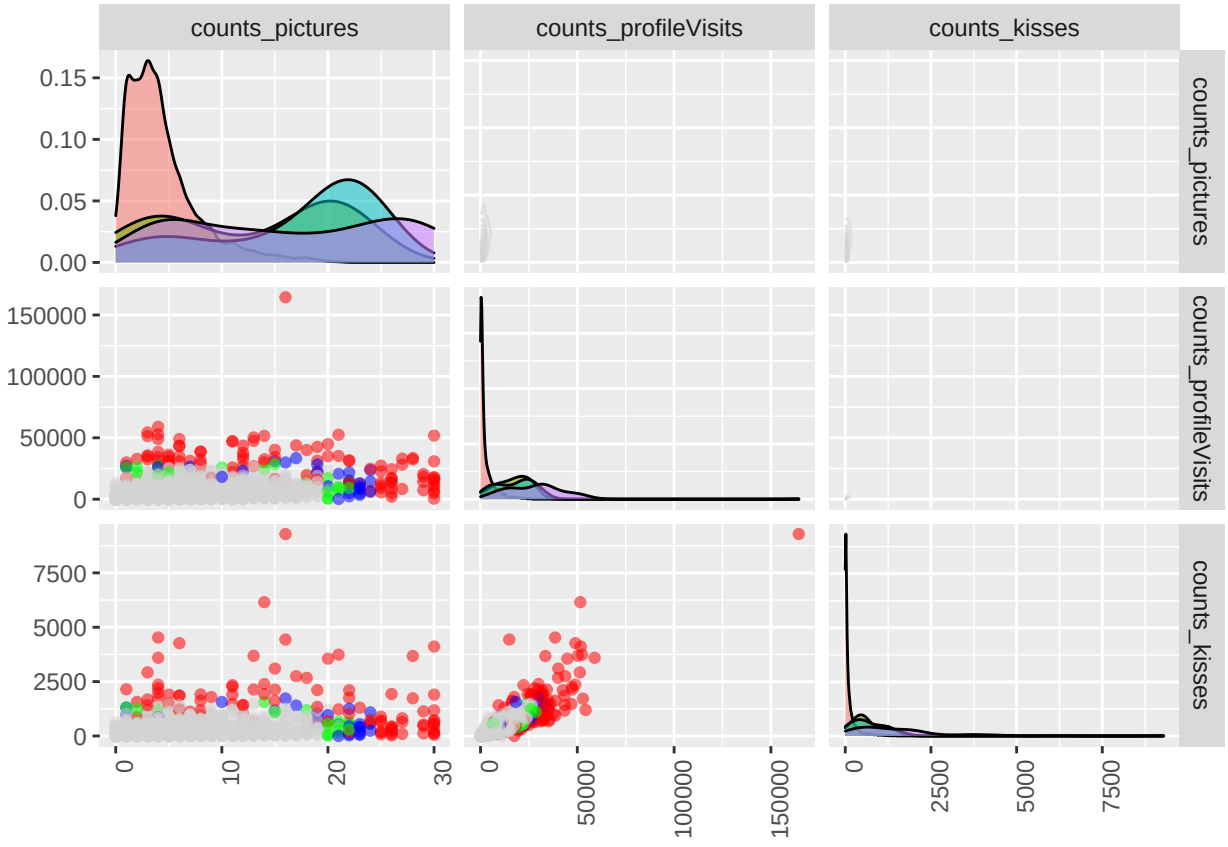


Figure 19: Pairs plot of count pictures, profile visits, and kisses

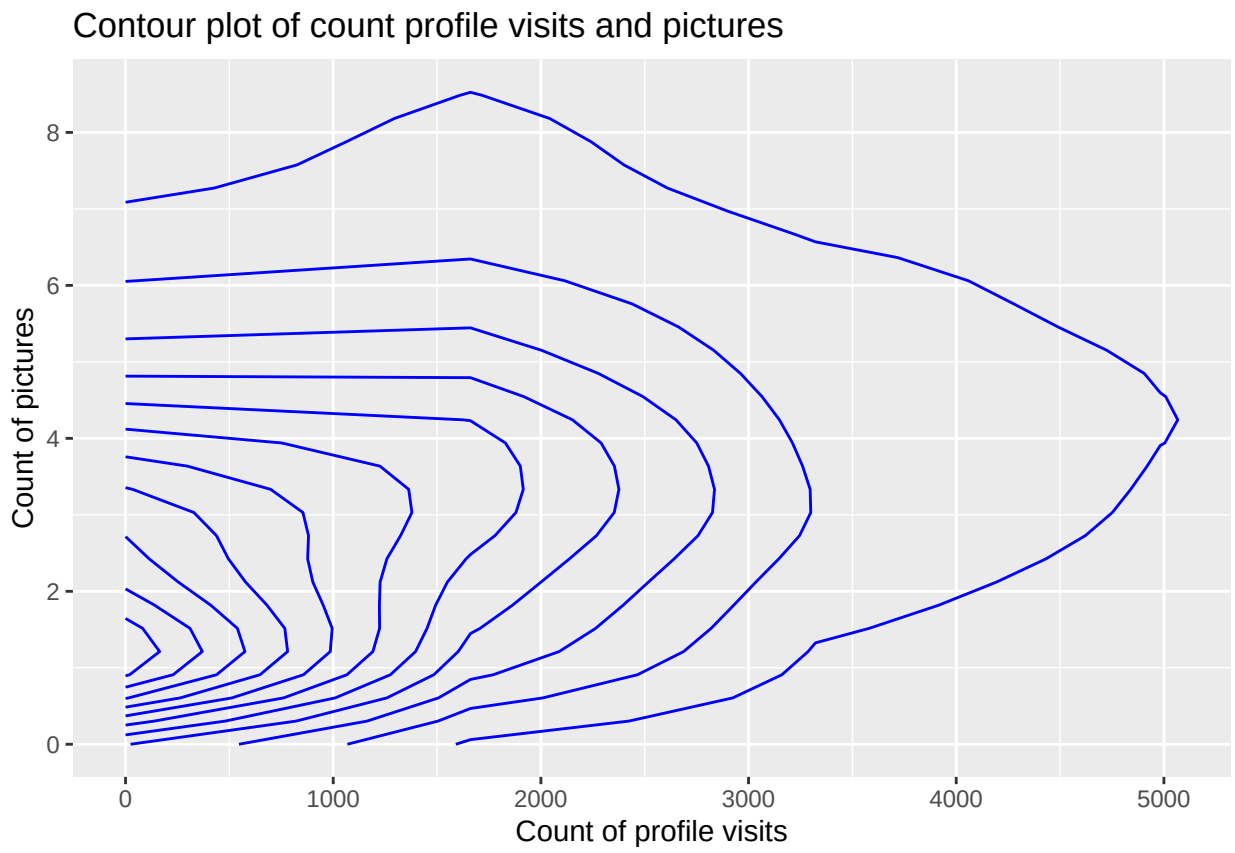


Figure 20: Contour plot of count profile visits and pictures

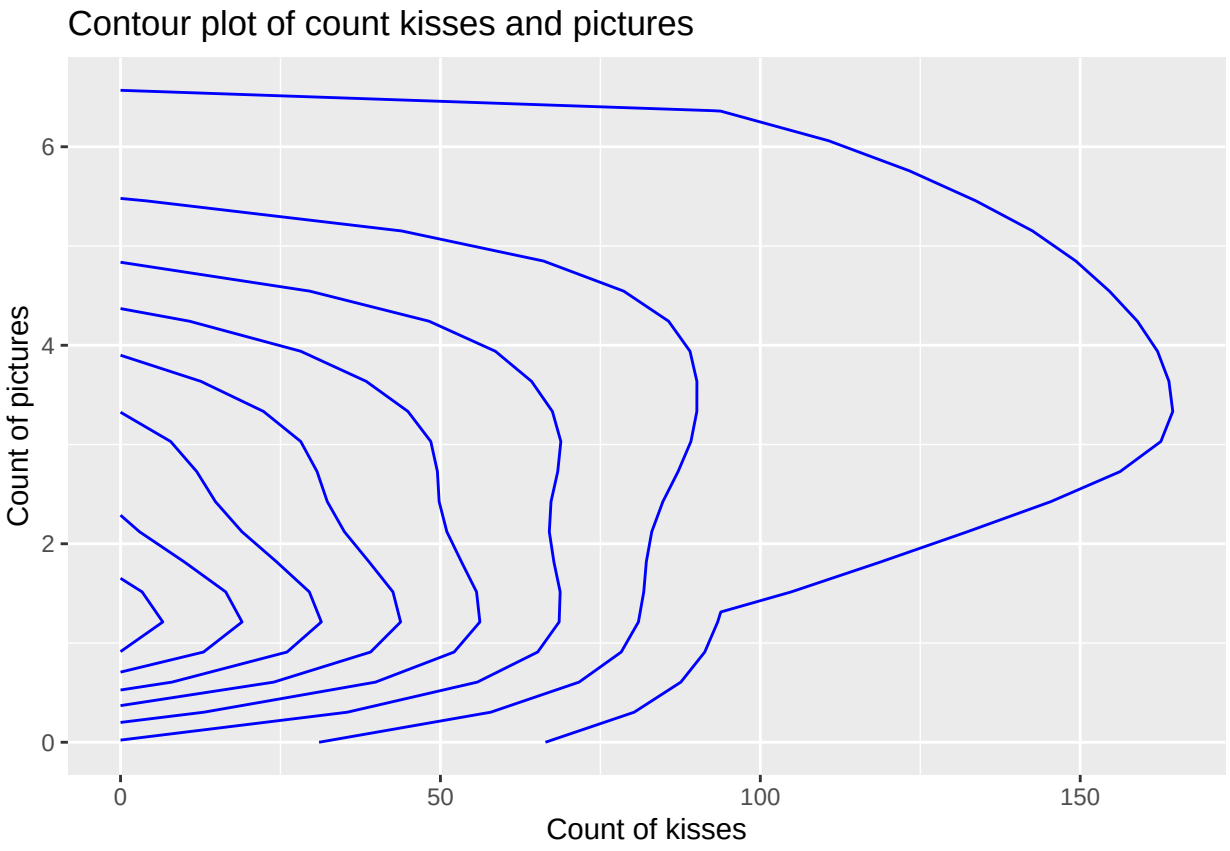


Figure 21: Contour plot of count kisses and pictures

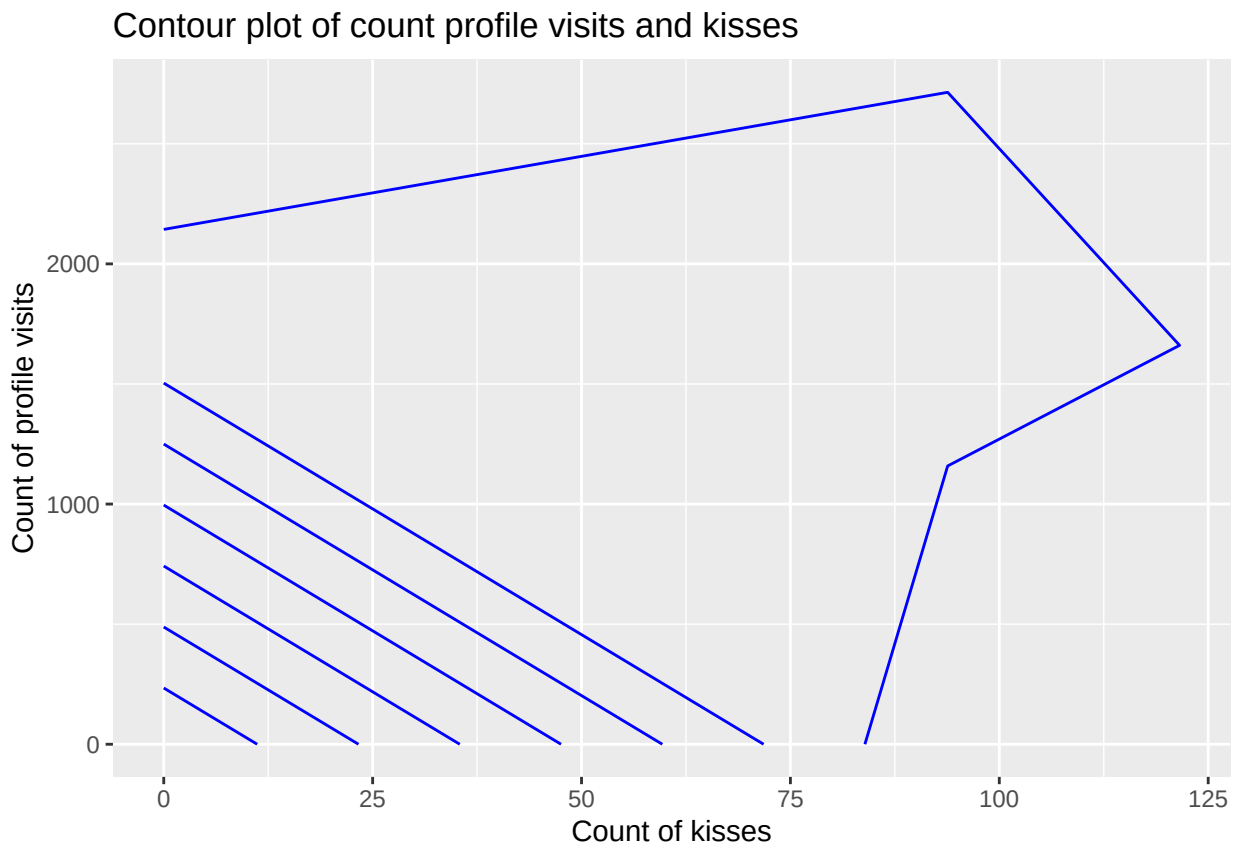


Figure 22: Contour plot of count profile visits and kisses

9 Age, country and number of pictures

Figure 23 shows the number of users by age.

It appears that a large number of countries have a very small number of users, while a few countries such as Switzerland, Germany, France and Italy have larger numbers of users. This can be seen in 9. Table 7 shows us a breakdown of the number of users per country by age. Again this supports higher numbers in Switzerland, Germany, France and Italy. Taking a closer look at the countries - Switzerland, Germany, France and Italy we can look at the number of users per country by age. Figure 24 shows that Switzerland has the larger number of users with ages 19-21 and Germany has the larger number of users for ages 22-24. Looking at the boxplot in figure 25 we can see that there are a few very large outliers in the data and it is very hard to read the boxplots. By taking the log of the data we can see the median more clearly. Figure 26 shows that the medians vary, however there is not much variability within a country across the different ages.

10 Update file path

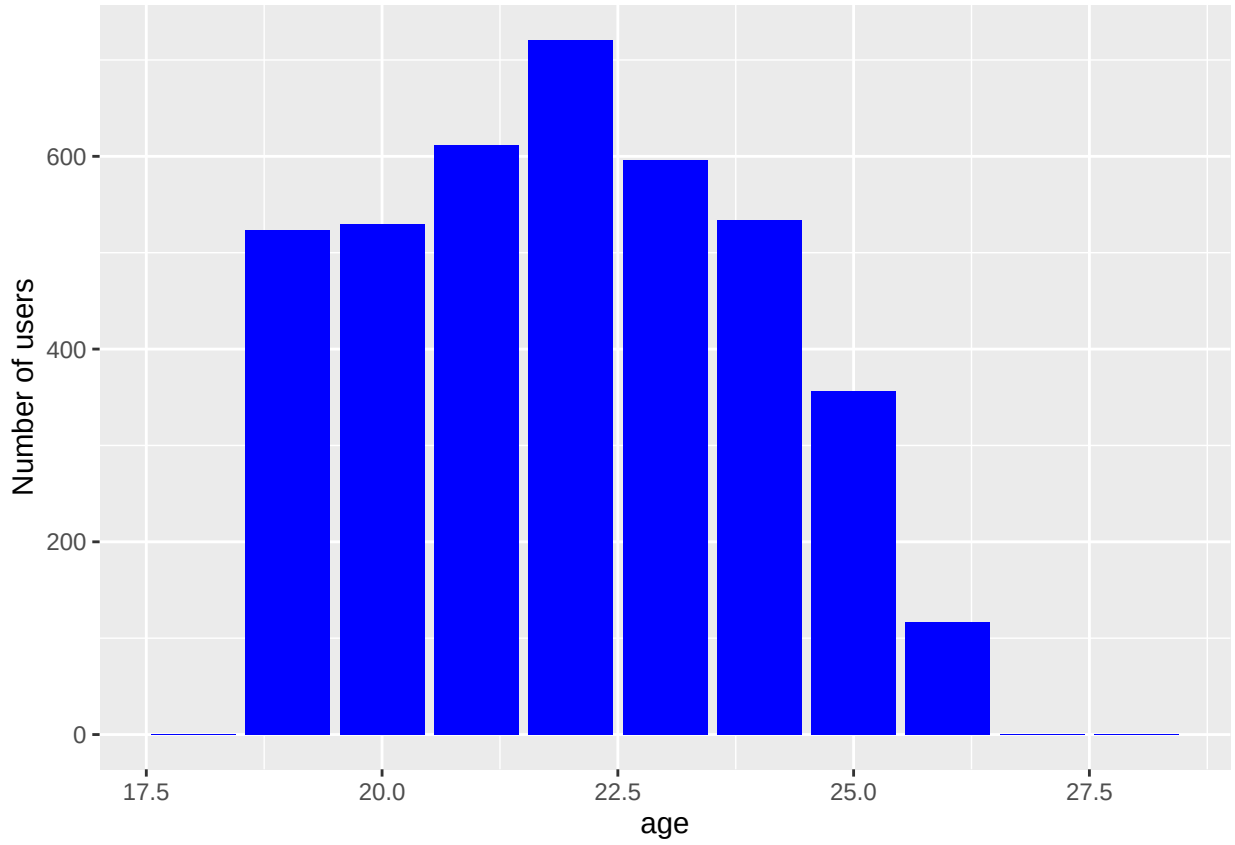


Figure 23: Number of users by age

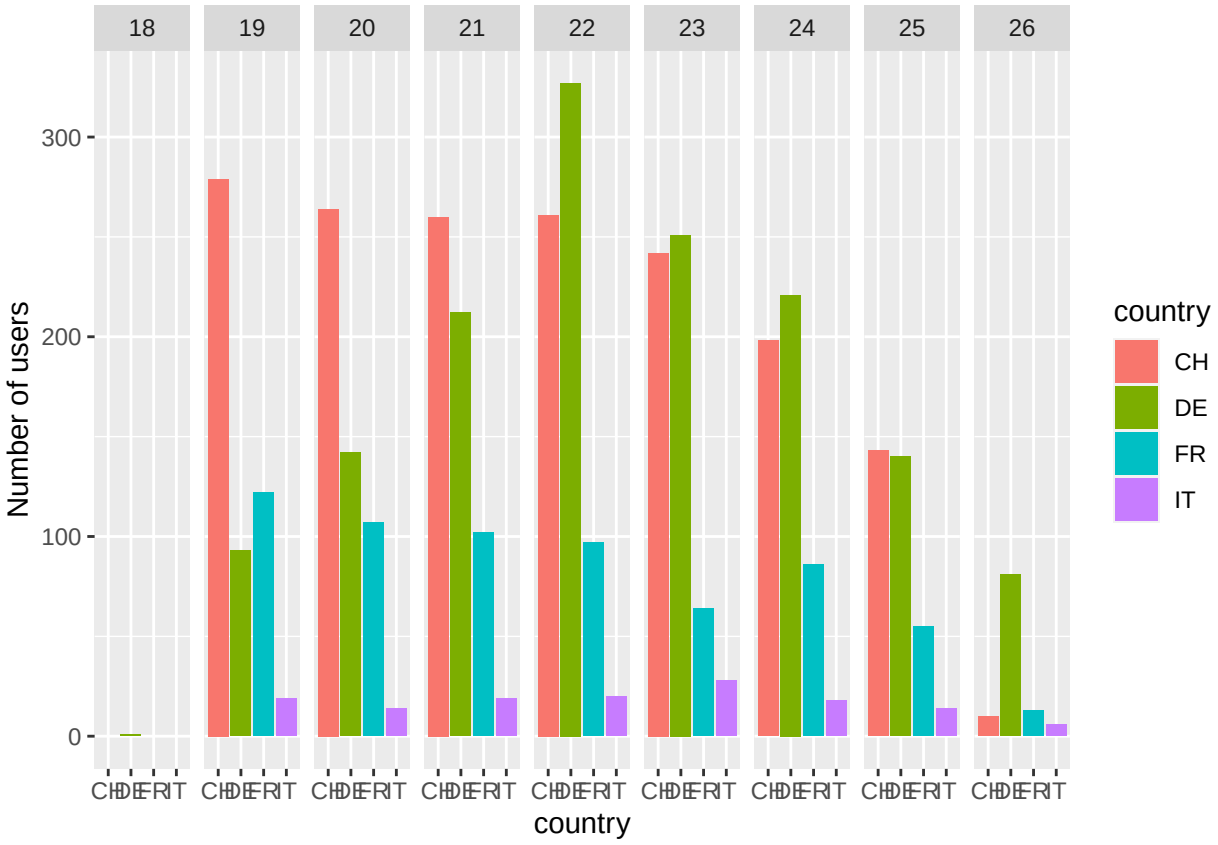


Figure 24: Number of users by age and country

```
## Warning: Ignoring unknown parameters: notched
```

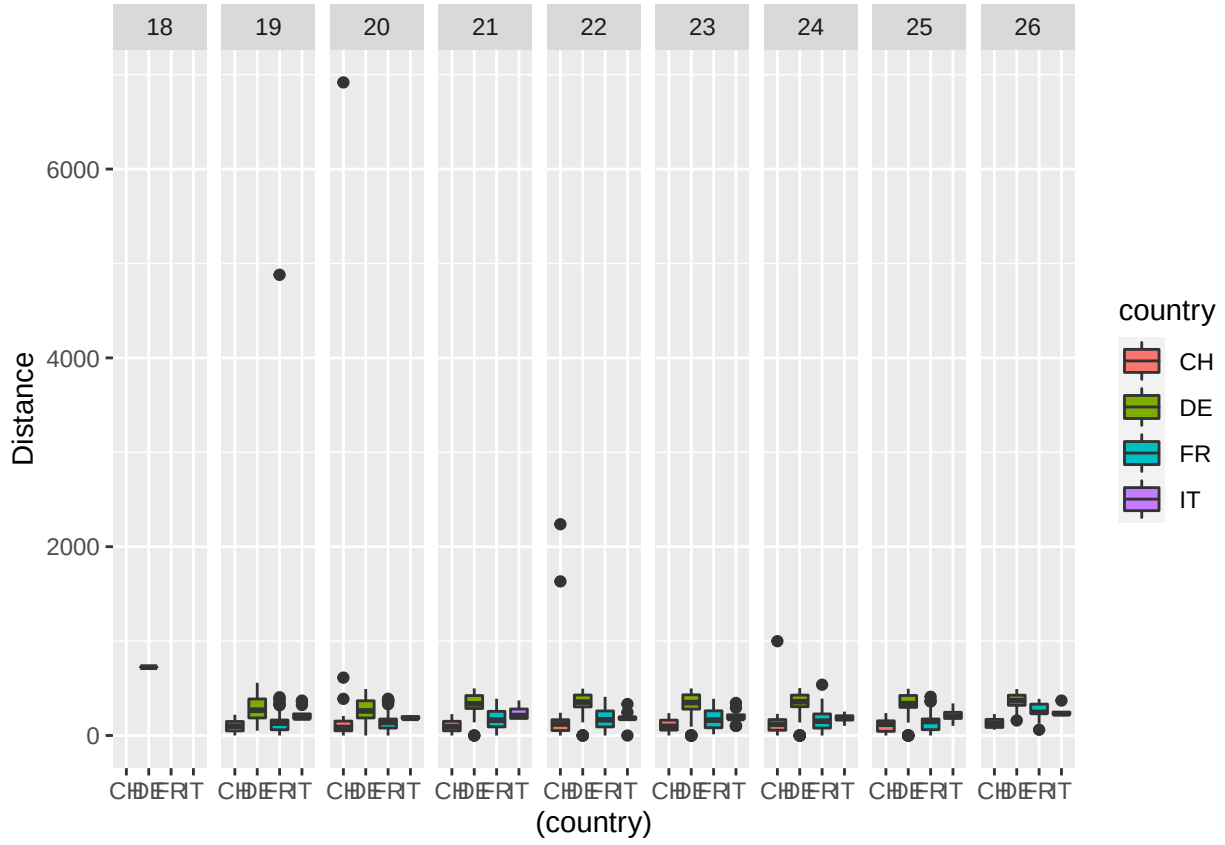


Figure 25: Side by side boxplots of age, country and distance

```
## Warning: Ignoring unknown parameters: notched

## Warning: Transformation introduced infinite values in continuous y-axis

## Warning: Removed 124 rows containing non-finite values (stat_boxplot).
```

11 References

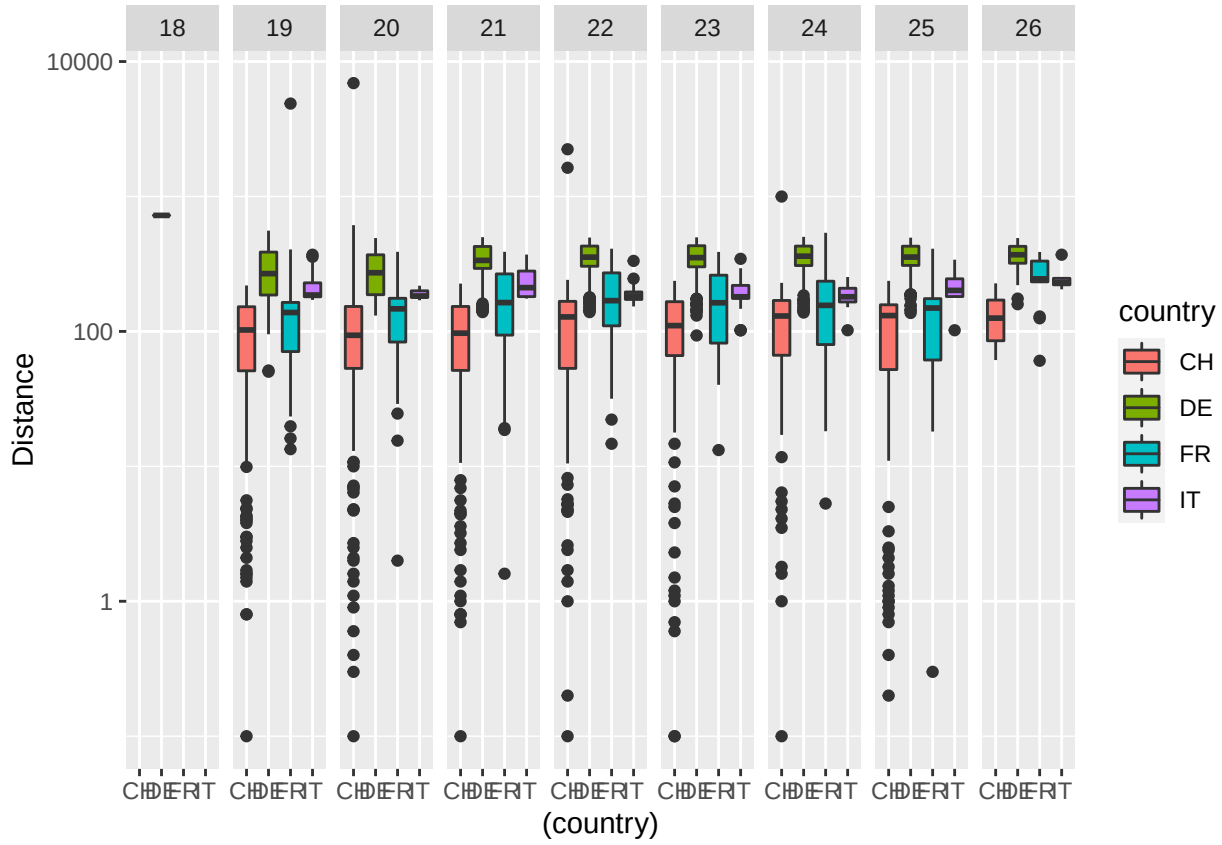


Figure 26: Side by side boxplots of age, country and distance (log scale)

Table 2: Summary Statistics - Countires

	Country	Count
1	AR	1
2	AT	20
3	AU	2
4	BA	3
5	BE	7
6	BR	2
7	CA	2
8	CF	1
9	CH	1657
10	CZ	1
11	DE	1468
12	ES	6
13	ET	1
14	FR	646
15	GB	2
16	HU	1
17	ID	1
18	IN	1
19	IT	138
20	JM	1
21	LI	1
22	LR	1
23	LU	5
24	NL	2
25	PE	1
26	PH	1
27	RO	2
28	RU	2
29	SC	2
30	TR	10
31	UA	1
32	US	3

Table 3: Summary Statistics - isVip

	isVip	Count
No	0	3901
Yes	1	91

Table 4: Summary Statistics - isVip (or not) by country

	isVip - No	isVip - No
AR	1	0
AT	20	0
AU	2	0
BA	3	0
BE	7	0
BR	1	1
CA	2	0
CF	1	0
CH	1611	46
CZ	1	0
DE	1443	25
ES	6	0
ET	1	0
FR	631	15
GB	2	0
HU	1	0
ID	1	0
IN	1	0
IT	134	4
JM	1	0
LI	1	0
LR	1	0
LU	5	0
NL	2	0
PE	1	0
PH	1	0
RO	2	0
RU	2	0
SC	2	0
TR	10	0
UA	1	0
US	3	0

Table 5: Covariance matrix

	counts_pictures	counts_profileVisits	counts_kisses
counts_pictures	19.54	12647.04	614.40
counts_profileVisits	12647.04	46854549.74	2289351.68
counts_kisses	614.40	2289351.68	142620.04

Table 6: Correlation matrix

	counts_pictures	counts_profileVisits	counts_kisses
counts_pictures	1.00	0.42	0.37
counts_profileVisits	0.42	1.00	0.89
counts_kisses	0.37	0.89	1.00

Table 7: Country vs age

	18	19	20	21	22	23	24	25	26	27	28
AR	0	0	0	0	0	0	1	0	0	0	0
AT	0	1	1	2	4	4	1	0	6	1	0
AU	0	1	0	0	1	0	0	0	0	0	0
BA	0	0	0	0	2	0	0	1	0	0	0
BE	0	1	0	0	3	0	1	1	1	0	0
BR	0	0	0	1	0	0	1	0	0	0	0
CA	0	0	1	1	0	0	0	0	0	0	0
CF	0	0	0	0	0	0	0	1	0	0	0
CH	0	279	264	260	261	242	198	143	10	0	0
CZ	0	0	0	0	1	0	0	0	0	0	0
DE	1	93	142	212	327	251	221	140	81	0	0
ES	0	0	1	1	2	1	1	0	0	0	0
ET	0	1	0	0	0	0	0	0	0	0	0
FR	0	122	107	102	97	64	86	55	13	0	0
GB	0	1	0	0	0	0	0	0	0	0	1
HU	0	0	0	1	0	0	0	0	0	0	0
ID	0	0	0	0	0	1	0	0	0	0	0
IN	0	0	0	1	0	0	0	0	0	0	0
IT	0	19	14	19	20	28	18	14	6	0	0
JM	0	1	0	0	0	0	0	0	0	0	0
LI	0	0	0	0	1	0	0	0	0	0	0
LR	0	1	0	0	0	0	0	0	0	0	0
LU	0	0	0	0	1	2	1	1	0	0	0
NL	0	0	0	0	0	2	0	0	0	0	0
PE	0	0	0	1	0	0	0	0	0	0	0
PH	0	1	0	0	0	0	0	0	0	0	0
RO	0	0	0	0	0	1	1	0	0	0	0
RU	0	0	0	1	0	0	1	0	0	0	0
SC	0	1	0	0	0	0	1	0	0	0	0
TR	0	0	0	10	0	0	0	0	0	0	0
UA	0	0	0	0	0	0	1	0	0	0	0
US	0	1	0	0	1	0	1	0	0	0	0