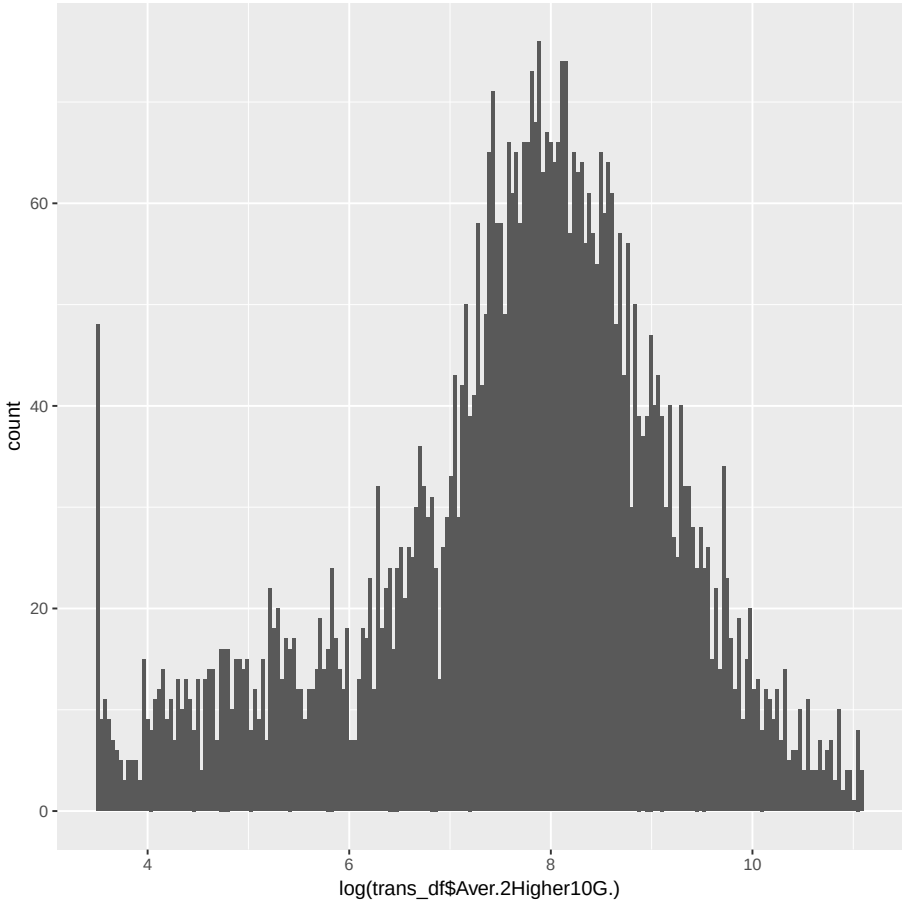
Acetilation

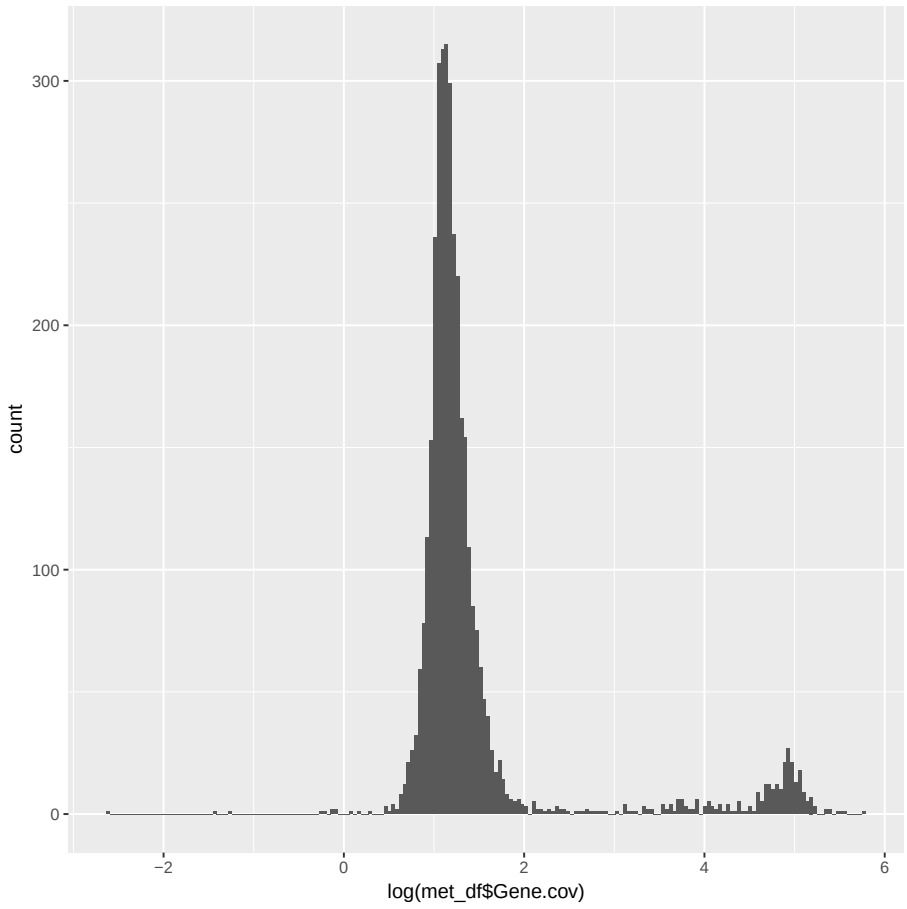
Lucas Michel Todó

January 12, 2018

Contents

```
##
## Attaching package:  'dplyr'
## The following objects are masked from 'package:stats':
##
##    filter, lag
## The following objects are masked from 'package:base':
##
##    intersect, setdiff, setequal, union
## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis
## Loading required package:  XLConnectJars
## XLConnect 0.2-13 by Mirai Solutions GmbH [aut],
##  Martin Studer [cre],
##  The Apache Software Foundation [ctb, cph] (Apache POI),
##  Graph Builder [ctb, cph] (Curvesapi Java library)
## http://www.mirai-solutions.com ,
## http://miraisolutions.wordpress.com
```

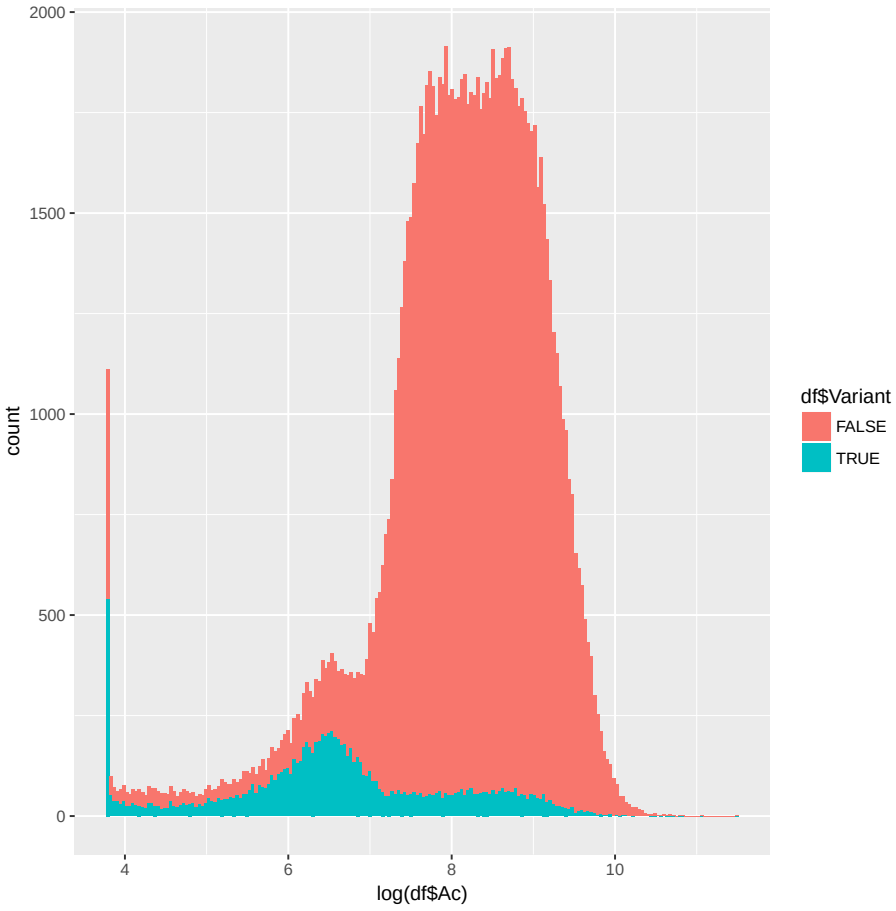

count
60
40
20
0
4
6
8
10
log(trans_df$Aver.2Higher10G.)



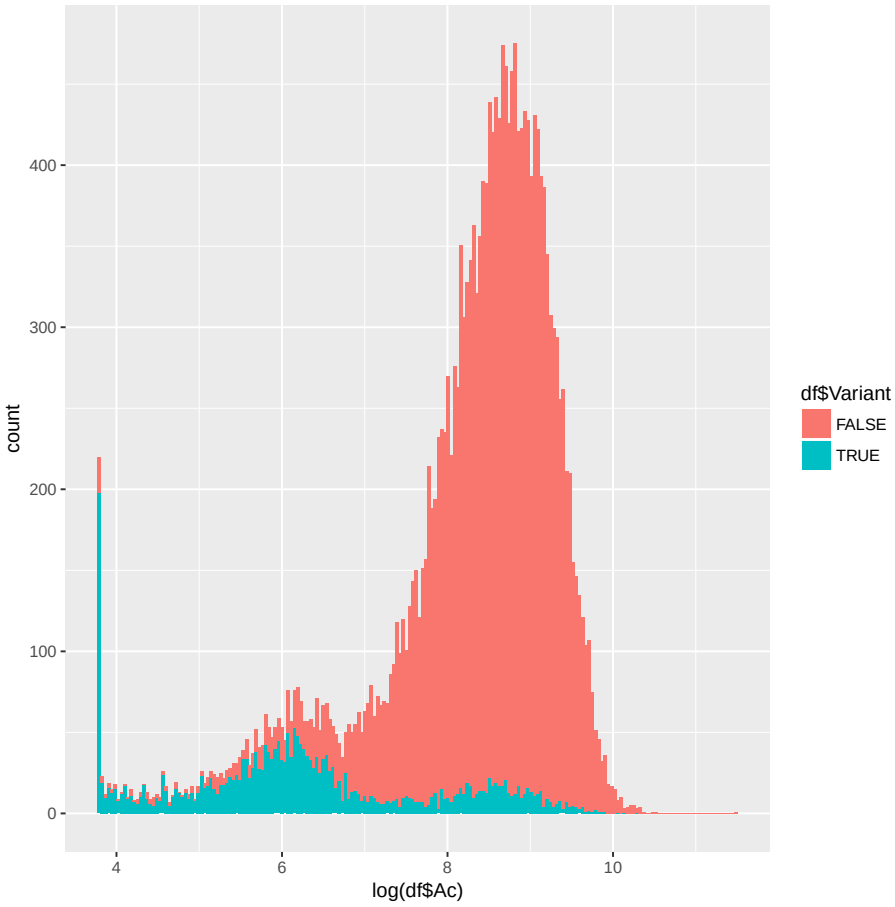

```
##
## FALSE  TRUE
##   214    74
##
## FALSE  TRUE
##    61    74
```

1 Density Plots

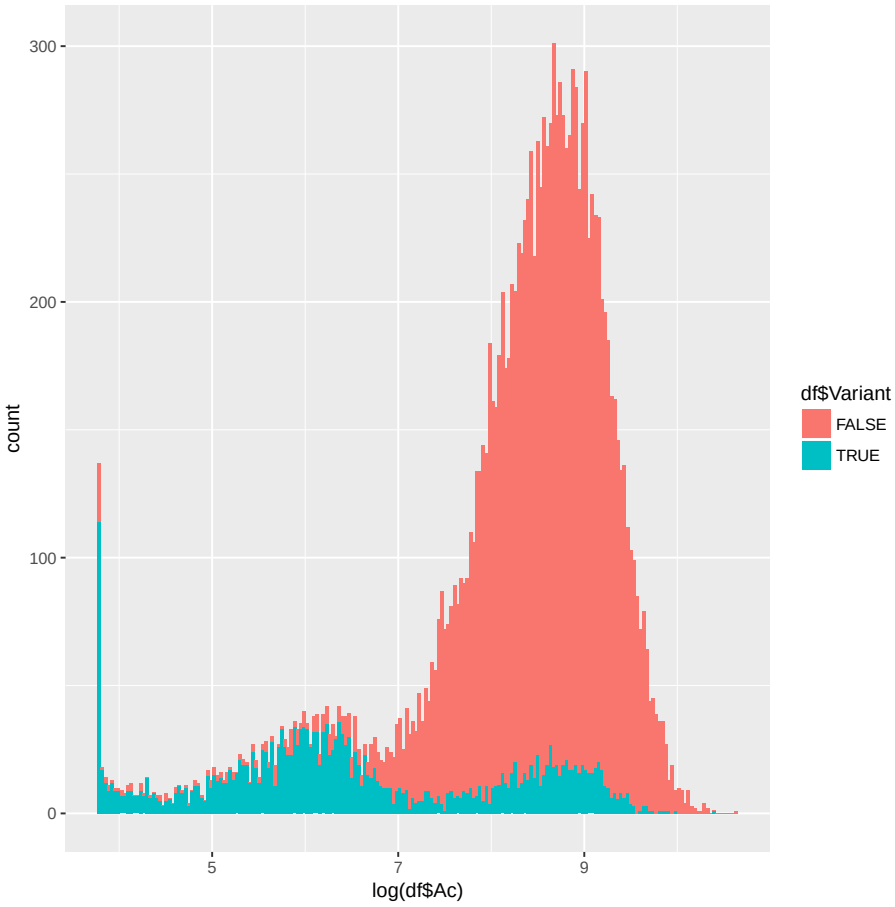
1.1 log(Ac) All


2000
1500
1000
500
0
count
4
6
8
10
log(df$Ac)
df$Variant
FALSE
TRUE

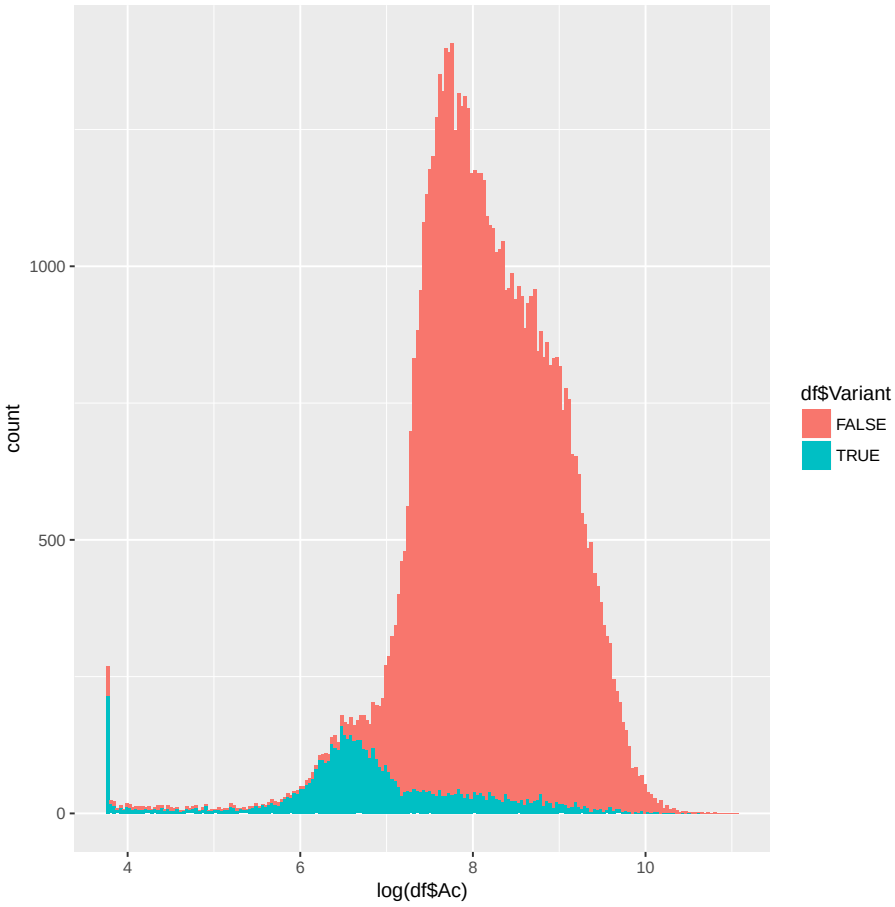
1.2 log(Ac) 5'



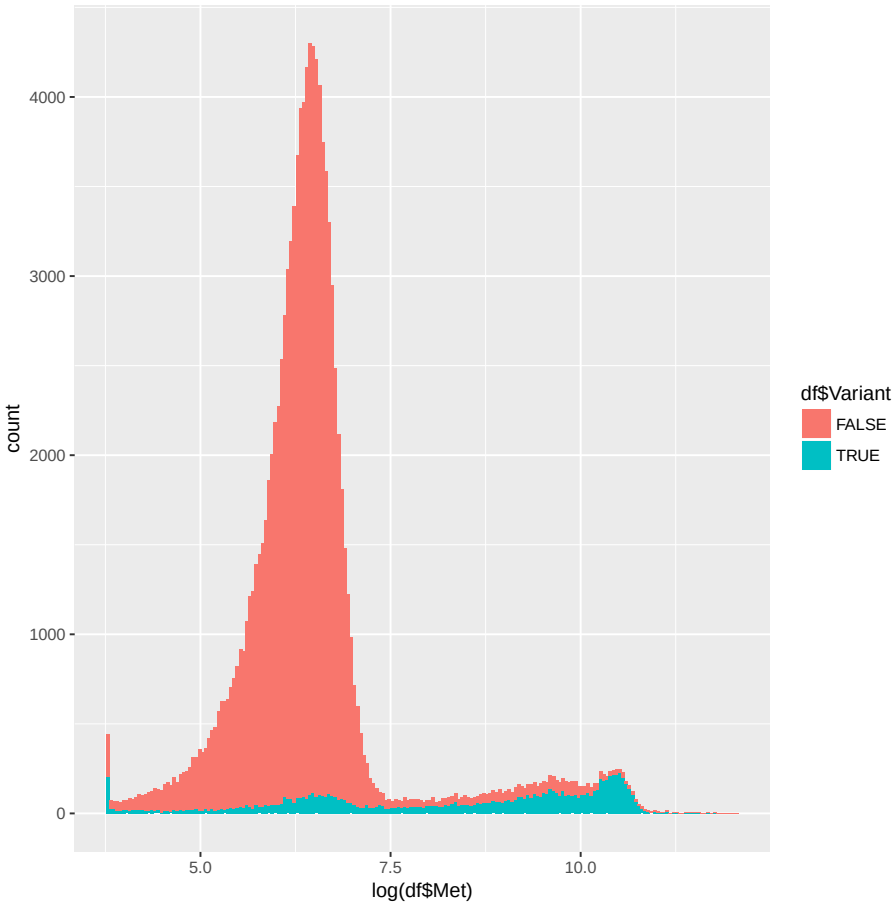
1.3 log(Ac) 3'



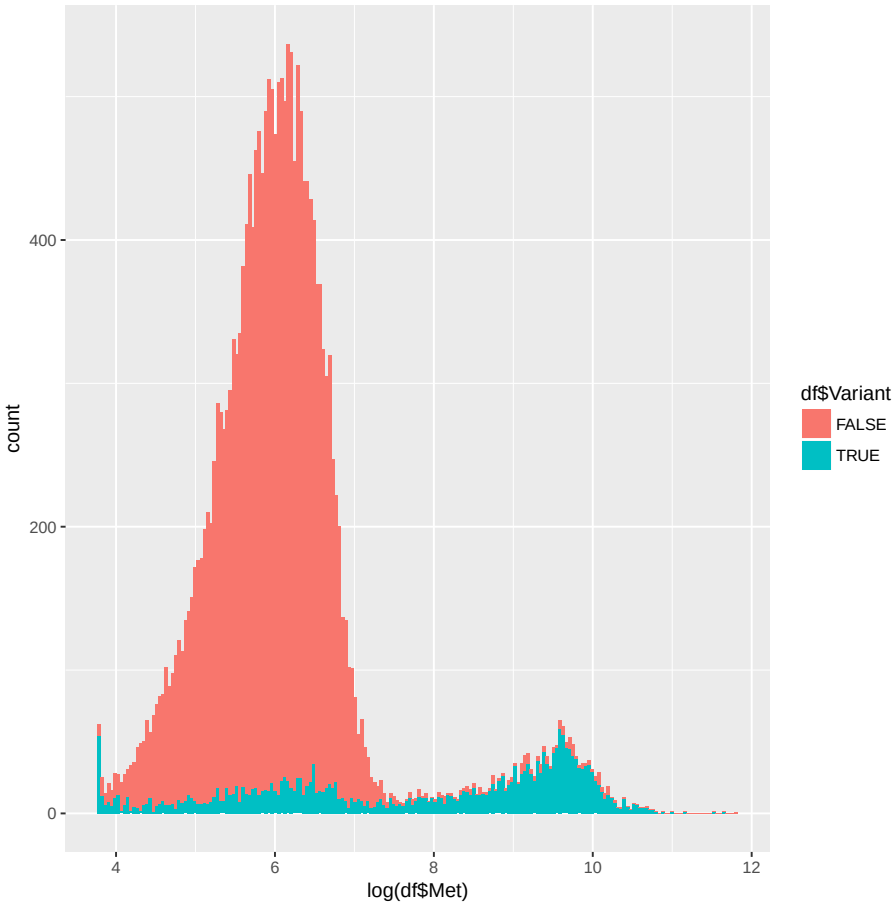
1.4 log(Ac) ORF



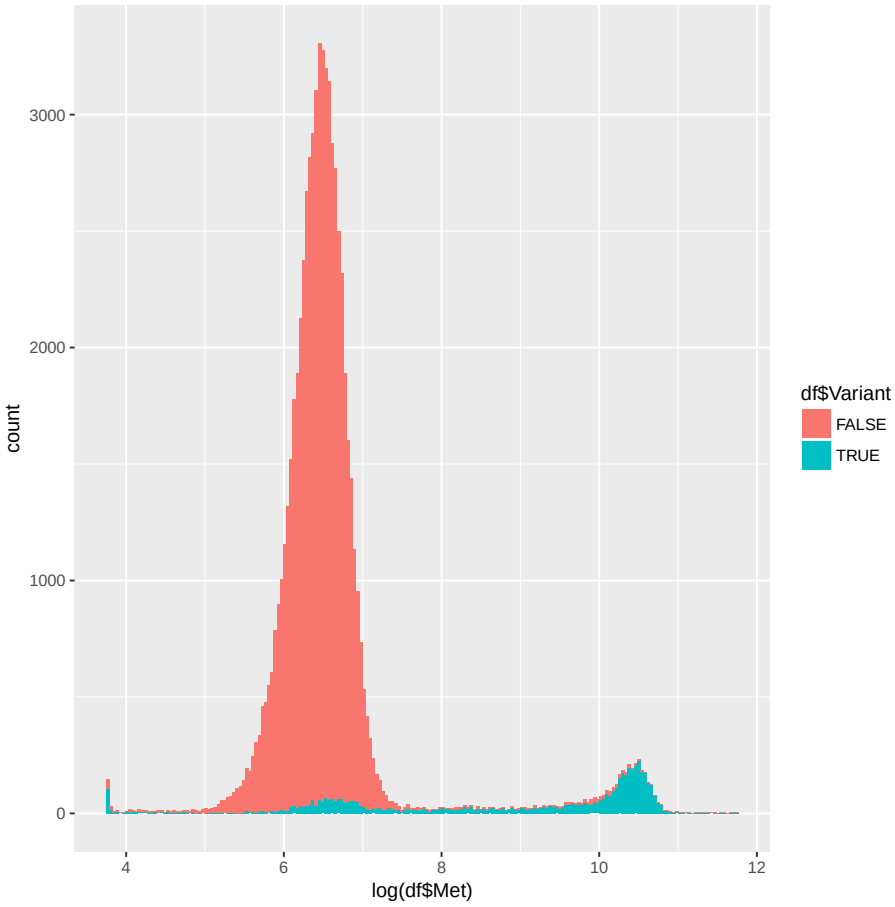
1.5 log(Met) All



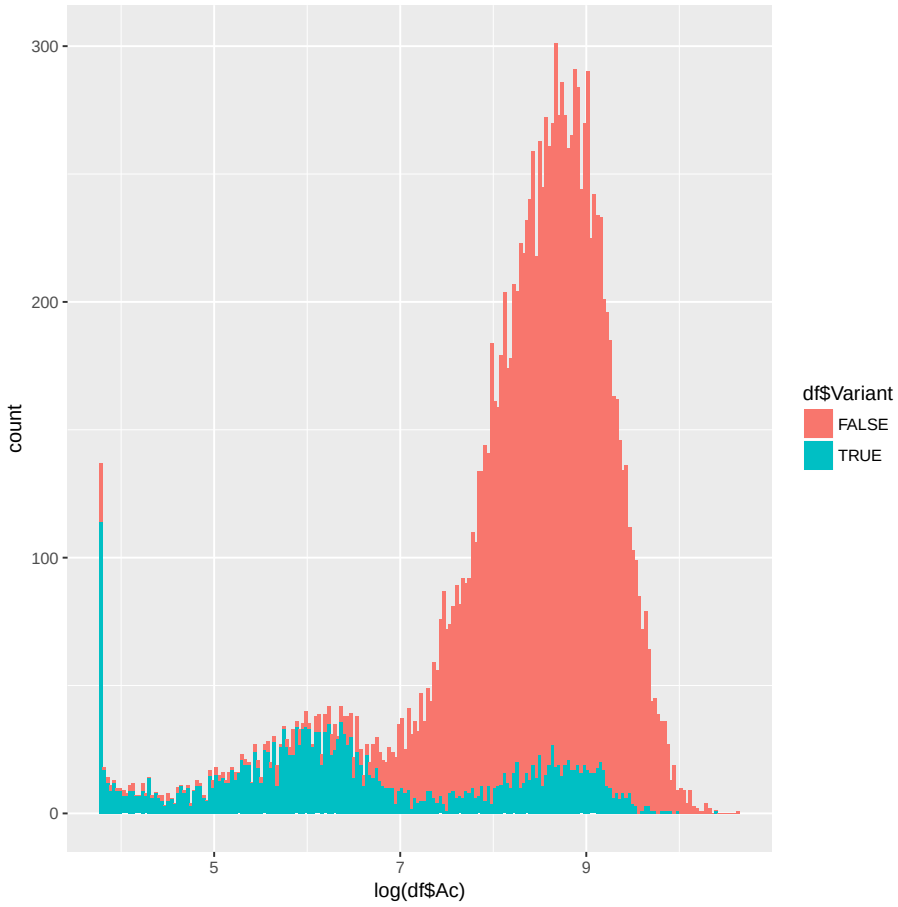
1.6 log(Met) 5'



1.7 log(Met) ORF

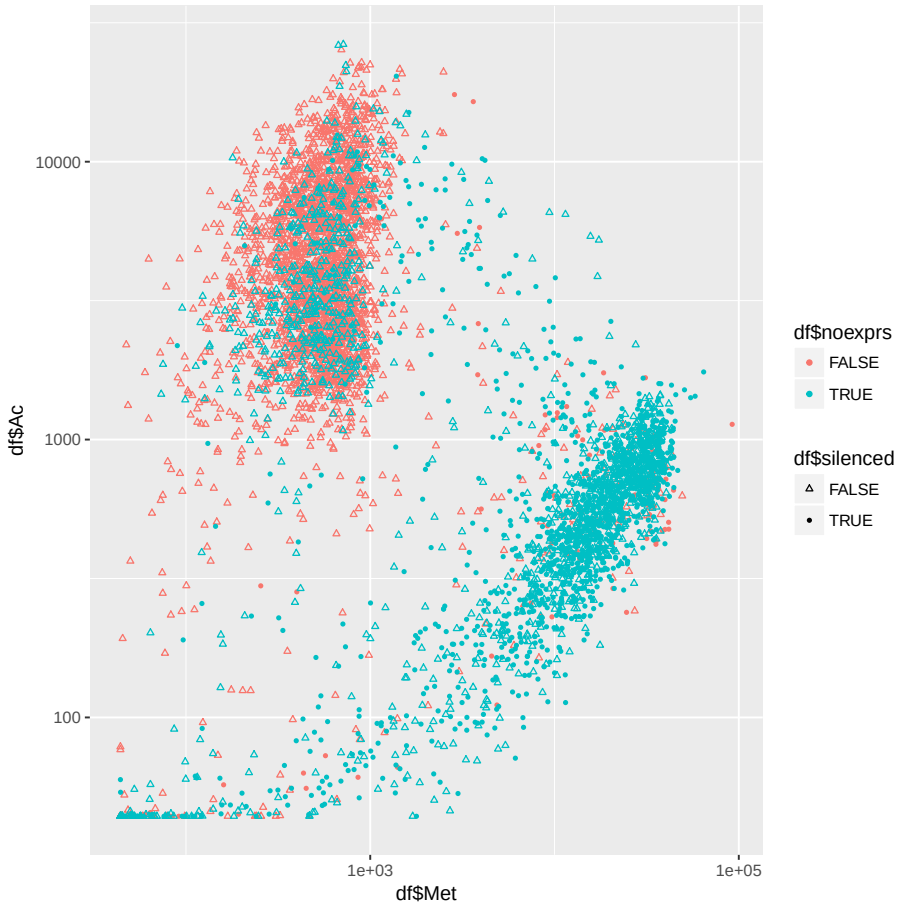


1.8 log(Ac) 3'




```
##
##          Regular   Variant-Active Variant-Silenced
##           105691             5530             5436
```

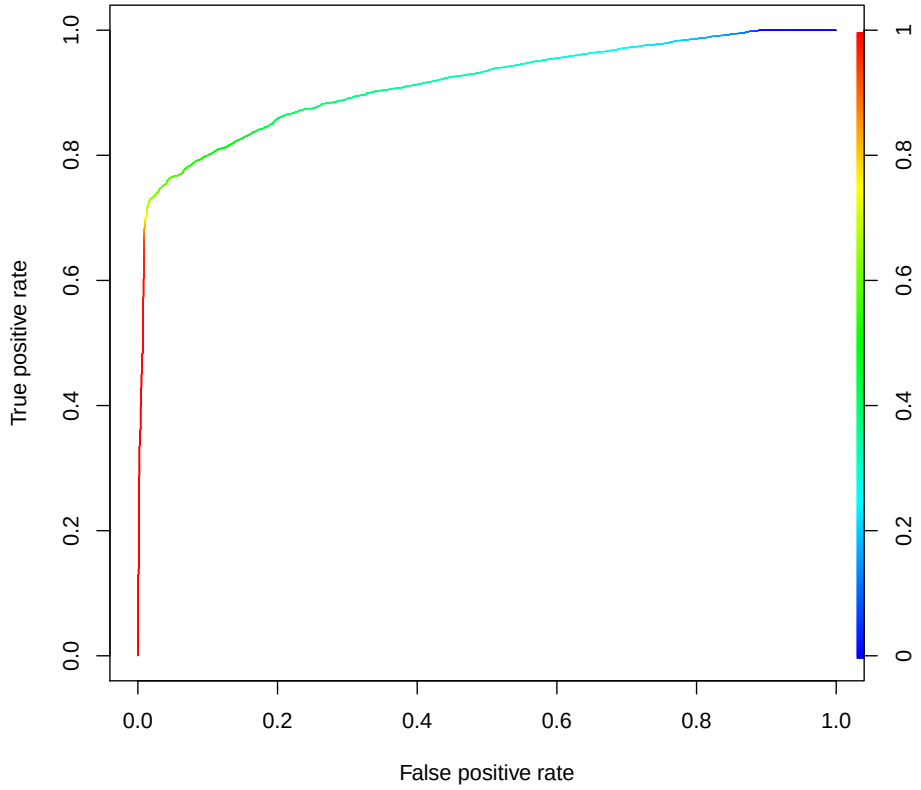

```
##
##   FALSE    TRUE
## 114212    2445
```




```
##
##  FALSE   TRUE
## 105691  10966
## Analysis of Deviance Table
##
## Model 1: Variant ~ Ac + Met + Type + Start + Stop + silenced + noexprs
## Model 2: Variant ~ Ac + Met + Type + Start + Stop
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      8477     5767.9
## 2      8479     6402.1 -2  -634.24 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Call:
## glm(formula = Variant ~ Ac + Met + Type + Start + Stop + silenced +
##     noexprs, family = binomial(link = "logit"), data = train_df)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.4070  -0.7659   0.0120   0.2261   2.8161
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   8.719e-01  9.888e-02   8.818  < 2e-16 ***
## Ac           -1.157e-04  9.326e-06 -12.409  < 2e-16 ***
## Met           2.663e-04  1.994e-05  13.352  < 2e-16 ***
## Type5prima   -4.064e-01  9.780e-02  -4.156 3.24e-05 ***
## TypeORF      -1.177e+00  8.561e-02 -13.750  < 2e-16 ***
## Typeother    -2.780e+01  1.516e+02  -0.183    0.854
## Start        -2.517e-07  4.194e-08  -6.002 1.94e-09 ***
## Stop                 NA         NA      NA       NA
## silencedTRUE  3.758e+00  2.672e-01  14.064  < 2e-16 ***
## noexprsTRUE   9.679e-01  2.330e-01   4.154 3.27e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 11127.3  on 8485  degrees of freedom
## Residual deviance:  5767.9  on 8477  degrees of freedom
## AIC: 5785.9
##
## Number of Fisher Scoring iterations: 16
##
##         FALSE TRUE
##   FALSE  2910  367
##   TRUE   1088 4480
## [1] "Accuracy 0.835500282645562"
## [1] "Accuracy of null model 0.502769926512154"

## Loading required package:  gplots
##
## Attaching package:  'gplots'
## The following object is masked from 'package:stats':
##
##    lowess
```




```
##
## 3prima 5prima    ORF  other
##    227    113     26      1
##
## 3prima 5prima    ORF
##    130    207    751
##           llh       llhNull            G2      McFadden          r2ML
## -2883.9323527 -5563.6381013  5359.4114972     0.4816463     0.4682380
##          r2CU
##     0.6409677
```