Parenthesis Notebook

Set up the initial libraries and dependant functions.

```
#Installation example:
#install.packages('hexbin', repos='http://cran.us.r-project.org')
#Knit doesn't seem to be working in RStudio, R command:
#require("knitr")
#opts_knit$set(root.dir = "/data/anon/SemanticTransformation")
#rmarkdown::render("./RSource/TransformationComparison.Rmd", params = list())

#Note: Either before this or in this R script, remove cases where the
#transformation = original (Only really need to check in the swap)
setwd("/data/anon/SemanticTransformation/")

library(lmerTest)
library(car)
library(sqldf)
library(effsize)
library(GGally)
library(compiler)
library(dplyr)
library(MuMIn) #Mixed effects R squared
library(Hmisc)
library(stargazer)
library(xtable)

source("./RSource/RegressionHelper.R")
source("./RSource/ComparisonTestHelper.R")
source("./RSource/FileLoadHelper.R")
source("./RSource/GitInfoTableFunctions.R")
source("./RSource/TransSummaryFunctions.R")
source("./RSource/TSumCompareFunctions.R")
source("./RSource/ColorBlind.R")
source("./RSource/ResultPrinter.R")
source("./RSource/LMDiffModels.R")
```

Parenthesis adding results

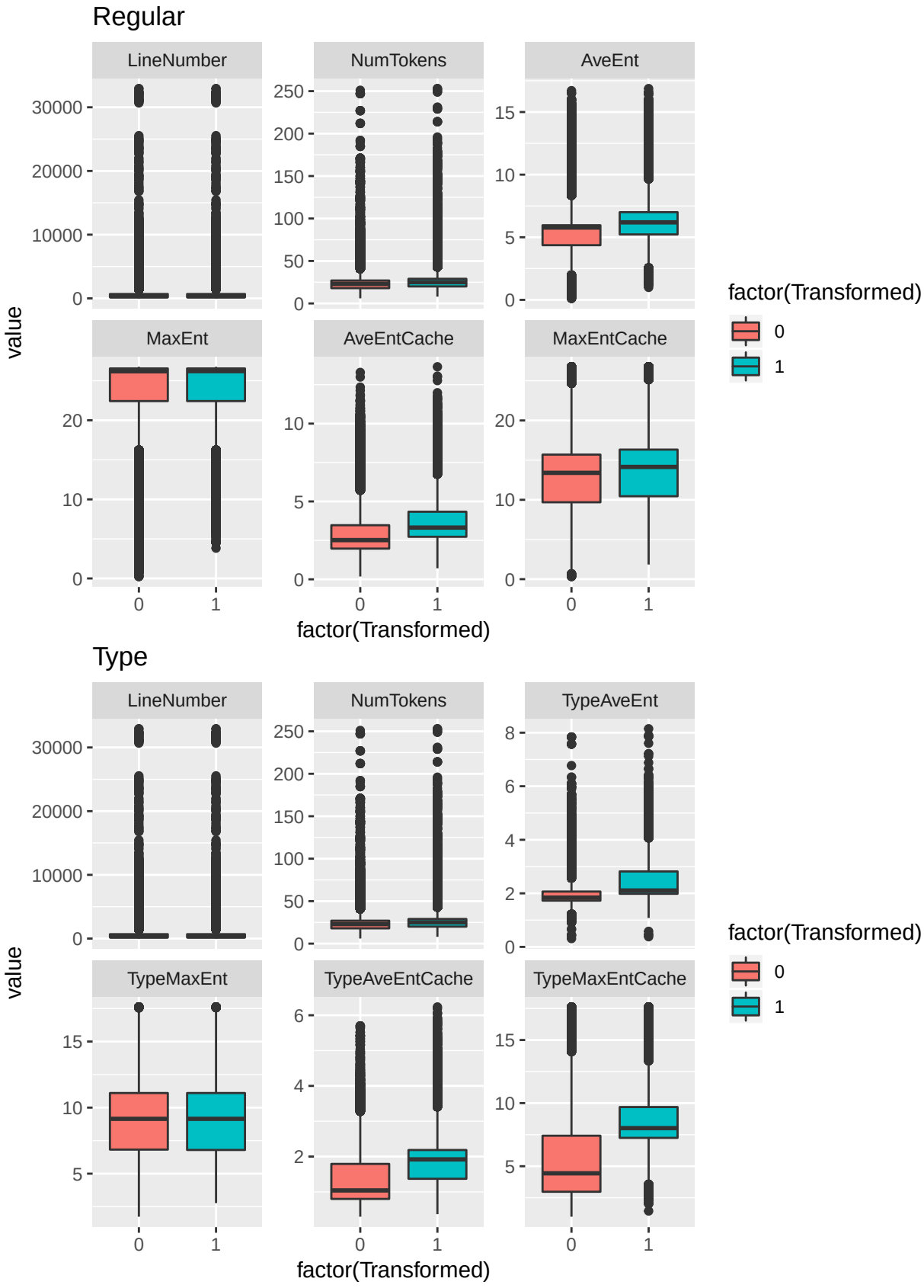
```
setwd("/data/anon/SemanticTransformation/")
#dap <- compareDepthSummary("add_parenthesis.csv", "AddParen", "ADD", FALSE)
dap <- compareDepthSummary("add_parenthesis_topstarred.csv", "AddParenTop", "ADD", FALSE)
```

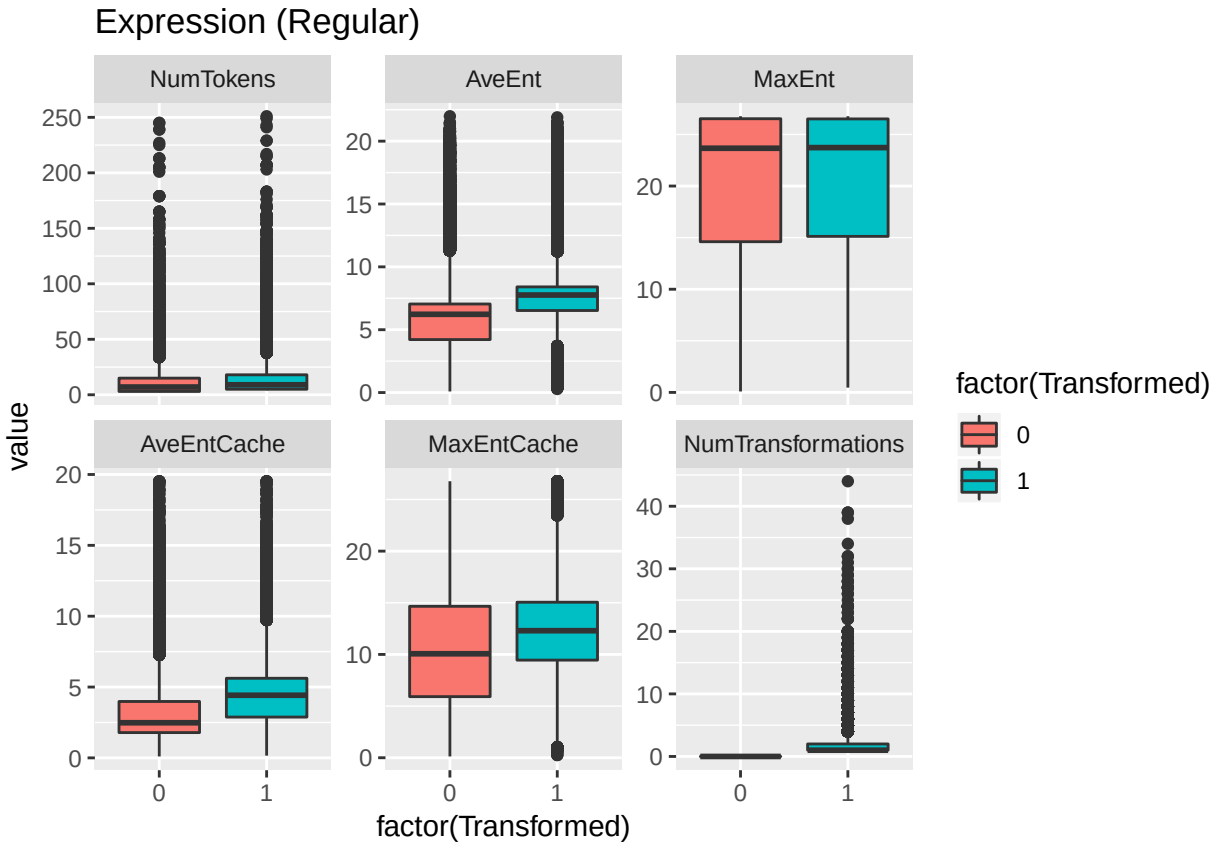
```
## [[1]]
## [1] "5"    "Bool"
```

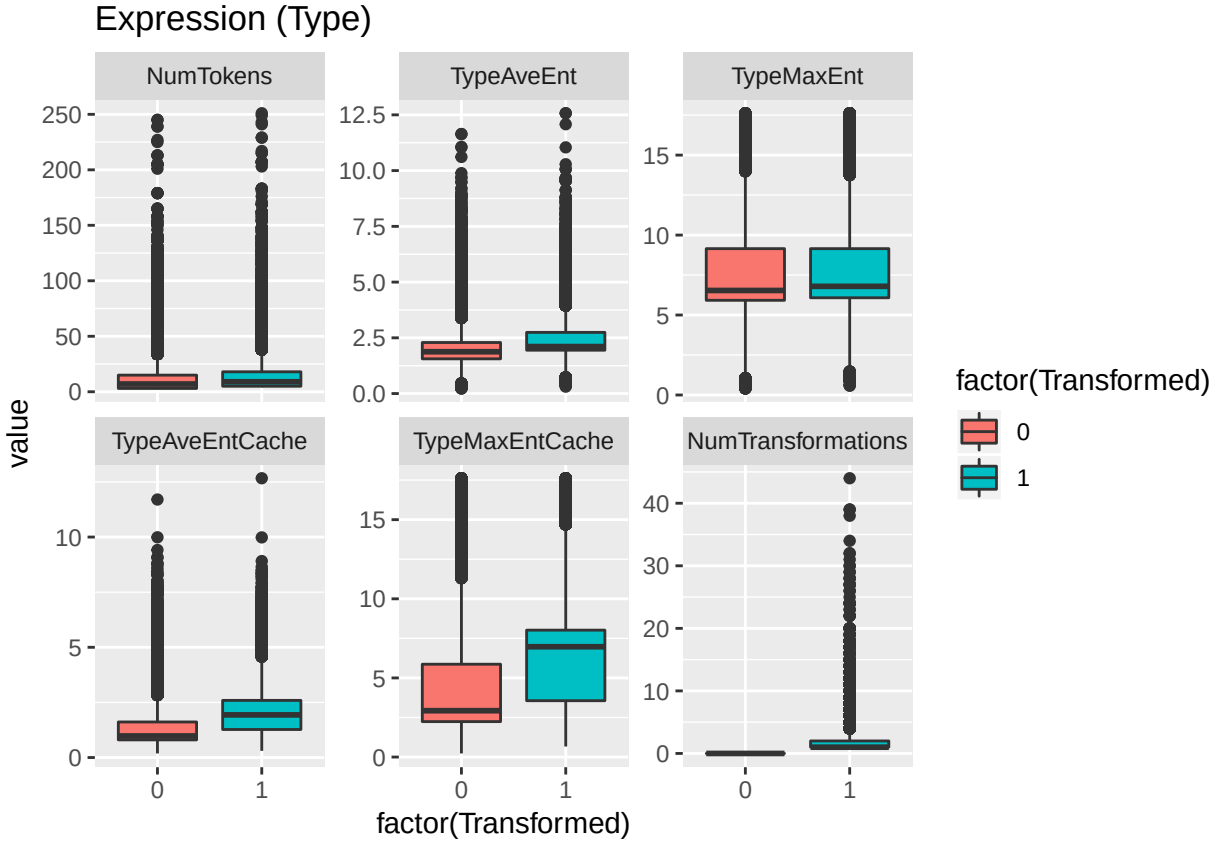
```
## Loading required package: tcltk
```

```
##  [1] "TransId"
##  [2] "Filepath"
##  [3] "LineNumber"
##  [4] "NumTokens"
##  [5] "Transformed"
##  [6] "Source"
##  [7] "CleanLexerNumTokens"
```

```
##  [8] "CleanLexerSource"
##  [9] "AveEnt"
## [10] "MaxEnt"
## [11] "AveEntCache"
## [12] "MaxEntCache"
## [13] "AveEntRev"
## [14] "MaxEntRev"
## [15] "TypeSource"
## [16] "TypeNumTokens"
## [17] "TypeAveEnt"
## [18] "TypeMaxEnt"
## [19] "TypeAveEntCache"
## [20] "TypeMaxEntCache"
## [21] "Depth"
## [22] "Expression"
## [23] "ExpressionNumTokens"
## [24] "ExpressionCleanLexerSource"
## [25] "ExpressionCleanLexerNumTokens"
## [26] "ExpressionForwardAverageEntropy"
## [27] "ExpressionForwardMaxEntropy"
## [28] "ExpressionForwardAverageEntropyCache"
## [29] "ExpressionForwardMaxEntropyCache"
## [30] "TypeExpression"
## [31] "ExpressionTypeNumTokens"
## [32] "ExpressionTypeAverageEntropy"
## [33] "ExpressionTypeMaxEntropy"
## [34] "ExpressionTypeAverageEntropyCache"
## [35] "ExpressionTypeMaxEntropyCache"
## [36] "NumTransformations"
## [37] "ParentOp"
## [38] "MostFreqOp"
## [39] "LeastFreqOp"
## [40] "MostFreqParentOp"
## [41] "ParentChildFreq"
## [42] "ParentParenChildFreq"
## [43] "PoolSize"
## [44] "TransSetNo"
## [45] "TransNo"
## [46] "Type"
## [47] "NumTypes"
## [48] "MethodName"
## [49] "rowID"
```

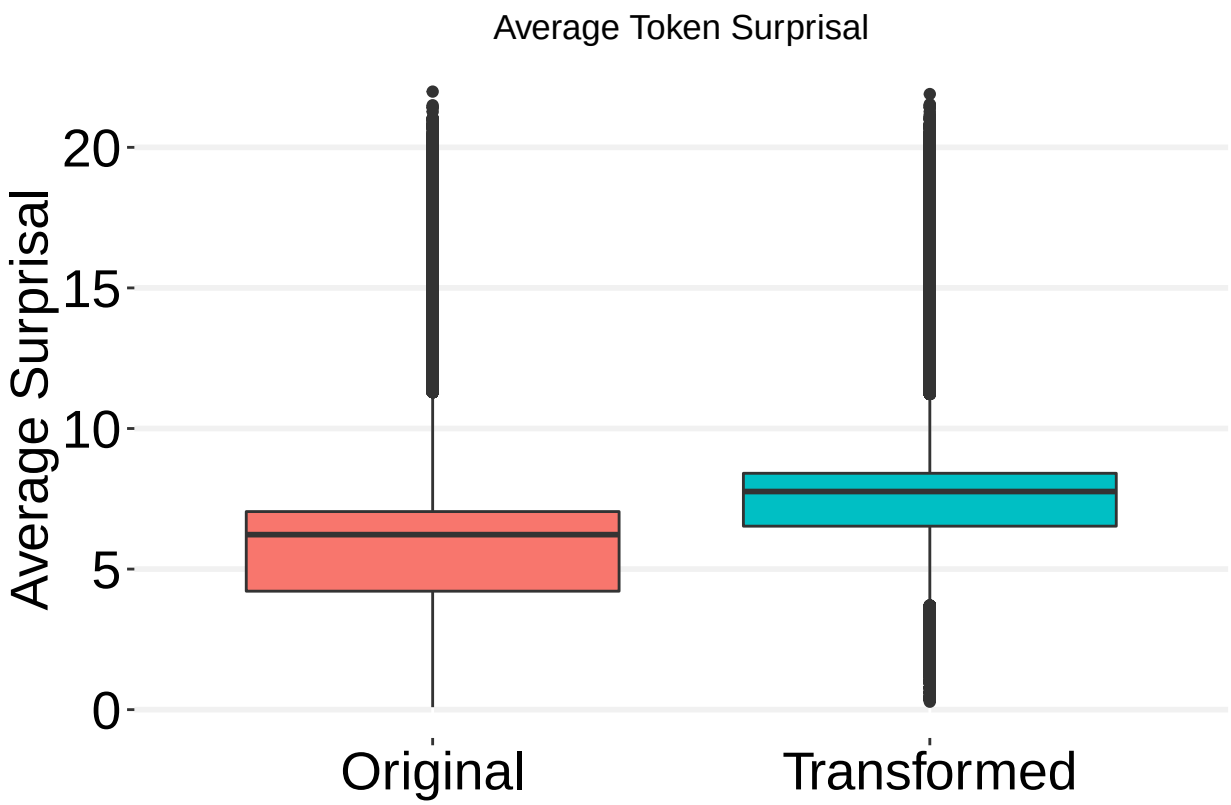

Regular
LineNumber
NumTokens
AveEnt
MaxEnt
AveEntCache
MaxEntCache
value
factor(Transformed)
factor(Transformed)
0
1
Type
LineNumber
NumTokens
TypeAveEnt
TypeMaxEnt
TypeAveEntCache
TypeMaxEntCache
value
factor(Transformed)
factor(Transformed)
0
1





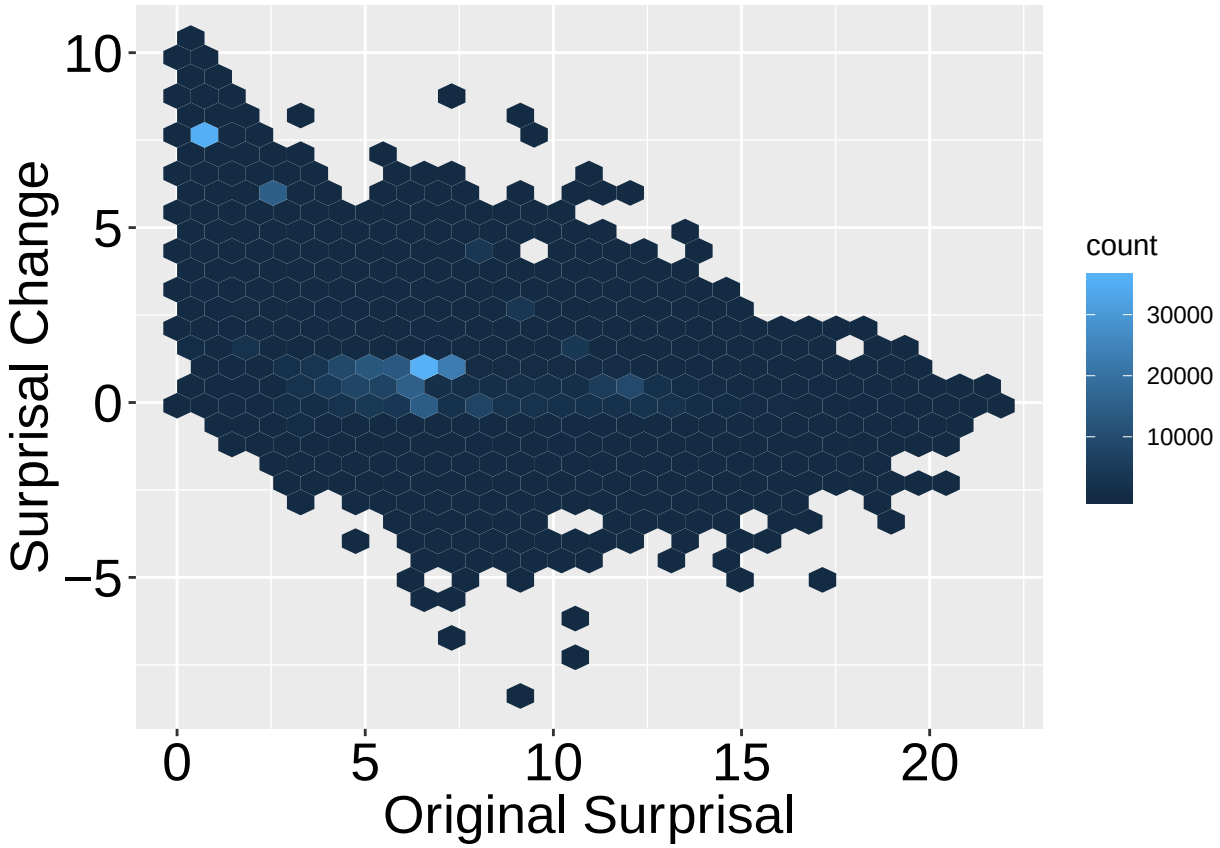

```
## [1] " ------ Expression Global Model ----- "
```

```
## [1] "AddParenTopGlobalExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -380.96, df = 297260, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##       -Inf -1.781152
## sample estimates:
## mean of the differences
##                -1.79477
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -380.96, df = 297260, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -1.809383 -1.780156
## sample estimates:
## mean of the differences
##                -1.79477

## Warning in n1 * n2: NAs produced by integer overflow

##
## Cohen's d
##
## d estimate: -0.6987372 (medium)
## 95 percent confidence interval:
## inf sup
##  NA  NA

## Warning in n1 * n2: NAs produced by integer overflow


## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1097700000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##       -Inf -0.856663
## sample estimates:
## (pseudo)median
##     -0.8683785
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1097700000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.8706618 -0.8566558
## sample estimates:
## (pseudo)median
##     -0.8683785
##
##
## Cliff's Delta
##
## delta estimate: -0.3781996 (medium)
```

```
## 95 percent confidence interval:
##        inf        sup
## -0.3809178 -0.3754749
```




```
## [1] " ------ Expression Cache Model ----- "
## [1] "AddParenTopCacheExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -475.01, df = 297260, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##       -Inf -1.506851
## sample estimates:
## mean of the differences
##               -1.516076
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -475.01, df = 297260, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -1.525977 -1.506176
## sample estimates:
```

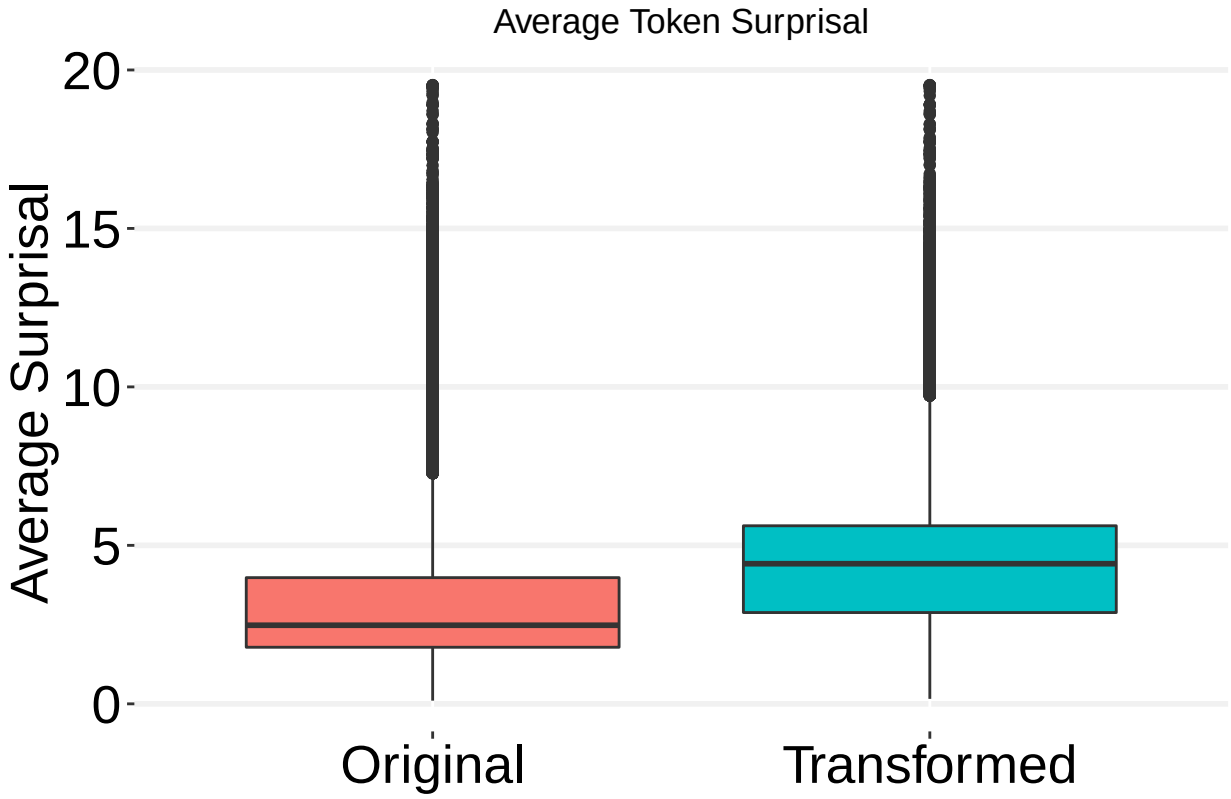
```
## mean of the differences
##               -1.516076
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.8712363 (large)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

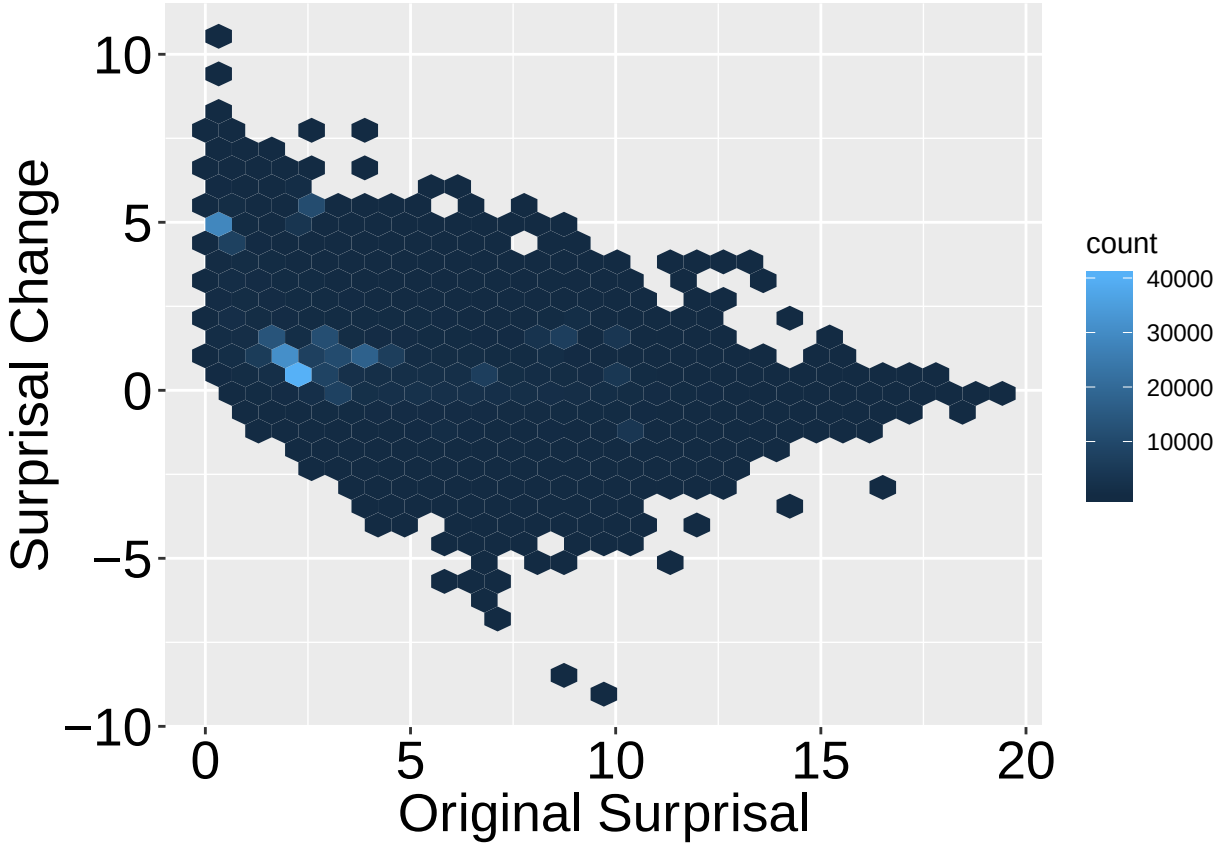
```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1950900000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##       -Inf -1.056307
## sample estimates:
## (pseudo)median
##      -1.061308
##
##
```

```
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1950900000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -1.066635 -1.055949
## sample estimates:
## (pseudo)median
##      -1.061308
##
##
## Cliff's Delta
##
## delta estimate: -0.4550117 (medium)
## 95 percent confidence interval:
##        inf        sup
## -0.4576501 -0.4523652
```




```
## [1] " ------ Expression Global Type Model ----- "
## [1] "AddParenTopGlobalTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -359.59, df = 297260, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
```

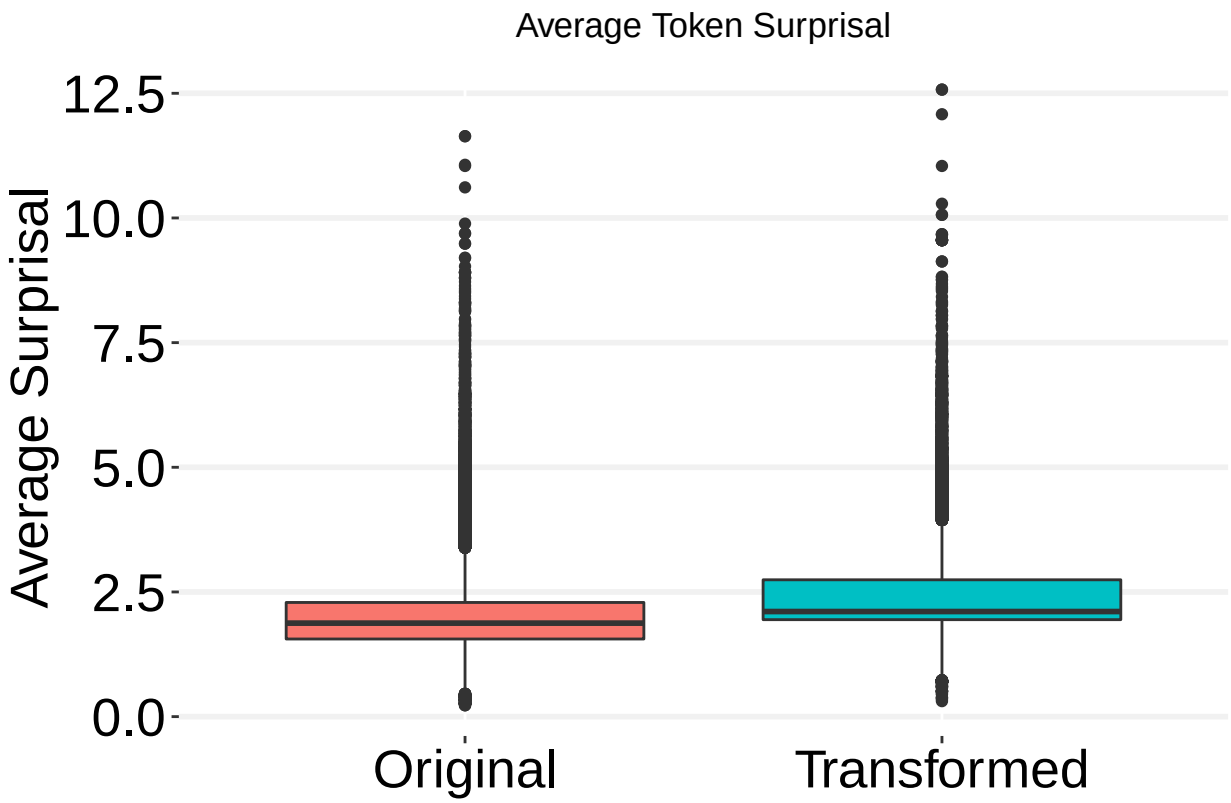
```
## 99.80769 percent confidence interval:
##        -Inf -0.3594836
## sample estimates:
## mean of the differences
##              -0.3623967
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -359.59, df = 297260, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.3655228 -0.3592705
## sample estimates:
## mean of the differences
##              -0.3623967
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.6595324 (medium)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

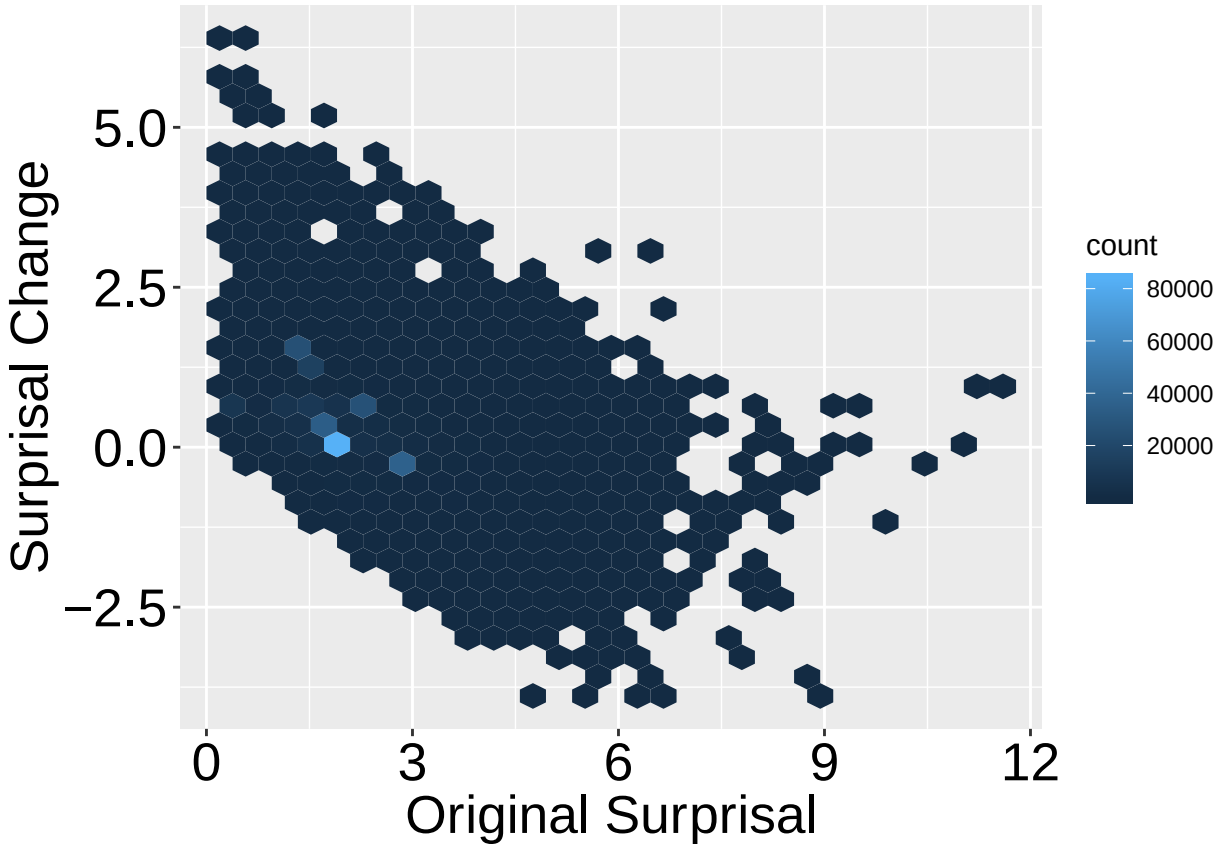
```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 6240300000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3117621
## sample estimates:
## (pseudo)median
##      -0.317461
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 6240300000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.3180092 -0.3111457
## sample estimates:
## (pseudo)median
##      -0.317461
##
##
## Cliff's Delta
##
## delta estimate: -0.4068039 (medium)
```

```
## 95 percent confidence interval:
##        inf        sup
## -0.4094816 -0.4041191
```




```
## [1] " ------ Expression Cache Type Model ----- "
## [1] "AddParenTopCacheTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -565.6, df = 297260, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.7199232
## sample estimates:
## mean of the differences
##              -0.7236213
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -565.6, df = 297260, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.7275898 -0.7196528
## sample estimates:
```

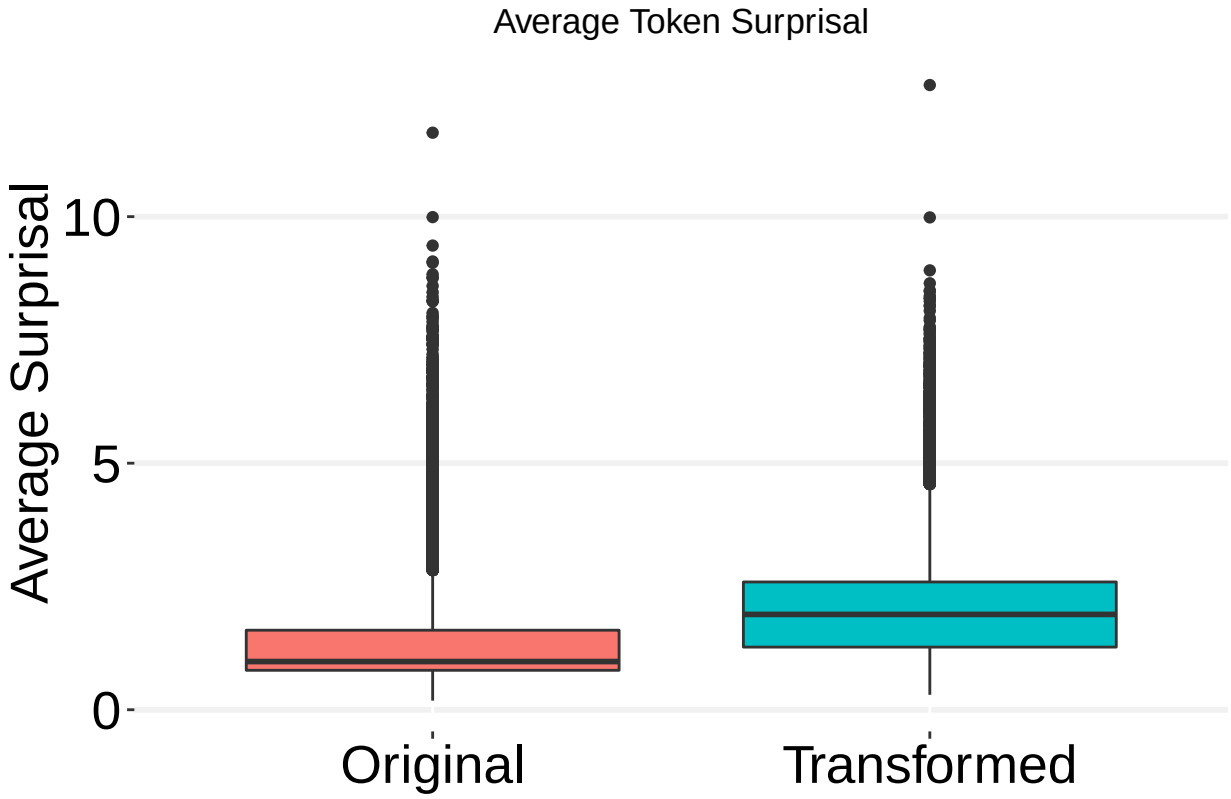
```
## mean of the differences
##              -0.7236213
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -1.037395 (large)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```



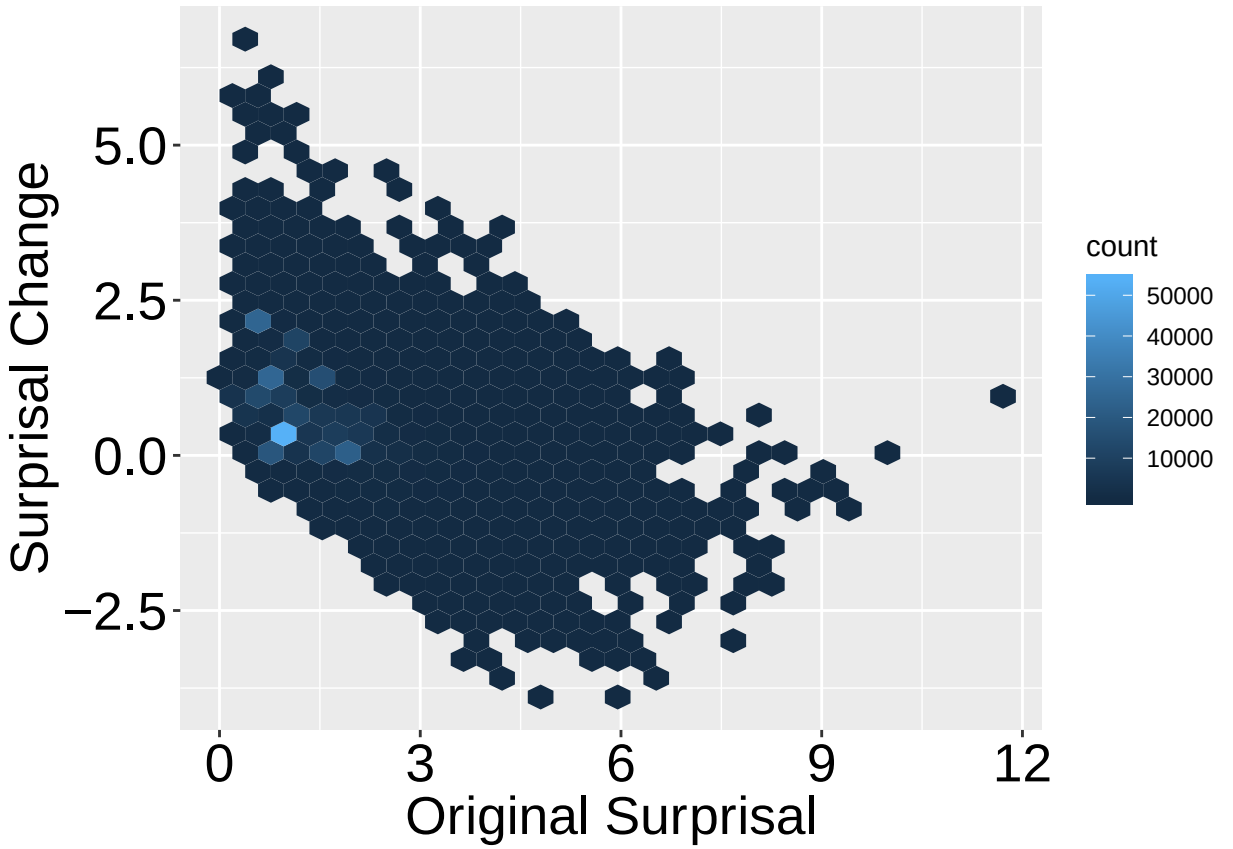

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1210300000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.6795346
## sample estimates:
## (pseudo)median
##      -0.682251
##
##
```

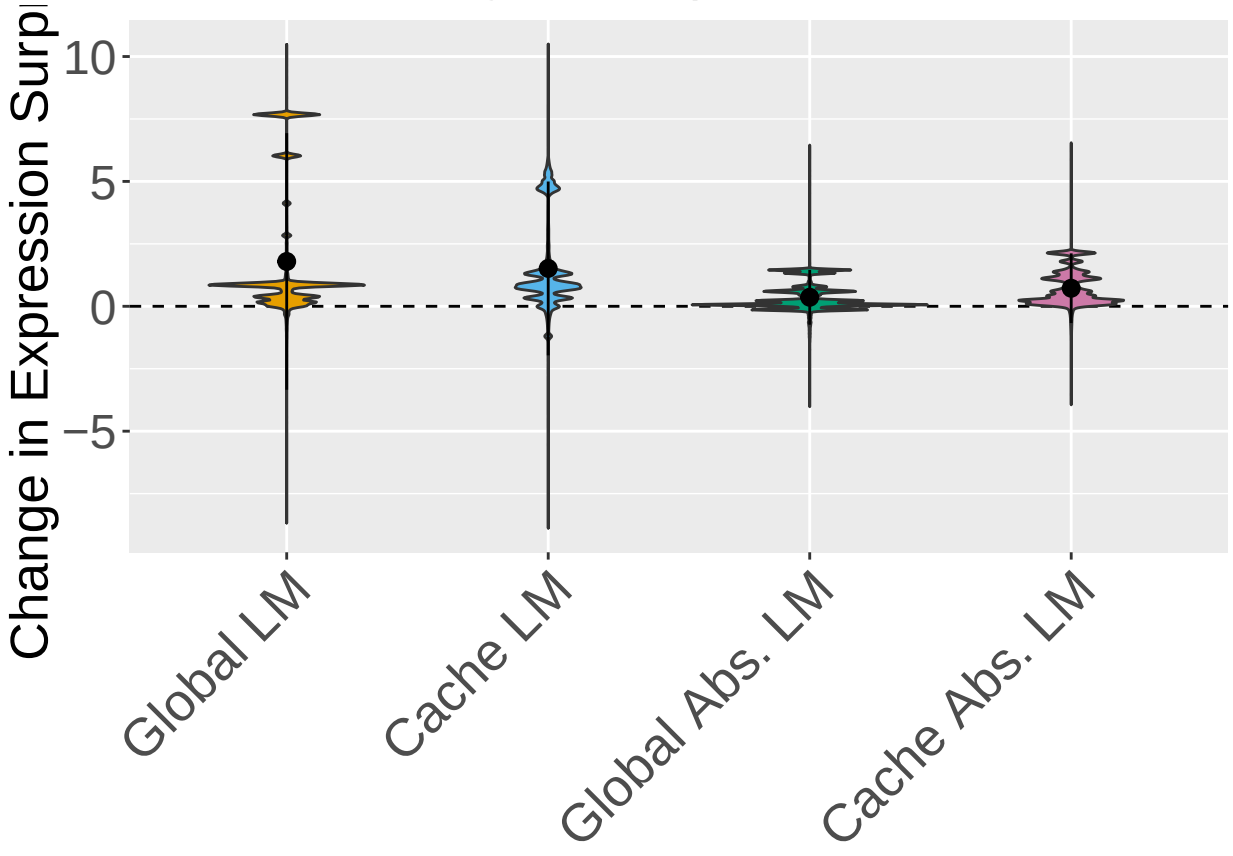
```
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1210300000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.6851791 -0.6793239
## sample estimates:
## (pseudo)median
##      -0.682251
##
##
## Cliff's Delta
##
## delta estimate: -0.5989664 (large)
## 95 percent confidence interval:
##        inf        sup
## -0.6011860 -0.5967375
```

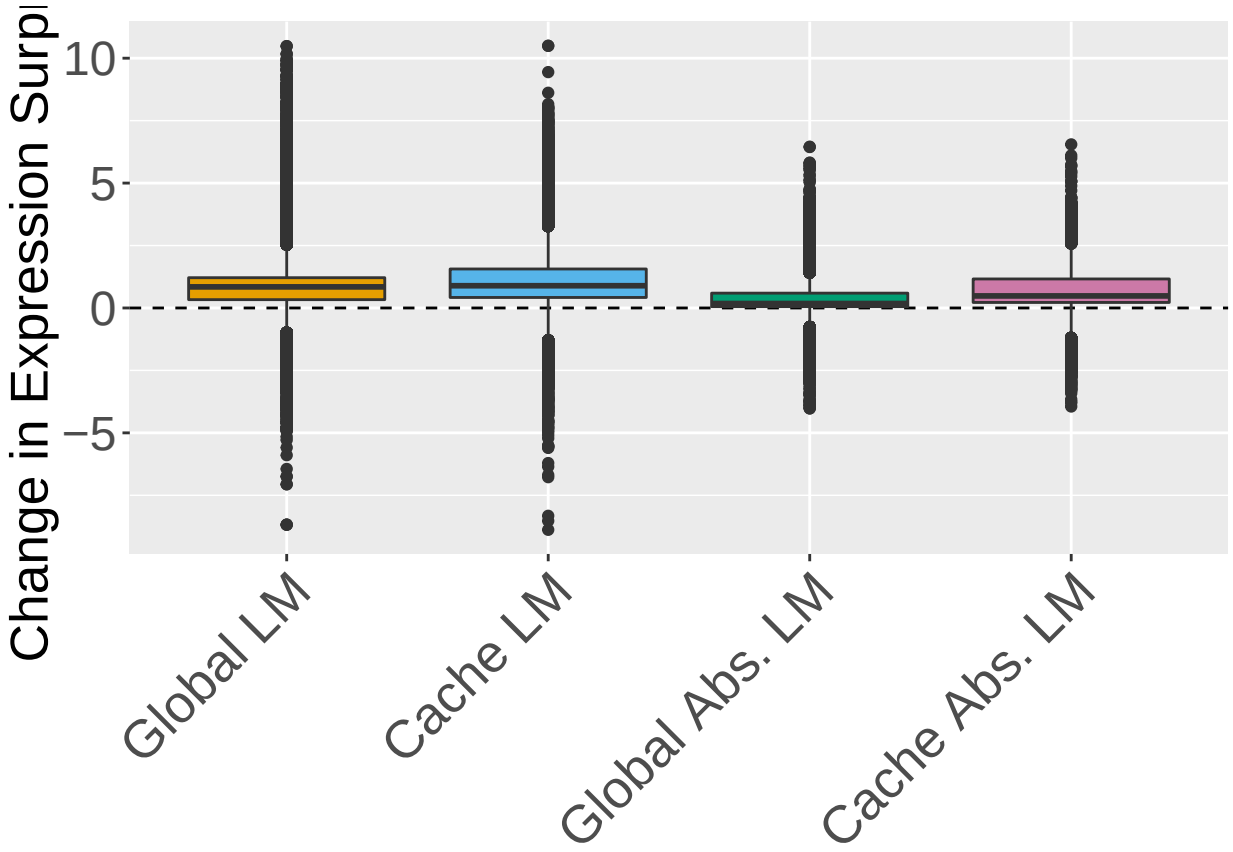
```
## [1] "Binary differences"
##
##  FALSE   TRUE
##  14216 283042
##
##  FALSE   TRUE
##  27982 269276
##
##  FALSE   TRUE
##  54962 242296
##
##  FALSE   TRUE
##  12022 285236
```

```
## No id variables; using all as measure variables
```

```
## Warning: Ignoring unknown parameters: mult
```


Surprisal Change
5.0
2.5
0.0
−2.5
0
3
6
9
12
Original Surprisal
count
50000
40000
30000
20000
10000


Change in Expression Surpl
10
5
0
−5
Global LM
Cache LM
Global Abs. LM
Cache Abs. LM




```
setwd("/data/anon/SemanticTransformation/")

#Plot results on everything (for paired table)
#dsp <- compareDepthSummary("swap.csv", "Swap", "SAME", FALSE)
#Plot for the Large dataset
#Give some sense of robustness to n for these values...
dapFiltered <- generateFilteredResults(dap, "Paren", "AddParenTopFiltered100", 100)
```

```
## [1] " ------ Expression Global Model ----- "
## [1] "AddParenTopFiltered100GlobalExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -106.12, df = 63624, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.4558477
## sample estimates:
## mean of the differences
##              -0.4686123
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -106.12, df = 63624, p-value < 2.2e-16
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.4823102 -0.4549143
## sample estimates:
## mean of the differences
##              -0.4686123
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.4207116 (small)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 314690000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3202462
## sample estimates:
## (pseudo)median
##     -0.3287996
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 314690000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.3341131 -0.3201590
## sample estimates:
## (pseudo)median
##     -0.3287996
##
##
## Cliff's Delta
##
## delta estimate: -0.06350059 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.06982758 -0.05716850
```

```
## [1] " ------ Expression Cache Model ----- "
## [1] "AddParenTopFiltered100CacheExp Original < Transformed"
##
```

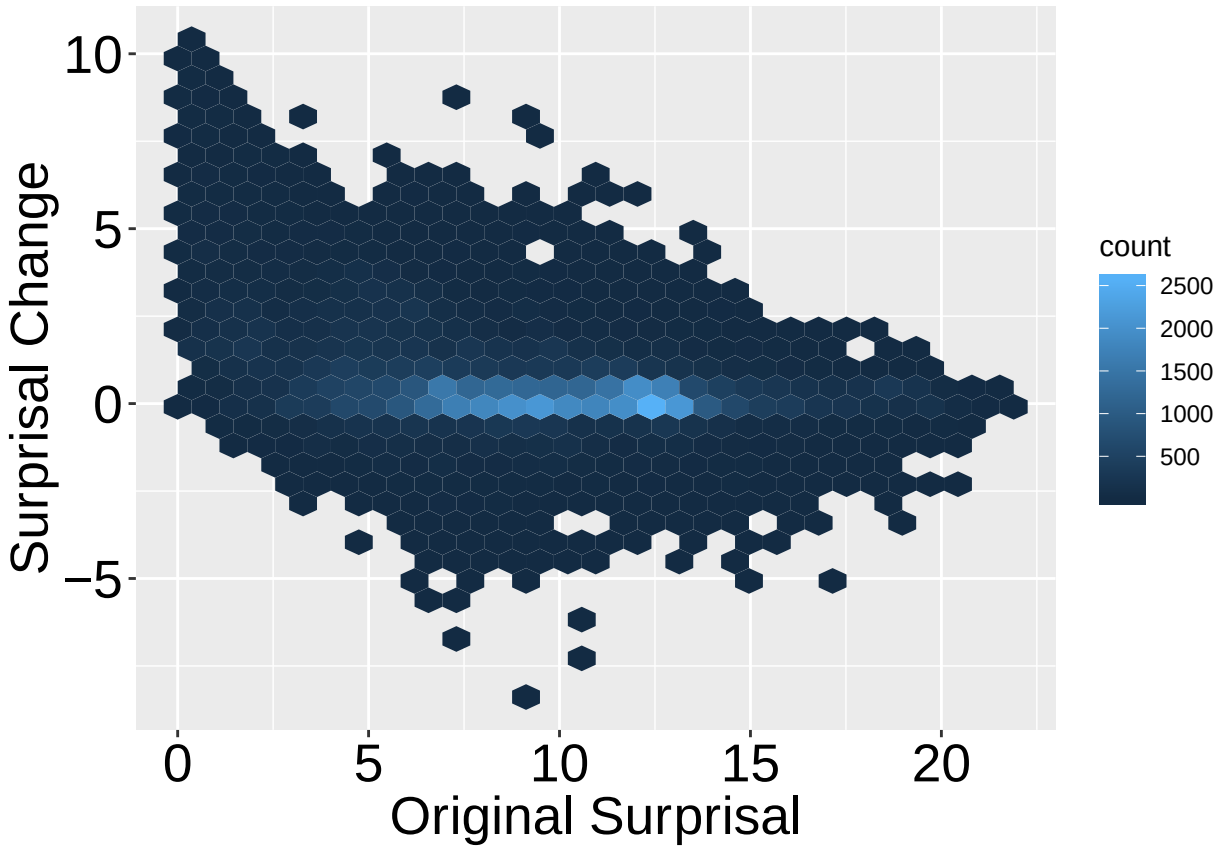
```
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -135.08, df = 63624, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.7334927
## sample estimates:
## mean of the differences
##              -0.7495321
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -135.08, df = 63624, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.7667443 -0.7323198
## sample estimates:
## mean of the differences
##              -0.7495321
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.535524 (medium)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 427430000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.5447139
## sample estimates:
## (pseudo)median
##     -0.5600593
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 427430000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.5767162 -0.5436034
## sample estimates:
## (pseudo)median
##     -0.5600593
##
##
## Cliff's Delta
##
```

```
## delta estimate: -0.151858 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.1581393 -0.1455645
```

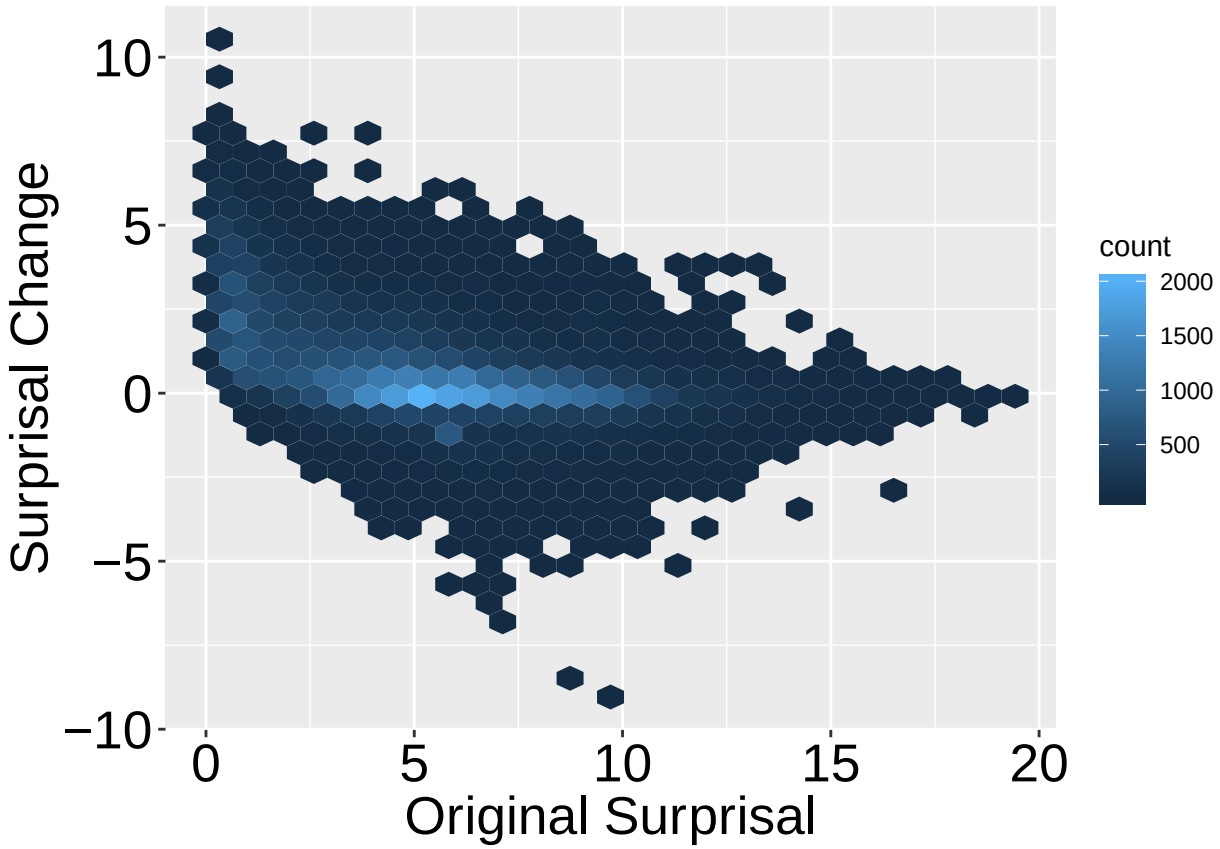
```
## [1] " ------ Expression Global Type Model ----- "
## [1] "AddParenTopFiltered100GlobalTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -89.868, df = 63624, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.2079966
## sample estimates:
## mean of the differences
##              -0.2149092
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -89.868, df = 63624, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.2223273 -0.2074912
## sample estimates:
## mean of the differences
##              -0.2149092
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.3562788 (small)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 482600000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.2242366
## sample estimates:
## (pseudo)median
##     -0.2293936
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 482600000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.2349309 -0.2238430
## sample estimates:
## (pseudo)median
##     -0.2293936
##
##
## Cliff's Delta
##
```

```
## delta estimate: -0.1675328 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.1737831 -0.1612690
```

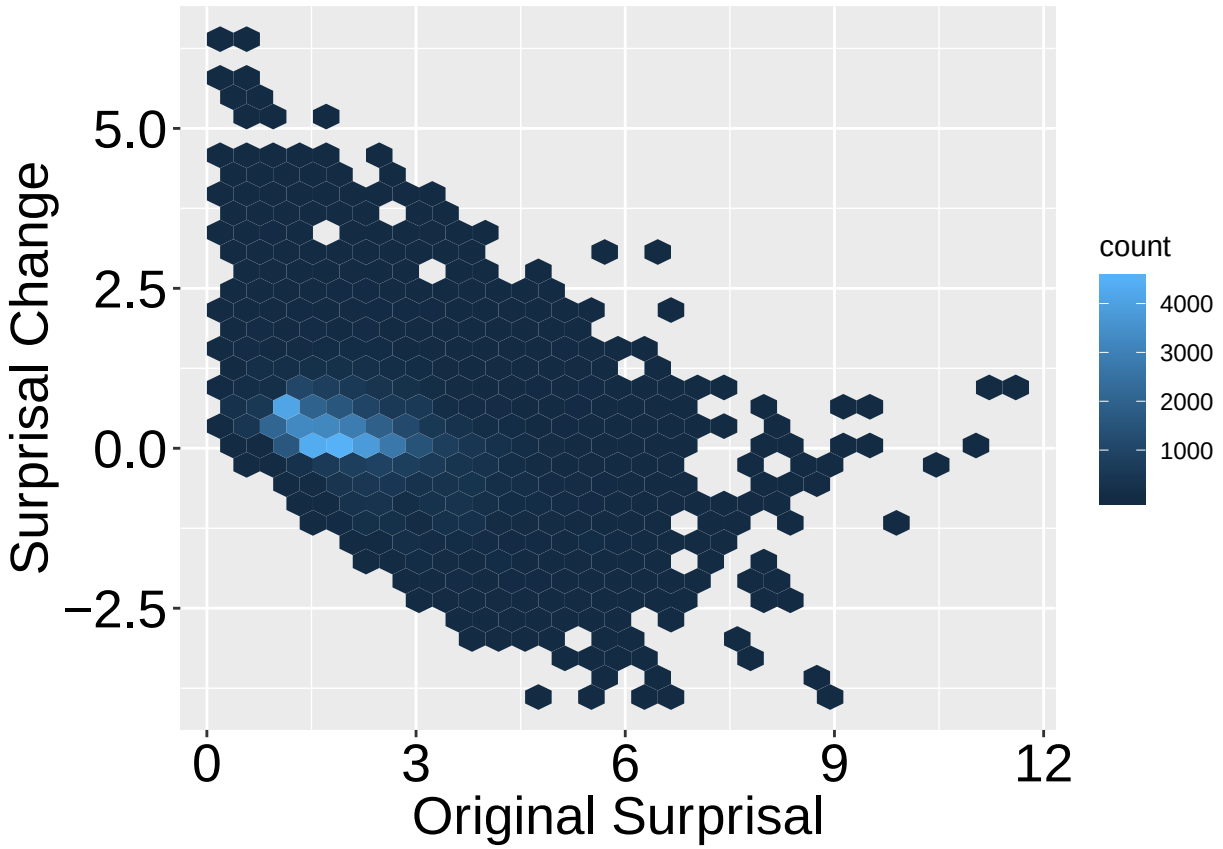
```
## [1] " ------ Expression Cache Type Model ----- "
## [1] "AddParenTopFiltered100CacheTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -150.93, df = 63624, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3925301
## sample estimates:
## mean of the differences
##              -0.4001948
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -150.93, df = 63624, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.4084201 -0.3919696
## sample estimates:
## mean of the differences
##              -0.4001948
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.5983396 (medium)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

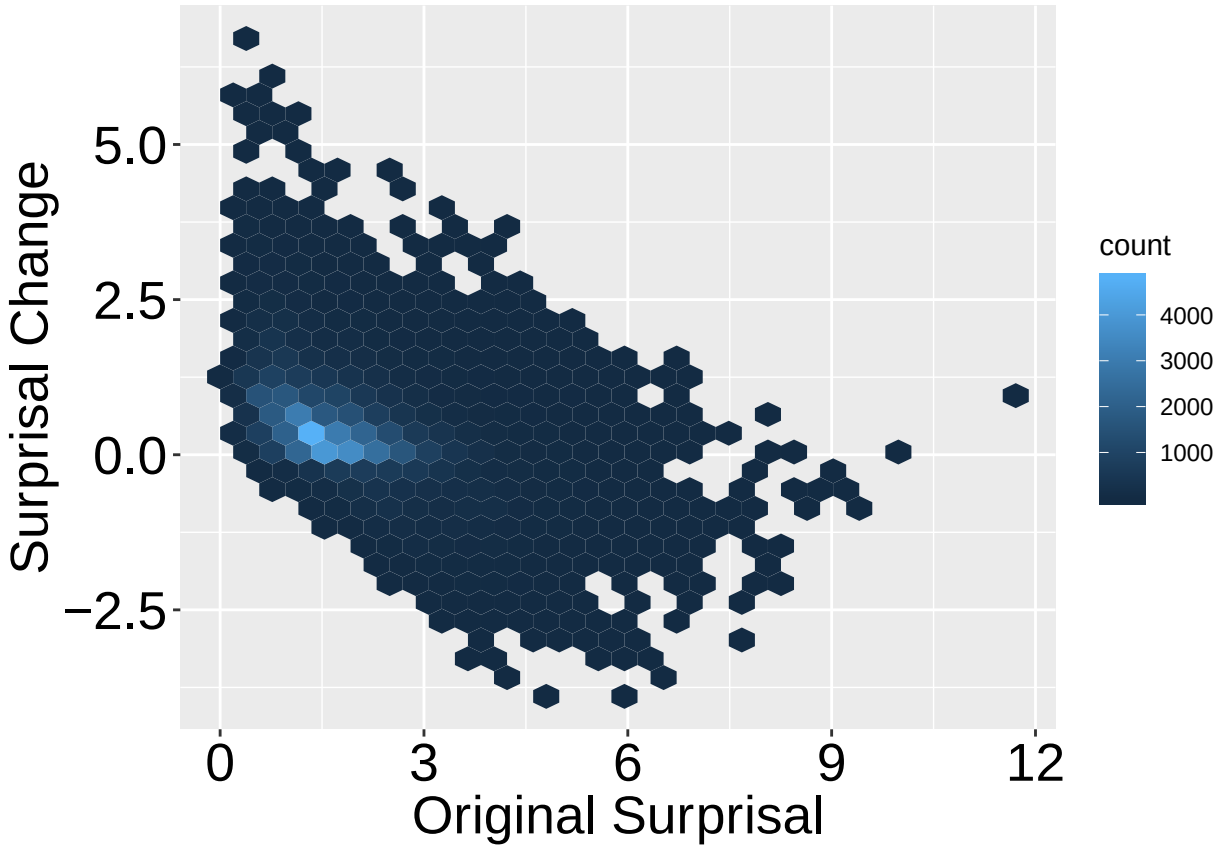


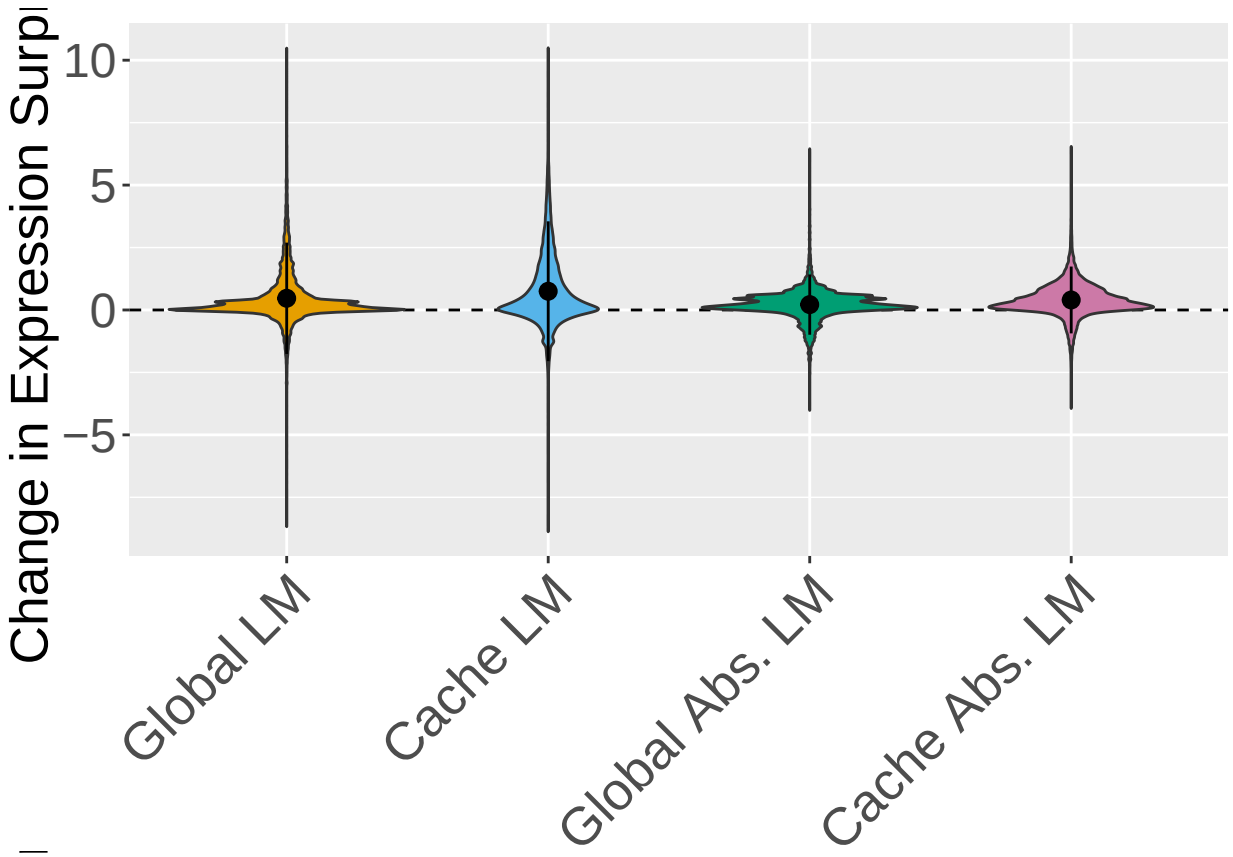

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 301870000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3646872
## sample estimates:
## (pseudo)median
##      -0.371148
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 301870000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.3781098 -0.3642093
## sample estimates:
## (pseudo)median
##      -0.371148
##
##
## Cliff's Delta
##
```

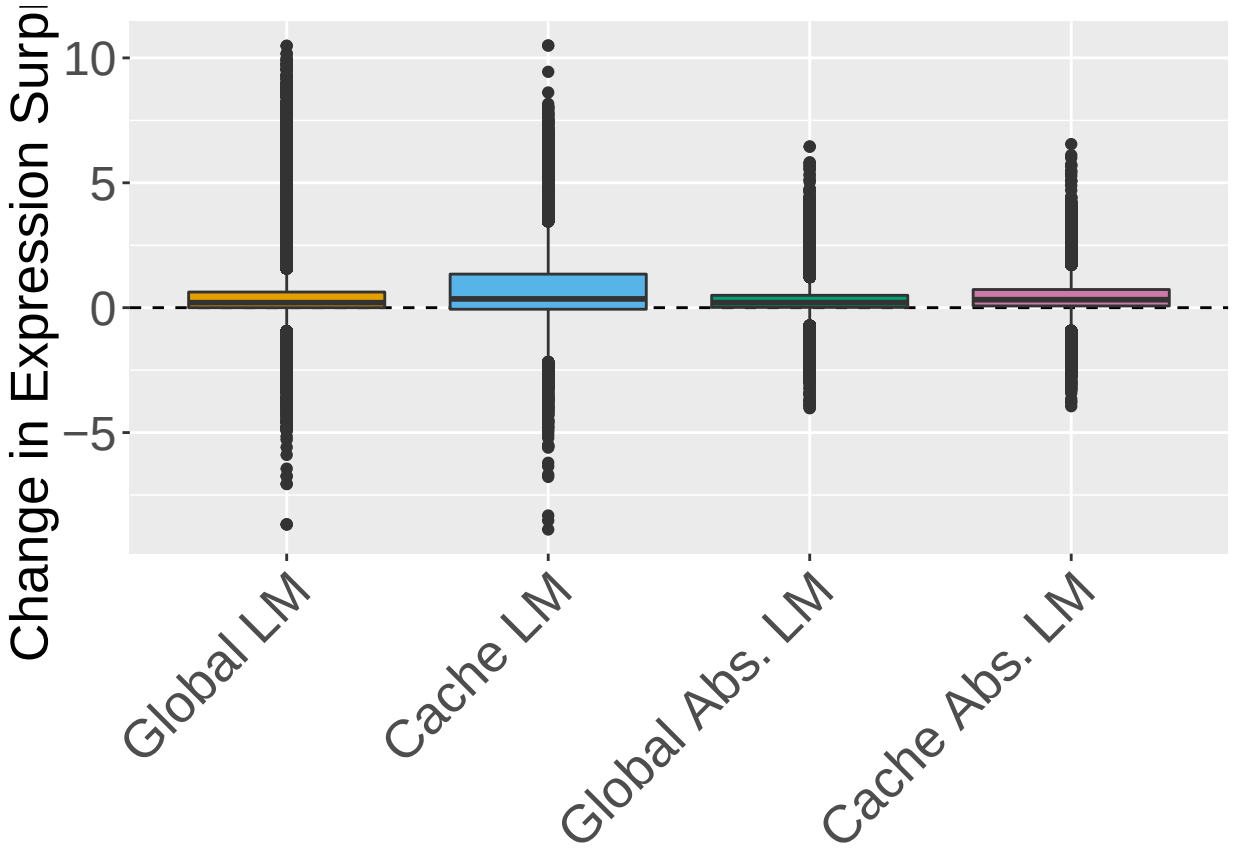
```
## delta estimate: -0.3106422 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.3166398 -0.3046196
```

```
## No id variables; using all as measure variables
```

```
## Warning: Ignoring unknown parameters: mult
```




Change in Expression Surpl
10
5
0
−5
Global LM
Cache LM
Global Abs. LM
Cache Abs. LM


Change in Expression Surpl
10
5
0
−5
Global LM
Cache LM
Global Abs. LM
Cache Abs. LM

```
dapFiltered2 <- generateFilteredResults(dap, "Paren", "AddParenTopFiltered10", 10)
```

```
## [1] " ------ Expression Global Model ----- "
## [1] "AddParenTopFiltered10GlobalExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -100.72, df = 60565, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.4406778
## sample estimates:
## mean of the differences
##              -0.4536982
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -100.72, df = 60565, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.4676707 -0.4397257
## sample estimates:
## mean of the differences
##              -0.4536982
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.4092803 (small)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

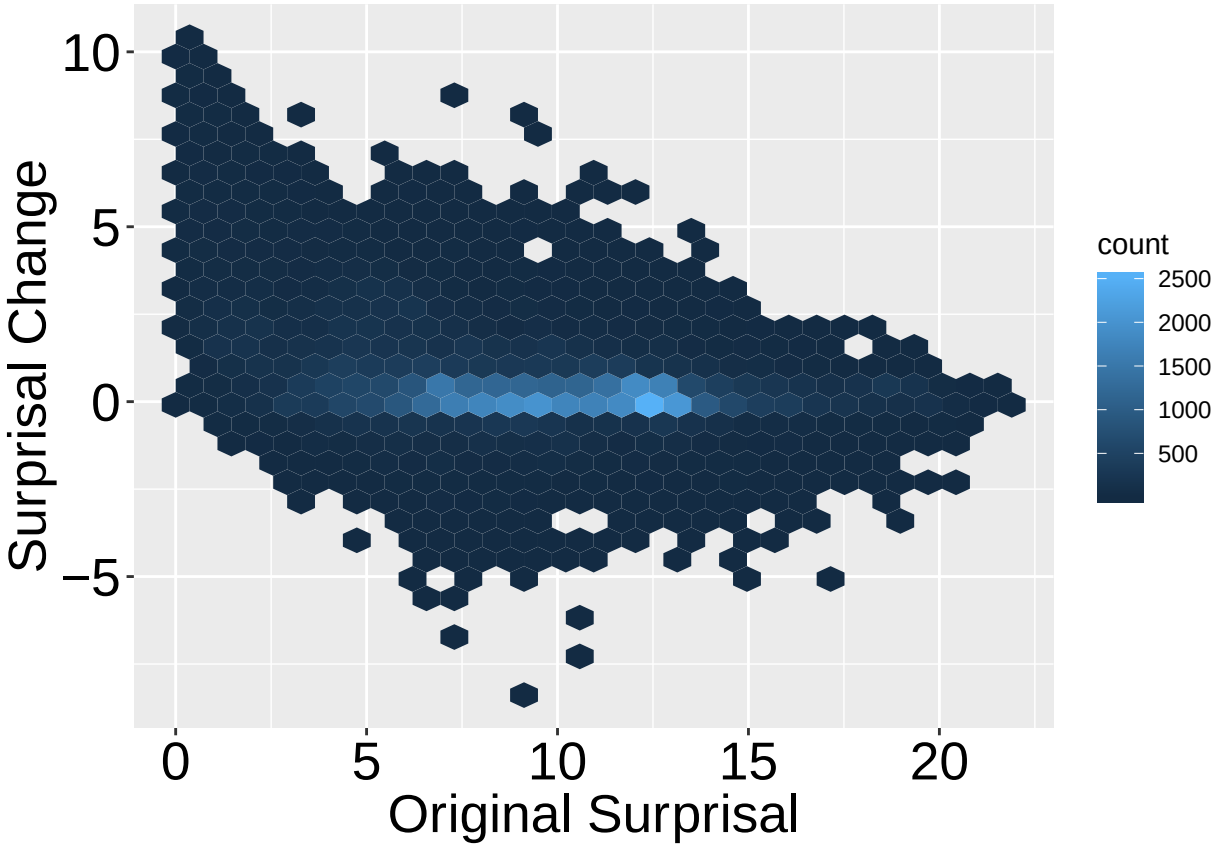
```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 290140000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3100823
## sample estimates:
## (pseudo)median
##     -0.3192056
##
##
##  Wilcoxon signed rank test with continuity correction
```

```
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 290140000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.3273266 -0.3093891
## sample estimates:
## (pseudo)median
##     -0.3192056
##
##
## Cliff's Delta
##
## delta estimate: -0.06216005 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.06864536 -0.05566949
## [1] " ------ Expression Cache Model ----- "
## [1] "AddParenTopFiltered10CacheExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -126.67, df = 60565, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.6987675
## sample estimates:
## mean of the differences
##              -0.7150862
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -126.67, df = 60565, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.7325982 -0.6975742
## sample estimates:
## mean of the differences
##              -0.7150862
## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -0.5146945 (medium)
## 95 percent confidence interval:
## inf sup
##  NA  NA
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 406960000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.5070099
## sample estimates:
## (pseudo)median
##     -0.5221597
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 406960000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.5386377 -0.5058837
## sample estimates:
## (pseudo)median
##     -0.5221597
##
##
```

```
## Cliff's Delta
##
## delta estimate: -0.1451544 (negligible)
## 95 percent confidence interval:
##        inf        sup
## -0.1515967 -0.1386997
```

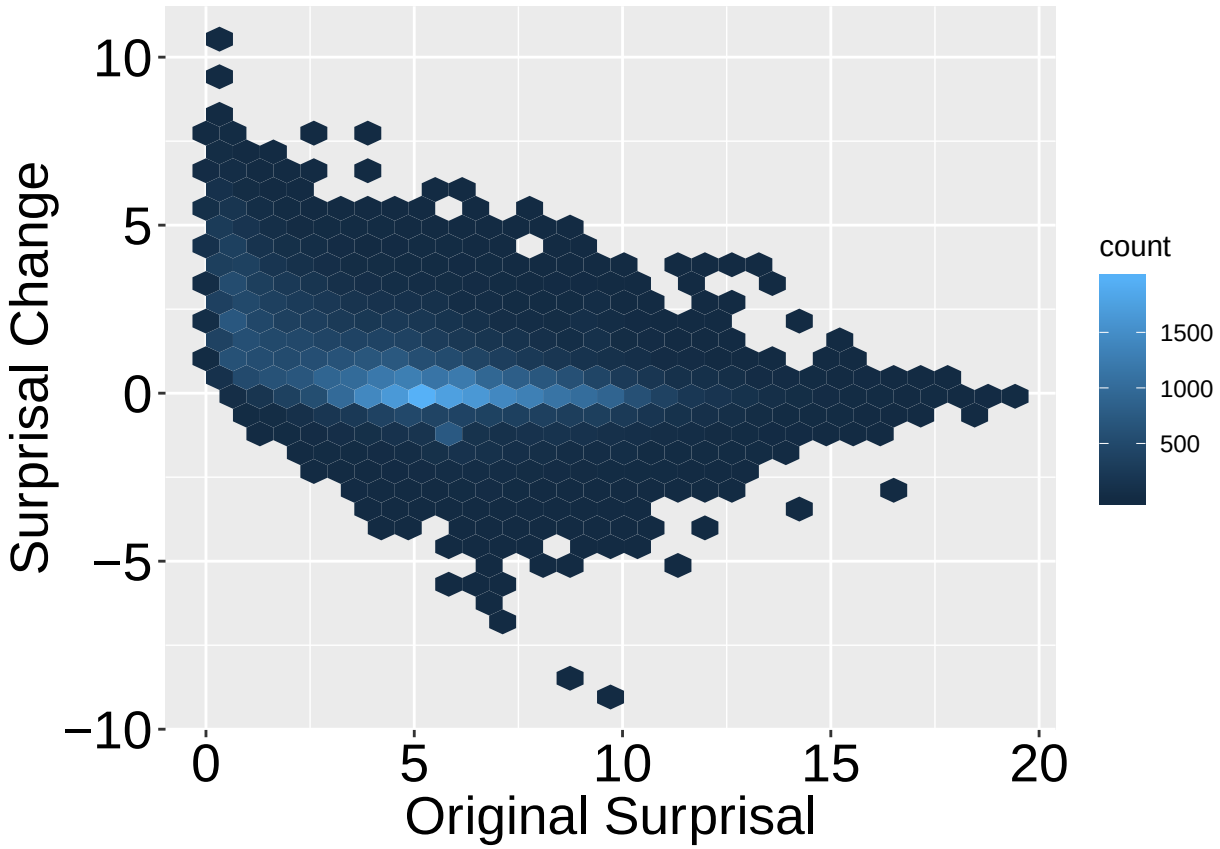
```
## [1] " ------ Expression Global Type Model ----- "
## [1] "AddParenTopFiltered10GlobalTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -85.405, df = 60565, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.2038846
## sample estimates:
## mean of the differences
##               -0.211027
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -85.405, df = 60565, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.2186918 -0.2033623
## sample estimates:
## mean of the differences
##               -0.211027
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.3470299 (small)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 443720000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.2205885
## sample estimates:
## (pseudo)median
##      -0.225933
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 443720000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.2315757 -0.2201663
## sample estimates:
## (pseudo)median
##      -0.225933
##
##
## Cliff's Delta
##
```

```
## delta estimate: -0.1653194 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.1717280 -0.1588968
```

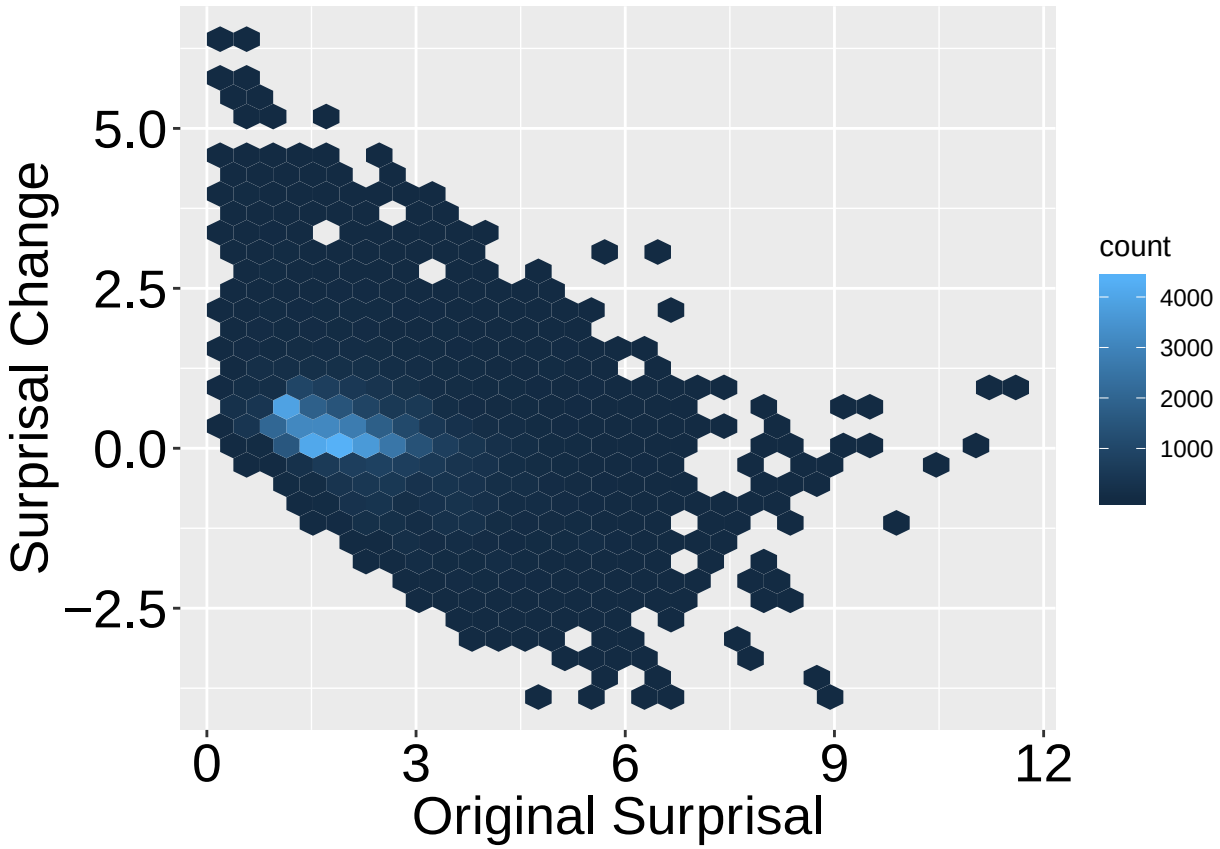
```
## [1] " ------ Expression Cache Type Model ----- "
## [1] "AddParenTopFiltered10CacheTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -143.22, df = 60565, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3811657
## sample estimates:
## mean of the differences
##              -0.3890172
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -143.22, df = 60565, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.3974429 -0.3805915
## sample estimates:
## mean of the differences
##              -0.3890172
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.5819552 (medium)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

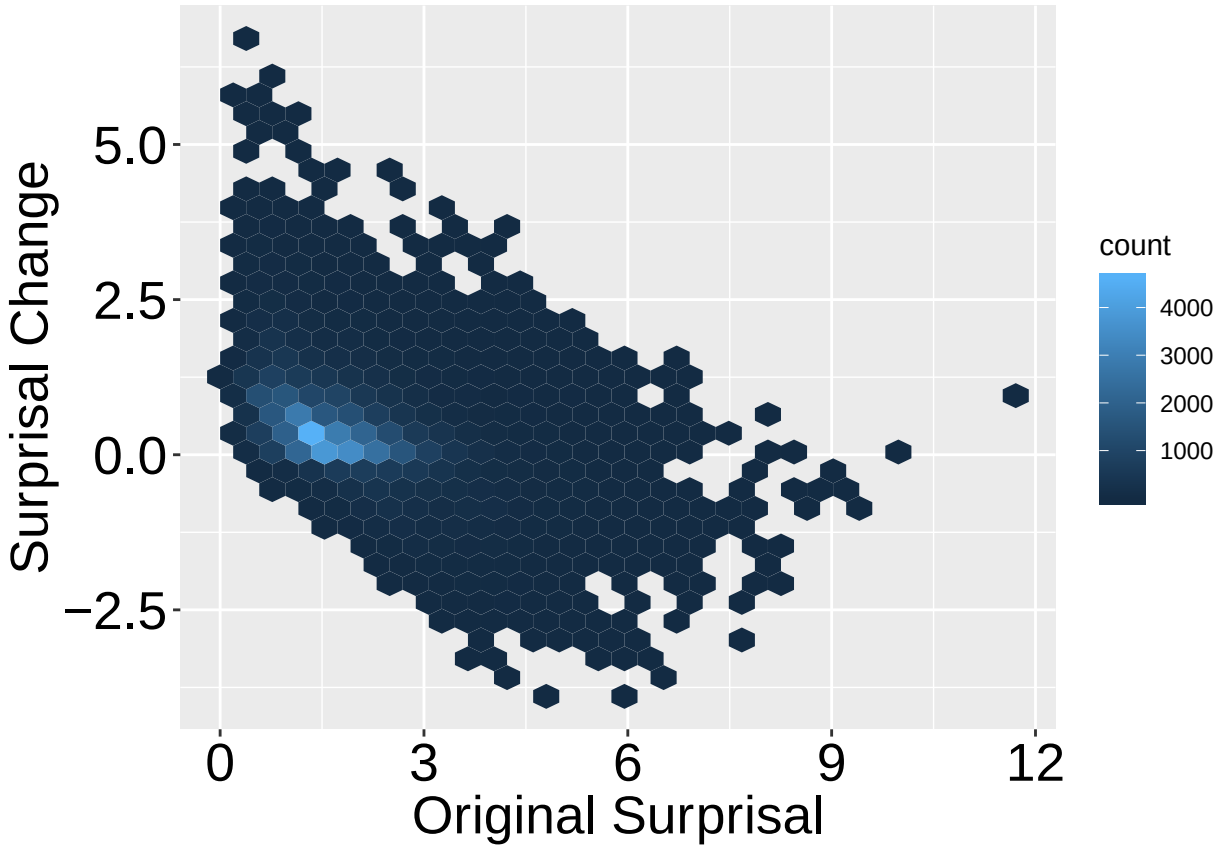


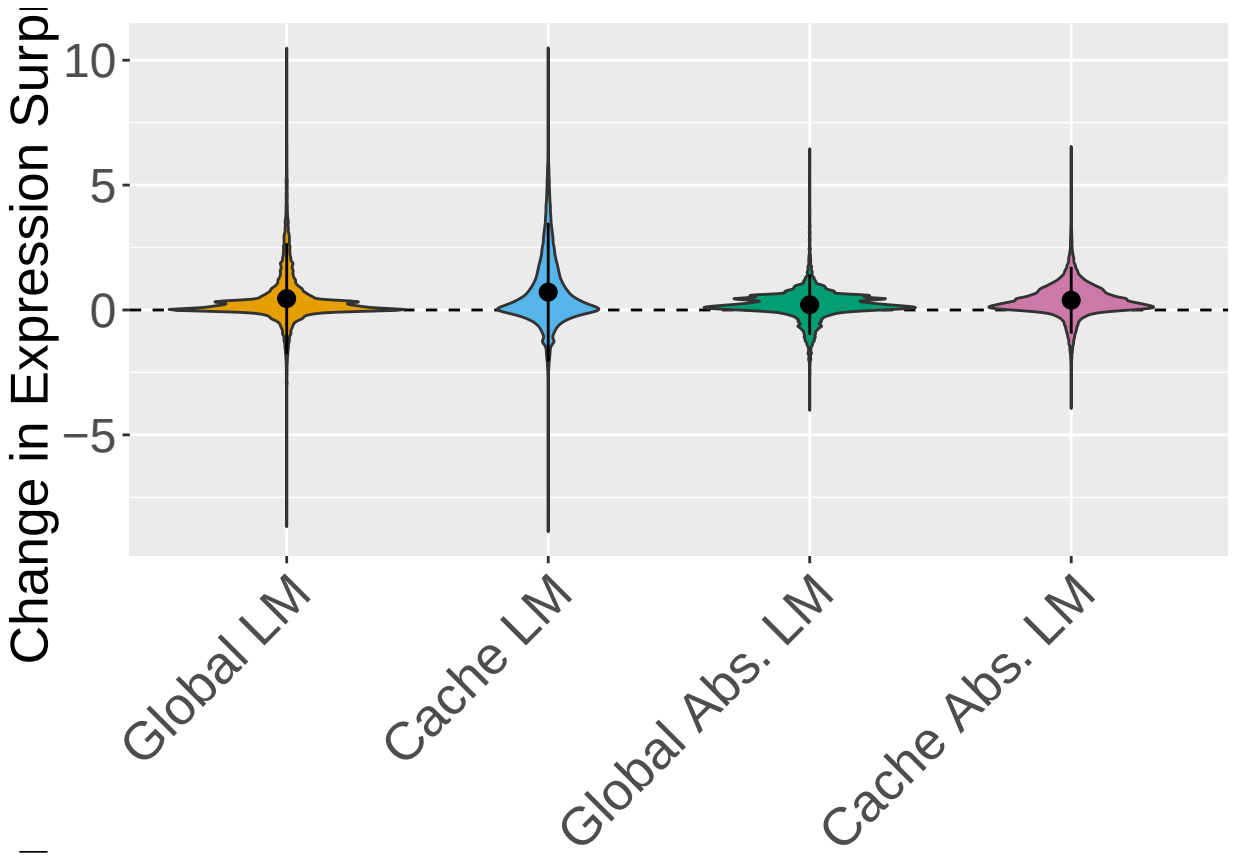

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 283060000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3537775
## sample estimates:
## (pseudo)median
##     -0.3603238
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 283060000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.3673645 -0.3533099
## sample estimates:
## (pseudo)median
##     -0.3603238
##
##
## Cliff's Delta
##
```

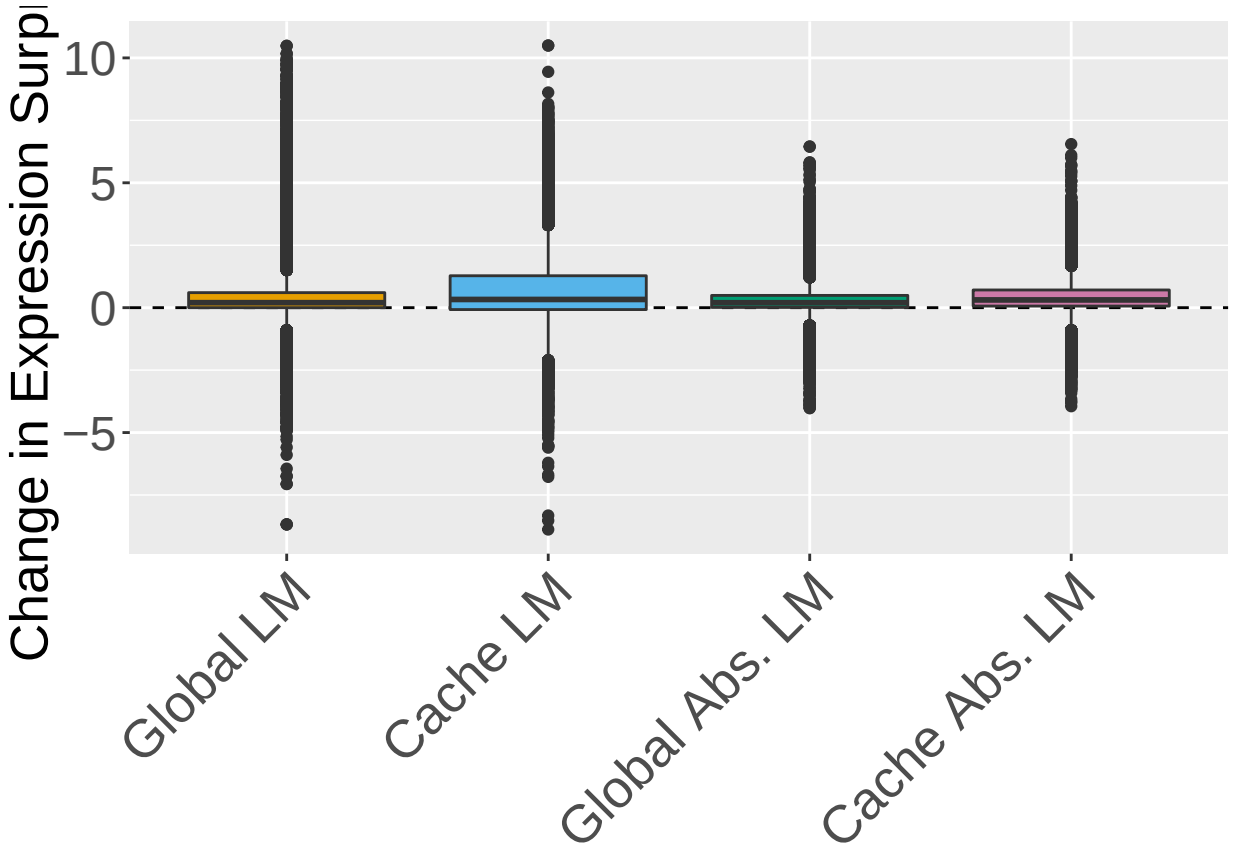
```
## delta estimate: -0.3020233 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.3081896 -0.2958316
```

```
## No id variables; using all as measure variables
```

```
## Warning: Ignoring unknown parameters: mult
```




Change in Expression Surpl
10
5
0
−5
Global LM
Cache LM
Global Abs. LM
Cache Abs. LM


Change in Expression Surpl
10
5
0
−5
Global LM
Cache LM
Global Abs. LM
Cache Abs. LM

Regression models

```
m_add_no_out <- modelGlobal(dapFiltered, "==", "")
```

```
##
## Call:
## lm(formula = AverageEntChangeExp ~ BaseAveEntExp + log(NumTokens) +
##     factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1197 -0.4426 -0.0852  0.3366  5.6377
##
## Coefficients:
##                                      Estimate Std. Error t value
## (Intercept)                          2.179050   0.113180  19.253
## BaseAveEntExp                       -0.093778   0.001138 -82.429
## log(NumTokens)                      -0.369699   0.012215 -30.266
## factor(ParentOp)<                    0.111978   0.057960   1.932
## factor(ParentOp)<=                   0.276417   0.103109   2.681
## factor(ParentOp)>                    0.228466   0.068345   3.343
## factor(ParentOp)>=                   0.040836   0.096128   0.425
## factor(ParentOp)|                    0.172284   0.520443   0.331
## factor(ParentOp)||                   0.300284   0.105989   2.833
## factor(ParentOp)-                    0.422530   0.059053   7.155
## factor(ParentOp)!=                  -0.173927   0.086000  -2.022
## factor(ParentOp)/                    0.478598   0.074655   6.411
## factor(ParentOp)*                    0.659740   0.107024   6.164
## factor(ParentOp)&&                   0.244459   0.105775   2.311
## factor(ParentOp)+                    0.525209   0.053560   9.806
## factor(ParentOp)ArrayAccess          0.226481   0.062599   3.618
## factor(ParentOp)ConditionalExpression  0.210979   0.099888   2.112
## factor(ParentOp)MethodInvocation     0.424620   0.056038   7.577
## factor(MostFreqOp)<<                -0.599336   0.515630  -1.162
## factor(MostFreqOp)<=                 0.090879   0.044267   2.053
## factor(MostFreqOp)==                -0.110239   0.023170  -4.758
## factor(MostFreqOp)>                 -0.191312   0.029162  -6.560
## factor(MostFreqOp)>=                -0.070187   0.034787  -2.018
## factor(MostFreqOp)||                -0.233880   0.049961  -4.681
## factor(MostFreqOp)-                 -0.256936   0.098237  -2.615
## factor(MostFreqOp)!=                -0.035203   0.023268  -1.513
## factor(MostFreqOp)/                 -0.466656   0.100341  -4.651
## factor(MostFreqOp)*                 -0.425243   0.097566  -4.359
## factor(MostFreqOp)&&                -0.171439   0.033011  -5.193
## factor(MostFreqOp)%                  0.100478   0.121964   0.824
## factor(MostFreqOp)+                 -0.158921   0.094514  -1.681
##                                     Pr(>|t|)
## (Intercept)                          < 2e-16 ***
## BaseAveEntExp                        < 2e-16 ***
## log(NumTokens)                       < 2e-16 ***
## factor(ParentOp)<                   0.053367 .
## factor(ParentOp)<=                  0.007347 **
## factor(ParentOp)>                   0.000830 ***
## factor(ParentOp)>=                  0.670977
## factor(ParentOp)|                   0.740621
```

```
## factor(ParentOp)||                    0.004611 **
## factor(ParentOp)-                     8.52e-13 ***
## factor(ParentOp)!=                    0.043141 *
## factor(ParentOp)/                     1.46e-10 ***
## factor(ParentOp)*                     7.15e-10 ***
## factor(ParentOp)&&                    0.020831 *
## factor(ParentOp)+                      < 2e-16 ***
## factor(ParentOp)ArrayAccess           0.000297 ***
## factor(ParentOp)ConditionalExpression 0.034679 *
## factor(ParentOp)MethodInvocation      3.61e-14 ***
## factor(MostFreqOp)<<                  0.245106
## factor(MostFreqOp)<=                  0.040082 *
## factor(MostFreqOp)==                  1.96e-06 ***
## factor(MostFreqOp)>                   5.44e-11 ***
## factor(MostFreqOp)>=                  0.043637 *
## factor(MostFreqOp)||                  2.86e-06 ***
## factor(MostFreqOp)-                   0.008914 **
## factor(MostFreqOp)!=                  0.130312
## factor(MostFreqOp)/                   3.32e-06 ***
## factor(MostFreqOp)*                   1.31e-05 ***
## factor(MostFreqOp)&&                  2.08e-07 ***
## factor(MostFreqOp)%                   0.410036
## factor(MostFreqOp)+                   0.092682 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8824 on 37390 degrees of freedom
## Multiple R-squared:  0.1785, Adjusted R-squared:  0.1778
## F-statistic: 270.8 on 30 and 37390 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: AverageEntChangeExp
##                       Df  Sum Sq Mean Sq  F value    Pr(>F)
## BaseAveEntExp          1  4957.2  4957.2 6366.774 < 2.2e-16 ***
## log(NumTokens)         1   885.1   885.1 1136.833 < 2.2e-16 ***
## factor(ParentOp)      15   259.5    17.3   22.222 < 2.2e-16 ***
## factor(MostFreqOp)    13   222.9    17.1   22.023 < 2.2e-16 ***
## Residuals          37390 29112.0     0.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.139888402" "0.024978077" "0.007323721" "0.006290581" "0.821519219"
##                            GVIF Df GVIF^(1/(2*Df))
## BaseAveEntExp          1.046128  1        1.022804
## log(NumTokens)         1.193014  1        1.092252
## factor(ParentOp)   10555.281042 15        1.361807
## factor(MostFreqOp) 10394.921281 13        1.427227
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Feb 19, 2019 - 10:43:52 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
```

```
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & AverageEntChangeExp \\
## \hline \\[-1.8ex]
##  BaseAveEntExp & $-$0.094$^{***}$ (0.001) \\
##   log(NumTokens) & $-$0.370$^{***}$ (0.012) \\
##   factor(ParentOp)\textless  & 0.112$^{*}$ (0.058) \\
##   factor(ParentOp)\textless = & 0.276$^{***}$ (0.103) \\
##   factor(ParentOp)\textgreater  & 0.228$^{***}$ (0.068) \\
##   factor(ParentOp)\textgreater = & 0.041 (0.096) \\
##   factor(ParentOp)\textbar  & 0.172 (0.520) \\
##   factor(ParentOp)\textbar \textbar  & 0.300$^{***}$ (0.106) \\
##   factor(ParentOp)- & 0.423$^{***}$ (0.059) \\
##   factor(ParentOp)!= & $-$0.174$^{**}$ (0.086) \\
##   factor(ParentOp)/ & 0.479$^{***}$ (0.075) \\
##   factor(ParentOp)\textasteriskcentered  & 0.660$^{***}$ (0.107) \\
##   factor(ParentOp)&& & 0.244$^{**}$ (0.106) \\
##   factor(ParentOp)+ & 0.525$^{***}$ (0.054) \\
##   factor(ParentOp)ArrayAccess & 0.226$^{***}$ (0.063) \\
##   factor(ParentOp)ConditionalExpression & 0.211$^{**}$ (0.100) \\
##   factor(ParentOp)MethodInvocation & 0.425$^{***}$ (0.056) \\
##   factor(MostFreqOp)\textless \textless  & $-$0.599 (0.516) \\
##   factor(MostFreqOp)\textless = & 0.091$^{**}$ (0.044) \\
##   factor(MostFreqOp)== & $-$0.110$^{***}$ (0.023) \\
##   factor(MostFreqOp)\textgreater  & $-$0.191$^{***}$ (0.029) \\
##   factor(MostFreqOp)\textgreater = & $-$0.070$^{**}$ (0.035) \\
##   factor(MostFreqOp)\textbar \textbar  & $-$0.234$^{***}$ (0.050) \\
##   factor(MostFreqOp)- & $-$0.257$^{***}$ (0.098) \\
##   factor(MostFreqOp)!= & $-$0.035 (0.023) \\
##   factor(MostFreqOp)/ & $-$0.467$^{***}$ (0.100) \\
##   factor(MostFreqOp)\textasteriskcentered  & $-$0.425$^{***}$ (0.098) \\
##   factor(MostFreqOp)&& & $-$0.171$^{***}$ (0.033) \\
##   factor(MostFreqOp)\% & 0.100 (0.122) \\
##   factor(MostFreqOp)+ & $-$0.159$^{*}$ (0.095) \\
##   Constant & 2.179$^{***}$ (0.113) \\
##  \hline \\[-1.8ex]
## Observations & 37,421 \\
## R$^{2}$ & 0.178 \\
## Adjusted R$^{2}$ & 0.178 \\
## Residual Std. Error & 0.882 (df = 37390) \\
## F Statistic & 270.775$^{***}$ (df = 30; 37390) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:43:53 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
```

```
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## BaseAveEntExp & 1 & 4957.19 & 4957.19 & 6366.77 & 0.0000 \\
##   log(NumTokens) & 1 & 885.14 & 885.14 & 1136.83 & 0.0000 \\
##   factor(ParentOp) & 15 & 259.53 & 17.30 & 22.22 & 0.0000 \\
##   factor(MostFreqOp) & 13 & 222.92 & 17.15 & 22.02 & 0.0000 \\
##   Residuals & 37390 & 29111.98 & 0.78 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
m_add_cache_no_out <- modelCache(dapFiltered, "==", "")
```

```
##
## Call:
## lm(formula = CacheAverageEntChangeExp ~ BaseCacheAveEntExp +
##     log(NumTokens) + factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3811 -0.7162 -0.0678  0.6486  5.5433
##
## Coefficients:
##                                      Estimate Std. Error  t value
## (Intercept)                          3.559737   0.167051   21.309
## BaseCacheAveEntExp                  -0.276120   0.001836 -150.353
## log(NumTokens)                      -0.733055   0.015522  -47.227
## factor(ParentOp)<                    0.193517   0.078679    2.460
## factor(ParentOp)<=                   0.289303   0.147313    1.964
## factor(ParentOp)>                    0.364843   0.091974    3.967
## factor(ParentOp)>=                   0.093842   0.128618    0.730
## factor(ParentOp)|                   -0.169086   0.655847   -0.258
## factor(ParentOp)||                   0.715220   0.160195    4.465
## factor(ParentOp)-                    0.457827   0.079690    5.745
## factor(ParentOp)!=                  -0.383989   0.114551   -3.352
## factor(ParentOp)/                    0.542584   0.101035    5.370
## factor(ParentOp)*                    1.103179   0.149338    7.387
## factor(ParentOp)&&                   0.815167   0.159981    5.095
## factor(ParentOp)+                    0.850937   0.073316   11.606
## factor(ParentOp)ArrayAccess          1.024024   0.085612   11.961
## factor(ParentOp)ConditionalExpression  0.866899   0.154000    5.629
## factor(ParentOp)MethodInvocation     0.770796   0.075838   10.164
## factor(MostFreqOp)<<                 0.260390   0.664397    0.392
## factor(MostFreqOp)<=                 0.289312   0.056351    5.134
## factor(MostFreqOp)==                -0.223372   0.028949   -7.716
## factor(MostFreqOp)>                 -0.241401   0.036583   -6.599
## factor(MostFreqOp)>=                -0.123301   0.043084   -2.862
## factor(MostFreqOp)||                -0.167267   0.063657   -2.628
## factor(MostFreqOp)-                 -0.078692   0.149502   -0.526
## factor(MostFreqOp)!=                -0.068890   0.029073   -2.370
## factor(MostFreqOp)/                 -0.165868   0.152022   -1.091
## factor(MostFreqOp)*                  0.155950   0.148847    1.048
## factor(MostFreqOp)&&                -0.281381   0.041484   -6.783
## factor(MostFreqOp)%                  0.738028   0.178019    4.146
## factor(MostFreqOp)+                  0.538602   0.145685    3.697
```

```
##                                      Pr(>|t|)
## (Intercept)                           < 2e-16 ***
## BaseCacheAveEntExp                    < 2e-16 ***
## log(NumTokens)                        < 2e-16 ***
## factor(ParentOp)<                    0.013915 *
## factor(ParentOp)<=                   0.049552 *
## factor(ParentOp)>                    7.30e-05 ***
## factor(ParentOp)>=                   0.465628
## factor(ParentOp)|                    0.796553
## factor(ParentOp)||                   8.04e-06 ***
## factor(ParentOp)-                    9.26e-09 ***
## factor(ParentOp)!=                   0.000803 ***
## factor(ParentOp)/                    7.91e-08 ***
## factor(ParentOp)*                    1.53e-13 ***
## factor(ParentOp)&&                   3.50e-07 ***
## factor(ParentOp)+                     < 2e-16 ***
## factor(ParentOp)ArrayAccess           < 2e-16 ***
## factor(ParentOp)ConditionalExpression 1.82e-08 ***
## factor(ParentOp)MethodInvocation      < 2e-16 ***
## factor(MostFreqOp)<<                 0.695120
## factor(MostFreqOp)<=                 2.85e-07 ***
## factor(MostFreqOp)==                 1.23e-14 ***
## factor(MostFreqOp)>                  4.20e-11 ***
## factor(MostFreqOp)>=                 0.004214 **
## factor(MostFreqOp)||                 0.008602 **
## factor(MostFreqOp)-                  0.598641
## factor(MostFreqOp)!=                 0.017815 *
## factor(MostFreqOp)/                  0.275244
## factor(MostFreqOp)*                  0.294776
## factor(MostFreqOp)&&                 1.19e-11 ***
## factor(MostFreqOp)%                  3.39e-05 ***
## factor(MostFreqOp)+                  0.000218 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.114 on 37252 degrees of freedom
## Multiple R-squared:  0.405,  Adjusted R-squared:  0.4045
## F-statistic: 845.3 on 30 and 37252 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: CacheAverageEntChangeExp
##                       Df Sum Sq Mean Sq   F value    Pr(>F)
## BaseCacheAveEntExp     1  26561 26561.4 21385.585 < 2.2e-16 ***
## log(NumTokens)         1   2617  2617.5  2107.425 < 2.2e-16 ***
## factor(ParentOp)      15   1399    93.3    75.101 < 2.2e-16 ***
## factor(MostFreqOp)    13    918    70.6    56.851 < 2.2e-16 ***
## Residuals          37252  46268     1.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.34156499" "0.03365924" "0.01799233" "0.01180421" "0.59497923"
##                            GVIF Df GVIF^(1/(2*Df))
## BaseCacheAveEntExp     1.042653  1        1.021104
## log(NumTokens)         1.183536  1        1.087904
```

```
## factor(ParentOp)   16661.378773 15        1.382686
## factor(MostFreqOp) 16624.451669 13        1.453237
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Feb 19, 2019 - 10:44:05 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & CacheAverageEntChangeExp \\
## \hline \\[-1.8ex]
##  BaseCacheAveEntExp & $-$0.276$^{***}$ (0.002) \\
##   log(NumTokens) & $-$0.733$^{***}$ (0.016) \\
##   factor(ParentOp)\textless  & 0.194$^{**}$ (0.079) \\
##   factor(ParentOp)\textless = & 0.289$^{**}$ (0.147) \\
##   factor(ParentOp)\textgreater  & 0.365$^{***}$ (0.092) \\
##   factor(ParentOp)\textgreater = & 0.094 (0.129) \\
##   factor(ParentOp)\textbar  & $-$0.169 (0.656) \\
##   factor(ParentOp)\textbar \textbar  & 0.715$^{***}$ (0.160) \\
##   factor(ParentOp)- & 0.458$^{***}$ (0.080) \\
##   factor(ParentOp)!= & $-$0.384$^{***}$ (0.115) \\
##   factor(ParentOp)/ & 0.543$^{***}$ (0.101) \\
##   factor(ParentOp)\textasteriskcentered  & 1.103$^{***}$ (0.149) \\
##   factor(ParentOp)&& & 0.815$^{***}$ (0.160) \\
##   factor(ParentOp)+ & 0.851$^{***}$ (0.073) \\
##   factor(ParentOp)ArrayAccess & 1.024$^{***}$ (0.086) \\
##   factor(ParentOp)ConditionalExpression & 0.867$^{***}$ (0.154) \\
##   factor(ParentOp)MethodInvocation & 0.771$^{***}$ (0.076) \\
##   factor(MostFreqOp)\textless \textless  & 0.260 (0.664) \\
##   factor(MostFreqOp)\textless = & 0.289$^{***}$ (0.056) \\
##   factor(MostFreqOp)== & $-$0.223$^{***}$ (0.029) \\
##   factor(MostFreqOp)\textgreater  & $-$0.241$^{***}$ (0.037) \\
##   factor(MostFreqOp)\textgreater = & $-$0.123$^{***}$ (0.043) \\
##   factor(MostFreqOp)\textbar \textbar  & $-$0.167$^{***}$ (0.064) \\
##   factor(MostFreqOp)- & $-$0.079 (0.150) \\
##   factor(MostFreqOp)!= & $-$0.069$^{**}$ (0.029) \\
##   factor(MostFreqOp)/ & $-$0.166 (0.152) \\
##   factor(MostFreqOp)\textasteriskcentered  & 0.156 (0.149) \\
##   factor(MostFreqOp)&& & $-$0.281$^{***}$ (0.041) \\
##   factor(MostFreqOp)\% & 0.738$^{***}$ (0.178) \\
##   factor(MostFreqOp)+ & 0.539$^{***}$ (0.146) \\
##   Constant & 3.560$^{***}$ (0.167) \\
##  \hline \\[-1.8ex]
## Observations & 37,283 \\
## R$^{2}$ & 0.405 \\
## Adjusted R$^{2}$ & 0.405 \\
## Residual Std. Error & 1.114 (df = 37252) \\
## F Statistic & 845.286$^{***}$ (df = 30; 37252) \\
## \hline
## \hline \\[-1.8ex]
```

```
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:44:05 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## BaseCacheAveEntExp & 1 & 26561.37 & 26561.37 & 21385.59 & 0.0000 \\
##   log(NumTokens) & 1 & 2617.47 & 2617.47 & 2107.42 & 0.0000 \\
##   factor(ParentOp) & 15 & 1399.15 & 93.28 & 75.10 & 0.0000 \\
##   factor(MostFreqOp) & 13 & 917.94 & 70.61 & 56.85 & 0.0000 \\
##   Residuals & 37252 & 46267.81 & 1.24 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
mt_add_no_out <- modelGlobalType(dapFiltered, "==", "")
```

```
##
## Call:
## lm(formula = TypeAverageEntChangeExp ~ BaseTypeAveEntExp + log(NumTokens) +
##     factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.14732 -0.18968  0.03651  0.17409  2.61040
##
## Coefficients:
##                                      Estimate Std. Error  t value
## (Intercept)                          0.823503   0.055192   14.921
## BaseTypeAveEntExp                   -0.317708   0.002344 -135.536
## log(NumTokens)                      -0.022613   0.005077   -4.454
## factor(ParentOp)<                   -0.153453   0.027295   -5.622
## factor(ParentOp)<=                   0.215647   0.058824    3.666
## factor(ParentOp)>                    0.120390   0.031932    3.770
## factor(ParentOp)>=                   0.069732   0.046294    1.506
## factor(ParentOp)|                    0.561099   0.167277    3.354
## factor(ParentOp)||                   0.214276   0.052727    4.064
## factor(ParentOp)-                    0.494256   0.027986   17.661
## factor(ParentOp)!=                   0.060757   0.038403    1.582
## factor(ParentOp)/                    0.706673   0.033993   20.789
## factor(ParentOp)*                    0.810717   0.046511   17.431
## factor(ParentOp)&&                   0.186778   0.052633    3.549
## factor(ParentOp)+                    0.647730   0.025722   25.182
## factor(ParentOp)ArrayAccess          0.227702   0.029408    7.743
## factor(ParentOp)ConditionalExpression  0.092723   0.050422    1.839
## factor(ParentOp)MethodInvocation     0.358639   0.026401   13.584
## factor(MostFreqOp)<<                -0.509585   0.167703   -3.039
## factor(MostFreqOp)<=                 0.603568   0.023353   25.845
## factor(MostFreqOp)==                -0.213780   0.009521  -22.454
## factor(MostFreqOp)>                 -0.013938   0.012007   -1.161
```

```
## factor(MostFreqOp)>=                  -0.039361   0.014181   -2.776
## factor(MostFreqOp)||                  -0.266655   0.020871  -12.776
## factor(MostFreqOp)-                   -0.508790   0.048473  -10.496
## factor(MostFreqOp)!=                  -0.170760   0.009567  -17.849
## factor(MostFreqOp)/                   -0.461595   0.049392   -9.346
## factor(MostFreqOp)*                   -0.608626   0.048405  -12.574
## factor(MostFreqOp)&&                  -0.283986   0.013688  -20.747
## factor(MostFreqOp)%                   -0.047808   0.058622   -0.816
## factor(MostFreqOp)+                   -0.319927   0.047262   -6.769
##                                      Pr(>|t|)
## (Intercept)                           < 2e-16 ***
## BaseTypeAveEntExp                     < 2e-16 ***
## log(NumTokens)                       8.47e-06 ***
## factor(ParentOp)<                    1.90e-08 ***
## factor(ParentOp)<=                   0.000247 ***
## factor(ParentOp)>                    0.000163 ***
## factor(ParentOp)>=                   0.132002
## factor(ParentOp)|                    0.000796 ***
## factor(ParentOp)||                   4.84e-05 ***
## factor(ParentOp)-                     < 2e-16 ***
## factor(ParentOp)!=                   0.113631
## factor(ParentOp)/                     < 2e-16 ***
## factor(ParentOp)*                     < 2e-16 ***
## factor(ParentOp)&&                   0.000388 ***
## factor(ParentOp)+                     < 2e-16 ***
## factor(ParentOp)ArrayAccess          9.97e-15 ***
## factor(ParentOp)ConditionalExpression 0.065931 .
## factor(ParentOp)MethodInvocation      < 2e-16 ***
## factor(MostFreqOp)<<                 0.002378 **
## factor(MostFreqOp)<=                  < 2e-16 ***
## factor(MostFreqOp)==                  < 2e-16 ***
## factor(MostFreqOp)>                  0.245717
## factor(MostFreqOp)>=                 0.005513 **
## factor(MostFreqOp)||                  < 2e-16 ***
## factor(MostFreqOp)-                   < 2e-16 ***
## factor(MostFreqOp)!=                  < 2e-16 ***
## factor(MostFreqOp)/                   < 2e-16 ***
## factor(MostFreqOp)*                   < 2e-16 ***
## factor(MostFreqOp)&&                  < 2e-16 ***
## factor(MostFreqOp)%                  0.414766
## factor(MostFreqOp)+                  1.31e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.367 on 36706 degrees of freedom
## Multiple R-squared:  0.4779, Adjusted R-squared:  0.4775
## F-statistic:  1120 on 30 and 36706 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: TypeAverageEntChangeExp
##                       Df Sum Sq Mean Sq  F value    Pr(>F)
## BaseTypeAveEntExp      1 3544.1  3544.1 26309.83 < 2.2e-16 ***
## log(NumTokens)         1    1.5     1.5    11.21 0.0008144 ***
```

```
## factor(ParentOp)      15  540.4    36.0   267.43 < 2.2e-16 ***
## factor(MostFreqOp)    13  440.3    33.9   251.44 < 2.2e-16 ***
## Residuals          36706 4944.5     0.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.3742122887" "0.0001594444" "0.0570560841" "0.0464922639"
## [5] "0.5220799189"
##                           GVIF Df GVIF^(1/(2*Df))
## BaseTypeAveEntExp     1.403241  1        1.184585
## log(NumTokens)        1.164802  1        1.079260
## factor(ParentOp)   9591.168409 15        1.357466
## factor(MostFreqOp) 9689.232169 13        1.423373
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Feb 19, 2019 - 10:44:11 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & TypeAverageEntChangeExp \\
## \hline \\[-1.8ex]
##  BaseTypeAveEntExp & $-$0.318$^{***}$ (0.002) \\
##   log(NumTokens) & $-$0.023$^{***}$ (0.005) \\
##   factor(ParentOp)\textless  & $-$0.153$^{***}$ (0.027) \\
##   factor(ParentOp)\textless = & 0.216$^{***}$ (0.059) \\
##   factor(ParentOp)\textgreater  & 0.120$^{***}$ (0.032) \\
##   factor(ParentOp)\textgreater = & 0.070 (0.046) \\
##   factor(ParentOp)\textbar  & 0.561$^{***}$ (0.167) \\
##   factor(ParentOp)\textbar \textbar  & 0.214$^{***}$ (0.053) \\
##   factor(ParentOp)- & 0.494$^{***}$ (0.028) \\
##   factor(ParentOp)!= & 0.061 (0.038) \\
##   factor(ParentOp)/ & 0.707$^{***}$ (0.034) \\
##   factor(ParentOp)\textasteriskcentered  & 0.811$^{***}$ (0.047) \\
##   factor(ParentOp)&& & 0.187$^{***}$ (0.053) \\
##   factor(ParentOp)+ & 0.648$^{***}$ (0.026) \\
##   factor(ParentOp)ArrayAccess & 0.228$^{***}$ (0.029) \\
##   factor(ParentOp)ConditionalExpression & 0.093$^{*}$ (0.050) \\
##   factor(ParentOp)MethodInvocation & 0.359$^{***}$ (0.026) \\
##   factor(MostFreqOp)\textless \textless  & $-$0.510$^{***}$ (0.168) \\
##   factor(MostFreqOp)\textless = & 0.604$^{***}$ (0.023) \\
##   factor(MostFreqOp)== & $-$0.214$^{***}$ (0.010) \\
##   factor(MostFreqOp)\textgreater  & $-$0.014 (0.012) \\
##   factor(MostFreqOp)\textgreater = & $-$0.039$^{***}$ (0.014) \\
##   factor(MostFreqOp)\textbar \textbar  & $-$0.267$^{***}$ (0.021) \\
##   factor(MostFreqOp)- & $-$0.509$^{***}$ (0.048) \\
##   factor(MostFreqOp)!= & $-$0.171$^{***}$ (0.010) \\
##   factor(MostFreqOp)/ & $-$0.462$^{***}$ (0.049) \\
##   factor(MostFreqOp)\textasteriskcentered  & $-$0.609$^{***}$ (0.048) \\
##   factor(MostFreqOp)&& & $-$0.284$^{***}$ (0.014) \\
##   factor(MostFreqOp)\% & $-$0.048 (0.059) \\
```

```
##   factor(MostFreqOp)+ & $-$0.320$^{***}$ (0.047) \\
##   Constant & 0.824$^{***}$ (0.055) \\
##  \hline \\[-1.8ex]
## Observations & 36,737 \\
## R$^{2}$ & 0.478 \\
## Adjusted R$^{2}$ & 0.477 \\
## Residual Std. Error & 0.367 (df = 36706) \\
## F Statistic & 1,120.041$^{***}$ (df = 30; 36706) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:44:12 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## BaseTypeAveEntExp & 1 & 3544.07 & 3544.07 & 26309.83 & 0.0000 \\
##   log(NumTokens) & 1 & 1.51 & 1.51 & 11.21 & 0.0008 \\
##   factor(ParentOp) & 15 & 540.36 & 36.02 & 267.43 & 0.0000 \\
##   factor(MostFreqOp) & 13 & 440.32 & 33.87 & 251.44 & 0.0000 \\
##   Residuals & 36706 & 4944.48 & 0.13 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
mt_add_cache_no_out <- modelCacheType(dapFiltered, "==", "")
```

```
##
## Call:
## lm(formula = CacheTypeAverageEntChangeExp ~ BaseCacheTypeAveEntExp +
##     log(NumTokens) + factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1449 -0.2522 -0.0267  0.2188  3.0280
##
## Coefficients:
##                                    Estimate Std. Error  t value
## (Intercept)                         1.579380   0.063476   24.881
## BaseCacheTypeAveEntExp             -0.438460   0.002616 -167.613
## log(NumTokens)                     -0.158296   0.005877  -26.933
## factor(ParentOp)<                  -0.077276   0.030703   -2.517
## factor(ParentOp)<=                  0.097373   0.065008    1.498
## factor(ParentOp)>                   0.061752   0.036461    1.694
## factor(ParentOp)>=                 -0.034279   0.050925   -0.673
## factor(ParentOp)|                   0.118288   0.192870    0.613
## factor(ParentOp)||                  0.091738   0.060702    1.511
## factor(ParentOp)-                   0.332588   0.031372   10.602
## factor(ParentOp)!=                 -0.020785   0.044418   -0.468
## factor(ParentOp)/                   0.552651   0.039666   13.933
```

```
## factor(ParentOp)*                      0.829813   0.063306   13.108
## factor(ParentOp)&&                     0.063701   0.060611    1.051
## factor(ParentOp)+                      0.601605   0.028837   20.862
## factor(ParentOp)ArrayAccess            0.233443   0.033535    6.961
## factor(ParentOp)ConditionalExpression  0.083967   0.058094    1.445
## factor(ParentOp)MethodInvocation       0.554675   0.029787   18.621
## factor(MostFreqOp)<<                  -0.222762   0.195165   -1.141
## factor(MostFreqOp)<=                   0.634745   0.025519   24.874
## factor(MostFreqOp)==                  -0.238528   0.010980  -21.723
## factor(MostFreqOp)>                   -0.001604   0.013849   -0.116
## factor(MostFreqOp)>=                  -0.010240   0.016322   -0.627
## factor(MostFreqOp)||                  -0.275644   0.024096  -11.440
## factor(MostFreqOp)-                   -0.499871   0.056257   -8.885
## factor(MostFreqOp)!=                  -0.164135   0.011047  -14.858
## factor(MostFreqOp)/                   -0.428117   0.057434   -7.454
## factor(MostFreqOp)*                   -0.418365   0.056048   -7.464
## factor(MostFreqOp)&&                  -0.284269   0.015746  -18.053
## factor(MostFreqOp)%                    0.068405   0.068074    1.005
## factor(MostFreqOp)+                   -0.062076   0.054791   -1.133
##                                      Pr(>|t|)
## (Intercept)                           < 2e-16 ***
## BaseCacheTypeAveEntExp                < 2e-16 ***
## log(NumTokens)                        < 2e-16 ***
## factor(ParentOp)<                      0.0118 *
## factor(ParentOp)<=                     0.1342
## factor(ParentOp)>                      0.0903 .
## factor(ParentOp)>=                     0.5009
## factor(ParentOp)|                      0.5397
## factor(ParentOp)||                     0.1307
## factor(ParentOp)-                     < 2e-16 ***
## factor(ParentOp)!=                     0.6398
## factor(ParentOp)/                     < 2e-16 ***
## factor(ParentOp)*                     < 2e-16 ***
## factor(ParentOp)&&                     0.2933
## factor(ParentOp)+                     < 2e-16 ***
## factor(ParentOp)ArrayAccess          3.43e-12 ***
## factor(ParentOp)ConditionalExpression  0.1484
## factor(ParentOp)MethodInvocation      < 2e-16 ***
## factor(MostFreqOp)<<                   0.2537
## factor(MostFreqOp)<=                  < 2e-16 ***
## factor(MostFreqOp)==                  < 2e-16 ***
## factor(MostFreqOp)>                    0.9078
## factor(MostFreqOp)>=                   0.5304
## factor(MostFreqOp)||                  < 2e-16 ***
## factor(MostFreqOp)-                   < 2e-16 ***
## factor(MostFreqOp)!=                  < 2e-16 ***
## factor(MostFreqOp)/                  9.25e-14 ***
## factor(MostFreqOp)*                  8.55e-14 ***
## factor(MostFreqOp)&&                  < 2e-16 ***
## factor(MostFreqOp)%                    0.3150
## factor(MostFreqOp)+                    0.2572
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4217 on 36527 degrees of freedom
## Multiple R-squared:  0.4963, Adjusted R-squared:  0.4959
## F-statistic:  1200 on 30 and 36527 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: CacheTypeAverageEntChangeExp
##                            Df Sum Sq Mean Sq  F value    Pr(>F)
## BaseCacheTypeAveEntExp      1 4924.4  4924.4 27687.87 < 2.2e-16 ***
## log(NumTokens)              1   58.7    58.7   329.98 < 2.2e-16 ***
## factor(ParentOp)           15  731.0    48.7   273.99 < 2.2e-16 ***
## factor(MostFreqOp)         13  687.6    52.9   297.38 < 2.2e-16 ***
## Residuals               36527 6496.5     0.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.381792955" "0.004550136" "0.056671021" "0.053308525" "0.503677364"
##                                GVIF Df GVIF^(1/(2*Df))
## BaseCacheTypeAveEntExp     1.176848  1        1.084826
## log(NumTokens)             1.162094  1        1.078005
## factor(ParentOp)        5601.211062 15        1.333346
## factor(MostFreqOp)      5718.624958 13        1.394798
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Feb 19, 2019 - 10:44:18 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & CacheTypeAverageEntChangeExp \\
## \hline \\[-1.8ex]
##  BaseCacheTypeAveEntExp & $-$0.438$^{***}$ (0.003) \\
##   log(NumTokens) & $-$0.158$^{***}$ (0.006) \\
##   factor(ParentOp)\textless  & $-$0.077$^{**}$ (0.031) \\
##   factor(ParentOp)\textless = & 0.097 (0.065) \\
##   factor(ParentOp)\textgreater  & 0.062$^{*}$ (0.036) \\
##   factor(ParentOp)\textgreater = & $-$0.034 (0.051) \\
##   factor(ParentOp)\textbar  & 0.118 (0.193) \\
##   factor(ParentOp)\textbar \textbar  & 0.092 (0.061) \\
##   factor(ParentOp)- & 0.333$^{***}$ (0.031) \\
##   factor(ParentOp)!= & $-$0.021 (0.044) \\
##   factor(ParentOp)/ & 0.553$^{***}$ (0.040) \\
##   factor(ParentOp)\textasteriskcentered  & 0.830$^{***}$ (0.063) \\
##   factor(ParentOp)&& & 0.064 (0.061) \\
##   factor(ParentOp)+ & 0.602$^{***}$ (0.029) \\
##   factor(ParentOp)ArrayAccess & 0.233$^{***}$ (0.034) \\
##   factor(ParentOp)ConditionalExpression & 0.084 (0.058) \\
##   factor(ParentOp)MethodInvocation & 0.555$^{***}$ (0.030) \\
##   factor(MostFreqOp)\textless \textless  & $-$0.223 (0.195) \\
##   factor(MostFreqOp)\textless = & 0.635$^{***}$ (0.026) \\
##   factor(MostFreqOp)== & $-$0.239$^{***}$ (0.011) \\
```

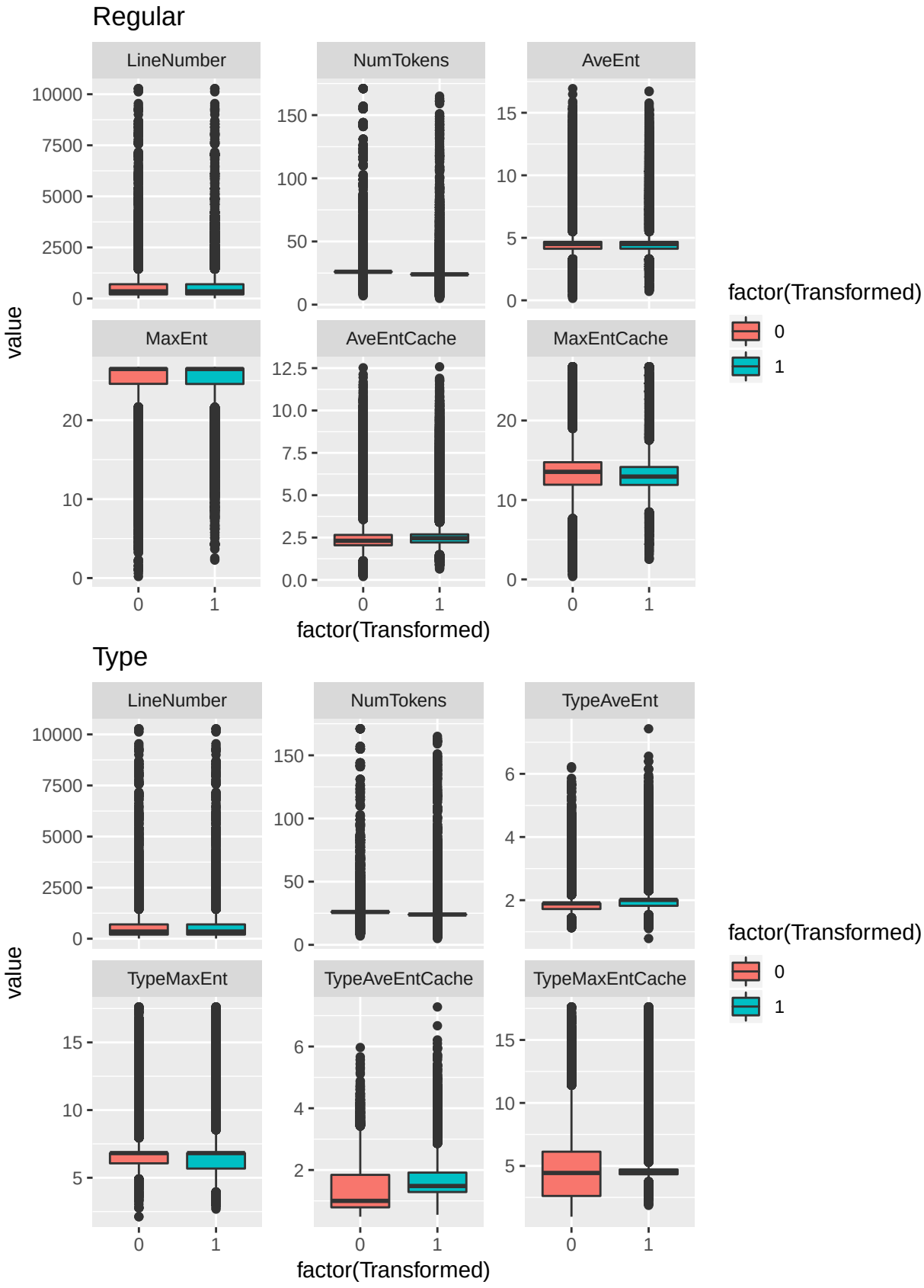
```
##   factor(MostFreqOp)\textgreater  & $-$0.002 (0.014) \\
##   factor(MostFreqOp)\textgreater = & $-$0.010 (0.016) \\
##   factor(MostFreqOp)\textbar \textbar  & $-$0.276$^{***}$ (0.024) \\
##   factor(MostFreqOp)- & $-$0.500$^{***}$ (0.056) \\
##   factor(MostFreqOp)!= & $-$0.164$^{***}$ (0.011) \\
##   factor(MostFreqOp)/ & $-$0.428$^{***}$ (0.057) \\
##   factor(MostFreqOp)\textasteriskcentered  & $-$0.418$^{***}$ (0.056) \\
##   factor(MostFreqOp)&& & $-$0.284$^{***}$ (0.016) \\
##   factor(MostFreqOp)\% & 0.068 (0.068) \\
##   factor(MostFreqOp)+ & $-$0.062 (0.055) \\
##   Constant & 1.579$^{***}$ (0.063) \\
##  \hline \\[-1.8ex]
## Observations & 36,558 \\
## R$^{2}$ & 0.496 \\
## Adjusted R$^{2}$ & 0.496 \\
## Residual Std. Error & 0.422 (df = 36527) \\
## F Statistic & 1,199.788$^{***}$ (df = 30; 36527) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:44:18 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## BaseCacheTypeAveEntExp & 1 & 4924.42 & 4924.42 & 27687.87 & 0.0000 \\
##   log(NumTokens) & 1 & 58.69 & 58.69 & 329.98 & 0.0000 \\
##   factor(ParentOp) & 15 & 730.95 & 48.73 & 273.99 & 0.0000 \\
##   factor(MostFreqOp) & 13 & 687.58 & 52.89 & 297.38 & 0.0000 \\
##   Residuals & 36527 & 6496.51 & 0.18 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

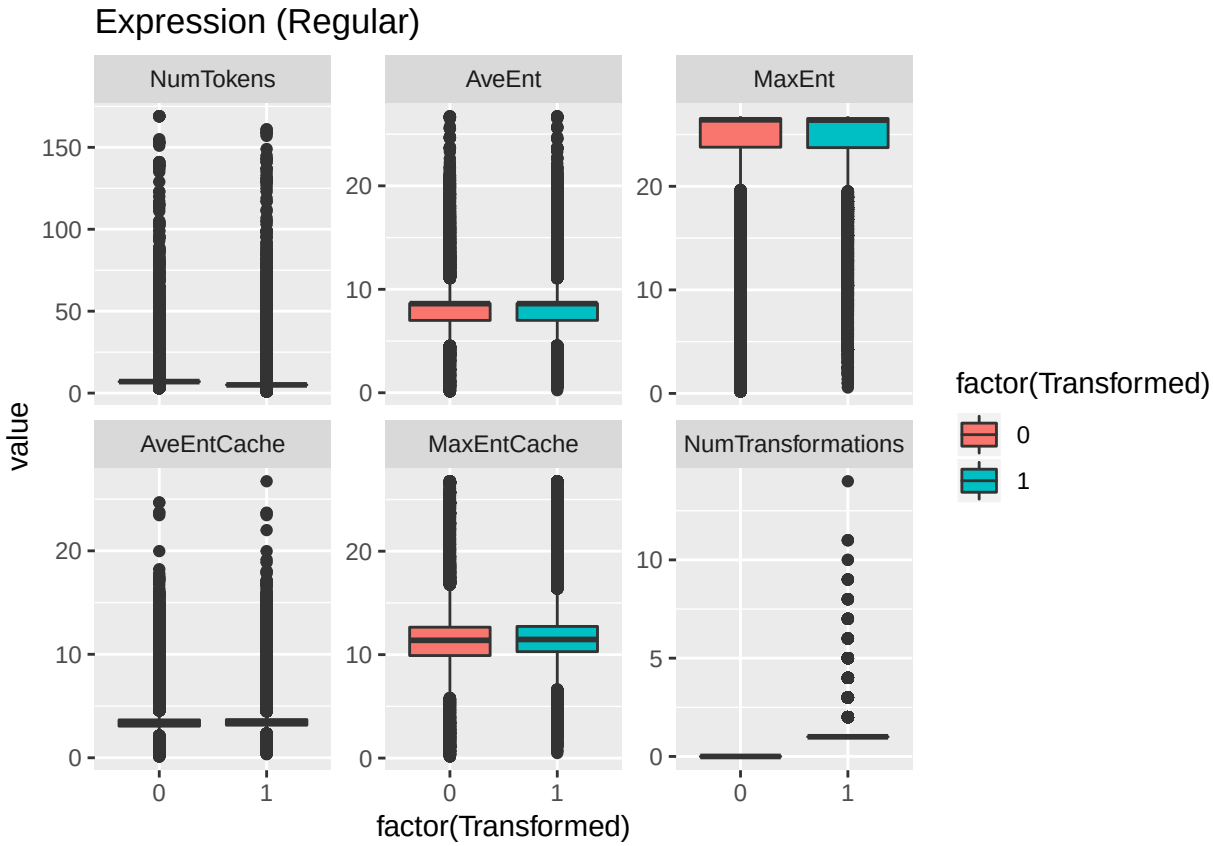
Parenthesis removing results

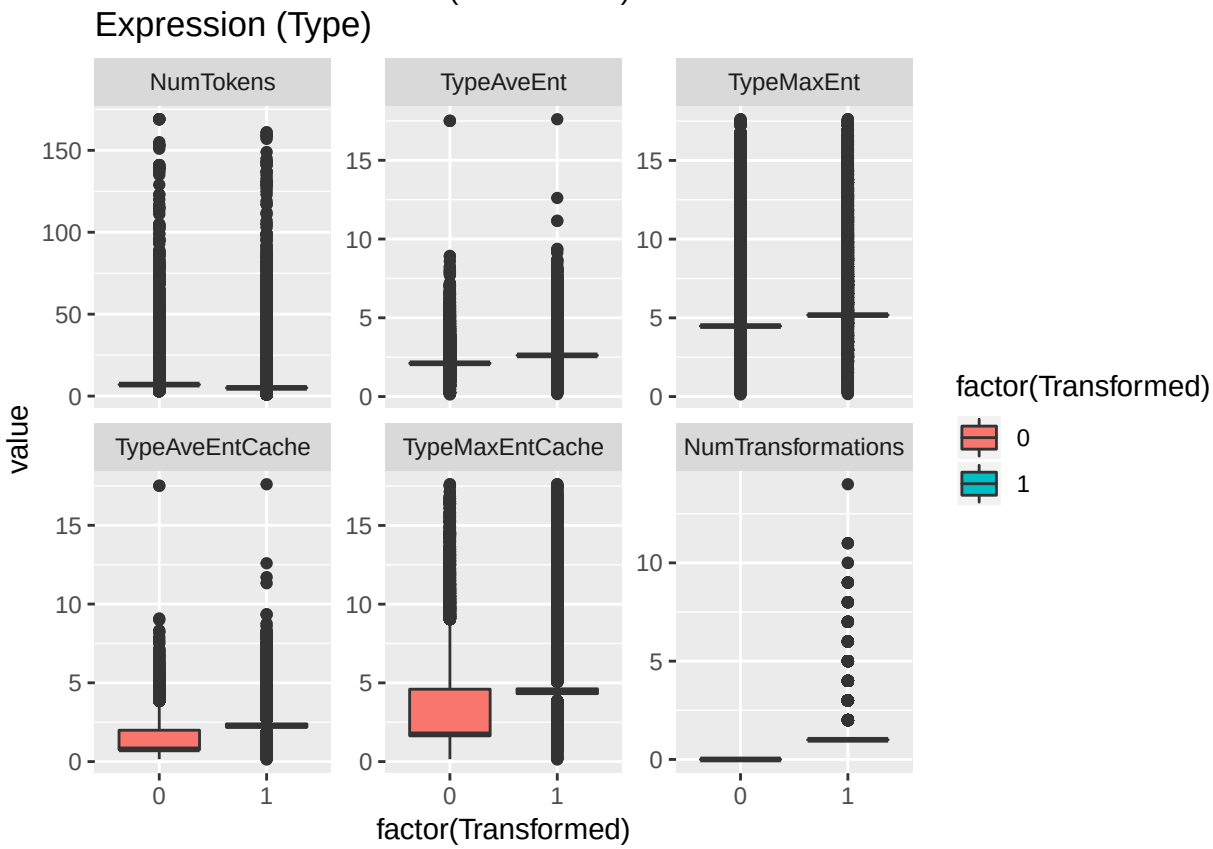
```
setwd("/data/anon/SemanticTransformation/")
#drm <- compareDepthSummary("remove_parenthesis.csv", "RmParen", "REMOVE")
drm <- compareDepthSummary("remove_parenthesis_topstarred.csv", "RmParenTop", "REMOVE")
```

```
## [[1]]
## [1] "5"    "Bool"
##
##  [1] "TransId"
##  [2] "Filepath"
##  [3] "LineNumber"
##  [4] "NumTokens"
##  [5] "Transformed"
##  [6] "Source"
##  [7] "CleanLexerNumTokens"
```

```
##  [8] "CleanLexerSource"
##  [9] "AveEnt"
## [10] "MaxEnt"
## [11] "AveEntCache"
## [12] "MaxEntCache"
## [13] "AveEntRev"
## [14] "MaxEntRev"
## [15] "TypeSource"
## [16] "TypeNumTokens"
## [17] "TypeAveEnt"
## [18] "TypeMaxEnt"
## [19] "TypeAveEntCache"
## [20] "TypeMaxEntCache"
## [21] "Depth"
## [22] "Expression"
## [23] "ExpressionNumTokens"
## [24] "ExpressionCleanLexerSource"
## [25] "ExpressionCleanLexerNumTokens"
## [26] "ExpressionForwardAverageEntropy"
## [27] "ExpressionForwardMaxEntropy"
## [28] "ExpressionForwardAverageEntropyCache"
## [29] "ExpressionForwardMaxEntropyCache"
## [30] "TypeExpression"
## [31] "ExpressionTypeNumTokens"
## [32] "ExpressionTypeAverageEntropy"
## [33] "ExpressionTypeMaxEntropy"
## [34] "ExpressionTypeAverageEntropyCache"
## [35] "ExpressionTypeMaxEntropyCache"
## [36] "NumTransformations"
## [37] "ParentOp"
## [38] "MostFreqOp"
## [39] "LeastFreqOp"
## [40] "MostFreqParentOp"
## [41] "ParentChildFreq"
## [42] "ParentParenChildFreq"
## [43] "PoolSize"
## [44] "TransSetNo"
## [45] "TransNo"
## [46] "Type"
## [47] "NumTypes"
## [48] "MethodName"
## [49] "rowID"
```

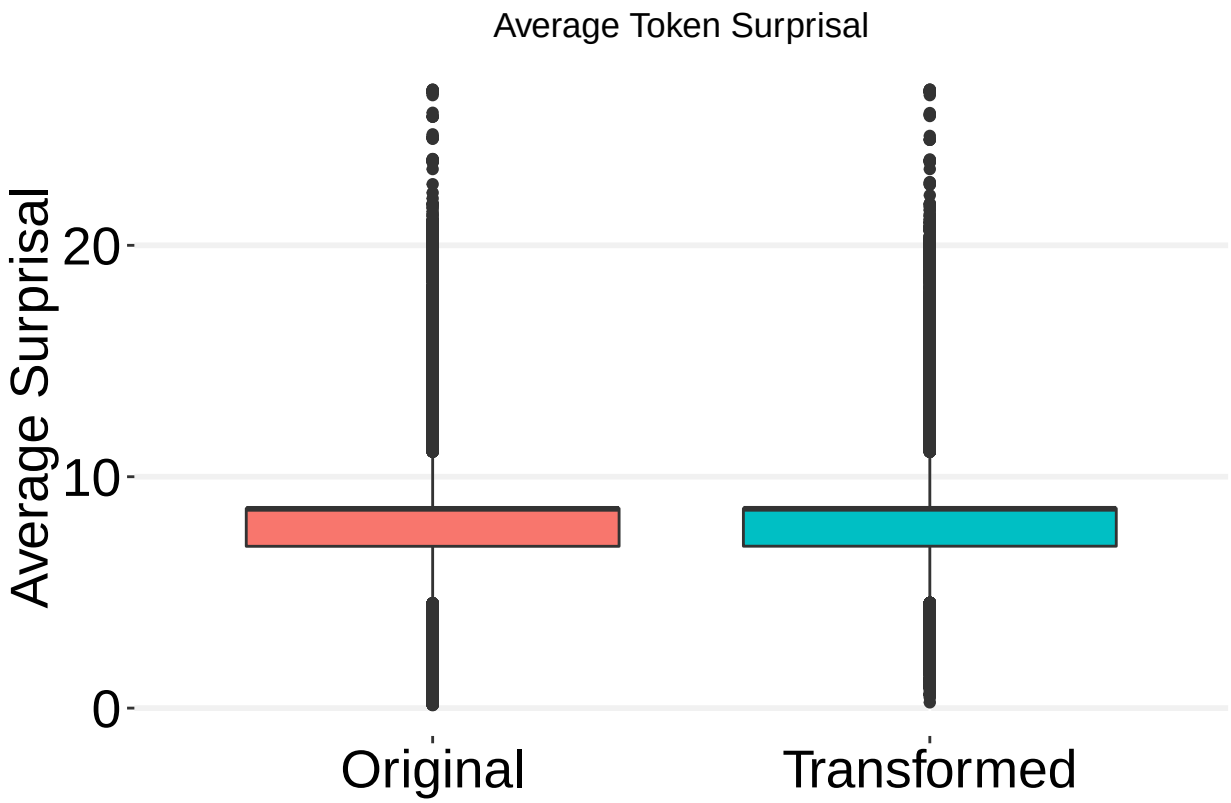

Regular
LineNumber
NumTokens
AveEnt
MaxEnt
AveEntCache
MaxEntCache
value
factor(Transformed)
factor(Transformed)
0
1
Type
LineNumber
NumTokens
TypeAveEnt
TypeMaxEnt
TypeAveEntCache
TypeMaxEntCache
value
factor(Transformed)
factor(Transformed)
0
1





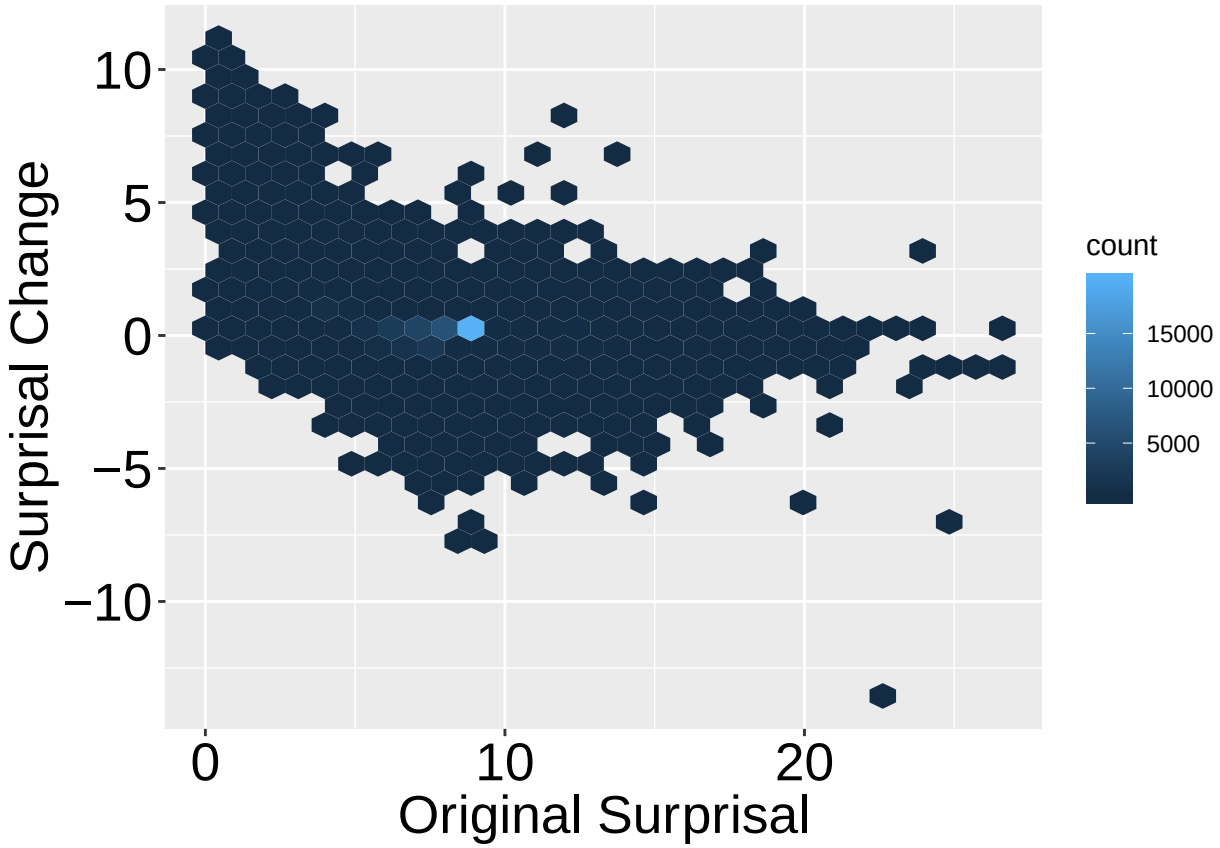

```
## [1] " ------ Expression Global Model ----- "
```

```
## [1] "RmParenTopGlobalExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -18.695, df = 48111, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##         -Inf -0.05623817
## sample estimates:
## mean of the differences
##             -0.06652413
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -18.695, df = 48111, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.07756226 -0.05548600
## sample estimates:
## mean of the differences
##             -0.06652413
## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -0.08523214 (negligible)
## 95 percent confidence interval:
## inf sup
##  NA  NA
## Warning in n1 * n2: NAs produced by integer overflow

## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 30998000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##         -Inf -0.09766848
## sample estimates:
## (pseudo)median
##     -0.1258509
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 30998000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.15020199 -0.09536597
## sample estimates:
## (pseudo)median
##     -0.1258509
##
##
## Cliff's Delta
##
## delta estimate: -0.006183817 (negligible)
```

```
## 95 percent confidence interval:
##          inf          sup
## -0.013479798  0.001112822
```




```
## [1] " ------ Expression Cache Model ----- "
## [1] "RmParenTopCacheExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -40.037, df = 48111, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.1342964
## sample estimates:
## mean of the differences
##              -0.1447471
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -40.037, df = 48111, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.1559619 -0.1335322
## sample estimates:
```

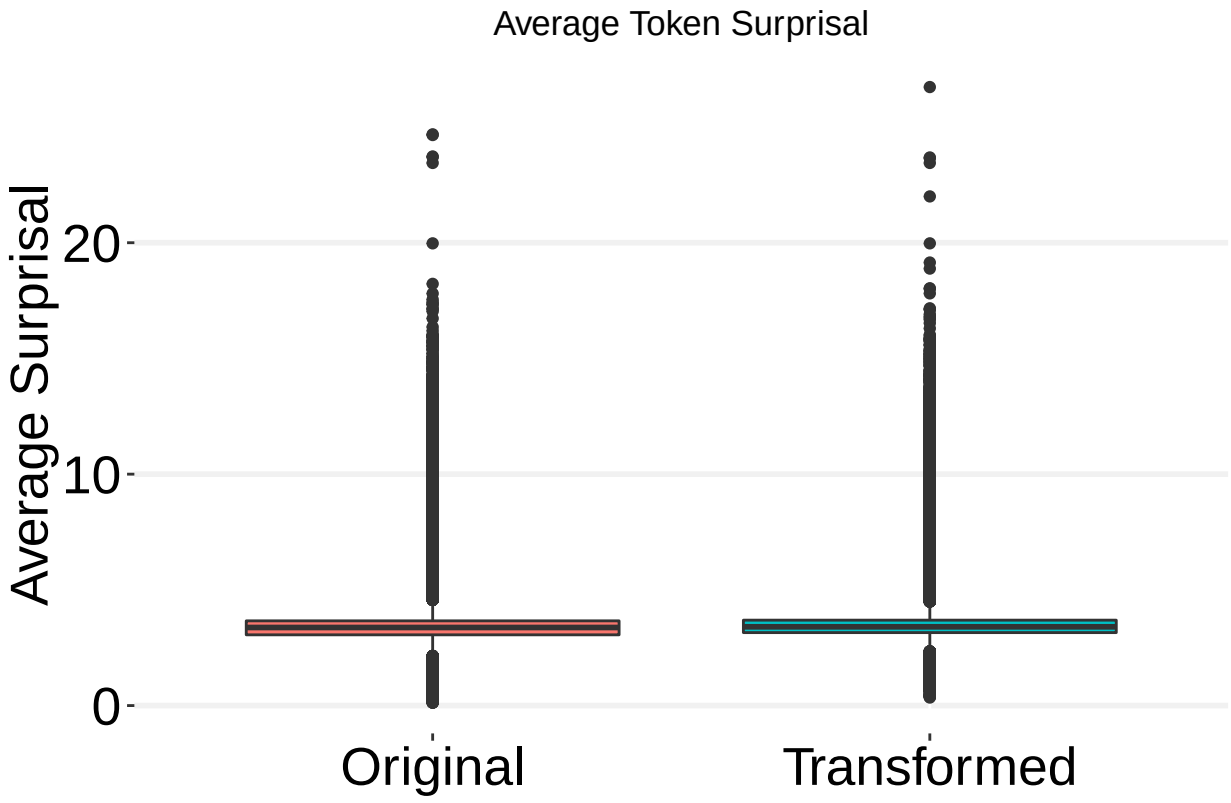
```
## mean of the differences
##              -0.1447471
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.1825307 (negligible)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

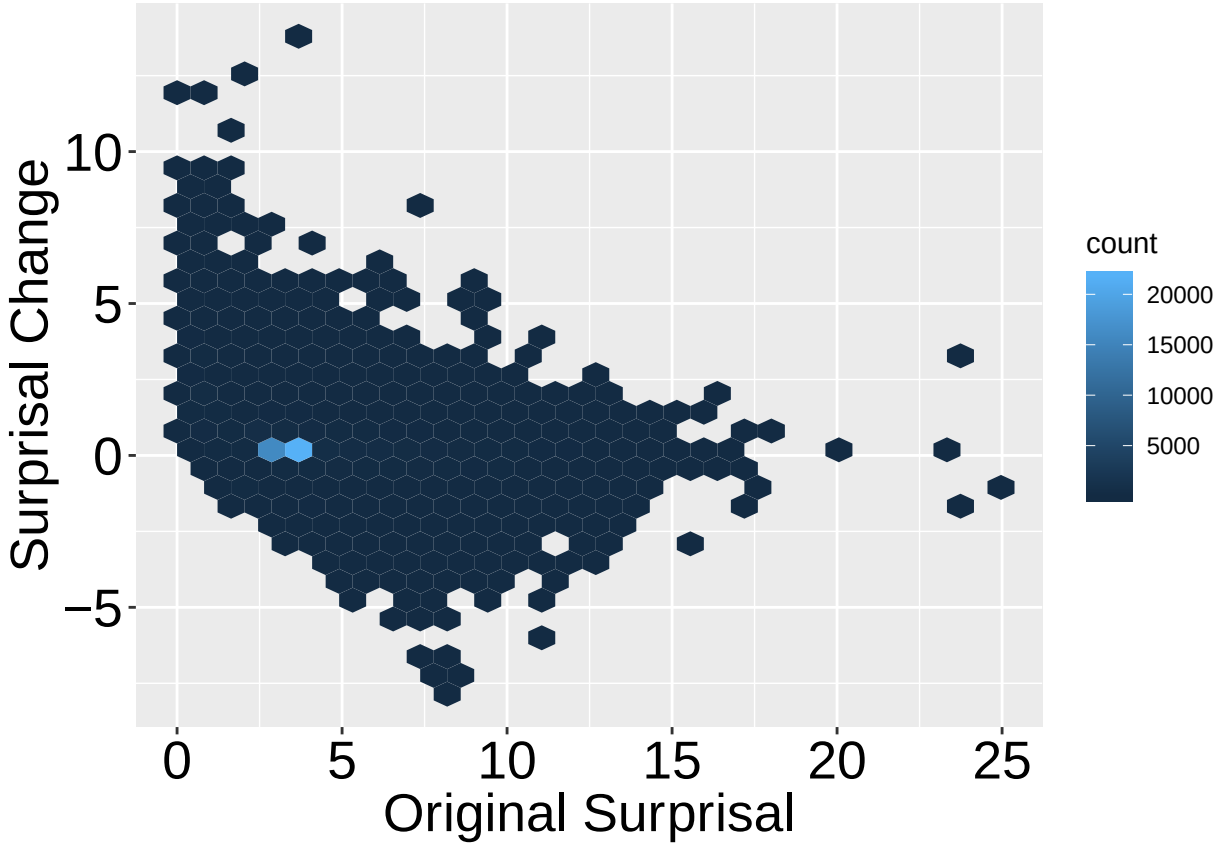
```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 778620000, p-value = 1
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##          -Inf 0.0004902416
## sample estimates:
## (pseudo)median
##   0.0004725755
##
##
```

```
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 778620000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  0.0004517102 0.0004749438
## sample estimates:
## (pseudo)median
##   0.0004725755
##
##
## Cliff's Delta
##
## delta estimate: -0.05107929 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.05835986 -0.04379329
```




```
## [1] " ------ Expression Global Type Model ----- "
## [1] "RmParenTopGlobalTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -207.88, df = 48111, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
```

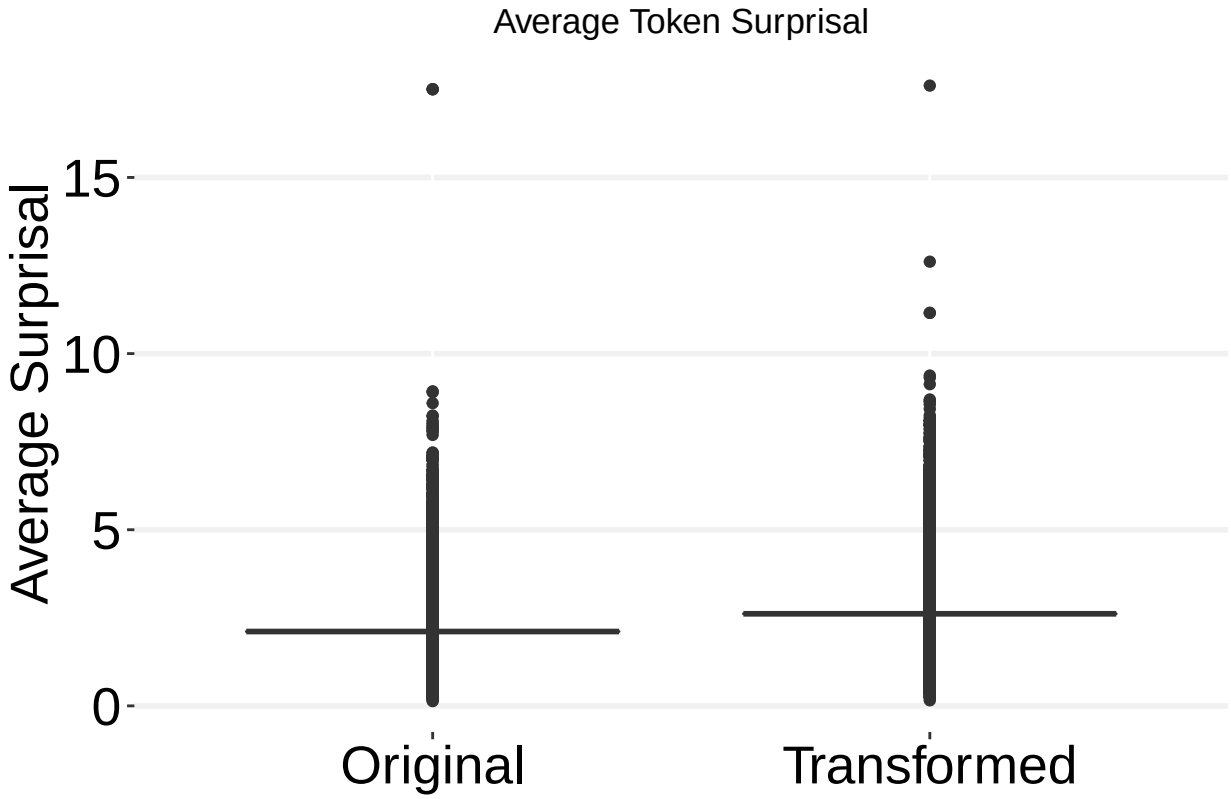
```
## 99.80769 percent confidence interval:
##        -Inf -0.4624819
## sample estimates:
## mean of the differences
##              -0.4690034
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -207.88, df = 48111, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.4760019 -0.4620050
## sample estimates:
## mean of the differences
##              -0.4690034
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -0.9477515 (large)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

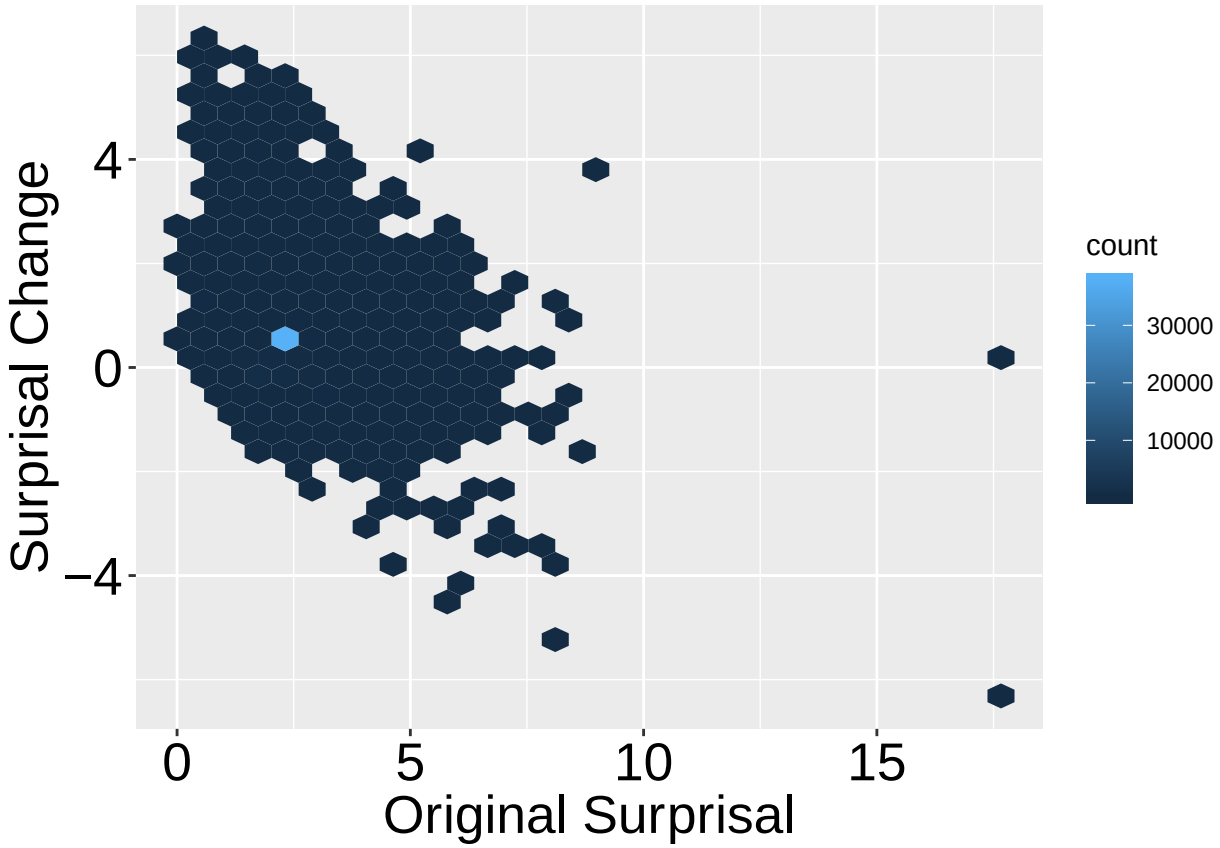
```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 61231000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.5032838
## sample estimates:
## (pseudo)median
##     -0.5032564
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 61231000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.5032592 -0.5032750
## sample estimates:
## (pseudo)median
##     -0.5032564
##
##
## Cliff's Delta
##
## delta estimate: -0.7222819 (large)
```

```
## 95 percent confidence interval:
##        inf        sup
## -0.7281009 -0.7163588
```




```
## [1] " ------ Expression Cache Type Model ----- "
## [1] "RmParenTopCacheTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -316.4, df = 48111, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##       -Inf -1.014165
## sample estimates:
## mean of the differences
##               -1.023516
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -316.4, df = 48111, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -1.033551 -1.013482
## sample estimates:
```

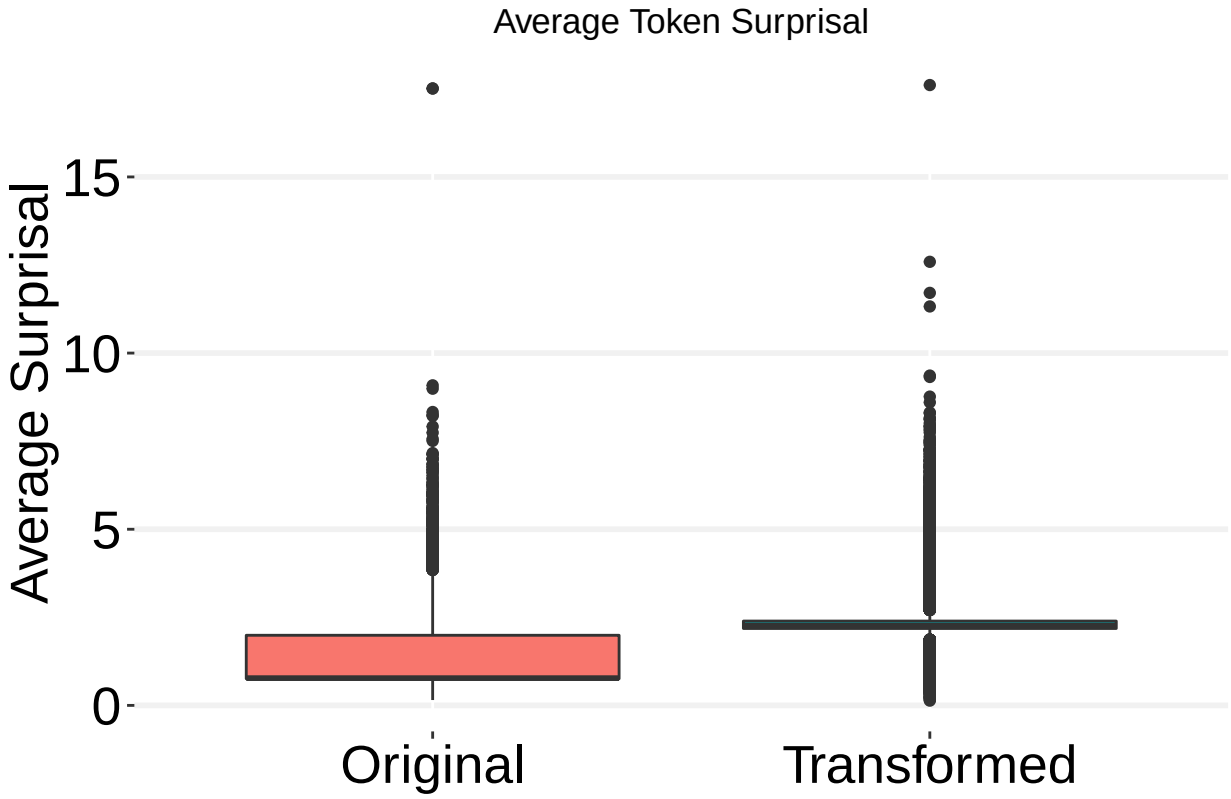
```
## mean of the differences
##               -1.023516
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
##
## Cohen's d
##
## d estimate: -1.442461 (large)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```

```
## Warning in n1 * n2: NAs produced by integer overflow
```



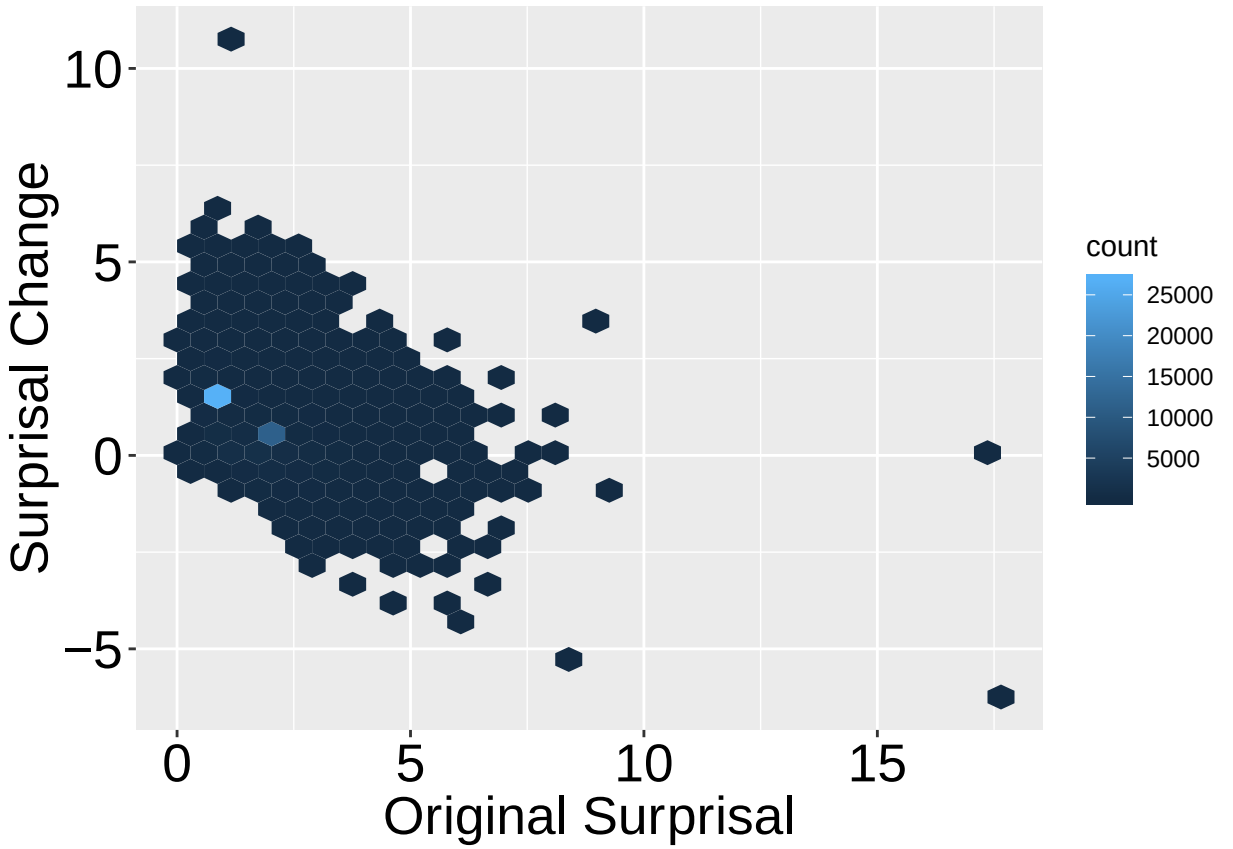

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 26615000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -0.95392
## sample estimates:
## (pseudo)median
##      -0.955528
##
##
```

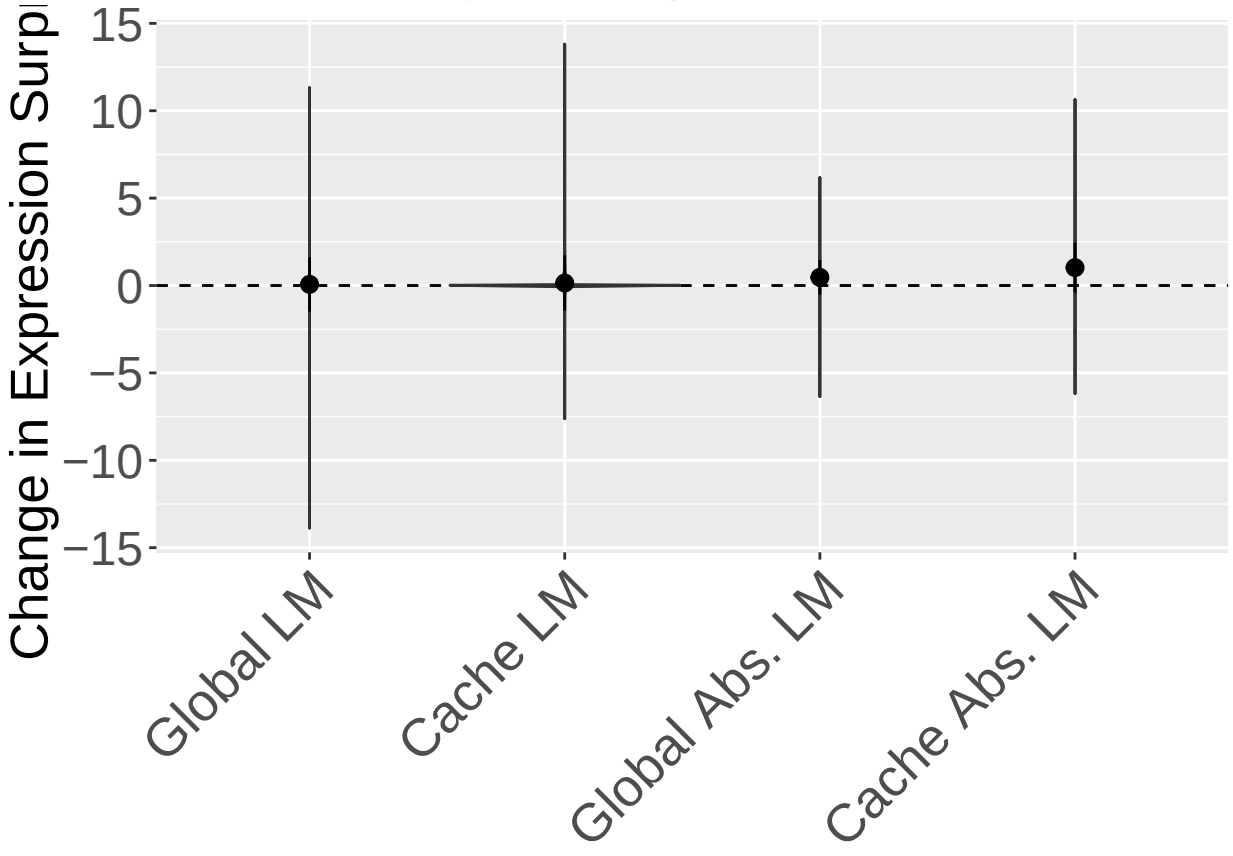
```
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 26615000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.9579815 -0.9537087
## sample estimates:
## (pseudo)median
##      -0.955528
##
##
## Cliff's Delta
##
## delta estimate: -0.8226952 (large)
## 95 percent confidence interval:
##        inf        sup
## -0.8269347 -0.8183621
```

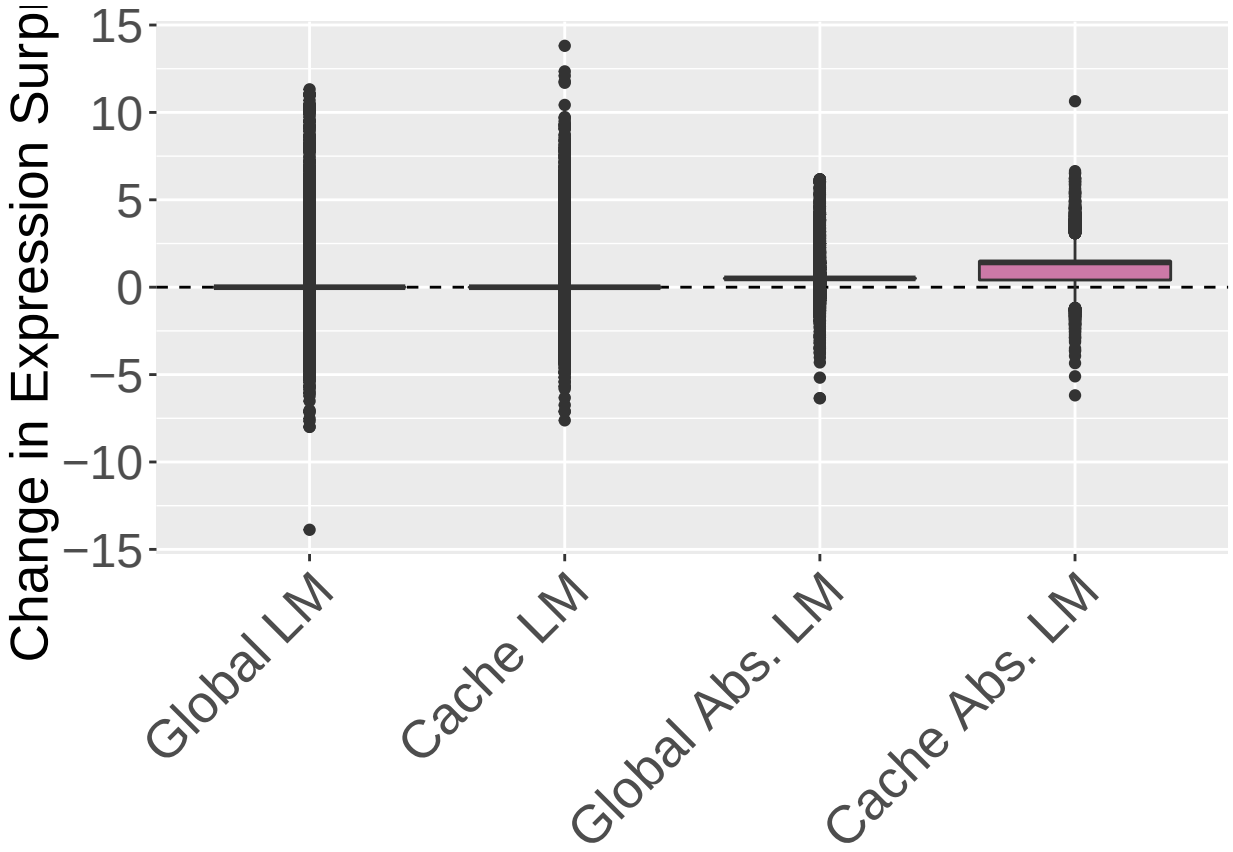
```
## [1] "Binary differences"
##
## FALSE  TRUE
##  5956 42156
##
## FALSE  TRUE
## 39118  8994
##
## FALSE  TRUE
##  4541 43571
##
## FALSE  TRUE
##  3433 44679
```

```
## No id variables; using all as measure variables
```

```
## Warning: Ignoring unknown parameters: mult
```


Surprisal Change
10
5
0
−5
0
5
10
15
Original Surprisal
count
25000
20000
15000
10000
5000

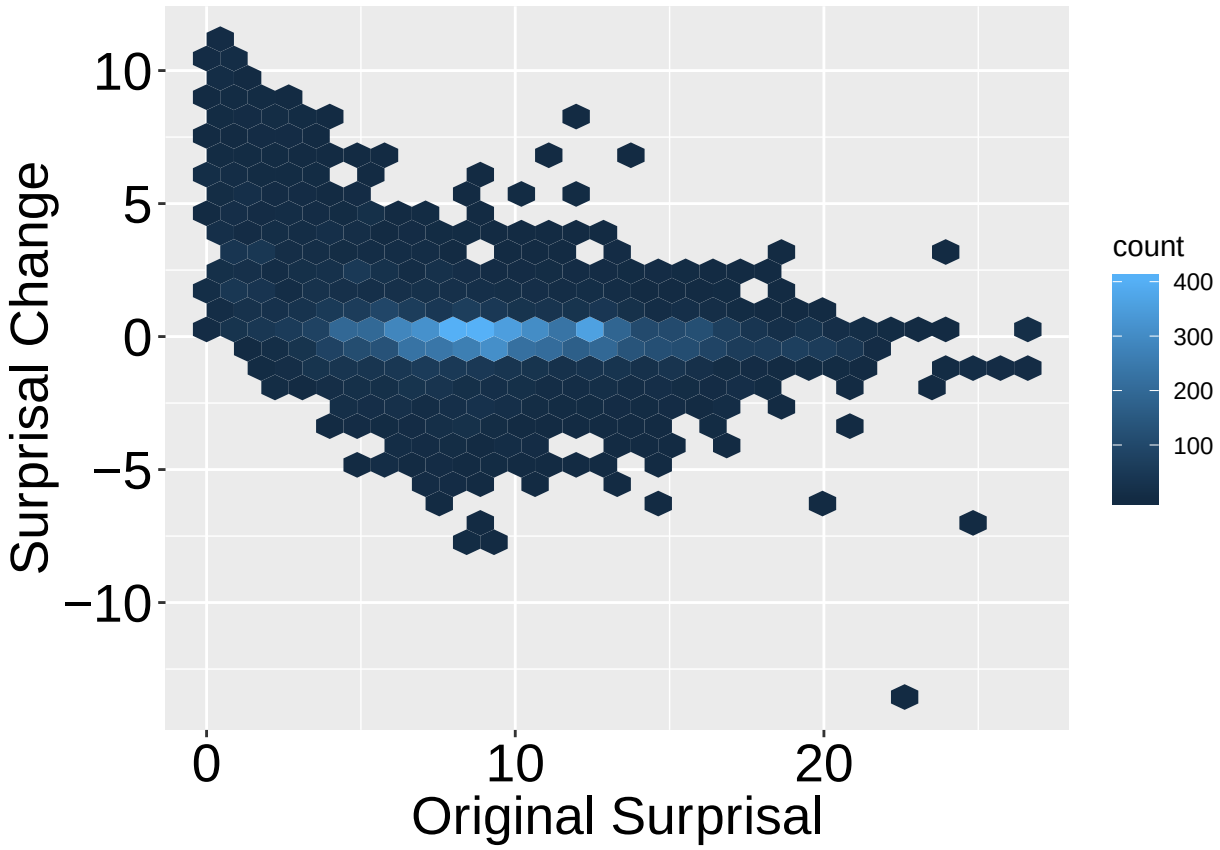

Change in Expression Surpl
15
10
5
0
−5
−10
−15
Global LM
Cache LM
Global Abs. LM
Cache Abs. LM




```
setwd("/data/anon/SemanticTransformation/")

#Plot results on everything (for paired table)
#dsp <- compareDepthSummary("swap.csv", "Swap", "SAME", FALSE)
#Plot for the Large dataset
#Give some sense of robustness to n for these values...
#Also remove multiple transform lines?
drmFiltered <- generateFilteredResults(drm, "Paren", "RmParenTopFiltered100", 100)
```

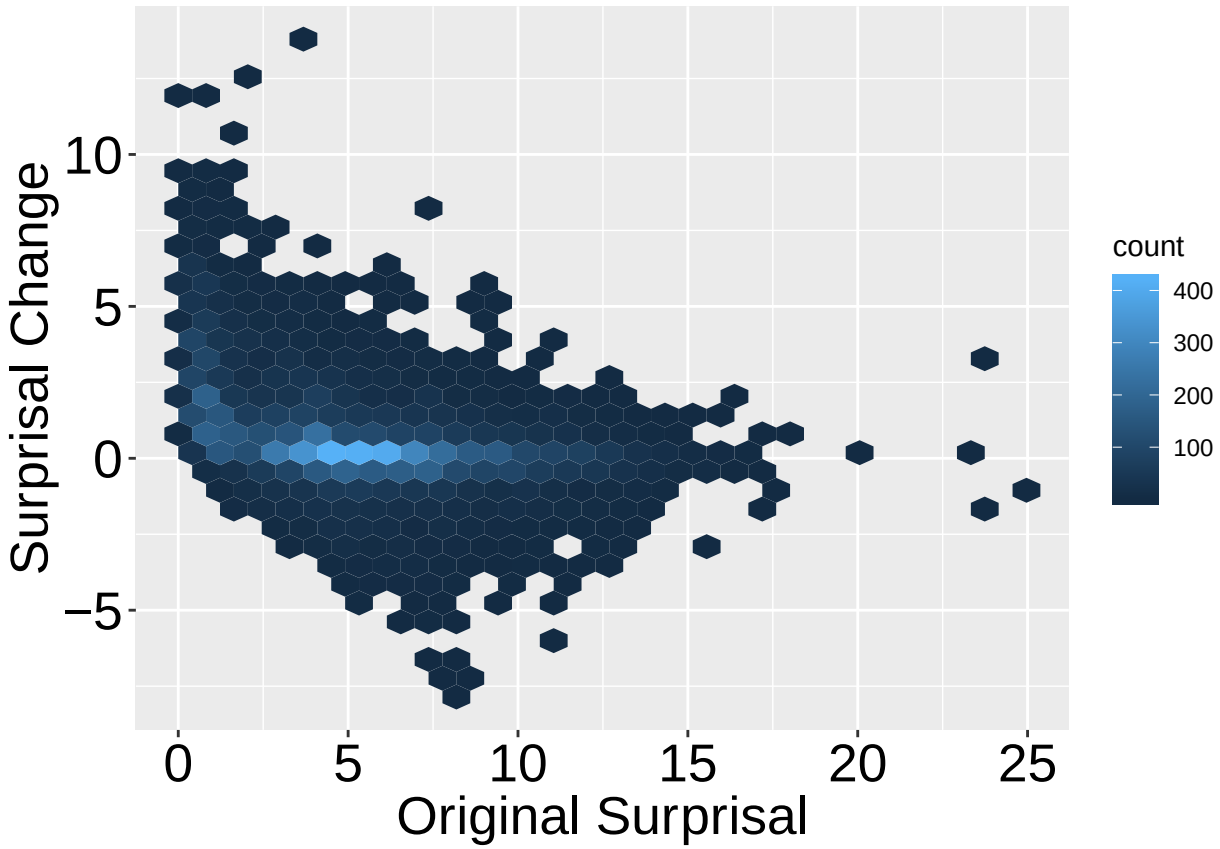
```
## [1] " ------ Expression Global Model ----- "
## [1] "RmParenTopFiltered100GlobalExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -9.5914, df = 9716, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.1039859
## sample estimates:
## mean of the differences
##              -0.1488566
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
```

```
## t = -9.5914, df = 9716, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.1970099 -0.1007034
## sample estimates:
## mean of the differences
##              -0.1488566
##
##
## Cohen's d
##
## d estimate: -0.09730113 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.12543833 -0.06916393
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 20338000, p-value = 0.9586
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf 0.03265256
## sample estimates:
## (pseudo)median
##     0.01227168
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 20338000, p-value = 0.0828
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.009978882  0.034086419
## sample estimates:
## (pseudo)median
##     0.01227168
##
##
## Cliff's Delta
##
## delta estimate: -0.016188 (negligible)
## 95 percent confidence interval:
##           inf           sup
## -3.241956e-02  5.210688e-05
```



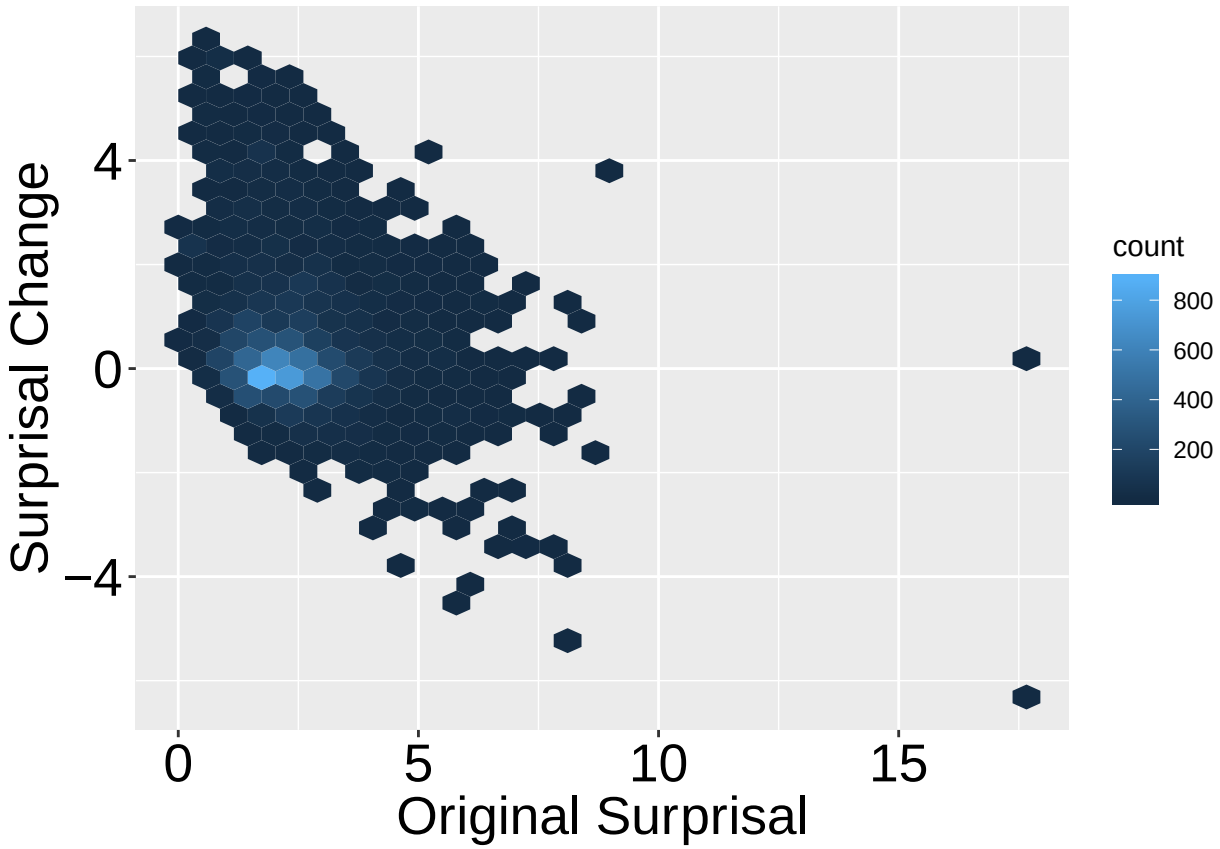

```
## [1] " ------ Expression Cache Model ----- "
## [1] "RmParenTopFiltered100CacheExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -37.744, df = 9716, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.5753082
## sample estimates:
## mean of the differences
##              -0.6230331
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -37.744, df = 9716, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.6742491 -0.5718170
## sample estimates:
## mean of the differences
##              -0.6230331
##
##
```

```
## Cohen's d
##
## d estimate: -0.3828952 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.4112723 -0.3545182
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 13017000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##       -Inf -0.363969
## sample estimates:
## (pseudo)median
##     -0.3990231
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 13017000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.4382826 -0.3614667
## sample estimates:
## (pseudo)median
##     -0.3990231
##
##
## Cliff's Delta
##
## delta estimate: -0.1220144 (negligible)
## 95 percent confidence interval:
##        inf        sup
## -0.1381101 -0.1058542
```



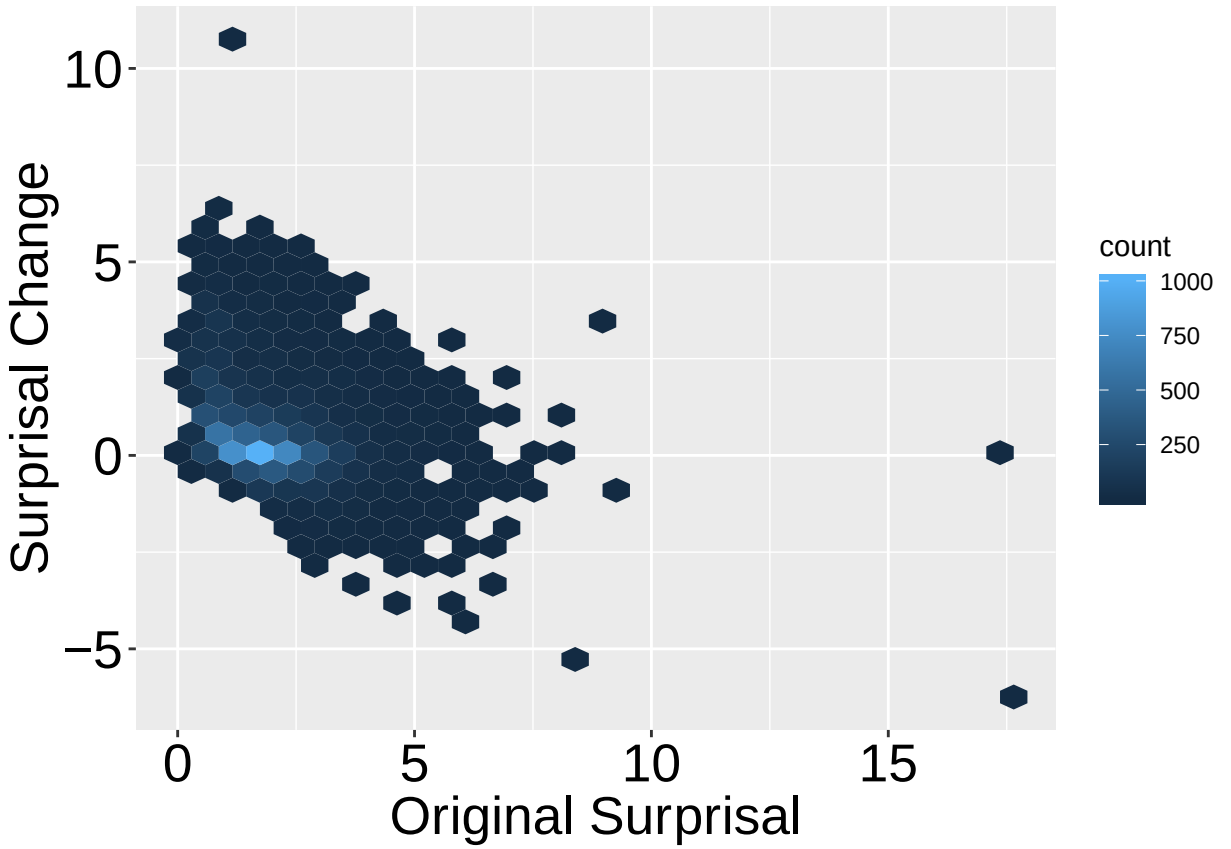

```
## [1] " ------ Expression Global Type Model ----- "
## [1] "RmParenTopFiltered100GlobalTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -32.498, df = 9716, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3214746
## sample estimates:
## mean of the differences
##               -0.352868
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -32.498, df = 9716, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.3865578 -0.3191781
## sample estimates:
## mean of the differences
##               -0.352868
##
##
```

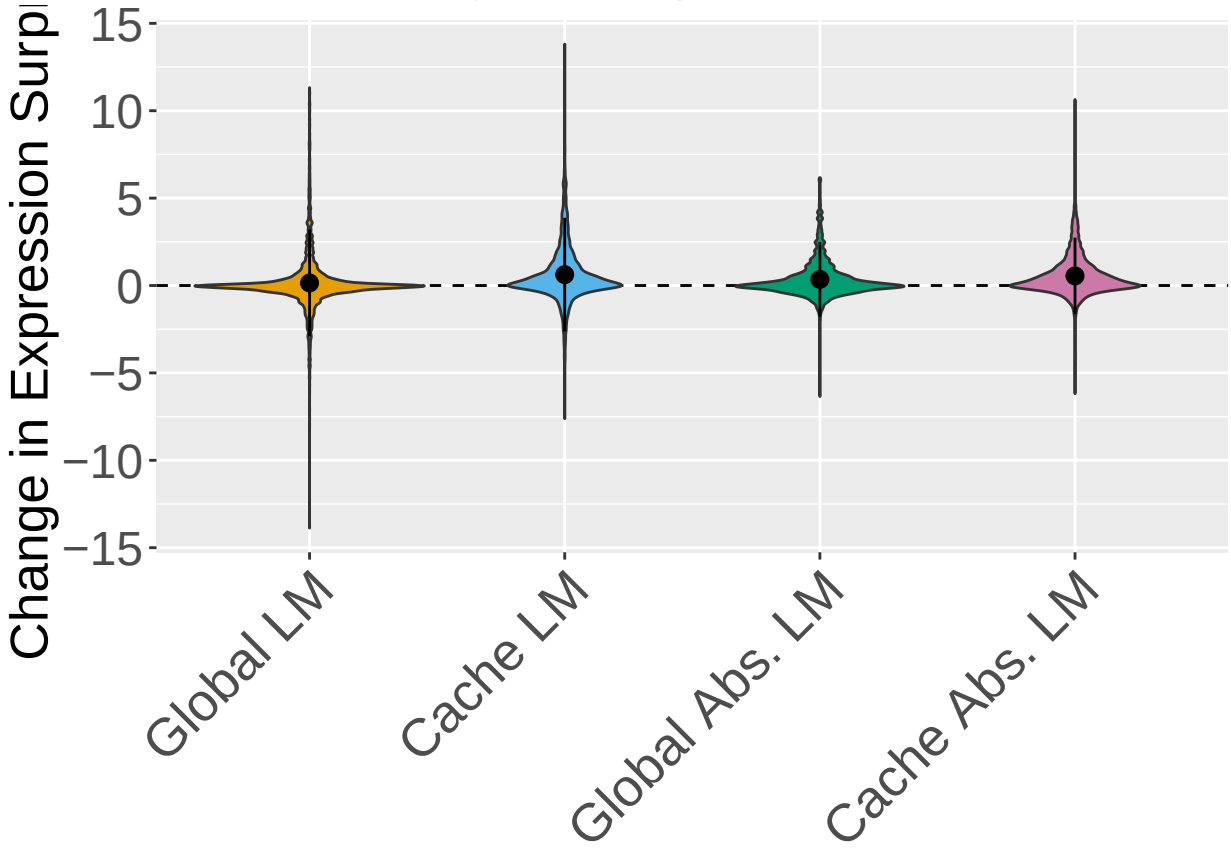
```
## Cohen's d
##
## d estimate: -0.3296764 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.3579873 -0.3013655
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 16746000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.1517208
## sample estimates:
## (pseudo)median
##     -0.1746742
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 16746000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.2000501 -0.1500525
## sample estimates:
## (pseudo)median
##     -0.1746742
##
##
## Cliff's Delta
##
## delta estimate: -0.136802 (negligible)
## 95 percent confidence interval:
##        inf        sup
## -0.1528274 -0.1207046
```

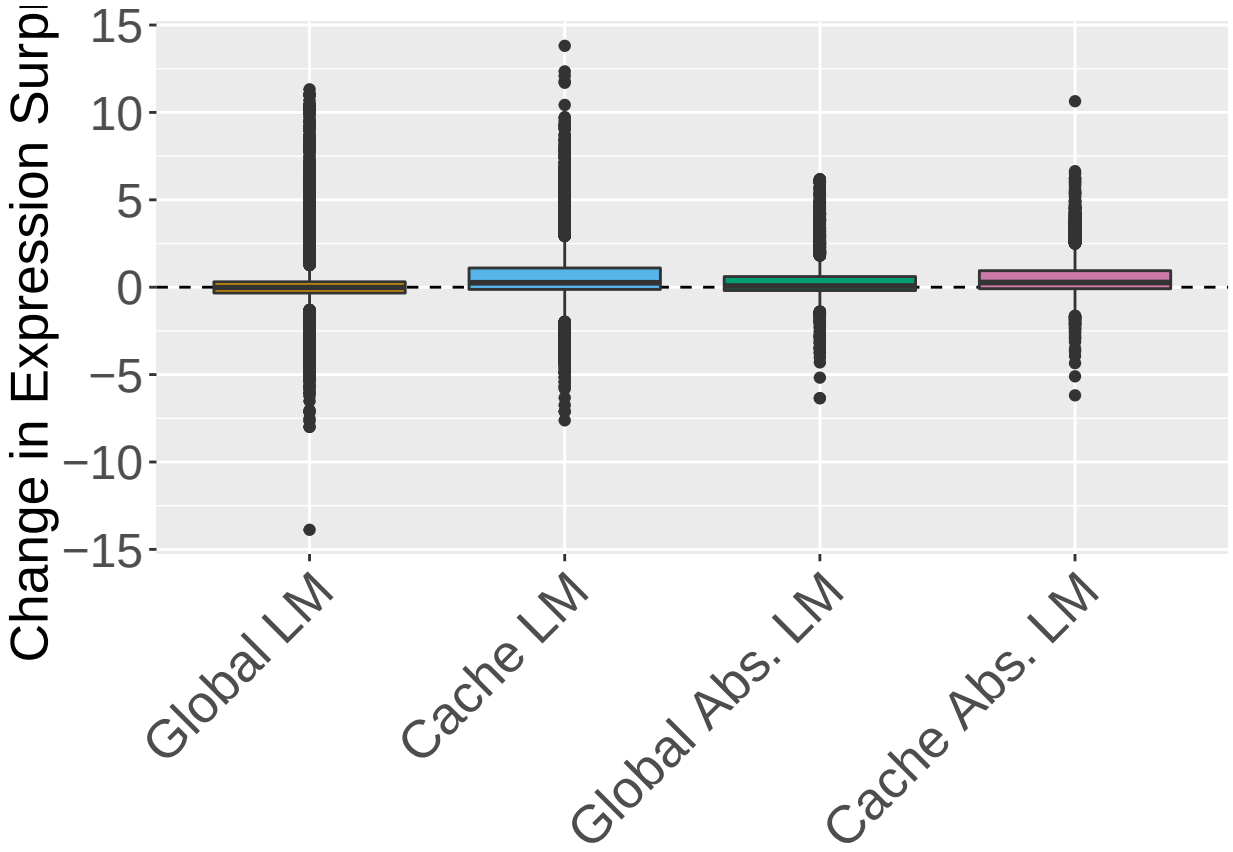



```
## [1] " ------ Expression Cache Type Model ----- "
## [1] "RmParenTopFiltered100CacheTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -49.679, df = 9716, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.5200705
## sample estimates:
## mean of the differences
##              -0.5522079
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -49.679, df = 9716, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.5866963 -0.5177195
## sample estimates:
## mean of the differences
##              -0.5522079
##
##
```

```
## Cohen's d
##
## d estimate: -0.5039701 (medium)
## 95 percent confidence interval:
##        inf        sup
## -0.5325335 -0.4754066
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10801000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3652943
## sample estimates:
## (pseudo)median
##     -0.3925125
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10801000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.4224886 -0.3633293
## sample estimates:
## (pseudo)median
##     -0.3925125
##
##
## Cliff's Delta
##
## delta estimate: -0.2770034 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.2923105 -0.2615544

## No id variables; using all as measure variables

## Warning: Ignoring unknown parameters: mult
```


Surprisal Change
10
5
0
−5
0
5
10
15
Original Surprisal
count
1000
750
500
250

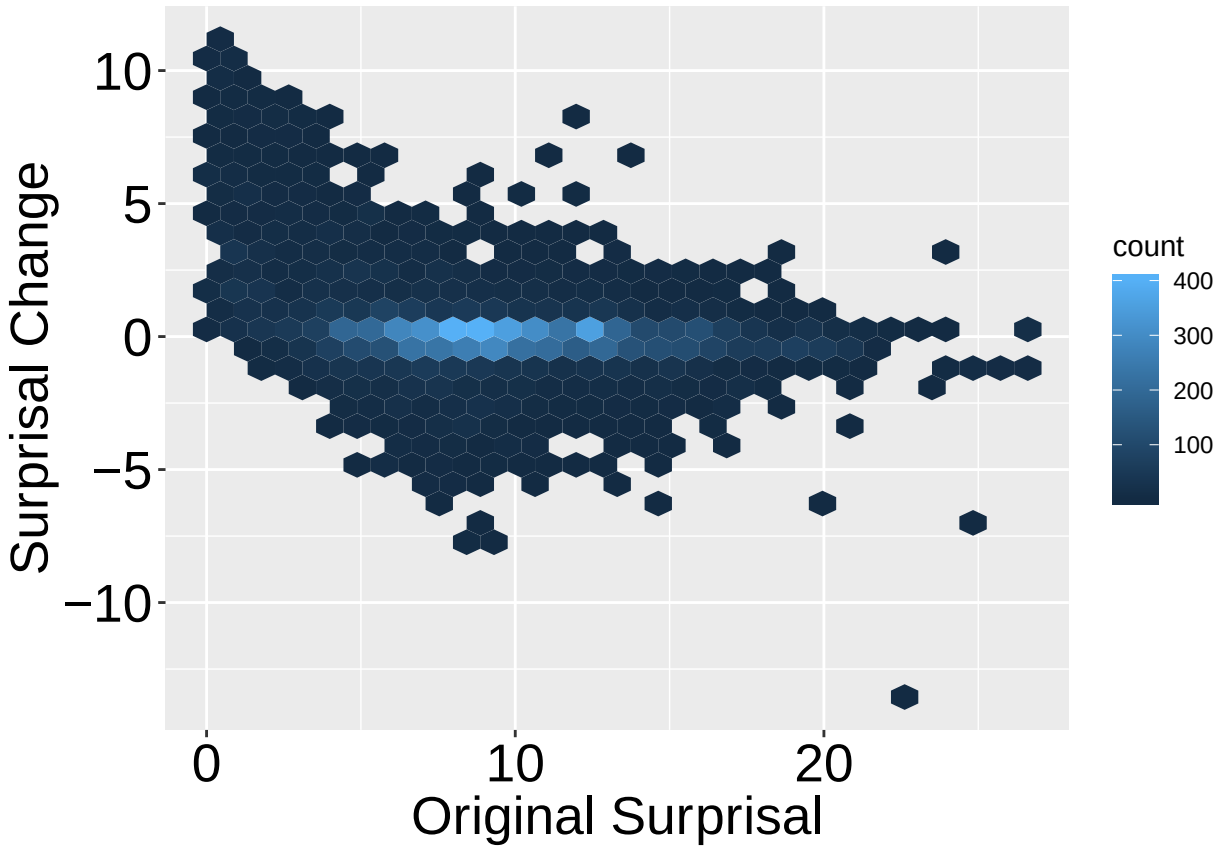

Change in Expression Surpl
15
10
5
0
−5
−10
−15
Global LM
Cache LM
Global Abs. LM
Cache Abs. LM




```
drmFiltered2 <- generateFilteredResults(drm, "Paren", "RmParenTopFiltered10", 10)
```

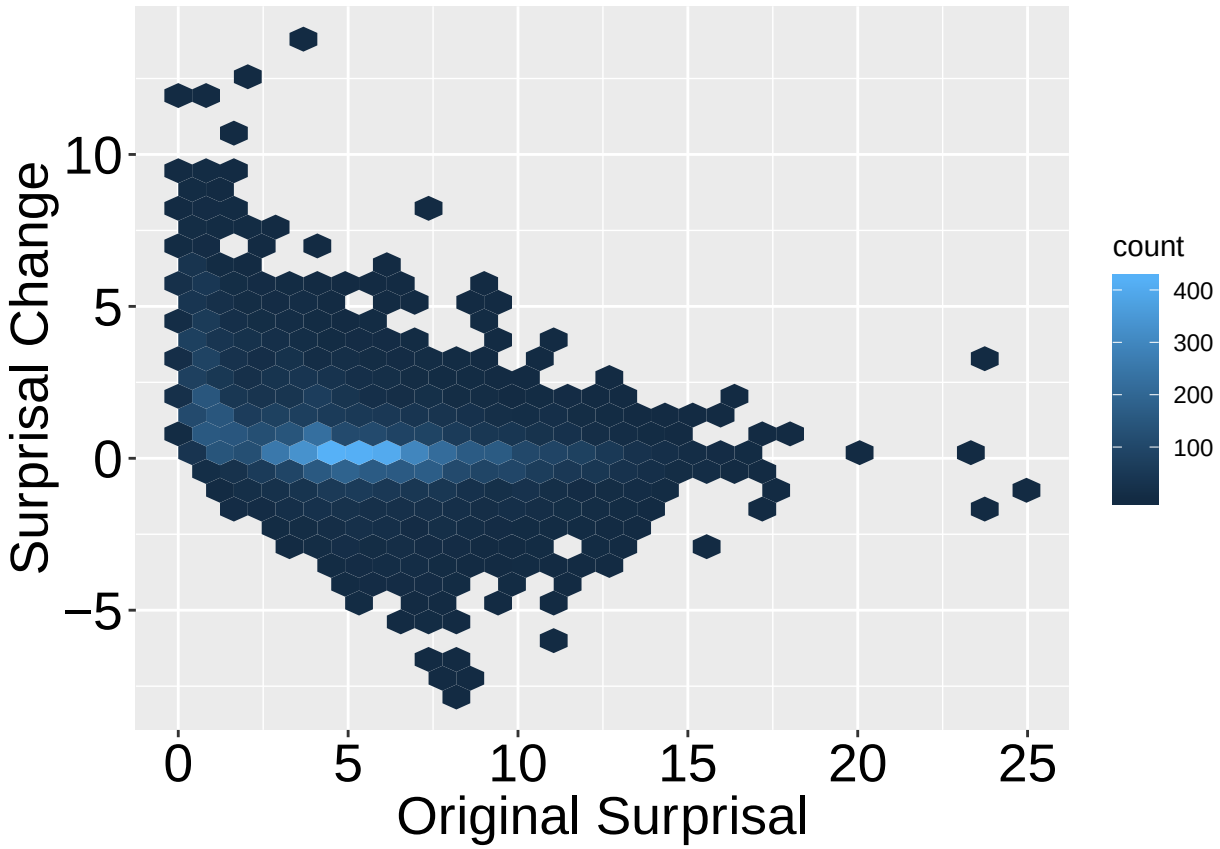
```
## [1] " ------ Expression Global Model ----- "
## [1] "RmParenTopFiltered10GlobalExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -8.7555, df = 9456, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##         -Inf -0.09179208
## sample estimates:
## mean of the differences
##              -0.1370478
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -8.7555, df = 9456, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.18561424 -0.08848141
## sample estimates:
## mean of the differences
##              -0.1370478
```

```
##
##
## Cohen's d
##
## d estimate: -0.09003344 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.11855243 -0.06151445
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 19390000, p-value = 0.996
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf 0.03801913
## sample estimates:
## (pseudo)median
##     0.01836747
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 19390000, p-value = 0.00802
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.003087245  0.039381689
## sample estimates:
## (pseudo)median
##     0.01836747
##
##
## Cliff's Delta
##
## delta estimate: -0.01446492 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.03091801  0.00199601
```



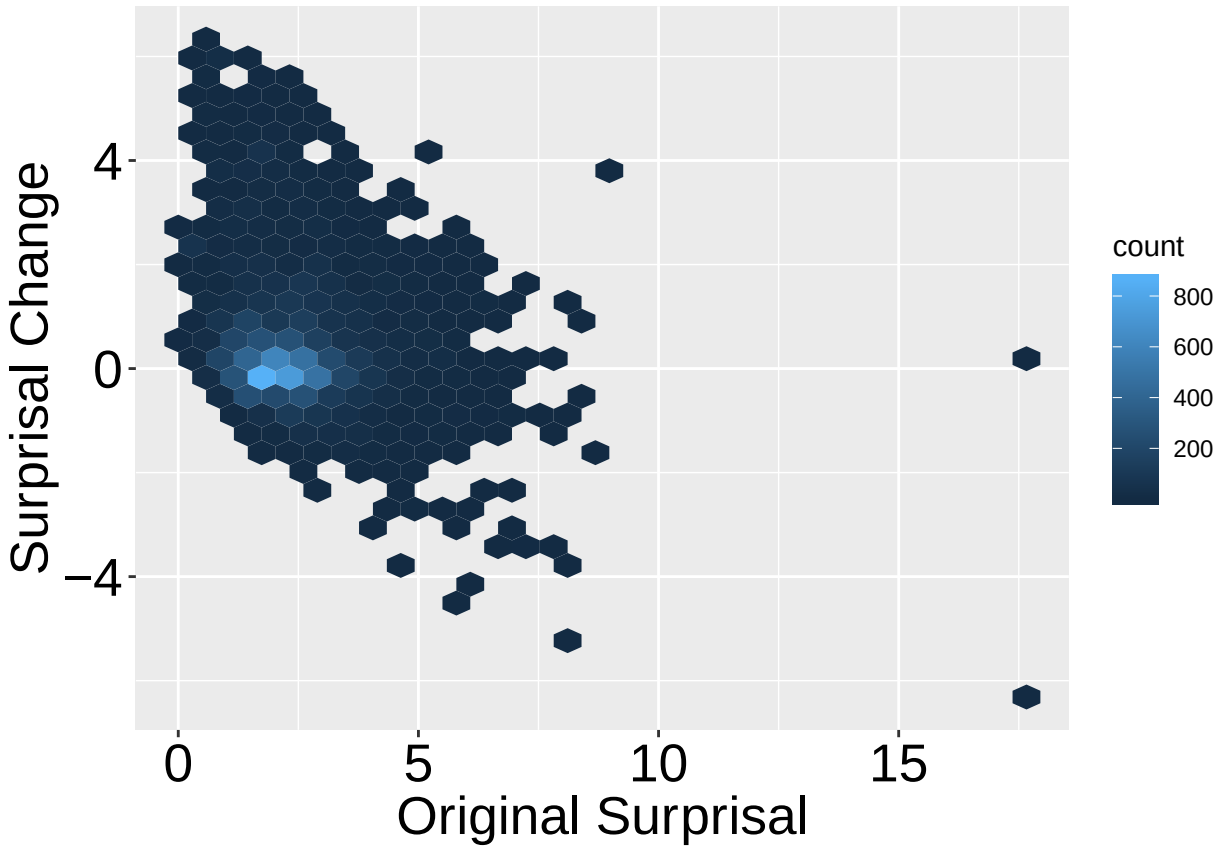

```
## [1] " ------ Expression Cache Model ----- "
## [1] "RmParenTopFiltered10CacheExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -36.319, df = 9456, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.5583273
## sample estimates:
## mean of the differences
##              -0.6066179
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -36.319, df = 9456, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.6584411 -0.5547946
## sample estimates:
## mean of the differences
##              -0.6066179
##
##
```

```
## Cohen's d
##
## d estimate: -0.373472 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.4022240 -0.3447201
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 12525000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3430005
## sample estimates:
## (pseudo)median
##     -0.3769593
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 12525000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.4152379 -0.3405314
## sample estimates:
## (pseudo)median
##     -0.3769593
##
##
## Cliff's Delta
##
## delta estimate: -0.1186437 (negligible)
## 95 percent confidence interval:
##       inf       sup
## -0.134963 -0.102260
```



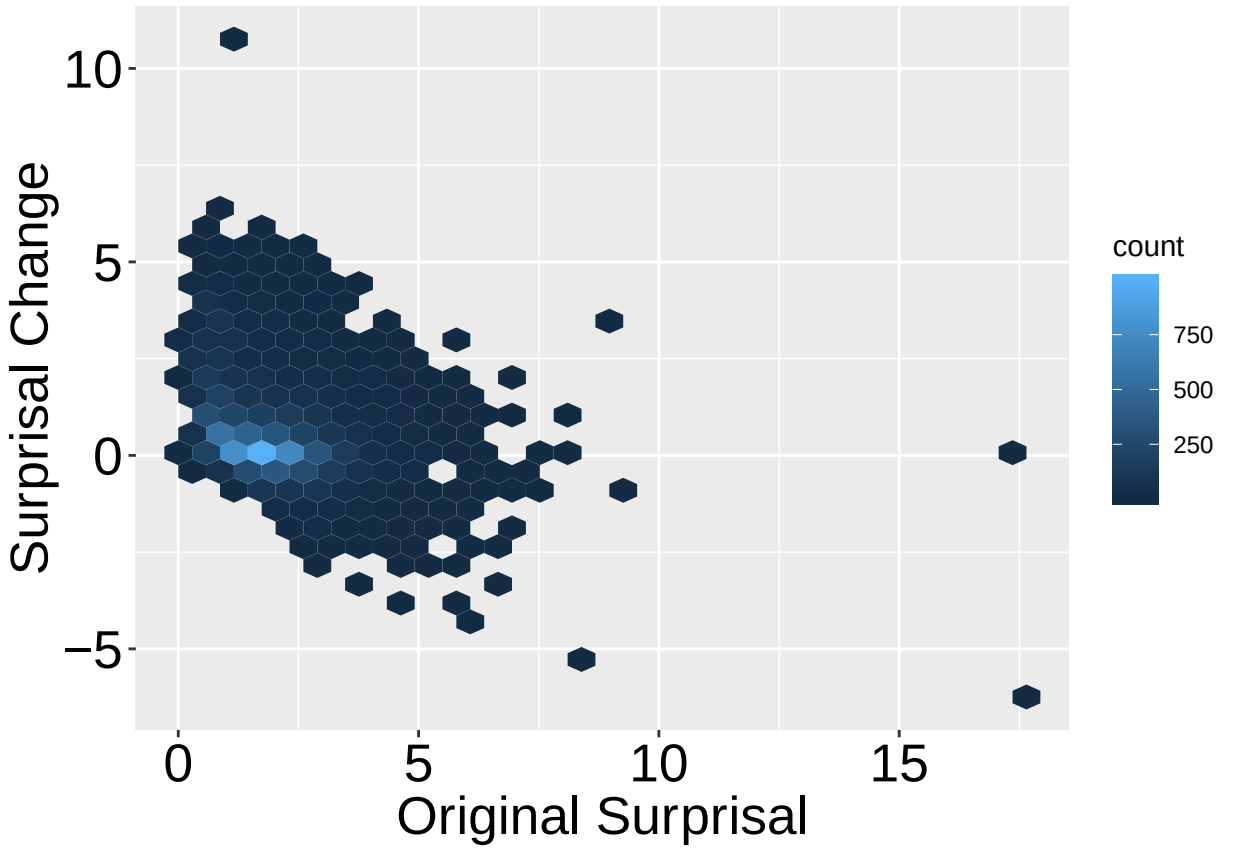

```
## [1] " ------ Expression Global Type Model ----- "
## [1] "RmParenTopFiltered10GlobalTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -31.507, df = 9456, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3144249
## sample estimates:
## mean of the differences
##              -0.3461929
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -31.507, df = 9456, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.3802849 -0.3121009
## sample estimates:
## mean of the differences
##              -0.3461929
##
##
```

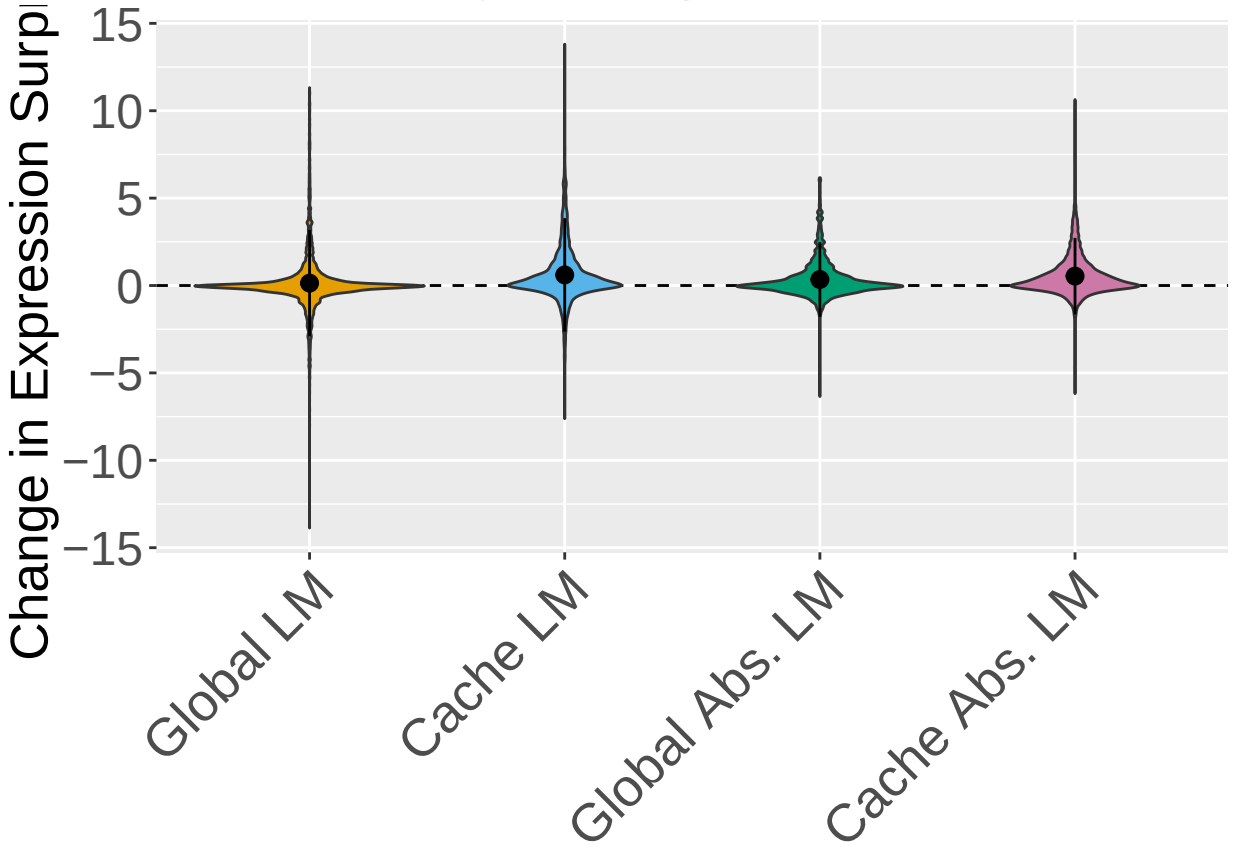
```
## Cohen's d
##
## d estimate: -0.3239918 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.3526827 -0.2953008
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 15998000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.1444262
## sample estimates:
## (pseudo)median
##     -0.1671991
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 15998000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.1922132 -0.1428164
## sample estimates:
## (pseudo)median
##     -0.1671991
##
##
## Cliff's Delta
##
## delta estimate: -0.1337108 (negligible)
## 95 percent confidence interval:
##        inf        sup
## -0.1499588 -0.1173905
```

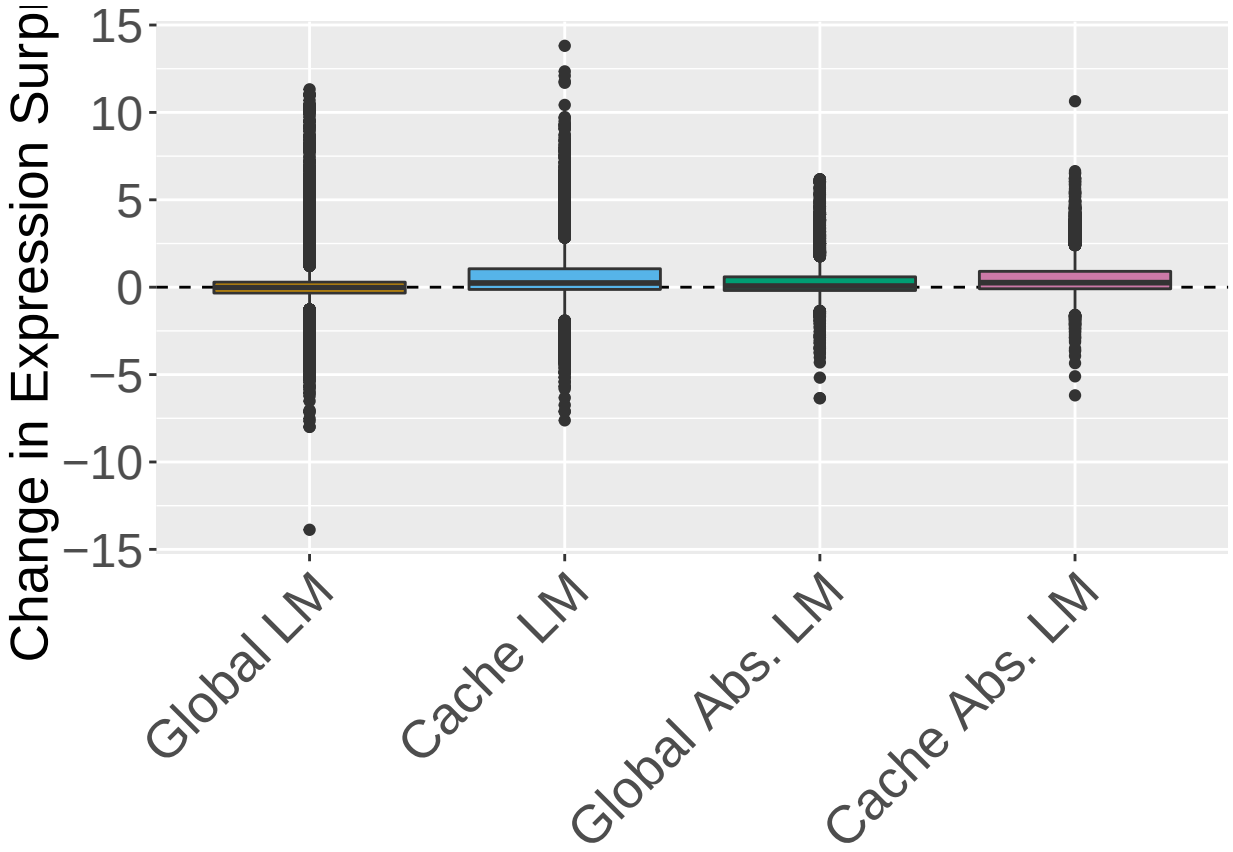



```
## [1] " ------ Expression Cache Type Model ----- "
## [1] "RmParenTopFiltered10CacheTypeExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -47.89, df = 9456, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.5063963
## sample estimates:
## mean of the differences
##              -0.5389326
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -47.89, df = 9456, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.5738490 -0.5040161
## sample estimates:
## mean of the differences
##              -0.5389326
##
##
```

```
## Cohen's d
##
## d estimate: -0.4924616 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.5213950 -0.4635282
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10480000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.3491612
## sample estimates:
## (pseudo)median
##     -0.3760252
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10480000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.4058028 -0.3472347
## sample estimates:
## (pseudo)median
##     -0.3760252
##
##
## Cliff's Delta
##
## delta estimate: -0.2678855 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.2834551 -0.2521747

## No id variables; using all as measure variables

## Warning: Ignoring unknown parameters: mult
```


Surprisal Change
10
5
0
−5
0
5
10
15
Original Surprisal
count
750
500
250


Change in Expression Surpl
15
10
5
0
−5
−10
−15
Global LM
Cache LM
Global Abs. LM
Cache Abs. LM



Regression models

```
m_rm_no_out <- modelGlobal(drmFiltered, "==", "", TRUE)
```

```
##
## Call:
## lm(formula = AverageEntChangeExp ~ BaseAveEntExp + log(NumTokens) +
##     factor(ParentOp), data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5266 -0.3827  0.0434  0.3578  3.6531
##
## Coefficients:
##                                  Estimate Std. Error t value
## (Intercept)                      1.242791   0.124145  10.011
## BaseAveEntExp                   -0.067882   0.002700 -25.143
## log(NumTokens)                  -0.150794   0.031164  -4.839
## factor(ParentOp)^                0.610802   0.115321   5.297
## factor(ParentOp)<                0.176860   0.110513   1.600
## factor(ParentOp)<<               0.193116   0.099081   1.949
## factor(ParentOp)>                0.067244   0.109193   0.616
## factor(ParentOp)|                0.612816   0.084221   7.276
## factor(ParentOp)||              -0.241890   0.079184  -3.055
## factor(ParentOp)-               -0.085703   0.100229  -0.855
## factor(ParentOp)/               -0.204383   0.098783  -2.069
## factor(ParentOp)*               -0.178367   0.115557  -1.544
## factor(ParentOp)&                0.373909   0.084765   4.411
```

```
## factor(ParentOp)&&                    -0.335281   0.079172  -4.235
## factor(ParentOp)+                      0.044635   0.090261   0.495
## factor(ParentOp)ConditionalExpression -0.002151   0.123447  -0.017
##                                       Pr(>|t|)
## (Intercept)                            < 2e-16 ***
## BaseAveEntExp                          < 2e-16 ***
## log(NumTokens)                        1.34e-06 ***
## factor(ParentOp)^                     1.22e-07 ***
## factor(ParentOp)<                      0.10957
## factor(ParentOp)<<                     0.05133 .
## factor(ParentOp)>                      0.53803
## factor(ParentOp)|                     3.85e-13 ***
## factor(ParentOp)||                     0.00226 **
## factor(ParentOp)-                      0.39255
## factor(ParentOp)/                      0.03859 *
## factor(ParentOp)*                      0.12275
## factor(ParentOp)&                     1.05e-05 ***
## factor(ParentOp)&&                    2.32e-05 ***
## factor(ParentOp)+                      0.62096
## factor(ParentOp)ConditionalExpression  0.98610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9174 on 6304 degrees of freedom
## Multiple R-squared:  0.1931, Adjusted R-squared:  0.1911
## F-statistic: 100.6 on 15 and 6304 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: AverageEntChangeExp
##                     Df Sum Sq Mean Sq F value    Pr(>F)
## BaseAveEntExp        1  557.2  557.21 662.000 < 2.2e-16 ***
## log(NumTokens)       1   77.3   77.27  91.797 < 2.2e-16 ***
## factor(ParentOp)    13  635.0   48.85  58.035 < 2.2e-16 ***
## Residuals         6304 5306.1    0.84
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.08473867" "0.01175044" "0.09657399" "0.80693690"
##                        GVIF Df GVIF^(1/(2*Df))
## BaseAveEntExp     1.067698  1        1.033295
## log(NumTokens)    1.233212  1        1.110501
## factor(ParentOp) 1.200847 13        1.007064
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Feb 19, 2019 - 10:46:41 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & AverageEntChangeExp \\
```

```
## \hline \\[-1.8ex]
##  BaseAveEntExp & $-$0.068$^{***}$ (0.003) \\
##   log(NumTokens) & $-$0.151$^{***}$ (0.031) \\
##   factor(ParentOp)$\hat{\mkern6mu}$ & 0.611$^{***}$ (0.115) \\
##   factor(ParentOp)\textless  & 0.177 (0.111) \\
##   factor(ParentOp)\textless \textless  & 0.193$^{*}$ (0.099) \\
##   factor(ParentOp)\textgreater  & 0.067 (0.109) \\
##   factor(ParentOp)\textbar  & 0.613$^{***}$ (0.084) \\
##   factor(ParentOp)\textbar \textbar  & $-$0.242$^{***}$ (0.079) \\
##   factor(ParentOp)- & $-$0.086 (0.100) \\
##   factor(ParentOp)/ & $-$0.204$^{**}$ (0.099) \\
##   factor(ParentOp)\textasteriskcentered  & $-$0.178 (0.116) \\
##   factor(ParentOp)& & 0.374$^{***}$ (0.085) \\
##   factor(ParentOp)&& & $-$0.335$^{***}$ (0.079) \\
##   factor(ParentOp)+ & 0.045 (0.090) \\
##   factor(ParentOp)ConditionalExpression & $-$0.002 (0.123) \\
##   Constant & 1.243$^{***}$ (0.124) \\
##  \hline \\[-1.8ex]
## Observations & 6,320 \\
## R$^{2}$ & 0.193 \\
## Adjusted R$^{2}$ & 0.191 \\
## Residual Std. Error & 0.917 (df = 6304) \\
## F Statistic & 100.551$^{***}$ (df = 15; 6304) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:46:41 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## BaseAveEntExp & 1 & 557.21 & 557.21 & 662.00 & 0.0000 \\
##   log(NumTokens) & 1 & 77.27 & 77.27 & 91.80 & 0.0000 \\
##   factor(ParentOp) & 13 & 635.03 & 48.85 & 58.04 & 0.0000 \\
##   Residuals & 6304 & 5306.12 & 0.84 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
m_rm_cache_no_out <- modelCache(drmFiltered, "==", "", TRUE)
```

```
##
## Call:
## lm(formula = CacheAverageEntChangeExp ~ BaseCacheAveEntExp +
##     log(NumTokens) + factor(ParentOp), data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.9701 -0.6053 -0.0823  0.5520  4.8528
##
```

```
## Coefficients:
##                                    Estimate Std. Error t value
## (Intercept)                         3.491150   0.153683  22.717
## BaseCacheAveEntExp                 -0.200488   0.004409 -45.477
## log(NumTokens)                     -0.653795   0.038251 -17.092
## factor(ParentOp)^                   0.793071   0.139694   5.677
## factor(ParentOp)<                   0.268020   0.143111   1.873
## factor(ParentOp)<<                  0.130011   0.125925   1.032
## factor(ParentOp)>                   0.272046   0.139370   1.952
## factor(ParentOp)|                   0.513493   0.107135   4.793
## factor(ParentOp)||                 -0.004310   0.101404  -0.043
## factor(ParentOp)-                   0.062882   0.128167   0.491
## factor(ParentOp)/                  -0.054853   0.124881  -0.439
## factor(ParentOp)*                  -0.013017   0.147651  -0.088
## factor(ParentOp)&                   0.725524   0.107665   6.739
## factor(ParentOp)&&                  0.032654   0.101489   0.322
## factor(ParentOp)+                   0.158193   0.114825   1.378
## factor(ParentOp)ConditionalExpression  0.271331   0.163123   1.663
##                                   Pr(>|t|)
## (Intercept)                        < 2e-16 ***
## BaseCacheAveEntExp                 < 2e-16 ***
## log(NumTokens)                     < 2e-16 ***
## factor(ParentOp)^                 1.43e-08 ***
## factor(ParentOp)<                   0.0611 .
## factor(ParentOp)<<                  0.3019
## factor(ParentOp)>                   0.0510 .
## factor(ParentOp)|                 1.68e-06 ***
## factor(ParentOp)||                  0.9661
## factor(ParentOp)-                   0.6237
## factor(ParentOp)/                   0.6605
## factor(ParentOp)*                   0.9298
## factor(ParentOp)&                 1.74e-11 ***
## factor(ParentOp)&&                  0.7477
## factor(ParentOp)+                   0.1683
## factor(ParentOp)ConditionalExpression   0.0963 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.117 on 6240 degrees of freedom
## Multiple R-squared:  0.3038, Adjusted R-squared:  0.3021
## F-statistic: 181.5 on 15 and 6240 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: CacheAverageEntChangeExp
##                      Df Sum Sq Mean Sq  F value    Pr(>F)
## BaseCacheAveEntExp    1 2393.0 2393.00 1919.576 < 2.2e-16 ***
## log(NumTokens)        1  603.7  603.71  484.276 < 2.2e-16 ***
## factor(ParentOp)     13  397.7   30.59   24.541 < 2.2e-16 ***
## Residuals          6240 7779.0    1.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.21416956" "0.05403127" "0.03559452" "0.69620466"
##                        GVIF Df GVIF^(1/(2*Df))
```

```
## BaseCacheAveEntExp 1.083181  1        1.040760
## log(NumTokens)     1.232415  1        1.110142
## factor(ParentOp)   1.202736 13        1.007125
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Feb 19, 2019 - 10:46:42 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & CacheAverageEntChangeExp \\
## \hline \\[-1.8ex]
##  BaseCacheAveEntExp & $-$0.200$^{***}$ (0.004) \\
##   log(NumTokens) & $-$0.654$^{***}$ (0.038) \\
##   factor(ParentOp)$\hat{\mkern6mu}$ & 0.793$^{***}$ (0.140) \\
##   factor(ParentOp)\textless  & 0.268$^{*}$ (0.143) \\
##   factor(ParentOp)\textless \textless  & 0.130 (0.126) \\
##   factor(ParentOp)\textgreater  & 0.272$^{*}$ (0.139) \\
##   factor(ParentOp)\textbar  & 0.513$^{***}$ (0.107) \\
##   factor(ParentOp)\textbar \textbar  & $-$0.004 (0.101) \\
##   factor(ParentOp)- & 0.063 (0.128) \\
##   factor(ParentOp)/ & $-$0.055 (0.125) \\
##   factor(ParentOp)\textasteriskcentered  & $-$0.013 (0.148) \\
##   factor(ParentOp)& & 0.726$^{***}$ (0.108) \\
##   factor(ParentOp)&& & 0.033 (0.101) \\
##   factor(ParentOp)+ & 0.158 (0.115) \\
##   factor(ParentOp)ConditionalExpression & 0.271$^{*}$ (0.163) \\
##   Constant & 3.491$^{***}$ (0.154) \\
##  \hline \\[-1.8ex]
## Observations & 6,256 \\
## R$^{2}$ & 0.304 \\
## Adjusted R$^{2}$ & 0.302 \\
## Residual Std. Error & 1.117 (df = 6240) \\
## F Statistic & 181.525$^{***}$ (df = 15; 6240) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:46:42 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## BaseCacheAveEntExp & 1 & 2393.00 & 2393.00 & 1919.58 & 0.0000 \\
##   log(NumTokens) & 1 & 603.71 & 603.71 & 484.28 & 0.0000 \\
##   factor(ParentOp) & 13 & 397.71 & 30.59 & 24.54 & 0.0000 \\
```

```
##   Residuals & 6240 & 7778.97 & 1.25 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
mt_rm_no_out <- modelGlobalType(drmFiltered, "==", "", TRUE)
```

```
##
## Call:
## lm(formula = TypeAverageEntChangeExp ~ BaseTypeAveEntExp + log(NumTokens) +
##     factor(ParentOp), data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0797 -0.4401 -0.0946  0.2938  3.2013
##
## Coefficients:
##                                     Estimate Std. Error t value
## (Intercept)                         1.611991   0.088185  18.280
## BaseTypeAveEntExp                  -0.138897   0.008841 -15.711
## log(NumTokens)                     -0.322985   0.021685 -14.894
## factor(ParentOp)^                   1.593030   0.085734  18.581
## factor(ParentOp)<                   0.723698   0.081707   8.857
## factor(ParentOp)<<                  0.226675   0.076686   2.956
## factor(ParentOp)>                   0.207075   0.080512   2.572
## factor(ParentOp)|                   0.548759   0.062982   8.713
## factor(ParentOp)||                 -0.136368   0.059421  -2.295
## factor(ParentOp)-                  -0.069188   0.072916  -0.949
## factor(ParentOp)/                  -0.323050   0.072172  -4.476
## factor(ParentOp)*                  -0.294583   0.084879  -3.471
## factor(ParentOp)&                   0.231751   0.062980   3.680
## factor(ParentOp)&&                 -0.270806   0.059256  -4.570
## factor(ParentOp)+                  -0.014911   0.066535  -0.224
## factor(ParentOp)ConditionalExpression  0.006812   0.092897   0.073
##                                    Pr(>|t|)
## (Intercept)                         < 2e-16 ***
## BaseTypeAveEntExp                   < 2e-16 ***
## log(NumTokens)                      < 2e-16 ***
## factor(ParentOp)^                   < 2e-16 ***
## factor(ParentOp)<                   < 2e-16 ***
## factor(ParentOp)<<                 0.003130 **
## factor(ParentOp)>                  0.010135 *
## factor(ParentOp)|                   < 2e-16 ***
## factor(ParentOp)||                 0.021769 *
## factor(ParentOp)-                  0.342719
## factor(ParentOp)/                  7.74e-06 ***
## factor(ParentOp)*                  0.000523 ***
## factor(ParentOp)&                  0.000235 ***
## factor(ParentOp)&&                 4.97e-06 ***
## factor(ParentOp)+                  0.822684
## factor(ParentOp)ConditionalExpression 0.941547
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6512 on 6168 degrees of freedom
```

```
## Multiple R-squared:  0.2821, Adjusted R-squared:  0.2804
## F-statistic: 161.6 on 15 and 6168 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: TypeAverageEntChangeExp
##                     Df  Sum Sq Mean Sq F value    Pr(>F)
## BaseTypeAveEntExp    1  101.22 101.216  238.66 < 2.2e-16 ***
## log(NumTokens)       1  204.15 204.150  481.37 < 2.2e-16 ***
## factor(ParentOp)    13  722.71  55.593  131.08 < 2.2e-16 ***
## Residuals         6168 2615.86   0.424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.02777661" "0.05602459" "0.19833254" "0.71786626"
##                        GVIF Df GVIF^(1/(2*Df))
## BaseTypeAveEntExp 1.111869  1        1.054452
## log(NumTokens)    1.188698  1        1.090274
## factor(ParentOp)  1.305199 13        1.010297
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Feb 19, 2019 - 10:46:44 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & TypeAverageEntChangeExp \\
## \hline \\[-1.8ex]
##  BaseTypeAveEntExp & $-$0.139$^{***}$ (0.009) \\
##   log(NumTokens) & $-$0.323$^{***}$ (0.022) \\
##   factor(ParentOp)$\hat{\mkern6mu}$ & 1.593$^{***}$ (0.086) \\
##   factor(ParentOp)\textless  & 0.724$^{***}$ (0.082) \\
##   factor(ParentOp)\textless \textless  & 0.227$^{***}$ (0.077) \\
##   factor(ParentOp)\textgreater  & 0.207$^{**}$ (0.081) \\
##   factor(ParentOp)\textbar  & 0.549$^{***}$ (0.063) \\
##   factor(ParentOp)\textbar \textbar  & $-$0.136$^{**}$ (0.059) \\
##   factor(ParentOp)- & $-$0.069 (0.073) \\
##   factor(ParentOp)/ & $-$0.323$^{***}$ (0.072) \\
##   factor(ParentOp)\textasteriskcentered  & $-$0.295$^{***}$ (0.085) \\
##   factor(ParentOp)& & 0.232$^{***}$ (0.063) \\
##   factor(ParentOp)&& & $-$0.271$^{***}$ (0.059) \\
##   factor(ParentOp)+ & $-$0.015 (0.067) \\
##   factor(ParentOp)ConditionalExpression & 0.007 (0.093) \\
##   Constant & 1.612$^{***}$ (0.088) \\
##  \hline \\[-1.8ex]
## Observations & 6,184 \\
## R$^{2}$ & 0.282 \\
## Adjusted R$^{2}$ & 0.280 \\
## Residual Std. Error & 0.651 (df = 6168) \\
## F Statistic & 161.609$^{***}$ (df = 15; 6168) \\
## \hline
```

```
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:46:44 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## BaseTypeAveEntExp & 1 & 101.22 & 101.22 & 238.66 & 0.0000 \\
##   log(NumTokens) & 1 & 204.15 & 204.15 & 481.37 & 0.0000 \\
##   factor(ParentOp) & 13 & 722.71 & 55.59 & 131.08 & 0.0000 \\
##   Residuals & 6168 & 2615.86 & 0.42 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
mt_rm_cache_no_out <- modelCacheType(drmFiltered, "==", "", TRUE)
```

```
##
## Call:
## lm(formula = CacheTypeAverageEntChangeExp ~ BaseCacheTypeAveEntExp +
##     log(NumTokens) + factor(ParentOp), data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1218 -0.5000 -0.1150  0.3568  3.1738
##
## Coefficients:
##                                     Estimate Std. Error t value
## (Intercept)                         2.886181   0.102051  28.282
## BaseCacheTypeAveEntExp             -0.335203   0.009861 -33.993
## log(NumTokens)                     -0.548354   0.025069 -21.874
## factor(ParentOp)^                   1.137812   0.095520  11.912
## factor(ParentOp)<                   0.527590   0.095230   5.540
## factor(ParentOp)<<                  0.048534   0.088044   0.551
## factor(ParentOp)>                   0.106435   0.095797   1.111
## factor(ParentOp)|                   0.566798   0.073428   7.719
## factor(ParentOp)||                 -0.257851   0.069543  -3.708
## factor(ParentOp)-                  -0.165179   0.086317  -1.914
## factor(ParentOp)/                  -0.344881   0.085489  -4.034
## factor(ParentOp)*                  -0.374123   0.101082  -3.701
## factor(ParentOp)&                   0.362161   0.073760   4.910
## factor(ParentOp)&&                 -0.339449   0.069511  -4.883
## factor(ParentOp)+                   0.018759   0.078265   0.240
## factor(ParentOp)ConditionalExpression  0.033441   0.108312   0.309
##                                    Pr(>|t|)
## (Intercept)                         < 2e-16 ***
## BaseCacheTypeAveEntExp              < 2e-16 ***
## log(NumTokens)                      < 2e-16 ***
## factor(ParentOp)^                   < 2e-16 ***
## factor(ParentOp)<                  3.15e-08 ***
```

```
## factor(ParentOp)<<                    0.581486
## factor(ParentOp)>                     0.266589
## factor(ParentOp)|                     1.36e-14 ***
## factor(ParentOp)||                    0.000211 ***
## factor(ParentOp)-                     0.055713 .
## factor(ParentOp)/                     5.54e-05 ***
## factor(ParentOp)*                     0.000216 ***
## factor(ParentOp)&                     9.34e-07 ***
## factor(ParentOp)&&                    1.07e-06 ***
## factor(ParentOp)+                     0.810583
## factor(ParentOp)ConditionalExpression 0.757522
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7522 on 6221 degrees of freedom
## Multiple R-squared:  0.3624, Adjusted R-squared:  0.3609
## F-statistic: 235.8 on 15 and 6221 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: CacheTypeAverageEntChangeExp
##                           Df Sum Sq Mean Sq F value    Pr(>F)
## BaseCacheTypeAveEntExp     1  706.6  706.59 1248.70 < 2.2e-16 ***
## log(NumTokens)             1  517.7  517.70  914.89 < 2.2e-16 ***
## factor(ParentOp)          13  776.9   59.76  105.61 < 2.2e-16 ***
## Residuals               6221 3520.2    0.57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.12797335" "0.09376274" "0.14070290" "0.63756100"
##                             GVIF Df GVIF^(1/(2*Df))
## BaseCacheTypeAveEntExp 1.065166  1        1.032069
## log(NumTokens)         1.185127  1        1.088636
## factor(ParentOp)       1.238730 13        1.008268
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Tue, Feb 19, 2019 - 10:46:45 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & CacheTypeAverageEntChangeExp \\
## \hline \\[-1.8ex]
##  BaseCacheTypeAveEntExp & $-$0.335$^{***}$ (0.010) \\
##   log(NumTokens) & $-$0.548$^{***}$ (0.025) \\
##   factor(ParentOp)$\hat{\mkern6mu}$ & 1.138$^{***}$ (0.096) \\
##   factor(ParentOp)\textless  & 0.528$^{***}$ (0.095) \\
##   factor(ParentOp)\textless \textless  & 0.049 (0.088) \\
##   factor(ParentOp)\textgreater  & 0.106 (0.096) \\
##   factor(ParentOp)\textbar  & 0.567$^{***}$ (0.073) \\
##   factor(ParentOp)\textbar \textbar  & $-$0.258$^{***}$ (0.070) \\
```

```
##   factor(ParentOp)- & $-$0.165$^{*}$ (0.086) \\
##   factor(ParentOp)/ & $-$0.345$^{***}$ (0.085) \\
##   factor(ParentOp)\textasteriskcentered  & $-$0.374$^{***}$ (0.101) \\
##   factor(ParentOp)& & 0.362$^{***}$ (0.074) \\
##   factor(ParentOp)&& & $-$0.339$^{***}$ (0.070) \\
##   factor(ParentOp)+ & 0.019 (0.078) \\
##   factor(ParentOp)ConditionalExpression & 0.033 (0.108) \\
##   Constant & 2.886$^{***}$ (0.102) \\
##  \hline \\[-1.8ex]
## Observations & 6,237 \\
## R$^{2}$ & 0.362 \\
## Adjusted R$^{2}$ & 0.361 \\
## Residual Std. Error & 0.752 (df = 6221) \\
## F Statistic & 235.767$^{***}$ (df = 15; 6221) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:46:45 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## BaseCacheTypeAveEntExp & 1 & 706.59 & 706.59 & 1248.70 & 0.0000 \\
##   log(NumTokens) & 1 & 517.70 & 517.70 & 914.89 & 0.0000 \\
##   factor(ParentOp) & 13 & 776.87 & 59.76 & 105.61 & 0.0000 \\
##   Residuals & 6221 & 3520.20 & 0.57 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

Output the effect sizes

```
ParenOut <- printEffTable(pairedResults)
```

```
## [1] "Type,PTOne,PTTwo,CITTwo,CohensD,PWilcoxOne,PWilcoxTwo,CIWilcoxTwo,CliffDelta"
## [1] "AddParenTopCacheExp,0,0,-1.526 -1.5062,-0.8712,0,0,-1.0666 -1.0559,-0.455"
## [1] "AddParenTopCacheTypeExp,0,0,-0.7276 -0.7197,-1.0374,0,0,-0.6852 -0.6793,-0.599"
## [1] "AddParenTopFiltered100CacheExp,0,0,-0.7667 -0.7323,-0.5355,0,0,-0.5767 -0.5436,-0.1519"
## [1] "AddParenTopFiltered100CacheTypeExp,0,0,-0.4084 -0.392,-0.5983,0,0,-0.3781 -0.3642,-0.3106"
## [1] "AddParenTopFiltered100GlobalExp,0,0,-0.4823 -0.4549,-0.4207,0,0,-0.3341 -0.3202,-0.0635"
## [1] "AddParenTopFiltered100GlobalTypeExp,0,0,-0.2223 -0.2075,-0.3563,0,0,-0.2349 -0.2238,-0.1675"
## [1] "AddParenTopFiltered10CacheExp,0,0,-0.7326 -0.6976,-0.5147,0,0,-0.5386 -0.5059,-0.1452"
## [1] "AddParenTopFiltered10CacheTypeExp,0,0,-0.3974 -0.3806,-0.582,0,0,-0.3674 -0.3533,-0.302"
## [1] "AddParenTopFiltered10GlobalExp,0,0,-0.4677 -0.4397,-0.4093,0,0,-0.3273 -0.3094,-0.0622"
## [1] "AddParenTopFiltered10GlobalTypeExp,0,0,-0.2187 -0.2034,-0.347,0,0,-0.2316 -0.2202,-0.1653"
## [1] "AddParenTopGlobalExp,0,0,-1.8094 -1.7802,-0.6987,0,0,-0.8707 -0.8567,-0.3782"
## [1] "AddParenTopGlobalTypeExp,0,0,-0.3655 -0.3593,-0.6595,0,0,-0.318 -0.3111,-0.4068"
## [1] "RmParenTopCacheExp,0,0,-0.156 -0.1335,-0.1825,1,0,5e-04 5e-04,-0.0511"
## [1] "RmParenTopCacheTypeExp,0,0,-1.0336 -1.0135,-1.4425,0,0,-0.958 -0.9537,-0.8227"
## [1] "RmParenTopFiltered100CacheExp,0,0,-0.6742 -0.5718,-0.3829,0,0,-0.4383 -0.3615,-0.122"
```

```
## [1] "RmParenTopFiltered100CacheTypeExp,0,0,-0.5867 -0.5177,-0.504,0,0,-0.4225 -0.3633,-0.277"
## [1] "RmParenTopFiltered100GlobalExp,0,0,-0.197 -0.1007,-0.0973,0.9586,0.0828,-0.01 0.0341,-0.0162"
## [1] "RmParenTopFiltered100GlobalTypeExp,0,0,-0.3866 -0.3192,-0.3297,0,0,-0.2001 -0.1501,-0.1368"
## [1] "RmParenTopFiltered10CacheExp,0,0,-0.6584 -0.5548,-0.3735,0,0,-0.4152 -0.3405,-0.1186"
## [1] "RmParenTopFiltered10CacheTypeExp,0,0,-0.5738 -0.504,-0.4925,0,0,-0.4058 -0.3472,-0.2679"
## [1] "RmParenTopFiltered10GlobalExp,0,0,-0.1856 -0.0885,-0.09,0.996,0.008,-0.0031 0.0394,-0.0145"
## [1] "RmParenTopFiltered10GlobalTypeExp,0,0,-0.3803 -0.3121,-0.324,0,0,-0.1922 -0.1428,-0.1337"
## [1] "RmParenTopGlobalExp,0,0,-0.0776 -0.0555,-0.0852,0,0,-0.1502 -0.0954,-0.0062"
## [1] "RmParenTopGlobalTypeExp,0,0,-0.476 -0.462,-0.9478,0,0,-0.5033 -0.5033,-0.7223"
```

```
save(ParenOut, file = "/data/anon/SemanticTransformation/sample/ParenOut.RDat")
save(dap, file = "/data/anon/SemanticTransformation/sample/dap.RDat")
save(dapFiltered, file = "/data/anon/SemanticTransformation/sample/dapFiltered.RDat")
save(dapFiltered2, file = "/data/anon/SemanticTransformation/sample/dapFiltered2.RDat")
save(drm, file = "/data/anon/SemanticTransformation/sample/drm.RDat")
save(drmFiltered, file = "/data/anon/SemanticTransformation/sample/drmFiltered.RDat")
save(drmFiltered2, file = "/data/anon/SemanticTransformation/sample/drmFiltered2.RDat")
```