Variable Swaping Comparison 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/VariableSwapping.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/ColorBlind.R")
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/ResultPrinter.R")
source("./RSource/ResultPrinter.R")
source("./RSource/LMDiffModels.R")
```

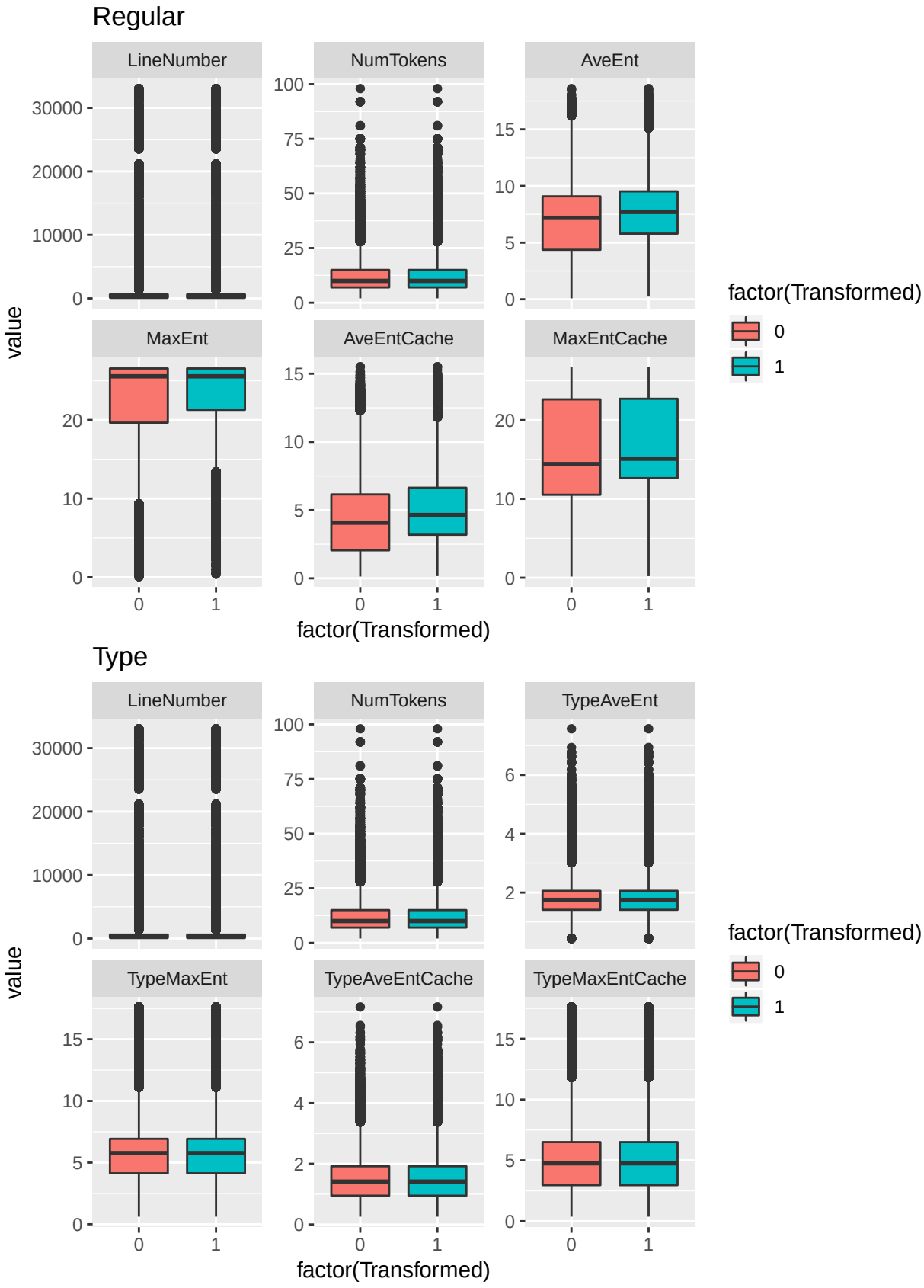
```
setwd("/data/anon/SemanticTransformation")
#dVsp <- compareDepthSummaryVar("variable_rename_type.csv", "VarSwapWithin", "SAME")
dVsp <- compareDepthSummaryVar("variable_rename_type_topstarred.csv", "VarSwapWithinTop", "SAME", FALSE
```

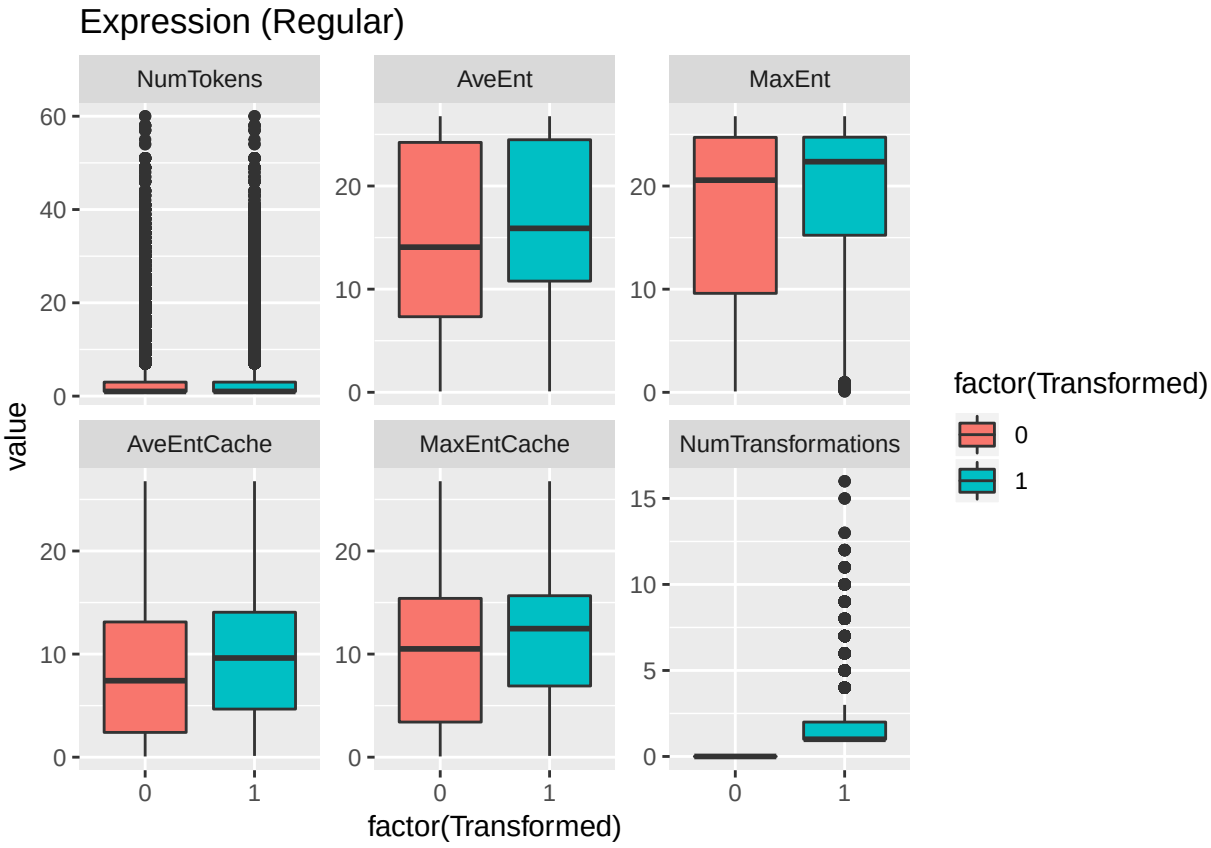
```
## [[1]]
## [1] "5"    "Bool"
```

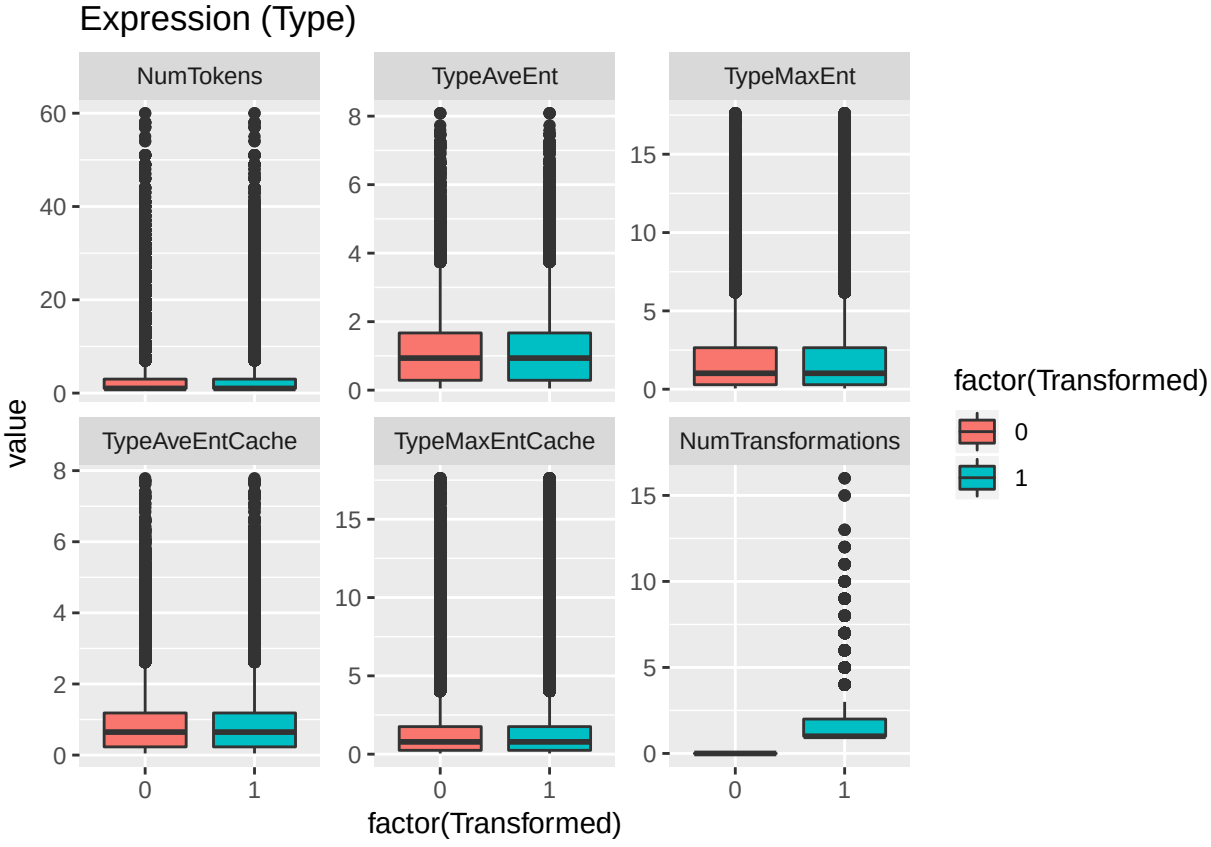
```
## Loading required package: tcltk
```

```
##  [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"
```

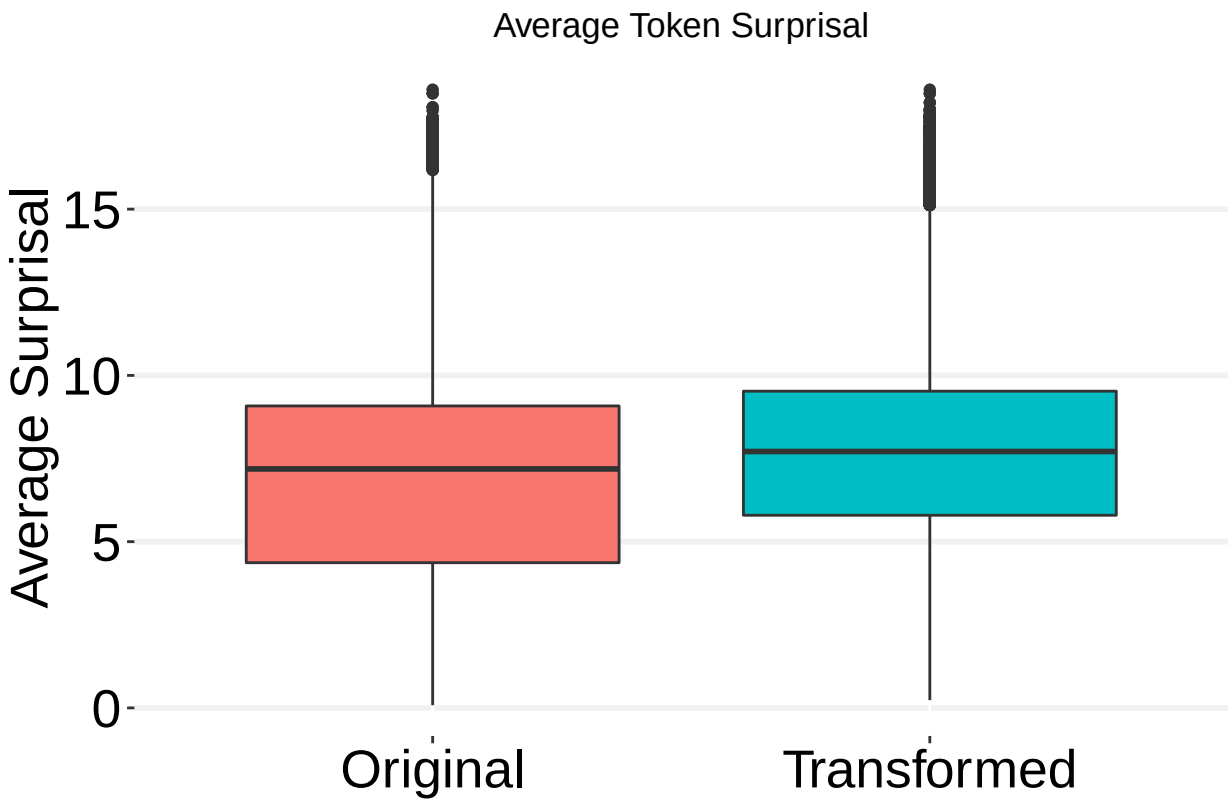

Regular
LineNumber
NumTokens
AveEnt
MaxEnt
AveEntCache
MaxEntCache
value
factor(Transformed)
factor(Transformed)
0
1
Type
LineNumber
NumTokens
TypeAveEnt
TypeMaxEnt
TypeAveEntCache
TypeMaxEntCache
value
factor(Transformed)
factor(Transformed)
0
1





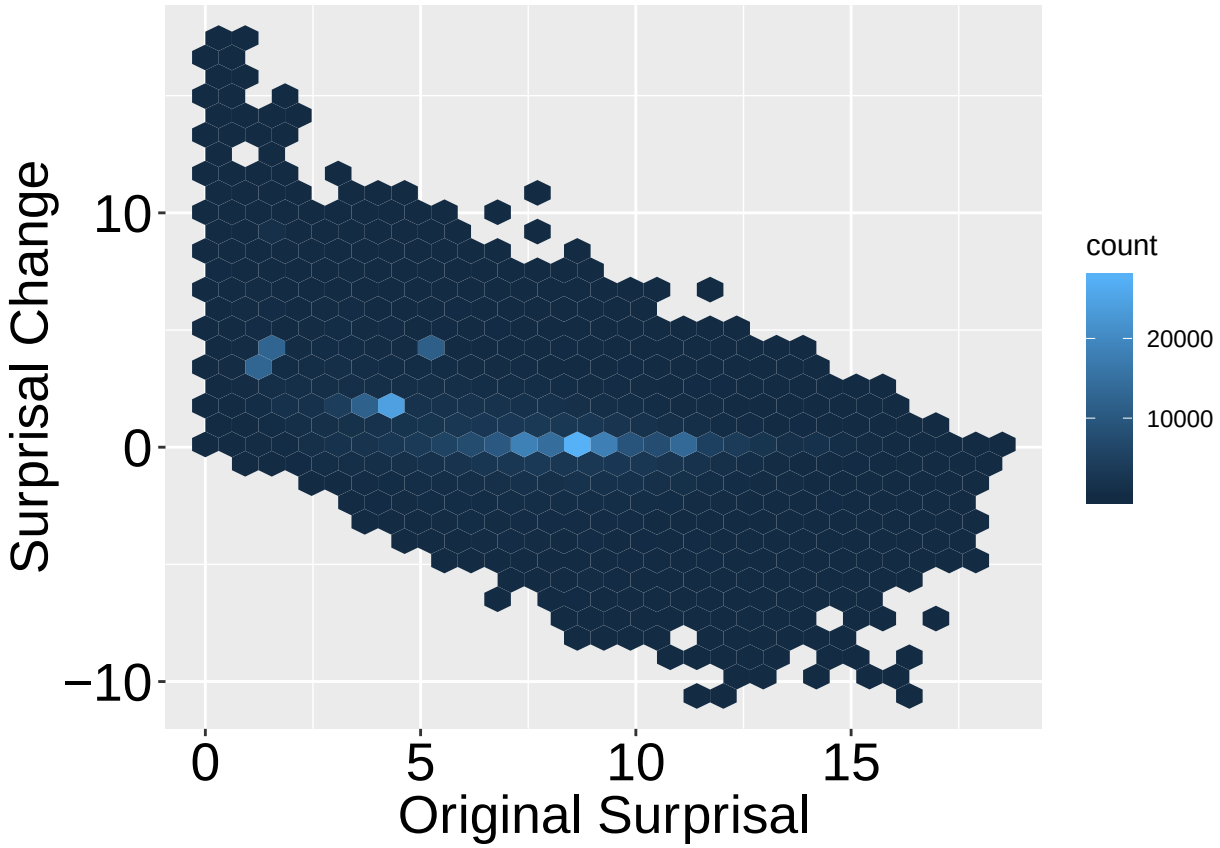

```
## [1] " ------ Global Model ----- "
```

```
## [1] "VarSwapWithinTopGlobal Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -260.61, df = 384790, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.7989391
## sample estimates:
## mean of the differences
##              -0.8078999
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -260.61, df = 384790, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.8175159 -0.7982839
## sample estimates:
## mean of the differences
##              -0.8078999
## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -0.420125 (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 = 9801600000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.9667575
## sample estimates:
## (pseudo)median
##     -0.9767836
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 9801600000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.9876339 -0.9660876
## sample estimates:
## (pseudo)median
##     -0.9767836
##
##
## Cliff's Delta
##
## delta estimate: -0.1426076 (negligible)
```

```
## 95 percent confidence interval:
##        inf        sup
## -0.1451713 -0.1400419
```




```
## [1] " ------ Cache Model ----- "
## [1] "VarSwapWithinTopCache Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -249.77, df = 384790, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.6331414
## sample estimates:
## mean of the differences
##              -0.6405544
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -249.77, df = 384790, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.6485095 -0.6325993
## sample estimates:
```

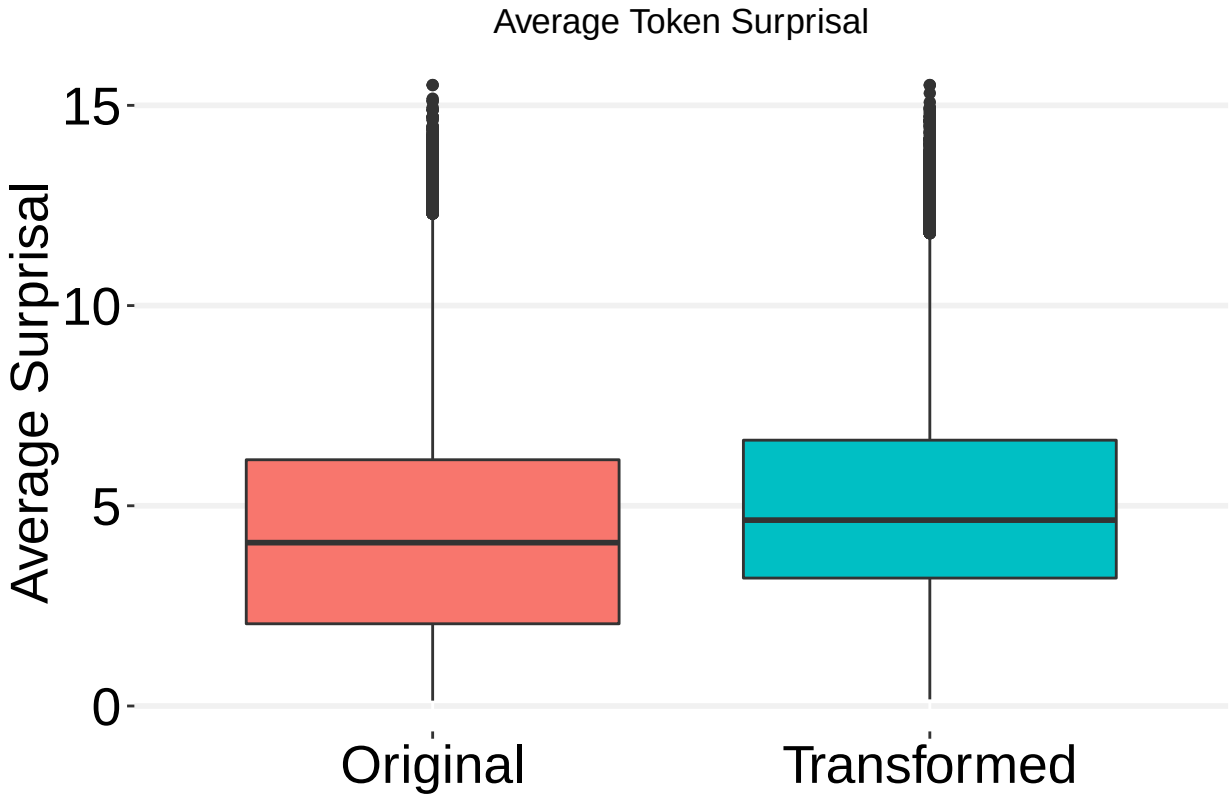
```
## mean of the differences
##              -0.6405544
```

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

```
##
## Cohen's d
##
## d estimate: -0.4026489 (small)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

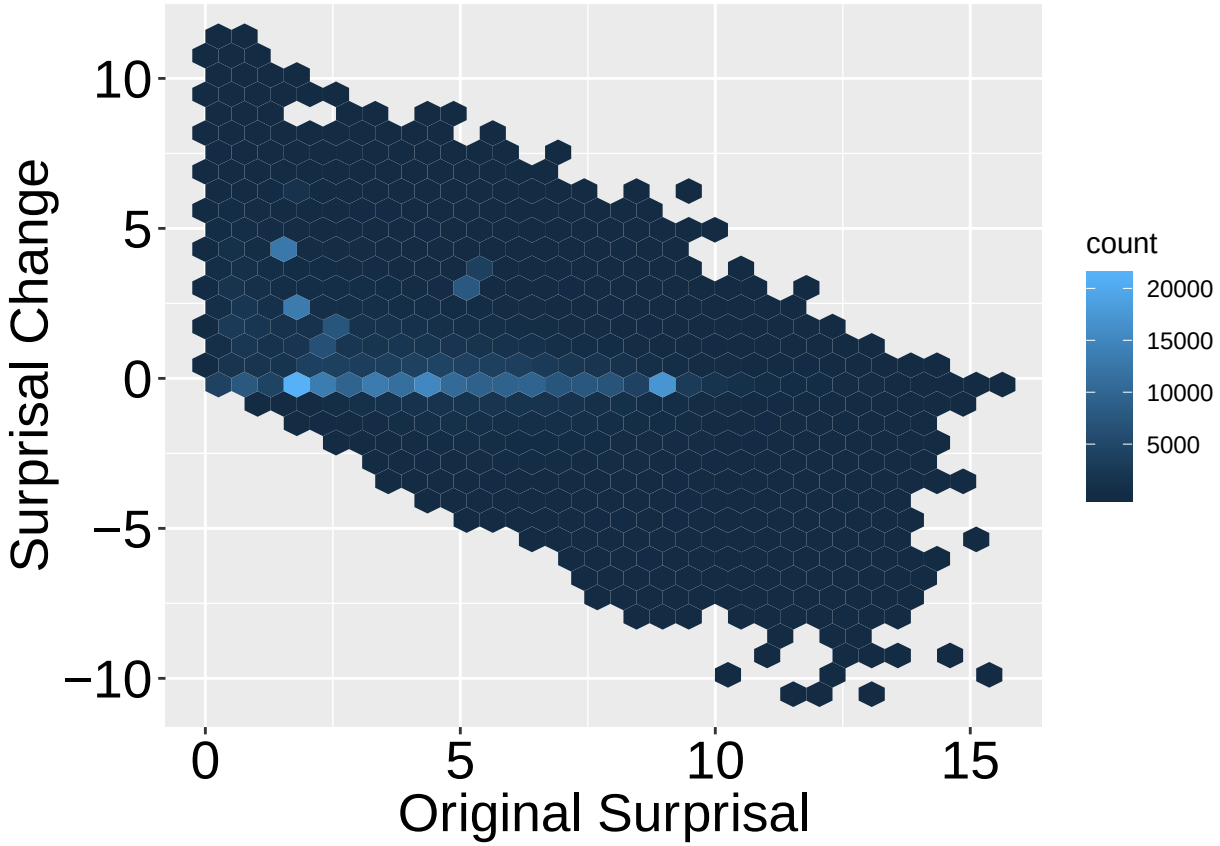
```
## Warning in n1 * n2: NAs produced by integer overflow
```

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




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.1107e+10, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.6410837
## sample estimates:
## (pseudo)median
##     -0.6484666
##
##
```

```
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.1107e+10, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.6567772 -0.6403980
## sample estimates:
## (pseudo)median
##     -0.6484666
##
##
## Cliff's Delta
##
## delta estimate: -0.1577488 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.1602974 -0.1551982
```




```
## [1] " ------ Expression Global Model ----- "
## [1] "VarSwapWithinTopGlobalExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -189.14, df = 384790, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
```

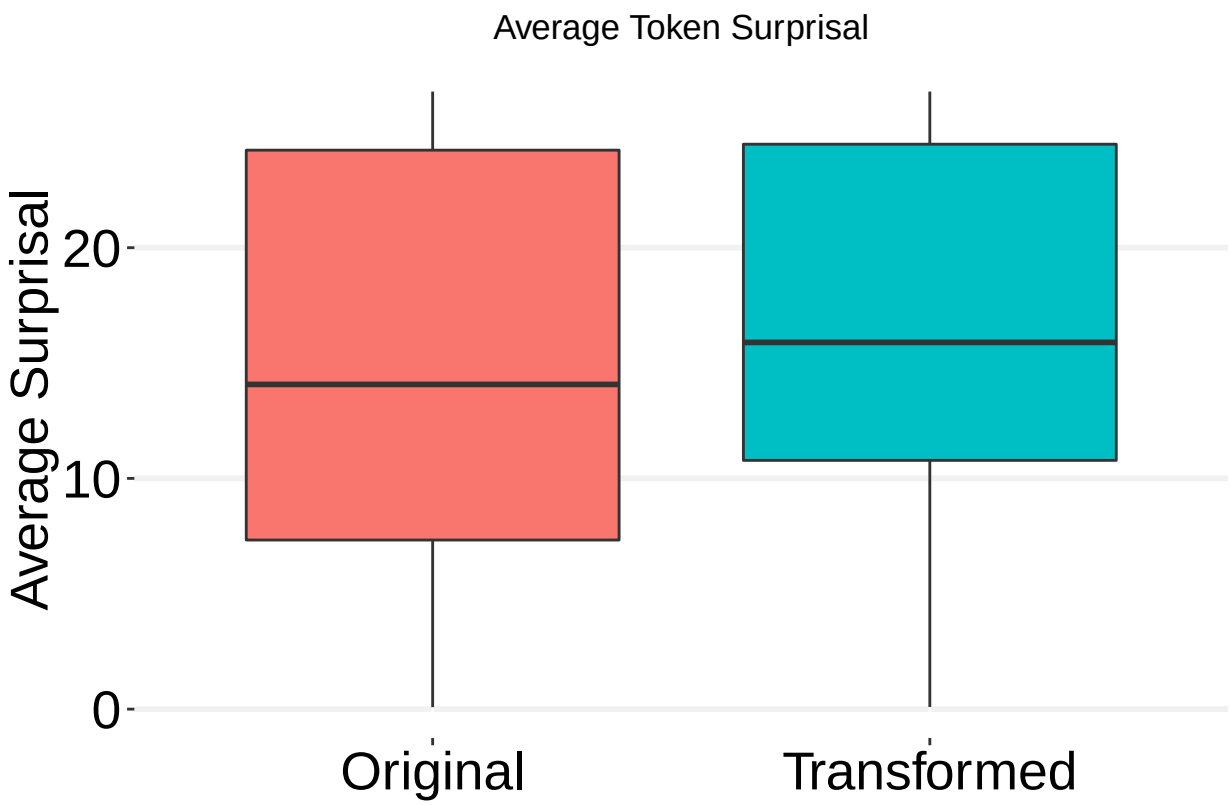
```
## 99.80769 percent confidence interval:
##       -Inf -2.093226
## sample estimates:
## mean of the differences
##               -2.125712
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -189.14, df = 384790, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -2.160573 -2.090850
## sample estimates:
## mean of the differences
##               -2.125712
```

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

```
##
## Cohen's d
##
## d estimate: -0.3049133 (small)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

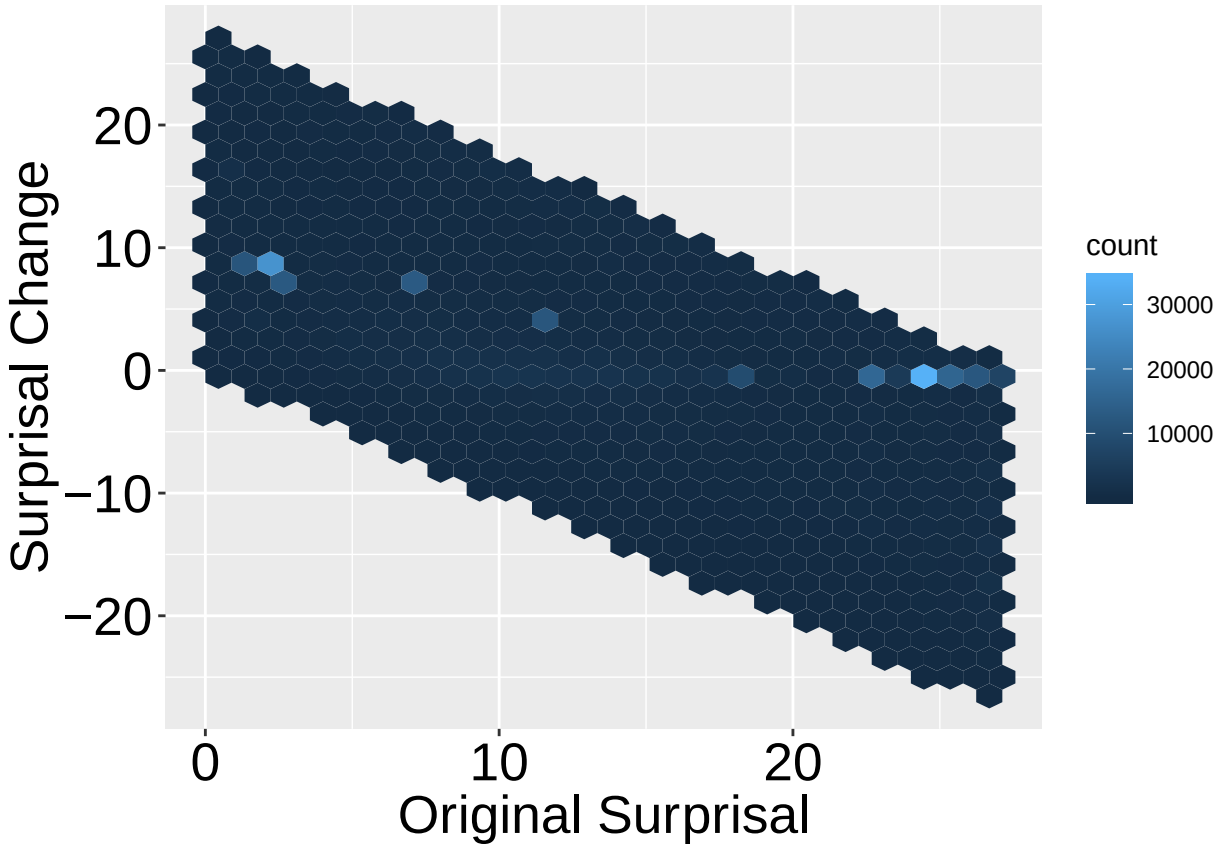
```
## Warning in n1 * n2: NAs produced by integer overflow
```

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




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.1887e+10, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##       -Inf -3.296511
## sample estimates:
## (pseudo)median
##      -3.339439
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.1887e+10, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -3.384040 -3.293456
## sample estimates:
## (pseudo)median
##      -3.339439
##
##
## Cliff's Delta
##
## delta estimate: -0.1415202 (negligible)
```

```
## 95 percent confidence interval:
##        inf        sup
## -0.1441017 -0.1389368
```




```
## [1] " ------ Expression Cache Model ----- "
## [1] "VarSwapWithinTopCacheExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -158.93, df = 384790, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##       -Inf -1.334067
## sample estimates:
## mean of the differences
##                -1.35878
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -158.93, df = 384790, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -1.38530 -1.33226
## sample estimates:
```

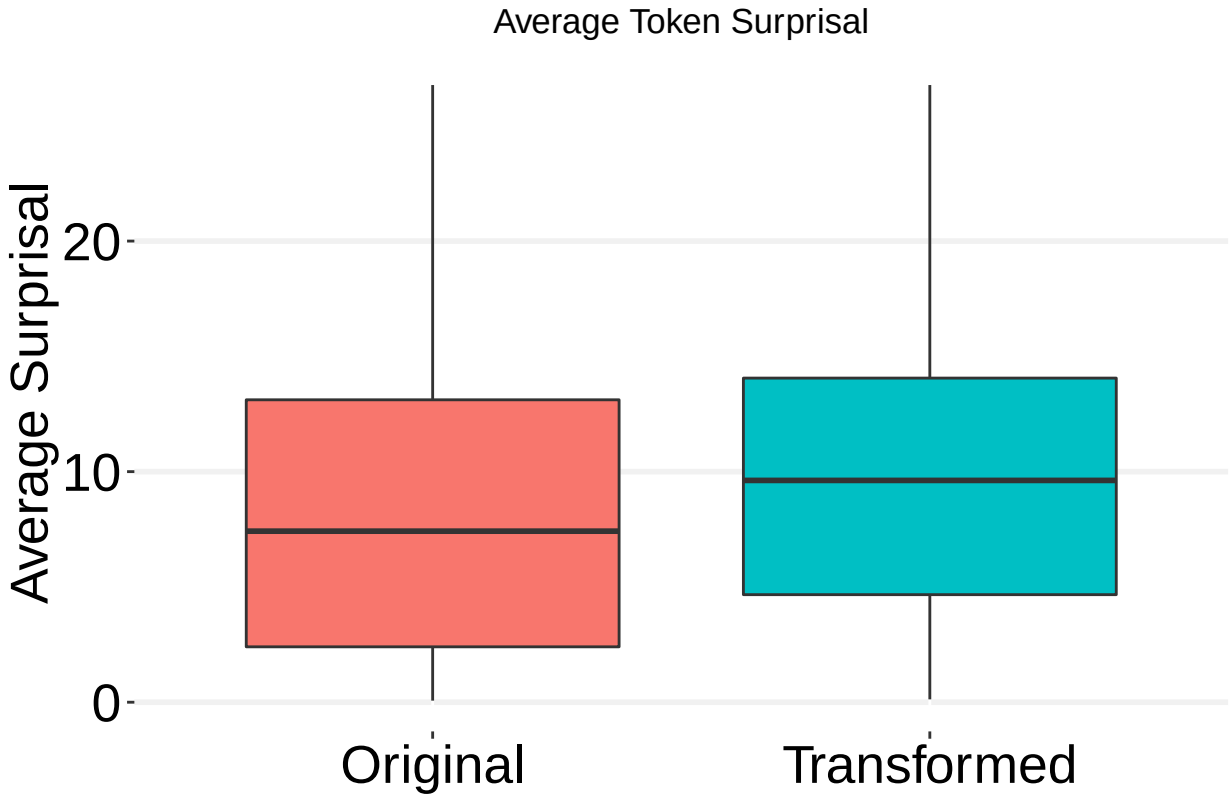
```
## mean of the differences
##               -1.35878
```

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

```
##
## Cohen's d
##
## d estimate: -0.2562058 (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
```



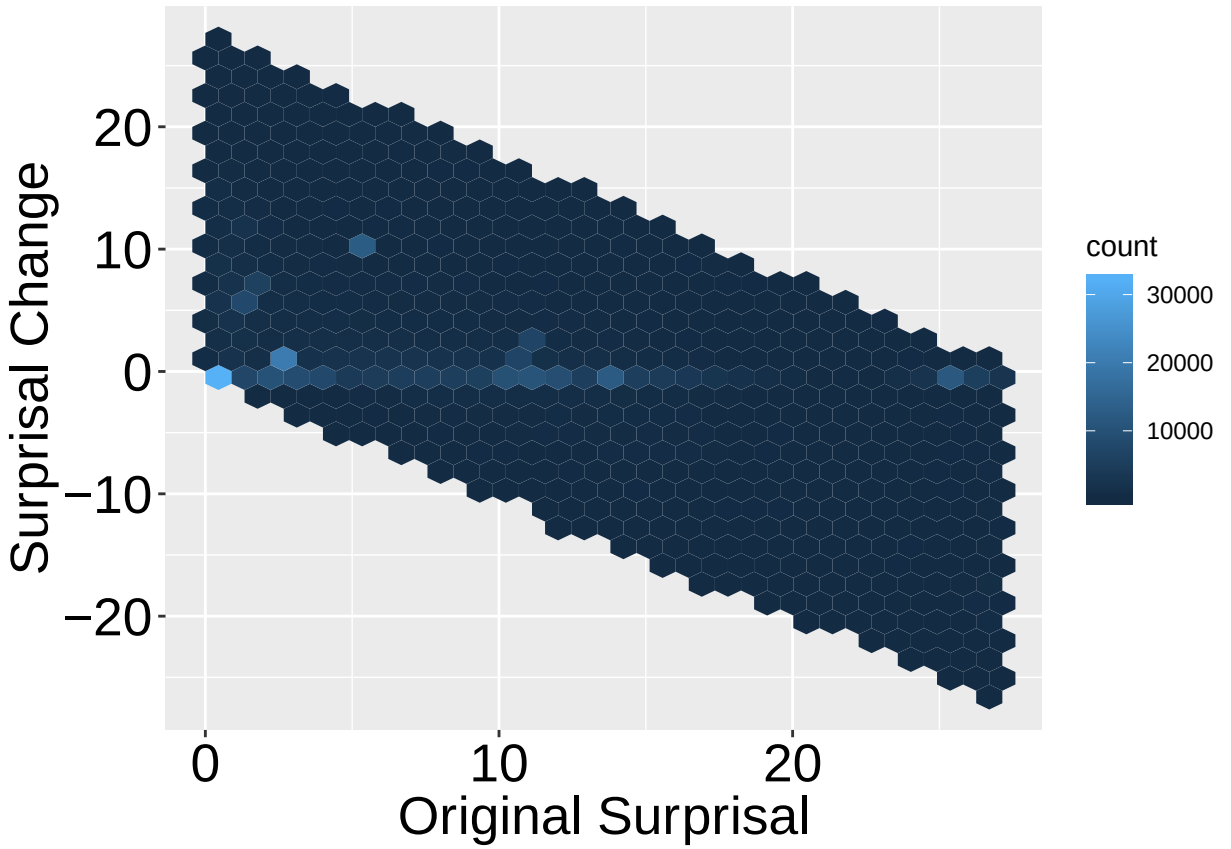

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.2e+10, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##       -Inf -1.151606
## sample estimates:
## (pseudo)median
##      -1.178904
##
##
```

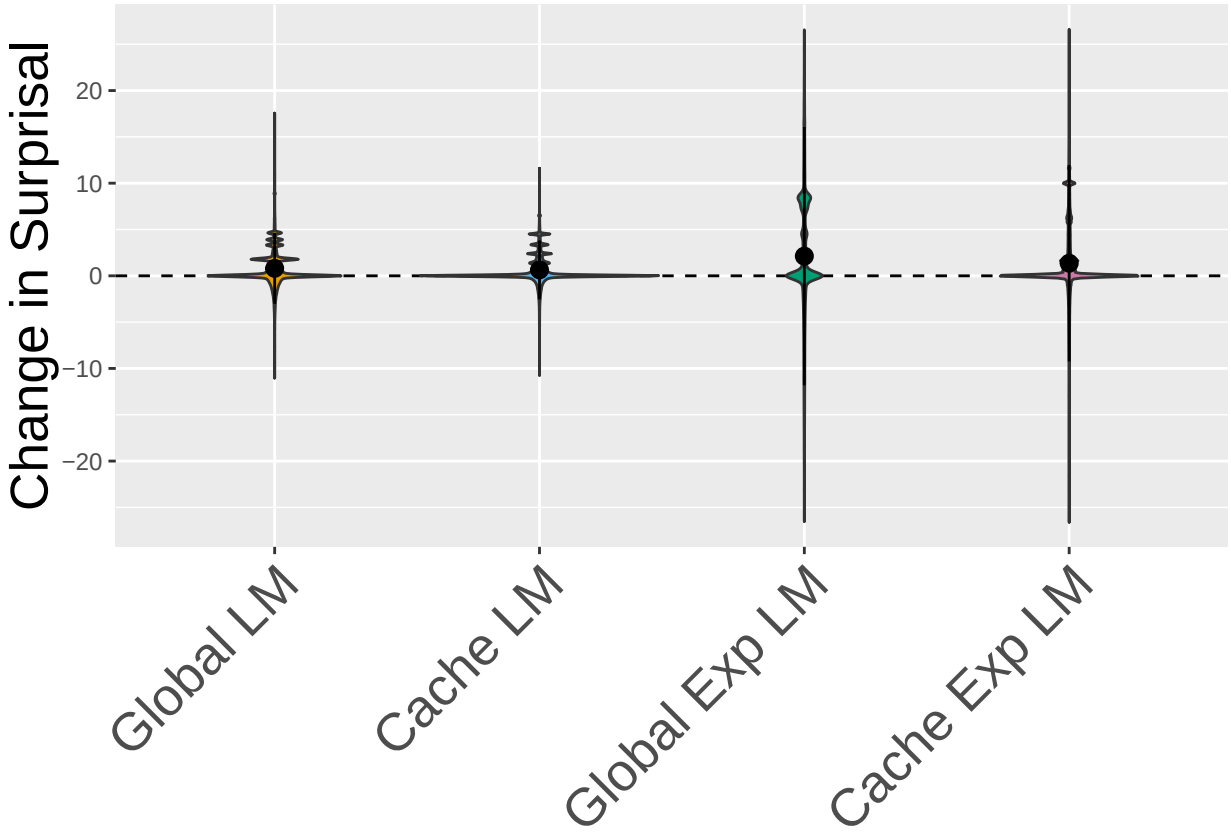
```
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.2e+10, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -1.208912 -1.149633
## sample estimates:
## (pseudo)median
##      -1.178904
##
##
## Cliff's Delta
##
## delta estimate: -0.1404018 (negligible)
## 95 percent confidence interval:
##        inf        sup
## -0.1429609 -0.1378408
```

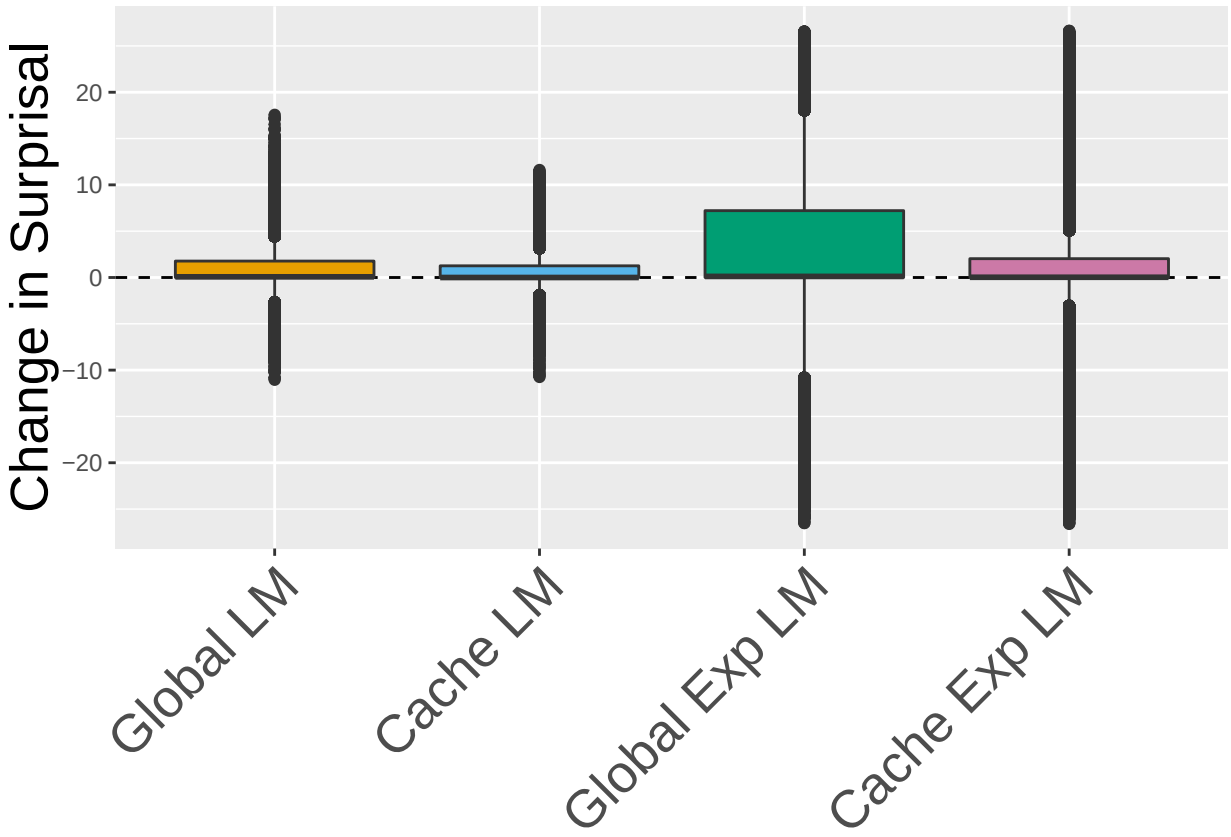
```
## [1] "Binary differences"
##
##  FALSE   TRUE
##  90663 294124
##
##  FALSE   TRUE
## 106138 278649
##
##  FALSE   TRUE
##  92520 292267
##
##  FALSE   TRUE
##  88472 296315
```

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

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


Surprisal Change
20
10
0
−10
−20
0
10
20
Original Surprisal
count
30000
20000
10000

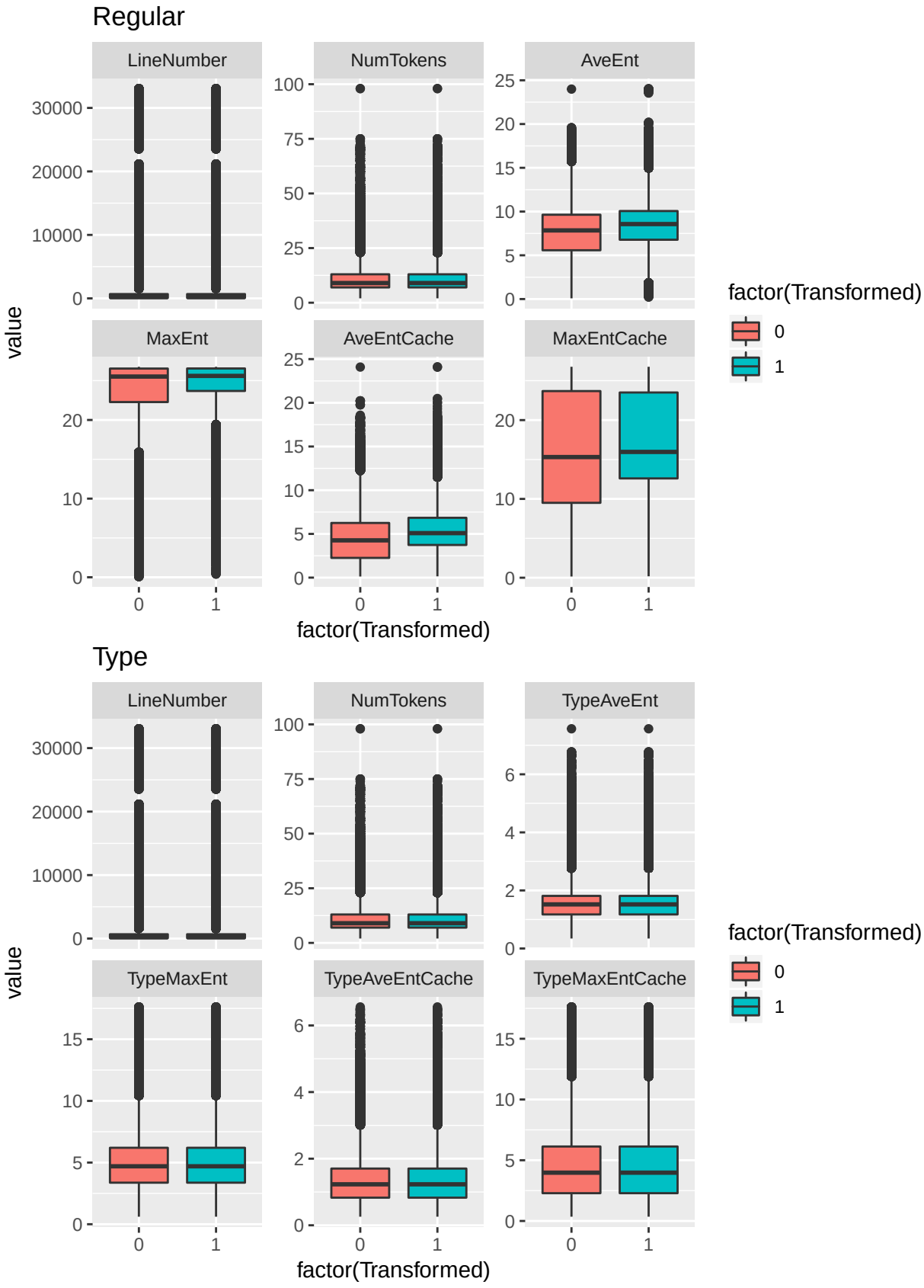

Change in Surprisal
20
10
0
−10
−20
Global LM
Cache LM
Global Exp LM
Cache Exp LM

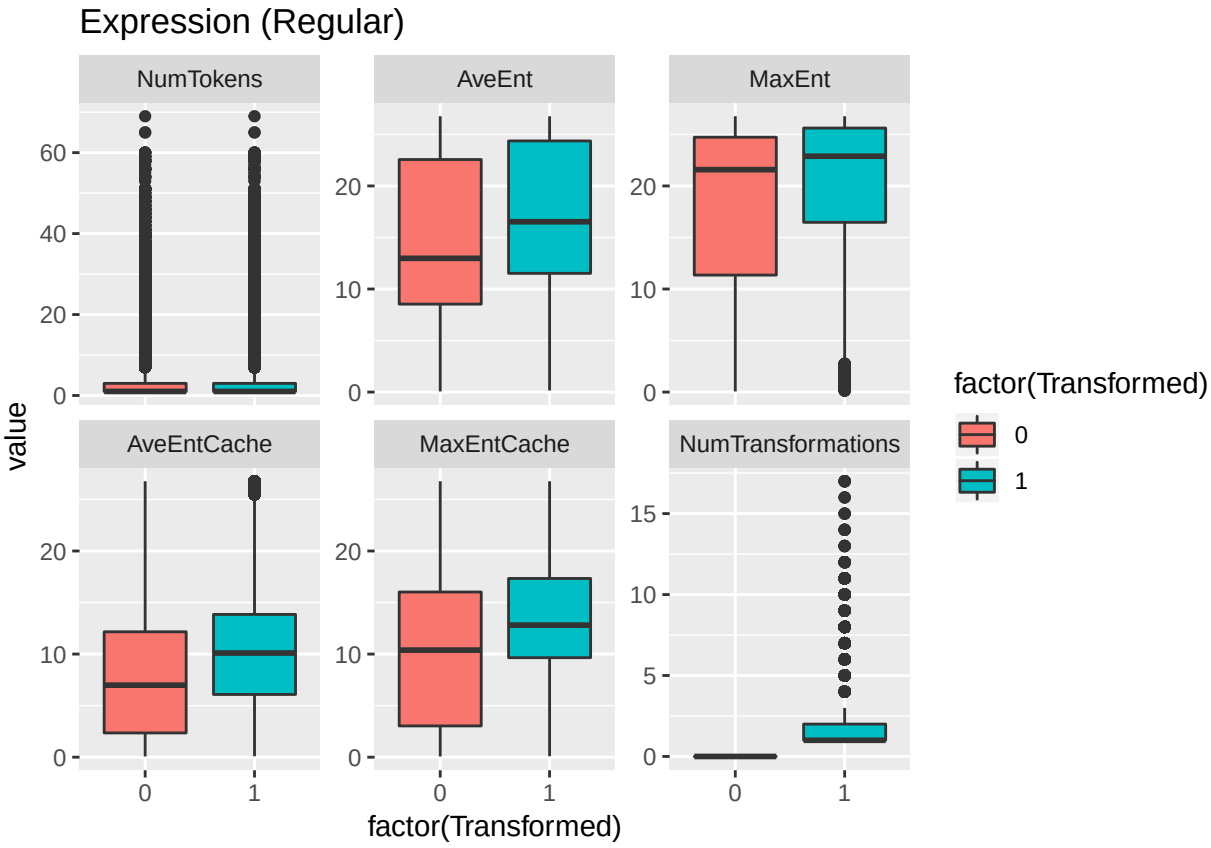


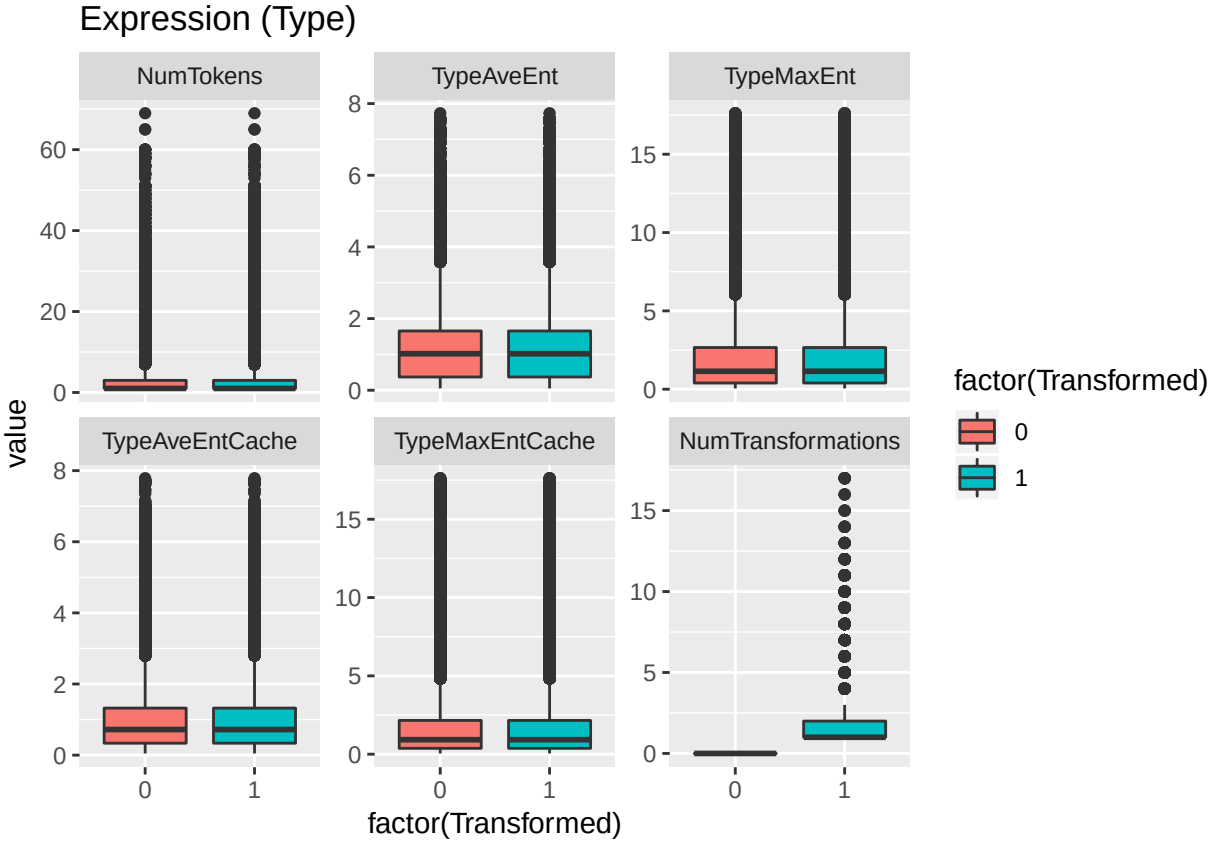

```
setwd("/data/anon/SemanticTransformation")
#dVsp2 <- compareDepthSummaryVar("variable_rename.csv", "VarSwapBetween", "SAME")
dVsp2 <- compareDepthSummaryVar("variable_rename_topstarred.csv", "VarSwapBetweenTop", "SAME", FALSE)
```

```
## [[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"
```

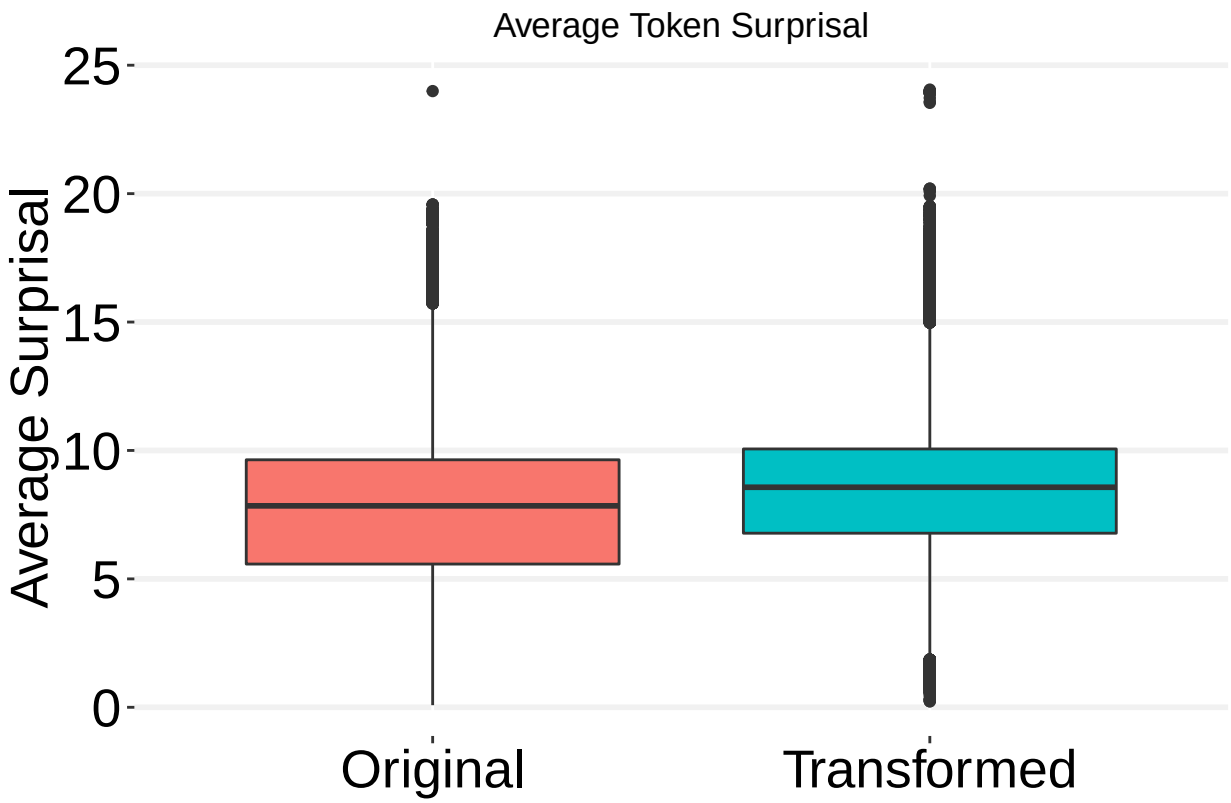

Regular
LineNumber
NumTokens
AveEnt
MaxEnt
AveEntCache
MaxEntCache
value
factor(Transformed)
factor(Transformed)
0
1
Type
LineNumber
NumTokens
TypeAveEnt
TypeMaxEnt
TypeAveEntCache
TypeMaxEntCache
value
factor(Transformed)
factor(Transformed)
0
1





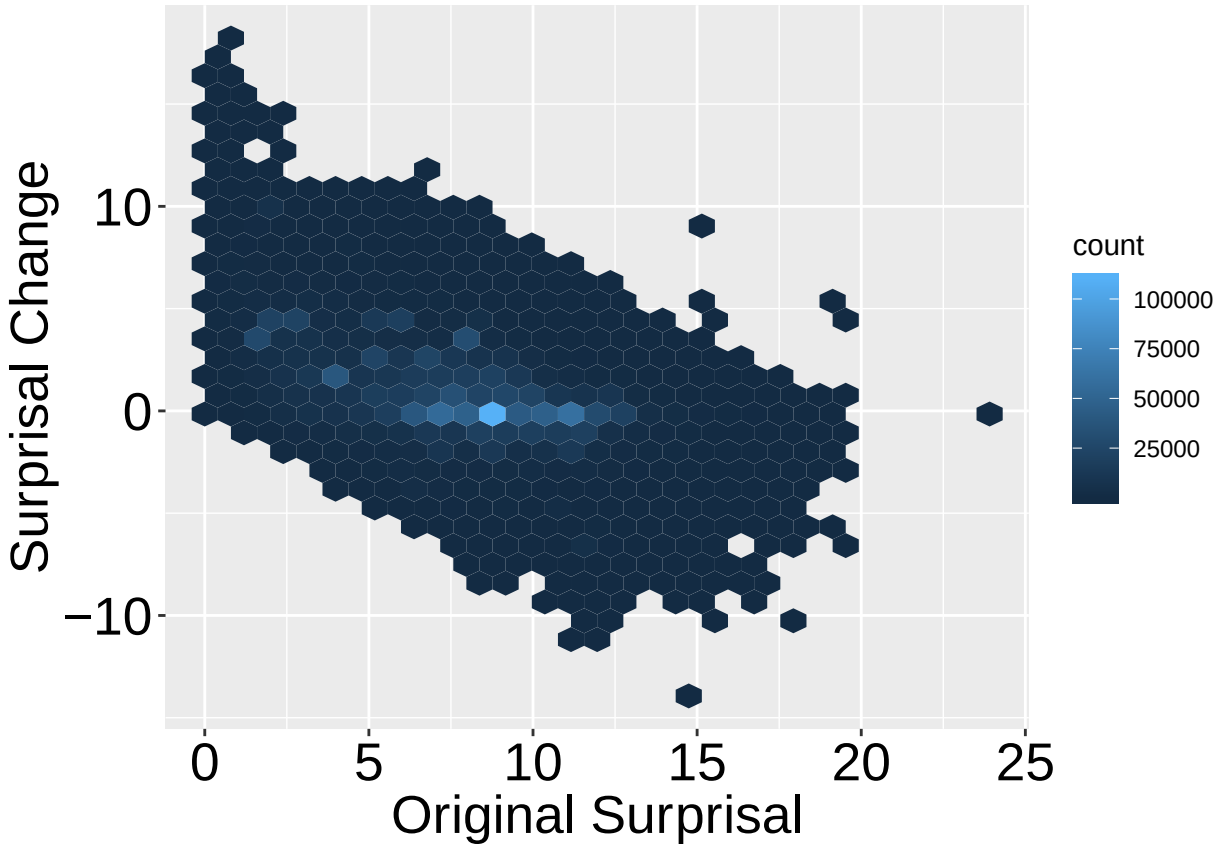

```
## [1] " ------ Global Model ----- "
```

```
## [1] "VarSwapBetweenTopGlobal Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -516.49, df = 1458300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.8671276
## sample estimates:
## mean of the differences
##              -0.8720078
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -516.49, df = 1458300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.8772448 -0.8667708
## sample estimates:
## mean of the differences
##              -0.8720078
## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -0.427695 (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 = 1.8595e+11, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.9381055
## sample estimates:
## (pseudo)median
##     -0.9436151
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.8595e+11, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.9491099 -0.9376713
## sample estimates:
## (pseudo)median
##     -0.9436151
##
##
## Cliff's Delta
##
## delta estimate: -0.1610196 (small)
```

```
## 95 percent confidence interval:
##        inf        sup
## -0.1623245 -0.1597141
```



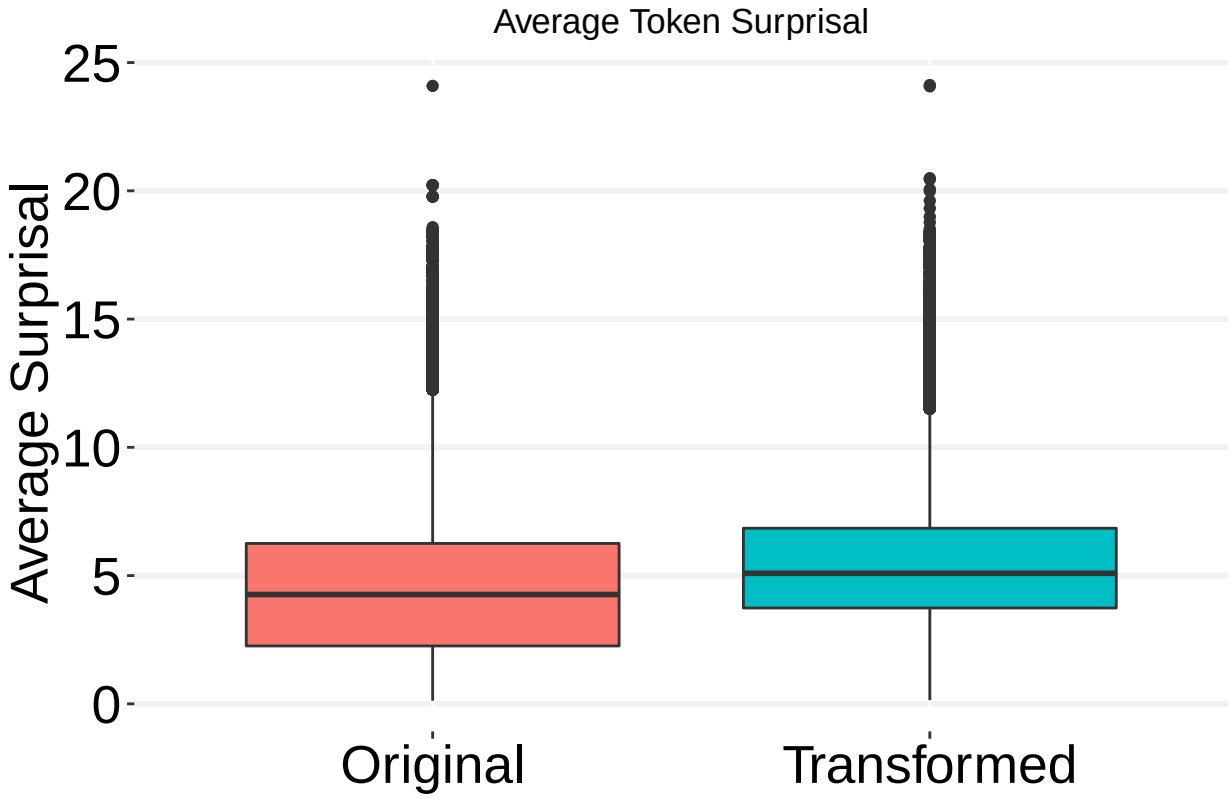

```
## [1] " ------ Cache Model ----- "
## [1] "VarSwapBetweenTopCache Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -630.64, df = 1458300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.8499246
## sample estimates:
## mean of the differences
##              -0.8538381
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -630.64, df = 1458300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.8580378 -0.8496384
## sample estimates:
```

```
## mean of the differences
##              -0.8538381
```

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

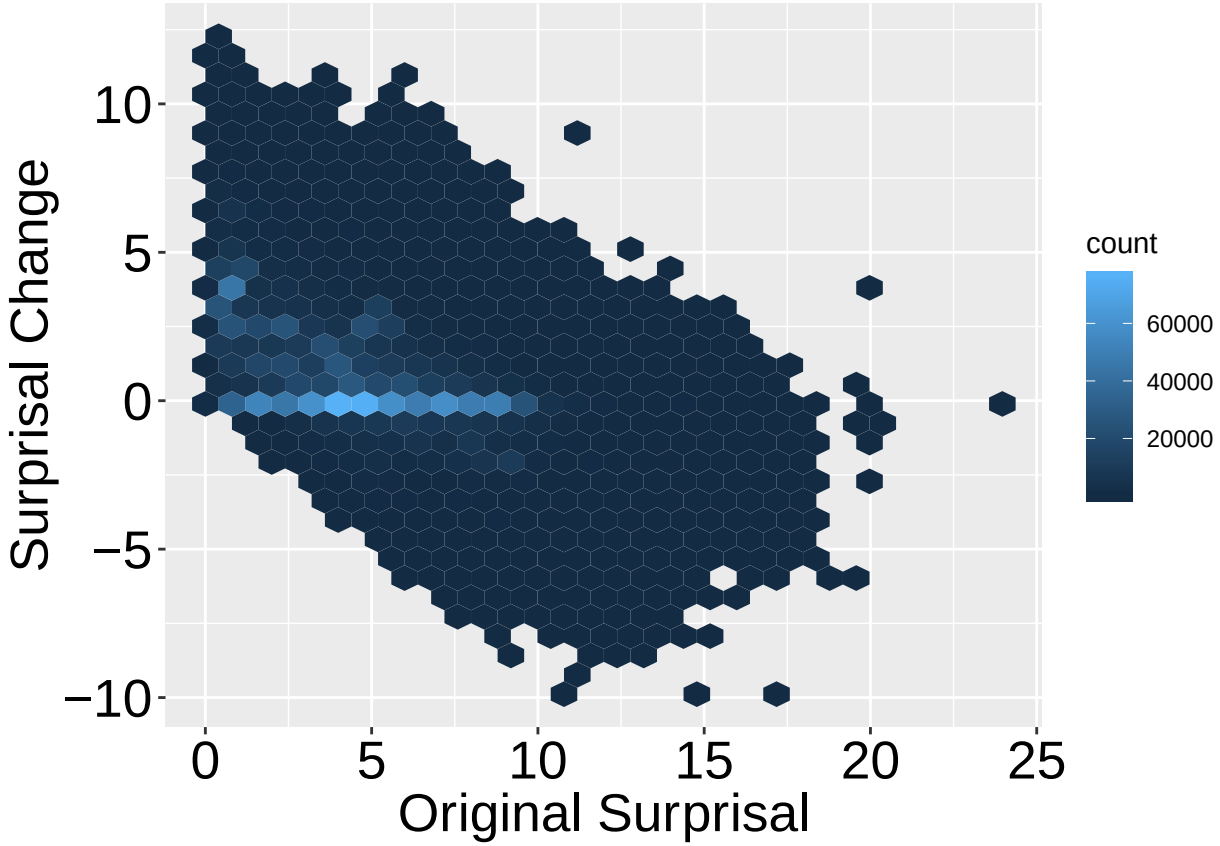
```
##
## Cohen's d
##
## d estimate: -0.522221 (medium)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

```
## 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 = 1.6759e+11, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.8741584
## sample estimates:
## (pseudo)median
##     -0.8797104
##
##
```

```
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.6759e+11, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.8857167 -0.8737523
## sample estimates:
## (pseudo)median
##     -0.8797104
##
##
## Cliff's Delta
##
## delta estimate: -0.2001322 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.2014318 -0.1988319
```




```
## [1] " ------ Expression Global Model ----- "
## [1] "VarSwapBetweenTopGlobalExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -408.63, df = 1458300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
```

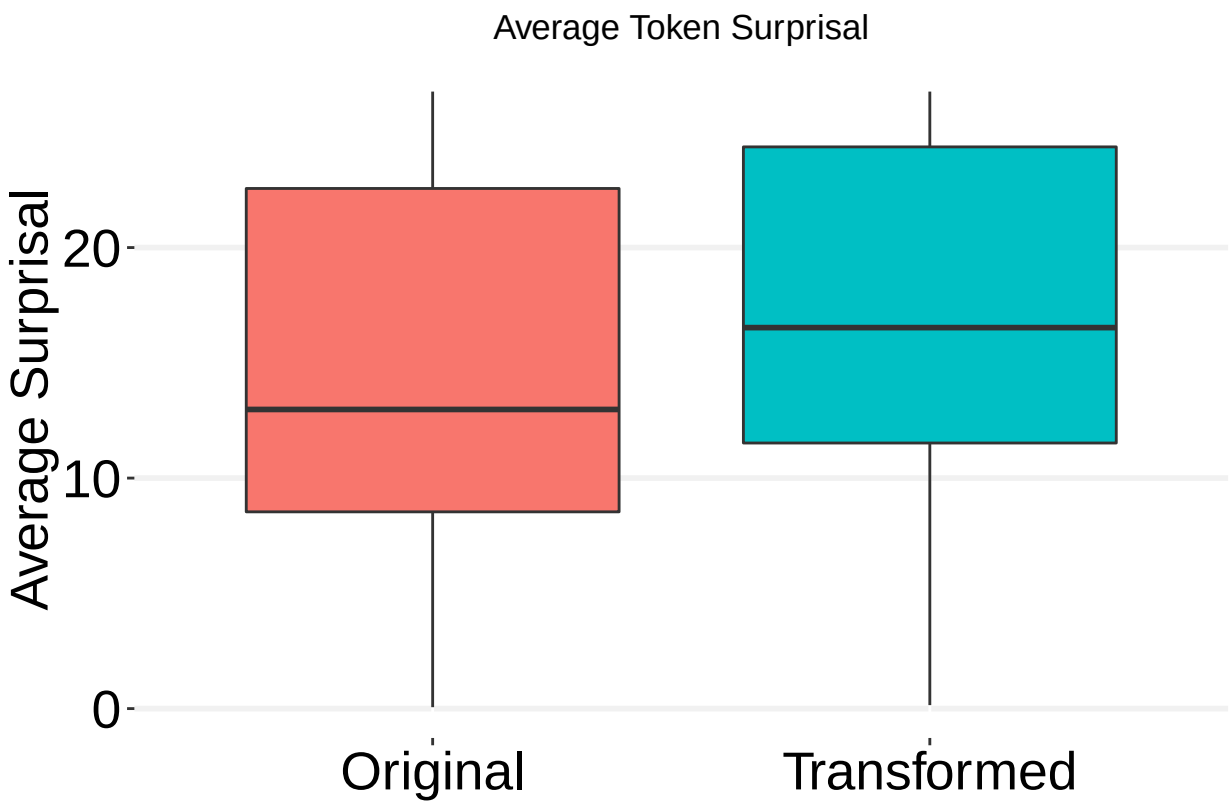
```
## 99.80769 percent confidence interval:
##      -Inf -2.77364
## sample estimates:
## mean of the differences
##                 -2.7934
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -408.63, df = 1458300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -2.814605 -2.772195
## sample estimates:
## mean of the differences
##                 -2.7934
```

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

```
##
## Cohen's d
##
## d estimate: -0.3383737 (small)
## 95 percent confidence interval:
## inf sup
##  NA  NA
```

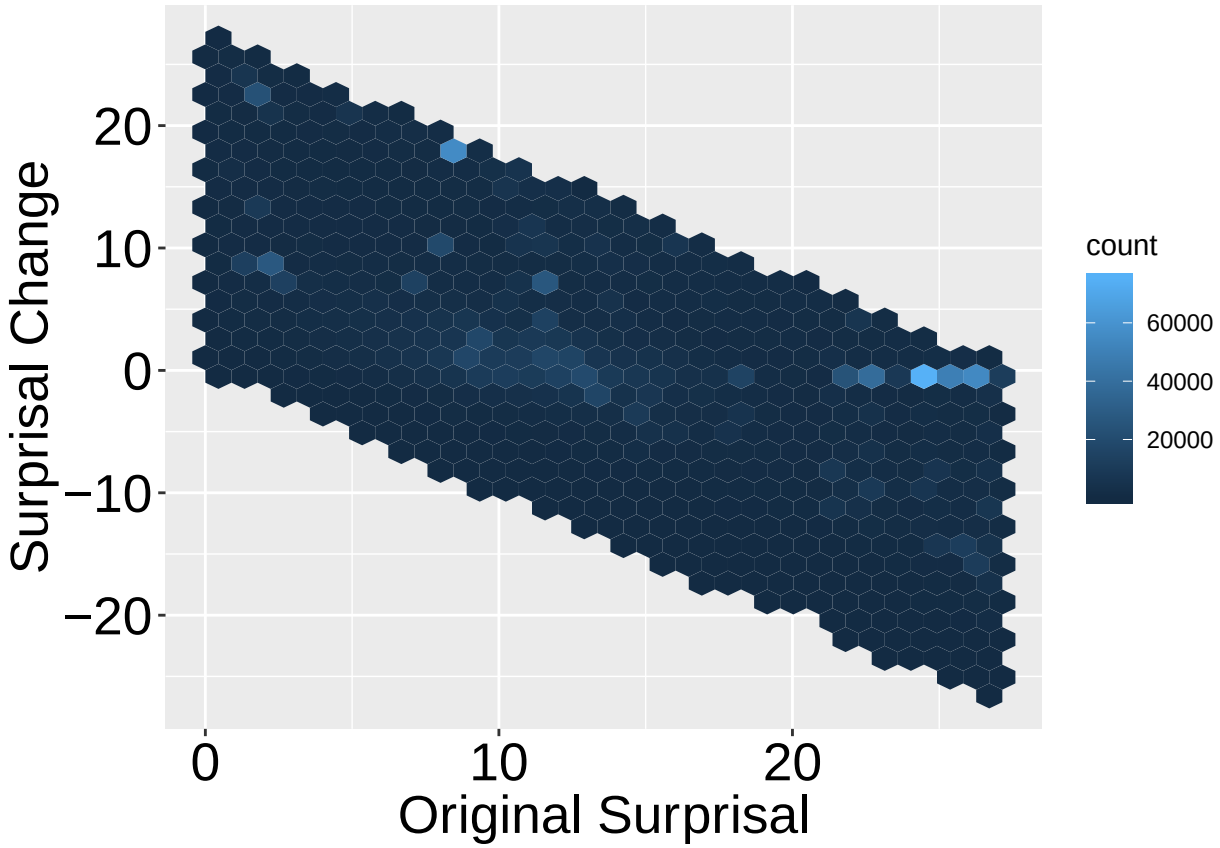
```
## Warning in n1 * n2: NAs produced by integer overflow
```

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




```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 2.1951e+11, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##       -Inf -3.108418
## sample estimates:
## (pseudo)median
##      -3.128779
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 2.1951e+11, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -3.154807 -3.106754
## sample estimates:
## (pseudo)median
##      -3.128779
##
##
## Cliff's Delta
##
## delta estimate: -0.2182133 (small)
```

```
## 95 percent confidence interval:
##        inf        sup
## -0.2195065 -0.2169194
```




```
## [1] " ------ Expression Cache Model ----- "
## [1] "VarSwapBetweenTopCacheExp Original < Transformed"
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -479.07, df = 1458300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##       -Inf -2.394541
## sample estimates:
## mean of the differences
##               -2.409077
##
##
##  Paired t-test
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -479.07, df = 1458300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -2.424675 -2.393479
## sample estimates:
```

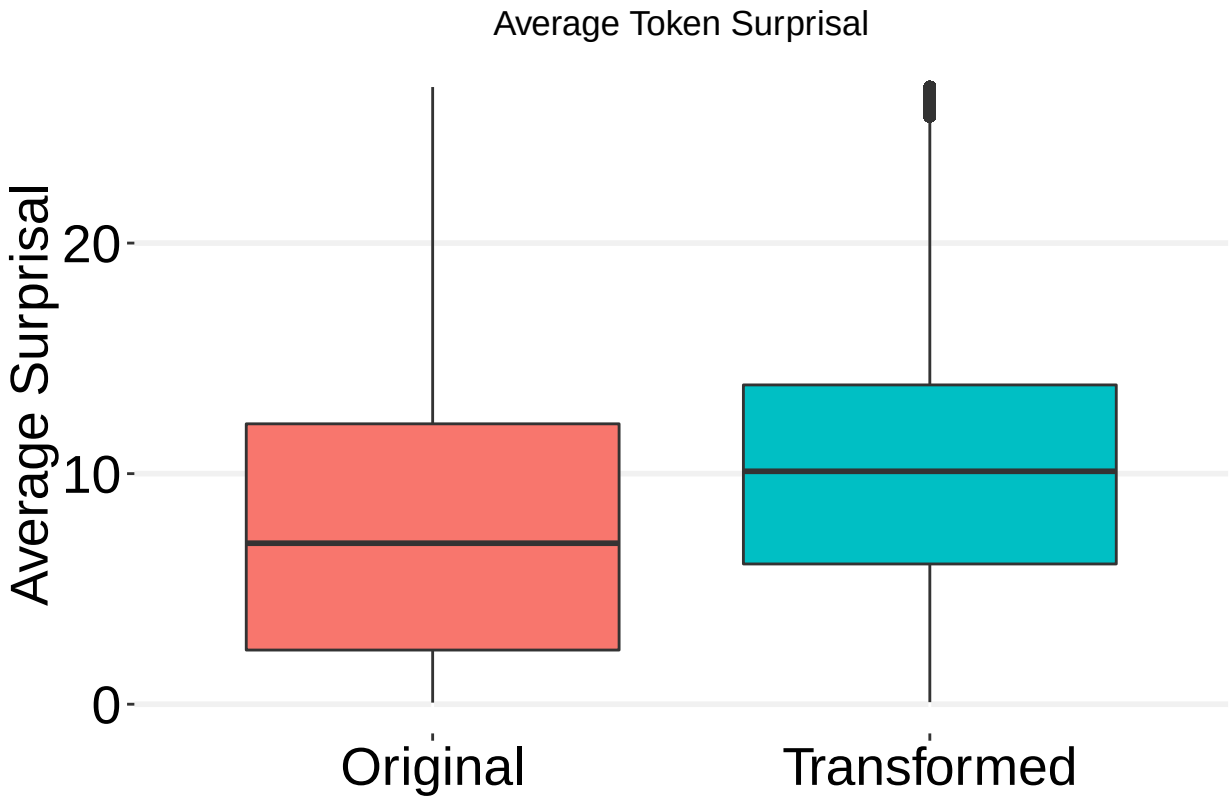
```
## mean of the differences
##               -2.409077
```

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

```
##
## Cohen's d
##
## d estimate: -0.3967104 (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
```



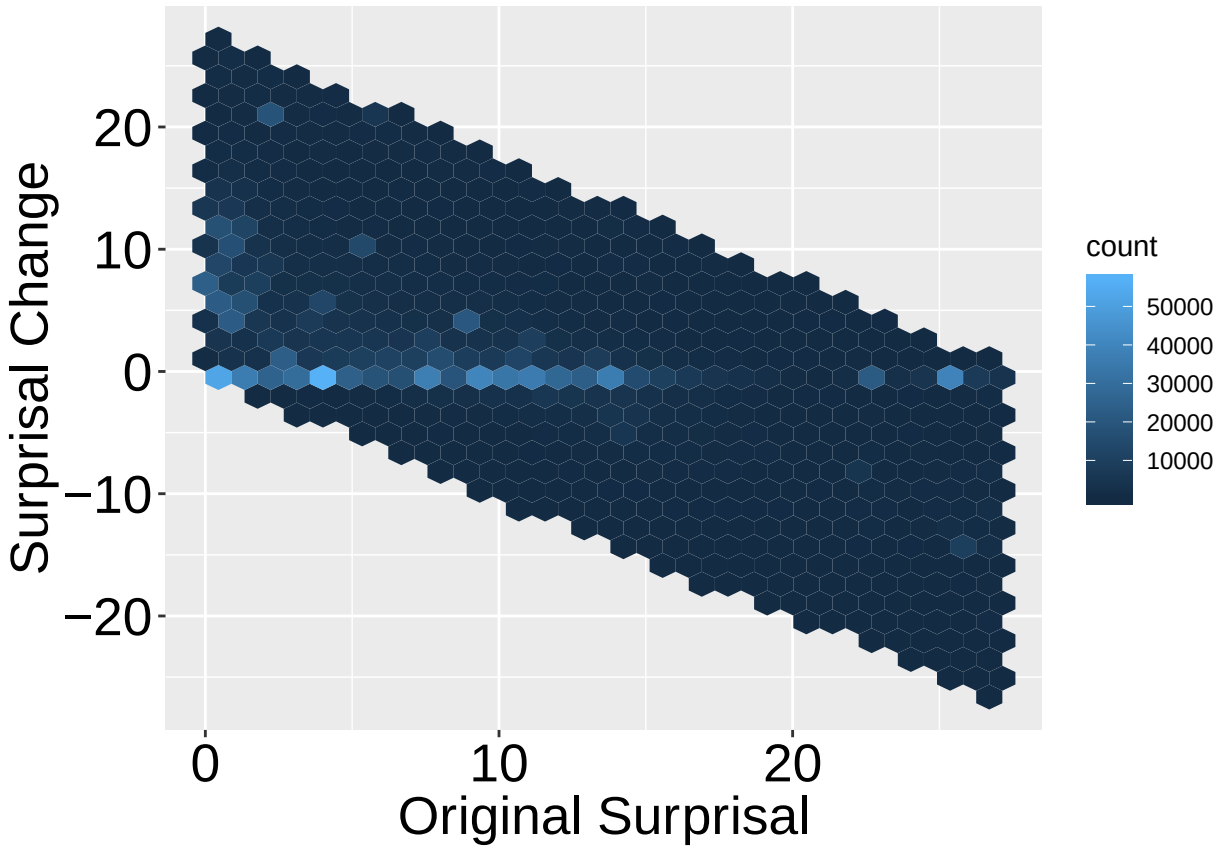

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.8568e+11, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##       -Inf -2.234505
## sample estimates:
## (pseudo)median
##      -2.247741
##
##
```

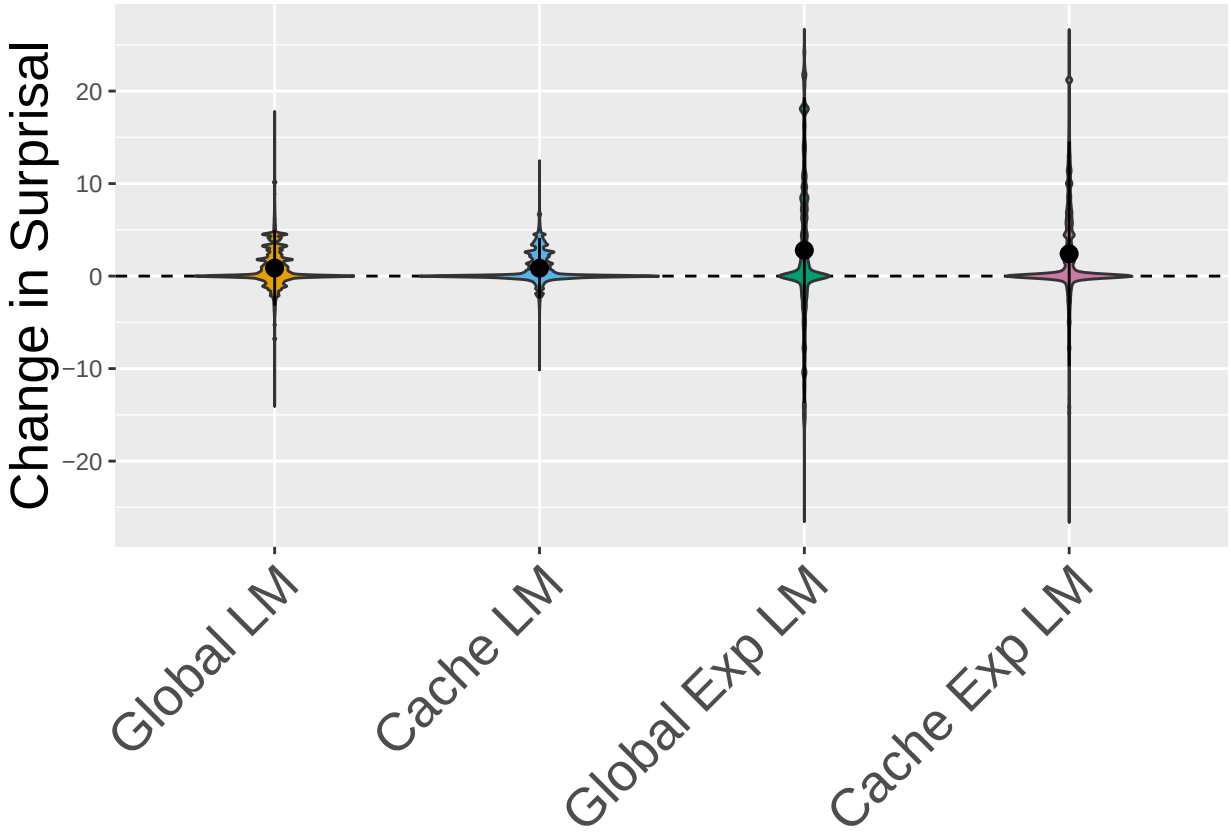
```
##  Wilcoxon signed rank test with continuity correction
##
## data:  diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1.8568e+11, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -2.262323 -2.233519
## sample estimates:
## (pseudo)median
##      -2.247741
##
##
## Cliff's Delta
##
## delta estimate: -0.2462787 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.2475647 -0.2449918
```

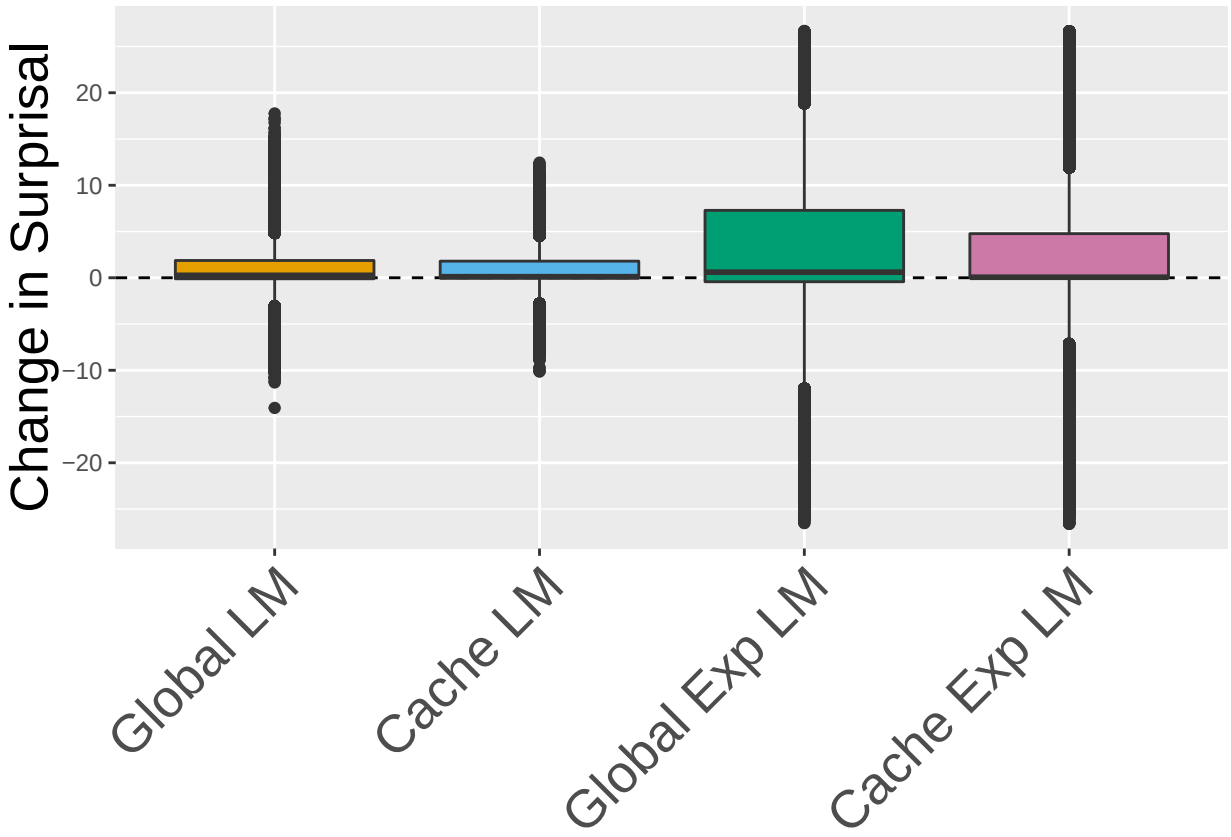
```
## [1] "Binary differences"
##
##   FALSE    TRUE
##  406792 1051548
##
##   FALSE    TRUE
##  415332 1043008
##
##   FALSE    TRUE
##  424894 1033446
##
##   FALSE    TRUE
##  381709 1076631
```

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

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


Surprisal Change
20
10
0
−10
−20
0
10
20
Original Surprisal
count
50000
40000
30000
20000
10000


Change in Surprisal
20
10
0
−10
−20
Global LM
Cache LM
Global Exp LM
Cache Exp 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?
dVspFiltered <- generateFilteredResults(dVsp, "Var", "VarWithinTopFiltered100", 100)
dVspFiltered2 <- generateFilteredResults(dVsp, "Var", "VarWithinTopFiltered10", 10)
```

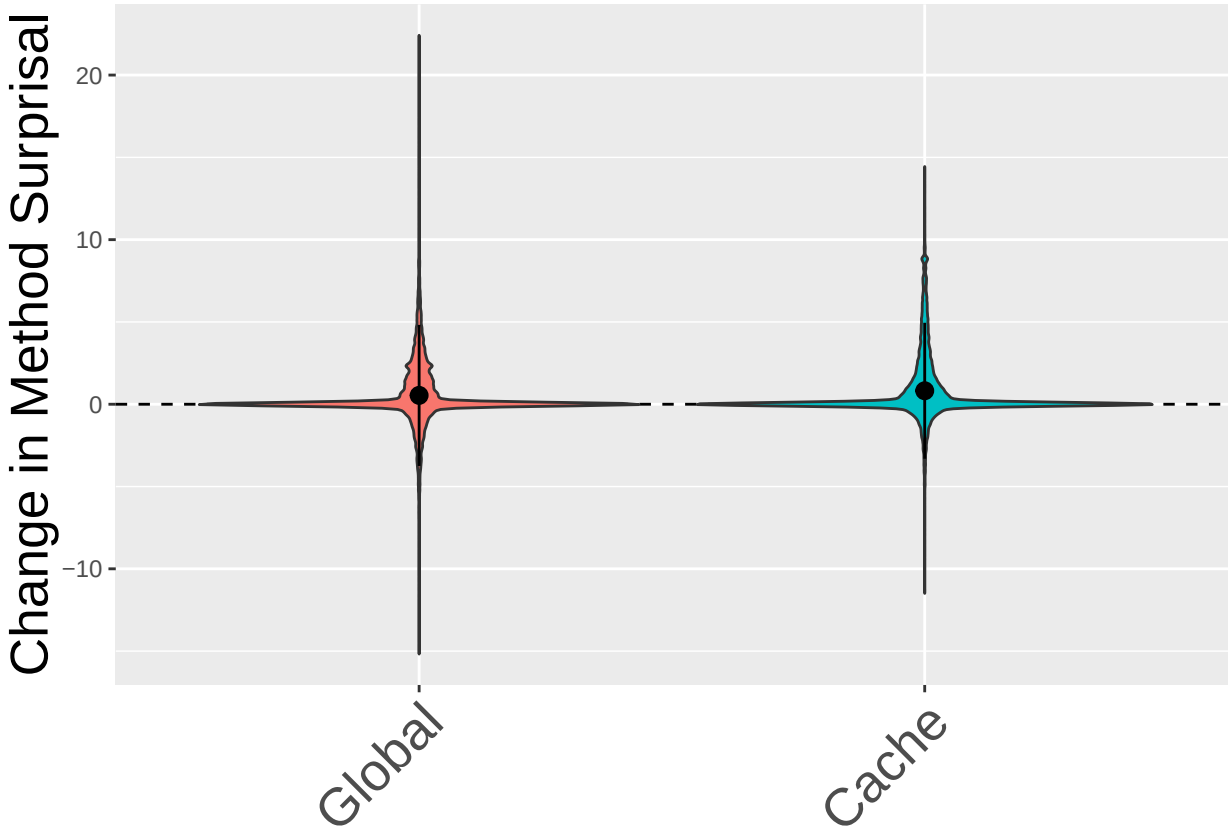
```
setwd("/data/anon/SemanticTransformation/")

#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?
dVsp2Filtered <- generateFilteredResults(dVsp2, "Var", "VarBetweenTopFiltered100", 100)
dVsp2Filtered2 <- generateFilteredResults(dVsp2, "Var", "VarBetweenTopFiltered10", 10)
```

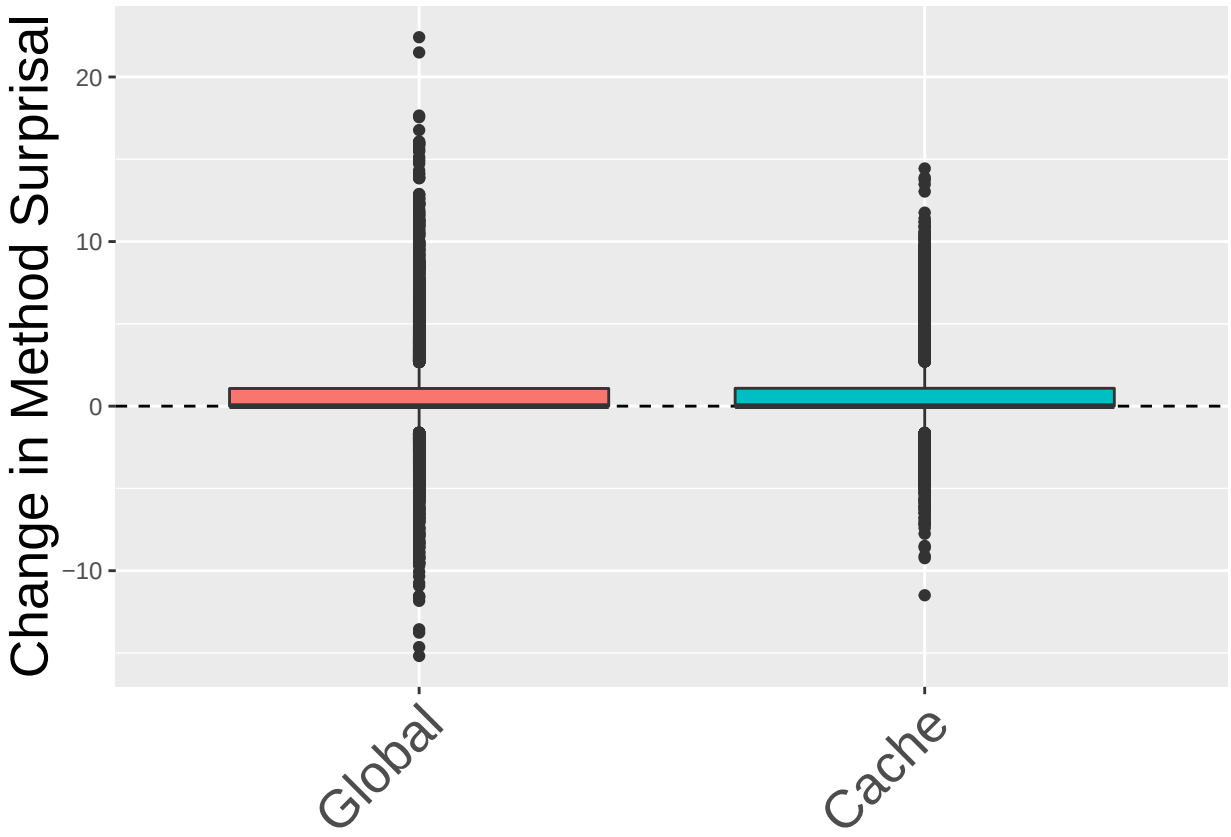
```
setwd("/data/anon/SemanticTransformation")
#Do two the || nature of this - what is the average change for these IN THE WHOLE method
#usByFile = sqldf("SELECT filepath, avg(BaseAveEnt), avg(ChangeAveEnt), avg(AverageEntChange) FROM dVsp

fWithin1 <- getMethodAverage(dVspFiltered, "BaseAveEnt", "ChangeAveEnt", "AverageEntChange")
fWithin2 <- getMethodAverage(dVspFiltered, "BaseCacheAveEnt", "ChangeCacheAveEnt", "CacheAverageEntChan
fWithin3 <- getMethodAverage(dVspFiltered, "BaseAveEntExp", "ChangeAveEntExp", "AverageEntChangeExp")
fWithin4 <- getMethodAverage(dVspFiltered, "BaseCacheAveEntExp", "ChangeCacheAveEntExp", "CacheAverageE
```

```
fBetween1 <- getMethodAverage(dVsp2Filtered, "BaseAveEnt", "ChangeAveEnt", "AverageEntChange")
fBetween2 <- getMethodAverage(dVsp2Filtered, "BaseCacheAveEnt", "ChangeCacheAveEnt", "CacheAverageEntCha
fBetween3 <- getMethodAverage(dVsp2Filtered, "BaseAveEntExp", "ChangeAveEntExp", "AverageEntChangeExp")
fBetween4 <- getMethodAverage(dVsp2Filtered, "BaseCacheAveEntExp", "ChangeCacheAveEntExp", "CacheAverag

#Boxplot
diffMerge = fWithin1
#Add in averages/maxes of base + change + diff
diffMerge$AveCacheBefore = fWithin2$AveBefore
diffMerge$AveCacheAfter = fWithin2$AveAfter
diffMerge$AveCacheDiff = fWithin2$AveDiff

diffMerge$AveExpBefore = fWithin3$AveBefore
diffMerge$AveExpAfter = fWithin3$AveAfter
diffMerge$AveExpDiff = fWithin3$AveDiff


diffMerge$AveCacheExpBefore = fWithin4$AveBefore
diffMerge$AveCacheExpAfter = fWithin4$AveAfter
diffMerge$AveCacheExpDiff = fWithin4$AveDiff
boxplotAvgCompareExp(diffMerge, "VarSwapWithin", "Method")
```

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

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

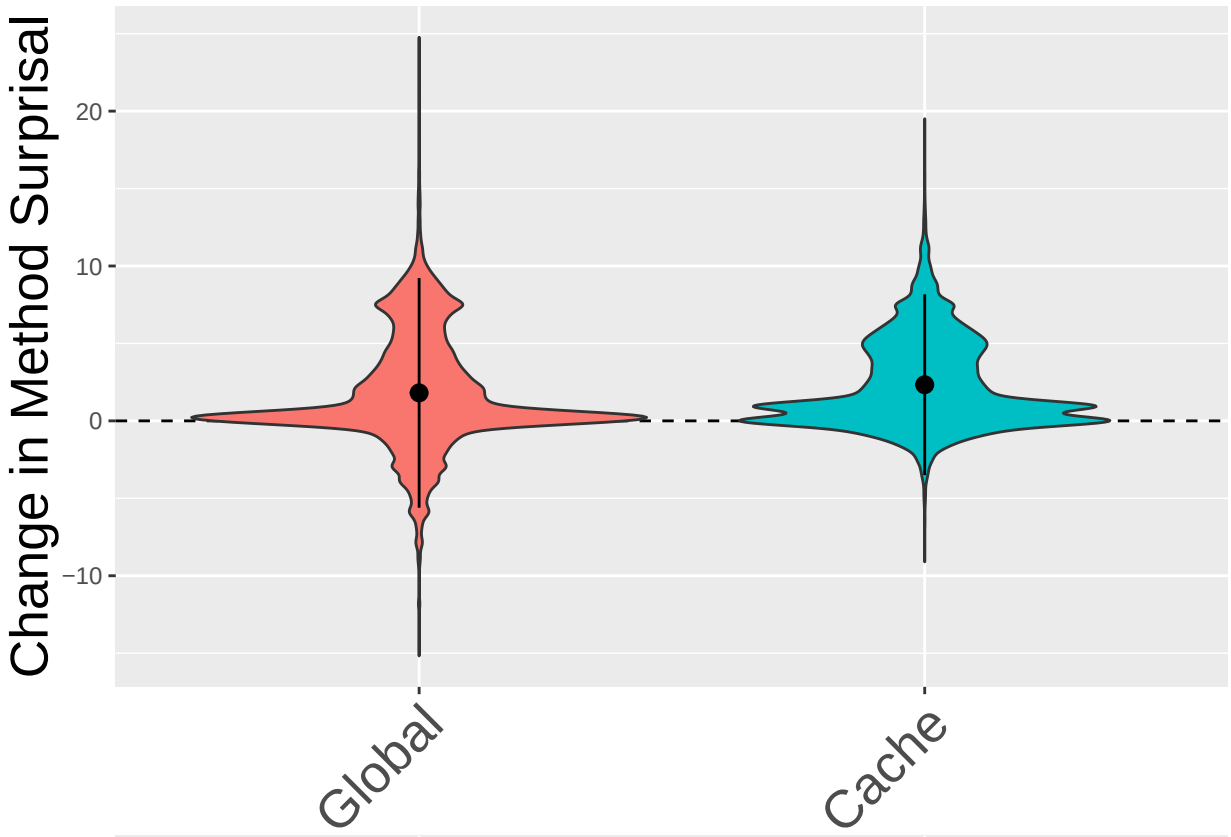


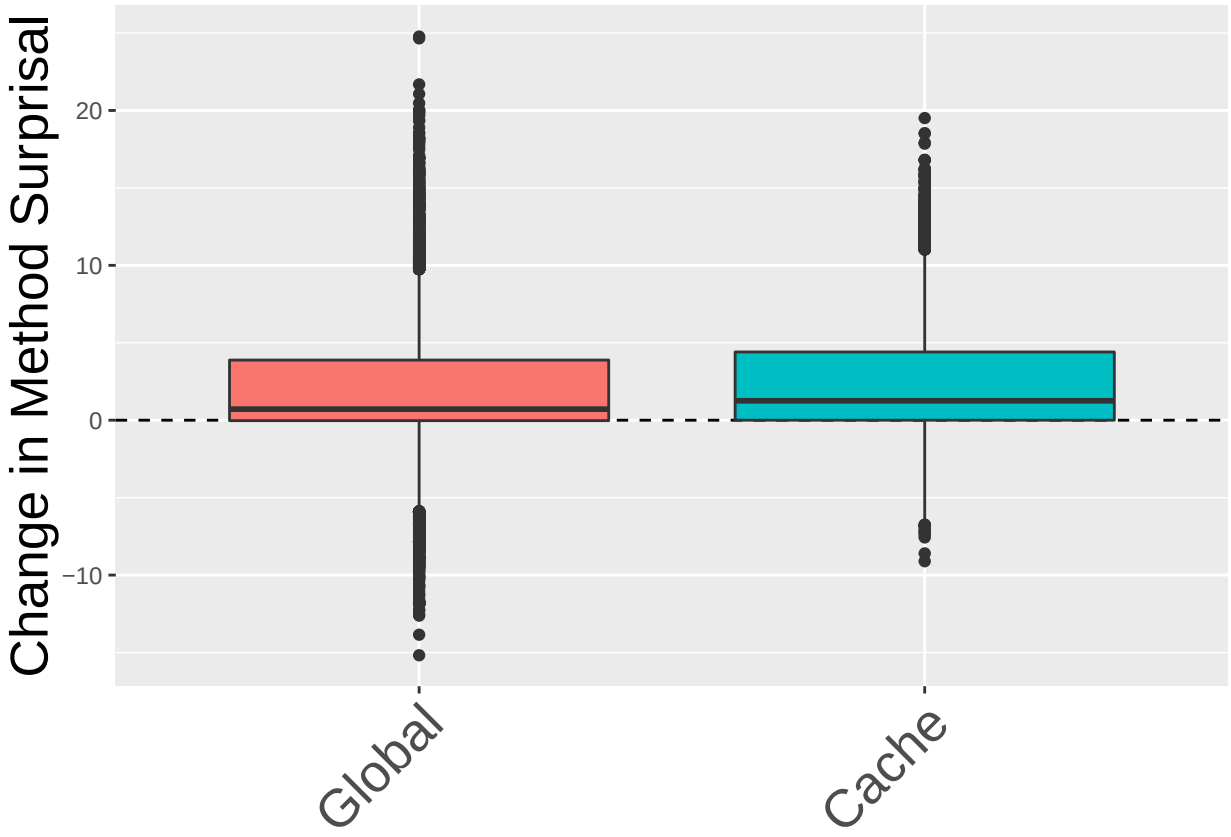



```
diffMergeBet = fBetween1
diffMergeBet$AveCacheBefore = fBetween2$AveBefore
diffMergeBet$AveCacheAfter = fBetween2$AveAfter
diffMergeBet$AveCacheDiff = fBetween2$AveDiff

diffMergeBet$AveExpBefore = fBetween3$AveBefore
diffMergeBet$AveExpAfter = fBetween3$AveAfter
diffMergeBet$AveExpDiff = fBetween3$AveDiff


diffMergeBet$AveCacheExpBefore = fBetween4$AveBefore
diffMergeBet$AveCacheExpAfter = fBetween4$AveAfter
diffMergeBet$AveCacheExpDiff = fBetween4$AveDiff
boxplotAvgCompareExp(diffMergeBet, "VarSwapBetween", "Method")
```

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

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

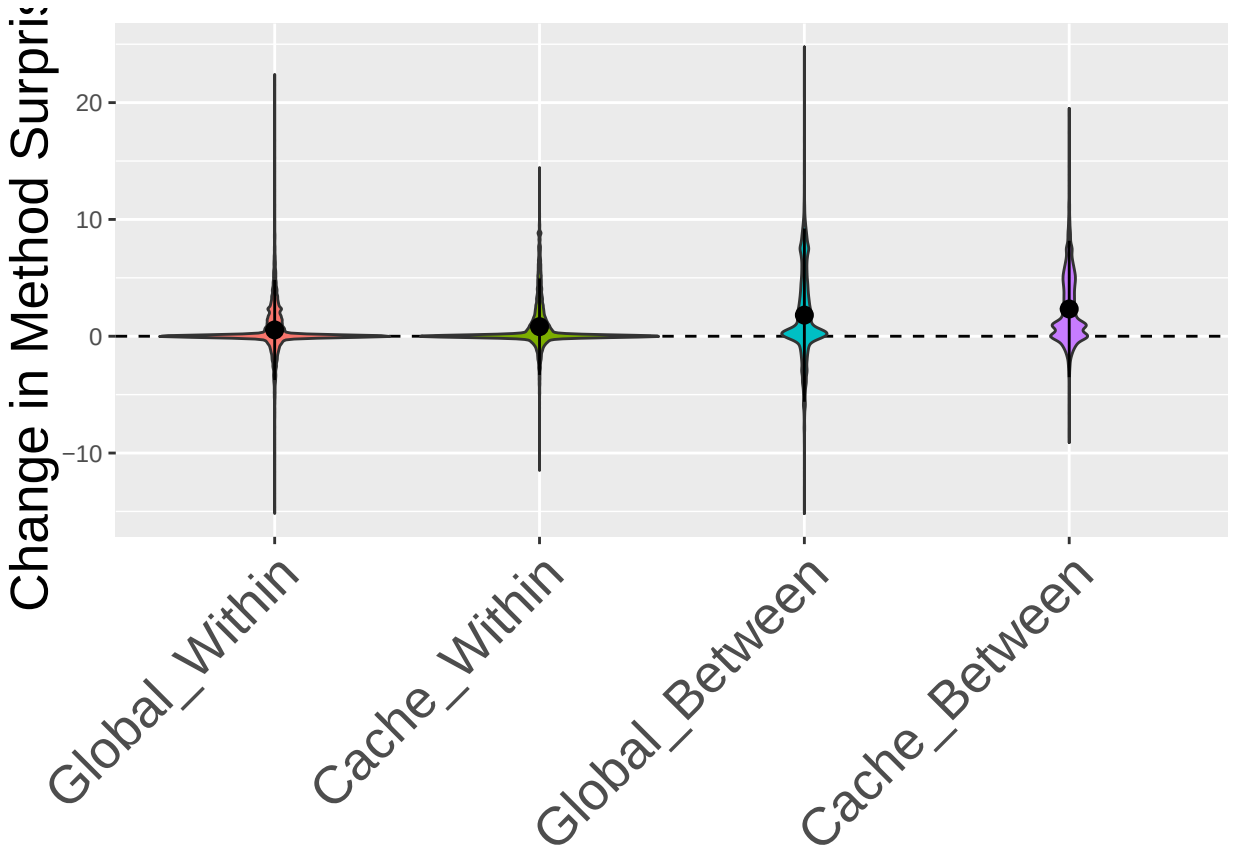

Change in Method Surprisal
20
10
0
−10
Global
Cache

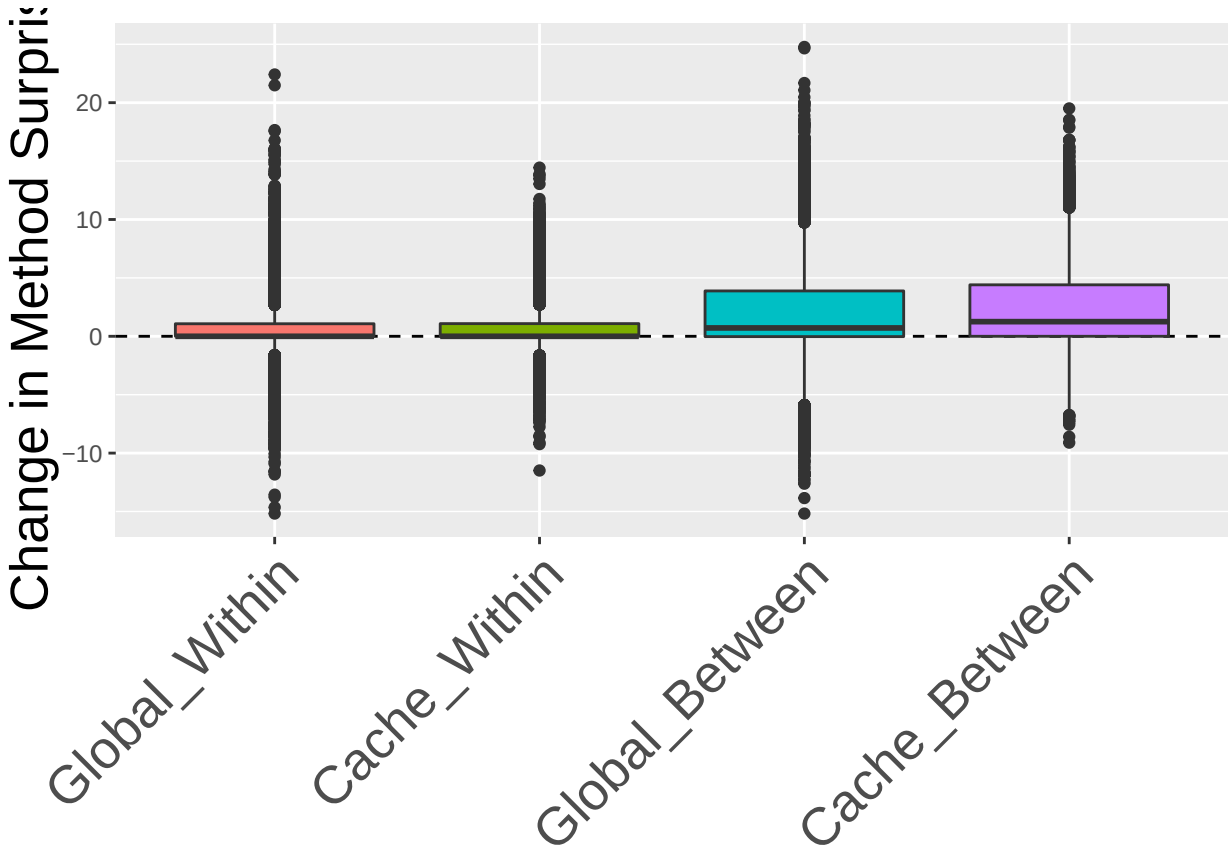

Change in Method Surprisal
20
10
0
−10
Global
Cache

```
boxplotAvgVarSwap(diffMerge, diffMergeBet, "VarSwapAll", "Method")
```

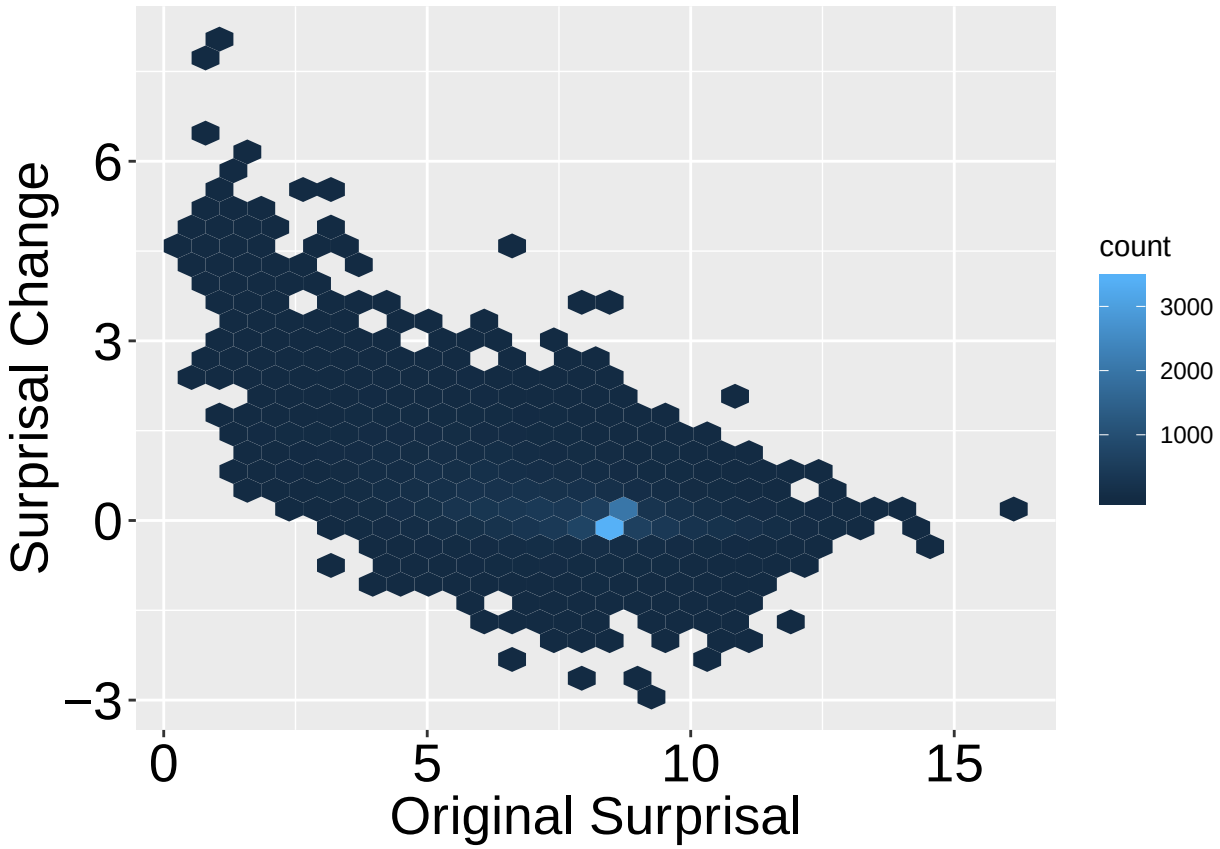
```
## No id variables; using all as measure variables
## No id variables; using all as measure variables
```

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

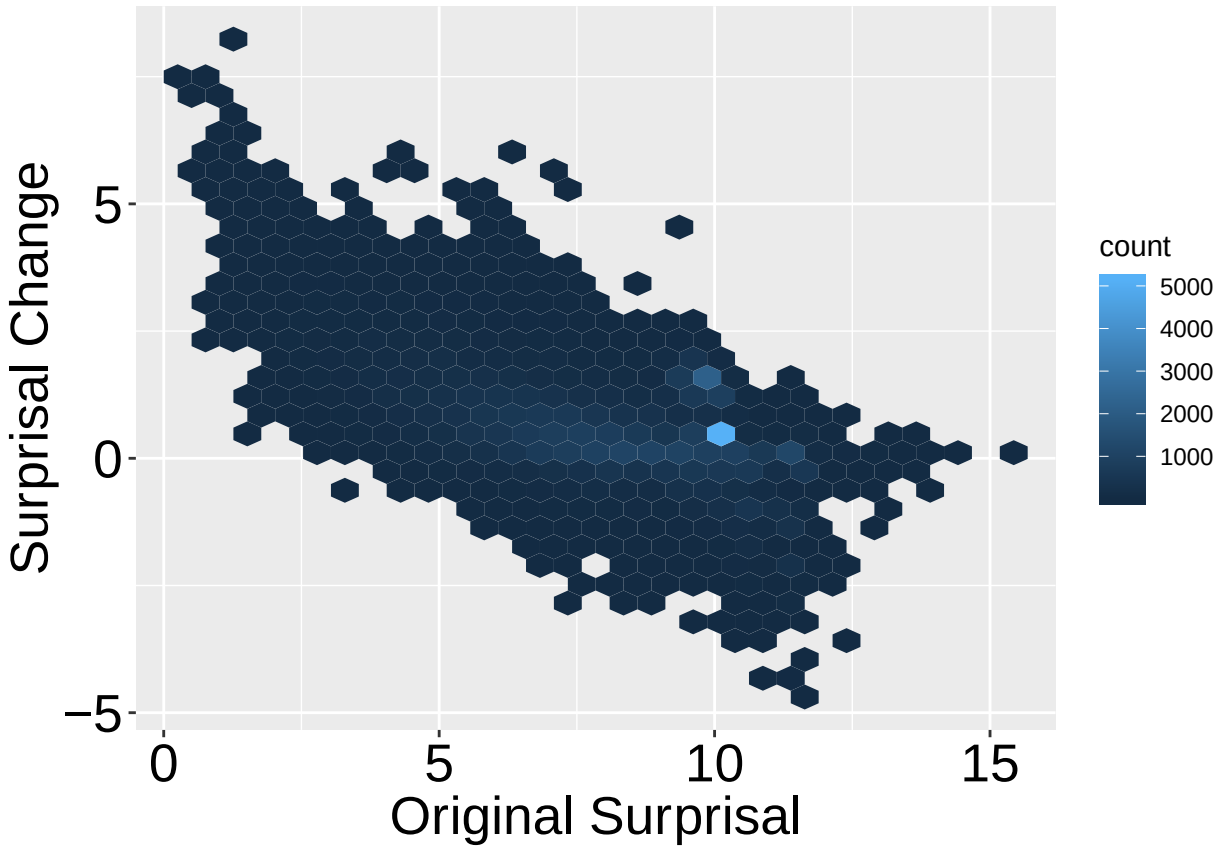




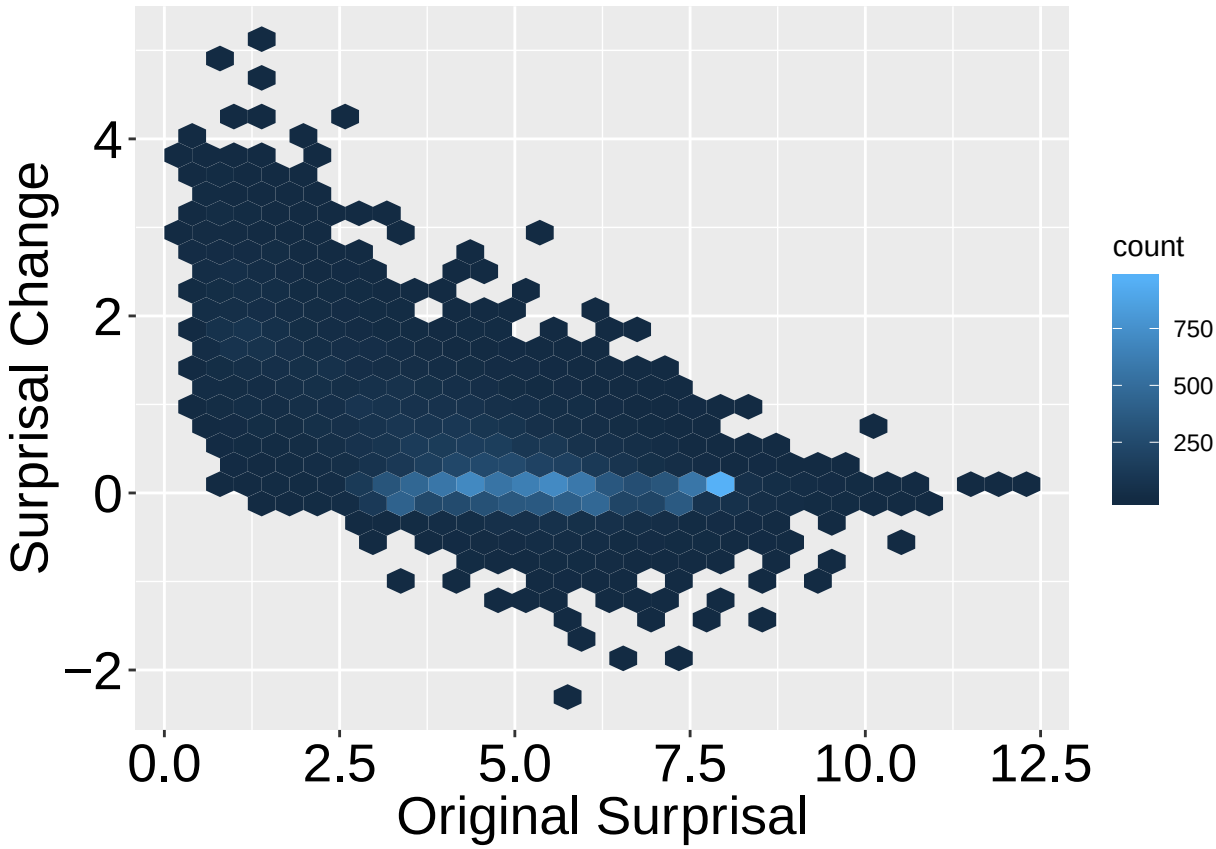

```
#Original Vs change
hexDiffCompare(fWithin1, "VarSwapWithinGlobalMethod")
```



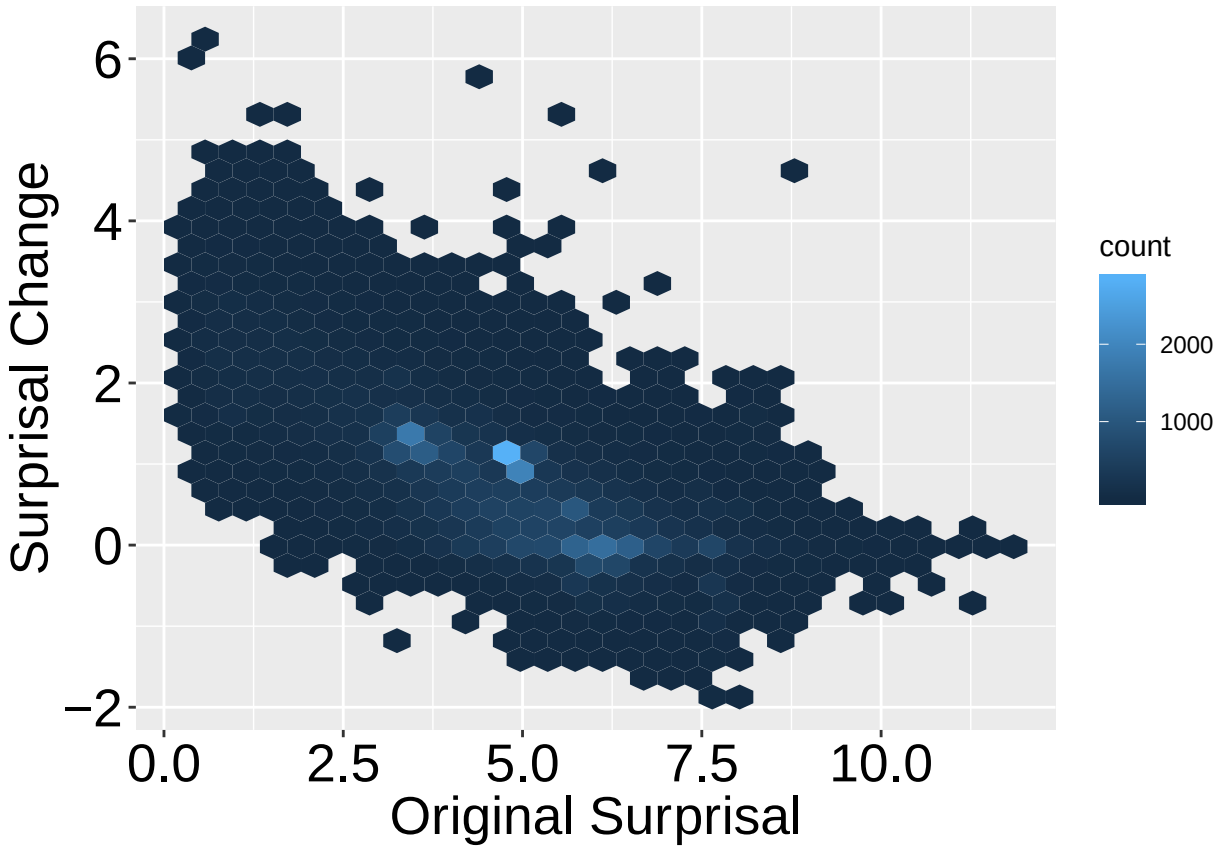

```
hexDiffCompare(fBetween1, "VarSwapBetweenGlobalMethod")
```



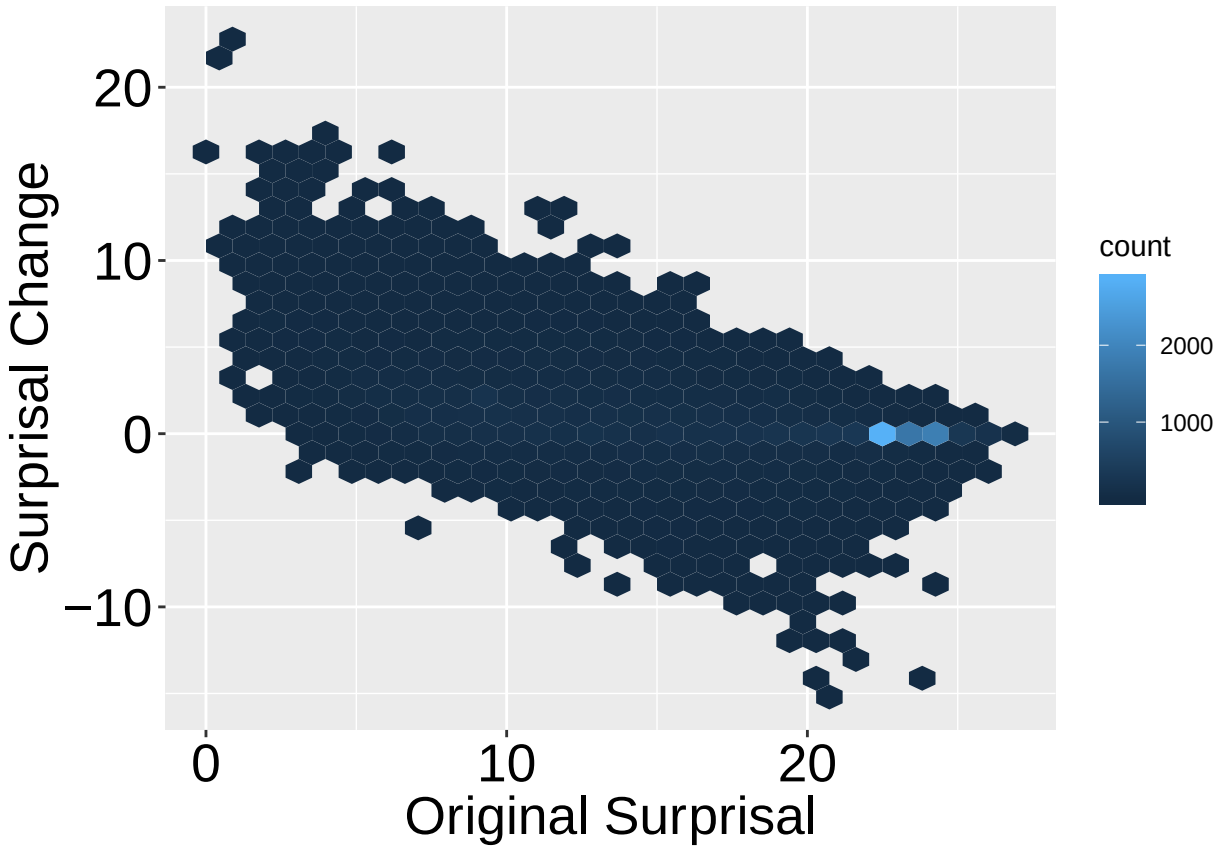

```
hexDiffCompare(fWithin2, "VarSwapWithinCacheMethod")
```



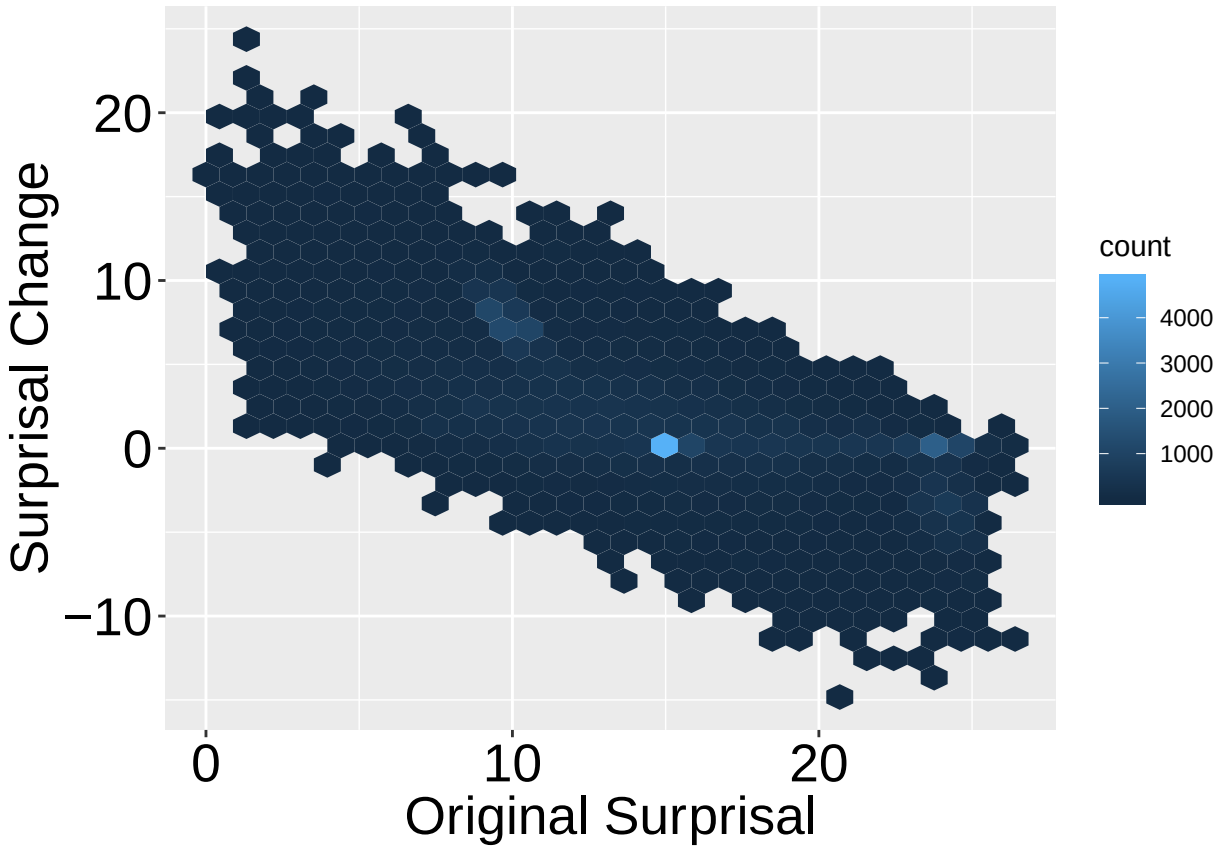

```
hexDiffCompare(fBetween2, "VarSwapBetweenCacheMethod")
```



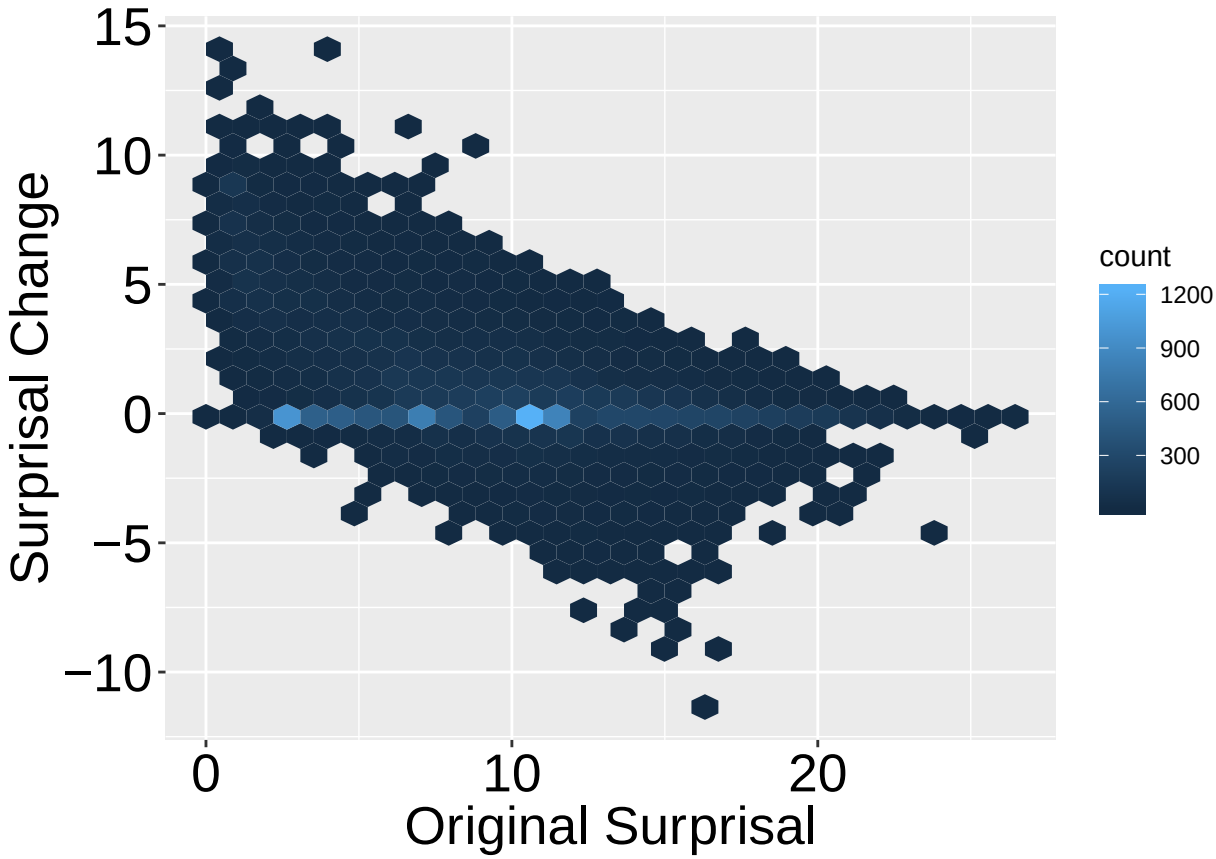

```
hexDiffCompare(fWithin3, "VarSwapWithinGlobalExpMethod")
```



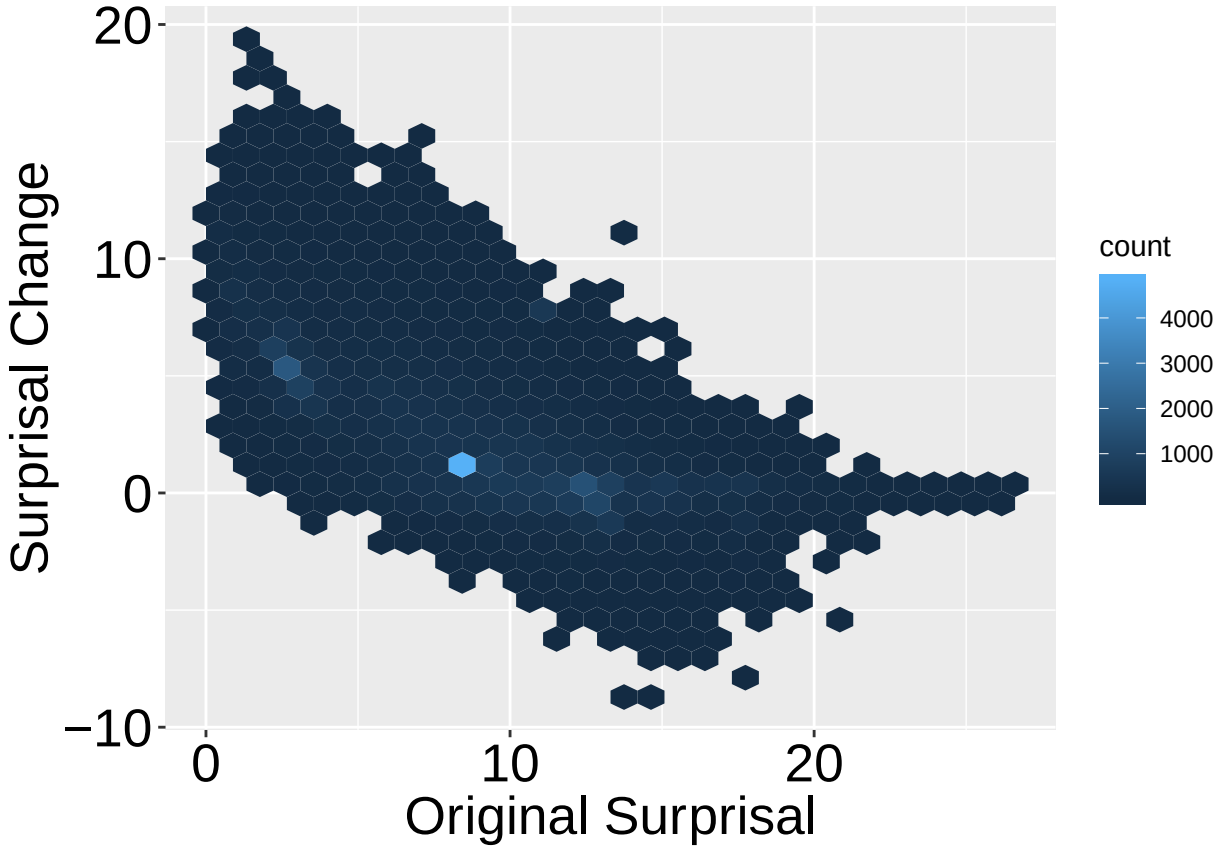

```
hexDiffCompare(fBetween3, "VarSwapBetweenGlobalExpMethod")
```




```
hexDiffCompare(fWithin4, "VarSwapWithinCacheExpMethod")
```




```
hexDiffCompare(fBetween4, "VarSwapBetweenCacheExpMethod")
```




```
#Modelling...
#cohen.d(fWithin1$AveBefore, fWithin1$AveAfter, paired=TRUE)
#cohen.d(fWithin2$AveBefore, fWithin2$AveAfter, paired=TRUE)
#cohen.d(fWithin3$AveBefore, fWithin3$AveAfter, paired=TRUE)
#cohen.d(fWithin4$AveBefore, fWithin4$AveAfter, paired=TRUE)
#cohen.d(fBetween1$AveBefore, fBetween1$AveAfter, paired=TRUE)
#cohen.d(fBetween2$AveBefore, fBetween2$AveAfter, paired=TRUE)
#cohen.d(fBetween3$AveBefore, fBetween3$AveAfter, paired=TRUE)
#cohen.d(fBetween4$AveBefore, fBetween4$AveAfter, paired=TRUE)

addToEffectTable(fWithin3, "VarSwapWithinGlobalExpMethod")
```

```
## [1] "VarSwapWithinGlobalExpMethod Original < Transformed"
##
##  Paired t-test
##
## data:  dataset$AveBefore and dataset$AveAfter
## t = -33.632, df = 17929, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.4921809
## sample estimates:
## mean of the differences
##               -0.538466
##
##
##  Paired t-test
```

```
##
## data:  dataset$AveBefore and dataset$AveAfter
## t = -33.632, df = 17929, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.5881363 -0.4887957
## sample estimates:
## mean of the differences
##               -0.538466
##
##
## Cohen's d
##
## d estimate: -0.2511646 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.2719469 -0.2303823
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  dataset$AveBefore and dataset$AveAfter
## V = 18587000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.6812375
## sample estimates:
## (pseudo)median
##     -0.7397495
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  dataset$AveBefore and dataset$AveAfter
## V = 18587000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.8038566 -0.6769026
## sample estimates:
## (pseudo)median
##     -0.7397495
##
##
## Cliff's Delta
##
## delta estimate: -0.03628699 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.04823105 -0.02433255
addToEffectTable(fWithin4,  "VarSwapWithinCacheExpMethod")

## [1] "VarSwapWithinCacheExpMethod Original < Transformed"
##
##  Paired t-test
##
```

```
## data:  dataset$AveBefore and dataset$AveAfter
## t = -53.116, df = 17929, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.7745963
## sample estimates:
## mean of the differences
##              -0.8191812
##
##
##  Paired t-test
##
## data:  dataset$AveBefore and dataset$AveAfter
## t = -53.116, df = 17929, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -0.8670269 -0.7713354
## sample estimates:
## mean of the differences
##              -0.8191812
##
##
## Cohen's d
##
## d estimate: -0.3966736 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.4175770 -0.3757702
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  dataset$AveBefore and dataset$AveAfter
## V = 14226000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##        -Inf -0.8462882
## sample estimates:
## (pseudo)median
##     -0.9060937
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  dataset$AveBefore and dataset$AveAfter
## V = 14226000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -0.9728920 -0.8419734
## sample estimates:
## (pseudo)median
##     -0.9060937
##
##
## Cliff's Delta
```

```
##
## delta estimate: -0.1033953 (negligible)
## 95 percent confidence interval:
##         inf         sup
## -0.11528143 -0.09147947
```

```
addToEffectTable(fBetween3, "VarSwapBetweenGlobalExpMethod")
```

```
## [1] "VarSwapBetweenGlobalExpMethod Original < Transformed"
##
##  Paired t-test
##
## data:  dataset$AveBefore and dataset$AveAfter
## t = -106.97, df = 48159, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##       -Inf -1.758681
## sample estimates:
## mean of the differences
##               -1.807528
##
##
##  Paired t-test
##
## data:  dataset$AveBefore and dataset$AveAfter
## t = -106.97, df = 48159, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -1.859947 -1.755109
## sample estimates:
## mean of the differences
##               -1.807528
```

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

```
##
## Cohen's d
##
## d estimate: -0.4874172 (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:  dataset$AveBefore and dataset$AveAfter
## V = 229310000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##       -Inf -1.616876
## sample estimates:
## (pseudo)median
```

```
##      -1.670974
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  dataset$AveBefore and dataset$AveAfter
## V = 229310000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -1.730130 -1.612827
## sample estimates:
## (pseudo)median
##      -1.670974
##
##
## Cliff's Delta
##
## delta estimate: -0.2343372 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.2415792 -0.2270690
```

```
addToEffectTable(fBetween4, "VarSwapBetweenCacheExpMethod")
```

```
## [1] "VarSwapBetweenCacheExpMethod Original < Transformed"
##
##  Paired t-test
##
## data:  dataset$AveBefore and dataset$AveAfter
## t = -175.4, df = 48159, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##       -Inf -2.294283
## sample estimates:
## mean of the differences
##               -2.332727
##
##
##  Paired t-test
##
## data:  dataset$AveBefore and dataset$AveAfter
## t = -175.4, df = 48159, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##  -2.373983 -2.291472
## sample estimates:
## mean of the differences
##               -2.332727
```

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

```
##
## Cohen's d
##
## d estimate: -0.7992525 (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:  dataset$AveBefore and dataset$AveAfter
## V = 104430000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##       -Inf -2.235278
## sample estimates:
## (pseudo)median
##      -2.279101
##
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  dataset$AveBefore and dataset$AveAfter
## V = 104430000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##  -2.326184 -2.232054
## sample estimates:
## (pseudo)median
##      -2.279101
##
##
## Cliff's Delta
##
## delta estimate: -0.3122258 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.3191225 -0.3052961
```

```
#Examples
# View(fWithin1)
# View(sqldf("SELECT * FROM dVsp WHERE Filepath=\'spring-framework-5.0.6.RELEASE/libs/spring-web-5.0.6.
# View(sqldf("SELECT * FROM dVsp WHERE Filepath=\'spring-framework-5.0.6.RELEASE/libs/spring-aop-5.0.6.
```

Regression Models

```
m_var_within_global <- modelVar(fWithin3)
```

```
##
## Call:
## lm(formula = AveDiff ~ AveBefore, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.2051  -0.6746   0.2131   0.5088  19.3482
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.137327   0.043341   72.39   <2e-16 ***
## AveBefore   -0.148479   0.002334  -63.61   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.937 on 17928 degrees of freedom
## Multiple R-squared:  0.1841, Adjusted R-squared:  0.1841
## F-statistic:  4046 on 1 and 17928 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: AveDiff
##              Df Sum Sq Mean Sq F value    Pr(>F)
## AveBefore     1  15173 15173.1    4046 < 2.2e-16 ***
## Residuals 17928  67232     3.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.1841271" "0.8158729"
##
## % 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 - 11:44:06 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] & AveDiff \\
## \hline \\[-1.8ex]
##  AveBefore & $-$0.148$^{***}$ (0.002) \\
##   Constant & 3.137$^{***}$ (0.043) \\
##  \hline \\[-1.8ex]
## Observations & 17,930 \\
## R$^{2}$ & 0.184 \\
## Adjusted R$^{2}$ & 0.184 \\
## Residual Std. Error & 1.937 (df = 17928) \\
## F Statistic & 4,046.011$^{***}$ (df = 1; 17928) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 23:44:07 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
```

```
## AveBefore & 1 & 15173.07 & 15173.07 & 4046.01 & 0.0000 \\
##   Residuals & 17928 & 67232.34 & 3.75 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
m_var_within_cache <- modelVar(fWithin4)
```

```
##
## Call:
## lm(formula = AveDiff ~ AveBefore, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9912  -1.2432  -0.3168   0.8369  12.5844
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.574095   0.029265   87.96   <2e-16 ***
## AveBefore   -0.192874   0.002839  -67.94   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.842 on 17928 degrees of freedom
## Multiple R-squared:  0.2047, Adjusted R-squared:  0.2047
## F-statistic:  4615 on 1 and 17928 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: AveDiff
##              Df Sum Sq Mean Sq F value    Pr(>F)
## AveBefore     1  15655 15654.7  4615.5 < 2.2e-16 ***
## Residuals 17928  60808     3.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.2047369" "0.7952631"
##
## % 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 - 11:44:13 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] & AveDiff \\
## \hline \\[-1.8ex]
##  AveBefore & $-$0.193$^{***}$ (0.003) \\
##   Constant & 2.574$^{***}$ (0.029) \\
##  \hline \\[-1.8ex]
## Observations & 17,930 \\
## R$^{2}$ & 0.205 \\
## Adjusted R$^{2}$ & 0.205 \\
```

```
## Residual Std. Error & 1.842 (df = 17928) \\
## F Statistic & 4,615.482$^{***}$ (df = 1; 17928) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 23:44:14 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## AveBefore & 1 & 15654.73 & 15654.73 & 4615.48 & 0.0000 \\
##   Residuals & 17928 & 60807.94 & 3.39 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
m_var_between_global <- modelVar(fBetween3)
```

```
##
## Call:
## lm(formula = AveDiff ~ AveBefore, data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.280  -1.558  -0.237   2.059  16.722
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.811067   0.035217   250.2   <2e-16 ***
## AveBefore   -0.463578   0.002188  -211.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.668 on 48158 degrees of freedom
## Multiple R-squared:  0.4825, Adjusted R-squared:  0.4825
## F-statistic: 4.49e+04 on 1 and 48158 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: AveDiff
##              Df Sum Sq Mean Sq F value    Pr(>F)
## AveBefore     1 319543  319543   44898 < 2.2e-16 ***
## Residuals 48158 342743       7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.482485" "0.517515"
##
## % 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 - 11:44:14 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] & AveDiff \\
## \hline \\[-1.8ex]
##  AveBefore & $-$0.464$^{***}$ (0.002) \\
##   Constant & 8.811$^{***}$ (0.035) \\
##  \hline \\[-1.8ex]
## Observations & 48,160 \\
## R$^{2}$ & 0.482 \\
## Adjusted R$^{2}$ & 0.482 \\
## Residual Std. Error & 2.668 (df = 48158) \\
## F Statistic & 44,898.240$^{***}$ (df = 1; 48158) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 23:44:14 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## AveBefore & 1 & 319543.30 & 319543.30 & 44898.24 & 0.0000 \\
##   Residuals & 48158 & 342743.19 & 7.12 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

```
m_var_between_cache <- modelVar(fBetween4)
```

```
##
## Call:
## lm(formula = AveDiff ~ AveBefore, data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7164 -1.5517 -0.3201  1.0848 13.8827
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.445444   0.021947   293.7   <2e-16 ***
## AveBefore   -0.462164   0.002216  -208.6   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.115 on 48158 degrees of freedom
## Multiple R-squared:  0.4746, Adjusted R-squared:  0.4746
```

```
## F-statistic: 4.351e+04 on 1 and 48158 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: AveDiff
##              Df Sum Sq Mean Sq F value    Pr(>F)
## AveBefore     1 194717  194717   43509 < 2.2e-16 ***
## Residuals 48158 215522       4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.4746426" "0.5253574"
##
## % 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 - 11:44:14 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] & AveDiff \\
## \hline \\[-1.8ex]
##  AveBefore & $-$0.462$^{***}$ (0.002) \\
##   Constant & 6.445$^{***}$ (0.022) \\
##  \hline \\[-1.8ex]
## Observations & 48,160 \\
## R$^{2}$ & 0.475 \\
## Adjusted R$^{2}$ & 0.475 \\
## Residual Std. Error & 2.115 (df = 48158) \\
## F Statistic & 43,509.120$^{***}$ (df = 1; 48158) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 23:44:14 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
##   \hline
##  & Df & Sum Sq & Mean Sq & F value & Pr($>$F) \\
##   \hline
## AveBefore & 1 & 194717.10 & 194717.10 & 43509.12 & 0.0000 \\
##   Residuals & 48158 & 215522.33 & 4.48 &  &  \\
##    \hline
## \end{tabular}
## \end{table}
```

Output the effect sizes

```
VarOut <- printEffTable(pairedResults)
```

```
## [1] "Type,PTOne,PTTwo,CITTwo,CohensD,PWilcoxOne,PWilcoxTwo,CIWilcoxTwo,CliffDelta"
## [1] "VarSwapBetweenCacheExpMethod,0,0,-2.374 -2.2915,-0.7993,0,0,-2.3262 -2.2321,-0.3122"
## [1] "VarSwapBetweenGlobalExpMethod,0,0,-1.8599 -1.7551,-0.4874,0,0,-1.7301 -1.6128,-0.2343"
## [1] "VarSwapBetweenTopCache,0,0,-0.858 -0.8496,-0.5222,0,0,-0.8857 -0.8738,-0.2001"
## [1] "VarSwapBetweenTopCacheExp,0,0,-2.4247 -2.3935,-0.3967,0,0,-2.2623 -2.2335,-0.2463"
## [1] "VarSwapBetweenTopGlobal,0,0,-0.8772 -0.8668,-0.4277,0,0,-0.9491 -0.9377,-0.161"
## [1] "VarSwapBetweenTopGlobalExp,0,0,-2.8146 -2.7722,-0.3384,0,0,-3.1548 -3.1068,-0.2182"
## [1] "VarSwapWithinCacheExpMethod,0,0,-0.867 -0.7713,-0.3967,0,0,-0.9729 -0.842,-0.1034"
## [1] "VarSwapWithinGlobalExpMethod,0,0,-0.5881 -0.4888,-0.2512,0,0,-0.8039 -0.6769,-0.0363"
## [1] "VarSwapWithinTopCache,0,0,-0.6485 -0.6326,-0.4026,0,0,-0.6568 -0.6404,-0.1577"
## [1] "VarSwapWithinTopCacheExp,0,0,-1.3853 -1.3323,-0.2562,0,0,-1.2089 -1.1496,-0.1404"
## [1] "VarSwapWithinTopGlobal,0,0,-0.8175 -0.7983,-0.4201,0,0,-0.9876 -0.9661,-0.1426"
## [1] "VarSwapWithinTopGlobalExp,0,0,-2.1606 -2.0909,-0.3049,0,0,-3.384 -3.2935,-0.1415"
```

```
save(VarOut, file = "/data/anon/SemanticTransformation/sample/VarOut.RDat")

save(dVsp, file = "/data/anon/SemanticTransformation/sample/dVsp.RDat")
save(dVsp2, file = "/data/anon/SemanticTransformation/sample/dVsp2.RDat")

save(dVspFiltered, file = "/data/anon/SemanticTransformation/sample/dVspFiltered.RDat")
save(dVspFiltered2, file = "/data/anon/SemanticTransformation/sample/dVspFiltered2.RDat")
save(dVsp2Filtered, file = "/data/anon/SemanticTransformation/sample/dVsp2Filtered.RDat")
save(dVsp2Filtered2, file = "/data/anon/SemanticTransformation/sample/dVsp2Filtered2.RDat")
```