

Python Relational Operator Swap

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)

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

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

Read in swapping files

```
setwd("/data/anon/SemanticTransformation")
dsp <- compareDepthSummary("python_swap_operators.csv", "PythonLogicalSwapTop", "SAME")

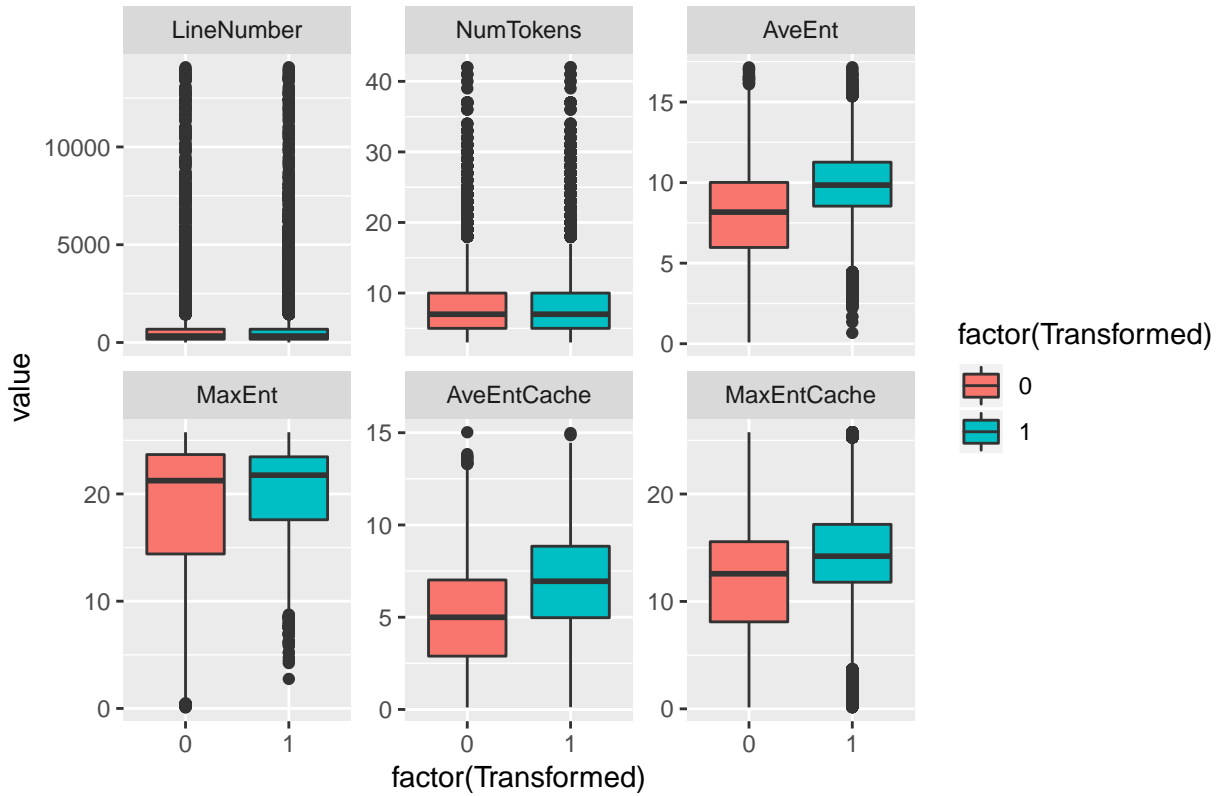
## [[1]]
## [1] "4"      "Bool"

## Loading required package: tcltk

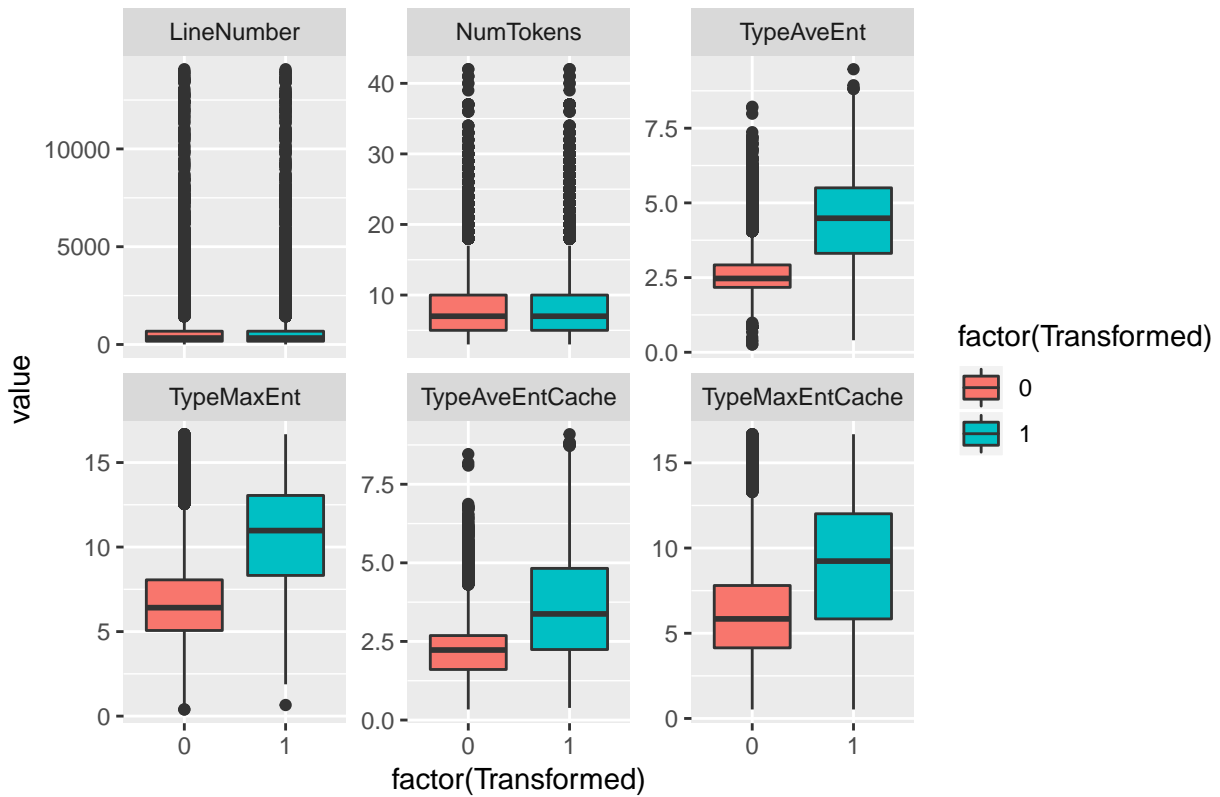
## [1] "Filepath"
## [2] "LineNumber"
## [3] "NumTokens"
## [4] "Transformed"
## [5] "Source"
## [6] "CleanLexerNumTokens"
## [7] "CleanLexerSource"
## [8] "AveEnt"
```

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

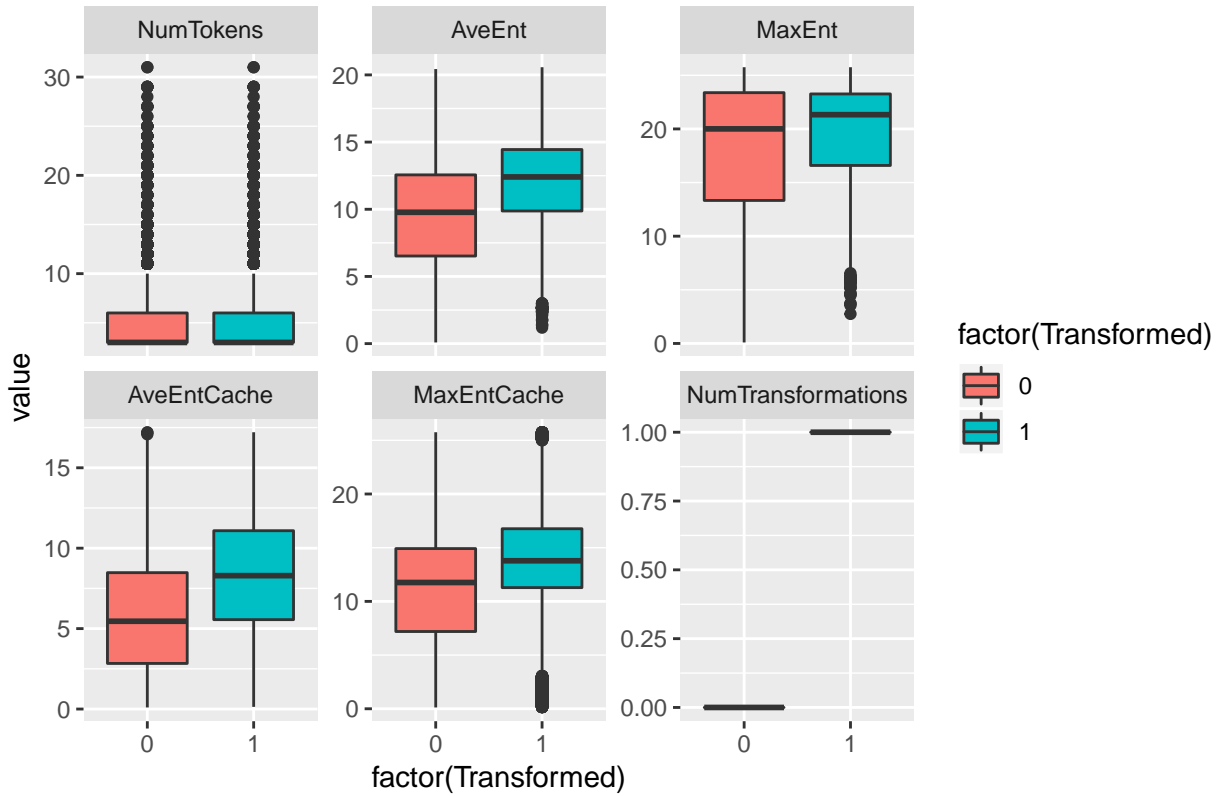
Regular



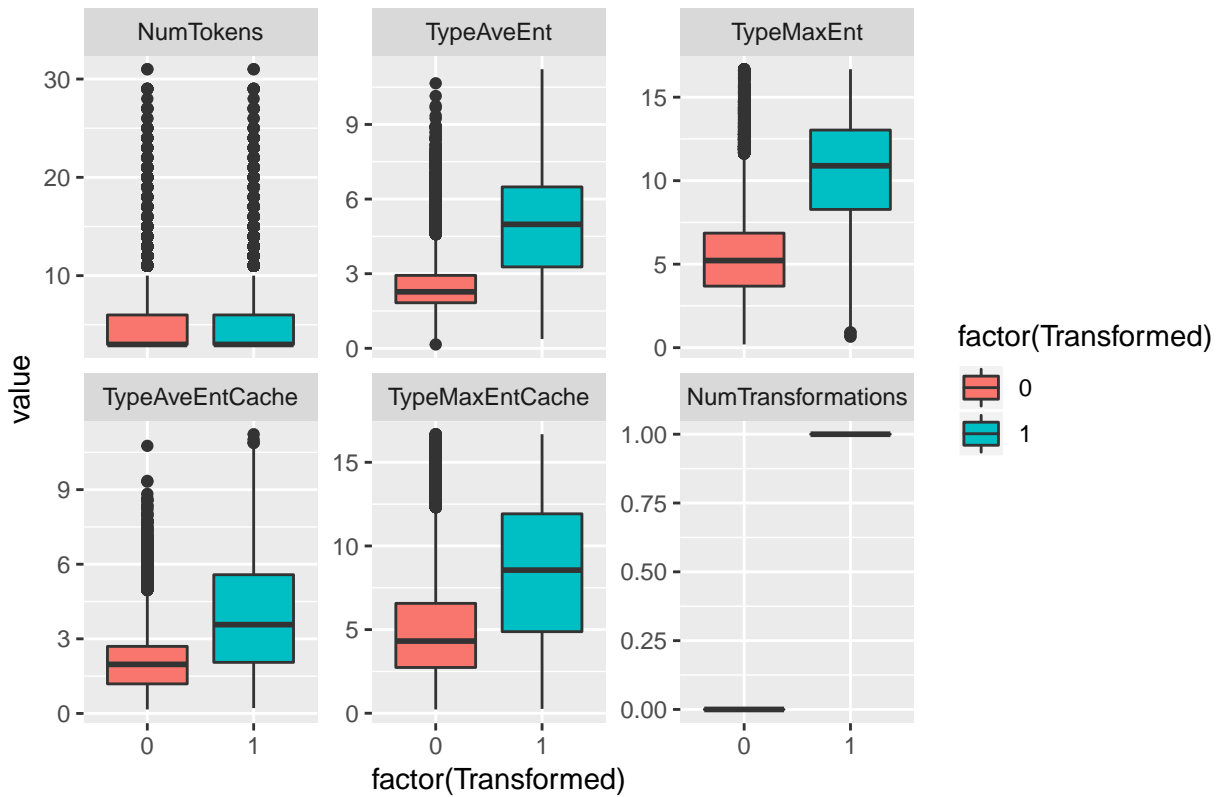
Type



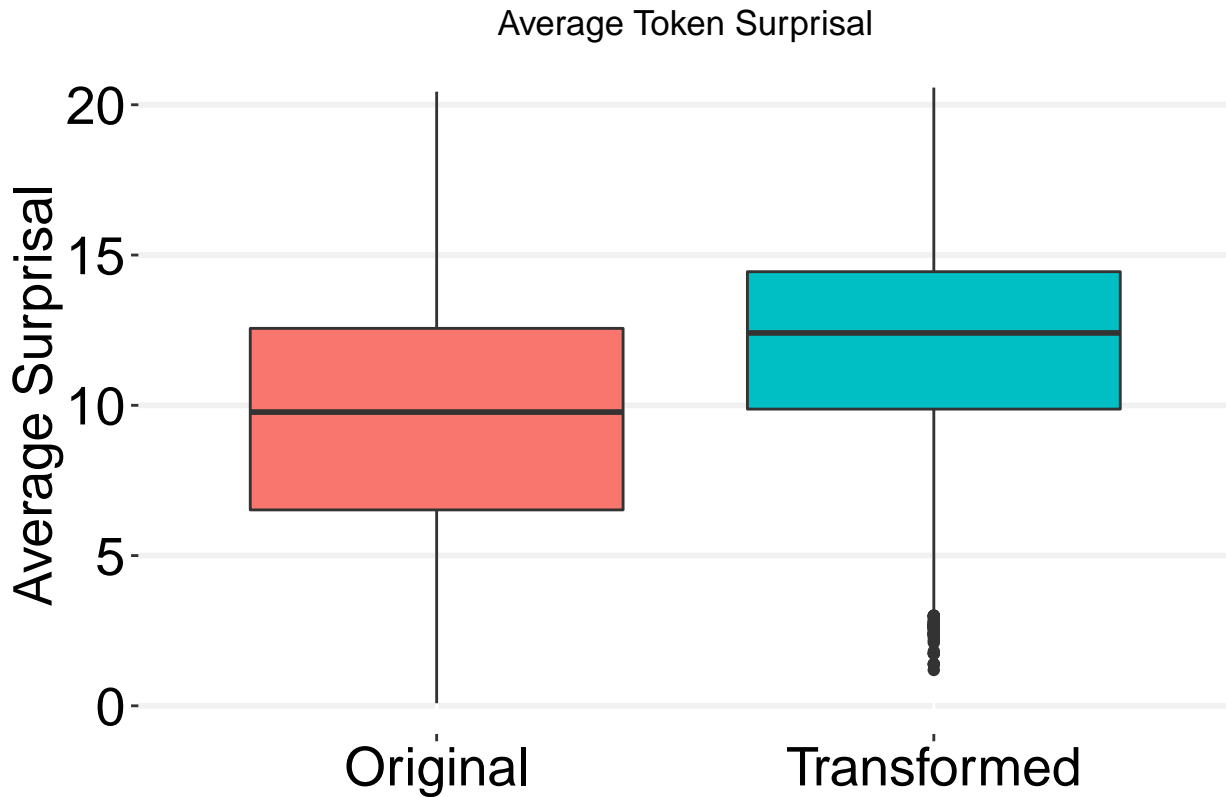
Expression (Regular)



Expression (Type)



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

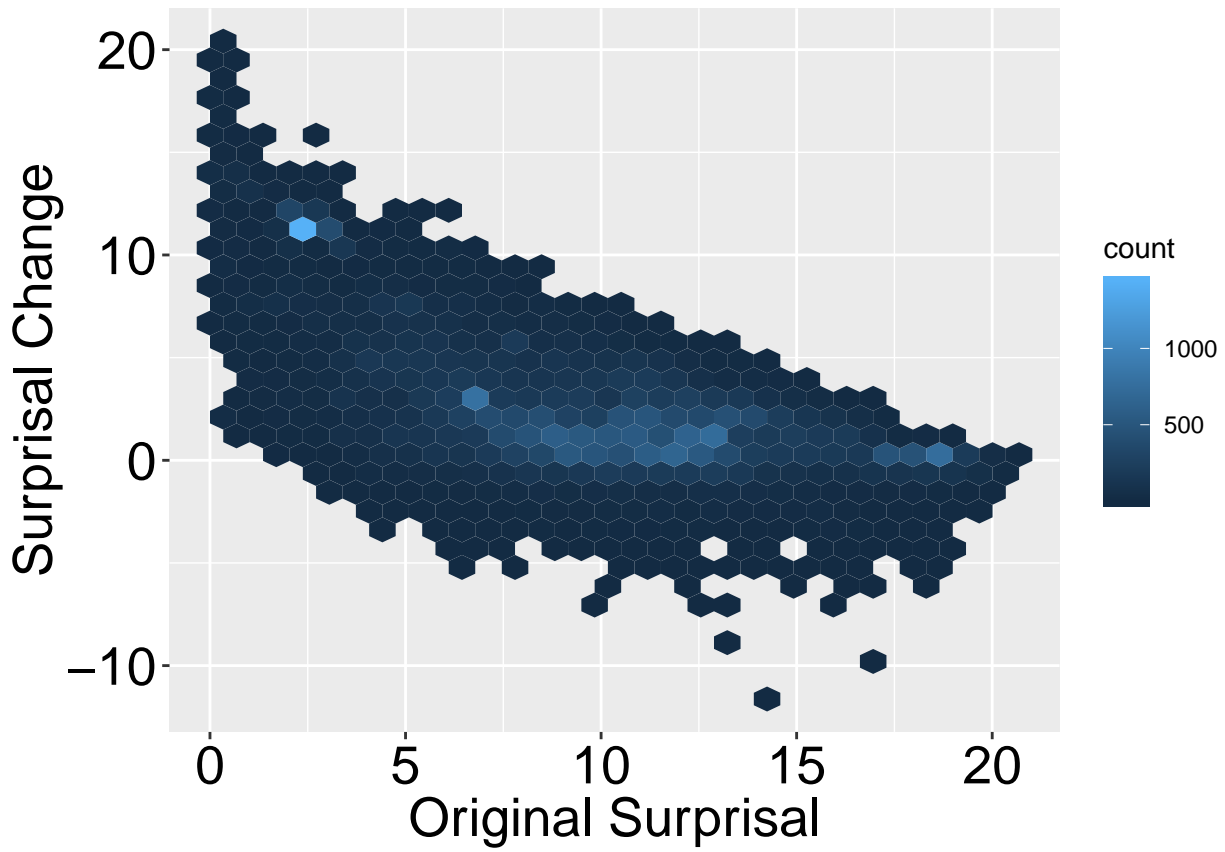


```
## [1] "PythonLogicalSwapTopGlobalExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -150.12, df = 36679, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.597236
## sample estimates:
## mean of the differences
##      -2.64823
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -150.12, df = 36679, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -2.702954 -2.593507
## sample estimates:
## mean of the differences
##      -2.64823
##
## Cohen's d
##
```

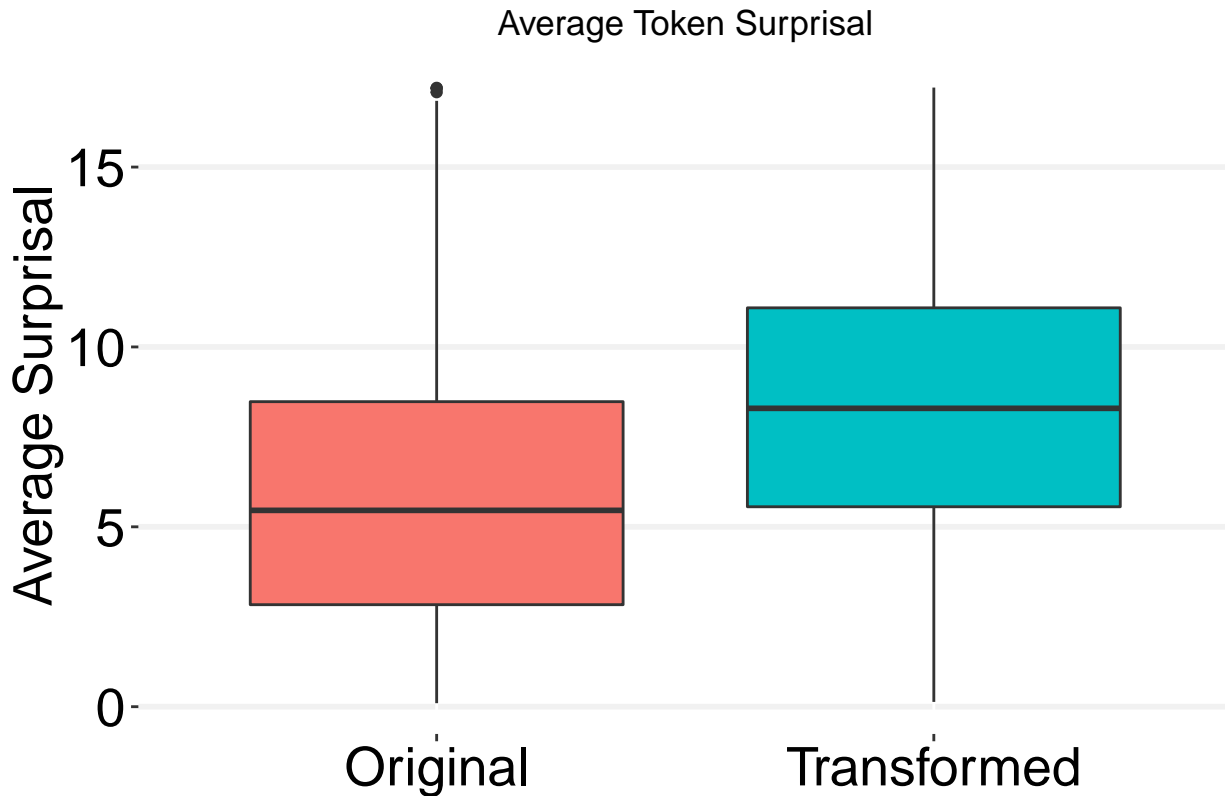
```

## d estimate: -0.7838287 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.7988470 -0.7688103
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 38141000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.061838
## sample estimates:
## (pseudo)median
##      -2.105549
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 38141000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -2.153615 -2.058898
## sample estimates:
## (pseudo)median
##      -2.105549
##
## Cliff's Delta
##
## delta estimate: -0.3689679 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.3766235 -0.3612618

```



[1] " ----- Expression Cache Model ----- "



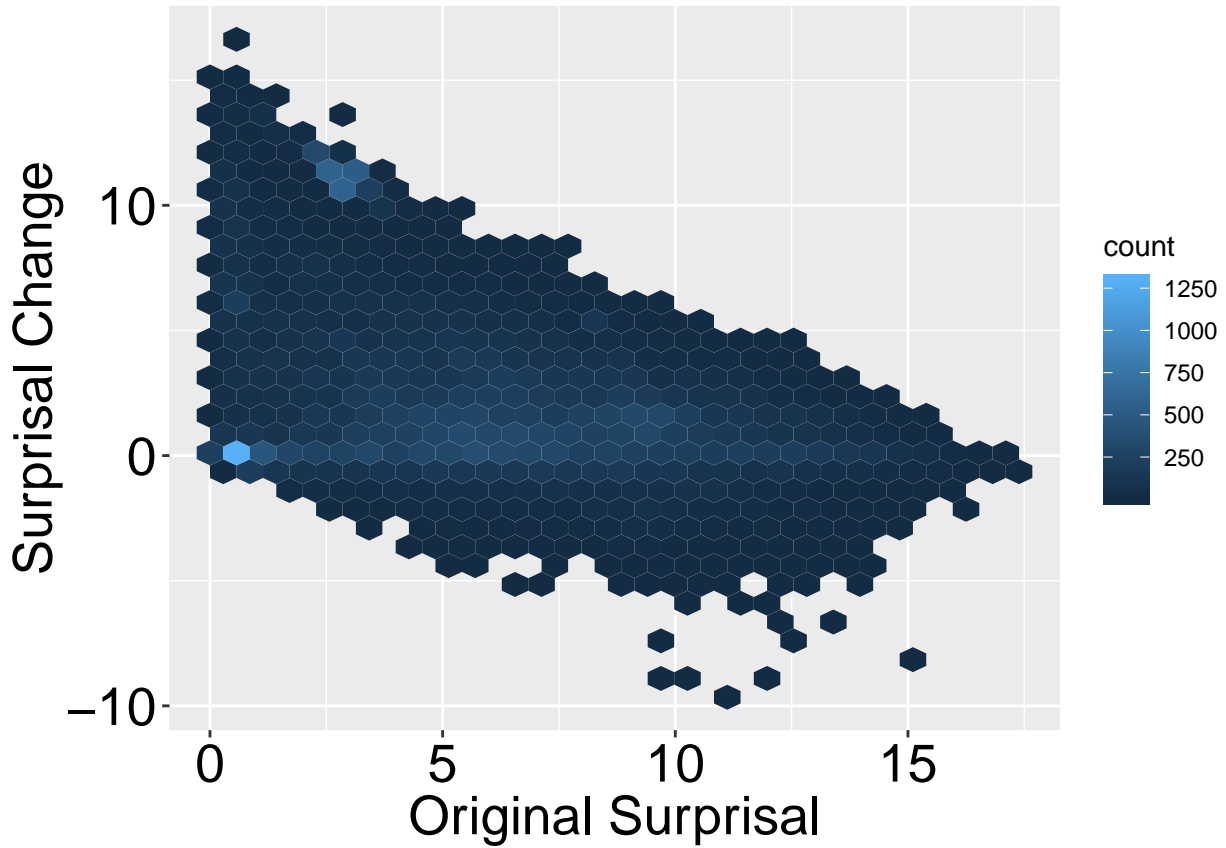
```

## [1] "PythonLogicalSwapTopCacheExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -134.55, df = 36679, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.363454
## sample estimates:
## mean of the differences
##      -2.415346
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -134.55, df = 36679, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -2.471033 -2.359659
## sample estimates:
## mean of the differences
##      -2.415346
##
## Cohen's d
##
## d estimate: -0.7025298 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.7174425 -0.6876172
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 75803000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.749976
## sample estimates:
## (pseudo)median
##      -1.796357
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 75803000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
## -1.846997 -1.746608
## sample estimates:
## (pseudo)median

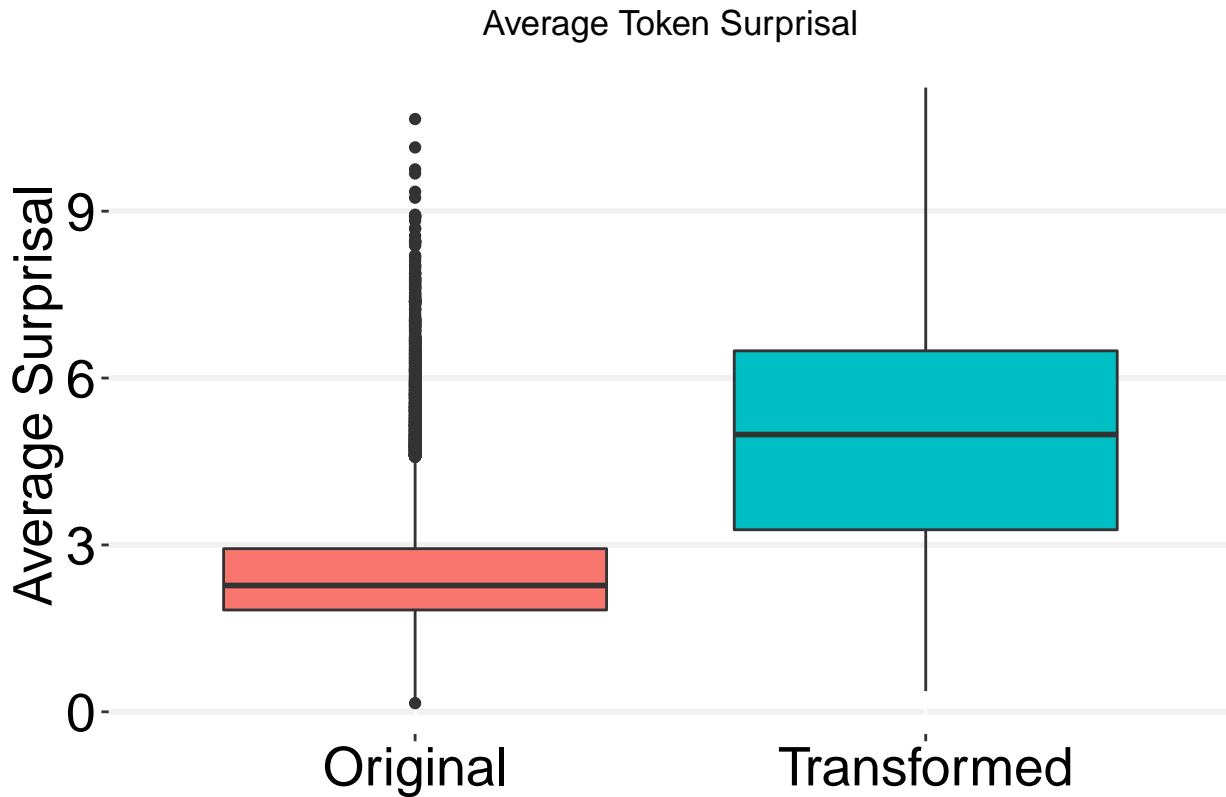
```



```
##      -1.796357
##
##
## Cliff's Delta
##
## delta estimate: -0.3533892 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.3610782 -0.3456520
```



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

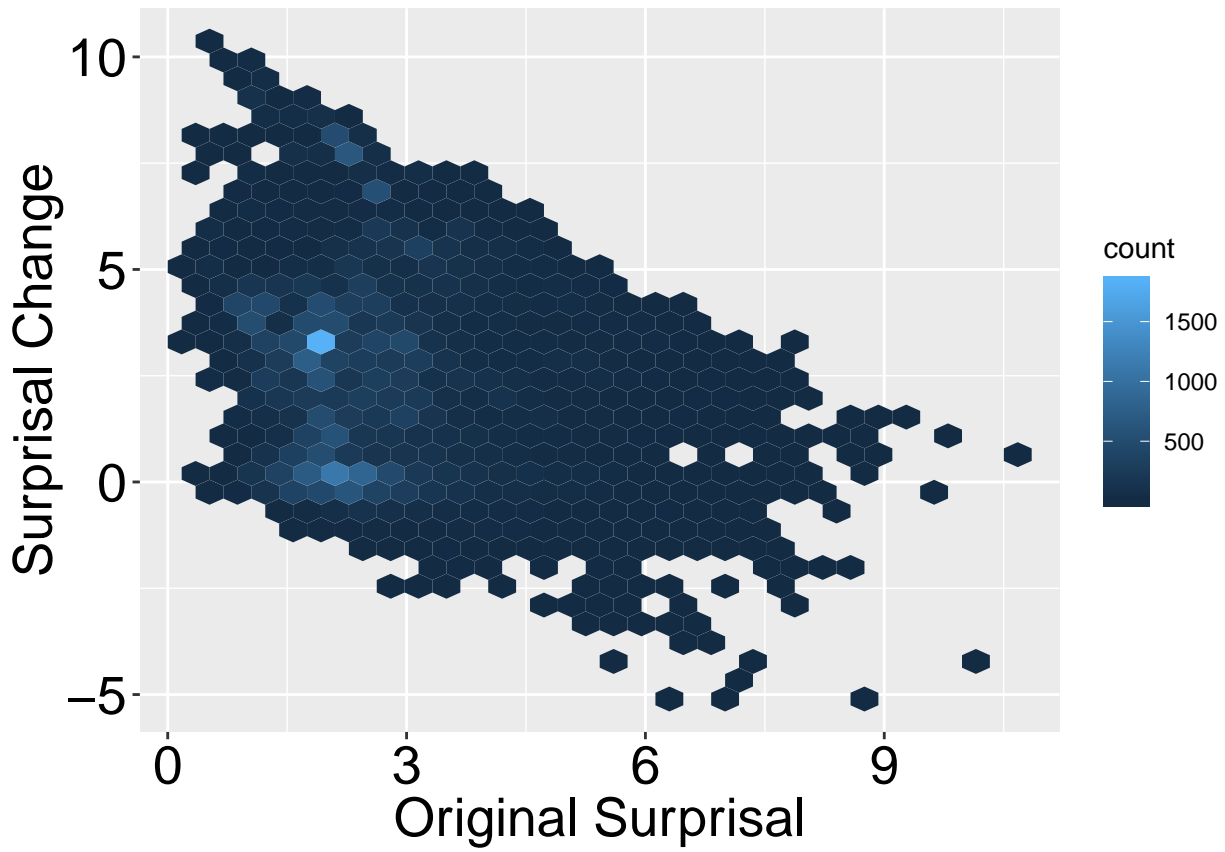


```
## [1] "PythonLogicalSwapTopGlobalTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -225.83, df = 36679, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.656182
## sample estimates:
## mean of the differences
##      -2.690622
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -225.83, df = 36679, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -2.727582 -2.653663
## sample estimates:
## mean of the differences
##      -2.690622
##
## Cohen's d
##
```

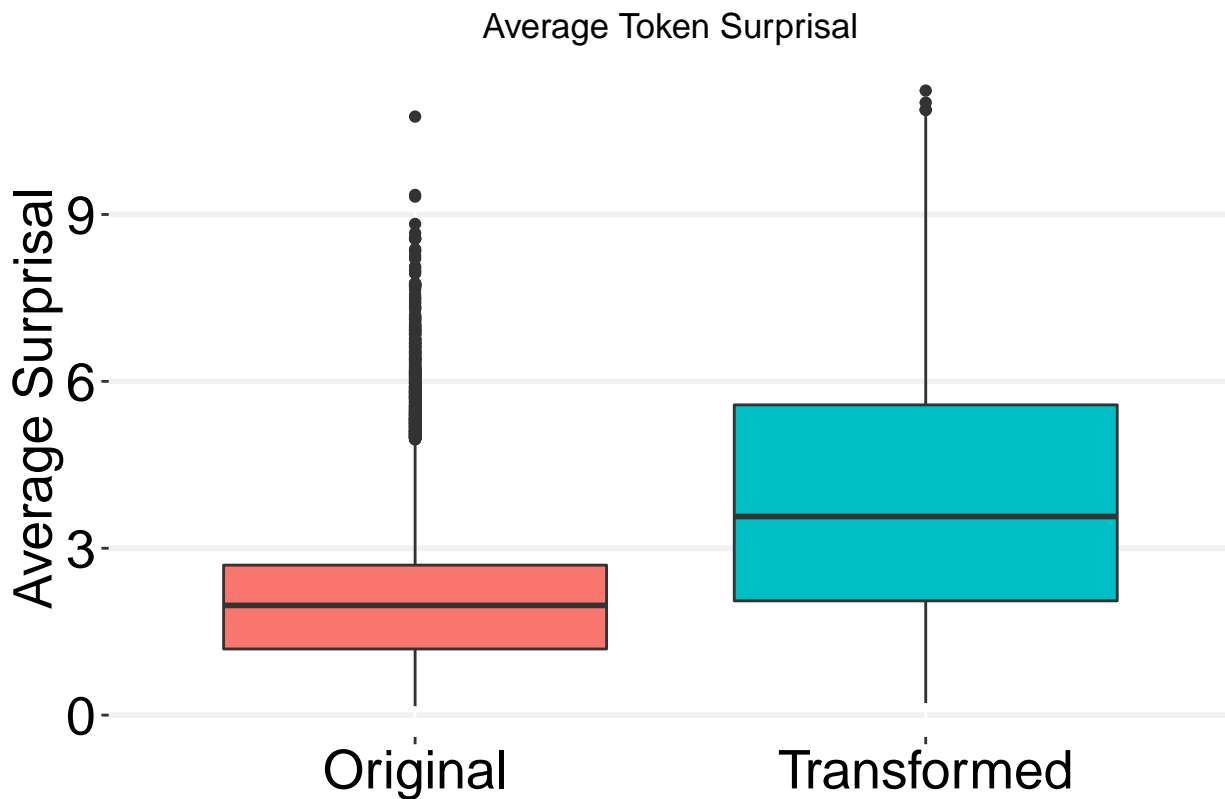
```

## d estimate: -1.179144 (large)
## 95 percent confidence interval:
##      inf      sup
## -1.194824 -1.163464
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10721000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.768967
## sample estimates:
## (pseudo)median
##      -2.804761
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10721000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -2.844280 -2.766274
## sample estimates:
## (pseudo)median
##      -2.804761
##
## Cliff's Delta
##
## delta estimate: -0.6999529 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.7055326 -0.6942864

```



```
## [1] " ----- Expression Cache Type Model ----- "
```

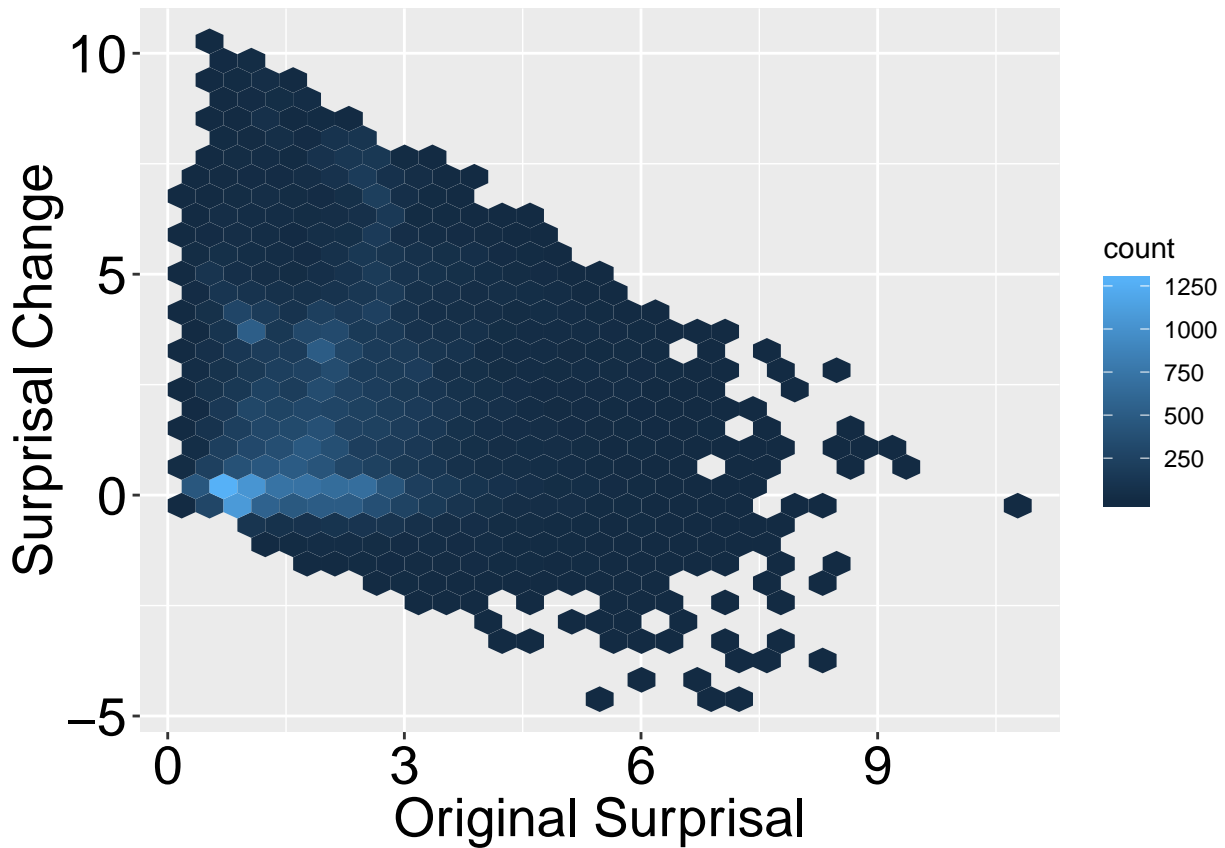


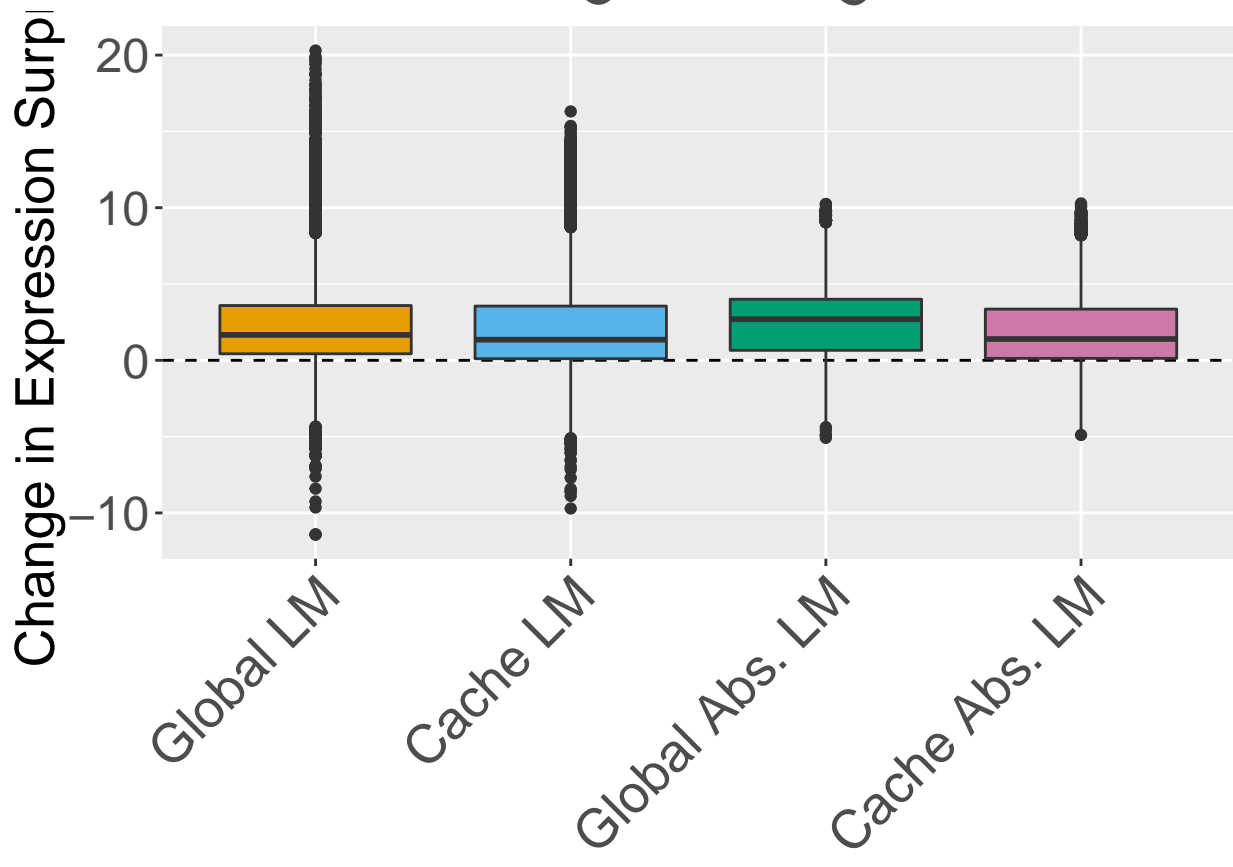
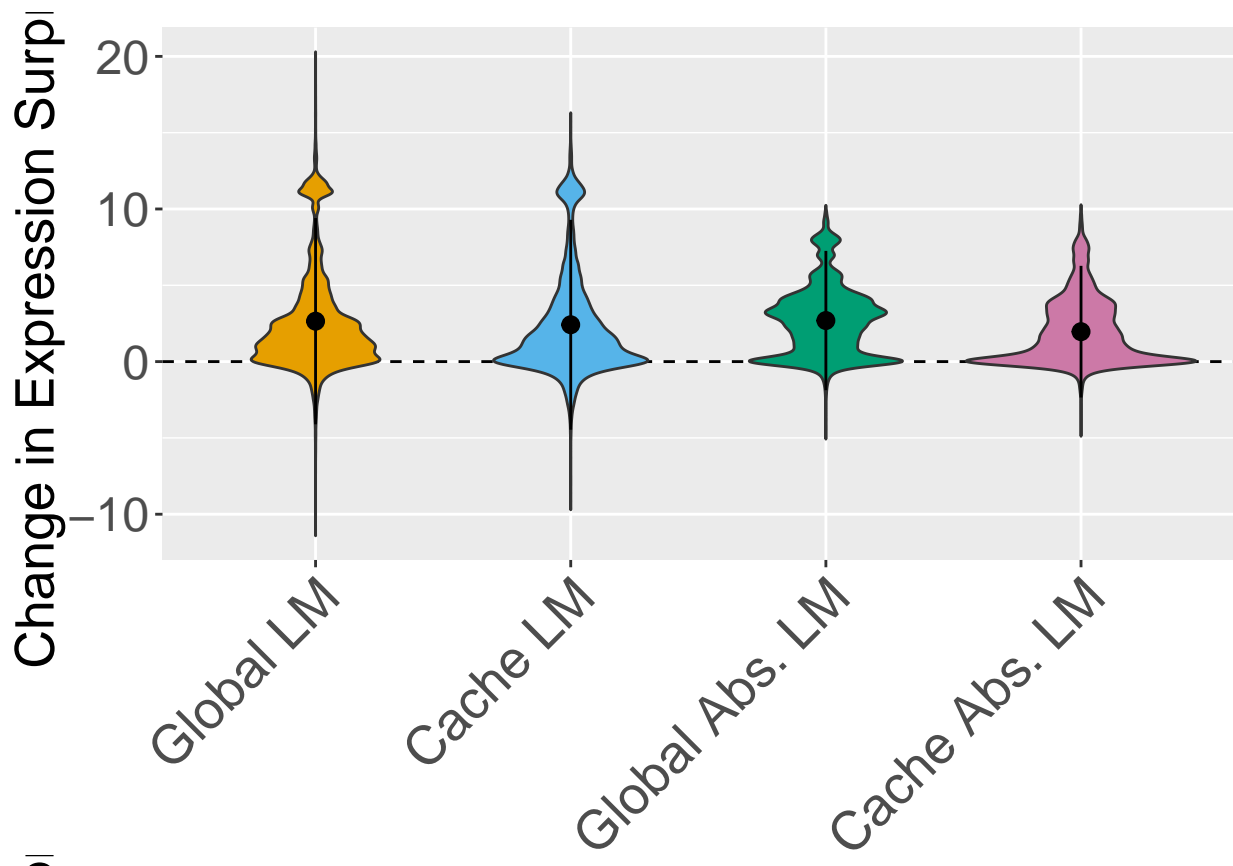
```

## [1] "PythonLogicalSwapTopCacheTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -174.18, df = 36679, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.929248
## sample estimates:
## mean of the differences
##      -1.961805
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -174.18, df = 36679, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -1.996744 -1.926867
## sample estimates:
## mean of the differences
##      -1.961805
##
## Cohen's d
##
## d estimate: -0.909475 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.9246777 -0.8942723
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 37315000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.811855
## sample estimates:
## (pseudo)median
##      -1.84456
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 37315000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
## -1.879471 -1.809425
## sample estimates:
## (pseudo)median

```

```
##      -1.84456
##
##
## Cliff's Delta
##
## delta estimate: -0.4801851 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.4874409 -0.4728630
## [1] "Binary differences"
##
## FALSE TRUE
## 4910 31770
##
## FALSE TRUE
## 7780 28900
##
## FALSE TRUE
## 2799 33881
##
## FALSE TRUE
## 5895 30785
##
## No id variables; using all as measure variables
## Warning: Ignoring unknown parameters: mult
```





Filter out the possible auto-generated code.

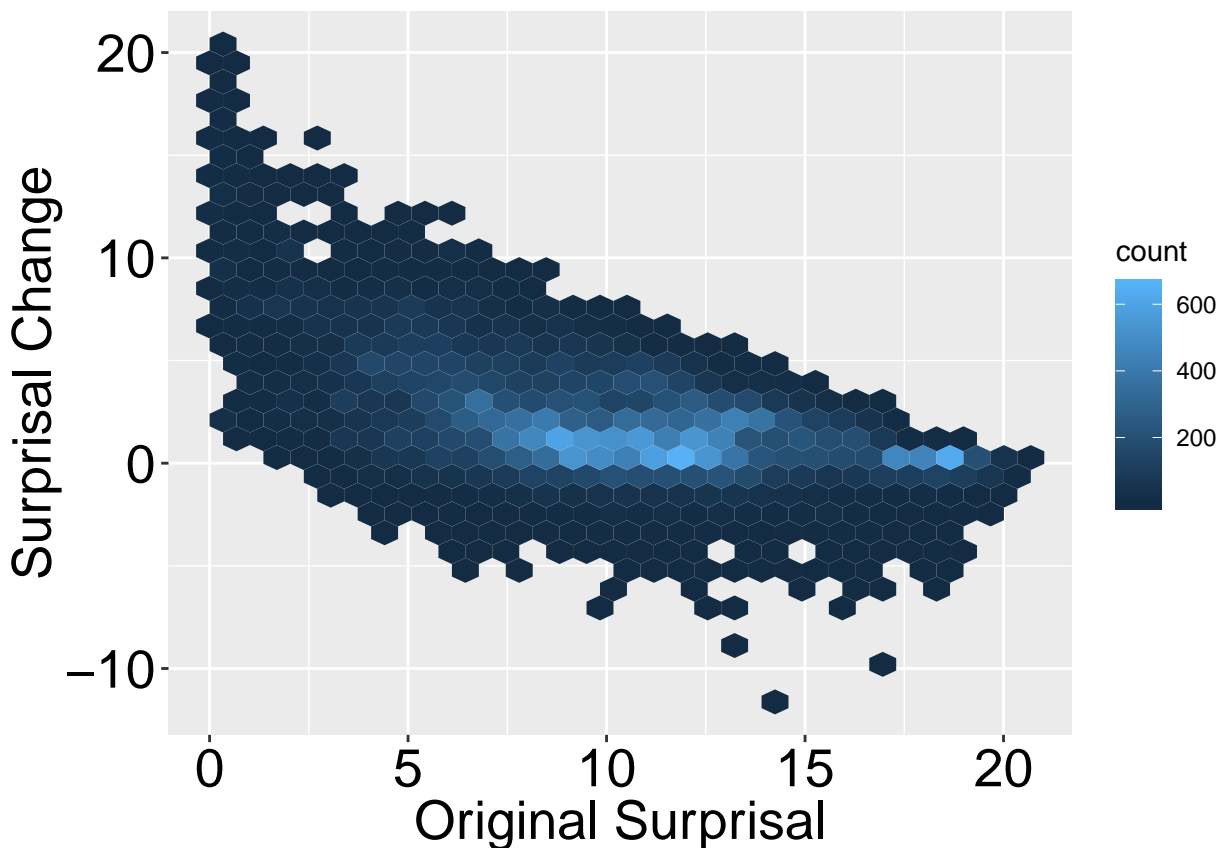
```
setwd("/data/anon/SemanticTransformation/")

#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?
dspFiltered <- generateFilteredResults(dsp, "PyLogic", "PyLogicSwapTopFiltered100", 100)

## [1] " ----- Expression Global Model ----- "
## [1] "PyLogicSwapTopFiltered100GlobalExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -141.68, df = 32218, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.883869
## sample estimates:
## mean of the differences
##      -1.923107
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -141.68, df = 32218, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -1.965215 -1.881000
## sample estimates:
## mean of the differences
##      -1.923107
##
## Cohen's d
##
## d estimate: -0.7893058 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.8053383 -0.7732734
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 37177000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.687368
## sample estimates:
## (pseudo)median
##      -1.723773
```



```
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 37177000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
## -1.763307 -1.684673
## sample estimates:
## (pseudo)median
## -1.723773
##
##
## Cliff's Delta
##
## delta estimate: -0.2834245 (small)
## 95 percent confidence interval:
## inf sup
## -0.2918864 -0.2749182
```



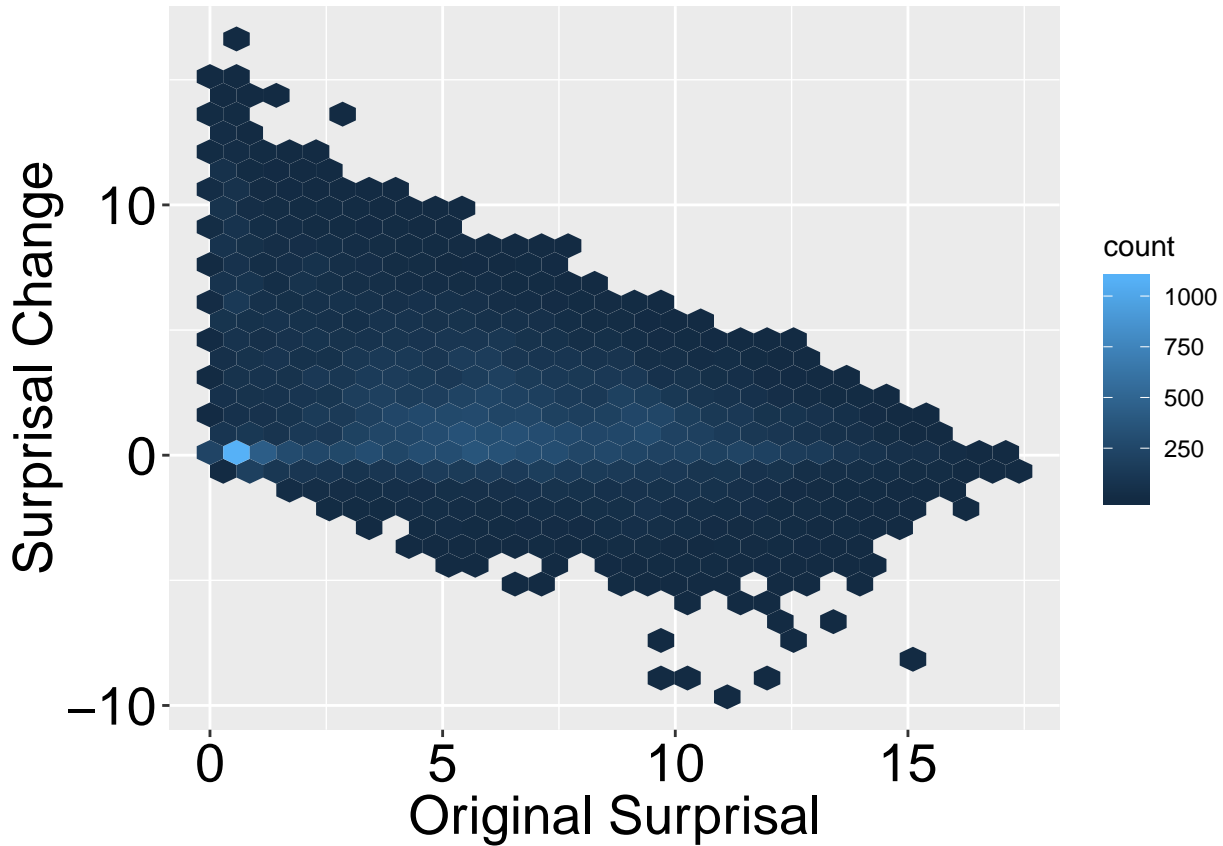
```
## [1] "----- Expression Cache Model -----"
## [1] "PyLogicSwapTopFiltered100CacheExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
```

```

## t = -123.21, df = 32218, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.665221
## sample estimates:
## mean of the differences
##      -1.705229
##
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -123.21, df = 32218, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -1.748162 -1.662296
## sample estimates:
## mean of the differences
##      -1.705229
##
##
## Cohen's d
##
## d estimate: -0.6864242 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.7023149 -0.6705335
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 69442000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.381569
## sample estimates:
## (pseudo)median
##      -1.419427
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 69442000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
## -1.460570 -1.378819
## sample estimates:
## (pseudo)median
##      -1.419427
##
##
## Cliff's Delta
##

```

```
## delta estimate: -0.269282 (small)
## 95 percent confidence interval:
##      inf      sup
## -0.2777877 -0.2607341
```

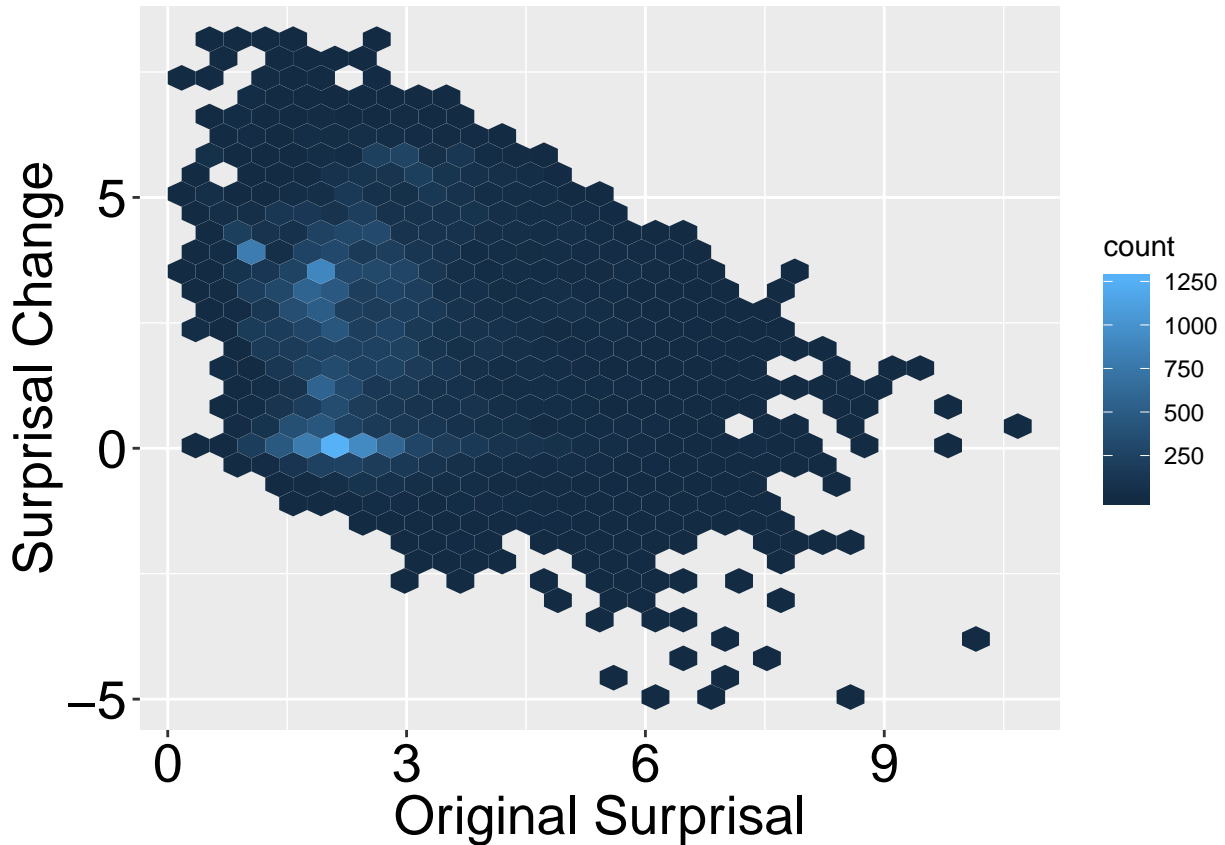


```
## [1] "----- Expression Global Type Model ----- "
## [1] "PyLogicSwapTopFiltered100GlobalTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -216.72, df = 32218, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.221463
## sample estimates:
## mean of the differences
##      -2.251495
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -216.72, df = 32218, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##      -2.283724 -2.219267
```

```

## sample estimates:
## mean of the differences
##           -2.251495
##
##
## Cohen's d
##
## d estimate: -1.207363 (large)
## 95 percent confidence interval:
##      inf      sup
## -1.224153 -1.190572
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10683000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.409512
## sample estimates:
## (pseudo)median
##      -2.443928
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10683000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -2.480461 -2.407098
## sample estimates:
## (pseudo)median
##      -2.443928
##
##
## Cliff's Delta
##
## delta estimate: -0.6480307 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.6544292 -0.6415395

```

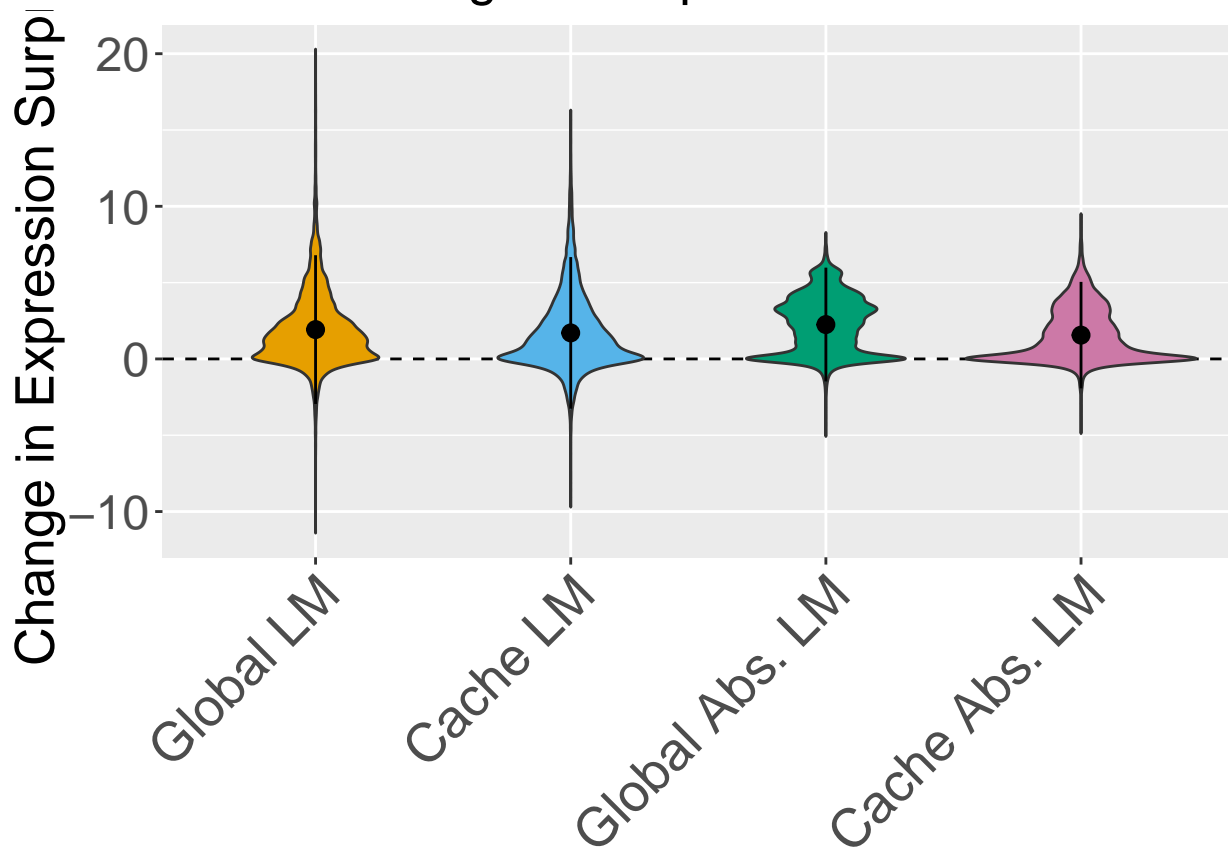
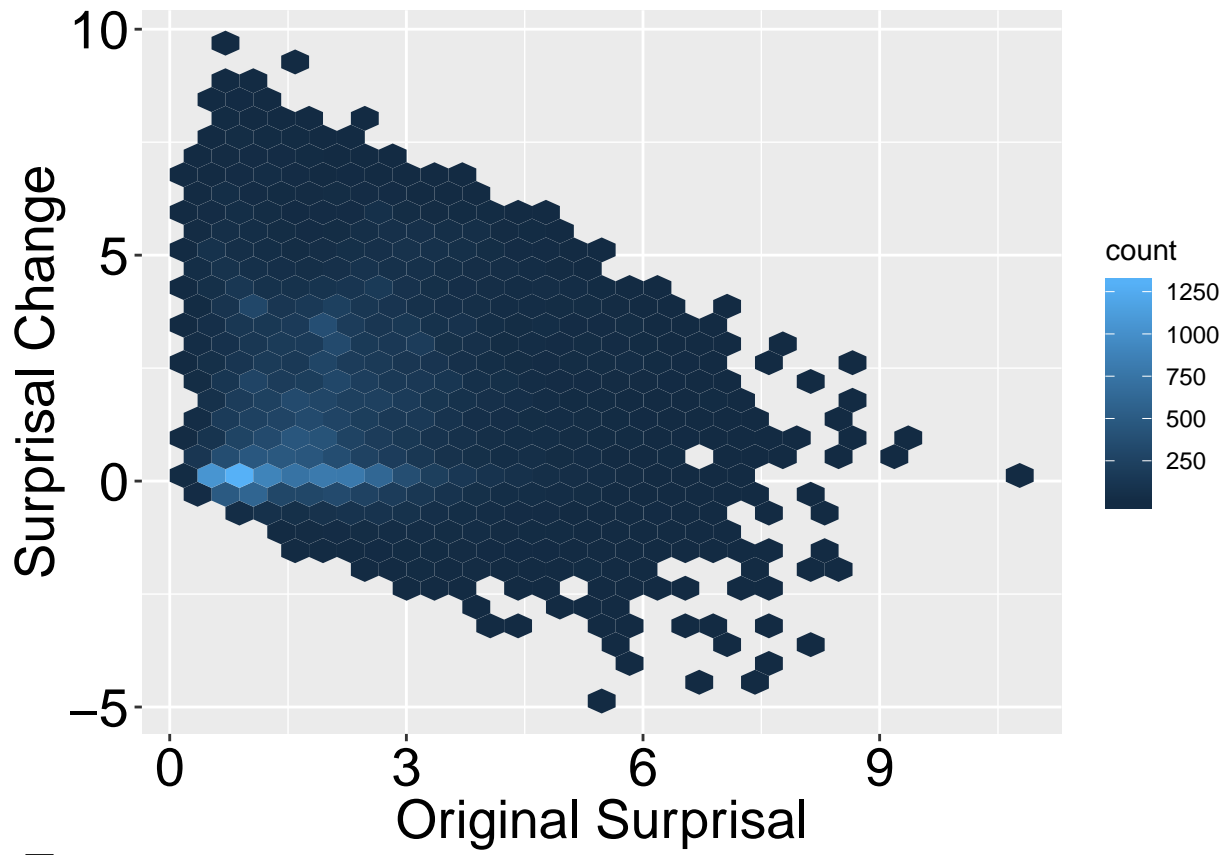


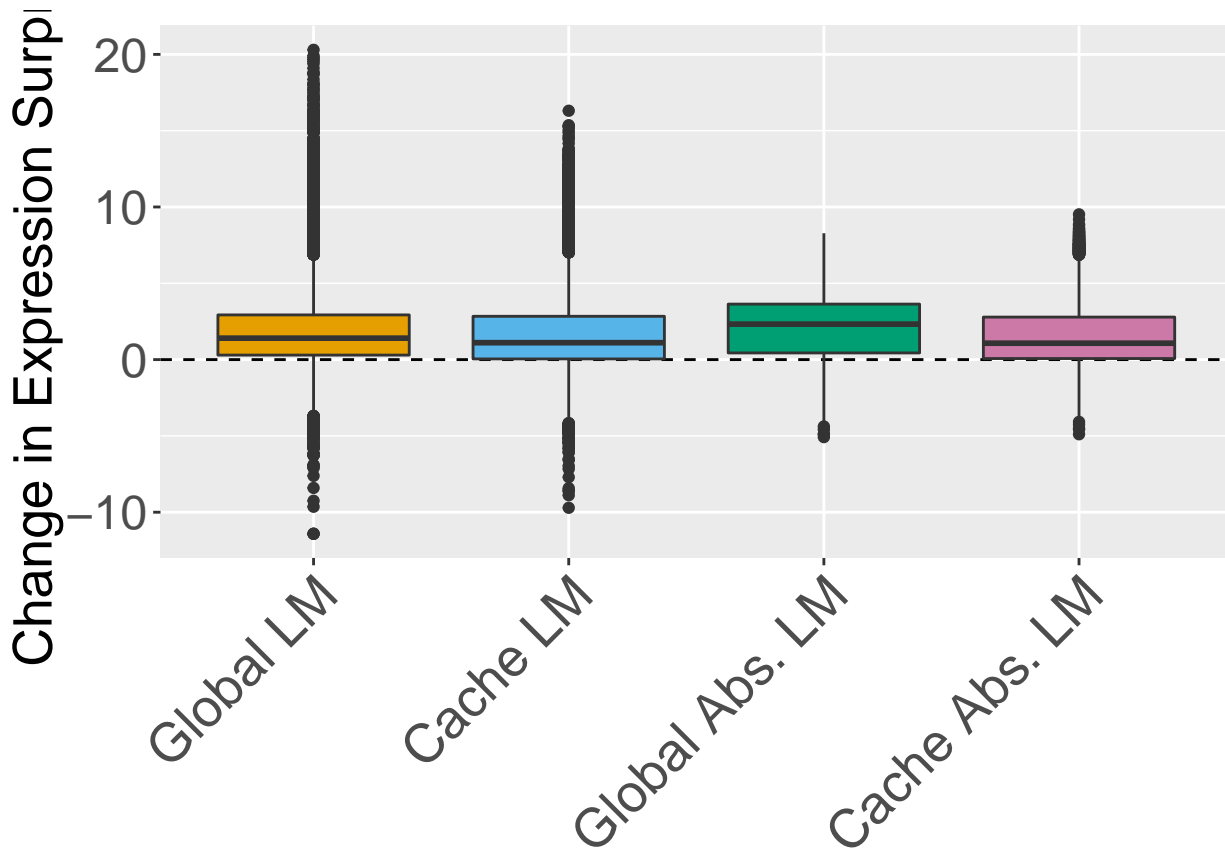
```
## [1] " ----- Expression Cache Type Model ----- "
## [1] "PyLogicSwapTopFiltered100CacheTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -160.49, df = 32218, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.531501
## sample estimates:
## mean of the differences
##      -1.559591
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -160.49, df = 32218, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -1.589736 -1.529447
## sample estimates:
## mean of the differences
##      -1.559591
##
##
```

```

## Cohen's d
##
## d estimate: -0.894132 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.9103277 -0.8779364
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 35074000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.477243
## sample estimates:
## (pseudo)median
##      -1.509505
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 35074000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.543917 -1.474885
## sample estimates:
## (pseudo)median
##      -1.509505
##
## Cliff's Delta
##
## delta estimate: -0.4314041 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.4393389 -0.4234020
##
## No id variables; using all as measure variables
## Warning: Ignoring unknown parameters: mult

```





```
dsppFiltered2 <- generateFilteredResults(dspp, "PyLogic", "PyLogicSwapTopFiltered10", 10)
```

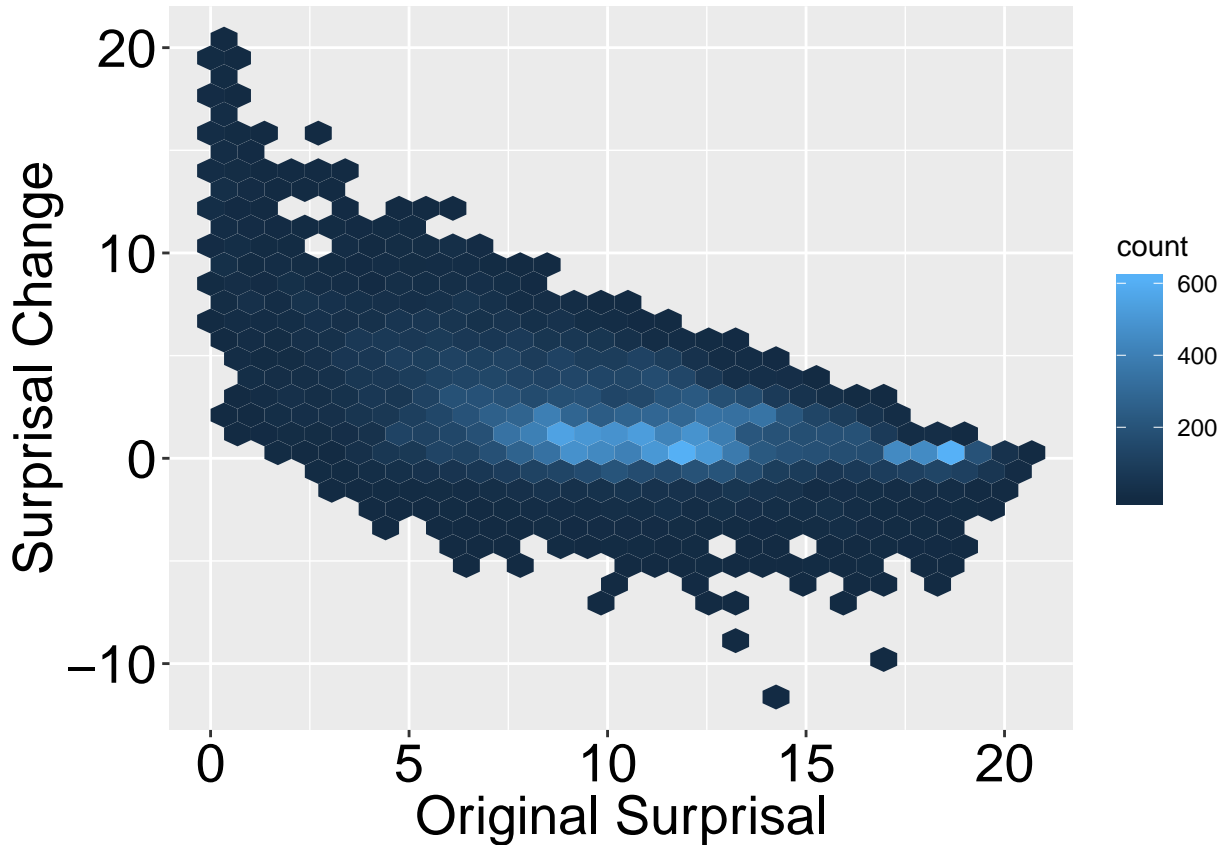
```
## [1] " ----- Expression Global Model ----- "
## [1] "PyLogicSwapTopFiltered10GlobalExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -128.83, df = 29139, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.764487
## sample estimates:
## mean of the differences
##      -1.804988
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -128.83, df = 29139, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -1.848452 -1.761525
## sample estimates:
## mean of the differences
##      -1.804988
```



```

##
##
## Cohen's d
##
## d estimate: -0.7546884 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.7714943 -0.7378826
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 33594000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.575088
## sample estimates:
## (pseudo)median
##      -1.611943
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 33594000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.652068 -1.572409
## sample estimates:
## (pseudo)median
##      -1.611943
##
##
## Cliff's Delta
##
## delta estimate: -0.268855 (small)
## 95 percent confidence interval:
##      inf      sup
## -0.2777984 -0.2598651

```

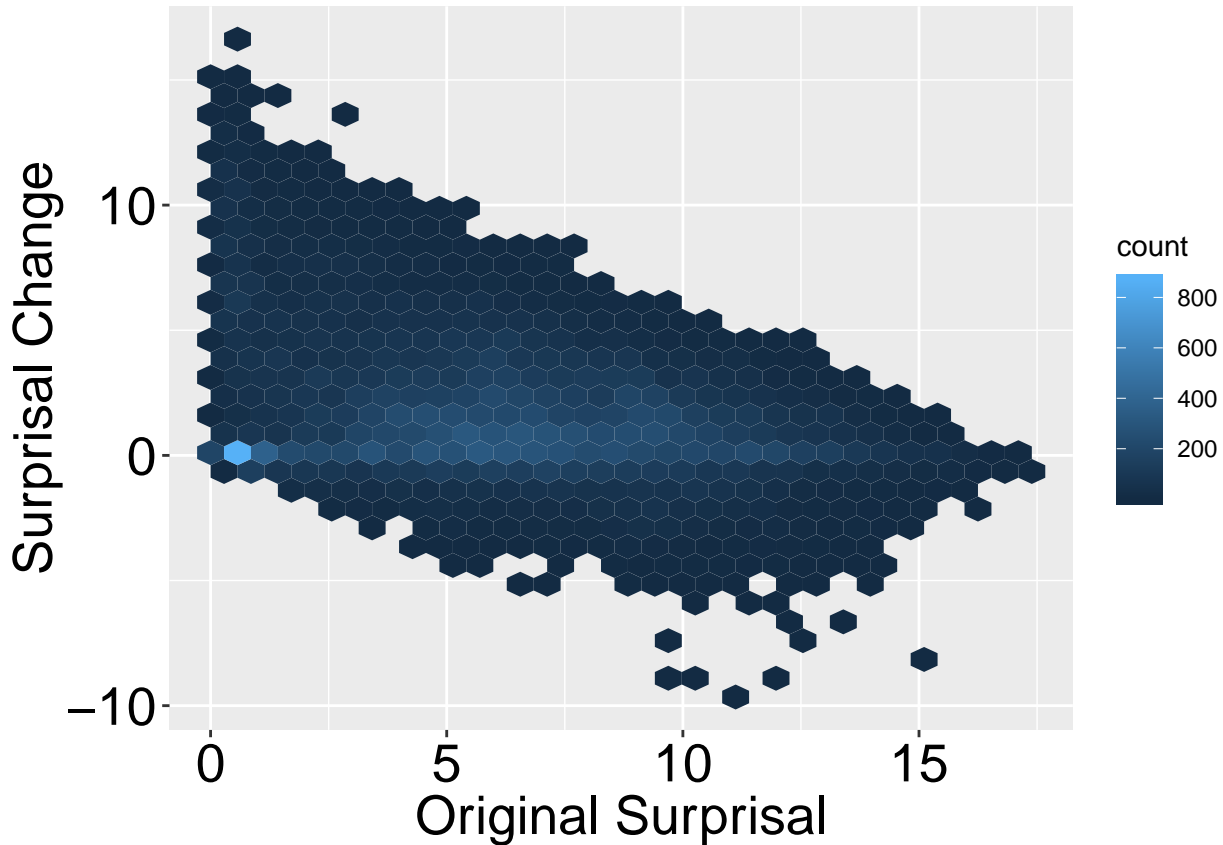


```
## [1] " ----- Expression Cache Model ----- "
## [1] "PyLogicSwapTopFiltered10CacheExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -113.54, df = 29139, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.589186
## sample estimates:
## mean of the differences
##      -1.630704
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -113.54, df = 29139, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##      -1.675258 -1.586149
## sample estimates:
## mean of the differences
##      -1.630704
##
##
```

```

## Cohen's d
##
## d estimate: -0.6651216 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.6818024 -0.6484409
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 60617000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.316658
## sample estimates:
## (pseudo)median
##      -1.355293
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 60617000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.397243 -1.313837
## sample estimates:
## (pseudo)median
##      -1.355293
##
## Cliff's Delta
##
## delta estimate: -0.2588168 (small)
## 95 percent confidence interval:
##      inf      sup
## -0.2677832 -0.2498056

```

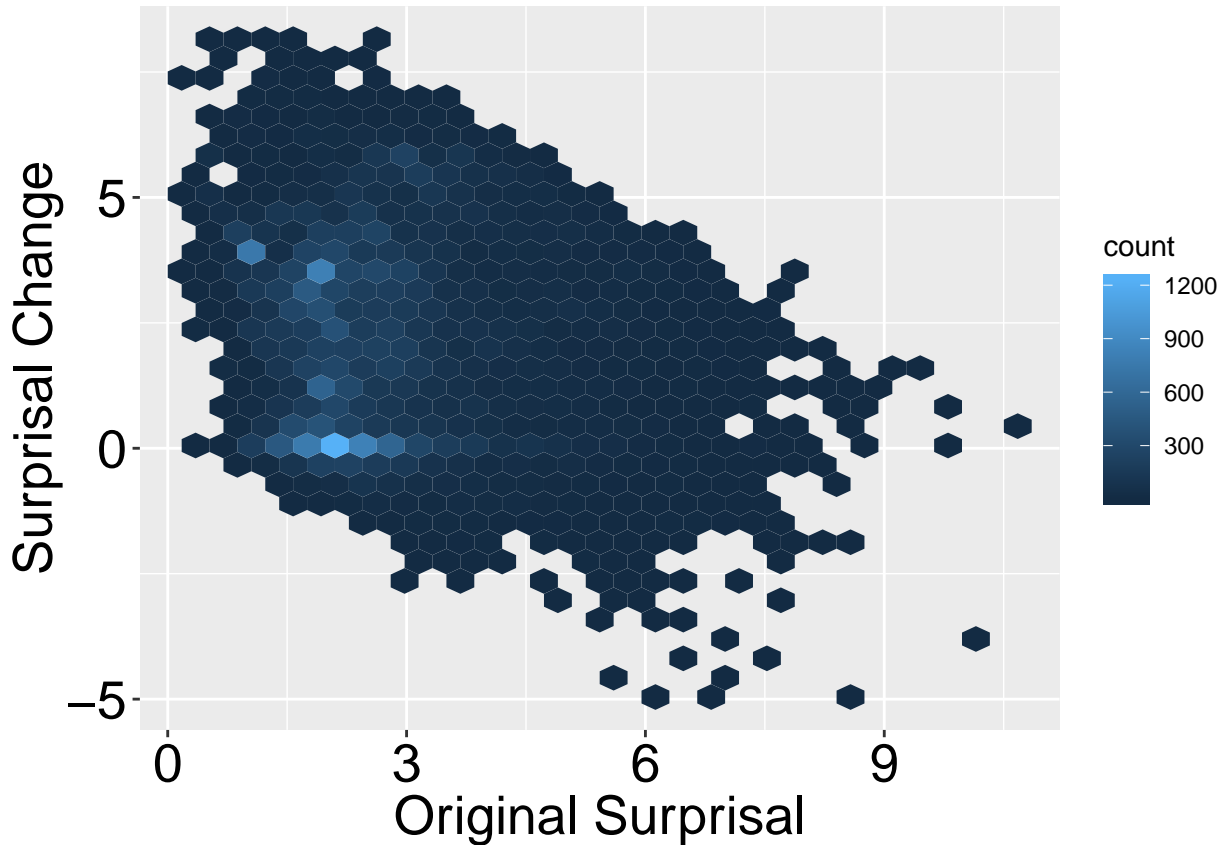


```
## [1] " ----- Expression Global Type Model ----- "
## [1] "PyLogicSwapTopFiltered10GlobalTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -198.06, df = 29139, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.149526
## sample estimates:
## mean of the differences
##      -2.181365
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -198.06, df = 29139, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##      -2.215532 -2.147198
## sample estimates:
## mean of the differences
##      -2.181365
##
##
```

```

## Cohen's d
##
## d estimate: -1.160223 (large)
## 95 percent confidence interval:
##      inf      sup
## -1.177774 -1.142672
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10068000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.33403
## sample estimates:
## (pseudo)median
##      -2.369978
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 10068000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -2.408596 -2.331489
## sample estimates:
## (pseudo)median
##      -2.369978
##
## Cliff's Delta
##
## delta estimate: -0.6231244 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.6300651 -0.6160842

```

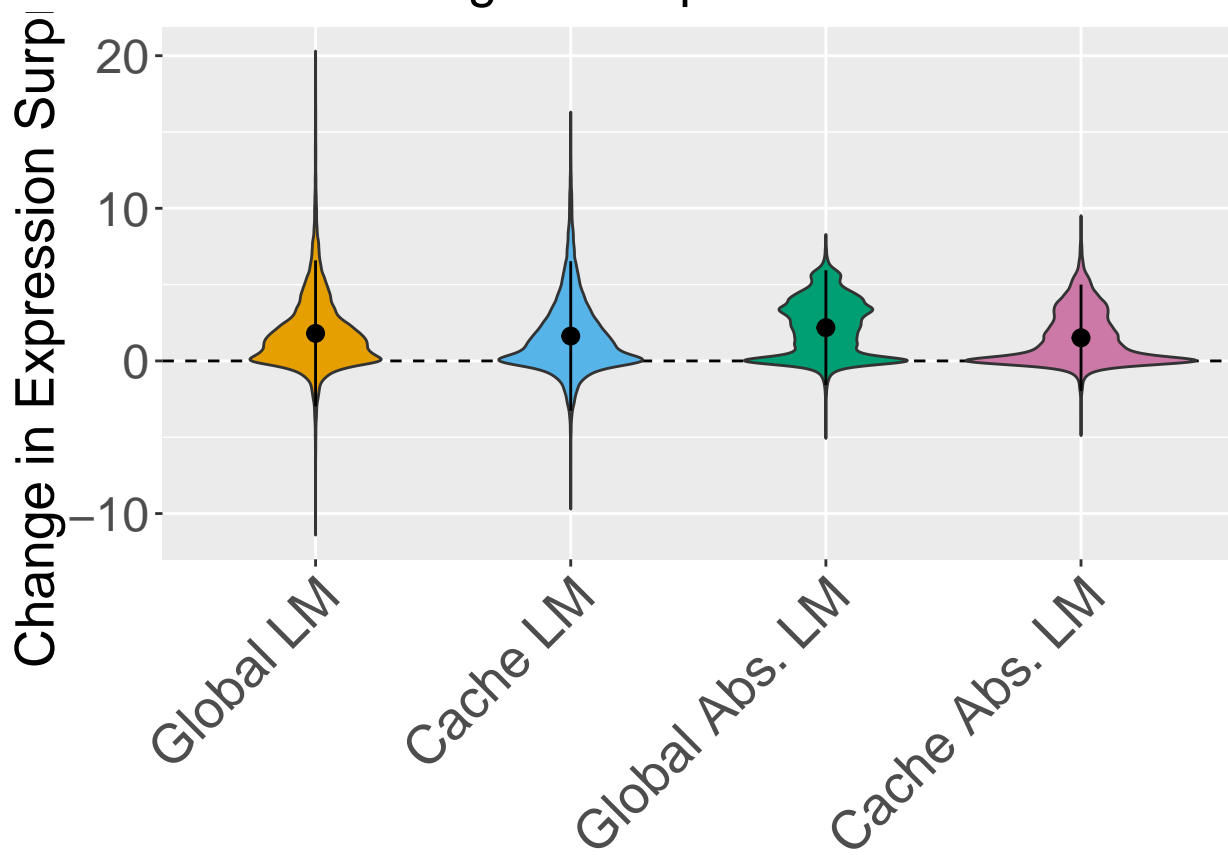
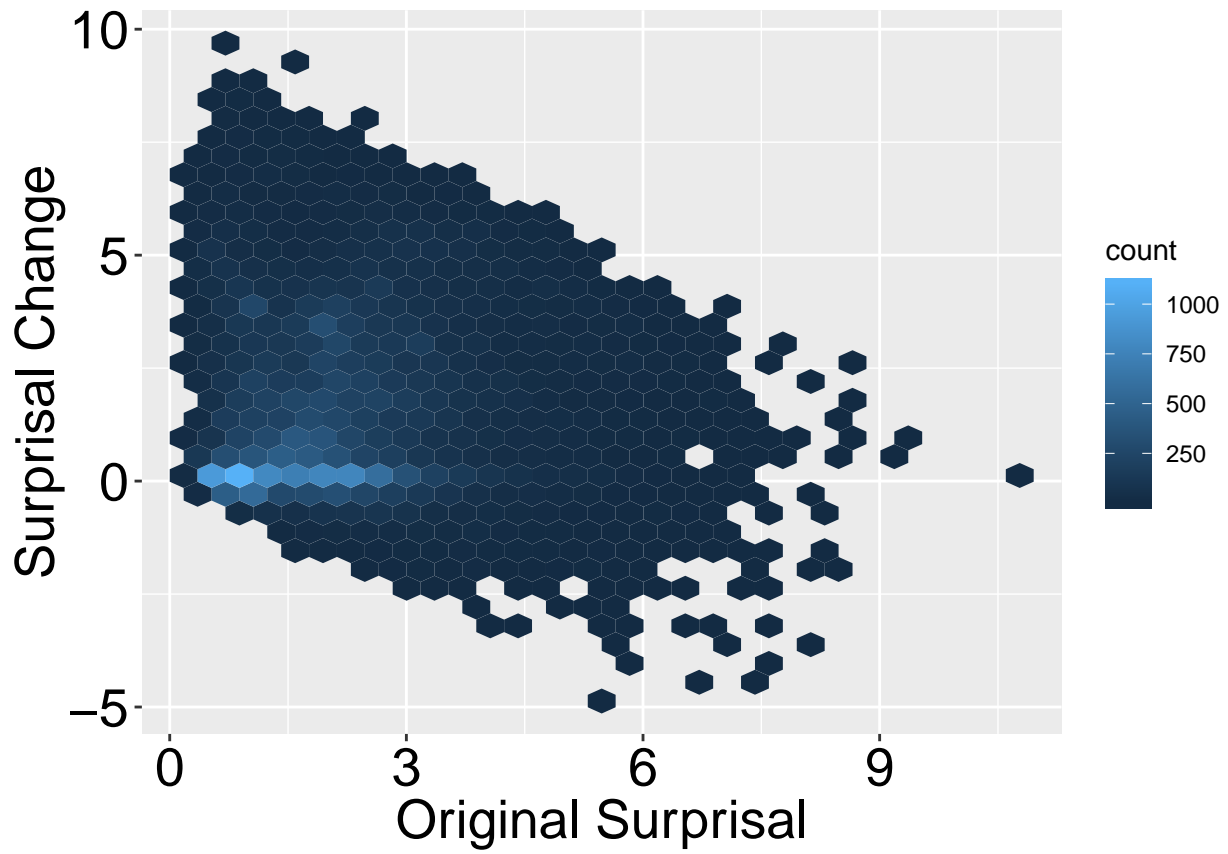


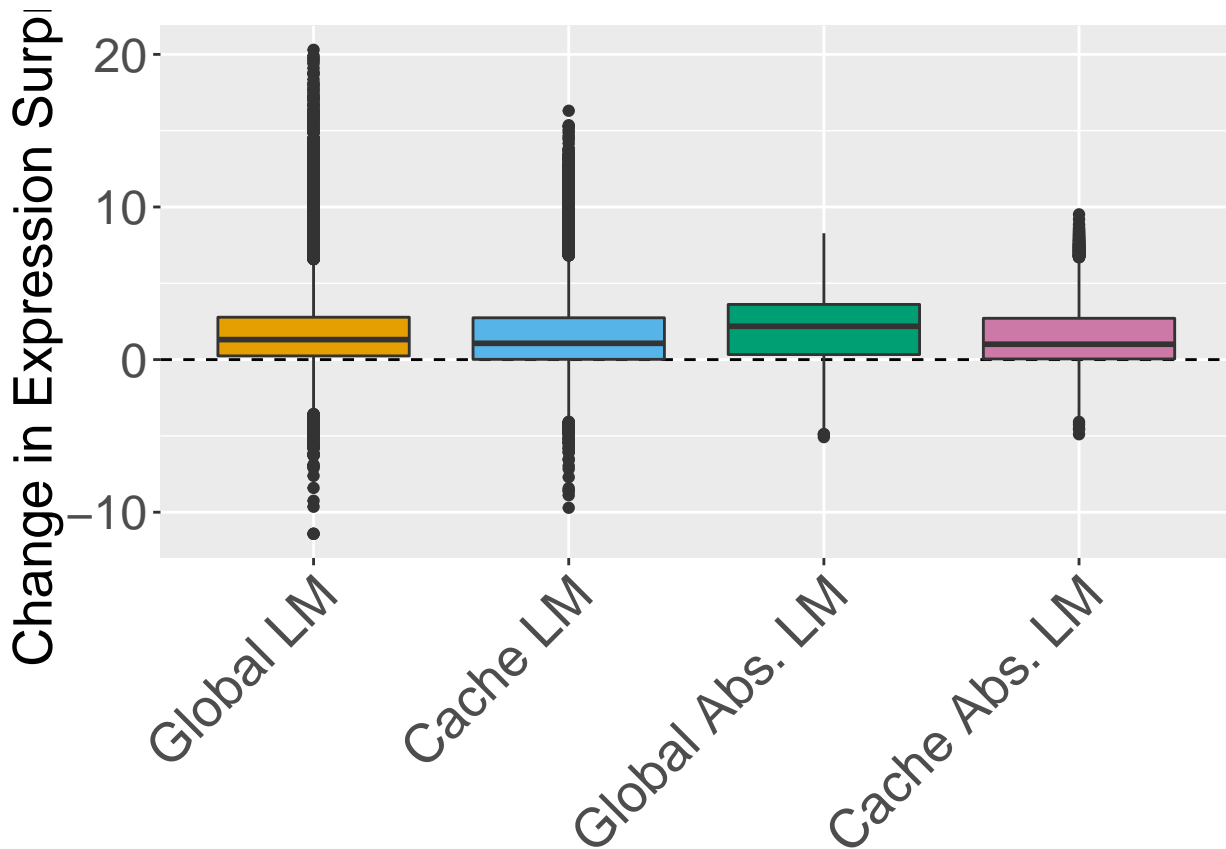
```
## [1] " ----- Expression Cache Type Model ----- "
## [1] "PyLogicSwapTopFiltered10CacheTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -148.46, df = 29139, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.486032
## sample estimates:
## mean of the differences
##      -1.515541
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -148.46, df = 29139, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -1.547208 -1.483874
## sample estimates:
## mean of the differences
##      -1.515541
##
##
```

```

## Cohen's d
##
## d estimate: -0.8697131 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.8867012 -0.8527250
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 31224000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.422483
## sample estimates:
## (pseudo)median
##      -1.457267
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 31224000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.494392 -1.419912
## sample estimates:
## (pseudo)median
##      -1.457267
##
## Cliff's Delta
##
## delta estimate: -0.4153413 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.4237574 -0.4068535
## No id variables; using all as measure variables
## Warning: Ignoring unknown parameters: mult

```





Regression models

```
m_plswap_no_out <- modelGlobal(dsppFiltered, "", "===")
```

```
##
## Call:
## lm(formula = AverageEntChangeExp ~ BaseAveEntExp + log(NumTokens) +
##     factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.1883 -0.9546 -0.0871  0.8955  6.7654
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    10.051280   0.079942  125.733 < 2e-16 ***
## BaseAveEntExp  -0.349030   0.002291 -152.364 < 2e-16 ***
## log(NumTokens) -2.074786   0.021474 -96.617 < 2e-16 ***
## factor(ParentOp)Call    0.357580   0.207689   1.722  0.08513 .
## factor(ParentOp)If     -0.232375   0.050050  -4.643  3.45e-06 ***
## factor(ParentOp)Return -0.545396   0.089588  -6.088  1.16e-09 ***
## factor(ParentOp)While  -0.202233   0.065479  -3.088  0.00201 **
## factor(MostFreqOp)<    -0.286692   0.034331  -8.351 < 2e-16 ***
## factor(MostFreqOp)<=  -0.406658   0.068824  -5.909  3.49e-09 ***
## factor(MostFreqOp)>    -0.491206   0.032711 -15.016 < 2e-16 ***
## factor(MostFreqOp)>=  -0.637562   0.062999 -10.120 < 2e-16 ***
## factor(MostFreqOp)!=   -0.192844   0.023830  -8.093  6.06e-16 ***
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.498 on 30127 degrees of freedom
## Multiple R-squared:  0.4862, Adjusted R-squared:  0.486
## F-statistic: 2592 on 11 and 30127 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: AverageEntChangeExp
##
##           Df Sum Sq Mean Sq    F value    Pr(>F)
## BaseAveEntExp      1  39793   39793 17743.3769 < 2.2e-16 ***
## log(NumTokens)     1  23203   23203 10346.0685 < 2.2e-16 ***
## factor(ParentOp)   4     84     21    9.3624  1.47e-07 ***
## factor(MostFreqOp) 5    864    173   77.0701 < 2.2e-16 ***
## Residuals         30127  67566     2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.3025853540" "0.1764359062" "0.0006386447" "0.0065715499"
## [5] "0.5137685452"
##
##           GVIF Df GVIF^(1/(2*Df))
## BaseAveEntExp  1.103404  1      1.050430
## log(NumTokens) 1.164535  1      1.079136
## factor(ParentOp) 1.081555  4      1.009848
## factor(MostFreqOp) 1.066002  5      1.006412
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: Tue, Feb 19, 2019 - 11:48:26 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lc}
##     \hline
##     \hline \hline
##     & \multicolumn{1}{c}{\textit{Dependent variable:}} & \\
##     \cline{2-2}
##     \hline & AverageEntChangeExp & \\
##     \hline \hline
##     BaseAveEntExp & $-0.349$^{***}$ (0.002) & \\
##     log(NumTokens) & $-2.075$^{***}$ (0.021) & \\
##     factor(ParentOp)Call & 0.358$^{*}$ (0.208) & \\
##     factor(ParentOp)If & $-0.232$^{***}$ (0.050) & \\
##     factor(ParentOp)Return & $-0.545$^{***}$ (0.090) & \\
##     factor(ParentOp)While & $-0.202$^{***}$ (0.065) & \\
##     factor(MostFreqOp)\textless & $-0.287$^{***}$ (0.034) & \\
##     factor(MostFreqOp)\textless = & $-0.407$^{***}$ (0.069) & \\
##     factor(MostFreqOp)\textgreater & $-0.491$^{***}$ (0.033) & \\
##     factor(MostFreqOp)\textgreater = & $-0.638$^{***}$ (0.063) & \\
##     factor(MostFreqOp)!= & $-0.193$^{***}$ (0.024) & \\
##     Constant & 10.051$^{***}$ (0.080) & \\
##     \hline \hline
##     Observations & 30,139 & \\
##     R$^2$ & 0.486 & \\
##     Adjusted R$^2$ & 0.486 & \\
##     Residual Std. Error & 1.498 (df = 30127) & \\

```

```

## F Statistic & 2,592.022$^{***}$ (df = 11; 30127) \\
## \hline
## \hline \[-1.8ex]
## \textit{Note:} & \multicolumn{1}{r}{\${^*}$p$<$0.1; \${^{**}$p$<$0.05; \${^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 23:48:27 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
## \hline
## & Df & Sum Sq & Mean Sq & F value & Pr(>$F) \\
## \hline
## BaseAveEntExp & 1 & 39792.90 & 39792.90 & 17743.38 & 0.0000 \\
## log(NumTokens) & 1 & 23203.03 & 23203.03 & 10346.07 & 0.0000 \\
## factor(ParentOp) & 4 & 83.99 & 21.00 & 9.36 & 0.0000 \\
## factor(MostFreqOp) & 5 & 864.22 & 172.84 & 77.07 & 0.0000 \\
## Residuals & 30127 & 67565.53 & 2.24 & & \\
## \hline
## \end{tabular}
## \end{table}
m_plswap_cache_no_out <- modelCache(dsppFiltered, "", "==" )

##
## Call:
## lm(formula = CacheAverageEntChangeExp ~ BaseCacheAveEntExp +
## log(NumTokens) + factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
## Min 1Q Median 3Q Max
## -7.1123 -1.3213 -0.1594 1.1577 8.2429
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.353707 0.099488 53.812 < 2e-16 ***
## BaseCacheAveEntExp -0.196068 0.003342 -58.663 < 2e-16 ***
## log(NumTokens) -1.093315 0.027843 -39.267 < 2e-16 ***
## factor(ParentOp)Call -0.134520 0.302420 -0.445 0.65646
## factor(ParentOp)If -0.162421 0.065165 -2.492 0.01269 *
## factor(ParentOp)Return -0.050558 0.109504 -0.462 0.64430
## factor(ParentOp)While -0.229030 0.086945 -2.634 0.00844 **
## factor(MostFreqOp)< -0.525304 0.045198 -11.622 < 2e-16 ***
## factor(MostFreqOp)<= -0.673600 0.091346 -7.374 1.7e-13 ***
## factor(MostFreqOp)> -0.675480 0.042988 -15.713 < 2e-16 ***
## factor(MostFreqOp)>= -0.747916 0.077244 -9.683 < 2e-16 ***
## factor(MostFreqOp)!= -0.470097 0.031415 -14.964 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.985 on 30364 degrees of freedom
## Multiple R-squared: 0.153, Adjusted R-squared: 0.1527
## F-statistic: 498.5 on 11 and 30364 DF, p-value: < 2.2e-16
##

```

```

## Analysis of Variance Table
##
## Response: CacheAverageEntChangeExp
##           Df Sum Sq Mean Sq  F value  Pr(>F)
## BaseCacheAveEntExp      1  11884  11884.3  3017.4049 < 2e-16 ***
## log(NumTokens)          1   7508   7508.5  1906.3871 < 2e-16 ***
## factor(ParentOp)         4     43    10.7    2.7245 0.02775 *
## factor(MostFreqOp)       5   2163   432.6  109.8467 < 2e-16 ***
## Residuals                30364 119591     3.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.0841723751" "0.0531798482" "0.0003040035" "0.0153212125"
## [5] "0.8470225607"
##           GVIF Df GVIF^(1/(2*Df))
## BaseCacheAveEntExp 1.079897 1      1.039181
## log(NumTokens)     1.136888 1      1.066250
## factor(ParentOp)   1.094610 4      1.011364
## factor(MostFreqOp) 1.069089 5      1.006703
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: Tue, Feb 19, 2019 - 11:48:35 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lc}
##     \hline
##     \hline \hline
##     & \multicolumn{1}{c}{\textit{Dependent variable:}} & \\
##     \cline{2-2}
##     \hline \hline
##     BaseCacheAveEntExp &  $-\$0.196^{\{***\}}$  & (0.003) \\
##     log(NumTokens) &  $-\$1.093^{\{***\}}$  & (0.028) \\
##     factor(ParentOp)Call &  $-\$0.135$  & (0.302) \\
##     factor(ParentOp)If &  $-\$0.162^{\{**\}}$  & (0.065) \\
##     factor(ParentOp)Return &  $-\$0.051$  & (0.110) \\
##     factor(ParentOp)While &  $-\$0.229^{\{***\}}$  & (0.087) \\
##     factor(MostFreqOp)\textless &  $-\$0.525^{\{***\}}$  & (0.045) \\
##     factor(MostFreqOp)\textless = &  $-\$0.674^{\{***\}}$  & (0.091) \\
##     factor(MostFreqOp)\textgreater &  $-\$0.675^{\{***\}}$  & (0.043) \\
##     factor(MostFreqOp)\textgreater = &  $-\$0.748^{\{***\}}$  & (0.077) \\
##     factor(MostFreqOp)\textgreater != &  $-\$0.470^{\{***\}}$  & (0.031) \\
##     Constant &  $5.354^{\{***\}}$  & (0.099) \\
##     \hline \hline
##     Observations & 30,376 & \\
##     R2 & 0.153 & \\
##     Adjusted R2 & 0.153 & \\
##     Residual Std. Error & 1.985 (df = 30364) & \\
##     F Statistic &  $498.539^{\{***\}}$  (df = 11; 30364) & \\
##     \hline
##     \hline \hline
##     \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 23:48:35 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
## \hline
## & Df & Sum Sq & Mean Sq & F value & Pr(>F) \\
## \hline
## BaseCacheAveEntExp & 1 & 11884.29 & 11884.29 & 3017.40 & 0.0000 \\
## log(NumTokens) & 1 & 7508.46 & 7508.46 & 1906.39 & 0.0000 \\
## factor(ParentOp) & 4 & 42.92 & 10.73 & 2.72 & 0.0278 \\
## factor(MostFreqOp) & 5 & 2163.20 & 432.64 & 109.85 & 0.0000 \\
## Residuals & 30364 & 119591.04 & 3.94 & & \\
## \hline
## \end{tabular}
## \end{table}

```

```
mt_plswap_no_out <- modelGlobalType(dsppFiltered, "", "==" )
```

```

##
## Call:
## lm(formula = TypeAverageEntChangeExp ~ BaseTypeAveEntExp + log(NumTokens) +
##     factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.8958 -1.0763  0.0999  1.0324  5.6171
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.948730   0.077999   63.446 < 2e-16 ***
## BaseTypeAveEntExp  0.203249   0.009371   21.690 < 2e-16 ***
## log(NumTokens)    -1.580189   0.021097  -74.903 < 2e-16 ***
## factor(ParentOp)Call  0.682959   0.209070    3.267  0.00109 **
## factor(ParentOp)If    0.382300   0.051970    7.356 1.94e-13 ***
## factor(ParentOp)Return -0.043481   0.091352   -0.476  0.63410
## factor(ParentOp)While  0.258895   0.069844    3.707  0.00021 ***
## factor(MostFreqOp)<   -1.049675   0.036619  -28.665 < 2e-16 ***
## factor(MostFreqOp)<=  -0.952502   0.083721  -11.377 < 2e-16 ***
## factor(MostFreqOp)>   -1.124954   0.034486  -32.620 < 2e-16 ***
## factor(MostFreqOp)>=  -1.943099   0.058127  -33.428 < 2e-16 ***
## factor(MostFreqOp)!=  -0.718510   0.024769  -29.008 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.53 on 30047 degrees of freedom
## Multiple R-squared:  0.2446, Adjusted R-squared:  0.2443
## F-statistic: 884.4 on 11 and 30047 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: TypeAverageEntChangeExp
##              Df Sum Sq Mean Sq  F value    Pr(>F)
## BaseTypeAveEntExp      1      90      90.2   38.521 5.488e-10 ***
## log(NumTokens)         1  16035 16034.9  6848.564 < 2.2e-16 ***

```

```

## factor(ParentOp)          4    283    70.7    30.201 < 2.2e-16 ***
## factor(MostFreqOp)       5   6370  1274.1   544.164 < 2.2e-16 ***
## Residuals                 30047  70350    2.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.0009684507" "0.1721795536" "0.0030371717" "0.0684040486"
## [5] "0.7554107754"
##
##              GVIF Df GVIF^(1/(2*Df))
## BaseTypeAveEntExp  1.234996  1      1.111304
## log(NumTokens)    1.076845  1      1.037711
## factor(ParentOp)  1.137130  4      1.016193
## factor(MostFreqOp) 1.235300  5      1.021356
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: Tue, Feb 19, 2019 - 11:48:38 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lc}
##     \hline
##     \hline \hline \hline
##     & \multicolumn{1}{c}{\textit{Dependent variable:}} & \hline
##     \cline{2-2}
##     \hline \hline & TypeAverageEntChangeExp & \hline
##     \hline \hline
##     BaseTypeAveEntExp & 0.203$^{***}$ (0.009) & \hline
##     log(NumTokens) & $-1.580$^{***}$ (0.021) & \hline
##     factor(ParentOp)Call & 0.683$^{***}$ (0.209) & \hline
##     factor(ParentOp)If & 0.382$^{***}$ (0.052) & \hline
##     factor(ParentOp)Return & $-0.043$ (0.091) & \hline
##     factor(ParentOp)While & 0.259$^{***}$ (0.070) & \hline
##     factor(MostFreqOp)\textless & $-1.050$^{***}$ (0.037) & \hline
##     factor(MostFreqOp)\textless = & $-0.953$^{***}$ (0.084) & \hline
##     factor(MostFreqOp)\textgreater & $-1.125$^{***}$ (0.034) & \hline
##     factor(MostFreqOp)\textgreater = & $-1.943$^{***}$ (0.058) & \hline
##     factor(MostFreqOp)!= & $-0.719$^{***}$ (0.025) & \hline
##     Constant & 4.949$^{***}$ (0.078) & \hline
##     \hline \hline
##     Observations & 30,059 & \hline
##     R$^{2}$ & 0.245 & \hline
##     Adjusted R$^{2}$ & 0.244 & \hline
##     Residual Std. Error & 1.530 (df = 30047) & \hline
##     F Statistic & 884.428$^{***}$ (df = 11; 30047) & \hline
##     \hline
##     \hline \hline
##     \textit{Note:} & \multicolumn{1}{r}{$^{*}$p<$0.1; $^{**}$p<$0.05; $^{***}$p<$0.01} & \hline
##     \end{tabular}
##   \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 23:48:38 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
## \hline

```

```

## & Df & Sum Sq & Mean Sq & F value & Pr(>F) \\
## \hline
## BaseTypeAveEntExp & 1 & 90.19 & 90.19 & 38.52 & 0.0000 \\
## log(NumTokens) & 1 & 16034.86 & 16034.86 & 6848.56 & 0.0000 \\
## factor(ParentOp) & 4 & 282.85 & 70.71 & 30.20 & 0.0000 \\
## factor(MostFreqOp) & 5 & 6370.38 & 1274.08 & 544.16 & 0.0000 \\
## Residuals & 30047 & 70350.41 & 2.34 & & \\
## \hline
## \end{tabular}
## \end{table}

```

```

mt_plswap_cache_no_out <- modelCacheType(dsppFiltered, "", "==" )

```

```

##
## Call:
## lm(formula = CacheTypeAverageEntChangeExp ~ BaseCacheTypeAveEntExp +
## log(NumTokens) + factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.4698 -1.1541 -0.3087  1.0494  7.1971
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.161800   0.073034  29.600 <2e-16 ***
## BaseCacheTypeAveEntExp  0.361826   0.008201  44.121 <2e-16 ***
## log(NumTokens)     -0.769710   0.020420 -37.694 <2e-16 ***
## factor(ParentOp)Call  0.172234   0.195770   0.880  0.3790
## factor(ParentOp)If    0.476985   0.049402   9.655 <2e-16 ***
## factor(ParentOp)Return 0.146853   0.083279   1.763  0.0778 .
## factor(ParentOp)While 0.114202   0.064939   1.759  0.0787 .
## factor(MostFreqOp)<  -0.821959   0.034652 -23.721 <2e-16 ***
## factor(MostFreqOp)<= -0.886986   0.073932 -11.997 <2e-16 ***
## factor(MostFreqOp)>  -0.855237   0.032924 -25.976 <2e-16 ***
## factor(MostFreqOp)>= -1.377602   0.056737 -24.281 <2e-16 ***
## factor(MostFreqOp)!= -0.629947   0.023872 -26.389 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.494 on 30421 degrees of freedom
## Multiple R-squared:  0.1439, Adjusted R-squared:  0.1436
## F-statistic:  465 on 11 and 30421 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: CacheTypeAverageEntChangeExp
##              Df Sum Sq Mean Sq  F value    Pr(>F)
## BaseCacheTypeAveEntExp      1    2012   2012.4   901.540 < 2.2e-16 ***
## log(NumTokens)              1    4618   4617.7  2068.742 < 2.2e-16 ***
## factor(ParentOp)            4     622    155.5   69.668 < 2.2e-16 ***
## factor(MostFreqOp)          5    4165    833.1  373.226 < 2.2e-16 ***
## Residuals                   30421  67904     2.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.02536972" "0.05821526" "0.00784192" "0.05251373" "0.85605938"

```

```

##                               GVIF Df GVIF^(1/(2*Df))
## BaseCacheTypeAveEntExp 1.116132 1      1.056472
## log(NumTokens)          1.078696 1      1.038603
## factor(ParentOp)        1.113482 4      1.013527
## factor(MostFreqOp)      1.149316 5      1.014014
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: Tue, Feb 19, 2019 - 11:48:42 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lc}
##     \hline
##     \hline \hline \hline
##     & \multicolumn{1}{c}{\textit{Dependent variable:}} & \\
##     \cline{2-2}
##     \hline \hline \hline \hline
##     BaseCacheTypeAveEntExp & 0.362$^{***}$ (0.008) & \\
##     log(NumTokens) & -$0.770$^{***}$ (0.020) & \\
##     factor(ParentOp)Call & 0.172 (0.196) & \\
##     factor(ParentOp)If & 0.477$^{***}$ (0.049) & \\
##     factor(ParentOp)Return & 0.147$^{*}$ (0.083) & \\
##     factor(ParentOp)While & 0.114$^{*}$ (0.065) & \\
##     factor(MostFreqOp)\textless & -$0.822$^{***}$ (0.035) & \\
##     factor(MostFreqOp)\textless = & -$0.887$^{***}$ (0.074) & \\
##     factor(MostFreqOp)\textgreater & -$0.855$^{***}$ (0.033) & \\
##     factor(MostFreqOp)\textgreater = & -$1.378$^{***}$ (0.057) & \\
##     factor(MostFreqOp)!= & -$0.630$^{***}$ (0.024) & \\
##     Constant & 2.162$^{***}$ (0.073) & \\
##     \hline \hline \hline \hline
##     Observations & 30,433 & \\
##     R$^{2}$ & 0.144 & \\
##     Adjusted R$^{2}$ & 0.144 & \\
##     Residual Std. Error & 1.494 (df = 30421) & \\
##     F Statistic & 465.008$^{***}$ (df = 11; 30421) & \\
##     \hline
##     \hline \hline \hline \hline
##     \textit{Note:} & \multicolumn{1}{r}{$^{*}$p<$0.1; $^{**}$p<$0.05; $^{***}$p<$0.01} & \\
##     \end{tabular}
##   \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 23:48:42 2019
## \begin{table}[ht]
##   \centering
##   \begin{tabular}{lrrrrr}
##     \hline
##     & Df & Sum Sq & Mean Sq & F value & Pr(>$F) & \\
##     \hline
##     BaseCacheTypeAveEntExp & 1 & 2012.36 & 2012.36 & 901.54 & 0.0000 & \\
##     log(NumTokens) & 1 & 4617.71 & 4617.71 & 2068.74 & 0.0000 & \\
##     factor(ParentOp) & 4 & 622.03 & 155.51 & 69.67 & 0.0000 & \\
##     factor(MostFreqOp) & 5 & 4165.46 & 833.09 & 373.23 & 0.0000 & \\
##     Residuals & 30421 & 67903.71 & 2.23 & & & \\

```



```
## \hline
## \end{tabular}
## \end{table}
```

Output the effect sizes

```
PyLogicOut <- printEffTable(pairedResults)
```

```
## [1] "Type,PTOne,PTTwo,CITTwo,CohensD,PWilcoxOne,PWilcoxTwo,CIWilcoxTwo,CliffDelta"
## [1] "PyLogicSwapTopFiltered100CacheExp,0,0,-1.7482 -1.6623,-0.6864,0,0,-1.4606 -1.3788,-0.2693"
## [1] "PyLogicSwapTopFiltered100CacheTypeExp,0,0,-1.5897 -1.5294,-0.8941,0,0,-1.5439 -1.4749,-0.4314"
## [1] "PyLogicSwapTopFiltered100GlobalExp,0,0,-1.9652 -1.881,-0.7893,0,0,-1.7633 -1.6847,-0.2834"
## [1] "PyLogicSwapTopFiltered100GlobalTypeExp,0,0,-2.2837 -2.2193,-1.2074,0,0,-2.4805 -2.4071,-0.648"
## [1] "PyLogicSwapTopFiltered10CacheExp,0,0,-1.6753 -1.5861,-0.6651,0,0,-1.3972 -1.3138,-0.2588"
## [1] "PyLogicSwapTopFiltered10CacheTypeExp,0,0,-1.5472 -1.4839,-0.8697,0,0,-1.4944 -1.4199,-0.4153"
## [1] "PyLogicSwapTopFiltered10GlobalExp,0,0,-1.8485 -1.7615,-0.7547,0,0,-1.6521 -1.5724,-0.2689"
## [1] "PyLogicSwapTopFiltered10GlobalTypeExp,0,0,-2.2155 -2.1472,-1.1602,0,0,-2.4086 -2.3315,-0.6231"
## [1] "PythonLogicalSwapTopCacheExp,0,0,-2.471 -2.3597,-0.7025,0,0,-1.847 -1.7466,-0.3534"
## [1] "PythonLogicalSwapTopCacheTypeExp,0,0,-1.9967 -1.9269,-0.9095,0,0,-1.8795 -1.8094,-0.4802"
## [1] "PythonLogicalSwapTopGlobalExp,0,0,-2.703 -2.5935,-0.7838,0,0,-2.1536 -2.0589,-0.369"
## [1] "PythonLogicalSwapTopGlobalTypeExp,0,0,-2.7276 -2.6537,-1.1791,0,0,-2.8443 -2.7663,-0.7"

save(PyLogicOut, file = "/data/anon/SemanticTransformation/sample/PyLogicOut.RDat")
save(dspp, file = "/data/anon/SemanticTransformation/sample/dspp.RDat")
save(dsppFiltered, file = "/data/anon/SemanticTransformation/sample/dsppFiltered.RDat")
save(dsppFiltered2, file = "/data/anon/SemanticTransformation/sample/dsppFiltered2.RDat")
```