

Java 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(effsize)
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")
```

Java

```
setwd("/data/anon/SemanticTransformation")
#dlsp <- compareDepthSummary("swap_operators.csv", "LogicalSwap", "SAME")
dlsp <- compareDepthSummary("swap_operators_topstarred.csv", "LogicalSwapTop", "SAME")

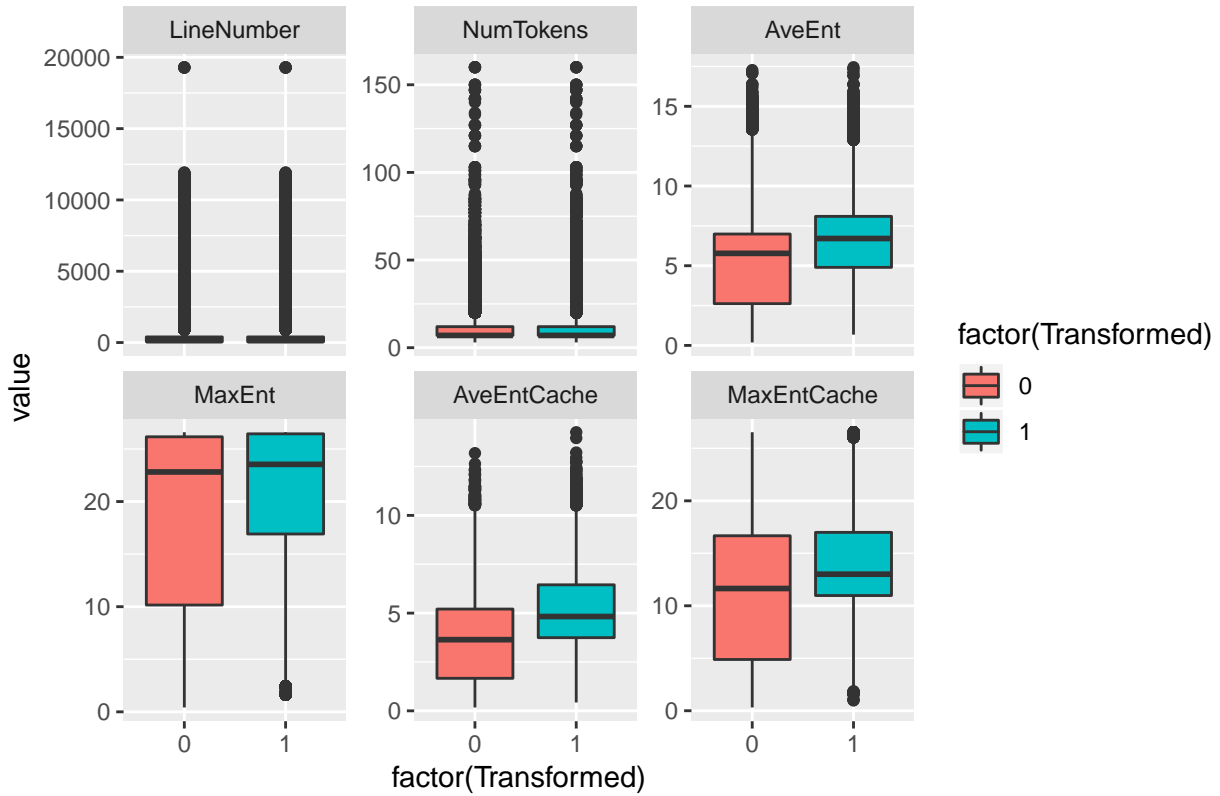
## [[1]]
## [1] "5"      "Bool"

## Loading required package: tcltk

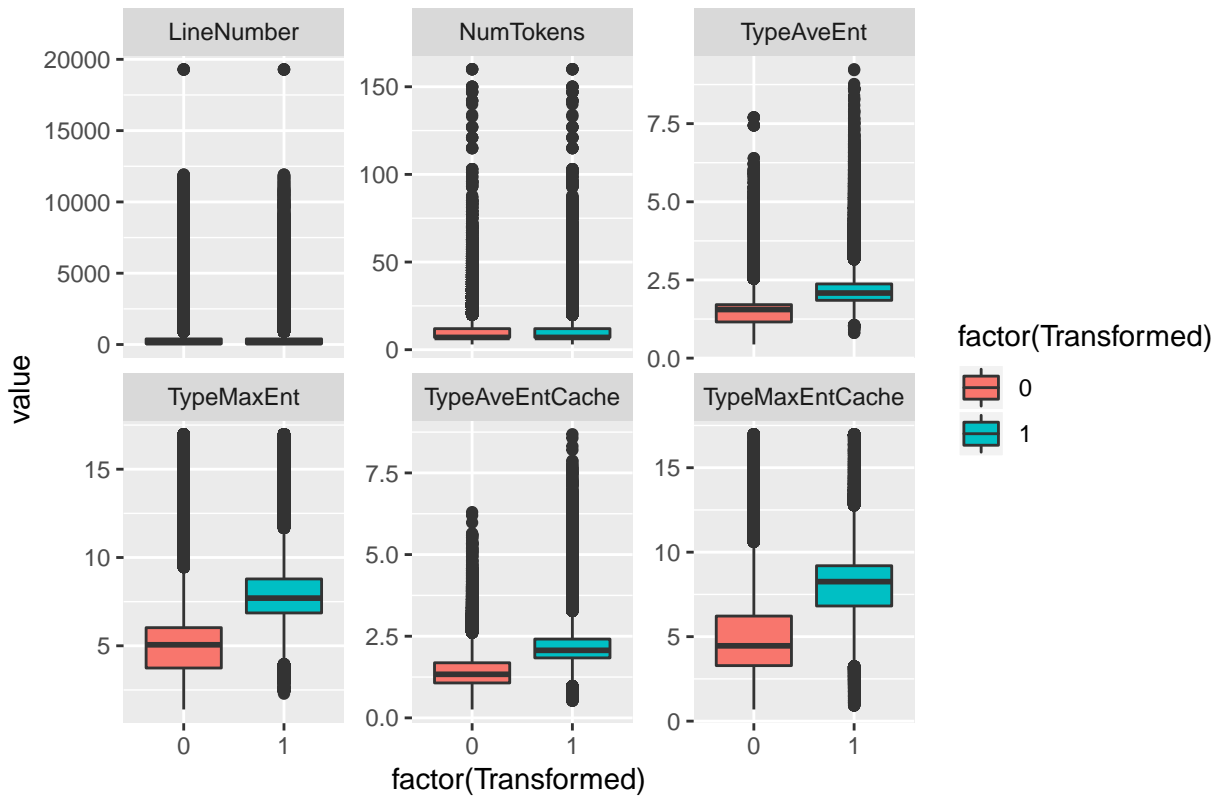
## [1] "TransId"
## [2] "Filepath"
## [3] "LineNumber"
## [4] "NumTokens"
## [5] "Transformed"
## [6] "Source"
## [7] "CleanLexerNumTokens"
```

```
## [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"
```

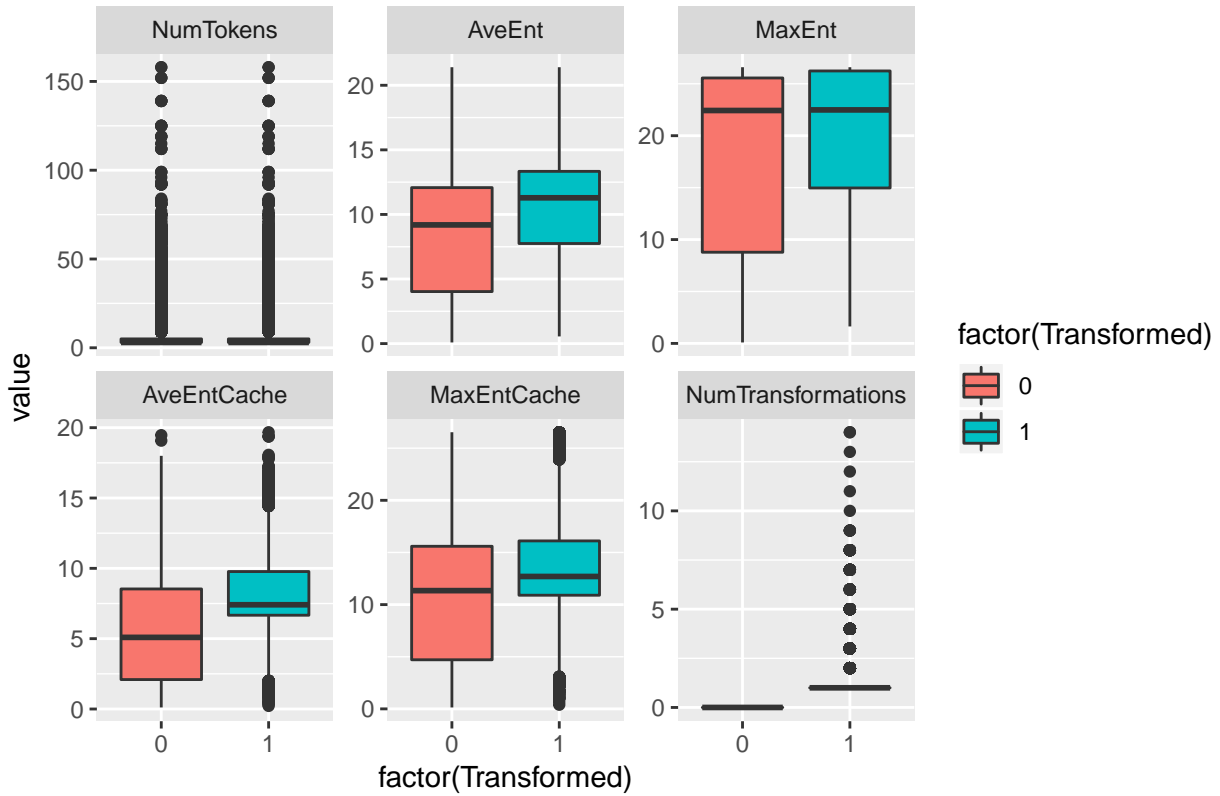
Regular



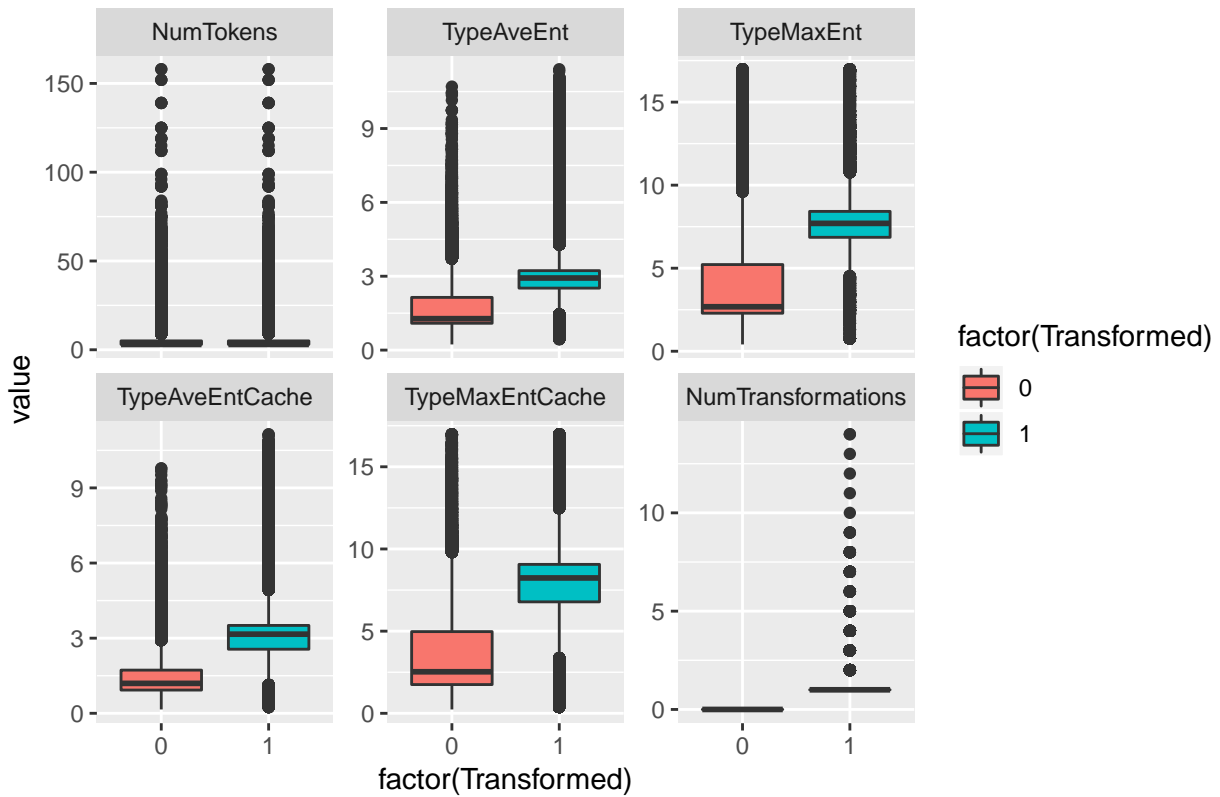
Type



Expression (Regular)



Expression (Type)

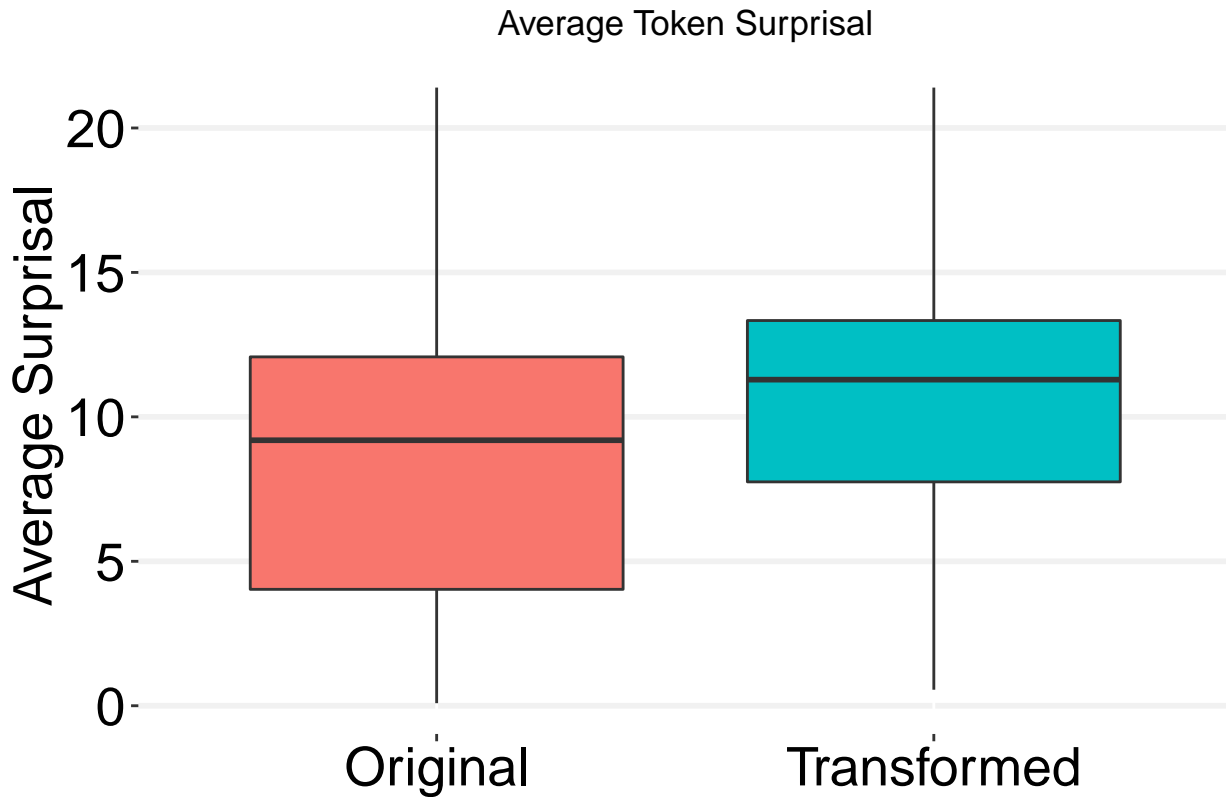


[1] " ----- Expression Global Model ----- "


```

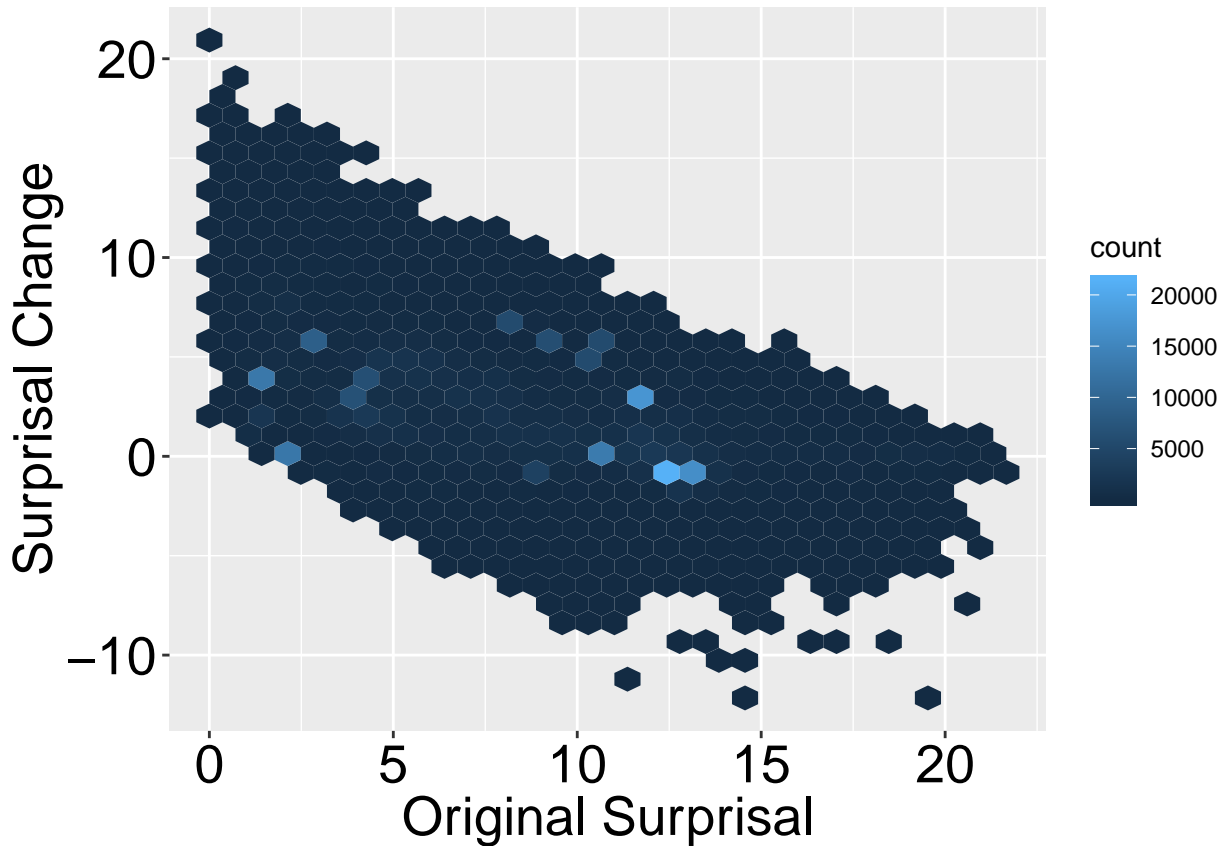
## [1] "LogicalSwapTopGlobalExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -403.71, df = 234810, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.17643
## sample estimates:
## mean of the differences
##      -2.192126
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -403.71, df = 234810, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -2.208969 -2.175282
## sample estimates:
## mean of the differences
##      -2.192126
##
## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -0.8331161 (large)
## 95 percent confidence interval:
## inf sup
## NA NA
##
## Warning in n1 * n2: NAs produced by integer overflow
##
## Warning in n1 * n2: NAs produced by integer overflow

```



```
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 3828400000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.10744
## sample estimates:
## (pseudo)median
##      -2.125434
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 3828400000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -2.151232 -2.106407
## sample estimates:
## (pseudo)median
##      -2.125434
##
##
## Cliff's Delta
##
## delta estimate: -0.2645336 (small)
```

```
## 95 percent confidence interval:
##      inf      sup
## -0.2676890 -0.2613726
```



```
## [1] "----- Expression Cache Model -----"
## [1] "LogicalSwapTopCacheExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -470.92, df = 234810, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.600542
## sample estimates:
## mean of the differences
##      -2.616603
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -470.92, df = 234810, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -2.633838 -2.599367
## sample estimates:
```

```

## mean of the differences
##          -2.616603

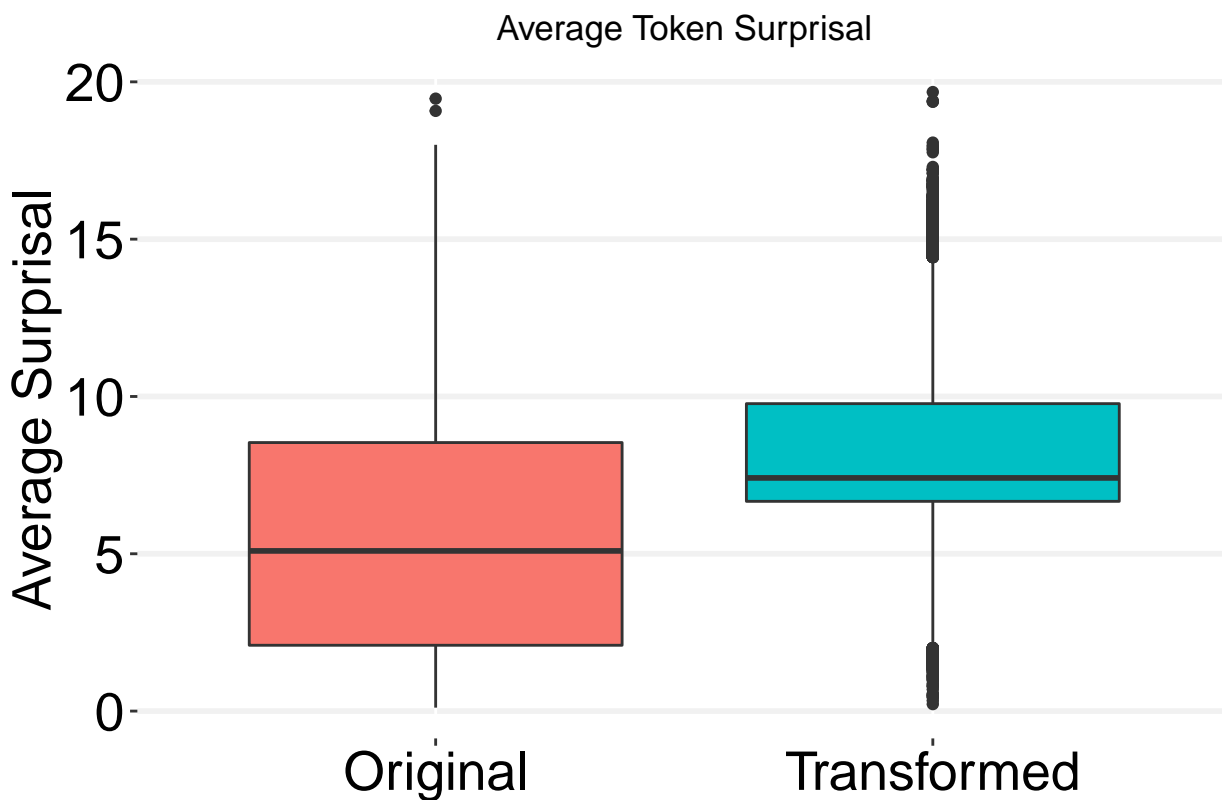
## Warning in n1 * n2: NAs produced by integer overflow

##
## Cohen's d
##
## d estimate: -0.971823 (large)
## 95 percent confidence interval:
## inf sup
## NA NA

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

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

```

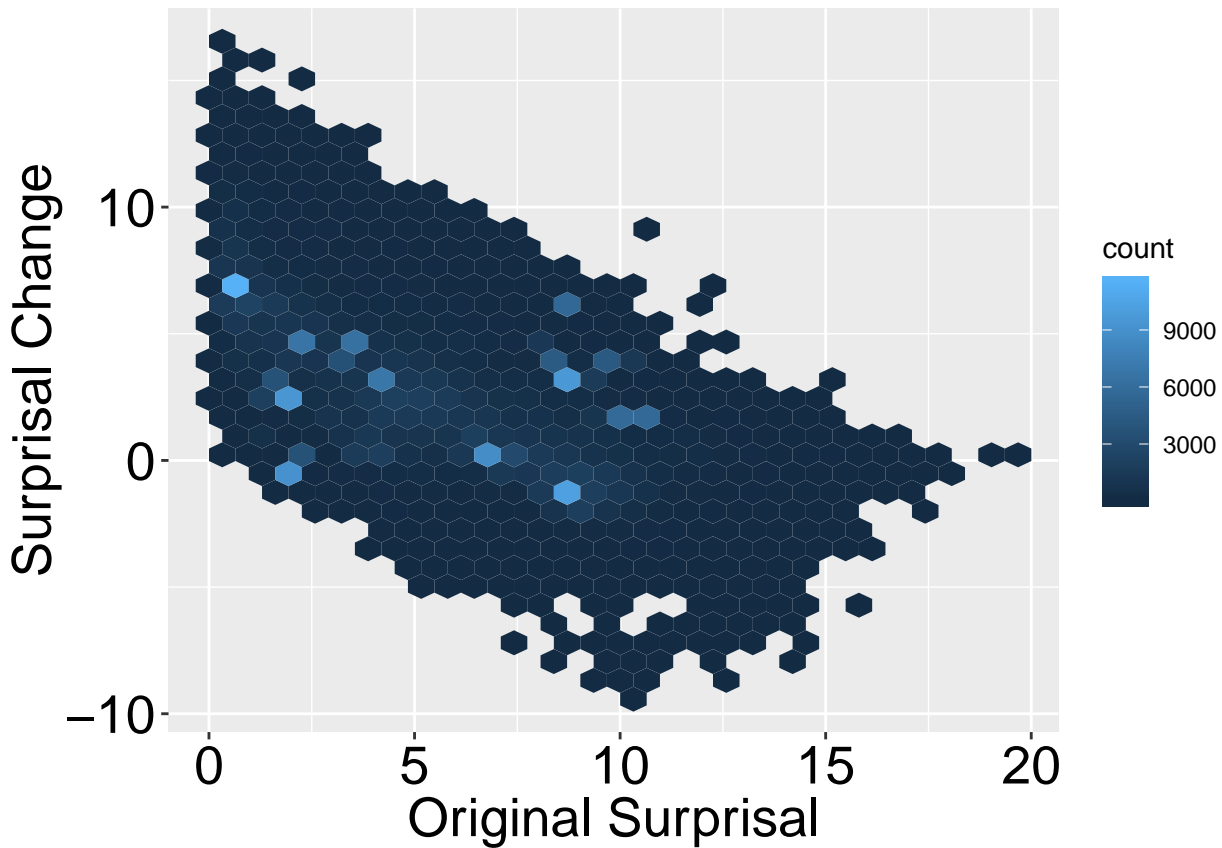


```

##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 2221900000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.522148
## sample estimates:
## (pseudo)median
##      -2.536013
##
##

```

```
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 2221900000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
## -2.556848 -2.520510
## sample estimates:
## (pseudo)median
## -2.536013
##
##
## Cliff's Delta
##
## delta estimate: -0.3816724 (medium)
## 95 percent confidence interval:
## inf sup
## -0.3847078 -0.3786287
```



```
## [1] "----- Expression Global Type Model ----- "
## [1] "LogicalSwapTopGlobalTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -660.38, df = 234810, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
```

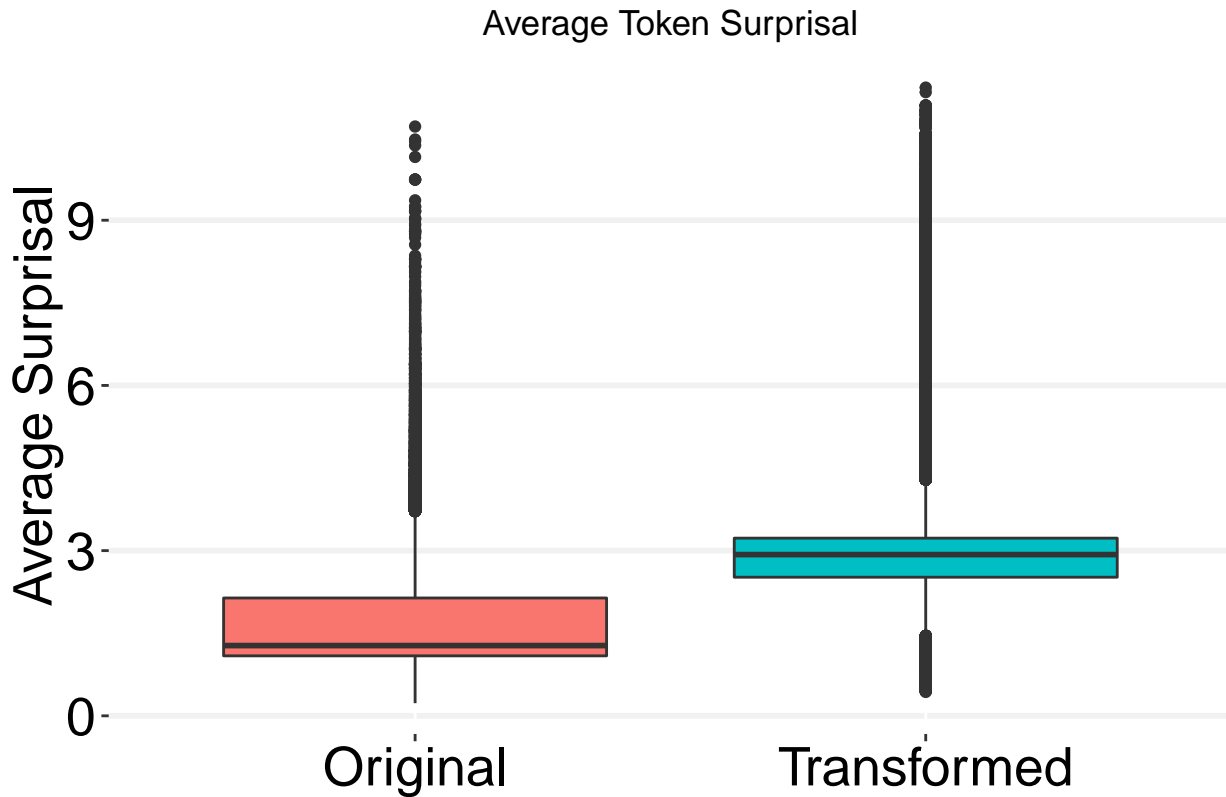
```
## 99.80769 percent confidence interval:
##      -Inf -1.385079
## sample estimates:
## mean of the differences
##      -1.391169
##
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -660.38, df = 234810, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -1.397703 -1.384634
## sample estimates:
## mean of the differences
##      -1.391169

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

##
## Cohen's d
##
## d estimate: -1.362813 (large)
## 95 percent confidence interval:
## inf sup
## NA NA

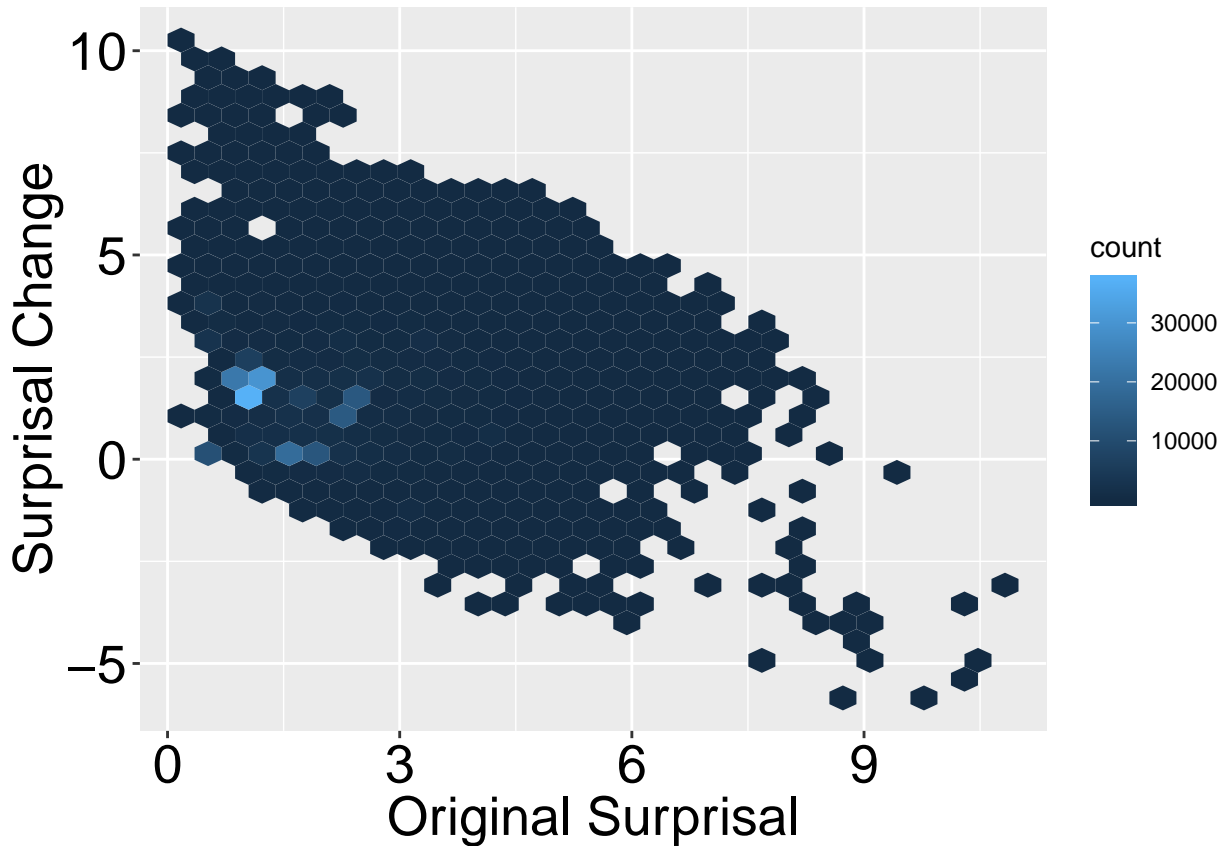
## Warning in n1 * n2: NAs produced by integer overflow

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



```
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 114320000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.755948
## sample estimates:
## (pseudo)median
##      -1.756701
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 114320000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.757989 -1.755985
## sample estimates:
## (pseudo)median
##      -1.756701
##
##
## Cliff's Delta
##
## delta estimate: -0.6957594 (large)
```

```
## 95 percent confidence interval:
##      inf      sup
## -0.6981173 -0.6933864
```



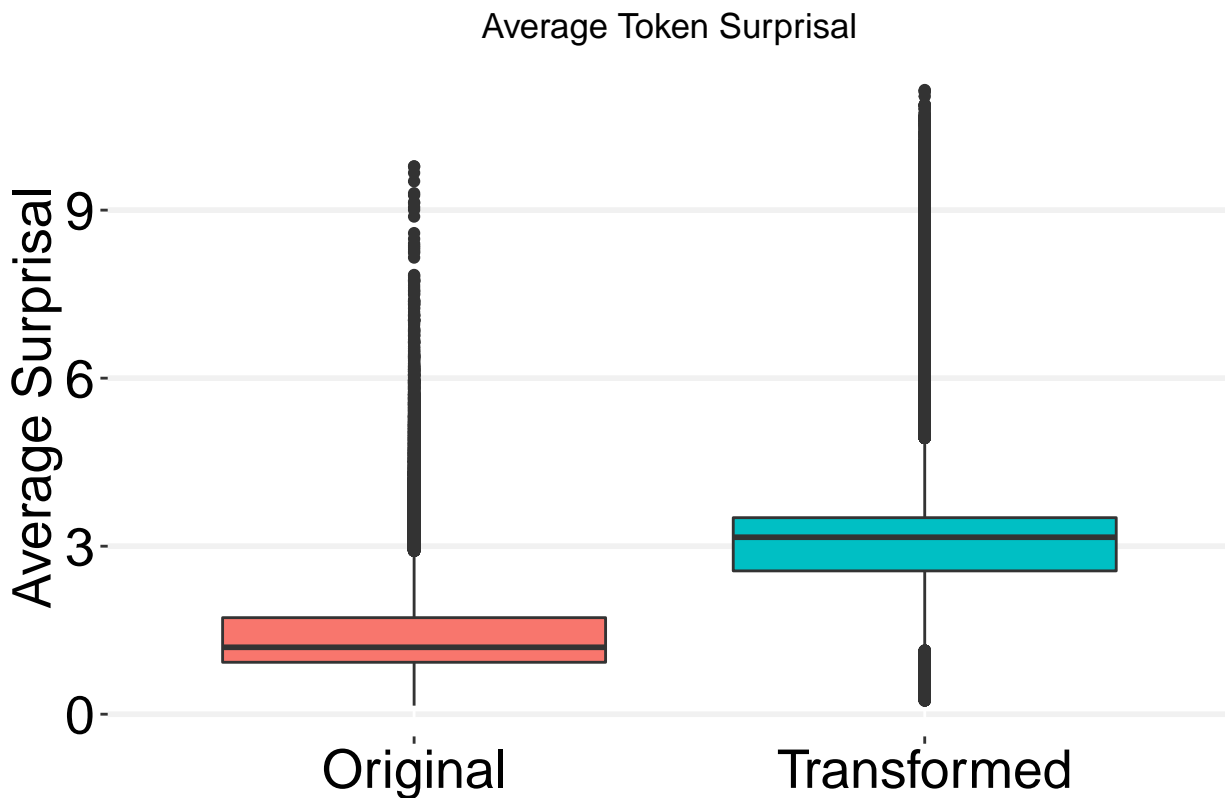
```
## [1] "----- Expression Cache Type Model ----- "
## [1] "LogicalSwapTopCacheTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -657.28, df = 234810, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.626218
## sample estimates:
## mean of the differences
##      -1.633401
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -657.28, df = 234810, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -1.641110 -1.625693
## sample estimates:
```



```

## mean of the differences
##           -1.633401
## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -1.356412 (large)
## 95 percent confidence interval:
## inf sup
## NA NA
## Warning in n1 * n2: NAs produced by integer overflow
## Warning in n1 * n2: NAs produced by integer overflow

```



```

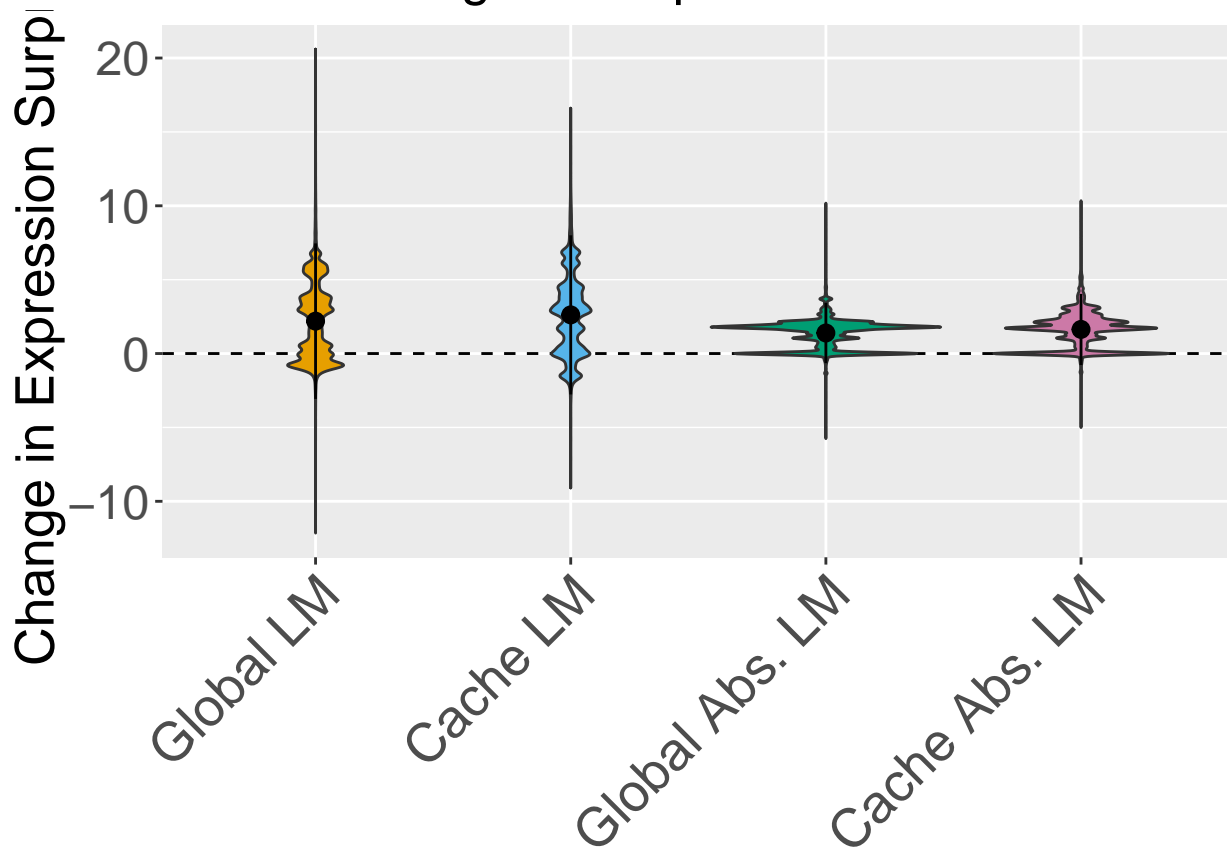
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 79875000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.95102
## sample estimates:
## (pseudo)median
##      -1.957789
##
##

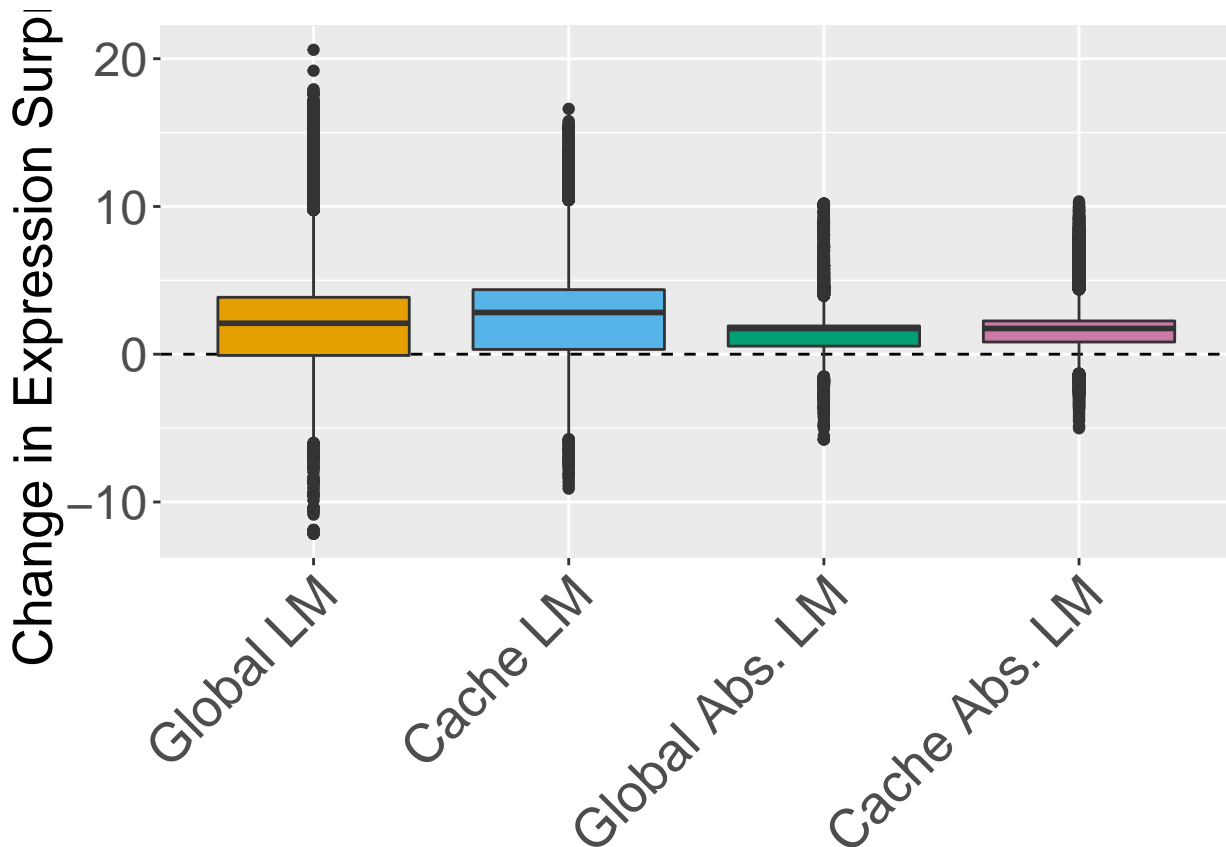
```

```

## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 79875000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
## -1.964316 -1.950562
## sample estimates:
## (pseudo)median
## -1.957789
##
##
## Cliff's Delta
##
## delta estimate: -0.711537 (large)
## 95 percent confidence interval:
## inf sup
## -0.7138471 -0.7092113
## [1] "Binary differences"
##
## FALSE TRUE
## 69400 165412
##
## FALSE TRUE
## 47817 186995
##
## FALSE TRUE
## 4943 229869
##
## FALSE TRUE
## 4068 230744
## No id variables; using all as measure variables
## Warning: Ignoring unknown parameters: mult

```





```

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?
dlspFiltered <- generateFilteredResults(dlsp, "Logic", "LogicSwapTopFiltered100", 100)

## [1] " ----- Expression Global Model ----- "
## [1] "LogicSwapTopFiltered100GlobalExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -208.58, df = 133840, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.441556
## sample estimates:
## mean of the differences
##      -1.461814
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt

```

```

## t = -208.58, df = 133840, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -1.483554 -1.440075
## sample estimates:
## mean of the differences
## -1.461814

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

##
## Cohen's d
##
## d estimate: -0.5701241 (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 = 2064900000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
## -Inf -1.169922
## sample estimates:
## (pseudo)median
## -1.192052
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 2064900000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
## -1.216525 -1.168637
## sample estimates:
## (pseudo)median
## -1.192052
##
##
## Cliff's Delta
##
## delta estimate: -0.1613614 (small)
## 95 percent confidence interval:
## inf sup
## -0.1657736 -0.1569427

## [1] " ----- Expression Cache Model ----- "
## [1] "LogicSwapTopFiltered100CacheExp Original < Transformed"

```

```

##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -258.69, df = 133840, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -2.099206
## sample estimates:
## mean of the differences
##      -2.122927
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -258.69, df = 133840, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -2.148383 -2.097471
## sample estimates:
## mean of the differences
##      -2.122927

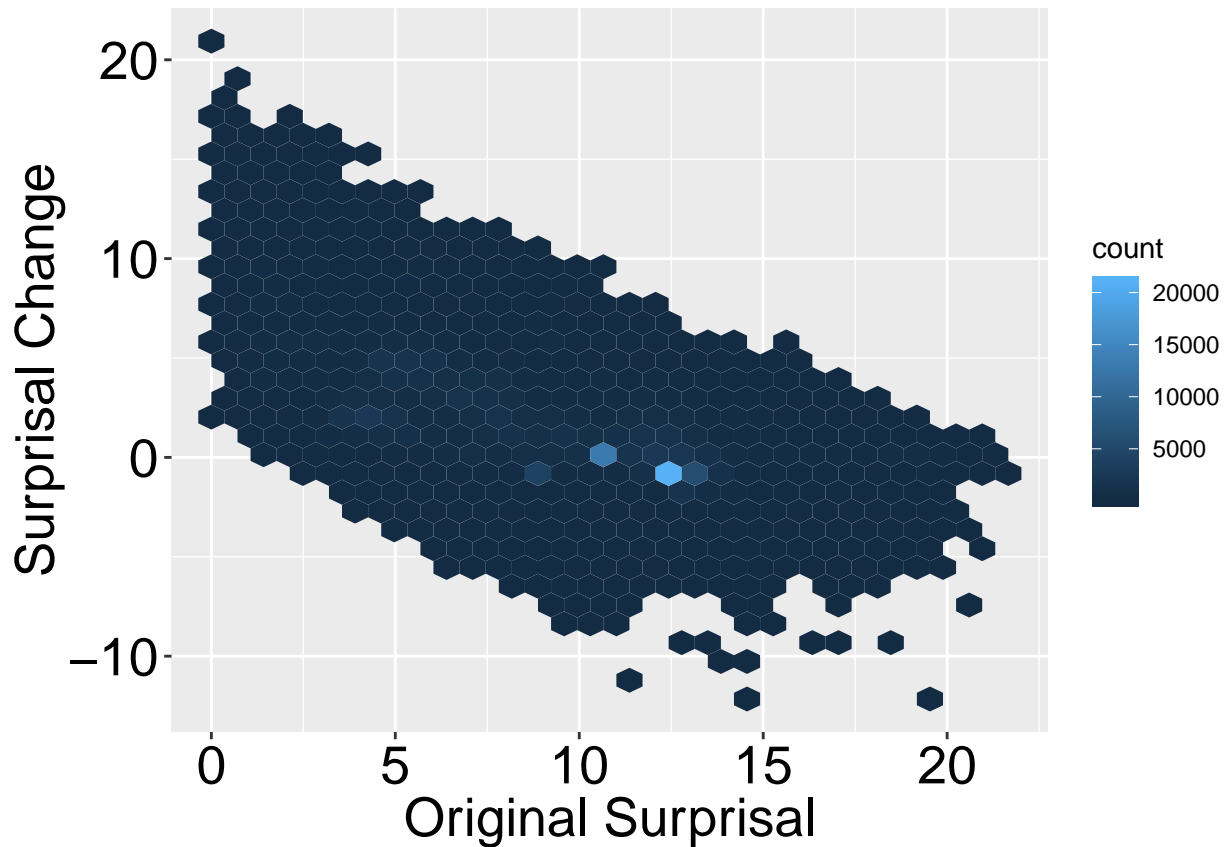
## Warning in n1 * n2: NAs produced by integer overflow

##
## Cohen's d
##
## d estimate: -0.7070878 (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 = 1470100000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.907662
## sample estimates:
## (pseudo)median
##      -1.934393
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1470100000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.963284 -1.905724
## sample estimates:
## (pseudo)median
##      -1.934393
##
##
## Cliff's Delta
##
```

```

## delta estimate: -0.3914333 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.3954837 -0.3873678

## [1] "----- Expression Global Type Model ----- "
## [1] "LogicSwapTopFiltered100GlobalTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -584.45, df = 133840, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.622831
## sample estimates:
## mean of the differences
##      -1.630897
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -584.45, df = 133840, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##      -1.639553 -1.622241
## sample estimates:
## mean of the differences
##      -1.630897

## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -1.597516 (large)
## 95 percent confidence interval:
## inf sup
## NA NA

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

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

```




```
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 82953000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.701947
## sample estimates:
## (pseudo)median
##      -1.711582
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 82953000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.720089 -1.701265
## sample estimates:
## (pseudo)median
##      -1.711582
##
##
## Cliff's Delta
##
```

```

## delta estimate: -0.7687824 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.7714124 -0.7661262

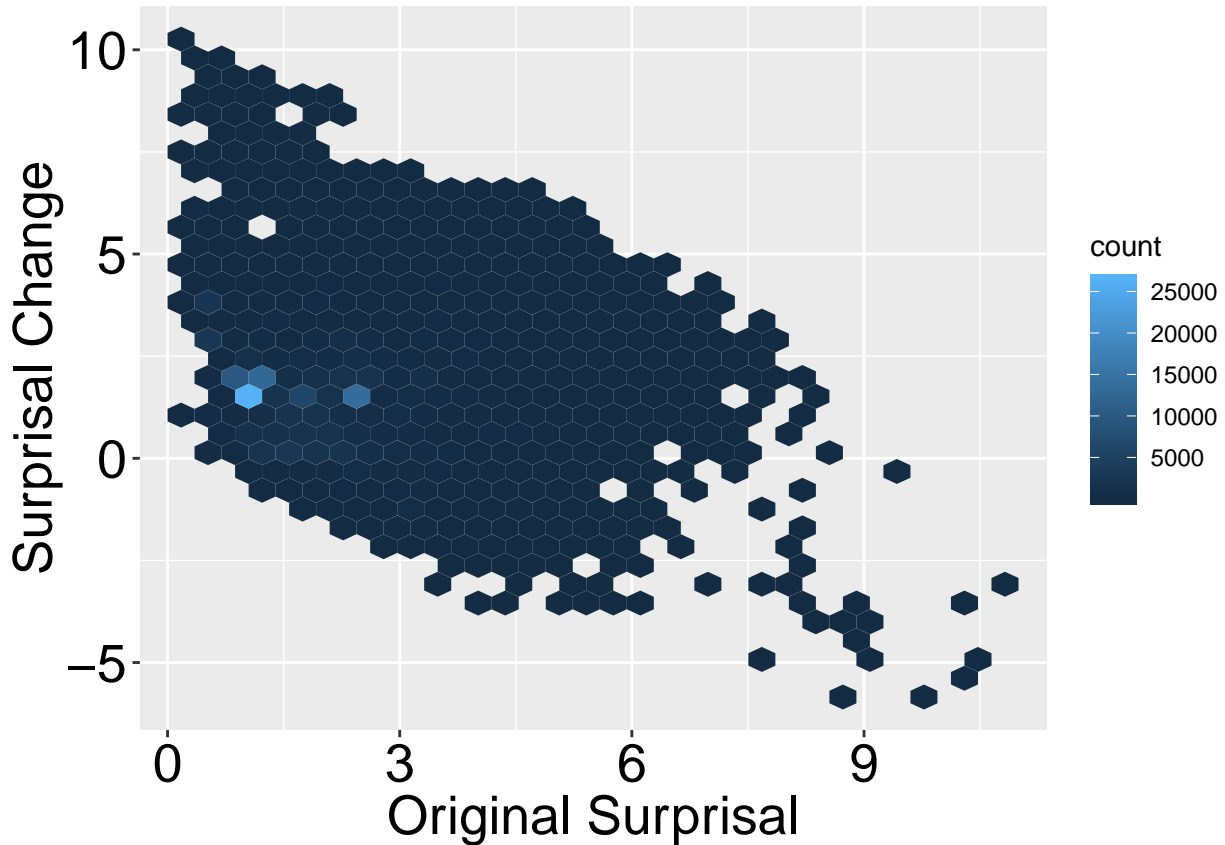
## [1] "----- Expression Cache Type Model ----- "
## [1] "LogicSwapTopFiltered100CacheTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -603.45, df = 133840, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.903285
## sample estimates:
## mean of the differences
##           -1.912445
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -603.45, df = 133840, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##      -1.922276 -1.902615
## sample estimates:
## mean of the differences
##           -1.912445

## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -1.649446 (large)
## 95 percent confidence interval:
## inf sup
## NA NA

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

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

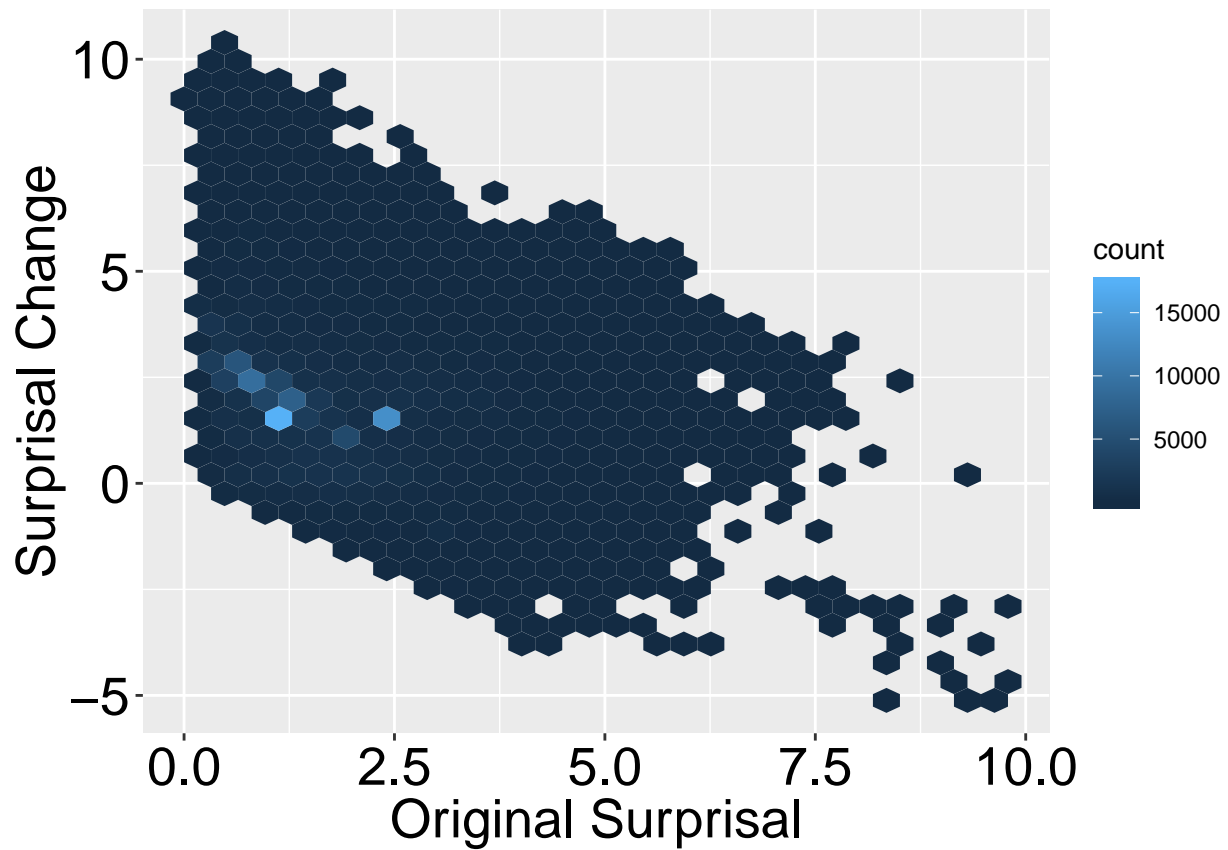
```

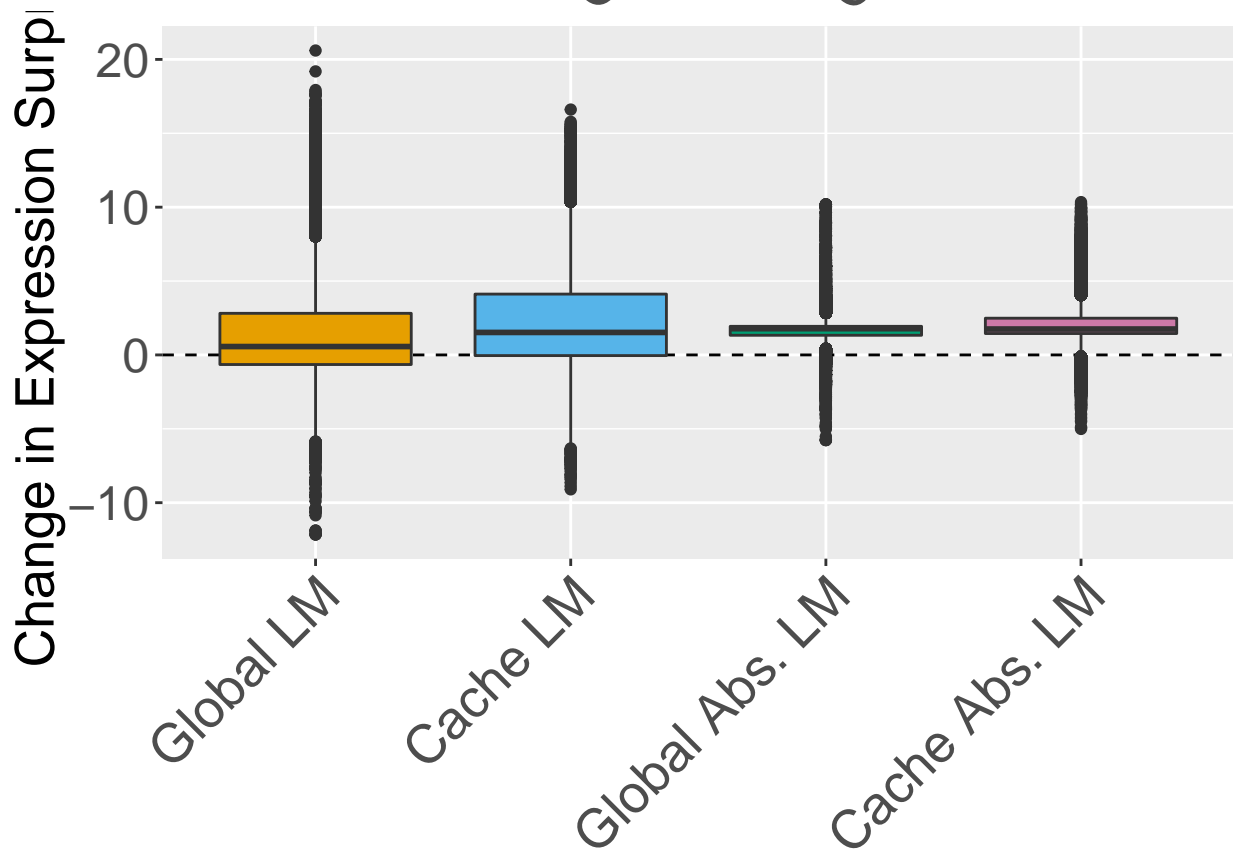
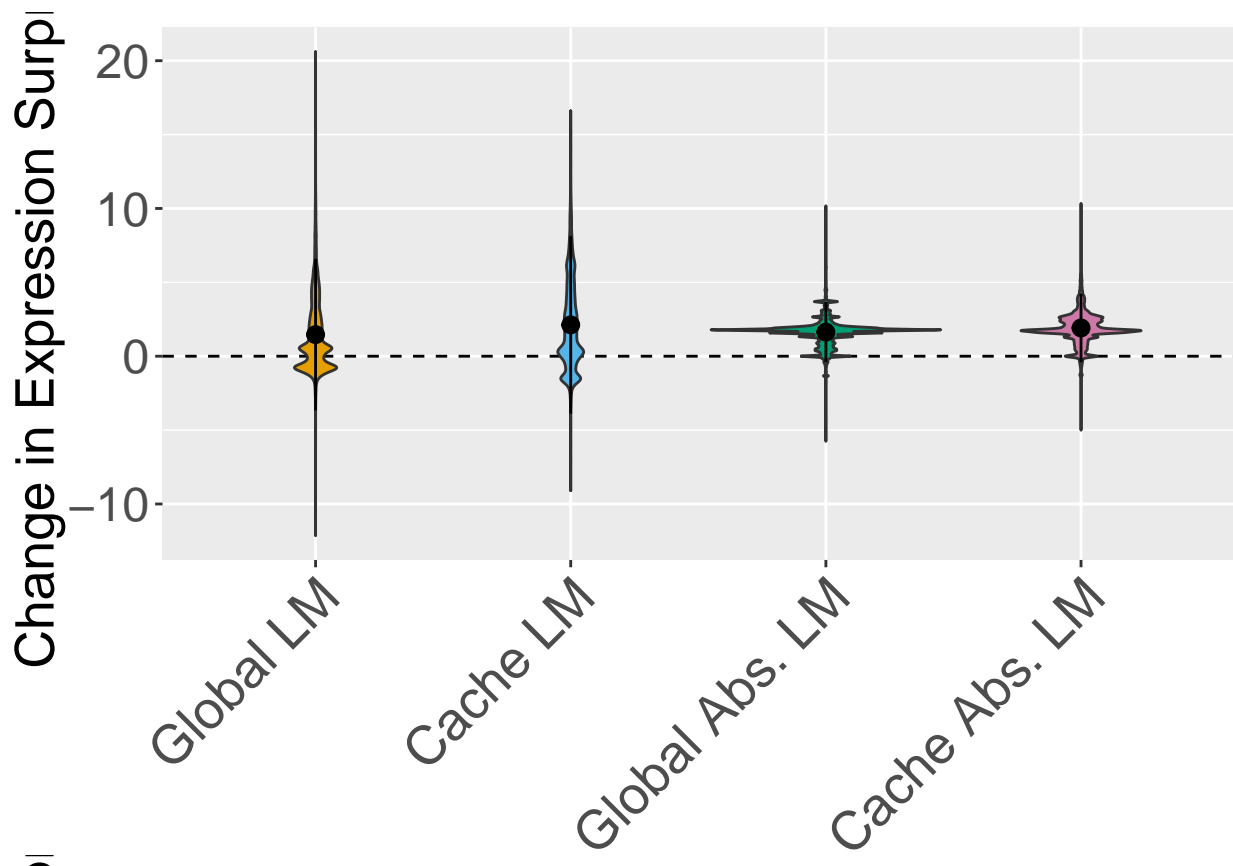


```
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 56815000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.936834
## sample estimates:
## (pseudo)median
##      -1.944973
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 56815000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.953744 -1.936244
## sample estimates:
## (pseudo)median
##      -1.944973
##
##
## Cliff's Delta
##
```

```
## delta estimate: -0.8255308 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.8278032 -0.8232313

## No id variables; using all as measure variables
## Warning: Ignoring unknown parameters: mult
```





```
dlsplFiltered2 <- generateFilteredResults(dlspl, "Logic", "LogicSwapTopFiltered10", 10)
```

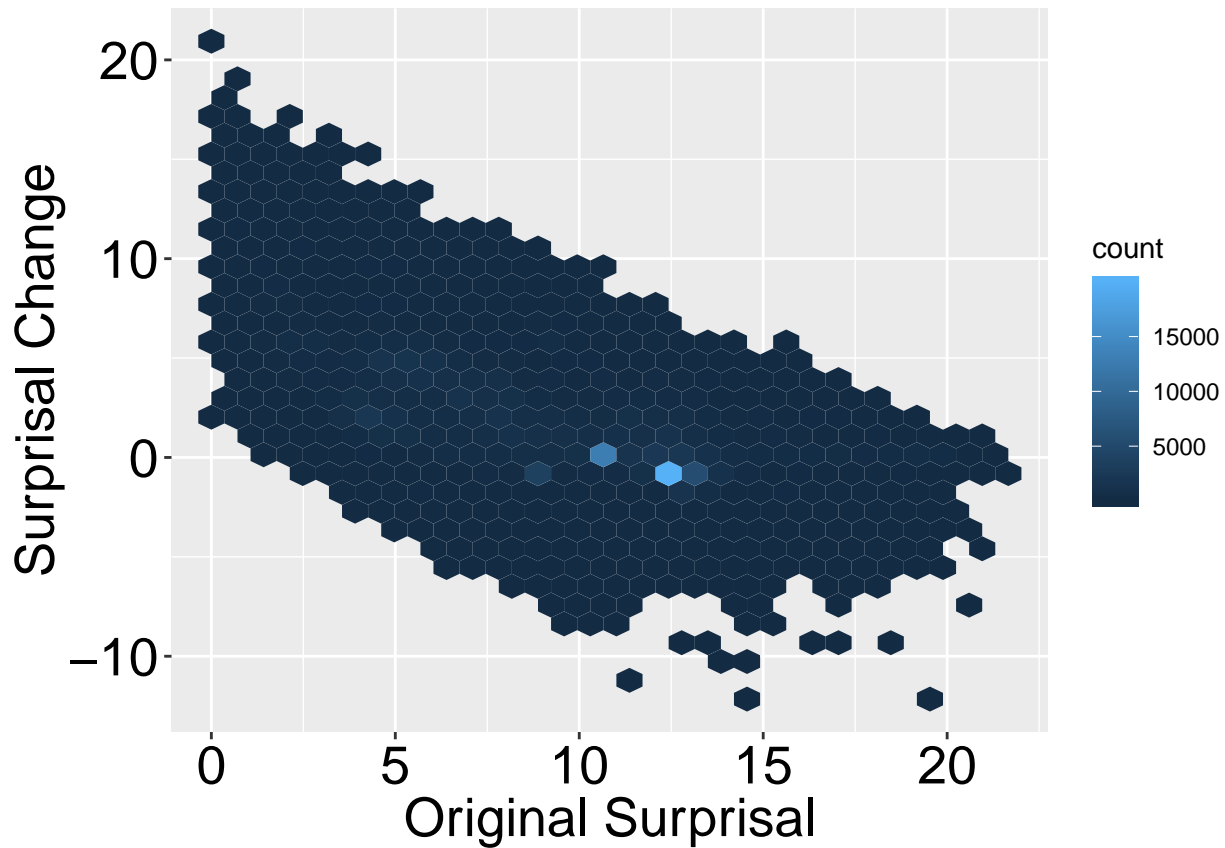
```
## [1] "----- Expression Global Model ----- "  
## [1] "LogicSwapTopFiltered10GlobalExp Original < Transformed"  
##  
## Paired t-test  
##  
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt  
## t = -190.35, df = 121800, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is less than 0  
## 99.80769 percent confidence interval:  
##      -Inf -1.358512  
## sample estimates:  
## mean of the differences  
##      -1.37946  
##  
## Paired t-test  
##  
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt  
## t = -190.35, df = 121800, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 99.80769 percent confidence interval:  
## -1.401939 -1.356981  
## sample estimates:  
## mean of the differences  
##      -1.37946  
  
## Warning in n1 * n2: NAs produced by integer overflow  
##  
## Cohen's d  
##  
## d estimate: -0.5454331 (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 = 1827900000, p-value < 2.2e-16  
## alternative hypothesis: true location shift is less than 0  
## 99.80769 percent confidence interval:  
##      -Inf -1.073024  
## sample estimates:  
## (pseudo)median  
##      -1.096647  
##  
##  
## Wilcoxon signed rank test with continuity correction
```

```

##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 1827900000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
## -1.118077 -1.071185
## sample estimates:
## (pseudo)median
## -1.096647
##
##
## Cliff's Delta
##
## delta estimate: -0.1520483 (small)
## 95 percent confidence interval:
## inf sup
## -0.1566868 -0.1474030
##
## [1] "----- Expression Cache Model ----- "
## [1] "LogicSwapTopFiltered10CacheExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -233.76, df = 121800, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
## -Inf -1.972765
## sample estimates:
## mean of the differences
## -1.997465
##
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -233.76, df = 121800, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
## -2.023972 -1.970959
## sample estimates:
## mean of the differences
## -1.997465
##
## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -0.6697956 (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 = 1332200000, p-value < 2.2e-16  
## alternative hypothesis: true location shift is less than 0  
## 99.80769 percent confidence interval:  
##      -Inf -1.749935  
## sample estimates:  
## (pseudo)median  
##      -1.777505  
##  
##  
## Wilcoxon signed rank test with continuity correction  
##  
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt  
## V = 1332200000, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0  
## 99.80769 percent confidence interval:  
##      -1.807315 -1.747941  
## sample estimates:  
## (pseudo)median  
##      -1.777505  
##  
##
```



```

## Cliff's Delta
##
## delta estimate: -0.3742486 (medium)
## 95 percent confidence interval:
##      inf      sup
## -0.3785234 -0.3699579

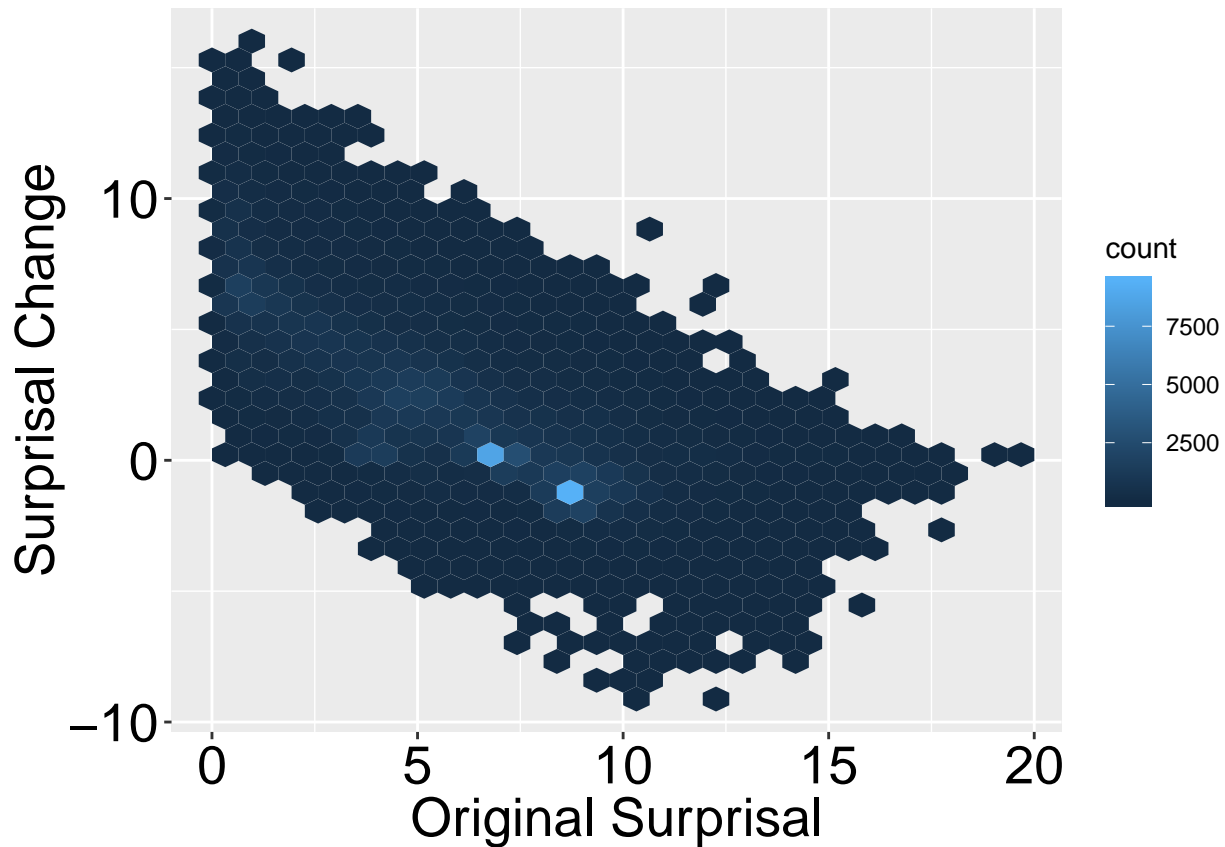
## [1] " ----- Expression Global Type Model ----- "
## [1] "LogicSwapTopFiltered10GlobalTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -552.19, df = 121800, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.62047
## sample estimates:
## mean of the differences
##              -1.628998
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -552.19, df = 121800, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##      -1.638149 -1.619847
## sample estimates:
## mean of the differences
##              -1.628998

## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -1.582231 (large)
## 95 percent confidence interval:
## inf sup
## NA NA

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

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

```



```
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 70058000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.695616
## sample estimates:
## (pseudo)median
##      -1.703641
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 70058000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.713673 -1.695522
## sample estimates:
## (pseudo)median
##      -1.703641
##
##
## Cliff's Delta
##
```

```

## delta estimate: -0.765483 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.7682490 -0.7626885

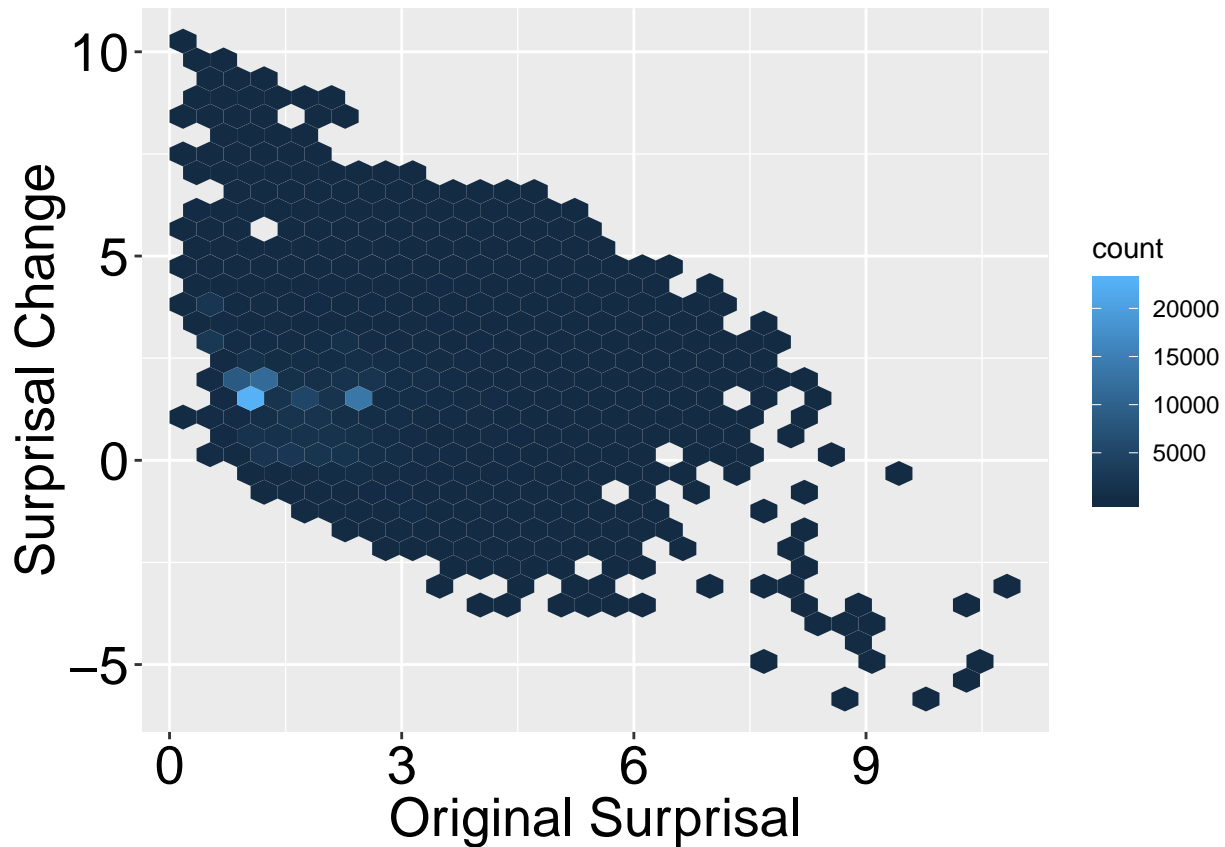
## [1] "----- Expression Cache Type Model ----- "
## [1] "LogicSwapTopFiltered10CacheTypeExp Original < Transformed"
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -571.49, df = 121800, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.885345
## sample estimates:
## mean of the differences
##      -1.894929
##
## Paired t-test
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## t = -571.49, df = 121800, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99.80769 percent confidence interval:
##      -1.905215 -1.884644
## sample estimates:
## mean of the differences
##      -1.894929

## Warning in n1 * n2: NAs produced by integer overflow
##
## Cohen's d
##
## d estimate: -1.637514 (large)
## 95 percent confidence interval:
## inf sup
## NA NA

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

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

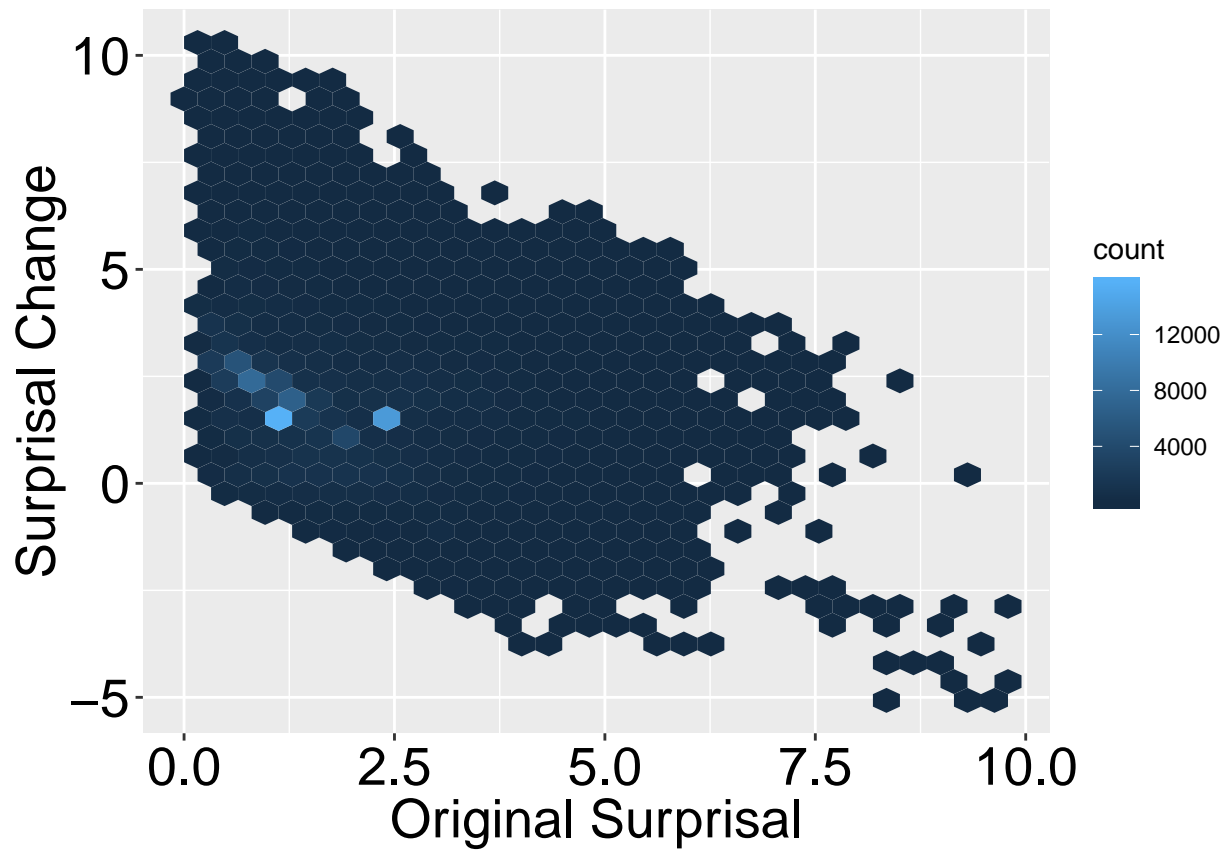
```

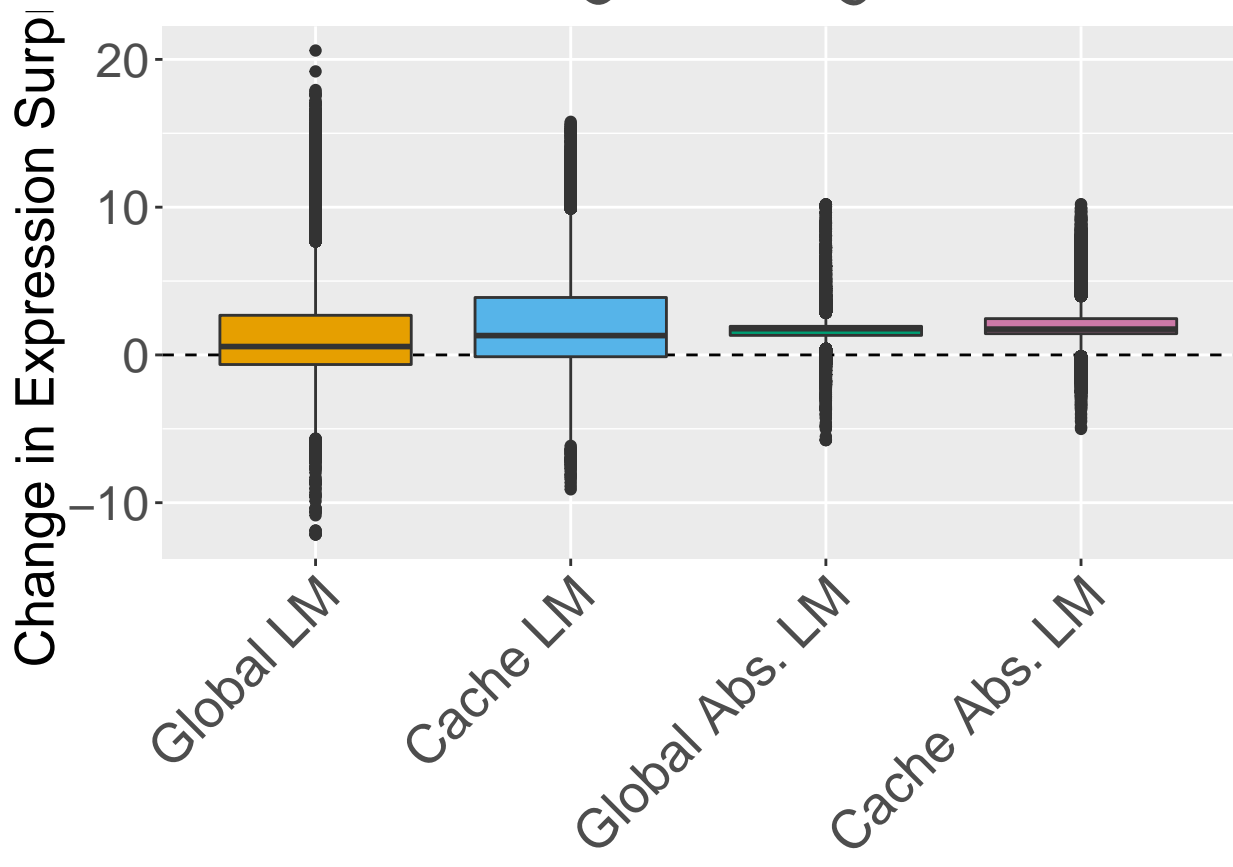
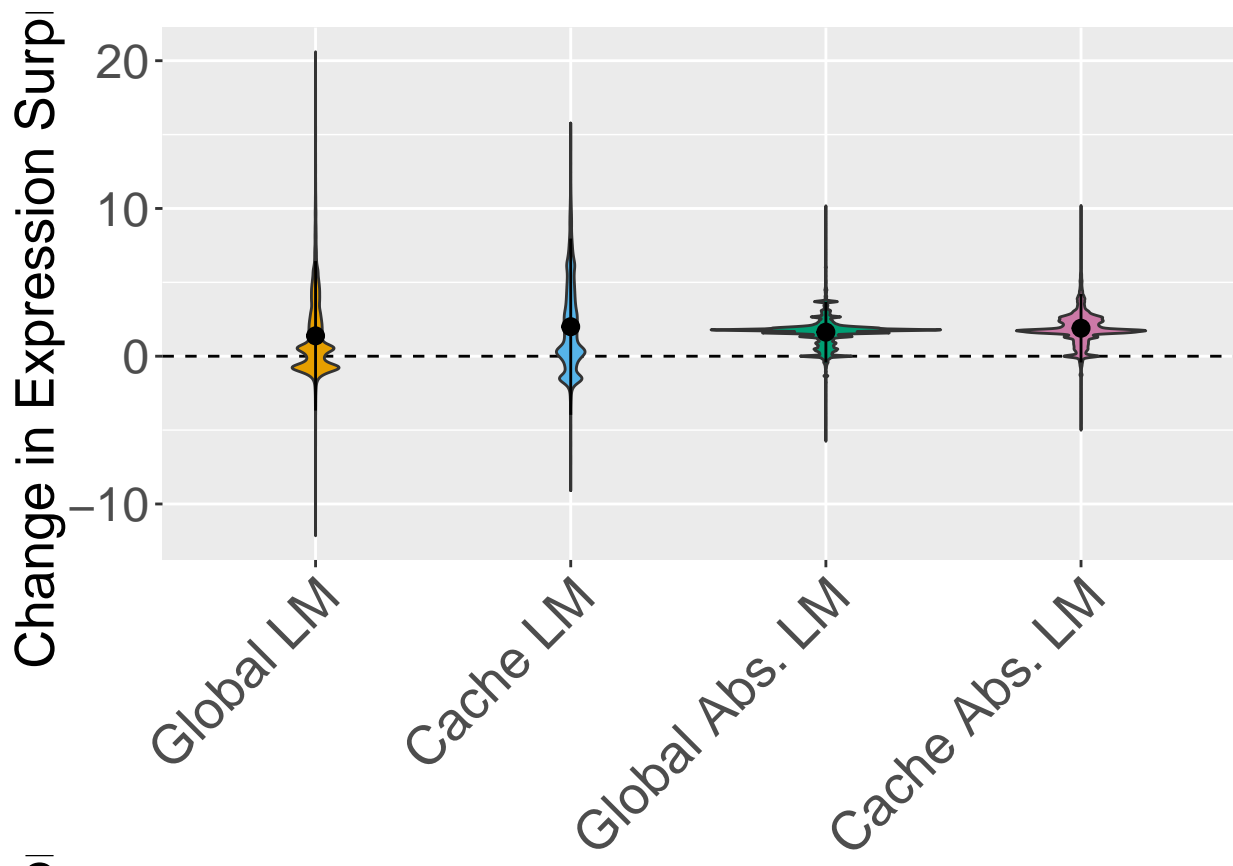


```
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 46276000, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
## 99.80769 percent confidence interval:
##      -Inf -1.913872
## sample estimates:
## (pseudo)median
##      -1.922376
##
##
## Wilcoxon signed rank test with continuity correction
##
## data: diffClean$BaseAveEnt and diffClean$ChangeAveEnt
## V = 46276000, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
## 99.80769 percent confidence interval:
##      -1.931525 -1.913245
## sample estimates:
## (pseudo)median
##      -1.922376
##
##
## Cliff's Delta
##
```

```
## delta estimate: -0.8193201 (large)
## 95 percent confidence interval:
##      inf      sup
## -0.8217430 -0.8168676

## No id variables; using all as measure variables
## Warning: Ignoring unknown parameters: mult
```





Regression models

```
m_lswap_no_out <- modelGlobal(dlspFiltered, "", "==" )
```

```
##
## Call:
## lm(formula = AverageEntChangeExp ~ BaseAveEntExp + log(NumTokens) +
##     factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.2686 -0.7721 -0.3164  0.6680  8.5360
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)      7.590573   0.053260 142.520
## BaseAveEntExp    -0.415319   0.001143 -363.339
## log(NumTokens)   -0.699800   0.017833  -39.242
## factor(ParentOp)&&
## factor(ParentOp)AssertStatement  -1.810392   0.037268 -48.577
## factor(ParentOp)Assignment        -0.552046   0.260716  -2.117
## factor(ParentOp)ConditionalExpression  0.190601   0.026642   7.154
## factor(ParentOp)ForStatement        0.390794   0.034571  11.304
## factor(ParentOp)IfStatement        -1.283166   0.022088 -58.094
## factor(ParentOp)MethodInvocation    -0.662224   0.045248 -14.635
## factor(ParentOp)ParenthesizedExpression -0.554134   0.036008 -15.389
## factor(ParentOp)ReturnStatement     0.162947   0.053384   3.052
## factor(ParentOp)VariableDeclarationFragment 0.024149   0.219209   0.110
## factor(ParentOp)WhileStatement     -1.387575   0.083235 -16.671
## factor(MostFreqOp)<                1.587466   0.025180  63.044
## factor(MostFreqOp)<=                2.388189   0.052463  45.521
## factor(MostFreqOp)>                 0.849911   0.023653  35.933
## factor(MostFreqOp)>=                0.429873   0.036743  11.699
## factor(MostFreqOp)!=                0.095593   0.010371   9.217
##
##              Pr(>|t|)
## (Intercept)      < 2e-16 ***
## BaseAveEntExp    < 2e-16 ***
## log(NumTokens)   < 2e-16 ***
## factor(ParentOp)&&
## factor(ParentOp)AssertStatement  < 2e-16 ***
## factor(ParentOp)Assignment        0.03423 *
## factor(ParentOp)ConditionalExpression 8.46e-13 ***
## factor(ParentOp)ForStatement        < 2e-16 ***
## factor(ParentOp)IfStatement        < 2e-16 ***
## factor(ParentOp)MethodInvocation    < 2e-16 ***
## factor(ParentOp)ParenthesizedExpression < 2e-16 ***
## factor(ParentOp)ReturnStatement     0.00227 **
## factor(ParentOp)VariableDeclarationFragment 0.91228
## factor(ParentOp)WhileStatement     < 2e-16 ***
## factor(MostFreqOp)<                < 2e-16 ***
## factor(MostFreqOp)<=                < 2e-16 ***
## factor(MostFreqOp)>                 < 2e-16 ***
## factor(MostFreqOp)>=                < 2e-16 ***
## factor(MostFreqOp)!=                < 2e-16 ***
## ---
```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.448 on 117235 degrees of freedom
## Multiple R-squared:  0.6107, Adjusted R-squared:  0.6106
## F-statistic: 1.021e+04 on 18 and 117235 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: AverageEntChangeExp
##
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## BaseAveEntExp      1 320740   320740 153076.8 < 2.2e-16 ***
## log(NumTokens)     1   9614    9614   4588.3 < 2.2e-16 ***
## factor(ParentOp)   11  40944    3722   1776.4 < 2.2e-16 ***
## factor(MostFreqOp)  5  13963    2793   1332.8 < 2.2e-16 ***
## Residuals         117235 245641         2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.50838358" "0.01523815" "0.06489744" "0.02213141" "0.38934942"
##           GVIF Df GVIF^(1/(2*Df))
## BaseAveEntExp      1.070754  1         1.034773
## log(NumTokens)     3.053357  1         1.747386
## factor(ParentOp)   8.354532 11         1.101300
## factor(MostFreqOp) 3.033773  5         1.117373
##
## % 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 - 10:16:31 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lc}
##     \hline
##     & \multicolumn{1}{c}{\textit{Dependent variable:}} & \\
##     \hline
##     & AverageEntChangeExp & \\
##     \hline
##     BaseAveEntExp &  $-\$0.415^{\{***\}}\$$  (0.001) & \\
##     log(NumTokens) &  $-\$0.700^{\{***\}}\$$  (0.018) & \\
##     factor(ParentOp)&& &  $-\$0.364^{\{***\}}\$$  (0.023) & \\
##     factor(ParentOp)AssertStatement &  $-\$1.810^{\{***\}}\$$  (0.037) & \\
##     factor(ParentOp)Assignment &  $-\$0.552^{\{**\}}\$$  (0.261) & \\
##     factor(ParentOp)ConditionalExpression &  $0.191^{\{***\}}\$$  (0.027) & \\
##     factor(ParentOp)ForStatement &  $0.391^{\{***\}}\$$  (0.035) & \\
##     factor(ParentOp)IfStatement &  $-\$1.283^{\{***\}}\$$  (0.022) & \\
##     factor(ParentOp)MethodInvocation &  $-\$0.662^{\{***\}}\$$  (0.045) & \\
##     factor(ParentOp)ParenthesizedExpression &  $-\$0.554^{\{***\}}\$$  (0.036) & \\
##     factor(ParentOp)ReturnStatement &  $0.163^{\{***\}}\$$  (0.053) & \\
##     factor(ParentOp)VariableDeclarationFragment &  $0.024$  (0.219) & \\
##     factor(ParentOp)WhileStatement &  $-\$1.388^{\{***\}}\$$  (0.083) & \\
##     factor(MostFreqOp)\textless &  $1.587^{\{***\}}\$$  (0.025) & \\
##     factor(MostFreqOp)\textless = &  $2.388^{\{***\}}\$$  (0.052) & \\
##     factor(MostFreqOp)\textgreater &  $0.850^{\{***\}}\$$  (0.024) & \\
##     factor(MostFreqOp)\textgreater = &  $0.430^{\{***\}}\$$  (0.037) & \\
##     factor(MostFreqOp)!= &  $0.096^{\{***\}}\$$  (0.010) & \\

```



```

## Constant & 7.591*** (0.053) \
## \hline \[-1.8ex]
## Observations & 117,254 \
## R2 & 0.611 \
## Adjusted R2 & 0.611 \
## Residual Std. Error & 1.448 (df = 117235) \
## F Statistic & 10,214.990*** (df = 18; 117235) \
## \hline
## \hline \[-1.8ex]
## \textit{Note:} & \multicolumn{1}{r}{*p<$0.1; **p<$0.05; ***p<$0.01} \
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:16:32 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
## \hline
## & Df & Sum Sq & Mean Sq & F value & Pr(>$F) \
## \hline
## BaseAveEntExp & 1 & 320740.31 & 320740.31 & 153076.76 & 0.0000 \
## log(NumTokens) & 1 & 9613.78 & 9613.78 & 4588.28 & 0.0000 \
## factor(ParentOp) & 11 & 40943.94 & 3722.18 & 1776.45 & 0.0000 \
## factor(MostFreqOp) & 5 & 13962.75 & 2792.55 & 1332.77 & 0.0000 \
## Residuals & 117235 & 245641.40 & 2.10 & & \
## \hline
## \end{tabular}
## \end{table}

```

```
m_lswap_cache_no_out <- modelCache(dlspFiltered, "", "==" )
```

```

##
## Call:
## lm(formula = CacheAverageEntChangeExp ~ BaseCacheAveEntExp +
## log(NumTokens) + factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
## Min 1Q Median 3Q Max
## -6.8792 -0.7863 -0.1567 0.7435 8.0370
##
## Coefficients:
## Estimate Std. Error t value
## (Intercept) 9.966481 0.053890 184.942
## BaseCacheAveEntExp -0.788019 0.001438 -547.901
## log(NumTokens) -1.211262 0.018124 -66.833
## factor(ParentOp)&&
## factor(ParentOp)AssertStatement -1.788833 0.036845 -48.550
## factor(ParentOp)Assignment -0.571288 0.219195 -2.606
## factor(ParentOp)ConditionalExpression 0.725802 0.026570 27.317
## factor(ParentOp)ForStatement -0.489268 0.034512 -14.177
## factor(ParentOp)IfStatement -1.620985 0.022098 -73.354
## factor(ParentOp)MethodInvocation -0.460945 0.046901 -9.828
## factor(ParentOp)ParenthesizedExpression -0.226352 0.037142 -6.094
## factor(ParentOp)ReturnStatement -0.337737 0.056294 -6.000
## factor(ParentOp)VariableDeclarationFragment -0.101067 0.259065 -0.390

```

```

## factor(ParentOp)WhileStatement          -1.555257  0.075137 -20.699
## factor(MostFreqOp)<                     1.989859  0.025691  77.455
## factor(MostFreqOp)<=                    2.861428  0.054252  52.743
## factor(MostFreqOp)>                     1.354407  0.024034  56.354
## factor(MostFreqOp)>=                    1.165863  0.036264  32.149
## factor(MostFreqOp)!=                    0.471423  0.010184  46.291
##                                         Pr(>|t|)
## (Intercept)                            < 2e-16 ***
## BaseCacheAveEntExp                     < 2e-16 ***
## log(NumTokens)                         < 2e-16 ***
## factor(ParentOp)&&                       < 2e-16 ***
## factor(ParentOp)AssertStatement        < 2e-16 ***
## factor(ParentOp)Assignment             0.00915 **
## factor(ParentOp)ConditionalExpression  < 2e-16 ***
## factor(ParentOp)ForStatement           < 2e-16 ***
## factor(ParentOp)IfStatement            < 2e-16 ***
## factor(ParentOp)MethodInvocation       < 2e-16 ***
## factor(ParentOp)ParenthesizedExpression 1.10e-09 ***
## factor(ParentOp)ReturnStatement        1.98e-09 ***
## factor(ParentOp)VariableDeclarationFragment 0.69645
## factor(ParentOp)WhileStatement        < 2e-16 ***
## factor(MostFreqOp)<                   < 2e-16 ***
## factor(MostFreqOp)<=                   < 2e-16 ***
## factor(MostFreqOp)>                   < 2e-16 ***
## factor(MostFreqOp)>=                   < 2e-16 ***
## factor(MostFreqOp)!=                   < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.414 on 115674 degrees of freedom
## Multiple R-squared:  0.7616, Adjusted R-squared:  0.7616
## F-statistic: 2.053e+04 on 18 and 115674 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: CacheAverageEntChangeExp
##              Df Sum Sq Mean Sq F value    Pr(>F)
## BaseCacheAveEntExp  1 664845  664845 332403.9 < 2.2e-16 ***
## log(NumTokens)      1   6646   6646  3322.7 < 2.2e-16 ***
## factor(ParentOp)    11  44029   4003  2001.2 < 2.2e-16 ***
## factor(MostFreqOp)  5  23586   4717  2358.5 < 2.2e-16 ***
## Residuals          115674 231361     2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.685077682" "0.006847953" "0.045368411" "0.024304215" "0.238401740"
##              GVIF Df GVIF^(1/(2*Df))
## BaseCacheAveEntExp 1.073138  1  1.035924
## log(NumTokens)    3.139843  1  1.771960
## factor(ParentOp)  8.914642 11  1.104553
## factor(MostFreqOp) 3.204714  5  1.123515
##
## % 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 - 10:16:50 PM
## \begin{table}[!htbp] \centering

```

```

## \caption{}
## \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \[-1.8ex]\hline
## \hline \[-1.8ex]
## & \multicolumn{1}{c}{\textit{Dependent variable:}} \\\
## \cline{2-2}
## \[-1.8ex] & CacheAverageEntChangeExp \\\
## \hline \[-1.8ex]
## BaseCacheAveEntExp &  $-\$0.788^{***}$  (0.001) \\\
## \log(NumTokens) &  $-\$1.211^{***}$  (0.018) \\\
## \factor(ParentOp)&& &  $-\$0.193^{***}$  (0.023) \\\
## \factor(ParentOp)AssertStatement &  $-\$1.789^{***}$  (0.037) \\\
## \factor(ParentOp)Assignment &  $-\$0.571^{***}$  (0.219) \\\
## \factor(ParentOp)ConditionalExpression &  $0.726^{***}$  (0.027) \\\
## \factor(ParentOp)ForStatement &  $-\$0.489^{***}$  (0.035) \\\
## \factor(ParentOp)IfStatement &  $-\$1.621^{***}$  (0.022) \\\
## \factor(ParentOp)MethodInvocation &  $-\$0.461^{***}$  (0.047) \\\
## \factor(ParentOp)ParenthesizedExpression &  $-\$0.226^{***}$  (0.037) \\\
## \factor(ParentOp)ReturnStatement &  $-\$0.338^{***}$  (0.056) \\\
## \factor(ParentOp)VariableDeclarationFragment &  $-\$0.101$  (0.259) \\\
## \factor(ParentOp)WhileStatement &  $-\$1.555^{***}$  (0.075) \\\
## \factor(MostFreqOp)\textless &  $1.990^{***}$  (0.026) \\\
## \factor(MostFreqOp)\textless = &  $2.861^{***}$  (0.054) \\\
## \factor(MostFreqOp)\textgreater &  $1.354^{***}$  (0.024) \\\
## \factor(MostFreqOp)\textgreater = &  $1.166^{***}$  (0.036) \\\
## \factor(MostFreqOp)!= &  $0.471^{***}$  (0.010) \\\
## \Constant &  $9.966^{***}$  (0.054) \\\
## \hline \[-1.8ex]
## Observations & 115,693 \\\
## R2 & 0.762 \\\
## Adjusted R2 & 0.762 \\\
## Residual Std. Error & 1.414 (df = 115674) \\\
## F Statistic & 20,529.570*** (df = 18; 115674) \\\
## \hline
## \hline \[-1.8ex]
## \textit{Note:} & \multicolumn{1}{r}{ $^{***}p < 0.01$ ;  $^{**}p < 0.05$ ;  $^{*}p < 0.01$ } \\\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:16:51 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
## \hline
## & Df & Sum Sq & Mean Sq & F value & Pr(>F) \\\
## \hline
## BaseCacheAveEntExp & 1 & 664844.91 & 664844.91 & 332403.93 & 0.0000 \\\
## \log(NumTokens) & 1 & 6645.71 & 6645.71 & 3322.67 & 0.0000 \\\
## \factor(ParentOp) & 11 & 44028.52 & 4002.59 & 2001.18 & 0.0000 \\\
## \factor(MostFreqOp) & 5 & 23586.43 & 4717.29 & 2358.51 & 0.0000 \\\
## Residuals & 115674 & 231360.89 & 2.00 & & \\\
## \hline
## \end{tabular}
## \end{table}

```

```

## \end{table}
mt_lswap_no_out <- modelGlobalType(dlspFiltered, "", "==" )

##
## Call:
## lm(formula = TypeAverageEntChangeExp ~ BaseTypeAveEntExp + log(NumTokens) +
##     factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1433 -0.1054  0.1014  0.3085  2.5919
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)      4.582061   0.022160  206.776
## BaseTypeAveEntExp -0.208584   0.002667  -78.210
## log(NumTokens)   -0.978911   0.007687 -127.346
## factor(ParentOp)&&  0.056148   0.009891   5.677
## factor(ParentOp)AssertStatement -1.146844   0.015721 -72.952
## factor(ParentOp)Assignment -0.803217   0.108712  -7.388
## factor(ParentOp)ConditionalExpression  0.561490   0.011965  46.926
## factor(ParentOp)ForStatement  0.608617   0.016104  37.794
## factor(ParentOp)IfStatement -0.772738   0.009655 -80.038
## factor(ParentOp)MethodInvocation  0.058105   0.018394   3.159
## factor(ParentOp)ParenthesizedExpression -0.102102   0.016052  -6.361
## factor(ParentOp)ReturnStatement -0.970784   0.028878 -33.617
## factor(ParentOp)VariableDeclarationFragment -1.239162   0.073559 -16.846
## factor(ParentOp)WhileStatement -1.228226   0.031756 -38.677
## factor(MostFreqOp)<  0.678541   0.012570  53.982
## factor(MostFreqOp)<=  1.072422   0.032941  32.556
## factor(MostFreqOp)>  0.220769   0.010645  20.740
## factor(MostFreqOp)>= -0.096387   0.013629  -7.072
## factor(MostFreqOp)!=  0.165325   0.004297  38.472
##
##              Pr(>|t|)
## (Intercept)      < 2e-16 ***
## BaseTypeAveEntExp < 2e-16 ***
## log(NumTokens)   < 2e-16 ***
## factor(ParentOp)&& 1.38e-08 ***
## factor(ParentOp)AssertStatement < 2e-16 ***
## factor(ParentOp)Assignment 1.50e-13 ***
## factor(ParentOp)ConditionalExpression < 2e-16 ***
## factor(ParentOp)ForStatement < 2e-16 ***
## factor(ParentOp)IfStatement < 2e-16 ***
## factor(ParentOp)MethodInvocation 0.00158 **
## factor(ParentOp)ParenthesizedExpression 2.01e-10 ***
## factor(ParentOp)ReturnStatement < 2e-16 ***
## factor(ParentOp)VariableDeclarationFragment < 2e-16 ***
## factor(ParentOp)WhileStatement < 2e-16 ***
## factor(MostFreqOp)< < 2e-16 ***
## factor(MostFreqOp)<= < 2e-16 ***
## factor(MostFreqOp)> < 2e-16 ***
## factor(MostFreqOp)>= 1.53e-12 ***
## factor(MostFreqOp)!= < 2e-16 ***
## ---

```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.583 on 113867 degrees of freedom
## Multiple R-squared:  0.3791, Adjusted R-squared:  0.379
## F-statistic: 3862 on 18 and 113867 DF,  p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: TypeAverageEntChangeExp
##
##      Df Sum Sq Mean Sq  F value    Pr(>F)
## BaseTypeAveEntExp      1   5337   5337.1 15700.76 < 2.2e-16 ***
## log(NumTokens)         1    287    286.7   843.47 < 2.2e-16 ***
## factor(ParentOp)      11 16175   1470.5  4325.82 < 2.2e-16 ***
## factor(MostFreqOp)     5   1833    366.5  1078.19 < 2.2e-16 ***
## Residuals             113867 38707     0.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.085615794" "0.004599393" "0.259474488" "0.029396669" "0.620913657"
##
##      GVIF Df GVIF^(1/(2*Df))
## BaseTypeAveEntExp  1.624969  1      1.274743
## log(NumTokens)     3.276362  1      1.810072
## factor(ParentOp)  15.820965 11      1.133732
## factor(MostFreqOp) 4.947884  5      1.173389
##
## % 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 - 10:17:04 PM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
##   \begin{tabular}{@{\extracolsep{5pt}}lc}
##     \hline
##     & \multicolumn{1}{c}{\textit{Dependent variable:}} & \\
##     \hline
##     \hline & TypeAverageEntChangeExp & \\
##     \hline
##     BaseTypeAveEntExp &  $0.209^{***}$  & (0.003) \\
##     log(NumTokens) &  $0.979^{***}$  & (0.008) \\
##     factor(ParentOp) &  $0.056^{***}$  & (0.010) \\
##     factor(ParentOp)AssertStatement &  $1.147^{***}$  & (0.016) \\
##     factor(ParentOp)Assignment &  $0.803^{***}$  & (0.109) \\
##     factor(ParentOp)ConditionalExpression &  $0.561^{***}$  & (0.012) \\
##     factor(ParentOp)ForStatement &  $0.609^{***}$  & (0.016) \\
##     factor(ParentOp)IfStatement &  $0.773^{***}$  & (0.010) \\
##     factor(ParentOp)MethodInvocation &  $0.058^{***}$  & (0.018) \\
##     factor(ParentOp)ParenthesizedExpression &  $0.102^{***}$  & (0.016) \\
##     factor(ParentOp)ReturnStatement &  $0.971^{***}$  & (0.029) \\
##     factor(ParentOp)VariableDeclarationFragment &  $1.239^{***}$  & (0.074) \\
##     factor(ParentOp)WhileStatement &  $1.228^{***}$  & (0.032) \\
##     factor(MostFreqOp)\textless &  $0.679^{***}$  & (0.013) \\
##     factor(MostFreqOp)\textless = &  $1.072^{***}$  & (0.033) \\
##     factor(MostFreqOp)\textgreater &  $0.221^{***}$  & (0.011) \\
##     factor(MostFreqOp)\textgreater = &  $0.096^{***}$  & (0.014) \\
##     factor(MostFreqOp)!= &  $0.165^{***}$  & (0.004)

```

```

## Constant & 4.582 $\hat{\{***\}}$ $ (0.022) \\
## \hline \\[-1.8ex]
## Observations & 113,886 \\
## R $\hat{\{2\}}$ $ & 0.379 \\
## Adjusted R $\hat{\{2\}}$ $ & 0.379 \\
## Residual Std. Error & 0.583 (df = 113867) \\
## F Statistic & 3,862.178 $\hat{\{***\}}$ $ (df = 18; 113867) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:} & \multicolumn{1}{r}{ $\hat{\{*\}}$ $p$<$0.1;  $\hat{\{**\}}$ $p$<$0.05;  $\hat{\{***\}}$ $p$<$0.01} \\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:17:04 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
## \hline
## & Df & Sum Sq & Mean Sq & F value & Pr(>$F) \\
## \hline
## BaseTypeAveEntExp & 1 & 5337.15 & 5337.15 & 15700.76 & 0.0000 \\
## log(NumTokens) & 1 & 286.72 & 286.72 & 843.47 & 0.0000 \\
## factor(ParentOp) & 11 & 16175.21 & 1470.47 & 4325.82 & 0.0000 \\
## factor(MostFreqOp) & 5 & 1832.54 & 366.51 & 1078.19 & 0.0000 \\
## Residuals & 113867 & 38706.73 & 0.34 & & \\
## \hline
## \end{tabular}
## \end{table}

```

```
mt_lswap_cache_no_out <- modelCacheType(dlspFiltered, "", "==" )
```

```

##
## Call:
## lm(formula = CacheTypeAverageEntChangeExp ~ BaseCacheTypeAveEntExp +
## log(NumTokens) + factor(ParentOp) + factor(MostFreqOp), data = dataset)
##
## Residuals:
## Min 1Q Median 3Q Max
## -3.4280 -0.1879 0.0181 0.3889 5.9809
##
## Coefficients:
## Estimate Std. Error t value
## (Intercept) 4.728175 0.025518 185.288
## BaseCacheTypeAveEntExp -0.466017 0.002718 -171.428
## log(NumTokens) -0.820527 0.008769 -93.570
## factor(ParentOp)&& 0.078519 0.011391 6.893
## factor(ParentOp)AssertStatement -0.951294 0.018220 -52.210
## factor(ParentOp)Assignment -0.260297 0.124590 -2.089
## factor(ParentOp)ConditionalExpression 1.063816 0.013525 78.658
## factor(ParentOp)ForStatement 0.215905 0.018020 11.981
## factor(ParentOp)IfStatement -0.661746 0.011080 -59.725
## factor(ParentOp)MethodInvocation 0.674813 0.022331 30.219
## factor(ParentOp)ParenthesizedExpression -0.014355 0.018756 -0.765
## factor(ParentOp)ReturnStatement -0.741006 0.037526 -19.746
## factor(ParentOp)VariableDeclarationFragment -1.039330 0.100405 -10.351

```

```

## factor(ParentOp)WhileStatement          -0.998477  0.038970 -25.622
## factor(MostFreqOp)<                     0.733232  0.014146  51.834
## factor(MostFreqOp)<=                    1.298375  0.035279  36.803
## factor(MostFreqOp)>                     0.399199  0.012049  33.131
## factor(MostFreqOp)>=                    0.266611  0.016211  16.446
## factor(MostFreqOp)!=                    0.206397  0.004959  41.619
##                                         Pr(>|t|)
## (Intercept)                             < 2e-16 ***
## BaseCacheTypeAveEntExp                   < 2e-16 ***
## log(NumTokens)                           < 2e-16 ***
## factor(ParentOp)&&                        5.48e-12 ***
## factor(ParentOp)AssertStatement           < 2e-16 ***
## factor(ParentOp)Assignment                0.0367 *
## factor(ParentOp)ConditionalExpression     < 2e-16 ***
## factor(ParentOp)ForStatement              < 2e-16 ***
## factor(ParentOp)IfStatement              < 2e-16 ***
## factor(ParentOp)MethodInvocation         < 2e-16 ***
## factor(ParentOp)ParenthesizedExpression  0.4441
## factor(ParentOp)ReturnStatement          < 2e-16 ***
## factor(ParentOp)VariableDeclarationFragment < 2e-16 ***
## factor(ParentOp)WhileStatement           < 2e-16 ***
## factor(MostFreqOp)<                      < 2e-16 ***
## factor(MostFreqOp)<=                     < 2e-16 ***
## factor(MostFreqOp)>                      < 2e-16 ***
## factor(MostFreqOp)>=                     < 2e-16 ***
## factor(MostFreqOp)!=                     < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6688 on 113147 degrees of freedom
## Multiple R-squared:  0.4037, Adjusted R-squared:  0.4036
## F-statistic: 4256 on 18 and 113147 DF, p-value: < 2.2e-16
##
## Analysis of Variance Table
##
## Response: CacheTypeAverageEntChangeExp
##
##              Df Sum Sq Mean Sq F value    Pr(>F)
## BaseCacheTypeAveEntExp    1  16065  16064.9  35916.9 < 2.2e-16 ***
## log(NumTokens)            1     717    716.6  1602.2 < 2.2e-16 ***
## factor(ParentOp)          11  14949   1359.0  3038.3 < 2.2e-16 ***
## factor(MostFreqOp)         5   2534    506.8  1133.0 < 2.2e-16 ***
## Residuals                 113147  50608     0.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "0.189282850" "0.008443544" "0.176133260" "0.029853655" "0.596286691"
##
##              GVIF Df GVIF^(1/(2*Df))
## BaseCacheTypeAveEntExp  1.328143  1     1.152451
## log(NumTokens)          3.216374  1     1.793425
## factor(ParentOp)        13.169249 11     1.124318
## factor(MostFreqOp)       4.598166  5     1.164819
##
## % 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 - 10:17:16 PM
## \begin{table}[!htbp] \centering

```

```

## \caption{}
## \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \[-1.8ex]\hline
## \hline \[-1.8ex]
## & \multicolumn{1}{c}{\textit{Dependent variable:}} \\\
## \cline{2-2}
## \[-1.8ex] & CacheTypeAverageEntChangeExp \\\
## \hline \[-1.8ex]
## BaseCacheTypeAveEntExp &  $-\$0.466^{\{***\}}\$$  (0.003) \\\
## \log(NumTokens) &  $-\$0.821^{\{***\}}\$$  (0.009) \\\
## \factor(ParentOp)&& &  $0.079^{\{***\}}\$$  (0.011) \\\
## \factor(ParentOp)AssertStatement &  $-\$0.951^{\{***\}}\$$  (0.018) \\\
## \factor(ParentOp)Assignment &  $-\$0.260^{\{**\}}\$$  (0.125) \\\
## \factor(ParentOp)ConditionalExpression &  $1.064^{\{***\}}\$$  (0.014) \\\
## \factor(ParentOp)ForStatement &  $0.216^{\{***\}}\$$  (0.018) \\\
## \factor(ParentOp)IfStatement &  $-\$0.662^{\{***\}}\$$  (0.011) \\\
## \factor(ParentOp)MethodInvocation &  $0.675^{\{***\}}\$$  (0.022) \\\
## \factor(ParentOp)ParenthesizedExpression &  $-\$0.014$  (0.019) \\\
## \factor(ParentOp)ReturnStatement &  $-\$0.741^{\{***\}}\$$  (0.038) \\\
## \factor(ParentOp)VariableDeclarationFragment &  $-\$1.039^{\{***\}}\$$  (0.100) \\\
## \factor(ParentOp)WhileStatement &  $-\$0.998^{\{***\}}\$$  (0.039) \\\
## \factor(MostFreqOp)\textless &  $0.733^{\{***\}}\$$  (0.014) \\\
## \factor(MostFreqOp)\textless = &  $1.298^{\{***\}}\$$  (0.035) \\\
## \factor(MostFreqOp)\textgreater &  $0.399^{\{***\}}\$$  (0.012) \\\
## \factor(MostFreqOp)\textgreater = &  $0.267^{\{***\}}\$$  (0.016) \\\
## \factor(MostFreqOp)!= &  $0.206^{\{***\}}\$$  (0.005) \\\
## Constant &  $4.728^{\{***\}}\$$  (0.026) \\\
## \hline \[-1.8ex]
## Observations & 113,166 \\\
## R2 & 0.404 \\\
## Adjusted R2 & 0.404 \\\
## Residual Std. Error & 0.669 (df = 113147) \\\
## F Statistic & 4,255.871{***} (df = 18; 113147) \\\
## \hline
## \hline \[-1.8ex]
## \textit{Note:} & \multicolumn{1}{r}{ $^{\{*\}}\$p < \$0.1$ ;  $^{\{**\}}\$p < \$0.05$ ;  $^{\{***\}}\$p < \$0.01$ } \\\
## \end{tabular}
## \end{table}
## % latex table generated in R 3.4.4 by xtable 1.8-3 package
## % Tue Feb 19 22:17:16 2019
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
## \hline
## & Df & Sum Sq & Mean Sq & F value & Pr(>F) \\\
## \hline
## BaseCacheTypeAveEntExp & 1 & 16064.95 & 16064.95 & 35916.93 & 0.0000 \\\
## \log(NumTokens) & 1 & 716.63 & 716.63 & 1602.19 & 0.0000 \\\
## \factor(ParentOp) & 11 & 14948.91 & 1358.99 & 3038.34 & 0.0000 \\\
## \factor(MostFreqOp) & 5 & 2533.76 & 506.75 & 1132.96 & 0.0000 \\\
## Residuals & 113147 & 50608.46 & 0.45 & & \\\
## \hline
## \end{tabular}
## \end{table}

```



```
## \end{table}
```

Output the effect sizes

```
JLogicOut <- printEffTable(pairedResults)
```

```
## [1] "Type,PTOne,PTTwo,CITTwo,CohensD,PWilcoxOne,PWilcoxTwo,CIWilcoxTwo,CliffDelta"  
## [1] "LogicalSwapTopCacheExp,0,0,-2.6338 -2.5994,-0.9718,0,0,-2.5568 -2.5205,-0.3817"  
## [1] "LogicalSwapTopCacheTypeExp,0,0,-1.6411 -1.6257,-1.3564,0,0,-1.9643 -1.9506,-0.7115"  
## [1] "LogicalSwapTopGlobalExp,0,0,-2.209 -2.1753,-0.8331,0,0,-2.1512 -2.1064,-0.2645"  
## [1] "LogicalSwapTopGlobalTypeExp,0,0,-1.3977 -1.3846,-1.3628,0,0,-1.758 -1.756,-0.6958"  
## [1] "LogicalSwapTopFiltered100CacheExp,0,0,-2.1484 -2.0975,-0.7071,0,0,-1.9633 -1.9057,-0.3914"  
## [1] "LogicalSwapTopFiltered100CacheTypeExp,0,0,-1.9223 -1.9026,-1.6494,0,0,-1.9537 -1.9362,-0.8255"  
## [1] "LogicalSwapTopFiltered100GlobalExp,0,0,-1.4836 -1.4401,-0.5701,0,0,-1.2165 -1.1686,-0.1614"  
## [1] "LogicalSwapTopFiltered100GlobalTypeExp,0,0,-1.6396 -1.6222,-1.5975,0,0,-1.7201 -1.7013,-0.7688"  
## [1] "LogicalSwapTopFiltered10CacheExp,0,0,-2.024 -1.971,-0.6698,0,0,-1.8073 -1.7479,-0.3742"  
## [1] "LogicalSwapTopFiltered10CacheTypeExp,0,0,-1.9052 -1.8846,-1.6375,0,0,-1.9315 -1.9132,-0.8193"  
## [1] "LogicalSwapTopFiltered10GlobalExp,0,0,-1.4019 -1.357,-0.5454,0,0,-1.1181 -1.0712,-0.152"  
## [1] "LogicalSwapTopFiltered10GlobalTypeExp,0,0,-1.6381 -1.6198,-1.5822,0,0,-1.7137 -1.6955,-0.7655"
```

```
save(JLogicOut, file = "/data/anon/SemanticTransformation/sample/JLogicOut.RDat")  
save(dlsp, file = "/data/anon/SemanticTransformation/sample/dlsp.RDat")  
save(dlspFiltered, file = "/data/anon/SemanticTransformation/sample/dlspFiltered.RDat")  
save(dlspFiltered2, file = "/data/anon/SemanticTransformation/sample/dlspFiltered2.RDat")
```