Parkinson: Decorrelation-Based Feature Discovery

Jose Tamez

2022-10-04

Table of Contents

# Outputs Analysis and Graphics: Effect of GDSTM-Based Decorrelation on Feature Discovery

This will load the results of the Parkinson Analysis and create the outputs.

## From the original

Here I showcase of to use BSWiMS feature selection/modeling function coupled with Goal Driven Sparse Transformation Matrix (GDSTM) as a pre-processing step to decorrelate highly correlated features. The aim(s) are:

1. To improve model performance by uncovering the hidden information between correlated features.
2. To simplify the interpretation of the machine learning models.

This demo will use:

* FRESA.CAD::GDSTMDecorrelation(). For Decorrelation of Multidimensional data sets
  + FRESA.CAD::getDerivedCoefficients(). For the extraction of the decorrelated features.
* FRESA.CAD::randomCV() For the cross-validation of the Machine Learning models
* FRESA.CAD::BSWiMS.model(). For the generation of bootstrapped logistic models
  + FRESA.CAD::summary(). For the summary description of the BSWiMS model
* FRESA.CAD::predictionStats\_binary(). For describing the performance of the model
* heatmap.2(). For displaying the correlation matrix
* vioplot::vioplot(). For the display of the z-distribution of significant features.

### Loading the libraries

library("FRESA.CAD")  
library(readxl)  
library(vioplot)  
library(igraph)  
  
op <- par(no.readonly = TRUE)  
pander::panderOptions('digits', 3)  
pander::panderOptions('table.split.table', 400)  
pander::panderOptions('keep.trailing.zeros',TRUE)

## Material and Methods

### Signed Log Transform

The function will be used to transform all the continuous features of the data

signedlog <- function(x) { return (sign(x)\*log(abs(x)+1.0e-12))}

## Data: The Parkinson Data-Set

The data to process is described in:

Erdogdu Sakar, Betul, Gorkem Serbes, and C. Okan Sakar. “Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson’s disease.” *PloS one* 12, no. 8 (2017): e0182428.

The data was obtained from the UCI ML repository:

<https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification>

I added a column to the data identifying the repeated experiments.

load(file="~/GitHub/FCA/ParkinsonDemo.RData")  
  
namecode <- read.csv("~/GitHub/FCA/Data/Parkinson\_names.csv")

### The Average of the Three Repetitions

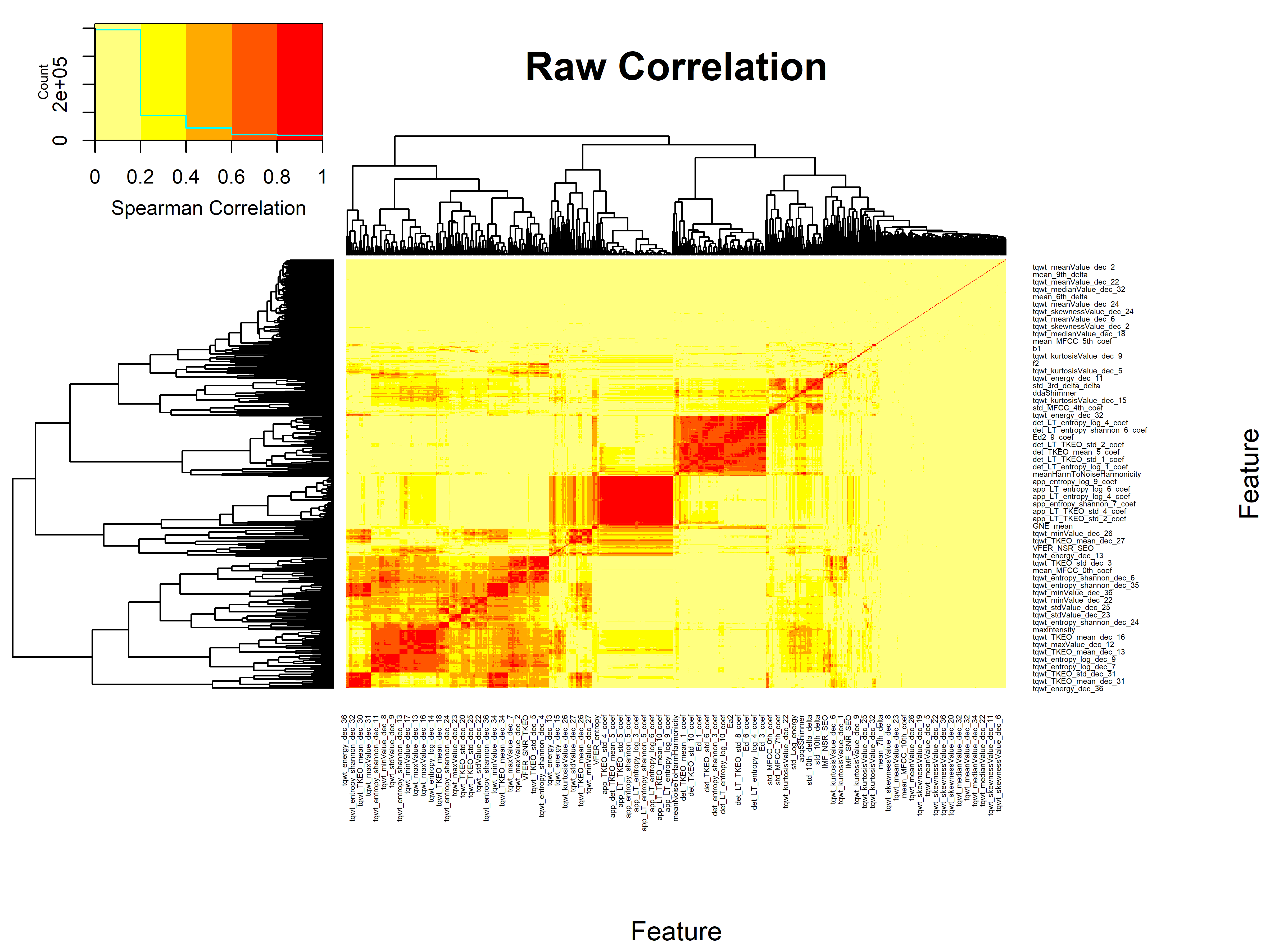
Each subject had three repeated observations. Here I’ll use the average of the three experiments per subject.

# rep1Parkison <- subset(pd\_speech\_features,RID==1)  
# rownames(rep1Parkison) <- rep1Parkison$id  
# rep1Parkison$id <- NULL  
# rep1Parkison$RID <- NULL  
# rep1Parkison[,1:ncol(rep1Parkison)] <- sapply(rep1Parkison,as.numeric)  
#   
# rep2Parkison <- subset(pd\_speech\_features,RID==2)  
# rownames(rep2Parkison) <- rep2Parkison$id  
# rep2Parkison$id <- NULL  
# rep2Parkison$RID <- NULL  
# rep2Parkison[,1:ncol(rep2Parkison)] <- sapply(rep2Parkison,as.numeric)  
#   
# rep3Parkison <- subset(pd\_speech\_features,RID==3)  
# rownames(rep3Parkison) <- rep3Parkison$id  
# rep3Parkison$id <- NULL  
# rep3Parkison$RID <- NULL  
# rep3Parkison[,1:ncol(rep3Parkison)] <- sapply(rep3Parkison,as.numeric)  
#   
# whof <- !(colnames(rep1Parkison) %in% c("gender","class"));  
# avgParkison <- rep1Parkison;  
# avgParkison[,whof] <- (rep1Parkison[,whof] + rep2Parkison[,whof] + rep3Parkison[,whof])/3  
# ## I apply the log transform to the data  
# avgParkison[,whof] <- signedlog(avgParkison[,whof])  
# pander::pander(table(avgParkison$class))

### Correlation Matrix of the Parkinson Data

The heat-map of the correlation:

cormat <- cor(avgParkison,method="spearman")  
gplots::heatmap.2(abs(cormat),  
 trace = "none",  
 scale = "none",  
 mar = c(10,10),  
 col=rev(heat.colors(5)),  
 main = "Raw Correlation",  
 cexRow = 0.35,  
 cexCol = 0.35,  
 key.title=NA,  
 key.xlab="Spearman Correlation",  
 xlab="Feature", ylab="Feature",  
# srtRow = 45,  
# srtCol = 45  
)



### Training and Testing Sets

We divided the data into training and testing sets.

# set.seed(2)  
# caseSet <- subset(avgParkison, class == 1)  
# controlSet <- subset(avgParkison, class == 0)  
# caseTrainSize <- nrow(caseSet)\*trainFraction;  
# controlTrainSize <- nrow(controlSet)\*trainFraction;  
# sampleCaseTrain <- sample(nrow(caseSet),caseTrainSize)  
# sampleControlTrain <- sample(nrow(controlSet),controlTrainSize)  
# trainSet <- rbind(caseSet[sampleCaseTrain,], controlSet[sampleControlTrain,])  
# testSet <- rbind(caseSet[-sampleCaseTrain,],controlSet[-sampleControlTrain,])  
pander::pander(table(trainSet$class))

| 0 | 1 |
| --- | --- |
| 41 | 122 |

pander::pander(table(testSet$class))

| 0 | 1 |
| --- | --- |
| 23 | 66 |

#### Decorrelation: Training and Testing Sets Creation

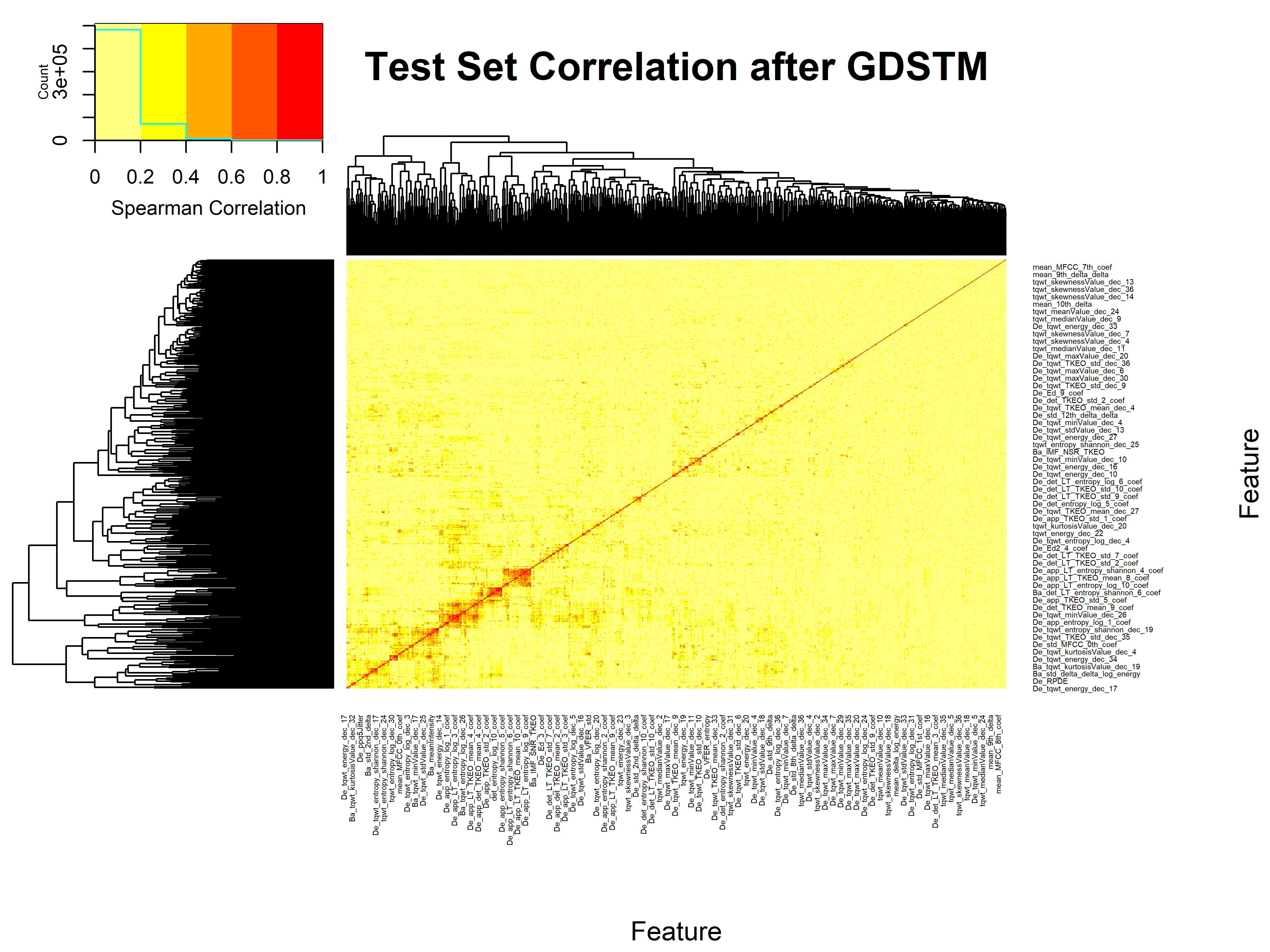
I compute a decorrelated version of the training and testing sets using the *GDSTMDecorrelation()* function of FRESA.CAD. The first decorrelation will be driven by features associated with the outcome. The second decorrelation will find the GDSTM without the outcome restriction.

## The GDSTM transformation driven by the Outcome  
  
# deTrain <- GDSTMDecorrelation(trainSet,Outcome="class",thr=0.8,verbose = TRUE)  
# deTest <- predictDecorrelate(deTrain,testSet)  
  
## The GDSTM transformation without outcome  
  
# deTrainU <- GDSTMDecorrelation(trainSet,thr=0.8,verbose = TRUE)  
# deTestU <- predictDecorrelate(deTrainU,testSet)

#### Correlation Matrix of the Decorrelated Test Data

The heat map of the testing set.

cormat <- cor(deTest,method="spearman")  
gplots::heatmap.2(abs(cormat),  
 trace = "none",  
 scale = "none",  
 mar = c(10,10),  
 col=rev(heat.colors(5)),  
 main = "Test Set Correlation after GDSTM",  
 cexRow = 0.35,  
 cexCol = 0.35,  
 key.title=NA,  
 key.xlab="Spearman Correlation",  
 xlab="Feature", ylab="Feature")



### Holdout Cross-Validation

Before doing the feature analysis. I’ll explore BSWiMS modeling using the Holdout cross validation method of FRESA.CAD. The purpose of the cross-validation is to observe and estimate the performance gain of decorrelation.

par(op)  
par(mfrow=c(1,3))  
  
## The Raw validation  
# cvBSWiMSRaw <- randomCV(avgParkison,  
# "class",  
# fittingFunction= BSWiMS.model,  
# classSamplingType = "Pro",  
# trainFraction = trainFraction,  
# repetitions = 150  
# )  
  
bpraw <- predictionStats\_binary(cvBSWiMSRaw$medianTest,"BSWiMS RAW",cex=0.60)

BSWiMS RAW

pander::pander(bpraw$CM.analysis$tab)

|  | Outcome + | Outcome - | Total |
| --- | --- | --- | --- |
| **Test +** | 145 | 18 | 163 |
| **Test -** | 43 | 46 | 89 |
| **Total** | 188 | 64 | 252 |

pander::pander(bpraw$accc)

|  | est | lower | upper |
| --- | --- | --- | --- |
| **5** | 0.758 | 0.7 | 0.809 |

pander::pander(bpraw$aucs)

| est | lower | upper |
| --- | --- | --- |
| 0.838 | 0.782 | 0.894 |

pander::pander(bpraw$berror)

| 50% | 2.5% | 97.5% |
| --- | --- | --- |
| 0.256 | 0.197 | 0.326 |

## The validation with Outcome-driven Decorrelation  
# cvBSWiMSDeCor <- randomCV(avgParkison,  
# "class",  
# trainSampleSets= cvBSWiMSRaw$trainSamplesSets,  
# fittingFunction= filteredFit,  
# fitmethod=BSWiMS.model,  
# filtermethod=NULL,  
# DECOR = TRUE,  
# DECOR.control=list(Outcome="class",thr=0.8)  
# )  
  
bpDecor <- predictionStats\_binary(cvBSWiMSDeCor$medianTest,"BSWiMS Outcome-Driven GDSTM",cex=0.60)

BSWiMS Outcome-Driven GDSTM

pander::pander(bpDecor$CM.analysis$tab)

|  | Outcome + | Outcome - | Total |
| --- | --- | --- | --- |
| **Test +** | 167 | 13 | 180 |
| **Test -** | 21 | 51 | 72 |
| **Total** | 188 | 64 | 252 |

pander::pander(bpDecor$accc)

|  | est | lower | upper |
| --- | --- | --- | --- |
| **5** | 0.865 | 0.817 | 0.905 |

pander::pander(bpDecor$aucs)

| est | lower | upper |
| --- | --- | --- |
| 0.874 | 0.816 | 0.932 |

pander::pander(bpDecor$berror)

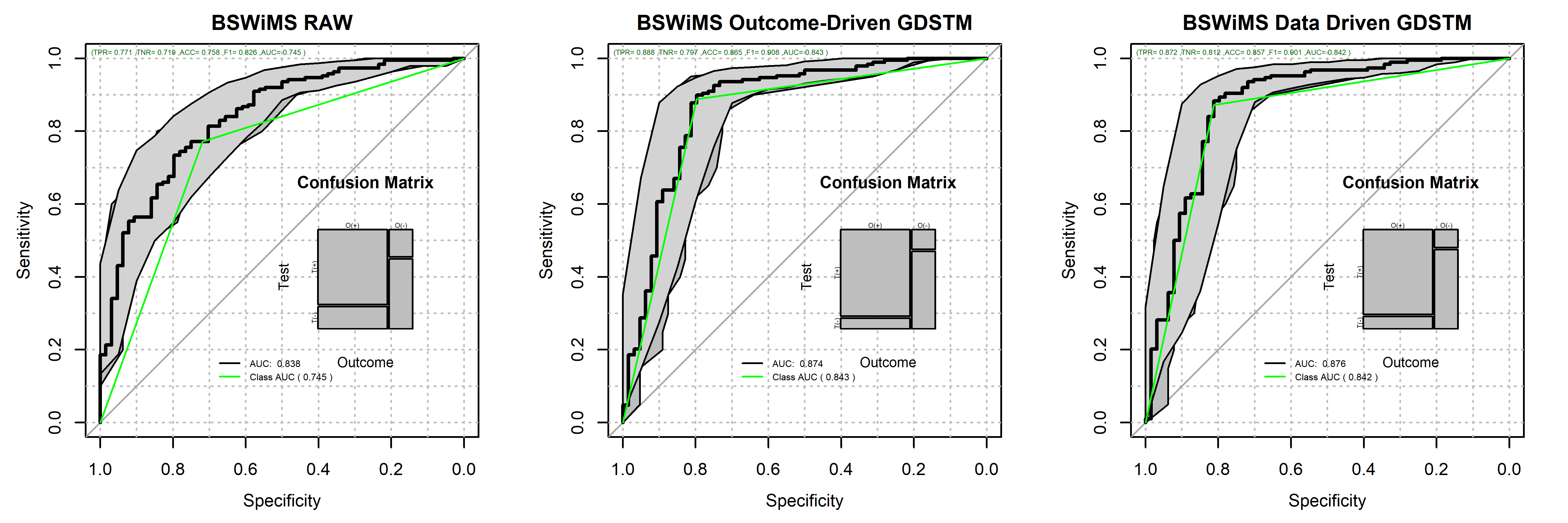
| 50% | 2.5% | 97.5% |
| --- | --- | --- |
| 0.157 | 0.105 | 0.214 |

### Here we compute the probability that the outcome-driven decorrelation ROC is superior to the RAW ROC.   
pander::pander(roc.test(bpDecor$ROC.analysis$roc.predictor,bpraw$ROC.analysis$roc.predictor))

DeLong’s test for two correlated ROC curves: bpDecor$ROC.analysis$roc.predictor and bpraw$ROC.analysis$roc.predictor

| Test statistic | P value | Alternative hypothesis | AUC of roc1 | AUC of roc2 |
| --- | --- | --- | --- | --- |
| 2.41 | 0.0159 \* | two.sided | 0.874 | 0.838 |

## The validation of Decorrelation without the outcome restriction  
# cvBSWiMSDeCorU <- randomCV(avgParkison,  
# "class",  
# trainSampleSets= cvBSWiMSRaw$trainSamplesSets,  
# fittingFunction= filteredFit,  
# fitmethod=BSWiMS.model,  
# filtermethod=NULL,  
# DECOR = TRUE,  
# DECOR.control=list(thr=0.8)  
# )  
  
bpDecorU <- predictionStats\_binary(cvBSWiMSDeCorU$medianTest,"BSWiMS Data Driven GDSTM",cex=0.60)

BSWiMS Data Driven GDSTM 

pander::pander(bpDecorU$CM.analysis$tab)

|  | Outcome + | Outcome - | Total |
| --- | --- | --- | --- |
| **Test +** | 164 | 12 | 176 |
| **Test -** | 24 | 52 | 76 |
| **Total** | 188 | 64 | 252 |

pander::pander(bpDecorU$accc)

|  | est | lower | upper |
| --- | --- | --- | --- |
| **5** | 0.857 | 0.808 | 0.898 |

pander::pander(bpDecorU$aucs)

| est | lower | upper |
| --- | --- | --- |
| 0.876 | 0.818 | 0.934 |

pander::pander(bpDecorU$berror)

| 50% | 2.5% | 97.5% |
| --- | --- | --- |
| 0.157 | 0.105 | 0.212 |

### Here we compute the probability that the blind decorrelation ROC is superior to the RAW ROC.   
pander::pander(roc.test(bpDecorU$ROC.analysis$roc.predictor,bpraw$ROC.analysis$roc.predictor))

DeLong’s test for two correlated ROC curves: bpDecorU$ROC.analysis$roc.predictor and bpraw$ROC.analysis$roc.predictor

| Test statistic | P value | Alternative hypothesis | AUC of roc1 | AUC of roc2 |
| --- | --- | --- | --- | --- |
| 2.59 | 0.00948 \* \* | two.sided | 0.876 | 0.838 |

par(op)

## The Raw Model *vs.* the Decorrelated-Based Model

After demonstrating that decorrelation is able to improve BSWiMS model performance, I’ll focus is showcasing the ability to discover new features associated with the outcome.

First, I’ll compute the BSWiMS models for the original data, and for the decorrelated data-set. The model estimation will be done using the training set and tested on the holdout test set, and repeated 10 times. After that, I’ll compare the statistical difference of both ROC curves.

par(op)  
par(mfrow=c(1,3))  
  
# bm <- BSWiMS.model(class~.,trainSet,NumberofRepeats = 20)  
bpraw <- predictionStats\_binary(cbind(testSet$class,predict(bm,testSet)),"BSWiMS RAW",cex=0.60)

BSWiMS RAW

# bmd <- BSWiMS.model(class~.,deTrain,NumberofRepeats = 20)  
bpdecor <- predictionStats\_binary(cbind(deTest$class,predict(bmd,deTest)),"Outcome-Driven Decorrelation",cex=0.60)

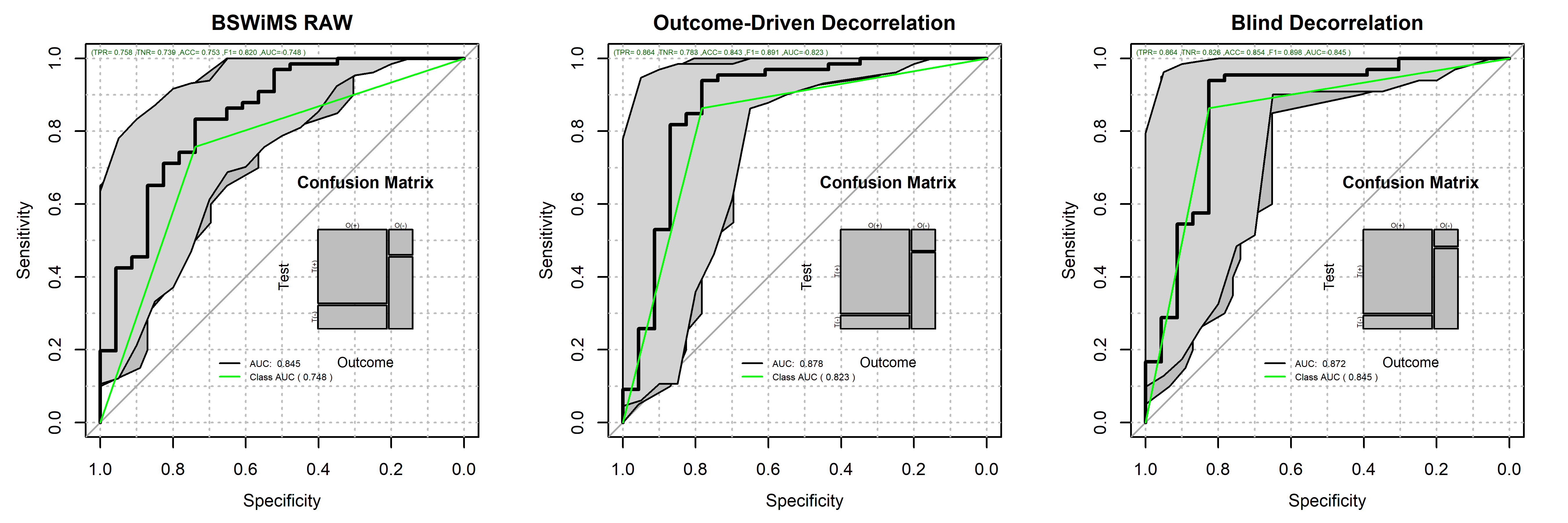
Outcome-Driven Decorrelation

pander::pander(roc.test(bpraw$ROC.analysis$roc.predictor,bpdecor$ROC.analysis$roc.predictor))

DeLong’s test for two correlated ROC curves: bpraw$ROC.analysis$roc.predictor and bpdecor$ROC.analysis$roc.predictor

| Test statistic | P value | Alternative hypothesis | AUC of roc1 | AUC of roc2 |
| --- | --- | --- | --- | --- |
| -1.42 | 0.157 | two.sided | 0.845 | 0.878 |

# bmdU <- BSWiMS.model(class~.,deTrainU,NumberofRepeats = 20)  
bpdecorU <- predictionStats\_binary(cbind(deTest$class,predict(bmdU,deTestU)),"Blind Decorrelation",cex=0.60)

Blind Decorrelation 

pander::pander(roc.test(bpraw$ROC.analysis$roc.predictor,bpdecorU$ROC.analysis$roc.predictor))

DeLong’s test for two correlated ROC curves: bpraw$ROC.analysis$roc.predictor and bpdecorU$ROC.analysis$roc.predictor

| Test statistic | P value | Alternative hypothesis | AUC of roc1 | AUC of roc2 |
| --- | --- | --- | --- | --- |
| -1.16 | 0.246 | two.sided | 0.845 | 0.872 |

par(op)

## The feature associations

par(op)  
par(mfrow=c(1,3))  
### The raw model  
  
pander::pander(nrow(bm$bagging$formulaNetwork))

*67*

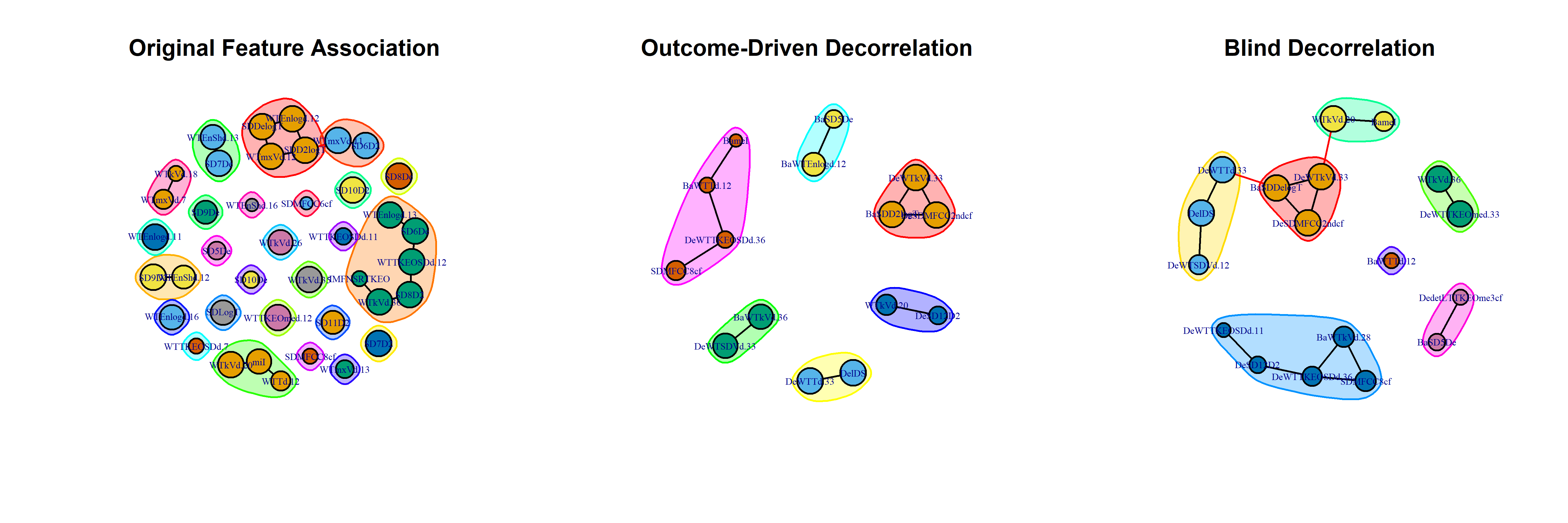
cmax <- apply(bm$bagging$formulaNetwork,2,max)  
cnames <- names(cmax[cmax>=0.5])  
cmax <- cmax[cmax>=0.5]  
adma <- bm$bagging$formulaNetwork[cnames,cnames]  
  
for (cx in c(1:nrow(namecode)))  
{  
 cnames <- str\_replace\_all(cnames,namecode[cx,1],namecode[cx,2])  
}  
cnames <- str\_replace\_all(cnames,"\_","")  
cnames <- str\_replace\_all(cnames,"th","")  
rownames(adma) <- cnames  
colnames(adma) <- cnames  
names(cmax) <- cnames  
adma[adma<0.25] <- 0;  
gr <- graph\_from\_adjacency\_matrix(adma,mode = "undirected",diag = FALSE,weighted=TRUE)  
gr$layout <- layout\_with\_fr  
  
fc <- cluster\_optimal(gr)  
plot(fc, gr,  
 vertex.size=20\*cmax,  
 vertex.label.cex=0.5,  
 vertex.label.dist=0,  
 main="Original Feature Association")  
  
  
  
### The Outcome Driven Model  
  
pander::pander(nrow(bmd$bagging$formulaNetwork))

*36*

cmax <- apply(bmd$bagging$formulaNetwork,2,max)  
cnames <- names(cmax[cmax>=0.5])  
cmax <- cmax[cmax>=0.5]  
outcomeNames <- cnames  
adma <- bmd$bagging$formulaNetwork[cnames,cnames]  
  
for (cx in c(1:nrow(namecode)))  
{  
 cnames <- str\_replace\_all(cnames,namecode[cx,1],namecode[cx,2])  
}  
cnames <- str\_replace\_all(cnames,"\_","")  
cnames <- str\_replace\_all(cnames,"th","")  
rownames(adma) <- cnames  
colnames(adma) <- cnames  
names(cmax) <- cnames  
adma[adma<0.25] <- 0;  
gr <- graph\_from\_adjacency\_matrix(adma,mode = "undirected",diag = FALSE,weighted=TRUE)  
gr$layout <- layout\_with\_fr  
  
fc <- cluster\_optimal(gr)  
clusterOutcome <- fc  
clusterOutcome$names <- outcomeNames  
plot(fc, gr,  
 vertex.size=20\*cmax,  
 vertex.label.cex=0.5,  
 vertex.label.dist=0,  
 main="Outcome-Driven Decorrelation")  
  
  
### The Blind Decorrelation  
  
pander::pander(nrow(bmdU$bagging$formulaNetwork))

*58*

cmax <- apply(bmdU$bagging$formulaNetwork,2,max)  
cnames <- names(cmax[cmax>=0.5])  
cmax <- cmax[cmax>=0.5]  
adma <- bmdU$bagging$formulaNetwork[cnames,cnames]  
  
for (cx in c(1:nrow(namecode)))  
{  
 cnames <- str\_replace\_all(cnames,namecode[cx,1],namecode[cx,2])  
}  
cnames <- str\_replace\_all(cnames,"\_","")  
cnames <- str\_replace\_all(cnames,"th","")  
rownames(adma) <- cnames  
colnames(adma) <- cnames  
names(cmax) <- cnames  
adma[adma<0.25] <- 0;  
gr <- graph\_from\_adjacency\_matrix(adma,mode = "undirected",diag = FALSE,weighted=TRUE)  
gr$layout <- layout\_with\_fr  
  
fc <- cluster\_optimal(gr)  
plot(fc, gr,  
 vertex.size=20\*cmax,  
 vertex.label.cex=0.5,  
 vertex.label.dist=0,  
 main="Blind Decorrelation")



### Feature Analysis of Models

The analysis of the features required to predict the outcome will use the following:

1. Analysis of the BSWiMS bagged model using the summary function.
2. Analysis of the sparse GDSMT
3. Analysis of the univariate association of the model features of both models
4. Report the new features not found by the Original data analysis

par(op)  
par(mfrow=c(1,1))  
## 1 Get the Model Features  
smOriginal <- summary(bm)  
rawnames <- rownames(smOriginal$coefficients)  
  
### From Drived Decorrelation  
smDecor <- summary(bmd)  
decornames <- rownames(smDecor$coefficients)  
  
### From Blind Decorrelation  
smDecorU <- summary(bmdU)  
decornamesU <- rownames(smDecorU$coefficients)  
  
  
  
## 2 Get the decorrelation matrix formulas  
dc <- getDerivedCoefficients(deTrain)  
### 2a Get only the ones that were decorrelated by the decorrelation-based model  
deNames\_in\_dc <- decornames[decornames %in% names(dc)]  
selectedlist <- dc[deNames\_in\_dc]  
theDeFormulas <- selectedlist  
  
pander::pander(selectedlist)

* **De\_tqwt\_energy\_dec\_33**:

| * tqwt\_energy\_dec\_31 | * tqwt\_energy\_dec\_33 |
| --- | --- |
| * -0.907 | * 1 |

* **De\_tqwt\_stdValue\_dec\_33**:

| * tqwt\_TKEO\_mean\_dec\_35 | * tqwt\_stdValue\_dec\_32 | * tqwt\_stdValue\_dec\_33 |
| --- | --- | --- |
| * -0.131 | * -0.715 | * 1 |

* **De\_locDbShimmer**:

| * locDbShimmer | * ddaShimmer |
| --- | --- |
| * 1 | * -0.963 |

* **De\_std\_MFCC\_2nd\_coef**:

| * std\_MFCC\_2nd\_coef | * std\_2nd\_delta |
| --- | --- |
| * 1 | * -0.828 |

* **De\_tqwt\_TKEO\_std\_dec\_11**:

| * tqwt\_TKEO\_mean\_dec\_10 | * tqwt\_TKEO\_mean\_dec\_11 | * tqwt\_TKEO\_std\_dec\_11 | * tqwt\_stdValue\_dec\_10 | * tqwt\_minValue\_dec\_17 |
| --- | --- | --- | --- | --- |
| * 0.713 | * -0.644 | * 1 | * -1.64 | * 0.314 |

* **De\_tqwt\_minValue\_dec\_11**:

| * tqwt\_minValue\_dec\_11 | * tqwt\_minValue\_dec\_17 |
| --- | --- |
| * 1 | * -0.918 |

* **De\_std\_12th\_delta\_delta**:

| * std\_MFCC\_12th\_coef | * std\_12th\_delta\_delta |
| --- | --- |
| * -0.914 | * 1 |

* **De\_tqwt\_minValue\_dec\_7**:

| * tqwt\_minValue\_dec\_7 | * tqwt\_minValue\_dec\_17 | * tqwt\_maxValue\_dec\_8 |
| --- | --- | --- |
| * 1 | * -0.0186 | * 0.967 |

* **De\_tqwt\_kurtosisValue\_dec\_33**:

| * tqwt\_kurtosisValue\_dec\_32 | * tqwt\_kurtosisValue\_dec\_33 |
| --- | --- |
| * -0.883 | * 1 |

* **De\_tqwt\_TKEO\_std\_dec\_36**:

| * tqwt\_TKEO\_mean\_dec\_35 | * tqwt\_TKEO\_std\_dec\_36 |
| --- | --- |
| * -0.926 | * 1 |

* **De\_tqwt\_TKEO\_std\_dec\_7**:

| * tqwt\_TKEO\_std\_dec\_7 | * tqwt\_stdValue\_dec\_7 | * tqwt\_minValue\_dec\_17 |
| --- | --- | --- |
| * 1 | * -1.66 | * 0.145 |

* **De\_det\_LT\_TKEO\_mean\_3\_coef**:

| * Ed2\_3\_coef | * det\_LT\_TKEO\_mean\_1\_coef | * det\_LT\_TKEO\_mean\_3\_coef |
| --- | --- | --- |
| * -0.89 | * -0.0834 | * 1 |

* **De\_tqwt\_entropy\_log\_dec\_1**:

| * tqwt\_entropy\_log\_dec\_1 | * tqwt\_entropy\_log\_dec\_2 |
| --- | --- |
| * 1 | * -0.849 |

* **De\_tqwt\_TKEO\_std\_dec\_17**:

| * tqwt\_TKEO\_std\_dec\_17 | * tqwt\_minValue\_dec\_17 |
| --- | --- |
| * 1 | * 2.13 |

names(selectedlist) <- NULL  
### 2b Get the the names of the original features  
  
allDevar <- unique(c(names(unlist(selectedlist)),decornames))  
allDevar <- allDevar[!str\_detect(allDevar,"De\_")]  
allDevar <- str\_remove(allDevar,"Ba\_")  
allDevar <- unique(allDevar)  
  
  
# The analysis of the blind decorrelation  
  
dcU <- getDerivedCoefficients(deTrainU)  
### 2a Get only the ones that were decorrelated by the decorrelation-based model  
deNames\_in\_dcU <- decornamesU[decornamesU %in% names(dcU)]  
selectedlistU <- dcU[deNames\_in\_dcU]  
pander::pander(selectedlistU)

* **De\_tqwt\_TKEO\_std\_dec\_27**:

| * tqwt\_TKEO\_std\_dec\_27 | * tqwt\_stdValue\_dec\_27 | * tqwt\_stdValue\_dec\_28 |
| --- | --- | --- |
| * 1 | * -1.34 | * -0.141 |

* **De\_tqwt\_TKEO\_mean\_dec\_17**:

| * tqwt\_TKEO\_mean\_dec\_17 | * tqwt\_minValue\_dec\_17 |
| --- | --- |
| * 1 | * 2.32 |

* **De\_tqwt\_energy\_dec\_33**:

| * tqwt\_energy\_dec\_31 | * tqwt\_energy\_dec\_33 |
| --- | --- |
| * -0.907 | * 1 |

* **De\_std\_MFCC\_2nd\_coef**:

| * std\_MFCC\_2nd\_coef | * std\_2nd\_delta |
| --- | --- |
| * 1 | * -0.828 |

* **De\_tqwt\_kurtosisValue\_dec\_33**:

| * tqwt\_kurtosisValue\_dec\_32 | * tqwt\_kurtosisValue\_dec\_33 |
| --- | --- |
| * -0.883 | * 1 |

* **De\_tqwt\_maxValue\_dec\_28**:

| * tqwt\_stdValue\_dec\_28 | * tqwt\_maxValue\_dec\_28 |
| --- | --- |
| * -0.703 | * 1 |

* **De\_tqwt\_entropy\_log\_dec\_1**:

| * tqwt\_entropy\_log\_dec\_1 | * tqwt\_entropy\_log\_dec\_2 |
| --- | --- |
| * 1 | * -0.849 |

* **De\_locDbShimmer**:

| * locDbShimmer | * ddaShimmer |
| --- | --- |
| * 1 | * -0.963 |

* **De\_tqwt\_TKEO\_std\_dec\_17**:

| * tqwt\_TKEO\_std\_dec\_17 | * tqwt\_minValue\_dec\_17 |
| --- | --- |
| * 1 | * 2.13 |

* **De\_tqwt\_entropy\_shannon\_dec\_17**:

| * tqwt\_entropy\_shannon\_dec\_17 | * tqwt\_minValue\_dec\_17 |
| --- | --- |
| * 1 | * 1.62 |

* **De\_tqwt\_maxValue\_dec\_29**:

| * tqwt\_stdValue\_dec\_28 | * tqwt\_maxValue\_dec\_29 |
| --- | --- |
| * -0.555 | * 1 |

* **De\_tqwt\_minValue\_dec\_11**:

| * tqwt\_minValue\_dec\_11 | * tqwt\_minValue\_dec\_17 |
| --- | --- |
| * 1 | * -0.918 |

* **De\_tqwt\_TKEO\_std\_dec\_36**:

| * tqwt\_TKEO\_mean\_dec\_35 | * tqwt\_TKEO\_std\_dec\_36 |
| --- | --- |
| * -0.926 | * 1 |

* **De\_tqwt\_entropy\_log\_dec\_11**:

| * tqwt\_entropy\_log\_dec\_9 | * tqwt\_entropy\_log\_dec\_11 |
| --- | --- |
| * -0.97 | * 1 |

* **De\_det\_LT\_TKEO\_mean\_3\_coef**:

| * Ed2\_3\_coef | * det\_LT\_TKEO\_mean\_1\_coef | * det\_LT\_TKEO\_mean\_3\_coef |
| --- | --- | --- |
| * -0.89 | * -0.0834 | * 1 |

* **De\_tqwt\_TKEO\_std\_dec\_11**:

| * tqwt\_TKEO\_std\_dec\_11 | * tqwt\_stdValue\_dec\_10 | * tqwt\_minValue\_dec\_17 |
| --- | --- | --- |
| * 1 | * -1.64 | * 0.314 |

* **De\_tqwt\_TKEO\_mean\_dec\_33**:

| * tqwt\_TKEO\_mean\_dec\_33 | * tqwt\_TKEO\_mean\_dec\_35 | * tqwt\_stdValue\_dec\_32 |
| --- | --- | --- |
| * 1 | * -0.248 | * -1.49 |

* **De\_tqwt\_TKEO\_std\_dec\_7**:

| * tqwt\_TKEO\_std\_dec\_7 | * tqwt\_stdValue\_dec\_7 | * tqwt\_minValue\_dec\_17 |
| --- | --- | --- |
| * 1 | * -1.66 | * 0.145 |

* **De\_tqwt\_minValue\_dec\_7**:

| * tqwt\_minValue\_dec\_7 | * tqwt\_minValue\_dec\_17 | * tqwt\_maxValue\_dec\_8 |
| --- | --- | --- |
| * 1 | * -0.0186 | * 0.967 |

* **De\_tqwt\_entropy\_log\_dec\_35**:

| * tqwt\_entropy\_log\_dec\_35 | * tqwt\_TKEO\_mean\_dec\_35 |
| --- | --- |
| * 1 | * -0.106 |

names(selectedlistU) <- NULL  
### 2b Get the the names of the original features  
  
allDevarU <- unique(c(names(unlist(selectedlistU)),decornamesU))  
allDevarU <- allDevarU[!str\_detect(allDevarU,"De\_")]  
allDevarU <- str\_remove(allDevarU,"Ba\_")  
allDevarU <- unique(allDevarU)  
  
pander::pander(c(length(rawnames),length(decornames),length(decornamesU)))

*59*, *29* and *42*

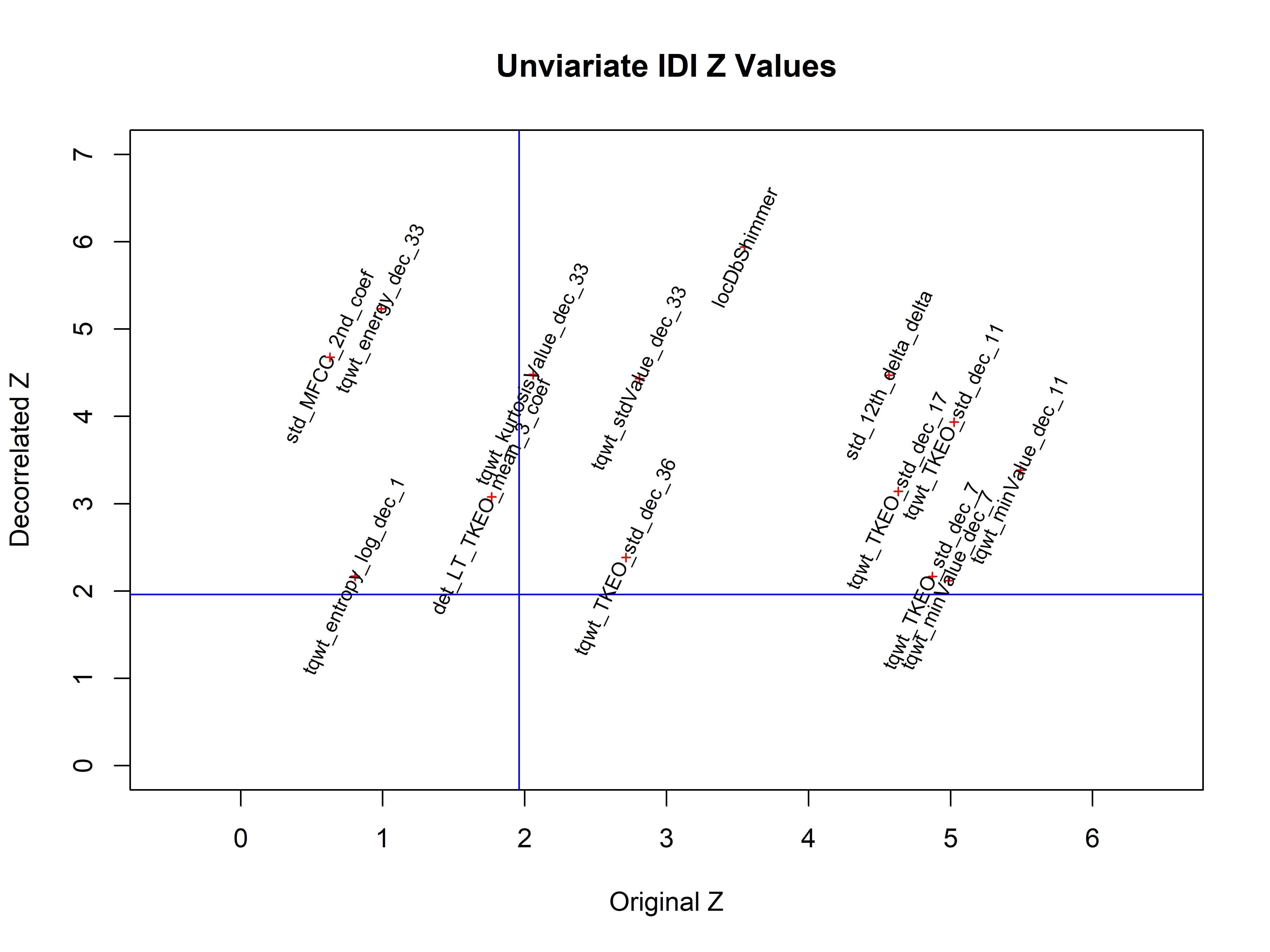
pander::pander(c(length(rawnames),length(allDevar),length(allDevarU)))

*59*, *43* and *53*

### 2c Get only the new feautres not found in the original analysis  
dvar <- allDevar[!(allDevar %in% rawnames)]   
  
### 2d Get the decorrelated variables that have new features  
newvars <- character();  
for (cvar in deNames\_in\_dc)  
{  
 lvar <- dc[cvar]  
 names(lvar) <- NULL  
 lvar <- names(unlist(lvar))  
 if (length(lvar[lvar %in% dvar]) > 0)  
 {  
 newvars <- append(newvars,cvar)  
 }  
}  
  
## 3 Here is the univariate z values of the orignal set  
#pander::pander(bm$univariate[dvar,])  
## 4 Here is the univariate z values of the decorrelated set  
#pander::pander(bmd$univariate[newvars,])  
  
## 4a The scater plot of the decorrelated vs original Univariate values  
  
zvalueNew <- bmd$univariate[newvars,]  
rownames(zvalueNew) <- str\_remove(rownames(zvalueNew),"De\_")  
rownames(zvalueNew) <- str\_remove(rownames(zvalueNew),"Ba\_")  
  
zvaluePrePost <- bm$univariate[rownames(zvalueNew),c(1,3)]  
zvaluePrePost$Name <- NULL  
zvaluePrePost$NewZ <- zvalueNew[rownames(zvaluePrePost),"ZUni"]  
pander::pander(zvaluePrePost)

|  | ZUni | NewZ |
| --- | --- | --- |
| **tqwt\_energy\_dec\_33** | 0.990 | 5.23 |
| **tqwt\_stdValue\_dec\_33** | 2.810 | 4.43 |
| **locDbShimmer** | 3.548 | 5.94 |
| **std\_MFCC\_2nd\_coef** | 0.628 | 4.68 |
| **tqwt\_TKEO\_std\_dec\_11** | 5.025 | 3.94 |
| **tqwt\_minValue\_dec\_11** | 5.493 | 3.38 |
| **std\_12th\_delta\_delta** | 4.566 | 4.47 |
| **tqwt\_minValue\_dec\_7** | 4.986 | 2.12 |
| **tqwt\_kurtosisValue\_dec\_33** | 2.060 | 4.48 |
| **tqwt\_TKEO\_std\_dec\_36** | 2.714 | 2.38 |
| **tqwt\_TKEO\_std\_dec\_7** | 4.872 | 2.17 |
| **det\_LT\_TKEO\_mean\_3\_coef** | 1.767 | 3.08 |
| **tqwt\_entropy\_log\_dec\_1** | 0.805 | 2.17 |
| **tqwt\_TKEO\_std\_dec\_17** | 4.631 | 3.14 |

plot(zvaluePrePost,  
 xlim=c(-0.5,6.5),  
 ylim=c(0,7),  
 xlab="Original Z",  
 ylab="Decorrelated Z",  
 main="Unviariate IDI Z Values",  
 pch=3,cex=0.5,  
 col="red")  
abline(v=1.96,col="blue")  
abline(h=1.96,col="blue")  
text(zvaluePrePost$ZUni,zvaluePrePost$NewZ,rownames(zvaluePrePost),srt=65,cex=0.75)



### The Summary of the Decorrelated-based Model

Here I will print the summary statistics of the Logistic models found by BSWiMS, using the original and transformed dataset. After that, I will show the characteristics of the features not found by the original analysis.

pander::pander(smOriginal$coefficients)

|  | Estimate | lower | OR | upper | u.Accuracy | r.Accuracy | full.Accuracy | u.AUC | r.AUC | full.AUC | IDI | NRI | z.IDI | z.NRI | Frequency |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **tqwt\_entropy\_shannon\_dec\_16** | -0.026241 | 0.968 | 0.974 | 0.980 | 0.676 | 0.720 | 0.684 | 0.735 | 0.552 | 0.736 | 0.1558 | 0.842 | 6.66 | 7.69 | 0.50 |
| **minIntensity** | -0.774840 | 0.371 | 0.461 | 0.573 | 0.609 | 0.688 | 0.699 | 0.683 | 0.676 | 0.735 | 0.1460 | 0.708 | 6.44 | 6.28 | 0.95 |
| **std\_delta\_delta\_log\_energy** | 0.116023 | 1.083 | 1.123 | 1.164 | 0.715 | 0.715 | 0.756 | 0.724 | 0.725 | 0.770 | 0.1354 | 0.694 | 6.11 | 5.87 | 1.00 |
| **std\_6th\_delta\_delta** | 0.299966 | 1.232 | 1.350 | 1.478 | 0.670 | 0.702 | 0.752 | 0.701 | 0.715 | 0.777 | 0.1337 | 0.695 | 6.10 | 5.86 | 1.00 |
| **tqwt\_TKEO\_mean\_dec\_11** | -0.010087 | 0.987 | 0.990 | 0.993 | 0.646 | 0.632 | 0.711 | 0.648 | 0.648 | 0.737 | 0.1329 | 0.650 | 6.09 | 5.50 | 0.30 |
| **tqwt\_stdValue\_dec\_11** | -0.011422 | 0.985 | 0.989 | 0.992 | 0.671 | 0.650 | 0.720 | 0.681 | 0.650 | 0.733 | 0.1280 | 0.683 | 5.91 | 5.77 | 0.20 |
| **std\_11th\_delta\_delta** | 0.263474 | 1.210 | 1.301 | 1.400 | 0.688 | 0.658 | 0.732 | 0.707 | 0.668 | 0.747 | 0.1282 | 0.637 | 5.87 | 5.40 | 0.90 |
| **std\_delta\_log\_energy** | 0.118481 | 1.085 | 1.126 | 1.168 | 0.709 | 0.722 | 0.745 | 0.715 | 0.728 | 0.762 | 0.1276 | 0.634 | 5.85 | 5.31 | 1.00 |
| **tqwt\_TKEO\_std\_dec\_11** | -0.021164 | 0.973 | 0.979 | 0.985 | 0.664 | 0.662 | 0.727 | 0.677 | 0.659 | 0.751 | 0.1241 | 0.587 | 5.85 | 4.91 | 0.65 |
| **std\_7th\_delta\_delta** | 0.252877 | 1.192 | 1.288 | 1.391 | 0.688 | 0.712 | 0.725 | 0.717 | 0.704 | 0.740 | 0.1276 | 0.725 | 5.82 | 6.21 | 1.00 |
| **std\_Log\_energy** | 0.122681 | 1.092 | 1.131 | 1.170 | 0.670 | 0.719 | 0.705 | 0.691 | 0.651 | 0.717 | 0.1291 | 0.569 | 5.72 | 4.85 | 0.95 |
| **tqwt\_maxValue\_dec\_11** | -0.074605 | 0.906 | 0.928 | 0.951 | 0.692 | 0.687 | 0.746 | 0.705 | 0.709 | 0.773 | 0.1105 | 0.559 | 5.67 | 4.63 | 1.00 |
| **std\_8th\_delta\_delta** | 0.284006 | 1.218 | 1.328 | 1.449 | 0.689 | 0.705 | 0.733 | 0.712 | 0.705 | 0.754 | 0.1244 | 0.722 | 5.65 | 6.16 | 1.00 |
| **std\_6th\_delta** | 0.230671 | 1.161 | 1.259 | 1.366 | 0.636 | 0.721 | 0.731 | 0.665 | 0.717 | 0.751 | 0.1161 | 0.618 | 5.51 | 5.14 | 1.00 |
| **tqwt\_entropy\_log\_dec\_16** | -0.116420 | 0.859 | 0.890 | 0.922 | 0.688 | 0.673 | 0.705 | 0.697 | 0.645 | 0.729 | 0.1120 | 0.657 | 5.49 | 5.53 | 0.95 |
| **std\_7th\_delta** | 0.214150 | 1.156 | 1.239 | 1.327 | 0.674 | 0.695 | 0.706 | 0.708 | 0.700 | 0.729 | 0.1164 | 0.790 | 5.45 | 6.85 | 1.00 |
| **tqwt\_TKEO\_std\_dec\_12** | -0.023313 | 0.969 | 0.977 | 0.985 | 0.729 | 0.669 | 0.728 | 0.722 | 0.696 | 0.750 | 0.1099 | 0.707 | 5.44 | 6.08 | 1.00 |
| **std\_MFCC\_6th\_coef** | 0.142335 | 1.096 | 1.153 | 1.213 | 0.578 | 0.665 | 0.731 | 0.603 | 0.678 | 0.756 | 0.1075 | 0.536 | 5.38 | 4.42 | 0.50 |
| **tqwt\_TKEO\_mean\_dec\_12** | -0.020456 | 0.973 | 0.980 | 0.986 | 0.728 | 0.680 | 0.718 | 0.707 | 0.679 | 0.728 | 0.1077 | 0.674 | 5.35 | 5.69 | 1.00 |
| **tqwt\_maxValue\_dec\_7** | -0.038438 | 0.949 | 0.962 | 0.975 | 0.639 | 0.670 | 0.686 | 0.680 | 0.665 | 0.710 | 0.1069 | 0.670 | 5.34 | 5.68 | 0.75 |
| **tqwt\_maxValue\_dec\_12** | -0.060547 | 0.921 | 0.941 | 0.961 | 0.716 | 0.701 | 0.742 | 0.734 | 0.716 | 0.764 | 0.0991 | 0.618 | 5.34 | 5.20 | 1.00 |
| **tqwt\_maxValue\_dec\_13** | -0.031601 | 0.959 | 0.969 | 0.979 | 0.680 | 0.680 | 0.697 | 0.714 | 0.645 | 0.725 | 0.1033 | 0.620 | 5.29 | 5.26 | 0.70 |
| **tqwt\_TKEO\_std\_dec\_7** | -0.015262 | 0.979 | 0.985 | 0.990 | 0.646 | 0.675 | 0.716 | 0.678 | 0.677 | 0.743 | 0.1054 | 0.578 | 5.26 | 4.79 | 0.60 |
| **tqwt\_maxValue\_dec\_1** | -0.006163 | 0.992 | 0.994 | 0.996 | 0.632 | 0.715 | 0.653 | 0.672 | 0.587 | 0.677 | 0.1157 | 0.548 | 5.25 | 4.55 | 0.10 |
| **tqwt\_kurtosisValue\_dec\_28** | -0.012651 | 0.983 | 0.987 | 0.992 | 0.680 | 0.672 | 0.720 | 0.673 | 0.678 | 0.734 | 0.1026 | 0.611 | 5.24 | 5.09 | 0.40 |
| **std\_8th\_delta** | 0.217709 | 1.150 | 1.243 | 1.344 | 0.664 | 0.716 | 0.723 | 0.684 | 0.708 | 0.737 | 0.1111 | 0.658 | 5.23 | 5.51 | 1.00 |
| **tqwt\_entropy\_log\_dec\_13** | -0.133759 | 0.833 | 0.875 | 0.919 | 0.712 | 0.660 | 0.716 | 0.714 | 0.688 | 0.736 | 0.0995 | 0.696 | 5.22 | 5.93 | 1.00 |
| **std\_9th\_delta** | 0.165252 | 1.112 | 1.180 | 1.252 | 0.665 | 0.690 | 0.710 | 0.705 | 0.697 | 0.732 | 0.1066 | 0.663 | 5.20 | 5.62 | 0.90 |
| **std\_9th\_delta\_delta** | 0.198143 | 1.133 | 1.219 | 1.312 | 0.689 | 0.713 | 0.727 | 0.719 | 0.715 | 0.742 | 0.1049 | 0.651 | 5.18 | 5.47 | 1.00 |
| **tqwt\_kurtosisValue\_dec\_18** | 0.143848 | 1.101 | 1.155 | 1.212 | 0.652 | 0.644 | 0.690 | 0.677 | 0.683 | 0.711 | 0.1045 | 0.655 | 5.18 | 5.49 | 0.60 |
| **tqwt\_energy\_dec\_5** | -0.005528 | 0.992 | 0.994 | 0.997 | 0.575 | 0.693 | 0.756 | 0.596 | 0.704 | 0.773 | 0.0960 | 0.643 | 5.16 | 5.35 | 0.10 |
| **std\_10th\_delta\_delta** | 0.173090 | 1.115 | 1.189 | 1.267 | 0.678 | 0.692 | 0.713 | 0.691 | 0.696 | 0.729 | 0.1042 | 0.592 | 5.15 | 4.95 | 1.00 |
| **tqwt\_entropy\_log\_dec\_11** | -0.208819 | 0.755 | 0.812 | 0.873 | 0.696 | 0.665 | 0.716 | 0.693 | 0.693 | 0.737 | 0.0966 | 0.686 | 5.15 | 5.82 | 1.00 |
| **tqwt\_energy\_dec\_6** | -0.011565 | 0.984 | 0.989 | 0.993 | 0.599 | 0.694 | 0.742 | 0.629 | 0.712 | 0.751 | 0.0995 | 0.665 | 5.14 | 5.62 | 0.25 |
| **meanIntensity** | -0.085067 | 0.890 | 0.918 | 0.948 | 0.613 | 0.674 | 0.715 | 0.692 | 0.668 | 0.724 | 0.0951 | 0.657 | 5.12 | 5.80 | 0.15 |
| **tqwt\_entropy\_shannon\_dec\_12** | -0.017886 | 0.976 | 0.982 | 0.989 | 0.724 | 0.668 | 0.714 | 0.719 | 0.695 | 0.728 | 0.0970 | 0.678 | 5.06 | 5.71 | 0.90 |
| **tqwt\_kurtosisValue\_dec\_34** | 0.013536 | 1.008 | 1.014 | 1.019 | 0.622 | 0.697 | 0.760 | 0.641 | 0.702 | 0.766 | 0.0876 | 0.617 | 5.03 | 5.14 | 0.20 |
| **tqwt\_energy\_dec\_12** | -0.018327 | 0.975 | 0.982 | 0.989 | 0.675 | 0.631 | 0.712 | 0.675 | 0.684 | 0.729 | 0.0890 | 0.617 | 5.02 | 5.12 | 0.75 |
| **tqwt\_entropy\_log\_dec\_35** | -0.013035 | 0.982 | 0.987 | 0.992 | 0.661 | 0.664 | 0.678 | 0.644 | 0.600 | 0.682 | 0.0940 | 0.611 | 5.02 | 5.05 | 0.15 |
| **tqwt\_entropy\_log\_dec\_12** | -0.154394 | 0.807 | 0.857 | 0.910 | 0.742 | 0.698 | 0.747 | 0.731 | 0.713 | 0.758 | 0.0891 | 0.669 | 4.93 | 5.65 | 1.00 |
| **std\_5th\_delta** | 0.116534 | 1.073 | 1.124 | 1.176 | 0.622 | 0.674 | 0.705 | 0.654 | 0.681 | 0.733 | 0.0933 | 0.483 | 4.88 | 3.94 | 0.70 |
| **std\_MFCC\_8th\_coef** | 0.133751 | 1.084 | 1.143 | 1.205 | 0.657 | 0.670 | 0.722 | 0.675 | 0.693 | 0.742 | 0.0912 | 0.628 | 4.85 | 5.24 | 0.60 |
| **tqwt\_entropy\_shannon\_dec\_13** | -0.020386 | 0.973 | 0.980 | 0.987 | 0.689 | 0.673 | 0.703 | 0.714 | 0.687 | 0.730 | 0.0890 | 0.624 | 4.84 | 5.23 | 0.95 |
| **tqwt\_kurtosisValue\_dec\_26** | -0.037243 | 0.949 | 0.963 | 0.978 | 0.762 | 0.656 | 0.720 | 0.663 | 0.690 | 0.725 | 0.0891 | 0.563 | 4.81 | 5.04 | 0.95 |
| **tqwt\_kurtosisValue\_dec\_20** | 0.193267 | 1.123 | 1.213 | 1.310 | 0.671 | 0.663 | 0.718 | 0.689 | 0.702 | 0.742 | 0.0878 | 0.679 | 4.76 | 5.72 | 1.00 |
| **tqwt\_kurtosisValue\_dec\_33** | 0.009416 | 1.006 | 1.009 | 1.013 | 0.531 | 0.713 | 0.722 | 0.586 | 0.731 | 0.724 | 0.0844 | 0.618 | 4.74 | 5.17 | 0.10 |
| **tqwt\_TKEO\_std\_dec\_16** | -0.002452 | 0.997 | 0.998 | 0.999 | 0.651 | 0.633 | 0.674 | 0.710 | 0.659 | 0.723 | 0.0826 | 0.613 | 4.72 | 5.17 | 0.15 |
| **tqwt\_kurtosisValue\_dec\_36** | 0.037035 | 1.022 | 1.038 | 1.054 | 0.678 | 0.711 | 0.758 | 0.708 | 0.726 | 0.768 | 0.0839 | 0.727 | 4.62 | 6.21 | 1.00 |
| **tqwt\_entropy\_log\_dec\_34** | -0.020112 | 0.972 | 0.980 | 0.988 | 0.674 | 0.687 | 0.726 | 0.665 | 0.698 | 0.742 | 0.0825 | 0.608 | 4.62 | 5.05 | 0.15 |
| **tqwt\_kurtosisValue\_dec\_27** | -0.005812 | 0.992 | 0.994 | 0.997 | 0.691 | 0.654 | 0.722 | 0.662 | 0.686 | 0.717 | 0.0803 | 0.520 | 4.57 | 4.27 | 0.25 |
| **std\_12th\_delta** | 0.010771 | 1.006 | 1.011 | 1.015 | 0.650 | 0.663 | 0.688 | 0.683 | 0.672 | 0.702 | 0.0776 | 0.541 | 4.57 | 4.52 | 0.10 |
| **IMF\_NSR\_TKEO** | -0.059900 | 0.918 | 0.942 | 0.966 | 0.623 | 0.721 | 0.769 | 0.593 | 0.736 | 0.784 | 0.0721 | 0.442 | 4.50 | 3.61 | 0.60 |
| **std\_10th\_delta** | 0.084396 | 1.049 | 1.088 | 1.128 | 0.657 | 0.688 | 0.706 | 0.681 | 0.704 | 0.733 | 0.0798 | 0.395 | 4.50 | 3.18 | 0.65 |
| **tqwt\_meanValue\_dec\_11** | 0.000318 | 1.000 | 1.000 | 1.000 | 0.613 | 0.699 | 0.724 | 0.653 | 0.707 | 0.738 | 0.0728 | 0.620 | 4.39 | 5.21 | 0.20 |
| **tqwt\_kurtosisValue\_dec\_35** | 0.053102 | 1.030 | 1.055 | 1.079 | 0.643 | 0.688 | 0.731 | 0.667 | 0.706 | 0.745 | 0.0731 | 0.585 | 4.35 | 4.86 | 1.00 |
| **mean\_MFCC\_2nd\_coef** | 0.006274 | 1.003 | 1.006 | 1.009 | 0.666 | 0.729 | 0.758 | 0.643 | 0.744 | 0.760 | 0.0632 | 0.485 | 4.25 | 3.99 | 0.15 |
| **std\_12th\_delta\_delta** | 0.010420 | 1.006 | 1.010 | 1.015 | 0.657 | 0.637 | 0.674 | 0.675 | 0.661 | 0.704 | 0.0681 | 0.562 | 4.20 | 4.60 | 0.10 |
| **tqwt\_kurtosisValue\_dec\_1** | -0.004130 | 0.994 | 0.996 | 0.998 | 0.622 | 0.756 | 0.789 | 0.619 | 0.766 | 0.792 | 0.0537 | 0.464 | 4.05 | 3.77 | 0.20 |
| **std\_4th\_delta** | 0.016366 | 1.008 | 1.017 | 1.025 | 0.599 | 0.664 | 0.675 | 0.622 | 0.708 | 0.724 | 0.0626 | 0.419 | 3.84 | 3.41 | 0.15 |

pander::pander(smDecor$coefficients)

|  | Estimate | lower | OR | upper | u.Accuracy | r.Accuracy | full.Accuracy | u.AUC | r.AUC | full.AUC | IDI | NRI | z.IDI | z.NRI | Frequency |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ba\_tqwt\_kurtosisValue\_dec\_36** | 0.14505 | 1.116 | 1.156 | 1.198 | 0.678 | 0.687 | 0.763 | 0.706 | 0.678 | 0.758 | 0.1586 | 0.831 | 6.87 | 7.34 | 0.95 |
| **Ba\_tqwt\_entropy\_log\_dec\_12** | -0.64945 | 0.444 | 0.522 | 0.614 | 0.741 | 0.703 | 0.765 | 0.730 | 0.688 | 0.770 | 0.1488 | 0.871 | 6.49 | 7.72 | 0.90 |
| **Ba\_tqwt\_kurtosisValue\_dec\_28** | -0.03080 | 0.961 | 0.970 | 0.978 | 0.678 | 0.706 | 0.750 | 0.672 | 0.719 | 0.758 | 0.1440 | 0.823 | 6.48 | 7.15 | 0.20 |
| **Ba\_std\_delta\_delta\_log\_energy** | 0.43367 | 1.361 | 1.543 | 1.749 | 0.717 | 0.767 | 0.824 | 0.724 | 0.758 | 0.823 | 0.1397 | 0.795 | 6.33 | 6.88 | 1.00 |
| **De\_tqwt\_energy\_dec\_33** | -0.21257 | 0.760 | 0.809 | 0.860 | 0.747 | 0.719 | 0.781 | 0.736 | 0.739 | 0.779 | 0.1437 | 0.802 | 6.26 | 6.94 | 1.00 |
| **Ba\_tqwt\_energy\_dec\_12** | -0.06308 | 0.922 | 0.939 | 0.956 | 0.672 | 0.679 | 0.757 | 0.673 | 0.702 | 0.776 | 0.1211 | 0.644 | 5.95 | 5.41 | 0.60 |
| **De\_tqwt\_TKEO\_mean\_dec\_11** | -0.02999 | 0.962 | 0.970 | 0.979 | 0.635 | 0.648 | 0.725 | 0.645 | 0.661 | 0.749 | 0.1280 | 0.605 | 5.94 | 5.02 | 0.45 |
| **De\_tqwt\_stdValue\_dec\_33** | -0.92200 | 0.296 | 0.398 | 0.534 | 0.680 | 0.693 | 0.777 | 0.683 | 0.717 | 0.767 | 0.1142 | 0.658 | 5.85 | 5.53 | 0.95 |
| **Ba\_locAbsJitter** | 0.01770 | 1.012 | 1.018 | 1.024 | 0.638 | 0.641 | 0.720 | 0.655 | 0.658 | 0.732 | 0.1147 | 0.602 | 5.74 | 4.97 | 0.15 |
| **std\_MFCC\_8th\_coef** | 0.54194 | 1.431 | 1.719 | 2.066 | 0.656 | 0.686 | 0.743 | 0.675 | 0.687 | 0.756 | 0.1161 | 0.669 | 5.67 | 5.62 | 0.75 |
| **Ba\_tqwt\_kurtosisValue\_dec\_26** | -0.02685 | 0.965 | 0.974 | 0.982 | 0.763 | 0.680 | 0.761 | 0.663 | 0.695 | 0.761 | 0.1166 | 0.657 | 5.67 | 6.05 | 0.25 |
| **De\_locDbShimmer** | 1.08762 | 2.098 | 2.967 | 4.197 | 0.714 | 0.744 | 0.779 | 0.736 | 0.734 | 0.777 | 0.0977 | 0.788 | 5.66 | 6.86 | 1.00 |
| **De\_std\_MFCC\_2nd\_coef** | -1.37310 | 0.157 | 0.253 | 0.408 | 0.703 | 0.787 | 0.827 | 0.685 | 0.781 | 0.826 | 0.1115 | 0.691 | 5.57 | 5.85 | 1.00 |
| **Ba\_std\_5th\_delta** | 0.49671 | 1.385 | 1.643 | 1.949 | 0.623 | 0.739 | 0.758 | 0.658 | 0.732 | 0.770 | 0.1117 | 0.588 | 5.40 | 4.94 | 0.70 |
| **De\_tqwt\_TKEO\_std\_dec\_11** | -0.04899 | 0.936 | 0.952 | 0.969 | 0.699 | 0.692 | 0.767 | 0.701 | 0.691 | 0.772 | 0.1031 | 0.757 | 5.35 | 6.46 | 0.10 |
| **De\_tqwt\_minValue\_dec\_11** | 0.02668 | 1.017 | 1.027 | 1.037 | 0.637 | 0.706 | 0.748 | 0.609 | 0.721 | 0.758 | 0.0994 | 0.592 | 5.32 | 4.93 | 0.10 |
| **De\_std\_12th\_delta\_delta** | 0.67639 | 1.538 | 1.967 | 2.515 | 0.694 | 0.724 | 0.756 | 0.675 | 0.730 | 0.761 | 0.1078 | 0.643 | 5.31 | 5.36 | 0.70 |
| **De\_tqwt\_minValue\_dec\_7** | 0.19404 | 1.130 | 1.214 | 1.305 | 0.563 | 0.695 | 0.746 | 0.591 | 0.709 | 0.763 | 0.0969 | 0.689 | 5.14 | 5.83 | 0.25 |
| **De\_tqwt\_kurtosisValue\_dec\_33** | 0.63238 | 1.493 | 1.882 | 2.373 | 0.694 | 0.780 | 0.827 | 0.679 | 0.773 | 0.826 | 0.0868 | 0.541 | 5.14 | 4.49 | 1.00 |
| **Ba\_tqwt\_kurtosisValue\_dec\_17** | 0.06951 | 1.044 | 1.072 | 1.100 | 0.629 | 0.752 | 0.774 | 0.675 | 0.755 | 0.786 | 0.0942 | 0.605 | 5.12 | 5.02 | 0.10 |
| **tqwt\_kurtosisValue\_dec\_20** | 0.53296 | 1.409 | 1.704 | 2.060 | 0.668 | 0.718 | 0.766 | 0.688 | 0.727 | 0.777 | 0.0957 | 0.743 | 5.07 | 6.36 | 0.80 |
| **Ba\_meanIntensity** | -1.19764 | 0.191 | 0.302 | 0.478 | 0.608 | 0.718 | 0.753 | 0.689 | 0.724 | 0.779 | 0.0895 | 0.628 | 5.00 | 5.44 | 0.50 |
| **De\_std\_MFCC\_6th\_coef** | 0.25653 | 1.170 | 1.292 | 1.428 | 0.578 | 0.696 | 0.743 | 0.601 | 0.713 | 0.765 | 0.0952 | 0.550 | 4.92 | 4.57 | 0.40 |
| **De\_tqwt\_TKEO\_std\_dec\_36** | 0.13162 | 1.083 | 1.141 | 1.202 | 0.640 | 0.704 | 0.753 | 0.609 | 0.723 | 0.767 | 0.0770 | 0.576 | 4.82 | 4.80 | 0.65 |
| **De\_tqwt\_TKEO\_std\_dec\_7** | -0.11331 | 0.852 | 0.893 | 0.936 | 0.595 | 0.717 | 0.747 | 0.596 | 0.745 | 0.770 | 0.0839 | 0.644 | 4.66 | 5.38 | 0.25 |
| **tqwt\_meanValue\_dec\_11** | 0.00131 | 1.001 | 1.001 | 1.002 | 0.613 | 0.728 | 0.763 | 0.653 | 0.721 | 0.767 | 0.0688 | 0.615 | 4.24 | 5.19 | 0.30 |
| **De\_det\_LT\_TKEO\_mean\_3\_coef** | -0.18463 | 0.762 | 0.831 | 0.908 | 0.610 | 0.721 | 0.742 | 0.621 | 0.726 | 0.750 | 0.0610 | 0.585 | 4.05 | 4.85 | 0.35 |
| **De\_tqwt\_entropy\_log\_dec\_1** | 0.83992 | 1.528 | 2.316 | 3.512 | 0.607 | 0.751 | 0.770 | 0.631 | 0.763 | 0.779 | 0.0583 | 0.522 | 3.88 | 4.30 | 0.25 |
| **De\_tqwt\_TKEO\_std\_dec\_17** | -0.06260 | 0.910 | 0.939 | 0.970 | 0.623 | 0.696 | 0.745 | 0.624 | 0.720 | 0.759 | 0.0504 | 0.495 | 3.81 | 4.03 | 0.10 |

pander::pander(smDecorU$coefficients)

|  | Estimate | lower | OR | upper | u.Accuracy | r.Accuracy | full.Accuracy | u.AUC | r.AUC | full.AUC | IDI | NRI | z.IDI | z.NRI | Frequency |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **tqwt\_kurtosisValue\_dec\_36** | 0.15140 | 1.126 | 1.163 | 1.202 | 0.679 | 0.672 | 0.759 | 0.709 | 0.686 | 0.766 | 0.1734 | 0.871 | 7.20 | 7.76 | 1.00 |
| **De\_tqwt\_TKEO\_std\_dec\_27** | -0.04885 | 0.939 | 0.952 | 0.966 | 0.691 | 0.648 | 0.718 | 0.687 | 0.674 | 0.724 | 0.1407 | 0.831 | 6.35 | 7.19 | 0.25 |
| **Ba\_std\_5th\_delta** | 0.43854 | 1.380 | 1.550 | 1.743 | 0.621 | 0.692 | 0.733 | 0.659 | 0.692 | 0.739 | 0.1462 | 0.742 | 6.34 | 6.36 | 0.65 |
| **Ba\_std\_delta\_log\_energy** | 0.33390 | 1.264 | 1.396 | 1.543 | 0.709 | 0.754 | 0.814 | 0.713 | 0.748 | 0.812 | 0.1364 | 0.691 | 6.22 | 5.83 | 1.00 |
| **De\_std\_12th\_delta\_delta** | 0.40227 | 1.324 | 1.495 | 1.688 | 0.657 | 0.708 | 0.741 | 0.672 | 0.707 | 0.752 | 0.1293 | 0.743 | 5.91 | 6.38 | 0.65 |
| **De\_tqwt\_TKEO\_mean\_dec\_17** | -0.17250 | 0.794 | 0.842 | 0.892 | 0.643 | 0.679 | 0.723 | 0.669 | 0.695 | 0.727 | 0.1205 | 0.803 | 5.76 | 6.96 | 0.35 |
| **De\_tqwt\_stdValue\_dec\_12** | -0.07985 | 0.904 | 0.923 | 0.943 | 0.729 | 0.760 | 0.773 | 0.721 | 0.676 | 0.768 | 0.1223 | 0.668 | 5.68 | 5.66 | 0.75 |
| **Ba\_tqwt\_energy\_dec\_12** | -0.04536 | 0.941 | 0.956 | 0.970 | 0.672 | 0.665 | 0.738 | 0.674 | 0.692 | 0.749 | 0.1128 | 0.628 | 5.64 | 5.24 | 0.60 |
| **De\_std\_MFCC\_6th\_coef** | 0.25436 | 1.187 | 1.290 | 1.402 | 0.576 | 0.703 | 0.742 | 0.602 | 0.714 | 0.751 | 0.1130 | 0.643 | 5.60 | 5.47 | 0.35 |
| **Ba\_tqwt\_kurtosisValue\_dec\_28** | -0.07042 | 0.910 | 0.932 | 0.954 | 0.680 | 0.705 | 0.753 | 0.673 | 0.713 | 0.763 | 0.1132 | 0.689 | 5.52 | 5.82 | 0.80 |
| **De\_tqwt\_energy\_dec\_33** | -0.14409 | 0.823 | 0.866 | 0.910 | 0.748 | 0.734 | 0.793 | 0.734 | 0.744 | 0.787 | 0.1113 | 0.685 | 5.47 | 5.80 | 1.00 |
| **tqwt\_kurtosisValue\_dec\_20** | 0.53899 | 1.420 | 1.714 | 2.069 | 0.669 | 0.702 | 0.758 | 0.688 | 0.718 | 0.777 | 0.1072 | 0.723 | 5.46 | 6.20 | 1.00 |
| **tqwt\_kurtosisValue\_dec\_26** | -0.04392 | 0.943 | 0.957 | 0.971 | 0.761 | 0.692 | 0.751 | 0.662 | 0.706 | 0.755 | 0.1104 | 0.585 | 5.43 | 5.31 | 0.35 |
| **Ba\_meanIntensity** | -1.31455 | 0.170 | 0.269 | 0.424 | 0.608 | 0.693 | 0.739 | 0.689 | 0.699 | 0.764 | 0.1056 | 0.654 | 5.39 | 5.77 | 0.75 |
| **mean\_MFCC\_2nd\_coef** | 0.01693 | 1.011 | 1.017 | 1.023 | 0.674 | 0.696 | 0.718 | 0.648 | 0.631 | 0.714 | 0.1090 | 0.656 | 5.36 | 5.51 | 0.15 |
| **std\_MFCC\_8th\_coef** | 0.48676 | 1.372 | 1.627 | 1.930 | 0.656 | 0.715 | 0.754 | 0.673 | 0.719 | 0.757 | 0.1084 | 0.670 | 5.35 | 5.64 | 0.80 |
| **Ba\_tqwt\_kurtosisValue\_dec\_17** | 0.20212 | 1.137 | 1.224 | 1.317 | 0.633 | 0.709 | 0.746 | 0.675 | 0.718 | 0.764 | 0.1022 | 0.665 | 5.31 | 5.62 | 0.40 |
| **De\_std\_MFCC\_2nd\_coef** | -0.97463 | 0.265 | 0.377 | 0.537 | 0.703 | 0.764 | 0.816 | 0.686 | 0.761 | 0.811 | 0.1078 | 0.684 | 5.31 | 5.77 | 1.00 |
| **Ba\_tqwt\_minValue\_dec\_17** | 0.01876 | 1.012 | 1.019 | 1.026 | 0.600 | 0.696 | 0.729 | 0.660 | 0.681 | 0.748 | 0.0984 | 0.664 | 5.28 | 5.66 | 0.10 |
| **De\_tqwt\_kurtosisValue\_dec\_33** | 0.45260 | 1.335 | 1.572 | 1.852 | 0.689 | 0.752 | 0.810 | 0.678 | 0.757 | 0.807 | 0.0929 | 0.568 | 5.22 | 4.72 | 1.00 |
| **De\_tqwt\_maxValue\_dec\_28** | -0.08953 | 0.885 | 0.914 | 0.945 | 0.650 | 0.697 | 0.744 | 0.647 | 0.701 | 0.750 | 0.0984 | 0.636 | 5.19 | 5.32 | 0.30 |
| **De\_tqwt\_entropy\_log\_dec\_1** | 1.11054 | 2.014 | 3.036 | 4.577 | 0.607 | 0.712 | 0.772 | 0.632 | 0.720 | 0.772 | 0.0980 | 0.748 | 5.16 | 6.42 | 0.20 |
| **De\_locDbShimmer** | 0.75268 | 1.629 | 2.123 | 2.766 | 0.716 | 0.755 | 0.791 | 0.737 | 0.747 | 0.784 | 0.0804 | 0.742 | 5.15 | 6.43 | 1.00 |
| **Ba\_tqwt\_TKEO\_mean\_dec\_35** | -0.00529 | 0.993 | 0.995 | 0.997 | 0.657 | 0.700 | 0.720 | 0.653 | 0.711 | 0.739 | 0.0967 | 0.604 | 5.15 | 5.01 | 0.15 |
| **De\_tqwt\_TKEO\_std\_dec\_17** | -0.12032 | 0.847 | 0.887 | 0.928 | 0.622 | 0.673 | 0.725 | 0.625 | 0.702 | 0.744 | 0.0992 | 0.638 | 5.10 | 5.31 | 0.20 |
| **tqwt\_meanValue\_dec\_11** | 0.00190 | 1.001 | 1.002 | 1.003 | 0.613 | 0.717 | 0.764 | 0.653 | 0.722 | 0.772 | 0.0936 | 0.656 | 5.08 | 5.55 | 0.35 |
| **tqwt\_kurtosisValue\_dec\_27** | -0.00856 | 0.988 | 0.991 | 0.995 | 0.692 | 0.662 | 0.733 | 0.666 | 0.683 | 0.737 | 0.0973 | 0.659 | 5.04 | 5.51 | 0.15 |
| **De\_tqwt\_entropy\_shannon\_dec\_17** | -0.03367 | 0.954 | 0.967 | 0.980 | 0.604 | 0.656 | 0.696 | 0.615 | 0.666 | 0.711 | 0.0958 | 0.598 | 4.96 | 4.93 | 0.10 |
| **De\_tqwt\_maxValue\_dec\_29** | -0.05682 | 0.924 | 0.945 | 0.966 | 0.634 | 0.679 | 0.714 | 0.655 | 0.700 | 0.731 | 0.0926 | 0.618 | 4.91 | 5.18 | 0.30 |
| **De\_tqwt\_minValue\_dec\_11** | 0.04927 | 1.033 | 1.051 | 1.069 | 0.648 | 0.681 | 0.735 | 0.615 | 0.695 | 0.743 | 0.0937 | 0.523 | 4.91 | 4.39 | 0.25 |
| **De\_tqwt\_TKEO\_std\_dec\_36** | 0.13568 | 1.085 | 1.145 | 1.209 | 0.640 | 0.706 | 0.760 | 0.611 | 0.721 | 0.769 | 0.0809 | 0.616 | 4.83 | 5.12 | 0.75 |
| **De\_tqwt\_entropy\_log\_dec\_11** | -0.18590 | 0.771 | 0.830 | 0.895 | 0.664 | 0.692 | 0.742 | 0.650 | 0.706 | 0.750 | 0.0845 | 0.684 | 4.81 | 5.75 | 0.25 |
| **Ba\_IMF\_SNR\_SEO** | 0.01809 | 1.011 | 1.018 | 1.026 | 0.608 | 0.706 | 0.755 | 0.606 | 0.722 | 0.762 | 0.0874 | 0.578 | 4.72 | 4.76 | 0.20 |
| **De\_det\_LT\_TKEO\_mean\_3\_coef** | -0.29339 | 0.661 | 0.746 | 0.842 | 0.608 | 0.675 | 0.722 | 0.622 | 0.696 | 0.736 | 0.0826 | 0.598 | 4.62 | 4.96 | 0.60 |
| **De\_tqwt\_TKEO\_std\_dec\_11** | -0.08033 | 0.892 | 0.923 | 0.954 | 0.662 | 0.685 | 0.743 | 0.663 | 0.703 | 0.753 | 0.0809 | 0.677 | 4.60 | 5.68 | 0.55 |
| **Ba\_tqwt\_stdValue\_dec\_5** | -0.01078 | 0.985 | 0.989 | 0.994 | 0.615 | 0.683 | 0.713 | 0.645 | 0.698 | 0.726 | 0.0832 | 0.533 | 4.59 | 4.42 | 0.15 |
| **De\_tqwt\_TKEO\_mean\_dec\_33** | -0.26181 | 0.690 | 0.770 | 0.859 | 0.655 | 0.711 | 0.766 | 0.664 | 0.735 | 0.773 | 0.0745 | 0.551 | 4.56 | 4.57 | 1.00 |
| **De\_tqwt\_TKEO\_std\_dec\_7** | -0.09248 | 0.877 | 0.912 | 0.948 | 0.594 | 0.714 | 0.758 | 0.600 | 0.736 | 0.778 | 0.0795 | 0.597 | 4.56 | 4.96 | 0.30 |
| **De\_tqwt\_minValue\_dec\_7** | 0.16090 | 1.099 | 1.175 | 1.256 | 0.565 | 0.712 | 0.744 | 0.584 | 0.736 | 0.761 | 0.0812 | 0.674 | 4.54 | 5.69 | 0.35 |
| **f1** | -0.08060 | 0.890 | 0.923 | 0.956 | 0.598 | 0.697 | 0.735 | 0.596 | 0.711 | 0.748 | 0.0724 | 0.352 | 4.37 | 2.82 | 0.20 |
| **Ba\_tqwt\_kurtosisValue\_dec\_34** | 0.00635 | 1.003 | 1.006 | 1.009 | 0.610 | 0.642 | 0.693 | 0.636 | 0.659 | 0.712 | 0.0597 | 0.503 | 3.97 | 4.11 | 0.10 |
| **De\_tqwt\_entropy\_log\_dec\_35** | -0.15669 | 0.789 | 0.855 | 0.926 | 0.560 | 0.745 | 0.779 | 0.556 | 0.759 | 0.785 | 0.0524 | 0.553 | 3.68 | 4.58 | 0.15 |

## Let focus on the new features  
  
decorCoeff <- smDecor$coefficients[newvars,];  
ncoef <- dc[newvars]  
cnames <- lapply(ncoef,names)  
names(cnames) <- NULL;  
decorCoeff$Elements <- lapply(cnames,paste,collapse="+")  
pander::pander(decorCoeff)

|  | Estimate | lower | OR | upper | u.Accuracy | r.Accuracy | full.Accuracy | u.AUC | r.AUC | full.AUC | IDI | NRI | z.IDI | z.NRI | Frequency | Elements |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **De\_tqwt\_energy\_dec\_33** | -0.2126 | 0.760 | 0.809 | 0.860 | 0.747 | 0.719 | 0.781 | 0.736 | 0.739 | 0.779 | 0.1437 | 0.802 | 6.26 | 6.94 | 1.00 | tqwt\_energy\_dec\_31+tqwt\_energy\_dec\_33 |
| **De\_tqwt\_stdValue\_dec\_33** | -0.9220 | 0.296 | 0.398 | 0.534 | 0.680 | 0.693 | 0.777 | 0.683 | 0.717 | 0.767 | 0.1142 | 0.658 | 5.85 | 5.53 | 0.95 | tqwt\_TKEO\_mean\_dec\_35+tqwt\_stdValue\_dec\_32+tqwt\_stdValue\_dec\_33 |
| **De\_locDbShimmer** | 1.0876 | 2.098 | 2.967 | 4.197 | 0.714 | 0.744 | 0.779 | 0.736 | 0.734 | 0.777 | 0.0977 | 0.788 | 5.66 | 6.86 | 1.00 | locDbShimmer+ddaShimmer |
| **De\_std\_MFCC\_2nd\_coef** | -1.3731 | 0.157 | 0.253 | 0.408 | 0.703 | 0.787 | 0.827 | 0.685 | 0.781 | 0.826 | 0.1115 | 0.691 | 5.57 | 5.85 | 1.00 | std\_MFCC\_2nd\_coef+std\_2nd\_delta |
| **De\_tqwt\_TKEO\_std\_dec\_11** | -0.0490 | 0.936 | 0.952 | 0.969 | 0.699 | 0.692 | 0.767 | 0.701 | 0.691 | 0.772 | 0.1031 | 0.757 | 5.35 | 6.46 | 0.10 | tqwt\_TKEO\_mean\_dec\_10+tqwt\_TKEO\_mean\_dec\_11+tqwt\_TKEO\_std\_dec\_11+tqwt\_stdValue\_dec\_10+tqwt\_minValue\_dec\_17 |
| **De\_tqwt\_minValue\_dec\_11** | 0.0267 | 1.017 | 1.027 | 1.037 | 0.637 | 0.706 | 0.748 | 0.609 | 0.721 | 0.758 | 0.0994 | 0.592 | 5.32 | 4.93 | 0.10 | tqwt\_minValue\_dec\_11+tqwt\_minValue\_dec\_17 |
| **De\_std\_12th\_delta\_delta** | 0.6764 | 1.538 | 1.967 | 2.515 | 0.694 | 0.724 | 0.756 | 0.675 | 0.730 | 0.761 | 0.1078 | 0.643 | 5.31 | 5.36 | 0.70 | std\_MFCC\_12th\_coef+std\_12th\_delta\_delta |
| **De\_tqwt\_minValue\_dec\_7** | 0.1940 | 1.130 | 1.214 | 1.305 | 0.563 | 0.695 | 0.746 | 0.591 | 0.709 | 0.763 | 0.0969 | 0.689 | 5.14 | 5.83 | 0.25 | tqwt\_minValue\_dec\_7+tqwt\_minValue\_dec\_17+tqwt\_maxValue\_dec\_8 |
| **De\_tqwt\_kurtosisValue\_dec\_33** | 0.6324 | 1.493 | 1.882 | 2.373 | 0.694 | 0.780 | 0.827 | 0.679 | 0.773 | 0.826 | 0.0868 | 0.541 | 5.14 | 4.49 | 1.00 | tqwt\_kurtosisValue\_dec\_32+tqwt\_kurtosisValue\_dec\_33 |
| **De\_tqwt\_TKEO\_std\_dec\_36** | 0.1316 | 1.083 | 1.141 | 1.202 | 0.640 | 0.704 | 0.753 | 0.609 | 0.723 | 0.767 | 0.0770 | 0.576 | 4.82 | 4.80 | 0.65 | tqwt\_TKEO\_mean\_dec\_35+tqwt\_TKEO\_std\_dec\_36 |
| **De\_tqwt\_TKEO\_std\_dec\_7** | -0.1133 | 0.852 | 0.893 | 0.936 | 0.595 | 0.717 | 0.747 | 0.596 | 0.745 | 0.770 | 0.0839 | 0.644 | 4.66 | 5.38 | 0.25 | tqwt\_TKEO\_std\_dec\_7+tqwt\_stdValue\_dec\_7+tqwt\_minValue\_dec\_17 |
| **De\_det\_LT\_TKEO\_mean\_3\_coef** | -0.1846 | 0.762 | 0.831 | 0.908 | 0.610 | 0.721 | 0.742 | 0.621 | 0.726 | 0.750 | 0.0610 | 0.585 | 4.05 | 4.85 | 0.35 | Ed2\_3\_coef+det\_LT\_TKEO\_mean\_1\_coef+det\_LT\_TKEO\_mean\_3\_coef |
| **De\_tqwt\_entropy\_log\_dec\_1** | 0.8399 | 1.528 | 2.316 | 3.512 | 0.607 | 0.751 | 0.770 | 0.631 | 0.763 | 0.779 | 0.0583 | 0.522 | 3.88 | 4.30 | 0.25 | tqwt\_entropy\_log\_dec\_1+tqwt\_entropy\_log\_dec\_2 |
| **De\_tqwt\_TKEO\_std\_dec\_17** | -0.0626 | 0.910 | 0.939 | 0.970 | 0.623 | 0.696 | 0.745 | 0.624 | 0.720 | 0.759 | 0.0504 | 0.495 | 3.81 | 4.03 | 0.10 | tqwt\_TKEO\_std\_dec\_17+tqwt\_minValue\_dec\_17 |

## Differences Between Blind *vs.* Outcome-Driven Decorrelation

In this section I will show the differences in unaltered basis vectors between the Outcome driven Transformation vs the blind decorrelated transformation

par(op)  
par(mfrow=c(1,1))  
  
  
smDecorU <- summary(bmdU)  
decornamesU <- rownames(smDecorU$coefficients)  
  
get\_De\_names <- decornames[!str\_detect(decornames,"De\_")]  
get\_De\_namesU <- decornamesU[!str\_detect(decornamesU,"De\_")]  
  
unn <- bmd$univariate[,3]  
names(unn) <- rownames(bmd$univariate)  
pander::pander(as.matrix(unn[get\_De\_names]))

|  |  |
| --- | --- |
| **Ba\_tqwt\_kurtosisValue\_dec\_36** | 5.83 |
| **Ba\_tqwt\_entropy\_log\_dec\_12** | 5.38 |
| **Ba\_tqwt\_kurtosisValue\_dec\_28** | 4.51 |
| **Ba\_std\_delta\_delta\_log\_energy** | 6.15 |
| **Ba\_tqwt\_energy\_dec\_12** | 4.45 |
| **Ba\_locAbsJitter** | 4.11 |
| **std\_MFCC\_8th\_coef** | 4.40 |
| **Ba\_tqwt\_kurtosisValue\_dec\_26** | 3.88 |
| **Ba\_std\_5th\_delta** | 4.28 |
| **Ba\_tqwt\_kurtosisValue\_dec\_17** | 4.27 |
| **tqwt\_kurtosisValue\_dec\_20** | 4.85 |
| **Ba\_meanIntensity** | 5.34 |
| **tqwt\_meanValue\_dec\_11** | 3.67 |

pander::pander(summary(unn[get\_De\_names]))

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| --- | --- | --- | --- | --- | --- |
| 3.67 | 4.27 | 4.45 | 4.7 | 5.34 | 6.15 |

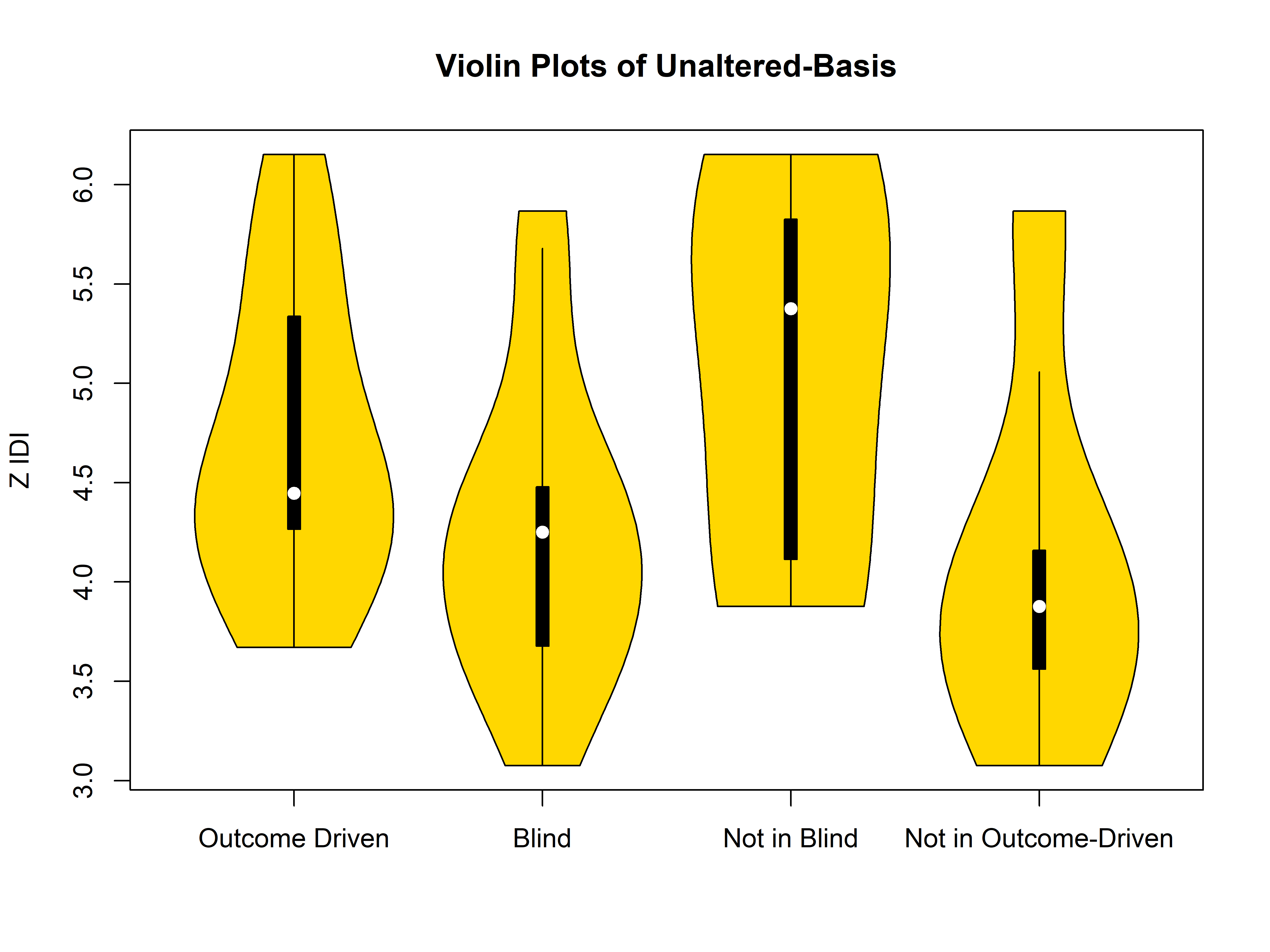
unnU <- bmdU$univariate[,3]  
names(unnU) <- rownames(bmdU$univariate)  
pander::pander(as.matrix(unnU[get\_De\_namesU]))

|  |  |
| --- | --- |
| **tqwt\_kurtosisValue\_dec\_36** | 5.83 |
| **Ba\_std\_5th\_delta** | 4.28 |
| **Ba\_std\_delta\_log\_energy** | 5.87 |
| **Ba\_tqwt\_energy\_dec\_12** | 4.45 |
| **Ba\_tqwt\_kurtosisValue\_dec\_28** | 4.51 |
| **tqwt\_kurtosisValue\_dec\_20** | 4.85 |
| **tqwt\_kurtosisValue\_dec\_26** | 3.88 |
| **Ba\_meanIntensity** | 5.34 |
| **mean\_MFCC\_2nd\_coef** | 3.54 |
| **std\_MFCC\_8th\_coef** | 4.40 |
| **Ba\_tqwt\_kurtosisValue\_dec\_17** | 4.27 |
| **Ba\_tqwt\_minValue\_dec\_17** | 4.07 |
| **Ba\_tqwt\_TKEO\_mean\_dec\_35** | 3.49 |
| **tqwt\_meanValue\_dec\_11** | 3.67 |
| **tqwt\_kurtosisValue\_dec\_27** | 4.25 |
| **Ba\_IMF\_SNR\_SEO** | 3.08 |
| **Ba\_tqwt\_stdValue\_dec\_5** | 3.69 |
| **f1** | 4.02 |
| **Ba\_tqwt\_kurtosisValue\_dec\_34** | 3.58 |

pander::pander(summary(unnU[get\_De\_namesU]))

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| --- | --- | --- | --- | --- | --- |
| 3.08 | 3.68 | 4.25 | 4.27 | 4.48 | 5.87 |

#boxplot(unn[get\_De\_names],unnU[get\_De\_namesU],xlab=c("Method"),ylab="Z",main="Z Values of Basis Features")  
  
x1 <- unn[get\_De\_names]  
x2 <- unnU[get\_De\_namesU]  
X3 <- x1[!(get\_De\_names %in% get\_De\_namesU)]  
X4 <- x2[!(get\_De\_namesU %in% get\_De\_names)]  
vioplot(x1, x2, X3,X4, names=c("Outcome Driven", "Blind","Not in Blind","Not in Outcome-Driven"),ylab="Z IDI",  
 col="gold")  
title("Violin Plots of Unaltered-Basis")



sameFeatures <- get\_De\_names[get\_De\_names %in% get\_De\_namesU]  
pander::pander(as.matrix(unn[sameFeatures]))

|  |  |
| --- | --- |
| **Ba\_tqwt\_kurtosisValue\_dec\_28** | 4.51 |
| **Ba\_tqwt\_energy\_dec\_12** | 4.45 |
| **std\_MFCC\_8th\_coef** | 4.40 |
| **Ba\_std\_5th\_delta** | 4.28 |
| **Ba\_tqwt\_kurtosisValue\_dec\_17** | 4.27 |
| **tqwt\_kurtosisValue\_dec\_20** | 4.85 |
| **Ba\_meanIntensity** | 5.34 |
| **tqwt\_meanValue\_dec\_11** | 3.67 |

## The features by Outcome Drive not in blind  
pander::pander(as.matrix(x1[!(get\_De\_names %in% get\_De\_namesU)]))

|  |  |
| --- | --- |
| **Ba\_tqwt\_kurtosisValue\_dec\_36** | 5.83 |
| **Ba\_tqwt\_entropy\_log\_dec\_12** | 5.38 |
| **Ba\_std\_delta\_delta\_log\_energy** | 6.15 |
| **Ba\_locAbsJitter** | 4.11 |
| **Ba\_tqwt\_kurtosisValue\_dec\_26** | 3.88 |

## The features not in outcome driven  
pander::pander(as.matrix(x2[!(get\_De\_namesU %in% get\_De\_names)]))

|  |  |
| --- | --- |
| **tqwt\_kurtosisValue\_dec\_36** | 5.83 |
| **Ba\_std\_delta\_log\_energy** | 5.87 |
| **tqwt\_kurtosisValue\_dec\_26** | 3.88 |
| **mean\_MFCC\_2nd\_coef** | 3.54 |
| **Ba\_tqwt\_minValue\_dec\_17** | 4.07 |
| **Ba\_tqwt\_TKEO\_mean\_dec\_35** | 3.49 |
| **tqwt\_kurtosisValue\_dec\_27** | 4.25 |
| **Ba\_IMF\_SNR\_SEO** | 3.08 |
| **Ba\_tqwt\_stdValue\_dec\_5** | 3.69 |
| **f1** | 4.02 |
| **Ba\_tqwt\_kurtosisValue\_dec\_34** | 3.58 |

### The Final Table

I’ll create a table subset of the logistic model from the Outcome-Driven decorrelated data.

The table will have:

1. The top associated features described by the feature network, as well as, and the new features.
   1. For Decorrelated features it will provide the decorrelation formula
2. Nugget labels
   1. The label of nugget as found by the clustering procedure
3. The feature coefficient
4. The feature Odd ratios and their corresponding 95%CI

## The features in top nugget  
clusterFeatures <- clusterOutcome$names  
## The new features   
discoveredFeatures <- newvars[zvaluePrePost$ZUni<1.96]  
  
tablefinal <- smDecor$coefficients[unique(c(clusterFeatures,discoveredFeatures)),  
 c("Estimate","lower","OR","upper","z.IDI")]  
  
nugget <- clusterOutcome$membership  
names(nugget) <- clusterOutcome$names  
tablefinal$Nugget <- nugget[rownames(tablefinal)]  
tablefinal$Nugget[is.na(tablefinal$Nugget)] <- "D"  
deFromula <- character(length(theDeFormulas))  
names(deFromula) <- names(theDeFormulas)  
for (dx in names(deFromula))  
{  
 coef <- theDeFormulas[[dx]]  
 cname <- names(theDeFormulas[[dx]])  
 names(cname) <- cname  
 for (cf in names(coef))  
 {  
 if (cf != dx)  
 {  
 if (coef[cf]>0)  
 {  
 deFromula[dx] <- paste(deFromula[dx],  
 sprintf("+ %5.3f\*%s",coef[cf],cname[cf]))  
 }  
 else  
 {  
 deFromula[dx] <- paste(deFromula[dx],  
 sprintf("%5.3f\*%s",coef[cf],cname[cf]))  
 }  
 }  
 }  
}  
tablefinal$DecorFormula <- deFromula[rownames(tablefinal)]  
pander::pander(tablefinal)

|  | Estimate | lower | OR | upper | z.IDI | Nugget | DecorFormula |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Ba\_std\_delta\_delta\_log\_energy** | 0.4337 | 1.361 | 1.543 | 1.749 | 6.33 | 1 | NA |
| **De\_std\_MFCC\_2nd\_coef** | -1.3731 | 0.157 | 0.253 | 0.408 | 5.57 | 1 | + 1.000*std\_MFCC\_2nd\_coef -0.828*std\_2nd\_delta |
| **De\_tqwt\_kurtosisValue\_dec\_33** | 0.6324 | 1.493 | 1.882 | 2.373 | 5.14 | 1 | -0.883*tqwt\_kurtosisValue\_dec\_32 + 1.000*tqwt\_kurtosisValue\_dec\_33 |
| **De\_tqwt\_energy\_dec\_33** | -0.2126 | 0.760 | 0.809 | 0.860 | 6.26 | 2 | -0.907*tqwt\_energy\_dec\_31 + 1.000*tqwt\_energy\_dec\_33 |
| **De\_locDbShimmer** | 1.0876 | 2.098 | 2.967 | 4.197 | 5.66 | 2 | + 1.000*locDbShimmer -0.963*ddaShimmer |
| **De\_tqwt\_stdValue\_dec\_33** | -0.9220 | 0.296 | 0.398 | 0.534 | 5.85 | 3 | -0.131*tqwt\_TKEO\_mean\_dec\_35 -0.715*tqwt\_stdValue\_dec\_32 + 1.000\*tqwt\_stdValue\_dec\_33 |
| **Ba\_tqwt\_kurtosisValue\_dec\_36** | 0.1450 | 1.116 | 1.156 | 1.198 | 6.87 | 3 | NA |
| **Ba\_tqwt\_entropy\_log\_dec\_12** | -0.6495 | 0.444 | 0.522 | 0.614 | 6.49 | 4 | NA |
| **tqwt\_kurtosisValue\_dec\_20** | 0.5330 | 1.409 | 1.704 | 2.060 | 5.07 | 5 | NA |
| **De\_std\_12th\_delta\_delta** | 0.6764 | 1.538 | 1.967 | 2.515 | 5.31 | 5 | -0.914*std\_MFCC\_12th\_coef + 1.000*std\_12th\_delta\_delta |
| **Ba\_std\_5th\_delta** | 0.4967 | 1.385 | 1.643 | 1.949 | 5.40 | 4 | NA |
| **Ba\_meanIntensity** | -1.1976 | 0.191 | 0.302 | 0.478 | 5.00 | 6 | NA |
| **std\_MFCC\_8th\_coef** | 0.5419 | 1.431 | 1.719 | 2.066 | 5.67 | 6 | NA |
| **De\_tqwt\_TKEO\_std\_dec\_36** | 0.1316 | 1.083 | 1.141 | 1.202 | 4.82 | 6 | -0.926*tqwt\_TKEO\_mean\_dec\_35 + 1.000*tqwt\_TKEO\_std\_dec\_36 |
| **Ba\_tqwt\_energy\_dec\_12** | -0.0631 | 0.922 | 0.939 | 0.956 | 5.95 | 6 | NA |
| **De\_det\_LT\_TKEO\_mean\_3\_coef** | -0.1846 | 0.762 | 0.831 | 0.908 | 4.05 | D | -0.890*Ed2\_3\_coef -0.083*det\_LT\_TKEO\_mean\_1\_coef + 1.000\*det\_LT\_TKEO\_mean\_3\_coef |
| **De\_tqwt\_entropy\_log\_dec\_1** | 0.8399 | 1.528 | 2.316 | 3.512 | 3.88 | D | + 1.000*tqwt\_entropy\_log\_dec\_1 -0.849*tqwt\_entropy\_log\_dec\_2 |