Test

2023-04-30

```
# load required libraries
library(parallel)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stats)
library(knitr)
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
library(ggplot2)
library(foreach)
```

Load RDAs

```
#load(file = file.path(root_dir,"R_data","X_wm.Rda"))
#load(file = file.path(root_dir,"R_data","X_gm.Rda"))
load(file = file.path(root_dir,"R_data","X_cb.Rda"))
```

PCA

```
# fnto perform PCA and save output
pca_cb <- perform_pca(X_cb)
```

```
## [1] "n components: 158"
```

```
save(pca_cb, file = file.path(root_dir, "R_data", "pca_cb.Rda"))


head(pca_cb$X_train_pca[,1:5])
```

```
##        PC1        PC2       PC3       PC4      PC5
## 2 -6495908  1161601.3 -11211910   4265114 -5694802
## 3 -6466968  1241209.4 -10062481   3574993 -6114345
## 4 -1359033 -2727462.8   3721284 -10125874  6676974
```

```
## 5 -6066214   203555.3 -10608271   2938972 -5064700
## 6 -6111401   391510.8 -10256438   3490440 -5954710
## 7 -6774145   510879.9 -11552593   3841450 -6437806
```

```
dim(pca_cb$X_train_pca)
```

```
## [1] 348 158
```

```
head(pca_cb$y_train)
```

```
## [1] 2 1 1 1 1 1
```

```
length(pca_cb$y_train)
```

```
## [1] 348
```

```
length(pca_cb$y_test)
```

```
## [1] 86
```

```
head(pca_cb$X_test_pca[,1:5])
```

```
##         PC1        PC2       PC3     PC4      PC5
## 1  -6614385  -63108.68 -11420074 3723285 -6695565
## 9  -6766838  822584.88 -11153325 3641374 -6585939
## 15 -6687820  189298.49 -12046677 3601003 -6796936
## 16 -6222846 2283190.64 -11182672 4280336 -4400092
## 18 -6644409  766523.00 -11404098 3857188 -6234436
## 32 -5845090 -290050.39 -10635865 3039552 -6040057
```

```
dim(pca_cb$X_test_pca)
```

```
## [1]  86 158
```