

# R Notebook

## Different naming of bx for correction

```
-SFactor7 = PB23001 -> bx.identifier = "NEW"  
-SFactor7 = PB25001 -> bx.identifier = "DB"  
-bx_correct = bx.identifier  
-bx_bg = SFactor7  
  
## Warning: replacing previous import 'data.table::first' by 'dplyr::first' when  
## loading 'dascombat'  
  
## Warning: replacing previous import 'data.table::last' by 'dplyr::last' when  
## loading 'dascombat'  
  
## Warning: replacing previous import 'data.table::between' by 'dplyr::between'  
## when loading 'dascombat'  
  
## [1] "C:/Users/rdwijn/OneDrive - Pamgene/Documents/101-100 DAS/PB25001"
```

## basic processing

```
## This batch took 3 iterations until convergence  
## This batch took 4 iterations until convergence  
## This batch took 5 iterations until convergence  
## This batch took 5 iterations until convergence  
## This batch took 5 iterations until convergence
```

### 1. PCA visual of uncorrected data

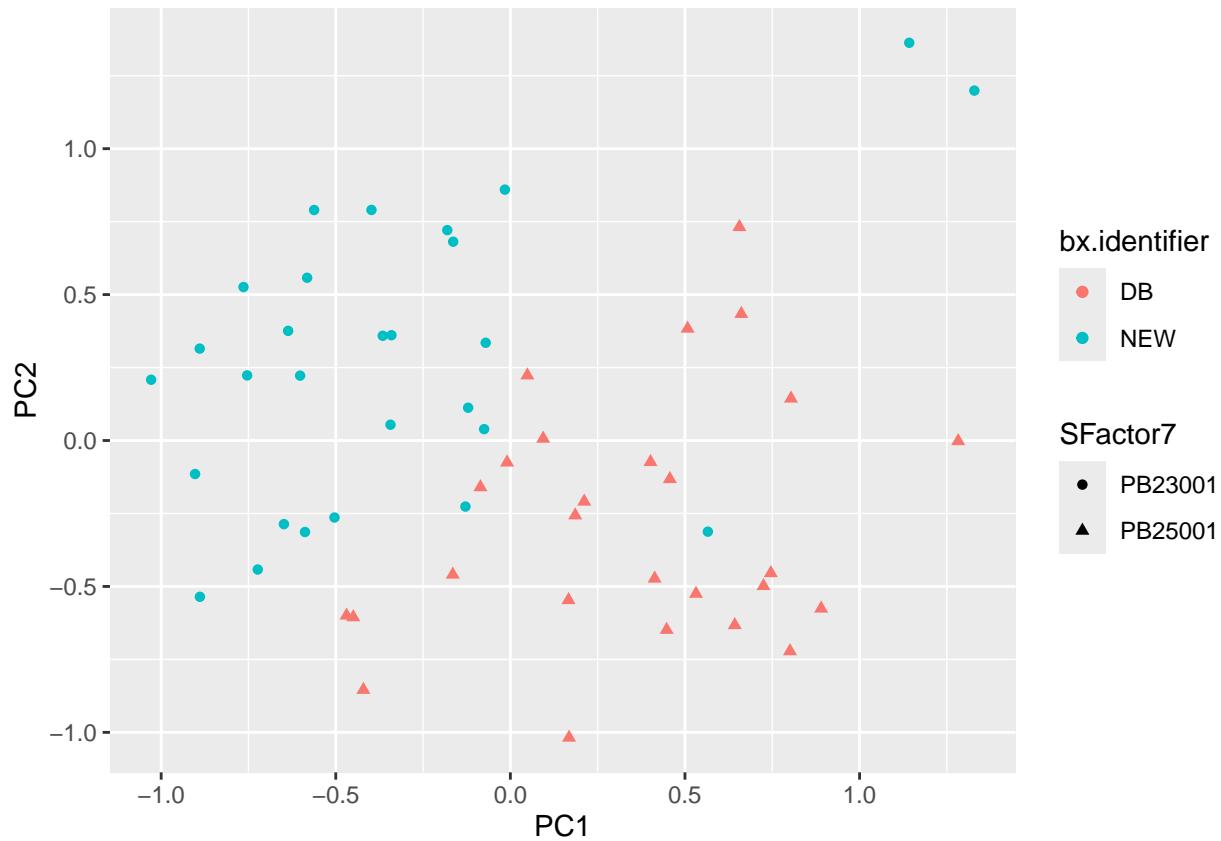
```
aPca = prcomp(t(Y))  
  
dfs = aPca$x %>%  
  as.data.frame() %>%  
  rownames_to_column("Sample") %>%  
  separate(Sample, into = c("Barcode", "Array"), sep = "_")  
  
meta = pb %>%  
  filter(rowSeq == 1) %>%  
  select(Barcode, Array, SFactor7, bx.identifier, pcRun, Date, measID)
```

```

dfs = dfs %>%
  left_join(meta, by = c("Barcode", "Array"))

dfs %>%
  ggplot(aes(x = PC1, y = PC2, color = bx.identifier, shape = SFactor7)) +
  geom_point()

```



## 2. PCA visual of corrected data

- Using `bx_correct` for ComBat without reference batch

### correct

```
model2 = dascombat::fit(Y, bx_correct, mean.only = FALSE, ref.batch = NULL)
```

```
## This batch took 4 iterations until convergence
## This batch took 4 iterations until convergence
```

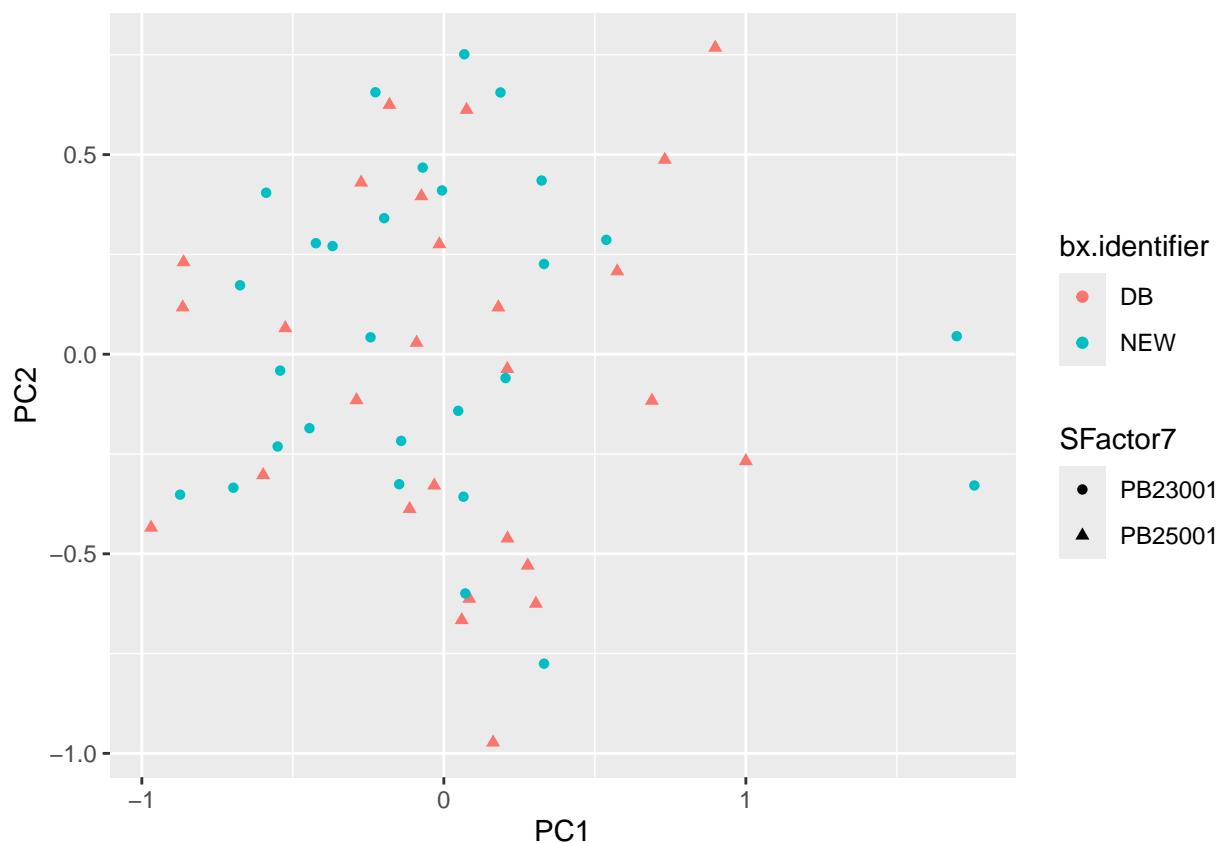
```

Y2 = dascombat::applyModel(Y, model2, model2$bx)

aPca2 = prcomp(t(Y2))
dfs2 = aPca2$x %>%
  as.data.frame() %>%
  rownames_to_column("Sample") %>%
  separate(Sample, into = c("Barcode", "Array"), sep = "_")
dfs2 = dfs2 %>%
  left_join(meta, by = c("Barcode", "Array"))

ggplot(dfs2, aes(x = PC1, y = PC2, color = bx.identifier, shape = SFactor7)) +
  geom_point()

```



### 3. PCA visual of corrected data

- Using bx\_bg for ComBat without reference batch

**correct**

```

model3 = dascombat::fit(Y, bx_bg, mean.only = FALSE, ref.batch = NULL)

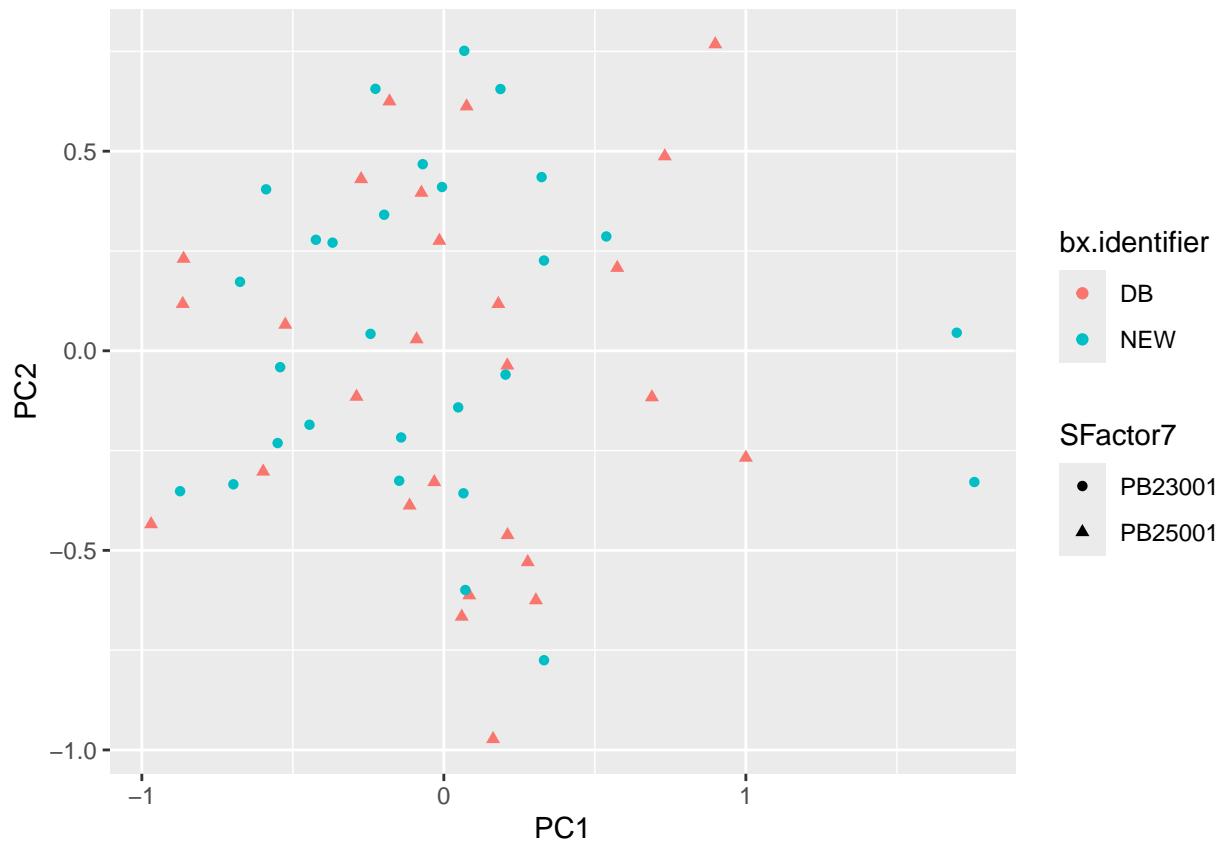
```

```

## This batch took 4 iterations until convergence
## This batch took 4 iterations until convergence

Y3 = dascombat::applyModel(Y, model3, model3$bx)
aPca3 = prcomp(t(Y3))
dfs3 = aPca3$x %>%
  as.data.frame() %>%
  rownames_to_column("Sample") %>%
  separate(Sample, into = c("Barcode", "Array"), sep = "_")
dfs3 = dfs3 %>%
  left_join(meta, by = c("Barcode", "Array"))
ggplot(dfs3, aes(x = PC1, y = PC2, color = bx.identifier, shape = SFactor7)) +
  geom_point()

```



#### 4. PCA visual of corrected data

- Using `bx_correct` for ComBat with reference batch = “DB”

`correct`

```

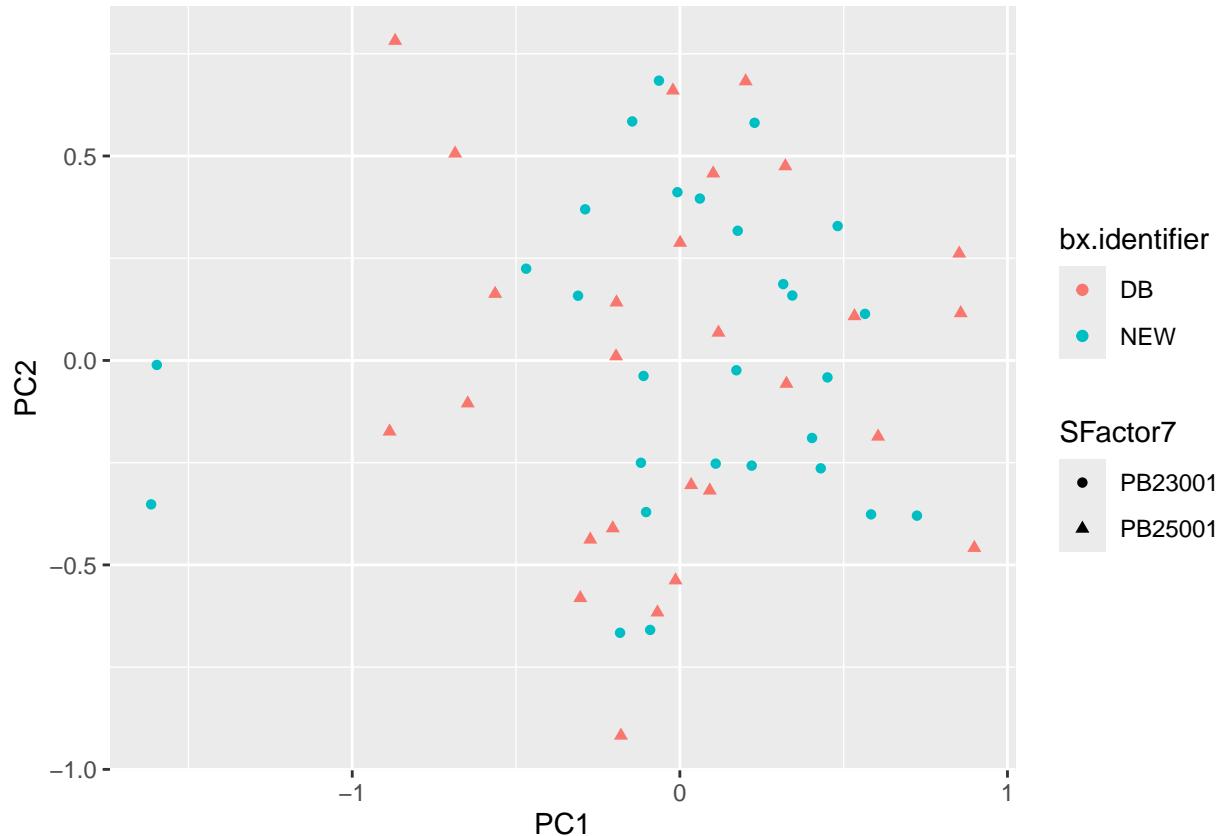
model4 = dascombat::fit(Y, bx_correct, mean.only = FALSE, ref.batch = "DB")

## This batch took 2 iterations until convergence
## This batch took 4 iterations until convergence

Y4 = dascombat::applyModel(Y, model4, model4$bx)
aPca4 = prcomp(t(Y4))
dfs4 = aPca4$x %>%
  as.data.frame() %>%
  rownames_to_column("Sample") %>%
  separate(Sample, into = c("Barcode", "Array"), sep = "_")
dfs4 = dfs4 %>%
  left_join(meta, by = c("Barcode", "Array"))

ggplot(dfs4, aes(x = PC1, y = PC2, color = bx.identifier, shape = SFactor7)) +
  geom_point()

```



## 5. PCA visual of corrected data

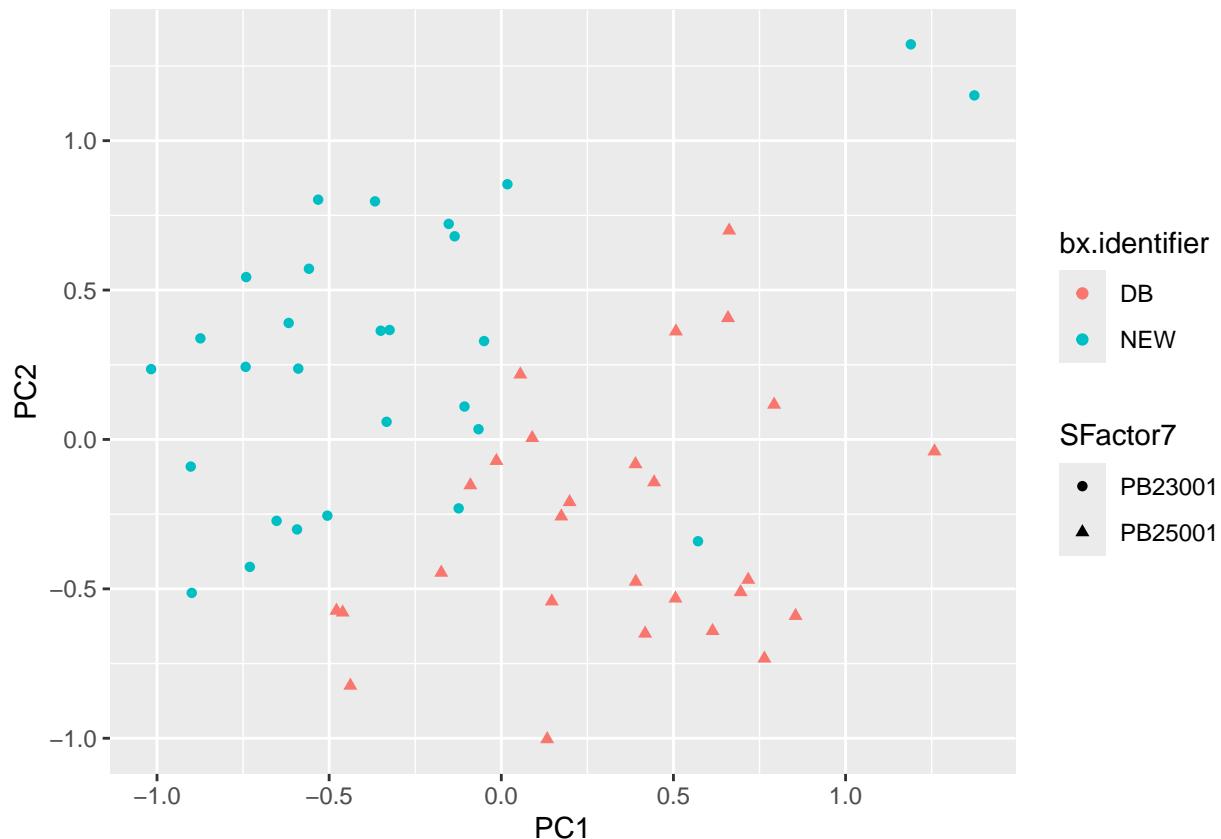
- Using `bx_correct` for ComBat with reference batch = “NEW”

## FAIL

```
model5 = dascombat::fit(Y, bx_correct, mean.only = FALSE, ref.batch = "NEW")
```

```
## This batch took 4 iterations until convergence  
## This batch took 2 iterations until convergence
```

```
Y5 = dascombat::applyModel(Y, model5, model5$bx)  
aPca5 = prcomp(t(Y5))  
dfs5 = aPca5$x %>%  
  as.data.frame() %>%  
  rownames_to_column("Sample") %>%  
  separate(Sample, into = c("Barcode", "Array"), sep = "_")  
dfs5 = dfs5 %>%  
  left_join(meta, by = c("Barcode", "Array"))  
ggplot(dfs5, aes(x = PC1, y = PC2, color = bx.identifier, shape = SFactor7)) +  
  geom_point()
```



## 6: PCA visual of corrected data

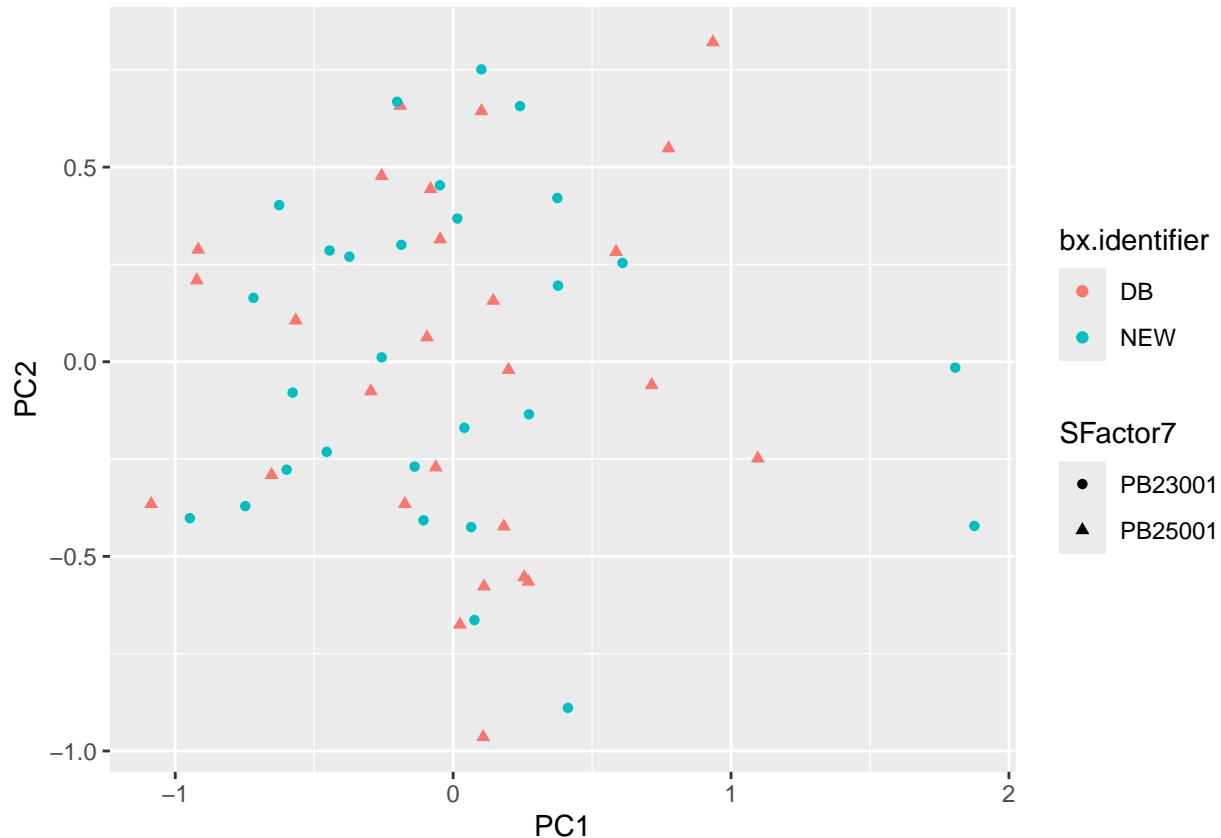
- Using `bx_bg` for ComBat with reference batch = “PB23001”

## correct

```
model6 = dascombat::fit(Y, bx_bg, mean.only = FALSE, ref.batch = "PB23001")
```

```
## This batch took 2 iterations until convergence  
## This batch took 4 iterations until convergence
```

```
Y6 = dascombat::applyModel(Y, model6, model6$bx)  
aPca6 = prcomp(t(Y6))  
dfs6 = aPca6$x %>%  
  as.data.frame() %>%  
  rownames_to_column("Sample") %>%  
  separate(Sample, into = c("Barcode", "Array"), sep = "_")  
dfs6 = dfs6 %>%  
  left_join(meta, by = c("Barcode", "Array"))  
  
ggplot(dfs6, aes(x = PC1, y = PC2, color = bx.identifier, shape = SFactor7)) +  
  geom_point()
```



```
# 7: PCA visual of corrected data - Using bx_bg for ComBat with reference batch = "PB25001"
```

## FAIL

```
model7 = dascombat::fit(Y, bx_bg, mean.only = FALSE, ref.batch= "PB25001")
```

```
## This batch took 4 iterations until convergence  
## This batch took 2 iterations until convergence
```

```
Y7 = dascombat::applyModel(Y, model7, model7$bx)  
aPca7 = prcomp(t(Y7))  
dfs7 = aPca7$x %>%  
  as.data.frame() %>%  
  rownames_to_column("Sample") %>%  
  separate(Sample, into = c("Barcode", "Array"), sep = "_")  
dfs7 = dfs7 %>%  
  left_join(meta, by = c("Barcode", "Array"))  
ggplot(dfs7, aes(x = PC1, y = PC2, color = bx.identifier, shape = SFactor7)) +  
  geom_point()
```

