Results based on the outlines

February, 2019

Contents

0.1   1. Load data

The *'mice.data.20181031.RData'* dataset contains 37 kinds of microbiomes and different numbers of subjects in control group and PAT group at week 3, 6, 10 ,13.

Here is the number of subjects in different groups at different time.

	Week 3	Week 6	Week 10	Week 13
Control	47	45	37	36
PAT	42	42	31	32

Here is the number of 0s in the counts for different groups at different time.

	Week 3	Week 6	Week 10	Week 13
Control	694	641	444	411
PAT	1128	1081	650	648

Since too many 0s may affect the results. I would like to use the subset at week 13.

I feel weird here. Why there are more non-zeros at the end of the trial?

0.2   2. Pure real data (without any change on the real data)

0.2.0.1   2.1 Whether the control group and the PAT group are different?

Here we can see, the groups ('control','PAT') do not have significant differences.

```
##              Df    Sum Sq Mean Sq F value Pr(>F)
## group         1 2.356e+05  235602   0.161  0.689
## Residuals  2446 3.590e+09 1467628
```

When adjusted with group, the microbiomes have significant differences in counts. This makes sense since we observed big differences.

```
##              Df    Sum Sq  Mean Sq F value Pr(>F)
## micro        35 3.074e+09 87824190 410.379 <2e-16 ***
## group         1 2.356e+05   235602   1.101  0.294
## Residuals  2411 5.160e+08   214008
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To show it more clear, we can draw a heatmap:

Fig 1: Raw data (control & PAT)
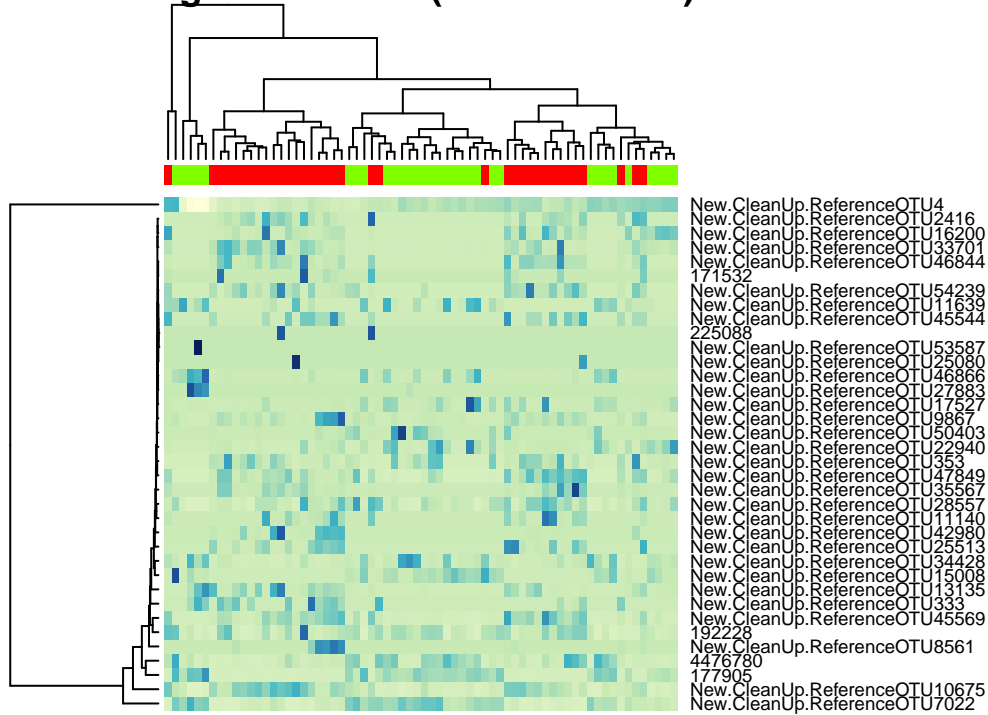
0.2.0.2 2.2 Test whether the clustering method can separate those two groups

Let's then try the biclustering methods:

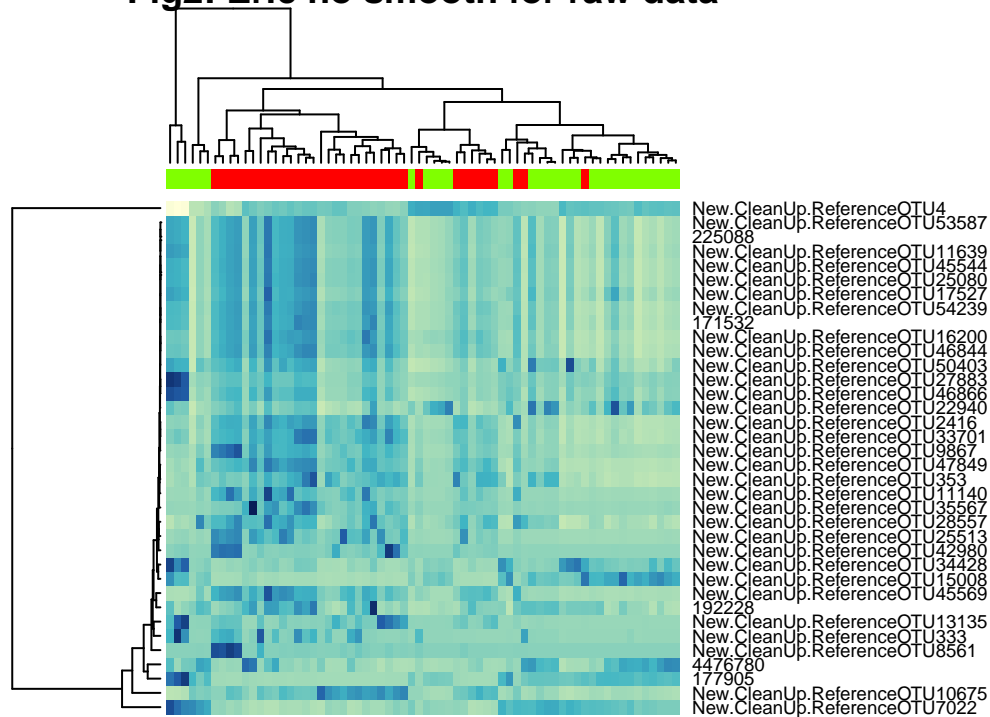### 0.2.0.2.1 2.1.1 Eric's method
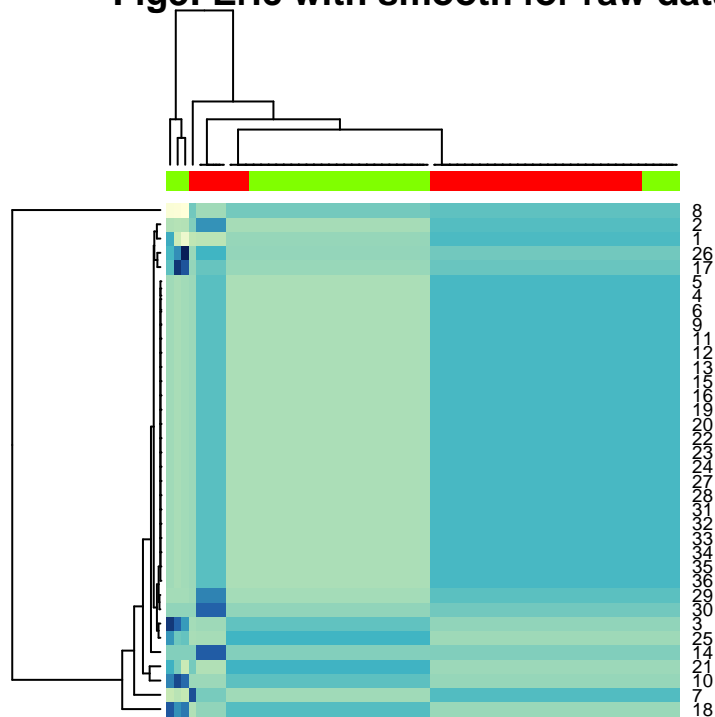
Fig2: Eric no smooth for raw data

New.CleanUp.ReferenceOTU4
New.CleanUp.ReferenceOTU53587
225088
New.CleanUp.ReferenceOTU11639
New.CleanUp.ReferenceOTU45544
New.CleanUp.ReferenceOTU25080
New.CleanUp.ReferenceOTU17527
New.CleanUp.ReferenceOTU54239
171532
New.CleanUp.ReferenceOTU16200
New.CleanUp.ReferenceOTU46844
New.CleanUp.ReferenceOTU50403
New.CleanUp.ReferenceOTU27883
New.CleanUp.ReferenceOTU46866
New.CleanUp.ReferenceOTU22940
New.CleanUp.ReferenceOTU2416
New.CleanUp.ReferenceOTU33701
New.CleanUp.ReferenceOTU9867
New.CleanUp.ReferenceOTU47849
New.CleanUp.ReferenceOTU353
New.CleanUp.ReferenceOTU11140
New.CleanUp.ReferenceOTU35567
New.CleanUp.ReferenceOTU28557
New.CleanUp.ReferenceOTU25513
New.CleanUp.ReferenceOTU42980
New.CleanUp.ReferenceOTU34428
New.CleanUp.ReferenceOTU15008
New.CleanUp.ReferenceOTU45569
192228
New.CleanUp.ReferenceOTU13135
New.CleanUp.ReferenceOTU333
New.CleanUp.ReferenceOTU8561
4476780
177905
New.CleanUp.ReferenceOTU10675
New.CleanUp.ReferenceOTU7022

Fig3: Eric with smooth for raw data

0.2.0.2.2   2.1.2 Our ADMM

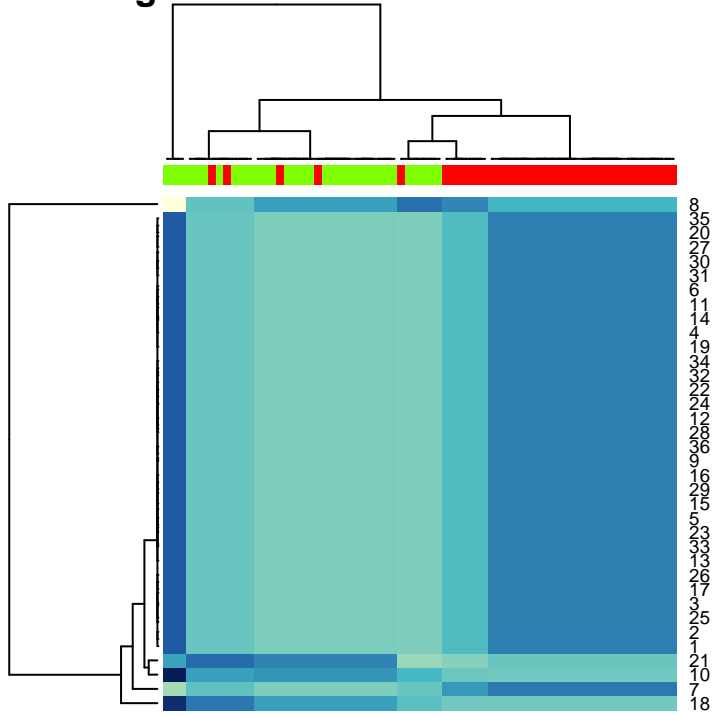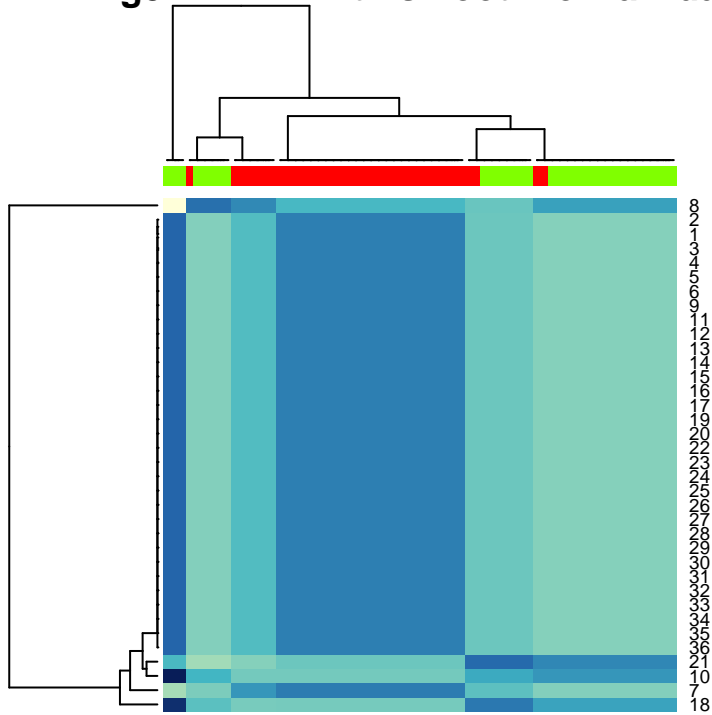**Fig4: ADMM no smooth for raw data**

Fig5: ADMM with smooth for raw data

0.3   3. Semi real data on the subjects.

Select one group, e.g. the control group.

First, test whether the distributions of microbiomes are different within the group. If not, select half of the

subjects as group 1 and the rest as group 2. Increase or decrease some microbiome values in group 1. Then apply our convex bi-clustering method to test wether we can separate group 1 and group 2.

0.3.0.1 3.1 comparision of the microbiome counts for subjects within the control group

The smallest p value in those comparision is:

```
## [1] 0.3533198
```

which is larger than 0.05, that is, there is no significant differences between two subjects.

0.3.0.2 3.2 Assign values to part of the counts

```
dim(dat2_temp)
```

```
## [1] 34 36
```

```
dat2_temp[18:34,1:18] = dat2_temp[18:34,1:18] * 50
```

- the data were sorted before the mutiplication

The heatmaps:

- Fig 6 is the heatmap for the raw control group data
- Fig 7 is the heatmap for the control group after mutiplying some arbitrary value.

Fig6: Raw data (control)

Fig7: Semi−raw data1 (control)
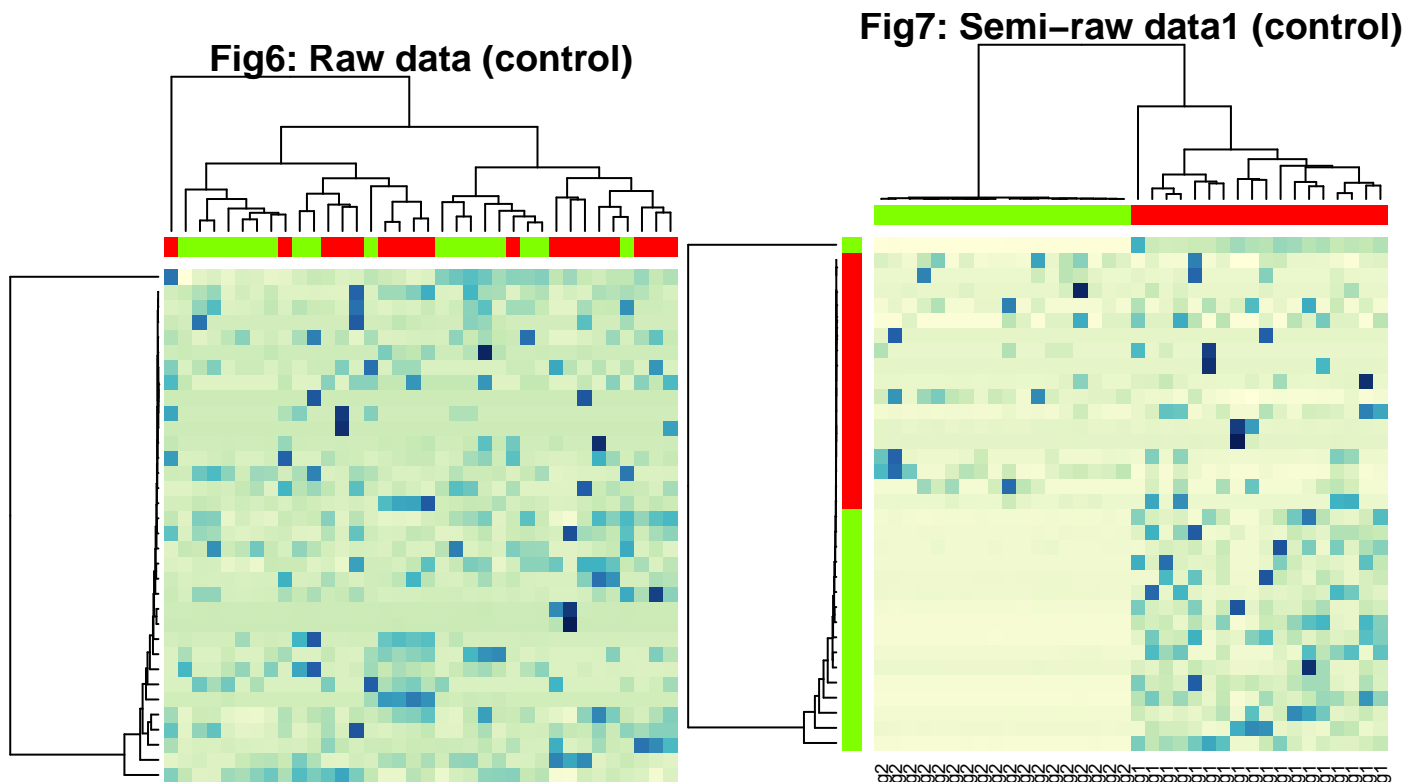
0.3.0.3 3.2 Test whether the clustering method can separate those two groups
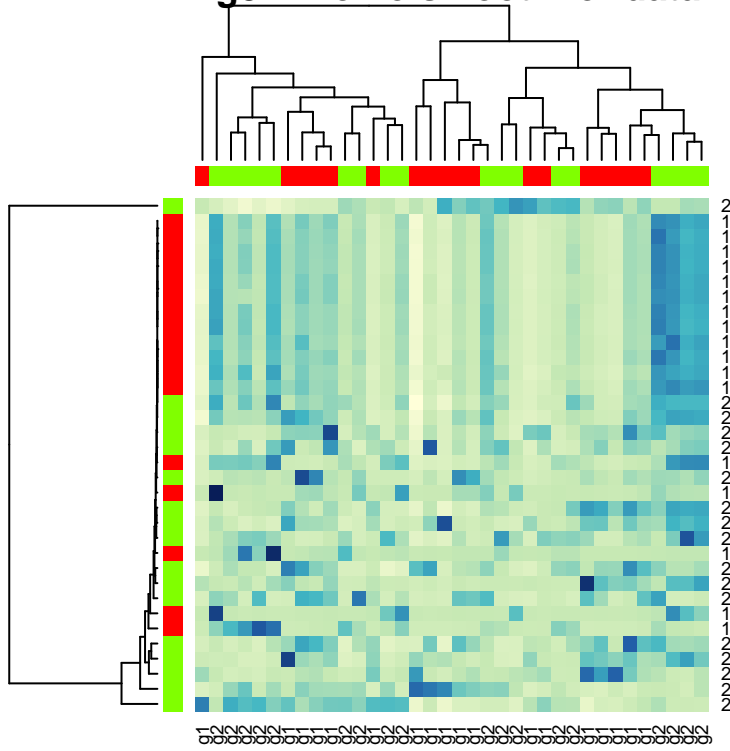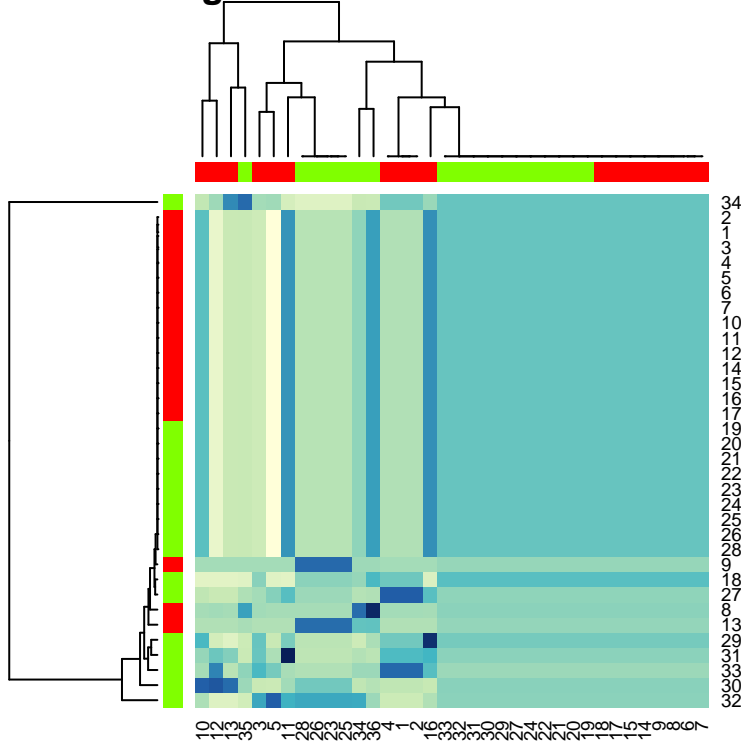
5

Fig8: Eric no smooth for data2

Fig9: Eric with smooth for data2

0.3.0.3.2 3.2.2 Our ADMM

nu1 = nu2 =2 gamma_1 = gamma_2 = 0.5
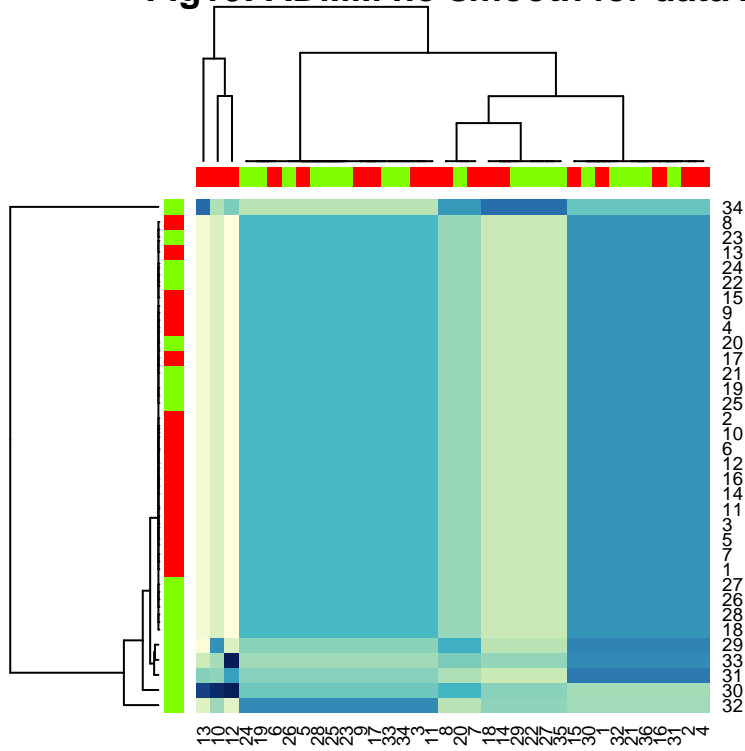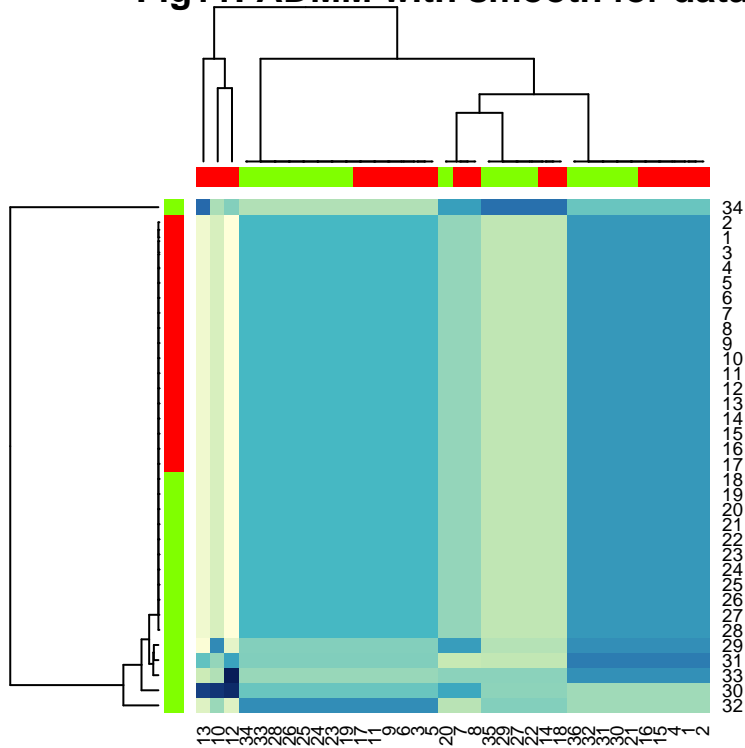
Fig10: ADMM no smooth for data 2

Fig11: ADMM with smooth for data2

0.4   4. Semi real data on both subjects and microbiomes.

We would like to see whether our method can this help us separate microbiome groups when we have already known the treatment groups.
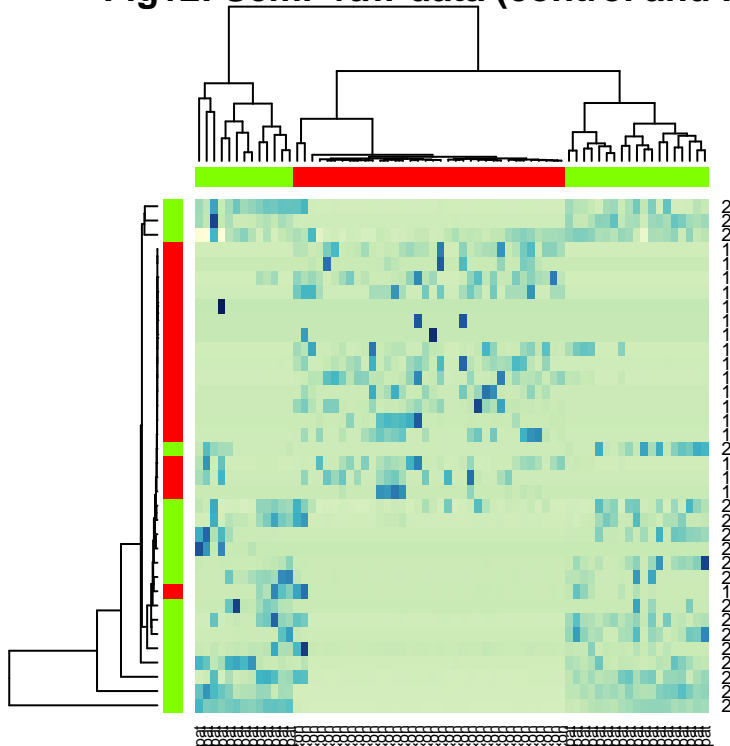
multiple some value for part of the data

```r
dim(dat3_temp)
```

```
## [1] 36 68
```

```r
dat3_temp[18:34,1:34] = dat3_temp[18:34,1:34] * 50
```

The heatmap of the data after multiplication

Fig12: Semi−raw data (control and PAT)

0.4.0.1   4.2 Test whether the clustering method can separate those two groups

0.4.0.1.1   4.2.1 Eric's method
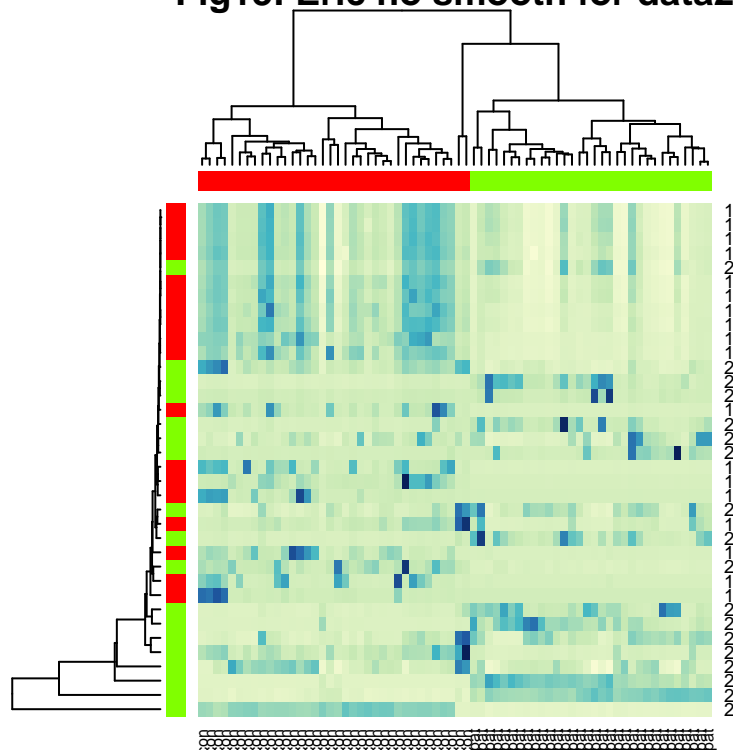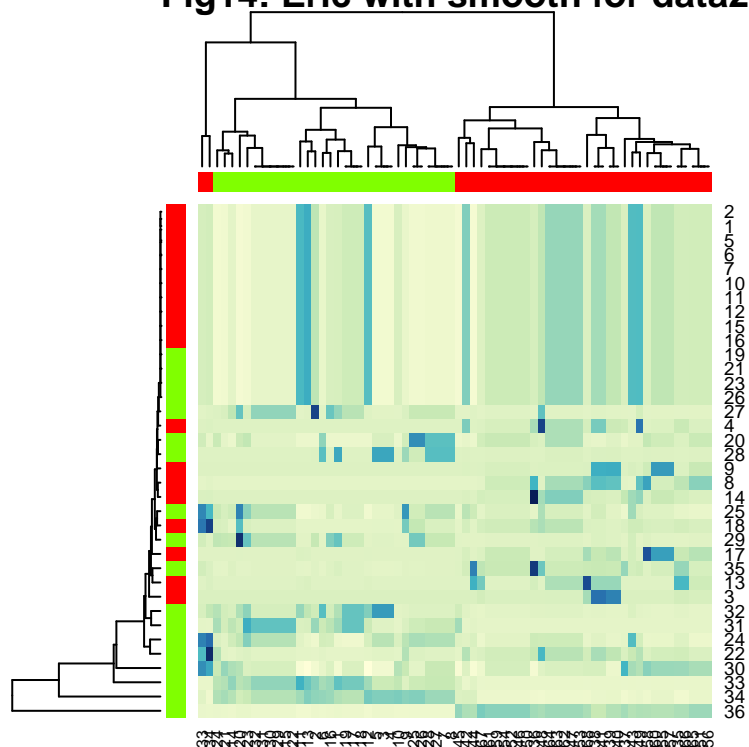
Fig13: Eric no smooth for data2

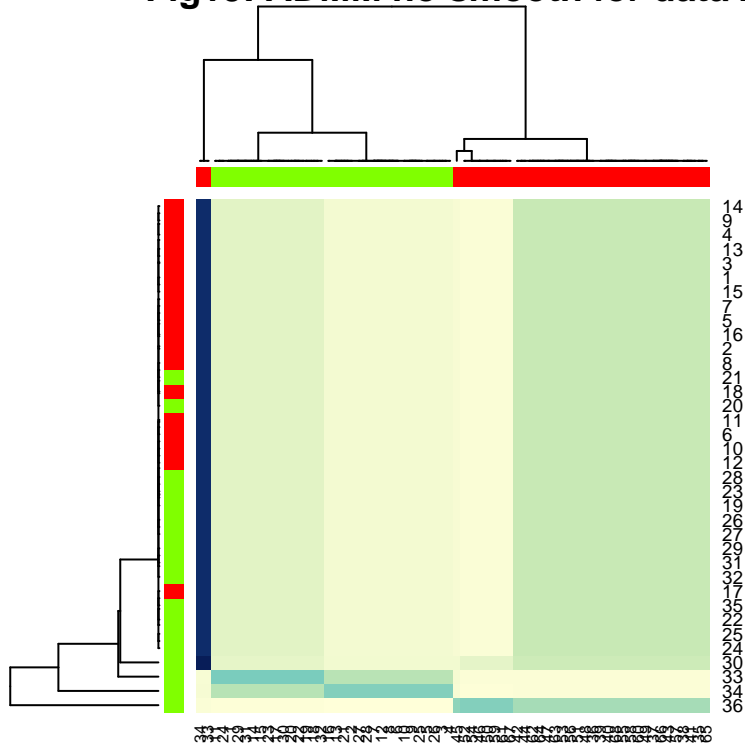**Fig14: Eric with smooth for data2**

0.4.0.1.2   4.2.2 ADMM

Fig15: ADMM no smooth for data 2

Fig16: ADMM with smooth for data2