

Test

Erica Baller

3/21/2018

R Markdown

This script goes through demographics, cnb scores, health, and psych summaries, adds clustering information, runs statistics and makes graphs from results.

Part 1 : Read in csv.s -This script reads in demographics, cnb_scores, health and psych summaries, merges them, removes NAs, codes and separates by depression.

Part 2 : merge with hydra -It then merges these documents with hydra output (made in cbica), adding Hydra_k1 through Hydra_k10 columns (which represent the number of clusters) -The script reads in 3 different types of groups (matched, unmatched, and residualized unmatched groups), and also does all gender together as well as separating them by gender.

Part 3 : Demographics tables - Demographics tables for each group (matched, unmatched, resid) were produced

Part 4 : Graphing - Graphs were then made.

For continuous variables(age, medu1), the graphs represent means, with SEM as error bars For categorical variables (race, sex) the graphs are percentages (caucasian, male) per group, with chisq used to calculate significance

Part 5 : LM -The script then runs LM on each cognitive score (cnb_measure ~ hydra_group).

-There is a test option that does this for all cnb measures and all hydra groups, but for the remainder of the analysis, Hydra_k2 was the only classification more deeply explored.

Part 6: Visreg : Look at results of linear model graphically -Allows you to visualize each cluster by cognitive measure

Part 7 : Anova -Anovas were also run on the results of the LM of each cnb value by cluster.

Part 8 : FDR Correction -FDR correction was calculated for each cnb measure ANOVA output -A table of the results was extracted

```
## Loading required package: nlme
## This is mgcv 1.8-22. For overview type 'help("mgcv-package")'.
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:nlme':
##   collapse
##
## The following objects are masked from 'package:stats':
##   filter, lag
##
## The following objects are masked from 'package:base':
##   intersect, setdiff, setequal, union
## Loading required package: Formula
```

```

## 
## Attaching package: 'plm'
## The following objects are masked from 'package:dplyr':
##   between, lag, lead
##
## Attaching package: 'reshape'
## The following objects are masked from 'package:tidyverse':
##   expand, smiths
## The following object is masked from 'package:dplyr':
##   rename

```

PREPPING DATA (part 1 and 2)

(read in csvs, merge with hydra)

Part 3: Demographics

Stratified by Cluster			
	level	-1	1
## n		712	376
## Race (%)	Caucasian	393 (55.2)	155 (41.2)
##	Non-caucasian	319 (44.8)	221 (58.8)
## Sex (%)	Female	477 (67.0)	270 (71.8)
##	Male	235 (33.0)	106 (28.2)
## Maternal Ed (mean (sd))		14.14 (2.26)	13.75 (2.23)
## Age (mean (sd))		16.11 (2.99)	15.66 (3.14)
## Depression (%)	Depressed	0 (0.0)	376 (100.0)
##	Non-depressed	712 (100.0)	0 (0.0)
## Cluster (%)	-1	712 (100.0)	0 (0.0)
##	1	0 (0.0)	376 (100.0)
##	2	0 (0.0)	0 (0.0)
Stratified by Cluster			
	2	p	test
## n	336		
## Race (%)	238 (70.8)	<0.001	
##	98 (29.2)		
## Sex (%)	207 (61.6)	0.015	
##	129 (38.4)		
## Maternal Ed (mean (sd))	14.53 (2.29)	<0.001	
## Age (mean (sd))	16.66 (2.50)	<0.001	
## Depression (%)	336 (100.0)	<0.001	
##	0 (0.0)		
## Cluster (%)	0 (0.0)	<0.001	
##	0 (0.0)		
##	336 (100.0)		
Stratified by Cluster			

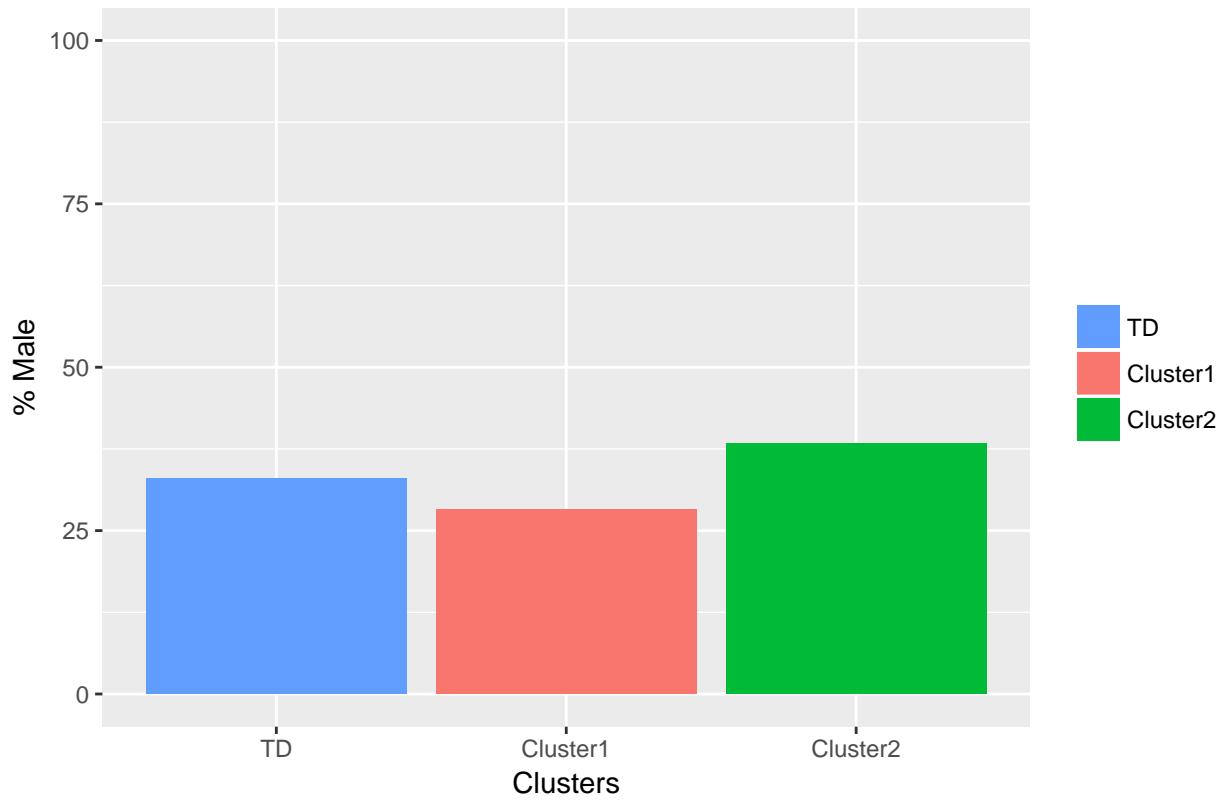
```

##          level      -1        1
##    n            2305      376
##    Race (%)     Caucasian  1475 ( 64.0)  177 ( 47.1)
##                  Non-caucasian 830 ( 36.0)  199 ( 52.9)
##    Sex (%)      Female    1107 ( 48.0)  264 ( 70.2)
##                  Male      1198 ( 52.0)  112 ( 29.8)
##    Maternal Ed (mean (sd)) 14.93 (2.45) 13.89 (2.24)
##    Age (mean (sd))       13.83 (3.71) 16.27 (2.84)
##    Depression (%)       Depressed   0 ( 0.0)  376 (100.0)
##                  Non-depressed 2305 (100.0) 0 ( 0.0)
##    Cluster (%)        -1        2305 (100.0) 0 ( 0.0)
##                  1           0 ( 0.0)  376 (100.0)
##                  2           0 ( 0.0)  0 ( 0.0)
##          Stratified by Cluster
##          2          p      test
##    n            341
##    Race (%)     218 ( 63.9) <0.001
##                  123 ( 36.1)
##    Sex (%)      217 ( 63.6) <0.001
##                  124 ( 36.4)
##    Maternal Ed (mean (sd)) 14.37 (2.31) <0.001
##    Age (mean (sd))       15.96 (2.95) <0.001
##    Depression (%)       341 (100.0) <0.001
##                  0 ( 0.0)
##    Cluster (%)       0 ( 0.0) <0.001
##                  0 ( 0.0)
##                  341 (100.0)
##          Stratified by Cluster
##          level      -1        1
##    n            2305      346
##    Race (%)     Caucasian  1475 ( 64.0)  211 ( 61.0)
##                  Non-caucasian 830 ( 36.0)  135 ( 39.0)
##    Sex (%)      Female    1107 ( 48.0)  219 ( 63.3)
##                  Male      1198 ( 52.0)  127 ( 36.7)
##    Maternal Ed (mean (sd)) 14.93 (2.45) 14.34 (2.31)
##    Age (mean (sd))       13.83 (3.71) 16.02 (3.03)
##    Depression (%)       Depressed   0 ( 0.0)  346 (100.0)
##                  Non-depressed 2305 (100.0) 0 ( 0.0)
##    Cluster (%)        -1        2305 (100.0) 0 ( 0.0)
##                  1           0 ( 0.0)  346 (100.0)
##                  2           0 ( 0.0)  0 ( 0.0)
##          Stratified by Cluster
##          2          p      test
##    n            371
##    Race (%)     184 ( 49.6) <0.001
##                  187 ( 50.4)
##    Sex (%)      262 ( 70.6) <0.001
##                  109 ( 29.4)
##    Maternal Ed (mean (sd)) 13.92 (2.25) <0.001
##    Age (mean (sd))       16.21 (2.77) <0.001
##    Depression (%)       371 (100.0) <0.001
##                  0 ( 0.0)
##    Cluster (%)       0 ( 0.0) <0.001

```

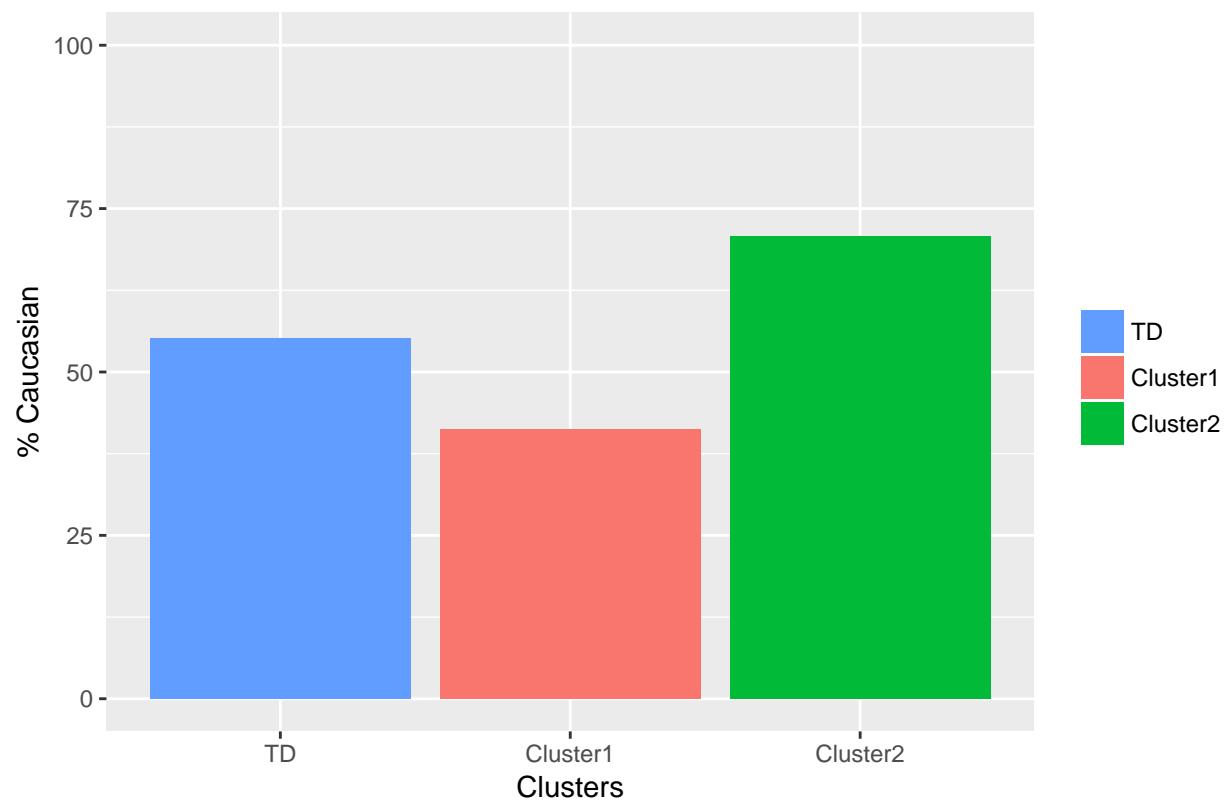
```
##          0 ( 0.0)
## 371 (100.0)
```

Hydra_k2 % Male, p = 0.0154



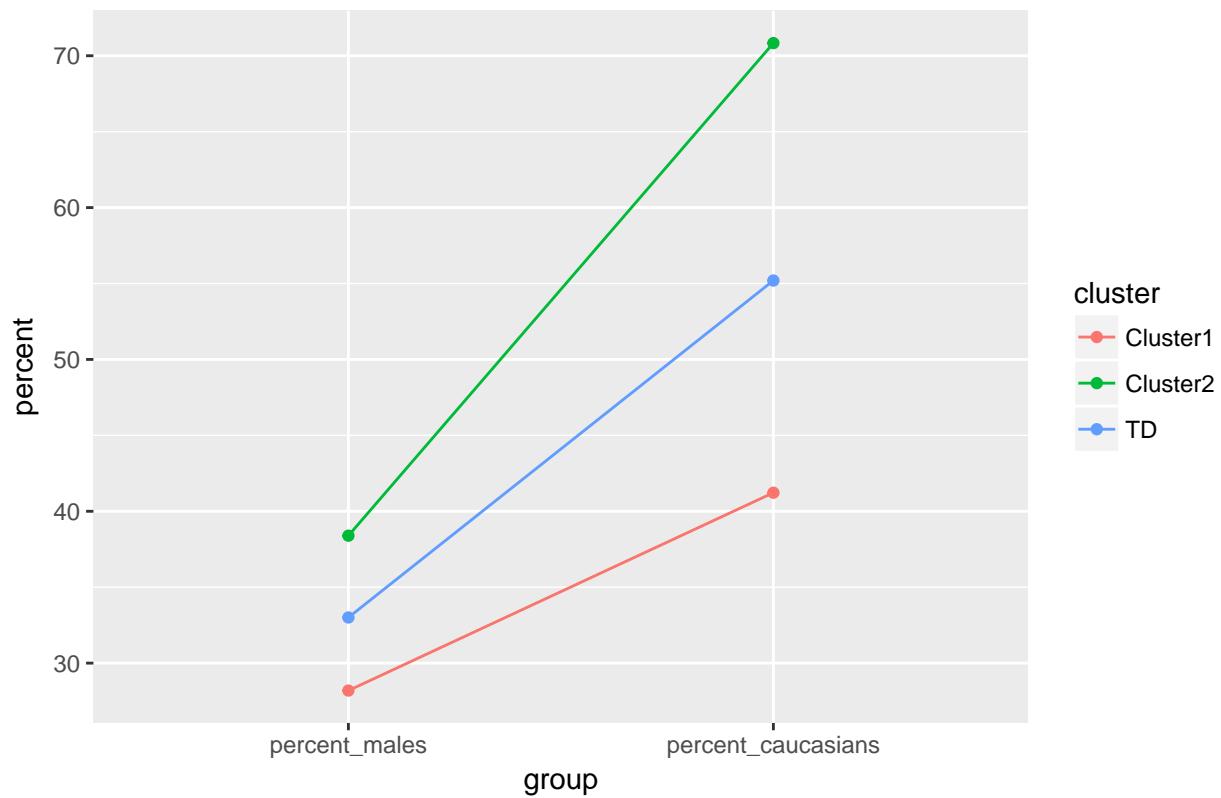
Part 4: Graphing

Hydra_k2 % Caucasian, p = 2.187e-14



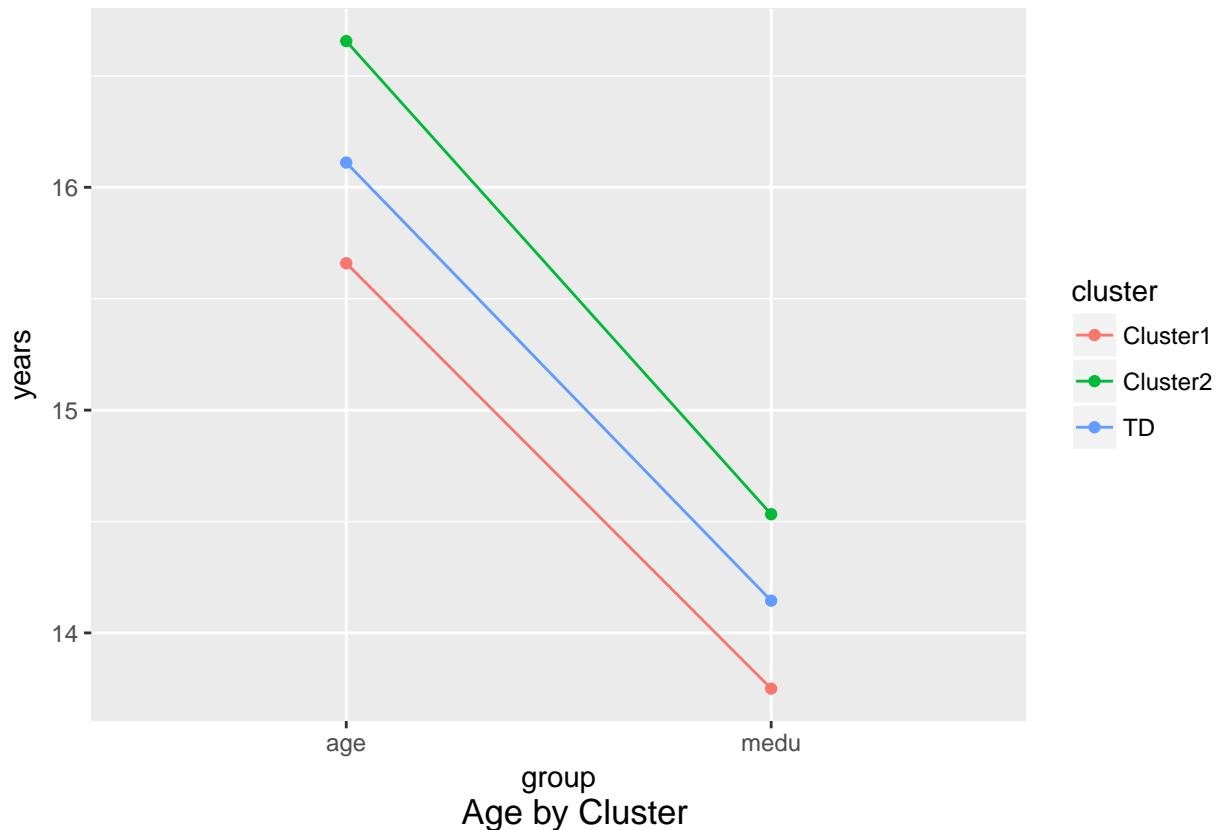
Using cl as id variables

Hydra_k2 Percentages

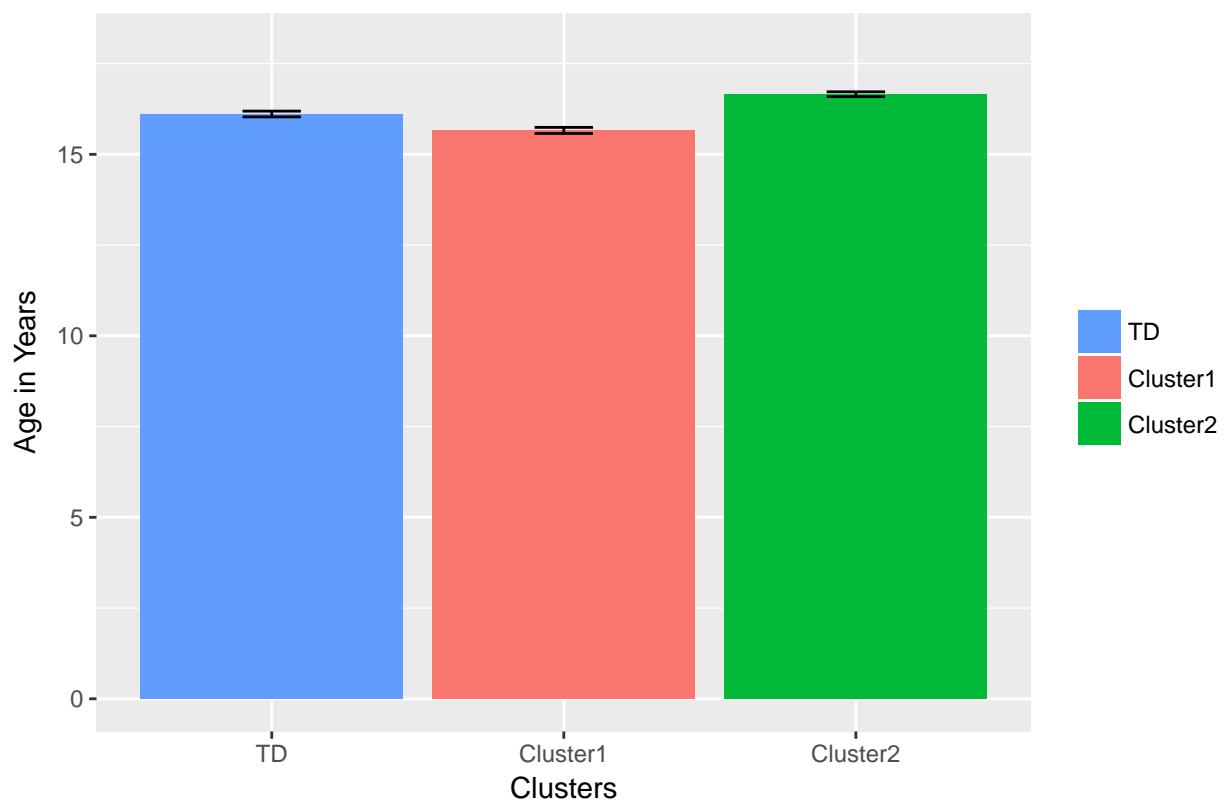


```
## Using cl as id variables
```

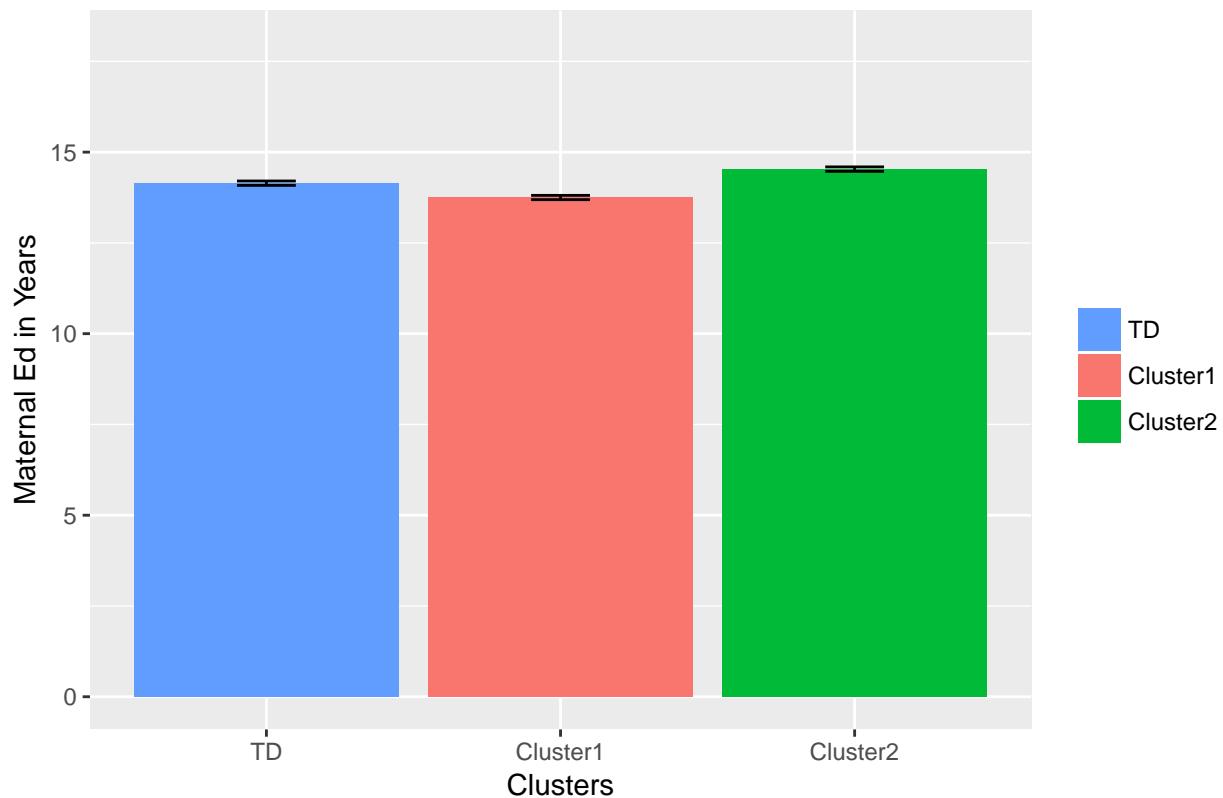
Hydra_k2 Ages



Age by Cluster



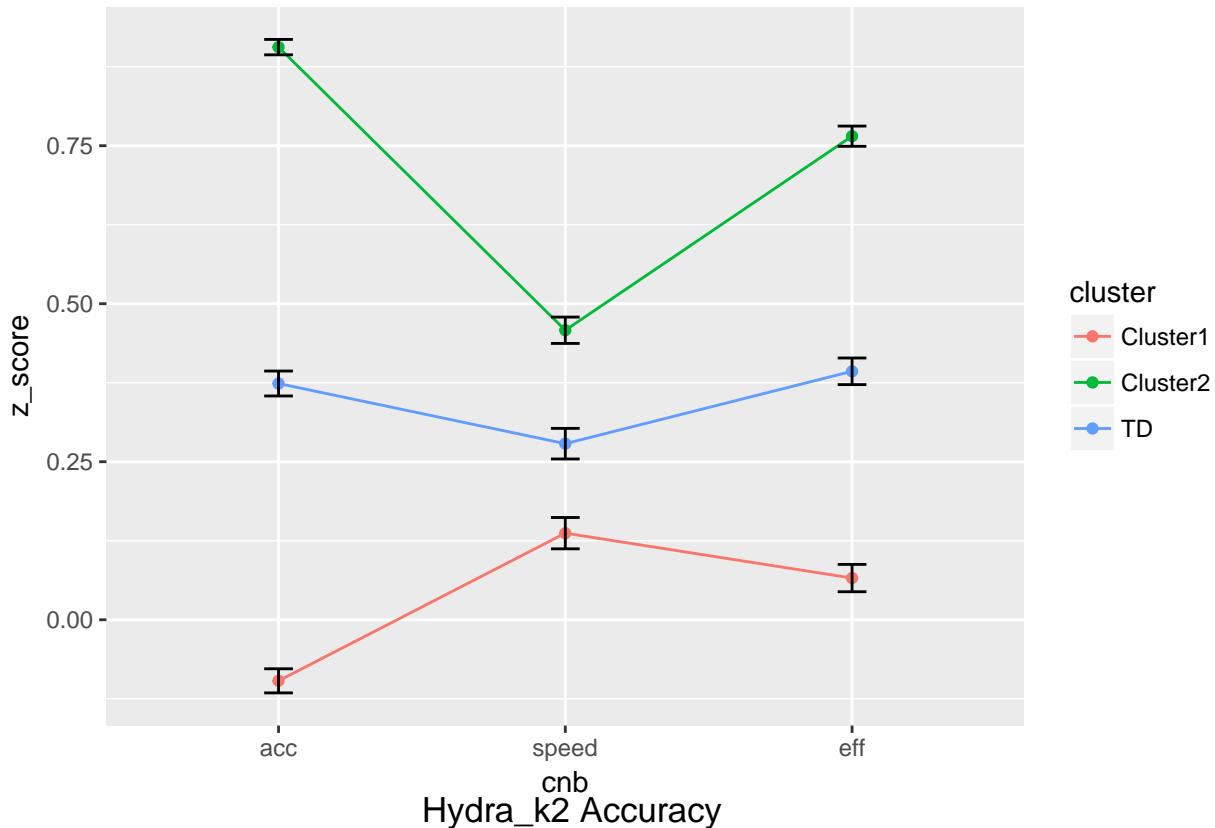
Maternal Edu by Cluster



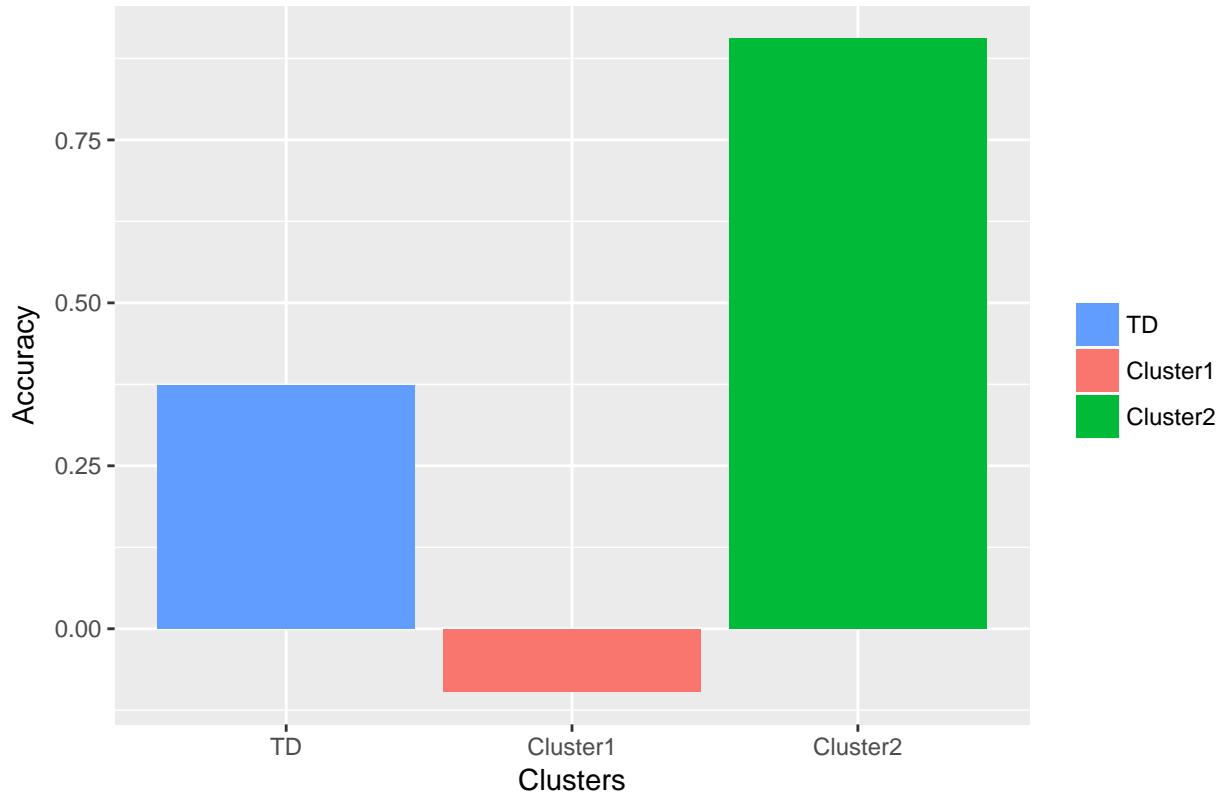
Part 5-8: linear model with visreg, anova, and FDR correction

```
##                                     TD   Cluster 1 Cluster 2
## df_mean_accuracy_z      0.3737861 -0.09663631 0.9059611
## df_mean_processing_speed_z 0.2785680  0.13700370 0.4580268
## df_mean_efficiency_z     0.3930667  0.06595696 0.7650560
## Using cl as id variables
```

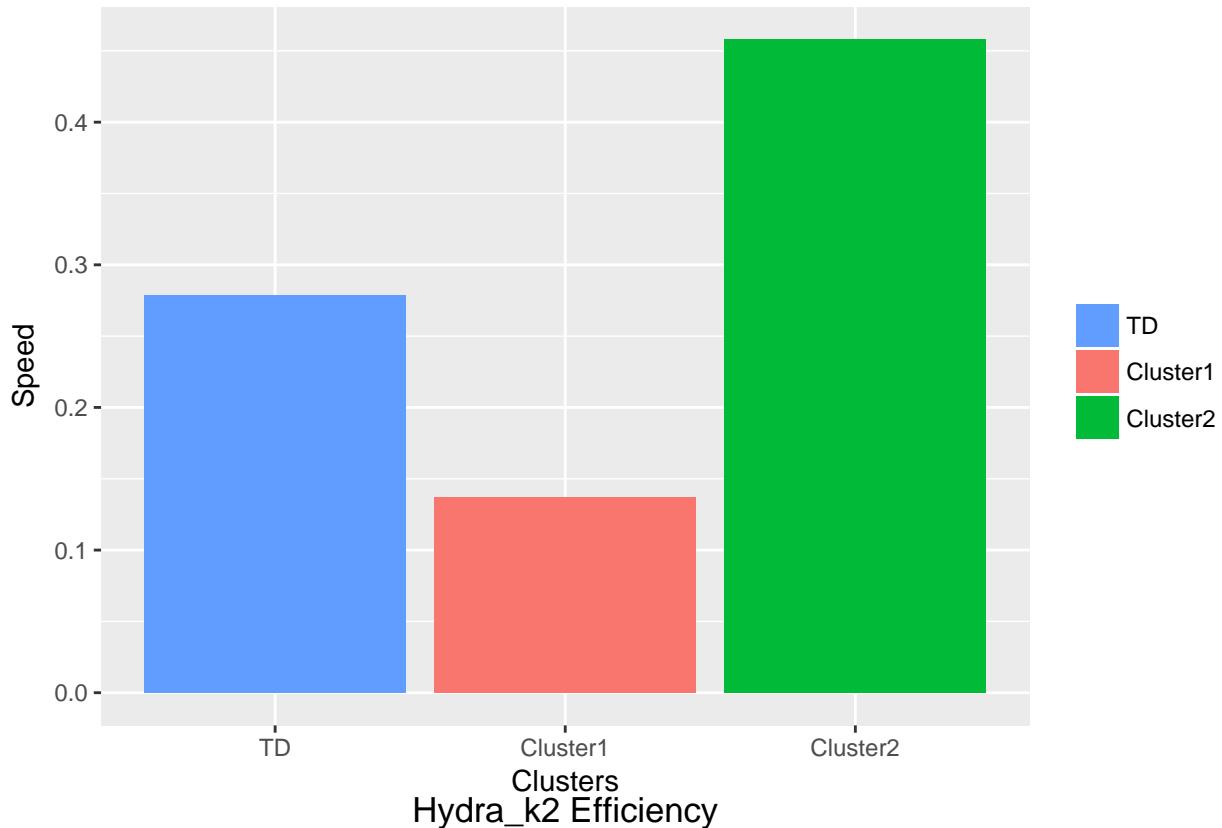
Hydra_k2 Cognitive Measures



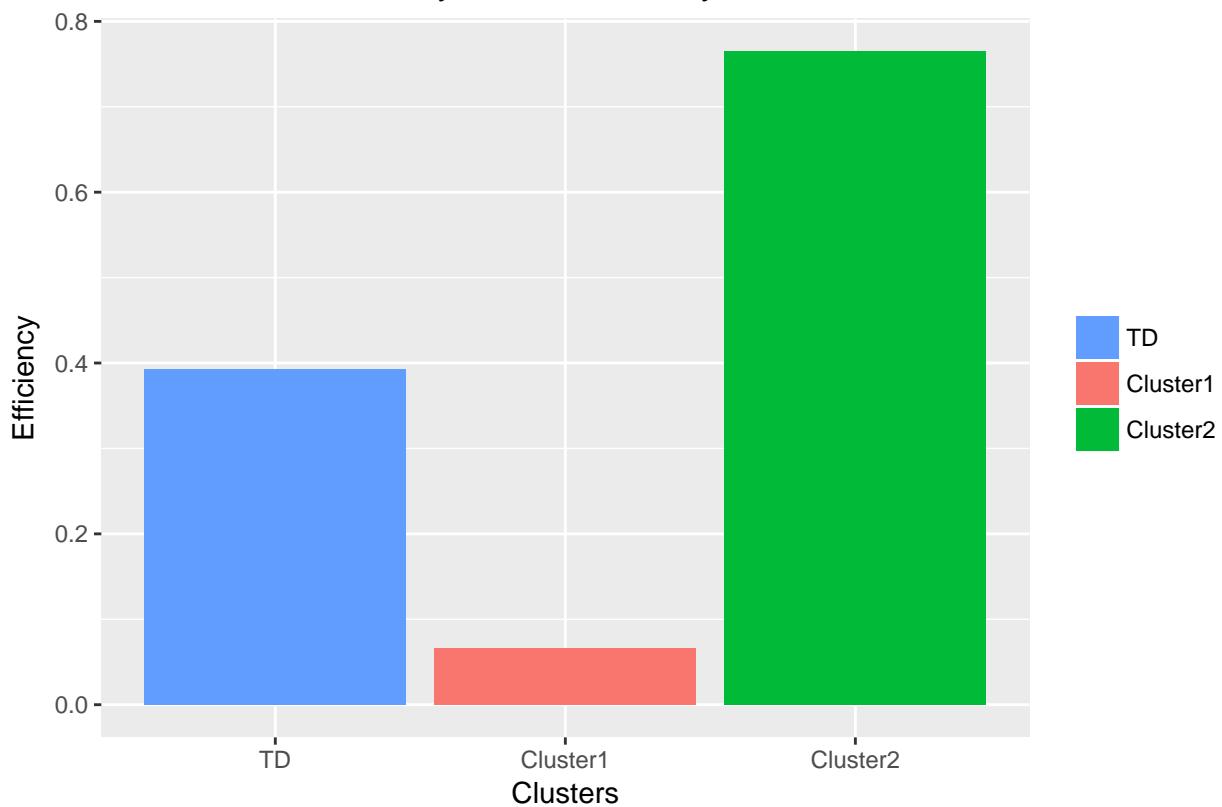
Hydra_k2 Accuracy



Hydra_k2 Processing Speed



Hydra_k2 Efficiency

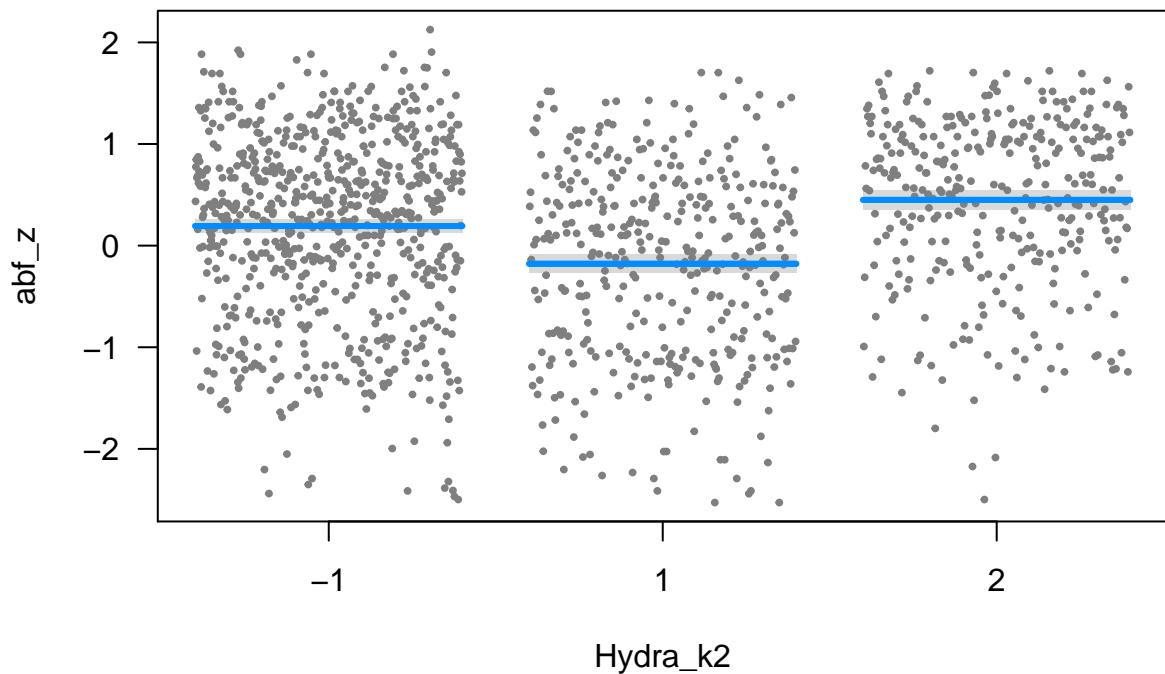


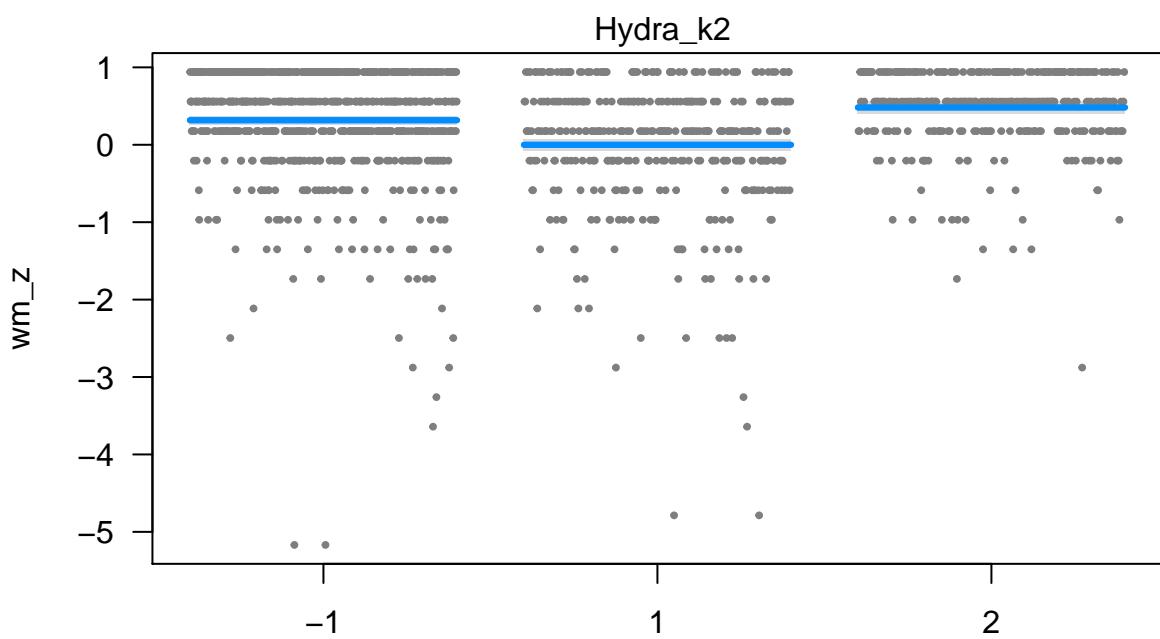
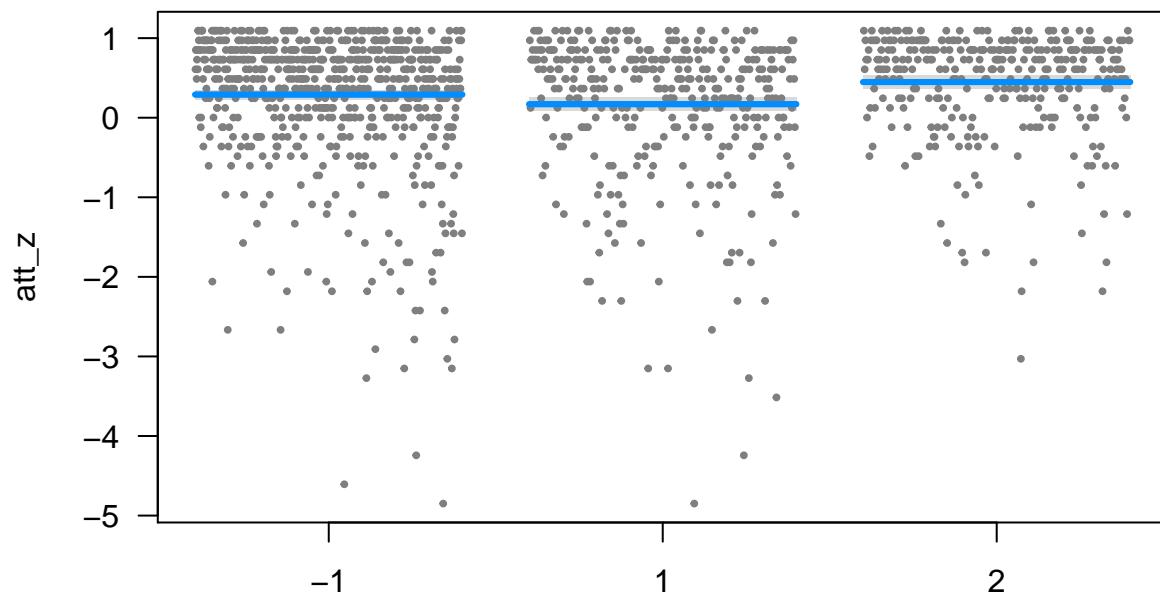
```
## [1] "CNB names and FDR correction values"
```

```

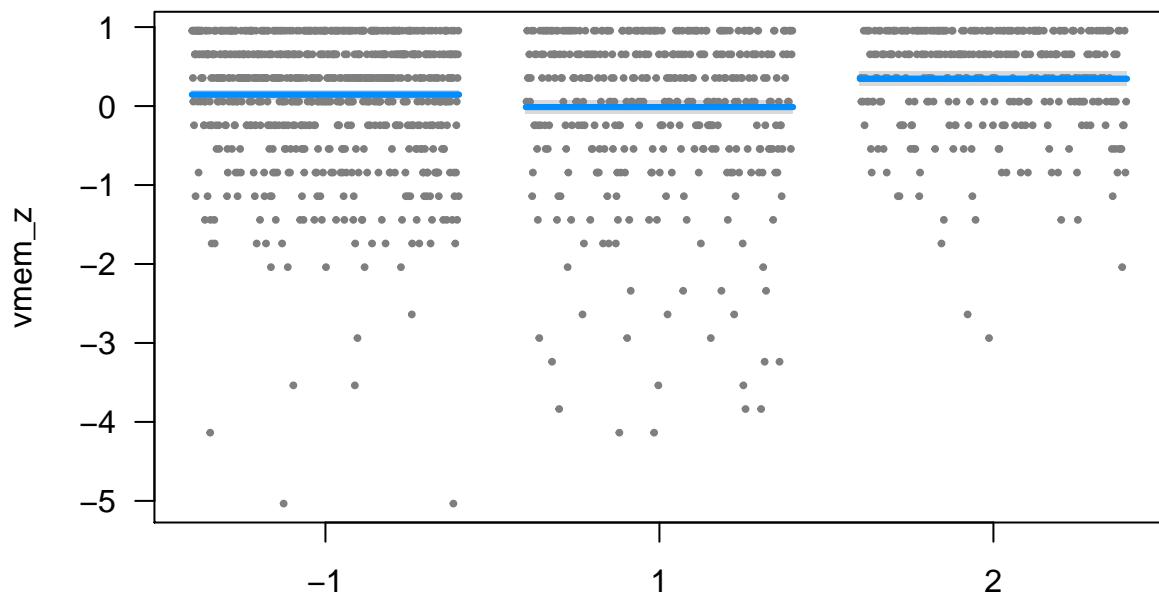
##      CNB_measure p_FDR_corr
## 1          abf_z      0
## 2          att_z      0
## 3          wm_z      0
## 4          vmem_z     0
## 5          fmem_z     0
## 6          smem_z     0
## 7          lan_z      0
## 8          nvr_z      0
## 9          spa_z      0
## 10         eid_z     0.003
## 11         edi_z      0
## 12         adi_z      0
## 13         abf_s_z     0
## 14         att_s_z    0.008
## 15         wm_s_z    0.001
## 16         vmem_s_z   0.002
## 17         smem_s_z   0.003
## 18         lan_s_z     0
## 19         nvr_s_z     0
## 20         eid_s_z    0.003
## 21         adi_s_z    0.041
## 22         mot_s_z     0

```

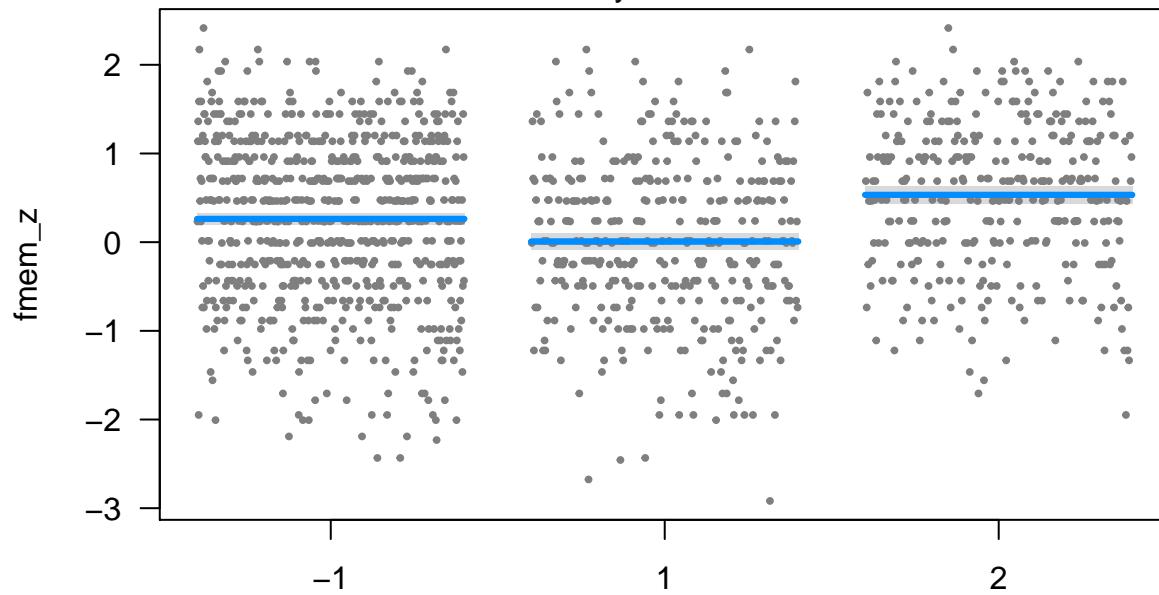




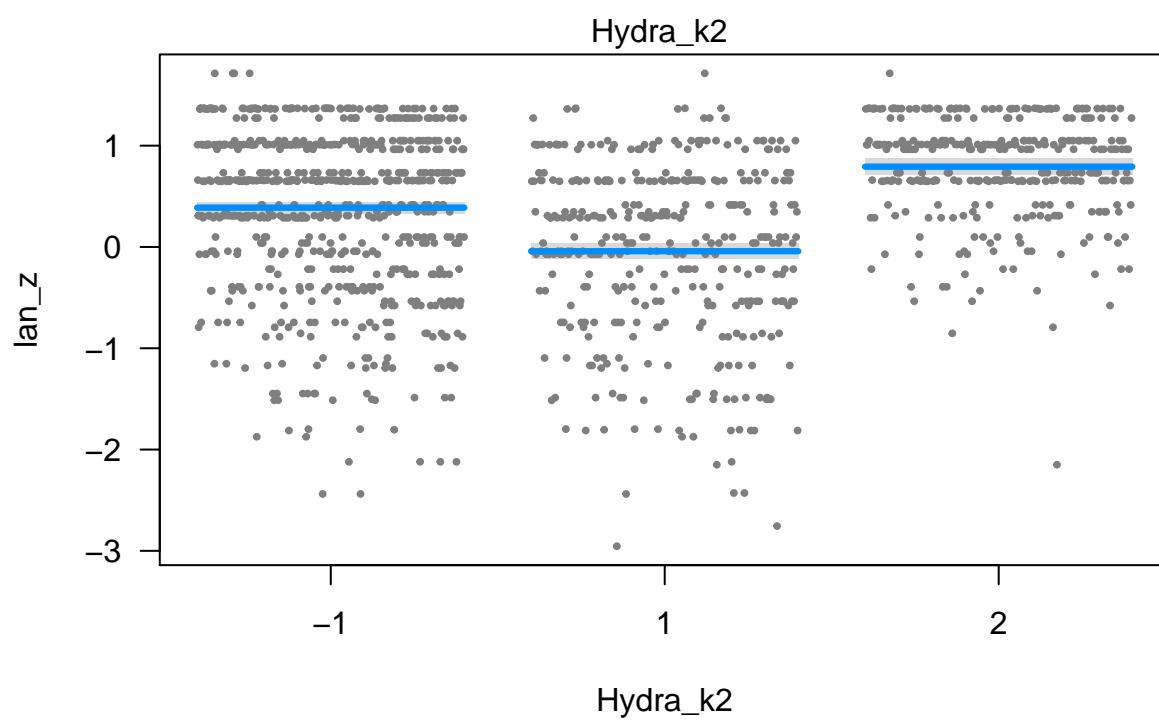
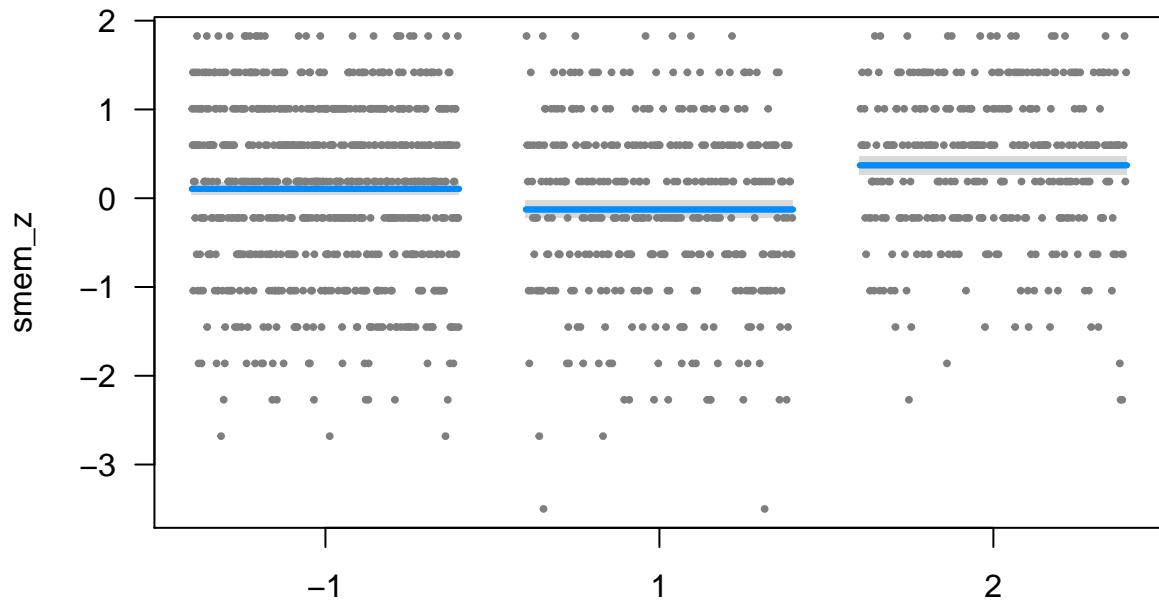
Hydra_k2

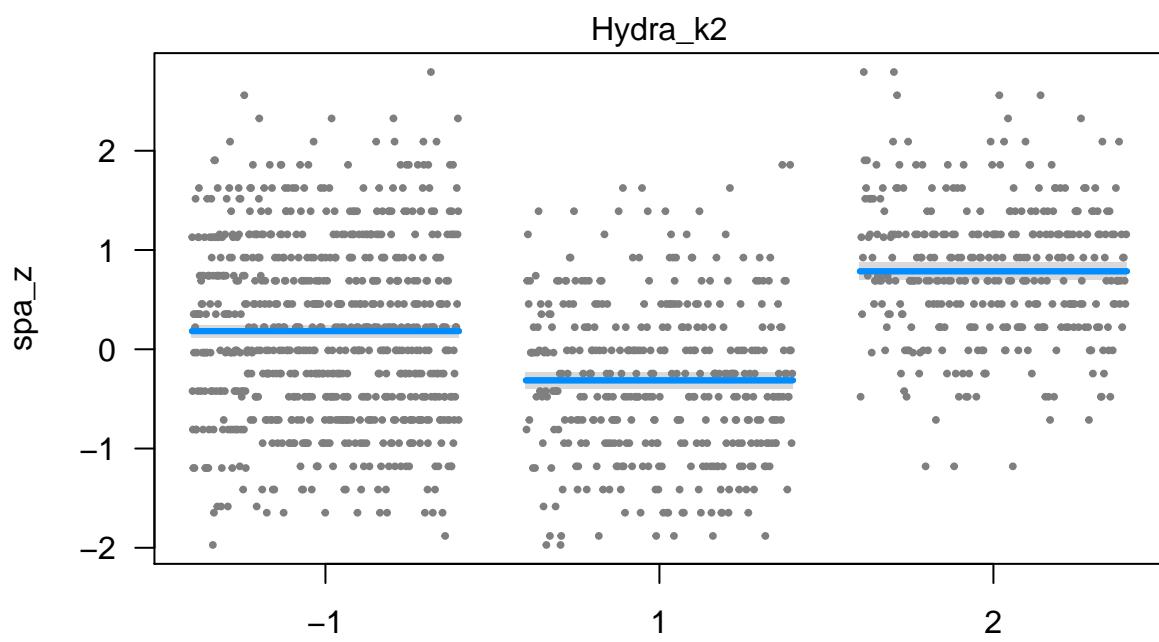
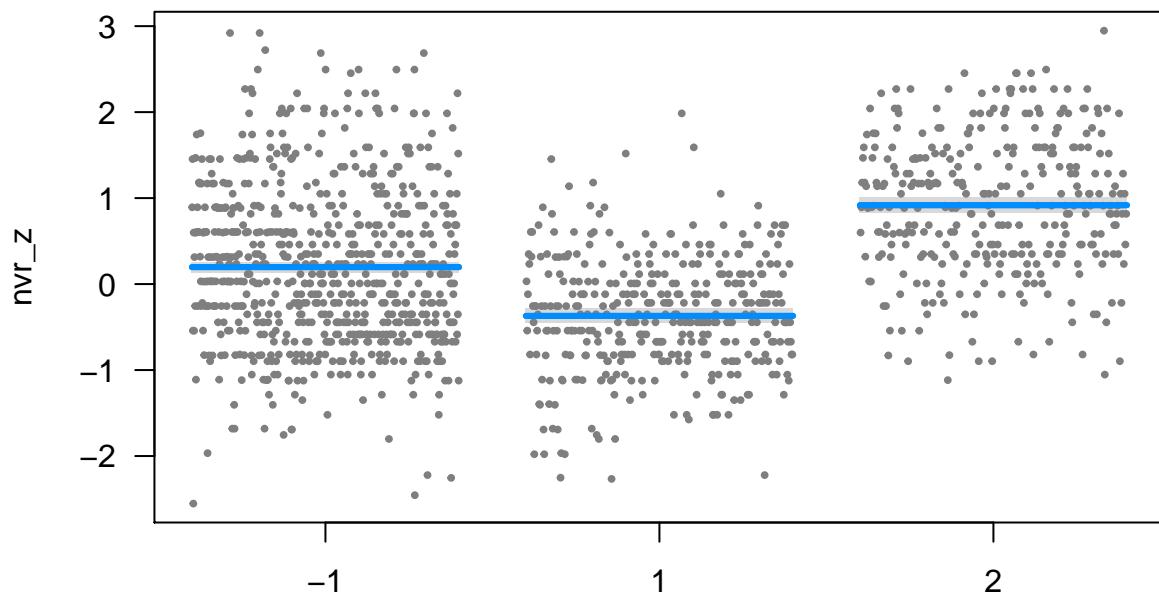


Hydra_k2



Hydra_k2





Hydra_k2

