

# Data Madness

*Martin Gassner, Henry Mauranen, Matteo Maggiolo*

*17 March 2018*

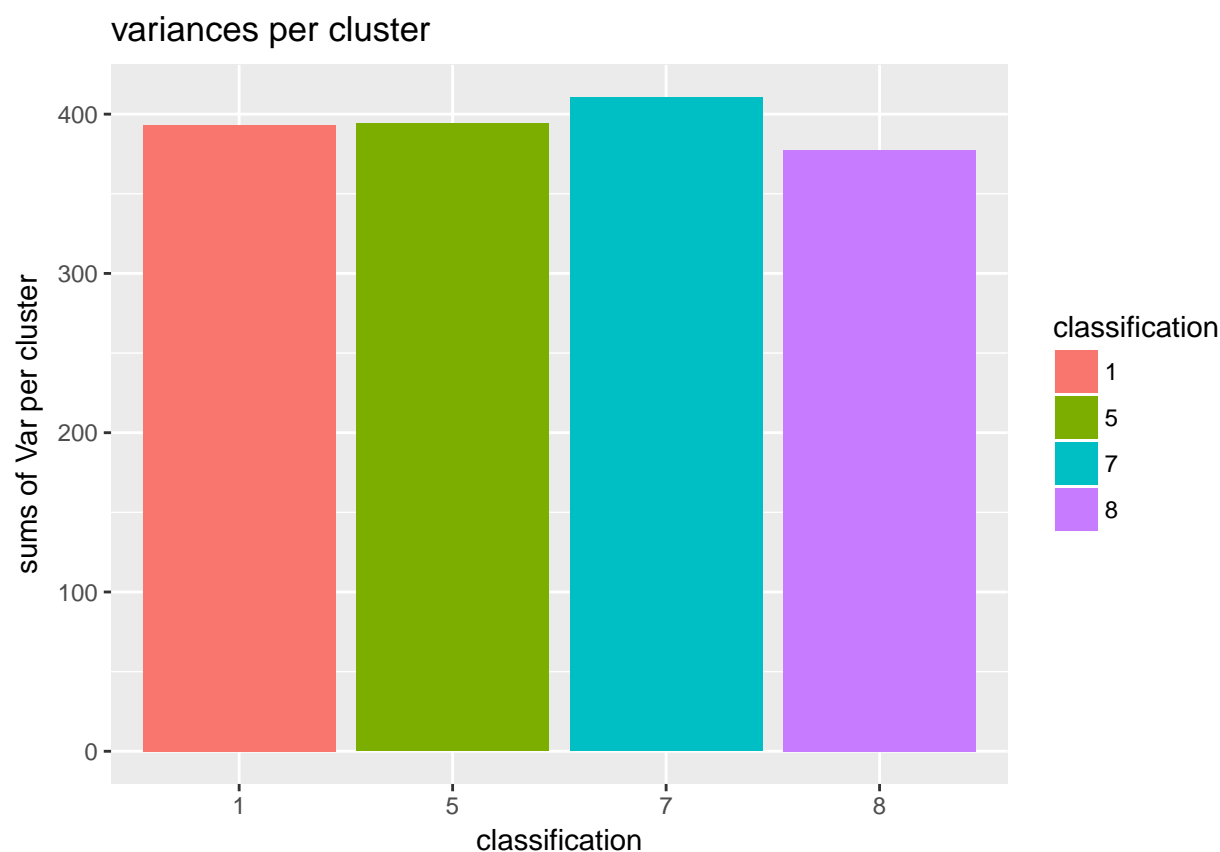
## Introduction

We worked on a dataset containing the recorded answers to a survey about morality.

The questions we would like to adress are as follows:

1. Can we find out ways to interpret the answers of each block as one value?
2. Can we identify distinct clusters within different blocks of questions?
3. Based on these values and/or clusters, are there clear correlations between interpretable blocks?
4. Are there interpretable types of individuals?
5. Can we link the clusters to personal metrics?

## EM clustering

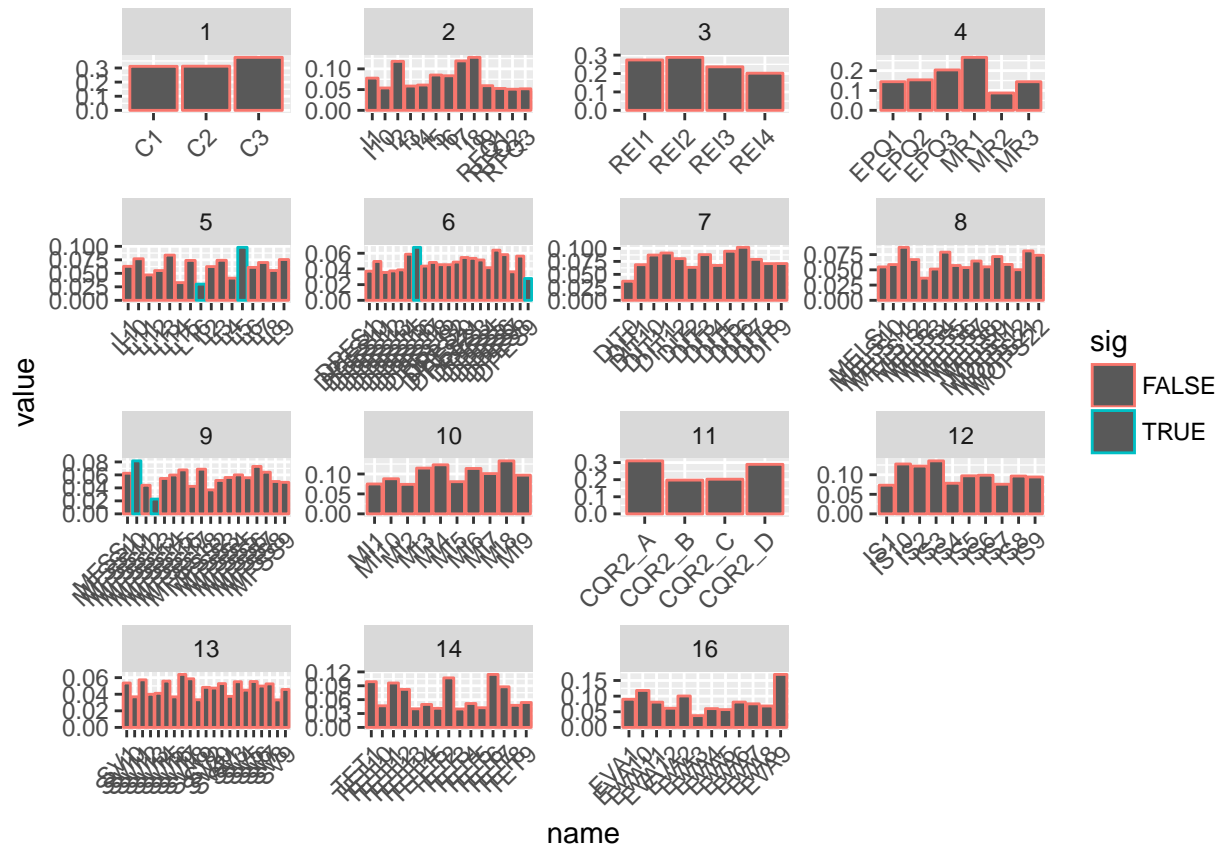


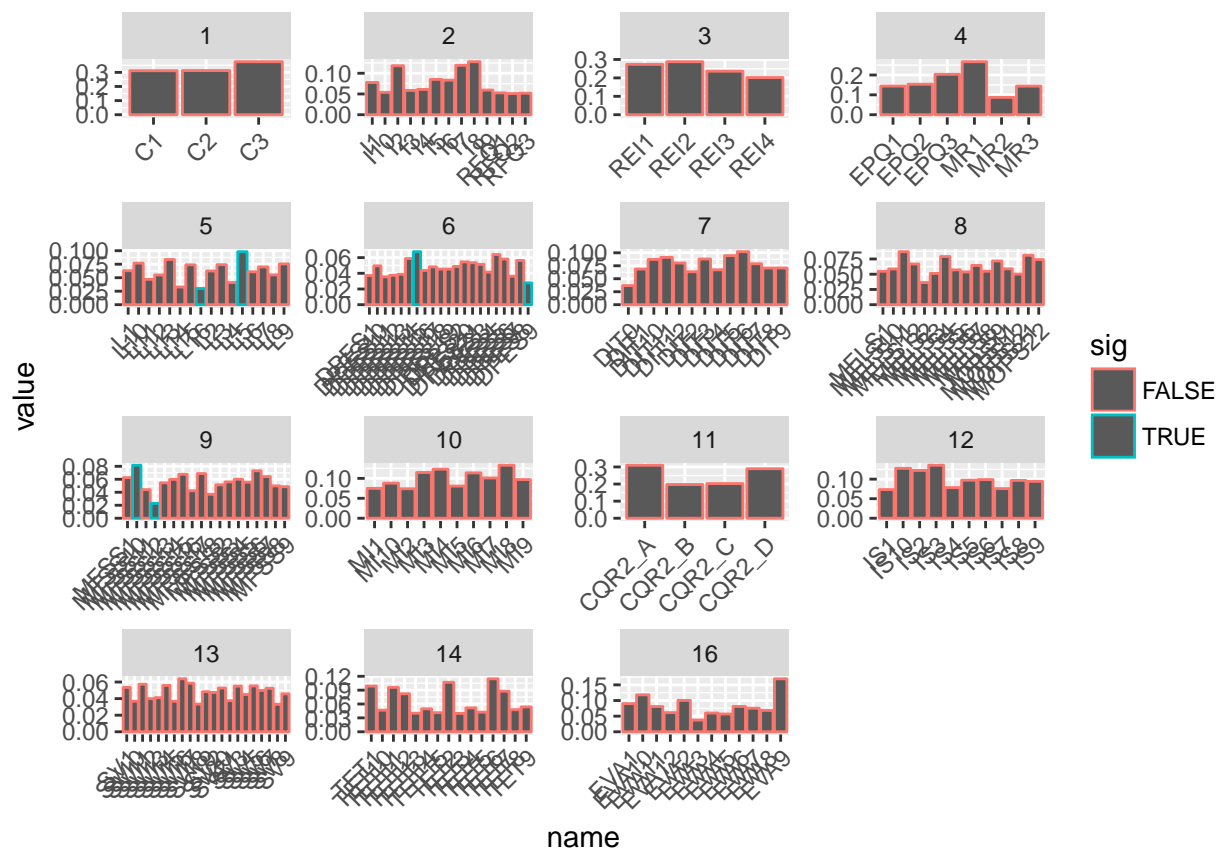
## PCA weighting + kmeans

To obtain a weighting for each block of questions

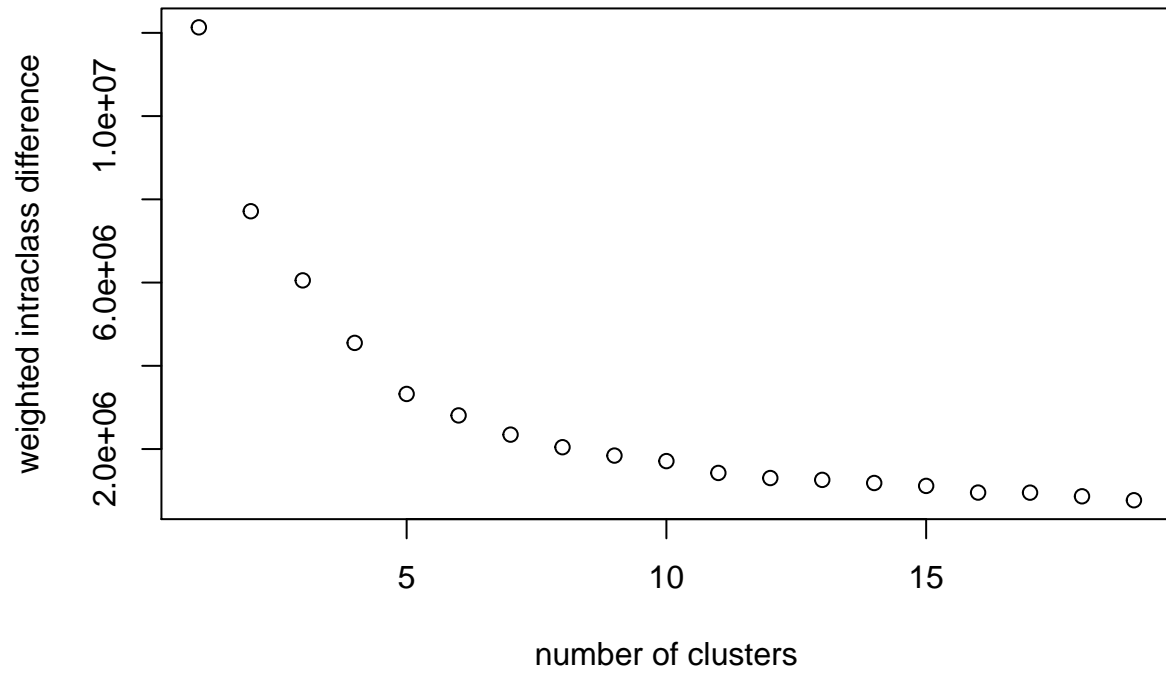
1. obtain the principal components for each block
2. keep the ones explaining 85% of the variance
3. for each variable: sum the loadings of the most important principal components, weighted by the variance explained.

Additionally we tested the weights for significance within each block.

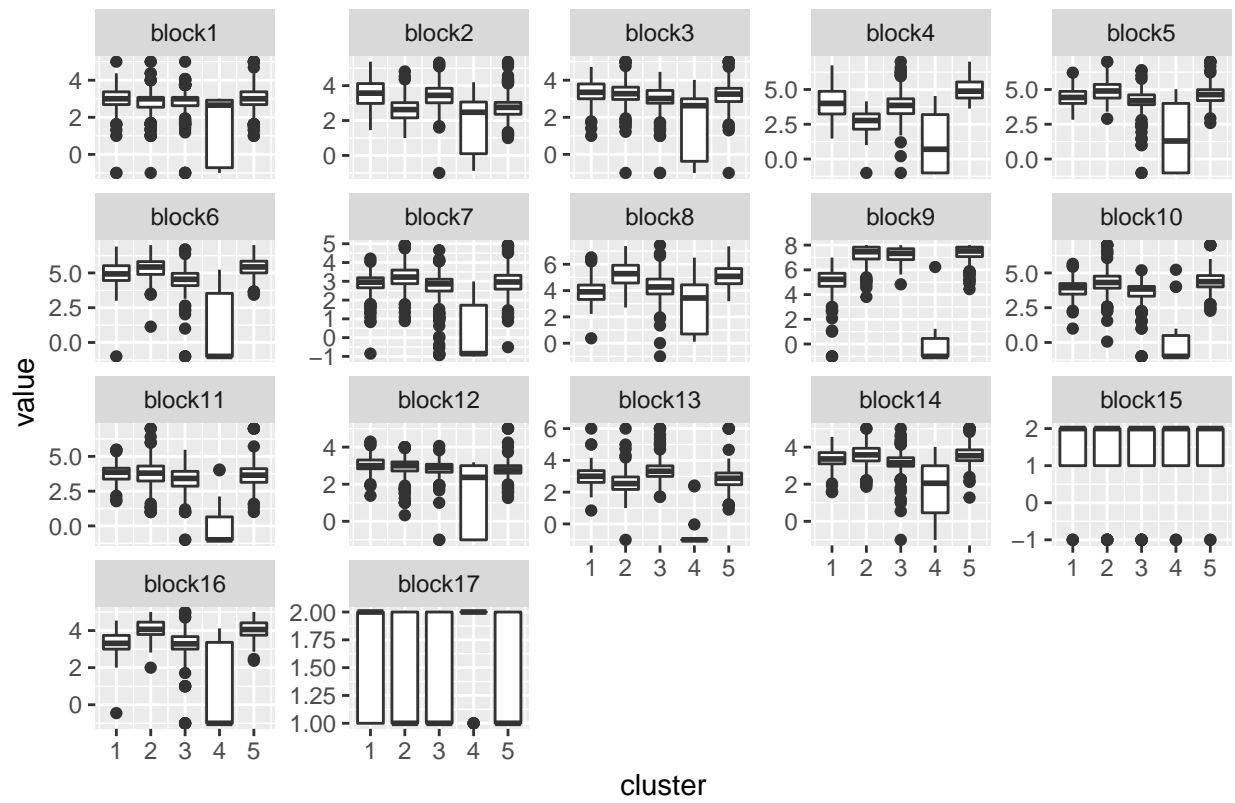




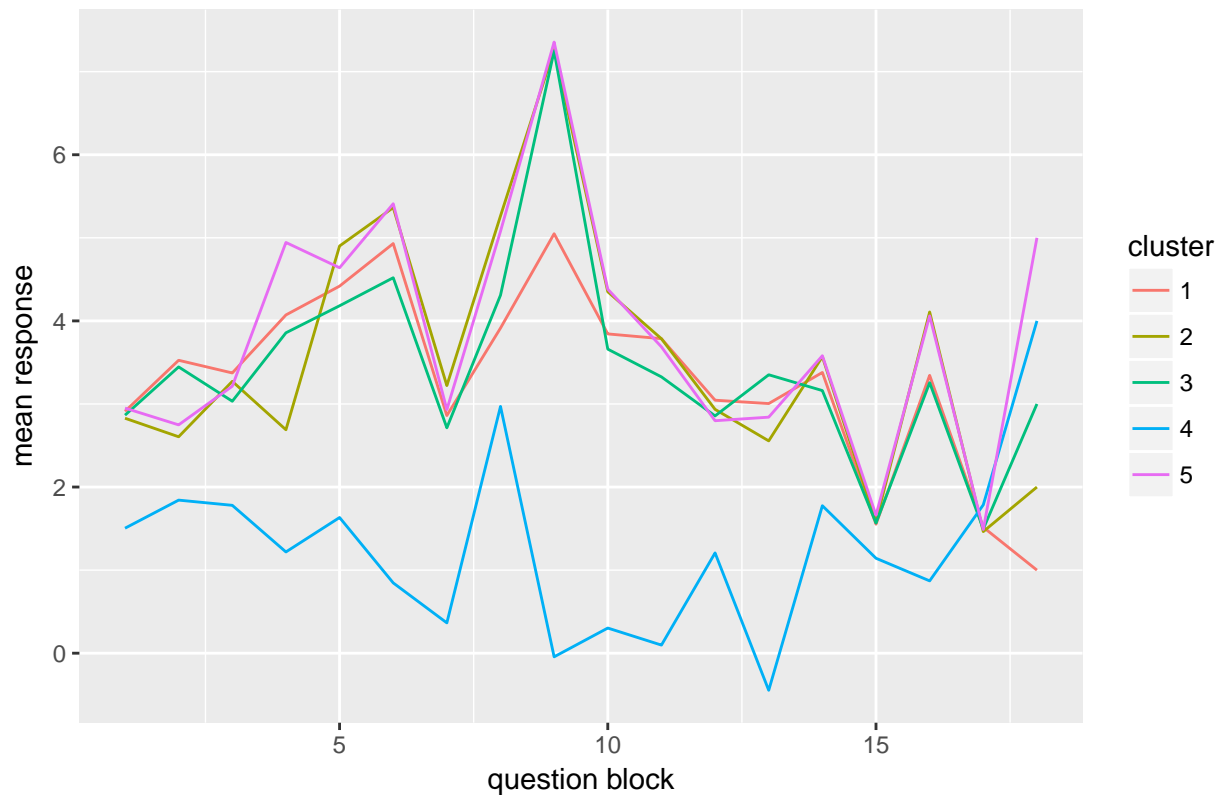
**difference per cluster size**



## distribution of responses



## Overview of mean responses



big differences Block 2: Identity measures - clusters 2,5 (less important) vs. 1,3 (very important)

highly weighted: job, political party, favorite sports team

Block 4: Moral relativism - cluster 2 (relative) vs 5 (disagreeing)

highly weighted: What is right and wrong is up to each person to decide

Block 9: Sacredness of moral principles - cluster 1 (less) vs 2,3,5 (more)

highly weighted: Sign a piece of paper that says "I hereby sell my soul, after my death, to whoever has this piece of paper" (SIGNIFICANT)

low weighted: Attend a performance art piece in which all participants (including you) have to act like animals for 30 minutes, including crawling around naked and urinating on stage

Block 17: all different; rank 1, 2, 3, 4, 5

```
## [1] "Block 2: Identity measures - clusters 2,5 (less important) vs. 1,3 (very important)"
```

```
## [1] 0.2268667
```

```
## [1] "Block 4: Moral relativism - cluster 2 (relative) vs 5 (disagreeing)"
```

```
## [1] 8.384556e-214
```

```
## [1] "Block 9: Sacredness of moral principles - cluster 1 (less) vs 2,3,5 (more)"
```

```
## [1] 0.004285064
```

```
## [1] "Block 17: Dictator game - cluster 1 vs 2, 3, 4, 5"
```

```
## [1] 0.001281282
```

```
## [1] "Block 17: Dictator game - cluster 2 vs 1, 3, 4, 5"
## [1] 5.086156e-153
## [1] "Block 17: Dictator game - cluster 3 vs 1, 2, 4, 5"
## [1] 0.7656305
## [1] "Block 17: Dictator game - cluster 4 vs 1, 2, 3, 5"
## [1] 0.0008385006
## [1] "Block 17: Dictator game - cluster 5 vs 1, 2, 3, 4"
## [1] 1.022094e-149
```

Cluster 1

- find many things (country, religion etc) relevant to their identity
- do not hold moral values high/more pragmatic
- give the least in the dictator game

Cluster 2

- find less things relevant to their identity
- moral relativists
- do think of their moral values sacred/would take a lot of money
- second to last in dictator game

Cluster 3

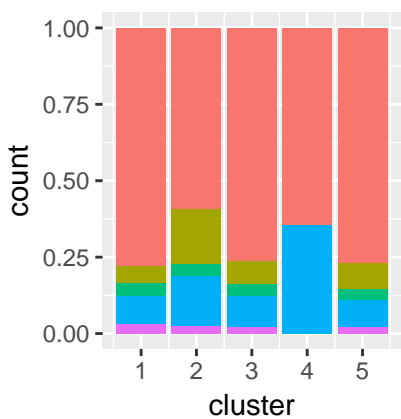
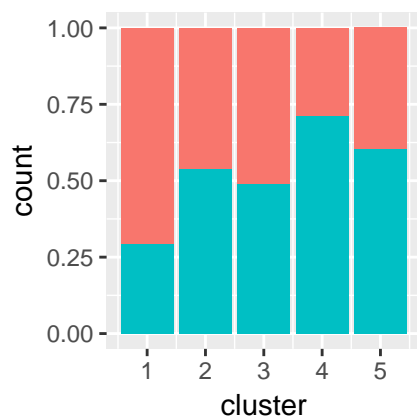
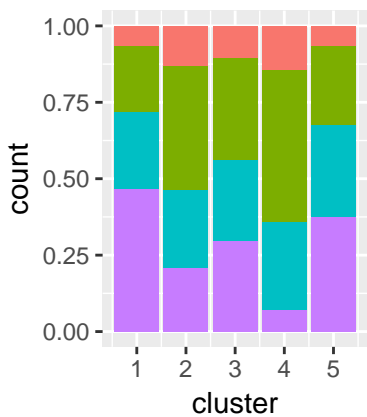
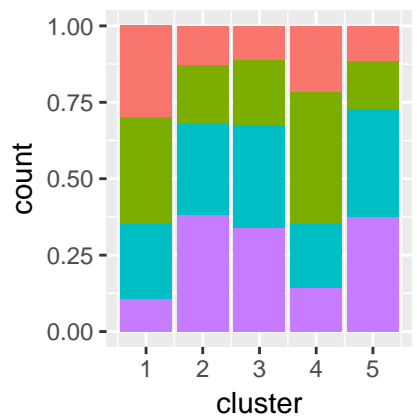
- average, really everywhere

Cluster 4

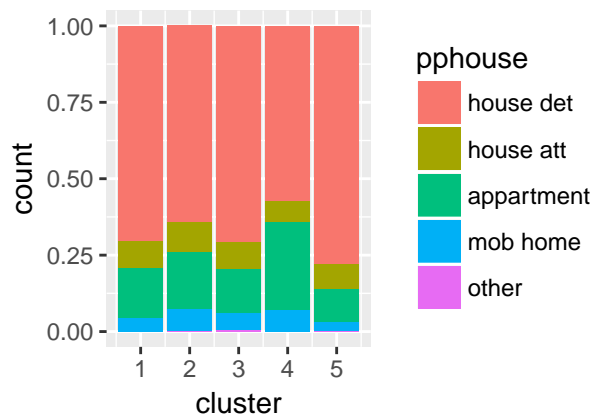
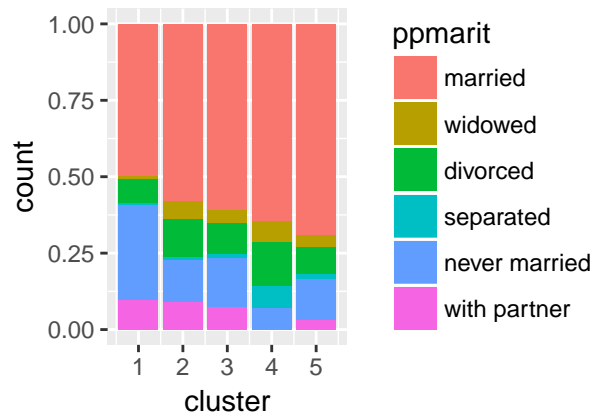
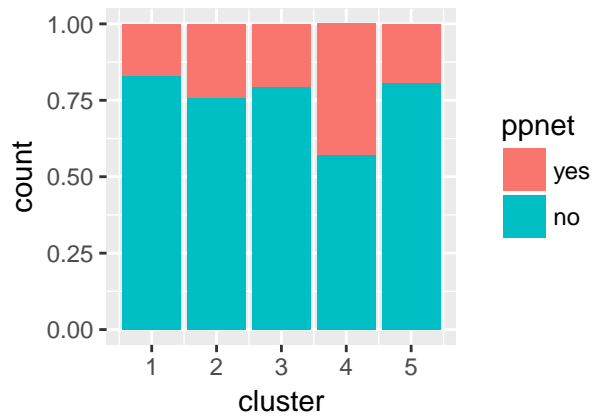
- irrelevant, too small, collects refused answers

Cluster 5

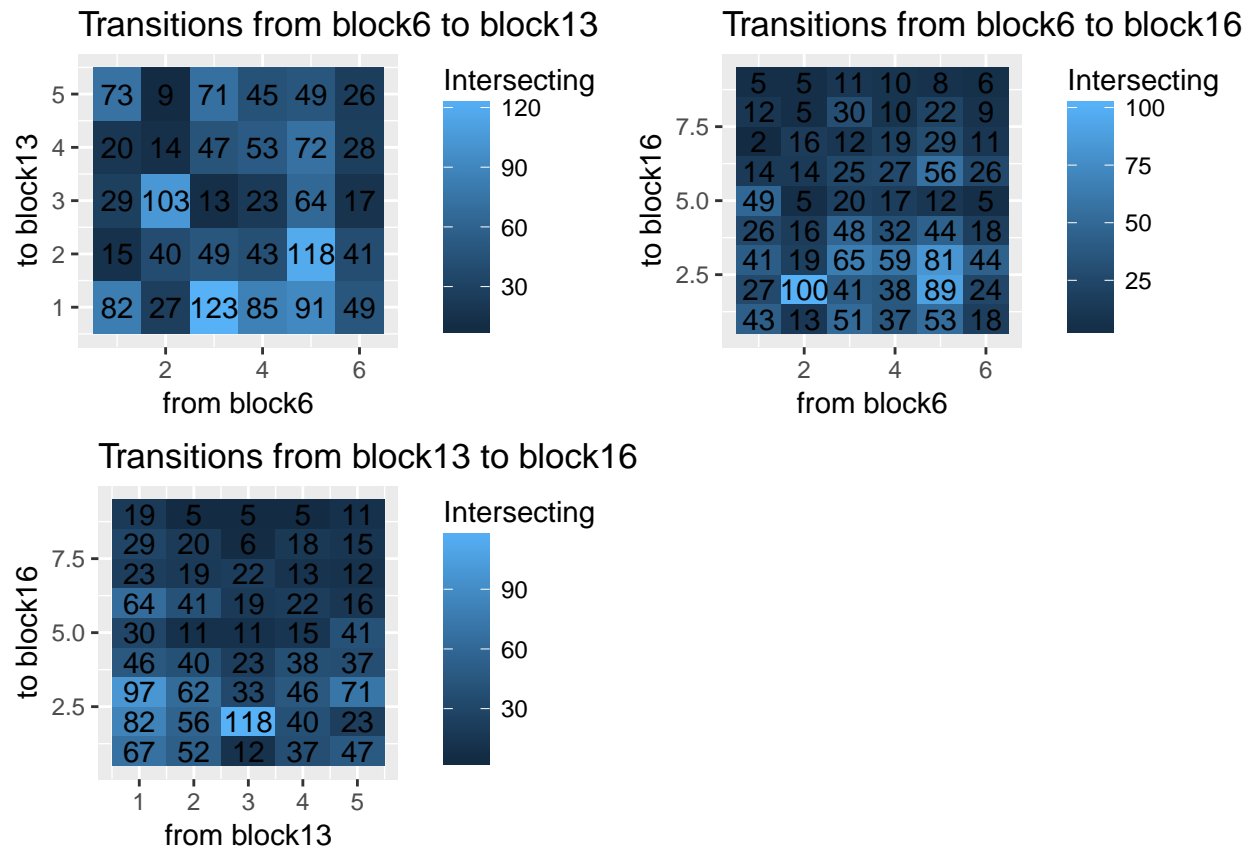
- find less things relevant to their identity
- oppose moral relativism
- do hold their morals dear
- give the most in the dictator game (exactly average)



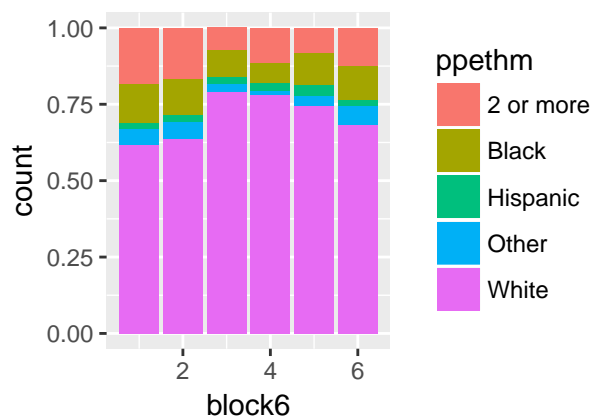
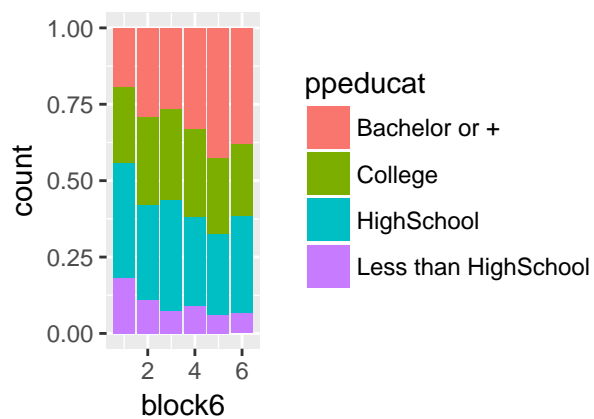
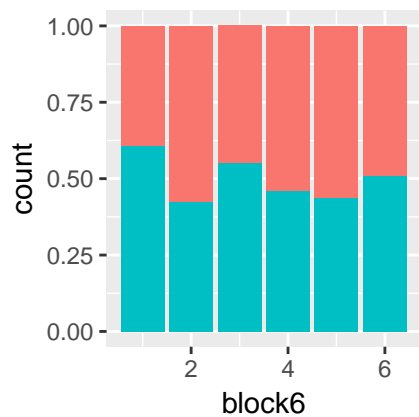
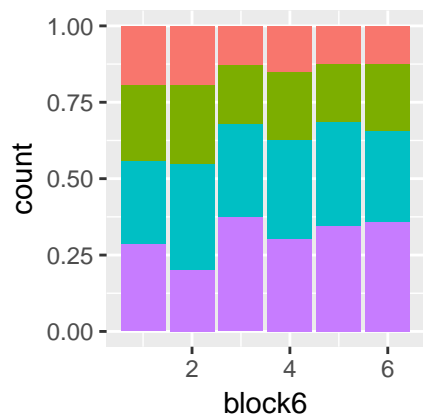




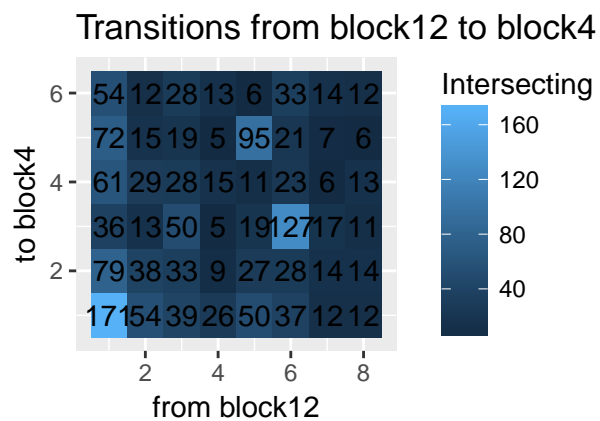
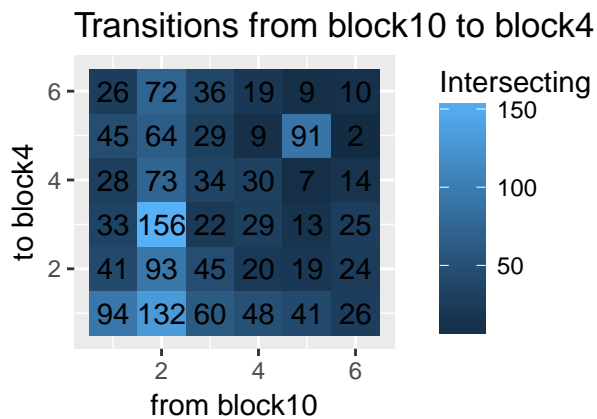
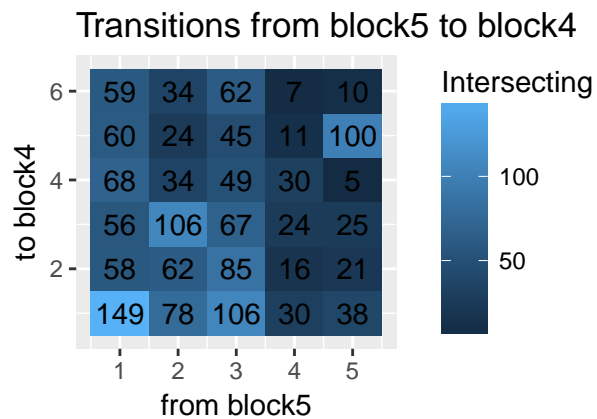
## k-modes



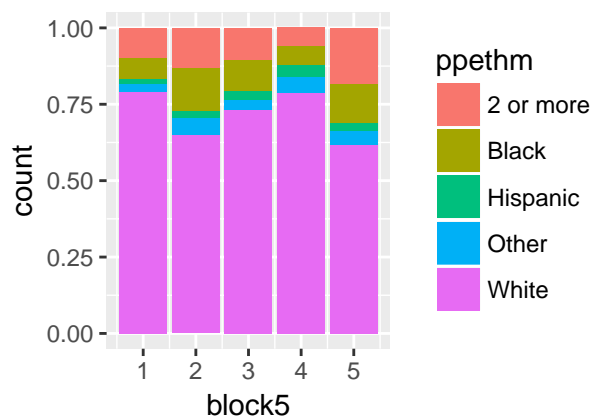
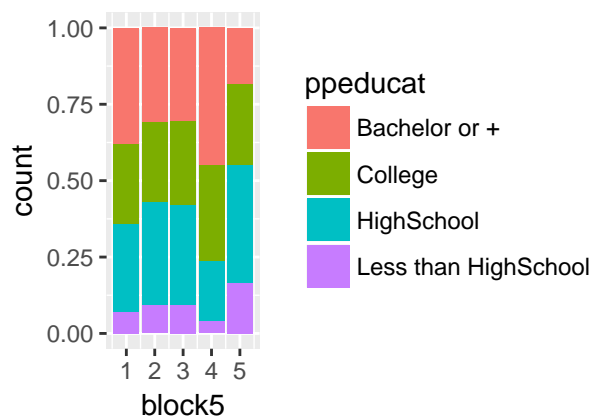
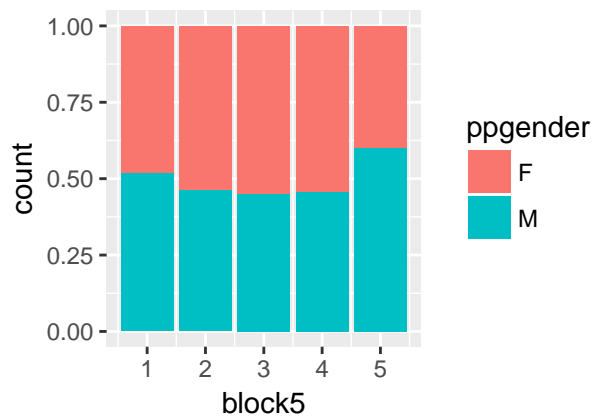
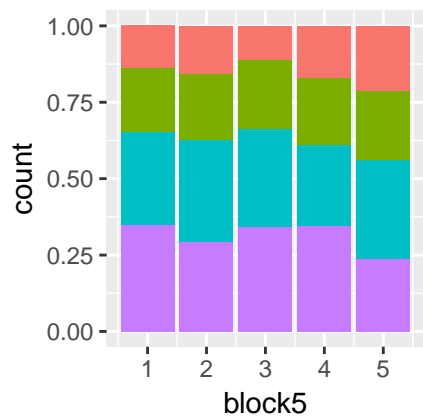
```
## [1] "Interesting transition from block6(2) to block13(3), from block16(2) [The 'hippies' (all 7 in b
```



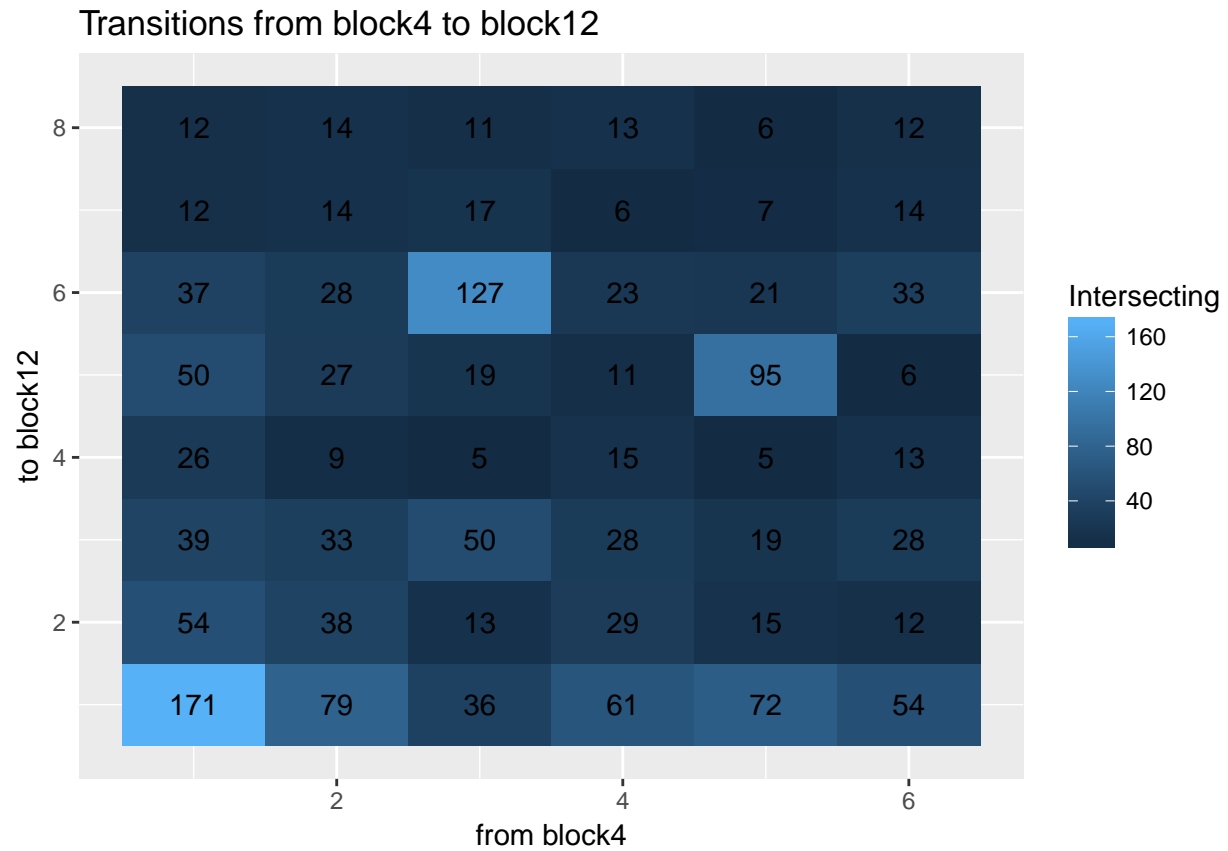
```
## [1] "Plotting only block 6 info [check cluster 2]"
```



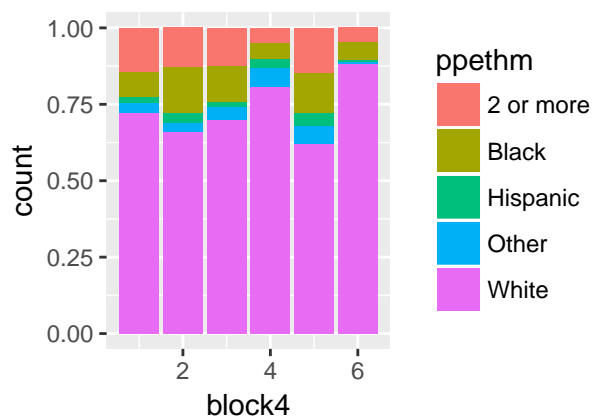
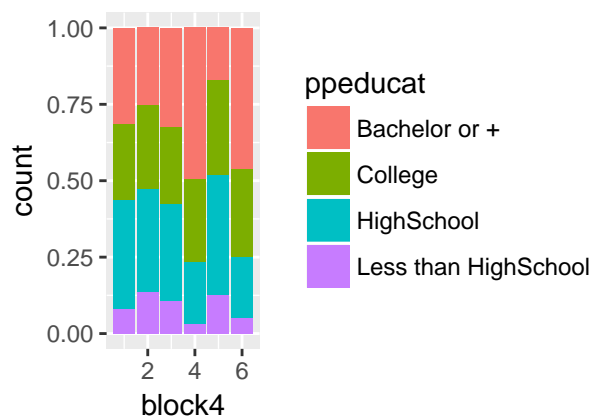
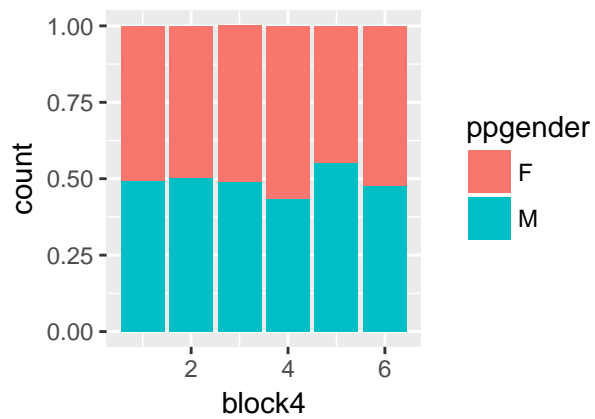
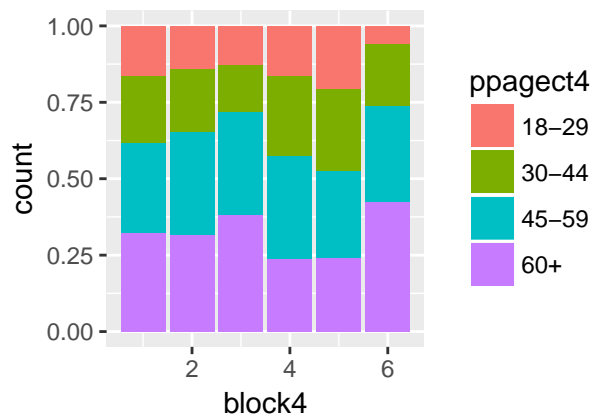
## [1] "Interesting transition from block5(5) to block4(5), block12(5) to block4(5), block10(5) to block4(5)"

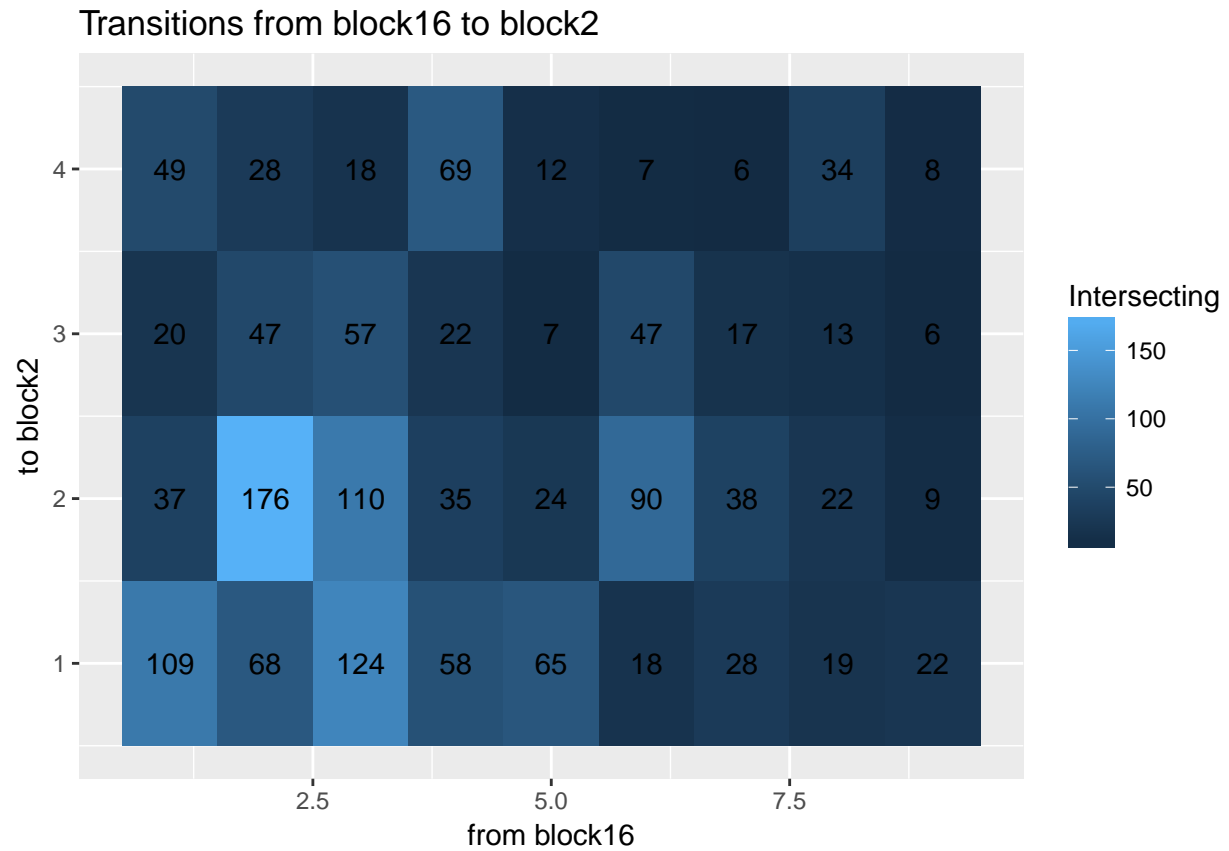


```
## [1] "Plotting only block 5 info [check cluster 5]"
```



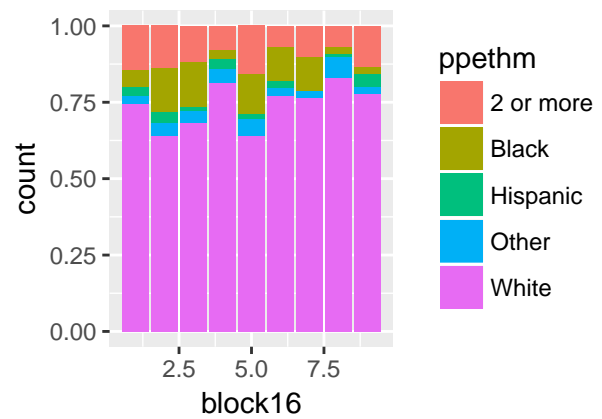
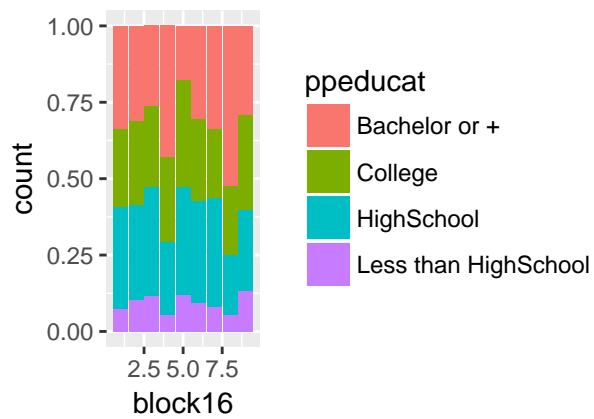
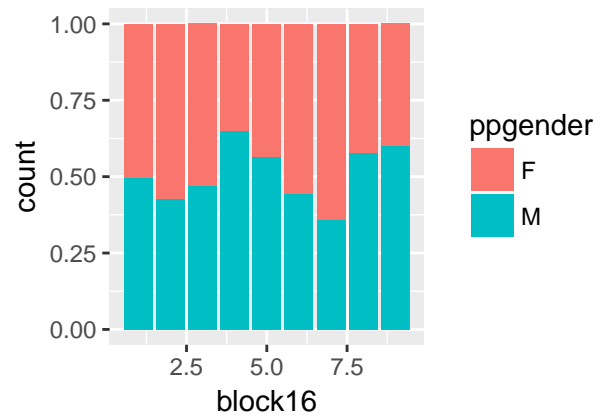
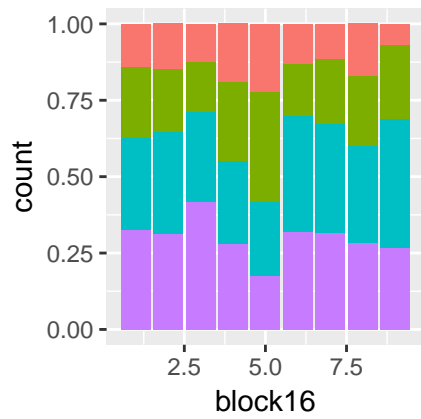
```
## [1] "Interesting transition between block4(1), block12(1) [The 'Believe in Jesus but don't like the v
## [1] "Plotting only block 4 info [check cluster 1]"
```



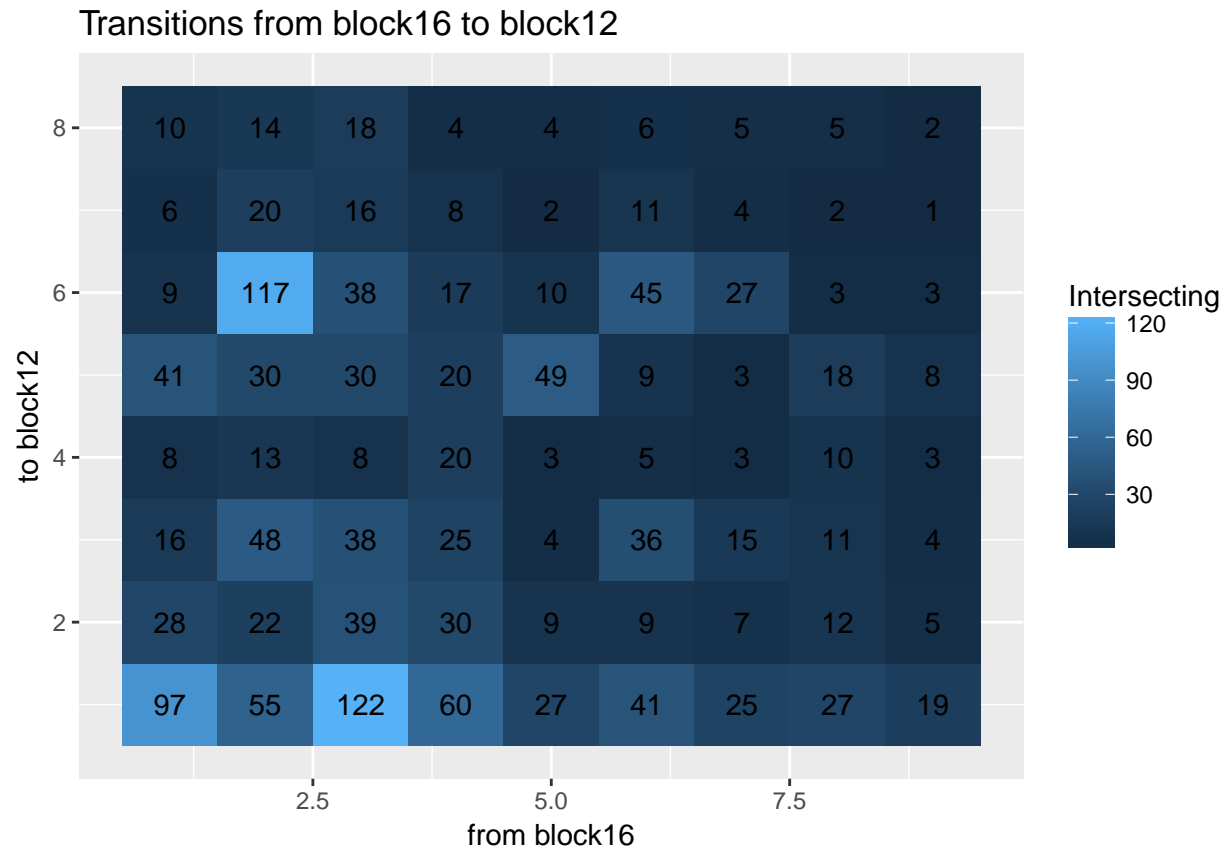


```
## [1] "Interesting transition from block16(2) and block2(2) [The non religious] [about 275 people in b
```

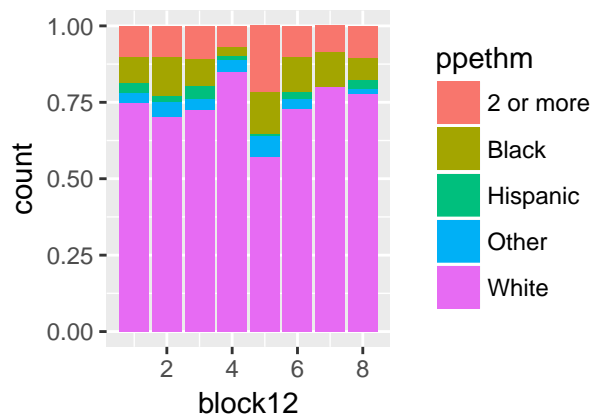
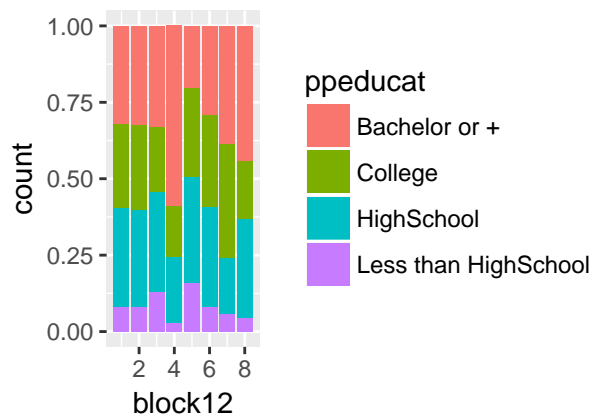
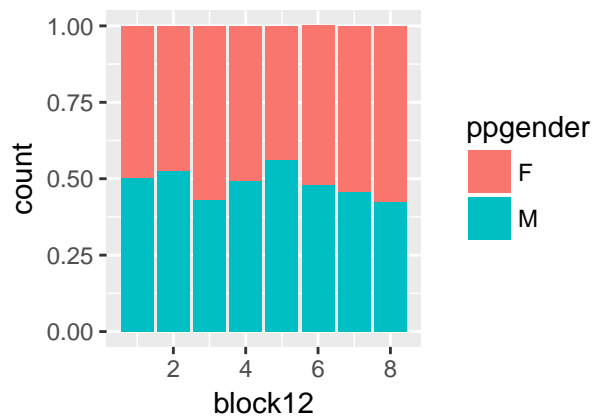
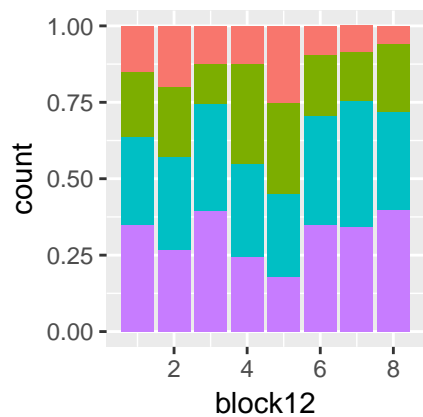




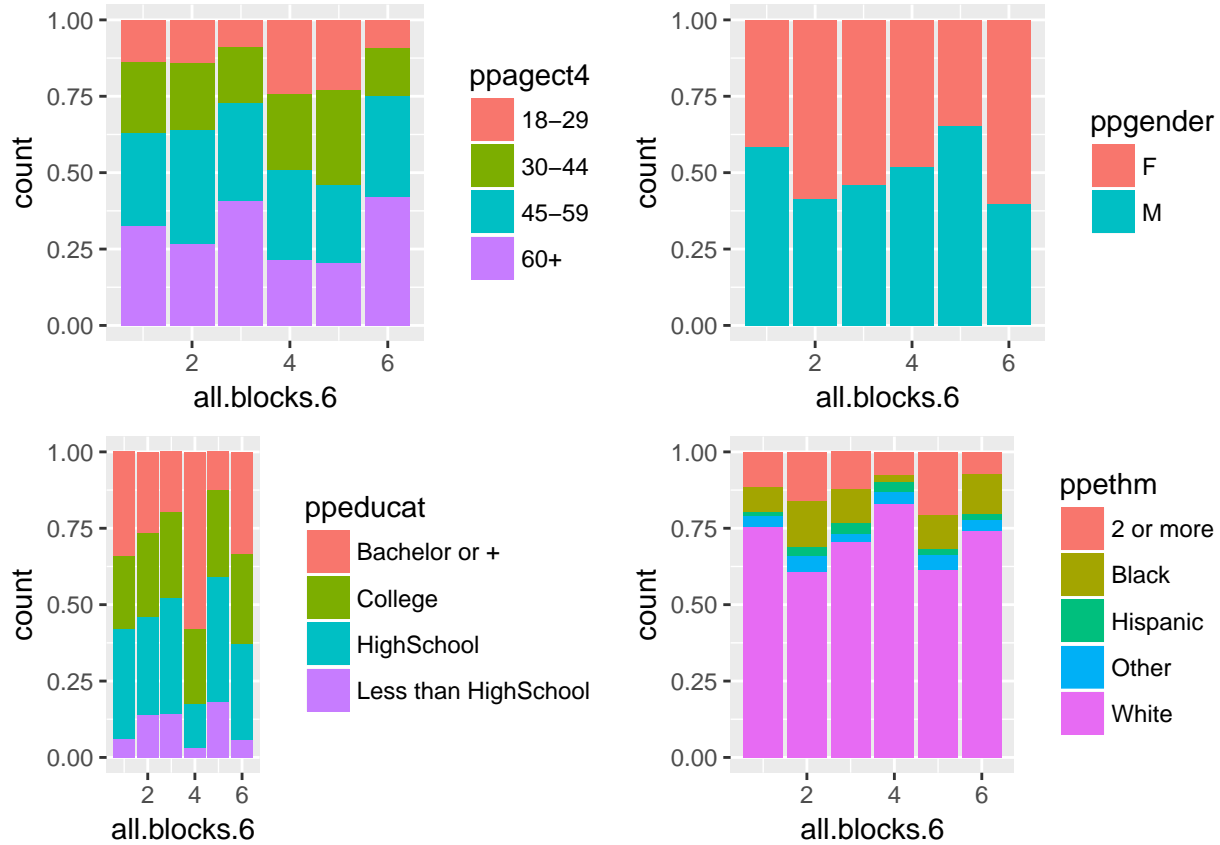
```
## [1] "Plotting only block 16 info [check cluster 2]"
```



```
## [1] "Interesting transition between block16(3) and block12(1) [High integrity and moral values (chri
```



```
## [1] "Plotting only block 12 info [check cluster 6]"
```



## [1] "Clusters in all.blocks.6 => {6: [Non religious], 4: [Hippies/High integrity and moral], 1: [Alw

## Conclusions

1. Can we find out ways to interpret the answers of each block as one value?

Yes we can! We used PCA to obtain weights from the sum of loadings per variable weighted by the variance explained by each principal component.

2. Can we identify distinct clusters within different blocks of questions?

Yes we can! We performed k-modes clustering on a per block basis and obtained the centroids of each cluster. We then related the results with the questions presented in the questionnaire.

3. Based on these values and/or clusters, are there clear correlations between interpretable blocks?

Sometimes. Our results show that the correlations between variables are not particularly strong. However we could find correlations between individual clusters for specific blocks. Additionally, there was often overlap between the clusters and only few variables clearly separated a cluster from the others.

4. Are there interpretable types of individuals?

Based on the per-block basis clustering we could find pairs of clusters correlated over multiple blocks. This established a link between blocks of questions for a specific cluster, such as:

- a group of individuals never taking a side
- a group of individuals with high answers to integrity and ethical values questions

Given our clustering on the entire dataset we could find 4 types

- Materialistic pragmatists
- Individualists
- Average people
- Saints
- Can we link the clusters to personal metrics?

Yes, we can! Corresponding to the previous 4 types

- Materialistic pragmatists
  - mostly young
  - educated
  - male
  - abstain from marriage
- Individualists
  - ethnically more diverse
  - less tertiary education
  - fewer living in detached houses
- Average people
  - guess what... average - everywhere!
- Saints
  - most likely to be married
  - least likely to have an apartment
  - most likely to be living in detached houses
  - more women