1 Eurobarometer Data

1.1 Data Overview

We examine rank data collected in Eurobarometer 34.1. The Eurobarometer is series of surveys conducted by the European Commission to measure public opinion and sentiment across a variety of topics. The standard Eurobarometer survey has been carried out biannually since 1973, and special modules to study specific topics are commissioned frequently as well. In the fall of 1990, Eurobarometer 34.1 was a special module commissioned to study European perspectives on various public health issues. The survey included various questions across a broad spectrum of topics including alcohol consumption, childcare, employment, tobacco use. In particular, we examine the rank data generated by four specific questions. Three separate questions asked respondents to rank policy priorities for addressing AIDs, alcoholism, and illegal drugs. The data we have includes 11872 individuals across the following 13 countries: France, Belgium, Netherlands, West Germany, Italy, Luxembourg, Denmark, Ireland, UK, Greece, Spain, Portual and East Germany.

	FR	BG	NL	WG	IT	LX	DN	IR	UK	GE	SP	PR	EG	Total
Individual Counts	967	812	1016	975	1048	242	950	879	1223	980	864	978	938	11872
Relative Proportion	0.08	0.07	0.09	0.08	0.09	0.02	0.08	0.07	0.10	0.08	0.07	0.08	0.08	1.00

Table 1: Nationalities of Respondents in Eurobarometer 34.1

1.2 Survey Questions

The specific questions we are interested in are as follows:

There are various actions that could be taken to eliminate the drugs problem. In your opinion, what is the first priority? And the next most urgent? (Ask respondent to rank all 7, with 1 as the most urgent).

1. Information campaigns about the dangers of drugs

2. Hunting down drug pushers and distributors

3. Legal penalty for drug taking

4. Looking after and treating drug addicts and rehabilitating them

5. Funding research into drug substitutes, and into the treatment of drug addiction

6. Fighting the social causes of drug addiction

7. Reinforcing the control or distribution and usage of addictive medicines

There are various actions that could be taken in order to ease the problem of alcoholism and its consequences. Looking at this card, which is the main priority in your view? and the next? (Rank up to 5)

1. Information campaigns about the danger of alcoholism

2. Stiffer penalties for offences committed under the influence of alcohol

3. Banning advertising for alcoholic drinks

4. Increasing taxes on alcohol

5. Restricting the sale of alcohol. especially to young people

6. Putting lower legal limits on alcohol content

7. Making social outcasts of alcoholics

8. Helping alcoholics to submit to "drying out"

9. Funding medical research to develop more effective treatments

10. Setting up more reception centers, drying out treatment centers

There are various actions that could be taken in order to eliminate the problem of AIDS or at least to slow down its development. Looking at this card, which is the main priority in your view? And the next? (Rank all by giving a number from 1 to 5. with 1 as a top priority)

1. Information campaign about the danger

2. Punishment for behavior which increases the risk

3. Identifying and isolating those with AIDS or those who are HIV positive

4. Treating of those with AIDS and looking after them

5. Funding research to find a vaccine

2   Model

To model the data, we use a mixed membership model. Mixed membership models extend the mixture model framework by allowing each individual to belong to multiple sub-populations instead of only a single sub-population. In particular, we use the Plackett-Luce model that assumes that individuals select a permutations from a set of V candidates given non-negative support parameters θ_v for each candidate. For identifiability, the support parameters must sum to 1. Each individual i then selects $N_i \leq V$ candidates sequentially, beginning with their highest preference candidate where the probability of selecting one of the remaining candidates at a given ranking level is proportional to the candidates' support parameter. For a set of support parameters $\{\theta_v\}_{v=1:V}$ such that $\sum_{v=1}^{V} \theta_v = 1$ and $\theta_v \geq 0$ for $v = 1, \ldots, V$, the mass function for the Plackett-Luce Model is then defined as

$$P(Y_i) = \prod_{n=1}^{N_i} \frac{\theta_{c(n)}}{1 - \sum_{\eta=0}^{n-1} \theta_{c(\eta)}} \tag{1}$$

where Y_i is a permutation of length N_i of the possible candidates, c(n) indexes the candidate selected at the nth ranking level, and $\theta_{c(0)} = 0$. This process can be thought of as a "multinomial without replacement". With this view, support parameters can also be interpreted as the probabilities of a candidate being ranked first.

The full hierarchical generative model is described below. This is a multivariate generalization of the univariate model proposed by Gormley and Murph 2009. We assume the following fixed quantities: J observed variables with V_j candidates where V_j is known, K latent sub-populations, $\alpha = \{\alpha_1, \ldots, \alpha_K\}$ where each element is positive, and a set of support parameters $\theta_{j,k}$ for each variable j and sub-population k where $\theta_{j,k}$ is a vector which lies in the V_j simplex.

For each of the $i \in [T]$ individuals:

(a) Draw a vector $\lambda_i \sim \text{Dirichlet}(\alpha)$. The vector λ_i lies in the K simplex and indicates the degree to which individual i belongs to each of the K sub-populations.

(b) For each $j \in [J]$ observed variables, $r \in [R_j]$ repeated measures for variable j, and $n \in [N_{i,j,r}]$ ranking level for individual i on variable j and repeated measure r:

(1) Draw $Z_{i,j,r,n} \sim multinomial(1, \lambda_i)$. Z is an indicator vector of length K, such that $Z_{i,j,r,n,k_0} = 1$ indicates that the support parameters of the k_0^{th} sub-population govern the candidate selection at level n.

(2) Draw a candidate $X_{i,j,r,n} \sim Plackett - Luce(\theta|Z_{i,j,r,n}, X_{i,j,r,\eta<n})$ for ranking level n. $X_{i,j,r,n}$ is an indicator vector of length V_j, such that $X_{i,j,r,n,v_0} = 1$ indicates that candidate v_0 was chosen at ranking level n.

In the described model, intra-individual mixing of sub-population preferences occurs at each ranking level, not just between variables or repeated measures. Thus, an individual may select their first rank candidate according to the preferences of one sub-population, but select their second rank candidate (from the remaining candidates) according to the preferences of another sub-population. In addition, each individual may provide an incomplete ranking where $N_{i,j,r} \leq V_j$.

3 Model Fitting and Selection

Direct estimation of mixed membership models is intractable in general, so we use a variational EM approach which optimizes a lower bound to the log-likelihood (often the lower bound is called the ELBO). To select the number of sub-populations, we use a held out ELBO criterion. We first select a training set of 6000 individuals and leave the remaining 5782 individuals in a test set. We fit a mixed membership model to the training set, and use the estimated global parameters, α and θ, to calculate the ELBO for the held out test set. Unfortunately, in these models, we find that the ELBO has many local modes so we use 120 different initializations at each level of K = 3, ... 10 to increase the probability of finding a global maxiumum and not just a local model. The model with the highest held out ELBO is selected and we then use then use the corresponding intialization point to fit a model to the entire 11872 indivudals.

We repeated this procedure three times with randomly selected training/test sets and the result from a single batch is shown below. Two of the batches selected similar 4 component models, while 1 batch selected a 5 component model. Presently, we are still thinking about a more principled model selection procedure. Nonetheless, one of the selected 4 component models and the selected 5 component model are presented here.

```
## Warning:  package 'vioplot' was built under R version 3.1.3
```


Held Out ELBO
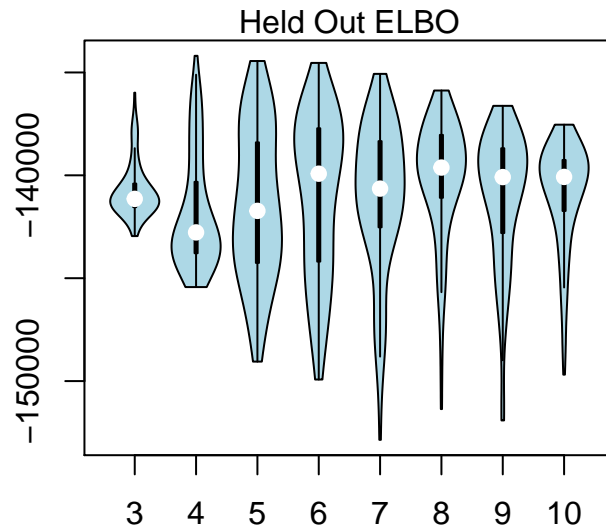
4   Data Analysis

In our exploratory analysis, we saw a disproportionately large number of ranked responses which simply corresponded to the ordering in which the priorities were presented in the physical survey. Thus, we include two fixed sub-populations. The first sub-population has uniform support over all priorities, the second sub-population has extremely high support corresponding to the ordering of the priorities in the survey.

4.1   Model 1: 4 Component Model

The optimal selected model here has 4 sub-populations including the 2 fixed groups

First we examine the estimated relative frequency of each of the identified sub-populations. The uniform group is group 2 and the rank ordering group is group 3. The other two fitted groups are discussed below. The uniform group (which is meant to simply model noisy responses) is the largest identified group.

We can also see when we plot the grade of membership for each individual's mode sub-population that while some individuals have fairly dispersed sub-population memberships, a large majority of the individuals have very high membership in a single sub-population.

```
rel.freq <- round(out$alpha/sum(out$alpha),3)
names(rel.freq) <- paste("Group", c(1:4))
rel.freq

## Group 1 Group 2 Group 3 Group 4
##   0.203   0.269   0.374   0.154
```
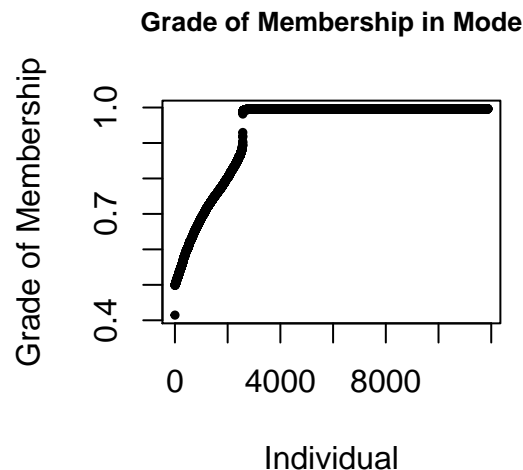

When examining the fitted support parameters, we see that group 1 highly favors information campaigns and social programs while group 2 is in favor of more punitive fixes. First we print tables with the estimated support parameters. We also show the results graphically by plotting the support parameters of each group. The red line shows the weighted average of the support parameters for the 2 non-fixed groups where the weights are the relative frequency of each group. This essentially shows the typical response for those not in the 2 fixed groups.

| Priorities		Group 1	Group 2	Group 3	Group 4
1	Information campaigns about the dangers of drugs	0.24	0.06	0.14	0.99
2	Hunting down drug pushers and distributors	0.22	0.62	0.14	0.01
3	Legal penalty for drug taking	0.05	0.07	0.14	0
4	Looking after and treating drug addicts and rehabilitating them	0.14	0.05	0.14	0
5	Funding research into drug substitutes, and into the treatment of drug addiction	0.06	0.04	0.14	0
6	Fighting the social causes of drug addiction	0.24	0.06	0.14	0
7	Reinforcing the control or distribution and usage of addictive medicines	0.05	0.1	0.14	0

Table 2: Question 1

| Priorities		Group 1	Group 2	Group 3	Group 4
1	Information campaigns about the danger of alcoholism	0.4	0.12	0.1	0.99
2	Stiffer penalties for offences committed under the influence of alcohol	0.1	0.3	0.1	0.01
3	Banning advertising for alcoholic drinks	0.06	0.07	0.1	0
4	Increasing taxes on alcohol	0.02	0.03	0.1	0
5	Restricting the sale of alcohol, especially to young people	0.03	0.16	0.1	0
6	Putting lower legal limits on alcohol content	0.04	0.02	0.1	0
7	Making social outcasts of alcoholics	0	0.01	0.1	0
8	Helping alcoholics to submit to drying out	0.15	0.12	0.1	0
9	Funding medical research to develop more effective treatments	0.11	0.07	0.1	0
10	Setting up more reception centers, drying out treatment centers	0.08	0.1	0.1	0

Table 3: Question 2

Priorities		Group 1	Group 2	Group 3	Group 4
1	Information campaign about the danger	0.34	0.26	0.2	0.99
2	Punishment for behavior which increases the risk	0.03	0.07	0.2	0.01
3	Identifying and isolating those with AIDS or those who are HIV positive	0.03	0.06	0.2	0
4	Treating of those with AIDS and looking after them	0.16	0.19	0.2	0
5	Funding research to find a vaccine	0.44	0.41	0.2	0

Table 4: Question 3

Estimated Theta
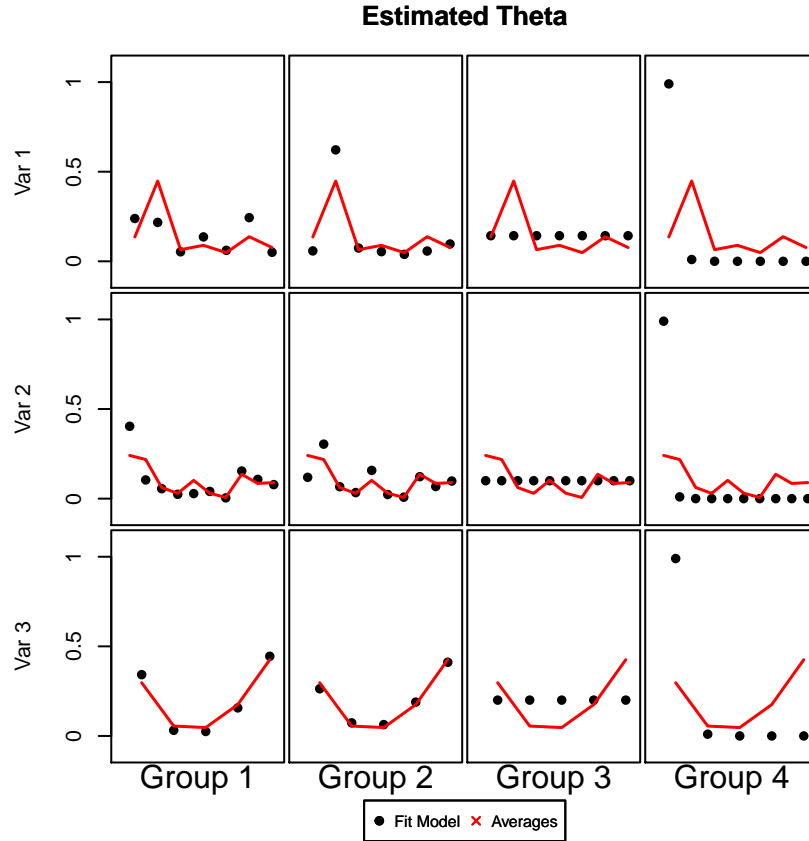
We can also view the results on a national level where we take the average individual membership for each and project each nation in a simplex. We see structure that seems somewhat linked to national political ideologies. When we view only the 2 non-fixed groups, we see an even clearer pattern. The UK and Ireland seem the most punitive, while more socialist countries like Denmark, Greece, and Portugal are on the other end of the spectrum.
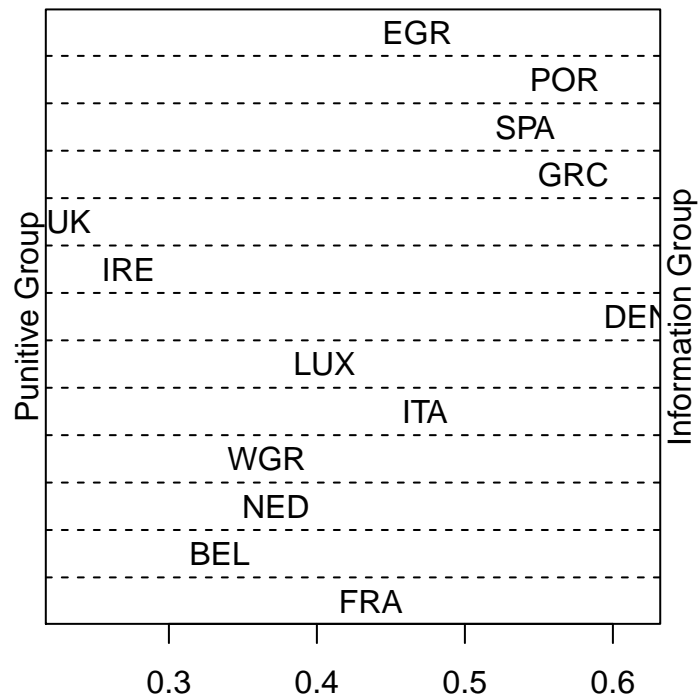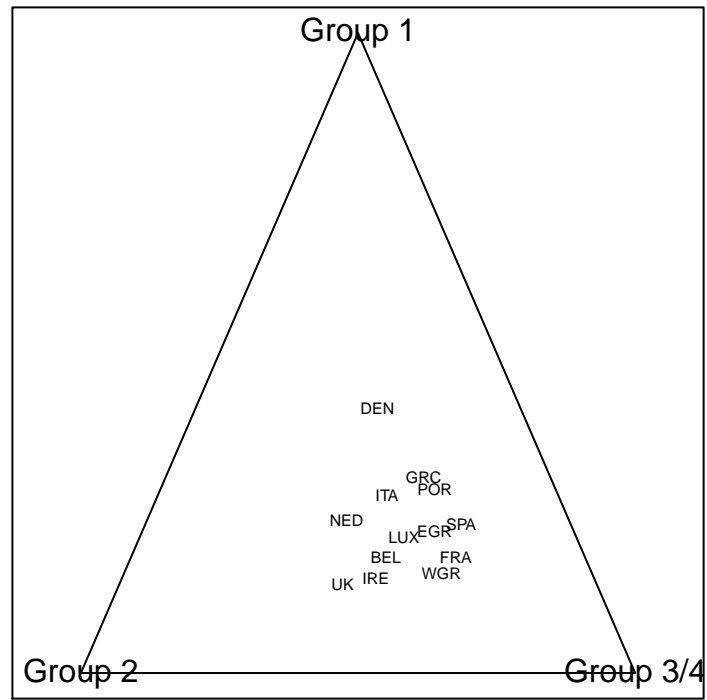
4.2 Model 2: 5 Component Model

The optimal selected model here has 5 sub-populations including the 2 fixed groups

First we examine the estimated relative frequency of each of the identified sub-populations. The uniform group is group 4 and the rank ordering group is group 5. The other three fitted groups are discussed below. Here we see that the uniform group has a much smaller relative frequency when compared to the previous 4 component model.

We can also see when we plot the grade of membership for each individual's mode sub-population that while some individuals have fairly dispersed sub-population memberships, a large majority of the individuals have very high membership in a single sub-population. This agrees with the previously presented result.

```
rel.freq <- round(out$alpha/sum(out$alpha),3)
names(rel.freq) <- paste("Group", c(1:5))
rel.freq

## Group 1 Group 2 Group 3 Group 4 Group 5
##   0.390   0.237   0.079   0.148   0.146
```
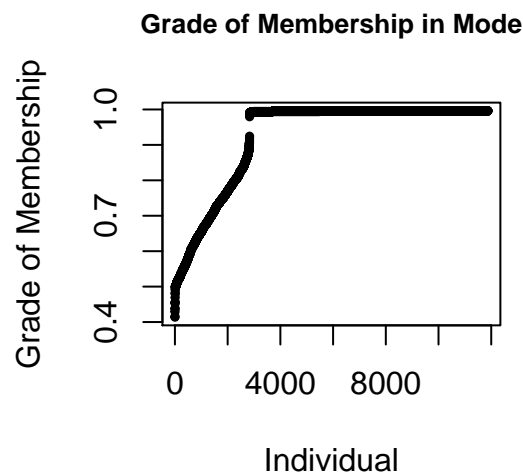
Grade of Membership in Mode

When examining the fitted support parameters, we see that group 1 highly favors information campaigns, group 2 is in favor of more punitive fixes and group 3 seems to support research and rehabilitation. First we print tables with the estimated support parameters. We also show the results graphically by plotting the support parameters of each group. The red line shows the weighted average of the support parameters for the 2 non-fixed groups where the weights are the relative frequency of each group. This essentially shows the typical response for those not in the 2 fixed groups.

|	Priorities	Group 1	Group 2	Group 3	Group 4	Group 5
1	Information campaigns about the dangers of drugs	0.27	0.05	0.16	0.14	0.99
2	Hunting down drug pushers and distributors	0.2	0.72	0.19	0.14	0.01
3	Legal penalty for drug taking	0.04	0.06	0.04	0.14	0
4	Looking after and treating drug addicts and rehabilitating them	0.18	0.04	0.14	0.14	0
5	Funding research into drug substitutes, and into the treatment of drug addiction	0.08	0.03	0.16	0.14	0
6	Fighting the social causes of drug addiction	0.16	0.06	0.19	0.14	0
7	Reinforcing the control or distribution and usage of addictive medicines	0.07	0.04	0.12	0.14	0

Table 5: Question 1

	Priorities	Group 1	Group 2	Group 3	Group 4	Group 5
1	Information campaigns about the danger of alcoholism	0.33	0.06	0.04	0.1	0.99
2	Stiffer penalties for offences committed under the influence of alcohol	0.09	0.3	0.03	0.1	0.01
3	Banning advertising for alcoholic drinks	0.08	0.14	0.01	0.1	0
4	Increasing taxes on alcohol	0.03	0.05	0.01	0.1	0
5	Restricting the sale of alcohol. especially to young people	0.08	0.24	0.1	0.1	0
6	Putting lower legal limits on alcohol content	0.03	0.05	0.03	0.1	0
7	Making social outcasts of alcoholics	0	0.01	0.01	0.1	0
8	Helping alcoholics to submit to drying out	0.15	0.06	0.48	0.1	0
9	Funding medical research to develop more effective treatments	0.09	0.04	0.15	0.1	0
10	Setting up more reception centers, drying out treatment centers	0.1	0.03	0.14	0.1	0

Table 6: Question 2

	Priorities	Group 1	Group 2	Group 3	Group 4	Group 5
1	Information campaign about the danger	0.38	0.22	0.15	0.2	0.99
2	Punishment for behavior which increases the risk	0.02	0.1	0.08	0.2	0.01
3	Identifying and isolating those with AIDS or those who are HIV positive	0.06	0.16	0.16	0.2	0
4	Treating of those with AIDS and looking after them	0.19	0.17	0.32	0.2	0
5	Funding research to find a vaccine	0.35	0.35	0.3	0.2	0

Table 7: Question 3

Estimated Theta
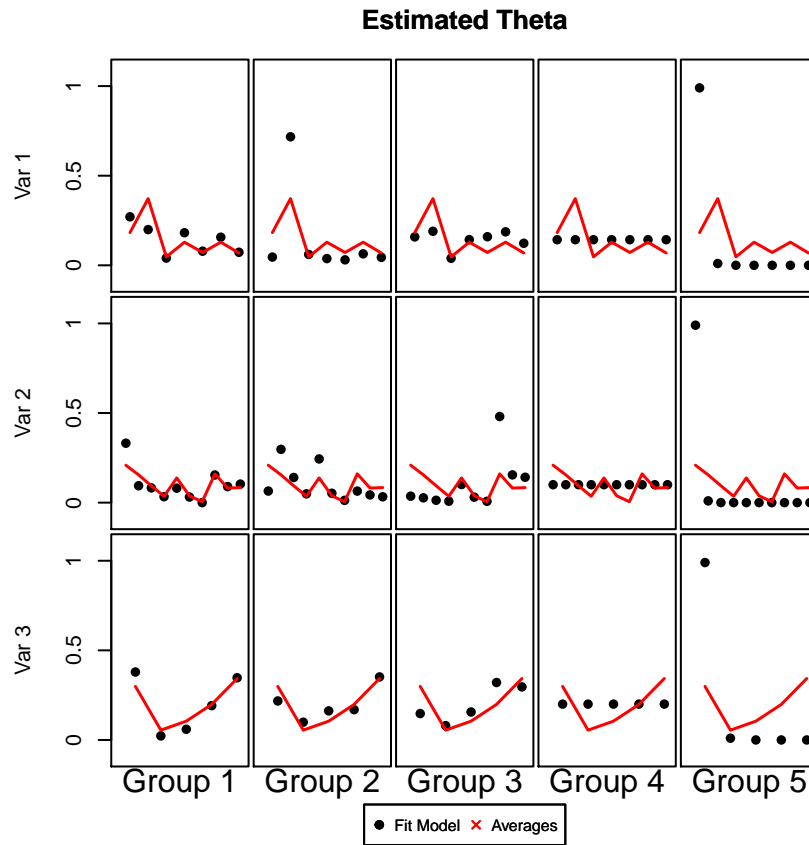
We can also view the results on a national level where we take the average national membership and plot each nation in a simplex. Note that since group 3 has relatively low relative frequency, the countries are differ primarily on the Group 1 vs Group 2 scale, so we plot the memberships in group 1 vs group 2 as well. Again, we see structure that seems very similar to the previous results.