Bayesian Factor Analysis for Mixed Ordinal and Continuous Responses

Kevin M. Quinn (2004)

Political Analysis

April 19, 2017

Outline

Factor Analysis

- Motivation for factor analysis

Factor Analysis

- Motivation for factor analysis
- Explain patterns influence the covariance of observed variables through underlying latent factors

Factor Analysis

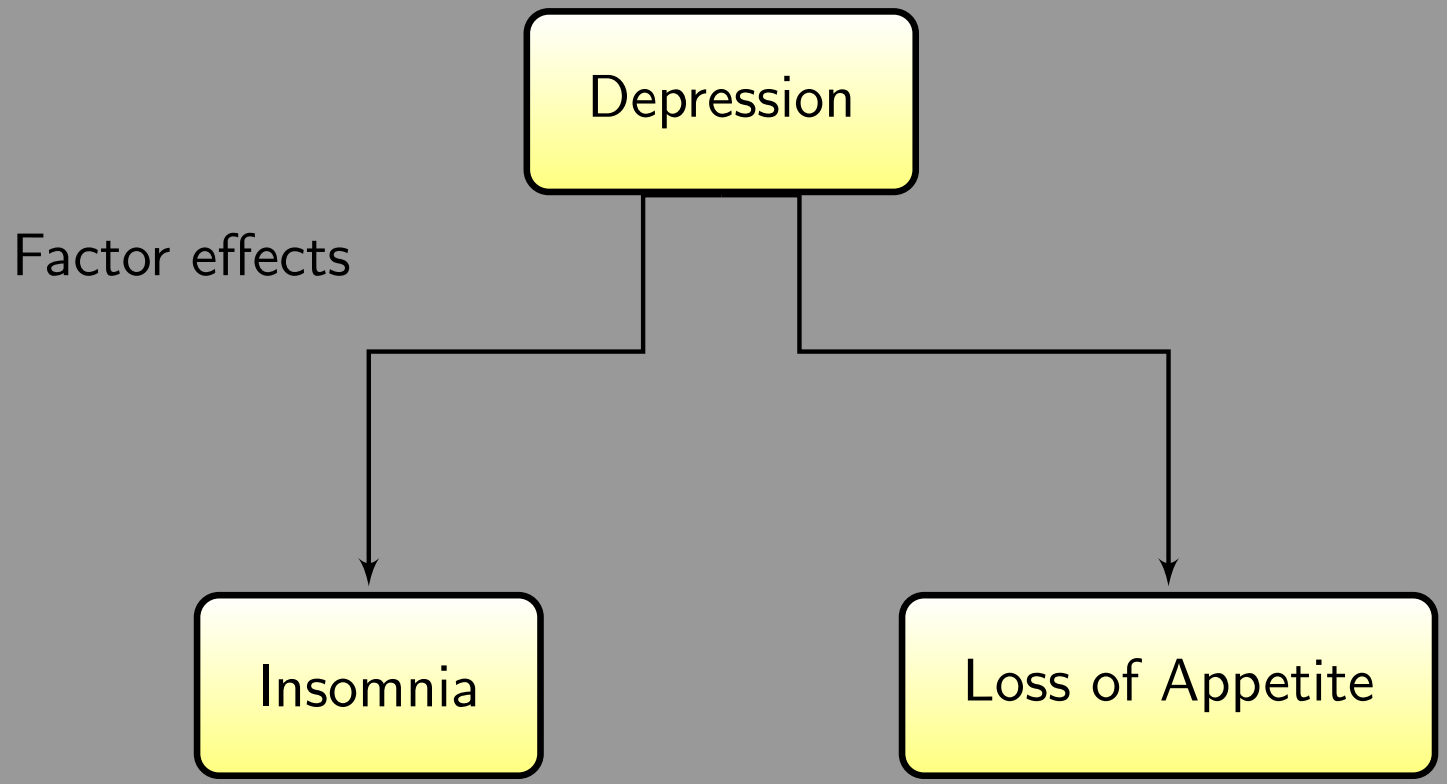


Figure 1: Factor Effect on Observables

Factor Analysis

Two methods traditions being integrated:

- Normal theory factor analysis (continuous variables)

Factor Analysis

Two methods traditions being integrated:

- Normal theory factor analysis (continuous variables)
- Item Response Theory (Ordinal/Likert-type variables)

Factor Analysis

Two methods traditions being integrated:

- Normal theory factor analysis (continuous variables)
- Item Response Theory (Ordinal/Likert-type variables)
-

$$\mathbf{x}^* = \phi\lambda' + \mathbf{E} \tag{1}$$

Factor Analysis

Two methods traditions being integrated:

- Normal theory factor analysis (continuous variables)
- Item Response Theory (Ordinal/Likert-type variables)
- $$\mathbf{x}^* = \phi\lambda' + \mathbf{E} \tag{1}$$
- Problem: posterior density of X would have to come from different types of distribution

Factor Analysis for Mixed Factors

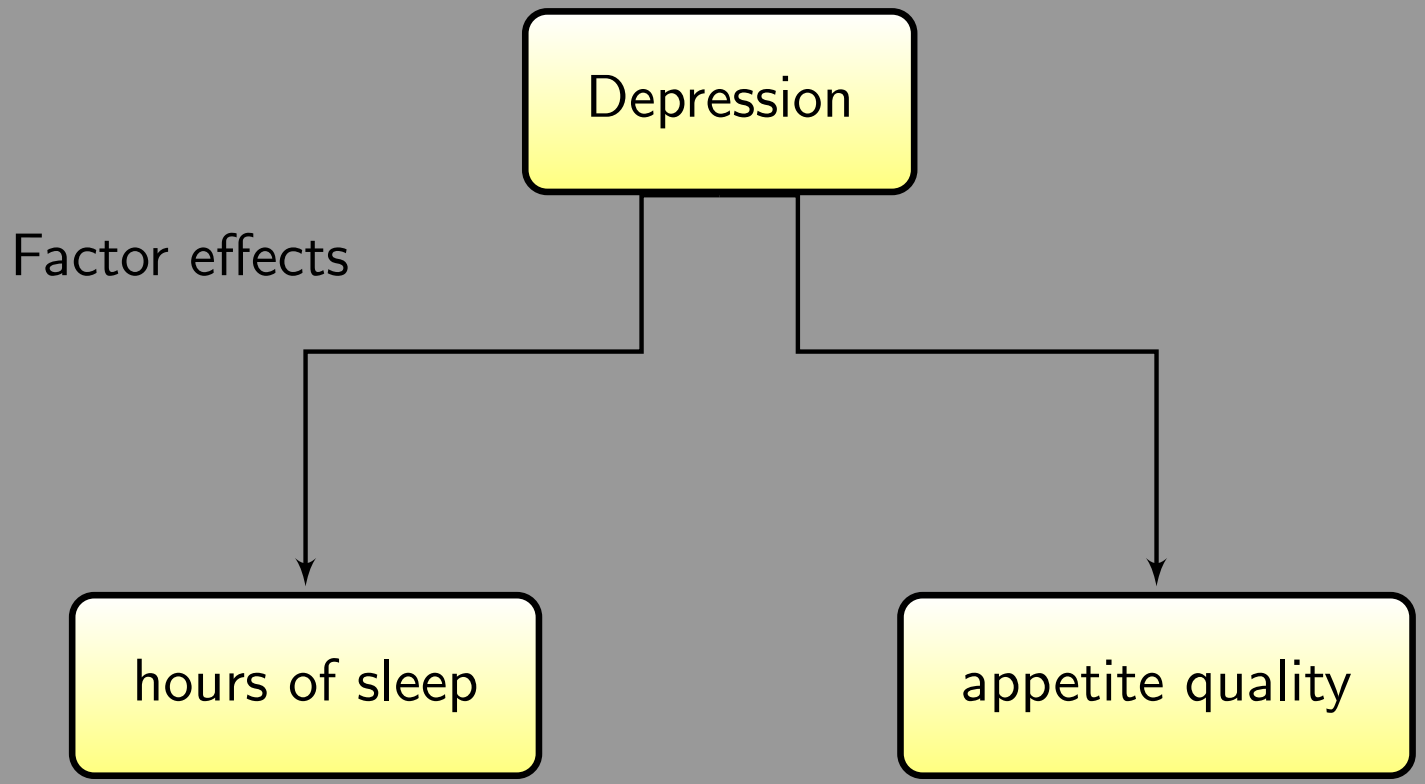


Figure 2: Mixed Factors

Standard Solutions

- Make everything continuous (standard normal theory)

Standard Solutions

- Make everything continuous (standard normal theory)
- Discretize everything (IRT)

Standard Solutions

- Make everything continuous (standard normal theory)
- Discretize everything (IRT)
- Discard one set of the variables

Standard Solutions

- Make everything continuous (standard normal theory)
- Discretize everything (IRT)
- Discard one set of the variables
- Do something else that isn't model based

Proposed Solution

- Generalize the IRT and standard models

Proposed Solution

- Generalize the IRT and standard models
- Should look familiar:

$$\mathbf{X}^* = \phi\lambda' + \mathbf{E} \tag{2}$$

Posterior Density for Mixed FA

$$p(\mathbf{X}^*, \gamma, \Lambda, \phi, \Psi | \mathbf{X}) \propto p(\mathbf{X}|\mathbf{X}^*, \gamma) p(\mathbf{X}^*|\Lambda, \phi, \Psi) p(\gamma) p(\Lambda) p(\Phi) p(\Psi)$$

$$\propto \left\{ \prod_{i=1}^{N} \prod_{j=1}^{J} \left\{ \mathbb{I}(x_{ij} = x_{ij}^*) \mathbb{I}(X_j \text{ continuous}) \right.\right.$$

$$\left. + \sum_{c=1}^{C_j} \mathbb{I}(x_{ij} = c) \mathbb{I}(x_{ij}^* \in (\gamma_{j(c-1)}, \gamma_{jc}]) \mathbb{I}(X_j \text{ ordinal}) \right\}$$

$$\left. \times p_{\mathcal{N}}(\mathbf{x}_i^* | \Lambda \phi_i, \Psi) \right\} p(\Lambda) p(\Phi) p(\Psi),$$

Posterior Density for Mixed FA

$$\alpha = \prod_{i=1}^{N} \frac{\Phi\left(\gamma_{jx_{ij}}^{(can)} - \lambda_j' \phi_i\right) - \Phi\left(\gamma_{j(x_{ij}-1)}^{(can)} - \lambda_j' \phi_i\right)}{\Phi(\gamma_{jx_{ij}} - \lambda_j' \phi_i) - \Phi(\gamma_{j(x_{ij}-1)} - \lambda_j' \phi_i)}$$
$$\times \prod_{c=2}^{C-1} \frac{\Phi((\gamma_{c+1} - \gamma_c)/t_j) - \Phi\left(\left(\gamma_{c-1}^{(can)} - \gamma_c\right)/t_j\right)}{\Phi\left(\left(\gamma_{c+1}^{(can)} - \gamma_c^{(can)}\right)/t_j\right) - \Phi\left(\left(\gamma_{c-1} - \gamma_c^{(can)}\right)/t_j\right)}.$$

Data and Context

- Survey of SOC204 students (n=130)

Data and Context

- Survey of SOC204 students (n=130)
- Mixture of survey questions designed to be easily influenced by political content

Data and Context

- Survey of SOC204 students (n=130)
- Mixture of survey questions designed to be easily influenced by political content
- Each precept wrote a question, so there's a mixture of ordinal, continuous, and open ended responses

Data and Context

- Survey of SOC204 students (n=130)
- Mixture of survey questions designed to be easily influenced by political content
- Each precept wrote a question, so there's a mixture of ordinal, continuous, and open ended responses
- So we might assume at least one underlying factor: liberalness (conservativeness)

Code Pointers

- Variables have to be numeric or ordered factors

Code Pointers

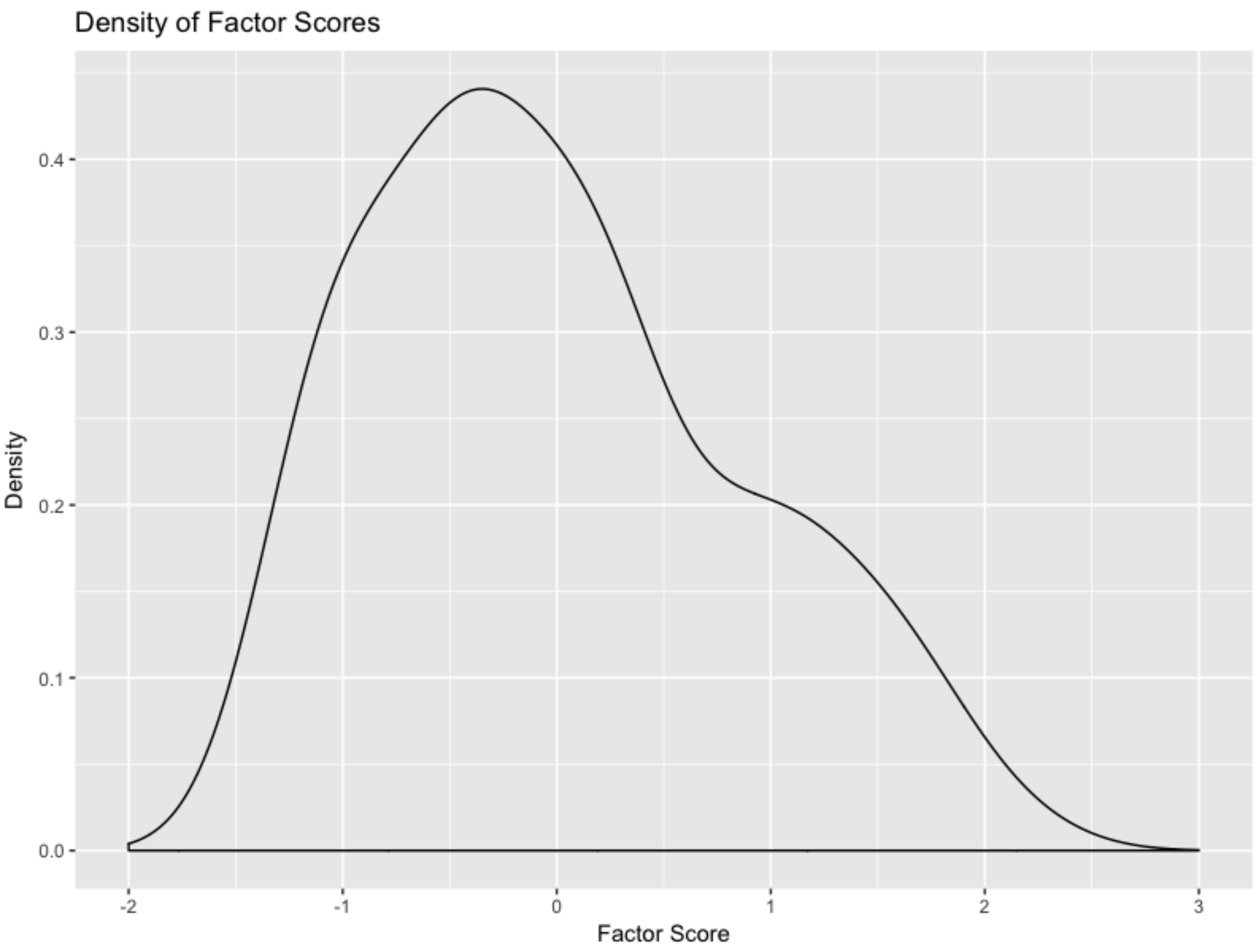
- Variables have to be numeric or ordered factors
- All factor levels have to be represented

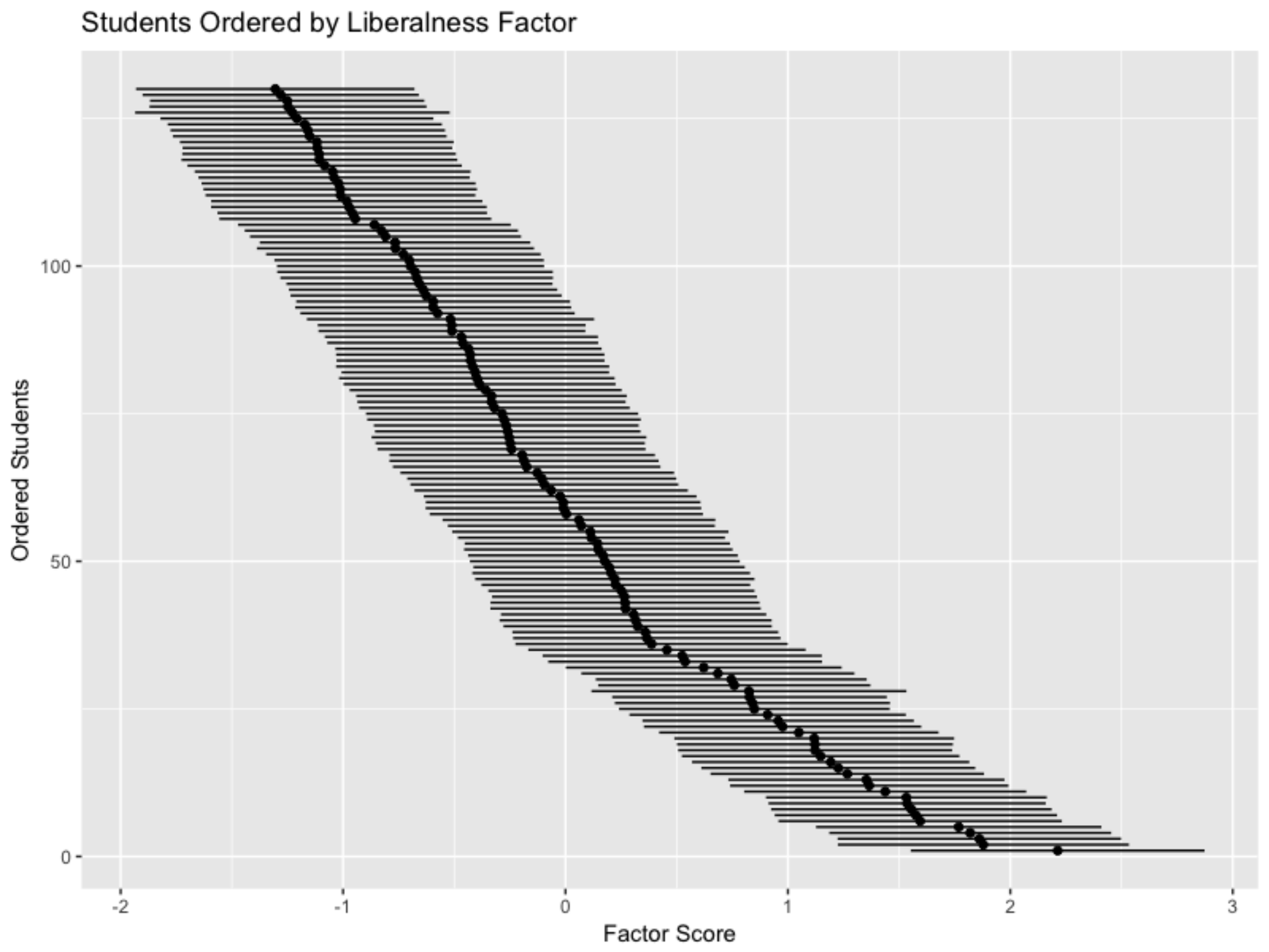
Code

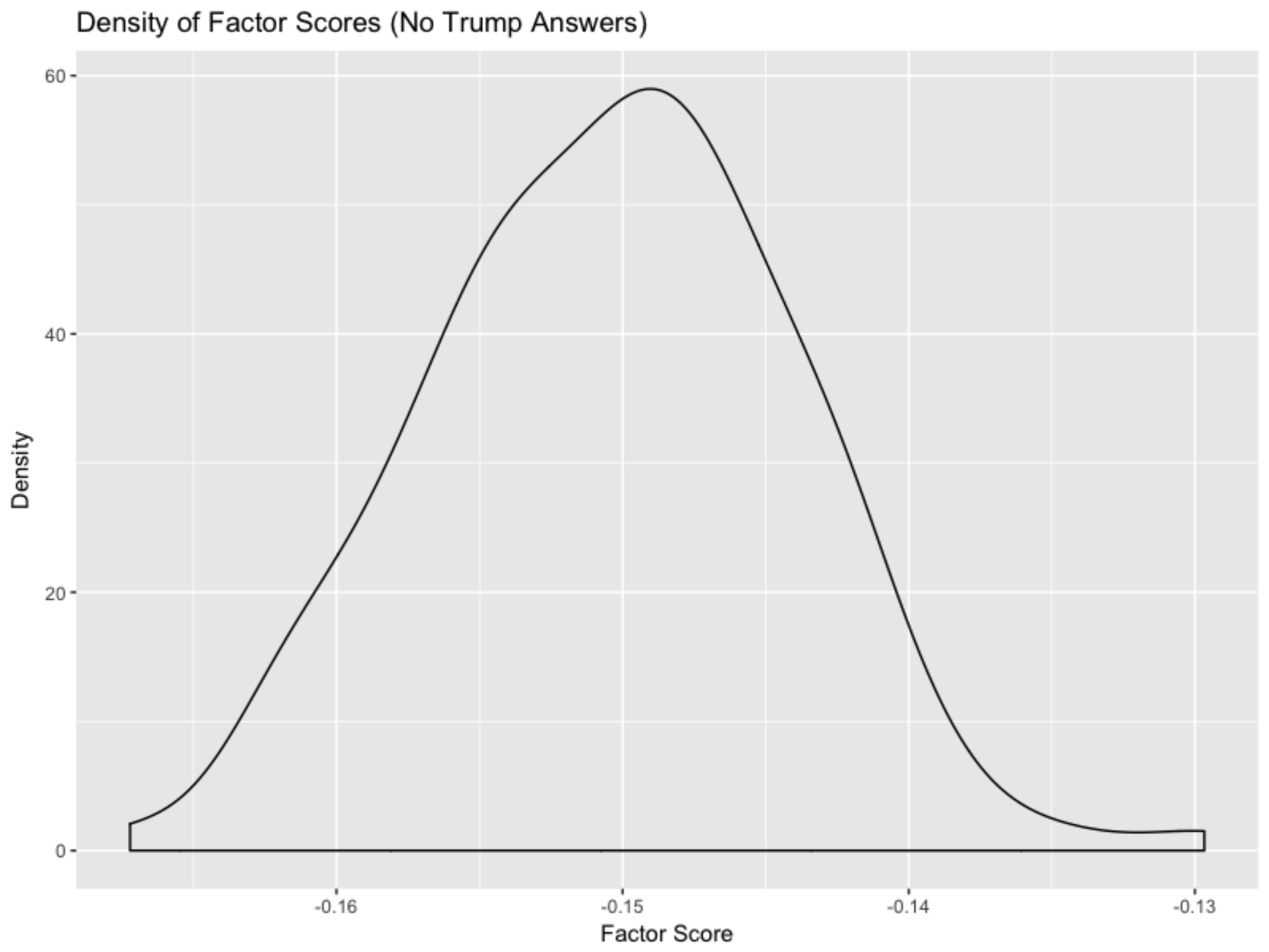
$$\text{mcmc.out} = \text{MCMCmixfactanal}(\text{variables}, \text{factors} = 1,$$
$$\text{data} = \text{dataframe}, \text{store.scores} = \text{TRUE})$$

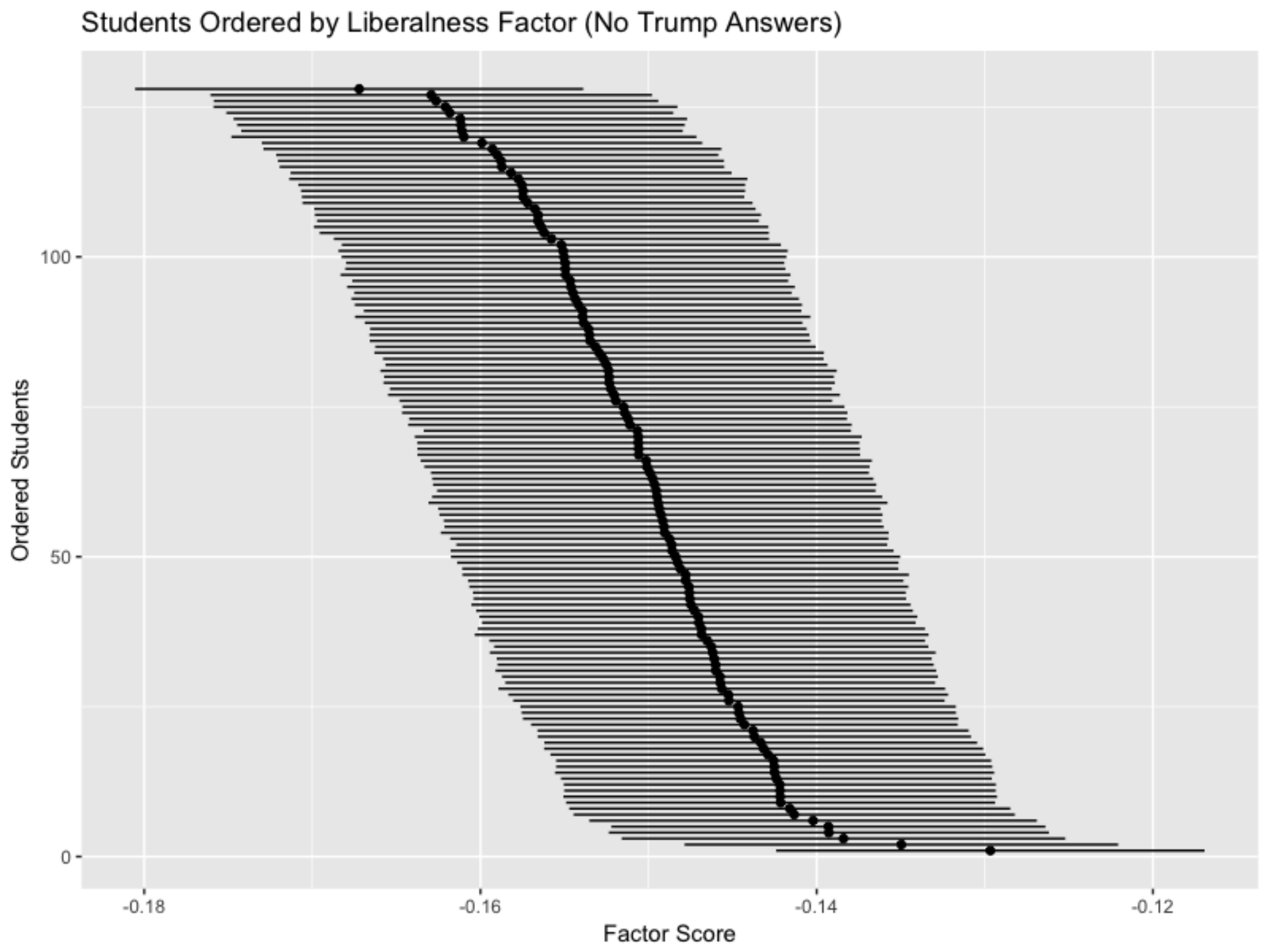
Results

	λ_1	λ_2
Trump Approval	-	-0.09
Islam Threat	0.71	0.15
Trump Score	-	-0.83
Econ Score	-	-0.84
Foreign Score	-	-0.80
Social Score	-	-0.89
Env Score	-	-0.80
Min Wage	2.73	0.72
Crime Issue	0.40	0.05
Gun Rights	-	-0.17
Immigration	-	0.41
Health	-	0.55
Outsourcing	-	-0.03
Tax Reform	-	-0.03
Dakota Access	-	-0.02
Ed Access	-	0.35
Military Involvement	-	0.03
Climate Role	0.12	-0.25
Facebook Use	2.92	-0.06
Facebook Posts	0.47	0.11
Johnson Approval	2.92	-0.16


Density of Factor Scores
Density
0.4
0.3
0.2
0.1
0.0
-2
-1
0
1
2
3
Factor Score


Students Ordered by Liberalness Factor
Ordered Students
100
50
0
-2
-1
0
1
2
3
Factor Score


Density of Factor Scores (No Trump Answers)
60
40
20
0
Density
-0.16
-0.15
-0.14
-0.13
Factor Score


Students Ordered by Liberalness Factor (No Trump Answers)
Ordered Students
100
50
0
-0.18
-0.16
-0.14
-0.12
Factor Score