

Subjectively Perceived Quality of Life

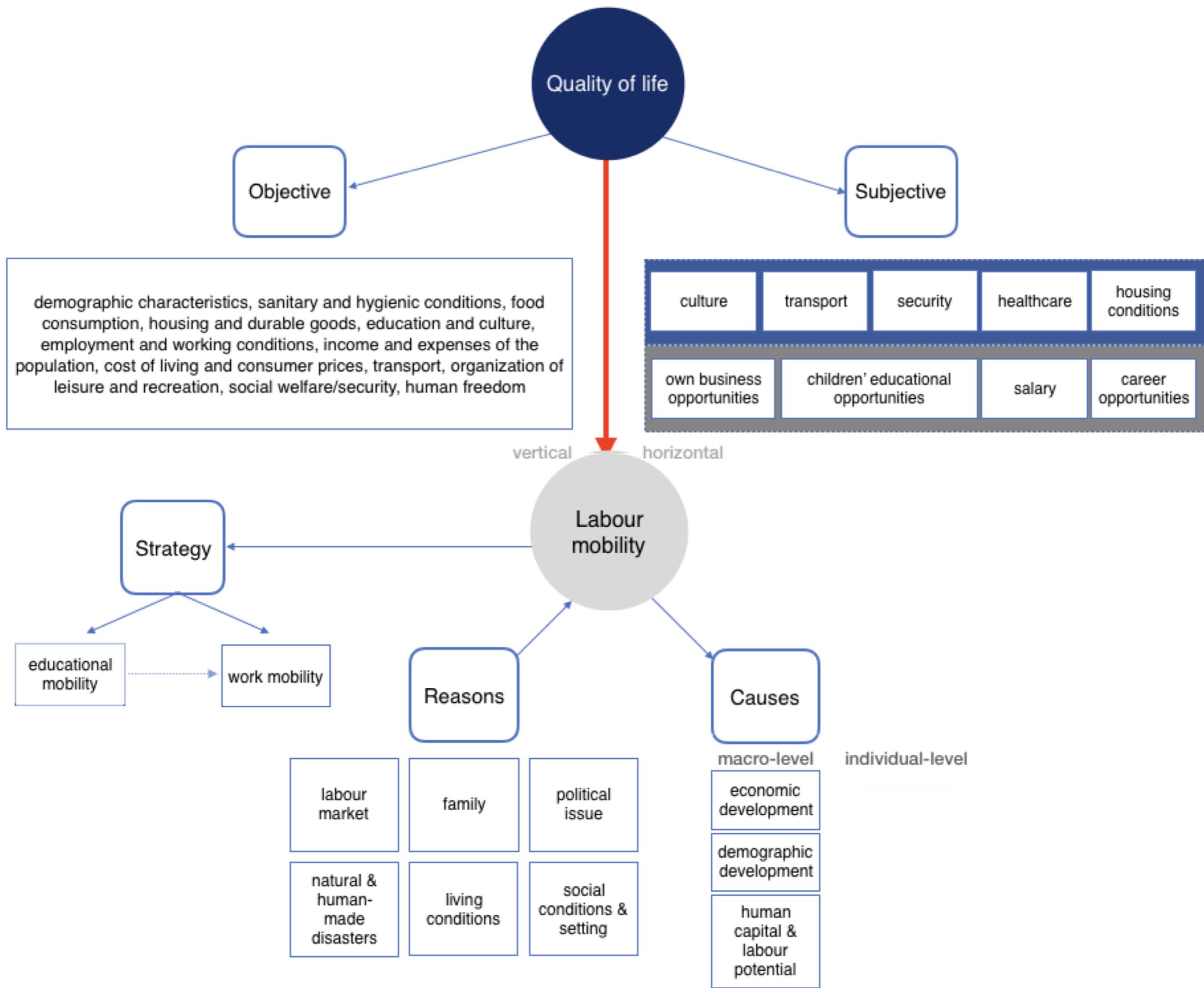
Student: Daria Shcherbakova

Research question

- **The aim:** to investigate quality of life as the factor causing labour migration among young professionals in Russia from different degree background.
- **The research question:** *How does the level of subjective quality of life in the place of studying determine the labour migration among Russian graduates from different fields of study?*
- **Hypothesis:**
 - The decision to relocate for work is influenced by the assessment of various aspects of quality of life, and the satisfaction or dissatisfaction with these aspects plays a crucial role in determining labour migration. Consequently, the higher estimates of quality of life in the locality of study decrease the likelihood of migration in another locality for work;
 - Economic components of life quality exert a greater influence on the decision to relocate than social components;
 - Additionally, graduates from economic backgrounds are more likely to relocate for work due to their diverse skill sets, which allow for greater mobility in career and job seeking.



Theoretical framework



Theoretical framework

- Previous results are focused only on specific aspects of life, not at all indicators of quality of life in order to measure the marginal effect of each particular indicator of quality of life on the decision on labour mobility.
- Former evidence shows consideration of the satisfaction levels of some aspects of life but without conceptualisation and comparative perspective.



Methodology

- **Data:** The panel study “Trajectories in Education and Careers” (9th wave, 2020). The sample of the study is graduates (who are able to work, including those who are working and on parental leave, unpaid or paid leave)
- **Novelty:**
 - The comparison approach based on the different fields of graduates’ studies and gender;
 - The contribution to study quality of life as the factors influencing on the decision on migration among junior specialists, performing binary choice model in terms of the migration prediction. This approach helps to clarify particular determinants, in other words, factors and characteristics of regions, which could affect the decision on migration.

Variables

- **Dependent variable (binary):** the fact of relocation from one locality to another one (“*Do you now live in the same locality where you graduated from a university (college/technical school), or in another?*”)
- **Independent variables:**
 - **The characteristics of quality life (5-point scale):**

HealthCare	
Transport	
Culture	
Housing	
Security	
Career	
Salary	
Education	
Business	

“To which extent do you satisfy with the following aspects of the life in the place where you studied”
 - **Specialisation of the graduates**
 - **Gender**
- **Control variables:** Educational level of parents (categorical), Subjective material status (categorical), Level of Happiness (quasi-interval)

Data Description

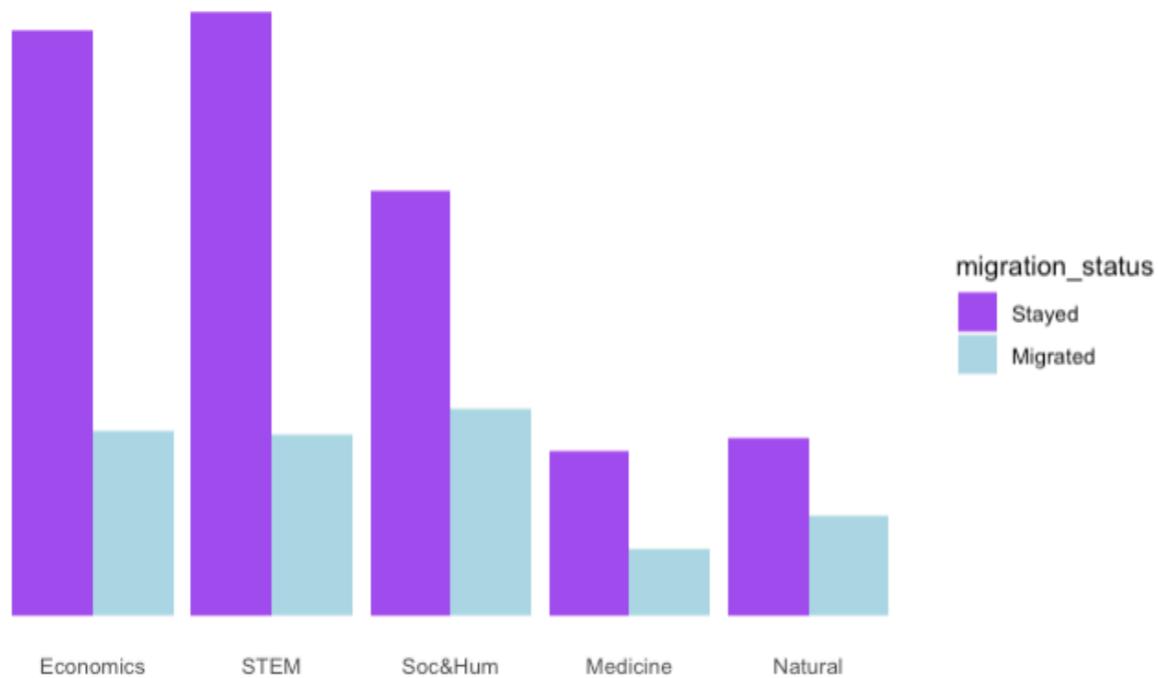
N = 756 (missing values have been dropped due to lack of possibility to impute the data, since the group of respondents has been ignoring the block of question on quality of life).

Groups by dependent variable (migration status) are unbalanced, but it would not affect the results of GLM.

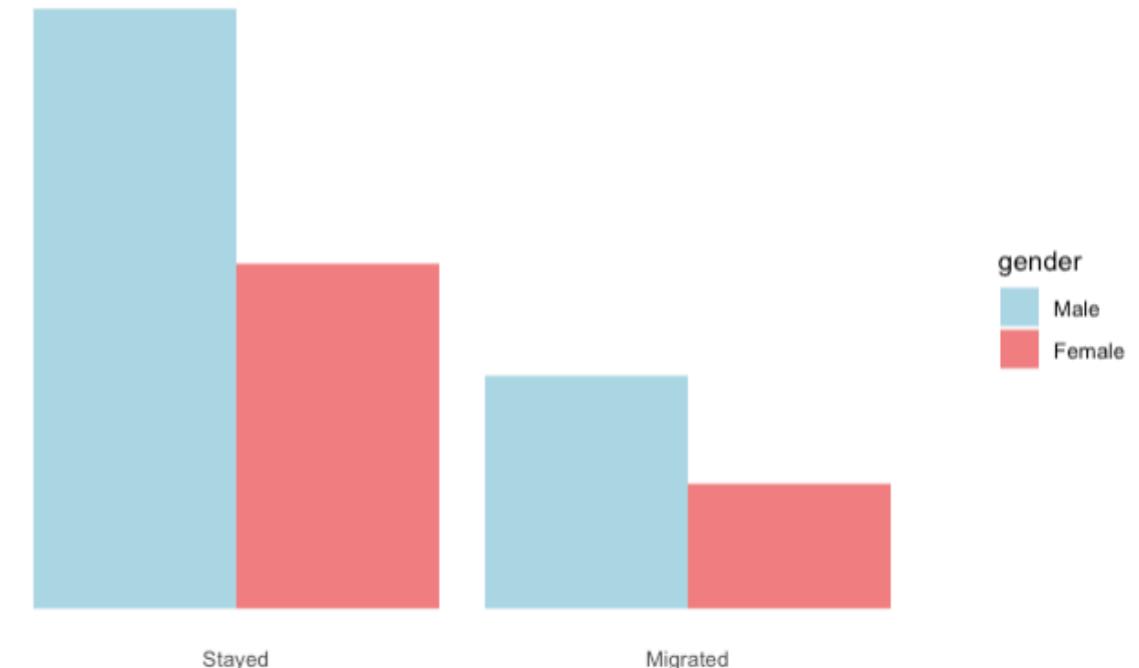
The share of DV

Live in the same locality	Live in another locality
72 %	28 %

Distribution of spec by Migration Status



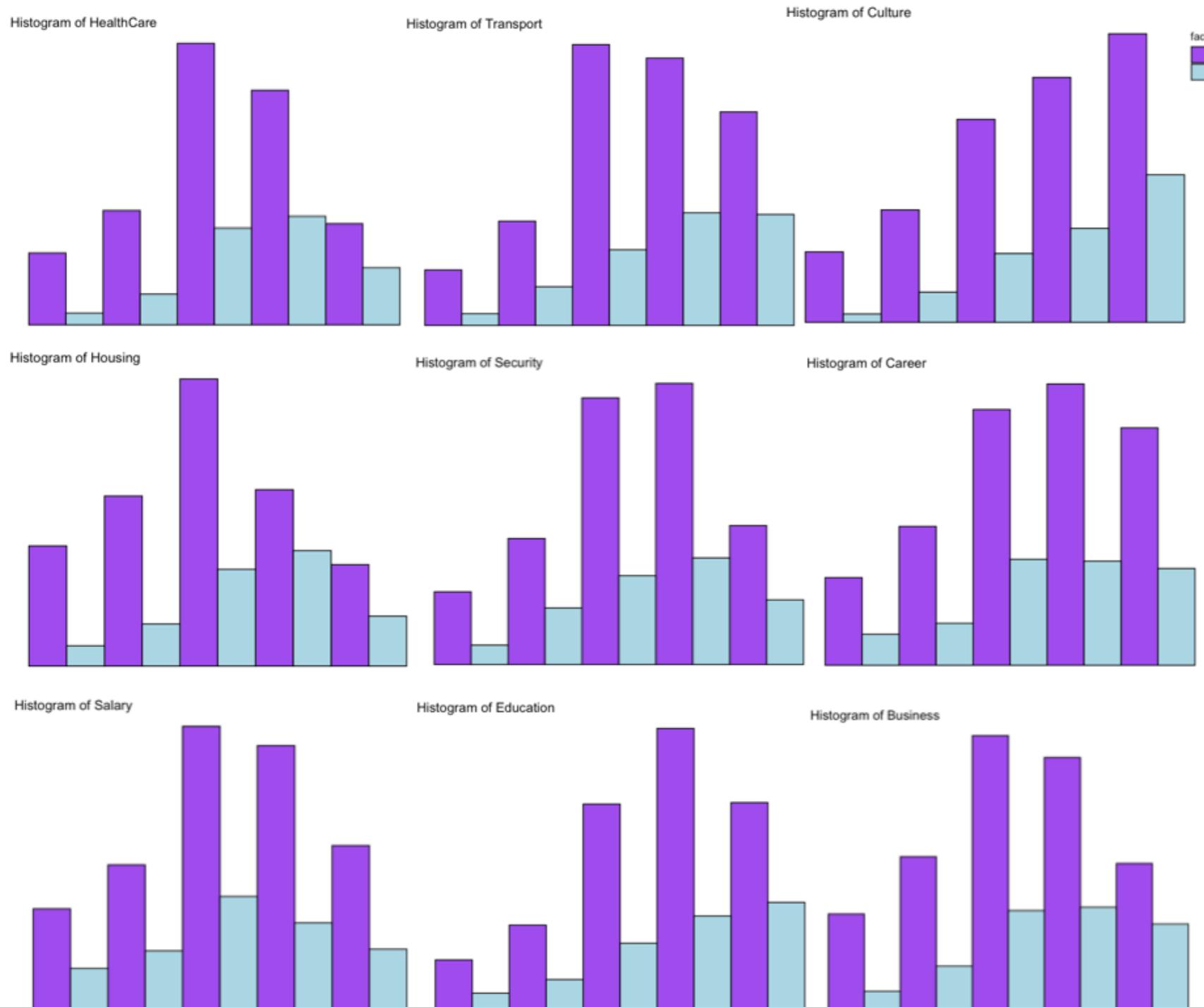
Gender Distribution by DV



According to the distribution of the DV by specialisations, groups are unbalanced (but it would not affect the results). The number of migrated Social Science graduates is the largest.

According to the distribution of the DV by gender, groups are unbalanced (but it would not affect the results). The number of migrated males graduates are higher than among females.

Data Description

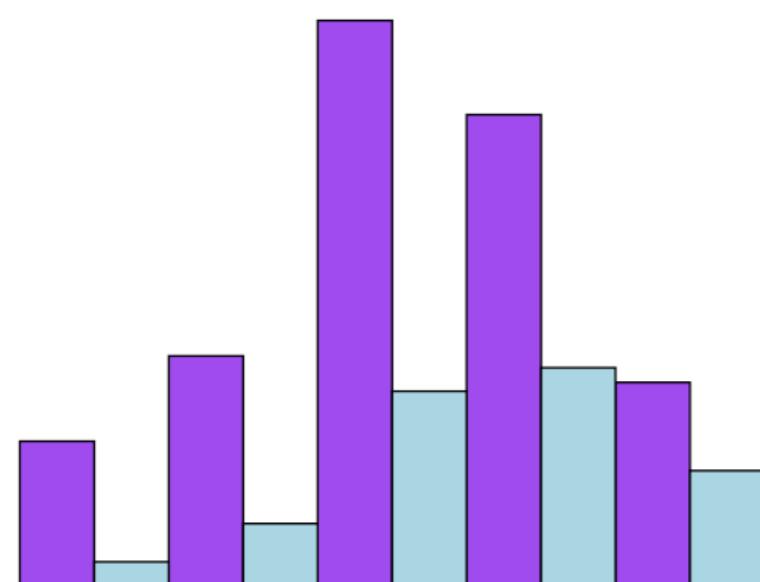


Subjective levels of satisfaction with different characteristics of the region of study are not normal distributed for groups of migrated and not graduates.

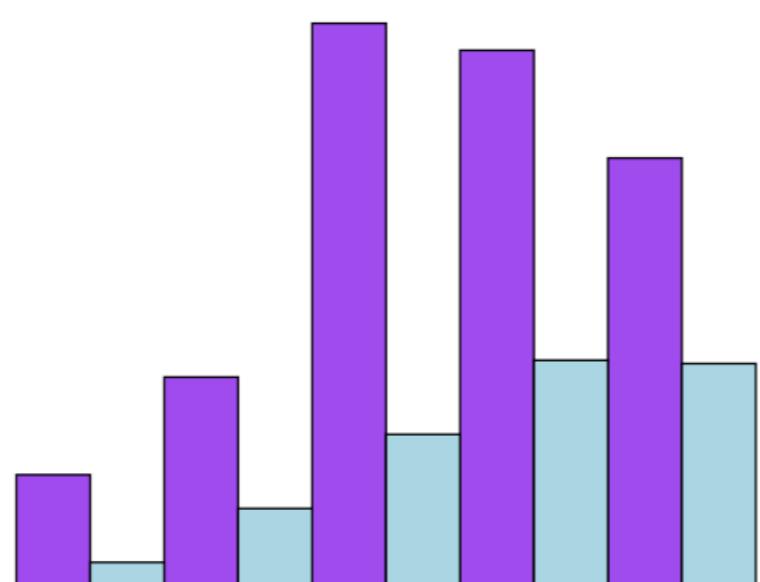
Majority of satisfaction level is based almost between 3-4 points. Housing condition for non-migrated group seems normal distributes, while distribution of Culture is skewed to the right.

However, the number of sample is enough to apply CLT.

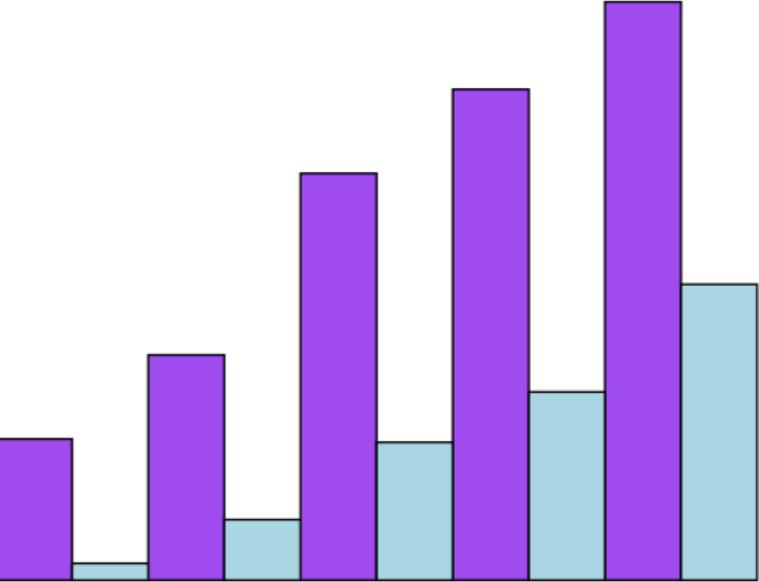
Histogram of HealthCare



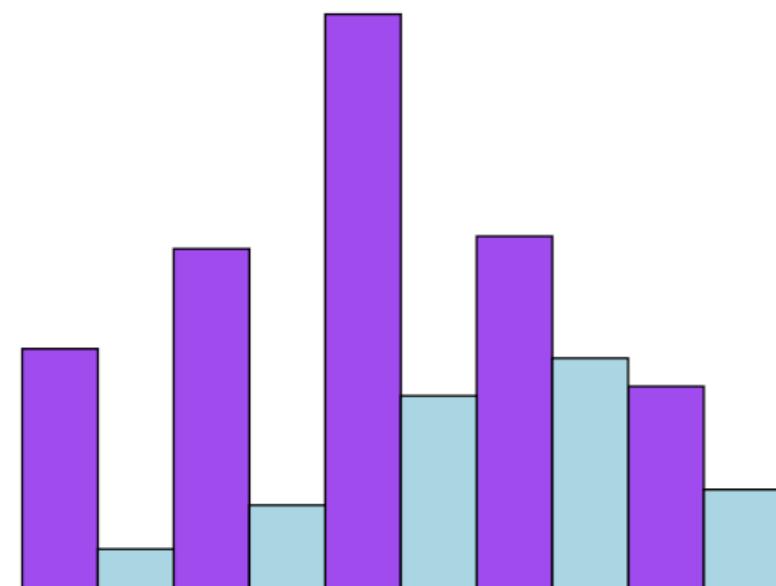
Histogram of Transport



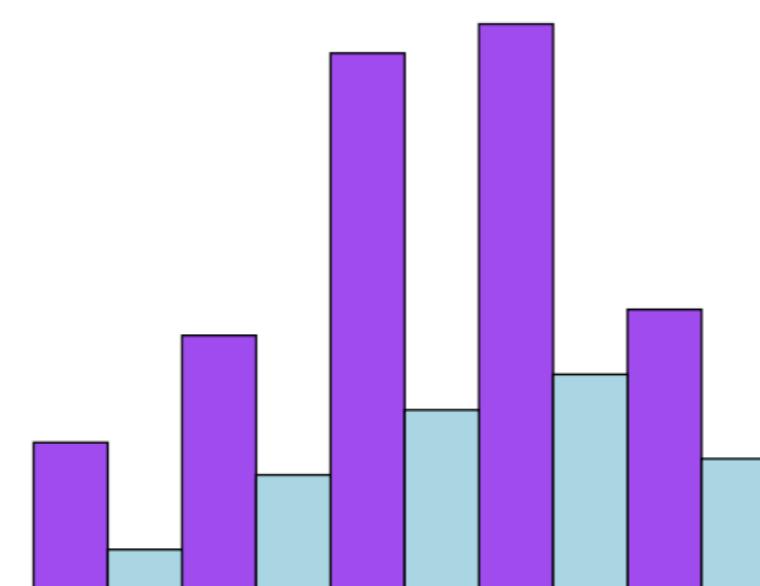
Histogram of Culture



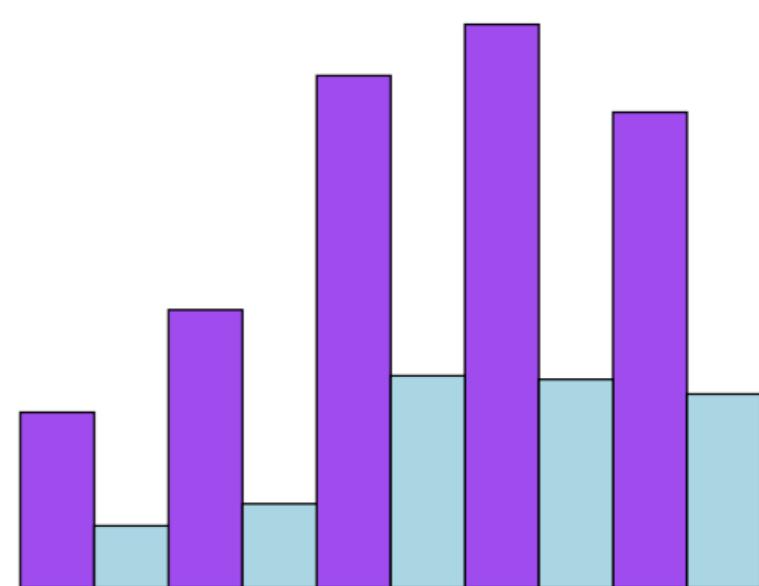
Histogram of Housing



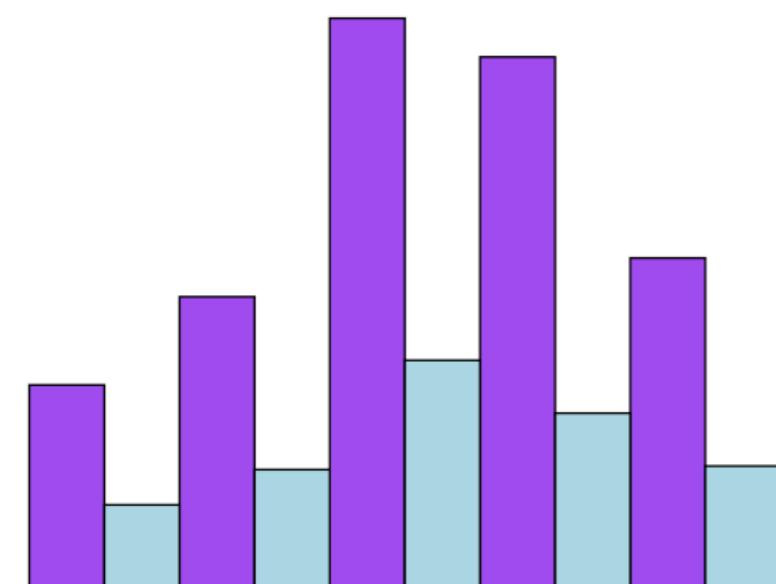
Histogram of Security



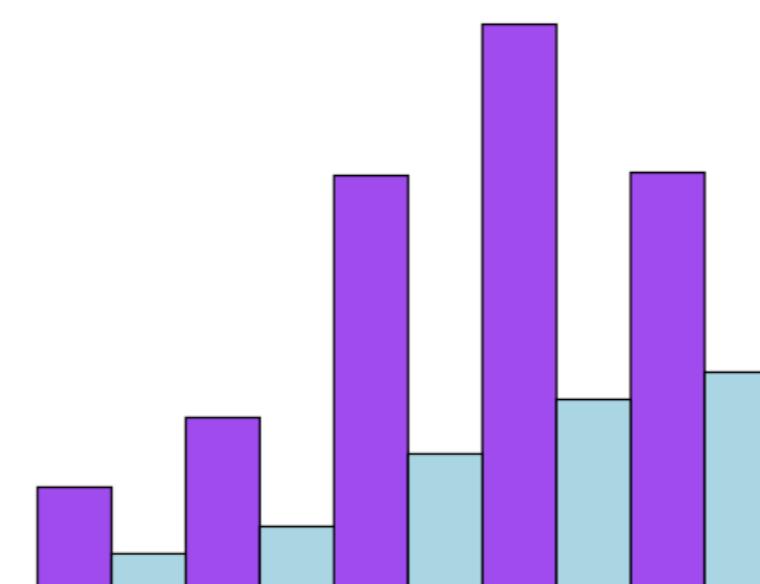
Histogram of Career



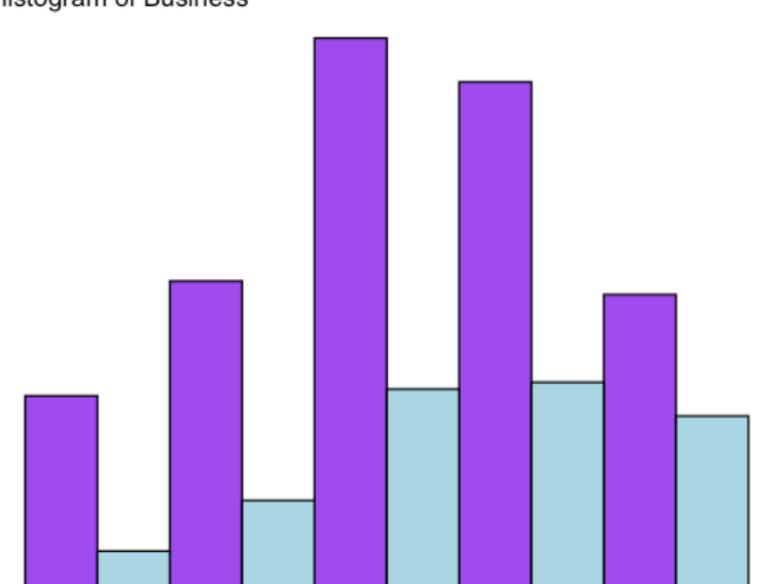
Histogram of Salary



Histogram of Education



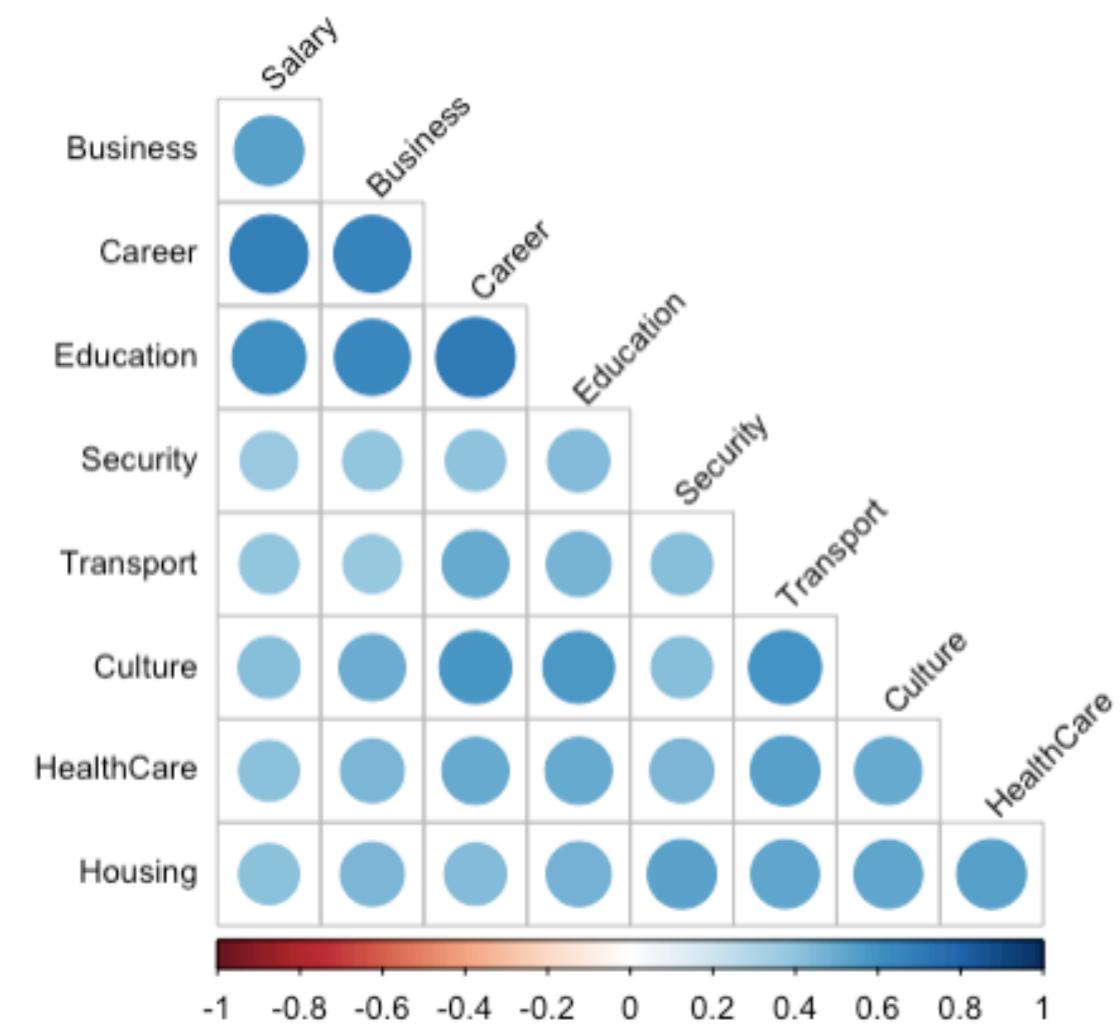
Histogram of Business



factor(migration_status)

- Stayed (Purple)
- Migrated (Light Blue)

Data Description



Basically, there is no extremely high correlation between characteristics. Also, based on the idea that quality of life is the common concept including all mentioned characteristics there should not be any issues with performing analysis.

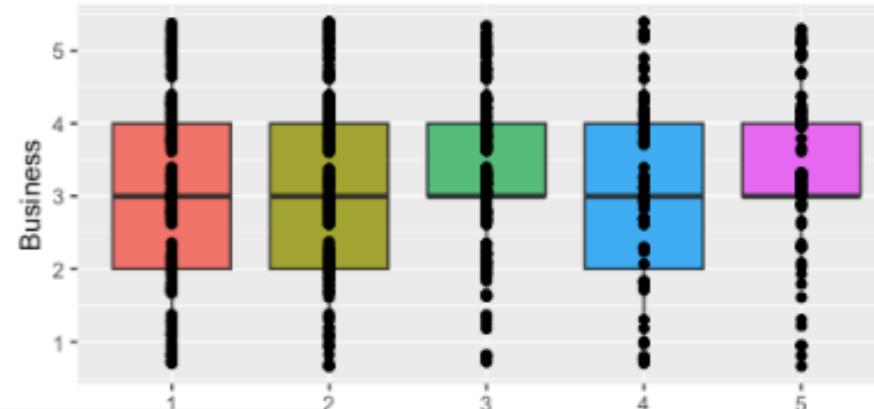
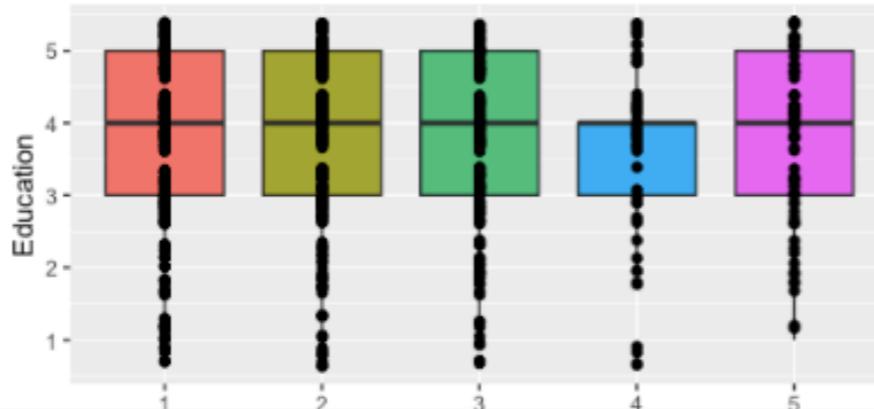
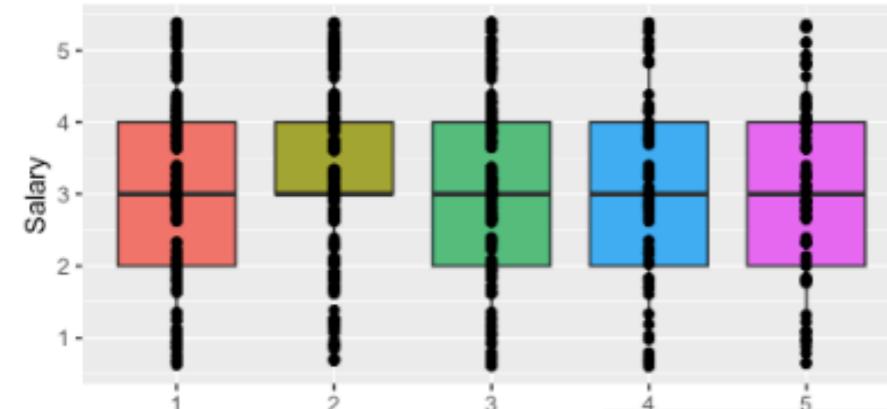
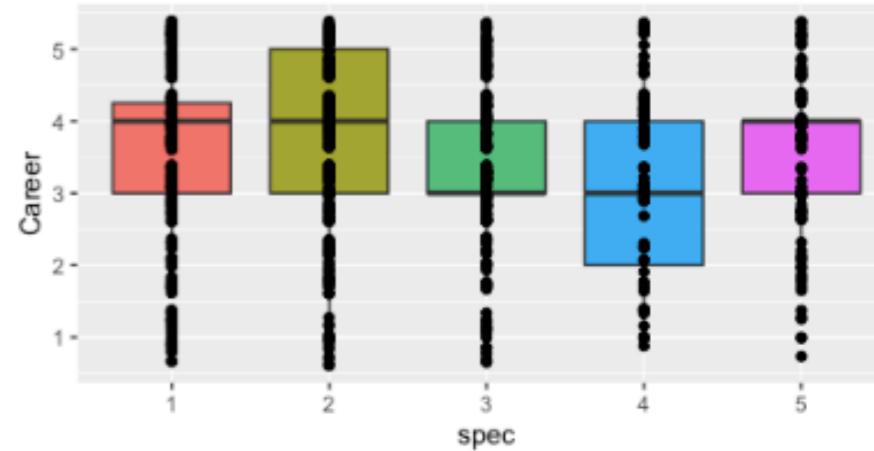
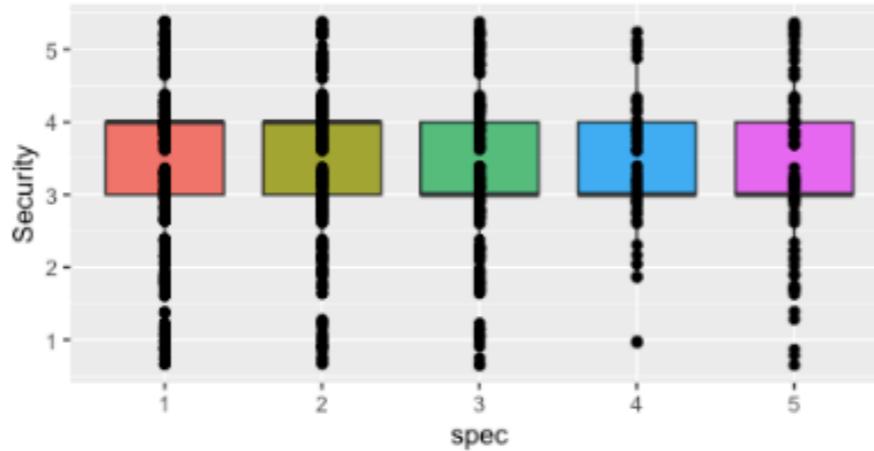
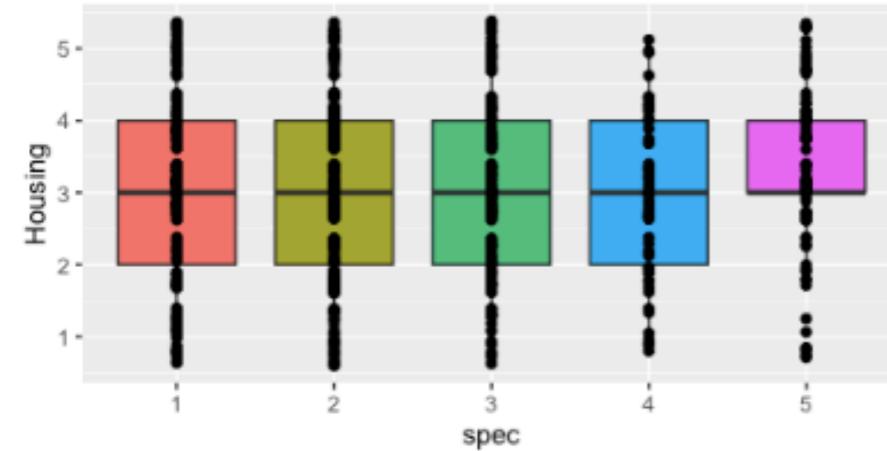
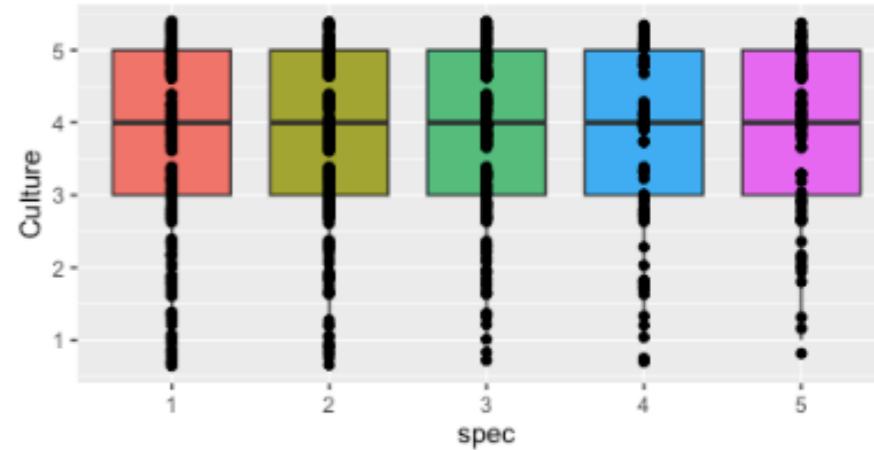
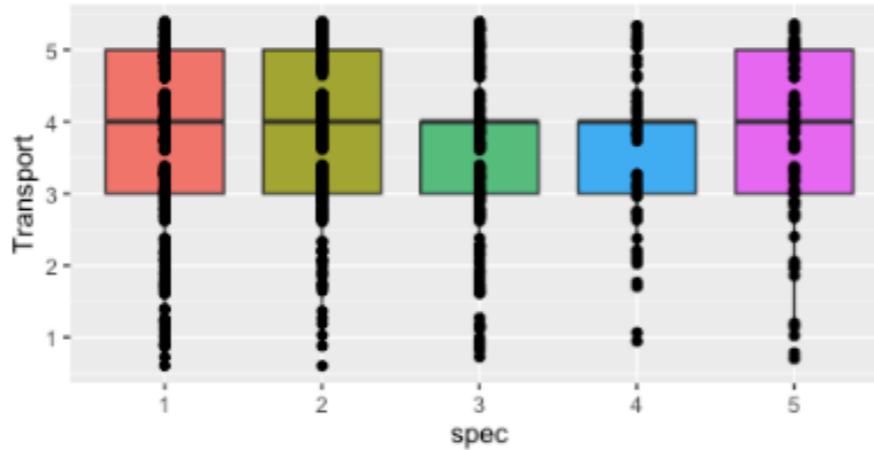
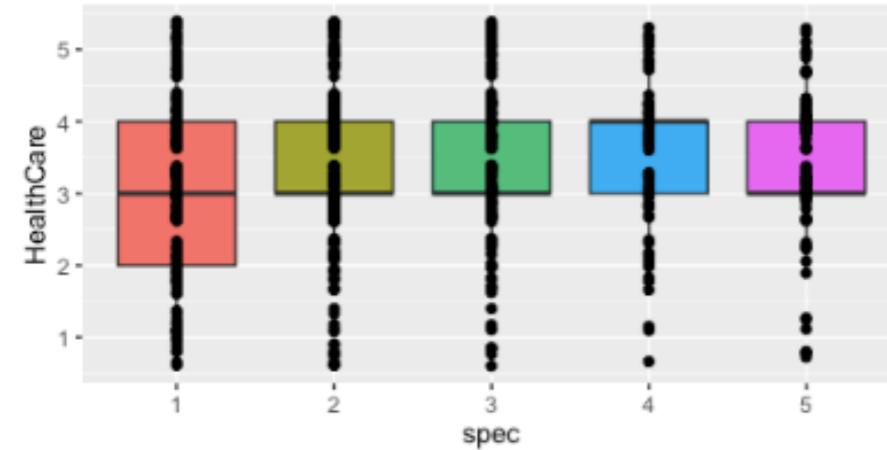
Characteristic	Group	Mean	Standard Deviation
HealthCare	Stayed	3.222628	1.117289
	Migrated	3.552885	1.029559
Transport	Stayed	3.52007	1.141521
	Migrated	3.77884	1.111701
Culture	Stayed	3.618613	1.253361
	Migrated	3.980769	1.089970
Housing	Stayed	2.963504	1.201211
	Migrated	3.408654	1.090726
Security	Stayed	3.324818	1.145839
	Migrated	3.418269	1.147588
Career	Stayed	3.441606	1.235617
	Migrated	3.509615	1.215799
Salary	Stayed	3.255474	1.214051
	Migrated	3.182692	1.229779
Education	Stayed	3.616788	1.142118
	Migrated	3.778846	1.170959
Business	Stayed	3.217153	1.202647
	Migrated	3.552885	1.145076

Data Description

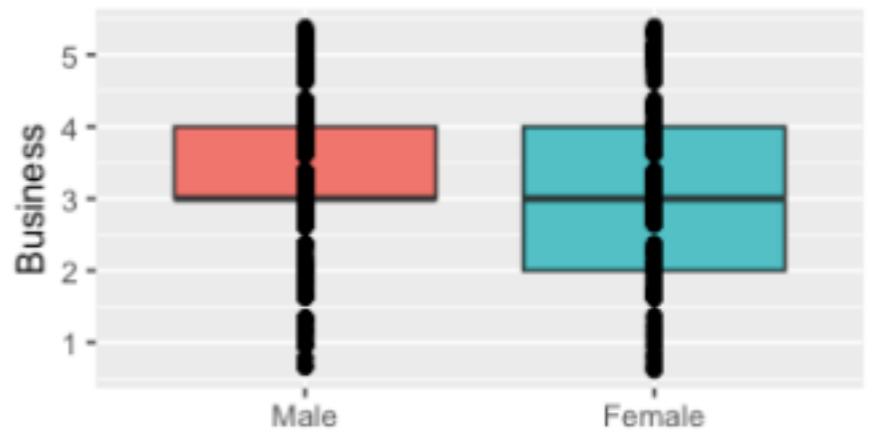
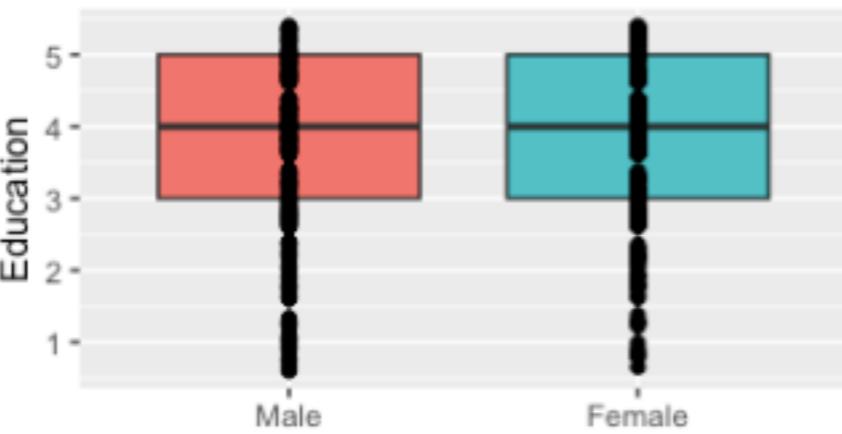
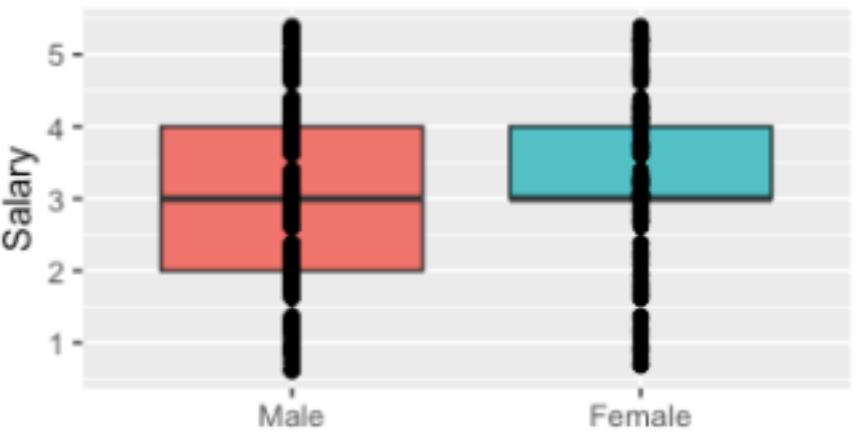
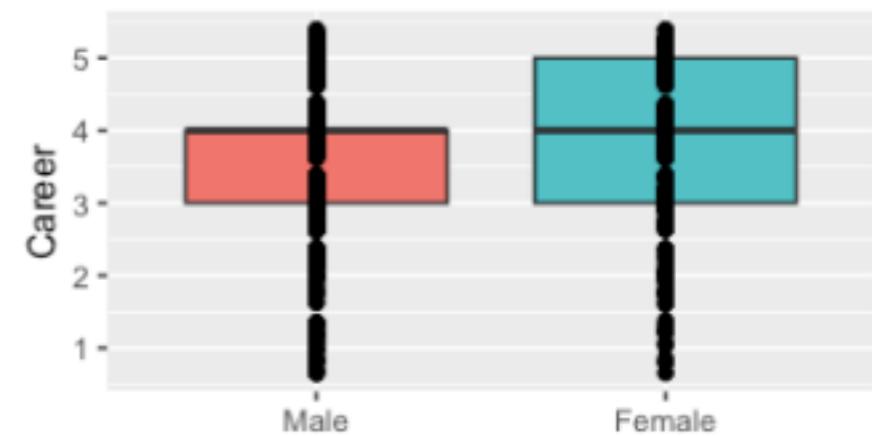
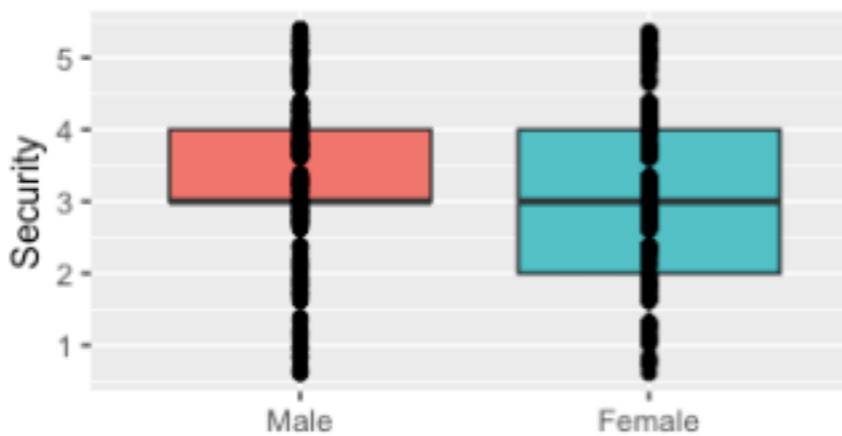
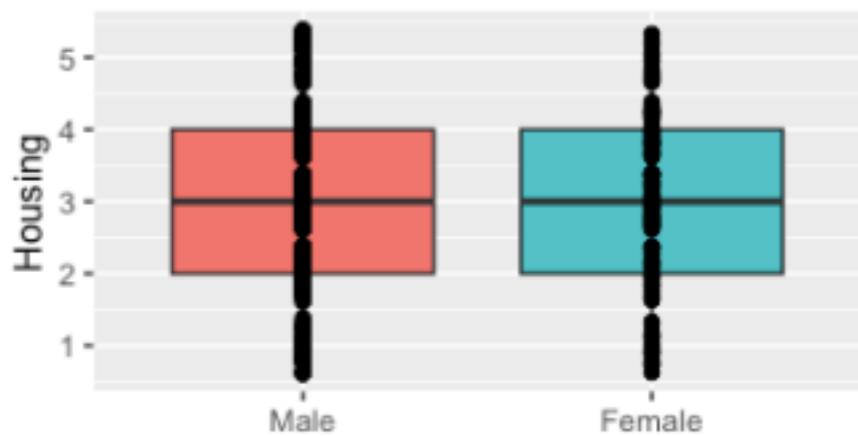
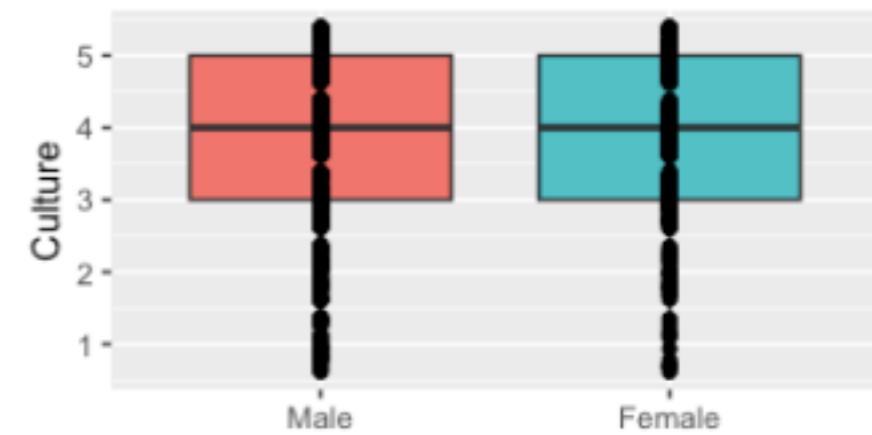
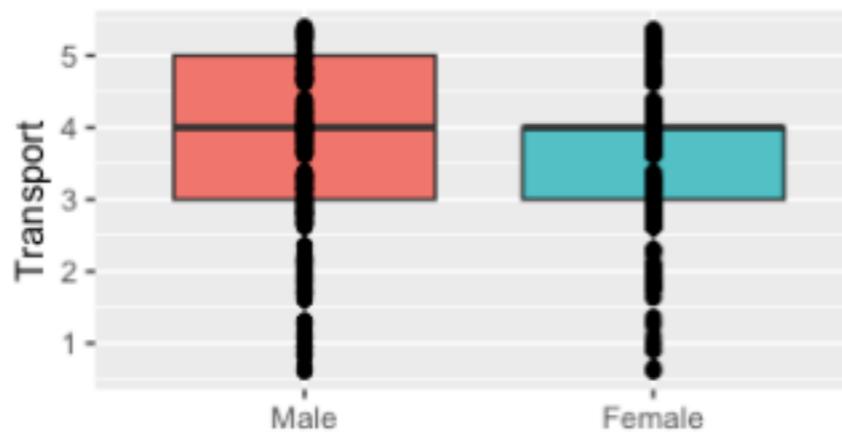
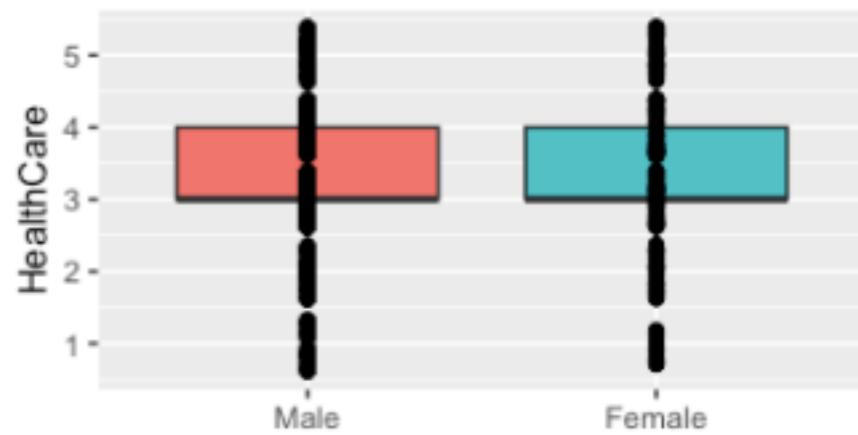
spec 1 2 3 4 5

spec 1 2 3 4 5

spec 1 2 3 4 5



Data Description



MANOVA: the difference by specialisation

To test the difference on the levels of satisfaction with quality of life characteristics in the place of studying between graduates from different specialisations the MANOVA test has been implemented due to using multiple dependent variables (quality of life characteristics) and categorical factor (specialisation)

	Df	Pillai	approx F	num Df	den Df	Pr(>F)
spec	4	0.057896	1.2173	36	2984	0.1755
Response Security :						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spec	4	0.37	0.09215	0.0698	0.9911	
Residuals	751	991.74	1.32056			
Response HealthCare :						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spec	4	8.51	2.1282	1.756	0.1359	
Residuals	751	910.19	1.2120			
Response Transport :						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spec	4	6.14	1.5348	1.1851	0.3159	
Residuals	751	972.56	1.2950			
Response Culture :						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spec	4	6.99	1.7464	1.1731	0.3213	
Residuals	751	1118.00	1.4887			
Response Housing :						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spec	4	6.15	1.5377	1.0902	0.3602	
Residuals	751	1059.26	1.4105			
Response Career :						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spec	4	2.65	0.66337	0.4373	0.7817	
Residuals	751	1139.16	1.51685			
Response Salary :						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spec	4	6.09	1.5218	1.0259	0.3929	
Residuals	751	1114.00	1.4834			
Response Education :						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spec	4	3.33	0.83153	0.6257	0.6443	
Residuals	751	997.99	1.32888			
Response Business :						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spec	4	2.52	0.62949	0.4389	0.7805	
Residuals	751	1077.05	1.43416			

According to MANOVA, there is no statistically significant differences between average levels of satisfaction with quality of life characteristics in the place of studying between graduates from different specialisations

MANOVA: the difference by gender

To test the difference on the levels of satisfaction with quality of life characteristics in the place of studying between males and females the MANOVA test has been implemented due to using multiple dependent variables (quality of life characteristics) and binary factor (gender)

```
Df Pillai approx F num Df den Df Pr(>F)
gender      1 0.030264   2.5868     9    746 0.006123 **

Response Salary :
Df Sum Sq Mean Sq F value Pr(>F)
gender      1   4.09   4.0941  2.7661 0.0967 .
Residuals  754 1116.00  1.4801
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response Business :
Df Sum Sq Mean Sq F value Pr(>F)
gender      1   4.03   4.0263  2.8226 0.09336 .
Residuals  754 1075.55  1.4265
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to MANOVA, there is statistically significant differences between average levels of satisfaction with quality of life characteristics in the place of studying between males and females.

On average, women (3.33) are more satisfied with salary at the place of study than men (3.18). Moreover, average level of satisfaction with the opportunity to launch own business at the place of study for men are higher (3.36) than for women (3.21).

Logistic regression: propensity to relocate

We want to predict the **propensity of relocation** based on individual profile and **subjectively perceived quality of life**.

M0: the initial model includes all variables

```
glm(DV ~ HealthCare + Transport + Culture + Housing + Security + Career + Salary + Education + Business +
as.factor(spec) + as.factor(mother_education) + as.factor(father_education) + as.factor(material_status) +
as.factor(gender), family = «binomial»)
```

With STEP ('backward') algorithm we have the best model according to AIC:

```
glm(DV ~ HealthCare + Transport + Culture + Housing + Security + Career + Salary + Education + Business +
as.factor(gender), family = «binomial»)
```

This model does not have the problem of **multicollinearity**, but has **heteroscedasticity** ($p\text{-value} > .05$ for the studentized Breusch-Pagan test), meaning that our estimates of logistic regression are still unbiased, but are not consistent, so that, we could not use these estimates to test the hypothesis.

To cope with heteroscedasticity robust standard errors are applied. There is statistically significant associations in the model between the fact of labour mobility and Housing conditions, Security level, Salary by profession and Business opportunities.

Thus, the higher levels of satisfaction with Housing conditions and Business opportunities (separately), increases the propensity to migrate, while the higher levels of satisfaction with Security and Salary (separately), decreases the propensity to migrate.

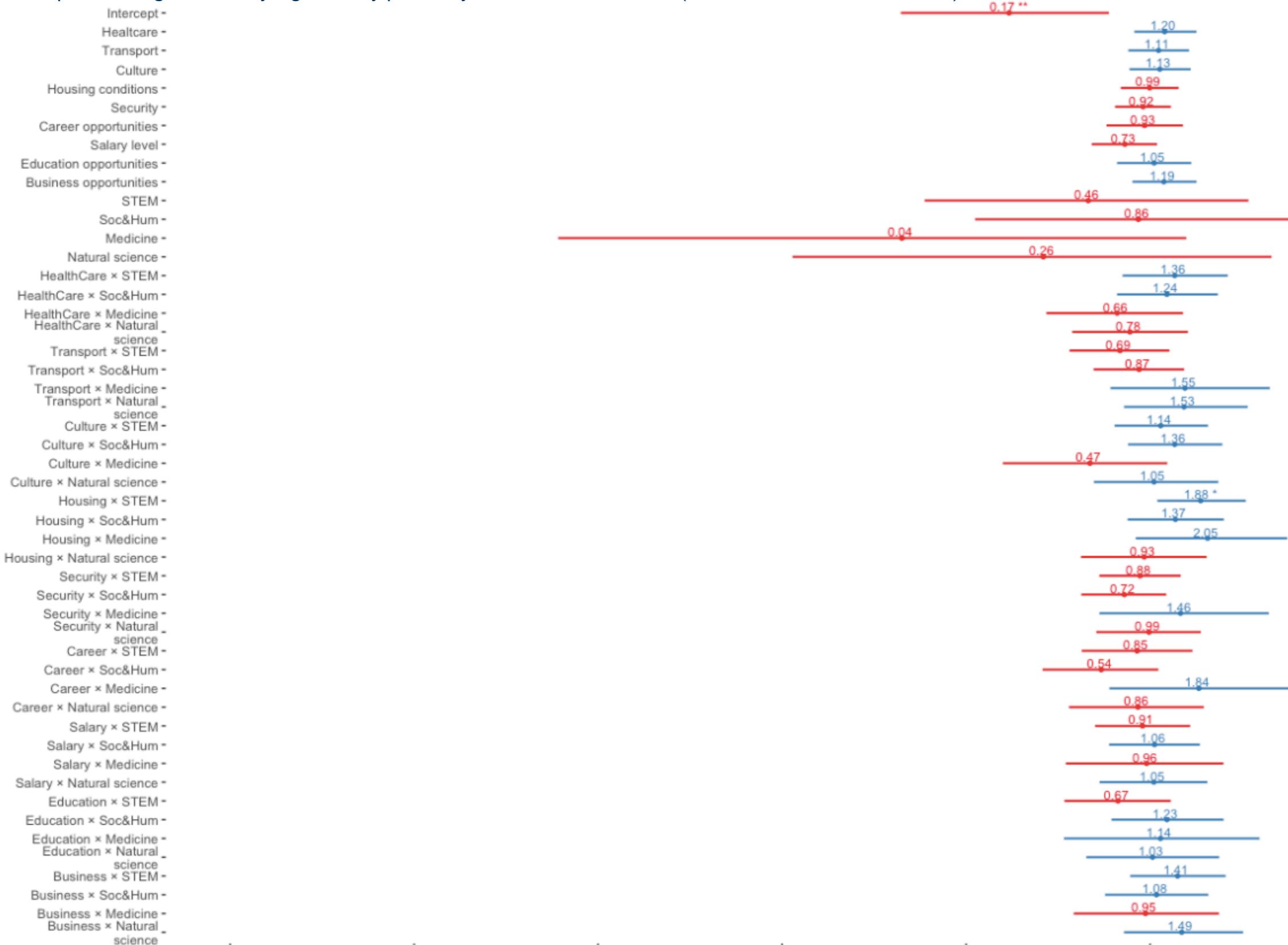
Predictors	Odds Ratios	CI	p
(Intercept)	0.26	0.01 – 3.96	0.346
HealthCare	1.22	0.98 – 1.52	0.073
Transport	1.02	0.82 – 1.27	0.848
Culture	1.24	1.00 – 1.54	0.047
Housing	1.44	1.17 – 1.78	0.001
Security	0.81	0.67 – 0.98	0.028
Career	0.83	0.64 – 1.07	0.148
Salary	0.76	0.61 – 0.93	0.009
Education	0.91	0.71 – 1.17	0.452
Business	1.39	1.12 – 1.74	0.003

Predictors	Odds Ratios	CI	p
(Intercept)	0.12	0.05 – 0.24	<0.001
HealthCare	1.22	0.99 – 1.51	0.066
Transport	1.03	0.84 – 1.27	0.776
Culture	1.22	0.99 – 1.50	0.060
Housing	1.38	1.13 – 1.70	0.002
Security	0.82	0.68 – 0.98	0.034
Career	0.80	0.62 – 1.02	0.069
Salary	0.75	0.61 – 0.91	0.005
Education	0.95	0.75 – 1.21	0.701
Business	1.41	1.14 – 1.75	0.002
gender [Female]	0.99	0.69 – 1.40	0.950

LR with interaction: propensity to relocate

`glm(DV ~ (HealthCare + Transport + Culture + Housing + Security + Career + Salary + Education + Business) * as.factor(spec), family = "binomial", data = db)`

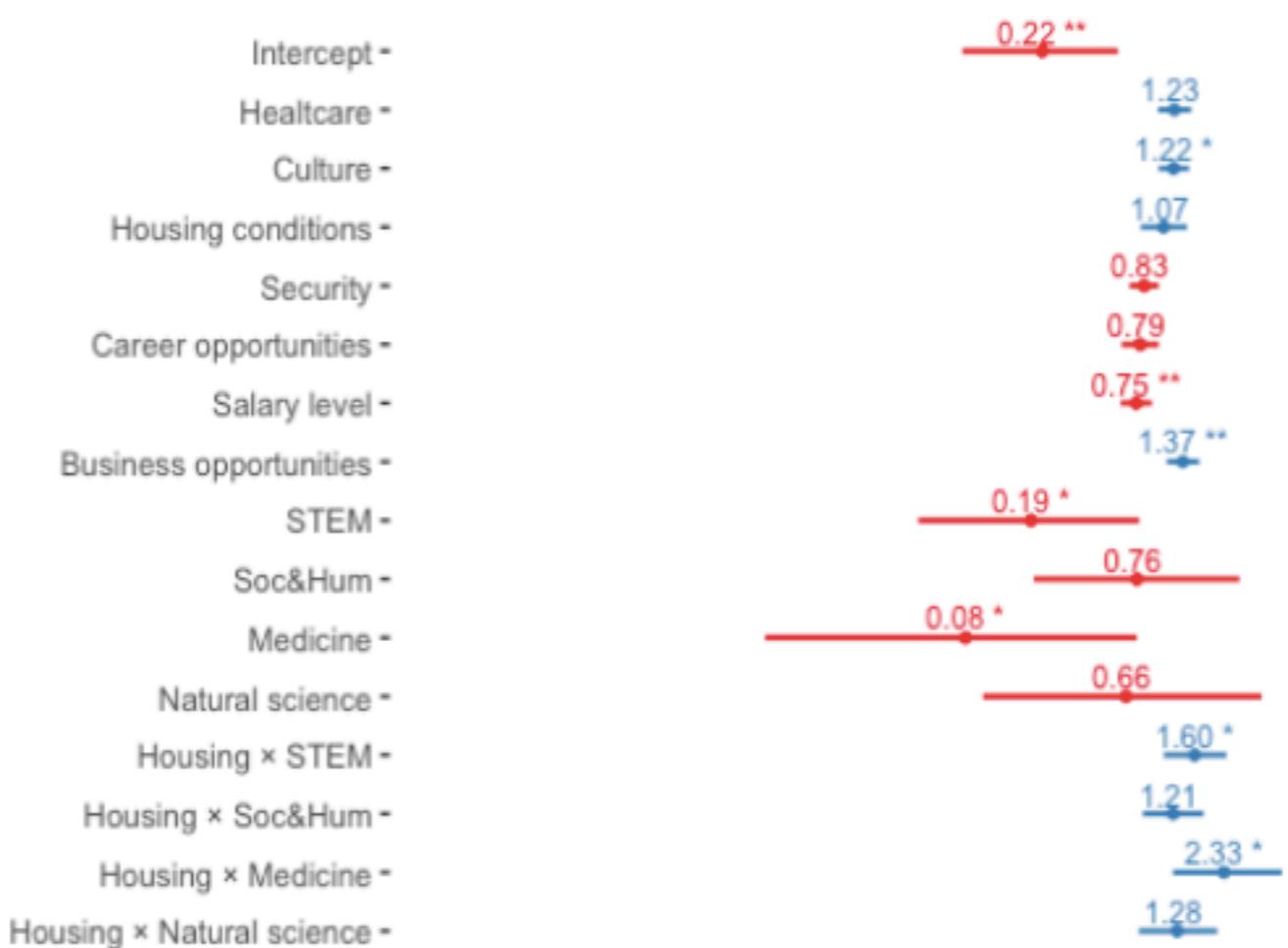
According to results we can claim that for students of STEM specialisation compared to graduates from Economics the satisfaction with housing condition path the region of study significantly positively leads to the relocation (the chance to relocate is 88%).



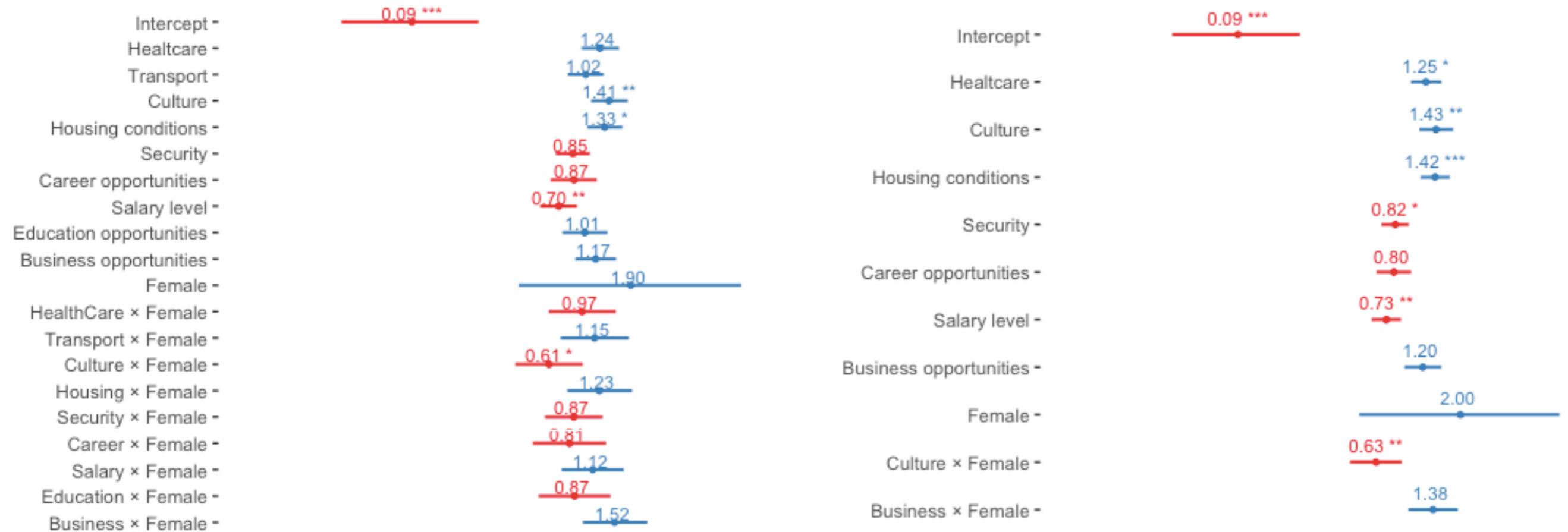
LR with interaction: propensity to relocate

Performing the interaction effect of specialisation on the level of satisfaction with Housing condition at the region of study there is a positive significant associations between the fact of the relocation at satisfaction with Culture, Business opportunity (separately) and negative association between the fact of the relocation and Salary level.

Moreover, for STEM graduates the propensity of migration is increasing with higher level of Housing conditions compared to Economic graduates, and the same for Medicine graduates compared to Economic graduates



LR with interaction: propensity to relocate



Interaction between quality of life characteristics and gender shows some results. The stepwise algorithm has been applied to choose the best model.

For female graduates the higher estimates of satisfaction with culture to less extant increase the chance of relocation comparing to male graduates. However, there is no enough evidence to observe significant difference between females and males in chances to relocate based on estimates of others place characteristics

Limitations

- Lack of opportunity generalise results on regional level because of provided data (only individual level)
- Case-bias (Russian sample) with open-source data demanding full rely on data from the researcher side

Correspondence analysis (gender and indicators)

	Gender		
	Male	Female	Sum
HealthCare			
Not satisfied	5.1	2.7	7.8
Quite not satisfied	8.0	5.5	13.5
Somehow satisfied, somehow no	21.6	11.4	33.0
Quite satisfied	17.6	12.6	30.2
Satisfied	8.5	7.0	15.5
Sum	60.9	39.1	100.0
	Gender		
	Male	Female	Sum

	Gender		
	Male	Female	Sum
Housing			
Not satisfied	7.6	5.4	13.0
Quite not satisfied	10.6	6.2	16.8
Somehow satisfied, somehow no	20.5	12.2	32.6
Quite satisfied	14.1	10.1	24.2
Satisfied	8.1	5.2	13.3
Sum	60.9	39.1	100.0
	Gender		
	Male	Female	Sum

	Gender		
	Male	Female	Sum
Salary			
Not satisfied	6.2	3.1	9.3
Quite not satisfied	10.0	5.1	15.1
Somehow satisfied, somehow no	19.3	11.3	30.6
Quite satisfied	14.8	10.8	25.5
Satisfied	10.6	8.9	19.5
Sum	60.9	39.1	100.0
	Gender		
	Male	Female	Sum

Does gender have any influence on satisfaction level?

Seems that yes, and we have already seen it at MANOVA and Logreg steps

	Male	Female	Housing	Not satisfied	243	166	Salary	Not satisfied	76	106
HealthCare				Quite not satisfied	223	129		Quite not satisfied	26	85
Not satisfied	71	45		Somehow satisfied, somehow no	91	70		Somehow satisfied, somehow no	309	123
Quite not satisfied	97	203		Quite satisfied	105	26		Quite satisfied	195	182
Somehow satisfied, somehow no	167	204		Satisfied	72	159		Satisfied	269	165
Quite satisfied	120	211								
Satisfied	86	114	Security	Not satisfied	117	99	Education	Not satisfied	157	228
				Quite not satisfied	245	143		Quite not satisfied	67	138
Transport				Somehow satisfied, somehow no	176	156		Somehow satisfied, somehow no	27	50
Not satisfied	111	76		Quite satisfied	146	54		Quite satisfied	112	153
Quite not satisfied	50	40		Satisfied	64	54		Satisfied	261	131
Somehow satisfied, somehow no	147	147	Career	Not satisfied	37	68	Business	Not satisfied	174	236
Quite satisfied	156	266		Quite not satisfied	226	162		Quite not satisfied	134	267
Satisfied	168	251		Somehow satisfied, somehow no	204	149		Somehow satisfied, somehow no	48	77
				Quite satisfied	283	126		Quite satisfied	75	68
Culture				Satisfied	108	90		Satisfied	139	101
Not satisfied	299	155								
Quite not satisfied	97	67								
Somehow satisfied, somehow no	61	43								
Quite satisfied	190	237								
Satisfied	140	249								

There is some variations in satisfaction levels for Males and Females

	Gender	HealthCare	Transport	Culture	Housing	Security	Career	Salary	Education	Business
Satisfaction level		Male	Female	Male	Female	Male	Female	Male	Female	Male
Not satisfied	Male	71	50	61	105	64	68	85	50	68
	Female	37	26	27	75	45	40	43	26	54
Quite not satisfied	Male	111	97	91	146	108	106	138	77	114
	Female	76	67	48	86	76	67	70	54	90
Somehow satisfied&no	Male	299	223	176	283	269	228	267	211	251
	Female	157	134	120	168	155	129	156	126	165
Quite satisfied	Male	243	245	204	195	261	236	204	266	249
	Female	174	167	156	140	166	143	149	182	131
Satisfied	Male	117	226	309	112	139	203	147	237	159
	Female	97	147	190	72	99	162	123	153	101

	Male	Female
HealthCare	541	777
Transport	632	780
Culture	787	751
Housing	734	550
Security	748	506
Career	858	595
Salary	875	661
Education	624	700
Business	570	749

Correspondence analysis (gender and indicators)

Does gender have any influence on satisfaction level of life quality indicators?
Seems that yes, and we have already seen it at MANOVA and Logreg steps

		Gender					Gender					Gender			
		Male	Female	Sum			Male	Female	Sum			Male	Female	Sum	
HealthCare					Housing					Salary					
Not satisfied		5.1	2.7	7.8	Not satisfied		7.6	5.4	13.0	Not satisfied		6.2	3.1	9.3	
Quite not satisfied		8.0	5.5	13.5	Quite not satisfied		10.6	6.2	16.8	Quite not satisfied		10.0	5.1	15.1	
Somehow satisfied, somehow no	21.6	11.4	33.0	Somehow satisfied, somehow no	20.5	12.2	32.6	Somehow satisfied, somehow no	19.3	11.3	30.6	Somehow satisfied, somehow no	19.3	11.3	30.6
Quite satisfied	17.6	12.6	30.2	Quite satisfied	14.1	10.1	24.2	Quite satisfied	14.8	10.8	25.5	Quite satisfied	14.8	10.8	25.5
Satisfied	8.5	7.0	15.5	Satisfied	8.1	5.2	13.3	Satisfied	10.6	8.9	19.5	Satisfied	10.6	8.9	19.5
Sum	60.9	39.1	100.0	Sum	60.9	39.1	100.0	Sum	60.9	39.1	100.0	Sum	60.9	39.1	100.0
		Gender					Gender					Gender			
Transport		Male	Female	Sum			Security					Education			
Not satisfied		3.6	1.9	5.5	Not satisfied		4.6	3.3	7.9	Not satisfied		3.6	1.9	5.5	
Quite not satisfied		7.0	4.8	11.9	Quite not satisfied		7.8	5.5	13.3	Quite not satisfied		5.6	3.9	9.5	
Somehow satisfied, somehow no	16.1	9.7	25.8	Somehow satisfied, somehow no	19.5	11.2	30.7	Somehow satisfied, somehow no	15.3	9.1	24.4	Somehow satisfied, somehow no	15.3	9.1	24.4
Quite satisfied	17.7	12.1	29.8	Quite satisfied	18.9	12.0	30.9	Quite satisfied	19.2	13.2	32.4	Quite satisfied	19.2	13.2	32.4
Satisfied	16.4	10.6	27.0	Satisfied	10.1	7.2	17.2	Satisfied	17.1	11.1	28.2	Satisfied	17.1	11.1	28.2
Sum	60.9	39.1	100.0	Sum	60.9	39.1	100.0	Sum	60.9	39.1	100.0	Sum	60.9	39.1	100.0
		Gender					Gender					Gender			
Culture		Male	Female	Sum			Career					Business			
Not satisfied		4.4	2.0	6.4	Not satisfied		4.9	2.9	7.8	Not satisfied		4.9	3.9	8.8	
Quite not satisfied		6.6	3.5	10.1	Quite not satisfied		7.7	4.8	12.5	Quite not satisfied		8.2	6.5	14.8	
Somehow satisfied, somehow no	12.7	8.7	21.4	Somehow satisfied, somehow no	16.5	9.3	25.8	Somehow satisfied, somehow no	18.2	11.9	30.1	Somehow satisfied, somehow no	18.2	11.9	30.1
Quite satisfied	14.8	11.3	26.0	Quite satisfied	17.1	10.3	27.4	Quite satisfied	18.0	9.5	27.5	Quite satisfied	18.0	9.5	27.5
Satisfied	22.4	13.7	36.1	Satisfied	14.7	11.7	26.4	Satisfied	11.5	7.3	18.8	Satisfied	11.5	7.3	18.8
Sum	60.9	39.1	100.0	Sum	60.9	39.1	100.0	Sum	60.9	39.1	100.0	Sum	60.9	39.1	100.0

Satisfaction levels for Males and Females vary among life quality indicators

		HealthCare Transport Culture Housing Security Career Salary Education Business									
Satisfaction level	Gender										
Not satisfied	Male	71	50	61	105	64	68	85	50	68	
	Female	37	26	27	75	45	40	43	26	54	
Quite not satisfied	Male	111	97	91	146	108	106	138	77	114	
	Female	76	67	48	86	76	67	70	54	90	
Somehow satisfied&no	Male	299	223	176	283	269	228	267	211	251	
	Female	157	134	120	168	155	129	156	126	165	
Quite satisfied	Male	243	245	204	195	261	236	204	266	249	
	Female	174	167	156	140	166	143	149	182	131	
Satisfied	Male	117	226	309	112	139	203	147	237	159	
	Female	97	147	190	72	99	162	123	153	101	

Correspondence analysis (gender and indicators)

		HealthCare						
Gender		Not satisfied	Quite not satisfied	Somehow satisfied, somehow no	Quite satisfied	Satisfied	Sum	
Male		5.1	8.0		21.6	17.6	8.5	60.9
Female		2.7	5.5		11.4	12.6	7.0	39.1
Sum		7.8	13.5		33.0	30.2	15.5	100.0
		Transport						
Gender		Not satisfied	Quite not satisfied	Somehow satisfied, somehow no	Quite satisfied	Satisfied	Sum	
Male		3.6	7.0		16.1	17.7	16.4	60.9
Female		1.9	4.8		9.7	12.1	10.6	39.1
Sum		5.5	11.9		25.8	29.8	27.0	100.0
		Culture						
Gender		Not satisfied	Quite not satisfied	Somehow satisfied, somehow no	Quite satisfied	Satisfied	Sum	
Male		4.4	6.6		12.7	14.8	22.4	60.9
Female		2.0	3.5		8.7	11.3	13.7	39.1
Sum		6.4	10.1		21.4	26.0	36.1	100.0
		Housing						
Gender		Not satisfied	Quite not satisfied	Somehow satisfied, somehow no	Quite satisfied	Satisfied	Sum	
Male		7.6	10.6		20.5	14.1	8.1	60.9
Female		5.4	6.2		12.2	10.1	5.2	39.1
Sum		13.0	16.8		32.6	24.2	13.3	100.0
		Security						
Gender		Not satisfied	Quite not satisfied	Somehow satisfied, somehow no	Quite satisfied	Satisfied	Sum	
Male		4.6	7.8		19.5	18.9	10.1	60.9
Female		3.3	5.5		11.2	12.0	7.2	39.1
Sum		7.9	13.3		30.7	30.9	17.2	100.0
		Career						
Gender		Not satisfied	Quite not satisfied	Somehow satisfied, somehow no	Quite satisfied	Satisfied	Sum	
Male		4.9	7.7		16.5	17.1	14.7	60.9
Female		2.9	4.8		9.3	10.3	11.7	39.1
Sum		7.8	12.5		25.8	27.4	26.4	100.0
		Salary						
Gender		Not satisfied	Quite not satisfied	Somehow satisfied, somehow no	Quite satisfied	Satisfied	Sum	
Male		6.2	10.0		19.3	14.8	10.6	60.9
Female		3.1	5.1		11.3	10.8	8.9	39.1
Sum		9.3	15.1		30.6	25.5	19.5	100.0
		Education						
Gender		Not satisfied	Quite not satisfied	Somehow satisfied, somehow no	Quite satisfied	Satisfied	Sum	
Male		3.6	5.6		15.3	19.2	17.1	60.9
Female		1.9	3.9		9.1	13.2	11.1	39.1
Sum		5.5	9.5		24.4	32.4	28.2	100.0
		Business						
Gender		Not satisfied	Quite not satisfied	Somehow satisfied, somehow no	Quite satisfied	Satisfied	Sum	
Male		4.9	8.2		18.2	18.0	11.5	60.9
Female		3.9	6.5		11.9	9.5	7.3	39.1
Sum		8.8	14.8		30.1	27.5	18.8	100.0

The chi square of independence between the two variables = 10.1138
(p-value = 0.03855368)

The chi square of independence between the two variables = 1.710934
(p-value = 0.7887305)

The chi square of independence between the two variables = 7.019646
(p-value = 0.1348537)

The chi square of independence between the two variables = 2.564295
(p-value = 0.633161)

The chi square of independence between the two variables = 2.375539
(p-value = 0.6670523)

The chi square of independence between the two variables = 6.095219
(p-value = 0.1921492)

The chi square of independence between the two variables = 11.24951
(p-value = 0.02389845)

The chi square of independence between the two variables = 1.863389
(p-value = 0.7608668)

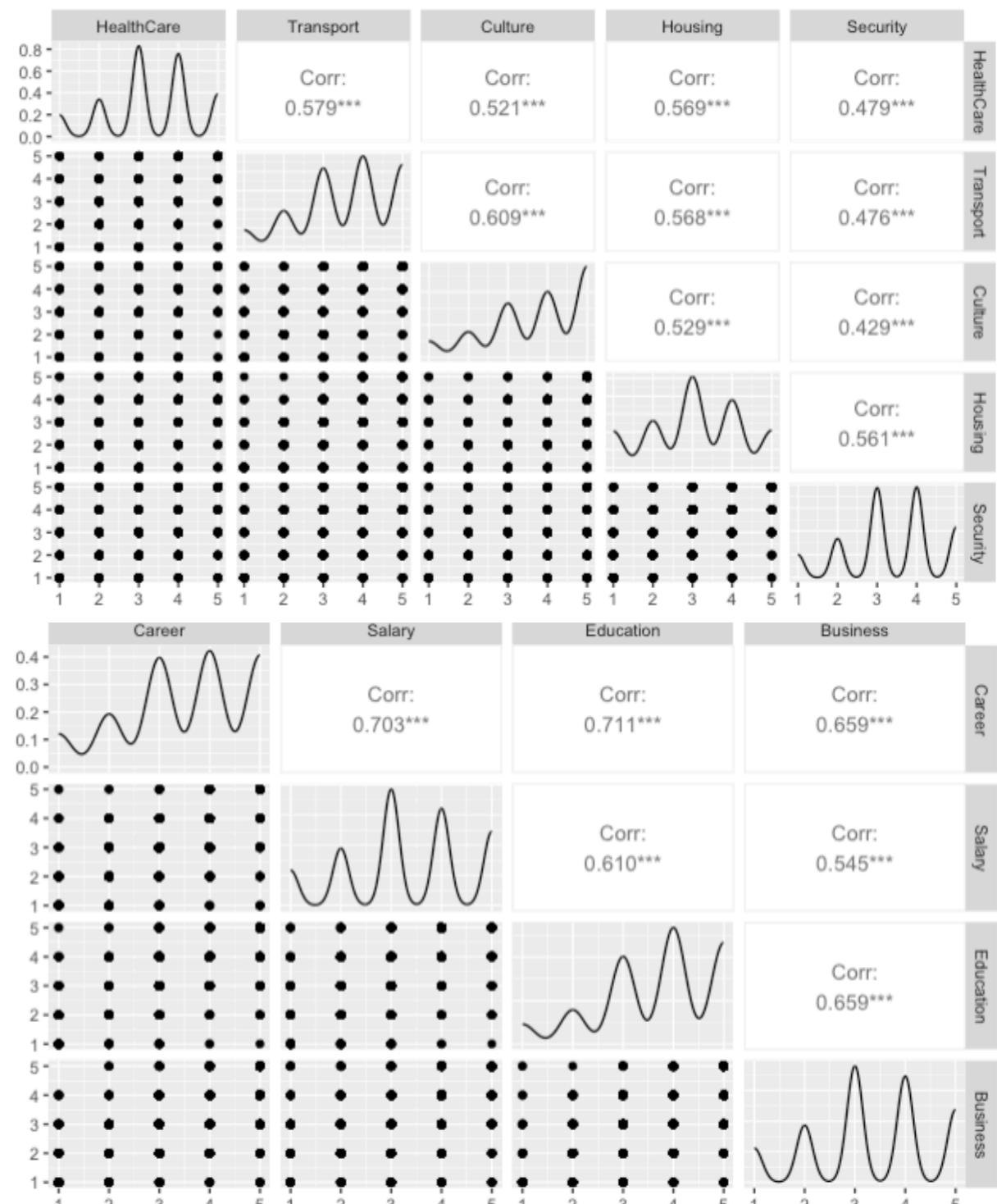
The chi square of independence between the two variables = 6.996165
(p-value = 0.136091)

Canonical correlation analysis

There are 2 «sets» of variables corresponding to ***social infrastructure*** and ***economical infrastructure***, and we might be interested in assessing the relationships between these sets of variables. So we want to find the linear functions of the variables in one set that maximally correlate with linear functions of variables in the other set.

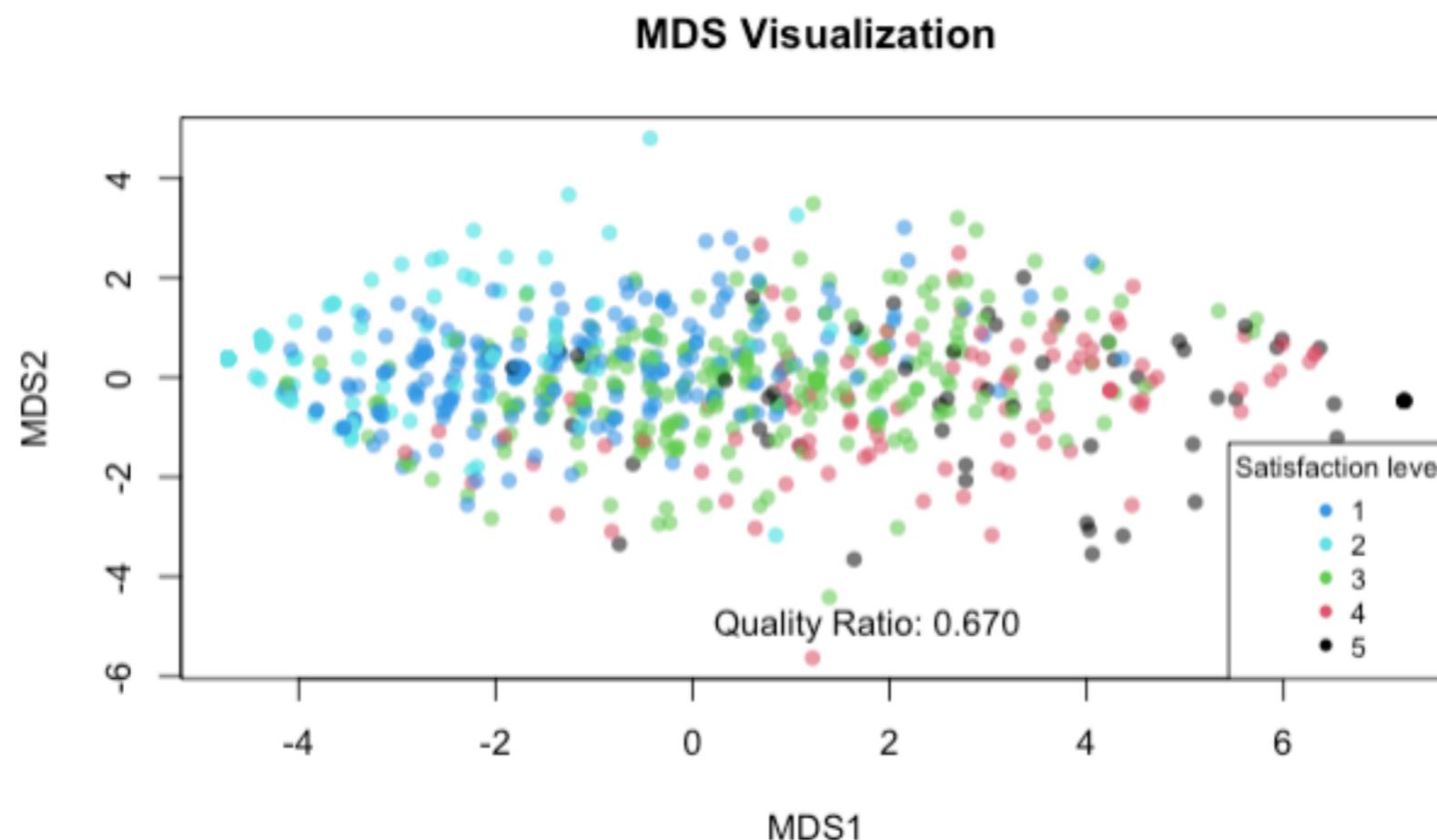
Tests of Canonical Dimensions					
Dim	Canonical_Corr	F_statistic	df1	df2	p_value
1	0.7198293	60.4284704	20	4554.676	0.0000000
2	0.1983840	5.8294334	12	3635.554	0.0000000
3	0.0969366	2.3602423	6	2750.000	0.0281599
4	0.0288329	0.5724357	2	1376.000	0.5642840

Standardized Canonical Coefficients			
Variable	Dim_1	Dim_2	Dim_3
HealthCare	-0.2576939	0.0688693	0.0078156
Transport	-0.1600523	-0.7717832	1.1175988
Culture	-0.5200579	-0.4726919	-1.0995345
Housing	-0.1367694	0.9256999	0.0351809
Security	-0.1632234	0.4322446	0.1172623
Career	-0.4185853	-1.2679507	0.0575686
Salary	-0.0516803	0.5480185	1.2580896
Education	-0.4596466	-0.1998430	-0.7130622
Business	-0.1969018	1.1665452	-0.3985529



Multidimensional scaling

At this step we want to explain structure or pattern amongst the subjectively perceived quality of life indicators: *Can we say that some of indicators seem alike or seem different from subjectively perceived satisfaction levels?*



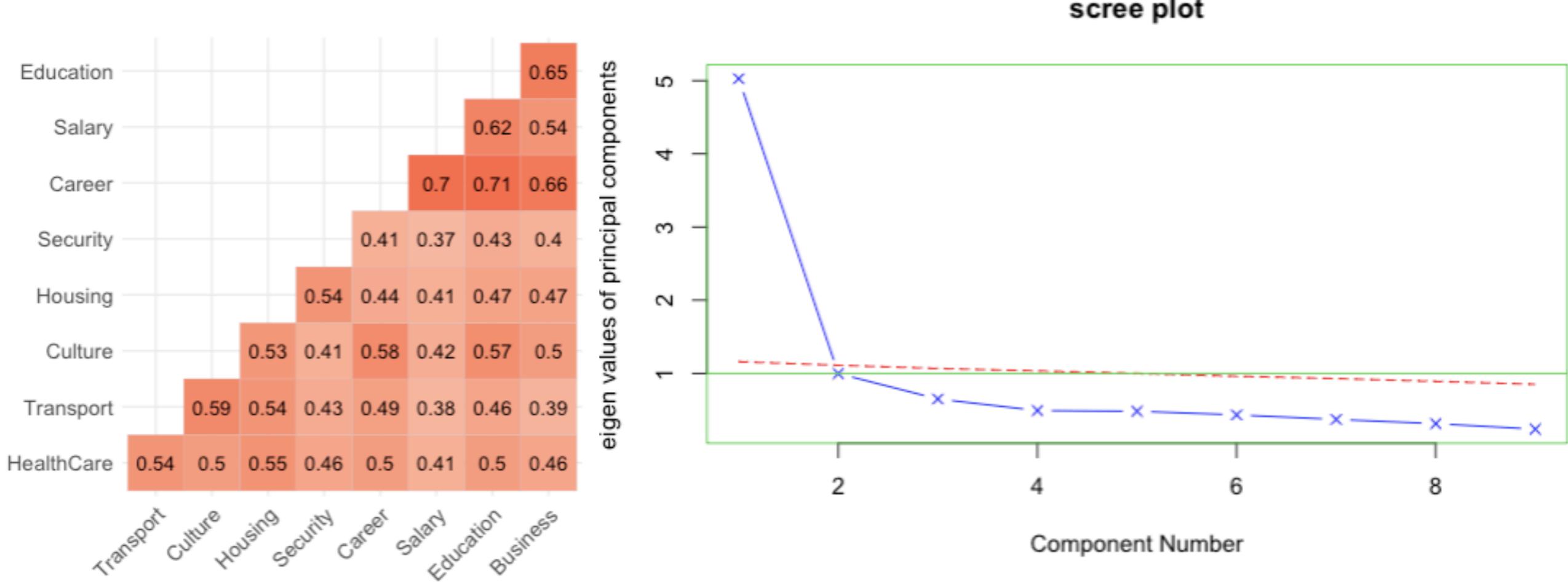
K = 2 (from CCorr and PCA insights)

K = 3; 9 also work

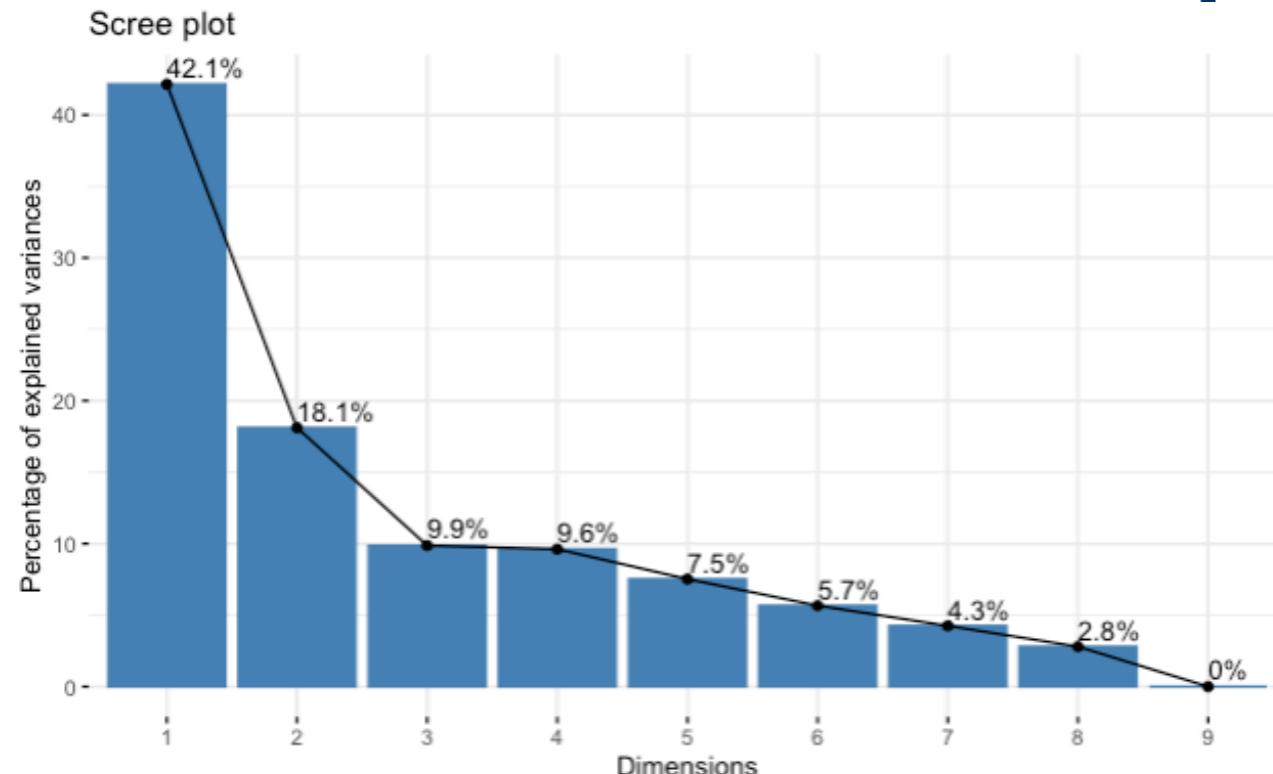
PCA

We want to describe the variation in set of correlated variables (*HealthCare*, *Culture*, *Transport*, *Housing*, *Security*, *Career*, *Salary*, *Education for kids*, *Business opportunities*) in terms of new uncorrelated variables, each of which is a linear combination of initial variables.

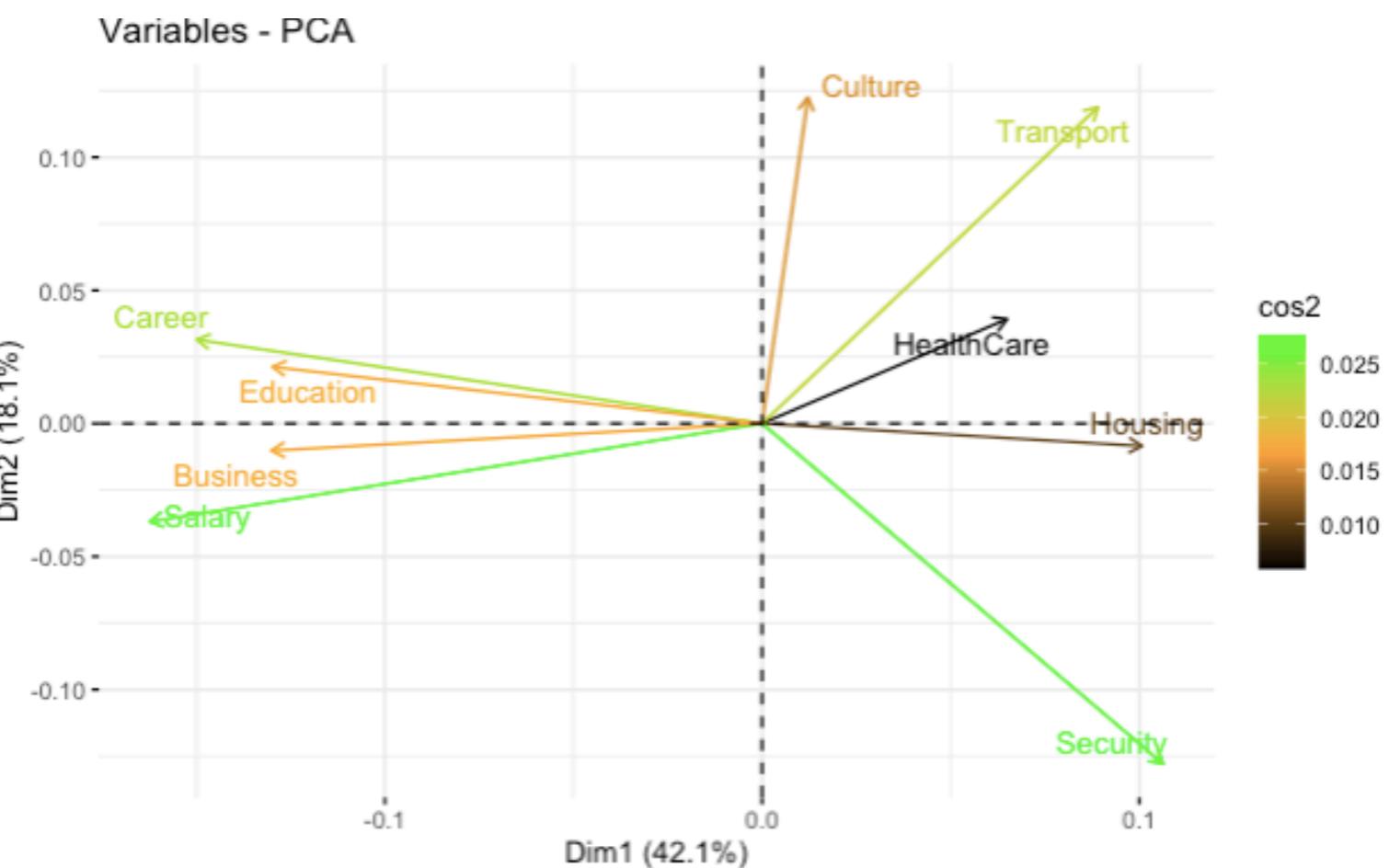
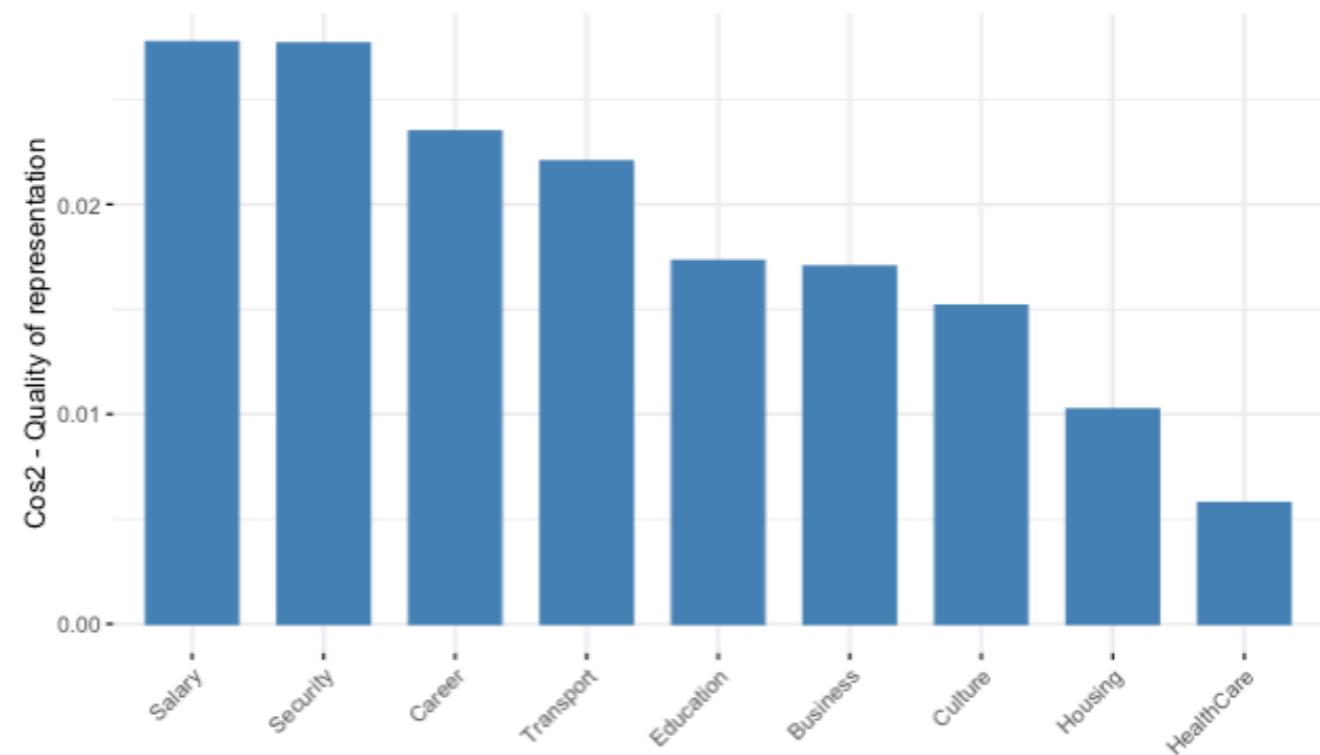
We assume that the first few components will account for a substantial proportion of the variation in the original variables and can be used to provide a convenient lower-dimensional summary of these variables that might prove useful for a variety of reasons.



PCA



Cos2 of variables to Dim-1-2



PCA : varimax rotation

	RC1 <S3: AsIs>	RC2 <S3: AsIs>
HealthCare	0.32	0.71
Transport	0.25	0.76
Culture	0.44	0.64
Housing	0.25	0.79
Security	0.20	0.71
Career	0.84	0.33
Salary	0.82	0.21
Education	0.78	0.37
Business	0.76	0.32

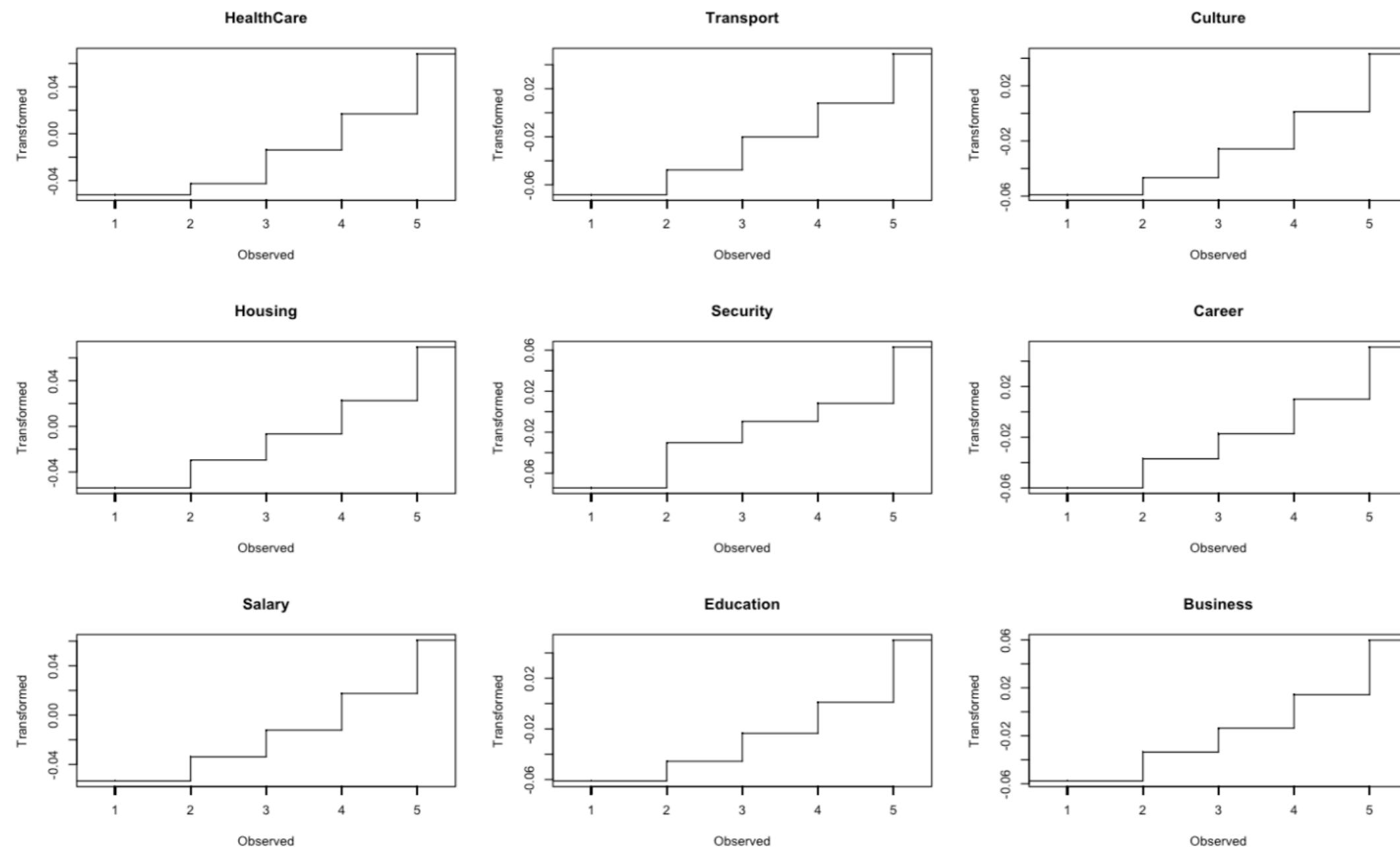
We also transform computed components into new ones for easier interpretation with the orthogonal «varimax» rotation, limiting the correlation between components.

Component 1: ***social infrastructure*** (Healthcare, Transport, Culture, Housing, Security)

Component 2: ***economical infrastructure*** (Career opportunities, Salary, Education for kids, Business opportunities)

NLPCA

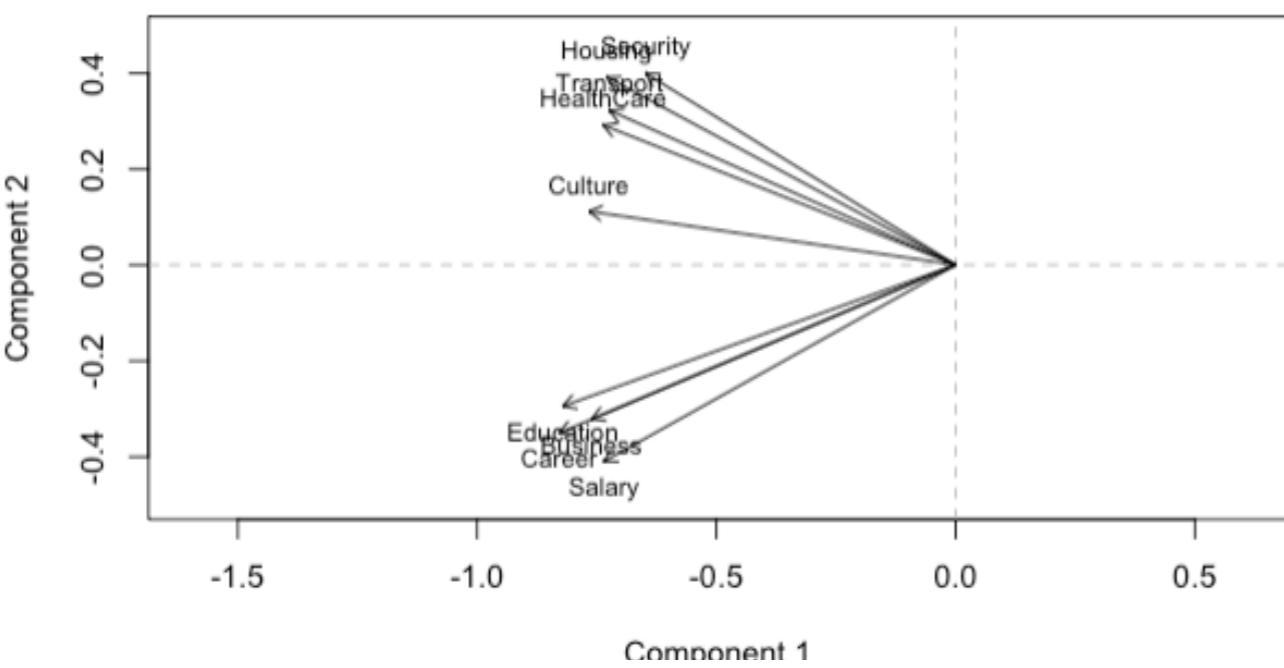
However, the subjectively perceived estimates of quality of life estimates are vulnerable to indicate the same distances between levels among the sample for different regions' characteristics. To cope with it we can assign quantitative values to quantitative scales.



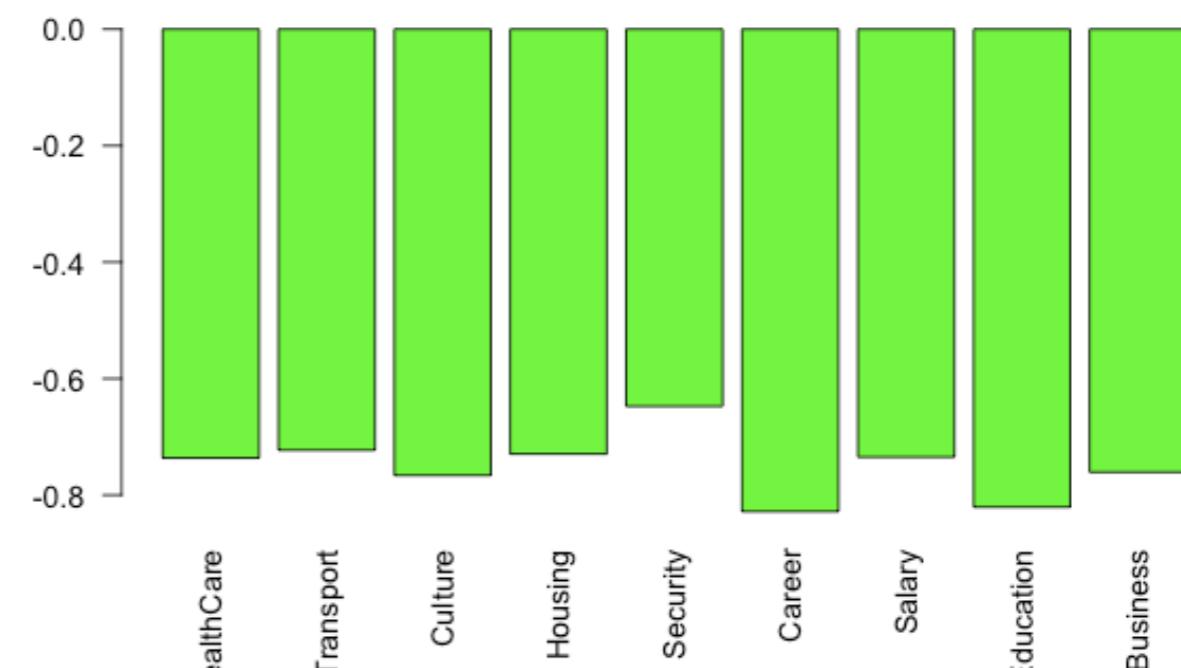
NLPCA

The component division is similar to classical PCA on the same data. The given components might be considered as ***life comfort*** (the 1st component) and ***professional & self-development*** (the 2nd component)

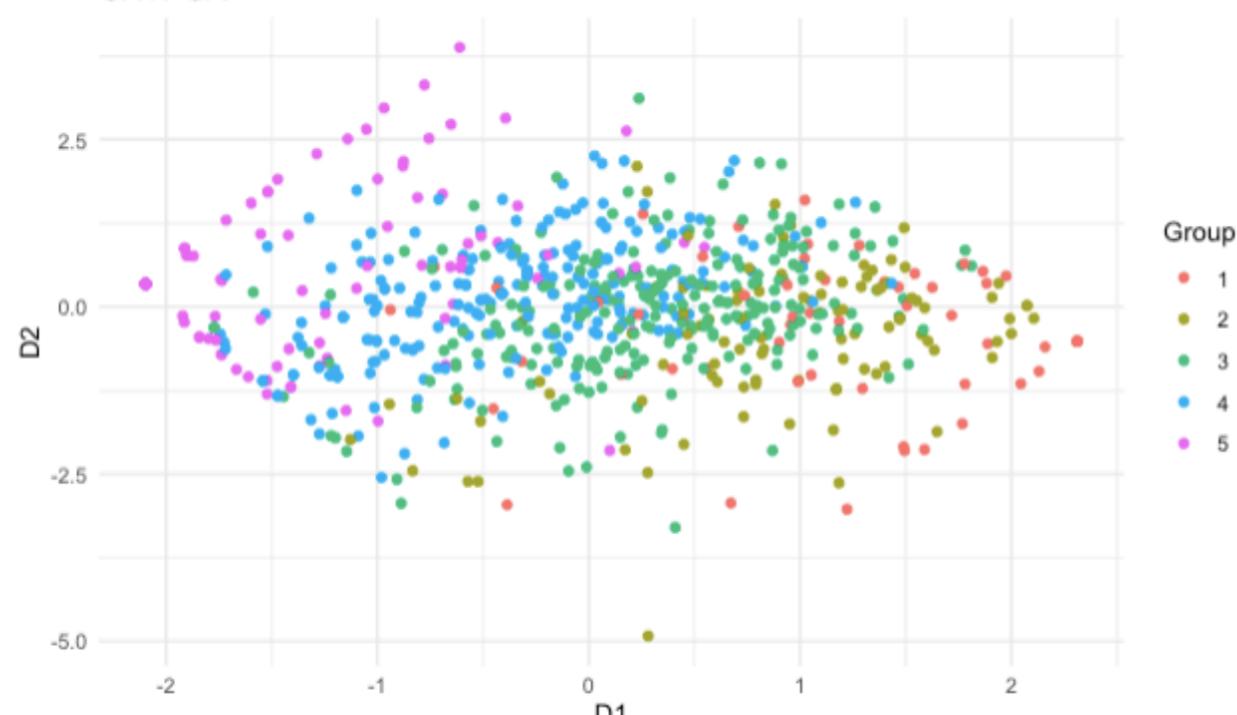
Loadings Plot



Contribution to PC1



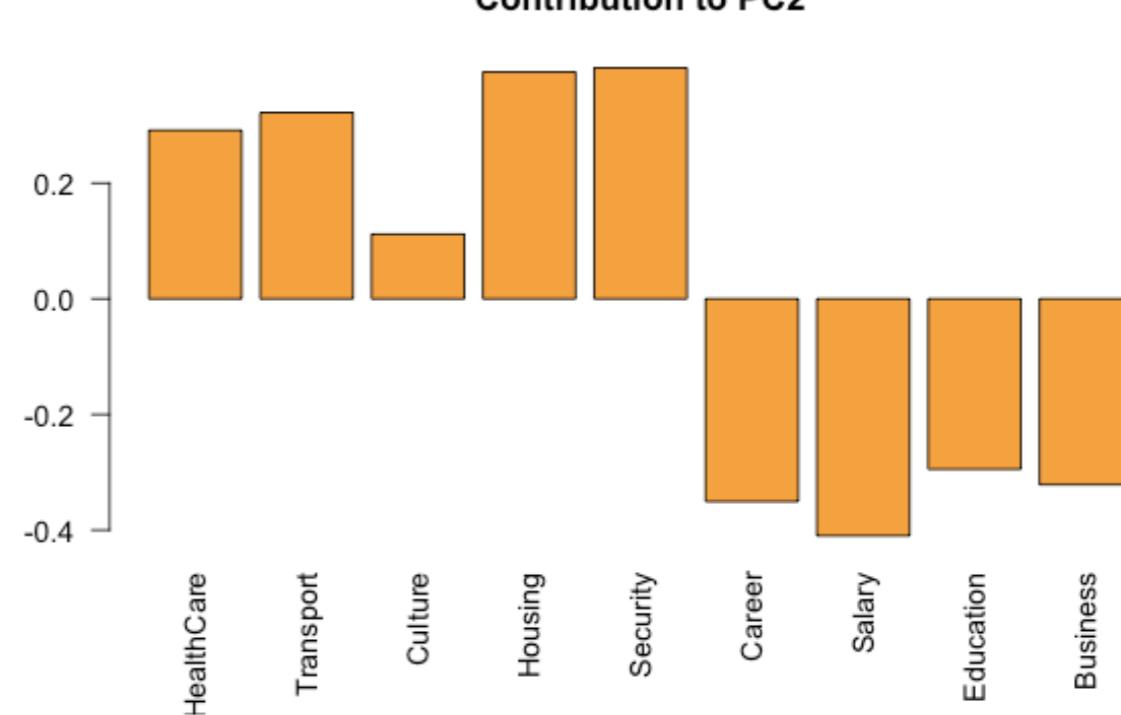
CATPCA



Group

- 1
- 2
- 3
- 4
- 5

Contribution to PC2



Data preparation for Clustering

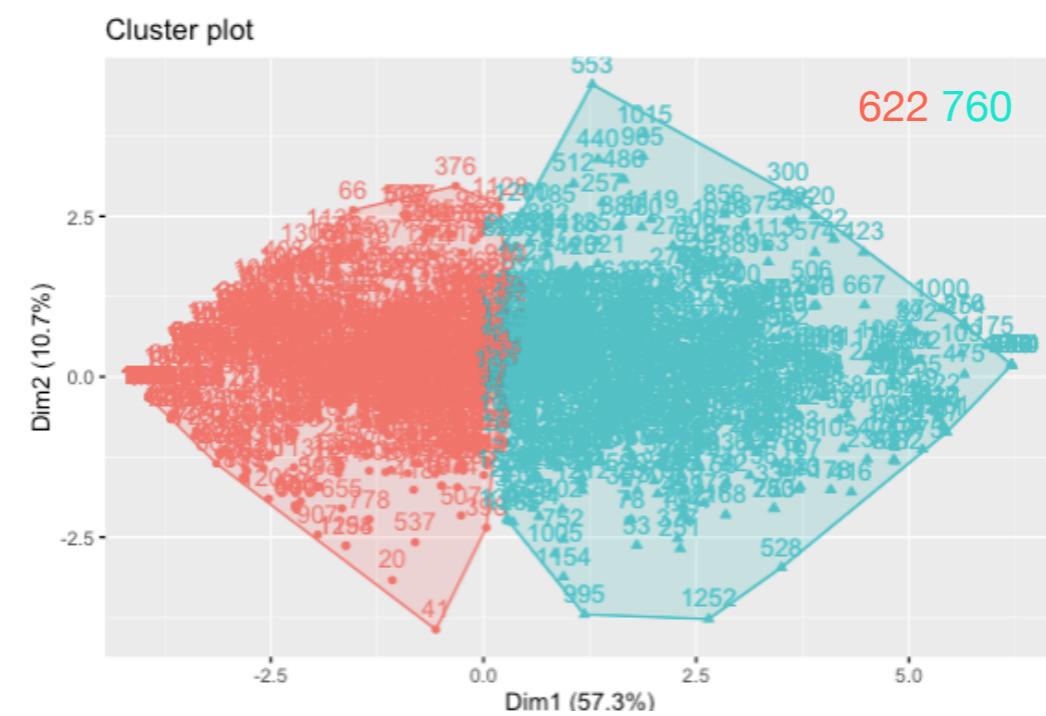
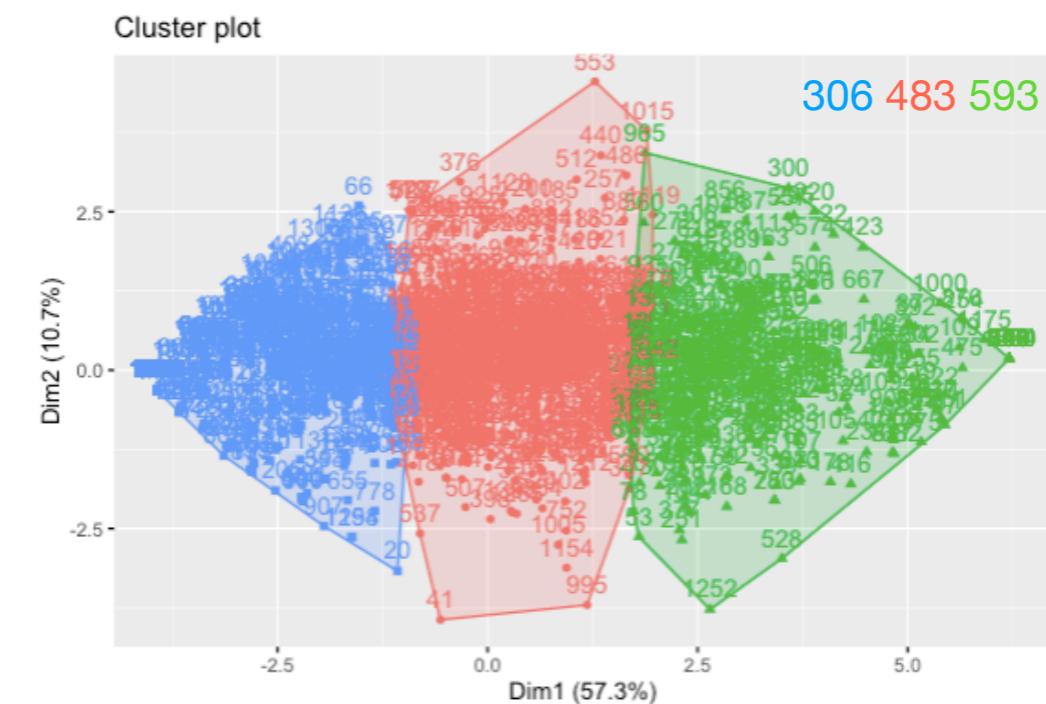
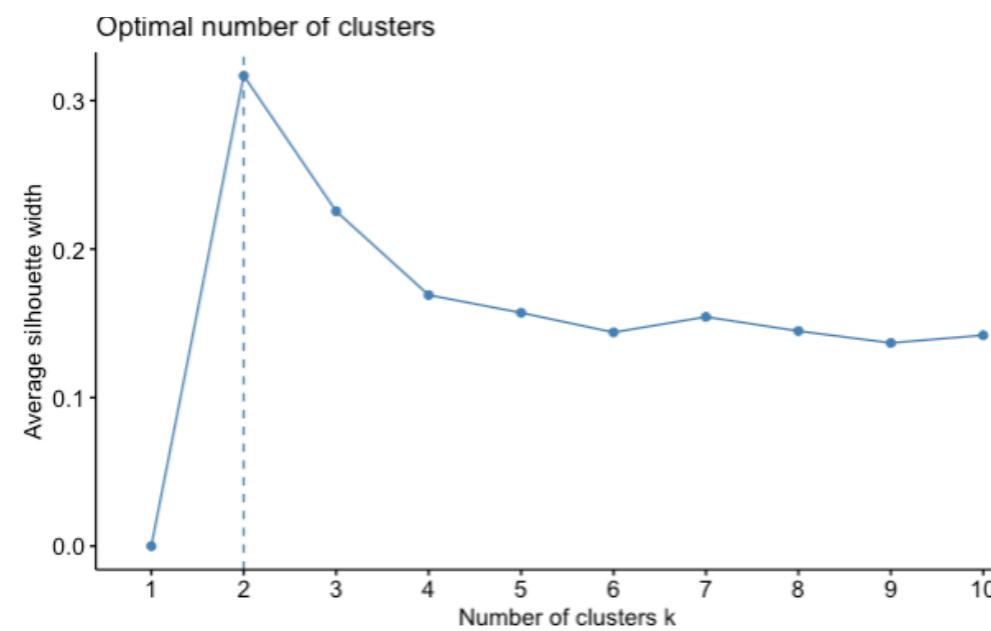
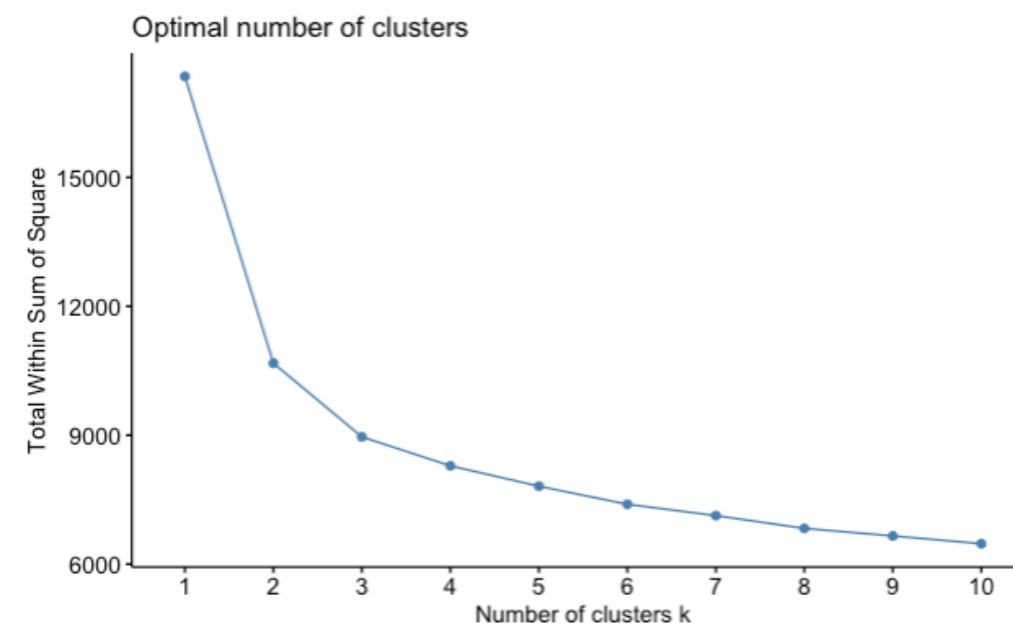
- **Sub-datasets:**
 1. initial indicators (pseudo-interval);
 2. indicators + control variables (categorical);
 3. PCA + control variables (categorical);
 4. NLPCA + control variables(categorical)
- **Dummy + scaling**

Specialisation: Economics, Social, Medicine, STEM, Natural Science
Income (recoded): poor, medium, high, rich
Gender: Female, Male



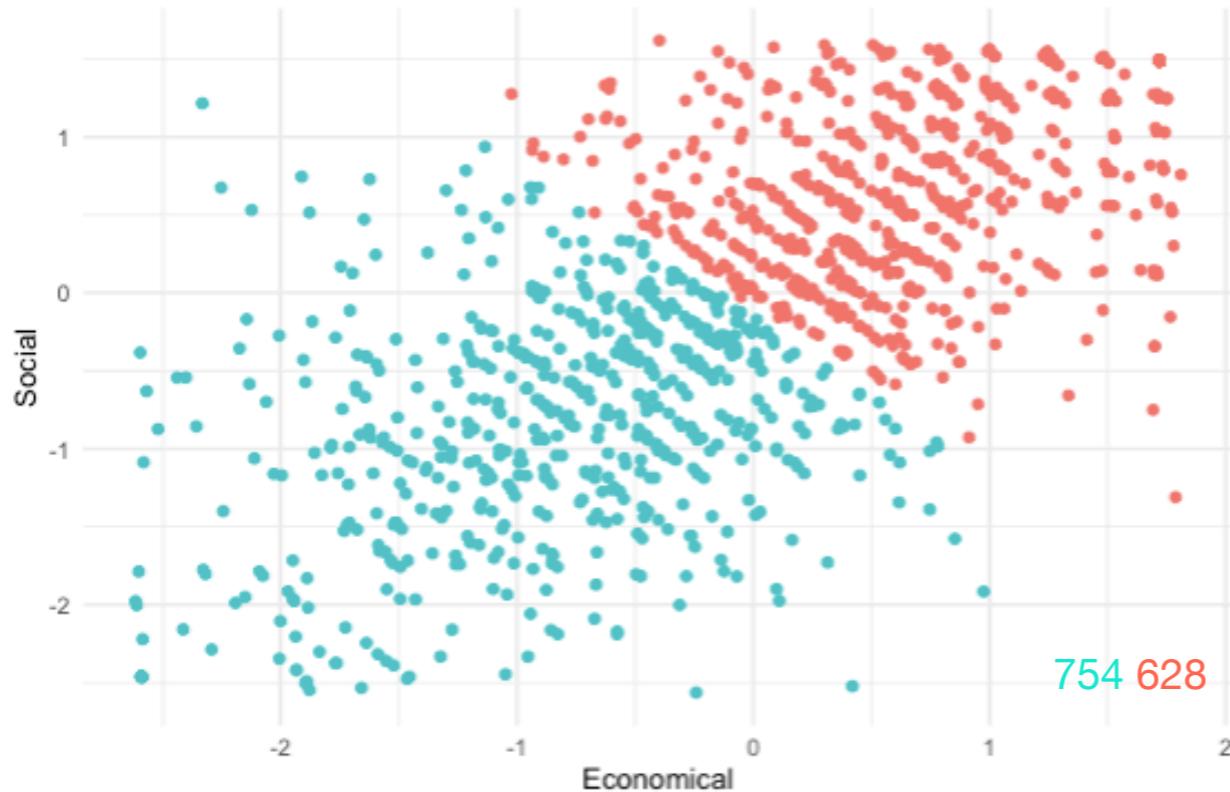
Clustering: KMeans (sat. levels) + kmmedoids

We want to discover groups of observations that are homogeneous and separated from other groups, identified by the assessment of the relative distances between points.

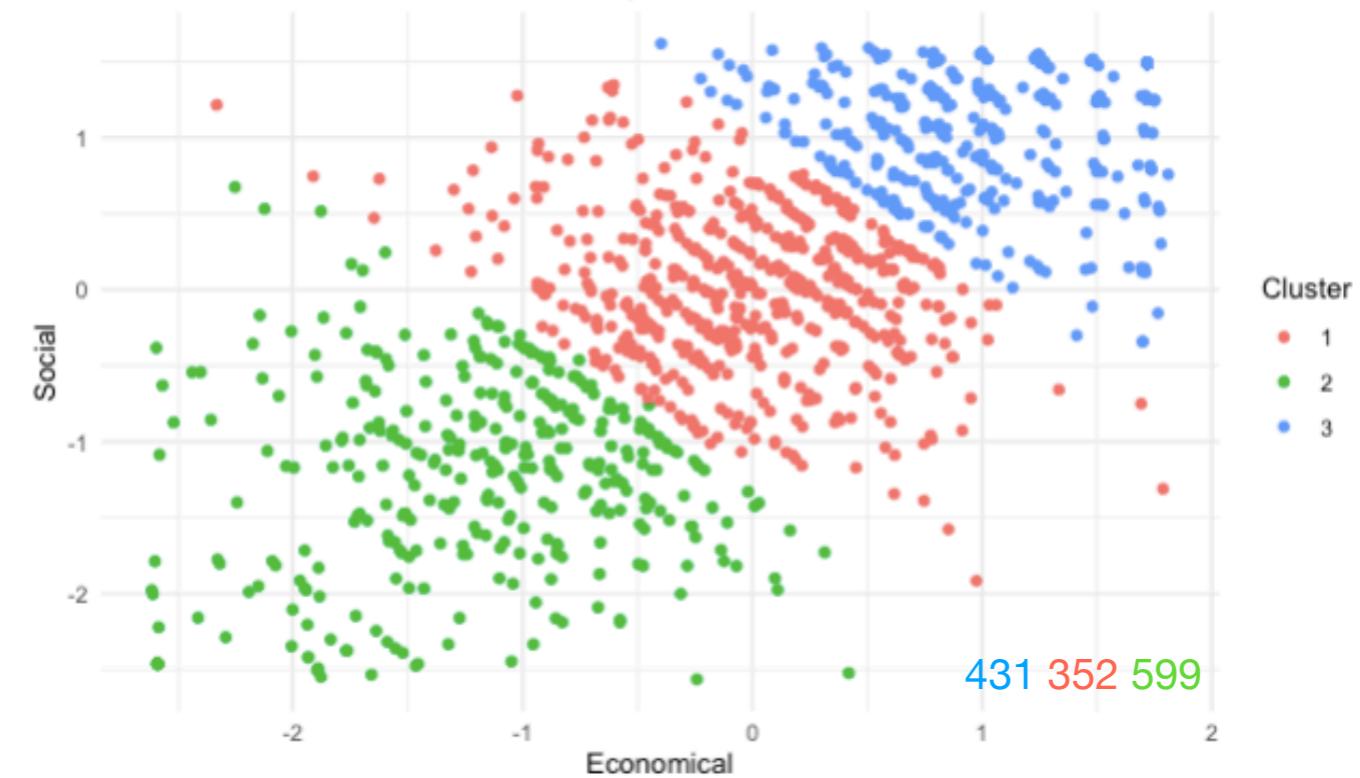


Clustering (KMeans + kmmedoids): satisfaction levels + PCA(NL)

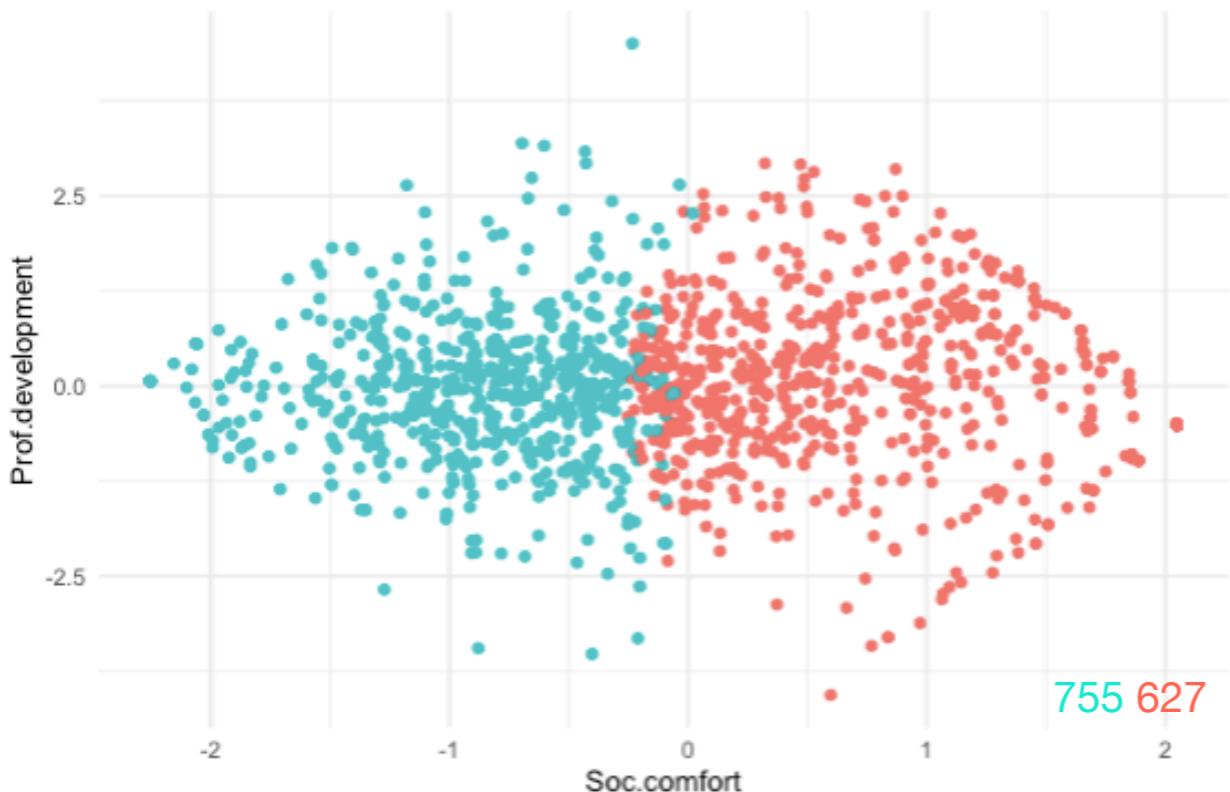
Cluster Scatter Plot with PCA Components



Cluster Scatter Plot with PCA Components



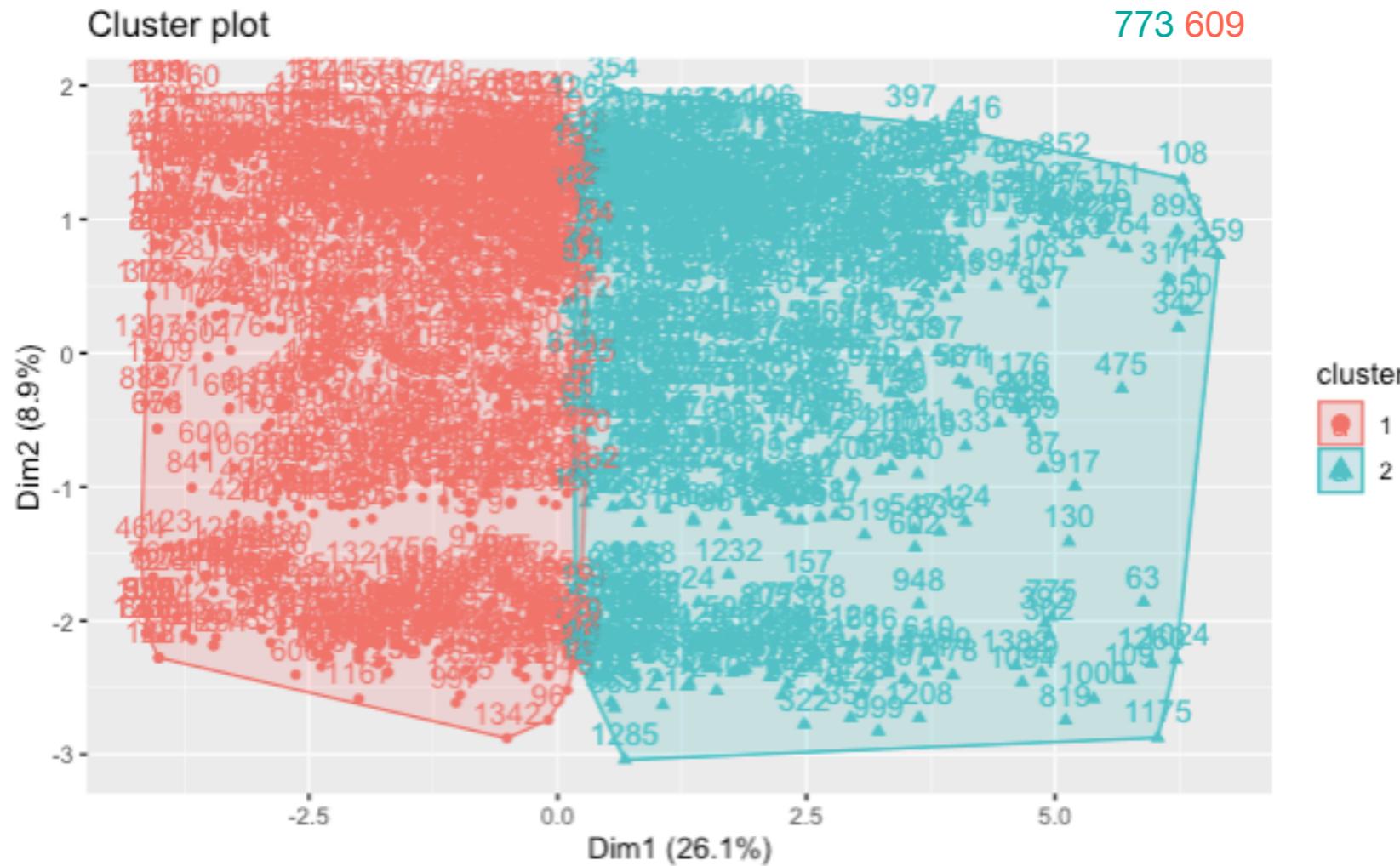
Cluster Scatter Plot with PCA Components



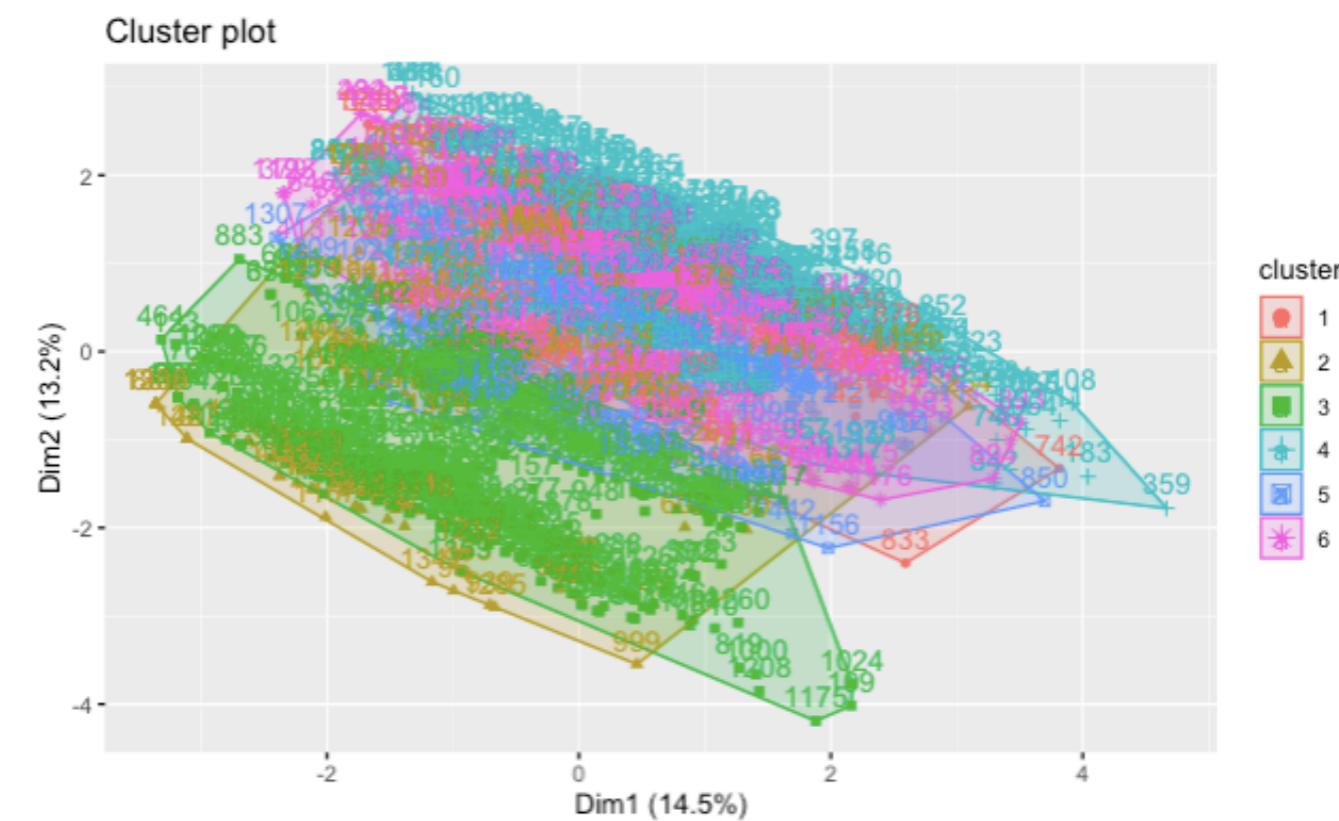
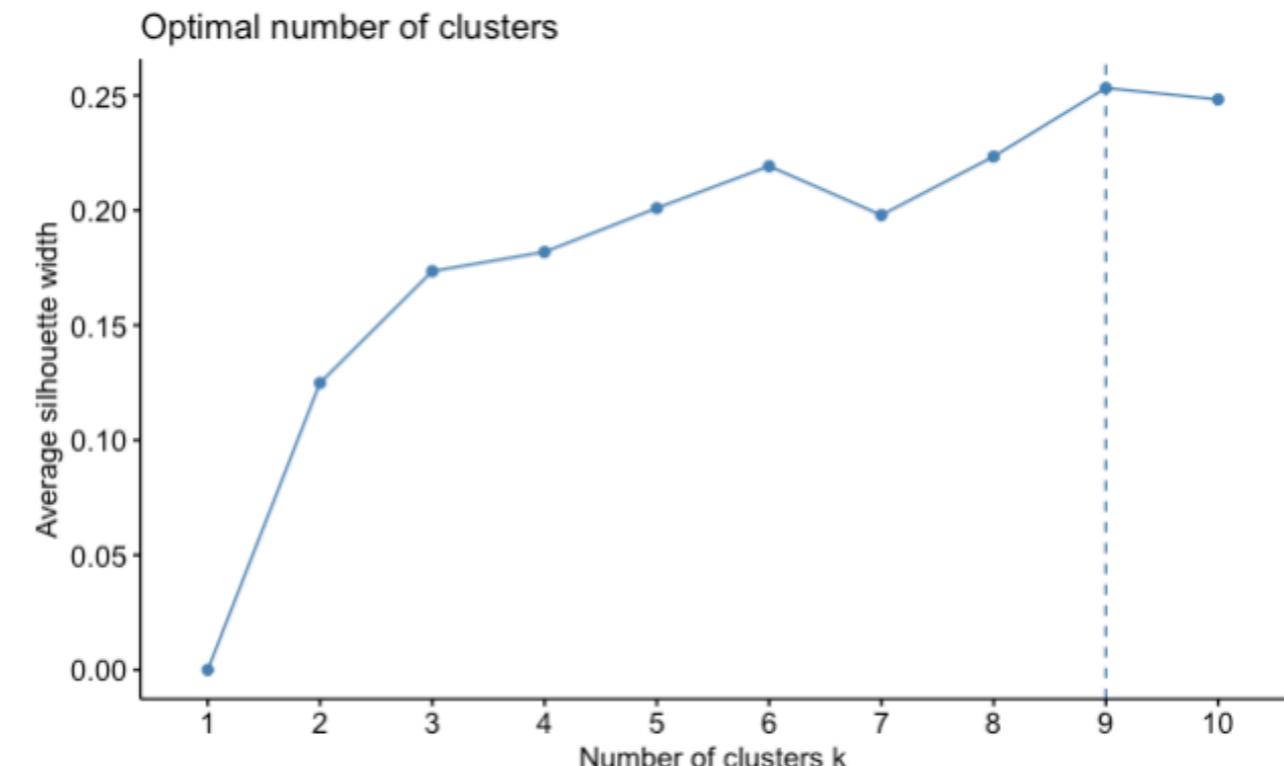
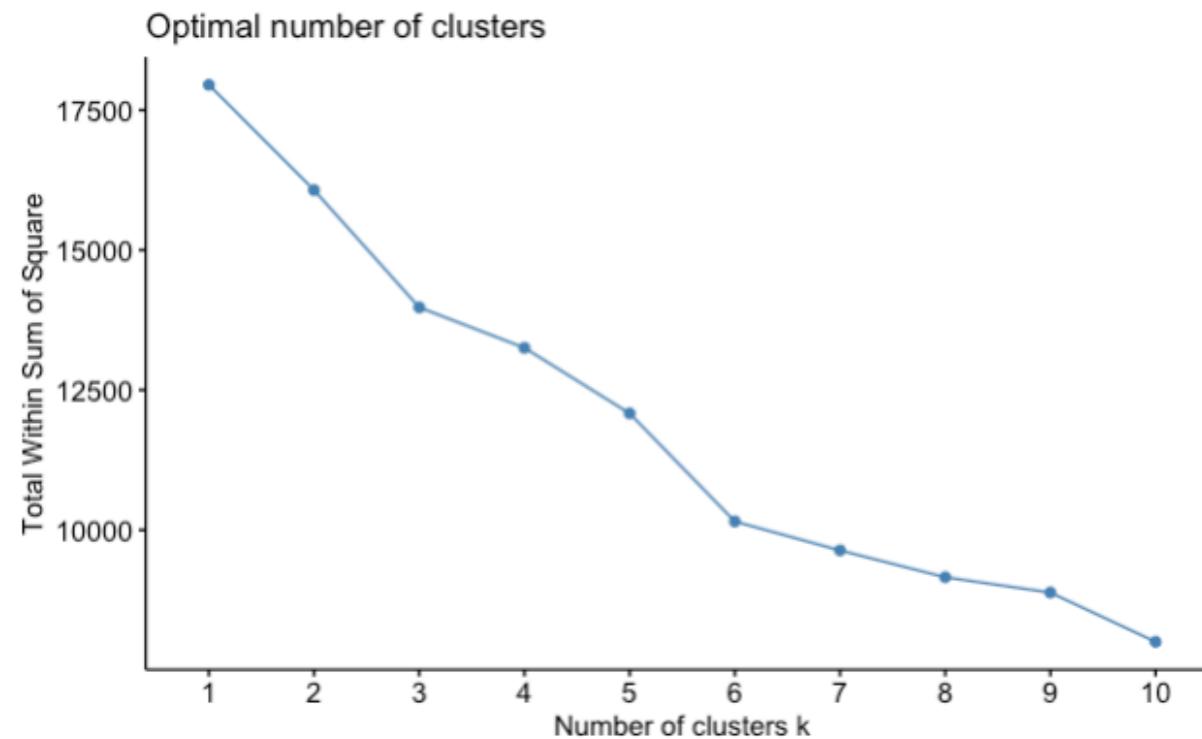
Cluster Scatter Plot with PCA Components



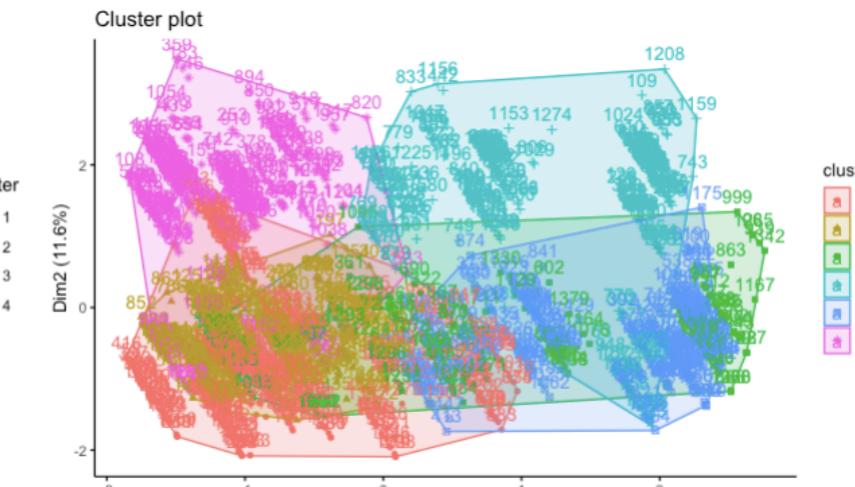
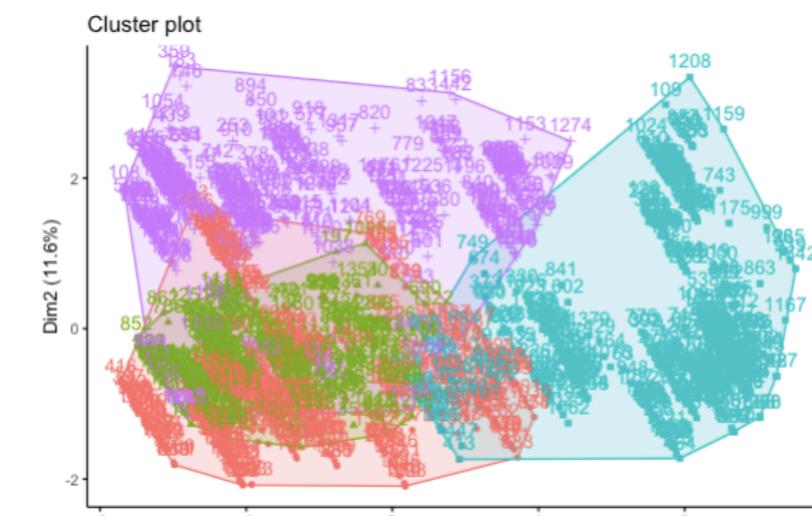
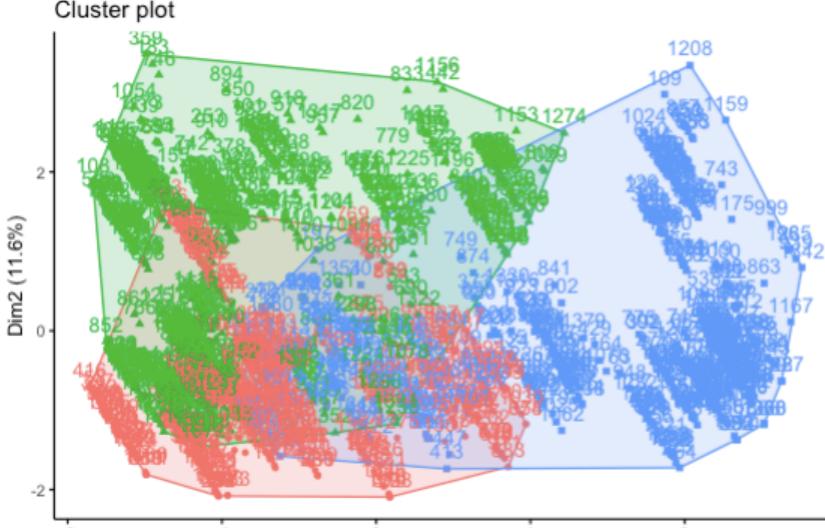
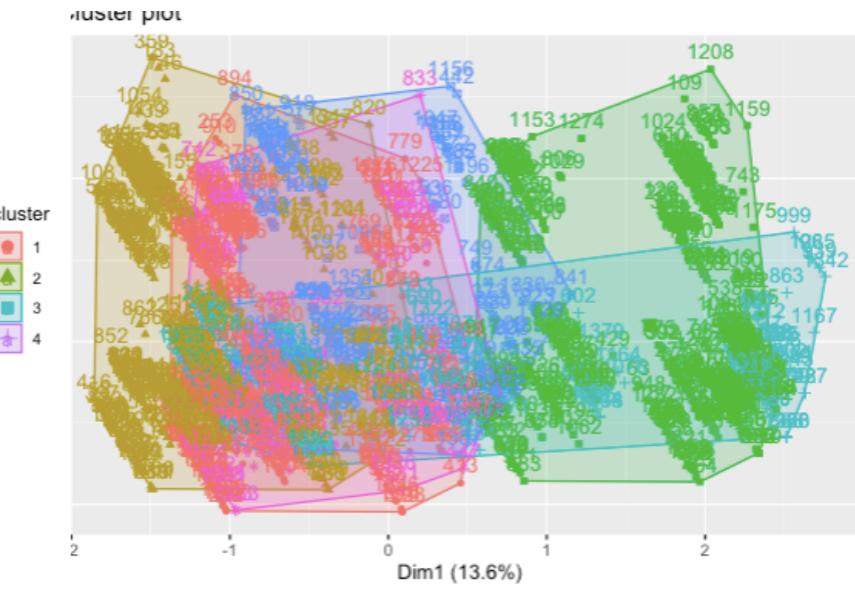
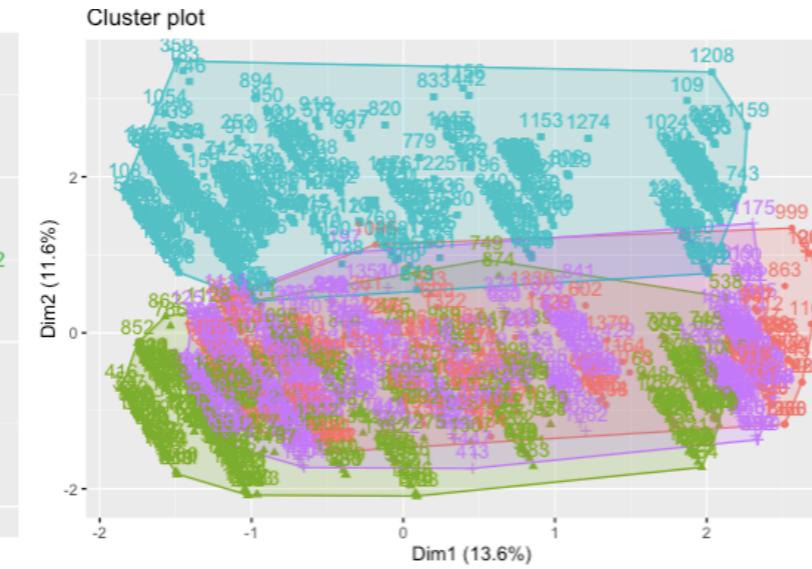
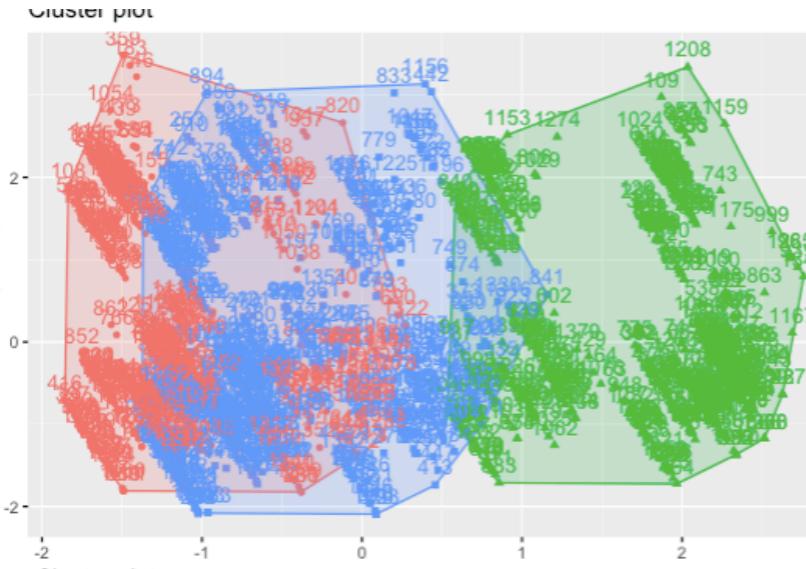
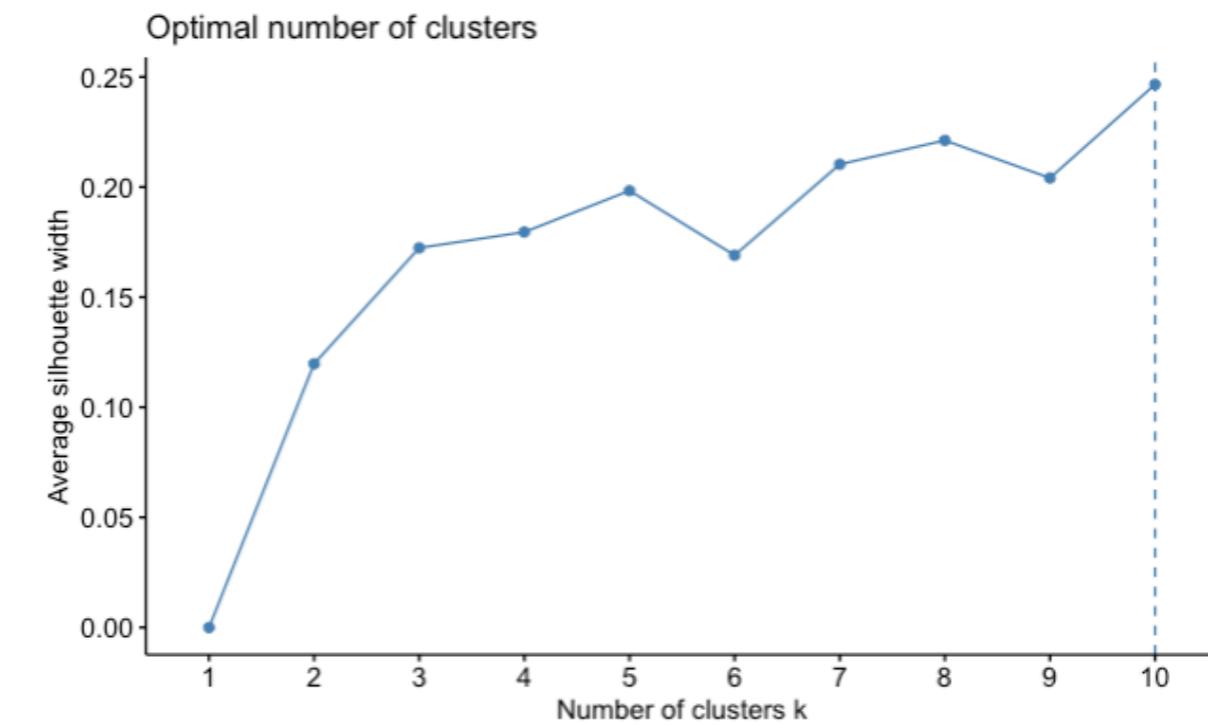
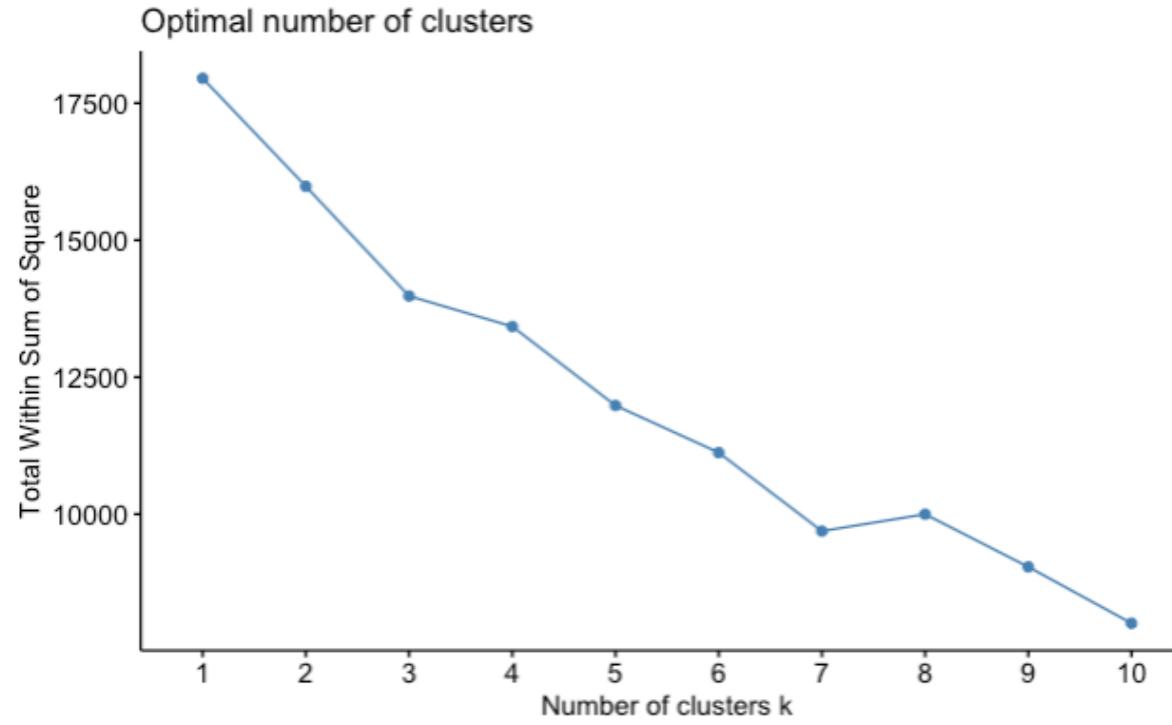
Clustering(KMeans+kmedoids): sat.levels + controls



Clustering(KMeans+kmedoids): PCA + controls



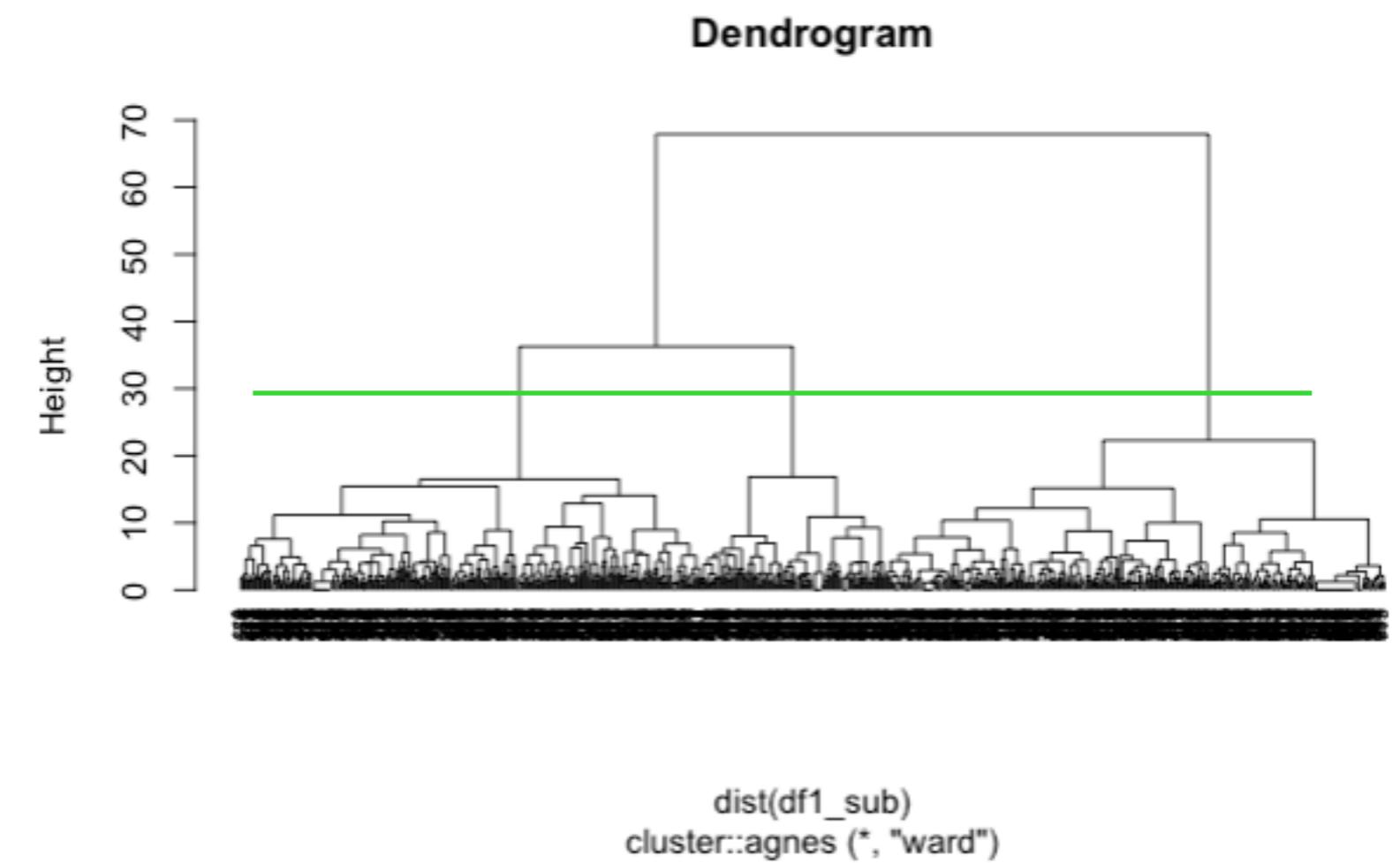
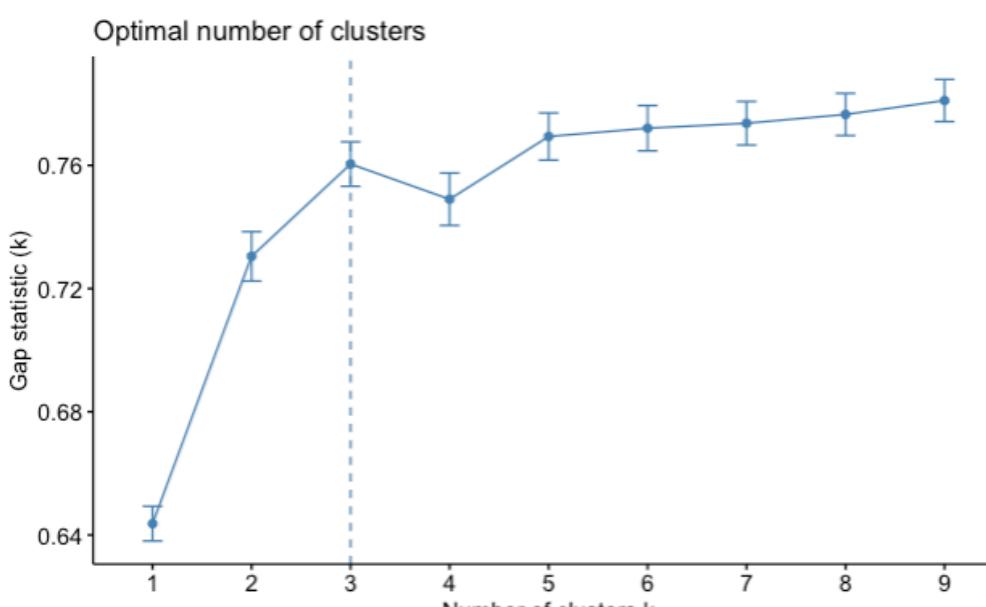
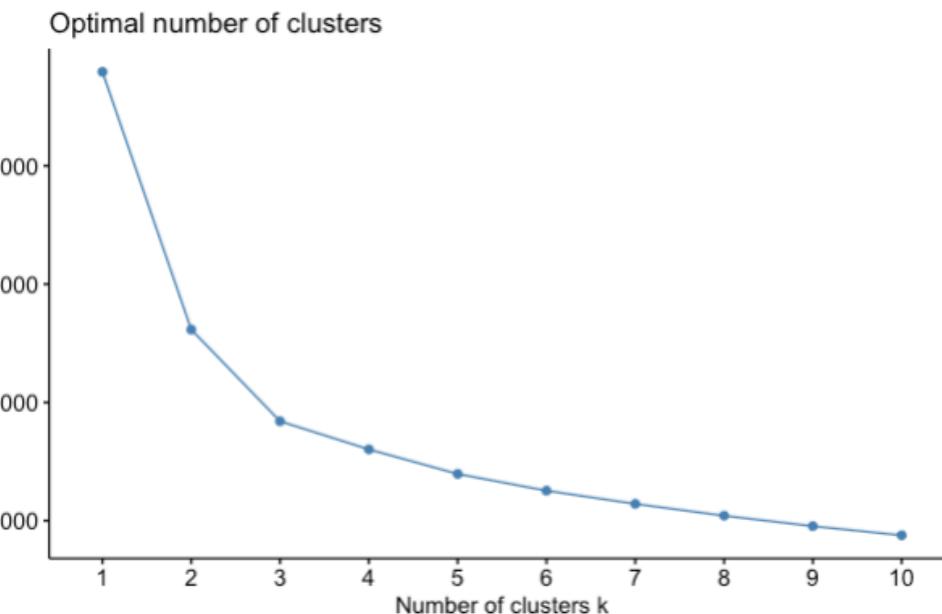
Clustering(KMeans+kmedoids): NLPCA + controls



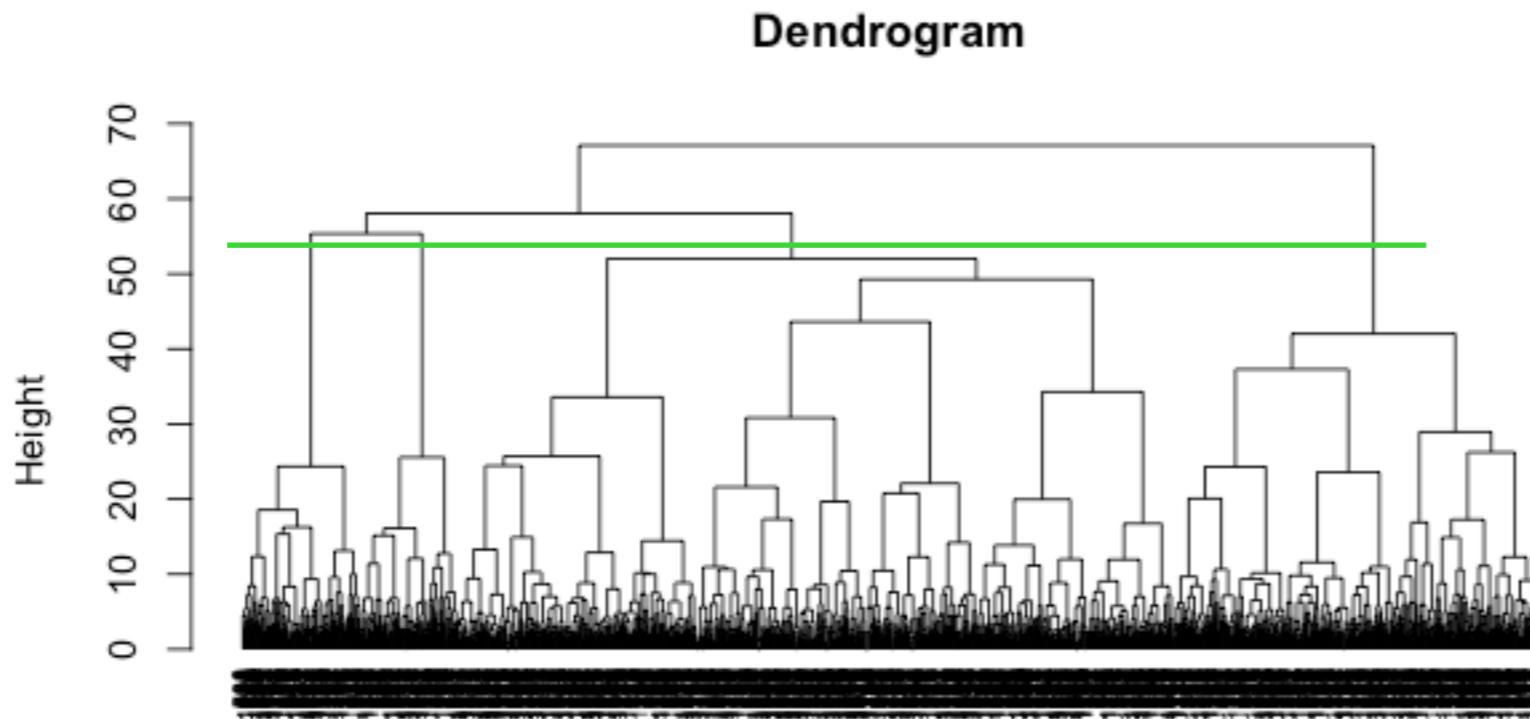
Clustering (Hierarchical): sat.levels

Agglomerative hierarchical clustering techniques produce partitions by a series of successive fusions of the n individuals into groups. It's based on distance matrix which can be calculated with Euclidean distances due to from the raw data may not be sensible when the variables are on very different scales.

The agglomerative coefficient for each clustering linkage method has been calculated, as a result, Ward's method better suits to data.



Clustering (Hierarchical): all data



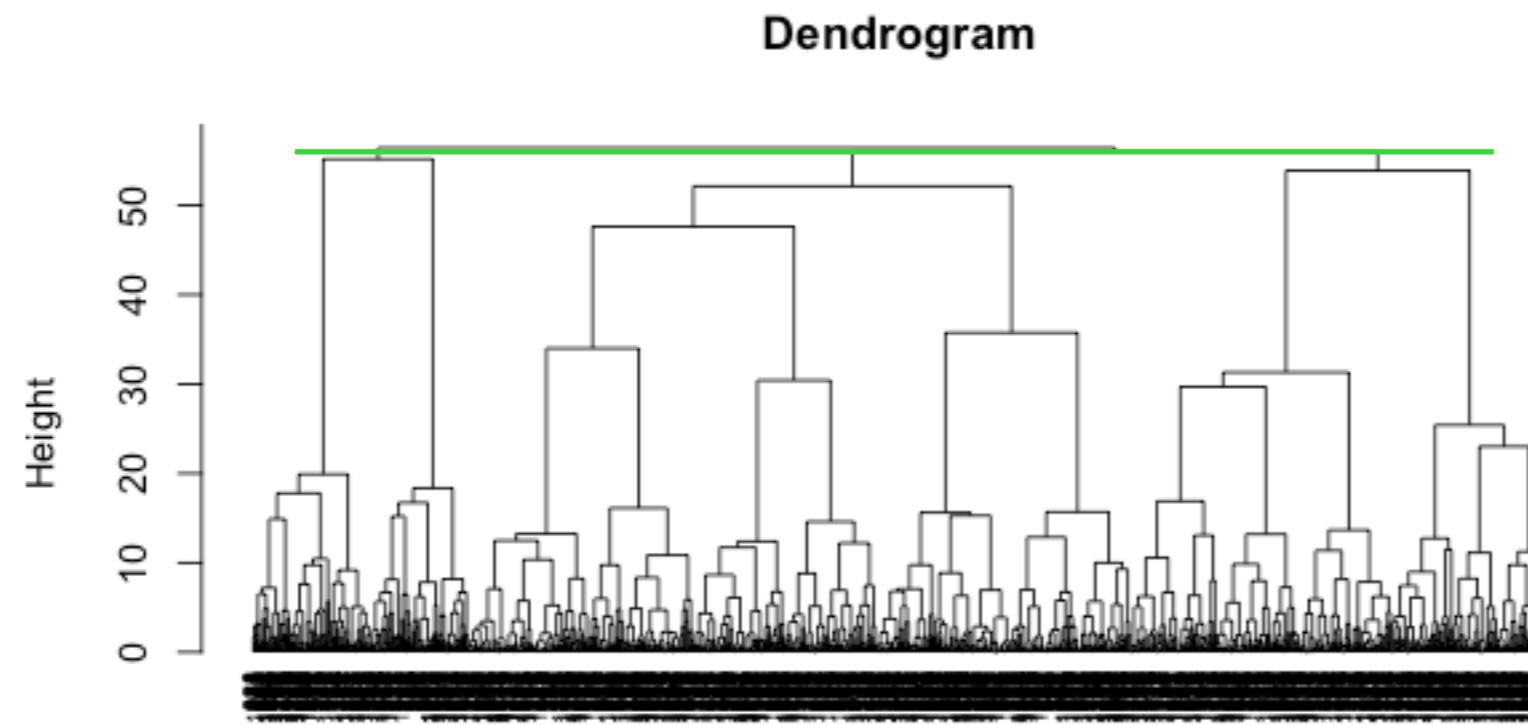
1st cluster: relocated unsatisfied with region of study male graduates from Natural Science with poor family's income

2nd cluster: relocated satisfied with region of study female graduates from Economics or STEM with medium family's income

3d cluster: stayed completely unsatisfied with region of study female graduates from Social Science with high family's income

4th cluster: stayed somehow satisfied with region of study male graduates from Medicine with high family's income

Clustering (Hierarchical): PCA + controls

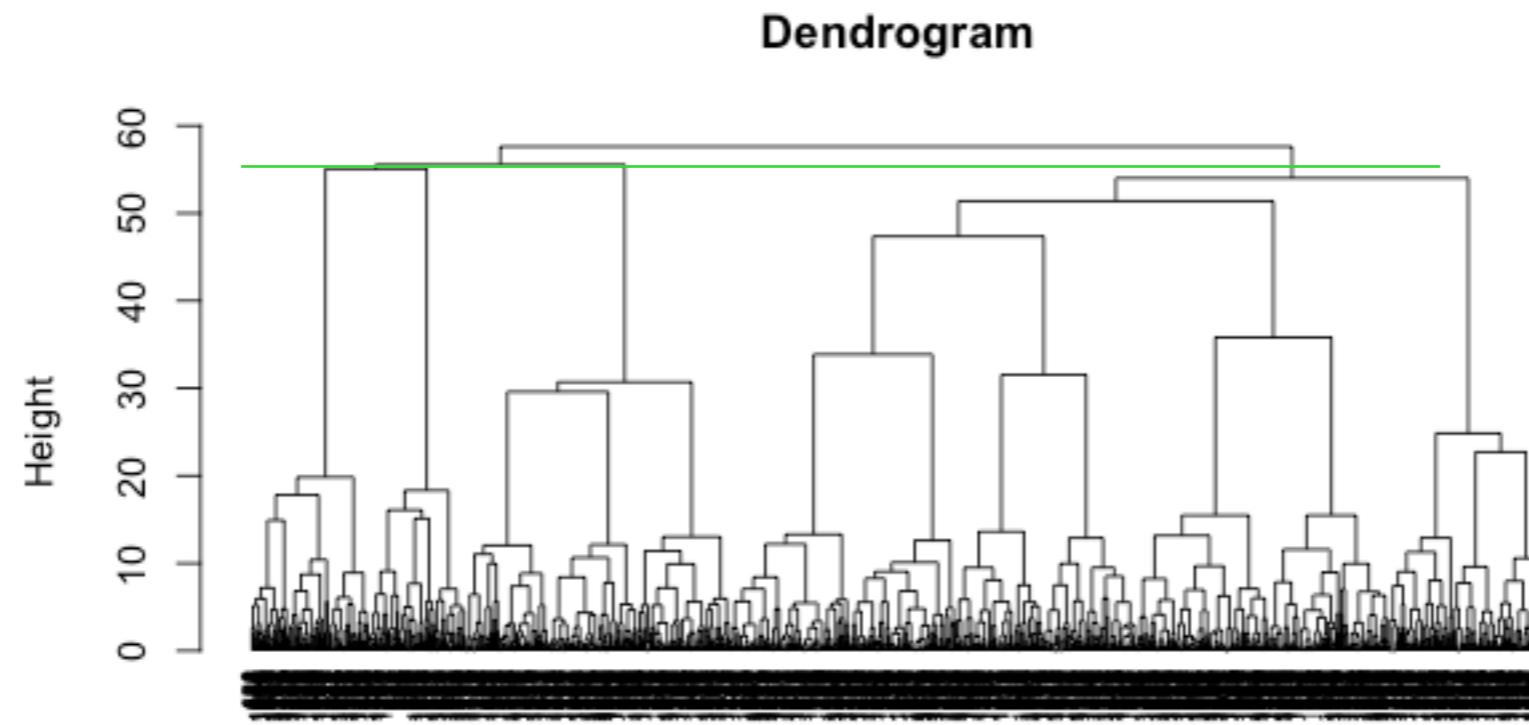


1st cluster: less satisfied with social factor and more satisfied with economical factor
male graduates from Natural or Medicine studies

2nd cluster: not really satisfied with social and economical factors male graduates from
Economics or STEM with poor or medium family's income

3d cluster: completely unsatisfied with region of study female graduates from Social
Science with high or rich family's income

Clustering (Hierarchical): NLPCA + controls

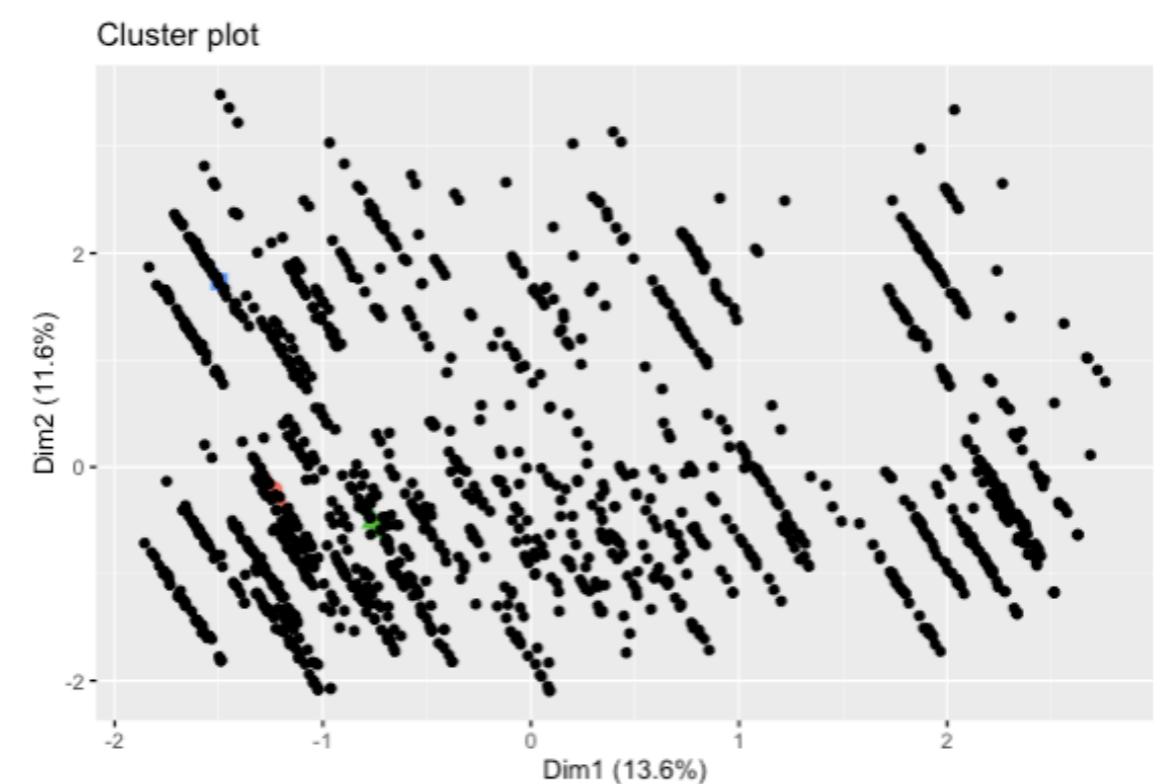
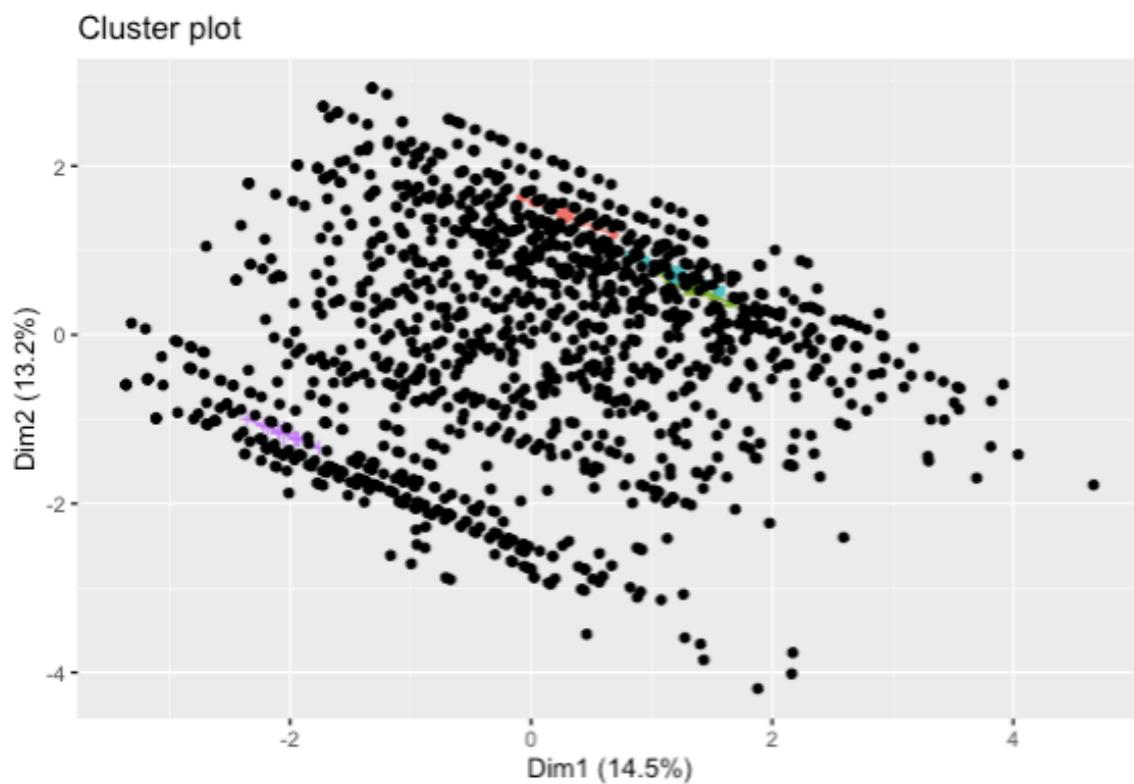
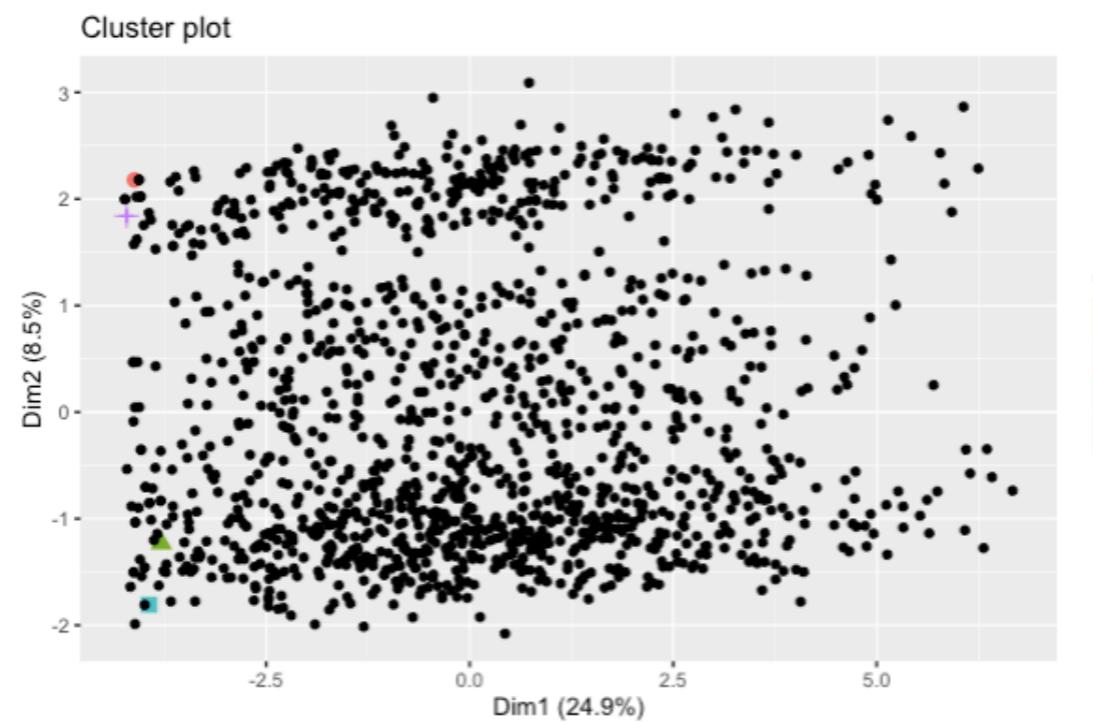


1st cluster: not really satisfied with life comfort and self-development factors graduates from Natural or Medicine studies with high family's income

2nd cluster: somehow satisfied with life comfort and self-development factors male graduates from Economics or STEM with rich family's income

3d cluster: less satisfied with life comfort factor and more satisfied with self-development factor male graduates from Social studies with poor family's income

Clustering: DBSCAN



Clustering: model-based

The approach to clustering is based on probability distribution. The best model is EEE (ellipsoidal, equal volume, shape, and orientation) with 7 components

