

Факультет экономических наук

ПРОЕКТ НА ТЕМУ: «ИССЛЕДОВАНИЕ ДЕТЕРМИНАНТ УРОВНЯ ВВП НА ДУШУ НАСЕЛЕНИЯ»



Подготовили:

Джумаева Ш.Х.,

Золотарев А.О.,

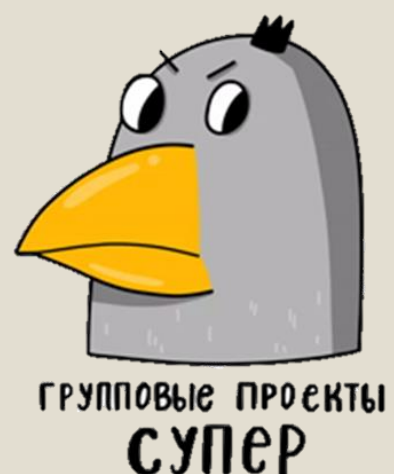
Кабанова Е.Ю.,

Морина Е.К.,

Тислюк А.А.

Хисяметдинова Р.Р.

Москва, 2020





ОБЪЯСНЯЮЩИЕ ПЕРЕМЕННЫЕ

Факультет экономических наук

Обозначение

Название

Export

Экспорт товаров и услуг (в % от ВВП)

Consumption

Расходы на конечное потребление (в % от ВВП)

Labor among youth

Коэффициент участия рабочей силе в возрасте 15-24 лет (%)

Employment in services

Доля занятых в сфере услуг (%)

Employment in industry

Доля занятых в промышленном секторе (%)

Mortality rate under 5

Уровень смертности детей в возрасте до 5 лет (%)

Population in largest city

Население крупнейшего города (в % от городского населения страны)

University enrollment

Доля вовлеченности в профессиональное образование (%)

Unemployment

Общий уровень безработицы (%)

Urban population

Городское население (% от общего населения)



ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Факультет экономических наук

Переменные	Среднее	Медиана	Стандартное отклонение	Асимметрия	Эксцесс
Экспорт	44,27	36,03	33,53	2,42	7,43
Потребление	78,72	78,07	15,55	0,64	3,58
Участие молодежи в рабочей силе	43,65	42,15	13,68	0,41	-0,55
Занятость в секторе услуг	55,94	57,61	17,73	-0,41	-0,53
Занятость в секторе промышленности	20,16	19,77	8,23	0,40	1,10
Доля населения в крупнейшем городе от всего городского населения	34,26	31,95	18,42	1,03	1,37
Смертность среди детей до 5 лет	27,98	14,30	30,53	1,30	0,64
Общий уровень безработицы	6,66	5,10	5,06	1,62	2,80
Коэффициент вовлеченности в профессиональное образование	43,67	43,69	29,20	0,18	-1,36
Доля городского населения от всего населения страны	62,62	64,52	21,93	-0,32	-0,76

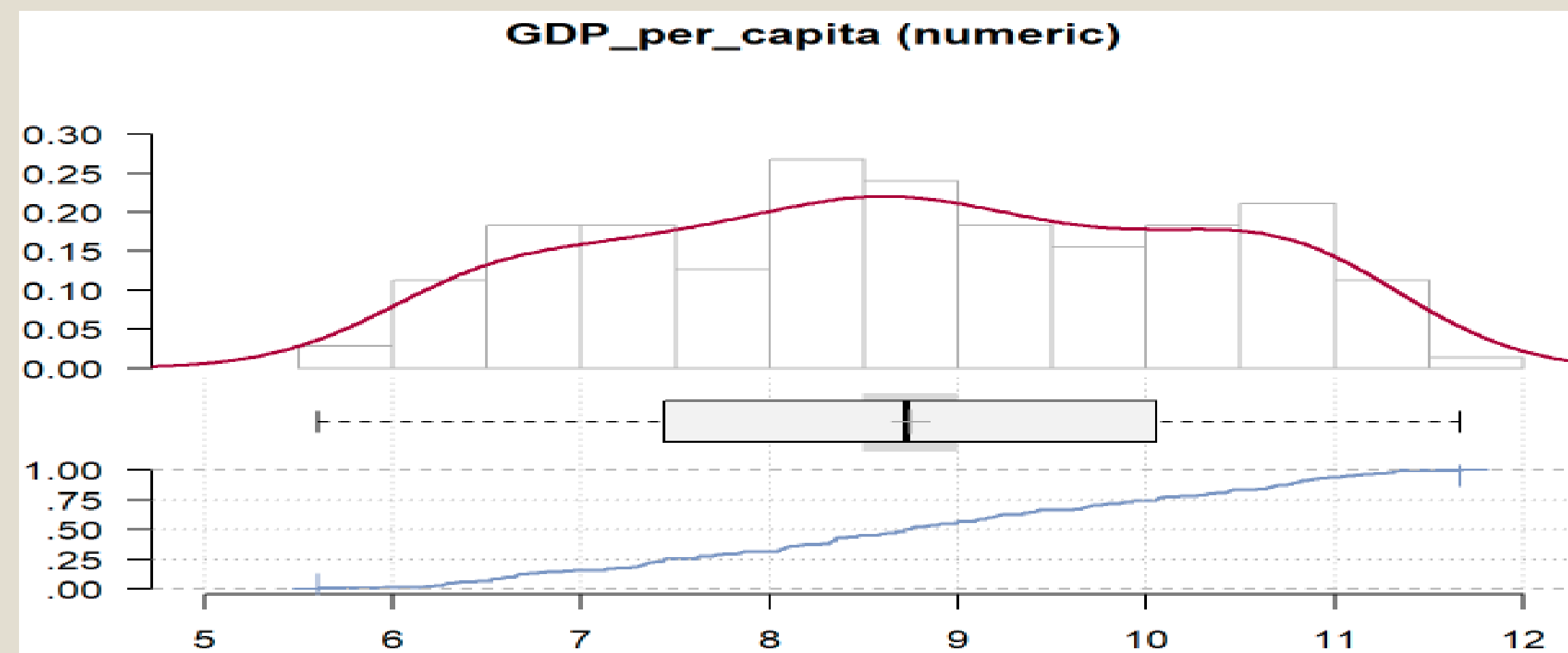


РИС. 1, ВИЗУАЛИЗАЦИЯ ДАННЫХ
ПОЛУЧЕННАЯ ИСПОЛЬЗОВАНИЕМ
ФУНКЦИИ DESC

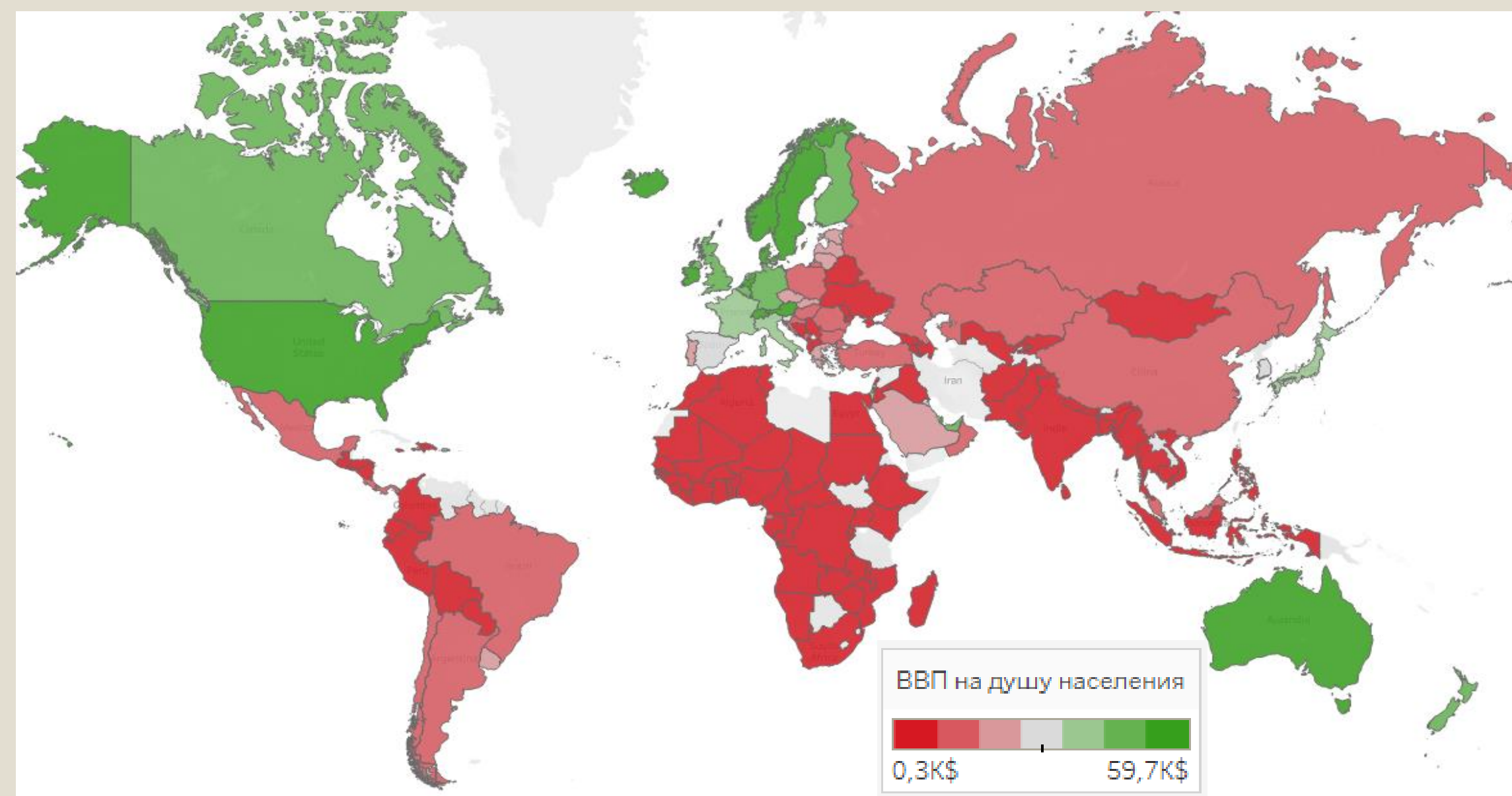


РИС. 2. ВИЗУАЛИЗАЦИЯ УРОВНЯ ВВП НА ДУШУ
НАСЕЛЕНИЯ

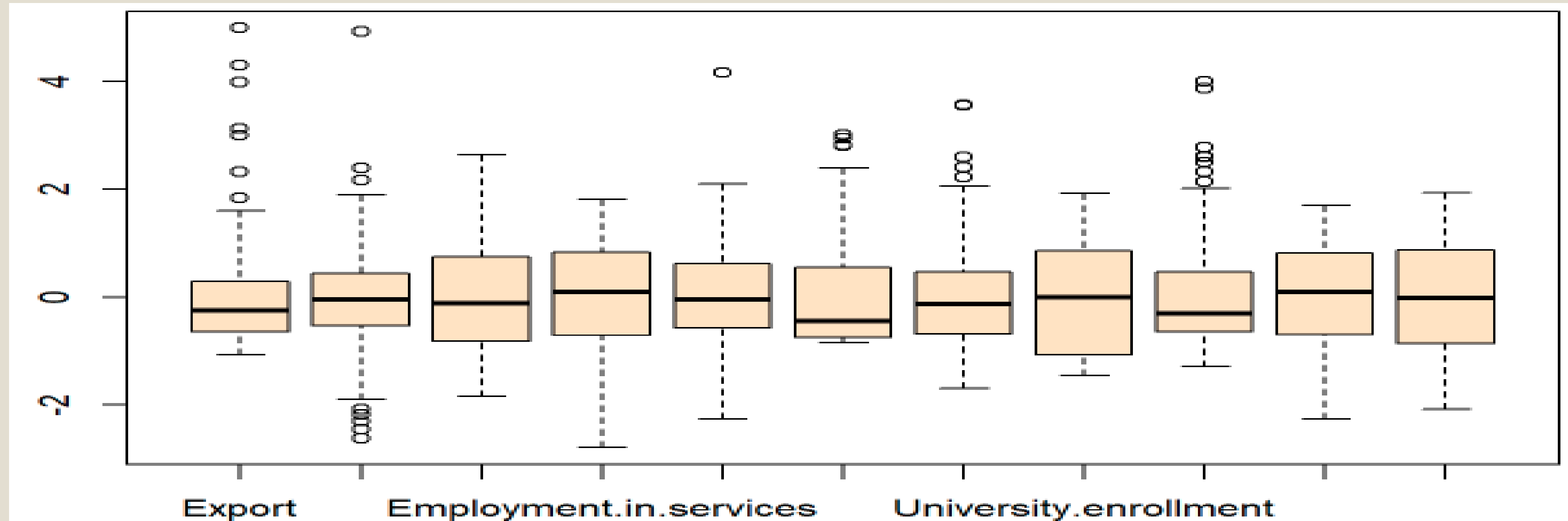
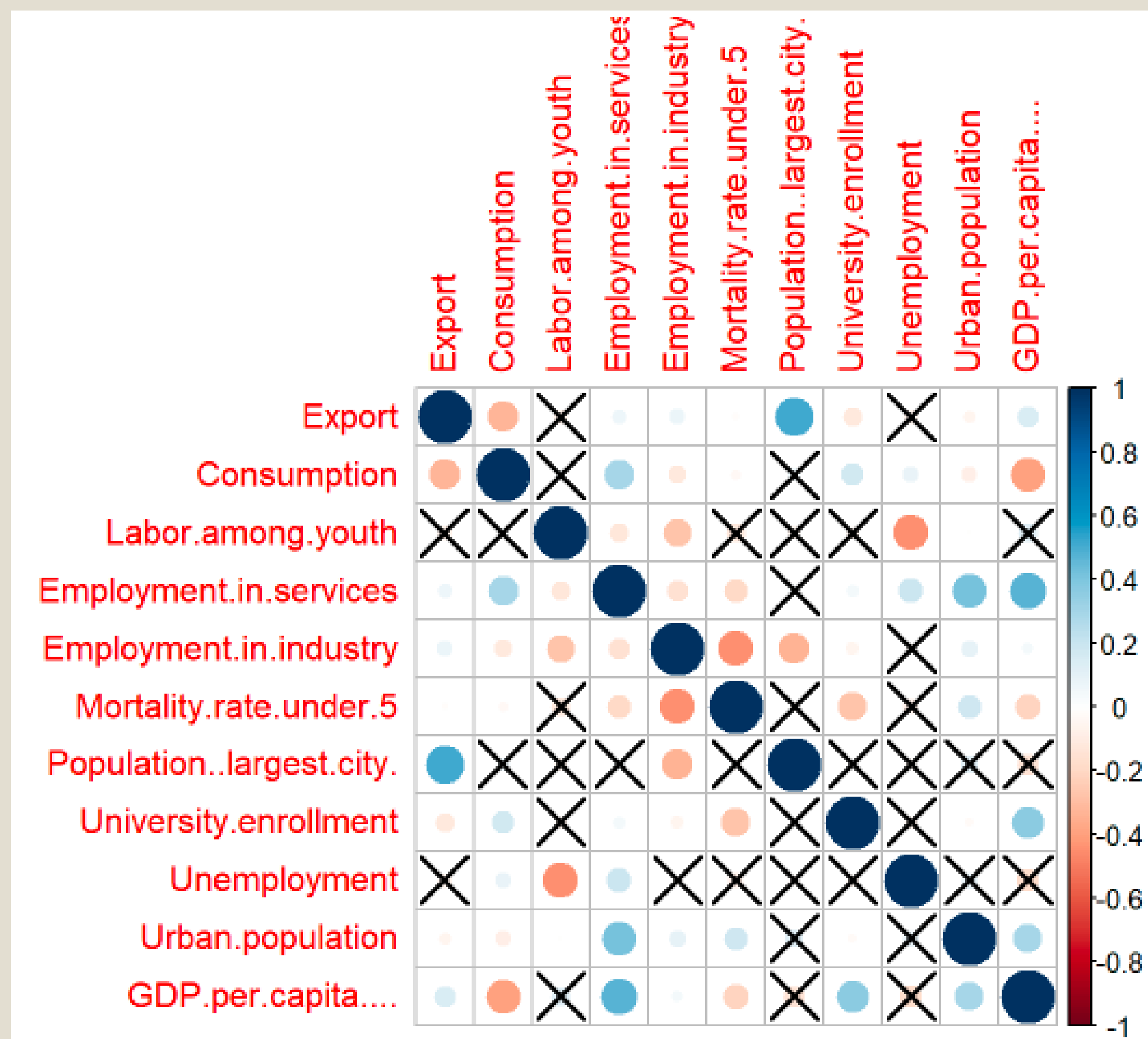


РИС.3. ЯЩИЧКОВАЯ ДИАГРАММА ДЛЯ ОБЪЯСНЯЮЩИХ И ЦЕЛЕВОЙ ПЕРЕМЕННЫХ

Логарифмирование целевой переменной дало полное отсутствие выбросов в ВВП на душу населения



Множественный коэффициент корреляции для нашей целевой переменной составляет 0,9406

РИС. 4. МАТРИЦА ЧАСТНЫХ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ



РЕГРЕССИОННЫЙ АНАЛИЗ. ЛИНЕЙНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.014703   0.541878  12.945  < 2e-16 ***
Export         0.003160   0.001868   1.692  0.093056 .
Consumption   -0.018081   0.003621  -4.993  1.86e-06 ***
Labor.among.youth  0.005254   0.003886   1.352  0.178615
Employment.in.services  0.034249   0.005529   6.195  7.01e-09 ***
Employment.in.industry  0.003753   0.007577   0.495  0.621204
Mortality.rate.under.5 -0.006573   0.002598  -2.530  0.012600 *
Population..largest.city -0.005214   0.003023  -1.724  0.086985 .
University.enrollment  0.010667   0.002395   4.453  1.79e-05 ***
Unemployment   -0.021707   0.010573  -2.053  0.042048 *
Urban.population  0.013359   0.003779   3.536  0.000563 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5184 on 131 degrees of freedom
Multiple R-squared:  0.8911, Adjusted R-squared:  0.8828

```

РИС. 5. КОЭФФИЦИЕНТЫ ЛИНЕЙНОЙ РЕГРЕССИОННОЙ МОДЕЛИ

Наибольшее влияние на логарифм ВВП на душу населения оказывают такие переменные как занятость в сфере услуг, уровень безработицы и уровень потребления.

Модель имеет сильную предсказательную силу (велика доля объясненной дисперсии)



НЕЛИНЕЙНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	8.504260	1.071744	7.935	8.24e-13	***
Export	-0.026541	0.083091	-0.319	0.749916	
Consumption	-0.018930	0.003111	-6.084	1.20e-08	***
Labor.among.youth	-0.029703	0.144464	-0.206	0.837414	
Employment.in.services	0.030321	0.004898	6.190	7.16e-09	***
Employment.in.industry	-0.240196	0.100396	-2.392	0.018154	*
Mortality.rate.under.5	-0.564793	0.068435	-8.253	1.44e-13	***
Population..largest.city.	-0.096568	0.073703	-1.310	0.192413	
University.enrollment	0.032913	0.073421	0.448	0.654700	
Unemployment	-0.211009	0.060503	-3.488	0.000664	***
Urban.population	0.742579	0.159748	4.648	8.05e-06	***

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' ' 1
Residual standard error:	0.4502 on 131 degrees of freedom				
Multiple R-squared:	0.9179, Adjusted R-squared: 0.9116				

Наиболее высокую предсказательную силу имеет модель, в которой прологарифмированы все 10 объясняющих переменных за исключением потребления и занятости в сфере услуг

РИС. 6. РЕЗУЛЬТАТЫ ВЫЧИСЛЕНИЯ ОПТИМАЛЬНОЙ НЕЛИНЕЙНОЙ РЕГРЕССИОННОЙ МОДЕЛИ

ВЫДЕЛЕНИЕ ГЛАВНЫХ КОМПОНЕНТ

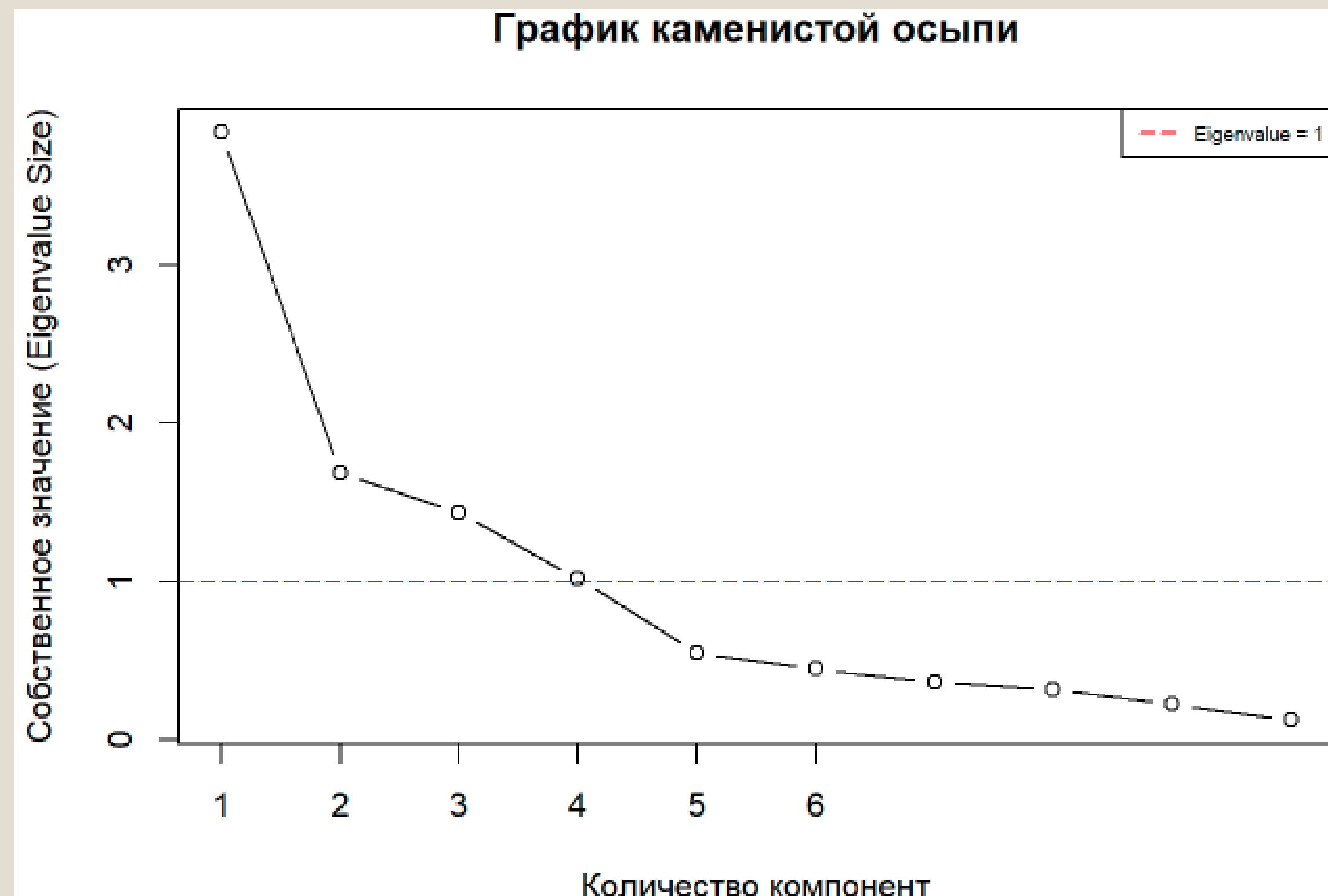


РИС. 7. ГРАФИК КАМЕНИСТОЙ ОСЫПИ ДЛЯ ОПРЕДЕЛЕНИЯ ОПТИМАЛЬНОГО КОЛИЧЕСТВА КОМПОНЕНТ

	PC1	PC2	PC3	PC4
Export	0.53	0.59	0.33	-0.28
Consumption	-0.51	-0.44	0.17	0.54
Labor.among.youth	-0.26	0.52	-0.55	0.4
Employment.in.services	0.88	-0.05	0.11	0.3
Employment.in.industry	0.61	-0.3	-0.36	-0.45
Mortality.rate.under.5	-0.84	0.1	0.23	-0.1
Population..largest.city.	0.04	0.58	0.68	0.14
University.enrollment	0.79	-0.06	-0.16	0.36
Unemployment	0.16	-0.65	0.56	-0.02
Urban.population	0.84	0.04	0.1	0.2

РИС. 8. ЗАВИСИМОСТЬ ЧЕТЫРЁХ ГЛАВНЫХ КОМПОНЕНТ ОТ ОБЪЯСНЯЮЩИХ ПЕРЕМЕННЫХ

ПОСТРОЕНИЕ УРАВНЕНИЯ РЕГРЕССИИ С ИСПОЛЬЗОВАНИЕМ ГЛАВНЫХ КОМПОНЕНТ

Наилучшей моделью будет линейная модель на главных компонентах

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.74732    0.04620  189.338  < 2e-16 ***
components$Dim.1  0.70013    0.02359   29.679  < 2e-16 ***
components$Dim.2  0.13622    0.03562    3.824 0.000199 ***
components$Dim.3 -0.15550    0.03857   -4.032 9.15e-05 ***
components$Dim.4  0.19349    0.04579    4.225 4.33e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5505 on 137 degrees of freedom
Multiple R-squared:  0.8716, Adjusted R-squared:  0.8678
```

РИС. 9. РЕЗУЛЬТАТЫ ВЫЧИСЛЕНИЯ РЕГРЕССИИ НА ГЛАВНЫХ КОМПОНЕНТАХ

	Информационный критерий Акайке	Байесовский критерий Шварца
Линейная модель	196,46	212,76
Нелинейная модель	168,41	194,75
Модель на главных компонентах	135,78	149,04

ТАБЛИЦА 4, СРАВНЕНИЕ КАЧЕСТВА ПОСТРОЕННЫХ МОДЕЛЕЙ

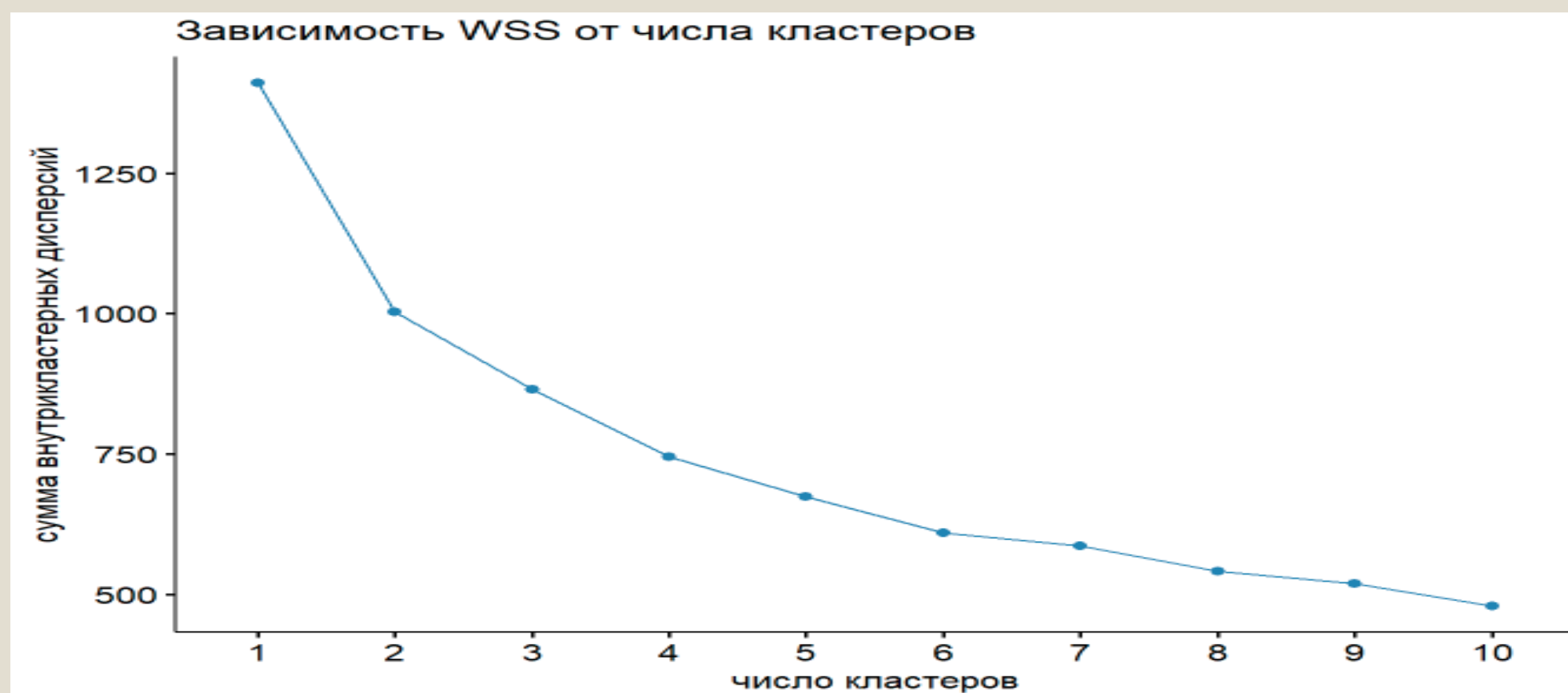


РИС. 10. МЕТОД ЛОКТЯ



РИС.11 СТАТИСТИКИ РАЗРЫВА

Метод локтя и статистики разрыва: Оптимальное количество кластеров – 2 или 4

Командирским решением было принято взять 4 кластера.



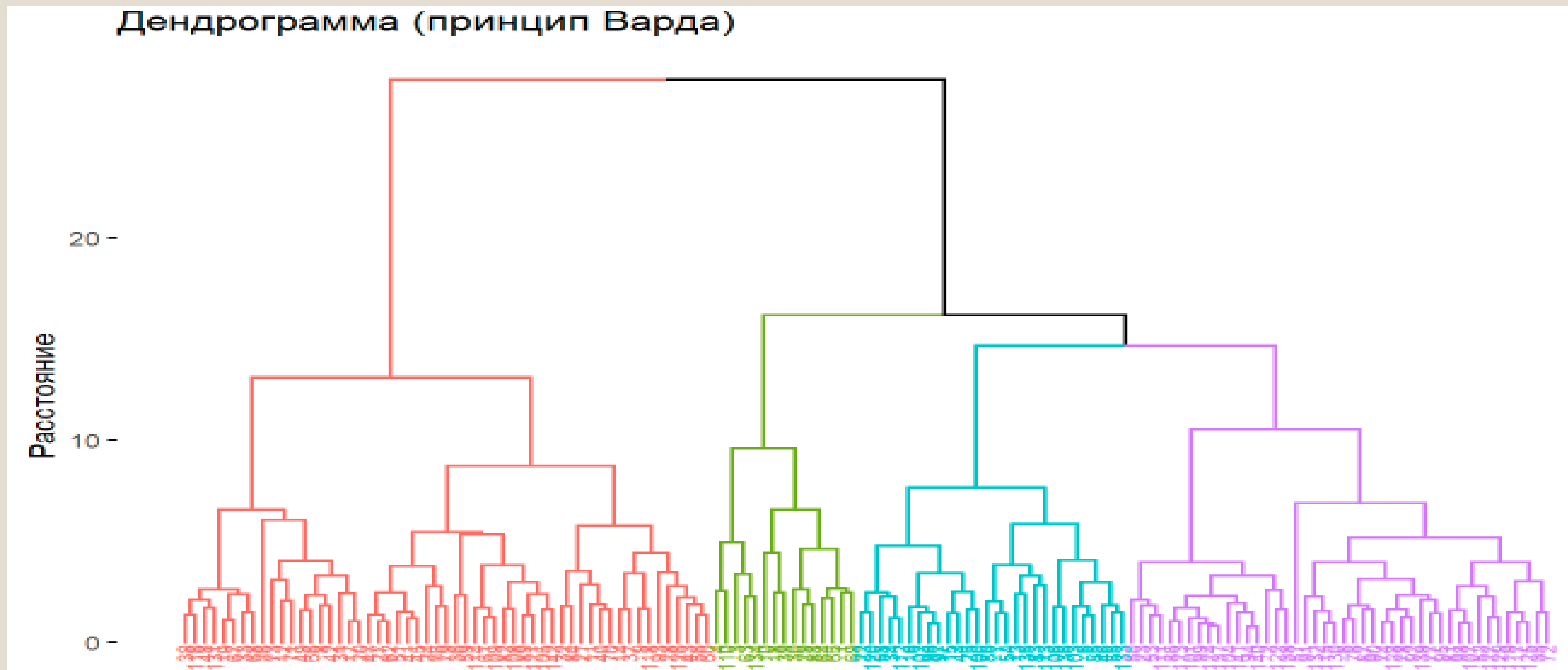
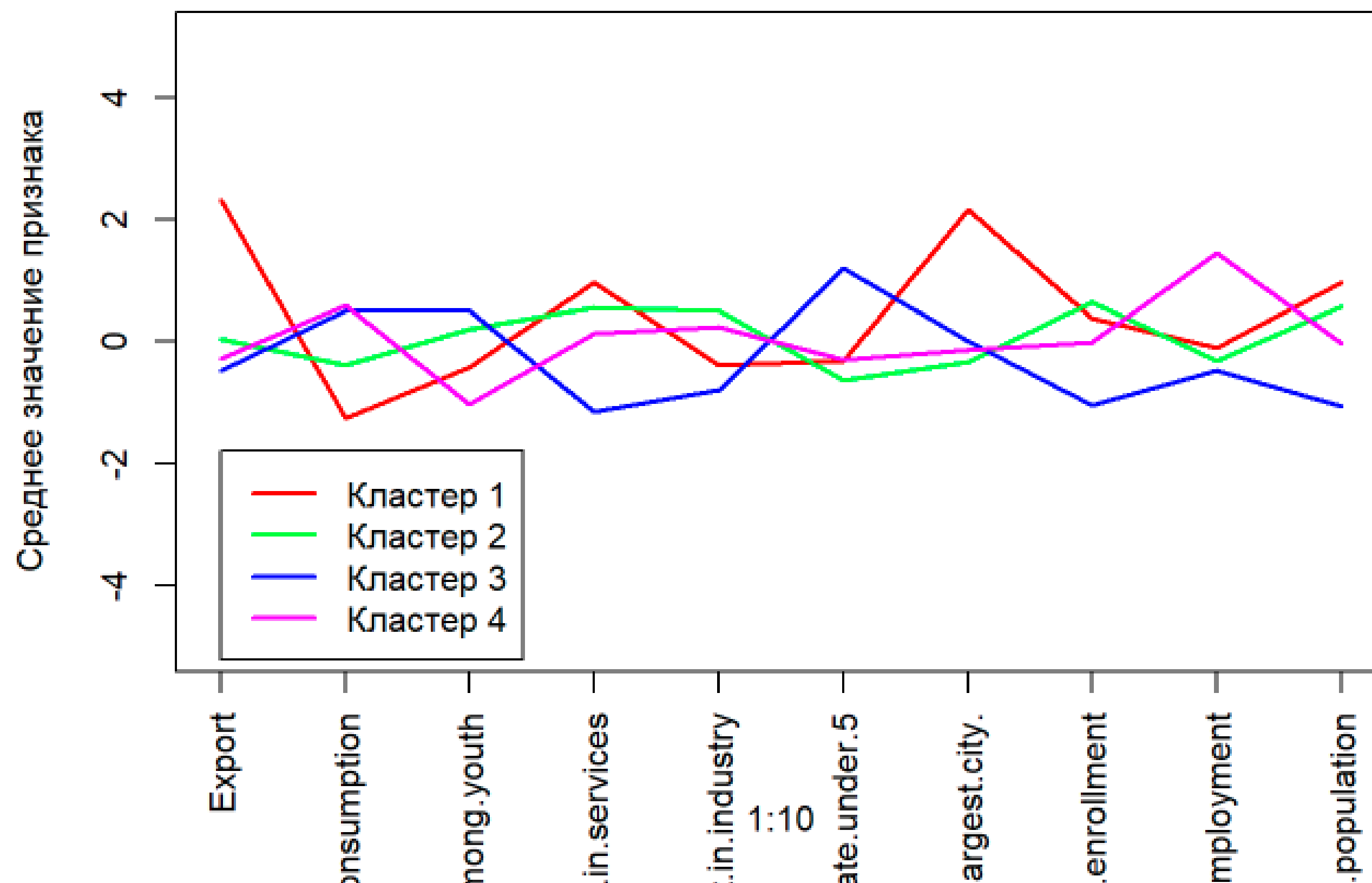


РИС. 12. ДЕНДРОГРАММА, ПОСТРОЕННАЯ ПО МЕТОДУ ВАРДА

ГРАФИК СРЕДНИХ ЗНАЧЕНИЙ В КЛАСТЕРАХ

График средних (признаки стандартизованы)



При проверке гипотезы о равенстве средних лишь 13 из 60 тестов не отверглись на уровне значимости 0.05, что даёт нам возможность сделать вывод о наличии значимых различий между полученными кластерами.

РИС. 13. ВИЗУАЛИЗАЦИЯ РАЗБИЕНИЯ СТРАН ПО КЛАСТЕРАМ

ВИЗУАЛИЗАЦИЯ КЛАСТЕРИЗАЦИИ

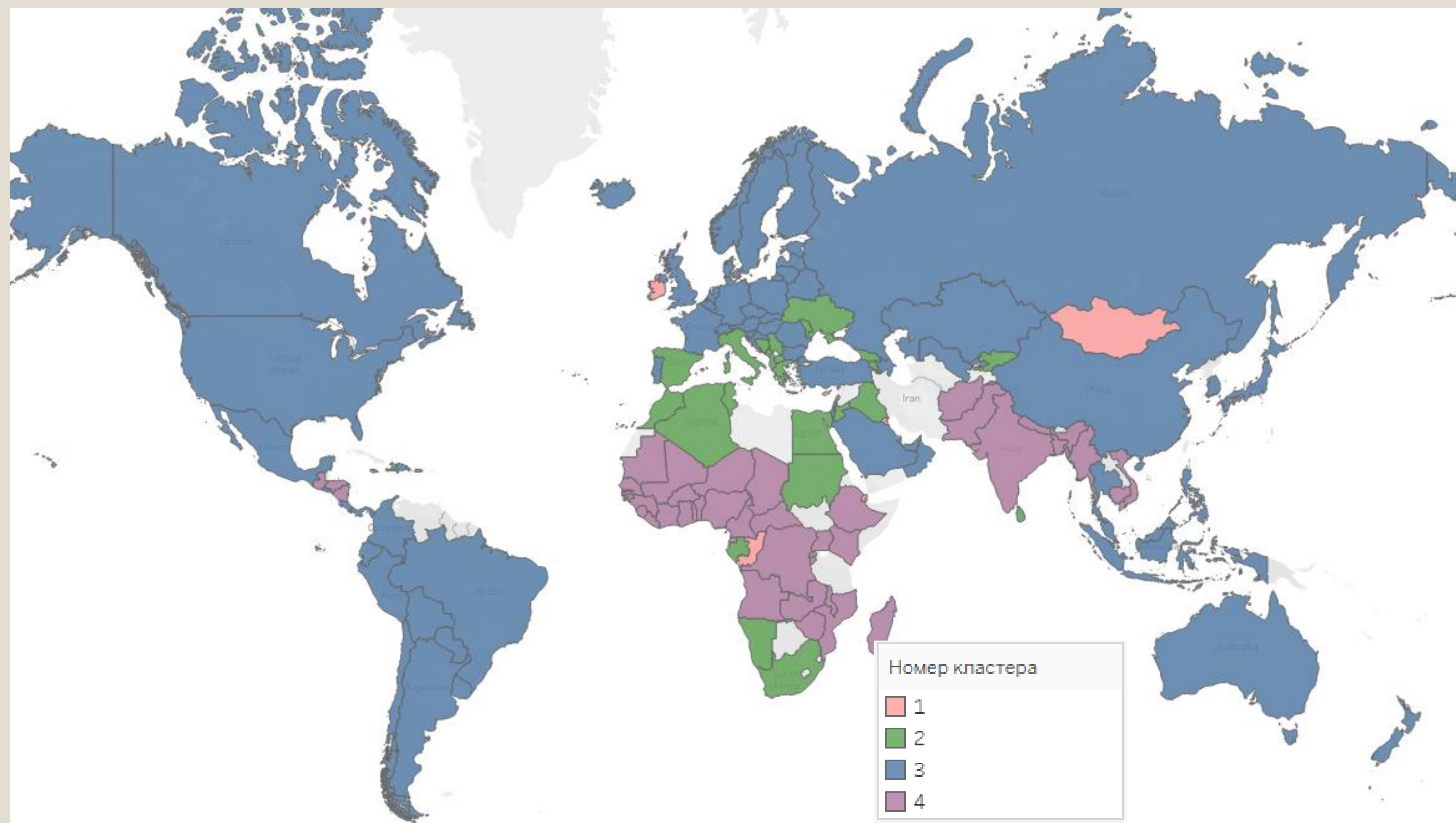


РИС. 14. ВИЗУАЛИЗАЦИЯ РАЗБИЕНИЯ СТРАН ПО КЛАСТЕРАМ

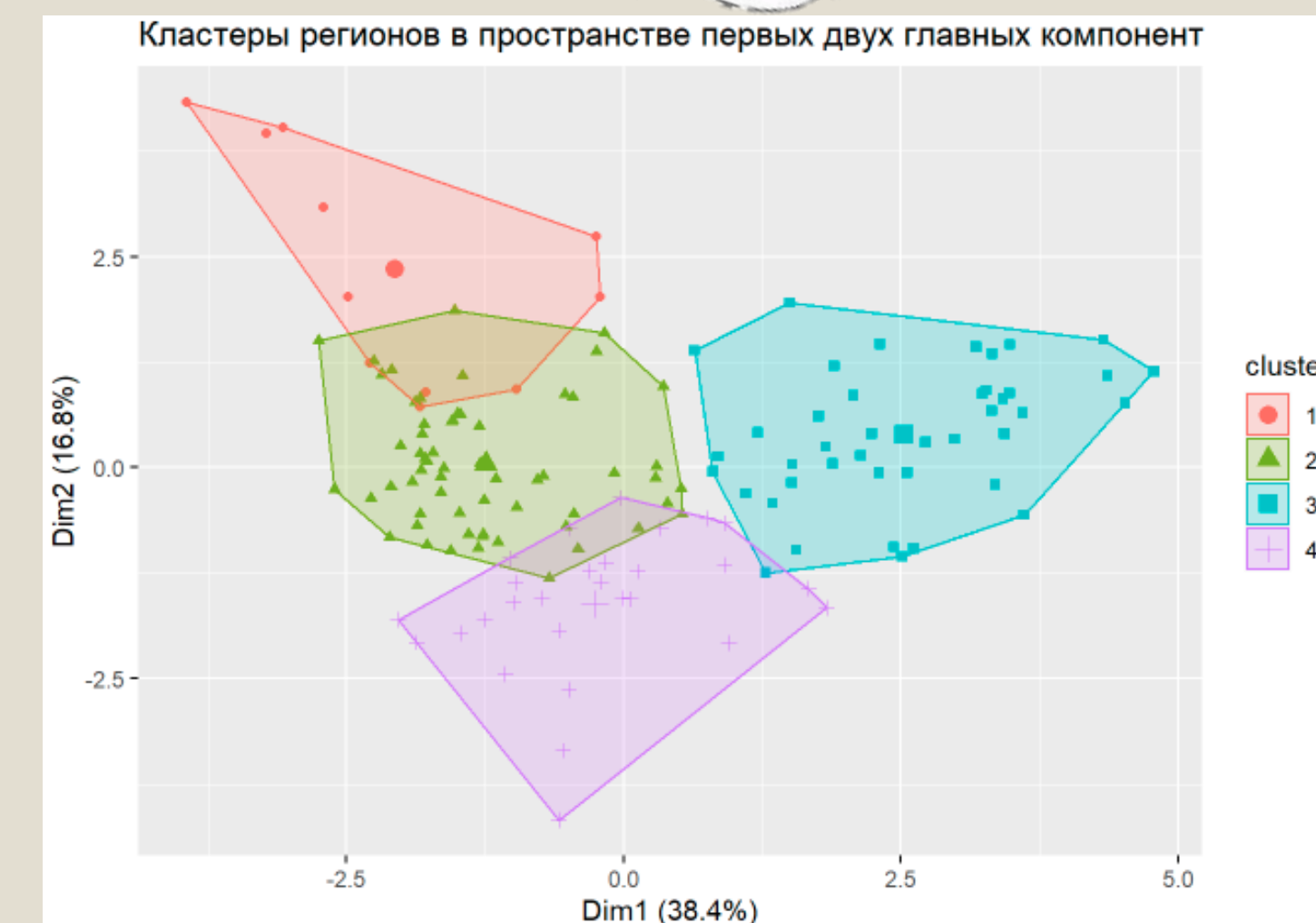


РИС. 15. РАЗБИЕНИЕ СТРАН ПО КЛАСТЕРАМ В ПРОСТРАНСТВЕ ПЕРВЫХ ДВУХ КОМПОНЕНТ



ТИПОЛОГИЧЕСКАЯ РЕГРЕССИЯ

Факультет экономических наук

Используемые переменные	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Экспорт				
Потребление	-0,261	-0,220	-0,189	-0,232
Участие молодежи в рабочей силе		0,140		
Занятость в секторе услуг	0,908	0,706		0,569
Занятость в секторе промышленности				
Доля населения в крупнейшем городе от всего городского населения				
Смертность среди детей до 5 лет	-0,224	-0,925		
Общий уровень безработицы				
Коэффициент вовлеченности в профессиональное образование			0,445	
Доля городского населения от всего населения страны			0,471	

ТАБЛИЦА 5, СРАВНЕНИЕ КОЭФФИЦИЕНТОВ ДЛЯ ТИПОЛОГИЧЕСКИХ РЕГРЕССИЙ

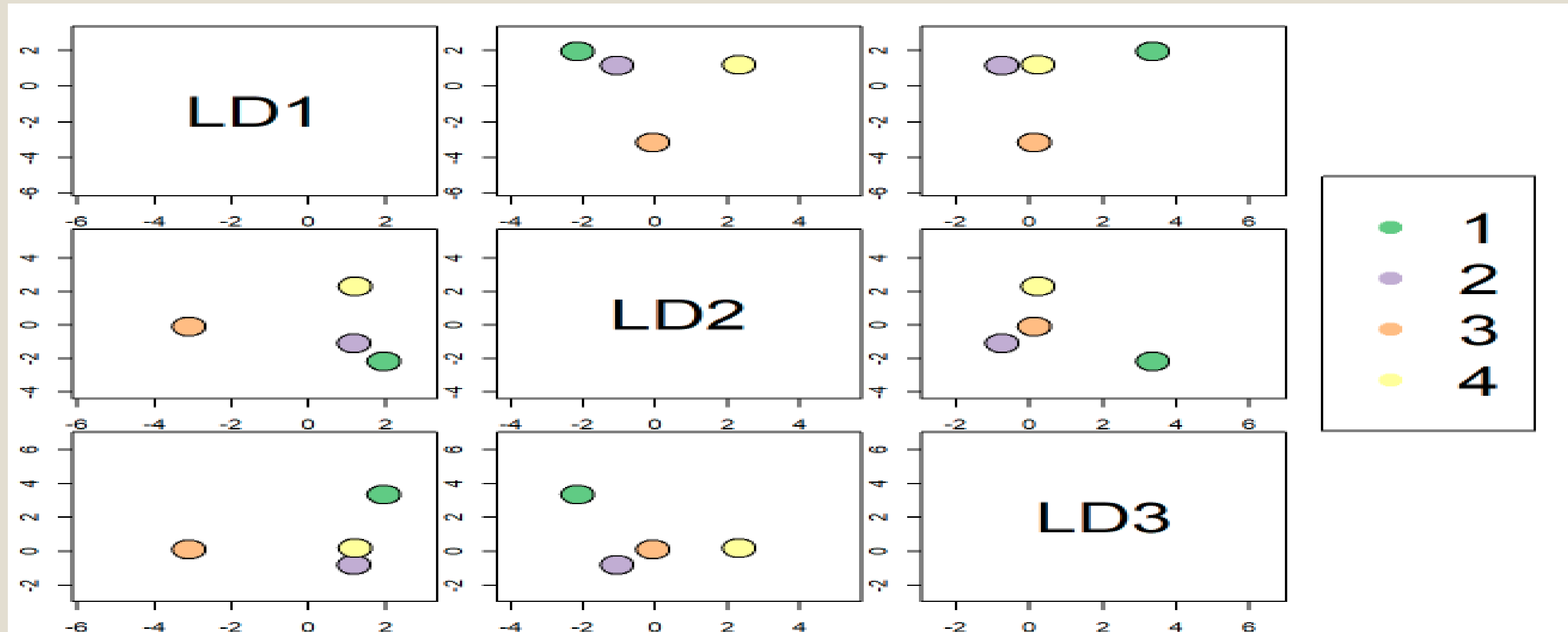


РИС. 16. ЦЕНТРОИДЫ ГРУПП В ПРОСТРАНСТВЕ ТРЕХ ДИСКРИМИНАНТНЫХ ФУНКЦИЙ

В целом, на большинстве графиков можно увидеть довольно репрезентативную визуализацию центроидов, что свидетельствует о высоком уровне классификации.

Classification table:

obs				
pred	1	2	3	4
1	5	0	0	0
2	0	22	1	0
3	0	0	13	0
4	0	1	1	4

Misclassification errors:

1	2	3	4
0.00	4.35	13.33	0.00

Результаты классификации наблюдений из тестовой выборки дали весьма хорошие результаты, позволяющие с высокой точностью отнести неизвестную страну к тому или иному кластеру.

Значение Лямбды Уилкса, равное 0,017, увеличивает нашу уверенность в качестве полученного результата.

РИС. 17. ТАБЛИЦА СООТВЕТСТВИЯ
ПРЕДСКАЗАННЫХ КЛАССОВ ИСХОДНЫМ

СПАСИБО ЗА ВНИМАНИЕ!



ГОТОВЫ ОТВЕТИТЬ НА ВАШИ ВОПРОСЫ!

