National Research University Higher School of Economics
International Centre of Decision Choice and Analysis

NEW METHODS OF PATTERN ANALYSIS: THEORETICAL ASPECTS AND PRACTICAL APPLICATION

Alexey Myachin
amyachin@hse.ru

Autumn School "Advances in Decision Analysis"

Internet Live Stats (2022):

5,2

billion Internet users worldwide

1,9

billion sites

5,1

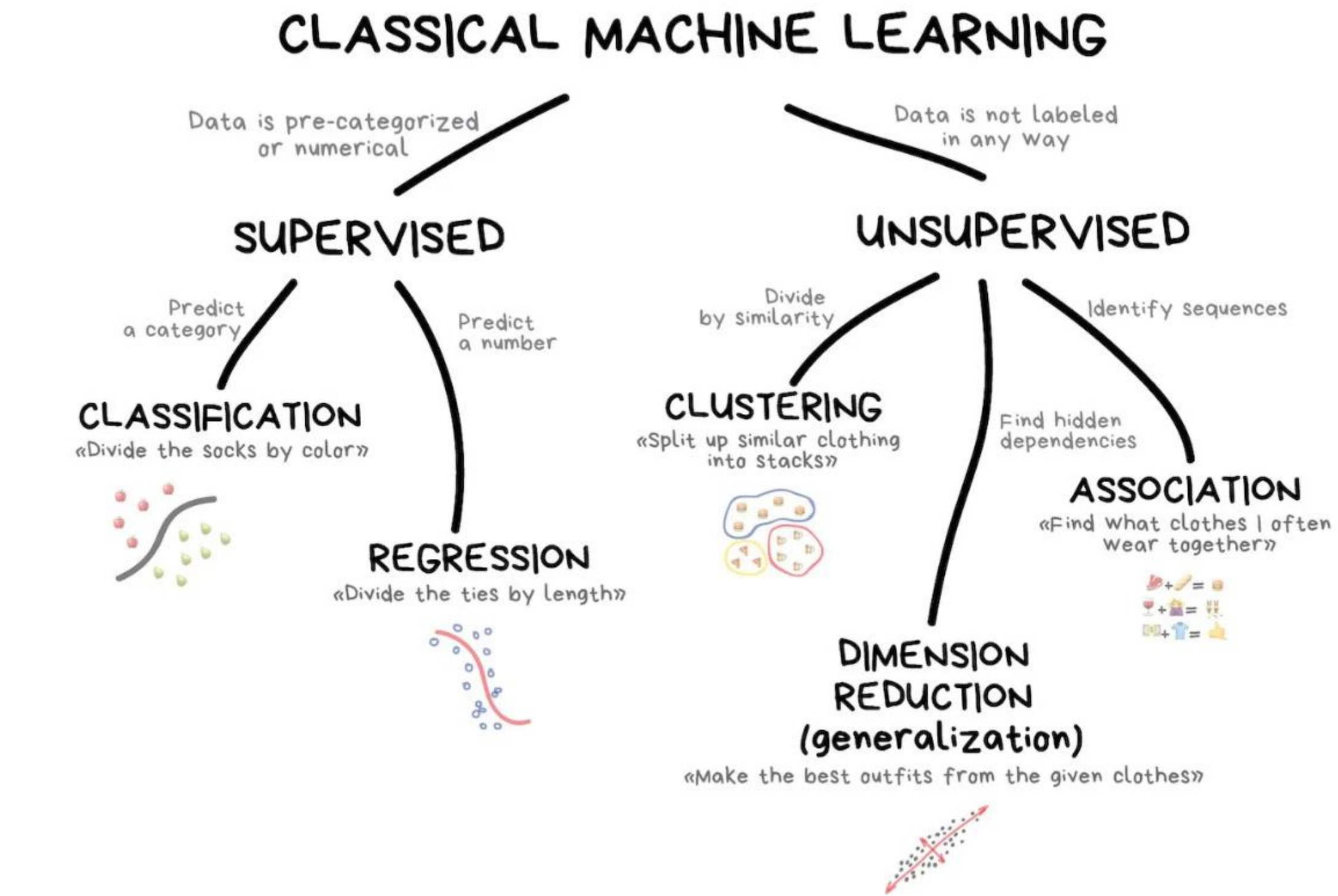
billion videos viewed on YouTube (for one day)

176

billion e-mail sent (for one day)

8

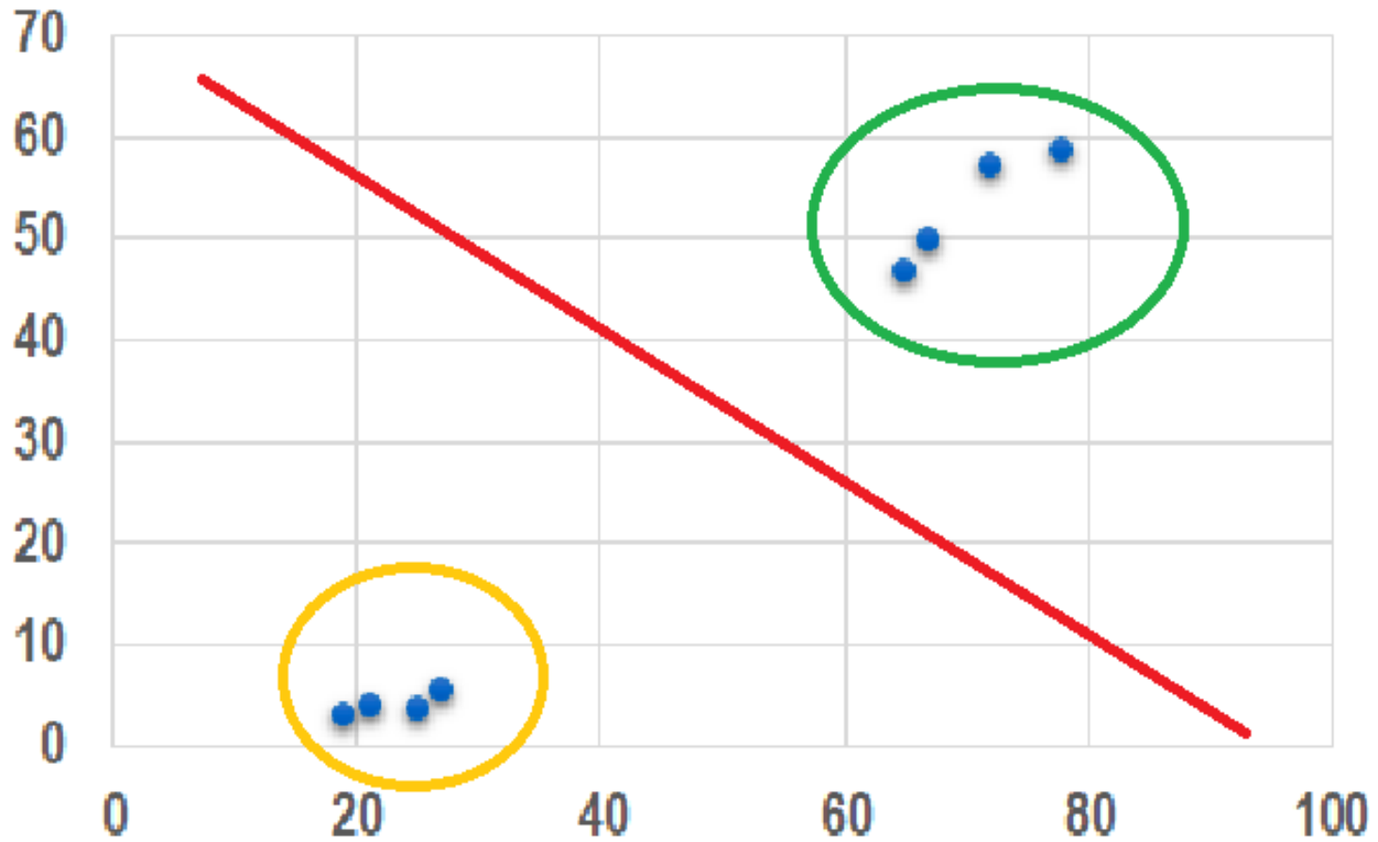
billion GB of Internet traffic (for one day)

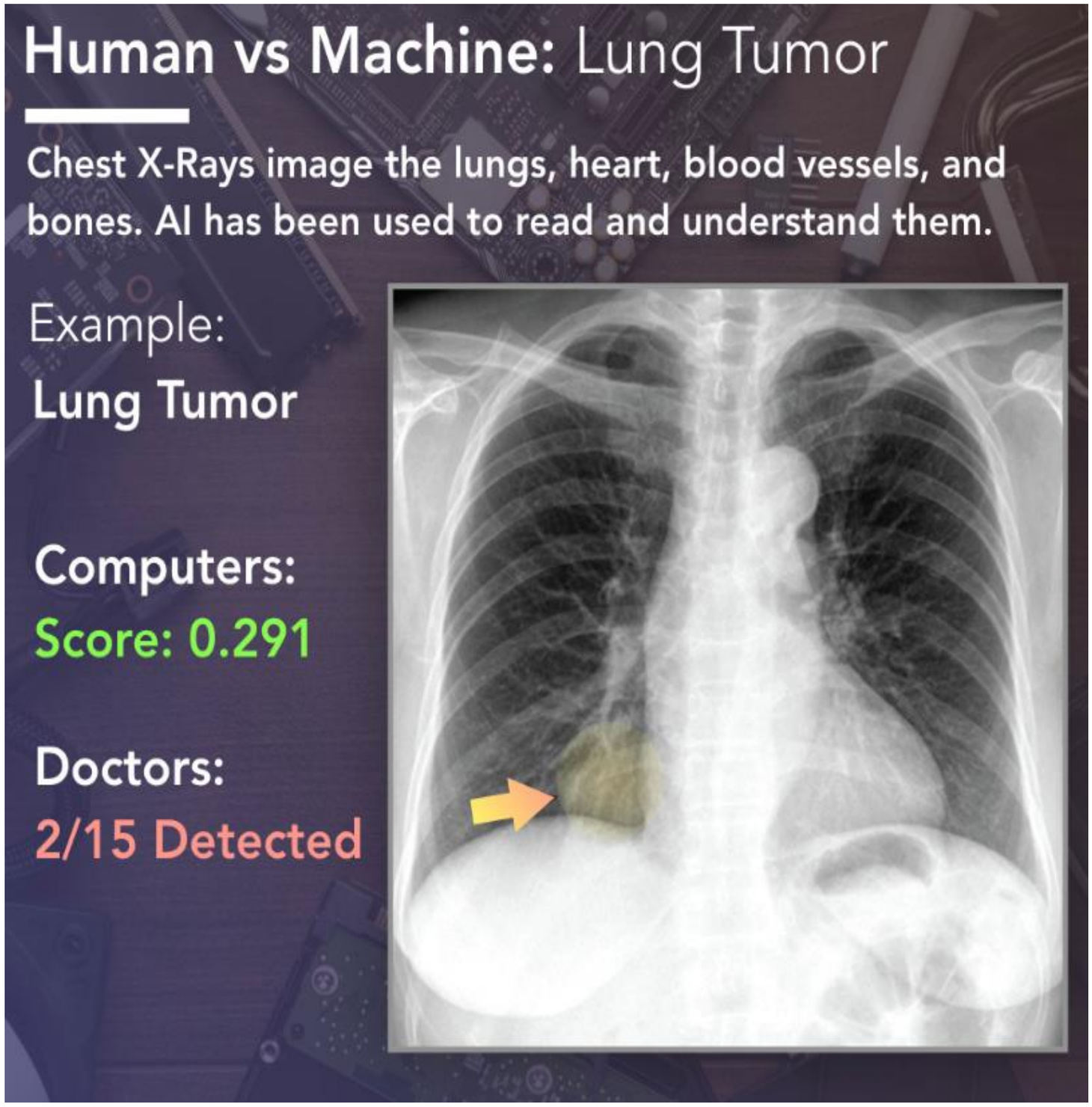


https://vitalflux.com/great-mind-maps-for-learning-machine-learning/

CLASSIFICATION

	Height	Weight
Dog	65	46,8
Dog	67	49,9
???	72	57
Dog	78	58,6
Cat	25	3,5
Cat	27	5,5
???	21	4
Cat	19	3

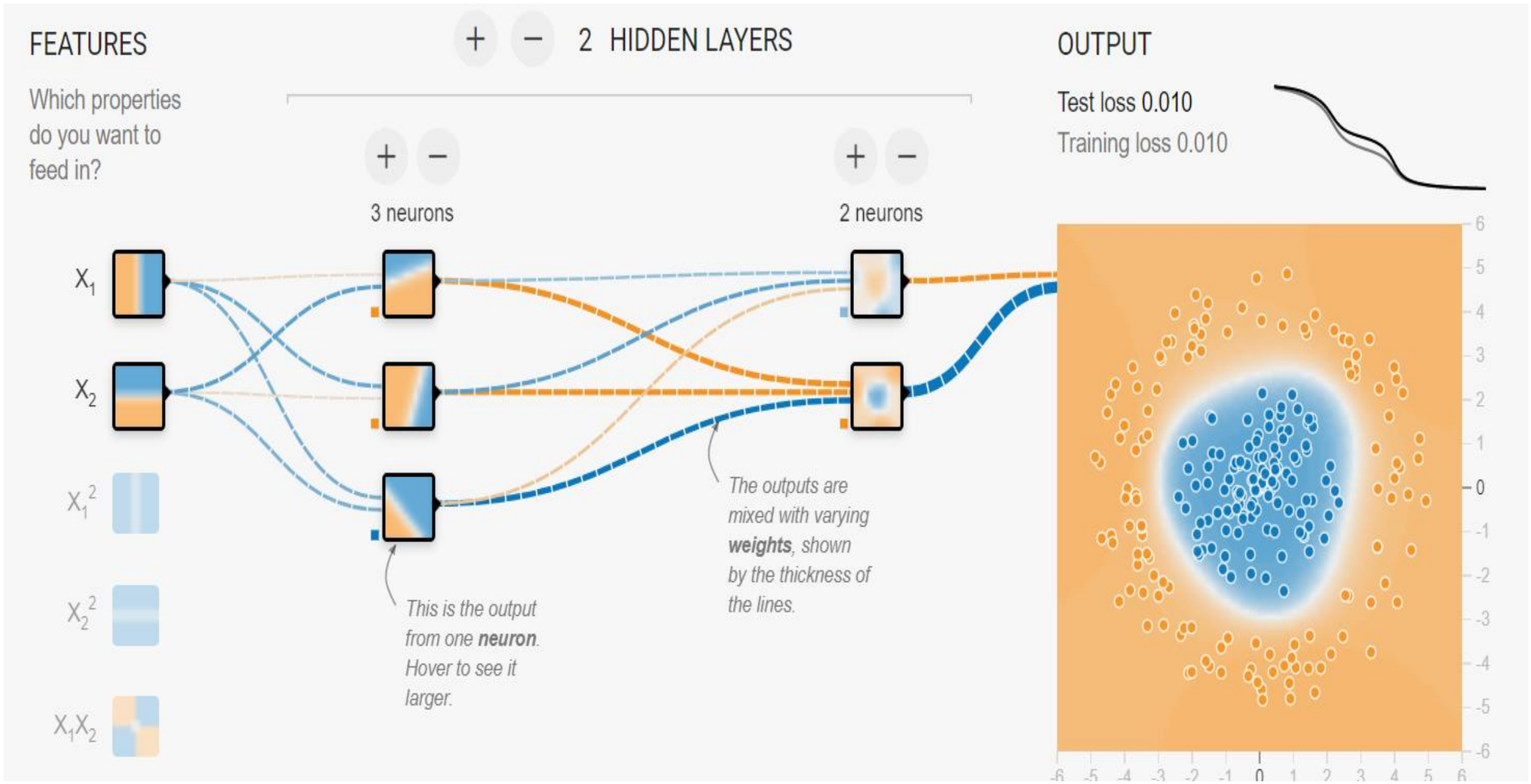




https://www.clearvuehealth.com/b/ai-radiology-xray/

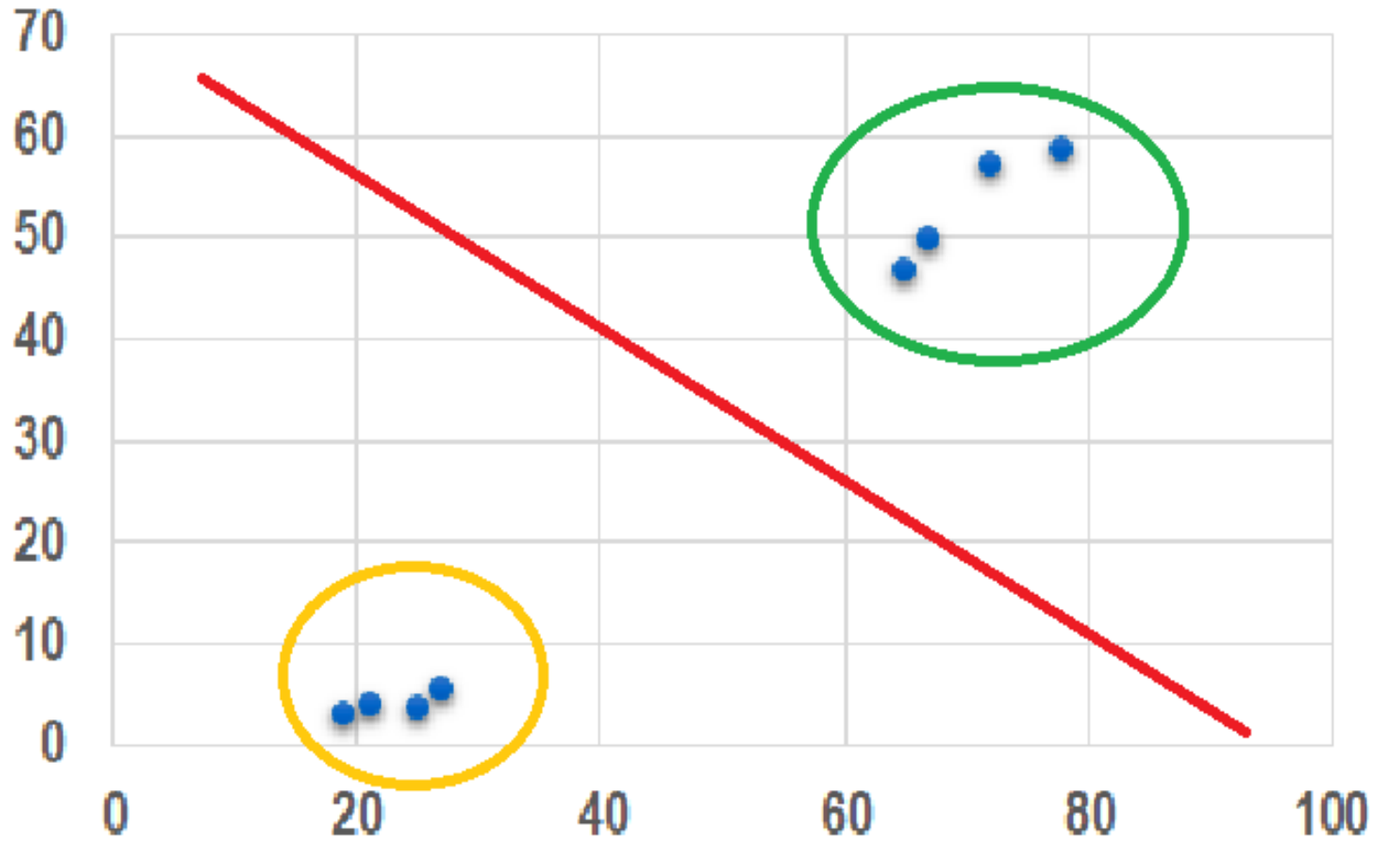
EXAMPLE

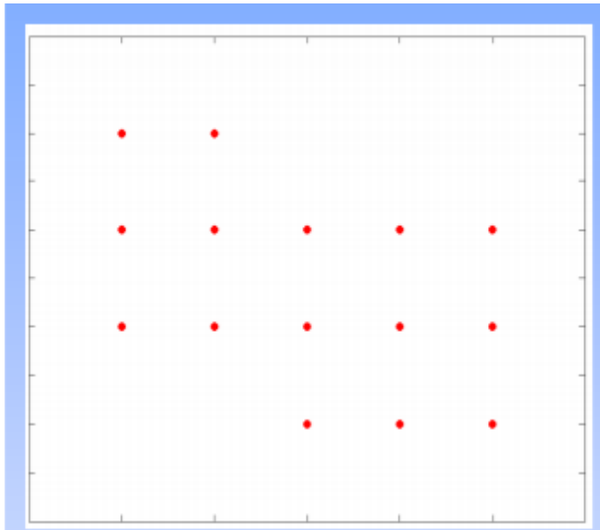
http://playground.tensorflow.org



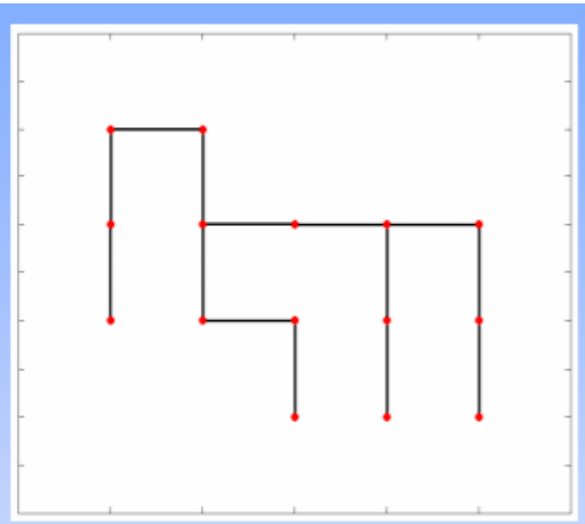
CLUSTERING

	Height	Weight
?	65	46,8
?	67	49,9
?	72	57
?	78	58,6
?	25	3,5
?	27	5,5
?	21	4
?	19	3

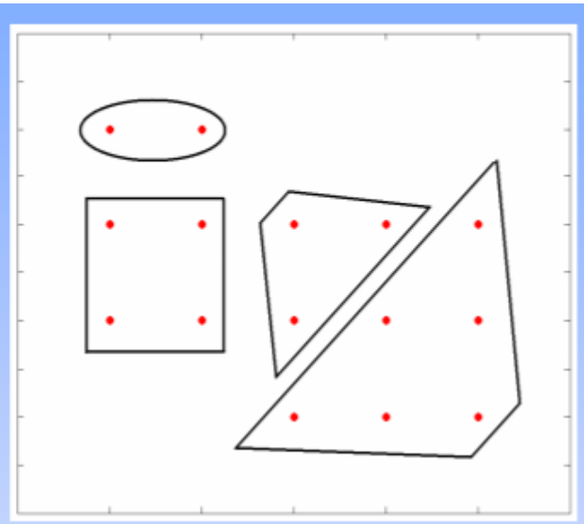




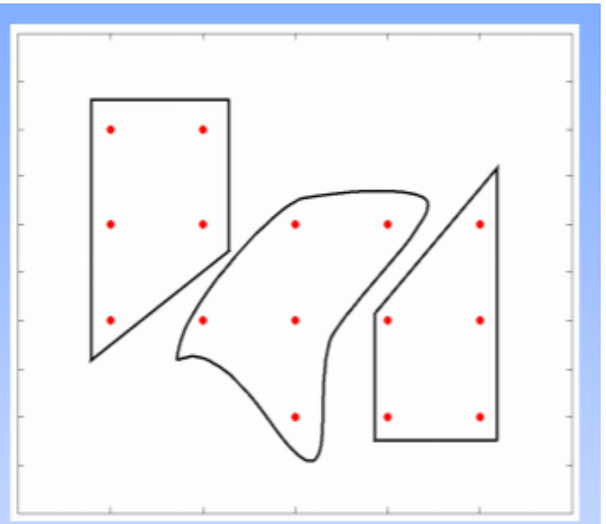
15 Data points



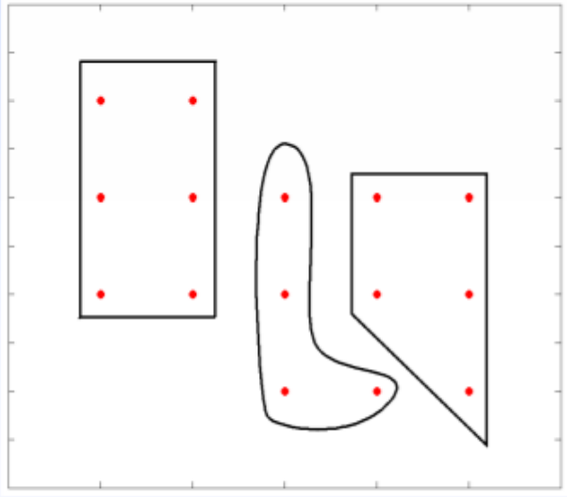
MST



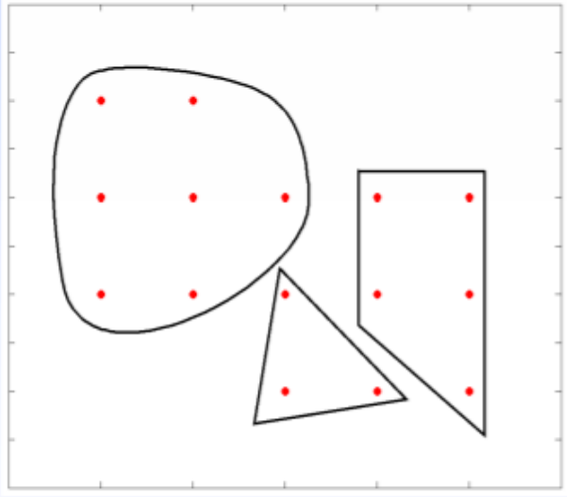
FORGY



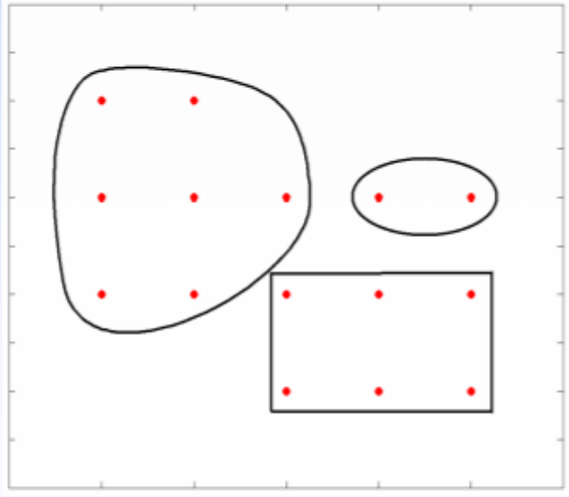
ISODATA



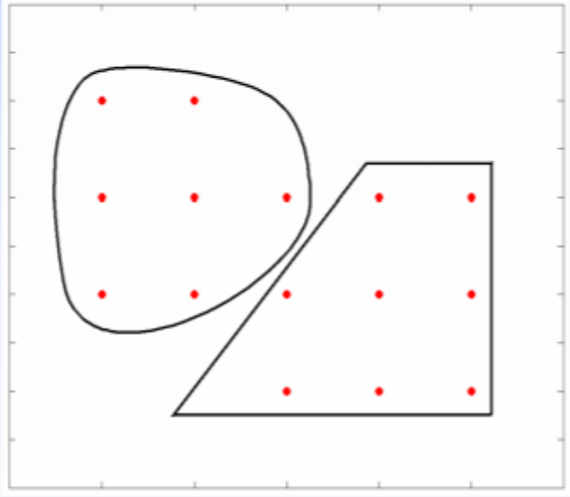
WISH



CLUSTER

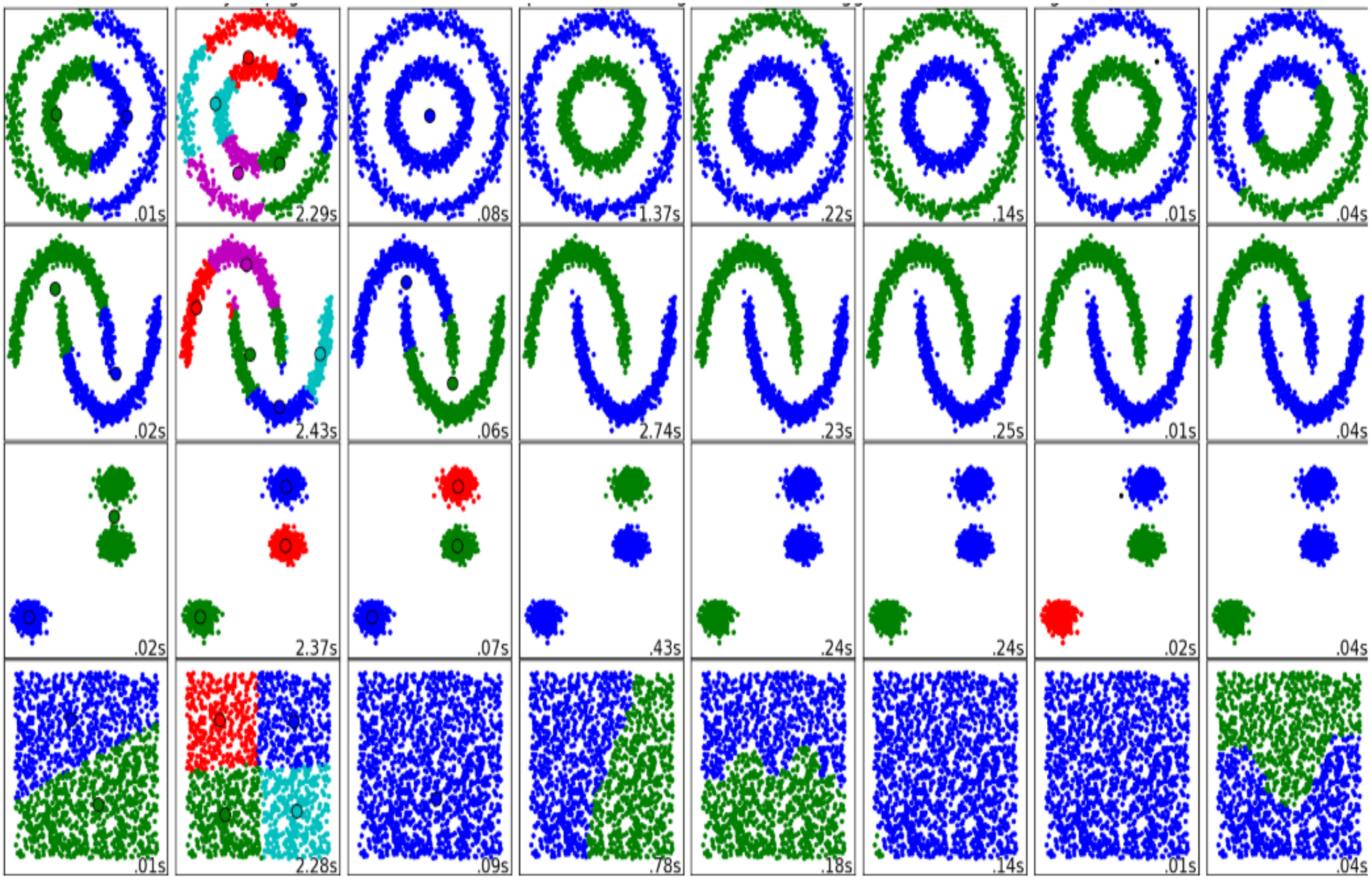


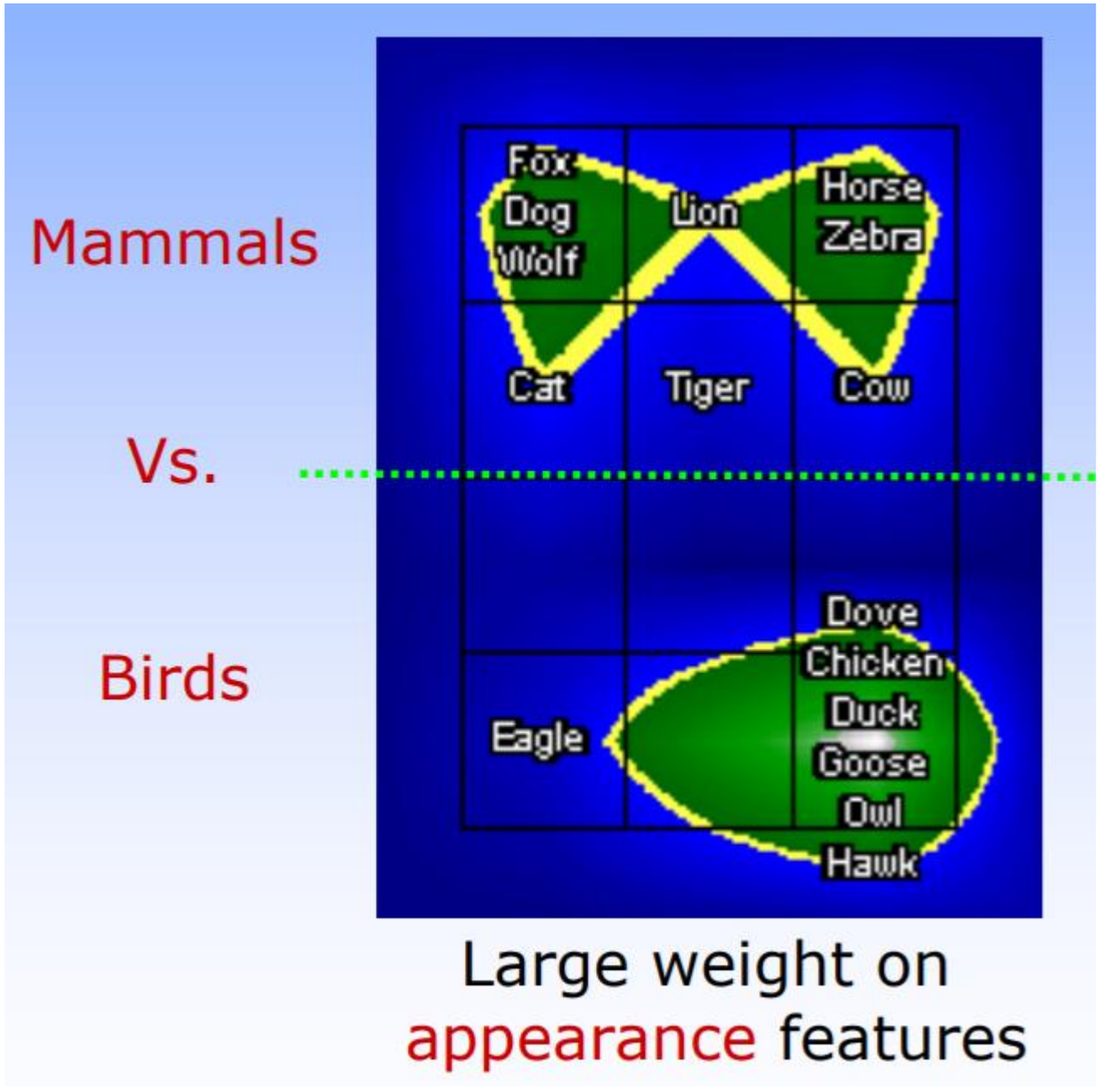
Complete Link

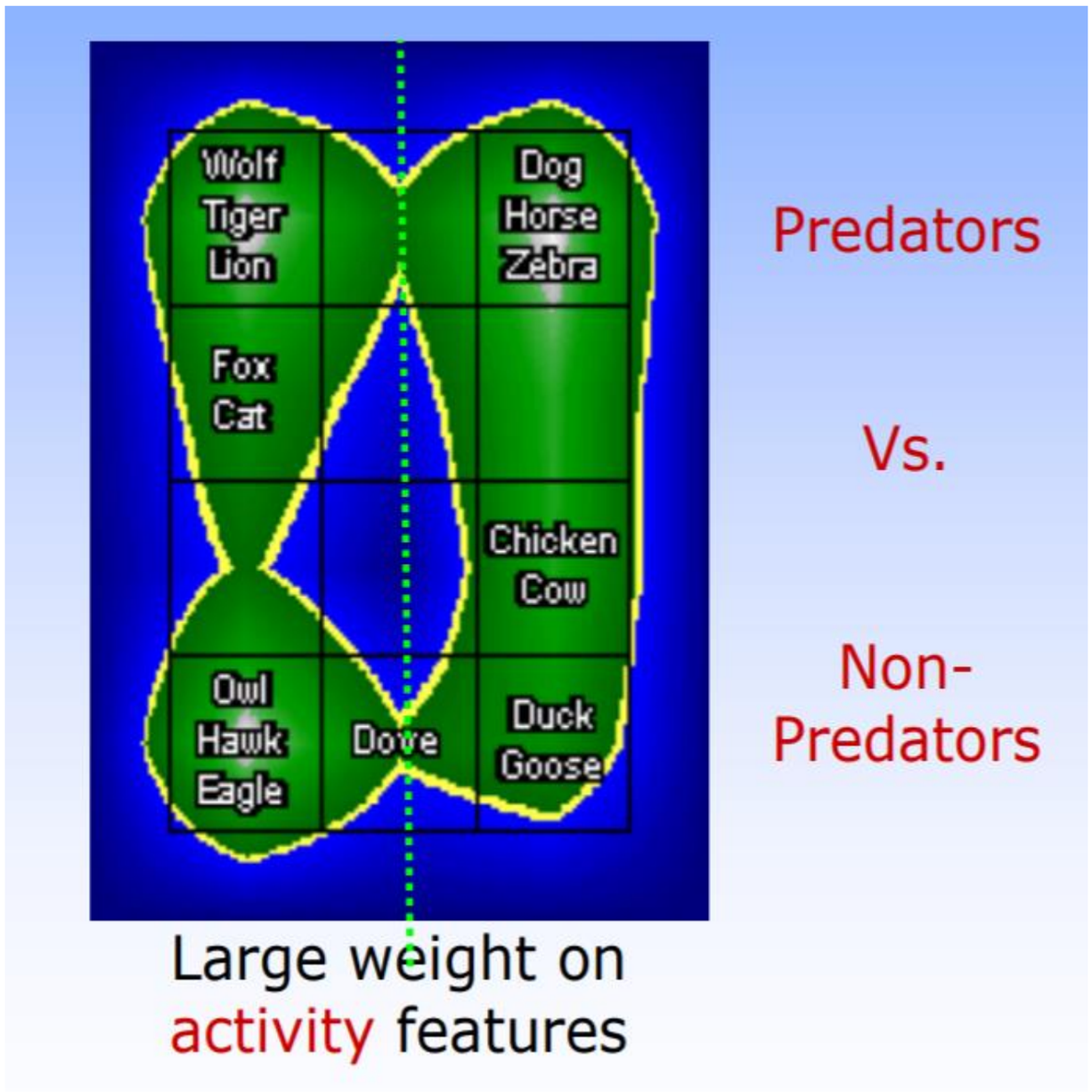


JP

Dubes and Jain, "Clustering Techniques: User's Dilemma", *Pattern Recognition*, 1976


.01s
2.29s
.08s
1.37s
.22s
.14s
.01s
.04s
.02s
2.43s
.06s
2.74s
.23s
.25s
.01s
.04s
.02s
2.37s
.07s
.43s
.24s
.24s
.02s
.04s
.01s
2.28s
.09s
.78s
.18s
.14s
.01s
.04s

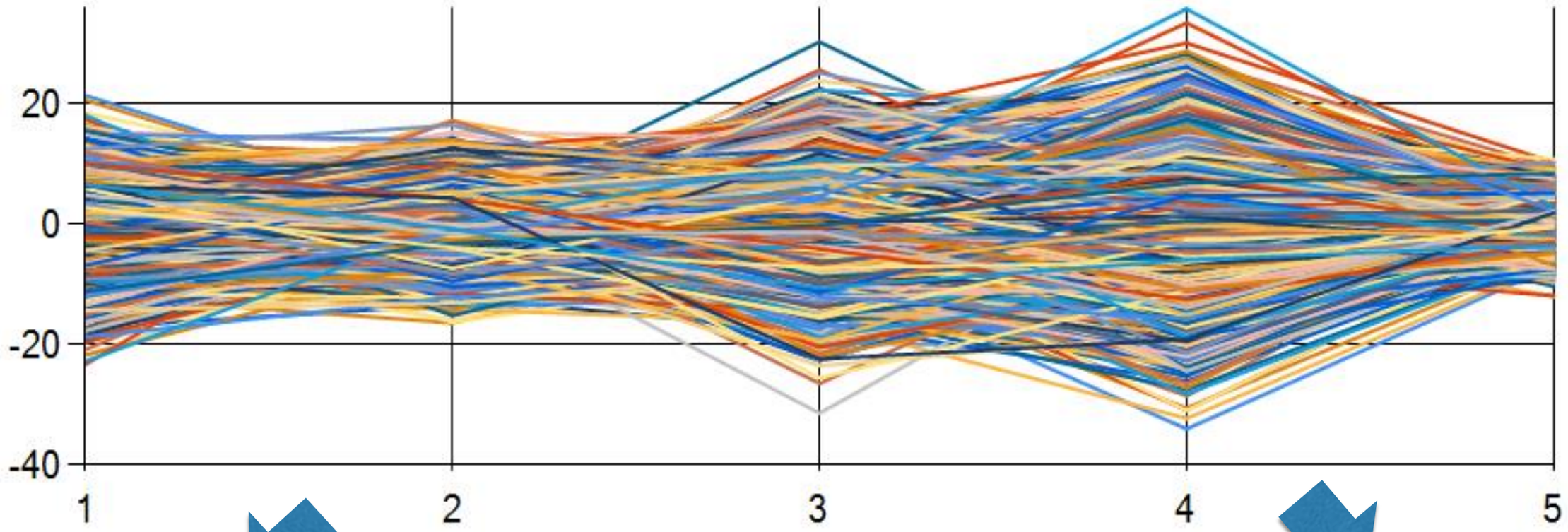


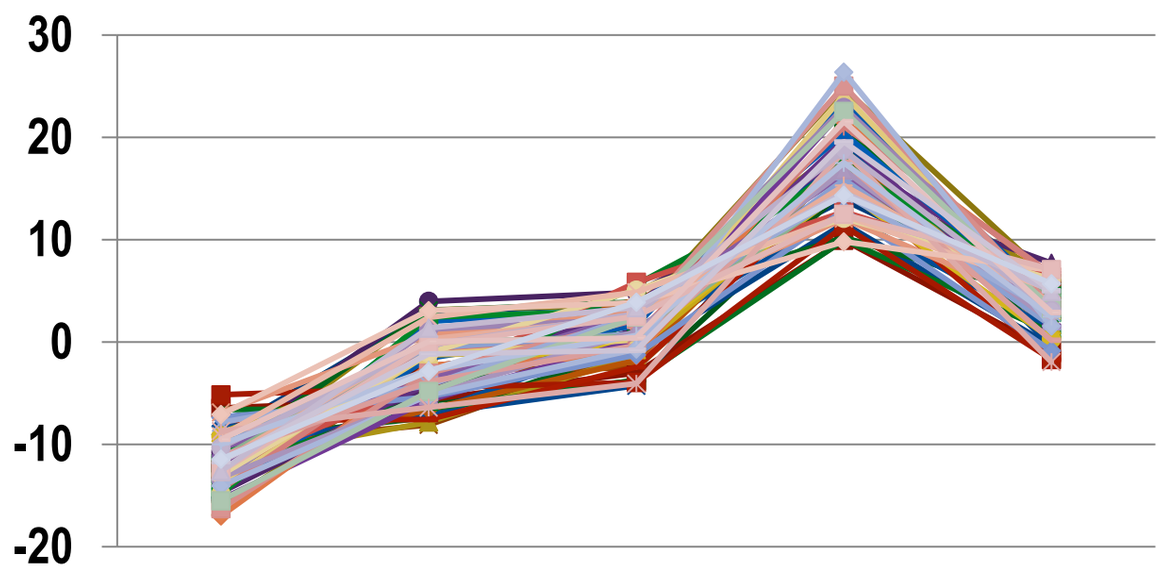


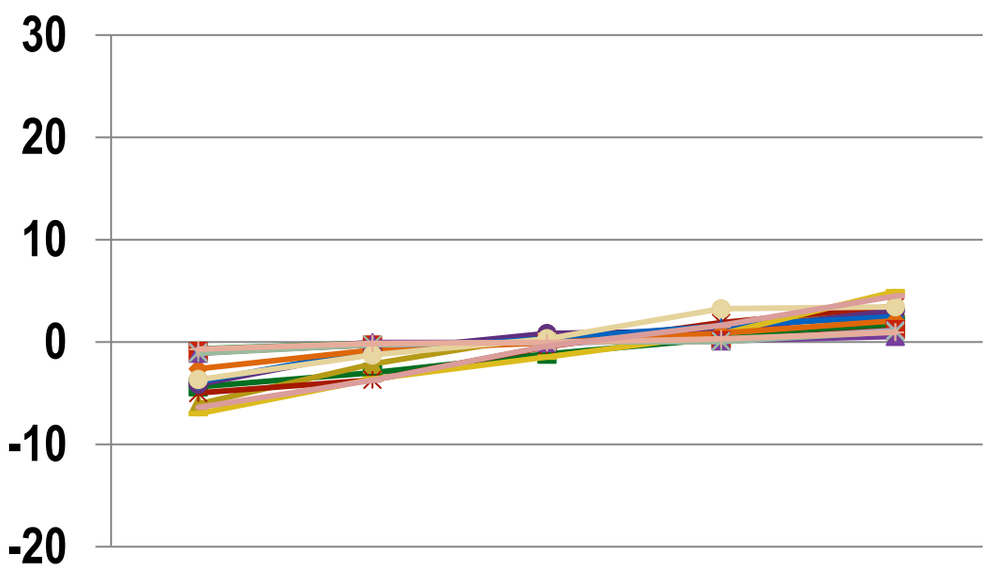
https://pdfs.semanticscholar.org/1823/98ca4d85c25c64ba238dad10cafc92203660.pdf

http://www.ofai.at/~elias.pampalk/kdd03/animals/

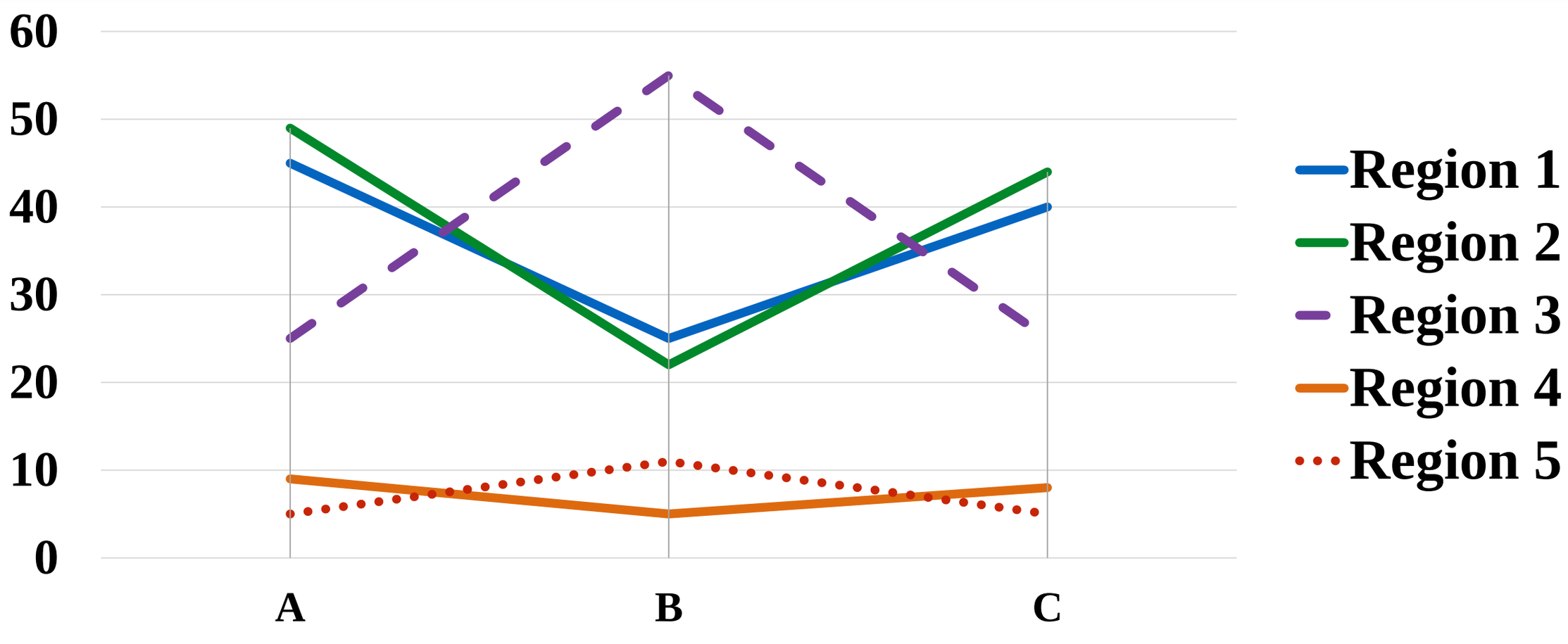
PATTERN ANALYSIS



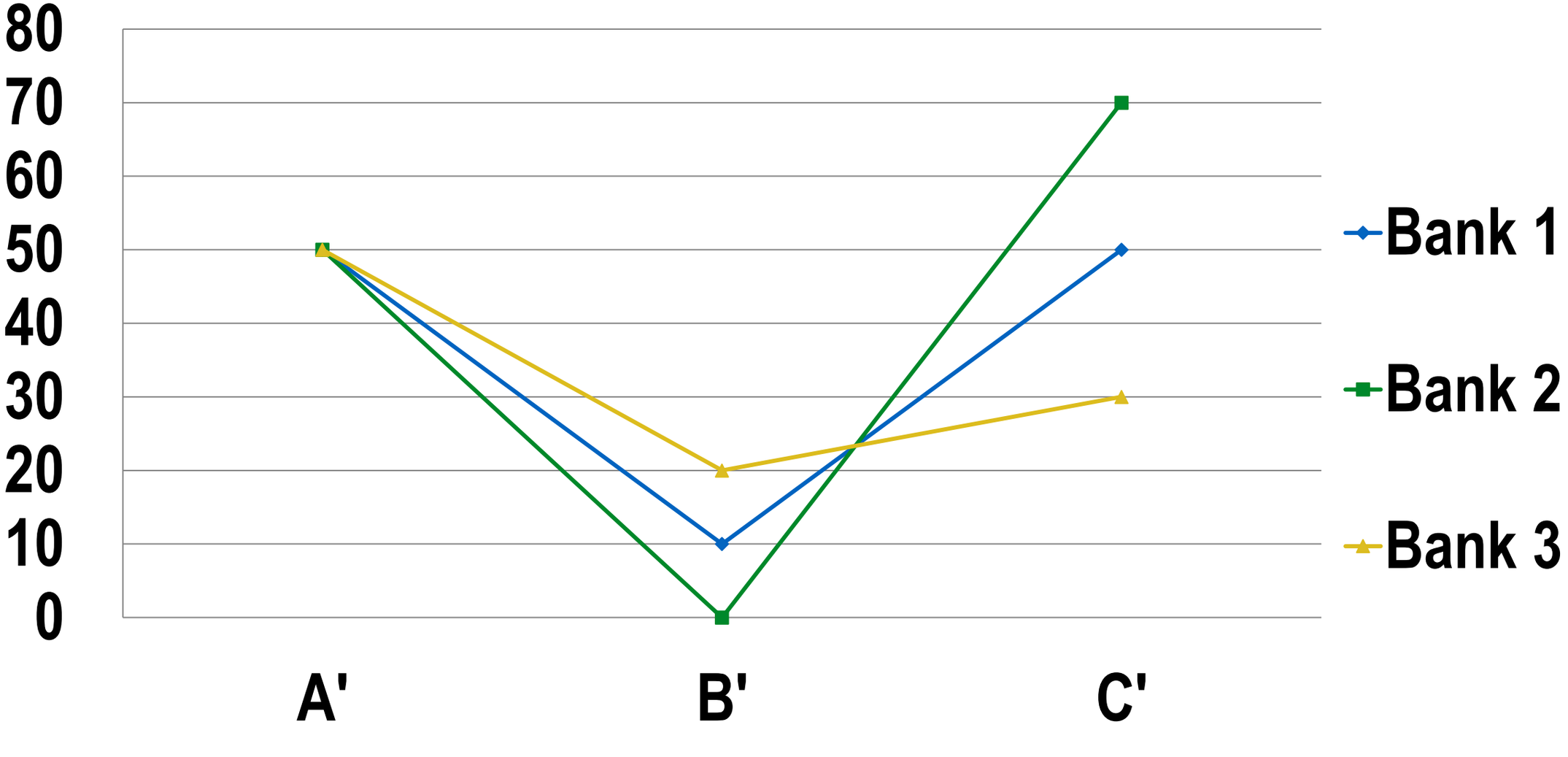


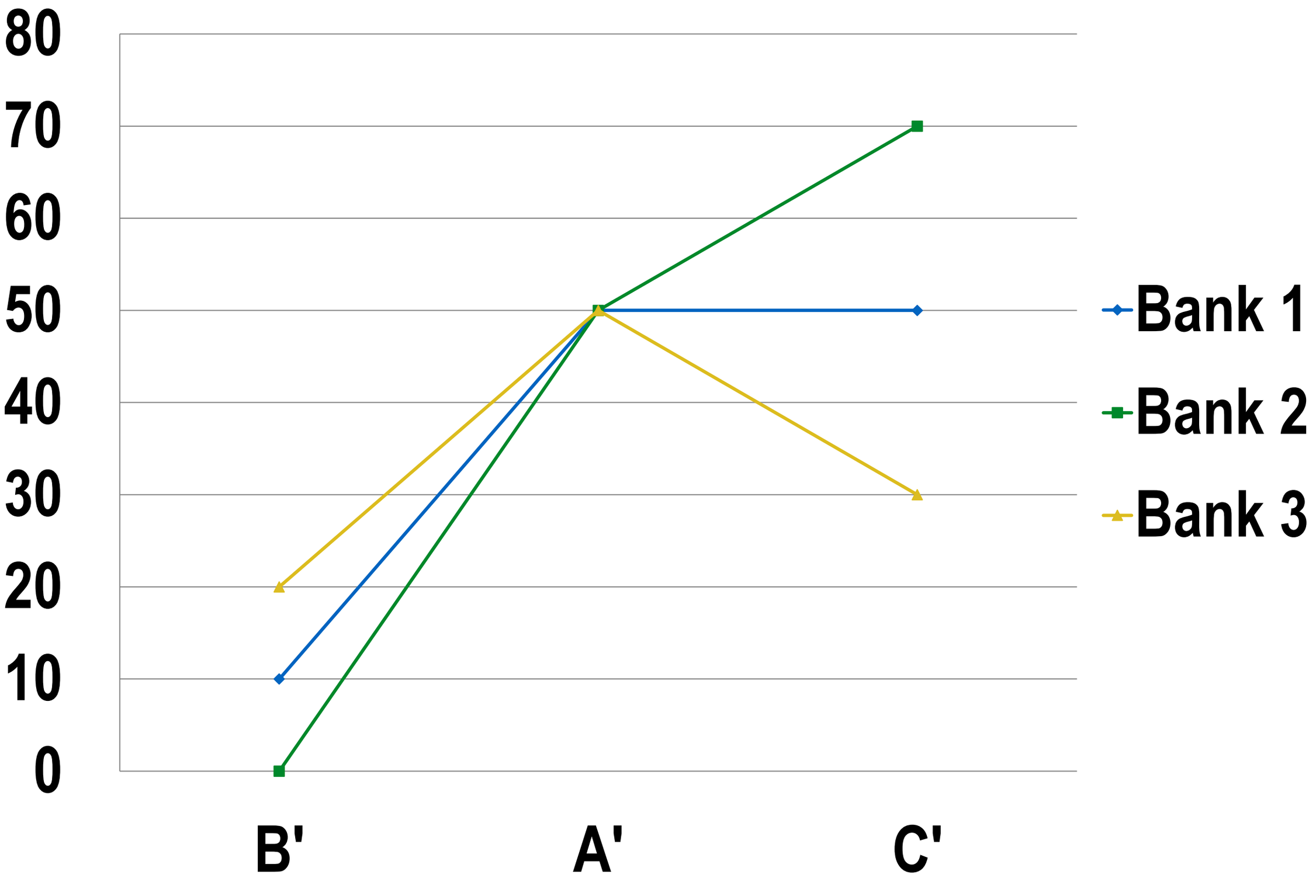


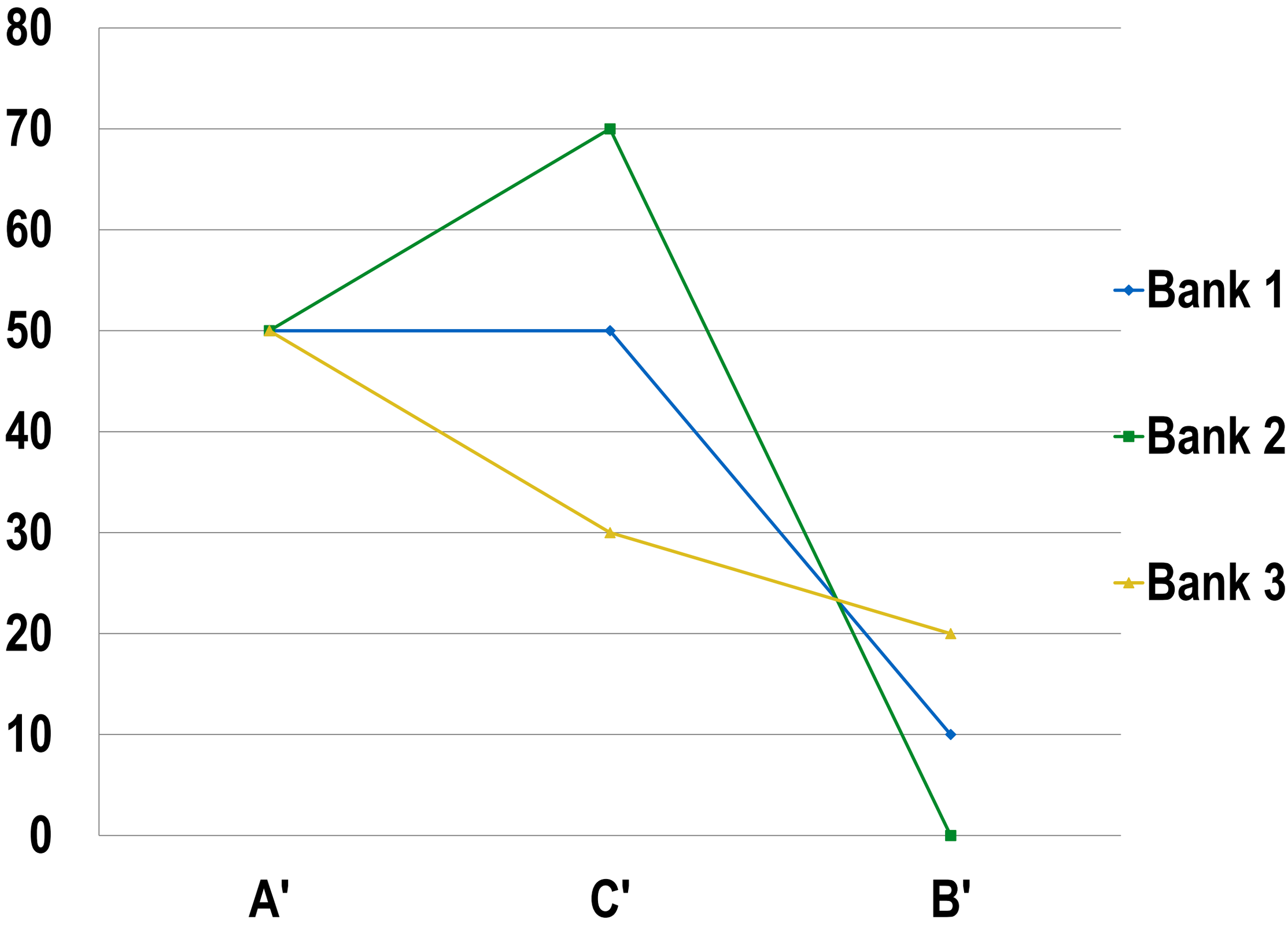
Regions	A	B	C
Region 1	45	25	40
Region 2	49	22	44
Region 3	25	55	25
Region 4	9	5	8
Region 5	5	11	5

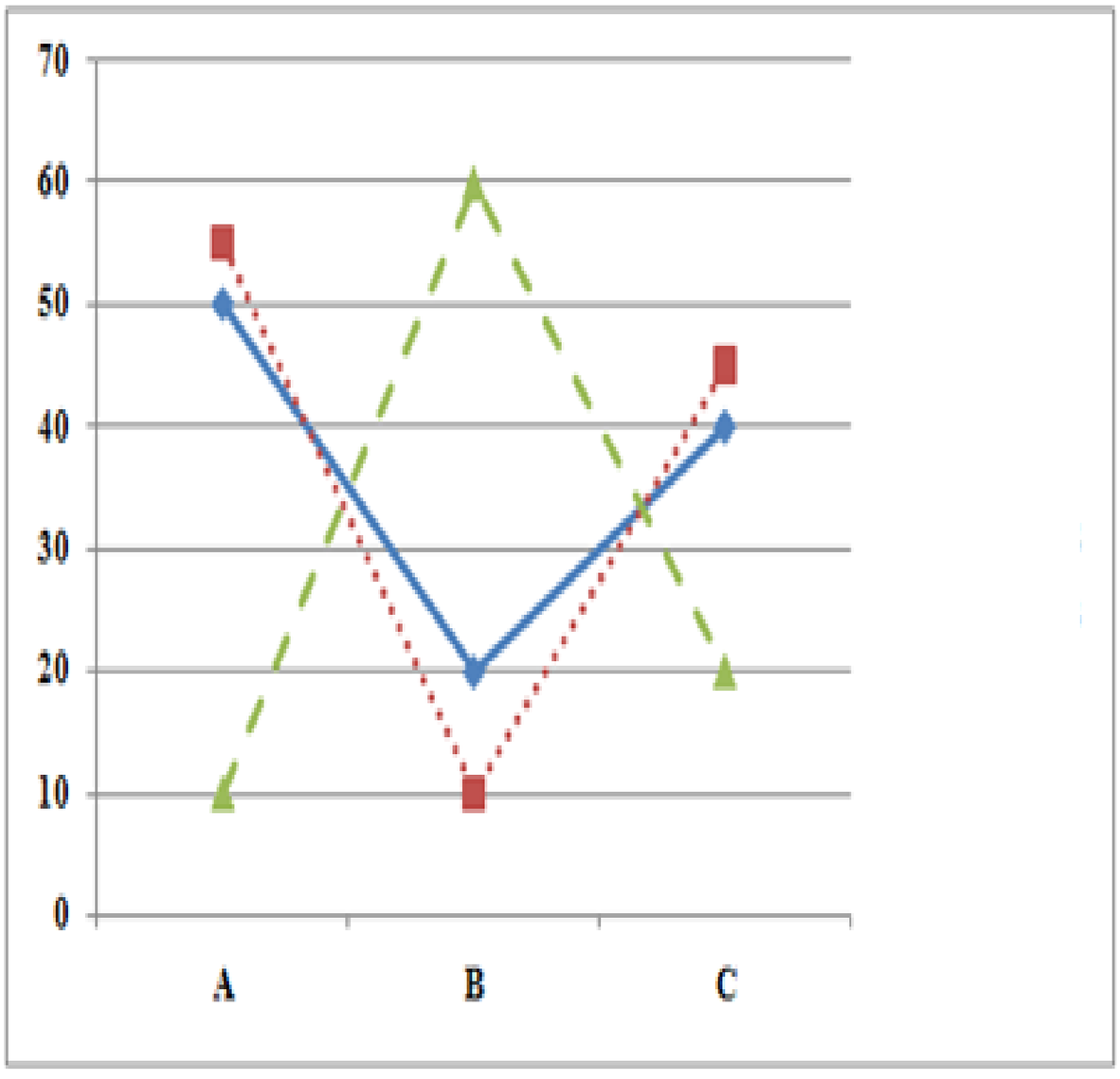


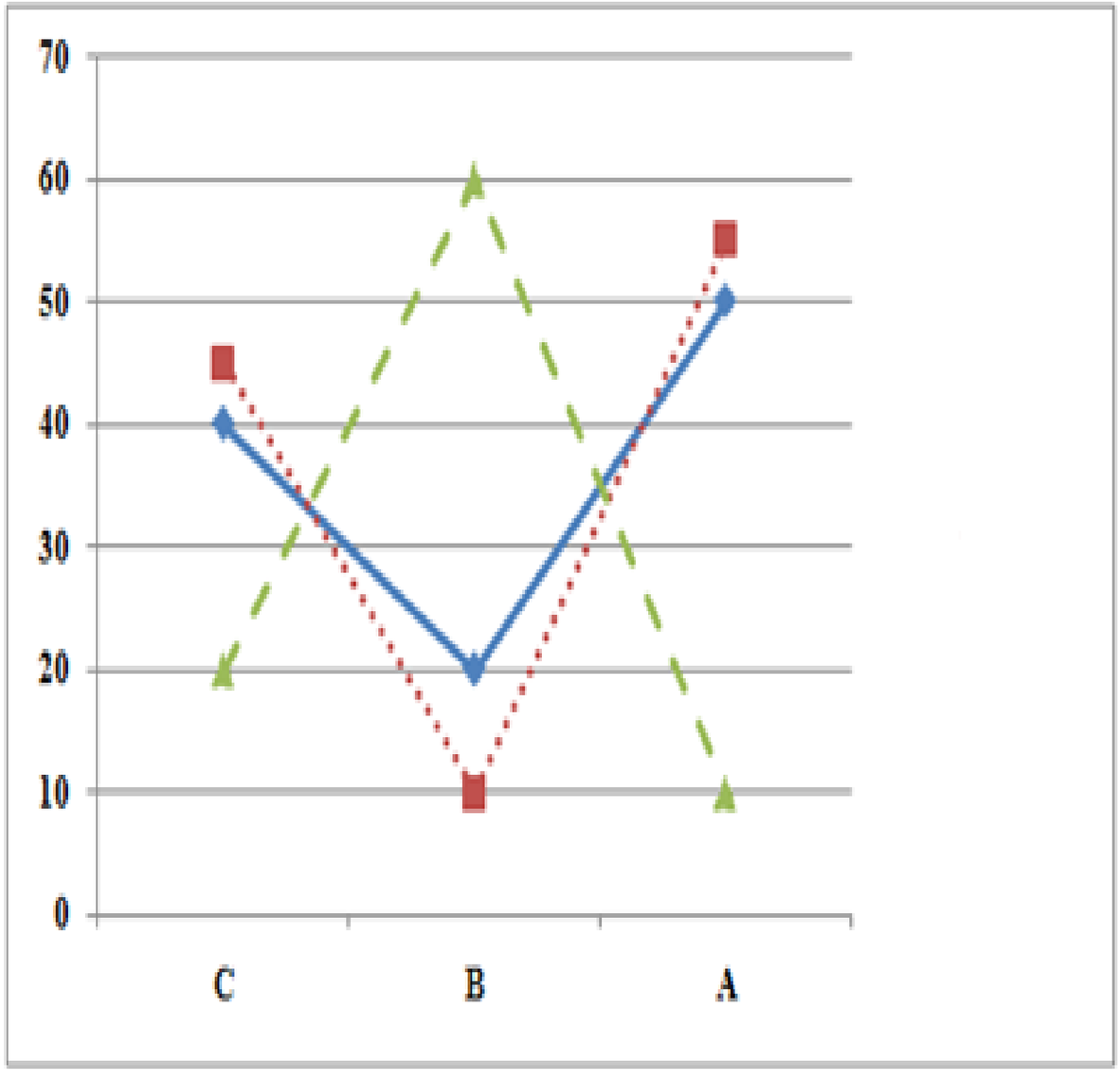
Bank	A’	B’	C’
Bank 1	50	10	50
Bank 2	50	0	70
Bank 3	50	20	30




80
70
60
50
40
30
20
10
0
B'
A'
C'
Bank 1
Bank 2
Bank 3


80
70
60
50
40
30
20
10
0
A'
C'
B'
Bank 1
Bank 2
Bank 3


70
60
50
40
30
20
10
0
A
B
C


70
60
50
40
30
20
10
0
C
B
A

METHODS

1. Ordinal-fixed pattern clustering;

2. Ordinal-invariant pattern clustering;

3. Diffusion-invariant pattern clustering;

4. Ordinal-interval pattern clustering;

5. Bayesian pattern analysis

ORDINAL-INVARIANT PATTERN CLUSTERING

$$q_i = \sum_{s=1}^{n-1} 10^{s-1} z_i^{n-s}$$

$$q_i^{dop} = \sum_{s=1}^{n-2} \sum_{j=s+2}^{n} 10^{j-(s+2)} e_i^{sj}$$

$$z_i^s = \begin{cases} 1, \text{if } x_{is} < x_{is+1} \\ 0, \text{if } x_{is} = x_{is+1} \\ 2, \text{if } x_{is} > x_{is+1} \end{cases}$$

$$e_i^{sj} = \begin{cases} 1, \text{if } x_{is} < x_{is+1} \\ 0, \text{if } x_{is} = x_{is+1} \\ 2, \text{if } x_{is} > x_{is+1} \end{cases}$$

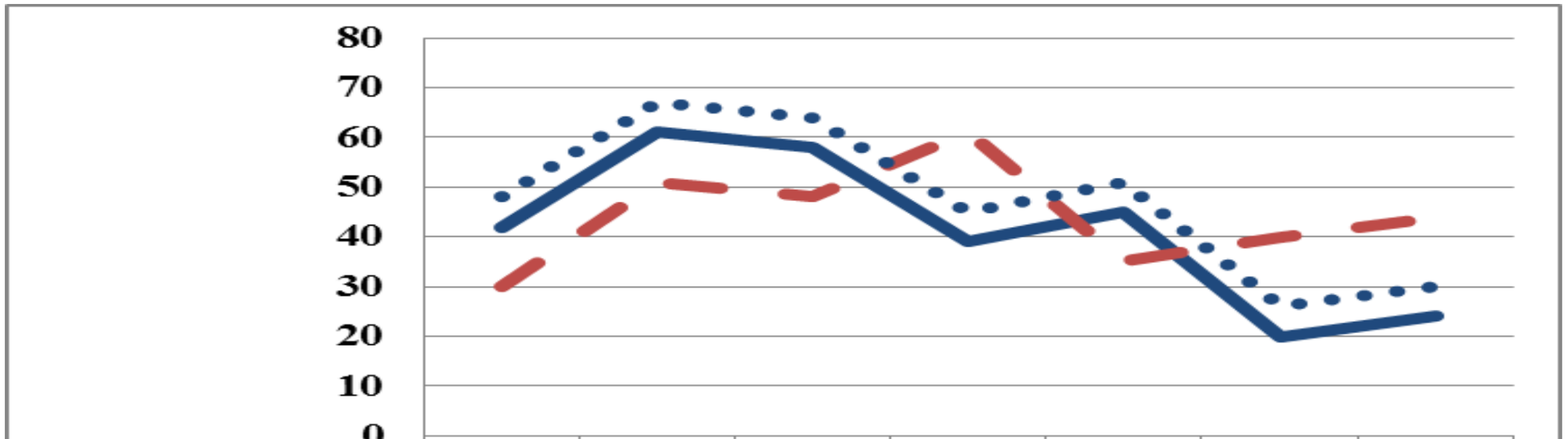
$$Z_{fix} = \frac{m^2(m-1)(n-1)}{2}$$

$$Z_{inv} = \frac{m^2 n(m-1)(n-1)}{4}$$

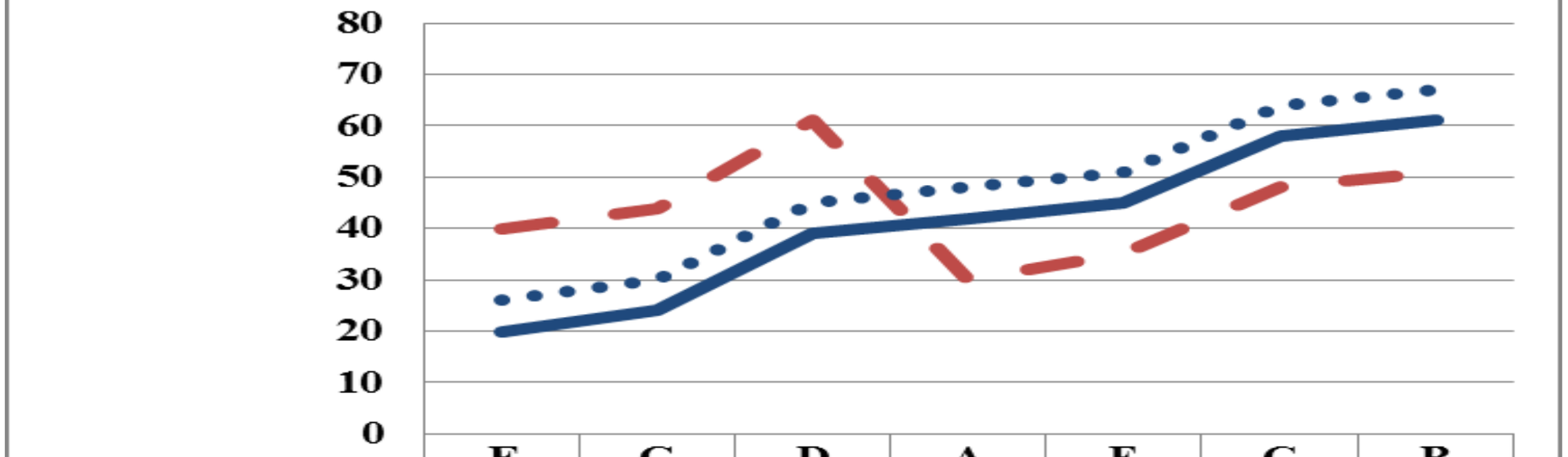
BASIC PROPERTIES

1. Clusters obtained using ordinal-invariant pattern clustering do not intersect.

2. If there exists a sequence of indicators in which their values form a strictly monotonically increasing / decreasing sequence for each object of the original set *X*, then this set is an ordinal-invariant pattern cluster.

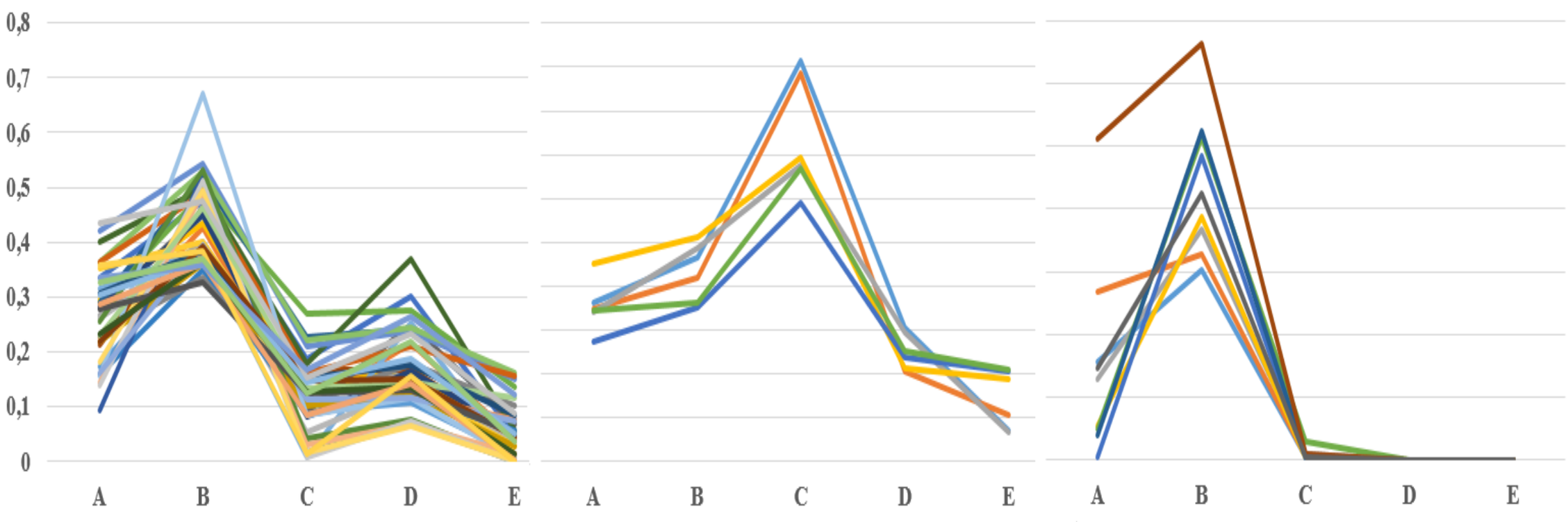


	A	B	C	D	E	F	G
Объект 1	42	61	58	39	45	20	24
Объект 2	30	51	48	61	35	40	44
Объект 3	48	67	64	45	51	26	30

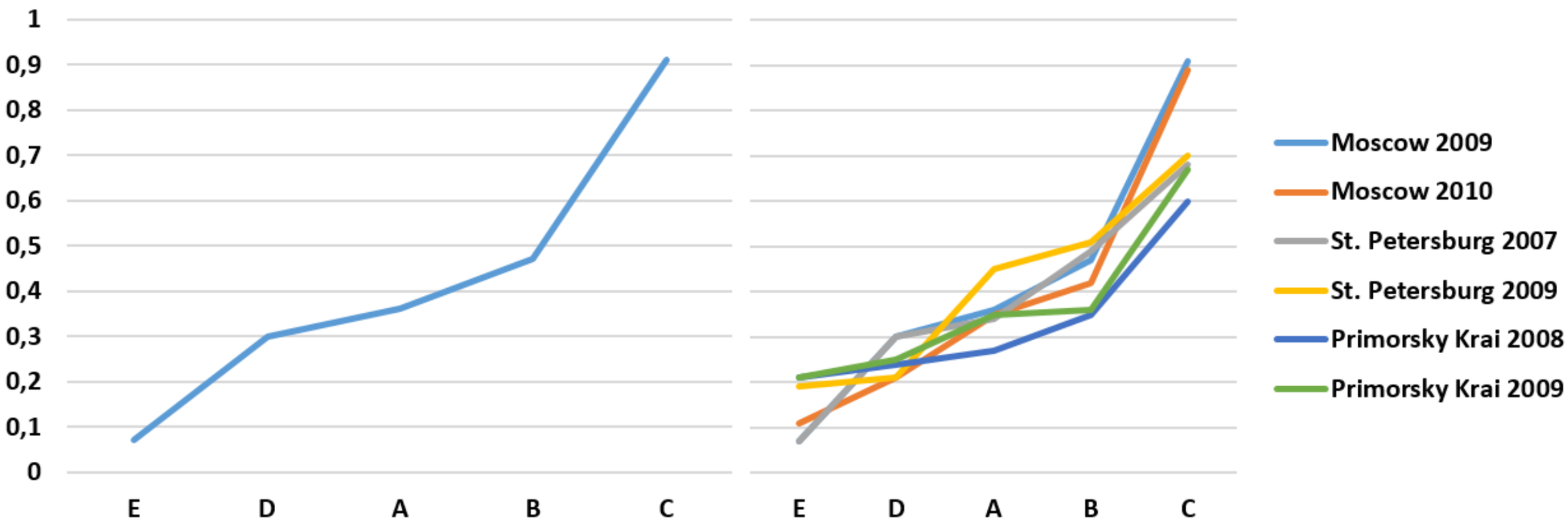


	F	G	D	A	E	C	B
Объект 1	20	24	39	42	45	58	61
Объект 2	40	44	61	30	35	48	51
Объект 3	26	30	45	48	51	64	67

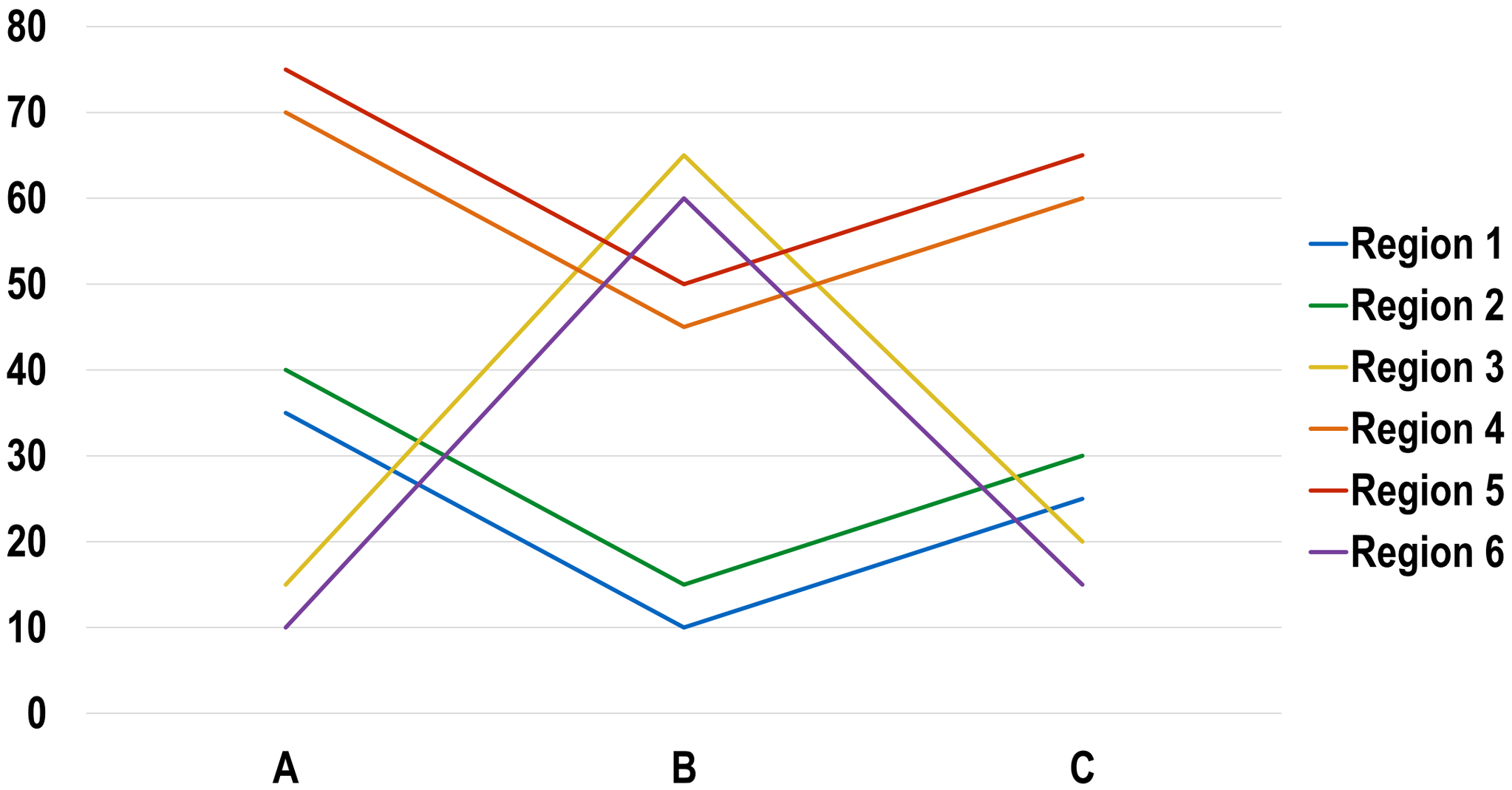
3. There is an order of the indicators for objects of ordinal-invariant pattern-cluster, in which their values are monotonous non decreasing/non increasing sequence for each object.

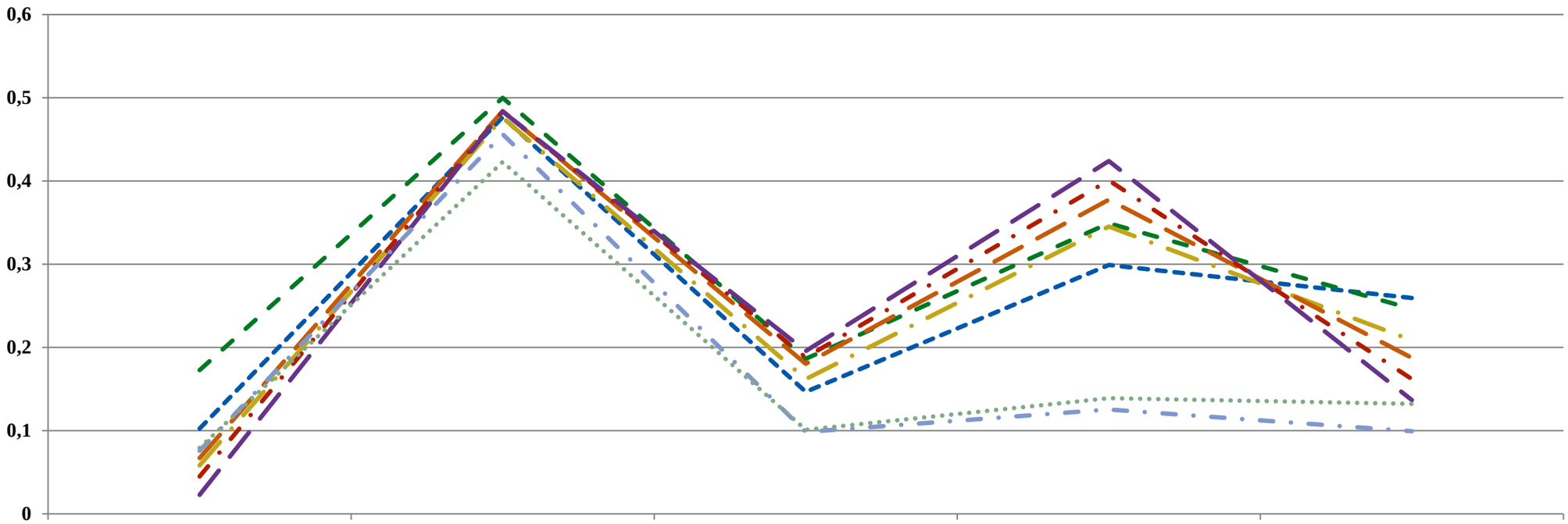


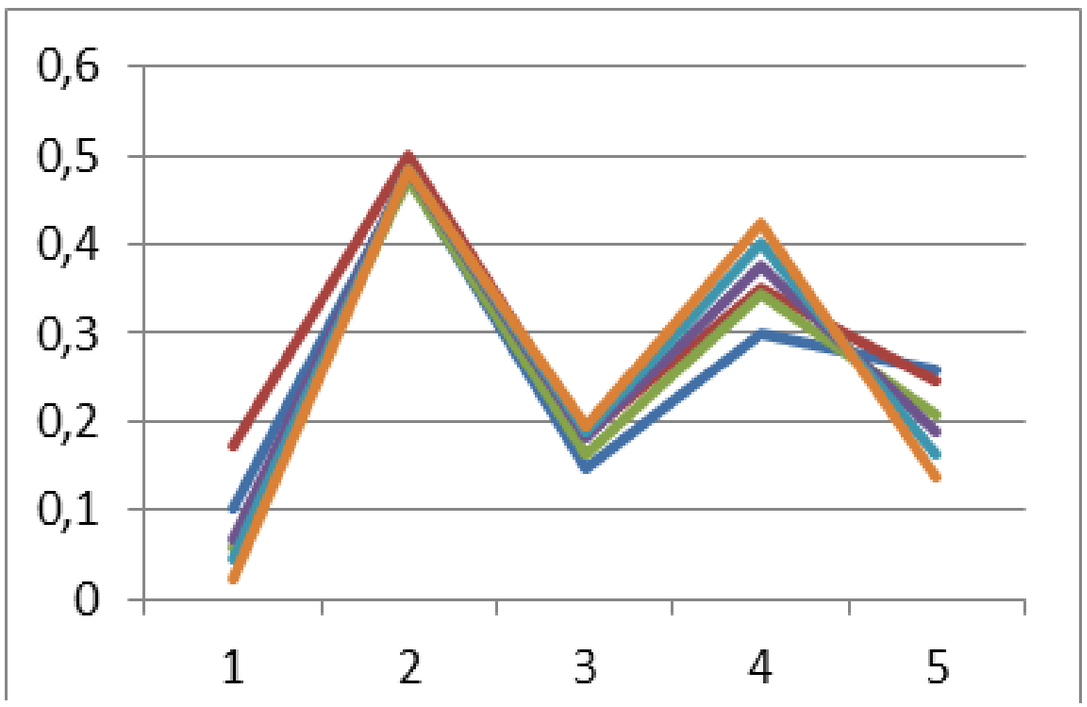
Region/year	A	B	C	D	E
Moscow 2009	0,36	0,47	0,91	0,3	0,07
Moscow 2010	0,35	0,42	0,89	0,21	0,11
St. Petersburg 2007	0,34	0,49	0,68	0,3	0,07
St. Petersburg 2009	0,45	0,51	0,7	0,21	0,19
Primorsky Krai 2008	0,27	0,35	0,6	0,24	0,21
Primorsky Krai 2009	0,35	0,36	0,67	0,25	0,21

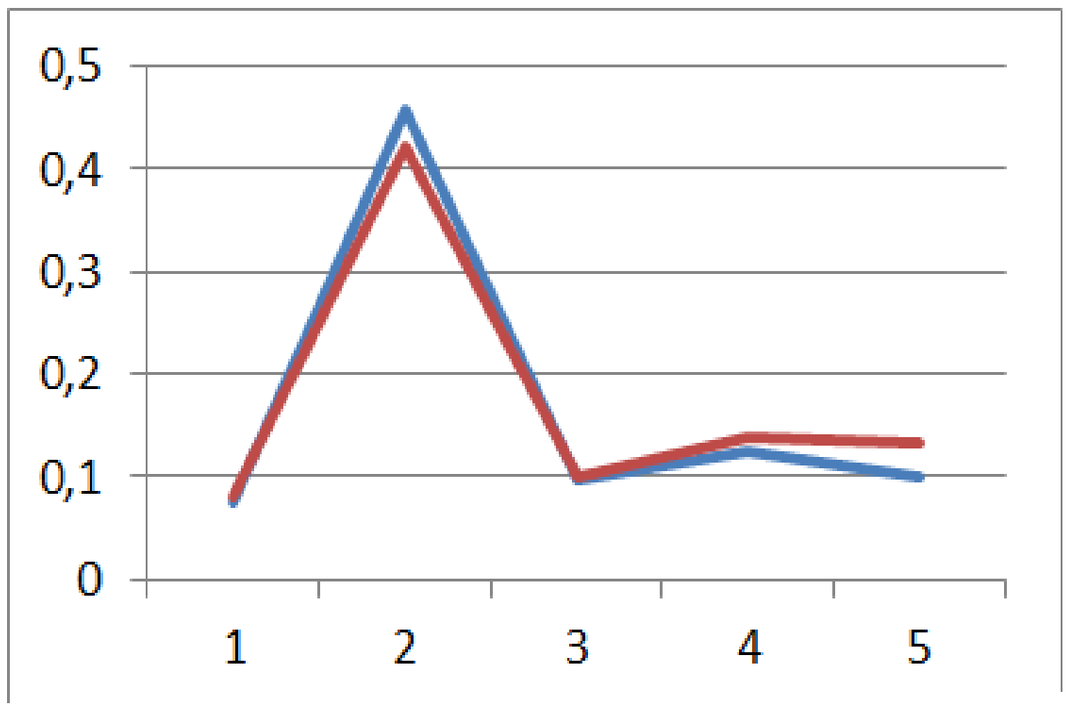


DIFFUSION-INVARIANT PATTERN CLUSTERING




0,6
0,5
0,4
0,3
0,2
0,1
0
1
2
3
4
5


0,6
0,5
0,4
0,3
0,2
0,1
0
1
2
3
4
5


0,5
0,4
0,3
0,2
0,1
0
1
2
3
4
5

Theorem. Let two objects x_1 и x_2, described by the vectors $x_1 = (x_{11}, x_{12}, \dots, x_{1m})$ and $x_2 = (x_{21}, x_{22}, \dots, x_{2m})$ respectively, belong to the same Euclidean cluster distances – $d(x_i, z) < R_{v_{dif}}$ between the object and the center of the cluster z, described by the vector $z = (z_1, z_2, \dots, z_m)$, where $R_{v_{dif}}$ is the cluster radius. We form new objects x_1^* and x_2^* by exchanging the corresponding values of one of the coordinates (for definiteness, the first) of the original objects x_1 and x_2 : $x_1^* = (x_{21}, x_{12}, \dots, x_{1m})$; $x_2^* = (x_{11}, x_{22}, \dots, x_{2m})$. Then:

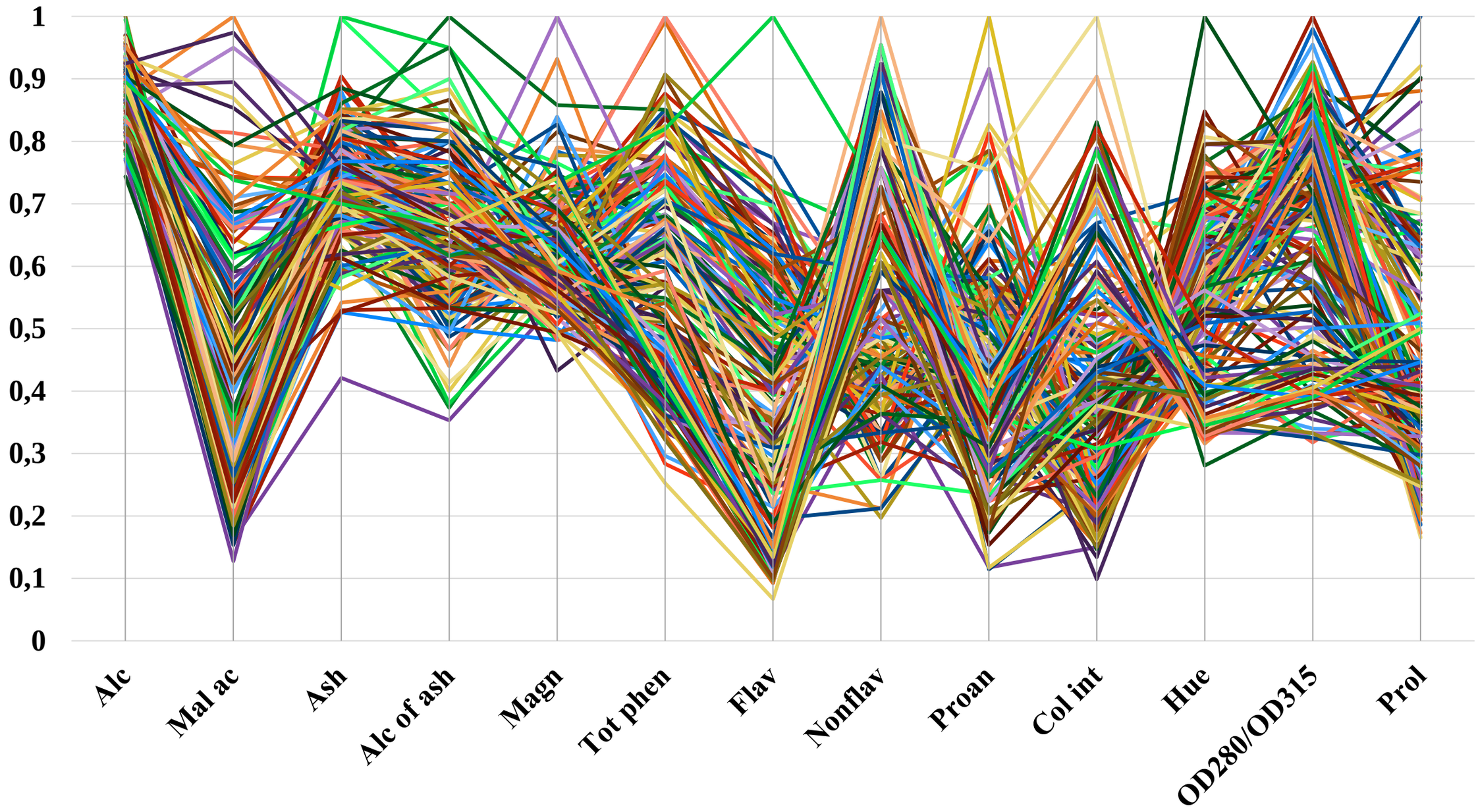
1) The Euclidean distances between the original objects $d(x_1, x_2)$ and between the newly formed objects $d(x_1^*, x_2^*)$ are equal.

2) At least one of the newly formed objects belongs to this cluster v_{dif}.

“In most cases, calculations are made with approximate numbers and, moreover, approximately. Therefore, even for the exact method of solving the problem, at each stage of the calculation, an error of actions and a rounding error arise... Under adverse circumstances, the total error can be so great that the result obtained will have only an illusory value” *

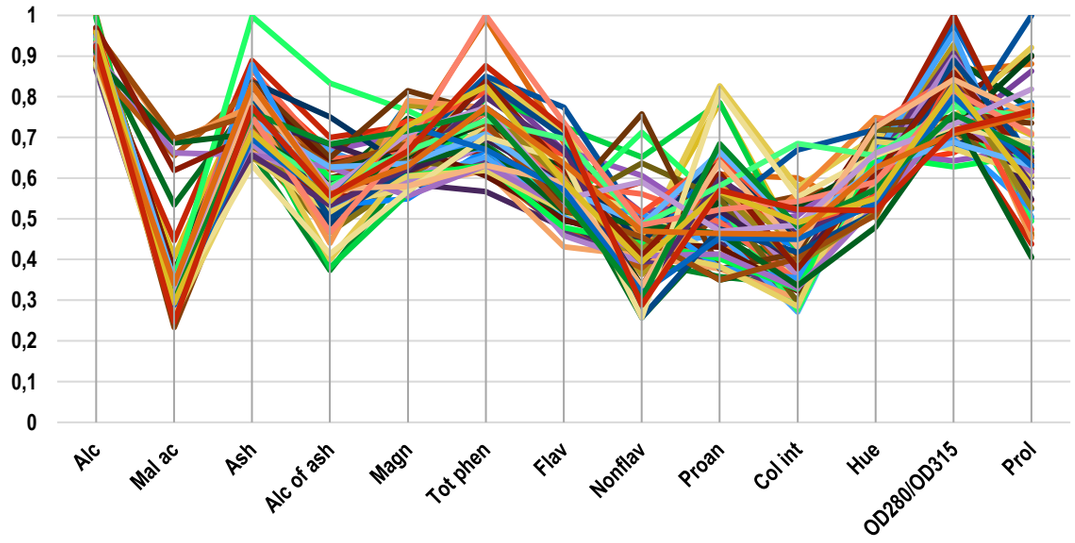
*DEMIDOVICH B.P., MARON I.A. Fundamentals of Computational Mathematics (in Russian)

WINE DATA SET

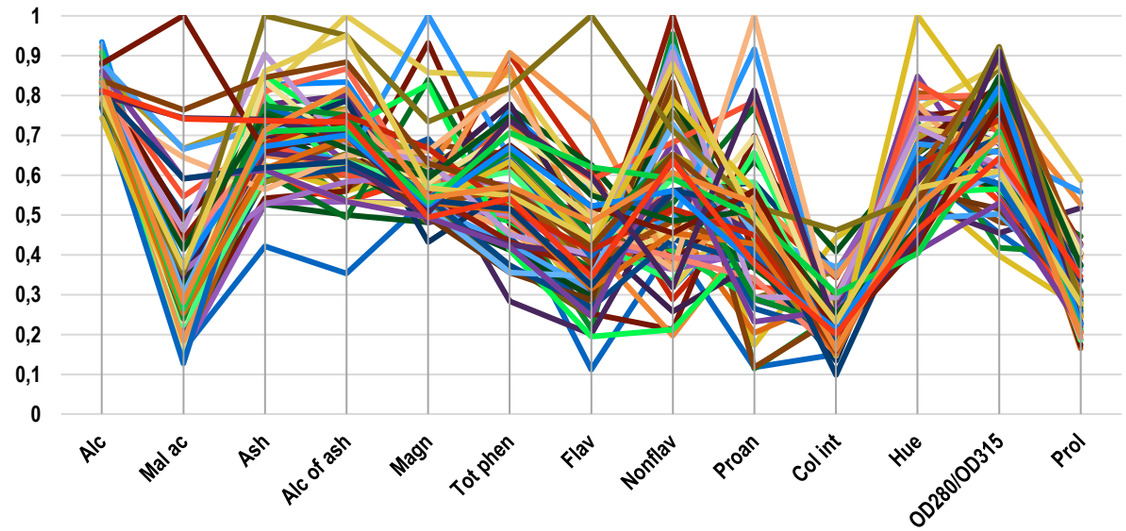


VINE DATA SET

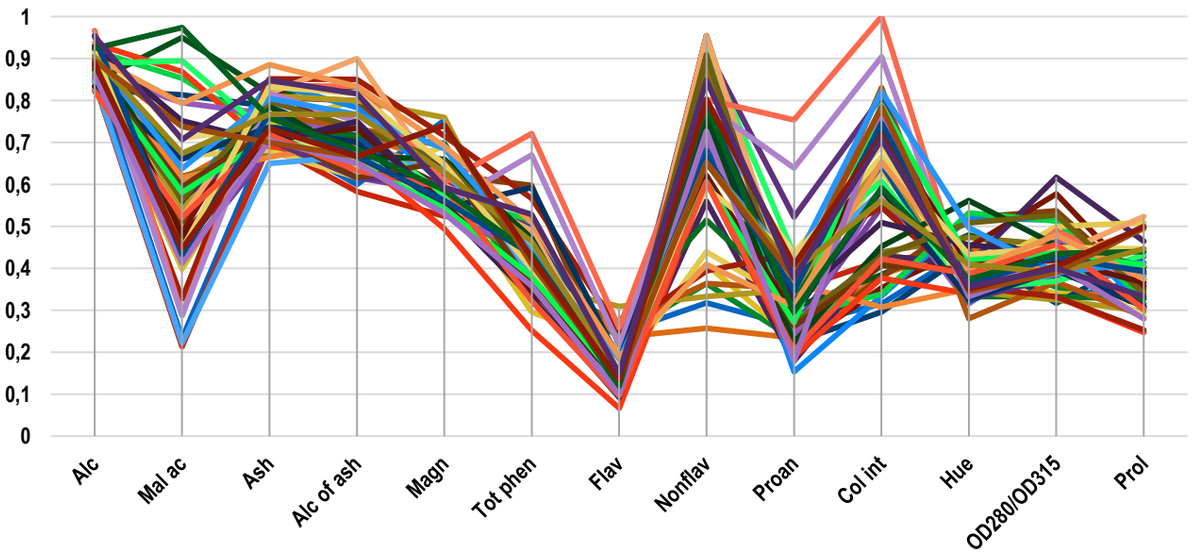
1

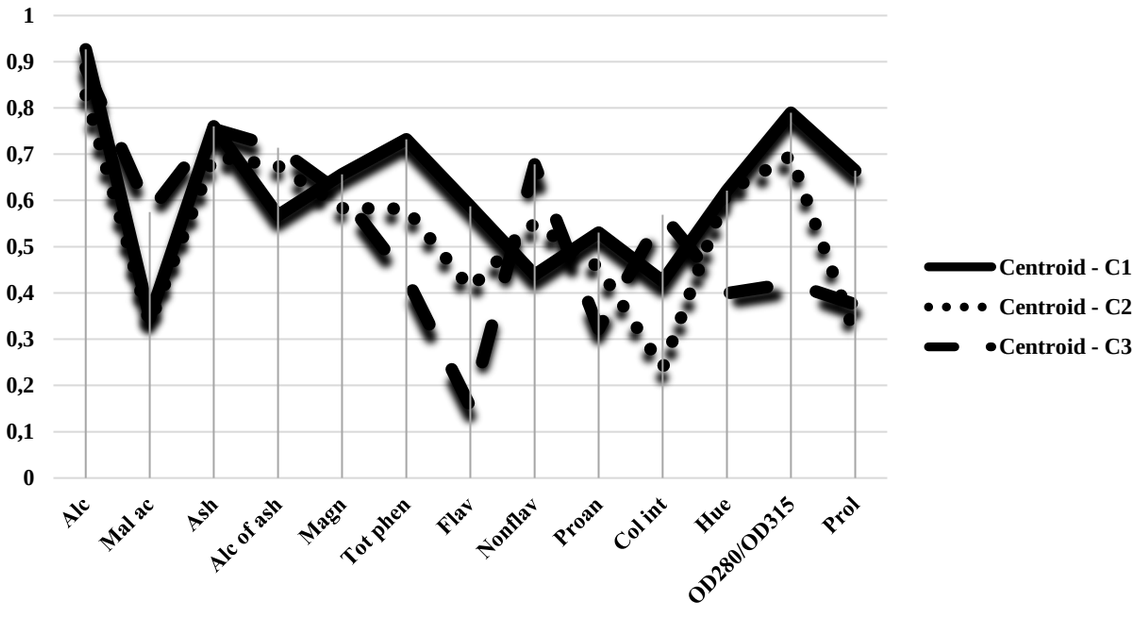


2

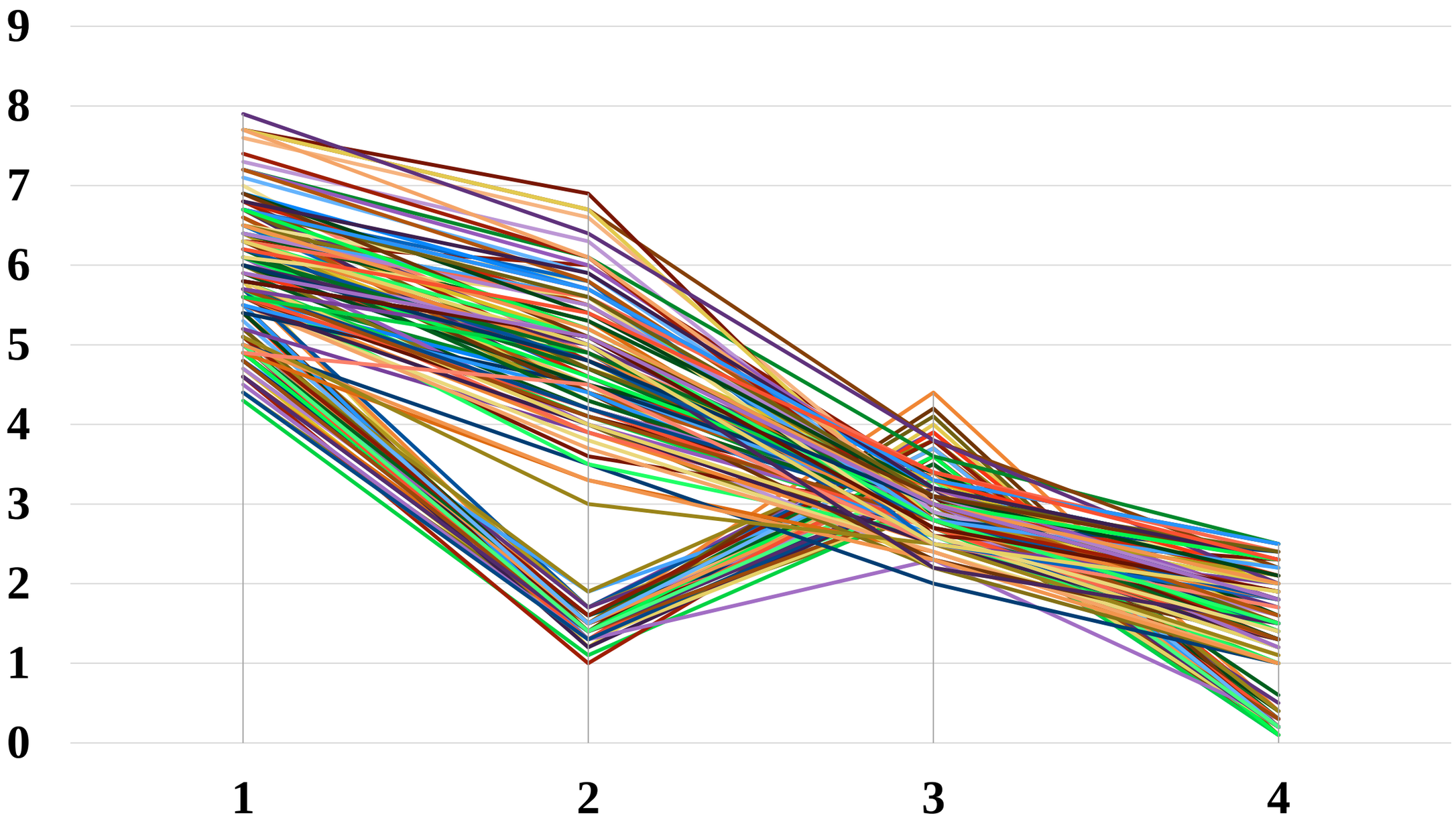


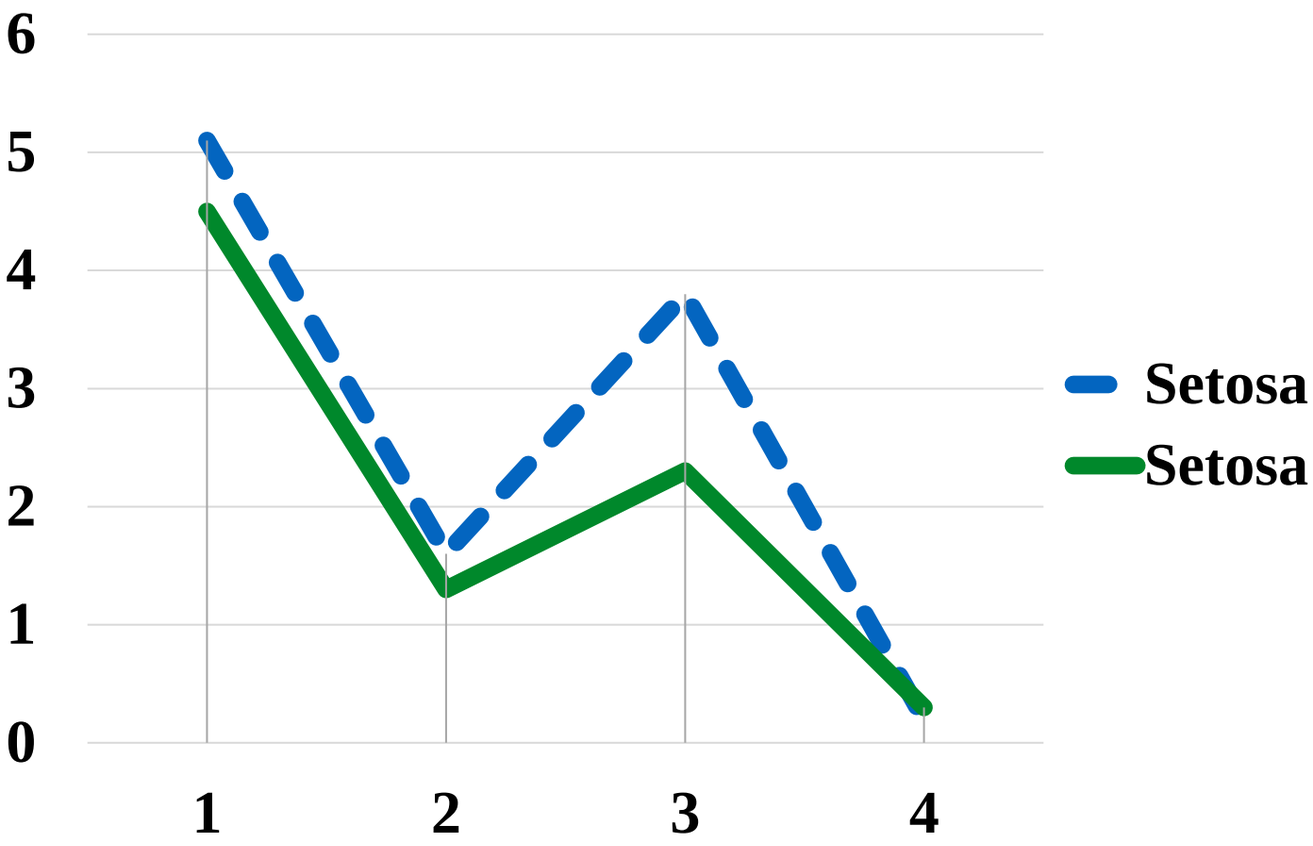
3

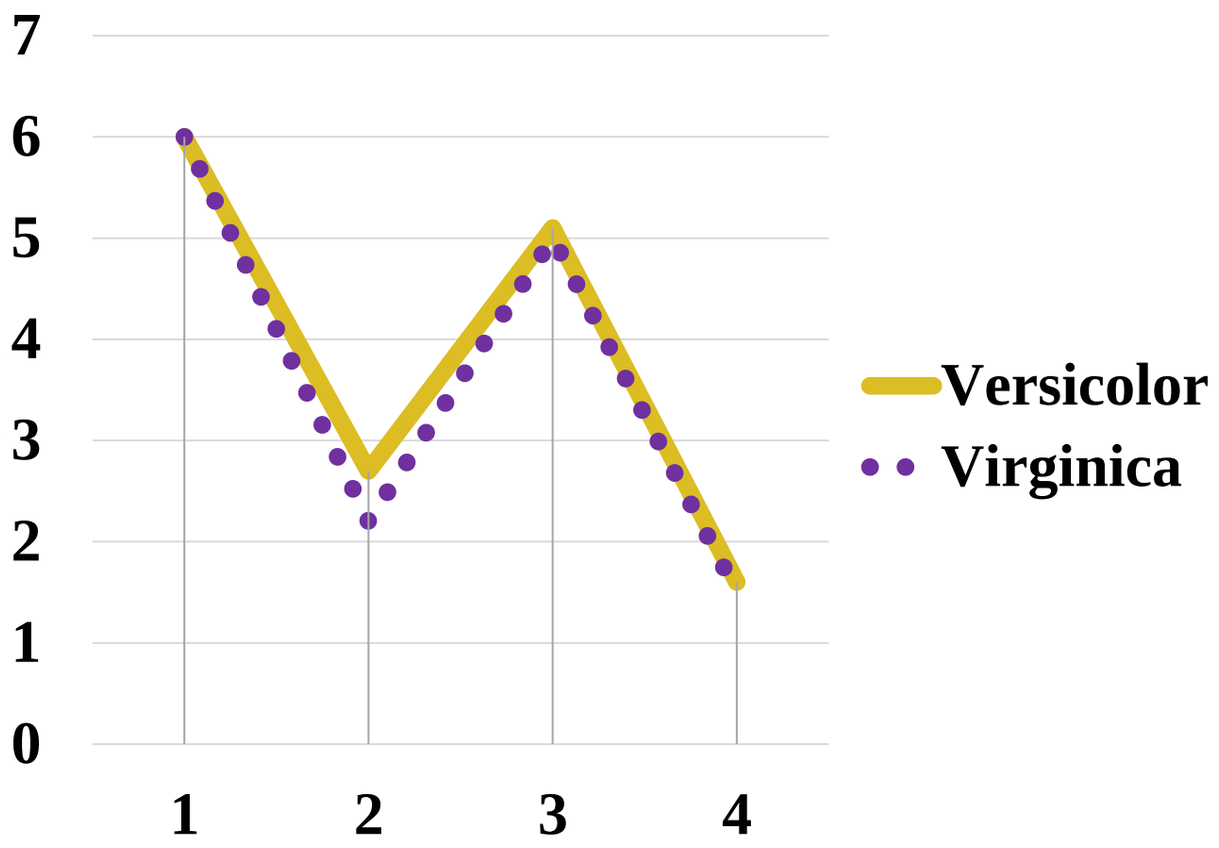




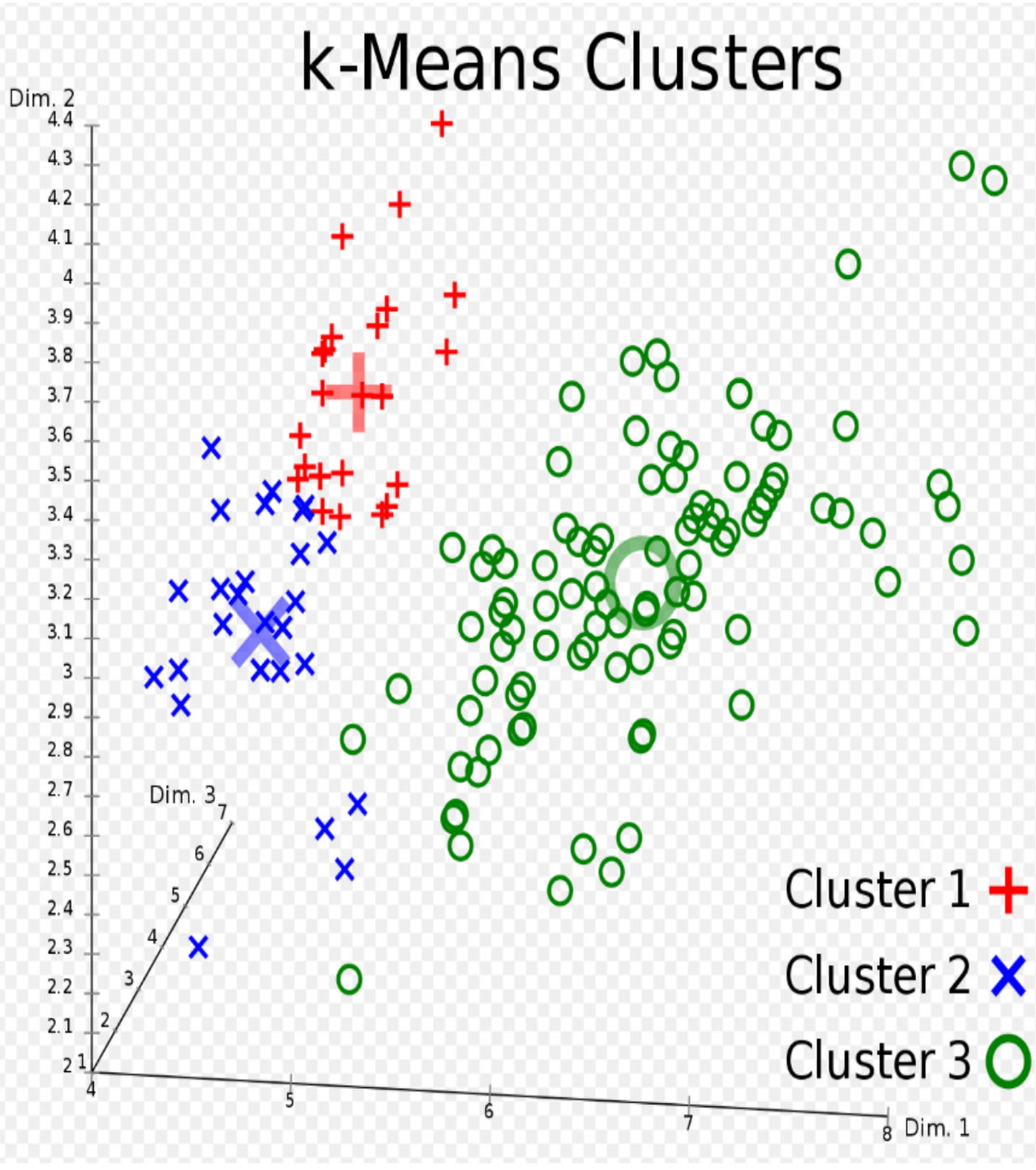
IRIS DATA

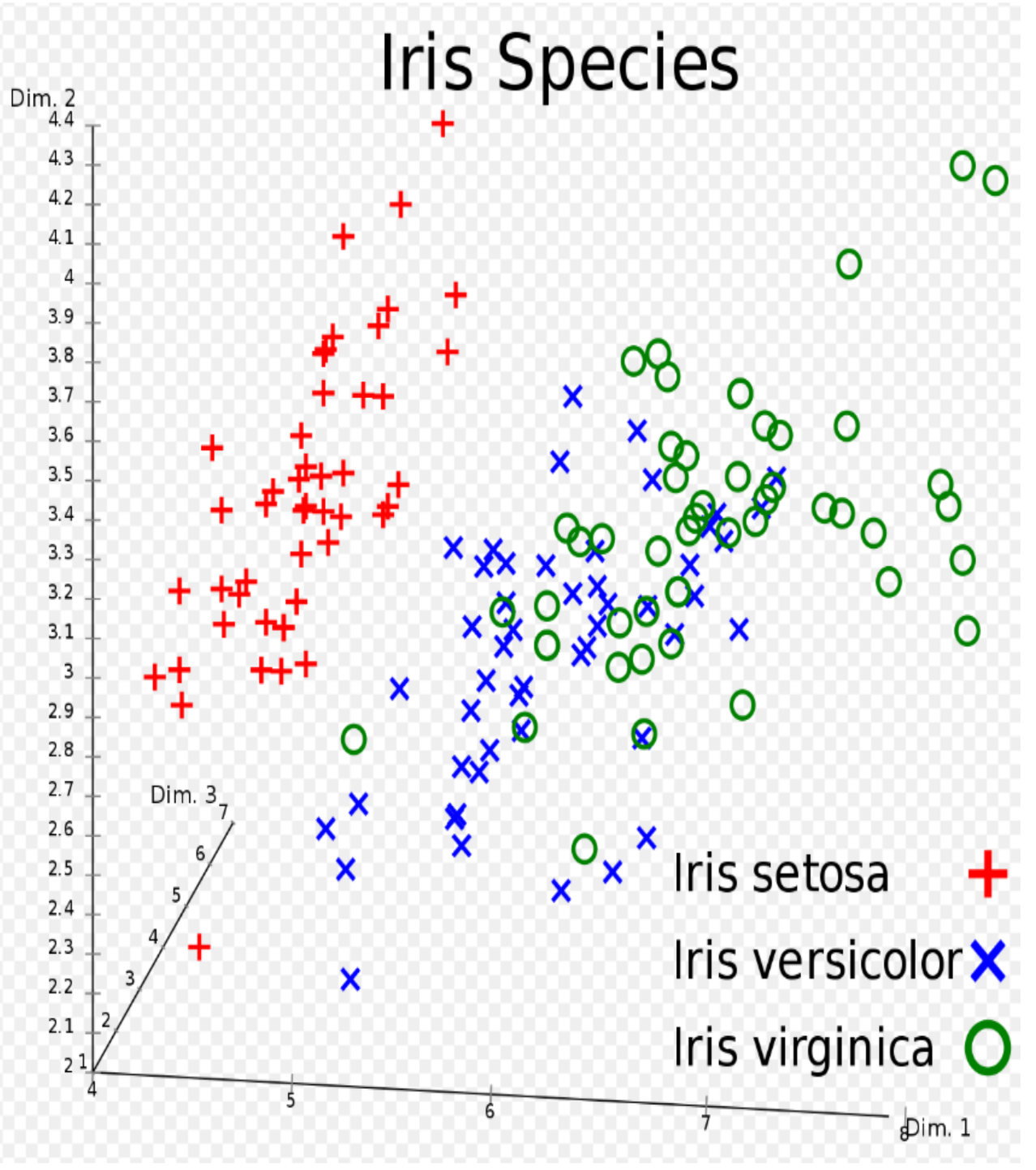


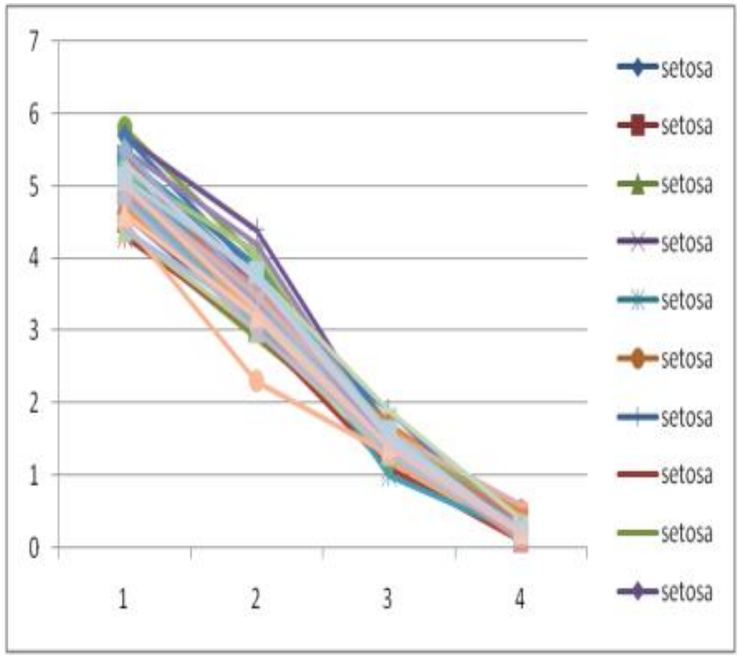
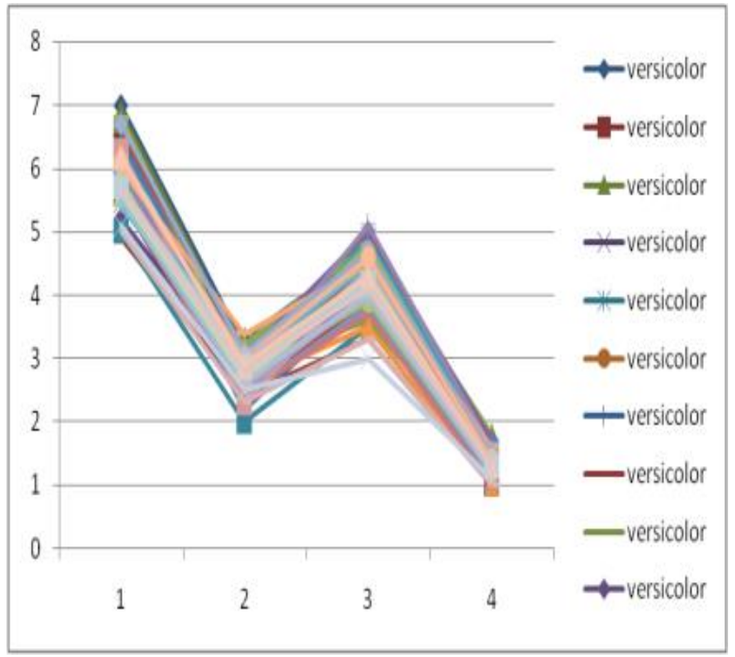
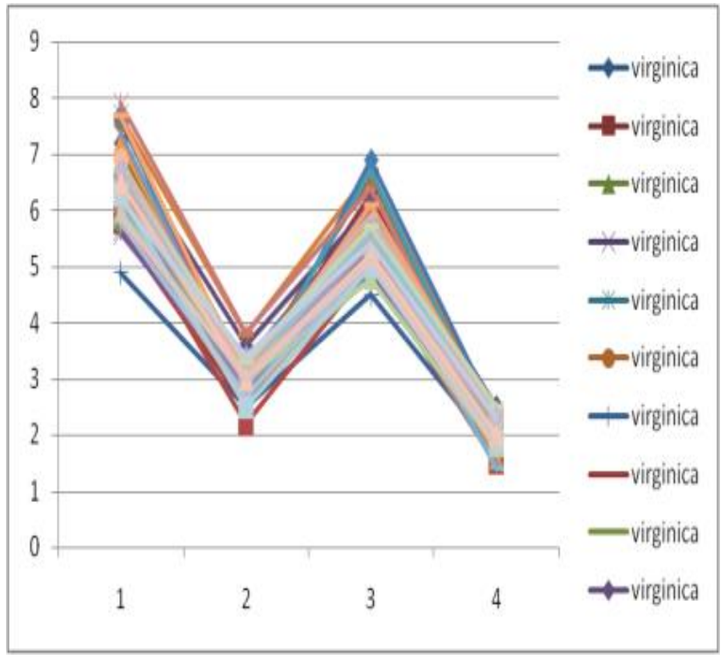

6
5
4
3
2
1
0
1
2
3
4
Setosa
Setosa


7
6
5
4
3
2
1
0
1
2
3
4
Versicolor
Virginica

K-means


k-Means Clusters
Dim. 2
Dim. 3
Dim. 1
Cluster 1
Cluster 2
Cluster 3


Iris Species
Dim. 2
Dim. 3
Dim. 1
Iris setosa
Iris versicolor
Iris virginica

Setosa	*Versicolor*	*Virginica*
		

IRIS	NUMBER OF IRIS	MISTAKES
Iris Setosa	50	0
Iris Versicolor	50	1
Iris Virginica	50	4

Balance Scale Data Set
Siegler (1976)

The “Balance Scale Data set” data set is modeled for psychological experiments and contains 625 lines. This kit is also often used to test cluster analysis methods. Each line reflects the values of five indicators (p_0; p_1; p_2; p_3; p_4), which characterize the indicators and the state of the lever weights. The indicator p_0 = {"B", "L", "R"} characterizes the state of the weights: "B" - balance; the deviation of the scale arrow to the left ("L" - left) or to the right ("R" - right).

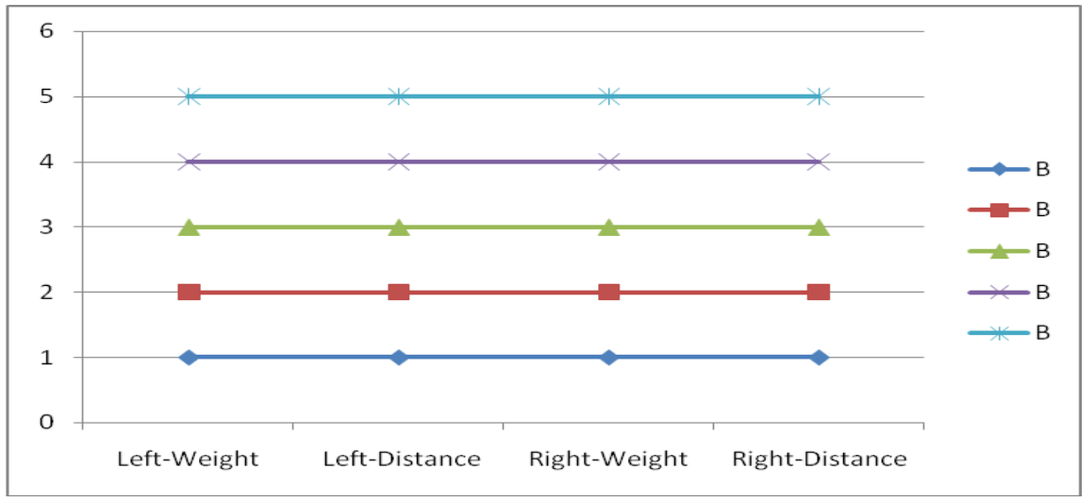
The remaining parameters p_1, p_2, p_3, and p_4 take integer values from one to five and characterize:

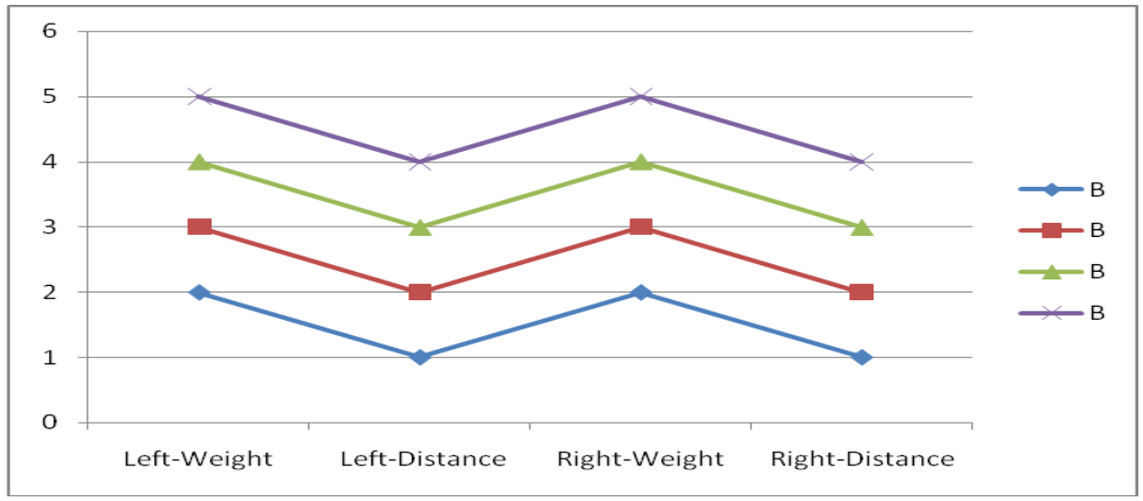
p_1 = {1, 2, 3, 4, 5} - weight of the left bowl of scales;

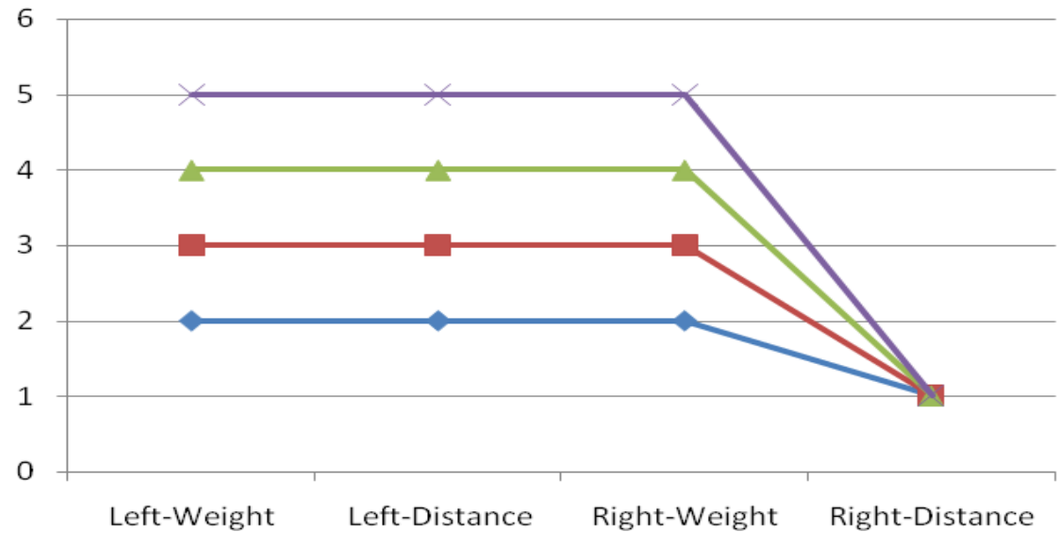
p_2 = {1, 2, 3, 4, 5} - the length of the lever on which the left bowl is suspended;

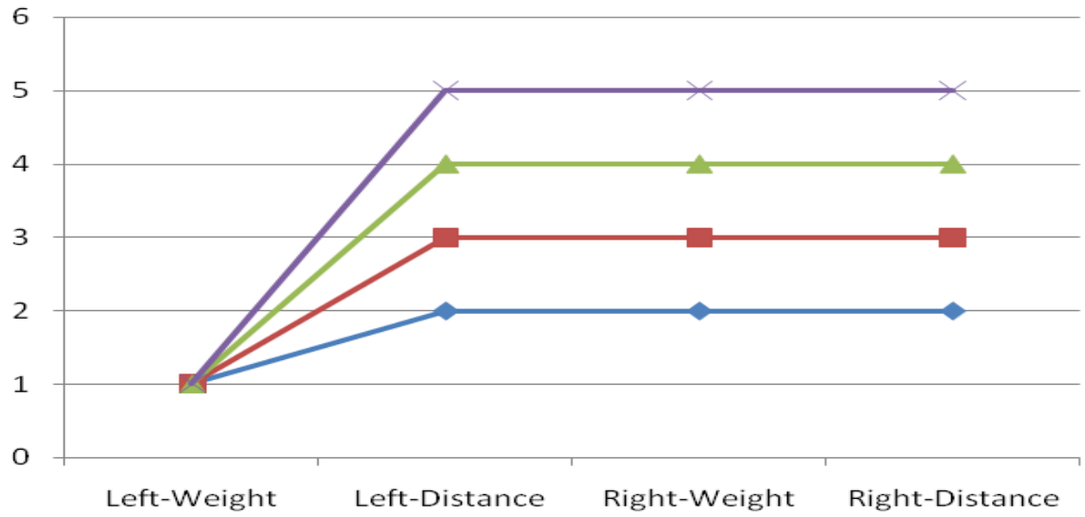
p_3 = {1, 2, 3, 4, 5} - weight of the right bowl of scales; and

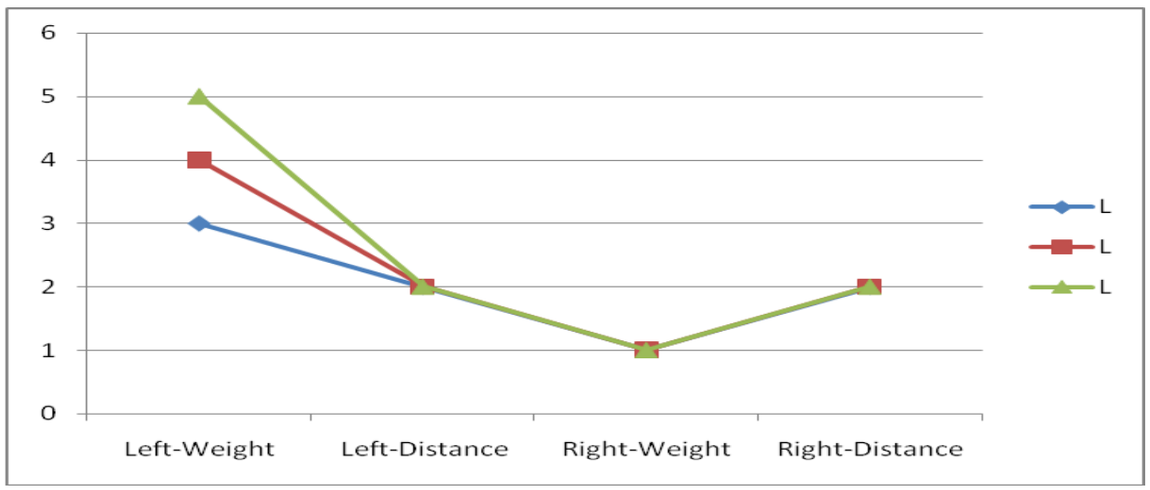
p_4 = {1, 2, 3, 4, 5} - the length of the lever on which the right bowl is suspended.

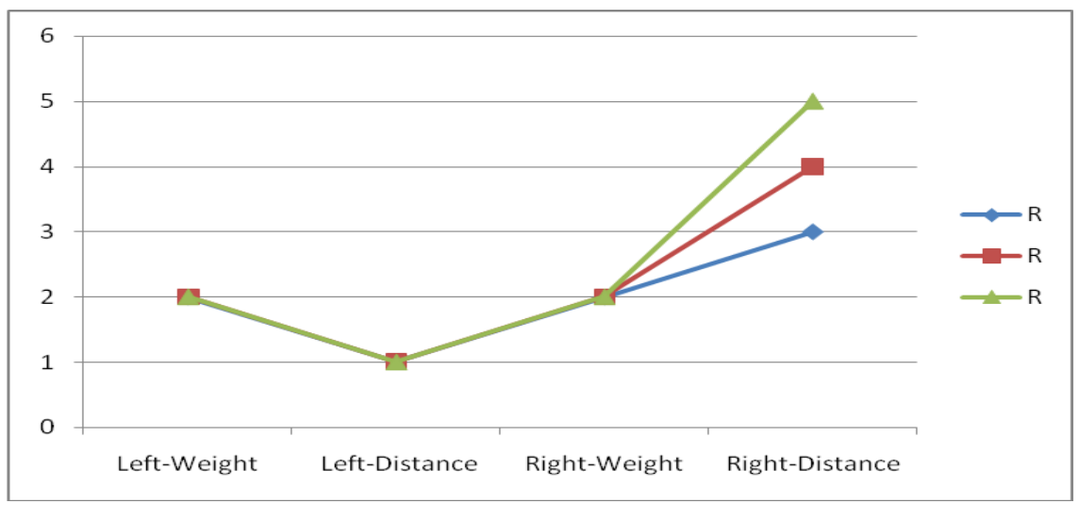

6
5
4
3
2
1
0
Left-Weight
Left-Distance
Right-Weight
Right-Distance
B
B
B
B
B


6
5
4
3
2
1
0
Left-Weight
Left-Distance
Right-Weight
Right-Distance
B
B
B
B


6
5
4
3
2
1
0
Left-Weight
Left-Distance
Right-Weight
Right-Distance


6
5
4
3
2
1
0
Left-Weight
Left-Distance
Right-Weight
Right-Distance


6
5
4
3
2
1
0
Left-Weight
Left-Distance
Right-Weight
Right-Distance
L
L
L


6
5
4
3
2
1
0
Left-Weight
Left-Distance
Right-Weight
Right-Distance
R
R
R

STATE CAPACITY

State Capacity - the ability of the state to perform basic functions (external security, internal order, legitimacy, good governance, ensuring conditions for development).

BASIC SYSTEM OF INDICATORS (1995-2005-2015)

UNSC_NY	Membership in the Security Council
IMF_vote	Share of votes in the IMF
Military_exp	Military expenditures
army	Number of military
Nuclear_Weapon	Availability of nuclear weapons
export_share	Share in world exports
GDP_share	Share in world GDP
R&D_worldshare	Share in world Research & Development
Territory	Share in world territory
POPUL_share	Share in world population
Oil_Gas	Share in world oil and gas reserves
University	Number of universities in the top-500 (Shanghai ranking)
Nobel	Number of Nobel Prizes
Non-resident patents	Number of applications for patents from non-residents

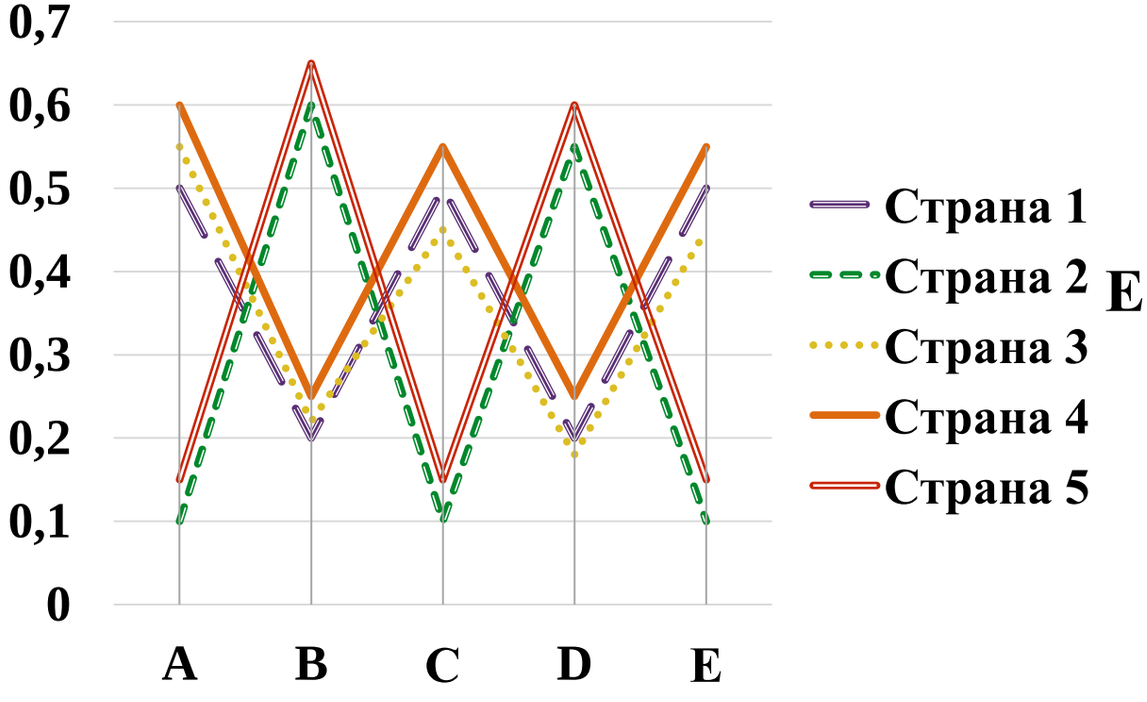
CORRELATION ANALYSIS (2005)

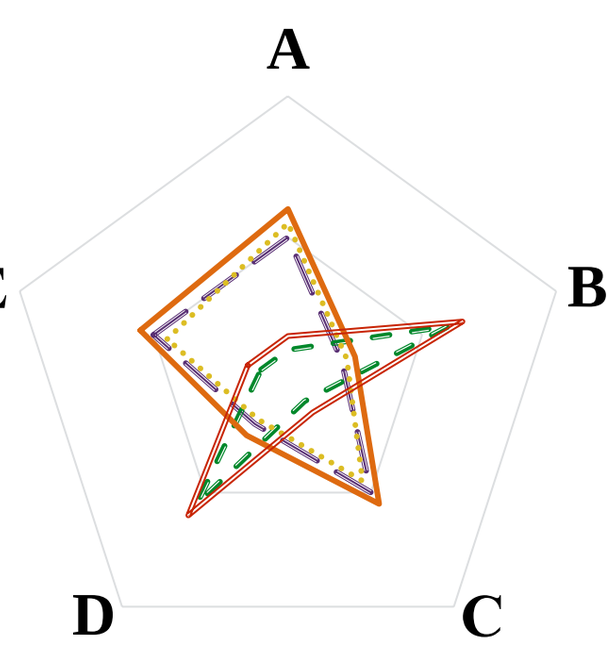
_	*UNSC_ NY2005*	*IMF_vot e2005*	*Military _exp20 05*	*army20 05*	*Nuclear _Weap ons200 5*	*export_ share20 05*	*GDP_sh are2005*	*R&D_w orldshar e2005*	*Territor y 2005*	*POPUL_ share20 05*	*Oil_Gas 2005*	*Universi ty2005*	*Nobel2 005*	*Non-resident patents 2005*
UNSC_NY2005	1													
IMF_vote2005	0,73	1												
Military_exp2005	0,59	0,9	1											
army2005	0,57	0,52	0,5	1										
Nuclear_Weapons200 5	0,86	0,66	0,64	0,68	1									
export_share2005	0,71	0,9	0,7	0,57	0,59	1								
GDP_share2005	0,66	0,97	0,95	0,52	0,63	0,84	1							
R&D_worldshare2005	0,58	0,93	0,93	0,45	0,56	0,79	0,98	1						
Territory 2005	0,62	0,48	0,41	0,6	0,63	0,44	0,43	0,36	1					
POPUL_share2005	0,39	0,32	0,25	0,81	0,5	0,41	0,31	0,23	0,46	1				
Oil_Gas2005	0,33	0,25	0,14	0,3	0,37	0,19	0,13	0,09	0,59	0,11	1			
University2005	0,63	0,95	0,96	0,47	0,62	0,83	0,98	0,96	0,4	0,25	0,1	1		
Nobel2005	0,66	0,91	0,94	0,4	0,65	0,76	0,92	0,89	0,34	0,16	0,1	0,96	1	
Non-resident patents 2005	0,57	0,85	0,9	0,65	0,61	0,75	0,92	0,91	0,54	0,48	0,14	0,9	0,79	1

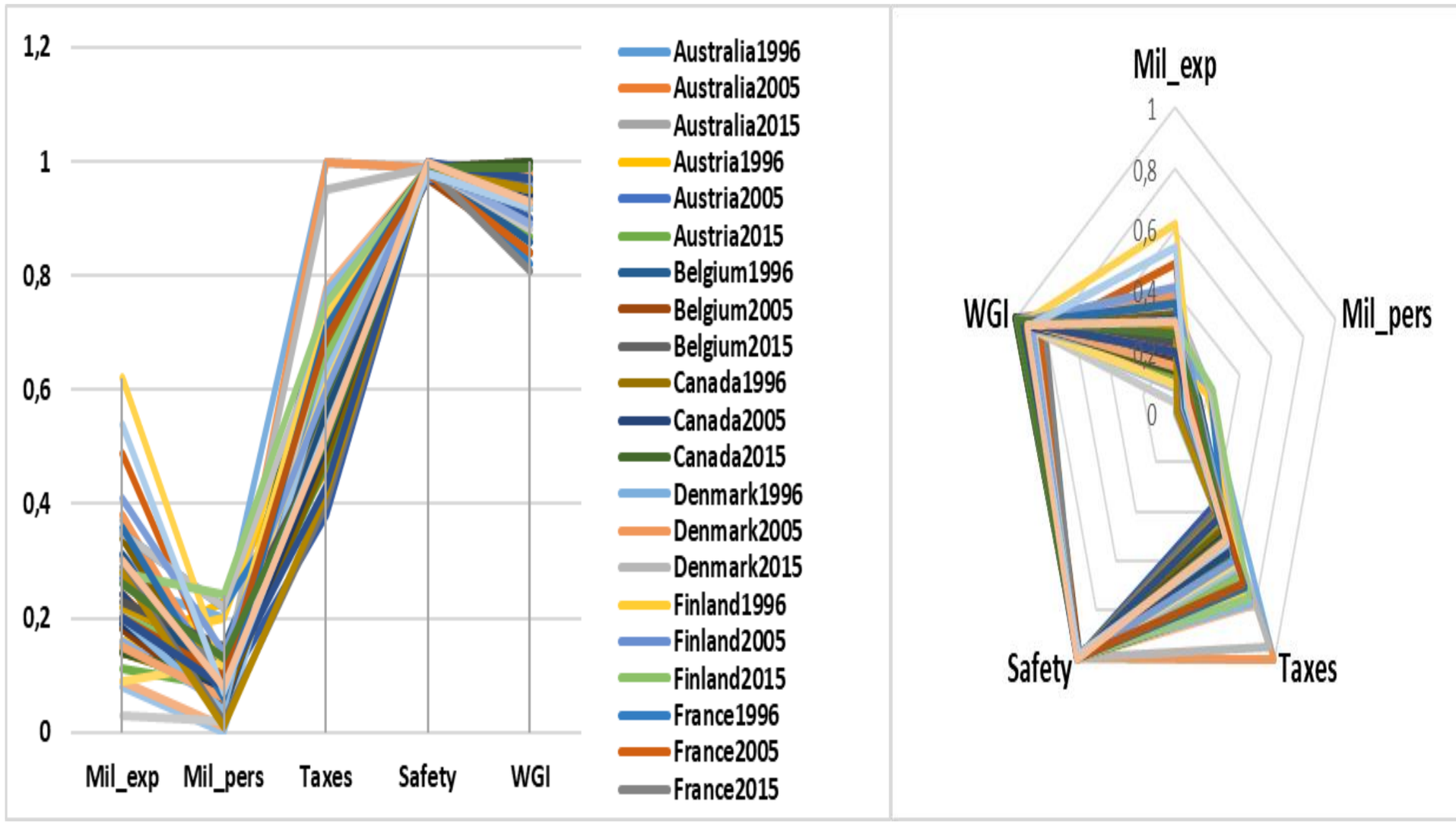
BASIC SYSTEM OF INDICATORS

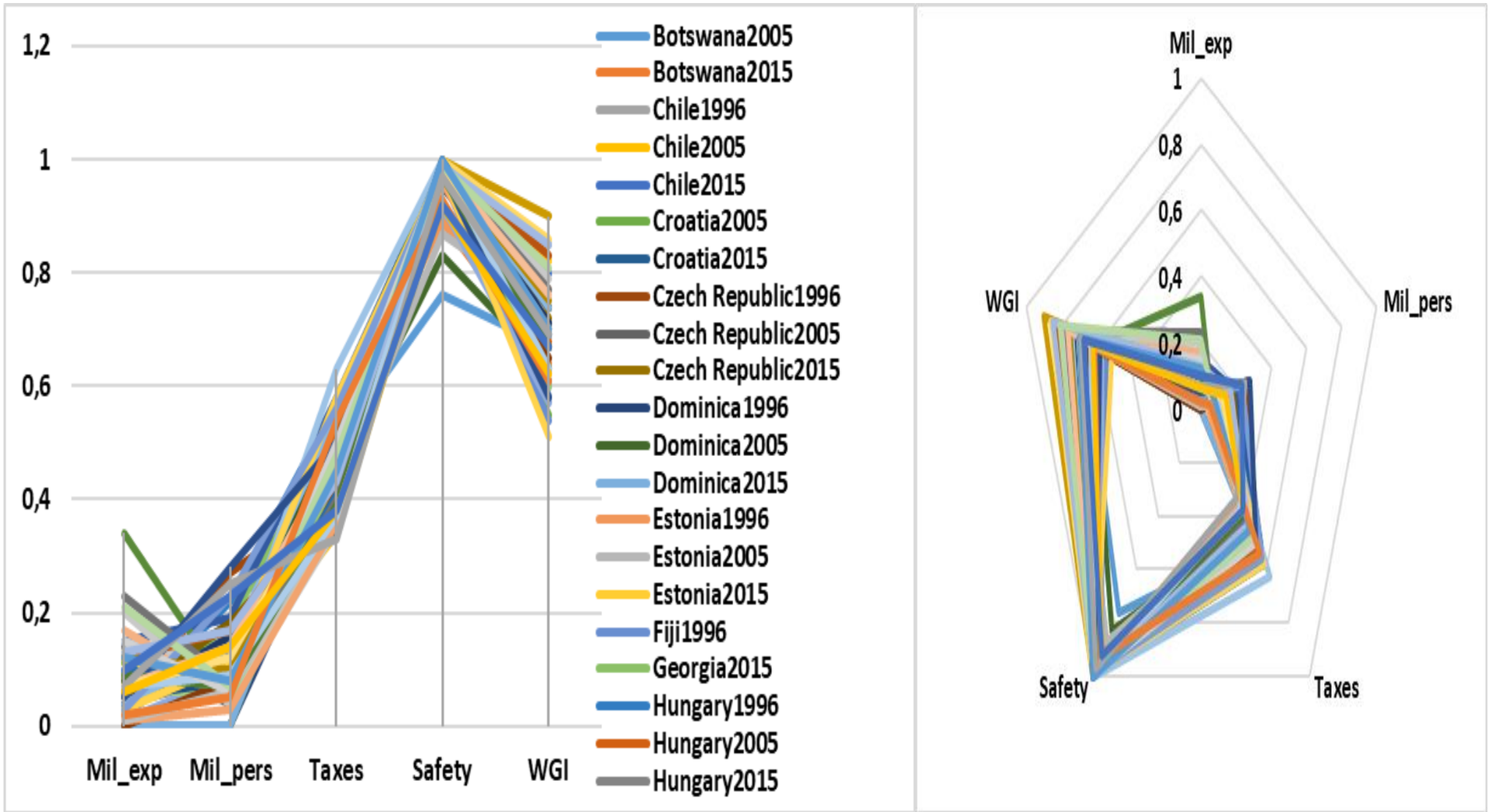
- *Mil_exp* – share of military spending in a country's GDP;
- *Mil_pers* – share of military personnel in the total population;
- *Taxes* – share of tax revenue in a country's GDP;
- *Safety* – parameter, calculated as the inverse of the total level of homicides and victims of conflicts;
- *WGI* – indicator characterizing the quality of state institutions.

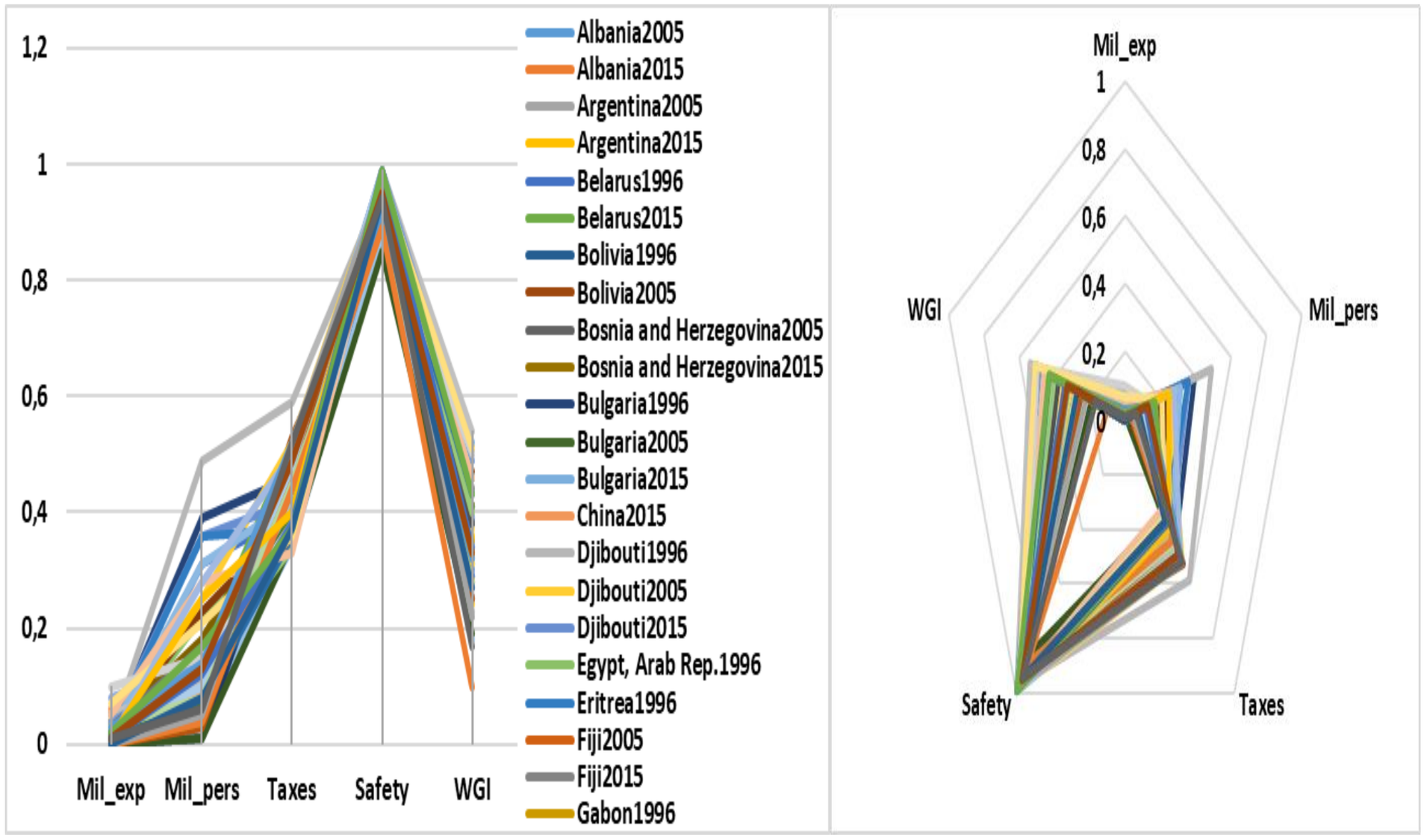
Country	*A*	*B*	*C*	*D*	*E*
Country 1	0,5	0,2	0,5	0,2	0,5
Country 2	0,1	0,6	0,1	0,55	0,1
Country 3	0,55	0,22	0,45	0,18	0,45
Country 4	0,6	0,25	0,55	0,25	0,55
Country 5	0,15	0,65	0,15	0,6	0,15

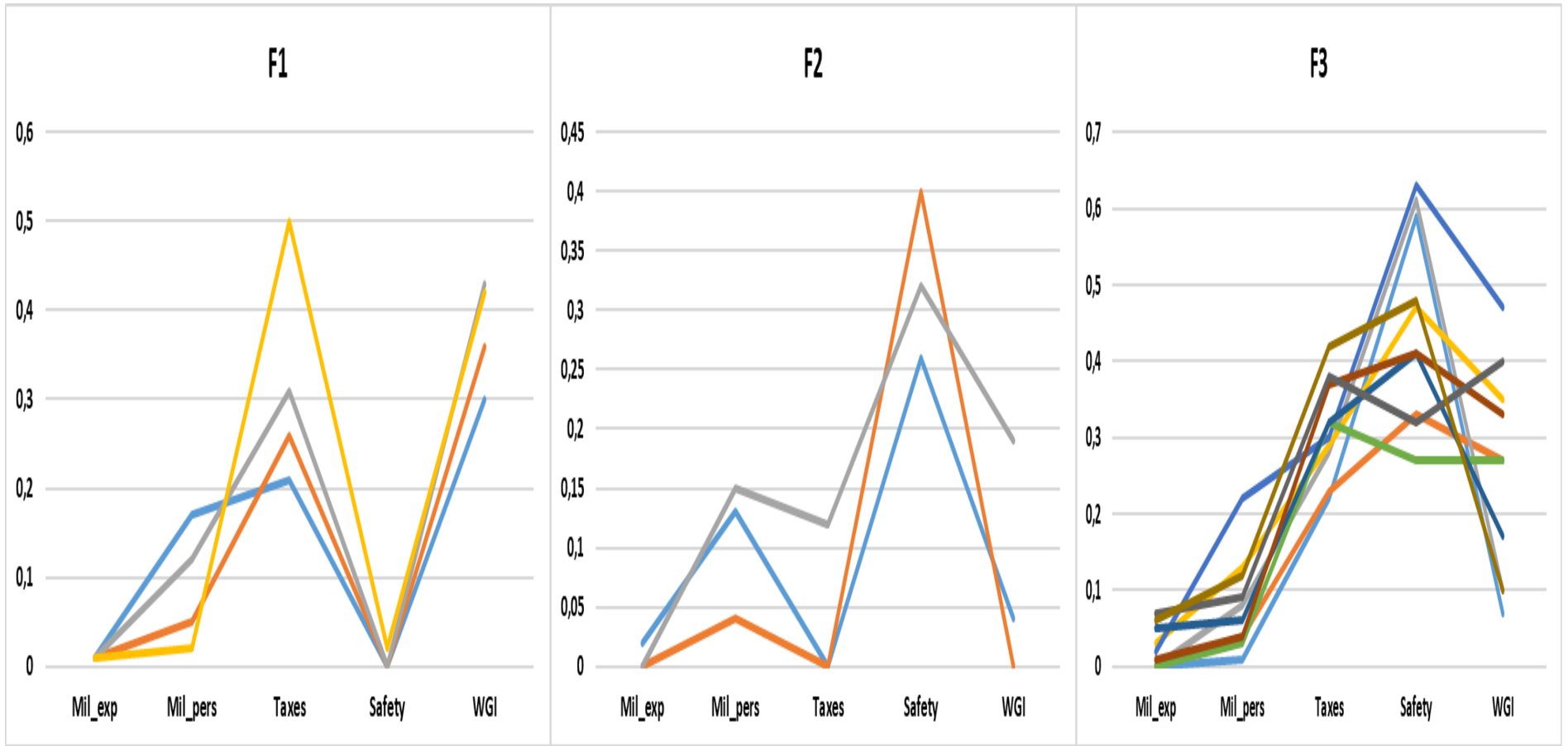





1,2
1
0,8
0,6
0,4
0,2
0
Mil_exp
Mil_pers
Taxes
Safety
WGI
Australia1996
Australia2005
Australia2015
Austria1996
Austria2005
Austria2015
Belgium1996
Belgium2005
Belgium2015
Canada1996
Canada2005
Canada2015
Denmark1996
Denmark2005
Denmark2015
Finland1996
Finland2005
Finland2015
France1996
France2005
France2015
Mil_exp
1
0,8
0,6
0,4
0,2
0
WGI
Mil_pers
Safety
Taxes


1,2
1
0,8
0,6
0,4
0,2
0
Mil_exp
Mil_pers
Taxes
Safety
WGI
Botswana2005
Botswana2015
Chile1996
Chile2005
Chile2015
Croatia2005
Croatia2015
Czech Republic1996
Czech Republic2005
Czech Republic2015
Dominica1996
Dominica2005
Dominica2015
Estonia1996
Estonia2005
Estonia2015
Fiji1996
Georgia2015
Hungary1996
Hungary2005
Hungary2015
Mil_exp
1
0,8
0,6
0,4
0,2
0
WGI
Mil_pers
Safety
Taxes


1,2
1
0,8
0,6
0,4
0,2
0
Mil_exp
Mil_pers
Taxes
Safety
WGI
Albania2005
Albania2015
Argentina2005
Argentina2015
Belarus1996
Belarus2015
Bolivia1996
Bolivia2005
Bosnia and Herzegovina2005
Bosnia and Herzegovina2015
Bulgaria1996
Bulgaria2005
Bulgaria2015
China2015
Djibouti1996
Djibouti2005
Djibouti2015
Egypt, Arab Rep.1996
Eritrea1996
Fiji2005
Fiji2015
Gabon1996
Mil_exp
1
0,8
0,6
0,4
0,2
0
WGI
Mil_pers
Safety
Taxes


F1
0,6
0,5
0,4
0,3
0,2
0,1
0
Mil_exp
Mil_pers
Taxes
Safety
WGI
F2
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
Mil_exp
Mil_pers
Taxes
Safety
WGI
F3
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
Mil_exp
Mil_pers
Taxes
Safety
WGI

РРИИ

- Социально-экономические условия инновационной деятельности регионов
  - Основные макроэкономические показатели
    - ВРП к численности занятых
    - ВРП к стоимости основных фон-дов
    - Уд. вес занятых в высокотехнологичных и среднетехнологичных отраслях
    - Уд. вес занятых в наукоемких отраслях сферы услуг
  - Образовательный потенциал населения
    - Уд. вес населения в возрасте 25-64 лет с высшим и послевузовским профессиональным образованием
    - Числен-ность студентов вузов в расчете на 10 000 человек населения
  - Развитие информационного общества
    - Уд. вес организаций, использующих широкополосный доступ к Интернету
    - Уд. вес населения, имеющего доступ к Интернету в домашних хозяйствах
- Научно-технический потенциал регионов
  - Исследования и разработки: финансирование
    - Внутренние затраты на ИР к ВРП
    - Внутренние затраты на ИР в расчете на 1 исследователя
    - Уд. вес средств организаций предпринимательского сектора
  - Исследования и разработки: кадры
    - Уд. вес среднемесячной заработной платы работников, занятых ИР
    - Уд. вес персонала, занятого ИР в расчете до 39 лет
    - Уд. вес лиц в возрасте до 39 лет, имеющую ученую степень
    - Уд. вес исследователей, имеющих ученую степень
  - Публикационная активность
    - Число статей в РИНЦ на 10 исследователей
  - Патентная активность
    - Патентные заявки на изобретения, поданные в Роспатент на 1 млн населения
    - Число изобретений, созданных в регионе, в расчете на 1 млн населения
  - Создание технологий
    - Число технологий,
  - Торговля технологиями
    - Объем экспорта к ВРП
- Качество инновационной политики регионов
  - Затраты консолидированного бюджета
    - Уд. вес средств бюджетов субъектов РФ и местных бюджетов в объеме внутренних затрат на ИР
    - Уд. вес средств бюджетов субъектов РФ и местных бюджетов в затратах на технологичес-кие инновации
  - Организационное обеспечение инновационной политики
    - Наличие региональных институтов развития
    - Наличие координационных органов по инновационной политике
  - Нормативно-правовая база инновационной политики
    - Наличие программы государственной поддержки развития инноваций
    - Наличие правовой базы государственной поддержки ИД
    - Наличие стратегии (концепции) инновационного развития
    - Выделение зон приоритетного развития ИД
    - Наличие раздела по инновационному развитию в стратегии развития региона
- Инновационная деятельность в регионах
  - Результативность инновационной деятельности
    - Уд. вес организаций, оценивших сокращение материаль-ных и энергозатрат как основной результат ИД
    - Уд. вес инновационных товаров, работ, услуг новых для рынка
    - Уд. вес инновационных товаров работ, услуг
  - Затраты на технологические инновации
    - Интенсивность затрат на технологические инновации
  - Малый инновационный бизнес
    - Уд. вес малых предприятий, осуществлявших технологические инновации
    - Уд. вес организаций, имевших готовые технологические инновации
  - Активность в сфере технологических и нетехнологических инноваций
    - Уд. вес организаций, осуществлявших технологические инновации
    - Уд. вес организаций, осуществлявших нетехнологические инновации

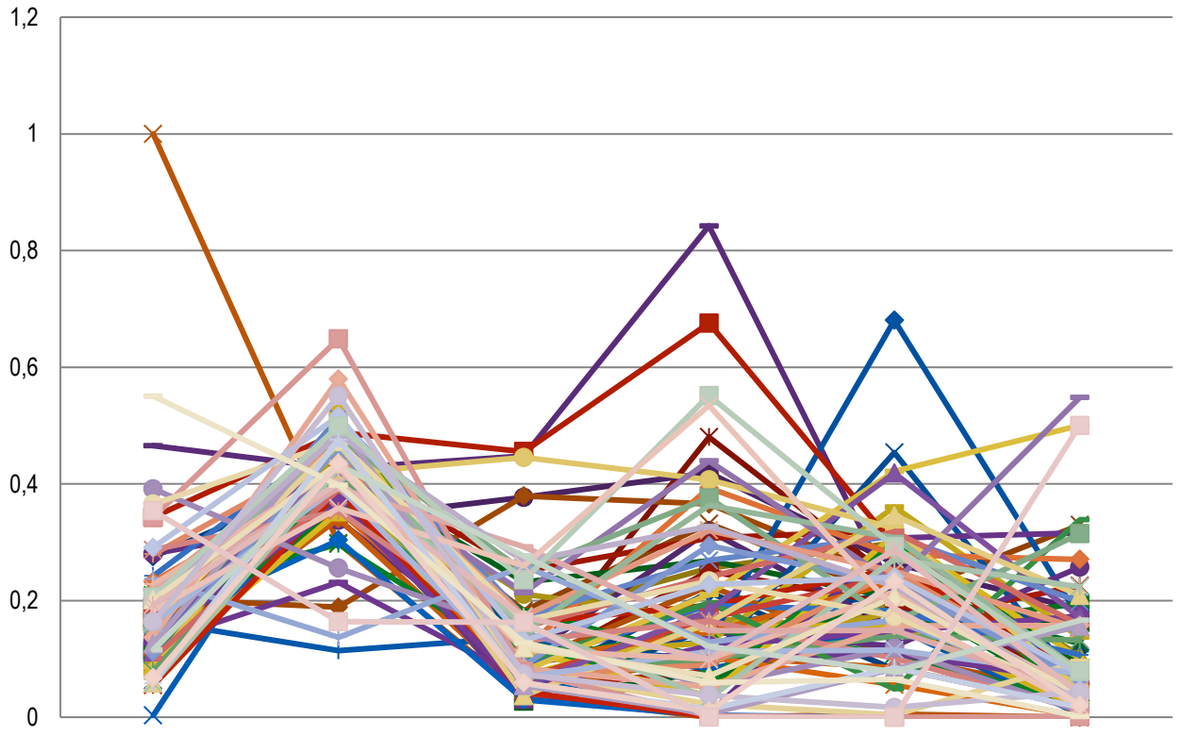
Высшая школа экономики, Москва, 2013

Basic system of indicators

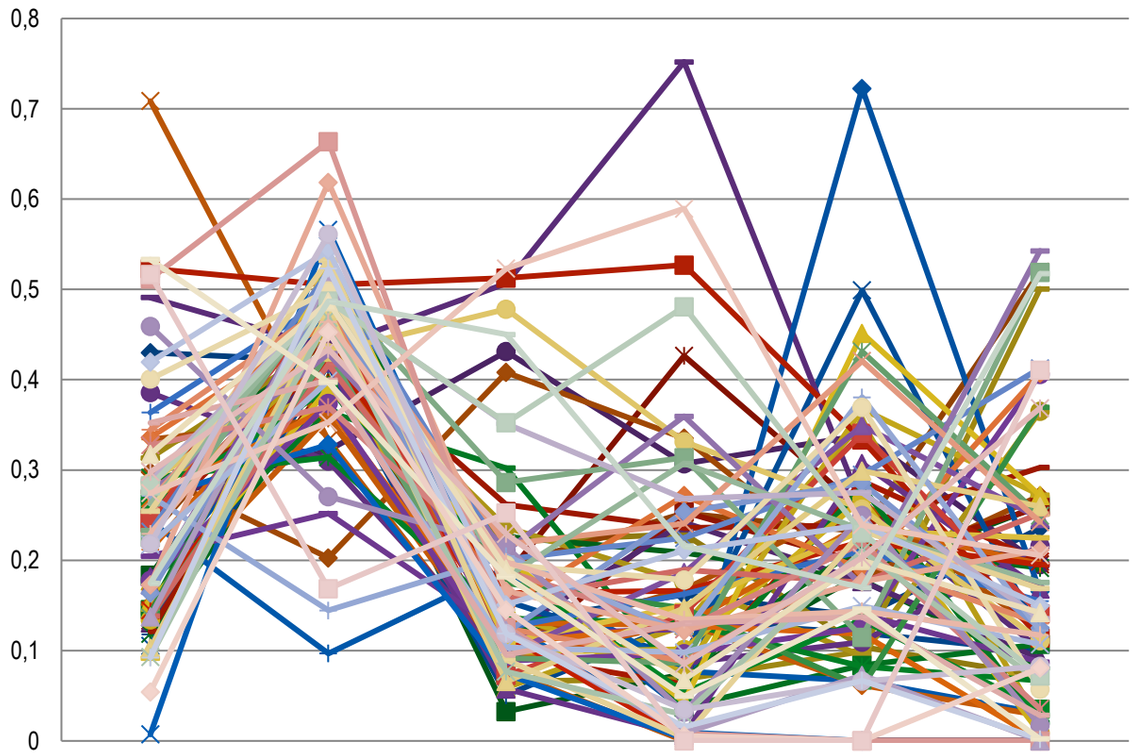
Block 1. Socio-economic conditions
Block 2. Educational potential
Block 3. The effectiveness of research and development
Block 4. The potential of innovation
Block 5. The effectiveness of innovation

2007	*Block 1*	*Block 2*	*Block X*	*Block 3*	*Block 4*	*Block 5*
Block 1	1					
Block 2	-0,2	1				
Block X	0,16	-0,1	1			
Block 4	0,11	0,11	0,7	1		
Block 5	-0,2	0,14	0,22	0,33	1	
Block 6	-0,1	0,05	0,05	0,13	0,21	1

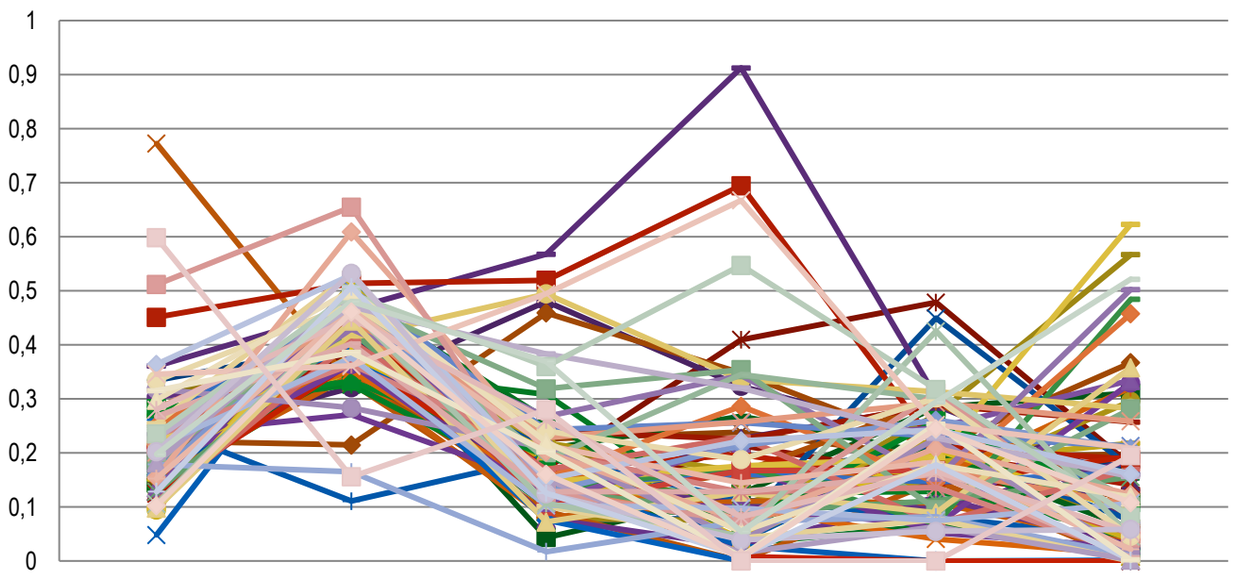
2007



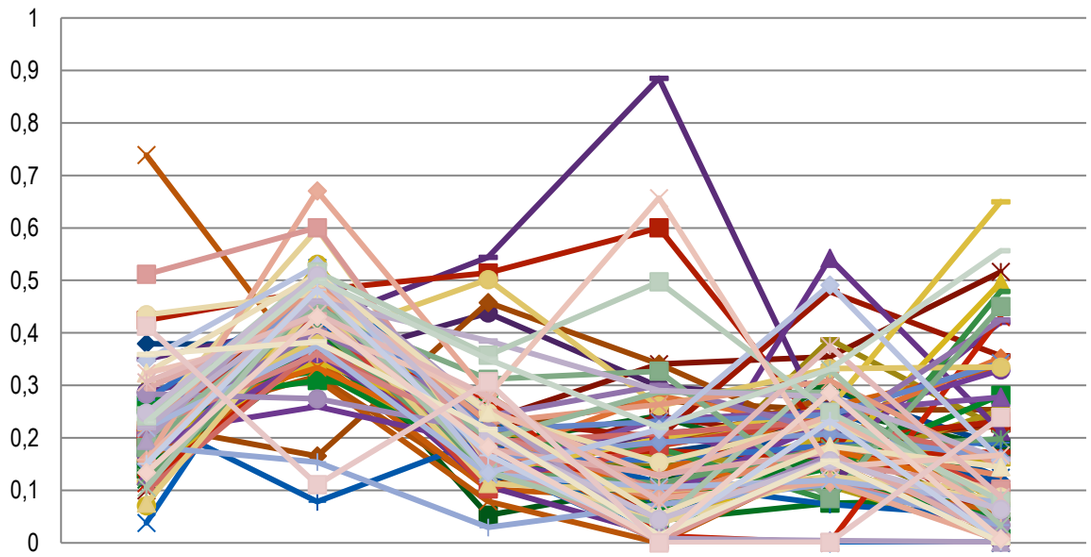
2008



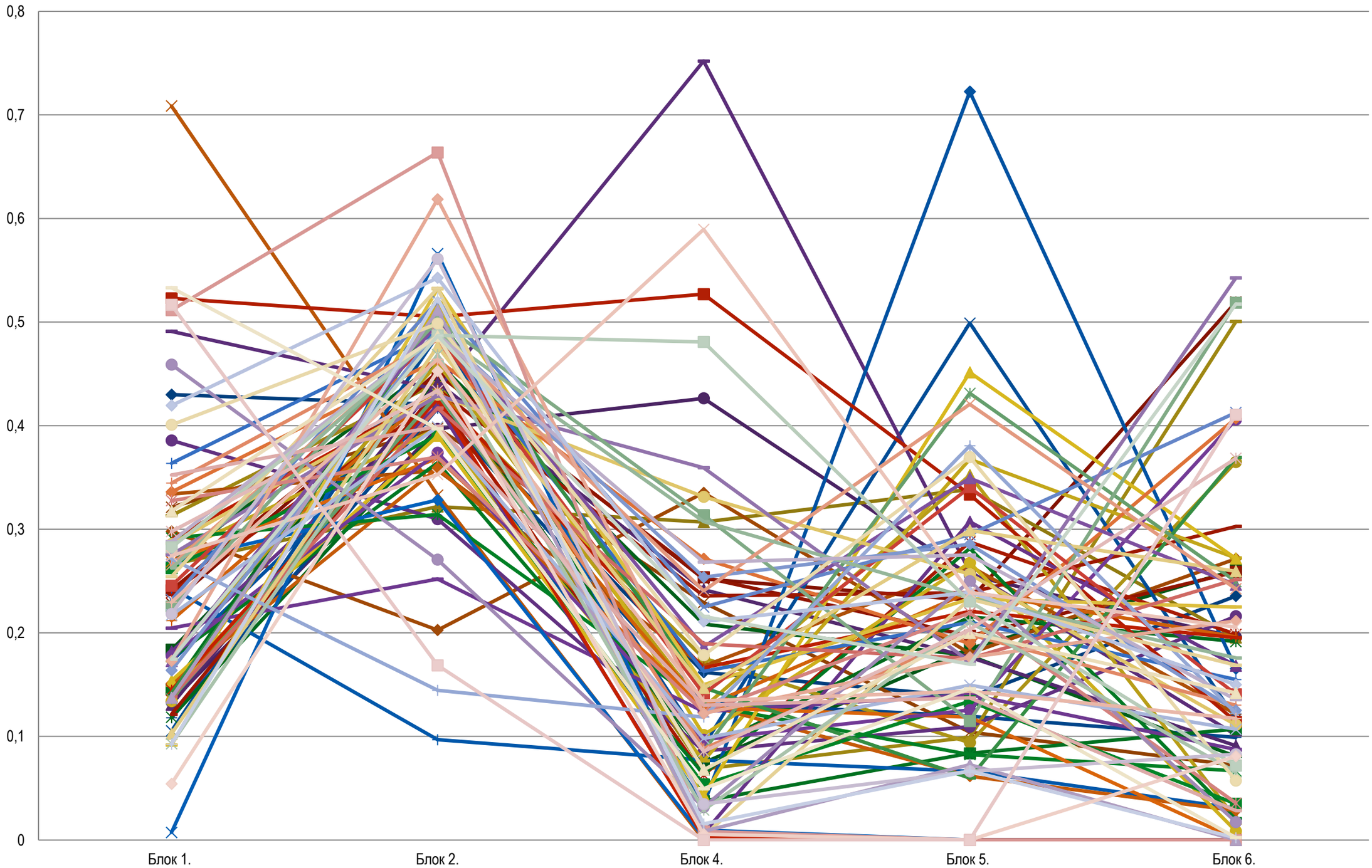
2009

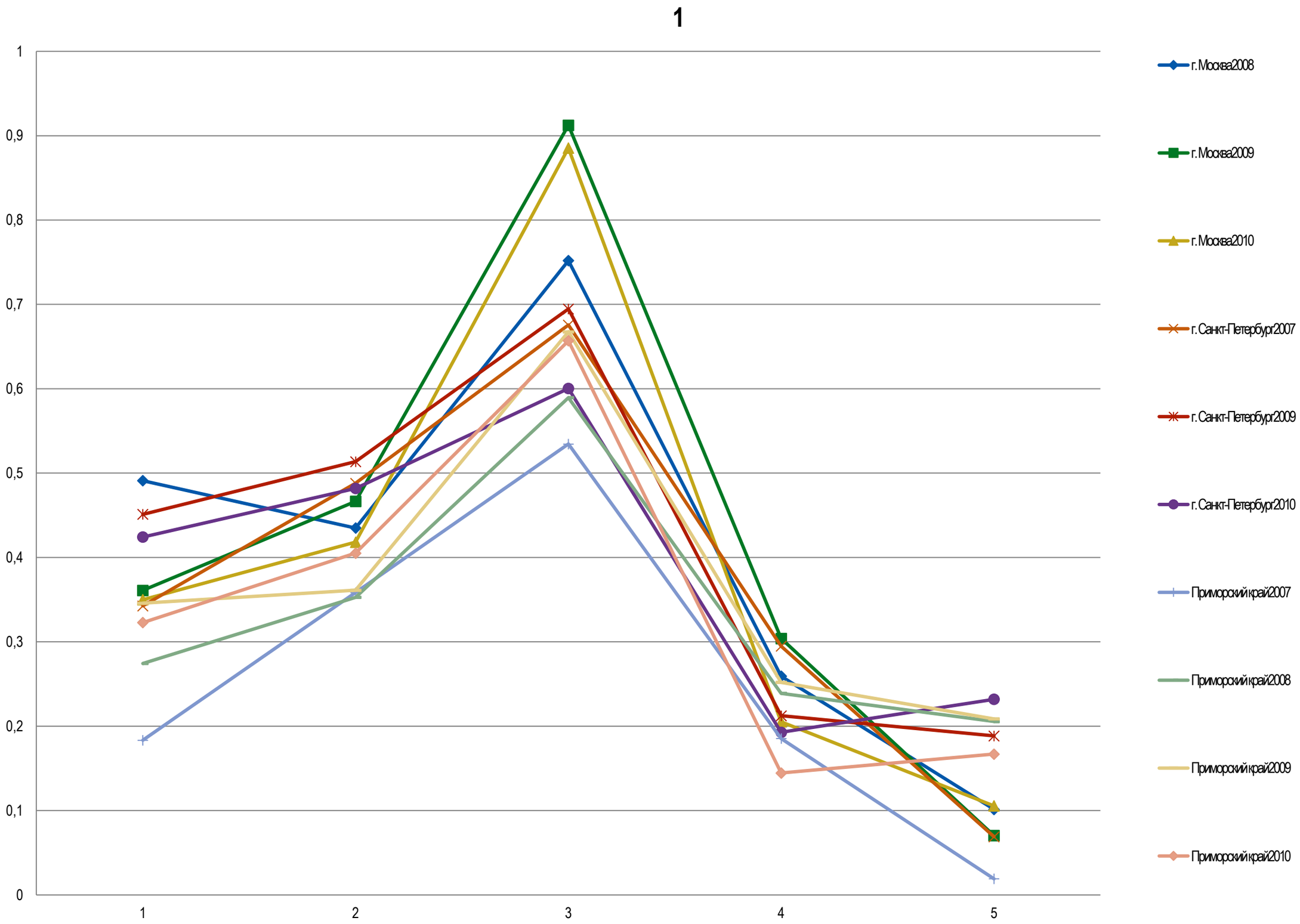


2010

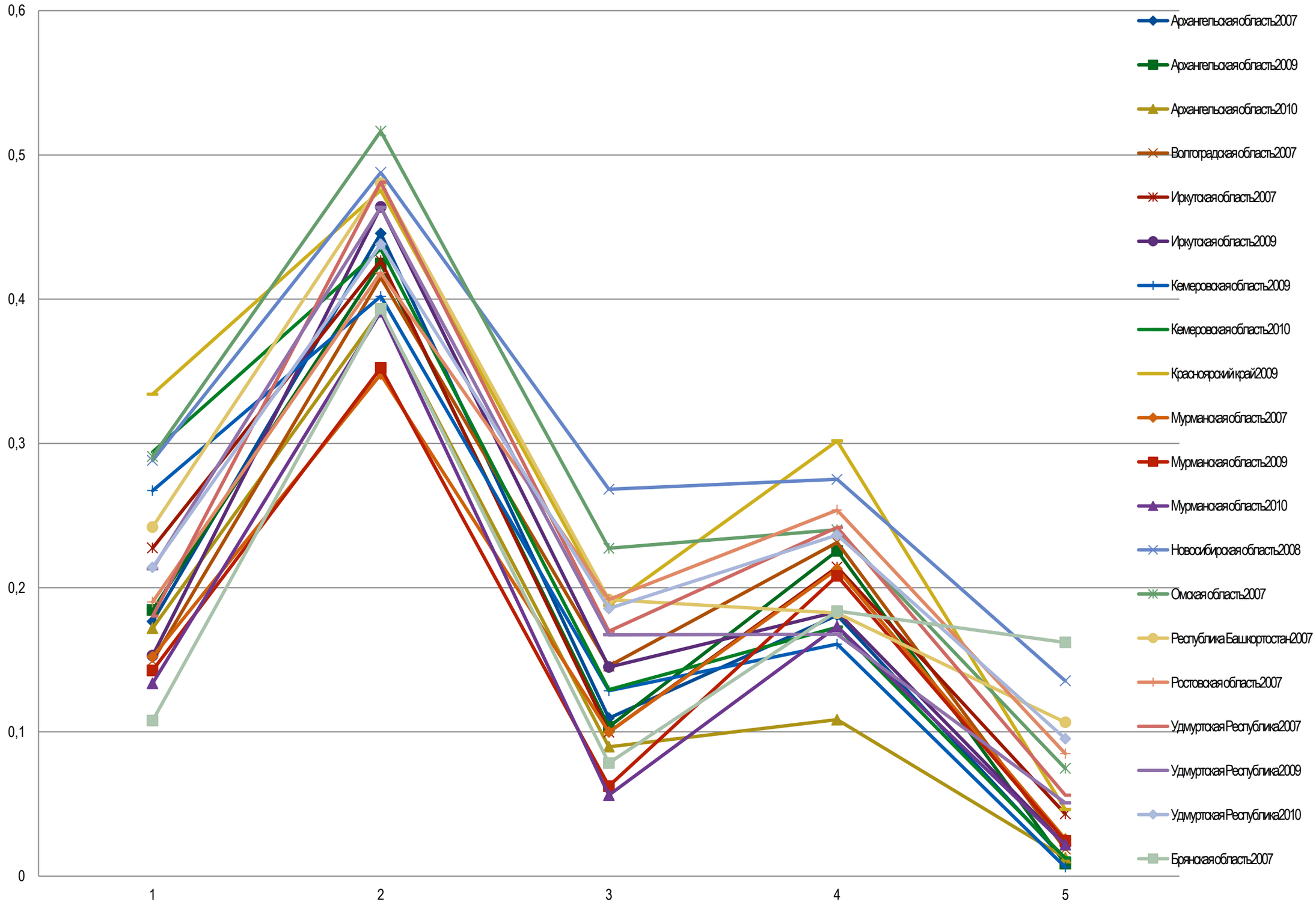


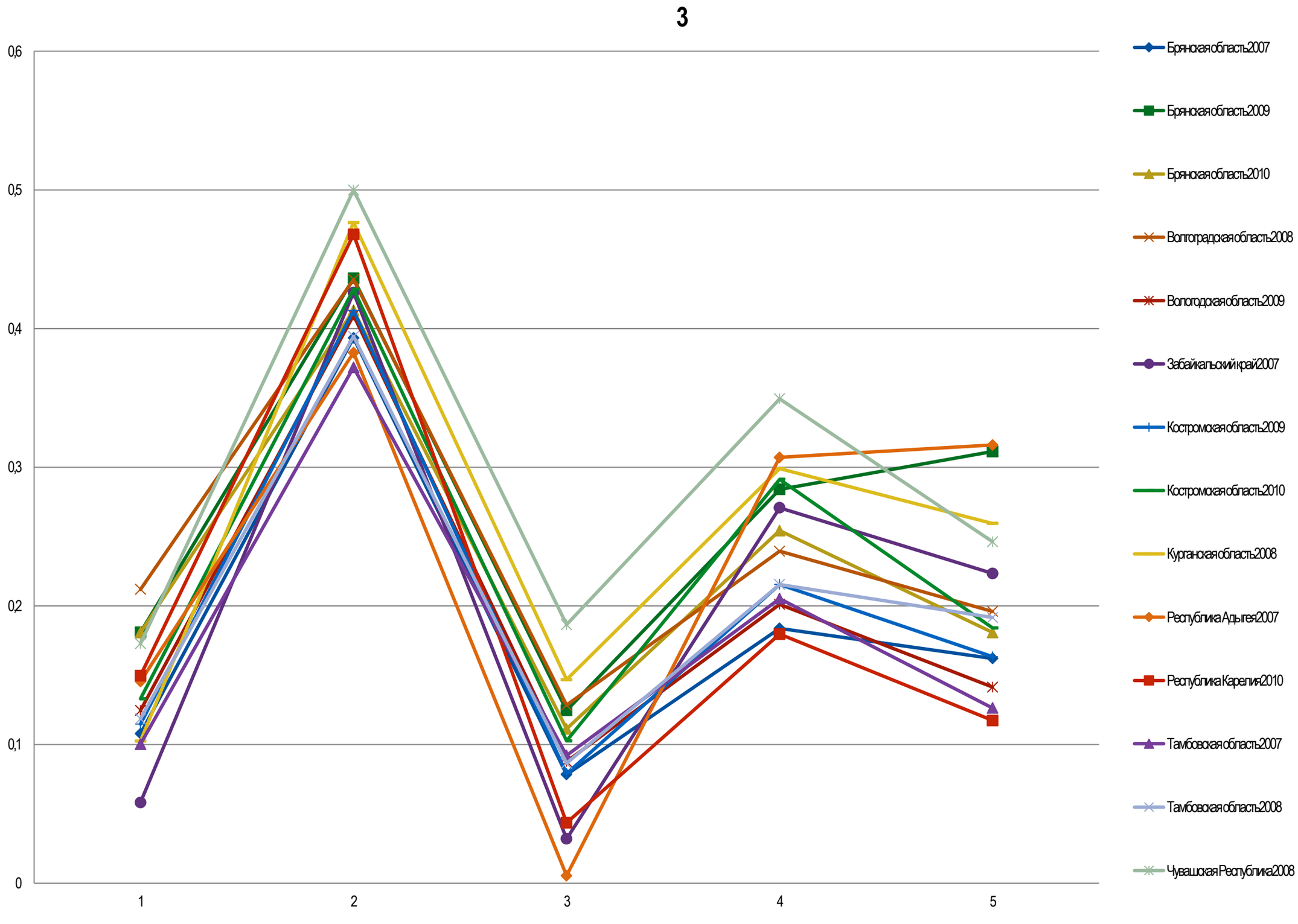
2008

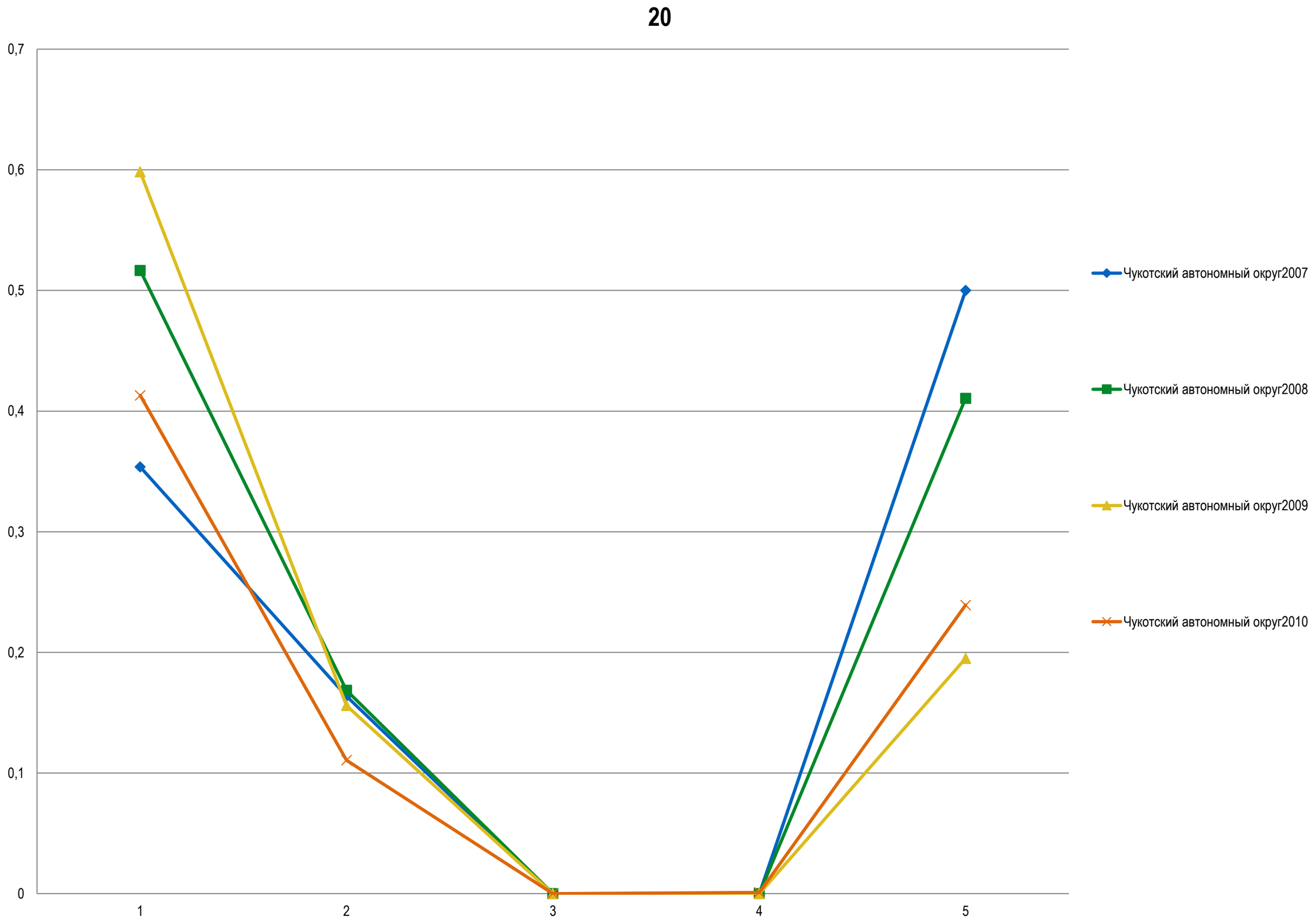



1
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
1
2
3
4
5
г. Москва2008
г. Москва2009
г. Москва2010
г. Санкт-Петербург2007
г. Санкт-Петербург2009
г. Санкт-Петербург2010
Приморский край2007
Приморский край2008
Приморский край2009
Приморский край2010

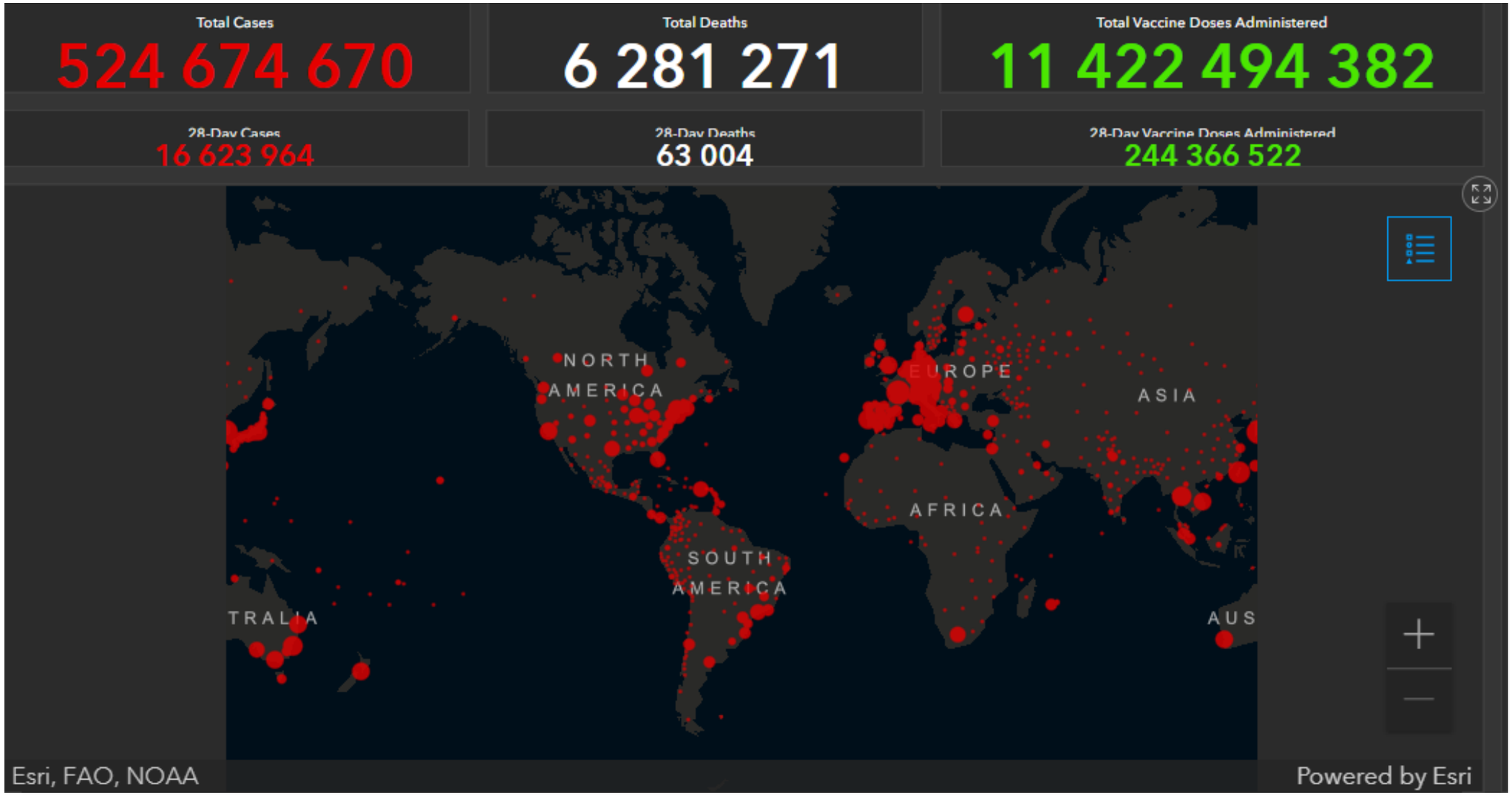
2


0,6
0,5
0,4
0,3
0,2
0,1
0
1
2
3
4
5
Архангельская область2007
Архангельская область2009
Архангельская область2010
Волгоградская область2007
Иркутская область2007
Иркутская область2009
Кемеровская область2009
Кемеровская область2010
Красноярский край2009
Мурманская область2007
Мурманская область2009
Мурманская область2010
Новосибирская область2008
Омская область2007
Республика Башкортостан2007
Ростовская область2007
Удмуртская Республика2007
Удмуртская Республика2009
Удмуртская Республика2010
Брянская область2007


3
0,6
0,5
0,4
0,3
0,2
0,1
0
1
2
3
4
5
Брянская область2007
Брянская область2009
Брянская область2010
Волгоградская область2008
Вологодская область2009
Забайкальский край2007
Костромская область2009
Костромская область2010
Курганская область2008
Республика Адыгея2007
Республика Карелия2010
Тамбовская область2007
Тамбовская область2008
Чувашская Республика2008


20
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
1
2
3
4
5
Чукотский автономный округ2007
Чукотский автономный округ2008
Чукотский автономный округ2009
Чукотский автономный округ2010

COVID-19 CASES



COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU): https://coronavirus.jhu.edu/map.html
Accessed 18 May 2022

DATA USED

- The intensity of separate quarantine measures
- Number of infected people
- Education Level
- Law-abidingness

The source: The Oxford COVID-19 Government Response Tracker

On everyday basis from January 1, 2020
Covers more than 180 countries

https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker

DATA USED

Quarantine measures, Criteria, Grades

Criterion	Grades
C1. School closing	4
C2. Workplace closing	4
C3. Cancel public events	3
C4. Restrictions on gathering size	5
C5. Close public transport	3
C6. Stay at home requirements	4
C7. Restrictions on internal movement	3
C8. Restrictions on international travel	5

Law-abidingness – to which extent people follow the rules stated by government (expert estimations)

AGGREGATION MODEL OF QUARANTINE MEASURES

National Research University
Higher School of Economics

The threshold aggregation of quarantine measure is a lexicographic ordering of criterial values. Countries are compared in a pairwise way beginning from criteria with minimal values being equal 0.

A country for which the number of criteria with 0 values is lower dominates a country with more 0 values. If for a pair of countries the number of criteria with 0 values is equal, then the criteria with value 1 are compared. The process proceeds until maximal values of criteria.

Gathering restrictions

C2. Workplace closing
C3. Cancel public events
C4. Restrictions on gathering size
C6. Stay at home requirements

Government prohibitions

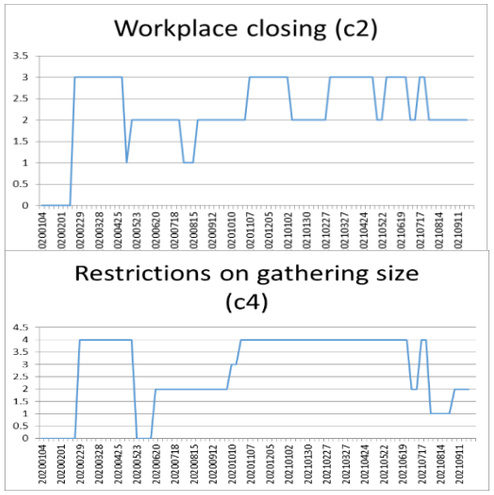
C1. School closing
C5. Close public transport
C7. Restrictions on internal movement
C8. Restrictions on international travel

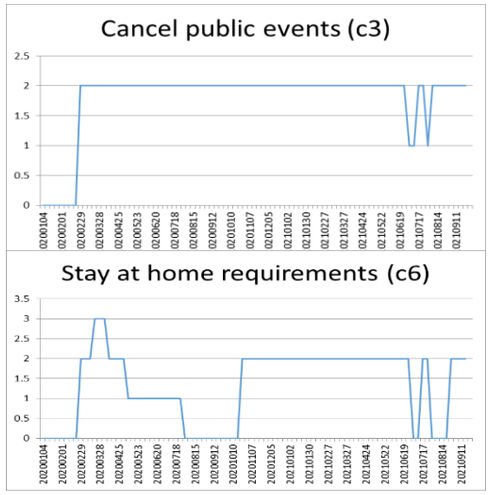
Threshold aggregation of quarantine measures
Example for 3 parameters

	Measure X	Measure Y	Measure Z
Country A	average	average	average
Country B	tough	tough	mild

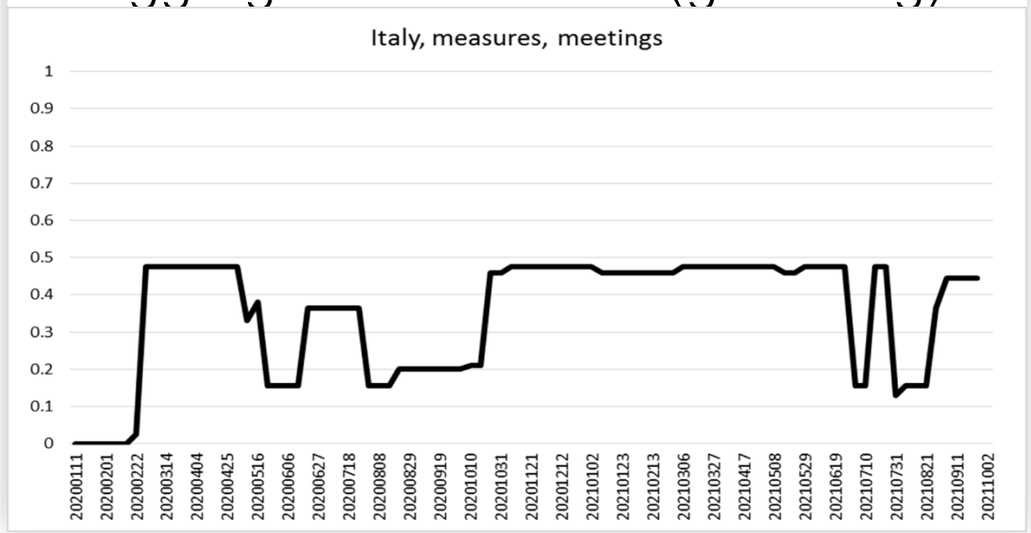
Aggregation – country A has stronger measures than country B

Measures (gathering)

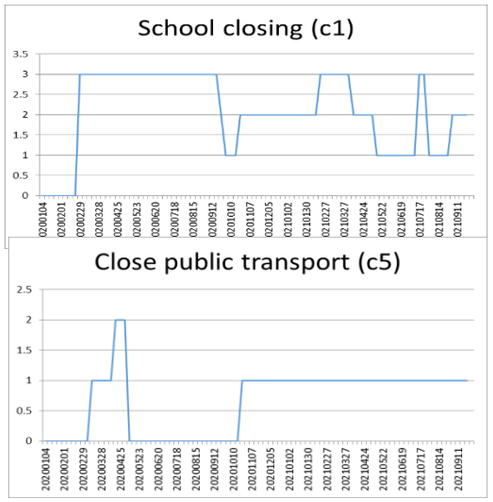


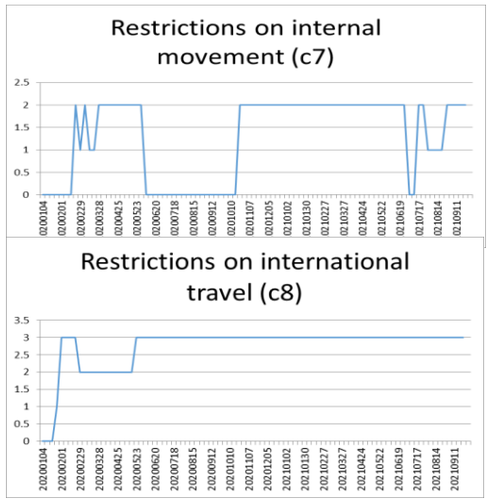


Aggregated measures (gathering)

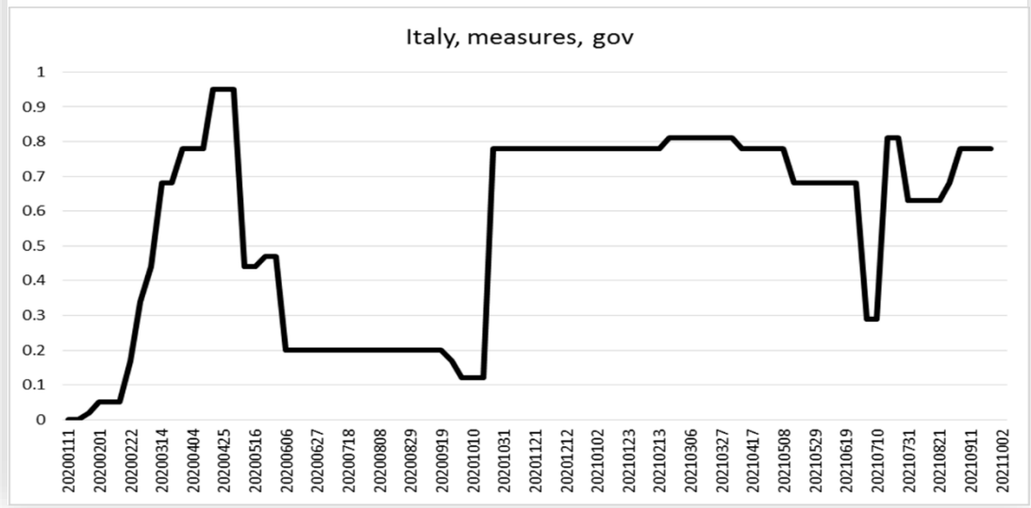


Measures (government)



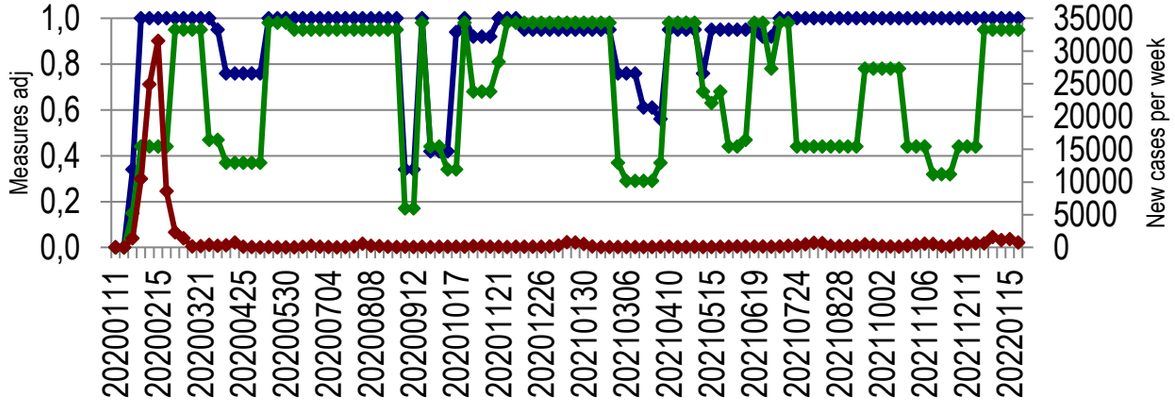


Aggregated measures (government)

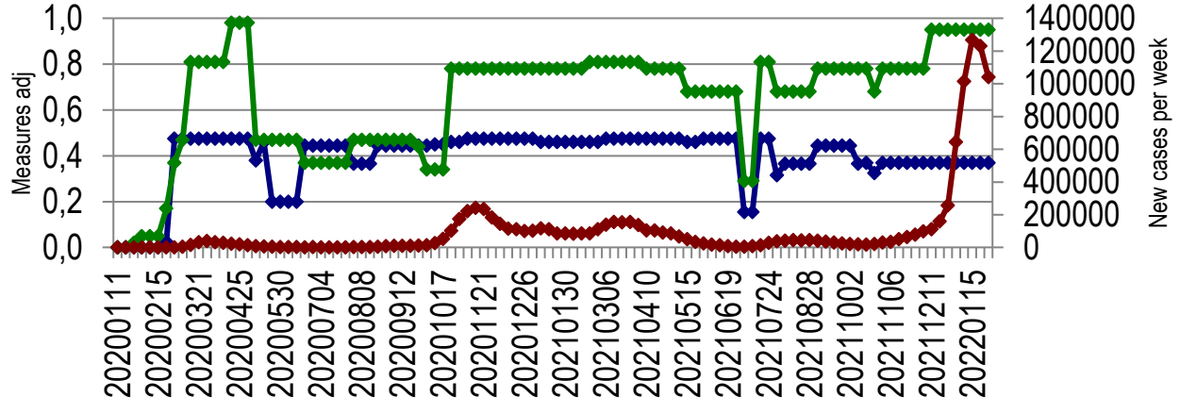


AGGREGATED MEASURES: GATHERING AND ENTRANCE RESTRICTION AND NEW CASES PER WEEK

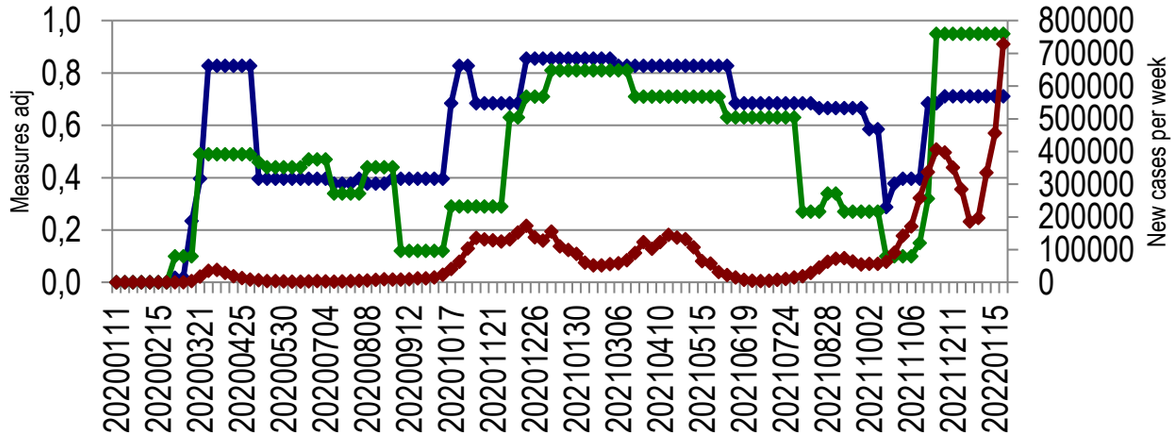
China



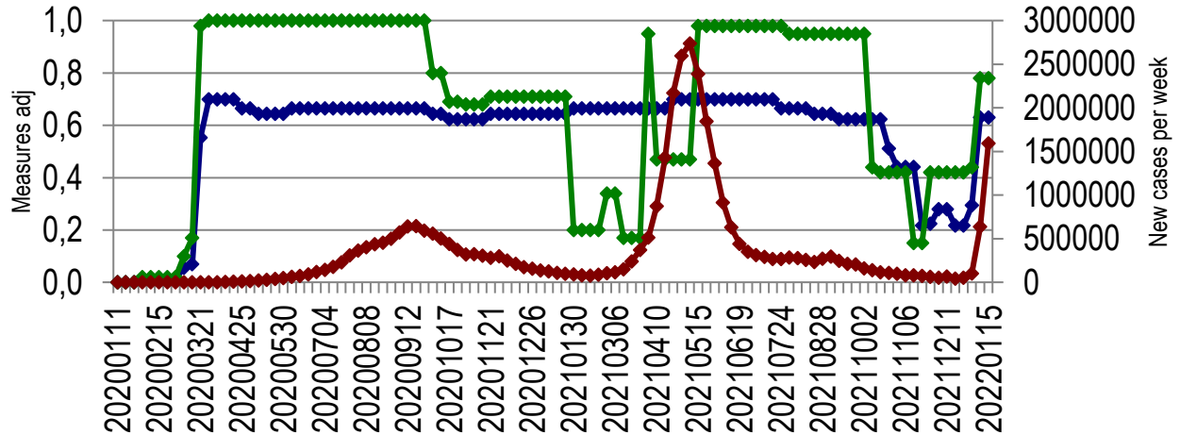
Italy



Germany



India



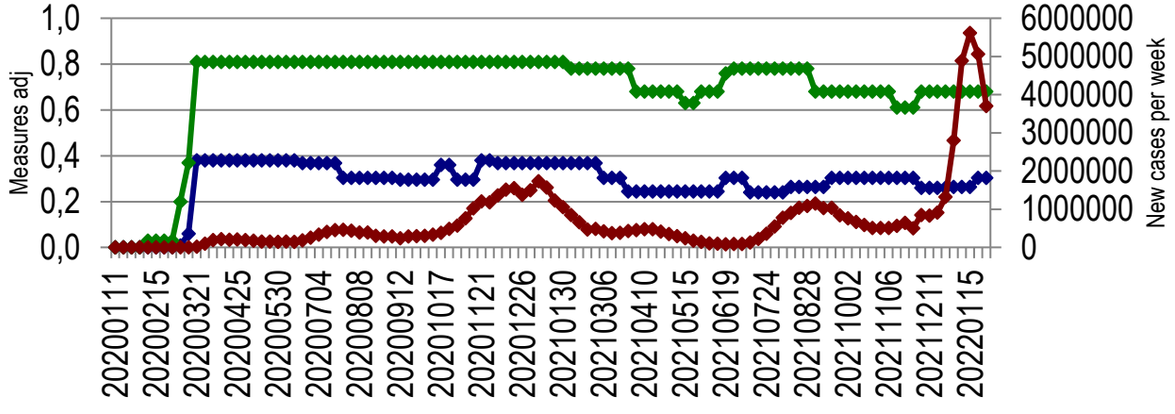
New cases

Measures: Meet.

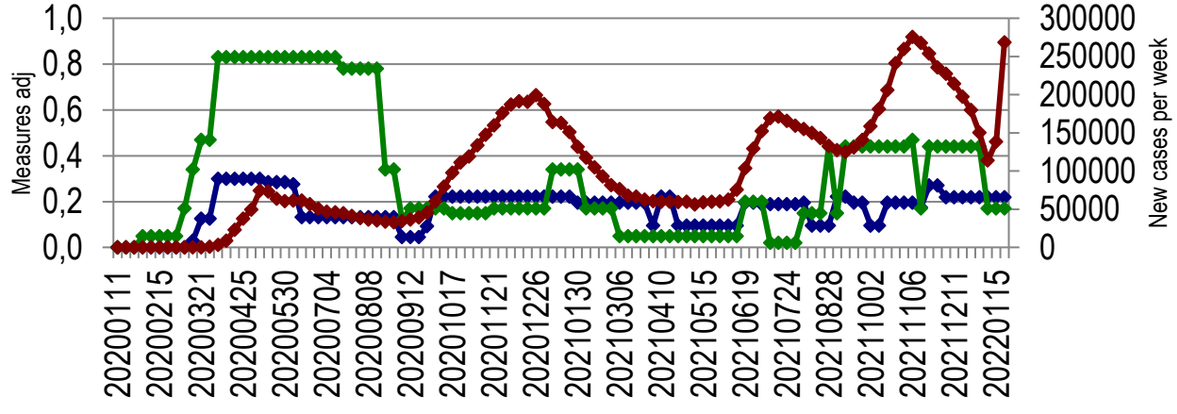
Measures: Gov.

AGGREGATED MEASURES: GATHERING AND ENTRANCE RESTRICTION AND NEW CASES PER WEEK

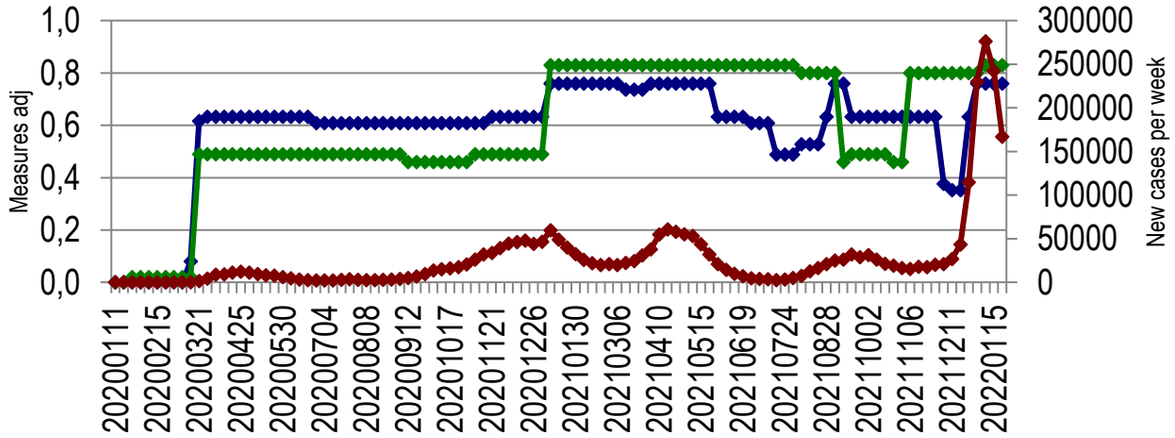
USA



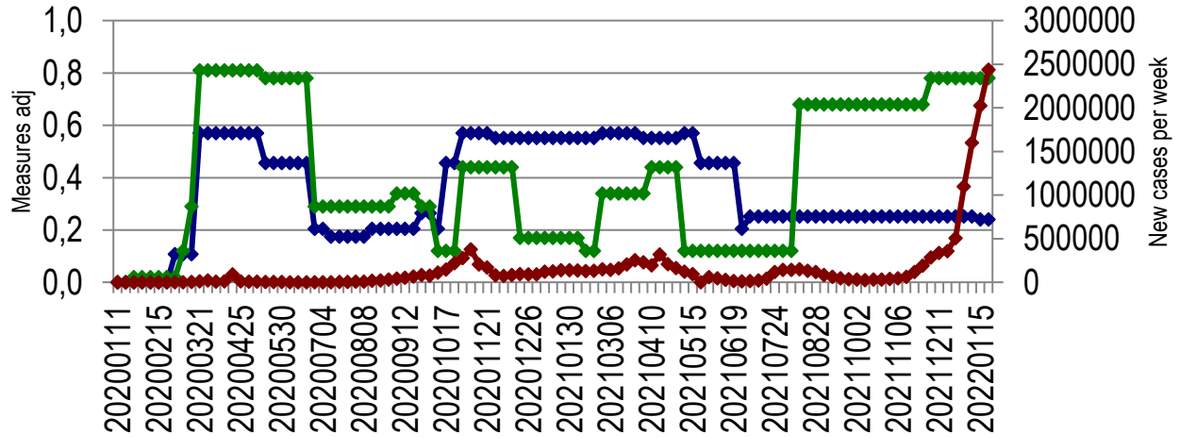
Russia



Canada



France



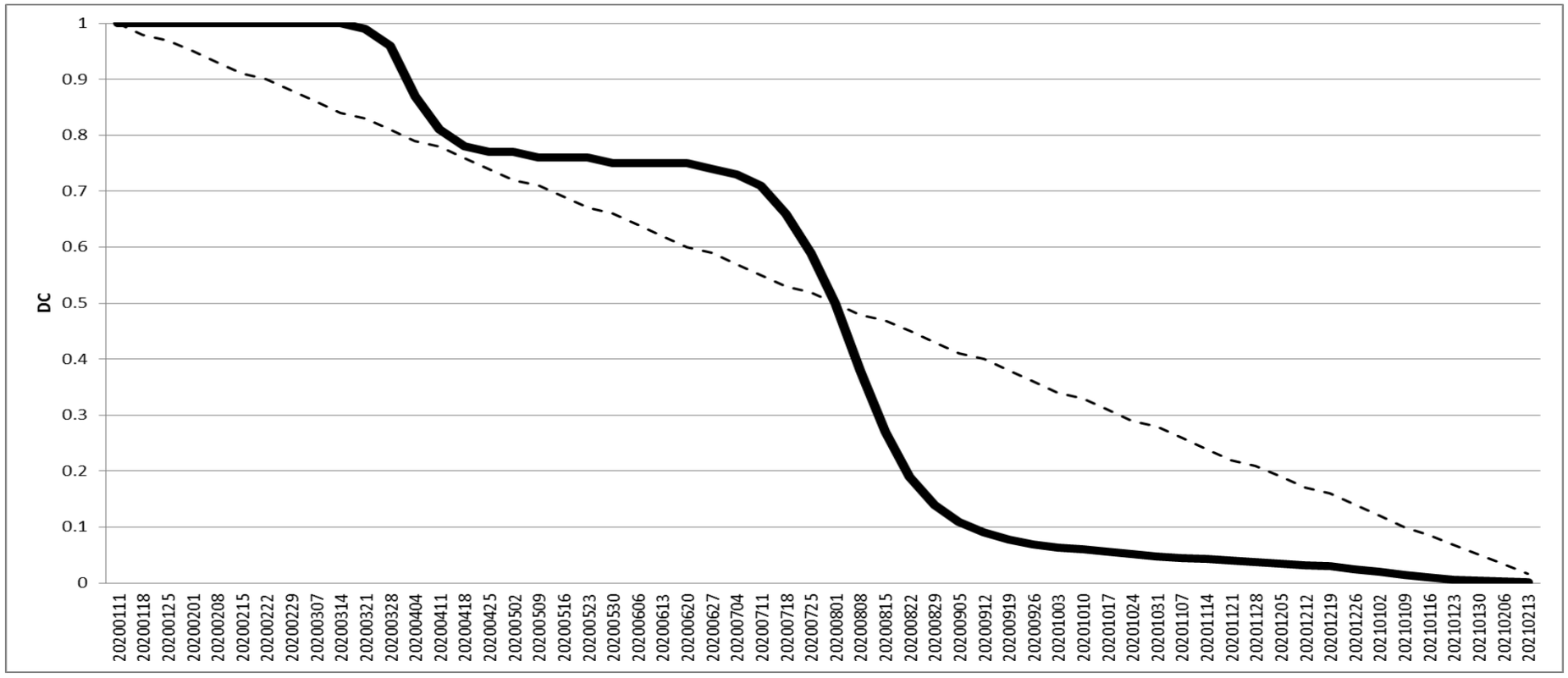
New cases

Measures: Meet.

Measures: Gov.

THE MODEL OF TIME-SERIES ANALYSIS OF THE NUMBER OF INFECTED PEOPLE

For the selected time interval we evaluate the sum of the values of a series for all period (of the length n), and for a current k-th moment we evaluate the share of values from this moment up to the end of the period.



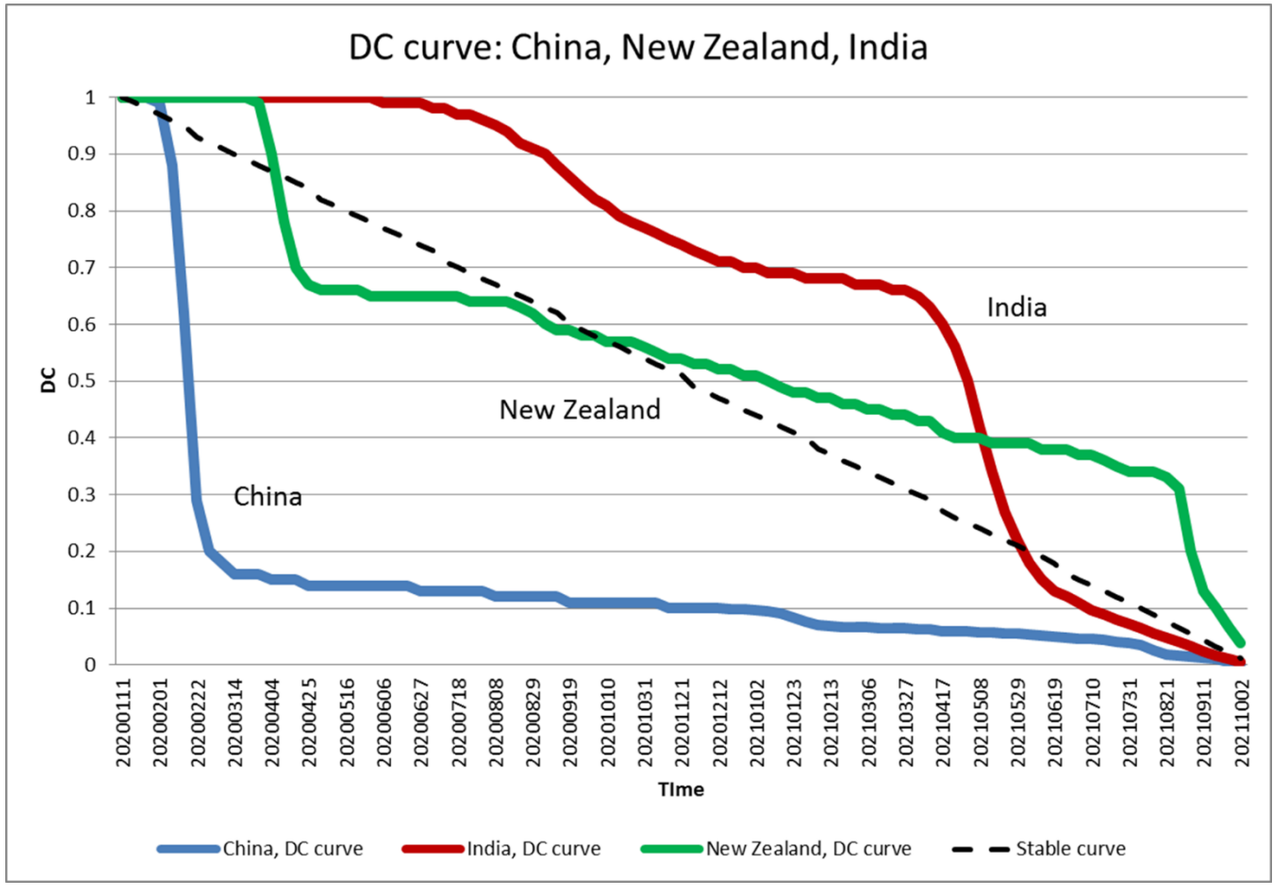
THE MODEL OF TIME-SERIES ANALYSIS OF THE NUMBER OF INFECTED PEOPLE

$$DC(k) = \frac{\sum_{t=k}^{n} v_t}{\sum_{t=1}^{n} v_t},$$

where v_t is the value in the period t

$v\_t = a\_t - a\_(t-1)$, where a_t is the number of new cases in the moment t (from 1 to n).

Non-stationarity is measured as the deviation from the line



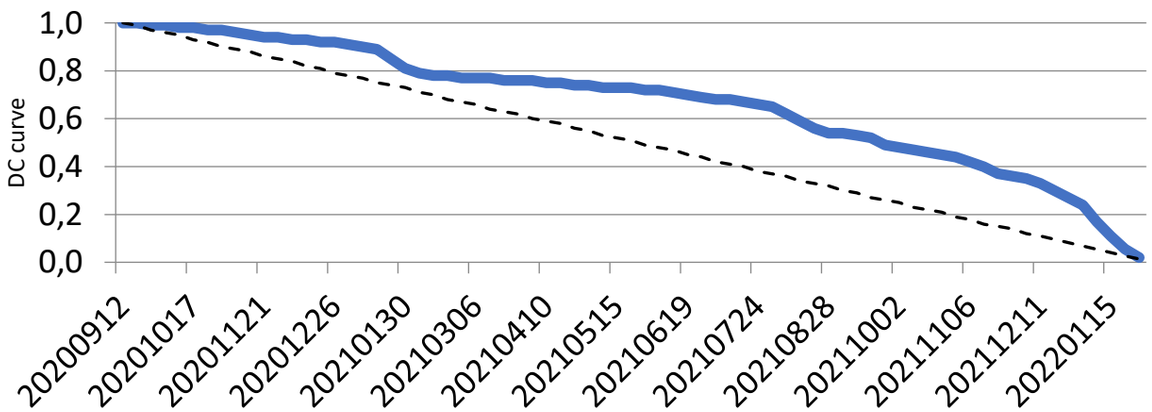
Peak in January – the graph is under diagonal (China)

Peak in September – the graph is above diagonal (India)

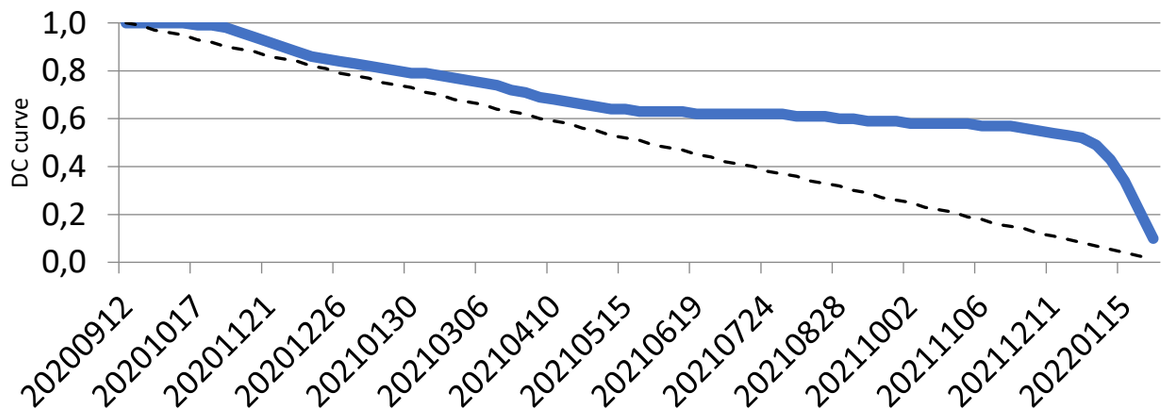
THE MODEL OF NON-STATIONARY TIME-SERIES (INITIAL)

National Research University
Higher School of Economics

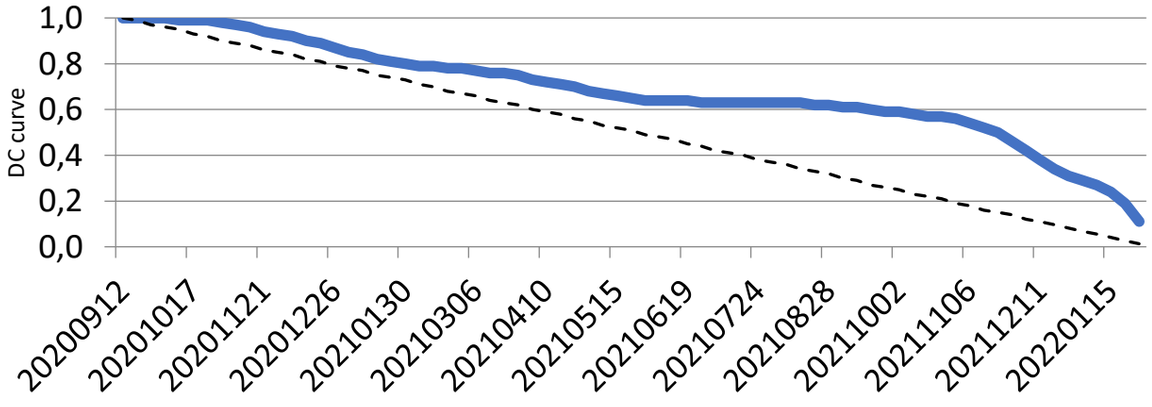
China



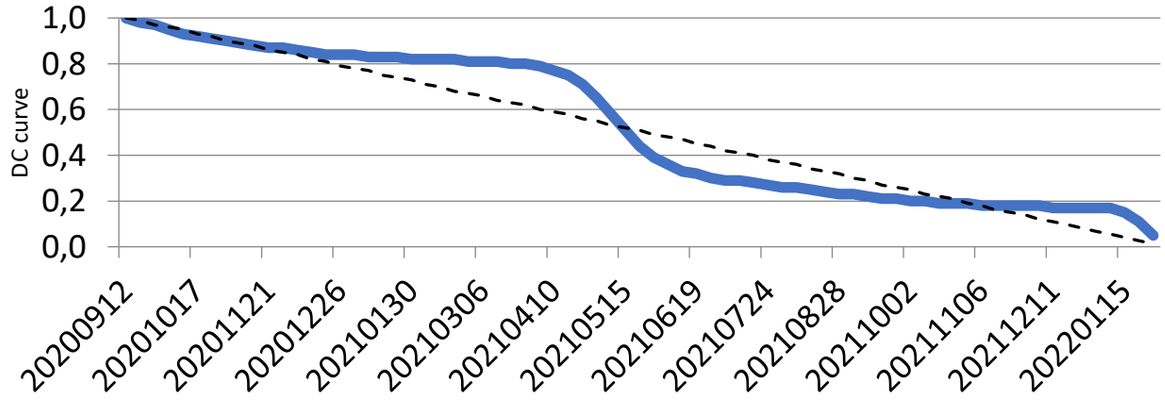
Italy



Germany



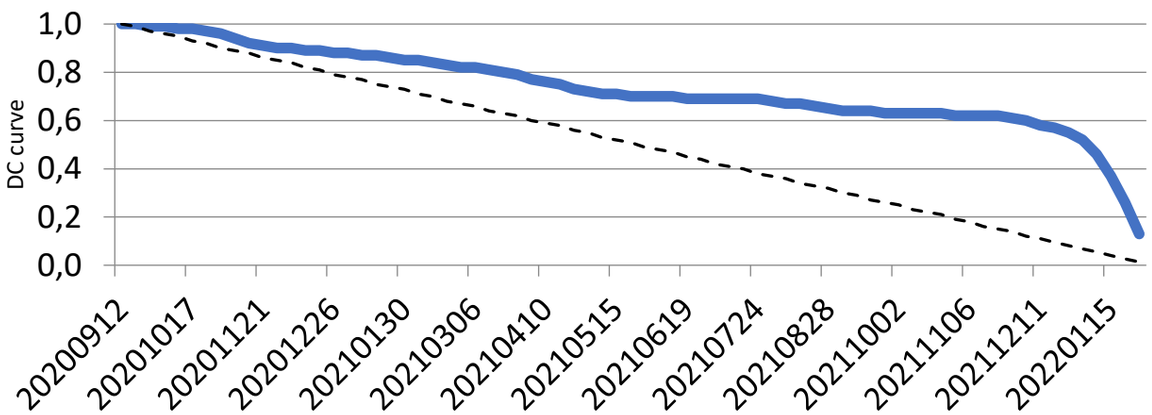
India



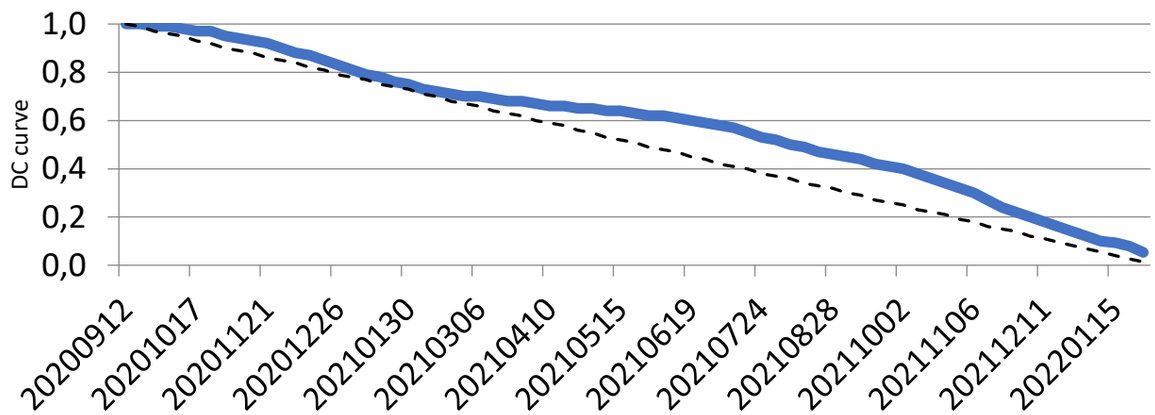
THE MODEL OF NON-STATIONARY TIME-SERIES (INITIAL)

National Research University
Higher School of Economics

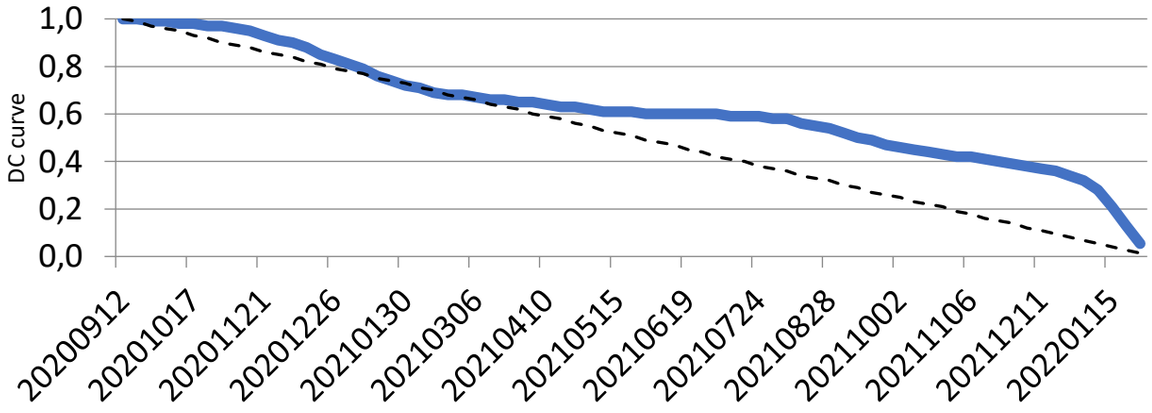
France



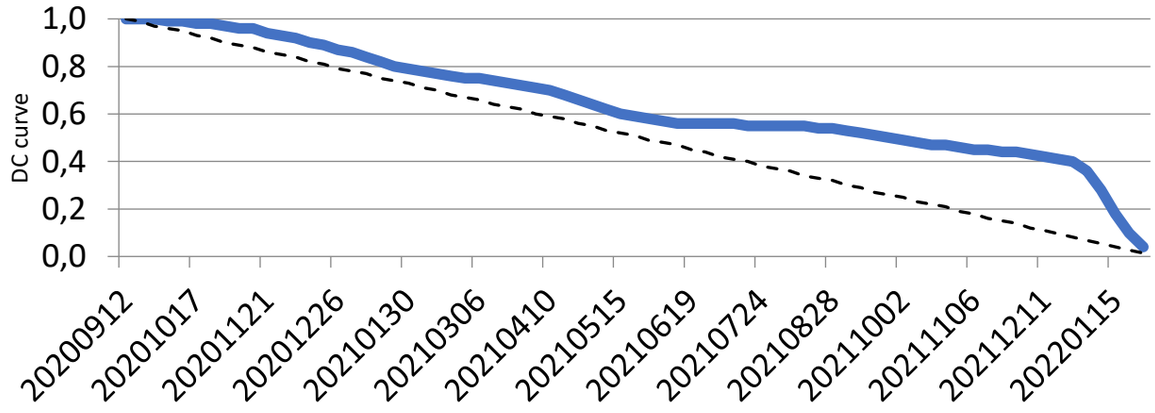
Russia



USA

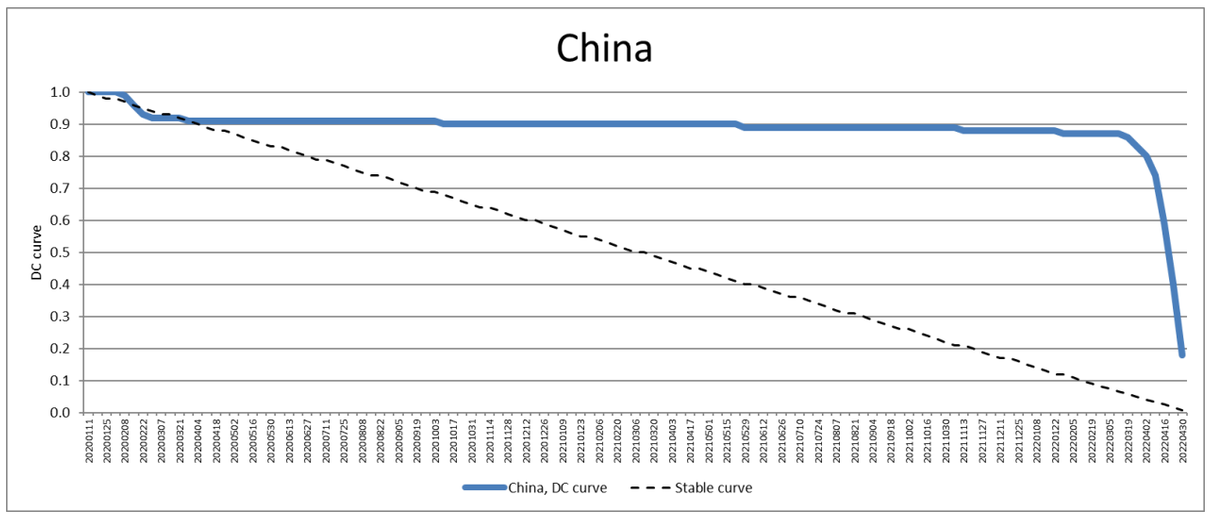


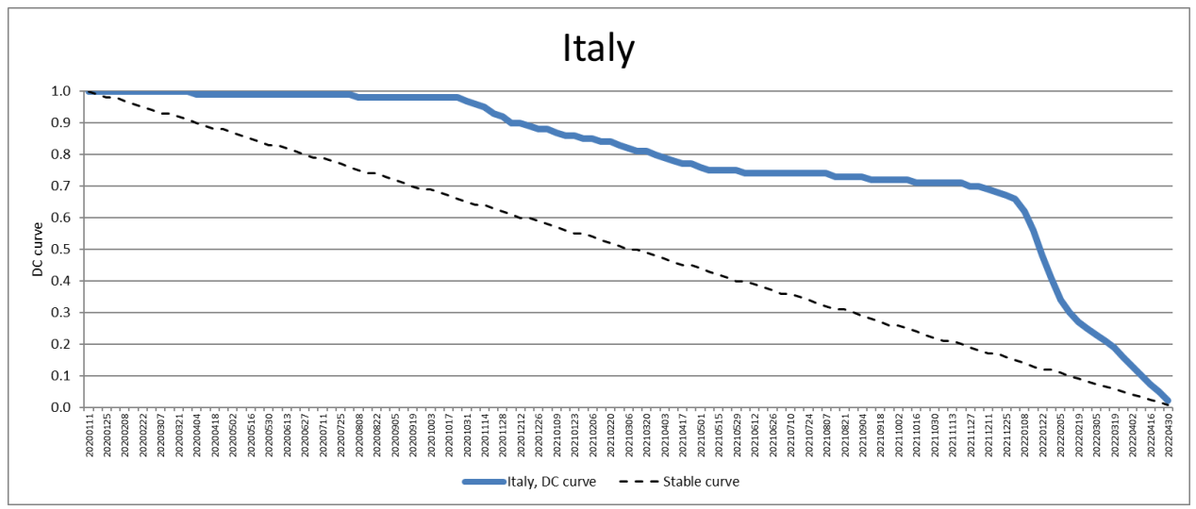
Canada

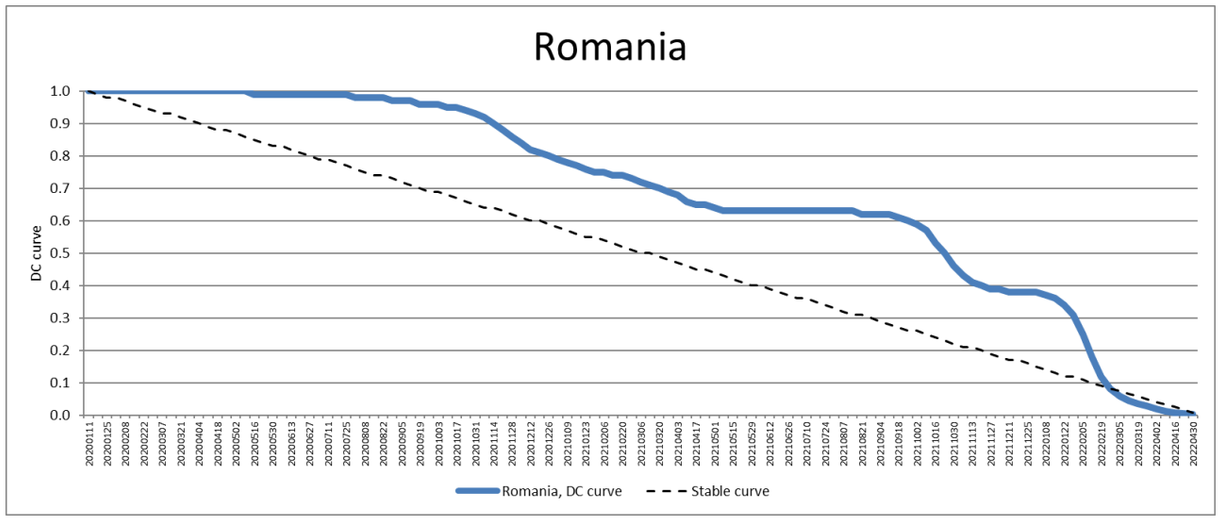


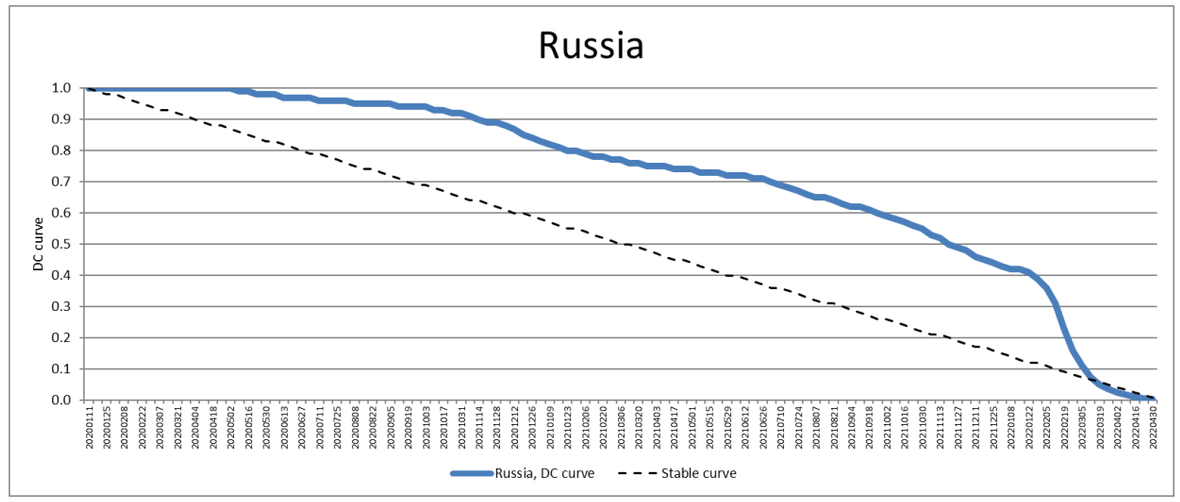
THE MODEL OF NON-STATIONARY TIME-SERIES (NOW)

National Research University
Higher School of Economics









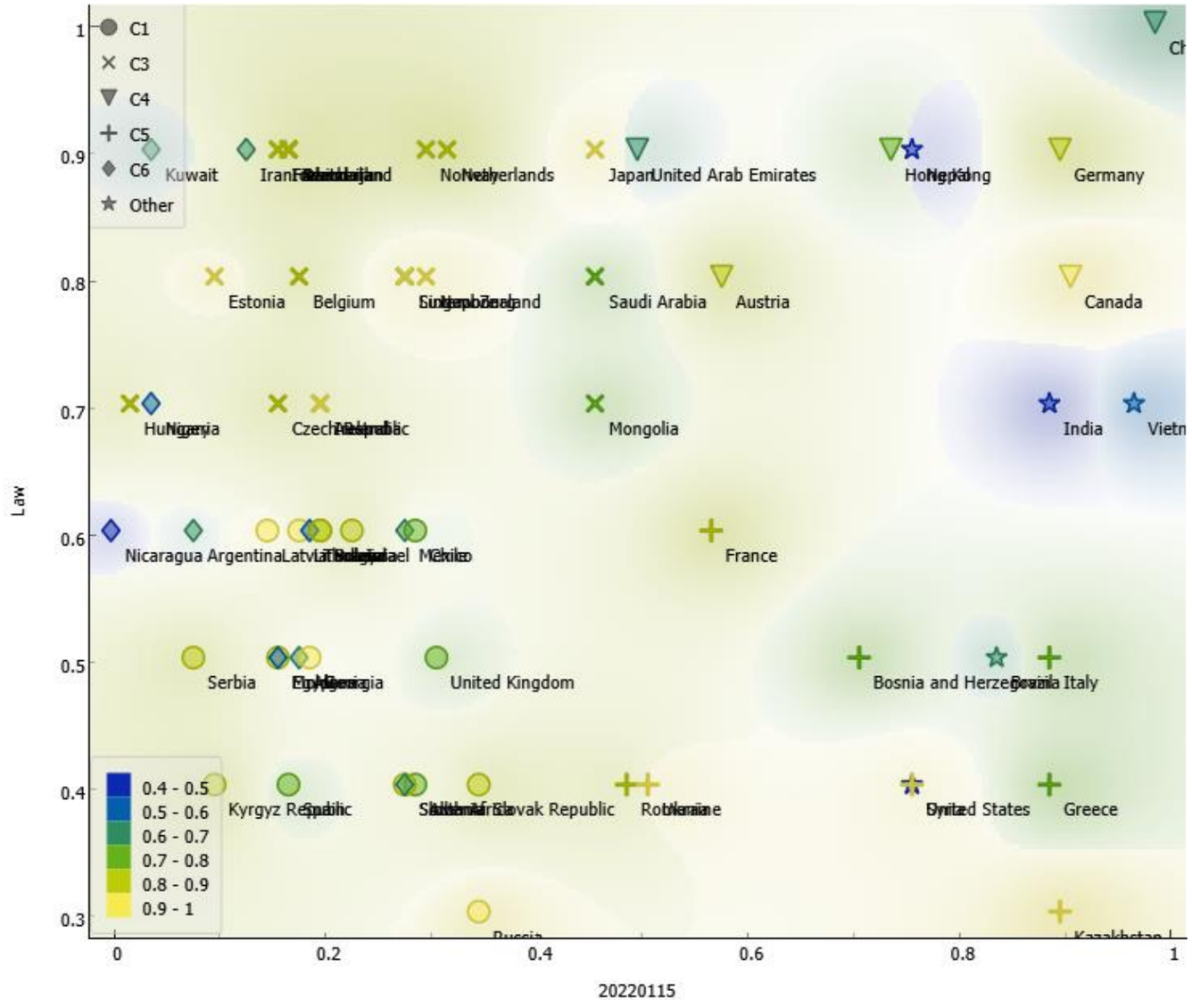
GROUPS OF COUNTRIES WITH SIMILAR STRUCTURE OF EDUCATION LEVEL, LAW-ABIDINGNESS, MEASURES AND INFECTED PEOPLE

Initial data:
66 countries
Basic system of indices

Period under study:
February 1, 2020 – May 18, 2022

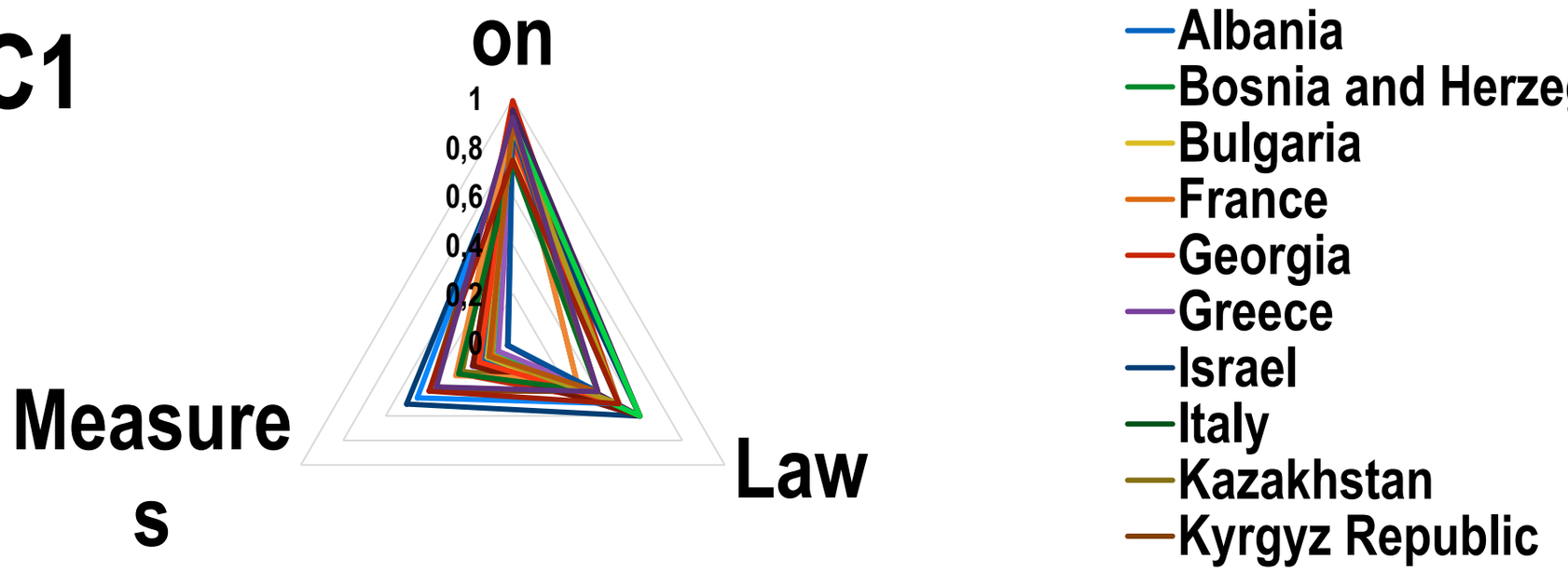
Methods used: Cluster analysis

Example 1: 15 January 2022

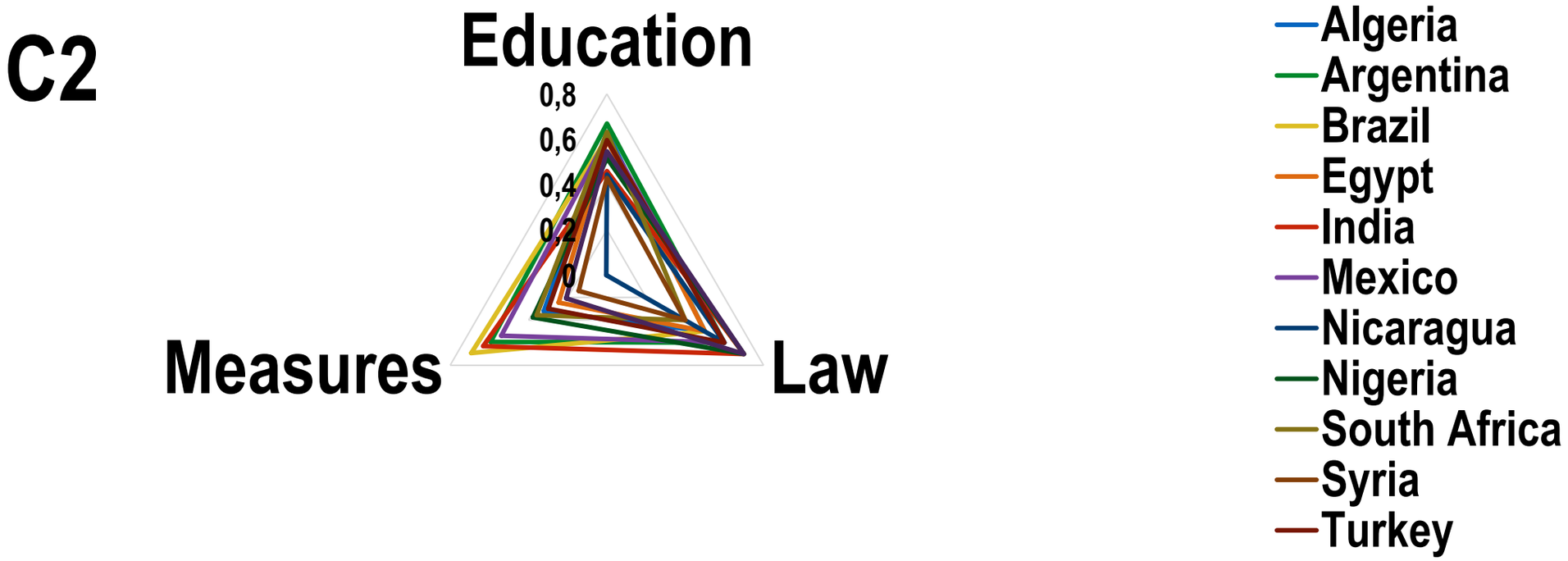


REGRESSION MODELS

$$N_t(Cluster\ 1) = 0{,}015 + 0{,}523N_{t-1} - 0{,}715N_{t-2} - 0{,}009N_{t-3} + 0{,}868N_{t-4} + 0{,}078N_{t-5}$$



$$N_t(Cluster\ 2) = 0{,}036 - 0{,}356N_{t-1} - 0{,}328N_{t-2} - 0{,}461N_{t-3} - 0{,}44N_{t-4} + 1{,}737N_{t-5}$$



REGRESSION EQUATIONS

National Research University
Higher School of Economics

$N_t(\mathrm{C1}) =$

Intercept	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12
+0,23	+0,53	-0,01	-0,69	-0,48	+0,66	-0,95	+1,08	-0,54	+0,48	-0,38	-1,21	+1,00

$N_t(\mathrm{C2}) =$

Intercept	t-1	t-2	t-3
+0,02	+1,10	-0,93	+0,80

$N_t(\mathrm{C3}) =$

Intercept	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11
+0,33	-0,09	+0,08	-0,37	-1,54	+0,27	+2,50	-1,74	-1,12	+0,53	+0,05	-0,63

$N_t(\mathrm{C4}) =$

Intercept	t-1	t-2	t-3	t-4
+0,88	+2,55	-0,71	-3,64	+1,54

$N_t(\mathrm{C5}) =$

Intercept	t-1	t-2	t-3	t-4	t-5	t-6
+0,67	-0,88	+1,65	-0,98	-2,85	+0,83	+2,97

$N_t(\mathrm{C6}) =$

Intercept	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8
+0,09	+0,05	-2,95	+1,35	+3,38	-3,62	-2,78	+2,79	+2,95

RESULTS

Cluster	All	Average	1	2	3	4	5	6
Multiple R	0,765	0,959	0,992	0,996	0,793	0,980	0,999	0,994
R Square	0,585	0,925	0,984	0,992	0,630	0,961	0,997	0,988
Adjusted R Square	0,455	0,769	0,921	0,969	0,047	0,807	0,980	0,889
Standard Error	0,318	0,176	0,068	0,053	0,539	0,187	0,040	0,167
Observations	64	10,667	16	5	19	6	8	10
			Albania Bulgaria Chile Georgia Israel Kyrgyz Republic Latvia Lithuania Moldova Poland Russia Serbia Slovak Republic Slovenia Spain United Kingdom	Brazil India Nepal Syria Vietnam	Australia Azerbaijan Belgium Czech Republic Denmark Estonia Finland Hungary Iceland Ireland Japan Luxembourg Mongolia Netherlands New Zealand Norway Saudi Arabia Singapore Switzerland	Austria Canada China Germany Hong Kong United Arab Emirates	Bosnia and Herzegovina France Greece Italy Kazakhstan Romania Ukraine United States	Algeria Argentina Egypt Iran Kuwait Mexico Nicaragua Nigeria South Africa Turkey

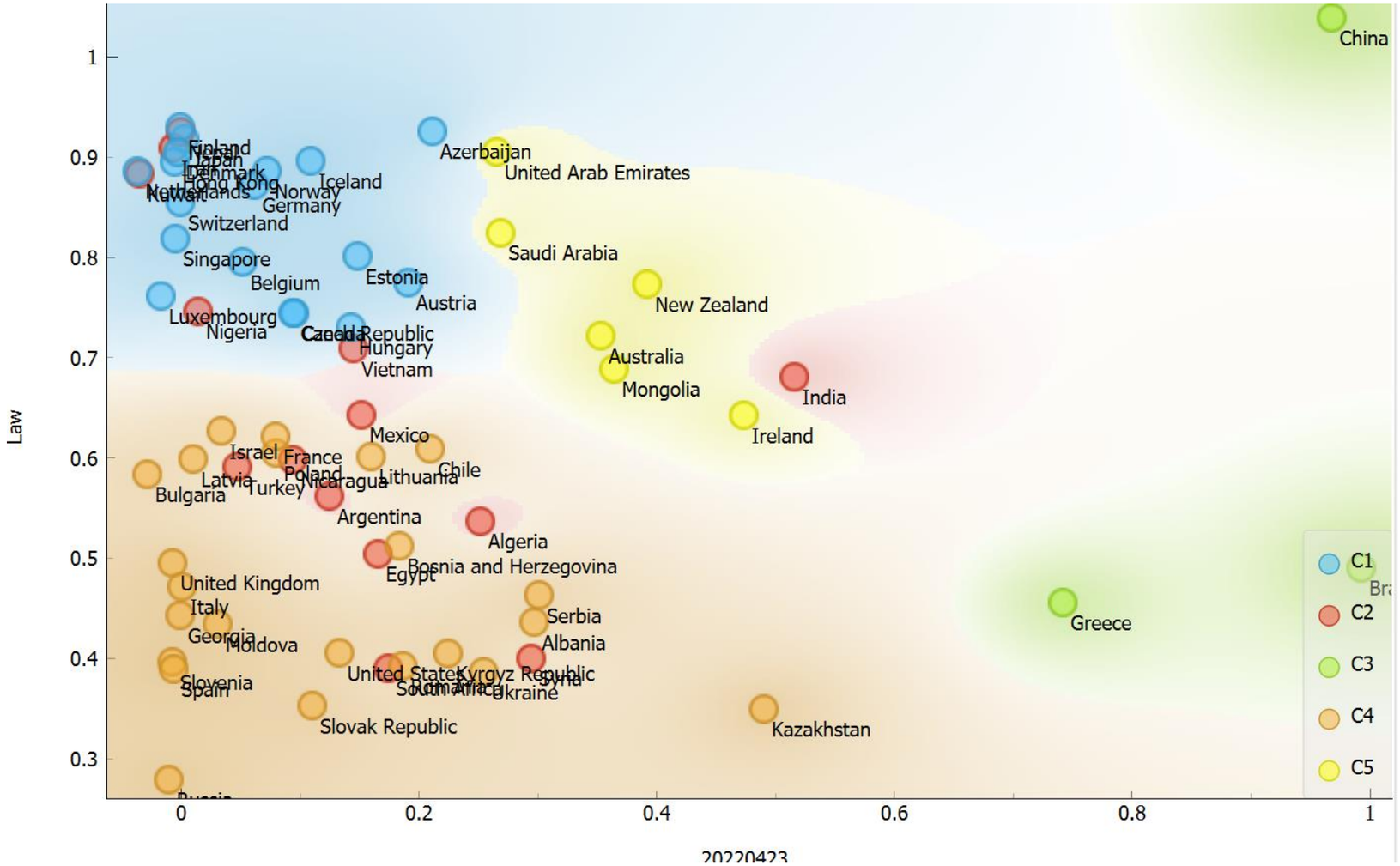
Comparison of the results as of July 25, 2020 for Cluster 1 with the entire sample

Number of weeks	R-squared for Cluster 1	R-squared for all countries
1	0.487	0.346
2	0.893	0.644
3	0.896	0.655
4	0.896	0.679
5	0.911	0.678

Example 2: 23 April 2022

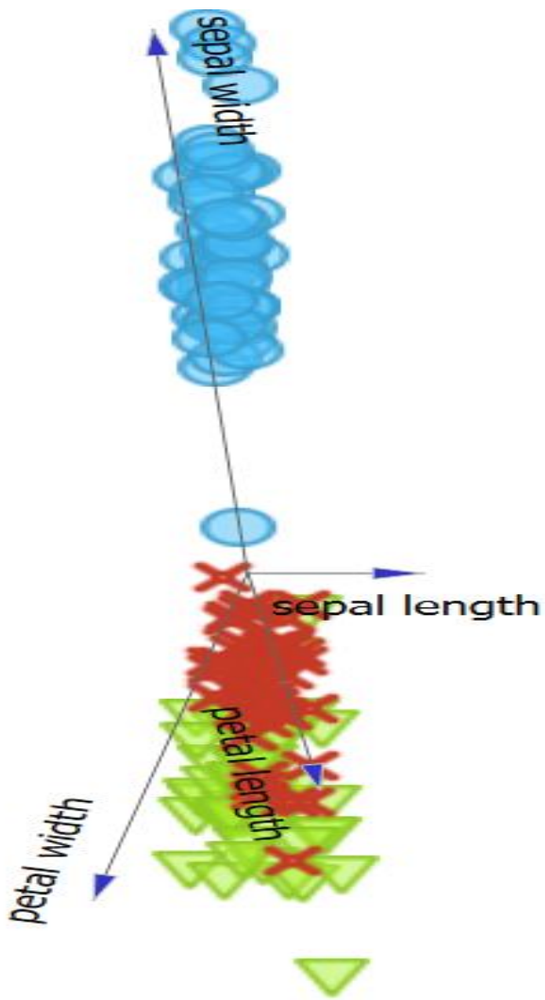
Initial data:
66 countries
Basic system of indices

Average R^2: 0.8



Reducing dimensionality in Data

1. PCA
2. MDM;
3. Expert Analysis;
4. Neural Networks;
5. tSNE.



Power Indices

- $\boldsymbol{N}$ – Set of parties;
- $\boldsymbol{w_i}$ – number of votes;
- $\boldsymbol{q}$ – quota;
- **Wining coalition:**

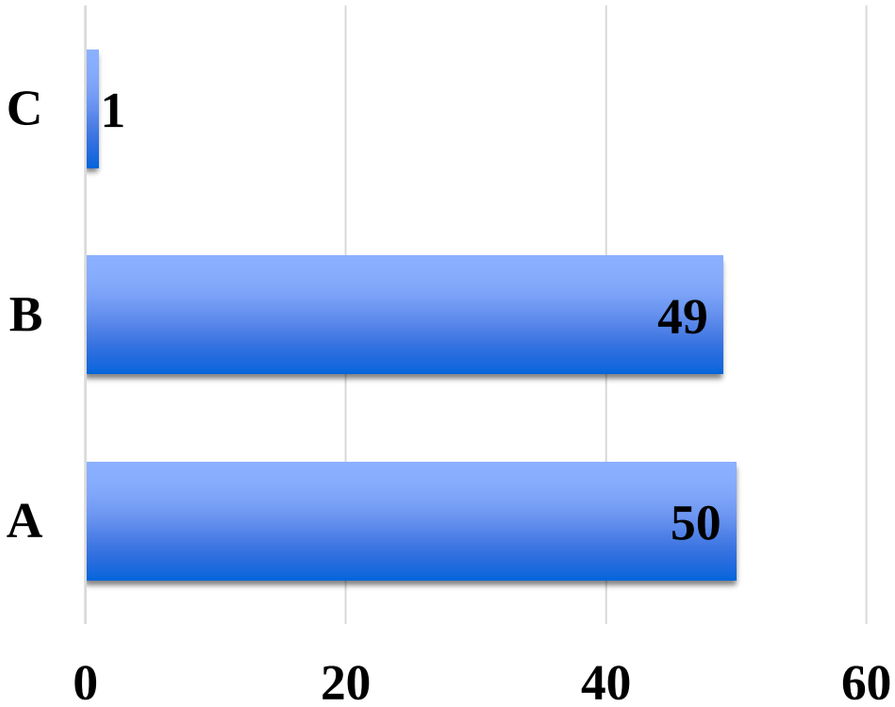
$S \subseteq 2^N, if \ \sum_{i \in S} w_i \geq q$;

- **Pivotal paryt:**

Party $i \in S$ is pivotal if:

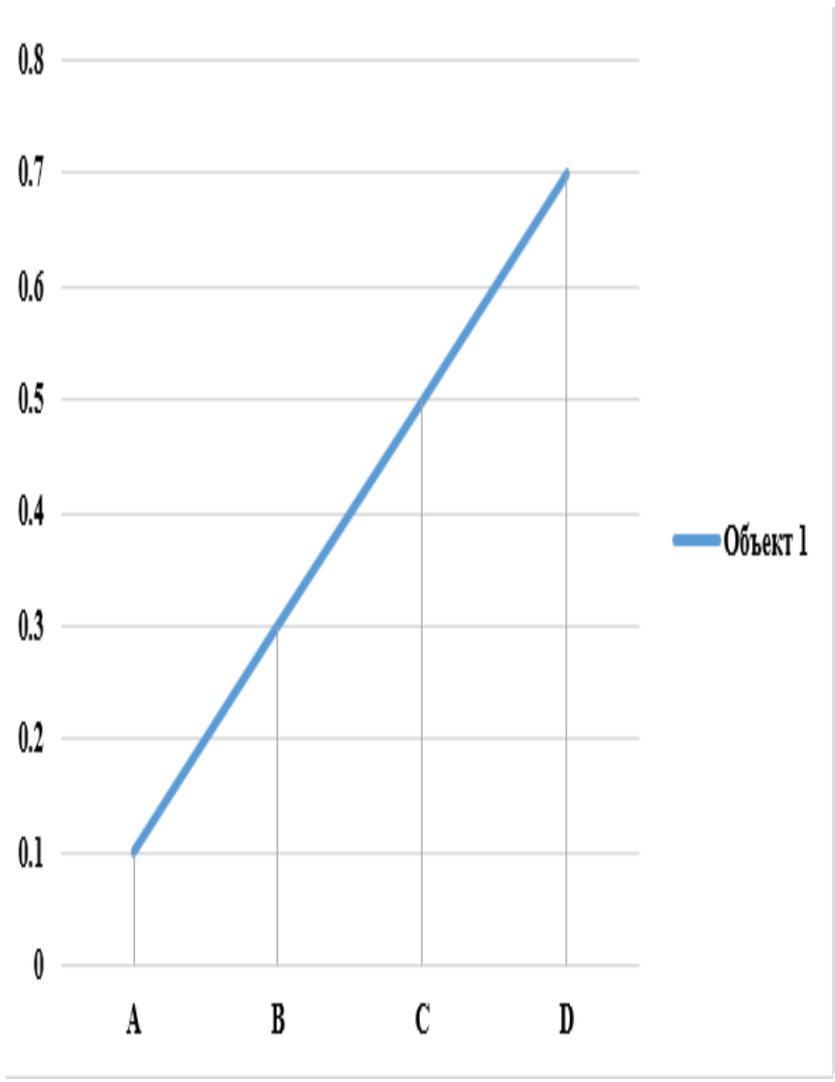
1. S – *wining coalitions*;
2. $S \backslash \{i\}$ $is\ not\ wining\ coalition$.

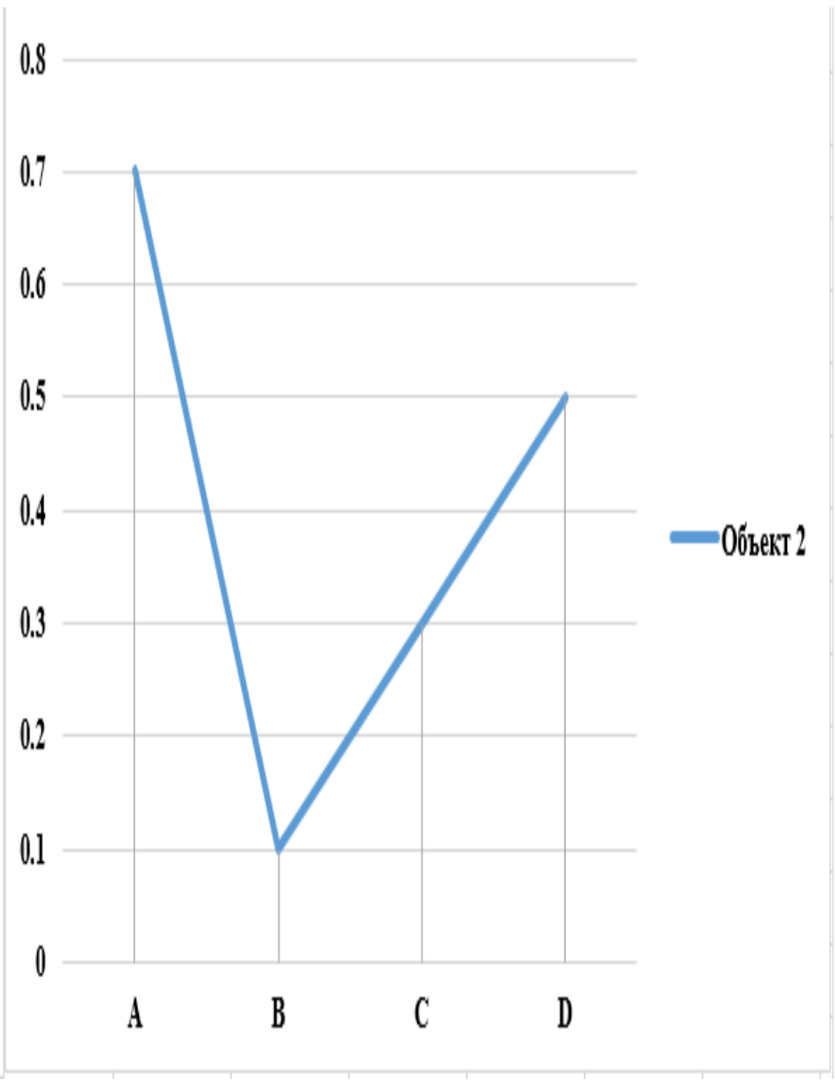
Banzhaf Index

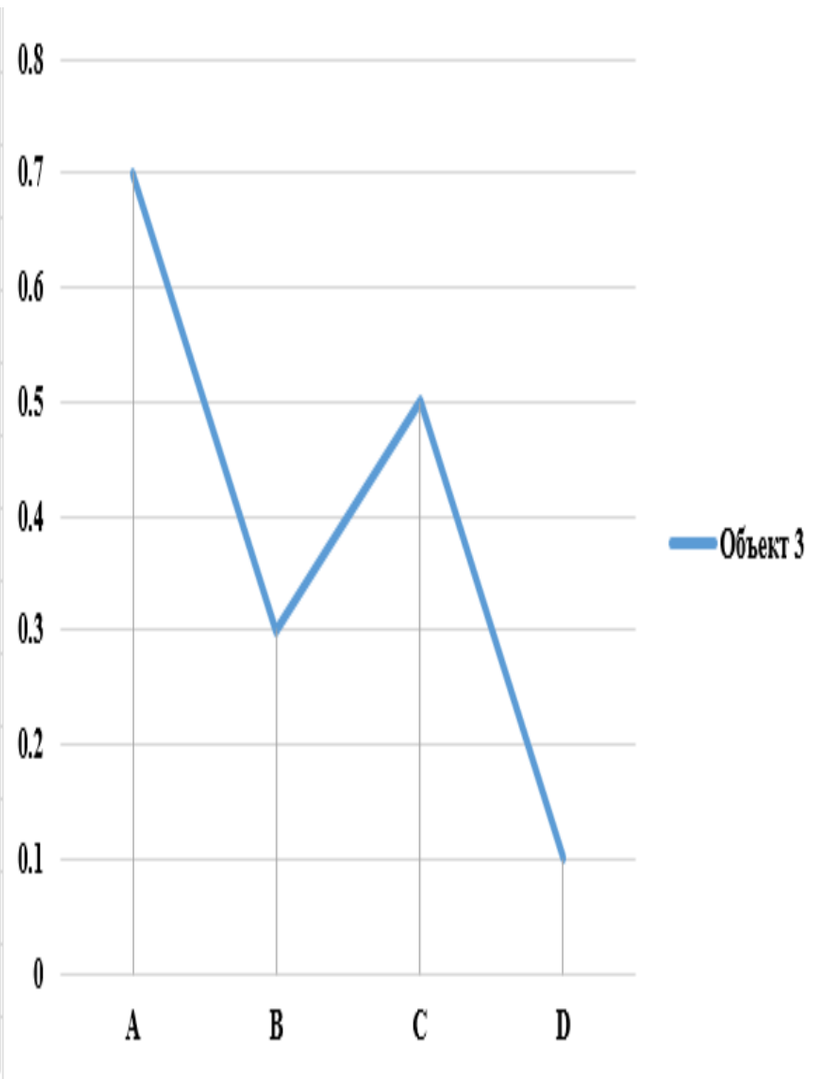


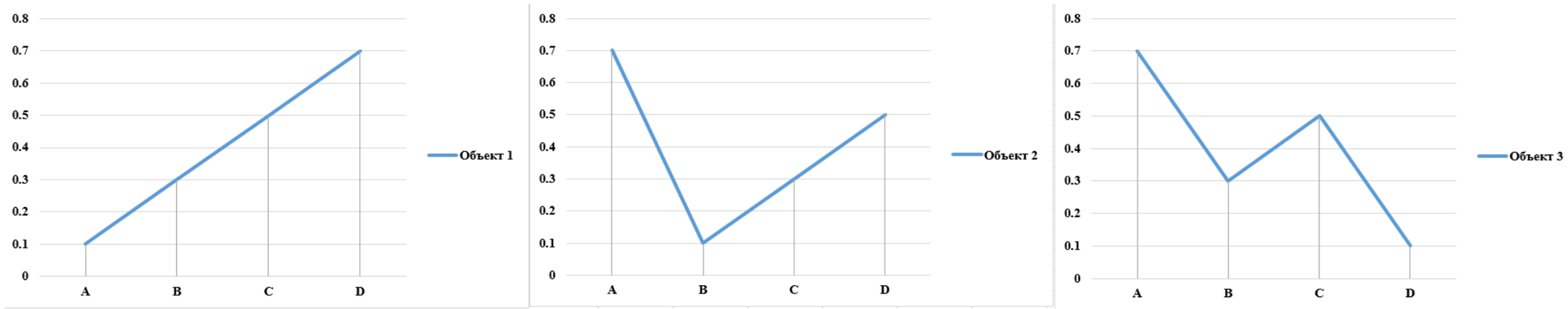
$$\beta(i) = \frac{b_i}{\sum_j b_j}$$

$$\beta(A) = \frac{3}{5}; \beta(B) = \frac{1}{5}; \beta(C) = \frac{1}{5}.$$


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
A
B
C
D
Объект 1


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
A
B
C
D
Объект 2


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
A
B
C
D
Объект 3



AB: $\{x_1\}, \{x_2, x_3\}$;

AC: $\{x_1\}, \{x_2, x_3\}$;

AD: $\{x_1\}, \{x_2, x_3\}$;

BC: $\{x_1, x_2, x_3\}$;

BD: $\{x_1, x_2\}, \{x_3\}$;

CD: $\{x_1, x_2\}, \{x_3\}$;

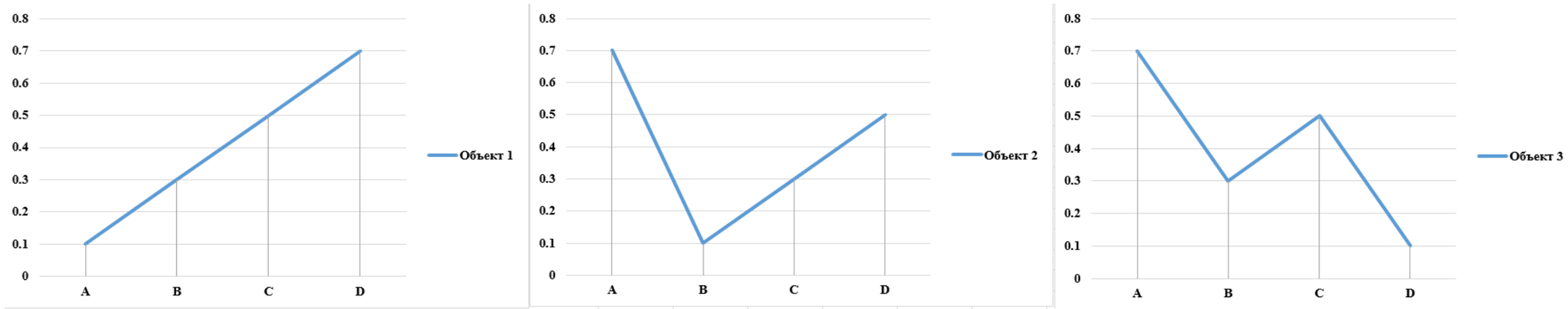
ABC: $\{x_1\}, \{x_2, x_3\}$;

ABD: $\{x_1\}, \{x_2\}, \{x_3\}$;

ACD: $\{x_1\}, \{x_2\}, \{x_3\}$;

BCD: $\{x_1, x_2\}, \{x_3\}$;

ABCD: $\{x_1\}, \{x_2\}, \{x_3\}$.



AB: $\{x_1\},\{x_2,x_3\}$;

AC: $\{x_1\},\{x_2,x_3\}$;

AD: $\{x_1\},\{x_2,x_3\}$;

BC: $\{x_1,x_2,x_3\}$;

BD: $\{x_1,x_2\},\{x_3\}$;

CD: $\{x_1,x_2\},\{x_3\}$;

ABC: $\{x_1\},\{x_2,x_3\}$;

ABD: $\{x_1\},\{x_2\},\{x_3\}$;

ACD: $\{x_1\},\{x_2\},\{x_3\}$;

BCD: $\{x_1,x_2\},\{x_3\}$;

ABCD: $\{x_1\},\{x_2\},\{x_3\}$.

$\mu(A) = \frac{6}{20}$ $\mu(C) = \frac{4}{20}$

$\mu(B) = \frac{4}{20}$ $\mu(D) = \frac{6}{20}$

THANK YOU!

NATIONAL RESEARCH UNIVERSITY

amyachin@hse.ru
www.hse.ru