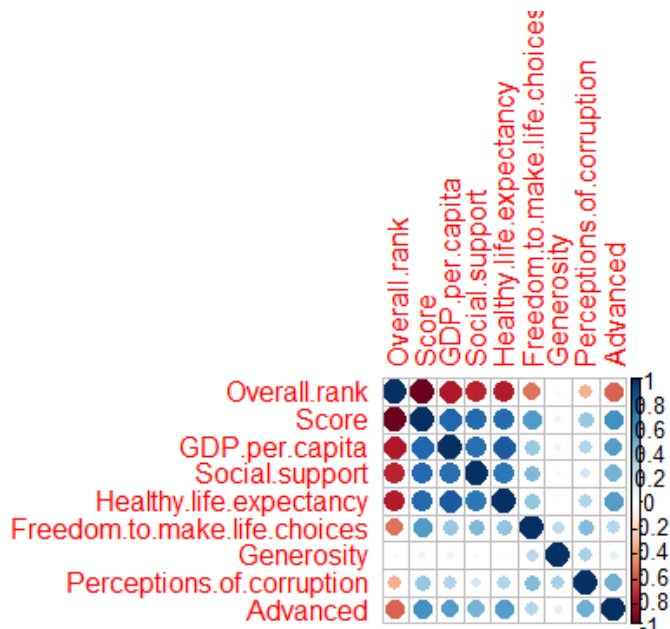


Исследование характеристик Польши по методу ближайшего соседа (KNN)

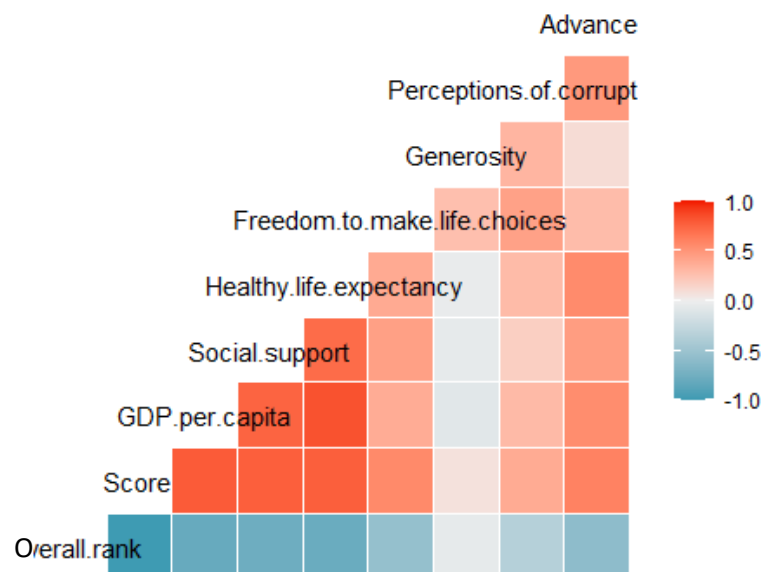
1. Описательная статистика:

Overall.rank	Country.or.region	Score	GDP.per.capita
Min. : 1.00	Afghanistan: 1	Min. :2.853	Min. :0.0000
1st Qu.: 39.75	Albania : 1	1st Qu.:4.545	1st Qu.:0.6028
Median : 78.50	Algeria : 1	Median :5.380	Median :0.9600
Mean : 78.50	Argentina : 1	Mean :5.407	Mean :0.9051
3rd Qu.:117.25	Armenia : 1	3rd Qu.:6.184	3rd Qu.:1.2325
Max. :156.00	Australia : 1	Max. :7.769	Max. :1.6840
	(Other) :150		
Social.support	Healthy.life.expectancy	Freedom.to.make.life.choices	
Min. :0.000	Min. :0.0000	Min. :0.0000	
1st Qu.:1.056	1st Qu.:0.5477	1st Qu.:0.3080	
Median :1.272	Median :0.7890	Median :0.4170	
Mean :1.209	Mean :0.7252	Mean :0.3926	
3rd Qu.:1.452	3rd Qu.:0.8818	3rd Qu.:0.5072	
Max. :1.624	Max. :1.1410	Max. :0.6310	
Generosity	Perceptions.of.corruption	Advanced	
Min. :0.0000	Min. :0.0000	Min. :0.0000	
1st Qu.:0.1087	1st Qu.:0.0470	1st Qu.:0.0000	
Median :0.1775	Median :0.0855	Median :0.0000	
Mean :0.1848	Mean :0.1106	Mean :0.1987	
3rd Qu.:0.2482	3rd Qu.:0.1412	3rd Qu.:0.0000	
Max. :0.5660	Max. :0.4530	Max. :1.0000	



	vars	n	mean	sd	median	trimmed	mad
Overall.rank	1	156	78.50	45.18	78.50	78.50	57.82
Country.or.region*	2	156	78.50	45.18	78.50	78.50	57.82
Score	3	156	5.41	1.11	5.38	5.41	1.21
GDP.per.capita	4	156	0.91	0.40	0.96	0.92	0.45
Social.support	5	156	1.21	0.30	1.27	1.24	0.28
Healthy.life.expectancy	6	156	0.73	0.24	0.79	0.74	0.26

Freedom.to.make.life.choices	7	156	0.39	0.14	0.42	0.40	0.15
Generosity	8	156	0.18	0.10	0.18	0.18	0.10
Perceptions.of.corruption	9	156	0.11	0.09	0.09	0.09	0.07
Advanced	10	156	0.20	0.40	0.00	0.13	0.00
	min	max	range	skew	kurtosis	se	
Overall.rank	1.00	156.00	155.00	0.00	-1.22	3.62	
Country.or.region*	1.00	156.00	155.00	0.00	-1.22	3.62	
Score	2.85	7.77	4.92	0.01	-0.66	0.09	
GDP.per.capita	0.00	1.68	1.68	-0.38	-0.81	0.03	
Social.support	0.00	1.62	1.62	-1.11	1.10	0.02	
Healthy.life.expectancy	0.00	1.14	1.14	-0.60	-0.37	0.02	
Freedom.to.make.life.choices	0.00	0.63	0.63	-0.67	-0.14	0.01	
Generosity	0.00	0.57	0.57	0.73	1.05	0.01	
Perceptions.of.corruption	0.00	0.45	0.45	1.62	2.23	0.01	
Advanced	0.00	1.00	1.00	1.50	0.24	0.03	



2. Пять ближайших соседей Польши при помощи алгоритма kd-tree:

Ответ: Panama, Czech Republic, Argentina, Uruguay, Portugal
 Развитые соседи: Czech Republic, Portugal

3. Расчёт и сравнение времени работы

Полный перебор:

Алгоритм kd-tree: 0.01 sec elapsed

Алгоритм полный перебор (brute): 0.02 sec elapsed

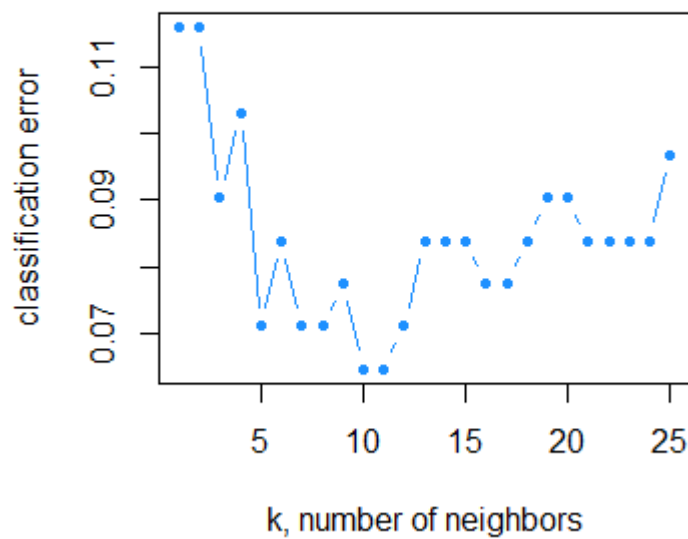
kd-tree оказался быстрее, что согласуется с логикой.

4. Согласно анализу, для Польши уровень счастья соотносится с уровнем её экономического развития. Класс страны – развивающаяся (вероятность 0,6), что совпадает с действительностью.

5. Ошибка прогноза класса составляет 0.07096774 (то есть правильно примерно 93%) для случая, когда выбирают пятерых соседей.

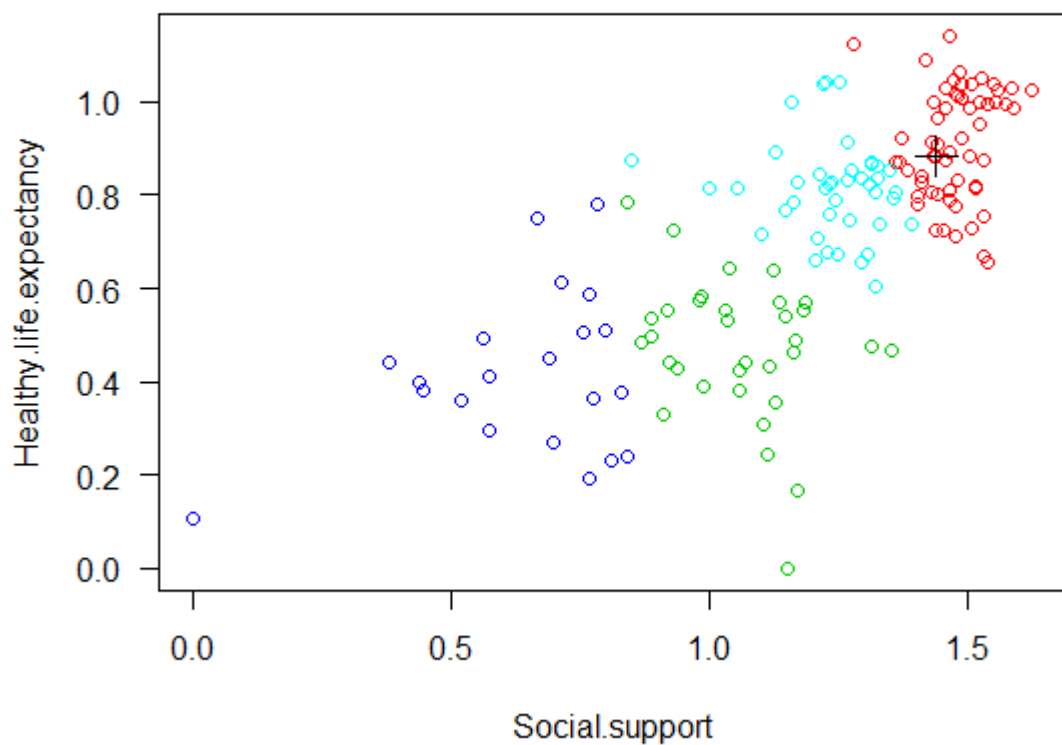
чтобы минимизировать ошибку прогноза класса нужно выбрать $k = 10$

(Test) Error Rate vs Neighbors

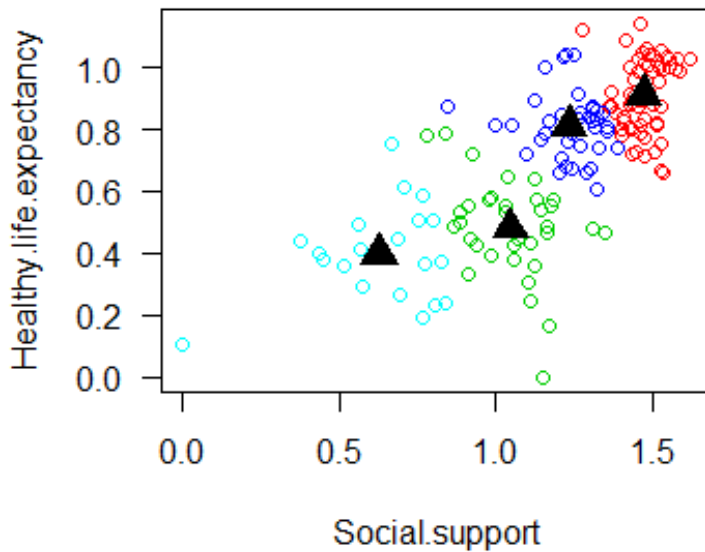


6. Кластеризовали страны на 4 кластера по компонентам "Социальная поддержка" и "Ожидание здоровой жизни". Содержательно описать кластеры можно следующим образом. Оба показателя символизируют счастье пропорционально, то есть чем выше каждый из показателей, тем лучше при прочих равных там жить. Заметно, что кластеры рассредоточены пропорционально увеличению каждого из параметров.

k-means result with 4 clusters



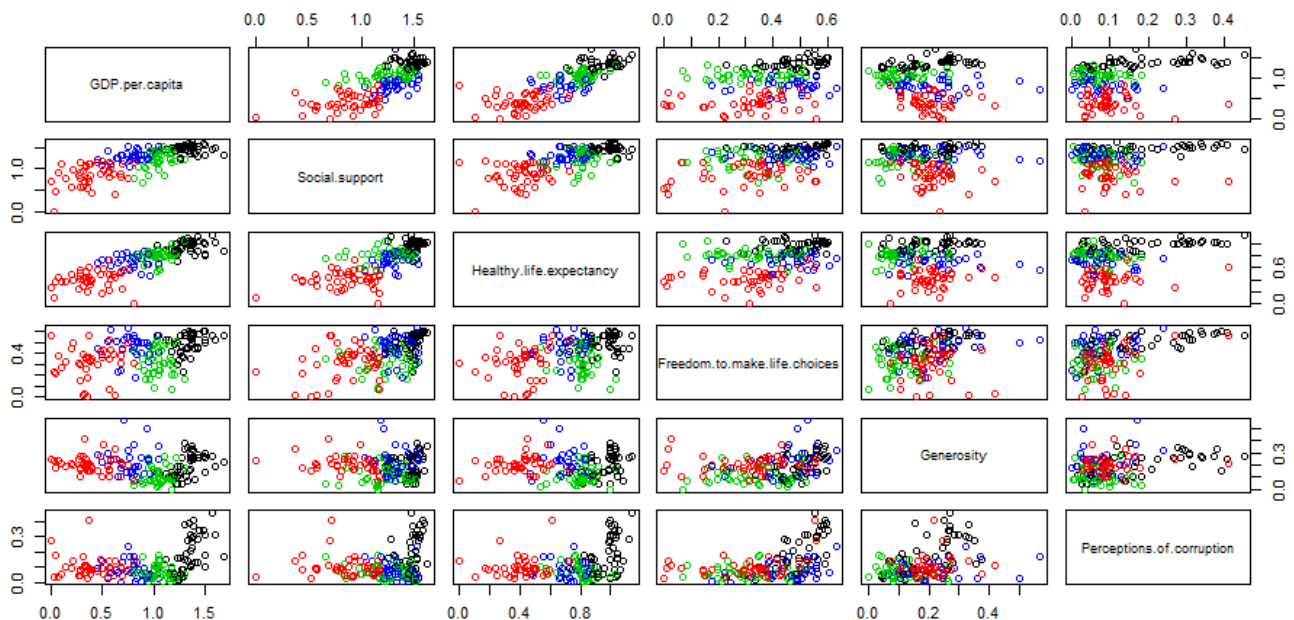
k-means result with 4 clusters



К центрам кластеров оказались ближе всего следующие типичные страны представители: Haiti, Uganda, Czech Republic, Jordan.

Польша оказалась во втором (синем) кластере.

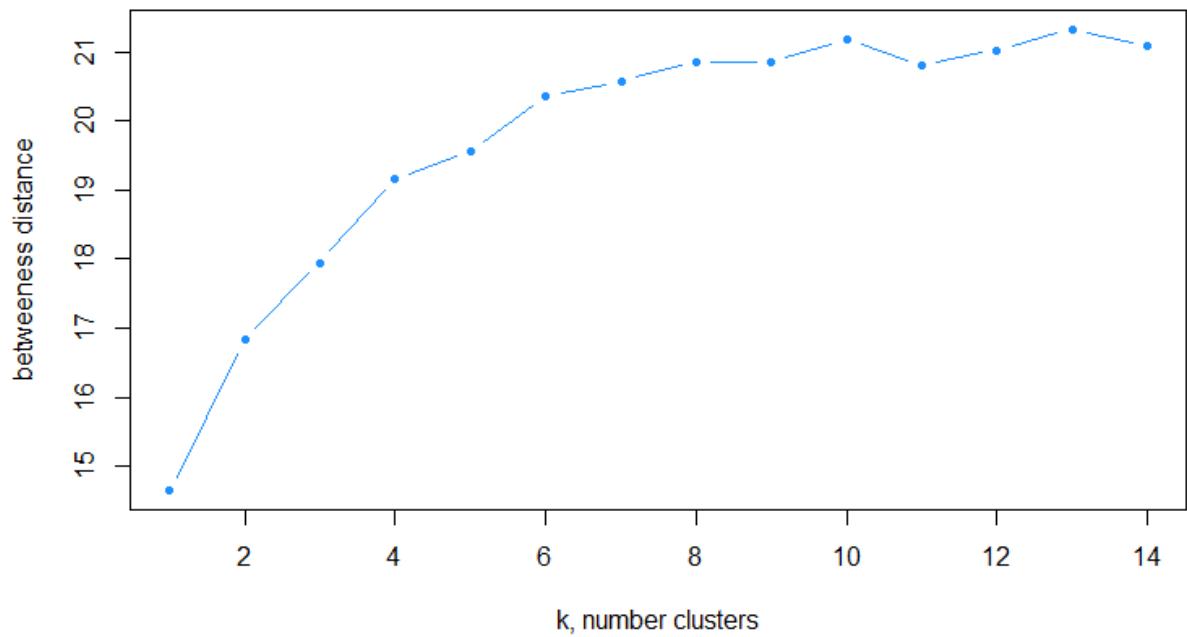
7. Кластеризация по всем компонентам индекса счастья.



Согласно такой кластеризации, Польша располагается в первом (чёрном) кластере. Если описывать кластер содержательно, то можно сказать, что по всем показателям данный кластер находится в более выигрышном положении и относительно остальных кластеров.

8. Минимизация суммарного внутригруппового расстояния.

(Test) betweenness distance vs Clusters



Согласно графику для минимизации суммарного внутригруппового расстояния достаточно разделять на 6 кластеров (далее прирост в расстоянии не такой большой).