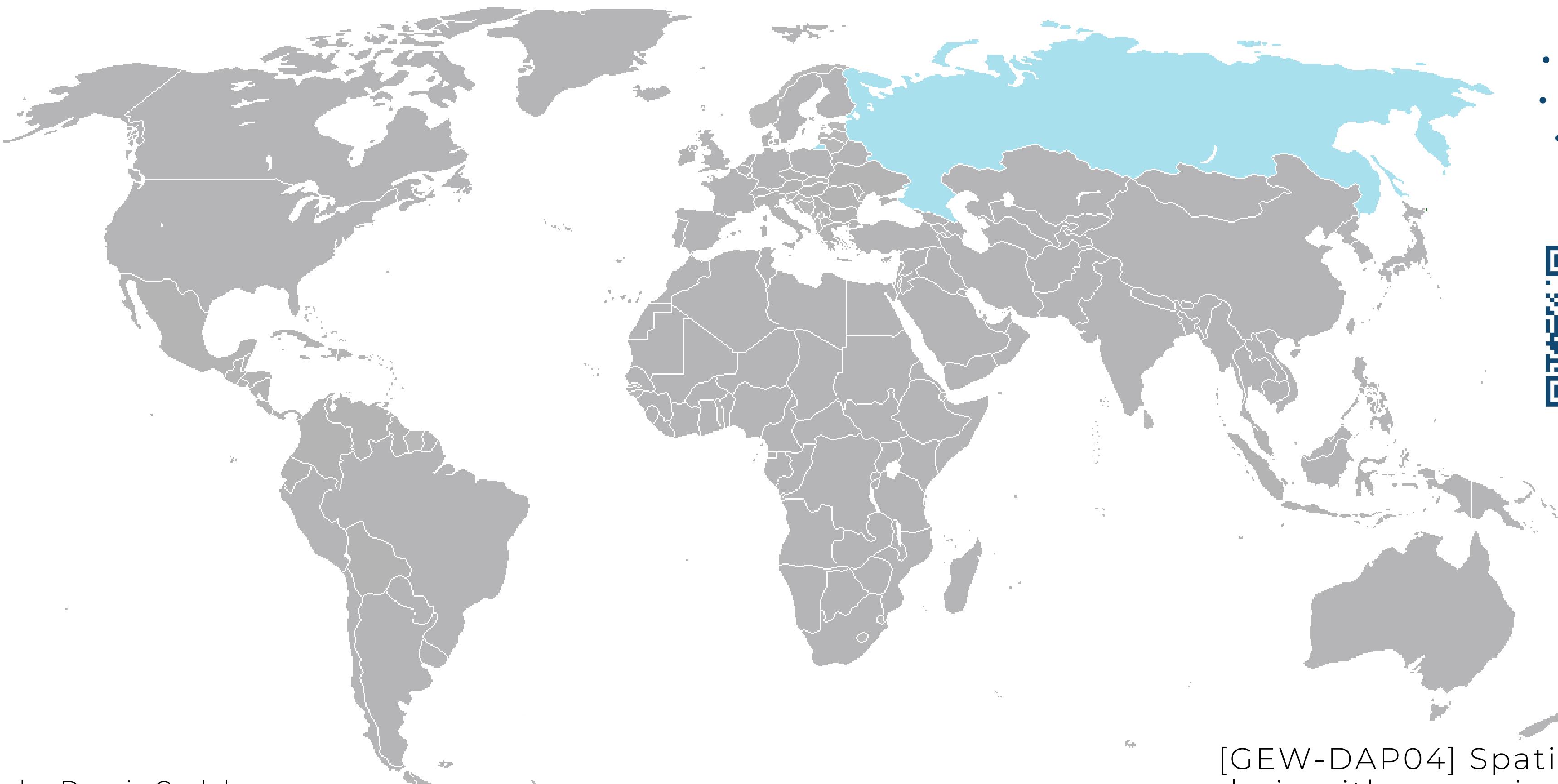


# Unsupervised Machine Learning clustering of Russian regions



[GEW-DAP04] Spatial data  
analysis with numerical methods

Motivation



85\* regions

pongkit

# Data sources

Russian Federal State  
Statistics Service



gks

Prosecutor General's Office of Russia



genproc.gov

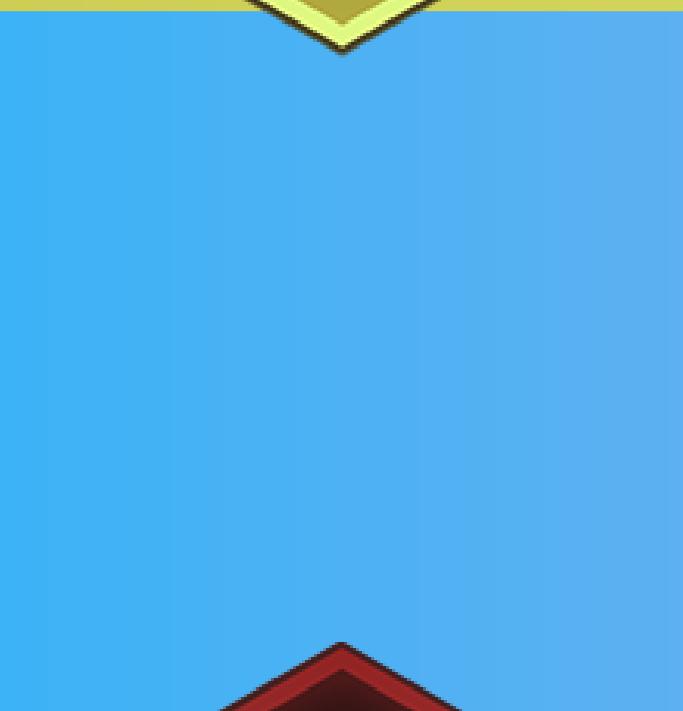


Gross regional product per capita, 2016, rubles

Median salary, 2019, rubles

Unemployment rate, 2017, %

Share of poor people, 2017, %



Share of people with higher education, 2010, %

Museum attendance, 2010, # / population

Fertility rate, 2017

Share of people using internet everyday, 2016, %



Criminal possession of weapons per 10k, 2017

Murders and attempted murders, 2017

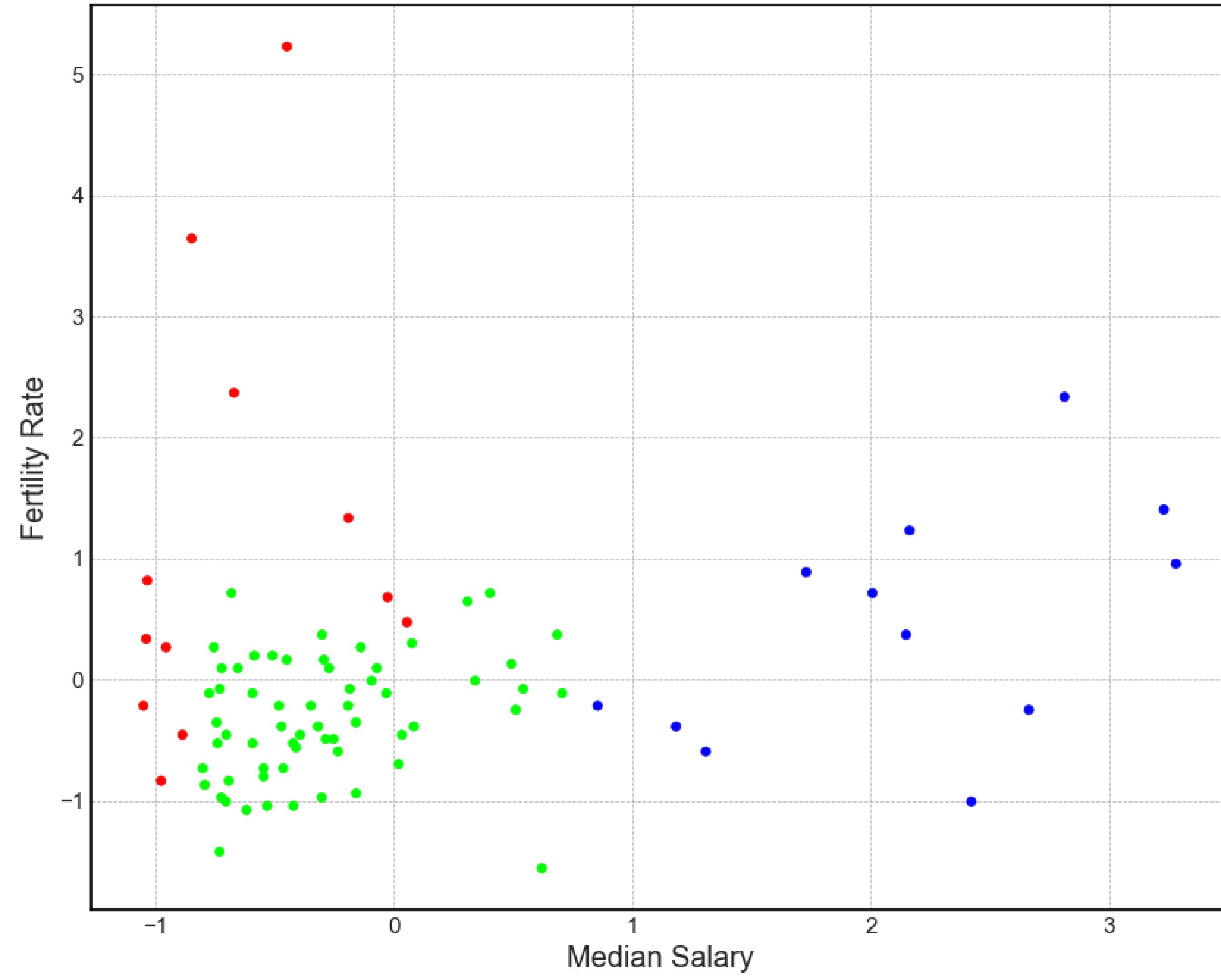
Terrorist crimes per 10k, 2017

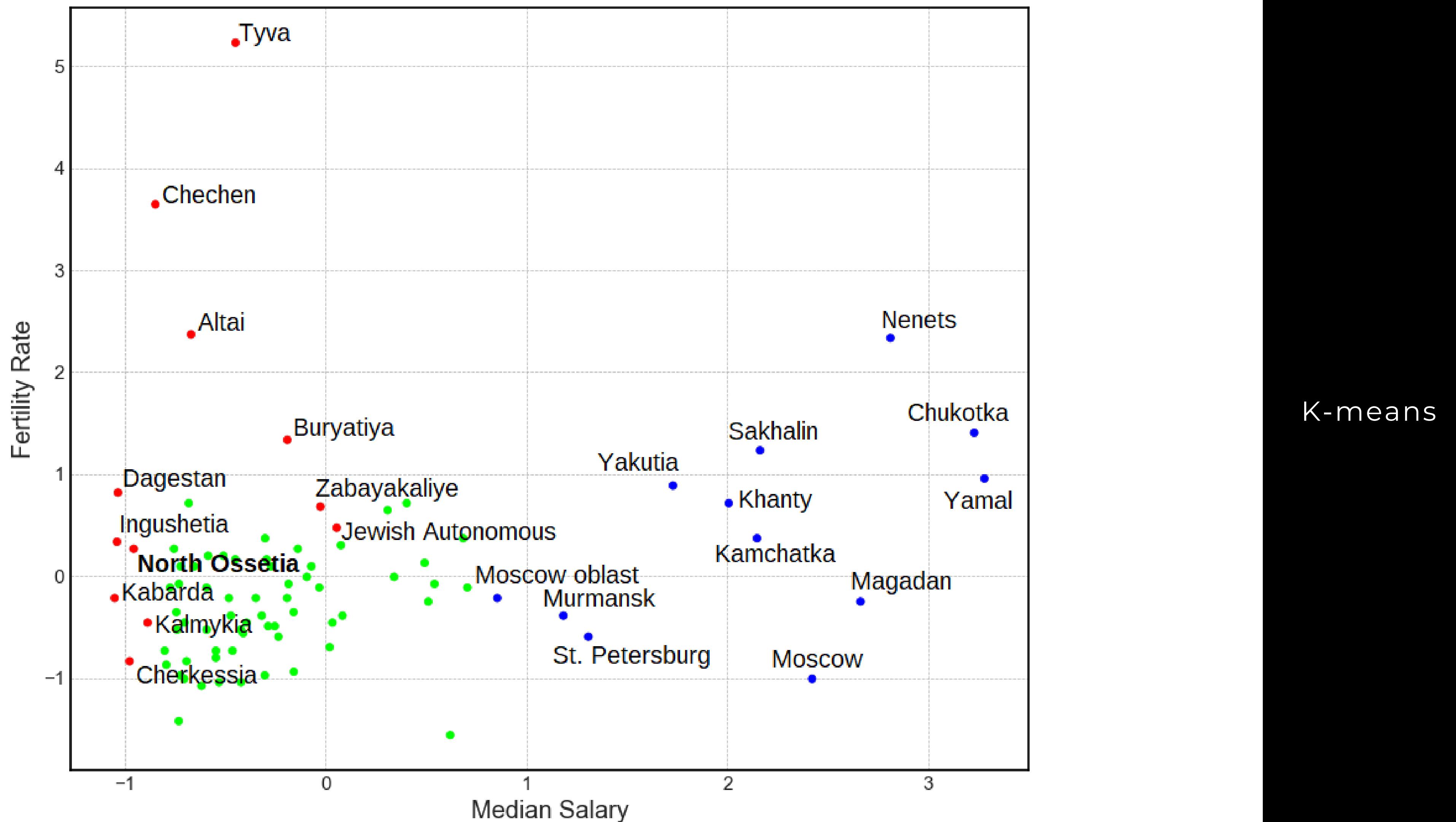
# Methods

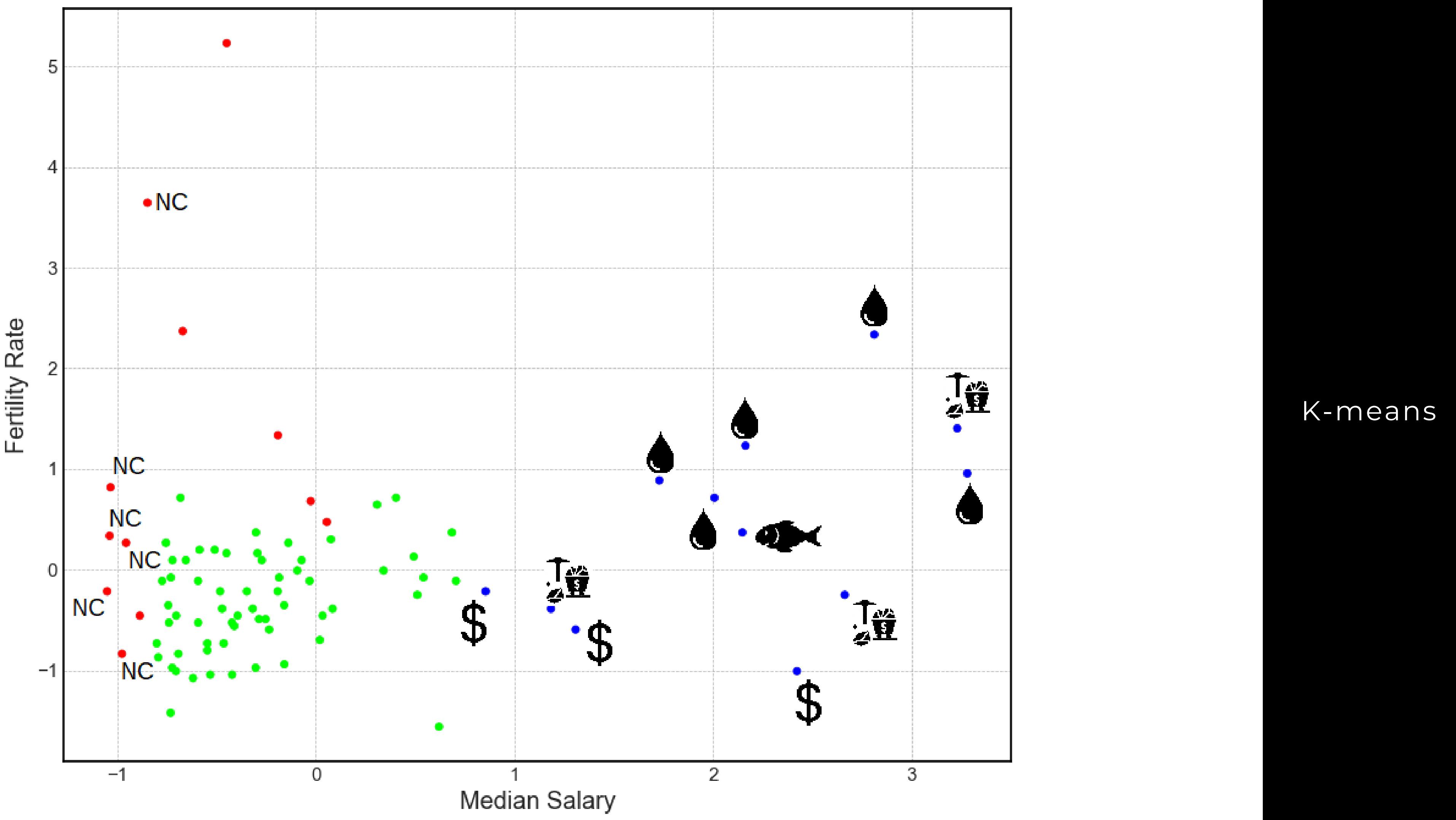
Clustering method	Usecase	Geometry (metric used)
K-Means	General-purpose, even cluster size, flat geometry, not too many clusters	Distances between points
Spectral clustering	Few clusters, even cluster size, non-flat geometry	Graph distance (e.g. nearest-neighbor graph)
Ward Tree	Many clusters, possibly connectivity constraints	Distances between points
Gaussian mixture	Flat geometry, good for density estimation	Mahalanobis distances to centers

# Results

K-means





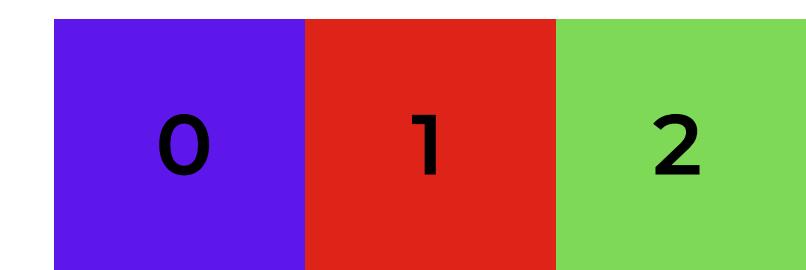


# Labels

Share of people with higher education. 2010. %	Median salary. 2019. rubles	Unemployment rate. 2017. %	Share of poor people. 2017. %	Gross regional product per capita. 2016. rubles	Terrorist crimes_2017	Murders and attempted murders. 2017	Criminal possession of weapons. 2017	Fertility rate. 2017	Museum attendance. 2010. per pop	Share of people using internet everyday, 2016, %
--	-----------------------------	----------------------------	-------------------------------	---	-----------------------	-------------------------------------	--------------------------------------	----------------------	----------------------------------	--

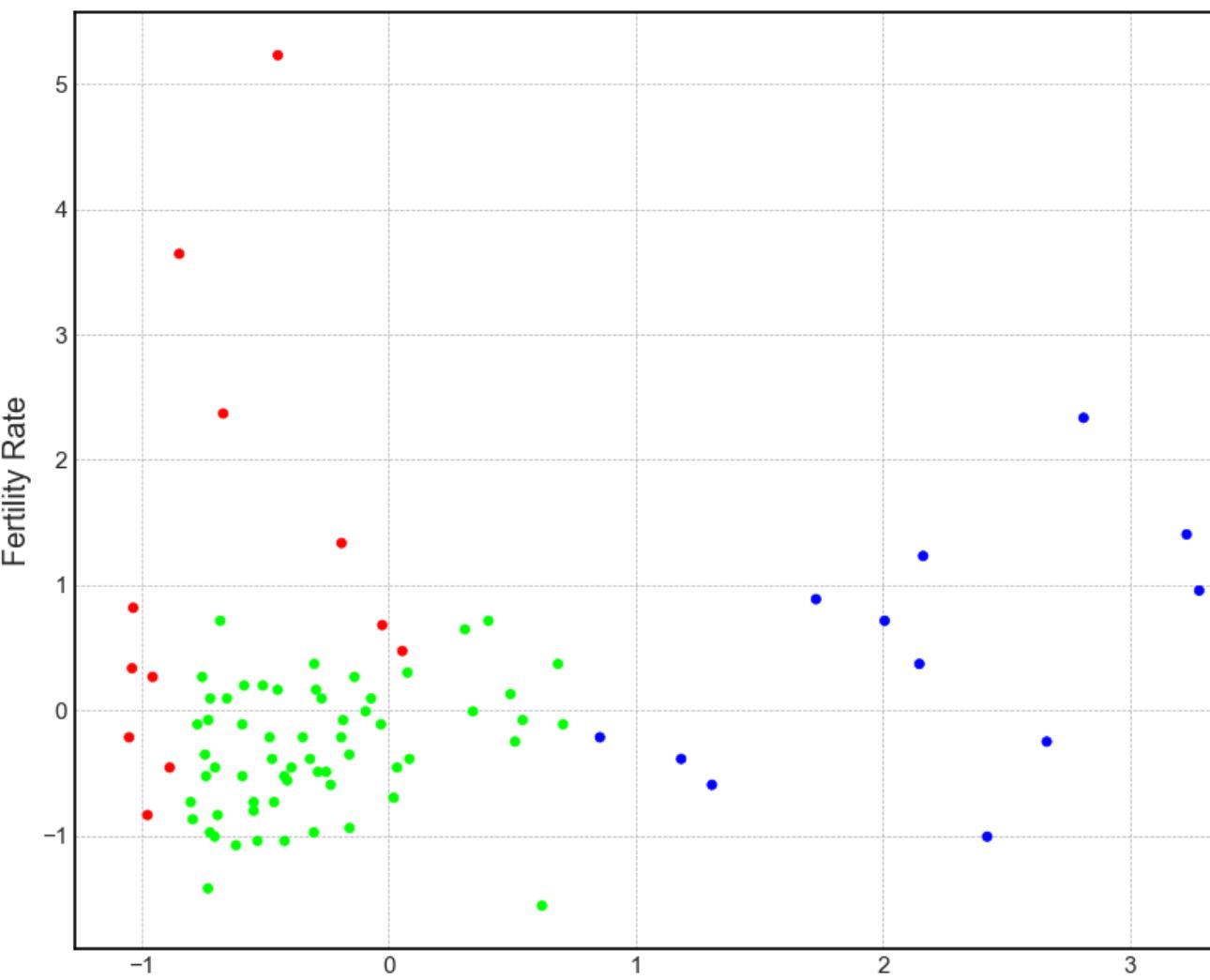
labels

0	25.000000	62474.250000	-0.555951	-0.648022	1.541223	-0.263642	0.358297	0.084501	0.454058	0.521514	1.509511
1	18.166667	24837.250000	1.878018	1.556692	-0.429212	1.573727	0.630475	1.471655	1.137867	-0.663739	-0.137616
2	19.672131	29913.229508	-0.260079	-0.178755	-0.218756	-0.257722	-0.194513	-0.306129	-0.313166	0.027979	-0.269881

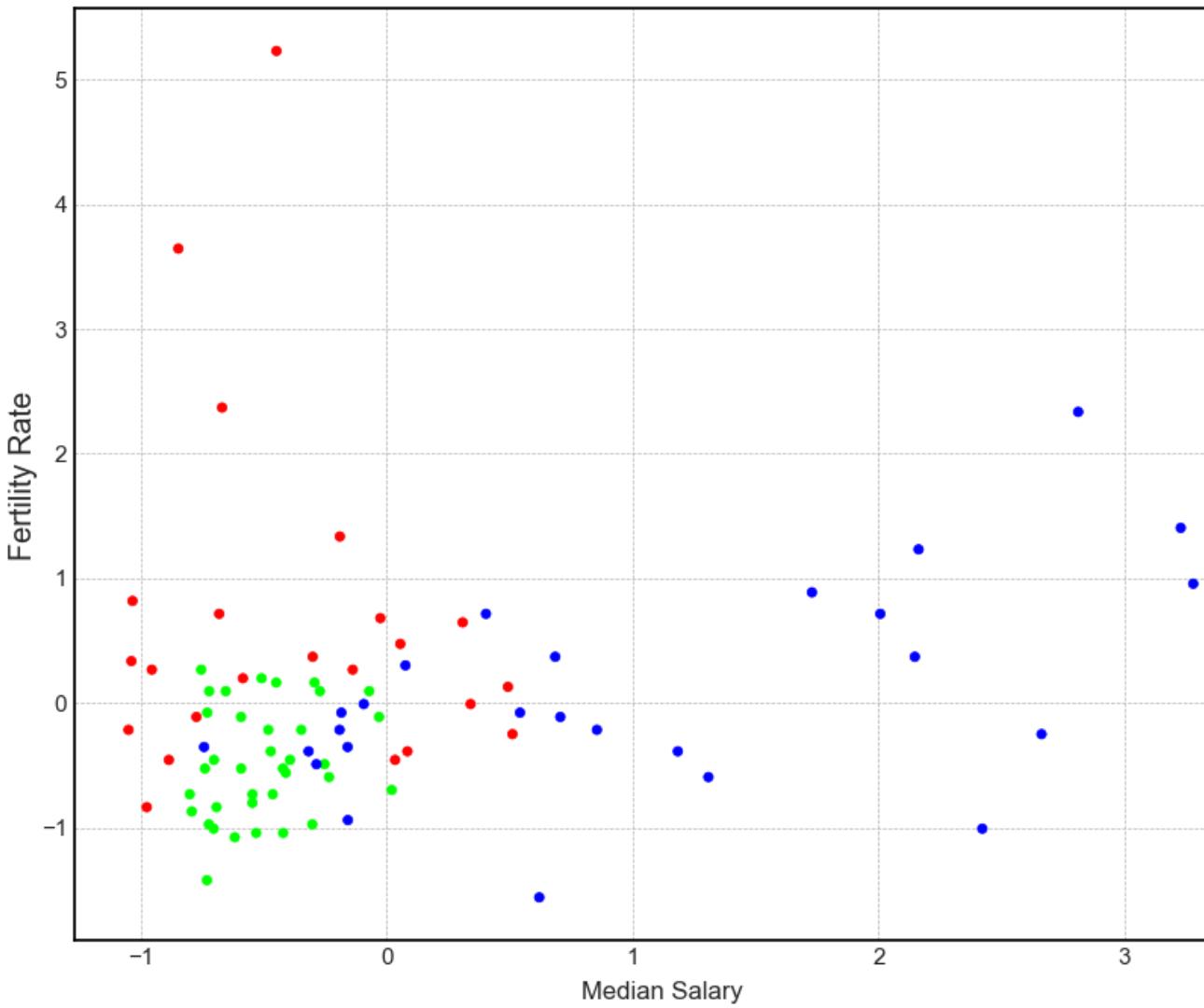




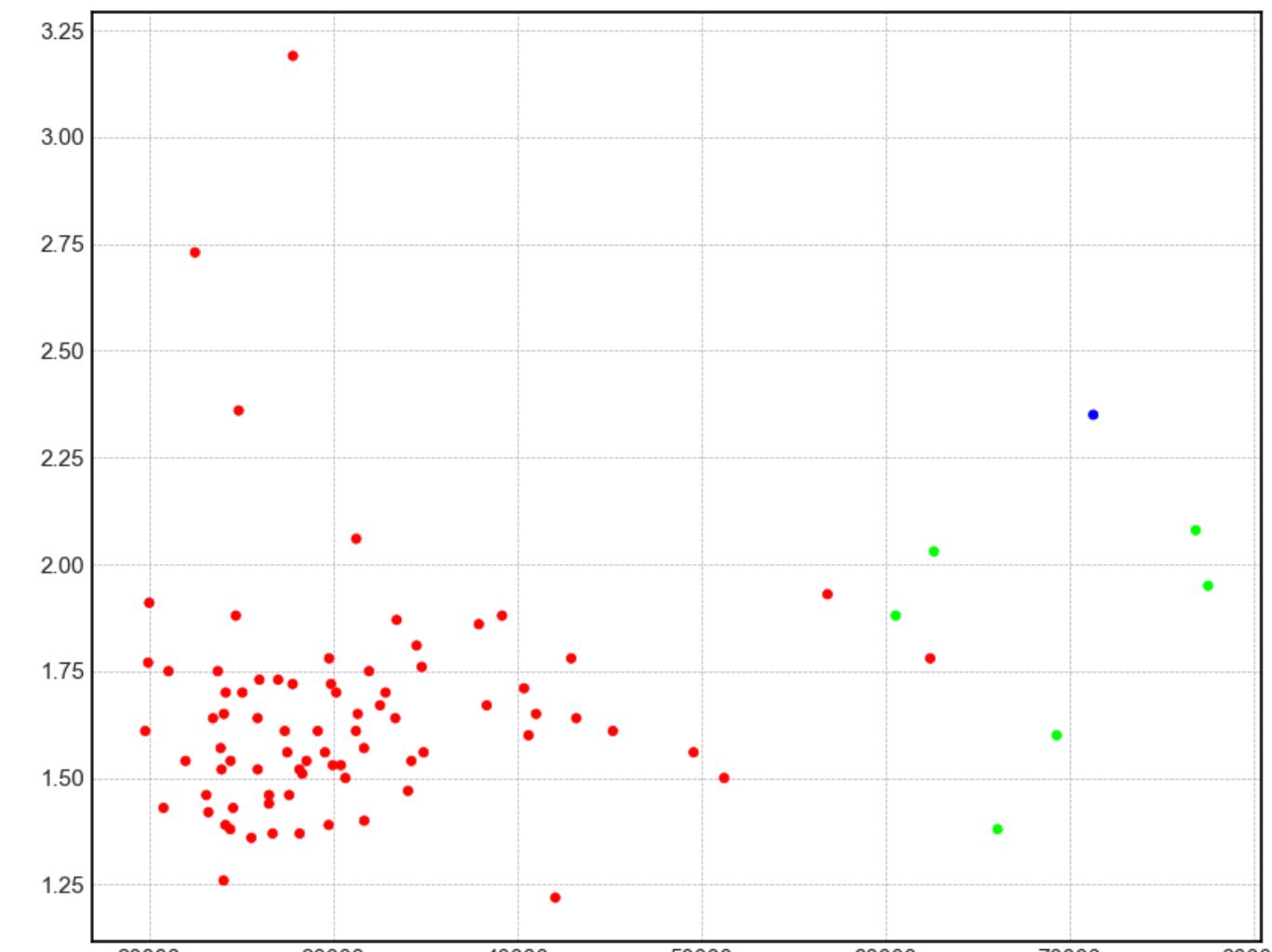
K-means



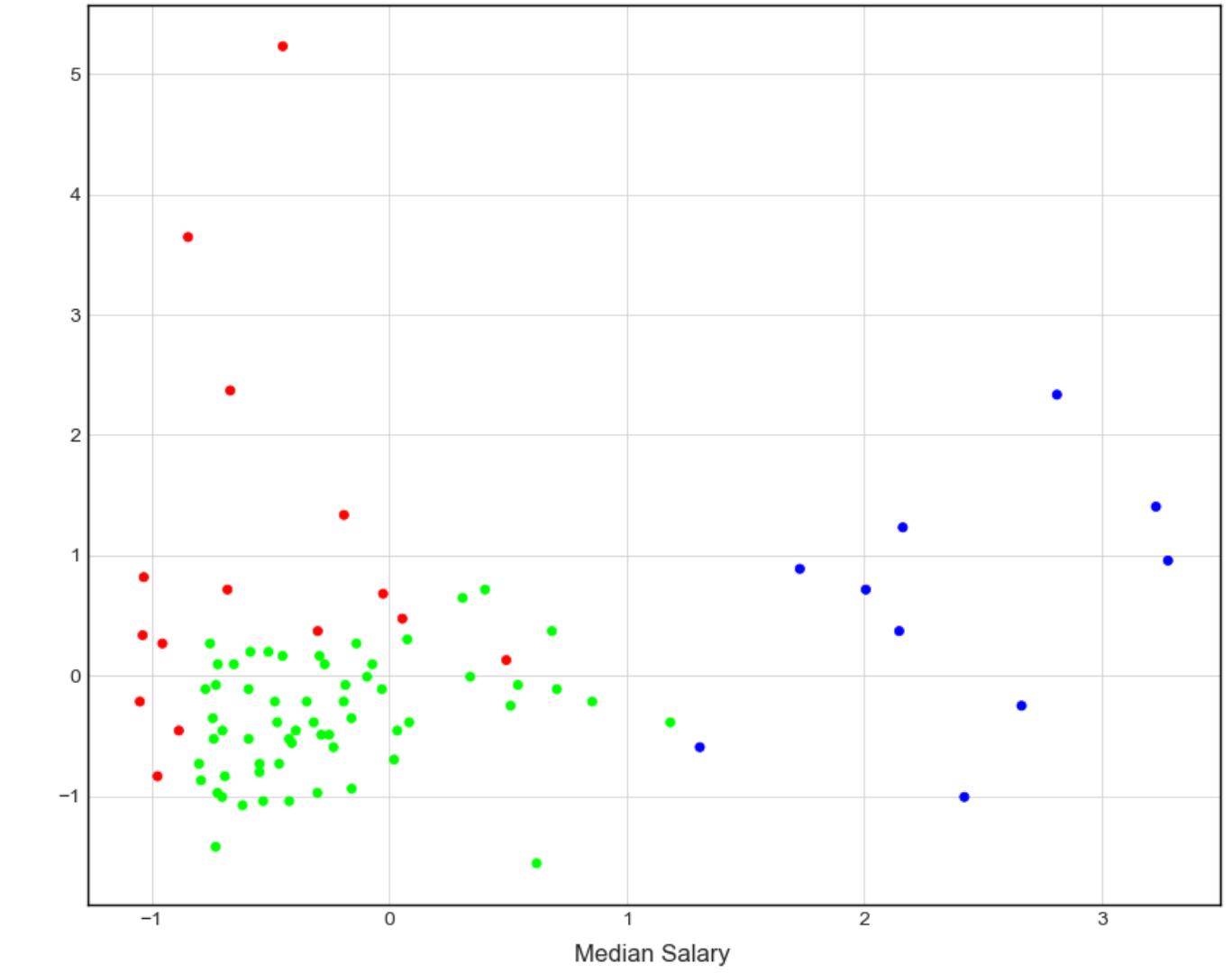
Spectral  
Clustering

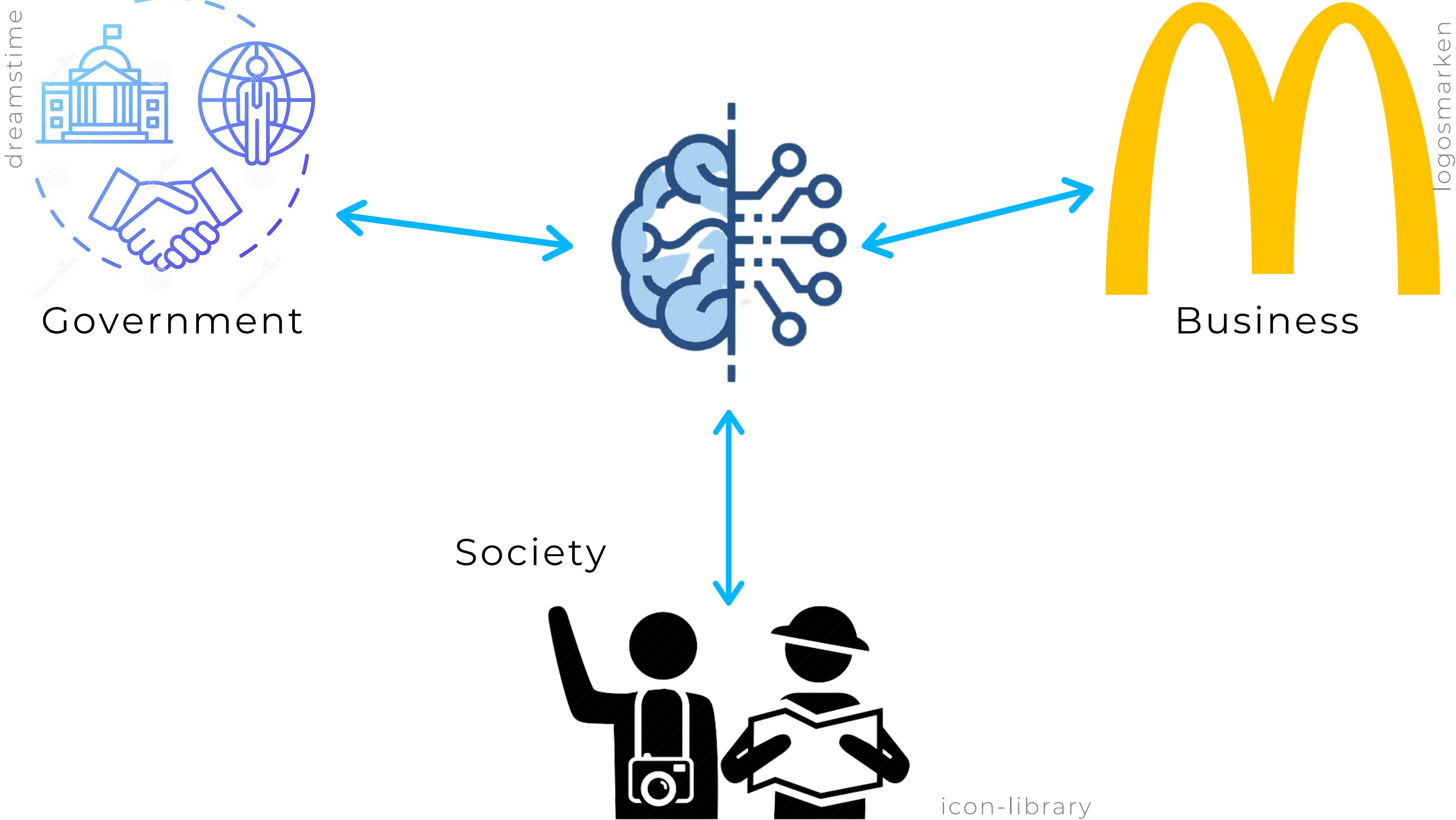


Gaussian  
Mixture

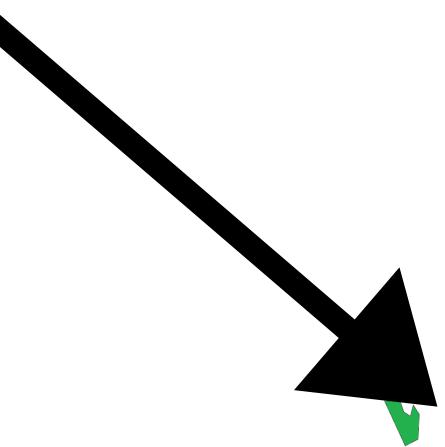


Ward  
Tree

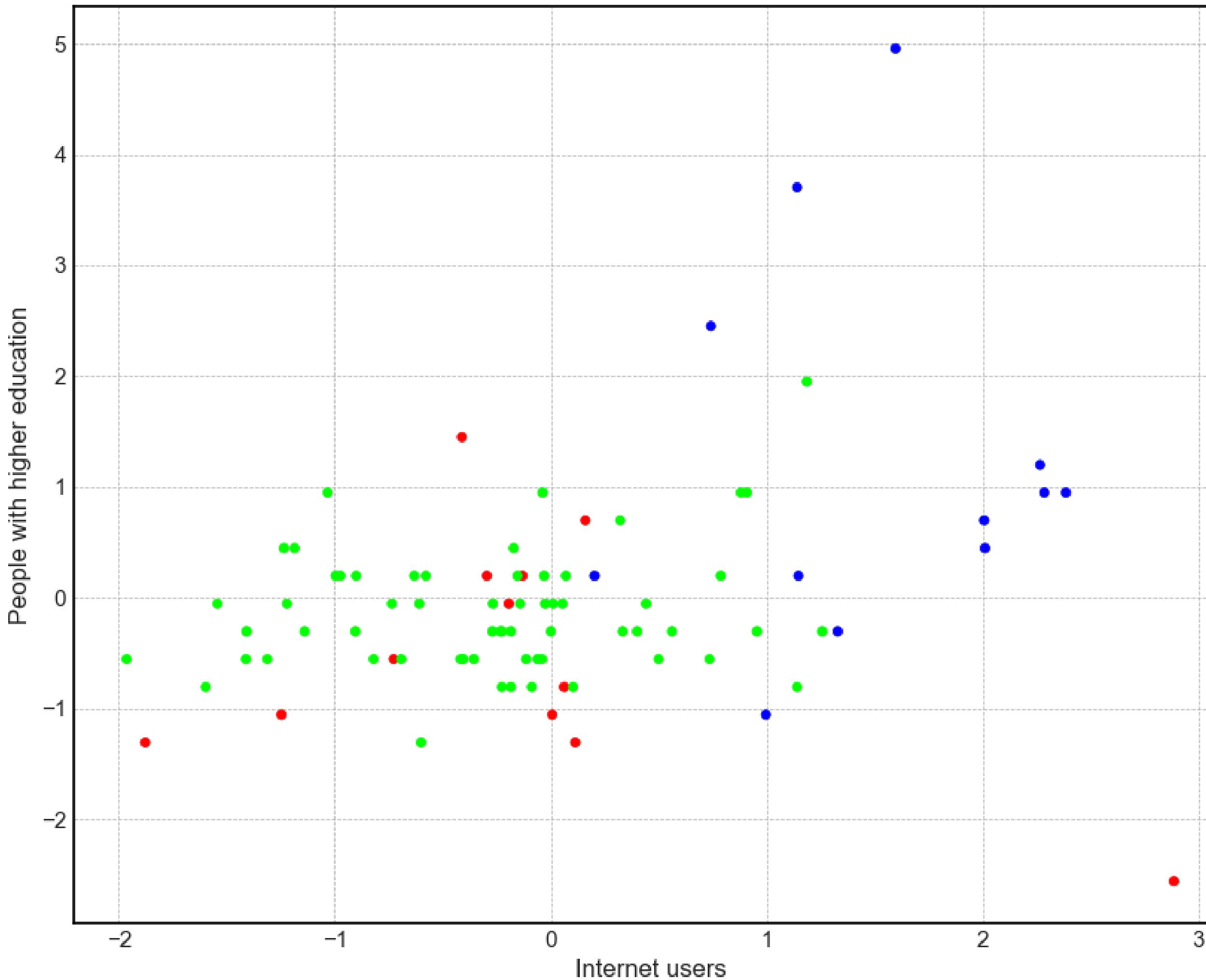




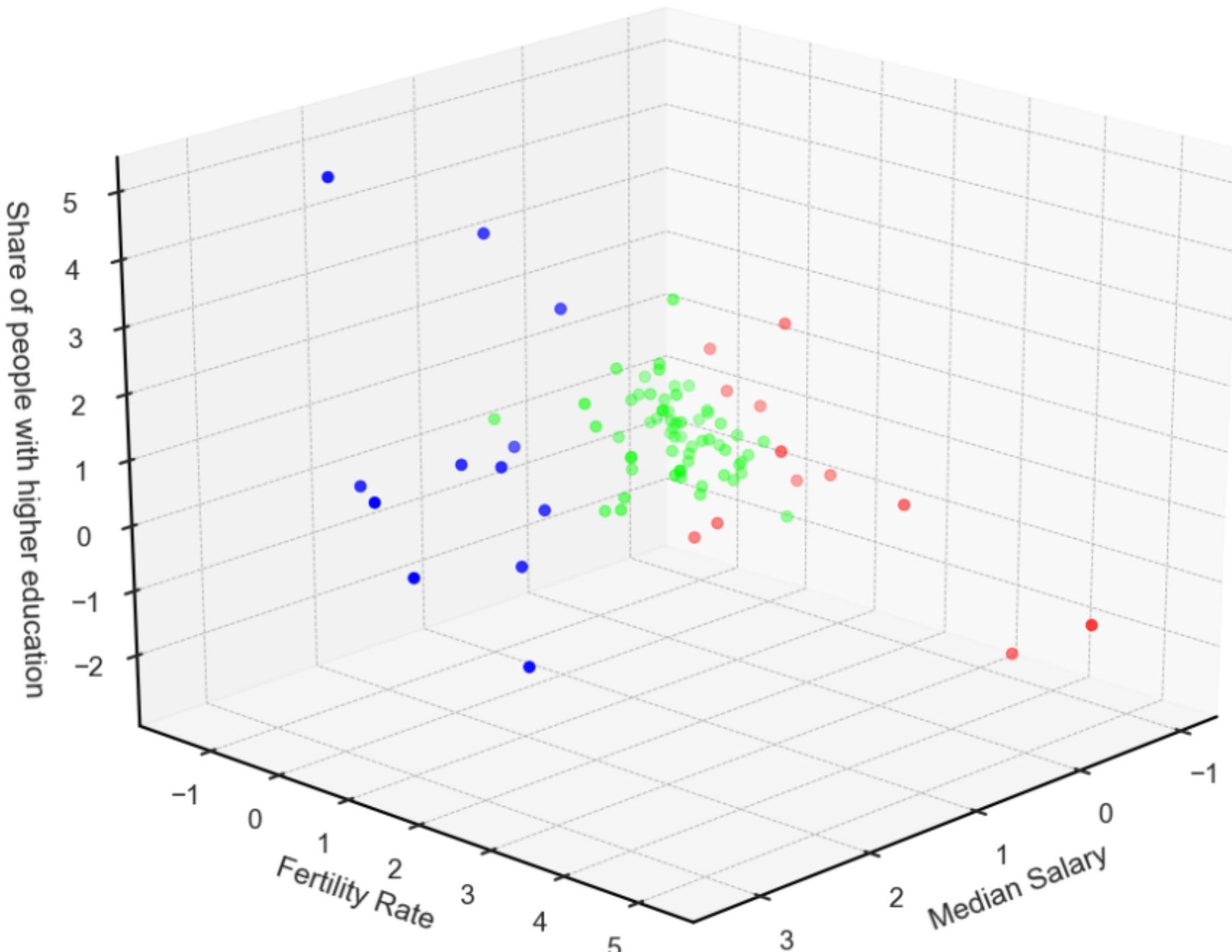
pikabu



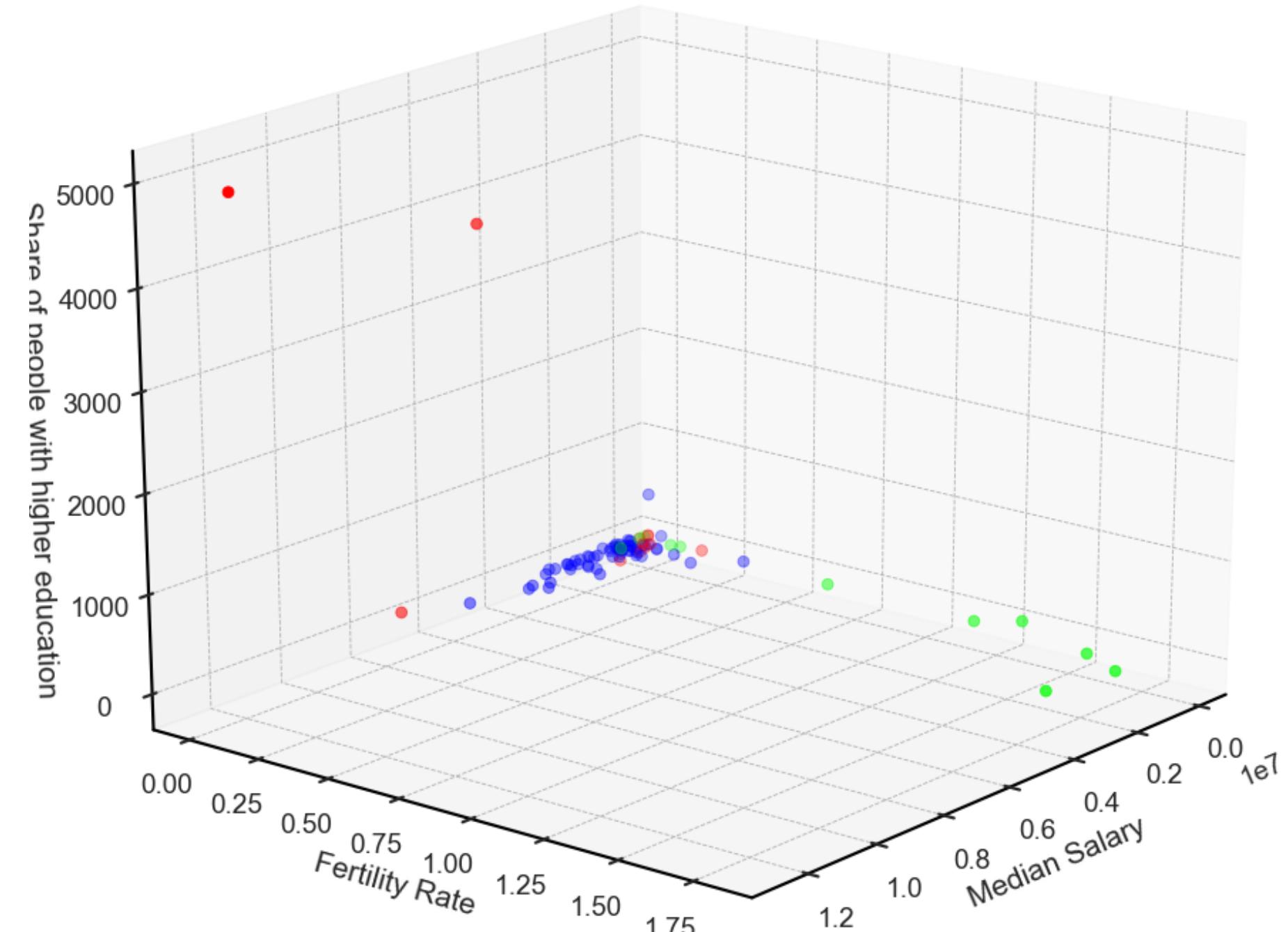
K-means Z score



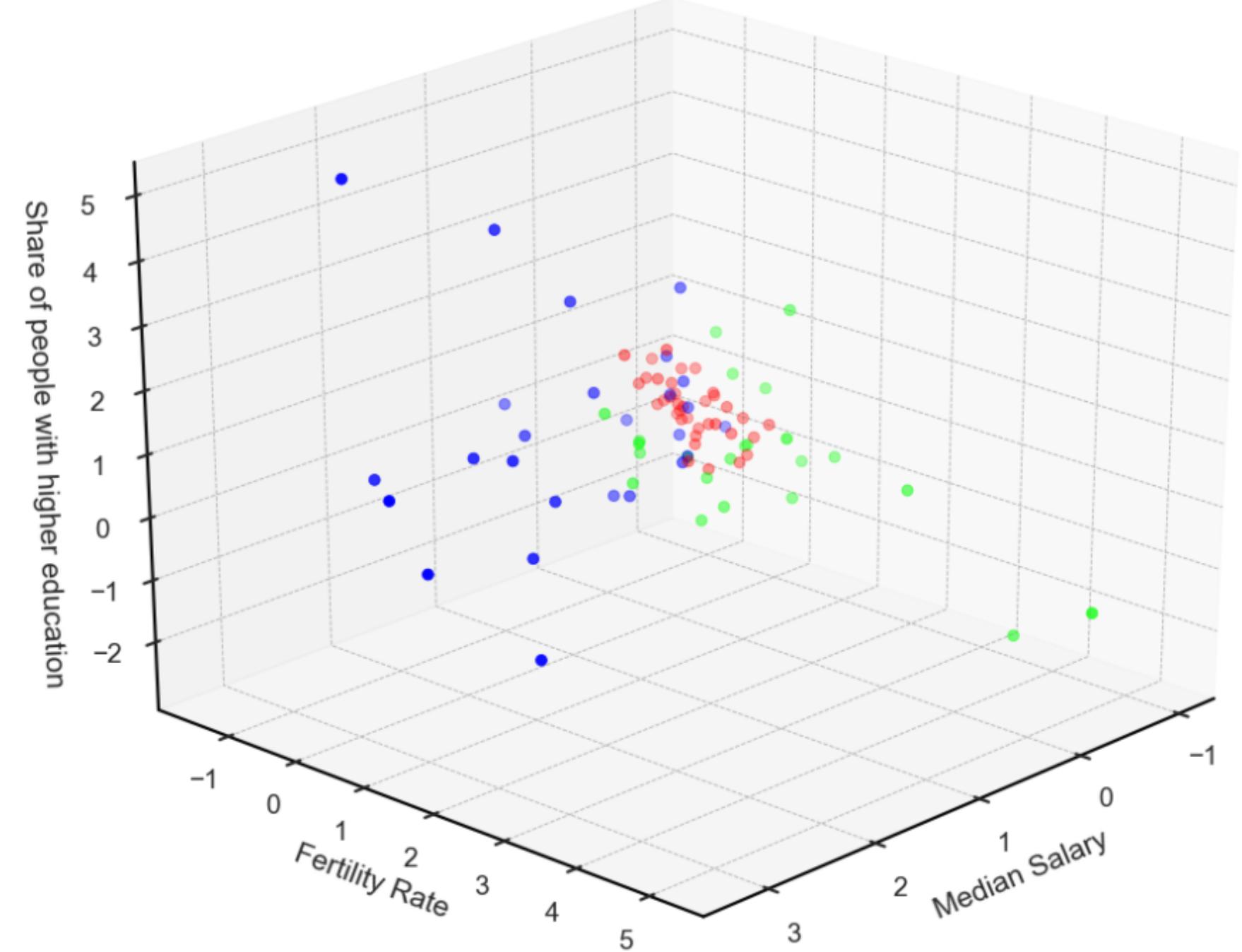
K-means Z score 3D



Gaussian mixture 3D



Spectral clustering 3D



Ward tree 3D

