ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ανοικτάμαθήματα
opencourses

Thematic cartography

Session 4: Thematic data classification methods – The role of color

Michail Agorastakis

Department of Planning and Regional Development

BY NC SA

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Θεσσαλίας**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Outline

- Thematic data classification methods
- The role of colour

Thematic data classification methods (1)

The objective of classification is to group (*classify*) features or data in such way that not only are the data/ features within a class similar but also the classes themselves are dissimilar.

Regarding thematic overlay, data classification…

"In more technical terms, the goal is to find the optimal number of classes—and where to put the breaks between those classes—so as to minimize within-group variance and maximize between-group differences."

source: http://axismaps.github.io/thematic-cartography/articles/classification.html

Thematic data classification methods (2)

The objective of classification is to group (*classify*) features or data in such way that not only are the data/ features within a class similar but also the classes themselves are dissimilar.

Regarding thematic overlay, data classification…

"*In more technical terms, the goal is to find the optimal number of classes—and where to put the breaks between those classes—so as to minimize within-group variance and maximize between-group differences.*" **source**: http://axismaps.github.io/thematic-cartography/articles/classification.html

Too few classes - strange patterns (loss of information)
Too many - confusion (too much "noise")
Difficult to recognize more than seven classes

Thematic data classification methods (3)

Common methods of thematic data classification

- Equal intervals
- Quantiles
- Mean – standard deviation
- Natural breaks
- Geometric interval

Equal intervals (1)

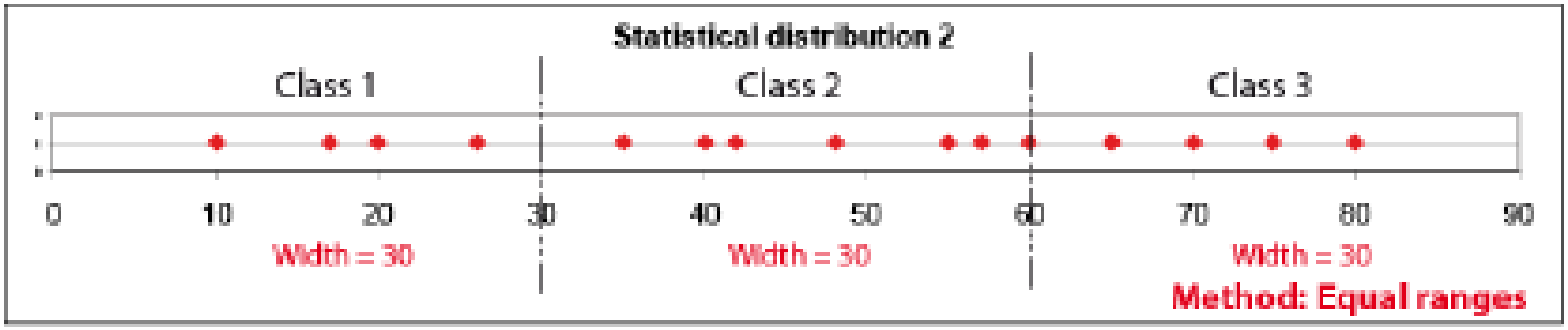
"Divides the data into equal size classes (e.g., 0-10, 10-20, 20-30, etc.) and works best on data that is generally spread across the entire range."

Source: http://axismaps.github.io/thematic-cartography/articles/classification.html

"The difference between the top and bottom values in each range is the same. This means that we can use values like 0-20; 20-40 etc. or calculate the width of the dataset, and divide by the number of classes wanted. In this case the lowest class will start with the lowest value; the width between the classes will be the same, and the top of the highest value in the dataset. This method is suited for datasets with a smooth linear distribution. If the method is used on dataset that are not linear distributed, you will have some classes with many values and others with few or no values."

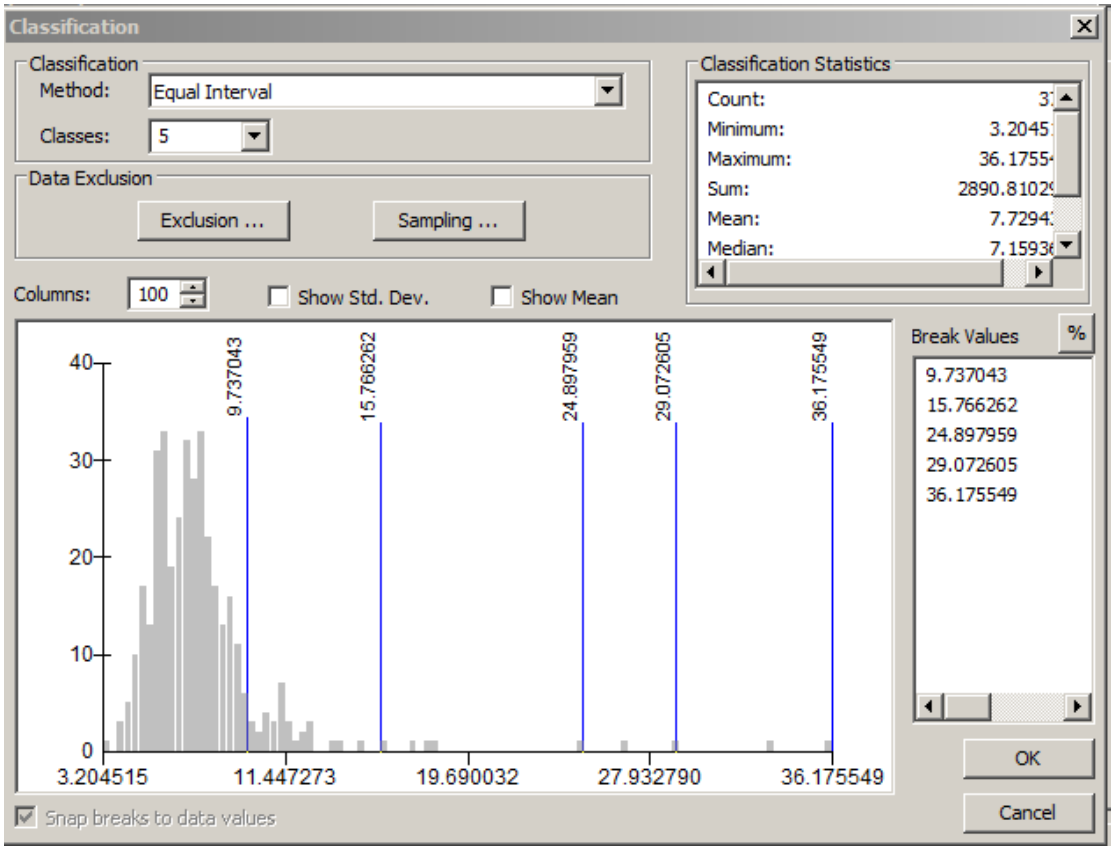
Source: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

Equal intervals (2)



Source: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

Comments??



Quantiles (1)

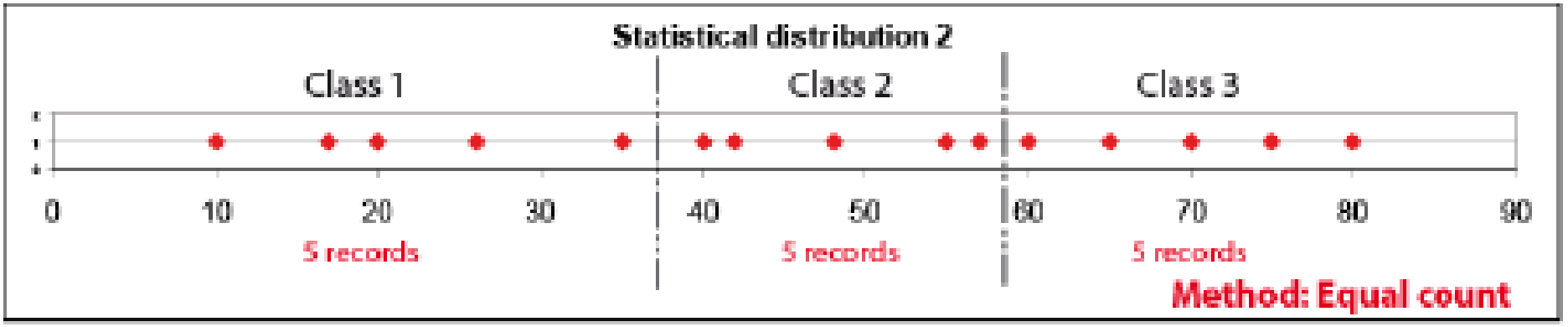
"…Maps that place an equal number of observations in each class: If you have 30 counties and 6 data classes, you'll have 5 counties in each class. The problem with quantiles is that you can end up with classes that have very different numerical ranges (e.g., 1-4, 4-9, 9-250…the last class is huge). Quantiles can also separate locations with very similar rates and group together places that have very different rates, which is very undesirable, so use the histogram to see if this is happening."

Source: http://axismaps.github.io/thematic-cartography/articles/classification.html

"Equal range contains approximately the same number of records. With 5 classes, each contains 20 % of the total number of the data values. This method is suited for comparing one dataset with datasets from other themes. If the data deviate from a linear distribution, the absolute class width will show large variations. Equal count methodology does not take into account exceptional values in the distribution. "

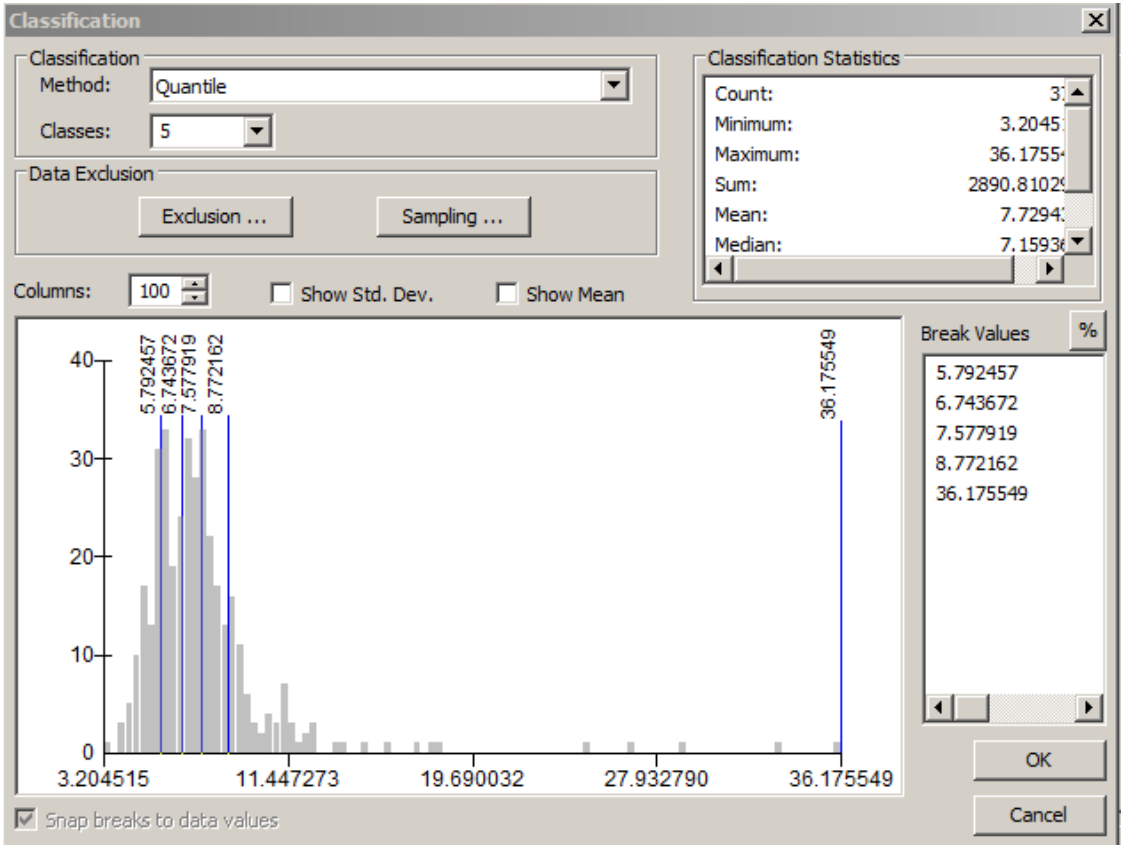
Source: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

Quantiles (2)



Source: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

Comments??

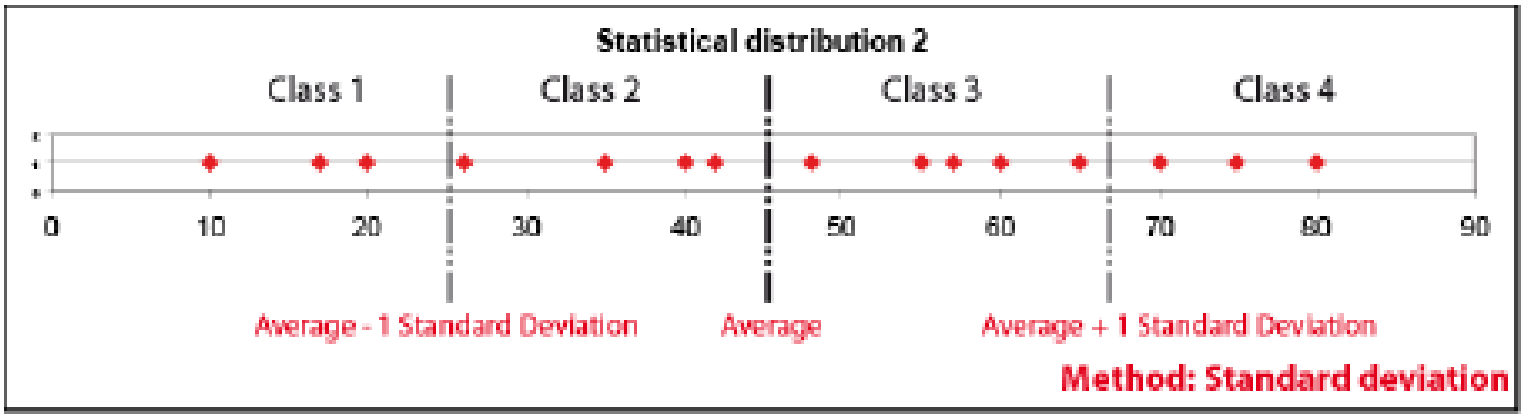


Mean – standard deviation (1)

"The class borders are calculated from the mean value and the standard deviation. Standard deviation is a way to describe statistical dispersion. The width of the class is equal to the standard dispersion (or an half depending on the number of classes expected). This method is suited for normal distributed datasets only."

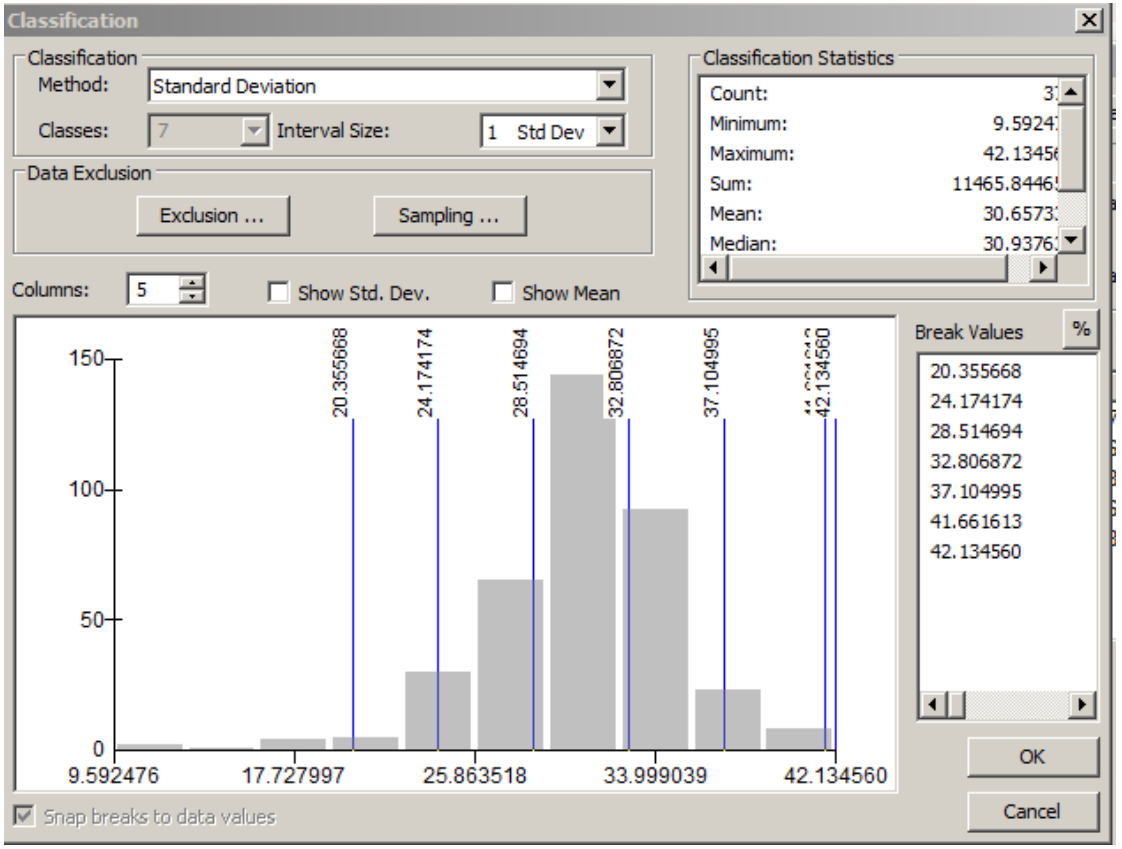
Source: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

Mean – standard deviation (2)



Source: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

Comments??



Natural breaks (1)

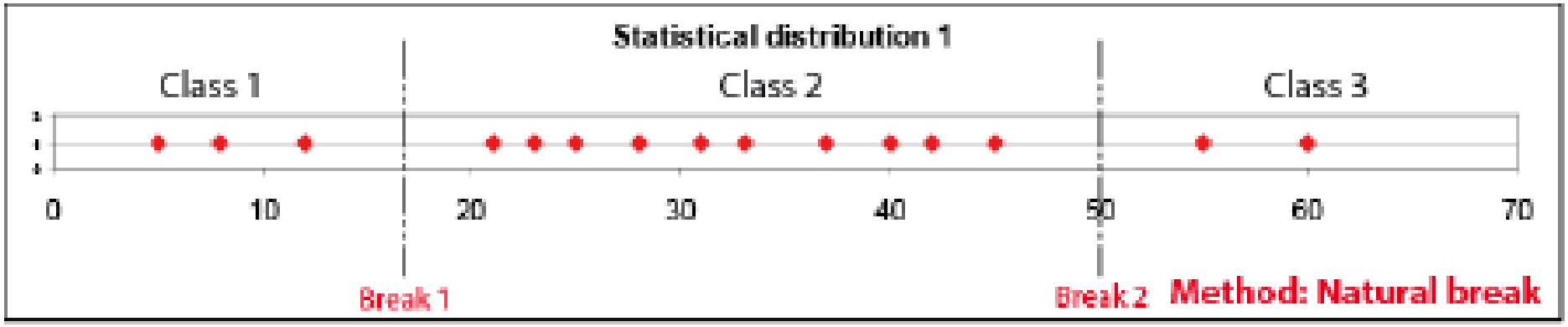
"….a kind of "optimal" classification scheme that finds class breaks that (for a given number of classes) will minimize within-class variance and maximize between-class differences. One drawback of this approach is each dataset generates a unique classification solution, and if you need to make comparison across maps, such as in an atlas or a series (e.g., one map each for 1980, 1990, 2000) you might want to use a single scheme that can be applied across all of the maps."

Source: http://axismaps.github.io/thematic-cartography/articles/classification.html

"This method sets the breakpoint to "natural points" in the dataset. The strength of this method is that it increases the information content. This method is suited when important breaks describe the dataset."

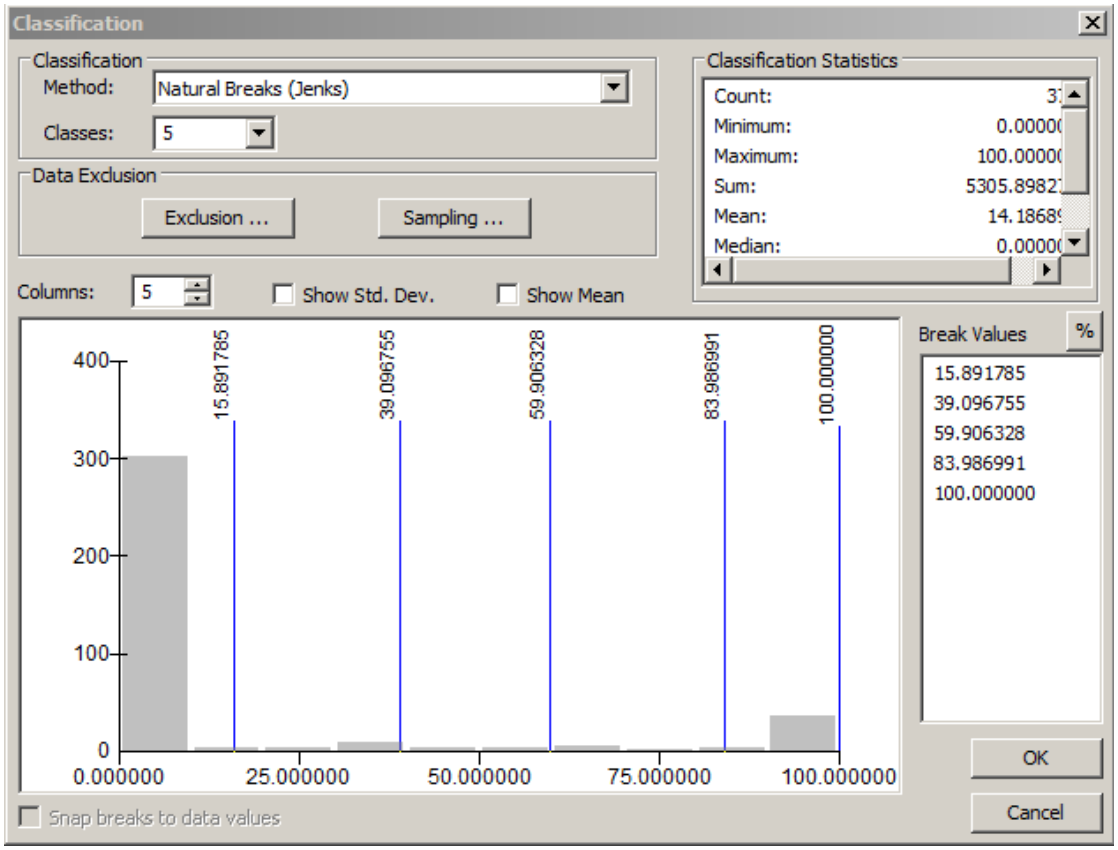
Source: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

Natural breaks (2)



Source: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

Comments??



Geometric interval (1)

"….a kind of "optimal" classification scheme that finds class breaks that (for a given number of classes) will minimize within-class variance and maximize between-class differences. One drawback of this approach is each dataset generates a unique classification solution, and if you need to make comparison across maps, such as in an atlas or a series (e.g., one map each for 1980, 1990, 2000) you might want to use a single scheme that can be applied across all of the maps."

Source: http://axismaps.github.io/thematic-cartography/articles/classification.html

The widths of the class follow a geometric progression. To calculate the width of the different class, it is necessarily to estimate the geometric ratio, such as:

log R = (log10 Max – log10 Min) / number of classes wanted

R = 10log r

Width of the Classes = (min, min x R); (min x R; min x R x R) and so on. This method is suited for uneven distribution and particularly distribution described by a lot of low values and few high values, such as density of population distribution. ***Source***: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

Example (1)

Figure 12 shows the importance of the choice of data range on the visualisation of phenomena.

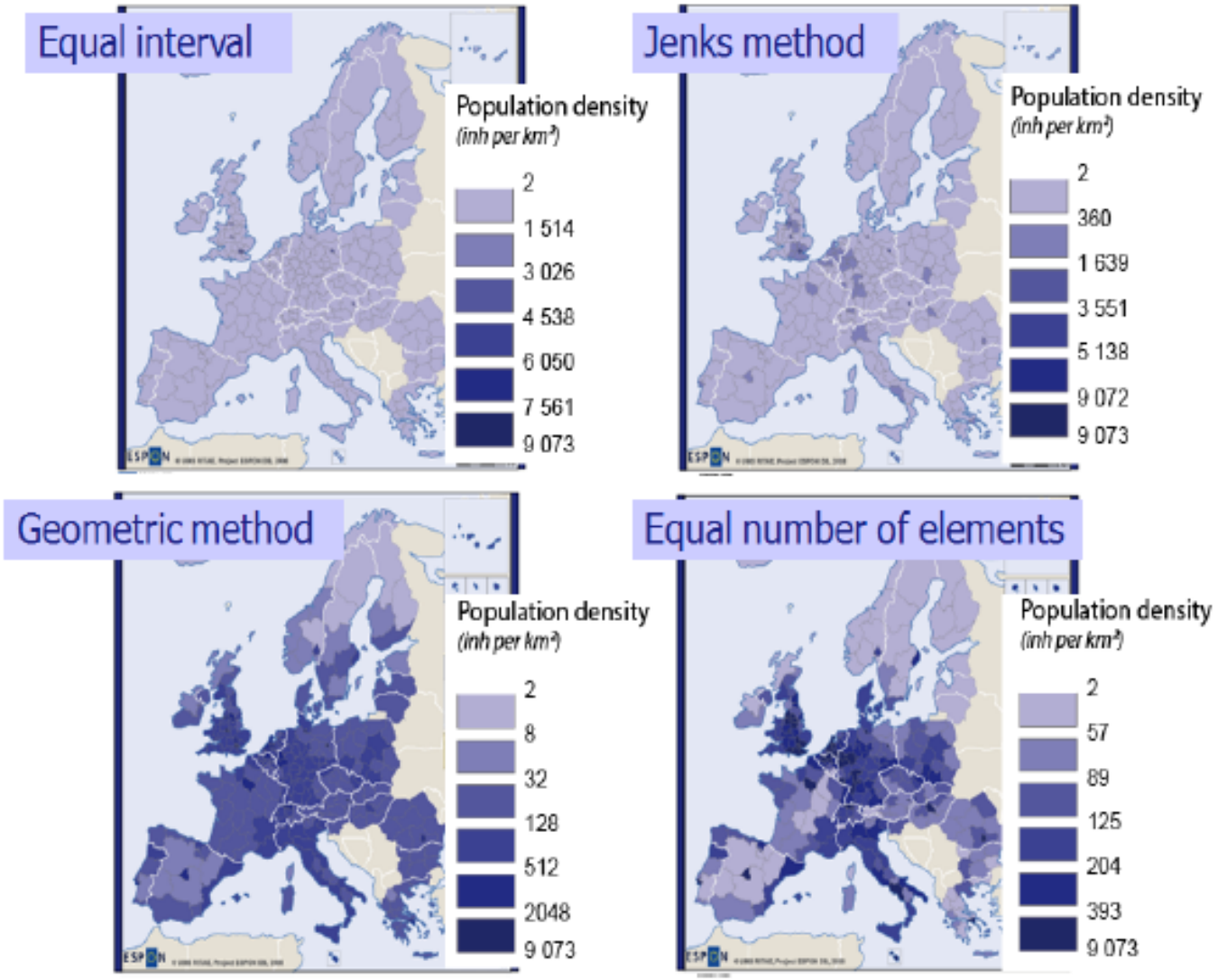


***Figure 12**: Result and efficiency are dependent upon the data classification method*

Source: Zanin et al. (2013), http://www.espon.eu/export/sites/default/Documents/ToolsandMaps/MappingGuide/MAPPING_GUIDE_EXTERNAL.pdf

The role of colour (1)

Here are some cultural colour associations

Red

China: the colour of brides, good luck, celebration, summoning
Cherokees: triumph, success
India: purity
South Africa: mourning colour
Eastern: worn by brides, joy (in combination with white)
Western: excitement, love, passion, stop
Europe/USA: Christmas (with green), Valentine's day (with white)
Hebrew: sacrifice, sin
Japan: life
Christian: sacrifice, passion, love

Blue

European: soothing, 'something blue' bridal tradition, depression
Cherokees: defeat, trouble
Iran: mourning, colour of heaven and spirituality
China: immortality
Colombia: soap
Hinduism: the colour of Krishna
Judaism: holiness
Christianity: Christ's colour, colour of Mary's robe
Middle East: protection
Worldwide: 'safe' colour

Yellow

European: Easter, happiness, hope, joy, cowardice, hazards, weakness
Asia: imperial, sacred
China: royalty, nourishing
Egypt: mourning
Japan: courage
India: merchants
Buddhism: wisdom

Orange

European: autumn, harvest, creativity
Netherlands: favourite colour (House of Orange)
Ireland: protestants
USA: Halloween (with black), cheap goods
Hinduism: saffron (peachy orange) is a sacred colour

Brown

Colombia: discourages sales
Australian Aboriginals: colour of the land, ceremonial ochre

Green

China: exorcism, green hats indicate a man's wife is cheating on him
Japan: life
Islam: hope - the cloak of the prophet was thought to be green, virtue - only those of perfect faith can wear green
Ireland: symbol of the entire country, Catholics
Western: spring, new birth, go, safe, environmental awareness, Saint Patrick's Day, Christmas (with red)
USA: money

Purple

Thailand: mourning (widows)
European: royalty
Catholicism: death, mourning, crucifixion

White

European: marriage, angels, doctors, hospitals, peace
Japan: mourning, white carnation symbolizes death
China: mourning, death,
India: unhappiness
Eastern: funerals

Black

European: mourning, funerals, death, rebellion, cool, restfulness
China: colour for young boys
Thailand: bad luck, evil, unhappiness,
Judaism: unhappiness, bad luck, evil
Australian Aboriginals: colour of the people, ceremonial ochre

Source: http://www.chamberlaindunn.com/documents/Themeaningofcolour.pdf

The role of colour (2)

“ Nominal data are categories that are inherently unorderable (like soils or landuse) and should only be mapped with nominal color schemes.

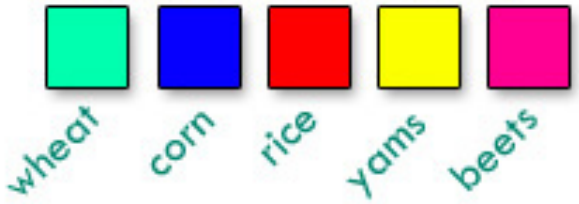
If you have orderable categories (such as low/med/high) or if you have numerical data, a sequential color scheme is what you need. Sequential color schemes can be single or multi-hue, but they are dominated and ordered by differences in lightness/saturation.

Diverging schemes should only be used when your data has a natural mid-point such as a zero (e.g., positive and negative change/growth) or if you want to compare places to something like the national average (e.g., county data showing places that are above and below the national average for a variable like per capita income).”

Source: Adopted by http://axismaps.github.io/thematic-cartography/articles/proportional_symbols.html,

The role of colour (3)

Nominal Color Scheme



different hues that keep lightness and saturation constant should be used for **nominal data** (i.e., un-orderable categories, not numerical data).

Sequential Color Scheme

any sequence that is **dominated by changes in lightness** can be used with orderable (rankable) categories (low/med/high) or with numerical data.

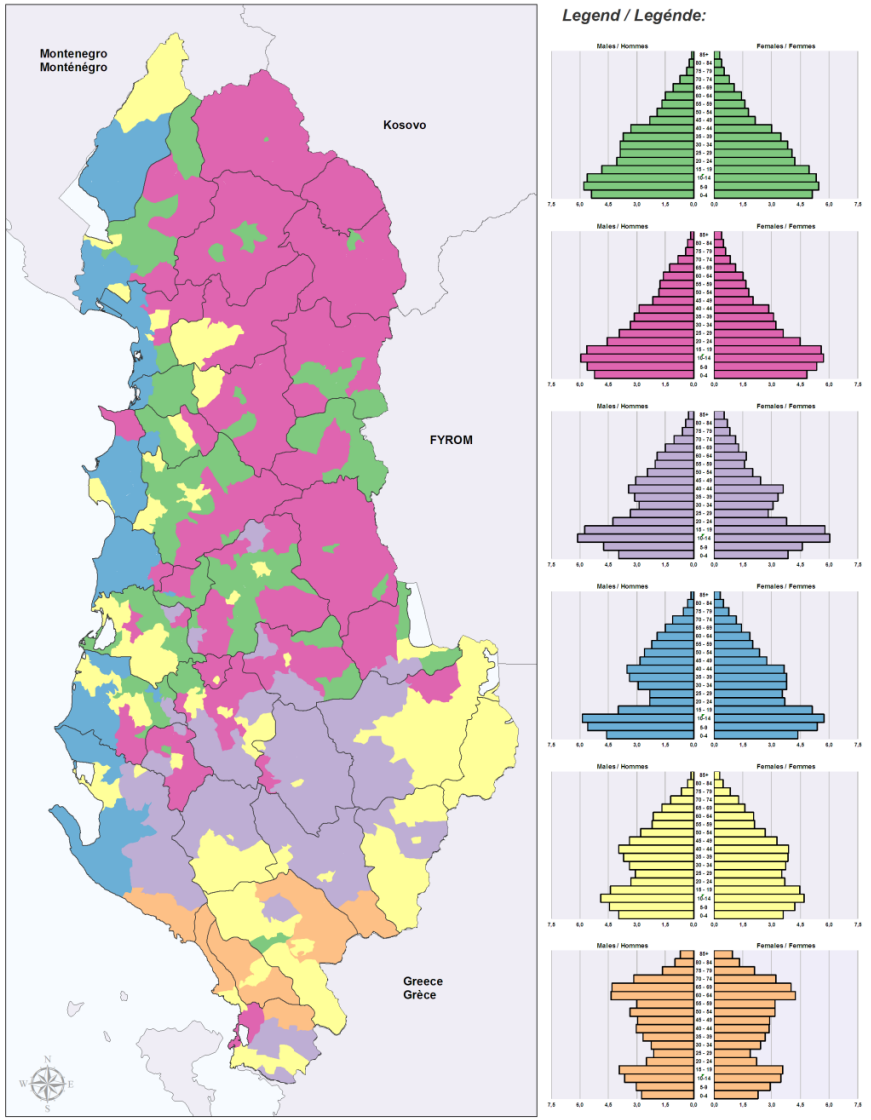
Diverging Color Scheme

any numerical data **that can be divided meaningful at a mid-point** (e.g., national average, zero) can use a diverging scheme: the data are split in two around the lightest, middle color/class.

Source: Adopted by http://axismaps.github.io/thematic-cartography/articles/proportional_symbols.html,

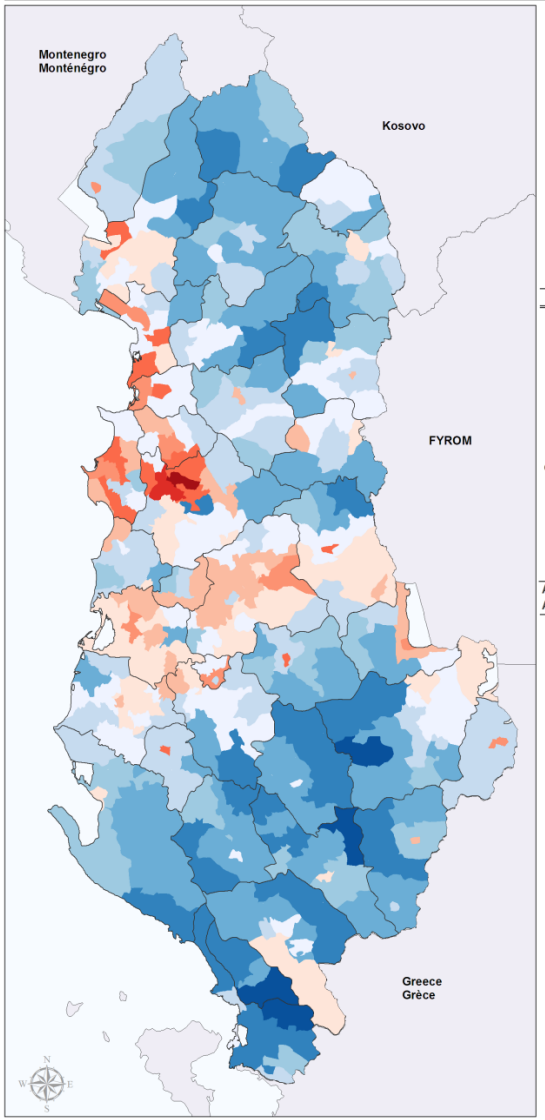
The role of colour (4)

Typology of age structure (population pyramids) 2001
Typologie de la structure d'âge (pyramide des âges) 2001

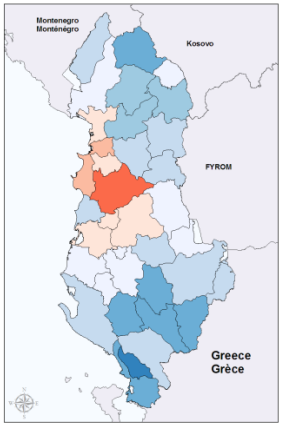


Source: Agorastakis M. (2012). "Reconstruction of the population map of western Balkans: Crises and population mobility during 1991-2001. The case of Albania". PhD Thesis. University of Thessaly, University of Thessaly, Department of Planning and Regional Development

Population Change 1989-2001 (%)
Taux de croissement (%)



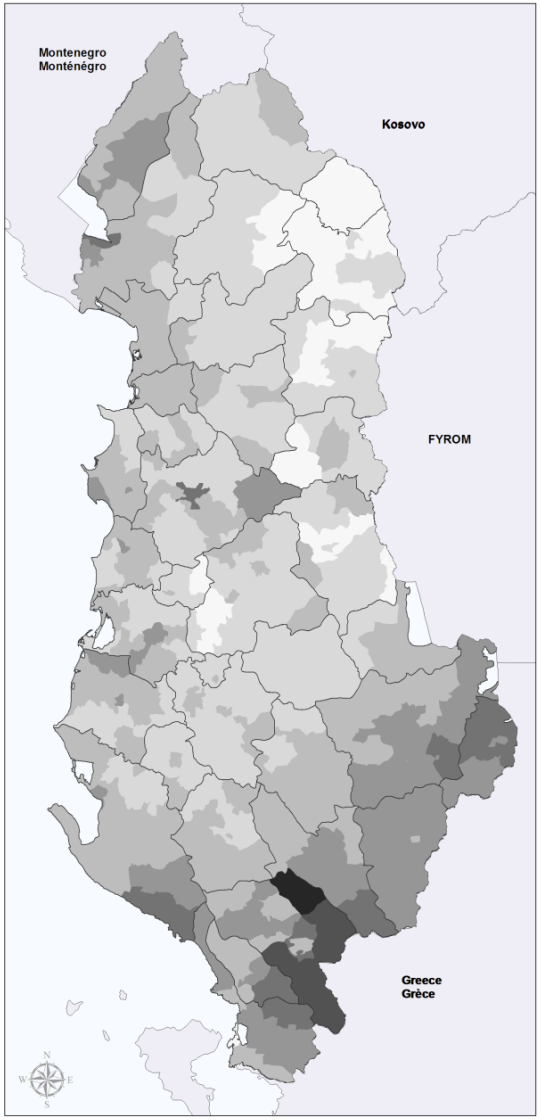
	aa	Group / Groupe	#	%		#	%
Commune	12	> 70	2	0,53	District	-	-
	11	50,01 - 70	2	0,53		-	-
	10	30,01 - 50	14	3,74		1	2,78
	9	20,01 - 30	10	2,67		-	-
	8	10,01 - 20	30	8,02		2	5,56
	7	0 - 10	45	12,03		5	13,89
	6	-10 - 0,01	59	15,78		8	22,22
	5	-20 - -10,01	65	17,38		10	27,78
	4	-30 - -20,01	39	10,43		3	8,33
	3	-50 - -30,01	69	18,45		6	16,67
	2	-70 - -50,01	34	9,09		1	2,78
	1	< - 70	5	1,34		-	-
		Total / Totale	374	100,00		36	100,00
	Min	-85,1 Finiqi				-59,51	Delvina
	Max	506,4 Pasku qani				41,15	Tirana
Albania / Albanie				-3,55			



Source: Agorastakis M. (2012). "Reconstruction of the population map of western Balkans: Crises and population mobility during 1991-2001. The case of Albania". PhD Thesis. University of Thessaly, University of Thessaly, Department of Planning and Regional Development

The role of colour (5)

Mean Age (total population, 1989)
Âge moyen (population totale, 1989)

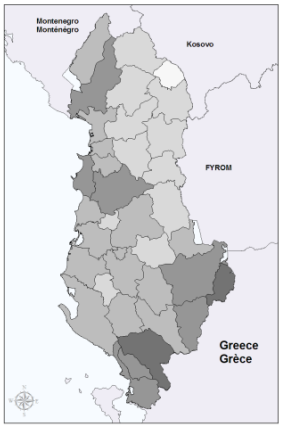


Legend / Legénde:

> 35	26,01 - 28
32,01 - 35	24,01 - 26
30,01 - 32	<= 24
28,01 - 30	

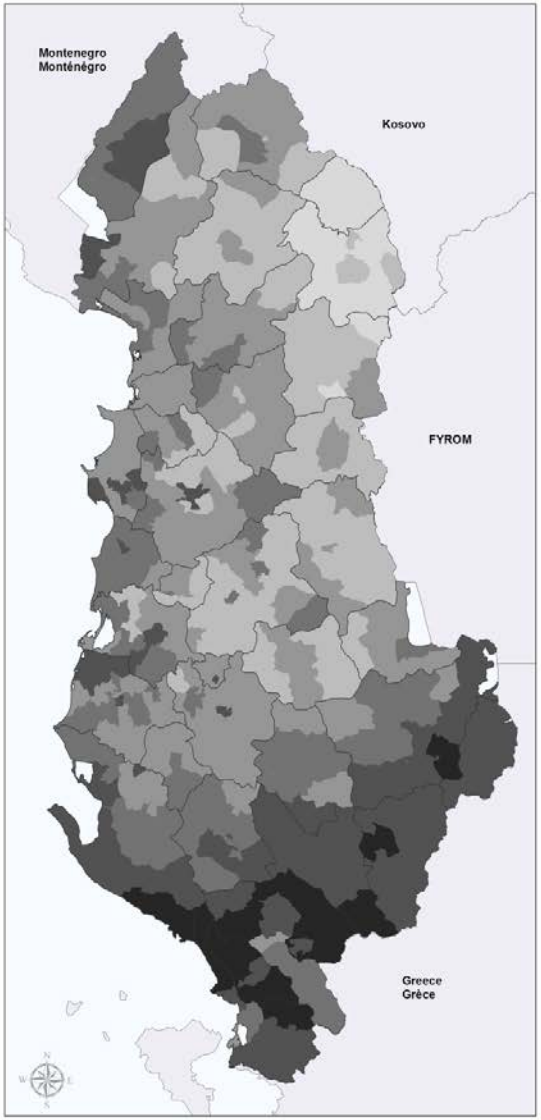
	aa	Group / Groupe	#	%
Commune	7	> 35	1	0,27
	6	32,01 - 35	3	0,80
	5	30,01 - 32	12	3,21
	4	28,01 - 30	42	11,23
	3	26,01 -28	133	35,56
	2	24,01 - 26	153	40,91
	1	<= 24	30	8,02
		Total / Totale	374	100,00
	Min	22,02 Kolshi		
	Max	35,81 Zagoria		

	aa	Group / Groupe	#	%
District	7	> 35	-	-
	6	32,01 - 35	-	-
	5	30,01 - 32	2	5,56
	4	28,01 - 30	7	19,44
	3	26,01 -28	15	41,67
	2	24,01 - 26	11	30,56
	1	<= 24	1	2,78
		Total / Totale	36	100,00
	Min	22,96 Hasi		
	Max	30,27 Devolli		
Albania /			27,21	



Source: Agorastakis M. (2012). "Reconstruction of the population map of western Balkans: Crises and population mobility during 1991-2001. The case of Albania". PhD Thesis. University of Thessaly, University of Thessaly, Department of Planning and Regional Development

Mean Age (total population, 2001)
Âge moyen (population totale, 2001)



Legend / Legénde:

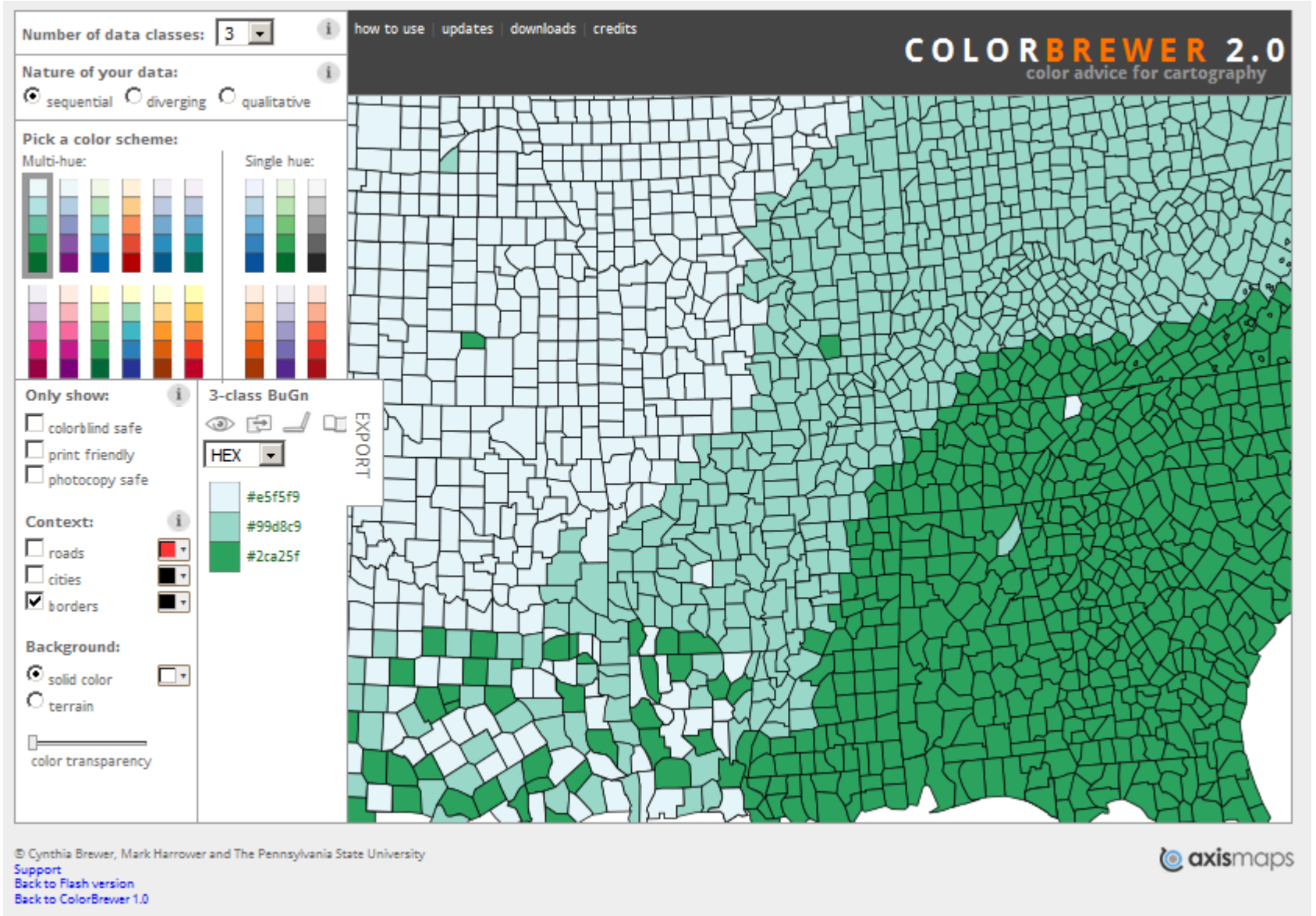
> 35	26,01 - 28
32,01 - 35	24,01 - 26
30,01 - 32	<= 24
28,01 - 30	

	aa	Group / Groupe	#	%
Commune	7	> 35	15	4,01
	6	32,01 - 35	62	16,58
	5	30,01 - 32	79	21,12
	4	28,01 - 30	121	32,35
	3	26,01 -28	79	21,12
	2	24,01 - 26	18	4,81
	1	<= 24	-	-
		Total / Totale	374	100,00
	Min	24,20 Gjinaj		
	Max	52,75 Mesopotami		

	aa	Group / Groupe	#	%
District	7	> 35	1	2,78
	6	32,01 - 35	7	19,44
	5	30,01 - 32	13	36,11
	4	28,01 - 30	9	25,00
	3	26,01 -28	5	13,89
	2	24,01 - 26	1	2,78
	1	<= 24	-	-
		Total / Totale	36	100,00
	Min	24,99 Hasi		
	Max	37,68 Delvina		
Albania /			30,73	

Source: Agorastakis M. (2012). "Reconstruction of the population map of western Balkans: Crises and population mobility during 1991-2001. The case of Albania". PhD Thesis. University of Thessaly, University of Thessaly, Department of Planning and Regional Development

The role of colour (6)



Source: http://colorbrewer2.org/

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

End of session