Unit 1 - Level of Detail

The term level of detail (lod) describes the level of aggregation of a spatial unit, i.e. data from a multitude of smaller administrative levels will be aggregated into a larger unit. This way, data is more detailed and represents reality more closely. By aggregating data, individual values will be lost and a loss of information occurs.
Taking a look at France, observe the influence of aggregation on the visualization of the population density.

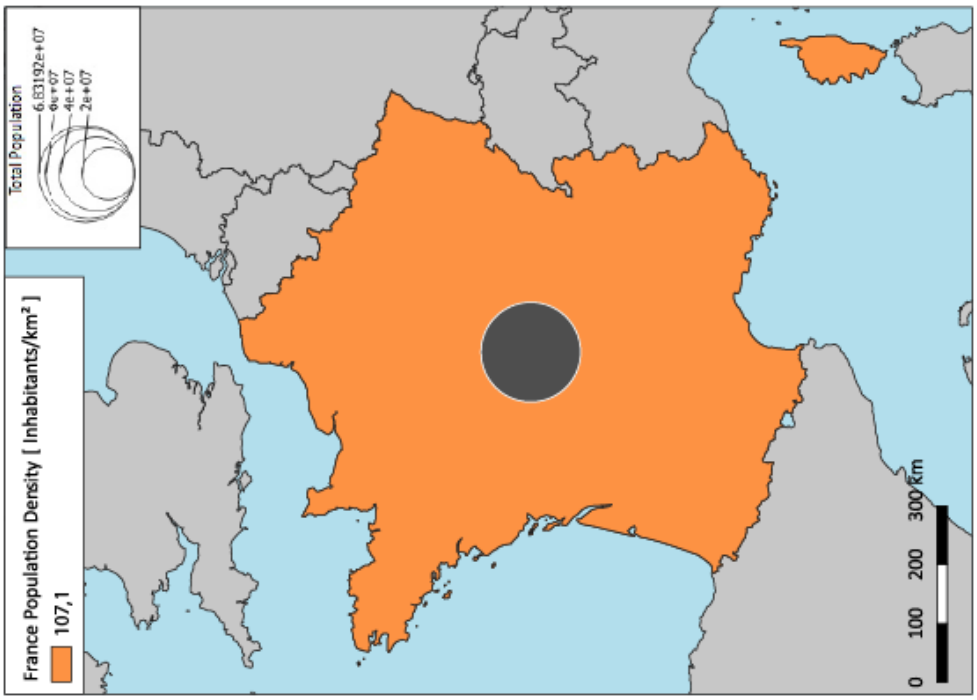
Exercises

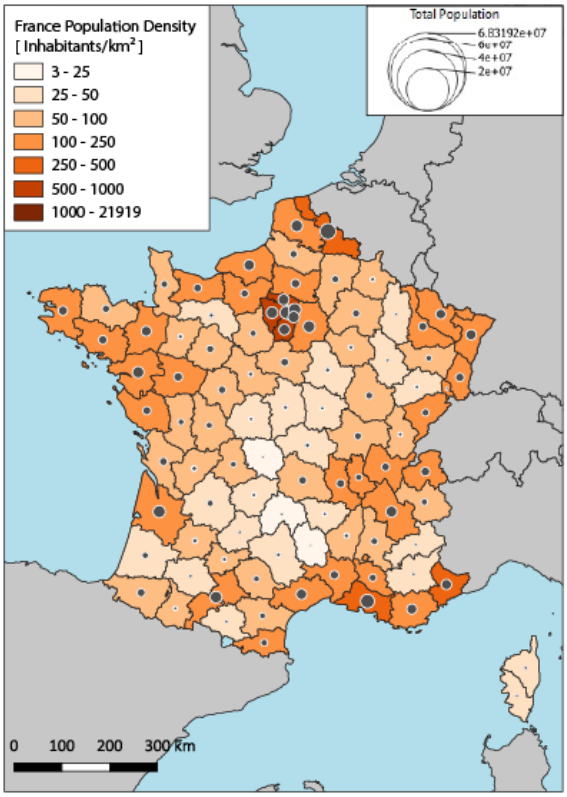
(I) The French region Normandie is made up of five ‘departements’ which are listed in the table below. Calculate the population density per square kilometer for each ‘departement’.

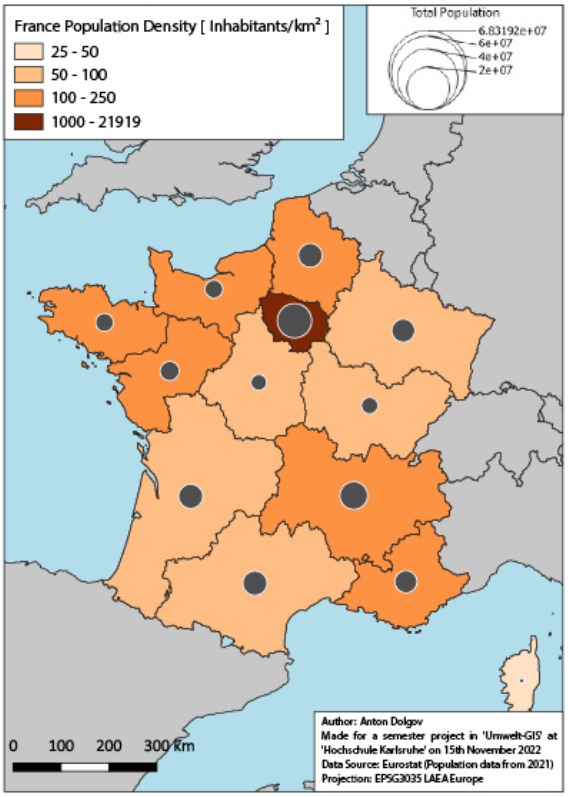
Departement	Area	Population
Calvados	5 534 km²	695 310
Eure	6 040 km²	593 885
Manche	5 951 km²	491 281
Orne	6 103 km²	273 214
Seine-Maritime	6 278 km²	1 253 596

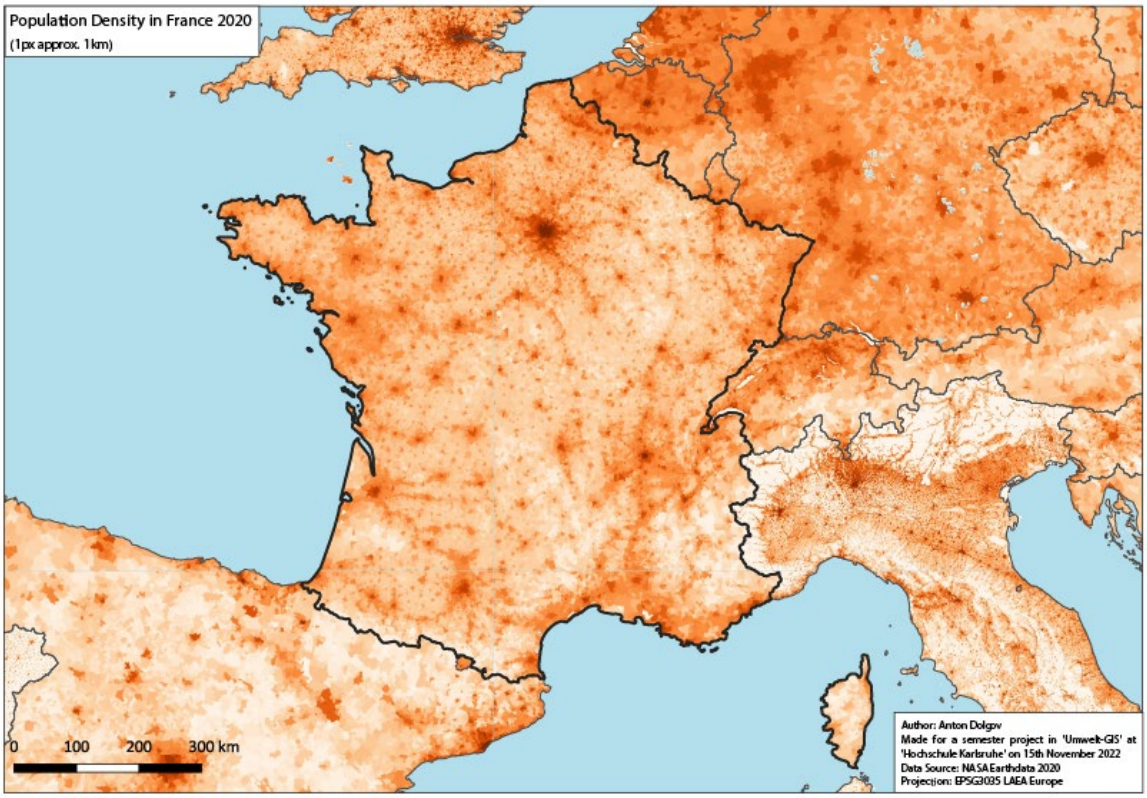
(I) In which administrative levels is France divided? Compare its administrative levels with those of Germany, Kenya and South Africa. Take a look at the maps. Are absolute or relative values shown?
(II) Which administrative area is the most / least populous of each administrative level?
(III) How is the population density affected by the level of detail? Which advantages and disadvantages do different levels of detail have? Compare the different maps with the heatmap. What differences are there?
(IV) Topic Specialists: Research this unit's topic. Create a handout and present your topic in class.

Attachments




France Population Density
[Inhabitants/km²]
3 - 25
25 - 50
50 - 100
100 - 250
250 - 500
500 - 1000
1000 - 21919
Total Population
6.83192e+07
6e+07
4e+07
2e+07
0 100 200 300 km


France Population Density [Inhabitants/km²]
25 - 50
50 - 100
100 - 250
1000 - 21919
Total Population
6.83192e+07
6e+07
4e+07
2e+07
0 100 200 300 km
Author: Anton Dolgov
Made for a semester project in 'Umwelt-GIS' at 'Hochschule Karlsruhe' on 15th November 2022
Data Source: Eurostat (Population data from 2021)
Projection: EPSG3035 LAEA Europe


Population Density in France 2020
(1px approx. 1km)
0 100 200 300 km
Author: Anton Dolgov
Made for a semester project in 'Umwelt-GIS' at 'Hochschule Karlsruhe' on 15th November 2022
Data Source: NASA Earthdata 2020
Projection: EPSG3035 LAEA Europe

Unit 2 - Frequency

A frequency distribution
• Increases the clarity of the dataset
• Emphasizes important information
This helps answer the following questions:
• How do the values of the variables vary?
• How are the observations distributed spatially?

There are multiple kinds of frequency:
• Absolute frequency
• Relative frequency
• Absolute cumulative frequency
• Relative cumulative frequency

Absolute Frequency

The absolute frequency describes how often an observation or a value occurs in the dataset.

Relative Frequency

The relative frequency is the absolute frequency in proportion to the total number of observations.

Cumulative frequency

The cumulative frequency shows the occurrence of observations up to a predefined limit. To calculate it, ordinal data is required.

Absolute cumulative frequency

The absolute cumulative frequency is a cumulative frequency which is made up exclusively of absolute values.

Relative cumulative frequency

The relative cumulative frequency is a cumulative frequency which is made up exclusively of relative values.

Exercises

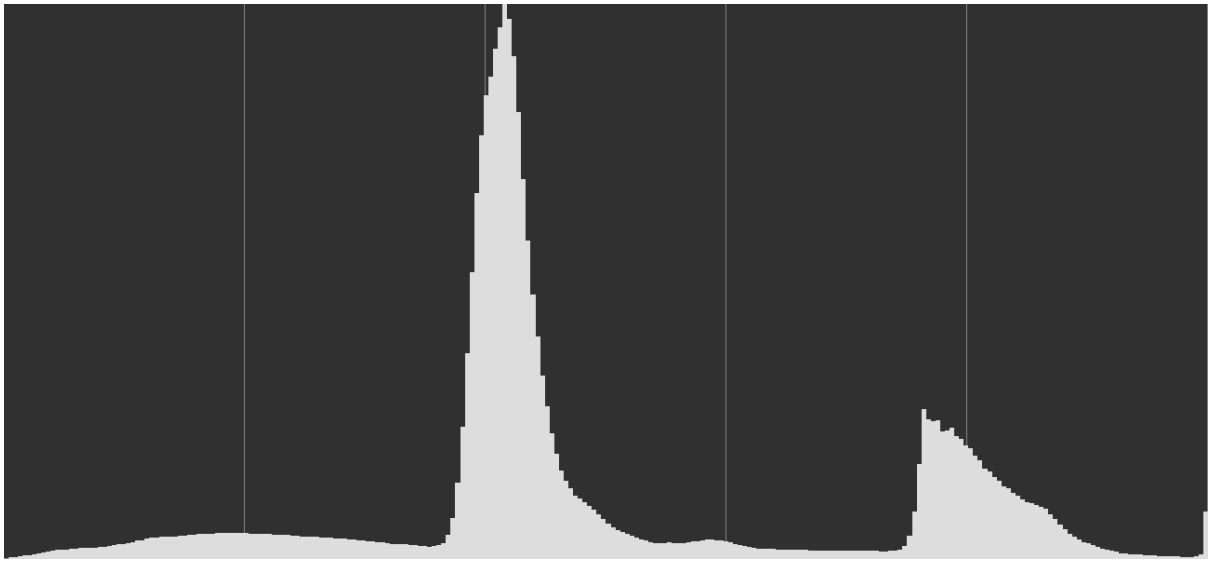
(I) The following table shows the amount of French men and women with degrees in proportion to the total population. France's population is made up of 32.132.852 men and 33.885.763 women. How many of them have earned a degree?

Sex	Degree
Male	46%
Female	50,4%

(II) Calculate the absolute and relative cumulative frequency for both sexes and France as a whole.
(III) The following histogram shows the pixel values of a grayscale image.

Describe it using the following terms: *absolute frequency, relative frequency, absolute cumulative frequency und relative cumulative frequency*.
(IV) Topic Specialists: Research this unit's topic. Create a handout and present your topic in class.

Attachments



Unit 3 - Topic

The goal of the blended learning experience is to give you a basic understanding of statistics using the Sustainable Development Goals (SDGs) as established by the United Nations. This worksheet will be an introduction to Indicators 1.2.1 and 4.2.2 which will be used in the webtool.

UN-Sustainable Development Goals

The SDGs are political goals guaranteeing sustainable development on a social, economical and ecological level. There are 17 SDGs which are to be achieved by 2030.

Indicator 1.2.1.

This indicator aims to halve the proportion of people living under the poverty line. For this, data is dissected by age and sex. It is part of SDG 1 - “No Poverty”.

Indicator 4.2.2.

This indicator aims to ensure that all boys and girls have access to pre-primary education. For this, data is dissected by age and sex. It is part of SDG 4 - “Quality Education”.

Exercises

(I) While their predecessors, the Millennium Development Goals were only relevant to developing countries, the SDGs are to be achieved by all member states of the United Nations.What problems could result from this? Which differences might occur on a transnational level? Discuss.
(II) Indicator 1.2.1 is part of SDG 1 - “No Poverty”. What actions can be taken to reach this goal? Think of two measures and present them to your group.
(III) Indicator 4.2.2 is part of SDG 4 -”Quality Education”. Research about institutions for pre-primary education in your hometown. How many are there? How expensive are they? Did you participate in pre-primary education? What advantages are there for children attending pre-primary education? Discuss.
(IV) Topic Specialists: Research this unit's topic. Create a handout and present your topic in class.

Unit 4 - Correlation and regression

Correlation describes the degree of interdependency of two variables. Regression measures how one variable affects another.

Covariance is a measure of the joint variability of two variables. Its sign shows the tendency of the linear relationship between the two variables. It is not normalized and therefore hard to interpret.
This is why the correlation of two variables is given through the correlation coefficient r.

Correlation

The correlation coefficient r describes the grade of correlation. It ranges from -1 to 1.
It can be calculated by dividing the covariance of the two variables by dividing the product of their standard deviation:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Two variables can be correlated positively (r=1),negatively (r=-1) or not at all (r=0). A positive correlation means that an increase of the first variable results in an increase of the second. A negative correlation means the opposite, i.e. that an increase in the first variable results in a decrease of the second.

A correlation between two variables does not automatically mean that they influence each other, as there might be a third variable influencing both. This is where the regression comes in.

Linear bivariate regression

The trend of individual values can be illustrated using a regression line. Using the least-squares estimation a line fitting all observations can be created. It is defined by its slope a and axis-intersect b.

$$y = ax + b$$

This means that simple predictions can be made using the regression line.

Exercises

(I) The following table grades subjects' dental hygiene and their dental health.

Patient	Dental hygiene	Dental health
A	9,5	9,0
B	6,5	7,5
C	4,5	6,0
D	8,5	8,0
E	1,5	2,0
M	6,1	6,5
SD	3,21	2,74

Calculate the covariance and the correlation of the variables.
(II) Topic Specialists: Research this unit's topic. Create a handout and present your topic in class.