Take-home Exam

Spatial Data (GIS Tools Laboratory)

Giacomo Novelli

08-02-2025

Please follow the guidelines below:

- use R to replicate the tasks;
- submit a compressed archive (.zip) including all necessary files for replicating your code. The archive name must be: SURNAME_NAME.zip;
- the compressed archive must contain your complete working directory, including the Rmd file, the corresponding pdf file, and the corresponding R file, such that I can run your code;
- organize your Rmd file clearly, numbering the task you are replicating;
- use the NetCDF_NASA.Rmd file as template;
- the datasets can be find in R classes and/or in the shared One Drive directory and/or online following the provided links. Please contact me if you cannot access the shared directory or find the data.

Start your Take-home Exam assessing your name, surname and student ID (matricola).

The aim of the exam is testing your GIS skills and your analytical reasoning.

The total points you can score in this exam is 40.

You get 30L with more than 30 points.

Section A: Research Proposal [4 points]

- Write a research proposal involving Spatial Data.
  You can use the following structure:
  Research Question, Data & Methods, Expected Outcomes/Contribution. 450 words max.

Section B: Task replication [36 points]

- Section B consists in handling Croatian data.
  It is noteworthy that no prior knowledge of Croatia is required.
  You will mostly deal with "fake" spatial data, i.e. spatial data arbitrarily modified for the purposes of this exam, such that the nature of data is irrelevant.

Section B

1. Load the shapefile of countries, keep Croatia, and call it "shp". Then, load the shapefile of European NUTS 3 regions, keep the NUTS 3 of Croatia, and call it "shp3". Finally, create a boundary box containing the shapefile of Croatia with a small margin. [0.5 points]

```
# Hint:
box <- st_bbox(shp)
box <- box+c(-.5,-.5,.5,.5)]
```

2. Load SPEI data and crop it around the box using `crop()`. Some grid cells will (partly or completely) lay outside the Croatian border. Then, keep the last six layers. Four our purposes, those layers will be the SPEI values for the following years: 2015,2016,2017,2018,2019,2020. [0.5 points]

3. Generate a vector of 6 numbers from a uniform distribution (min = 0, max = 1) after setting the seed as follows, then multiply the 6 SPEI layers by those random numbers. [0.5 points]

```
# Set seed for reproducibility
set.seed(XXX) # insert here the last 3 figures of student ID (matricola)
```

```
# Example:
reduced_spei[[1]] <- reduced_spei[[1]] * random_vector[1]
reduced_spei[[2]] <- reduced_spei[[2]] * random_vector[2] # and so on
```

4. Extract the grid points of the SPEI raster data. [1.5 points]

```
# Hint:
points <- as.points(reduced_spei)
points
points_sf <- st_as_sf(points)
```

5. Download CRU precipitation data from: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.08/cruts.2406270035.v4.08/pre/. The specific file is "cru_ts4.08.2011.2020.pre.dat.nc.gz". Keep data for years 2015-2020 and crop it around the boundary box already defined. [2.5 points]

6. Load the provided Aqueduct 4.0 shapefile. Attribute "bws_raw" is the raw value of the Aqueduct water stress index. Four our purposes, we will assume "bws_raw" is the water stress index of each watershed in 2015. [1 point] [More info on water stress in the provided Aqueduct 4.0 documentation.]

7. Adapt the following lines to generate a water stress time series. [1 point]

```
# Set seed for reproducibility
set.seed(XXX) # insert here the last 3 figures of student ID (matricola)
# Depending on your choices, the number of features may vary
# Assuming aqueduct_shapefile$bws_raw has got 99 elements
length(aqueduct_shapefile$bws_raw) # 99
water_stress_2015 <- aqueduct_shapefile$bws_raw  # A vector of length 99
# Add this vector to the Aqueduct shapefile
aqueduct_shapefile$water_stress_2015 <- water_stress_2015
# Generate a matrix of random multipliers for 5 years (2016 to 2020) for 99 elements
multipliers <- matrix(rnorm(5 * 99, mean = 1, sd = 1), nrow = 99, ncol = 5)
# Calculate the annual values by multiplying the 2015 values by the random multipliers
```

```
water_stress_ts <- water_stress_2015 * multipliers
# Add those attributes to the Aqueduct shapefile
aqueduct_shapefile$water_stress_2016 <- water_stress_ts[, 1]  # Values for 2016
aqueduct_shapefile$water_stress_2017 <- water_stress_ts[, 2]  # Values for 2017
aqueduct_shapefile$water_stress_2018 <- water_stress_ts[, 3]  # Values for 2018
aqueduct_shapefile$water_stress_2019 <- water_stress_ts[, 4]  # Values for 2019
aqueduct_shapefile$water_stress_2020 <- water_stress_ts[, 5]  # Values for 2020
```

8. Rasterize the values of "water_stress_20XX" attributes using `terra`, and then extract them to the shapefiles of Croatia and Croatian NUTS 3. What is the average water stress level in Croatia in 2020? Is this value equal to the simple average of water stress in Croatian NUTS 3 districts? Discuss this shortly (a few lines). [4 points]

9. Load the provided population density data for 2015, crop it around the boundary box already defined. What is the average population density level in Croatia? Now, resample it on the resolution of SPEI data using the "bilinear" method. Is the Croatian average population density changed? Discuss this shortly (a few lines). [3 points]

- First Not Mandatory Task: as discussed in the following paper, create a population-weighted version of SPEI and water stress data using the gridded data of population density. [3 points]

- Gortan, M., Testa, L., Fagiolo, G., & Lamperti, F. (2024). A unified dataset for pre-processed climate indicators weighted by gridded economic activity. Scientific Data, 11(1), 533.

```
# Hint:
# Step 1: calculate the average pop.density in Croatia.
# Step 2: create the following ratio for each grid point:
#         pop.density in point i / average pop.density in Croatia
# Step 3: multiply the SPEI value in 2015 in point i
#         by the ratio calculated at Step 2.
# Step 4: multiply the SPEI value in 2016 in point i
#         by the ratio calculated at Step 2.
# Step 5: the same for the remaining years.
# Step 6: the same for the water stress index.
# Step 7: the same for precipitation data.
# Step 8: from now on use the newly created pop.weighted climate variables.
```

10. Some econ papers do not use population density to weight climate data. In your opinion is it relevant to use population-weighted climate variables when analyzing GDP data? Are they relevant when investigating Construction GVA? And what about Agricultural and Industrial GVA? Discuss this shortly (a few lines). [2 points]

11. Assume SPEI resolution is c(1,1) and population density resolution is c(0.25,0.25). In your opinion is it preferable to reduce the resolution of SPEI to the resolution of population density and then create a pop-weighted SPEI index, or viceversa? Discuss this shortly (a few lines). [2 points]

12. Download two datasets from the following link: https://urban.jrc.ec.europa.eu/ardeco/explorer?lng=en.
The first dataset is "Hours Worked by industry (10 NACE sectors)" (code: RNLHZ) and the second one is "GVA at basic prices by industry (10 NACE sectors)" (code: SUVGZ). Keep the data for sector F, which is CONSTRUCTION, from 2015 to 2020 for Croatian NUTS 3 regions only. Merge them with the sf object called "shp3". Before merging the data, drop the Croatian NUTS whose name ends

with "ZZZ". [2 points]
[More info on the datasets here: https://urban.jrc.ec.europa.eu/map-view?lng=en&ctx=udp&ts=EU&pil=indicator-level&is=Default]

13. Generate a new simple feature object called "final_sf" whose features are the Croatian NUTS 3 polygons and whose fields are the average values of SPEI, water stress, precipiation, and Construction GVA in each considered year. [0.5 points]

```
# Hint: final_sf should have the following fields: nuts_id, spei_2015, spei_2016,
# spei_2017, spei_2018, spei_2019, spei_2020, water_stress_2015, water_stress_2016,
# water_stress_2017, water_stress_2018, water_stress_2019, water_stress_2020,
# constr_gva_2015, constr_gva_2016, constr_gva_2017,
# constr_gva_2018, constr_gva_2019, constr_gva_2020.
# Obviously, this is just an example: you can name attributes as you prefer.
```

14. Plot the time series of SPEI, water stress, and precipitation in Croatia from 2015 to 2020. [0.5 points]

15. Plot the time series of SPEI, water stress, precipitation, and Construction GVA of each Croatian NUTS 3 region. [0.5 points]

```
# Hint: look at the hands-in part of Lecture 5
```

16. Assume

$$speigr = \frac{(spei_{2020} - spei_{2015})}{spei_{2015}}$$

and

$$wsgr = \frac{(ws_{2020} - ws_{2015})}{ws_{2015}}$$

Plot a map showing the growth rate of SPEI from 2015 to 2020 of each Croatian NUTS 3 region, in CRS 4326. Then, plot a map showing the growth rate of the water stress index from 2015 to 2020 of each Croatian NUTS 3 region, in CRS 4326. [0.5 points]

17. Plot a map showing the growth rate of precipitation from 2015 to 2020 of each Croatian NUTS 3 region, in CRS 32634. [0.5 points]

18. Assume

$$constructiongvagrowth = \frac{(constructiongva2020 - constructiongva2015)}{constructiongva2015}$$

Plot a map showing the growth rate of Construction GVA from 2015 to 2020 of each Croatian NUTS 3 region, in CRS 3035. Then, plot the map again using blue for regions experiencing a negative growth, lightblue for those having a growth rate between 0% and 5%, and white for regions with a growth rate >5%. [1 point]

19. Let's define a new dataset with monthly frequency. The study area is Europe and the dataset shows consumption, investment and GDP per capita for each NUTS 3 region from 2015 to 2020. Could GDP per capita in one region be affected by the lagged investment in a neighboring region? How could you check it? Discuss this shortly (a few lines). [2 points]

- Second Not Mandatory Task: [3 points]

  using the "queen" criterion,

  - compute the Moran's I Test for SPEI in 2015, and for Construction GVA in 2015.

- create a spatial autocorrelation scatterplot for the above-mentioned variables. [Hint: see class5.R]

- Let's recall the previous NUTS 3 European dataset with monthly frequency for consumption, investment and GDP per capita from 2015 to 2020. Can the Moran's I Test be used to test the effect of a lagged variable on the dependent variable in a neighboring region? For instance, the effect of investment at t-1 in region j on gdp at t in region i? Discuss this shortly (a few lines).
- Third Not Mandatory Task: [4 points]
- download Aqueduct gdb data from: https://www.wri.org/data/aqueduct-global-maps-40-data (direct download, no form, at the bottom of the webpage).
- generate the shapefile of Aqueduct data using `R` and the .gdb file.
  [Hint: you may want to crop the data around Croatia ("gid_0" is the variable containing country names) and remove any NAs at this stage in order to reduce the computational burden]