

MA22019 2025 - Problem Sheet 6

Visualization of point-referenced data and inverse distance weighting

Overview

This week's exercises help you with revising Sections 4.1 and 4.2 in the lecture notes. Exercises 1 asks you to identify the correct type of spatial data. Exercises 2 and 3 provide additional opportunities to visualize spatial data using shapefiles and maps, and to interpret the plots.

Tutorial Question 1 considers an example similar to that in the lecture notes, and Tutorial Question 2 focuses on the use of projections for visualizing data.

The Homework question is inspired by one of last year's coursework questions. Besides the creation of plots, the question asks you to carefully discuss the assumptions we make when applying the tools covered in Sections 4.1 and 4.2.

Your answer to the Homework Question can be submitted on Moodle to your tutor for feedback. The submission deadline is 17:00 on Thursday 27 March 2025. You should submit a single Word, PDF or HTML file that provides your R code, any created R output and all your comments.

You may want to load the following packages before starting the exercises:

```
library( dplyr )  
library( ggplot2 )  
library( sf )  
library( ggspatial )  
library( prettymapr )
```

Exercise 1 - Types of spatial data

For each of the following applications, decide whether the data will be point-referenced, point pattern or lattice/areal data:

- a) Occurrence of tornadoes across Europe
- b) Air pollution levels across a city
- c) Strength of mobile phone signal
- d) Damage caused by the great spruce bark beetle across Western England
- e) Sightings of Type Ia supernova

Exercise 2 - Rainfall across New Zealand

The file “NewZealandRain.csv” provides the aggregated rainfall between 01/09/2019 and 30/11/2019 for various locations across New Zealand. A shapefile for New Zealand is provided in the file “ShapeFileNZ.Rdata”.

- a) Load the shapefile and create a plot of New Zealand.
- b) Load the data in “NewZealandRain.csv”. Add points to the plot in part a) which represent the locations contained in “NewZealandRain.csv”, and use the visual cue colour to visualize the amount of precipitation on logarithmic scale recorded at the various locations
- c) Use inverse distance weighting with power parameter $p = 2$ to estimate the amount of precipitation between 01/09/2019 and 30/11/2019 for the following two locations: (i) 177°E Longitude and 38°S Latitude; (ii) 166°E Longitude and 51°S Latitude. You should use the original data and not apply the logarithmic scale from part b). Submit your answers (to two decimal places) in the Moodle quiz.

Exercise 3 - Earthquakes across Japan

The files “Japan earthquakes.csv” contains data for earthquakes which affected Japan between 2011 and 2018. The data has been collected by the United States Geological Survey. We want to use spatial data analysis to investigate the locations of the strongest earthquakes.

- a) Load the data and extract the earthquakes with a magnitude of 6 or higher.
- b) Create a map to visualize the locations and magnitude of the earthquakes extracted in part a).
- c) Go to Moodle and complete the quiz.

Tutorial Question 1 - Temperature across north-eastern Brazil

The file “Brazil.csv” contains average daily temperatures between 2000 and 2021 for 149 weather stations in north-eastern Brazil. You also given a shapefile for Brazil. We want to analyze the data using the techniques for point-referenced data introduced in the lecture.

- a) Visualize the locations of the weather stations and average daily temperature using the provided shapefile. Use the functions `xlim()` and `ylim()` to zoom in on the area of Brazil the stations are located in.
- b) Create a map with the `ggspatial` R package that illustrates the spatial locations of the weather stations and the recorded average daily temperatures. What do you conclude?
- c) Use inverse distance weighting to predict average daily temperature across north-eastern Brazil. Make sure to choose a suitable power parameter. Comment on the reliability of our estimates.

Tutorial Question 2 - Alaskan horsehair crab landings

In Section 4.1.3 we covered the topic of projections. The following tasks showcase one important aspect we should keep in mind when working with projections and demonstrates how we may handle it.

The data file “Crabs.csv” contains catch per unit effort (CPUE) data of commercial horsehair crab landings for various locations across the Alaskan Eastern Bering Sea for 2010-2018. CPUE is an indirect measure of the abundance of a species. We want to visualize the data using the shapefile for Alaska:

```
Alaska <- read_sf( "Shapefiles/Alaska.shp" )
```

- a) Create a plot of the provided shapefile. Is the WGS84 coordinate reference system being used for the plot?

We now want to explore the spatial data contained in “Crabs.csv”. Run the following piece of R code

```
Crabs <- read.csv( "Crabs.csv" )
ggplot( data=Alaska ) + theme_bw() + geom_sf() +
  geom_point( data=Crabs, aes(x=longitude, y=latitude), colour=2 )
```

- b) Consider the data in “Crabs.csv” and discuss whether the map we produced is correct.

Let’s change the projection of the shapefile with the `coord_sf()` and `st_crs()` functions covered in Section 4.1.2:

```
ggplot( data=Alaska ) + theme_bw() + geom_sf() +
  coord_sf( crs=st_crs(4326) ) + xlim(-180,-140) +
  geom_point( data=Crabs, aes(x=longitude, y=latitude) )
```

- c) Is this plot more realistic than that in part b)? Adapt the code such that the recorded CPUE is shown in the plot. What do you conclude from the plot?

Instead of changing the projection of the map, let’s define spatial data points using the data in “Crabs.csv”:

```
Crabs_map <- Crabs %>%
  st_as_sf( coords = c("longitude", "latitude"), crs = "WGS84" )
ggplot( data=Alaska ) + theme_bw() + geom_sf() +
  geom_sf( data=Crabs_map, aes(color=cpue) ) +
  labs( x="Longitude", y="Latitude" )
```

- d) Do the locations of the points in the created plot match with the data shown in part c)? What can we learn from the considered R code about the use of projections?

Homework Question - Lead concentration across Amaurot

The local authorities in Amaurot, the capital of Utopia, have seen an alarming increase in lead concentration levels in the drinking water. They thus collected samples on lead concentration across the city. The data on the recorded lead levels and the spatial coordinates are provided in the file “Amaurot Lead.csv”.

The local authorities have now approached you to help them tackle the issue. They provided you with the collected data and a shapefile of Amaurot. To hide Amaurot’s location, the latitude and longitude coordinates have been manipulated, but the provided shapes are correct. You are asked to perform the following tasks:

- a) Visualize the measured lead concentrations. What do you conclude?
- b) Perform inverse distance weighting to predict lead levels for all points in the file “Amaurot Grid.csv”; the points form a regular grid over the whole city and there is also information in which district the point lies. Comment on the reliability of your predictions.
- c) The authorities want to reduce the occurrences of lead concentrations exceeding a level of 10 parts per million. They have thus decided to deploy a team to one of the districts with the job of reducing lead levels to below this threshold for all households in that districts. Use your results in part b) to identify which district should be targeted.