

Understanding the Arne data

Duncan Golicher

14/12/2021

Introduction

The data collected from Arne and used in the assignment is quite complex, as it consists of non spatial data frames and spatial objects. The analysis of these data is shown in the book chapters. However some students have reported difficulties in finding the data. This worksheet is designed to clarify the issue.

Libraries needed

As some of the data analysis uses the GIS elements of R the analysis should usually begin with a chunk that loads the spatial libraries.

```
library(mapview)
library(leaflet.extras)
library(tidyverse)
library(sf)
## Used for splines
library(mgcv)
## The data itself is held in the aqm package
library(aqm)
```

The pine heights data

The worksheet for this is chapter 19

http://r.bournemouth.ac.uk:82/books/qs_21/_book/arne-data-analysis-part-one.html

These data were taken from two sites in 2019. Running this code chunk loads the data. Notice that the first line will ensure that your working environment is empty.

```
rm(list=ls()) # Or clear the environment with the sweeping brush
data(arne_pine_hts)
pine_hts<-st_transform(pine_hts,27700) # British national grid
```

If you click on this object you will see that it has a geometry column. This holds the positions of the trees.

Removing the spatial element

To just conduct a statistical analysis on the data you should remove the geometry column. This code does that and assigns the results to a data frame called “d”

```
pine_hts$Site<-as.factor(pine_hts$Site) # Convert the numerical value of the site to a factor
pine_hts %>% st_drop_geometry() -> d
```

Now you can work with the data frame using statistical techniques.

Pine densities

The worksheet for these data is chapter 20.

http://r.bournemouth.ac.uk:82/books/qs_21/_book/arne-data-analysis-part-two.html

An additional data frame contains the density of the pines and some variables that were derived from the Lidar data and “extracted” from the layers in order to provide variables for the quadrats.

If you run these lines you will obtain these data and assign them to a data frame called d. Note that in order to avoid confusion it is best to work on each data set in a separate markdown document. Remember to add the libraries that you require for anything you want to do with the data.

```
rm(list=ls())
data(arne_quads)
d<-arne_quads[,-c(1:3)]
str(d)
```

```
## 'data.frame': 184 obs. of 6 variables:
## $ pine_density: num 0.875 0.557 0.955 0.239 0.318 0.637 0.398 0.637 0.557 0.477 ...
## $ twi : num 4.24 3.96 3.64 4.14 4.26 ...
## $ sol : num 0.873 0.906 0.86 0.84 0.878 ...
## $ dtm : num 12.6 12.8 13 13.3 13.2 ...
## $ slope : num 1.282 1.934 1.945 0.861 1.22 ...
## $ min_dist : num 14.02 9.11 5.79 11.8 21.26 ...
```

Pine density = numbers of pines per meter squared
twi = Topographic wetness index (no units)
sol = Direct beam insolation
slope = slope in degrees
min_dist = minimum distance to nearest seed source.

Arne spatial analysis

The data provided in the last code chunk was produced through combining the measurements taken on the ground with lidar derived digital surface models and digital terrain models. GIS techniques were used to produce additional layers and to measure distances to large pines.

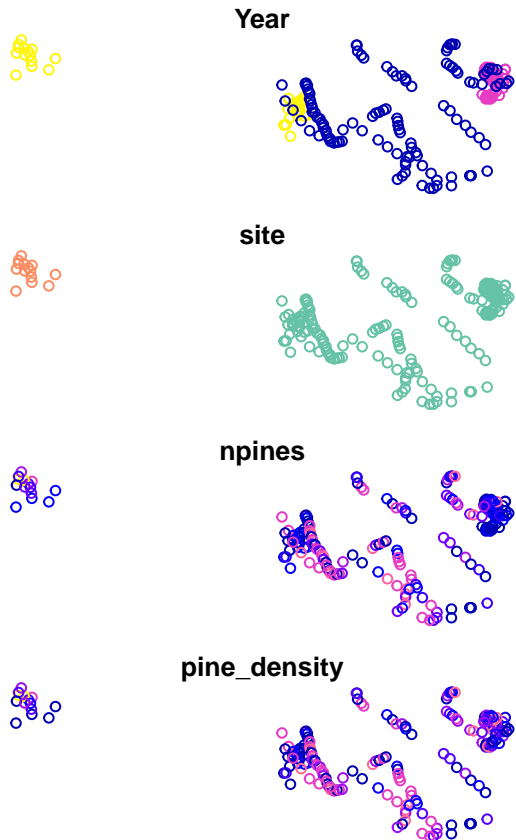
Arne pines quadrat data

The “arne_pines” data from AQM was provided in chapter 22 to show how the quadrats were turned into a spatial object.

```

rm(list=ls())
data("arne_pines")
d<-arne_pines
d<-st_as_sf(arne_pines,coords = c("lon","lat"),crs=4326)
d<-st_transform(d,27700)
plot(d)

```



Lidar data

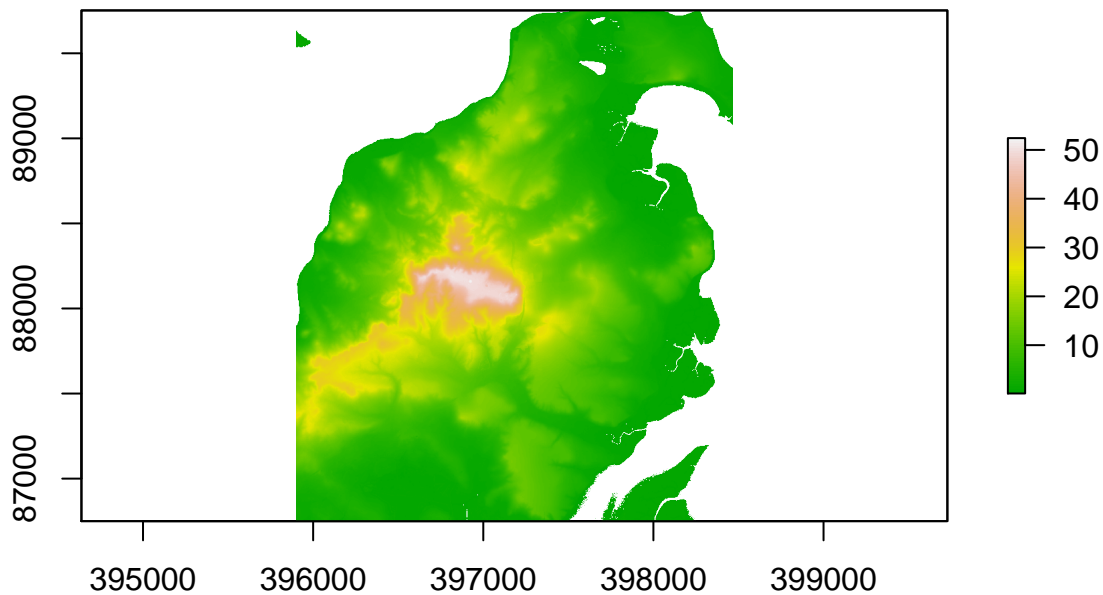
The Lidar layers start off with just two rasters. These are the digital surface model (dsm) and digital terrain model (dtm)

```

library(tidyverse)
library(aqm)
library(giscourse)
library(sf)
library(mapview)
library(raster)
library(leaflet.extras)
library(RColorBrewer)
pal1<-terrain.colors(100)
pal2<-gray(0:100 / 100)

```

```
data(arne_lidar)
plot(dtm, col=pal1)
```



Chapters 22 and 23 show some processing of these data

http://r.bournemouth.ac.uk:82/books/qs_21/_book/arne-pine-density-spatial-data-part-1.html http://r.bournemouth.ac.uk:82/books/qs_21/_book/more-on-distances-in-gis.html

You can also adapt some of the code from chapter 20 on terrain analysis to produce maps. Note that to do this you will have to understand how the operations can be used on a different data set to that shown in chapter 20.

http://r.bournemouth.ac.uk:82/books/qs_21/_book/terrain-analysis.html

Guidance for the write up

The data from Arne is of greater relevance for reserve management than for answering more general scientific questions. If you are using the data to make recommendations for management then some maps showing where the data was collected from and how representative the data may be of the area would be relevant. You may also use maps to explain how the derived variables were obtained and to discuss some other aspects of the reserve. So this section of the assignment provides a chance to include some maps into your report and provide an explanation of how spatial data was used.

Avoiding confusion

Confusions arise when you try to run a code chunk through copying and pasting without looking carefully at the data objects that you have loaded at the time. Understanding the data that you are working with is essential in order to use R. So look at what is in the environment tab and if in doubt click on the object and eyeball it (or use `str`) before running a code chunk. Some large objects such as raster layers cannot be visualised in spreadsheet format. This makes them appear more abstract and less tangible. However they are loaded into memory and can be seen at any point by using the `mapview` function appropriately. Some more information on `mapview` is found here https://r-spatial.github.io/mapview/articles/articles/mapview_01-basics.html

Mistakes and error messages occur when a code chunk is run without knowing beforehand what it does. This won't happen when you are working through the chapters the first time, as the code chunks are provided to show what they do using the data that's loaded at the start of the chapter. **Do always remember to clean your environment first!**

However confusions can, and frequently do, happen when you try to apply code without thinking about which elements in the code need to be changed in order to apply to new data that you have loaded or modified yourself. So do always try to think about what you want to happen before running code, and not afterwards. We all (me included) do run code chunks rather experimentally to see what they produce, but there should not be any major surprises.

Take particular care with chunks that assign results to objects. They won't produce any obvious output in the markdown document. However they do result in objects being added to the working environment. These will be needed for further analysis later on. Again, the GIS component of R can seem particularly difficult in this respect, as a lot of work is being done "behind the scenes". This is not always immediately apparent. You can always look at these objects, but you often don't really need to until you have results that are going to be presented to others. This distinguishes the R way of working from desktop GIS, that tends to produce large numbers of map layers as intermediate steps. Once the R concepts are mastered it becomes quicker and cleaner to work in R than in QGIS. However this advantage is not immediately apparent until you have tried using a desktop GIS to achieve the same sort of results.