Formative_Assesment_TurekJ.R

james

2025-11-07

```
#install required packages

#install.packages("dplyr")
#install.packages("sf")
#install.packages("terra")
#install.packages("tmap")
#install.packages("ggplot2")
#install.packages("viridis")
#install.packages("RColorBrewer")
#install.packages("R.utils")
#install.packages("exactextractr")

#Question 1

#Import the West Midlands and England shapefile and set the CRS to the British National Grid (27700).
#I will use the British National Grid or, EPSG:27700, as my CRS as both shapefiles pertain to places wi
#to me to use this CRS as to not potentially geographically skew any data points.

library(sf)
```

```
## Warning: package 'sf' was built under R version 4.5.2
```

```
## Linking to GEOS 3.13.1, GDAL 3.11.4, PROJ 9.7.0; sf_use_s2() is TRUE
```

```
path_to_data <- "C:/Users/james/OneDrive/Documentos/GY476_Data/GY476_2025_GIS_Files_20250930/Formative"

West_Midlands <- read_sf(paste0(path_to_data, "/Other_Data/west-midlands.gpkg")) %>%
  st_transform(st_crs(27700))

England_shp <- read_sf("C:/Users/james/OneDrive/Documentos/GY476_Data/GY476_2025_GIS_Files_20250930/For
  st_transform(st_crs(27700))


st_crs(West_Midlands) == st_crs(England_shp)
```

```
## [1] TRUE
```

```
#Question 2 (Figure 1)

#Intersect the two shapefiles
```

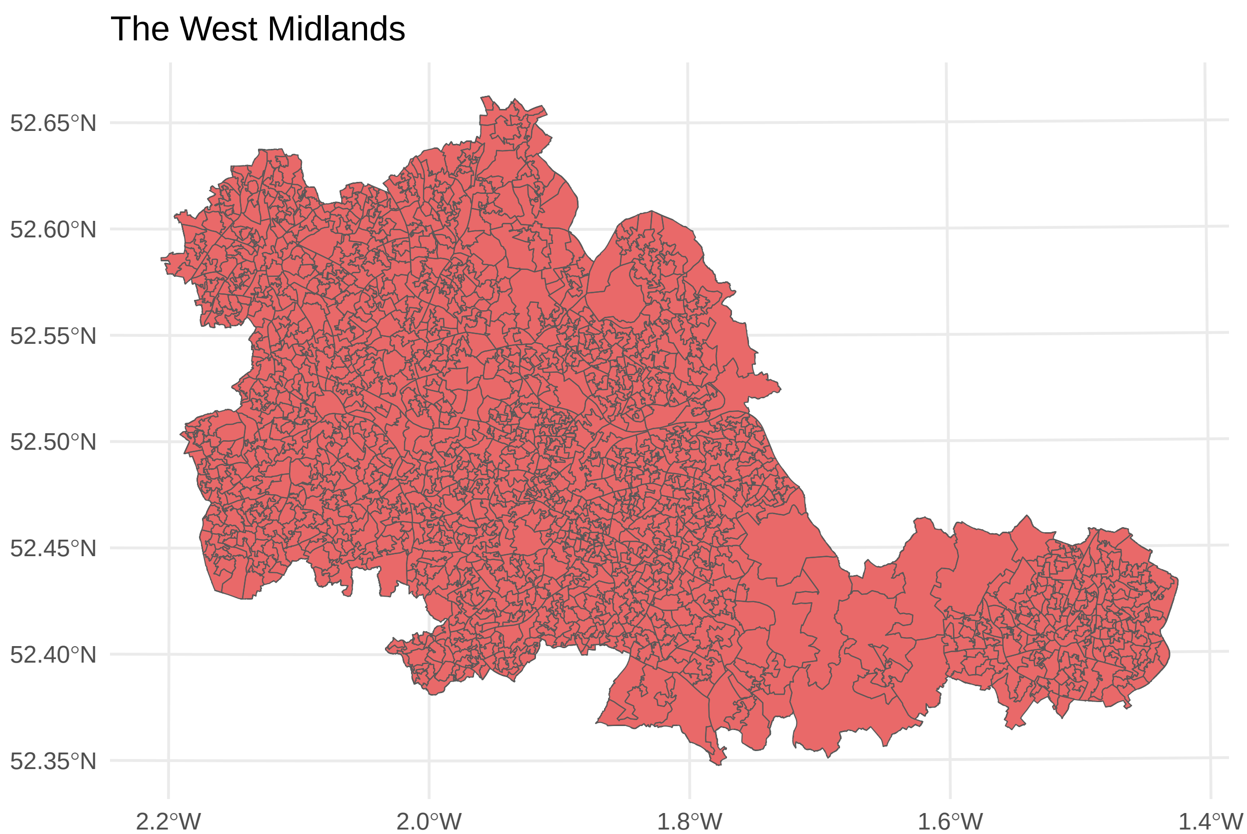
```r
West_Midlands_Intersection <- st_intersection(England_shp, West_Midlands)
```

```
## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.5.2
```

```r
ggplot() +
  geom_sf(data = West_Midlands_Intersection, fill = "lightgrey", color = "white") +
  geom_sf(data = West_Midlands_Intersection, fill = "red", alpha = 0.5) +
  theme_minimal() +
  labs(title = "The West Midlands")
```



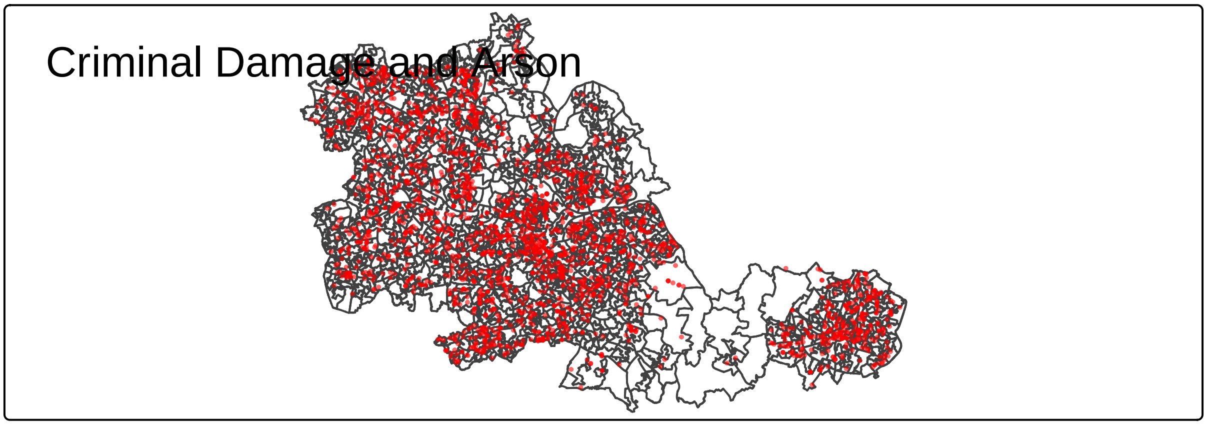

```r
# Import the West Midlands crime data and load/install any geopackages that you need

Crimes <- read.csv(paste0("2023-07-west-midlands-street.csv"))


library(tmap)
```

```
## Warning: package 'tmap' was built under R version 4.5.2
```

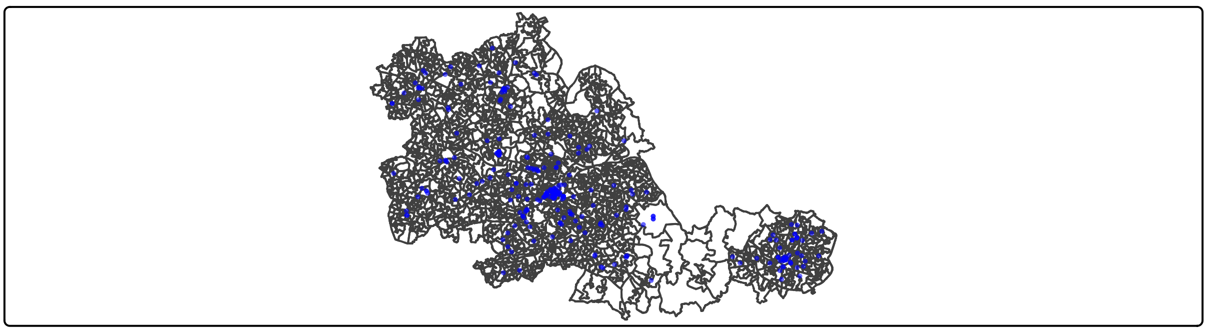
```r
#Filter the data for the crime types

criminal_damage <- Crimes[Crimes$Crime.type == "Criminal damage and arson",]
theft_person <- Crimes[Crimes$Crime.type == "Theft from the person",]

#Convert data to sf objects and remove any rows with missing lat/long data

criminal_damage_clean <- criminal_damage[!is.na(criminal_damage$Longitude) &
                                           !is.na(criminal_damage$Latitude), ]
theft_person_clean <- theft_person[!is.na(theft_person$Longitude) &
                                     !is.na(theft_person$Latitude), ]

#Create sf objects but with BNG (27700)

criminal_damage_sf <- st_as_sf(criminal_damage_clean,
                               coords = c("Longitude", "Latitude"),
                               crs = 4326) %>%
  st_transform(27700)

theft_person_sf <- st_as_sf(theft_person_clean,
                            coords = c("Longitude", "Latitude"),
                            crs = 4326) %>%
  st_transform(27700)

#Use tmap function to create two maps; one showing 'Criminal damage and arson' and another showing 'The

Criminal_Damage_And_Arson <- tm_shape(West_Midlands_Intersection) +
  tm_borders() +
  tm_shape(criminal_damage_sf) +
  tm_dots(col = "red", size = 0.1, alpha = 0.6) +
  tm_layout(title = "Criminal Damage and Arson")
```

```
##

## -- tmap v3 code detected -------------------------------------------------------

## [v3->v4] `tm_dots()`: use 'fill' for the fill color of polygons/symbols
## (instead of 'col'), and 'col' for the outlines (instead of 'border.col').
## [v3->v4] `tm_dots()`: use `fill_alpha` instead of `alpha`.
## [v3->v4] `tm_layout()`: use `tm_title()` instead of `tm_layout(title = )`
## This message is displayed once every 8 hours.
```

```r
Theft_From_Person <- tm_shape(West_Midlands_Intersection) +
  tm_borders() +
  tm_shape(theft_person_sf) +
  tm_dots(fill = "blue", size = 0.1, fill_alpha = 0.6) +
  tm_title("Theft from the Person")

tmap_arrange(Criminal_Damage_And_Arson, Theft_From_Person)
```



Theft from the Person



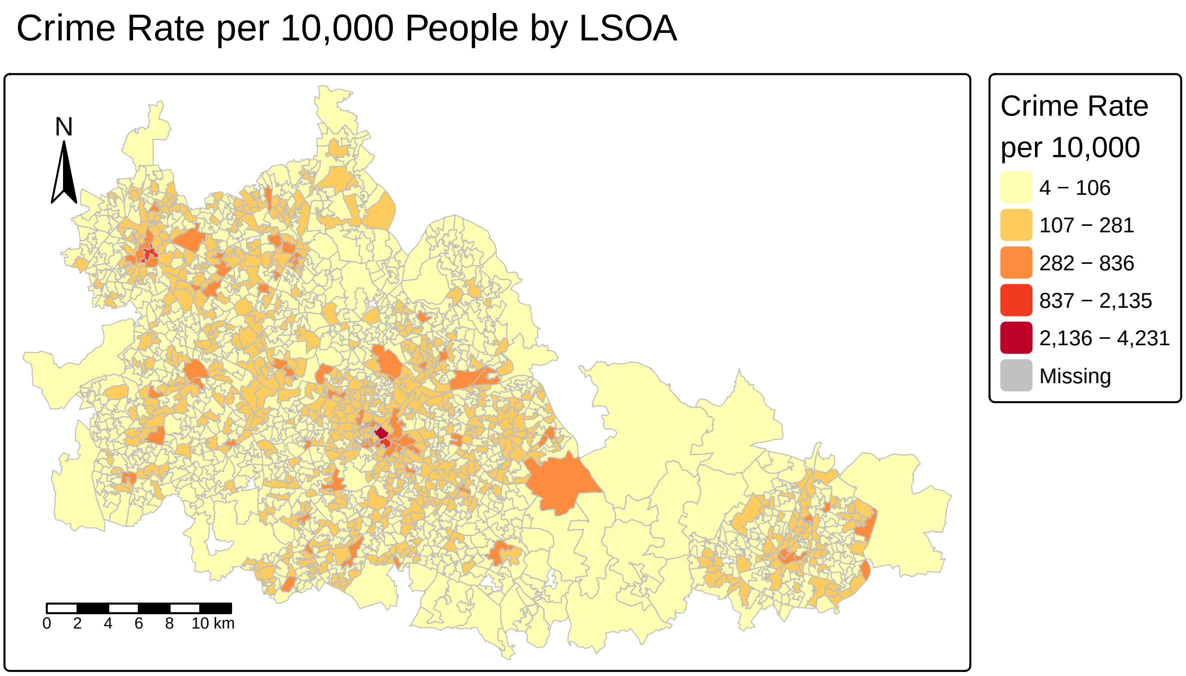
```r
#Analysis
#Looking at the maps, one can immediately see that Criminal Damage and Arson are are far more prevalent
#than Theft From Person. Criminal Damage and Arson seem to be a problem in almost every part of the bou
#Theft from Person, however, seems to be more or less located within the center of the region. This cou
#be due to a dense city in that area meaning more people to steal from.

#Question 3 (Figure 2)

#Load geopackages
library(tmap)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(sf)
```

```r
#Count number of crimes by each LSOA

crime_counts <- Crimes %>%
  group_by(LSOA.code) %>%
  summarise(crime_count = n())

#Import population data

LSOA.population <- read.csv("~/GY476_Data/GY476_2025_GIS_Files_20250930/Formative/Other_Data/TS001_usual

#Execute join by combining popuation data with crime data

crime_pop <- crime_counts %>%
  left_join(LSOA.population, by = c("LSOA.code" = "mnemonic"))

#Calculate the crime rate per 10,000 people

crime_pop <- crime_pop %>%
  mutate(crime_rate = (crime_count * 10000) / total)

#Execute another join, this time combining crime_pop data with the LSOA shapefile

lsoa_crime <- England_shp %>%
  left_join(crime_pop, by = c("lsoa21cd" = "LSOA.code"))

#Plot the choropleth map

lsoa_crime_filtered <- lsoa_crime %>%
  filter(!is.na(crime_rate))

tm_shape(lsoa_crime_filtered) +
  tm_fill("crime_rate",
          palette = "YlOrRd",
          title = "Crime Rate\nper 10,000",
          style = "jenks",
          n = 5) +
  tm_borders(col = "grey", lwd = 0.5) +
  tm_layout(main.title = "Crime Rate per 10,000 People by LSOA",
            main.title.size = 1.2,
            legend.inside = TRUE,
            legend.inside.position = "right") +
  tm_compass(position = c("left", "top")) +
  tm_scale_bar(position = c("left", "bottom"))
```

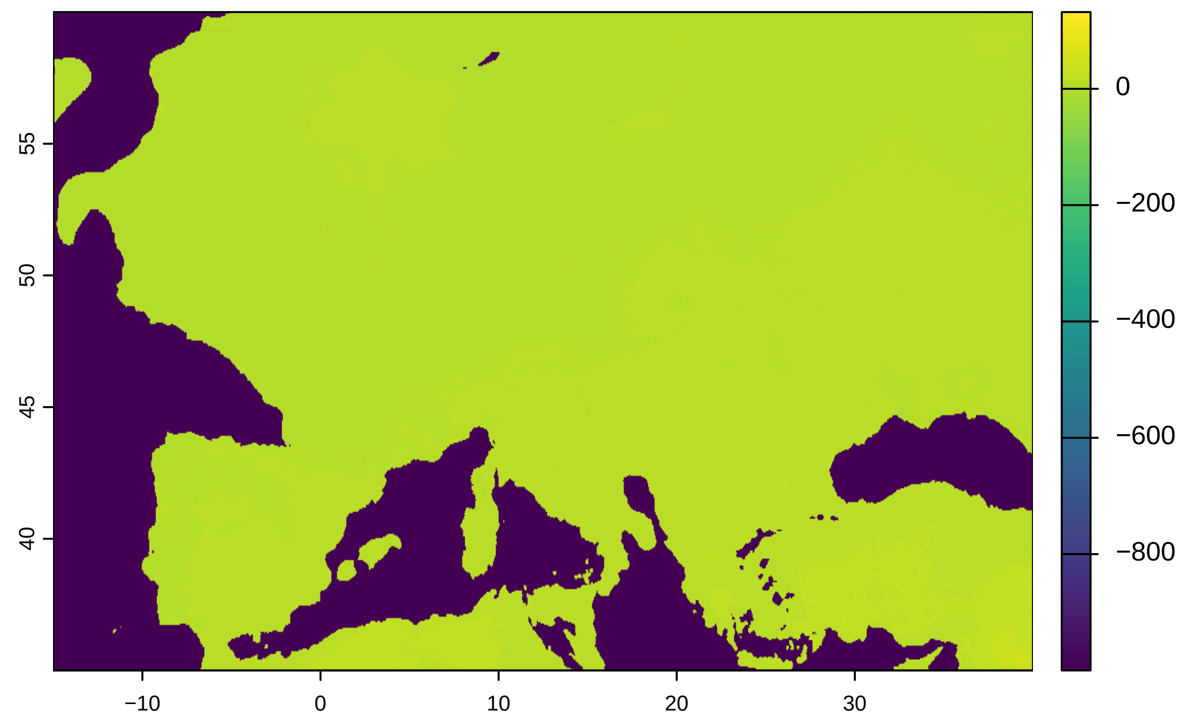
```
##
## -- tmap v3 code detected -------------------------------------------------------
## [v3->v4] `tm_fill()`: instead of `style = "jenks"`, use fill.scale =
## `tm_scale_intervals()`.
## i Migrate the argument(s) 'style', 'n', 'palette' (rename to 'values') to
##   'tm_scale_intervals(<HERE>)'[v3->v4] `tm_fill()`: migrate the argument(s) related to the legend of
## visual variable `fill` namely 'title' to 'fill.legend = tm_legend(<HERE>)'[v3->v4] `tm_layout()`: use
## `cols4all::c4a_gui()` to explore them. The old palette name "YlOrRd" is named
## "brewer.yl_or_rd"Multiple palettes called "yl_or_rd" found: "brewer.yl_or_rd", "matplotlib.yl_or_rd"
```




```r
#Judging from this choropleth map, there seems to be a spatial trend within the inner regions of the
#West Midlands where we can see darker colors, thus depicting a higher rate of crime. You also notice
#a large swath of land towards the east with a very low crime rate. This could suggest that the area is
#dominated by agricultural pastures or crop fields. My choice in selecting the yellow/orange/and red
#color scale is meant to show the high contrast between places with higher crime rates. Particularly,
# the more-or-les geographical center of the West Midlands is clearly shown by the deep red color as ha
#the highest rates of crime in the region.

#Question 4 (Figure 3)

#Load the necessary libraries

library(terra)
```

```
## Warning: package 'terra' was built under R version 4.5.2

## terra 1.8.70
```

```r
library(tmap)
library(sf)

#Import Raster data

pm25_raster <- rast("Other_Data/PM2_5_July2023.tif")
plot(pm25_raster)
```




```
#Import the West Midlands boundary shapefile in order to prepare it for processing with the PM25 Raster

west_midlands_boundary <- st_read("Other_Data/west-midlands.gpkg")
```

```
## Reading layer ‘westmidlands’ from data source
##   ‘C:\Users\james\OneDrive\Documentos\GY476_Data\GY476_2025_GIS_Files_20250930\Formative\Other_Data\
##   using driver ‘GPKG’
## Simple feature collection with 1 feature and 11 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XYZ
## Bounding box:  xmin: -2.206875 ymin: 52.34763 xmax: -1.42396 ymax: 52.66277
## z_range:       zmin: 0 zmax: 0
## Geodetic CRS:  WGS 84
```

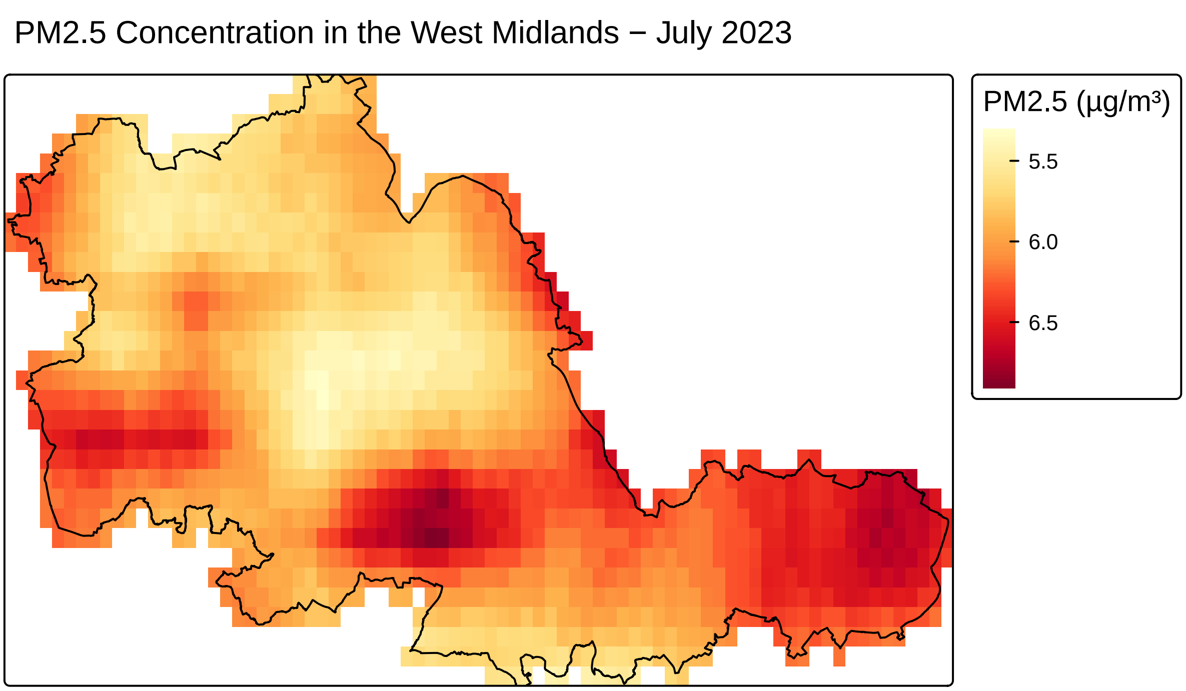
```
#Clip and mask the PM25 Raster to the boundary of the West Midlands

pm25_wm <- crop(pm25_raster, west_midlands_boundary)
pm25_wm <- mask(pm25_wm, west_midlands_boundary)

#Plot the continuous PM25 data

map1 <- tm_shape(pm25_wm) +
  tm_raster(palette = "YlOrRd",
            title = "PM2.5 (µg/m³)",
            style = "cont") +
```

```r
  tm_shape(west_midlands_boundary) +
  tm_borders(col = "black", lwd = 1) +
  tm_layout(main.title = "PM2.5 Concentration in the West Midlands - July 2023",
            main.title.size = 1,
            legend.outside = TRUE)
```

```
##
## -- tmap v3 code detected -------------------------------------------------------
## [v3->v4] `tm_raster()`: instead of `style = "cont"`, use col.scale =
## `tm_scale_continuous()`.
## i Migrate the argument(s) 'palette' (rename to 'values') to
##   'tm_scale_continuous(<HERE>)'[v3->v4] `tm_raster()`: migrate the argument(s) related to the legend
## visual variable `col` namely 'title' to 'col.legend = tm_legend(<HERE>)'[v3->v4] `tm_layout()`: use
```

```r
print(map1)
```

```
## [cols4all] color palettes: use palettes from the R package cols4all. Run
## `cols4all::c4a_gui()` to explore them. The old palette name "YlOrRd" is named
## "brewer.yl_or_rd"
```

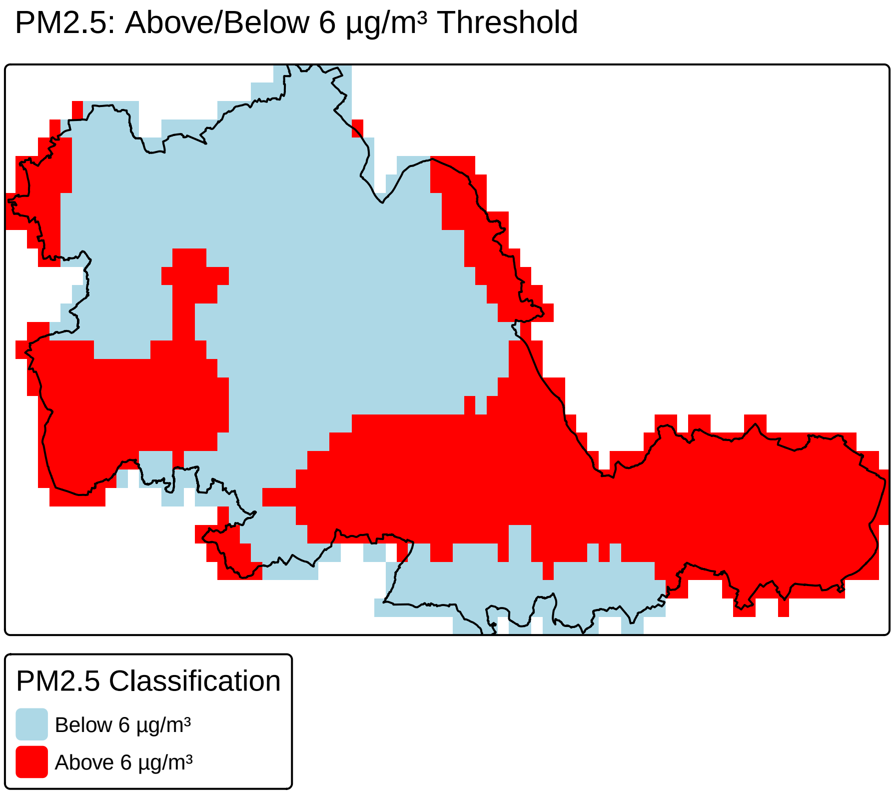



```r
#Reclassify te previous map to show areas with PM2.5 concentrations above and below 6 µg/m³

rcl_matrix <- matrix(c(0, 6, 1,       # Below 6 = 1
```

```
                     6, Inf, 2),    # Above 6 = 2
                   ncol = 3, byrow = TRUE)

pm25_classified <- classify(pm25_wm, rcl_matrix)

#Now plot the reclassified map and print

map2 <- tm_shape(pm25_classified) +
  tm_raster(palette = c("lightblue", "red"),
            labels = c("Below 6 µg/m³", "Above 6 µg/m³"),
            title = "PM2.5 Classification",
            style = "cat") +
  tm_shape(west_midlands_boundary) +
  tm_borders(col = "black", lwd = 1) +
  tm_layout(main.title = "PM2.5: Above/Below 6 µg/m³ Threshold",
            main.title.size = 1,
            legend.outside = TRUE)
```

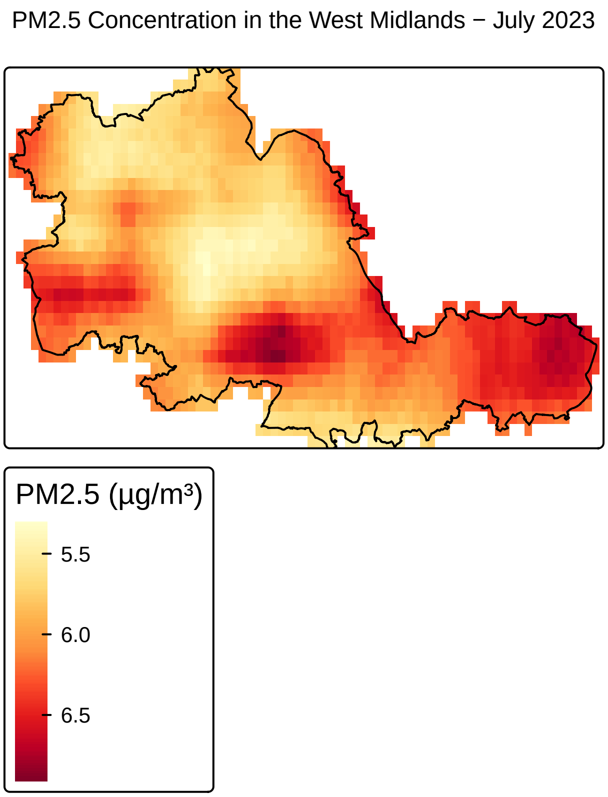
```
##
## -- tmap v3 code detected -------------------------------------------------------
## [v3->v4] `tm_raster()`: instead of `style = "cat"`, use col.scale =
## `tm_scale_categorical()`.
## i Migrate the argument(s) 'palette' (rename to 'values'), 'labels' to
##   'tm_scale_categorical(<HERE>)'[v3->v4] `tm_raster()`: migrate the argument(s) related to the legend
## visual variable `col` namely 'title' to 'col.legend = tm_legend(<HERE>)'[v3->v4] `tm_layout()`: use
```

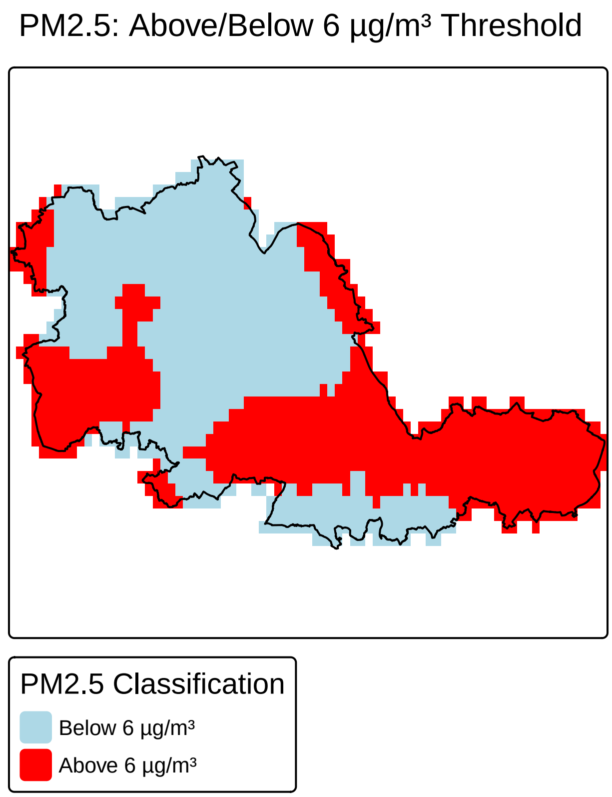
```
print(map2)
```




```r
#Use tmap function to plot them side by side

tmap_arrange(map1, map2, ncol = 2)
```

```
## [cols4all] color palettes: use palettes from the R package cols4all. Run
## ‘cols4all::c4a_gui()‘ to explore them. The old palette name "YlOrRd" is named
## "brewer.yl_or_rd"
## [plot mode] fit legend/component: Some legend items or map compoments do not
## fit well, and are therefore rescaled.
## i Set the tmap option ‘component.autoscale = FALSE‘ to disable rescaling.
```






```r
#By looking at both the crime maps and the PM2.5 concentration maps, one can see that there does
#seem to be some correlation between high crime happens and where there is a PM2.5 concentration above
#6 µg/m³ threshold. The high PM2.5 concentrations looks to be happening mostly around the more densely
#populated areas. Regarding visualization choices, I chose the yellow, orange, and red color palette
#to drive home the point that PM2.5 concentrations are high where the cities exist. For the threshold m
#I wanted to use two distinct colors but also ones that pop out where a viewer can clearly see the disc
#between areas above or below 6 µg/m³.
```

Appendix: R Code

```r
#install required packages

#install.packages("dplyr")
#install.packages("sf")
#install.packages("terra")
#install.packages("tmap")
#install.packages("ggplot2")
#install.packages("viridis")
#install.packages("RColorBrewer")
#install.packages("R.utils")
#install.packages("exactextractr")

#Question 1
```

```
#Import the West Midlands and England shapefile and set the CRS to the British National Grid (27700).
#I will use the British National Grid or, EPSG:27700, as my CRS as both shapefiles pertain to places wi
#to me to use this CRS as to not potentially geographically skew any data points.

library(sf)
path_to_data <- "C:/Users/james/OneDrive/Documentos/GY476_Data/GY476_2025_GIS_Files_20250930/Formative"

West_Midlands <- read_sf(paste0(path_to_data, "/Other_Data/west-midlands.gpkg")) %>%
  st_transform(st_crs(27700))

England_shp <- read_sf("C:/Users/james/OneDrive/Documentos/GY476_Data/GY476_2025_GIS_Files_20250930/Forn
  st_transform(st_crs(27700))


st_crs(West_Midlands) == st_crs(England_shp)

#Question 2 (Figure 1)

#Intersect the two shapefiles

West_Midlands_Intersection <- st_intersection(England_shp, West_Midlands)

library(ggplot2)
ggplot() +
  geom_sf(data = West_Midlands_Intersection, fill = "lightgrey", color = "white") +
  geom_sf(data = West_Midlands_Intersection, fill = "red", alpha = 0.5) +
  theme_minimal() +
  labs(title = "The West Midlands")


# Import the West Midlands crime data and load/install any geopackages that you need

Crimes <- read.csv(paste0("2023-07-west-midlands-street.csv"))


library(tmap)

#Filter the data for the crime types

criminal_damage <- Crimes[Crimes$Crime.type == "Criminal damage and arson",]
theft_person <- Crimes[Crimes$Crime.type == "Theft from the person",]

#Convert data to sf objects and remove any rows with missing lat/long data

criminal_damage_clean <- criminal_damage[!is.na(criminal_damage$Longitude) &
                                          !is.na(criminal_damage$Latitude), ]
theft_person_clean <- theft_person[!is.na(theft_person$Longitude) &
                                    !is.na(theft_person$Latitude), ]

#Create sf objects but with BNG (27700)

criminal_damage_sf <- st_as_sf(criminal_damage_clean,
                               coords = c("Longitude", "Latitude"),
```

```r
                               crs = 4326) %>%
  st_transform(27700)

theft_person_sf <- st_as_sf(theft_person_clean,
                            coords = c("Longitude", "Latitude"),
                            crs = 4326) %>%
  st_transform(27700)

#Use tmap function to create two maps; one showing 'Criminal damage and arson' and another showing 'The

Criminal_Damage_And_Arson <- tm_shape(West_Midlands_Intersection) +
  tm_borders() +
  tm_shape(criminal_damage_sf) +
  tm_dots(col = "red", size = 0.1, alpha = 0.6) +
  tm_layout(title = "Criminal Damage and Arson")

Theft_From_Person <- tm_shape(West_Midlands_Intersection) +
  tm_borders() +
  tm_shape(theft_person_sf) +
  tm_dots(fill = "blue", size = 0.1, fill_alpha = 0.6) +
  tm_title("Theft from the Person")

tmap_arrange(Criminal_Damage_And_Arson, Theft_From_Person)

#Analysis
#Looking at the maps, one can immediately see that Criminal Damage and Arson are are far more prevalent
#than Theft From Person. Criminal Damage and Arson seem to be a problem in almost every part of the bou
#Theft from Person, however, seems to be more or less located within the center of the region. This cou
#be due to a dense city in that area meaning more people to steal from.

#Question 3 (Figure 2)

#Load geopackages
library(tmap)
library(dplyr)
library(sf)

#Count number of crimes by each LSOA

crime_counts <- Crimes %>%
  group_by(LSOA.code) %>%
  summarise(crime_count = n())

#Import population data

LSOA.population <- read.csv("~/GY476_Data/GY476_2025_GIS_Files_20250930/Formative/Other_Data/TS001_usua

#Execute join by combining popuation data with crime data

crime_pop <- crime_counts %>%
  left_join(LSOA.population, by = c("LSOA.code" = "mnemonic"))

#Calculate the crime rate per 10,000 people
```

```r
crime_pop <- crime_pop %>%
  mutate(crime_rate = (crime_count * 10000) / total)

#Execute another join, this time combining crime_pop data with the LSOA shapefile

lsoa_crime <- England_shp %>%
  left_join(crime_pop, by = c("lsoa21cd" = "LSOA.code"))

#Plot the choropleth map

lsoa_crime_filtered <- lsoa_crime %>%
  filter(!is.na(crime_rate))

tm_shape(lsoa_crime_filtered) +
  tm_fill("crime_rate",
          palette = "YlOrRd",
          title = "Crime Rate\nper 10,000",
          style = "jenks",
          n = 5) +
  tm_borders(col = "grey", lwd = 0.5) +
  tm_layout(main.title = "Crime Rate per 10,000 People by LSOA",
            main.title.size = 1.2,
            legend.inside = TRUE,
            legend.inside.position = "right") +
  tm_compass(position = c("left", "top")) +
  tm_scale_bar(position = c("left", "bottom"))

#Judging from this choropleth map, there seems to be a spatial trend within the inner regions of the
#West Midlands where we can see darker colors, thus depicting a higher rate of crime. You also notice
#a large swath of land towards the east with a very low crime rate. This could suggest that the area is
#dominated by agricultural pastures or crop fields. My choice in selecting the yellow/orange/and red
#color scale is meant to show the high contrast between places with higher crime rates. Particularly,
# the more-or-les geographical center of the West Midlands is clearly shown by the deep red color as ha
#the highest rates of crime in the region.

#Question 4 (Figure 3)

#Load the necessary libraries

library(terra)
library(tmap)
library(sf)

#Import Raster data

pm25_raster <- rast("Other_Data/PM2_5_July2023.tif")
plot(pm25_raster)

#Import the West Midlands boundary shapefile in order to prepare it for processing with the PM25 Raster

west_midlands_boundary <- st_read("Other_Data/west-midlands.gpkg")

#Clip and mask the PM25 Raster to the boundary of the West Midlands
```

```r
pm25_wm <- crop(pm25_raster, west_midlands_boundary)
pm25_wm <- mask(pm25_wm, west_midlands_boundary)

#Plot the continuous PM25 data

map1 <- tm_shape(pm25_wm) +
  tm_raster(palette = "YlOrRd",
            title = "PM2.5 (µg/m³)",
            style = "cont") +
  tm_shape(west_midlands_boundary) +
  tm_borders(col = "black", lwd = 1) +
  tm_layout(main.title = "PM2.5 Concentration in the West Midlands - July 2023",
            main.title.size = 1,
            legend.outside = TRUE)
print(map1)


#Reclassify te previous map to show areas with PM2.5 concentrations above and below 6 µg/m³

rcl_matrix <- matrix(c(0, 6, 1,      # Below 6 = 1
                       6, Inf, 2),    # Above 6 = 2
                     ncol = 3, byrow = TRUE)

pm25_classified <- classify(pm25_wm, rcl_matrix)

#Now plot the reclassified map and print

map2 <- tm_shape(pm25_classified) +
  tm_raster(palette = c("lightblue", "red"),
            labels = c("Below 6 µg/m³", "Above 6 µg/m³"),
            title = "PM2.5 Classification",
            style = "cat") +
  tm_shape(west_midlands_boundary) +
  tm_borders(col = "black", lwd = 1) +
  tm_layout(main.title = "PM2.5: Above/Below 6 µg/m³ Threshold",
            main.title.size = 1,
            legend.outside = TRUE)
print(map2)

#Use tmap function to plot them side by side

tmap_arrange(map1, map2, ncol = 2)

#By looking at both the crime maps and the PM2.5 concentration maps, one can see that there does
#seem to be some correlation between high crime happens and where there is a PM2.5 concentration above
#6 µg/m³ threshold. The high PM2.5 concentrations looks to be happening mostly around the more densely
#populated areas. Regarding visualization choices, I chose the yellow, orange, and red color palette
#to drive home the point that PM2.5 concentrations are high where the cities exist. For the threshold m
#I wanted to use two distinct colors but also ones that pop out where a viewer can clearly see the disc
#between areas above or below 6 µg/m³.
```