Introduction to maps in R

Sumali Bajaj

2025-04-25

Introduction

Mapping is a powerful way to visualize spatial data. In this practical, you'll learn how to:

- Work with **shapefiles** using the `sf` package
- Visualize spatial data using `ggplot2`
- Combine data with spatial boundaries

What is a shapefile?

A **shapefile** is a widely used geospatial file format that contains the geometry (shapes) and attributes (metadata) of spatial features like countries, districts, or villages.

A shapefile typically consists of multiple files (`.shp`, `.shx`, `.dbf`, etc.) and must be kept together in the same folder.

We'll use a shapefile of **district boundaries in Meghalaya**, and join it with **survey data collected for STH**.

Packages Required

```
# Install if you haven't already
# install.packages(c("sf", "dplyr", "ggplot2"))

# Load libraries
library(sf)
library(dplyr)
library(ggplot2)
```

Step 1: Load your data and shapefile

- (Make sure you write the correct file path!)

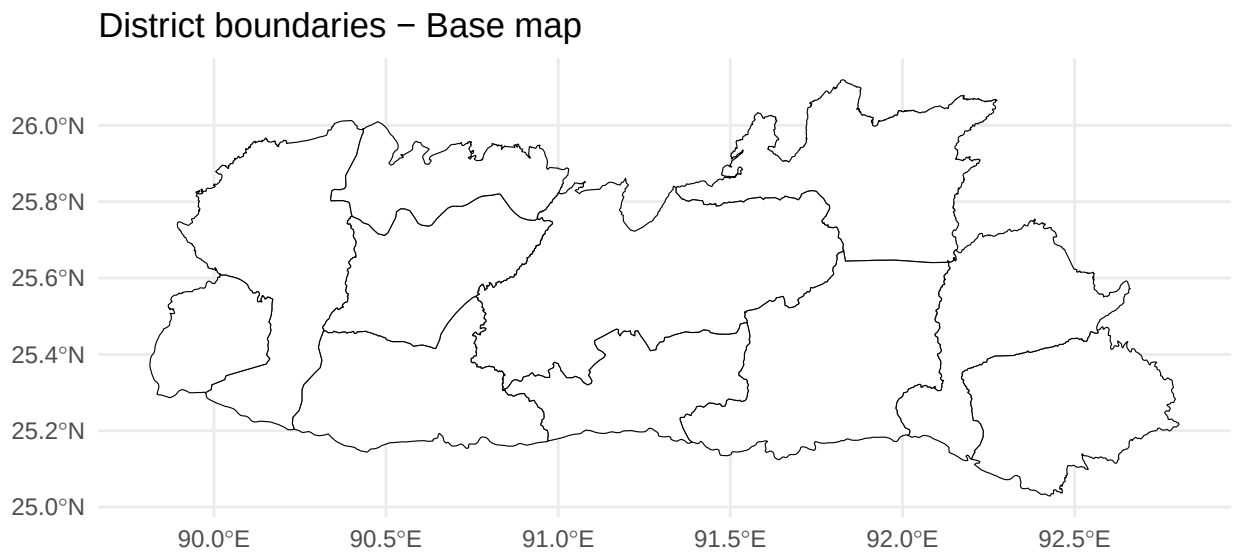
```r
# Load survey data
dat <- read.csv("../../data/processed/STH_data.csv")

# Load shapefile (district boundaries)
ml <- read_sf("../../data/raw/ML_shapefile/India_District_17_ML.shp")
```

Step 2: Visualize the shapefile map outline

- Before we overlay any data, let's plot just the **district boundaries** from the shapefile. This helps verify that the shapefile is loaded and projected correctly.

```r
# Load and plot the shapefile only
ggplot() +
  geom_sf(data = ml, fill = "white", color = "black") +
  theme_minimal() +
  labs(title = "District boundaries - Base map")
```



Step 3: Prepare the data

- We group the survey data by village and calculate a total count of microscopy-positive individuals.

```r
# Summarise data by village
dat_map <- dat %>%
  group_by(village) %>%
  mutate(positive_total = sum(overall_sth_microscopy))
```

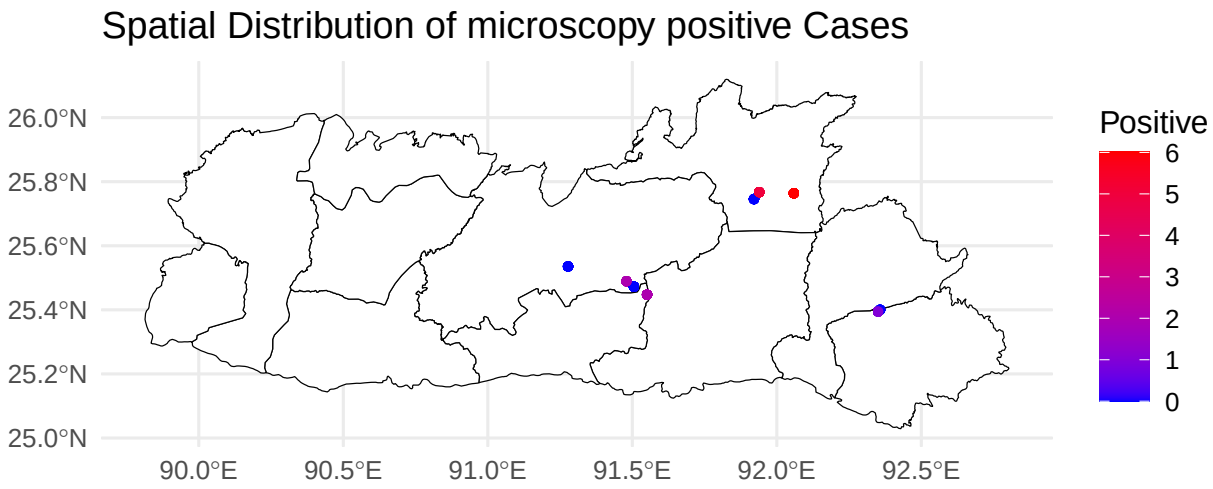
Step 4: Plotting the map with metadata

1. Convert the data to a spatial format

```r
# Convert to spatial object using coordinates
# WGS84 coordinate system
dat_sf <- st_as_sf(dat_map, coords = c("Longitude", "Lattitude"), crs = 4326)
```

2. Creating a map with data

```r
ggplot() +
  # Plot district polygons
  # a
  geom_sf(data = ml, fill = "white", color = "black") +
  # Plot points colored by total PCR-positive
  geom_sf(data = dat_sf, aes(color = positive_total), size = 1) +
  scale_color_gradient(low = "blue", high = "red") +  # Color scale
  theme_minimal() +
  labs(title = "Spatial Distribution of microscopy positive Cases",
       color = "Positive")
```

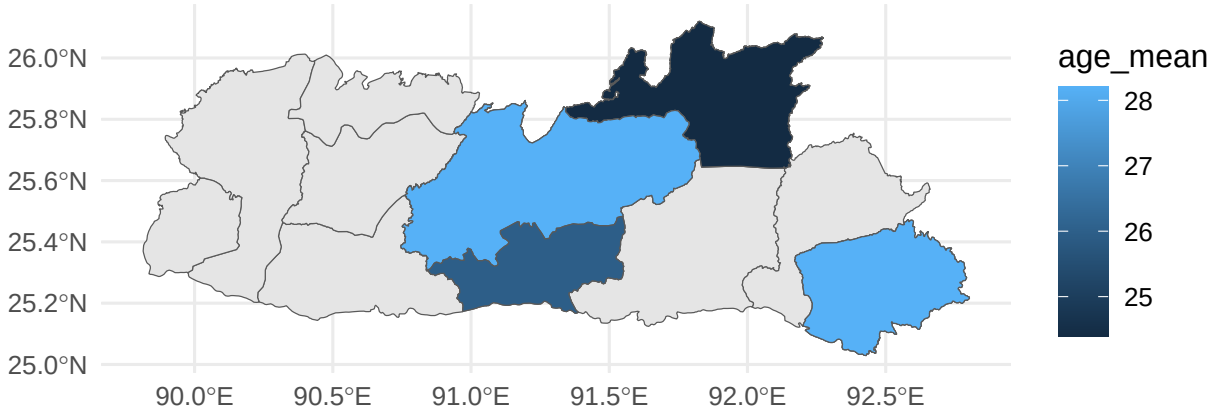


3. Create a new dataset of average age of males and females for each district. Then combine it with the `ml` shapefile and plot it.

```
dat_sf_summary <- dat_sf %>%
  group_by(district, gender) %>%
  summarise(age_mean = mean(age, na.rm = TRUE)) %>%
  na.omit()

districts_merged <- st_join(ml, dat_sf_summary)

ggplot(data = districts_merged %>% filter(gender == "Female")) +
  geom_sf(data = ml) +
  geom_sf(aes(fill = age_mean)) +
  theme_minimal()
```



Practical exercises

Create maps for some of the variables which might be associated with STH infection.

For example - percentage of people who wear footwear occasionally for each district.