

Introduction to maps in R

Sumali Bajaj

2025-04-16

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!)

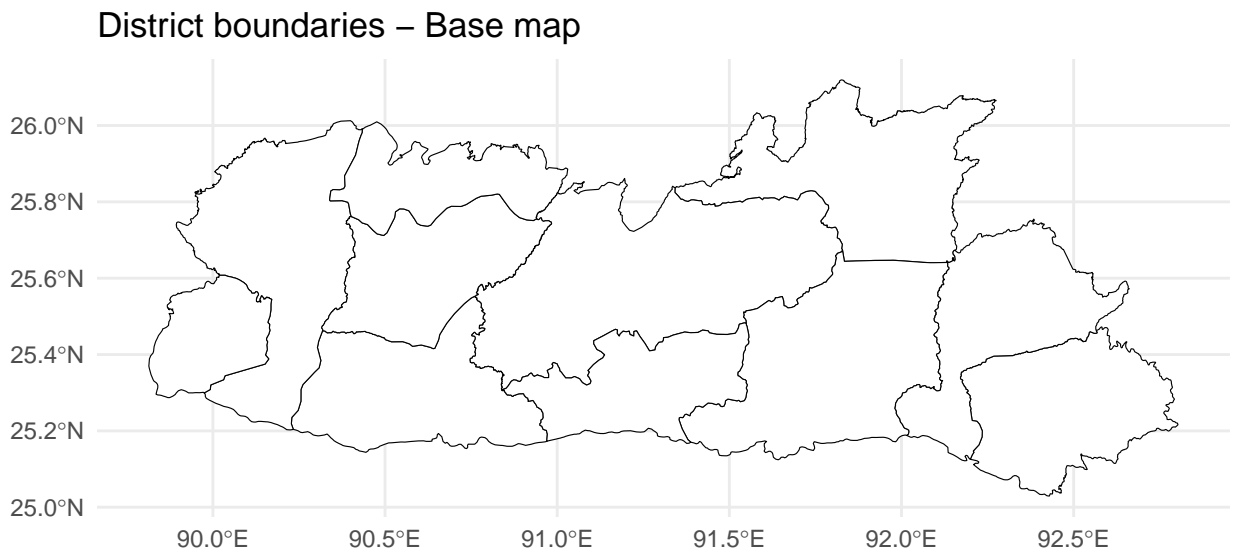
```
# Load survey data
dat <- read.csv("../data/processed/STH_data.csv")

# 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.

```
# 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 PCR-positive individuals.

```
# Summarise data by village
dat_map <- dat %>%
  group_by(village) %>%
  mutate(pcr_positive_total = sum(pcr_positive))
```

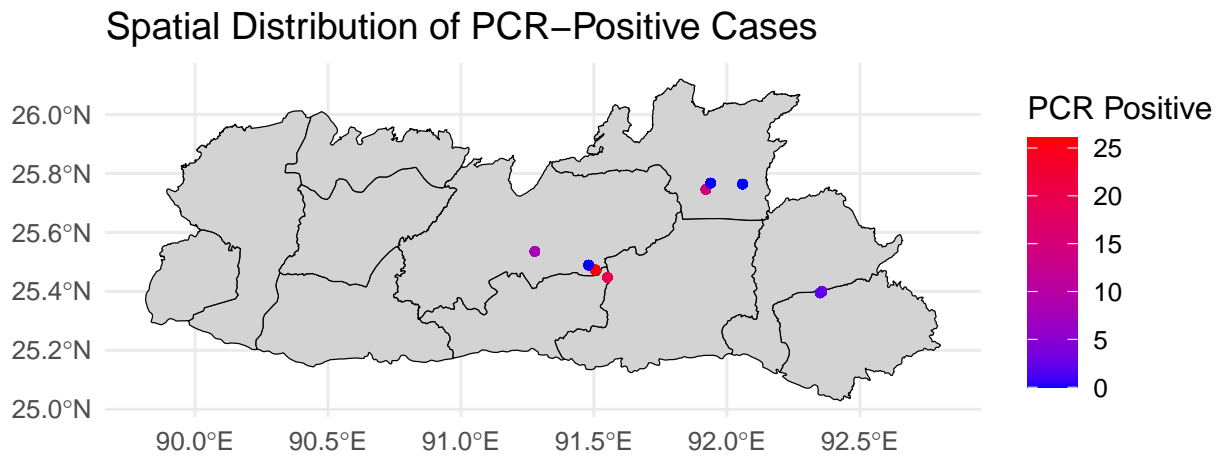
Step 4: Plotting the map with metadata

1. Convert the data to a spatial format

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

2. Creating a map with data

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



Practical exercises

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

For example - percentage of people who use manure for each district.