

Mapping Geospatial Data with R

Md Sakhawat Hossen

Former Data Analyst at Navana Group, Bangladesh
sakhawat3003@gmail.com

06/24/2022

- [Motivation](#)
- [Plotting the Data](#)
- [Geospatial Analysis](#)

Motivation

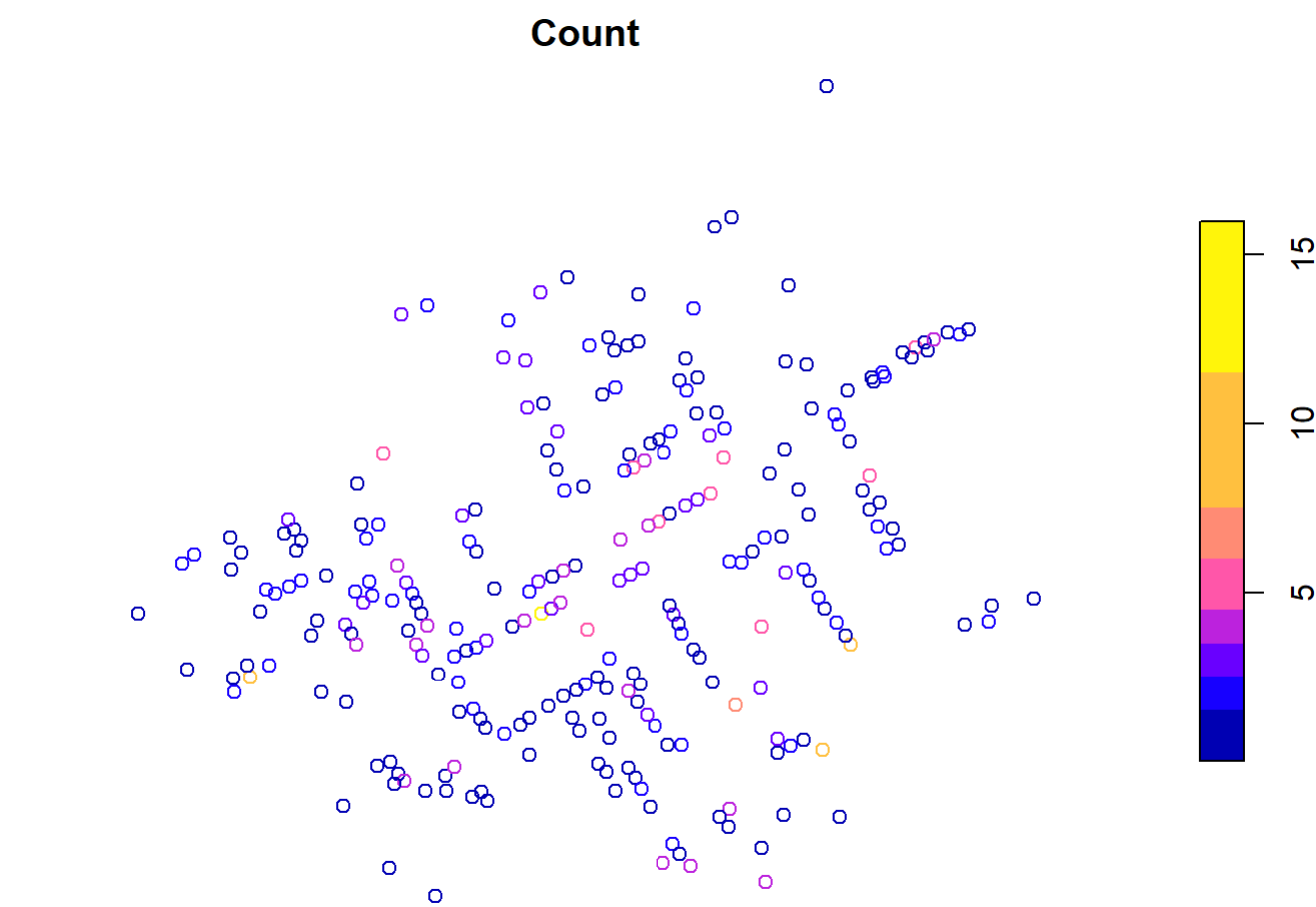
In the year 1854, there was an outbreak of Cholera disease in a London neighborhood killing many peoples. Everybody suspected it was an airborne disease except one: physician *John Snow*, not to be mixed up with the famous *Game of Thrones* character *John Snow*. Doctor Snow created a famous map depicting the location of the people died in the area. Curiously, the locations were all clustered around a water pump! From this inspection, Doctor Snow suggested that the Cholera was spreading through the water of this pump, not the air. And later his description was found to be exactly true!

In this short project, we will walk through reproducing the map that Doctor Snow created a long time ago.

Plotting the Data

The *CholeraDeaths* data collected by Doctor Snow is a part of *mdsr* library. We will plot the address of individuals died on the occasion.

```
library(tidyverse)
library(mdsr)
library(sf)
plot(CholeraDeaths["Count"])
```



Geospatial Analysis

Although, we might observe certain pattern in the plot but so far no context has been provided for making a judgment. We need to overlay the street information over the data to draw any implications. Snow's insight was driven by another set of data: the locations of the street-side water pumps.

There are many libraries in R specialized for geospatial analysis. But, we will focus mostly on the more advanced and sophisticated *sf* library. The shape files for Cholera Data has been downloaded.

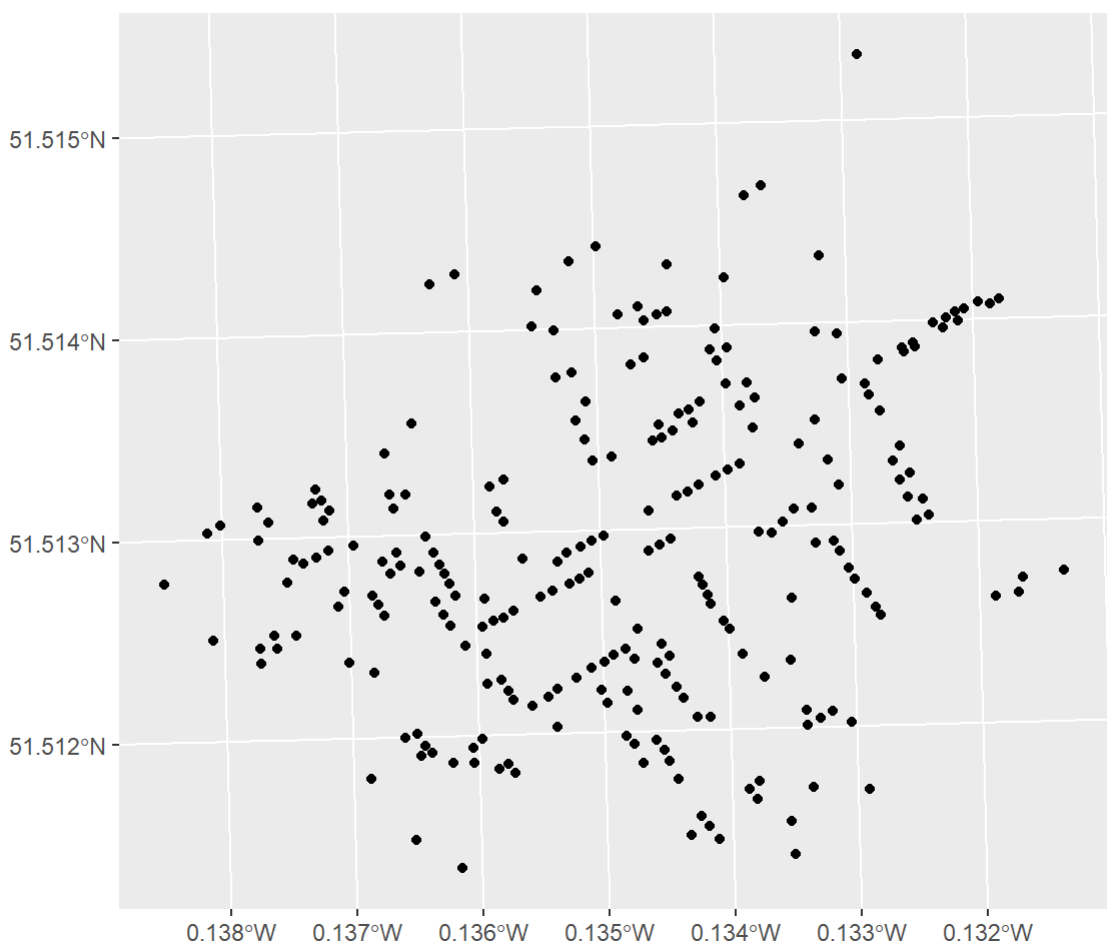
```
dsn<-fs::path("C:\\Users\\PC\\R Working Directory\\Datasets for Statistical Analysis\\SnowGIS_SHP",
              "SnowGIS_SHP")
list.files(dsn)
```

```
## [1] "Cholera_Deaths.dbf"      "Cholera_Deaths.prj"
## [3] "Cholera_Deaths.sbn"     "Cholera_Deaths.sbx"
## [5] "Cholera_Deaths.shp"     "Cholera_Deaths.shx"
## [7] "OSMap.tif"              "OSMap.tif"
## [9] "OSMap_Grayscale.tif"    "OSMap_Grayscale.tif"
## [11] "OSMap_Grayscale.tif.aux.xml" "OSMap_Grayscale.tif.ovr"
## [13] "Pumps.dbf"              "Pumps.prj"
## [15] "Pumps.sbx"              "Pumps.shp"
## [17] "Pumps.shx"              "README.txt"
## [19] "SnowMap.tifw"           "SnowMap.tif"
## [21] "SnowMap.tif.aux.xml"    "SnowMap.tif.ovr"
```

There are six files with the name Cholera_Deaths and another five with the name Pumps. These are related to two different sets of shape files called layers.

We can plot geospatial objects by using the function *geom_sf* which follows as an extension of *ggplot* object.

```
ggplot(CholeraDeaths) +
  geom_sf()
```



This is an improvement over which we got previously from the *plot* function. But still we have not provided any context for the data to be intelligible.

In the documentation of the *CholeraDeaths* original data, it is suggested to use the specification *epsg:27700*. So, we first follow that the CholeraDeaths data is under the *epsg:27700* specification. Then, we will project to *epsg:4326*.

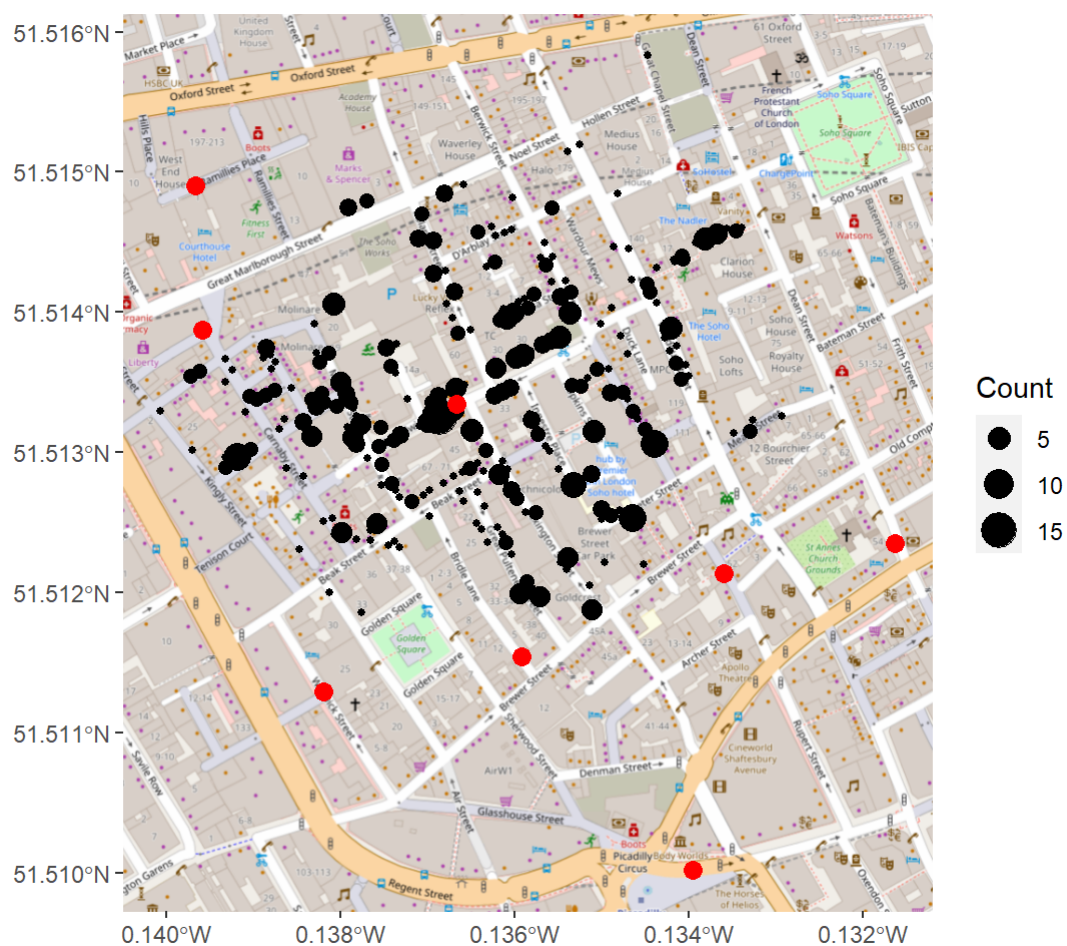
```
library(ggspatial)
cholera_latlong <- CholeraDeaths %>%
  st_set_crs(27700) %>%
  st_transform(4326)
snow <- ggplot(cholera_latlong) +
  annotation_map_tile(type = "osm", zoomin = 0) +
  geom_sf(aes(size = Count))
```

Now, we will add the locations of the water pumps around the area.

```
pumps <- st_read(dsn, layer = "Pumps")
```

```
## Reading layer `Pumps' from data source
##   `C:\Users\PC\R Working Directory\Datasets for Statistical Analysis\SnowGIS_SHP\SnowGIS_SHP'
##   using driver `ESRI Shapefile'
## Simple feature collection with 8 features and 1 field
## Geometry type: POINT
## Dimension:      XY
## Bounding box:   xmin: 529183.7 ymin: 180660.5 xmax: 529748.9 ymax: 181193.7
## Projected CRS: OSGB 1936 / British National Grid
```

```
pumps_latlong <- pumps %>%
  st_set_crs(27700) %>%
  st_transform(4326)
snow +
  geom_sf(data = pumps_latlong, size = 3, color = "red")
```



Finally, the map created by Doctor John Snow is reproduced here. The size of each black dot is proportional to the number of deaths by cholera. And the red dots indicate the locations of the water pump. From the map, it was discernible to Doctor Snow to draw the conclusion on the probable cause of Cholera outbreak and death of many people. It was the polluted water pump around which the locations of the deceased are clustered.