

STAT 209

Spatial/Geographic Data

December 10, 2019

Colin Reimer Dawson

Outline

Spatial Data

Projections

Cholera Outbreaks

Date	Last Name	First Name	Address	Age	Cause of death
Aug 31, 1854	Jones	Thomas	26 Broad St.	37	cholera
Aug 31, 1854	Jones	Mary	26 Broad St.	11	cholera
Oct 1, 1854	Warwick	Martin	14 Broad St.	23	cholera
⋮					

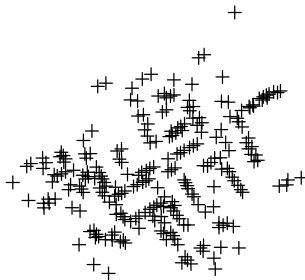
Table 14.1: Hypothetical data from 1854 cholera outbreak.

Figure: A partial list of patients dying of cholera in an 1854 outbreak (Source: MDSR p. 318)

Prevailing wisdom: cholera was transmitted through the air.
Can we use data visualization to find patterns in this outbreak?

Plotting the locations of the cases

```
library(mdsr)  
library(sp) # for spatial data  
plot(CholeraDeaths)
```



That's... not that comprehensible

Cholera Outbreaks

Would like something more like this...

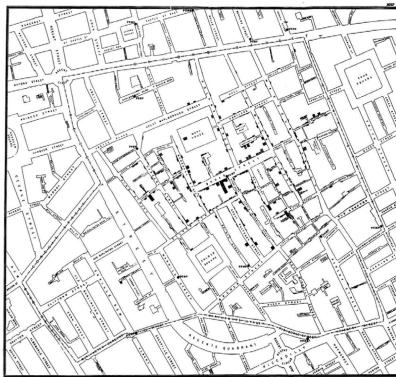


Figure: A hand-drawn map of cholera cases by John Snow, overlaying cases on streets and water pumps

Spatial Layers

```
## data archive at http://rtwilson.com/downloads/SnowGIS\_SHP.zip  
root <- "~/stat209/data/"  
dsn <- paste0(root, "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_Grayscale.tfw"           "OSMap_Grayscale.tif"  
[9] "OSMap_Grayscale.tif.aux.xml"   "OSMap_Grayscale.tif.ovr"  
[11] "OSMap.tfw"                     "OSMap.tif"  
[13] "Pumps.dbf"                     "Pumps.prj"  
[15] "Pumps.sbx"                     "Pumps.shp"  
[17] "Pumps.shx"                     "README.txt"  
[19] "SnowMap.tfw"                   "SnowMap.tif"  
[21] "SnowMap.tif.aux.xml"           "SnowMap.tif.ovr"
```

Note two groups: Cholera_Deaths.* and Pumps.*. These are “layers” of spatial data

Listing Layers

```
library(rgdal)
ogrListLayers(dsn)

[1] "Cholera_Deaths" "Pumps"
attr(,"driver")
[1] "ESRI Shapefile"
attr(,"nlayers")
[1] 2
```

Viewing Layer Info

```
ogrInfo(dsn, layer = "Cholera_Deaths")
```

```
Source: "/Users/cdawson/shared/teaching/stat209/data/SnowGIS_SHP", layer:  
Driver: ESRI Shapefile; number of rows: 250  
Feature type: wkbPoint with 2 dimensions  
Extent: (529160.3 180857.9) - (529655.9 181306.2)  
CRS: +proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000 +y_0=-100  
LDID: 87  
Number of fields: 2  
  name type length typeName  
1   Id    0       6   Integer  
2 Count  0       4   Integer
```

Contains 250 locations, each with a count (number of cases) and an Id

Read into R as a Spatial Data Frame

```
CholeraDeaths <- readOGR(dsn, layer = "Cholera_Deaths")
```

OGR data source with driver: ESRI Shapefile

Source: "/Users/cdawson/shared/teaching/stat209/data/SnowGIS_SHP", layer:
with 250 features

It has 2 fields

```
head(CholeraDeaths)
```

	coordinates	Id	Count
1	(529308.7, 181031.4)	0	3
2	(529312.2, 181025.2)	0	2
3	(529314.4, 181020.3)	0	1
4	(529317.4, 181014.3)	0	1
5	(529320.7, 181007.9)	0	4
6	(529336.7, 181006)	0	2

Display a Summary

```
summary(CholeraDeaths)
```

```
Object of class SpatialPointsDataFrame
```

```
Coordinates:
```

```
              min      max  
coords.x1 529160.3 529655.9  
coords.x2 180857.9 181306.2
```

```
Is projected: TRUE
```

```
proj4string :
```

```
[+proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000  
+y_0=-100000 +ellps=airy +units=m +no_defs]
```

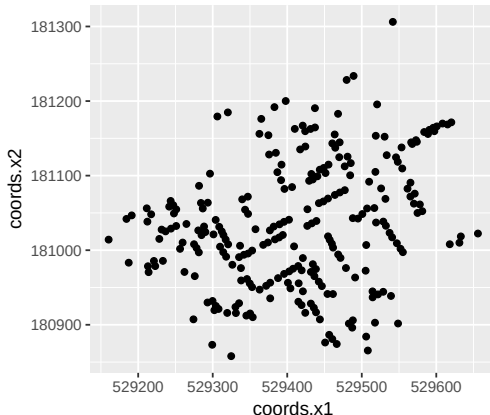
```
Number of points: 250
```

```
Data attributes:
```

	Id	Count
Min.	:0	Min. : 1.000
1st Qu.:	:0	1st Qu.: 1.000
Median	:0	Median : 1.000
Mean	:0	Mean : 1.956
3rd Qu.:	:0	3rd Qu.: 2.000
Max.	:0	Max. :15.000

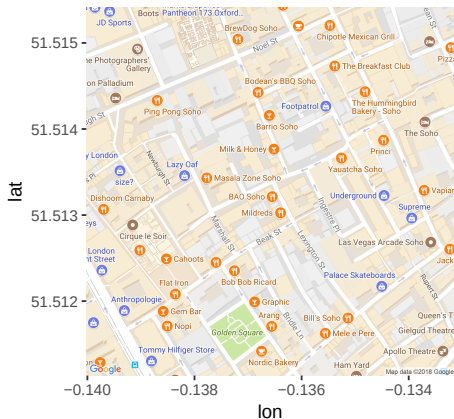
Quick-and-Dirty Map of Coordinates

```
cholera_coords <- CholeraDeaths %>% coordinates() %>% as.data.frame()
ggplot(cholera_coords) +
  geom_point(aes(x = coords.x1, y = coords.x2)) +
  coord_quickmap()
```



Street Map Context

```
library(ggmap)
m <- get_map("John Snow, London, England", zoom = 17, maptype = "roadmap")
ggmap(m)
```



Overlaying on a Street Map

```
ggmap(m) +  
  geom_point(data = as.data.frame(CholeraDeaths),  
    aes(x = coords.x1, y = coords.x2, size = Count))
```



... Where's the cholera data?

Different Coordinate Systems

```
## Cholera coordinates
CholeraDeaths %>% as.data.frame() %>% head()
```

	Id	Count	coords.x1	coords.x2
1	0	3	529308.7	181031.4
2	0	2	529312.2	181025.2
3	0	1	529314.4	181020.3
4	0	1	529317.4	181014.3
5	0	4	529320.7	181007.9
6	0	2	529336.7	181006.0

```
## Map coordinates (bb = "Bounding box")
attr(m, "bb")
```

	ll.lat	ll.lon	ur.lat	ur.lon
1	51.51113	-0.1400205	51.51541	-0.133154

Outline

Spatial Data

Projections

Geographic Projections

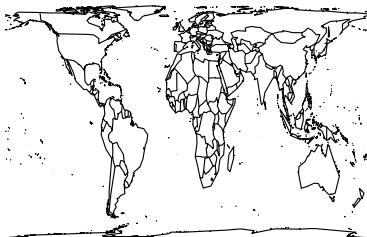
```
library(maps)  
map("world", projection = "mercator", wrap = TRUE)
```



Mercator projection preserves angles, but distorts areas

Another Projection

```
map("world", projection = "cylequalarea", param = 45, wrap = TRUE)
```



Gall-Peters projection preserves areas, distorts angles

Another Angle-Preserving Projection

```
map("state", projection = "lambert",  
    parameters = c(lat0 = 20, lat1 = 50), wrap = TRUE)
```



Another Angle-Preserving Projection

```
map("state", projection = "albers",  
    parameters = c(lat0 = 20, lat1 = 50), wrap = TRUE)
```



Distorts both a little, tries to minimize overall distortion

How are Cholera Deaths Projected?

```
proj4string(CholeraDeaths) %>% strwrap()  
  
[1] "+proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000"  
[2] "+y_0=-100000 +ellps=airy +units=m +no_defs"
```

Coordinate Reference Systems

EPSG:4326 Also known as WGS84, this is the standard for GPS systems and Google Earth.

EPSG:3857 A Mercator projection used in maps tiles⁴ by Google Maps, Open Street Maps, etc.

EPSG:27700 Also known as OSGB 1936, or the British National Grid: United Kingdom Ordnance Survey. It is commonly used in Britain.

Transforming the Data to GMaps format

```
cholera_latlong <- CholeraDeaths %>% spTransform(CRS("+init=epsg:4326"))  
bbox(cholera_latlong)
```

	min	max
coords.x1	-0.1384685	-0.1313274
coords.x2	51.5113460	51.5153252

```
ggmap(m) +  
  geom_point(aes(x = coords.x1, y = coords.x2, size = Count),  
             data = as.data.frame(cholera_latlong))
```

Not Quite Right...

```
proj4string(CholeraDeaths) <- CRS("+init=epsg:27700")
```

Need to dive into data documentation to find out that we need an extra transformation


```
cholera_latlong <- CholeraDeaths %>% spTransform(CRS("+init=epsg:4326"))
snow <- ggmap(m) +
  geom_point(aes(x = coords.x1, y = coords.x2,
                 size = Count), data = as.data.frame(cholera_latlong))
pumps <- readOGR(dsn, layer = "Pumps")
```

OGR data source with driver: ESRI Shapefile

Source: "/Users/cdawson/shared/teaching/stat209/data/SnowGIS_SHP", layer:
with 8 features

It has 1 fields

```
proj4string(pumps) <- CRS("+init=epsg:27700")
pumps_latlong <- pumps %>% spTransform(CRS("+init=epsg:4326"))
```

Finally...

```
SNOW +  
geom_point(data = as.data.frame(pumps_latlong),  
  aes(x = coords.x1, y = coords.x2), size = 3, color = "red")
```

