Thematic Maps

Introduction

Thematic maps portrait geographic patterns about a particular subject theme in a geographic area. In the context of epidemiological outbreaks, these maps can be called *Epidemic maps*.

In this lesson we are going to learn about the most iconic types of Thematic maps to visualize your spatial data: **Choropleth maps** and **Dot maps**.

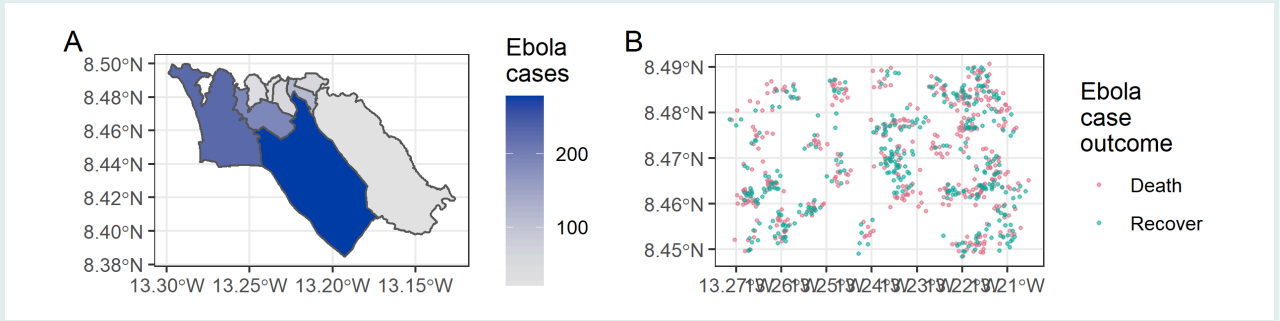


Figure 1. (A) Choropleth map with the number of cases in a simulated Ebola epidemic in Sierra Leona. (B) Dot map. Location of simulated Ebola cases in Sierra Leona, colored by each case outcome.

Learning objectives

1. Identify two types of **Thematic maps** (choropleth and dot maps) used by epidemiologist to visualize Geospatial data.

2. Create Thematic maps using `{ggplot2}` and the **`geom_sf()`** function.

3. Relate each Thematic map with a **Geometry** type.

Prerequisites

This lesson requires the following packages:

```
if(!require('pacman')) install.packages('pacman')
pacman::p_load_gh("afrimapr/afrilearndata")
pacman::p_load(rnaturalearth,
               ggspatial,
               ggplot2,
               mdsr,
               terra,
               spData,
               readr)
```

This lesson requires familiarity with `{ggplot2}`: if you need to brush up, have a look at our introductory course on data visualization.

Choropleth map

What is it?

A *Choropleth map* is a type of thematic map where *colors, shading, or patterns* are used to represent **geographic regions** in relation to the value of an attribute.

For instance a *larger value* could be indicated by a darker color, while a *smaller value* could be indicated by a lighter color.

How to plot it?

Geospatial data can be plotted with the `{ggplot2}` package, using the **`geom_sf()`** function. Information such as *colors* and *shapes* can be depicted using the `aes()` function with the `fill`, `color` and `size` arguments.

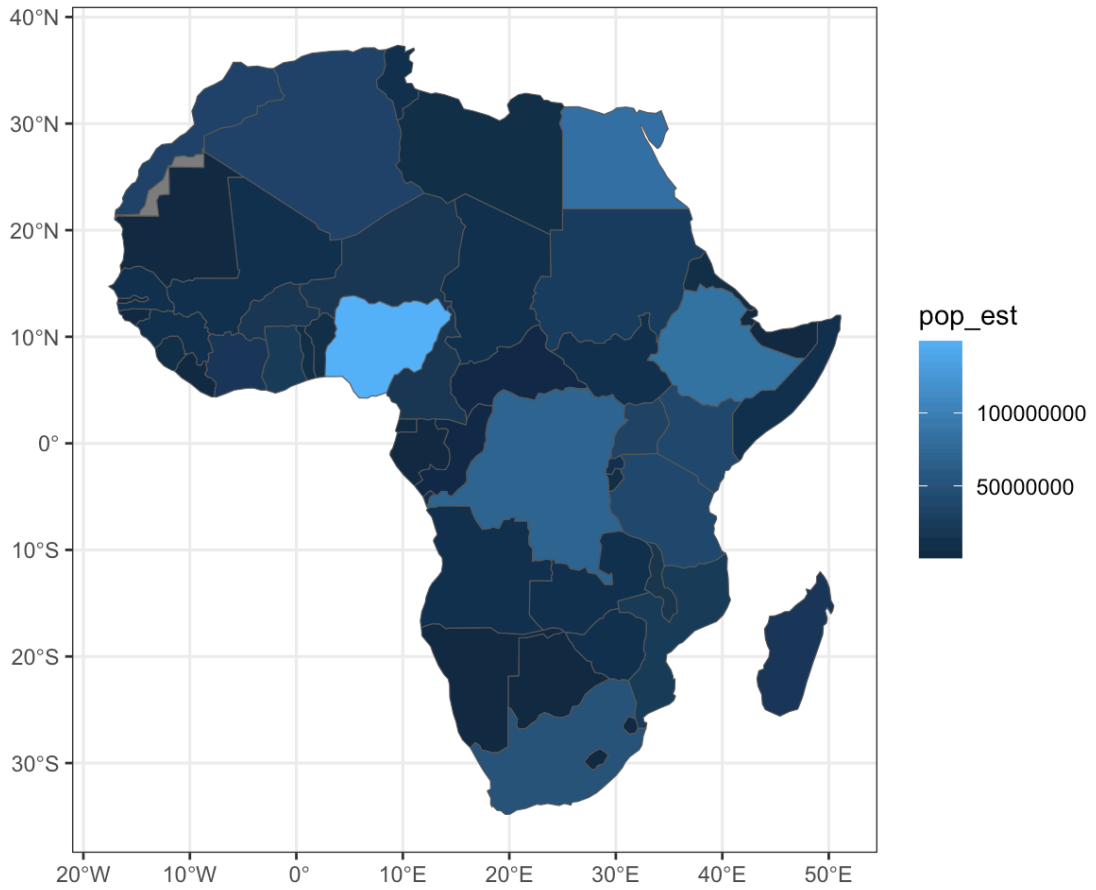
With Quantitative data

A *Choropleth map* will usually require using the `fill` argument. Let's create a Choropleth map!

We are going to use the `africountries` dataset from the `{afrilearndata}` package. It contains the administrative boundaries of all the countries in the African continent.

1. First, use `{ggplot2}` and the `geom_sf()` function to plot African countries,
2. and `fill` each of them in relation to the estimated population (`pop_est`) of each country:

```
ggplot(data = africountries) +
  geom_sf(mapping = aes(fill = pop_est))
```



SIDE NOTE

`sf` stands for **"simple features"**, an open standard used to represent a wide range of *geometric shapes*.

PRACTICE
(in RMD)

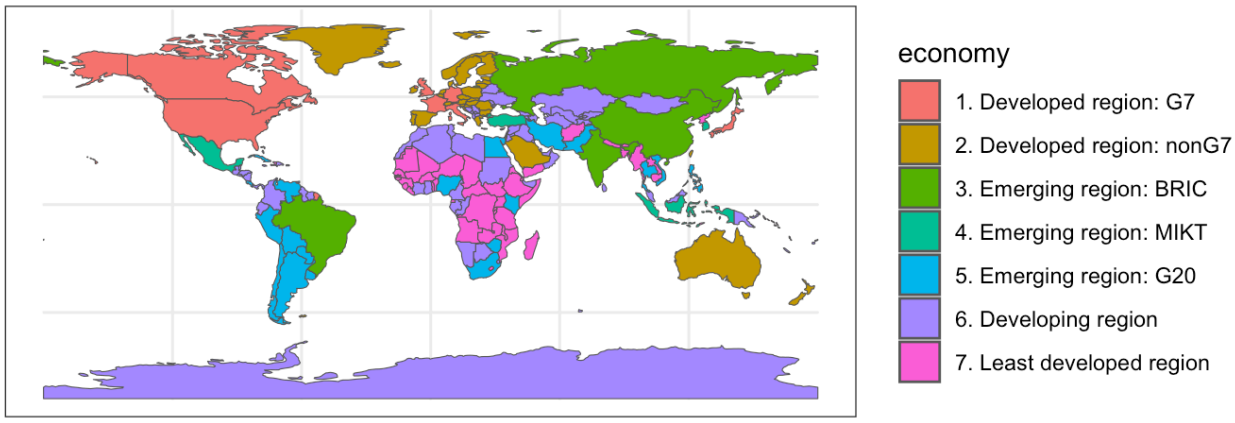
Create a Choropleth map with the `world` data from the `{spData}` package, using `geom_sf()`, to portrait its countries and `fill` them in relation to its population, available in the `pop` variable.


```
# Write and visualize your answer:
q1 <-
  ggplot(data = world) +
  ________(mapping = aes(fill = pop))
q1
```

With Categorical data

To create a map with countries *colored* by their Economical classification also use `fill`:

```
countries <- rnaturalearth::ne_countries(returnclass = "sf")

ggplot(data = countries) +
  geom_sf(aes(fill = economy))
```



Before using `geom_sf()` to create *Thematic maps*, you must verify that your *Spatial data* is an `"sf"` R object using the **`class()`** function:

```
class(africountries)
```


```
## [1] "sf"         "data.frame"
```

In the following lessons, we will learn how to get more of them and even convert foreign objects to `sf`!

How to use it?

This type of map is particularly useful when visualizing a variable and how it changes across defined regions or geopolitical areas.

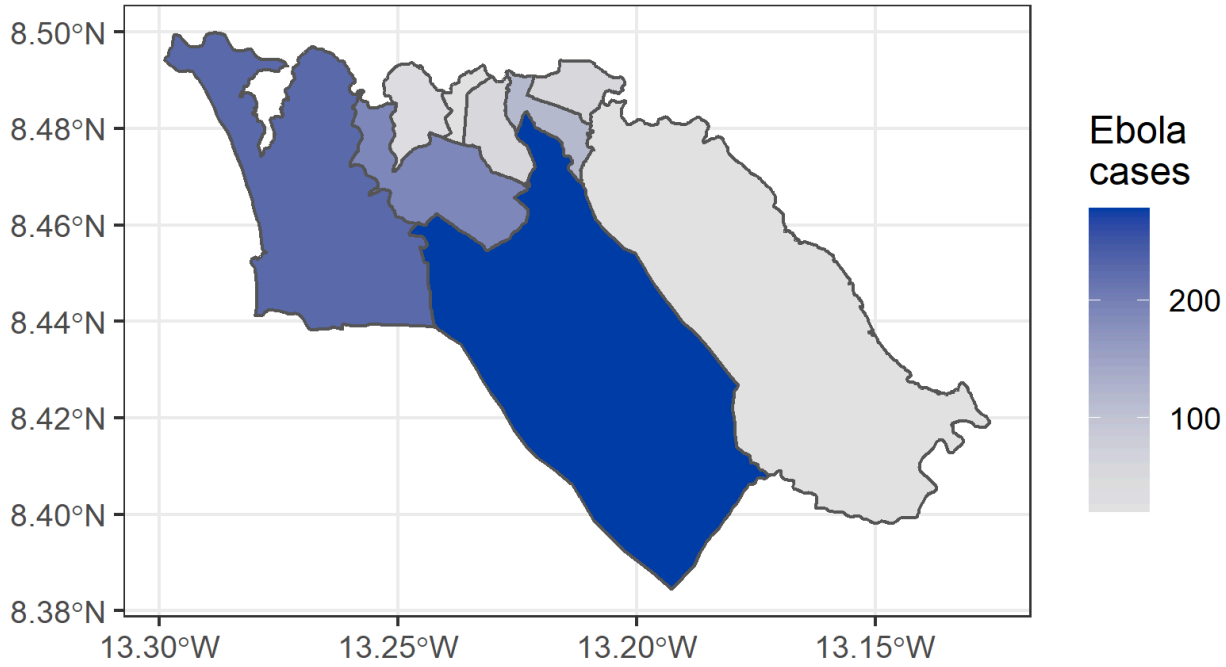


Figure 2. Choropleth map with the number of cases in a simulated Ebola epidemic in Sierra Leone.

In Figure 2, the region of interest (Sierra Leone) has been partitioned into a finite number of subregions (districts) at which the number of cases have been aggregated.

The type of shape that *Choropleth maps* visualize is called **Polygons**. This shape collects data that pertains to an enclosed region partitioned into a finite number of *areal units* with well-defined boundaries. For example, attributes collected by ZIP code, census tract, or the administrative boundary levels of a country (Figure 2).

Dot map

What is it?

A *Dot map* is a thematic map type that uses **dots** to represent attribute values in your data.

How to plot it?

The *Dot map* could use the `size` or `color` argument.

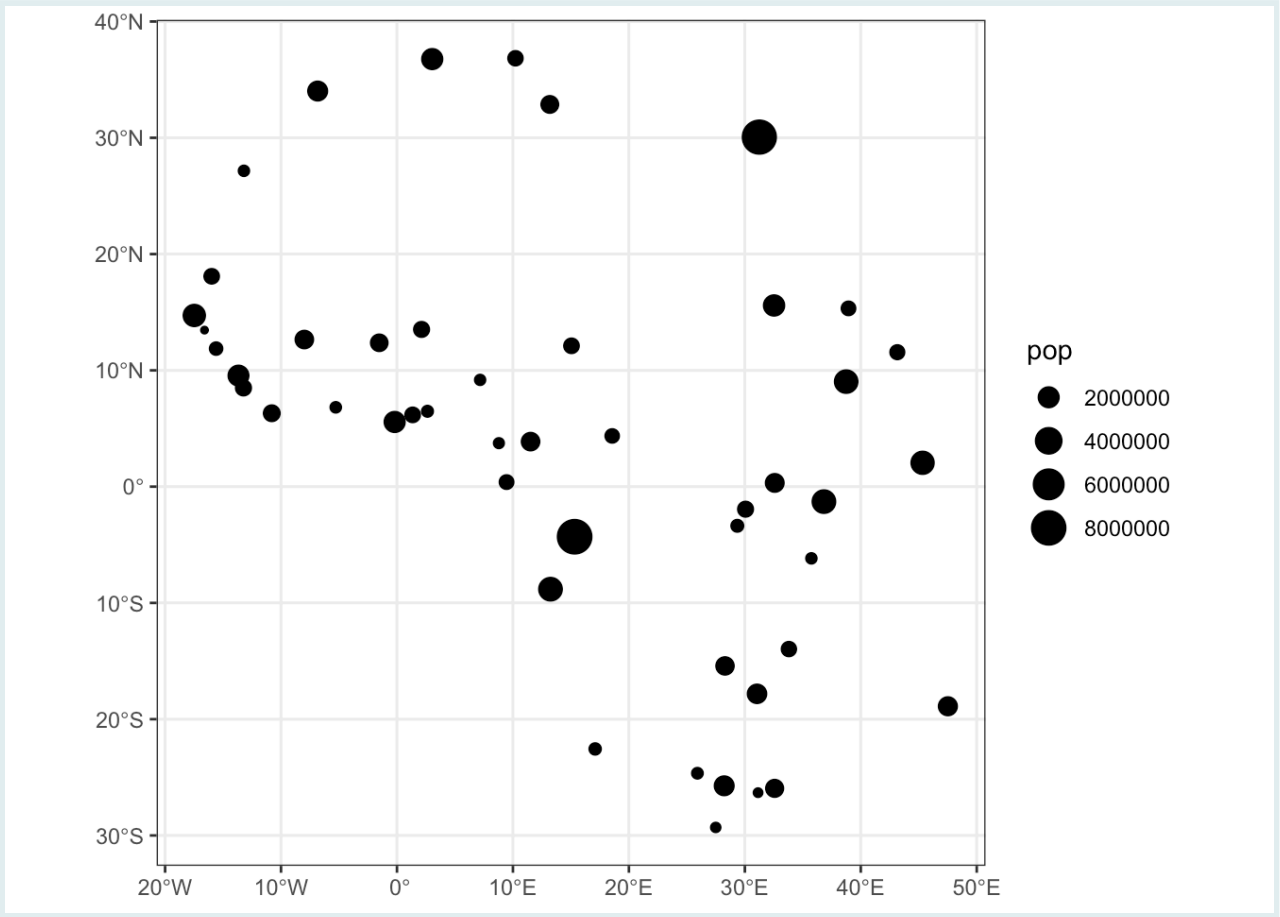
With Quantitative data

A Quantitative Dot map requires the `size` argument. Let's create a Dot map!

Let's use the `africapitals` dataset, also from the `{afrilearndata}` package, which contains the location of capital cities in the African continent.

1. First, use again `{ggplot2}` and `geom_sf()` to plot these locations,
2. and `size` each of them in relation to their number of inhabitants:

```
ggplot(data = africapitals) +
  geom_sf(mapping = aes(size = pop))
```



RSTUDIO CLOUD

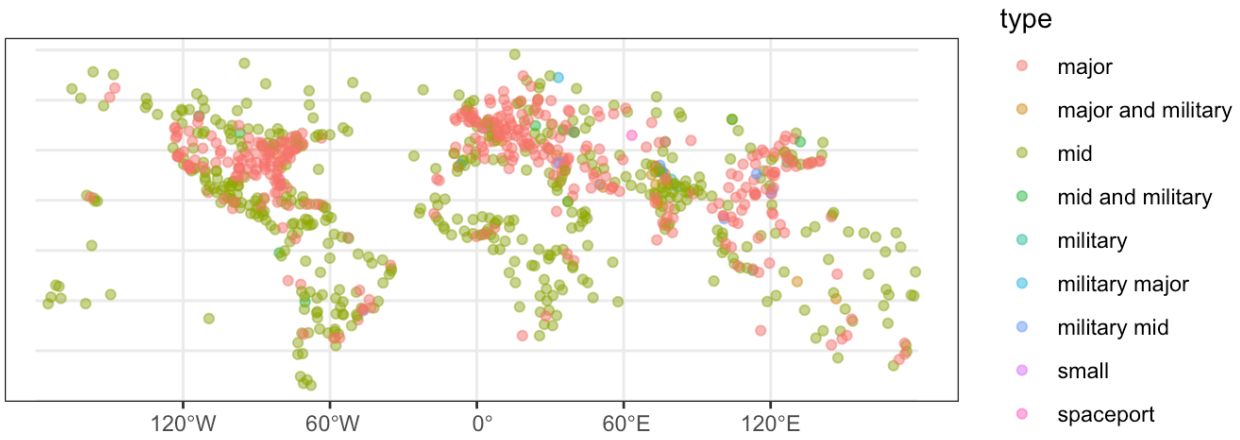
We can replicate John Snow's Dot map with the number of deaths per household from the 1854 London cholera outbreak:

```
cholera_deaths <-
  read_rds(here("data/cholera_deaths.rds"))

ggplot(data = cholera_deaths) +
  geom_sf(mapping = aes(_____ = Count), alpha = 0.7)
```

With Categorical data

To visualize airports *classified* by their type we need the `color` argument:

```
airports <- rnaturalearth::ne_download(scale = 10,
                                       type = "airports",
                                       returnclass = "sf")

ggplot(data = airports) +
  geom_sf(aes(color = type), alpha = 0.5)
```



Create a Thematic map with the `afriairports` object to portrait all its airport locations, using `geom_sf()`, and `color` them in relation to the `type` variable.

```
# Write and visualize your answer:
q2 <-
  ggplot(data = afriairports) +
  ________(mapping = aes(________ = ________))
q2
```

How to use it?

This type of map is best used to visualize the *scatter of your data* and visually *scan for clusters*.

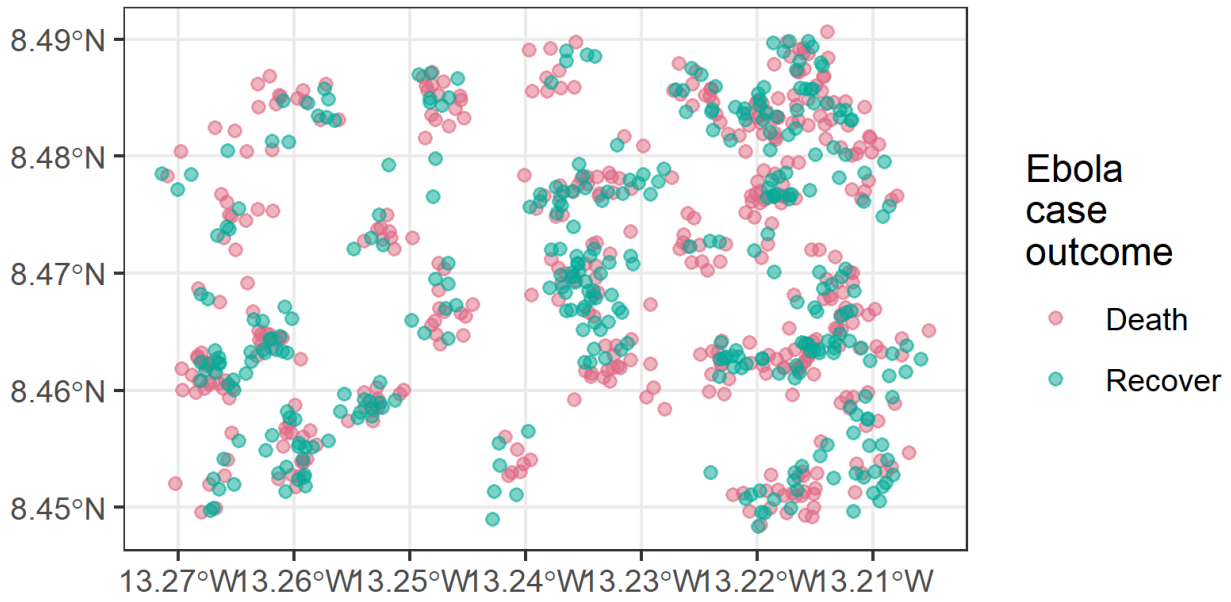


Figure 3. Dot map. Location of simulated Ebola cases in Sierra Leona, colored by each case outcome.

VOCAB

The type of shape that *Dot maps* visualize is called **Point**. This shape collects data that register the *locations* of random events. For example, collecting geographical coordinates of individuals with a given diagnosis (Figure 3): the Ebola cases in Sierra Leone.

SIDE NOTE

Are you bothered by the fact of having just dots and no country lines or any geographical context? That's good! We will see how to add those using roads and rivers, also known as **Physical features**, very soon.

RECAP

Thematic maps visualize specific *shapes* or **Geometry** types:

- Choropleth maps visualize *Polygons*.
- Dot maps visualize *Points*.

RECAP

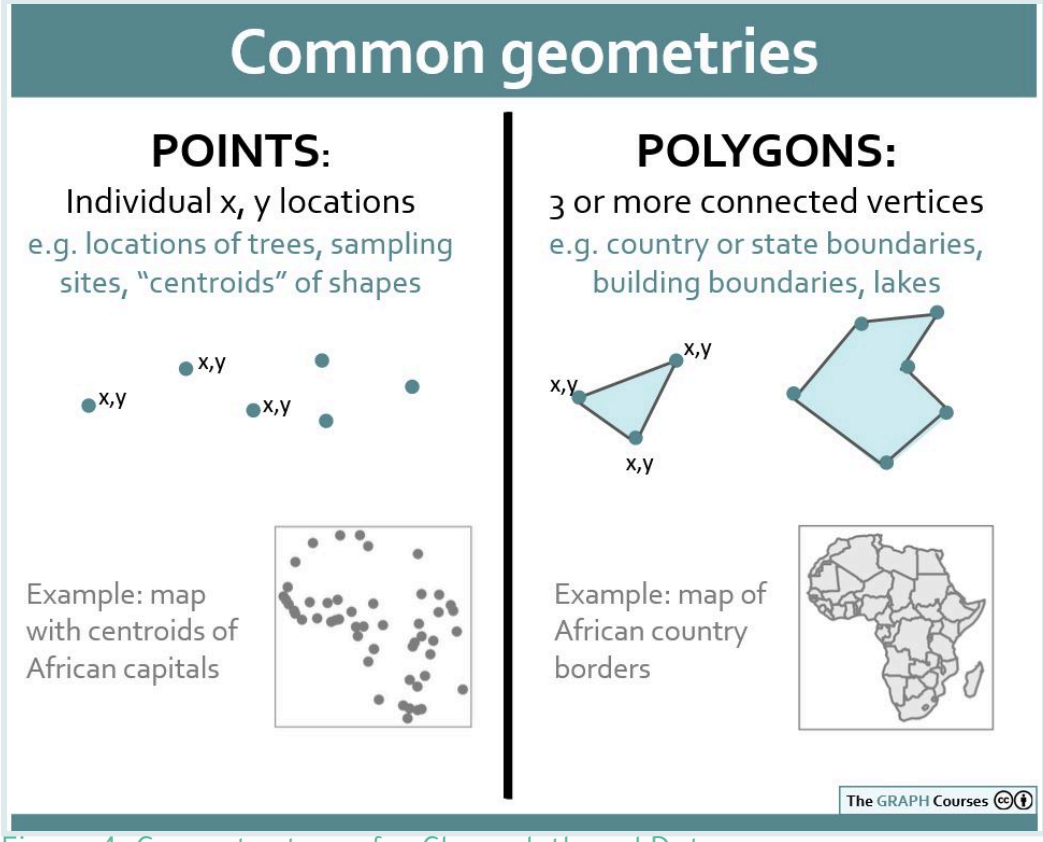


Figure 4. Geometry types for Choropleth and Dot maps.

PRACTICE (in RMD)

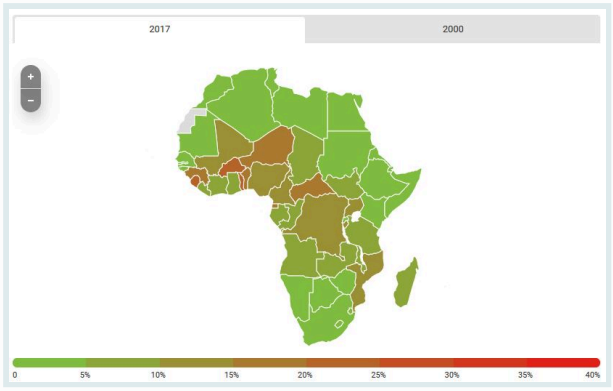
Which of the following options of *Thematic map* types:

a. `"choropleth_map"`
b. `"dot_distribution_map"`

...corresponds to each of these *Epidemic map* figures?

Your answer should be either "choropleth_map" or "dot_distribution_map".

Malaria cases in Africa:



COVID-19 cases in the world:

Wrap up

In this lesson, we learned about *Thematic maps*, how to create them using *{ggplot2}* and the `geom_sf()` function, and which type of *Geometry* they visualize.

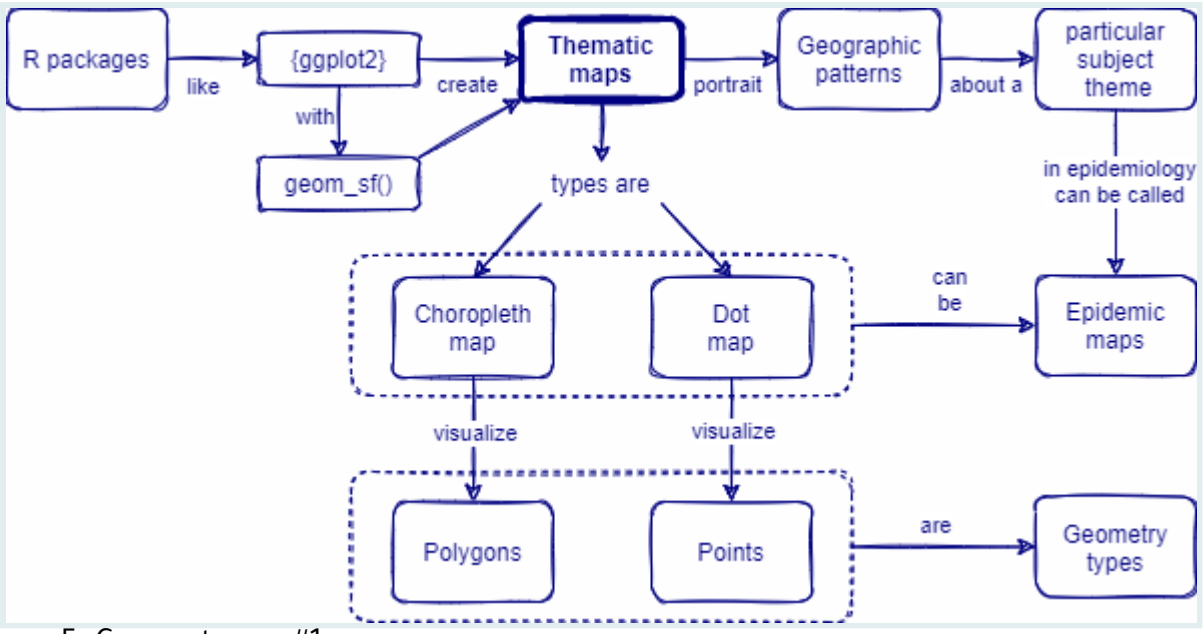


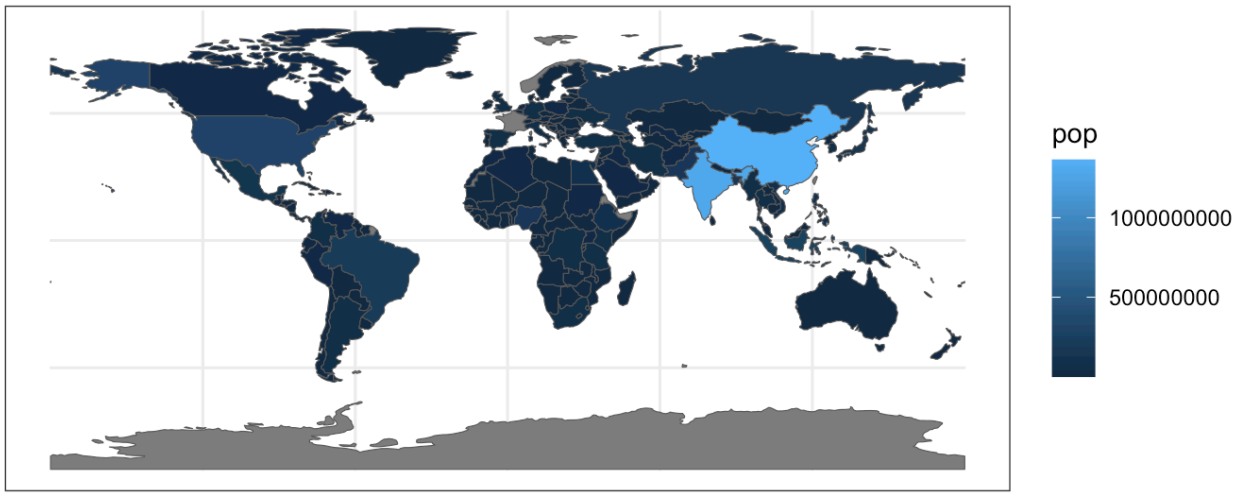
Figure 5. Concept map #1.

But, how can we complement Thematic maps with *geographic context*? Or how can we avoid *overlapped points* when needed? In the following lessons, we are going to learn about how to add *Physical features* to our maps and use *Density maps* to avoid overlaps in them!

Answer Key

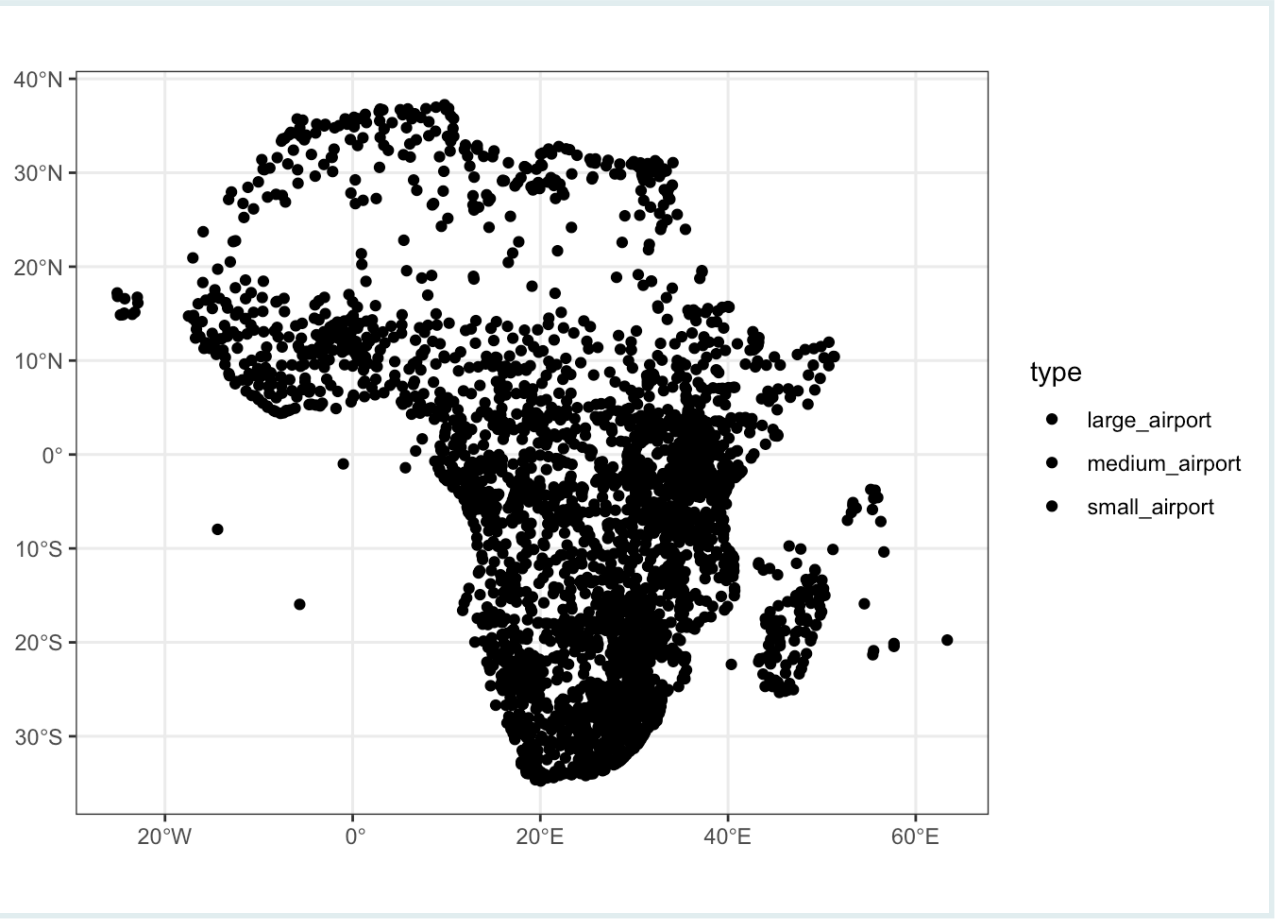
Q1

```
q1 <- ggplot(data = world) +
  geom_sf(aes(fill = pop))
q1
```



Q2

```
q2 <- ggplot(data = afriairports) +
  geom_sf(aes(fill = type))
q2
```



Q3 & Q4

Which of the following options of *Thematic map* types corresponds to each of these *Epidemic map* figures? Your answer should be either "choropleth_map" or "dot_distribution_map".

1. [**Malaria cases in Africa]** is a CHROPLETH MAP
2. **[COVID-19 cases in the world]** is a DOT DISTRIBUTION MAP

Contributors

The following team members contributed to this lesson:

ANDREE VALLE CAMPOS

R Developer and Instructor, the GRAPH Network
Motivated by reproducible science and education

LAURE VANCAUWENBERGHE

Data analyst, the GRAPH Network
A firm believer in science for good, striving to ally programming, health and education

KENE DAVID NWOSU

Data analyst, the GRAPH Network
Passionate about world improvement

References

Some material in this lesson was adapted from the following sources:

- *Batra, Neale, et al. (2021). The Epidemiologist R Handbook. Chapter 28: GIS Basics*. (2021). Retrieved 01 April 2022, from https://epirhandbook.com/en/gis-basics.html
- *Lovelace, R., Nowosad, J., & Muenchow, J. Geocomputation with R. Chapter 2: Geographic data in R*. (2019). Retrieved 01 April 2022, from https://geocompr.robinlovelace.net/spatial-class.html
- *Moraga, Paula. Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny. Chapter 2: Spatial data and R packages for mapping*. (2019). Retrieved 01 April 2022, from https://www.paulamoraga.com/book-geospatial/sec-spatialdataandCRS.html
- *Baumer, Benjamin S., Kaplan, Daniel T., and Horton, Nicholas J. Modern Data Science with R. Chapter 17: Working with geospatial data*. (2021). Retrieved 05 June 2022, from https://mdsr-book.github.io/mdsr2e/ch-spatial.html