In-class Exercise 2: Geospatial Data Wrangling

1/16/23

Setting the Scene

Water is an important resource to mankind. Clean and accessible water is critical to human health. It provides a healthy environment, a sustainable economy, reduces poverty and ensures peace and security. Yet over 40% of the global population does not have access to sufficient clean water. By 2025, 1.8 billion people will be living in countries or regions with absolute water scarcity, according to UN-Water. The lack of water poses a major threat to several sectors, including food security. Agriculture uses about 70% of the world's accessible freshwater.

Developing countries are most affected by water shortages and poor water quality. Up to 80% of illnesses in the developing world are linked to inadequate water and sanitation. Despite technological advancement, providing clean water to the rural community is still a major development issues in many countries globally, especially countries in the Africa continent.

To address the issue of providing clean and sustainable water supply to the rural community, a global Water Point Data Exchange (WPdx) project has been initiated. The main aim of this initiative is to collect water point related data from rural areas at the water point or small water scheme level and share the data via WPdx Data Repository, a cloud-based data library. What is so special of this project is that data are collected based on WPDx Data Standard.

Objectives

Geospatial analytics hold tremendous potential to address complex problems facing society. In this study, you are tasked to apply appropriate geospatial data wrangling methods to prepare the data for water point mapping study. For the purpose of this study, Nigeria will be used as the study country.

The Data

Apstial data

For the purpose of this assignment, data from WPdx Global Data Repositories will be used. There are two versions of the data. They are: WPdx-Basic and WPdx+. You are required to use WPdx+ data set.

Geospatial data

Nigeria Level-2 Administrative Boundary (also known as Local Government Area) polygon features GIS data will be used in this take-home exercise. The data can be downloaded either from The Humanitarian Data Exchange portal or geoBoundaries.

The Task

The specific tasks of this take-home exercise are as follows:

- Using appropriate sf method, import the shapefile into R and save it in a simple feature data frame format. Note that there are three Projected Coordinate Systems of Nigeria, they are: EPSG: 26391, 26392, and 26303. You can use any one of them.
- Using appropriate tidyr and dplyr methods, derive the proportion of functional and non-functional water point at LGA level.
- Combining the geospatial and aspatial data frame into simple feature data frame.
- Visualising the distribution of water point by using appropriate analytical visualisation methods.

Getting started

For the purpose of this in-class exercise, three R packages will be used. They are: sf, tidyverse and funModeling.

> **i** Your turn
>
> Using the step you had learned, check if these three R packages have been installed in you laptop, if not install the missing R packages. If Yes, launch the R packages into R environment

```
pacman::p_load(sf, tidyverse, funModeling)
```

Importing Geospatial

> **Your turn**
>
> Using the step you had learned, import the LGA boundary GIS data of Nigeria downloaded from both sources recommend above.

The geoBoundaries data set

```
geoNGA <- st_read("data/geospatial/",
                  layer = "geoBoundaries-NGA-ADM2") %>%
  st_transform(crs = 26392)
```

```
Reading layer `geoBoundaries-NGA-ADM2' from data source
  `D:\tskam\IS415-GAA\In-class_Ex\In-class_Ex02\data\geospatial'
  using driver `ESRI Shapefile'
Simple feature collection with 774 features and 5 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: 2.668534 ymin: 4.273007 xmax: 14.67882 ymax: 13.89442
Geodetic CRS:  WGS 84
```

The NGA data set

```
NGA <- st_read("data/geospatial/",
               layer = "nga_admbnda_adm2") %>%
  st_transform(crs = 26392)
```

```
Reading layer `nga_admbnda_adm2' from data source
  `D:\tskam\IS415-GAA\In-class_Ex\In-class_Ex02\data\geospatial'
  using driver `ESRI Shapefile'
Simple feature collection with 774 features and 16 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: 2.668534 ymin: 4.273007 xmax: 14.67882 ymax: 13.89442
Geodetic CRS:  WGS 84
```

Importing Aspatial data

> **Your turn**
>
> Using the steps you had learned, import the downloaded water point data set into R.

```
wp_nga <- read_csv("data/aspatial/WPdx.csv") %>%
  filter(`#clean_country_name` == "Nigeria")
```

Write the extracted data into rds format

Converting Aspatial Data into Geospatial

> **Your turn**
>
> Using the steps you had learned, convert the newly extracted wp_NGA into point sf data frame

```
wp_nga$Geometry = st_as_sfc(wp_nga$`New Georeferenced Column`)
wp_nga
```

```
# A tibble: 95,008 x 71
   row_id `#source`      #lat_~1 #lon_~2 #repo~3 #stat~4 #wate~5 #wate~6 #wate~7
    <dbl> <chr>            <dbl>   <dbl> <chr>   <chr>   <chr>   <chr>   <chr>
 1 429068 GRID3             7.98    5.12 08/29/~ Unknown <NA>    <NA>    Tapsta~
 2 222071 Federal Minis~    6.96    3.60 08/16/~ Yes     Boreho~ Well    Mechan~
 3 160612 WaterAid          6.49    7.93 12/04/~ Yes     Boreho~ Well    Hand P~
 4 160669 WaterAid          6.73    7.65 12/04/~ Yes     Boreho~ Well    <NA>
 5 160642 WaterAid          6.78    7.66 12/04/~ Yes     Boreho~ Well    Hand P~
 6 160628 WaterAid          6.96    7.78 12/04/~ Yes     Boreho~ Well    Hand P~
 7 160632 WaterAid          7.02    7.84 12/04/~ Yes     Boreho~ Well    Hand P~
 8 642747 Living Water ~    7.33    8.98 10/03/~ Yes     Boreho~ Well    Mechan~
 9 642456 Living Water ~    7.17    9.11 10/03/~ Yes     Boreho~ Well    Hand P~
10 641347 Living Water ~    7.20    9.22 03/28/~ Yes     Boreho~ Well    Hand P~
# ... with 94,998 more rows, 62 more variables: `#water_tech_category` <chr>,
#   `#facility_type` <chr>, `#clean_country_name` <chr>, `#clean_adm1` <chr>,
#   `#clean_adm2` <chr>, `#clean_adm3` <chr>, `#clean_adm4` <chr>,
#   `#install_year` <dbl>, `#installer` <chr>, `#rehab_year` <lgl>,
#   `#rehabilitator` <lgl>, `#management_clean` <chr>, `#status_clean` <chr>,
```

```
#   `#pay` <chr>, `#fecal_coliform_presence` <chr>,
#   `#fecal_coliform_value` <dbl>, `#subjective_quality` <chr>, ...
```

```r
wp_sf <- st_sf(wp_nga, crs=4326)
wp_sf
```

```
Simple feature collection with 95008 features and 70 fields
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: 2.707441 ymin: 4.301812 xmax: 14.21828 ymax: 13.86568
Geodetic CRS:  WGS 84
# A tibble: 95,008 x 71
   row_id `#source`      #lat_~1 #lon_~2 #repo~3 #stat~4 #wate~5 #wate~6 #wate~7
 *  <dbl> <chr>            <dbl>   <dbl> <chr>   <chr>   <chr>   <chr>   <chr>
 1 429068 GRID3             7.98    5.12 08/29/~ Unknown <NA>    <NA>    Tapsta~
 2 222071 Federal Minis~    6.96    3.60 08/16/~ Yes     Boreho~ Well    Mechan~
 3 160612 WaterAid          6.49    7.93 12/04/~ Yes     Boreho~ Well    Hand P~
 4 160669 WaterAid          6.73    7.65 12/04/~ Yes     Boreho~ Well    <NA>
 5 160642 WaterAid          6.78    7.66 12/04/~ Yes     Boreho~ Well    Hand P~
 6 160628 WaterAid          6.96    7.78 12/04/~ Yes     Boreho~ Well    Hand P~
 7 160632 WaterAid          7.02    7.84 12/04/~ Yes     Boreho~ Well    Hand P~
 8 642747 Living Water ~    7.33    8.98 10/03/~ Yes     Boreho~ Well    Mechan~
 9 642456 Living Water ~    7.17    9.11 10/03/~ Yes     Boreho~ Well    Hand P~
10 641347 Living Water ~    7.20    9.22 03/28/~ Yes     Boreho~ Well    Hand P~
# ... with 94,998 more rows, 62 more variables: `#water_tech_category` <chr>,
#   `#facility_type` <chr>, `#clean_country_name` <chr>, `#clean_adm1` <chr>,
#   `#clean_adm2` <chr>, `#clean_adm3` <chr>, `#clean_adm4` <chr>,
#   `#install_year` <dbl>, `#installer` <chr>, `#rehab_year` <lgl>,
#   `#rehabilitator` <lgl>, `#management_clean` <chr>, `#status_clean` <chr>,
#   `#pay` <chr>, `#fecal_coliform_presence` <chr>,
#   `#fecal_coliform_value` <dbl>, `#subjective_quality` <chr>, ...
```

Projection transformation

> **Your turn**
>
> Using the steps you had learned, transform the projection from wgs84 to appropriate projected coordinate system of Nigeria.

```r
wp_sf <- wp_sf %>%
  st_transform(crs = 26392)
```