Buffer Zone Simulation with a Customized Function in R Studio

Chen Zhuge B36830

This function is to calculate some specific value in a certain buffer (similar as calculating how many people are living in the 2km main_road buffer zone).

1. Creating buffer_function

The code is also in attached .R file, in which it would be nicer to read.

```
buffer_function <- function(dfone, dftwo, threshold) {          #dfone and dftwo are the Data Frames. In this order, we will make a buffer zone around dftwo.
dfone$label <- "OK"          # label column is temporary, marking the rows which are in buffer zone.
for(i in 1:nrow(dfone))
{
    for(j in 1:nrow(dftwo))
    {
    lon_down_limit <- dftwo[j,]$lon - threshold
    lon_up_limit <- dftwo[j,]$lon + threshold
    lat_down_limit <- dftwo[j,]$lat - threshold
    lat_up_limit <- dftwo[j,]$lat + threshold
        if ((dfone[i,]$lon > lon_down_limit) && (dfone[i,]$lon < lon_up_limit) && (dfone[i,]$lat > lat_down_limit) && (dfone[i,]$lat < lat_up_limit))
        {
        dfone[i,]$label <- "Detected"
        break
        }
    }
}
generated <<- subset(dfone, label == "Detected")
dfone$label <- NULL
result <- "BufferComputationIsCompleted"
return(result)
}
```

As for the distance of buffer zone, I temporarily gave up using **km** or **m** as unit. Insteadly, I used the "raw" unit, the **lat** and **lon,** which are same as the units in CSV file. There is a formual in internet trasnferring **lat** and **lon** to **m**, but it's a bit complicated and I am not good at mathematics. For later usage, I just copy the formula below.

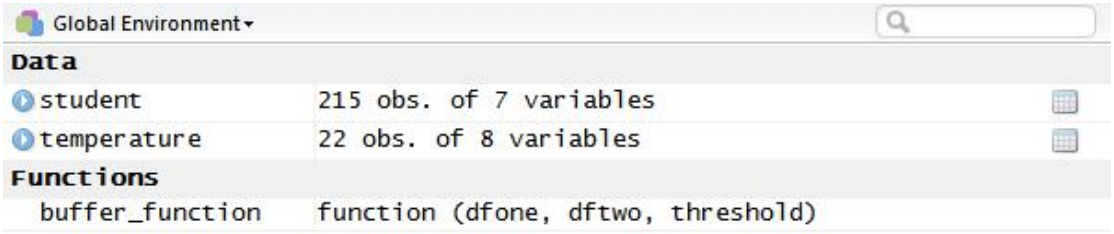
```
double s = 2 * Math.Asin(Math.Sqrt(Math.Pow(Math.Sin(a/2),2) +
Math.Cos(radLat1)*Math.Cos(radLat2)*Math.Pow(Math.Sin(b/2),2)));
```

2. **Testing the function**

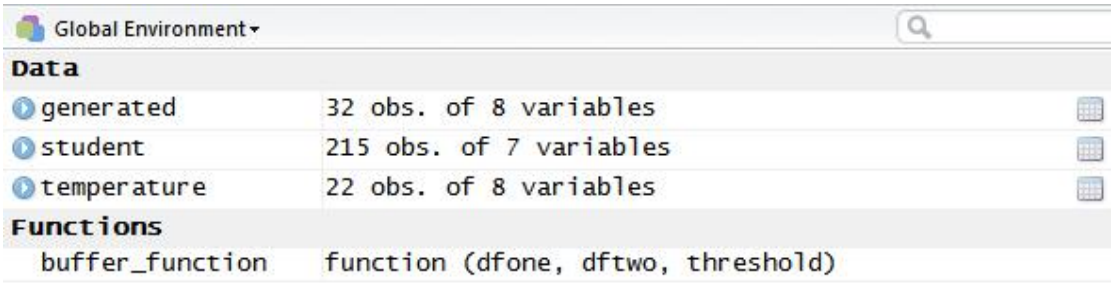
Two CSV files were picked, which are **study_migration.csv** and **estonia_air_temperature_2.csv**. The corresponding task could be described as "Find out how many students' homes locate around temperature stations within a certain distance". The threshold was set to 0.1, 0.2, 0.3, 0.4, 0.8 for test. Theoretically, when buffer zone is enlarged, the number of eligible homes would be increased.

Here is the operation recording:

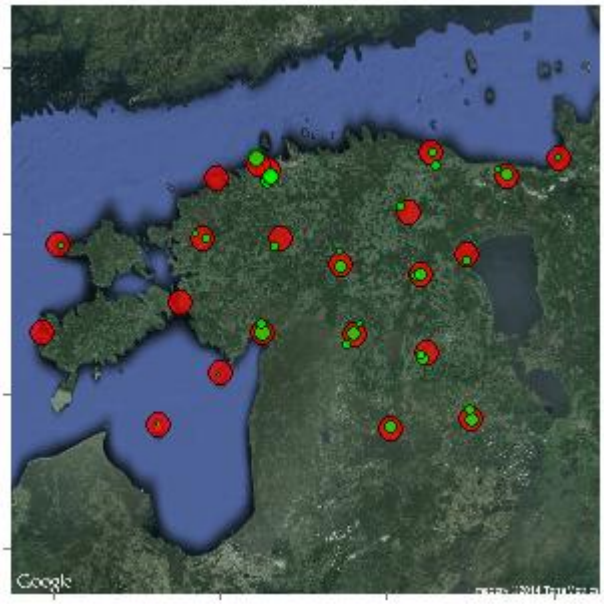
```
setwd("E:/1/currentcourse/GIS/RRelated/workplace4")

student <- read.csv("study_migration.csv", sep=";")

temperature <- read.csv("estonia_air_temperature_2.csv",
sep=",")

buffer_function <- function(dfone, dftwo, threshold) { ..}
```




```
buffer_function(student,temperature,0.1)
```



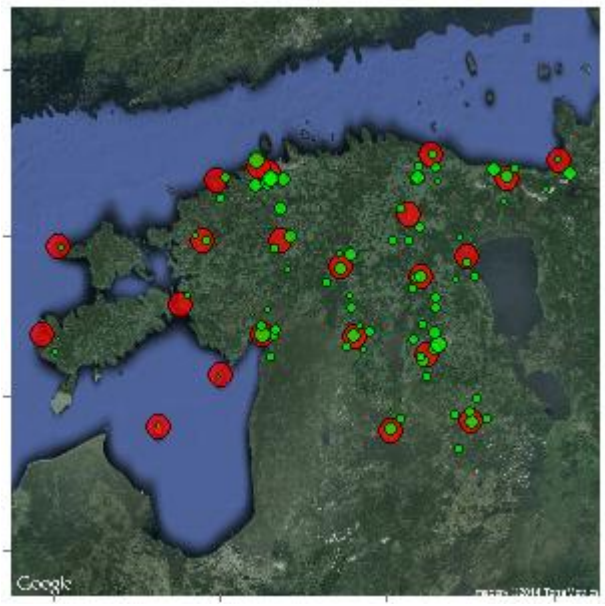

```
est_map <- ggmap(get_map(location = "estonia", maptype = "
satellite", zoom = 7))

est_map + geom_point(aes(x = lon, y = lat, size = 10), fill
 = "red", shape = 21, alpha = 0.7, data = temperature) + ge
om_point(aes(x = lon, y = lat, size = log(Sum)), fill = "gr
een", shape = 21, alpha = 0.7, data = generated) +guides(si
ze = FALSE)
```




```
buffer_function(student,temperature,0.2)
```

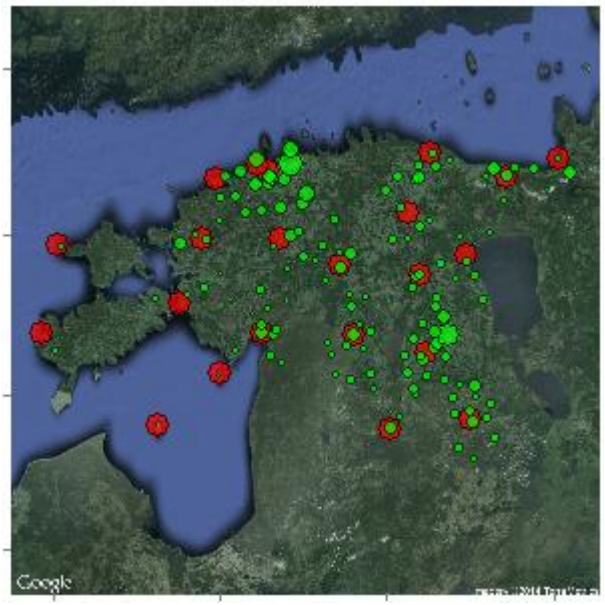
Data	
generated	85 obs. of 8 variables
student	215 obs. of 7 variables

```
est_map + geom_point(... data = temperature) + geom_point
(... data = generated) +guides(size = FALSE)
```



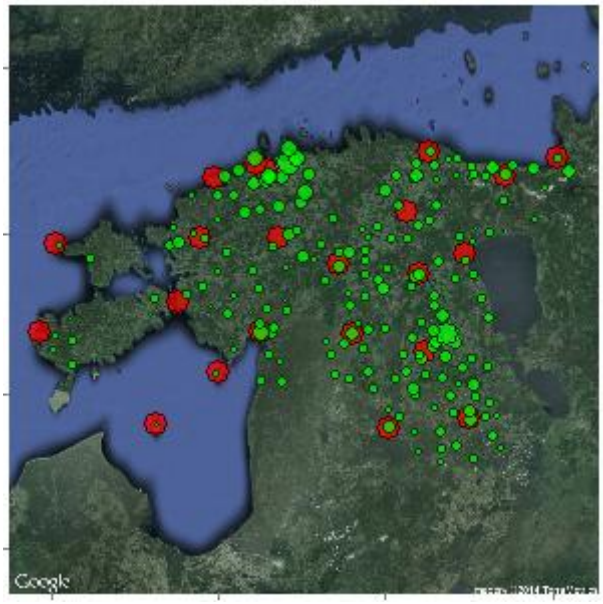

```
buffer_function(student,temperature,0.3)
```

Data	
generated	140 obs. of 8 variables
student	215 obs. of 7 variables



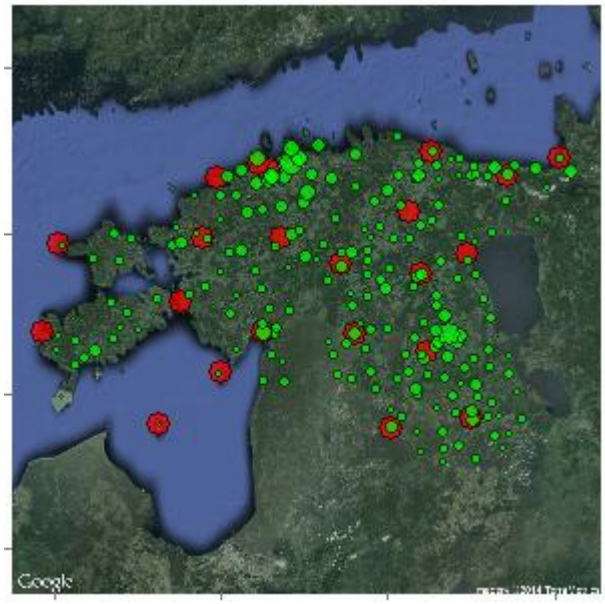
buffer_function(student,temperature,0.4)

Data	
generated	187 obs. of 8 variables
student	215 obs. of 7 variables



buffer_function(student,temperature,0.8)

Data	
generated	215 obs. of 8 variables
student	215 obs. of 7 variables



Red points are temperature stations, and green points are student's homes. As the parameter is increased, the number of green points become greater.

Through this generated Data Frame, we could calculate the numbers of homes.

3. Problems

Issues occurred when I tried to simulate the problem "How many people are living in the 2 km buffer zone of main roads" in R with this approach. That is because I can not find the lat&lon of main roads and the grids of population, even after transferring SHP file to CSV file by using the fortify function. I believe if I got the lat&lon of each population grid, I can calculate how many people lives around temperature stations, and then other places.