

HW3

Duc Le

9/20/2020

Loading the data frame created from task 5.

```
library(tibble)
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

load("map.coords.RData")
ccA = read.csv("ccA.csv")
ccB = read.csv("ccB.csv")
gdp = read.csv("gdp.csv")
df = merge(ccA, ccB, all = TRUE)
df2 = cbind(df, gdp)
```

Problem 1

Inner merge

```
inner = merge(df2, coords, all = FALSE)
dim(inner)
```

```
## [1] 1535    8
```

Left merge

```
left = merge(df2, coords, all.x = TRUE)
dim(left)
```

```
## [1] 1677    8
```

Right merge

```
right = merge(df2, coords, all.y = TRUE)
dim(right)
```

```
## [1] 1657    8
```

There's a clear difference in the # of rows after each merge. This is obviously caused by the number of unmatched rows merged between the data frames, depending on each method.

Problem 2

2a)

```
data = inner %>%  
  select(gdp, lat)
```

```
quantiles = quantile(data$gdp)  
quantiles
```

```
##           0%           25%           50%           75%           100%  
##    241.1659    1191.0260    3614.1013    9341.5210 113523.1329
```

2b)

```
gdp.q = findInterval(data$gdp, quantiles, rightmost.closed = TRUE)  
table(gdp.q)
```

```
## gdp.q  
##    1    2    3    4  
## 384 383 384 384
```

Yes the function returns 4 quartiles & their associated medians. It also provides the amount of elements in each quartile.

2c)

```
rev_gdp.q = factor(gdp.q, levels = 4:1)  
table(rev_gdp.q)
```

```
## rev_gdp.q  
##    4    3    2    1  
## 384 384 383 384
```

2d)

```
q1 = data %>%  
  filter(gdp <= quantiles[2])  
q1.lat = mean(abs(q1$lat))  
q1.lat
```

```
## [1] 21.93251
```

```
q2 = data %>%  
  filter(gdp >= quantiles[2] & gdp <= quantiles[3])  
q2.lat = mean(abs(q2$lat))  
q2.lat
```

```
## [1] 27.31258
```

```
q3 = data %>%  
  filter(gdp >= quantiles[3] & gdp <= quantiles[4])  
q3.lat = mean(abs(q3$lat))  
q3.lat
```

```
## [1] 27.70101
```

```
q4 = data %>%  
  filter(gdp >= quantiles[4] & gdp <= quantiles[5])  
q4.lat = mean(abs(q4$lat))  
q4.lat
```

```
## [1] 23.33047
```

There isn't a big discrepancy between the means of latitude associated to each quartile.