IS607: WEEK 3 ASSIGNMENT

MUSA T. GANIYU

February 9, 2016

This is the solution to IS607 (Week 3) assignment. Thanks for your time.

```r
#Before we proceed, lets install and load the packages below for easy accesibility to SQL.

#install.packages("RMySQL");
#install.packages("DBI");

library(DBI);
library("RMySQL");

# NOTE: my username = "root", password = "oracle", database_name = "diseases", and host = "localhost".

disease = dbConnect(MySQL(), user='root', password='oracle', dbname='DISEASES', host='localhost');

# viewing the table lists in the database.

dbListTables(disease);
```

```
## [1] "population"    "tb"            "tb_population"
```

```r
# Query from MySQL through R.

disease2 = dbSendQuery(disease, "SELECT T.SEX, T.CASE1, T.CASE2
FROM TB T
LEFT JOIN TB_POPULATION TP ON TP.COUNTRY = T.COUNTRY
LEFT JOIN POPULATION P ON TP.YEARS = P.YEARS
ORDER BY T.SEX, T.CASE1;");

# Viewing the sql query
disease3 = fetch(disease2, n=-1);
View(disease3);



# Now on R and Loading our databases (tb & population) from my local directory.

library(plyr);

tb1 = read.csv("C:/Data/tb.csv", head = TRUE, sep = ",");

population = read.csv("C:/Data/population.csv", head = TRUE, sep = ",");

# Renaming the tb column names

tb=rename(tb1, c("Afghanistan"="country", "X1995"="year", "female"="sex", "X.1"="case1", "X.1.1"="case2
```

```
# A glance at the output after renaming.

head(tb);
```

```
##       country year    sex case1 case2 case3
## 1 Afghanistan 1995   male    -1    -1    -1
## 2 Afghanistan 1996 female    -1    -1    -1
## 3 Afghanistan 1996   male    -1    -1    -1
## 4 Afghanistan 1997 female     5    96     1
## 5 Afghanistan 1997   male     0    26     0
## 6 Afghanistan 1998 female    45  1142    20
```

```
head(population);
```

```
##       country year population
## 1 Afghanistan 1995   17586073
## 2      Algeria 1995   29315463
## 3       Angola 1995   12104952
## 4    Argentina 1995   34833168
## 5   Azerbaijan 1995    7770806
## 6   Bangladesh 1995  119869585
```

```
# Merging the databases together

diseases5 <- merge(tb, population, by =c("country", "year"));


# obtaining the column summation.

sum_all <- cbind(diseases5$case1+diseases5$case2+diseases5$case3);

# cases/population as asked in the assignment!

rate1 <- cbind(sum_all/diseases5$population);

# Combing resulting data with the both tb and population

diseases5["rate"] <- rate1;
head(diseases5);
```

```
##       country year    sex case1 case2 case3 population          rate
## 1 Afghanistan 1995   male    -1    -1    -1   17586073 -1.705895e-07
## 2 Afghanistan 1996 female    -1    -1    -1   18415307 -1.629080e-07
## 3 Afghanistan 1996   male    -1    -1    -1   18415307 -1.629080e-07
## 4 Afghanistan 1997 female     5    96     1   19021226  5.362430e-06
## 5 Afghanistan 1997   male     0    26     0   19021226  1.366894e-06
## 6 Afghanistan 1998 female    45  1142    20   19496836  6.190748e-05
```

```
# some statistics, Hooray! Here is the results...
```

```
display <- diseases5[, c("country","year","rate")];

summary(display$rate);
```

```
##          V1
##  Min.   :-4.518e-06
##  1st Qu.: 5.825e-05
##  Median : 1.497e-04
##  Mean   : 2.709e-04
##  3rd Qu.: 3.433e-04
##  Max.   : 4.098e-03
```

```
boxplot(display$rate);
```
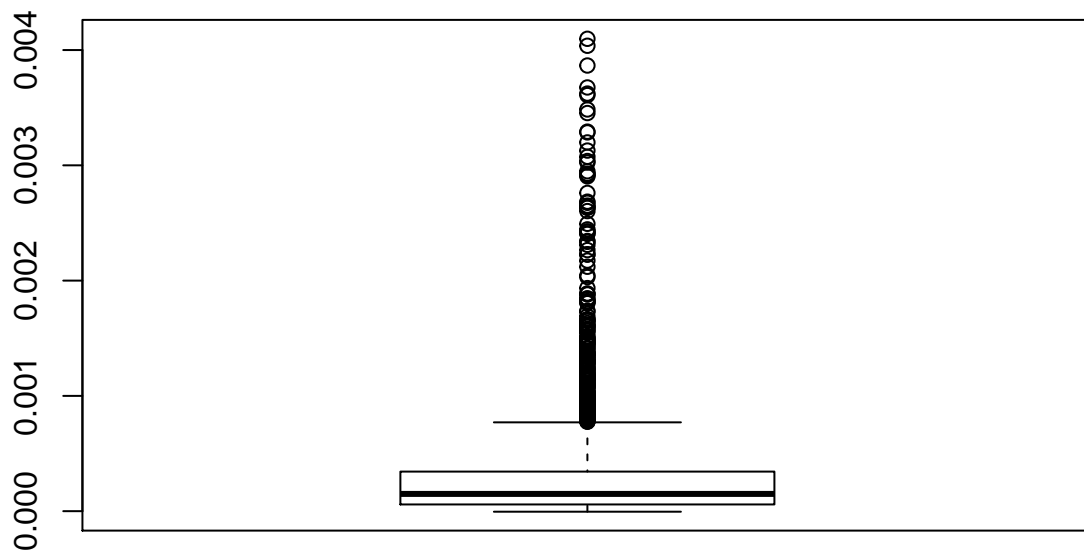


```
head(display);
```

```
##       country year          rate
## 1 Afghanistan 1995 -1.705895e-07
## 2 Afghanistan 1996 -1.629080e-07
## 3 Afghanistan 1996 -1.629080e-07
## 4 Afghanistan 1997  5.362430e-06
## 5 Afghanistan 1997  1.366894e-06
## 6 Afghanistan 1998  6.190748e-05
```