EUN SHIN DATA VISUALIZATION

Eun Kyong Shin

February 3, 2015

TASK

The Nigerian Government has committed $1.2 billion USD to achieving the United Nations Millennium Development Goals. Matching grants are given to Local Government Areas (LGAs) for approved projects and programmes geared towards reducing poverty and improving education and health. To promote the use of data in the local planning process, a rigorous, geo-referenced, baseline facility inventory was conducted for the entire country. The objective was to collect data for all of Nigeria's health, education and water facilities. The result is this online portal, the Nigeria MDG Information System (NMIS).

Using this rich dataset, we will assess a few key development statistics, previously unknowable.

- **[PART 1]** Go to this link to retrieve general metadata for Local Government Areas (LGAs) of Nigeria. LGAs are akin to counties in the U.S.Next, download the full NMIS dataset, and grab the [large] .csv file called "Health_Mopup_and_Baseline_NMIS_Facility.csv". This contains every health facility in Nigeria.
- **[PART 2]** Produce a new data.frame with the following properties:From the NMIS Health Facilities Inventory, select all facilities located in the Southern zones of Nigeria.Incorporate LGA metadata into this new dataframe.
- **[PART 3]** Calculate the total number of full time nurses and doctors for all health facilities, by state. Compute per capita and per area values, as well. Sort the resulting dataset by state population, in descending order. Show the results!

SOLUTION

For **PART 1** of the assignemnt,I downloaded the two data sets in my working directory. For **PART 2** I merged the two data files by the `unique_lga` variable. Then, to select all facilities located in the Southern zones of Nigeria, I made the subset of the data which I names as `newdata`. Also I omitted missing variables.

```
setwd("/Users/eunkyongshin/GitHub/data-viz/r")
lgas<-read.csv("lgas.csv")
nmis<- read.csv("Health_Mopup_and_Baseline_NMIS_Facility.csv")
data<-merge(nmis, lgas, by="unique_lga")
newdata<-subset(data, zone %in% c("South-South","Southeast","Southwest"))
summary(newdata$state)
```

```
##        Abia     Adamawa   Akwa Ibom     Anambra      Bauchi     Bayelsa
##         701           0         511         946           0         256
##       Benue       Borno Cross River       Delta      Ebonyi         Edo
##           0           0         949         806         597         685
##       Ekiti       Enugu         FCT       Gombe         Imo      Jigawa
##         490         811           0           0        1164           0
##      Kaduna        Kano     Katsina       Kebbi        Kogi       Kwara
##           0           0           0           0           0           0
##       Lagos    Nasarawa       Niger        Ogun        Ondo        Osun
##        1254           0           0        1151         729        1229
##         Oyo     Plateau      Rivers      Sokoto      Taraba        Yobe
```

```
##        1292           0         604           0           0           0
##     Zamfara
##           0
```

For coding efficiency, I costomized value names. Also by creating new variable `nurse` I added up the `num_nurses_fulltime` and `num_nursemidwives_fulltime` to use it as the total number of full time nurses.

```
newdata<-rename(newdata,c(num_nurses_fulltime="nurs"))
newdata<-rename(newdata,c(num_nursemidwives_fulltime="midw"))
newdata<-rename(newdata,c(num_doctors_fulltime="doctor"))
newdata<-rename(newdata,c(facility_name="facility"))
newdata<-rename(newdata,c(pop_2006="popu"))
newdata<-rename(newdata,c(area_sq_km="area"))
newdata$nurse<-newdata$nurs+newdata$midw
```

Next,in **PART 3** I created two subsets that willl merged as `statedata` at the end. The first subset, `mydata1` contains facilities only in the southern part of Nigeria, The `finaldata` contains informations of the total full time nurses and doctors for all health dacilities in the Southern Zones of Nigeria.

```
myvars<-c("facility","nurse","doctor","state","zone","area","popu")
mydata<-newdata[myvars]
mydata<-na.omit(mydata) ## Omit missing variables
head(mydata)
```

```
> head(mydata)
                                        facility nurse doctor state      zone     area   popu
1                          Asaokpuaja Health Centre     1      0  Abia Southeast 22.77506 107488
2                                  Uwalaka Hospital    10      1  Abia Southeast 22.77506 107488
3 St. Theresa's Maternity Home and Children's Care.     7      0  Abia Southeast 22.77506 107488
4                                      Todac clinic    12      3  Abia Southeast 22.77506 107488
5                                      M.C Hospital    10      1  Abia Southeast 22.77506 107488
7                                      Eziama P H C     4      0  Abia Southeast 22.77506 107488
```

To calculate total numer of full time nurses and doctors, I extracted `state`,`nurse`,`doctor` and created `statedata2`.

```
statedata1<-ddply(mydata,.(state),numcolwise(sum))
write.csv(statedata1, file="statedata1.csv")
states1<-read.csv("statedata1.csv")
newvars<-c("state","nurse","doctor")
statedata2<-states1[newvars]
head(statedata2)
```

```
> head(statedata2)
        state nurse doctor
1        Abia  1541    242
2   Akwa Ibom  1740    208
3     Anambra  3004    680
4     Bayelsa   547    125
5 Cross River   821    242
6       Delta  2737    590
```

To calculate the sum of population and area data which are given at lga level upto the state level, I sumed up unique data points in `area` and `popu` in each state. `statedata6` contains population and area data at the state level.

```
statedata3<-c("state","area","popu")
stateinfo<-mydata[statedata3]
stateinfo1<-unique(stateinfo)
statedata4<-ddply(stateinfo1,.(state),numcolwise(sum))
write.csv(statedata4, file="statedata4.csv")
statedata5<-read.csv("statedata4.csv")
newvars2<-c("state","area","popu")
statedata6<-statedata5[newvars2]
head(statedata6)
```

```
> head(statedata6)
        state      area    popu
1        Abia  4093.651 2488632
2   Akwa Ibom  6514.807 3791977
3     Anambra  4762.182 4182032
4     Bayelsa  9006.983 1626066
5 Cross River 20936.867 2856581
6       Delta 17105.597 4098391
```

Then, I merged `statedata2` and `statedata6` to generate final table.

```
findata<-merge(statedata2, statedata6, by="state") ## merge docs/nurse data with state level data
head(findata)
```

```
> head(findata)
        state nurse doctor      area    popu
1        Abia  1541    242  4093.651 2488632
2   Akwa Ibom  1740    208  6514.807 3791977
3     Anambra  3004    680  4762.182 4182032
4     Bayelsa   547    125  9006.983 1626066
5 Cross River   821    242 20936.867 2856581
6       Delta  2737    590 17105.597 4098391
```

Then to calculate numbers of full time nurses and doctors per cap and per area, I created new varaibles `nursepc` (number of full time nurses per capita), `nursepa` (number of full time nurses per area), `doctorpc` (number of full time doctors per capita), and `doctorpa` (number of full time doctors per area). As requested in the assignment, I arranged my finaldata in descending order of population.

```
findata$nursepc<-findata$nurse/findata$popu
findata$nursepa<-findata$nurse/findata$area
findata$doctorpc<-findata$doctor/findata$popu
findata$doctorpa<-findata$doctor/findata$area
findata
arrange(findata,desc(popu))
```

```
> arrange(findata,desc(popu))
         state nurse doctor      area    popu      nursepc    nursepa     doctorpc   doctorpa
1        Lagos  5746   2540  3479.717 9013534 0.0006374858 1.65128377 2.817985e-04 0.72994445
2          Oyo  3191    930 27900.302 5591589 0.0005706786 0.11437152 1.663212e-04 0.03333297
3       Rivers  2093    611 10105.266 5010351 0.0004177352 0.20711974 1.219475e-04 0.06046353
4      Anambra  3004    680  4762.182 4182032 0.0007183111 0.63080330 1.626004e-04 0.14279169
5        Delta  2737    590 17105.597 4098391 0.0006678231 0.16000611 1.439589e-04 0.03449164
6          Imo  3874    865  5049.308 3835652 0.0010099978 0.76723386 2.255158e-04 0.17131061
7    Akwa Ibom  1740    208  6514.807 3791977 0.0004588635 0.26708388 5.485265e-05 0.03192727
8         Ondo  1139    275 15031.599 3441024 0.0003310061 0.07577371 7.991807e-05 0.01829479
9         Osun  1799    451  8595.119 3423535 0.0005254802 0.20930483 1.317352e-04 0.05247164
10        Ogun  3683   1017 14153.857 3377422 0.0010904767 0.26021175 3.011172e-04 0.07185321
11       Enugu  1974    520  6766.377 3108524 0.0006350281 0.29173662 1.672820e-04 0.07685058
12         Edo  2019    496 16833.356 2965976 0.0006807203 0.11994043 1.672299e-04 0.02946531
13 Cross River   821    242 20936.867 2856581 0.0002874065 0.03921313 8.471666e-05 0.01155856
14        Abia  1541    242  4093.651 2488632 0.0006192157 0.37643661 9.724218e-05 0.05911594
15       Ekiti  1071    243  5801.502 2384212 0.0004492050 0.18460737 1.019205e-04 0.04188570
16      Ebonyi   845    215  6342.013 2173501 0.0003887737 0.13323845 9.891875e-05 0.03390091
17     Bayelsa   547    125  9006.983 1626066 0.0003363947 0.06073066 7.687265e-05 0.01387812
```

```
write.csv(findata, file="findata.csv") ## save the final data
```