EUN SHIN DATA VISUALIZATION

DATE: 2015-01-29

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

Before I run R for the assignment, I activated some packages using `library()` that might be needed ro run my codes.

```
library("ggplot2", lib.loc="/Library/Frameworks/R.framework/Versions/3.1/Resources/library")
library("plyr", lib.loc="/Library/Frameworks/R.framework/Versions/3.1/Resources/library")
library("xlsx", lib.loc="/Library/Frameworks/R.framework/Versions/3.1/Resources/library")
library("reshape2", lib.loc="/Library/Frameworks/R.framework/Versions/3.1/Resources/library")
library("markdown", lib.loc="/Library/Frameworks/R.framework/Versions/3.1/Resources/library")
```

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"))
```

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.

Then, I created tables to calculate the total numbers of nurses and doctors by states. Then, I merge statedata1 which contains the total numbers od nurses and doctors for each state and statedata4 which contains the area and population data for each state.

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

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

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

Then I will induce state level population and area value.

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

```
statedata5
newvars2<-c("state","area","popu")
statedata6<-statedata5[newvars2]
findata<-merge(statedata2, statedata6, by="state")
write.csv(findata, file="findata.csv")
finaldata.eks<-read.csv("findata.csv")

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))
```