

Yiding Lyu yl3248

January 28, 2015

R Basic Homework (Homework 1)

By Yiding Lyu, yl3248, Jan 30

Part 1 of the problem

In Part 1 of the problem, we should 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.

To solve this problem, first we should set the working directory and read the data:

```
setwd("/Users/Lichking/Desktop/Columbia/Courses/DataVisualization/Assignments/Assignment1")
LGA_Data=read.csv("lgas.csv")
Facility_Data=read.csv("Health_Mopup_and_Baseline_NMIS_Facility.csv")
```

Then, we can take a look at the data:

```
str(LGA_Data)
```

```
## 'data.frame':    774 obs. of  13 variables:
##  $ lga_id          : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ unique_lga      : Factor w/ 773 levels "abia_aba_north",...: 142 275 39 211 1 2 563 564 169 292 ...
##  $ pop_2006        : int  100180 58444 139090 151723 107488 423852 201329 250278 144802 195652 ...
##  $ lga             : Factor w/ 768 levels "Aba North","Aba South",...: 3 4 5 6 1 2 7 8 9 10 ...
##  $ state           : Factor w/ 37 levels "Abia","Adamawa",...: 8 15 3 11 1 1 28 28 9 17 ...
##  $ zone            : Factor w/ 6 levels "North-Central",...: 2 1 4 5 5 5 6 6 4 5 ...
##  $ area_sq_km      : num  3967.1 992.4 190 584.1 22.8 ...
##  $ surveying_effort : Factor w/ 4 levels "113","148","Others",...: 3 1 3 3 3 3 3 3 3 ...
##  $ pop_density_2006 : num  25.3 58.9 732.1 259.8 4719.5 ...
##  $ longitude        : num  13.33 6.85 7.77 8.22 7.37 ...
##  $ latitude         : num  13.43 8.88 5.02 6.26 5.11 ...
##  $ senatorial_district: Factor w/ 67 levels "", "Abia Central",...: 40 1 33 21 4 4 42 42 12 53 ...
##  $ TA_names         : Factor w/ 121 levels "Abdul Samad Isah Haruna",...: 75 57 62 83 19 114 108 109 ...
```

```
str(Facility_Data)
```

```
## 'data.frame':    34139 obs. of  30 variables:
##  $ facility_name    : Factor w/ 28550 levels "\nModel pry healthcare",...: 10496 1224 ...
##  $ facility_type_display : Factor w/ 18 levels "Basic Health Centre / Primary Health Clin...: ...
##  $ maternal_health_delivery_services: logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
##  $ emergency_transport : logi  TRUE TRUE TRUE FALSE FALSE TRUE ...
##  $ skilled_birth_attendant : logi  FALSE TRUE TRUE FALSE FALSE TRUE ...
##  $ num_chews_fulltime   : int   1 4 2 2 0 0 0 1 4 1 ...
##  $ phcn_electricity    : logi  TRUE TRUE TRUE FALSE FALSE FALSE ...
```

```
## $ c_section_yn : logi FALSE FALSE TRUE FALSE FALSE TRUE ...
## $ child_health_measles_immun_calc : logi TRUE TRUE TRUE TRUE TRUE FALSE ...
## $ num_nurses_fulltime : int 0 1 4 0 0 2 1 6 NA 0 ...
## $ num_nursemidwives_fulltime : int 0 1 11 0 0 0 0 12 4 0 ...
## $ num_doctors_fulltime : int 0 1 2 0 0 1 0 5 15 0 ...
## $ date_of_survey : Factor w/ 323 levels "2009-07-23","2010-10-20",...: 240 273 249
## $ facility_id : Factor w/ 34122 levels "AAABW","AAADE",...: 28724 10619 10978 1
## $ community : Factor w/ 25739 levels "\n KATSINAWA ",...: 12427 15598 23
## $ ward : Factor w/ 13988 levels "\n2"," 2"," Mjakin",...: 12580 6919
## $ management : Factor w/ 4 levels "dk","faith_based",...: 4 3 4 4 4 3 4 3 3
## $ improved_water_supply : logi FALSE FALSE TRUE TRUE FALSE TRUE ...
## $ improved_sanitation : logi FALSE TRUE TRUE FALSE FALSE TRUE ...
## $ vaccines_fridge_freezer : logi FALSE TRUE TRUE FALSE FALSE FALSE ...
## $ antenatal_care_yn : logi TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ family_planning_yn : logi FALSE FALSE TRUE TRUE TRUE TRUE ...
## $ malaria_treatment_artemisinin : logi TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ sector : Factor w/ 1 level "health": 1 1 1 1 1 1 1 1 1 1 ...
## $ formhub_photo_id : Factor w/ 33850 levels "1230768326188.jpg",...: 22087 21911 229
## $ gps : Factor w/ 33850 levels "10.00030365 11.89968052 231.0 5.0",...:
## $ survey_id : Factor w/ 34122 levels "00006404-aeb9-4010-96b8-db287ee3fad1",...:
## $ unique_lga : Factor w/ 772 levels "abia_aba_north",...: 180 527 690 172 172
## $ latitude : num 6.54 9.01 5.13 5.46 5.5 ...
## $ longitude : num 9.08 7.68 7.16 8.15 8.03 ...
```

We can see that there are one same column in the two data.frame, which is called “unique_lga”, which seems to be the name of the city. If we want to get all the facility in the southern part, we can first get the cities in the southern part from the LGA_Data and then get the facility in southern part according to the city list we get from Facility_Data:

```
LGA_South=subset(LGA_Data,zone %in% c("South-South","Southeast","Southwest"))
SouthCity=LGA_South$unique_lga
Facility_South=subset(Facility_Data,unique_lga %in% SouthCity)
```

Now that we have the data.frame Facility_South, we can construct the new data.frame by the merge operation:

```
Merge_Data=merge(LGA_South,Facility_South,by="unique_lga")
```

Part 2 of the Problem

In this part, we should 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

If we do this part on all the states First we can extract all the state from LGA_Data:

```
States=levels(LGA_Data$state)
States
```

```
## [1] "Abia"      "Adamawa"    "Akwa Ibom"  "Anambra"    "Bauchi"
## [6] "Bayelsa"    "Benue"      "Borno"      "Cross River" "Delta"
## [11] "Ebonyi"     "Edo"        "Ekiti"      "Enugu"      "FCT"
## [16] "Gombe"      "Imo"        "Jigawa"     "Kaduna"     "Kano"
## [21] "Katsina"    "Kebbi"      "Kogi"       "Kwara"      "Lagos"
## [26] "Nasarawa"   "Niger"      "Ogun"       "Ondo"       "Osun"
## [31] "Oyo"        "Plateau"    "Rivers"     "Sokoto"     "Taraba"
## [36] "Yobe"       "Zamfara"
```

Because we should do the work for every state, so it is better to do it in a loop. Before we do the loop, we may first construct the data.frame for the result:

```
FacilityStatistics=data.frame(state=States,num_of_fulltime_nurses_and_doctors=rep(NA,length(States)),pop
```

Which contains 6 columns:

```
## 'data.frame': 37 obs. of 6 variables:
## $ state : Factor w/ 37 levels "Abia","Adamawa",...: 1 2 3 4 5 6 7 8 9 10
## $ num_of_fulltime_nurses_and_doctors: logi NA NA NA NA NA NA NA ...
## $ population : logi NA NA NA NA NA NA NA ...
## $ area : logi NA NA NA NA NA NA NA ...
## $ facility_per_capita : logi NA NA NA NA NA NA NA ...
## $ facility_per_area : logi NA NA NA NA NA NA NA ...
```

Then we do the loop. In this part, we consider “all fulltime nurses and doctors” include fulltime nurses, fulltime nurse midwives and fulltime doctors:

```
for (n in c(1:length(States))) {
  StateData=LGA_Data[LGA_Data$State==States[n],]
  Population=sum(StateData$pop_2006)
  Area=sum(StateData$area_sq_km)
  FacilityStatistics[n,3]=Population
  FacilityStatistics[n,4]=Area
  StateCity=StateData$unique_lga
  StateFacility=subset(Facility_Data,unique_lga %in% StateCity)
  nurse1=na.omit(StateFacility$num_nurses_fulltime)
  nurse2=na.omit(StateFacility$num_nursemidwives_fulltime)
  doctor=na.omit(StateFacility$num_doctors_fulltime)
  FulltimeTotal=sum(nurse1)+sum(nurse2)+sum(doctor)
  FacilityStatistics[n,5]=FulltimeTotal/Population
  FacilityStatistics[n,6]=FulltimeTotal/Area
  FacilityStatistics[n,2]=FulltimeTotal
}
```

Now we have an unsorted result:

```
head(FacilityStatistics)
```

```
##      state num_of_fulltime_nurses_and_doctors population      area
## 1     Abia                                3562    2833999  4857.434
## 2   Adamawa                               1208    3168101 37906.415
## 3 Akwa Ibom                               2146    3920208  6720.579
```

```
## 4    Anambra                3807    4182032  4762.182
## 5     Bauchi                1685    4676438 48168.152
## 6     Bayelsa                709    1703358  9368.015
##   facility_per_capita facility_per_area
## 1      0.0012568812      0.73330893
## 2      0.0003813010      0.03186796
## 3      0.0005474199      0.31931772
## 4      0.0009103230      0.79942349
## 5      0.0003603170      0.03498162
## 6      0.0004162366      0.07568305
```

Then we may sort it. Here we use the `arrange()` function to sort it, which need to load plyr library:

```
library(plyr)
FacilityStatistics_Sorted=arrange(FacilityStatistics,desc(population))
FacilityStatistics_Sorted
```

```
##      state num_of_fulltime_nurses_and_doctors population      area
## 1      Kano                1332    9546529 20790.265
## 2      Lagos                9028    9013534  3479.717
## 3      Kaduna               2073    5913562 44220.770
## 4       Oyo                4417    5591589 27900.302
## 5     Katsina               916    5506778 23823.893
## 6     Rivers               2908    5185400 10363.981
## 7     Bauchi               1685    4676438 48168.152
## 8     Jigawa                497    4348649 23404.072
## 9     Borno                1777    4251193 72655.785
## 10    Benue                3199    4219244 30748.258
## 11    Anambra              3807    4182032  4762.182
## 12    Delta               3538    4098391 17105.597
## 13    Niger               1756    3950229 72116.147
## 14     Imo                4904    3934899  5136.055
## 15   Akwa Ibom             2146    3920208  6720.579
## 16     Ogun                4947    3728098 16873.168
## 17    Sokoto                752    3696999 32170.356
## 18     Ondo               1673    3441024 15031.599
## 19     Osun               2391    3423535  8595.119
## 20     Kogi               1858    3278487 29074.360
## 21    Zamfara               480    3259846 33684.188
## 22     Enugu              2621    3257298  7561.111
## 23     Kebbi               649    3227627 36357.052
## 24     Edo               3336    3218332 19596.676
## 25    Plateau             1550    3178712 26502.703
## 26    Adamawa             1208    3168101 37906.415
## 27 Cross River           1725    2856581 20936.867
## 28     Abia               3562    2833999  4857.434
## 29     Ekiti              2123    2384212  5801.502
## 30     Kwara              2073    2371089 33828.152
## 31     Gombe              1030    2353879 17411.735
## 32     Yobe                284    2321591 44833.942
## 33     Taraba              998    2280483 59130.787
## 34    Nasarawa            1005    2270109 26664.338
## 35     Ebonyi             1122    2173501  6342.013
```

## 36	Bayelsa	709	1703358	9368.015
## 37	FCT	905	1405201	7570.548
##	facility_per_capita	facility_per_area		
## 1	0.0001395272	0.064068448		
## 2	0.0010016049	2.594463957		
## 3	0.0003505501	0.046878424		
## 4	0.0007899365	0.158313700		
## 5	0.0001663405	0.038448795		
## 6	0.0005608053	0.280587147		
## 7	0.0003603170	0.034981620		
## 8	0.0001142884	0.021235621		
## 9	0.0004180003	0.024457791		
## 10	0.0007581927	0.104038414		
## 11	0.0009103230	0.799423488		
## 12	0.0008632656	0.206832891		
## 13	0.0004445312	0.024349609		
## 14	0.0012462836	0.954818360		
## 15	0.0005474199	0.319317717		
## 16	0.0013269501	0.293187388		
## 17	0.0002034082	0.023375557		
## 18	0.0004861925	0.111298872		
## 19	0.0006984009	0.278181127		
## 20	0.0005667248	0.063905104		
## 21	0.0001472462	0.014250010		
## 22	0.0008046547	0.346642191		
## 23	0.0002010765	0.017850732		
## 24	0.0010365618	0.170232947		
## 25	0.0004876189	0.058484600		
## 26	0.0003813010	0.031867957		
## 27	0.0006038688	0.082390553		
## 28	0.0012568812	0.733308929		
## 29	0.0008904410	0.365939718		
## 30	0.0008742818	0.061280320		
## 31	0.0004375756	0.059155506		
## 32	0.0001223299	0.006334487		
## 33	0.0004376266	0.016877841		
## 34	0.0004427100	0.037690791		
## 35	0.0005162178	0.176915430		
## 36	0.0004162366	0.075683054		
## 37	0.0006440360	0.119542205		

If we do this part on the southern states It is basically the same, we just need to write

```
SouthStates=LGA_South$state
SouthStates=factor(SouthStates)
SouthStates=levels(SouthStates)
```

Instead of

```
States=levels(LGA_Data$state)
```

Now we have the list of all the southern states:

```
## [1] "Abia"          "Akwa Ibom"    "Anambra"      "Bayelsa"      "Cross River"
## [6] "Delta"        "Ebonyi"       "Edo"          "Ekiti"        "Enugu"
## [11] "Imo"          "Lagos"        "Ogun"         "Ondo"         "Osun"
## [16] "Oyo"          "Rivers"
```

The other is almost the same:

```
States=SouthStates
FacilityStatistics=data.frame(state=States,num_of_fulltime_nurses_and_doctors=rep(-1,length(States)),pop
for (n in c(1:length(States))){
  StateData=LGA_Data[LGA_Data$State==States[n],]
  Population=sum(StateData$pop_2006)
  Area=sum(StateData$area_sq_km)
  FacilityStatistics[n,3]=Population
  FacilityStatistics[n,4]=Area
  StateCity=StateData$unique_lga
  StateFacility=subset(Facility_Data,unique_lga %in% StateCity)
  nurse1=na.omit(StateFacility$num_nurses_fulltime)
  nurse2=na.omit(StateFacility$num_nursemidwives_fulltime)
  doctor=na.omit(StateFacility$num_doctors_fulltime)
  FulltimeTotal=sum(nurse1)+sum(nurse2)+sum(doctor)
  FacilityStatistics[n,5]=FulltimeTotal/Population
  FacilityStatistics[n,6]=FulltimeTotal/Area
  FacilityStatistics[n,2]=FulltimeTotal
}
library(plyr)
FacilityStatistics_Sorted=arrange(FacilityStatistics,desc(population))
```

The result is as follows:

```
##      state num_of_fulltime_nurses_and_doctors population      area
## 1      Lagos                                9028      9013534 3479.717
## 2        Oyo                                4417      5591589 27900.302
## 3      Rivers                                2908      5185400 10363.981
## 4     Anambra                                3807      4182032  4762.182
## 5       Delta                                3538      4098391 17105.597
## 6        Imo                                4904      3934899  5136.055
## 7   Akwa Ibom                                2146      3920208  6720.579
## 8        Ogun                                4947      3728098 16873.168
## 9        Ondo                                1673      3441024 15031.599
## 10       Osun                                2391      3423535  8595.119
## 11      Enugu                                2621      3257298  7561.111
## 12       Edo                                3336      3218332 19596.676
## 13 Cross River                                1725      2856581 20936.867
## 14       Abia                                3562      2833999  4857.434
## 15      Ekiti                                2123      2384212  5801.502
## 16     Ebonyi                                1122      2173501  6342.013
## 17     Bayelsa                                709      1703358  9368.015
##      facility_per_capita facility_per_area
## 1      0.0010016049      2.59446396
## 2      0.0007899365      0.15831370
## 3      0.0005608053      0.28058715
## 4      0.0009103230      0.79942349
```

## 5	0.0008632656	0.20683289
## 6	0.0012462836	0.95481836
## 7	0.0005474199	0.31931772
## 8	0.0013269501	0.29318739
## 9	0.0004861925	0.11129887
## 10	0.0006984009	0.27818113
## 11	0.0008046547	0.34664219
## 12	0.0010365618	0.17023295
## 13	0.0006038688	0.08239055
## 14	0.0012568812	0.73330893
## 15	0.0008904410	0.36593972
## 16	0.0005162178	0.17691543
## 17	0.0004162366	0.07568305