Comparing\_Hospitals

Munazza

March 15, 2017

## Loading required package: gsubfn

## Loading required package: proto

## Loading required package: RSQLite

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, round.POSIXt, trunc.POSIXt, units

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:Hmisc':  
##   
## combine, src, summarize

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

## The following object is masked from 'package:Hmisc':  
##   
## combine

## getting data

state\_info <- read.table('C:/Users/naseem/Documents/CKME136/states\_info.txt',header=T ,sep = '\t',na.string=c("","NA"))  
state\_info =data.frame(state\_info)  
names(state\_info)= c("State\_name","area\_ranking","area\_sqmi","area\_sqkm","land\_area\_sqmi","land\_area\_sqkm","population\_ranking","population","State","capital\_city")  
  
data2 <- read.table('C:/Users/naseem/Documents/CKME136/data/Hospital General Information.csv',header=T ,sep = ',',na.string=c("","NA","Not Available","NaN"))  
data2 = data.frame(data2)  
names(data2)<- c("Provider\_ID","Hospital\_Name","Address","City","State","ZIP\_Code","County\_Name","Phone\_Number",  
 "Hospital\_Type","Hospital\_Ownership","Emergency\_Services","Meets\_criteria\_for\_meaningful\_use\_of\_EHRs",  
 "Hospital\_overall\_rating","footnote1","Mortality\_national\_comparison",  
 "footnote2","Safety\_of\_care\_national\_comparison","footnote3",  
 "Readmission\_national\_comparison","footnote4","Patient\_experience\_national\_comparison",  
 "footnote5","Effectiveness\_of\_care\_national\_comparison",  
 "footnote6","Timeliness\_of\_care\_national\_comparison",  
 "footnote7","Efficient\_use\_of\_medical\_imaging\_national\_comparison",  
 "footnote8")  
  
gen\_info<-data2[,c("Provider\_ID","Hospital\_Name","State",  
 "Hospital\_Type","Hospital\_Ownership","Emergency\_Services",  
 "Meets\_criteria\_for\_meaningful\_use\_of\_EHRs",  
 "Hospital\_overall\_rating","Mortality\_national\_comparison",  
 "Safety\_of\_care\_national\_comparison",  
 "Readmission\_national\_comparison","Patient\_experience\_national\_comparison",  
 "Effectiveness\_of\_care\_national\_comparison",  
 "Timeliness\_of\_care\_national\_comparison",  
 "Efficient\_use\_of\_medical\_imaging\_national\_comparison")]

describe(state\_info)

## state\_info   
##   
## 10 Variables 56 Observations  
## ---------------------------------------------------------------------------  
## State\_name   
## n missing distinct   
## 56 0 56   
##   
## lowest : Alabama Alaska American Samoa Arizona Arkansas   
## highest: Virginia Washington West Virginia Wisconsin Wyoming   
## ---------------------------------------------------------------------------  
## area\_ranking   
## n missing distinct Info Mean Gmd .05 .10   
## 55 1 54 1 27.93 18.53 3.7 6.4   
## .25 .50 .75 .90 .95   
## 14.5 28.0 41.5 49.6 51.3   
##   
## lowest : 1 2 3 4 5, highest: 50 51 52 53 54  
## ---------------------------------------------------------------------------  
## area\_sqmi   
## n missing distinct Info Mean Gmd .05 .10   
## 56 0 56 1 67916 71299 695 2232   
## .25 .50 .75 .90 .95   
## 12038 53499 82600 112279 151205   
##   
## lowest : 68.25 570.62 581.05 732.93 1545.00  
## highest: 121589.00 147042.00 163695.00 268580.00 663267.00  
## ---------------------------------------------------------------------------  
## area\_sqkm   
## n missing distinct Info Mean Gmd .05 .10   
## 56 0 56 1 175901 184663 1800 5782   
## .25 .50 .75 .90 .95   
## 31178 138561 213934 290803 391621   
##   
## lowest : 176.75 1478.00 1505.00 1898.00 4002.00  
## highest: 314915.00 380838.00 423970.00 695621.00 1717854.00  
## ---------------------------------------------------------------------------  
## land\_area\_sqmi   
## n missing distinct Info Mean Gmd .05 .10   
## 56 0 55 1 63199 66678 170.3 2686.2   
## .25 .50 .75 .90 .95   
## 9584.6 49631.6 77524.8 111687.6 148104.2   
##   
## lowest : 61.05 76.46 134.32 182.33 1033.81  
## highest: 121298.15 145545.80 155779.22 261231.71 570640.95  
## ---------------------------------------------------------------------------  
## land\_area\_sqkm   
## n missing distinct Info Mean Gmd .05 .10   
## 56 0 55 1 163685 172694 441 6958   
## .25 .50 .75 .90 .95   
## 24824 128546 200789 289270 383588   
##   
## lowest : 158 198 348 472 2678  
## highest: 314161 376962 403466 676587 1477953  
## ---------------------------------------------------------------------------  
## population\_ranking   
## n missing distinct Info Mean Gmd .05 .10   
## 56 0 56 1 28.59 19.17 3.75 6.50   
## .25 .50 .75 .90 .95   
## 14.75 28.50 42.25 50.50 54.25   
##   
## lowest : 1 2 3 4 5, highest: 53 54 55 56 57  
## ---------------------------------------------------------------------------  
## population   
## n missing distinct Info Mean Gmd .05 .10   
## 56 0 56 1 5716357 6306881 147232 636540   
## .25 .50 .75 .90 .95   
## 1327091 3732827 6643174 12172305 19577427   
##   
## lowest : 52344 54343 103574 161785 582658  
## highest: 12882135 19552860 19651127 26448193 38332521  
## ---------------------------------------------------------------------------  
## State   
## n missing distinct   
## 56 0 56   
##   
## lowest : AK AL AR AS AZ, highest: VT WA WI WV WY  
## ---------------------------------------------------------------------------  
## capital\_city   
## n missing distinct   
## 50 6 50   
##   
## lowest : Albany Annapolis Atlanta Augusta Austin   
## highest: Springfield St. Paul Tallahassee Topeka Trenton   
## ---------------------------------------------------------------------------

describe(gen\_info)

## gen\_info   
##   
## 15 Variables 4807 Observations  
## ---------------------------------------------------------------------------  
## Provider\_ID   
## n missing distinct Info Mean Gmd .05 .10   
## 4807 0 4807 1 268087 179660 41305 50568   
## .25 .50 .75 .90 .95   
## 140188 260034 390218 460051 510090   
##   
## lowest : 10001 10005 10006 10007 10008, highest: 670108 670109 670111 670112 670114  
## ---------------------------------------------------------------------------  
## Hospital\_Name   
## n missing distinct   
## 4807 0 4608   
##   
## lowest : ABBEVILLE AREA MEDICAL CENTER ABBEVILLE GENERAL HOSPITAL ABBOTT NORTHWESTERN HOSPITAL ABILENE REGIONAL MEDICAL CENTER ABINGTON MEMORIAL HOSPITAL   
## highest: YORK HOSPITAL YUKON KUSKOKWIM DELTA REG HOSPITAL YUMA DISTRICT HOSPITAL YUMA REGIONAL MEDICAL CENTER ZUNI COMPREHENSIVE COMMUNITY HEALTH CENTER  
## ---------------------------------------------------------------------------  
## State   
## n missing distinct   
## 4807 0 56   
##   
## lowest : AK AL AR AS AZ, highest: VT WA WI WV WY  
## ---------------------------------------------------------------------------  
## Hospital\_Type   
## n missing distinct   
## 4807 0 3   
##   
## Value Acute Care Hospitals Childrens  
## Frequency 3370 99  
## Proportion 0.701 0.021  
##   
## Value Critical Access Hospitals  
## Frequency 1338  
## Proportion 0.278  
## ---------------------------------------------------------------------------  
## Hospital\_Ownership   
## n missing distinct   
## 4807 0 10   
##   
## lowest : Government - Federal Government - Hospital District or Authority Government - Local Government - State Physician   
## highest: Proprietary Tribal Voluntary non-profit - Church Voluntary non-profit - Other Voluntary non-profit - Private   
## ---------------------------------------------------------------------------  
## Emergency\_Services   
## n missing distinct   
## 4807 0 2   
##   
## Value No Yes  
## Frequency 295 4512  
## Proportion 0.061 0.939  
## ---------------------------------------------------------------------------  
## Meets\_criteria\_for\_meaningful\_use\_of\_EHRs   
## n missing distinct value   
## 4373 434 1 Y   
##   
## Value Y  
## Frequency 4373  
## Proportion 1  
## ---------------------------------------------------------------------------  
## Hospital\_overall\_rating   
## n missing distinct Info Mean Gmd   
## 3584 1223 5 0.854 3.059 0.8541   
##   
## Value 1 2 3 4 5  
## Frequency 108 679 1774 941 82  
## Proportion 0.030 0.189 0.495 0.263 0.023  
## ---------------------------------------------------------------------------  
## Mortality\_national\_comparison   
## n missing distinct   
## 3477 1330 3   
##   
## Value Above the National average Below the National average  
## Frequency 400 341  
## Proportion 0.115 0.098  
##   
## Value Same as the National average  
## Frequency 2736  
## Proportion 0.787  
## ---------------------------------------------------------------------------  
## Safety\_of\_care\_national\_comparison   
## n missing distinct   
## 2654 2153 3   
##   
## Value Above the National average Below the National average  
## Frequency 787 665  
## Proportion 0.297 0.251  
##   
## Value Same as the National average  
## Frequency 1202  
## Proportion 0.453  
## ---------------------------------------------------------------------------  
## Readmission\_national\_comparison   
## n missing distinct   
## 3813 994 3   
##   
## Value Above the National average Below the National average  
## Frequency 811 869  
## Proportion 0.213 0.228  
##   
## Value Same as the National average  
## Frequency 2133  
## Proportion 0.559  
## ---------------------------------------------------------------------------  
## Patient\_experience\_national\_comparison   
## n missing distinct   
## 3454 1353 3   
##   
## Value Above the National average Below the National average  
## Frequency 1217 1084  
## Proportion 0.352 0.314  
##   
## Value Same as the National average  
## Frequency 1153  
## Proportion 0.334  
## ---------------------------------------------------------------------------  
## Effectiveness\_of\_care\_national\_comparison   
## n missing distinct   
## 3627 1180 3   
##   
## Value Above the National average Below the National average  
## Frequency 130 245  
## Proportion 0.036 0.068  
##   
## Value Same as the National average  
## Frequency 3252  
## Proportion 0.897  
## ---------------------------------------------------------------------------  
## Timeliness\_of\_care\_national\_comparison   
## n missing distinct   
## 3565 1242 3   
##   
## Value Above the National average Below the National average  
## Frequency 1096 905  
## Proportion 0.307 0.254  
##   
## Value Same as the National average  
## Frequency 1564  
## Proportion 0.439  
## ---------------------------------------------------------------------------  
## Efficient\_use\_of\_medical\_imaging\_national\_comparison   
## n missing distinct   
## 2790 2017 3   
##   
## Value Above the National average Below the National average  
## Frequency 358 373  
## Proportion 0.128 0.134  
##   
## Value Same as the National average  
## Frequency 2059  
## Proportion 0.738  
## ---------------------------------------------------------------------------

## Analyzing dataset

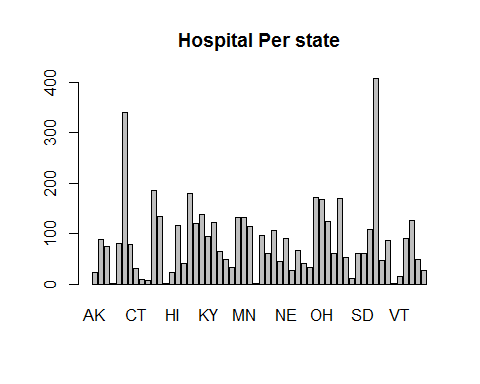
table(gen\_info$State)

##   
## AK AL AR AS AZ CA CO CT DC DE FL GA GU HI IA ID IL IN   
## 22 89 74 1 81 341 78 31 8 7 186 134 2 23 116 41 180 120   
## KS KY LA MA MD ME MI MN MO MP MS MT NC ND NE NH NJ NM   
## 137 94 122 64 49 33 131 131 114 1 96 61 106 44 90 26 66 41   
## NV NY OH OK OR PA PR RI SC SD TN TX UT VA VI VT WA WI   
## 33 172 168 123 60 170 52 11 60 60 109 407 46 87 2 14 91 126   
## WV WY   
## 49 27

max(table(gen\_info$State))

## [1] 407

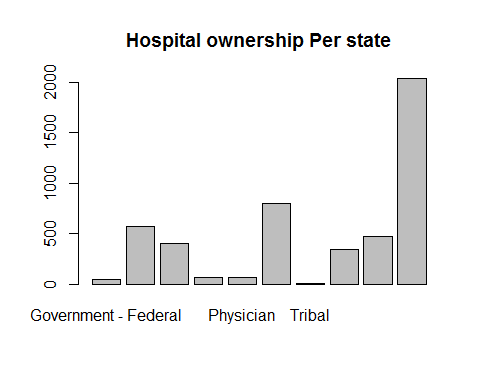
barplot(table(gen\_info$State),main="Hospital Per state")



table(gen\_info$Hospital\_Ownership)

##   
## Government - Federal   
## 46   
## Government - Hospital District or Authority   
## 566   
## Government - Local   
## 406   
## Government - State   
## 67   
## Physician   
## 64   
## Proprietary   
## 796   
## Tribal   
## 8   
## Voluntary non-profit - Church   
## 344   
## Voluntary non-profit - Other   
## 473   
## Voluntary non-profit - Private   
## 2037

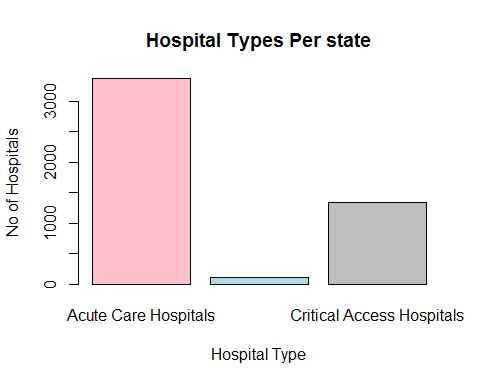
barplot(table(gen\_info$Hospital\_Ownership),main="Hospital ownership Per state",names.arg = NULL )



table(gen\_info$Hospital\_Type)

##   
## Acute Care Hospitals Childrens   
## 3370 99   
## Critical Access Hospitals   
## 1338

barplot(table(gen\_info$Hospital\_Type),main="Hospital Types Per state",names=c("Acute Care Hospitals","Childrens","Critical Access Hospitals"),col = c("pink","lightblue","grey"),xlab = "Hospital Type",ylab = "No of Hospitals")



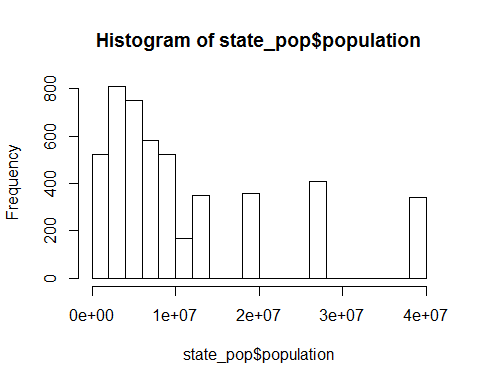
##Joining state\_info and gen\_info  
  
state\_pop<-left\_join(gen\_info,state\_info)

## Joining, by = "State"

head(state\_pop)

## Provider\_ID Hospital\_Name State Hospital\_Type  
## 1 10001 SOUTHEAST ALABAMA MEDICAL CENTER AL Acute Care Hospitals  
## 2 10005 MARSHALL MEDICAL CENTER SOUTH AL Acute Care Hospitals  
## 3 10006 ELIZA COFFEE MEMORIAL HOSPITAL AL Acute Care Hospitals  
## 4 10007 MIZELL MEMORIAL HOSPITAL AL Acute Care Hospitals  
## 5 10008 CRENSHAW COMMUNITY HOSPITAL AL Acute Care Hospitals  
## 6 10011 ST VINCENT'S EAST AL Acute Care Hospitals  
## Hospital\_Ownership Emergency\_Services  
## 1 Government - Hospital District or Authority Yes  
## 2 Government - Hospital District or Authority Yes  
## 3 Government - Hospital District or Authority Yes  
## 4 Voluntary non-profit - Private Yes  
## 5 Proprietary Yes  
## 6 Voluntary non-profit - Private Yes  
## Meets\_criteria\_for\_meaningful\_use\_of\_EHRs Hospital\_overall\_rating  
## 1 Y 3  
## 2 Y 3  
## 3 Y 2  
## 4 Y 2  
## 5 Y 3  
## 6 Y 2  
## Mortality\_national\_comparison Safety\_of\_care\_national\_comparison  
## 1 Same as the National average Above the National average  
## 2 Below the National average Same as the National average  
## 3 Below the National average Same as the National average  
## 4 Same as the National average <NA>  
## 5 Same as the National average <NA>  
## 6 Same as the National average Below the National average  
## Readmission\_national\_comparison Patient\_experience\_national\_comparison  
## 1 Same as the National average Below the National average  
## 2 Above the National average Same as the National average  
## 3 Same as the National average Below the National average  
## 4 Below the National average Same as the National average  
## 5 Same as the National average <NA>  
## 6 Same as the National average Below the National average  
## Effectiveness\_of\_care\_national\_comparison  
## 1 Same as the National average  
## 2 Same as the National average  
## 3 Same as the National average  
## 4 Below the National average  
## 5 Same as the National average  
## 6 Below the National average  
## Timeliness\_of\_care\_national\_comparison  
## 1 Same as the National average  
## 2 Above the National average  
## 3 Above the National average  
## 4 Above the National average  
## 5 Above the National average  
## 6 Same as the National average  
## Efficient\_use\_of\_medical\_imaging\_national\_comparison State\_name  
## 1 Same as the National average Alaska  
## 2 Below the National average Alaska  
## 3 Same as the National average Alaska  
## 4 <NA> Alaska  
## 5 <NA> Alaska  
## 6 Same as the National average Alaska  
## area\_ranking area\_sqmi area\_sqkm land\_area\_sqmi land\_area\_sqkm  
## 1 1 663267 1717854 570640.9 1477953  
## 2 1 663267 1717854 570640.9 1477953  
## 3 1 663267 1717854 570640.9 1477953  
## 4 1 663267 1717854 570640.9 1477953  
## 5 1 663267 1717854 570640.9 1477953  
## 6 1 663267 1717854 570640.9 1477953  
## population\_ranking population capital\_city  
## 1 23 4833722 Montgomery  
## 2 23 4833722 Montgomery  
## 3 23 4833722 Montgomery  
## 4 23 4833722 Montgomery  
## 5 23 4833722 Montgomery  
## 6 23 4833722 Montgomery

state\_pop$population<-as.numeric(state\_pop$population)  
hist(state\_pop$population)



str(state\_pop)

## 'data.frame': 4807 obs. of 24 variables:  
## $ Provider\_ID : int 10001 10005 10006 10007 10008 10011 10012 10016 10018 10019 ...  
## $ Hospital\_Name : Factor w/ 4608 levels "ABBEVILLE AREA MEDICAL CENTER",..: 3677 2217 1105 2573 911 3924 981 3584 505 1561 ...  
## $ State : Factor w/ 56 levels "AK","AL","AR",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ Hospital\_Type : Factor w/ 3 levels "Acute Care Hospitals",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Hospital\_Ownership : Factor w/ 10 levels "Government - Federal",..: 2 2 2 10 6 10 6 10 10 2 ...  
## $ Emergency\_Services : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ Meets\_criteria\_for\_meaningful\_use\_of\_EHRs : Factor w/ 1 level "Y": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Hospital\_overall\_rating : int 3 3 2 2 3 2 3 3 NA 2 ...  
## $ Mortality\_national\_comparison : Factor w/ 3 levels "Above the National average",..: 3 2 2 3 3 3 2 3 NA 2 ...  
## $ Safety\_of\_care\_national\_comparison : Factor w/ 3 levels "Above the National average",..: 1 3 3 NA NA 2 3 1 NA 2 ...  
## $ Readmission\_national\_comparison : Factor w/ 3 levels "Above the National average",..: 3 1 3 2 3 3 3 1 NA 2 ...  
## $ Patient\_experience\_national\_comparison : Factor w/ 3 levels "Above the National average",..: 2 3 2 3 NA 2 3 2 NA 2 ...  
## $ Effectiveness\_of\_care\_national\_comparison : Factor w/ 3 levels "Above the National average",..: 3 3 3 2 3 2 3 3 3 3 ...  
## $ Timeliness\_of\_care\_national\_comparison : Factor w/ 3 levels "Above the National average",..: 3 1 1 1 1 3 1 2 3 1 ...  
## $ Efficient\_use\_of\_medical\_imaging\_national\_comparison: Factor w/ 3 levels "Above the National average",..: 3 2 3 NA NA 3 3 2 NA 2 ...  
## $ State\_name : Factor w/ 56 levels "Alabama","Alaska",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ area\_ranking : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ area\_sqmi : num 663267 663267 663267 663267 663267 ...  
## $ area\_sqkm : num 1717854 1717854 1717854 1717854 1717854 ...  
## $ land\_area\_sqmi : num 570641 570641 570641 570641 570641 ...  
## $ land\_area\_sqkm : num 1477953 1477953 1477953 1477953 1477953 ...  
## $ population\_ranking : int 23 23 23 23 23 23 23 23 23 23 ...  
## $ population : num 4833722 4833722 4833722 4833722 4833722 ...  
## $ capital\_city : Factor w/ 50 levels "Albany","Annapolis",..: 32 32 32 32 32 32 32 32 32 32 ...

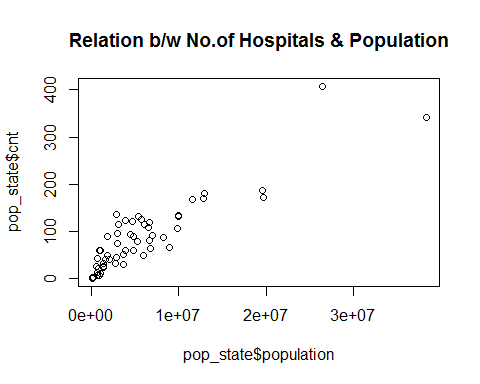
## Calculating the correlation between no.of hospitals per state and population,  
## no. of hospitals per state and area,and area of a state and population  
pop\_state<-sqldf('select State,count(\*) as cnt,population,land\_area\_sqkm as area from state\_pop group by State order by cnt')

## Warning: Quoted identifiers should have class SQL, use DBI::SQL() if the  
## caller performs the quoting.

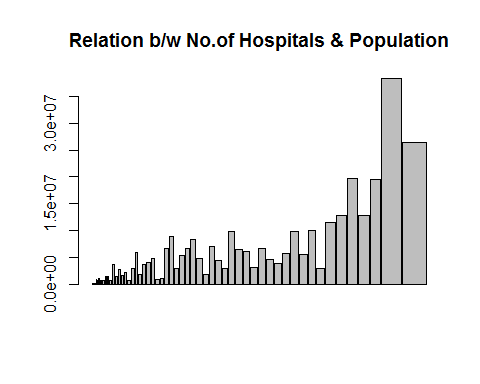
cor(pop\_state$cnt,pop\_state$population)

## [1] 0.8915243

plot(pop\_state$population,pop\_state$cnt,main = "Relation b/w No.of Hospitals & Population")



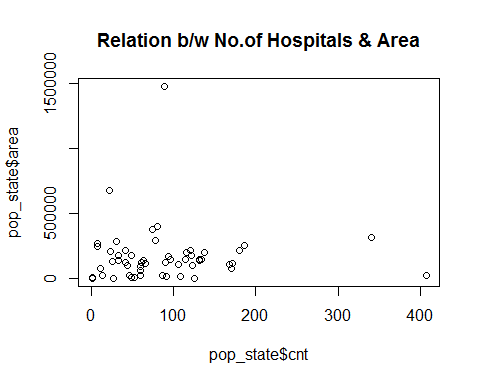
barplot(pop\_state$population,pop\_state$cnt,main = "Relation b/w No.of Hospitals & Population")



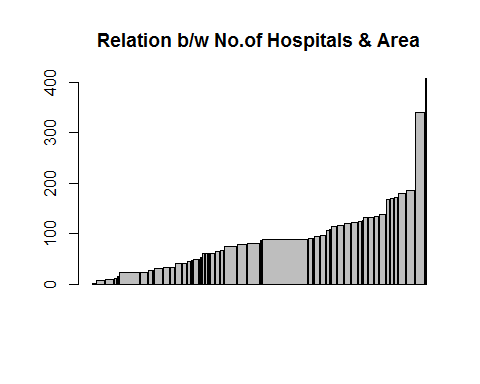
## 0.891 shows that there is a strong positive relation between the no of hospitals and population  
  
cor(pop\_state$cnt,pop\_state$area)

## [1] 0.03546556

plot(pop\_state$cnt,pop\_state$area, main = "Relation b/w No.of Hospitals & Area")



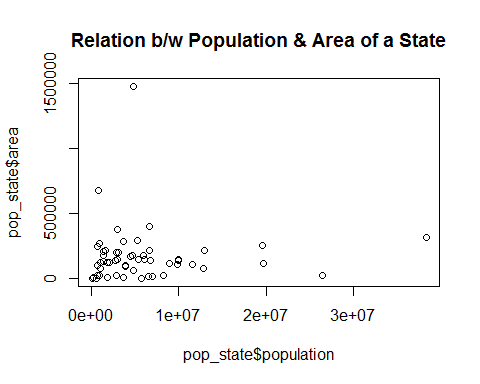
barplot(pop\_state$cnt,pop\_state$area, main = "Relation b/w No.of Hospitals & Area")



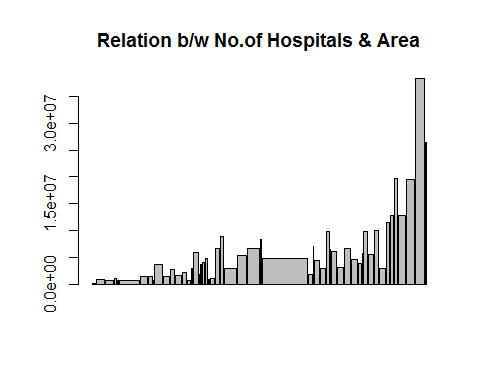
## .0354 means there is a positive but weak relation between no. of hospitals and area of a state  
## Finding out if there is a relation between area and population.  
cor(pop\_state$population,pop\_state$area)

## [1] 0.04492212

plot(pop\_state$population,pop\_state$area, main = "Relation b/w Population & Area of a State")



barplot(pop\_state$population,pop\_state$area, main = "Relation b/w No.of Hospitals & Area")



## No of hospitals is related to population of state and its a positive relation which mean higher the no. of hospitals greater the poulation would be.

## Comparing Hospitals in terms of Ratings

rating\_state<-sqldf('select State,count(\*) as cnt,sum(Hospital\_overall\_rating) as rating from gen\_info group by State order by cnt')  
rating\_state$rating= as.numeric(rating\_state$rating)  
rating\_state=mutate(rating\_state,state.rating=rating/cnt)  
top\_state\_rating=(arrange(rating\_state,desc(state.rating)))  
head(top\_state\_rating,n=10)

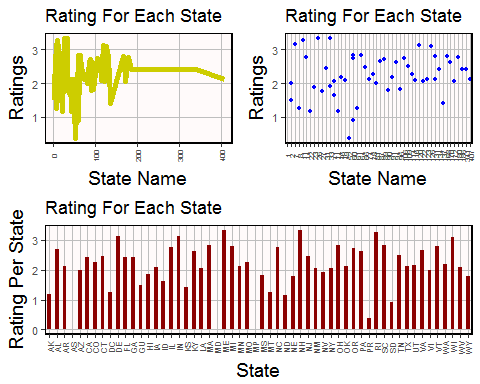
## State cnt rating state.rating  
## 1 NH 26 87 3.346154  
## 2 ME 33 110 3.333333  
## 3 RI 11 36 3.272727  
## 4 DE 7 22 3.142857  
## 5 IN 120 376 3.133333  
## 6 WI 126 390 3.095238  
## 7 MA 64 182 2.843750  
## 8 SC 60 170 2.833333  
## 9 OH 168 473 2.815476  
## 10 MI 131 367 2.801527

hospital.theme<-theme(  
 axis.text = element\_text(size = 8),  
 axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust = 0.5),  
 axis.title = element\_text(size = 14),  
 panel.grid.major = element\_line(color = "grey"),  
 panel.grid.minor = element\_blank(),  
 panel.background = element\_rect(fill = "snow1"),  
 legend.position = "right",  
 legend.justification = "top",   
 legend.background = element\_blank(),  
 panel.border = element\_rect(color = "black", fill = NA, size = 1))  
  
plot.rating.cnt<-ggplot(data = rating\_state,  
 aes(x=as.numeric(cnt),  
 y=state.rating))+  
 geom\_line(size=2,col="yellow3")+  
 hospital.theme+  
 ggtitle("Rating For Each State")+  
 labs(x="State Name",  
 y="Ratings")+  
 theme(axis.text.x=element\_text(size= 5, angle=90,hjust = 0.5))  
  
plot.rating.cnt.points<-ggplot(data = rating\_state,  
 aes(x=as.factor(cnt),  
 y=state.rating))+  
 geom\_point(size=1,col="blue")+  
 hospital.theme+  
 ggtitle("Rating For Each State")+  
 labs(x="State Name",  
 y="Ratings")+  
 theme(axis.text.x=element\_text(size= 5, angle=90,hjust = 0.5))  
plot.by.state <- ggplot(data = rating\_state,  
 aes(x=as.factor(State),  
 y=state.rating))+  
 geom\_bar(stat= "identity", fill="darkred", width=0.5 )+  
 hospital.theme+  
 ggtitle("Rating For Each State")+  
 labs(x="State",  
 y="Rating Per State")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))  
  
grid.arrange(arrangeGrob(plot.rating.cnt,plot.rating.cnt.points,ncol=2),  
 plot.by.state)

## Warning: Removed 2 rows containing missing values (geom\_path).

## Warning: Removed 3 rows containing missing values (geom\_point).

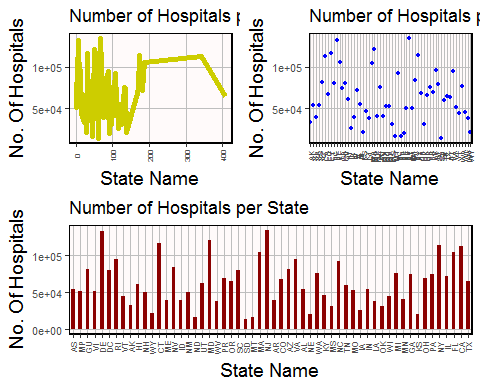
## Warning: Removed 3 rows containing missing values (position\_stack).

 ## Top 10 States in terms of No of Hospitals

## Finding the Best 10 State in term of no. of hospitals provided by State over population and area per state  
  
pop\_state<-sqldf('select State,count(\*) as cnt,population,land\_area\_sqkm as area from state\_pop group by State order by cnt')  
pop\_state$cnt= as.numeric(pop\_state$cnt)  
pop\_state=mutate(pop\_state,hos\_pop\_ratio=population/cnt,hos\_area\_ratio=area/cnt)  
top\_state\_pop\_ratio=(arrange(pop\_state,desc(hos\_pop\_ratio)))  
head(top\_state\_pop\_ratio,n=10)

## State cnt population area hos\_pop\_ratio hos\_area\_ratio  
## 1 NJ 66 8899339 111898 134838.47 1695.4242  
## 2 DE 7 925749 268431 132249.86 38347.2857  
## 3 MD 49 5928814 178040 120996.20 3633.4694  
## 4 CT 31 3596080 284332 116002.58 9172.0000  
## 5 NY 172 19651127 115883 114250.74 673.7384  
## 6 CA 341 38332521 314161 112412.09 921.2933  
## 7 FL 186 19552860 251470 105122.90 1351.9892  
## 8 MA 64 6692824 138887 104575.38 2170.1094  
## 9 RI 11 1051511 77857 95591.91 7077.9091  
## 10 VA 87 8260405 19047 94947.18 218.9310

## theme of the results  
  
  
plot.hos<-ggplot(data = pop\_state,  
 aes(x=as.numeric(cnt),  
 y=hos\_pop\_ratio))+  
 geom\_line(size=2,col="yellow3")+  
 hospital.theme+  
 ggtitle("Number of Hospitals per State")+  
 labs(x="State Name",  
 y="No. Of Hospitals")+  
 theme(axis.text.x=element\_text(size= 5, angle=90,hjust = 0.5))  
  
plot.hos.points<-ggplot(data = pop\_state,  
 aes(x=as.factor(State),  
 y=hos\_pop\_ratio))+  
 geom\_point(size=1,col="blue")+  
 hospital.theme+  
 ggtitle("Number of Hospitals per State")+  
 labs(x="State Name",  
 y="No. Of Hospitals")+  
 theme(axis.text.x=element\_text(size= 5, angle=90,hjust = 0.5))  
pop\_state$State <- fct\_inorder(pop\_state$State)  
plot.by.state <- ggplot(data = pop\_state,  
 aes(x=as.factor(State),  
 y=hos\_pop\_ratio))+  
 geom\_bar(stat= "identity", fill="darkred", width=0.5 )+  
 hospital.theme+  
 ggtitle("Number of Hospitals per State")+  
 labs(x="State Name",  
 y="No. Of Hospitals")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))  
  
grid.arrange(arrangeGrob(plot.hos,plot.hos.points,ncol=2),  
 plot.by.state)

 ## Finding NA's in other variables of gen\_info

sum(is.na(gen\_info$Meets\_criteria\_for\_meaningful\_use\_of\_EHRs) == TRUE)

## [1] 434

length(gen\_info$Meets\_criteria\_for\_meaningful\_use\_of\_EHRs)

## [1] 4807

table(gen\_info$Meets\_criteria\_for\_meaningful\_use\_of\_EHRs,useNA="always")

##   
## Y <NA>   
## 4373 434

gen\_info$Meets\_criteria\_for\_meaningful\_use\_of\_EHRs[is.na(gen\_info$Meets\_criteria\_for\_meaningful\_use\_of\_EHRs)]="Y"  
table(gen\_info$Meets\_criteria\_for\_meaningful\_use\_of\_EHRs,useNA="always")

##   
## Y <NA>   
## 4807 0

##Meets\_criteria\_for\_meaningful\_use\_of\_EHRs cannot be used for any purpose as there is no variance.  
  
sum(is.na(gen\_info$Hospital\_overall\_rating) == TRUE)

## [1] 1223

table(gen\_info$Hospital\_overall\_rating,useNA="always")

##   
## 1 2 3 4 5 <NA>   
## 108 679 1774 941 82 1223

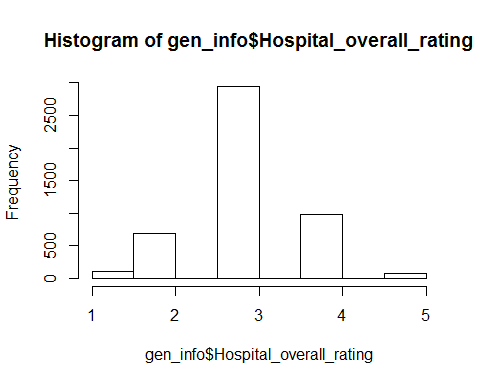
gen\_info$Hospital\_overall\_rating=as.numeric(gen\_info$Hospital\_overall\_rating)  
xx=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)  
x=select(xx,Hospital\_Ownership,Hospital\_overall\_rating)  
group.x=group\_by(x,Hospital\_Ownership)  
summarize(group.x,mean=mean(Hospital\_overall\_rating))

## Source: local data frame [10 x 2]  
##   
## Hospital\_Ownership mean  
## <fctr> <dbl>  
## 1 Government - Federal 2.941176  
## 2 Government - Hospital District or Authority 3.028736  
## 3 Government - Local 2.954733  
## 4 Government - State 2.638298  
## 5 Physician 3.888889  
## 6 Proprietary 2.845659  
## 7 Tribal 2.500000  
## 8 Voluntary non-profit - Church 3.186495  
## 9 Voluntary non-profit - Other 3.157068  
## 10 Voluntary non-profit - Private 3.120452

gen\_info$Hospital\_overall\_rating[grepl("Physician",gen\_info$Hospital\_Ownership) & is.na(gen\_info$Hospital\_overall\_rating)]=4  
gen\_info$Hospital\_overall\_rating[grepl("Tribal",gen\_info$Hospital\_Ownership) & is.na(gen\_info$Hospital\_overall\_rating)]=2  
gen\_info$Hospital\_overall\_rating[is.na(gen\_info$Hospital\_overall\_rating)]=3  
  
table(gen\_info$Hospital\_overall\_rating,useNA="always")

##   
## 1 2 3 4 5 <NA>   
## 108 685 2945 987 82 0

hist(gen\_info$Hospital\_overall\_rating)



sum(is.na(gen\_info$Safety\_of\_care\_national\_comparison) == TRUE)

## [1] 2153

table(gen\_info$Safety\_of\_care\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 787 665   
## Same as the National average <NA>   
## 1202 2153

sum(is.na(gen\_info$Mortality\_national\_comparison) == TRUE)

## [1] 1330

table(gen\_info$Mortality\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 400 341   
## Same as the National average <NA>   
## 2736 1330

sum(is.na(gen\_info$Readmission\_national\_comparison) == TRUE)

## [1] 994

table(gen\_info$Readmission\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 811 869   
## Same as the National average <NA>   
## 2133 994

sum(is.na(gen\_info$Patient\_experience\_national\_comparison) == TRUE)

## [1] 1353

table(gen\_info$Patient\_experience\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 1217 1084   
## Same as the National average <NA>   
## 1153 1353

sum(is.na(gen\_info$Effectiveness\_of\_care\_national\_comparison) == TRUE)

## [1] 1180

table(gen\_info$Effectiveness\_of\_care\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 130 245   
## Same as the National average <NA>   
## 3252 1180

sum(is.na(gen\_info$Timeliness\_of\_care\_national\_comparison) == TRUE)

## [1] 1242

table(gen\_info$Timeliness\_of\_care\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 1096 905   
## Same as the National average <NA>   
## 1564 1242

sum(is.na(gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison) == TRUE)

## [1] 2017

table(gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 358 373   
## Same as the National average <NA>   
## 2059 2017

nrow(gen\_info)

## [1] 4807

##Finding the combination of variables in order to remove max. no of NA's  
nrow(gen\_info[is.na(  
 gen\_info$Readmission\_national\_comparison==T&  
 gen\_info$Mortality\_national\_comparison==T&  
 gen\_info$Effectiveness\_of\_care\_national\_comparison==T&  
 gen\_info$Patient\_experience\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),])

## [1] 575

nrow(gen\_info[is.na(  
 gen\_info$Mortality\_national\_comparison==T&  
 gen\_info$Effectiveness\_of\_care\_national\_comparison==T&  
 gen\_info$Patient\_experience\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),])

## [1] 687

nrow(gen\_info[is.na(  
 gen\_info$Readmission\_national\_comparison==T&  
 gen\_info$Mortality\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),])

## [1] 733

nrow(gen\_info[is.na(  
 gen\_info$Effectiveness\_of\_care\_national\_comparison==T&  
 gen\_info$Patient\_experience\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),])

## [1] 817

nrow(gen\_info[is.na(  
 gen\_info$Readmission\_national\_comparison==T&  
 gen\_info$Patient\_experience\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),])

## [1] 613

nrow(gen\_info[is.na(  
 gen\_info$Mortality\_national\_comparison==T&  
 gen\_info$Effectiveness\_of\_care\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),])

## [1] 758

nrow(gen\_info[is.na(  
 gen\_info$Mortality\_national\_comparison==T&  
 gen\_info$Effectiveness\_of\_care\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),])

## [1] 758

nrow(gen\_info[is.na(  
   
 gen\_info$Mortality\_national\_comparison==T&  
 gen\_info$Effectiveness\_of\_care\_national\_comparison==T&  
 gen\_info$Patient\_experience\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),])

## [1] 752

nrow(gen\_info[is.na(  
 gen\_info$Mortality\_national\_comparison==T&  
 gen\_info$Effectiveness\_of\_care\_national\_comparison==T&  
 gen\_info$Patient\_experience\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T),])

## [1] 697

nrow(gen\_info[is.na(  
 gen\_info$Readmission\_national\_comparison==T&  
 gen\_info$Effectiveness\_of\_care\_national\_comparison==T&  
 gen\_info$Patient\_experience\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T),])

## [1] 582

##Using given combination of variables will remove max. rows containing NA's  
nrow(gen\_info)

## [1] 4807

gen\_info\_NA=gen\_info[is.na(  
 gen\_info$Readmission\_national\_comparison==T&  
 gen\_info$Mortality\_national\_comparison==T&  
 gen\_info$Timeliness\_of\_care\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),]  
nrow(gen\_info\_NA)

## [1] 733

gen\_info<- setdiff(gen\_info,gen\_info\_NA)  
nrow(gen\_info)

## [1] 4074

gen\_info\_NA\_1=gen\_info[is.na(  
 gen\_info$Mortality\_national\_comparison==T&  
 gen\_info$Effectiveness\_of\_care\_national\_comparison==T&  
 gen\_info$Patient\_experience\_national\_comparison==T&  
 gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison==T),]  
nrow(gen\_info\_NA\_1)

## [1] 177

gen\_info<- setdiff(gen\_info,gen\_info\_NA\_1)  
nrow(gen\_info)

## [1] 3897

## Recalculating NA's after filtering rows and assigning values to remaining NA's

gen\_info$Safety\_of\_care\_national\_comparison[is.na(gen\_info$Safety\_of\_care\_national\_comparison)]="Same as the National average"  
table(gen\_info$Safety\_of\_care\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 779 663   
## Same as the National average <NA>   
## 2455 0

gen\_info$Mortality\_national\_comparison[is.na(gen\_info$Mortality\_national\_comparison)]="Same as the National average"  
table(gen\_info$Mortality\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 400 341   
## Same as the National average <NA>   
## 3156 0

gen\_info$Readmission\_national\_comparison[is.na(gen\_info$Readmission\_national\_comparison)]="Same as the National average"  
table(gen\_info$Readmission\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 811 866   
## Same as the National average <NA>   
## 2220 0

gen\_info$Patient\_experience\_national\_comparison[is.na(gen\_info$Patient\_experience\_national\_comparison)]="Same as the National average"  
table(gen\_info$Patient\_experience\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 1121 1077   
## Same as the National average <NA>   
## 1699 0

gen\_info$Effectiveness\_of\_care\_national\_comparison[is.na(gen\_info$Effectiveness\_of\_care\_national\_comparison)]="Same as the National average"  
table(gen\_info$Effectiveness\_of\_care\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 126 238   
## Same as the National average <NA>   
## 3533 0

gen\_info$Timeliness\_of\_care\_national\_comparison[is.na(gen\_info$Timeliness\_of\_care\_national\_comparison)]="Same as the National average"  
table(gen\_info$Timeliness\_of\_care\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 1076 905   
## Same as the National average <NA>   
## 1916 0

gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison[is.na(gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison)]="Same as the National average"  
table(gen\_info$Efficient\_use\_of\_medical\_imaging\_national\_comparison,useNA="always")

##   
## Above the National average Below the National average   
## 358 373   
## Same as the National average <NA>   
## 3166 0

## Filtering Rows with the value "Same as National average" as it will effect the final results and filtering the rows which doesnt have "Same as National average" in any variable.

nrow(gen\_info)

## [1] 3897

nrow(filter(gen\_info,  
 Mortality\_national\_comparison=="Same as the National average"&  
 Effectiveness\_of\_care\_national\_comparison=="Same as the National average"&  
 Safety\_of\_care\_national\_comparison=="Same as the National average"&  
 Readmission\_national\_comparison=="Same as the National average"&  
 Patient\_experience\_national\_comparison=="Same as the National average"&  
 Timeliness\_of\_care\_national\_comparison=="Same as the National average"&  
 Efficient\_use\_of\_medical\_imaging\_national\_comparison=="Same as the National average"))

## [1] 566

nrow(filter(gen\_info,  
 Mortality\_national\_comparison=="Same as the National average"|  
 Effectiveness\_of\_care\_national\_comparison=="Same as the National average"|  
 Safety\_of\_care\_national\_comparison=="Same as the National average"|  
 Readmission\_national\_comparison=="Same as the National average"|  
 Patient\_experience\_national\_comparison=="Same as the National average"|  
 Timeliness\_of\_care\_national\_comparison=="Same as the National average"|  
 Efficient\_use\_of\_medical\_imaging\_national\_comparison=="Same as the National average"))

## [1] 3874

nrow(filter(gen\_info,  
 Mortality\_national\_comparison!="Same as the National average"&  
 Effectiveness\_of\_care\_national\_comparison!="Same as the National average"&  
 Safety\_of\_care\_national\_comparison!="Same as the National average"&  
 Readmission\_national\_comparison!="Same as the National average"&  
 Patient\_experience\_national\_comparison!="Same as the National average"&  
 Timeliness\_of\_care\_national\_comparison!="Same as the National average"&  
 Efficient\_use\_of\_medical\_imaging\_national\_comparison!="Same as the National average"))

## [1] 23

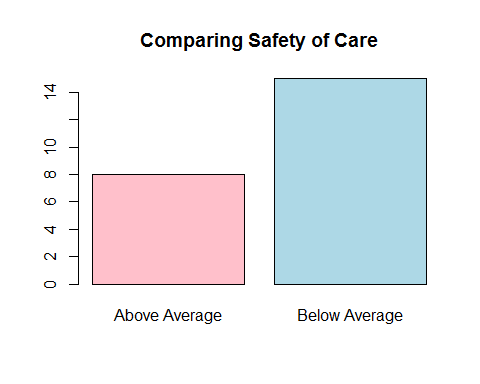
gen\_info\_filtr=filter(gen\_info,  
 Mortality\_national\_comparison!="Same as the National average"&  
 Effectiveness\_of\_care\_national\_comparison!="Same as the National average"&  
 Safety\_of\_care\_national\_comparison!="Same as the National average"&  
 Readmission\_national\_comparison!="Same as the National average"&  
 Patient\_experience\_national\_comparison!="Same as the National average"&  
 Timeliness\_of\_care\_national\_comparison!="Same as the National average"&  
 Efficient\_use\_of\_medical\_imaging\_national\_comparison!="Same as the National average")

## Ananlysing Hospitals which are either above or below the national average in all variables

gen\_info\_filtr=filter(gen\_info,  
 Mortality\_national\_comparison!="Same as the National average"&  
 Effectiveness\_of\_care\_national\_comparison!="Same as the National average"&  
 Safety\_of\_care\_national\_comparison!="Same as the National average"&  
 Readmission\_national\_comparison!="Same as the National average"&  
 Patient\_experience\_national\_comparison!="Same as the National average"&  
 Timeliness\_of\_care\_national\_comparison!="Same as the National average"&  
 Efficient\_use\_of\_medical\_imaging\_national\_comparison!="Same as the National average")  
  
table(gen\_info\_filtr$Safety\_of\_care\_national\_comparison)

##   
## Above the National average Below the National average   
## 8 15   
## Same as the National average   
## 0

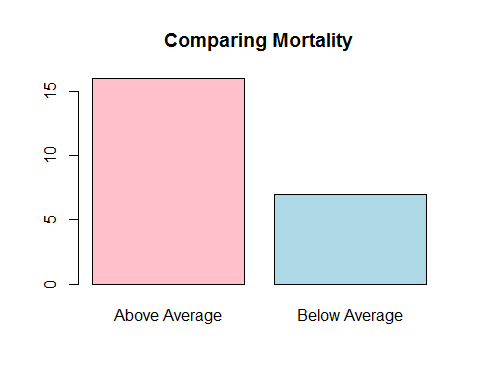
filter\_1 = sqldf('select Safety\_of\_care\_national\_comparison as safety\_of\_care,count(\*) as cnt from gen\_info\_filtr group by Safety\_of\_care\_national\_comparison')  
  
plot1=barplot(filter\_1$cnt,main = "Comparing Safety of Care",names=c("Above Average","Below Average"),col = c("pink","lightblue"))



table(gen\_info\_filtr$Mortality\_national\_comparison)

##   
## Above the National average Below the National average   
## 16 7   
## Same as the National average   
## 0

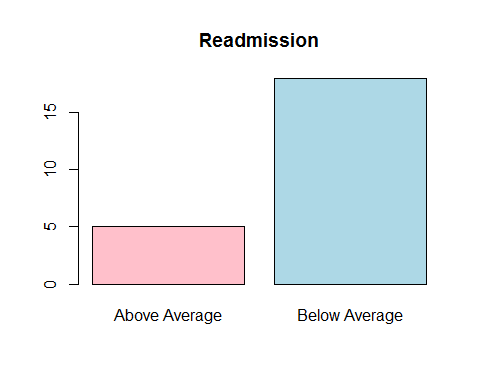
filter\_2 = sqldf('select Mortality\_national\_comparison as Mortality,count(\*) as cnt from gen\_info\_filtr group by Mortality\_national\_comparison')  
  
plot2=barplot(filter\_2$cnt,main = "Comparing Mortality",names=c("Above Average","Below Average"),col = c("pink","lightblue"))



table(gen\_info\_filtr$Readmission\_national\_comparison)

##   
## Above the National average Below the National average   
## 5 18   
## Same as the National average   
## 0

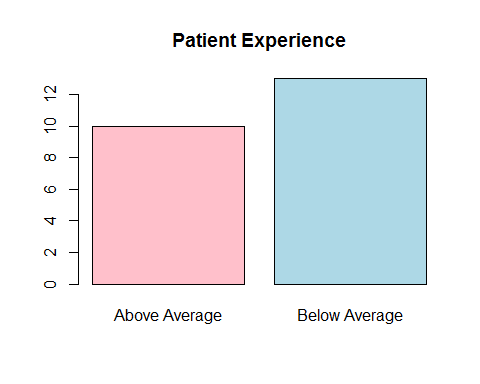
filter\_3 = sqldf('select Readmission\_national\_comparison as Readmission,count(\*) as cnt from gen\_info\_filtr group by Readmission\_national\_comparison')  
  
plot3=barplot(filter\_3$cnt,main = "Readmission",names=c("Above Average","Below Average"),col = c("pink","lightblue"))



table(gen\_info\_filtr$Patient\_experience\_national\_comparison)

##   
## Above the National average Below the National average   
## 10 13   
## Same as the National average   
## 0

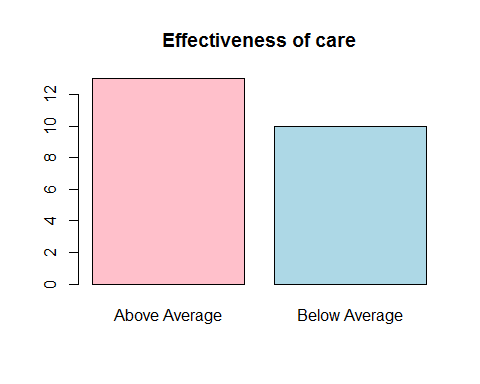
filter\_4 = sqldf('select Patient\_experience\_national\_comparison,count(\*) as cnt from gen\_info\_filtr group by Patient\_experience\_national\_comparison')  
  
plot4=barplot(filter\_4$cnt,main = "Patient Experience",names=c("Above Average","Below Average"),col = c("pink","lightblue"))



table(gen\_info\_filtr$Effectiveness\_of\_care\_national\_comparison)

##   
## Above the National average Below the National average   
## 13 10   
## Same as the National average   
## 0

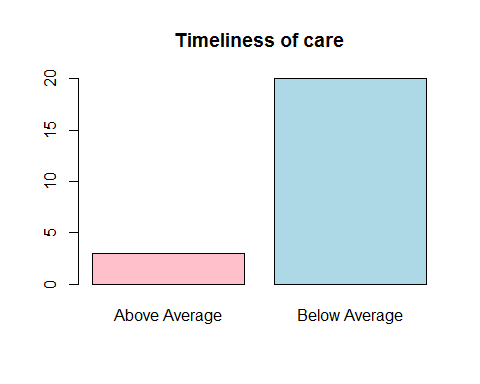
filter\_5 = sqldf('select Effectiveness\_of\_care\_national\_comparison,count(\*) as cnt from gen\_info\_filtr group by Effectiveness\_of\_care\_national\_comparison')  
  
plot5=barplot(filter\_5$cnt,main = "Effectiveness of care",names=c("Above Average","Below Average"),col = c("pink","lightblue"))



table(gen\_info\_filtr$Timeliness\_of\_care\_national\_comparison)

##   
## Above the National average Below the National average   
## 3 20   
## Same as the National average   
## 0

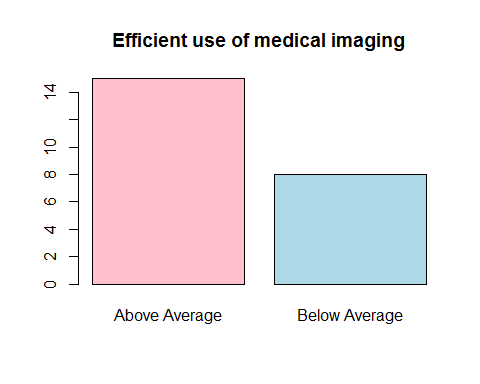
filter\_6 = sqldf('select Timeliness\_of\_care\_national\_comparison,count(\*) as cnt from gen\_info\_filtr group by Timeliness\_of\_care\_national\_comparison')  
  
plot6=barplot(filter\_6$cnt,main = "Timeliness of care",names=c("Above Average","Below Average"),col = c("pink","lightblue"))



table(gen\_info\_filtr$Efficient\_use\_of\_medical\_imaging\_national\_comparison)

##   
## Above the National average Below the National average   
## 15 8   
## Same as the National average   
## 0

filter\_7 = sqldf('select Efficient\_use\_of\_medical\_imaging\_national\_comparison,count(\*) as cnt from gen\_info\_filtr group by Efficient\_use\_of\_medical\_imaging\_national\_comparison')  
  
plot7=barplot(filter\_7$cnt,main = "Efficient use of medical imaging",names=c("Above Average","Below Average"),col = c("pink","lightblue"))



nrow(filter(gen\_info,  
 Mortality\_national\_comparison=="Above the National average"&  
 Effectiveness\_of\_care\_national\_comparison=="Above the National average"&  
 Safety\_of\_care\_national\_comparison=="Above the National average"&  
 Readmission\_national\_comparison=="Above the National average"&  
 Patient\_experience\_national\_comparison=="Above the National average"&  
 Timeliness\_of\_care\_national\_comparison=="Above the National average"&  
 Efficient\_use\_of\_medical\_imaging\_national\_comparison=="Same as the National average"))

## [1] 0

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.