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

## Loading required package: raster

## Loading required package: sp

##   
## Attaching package: 'raster'

## The following object is masked from 'package:tidyr':  
##   
## extract

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

## The following objects are masked from 'package:Hmisc':  
##   
## mask, zoom

##   
## Attaching package: 'dismo'

## The following object is masked from 'package:ggmap':  
##   
## geocode

## rgdal: version: 1.2-5, (SVN revision 648)  
## Geospatial Data Abstraction Library extensions to R successfully loaded  
## Loaded GDAL runtime: GDAL 2.0.1, released 2015/09/15  
## Path to GDAL shared files: C:/Users/naseem/Documents/R/win-library/3.2/rgdal/gdal  
## GDAL does not use iconv for recoding strings.  
## Loaded PROJ.4 runtime: Rel. 4.9.1, 04 March 2015, [PJ\_VERSION: 491]  
## Path to PROJ.4 shared files: C:/Users/naseem/Documents/R/win-library/3.2/rgdal/proj  
## Linking to sp version: 1.2-4

## Loading required package: rJava

## Rattle: A free graphical interface for data mining with R.  
## Version 4.1.0 Copyright (c) 2006-2015 Togaware Pty Ltd.  
## Type 'rattle()' to shake, rattle, and roll your data.

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

##   
## Attaching package: 'modeltools'

## The following object is masked from 'package:rJava':  
##   
## clone

## Loading required package: strucchange

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

##   
## Attaching package: 'party'

## The following object is masked from 'package:dismo':  
##   
## response

##   
## Attaching package: 'e1071'

## The following object is masked from 'package:raster':  
##   
## interpolate

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

## getting data

data2 <- read.table('C:/Users/naseem/Documents/CKME136/comparing hospitals/hospital\_information.txt',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","County\_Name","Phone\_Number",  
 "Hospital\_Type","Hospital\_Ownership","Emergency\_Services","Meets\_criteria\_for\_meaningful\_use\_of\_EHRs",  
 "Hospital\_overall\_rating","footnote1","mortality",  
 "footnote2","Safety\_of\_care","footnote3",  
 "Readmissions","footnote4","Patient\_experience",  
 "footnote5","Effectiveness\_of\_care",  
 "footnote6","Timeliness\_of\_care",  
 "footnote7","Efficient\_use\_of\_medical\_imaging",  
 "footnote8")  
  
gen\_info<-data2[,c("Provider\_ID","Hospital\_Name","City","State","zip",  
 "Hospital\_Type","Hospital\_Ownership","Emergency\_Services",  
 "Meets\_criteria\_for\_meaningful\_use\_of\_EHRs",  
 "Hospital\_overall\_rating","mortality",  
 "Safety\_of\_care",  
 "Readmissions","Patient\_experience",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care",  
 "Efficient\_use\_of\_medical\_imaging")]

## Data Intergration Steps

## Adding more Variables to gen\_info

library(noncensus)  
  
data(zip\_codes)## adding latitude and longitude based on zip cods  
head(zip\_codes,3)

## zip city state latitude longitude fips  
## 1 00210 Portsmouth NH 43.0059 -71.0132 33015  
## 2 00211 Portsmouth NH 43.0059 -71.0132 33015  
## 3 00212 Portsmouth NH 43.0059 -71.0132 33015

describe(zip\_codes)

## zip\_codes   
##   
## 6 Variables 43524 Observations  
## ---------------------------------------------------------------------------  
## zip   
## n missing distinct   
## 43524 0 43524   
##   
## lowest : 00210 00211 00212 00213 00214, highest: 99926 99927 99928 99929 99950  
## ---------------------------------------------------------------------------  
## city   
## n missing distinct   
## 43524 0 19071   
##   
## lowest : Aaronsburg Abbeville Abbot Abbotsford Abbott   
## highest: Zumbrota Zuni Zurich Zwingle Zwolle   
## ---------------------------------------------------------------------------  
## state   
## n missing distinct   
## 43524 0 52   
##   
## lowest : AK AL AR AZ CA, highest: VT WA WI WV WY  
## ---------------------------------------------------------------------------  
## latitude   
## n missing distinct Info Mean Gmd .05 .10   
## 43524 0 37890 1 38.51 5.86 29.70 31.65   
## .25 .50 .75 .90 .95   
## 35.00 39.13 41.88 44.49 46.27   
##   
## lowest : 17.96453 17.96977 17.97011 17.97637 17.97952  
## highest: 70.13346 70.21520 70.47766 70.64090 71.29953  
## ---------------------------------------------------------------------------  
## longitude   
## n missing distinct Info Mean Gmd .05 .10   
## 43524 0 37789 1 -90.86 16.31 -121.25 -116.56   
## .25 .50 .75 .90 .95   
## -97.22 -87.80 -80.06 -74.55 -72.67   
##   
## lowest : -176.63675 -174.19628 -171.70088 -170.40870 -170.27203  
## highest: -65.62761 -65.59477 -65.45603 -65.30473 -65.29257  
## ---------------------------------------------------------------------------  
## fips   
## n missing distinct Info Mean Gmd .05 .10   
## 43524 0 3218 1 29610 18155 5099 6073   
## .25 .50 .75 .90 .95   
## 17141 29179 42063 51001 54013   
##   
## lowest : 1001 1003 1005 1007 1009, highest: 72145 72147 72149 72151 72153  
## ---------------------------------------------------------------------------

nrow(zip\_codes)

## [1] 43524

zip\_codes=zip\_codes[,c(1,4,5,6)]##dropping column state as its already a part of gen\_info  
head(zip\_codes,n=3)

## zip latitude longitude fips  
## 1 00210 43.0059 -71.0132 33015  
## 2 00211 43.0059 -71.0132 33015  
## 3 00212 43.0059 -71.0132 33015

zip\_codes$zip=as.numeric(zip\_codes$zip)  
gen\_info=left\_join(gen\_info,zip\_codes)

## Joining, by = "zip"

head(gen\_info,3)

## Provider\_ID Hospital\_Name City State zip  
## 1 10005 MARSHALL MEDICAL CENTER SOUTH BOAZ AL 35957  
## 2 10006 ELIZA COFFEE MEMORIAL HOSPITAL FLORENCE AL 35631  
## 3 10007 MIZELL MEMORIAL HOSPITAL OPP AL 36467  
## Hospital\_Type Hospital\_Ownership  
## 1 Acute Care Hospitals Government - Hospital District or Authority  
## 2 Acute Care Hospitals Government - Hospital District or Authority  
## 3 Acute Care Hospitals Voluntary non-profit - Private  
## Emergency\_Services Meets\_criteria\_for\_meaningful\_use\_of\_EHRs  
## 1 Yes Y  
## 2 Yes Y  
## 3 Yes Y  
## Hospital\_overall\_rating mortality  
## 1 3 Below the National average  
## 2 2 Below the National average  
## 3 2 Same as the National average  
## Safety\_of\_care Readmissions  
## 1 Same as the National average Above the National average  
## 2 Same as the National average Same as the National average  
## 3 <NA> Below the National average  
## Patient\_experience Effectiveness\_of\_care  
## 1 Same as the National average Same as the National average  
## 2 Below the National average Same as the National average  
## 3 Same as the National average Below the National average  
## Timeliness\_of\_care Efficient\_use\_of\_medical\_imaging latitude  
## 1 Above the National average Below the National average 34.20362  
## 2 Above the National average Same as the National average 34.86996  
## 3 Above the National average <NA> 31.28005  
## longitude fips  
## 1 -86.17280 1095  
## 2 -87.70686 1077  
## 3 -86.25484 1039

library("ggcounty")  
data("population")## getting population of cities  
head(population)

## FIPS count  
## 2 01001 53155  
## 3 01003 175791  
## 4 01005 27699  
## 5 01007 22610  
## 6 01009 56692  
## 7 01011 10923

describe(population)

## population   
##   
## 2 Variables 3221 Observations  
## ---------------------------------------------------------------------------  
## FIPS   
## n missing distinct   
## 3221 0 3221   
##   
## lowest : 01001 01003 01005 01007 01009, highest: 72145 72147 72149 72151 72153  
## ---------------------------------------------------------------------------  
## count   
## n missing distinct Info Mean Gmd .05 .10   
## 3221 0 3160 1 95538 142428 2976 5235   
## .25 .50 .75 .90 .95   
## 11377 26076 65217 190346 407435   
##   
## lowest : 41 74 219 241 426  
## highest: 3022468 3751410 3950999 5172848 9758256  
## ---------------------------------------------------------------------------

nrow(population)

## [1] 3221

names(population)=c("fips","population")  
head(population)

## fips population  
## 2 01001 53155  
## 3 01003 175791  
## 4 01005 27699  
## 5 01007 22610  
## 6 01009 56692  
## 7 01011 10923

nrow(table(gen\_info$fips))

## [1] 2499

class(gen\_info$fips)

## [1] "numeric"

class(population$fips)

## [1] "character"

population$fips=as.numeric(population$fips)  
gen\_info=left\_join(gen\_info,population)

## Joining, by = "fips"

names(gen\_info)

## [1] "Provider\_ID"   
## [2] "Hospital\_Name"   
## [3] "City"   
## [4] "State"   
## [5] "zip"   
## [6] "Hospital\_Type"   
## [7] "Hospital\_Ownership"   
## [8] "Emergency\_Services"   
## [9] "Meets\_criteria\_for\_meaningful\_use\_of\_EHRs"  
## [10] "Hospital\_overall\_rating"   
## [11] "mortality"   
## [12] "Safety\_of\_care"   
## [13] "Readmissions"   
## [14] "Patient\_experience"   
## [15] "Effectiveness\_of\_care"   
## [16] "Timeliness\_of\_care"   
## [17] "Efficient\_use\_of\_medical\_imaging"   
## [18] "latitude"   
## [19] "longitude"   
## [20] "fips"   
## [21] "population"

sum(is.na(gen\_info$population)==T)

## [1] 9

sum(is.na(gen\_info$latitude)==T)

## [1] 9

sum(is.na(gen\_info$fips))

## [1] 9

sum(is.na(gen\_info$longitude))

## [1] 9

data("us.cities")##Finding out the Capital City  
head(us.cities)

## name country.etc pop lat long capital  
## 1 Abilene TX TX 113888 32.45 -99.74 0  
## 2 Akron OH OH 206634 41.08 -81.52 0  
## 3 Alameda CA CA 70069 37.77 -122.26 0  
## 4 Albany GA GA 75510 31.58 -84.18 0  
## 5 Albany NY NY 93576 42.67 -73.80 2  
## 6 Albany OR OR 45535 44.62 -123.09 0

describe(us.cities)

## us.cities   
##   
## 6 Variables 1005 Observations  
## ---------------------------------------------------------------------------  
## name   
## n missing distinct   
## 1005 0 1005   
##   
## lowest : Abilene TX Akron OH Alameda CA Albany GA Albany NY   
## highest: Yorba Linda CA Youngstown OH Yuba City CA Yucaipa CA Yuma AZ   
## ---------------------------------------------------------------------------  
## country.etc   
## n missing distinct   
## 1005 0 52   
##   
## lowest : AK AL AR AZ CA, highest: VT WA WI WV WY  
## ---------------------------------------------------------------------------  
## pop   
## n missing distinct Info Mean Gmd .05 .10   
## 1005 0 1000 1 125548 126933 40719 43024   
## .25 .50 .75 .90 .95   
## 50116 66628 103782 197888 349475   
##   
## 0 (249, 0.248), 1e+05 (601, 0.598), 2e+05 (85, 0.085), 3e+05 (19, 0.019),  
## 4e+05 (13, 0.013), 5e+05 (12, 0.012), 6e+05 (10, 0.010), 7e+05 (3, 0.003),  
## 8e+05 (2, 0.002), 9e+05 (2, 0.002), 1200000 (1, 0.001), 1300000 (2,  
## 0.002), 1400000 (1, 0.001), 1500000 (1, 0.001), 2e+06 (1, 0.001), 2800000  
## (1, 0.001), 3900000 (1, 0.001), 8100000 (1, 0.001)  
## ---------------------------------------------------------------------------  
## lat   
## n missing distinct Info Mean Gmd .05 .10   
## 1005 0 724 1 37.5 5.905 27.29 30.12   
## .25 .50 .75 .90 .95   
## 33.87 38.51 41.60 42.97 44.93   
##   
## lowest : 19.70 21.32 25.67 25.70 25.71, highest: 47.92 47.96 48.74 58.30 61.18  
## ---------------------------------------------------------------------------  
## long   
## n missing distinct Info Mean Gmd .05 .10   
## 1005 0 839 1 -95.1 20.01 -122.16 -121.27   
## .25 .50 .75 .90 .95   
## -116.40 -90.21 -80.30 -74.04 -71.57   
##   
## lowest : -157.80 -155.09 -149.19 -134.42 -123.28  
## highest: -70.30 -70.28 -70.20 -70.08 -69.77  
## ---------------------------------------------------------------------------  
## capital   
## n missing distinct Info Mean Gmd   
## 1005 0 2 0.142 0.0995 0.1893   
##   
## Value 0 2  
## Frequency 955 50  
## Proportion 0.95 0.05  
## ---------------------------------------------------------------------------

table(us.cities$capital)

##   
## 0 2   
## 955 50

filter(us.cities,capital==2)## the cities with Capital=2 are the capital cities Verified on the internet

## name country.etc pop lat long capital  
## 1 Albany NY NY 93576 42.67 -73.80 2  
## 2 Annapolis MD MD 36300 38.98 -76.49 2  
## 3 Atlanta GA GA 424096 33.76 -84.42 2  
## 4 Augusta ME ME 18626 44.32 -69.77 2  
## 5 Austin TX TX 683404 30.31 -97.75 2  
## 6 Baton Rouge LA LA 222217 30.45 -91.13 2  
## 7 Bismarck ND ND 56927 46.81 -100.77 2  
## 8 Boise ID ID 193628 43.61 -116.23 2  
## 9 Boston MA MA 567759 42.34 -71.02 2  
## 10 Carson City NV NV 58350 39.15 -119.74 2  
## 11 Charleston WV WV 49804 38.35 -81.63 2  
## 12 Cheyenne WY WY 55833 41.15 -104.79 2  
## 13 Columbia SC SC 118020 34.04 -80.89 2  
## 14 Columbus OH OH 741677 39.99 -82.99 2  
## 15 Concord NH NH 42967 43.23 -71.56 2  
## 16 Denver CO CO 556575 39.77 -104.87 2  
## 17 Des Moines IA IA 192050 41.58 -93.62 2  
## 18 Dover DE DE 34288 39.16 -75.53 2  
## 19 Frankfort KY KY 27210 38.20 -84.86 2  
## 20 Harrisburg PA PA 47576 40.28 -76.88 2  
## 21 Hartford CT CT 123836 41.77 -72.68 2  
## 22 Helena MT MT 27383 46.60 -112.03 2  
## 23 Honolulu HI HI 386345 21.32 -157.80 2  
## 24 Indianapolis IN IN 771725 39.78 -86.15 2  
## 25 Jackson MS MS 175085 32.32 -90.21 2  
## 26 Jefferson City MO MO 39062 38.57 -92.19 2  
## 27 Juneau AK AK 31187 58.30 -134.42 2  
## 28 Lansing MI MI 117236 42.71 -84.55 2  
## 29 Lincoln NE NE 245301 40.82 -96.69 2  
## 30 Little Rock AR AR 184323 34.72 -92.35 2  
## 31 Madison WI WI 227642 43.08 -89.39 2  
## 32 Montgomery AL AL 197653 32.35 -86.28 2  
## 33 Montpelier VT VT 8003 44.26 -72.57 2  
## 34 Nashville TN TN 523547 36.17 -86.78 2  
## 35 Oklahoma City OK OK 538141 35.47 -97.51 2  
## 36 Olympia WA WA 45403 47.04 -122.89 2  
## 37 Phoenix AZ AZ 1450884 33.54 -112.07 2  
## 38 Pierre SD SD 14052 44.37 -100.34 2  
## 39 Providence RI RI 178295 41.82 -71.42 2  
## 40 Raleigh NC NC 350822 35.82 -78.66 2  
## 41 Richmond VA VA 189498 37.53 -77.47 2  
## 42 Sacramento CA CA 480392 38.57 -121.47 2  
## 43 Saint Paul MN MN 272469 44.95 -93.10 2  
## 44 Salem OR OR 148942 44.92 -123.02 2  
## 45 Salt Lake City UT UT 177318 40.78 -111.93 2  
## 46 Santa Fe NM NM 70598 35.68 -105.95 2  
## 47 Springfield IL IL 116725 39.78 -89.64 2  
## 48 Tallahassee FL FL 154114 30.46 -84.28 2  
## 49 Topeka KS KS 121229 39.04 -95.69 2  
## 50 Trenton NJ NJ 84653 40.22 -74.76 2

nrow(us.cities)

## [1] 1005

us.cities=us.cities[,c(1,6)]  
gen\_info=mutate(gen\_info,name=paste(City,State))  
gen\_info=mutate(gen\_info,name=paste(City,State))  
gen\_info$name=tolower(gen\_info$name)  
us.cities$name=tolower(us.cities$name)  
   
class(us.cities$name)

## [1] "character"

class(gen\_info$name)

## [1] "character"

gen\_info=left\_join(gen\_info,us.cities)

## Joining, by = "name"

table(gen\_info$capital,useNA = "always")

##   
## 0 2 <NA>   
## 1398 206 3202

gen\_info$capital[is.na(gen\_info$capital)]=0##assigning 0 to NA's as they are not capital cities  
class(gen\_info$capital)

## [1] "numeric"

gen\_info$capital=as.factor(gen\_info$capital)  
  
library(USAboundaries)  
state\_boundaries <- us\_states()##addiing land Area of the cities  
sapply(state\_boundaries@data, class)

## statefp statens affgeoid geoid stusps name   
## "character" "character" "character" "character" "character" "character"   
## lsad aland awater   
## "character" "character" "character"

head(state\_boundaries@data, n = 2)

## statefp statens affgeoid geoid stusps name lsad  
## 0 06 01779778 0400000US06 06 CA California 00  
## 1 11 01702382 0400000US11 11 DC District of Columbia 00  
## aland awater  
## 0 403483823181 20483271881  
## 1 158350578 18633500

describe(state\_boundaries@data)

## state\_boundaries@data   
##   
## 9 Variables 52 Observations  
## ---------------------------------------------------------------------------  
## statefp   
## n missing distinct   
## 52 0 52   
##   
## lowest : 01 02 04 05 06, highest: 53 54 55 56 72  
## ---------------------------------------------------------------------------  
## statens   
## n missing distinct   
## 52 0 52   
##   
## lowest : 00068085 00294478 00448508 00481813 00606926  
## highest: 01779806 01779807 01779808 01785533 01785534  
## ---------------------------------------------------------------------------  
## affgeoid   
## n missing distinct   
## 52 0 52   
##   
## lowest : 0400000US01 0400000US02 0400000US04 0400000US05 0400000US06  
## highest: 0400000US53 0400000US54 0400000US55 0400000US56 0400000US72  
## ---------------------------------------------------------------------------  
## geoid   
## n missing distinct   
## 52 0 52   
##   
## lowest : 01 02 04 05 06, highest: 53 54 55 56 72  
## ---------------------------------------------------------------------------  
## stusps   
## n missing distinct   
## 52 0 52   
##   
## lowest : AK AL AR AZ CA, highest: VT WA WI WV WY  
## ---------------------------------------------------------------------------  
## name   
## n missing distinct   
## 52 0 52   
##   
## lowest : Alabama Alaska Arizona Arkansas California   
## highest: Virginia Washington West Virginia Wisconsin Wyoming   
## ---------------------------------------------------------------------------  
## lsad   
## n missing distinct value   
## 52 0 1 00   
##   
## Value 0  
## Frequency 52  
## Proportion 1  
## ---------------------------------------------------------------------------  
## aland   
## n missing distinct   
## 52 0 52   
##   
## lowest : 102262419204 102283343474 105831263791 106800130794 111901043977  
## highest: 676601887070 77857913931 79885244456 8867837128 92790411854   
## ---------------------------------------------------------------------------  
## awater   
## n missing distinct   
## 52 0 52   
##   
## lowest : 1026252314 10266413579 1027790845 1034369068 104031344385  
## highest: 6983455225 7000211755 7131805598 755712514 8504608966   
## ---------------------------------------------------------------------------

state\_area=state\_boundaries@data[,c("geoid","stusps","aland")]  
class(state\_area)

## [1] "data.frame"

state\_area=data.frame(state\_area)  
names(state\_area)=c("geoid","State","land\_area")  
head(state\_area)

## geoid State land\_area  
## 0 06 CA 403483823181  
## 1 11 DC 158350578  
## 2 12 FL 138903200855  
## 3 13 GA 148963503399  
## 4 16 ID 214045425549  
## 5 17 IL 143794747023

state\_area$State=as.character(state\_area$State)  
gen\_info$State=as.character(gen\_info$State)  
gen\_info=left\_join(gen\_info,state\_area)

## Joining, by = "State"

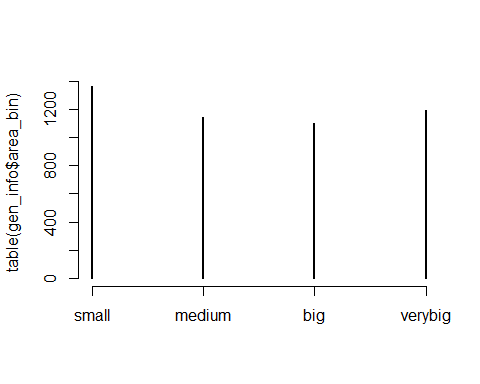
head(gen\_info,3)

## Provider\_ID Hospital\_Name City State zip  
## 1 10005 MARSHALL MEDICAL CENTER SOUTH BOAZ AL 35957  
## 2 10006 ELIZA COFFEE MEMORIAL HOSPITAL FLORENCE AL 35631  
## 3 10007 MIZELL MEMORIAL HOSPITAL OPP AL 36467  
## Hospital\_Type Hospital\_Ownership  
## 1 Acute Care Hospitals Government - Hospital District or Authority  
## 2 Acute Care Hospitals Government - Hospital District or Authority  
## 3 Acute Care Hospitals Voluntary non-profit - Private  
## Emergency\_Services Meets\_criteria\_for\_meaningful\_use\_of\_EHRs  
## 1 Yes Y  
## 2 Yes Y  
## 3 Yes Y  
## Hospital\_overall\_rating mortality  
## 1 3 Below the National average  
## 2 2 Below the National average  
## 3 2 Same as the National average  
## Safety\_of\_care Readmissions  
## 1 Same as the National average Above the National average  
## 2 Same as the National average Same as the National average  
## 3 <NA> Below the National average  
## Patient\_experience Effectiveness\_of\_care  
## 1 Same as the National average Same as the National average  
## 2 Below the National average Same as the National average  
## 3 Same as the National average Below the National average  
## Timeliness\_of\_care Efficient\_use\_of\_medical\_imaging latitude  
## 1 Above the National average Below the National average 34.20362  
## 2 Above the National average Same as the National average 34.86996  
## 3 Above the National average <NA> 31.28005  
## longitude fips population name capital geoid land\_area  
## 1 -86.17280 1095 90867 boaz al 0 01 131172403111  
## 2 -87.70686 1077 91762 florence al 0 01 131172403111  
## 3 -86.25484 1039 37608 opp al 0 01 131172403111

nrow(table(gen\_info$geoid))

## [1] 52

gen\_info=filter(gen\_info,is.na(gen\_info$land\_area)==FALSE)  
gen\_info$land\_area=as.numeric(gen\_info$land\_area)  
q <- quantile(gen\_info$land\_area, c(0, .25, .50, .75, 1))  
gen\_info = data.frame(gen\_info, area\_bin=cut(gen\_info$land\_area, q,labels = c("small","medium","big","verybig"), ordered=TRUE, include.lowest=TRUE))  
  
plot(table(gen\_info$area\_bin))



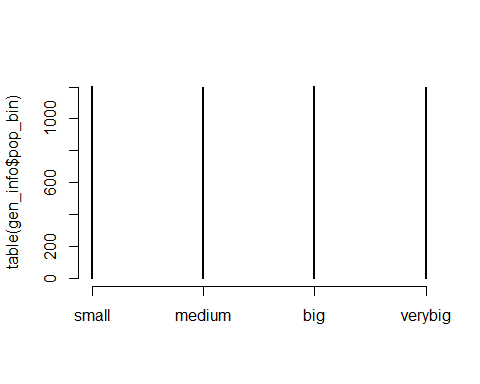
gen\_info$area\_bin=as.factor(gen\_info$area\_bin)  
  
sum(is.na(gen\_info$population))

## [1] 3

gen\_info=filter(gen\_info,is.na(gen\_info$population)==FALSE)  
gen\_info$population=as.numeric(gen\_info$population)  
q <- quantile(gen\_info$population, c(0, .25, .50, .75, 1))  
class(gen\_info$pop\_bin)

## [1] "NULL"

gen\_info = data.frame(gen\_info, pop\_bin=cut(gen\_info$population, q,labels=c(  
 "small","medium","big","verybig"),ordered=T ,include.lowest=TRUE))  
plot(table(gen\_info$pop\_bin))



gen\_info$pop\_bin=as.factor(gen\_info$pop\_bin)  
  
  
cols <- c("State","City","Hospital\_Type","Hospital\_Ownership",  
 "capital","Meets\_criteria\_for\_meaningful\_use\_of\_EHRs",  
 "Hospital\_overall\_rating","mortality",  
 "Safety\_of\_care",  
 "Readmissions","Patient\_experience",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care",  
 "Efficient\_use\_of\_medical\_imaging")  
gen\_info[cols] <- lapply(gen\_info[cols], as.factor)  
  
factor\_var=c("mortality",  
 "Safety\_of\_care",  
 "Readmissions","Patient\_experience",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care",  
 "Efficient\_use\_of\_medical\_imaging",  
 "Patient\_experience")  
table(gen\_info$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 677 1772 941 81

levels(gen\_info$mortality)

## [1] "Above the National average" "Below the National average"   
## [3] "Same as the National average"

table(gen\_info$capital)

##   
## 0 2   
## 4591 206

gen\_info$capital=factor(gen\_info$capital,labels = c("No","Yes"))  
  
names(gen\_info)

## [1] "Provider\_ID"   
## [2] "Hospital\_Name"   
## [3] "City"   
## [4] "State"   
## [5] "zip"   
## [6] "Hospital\_Type"   
## [7] "Hospital\_Ownership"   
## [8] "Emergency\_Services"   
## [9] "Meets\_criteria\_for\_meaningful\_use\_of\_EHRs"  
## [10] "Hospital\_overall\_rating"   
## [11] "mortality"   
## [12] "Safety\_of\_care"   
## [13] "Readmissions"   
## [14] "Patient\_experience"   
## [15] "Effectiveness\_of\_care"   
## [16] "Timeliness\_of\_care"   
## [17] "Efficient\_use\_of\_medical\_imaging"   
## [18] "latitude"   
## [19] "longitude"   
## [20] "fips"   
## [21] "population"   
## [22] "name"   
## [23] "capital"   
## [24] "geoid"   
## [25] "land\_area"   
## [26] "area\_bin"   
## [27] "pop\_bin"

describe(gen\_info)

## gen\_info   
##   
## 27 Variables 4797 Observations  
## ---------------------------------------------------------------------------  
## Provider\_ID   
## n missing distinct Info Mean Gmd .05 .10   
## 4797 0 4797 1 267724 179246 41306 50569   
## .25 .50 .75 .90 .95   
## 140187 260025 390199 460042 510073   
##   
## lowest : 10005 10006 10007 10008 10011, highest: 670108 670109 670111 670112 670114  
## ---------------------------------------------------------------------------  
## Hospital\_Name   
## n missing distinct   
## 4797 0 4598   
##   
## 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  
## ---------------------------------------------------------------------------  
## City   
## n missing distinct   
## 4797 0 2941   
##   
## lowest : ABBEVILLE ABERDEEN ABILENE ABINGDON ABINGTON   
## highest: ZANESVILLE ZEELAND ZEPHYRHILLS ZION ZUNI   
## ---------------------------------------------------------------------------  
## State   
## n missing distinct   
## 4797 0 52   
##   
## lowest : AK AL AR AZ CA, highest: VT WA WI WV WY  
## ---------------------------------------------------------------------------  
## zip   
## n missing distinct Info Mean Gmd .05 .10   
## 4797 0 4403 1 53938 30901 7108 15216   
## .25 .50 .75 .90 .95   
## 33013 55379 76012 91314 95523   
##   
## lowest : 603 613 614 631 641, highest: 99801 99833 99835 99901 99929  
## ---------------------------------------------------------------------------  
## Hospital\_Type   
## n missing distinct   
## 4797 0 3   
##   
## Value Acute Care Hospitals Childrens  
## Frequency 3360 99  
## Proportion 0.700 0.021  
##   
## Value Critical Access Hospitals  
## Frequency 1338  
## Proportion 0.279  
## ---------------------------------------------------------------------------  
## Hospital\_Ownership   
## n missing distinct   
## 4797 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   
## 4797 0 2   
##   
## Value No Yes  
## Frequency 295 4502  
## Proportion 0.061 0.939  
## ---------------------------------------------------------------------------  
## Meets\_criteria\_for\_meaningful\_use\_of\_EHRs   
## n missing distinct value   
## 4367 430 1 Y   
##   
## Value Y  
## Frequency 4367  
## Proportion 1  
## ---------------------------------------------------------------------------  
## Hospital\_overall\_rating   
## n missing distinct   
## 3579 1218 5   
##   
## Value 1 2 3 4 5  
## Frequency 108 677 1772 941 81  
## Proportion 0.030 0.189 0.495 0.263 0.023  
## ---------------------------------------------------------------------------  
## mortality   
## n missing distinct   
## 3471 1326 3   
##   
## Value Above the National average Below the National average  
## Frequency 400 340  
## Proportion 0.115 0.098  
##   
## Value Same as the National average  
## Frequency 2731  
## Proportion 0.787  
## ---------------------------------------------------------------------------  
## Safety\_of\_care   
## n missing distinct   
## 2650 2147 3   
##   
## Value Above the National average Below the National average  
## Frequency 785 665  
## Proportion 0.296 0.251  
##   
## Value Same as the National average  
## Frequency 1200  
## Proportion 0.453  
## ---------------------------------------------------------------------------  
## Readmissions   
## n missing distinct   
## 3805 992 3   
##   
## Value Above the National average Below the National average  
## Frequency 810 868  
## Proportion 0.213 0.228  
##   
## Value Same as the National average  
## Frequency 2127  
## Proportion 0.559  
## ---------------------------------------------------------------------------  
## Patient\_experience   
## n missing distinct   
## 3450 1347 3   
##   
## Value Above the National average Below the National average  
## Frequency 1216 1081  
## Proportion 0.352 0.313  
##   
## Value Same as the National average  
## Frequency 1153  
## Proportion 0.334  
## ---------------------------------------------------------------------------  
## Effectiveness\_of\_care   
## n missing distinct   
## 3622 1175 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 3247  
## Proportion 0.896  
## ---------------------------------------------------------------------------  
## Timeliness\_of\_care   
## n missing distinct   
## 3563 1234 3   
##   
## Value Above the National average Below the National average  
## Frequency 1096 905  
## Proportion 0.308 0.254  
##   
## Value Same as the National average  
## Frequency 1562  
## Proportion 0.438  
## ---------------------------------------------------------------------------  
## Efficient\_use\_of\_medical\_imaging   
## n missing distinct   
## 2787 2010 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 2056  
## Proportion 0.738  
## ---------------------------------------------------------------------------  
## latitude   
## n missing distinct Info Mean Gmd .05 .10   
## 4797 0 4386 1 37.89 6.325 29.01 30.57   
## .25 .50 .75 .90 .95   
## 33.92 38.61 41.76 44.53 46.34   
##   
## lowest : 17.98729 18.00430 18.01882 18.01933 18.06547  
## highest: 61.59820 64.53262 64.83507 66.89719 71.29953  
## ---------------------------------------------------------------------------  
## longitude   
## n missing distinct Info Mean Gmd .05 .10   
## 4797 0 4383 1 -92.89 16.37 -121.45 -117.70   
## .25 .50 .75 .90 .95   
## -98.61 -90.26 -82.38 -75.62 -73.47   
##   
## lowest : -165.10883 -162.58544 -161.88006 -159.71870 -159.37746  
## highest: -66.03806 -65.97249 -65.82269 -65.78850 -65.66116  
## ---------------------------------------------------------------------------  
## fips   
## n missing distinct Info Mean Gmd .05 .10   
## 4797 0 2499 1 29418 18518 5112 6067   
## .25 .50 .75 .90 .95   
## 17075 29023 42085 49011 54030   
##   
## lowest : 1001 1003 1005 1007 1009, highest: 72125 72127 72141 72145 72153  
## ---------------------------------------------------------------------------  
## population   
## n missing distinct Info Mean Gmd .05 .10   
## 4797 0 2477 1 618976 964574 6899 11050   
## .25 .50 .75 .90 .95   
## 27575 101994 540583 1504950 2965525   
##   
## lowest : 1089 1224 1289 1294 1434  
## highest: 3022468 3751410 3950999 5172848 9758256  
## ---------------------------------------------------------------------------  
## name   
## n missing distinct   
## 4797 0 3724   
##   
## lowest : abbeville la abbeville sc aberdeen ms aberdeen sd aberdeen wa   
## highest: zanesville oh zeeland mi zephyrhills fl zion il zuni nm   
## ---------------------------------------------------------------------------  
## capital   
## n missing distinct   
## 4797 0 2   
##   
## Value No Yes  
## Frequency 4591 206  
## Proportion 0.957 0.043  
## ---------------------------------------------------------------------------  
## geoid   
## n missing distinct   
## 4797 0 52   
##   
## lowest : 01 02 04 05 06, highest: 53 54 55 56 72  
## ---------------------------------------------------------------------------  
## land\_area   
## n missing distinct Info Mean Gmd .05   
## 4797 0 52 0.999 2.144e+11 1.737e+11 2.020e+10   
## .10 .25 .50 .75 .90 .95   
## 7.989e+10 1.159e+11 1.447e+11 2.129e+11 4.035e+11 6.766e+11   
##   
## lowest : 1.583506e+08 2.677751e+09 5.047094e+09 8.867837e+09 1.254240e+10  
## highest: 3.141614e+11 3.769638e+11 4.034838e+11 6.766019e+11 1.477849e+12  
## ---------------------------------------------------------------------------  
## area\_bin   
## n missing distinct   
## 4797 0 4   
##   
## Value small medium big verybig  
## Frequency 1361 1144 1101 1191  
## Proportion 0.284 0.238 0.230 0.248  
## ---------------------------------------------------------------------------  
## pop\_bin   
## n missing distinct   
## 4797 0 4   
##   
## Value small medium big verybig  
## Frequency 1201 1198 1204 1194  
## Proportion 0.250 0.250 0.251 0.249  
## ---------------------------------------------------------------------------

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

##   
## AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS  
## 1 0 0 2 3 9 0 0 5 0 7 4 0 0 0 4 1 0  
## 2 1 15 14 12 91 1 9 1 0 66 19 2 4 0 29 8 3  
## 3 8 47 31 27 133 27 18 1 2 76 69 6 37 10 73 43 34  
## 4 0 13 7 11 51 22 1 0 4 20 16 4 28 8 48 50 19  
## 5 0 3 1 2 5 1 0 0 0 1 2 1 2 1 5 6 2  
## <NA> 13 10 19 26 51 27 3 1 1 16 24 10 45 22 21 12 79  
##   
## KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV  
## 1 2 1 1 0 0 4 0 2 2 0 0 1 0 0 6 1 6  
## 2 15 10 5 0 2 21 1 18 20 1 16 0 2 0 27 10 5  
## 3 49 48 34 0 15 44 36 44 31 13 53 10 33 13 21 17 13  
## 4 17 20 16 0 14 41 41 21 10 9 24 5 12 12 10 3 2  
## 5 0 1 1 0 1 5 1 1 0 0 1 0 2 0 0 0 0  
## <NA> 11 42 7 49 1 16 52 28 33 38 12 28 41 1 2 10 7  
##   
## NY OH OK OR PA PR RI SC SD TN TX UT VA VT WA WI WV  
## 1 34 2 1 1 3 0 0 1 0 2 2 0 1 0 0 0 0  
## 2 55 15 10 4 37 8 1 12 0 22 37 1 16 2 13 3 10  
## 3 49 56 52 33 66 1 7 19 4 57 135 14 41 5 41 40 26  
## 4 15 62 16 14 40 0 2 17 8 14 81 14 19 5 10 61 1  
## 5 1 5 4 0 2 0 1 4 2 0 11 0 0 0 2 4 0  
## <NA> 18 28 40 8 22 41 0 7 46 14 141 17 10 2 25 18 12  
##   
## WY <NA>  
## 1 0 0  
## 2 3 0  
## 3 10 0  
## 4 3 0  
## 5 0 0  
## <NA> 11 0

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

##   
## Acute Care Hospitals Childrens Critical Access Hospitals <NA>  
## 1 107 0 1 0  
## 2 653 0 24 0  
## 3 1417 0 355 0  
## 4 726 0 215 0  
## 5 79 0 2 0  
## <NA> 378 99 741 0

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

##   
## Government - Federal Government - Hospital District or Authority  
## 1 0 8  
## 2 5 60  
## 3 8 198  
## 4 4 76  
## 5 0 5  
## <NA> 29 217  
##   
## Government - Local Government - State Physician Proprietary Tribal  
## 1 13 8 0 26 0  
## 2 47 10 2 174 1  
## 3 120 19 5 301 1  
## 4 59 8 4 112 0  
## 5 2 1 7 9 0  
## <NA> 163 20 46 173 6  
##   
## Voluntary non-profit - Church Voluntary non-profit - Other  
## 1 4 9  
## 2 55 57  
## 3 145 192  
## 4 93 113  
## 5 14 11  
## <NA> 32 91  
##   
## Voluntary non-profit - Private <NA>  
## 1 40 0  
## 2 266 0  
## 3 783 0  
## 4 472 0  
## 5 32 0  
## <NA> 441 0

## Analysing Data

## visualizing the density of Hospitals on US map based on different variables

##visualizing the density of Hospitals on US using different variables  
  
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 = 9),  
 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))  
  
theme\_set(theme\_bw(16))  
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "bottomleft")

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false

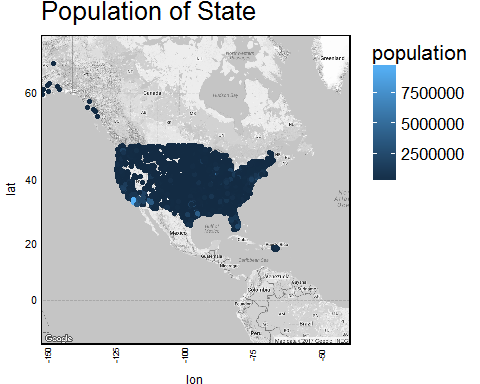
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = population), data = gen\_info,title="Population")+hospital.theme+  
 ggtitle("Population of State")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Ignoring unknown parameters: title

## Warning: Removed 29 rows containing missing values (geom\_point).



StateMap <- qmap("USA", zoom = 4, color = "bw", legend = "bottomleft")

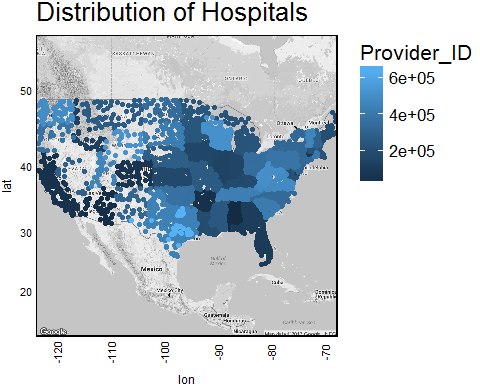
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=4&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour =Provider\_ID), data = gen\_info,title="Population")+hospital.theme+  
 ggtitle("Distribution of Hospitals")

## Warning: Ignoring unknown parameters: title

## Warning: Removed 115 rows containing missing values (geom\_point).



StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "bottomleft")

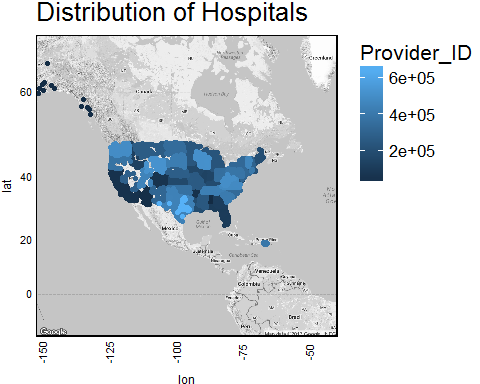
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour =Provider\_ID), data = gen\_info,title="Population")+hospital.theme+  
 ggtitle("Distribution of Hospitals")

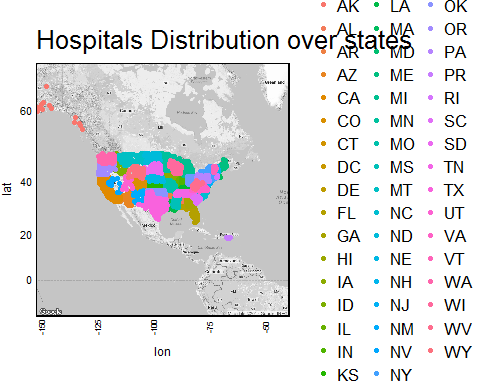
## Warning: Ignoring unknown parameters: title

## Warning: Removed 29 rows containing missing values (geom\_point).



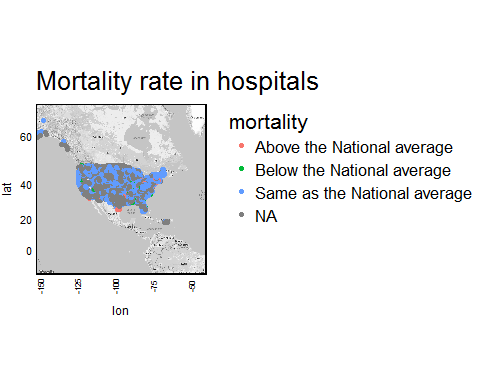
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = State), data = gen\_info)+hospital.theme+  
 ggtitle("Hospitals Distribution over states")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



StateMap + geom\_point(aes(x = longitude, y = latitude, colour = mortality), data = gen\_info)+hospital.theme+  
 ggtitle("Mortality rate in hospitals")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



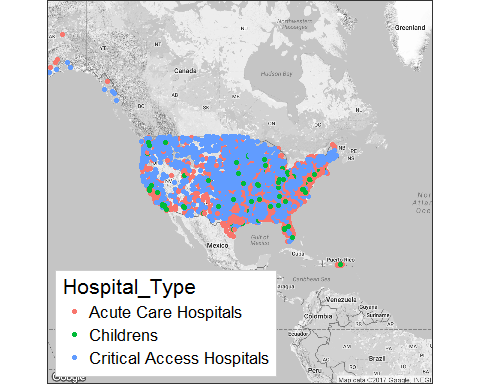
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "bottomleft")

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Hospital\_Type), data = gen\_info,title="Types of Hospitals")

## Warning: Ignoring unknown parameters: title  
  
## Warning: Removed 29 rows containing missing values (geom\_point).



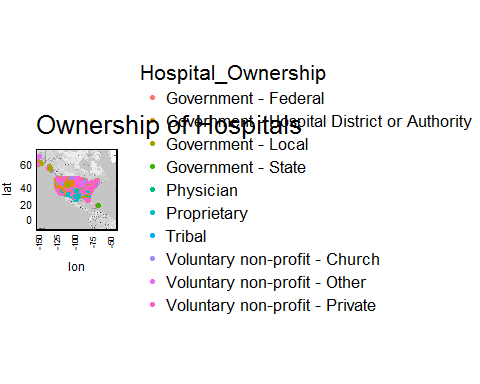
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Hospital\_Ownership), data = gen\_info)+hospital.theme+  
 ggtitle("Ownership of Hospitals")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



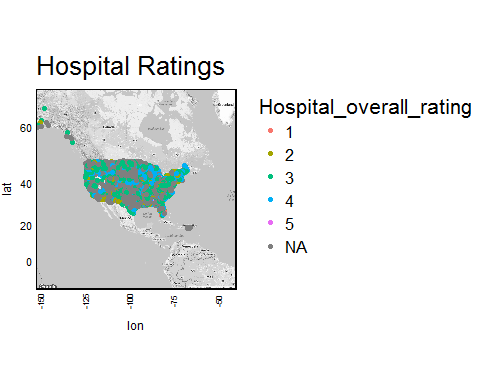
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Hospital\_overall\_rating), data = gen\_info)+hospital.theme+  
 ggtitle("Hospital Ratings ")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



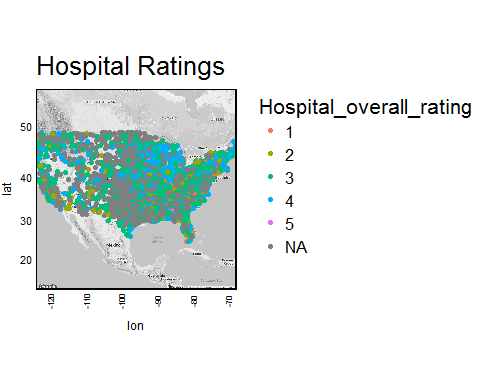
StateMap <- qmap("USA", zoom = 4, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=4&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Hospital\_overall\_rating), data = gen\_info)+hospital.theme+  
 ggtitle("Hospital Ratings ")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 115 rows containing missing values (geom\_point).



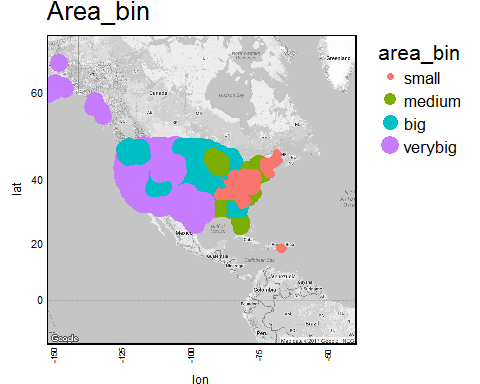
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = area\_bin,size=area\_bin), data = gen\_info)+hospital.theme+  
 ggtitle("Area\_bin")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Using size for a discrete variable is not advised.  
  
## Warning: Removed 29 rows containing missing values (geom\_point).



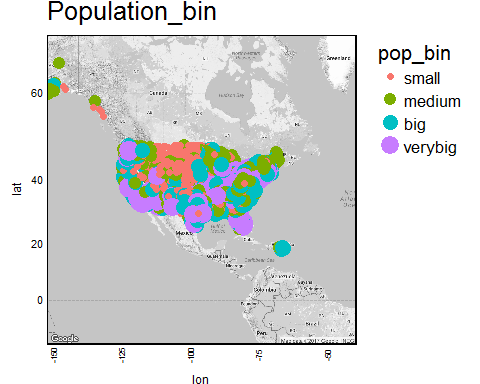
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = pop\_bin,size=pop\_bin), data = gen\_info)+hospital.theme+  
 ggtitle("Population\_bin")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Using size for a discrete variable is not advised.  
  
## Warning: Removed 29 rows containing missing values (geom\_point).



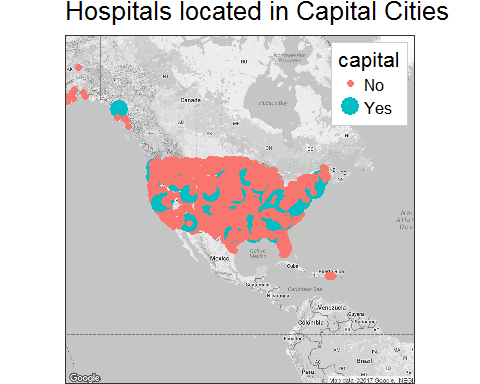
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = capital, size = capital), data = gen\_info)+  
 ggtitle("Hospitals located in Capital Cities")

## Warning: Using size for a discrete variable is not advised.  
  
## Warning: Removed 29 rows containing missing values (geom\_point).



StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

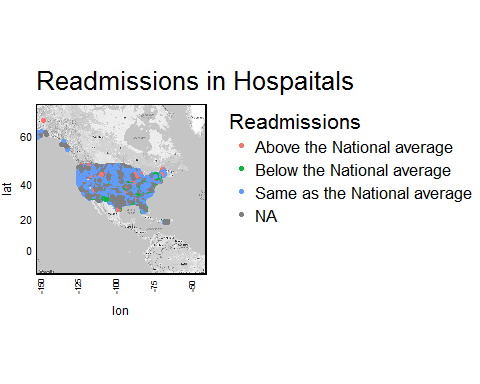
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Emergency\_Services), data = gen\_info)+  
 ggtitle("Emergency services in Hospitals")

## Warning: Removed 29 rows containing missing values (geom\_point).



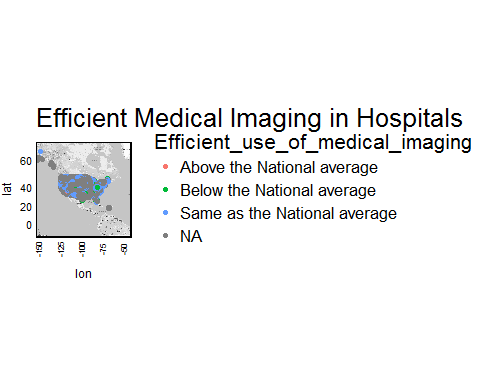
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Readmissions), data = gen\_info)+hospital.theme+  
 ggtitle("Readmissions in Hospaitals")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



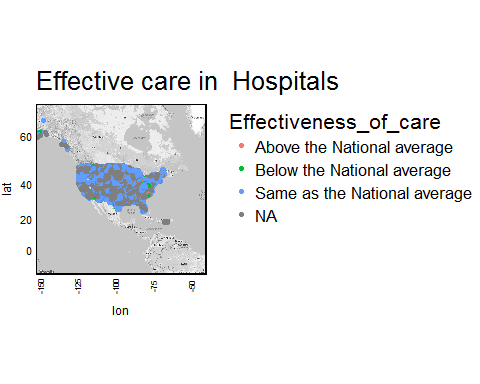
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Efficient\_use\_of\_medical\_imaging), data = gen\_info)+hospital.theme+  
 ggtitle("Efficient Medical Imaging in Hospitals")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



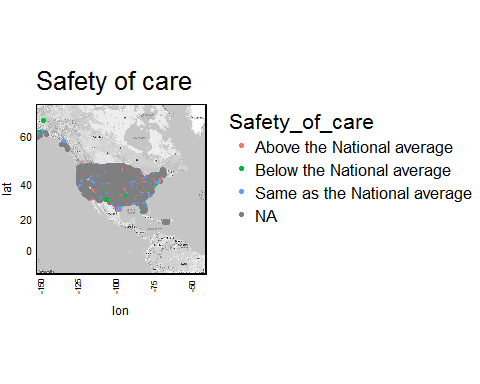
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Effectiveness\_of\_care), data = gen\_info)+hospital.theme+  
 ggtitle("Effective care in Hospitals")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



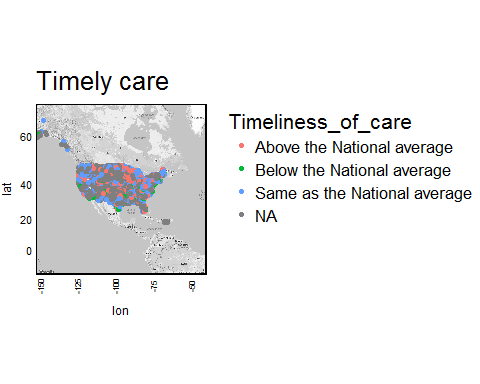
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Safety\_of\_care), data = gen\_info)+hospital.theme+hospital.theme+  
 ggtitle("Safety of care")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



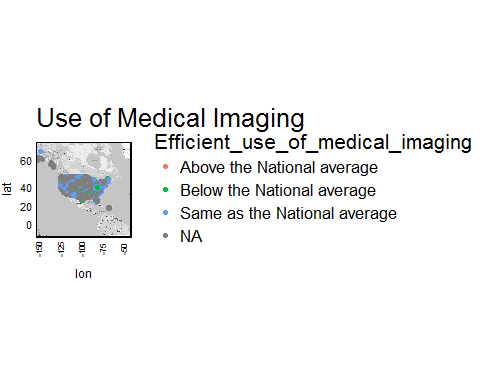
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Timeliness\_of\_care), data = gen\_info)+hospital.theme+hospital.theme+  
 ggtitle("Timely care")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



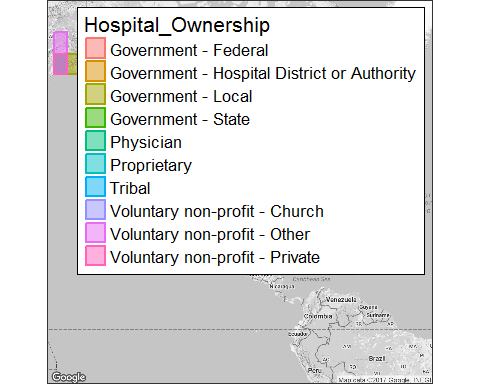
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Efficient\_use\_of\_medical\_imaging), data = gen\_info)+hospital.theme+hospital.theme+  
 ggtitle("Use of Medical Imaging")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 29 rows containing missing values (geom\_point).



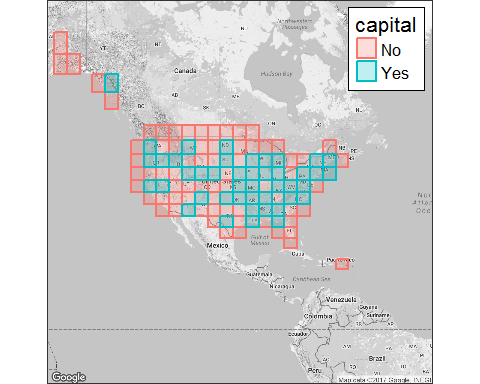
StateMap + stat\_bin2d( aes(x = longitude, y = latitude,  
 colour = Hospital\_Ownership, fill = Hospital\_Ownership),   
 size = 1, bins = 30, alpha = 1/2, data = gen\_info)+   
 theme(legend.background = element\_rect(colour = "black"))

## Warning: Removed 29 rows containing non-finite values (stat\_bin2d).



StateMap + stat\_bin2d( aes(x = longitude, y = latitude,  
 colour = capital, fill = capital),   
 size = 1, bins = 30, alpha = 1/4, data = gen\_info)+   
 theme(legend.background = element\_rect(colour = "black"))

## Warning: Removed 29 rows containing non-finite values (stat\_bin2d).

 ##Analyzing dataset ##Univariate Analysis

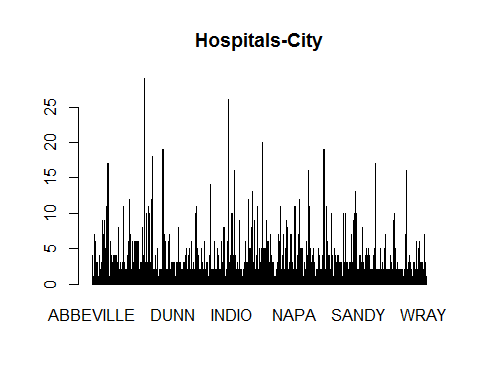
palette=getcolors(6, "rgb")  
table(gen\_info$State)

##   
## AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY   
## 22 88 74 81 340 78 31 8 7 186 134 23 116 41 180 120 137 94   
## LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH   
## 122 64 49 33 131 131 114 96 61 106 44 90 26 66 41 33 172 168   
## OK OR PA PR RI SC SD TN TX UT VA VT WA WI WV WY   
## 123 60 170 50 11 60 60 109 407 46 87 14 91 126 49 27

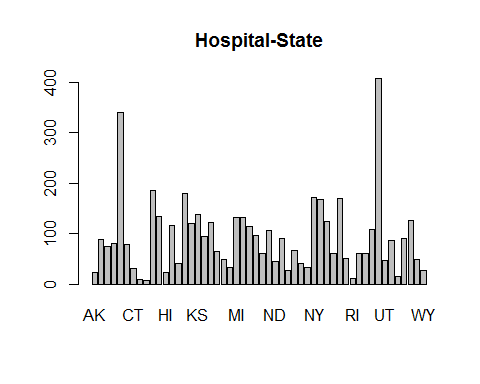
max(table(gen\_info$State))

## [1] 407

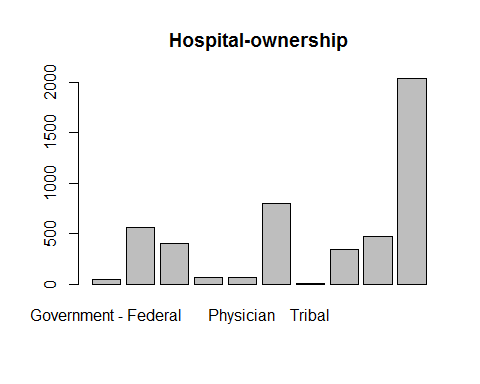
barplot(table(gen\_info$City),main="Hospitals-City")



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



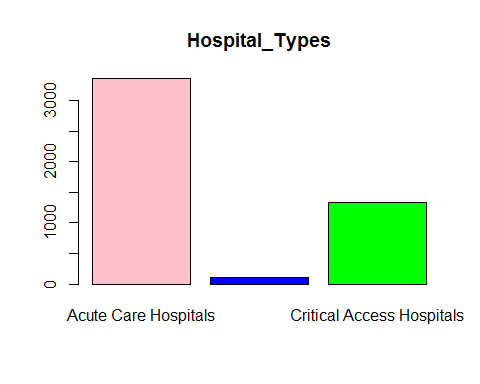
barplot(table(gen\_info$Hospital\_Ownership),main="Hospital-ownership")



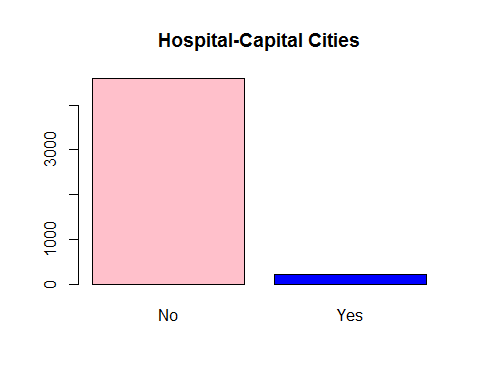
table(gen\_info$Hospital\_Ownership)

##   
## Government - Federal   
## 46   
## Government - Hospital District or Authority   
## 564   
## Government - Local   
## 404   
## Government - State   
## 66   
## Physician   
## 64   
## Proprietary   
## 795   
## Tribal   
## 8   
## Voluntary non-profit - Church   
## 343   
## Voluntary non-profit - Other   
## 473   
## Voluntary non-profit - Private   
## 2034

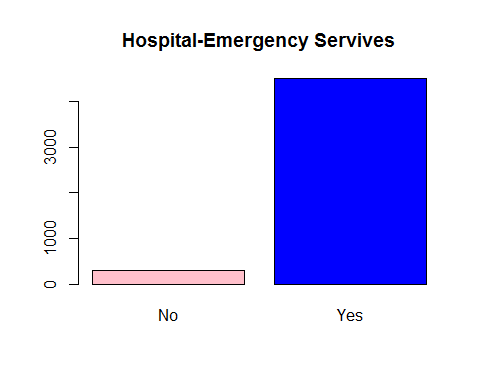
barplot(table(gen\_info$Hospital\_Type),main="Hospital\_Types",col=c("pink","blue","green"))



barplot(table(gen\_info$capital),main="Hospital-Capital Cities",col=c("pink","blue","green"))



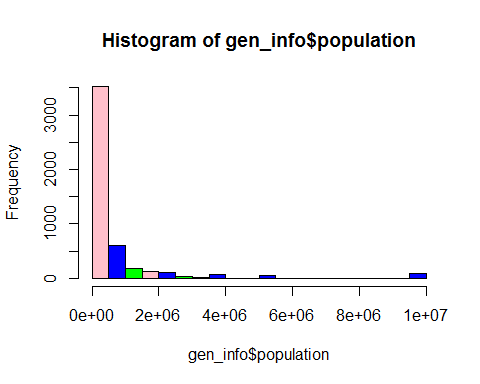
barplot(table(gen\_info$Emergency\_Services),main="Hospital-Emergency Servives",col=c("pink","blue","green"))



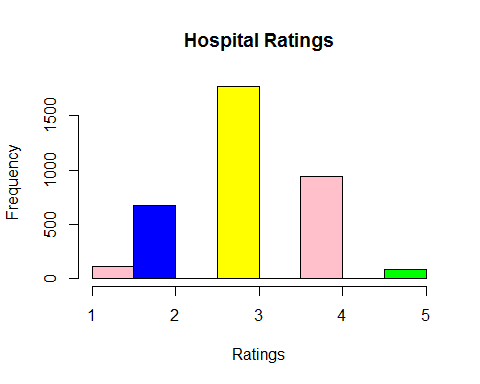
table(gen\_info$Emergency\_Services,useNA = 'always')

##   
## No Yes <NA>   
## 295 4502 0

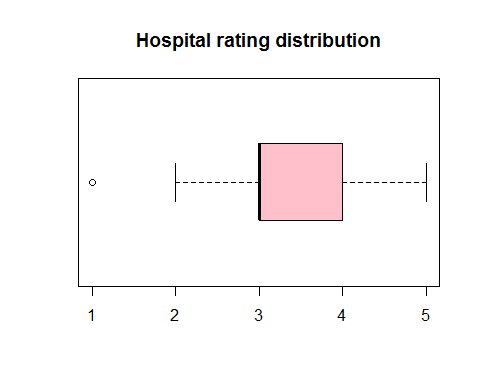
hist(gen\_info$population,col=c("pink","blue","green"))



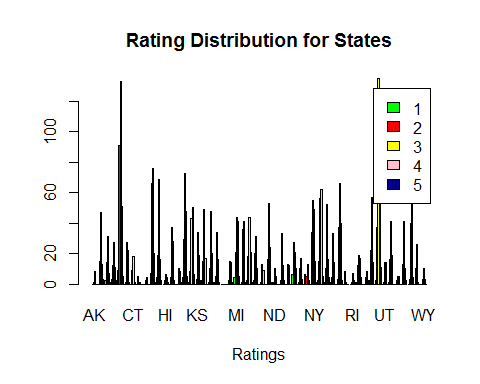
gen\_info$Hospital\_overall\_rating=as.numeric(gen\_info$Hospital\_overall\_rating)  
hist(gen\_info$Hospital\_overall\_rating,col=c("pink","blue","green","yellow","grey"),main="Hospital Ratings",xlab="Ratings")



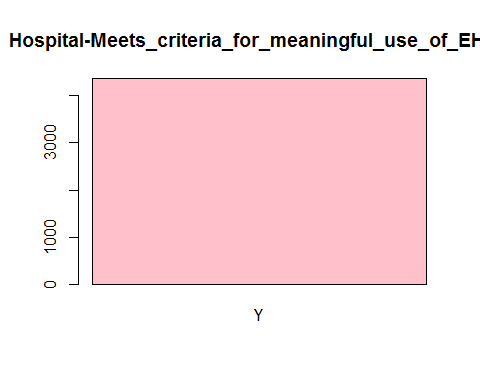
boxplot(gen\_info$Hospital\_overall\_rating,horizontal=T,main = "Hospital rating distribution",col=c("pink","blue","green","yellow","grey"))



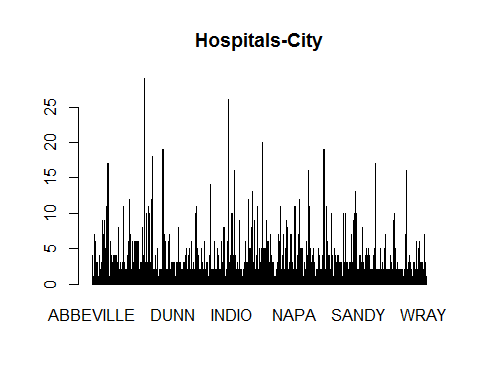
counts <- table(gen\_info$Hospital\_overall\_rating, gen\_info$State)  
barplot(counts, main="Rating Distribution for States",  
xlab="Ratings", col=c("green","red","yellow","pink","darkblue"),  
 legend = rownames(counts), beside=TRUE)



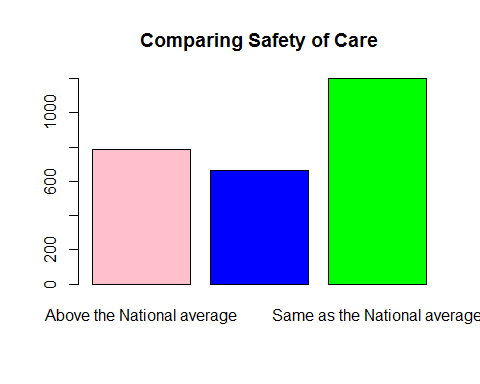
barplot(table(gen\_info$Meets\_criteria\_for\_meaningful\_use\_of\_EHRs),  
 main="Hospital-Meets\_criteria\_for\_meaningful\_use\_of\_EHRs",col=c("pink","blue","green","yellow","grey"))



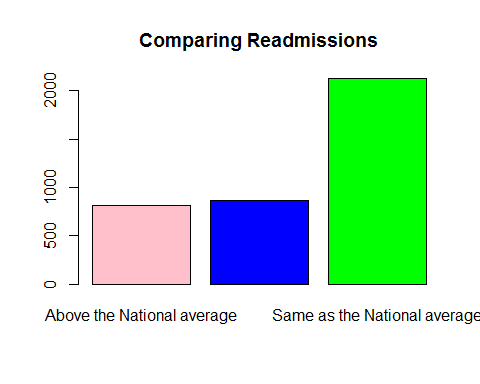
barplot(table(gen\_info$City),main="Hospitals-City",col=c("pink","blue","green","yellow","grey"))



barplot(table(gen\_info$Safety\_of\_care),main = "Comparing Safety of Care",col=c("pink","blue","green","yellow","grey"))



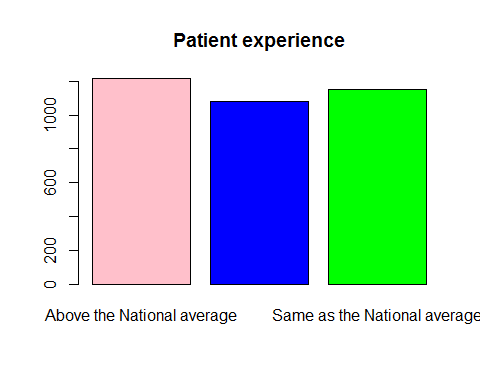
barplot(table(gen\_info$Readmissions),main = "Comparing Readmissions",col=c("pink","blue","green","yellow","grey"))



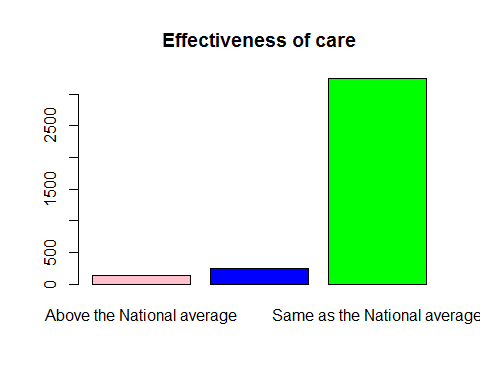
table(gen\_info$Readmissions)

##   
## Above the National average Below the National average   
## 810 868   
## Same as the National average   
## 2127

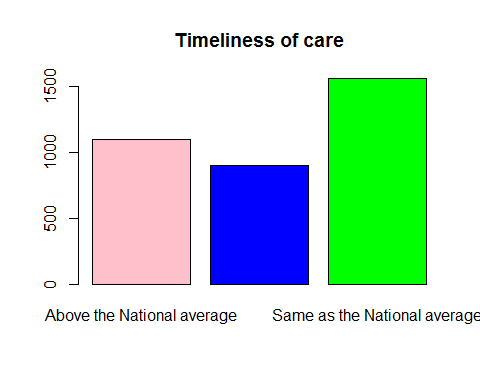
barplot(table(gen\_info$Patient\_experience),main = "Patient experience",col=c("pink","blue","green","yellow","grey"))



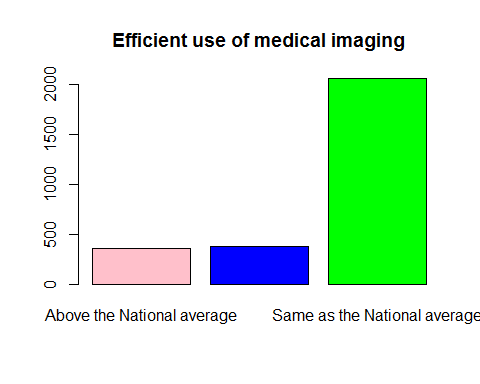
barplot(table(gen\_info$Effectiveness\_of\_care),main = "Effectiveness of care",col=c("pink","blue","green","yellow","grey"))



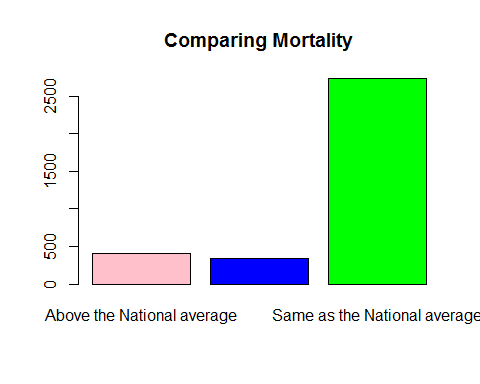
barplot(table(gen\_info$Timeliness\_of\_care),main = "Timeliness of care",col=c("pink","blue","green","yellow","grey"))



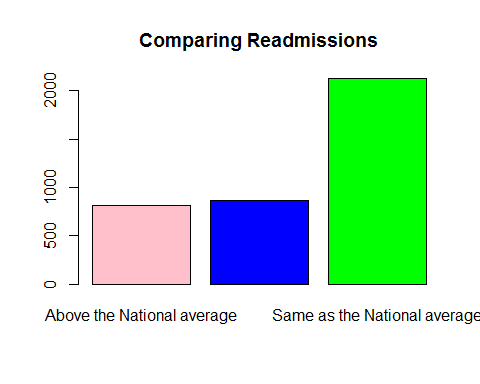
barplot(table(gen\_info$Efficient\_use\_of\_medical\_imaging),main = "Efficient use of medical imaging",col=c("pink","blue","green","yellow","grey"))



barplot(table(gen\_info$mortality),main = "Comparing Mortality",col=c("pink","blue","green","yellow","grey"))



barplot(table(gen\_info$Readmissions),main = "Comparing Readmissions",col=c("pink","blue","green","yellow","grey"))



table(gen\_info$capital)

##   
## No Yes   
## 4591 206

## Analyzing dataset

## Bivariate Analysis

## Chi Square Analysis

##Test the hypothesis whether the Hospital rating is independent of Hospital Type level at .05 significance level.  
tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Hospital\_Type)  
class(tbl)

## [1] "table"

tbl=tbl[,-2]  
tbl

##   
## Acute Care Hospitals Critical Access Hospitals  
## 1 107 1  
## 2 653 24  
## 3 1417 355  
## 4 726 215  
## 5 79 2

chisq.test(tbl)

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 155.2, df = 4, p-value < 2.2e-16

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Hospital\_Ownership)   
chisq.test(tbl)

## Warning in chisq.test(tbl): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 241.71, df = 36, p-value < 2.2e-16

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Emergency\_Services)   
chisq.test(tbl)

## Warning in chisq.test(tbl): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 26.258, df = 4, p-value = 2.807e-05

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$mortality)   
chisq.test(tbl)

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 263.73, df = 8, p-value < 2.2e-16

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Meets\_criteria\_for\_meaningful\_use\_of\_EHRs)   
chisq.test(tbl)

##   
## Chi-squared test for given probabilities  
##   
## data: tbl  
## X-squared = 2549.2, df = 4, p-value < 2.2e-16

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Safety\_of\_care)   
chisq.test(tbl)

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 620.13, df = 8, p-value < 2.2e-16

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Readmissions)   
chisq.test(tbl)

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 1208.6, df = 8, p-value < 2.2e-16

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Patient\_experience)   
chisq.test(tbl)

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 1565.4, df = 8, p-value < 2.2e-16

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Effectiveness\_of\_care)   
chisq.test(tbl)

## Warning in chisq.test(tbl): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 113.54, df = 8, p-value < 2.2e-16

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Timeliness\_of\_care)   
chisq.test(tbl)

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 462.18, df = 8, p-value < 2.2e-16

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$Efficient\_use\_of\_medical\_imaging)   
chisq.test(tbl)

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 36.047, df = 8, p-value = 1.722e-05

tbl = table(gen\_info$Hospital\_overall\_rating, gen\_info$capital)   
chisq.test(tbl)

## Warning in chisq.test(tbl): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: tbl  
## X-squared = 20.722, df = 4, p-value = 0.0003595

cor(gen\_info$Hospital\_overall\_rating,gen\_info$population)

## [1] NA

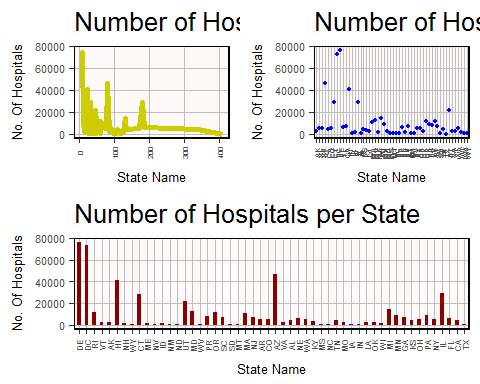
## Top 10 States in terms of No of Hospitals  
  
## Finding the Best 10 State in term of no. of hospitals provided by State over population and area per state  
  
pop\_state<-sqldf('select State,count(\*) as cnt,population,land\_area as area from gen\_info group by State order by cnt')

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

pop\_state=data.frame(pop\_state)  
pop\_state$cnt= as.numeric(pop\_state$cnt)  
pop\_state$population=as.numeric(pop\_state$population)  
pop\_state$area=as.numeric(pop\_state$area)  
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 DE 7 533514 5047093738 76216.29 721013391  
## 2 DC 8 584400 158350578 73050.00 19793822  
## 3 AZ 81 3751410 294205282243 46313.70 3632163978  
## 4 HI 23 936984 16634306891 40738.43 723230734  
## 5 IL 180 5172848 143794747023 28738.04 798859706  
## 6 CT 31 887976 12542396439 28644.39 404593434  
## 7 UT 46 1000155 212883424892 21742.50 4627900541  
## 8 MI 131 1870362 146455316950 14277.57 1117979519  
## 9 MD 49 620538 25147575220 12664.04 513215821  
## 10 OR 60 712036 248608666869 11867.27 4143477781

## theme of the results  
  
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 = 10),  
 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.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)



head(pop\_state)

## State cnt population area hos\_pop\_ratio hos\_area\_ratio  
## 1 DE 7 533514 5.047094e+09 76216.286 721013391  
## 2 DC 8 584400 1.583506e+08 73050.000 19793822  
## 3 RI 11 126987 2.677751e+09 11544.273 243431932  
## 4 VT 14 36841 2.387190e+10 2631.500 1705135458  
## 5 AK 22 53917 1.477849e+12 2450.773 67174970889  
## 6 HI 23 936984 1.663431e+10 40738.435 723230734

gen\_info1=left\_join(gen\_info,pop\_state)

## Joining, by = c("State", "population")

## Warning in left\_join\_impl(x, y, by$x, by$y, suffix$x, suffix$y): joining  
## factors with different levels, coercing to character vector

StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

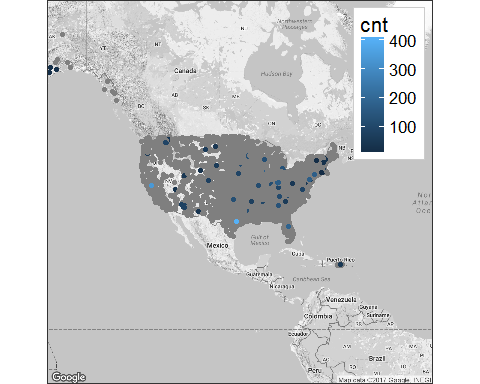
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false

## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

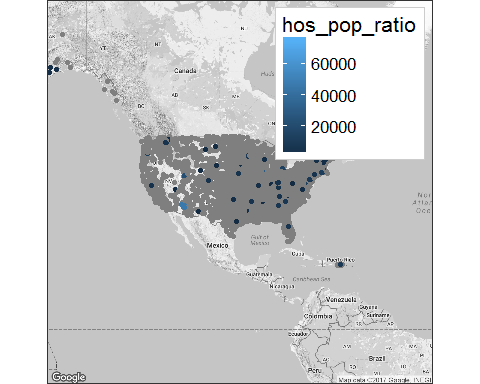
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = cnt), data = gen\_info1)

## Warning: Removed 29 rows containing missing values (geom\_point).



StateMap + geom\_point(aes(x = longitude, y = latitude, colour = hos\_pop\_ratio), data = gen\_info1)

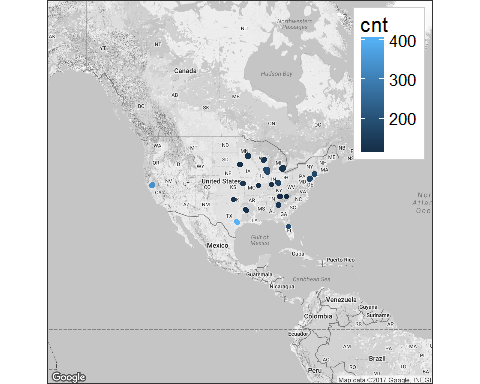
## Warning: Removed 29 rows containing missing values (geom\_point).



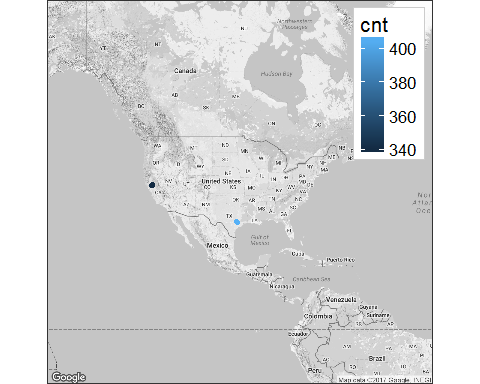
nrow(table(gen\_info1$cnt))

## [1] 46

gen\_info1=filter(gen\_info1,cnt>100)  
  
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = cnt), data = gen\_info1)



gen\_info1=filter(gen\_info1,cnt>300)  
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = cnt), data = gen\_info1)



## Comparing Hospitals in terms of Ratings  
  
head(gen\_info)

## Provider\_ID Hospital\_Name City State zip  
## 1 10005 MARSHALL MEDICAL CENTER SOUTH BOAZ AL 35957  
## 2 10006 ELIZA COFFEE MEMORIAL HOSPITAL FLORENCE AL 35631  
## 3 10007 MIZELL MEMORIAL HOSPITAL OPP AL 36467  
## 4 10008 CRENSHAW COMMUNITY HOSPITAL LUVERNE AL 36049  
## 5 10011 ST VINCENT'S EAST BIRMINGHAM AL 35235  
## 6 10012 DEKALB REGIONAL MEDICAL CENTER FORT PAYNE AL 35968  
## Hospital\_Type Hospital\_Ownership  
## 1 Acute Care Hospitals Government - Hospital District or Authority  
## 2 Acute Care Hospitals Government - Hospital District or Authority  
## 3 Acute Care Hospitals Voluntary non-profit - Private  
## 4 Acute Care Hospitals Proprietary  
## 5 Acute Care Hospitals Voluntary non-profit - Private  
## 6 Acute Care Hospitals Proprietary  
## Emergency\_Services Meets\_criteria\_for\_meaningful\_use\_of\_EHRs  
## 1 Yes Y  
## 2 Yes Y  
## 3 Yes Y  
## 4 Yes Y  
## 5 Yes Y  
## 6 Yes Y  
## Hospital\_overall\_rating mortality  
## 1 3 Below the National average  
## 2 2 Below the National average  
## 3 2 Same as the National average  
## 4 3 Same as the National average  
## 5 2 Same as the National average  
## 6 3 Below the National average  
## Safety\_of\_care Readmissions  
## 1 Same as the National average Above the National average  
## 2 Same as the National average Same as the National average  
## 3 <NA> Below the National average  
## 4 <NA> Same as the National average  
## 5 Below the National average Same as the National average  
## 6 Same as the National average Same as the National average  
## Patient\_experience Effectiveness\_of\_care  
## 1 Same as the National average Same as the National average  
## 2 Below the National average Same as the National average  
## 3 Same as the National average Below the National average  
## 4 <NA> Same as the National average  
## 5 Below the National average Below the National average  
## 6 Same as the National average Same as the National average  
## Timeliness\_of\_care Efficient\_use\_of\_medical\_imaging latitude  
## 1 Above the National average Below the National average 34.20362  
## 2 Above the National average Same as the National average 34.86996  
## 3 Above the National average <NA> 31.28005  
## 4 Above the National average <NA> 31.73741  
## 5 Same as the National average Same as the National average 33.60978  
## 6 Above the National average Same as the National average 34.46054  
## longitude fips population name capital geoid land\_area  
## 1 -86.17280 1095 90867 boaz al No 01 131172403111  
## 2 -87.70686 1077 91762 florence al No 01 131172403111  
## 3 -86.25484 1039 37608 opp al No 01 131172403111  
## 4 -86.26946 1041 13928 luverne al No 01 131172403111  
## 5 -86.65507 1073 656912 birmingham al No 01 131172403111  
## 6 -85.76744 1049 70038 fort payne al No 01 131172403111  
## area\_bin pop\_bin  
## 1 medium medium  
## 2 medium medium  
## 3 medium medium  
## 4 medium small  
## 5 medium verybig  
## 6 medium medium

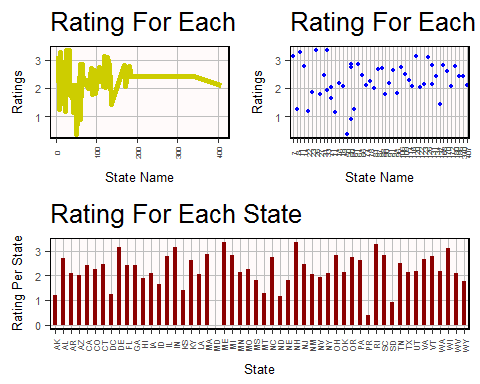
rating\_state<-sqldf('select State,count(\*) as cnt,sum(Hospital\_overall\_rating) as rating from gen\_info group by State order by cnt')  
rating\_state$rating= as.numeric(rating\_state$rating)  
rating\_state=mutate(rating\_state,state.rating=rating/cnt)  
top\_state\_rating=(arrange(rating\_state,desc(state.rating)))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

## Warning: Removed 1 rows containing missing values (position\_stack).



gen\_info2=left\_join(gen\_info,top\_state\_rating)

## Joining, by = "State"

StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

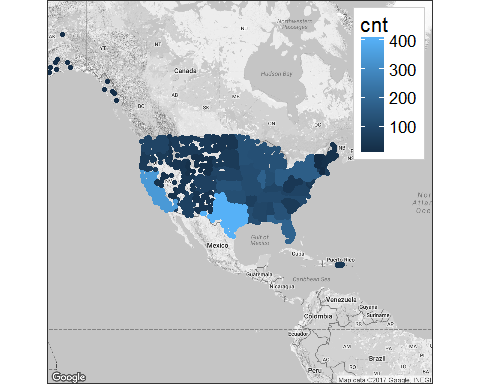
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false

## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = cnt), data = gen\_info2)

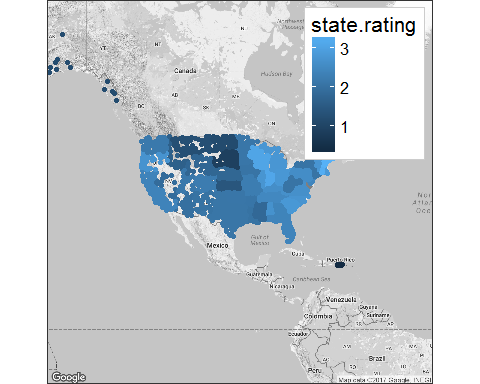
## Warning: Removed 29 rows containing missing values (geom\_point).



StateMap + geom\_point(aes(x = longitude, y = latitude, colour = state.rating), data = gen\_info2,title="Average Ratings For each State")

## Warning: Ignoring unknown parameters: title

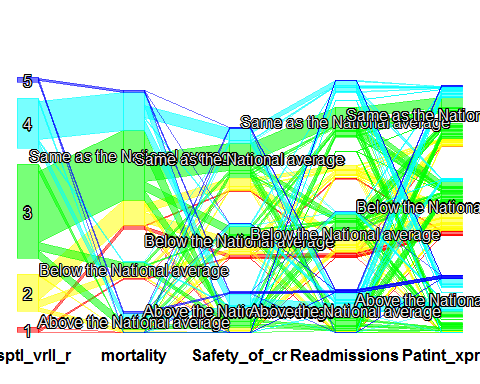
## Warning: Removed 29 rows containing missing values (geom\_point).

 ##Analyzing dataset ##Multivariate Analysis ##Correlation Analysis and Parallel Coordinates

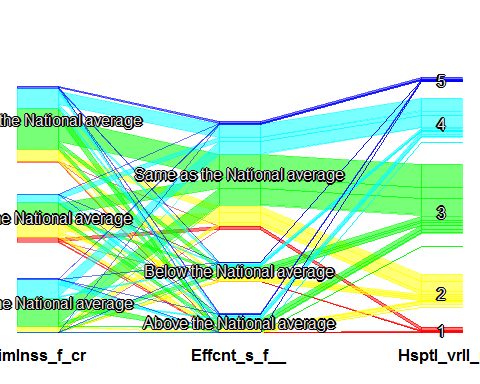
## Parallel Coordinate Analysis  
names(gen\_info)

## [1] "Provider\_ID"   
## [2] "Hospital\_Name"   
## [3] "City"   
## [4] "State"   
## [5] "zip"   
## [6] "Hospital\_Type"   
## [7] "Hospital\_Ownership"   
## [8] "Emergency\_Services"   
## [9] "Meets\_criteria\_for\_meaningful\_use\_of\_EHRs"  
## [10] "Hospital\_overall\_rating"   
## [11] "mortality"   
## [12] "Safety\_of\_care"   
## [13] "Readmissions"   
## [14] "Patient\_experience"   
## [15] "Effectiveness\_of\_care"   
## [16] "Timeliness\_of\_care"   
## [17] "Efficient\_use\_of\_medical\_imaging"   
## [18] "latitude"   
## [19] "longitude"   
## [20] "fips"   
## [21] "population"   
## [22] "name"   
## [23] "capital"   
## [24] "geoid"   
## [25] "land\_area"   
## [26] "area\_bin"   
## [27] "pop\_bin"

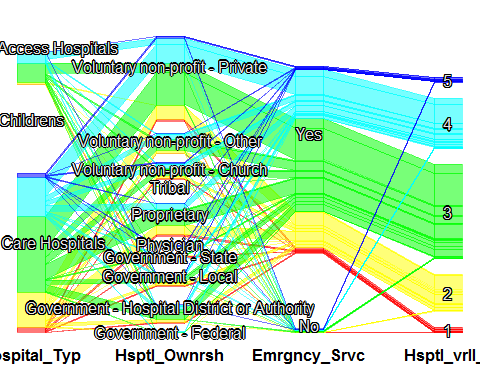
palette=getcolors(6, "rgb")  
scpcp(gen\_info[,10:14],sel = gen\_info[,10] ,sel.palette=palette)



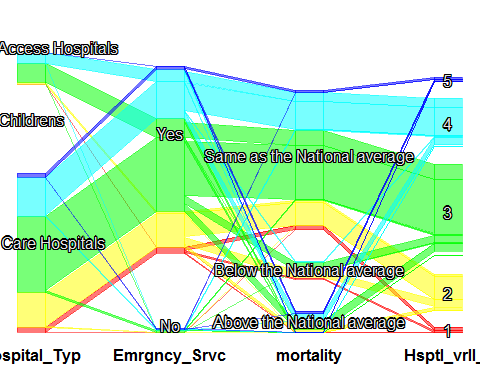
scpcp(gen\_info[,c(16,17,10)],sel = gen\_info[,10] ,sel.palette=palette)



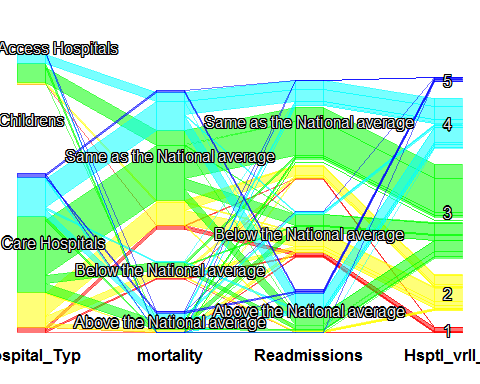
scpcp(gen\_info[,c(6,7,8,10)], sel = gen\_info[,10] ,sel.palette=palette)



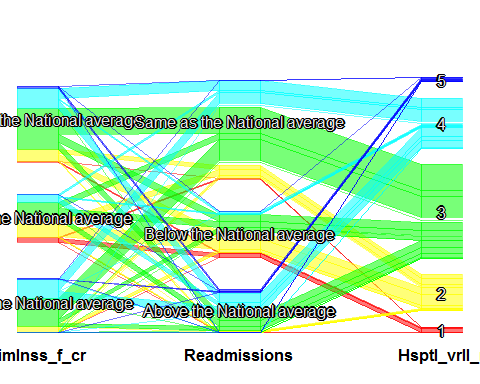
scpcp(gen\_info[,c(6,8,11,10)], sel = gen\_info[,10] ,sel.palette=palette)



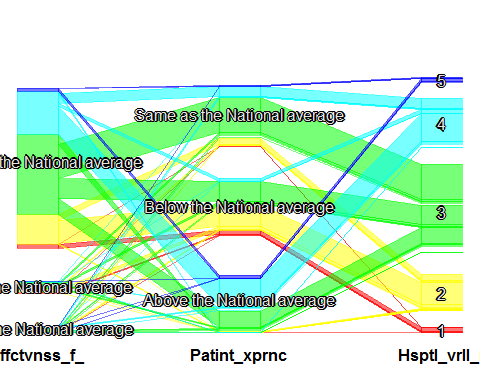
scpcp(gen\_info[,c(6,11,13,10)], sel = gen\_info[,10] ,sel.palette=palette)



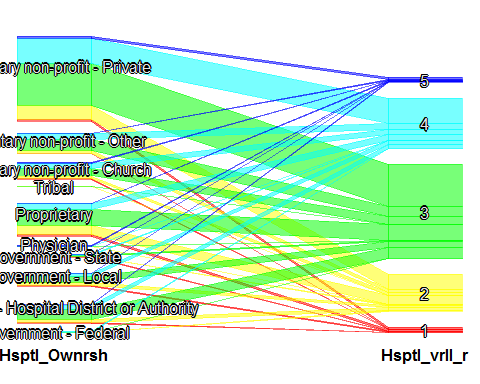
scpcp(gen\_info[,c(16,13,10)], sel = gen\_info[,10] ,sel.palette=palette)



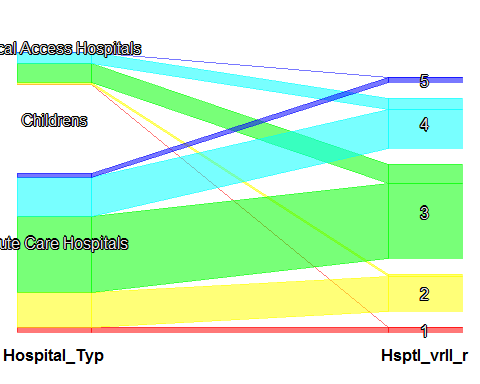
scpcp(gen\_info[,c(15,14,10)], sel = gen\_info[,10] ,sel.palette=palette)



scpcp(gen\_info[,c(7,10)], sel = gen\_info[,10] ,sel.palette=palette)



scpcp(gen\_info[,c(6,10)], sel = gen\_info[,10] ,sel.palette=palette)



names(gen\_info)

## [1] "Provider\_ID"   
## [2] "Hospital\_Name"   
## [3] "City"   
## [4] "State"   
## [5] "zip"   
## [6] "Hospital\_Type"   
## [7] "Hospital\_Ownership"   
## [8] "Emergency\_Services"   
## [9] "Meets\_criteria\_for\_meaningful\_use\_of\_EHRs"  
## [10] "Hospital\_overall\_rating"   
## [11] "mortality"   
## [12] "Safety\_of\_care"   
## [13] "Readmissions"   
## [14] "Patient\_experience"   
## [15] "Effectiveness\_of\_care"   
## [16] "Timeliness\_of\_care"   
## [17] "Efficient\_use\_of\_medical\_imaging"   
## [18] "latitude"   
## [19] "longitude"   
## [20] "fips"   
## [21] "population"   
## [22] "name"   
## [23] "capital"   
## [24] "geoid"   
## [25] "land\_area"   
## [26] "area\_bin"   
## [27] "pop\_bin"

x=filter(gen\_info,Hospital\_Type=="Childrens")  
  
table(x$Hospital\_overall\_rating,useNA = "always")##which means there is no information about performane of childrens hospital available

##   
## <NA>   
## 99

gen\_info=filter(gen\_info,Hospital\_Type!="Childrens")  
table(gen\_info$Hospital\_Type,useNA = "always")

##   
## Acute Care Hospitals Childrens   
## 3360 0   
## Critical Access Hospitals <NA>   
## 1338 0

gen\_info$Hospital\_Type <- factor(gen\_info$Hospital\_Type,levels = c("Acute Care Hospitals","Critical Access Hospitals"))  
table(gen\_info$Hospital\_Type,useNA = "always")

##   
## Acute Care Hospitals Critical Access Hospitals   
## 3360 1338   
## <NA>   
## 0

##analysing data on Hospital level  
  
cols <- c("State","City","Hospital\_Type","Hospital\_Ownership",  
 "capital","Meets\_criteria\_for\_meaningful\_use\_of\_EHRs",  
 "Hospital\_overall\_rating","mortality",  
 "Safety\_of\_care",  
 "Readmissions","Patient\_experience",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care",  
 "Efficient\_use\_of\_medical\_imaging")  
gen\_info[cols] <- lapply(gen\_info[cols], as.factor)  
names(gen\_info)

## [1] "Provider\_ID"   
## [2] "Hospital\_Name"   
## [3] "City"   
## [4] "State"   
## [5] "zip"   
## [6] "Hospital\_Type"   
## [7] "Hospital\_Ownership"   
## [8] "Emergency\_Services"   
## [9] "Meets\_criteria\_for\_meaningful\_use\_of\_EHRs"  
## [10] "Hospital\_overall\_rating"   
## [11] "mortality"   
## [12] "Safety\_of\_care"   
## [13] "Readmissions"   
## [14] "Patient\_experience"   
## [15] "Effectiveness\_of\_care"   
## [16] "Timeliness\_of\_care"   
## [17] "Efficient\_use\_of\_medical\_imaging"   
## [18] "latitude"   
## [19] "longitude"   
## [20] "fips"   
## [21] "population"   
## [22] "name"   
## [23] "capital"   
## [24] "geoid"   
## [25] "land\_area"   
## [26] "area\_bin"   
## [27] "pop\_bin"

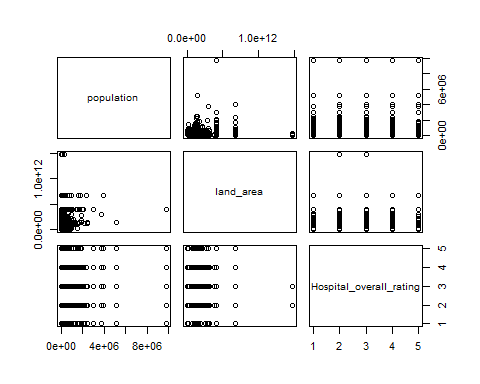
x=sqldf('select population,land\_area,Hospital\_overall\_rating from gen\_info')  
head(x)

## population land\_area Hospital\_overall\_rating  
## 1 90867 131172403111 3  
## 2 91762 131172403111 2  
## 3 37608 131172403111 2  
## 4 13928 131172403111 3  
## 5 656912 131172403111 2  
## 6 70038 131172403111 3

x$Hospital\_overall\_rating=as.numeric(x$Hospital\_overall\_rating)  
xx=filter(x,is.na(x$Hospital\_overall\_rating)==FALSE)  
mean=floor(mean(xx$Hospital\_overall\_rating))  
x$Hospital\_overall\_rating[is.na(x$Hospital\_overall\_rating)]=mean  
x$land\_area=as.numeric(x$land\_area)  
x$population=as.numeric(x$population)  
cor(x)

## population land\_area Hospital\_overall\_rating  
## population 1.00000000 0.19621825 -0.08644514  
## land\_area 0.19621825 1.00000000 0.02116817  
## Hospital\_overall\_rating -0.08644514 0.02116817 1.00000000

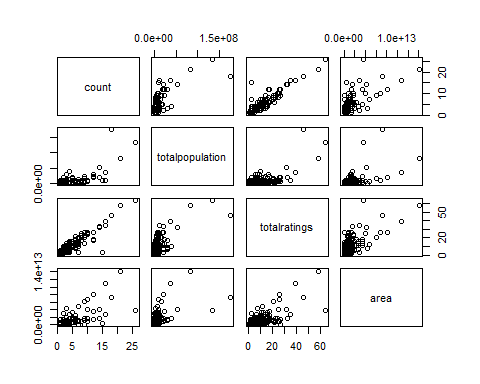
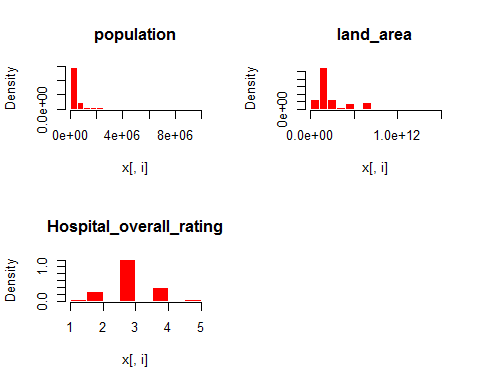
plot(x)



par(mfrow=c(2, 2))  
colnames <- dimnames(x)[[2]]  
for (i in 1:3) {  
 hist(x[,i], main=colnames[i], probability=TRUE, col="red", border="white")  
}  
  
##Analyzing data on City level  
  
hos\_type<-sqldf('select name ,count(\*) as count, sum(population) as totalpopulation,sum(land\_area) as area,sum(Hospital\_overall\_rating) as totalratings from gen\_info group by name order by count desc' )##name is more useful than city as there may be more cities with same name but different state as name variable has city and state information  
hos\_type=data.frame(hos\_type)  
hos\_type=hos\_type %>% mutate(hospitals\_person=totalpopulation/count,hospital\_area=area/count,average\_rating=totalratings/count)  
  
hos\_type$totalratings=as.numeric(hos\_type$totalratings)  
hos\_type$count=as.numeric(hos\_type$count)  
xx=filter(hos\_type,is.na(hos\_type$totalratings)==FALSE)##replacing NA's with mean  
mean=floor(mean(xx$totalratings))  
hos\_type$totalratings[is.na(hos\_type$totalratings)]=mean  
  
cor(hos\_type[,c("count","totalpopulation","totalratings","area")])

## count totalpopulation totalratings area  
## count 1.0000000 0.6660653 0.9092783 0.6818805  
## totalpopulation 0.6660653 1.0000000 0.6486225 0.5678046  
## totalratings 0.9092783 0.6486225 1.0000000 0.6702313  
## area 0.6818805 0.5678046 0.6702313 1.0000000

plot(hos\_type[,c("count","totalpopulation","totalratings","area")])



hospital\_type=hos\_type[,c(1,2,6,7,8)]  
gen\_info=left\_join(gen\_info,hospital\_type)

## Joining, by = "name"

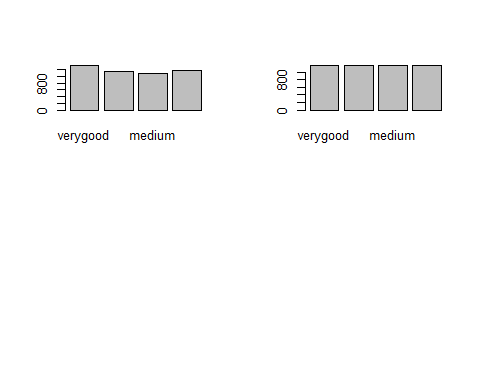
gen\_info$hospitals\_person=as.numeric(gen\_info$hospitals\_person)  
q <- quantile(gen\_info$hospitals\_person, c(0, .25, .50, .75, 1))  
gen\_info = data.frame(gen\_info, coverage\_person=cut(gen\_info$hospitals\_person, q,labels = c("verygood","good","medium","less"), ordered=TRUE, include.lowest=TRUE))  
levels(gen\_info$hospitals\_person)

## NULL

gen\_info$hospital\_area=as.numeric(gen\_info$hospital\_area)  
q <- quantile(gen\_info$hospital\_area, c(0, .25, .50, .75, 1))  
gen\_info = data.frame(gen\_info, coverage\_area=cut(gen\_info$hospital\_area, q,labels = c("verygood","good","medium","less"), ordered=TRUE, include.lowest=TRUE))  
plot(gen\_info$coverage\_area)  
levels(gen\_info$coverage\_area)

## [1] "verygood" "good" "medium" "less"

plot(gen\_info$coverage\_person)  
  
hos\_type$count=as.numeric(hos\_type$count)  
hos\_type$totalpopulation=as.numeric(hos\_type$totalpopulation)  
hos\_type$area=as.numeric(hos\_type$area)  
hos\_type$totalratings=as.numeric(hos\_type$totalratings)  
par(mfrow=c(3, 3))



colnames <- dimnames(hos\_type)[[2]]  
for (i in 2:5) {  
 hist(hos\_type[,i], main=colnames[i], probability=TRUE, col="red", border="white")  
}  
  
  
sum(is.na(hos\_type$totalratings))

## [1] 0

hos\_type=data.frame(hos\_type)  
hos\_type$totalratings=as.numeric(hos\_type$totalratings)  
hist(hos\_type$totalratings)  
##replacing the NA's with mean of rating  
xx=filter(hos\_type,is.na(hos\_type$totalratings)==FALSE)  
mean=floor(mean(xx$totalratings))  
hos\_type$totalratings[is.na(hos\_type$totalratings)]=mean  
cor(hos\_type[,-1])

## count totalpopulation area totalratings  
## count 1.00000000 0.66606535 0.6818805 0.90927834  
## totalpopulation 0.66606535 1.00000000 0.5678046 0.64862249  
## area 0.68188053 0.56780458 1.0000000 0.67023126  
## totalratings 0.90927834 0.64862249 0.6702313 1.00000000  
## hospitals\_person 0.16548044 0.50503981 0.1975458 0.14801418  
## hospital\_area 0.04710143 0.08989684 0.5315157 0.04986949  
## average\_rating NA NA NA NA  
## hospitals\_person hospital\_area average\_rating  
## count 0.1654804 0.04710143 NA  
## totalpopulation 0.5050398 0.08989684 NA  
## area 0.1975458 0.53151569 NA  
## totalratings 0.1480142 0.04986949 NA  
## hospitals\_person 1.0000000 0.14936482 NA  
## hospital\_area 0.1493648 1.00000000 NA  
## average\_rating NA NA 1

names(hos\_type)

## [1] "name" "count" "totalpopulation"   
## [4] "area" "totalratings" "hospitals\_person"  
## [7] "hospital\_area" "average\_rating"

scpcp(hos\_type[,c(2,5)],sel = hos\_type[,5])  
  
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false

## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = coverage\_area), data = gen\_info)+hospital.theme+  
 ggtitle("Hospital over area of city")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 27 rows containing missing values (geom\_point).

StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = coverage\_person), data = gen\_info)+hospital.theme+  
 ggtitle("Indivisuals Served per hospital")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 27 rows containing missing values (geom\_point).

table(gen\_info$count)

##   
## 1 2 3 4 5 6 7 8 9 10 12 14 15 16 18   
## 3270 566 216 140 65 96 49 32 36 40 48 28 15 32 18   
## 21 26   
## 21 26

range(gen\_info$count)

## [1] 1 26

table(gen\_info$coverage\_person)

##   
## verygood good medium less   
## 1175 1174 1175 1174

table(gen\_info$coverage\_area)

##   
## verygood good medium less   
## 1325 1131 1081 1161

gen\_info$count=as.factor(gen\_info$count)  
  
  
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = count), data = gen\_info)+hospital.theme+  
 ggtitle("Indivisuals Served per hospital")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 27 rows containing missing values (geom\_point).

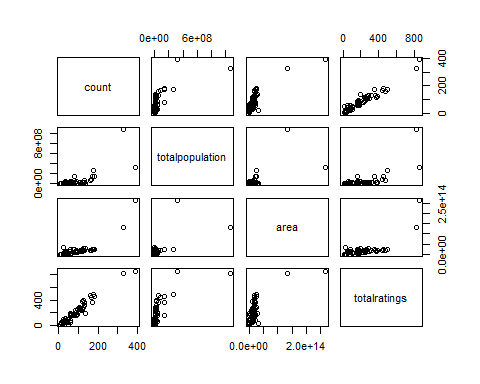
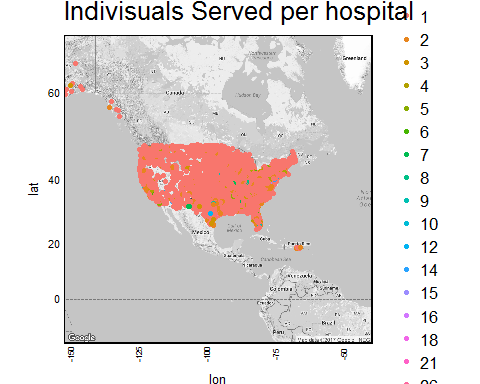
table(gen\_info$capital,gen\_info$count)

##   
## 1 2 3 4 5 6 7 8 9 10 12 14 15  
## No 3258 546 195 112 55 72 42 24 27 20 48 14 15  
## Yes 12 20 21 28 10 24 7 8 9 20 0 14 0  
##   
## 16 18 21 26  
## No 16 18 21 26  
## Yes 16 0 0 0

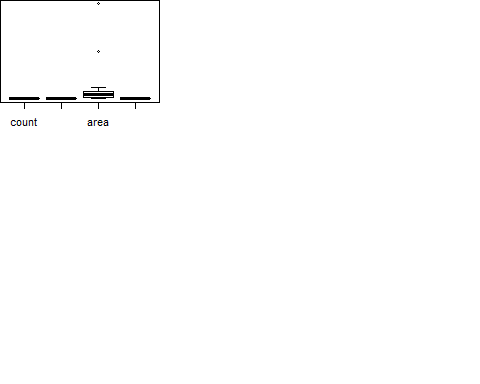
##Analyzing data on State level  
hos\_type2<-sqldf('select State ,count(\*) as count, sum(population) as totalpopulation,sum(land\_area) as area,sum(Hospital\_overall\_rating) as totalratings from gen\_info group by State order by count desc' )  
hos\_type2=data.frame(hos\_type2)  
hos\_type2$totalratings=as.numeric(hos\_type2$totalratings)  
hos\_type2$count=as.numeric(hos\_type2$count)  
class(hos\_type2$State)

## [1] "factor"

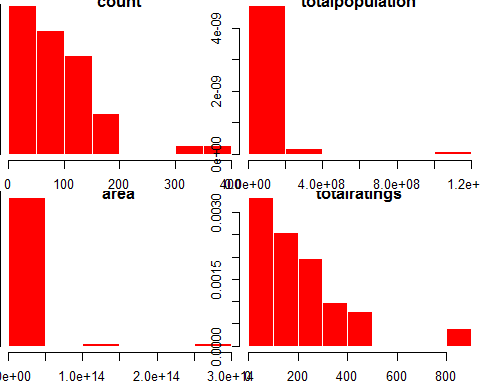
plot(hos\_type2[,-1])



boxplot(hos\_type2[,-1])  
par(mfrow=c(2, 2))



colnames <- dimnames(hos\_type2)[[2]]  
for (i in 2:5) {  
 hist(hos\_type2[,i], main=colnames[i], probability=TRUE, col="red", border="white")  
}



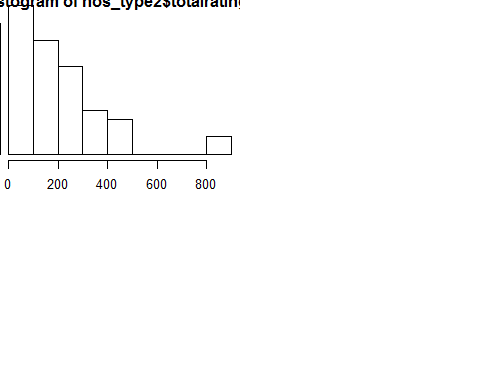
sum(is.na(hos\_type2$totalratings))

## [1] 1

hos\_type2=data.frame(hos\_type2)  
hos\_type2$ratings=as.numeric(hos\_type2$totalratings)  
hist(hos\_type2$totalratings)  
##replacing the NA's with mean of rating  
xx=filter(hos\_type2,is.na(hos\_type2$totalratings)==FALSE)  
mean=floor(mean(xx$totalratings))  
hos\_type2$totalratings[is.na(hos\_type2$totalratings)]=mean  
cor(hos\_type2[,-1])

## count totalpopulation area totalratings ratings  
## count 1.0000000 0.6970844 0.8185228 0.9629240 NA  
## totalpopulation 0.6970844 1.0000000 0.6271668 0.6967052 NA  
## area 0.8185228 0.6271668 1.0000000 0.7315663 NA  
## totalratings 0.9629240 0.6967052 0.7315663 1.0000000 NA  
## ratings NA NA NA NA 1

## 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.



a1=filter(gen\_info,Hospital\_Ownership=="Government - Federal")  
table(a1$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 0 5 8 4 0 29

a2=filter(gen\_info,Hospital\_Ownership=="Government - Hospital District or Authority")  
table(a2$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 8 60 198 76 5 216

a3=filter(gen\_info,Hospital\_Ownership=="Government - Local")  
table(a3$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 13 47 120 59 2 163

a10=filter(gen\_info,Hospital\_Ownership=="Government - State")  
table(a10$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 8 10 19 8 1 17

a4=filter(gen\_info,Hospital\_Ownership=="Physician")  
table(a4$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 0 2 5 4 7 46

a5=filter(gen\_info,Hospital\_Ownership=="Proprietary")  
table(a5$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 26 174 301 112 9 165

a6=filter(gen\_info,Hospital\_Ownership=="Tribal")  
table(a6$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 0 1 1 0 0 6

a7=filter(gen\_info,Hospital\_Ownership=="Voluntary non-profit - Church")  
table(a7$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 4 55 145 93 14 31

a8=filter(gen\_info,Hospital\_Ownership=="Voluntary non-profit - Other")  
table(a8$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 9 57 192 113 11 74

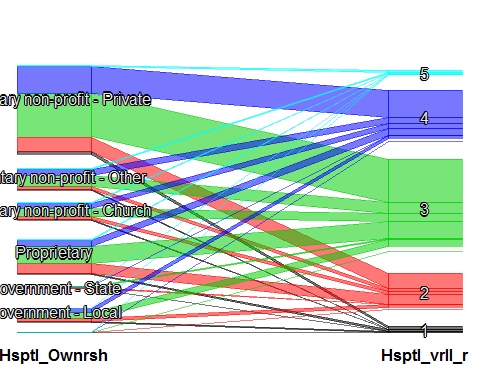
a9=filter(gen\_info,Hospital\_Ownership=="Voluntary non-profit - Private")  
table(a9$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 40 266 783 472 32 372

gen\_info=filter(gen\_info,  
 Hospital\_Ownership!="Government - Federal" &   
 Hospital\_Ownership!="Physician"&  
 Hospital\_Ownership!="Tribal" )  
  
  
gen\_info$Hospital\_Ownership <- factor(gen\_info$Hospital\_Ownership,levels = c("Hospital District or Authority","Government - Local","Government - State","Proprietary","Voluntary non-profit - Church","Voluntary non-profit - Other","Voluntary non-profit - Private"))  
table(gen\_info$Hospital\_Ownership,useNA="always")

##   
## Hospital District or Authority Government - Local   
## 0 404   
## Government - State Proprietary   
## 63 787   
## Voluntary non-profit - Church Voluntary non-profit - Other   
## 342 456   
## Voluntary non-profit - Private <NA>   
## 1965 563

scpcp(gen\_info[,c(7,10)], sel = gen\_info[,10] )



table(gen\_info$count)

##   
## 1 2 3 4 5 6 7 8 9 10 12 14 15 16 18   
## 3217 546 205 131 62 93 49 31 34 36 48 27 15 23 18   
## 21 26   
## 19 26

Data Cleaning

table(gen\_info$Hospital\_Ownership)

##   
## Hospital District or Authority Government - Local   
## 0 404   
## Government - State Proprietary   
## 63 787   
## Voluntary non-profit - Church Voluntary non-profit - Other   
## 342 456   
## Voluntary non-profit - Private   
## 1965

table(gen\_info$Hospital\_Type)

##   
## Acute Care Hospitals Critical Access Hospitals   
## 3257 1323

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

##   
## Y <NA>   
## 4184 396

table(gen\_info$capital,useNA="always")

##   
## No Yes <NA>   
## 4406 174 0

nrow(table(gen\_info$name,useNA="always"))

## [1] 3668

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

##   
## Above the National average Below the National average   
## 779 661   
## Same as the National average <NA>   
## 1185 1955

table(gen\_info$mortality,useNA="always")

##   
## Above the National average Below the National average   
## 395 340   
## Same as the National average <NA>   
## 2708 1137

table(gen\_info$Readmissions,useNA="always")

##   
## Above the National average Below the National average   
## 804 867   
## Same as the National average <NA>   
## 2099 810

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

##   
## Above the National average Below the National average   
## 1172 1071   
## Same as the National average <NA>   
## 1139 1198

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

##   
## Above the National average Below the National average   
## 128 240   
## Same as the National average <NA>   
## 3187 1025

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

##   
## Above the National average Below the National average   
## 1086 899   
## Same as the National average <NA>   
## 1540 1055

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

##   
## Above the National average Below the National average   
## 357 366   
## Same as the National average <NA>   
## 2040 1817

## Assigning values to NA's  
  
gen\_info$Safety\_of\_care[is.na(gen\_info$Safety\_of\_care)]="Same as the National average"  
  
gen\_info$mortality[is.na(gen\_info$mortality)]="Same as the National average"  
  
gen\_info$Readmissions[is.na(gen\_info$Readmissions)]="Same as the National average"  
  
gen\_info$Patient\_experience[is.na(gen\_info$Patient\_experience)]="Above the National average"  
  
gen\_info$Effectiveness\_of\_care[is.na(gen\_info$Effectiveness\_of\_care)]="Same as the National average"  
  
gen\_info$Timeliness\_of\_care[is.na(gen\_info$Timeliness\_of\_care)]="Same as the National average"  
  
gen\_info$Efficient\_use\_of\_medical\_imaging[is.na(gen\_info$Efficient\_use\_of\_medical\_imaging)]="Same as the National average"

##dividing gen\_info into two   
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
dataset2=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==TRUE)## 24% od data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type","Emergency\_Services",  
 "mortality","coverage\_area",  
 "Safety\_of\_care","count","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

class(dataset1$Hospital\_overall\_rating)

## [1] "factor"

dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

crx <- dataset1[ sample( nrow( dataset1 ) ), ]  
X <- crx[,2:11]  
y <- crx[,1]  
trainX <- X[1:3200,]  
trainy <- y[1:3200]  
testX <- X[3201:3579,]  
testy <- y[3201:3579]  
  
model <- C50::C5.0( trainX, trainy )  
summary( model )

##   
## Call:  
## C5.0.default(x = trainX, y = trainy)  
##   
##   
## C5.0 [Release 2.07 GPL Edition] Mon Apr 10 10:51:05 2017  
## -------------------------------  
##   
## Class specified by attribute `outcome'  
##   
## Read 3200 cases (11 attributes) from undefined.data  
##   
## Decision tree:  
##   
## Safety\_of\_care = Below the National average:  
## :...State in {AK,CO,IA,ID,KS,MA,MN,MT,SD}: 3 (50/11)  
## : State in {AR,DC,ND}: 1 (12/5)  
## : State in {DE,HI,NH,RI}: 4 (6/1)  
## : State in {FL,IN,LA,MD,ME,MS,NM,PA,PR,VT,WV,WY}: 2 (110/46)  
## : State = CT:  
## : :...Timeliness\_of\_care = Below the National average: 2 (6/2)  
## : : Timeliness\_of\_care in {Above the National average,  
## : : Same as the National average}: 3 (5)  
## : State = KY:  
## : :...count in {1,2,5,6,7,8,9,10,12,14,15,16,18,21,26}: 3 (5)  
## : : count in {3,4}: 2 (4/1)  
## : State = MO:  
## : :...Effectiveness\_of\_care = Above the National average: 4 (1)  
## : : Effectiveness\_of\_care in {Below the National average,  
## : : Same as the National average}: 2 (8/2)  
## : State = NC:  
## : :...count in {1,4,5,6,7,8,9,10,12,14,15,16,18,21,26}: 2 (11/2)  
## : : count in {2,3}: 3 (3)  
## : State = NE:  
## : :...mortality = Below the National average: 2 (1)  
## : : mortality in {Above the National average,  
## : : Same as the National average}: 3 (4)  
## : State = NJ:  
## : :...count in {1,2,4,5,6,7,8,9,10,12,14,15,16,18,21,26}: 2 (19/7)  
## : : count = 3: 1 (2)  
## : State = NV:  
## : :...mortality = Below the National average: 1 (4)  
## : : mortality = Same as the National average: 2 (3/1)  
## : : mortality = Above the National average: 3 (1)  
## : State = UT:  
## : :...Timeliness\_of\_care = Above the National average: 2 (2/1)  
## : : Timeliness\_of\_care in {Below the National average,  
## : : Same as the National average}: 3 (4)  
## : State = AL:  
## : :...coverage\_person in [medium-less]: 2 (10/2)  
## : : coverage\_person in [verygood-good]:  
## : : :...mortality = Below the National average: 2 (1)  
## : : mortality in {Above the National average,  
## : : Same as the National average}: 3 (3)  
## : State = AZ:  
## : :...Timeliness\_of\_care = Above the National average: 3 (2)  
## : : Timeliness\_of\_care = Same as the National average: 2 (5/1)  
## : : Timeliness\_of\_care = Below the National average:  
## : : :...coverage\_person in [verygood-medium]: 1 (2/1)  
## : : coverage\_person = less: 3 (6/3)  
## : State = CA:  
## : :...mortality in {Below the National average,  
## : : : Same as the National average}: 2 (34/15)  
## : : mortality = Above the National average:  
## : : :...Effectiveness\_of\_care in {Above the National average,  
## : : : Below the National average}: 4 (4/1)  
## : : Effectiveness\_of\_care = Same as the National average: 3 (9/3)  
## : State = GA:  
## : :...count = 2: 3 (2)  
## : : count in {3,5,6,8,9,10,12,14,15,16,18,21,26}: 2 (4/1)  
## : : count = 4: 1 (2/1)  
## : : count = 1:  
## : : :...Effectiveness\_of\_care = Below the National average: 1 (1)  
## : : : Effectiveness\_of\_care in {Above the National average,  
## : : : Same as the National average}: 2 (8/3)  
## : : count = 7:  
## : : :...Timeliness\_of\_care in {Above the National average,  
## : : : Below the National average}: 1 (3/1)  
## : : Timeliness\_of\_care = Same as the National average: 3 (2/1)  
## : State = IL:  
## : :...count in {3,4,5,6,7,8,9,10,12,14,15,16,18,21}: 3 (0)  
## : : count = 1:  
## : : :...Effectiveness\_of\_care = Above the National average: 4 (1)  
## : : : Effectiveness\_of\_care = Below the National average: 2 (3/1)  
## : : : Effectiveness\_of\_care = Same as the National average: 3 (10/5)  
## : : count = 2:  
## : : :...Effectiveness\_of\_care = Above the National average: 2 (1)  
## : : : Effectiveness\_of\_care in {Below the National average,  
## : : : Same as the National average}: 3 (7)  
## : : count = 26:  
## : : :...Effectiveness\_of\_care = Above the National average: 3 (1)  
## : : Effectiveness\_of\_care = Below the National average: 1 (1)  
## : : Effectiveness\_of\_care = Same as the National average: 2 (8/3)  
## : State = MI:  
## : :...Timeliness\_of\_care = Above the National average: 3 (2/1)  
## : : Timeliness\_of\_care = Below the National average:  
## : : :...count in {1,2,3,4,5,7,8,9,10,12,14,15,16,18,21,26}: 2 (13/4)  
## : : : count = 6: 1 (3)  
## : : Timeliness\_of\_care = Same as the National average:  
## : : :...count in {1,2}: 3 (5/1)  
## : : count in {3,4,5,6,7,8,9,10,12,14,15,16,18,21,26}: 2 (5/1)  
## : State = NY:  
## : :...count = 1: 2 (21/8)  
## : : count = 5: 3 (1)  
## : : count in {6,12}: 1 (14/4)  
## : : count = 3:  
## : : :...mortality in {Above the National average,  
## : : : : Below the National average}: 3 (2)  
## : : : mortality = Same as the National average: 1 (2/1)  
## : : count in {2,4,7,8,9,10,14,15,16,18,21,26}:  
## : : :...count in {2,7,8,9,10,14,15,16,18,21,26}: 1 (8/2)  
## : : count = 4: 2 (2)  
## : State = OK:  
## : :...Effectiveness\_of\_care = Below the National average: 2 (0)  
## : : Effectiveness\_of\_care = Above the National average: 1 (1)  
## : : Effectiveness\_of\_care = Same as the National average:  
## : : :...count in {1,2,4,5,6,7,8,9,12,14,15,16,18,21,26}: 2 (4/1)  
## : : count in {3,10}: 3 (2)  
## : State = OR:  
## : :...mortality = Above the National average: 3 (0)  
## : : mortality = Below the National average: 1 (1)  
## : : mortality = Same as the National average:  
## : : :...Timeliness\_of\_care = Above the National average: 2 (1)  
## : : Timeliness\_of\_care in {Below the National average,  
## : : Same as the National average}: 3 (5/2)  
## : State = SC:  
## : :...count = 2: 3 (1)  
## : : count in {3,4,5,6,7,8,9,10,12,14,15,16,18,21,26}: 2 (3)  
## : : count = 1:  
## : : :...Timeliness\_of\_care in {Above the National average,  
## : : : Below the National average}: 3 (2)  
## : : Timeliness\_of\_care = Same as the National average: 2 (3/1)  
## : State = TN:  
## : :...count in {1,3,4,5,6,7,8,9,10,12,14,15,16,18,21,26}: 2 (8/1)  
## : : count = 2:  
## : : :...mortality = Below the National average: 1 (1)  
## : : mortality in {Above the National average,  
## : : Same as the National average}: 3 (3)  
## : State = VA:  
## : :...mortality = Above the National average: 2 (2)  
## : : mortality = Below the National average: 1 (1)  
## : : mortality = Same as the National average:  
## : : :...count = 1: 2 (4/1)  
## : : count in {2,3,4,5,6,7,8,9,10,12,14,15,16,18,21,26}: 3 (3)  
## : State = WI:  
## : :...coverage\_person in [verygood-good]:  
## : : :...Timeliness\_of\_care in {Above the National average,  
## : : : : Below the National average}: 4 (4/2)  
## : : : Timeliness\_of\_care = Same as the National average: 5 (1)  
## : : coverage\_person in [medium-less]:  
## : : :...mortality = Above the National average: 4 (1)  
## : : mortality in {Below the National average,  
## : : Same as the National average}: 3 (5)  
## : State = TX:  
## : :...mortality = Above the National average:  
## : : :...count in {1,3}: 4 (2)  
## : : : count = 2: 3 (2)  
## : : : count in {4,5,6,7,8,9,10,12,14,15,16,18,21,26}: 2 (2/1)  
## : : mortality in {Below the National average,Same as the National average}:  
## : : :...count in {1,3,16}: 3 (13/4)  
## : : count in {4,5,6,8,9,10,14,15,18,21,26}: 2 (3)  
## : : count = 12: 1 (1)  
## : : count = 2:  
## : : :...Timeliness\_of\_care = Above the National average: 3 (2)  
## : : : Timeliness\_of\_care in {Below the National average,  
## : : : Same as the National average}: 2 (4)  
## : : count = 7:  
## : : :...mortality = Below the National average: 1 (1)  
## : : mortality = Same as the National average: 2 (3)  
## : State = WA:  
## : :...Effectiveness\_of\_care in {Above the National average,  
## : : : Below the National average}: 2 (2)  
## : : Effectiveness\_of\_care = Same as the National average:  
## : : :...mortality in {Above the National average,  
## : : : Same as the National average}: 3 (9/1)  
## : : mortality = Below the National average:  
## : : :...Timeliness\_of\_care in {Above the National average,  
## : : : Same as the National average}: 2 (5)  
## : : Timeliness\_of\_care = Below the National average: 3 (2)  
## : State = OH:  
## : :...count in {3,4,5,6,7,9,10,12,14,15,16,18,21,26}: 2 (8/2)  
## : count in {1,2,8}:  
## : :...Effectiveness\_of\_care = Above the National average: 3 (0)  
## : Effectiveness\_of\_care = Below the National average: 2 (1)  
## : Effectiveness\_of\_care = Same as the National average:  
## : :...mortality = Below the National average: 3 (0)  
## : mortality = Above the National average: 4 (3/1)  
## : mortality = Same as the National average: [S1]  
## Safety\_of\_care in {Above the National average,Same as the National average}:  
## :...mortality = Below the National average: 3 (216/105)  
## mortality in {Above the National average,Same as the National average}:  
## :...Timeliness\_of\_care = Below the National average:  
## :...mortality = Above the National average: 4 (102/54)  
## : mortality = Same as the National average:  
## : :...coverage\_person in [verygood-medium]: 3 (160/80)  
## : coverage\_person = less:  
## : :...Safety\_of\_care = Above the National average: 3 (62/30)  
## : Safety\_of\_care = Same as the National average: 2 (109/49)  
## Timeliness\_of\_care in {Above the National average,  
## : Same as the National average}:  
## :...Safety\_of\_care = Same as the National average: 3 (1560/625)  
## Safety\_of\_care = Above the National average:  
## :...State in {AL,CT,FL,MA,NH,NV,OR,TN,VA,WV}: 3 (74/30)  
## State in {AK,CO,DC,DE,GA,IA,ID,IL,IN,LA,MD,ME,MI,MN,MO,MT,ND,  
## : NJ,NM,NY,OH,PA,PR,UT,VT,WI}: 4 (171/59)  
## State in {HI,RI,SD}: 5 (5/1)  
## State = WY: 2 (1)  
## State = AR: [S2]  
## State = AZ: [S3]  
## State = CA:  
## :...Emergency\_Services = No: 4 (2/1)  
## : Emergency\_Services = Yes: 3 (28/15)  
## State = KS: [S4]  
## State = KY:  
## :...mortality = Above the National average: 4 (1)  
## : mortality = Same as the National average: 3 (10/4)  
## State = MS: [S5]  
## State = NE: [S6]  
## State = OK: [S7]  
## State = SC:  
## :...mortality = Above the National average: 3 (1)  
## : mortality = Same as the National average: 5 (5/2)  
## State = TX:  
## :...coverage\_person in [verygood-good]: 3 (4/2)  
## : coverage\_person in [medium-less]: 4 (33/13)  
## State = NC:  
## :...coverage\_person in [verygood-good]: 3 (3/1)  
## : coverage\_person in [medium-less]: [S8]  
## State = WA:  
## :...mortality = Above the National average: 5 (1)  
## mortality = Same as the National average:  
## :...count in {1,3,4,5,6,8,9,10,12,14,15,16,18,21,  
## : 26}: 3 (3)  
## count in {2,7}: 4 (2)  
##   
## SubTree [S1]  
##   
## Timeliness\_of\_care in {Above the National average,  
## : Same as the National average}: 3 (9/3)  
## Timeliness\_of\_care = Below the National average: 2 (2)  
##   
## SubTree [S2]  
##   
## Effectiveness\_of\_care = Above the National average: 4 (1)  
## Effectiveness\_of\_care in {Below the National average,  
## Same as the National average}: 3 (5/1)  
##   
## SubTree [S3]  
##   
## Effectiveness\_of\_care = Below the National average: 2 (1)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 4 (8/4)  
##   
## SubTree [S4]  
##   
## Effectiveness\_of\_care = Below the National average: 3 (1)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 4 (5/2)  
##   
## SubTree [S5]  
##   
## Timeliness\_of\_care = Above the National average: 4 (3/1)  
## Timeliness\_of\_care = Same as the National average: 2 (4/2)  
##   
## SubTree [S6]  
##   
## Timeliness\_of\_care = Above the National average: 4 (4/1)  
## Timeliness\_of\_care = Same as the National average: 3 (1)  
##   
## SubTree [S7]  
##   
## Effectiveness\_of\_care = Below the National average: 2 (1)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 3 (8/1)  
##   
## SubTree [S8]  
##   
## Effectiveness\_of\_care = Above the National average: 3 (1)  
## Effectiveness\_of\_care in {Below the National average,  
## Same as the National average}: 4 (5/1)  
##   
##   
## Evaluation on training data (3200 cases):  
##   
## Decision Tree   
## ----------------   
## Size Errors   
##   
## 134 1247(39.0%) <<  
##   
##   
## (a) (b) (c) (d) (e) <-classified as  
## ---- ---- ---- ---- ----  
## 45 45 17 (a): class 1  
## 9 291 281 21 (b): class 2  
## 4 90 1390 87 (c): class 3  
## 2 23 603 218 3 (d): class 4  
## 1 28 33 9 (e): class 5  
##   
##   
## Attribute usage:  
##   
## 100.00% Safety\_of\_care  
## 86.22% mortality  
## 77.81% Timeliness\_of\_care  
## 30.97% State  
## 12.81% coverage\_person  
## 8.41% count  
## 4.34% Effectiveness\_of\_care  
## 0.94% Emergency\_Services  
##   
##   
## Time: 0.0 secs

p<- predict( model, testX, type="class" )  
sum( p == testy ) / length( p )\*100##53% accuracy

## [1] NA

##tree 2 with different levels of ratings using C50 Tree  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1$Hospital\_overall\_rating=as.numeric(dataset1$Hospital\_overall\_rating)  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating<=2)] = "1"  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating==3)] = "2"  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating==4)|(dataset1$Hospital\_overall\_rating==5)] = "3"  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type",  
 "mortality","coverage\_area", "Readmissions","Patient\_experience",  
 "Safety\_of\_care","count","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3   
## 777 1758 1007

nrow(dataset1)

## [1] 3542

names(dataset1)

## [1] "Hospital\_overall\_rating" "State"   
## [3] "Hospital\_Type" "mortality"   
## [5] "coverage\_area" "Readmissions"   
## [7] "Patient\_experience" "Safety\_of\_care"   
## [9] "count" "coverage\_person"   
## [11] "Effectiveness\_of\_care" "Timeliness\_of\_care"

set.seed(12)  
crx <- dataset1[ sample( nrow( dataset1 ) ), ]  
X <- crx[,2:12]  
y <- crx[,1]  
trainX <- X[1:3200,]  
trainy <- y[1:3200]  
testX <- X[3201:3579,]  
testy <- y[3201:3579]  
  
model <- C50::C5.0( trainX, trainy )  
summary( model )

##   
## Call:  
## C5.0.default(x = trainX, y = trainy)  
##   
##   
## C5.0 [Release 2.07 GPL Edition] Mon Apr 10 10:51:06 2017  
## -------------------------------  
##   
## Class specified by attribute `outcome'  
##   
## Read 3200 cases (12 attributes) from undefined.data  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Readmissions = Below the National average:  
## : :...mortality in {Below the National average,  
## : : : Same as the National average}: 1 (325/53)  
## : : mortality = Above the National average:  
## : : :...Safety\_of\_care in {Above the National average,  
## : : : Same as the National average}: 2 (47/17)  
## : : Safety\_of\_care = Below the National average: 1 (40/7)  
## : Readmissions in {Above the National average,Same as the National average}:  
## : :...mortality = Above the National average:  
## : :...Safety\_of\_care = Above the National average: 3 (27/8)  
## : : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}: 2 (36/11)  
## : mortality in {Below the National average,Same as the National average}:  
## : :...Safety\_of\_care = Above the National average:  
## : :...Readmissions = Above the National average: 2 (45/16)  
## : : Readmissions = Same as the National average:  
## : : :...mortality = Same as the National average: 2 (47/9)  
## : : mortality = Below the National average:  
## : : :...count in {1,2,4,5,6,7,8,9,10,12,14,15,16,18,21,  
## : : : 26}: 1 (11/2)  
## : : count = 3: 2 (2)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...Safety\_of\_care = Below the National average:  
## : :...Readmissions = Same as the National average: 1 (79/11)  
## : : Readmissions = Above the National average:  
## : : :...mortality = Below the National average: 1 (16/3)  
## : : mortality = Same as the National average: 2 (33/12)  
## : Safety\_of\_care = Same as the National average:  
## : :...Readmissions = Above the National average: 2 (47/9)  
## : Readmissions = Same as the National average:  
## : :...mortality = Below the National average: 1 (15/2)  
## : mortality = Same as the National average: [S1]  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...mortality in {Below the National average,  
## : : : Same as the National average}: 2 (82/19)  
## : : mortality = Above the National average:  
## : : :...Patient\_experience = Above the National average: 3 (18/3)  
## : : Patient\_experience = Same as the National average: 2 (18/8)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Below the National average: 1 (28/5)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Safety\_of\_care = Below the National average: 1 (75/34)  
## : Safety\_of\_care = Same as the National average: [S2]  
## Readmissions in {Above the National average,Same as the National average}:  
## :...Safety\_of\_care = Above the National average:  
## :...mortality = Below the National average:  
## : :...Patient\_experience = Same as the National average: 2 (31/3)  
## : : Patient\_experience = Above the National average:  
## : : :...Readmissions = Above the National average: 3 (16/4)  
## : : Readmissions = Same as the National average: 2 (6/1)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Patient\_experience = Above the National average: 3 (156/4)  
## : Patient\_experience = Same as the National average:  
## : :...mortality = Above the National average: 3 (19)  
## : mortality = Same as the National average:  
## : :...Readmissions = Above the National average: 3 (54/7)  
## : Readmissions = Same as the National average:  
## : :...count in {1,4,5,6,7,8,10,14,15,18,21,  
## : : 26}: 2 (35/11)  
## : count in {2,3,9,12,16}: 3 (15/3)  
## Safety\_of\_care in {Below the National average,  
## : Same as the National average}:  
## :...Patient\_experience = Same as the National average:  
## :...mortality in {Below the National average,  
## : : Same as the National average}: 2 (601/121)  
## : mortality = Above the National average: [S3]  
## Patient\_experience = Above the National average:  
## :...mortality = Below the National average: [S4]  
## mortality in {Above the National average,  
## : Same as the National average}:  
## :...mortality = Above the National average: 3 (31/3)  
## mortality = Same as the National average: [S5]  
##   
## SubTree [S1]  
##   
## Timeliness\_of\_care in {Above the National average,  
## : Same as the National average}: 2 (132/42)  
## Timeliness\_of\_care = Below the National average:  
## :...coverage\_area = verygood: 2 (12/3)  
## coverage\_area in [good-less]: 1 (46/8)  
##   
## SubTree [S2]  
##   
## Effectiveness\_of\_care in {Above the National average,  
## : Same as the National average}: 2 (120/29)  
## Effectiveness\_of\_care = Below the National average: 1 (10/4)  
##   
## SubTree [S3]  
##   
## Effectiveness\_of\_care in {Above the National average,  
## : Below the National average}: 3 (5)  
## Effectiveness\_of\_care = Same as the National average: 2 (34/15)  
##   
## SubTree [S4]  
##   
## Effectiveness\_of\_care = Below the National average: 1 (3)  
## Effectiveness\_of\_care in {Above the National average,  
## : Same as the National average}:  
## :...Safety\_of\_care = Same as the National average: 2 (23/5)  
## Safety\_of\_care = Below the National average:  
## :...Readmissions = Same as the National average: 1 (5/1)  
## Readmissions = Above the National average:  
## :...Timeliness\_of\_care = Below the National average: 1 (2)  
## Timeliness\_of\_care in {Above the National average,  
## Same as the National average}: 2 (6)  
##   
## SubTree [S5]  
##   
## Effectiveness\_of\_care = Below the National average: 2 (42/14)  
## Effectiveness\_of\_care in {Above the National average,  
## : Same as the National average}:  
## :...Readmissions = Above the National average:  
## :...Safety\_of\_care = Below the National average: 2 (43/18)  
## : Safety\_of\_care = Same as the National average: 3 (104/9)  
## Readmissions = Same as the National average:  
## :...Safety\_of\_care = Below the National average: 2 (17/4)  
## Safety\_of\_care = Same as the National average:  
## :...count in {2,5,10,12}: 3 (46/16)  
## count in {3,6,7,8,9,14,15,16,18,21,26}: 2 (13/5)  
## count = 4:  
## :...coverage\_area = verygood: 2 (3)  
## : coverage\_area in [good-less]: 3 (5/1)  
## count = 1:  
## :...State in {AK,AR,CA,CO,CT,DC,DE,GA,HI,KS,KY,LA,MD,MO,MT,ND,NE,  
## : NJ,NM,NV,NY,OK,OR,PR,RI,SD,TN,TX,VT,WA,  
## : WV}: 2 (255/82)  
## State in {AZ,FL,IA,IL,MA,MI,MN,NC,NH,SC,UT,WI}: 3 (205/75)  
## State = AL:  
## :...coverage\_person = verygood: 2 (9/3)  
## : coverage\_person in [good-less]: 3 (6/1)  
## State = ID:  
## :...Hospital\_Type = Acute Care Hospitals: 2 (2)  
## : Hospital\_Type = Critical Access Hospitals: 3 (2)  
## State = IN:  
## :...coverage\_person = verygood: 3 (12/4)  
## : coverage\_person in [good-less]: 2 (12/4)  
## State = ME:  
## :...coverage\_person in [verygood-good]: 2 (8/3)  
## : coverage\_person in [medium-less]: 3 (3)  
## State = MS:  
## :...Hospital\_Type = Acute Care Hospitals: 3 (8/3)  
## : Hospital\_Type = Critical Access Hospitals: 2 (8/1)  
## State = OH:  
## :...Hospital\_Type = Acute Care Hospitals: 3 (8/2)  
## : Hospital\_Type = Critical Access Hospitals: 2 (10/4)  
## State = PA:  
## :...coverage\_person in [verygood-good]: 2 (9/3)  
## : coverage\_person in [medium-less]: 3 (7/1)  
## State = VA: [S6]  
## State = WY:  
## :...Hospital\_Type = Acute Care Hospitals: 3 (2)  
## Hospital\_Type = Critical Access Hospitals: 2 (4)  
##   
## SubTree [S6]  
##   
## Timeliness\_of\_care in {Above the National average,  
## : Below the National average}: 2 (2)  
## Timeliness\_of\_care = Same as the National average: 3 (2)  
##   
##   
## Evaluation on training data (3200 cases):  
##   
## Decision Tree   
## ----------------   
## Size Errors   
##   
## 69 746(23.3%) <<  
##   
##   
## (a) (b) (c) <-classified as  
## ---- ---- ----  
## 525 164 4 (a): class 1  
## 128 1322 140 (b): class 2  
## 2 308 607 (c): class 3  
##   
##   
## Attribute usage:  
##   
## 100.00% mortality  
## 100.00% Readmissions  
## 100.00% Patient\_experience  
## 89.84% Safety\_of\_care  
## 32.97% Effectiveness\_of\_care  
## 22.00% count  
## 17.94% State  
## 6.31% Timeliness\_of\_care  
## 2.06% coverage\_area  
## 2.06% coverage\_person  
## 1.38% Hospital\_Type  
##   
##   
## Time: 0.0 secs

p<- predict( model, testX, type="class" )  
sum( p == testy ) / length( p )\*100##74.406

## [1] NA

C5imp(model)

## Overall  
## mortality 100.00  
## Readmissions 100.00  
## Patient\_experience 100.00  
## Safety\_of\_care 89.84  
## Effectiveness\_of\_care 32.97  
## count 22.00  
## State 17.94  
## Timeliness\_of\_care 6.31  
## coverage\_area 2.06  
## coverage\_person 2.06  
## Hospital\_Type 1.38

##adding boosting  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
nrow(dataset1)

## [1] 3542

names(dataset1)

## [1] "Hospital\_overall\_rating" "State"   
## [3] "Hospital\_Type" "mortality"   
## [5] "coverage\_area" "Readmissions"   
## [7] "Patient\_experience" "Safety\_of\_care"   
## [9] "count" "coverage\_person"   
## [11] "Effectiveness\_of\_care" "Timeliness\_of\_care"

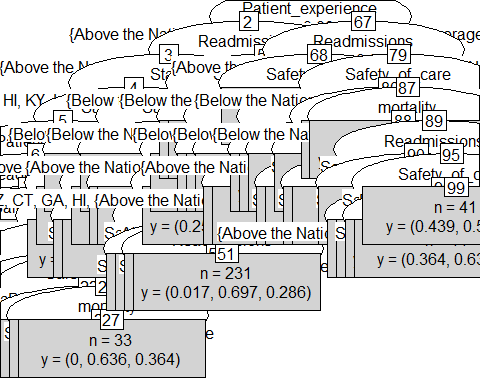
crx <- dataset1[ sample( nrow( dataset1 ) ), ]  
X <- crx[,2:10]  
y <- crx[,1]  
trainX <- X[1:3200,]  
trainy <- y[1:3200]  
testX <- X[3201:3579,]  
testy <- y[3201:3579]  
  
model <- C50::C5.0( trainX, trainy, trials=10 )  
summary( model )

##   
## Call:  
## C5.0.default(x = trainX, y = trainy, trials = 10)  
##   
##   
## C5.0 [Release 2.07 GPL Edition] Mon Apr 10 10:51:06 2017  
## -------------------------------  
##   
## Class specified by attribute `outcome'  
##   
## Read 3200 cases (10 attributes) from undefined.data  
##   
## ----- Trial 0: -----  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Readmissions = Below the National average:  
## : :...Safety\_of\_care = Above the National average:  
## : : :...mortality = Above the National average: 2 (30/11)  
## : : : mortality = Below the National average: 1 (18/2)  
## : : : mortality = Same as the National average:  
## : : : :...coverage\_person in [verygood-medium]: 2 (39/17)  
## : : : coverage\_person = less: 1 (45/17)  
## : : Safety\_of\_care in {Below the National average,  
## : : : Same as the National average}:  
## : : :...mortality in {Below the National average,  
## : : : Same as the National average}: 1 (222/15)  
## : : mortality = Above the National average:  
## : : :...Safety\_of\_care = Below the National average: 1 (41/7)  
## : : Safety\_of\_care = Same as the National average: 2 (20/8)  
## : Readmissions in {Above the National average,Same as the National average}:  
## : :...mortality = Above the National average:  
## : :...Safety\_of\_care = Above the National average: 3 (28/8)  
## : : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}: 2 (33/8)  
## : mortality in {Below the National average,Same as the National average}:  
## : :...Safety\_of\_care = Above the National average: 2 (103/30)  
## : Safety\_of\_care = Below the National average:  
## : :...Readmissions = Same as the National average: 1 (81/11)  
## : : Readmissions = Above the National average:  
## : : :...mortality = Below the National average: 1 (15/3)  
## : : mortality = Same as the National average: 2 (34/13)  
## : Safety\_of\_care = Same as the National average:  
## : :...Readmissions = Above the National average: 2 (48/10)  
## : Readmissions = Same as the National average:  
## : :...mortality = Below the National average: 1 (16/2)  
## : mortality = Same as the National average:  
## : :...coverage\_person in [verygood-medium]: 2 (116/38)  
## : coverage\_person = less: 1 (74/31)  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...mortality = Above the National average: 3 (35/12)  
## : : mortality in {Below the National average,  
## : : Same as the National average}: 2 (84/19)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Below the National average: 1 (28/3)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Safety\_of\_care = Below the National average: 1 (80/31)  
## : Safety\_of\_care = Same as the National average: 2 (129/35)  
## Readmissions in {Above the National average,Same as the National average}:  
## :...Safety\_of\_care = Above the National average:  
## :...mortality = Below the National average: 2 (55/16)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Patient\_experience = Above the National average: 3 (152/5)  
## : Patient\_experience = Same as the National average:  
## : :...mortality = Above the National average: 3 (18)  
## : mortality = Same as the National average: [S1]  
## Safety\_of\_care in {Below the National average,  
## : Same as the National average}:  
## :...Patient\_experience = Same as the National average: 2 (652/147)  
## Patient\_experience = Above the National average:  
## :...mortality = Below the National average: 2 (40/12)  
## mortality in {Above the National average,  
## : Same as the National average}:  
## :...Readmissions = Above the National average: [S2]  
## Readmissions = Same as the National average: [S3]  
##   
## SubTree [S1]  
##   
## Readmissions = Above the National average: 3 (53/7)  
## Readmissions = Same as the National average: 2 (46/21)  
##   
## SubTree [S2]  
##   
## Safety\_of\_care = Same as the National average: 3 (115/7)  
## Safety\_of\_care = Below the National average:  
## :...mortality = Above the National average: 3 (9)  
## mortality = Same as the National average: 2 (45/19)  
##   
## SubTree [S3]  
##   
## Safety\_of\_care = Below the National average: 2 (19/3)  
## Safety\_of\_care = Same as the National average:  
## :...State in {AR,AZ,CO,CT,GA,HI,IA,KS,KY,LA,MO,MS,MT,NC,ND,NE,NJ,NM,NV,NY,OK,  
## : OR,TN,VT,WA,WV}: 2 (312/106)  
## State in {FL,IN,MA,MI,MN,OH,SC,SD,UT,VA,WI}: 3 (196/68)  
## State in {AK,DC,DE,MD,PR,RI}: 1 (4/1)  
## State = AL:  
## :...coverage\_person = verygood: 2 (9/2)  
## : coverage\_person in [good-less]: 3 (5/1)  
## State = CA:  
## :...coverage\_person in [verygood-good]: 3 (2)  
## : coverage\_person in [medium-less]: 2 (13/4)  
## State = ID:  
## :...Hospital\_Type = Acute Care Hospitals: 2 (2)  
## : Hospital\_Type = Critical Access Hospitals: 3 (4/1)  
## State = IL:  
## :...count in {1,2,3,4,5,6,7,8,9,10,12,14,15,16,18,21}: 3 (28/11)  
## : count = 26: 2 (2)  
## State = ME:  
## :...coverage\_person in [verygood-good]: 2 (9/3)  
## : coverage\_person in [medium-less]: 3 (4)  
## State = NH:  
## :...Hospital\_Type = Acute Care Hospitals: 3 (2)  
## : Hospital\_Type = Critical Access Hospitals: 2 (6/2)  
## State = PA:  
## :...coverage\_person in [verygood-good]: 2 (8/2)  
## : coverage\_person in [medium-less]: 3 (9/2)  
## State = TX:  
## :...coverage\_person in [verygood-medium]: 2 (49/20)  
## : coverage\_person = less: 3 (6)  
## State = WY:  
## :...Hospital\_Type = Acute Care Hospitals: 3 (2)  
## Hospital\_Type = Critical Access Hospitals: 2 (5)  
##   
## ----- Trial 1: -----  
##   
## Decision tree:  
##   
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## : :...Safety\_of\_care = Above the National average:  
## : : :...mortality = Below the National average: 2 (15.4/4.5)  
## : : : mortality in {Above the National average,  
## : : : : Same as the National average}:  
## : : : :...Patient\_experience = Above the National average: 3 (47.1/19.5)  
## : : : Patient\_experience = Same as the National average: 2 (57.6/15.7)  
## : : Safety\_of\_care in {Below the National average,  
## : : : Same as the National average}:  
## : : :...mortality = Above the National average: 2 (55.3/15.2)  
## : : mortality = Below the National average: 1 (25.3/4.5)  
## : : mortality = Same as the National average:  
## : : :...Patient\_experience = Above the National average: 2 (62.2/18.1)  
## : : Patient\_experience = Same as the National average: 1 (101.2/40.8)  
## : Readmissions in {Above the National average,Same as the National average}:  
## : :...mortality = Below the National average: 2 (129/57.4)  
## : mortality in {Above the National average,Same as the National average}:  
## : :...Readmissions = Above the National average:  
## : :...Safety\_of\_care in {Above the National average,  
## : : : Same as the National average}: 3 (344.3/61.9)  
## : : Safety\_of\_care = Below the National average: [S1]  
## : Readmissions = Same as the National average:  
## : :...Safety\_of\_care = Below the National average: 2 (62.6/22.7)  
## : Safety\_of\_care = Above the National average: 3 (114.2/28.4)  
## : Safety\_of\_care = Same as the National average: [S2]  
## Patient\_experience = Below the National average:  
## :...Readmissions = Above the National average: 2 (172.5/81)  
## Readmissions in {Below the National average,Same as the National average}:  
## :...Safety\_of\_care = Below the National average: 1 (210.6/31.2)  
## Safety\_of\_care in {Above the National average,  
## : Same as the National average}:  
## :...mortality = Above the National average: 2 (83.4/39.4)  
## mortality = Below the National average: 1 (62.4/9.4)  
## mortality = Same as the National average:  
## :...Safety\_of\_care = Above the National average:  
## :...Readmissions = Below the National average: 1 (93/45.5)  
## : Readmissions = Same as the National average: 2 (45.2/13.6)  
## Safety\_of\_care = Same as the National average:  
## :...Readmissions = Below the National average: 1 (86.1/21.2)  
## Readmissions = Same as the National average:  
## :...coverage\_area = verygood: 2 (47.1/14.6)  
## coverage\_area in [good-less]:  
## :...coverage\_area = good: 1 (57.9/24)  
## coverage\_area in [medium-less]: 2 (100/44.8)  
##   
## SubTree [S1]  
##   
## Patient\_experience = Above the National average: 3 (57.8/21.6)  
## Patient\_experience = Same as the National average: 2 (55.3/21.2)  
##   
## SubTree [S2]  
##   
## Patient\_experience = Same as the National average: 2 (399.6/96.7)  
## Patient\_experience = Above the National average:  
## :...coverage\_person = verygood: 2 (332.6/153)  
## coverage\_person in [good-less]: 3 (382.2/176.5)  
##   
## ----- Trial 2: -----  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Readmissions in {Below the National average,Same as the National average}:  
## : :...Safety\_of\_care in {Below the National average,  
## : : : Same as the National average}: 1 (551.3/194.2)  
## : : Safety\_of\_care = Above the National average:  
## : : :...mortality in {Above the National average,  
## : : : Same as the National average}: 2 (196.1/91.6)  
## : : mortality = Below the National average: 1 (29/8)  
## : Readmissions = Above the National average:  
## : :...Safety\_of\_care = Above the National average: 3 (60.4/25.3)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Above the National average: 3 (17/8)  
## : mortality in {Below the National average,  
## : : Same as the National average}:  
## : :...Safety\_of\_care = Below the National average: 1 (56/19.3)  
## : Safety\_of\_care = Same as the National average: 2 (45.7/17.9)  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care = Above the National average: 2 (120.9/49.3)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Above the National average: 2 (54.6/19.4)  
## : mortality = Below the National average: 1 (23.6/5.4)  
## : mortality = Same as the National average:  
## : :...Patient\_experience = Above the National average: 2 (61.3/22.6)  
## : Patient\_experience = Same as the National average:  
## : :...Safety\_of\_care = Below the National average: 1 (30.1/9)  
## : Safety\_of\_care = Same as the National average: 2 (76.4/31.8)  
## Readmissions in {Above the National average,Same as the National average}:  
## :...Safety\_of\_care = Above the National average:  
## :...Patient\_experience = Above the National average: 3 (149.1/18.5)  
## : Patient\_experience = Same as the National average:  
## : :...mortality = Below the National average: 2 (24.4/5.4)  
## : mortality in {Above the National average,  
## : Same as the National average}: 3 (115/40.4)  
## Safety\_of\_care in {Below the National average,  
## : Same as the National average}:  
## :...mortality = Below the National average: 2 (73.8/39.4)  
## mortality in {Above the National average,  
## : Same as the National average}:  
## :...Patient\_experience = Same as the National average:  
## :...mortality = Above the National average: 3 (46.6/17.2)  
## : mortality = Same as the National average: 2 (550.7/192.4)  
## Patient\_experience = Above the National average:  
## :...mortality = Above the National average: 3 (21.9/3.6)  
## mortality = Same as the National average:  
## :...Readmissions = Above the National average: 3 (138.3/41.5)  
## Readmissions = Same as the National average:  
## :...coverage\_person in [good-less]: 2 (413.3/186.7)  
## coverage\_person = verygood:  
## :...coverage\_area = less: 3 (60/29.7)  
## coverage\_area in [verygood-medium]: [S1]  
##   
## SubTree [S1]  
##   
## Hospital\_Type = Acute Care Hospitals: 2 (65.1/27.7)  
## Hospital\_Type = Critical Access Hospitals: 3 (219.6/104.9)  
##   
## ----- Trial 3: -----  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Readmissions = Above the National average:  
## : :...mortality = Above the National average: 3 (30.2/13.1)  
## : : mortality in {Below the National average,Same as the National average}:  
## : : :...Safety\_of\_care = Below the National average: 1 (56.8/23.8)  
## : : Safety\_of\_care in {Above the National average,  
## : : Same as the National average}: 2 (95.8/42.4)  
## : Readmissions in {Below the National average,Same as the National average}:  
## : :...mortality = Below the National average: 1 (76.2/13.1)  
## : mortality in {Above the National average,Same as the National average}:  
## : :...Safety\_of\_care = Below the National average: 1 (162.4/41.5)  
## : Safety\_of\_care = Above the National average: 2 (199.9/105.6)  
## : Safety\_of\_care = Same as the National average:  
## : :...Readmissions = Below the National average: 1 (103.3/40.7)  
## : Readmissions = Same as the National average: 2 (234.3/96.2)  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average: 2 (363.9/176.6)  
## Readmissions in {Above the National average,Same as the National average}:  
## :...Safety\_of\_care = Below the National average:  
## :...mortality = Below the National average: 1 (31/9.2)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Readmissions = Same as the National average: 2 (62.7/29.3)  
## : Readmissions = Above the National average:  
## : :...mortality = Above the National average: 3 (22/6.4)  
## : mortality = Same as the National average: 2 (94.7/38.6)  
## Safety\_of\_care in {Above the National average,  
## : Same as the National average}:  
## :...Readmissions = Above the National average: 3 (381/105.3)  
## Readmissions = Same as the National average:  
## :...mortality = Below the National average: 2 (32.4/7.5)  
## mortality = Above the National average: 3 (32.4/7.9)  
## mortality = Same as the National average: [S1]  
##   
## SubTree [S1]  
##   
## Patient\_experience = Same as the National average: 2 (410.7/141.8)  
## Patient\_experience = Above the National average:  
## :...Safety\_of\_care = Above the National average: 3 (40.2/9.9)  
## Safety\_of\_care = Same as the National average:  
## :...count in {2,3,4,5,8,10,12}: 3 (78.5/30.7)  
## count in {6,7,9,14,15,16,18,21,26}: 2 (6.5/0.9)  
## count = 1:  
## :...coverage\_person = verygood:  
## :...State in {AR,IA,IL,NC,NY,WI}: 3 (87.8/38.7)  
## : State in {AK,AL,AZ,CA,CO,CT,DC,DE,FL,GA,HI,ID,IN,KS,KY,LA,MA,  
## : MD,ME,MI,MN,MO,MS,MT,ND,NE,NH,NJ,NM,NV,OH,OK,OR,PA,  
## : PR,RI,SC,SD,TN,TX,UT,VA,VT,WA,WV,WY}: 2 (268/126.3)  
## coverage\_person in [good-less]:  
## :...Hospital\_Type = Acute Care Hospitals:  
## :...coverage\_area in [verygood-good]: 3 (75.3/31.2)  
## : coverage\_area in [medium-less]: 2 (78/34.9)  
## Hospital\_Type = Critical Access Hospitals:  
## :...coverage\_person = good: 2 (140.9/64.5)  
## coverage\_person in [medium-less]: 3 (34.9/12.1)  
##   
## ----- Trial 4: -----  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...Readmissions = Below the National average: 1 (154.3/69.4)  
## : : Readmissions = Same as the National average: 2 (78.6/43.4)  
## : : Readmissions = Above the National average: 3 (62.6/29.2)  
## : Safety\_of\_care in {Below the National average,Same as the National average}:  
## : :...mortality = Above the National average: 2 (98.6/44.5)  
## : mortality = Below the National average: 1 (74/14.2)  
## : mortality = Same as the National average:  
## : :...Readmissions = Below the National average: 1 (134.9/32.4)  
## : Readmissions = Above the National average: 2 (73.2/32.1)  
## : Readmissions = Same as the National average:  
## : :...coverage\_area = verygood: 2 (71.2/26.4)  
## : coverage\_area in [good-less]: 1 (209.9/95.4)  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...Patient\_experience = Above the National average: 2 (89.4/45.1)  
## : : Patient\_experience = Same as the National average: 1 (161.8/67.3)  
## : Safety\_of\_care = Above the National average:  
## : :...mortality = Below the National average: 2 (14.3/6.5)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Patient\_experience = Above the National average: 3 (54.3/20.8)  
## : Patient\_experience = Same as the National average: 2 (55/24.8)  
## Readmissions in {Above the National average,Same as the National average}:  
## :...Safety\_of\_care = Below the National average: 2 (210.6/109.6)  
## Safety\_of\_care in {Above the National average,  
## : Same as the National average}:  
## :...mortality = Above the National average: 3 (64.9/12.1)  
## mortality in {Below the National average,  
## : Same as the National average}:  
## :...Readmissions = Above the National average: [S1]  
## Readmissions = Same as the National average:  
## :...mortality = Below the National average: 2 (30.9/8.3)  
## mortality = Same as the National average: [S2]  
##   
## SubTree [S1]  
##   
## Patient\_experience = Above the National average: 3 (158.3/31)  
## Patient\_experience = Same as the National average:  
## :...mortality = Below the National average: 2 (28/11.2)  
## mortality = Same as the National average: 3 (154.9/70.9)  
##   
## SubTree [S2]  
##   
## Patient\_experience = Above the National average: 3 (819/403.7)  
## Patient\_experience = Same as the National average: 2 (401.2/156.5)  
##   
## ----- Trial 5: -----  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Readmissions = Above the National average: 2 (185.4/91.7)  
## : Readmissions in {Below the National average,Same as the National average}:  
## : :...Safety\_of\_care = Above the National average:  
## : :...Readmissions = Below the National average: 1 (155.9/76.7)  
## : : Readmissions = Same as the National average: 2 (79.3/46.4)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Below the National average: 1 (42/4.5)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Safety\_of\_care = Below the National average: 1 (154.5/44)  
## : Safety\_of\_care = Same as the National average:  
## : :...Readmissions = Below the National average: 1 (101.2/46.4)  
## : Readmissions = Same as the National average: 2 (237.9/106.6)  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...mortality = Above the National average: 3 (42.1/13.9)  
## : : mortality in {Below the National average,  
## : : Same as the National average}: 2 (81.1/36.3)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Above the National average: 2 (56.2/24.9)  
## : mortality in {Below the National average,  
## : Same as the National average}: 1 (197.2/87.8)  
## Readmissions in {Above the National average,Same as the National average}:  
## :...Safety\_of\_care = Below the National average:  
## :...Readmissions = Above the National average: 2 (138.3/74)  
## : Readmissions = Same as the National average: 1 (73.3/34.1)  
## Safety\_of\_care in {Above the National average,  
## : Same as the National average}:  
## :...mortality = Above the National average: 3 (62.5/13.1)  
## mortality in {Below the National average,  
## : Same as the National average}:  
## :...Safety\_of\_care = Same as the National average:  
## :...Readmissions = Same as the National average: 2 (1145.8/528.4)  
## : Readmissions = Above the National average: [S1]  
## Safety\_of\_care = Above the National average: [S2]  
##   
## SubTree [S1]  
##   
## Patient\_experience = Above the National average: 3 (89.7/26.4)  
## Patient\_experience = Same as the National average: 2 (127/58.8)  
##   
## SubTree [S2]  
##   
## Patient\_experience = Above the National average: 3 (107.4/24.1)  
## Patient\_experience = Same as the National average:  
## :...mortality = Below the National average: 2 (22.8/5.7)  
## mortality = Same as the National average:  
## :...coverage\_person in [verygood-good]: 2 (30.7/11.8)  
## coverage\_person in [medium-less]: 3 (69.9/24.5)  
##   
## ----- Trial 6: -----  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...Readmissions = Below the National average: 1 (156.5/81.9)  
## : : Readmissions in {Above the National average,  
## : : : Same as the National average}:  
## : : :...coverage\_person = verygood: 1 (2.5)  
## : : coverage\_person in [good-less]:  
## : : :...mortality in {Above the National average,  
## : : : Same as the National average}: 3 (114.5/49.3)  
## : : mortality = Below the National average: 2 (30.9/13.2)  
## : Safety\_of\_care in {Below the National average,Same as the National average}:  
## : :...Readmissions = Above the National average:  
## : :...coverage\_person = verygood: 3 (2.5)  
## : : coverage\_person in [good-less]:  
## : : :...mortality = Above the National average: 2 (19.4/8.7)  
## : : mortality in {Below the National average,  
## : : : Same as the National average}:  
## : : :...Safety\_of\_care = Below the National average: 1 (60.2/25.4)  
## : : Safety\_of\_care = Same as the National average: 2 (42.2/21)  
## : Readmissions in {Below the National average,  
## : : Same as the National average}:  
## : :...Readmissions = Below the National average: 1 (174.4/31.9)  
## : Readmissions = Same as the National average:  
## : :...mortality = Above the National average: 2 (17.4/4.7)  
## : mortality = Below the National average: 1 (16.3)  
## : mortality = Same as the National average:  
## : :...Safety\_of\_care = Below the National average: 1 (43.1/13.1)  
## : Safety\_of\_care = Same as the National average:  
## : :...coverage\_person in [verygood-good]: 2 (80/33.5)  
## : coverage\_person in [medium-less]: 1 (152.9/69.2)  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...mortality = Below the National average: 1 (14.2/6.9)  
## : : mortality in {Above the National average,  
## : : : Same as the National average}:  
## : : :...Patient\_experience = Above the National average: 3 (54.6/22.7)  
## : : Patient\_experience = Same as the National average: 2 (55.2/26.7)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Below the National average: 1 (25.1/5.8)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Safety\_of\_care = Below the National average: 1 (92.1/39.2)  
## : Safety\_of\_care = Same as the National average: 2 (144.7/60.7)  
## Readmissions in {Above the National average,Same as the National average}:  
## :...Safety\_of\_care = Below the National average:  
## :...mortality = Below the National average: 1 (32.3/10.8)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Readmissions = Above the National average: 3 (121/59.1)  
## : Readmissions = Same as the National average: 2 (67.9/32.4)  
## Safety\_of\_care in {Above the National average,  
## : Same as the National average}:  
## :...Safety\_of\_care = Above the National average: 3 (233.7/66.3)  
## Safety\_of\_care = Same as the National average:  
## :...mortality = Below the National average: 2 (47.6/21.9)  
## mortality in {Above the National average,  
## : Same as the National average}:  
## :...Readmissions = Above the National average: 3 (183/61.1)  
## Readmissions = Same as the National average:  
## :...Patient\_experience = Above the National average:  
## :...coverage\_person = verygood: 2 (383.2/197.4)  
## : coverage\_person in [good-less]: 3 (444.3/209.3)  
## Patient\_experience = Same as the National average:  
## :...State in {AK,AL,AR,AZ,CA,CT,DC,DE,FL,HI,IL,KS,KY,  
## : LA,MA,MD,MO,MS,NC,NE,NH,NJ,NV,NY,OK,PR,  
## : RI,SD,TX,VT,WA,WI,WV,WY}: 2 (182.9/70)  
## State in {CO,GA,IA,ID,IN,ME,MI,MN,MT,ND,NM,OH,OR,  
## : PA,SC,TN,UT,VA}:  
## :...coverage\_person = verygood: 3 (61.5/24.8)  
## coverage\_person in [good-less]: 2 (102.8/44.2)  
##   
## ----- Trial 7: -----  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...Readmissions = Below the National average:  
## : : :...mortality in {Above the National average,  
## : : : : Same as the National average}: 2 (164.6/74.5)  
## : : : mortality = Below the National average: 1 (8.1)  
## : : Readmissions in {Above the National average,  
## : : : Same as the National average}:  
## : : :...coverage\_person = verygood: 1 (2.3)  
## : : coverage\_person in [good-less]:  
## : : :...mortality = Above the National average: 3 (38.7/11.6)  
## : : mortality in {Below the National average,  
## : : Same as the National average}: 2 (118.3/51.3)  
## : Safety\_of\_care in {Below the National average,Same as the National average}:  
## : :...Readmissions = Below the National average: 1 (156.7/20.7)  
## : Readmissions in {Above the National average,  
## : : Same as the National average}:  
## : :...mortality = Below the National average: 1 (48.6/11.3)  
## : mortality = Above the National average: 2 (35.6/13.3)  
## : mortality = Same as the National average:  
## : :...Safety\_of\_care = Below the National average: 1 (76.8/26.5)  
## : Safety\_of\_care = Same as the National average: 2 (256.3/110.1)  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...mortality = Above the National average: 3 (44.8/16.3)  
## : : mortality in {Below the National average,  
## : : Same as the National average}: 2 (66.7/16.7)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...Hospital\_Type = Critical Access Hospitals: 2 (5)  
## : Hospital\_Type = Acute Care Hospitals:  
## : :...Patient\_experience = Above the National average: 2 (73.6/23.9)  
## : Patient\_experience = Same as the National average:  
## : :...mortality = Above the National average: 2 (34.1/10.3)  
## : mortality in {Below the National average,  
## : Same as the National average}: 1 (121.2/42.8)  
## Readmissions in {Above the National average,Same as the National average}:  
## :...Patient\_experience = Same as the National average:  
## :...Safety\_of\_care = Below the National average: 2 (91.2/26.4)  
## : Safety\_of\_care in {Above the National average,  
## : : Same as the National average}:  
## : :...Readmissions = Same as the National average:  
## : :...mortality = Above the National average: 3 (18.6/6.5)  
## : : mortality in {Below the National average,  
## : : Same as the National average}: 2 (273.7/28.9)  
## : Readmissions = Above the National average:  
## : :...mortality = Above the National average: 3 (10.5)  
## : mortality in {Below the National average,  
## : : Same as the National average}: [S1]  
## Patient\_experience = Above the National average:  
## :...mortality = Above the National average: 3 (32/0.9)  
## mortality in {Below the National average,  
## : Same as the National average}:  
## :...Safety\_of\_care = Above the National average: 3 (93.6/18.7)  
## Safety\_of\_care in {Below the National average,  
## : Same as the National average}:  
## :...mortality = Below the National average: 2 (43.3/18.7)  
## mortality = Same as the National average:  
## :...Readmissions = Above the National average: 3 (114.1/34.7)  
## Readmissions = Same as the National average:  
## :...coverage\_person = verygood: 3 (415.5/200.1)  
## coverage\_person in [good-less]: 2 (490.6/230.4)  
##   
## SubTree [S1]  
##   
## Safety\_of\_care = Above the National average: 3 (42.5/11.2)  
## Safety\_of\_care = Same as the National average: 2 (139.9/61.4)  
##   
## ----- Trial 8: -----  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...Readmissions in {Above the National average,  
## : : : Same as the National average}: 2 (136.8/66.8)  
## : : Readmissions = Below the National average: 1 (170.9/81.1)  
## : Safety\_of\_care in {Below the National average,Same as the National average}:  
## : :...Readmissions = Below the National average: 1 (144.8/22.4)  
## : Readmissions in {Above the National average,  
## : : Same as the National average}:  
## : :...mortality = Above the National average: 2 (23.7/3.6)  
## : mortality = Below the National average: 1 (41.6/8)  
## : mortality = Same as the National average:  
## : :...Readmissions = Above the National average: 2 (67.4/19.5)  
## : Readmissions = Same as the National average:  
## : :...Safety\_of\_care = Below the National average: 1 (25.8)  
## : Safety\_of\_care = Same as the National average:  
## : :...coverage\_person = verygood: 2 (6.6)  
## : coverage\_person in [good-less]:  
## : :...State in {AK,AL,CO,CT,DC,DE,GA,HI,IA,ID,IN,KS,KY,  
## : : LA,MA,MD,ME,MI,MN,MT,NC,ND,NE,NH,NJ,OH,  
## : : PA,PR,RI,SC,SD,TN,TX,UT,VA,VT,WI,WV,  
## : : WY}: 2 (66.6/6.9)  
## : State in {AR,AZ,CA,FL,IL,MO,MS,NM,NV,NY,OK,OR,  
## : WA}: 1 (156/51.7)  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...mortality = Above the National average: 3 (39/13.4)  
## : : mortality in {Below the National average,  
## : : Same as the National average}: 2 (65.6/20.6)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Below the National average: 1 (18.8)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Safety\_of\_care = Below the National average: 1 (80.8/33)  
## : Safety\_of\_care = Same as the National average: 2 (129.9/39.3)  
## Readmissions in {Above the National average,Same as the National average}:  
## :...mortality = Below the National average: 2 (119.5/47.1)  
## mortality in {Above the National average,Same as the National average}:  
## :...State in {DC,MD}: 2 (0)  
## State = PR: 1 (9.2/1.6)  
## State in {AK,AL,AR,AZ,CA,CO,CT,DE,FL,GA,HI,IA,ID,IL,IN,KS,KY,LA,MA,  
## : ME,MI,MN,MO,MS,MT,NC,ND,NE,NH,NJ,NM,NV,NY,OH,OK,OR,PA,RI,  
## : SC,SD,TN,TX,UT,VA,VT,WA,WI,WV,WY}:  
## :...Safety\_of\_care = Above the National average: 3 (167.5/33)  
## Safety\_of\_care in {Below the National average,  
## : Same as the National average}:  
## :...Patient\_experience = Above the National average:  
## :...mortality = Above the National average: 3 (15.2/1.2)  
## : mortality = Same as the National average: [S1]  
## Patient\_experience = Same as the National average: [S2]  
##   
## SubTree [S1]  
##   
## Safety\_of\_care = Below the National average: 2 (80.7/27)  
## Safety\_of\_care = Same as the National average:  
## :...Readmissions = Above the National average: 3 (44.4)  
## Readmissions = Same as the National average:  
## :...coverage\_person = verygood: 2 (418.1/170.8)  
## coverage\_person in [good-less]: 3 (514.8/223.5)  
##   
## SubTree [S2]  
##   
## Readmissions = Same as the National average: 2 (211/7.2)  
## Readmissions = Above the National average:  
## :...Safety\_of\_care = Below the National average: 2 (44.5/12.3)  
## Safety\_of\_care = Same as the National average:  
## :...mortality = Above the National average: 3 (4.9)  
## mortality = Same as the National average:  
## :...Hospital\_Type = Critical Access Hospitals: 3 (11.5/2.8)  
## Hospital\_Type = Acute Care Hospitals:  
## :...coverage\_person = verygood: 3 (10.2/1.4)  
## coverage\_person in [good-less]:  
## :...State in {AK,AL,AR,AZ,CO,CT,FL,HI,ID,IL,IN,KS,MA,MI,MO,MS,  
## : MT,NC,ND,NE,NH,NJ,NM,NV,NY,RI,SD,UT,VA,WA,WV,  
## : WY}: 2 (37.3)  
## State in {CA,DE,GA,IA,KY,LA,ME,MN,OH,OK,OR,PA,SC,TN,TX,VT,  
## WI}: 3 (77.9/23.5)  
##   
## ----- Trial 9: -----  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...mortality = Above the National average:  
## : :...Readmissions in {Above the National average,  
## : : : Same as the National average}: 3 (72.7/31.4)  
## : : Readmissions = Below the National average: 1 (75.9/35.5)  
## : mortality in {Below the National average,Same as the National average}:  
## : :...Safety\_of\_care = Below the National average: 1 (141.8/30)  
## : Safety\_of\_care in {Above the National average,  
## : : Same as the National average}:  
## : :...Readmissions = Above the National average: 2 (51.7)  
## : Readmissions in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Below the National average: 1 (44.8/3.9)  
## : mortality = Same as the National average:  
## : :...Readmissions = Same as the National average: 2 (252.4/93.6)  
## : Readmissions = Below the National average:  
## : :...coverage\_person = less: 1 (48.1)  
## : coverage\_person in [verygood-medium]: [S1]  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care = Above the National average: 2 (82.9/22.4)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Below the National average: 1 (16.5)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Hospital\_Type = Critical Access Hospitals: 2 (3.9)  
## : Hospital\_Type = Acute Care Hospitals: [S2]  
## Readmissions in {Above the National average,Same as the National average}:  
## :...mortality = Above the National average: 3 (68/20.6)  
## mortality in {Below the National average,Same as the National average}:  
## :...Patient\_experience = Same as the National average:  
## :...mortality = Below the National average: 2 (34.2)  
## : mortality = Same as the National average:  
## : :...Readmissions = Same as the National average: 2 (253.1/32.2)  
## : Readmissions = Above the National average: [S3]  
## Patient\_experience = Above the National average:  
## :...Safety\_of\_care = Above the National average: 3 (81.5/9.7)  
## Safety\_of\_care in {Below the National average,  
## : Same as the National average}:  
## :...mortality = Below the National average: 2 (19.3)  
## mortality = Same as the National average:  
## :...count in {14,15,18,21}: 2 (0)  
## count in {4,5,10,12,16}: 3 (29.1/8.9)  
## count in {1,2,3,6,7,8,9,26}:  
## :...State in {AK,AL,AZ,CA,CO,CT,DC,DE,GA,HI,KS,KY,LA,  
## : MD,MO,MS,ND,NE,NJ,NM,OK,PR,RI,TN,TX,WA,  
## : WV}: 2 (396.2/87)  
## State in {AR,FL,IA,ID,IL,IN,MA,ME,MI,MN,MT,NC,NH,  
## : NV,NY,OH,OR,PA,SC,SD,UT,VA,VT,WI,WY}:  
## :...count in {2,3,6,7,8,9}: 3 (56.5/16.7)  
## count = 26: 2 (3.9)  
## count = 1: [S4]  
##   
## SubTree [S1]  
##   
## Safety\_of\_care = Above the National average: 2 (73.4/25.2)  
## Safety\_of\_care = Same as the National average: 1 (14.3)  
##   
## SubTree [S2]  
##   
## Patient\_experience = Above the National average: 2 (67/14.3)  
## Patient\_experience = Same as the National average:  
## :...mortality = Above the National average: 2 (39.2/11.4)  
## mortality = Same as the National average: 1 (126/49.7)  
##   
## SubTree [S3]  
##   
## Safety\_of\_care = Below the National average: 2 (21.1)  
## Safety\_of\_care = Above the National average: 3 (20)  
## Safety\_of\_care = Same as the National average:  
## :...coverage\_person = verygood: 3 (11.4/2.2)  
## coverage\_person in [good-less]:  
## :...Hospital\_Type = Acute Care Hospitals: 2 (118.2/47.6)  
## Hospital\_Type = Critical Access Hospitals: 3 (8.4/2.2)  
##   
## SubTree [S4]  
##   
## Readmissions = Above the National average: 3 (46.7/7.5)  
## Readmissions = Same as the National average:  
## :...coverage\_person in [medium-less]: 3 (72.6/18.4)  
## coverage\_person in [verygood-good]:  
## :...coverage\_person = good:  
## :...Hospital\_Type = Acute Care Hospitals: 3 (61.5/27.4)  
## : Hospital\_Type = Critical Access Hospitals: 2 (159/64.4)  
## coverage\_person = verygood:  
## :...Hospital\_Type = Acute Care Hospitals: 2 (12.5/2.3)  
## Hospital\_Type = Critical Access Hospitals:  
## :...coverage\_area in [verygood-medium]: 3 (192.3/66.6)  
## coverage\_area = less: 2 (15.6/4.5)  
##   
##   
## Evaluation on training data (3200 cases):  
##   
## Trial Decision Tree   
## ----- ----------------   
## Size Errors   
##   
## 0 54 791(24.7%)  
## 1 26 922(28.8%)  
## 2 25 1054(32.9%)  
## 3 26 956(29.9%)  
## 4 22 1080(33.8%)  
## 5 21 1028(32.1%)  
## 6 31 1047(32.7%)  
## 7 28 1026(32.1%)  
## 8 30 859(26.8%)  
## 9 36 904(28.3%)  
## boost 762(23.8%) <<  
##   
##   
## (a) (b) (c) <-classified as  
## ---- ---- ----  
## 501 188 4 (a): class 1  
## 119 1324 161 (b): class 2  
## 1 289 613 (c): class 3  
##   
##   
## Attribute usage:  
##   
## 100.00% mortality  
## 100.00% Readmissions  
## 100.00% Patient\_experience  
## 100.00% Safety\_of\_care  
## 60.47% State  
## 52.63% coverage\_person  
## 27.59% Hospital\_Type  
## 26.16% count  
## 21.97% coverage\_area  
##   
##   
## Time: 0.2 secs

p<- predict( model, testX, type="class" )  
sum( p == testy ) / length( p )\*100

## [1] NA

##using ctree  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
  
TreeModel = ctree(Hospital\_overall\_rating~., data = train)  
plot(TreeModel, type="simple")



x\_Pred <- predict(TreeModel,test)  
table(x\_Pred, test$Hospital\_overall\_rating)

##   
## x\_Pred 1 2 3  
## 1 109 34 1  
## 2 41 263 75  
## 3 3 53 130

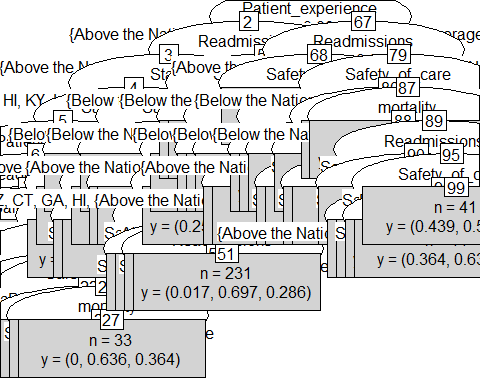
confusion\_matrix = table(x\_Pred, test$Hospital\_overall\_rating)  
TP = sum(diag(confusion\_matrix))  
Accuracy = TP/sum(confusion\_matrix)  
Accuracy\*100

## [1] 70.80395

## accuracy = 56.28  
  
## using 3 levels of ratings  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1$Hospital\_overall\_rating=as.numeric(dataset1$Hospital\_overall\_rating)  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating<=2)] = "1"  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating==3)] = "2"  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating==4)|(dataset1$Hospital\_overall\_rating==5)] = "3"  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type",  
 "mortality","coverage\_area", "Readmissions","Patient\_experience",  
 "Safety\_of\_care","count","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
TreeModel = ctree(Hospital\_overall\_rating~., data = train)  
plot(TreeModel, type="simple")



x\_Pred <- predict(TreeModel,test)  
table(x\_Pred, test$Hospital\_overall\_rating)

##   
## x\_Pred 1 2 3  
## 1 109 34 1  
## 2 41 263 75  
## 3 3 53 130

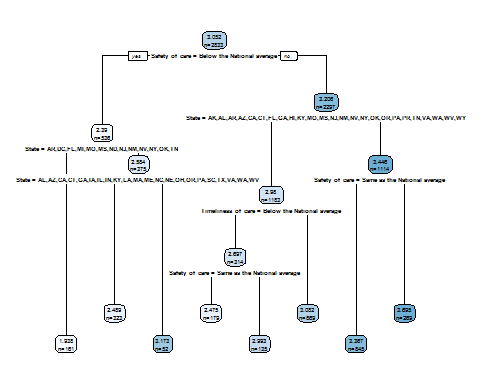
confusion\_matrix = table(x\_Pred, test$Hospital\_overall\_rating)  
TP = sum(diag(confusion\_matrix))  
Accuracy = TP/sum(confusion\_matrix)  
Accuracy\*100

## [1] 70.80395

##using rpart  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1$Hospital\_overall\_rating=as.numeric(dataset1$Hospital\_overall\_rating)  
dataset1<-dataset1[,c("State",  
 "Hospital\_Type","Hospital\_Ownership","Emergency\_Services",  
 "Hospital\_overall\_rating","mortality",  
 "Safety\_of\_care",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care","pop\_bin","capital")]  
  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
Tree <- rpart(Hospital\_overall\_rating ~ . , data = train, method = "anova")  
  
rpart.plot(Tree, extra=1, type=2, digits=4)



head(train)

## State Hospital\_Type Hospital\_Ownership  
## 941 IL Acute Care Hospitals Voluntary non-profit - Private  
## 1318 KY Acute Care Hospitals Proprietary  
## 2028 NJ Acute Care Hospitals Voluntary non-profit - Other  
## 3215 VA Acute Care Hospitals Government - State  
## 714 FL Acute Care Hospitals Voluntary non-profit - Other  
## 3178 UT Acute Care Hospitals Proprietary  
## Emergency\_Services Hospital\_overall\_rating  
## 941 Yes 4  
## 1318 Yes 4  
## 2028 Yes 4  
## 3215 Yes 2  
## 714 Yes 4  
## 3178 Yes 3  
## mortality Safety\_of\_care  
## 941 Same as the National average Same as the National average  
## 1318 Same as the National average Same as the National average  
## 2028 Same as the National average Same as the National average  
## 3215 Below the National average Above the National average  
## 714 Same as the National average Same as the National average  
## 3178 Same as the National average Same as the National average  
## Effectiveness\_of\_care Timeliness\_of\_care pop\_bin  
## 941 Same as the National average Above the National average small  
## 1318 Same as the National average Above the National average medium  
## 2028 Same as the National average Below the National average big  
## 3215 Same as the National average Below the National average medium  
## 714 Same as the National average Same as the National average verybig  
## 3178 Same as the National average Above the National average big  
## capital  
## 941 No  
## 1318 No  
## 2028 No  
## 3215 No  
## 714 No  
## 3178 No

# What variables are used in the best model?  
summary(Tree)

## Call:  
## rpart(formula = Hospital\_overall\_rating ~ ., data = train, method = "anova")  
## n= 2833   
##   
## CP nsplit rel error xerror xstd  
## 1 0.15722605 0 1.0000000 1.0003644 0.02652726  
## 2 0.06781832 1 0.8427739 0.8445452 0.02197220  
## 3 0.02554922 2 0.7749556 0.7869725 0.02069795  
## 4 0.01850394 3 0.7494064 0.7735483 0.02061123  
## 5 0.01194935 4 0.7309025 0.7590029 0.02020637  
## 6 0.01138272 5 0.7189531 0.7572606 0.02045524  
## 7 0.01120824 6 0.7075704 0.7564095 0.02046437  
## 8 0.01000000 7 0.6963622 0.7550791 0.02051643  
##   
## Variable importance  
## Safety\_of\_care State Timeliness\_of\_care   
## 53 33 8   
## Hospital\_Type pop\_bin mortality   
## 3 2 1   
##   
## Node number 1: 2833 observations, complexity param=0.1572261  
## mean=3.051535, MSE=0.6496561   
## left son=2 (536 obs) right son=3 (2297 obs)  
## Primary splits:  
## Safety\_of\_care splits as RLR, improve=0.15722610, (0 missing)  
## Timeliness\_of\_care splits as RLR, improve=0.10112620, (0 missing)  
## State splits as LLLLLRLLRLLRRRRRRLLR-RRRLLRRRRRLLLLRLRLLRRRLRRLRLRLL, improve=0.09418059, (0 missing)  
## mortality splits as RLR, improve=0.04371844, (0 missing)  
## pop\_bin splits as RRRL, improve=0.03369305, (0 missing)  
## Surrogate splits:  
## State splits as RRRRRRLLRRRRRRRRRRRR-RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR, agree=0.813, adj=0.009, (0 split)  
##   
## Node number 2: 536 observations, complexity param=0.02554922  
## mean=2.389925, MSE=0.6707194   
## left son=4 (161 obs) right son=5 (375 obs)  
## Primary splits:  
## State splits as RRLRRRRLRLR-RRRRRRRR-RLRLLRRLRRLLLLRLRR-RRRLRRR-RRR-, improve=0.13079830, (0 missing)  
## Hospital\_Ownership splits as -LLLRRR, improve=0.06427052, (39 missing)  
## Timeliness\_of\_care splits as RLR, improve=0.06028521, (0 missing)  
## mortality splits as RLR, improve=0.04591260, (0 missing)  
## Effectiveness\_of\_care splits as RLR, improve=0.01997790, (0 missing)  
##   
## Node number 3: 2297 observations, complexity param=0.06781832  
## mean=3.205921, MSE=0.5187634   
## left son=6 (1183 obs) right son=7 (1114 obs)  
## Primary splits:  
## State splits as LLLLLRL-RLLLRRRRRLRR-RRRLLRRRRRLLLLRLLLLRRRLRRLRLRLL, improve=0.10474830, (0 missing)  
## Timeliness\_of\_care splits as RLR, improve=0.05351132, (0 missing)  
## mortality splits as RLR, improve=0.03739073, (0 missing)  
## Hospital\_Ownership splits as -LLLRRR, improve=0.03325606, (238 missing)  
## Effectiveness\_of\_care splits as RLR, improve=0.02768733, (0 missing)  
## Surrogate splits:  
## Hospital\_Type splits as LR, agree=0.585, adj=0.145, (0 split)  
## pop\_bin splits as RRRL, agree=0.568, adj=0.109, (0 split)  
## Timeliness\_of\_care splits as RLR, agree=0.567, adj=0.107, (0 split)  
## Effectiveness\_of\_care splits as RLL, agree=0.522, adj=0.015, (0 split)  
## Emergency\_Services splits as RL, agree=0.516, adj=0.002, (0 split)  
##   
## Node number 4: 161 observations  
## mean=1.937888, MSE=0.5551483   
##   
## Node number 5: 375 observations, complexity param=0.01138272  
## mean=2.584, MSE=0.594944   
## left son=10 (323 obs) right son=11 (52 obs)  
## Primary splits:  
## State splits as RL-LLRL-R-L-LRLLRLLL-L-R--RL-LR----L-LL-RLR-LRL-LRL-, improve=0.09390068, (0 missing)  
## mortality splits as RLR, improve=0.07652588, (0 missing)  
## Timeliness\_of\_care splits as RLR, improve=0.05119869, (0 missing)  
## Hospital\_Ownership splits as -LLLRRR, improve=0.04518298, (29 missing)  
## Effectiveness\_of\_care splits as RLL, improve=0.01970111, (0 missing)  
##   
## Node number 6: 1183 observations, complexity param=0.01850394  
## mean=2.979713, MSE=0.5033923   
## left son=12 (314 obs) right son=13 (869 obs)  
## Primary splits:  
## Timeliness\_of\_care splits as RLR, improve=0.05718776, (0 missing)  
## State splits as RRLRR-L--LRR-----R------RR-----LLLL-RRRL---R--R-R-RR, improve=0.03967399, (0 missing)  
## mortality splits as RLR, improve=0.03703595, (0 missing)  
## Hospital\_Ownership splits as -LLLRRR, improve=0.03307330, (138 missing)  
## Safety\_of\_care splits as R-L, improve=0.02679752, (0 missing)  
## Surrogate splits:  
## State splits as RRRRR-L--RRR-----R------RR-----LRRL-RRRR---R--R-R-RR, agree=0.764, adj=0.111, (0 split)  
##   
## Node number 7: 1114 observations, complexity param=0.01194935  
## mean=3.44614, MSE=0.4230417   
## left son=14 (845 obs) right son=15 (269 obs)  
## Primary splits:  
## Safety\_of\_care splits as R-L, improve=0.04666660, (0 missing)  
## State splits as -----R--R---LRLRL-LL-RRR--LLLLR----R----RLR-LR-L-R--, improve=0.04341717, (0 missing)  
## mortality splits as RLL, improve=0.03662705, (0 missing)  
## Hospital\_Ownership splits as -RRLRRR, improve=0.02139459, (100 missing)  
## Effectiveness\_of\_care splits as RLR, improve=0.01125580, (0 missing)  
## Surrogate splits:  
## mortality splits as RLL, agree=0.771, adj=0.052, (0 split)  
## Timeliness\_of\_care splits as LRL, agree=0.764, adj=0.022, (0 split)  
## State splits as -----L--R---LLLLL-LL-LLL--LLLLL----L----RLL-LL-L-L--, agree=0.763, adj=0.019, (0 split)  
##   
## Node number 10: 323 observations  
## mean=2.489164, MSE=0.578056   
##   
## Node number 11: 52 observations  
## mean=3.173077, MSE=0.2969675   
##   
## Node number 12: 314 observations, complexity param=0.01120824  
## mean=2.697452, MSE=0.6122865   
## left son=24 (179 obs) right son=25 (135 obs)  
## Primary splits:  
## Safety\_of\_care splits as R-L, improve=0.10729590, (0 missing)  
## State splits as RR-RR-R--RRR-----R------RL-----RRLL-LRR----R--R-R-RL, improve=0.09556611, (0 missing)  
## mortality splits as RLL, improve=0.08947323, (0 missing)  
## Hospital\_Ownership splits as -LRRRRR, improve=0.06045515, (24 missing)  
## Effectiveness\_of\_care splits as RLR, improve=0.03972144, (0 missing)  
## Surrogate splits:  
## State splits as LR-RL-L--RLR-----L------RL-----LLLL-LLL----R--R-L-LL, agree=0.637, adj=0.156, (0 split)  
## mortality splits as RRL, agree=0.621, adj=0.119, (0 split)  
## Effectiveness\_of\_care splits as RLL, agree=0.580, adj=0.022, (0 split)  
##   
## Node number 13: 869 observations  
## mean=3.081703, MSE=0.4248551   
##   
## Node number 14: 845 observations  
## mean=3.366864, MSE=0.3695529   
##   
## Node number 15: 269 observations  
## mean=3.695167, MSE=0.5093075   
##   
## Node number 24: 179 observations  
## mean=2.47486, MSE=0.5063512   
##   
## Node number 25: 135 observations  
## mean=2.992593, MSE=0.5999451

x\_Pred <- predict(Tree, test)  
table(x\_Pred, test$Hospital\_overall\_rating)

##   
## x\_Pred 1 2 3 4 5  
## 1.93788819875776 12 24 10 3 0  
## 2.47486033519553 2 15 17 5 0  
## 2.48916408668731 4 32 28 3 0  
## 2.99259259259259 2 9 10 11 1  
## 3.08170310701956 1 29 146 52 4  
## 3.17307692307692 0 1 4 2 1  
## 3.36686390532544 1 16 113 69 5  
## 3.69516728624535 0 5 22 42 8

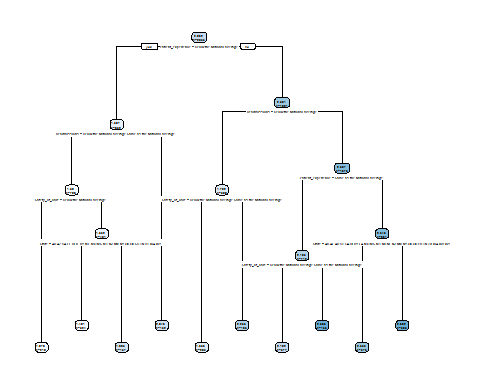
confusion\_matrix = table(x\_Pred, test$Hospital\_overall\_rating)  
TP = sum(diag(confusion\_matrix))  
Accuracy = TP/sum(confusion\_matrix)  
Accuracy\*100##Accuracy is very low 14.804%

## [1] 9.873061

## using 3 levels of ratings  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1$Hospital\_overall\_rating=as.numeric(dataset1$Hospital\_overall\_rating)  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating<=2)] = "1"  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating==3)] = "2"  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating==4)|(dataset1$Hospital\_overall\_rating==5)] = "3"  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type",  
 "mortality","coverage\_area", "Readmissions","Patient\_experience",  
 "Safety\_of\_care","count","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
Tree <- rpart(Hospital\_overall\_rating ~ . , data = train, method = "anova")  
  
rpart.plot(Tree, extra=1, type=2, digits=4)



head(train)

## Hospital\_overall\_rating State Hospital\_Type  
## 941 3 IL Acute Care Hospitals  
## 1318 3 KY Acute Care Hospitals  
## 2028 3 NJ Acute Care Hospitals  
## 3215 1 VA Acute Care Hospitals  
## 714 3 FL Acute Care Hospitals  
## 3178 2 UT Acute Care Hospitals  
## mortality coverage\_area  
## 941 Same as the National average good  
## 1318 Same as the National average verygood  
## 2028 Same as the National average verygood  
## 3215 Below the National average verygood  
## 714 Same as the National average good  
## 3178 Same as the National average medium  
## Readmissions Patient\_experience  
## 941 Same as the National average Same as the National average  
## 1318 Above the National average Same as the National average  
## 2028 Above the National average Below the National average  
## 3215 Below the National average Same as the National average  
## 714 Same as the National average Above the National average  
## 3178 Same as the National average Same as the National average  
## Safety\_of\_care count coverage\_person  
## 941 Same as the National average 1 verygood  
## 1318 Same as the National average 1 good  
## 2028 Same as the National average 1 medium  
## 3215 Above the National average 2 good  
## 714 Same as the National average 1 less  
## 3178 Same as the National average 1 medium  
## Effectiveness\_of\_care Timeliness\_of\_care  
## 941 Same as the National average Above the National average  
## 1318 Same as the National average Above the National average  
## 2028 Same as the National average Below the National average  
## 3215 Same as the National average Below the National average  
## 714 Same as the National average Same as the National average  
## 3178 Same as the National average Above the National average

# What variables are used in the best model?  
summary(Tree)

## Call:  
## rpart(formula = Hospital\_overall\_rating ~ ., data = train, method = "anova")  
## n= 2833   
##   
## CP nsplit rel error xerror xstd  
## 1 0.26585072 0 1.0000000 1.0003864 0.01883034  
## 2 0.06944924 1 0.7341493 0.7352121 0.01652921  
## 3 0.03801439 2 0.6647000 0.6663861 0.01456639  
## 4 0.03565260 3 0.6266857 0.6467873 0.01506596  
## 5 0.02221639 4 0.5910331 0.5979241 0.01451030  
## 6 0.02199463 5 0.5688167 0.5855127 0.01428449  
## 7 0.01948543 6 0.5468220 0.5572388 0.01359611  
## 8 0.01860339 7 0.5273366 0.5442918 0.01354464  
## 9 0.01204578 8 0.5087332 0.5227897 0.01337222  
## 10 0.01000000 9 0.4966874 0.5234410 0.01374157  
##   
## Variable importance  
## Patient\_experience Readmissions State   
## 45 19 14   
## Safety\_of\_care Timeliness\_of\_care count   
## 10 6 3   
## coverage\_person coverage\_area   
## 3 1   
##   
## Node number 1: 2833 observations, complexity param=0.2658507  
## mean=2.062478, MSE=0.4990969   
## left son=2 (852 obs) right son=3 (1981 obs)  
## Primary splits:  
## Patient\_experience splits as RLR, improve=0.26585070, (0 missing)  
## Readmissions splits as RLR, improve=0.20861810, (0 missing)  
## Safety\_of\_care splits as RLR, improve=0.13732400, (0 missing)  
## State splits as LLLLLRLLRLLRRRRRRLLR-RRRLLRRRRRLLLLRLRLLRRRLRRLRLRLL, improve=0.09645258, (0 missing)  
## Timeliness\_of\_care splits as RLR, improve=0.09223713, (0 missing)  
## Surrogate splits:  
## State splits as RRRLLRRLRLRRRRRRRRRR-RRRRRRRRRRLLLLRRRRRRRRRRRRRRRRR, agree=0.755, adj=0.185, (0 split)  
## Timeliness\_of\_care splits as RLR, agree=0.742, adj=0.142, (0 split)  
## Readmissions splits as RLR, agree=0.719, adj=0.067, (0 split)  
## count splits as RRLRRLRRRRLL-LLRL, agree=0.718, adj=0.062, (0 split)  
## coverage\_person splits as RRRL, agree=0.718, adj=0.062, (0 split)  
##   
## Node number 2: 852 observations, complexity param=0.03801439  
## mean=1.507042, MSE=0.3626265   
## left son=4 (700 obs) right son=5 (152 obs)  
## Primary splits:  
## Readmissions splits as RLL, improve=0.17397240, (0 missing)  
## Safety\_of\_care splits as RLL, improve=0.11897380, (0 missing)  
## State splits as RRLRRRLL-LLLRRLRRLRR-LRRRLRRRR-LLLLRRLR-RLRLRRRRLRRR, improve=0.08225411, (0 missing)  
## mortality splits as RLL, improve=0.04784650, (0 missing)  
## Timeliness\_of\_care splits as RLR, improve=0.04503690, (0 missing)  
## Surrogate splits:  
## State splits as LLLLLRLL-LLLLRLLLLLL-LLRLLLLLL-LLLLLLLL-LLRLLLLLRLLL, agree=0.837, adj=0.086, (0 split)  
##   
## Node number 3: 1981 observations, complexity param=0.06944924  
## mean=2.301363, MSE=0.3680395   
## left son=6 (308 obs) right son=7 (1673 obs)  
## Primary splits:  
## Readmissions splits as RLR, improve=0.13468520, (0 missing)  
## Patient\_experience splits as R-L, improve=0.11355350, (0 missing)  
## Safety\_of\_care splits as RLR, improve=0.08712373, (0 missing)  
## State splits as LLLLRRL-RLLRRRRRRLLL-RRRLLRLRLRLLLLRLLLLRRRLRRLRLRLL, improve=0.06272120, (0 missing)  
## mortality splits as RLR, improve=0.03885096, (0 missing)  
## Surrogate splits:  
## State splits as RRRRRRR-RRRRRRRRRRRL-RRRRRRRRRRLRRRRRRRRRRRRRRRRRRRR, agree=0.847, adj=0.016, (0 split)  
##   
## Node number 4: 700 observations, complexity param=0.02199463  
## mean=1.39, MSE=0.2893286   
## left son=8 (219 obs) right son=9 (481 obs)  
## Primary splits:  
## Safety\_of\_care splits as RLR, improve=0.15355310, (0 missing)  
## Readmissions splits as -LR, improve=0.09296893, (0 missing)  
## Timeliness\_of\_care splits as RLR, improve=0.07585395, (0 missing)  
## State splits as RRLLR-LL-LLLR-LRRLRR-LR-RLLRRR-LLLLRLLR-RL-LRRRRLRRL, improve=0.06227091, (0 missing)  
## mortality splits as RLL, improve=0.04832865, (0 missing)  
## Surrogate splits:  
## count splits as RRRLRLLLRLLR-RRRL, agree=0.713, adj=0.082, (0 split)  
## State splits as RRRRR-LL-RRRR-RRRRLR-RL-RRRRRR-RRRRRRRR-RR-RRRRRRRRR, agree=0.703, adj=0.050, (0 split)  
##   
## Node number 5: 152 observations  
## mean=2.046053, MSE=0.3465634   
##   
## Node number 6: 308 observations, complexity param=0.02221639  
## mean=1.782468, MSE=0.3715108   
## left son=12 (205 obs) right son=13 (103 obs)  
## Primary splits:  
## Safety\_of\_care splits as RLL, improve=0.27452560, (0 missing)  
## mortality splits as RLL, improve=0.11261090, (0 missing)  
## State splits as -LLLRLR-RRR-L-RLRLLR-RRRLL-L-LRR-LRRRRLLRL-RR-LL-RL-, improve=0.09269926, (0 missing)  
## Patient\_experience splits as R-L, improve=0.08208471, (0 missing)  
## coverage\_person splits as LLLR, improve=0.02991398, (0 missing)  
## Surrogate splits:  
## State splits as -LLLRLL-RLL-L-LRLLLL-LLLLL-L-LLL-LLLLRLLRL-LL-LL-LL-, agree=0.688, adj=0.068, (0 split)  
## count splits as LLLLLLLLLRLL-RL-L, agree=0.672, adj=0.019, (0 split)  
##   
## Node number 7: 1673 observations, complexity param=0.0356526  
## mean=2.396892, MSE=0.3087052   
## left son=14 (712 obs) right son=15 (961 obs)  
## Primary splits:  
## Patient\_experience splits as R-L, improve=0.09760739, (0 missing)  
## Safety\_of\_care splits as RLL, improve=0.08464203, (0 missing)  
## Readmissions splits as R-L, improve=0.07702134, (0 missing)  
## State splits as LLLLRRL-RRLRRRRRRLLR-RRRRLLRRLRLLRLRLLRLRRRLRRRRLRLL, improve=0.05854519, (0 missing)  
## mortality splits as RLL, improve=0.03774114, (0 missing)  
## Surrogate splits:  
## State splits as LRRLLRL-LLLRRLRRRRRL-RRRLRLRRRRLLRRRRLRRRRRLRRLRLRRR, agree=0.610, adj=0.083, (0 split)  
## Timeliness\_of\_care splits as RLR, agree=0.599, adj=0.058, (0 split)  
## Safety\_of\_care splits as RLR, agree=0.586, adj=0.028, (0 split)  
## count splits as RRRRRRRLLRLL-LRRR, agree=0.581, adj=0.015, (0 split)  
##   
## Node number 8: 219 observations  
## mean=1.077626, MSE=0.07159984   
##   
## Node number 9: 481 observations, complexity param=0.01204578  
## mean=1.532225, MSE=0.3238057   
## left son=18 (324 obs) right son=19 (157 obs)  
## Primary splits:  
## State splits as RRLLL-R--LRLR-LRRLRR-LR-LLLRRR-LLRLRLLR-RL-LRRRLLRRL, improve=0.10935460, (0 missing)  
## mortality splits as RLL, improve=0.08396887, (0 missing)  
## Readmissions splits as -LR, improve=0.08234016, (0 missing)  
## Timeliness\_of\_care splits as RLR, improve=0.06334829, (0 missing)  
## Safety\_of\_care splits as R-L, improve=0.05390310, (0 missing)  
## Surrogate splits:  
## coverage\_area splits as RLLL, agree=0.728, adj=0.166, (0 split)  
## coverage\_person splits as RRLL, agree=0.690, adj=0.051, (0 split)  
## count splits as LLLLLLLLRLLR-LLRL, agree=0.682, adj=0.025, (0 split)  
##   
## Node number 12: 205 observations  
## mean=1.556098, MSE=0.2566092   
##   
## Node number 13: 103 observations  
## mean=2.23301, MSE=0.295221   
##   
## Node number 14: 712 observations, complexity param=0.01860339  
## mean=2.195225, MSE=0.2610446   
## left son=28 (577 obs) right son=29 (135 obs)  
## Primary splits:  
## Safety\_of\_care splits as RLL, improve=0.14152360, (0 missing)  
## Readmissions splits as R-L, improve=0.11179910, (0 missing)  
## mortality splits as RLL, improve=0.06765730, (0 missing)  
## State splits as LLLLRRL-RLRRLRLRLRLR-LRLRLLLRLRRLLLRLLR-LLRRRRLRLLLL, improve=0.05310569, (0 missing)  
## count splits as LLLLLLLRRRRL-RRRL, improve=0.02849364, (0 missing)  
## Surrogate splits:  
## count splits as LLLLLLLLLLLL-RLRL, agree=0.815, adj=0.022, (0 split)  
## mortality splits as LRL, agree=0.812, adj=0.007, (0 split)  
##   
## Node number 15: 961 observations, complexity param=0.01948543  
## mean=2.546306, MSE=0.2915602   
## left son=30 (376 obs) right son=31 (585 obs)  
## Primary splits:  
## State splits as LLLRRRL-RRLLRRRRRLLR-RRRLLLRLLRLLRLRLLRLRRRLLRRRLRLL, improve=0.09833082, (0 missing)  
## Safety\_of\_care splits as RLL, improve=0.08198540, (0 missing)  
## Readmissions splits as R-L, improve=0.08012185, (0 missing)  
## Effectiveness\_of\_care splits as RLR, improve=0.04627965, (0 missing)  
## mortality splits as RLR, improve=0.03980122, (0 missing)  
## Surrogate splits:  
## coverage\_person splits as LRRR, agree=0.668, adj=0.152, (0 split)  
## coverage\_area splits as RRRL, agree=0.647, adj=0.098, (0 split)  
## count splits as RRRRRRLRLRLR-LRLR, agree=0.626, adj=0.045, (0 split)  
## Effectiveness\_of\_care splits as RLR, agree=0.626, adj=0.045, (0 split)  
##   
## Node number 18: 324 observations  
## mean=1.401235, MSE=0.2834553   
##   
## Node number 19: 157 observations  
## mean=1.802548, MSE=0.2985922   
##   
## Node number 28: 577 observations  
## mean=2.102253, MSE=0.2200469   
##   
## Node number 29: 135 observations  
## mean=2.592593, MSE=0.2414266   
##   
## Node number 30: 376 observations  
## mean=2.335106, MSE=0.3025973   
##   
## Node number 31: 585 observations  
## mean=2.682051, MSE=0.2373702

x\_Pred <- predict(Tree, test)  
table(x\_Pred, test$Hospital\_overall\_rating)

##   
## x\_Pred 1 2 3  
## 1.07762557077626 48 8 1  
## 1.40123456790123 43 30 3  
## 1.55609756097561 21 31 4  
## 1.80254777070064 19 26 3  
## 2.04605263157895 10 17 8  
## 2.10225303292894 6 111 23  
## 2.23300970873786 2 19 6  
## 2.33510638297872 0 47 61  
## 2.59259259259259 0 10 20  
## 2.68205128205128 4 51 77

confusion\_matrix = table(x\_Pred, test$Hospital\_overall\_rating)  
TP = sum(diag(confusion\_matrix))  
Accuracy = TP/sum(confusion\_matrix)

Naive Bayes Classifier

##Using Naive bayes Classifier  
  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type","Emergency\_Services",  
 "mortality","coverage\_area",  
 "Safety\_of\_care","count","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
  
model <- naiveBayes(Hospital\_overall\_rating ~ ., data = train)  
class(model)

## [1] "naiveBayes"

summary(model)

## Length Class Mode   
## apriori 5 table numeric   
## tables 10 -none- list   
## levels 5 -none- character  
## call 4 -none- call

print(model)

##   
## Naive Bayes Classifier for Discrete Predictors  
##   
## Call:  
## naiveBayes.default(x = X, y = Y, laplace = laplace)  
##   
## A-priori probabilities:  
## Y  
## 1 2 3 4 5   
## 0.03035651 0.18990469 0.49699965 0.26332510 0.01941405   
##   
## Conditional probabilities:  
## State  
## Y AK AL AR AZ CA  
## 1 0.0000000000 0.0000000000 0.0232558140 0.0232558140 0.0697674419  
## 2 0.0000000000 0.0223048327 0.0241635688 0.0185873606 0.1301115242  
## 3 0.0042613636 0.0284090909 0.0205965909 0.0163352273 0.0759943182  
## 4 0.0000000000 0.0120643432 0.0080428954 0.0080428954 0.0603217158  
## 5 0.0000000000 0.0181818182 0.0000000000 0.0363636364 0.0727272727  
## State  
## Y CO CT DC DE FL  
## 1 0.0000000000 0.0000000000 0.0116279070 0.0000000000 0.0697674419  
## 2 0.0018587361 0.0148698885 0.0018587361 0.0000000000 0.0966542751  
## 3 0.0149147727 0.0078125000 0.0007102273 0.0007102273 0.0411931818  
## 4 0.0281501340 0.0013404826 0.0000000000 0.0040214477 0.0241286863  
## 5 0.0181818182 0.0000000000 0.0000000000 0.0000000000 0.0181818182  
## State  
## Y GA HI IA ID IL  
## 1 0.0348837209 0.0000000000 0.0000000000 0.0000000000 0.0348837209  
## 2 0.0315985130 0.0037174721 0.0055762082 0.0000000000 0.0483271375  
## 3 0.0362215909 0.0035511364 0.0220170455 0.0049715909 0.0404829545  
## 4 0.0174262735 0.0026809651 0.0308310992 0.0080428954 0.0549597855  
## 5 0.0181818182 0.0181818182 0.0000000000 0.0181818182 0.0727272727  
## State  
## Y IN KS KY LA MA  
## 1 0.0116279070 0.0000000000 0.0232558140 0.0116279070 0.0116279070  
## 2 0.0130111524 0.0055762082 0.0223048327 0.0167286245 0.0074349442  
## 3 0.0255681818 0.0184659091 0.0255681818 0.0241477273 0.0198863636  
## 4 0.0495978552 0.0214477212 0.0160857909 0.0160857909 0.0160857909  
## 5 0.1090909091 0.0363636364 0.0000000000 0.0181818182 0.0000000000  
## State  
## Y MD ME MI MN MO  
## 1 0.0000000000 0.0000000000 0.0348837209 0.0000000000 0.0116279070  
## 2 0.0000000000 0.0037174721 0.0260223048 0.0000000000 0.0260223048  
## 3 0.0000000000 0.0085227273 0.0213068182 0.0234375000 0.0255681818  
## 4 0.0000000000 0.0134048257 0.0415549598 0.0469168901 0.0214477212  
## 5 0.0000000000 0.0181818182 0.0909090909 0.0181818182 0.0181818182  
## State  
## Y MS MT NC ND NE  
## 1 0.0232558140 0.0000000000 0.0000000000 0.0116279070 0.0000000000  
## 2 0.0278810409 0.0018587361 0.0241635688 0.0000000000 0.0037174721  
## 3 0.0170454545 0.0063920455 0.0326704545 0.0063920455 0.0170454545  
## 4 0.0107238606 0.0067024129 0.0281501340 0.0067024129 0.0134048257  
## 5 0.0000000000 0.0000000000 0.0181818182 0.0000000000 0.0000000000  
## State  
## Y NH NJ NM NV NY  
## 1 0.0000000000 0.0581395349 0.0116279070 0.0581395349 0.3255813953  
## 2 0.0000000000 0.0446096654 0.0130111524 0.0074349442 0.0873605948  
## 3 0.0085227273 0.0134943182 0.0099431818 0.0063920455 0.0241477273  
## 4 0.0147453083 0.0093833780 0.0000000000 0.0026809651 0.0160857909  
## 5 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0181818182  
## State  
## Y OH OK OR PA PR  
## 1 0.0232558140 0.0116279070 0.0116279070 0.0232558140 0.0000000000  
## 2 0.0204460967 0.0148698885 0.0055762082 0.0594795539 0.0130111524  
## 3 0.0312500000 0.0305397727 0.0184659091 0.0397727273 0.0007102273  
## 4 0.0723860590 0.0174262735 0.0107238606 0.0415549598 0.0000000000  
## 5 0.0545454545 0.0000000000 0.0000000000 0.0363636364 0.0000000000  
## State  
## Y RI SC SD TN TX  
## 1 0.0000000000 0.0116279070 0.0000000000 0.0232558140 0.0232558140  
## 2 0.0018587361 0.0167286245 0.0000000000 0.0297397770 0.0464684015  
## 3 0.0028409091 0.0127840909 0.0021306818 0.0291193182 0.0788352273  
## 4 0.0026809651 0.0174262735 0.0107238606 0.0147453083 0.0764075067  
## 5 0.0181818182 0.0363636364 0.0181818182 0.0000000000 0.1454545455  
## State  
## Y UT VA VT WA WI  
## 1 0.0000000000 0.0116279070 0.0000000000 0.0000000000 0.0000000000  
## 2 0.0000000000 0.0204460967 0.0037174721 0.0167286245 0.0055762082  
## 3 0.0099431818 0.0227272727 0.0021306818 0.0241477273 0.0205965909  
## 4 0.0147453083 0.0187667560 0.0053619303 0.0080428954 0.0723860590  
## 5 0.0000000000 0.0000000000 0.0000000000 0.0181818182 0.0363636364  
## State  
## Y WV WY  
## 1 0.0000000000 0.0000000000  
## 2 0.0092936803 0.0055762082  
## 3 0.0149147727 0.0063920455  
## 4 0.0013404826 0.0040214477  
## 5 0.0000000000 0.0000000000  
##   
## Hospital\_Type  
## Y Acute Care Hospitals Critical Access Hospitals  
## 1 0.98837209 0.01162791  
## 2 0.96282528 0.03717472  
## 3 0.80326705 0.19673295  
## 4 0.76675603 0.23324397  
## 5 0.98181818 0.01818182  
##   
## Emergency\_Services  
## Y No Yes  
## 1 0.02325581 0.97674419  
## 2 0.01486989 0.98513011  
## 3 0.01917614 0.98082386  
## 4 0.02278820 0.97721180  
## 5 0.07272727 0.92727273  
##   
## mortality  
## Y Above the National average Below the National average  
## 1 0.12790698 0.25581395  
## 2 0.08364312 0.19144981  
## 3 0.08664773 0.08238636  
## 4 0.14075067 0.02949062  
## 5 0.50909091 0.00000000  
## mortality  
## Y Same as the National average  
## 1 0.61627907  
## 2 0.72490706  
## 3 0.83096591  
## 4 0.82975871  
## 5 0.49090909  
##   
## coverage\_area  
## Y verygood good medium less  
## 1 0.2209302 0.4767442 0.1046512 0.1976744  
## 2 0.3029740 0.3420074 0.1245353 0.2304833  
## 3 0.2947443 0.2471591 0.2151989 0.2428977  
## 4 0.3176944 0.2573727 0.2198391 0.2050938  
## 5 0.3090909 0.1818182 0.2181818 0.2909091  
##   
## Safety\_of\_care  
## Y Above the National average Below the National average  
## 1 0.03488372 0.82558140  
## 2 0.12825279 0.42565056  
## 3 0.17968750 0.13707386  
## 4 0.31501340 0.05630027  
## 5 0.74545455 0.01818182  
## Safety\_of\_care  
## Y Same as the National average  
## 1 0.13953488  
## 2 0.44609665  
## 3 0.68323864  
## 4 0.62868633  
## 5 0.23636364  
##   
## count  
## Y 1 2 3 4 5  
## 1 0.290697674 0.139534884 0.093023256 0.058139535 0.034883721  
## 2 0.552044610 0.165427509 0.068773234 0.040892193 0.024163569  
## 3 0.710227273 0.129261364 0.049715909 0.026278409 0.012784091  
## 4 0.714477212 0.136729223 0.033512064 0.025469169 0.013404826  
## 5 0.472727273 0.127272727 0.109090909 0.072727273 0.036363636  
## count  
## Y 6 7 8 9 10  
## 1 0.104651163 0.046511628 0.011627907 0.023255814 0.011627907  
## 2 0.044609665 0.018587361 0.007434944 0.005576208 0.011152416  
## 3 0.017755682 0.007812500 0.004971591 0.006392045 0.007102273  
## 4 0.017426273 0.005361930 0.016085791 0.006702413 0.009383378  
## 5 0.036363636 0.000000000 0.000000000 0.018181818 0.018181818  
## count  
## Y 12 14 15 16 18  
## 1 0.151162791 0.000000000 0.000000000 0.011627907 0.000000000  
## 2 0.014869888 0.013011152 0.000000000 0.001858736 0.007434944  
## 3 0.002840909 0.005681818 0.000000000 0.006392045 0.004971591  
## 4 0.005361930 0.001340483 0.000000000 0.005361930 0.004021448  
## 5 0.018181818 0.018181818 0.000000000 0.000000000 0.000000000  
## count  
## Y 21 26  
## 1 0.000000000 0.023255814  
## 2 0.003717472 0.020446097  
## 3 0.002840909 0.004971591  
## 4 0.002680965 0.002680965  
## 5 0.072727273 0.000000000  
##   
## coverage\_person  
## Y verygood good medium less  
## 1 0.00000000 0.12790698 0.20930233 0.66279070  
## 2 0.05204461 0.21189591 0.31040892 0.42565056  
## 3 0.19531250 0.31178977 0.25284091 0.24005682  
## 4 0.19302949 0.27613941 0.30697051 0.22386059  
## 5 0.00000000 0.18181818 0.49090909 0.32727273  
##   
## Effectiveness\_of\_care  
## Y Above the National average Below the National average  
## 1 0.03488372 0.20930233  
## 2 0.02044610 0.12267658  
## 3 0.02485795 0.06036932  
## 4 0.05495979 0.02144772  
## 5 0.09090909 0.01818182  
## Effectiveness\_of\_care  
## Y Same as the National average  
## 1 0.75581395  
## 2 0.85687732  
## 3 0.91477273  
## 4 0.92359249  
## 5 0.89090909  
##   
## Timeliness\_of\_care  
## Y Above the National average Below the National average  
## 1 0.03488372 0.77906977  
## 2 0.15985130 0.49070632  
## 3 0.30752841 0.18892045  
## 4 0.37131367 0.13941019  
## 5 0.21818182 0.14545455  
## Timeliness\_of\_care  
## Y Same as the National average  
## 1 0.18604651  
## 2 0.34944238  
## 3 0.50355114  
## 4 0.48927614  
## 5 0.63636364

x\_pred <- predict(model, newdata = test)  
table(x\_Pred, test$Hospital\_overall\_rating)

##   
## x\_Pred 1 2 3 4 5  
## 1.07762557077626 14 34 8 1 0  
## 1.40123456790123 4 39 30 3 0  
## 1.55609756097561 1 20 31 4 0  
## 1.80254777070064 2 17 26 3 0  
## 2.04605263157895 1 9 17 8 0  
## 2.10225303292894 0 6 111 22 1  
## 2.23300970873786 0 2 19 6 0  
## 2.33510638297872 0 0 47 55 6  
## 2.59259259259259 0 0 10 19 1  
## 2.68205128205128 0 4 51 66 11

confusion\_matrix = table(x\_Pred, test$Hospital\_overall\_rating)  
TP = sum(diag(confusion\_matrix))  
Accuracy = TP/sum(confusion\_matrix)  
Accuracy\*100##35.89% accuracy

## [1] 12.2708

##using 3 levels of ratings  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1$Hospital\_overall\_rating=as.numeric(dataset1$Hospital\_overall\_rating)  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating<=2)] = "1"  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating==3)] = "2"  
levels(dataset1$Hospital\_overall\_rating)=dataset1$Hospital\_overall\_rating[(dataset1$Hospital\_overall\_rating==4)|(dataset1$Hospital\_overall\_rating==5)] = "3"  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type",  
 "mortality","coverage\_area", "Readmissions","Patient\_experience",  
 "Safety\_of\_care","count","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
  
model <- naiveBayes(Hospital\_overall\_rating ~ ., data = train)  
class(model)

## [1] "naiveBayes"

summary(model)

## Length Class Mode   
## apriori 3 table numeric   
## tables 11 -none- list   
## levels 3 -none- character  
## call 4 -none- call

print(model)

##   
## Naive Bayes Classifier for Discrete Predictors  
##   
## Call:  
## naiveBayes.default(x = X, y = Y, laplace = laplace)  
##   
## A-priori probabilities:  
## Y  
## 1 2 3   
## 0.2202612 0.4969996 0.2827391   
##   
## Conditional probabilities:  
## State  
## Y AK AL AR AZ CA  
## 1 0.0000000000 0.0192307692 0.0240384615 0.0192307692 0.1217948718  
## 2 0.0042613636 0.0284090909 0.0205965909 0.0163352273 0.0759943182  
## 3 0.0000000000 0.0124843945 0.0074906367 0.0099875156 0.0611735331  
## State  
## Y CO CT DC DE FL  
## 1 0.0016025641 0.0128205128 0.0032051282 0.0000000000 0.0929487179  
## 2 0.0149147727 0.0078125000 0.0007102273 0.0007102273 0.0411931818  
## 3 0.0274656679 0.0012484395 0.0000000000 0.0037453184 0.0237203496  
## State  
## Y GA HI IA ID IL  
## 1 0.0320512821 0.0032051282 0.0048076923 0.0000000000 0.0464743590  
## 2 0.0362215909 0.0035511364 0.0220170455 0.0049715909 0.0404829545  
## 3 0.0174781523 0.0037453184 0.0287141074 0.0087390762 0.0561797753  
## State  
## Y IN KS KY LA MA  
## 1 0.0128205128 0.0048076923 0.0224358974 0.0160256410 0.0080128205  
## 2 0.0255681818 0.0184659091 0.0255681818 0.0241477273 0.0198863636  
## 3 0.0536828964 0.0224719101 0.0149812734 0.0162297129 0.0149812734  
## State  
## Y MD ME MI MN MO  
## 1 0.0000000000 0.0032051282 0.0272435897 0.0000000000 0.0240384615  
## 2 0.0000000000 0.0085227273 0.0213068182 0.0234375000 0.0255681818  
## 3 0.0000000000 0.0137328340 0.0449438202 0.0449438202 0.0212234707  
## State  
## Y MS MT NC ND NE  
## 1 0.0272435897 0.0016025641 0.0208333333 0.0016025641 0.0032051282  
## 2 0.0170454545 0.0063920455 0.0326704545 0.0063920455 0.0170454545  
## 3 0.0099875156 0.0062421973 0.0274656679 0.0062421973 0.0124843945  
## State  
## Y NH NJ NM NV NY  
## 1 0.0000000000 0.0464743590 0.0128205128 0.0144230769 0.1201923077  
## 2 0.0085227273 0.0134943182 0.0099431818 0.0063920455 0.0241477273  
## 3 0.0137328340 0.0087390762 0.0000000000 0.0024968789 0.0162297129  
## State  
## Y OH OK OR PA PR  
## 1 0.0208333333 0.0144230769 0.0064102564 0.0544871795 0.0112179487  
## 2 0.0312500000 0.0305397727 0.0184659091 0.0397727273 0.0007102273  
## 3 0.0711610487 0.0162297129 0.0099875156 0.0411985019 0.0000000000  
## State  
## Y RI SC SD TN TX  
## 1 0.0016025641 0.0160256410 0.0000000000 0.0288461538 0.0432692308  
## 2 0.0028409091 0.0127840909 0.0021306818 0.0291193182 0.0788352273  
## 3 0.0037453184 0.0187265918 0.0112359551 0.0137328340 0.0811485643  
## State  
## Y UT VA VT WA WI  
## 1 0.0000000000 0.0192307692 0.0032051282 0.0144230769 0.0048076923  
## 2 0.0099431818 0.0227272727 0.0021306818 0.0241477273 0.0205965909  
## 3 0.0137328340 0.0174781523 0.0049937578 0.0087390762 0.0699126092  
## State  
## Y WV WY  
## 1 0.0080128205 0.0048076923  
## 2 0.0149147727 0.0063920455  
## 3 0.0012484395 0.0037453184  
##   
## Hospital\_Type  
## Y Acute Care Hospitals Critical Access Hospitals  
## 1 0.96634615 0.03365385  
## 2 0.80326705 0.19673295  
## 3 0.78152310 0.21847690  
##   
## mortality  
## Y Above the National average Below the National average  
## 1 0.08974359 0.20032051  
## 2 0.08664773 0.08238636  
## 3 0.16604245 0.02746567  
## mortality  
## Y Same as the National average  
## 1 0.70993590  
## 2 0.83096591  
## 3 0.80649189  
##   
## coverage\_area  
## Y verygood good medium less  
## 1 0.2916667 0.3605769 0.1217949 0.2259615  
## 2 0.2947443 0.2471591 0.2151989 0.2428977  
## 3 0.3171036 0.2521848 0.2197253 0.2109863  
##   
## Readmissions  
## Y Above the National average Below the National average  
## 1 0.05128205 0.61378205  
## 2 0.18821023 0.18039773  
## 3 0.44444444 0.04494382  
## Readmissions  
## Y Same as the National average  
## 1 0.33493590  
## 2 0.63139205  
## 3 0.51061174  
##   
## Patient\_experience  
## Y Above the National average Below the National average  
## 1 0.06730769 0.75000000  
## 2 0.33096591 0.23863636  
## 3 0.70911361 0.05992509  
## Patient\_experience  
## Y Same as the National average  
## 1 0.18269231  
## 2 0.43039773  
## 3 0.23096130  
##   
## Safety\_of\_care  
## Y Above the National average Below the National average  
## 1 0.1153846 0.4807692  
## 2 0.1796875 0.1370739  
## 3 0.3445693 0.0536829  
## Safety\_of\_care  
## Y Same as the National average  
## 1 0.4038462  
## 2 0.6832386  
## 3 0.6017478  
##   
## count  
## Y 1 2 3 4 5  
## 1 0.516025641 0.161858974 0.072115385 0.043269231 0.025641026  
## 2 0.710227273 0.129261364 0.049715909 0.026278409 0.012784091  
## 3 0.697877653 0.136079900 0.038701623 0.028714107 0.014981273  
## count  
## Y 6 7 8 9 10  
## 1 0.052884615 0.022435897 0.008012821 0.008012821 0.011217949  
## 2 0.017755682 0.007812500 0.004971591 0.006392045 0.007102273  
## 3 0.018726592 0.004993758 0.014981273 0.007490637 0.009987516  
## count  
## Y 12 14 15 16 18  
## 1 0.033653846 0.011217949 0.000000000 0.003205128 0.006410256  
## 2 0.002840909 0.005681818 0.000000000 0.006392045 0.004971591  
## 3 0.006242197 0.002496879 0.000000000 0.004993758 0.003745318  
## count  
## Y 21 26  
## 1 0.003205128 0.020833333  
## 2 0.002840909 0.004971591  
## 3 0.007490637 0.002496879  
##   
## coverage\_person  
## Y verygood good medium less  
## 1 0.04487179 0.20032051 0.29647436 0.45833333  
## 2 0.19531250 0.31178977 0.25284091 0.24005682  
## 3 0.17977528 0.26966292 0.31960050 0.23096130  
##   
## Effectiveness\_of\_care  
## Y Above the National average Below the National average  
## 1 0.02243590 0.13461538  
## 2 0.02485795 0.06036932  
## 3 0.05742821 0.02122347  
## Effectiveness\_of\_care  
## Y Same as the National average  
## 1 0.84294872  
## 2 0.91477273  
## 3 0.92134831  
##   
## Timeliness\_of\_care  
## Y Above the National average Below the National average  
## 1 0.1426282 0.5304487  
## 2 0.3075284 0.1889205  
## 3 0.3607990 0.1398252  
## Timeliness\_of\_care  
## Y Same as the National average  
## 1 0.3269231  
## 2 0.5035511  
## 3 0.4993758

x\_pred <- predict(model, newdata = test)  
table(x\_Pred, test$Hospital\_overall\_rating)

##   
## x\_Pred 1 2 3  
## 1.07762557077626 48 8 1  
## 1.40123456790123 43 30 3  
## 1.55609756097561 21 31 4  
## 1.80254777070064 19 26 3  
## 2.04605263157895 10 17 8  
## 2.10225303292894 6 111 23  
## 2.23300970873786 2 19 6  
## 2.33510638297872 0 47 61  
## 2.59259259259259 0 10 20  
## 2.68205128205128 4 51 77

confusion\_matrix = table(x\_Pred, test$Hospital\_overall\_rating)  
TP = sum(diag(confusion\_matrix))  
Accuracy = TP/sum(confusion\_matrix)  
Accuracy\*100##10.05Accuracy

## [1] 11.56559

## KNN

## using KNN Clustrering  
library(knncat)  
  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
dataset2=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==TRUE)## 24% od data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type","Emergency\_Services",  
 "mortality","coverage\_area",  
 "Safety\_of\_care","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
table(dataset1$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 108 669 1758 933 74 0

dataset1$Hospital\_Type <- factor(dataset1$Hospital\_Type,levels = c("Acute Care Hospitals","Critical Access Hospitals"))  
table(dataset1$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 108 669 1758 933 74 0

table(dataset1$State,useNA = "always")

##   
## AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL   
## 7 77 54 54 285 51 28 7 6 170 110 12 71 19 159   
## IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE   
## 107 58 82 80 56 0 32 114 79 86 62 23 94 16 48   
## NH NJ NM NV NY OH OK OR PA PR RI SC SD TN TX   
## 25 64 27 26 154 140 77 52 148 9 11 53 14 94 256   
## UT VA VT WA WI WV WY <NA>   
## 29 77 12 66 108 37 16 0

dataset1$State=factor(dataset1$State,levels = c("AK","AL","AR","AZ","CA","CO","CT","DC","DE","FL","GA",  
 "HI","IA","ID","IL","IN","KS","KY","LA","MA","ME",  
 "MI","MN","MO","MS","MT","NC","ND","NE","NH","NJ","NM",  
 "NV","NY","OH","OK","OR","PA","PR","RI","SC","SD","TN","TX",  
 "UT","VA","VT","WA","WI","WV","WY"))  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.7))  
train1 <- dataset1[x\_train,]  
test1<-dataset1[-x\_train,]  
predictions= knncat(train1,test1,k=3)  
TP = sum(diag(predictions$misclass.mat))  
Accuracy = TP/sum(predictions$misclass.mat)  
Accuracy\*100## accuracy=53.35

## [1] 53.3396

comparing models

##Cross Validating C5.0  
  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type","Emergency\_Services",  
 "mortality","coverage\_area",  
 "Safety\_of\_care","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
library(caret)

##   
## Attaching package: 'caret'

## The following object is masked from 'package:survival':  
##   
## cluster

library(klaR)

## Loading required package: MASS

##   
## Attaching package: 'MASS'

## The following objects are masked from 'package:raster':  
##   
## area, select

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

# define training control  
train\_control <- trainControl(method="cv", number=10, savePredictions = TRUE)  
  
# train the model  
model\_C5.0 <- train(Hospital\_overall\_rating~., data=train, trControl=train\_control, method="C5.0")

## Loading required package: plyr

## -------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

## -------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following object is masked from 'package:modeltools':  
##   
## empty

## The following object is masked from 'package:maps':  
##   
## ozone

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

## The following objects are masked from 'package:Hmisc':  
##   
## is.discrete, summarize

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 5 for this object. Predictions generated using 5  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 3 for this object. Predictions generated using 3  
## trials  
  
## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 3 for this object. Predictions generated using 3  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials  
  
## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials  
  
## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 3 for this object. Predictions generated using 3  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 5 for this object. Predictions generated using 5  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 3 for this object. Predictions generated using 3  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 1 for this object. Predictions generated using 1  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 6 for this object. Predictions generated using 6  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 7 for this object. Predictions generated using 7  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 6 for this object. Predictions generated using 6  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 7 for this object. Predictions generated using 7  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 5 for this object. Predictions generated using 5  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 7 for this object. Predictions generated using 7  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 1 for this object. Predictions generated using 1  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 6 for this object. Predictions generated using 6  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 1 for this object. Predictions generated using 1  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials  
  
## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 5 for this object. Predictions generated using 5  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials  
  
## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 1 for this object. Predictions generated using 1  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 5 for this object. Predictions generated using 5  
## trials  
  
## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 5 for this object. Predictions generated using 5  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 3 for this object. Predictions generated using 3  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 1 for this object. Predictions generated using 1  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 5 for this object. Predictions generated using 5  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 4 for this object. Predictions generated using 4  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 3 for this object. Predictions generated using 3  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 7 for this object. Predictions generated using 7  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 6 for this object. Predictions generated using 6  
## trials  
  
## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 6 for this object. Predictions generated using 6  
## trials

## Warning in predict.C5.0(modelFit, newdata, trial = submodels$trials[j]):  
## 'trials' should be <= 3 for this object. Predictions generated using 3  
## trials

# summarize results  
print(model\_C5.0)

## C5.0   
##   
## 2833 samples  
## 9 predictor  
## 5 classes: '1', '2', '3', '4', '5'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 2551, 2550, 2549, 2548, 2550, 2550, ...   
## Resampling results across tuning parameters:  
##   
## model winnow trials Accuracy Kappa   
## rules FALSE 1 0.5442683 0.2363714  
## rules FALSE 10 0.5580705 0.2583151  
## rules FALSE 20 0.5580705 0.2583151  
## rules TRUE 1 0.5509846 0.2372726  
## rules TRUE 10 0.5517125 0.2339797  
## rules TRUE 20 0.5517125 0.2339797  
## tree FALSE 1 0.5457003 0.2449648  
## tree FALSE 10 0.5616152 0.2631413  
## tree FALSE 20 0.5616152 0.2631413  
## tree TRUE 1 0.5488818 0.2361433  
## tree TRUE 10 0.5538488 0.2416984  
## tree TRUE 20 0.5538488 0.2416984  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final values used for the model were trials = 10, model = tree  
## and winnow = FALSE.

x\_pred1 <- predict(model\_C5.0, newdata = test)  
table(x\_pred1, test$Hospital\_overall\_rating)

##   
## x\_pred1 1 2 3 4 5  
## 1 1 2 2 0 0  
## 2 15 60 29 6 0  
## 3 5 63 257 111 8  
## 4 1 6 62 69 11  
## 5 0 0 0 1 0

confusionMatrix(data = x\_pred1, test$Hospital\_overall\_rating,mode = "prec\_recall")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 1 2 3 4 5  
## 1 1 2 2 0 0  
## 2 15 60 29 6 0  
## 3 5 63 257 111 8  
## 4 1 6 62 69 11  
## 5 0 0 0 1 0  
##   
## Overall Statistics  
##   
## Accuracy : 0.5458   
## 95% CI : (0.5084, 0.5829)  
## No Information Rate : 0.4937   
## P-Value [Acc > NIR] : 0.003045   
##   
## Kappa : 0.2512   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Precision 0.200000 0.54545 0.5788 0.46309 0.00000  
## Recall 0.045455 0.45802 0.7343 0.36898 0.00000  
## F1 0.074074 0.49793 0.6474 0.41071 NaN  
## Prevalence 0.031030 0.18477 0.4937 0.26375 0.02680  
## Detection Rate 0.001410 0.08463 0.3625 0.09732 0.00000  
## Detection Prevalence 0.007052 0.15515 0.6262 0.21016 0.00141  
## Balanced Accuracy 0.519816 0.68576 0.6067 0.60786 0.49928

## validating ctree  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type","Emergency\_Services",  
 "mortality","coverage\_area",  
 "Safety\_of\_care","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
library(caret)  
library(klaR)  
# define training control  
train\_control <- trainControl(method="cv", number=10, savePredictions = TRUE)  
# fix the parameters of the algorithm  
grid <- expand.grid(.fL=c(0), .usekernel=c(FALSE))  
# train the model  
model\_ctree <- train(Hospital\_overall\_rating~., data=train, trControl=train\_control, method="ctree")  
# summarize results  
print(model\_ctree)

## Conditional Inference Tree   
##   
## 2833 samples  
## 9 predictor  
## 5 classes: '1', '2', '3', '4', '5'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 2550, 2550, 2550, 2551, 2549, 2549, ...   
## Resampling results across tuning parameters:  
##   
## mincriterion Accuracy Kappa   
## 0.01 0.5429873 0.2236125  
## 0.50 0.5429611 0.2156058  
## 0.99 0.5369551 0.1927140  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was mincriterion = 0.01.

x\_pred1 <- predict(model\_ctree, newdata = test,type="raw")  
table(x\_pred1, test$Hospital\_overall\_rating)

##   
## x\_pred1 1 2 3 4 5  
## 1 3 2 0 0 0  
## 2 14 42 26 6 1  
## 3 5 81 274 121 7  
## 4 0 6 50 60 11  
## 5 0 0 0 0 0

confusionMatrix(data = x\_pred1, test$Hospital\_overall\_rating,mode = "prec\_recall")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 1 2 3 4 5  
## 1 3 2 0 0 0  
## 2 14 42 26 6 1  
## 3 5 81 274 121 7  
## 4 0 6 50 60 11  
## 5 0 0 0 0 0  
##   
## Overall Statistics  
##   
## Accuracy : 0.5346   
## 95% CI : (0.4971, 0.5718)  
## No Information Rate : 0.4937   
## P-Value [Acc > NIR] : 0.01613   
##   
## Kappa : 0.2105   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Precision 0.600000 0.47191 0.5615 0.47244 NA  
## Recall 0.136364 0.32061 0.7829 0.32086 0.0000  
## F1 0.222222 0.38182 0.6539 0.38217 NA  
## Prevalence 0.031030 0.18477 0.4937 0.26375 0.0268  
## Detection Rate 0.004231 0.05924 0.3865 0.08463 0.0000  
## Detection Prevalence 0.007052 0.12553 0.6883 0.17913 0.0000  
## Balanced Accuracy 0.566726 0.61965 0.5934 0.59625 0.5000

## Cross Validation  
##cross validating the model  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type","Emergency\_Services",  
 "mortality","coverage\_area",  
 "Safety\_of\_care","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
library(caret)  
library(klaR)  
# define training control  
train\_control <- trainControl(method="cv", number=10, savePredictions = TRUE)  
# fix the parameters of the algorithm  
grid <- expand.grid(.fL=c(0), .usekernel=c(FALSE))  
# train the model  
model\_rpart <- train(Hospital\_overall\_rating~., data=train, trControl=train\_control, method="rpart")  
# summarize results  
print(model\_rpart)

## CART   
##   
## 2833 samples  
## 9 predictor  
## 5 classes: '1', '2', '3', '4', '5'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 2550, 2550, 2550, 2551, 2549, 2549, ...   
## Resampling results across tuning parameters:  
##   
## cp Accuracy Kappa   
## 0.01169591 0.5248969 0.16161741  
## 0.01543860 0.5037464 0.09612018  
## 0.02526316 0.4952569 0.05652657  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was cp = 0.01169591.

x\_pred1 <- predict(model\_rpart, newdata = test,type="raw")  
table(x\_pred1, test$Hospital\_overall\_rating)

##   
## x\_pred1 1 2 3 4 5  
## 1 0 0 0 0 0  
## 2 16 34 19 3 0  
## 3 6 97 331 184 19  
## 4 0 0 0 0 0  
## 5 0 0 0 0 0

confusionMatrix(data = x\_pred1, test$Hospital\_overall\_rating,mode = "prec\_recall")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 1 2 3 4 5  
## 1 0 0 0 0 0  
## 2 16 34 19 3 0  
## 3 6 97 331 184 19  
## 4 0 0 0 0 0  
## 5 0 0 0 0 0  
##   
## Overall Statistics  
##   
## Accuracy : 0.5148   
## 95% CI : (0.4773, 0.5522)  
## No Information Rate : 0.4937   
## P-Value [Acc > NIR] : 0.138   
##   
## Kappa : 0.0977   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Precision NA 0.47222 0.5196 NA NA  
## Recall 0.00000 0.25954 0.9457 0.0000 0.0000  
## F1 NA 0.33498 0.6707 NA NA  
## Prevalence 0.03103 0.18477 0.4937 0.2638 0.0268  
## Detection Rate 0.00000 0.04795 0.4669 0.0000 0.0000  
## Detection Prevalence 0.00000 0.10155 0.8984 0.0000 0.0000  
## Balanced Accuracy 0.50000 0.59690 0.5467 0.5000 0.5000

##cross validating the nb model  
  
dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type","Emergency\_Services",  
 "mortality","coverage\_area",  
 "Safety\_of\_care","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care")]  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
library(caret)  
library(klaR)  
# define training control  
train\_control <- trainControl(method="cv", number=10, savePredictions = TRUE)  
# fix the parameters of the algorithm  
grid <- expand.grid(.fL=c(0), .usekernel=c(FALSE))  
# train the model  
model\_nb <- train(Hospital\_overall\_rating~., data=train, trControl=train\_control, method="nb")

## Warning: model fit failed for Fold01: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateCO, StateCT, StateDC, StateDE, StateHI, StateIA, StateID, StateKS, StateKY, StateMA, StateMD, StateME, StateMN, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateOK, StateOR, StatePR, StateRI, StateSD, StateTN, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, Hospital\_TypeCritical.Access.Hospitals, mortalityBelow.the.National.average

## Warning: model fit failed for Fold02: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateCO, StateCT, StateDC, StateDE, StateHI, StateIA, StateID, StateKS, StateKY, StateLA, StateMA, StateMD, StateME, StateMN, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateOK, StateOR, StatePR, StateRI, StateSC, StateSD, StateTN, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, Hospital\_TypeCritical.Access.Hospitals, mortalityBelow.the.National.average

## Warning: model fit failed for Fold03: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateAZ, StateCO, StateCT, StateDC, StateDE, StateHI, StateIA, StateID, StateKS, StateKY, StateMA, StateMD, StateME, StateMN, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateOK, StateOR, StatePR, StateRI, StateSD, StateTN, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, mortalityBelow.the.National.average

## Warning: model fit failed for Fold04: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateCO, StateCT, StateDC, StateDE, StateHI, StateIA, StateID, StateKS, StateKY, StateMA, StateMD, StateME, StateMN, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateOK, StateOR, StatePR, StateRI, StateSD, StateTN, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, mortalityBelow.the.National.average, Safety\_of\_careBelow.the.National.average

## Warning: model fit failed for Fold05: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateCO, StateCT, StateDC, StateDE, StateHI, StateIA, StateID, StateKS, StateKY, StateMA, StateMD, StateME, StateMN, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateNY, StateOK, StateOR, StatePR, StateRI, StateSD, StateTN, StateTX, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, mortalityBelow.the.National.average, Effectiveness\_of\_careBelow.the.National.average

## Warning: model fit failed for Fold06: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateCO, StateCT, StateDC, StateDE, StateHI, StateIA, StateID, StateKS, StateKY, StateMA, StateMD, StateME, StateMN, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateOK, StateOR, StatePR, StateRI, StateSD, StateTN, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, mortalityBelow.the.National.average

## Warning: model fit failed for Fold07: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateCO, StateCT, StateDC, StateDE, StateFL, StateHI, StateIA, StateID, StateKS, StateKY, StateMA, StateMD, StateME, StateMN, StateMO, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateOK, StateOR, StatePR, StateRI, StateSD, StateTN, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, mortalityBelow.the.National.average

## Warning: model fit failed for Fold08: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateCO, StateCT, StateDC, StateDE, StateHI, StateIA, StateID, StateKS, StateKY, StateMA, StateMD, StateME, StateMN, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateOK, StateOR, StatePR, StateRI, StateSD, StateTN, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, mortalityBelow.the.National.average

## Warning: model fit failed for Fold09: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateCO, StateCT, StateDC, StateDE, StateHI, StateIA, StateID, StateKS, StateKY, StateMA, StateMD, StateME, StateMN, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateOK, StateOR, StatePR, StateRI, StateSD, StateTN, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, mortalityBelow.the.National.average

## Warning: model fit failed for Fold10: usekernel=FALSE, fL=0, adjust=1 Error in NaiveBayes.default(x, y, usekernel = FALSE, fL = param$fL, ...) :   
## Zero variances for at least one class in variables: StateAL, StateAR, StateCO, StateCT, StateDC, StateDE, StateGA, StateHI, StateIA, StateID, StateIN, StateKS, StateKY, StateLA, StateMA, StateMD, StateME, StateMN, StateMS, StateMT, StateNC, StateND, StateNE, StateNH, StateNJ, StateNM, StateNV, StateOK, StateOR, StatePR, StateRI, StateSD, StateTN, StateUT, StateVA, StateVT, StateWA, StateWI, StateWV, StateWY, mortalityBelow.the.National.average

## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info =  
## trainInfo, : There were missing values in resampled performance measures.

## Warning in train.default(x, y, weights = w, ...): missing values found in  
## aggregated results

# summarize results  
print(model\_nb)

## Naive Bayes   
##   
## 2833 samples  
## 9 predictor  
## 5 classes: '1', '2', '3', '4', '5'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 2550, 2550, 2550, 2551, 2549, 2549, ...   
## Resampling results across tuning parameters:  
##   
## usekernel Accuracy Kappa  
## FALSE NaN NaN   
## TRUE 0.4970027 0   
##   
## Tuning parameter 'fL' was held constant at a value of 0  
## Tuning  
## parameter 'adjust' was held constant at a value of 1  
## Accuracy was used to select the optimal model using the largest value.  
## The final values used for the model were fL = 0, usekernel = TRUE  
## and adjust = 1.

x\_pred1 <- predict(model\_nb, newdata = test,type="raw")  
table(x\_pred1, test$Hospital\_overall\_rating)

##   
## x\_pred1 1 2 3 4 5  
## 1 0 0 0 0 0  
## 2 0 0 0 0 0  
## 3 22 131 350 187 19  
## 4 0 0 0 0 0  
## 5 0 0 0 0 0

confusionMatrix(data = x\_pred1, test$Hospital\_overall\_rating,mode = "prec\_recall")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 1 2 3 4 5  
## 1 0 0 0 0 0  
## 2 0 0 0 0 0  
## 3 22 131 350 187 19  
## 4 0 0 0 0 0  
## 5 0 0 0 0 0  
##   
## Overall Statistics  
##   
## Accuracy : 0.4937   
## 95% CI : (0.4562, 0.5311)  
## No Information Rate : 0.4937   
## P-Value [Acc > NIR] : 0.5149   
##   
## Kappa : 0   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Precision NA NA 0.4937 NA NA  
## Recall 0.00000 0.0000 1.0000 0.0000 0.0000  
## F1 NA NA 0.6610 NA NA  
## Prevalence 0.03103 0.1848 0.4937 0.2638 0.0268  
## Detection Rate 0.00000 0.0000 0.4937 0.0000 0.0000  
## Detection Prevalence 0.00000 0.0000 1.0000 0.0000 0.0000  
## Balanced Accuracy 0.50000 0.5000 0.5000 0.5000 0.5000

##cross validating the knn model  
  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
set.seed(1)  
x\_train <- sample(nrow(dataset1), floor(nrow(dataset1)\*0.8))  
train <- dataset1[x\_train,]  
test<-dataset1[-x\_train,]  
library(caret)  
library(klaR)  
# define training control  
train\_control <- trainControl(method="cv", number=10, savePredictions = TRUE)  
# fix the parameters of the algorithm  
grid <- expand.grid(.fL=c(0), .usekernel=c(FALSE))  
# train the model  
model\_knn <- train(Hospital\_overall\_rating~., data=train, trControl=train\_control, method="knn")  
# summarize results  
print(model\_knn)

## k-Nearest Neighbors   
##   
## 2833 samples  
## 9 predictor  
## 5 classes: '1', '2', '3', '4', '5'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 2550, 2550, 2550, 2551, 2549, 2549, ...   
## Resampling results across tuning parameters:  
##   
## k Accuracy Kappa   
## 5 0.5182058 0.1981315  
## 7 0.5270023 0.2052344  
## 9 0.5273943 0.2015296  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was k = 9.

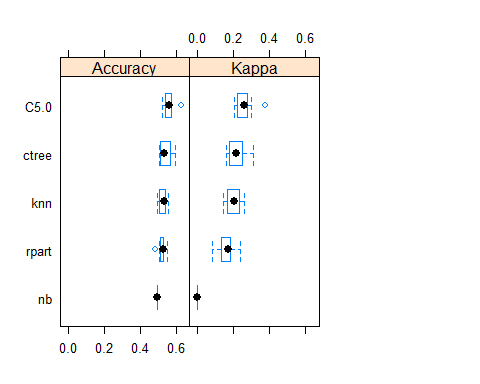
x\_pred1 <- predict(model\_knn, newdata = test,type="raw")  
table(x\_pred1, test$Hospital\_overall\_rating)

##   
## x\_pred1 1 2 3 4 5  
## 1 4 3 1 0 0  
## 2 14 49 34 7 1  
## 3 3 69 277 129 10  
## 4 1 10 38 51 8  
## 5 0 0 0 0 0

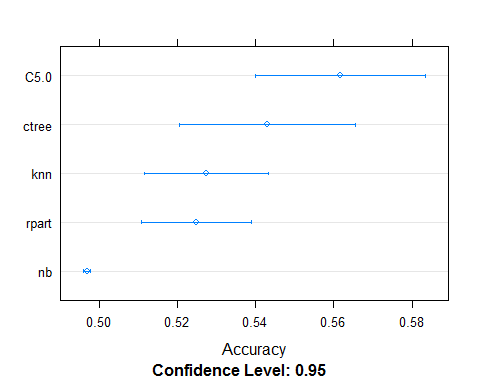
confusionMatrix(data = x\_pred1, test$Hospital\_overall\_rating,mode = "prec\_recall")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 1 2 3 4 5  
## 1 4 3 1 0 0  
## 2 14 49 34 7 1  
## 3 3 69 277 129 10  
## 4 1 10 38 51 8  
## 5 0 0 0 0 0  
##   
## Overall Statistics  
##   
## Accuracy : 0.5374   
## 95% CI : (0.4999, 0.5746)  
## No Information Rate : 0.4937   
## P-Value [Acc > NIR] : 0.01097   
##   
## Kappa : 0.219   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Precision 0.500000 0.46667 0.5676 0.47222 NA  
## Recall 0.181818 0.37405 0.7914 0.27273 0.0000  
## F1 0.266667 0.41525 0.6611 0.34576 NA  
## Prevalence 0.031030 0.18477 0.4937 0.26375 0.0268  
## Detection Rate 0.005642 0.06911 0.3907 0.07193 0.0000  
## Detection Prevalence 0.011283 0.14810 0.6883 0.15233 0.0000  
## Balanced Accuracy 0.587998 0.63858 0.6018 0.58177 0.5000

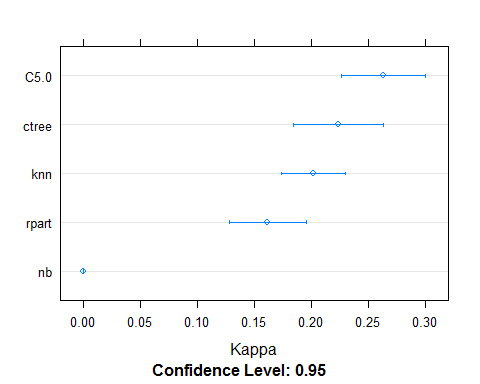
##identifying best model  
resamps <- resamples(list(nb = model\_nb,  
 ctree=model\_ctree,  
 rpart= model\_rpart,  
 knn = model\_knn,  
 C5.0 = model\_C5.0))  
  
bwplot(resamps, layout = c(3, 1))



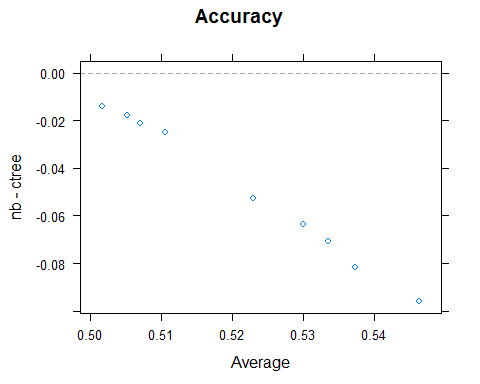
dotplot(resamps, metric = "Accuracy")



dotplot(resamps, metric = "Kappa")



xyplot(resamps, what = "BlandAltman")##best models

 ##Model Implementation

dataset1=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==FALSE)##76% of data  
names(dataset1)

## [1] "Provider\_ID"   
## [2] "Hospital\_Name"   
## [3] "City"   
## [4] "State"   
## [5] "zip"   
## [6] "Hospital\_Type"   
## [7] "Hospital\_Ownership"   
## [8] "Emergency\_Services"   
## [9] "Meets\_criteria\_for\_meaningful\_use\_of\_EHRs"  
## [10] "Hospital\_overall\_rating"   
## [11] "mortality"   
## [12] "Safety\_of\_care"   
## [13] "Readmissions"   
## [14] "Patient\_experience"   
## [15] "Effectiveness\_of\_care"   
## [16] "Timeliness\_of\_care"   
## [17] "Efficient\_use\_of\_medical\_imaging"   
## [18] "latitude"   
## [19] "longitude"   
## [20] "fips"   
## [21] "population"   
## [22] "name"   
## [23] "capital"   
## [24] "geoid"   
## [25] "land\_area"   
## [26] "area\_bin"   
## [27] "pop\_bin"   
## [28] "count"   
## [29] "hospitals\_person"   
## [30] "hospital\_area"   
## [31] "average\_rating"   
## [32] "coverage\_person"   
## [33] "coverage\_area"

dataset2=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==TRUE)## 24% od data  
table(dataset1$Hospital\_overall\_rating)

##   
## 1 2 3 4 5   
## 108 669 1758 933 74

dataset1$Hospital\_overall\_rating=as.numeric(dataset1$Hospital\_overall\_rating)  
dataset1<-dataset1[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type",  
 "mortality","coverage\_area", "Readmissions","Patient\_experience",  
 "Safety\_of\_care","count","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care","latitude","longitude","City","name")]  
dataset1$Hospital\_overall\_rating=as.factor(dataset1$Hospital\_overall\_rating)  
nrow(dataset1)

## [1] 3542

names(dataset1)

## [1] "Hospital\_overall\_rating" "State"   
## [3] "Hospital\_Type" "mortality"   
## [5] "coverage\_area" "Readmissions"   
## [7] "Patient\_experience" "Safety\_of\_care"   
## [9] "count" "coverage\_person"   
## [11] "Effectiveness\_of\_care" "Timeliness\_of\_care"   
## [13] "latitude" "longitude"   
## [15] "City" "name"

set.seed(12)  
crx <- dataset1[ sample( nrow( dataset1 ) ), ]  
X <- crx[,2:11]  
y <- crx[,1]  
trainX <- X[1:3200,]  
trainy <- y[1:3200]  
testX <- X[3201:3579,]  
testy <- y[3201:3579]  
library(C50)  
  
model <- C50::C5.0( trainX, trainy )  
summary( model )

##   
## Call:  
## C5.0.default(x = trainX, y = trainy)  
##   
##   
## C5.0 [Release 2.07 GPL Edition] Mon Apr 10 10:53:45 2017  
## -------------------------------  
##   
## Class specified by attribute `outcome'  
##   
## Read 3200 cases (11 attributes) from undefined.data  
##   
## Decision tree:  
##   
## Patient\_experience = Below the National average:  
## :...Safety\_of\_care = Below the National average:  
## : :...Readmissions = Above the National average:  
## : : :...mortality = Above the National average: [S1]  
## : : : mortality in {Below the National average,  
## : : : : Same as the National average}:  
## : : : :...mortality = Below the National average: 2 (16/4)  
## : : : mortality = Same as the National average: 3 (33/12)  
## : : Readmissions in {Below the National average,  
## : : : Same as the National average}:  
## : : :...Readmissions = Same as the National average: 2 (89/31)  
## : : Readmissions = Below the National average:  
## : : :...count in {1,2,3,4,5,10,14,15,18,21}: 2 (116/34)  
## : : count in {6,7,8,9,12,16,26}: 1 (47/16)  
## : Safety\_of\_care in {Above the National average,Same as the National average}:  
## : :...Readmissions = Above the National average:  
## : :...mortality in {Below the National average,  
## : : : Same as the National average}: 3 (92/25)  
## : : mortality = Above the National average: [S2]  
## : Readmissions in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Above the National average:  
## : :...Readmissions = Below the National average: 3 (47/17)  
## : : Readmissions = Same as the National average:  
## : : :...Safety\_of\_care = Same as the National average: 3 (10/1)  
## : : Safety\_of\_care = Above the National average:  
## : : :...State in {AK,AL,AR,AZ,CO,DC,DE,FL,GA,HI,IA,ID,IL,IN,KS,  
## : : : KY,LA,MA,MD,ME,MI,MN,MO,MS,MT,NC,ND,NE,NH,NJ,  
## : : : NM,NV,NY,OH,OK,OR,PA,PR,RI,SC,SD,TN,UT,VA,VT,  
## : : : WA,WI,WV,WY}: 4 (11)  
## : : State in {CA,CT,TX}: 3 (6)  
## : mortality in {Below the National average,  
## : : Same as the National average}:  
## : :...Readmissions = Below the National average: 2 (202/70)  
## : Readmissions = Same as the National average:  
## : :...mortality = Below the National average:  
## : :...count in {1,2,4,5,6,7,8,9,10,12,14,15,16,18,21,  
## : : : 26}: 2 (25/3)  
## : : count = 3: 3 (3)  
## : mortality = Same as the National average: [S3]  
## Patient\_experience in {Above the National average,Same as the National average}:  
## :...Readmissions = Below the National average:  
## :...Safety\_of\_care = Above the National average:  
## : :...mortality in {Below the National average,  
## : : : Same as the National average}: 3 (82/19)  
## : : mortality = Above the National average:  
## : : :...Patient\_experience = Above the National average: 4 (18/4)  
## : : Patient\_experience = Same as the National average: 3 (18/8)  
## : Safety\_of\_care in {Below the National average,  
## : : Same as the National average}:  
## : :...mortality = Below the National average: 2 (28/6)  
## : mortality in {Above the National average,  
## : : Same as the National average}: [S4]  
## Readmissions in {Above the National average,Same as the National average}:  
## :...Safety\_of\_care = Above the National average:  
## :...mortality = Below the National average:  
## : :...Patient\_experience = Same as the National average: 3 (31/3)  
## : : Patient\_experience = Above the National average:  
## : : :...Readmissions = Above the National average: 4 (16/4)  
## : : Readmissions = Same as the National average: 3 (6/1)  
## : mortality in {Above the National average,  
## : : Same as the National average}:  
## : :...Readmissions = Above the National average:  
## : :...mortality = Above the National average: 5 (31/11)  
## : : mortality = Same as the National average: 4 (130/31)  
## : Readmissions = Same as the National average: [S5]  
## Safety\_of\_care in {Below the National average,  
## : Same as the National average}:  
## :...Patient\_experience = Same as the National average:  
## :...mortality in {Below the National average,  
## : : Same as the National average}: 3 (601/121)  
## : mortality = Above the National average: [S6]  
## Patient\_experience = Above the National average:  
## :...mortality = Above the National average: 4 (31/9)  
## mortality = Below the National average:  
## :...Safety\_of\_care = Same as the National average: [S7]  
## : Safety\_of\_care = Below the National average: [S8]  
## mortality = Same as the National average:  
## :...Readmissions = Above the National average: [S9]  
## Readmissions = Same as the National average:  
## :...State in {AR,CA,CT,HI,KS,LA,MO,MT,ND,NJ,NM,SD,TN,VT,WA,  
## : WV}: 3 (140/41)  
## State in {AZ,FL,IA,MA,MN,UT,WI}: 4 (133/46)  
## State in {AK,DC,DE,MD,PR,RI}: 2 (5/1)  
## State = GA:  
## :...count in {1,3,4,5,6,7,8,9,10,12,14,15,16,18,21,  
## : : 26}: 3 (16/1)  
## : count = 2: 4 (1)  
## State = ID:  
## :...Hospital\_Type = Acute Care Hospitals: 3 (2)  
## : Hospital\_Type = Critical Access Hospitals: 4 (2)  
## State = IL:  
## :...count in {1,3,4,5,6,7,8,9,10,12,14,15,16,18,  
## : : 21}: 4 (30/12)  
## : count in {2,26}: 3 (3)  
## State = KY: [S10]  
## State = ME:  
## :...coverage\_person in [verygood-good]: 3 (8/3)  
## : coverage\_person in [medium-less]: 4 (4)  
## State = MI: [S11]  
## State = NE: [S12]  
## State = NH:  
## :...Hospital\_Type = Acute Care Hospitals: 4 (2)  
## : Hospital\_Type = Critical Access Hospitals: 3 (7/3)  
## State = NV:  
## :...count in {1,2,4,5,6,7,8,9,10,12,14,15,16,18,21,  
## : : 26}: 3 (3/1)  
## : count = 3: 4 (1)  
## State = NY:  
## :...count in {1,2,3,4,5,6,7,8,9,10,14,15,16,18,21,  
## : : 26}: 3 (8/3)  
## : count = 12: 5 (1)  
## State = OK: [S13]  
## State = OR: [S14]  
## State = SC: [S15]  
## State = VA:  
## :...coverage\_person in [verygood-good]: 4 (4/1)  
## : coverage\_person in [medium-less]: 3 (2)  
## State = WY:  
## :...Hospital\_Type = Acute Care Hospitals: 4 (2)  
## : Hospital\_Type = Critical Access Hospitals: 3 (5)  
## State = AL: [S16]  
## State = CO:  
## :...coverage\_person in [medium-less]: 3 (3)  
## : coverage\_person in [verygood-good]: [S17]  
## State = IN:  
## :...count in {2,3,4,5,6,7,8,9,10,12,14,15,16,18,21,  
## : : 26}: 4 (2)  
## : count = 1:  
## : :...coverage\_person = verygood: 4 (12/4)  
## : coverage\_person in [good-less]: 3 (12/4)  
## State = NC:  
## :...Hospital\_Type = Acute Care Hospitals: 3 (4/2)  
## : Hospital\_Type = Critical Access Hospitals:  
## : :...coverage\_person = verygood: 4 (5/1)  
## : coverage\_person in [good-less]: 3 (4/1)  
## State = MS: [S18]  
## State = OH: [S19]  
## State = PA: [S20]  
## State = TX: [S21]  
##   
## SubTree [S1]  
##   
## Effectiveness\_of\_care = Below the National average: 4 (2)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 3 (9/2)  
##   
## SubTree [S2]  
##   
## Effectiveness\_of\_care = Below the National average: 3 (1)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 4 (14/4)  
##   
## SubTree [S3]  
##   
## Safety\_of\_care = Above the National average: 3 (47/9)  
## Safety\_of\_care = Same as the National average:  
## :...State in {AK,AL,CO,CT,DC,DE,GA,IA,ID,KS,KY,LA,MA,MD,ME,MI,MN,NC,ND,NE,NH,  
## : NJ,NV,OH,PA,PR,RI,SC,SD,TX,UT,VA,VT,WI}: 3 (69/7)  
## State in {AR,CA,HI,IL,IN,MS,MT,NM,OK,OR,WA,WV,WY}: 2 (78/26)  
## State = AZ:  
## :...count in {1,2,3,4,5,6,7,8,9,10,12,15,16,18,21,26}: 3 (3)  
## : count = 14: 2 (2)  
## State = FL:  
## :...Effectiveness\_of\_care = Above the National average: 3 (1)  
## : Effectiveness\_of\_care in {Below the National average,  
## : Same as the National average}: 2 (10/3)  
## State = MO:  
## :...coverage\_person in [verygood-medium]: 3 (3/1)  
## : coverage\_person = less: 2 (2)  
## State = NY:  
## :...Effectiveness\_of\_care in {Above the National average,  
## : : Below the National average}: 3 (3)  
## : Effectiveness\_of\_care = Same as the National average: 2 (11/2)  
## State = TN:  
## :...count in {1,2,3,4,6,7,9,10,12,14,15,16,18,21,26}: 3 (6/1)  
## count in {5,8}: 2 (2)  
##   
## SubTree [S4]  
##   
## Effectiveness\_of\_care in {Above the National average,  
## : Same as the National average}: 3 (187/65)  
## Effectiveness\_of\_care = Below the National average: 2 (18/7)  
##   
## SubTree [S5]  
##   
## Patient\_experience = Above the National average: 4 (61/7)  
## Patient\_experience = Same as the National average:  
## :...mortality = Above the National average: 4 (7)  
## mortality = Same as the National average:  
## :...count in {1,4,5,6,7,8,10,14,15,18,21,26}: 3 (35/11)  
## count in {3,9,12,16}: 4 (6/1)  
## count = 2:  
## :...Effectiveness\_of\_care = Below the National average: 3 (1)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 4 (8/1)  
##   
## SubTree [S6]  
##   
## Effectiveness\_of\_care in {Above the National average,  
## : Below the National average}: 4 (5)  
## Effectiveness\_of\_care = Same as the National average: 3 (34/15)  
##   
## SubTree [S7]  
##   
## Effectiveness\_of\_care in {Above the National average,  
## : Same as the National average}: 3 (23/5)  
## Effectiveness\_of\_care = Below the National average: 2 (2)  
##   
## SubTree [S8]  
##   
## Readmissions = Same as the National average: 2 (6/1)  
## Readmissions = Above the National average:  
## :...coverage\_area = verygood: 2 (2)  
## coverage\_area in [good-less]: 3 (6)  
##   
## SubTree [S9]  
##   
## Safety\_of\_care = Below the National average: 3 (47/18)  
## Safety\_of\_care = Same as the National average: 4 (108/17)  
##   
## SubTree [S10]  
##   
## Effectiveness\_of\_care = Below the National average: 4 (1)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 3 (17/6)  
##   
## SubTree [S11]  
##   
## Safety\_of\_care = Below the National average: 2 (1)  
## Safety\_of\_care = Same as the National average: 4 (29/11)  
##   
## SubTree [S12]  
##   
## Effectiveness\_of\_care = Below the National average: 4 (1)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 3 (24/7)  
##   
## SubTree [S13]  
##   
## Effectiveness\_of\_care = Above the National average: 4 (1)  
## Effectiveness\_of\_care in {Below the National average,  
## Same as the National average}: 3 (24/6)  
##   
## SubTree [S14]  
##   
## Effectiveness\_of\_care = Below the National average: 4 (1)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 3 (13/6)  
##   
## SubTree [S15]  
##   
## Effectiveness\_of\_care = Below the National average: 3 (2)  
## Effectiveness\_of\_care in {Above the National average,  
## Same as the National average}: 4 (6/1)  
##   
## SubTree [S16]  
##   
## Safety\_of\_care = Below the National average: 3 (2)  
## Safety\_of\_care = Same as the National average:  
## :...coverage\_person = verygood: 3 (10/3)  
## coverage\_person in [good-less]: 4 (6/2)  
##   
## SubTree [S17]  
##   
## Hospital\_Type = Acute Care Hospitals: 4 (3)  
## Hospital\_Type = Critical Access Hospitals: 3 (6/1)  
##   
## SubTree [S18]  
##   
## Effectiveness\_of\_care in {Above the National average,  
## : Below the National average}: 3 (3)  
## Effectiveness\_of\_care = Same as the National average:  
## :...Hospital\_Type = Critical Access Hospitals: 3 (8/1)  
## Hospital\_Type = Acute Care Hospitals:  
## :...coverage\_person = verygood: 3 (3/1)  
## coverage\_person in [good-less]: 4 (5/1)  
##   
## SubTree [S19]  
##   
## Safety\_of\_care = Below the National average: 2 (1)  
## Safety\_of\_care = Same as the National average:  
## :...Hospital\_Type = Acute Care Hospitals: 4 (15/5)  
## Hospital\_Type = Critical Access Hospitals:  
## :...coverage\_person in [verygood-good]: 3 (8/2)  
## coverage\_person in [medium-less]: 4 (3/1)  
##   
## SubTree [S20]  
##   
## Effectiveness\_of\_care in {Above the National average,  
## : Below the National average}: 3 (2)  
## Effectiveness\_of\_care = Same as the National average:  
## :...coverage\_person in [medium-less]: 4 (7/1)  
## coverage\_person in [verygood-good]:  
## :...Hospital\_Type = Acute Care Hospitals: 4 (5/2)  
## Hospital\_Type = Critical Access Hospitals: 3 (4)  
##   
## SubTree [S21]  
##   
## Safety\_of\_care = Below the National average: 3 (3/1)  
## Safety\_of\_care = Same as the National average:  
## :...coverage\_person = less: 4 (6)  
## coverage\_person in [verygood-medium]:  
## :...count in {1,2,5,6,7,8,9,10,12,14,15,16,18,21,26}: 3 (48/20)  
## count in {3,4}: 4 (3)  
##   
##   
## Evaluation on training data (3200 cases):  
##   
## Decision Tree   
## ----------------   
## Size Errors   
##   
## 119 835(26.1%) <<  
##   
##   
## (a) (b) (c) (d) (e) <-classified as  
## ---- ---- ---- ---- ----  
## 31 64 3 (a): class 1  
## 12 428 152 3 (b): class 2  
## 4 122 1344 120 (c): class 3  
## 2 297 541 11 (d): class 4  
## 2 43 21 (e): class 5  
##   
##   
## Attribute usage:  
##   
## 100.00% Readmissions  
## 100.00% Patient\_experience  
## 100.00% Safety\_of\_care  
## 92.13% mortality  
## 28.09% State  
## 14.25% Effectiveness\_of\_care  
## 12.31% count  
## 5.50% coverage\_person  
## 2.91% Hospital\_Type  
## 0.25% coverage\_area  
##   
##   
## Time: 0.0 secs

p<- predict( model, testX, type="class" )  
sum( p == testy ) / length( p )\*100##74.406

## [1] NA

C5imp(model)

## Overall  
## Readmissions 100.00  
## Patient\_experience 100.00  
## Safety\_of\_care 100.00  
## mortality 92.13  
## State 28.09  
## Effectiveness\_of\_care 14.25  
## count 12.31  
## coverage\_person 5.50  
## Hospital\_Type 2.91  
## coverage\_area 0.25

##predicting ratings for dataset2  
dataset2=filter(gen\_info,is.na(gen\_info$Hospital\_overall\_rating)==TRUE)## 24% od data  
dataset2<-dataset2[,c("Hospital\_overall\_rating",  
 "State","Hospital\_Type",  
 "mortality","coverage\_area", "Readmissions","Patient\_experience",  
 "Safety\_of\_care","count","coverage\_person",  
 "Effectiveness\_of\_care",  
 "Timeliness\_of\_care","latitude","longitude","name","City")]  
  
p<- predict( model, dataset2[,2:11], type="class" )  
head(dataset2)

## Hospital\_overall\_rating State Hospital\_Type  
## 1 <NA> AL Acute Care Hospitals  
## 2 <NA> AL Acute Care Hospitals  
## 3 <NA> AL Acute Care Hospitals  
## 4 <NA> AL Acute Care Hospitals  
## 5 <NA> AL Acute Care Hospitals  
## 6 <NA> AL Critical Access Hospitals  
## mortality coverage\_area Readmissions  
## 1 Same as the National average good Same as the National average  
## 2 Same as the National average good Same as the National average  
## 3 Same as the National average good Same as the National average  
## 4 Same as the National average good Same as the National average  
## 5 Same as the National average good Same as the National average  
## 6 Same as the National average good Same as the National average  
## Patient\_experience Safety\_of\_care count  
## 1 Above the National average Same as the National average 7  
## 2 Above the National average Same as the National average 1  
## 3 Above the National average Same as the National average 1  
## 4 Above the National average Same as the National average 1  
## 5 Above the National average Same as the National average 1  
## 6 Above the National average Same as the National average 1  
## coverage\_person Effectiveness\_of\_care  
## 1 less Same as the National average  
## 2 verygood Same as the National average  
## 3 verygood Below the National average  
## 4 verygood Same as the National average  
## 5 verygood Same as the National average  
## 6 verygood Same as the National average  
## Timeliness\_of\_care latitude longitude name  
## 1 Same as the National average 33.50962 -86.79970 birmingham al  
## 2 Same as the National average 32.96349 -87.13867 centreville al  
## 3 Same as the National average 32.70353 -87.60177 greensboro al  
## 4 Same as the National average 32.00844 -87.25911 camden al  
## 5 Same as the National average 32.48773 -88.26790 york al  
## 6 Above the National average 31.46338 -88.26781 chatom al  
## City  
## 1 BIRMINGHAM  
## 2 CENTREVILLE  
## 3 GREENSBORO  
## 4 CAMDEN  
## 5 YORK  
## 6 CHATOM

class(p)

## [1] "factor"

p=data.frame(p)  
names(p)="Hospital\_overall\_rating"  
nrow(dataset2)

## [1] 1038

dataset=bind\_cols(p,dataset2)  
dataset=data.frame(dataset)  
head(dataset)

## Hospital\_overall\_rating Hospital\_overall\_rating.1 State  
## 1 4 <NA> AL  
## 2 3 <NA> AL  
## 3 3 <NA> AL  
## 4 3 <NA> AL  
## 5 3 <NA> AL  
## 6 3 <NA> AL  
## Hospital\_Type mortality coverage\_area  
## 1 Acute Care Hospitals Same as the National average good  
## 2 Acute Care Hospitals Same as the National average good  
## 3 Acute Care Hospitals Same as the National average good  
## 4 Acute Care Hospitals Same as the National average good  
## 5 Acute Care Hospitals Same as the National average good  
## 6 Critical Access Hospitals Same as the National average good  
## Readmissions Patient\_experience  
## 1 Same as the National average Above the National average  
## 2 Same as the National average Above the National average  
## 3 Same as the National average Above the National average  
## 4 Same as the National average Above the National average  
## 5 Same as the National average Above the National average  
## 6 Same as the National average Above the National average  
## Safety\_of\_care count coverage\_person  
## 1 Same as the National average 7 less  
## 2 Same as the National average 1 verygood  
## 3 Same as the National average 1 verygood  
## 4 Same as the National average 1 verygood  
## 5 Same as the National average 1 verygood  
## 6 Same as the National average 1 verygood  
## Effectiveness\_of\_care Timeliness\_of\_care latitude  
## 1 Same as the National average Same as the National average 33.50962  
## 2 Same as the National average Same as the National average 32.96349  
## 3 Below the National average Same as the National average 32.70353  
## 4 Same as the National average Same as the National average 32.00844  
## 5 Same as the National average Same as the National average 32.48773  
## 6 Same as the National average Above the National average 31.46338  
## longitude name City  
## 1 -86.79970 birmingham al BIRMINGHAM  
## 2 -87.13867 centreville al CENTREVILLE  
## 3 -87.60177 greensboro al GREENSBORO  
## 4 -87.25911 camden al CAMDEN  
## 5 -88.26790 york al YORK  
## 6 -88.26781 chatom al CHATOM

hospital\_info=bind\_rows(dataset,dataset1)  
class(hospital\_info)

## [1] "data.frame"

nrow(hospital\_info)

## [1] 4580

nrow(gen\_info)

## [1] 4580

hospitals=hospital\_info[,c(-2,-9,-13,-14,-15,-16)]  
head(hospitals)

## Hospital\_overall\_rating State Hospital\_Type  
## 1 4 AL Acute Care Hospitals  
## 2 3 AL Acute Care Hospitals  
## 3 3 AL Acute Care Hospitals  
## 4 3 AL Acute Care Hospitals  
## 5 3 AL Acute Care Hospitals  
## 6 3 AL Critical Access Hospitals  
## mortality coverage\_area Readmissions  
## 1 Same as the National average good Same as the National average  
## 2 Same as the National average good Same as the National average  
## 3 Same as the National average good Same as the National average  
## 4 Same as the National average good Same as the National average  
## 5 Same as the National average good Same as the National average  
## 6 Same as the National average good Same as the National average  
## Patient\_experience count coverage\_person  
## 1 Above the National average 7 less  
## 2 Above the National average 1 verygood  
## 3 Above the National average 1 verygood  
## 4 Above the National average 1 verygood  
## 5 Above the National average 1 verygood  
## 6 Above the National average 1 verygood  
## Effectiveness\_of\_care City  
## 1 Same as the National average BIRMINGHAM  
## 2 Same as the National average CENTREVILLE  
## 3 Below the National average GREENSBORO  
## 4 Same as the National average CAMDEN  
## 5 Same as the National average YORK  
## 6 Same as the National average CHATOM

hospitals$Hospital\_overall\_rating=as.factor(hospitals$Hospital\_overall\_rating)  
str(hospitals)

## 'data.frame': 4580 obs. of 11 variables:  
## $ Hospital\_overall\_rating: Factor w/ 5 levels "1","2","3","4",..: 4 3 3 3 3 3 4 3 2 2 ...  
## $ State : Factor w/ 52 levels "AK","AL","AR",..: 2 2 2 2 2 2 2 2 1 1 ...  
## $ Hospital\_Type : Factor w/ 2 levels "Acute Care Hospitals",..: 1 1 1 1 1 2 2 2 2 2 ...  
## $ mortality : Factor w/ 3 levels "Above the National average",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ coverage\_area : Ord.factor w/ 4 levels "verygood"<"good"<..: 2 2 2 2 2 2 2 2 4 4 ...  
## $ Readmissions : Factor w/ 3 levels "Above the National average",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ Patient\_experience : Factor w/ 3 levels "Above the National average",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ count : Factor w/ 17 levels "1","2","3","4",..: 7 1 1 1 1 1 1 1 2 1 ...  
## $ coverage\_person : Ord.factor w/ 4 levels "verygood"<"good"<..: 4 1 1 1 1 1 2 1 1 1 ...  
## $ Effectiveness\_of\_care : Factor w/ 3 levels "Above the National average",..: 3 3 2 3 3 3 3 3 3 3 ...  
## $ City : Factor w/ 2947 levels "ABBEVILLE","ABERDEEN",..: 236 432 1045 375 2935 447 2179 358 2443 1358 ...

describe(hospitals)

## hospitals   
##   
## 11 Variables 4580 Observations  
## ---------------------------------------------------------------------------  
## Hospital\_overall\_rating   
## n missing distinct   
## 4580 0 5   
##   
## Value 1 2 3 4 5  
## Frequency 108 763 2424 1209 76  
## Proportion 0.024 0.167 0.529 0.264 0.017  
## ---------------------------------------------------------------------------  
## State   
## n missing distinct   
## 4580 0 52   
##   
## lowest : AK AL AR AZ CA, highest: VT WA WI WV WY  
## ---------------------------------------------------------------------------  
## Hospital\_Type   
## n missing distinct   
## 4580 0 2   
##   
## Value Acute Care Hospitals Critical Access Hospitals  
## Frequency 3257 1323  
## Proportion 0.711 0.289  
## ---------------------------------------------------------------------------  
## mortality   
## n missing distinct   
## 4580 0 3   
##   
## Value Above the National average Below the National average  
## Frequency 395 340  
## Proportion 0.086 0.074  
##   
## Value Same as the National average  
## Frequency 3845  
## Proportion 0.840  
## ---------------------------------------------------------------------------  
## coverage\_area   
## n missing distinct   
## 4580 0 4   
##   
## Value verygood good medium less  
## Frequency 1302 1125 1046 1107  
## Proportion 0.284 0.246 0.228 0.242  
## ---------------------------------------------------------------------------  
## Readmissions   
## n missing distinct   
## 4580 0 3   
##   
## Value Above the National average Below the National average  
## Frequency 804 867  
## Proportion 0.176 0.189  
##   
## Value Same as the National average  
## Frequency 2909  
## Proportion 0.635  
## ---------------------------------------------------------------------------  
## Patient\_experience   
## n missing distinct   
## 4580 0 3   
##   
## Value Above the National average Below the National average  
## Frequency 2370 1071  
## Proportion 0.517 0.234  
##   
## Value Same as the National average  
## Frequency 1139  
## Proportion 0.249  
## ---------------------------------------------------------------------------  
## count   
## n missing distinct   
## 4580 0 17   
##   
## Value 1 2 3 4 5 6 7 8 9 10  
## Frequency 3217 546 205 131 62 93 49 31 34 36  
## Proportion 0.702 0.119 0.045 0.029 0.014 0.020 0.011 0.007 0.007 0.008  
##   
## Value 12 14 15 16 18 21 26  
## Frequency 48 27 15 23 18 19 26  
## Proportion 0.010 0.006 0.003 0.005 0.004 0.004 0.006  
## ---------------------------------------------------------------------------  
## coverage\_person   
## n missing distinct   
## 4580 0 4   
##   
## Value verygood good medium less  
## Frequency 1155 1144 1145 1136  
## Proportion 0.252 0.250 0.250 0.248  
## ---------------------------------------------------------------------------  
## Effectiveness\_of\_care   
## n missing distinct   
## 4580 0 3   
##   
## Value Above the National average Below the National average  
## Frequency 128 240  
## Proportion 0.028 0.052  
##   
## Value Same as the National average  
## Frequency 4212  
## Proportion 0.920  
## ---------------------------------------------------------------------------  
## City   
## n missing distinct   
## 4580 0 2892   
##   
## lowest : ABBEVILLE ABERDEEN ABILENE ABINGDON ABINGTON   
## highest: ZACHARY ZANESVILLE ZEELAND ZEPHYRHILLS ZION   
## ---------------------------------------------------------------------------

hospitals=filter(hospitals,is.na(hospitals$State)==FALSE)  
table(hospitals$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 108 763 2424 1209 76 0

hospitals=hospitals[,-11]  
names(hospitals)

## [1] "Hospital\_overall\_rating" "State"   
## [3] "Hospital\_Type" "mortality"   
## [5] "coverage\_area" "Readmissions"   
## [7] "Patient\_experience" "count"   
## [9] "coverage\_person" "Effectiveness\_of\_care"

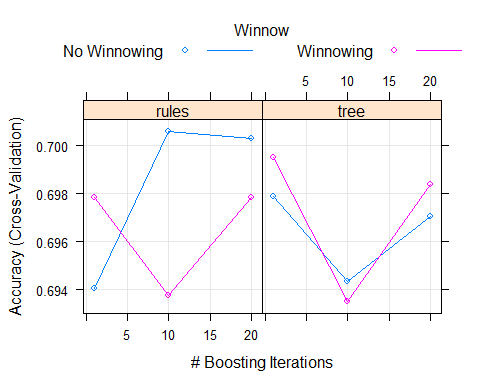
set.seed(1234)  
x\_train <- sample(nrow(hospitals), floor(nrow(hospitals)\*0.8))  
train <- hospitals[x\_train,]  
test<-hospitals[-x\_train,]  
library(caret)  
library(klaR)  
# define training control  
train\_control <- trainControl(method="cv", number=10, savePredictions = TRUE)  
names(hospitals)

## [1] "Hospital\_overall\_rating" "State"   
## [3] "Hospital\_Type" "mortality"   
## [5] "coverage\_area" "Readmissions"   
## [7] "Patient\_experience" "count"   
## [9] "coverage\_person" "Effectiveness\_of\_care"

# train the model  
model\_C5.0 <- train(Hospital\_overall\_rating~., data=train, trControl=train\_control, method="C5.0")  
# summarize results  
print(model\_C5.0)

## C5.0   
##   
## 3664 samples  
## 9 predictor  
## 5 classes: '1', '2', '3', '4', '5'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 3298, 3296, 3298, 3297, 3295, 3298, ...   
## Resampling results across tuning parameters:  
##   
## model winnow trials Accuracy Kappa   
## rules FALSE 1 0.6940466 0.4951728  
## rules FALSE 10 0.7005952 0.5102711  
## rules FALSE 20 0.7003115 0.5104614  
## rules TRUE 1 0.6978643 0.4973872  
## rules TRUE 10 0.6937570 0.4998052  
## rules TRUE 20 0.6978510 0.5037604  
## tree FALSE 1 0.6978666 0.5022154  
## tree FALSE 10 0.6943266 0.4900942  
## tree FALSE 20 0.6970417 0.4982709  
## tree TRUE 1 0.6995030 0.5012829  
## tree TRUE 10 0.6935046 0.4899482  
## tree TRUE 20 0.6983975 0.4983032  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final values used for the model were trials = 10, model = rules  
## and winnow = FALSE.

plot(model\_C5.0)



x\_pred1 <- predict(model\_C5.0, newdata = test)  
confusionMatrix(data = x\_pred1, test$Hospital\_overall\_rating)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 1 2 3 4 5  
## 1 5 6 0 0 0  
## 2 19 89 34 2 0  
## 3 2 55 381 64 1  
## 4 0 1 70 170 11  
## 5 0 0 1 4 1  
##   
## Overall Statistics  
##   
## Accuracy : 0.7052   
## 95% CI : (0.6745, 0.7346)  
## No Information Rate : 0.5306   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.517   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Sensitivity 0.192308 0.58940 0.7840 0.7083 0.076923  
## Specificity 0.993258 0.92810 0.7163 0.8787 0.994463  
## Pos Pred Value 0.454545 0.61806 0.7575 0.6746 0.166667  
## Neg Pred Value 0.976796 0.91969 0.7458 0.8946 0.986813  
## Prevalence 0.028384 0.16485 0.5306 0.2620 0.014192  
## Detection Rate 0.005459 0.09716 0.4159 0.1856 0.001092  
## Detection Prevalence 0.012009 0.15721 0.5491 0.2751 0.006550  
## Balanced Accuracy 0.592783 0.75875 0.7501 0.7935 0.535693

confusionMatrix(data = x\_pred1, test$Hospital\_overall\_rating,mode = "prec\_recall")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 1 2 3 4 5  
## 1 5 6 0 0 0  
## 2 19 89 34 2 0  
## 3 2 55 381 64 1  
## 4 0 1 70 170 11  
## 5 0 0 1 4 1  
##   
## Overall Statistics  
##   
## Accuracy : 0.7052   
## 95% CI : (0.6745, 0.7346)  
## No Information Rate : 0.5306   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.517   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Precision 0.454545 0.61806 0.7575 0.6746 0.166667  
## Recall 0.192308 0.58940 0.7840 0.7083 0.076923  
## F1 0.270270 0.60339 0.7705 0.6911 0.105263  
## Prevalence 0.028384 0.16485 0.5306 0.2620 0.014192  
## Detection Rate 0.005459 0.09716 0.4159 0.1856 0.001092  
## Detection Prevalence 0.012009 0.15721 0.5491 0.2751 0.006550  
## Balanced Accuracy 0.592783 0.75875 0.7501 0.7935 0.535693

table(x\_pred1, test$Hospital\_overall\_rating)

##   
## x\_pred1 1 2 3 4 5  
## 1 5 6 0 0 0  
## 2 19 89 34 2 0  
## 3 2 55 381 64 1  
## 4 0 1 70 170 11  
## 5 0 0 1 4 1

Final Results

nrow(hospital\_info)

## [1] 4580

hospital\_info=filter(hospital\_info,is.na(hospital\_info$State)==FALSE)  
names(hospital\_info)

## [1] "Hospital\_overall\_rating" "Hospital\_overall\_rating.1"  
## [3] "State" "Hospital\_Type"   
## [5] "mortality" "coverage\_area"   
## [7] "Readmissions" "Patient\_experience"   
## [9] "Safety\_of\_care" "count"   
## [11] "coverage\_person" "Effectiveness\_of\_care"   
## [13] "Timeliness\_of\_care" "latitude"   
## [15] "longitude" "name"   
## [17] "City"

StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

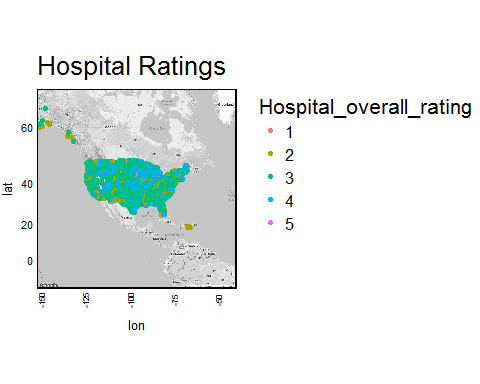
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false

## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

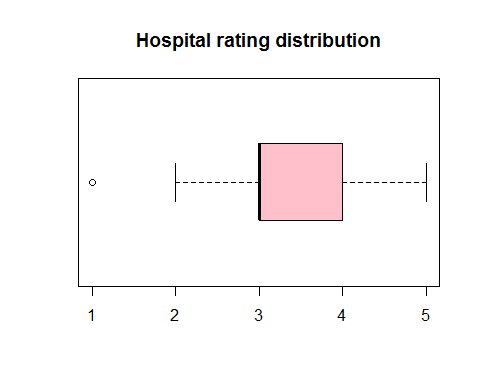
## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = Hospital\_overall\_rating), data = hospital\_info)+hospital.theme+  
 ggtitle("Hospital Ratings ")+  
 theme(axis.text.x=element\_text(size= 6, angle=90,hjust = 0.5))

## Warning: Removed 22 rows containing missing values (geom\_point).



gen\_info$Hospital\_overall\_rating=as.numeric(gen\_info$Hospital\_overall\_rating)  
boxplot(gen\_info$Hospital\_overall\_rating,horizontal=T,main = "Hospital rating distribution",col=c("pink","blue","green","yellow","grey"))



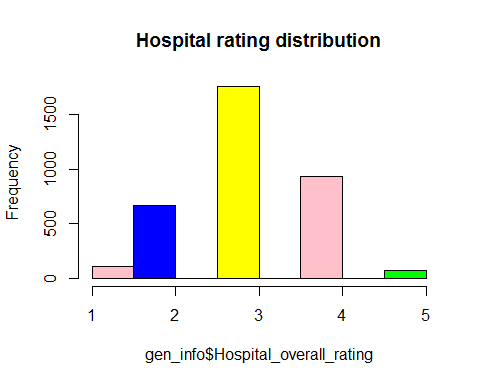
hist(gen\_info$Hospital\_overall\_rating,horizontal=T,main = "Hospital rating distribution",col=c("pink","blue","green","yellow","grey"))

## Warning in plot.window(xlim, ylim, "", ...): "horizontal" is not a  
## graphical parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):  
## "horizontal" is not a graphical parameter

## Warning in axis(1, ...): "horizontal" is not a graphical parameter

## Warning in axis(2, ...): "horizontal" is not a graphical parameter



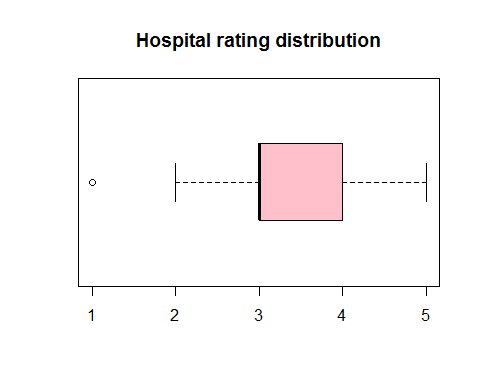
table(gen\_info$Hospital\_overall\_rating,useNA = "always")

##   
## 1 2 3 4 5 <NA>   
## 108 669 1758 933 74 1038

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

##   
## 1 2 3 4 5 <NA>   
## 108 763 2424 1209 76 0

hospital\_info$Hospital\_overall\_rating=as.numeric(hospital\_info$Hospital\_overall\_rating)  
boxplot(hospital\_info$Hospital\_overall\_rating,horizontal=T,main = "Hospital rating distribution",col=c("pink","blue","green","yellow","grey"))



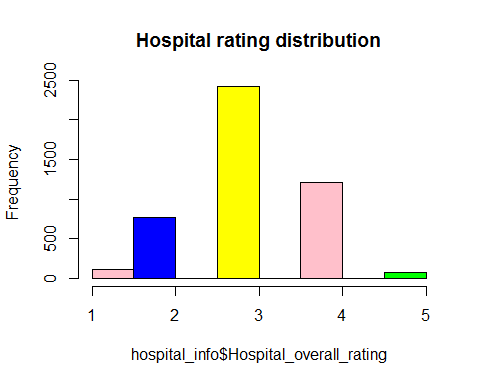
hist(hospital\_info$Hospital\_overall\_rating,horizontal=T,main = "Hospital rating distribution",col=c("pink","blue","green","yellow","grey"))

## Warning in plot.window(xlim, ylim, "", ...): "horizontal" is not a  
## graphical parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):  
## "horizontal" is not a graphical parameter

## Warning in axis(1, ...): "horizontal" is not a graphical parameter

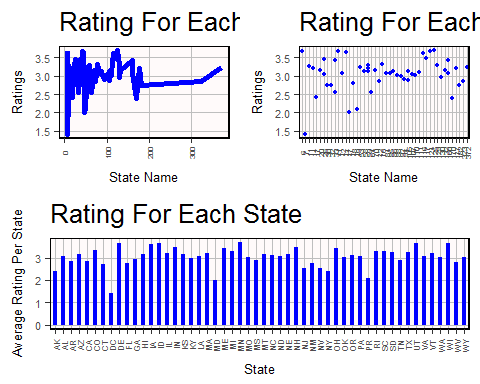
## Warning in axis(2, ...): "horizontal" is not a graphical parameter



rating\_state<-sqldf('select State,count(\*) as cnt,sum(Hospital\_overall\_rating) as rating from hospital\_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 MN 125 462 3.696000  
## 2 DE 6 22 3.666667  
## 3 WI 124 454 3.661290  
## 4 ID 41 150 3.658537  
## 5 UT 44 160 3.636364  
## 6 IA 116 420 3.620690  
## 7 IN 117 408 3.487179  
## 8 NH 26 90 3.461538  
## 9 ME 33 113 3.424242  
## 10 OH 160 547 3.418750

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 = 10),  
 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="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.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="blue", width=0.5 )+  
 hospital.theme+  
 ggtitle("Rating For Each State")+  
 labs(x="State",  
 y="Average 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)



gen\_info2=left\_join(gen\_info,top\_state\_rating)

## Joining, by = "State"

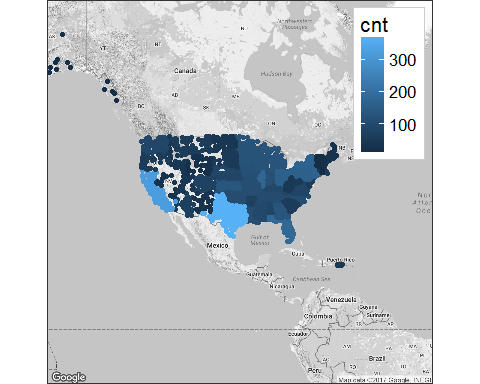
StateMap <- qmap("USA", zoom = 3, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=3&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

StateMap + geom\_point(aes(x = longitude, y = latitude, colour = cnt), data = gen\_info2)

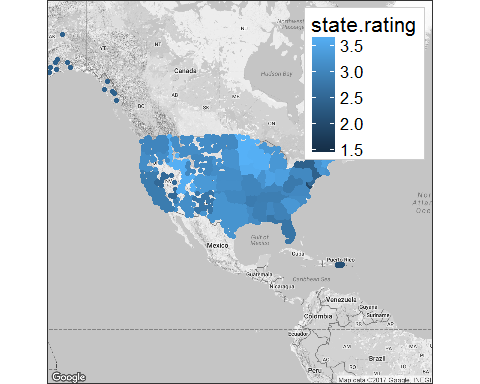
## Warning: Removed 22 rows containing missing values (geom\_point).



StateMap + geom\_point(aes(x = longitude, y = latitude, colour = state.rating), data = gen\_info2,title="Average Ratings For each State")

## Warning: Ignoring unknown parameters: title

## Warning: Removed 22 rows containing missing values (geom\_point).



nrow(top\_state\_rating)

## [1] 52

topten=head(top\_state\_rating,n=10)  
topten

## State cnt rating state.rating  
## 1 MN 125 462 3.696000  
## 2 DE 6 22 3.666667  
## 3 WI 124 454 3.661290  
## 4 ID 41 150 3.658537  
## 5 UT 44 160 3.636364  
## 6 IA 116 420 3.620690  
## 7 IN 117 408 3.487179  
## 8 NH 26 90 3.461538  
## 9 ME 33 113 3.424242  
## 10 OH 160 547 3.418750

bottomten=tail(top\_state\_rating,n=10)  
bottomten

## State cnt rating state.rating  
## 43 NM 32 88 2.750000  
## 44 FL 182 498 2.736264  
## 45 CT 30 82 2.733333  
## 46 NJ 64 163 2.546875  
## 47 NV 33 84 2.545455  
## 48 AK 17 41 2.411765  
## 49 NY 170 408 2.400000  
## 50 PR 49 103 2.102041  
## 51 MD 47 94 2.000000  
## 52 DC 7 10 1.428571

library(noncensus)  
data(zip\_codes)## adding latitude and longitude based on zip cods  
top10=filter(zip\_codes,state==c("DE","MN","NH","WI","ID","UT","IA","IN","ME","OH"))

## Warning in c("NH", "NH", "NH", "NH", "NH", "NH", "NH", "NH", "NH", "NH", :  
## longer object length is not a multiple of shorter object length

bottom10=filter(zip\_codes,state==c("DC","PR","NY","NV","NJ","NM","CT","WV","AR","FL"))

## Warning in c("NH", "NH", "NH", "NH", "NH", "NH", "NH", "NH", "NH", "NH", :  
## longer object length is not a multiple of shorter object length

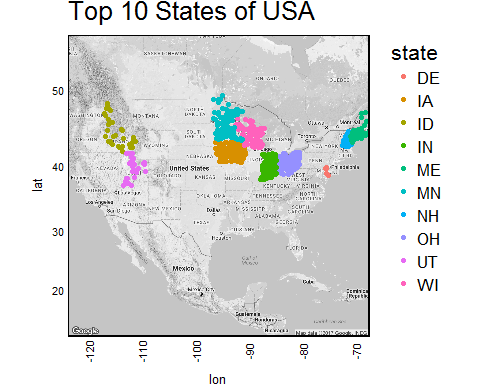
StateMap <- qmap("USA", zoom = 4, color = "bw", legend = "topright",ncol=2)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=USA&zoom=4&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false  
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=USA&sensor=false

## Warning: `panel.margin` is deprecated. Please use `panel.spacing` property  
## instead

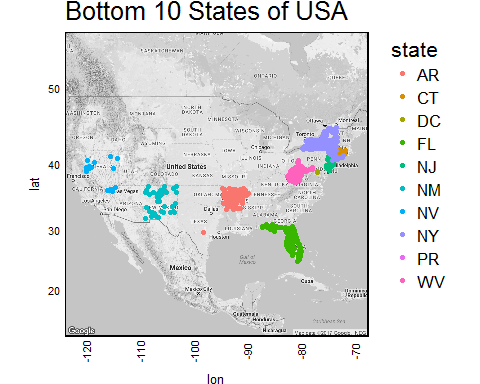
StateMap + geom\_point(aes(x = longitude, y = latitude, colour = state ), data = top10)+hospital.theme+  
 ggtitle("Top 10 States of USA")

## Warning: Removed 1 rows containing missing values (geom\_point).



StateMap + geom\_point(aes(x = longitude, y = latitude, colour = state ), data = bottom10)+hospital.theme+  
 ggtitle("Bottom 10 States of USA")

## Warning: Removed 21 rows containing missing values (geom\_point).



```