RASS data mapped on California Counties

Aditya V Chaturvedi

Saturday, January 3, 2014

This is an algorithm written in R, using RStudio, display counties certain indicators state wide on a map of California.

Last summer I worked as a data visualizaiton intern at for the Energy and Resources Group at UC Berkeley. I was involved with their “Information-Energy Nexus Researcg Project”.

The aim of the project was to plot countywise energy consumption data on California’s map and attempt to discover common trends. It was an exhilarating project and I learnt a lot.

The main source of data for this work is the Residential Applicance Saturation Study, funded and administered by the California Energy Commission in 2009. We were allowed access to the data on request as it is not freely available to the public.

You can find out more about the study at http://www.energy.ca.gov/appliances/rass/ (http://www.energy.ca.gov/appliances/rass/)

Beginning with the function **newrankfunc**. This function returns a numerical list, reordered according to the input in **choice**. I have allowed for two options - **rank** and **percentile**. **Rank** reorders the list based on its standing while **percentile** gives us an idea of the difference between two counties.

```
newrankfunc <- function(x, choice='rank')
{
  library(Hmisc)
  if ((tolower(choice))%nin%c('rank', 'percentile')){stop("Incorrect Input: Only rank o
r percentile accepted")}
  k <- length(x) - sum(is.na(x))
  z <- rep(-1, length(x))
  for (i in 1:length(x))
  {
    if (!is.na(x[i]))
    {
      if (choice=='percentile'){ ordered_temprory <- (x[i]/(max(x, na.rm=TRUE)))}
      else {ordered_temprory <- (sum(x<x[i], na.rm=TRUE)/k)}
      z[i] <- ordered_temprory
    }
  }
  z
}
```

mapcounty takes the input of county, which is a data frame with the county name and an associated rank numeral to decide the color intesity.

```
mapcounty <- function (county)
{
  if (as.numeric(county[2])==-1)
  {
    return()
  }
  else
  {
    col_code = as.numeric(county[2])
    col_code=rgb(1-col_code, 1-col_code, 1)
  }
  map("county", sub(" county", "", paste("california",tolower(county[1]), sep=",")), ad
d = TRUE, col=col_code, fill=TRUE)
}
```

onlycounty This returns the county's name based on zipcode entered. **http://www.unitedstateszipcodes.org (http://www.unitedstateszipcodes.org)** provides a nice database with the zipcodes listed with their county names. You can download the particular file for this code at http://www.unitedstateszipcodes.org/zip_code_database.csv (http://www.unitedstateszipcodes.org/zip_code_database.csv)

To run this code on your computer, you will need to modify the file path to wherever you have saved the file.

```
onlycounty <- function(zipno)
{
  countyzipdata <- read.csv("C:/Users/ADITYA/Desktop/R/RASS functions/zip_code_databas
e.csv", header=FALSE)
  z <- match(zipno, countyzipdata[,1])
  as.character(countyzipdata[z,7])
}
```

This is the final code. I believe it is fairly self explanatory. I have repeated the calls to the library because it helps to run the code independetly.

Quick Summary of the path:

-> Generate a map of the state, in this case California. The installed package has maps for other states too.

-> Load the rass file into the R database

-> Extract the zipcodes.

-> Generate the county names for those zipcodes.

-> Refine the data to account for incomplete map structure or incorrect county references

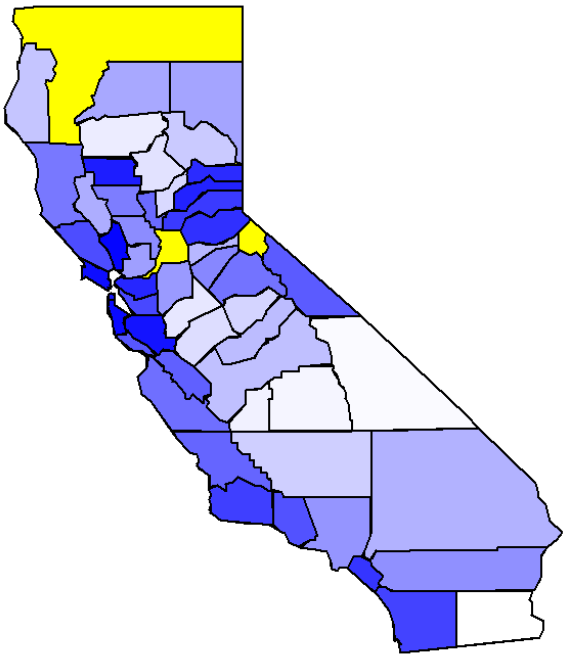
-> Generate a ranked table for the counties.

-> Generate the map.

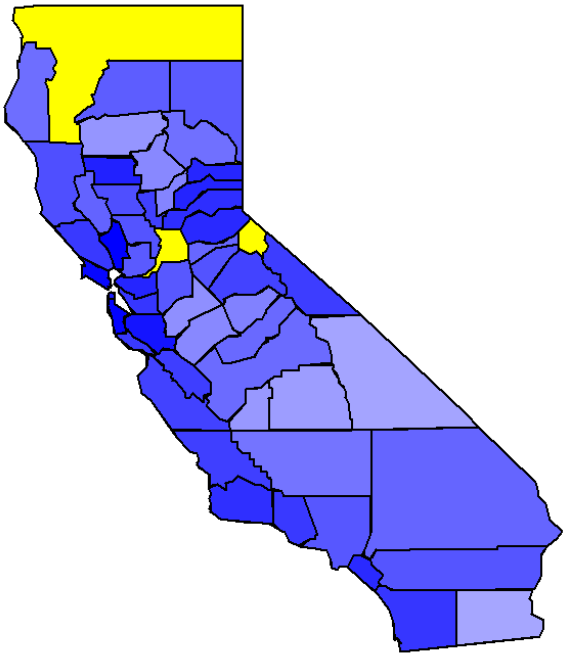
The following is an example of the outshowing the average income for each county dark blue signifying higher amount. The sections in yellow do not are sections with no available information.

```
{
  library(sp)
  library(maps)
  library(mapdata)
  library(maptools)
  library(scales)
  library(mapproj)
  library(Hmisc)
  library(plyr)
  rass <- read.csv("C:/Users/ADITYA/Desktop/R/RASS functions/Survdata.csv")
  countytemp <- ddply(rass, "servzip", summarize, avginc = mean(avginc, na.rm=TRUE))
  countytemp$servzip <- apply(data.frame(countytemp$servzip), 2, onlycounty)
  countytemp <- ddply(countytemp, "servzip", summarize, avgsal = mean(avginc, na.rm=TRU
E))
  countynames <- countytemp[,1]
  countyrank <- newrankfunc(countytemp$avgsal, "rank")
  countyrank2 <- newrankfunc(countytemp$avgsal, "percentile")
  county <- data.frame(countynames, countyrank)
  county2 <- data.frame(countynames, countyrank2)
  map("state", "california", col="yellow", fill=TRUE)
  k<-apply(county, 1, mapcounty)
  title("Ranked display of average income in California")
  map("state", "california", col="yellow", fill=TRUE)
  k<-apply(county2, 1, mapcounty)
  title("Percentile display of average income in California")
}
```

Ranked display of average income in California



Percentile display of average income in California

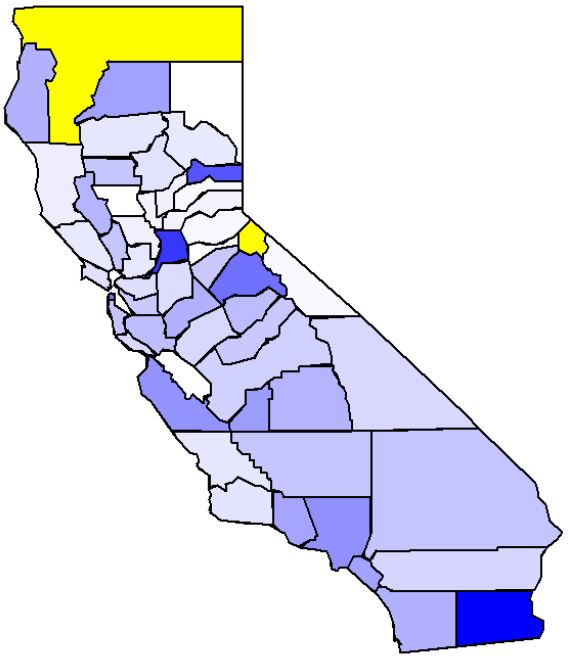


The percentile plot smoothens out the extremes visible in the rank plot. We see that the wealth is accumulated at the coast, which is in line with what we already know.

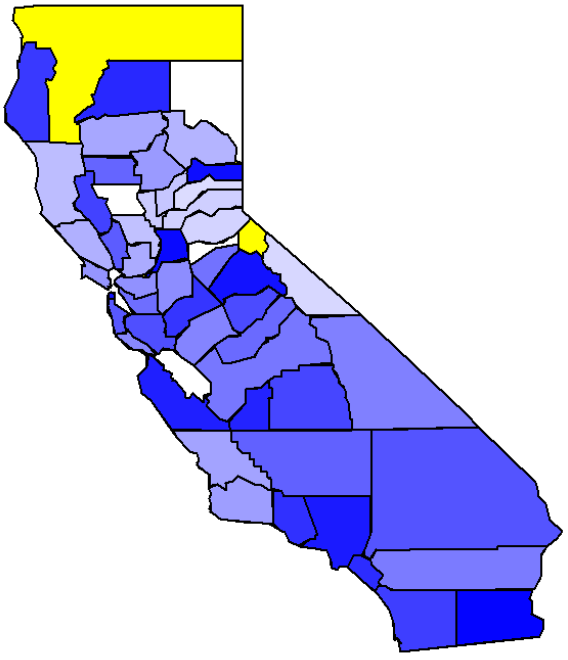
Next up is the solar energy use throughout the state.

NOTE: All code after this has been hidden and only the output map is displayed.

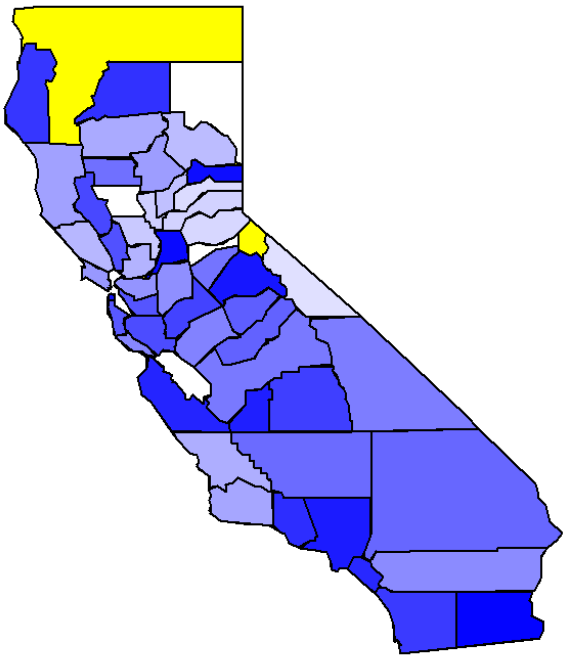
Use of heating facility in California



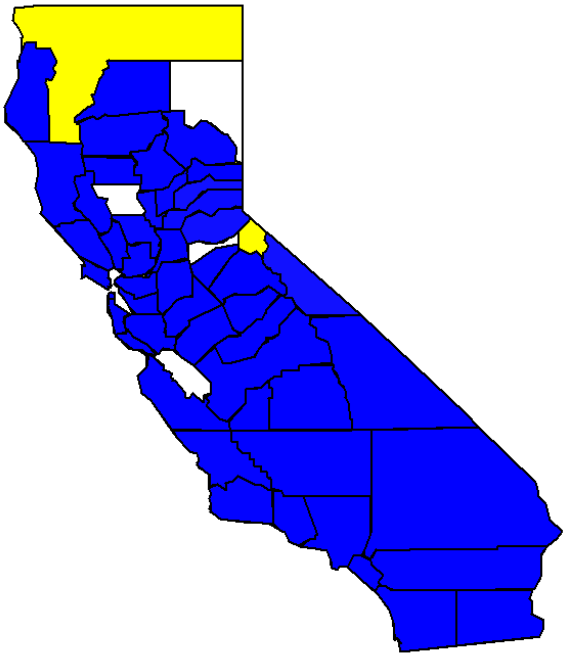
Ranked Use of Solar Energy heaters in California



Ranked Use of Solar heaters (w/o backup) as a fraction of others



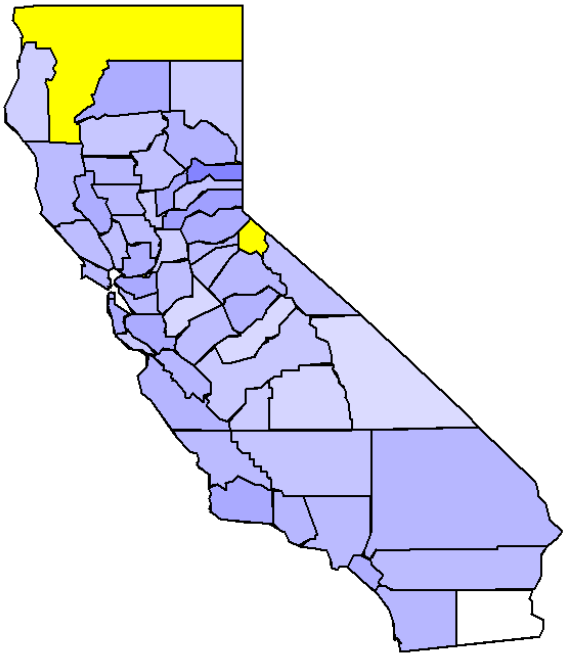
Percentile Use of Solar heaters (w/o backup) as a fraction of others



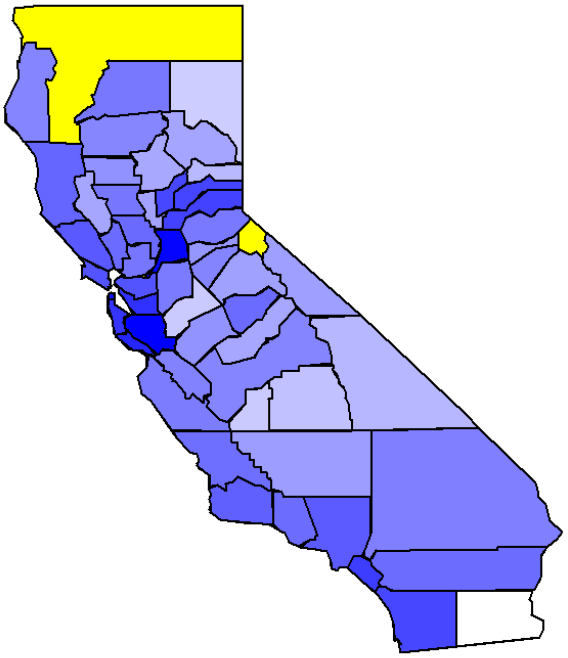
It seems from this data that that in general solar energy is not concentrated in any particular region of California but equally distributed across the state.However, we see that the percentile display of fraction of solar heaters shows almost equal use throughout the state. Therefore, though the ranked display shows wide variation, that variation is very little and the use of solar heaters without backup is also qually distirbuted across the state.

Lastly as an additional exercise, I considered the amount of laptop users vs. desktop PC users.

Desktop PC users in California



Laptop users in California



Understandably, the Desktop users are almost even numbered throughout the state while the Laptop users are concentrated around the Silicon Valley in Santa Clara County.

Hope you found this useful.