STA 141A HOMEWORK 3

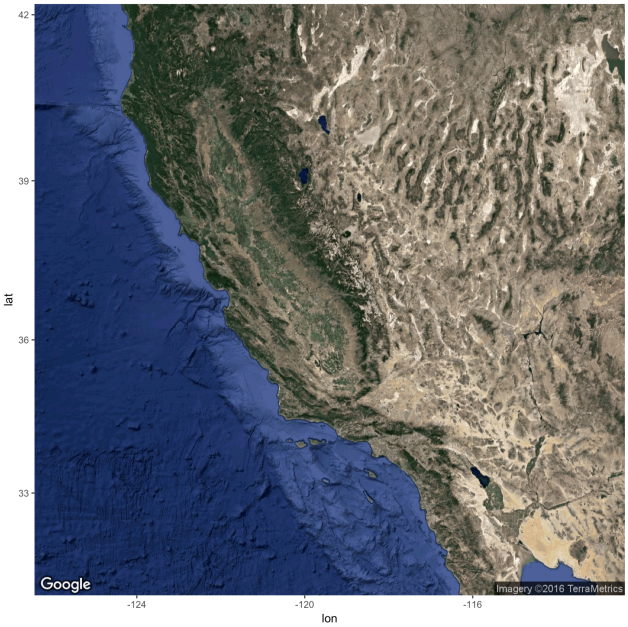
Janet Loyola
912092672

The codes and results derived by using these codes constitute my own work. I have consulted the following resources regarding this assignment: Stephanie Lam, Dhanya Chandrasekar

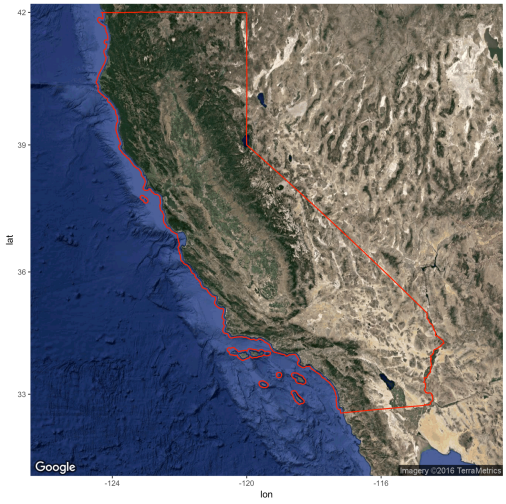
Question 1:

i) Draw a satellite map of California, together with state boundary, and indicate the location of the universities (cities where the universities are located) on the map.

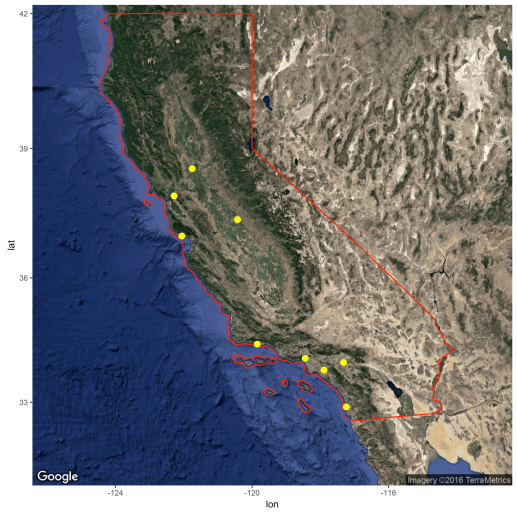
First I got the satellite map of California.



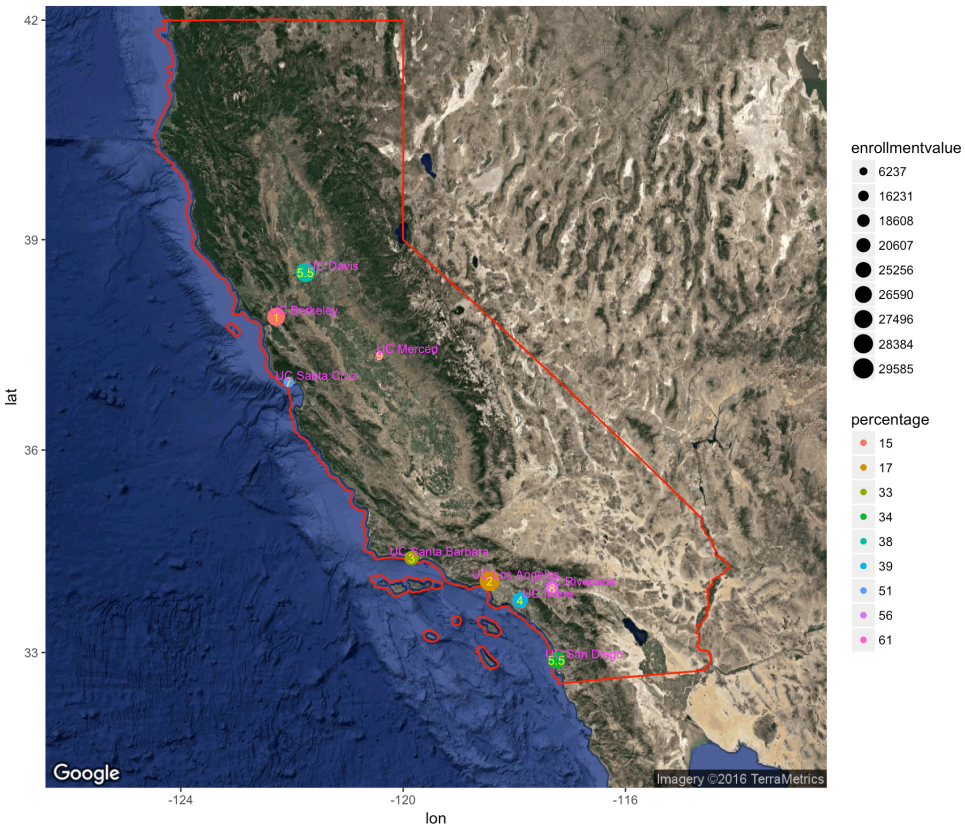
Then I got the CA boundary in the map.



Then I got the locations of the universities.

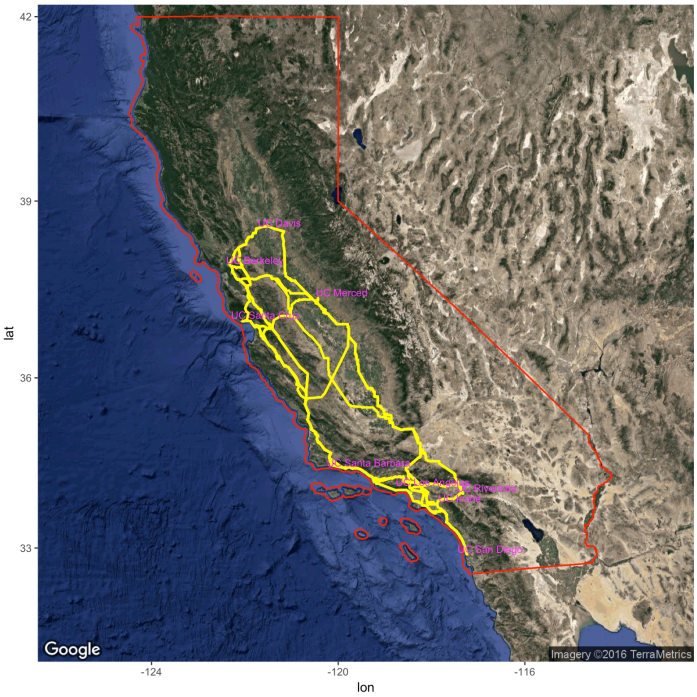


ii) In the same map, display the information on the relative rank (among the 9 UC campuses), student population size and acceptance rate by making use of various aesthetic features



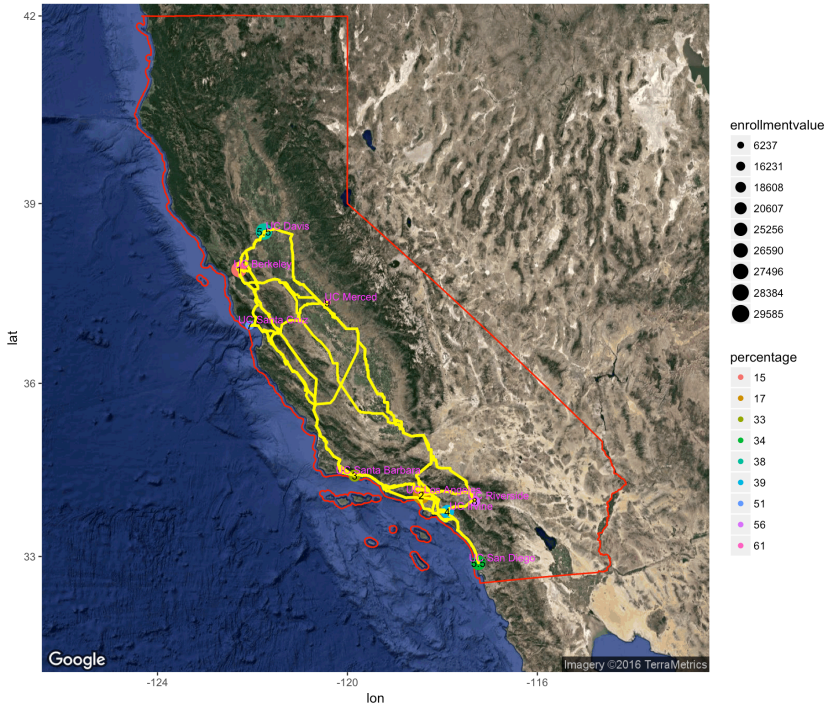
From this graph, we can see that UC Davis and UC San Diego have the same relative rank. This is because both these universities have the same rank. Also, we can see that UC Merced has very few enrollments in comparison to other universities and also has the highest acceptance rates (percentage).

iii) Find and depict, on the same map, bicycle routes connecting every pair of campuses (cities).



This graph tells us all the bike paths connecting each of the campuses to another.

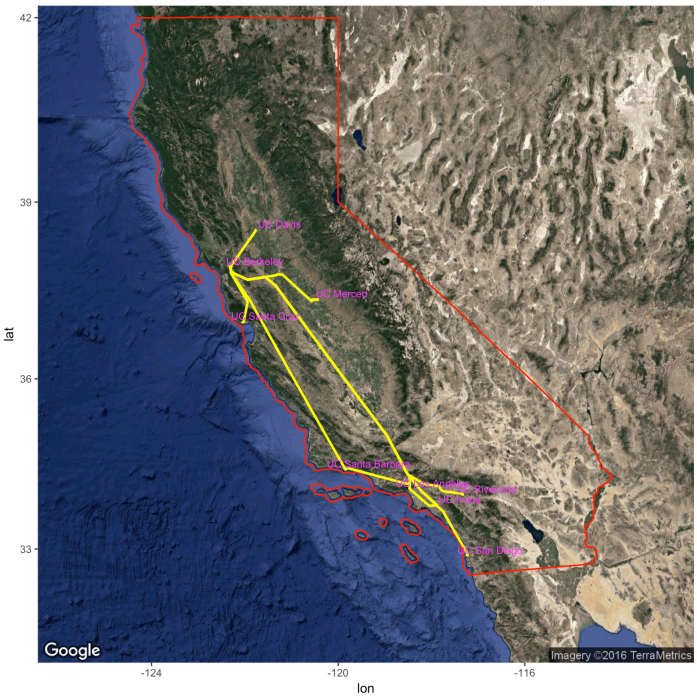
OVERALL GRAPH FOR Q1:



This graph depicts information of several UC campuses like the location, distance, time taken from one UC campus to another, rank, relative rank, correlation between enrollment and acceptance rates (percentage) and also shows all the possible bike routes from one UC campus to another.

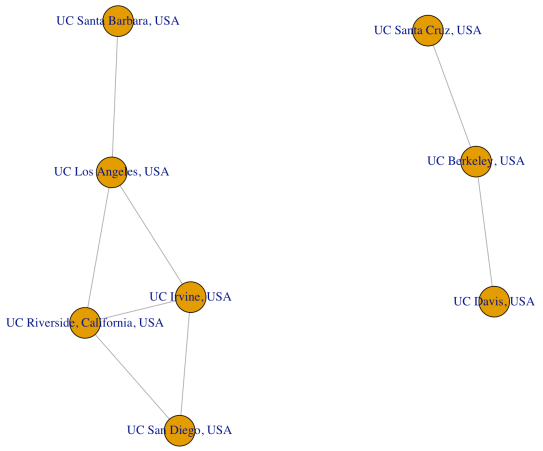
QUESTION 2:

i) Compute driving distances between each pair of campuses (cities).



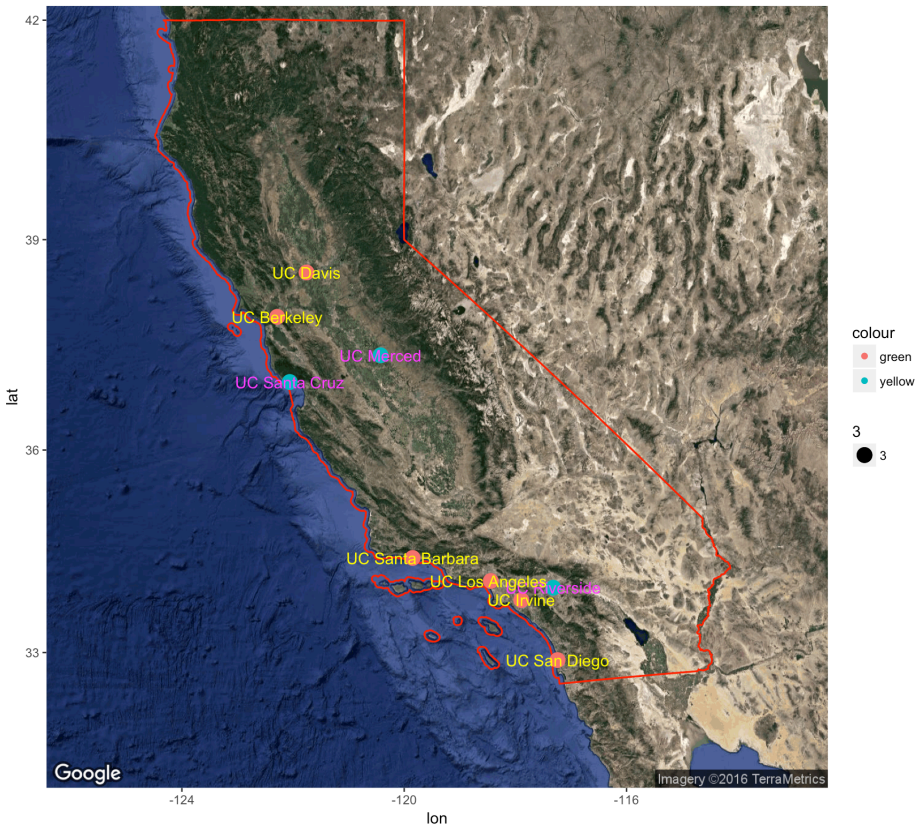
This graph depicts driving distances between each pair of campuses.

ii and iii) Use the rule that if the driving distance between a pair of campuses (cities) is at most 100 miles, then they are neighbors, i.e., there is an edge connecting those nodes. Depict the graph thus constructed.



From this graph, we can see which campuses are neighbors and which are not. We can also see that UC Merced is not even on this graph. This means that none of the UC campuses are within 100 miles to UC Merced.

iv) Use separate colors for depicting campuses depending on whether the student enrollment is less than 20,000 or greater than that.



The blue dots with the pink names represents the campuses with the student enrollment less than 20,000. The orange dots with yellow names represents the campuses with student enrollment greater than 20,000. This helps in visualizing the universities based on enrollment.

CODE APPENDIX:

```
#The codes and results derived by using these codes constitute my own work.
#I have consulted the following resources regarding this assignment:"Stephanie Lam, Dhanya
Chandrasekhar


###QUESTION 1
#Represent the information graphically on a map. Specific features are stated below.
#(i) Draw a satellite map of California, together with state boundary, and indicate the location of the
#universities (cities where the universities are located) on the map.
#(ii) In the same map, display the information on the relative rank (among the 9 UC campuses),
#student population size and acceptance rate by making use of various aesthetic features.
#(iii) Find and depict, on the same map, bicycle routes connecting every pair of campuses (cities).

packageurl <- "http://cran.r-project.org/src/contrib/Archive/ggplot2/ggplot2_2.1.0.tar.gz"
install.packages(packageurl, repos=NULL, type="source")
library(ggplot2)
install.packages("devtools")
library(devtools)
install_github("dkahle/ggmap")
library(ggmap)
install.packages("maptools")
library(maptools)
install.packages("sp")
libary(sp)

readShapePoly("usstate.shp")
stateshapefile=readShapeSpatial("usstate.shp")
states=fortify(stateshapefile)
cali=subset(states,states$id=="13")

#gets the satelline CA map
CA=get_map("California,USA",zoom=6,maptype="satellite")

#satellite CA map with CA boundary in red
ggmap(CA)+geom_polygon(aes(x = long, y = lat, group=group), data = cali, color = "red",size=0.6,
fill=NA)

#lon and lat of colleges
UC.location = rbind(
  geocode("UC Berkeley, USA",output="latlon"),
  geocode("UC Los Angeles, USA",output="latlon"),
  geocode("UC Santa Barbara, USA",output="latlon"),
  geocode("UC Irvine, USA",output="latlon"),
  geocode("UC Davis, USA",output="latlon"),
  geocode("UC San Diego, USA",output="latlon"),
  geocode("UC Santa Cruz, USA",output="latlon"),
  geocode("UC Riverside, California, USA",output="latlon"),
  geocode("UC Merced, USA",output="latlon"))

#To get the locations of the UC's
ggmap(CA) + geom_polygon(aes(x = long, y = lat, group=group), data = cali, color = "red",size=0.6,
fill=NA)+
```

```
 geom_point(aes(x = lon, y = lat), data = UC.location,color="yellow",size=3)

##To get information regarding lon,lan,relative rank,enrollment and percentage for each school
UC_info=read.csv("/Users/janetloyola/Downloads/UC_US_News_Ranking_2017.csv")

names(UC_info) = c("school", "rank", "enrollment", "percentage")
UC_info$school = as.character(UC_info$school)
percentage = UC_info$percentage
relativerank = rank(UC_info$rank)
enrollmentvalue = as.factor(UC_info$enrollment)
UCinfo = cbind(UC.location, rel.rank = paste(rank(UC_info$rank)),
enrollment = paste(UC_info$enrollment),
percentage = paste(UC_info$percentage),school = gsub("[[:space:]][[:space:]]", "", UC_info$school))

##Plotting the information based on the relative rank (among the 9 UC campuses),enrollment and
percentage
ggmap(CA) + geom_polygon(aes(x = long, y = lat, group=group), data = cali, color = "red",size=0.6,
fill=NA) +
geom_point(aes(x = lon, y = lat, size=enrollmentvalue, color=percentage), data = UCinfo) +
geom_text(mapping=aes(x = lon, y = lat, label=relativerank), data = UCinfo,  size=3,color="yellow") +
geom_text(mapping=aes(x = lon+0.5, y = lat+0.1, label=school),data = UCinfo,  size = 3, col='magenta')

#iii)

#names of the UC's
UC= c("UC Berkeley, USA", "UC Los Angeles, USA", "UC Santa Barbara, USA", "UC Irvine, USA", "UC
Davis, USA",
        "UC San Diego, USA", "UC Santa Cruz, USA", "UC Riverside, California, USA", "UC Merced, USA")
#all the possible from and to combinations
pair = combn(UC,2)
#data frame with 0 columns and 0 rows
path = data.frame()

##grouped based on all the possible bike paths from location to destination
for (i in 1:36){
  path = rbind(path, cbind(route(from = pair[1,i], to = pair[2,i], mode = "bicycling", structure =
"route"),
                  group=paste(pair[1,i],pair[2,i])))
}

#Map of bike paths from each of the locations to the respective destinations.


#OVERALL GRAPH FOR Q1
ggmap(CA) + geom_polygon(aes(x = long, y = lat, group=group), data = cali, color = "red",size=0.6,
fill=NA) +
geom_point(aes(x = lon, y = lat, size=enrollmentvalue, color=percentage), data = UCinfo) +
geom_path(aes(x = lon, y = lat, group=group), data = path, color = "yellow", size=1) +
geom_text(mapping=aes(x = lon, y = lat, label=relativerank), data = UCinfo,  size=3) +
geom_text(mapping=aes(x = lon+0.5, y = lat+0.1, label=school),data = UCinfo,  size = 3, col='magenta')

#####EXTRA CODE (tried different method)
#university=as.character(UC_info$University)
```

```
#test = sapply(1:length(university), function(i){
 # p=unlist(strsplit(university[i],' '))
  #s=geocode(paste(paste0(p,collapse = ''),",California, USA"),output=c("latlon"))
#}#end of function
#)#end of sapply
#test
#x=t(test)

#location=as.data.frame(x)
#k=cbind(location,university) ###created a dataframe

#names(UC_info) = c("school", "rank", "enrollment", "percentage")
#UC_info$school = as.character(UC_info$school)
#percentage = UC_info$percentage
#relrank = rank(UC_info$rank)
#enrollvalue = as.factor(UC_info$enrollment)

#UCinfo = cbind(x, rel.rank = paste("rank =", rank(UC_info$rank)),
#enrollment = paste("enrollment =", UC_info$enrollment),
#percentage = paste("percentage =", UC_info$percentage),
#school = gsub("[[:space:]][[:space:]]", "", UC_info$school))
#class(UCinfo) ##matrix
#Have to convert to data frame
#info=as.data.frame(UCinfo)
#class(info)  ###info=UC.info

####QUESTION 2
##getting distance
distance = data.frame()
for (i in 1:36){
  distance = mapdist(from = pair[1,i], to = pair[2,i], mode = "driving", structure = "route")
}

###Getting information for location and destinations
distance=data.frame()
for (i in 1:36)
{
  distance = rbind(distance, cbind(mapdist(pair[1,i],pair[2,i],mode="driving",structure="route"),
                   group=paste(pair[1,i],pair[2,i])))
}
##Getting the driving path
dpath=data.frame()
for (i in 1:36)
{
  dpath = rbind(dpath, cbind(route(pair[1,i],pair[2,i],mode="driving",structure="route"),
                             group=paste(pair[1,i],pair[2,i])))
}

# Shows driving distances between each pair of campuses
ggmap(CA) + geom_polygon(aes(x = long, y = lat, group=group), data = cali, color = "red",size=0.6,
fill=NA) +
  geom_path(aes(x = lon, y = lat, group=group), data = dpath, color = "yellow", size=1)+
  geom_text(mapping=aes(x = lon+0.5, y = lat+0.1, label=school),data = UCinfo,  size = 3,
col='magenta')
```

```
#To plot a graph where maximum distance is 100 miles between the campuses.This makes them
neighbors.
distance[distance$miles<=100,1:2]
mat=get.adjacency(graph.edgelist(as.matrix(distance[distance$miles<=100,1:2]), directed=FALSE))
net=graph.adjacency(adjmatrix=mat,mode="undirected",weighted=TRUE,diag=FALSE)
plot(net)

#To segregate the data into two different parts-one with enrollment less than 20000 students and
other more than 20000 students
UCVal = cbind(UC.location,enrollment = ranking$enrollment,
      percentage = ranking$percentage,school = gsub("[[:space:]][[:space:]]", "", ranking$school))

p20k=UCVal[UCVal$enrollment<20000,]
pa20k=UCVal[UCVal$enrollment>20000,]

# Graph depicting campuses depending on whether the student enrollment is less than 20,000 or
greater than that.
ggmap(CA) + geom_polygon(aes(x = long, y = lat, group=group), data = cali, color = "red",size=0.6,
fill=NA) +
geom_point(aes(x = lon, y = lat, size=3, color="yellow"), data = p20k) +
geom_point(aes(x = lon, y = lat, size=3, color="green"), data = pa20k) +
geom_text(mapping=aes(x = lon, y = lat, label=school), data = p20k,  size=4,color="magenta") +
geom_text(mapping=aes(x = lon, y = lat, label=school), data = pa20k,  size=4,color="yellow")
```