**Homework Assignment 3**

**STA 141A A02**

**Submitted By:**

#1

RDS1<-readRDS("sf\_bikeshare\_trips.rds")

View(RDS1)

#function1 read bay area data from the csv file and after modification store in rds

baybikesharetrip<-read.csv("sf\_bikeshare\_trips.csv")

baybikesharetrip$start\_date

View(baybikesharetrip)

h<-function(a,b){

BayTrip<-read.csv(a)

fmt="%Y-%m-%d %H:%M:%S"

BayTrip$start\_date=strptime(BayTrip$start\_date,fmt)

BayTrip$end\_date=strptime(BayTrip$end\_date,fmt)

BayTrip$subscriber\_type=factor(BayTrip$subscriber\_type)

saveRDS(BayTrip,b)

}

PathCsv<-"sf\_bikeshare\_trips.csv"

PathRds<-"sf\_bikeshare\_trips.rds"

h(PathCsv,PathRds)

c<-readRDS("sf\_bikeshare\_trips.rds")

View(c)

library(maps)

#function2 read data for bike share station from csv file and store data into rds after appropriate column transformation.

h1<-function(a,b){

baybikeshare<-read.csv(a)

baybikeshare$installation\_date=as.Date(baybikeshare$installation\_date,"%Y-%m-%d")

saveRDS(baybikeshare,b)

}

PathCsv2<-"sf\_bike\_share\_stations.csv"

PathRds2<-"sf\_bike\_share\_stations.rds"

h1(PathCsv2,PathRds2)

k<-readRDS(PathRds2)

View(k)

#2

read.csv("sf\_bikeshare\_trips.csv")

library(ggmap)

library(ggplot2)

library(ggmap)

no\_of\_trips<-readRDS("sf\_bikeshare\_trips.rds")

no\_of\_stations<- readRDS("sf\_bike\_share\_stations.rds")

no\_of\_stations$landmark

#remove duplicate stations

no\_of\_stations=no\_of\_stations[!duplicated(no\_of\_stations$station\_id),]

# for San Francisco stations latitude must be greater then 37.7

sanfran<-no\_of\_stations[no\_of\_stations$latitude>37.7,]

trip=table(sanfran$station\_id)

trip=as.data.frame(trip)

which.max(trip$freq)

names(trip)=c("sid","freq")

#merge table based on the trip frequency.

w<-merge(no\_of\_stations,trip,by.x="station\_id",by.y="sid")

get\_center =function(x){

sapply(x[c("longitude","latitude")],mean,na.rm=TRUE)

}

#we got the frequency for the station id:

map=get\_map(get\_center(w),zoom = 15)

library(ggmap)

ggmap(map,base\_layer = ggplot(w,aes(longitude,latitude)))+

geom\_point(aes(size=freq),pch=10)+

geom\_text(data =w,aes(x = longitude, y = latitude ,label =name),

position = position\_dodge(width = 0.00001),vjust = -0.5, size = 2

,check\_overlap = FALSE)

#3.

#function1 read bay area data from the 5 csv file and after appropriate modification store in rds

LosAngle<-function(input1,input2,input3,input4,input5,output)

{

a<-la\_metro\_16\_q3<-read.csv(input1)

b<-la\_metro\_16\_q4<-read.csv(input2)

c<-la\_metro\_17\_q1<-read.csv(input3)

d<-la\_metro\_17\_q2<-read.csv(input4)

e<-la\_metro\_17\_q3<-read.csv(input5)

library(plyr)

combined <- rbind.fill(a,b,c,d,e)

fmt = "%m/%d/%Y %H:%M"

combined$start\_time<- strptime(combined$start\_time,fmt)

combined$end\_time<- strptime(combined$end\_time,fmt)

combined$trip\_route\_category=factor(combined$trip\_route\_category)

combined$passholder\_type=factor(combined$passholder\_type)

saveRDS(combined,output)

}

#q<-read.csv("2016\_q3\_la\_metro\_trips.csv")

LosAngle("2016\_q3\_la\_metro\_trips.csv","2016\_q4\_la\_metro\_trips.csv",

"2017\_q1\_la\_metro\_trips.csv","2017\_q2\_la\_metro\_trips.csv",

"2017\_q3\_la\_metro\_trips.csv","Net\_Losangle\_metro\_trips.rds")

vb<-readRDS("Net\_Losangle\_metro\_trips.rds")

View(vb)

#function2 read bay area data from the csv file and after modification store in rds

LosAngleBikeStation<-function(input,output){

a<-read.csv(input)

names(a)=tolower(names(a))

a$go\_live\_date[[1]]="7/7/2016"

a$go\_live\_date<-as.Date(a$go\_live\_date,"%m/%d/%Y")

a$region<- factor(a$region)

a$region[a$region == "N/A"]=NA

a$region<- droplevels(a$region)

a$status<- factor(a$status)

saveRDS(a,output)

}

LosAngleBikeStation("metro-bike-share-stations-2017-10-20.csv","BiKeLos.rds")

s<-readRDS("BiKeLos.rds")

#4

no\_of\_stations<- readRDS("BiKeLos.rds")

trip<-readRDS("Net\_Losangle\_metro\_trips.rds")

View(trip)

View(no\_of\_stations)

wq<-merge(no\_of\_stations,trip,by.x="station\_id",by.y="start\_station\_id")

location<- trip[c("start\_station\_id","start\_lon","start\_lat")]

names(location)<-c("station\_id","longitude","latitude")

location<- location[!duplicated(location),]

location<- aggregate(location[-1],location["station\_id"],mean)

no\_of\_stations<-merge(no\_of\_stations,location,by="station\_id")

id<- no\_of\_stations$station\_id[is.na(no\_of\_stations$longitude)]

no\_of\_station1<-match(id,trip$end\_station\_id)

no\_of\_stations=no\_of\_stations[!duplicated(no\_of\_stations$station\_id),]

no\_of\_stations<-no\_of\_stations[no\_of\_stations$latitude>33.8,]

trips=table(no\_of\_trips$start\_station\_id)

trips=as.data.frame(trips)

names(trips)=c("sid","freq")

w<-merge(no\_of\_stations,trips,by="station\_id")

wq<-merge(no\_of\_stations,trip,by.x="station\_id",by.y="start\_station\_id")

table(wq$station\_id)

pi<-wq[wq$latitude>33.8,]

names(pi)

head(pi)

a<-pi[!is.na(pi),]

#pi<-pi[!is.na(pi)]

t1<-table(a$station\_id)

#t1=table(no\_of\_trips$start\_station\_id)

t1=as.data.frame(t1)

names(t1)<- c("sid","freq")

final<-merge(no\_of\_stations,t1,by.x="station\_id",by.y="sid")

map = get\_map(location = c(lon=mean(final$longitude),lat=mean(final$latitude)),zoom =14)

pi<-as.data.frame(pi)

library(maps)

library(ggmap)

ggmap(map , base\_layer = ggplot(final,aes(longitude,latitude))) +geom\_point(aes(size=freq),pch=10)+geom\_text(data =final,aes(x = longitude, y = latitude ,label =station\_name), position = position\_dodge(width = 0.00001),

vjust = -0.5, size = 3,check\_overlap = FALSE)

#5

library(lubridate)

library(geosphere)

hrs = function(x){

hr<-hour(x)

hrs\_con<- cut(hr,breaks = c(-Inf,6,10,15,19,Inf),right=FALSE,label=c("Fix","Morning","Noon","Evening","Night"))

levels(hrs\_con)[[1]]="Night"

hrs\_con

}

no\_of\_trips$hrs<-hrs(no\_of\_trips$start\_date)

trip$hrs<- hrs(final$start\_time)

View(trip)

a<- trip[!is.na(trip$start\_lat),]

a$start\_lat

b<- trip[!is.na(trip$start\_lon),]

c<- trip[!is.na(trip$end\_lat),]

table(c$end\_lat)

d<- trip[!is.na(trip$end\_lon),]

trip$dist<-distGeo(c(mean(b$start\_lon),mean(a$start\_lat)),c(mean(d$end\_lon),mean(c$end\_lat)))

trip$dist<-distGeo(cbind(trip$end\_lon,trip$end\_lat),cbind(trip$start\_lon,trip$start\_lat))

#no\_of\_trips$dist<-distGeo(c(mean(),mean(a$start\_lat)),c(mean(d$end\_lon),mean(c$end\_lat))

table(trip$hrs)

t2<-as.data.frame(table(trip$start\_station\_id))

names(t2)=c("station\_id","freq")

t4<-merge(trip,t2,by.x="start\_station\_id",by.y="station\_id")

View(t4)

table(t4$hrs)

m<-t4[!is.na(t4),]

?par

par(mfrow=c(1,3))

boxplot(freq~hrs,m ,ylab="Frequency",xlab="Time of the day",main="For Los Angles")

boxplot(dist~hrs,m,ylab="Distance",xlab="Time of the day",main="For Los Angles")

boxplot(duration~hrs,m,ylab="Duration",xlab="Time of the day",main="For Los Angles")

levels(hour(trip$start\_date))

no\_of\_trips$hrs<-hrs(no\_of\_trips$start\_date)

?distGeo

wq<-merge(no\_of\_stations,trip,by.x="station\_id",by.y="start\_station\_id")

#for san fransisco

tripp<-as.data.frame(table(no\_of\_trips$start\_station\_id))

names(tripp)<-c("sid","freq")

tripp

no\_of\_trips<-merge(no\_of\_trips,tripp,by.x="start\_station\_id",by.y="sid")

names(no\_of\_trips)

wq$dist<-distGeo(c(0,0),cbind(wq$longitude,wq$latitude))

wq$hrs

no\_of\_trips$freq

no\_of\_stations$dist<-distGeo(c(0,0),cbind(no\_of\_stations$longitude,no\_of\_stations$latitude))

no\_of\_trips$duration\_sec

par(mfrow=c(1,3))

w<-as.data.frame(table(no\_of\_trips$start\_station\_id))

names(w)<-c("sid","freq")

wq<-merge(wq,w,by.x="station\_id",by.y="sid")

boxplot(freq.x~hrs,wq,ylab="Frequency",xlab="Time of the day",main="For San Fransisco")

boxplot(duration~hrs,wq,ylab="Duration",xlab="Time of the day",main="For San Fransisco")

boxplot(dist~hrs,wq,ylab="Distance",xlab="Time of the day",main="For San Fransisco")

#6.

library(lubridate)

library(geosphere)

?bearing

hrs= function(no\_of\_stations){

hr<-bearing(cbind(no\_of\_stations$longitude,no\_of\_stations$latitude),c(0,0))

hrs\_con<- cut(hr,breaks = c(-180,-90,0,90,180),right=FALSE,label=c("NEG\_180","NEG 90","Neutral 0","POS 90"))

levels(hrs\_con)[[1]]="NEG\_ROT"

hrs\_con

}

no\_of\_stations$bearing\_angle<-hrs(no\_of\_stations)

table(no\_of\_stations$bearing\_angle)

wq<-merge(no\_of\_stations,trip,by.x="station\_id",by.y="start\_station\_id")

boxplot(bearing\_angle.x~hrs,wq)

class(wq$bearing\_angle.x)

class(wq$hrs)

par(mfrow=c(1,1))

plot(wq$bearing\_angle.x,wq$hrs,xlab="Bearing angle", lab="Time of the day", main="Bearing angle varies with different time of the day")