PradeepPoonati\_project1

pradeep poonati

2/23/2020

#Libraries Used   
  
library(dplyr)

##   
## Attaching package: 'dplyr'

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

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

library(ggplot2)  
library(tidyr)  
library(hrbrthemes)

## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.

## Please use hrbrthemes::import\_roboto\_condensed() to install Roboto Condensed and

## if Arial Narrow is not on your system, please see http://bit.ly/arialnarrow

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

library(ggplot2)  
library(reshape2)

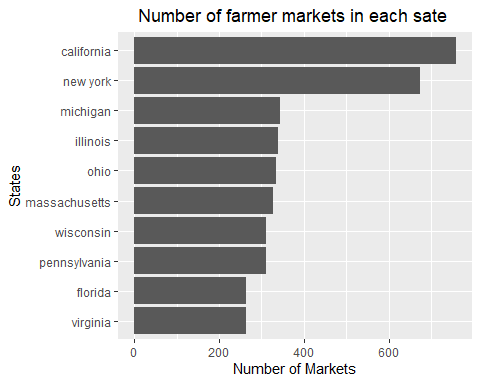
##   
## Attaching package: 'reshape2'

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

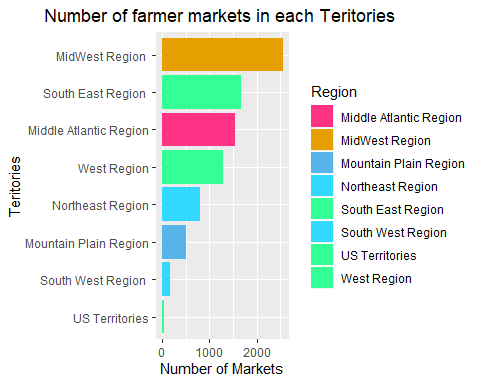
#importing the raw data(fmarket.csv) in to the R studio   
market\_data<-read.csv("fmarket.csv",stringsAsFactors =FALSE,na.strings="")  
Region\_data<-read.csv("RegionsUSA.CSV",stringsAsFactors =FALSE,na.strings="")

#data cleaning for required rows   
Region\_data$State<-tolower(Region\_data$State)  
  
market\_data$State<-tolower(market\_data$State)  
market\_data$Location<-tolower(market\_data$Location)  
  
market\_data$Credit[market\_data$Credit=="Y"]<-1  
market\_data$Credit[market\_data$Credit=="N"]<-0  
market\_data$Credit<-as.integer(market\_data$Credit)  
  
market\_data$WIC[market\_data$WIC=="Y"]<-1  
market\_data$WIC[market\_data$WIC=="N"]<-0  
market\_data$WIC<-as.integer(market\_data$WIC)  
  
market\_data$WICcash[market\_data$WICcash=="Y"]<-1  
market\_data$WICcash[market\_data$WICcash=="N"]<-0  
market\_data$WICcash<-as.integer(market\_data$WICcash)  
  
market\_data$SFMNP[market\_data$SFMNP=="Y"]<-1  
market\_data$SFMNP[market\_data$SFMNP=="N"]<-0  
market\_data$SFMNP<-as.integer(market\_data$SFMNP)  
  
market\_data$SNAP[market\_data$SNAP=="Y"]<-1  
market\_data$SNAP[market\_data$SNAP=="N"]<-0  
market\_data$SNAP<-as.integer(market\_data$SNAP)

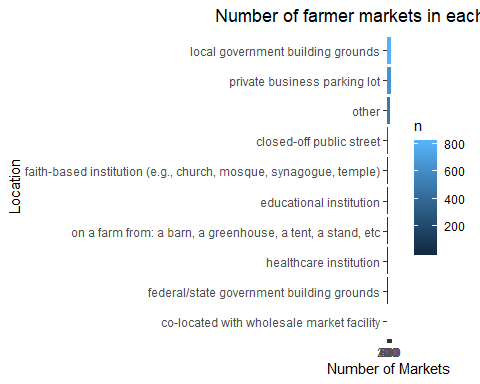
#plot based on geographical location  
  
market\_geographical<-market\_data%>%select(c("FMID","State"))  
market\_state<-market\_geographical%>%group\_by(State)%>%count(State)%>%arrange(desc(n))%>%head(10)  
ggplot(market\_state, aes(y = n, x = reorder(State, n)))+ geom\_bar(stat="identity")+ coord\_flip()+ labs(y="Number of Markets", x="States")+(ggtitle("Number of farmer markets in each sate "))+theme(plot.title = element\_text(hjust = 0.5))



#Graphical analysis by teritory   
  
  
  
#merging region data and farmer market data   
market\_region<-merge(market\_geographical,Region\_data,by.x = "State",by.y = "State")  
market\_region<-market\_region%>%group\_by(Region)%>%count(Region)  
ggplot(market\_region, aes(y = n, x = reorder(Region, n),fill=Region))+ geom\_bar(stat="identity")+ coord\_flip()+ labs(y="Number of Markets", x="Teritories")+(ggtitle("Number of farmer markets in each Teritories "))+theme(plot.title = element\_text(hjust = 0.5))+scale\_fill\_manual(values=c("#FF3386", "#E69F00", "#56B4E9","#33D8FF", "#33FF95","#33D8FF", "#33FF95","#33FF96"))



#plot based on residing area   
market\_area<-select(market\_data,c("FMID","Location"))  
market\_area<-na.omit(market\_area%>%group\_by(Location)%>%count(Location))%>%arrange(desc(n))  
ggplot(market\_area, aes(y = n, x = reorder(Location, n),fill=n))+ geom\_bar(stat="identity")+ coord\_flip()+ labs(y="Number of Markets", x="Location")+(ggtitle("Number of farmer markets in each Location "))+theme(plot.title = element\_text(hjust = 0.5))



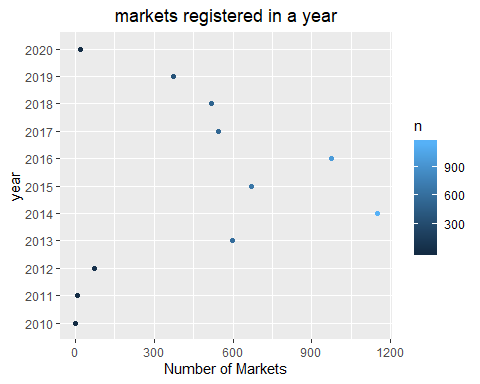
#No of markets registered in a year   
open\_year<-na.omit(select(market\_data,c("FMID","Season1Date")))  
open\_year<-open\_year%>%separate(Season1Date, c("Season1Date1","Season1Date2"), sep = "to")%>%select(c(FMID,Season1Date1))

## Warning: Expected 2 pieces. Additional pieces discarded in 345 rows [43, 51, 58,  
## 120, 127, 147, 151, 157, 158, 167, 175, 192, 210, 233, 234, 267, 269, 270, 275,  
## 301, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 24 rows [575,  
## 837, 1625, 1627, 1630, 1633, 1634, 1635, 1636, 1639, 1669, 2282, 2313, 2526,  
## 2648, 3234, 3607, 3900, 4305, 4435, ...].

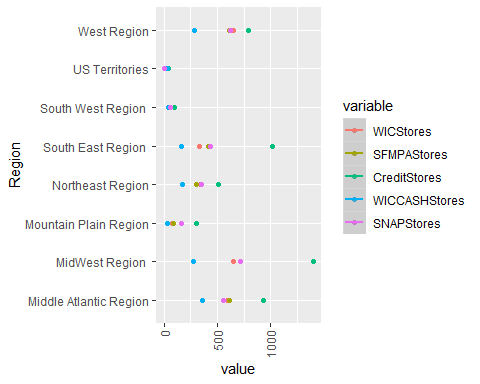
open\_year$Season1Date1<-(mdy(open\_year$Season1Date1,quiet = TRUE))  
open\_year$Season1Date1<-year(open\_year$Season1Date1)  
open\_year<-open\_year%>%filter(!is.na(Season1Date1))%>%count(Season1Date1)  
ggplot(open\_year, aes(x=factor(Season1Date1), y=n,group=n ,color=n)) +geom\_line()+ geom\_point()+ggtitle("markets registered in a year")+labs(y="Number of Markets", x="year")+theme(plot.title = element\_text(hjust = 0.5))+coord\_flip()

## geom\_path: Each group consists of only one observation. Do you need to adjust  
## the group aesthetic?



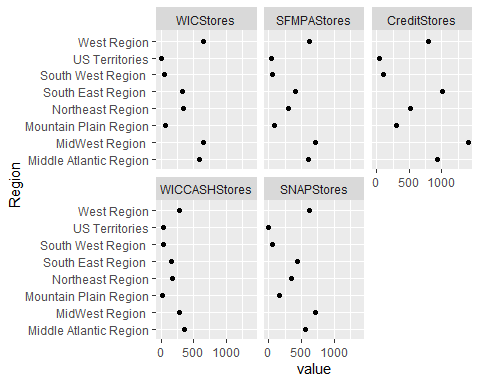
#markets offering card payments   
  
Payment\_System<-market\_data%>%select("State","Credit","WIC","WICcash","SFMNP","SNAP")%>%group\_by(State)%>%summarise(creditacceptingstores=sum(Credit,na.rm = TRUE),WICacceptingstores=sum(WIC,na.rm = TRUE),WICcashacceptingstores=sum(WICcash,na.rm = TRUE),SFMPAcceptingstores=sum(SFMNP,na.rm = TRUE),SNAPAcceptingstores=sum(SNAP,na.rm = TRUE))  
Payment\_System<-merge(Payment\_System,Region\_data,by.x = "State",by.y = "State")  
Payment\_System<-Payment\_System%>%select(-State)%>%group\_by(Region)%>%summarise(WICStores=sum(WICacceptingstores),SFMPAStores=sum(SFMPAcceptingstores),CreditStores=sum(creditacceptingstores),WICCASHStores=sum(WICcashacceptingstores),SNAPStores=sum(SNAPAcceptingstores))  
d <- melt(Payment\_System, id.vars="Region")  
  
# Everything on the same plot  
ggplot(d, aes(Region,value, col=variable)) + geom\_point() + stat\_smooth() +theme(axis.text.x = element\_text(angle = 90, hjust = 1, vjust = 0.5))+coord\_flip()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



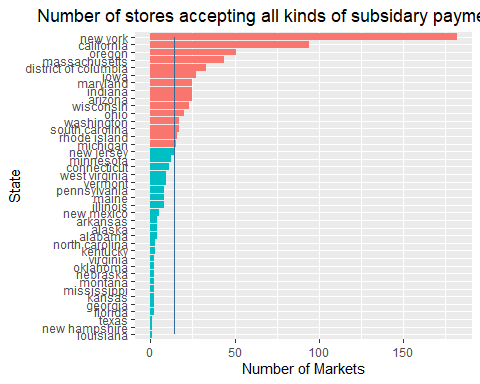
ggplot(d, aes(value,Region)) +   
 geom\_point() +   
 stat\_smooth() +  
 facet\_wrap(~variable)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

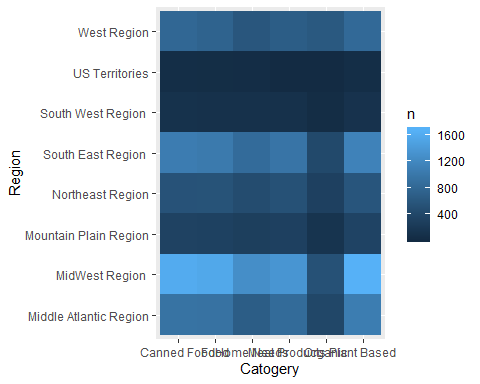


write.csv(Payment\_System,"payment\_system.csv")

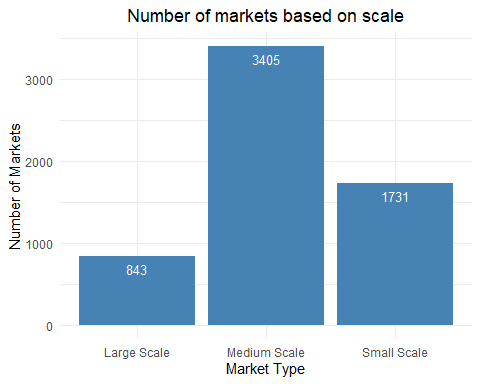
grant\_acceptance<-market\_data[c(11,25:28)]  
grant\_acceptance <- cbind(grant\_acceptance,sum=rowSums(grant\_acceptance[2:5]))  
grant\_acceptance<-grant\_acceptance%>%filter(sum==4)%>%count(State)  
grant\_acceptance$nationsaverage<-sum(grant\_acceptance$n/52)  
ggplot(grant\_acceptance, aes(y =n, x = reorder(State,n),fill=ifelse(n>sum(n/52),"Highlighted","Normal")))+ geom\_bar(stat="identity")+ coord\_flip()+ labs(y="Number of Markets", x="State")+(ggtitle("Number of stores accepting all kinds of subsidary payment systems"))+theme(plot.title = element\_text(hjust = 0.5),legend.position = 'none')+geom\_line(aes(State,nationsaverage,group=nationsaverage,color=nationsaverage))



products\_data<-market\_data[c(1,11,29:58)]  
products\_data<-products\_data%>%gather(key="Products",value = "Availability",-State,-FMID)  
products\_data$Products<-tolower(products\_data$Products)  
Product\_catogery<-read.csv("products .csv",stringsAsFactors =FALSE,na.strings="")  
Product\_catogery$Products<-tolower(Product\_catogery$Products)  
products\_data<-merge(products\_data,Product\_catogery,by.x = "Products",by.y = "Products")  
products\_data<-products\_data%>%filter(Availability=="Y")  
products\_data<-merge(products\_data,Region\_data,by.x = "State",by.y = "State")  
products\_data<-unique(products\_data %>% select(Region,Catogery,FMID))  
products\_data<-products\_data %>% count(Region,Catogery)  
ggplot(data = products\_data, aes(x=Catogery, y=Region, fill=n)) +   
 geom\_tile()



#sclaing the farmer markets according to the size   
market\_size<-market\_data[c(1,29:58)]  
market\_size<-market\_size%>%gather(key = "PRODCUTS",value = "Product",-FMID)  
market\_size<-filter(market\_size,Product=="Y")  
market\_size<-market\_size%>%count(FMID)  
market\_size$n<-ceiling(market\_size$n/10)  
  
market\_size$n[market\_size$n=="3"]<-"Large Scale"  
market\_size$n[market\_size$n=="2"]<-"Medium Scale"  
market\_size$n[market\_size$n=="1"]<-"Small Scale"  
market\_size<-market\_size%>%count(n)  
colnames(market\_size)<-c("markettype","count")  
  
ggplot(data=market\_size, aes(x=markettype, y=count,fill=markettype)) +  
 geom\_bar(stat="identity", fill="steelblue")+  
 geom\_text(aes(label=count), vjust=1.6, color="white", size=3.5)+  
 theme\_minimal()+labs(y="Number of Markets", x="Market Type")+(ggtitle("Number of markets based on scale"))+theme(plot.title = element\_text(hjust = 0.5),legend.position = 'none')



#sclaing the farmer markets according to the size   
market\_size<-market\_data[c(1,29:58)]  
market\_size<-market\_size%>%gather(key = "PRODCUTS",value = "Product",-FMID)  
market\_size<-filter(market\_size,Product=="Y")  
market\_size<-market\_size%>%count(FMID)  
market\_size$n<-ceiling(market\_size$n/10)  
  
market\_size$n[market\_size$n=="3"]<-"Large Scale"  
market\_size$n[market\_size$n=="2"]<-"Medium Scale"  
market\_size$n[market\_size$n=="1"]<-"Small Scale"  
market\_size<-market\_size%>%count(n)  
colnames(market\_size)<-c("markettype","count")  
  
ggplot(data=market\_size, aes(x=markettype, y=count,fill=markettype)) +  
 geom\_bar(stat="identity", fill="steelblue")+  
 geom\_text(aes(label=count), vjust=1.6, color="white", size=3.5)+  
 theme\_minimal()+labs(y="Number of Markets", x="Market Type")+(ggtitle("Number of markets based on scale"))+theme(plot.title = element\_text(hjust = 0.5),legend.position = 'none')

