group-project.R

HUANG JIANXIN

2021-12-03

setwd("C:/Users/47010/Downloads/New folder")  
data<-read.csv("hotel\_bookings.csv")  
model\_data<-data[,c(1,2,3,10,11,12,13,15,16,17,18,19,20,23,26,27,28,29)]  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

library(Rmisc)

## Warning: package 'Rmisc' was built under R version 4.0.5

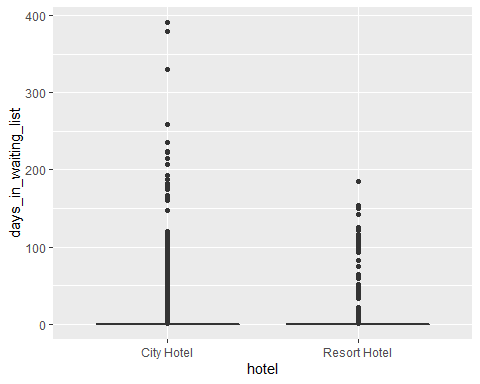
## Loading required package: lattice

## Warning: package 'lattice' was built under R version 4.0.5

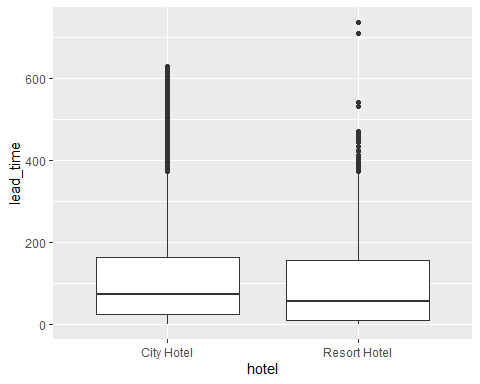
## Loading required package: plyr

## Warning: package 'plyr' was built under R version 4.0.5

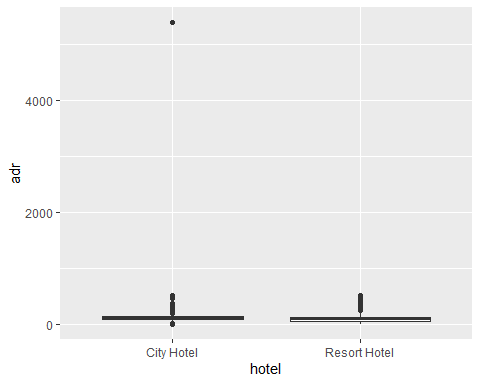
p1<-ggplot(model\_data,aes(x=hotel,y=days\_in\_waiting\_list))+geom\_boxplot()  
p1



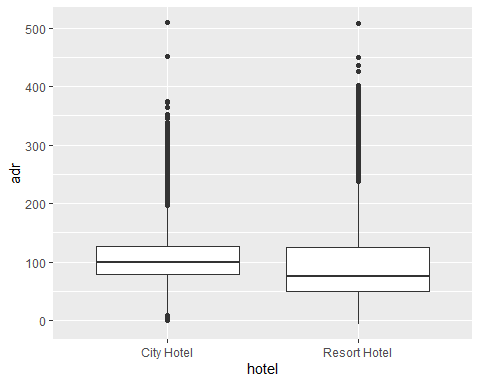
p2<-ggplot(model\_data,aes(x=hotel,y=lead\_time))+geom\_boxplot()  
p2



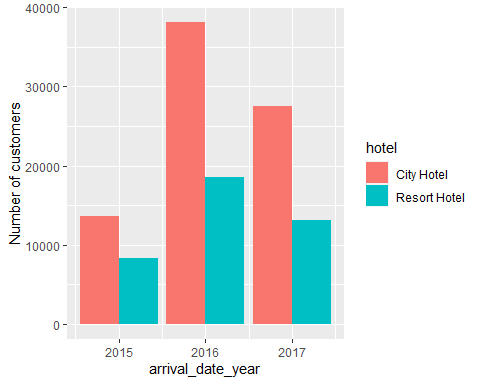
p3<-ggplot(model\_data,aes(x=hotel,y=adr))+geom\_boxplot()  
p3



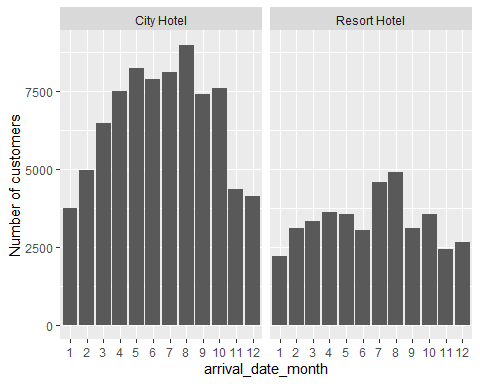
data\_remove\_adroutlier<-subset(model\_data,model\_data$adr<4000)  
p4<-ggplot(data\_remove\_adroutlier,aes(x=hotel,y=adr))+geom\_boxplot()  
p4



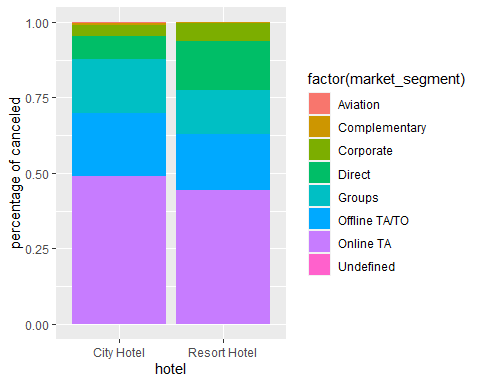
p5<-ggplot(data,aes(x=arrival\_date\_year,fill=hotel))+  
 geom\_bar(position='dodge')+ ylab("Number of customers")  
p5



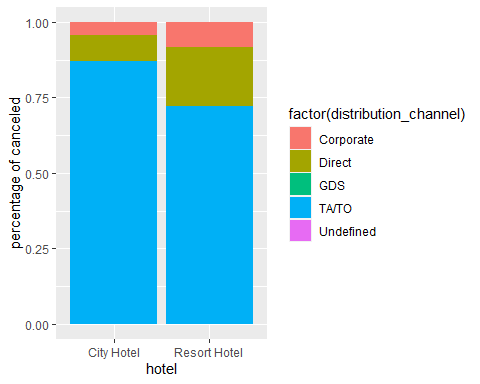
data$arrival\_date\_month<-factor(data$arrival\_date\_month,levels=c("January","February","March","April","May","June","July","August","September","October","November","December"),labels=c(1:12))  
p6<-ggplot(data,aes(x=arrival\_date\_month))+geom\_bar()+facet\_wrap(~hotel)+ ylab("Number of customers")  
p6



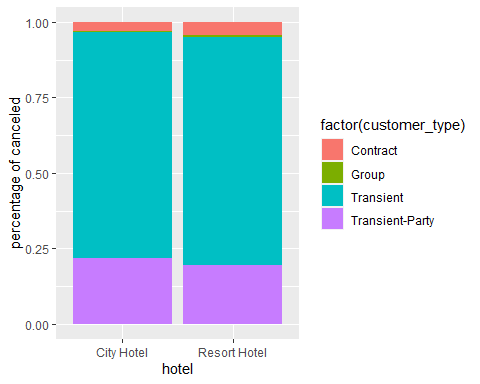
p7<-ggplot(data\_remove\_adroutlier,aes(hotel,fill=factor(market\_segment)))+  
 geom\_bar(position="fill")+ylab("percentage of canceled")  
p7



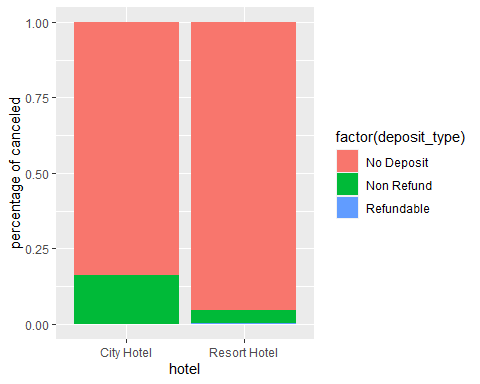
p8<-ggplot(data\_remove\_adroutlier,aes(hotel,fill=factor(distribution\_channel)))+  
 geom\_bar(position="fill")+ylab("percentage of canceled")  
p8



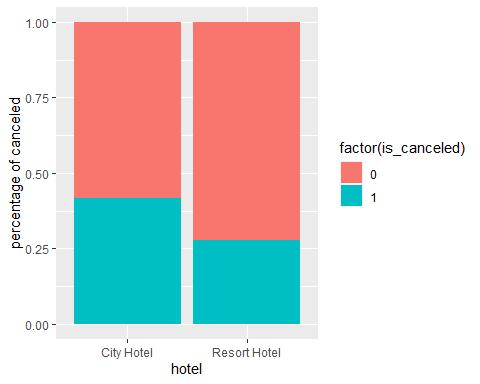
p9<-ggplot(data\_remove\_adroutlier,aes(hotel,fill=factor(customer\_type)))+  
 geom\_bar(position="fill")+ylab("percentage of canceled")  
p9



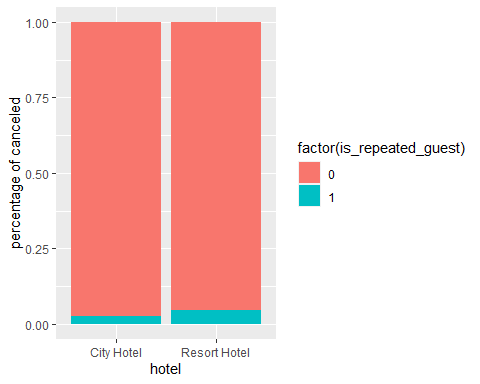
p10<-ggplot(data\_remove\_adroutlier,aes(hotel,fill=factor(deposit\_type)))+  
 geom\_bar(position="fill")+ylab("percentage of canceled")  
p10



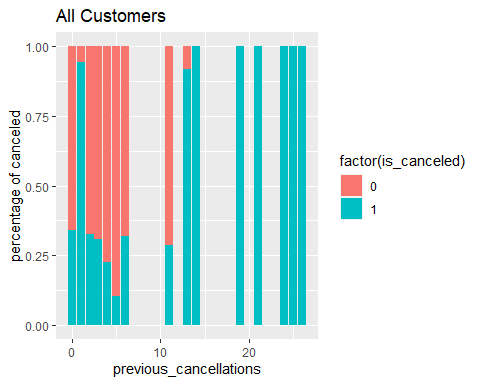
p11<-ggplot(data\_remove\_adroutlier,aes(hotel,fill=factor(is\_canceled)))+  
 geom\_bar(position="fill")+ylab("percentage of canceled")  
p11



p12<-ggplot(data\_remove\_adroutlier,aes(hotel,fill=factor(is\_repeated\_guest)))+  
 geom\_bar(position="fill")+ylab("percentage of canceled")  
p12



data\_repeated\_customer<-subset(data\_remove\_adroutlier,data\_remove\_adroutlier$is\_repeated\_guest==1)  
p13<-ggplot(data\_remove\_adroutlier,aes(previous\_cancellations,fill=factor(is\_canceled)))+  
 geom\_bar(position="fill")+ylab("percentage of canceled")+ggtitle("All Customers")  
p13



p14<-ggplot(data\_repeated\_customer,aes(previous\_cancellations,fill=factor(is\_canceled)))+  
 geom\_bar(position="fill")+ylab("percentage of canceled")+ggtitle("Repeated Customers")  
p14

