HW 2 problem 1

Carter Pearon

August 19, 2015

## 1

The graphic below depicts how the type of flight cancellation varied by month for the entire year of 2008. The cancellation codes are A = carrier, B = weather, C = NAS, D = security. It turns out there are 0 observations of cancellation code = D, i.e. a cancelled flight due to security reasons, and so we will on consider cancellation codes of A, B and C.

#Read in the data  
  
flights = read.csv('ABIA.csv', header = TRUE)  
  
# select only the flights that were cancelled  
cancelled = subset(flights, flights$Cancelled == 1)  
  
# now lets reduce the dataframe to only include our columns of interest  
cancelled = cancelled[,c(2,23)]  
  
# subset the data by cancellation code  
type\_A = subset(cancelled, CancellationCode == "A")  
type\_B = subset(cancelled, CancellationCode == "B")  
type\_C = subset(cancelled, CancellationCode == "C")  
  
# count function to be used in the next chunck of code to count the number of each cancellation type per month  
count = function(x) {  
 length(x)  
}  
  
total\_cancels\_by\_month = aggregate(CancellationCode ~ Month, cancelled, count)  
type\_A\_by\_month = aggregate(CancellationCode ~ Month, type\_A, count)  
type\_B\_by\_month = aggregate(CancellationCode ~ Month, type\_B, count)  
type\_C\_by\_month = aggregate(CancellationCode ~ Month, type\_C, count)  
  
newrow = c(11, 0)  
  
# needed to add a row for type C, month 11, as it had 0 observations  
type\_C\_by\_month = rbind(type\_C\_by\_month[1:10,],newrow,type\_C\_by\_month[(11),])  
  
#Combine all the months into one data frame, including the total # across all types  
type\_by\_month = cbind(type\_A\_by\_month,type\_B\_by\_month[,2],type\_C\_by\_month[,2],total\_cancels\_by\_month[,2])  
  
# make the column names neater  
names(type\_by\_month)[1] = paste('Month')  
names(type\_by\_month)[2] = paste('carrier')  
names(type\_by\_month)[3] = paste('weather')  
names(type\_by\_month)[4] = paste('NAS')  
names(type\_by\_month)[5] = paste('Total')  
  
# create an empty plot with axis labels and the right dimensions  
plot(1, type="n",main = "Cancellations per Month by Type", xlab="Month", ylab="# of Cancellations", xlim=c(1,12), ylim=c(0,235 ))  
  
# plot our lines  
lines(type\_by\_month$Month,type\_by\_month$carrier,col="purple", type = "b", main = "carrier", pch= 5)  
lines(type\_by\_month$Month,type\_by\_month$weather,col="yellow", type = "b", pch= 3)  
lines(type\_by\_month$Month,type\_by\_month$NAS,col="blue", type = "b", pch= 1)  
lines(type\_by\_month$Month,type\_by\_month$Total, col="black")

