INFM600 R-Plot

## R Markdown

setwd("E:\\Fall\_2016\\INFM\_600\\Project\\R-plot")  
d=read.csv("Airplane\_Crash\_Summary\_Final.csv")  
AirCrashTimeSlots=read.csv("AirCrashTimeSlots.csv")  
crash=read.csv("top\_5\_crash.csv")

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.2

library(extrafont)

## Warning: package 'extrafont' was built under R version 3.3.2

## Registering fonts with R

library(xkcd)

## Warning: package 'xkcd' was built under R version 3.3.2

font\_import(pattern="[C/c]omic")

## Importing fonts may take a few minutes, depending on the number of fonts and the speed of the system.  
## Continue? [y/n]

## Exiting.

fonts()

## [1] "xkcd" "Comic Sans MS"

loadfonts(device="win")

## xkcd already registered with windowsFonts().

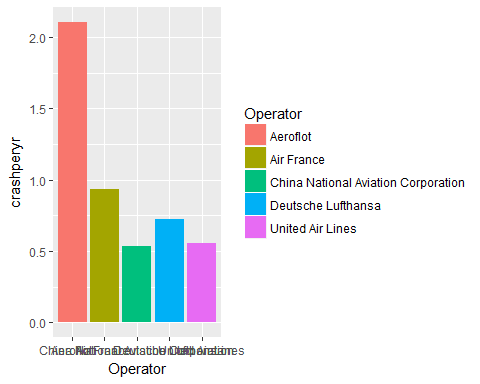
## Comic Sans MS already registered with windowsFonts().

theme\_xkcd <- theme(  
 panel.background = element\_rect(fill="whitesmoke"),  
 panel.grid = element\_line(colour="white"),  
 axis.text.x = element\_text(colour="black"),  
 text = element\_text(size=10, family="Comic Sans MS")  
)  
Legend<-crash$Operator  
ggplot(crash, aes(x=crash$Operator, y=crash$freq, fill=Legend))+stat\_summary(geom="bar") +  
 labs(y = "Number of air crashes") + labs(title = "Bar Graph: Number of air crashes since 1908 for top 5 operators with maximum number of crashes") +scale\_x\_discrete("Operators", waiver(), NULL)+  
 theme\_xkcd

## No summary function supplied, defaulting to `mean\_se()

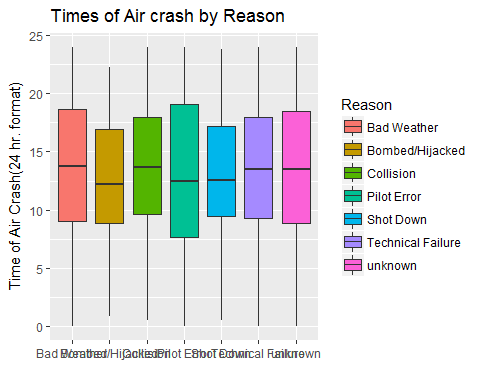
##This graph shows the total number of air crashes for the top 5 operators with maximum number of aircrashes since 1908.As we can see from this graph Aeroflot has the maximum number of aircrashes (around 180), which is huge compared to other airlines. We also found out that these airlines started opearting around the same time i.e. early 1920's. Further analysis will be required to help us answer the question what are the factors for such a high crash rate for Aeroloft.One important thing to note here is that we have excluded Military crashes for this analysis.

ggplot(crash, aes(x=Operator, y=crashperyr,fill=Operator)) +   
 geom\_bar(stat="identity")

 ##We decided to look into details regarding how many crashes happen per year for each of the top five operators. For this, we found out the year each operator began operating, and divided the total number of crashes for each by the number of years the operator has been active up till 2008. We found out that even though all of these airlines have started around the same time, a couple of them have significantly higher crash rate value per year as compared to others.For instace, Aeroflot has a value of almost 2 air crashes per year up to 2008 from the time it started operating. Further analysis can be done using these details to find out the exact reasons that are responsible for such a high crash rate for these airlines.

qplot(Reason, newTime, data=AirCrashTimeSlots, geom=c("boxplot"),fill=Reason, main="Times of Air crash by Reason",  
 xlab="", ylab="Time of Air Crash(24 hr. format)",ylim = c(0,24))

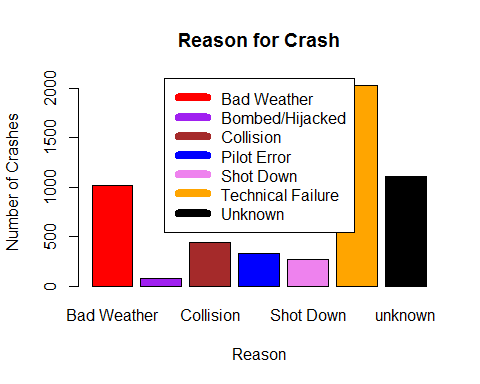
## Warning: Removed 7 rows containing non-finite values (stat\_boxplot).

 ##This is a box plot. The boxes represent the 1st quartile, median,3rd quartile and the minimum and maximum values for each Reason.The solid black line represents the median time of crash for each reason.This As can be seen from the box plots, the timings for the air-crash are least spread for Bombed/Hijacked, Shot Down and maximum forPilot error.We can make use of this graph to identify the time of crahes for each type of reason and then analyze them further to support the trends observed.

Legend=AirCrashTimeSlots$slot  
barplot(table(AirCrashTimeSlots$slot),main="Day v/s Night crashes",xlab="Time of the day",ylab="Number of Crashes",col=c("yellow","black"),ylim = c(0,2000))  
legend("topright",   
 legend = c("Day", "Night"),   
 fill = c("yellow", "black"))

##The above plot shows us the number of crashes that have happened during the day time versus the number of crashes that have occured during the night time. The graph reveals that more crashes have taken place during the day time as compared to the night.

catcolors=c("red","purple","brown","blue","violet","orange","black")  
barplot(table(d$Reason),main="Reason for Crash",xlab="Reason",ylab="Number of Crashes",col=catcolors,ylim = c(0,2100))  
legend("top",c("Bad Weather","Bombed/Hijacked","Collision","Pilot Error","Shot Down","Technical Failure","Unknown"),lwd=8,col=catcolors)

 ##The above plot shows details regarding the number of crashes that have happened due to the various reasons we have assigned. Reason of the crash was assigned to each crash after examining the explanation given in the crash summary. As seen from the graph, a majority of air crashes have taken place due to technical failure, followed by bad weather conditions. It is interesting to note that more than 350 air crashes have happened as a result of pilot/crew errors.