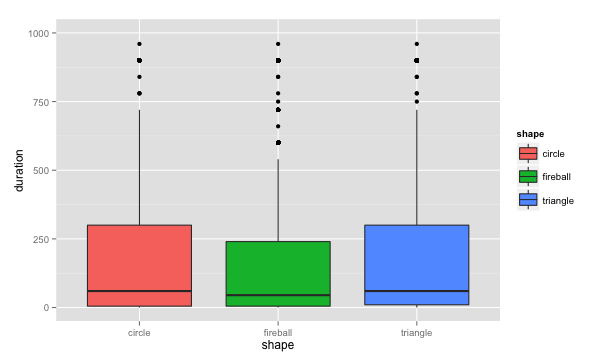
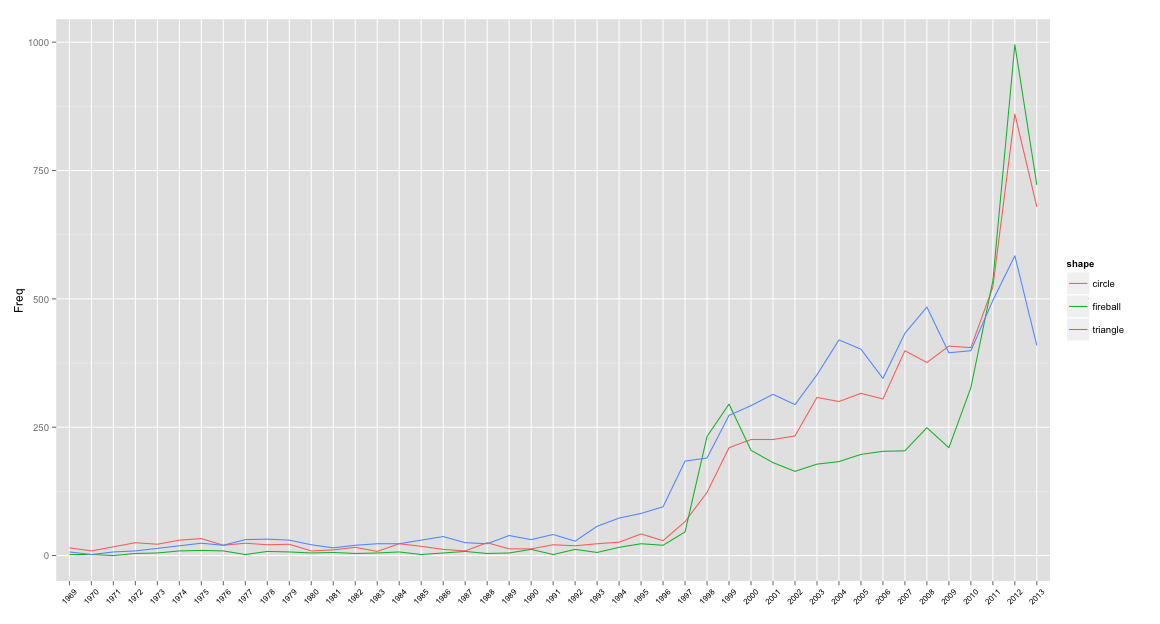
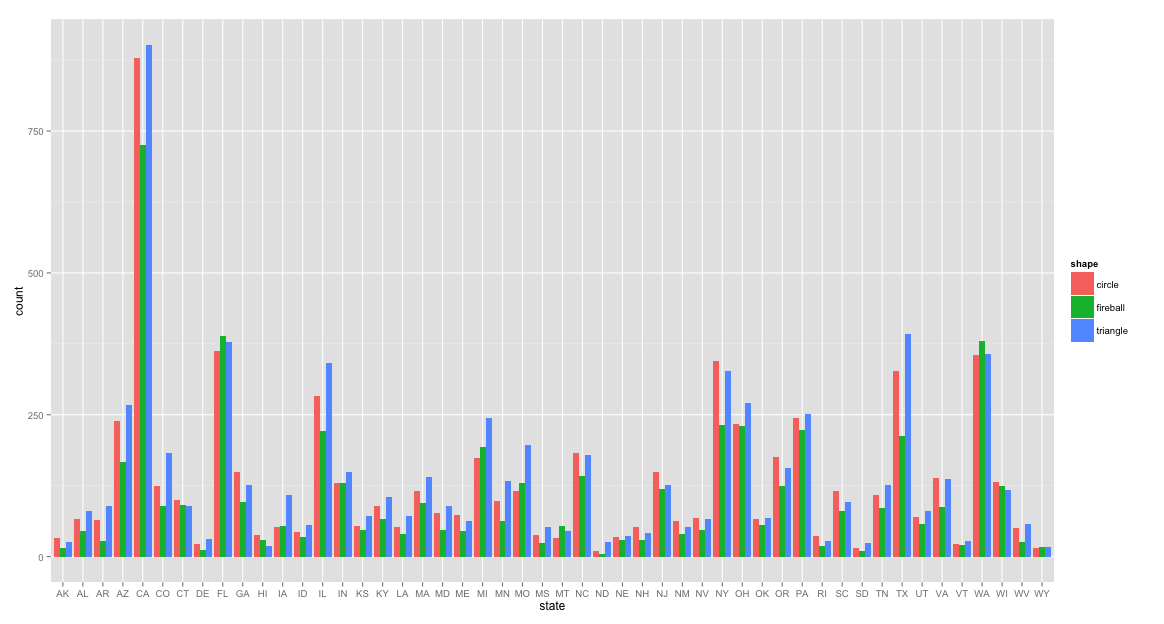
1.a.i) ggplot(UFOData, aes(y=duration,x=shape,fill=shape)) + ylim(0, 1000) + geom\_boxplot()



1.a.ii) ggplot(timeseries\_table,aes(x=year,y=Freq,colour=shape,group=shape)) + geom\_line() + theme(axis.text.x = element\_text(colour = 'black', angle = 45, size = 8, hjust = .5, vjust = .5),axis.title.x=element\_blank())

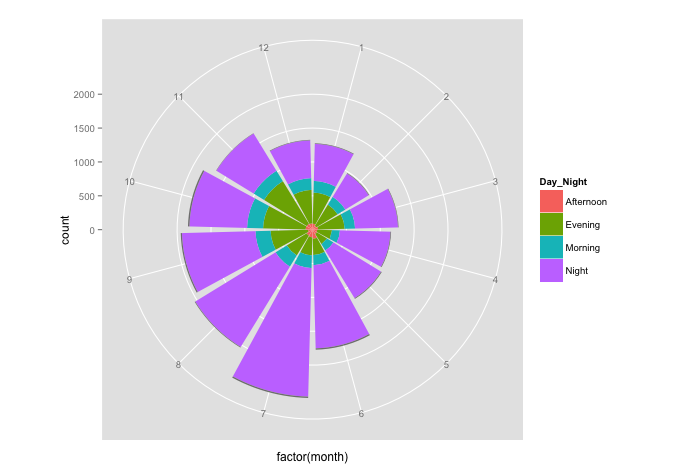


1.a.iii) ggplot(UFOData, aes(state, fill=shape)) + geom\_bar(position="dodge")



1.a.iv) Custom plot for the following question: In what months are the most sightings occur? J

ggplot(UFOData, aes(x=factor(month),fill=Day\_Night)) + geom\_bar() + coord\_polar(theta="x")



1.b.i) #normalize the data by state

shape\_state\_sub <- subset(UFOData,shape=="circle"|shape=="fireball"|shape=="triangle", select=c(shape,state))

state\_table <- table(shape\_state\_sub)

state\_table <- as.data.frame(state\_table)

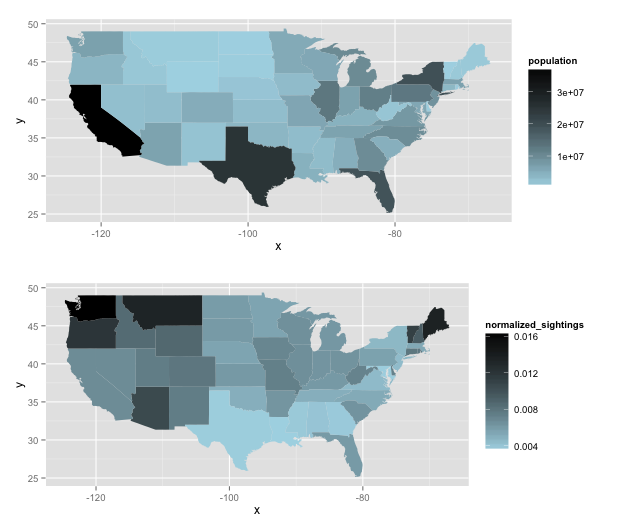
1.b.ii)

vplayout <- function(x, y) viewport(layout.pos.row = x, layout.pos.col = y)

* plot1 <- ggplot(sightings\_population, aes(map\_id = region)) + geom\_map(aes(fill = population), map = state\_map) + expand\_limits(x = state\_map$long, y = state\_map$lat) + scale\_fill\_gradient(low = "lightblue", high = "black")
* plot2 <- ggplot(sightings\_population, aes(map\_id = region)) + geom\_map(aes(fill = normalized\_sightings), map = state\_map) + expand\_limits(x = state\_map$long, y = state\_map$lat) + scale\_fill\_gradient(low = "lightblue", high = "black")

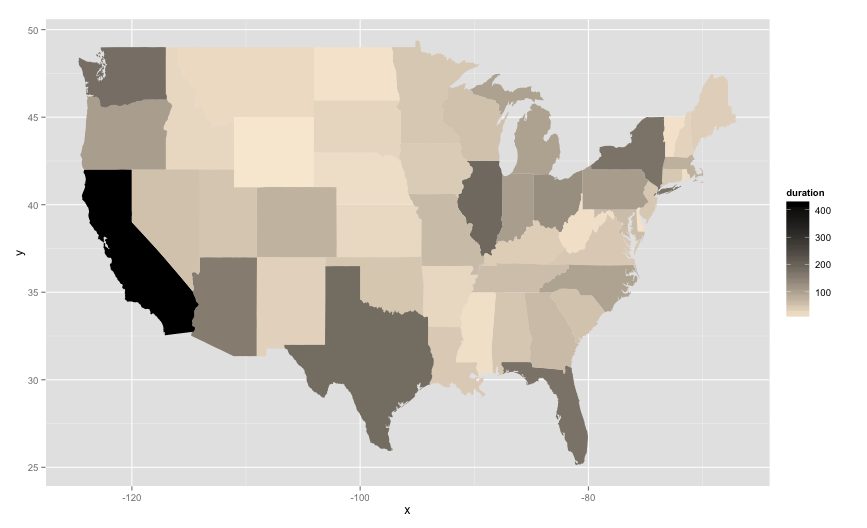
pushViewport(viewport(layout = grid.layout(2, 1)))

* print(plot1, vp = vplayout(1, 1))
* print(plot2, vp = vplayout(2, 1))

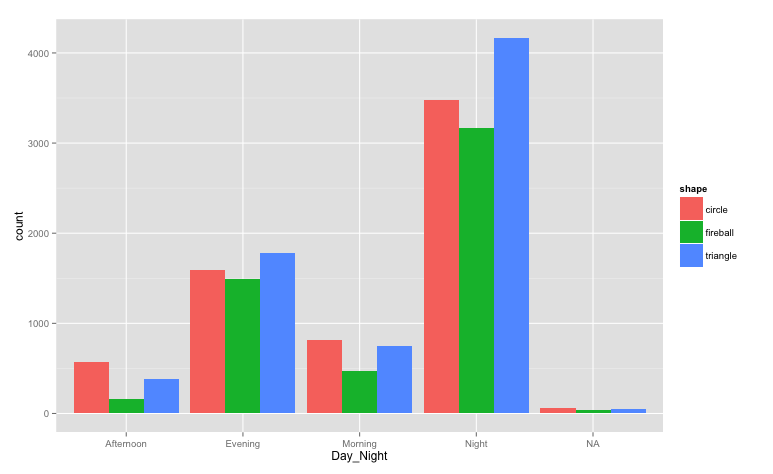


1.b.iii) Two additional questions

1) What is the distribution of the length of sightings? Based on the number of hours:



2) What shapes are seen at what part of the day?

ggplot(UFOData, aes(x=Day\_Night,fill=shape)) + geom\_bar(position="dodge

1.c.) I first used a simple linear regression model to predict the different shapes, then I used a randomForest. The variances using both methods were so high that it would be recommended to simply take 1-predicted\_value.

2.a) Describe infographic i would like to create

* Show what websites are popular
* Indicate what category the most popular websites are
* Show which websites are ranked as relevant
* The demand for new content

2.b)

* packages in R, such as rgl and ggplot 2
* microsoft word to organize the infograph

2.c) I created an infographic that shows the relationship between popular websites, their categorical classification, and their relevance, and the goal is to draw inferences from these relationships.

What’s evident is that popular media and social media websites, while they have a large amount of content and therefore frequently visited, aren’t always the most relevant. These rankings are specific to each user not all types of videos on youtube or articles on sportsillustrated will be of interest. However very specific and niche websites you will begin observing very high relevance rates.

As a result, there are some interesting inferences that can be drawn from this information. Based on what type of websites they have visited, the frequency of those visits, and the relevance of those visits , you can begin to make inferences about the users demographic information, such as their age, sex, and perhaps even their location.

2d) I began the visualization with some simple graphs and statistics about the websites that have been visited and ranked by the user and began describing all the information. It was very simple and rudimentary, and I realized that I needed more visuals to substitute for the text.

I then added a lot of colors and shading to the background of the page, however at a certain point it became too distracting, and began becoming more of a distraction to the reader (me). I decided to lose the shadings and backgrounds and add in descriptive pictures instead.

The pictures I added were of website logos to compliment and replace some of the text. I also added simple graphics to assist the reader in understanding the graphs and as well as pointing out the most important aspects of those graphs.

3) Creating a visualization system for people who can’t see can obviously be quite difficult. The easiest way (probably) for a person who can’t see to understand how something looks is by using the sense of touch and hearing.