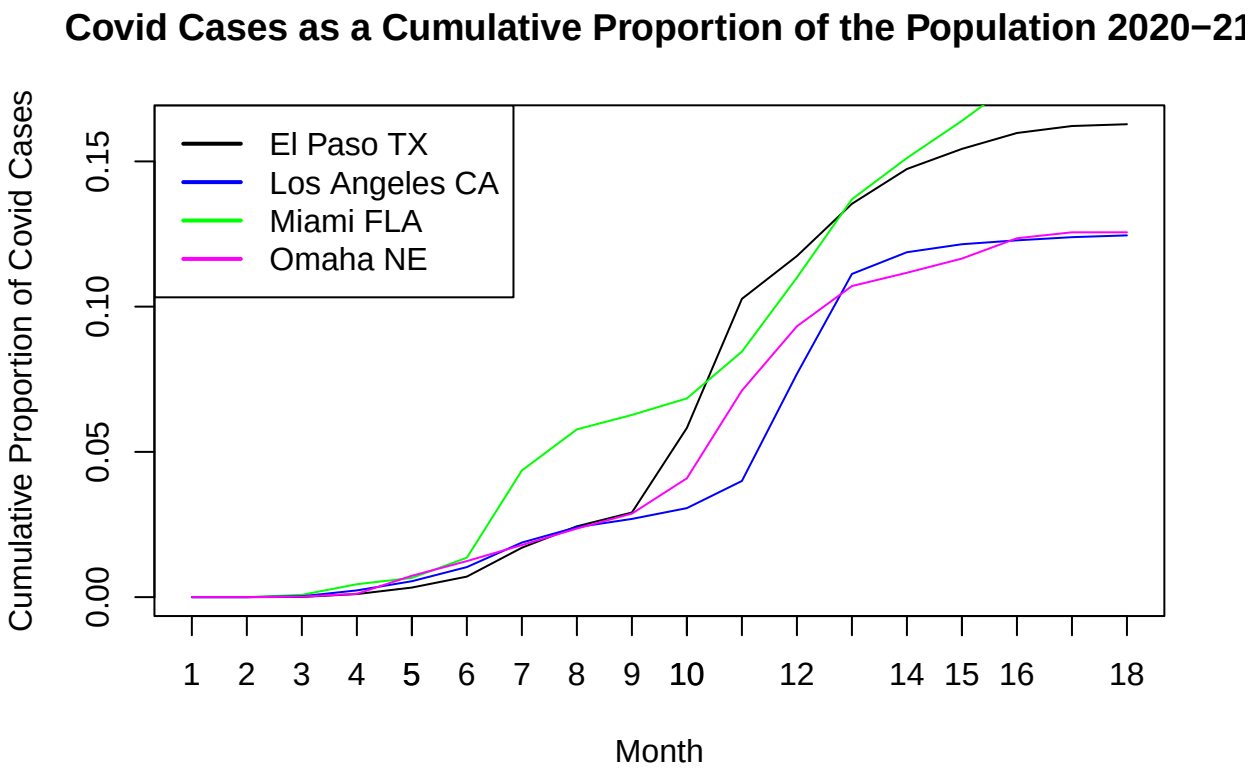
R Data Visualization

Alexander Shih

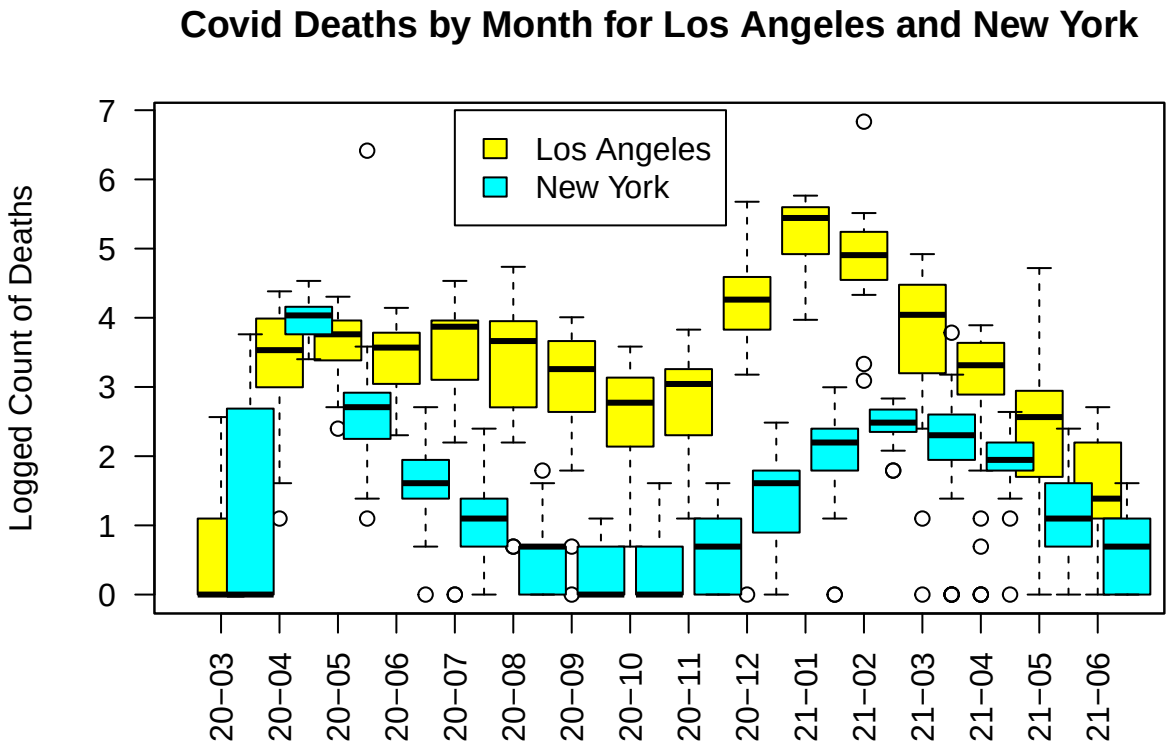
Graph of US COVID cases as a cumulative proportion of the US population 2020-2021 (1 represents January 2020)

```
load("~/R/covid_cases.RData")
plot(EP, type="l", ann=FALSE)
title(main = "Covid Cases as a Cumulative Proportion of the Population 2020-21",
      xlab = "Month", ylab = "Cumulative Proportion of Covid Cases")
lines(FL, col = "green")
lines(LA, col = "blue")
lines(OM, col = "magenta")
axis(1, at=1:18)
legend(x="topleft", legend=c("El Paso TX", "Los Angeles CA", "Miami FLA",
                             "Omaha NE"), col=c("black", "blue", "green",
                                                "magenta"), lwd=2)
```



Covid deaths by month for Los Angeles and New York (y-axis is log scale for better sense of scale)

```
load("~/R/covid_deaths.RData")
LAX$value = log(LAX$value + 1)
boxplot(value ~ y_m, data = LAX[ , c(17,14)], col="yellow", ann=FALSE,
        las=2)
NYC$value = log(NYC$value + 1)
boxplot(value ~ y_m, data = NYC[ , c(17,14)], col="cyan", add=TRUE,
        at=seq(1.5, 16.5, 1), xaxt="n", yaxt="n")
title(main = "Covid Deaths by Month for Los Angeles and New York", ylab =
        "Logged Count of Deaths")
legend(x=5, y=7, legend=c("Los Angeles", "New York"), fill=c("yellow",
                                                             "cyan"))
```



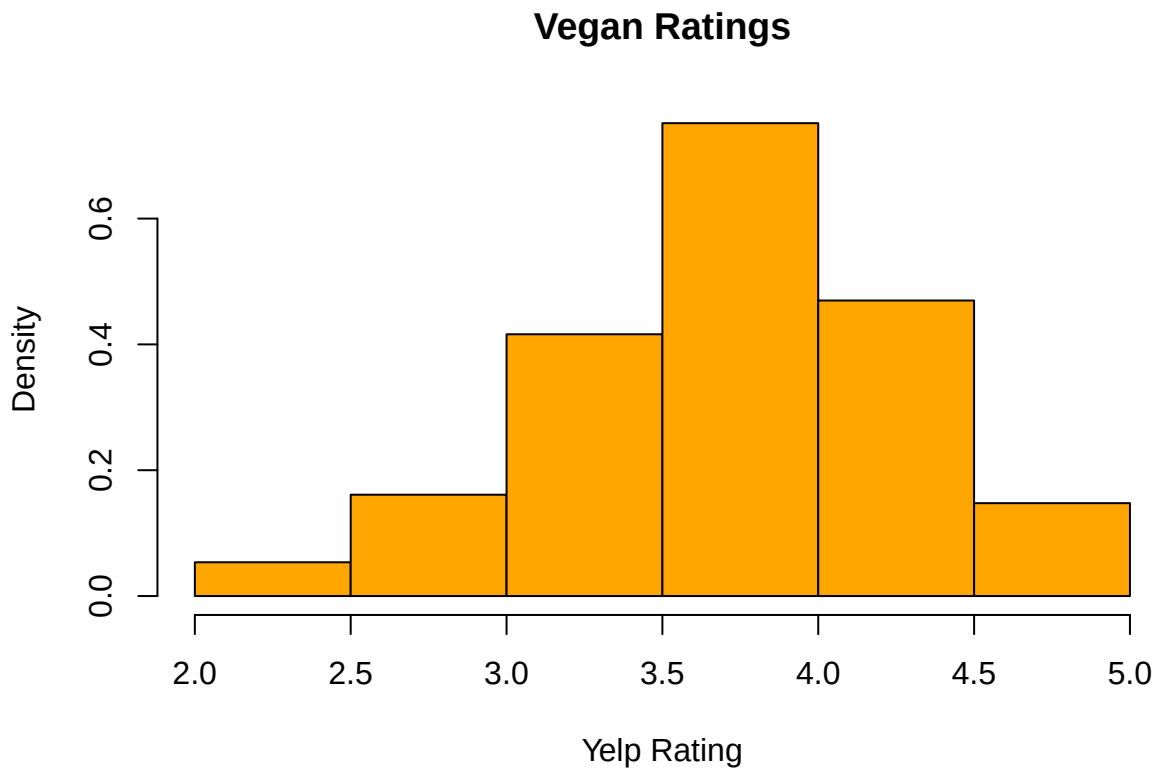
Histograms of Yelp reviews for various types of restaurants

```
library(readr)
yelp3 <- read_csv("~/R/yelp3.csv")
```

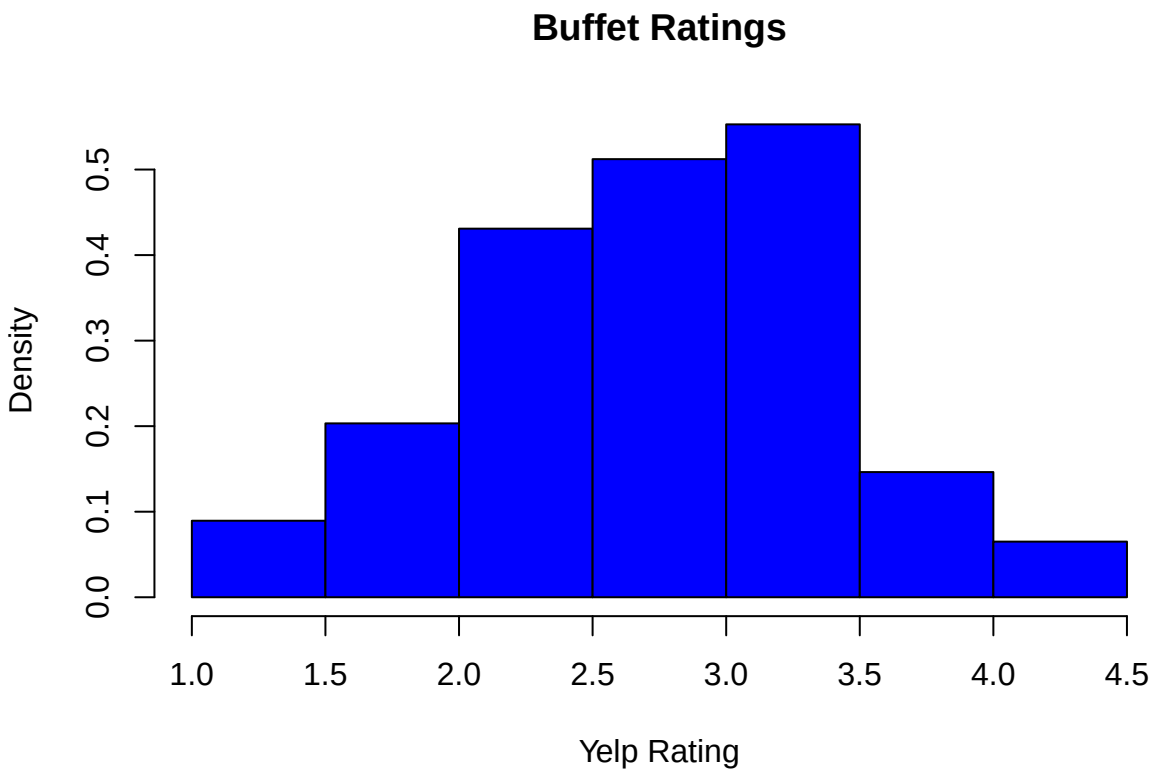
```
##
## -- Column specification --------------------------------------------------------
## cols(
##   business_id = col_character(),
##   categories.0 = col_character(),
##   city = col_character(),
##   latitude = col_double(),
##   longitude = col_double(),
##   name = col_character(),
##   stars = col_double()
```

```
## )
```

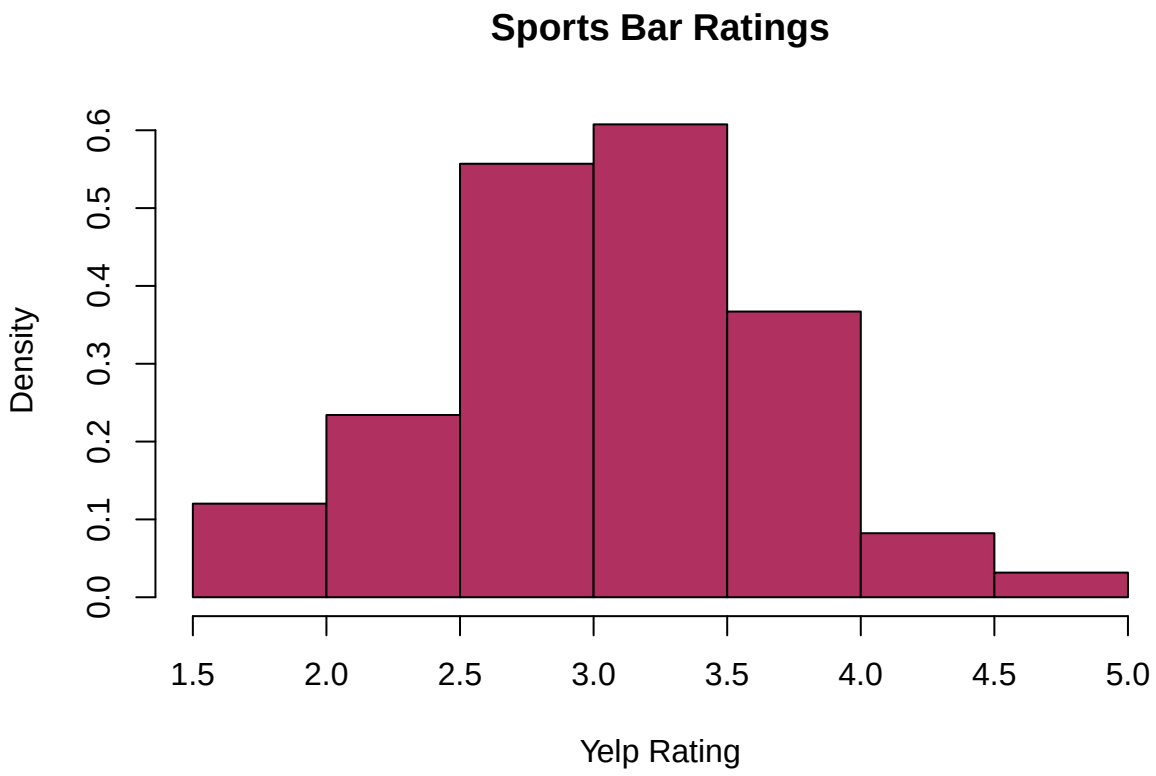
```
vegan <- yelp3[yelp3$categories.0 == "Vegan", 7]
buffet <- yelp3[yelp3$categories.0 == "Buffets", 7]
sports_bars <- yelp3[yelp3$categories.0 == "Sports Bars", 7]
##par(mfrow = c(3,1))
hist(vegan$stars, freq = FALSE, col = "orange", main = "Vegan Ratings",
     xlab = "Yelp Rating")
```



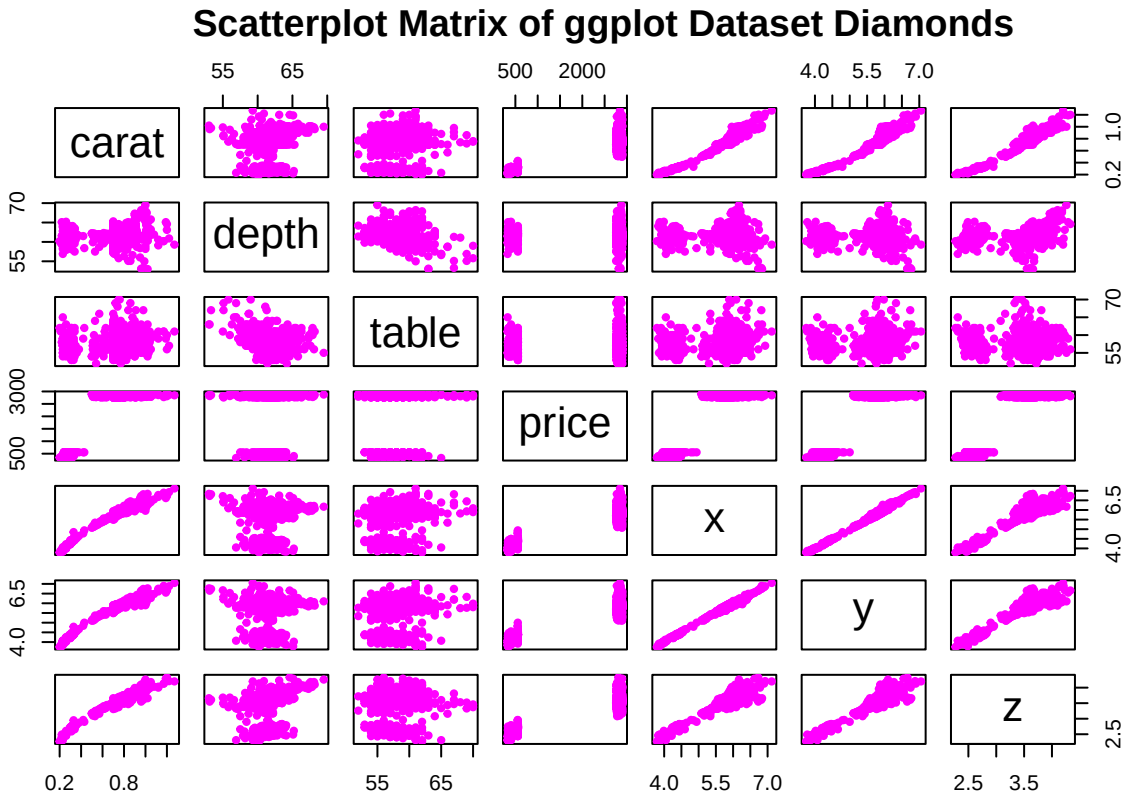

```
hist(buffet$stars, freq = FALSE, col = "blue", main = "Buffet Ratings",
     xlab = "Yelp Rating")
```




```
hist(sports_bars$stars, freq = FALSE, col = "maroon",
     main = "Sports Bar Ratings", xlab = "Yelp Rating")
```

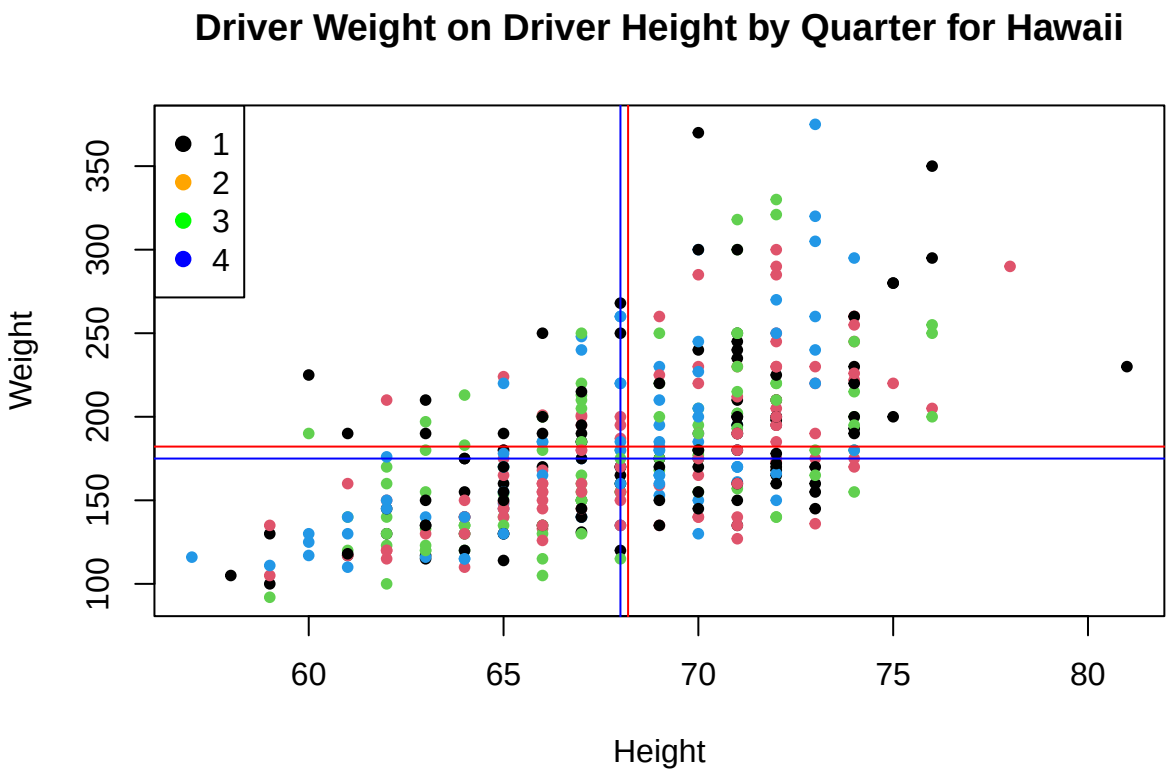



```
library(ggplot2)
d <- diamonds[1:1000, c(1, 5, 6, 7, 8, 9, 10)]
plot(d, col="magenta", pch=20)
title("Scatterplot Matrix of ggplot Dataset Diamonds", line = 3)
```



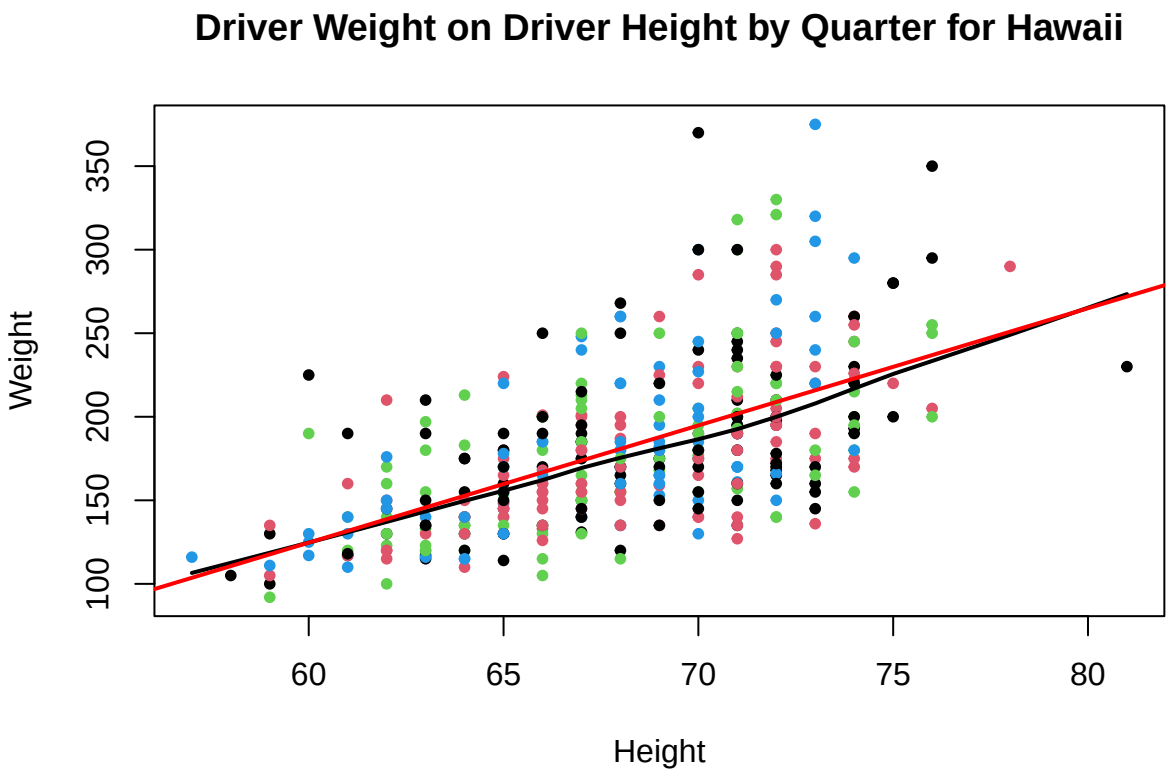
A graph of car accidents in Hawaii by height and weight of the driver (color represents the quarter of the year in which the accident occured). The red and blue lines are the means and medians, respectively, for each axis.

```
load("~/R/labdata20SU21.RData")
labdata20SU21$COLOR <- cut(labdata20SU21$MONTH, breaks=c(0, 4, 7, 10, 13),
                           labels=c("black", "orange", "green", "blue"))
temp <- labdata20SU21[labdata20SU21$STATENAME == "Hawaii" &
                        labdata20SU21$DR_HGT != 999 &
                        labdata20SU21$DR_WGT != 999, ]
plot(DR_WGT ~ DR_HGT, data=temp, col=temp$COLOR, pch=20,
     main="Driver Weight on Driver Height by Quarter for Hawaii",
     xlab="Height", ylab="Weight")
lines(x=rep(mean(temp$DR_HGT), 9), y=seq(0, 400, 50), col="red")
lines(x=rep(median(temp$DR_HGT), 9), y=seq(0, 400, 50), col="blue")
lines(x=seq(0,85,17), y=rep(mean(temp$DR_WGT),6), col="red")
lines(x=seq(0,85,17), y=rep(median(temp$DR_WGT),6), col="blue")
legend(x="topleft", legend=c(1,2,3,4), col=c("black", "orange", "green",
                                             "blue"), pch=19)
```



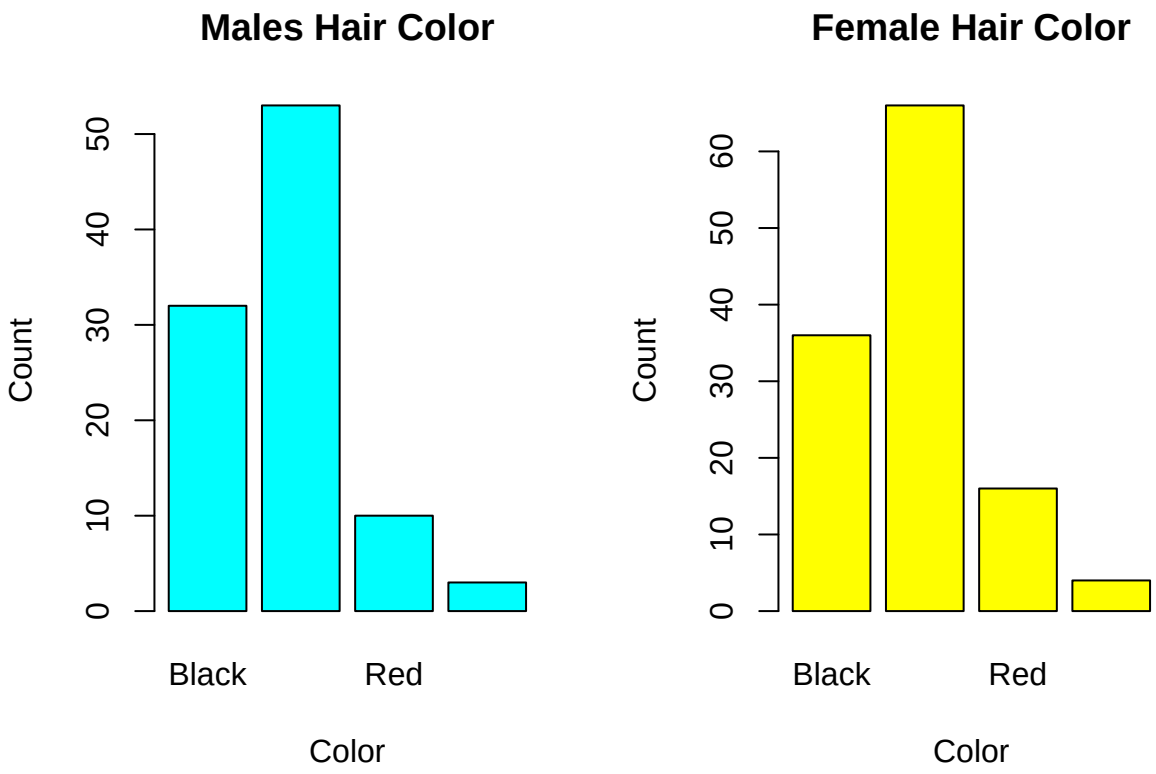
The same graph as before, but with a linear regression line (red) and a scatterplot smoothing line (black)

```
plot(DR_WGT ~ DR_HGT, data=temp, col=temp$COLOR, pch=20,
     main="Driver Weight on Driver Height by Quarter for Hawaii",
     xlab="Height", ylab="Weight")
lines(lowess(temp$DR_HGT, temp$DR_WGT), lwd=2)
abline(lm(DR_WGT ~ DR_HGT, data=temp), col="red", lwd=2)
```



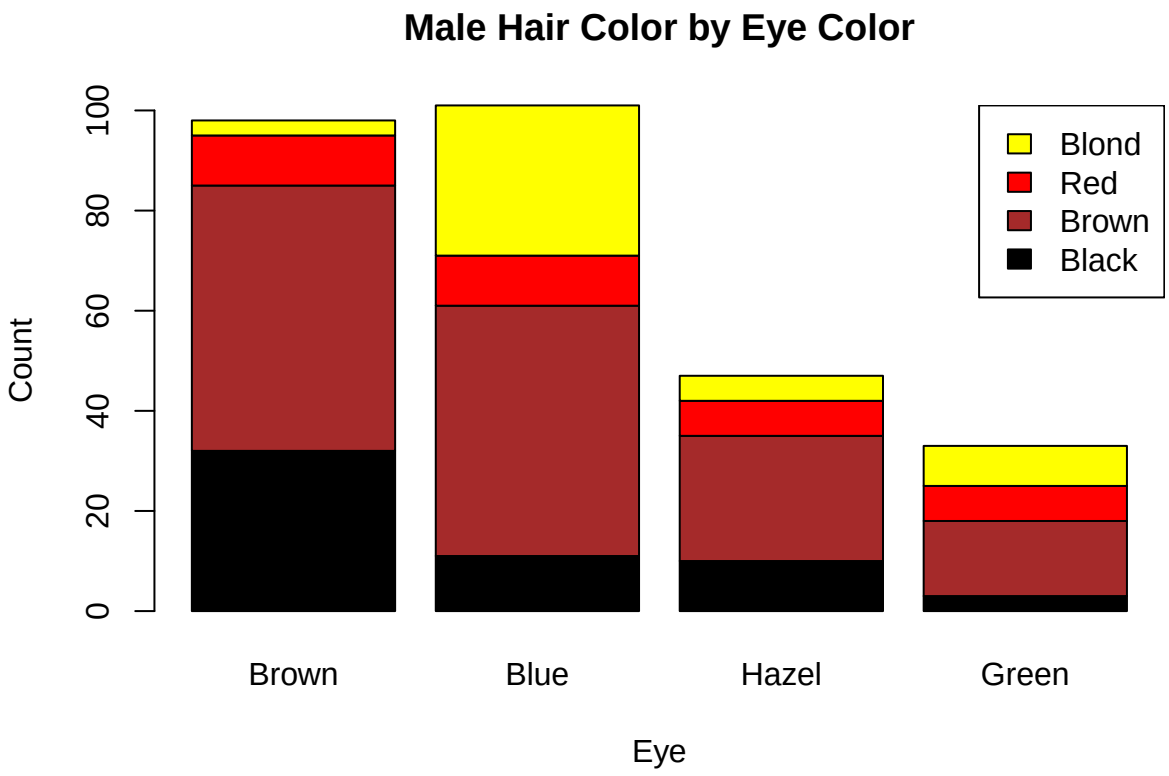
Bar plots of male/female hair color

```
par(mfrow=c(1,2))
barplot(HairEyeColor[,1,1], col = "cyan", main = "Males Hair Color",
        xlab = "Color", ylab = "Count")
barplot(HairEyeColor[,1,2], col = "yellow", main = "Female Hair Color",
        xlab = "Color", ylab = "Count")
```

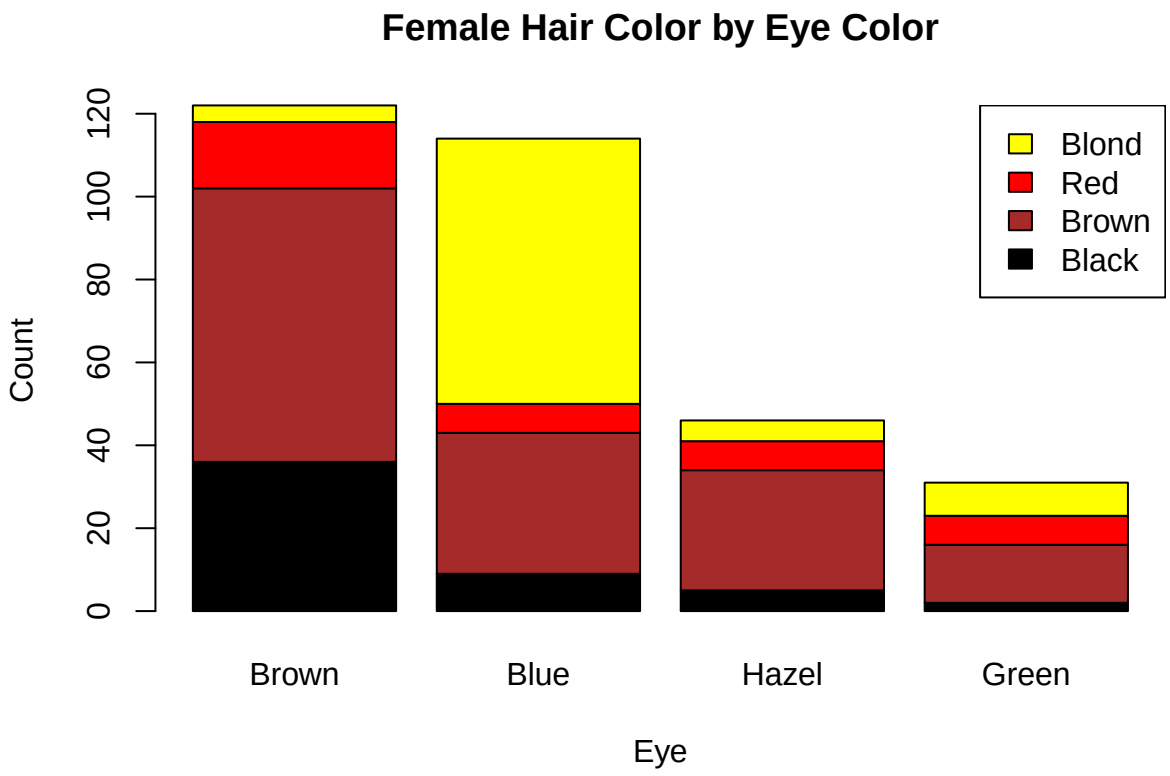


Bar plots of male/female hair color by eye color

```
par(mfrow=c(1,1))
barplot(HairEyeColor[,,1], col = c("black", "brown", "red", "yellow"),
        main = "Male Hair Color by Eye Color", xlab = "Eye",
        ylab = "Count", beside=FALSE)
legend(x="topright", legend=c("Blond", "Red", "Brown", "Black"),
       fill=c("yellow", "red", "brown", "black"))
```

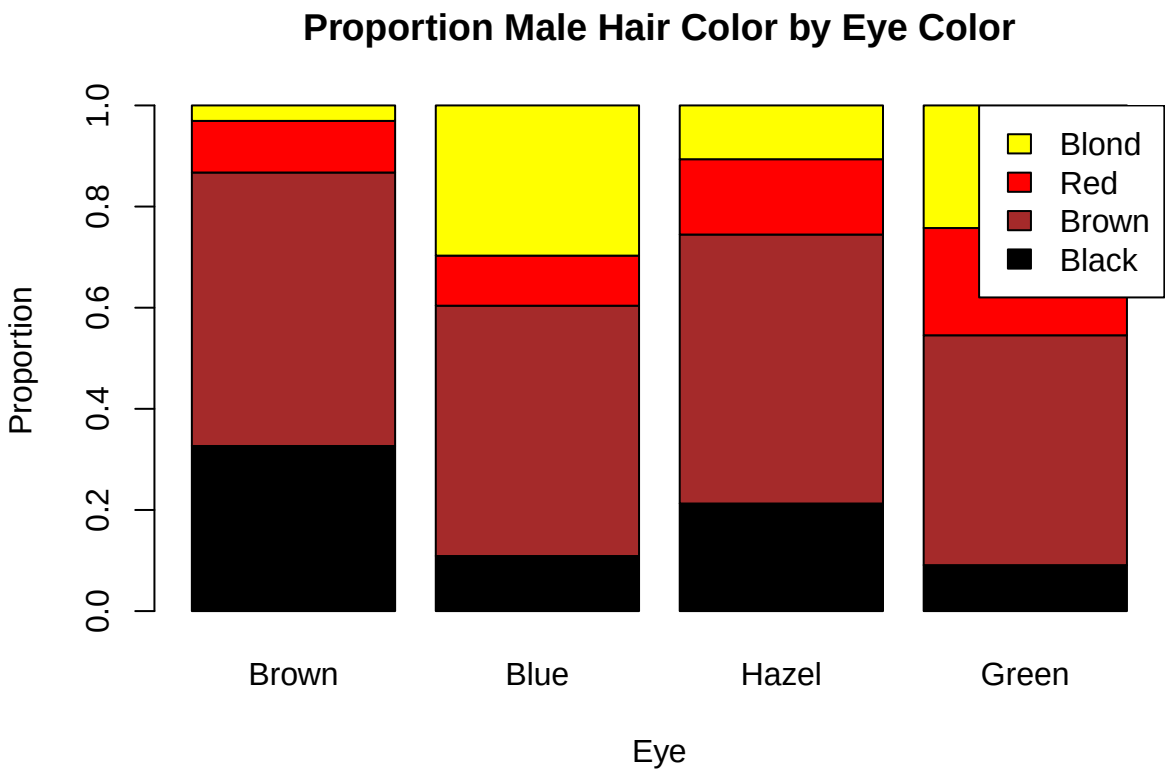



```
barplot(HairEyeColor[,,2], col = c("black", "brown", "red", "yellow"),
        main = "Female Hair Color by Eye Color", xlab = "Eye",
        ylab = "Count", beside = FALSE)
legend(x="topright", legend=c("Blond", "Red", "Brown", "Black"),
       fill=c("yellow", "red", "brown", "black"))
```

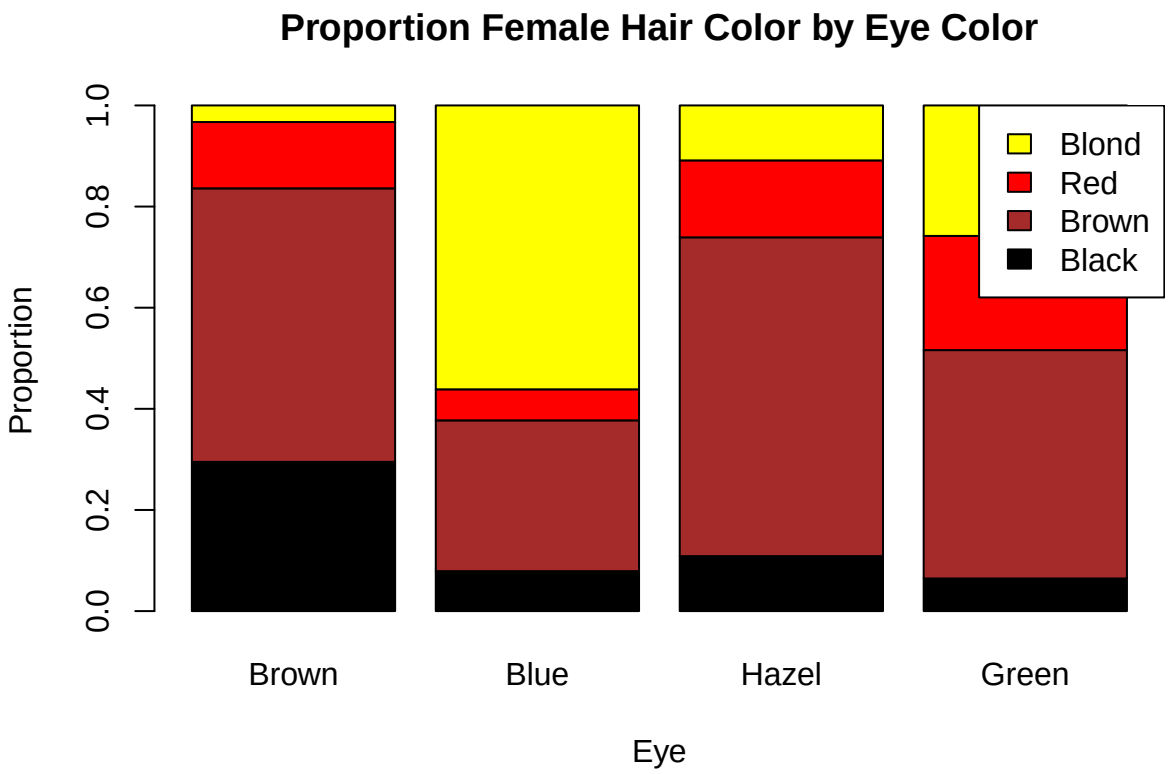


Bar plots of proportions of male/female hair color by eye color

```
barplot(prop.table(HairEyeColor[,,1], 2), col = c("black", "brown", "red",
                                                  "yellow"),
        main = "Proportion Male Hair Color by Eye Color", xlab = "Eye",
        ylab = "Proportion", beside=FALSE)
legend(x="topright", legend=c("Blond", "Red", "Brown", "Black"),
       fill=c("yellow", "red", "brown", "black"), bg = "white")
```




```
barplot(prop.table(HairEyeColor[,,2], 2), col = c("black", "brown", "red",
                                                  "yellow"),
        main = "Proportion Female Hair Color by Eye Color", xlab = "Eye",
        ylab = "Proportion", beside = FALSE)
legend(x="topright", legend=c("Blond", "Red", "Brown", "Black"),
       fill=c("yellow", "red", "brown", "black"), bg = "white")
```


Proportion Female Hair Color by Eye Color
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
Blond
Red
Brown
Black
Brown
Blue
Hazel
Green
Eye