Group work - week 4 - images

Stat 201: Statistics I

September 29, 2019

Question 1

(a) Create table

```
# Read data
ap <- read.csv("../../../Data/max_air_pol.csv")

# set appropriate cut points
ap.breaks <- (2:10)*5
ap$cls <- cut(ap$max.fpm, ap.breaks)

ap.tab <- table(ap$cls)
ap.tab <- cbind(Frequency=ap.tab, Cumulative=cumsum(ap.tab))
```

```
print(xtable(ap.tab),
      comment=FALSE)
```

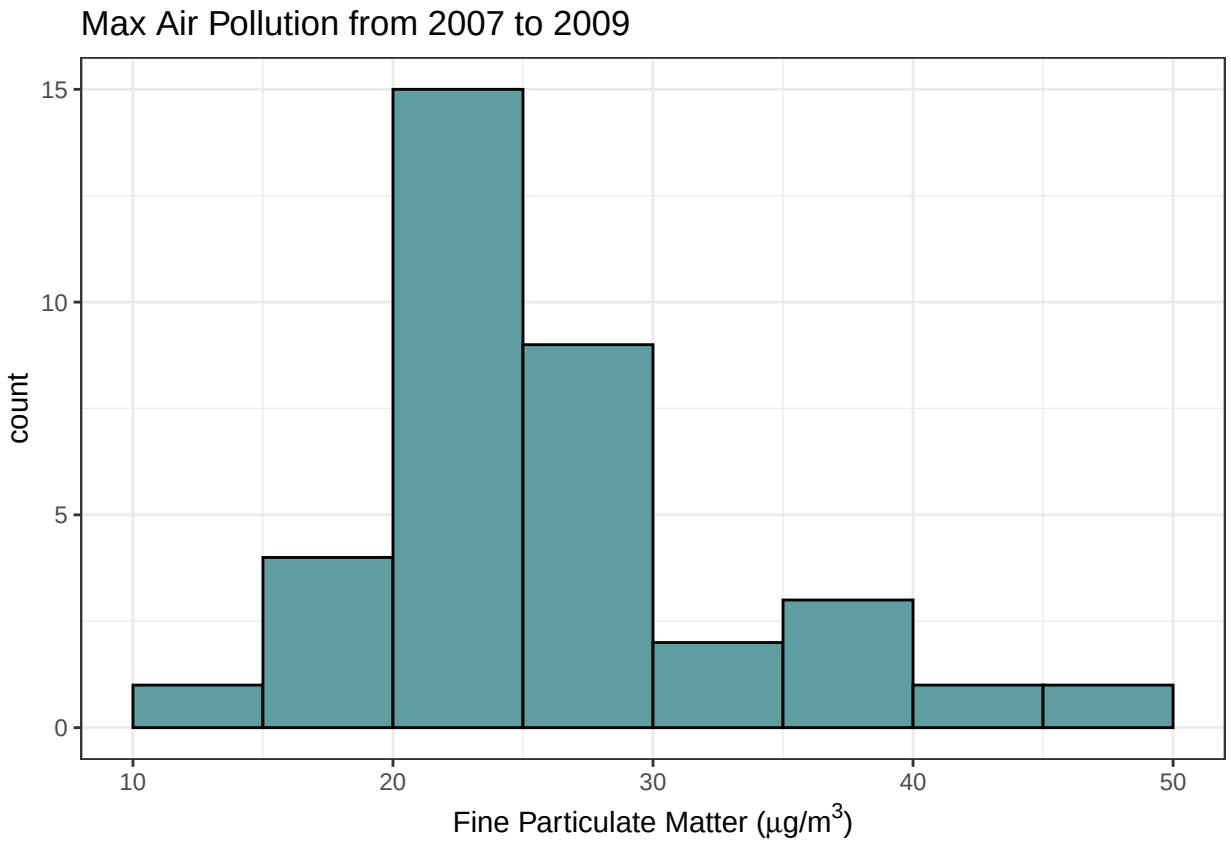
	Frequency	Cumulative
(10,15]	1	1
(15,20]	4	5
(20,25]	15	20
(25,30]	9	29
(30,35]	2	31
(35,40]	3	34
(40,45]	1	35
(45,50]	1	36

```
print(xtable(ap.tab),
      comment=FALSE)
```

```
\begin{table}[ht]
\centering
\begin{tabular}{rrr}
  \hline
 & Frequency & Cumulative \\
  \hline
(10,15] &   1 &   1 \\
  (15,20] &   4 &   5 \\
  (20,25] &  15 &  20 \\
  (25,30] &   9 &  29 \\
  (30,35] &   2 &  31 \\
  (35,40] &   3 &  34 \\
  (40,45] &   1 &  35 \\
  (45,50] &   1 &  36 \\
   \hline
\end{tabular}
\end{table}
```

(b) Create histogram

```
g <- ggplot(ap, aes(x=max.fpm))
g <- g + geom_histogram(breaks=ap.breaks, fill="cadetblue", col="black")
g <- g + theme_bw() + labs(x=expression(paste("Fine Particulate Matter (", mu, "g/", m^3,")")),
                           title="Max Air Pollution from 2007 to 2009")
g
```




```
ggsave('../images/group04_Q2_b.png', width=4, height=2.5, units = "in")
```

Question3

(a) Find measures of center

```
# Given a vector of numeric data, return vector of modes
find.mode <- function(x){
    x.tab <- table(x)
    x.modes <- names(x.tab[x.tab==max(x.tab)])
    if(length(x.modes)==length(x.tab)){
        x.modes <- c()
    }

    return(x.modes)
}
```

```r
hs <- read.csv('../../../Data/mpls_home_sales.csv')

# Find measures of center
hs.ctr <- data.frame(Mean=mean(hs$sale.price),
                     Median=median(hs$sale.price),
                     Mode=paste(find.mode(hs$sale.price), collapse=", "),
                     Midrange=mean(range(hs$sale.price)))

xt <- xtable(hs.ctr, align="ccccc")
```

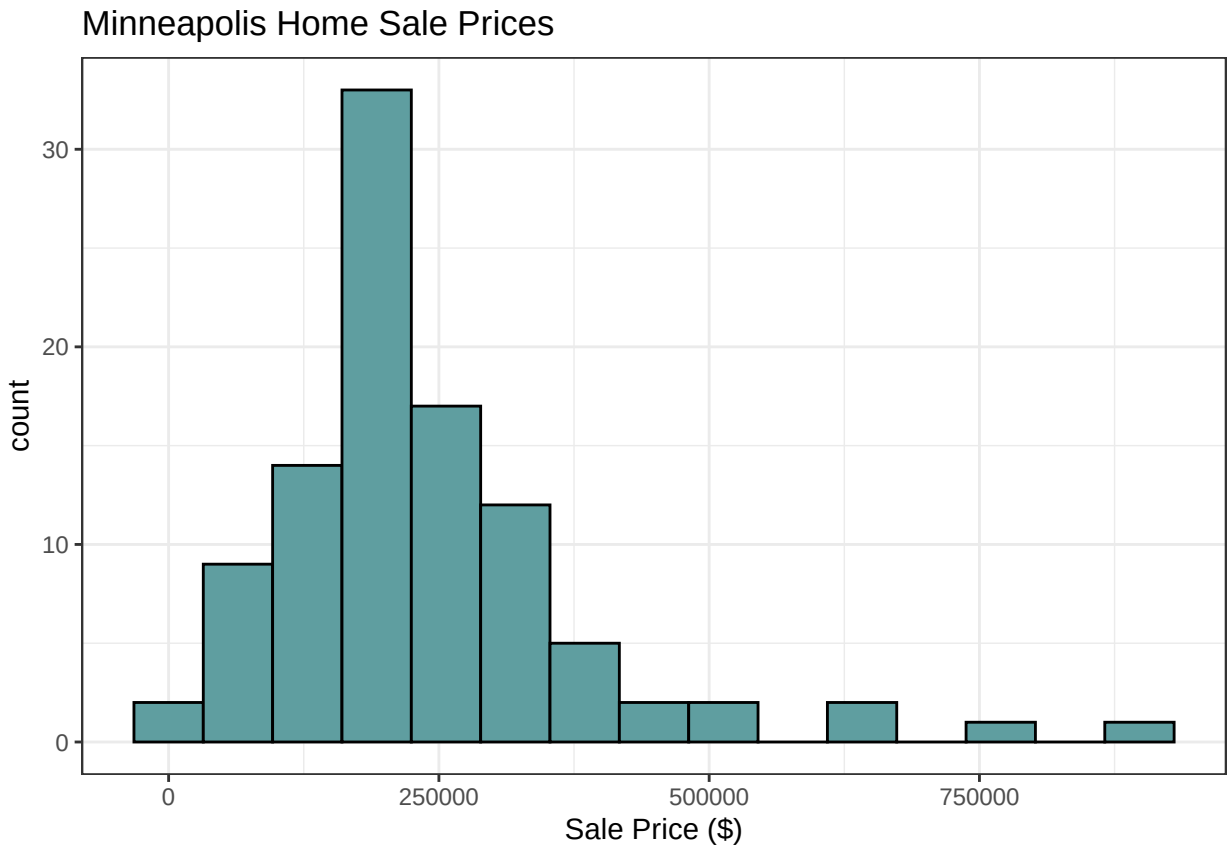
```r
print(xt,
      include.rownames=FALSE,
      comment=FALSE)
```

Mean	Median	Mode	Midrange
234980.59	202750.00	175000	451000.00

```r
print(xt,
      include.rownames=FALSE,
      comment=FALSE)
```

```
\begin{table}[ht]
\centering
\begin{tabular}{cccc}
  \hline
Mean & Median & Mode & Midrange \\
  \hline
234980.59 & 202750.00 & 175000 & 451000.00 \\
   \hline
\end{tabular}
\end{table}
```

```r
# Create histogram
g <- ggplot(hs, aes(x=sale.price))
g <- g + geom_histogram(fill="cadetblue", col="black", bins=15)
g <- g + theme_bw() + labs(x="Sale Price ($)", title="Minneapolis Home Sale Prices")
g
```




```
ggsave('../images/group04_Q2_a.png', width=4, height=2.5, units = "in")
```

(b) Find measures of variance

```
hs.var <- data.frame(Range=diff(range(hs$sale.price)),
                     Variance=var(hs$sale.price),
                     SD=sd(hs$sale.price))

xt <- xtable(hs.var, align="cccc")
```

```
print(xt,
      include.rownames=FALSE,
      comment=FALSE)
```

Range	Variance	SD
898000	21434941764.24	146406.77

```
print(xt,
      include.rownames=FALSE,
      comment=FALSE)
```

```
\begin{table}[ht]
\centering
\begin{tabular}{ccc}
  \hline
Range & Variance & SD \\
```

```
  \hline
898000 & 21434941764.24 & 146406.77 \\
   \hline
\end{tabular}
\end{table}
```

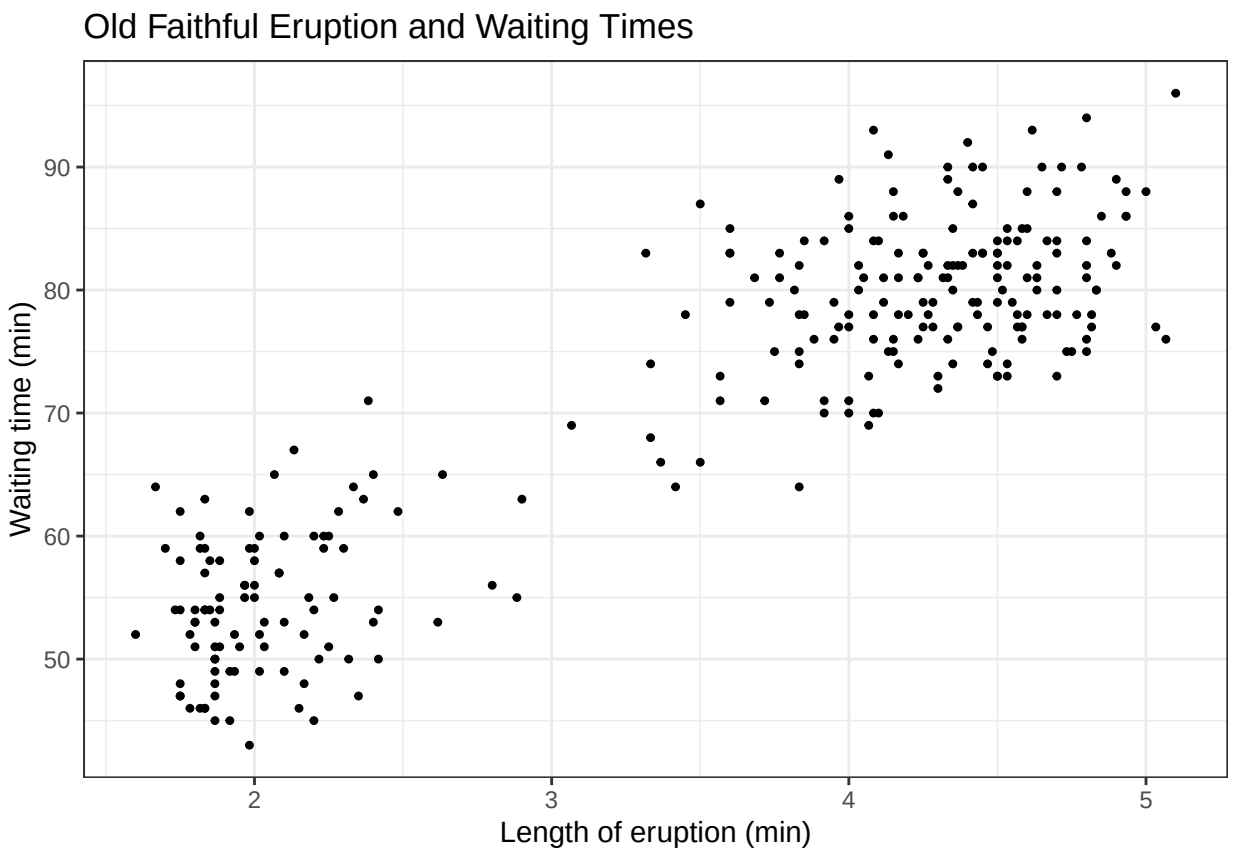
Quastion 4

(a) Faithful

```
faith <- read.csv('../../../Data/faithful.csv')

# Create scatterplot
g <- ggplot(faith, aes(x=eruptions, y=waiting))
g <- g + geom_point(shape=20)
g <- g + theme_bw() + labs(x="Length of eruption (min)",
                           y="Waiting time (min)",
                           title="Old Faithful Eruption and Waiting Times")

g
```



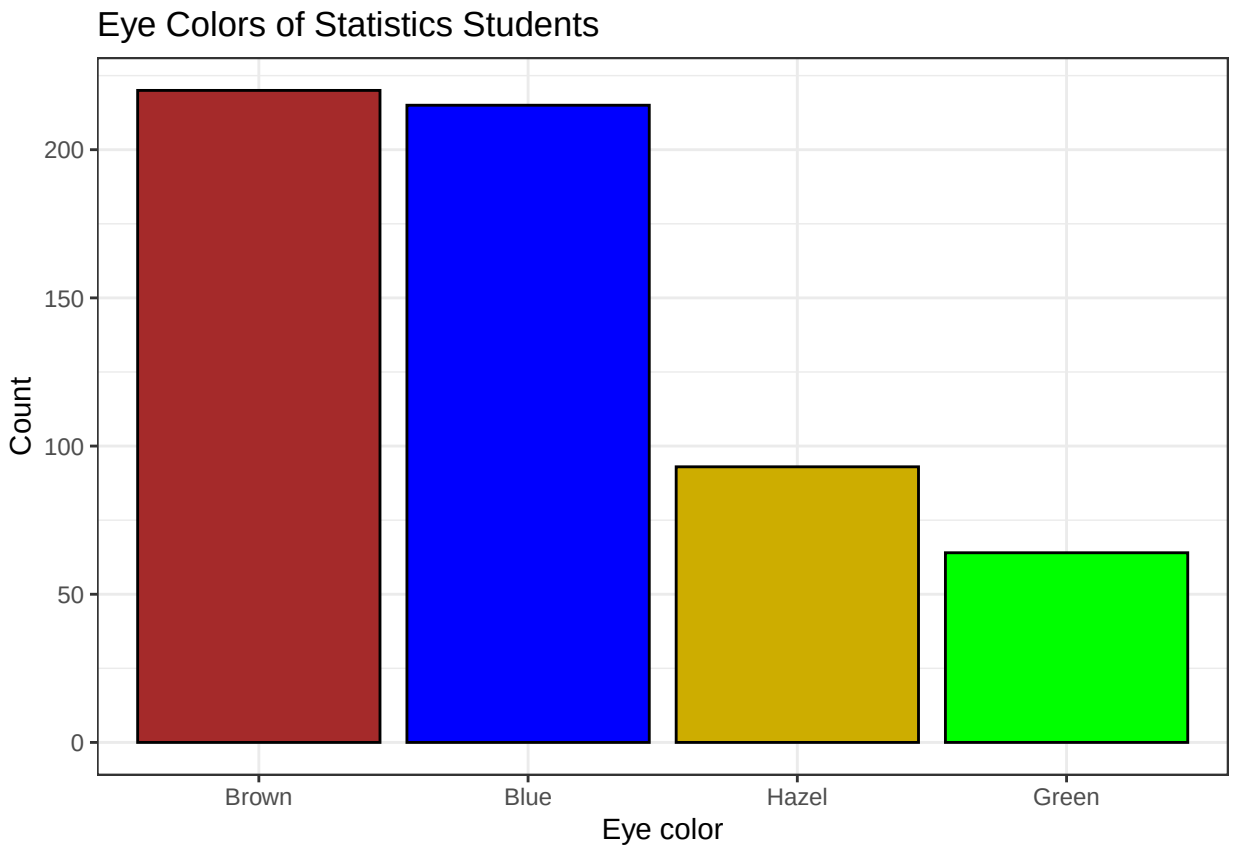

```
ggsave('../images/group04_Q4_a.png', width=4, height=2.5, units = "in")
```

(b) Eye color

```
eye <- read.csv('../../../Data/hair_eye.csv')

eye.tab <- sort(table(eye$Eye), decreasing = T)
eye.df <- data.frame(eye.col=factor(rownames(eye.tab), levels=rownames(eye.tab)), cnt=as.vector(eye.tab)

g <- ggplot(eye.df, aes(x=eye.col, y=cnt))
g <- g + geom_bar(stat="identity", fill=c("brown", "blue", "gold3", "green"), col="black")
g <- g + theme_bw() + labs(x="Eye color", y="Count",
                           title="Eye Colors of Statistics Students")
g
```




```
ggsave('../images/group04_Q4_b.png', width=4, height=2.5, units = "in")
```