Stats 10 Lab 1 Submission

Name: Brandon Truong

Section 1

1) a)

```
> names <- c("Brandon", "Samuel", "Bryan")
> print(names)
[1] "Brandon" "Samuel"  "Bryan"
```

b)

```
> heights <- c(68,65,72) #Height in inches
> print(heights)
[1] 68 65 72
```

c)

```
> cbind(heights,names)
     heights names
[1,] "68"    "Brandon"
[2,] "65"    "Samuel"
[3,] "72"    "Bryan"
```

The command cbind allowed us to bind two vectors of the same size into a matrix
with two columns: heights and names. We can determine the class by using
```
        > class(cbind(heights,names))
        [1] "matrix"
```
which gives us "matrix."

2) a)

```
> setwd("~/UCLA Coursework/STATS 10") #Directory containing births.csv

> NCbirths <- read.csv(file="births.csv")
```

b)

```
> head(NCbirths)
  Gender Premie weight Apgar1 Fage Mage Feduc Meduc TotPreg
1   Male     No    124      8   31   25    13    14       1
2 Female     No    177      8   36   26     9    12       2
3   Male     No    107      3   30   16    12     8       2
4 Female     No    144      6   33   37    12    14       2
```

```
5   Male     No     117     9    36    33    10    16        2
6 Female     No      98     4    31    29    14    16        3
  Visits    Marital Racemom Racedad Hispmom Hispdad Gained
1     13    Married   White   White NotHisp NotHisp     40
2     11  Unmarried   White   White Mexican Mexican     20
3     10  Unmarried   White Unknown Mexican Unknown     70
4     12  Unmarried   White   White NotHisp NotHisp     50
5     19    Married   White   Black NotHisp NotHisp     40
6     20    Married   White   White NotHisp NotHisp     21
      Habit MomPriorCond BirthDef    DelivComp BirthComp
1 NonSmoker         None     None At Least One      None
2 NonSmoker         None     None At Least One      None
3 NonSmoker At Least One     None At Least One      None
4 NonSmoker         None     None At Least One      None
5 NonSmoker At Least One     None         None      None
6 NonSmoker         None     None         None      None
```

3) a)

```
> install.packages("maps") #package installation
> find.package("maps")
[1] "C:/Users/btruo/OneDrive/Documents/R/win-library/3.6/maps"
```

b)

```
> library(maps) #package loading
> map("state")
```


4) a)

```
> weights<-NCbirths$weight
> head(weights) #Weight sample
[1] 124 177 107 144 117  98
```

b)

The weight unit is in ounces (oz)

c)

```
> weights_in_pounds<-weights/16
```

d)

```
> weights_in_pounds[1:20]
 [1]  7.7500 11.0625  6.6875  9.0000  7.3125  6.1250  9.1875  8.6250
 [9]  6.5000  7.6875  9.5625  8.0625  7.4375  6.7500  6.6250  7.8125
[17]  7.1875  8.0000  8.2500  5.1875
```

Section 2

1)

```
> mean(NCbirths$weight)
[1] 116.0512
```

2)

```
> tally(NCbirths$Habit, format = c("percent"))
X
NonSmoker    Smoker
 90.61245   9.38755
```
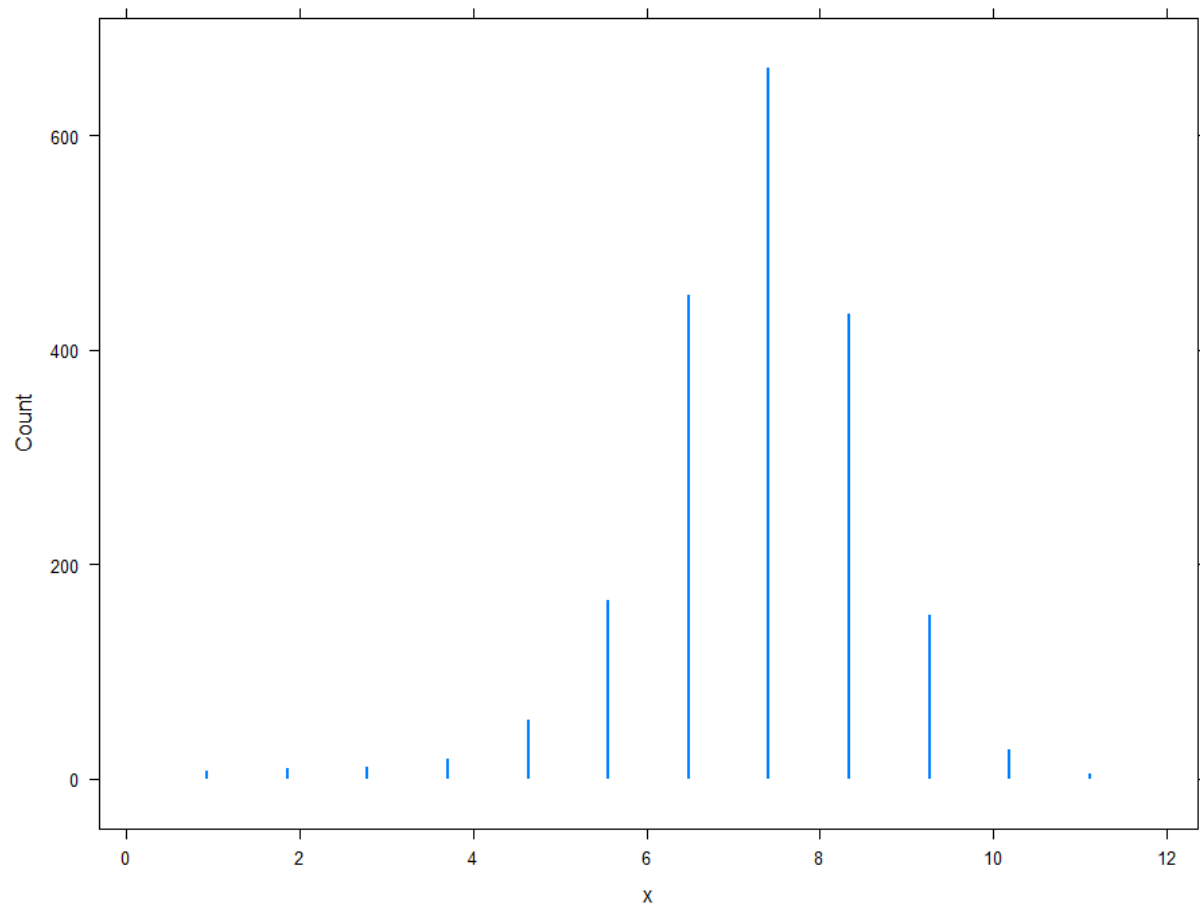
9.38755% of mothers smoke.

3)

21%-9.38755%=11.61245%

Our data is 11.61245% off from the CDC data.
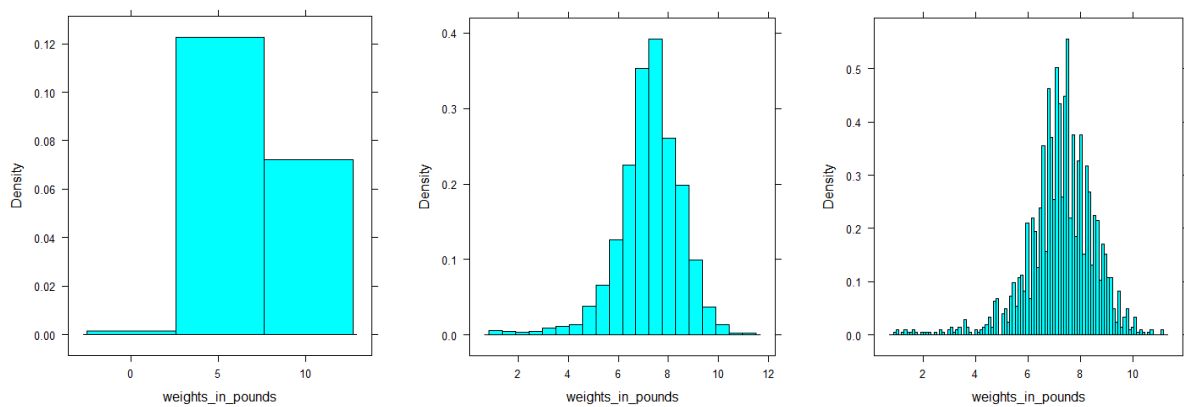
Section 3

1)

```
> dotPlot(weights_in_pounds)
```

2)

```
> histogram(weights_in_pounds, nint=3)

> histogram(weights_in_pounds, nint=20)

> histogram(weights_in_pounds, nint=100)
```

The histogram with 20 bins give the best visualization as it is neither too
clustered that we can't see trends nor is the data too generalized. 20 bins
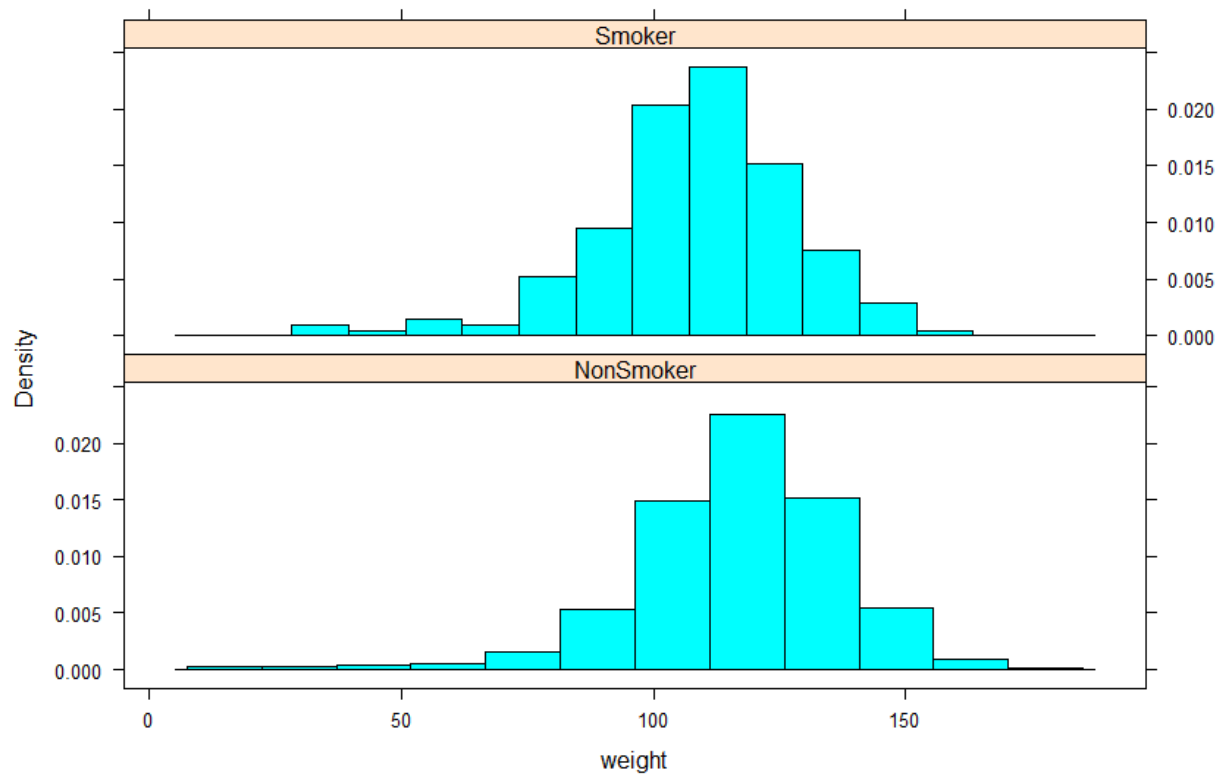allow us to see distribution and trends within the data.

3)

```
> boxplot(NCbirths$Fage, NCbirths$Mage)
```
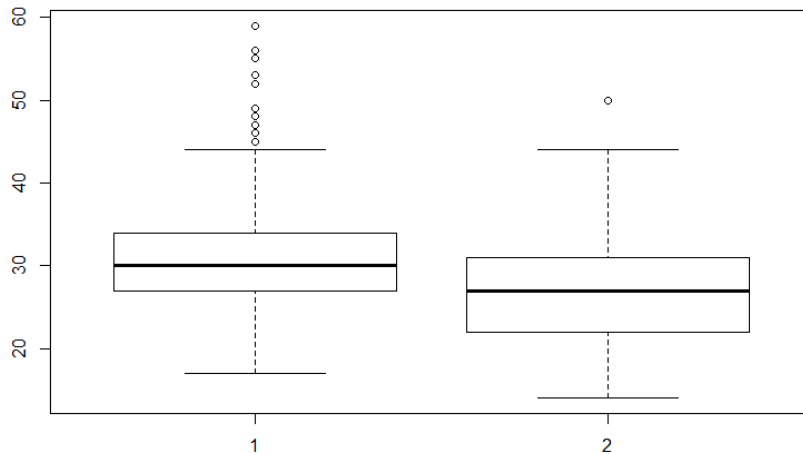

Fathers tend to be older in the dataset.
4)

```
> histogram(~weight|Habit,data=NCbirths,layout=c(1,2))
```

The code produces two histograms based on Weight and Smoking Habits, overlayed on top of each other. Smoking mothers result in lower and more variable baby weights in comparison to non-smoking mothers.

Section 4

1)

I hypothesize that Premie and BirthDef are related to smoking habits. This is because smoking results in the inhalation of nicotine, carbon monoxide, and other poisons, which is known to be related to health problems. We can test this by using

```
> tally(~Premie | Habit, data = NCbirths, format = "proportion")

> tally(~BirthDef | Habit, data = NCbirths, format = "proportion")
```

which gives us these two two-way tables


```
       Habit

Premie  NonSmoker      Smoker

   No  0.91191136 0.88235294

   Yes 0.08808864 0.11764706
```

```
          Habit

BirthDef         NonSmoker        Smoker

  At Least One 0.006648199 0.016042781

  None         0.993351801 0.983957219
```

We can see that there is a correlation with smoking habits and both BirthDef and Premie, showing an increase in proportion if the mother smokes. The proportions of at least one birth defects increase from .006 from nonsmokers to .016 from smokers, and the proportion of premature birth from .08 to .12 subsequently. This increase in Premie and BirthDef proportions in smokers prove this correlation.

Section 5

1)

```
> plot(NCbirths$weight ~ NCbirths$Mage, col = "blue", cex = 1.5, pch = 3, xlab=
"Mother's Age (years)", ylab = "Baby weight (oz.)", main = "Baby Weight vs.
Mother's Age")
```
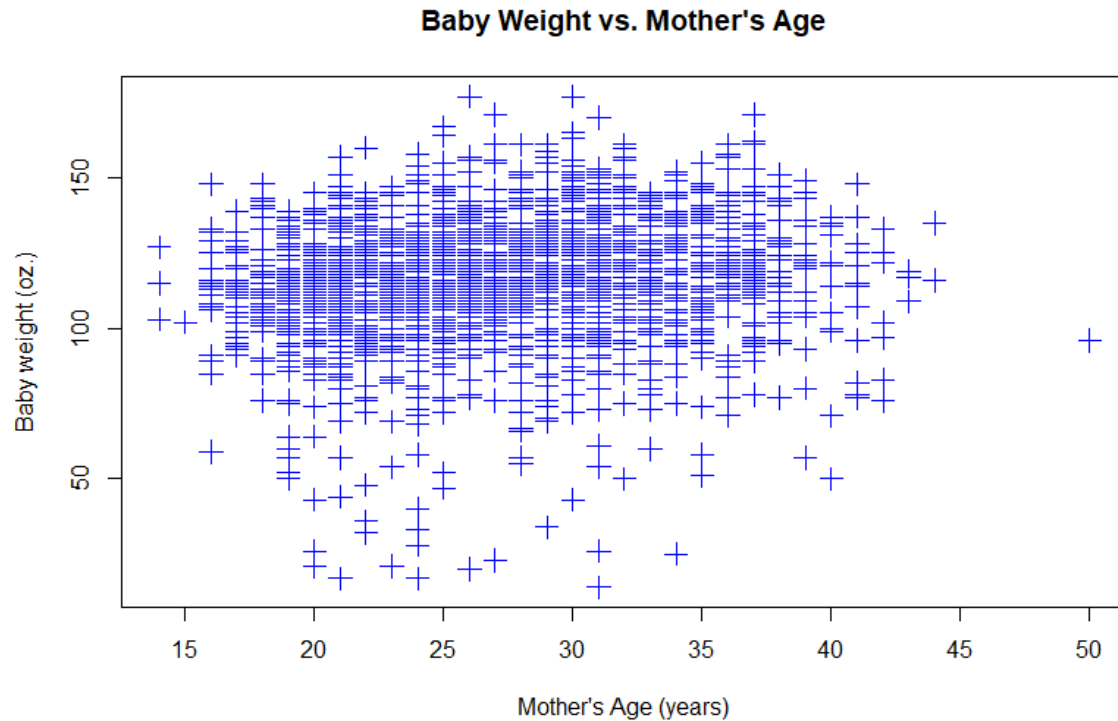

Section 6

1)

```
> a <- read.table(file = "ozone.txt", header=TRUE)

> AQI_colors<-c("pink","blue","orange","cyan","mahogany")

> AQI_levels<-cut(a$o3, c(0,0.06,0.075,0.104,0.115,0.374))

> as.numeric(AQI_levels)

  [1] 1 3 1 3 1 3 1 1 1 1 3 2 1 1 1 3 2 3 3 1 2 2 3 1 3 1 1 1 1 1 1 1 1 4 1 3 3
 33 2 3 3 3 1 1 1 3 1 1

 [50] 1 3 1 1 1 3 2 1 3 1 1 1 4 1 3 3 3 1 3 3 1 3 2 1 2 3 1 1 3 3 1 2 3 2 2 2 2
 1 2 1 1 3 2 2 1 1 1 2 2

 [99] 1 4 1 3 1 1 3 2 1 3 1 3 3 3 1 1 3 2 2 3 2 1 3 2 1 2 1 3 3 2 3 3 2 1 3 3 2
 3 3 1 1 2 3 4 3 1 1 2 2

[148] 2 3 2 1 2 2 1 3 1 1 1 1 3 3 4 3 1 4 1 1 1 1 1 3 1 1 1 1

>

> library(maps)

> plot(a$x,a$y, xlim=c(-125,-114),ylim=c(32,43), xlab="longitude",
ylab="latitude", main="California ozone bubble plot", "n")

> map("county", "ca",add=TRUE)

> points(a$x,a$y, cex=a$o3/mean(a$o3), col=AQI_colors[as.numeric(AQI_levels)],
pch=2)
```

California ozone bubble plot