Stats 12 Lab 1 Submission
Name: Brian Tehrani
UID: 604715464

Section 1

1)

a)

```
> heights <- c(71, 67, 66)
> print(heights)
[1] 71 67 66
```

b)

```
> names <- c("Brian", "Daisy", "Kim")
> print(names)
[1] "Brian" "Daisy" "Kim"
```

c)

```
> cbind(heights, names)
     heights names
[1,] "71"    "Brian"
[2,] "67"    "Daisy"
[3,] "66"    "Kim"
```

The command combined the vectors “heights” and “names” into a matrix with the first column being heights and the second entry being the names of the people of study. The class of this new object is a “matrix” (found using class(cbind(heights, names))).

2)

a) `> NCBirths <- read.csv(file.choose())`

b)

```
> head(NCBirths)
  Gender Premie weight Apgar1 Fage Mage Feduc Meduc TotPreg Visits   Marital Racemom
1   Male     No    124      8   31   25    13    14       1     13   Married   White
2 Female     No    177      8   36   26     9    12       2     11 Unmarried   White
3   Male     No    107      3   30   16    12     8       2     10 Unmarried   White
4 Female     No    144      6   33   37    12    14       2     12 Unmarried   White
5   Male     No    117      9   36   33    10    16       2     19   Married   White
6 Female     No     98      4   31   29    14    16       3     20   Married   White
  Racedad Hispmom Hispdad Gained     Habit MomPriorCond BirthDef    DelivComp BirthComp
1   White NotHisp NotHisp     40 NonSmoker          None     None At Least One      None
2   White Mexican Mexican     20 NonSmoker          None     None At Least One      None
3 Unknown Mexican Unknown     70 NonSmoker At Least One     None At Least One      None
4   White NotHisp NotHisp     50 NonSmoker          None     None At Least One      None
5   Black NotHisp NotHisp     40 NonSmoker At Least One     None          None      None
6   White NotHisp NotHisp     21 NonSmoker          None     None          None      None
```

3)

a) > install.packages("maps")

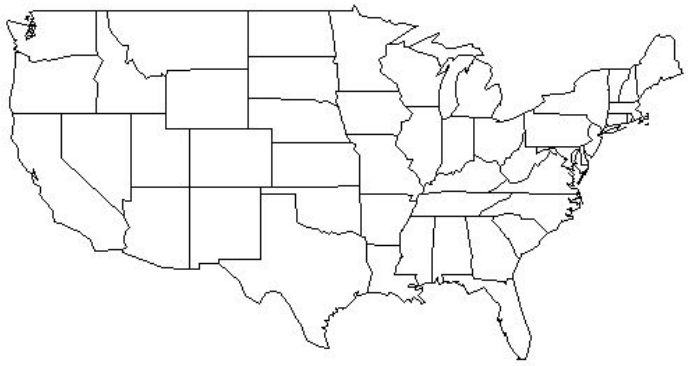
```
Installing package into ‘C:/Users/brian/Documents/R/win-library/3.5’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/maps_3.3.0.zip'
Content type 'application/zip' length 3694253 bytes (3.5 MB)
downloaded 3.5 MB

package ‘maps’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\brian\AppData\Local\Temp\Rtmp6ZtqJG\downloaded_packages
```

```
> find.package("maps")
[1] "C:/Users/brian/Documents/R/win-library/3.5/maps"
```

b)
```
> library(maps)
> map("state")
```



4)

a)
```
> weights <- NCBirths$weight
```

b) The unit of weights are in ounces.

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  6.5000
[10]  7.6875  9.5625  8.0625  7.4375  6.7500  6.6250  7.8125  7.1875  8.0000
[19]  8.2500  5.1875
```

Section 2

1) > mean(weights.in.pounds)
   [1] 7.2532

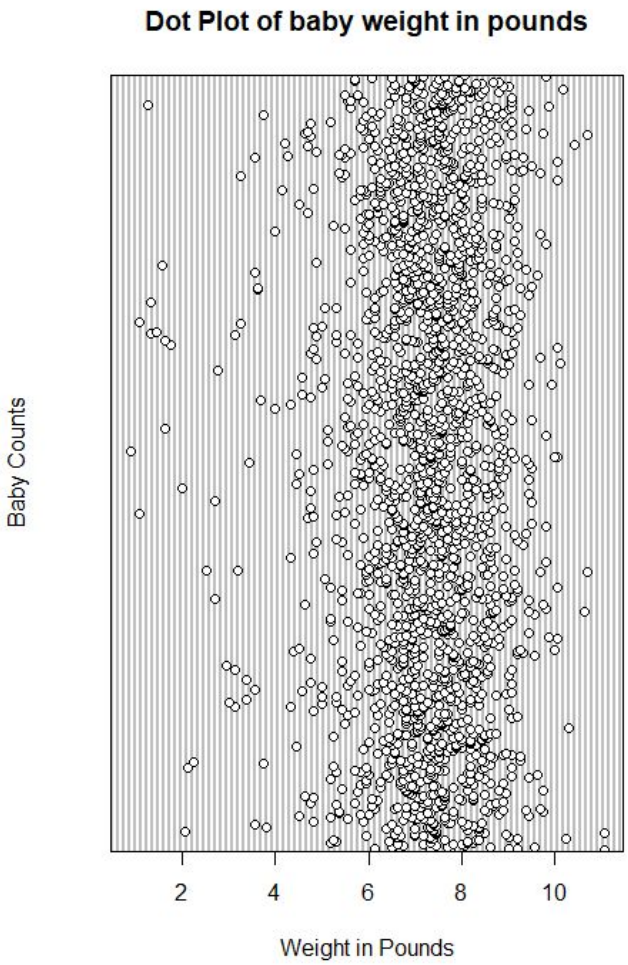
2) > tally(NCBirths$Habit, "percent")
   X
   NonSmoker Smoker
   90.61245 9.38755

   The percentage of mothers in the sample who smoke are 9.38755%.

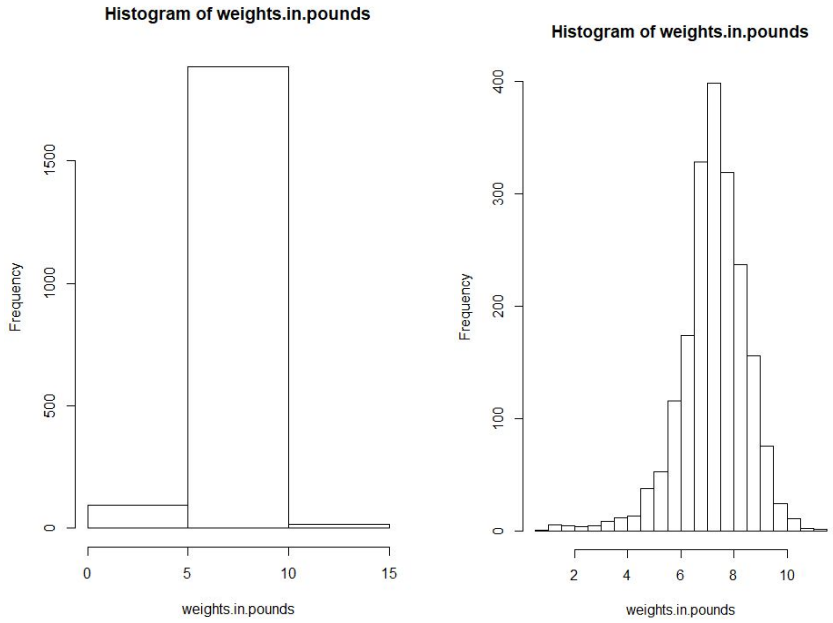
3) (21% - 9.38755%)/ 21% = 55.3% error factor, off by 11.61245%. 21% > 9.38755%.

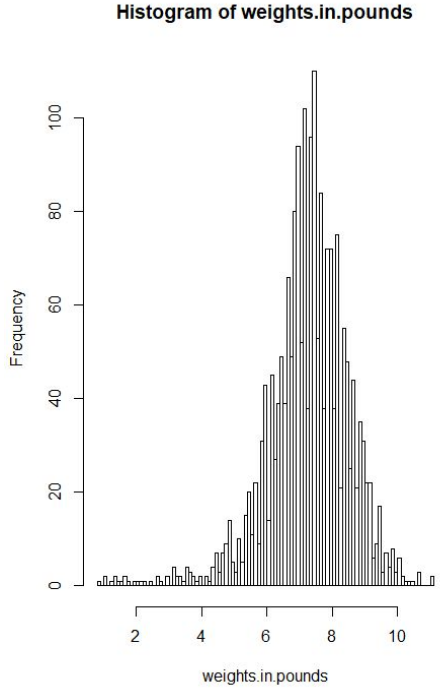
Section 3

1) > dotchart(weights.in.pounds, xlab = "Weight in Pounds", ylab = "Baby Counts",
   + main = "Dot Plot of baby weight in pounds")



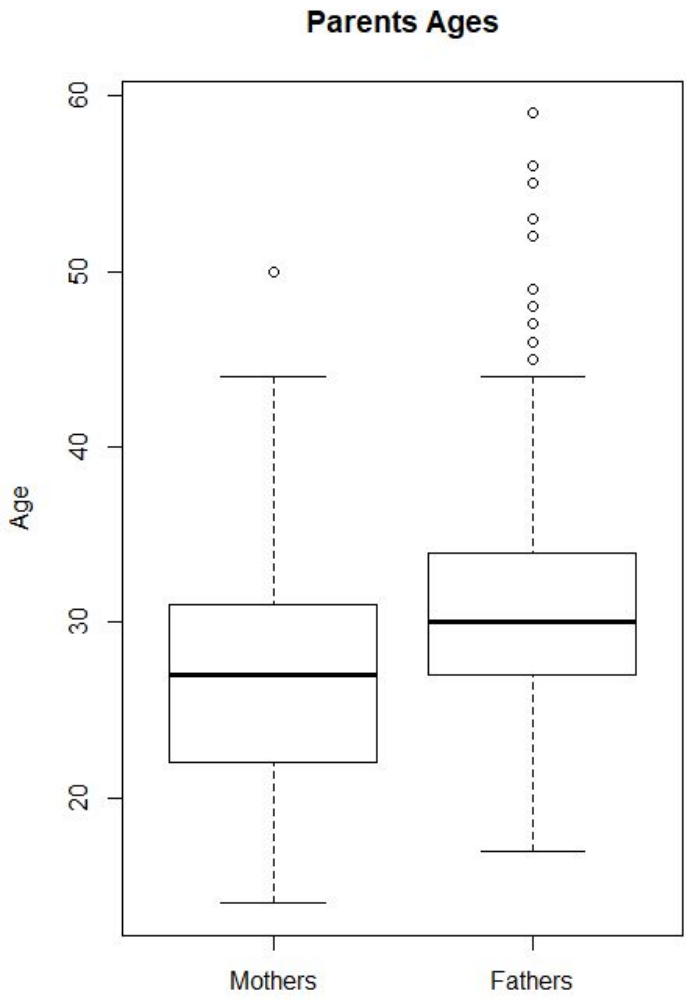
2) > hist(weights.in.pounds, breaks = 3)
> hist(weights.in.pounds, breaks = 20)
> hist(weights.in.pounds, breaks = 100)





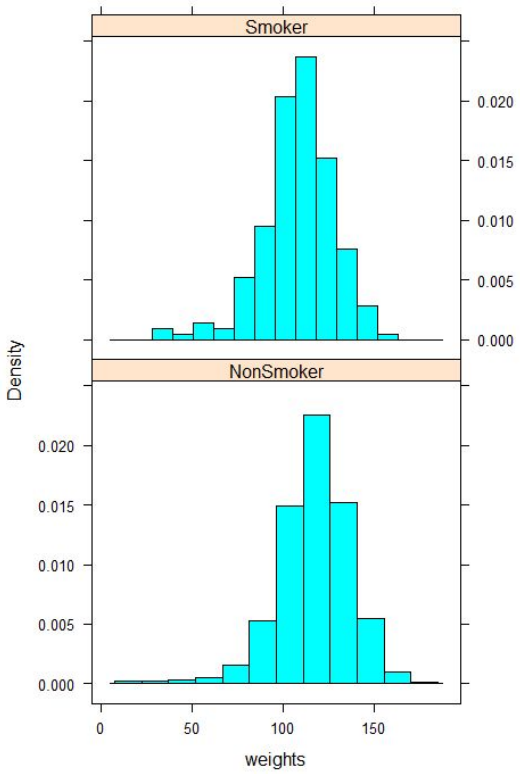
The histogram that gives the best visualization of the data is the histogram with the bins of 20 because the bars can be easily read as well as depicting a proper overall look at the distribution of data.

3)
```
> boxplot(NCBirths$Mage, NCBirths$Fage, main = "Parents Ages",
+         names = c("Mothers", "Fathers"), ylab = "Age")
```



The gender that tends to be older are the males as their median age is above the female median age. In addition most of the data points for males are above females.

4) > histogram(~weights|Habit, data = NCBirths, layout = c(1,2))



The code creates two different colored histograms that show the distributions of the baby weights between mothers who smoke on the top and the mothers who don't smoke on the bottom. The weights are shown in ounces. The differences are small, but noticeable in that the weights of the babies from the mothers who smoke are less than the babies from the mothers who don't smoke.

Section 4

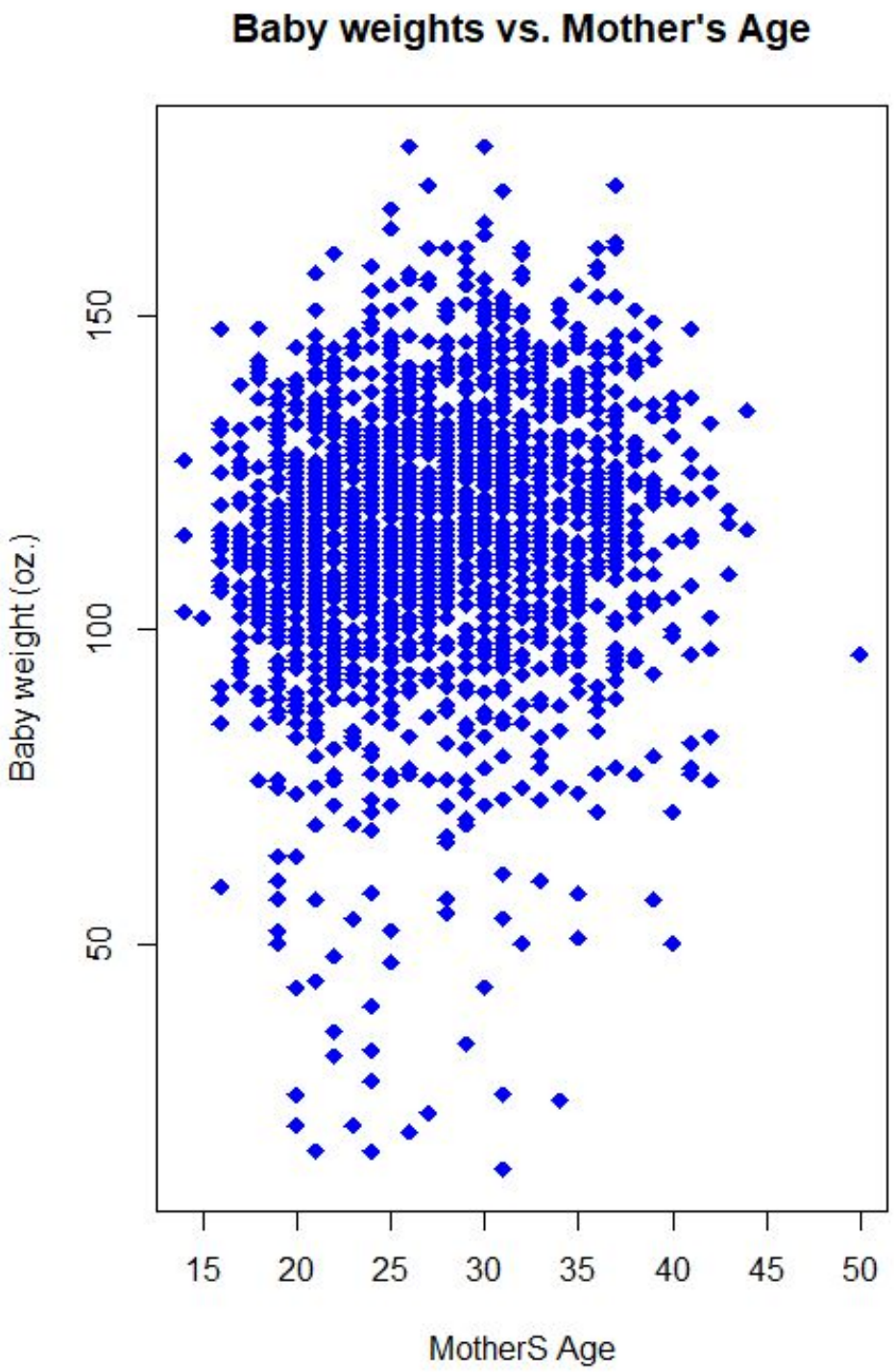
1) > tally(~Habit | DelivComp, data = NCBirths, format = "proportion")

Habit	At Least One	None
NonSmoker	0.8920188	0.9127864
Smoker	0.1079812	0.0872136

The hypothesis I had regarding the health of the baby to the mother's smoking is that the Habit of the mother is dependent on the delivery complications the mother may receive at birth and thus would harm the baby. The analysis of the comparison of the relative frequencies between the two categorical variables of Habit and Delivery Complications refutes this hypothesis as nonsmoking mothers show that they have at least one delivery complication or no complications with the birth of the baby.

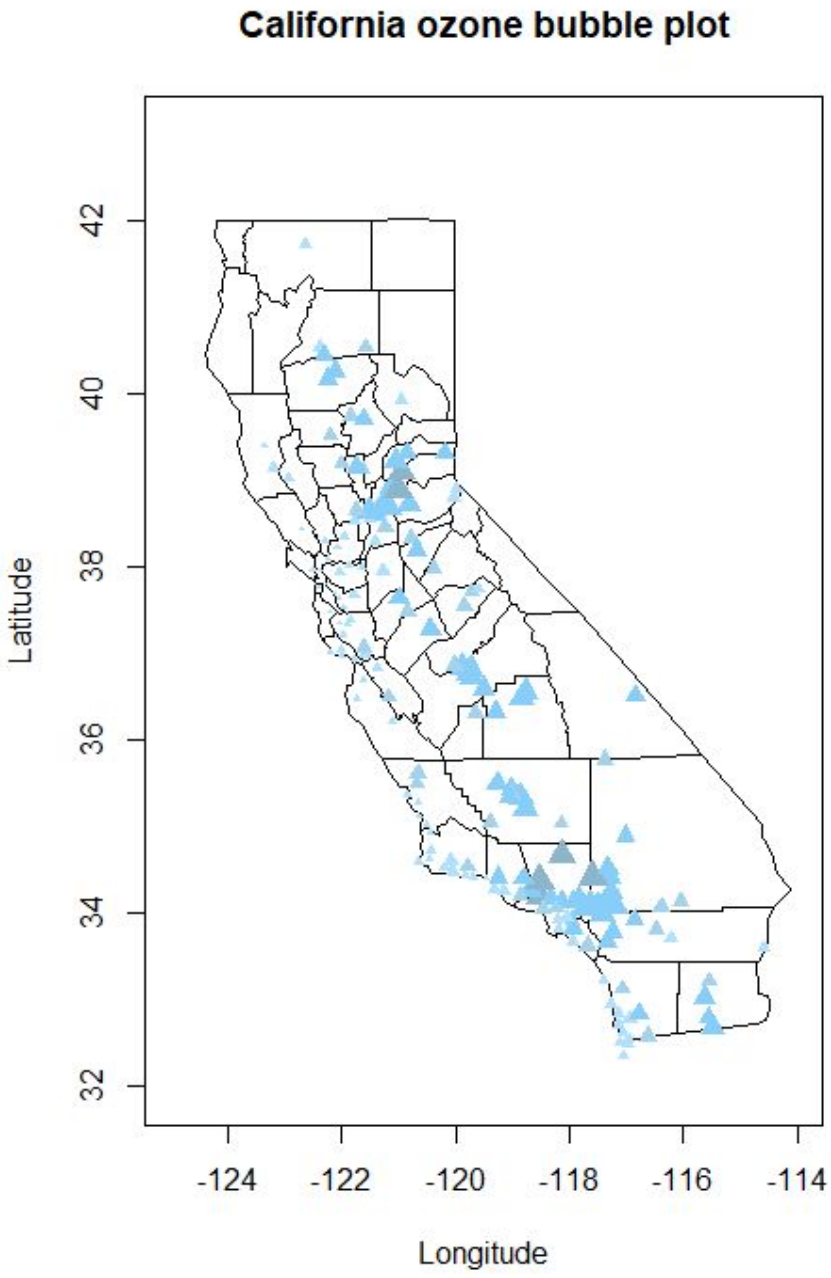
Section 5

1)
```
> plot(NCBirths$weight ~ NCBirths$Mage, col = "blue", ylab = "Baby weight (oz.)",
+      xlab = "MotherS Age", main = "Baby weights vs. Mother's Age", pch =18,
+      cex =1.25)
```



Section 6

1)
```
> a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics12/ozone.txt",
header=TRUE)
> library(maps)
>
> AQI_colors <- c("lightskyblue1","lightskyblue2",
"lightskyblue","lightskyblue3","lightskyblue4")
> AQI_levels <- cut(a$o3, c(0, 0.06, 0.075, 0.104, 0.115, 0.374))
>
> as.numeric(AQI_levels)
  [1] 1 3 1 3 1 3 1 1 1 1 3 2 1 1 1 3 2 3 3 1 2 2 3 1 3 1 1 1 1 1 1 1 4 1 3 3 3 3 2 3 3
 [43] 3 1 1 1 3 1 1 1 3 1 1 1 3 2 1 3 1 1 1 4 1 3 3 3 1 3 3 1 3 2 1 2 3 1 1 3 3 1 2 3 2 2
 [85] 2 2 1 2 1 1 3 2 2 1 1 1 2 2 1 4 1 3 1 1 3 2 1 3 1 3 3 3 1 1 3 2 2 3 2 1 3 2 1 2 1 3
[127] 3 2 3 3 2 1 3 3 2 3 3 1 1 2 3 4 3 1 1 2 2 2 3 2 1 2 2 1 3 1 1 1 1 3 3 4 3 1 4 1 1 1
[169] 1 1 3 1 1 1 1
>
> plot(a$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=17)
```


California ozone bubble plot
Latitude
42
40
38
36
34
32
-124
-122
-120
-118
-116
-114
Longitude