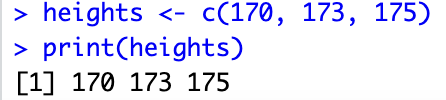
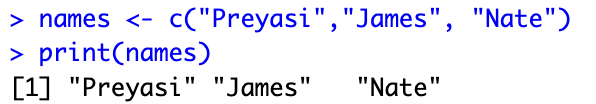
Stat 10 Lab Submission  
Name: Preyasi Gaur (UID: 705704939)  
Section: 4A

**Section 1**

1. Vectors
   1. > heights <- c(170, 173, 175)

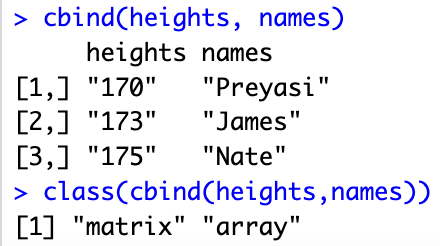
> print(heights)  


* 1. > names <- c("Preyasi","James", "Nate")

> print(names)  


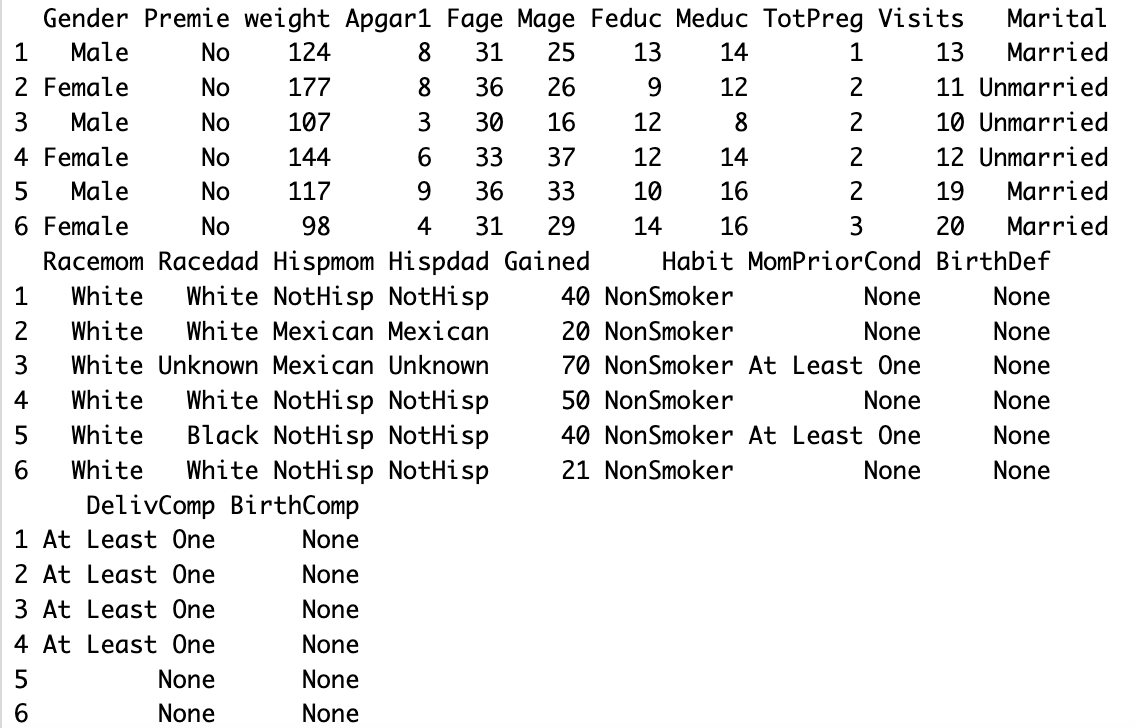
* 1. > cbind(heights, names)

> class(cbind(heights, names))



cbind combines two vectors as columns into one data object and each value in the data object is converted to a character. By using the class() function we see that the class of the new object is a matrix.

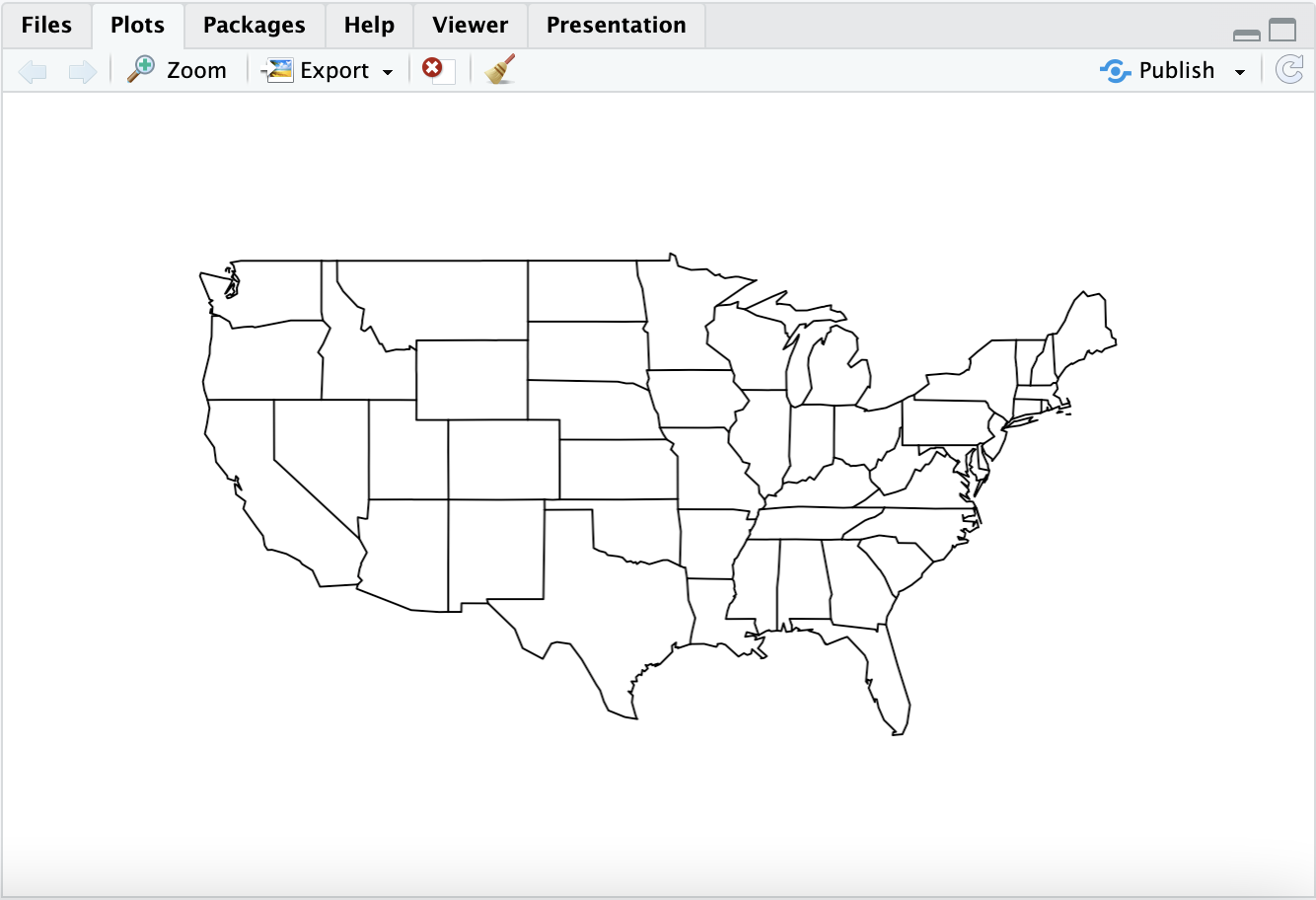
1. Downloading Data
   1. I first changed the working directory and then used the command: > NCbirths <- read.csv(“births.csv)  
      
   2. > head(NCbirths)   
      Output:

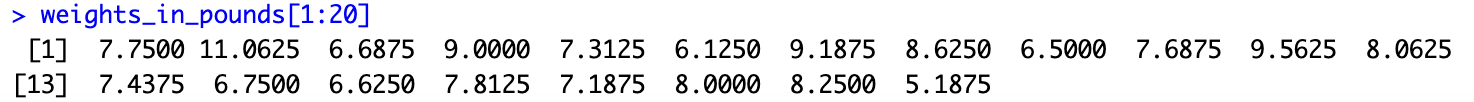


1. Load the Maps Package
   1. > install.packages(“maps”)  
      > find.package(“maps”)

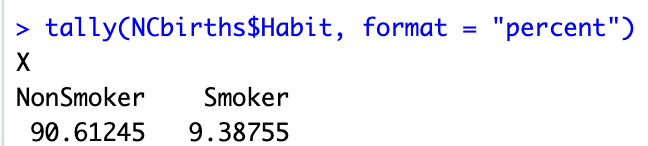


* 1. > library(maps)  
     > map(“state”)



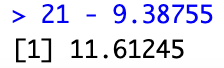
1. Perform Vector Operations
   1. > weights <- NCbirths$weights   
      Output:
   2. The weights are in ounces (oz)
   3. By looking up the conversion factor, I created a new variable weights\_in\_pounds by the following command:   
      > weights\_in\_pounds <- weights / 16  
      
   4. > weights\_in\_pounds[1:20]   
      Output: 

**Section 2: Summarizing Data (one variable)**

1. > mean(weights\_in\_pounds)  
   Output:  
   
2. > install.packages(“mosaic”)  
   > library(mosaic)  
   > tally(NCbirths$Habit, format = “percent”)  
   

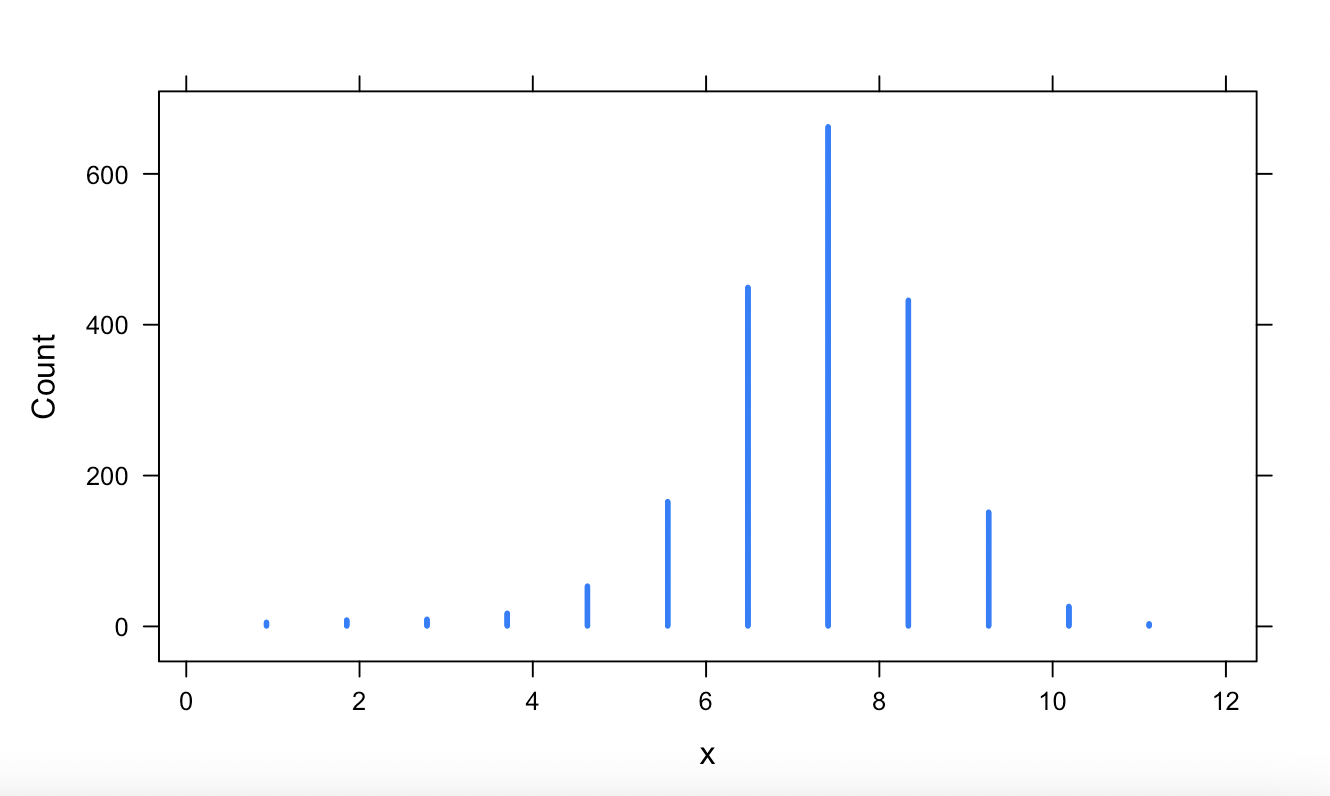
Thus, 9.38755% of the mothers in the sample smoke.

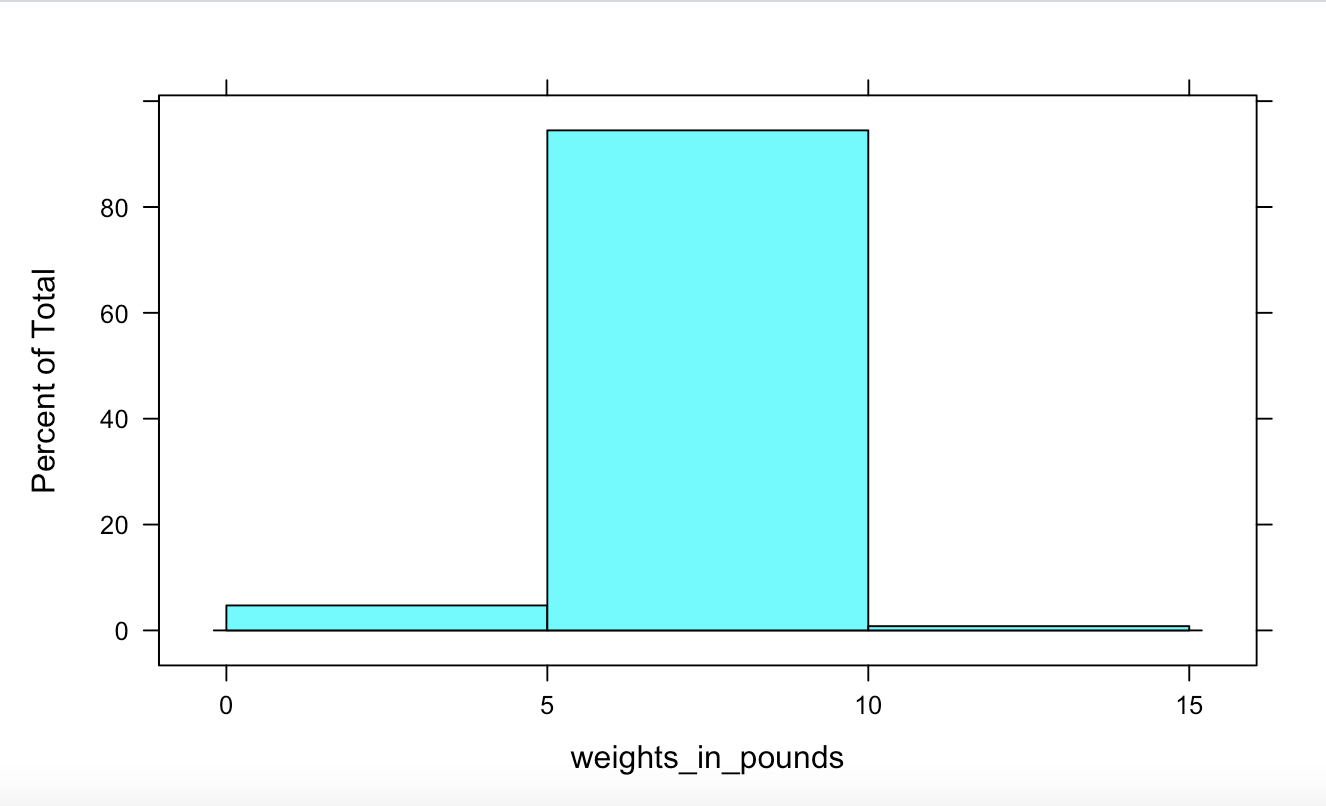
1. > 21 - 9.38755

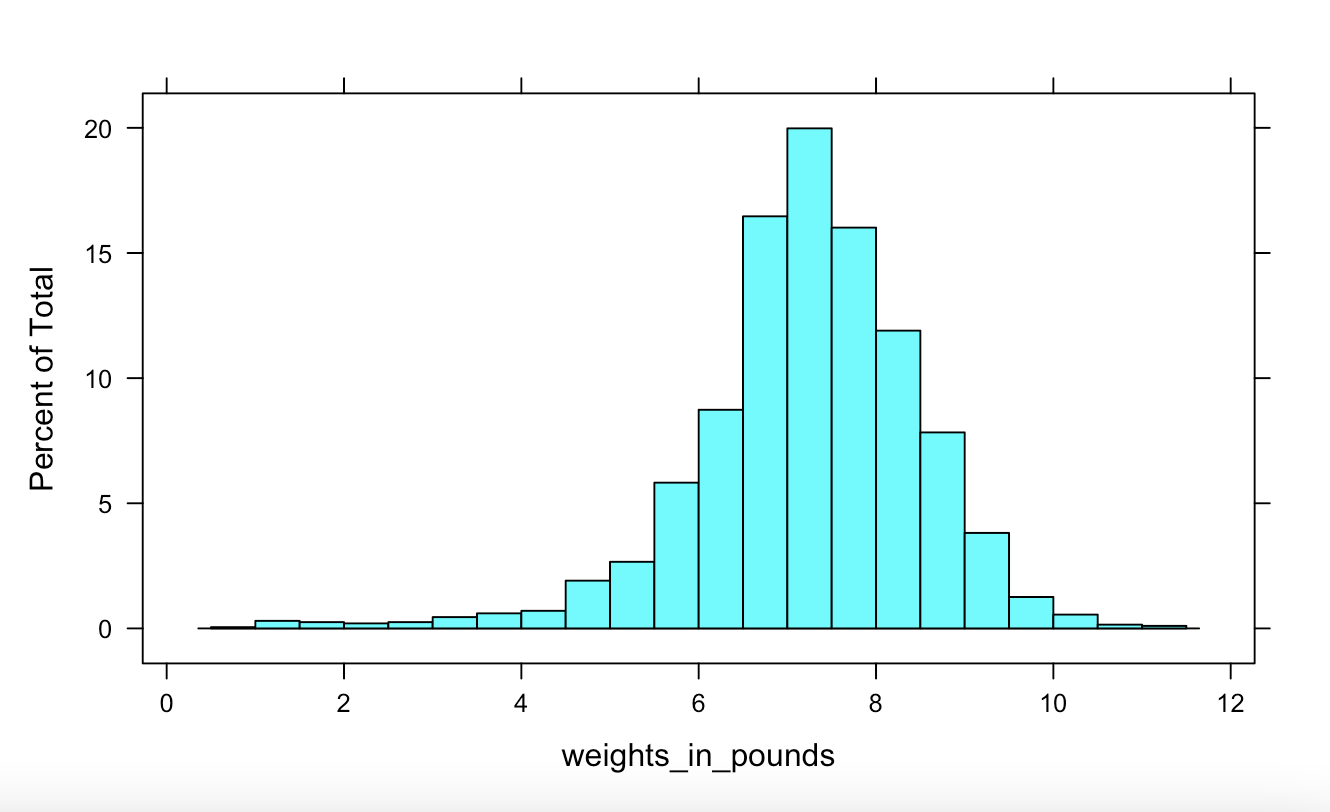
Output:  
 

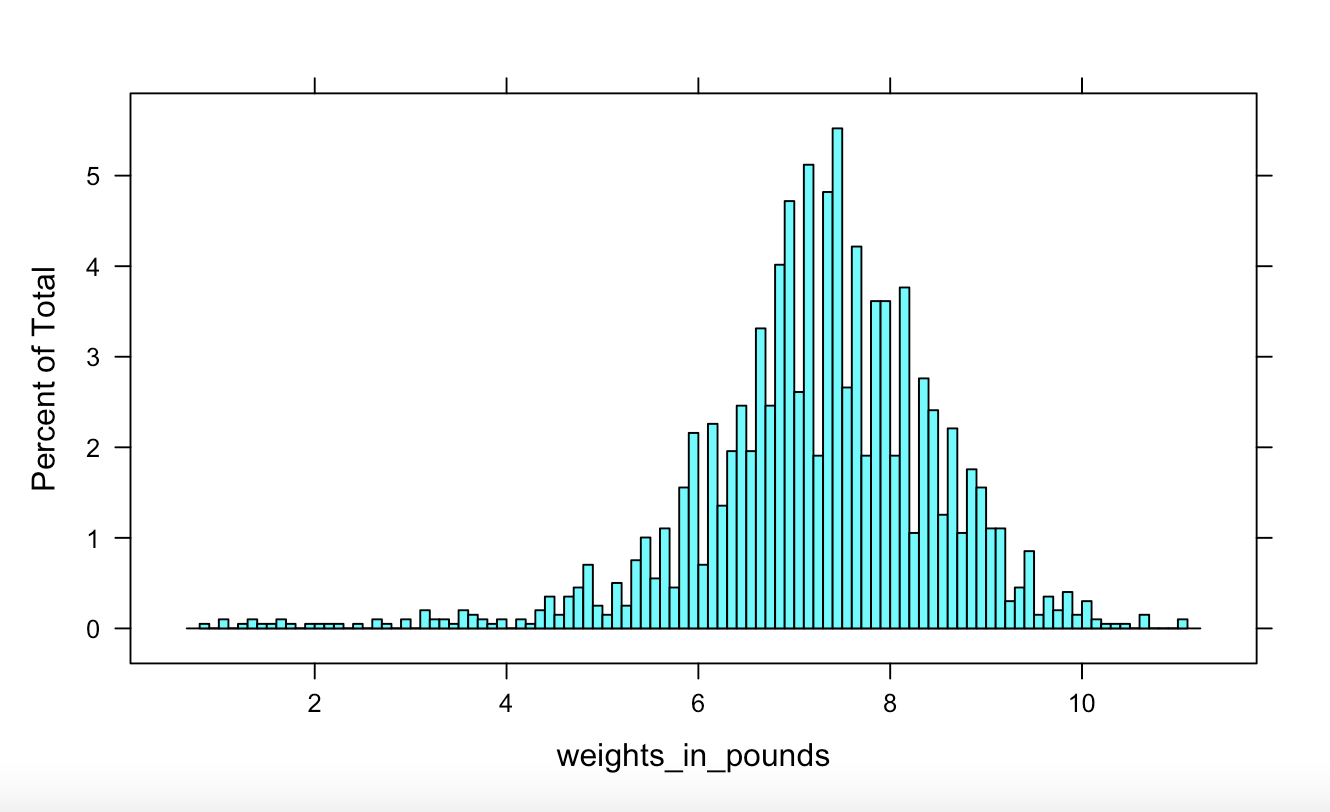
Thus, the percentage in (2) is 11.61245% off from the CDC’s report.

**Section 3: Visualizing Data (one quantitative variable)**

1. > dotPlot(weights\_in\_pounds, cex = 10)  
   
2. > histogram(weights\_in\_pounds, breaks = 3)

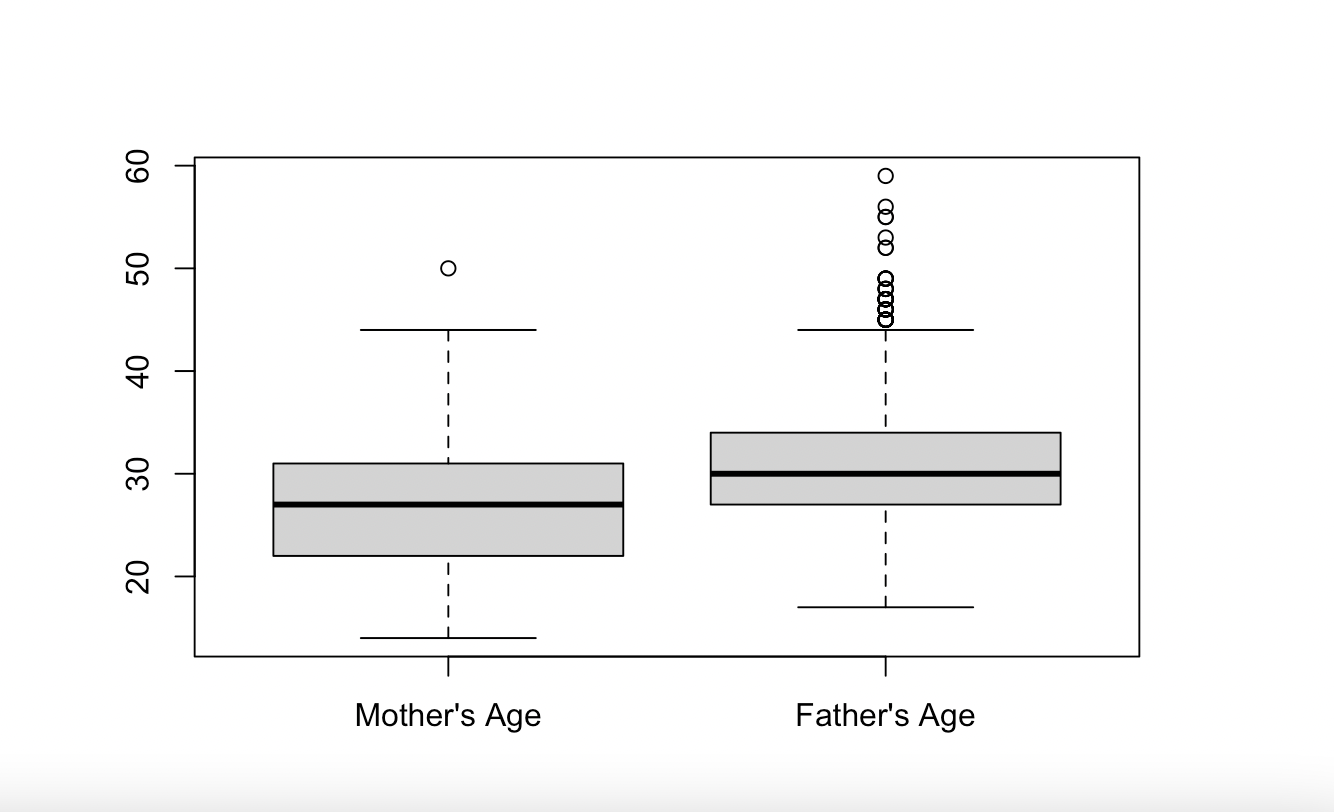


> histogram(weights\_in\_pounds, breaks = 20)  


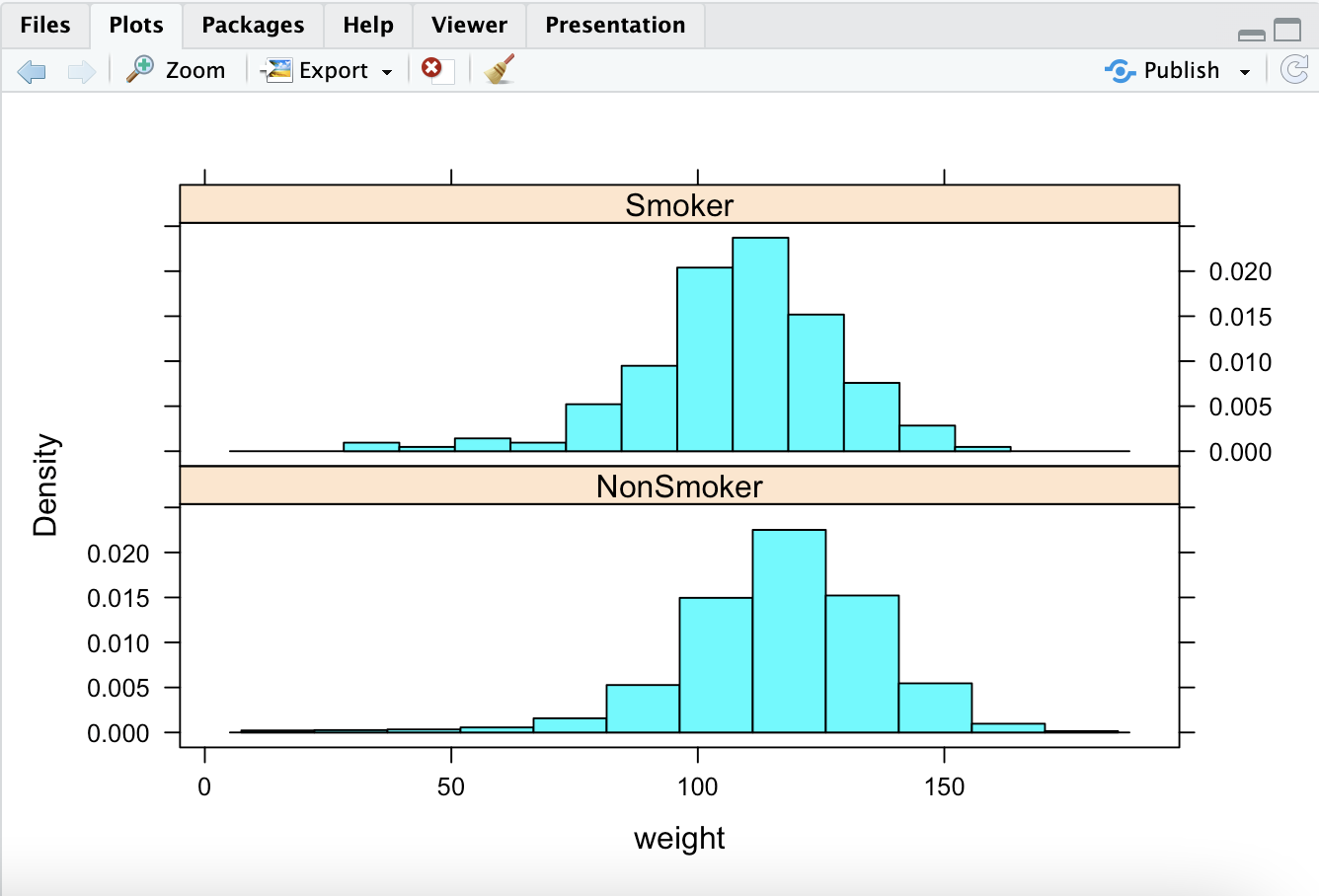
> histogram(weights\_in\_pounds, breaks = 100)  


The histogram with 20 bins seems to give the best visualization since it allows for us to observe the general pattern of the data without interrupting the smoothness of representation. The histogram with 3 bins is too broad and not informative, while the histogram with 100 bins is too specific.

1. > boxplot(NCbirths$Mage, NCbirths$Fage, names = c("Mother's Age", "Father's Age"))



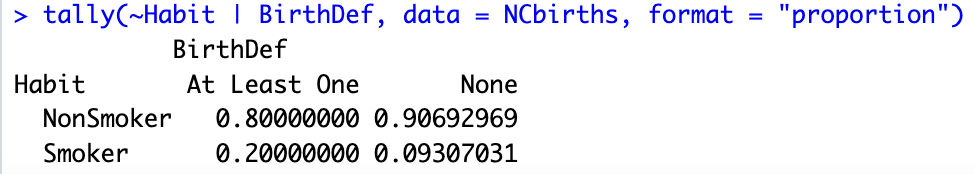
It is evident from the side by side box plots that the fathers tend to be older.

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

The above command creates two stacked density vs. weight histograms corresponding to two categories of the habit variable. Notice that the central tendency is higher for non-smoking mothers in comparison to smoking mothers, i.e. the variability in babies of smoking mothers is higher than non-smoking mothers.

**Section 4: Visualizing Data (one quantitative)**

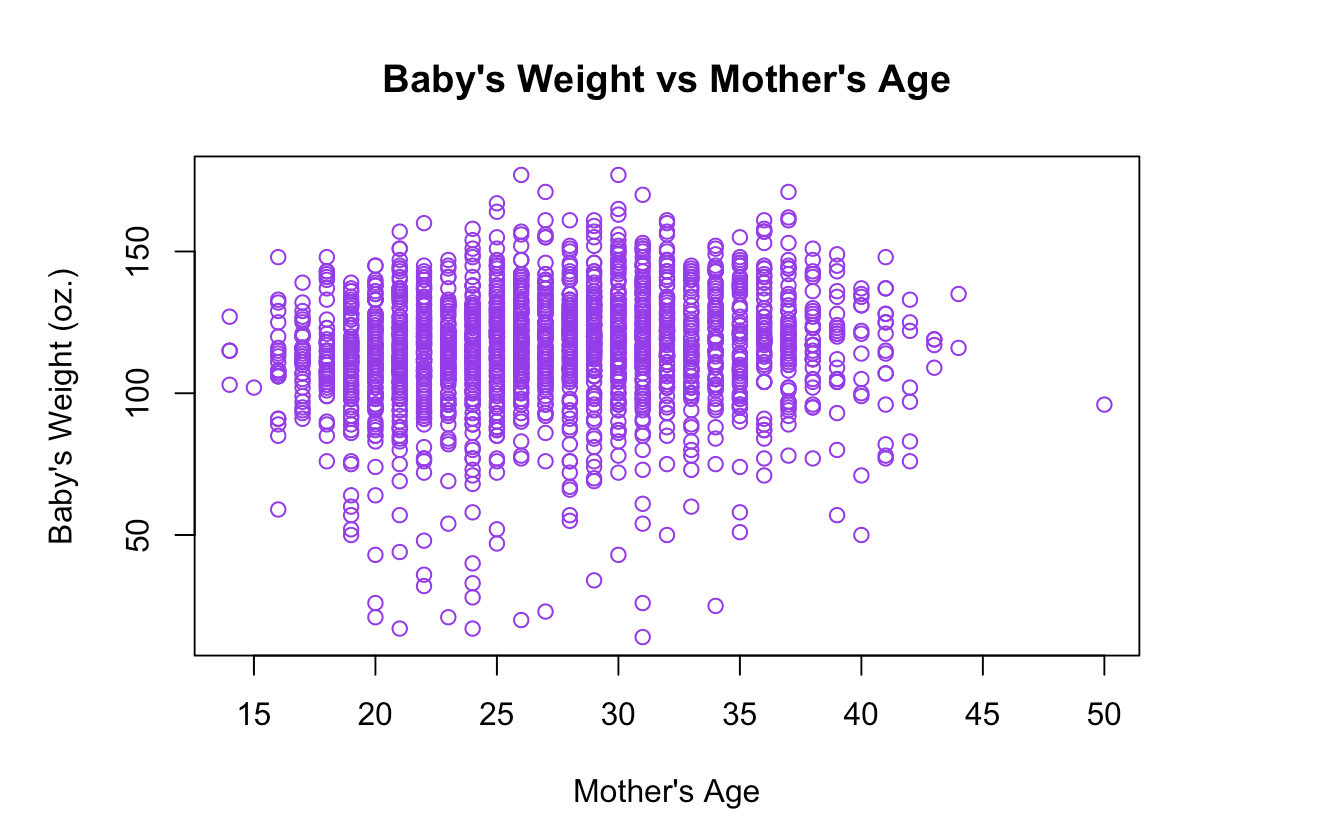
1. Smoking habits of the mother are likely to result in birth defects in children.   
   > tally(~Habit | BirthDef, data = NCbirths, format = "proportion")



It is evident from the two-way table that approximately 90% of the children born from non-smoking mothers have zero birth defects, while only a meager 9% of the children born from smoking mothers have zero birth defects. Due to the stark difference in the statistics, we can attribute this variable to smoking.

**Section 5: Visualizing Data (two quantitative variables)**

1. > plot(NCbirths$weight ~ NCbirths$Mage, col = "purple",cex = 1, pch = 1, xlab = "Mother's Age", ylab = "Baby's Weight (in oz.)", main = "Baby's Weight vs Mother's Age")



**Section 6: Visualizing Data (geographic data)**

1. > a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics12/ozone.txt", header=TRUE)  
   AQI\_colors <- c("aquamarine2", "gold", "darkmagenta", "steelblue2","firebrick1")

AQI\_levels <- cut(a$o3, c(0, 0.06, 0.075, 0.104, 0.115, 0.374))

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=15)

