

**Legend**

Code: blue

Output: black

**Section 1**

1) a)

```
> pinky <- c(2, 1.8, 2.2)
> pinky
[1] 2.0 1.8 2.2
```

b)

```
> cities <- c("LA", "Rowland Heights", "Diamond Bar")
> cities
[1] "LA" "Rowland Heights" "Diamond Bar"
```

c)

```
> cbind(pinky, cities)
      pinky cities
[1,] "2"    "LA"
[2,] "1.8"  "Rowland Heights"
[3,] "2.2"  "Diamond Bar"
> class(cbind(pinky, cities))
[1] "matrix" "array"
```

This command organized the vectors into columns. The class of this new object is matrix.

2) a)

```
> NCbirths <- read.csv(file = "births2022.csv")
```

b)

```
> head(NCbirths)
```

	Gender	Premie	weight	Apgar1	Fage	Mage	Feduc	Meduc	TotPreg	Visits	Marital
1	Female	No	118	8	30	20	12	11	3	13	Unmarried
2	Male	No	106	8	23	26	12	11	2	12	Unmarried

3 Female No 130 8 21 19 12 12 1 11 Unmarried

4 Male No 112 9 34 30 14 14 3 7 Married

5 Female No 102 9 25 26 12 14 2 14 Married

6 Female No 134 9 20 19 12 13 1 11 Married

Racemom Racedad Hispmom Hispdad Gained Habit MomPriorCond BirthDef

1 Black Unknown NotHisp Unknown 44 Smoker None None

2 Black Black NotHisp NotHisp 30 Smoker None None

3 White White NotHisp NotHisp 32 NonSmoker None None

4 White White NotHisp NotHisp 30 NonSmoker At Least One None

5 Black Black NotHisp NotHisp 53 NonSmoker None None

6 White White NotHisp NotHisp 68 NonSmoker None None

DelivComp BirthComp

1 At Least One None

2 None At Least One

3 None None

4 At Least One None

5 None None

6 None None

3) a)

```
> find.package("maps")
```

```
[1] "/Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/library/maps"
```

b)

```
> library(maps)
```

```
> map("state")
```



4) a)

```
> weights <- NCbirths$weight
```

b)

ounces

c)

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

d)

```
> weights_in_pounds[1:10]
```

```
[1] 7.3750 6.6250 8.1250 7.0000 6.3750 8.3750 7.2500 8.3750 7.3125
```

```
[10] 3.8125
```

## Section 2

1)

```
> mean(NCbirths$Fage)
```

```
[1] 30.26103
```

```
> sd(NCbirths$Fage)
```

```
[1] 6.056574
```

2)

```
> tally(NCbirths$Habit, format = "count")
```

X

NonSmoker	Smoker
-----------	--------

1747	180
------	-----

```
> tally(NCbirths$Habit, format = "percent")
```

X

NonSmoker	Smoker
-----------	--------

90.659056	9.340944
-----------	----------

180 mothers in the sample smoke and 9.340944% of the mothers in the sample smoke.

3)

```
> habit_tally <- tally(NCbirths$Habit, format = "percent")
```

```
> 17 - habit_tally[2]
```

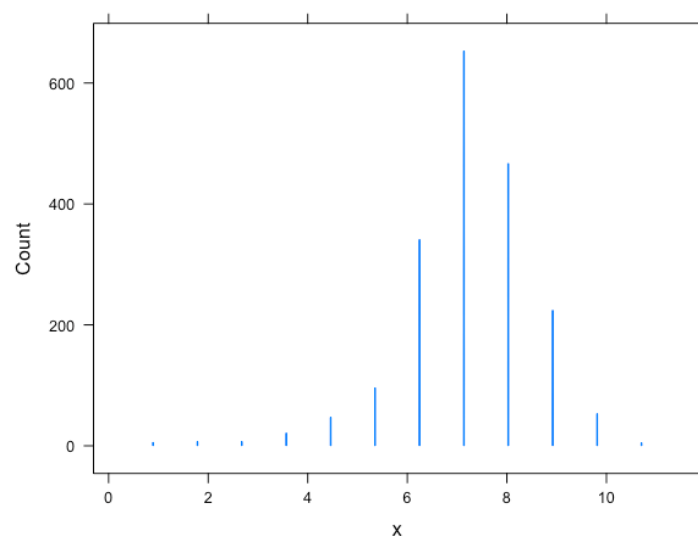
Smoker

7.659056

The percentage I found is 7.659056% off from the CDC's report in 2022

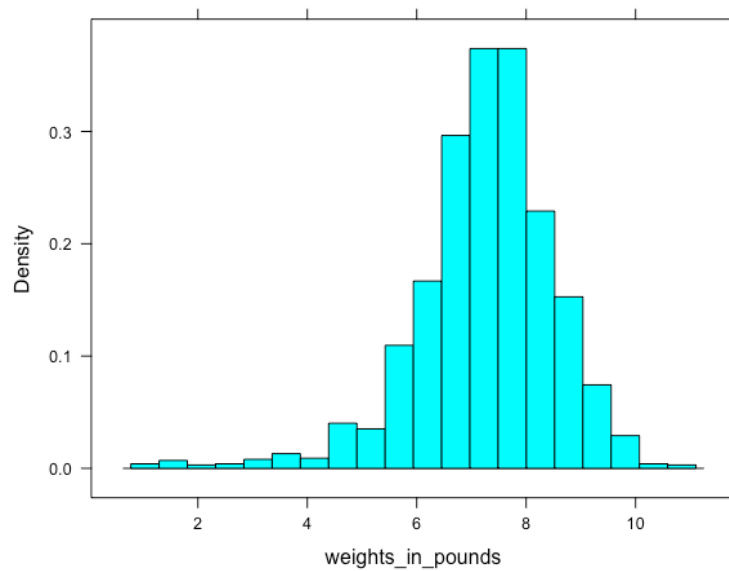
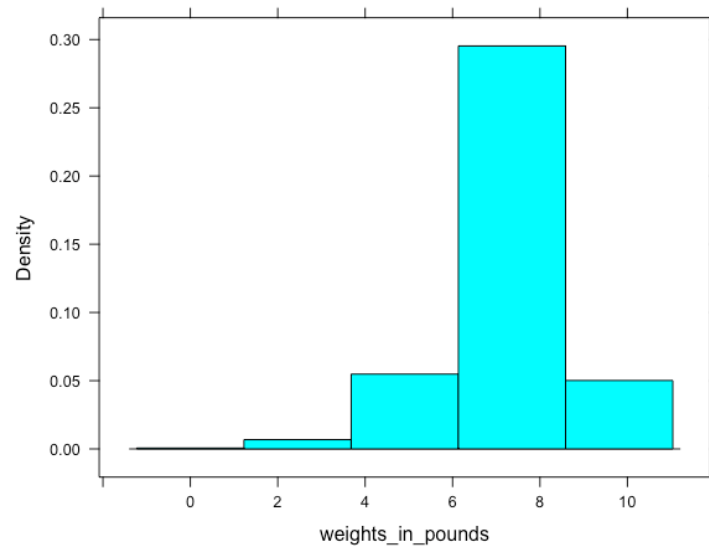
### Section 3

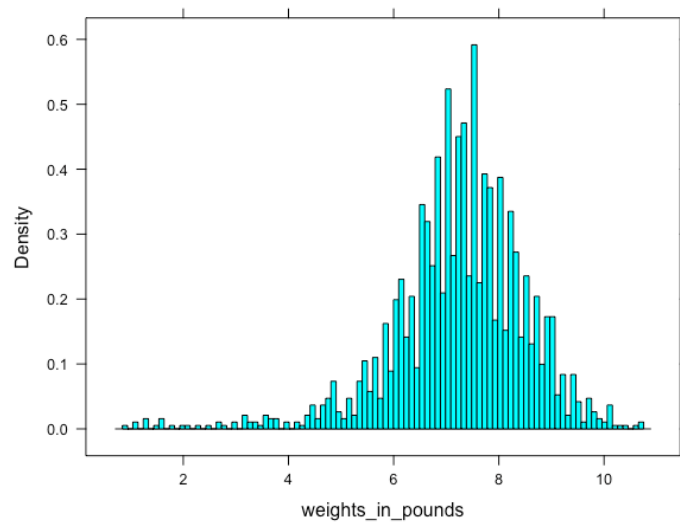
1)



This dot plot appears unimodal and left-skewed. The perceived values are not unusually small in the distribution so there are no outliers. Its typical value is around  $x=7$  so the spread is a bit small.

2)

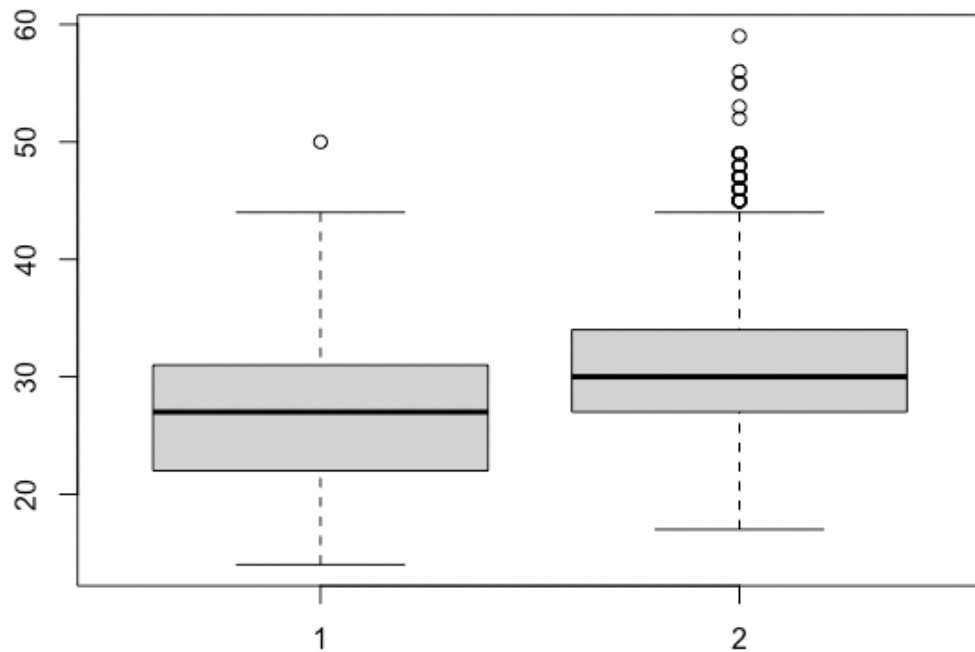




A histogram with more bins gives a more detailed representation and is generally better for data with some skewness or noticeable trends. On the other hand, if the data is more spread out and has smaller y-values, a histogram with fewer bins may be better so the viewer can see the range of the data. Therefore, the histogram with the best visualization is more of a personal choice.

3)

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

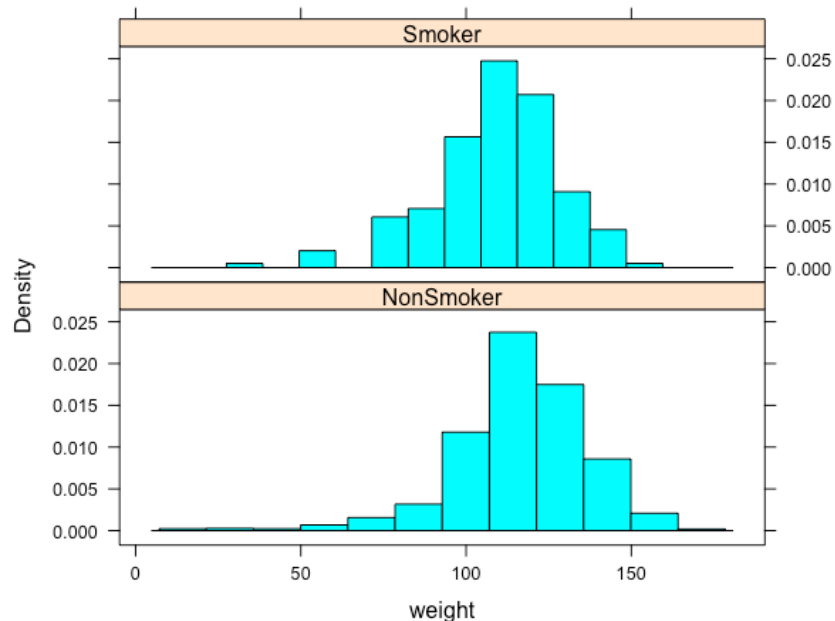


The boxplot below indicates that the father's ages tend to be older

4)

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

The code above generates a histogram from the category “Habit” and the baby weights associated with smoking or non-smoking moms. There are no major differences between baby weights from smoking moms and non-smoking moms since spread, central tendency, and the shape of both histograms are extremely similar.



## Section 4

1)

Hypothesis: I think the gender of the baby will be associated with the mother’s smoking because a mother’s failing health may affect her sex cells that are supposed to reproduce the baby.

```
> tally_condition_new <- tally(~Habit | Gender, data = NCbirths, format = "proportion")
```

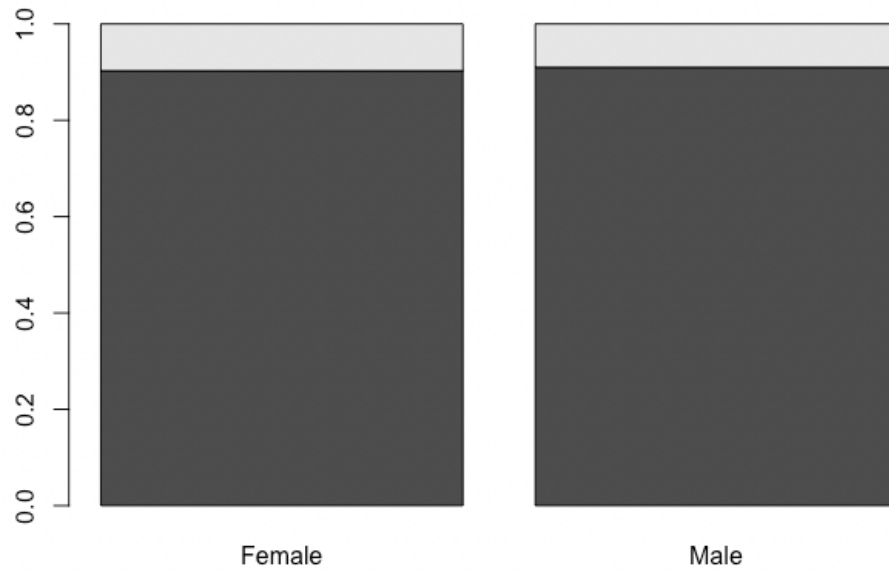
```
> tally_condition_new
```

Gender

Habit	Female	Male
NonSmoker	0.90291262	0.91000000
Smoker	0.09708738	0.09000000

There is no clear evidence or indication of the gender variable being associated with smoking because the probabilities for non-smokers and smokers in the two-way table are very similar.

2)



## Section 5

1)

```
> plot(NCbirths$weight ~ NCbirths$Mage, col = "green", cex = 1.5, pch = 4,  
xlab = "Mother's Age during Pregnancy", ylab = "Baby weight (oz.)",  
+      main = "Baby Weight vs. Mother's Age")
```

Based on the scatter plot below, there is a central tendency for the baby's weight to be roughly 115 ounces for most mothers during pregnancy.



**Baby Weight vs. Mother's Age**

