Problem Set 3

Kristina Finley
STAT 100, SECTION 0221

PROBLEM #1

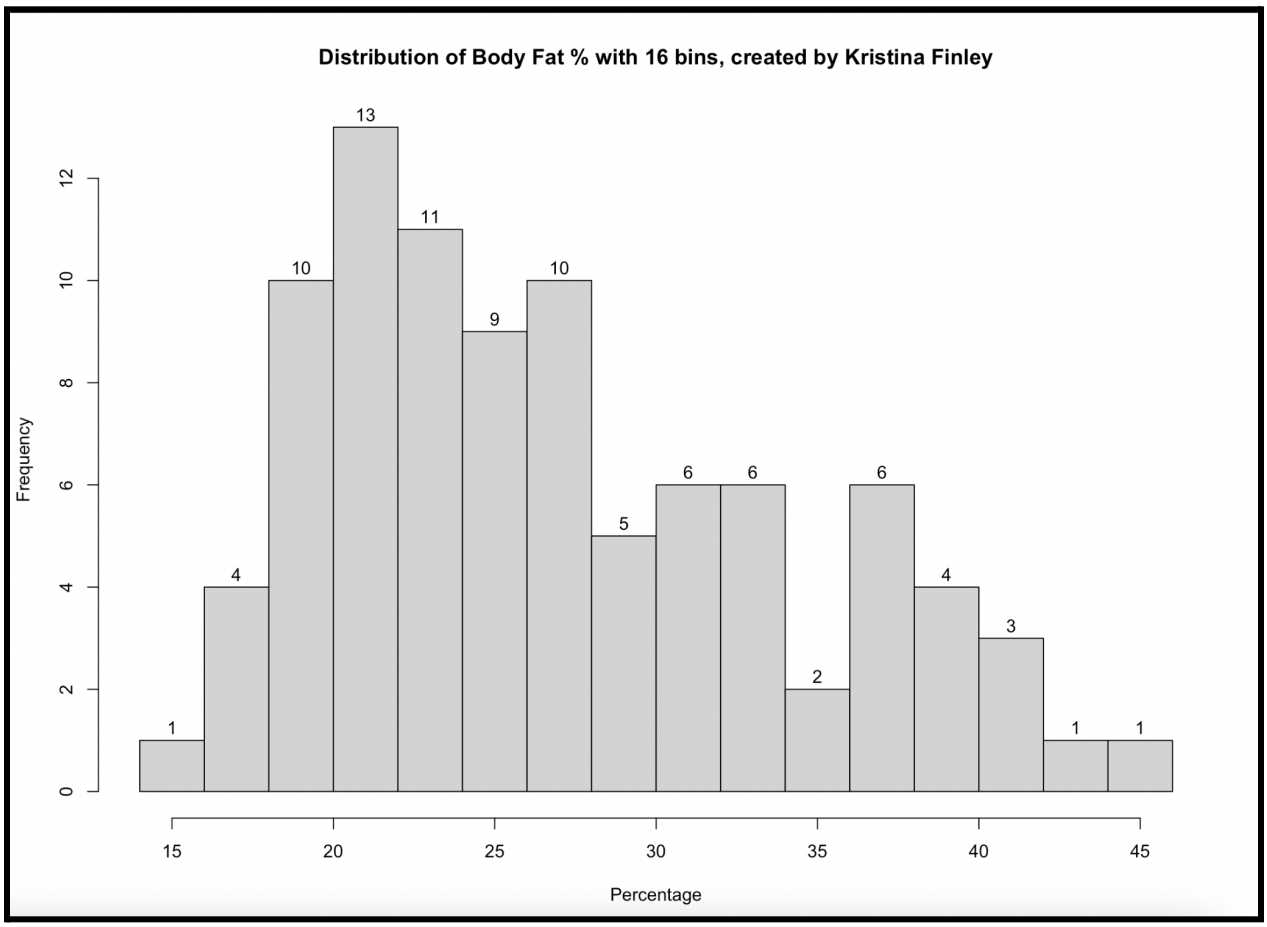
```
> #1(a-1) extract variable Pct_Fat from a data frame, by Kristina Finley
> Pct_Fat <- BodyFatPercentage$Pct_Fat
> Pct_Fat
 [1] 23.9 28.8 32.4 25.8 22.5 22.1 19.6 25.3 22.8 26.4 33.7 27.9 33.5 23.4 21.8
[16] 37.9 31.3 40.6 36.3 29.8 31.9 31.3 21.6 24.6 20.1 24.6 18.1 22.9 26.2 27.2
[31] 17.7 20.8 17.5 21.3 18.7 28.8 17.1 26.2 20.4 19.5 21.7 18.1 29.8 20.6 22.9
[46] 19.3 38.4 27.9 36.4 25.1 39.7 33.6 46.0 38.9 42.2 36.7 38.0 23.3 35.9 24.1
[61] 40.8 25.7 37.6 35.9 36.3 33.0 40.5 26.4 27.3 32.2 19.6 24.5 22.6 30.2 26.9
[76] 30.2 21.0 19.4 21.1 17.3 20.5 19.3 28.7 18.3 15.6 23.9 24.5 23.3 20.1 30.3
[91] 20.6 26.0
> #1(a-2) generate desriptive statistics for the variable Pct_Fat,
> #by Kristina Finley
> summary(Pct_Fat)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  15.60   21.07   25.50   26.96   31.98   46.00
>
> #1(b-1) extract variable BMI from a data frame, by Kristina Finley
> BMI <- BodyFatPercentage$BMI
> BMI
 [1] 19.3083 22.9642 27.7900 20.9174 20.3784 20.3862 19.6575 20.2983 20.6038
[10] 20.3064 21.2133 22.1067 28.6048 19.4964 20.4143 26.8534 21.4775 29.7552
[19] 23.9239 20.5508 21.6740 19.2703 18.1786 18.4596 17.0548 17.7009 16.6123
[28] 16.9420 18.7744 18.3869 17.8625 17.9756 15.3674 18.8354 15.8201 17.7127
[37] 14.9863 16.7468 16.4636 15.8716 18.0788 15.5805 17.1543 15.8181 18.6112
[46] 16.6625 25.1320 20.8334 24.5595 20.1940 24.1308 23.8650 33.5654 29.1450
[55] 28.5919 26.1749 34.4610 19.0836 23.5403 20.2406 29.1700 20.8476 22.6612
[64] 30.0781 22.6121 24.2981 31.3277 24.1429 23.4981 24.7797 19.4615 19.0188
[73] 20.1973 20.6894 19.1685 20.7282 16.1438 17.7113 16.0929 16.7395 18.9257
[82] 17.2245 18.3838 18.1828 17.0757 17.6386 18.6891 17.5179 16.1198 18.8322
[91] 15.4559 18.3936
> #1(b-2) generate desriptive statistics for the variable BMI,
> #by Kristina Finley
> summary(BMI)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  14.99   17.71   19.48   20.71   22.74   34.46
>
```

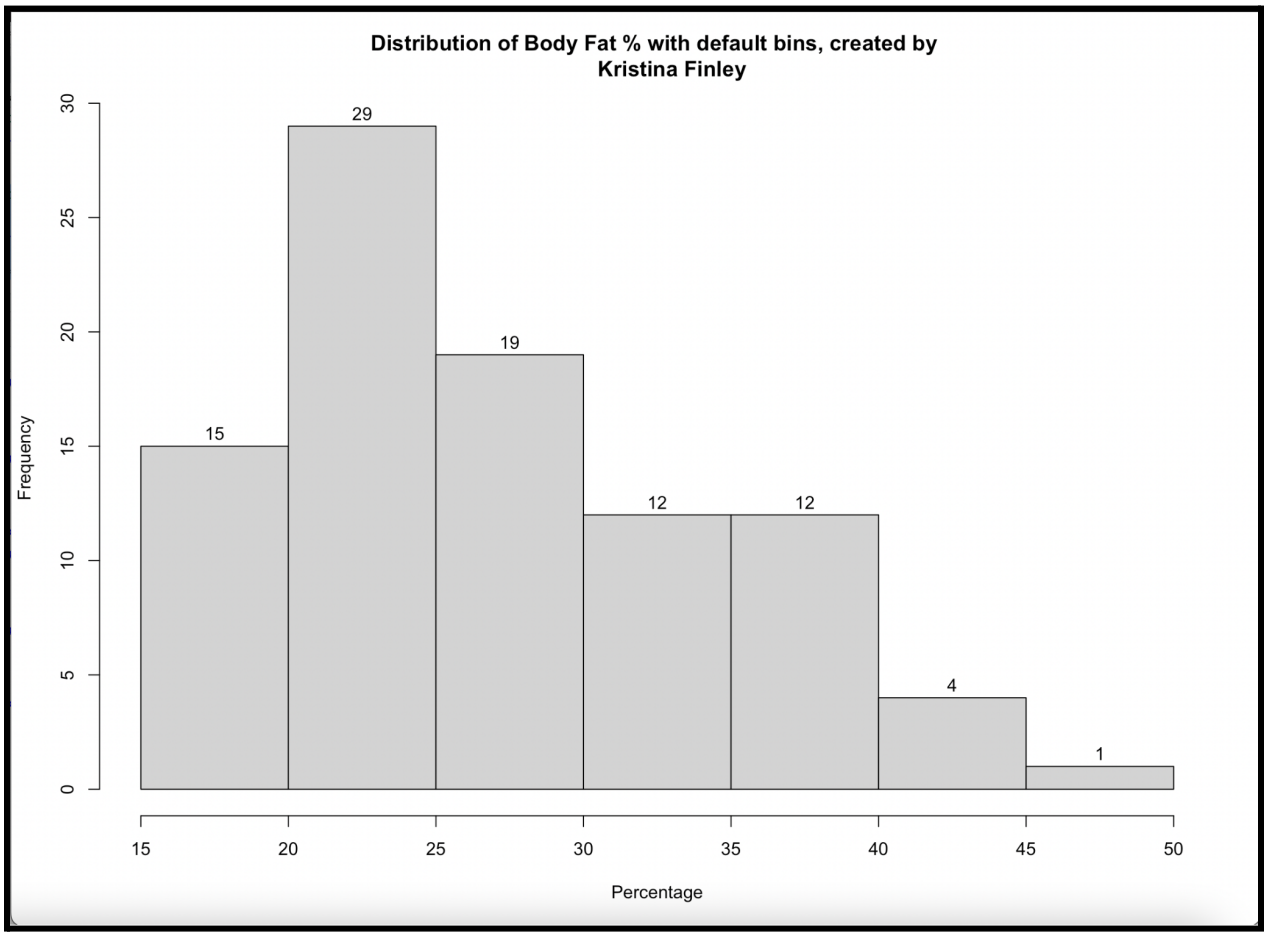
C.
IQR: 31.98 - 21.07 = 10.91

31.98 + 1.5(10.91) = 48.35
Yes, 49% would be considered an outlier because it is greater than 48.35.
D.
IQR: 22.74 - 17.71 = 5.03
17.71 - 1.5(5.03) = 10.17
No, a BMI of 12 would not be considered since the lower limit is less than 12.

PROBLEM #2

```
> #2(a-1) create a histogram for variable Pct_Fat with 16 bins, by Kristina Finley
> hist(x = Pct_Fat, right = FALSE, labels = TRUE, breaks = 16, main =
+        "Distribution of Body Fat % with 16 bins, created by Kristina Finley",
+      xlab = " Percentage ")
> #2(a-2) create a histogram for variable Pct_Fat with default number of bins,
> #by Kristina Finley
> hist(x = Pct_Fat, right = FALSE, labels = TRUE, main =
+        "Distribution of Body Fat % with default bins, created by
+      Kristina Finley", xlab = " Percentage ")
>
```





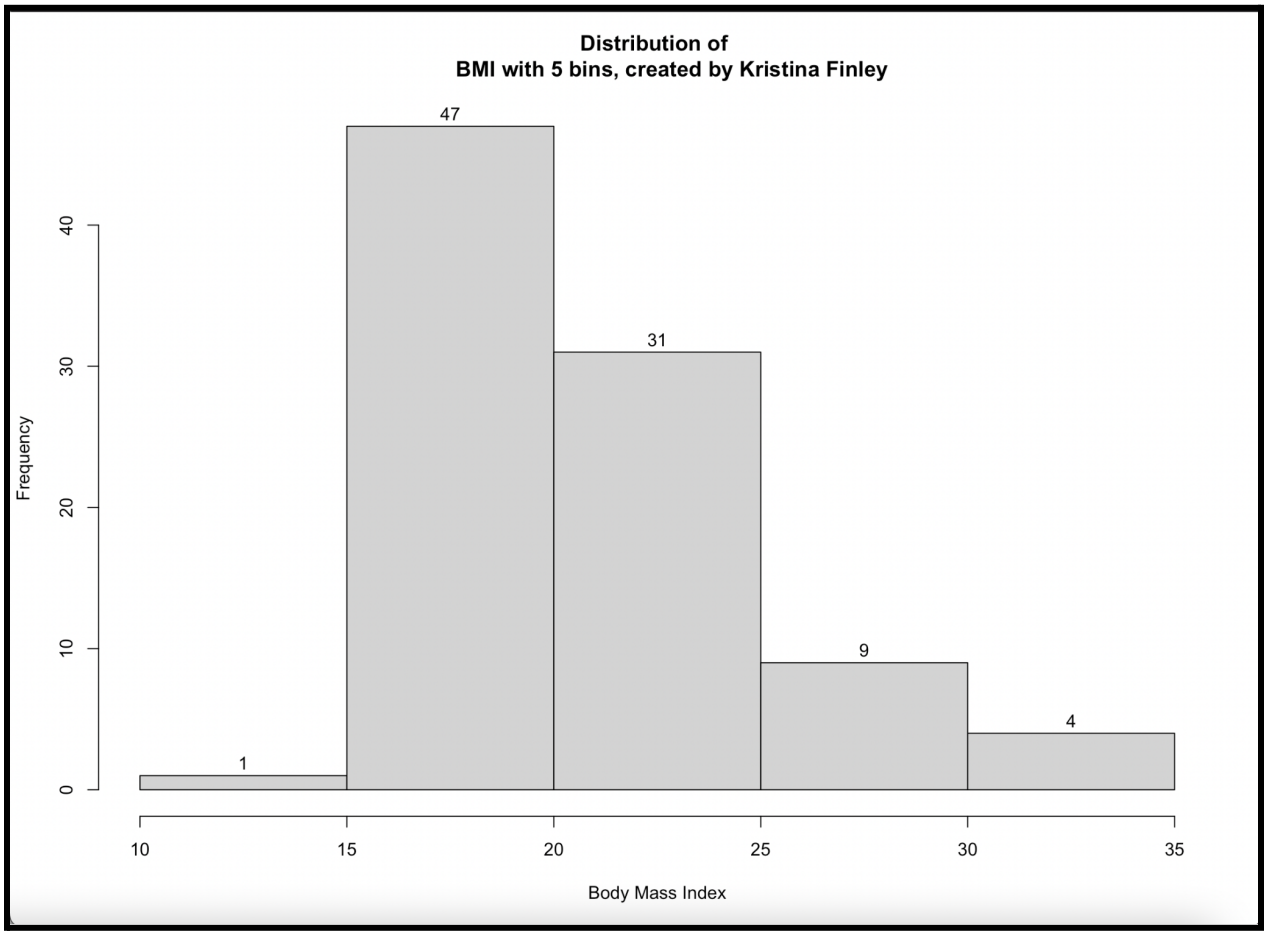
B.

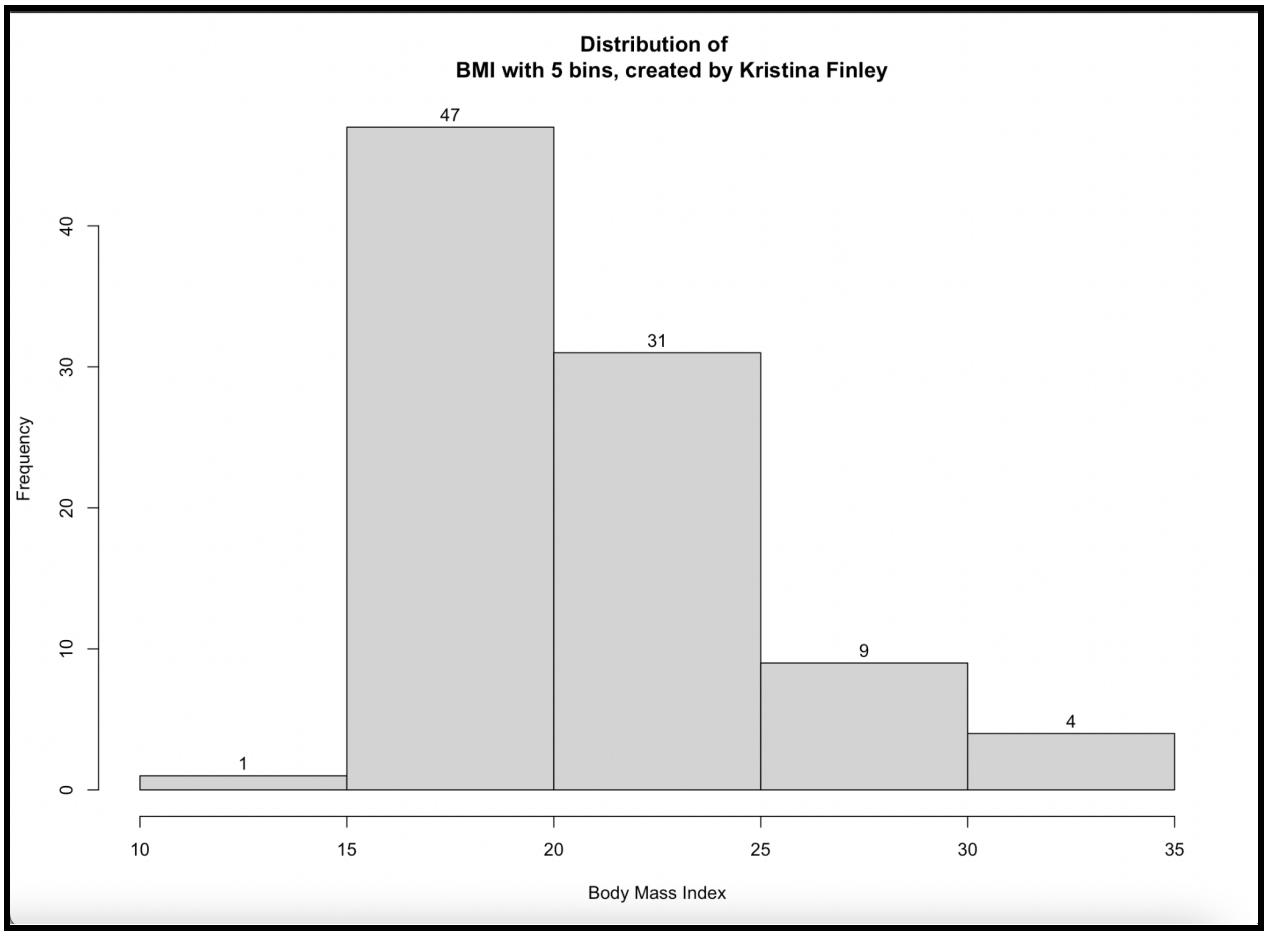
Both of the histograms are skewed right, and it is unimodal. To measure the center of both of the histograms, it is best to use the median. To find the spread of both of the histograms, it is best to use IQR. There are no outliers are both of the histograms.

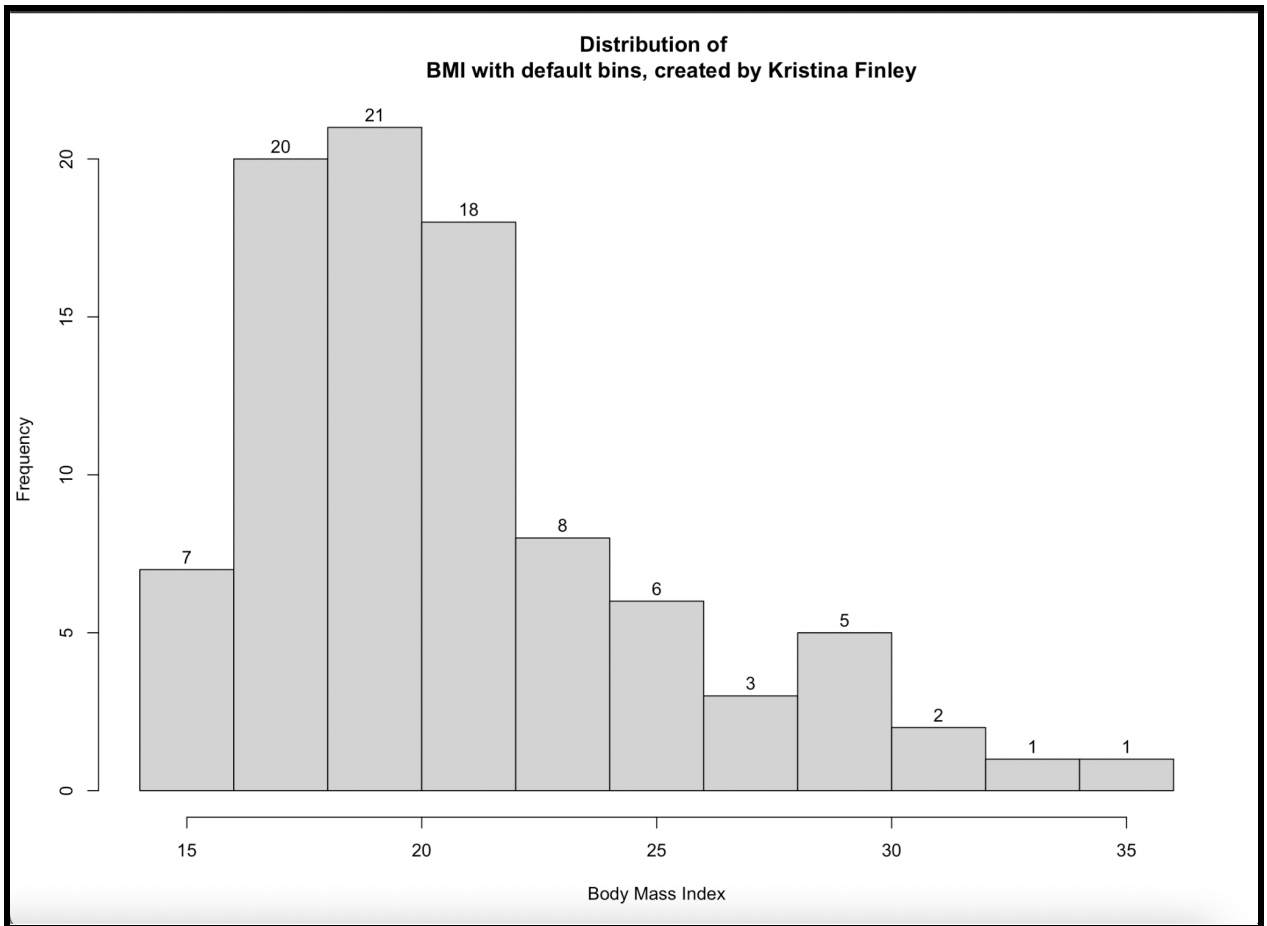
C.

To find the median of the histogram with default bins, I would say the median has to be between 25-30. Place all the numbers in a row, then figure out the middle of the data set for the median. The endpoints are 15 and 50.

```
> #2(d-1) create a histogram for variable BMI with 5 bins, by Kristina Finley
> hist(x = BMI, right = FALSE, labels = TRUE, breaks = 5, main = "Distribution of
+      BMI with 5 bins, created by Kristina Finley", xlab = "Body Mass Index")
> #2(d-2) create a histogram for variable BMI with default bins,
> #by Kristina Finley
> hist(x = BMI, right = FALSE, labels = TRUE, main = "Distribution of
+      BMI with default bins, created by Kristina Finley",
+      xlab = "Body Mass Index")
>
```




Distribution of
BMI with 5 bins, created by Kristina Finley
Frequency
0
10
20
30
40
47
31
9
4
1
10
15
20
25
30
35
Body Mass Index


Distribution of
BMI with default bins, created by Kristina Finley
Frequency
0
5
10
15
20
7
20
21
18
8
6
3
5
2
1
1
15
20
25
30
35
Body Mass Index

E.

Both of the histograms are skewed right, and it is unimodal. The one with 5 bins is slightly less obvious it is skewed right, but it is. To measure the center of both of the histograms, it is best to use the median. To find the spread of both of the histograms, it is best to use IQR. There are outliers are both of the histograms beyond 35.

F.

To find the median of the histogram with 5 bins, I would say the median has to be between 20-25. Place all the numbers in a row, then figure out the middle of the data set for the median. The endpoints are 10 and 35.