Untitled

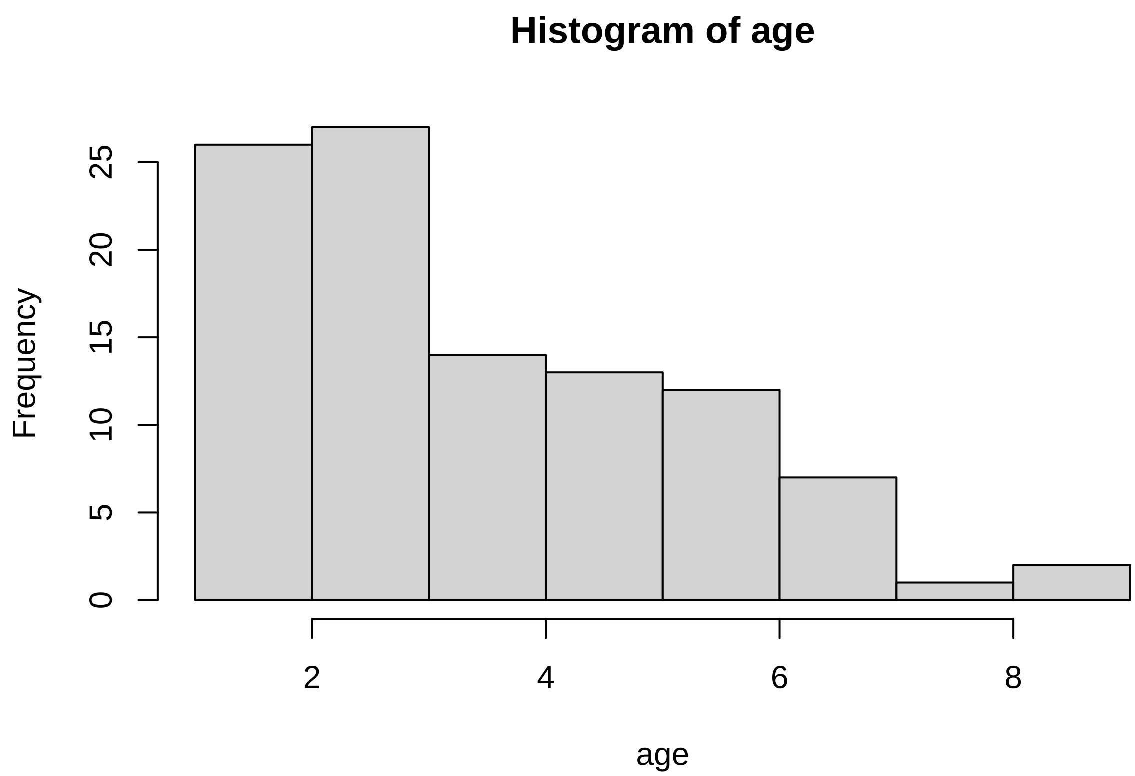
Daniel Briseno Servin

9/28/2020

HW Lecture 5

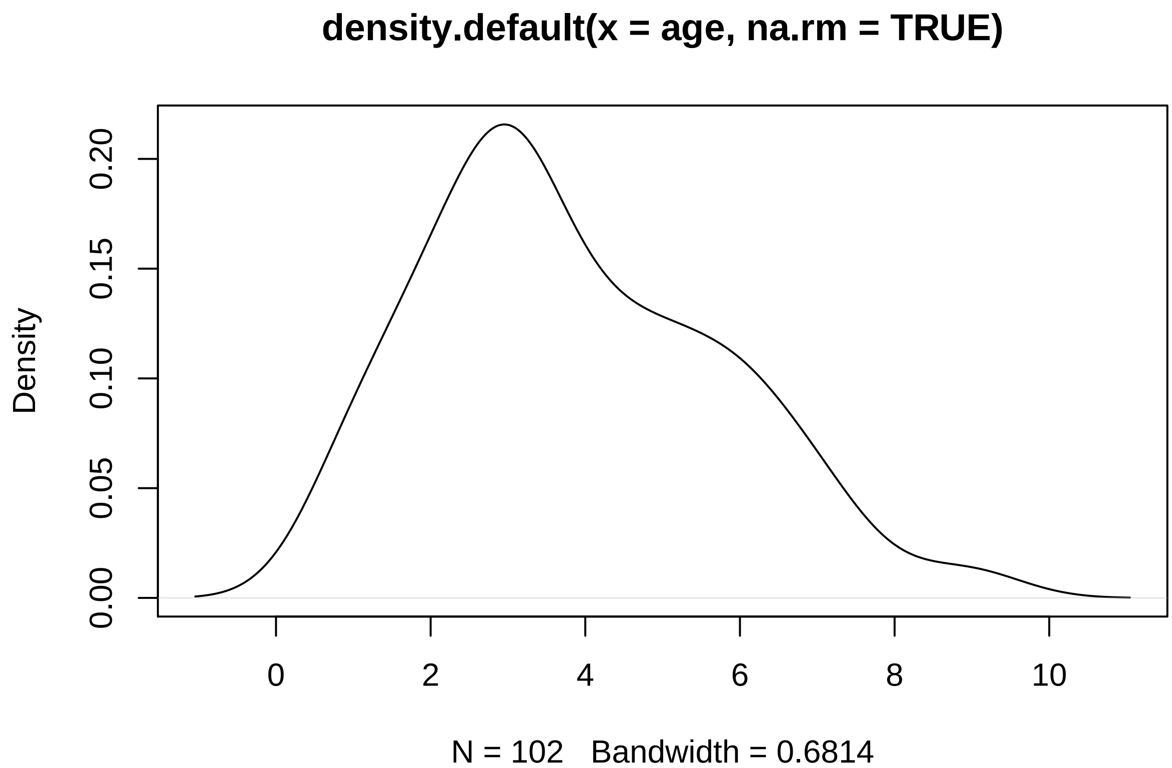
Question 1

```
hist(age)
```



```r
hist(age, seq(0,9,1.5))
```


```r
plot(density(age,na.rm=TRUE))
```

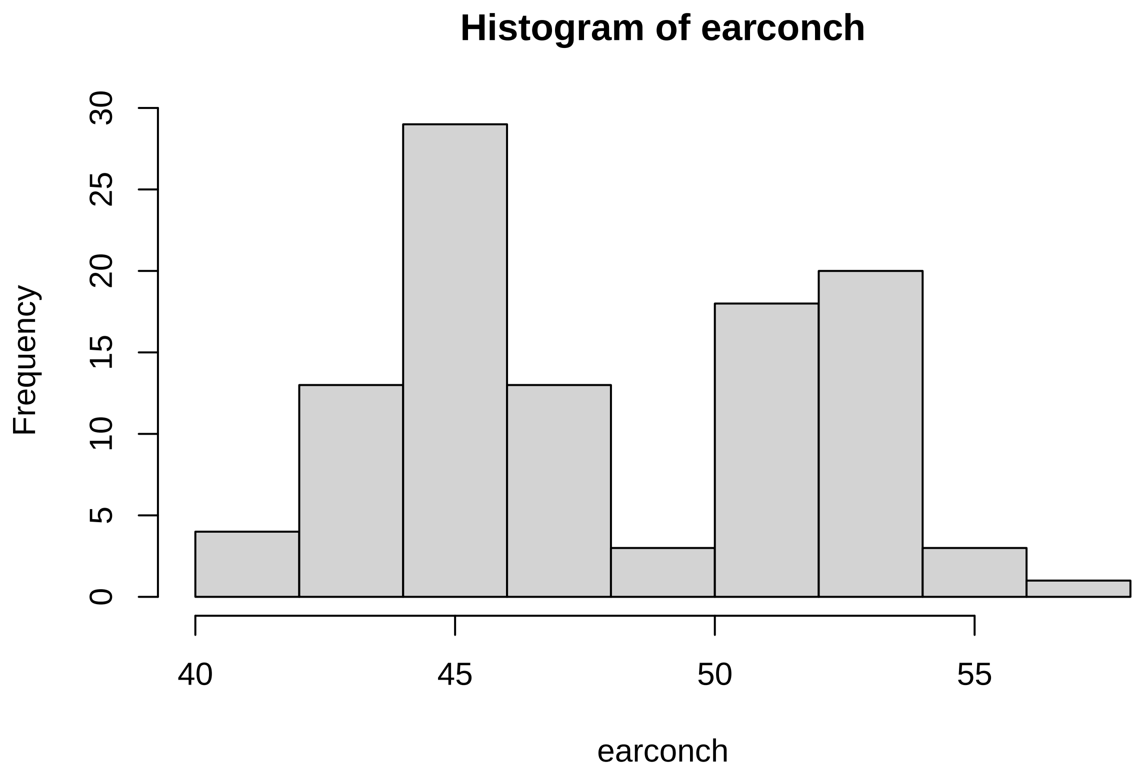


The dramatic difference between the first histogram and the second could be due to the first histogram having bin sizes that are too big. If we cut according to (0,2,4,. . .), then we see that in the 0-2 bin there is a large amount samples in that bin. This can give us the misconception that the data appears to be clustered in the lower extreme values. However, by reducing the bin size by 0.5 we can see that there are relatively few samples in the 0-0.15 bin, a feature of the data which remained hidden in the first histogram.

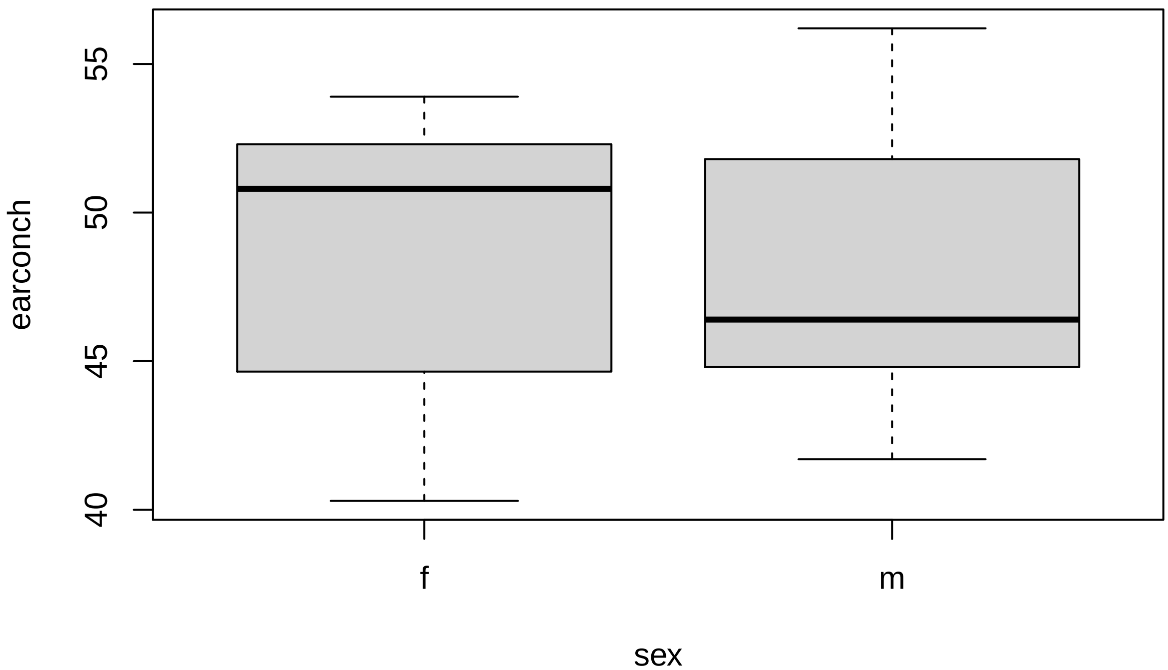
The advantages of using the kernel density estimate is that it more clearly shows trends in the data. The trade-off is that you lose the finer-grained categorical picture shown in the histogram, which can lead a researcher to observe "trends" in the data which could be an artifact of the bandwith rather than a real trend.

Question 2

```
hist(earconch)
```



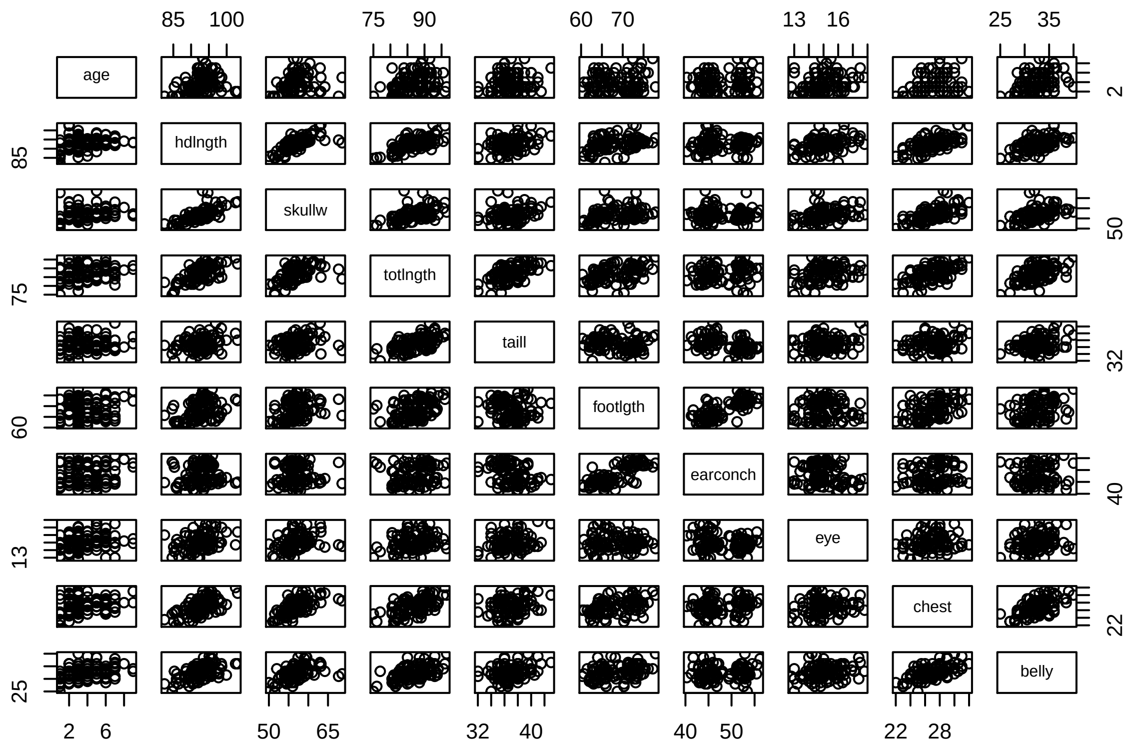

```
boxplot(earconch~sex)
```



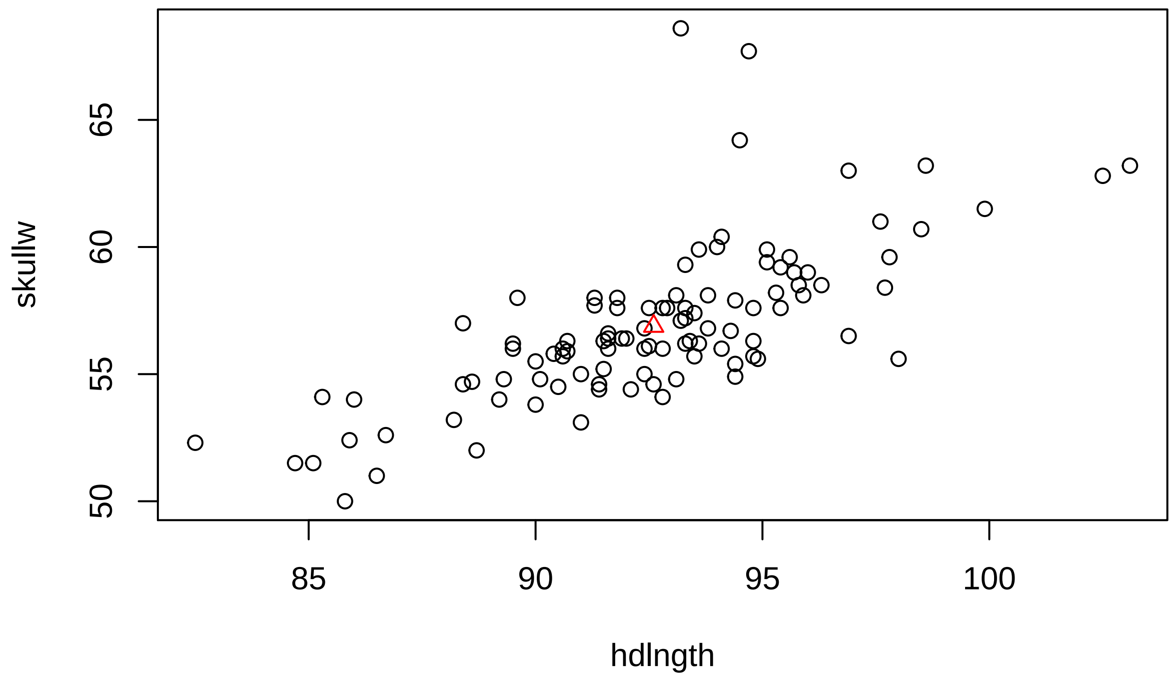
The histogram does indeed show that the data is bimodal, which could lead to suspicion of a 2-factor variable influencing ear conch size. The box-and-whisker plots show that while the interquartile range is not significantly different between males and females, females do have a higher median earconch size than males. Since the interquartile ranges are so similar, I suspect that more data analysis is needed to conclude that the differences in earconch median sizes is statistically significant.

Question 3

```
pairs(~age+hdlngth+skullw+totlngth+taill+footlgth+earconch+eye+chest+belly, data=possum)
```




```
plot(hdlngth,skullw)
points(mean(hdlngth),mean(skullw),pch=2,col="red")
```


skullw
65
60
55
50
85
90
95
100
hdlngth