

HW4

Duc Le

9/24/2020

Problem 1

```
library(DAAG)
```

```
## Loading required package: lattice
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

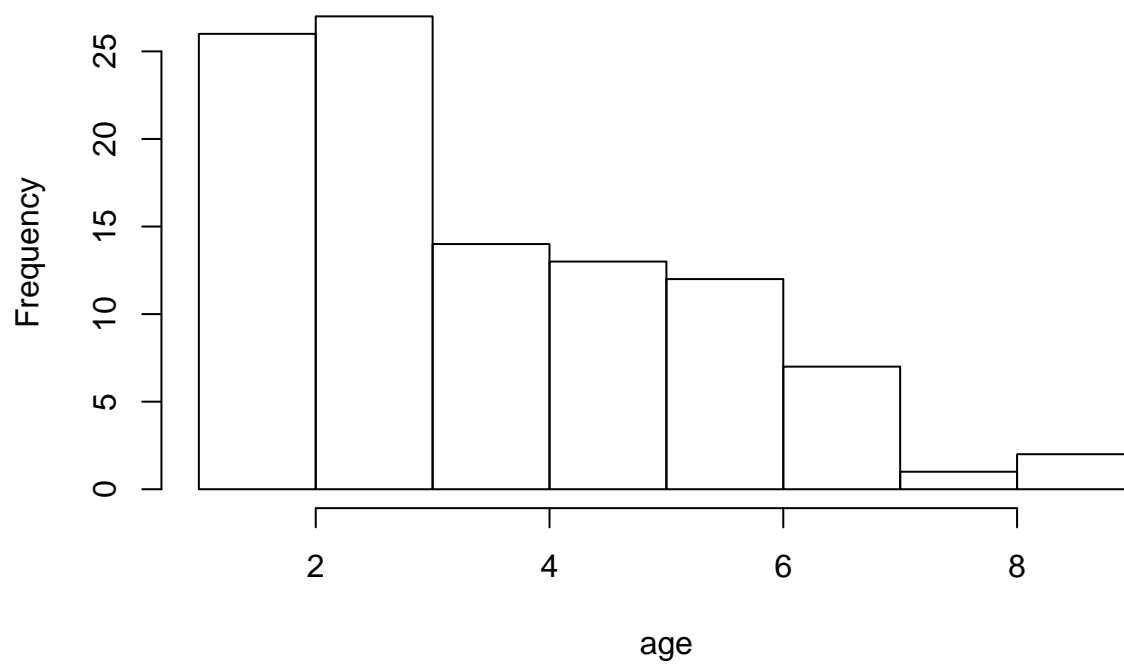
```
##      intersect, setdiff, setequal, union
```

```
data(possum)
```

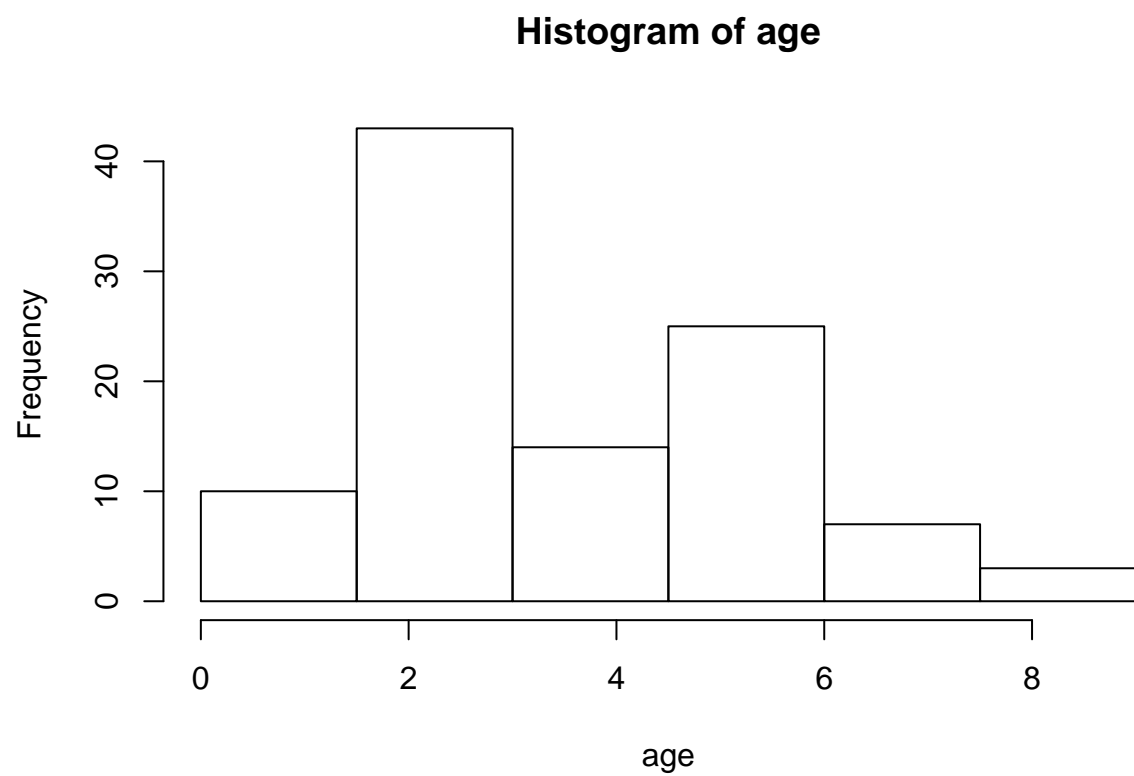
```
attach(possum)
```

```
hist1 = hist(age)
```

Histogram of age



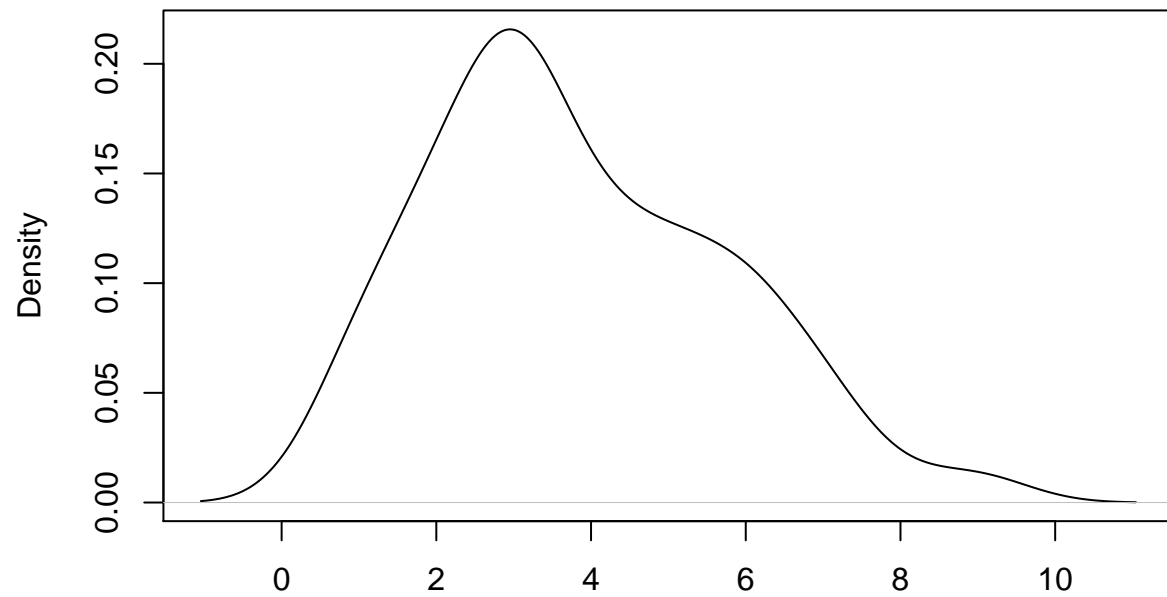
```
hist2 = hist(age, breaks = seq(0,9,1.5))
```



The difference we see in `hist2` with the custom bins is due to the fact that we attempt to split a discrete variable, `age`, into float bins.

```
plot(density(age, na.rm = T))
```

density.default(x = age, na.rm = T)

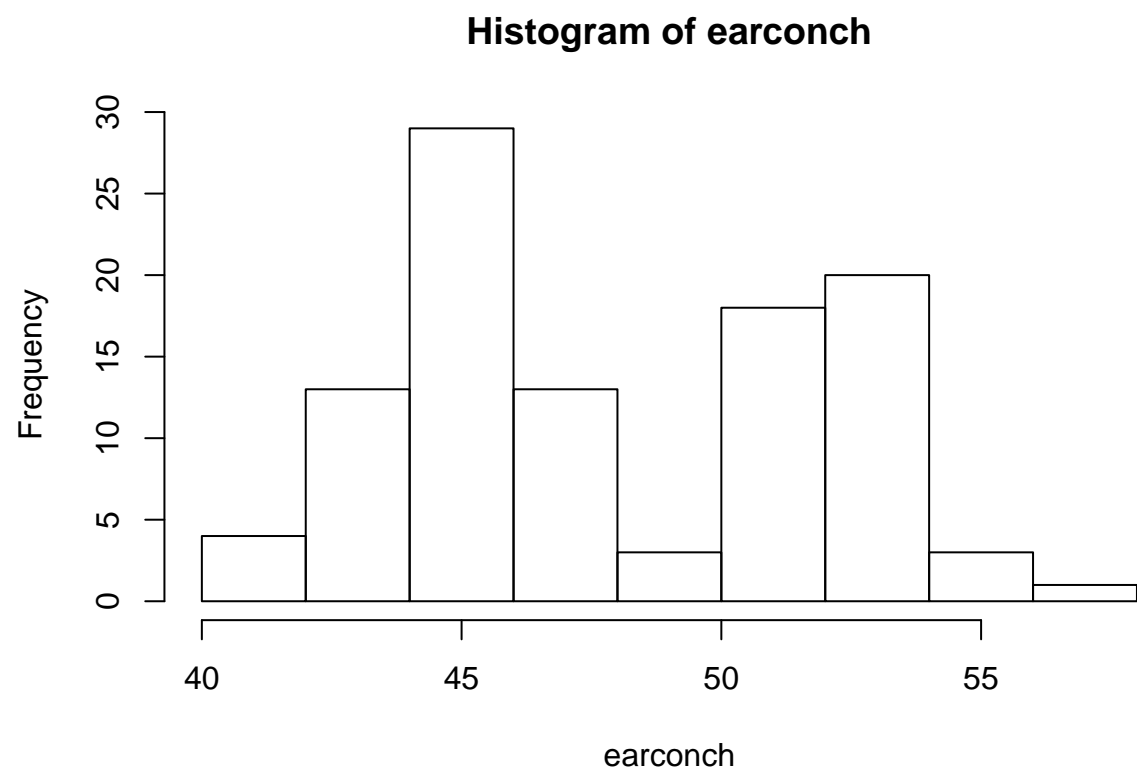


N = 102 Bandwidth = 0.6814

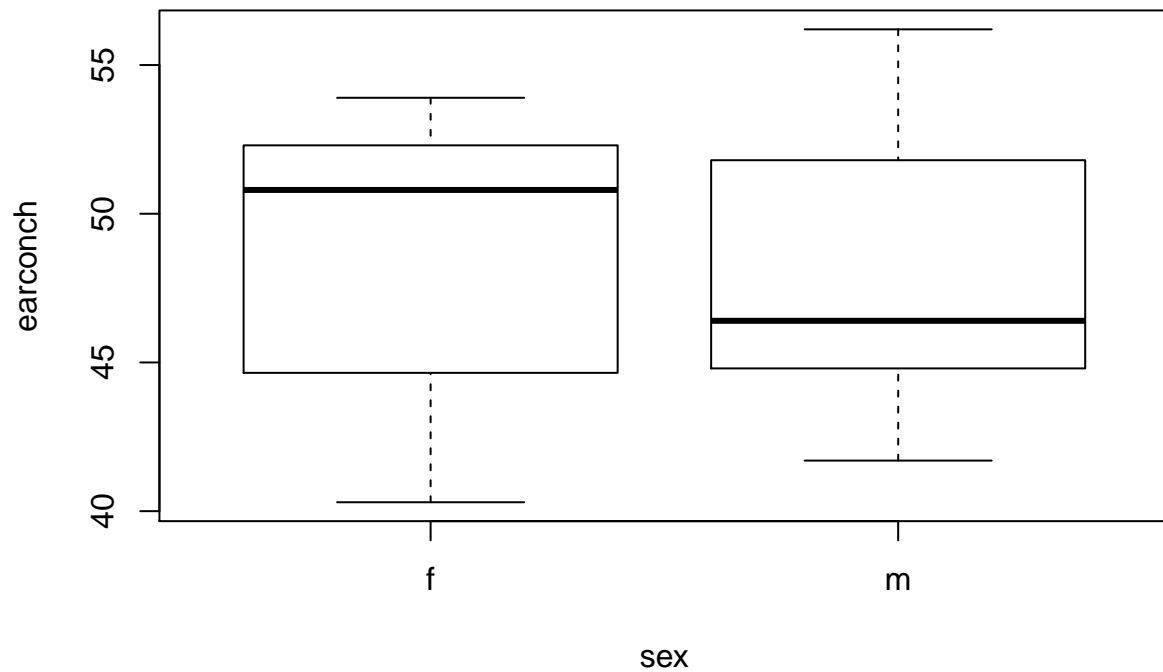
Histograms can be useful and easier to interpret discrete data and its frequency. KDE's on the other hand could come in handy when we need to visualize continuous data.

Problem 2

```
hist(earconch)
```



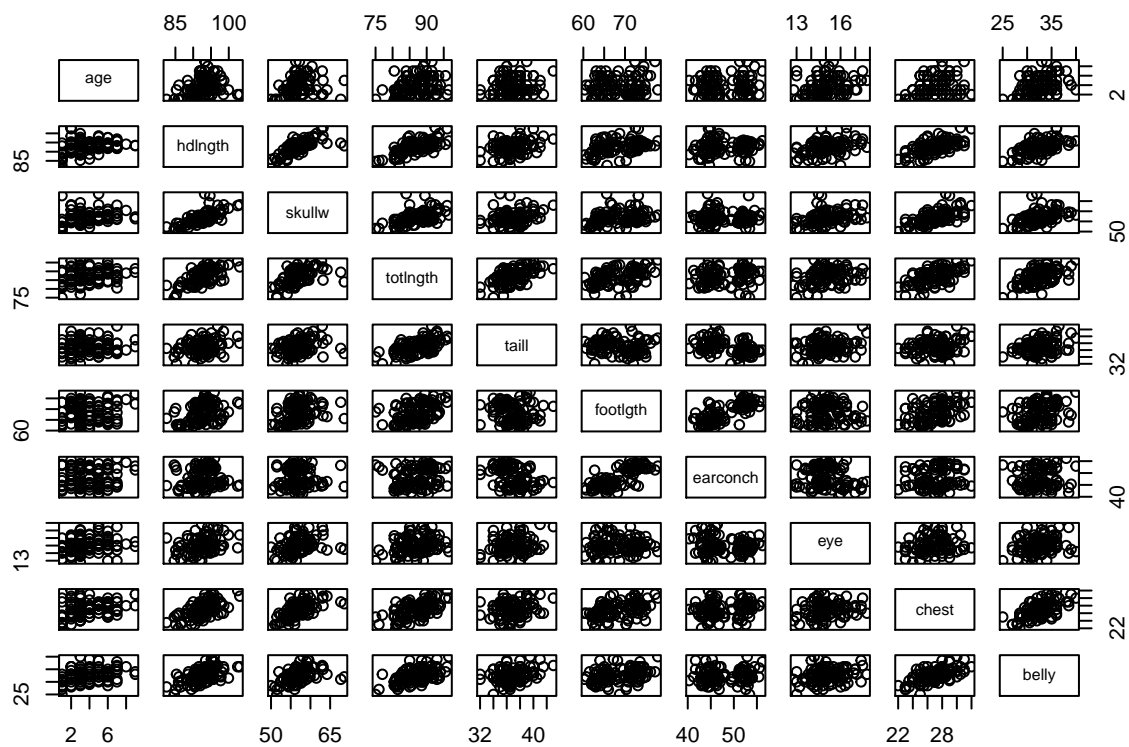
```
boxplot(earconch~sex)
```



The length of ear conch does seem to differ by sex. The median length of females is much higher than males'. Also, the males' data in the 75 percentile seems to be more spread out vs. the females' 75 percentile.

Problem 3

```
data = select(possum, -case, -site, -Pop, -sex)
plot(data)
```



```
plot(skullw, hdlngth)
points(x=mean(skullw), y=mean(hdlngth), col="red", pch=17, cex=1.5)
```

