1. Let's begin practicing real data analysis: look at, plot, and analyze the data as required below. Please describe the data, your analysis, and your findings using complete sentences. Use the data set named <u>finches</u> found on my website. All units are in millimeters.

   (a) Make and describe an appropriate plot for the a random variable of your choice.

   ```
   library(ggplot2)
   suppressMessages(library(dplyr))
   finch <- read.csv("https://roualdes.us/data/finches.csv")

   finch %>%
       ggplot(aes(beakwidth)) + geom_histogram(bins=31)
   ```
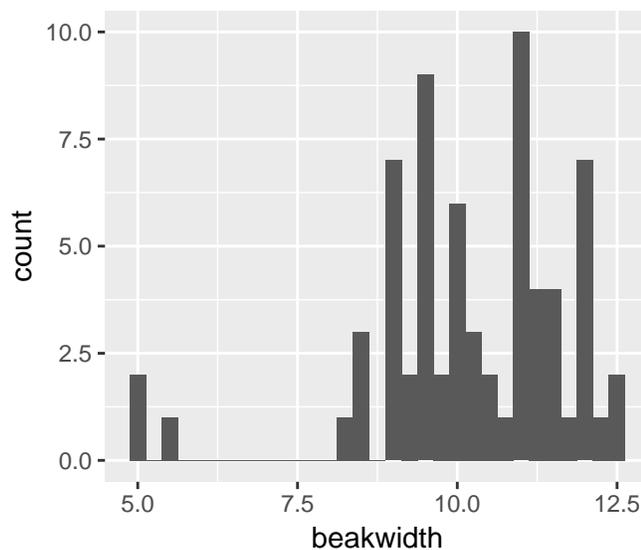

   Finches from the Galapagos islands appear to have beakwidths generally around 10 mm, save for a few observations who have a beak width of about 5. Maybe these smaller beak widths were measured on younger birds.

   (b) Calculate and interpret a 90% confidence interval for the numerical variable you chose above.

   ```
   with(finch, {
       n <- sum(!is.na(beakwidth))
       mean(beakwidth) + qt(c(0.05, 0.95), n-1) * sd(beakwidth)/sqrt(n)
   })

   ## [1]  9.859161 10.490839
   ```

   We are 90% confident that the population mean beak width for finches from the Galapagos islands is between 9.86 and 10.49 mm.

(c) Create and set up your own hypothesis test.
$H_0 : \mu = 10$ versus $H_1 : \mu \neq 0$ at $\alpha = 0.1$.

```
with(finch, {
    n <- sum(!is.na(beakwidth))
    t <- (mean(beakwidth)  - 10)/(sd(beakwidth)/sqrt(n))
    2*(1 - pt(abs(t), n-1))
})
```

```
## [1] 0.3587196
```

(d) Evaluate your hypothesis test appropriately. The p-value is larger than our level of significance, therefore we fail to reject H_0.

(e) Calculate and interpret an appropriately matching confidence interval. Did that.

(f) Does your confidence interval match the conclusion from your hypothesis test? Yes, the value from the null hypothesis is within our confidence interval, therefore both provide evidence in favor of the null hypothesis.

(g) What does the confidence interval tell you that the hypothesis test does not? The confidence interval tells us the most likely values of the true population mean of beak widths, whereas the hypothesis test only tells us that 10 is a likely value for the true mean.

(h) Describe at least three things about 14th observation in the data set.

```
finch[14,]
```

```
##        island winglength taillength beakwidth beakheight
## 14 santacruz         76       44.5      11.2         15
##    lowerbeaklength upperbeaklength tarsuslength
## 14              10            19.5         22.5
##    middletoelength
## 14            20.5
```

(i) Is the mean or median a more appropriate descriptor of your variable. Justify your answer. Because the data are not heavily skewed (despite a few potential outliers), the mean and median are equally valid here.

2. A 90% confidence interval for a population mean is given as $(17.68, 22.02)$. This confidence interval is based on a simple random sample of 36 observations. Calculate the sample mean and standard deviation. Use the t-distribution in any calculations.

```
s <- 6 * (22.02 - mean(c(17.68, 22.02)))/qt(0.95, 35)
mean(c(17.68, 22.02)) - qt(0.95, 35)*s/6
```

```
## [1] 17.68
```