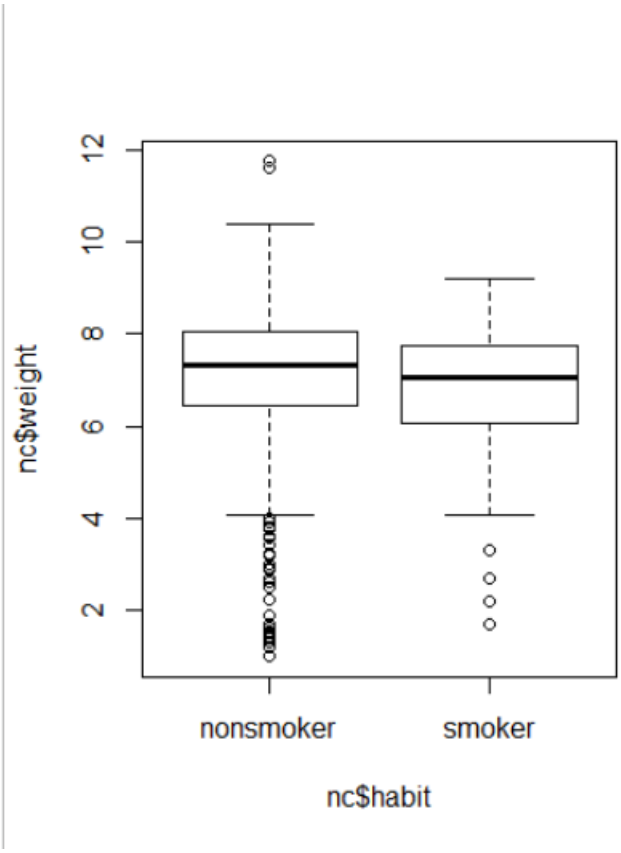
Name: Connor Lauerman Collaborators: (Any one you worked with)

R Lab 7: Inference for Numerical Data

Please answer all the Exercises and the questions from the "On Your Own" section. If you use any graphs or charts to justify your answer, please include them.

Exercise 1: What are the cases in this data set? How many cases are there in our sample?
The cases are expectant mothers and there are 1000 cases

Exercise 2: Make a side-by-side boxplot of `habit` and `weight`. What does the plot highlight about the relationship between these two variables? (Include plot.)
The mothers who smoke have on average lighter babies.



Exercise 3: Check if the conditions necessary for inference are satisfied. Note that you will need to obtain sample sizes to check the conditions. You can compute the group size using the same by command above but replacing `mean` with `length`.

We have necessary conditions since n>30 for both smokers and non-smokers

Exercise 4: Write the hypotheses for testing if the average weights of babies born to smoking and non-smoking mothers are different.
H0: The weight of babies born between non-smoking and smoking mothers are the same. u1=u2
H1:The weight of babies born to smoking mothers is less than the weight of non-smoking mothers babies u1>u2

Exercise 5: Change the `type` argument to `"ci"` to construct and record a confidence interval for the difference between the weights of babies born to smoking and non-smoking mothers.

```
95 % Confidence interval = ( 0.0534 , 0.5777 )
```

On Your Own:

1) Calculate a 95% confidence interval for the average length of pregnancies (`weeks`) and interpret it in context. Note that since you're doing inference on a single population parameter, there is no explanatory variable, so you can omit the `x` variable from the function.

   95 % Confidence interval = (38.1528 , 38.5165)
   We are 95% confident that the mean number of weeks to give birth is between 38.1528 and 38.5165

2) Calculate a new confidence interval for the same parameter at the 90% confidence level. You can change the confidence level by adding a new argument to the function: `conflevel = 0.90`.

   ```
   90 % Confidence interval = ( 38.182 , 38.4873 )
   ```

3) Conduct a hypothesis test evaluating whether the average weight gained by younger mothers is different than the average weight gained by mature mothers. (List the null and alternative hypotheses, report the test statistic and the p-value, state your decision about whether to reject the null hypothesis and state your conclusion in plain language.)

   Ho:average weight gained by younger mothers and older mothers is the same

   H1:average weight gained by younger mothers is different than mature mothers.

   ```
   Test statistic: Z =  -1.376
   p-value =  0.1686
   ```
   At a signifiganve level of 5%

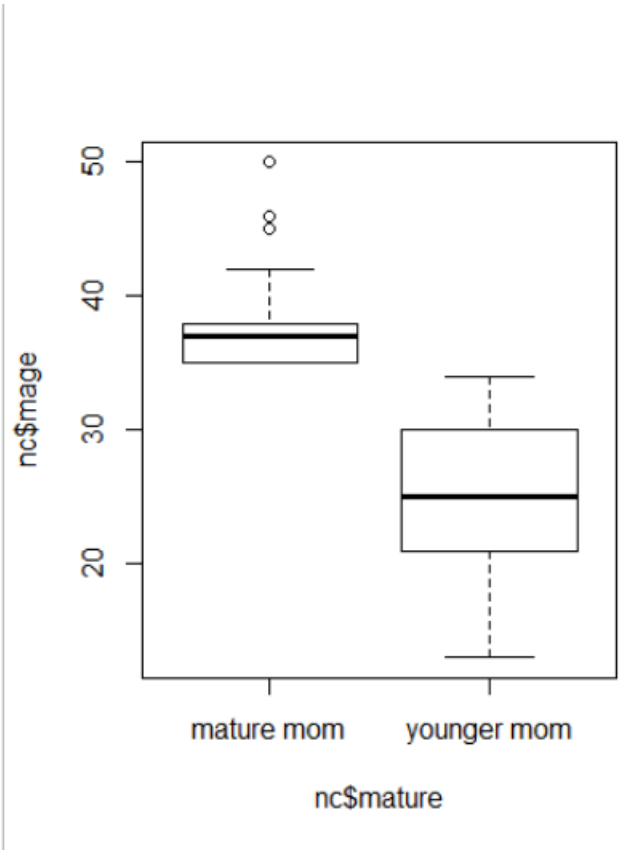
   .05<.1686

   Do not reject null hypothesis

   There is not enough evidence to conclude that the weight gained by younger and matured mothers is different

4) Now, a non-inference task: Determine the age cutoff for younger and mature mothers. Use a method of your choice, and explain how your method works.

Using the boxplot of mother age against mature and younger the cutoff can be seen to be at 35years



5) Pick a pair of numerical and categorical variables and come up with a research question evaluating the relationship between these variables. Formulate the question in a way that it can be answered using a hypothesis test and/or a confidence interval. Answer your question using the `inference` function, report the statistical results, and also provide an explanation in plain language. (List the null and alternative hypotheses, report the test statistic and the p-value, state your decision about whether to reject the null hypothesis and state your conclusion in plain language.)

Number of visits to mature.

Does being a mature mother or a young mother effect the number of visits a mother makes to the hospital?

H0: The number of visits to the hospital are the same for mature and younger mothers

H1:The number of visits to the hospital is different for mature and younger mothers

Sig level=5%

Test statistic: Z = 1.439
p-value = 0.15

.05<.15
Do not reject null hypothesis
There is not enough evidence to suggest that mature mothers and young mothers visit the hospital a different number of times