Lab 2: Analyzing Blackfoot River Fish
(25 points – Due Wednesday, June 10th)

From Canvas, download the dataset Fish.csv, and read it into R.

For this project, Montana FWP researchers used electrofishing equipment to attract fish to a boat, and then dipped them out of the water with nets, measured length in mm and weight in grams. They are often working in cold conditions in late autumn or early spring, so some measurement error is expected. Also, because this is real data, it might be a bit messy.

These data are not technically from a random sample. The goal is to catch all fish within a reach or section of the Blackfoot River every few years to assess the health of the population. Changes over years are important to the biologists.

Answer each question below using R (no Excel, other than as a check). Answers without supporting R code will not receive full credit. When you are finished, turn in both your write-up and your R script file in Canvas.

1. *[1pt]* How many total fish are in this dataset?

2. *[1pt]* How many different species are represented in this dataset?

3. *[2pts]* What are the names of the two river sections where the fish were collected?

4. *[2pts]* How many fish had either 0 or "NA" recorded for weight?

5. *[2pts]* Remove these 0 and "NA" weight fish from the dataset with the `subset()` function. (***Once this is complete, use this new dataset for the remainder of the lab)***

6. *[3pts]* How does the average weight of fish vary by time and species?

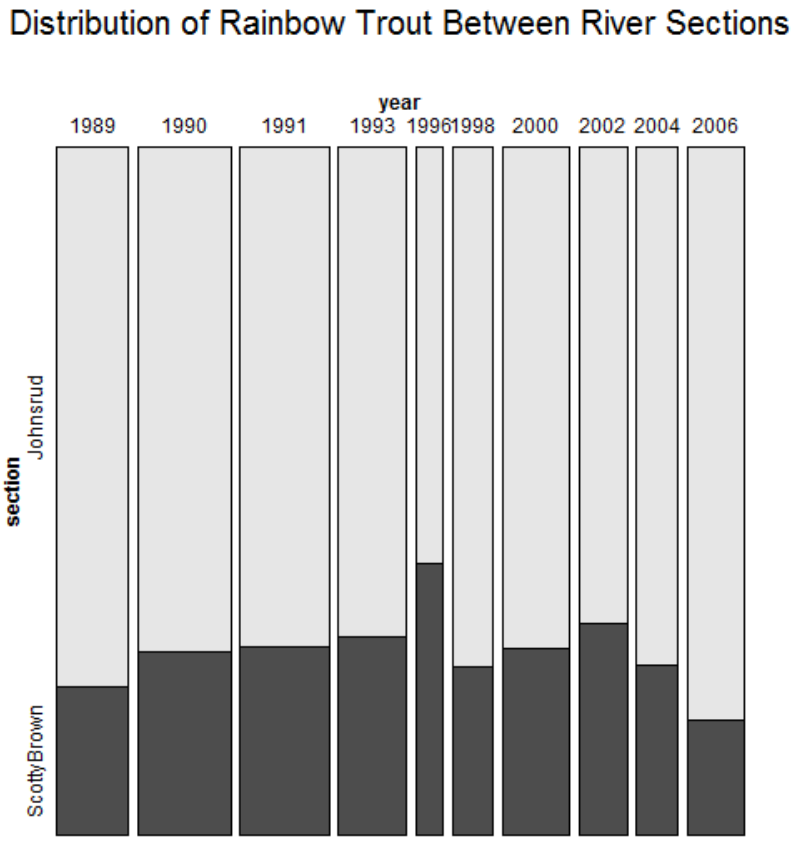
   Using your newly created subset, find the mean weight of fish by species, by year (Hint: try reading the help file on the `aggregate()` function. When finished, you should have an average for each of the species/year combinations).

7. (Subset your data from #5 to just Rainbow Trout. Name your new dataset RBT) Is the length of Rainbow Trout approximately normally distributed?

[3pts] Make and interpret a QQplot of the length of Rainbow Trout. That is, justify your comments about normality of the data by referring to your QQplot.

8. How are Rainbow Trout distributed between river sections?

   a. *[3pts]* Recreate the following plot including the title and axis labels. The shading doesn't need to match exactly so long as each section is a different color.
   (Hint: there are two similar commands to do this: `mosaic()` and `mosaicplot()`. I built mine using `mosaic()` which requires you to install the vcd package. To do this, go to Tools→Install Packages. Once installed, you will need to submit `require(vcd)` in the console window)



   b. *[1pt]* Which section of river typically supports a higher percentage of the Rainbow Trout population?

   c. *[2pts]* Looking at your mosaic plot, explain why the bar for 1996 is so much skinnier than the bars for 1991 or 1993. (Hint: try running `table(RBT$year)`)

9. For this particular habitat, as long as ScottyBrown can support at least 17% of the Rainbow Trout population, the ecosystem is considered healthy and biologists write off any change from year to year to a combination of sample variation and a natural

rise and fall of the fish population.

Back in 2006, however, some biologists were sounding alarm bells as the percent of Rainbow Trout in ScottyBrown fell to 16.7%. Others argued that, given the sample size, this proportion is not statistically different than the target of 17% and that they should wait to see the results of next year's sample before jumping to any conclusions.

[5pts] Conduct a hypothesis test to either support or refute the claim made by this second group of biologists. That is, conduct a hypothesis test for
H_0: P(RBT in $\text{ScottyBrown}_{2006}$) = 0.17 against the alternative that is in fact less. Report your test statistic and p-value and give a conclusion in the context of the problem.
(Hint: return to your data from #7 and subset it to just the year 2006)