Statistical Computing Tasks

We have data on fish caught in the Blackfoot River by Fish, Wildlife, & Parks personnel over a number of years. They used electrofishing equipment to attract the fish to the boat, then dipped them out of the water with nets, measured length in cm and weight in grams. They are often working in cold conditions in late autumn or early spring, so some measurement error is expected.

These data are not 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.

The data were collected by making two trips per section (Johnsrud or Scotty Brown) each sampling year. The fish caught each trip of a given year, had their weight, length, and species recorded.

```
head(blackfoot)
```

```
##   trip length weight year  section species
## 1    1    288    175 1989 Johnsrud     RBT
## 2    1    288    190 1989 Johnsrud     RBT
## 3    1    285    245 1989 Johnsrud     RBT
## 4    1    322    275 1989 Johnsrud     RBT
## 5    1    312    300 1989 Johnsrud     RBT
## 6    1    363    380 1989 Johnsrud     RBT
```

```
summary(blackfoot)
```

```
##       trip          length        weight          year
##  Min.   :1.0   Min.   : 16   Min.   :   0   Min.   :1989
##  1st Qu.:1.0   1st Qu.:186   1st Qu.:  65   1st Qu.:1991
##  Median :2.0   Median :250   Median : 150   Median :1996
##  Mean   :1.5   Mean   :262   Mean   : 246   Mean   :1997
##  3rd Qu.:2.0   3rd Qu.:330   3rd Qu.: 330   3rd Qu.:2002
##  Max.   :2.0   Max.   :986   Max.   :4677   Max.   :2006
##                              NA's   :1796
##    section            species
##  Length:18352       Length:18352
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

```
str(blackfoot)
```

```
## 'data.frame':    18352 obs. of  6 variables:
##  $ trip   : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ length : num  288 288 285 322 312 363 269 160 213 157 ...
##  $ weight : num  175 190 245 275 300 380 170 40 80 35 ...
##  $ year   : int  1989 1989 1989 1989 1989 1989 1989 1989 1989 1989 ...
##  $ section: chr  "Johnsrud" "Johnsrud" "Johnsrud" "Johnsrud" ...
##  $ species: chr  "RBT" "RBT" "RBT" "RBT" ...
```

What type of variable did `R` store species and section as? How would you change species and section to categorical variables?

If the researchers were only interested in Rainbow trout and Brown trout, how would you remove Bull trout and WCT (whitefish) from the data set?

Sometimes when sampling the fish, a technician fails to record one of the variables. How would you remove all the fish with missing values? How would this change if you instead removed the fish with only missing weight?

The sampling methods used by Fish, Wildlife, & Parks on the Blackfoot River has changed over the years. In the years 1989 - 1996 they used gill nets and since 1996 they have used electrofishing. How would you create a new variable named `method` to reflect these different sampling methods used over the years?

The researchers are interested in how many fish are caught each year that weigh over 1500 grams. How would you find these numbers to report?

Which pairs of (weight, length) combinations seem difficult to believe? One way to look for unusual pairs is to use what fisheries biologists call a "condition index" $\frac{w^{1/3}}{l} * 50$ **where** $w =$ **weight and** $l =$ **length of the fish. If fish are highly unusual in this scale, it would be best to remove them, but you might need to compare only within species.**

- **How would you calculate each trout's condition number?**
- **How would you summarize these condition numbers for each of the two species of trout (Rainbow and Brown)?**
- **How would you plot the condition numbers of each trout, making sure to differentiate between Rainbow and Brown trout?**

The researchers are interested in trends in fish size over the sampling period (1989-2006). How would you create a visualization of fish lengths over the sampling period?

Researchers are also interested in the number of fish from each species caught each year. How would you create a visualization of the number of fish caught from each species over the sampling period?

Lastly, the researchers are interested in trends in average fish weight over the sampling period. They want you to create a visualization of the average fish weight across years, differentiated by species of trout.

- **First, you need to create a data frame of the mean weight of fish caught each year for the two species of trout. The end product should look something like the data frame below. How would you create this data frame of mean weights?**

```
##   year species mean
## 1 1989   Brown  297
## 2 1989    Bull  429
## 3 1989     RBT  101
## 4 1989     WCT  120
## 5 1990   Brown  379
## 6 1990    Bull  422
```

- **Next, to plot these mean weights for each year you need to transform the data from the current long format to wide format. This process is done by spreading the `year` variable across 10 different columns, one for each year (1989, 1990, etc.). The end product should look something like the data frame below. How would you transform these data from long format to wide format?**

```
##   species 1989 1990 1991 1993 1996 1998 2000 2002 2004 2006
## 1   Brown  297  379  434  391  571  543  407  530  419  325
## 2     RBT  101  141  186  208  244  156  179  320  215  173
```

There are additional data about the sections of the Blackfoot river for the sampling days each year. Researchers wish to merge these data (shown below) with the data on the fish caught during the sampling period. The `year`, `trip`, and `section` variables are keys that connect the two data sets. How would you merge these two data sets together?

```
head(water)
```

```
##   trip year      section temp water_level
## 1    1 1989 Scotty Brown 73.5        3.69
## 2    2 1989     Johnsrud 58.4        3.91
## 3    1 1990 Scotty Brown 53.4        3.65
## 4    2 1990     Johnsrud 60.0        3.95
## 5    1 1991 Scotty Brown 48.0        4.03
## 6    2 1991     Johnsrud 65.6        3.84
```