FISH 549 - Assignment 5

Creating an expository figure

Marie Zahn

2023-02-11

Background

In this assignment, I am using data from the {FSAdata} package that contains age, sex, length, and mass data for Siscowet Lake Trout captured at four locations in Lake Superior.

Objective

I'm interested in seeing how total length (mm) and weight (g) of fish vary among the four sampling locations: Blind Sucker, Deer Park, Grand Marais, Little Lake Harbor. First, I will explore the data by plotting weight vs. length with observations colored by location. Then I will run a linear regression to analyze how these variables differ across the four sites.

1. Create exploratory figure

```
## load all packages needed for analysis/plotting
library(here)
```

```
## here() starts at C:/Users/marie/Documents/Coursework/Data_Science/Assignment-5
```

```
library(ggplot2)
library(tidyverse)
```

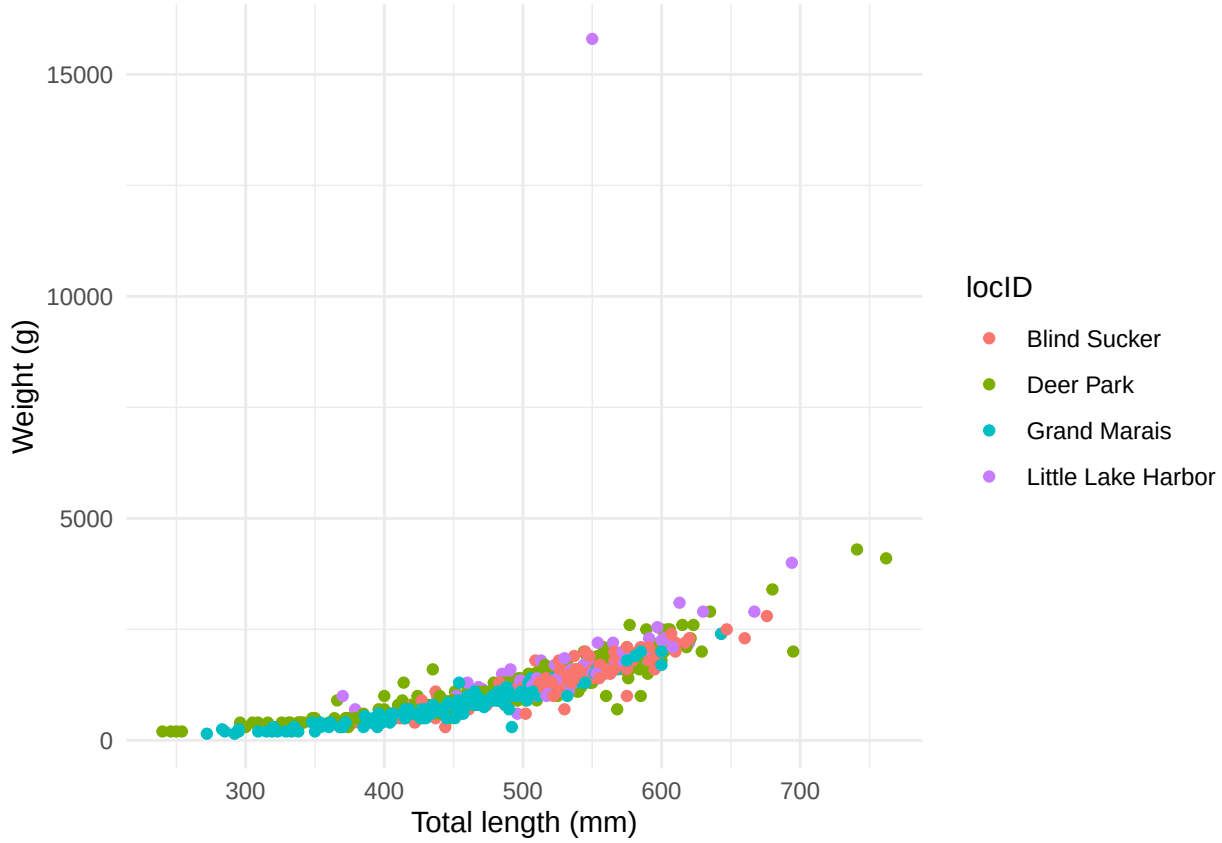
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2
## --

## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## v purrr   1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
trout_df <- read.csv(here('data/siscowet.csv'))

## inspect the data
head(trout_df)
```

```
##       locID pnldep mesh fishID  sex age len  wgt
## 1 Deer Park  36.74  2.5  19108 <NA>  NA 316  400
## 2 Deer Park  40.09  3.0  19109 <NA>  NA 396  700
## 3 Deer Park  41.46  5.0  19110    M  NA 590 1800
## 4 Deer Park  41.46  5.0  19111    M  NA 516 1500
## 5 Deer Park  43.45  5.5  19112 <NA>  NA 414  800
## 6 Deer Park  45.58  4.0  19113    M  NA 481 1000
```

```
## make simple plot showing length vs. weight and coloring by location
trout_df %>% ggplot(aes(x=len, y=wgt, color=locID)) +
  geom_point() +
  theme_minimal() +
  ylab("Weight (g)") +
  xlab("Total length (mm)")
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```

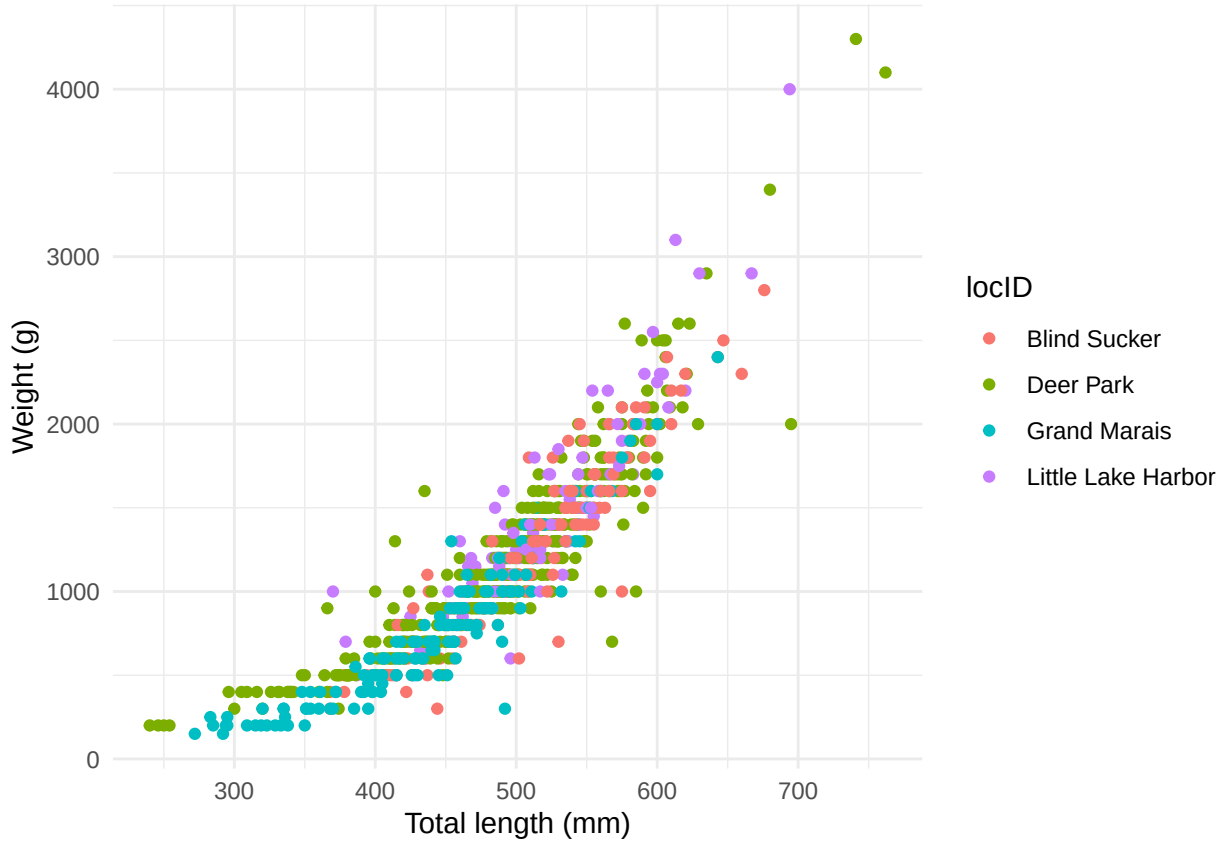


Based on this exploratory figure, we can see there are likely differences between sites but it's difficult to tell because the points are overlapping and there is one outlier point from Little Lake Harbor. This point has an exceedingly high weight measurement that is probably erroneous; I would go back to the original data to ensure this observation was correct. According to this record, the fish was ~33 lbs and ~22 inches - that's one FAT fish!

2. Create expository figure

Here I remove the outlier point, plot the data again (faceted by location) and added a model fit.

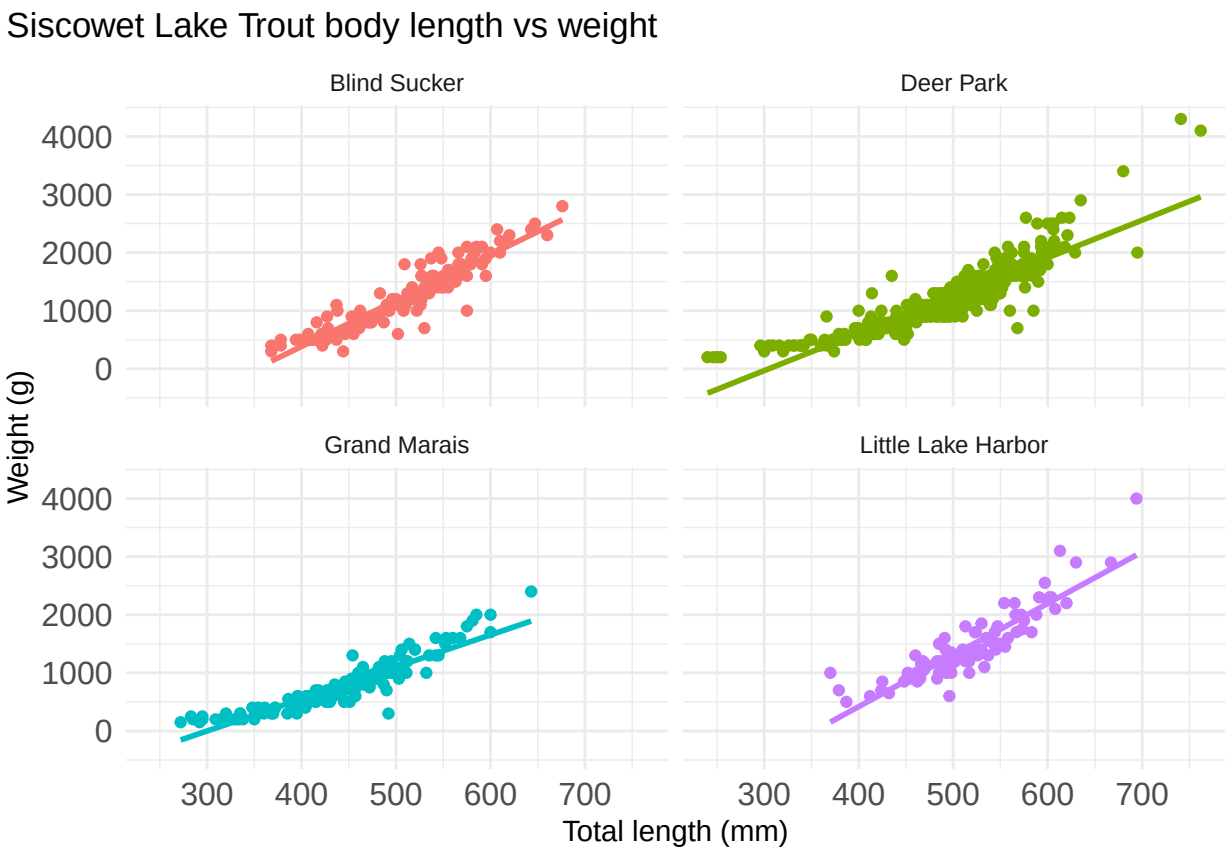
```
## identify and remove point with erroneous weight measurement
trout_filter_df <- trout_df %>% filter(wgt < 10000)

## plot to make sure this filter is correct
## make simple plot showing length vs. weight and coloring by location
trout_filter_df %>% ggplot(aes(x=len, y=wgt, color=locID)) +
  geom_point() +
  theme_minimal() +
  ylab("Weight (g)") +
  xlab("Total length (mm)")
```



Looks good so now we can make the final figure:

```
trout_filter_df %>% ggplot(aes(x=len, y=wgt, color=locID)) +
  geom_point() +
  theme_minimal() +
  ylab("Weight (g)") +
  xlab("Total length (mm)") +
  facet_wrap(~locID) +
  labs(color="Location",
       title = "Siscowet Lake Trout body length vs weight") +
  theme(
    plot.title.position = "plot",
    axis.text = element_text(size=12),
    legend.position = "none") +
  geom_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



As expected, there is a positive relationship between body length and weight across all four sites. The slope does vary between the four locations, with Grand Marais having the lowest slope and generally the smallest fish.