Assignment-5

Sara Faiad

2/17/2021

Assignment 5 - Creating an expository figure

Getting the Data

I created a data folder locally and downloaded "siscowet.csv".

```
# Reading in the data

siscowet<- read.csv("data/siscowet.csv")

# Checking things out

head(siscowet)
```

```
##         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
```

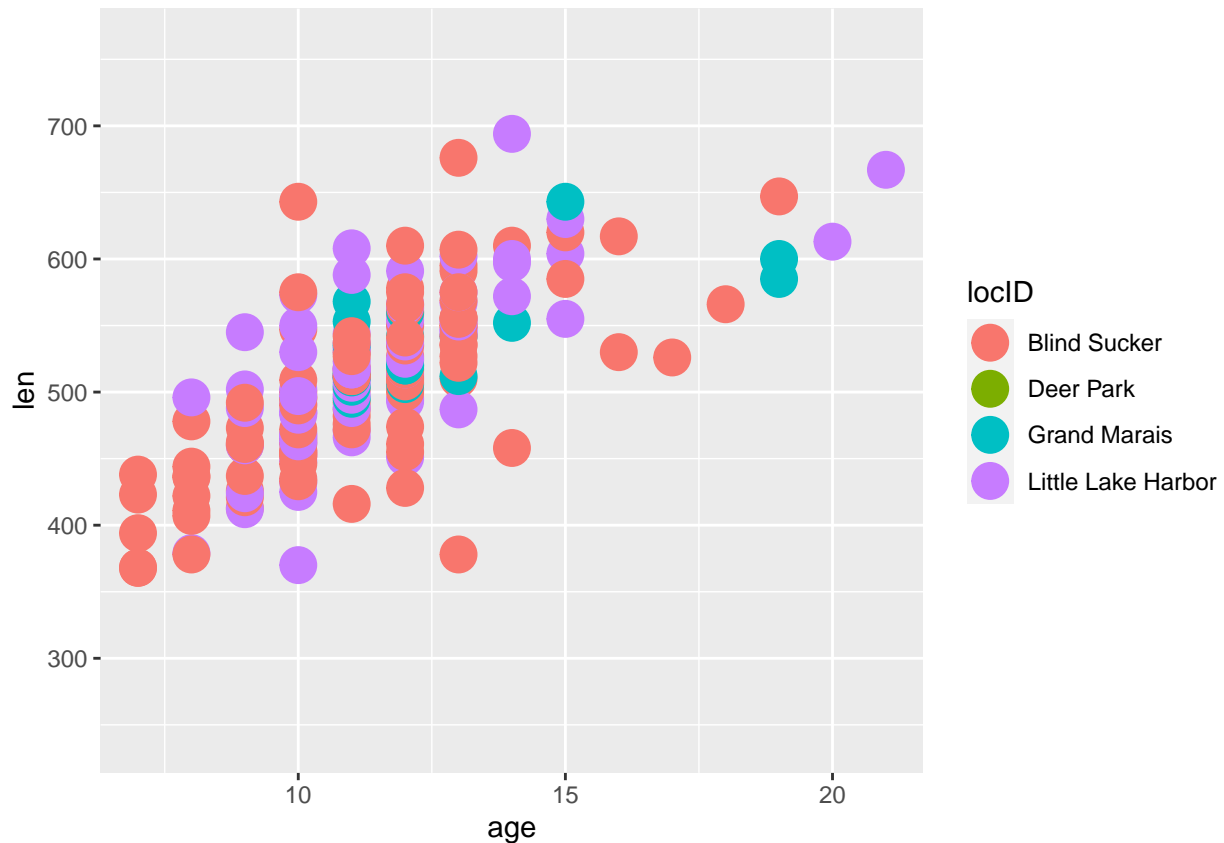### The Plot - Exploratory

```
# load ggplot2

library(ggplot2)


# A basic scatterplot showing the relationship between len and age. Points are colored depending on loc.

ggplot(data = siscowet, aes(x=age, y=len, color=locID)) +
    geom_point(size=6)
```

```
## Warning: Removed 580 rows containing missing values (geom_point).
```

That looks pretty bad, but it's a start. Time to refine.

Refining the Plot - Expository 1.0

```r
#PNW color pallette (because I love living in the PNW)

library(PNWColors)

#Building a scatterplot

siscowet_expository_plot<- ggplot(data = siscowet,
       aes(x = age,
       y= len)) +
  geom_point(aes(color = locID,
                shape = locID,
                alpha = locID),
            size = 3) +
  scale_color_manual(values = pnw_palette("Cascades",4)) +
  labs(title = "Siscowet Lake Trout from Michigan waters of Lake Superior",
       subtitle = "Length and weight from four locations: Blind Sucker, Deer Park, Grand Marais, and  Li
       x = "Assigned Ages (years)",
       y = "Total Length (mm)",
       color = "Location ID",
       shape = "Location ID",
```
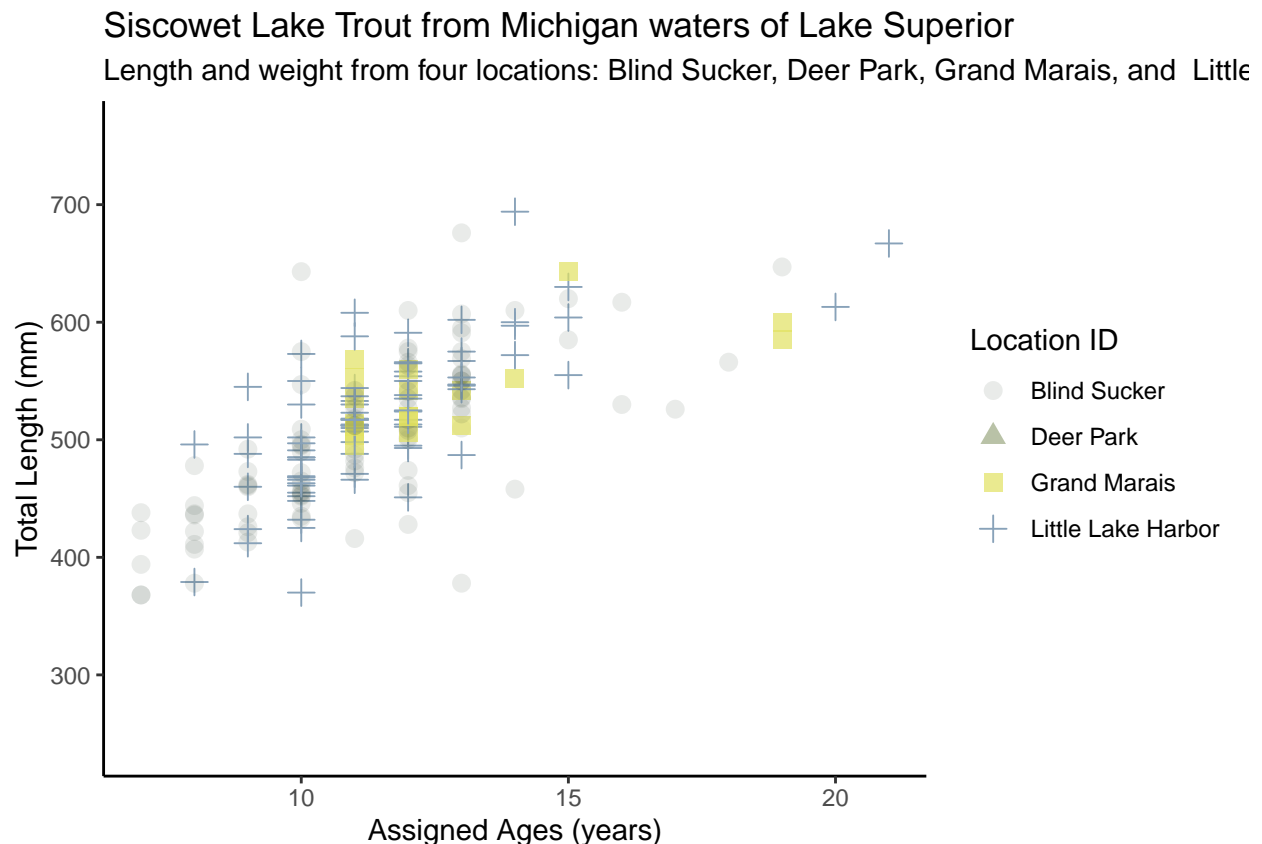
```
        alpha = "Location ID") +
  theme_classic()

siscowet_expository_plot
```

```
## Warning: Using alpha for a discrete variable is not advised.
```

```
## Warning: Removed 580 rows containing missing values (geom_point).
```


Better, but still not great. Many of the points overlap and adding some transperancy to the points doesn't really help matters. Additionally, the colors are not colorblind friendly. I'm going to switch up tactics and attempt a violin plot.

Violin Plot - Expository 2.0

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

ggplot(data = siscowet,
       aes(fill = sex, y = len, x = locID)) +
  geom_violin(position="dodge") +
  labs(title="Siscowet Lake Trout from Michigan waters of Lake Superior",
       x="Location ID",
       y = "Total Length (mm)",
       color = "Sex") +
  scale_fill_manual(values = pnw_palette("Cascades",3)) +
  theme_classic()
```
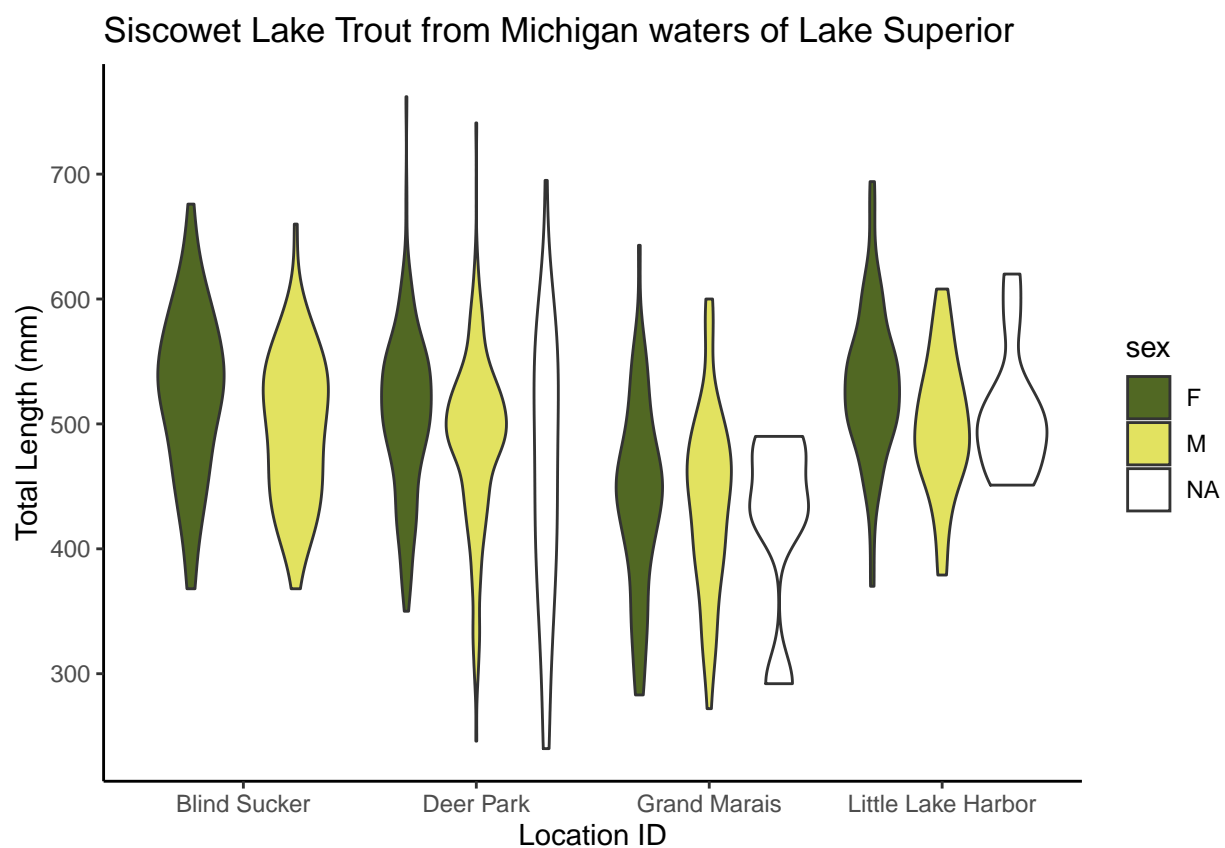

I wouldn't say this is the most beautiful thing that I have ever seen, but I think it's pretty good, especially for never making a violin plot before! The shape of the violin represents the total length of trout (mm), separated by sex, from each location. Sex is also reflected by the color of each violin. Green denotes females, yellow males, and white for NAs. The colors still appear distinct on the Color Blindness Simulator.