Assignment 5

Practice in data visualization

Jenny Stern

Creating an exploratory figure

I was interested in exploring the relationship between fish sampling location, length, and sex. Because I wanted to see how the distribution of lengths varied among sites and sex, I relied on two different density plot types - violin and rideline - to explore the data:

```r
## packages
library(ggplot2)
library(ggridges)

## load the data
laketrout <- read.csv("/Users/jhstern/Documents/Assignment-5/data/siscowet.csv")
summary(laketrout)
```

```
##     locID              pnldep            mesh          fishID
##  Length:780         Min.   : 15.40   Min.   :2.000   Min.   :19108
##  Class :character   1st Qu.: 45.20   1st Qu.:2.500   1st Qu.:19362
##  Mode  :character   Median : 59.60   Median :3.500   Median :19558
##                     Mean   : 56.23   Mean   :3.576   Mean   :19576
##                     3rd Qu.: 69.05   3rd Qu.:4.500   3rd Qu.:19816
##                     Max.   :108.69   Max.   :6.000   Max.   :20053
##
##      sex                 age             len             wgt
##  Length:780         Min.   : 7.00   Min.   :240.0   Min.   :  150
##  Class :character   1st Qu.:10.00   1st Qu.:443.0   1st Qu.:  775
##  Mode  :character   Median :11.00   Median :493.0   Median : 1100
##                     Mean   :11.45   Mean   :487.1   Mean   : 1175
##                     3rd Qu.:12.25   3rd Qu.:536.2   3rd Qu.: 1500
##                     Max.   :21.00   Max.   :762.0   Max.   :15800
##                     NA's   :580                     NA's   :1
```

```r
## remove NAs
laketrout <- laketrout[!is.na(laketrout$sex),]
summary(laketrout) #780 observations to 721 for sex
```

```
##     locID              pnldep            mesh          fishID
##  Length:721         Min.   : 17.99   Min.   :2.000   Min.   :19110
##  Class :character   1st Qu.: 46.80   1st Qu.:2.500   1st Qu.:19379
##  Mode  :character   Median : 60.37   Median :3.500   Median :19560
##                     Mean   : 57.64   Mean   :3.587   Mean   :19589
##                     3rd Qu.: 69.21   3rd Qu.:4.500   3rd Qu.:19830
```

1

```
##                       Max.    :108.69   Max.    :6.000   Max.     :20053
##
##      sex                 age             len              wgt
##   Length:721        Min.    : 7.00   Min.    :246.0   Min.     :  150
##   Class :character   1st Qu.:10.00   1st Qu.:446.0   1st Qu.:  800
##   Mode  :character   Median :11.00   Median :493.0   Median : 1100
##                       Mean    :11.44   Mean    :488.9   Mean     : 1184
##                       3rd Qu.:12.50   3rd Qu.:536.0   3rd Qu.: 1500
##                       Max.    :21.00   Max.    :762.0   Max.     :15800
##                       NA's   :522                      NA's    :1
```
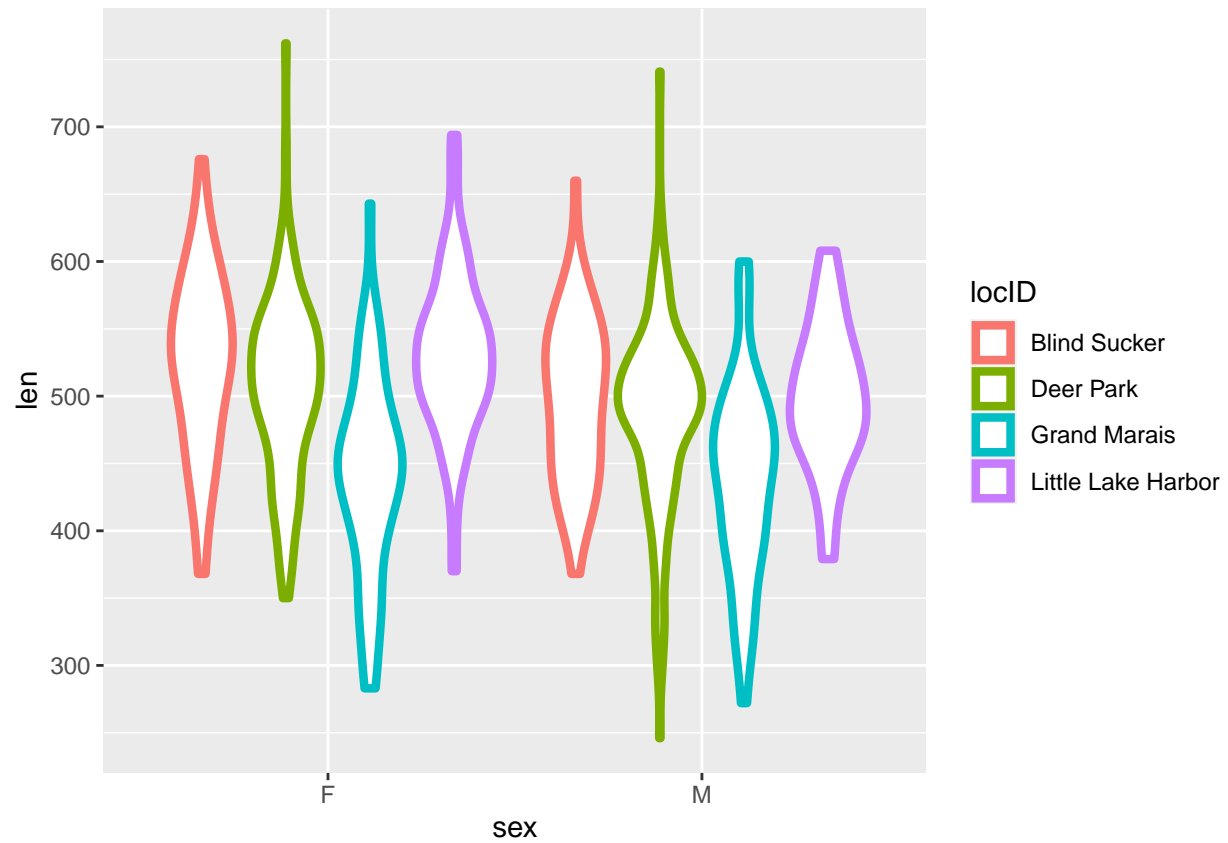
```r
## violin plots

loc_len_violin <- ggplot(laketrout, aes(x = locID, y = len, color = sex)) + geom_violin(size = 1.5)
loc_len_violin
```
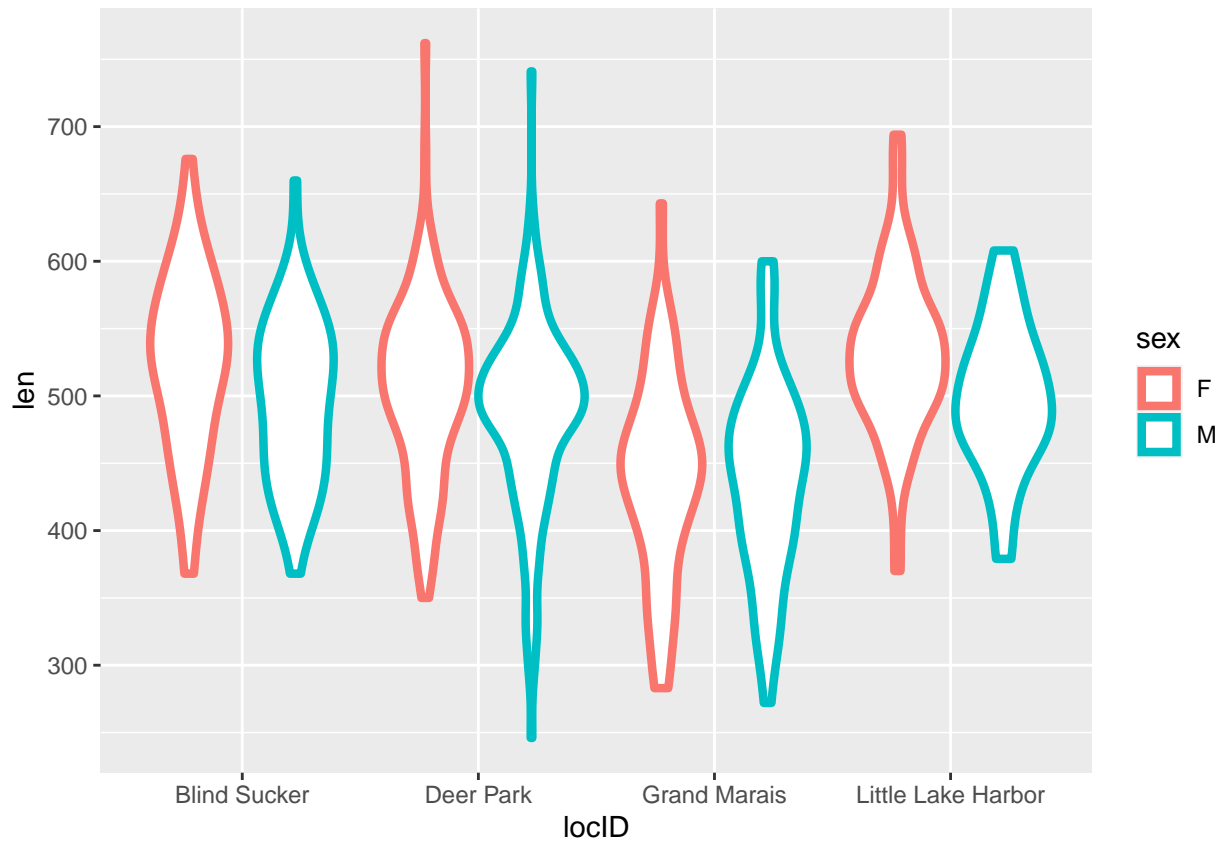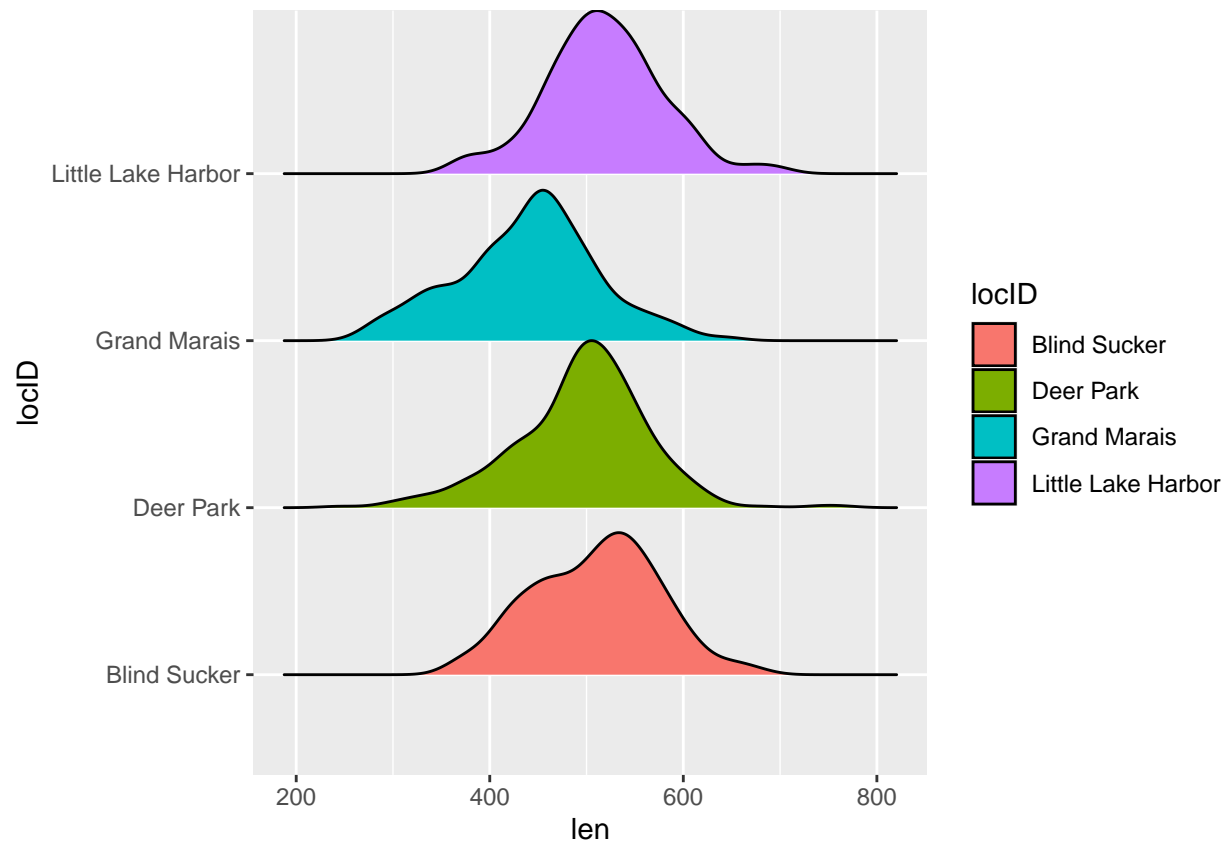


```r
sex_len_violin <- ggplot(laketrout, aes(x = sex, y = len, color = locID)) + geom_violin(size = 1.5)
sex_len_violin
```
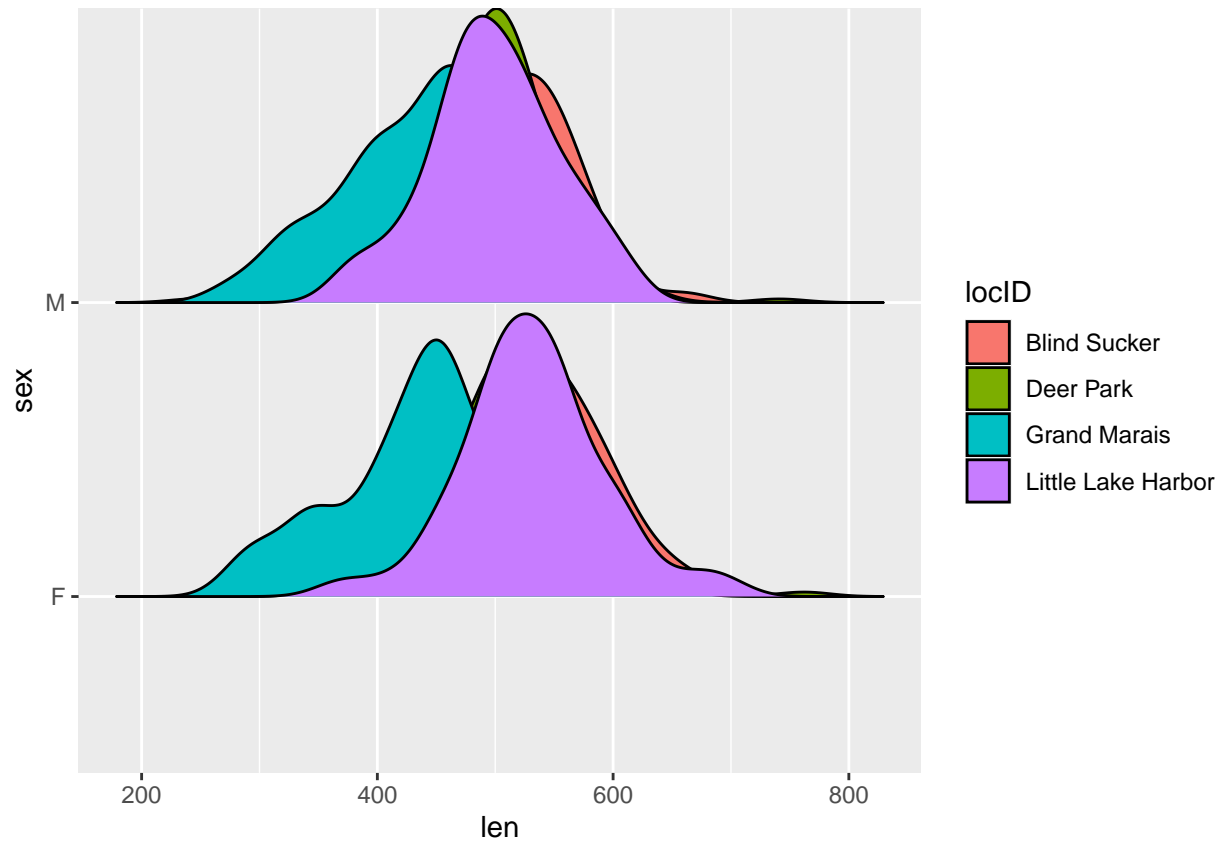
```
## ridgelines plots
len_loc_ridge <- ggplot(laketrout, aes(x = len , y = locID,  fill = locID)) +
  ggridges::geom_density_ridges(scale = 1)
len_loc_ridge
```

```
## Picking joint bandwidth of 19.5
```
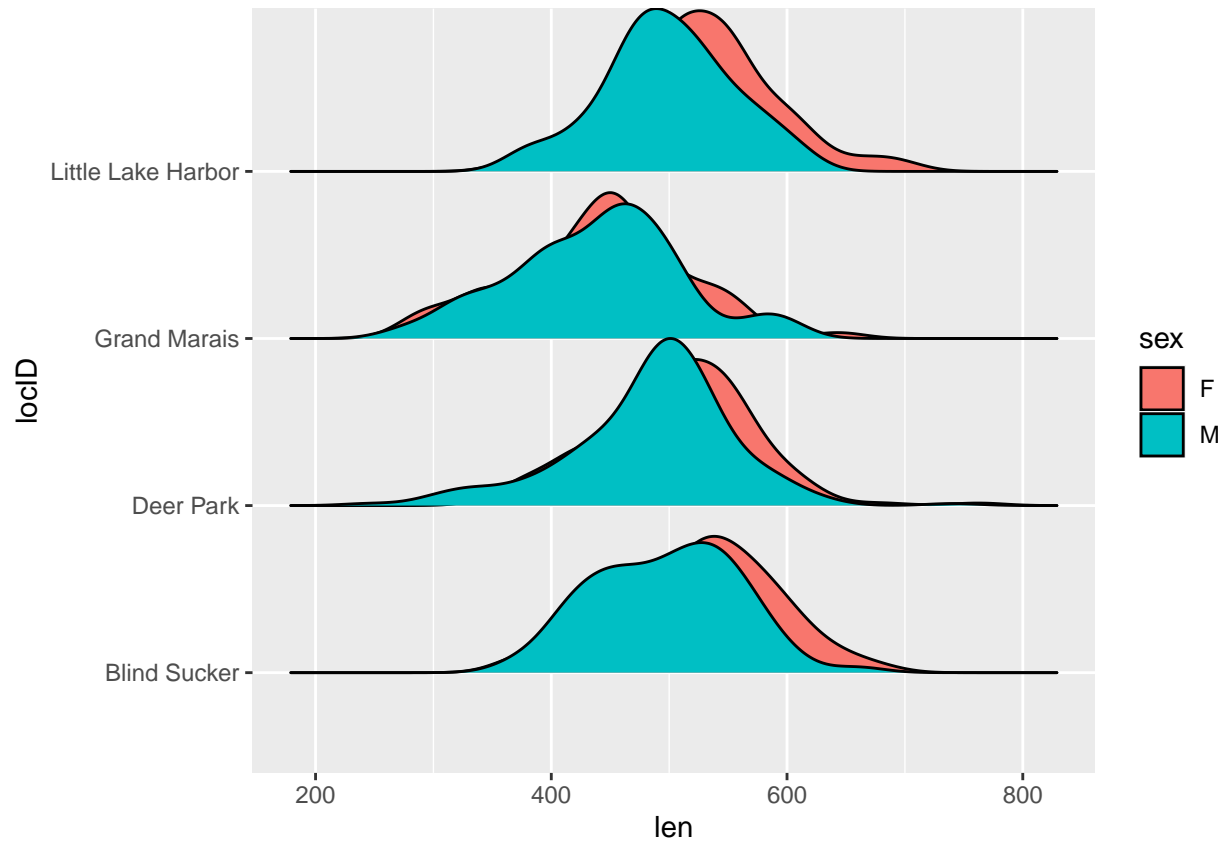
```
len_sex_ridge_loc <- ggplot(laketrout, aes(x = len , y = sex,  fill = locID)) +
  ggridges::geom_density_ridges(scale = 1)
len_sex_ridge_loc
```

```
## Picking joint bandwidth of 22.3
```

```
len_loc_ridge_sex <- ggplot(laketrout, aes(x = len , y = locID,  fill = sex)) +
  ggridges::geom_density_ridges(scale = 1)
len_loc_ridge_sex
```

```
## Picking joint bandwidth of 22.3
```

I have decided to build on the exploratory figure above because I have not made a ridgeline plot before and would like to practice. I also thought this plot did a nice job of showing how length distributions varied among sites and sexes.

Creating an expository figure

```
## loading additional packages
library(ggtext)

## defining my base plot, changing the scale (overlap) of the figure, and making the outline of the den
pp <- ggplot(laketrout, aes(x = len , y = locID,  fill = sex, color = sex)) +
  ggridges::geom_density_ridges(scale = 1.2, alpha = 0.4)
pp
```

```
## Picking joint bandwidth of 22.3
```
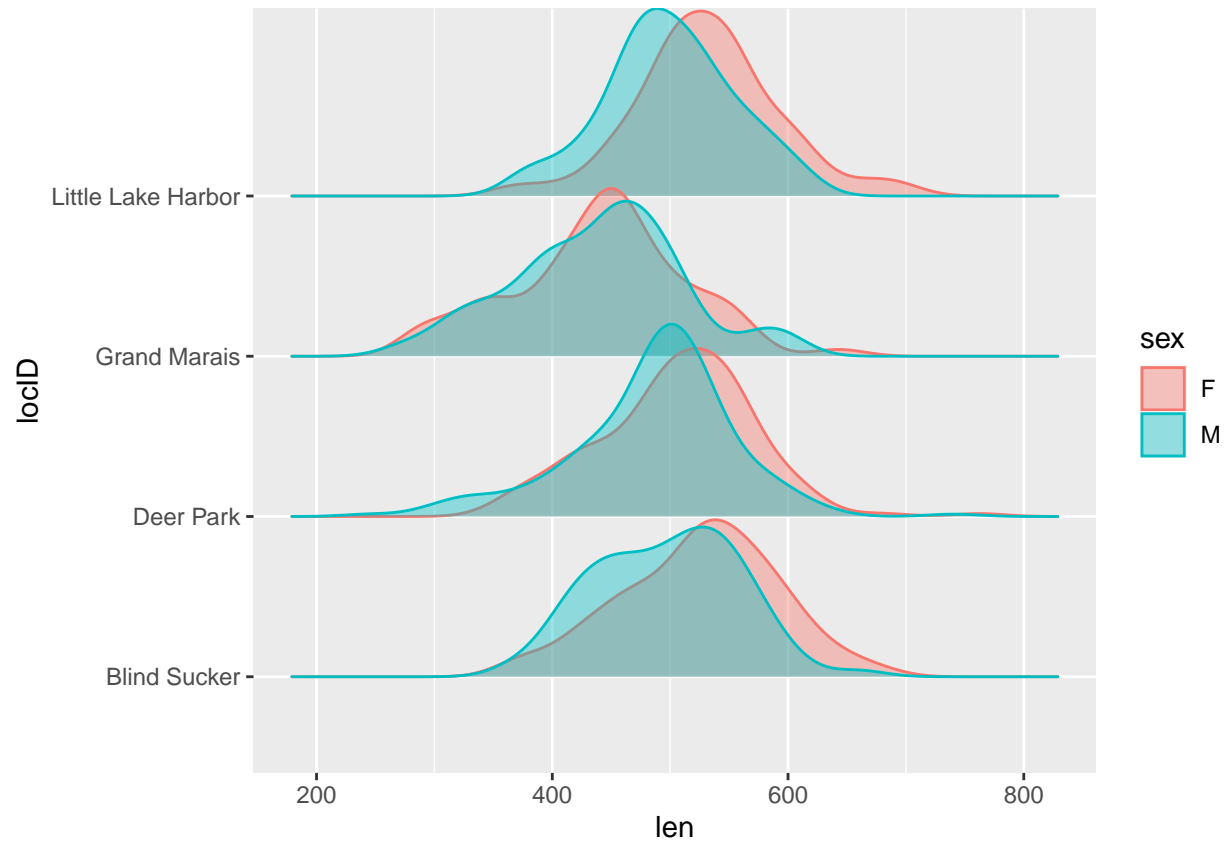
```
## changing axis labels and removing grey background & gridlines
p1 <- pp +
  scale_x_continuous(expand = c(0, 0), name = " Total length [mm]") +
  scale_y_discrete(expand = c(0, 0), name = "") + theme_ridges(grid = FALSE, center = TRUE)
p1
```

```
## Picking joint bandwidth of 22.3
```
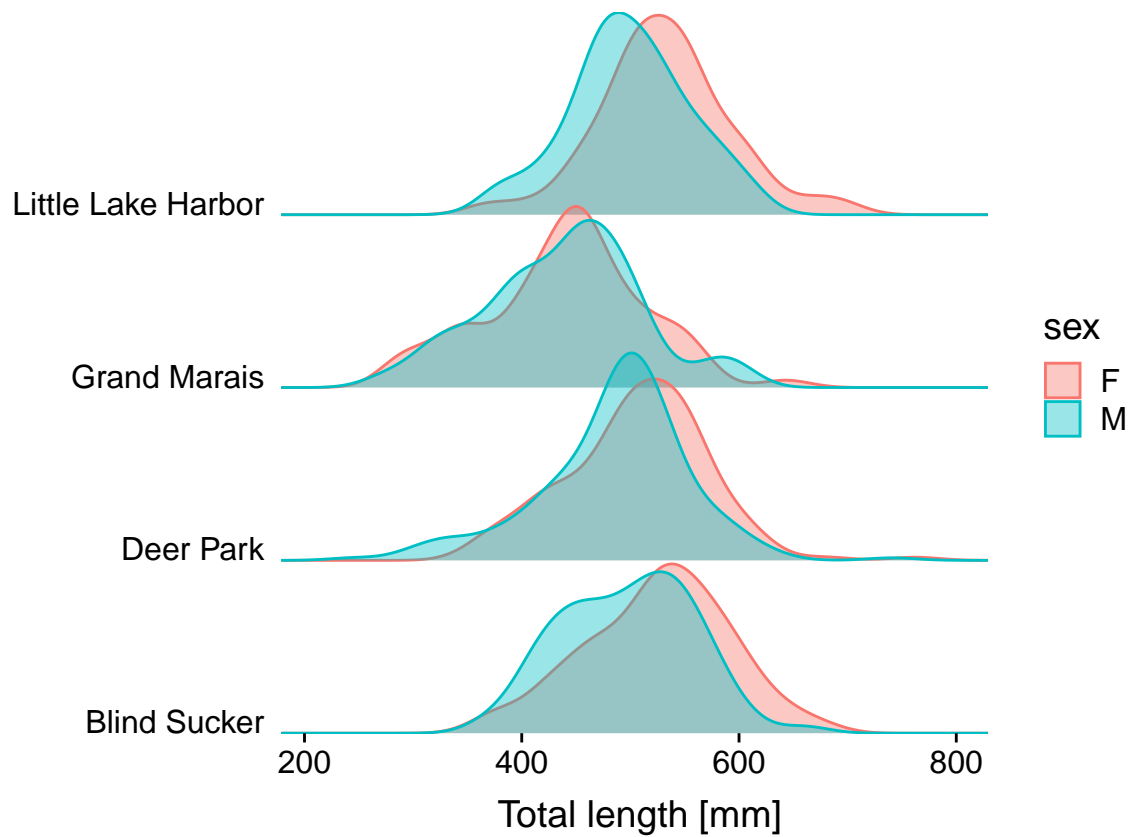
```
## using color brewer, changing the colors of male & female to be less stereotypical and more colorblin
 p2 <- p1 + scale_fill_manual(values = c("#b2df8a", "#a6cee3"), labels = c("female", "male")) + scale_c
p2
```

```
## Picking joint bandwidth of 22.3
## Picking joint bandwidth of 22.3
```
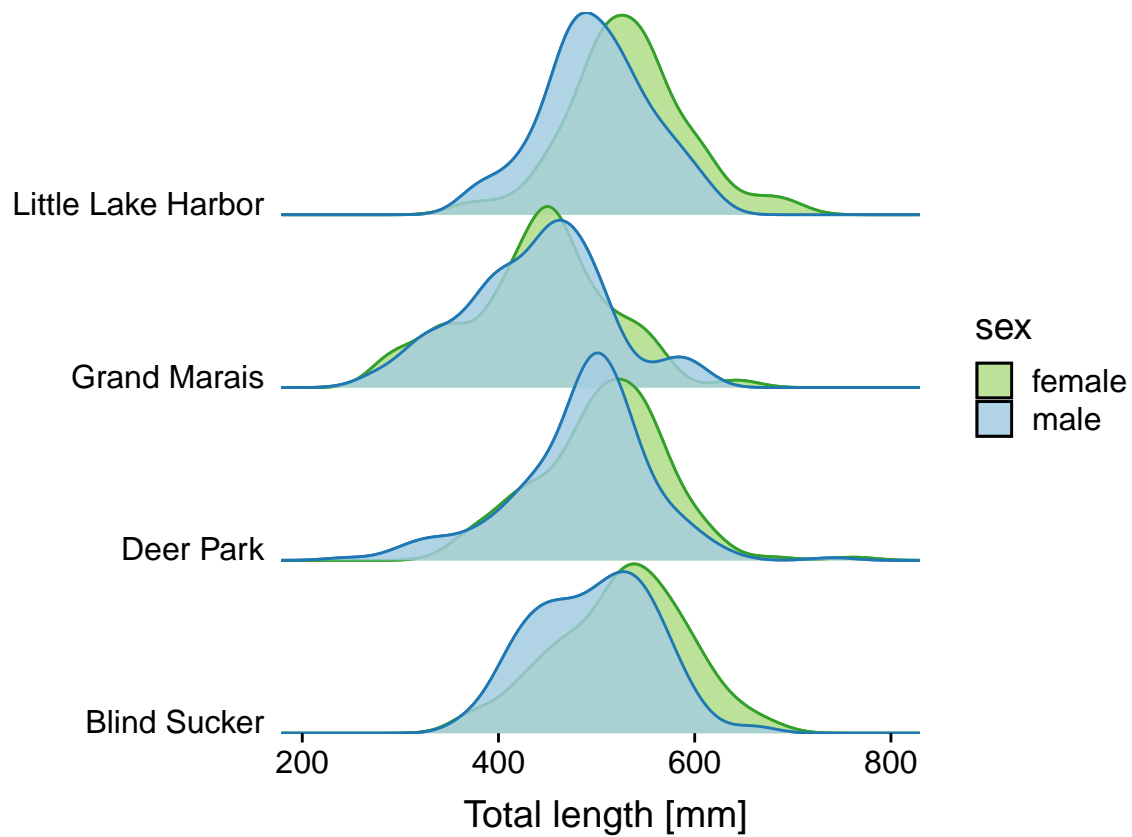
```
## taking away the legend, and labeling "male" and "female" within the plot
p3 <- p2 + annotate("text", x = 620, y = 4.75, label = "female",
        color = "#33a02c") + annotate("text", x = 410, y = 4.75, label = "male",
        color = "#1f78b4") + theme(legend.position="none")
p3
```

```
## Picking joint bandwidth of 22.3
## Picking joint bandwidth of 22.3
```
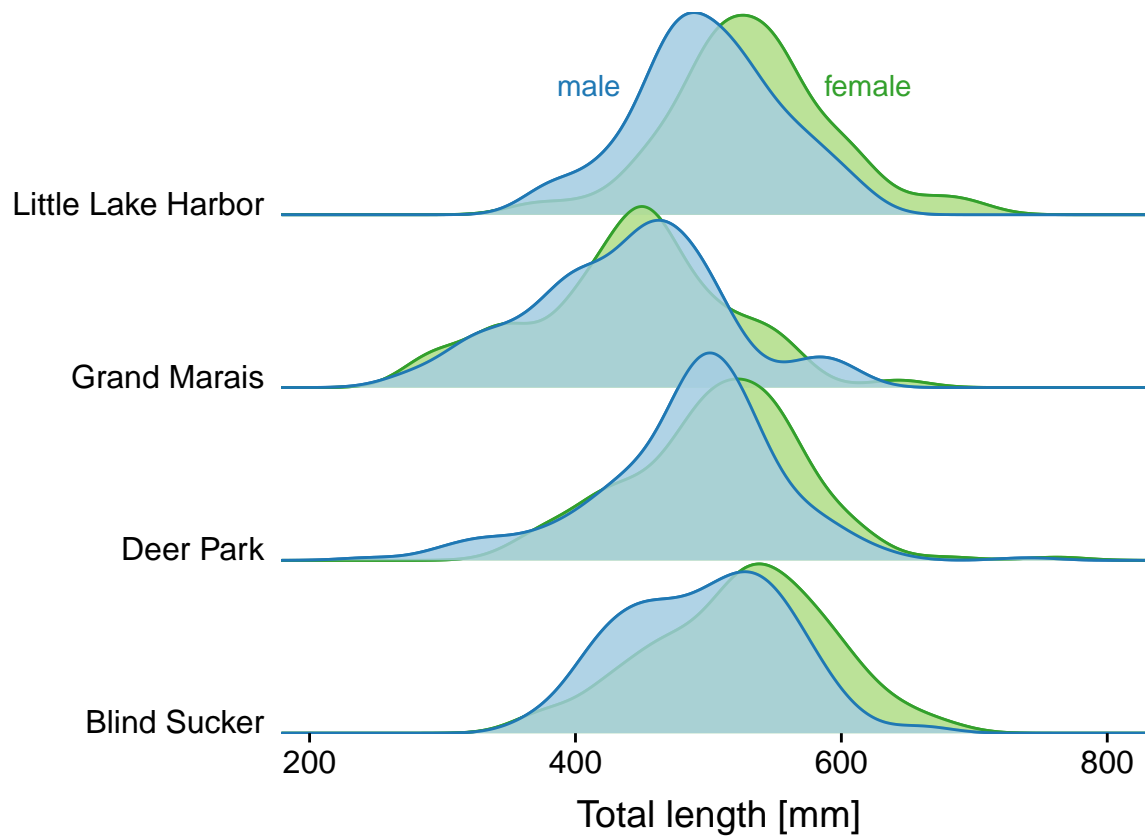
```
## adding a title with color
p4 <- p3 + labs(title = "<b> Total lengths of Siscowet Lake Trout </b><br> <span style = 'font-size:10pt
  theme(
    plot.title.position = "plot",
    plot.title = element_markdown(size = 25, lineheight = 1.2)
  )
 p4
```

```
## Picking joint bandwidth of 22.3
## Picking joint bandwidth of 22.3
```

Total lengths of Siscowet Lake Trout

Male and **female** lengths across four locations in Michigan waters of Lake Superior.
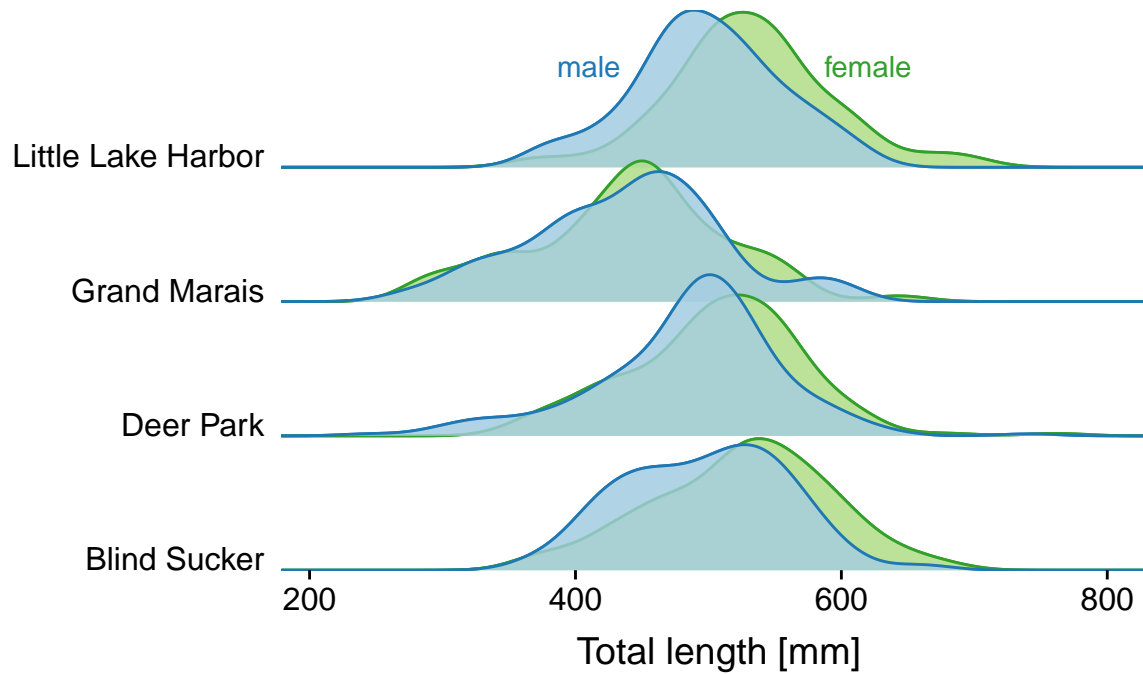
This is my final expository figure.