Assignment 5: Creating Figures

Aidan Coyle

2/15/2021

Assignment 5

In this assignment, we will be creating an exploratory figure, and then fine-tuning that into an explanatory figure.

Data is sourced from the FSAdata package authored by Derek Ogle. We will specifically be using the SiscowetMI2004 dataset, which contains age and length data for male and female Siscowet Lake Trout.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(FSAdata)
```

```
## ## FSAdata v0.3.8. See ?FSAdata to find data for specific fisheries analyses.
```

```
library(viridis)
```

```
## Loading required package: viridisLite
```

```
# Load Lake Siscowet dataset
fishdata <- FSAdata::SiscowetMI2004

# For better reproducibility, write dataset to CSV that we'll keep in our repo
write_csv(fishdata, file = "../data/SiscowetMI2004.csv")

# View the first few lines
head(fishdata)
```

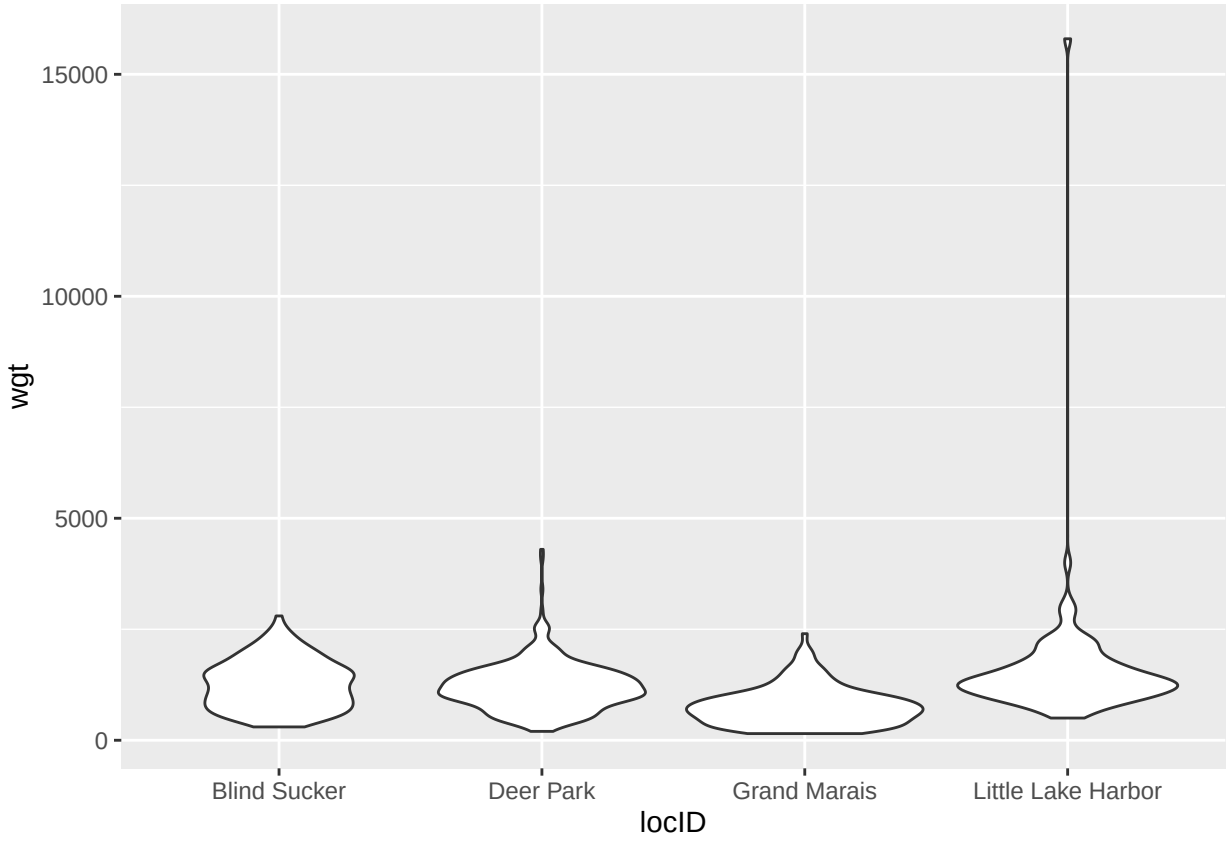
```
##       locID pnldep mesh fishID  sex age len  wgt
## 1 Deer Park  36.74  2.5  19108 <NA>  NA 316  400
## 2 Deer Park  40.09  3.0  19109 <NA>  NA 396  700
## 3 Deer Park  41.46  5.0  19110    M  NA 590 1800
## 4 Deer Park  41.46  5.0  19111    M  NA 516 1500
## 5 Deer Park  43.45  5.5  19112 <NA>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

Exploratory Plot

We will start exploring the data by plotting weight vs. location

```
fishdata %>%
  ggplot(aes(x = locID, y = wgt)) +
  geom_violin()
```

```
## Warning: Removed 1 rows containing non-finite values (stat_ydensity).
```



Based on this graph, we have either a data entry error or a truly massive fish. Let's check that line

```
fishdata[fishdata$wgt > 10000,]
```

```
##                  locID pnldep mesh fishID  sex age len   wgt
## NA                <NA>     NA   NA     NA <NA>  NA  NA    NA
## 541 Little Lake Harbor  40.85    3  19234    F  12 550 15800
```

```r
# Now let's examine fish lengths in the data
mean(fishdata$len)
```

```
## [1] 487.1359
```

```r
max(fishdata$len)
```

```
## [1] 762
```

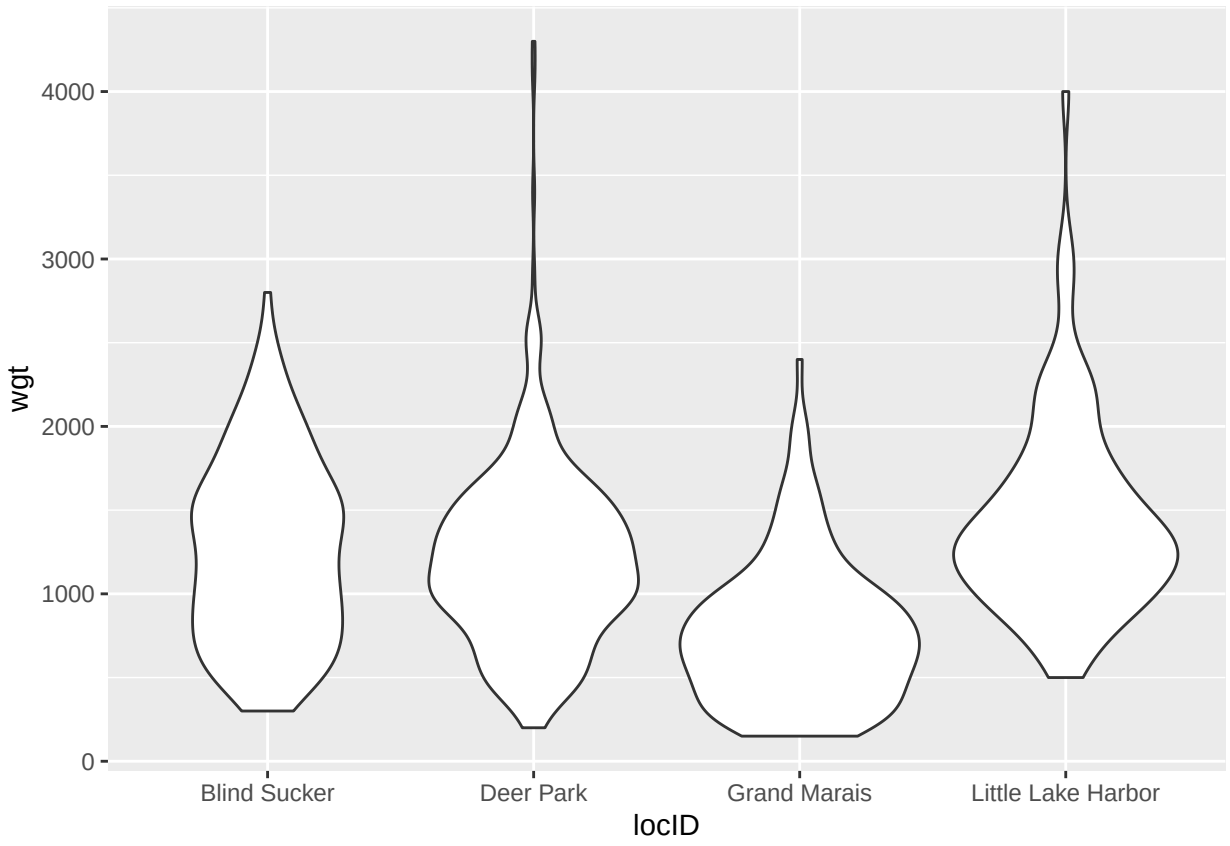
It is quite unlikely that a fish just slightly longer than the mean, and much smaller than the maximum length, is over 3x heavier than any other fish. Therefore, we will assume this is the result of a data entry error. We will therefore remove this fish from the analysis

```r
fishdata$wgt[fishdata$wgt == 15800] <- NA
```

Now let's make our exploratory graph again

```r
fishdata %>%
  ggplot(aes(x = locID, y = wgt)) +
  geom_violin()
```

```
## Warning: Removed 2 rows containing non-finite values (stat_ydensity).
```



Nice, looking good! Looks like we've got some differences in weight distribution among sites.

Expository Graph

We will now create our expository graph. We'll continue to look at location ID and weight, but will also be examining panel depth (the depth of the gillnet panel where the fish was caught).

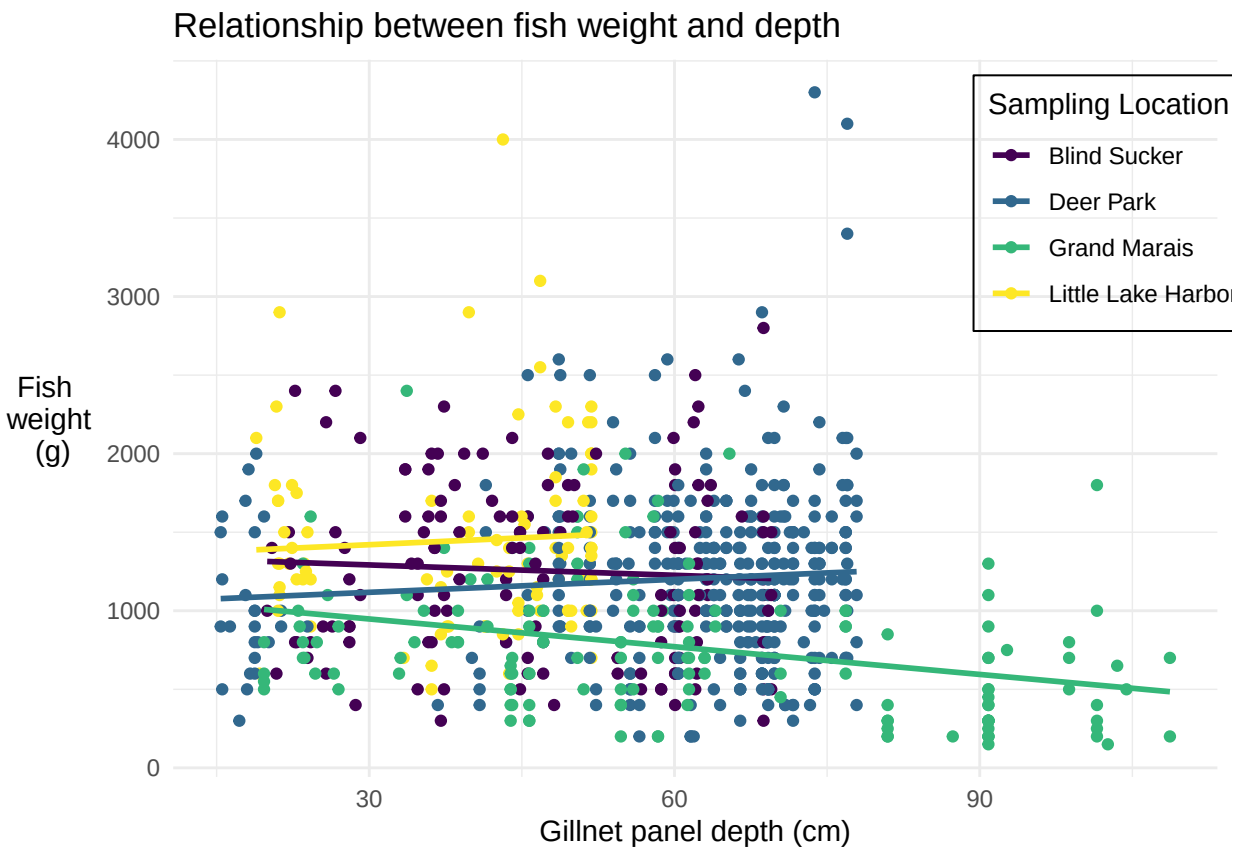
Units of panel depth are not provided either in the data or in the metadata. For the purposes of this examination, we will assume units are in cm for the following reasons - length and weight are both metric, and it is unlikely to be either mm (a 20mm difference is negligible when examining fish depth) or m (a gillnet 100m deep is implausible)

```
fishdata %>%
  ggplot(aes(x = pnldep, y = wgt, color = locID)) +
  geom_point() +
  ylab("Fish \n weight \n (g)") +
  xlab("Gillnet panel depth (cm)") +
  scale_color_viridis(discrete = "TRUE",
                      name = "Sampling Location",
                      option = "viridis") +
  theme_minimal() +
  theme(axis.title.y = element_text(angle = 0, vjust = 0.5),
        legend.position = c(0.9, 0.8),
        legend.box.background = element_rect(color = "black")) +
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("Relationship between fish weight and depth")
```

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

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



From a visual examination, it looks like we have no clear relationship between fish weight and depth! However, at Grand Marais, there does seem to be a negative trend. This is also the only location where the gillnet was deeper than ~75 cm. This indicates that we may want to gather additional data from deeper waters at the other locations to determine whether there is some amount of size segregation by depth