FISH 497 Homework #5

Jessica Diallo

20 February 2021

Background

This week we learned about some graphic design principles, which included an example of creating an *expository* figure from an initial *exploratory* figure. Your assignment this week is to use the provided dataset to

1) begin by creating an *exploratory* figure you'd like about *any* aspect of the data, and
2) create an *expository* figure that builds upon your original figure.

Your figure can include some aspect of data anlysis (eg, linear regression), but there is no requirement for you to do so. You can use either **base graphics** or **ggplot2** to complete this assignment.

Please make sure to take into consideration the concepts we discussed in class with respect to color palettes and accessibility. You may also want to consider the options that Jeff Leek used in his demonstration, but it's not necessary that your figure contain all of the elements and styling that he included.

Data description

The data come from the {FSAdata} package, which was created by Derek Ogle as part of his book *Introductory Fisheries Analysis with R*. The dataset contains information on the ages (subsample), and length and mass (all fish) for male and female Siscowet Lake Trout captured at four locations in Michigan waters of Lake Superior.

Format

A data frame with 780 observations on the following 8 variables.

1) `locID`: Locations (`Blind Sucker`, `Deer Park`, `Grand Marais`, `Little Lake Harbor`)
2) `pnldep`: Depth of gillnet panel in which the fish was captured
3) `mesh`: Gillnet stretch mesh measure
4) `fishID`: Unique fish identification code
5) `sex`: Sex (`F` = female; `M` = male)

6) `age`: Assigned ages (years)

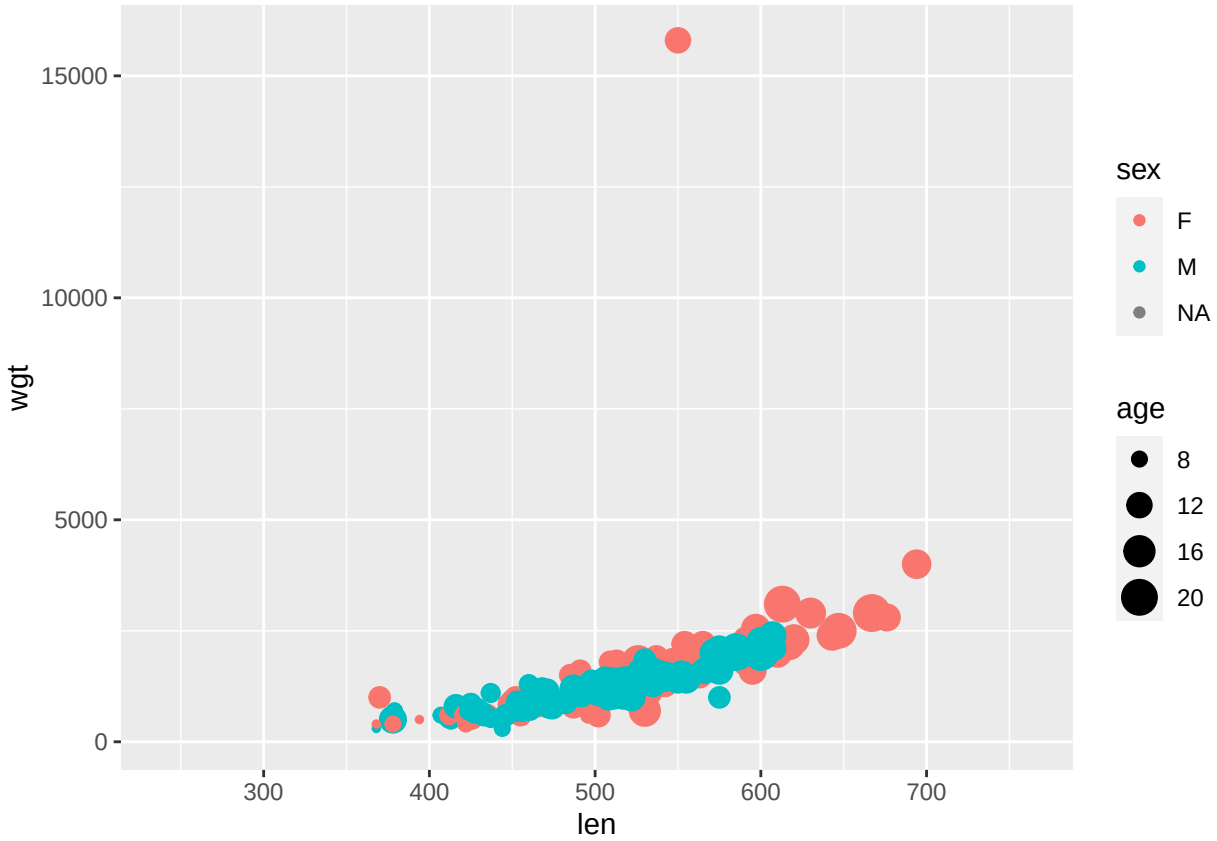
7) `len`: Total length (mm)

8) `wgt`: Weight (g)

Exploratory Figures

```
ggplot(data, aes(x=len, y=wgt, color=sex, size=age)) +
  geom_point()
```

```
## Warning: Removed 581 rows containing missing values (geom_point).
```



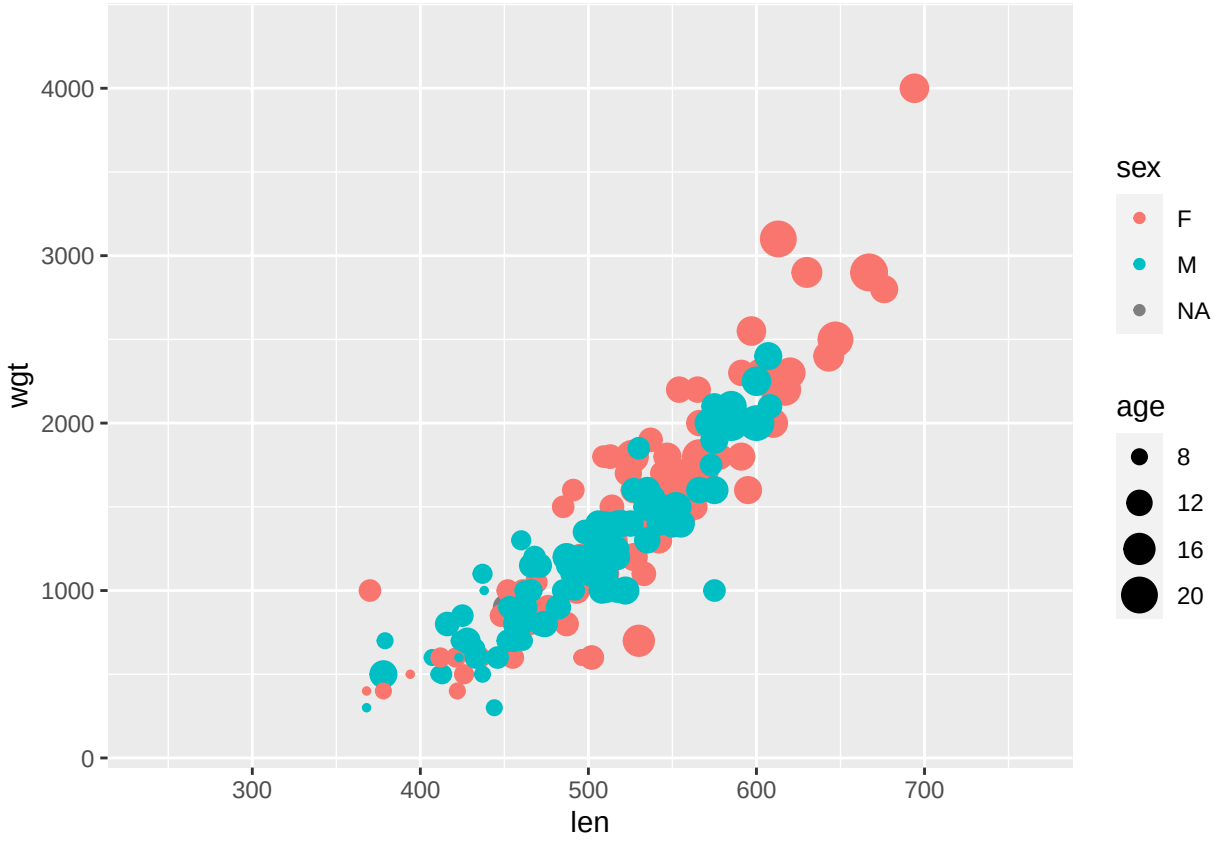

```
# there's one point that looks like a mistake because the weight value is almost
# an order of magnitude higher than the other points
# so I'm going to remove that point
summary(data$wgt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     150     775    1100    1175    1500   15800       1
```

```
data[which(data$wgt == 15800),"wgt"] <- NA

ggplot(data, aes(x=len, y=wgt, color=sex, size=age)) +
  geom_point()
```

```
## Warning: Removed 582 rows containing missing values (geom_point).
```



Expository Figure

I created this expository figure to show the relationship between length and weight of Siscowet Lake Trout. Males and females are separated by color, and the distribution of values for length and weight are shown as histograms. Linear regression lines are displayed separately for males and females.

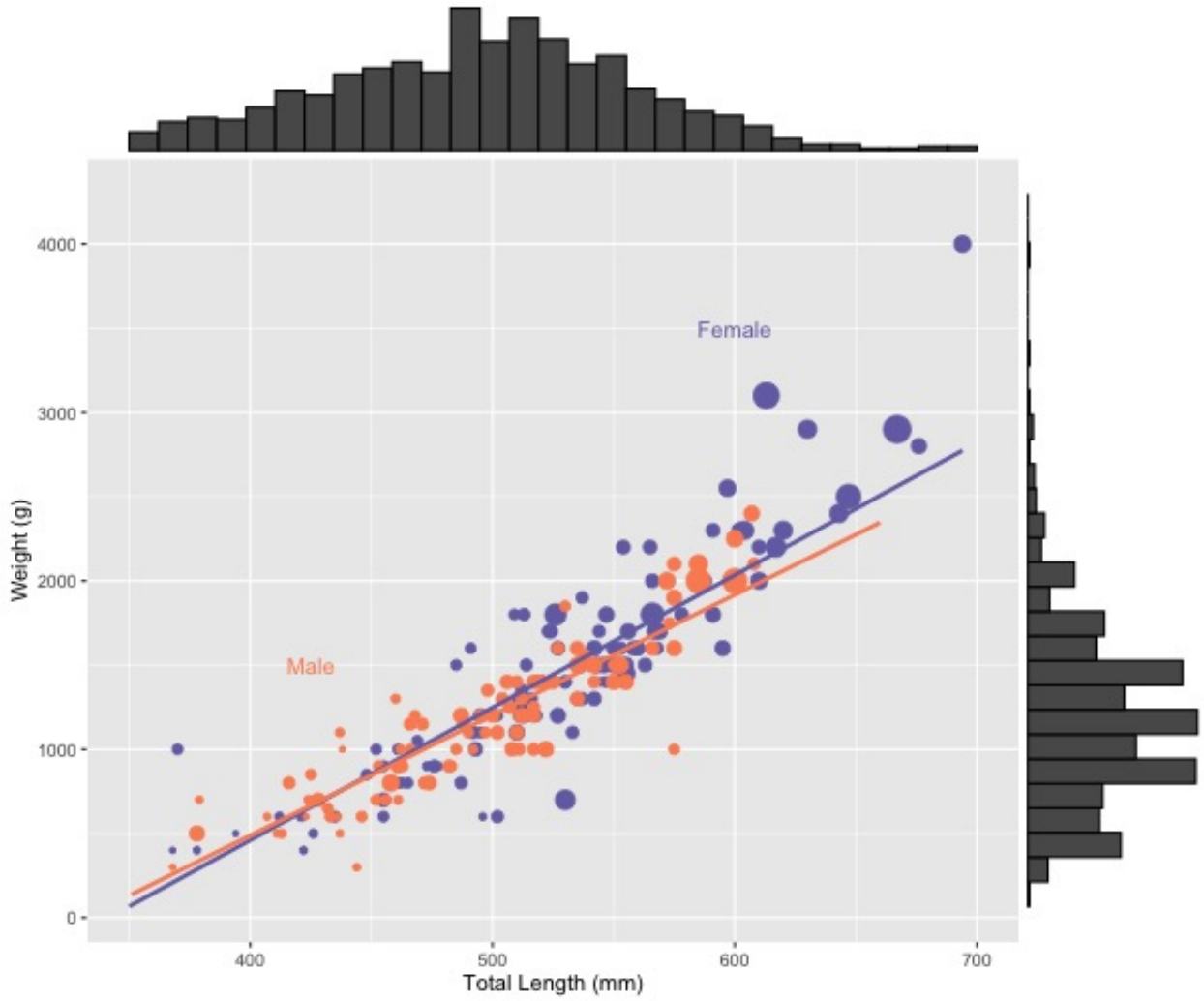


Figure 1: Expository figure of Siscowet Lake Trout weight as a function of length. Point size reflects relative fish age.