Assignment 8: Mapping

Rachel Bash

OVERVIEW

This exercise accompanies the lessons in Hydrologic Data Analysis on mapping

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single pdf file.
5. After Knitting, submit the completed exercise (pdf file) to the dropbox in Sakai. Add your last name into the file name (e.g., "A08_Salk.html") prior to submission.

The completed exercise is due on 23 October 2019 at 9:00 am.

Setup

1. Verify your working directory is set to the R project file,
2. Load the tidyverse, lubridate, cowplot, LAGOSNE, sf, maps, and viridis packages.
3. Set your ggplot theme (can be theme_classic or something else)
4. Load the lagos database, the USA rivers water features shape file, and the HUC6 watershed shape file.

```
getwd()
```

```
## [1] "C:/Users/19524/Documents/DUKE/Hydrologic Data Analytics/Hydrologic_Data_Analysis/Assignments"
```

```
library(tidyverse)
library(lubridate)
library(cowplot)
library(LAGOSNE)
library(sf)
library(maps)
library(viridis)

theme_set(theme_classic())
options(scipen = 100)


LAGOSdata <- lagosne_load()

waterfeatures <- st_read("../Data/Raw/hydrogl020.dbf")
```

```
## Reading layer `hydrogl020' from data source `C:\Users\19524\Documents\DUKE\Hydrologic Data Analytics
## Simple feature collection with 76975 features and 12 fields
## geometry type:  LINESTRING
## dimension:      XY
## bbox:           xmin: -179.9982 ymin: 17.67469 xmax: 179.9831 ymax: 71.39819
## epsg (SRID):    NA
## proj4string:    NA
```

```
HUC6 <- st_read("../Data/Raw/Watersheds_Spatial/WBDHU6.dbf")
```

```
## Reading layer `WBDHU6' from data source `C:\Users\19524\Documents\DUKE\Hydrologic Data Analytics\Hyd
## Simple feature collection with 33 features and 15 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -90.6235 ymin: 24.39533 xmax: -75.3981 ymax: 37.52103
## epsg (SRID):    4269
## proj4string:    +proj=longlat +datum=NAD83 +no_defs
```

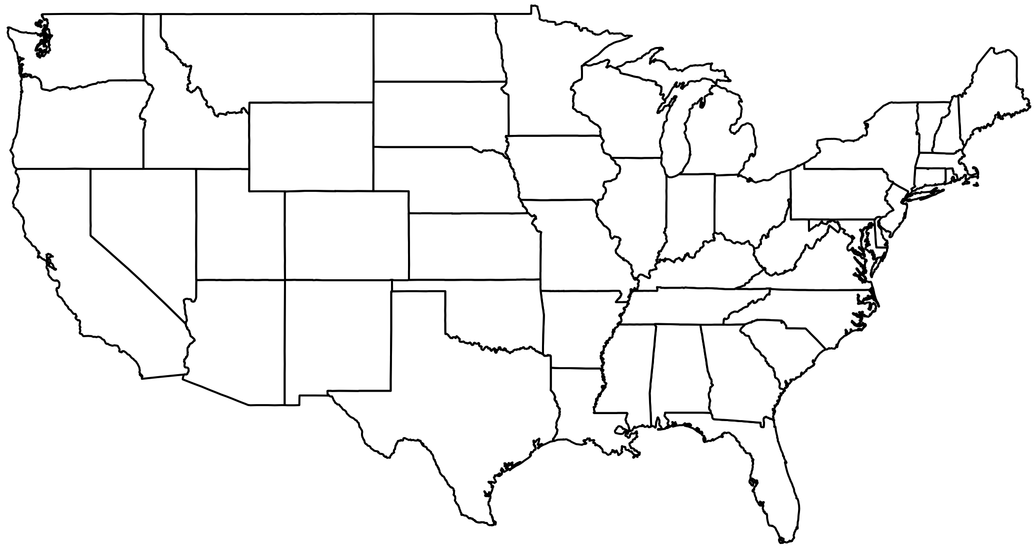
Mapping water quality in lakes

Complete the in-class exercise from lesson 15, to map average secchi depth measurements across states in Maine, considering lake area and lake depth as predictors for water clarity. Steps here are identical to the lesson, with the following edits:

- Make sure all your wrangling is done in this document (this includes basic wrangling of the LAGOS database)
- In your cowplot, do not adjust the legend items (even though they look ugly). Rather, reflect on how you would improve them with additional coding.
- For item 9, **do** run a regression on secchi depth by lake area and a separate regression on secchi depth by lake depth. Make scatterplots of these relationships. Note that log-transforming one of these items may be necessary.

5. Filter the states and secchi depth datasets so that they contain Maine only. For the secchi depth dataset, create a summary dataset with just the mean secchi depth.

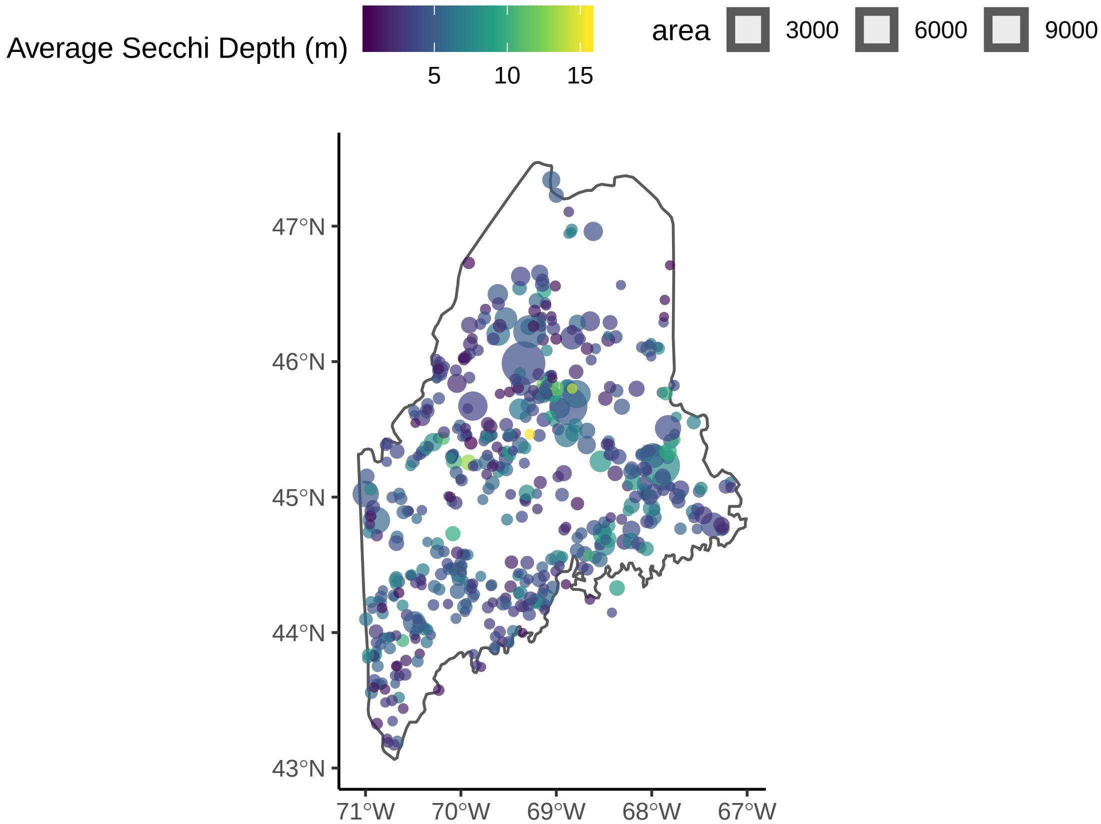
```r
# load LAGOSNE data frames
LAGOSlocus <- LAGOSdata$locus
LAGOSstate <- LAGOSdata$state
LAGOSnutrient <- LAGOSdata$epi_nutr
LAGOSlimno <- LAGOSdata$lakes_limno

# Join Lagos dataframes, select correct columns, and filter for Maine
LAGOSMaine <-
  left_join(LAGOSnutrient, LAGOSlocus) %>%
  left_join(., LAGOSlimno) %>%
  left_join(., LAGOSstate) %>%
  filter(!is.na(state)) %>%
  select(lagoslakeid, sampledate, secchi, lake_area_ha, maxdepth, nhd_lat, nhd_long,
         state) %>%
  filter(state=="ME")

#create secchi depth dataset
MEsecchi.summary <- LAGOSMaine %>%
  group_by(lagoslakeid) %>%
  summarise(secchi.mean = mean(secchi),
            area = mean(lake_area_ha),
            depth = mean(maxdepth),
            lat = mean(nhd_lat),
            long = mean(nhd_long)) %>%
  drop_na()

MEsecchi.spatial <- st_as_sf(MEsecchi.summary, coords = c("long", "lat"), crs = 4326)

#create Maine dataset from states subset
states.subset <- st_as_sf(map(database = "state", plot = TRUE, fill = TRUE, col = "white"))
```



```r
Mainemap <- states.subset %>%
  filter(ID=="maine")
```

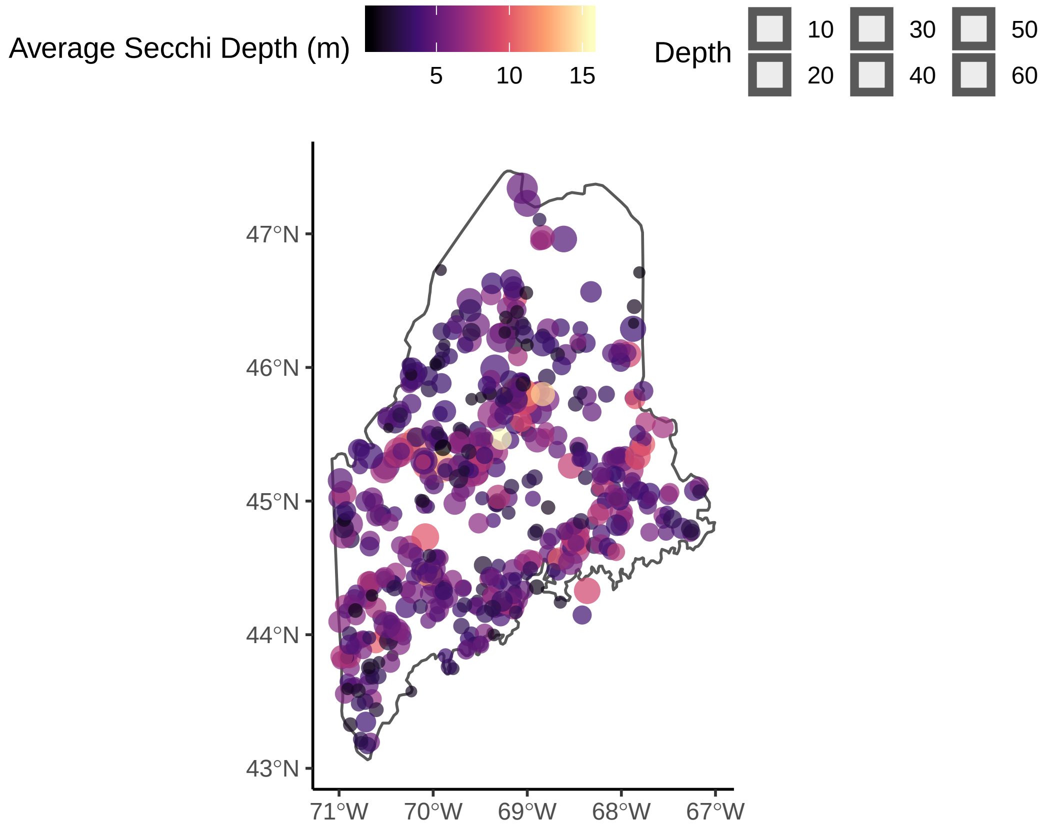
6. Create a plot of mean secchi depth for lakes in Maine, with mean secchi depth designated as color and the lake area as the size of the dot. Remember that you are using size in the aesthetics and should remove the size = 1 from the other part of the code. Adjust the transparency of points as needed.

```r
MEsecchiplot <- ggplot() +
  geom_sf(data = Mainemap, fill = "white") +
  geom_sf(data = MEsecchi.spatial, aes(color = secchi.mean, size = area),
          alpha = 0.7) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi Depth (m)") +
  theme(legend.position = "top") +
  scale_size_continuous(range = c(1, 7))
print(MEsecchiplot)
```



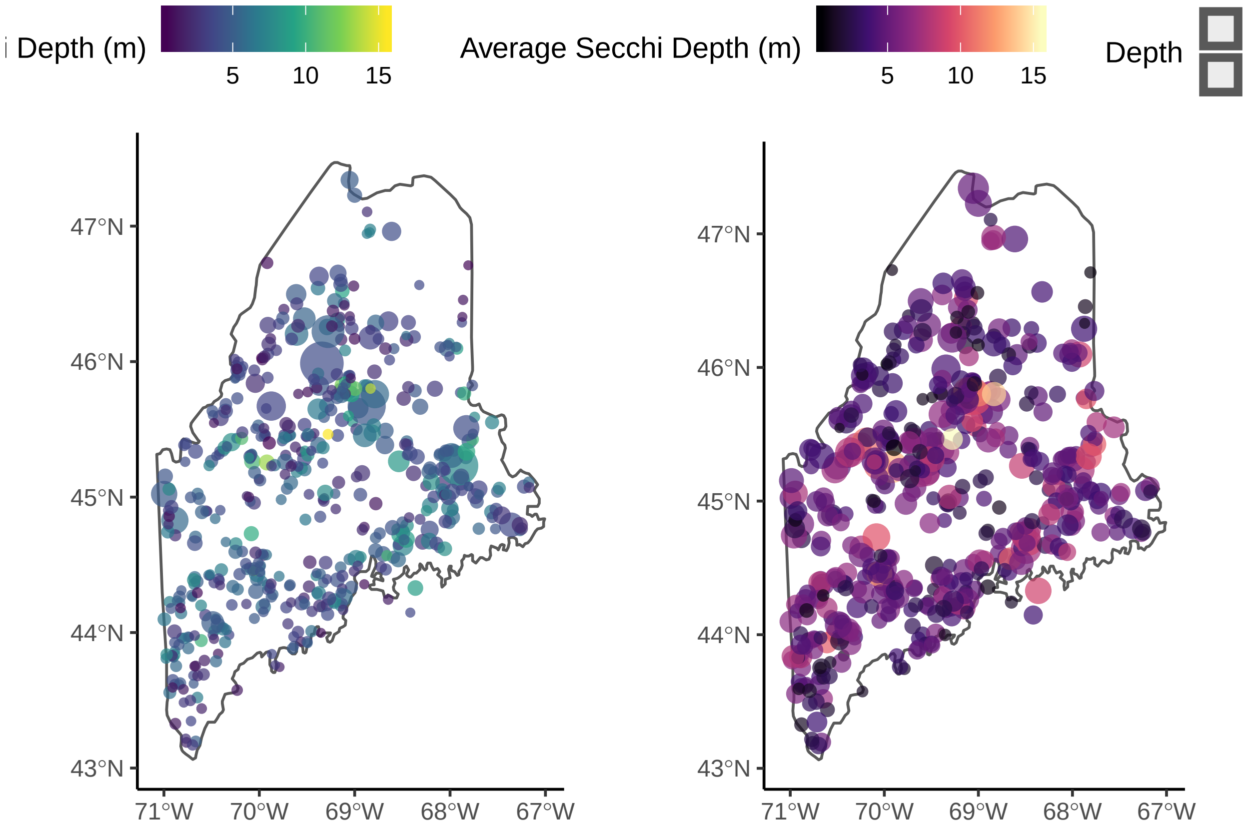
7. Create a second plot, but this time use maximum depth of the lake as the size of the dot.

```
MEsecchiplot2 <- ggplot() +
  geom_sf(data = Mainemap, fill = "white") +
  geom_sf(data = MEsecchi.spatial, aes(color = secchi.mean, size = depth),
          alpha = 0.7) +
  scale_color_viridis_c(option="magma") +
  labs(color="Average Secchi Depth (m)", size = "Depth") +
  theme(legend.position = "top")+
  scale_size_continuous(range = c(1, 5))
print(MEsecchiplot2)
```



8. Plot these maps in the same plot with the `plot_grid` function. Don't worry about adjusting the legends (if you have extra time this would be a good bonus task).

```
Both <- plot_grid(MEsecchiplot, MEsecchiplot2)
print(Both)
```



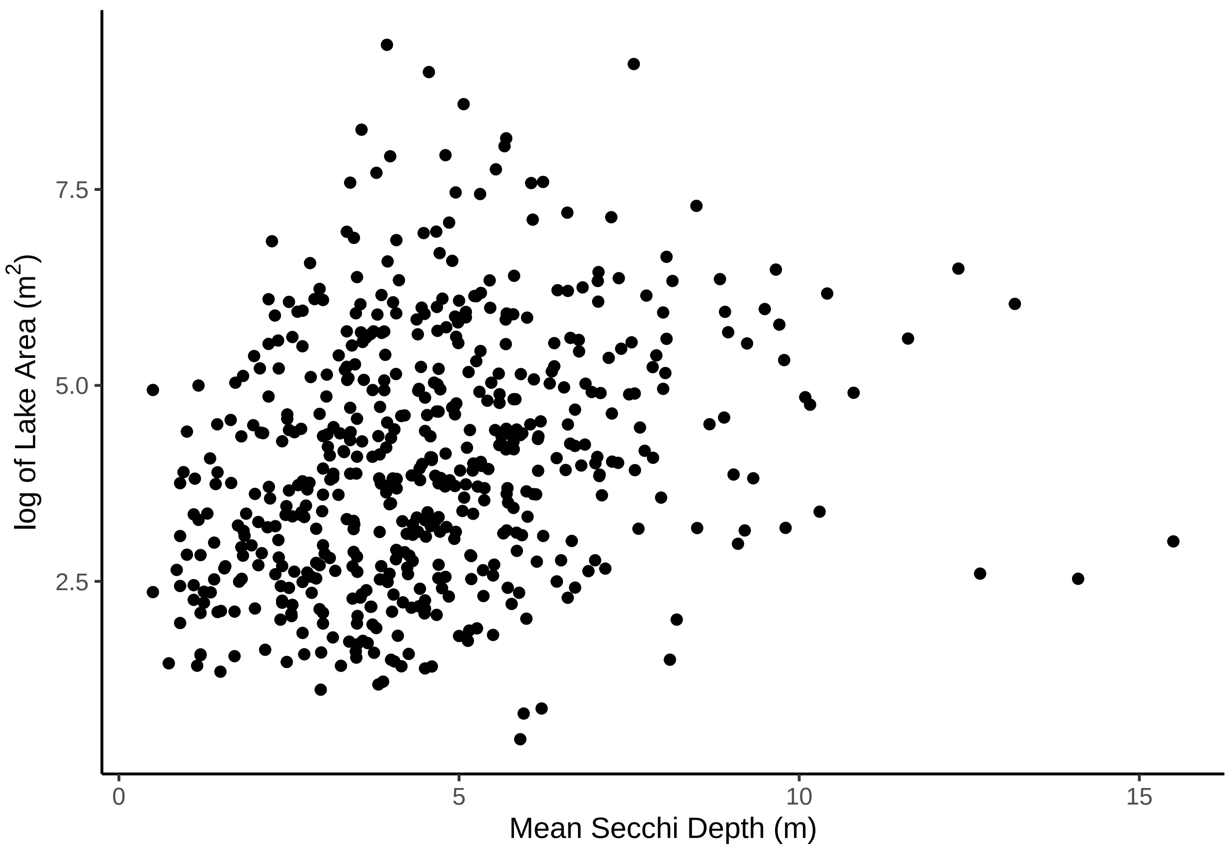
What would you change about the legend to make it a more effective visualization?

> I would write code so that the Depth and Area square sizes correctly reflect the depth and area measurements that the circles in the graphs are trying to reflect. I would also change the squares in the legends to circles to match the map

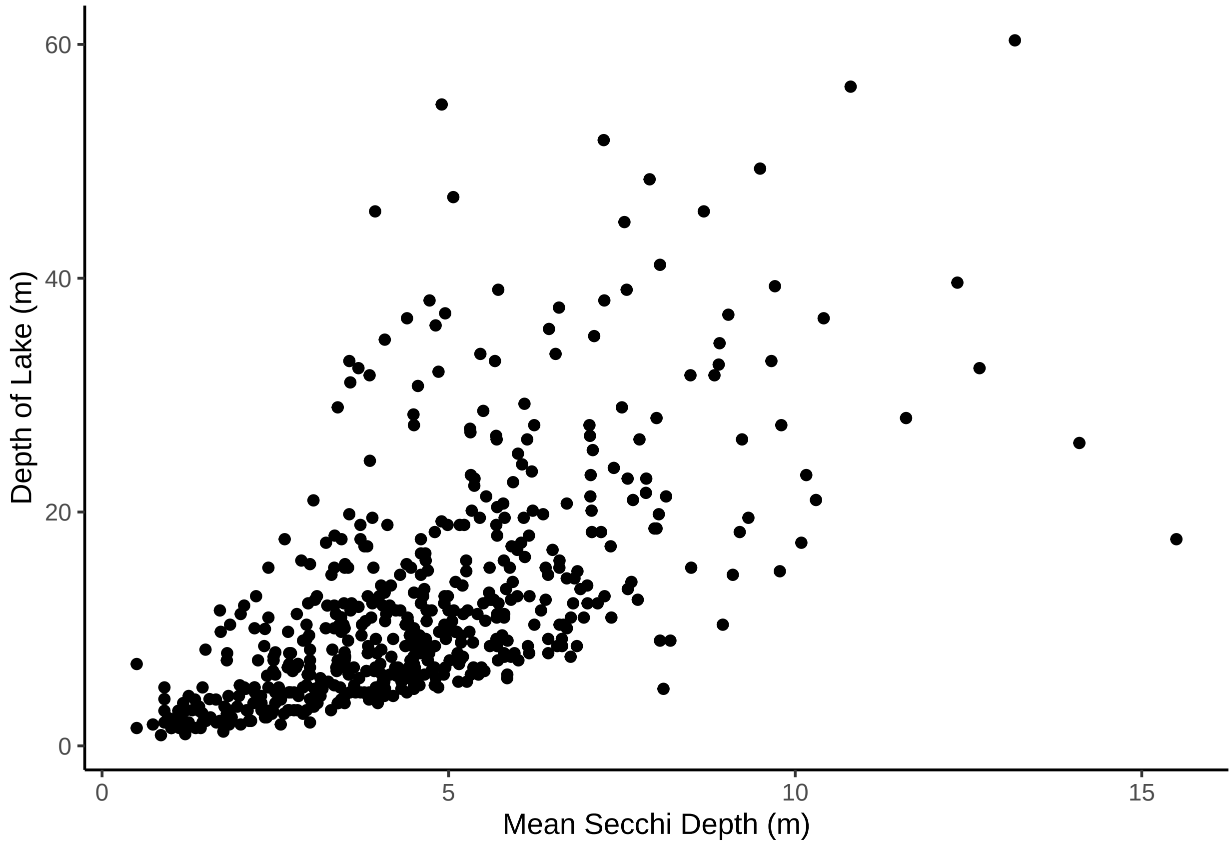
9. What relationships do you see between secchi depth, lake area, and lake depth? Which of the two lake variables seems to be a stronger determinant of secchi depth? (make a scatterplot and run a regression to test this)

Note: consider log-transforming a predictor variable if appropriate

```
#plot secchi depth vs log(area) for a linear relationship
area.plot <- ggplot(MEsecchi.spatial) +
  geom_point(aes(x=secchi.mean, y=log(area))) +
  labs(x="Mean Secchi Depth (m)", y=expression("log of Lake Area (m"^2*")"))
print(area.plot)
```




```r
#plot secchi depth vs lake depth
depth.plot <- ggplot(MEsecchi.spatial) +
  geom_point(aes(x=secchi.mean, y=depth)) +
  labs(x="Mean Secchi Depth (m)", y= "Depth of Lake (m)")
print(depth.plot)
```




```r
model1 <- lm(data=MEsecchi.summary, secchi.mean ~ log(area) + depth)
summary(model1)
```

```
##
## Call:
## lm(formula = secchi.mean ~ log(area) + depth, data = MEsecchi.summary)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5424 -1.0837 -0.0940  0.9511 10.0893
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept)  3.146493   0.199467  15.775 <0.0000000000000002 ***
## log(area)   -0.097802   0.052920  -1.848              0.0651 .
## depth        0.144728   0.008422  17.184 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.701 on 546 degrees of freedom
## Multiple R-squared:  0.3947, Adjusted R-squared:  0.3925
## F-statistic:   178 on 2 and 546 DF,  p-value: < 0.00000000000000022
```

Based on the maps, the plots, and the model, it is clear that secchi depth is better predicted by depth of the lake, while lake area is a weaker determinant of secchi depth. Lake area was log transformed in order to make the relationship between area and secchi depth linear. When area was log transformed, the model output showed that the log(area) was not a significant predictor of mean secchi depth ($p = 0.0651$) while lake depth is a significant predictor of secchi depth ($p < 0.001$). The model suggests that for every one meter increase in lake depth, mean secchi depth increases by 0.14m.

Mapping water features and watershed boundaries

10. Wrangle the USA rivers and HUC6 watershed boundaries dataset so that they include only the features present in Florida (FL). Adjust the coordinate reference systems if necessary to ensure they use the same projection.

```
# Filter for Florida
waterfeatures <- filter(waterfeatures, STATE == "FL")

# Remove a couple feature types we don't care about
waterfeatures <- filter(waterfeatures, FEATURE != "Apparent Limit" & FEATURE != "Closure Line")

#set reference system
waterfeatures <- st_set_crs(waterfeatures, 4269)

#filter for florida watersheds
summary(HUC6$States)
```

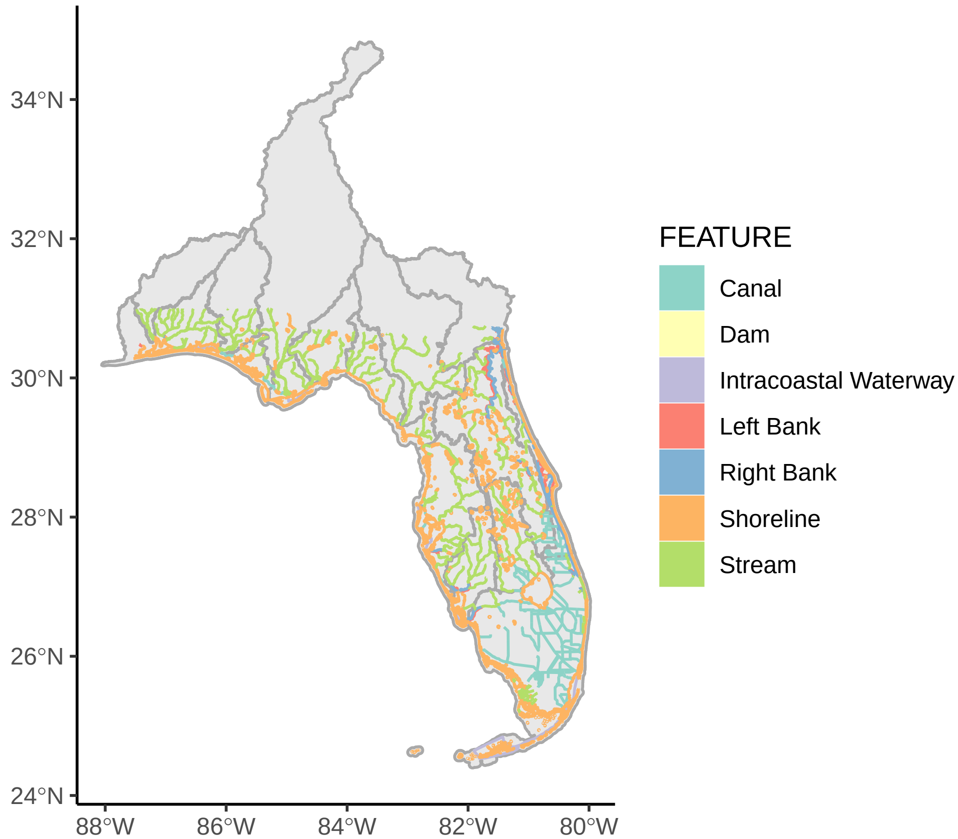
```
##       AL    AL,FL AL,FL,GA AL,GA,TN AL,LA,MS    AL,MS       FL    FL,GA
##        1        3        1        1        1        2        6        4
##       GA GA,NC,SC    LA,MS       NC    NC,SC NC,SC,VA    NC,VA       SC
##        2        1        1        4        2        1        2        1
```

```
HUC6.FL <- HUC6 %>%
  filter(States %in% c("AL,FL", "AL,FL,GA", "FL", "FL,GA"))
st_crs(HUC6.FL) #they match
```

```
## Coordinate Reference System:
##   EPSG: 4269
##   proj4string: "+proj=longlat +datum=NAD83 +no_defs"
```

11. Create a map of watershed boundaries in Florida, with the layer of water features on top. Color the watersheds gray (make sure the lines separating watersheds are still visible) and color the water features by type.

```
FLlayers <- ggplot() +
  geom_sf(data = HUC6.FL, fill = "lightgrey", color = "darkgray", alpha = 0.5) +
  geom_sf(data = waterfeatures, aes(fill = FEATURE, color = FEATURE)) +
  scale_fill_brewer(palette = "Set3") +
  scale_color_brewer(palette = "Set3")
print(FLlayers)
```



12. What are the dominant water features in Florida? How does this distribution differ (or not) compared to North Carolina?

    There are a lot more canals in Florida (concentrated mostly in the southern tip) and there seem to be more lakes in the interior than in North Carolina. Besides the southern tip of Florida, the interior is dominated by lots of streams. Just like North Carolina, the shore is very jagged, with many inlets and confusing shoreline designations.

Reflection

13. What are 2-3 conclusions or summary points about mapping you learned through your analysis?

    I really liked mapping in R. I learned it is relatively easy if you have the right data. You can convey a lot of information when you map the data instead of just plotting it on a graph.

14. What data, visualizations, and/or models supported your conclusions from 13?

    The plot grid of secchi depth in Maine for lake area and depth showed a lot of great information and can inform decisions regarding water quality.

15. Did hands-on data analysis impact your learning about mapping relative to a theory-based lesson? If so, how?

    Yes, there is no better way to communicate your findings than with mapping your spatial data, and learning how to do that ourselves was super helpful.

16. How did the real-world data compare with your expectations from theory?

    It unfortunately will take a lot more code and clipping in order to make really neat graphs (for example, clipping the rest of the HUC 6 watersheds that aren't in Florida would make the map look better)