Assignment 8: Mapping

Caroline Watson

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.

```
#verifying working directory
getwd()
```

```
## [1] "/Users/carolinewatson/Documents/Fall 2019/Hydrologic_Data_Analysis/Assignments"
```

```
#loading packages
library(tidyverse)
```

```
## -- Attaching packages ------------ tidyverse 1.2.1 --

## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

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

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

```r
library(cowplot)
```

```
##
## ********************************************************
## Note: As of version 1.0.0, cowplot does not change the
##   default ggplot2 theme anymore. To recover the previous
##   behavior, execute:
##   theme_set(theme_cowplot())
## ********************************************************
##
## Attaching package: 'cowplot'
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```r
library(LAGOSNE)
library(sf)
```

```
## Linking to GEOS 3.7.2, GDAL 2.4.2, PROJ 5.2.0
```

```r
library(maps)
```

```
##
## Attaching package: 'maps'
## The following object is masked from 'package:purrr':
##
##     map
```

```r
library(viridis)
```

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

```r
library(trend)

#setting ggplot theme
theme_set(theme_classic())

#loading LAGOS dataset
LAGOSdata <- lagosne_load()
```

```
## Warning in `_f`(version = version, fpath = fpath): LAGOSNE version
## unspecified, loading version: 1.087.3
```

```r
#loading USA rivers water features shape file
waterfeatures <- st_read("../Data/Raw/hydrogl020.dbf")
```

```
## Reading layer `hydrogl020' from data source `/Users/carolinewatson/Documents/Fall 2019/Hydrologic_Da
## 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
```

```r
#loading HUC6 watershed shape file
HUC6 <- st_read("../Data/Raw/Watersheds_Spatial/WBDHU6.dbf")
```

```
## Reading layer `WBDHU6' from data source `/Users/carolinewatson/Documents/Fall 2019/Hydrologic_Data_Ar
## 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 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +no_defs
```

Mapping water quality in lakes

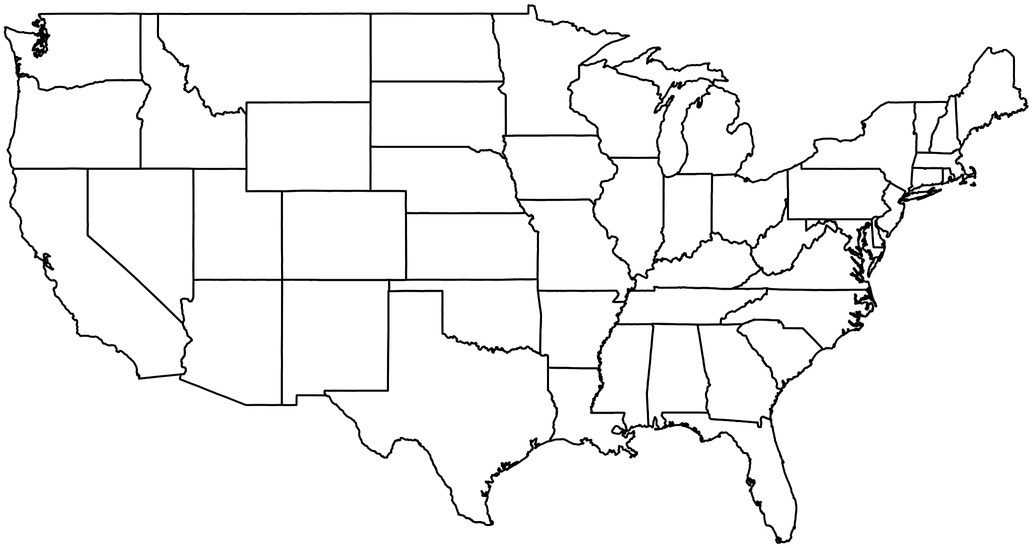
Complete the in-class exercise from lesson 15, to map average secchi depth measurements across states in Maine, considering lake area and lake depth as predictors for water clarity. Steps here are identical to the lesson, with the following edits:

- 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.

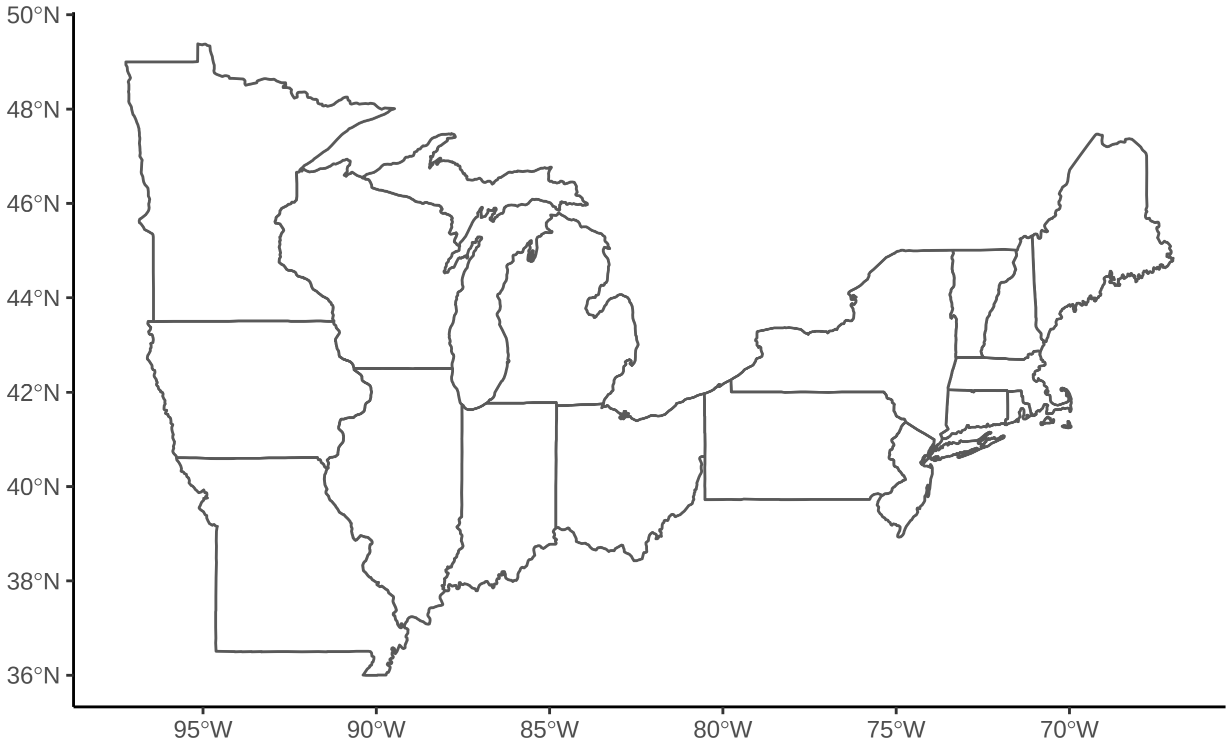
Data Wrangling

```r
# generate a map of U.S. states
states <- st_as_sf(map(database = "state", plot = TRUE, fill = TRUE, col = "white"))
```



```r
# filter only states that are included in the LAGOSNE database
states.subset <- filter(states, ID %in%
```

```r
                            c("minnesota", "iowa", "wisconsin", "illinois",
                              "missouri", "michigan", "indiana", "ohio",
                              "pennsylvania", "new york", "new jersey",
                              "connecticut", "new hampshire", "rhode island",
                              "massachusetts", "vermont", "maine"))
# visualize state plot
LAGOSstateplot <- ggplot(states.subset) +
  geom_sf(fill = "white")
print(LAGOSstateplot)
```




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

# Create a data frame to visualize secchi depth
LAGOScombined <-
  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)
```

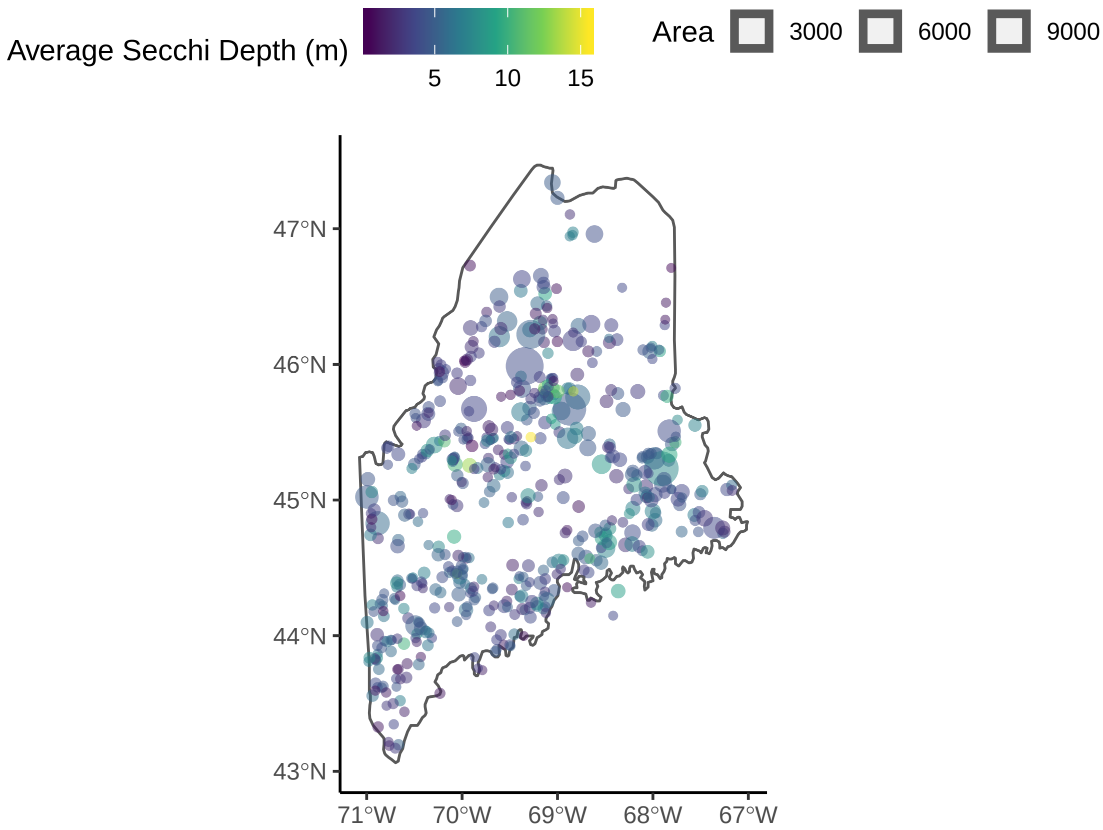
```
## Joining, by = "lagoslakeid"

## Joining, by = c("lagoslakeid", "nhdid", "nhd_lat", "nhd_long")

## Joining, by = "state_zoneid"
```

```r
#filtering for ME data only
Maine_secchi <- LAGOScombined %>%
```

```
  filter(state == "ME")

#Maine subset dataset
Maine_subset <- states.subset %>%
  filter(ID == "maine")

#ME secchi summary data
ME.secchi.summary <- Maine_secchi %>%
  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()
```

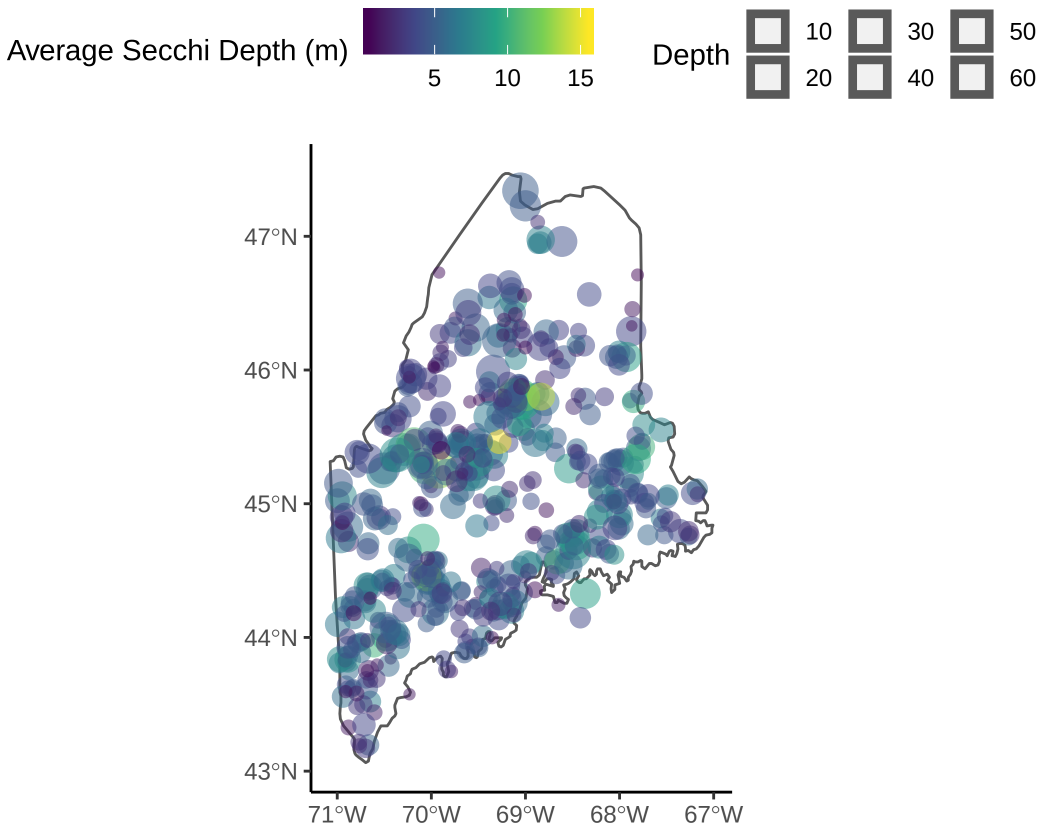
6. Create a plot of mean secchi depth for lakes in Maine, with mean secchi depth designated as color and the lake area as the size of the dot. Remember that you are using size in the aesthetics and should remove the size = 1 from the other part of the code. Adjust the transparency of points as needed.

```
#secchi spatial info for ME
secchi.spatial.ME <- st_as_sf(ME.secchi.summary, coords = c("long", "lat"), crs = 4326)

#plot of mean secchi depths for lakes in ME
ME.mean.secchi.plot <- ggplot() +
  geom_sf(data = Maine_subset, fill = "white") +
  geom_sf(data = secchi.spatial.ME, aes(color = secchi.mean, size = area),
          alpha = 0.5) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi Depth (m)", size = "Area") +
  theme(legend.position = "top")
print(ME.mean.secchi.plot)
```



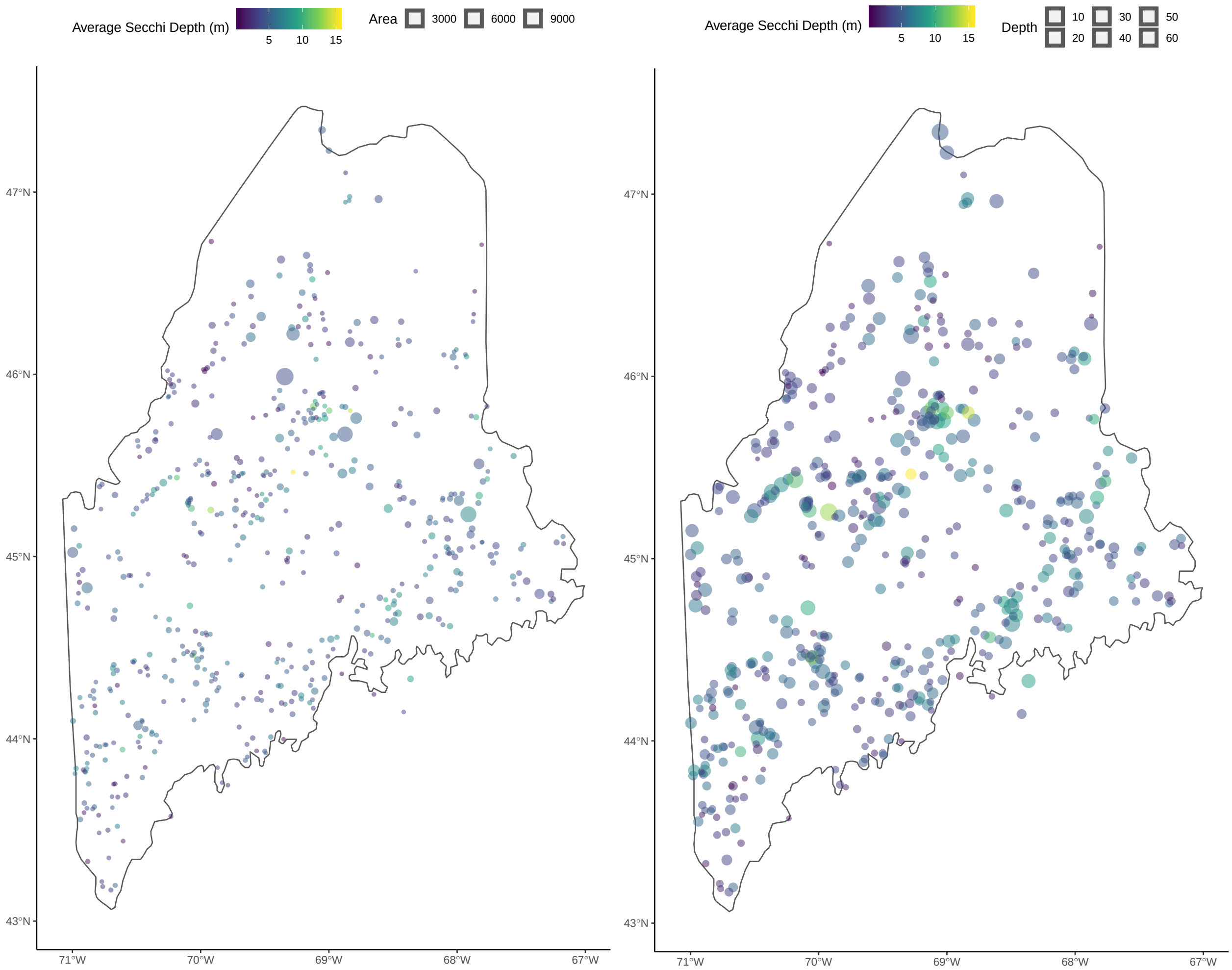
7. Create a second plot, but this time use maximum depth of the lake as the size of the dot.

```
#ggplot of data with max depth as size of the dot
ME.max.depth.plot <- ggplot() +
  geom_sf(data = Maine_subset, fill = "white") +
  geom_sf(data = secchi.spatial.ME, aes(color = secchi.mean, size = depth),
          alpha = 0.5) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi Depth (m)", size = "Depth") +
  theme(legend.position = "top")
print(ME.max.depth.plot)
```



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).

```
#plotting the graphs on the same plot
plot_grid(ME.mean.secchi.plot, ME.max.depth.plot)
```



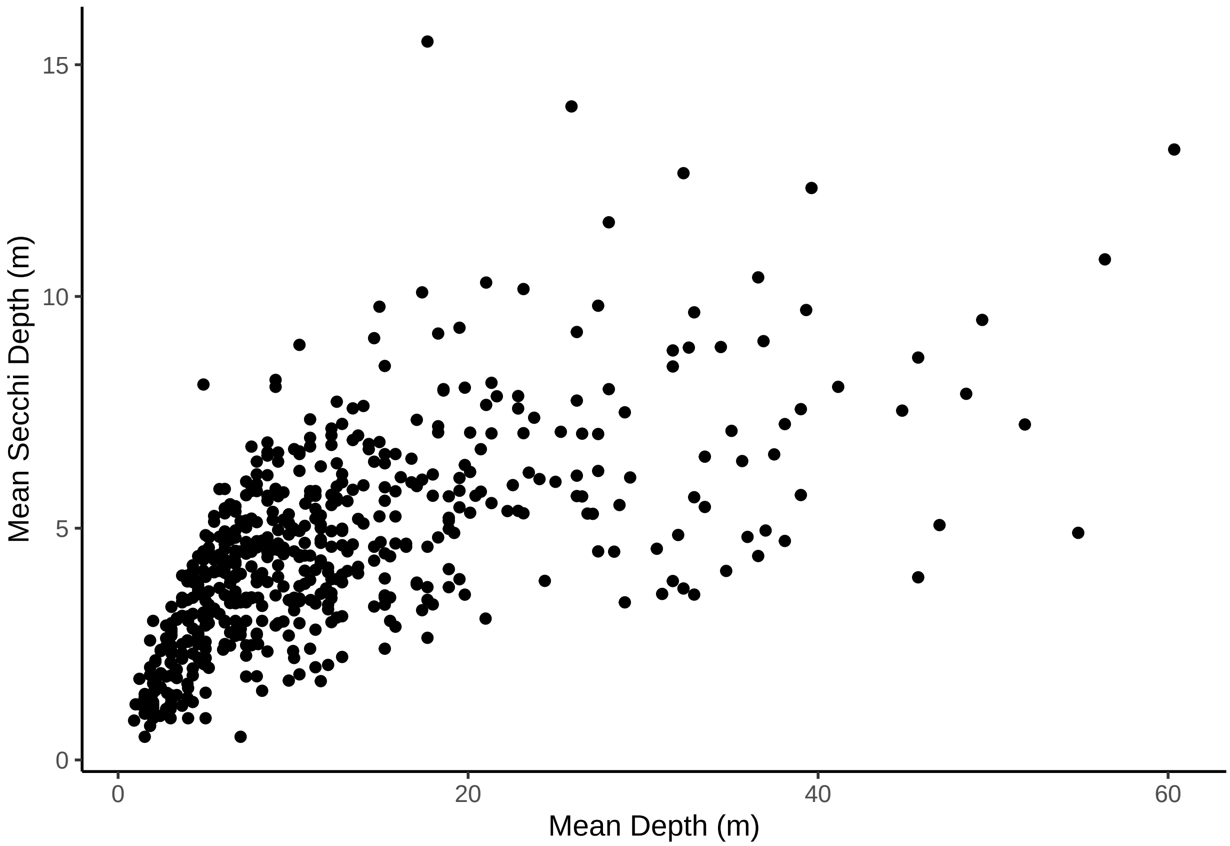
What would you change about the legend to make it a more effective visualization?

> To make this a more effective visualization, I would have one legend showing the average secchi depth, as opposed to having it appear twice. I would also decrease the box sizes and possibly get rid of the squares since the squares do not represent the size of the points.

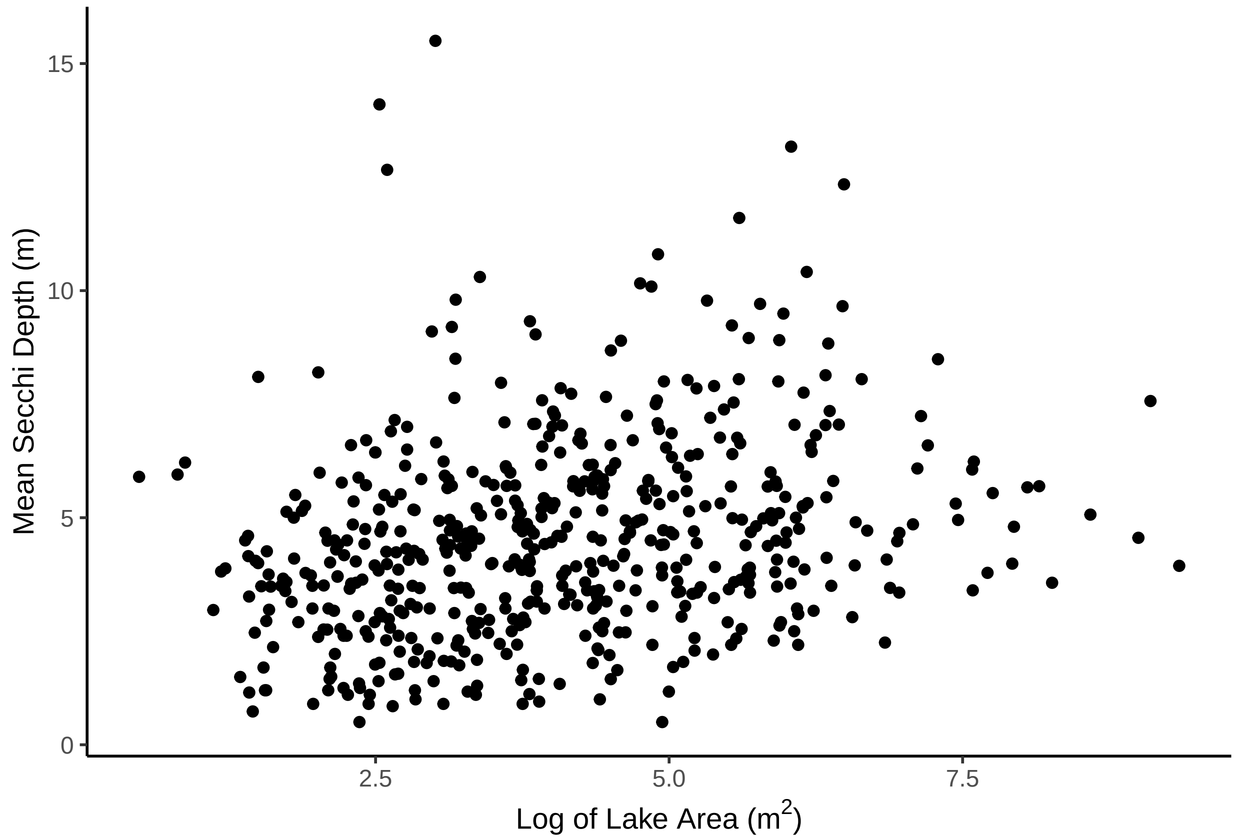
9. What relationships do you see between secchi depth, lake area, and lake depth? Which of the two lake variables seems to be a stronger determinant of secchi depth? (make a scatterplot and run a regression to test this)

Note: consider log-transforming a predictor variable if appropriate

```
#scatter plot of average secchi depth vs depth of lake
ME.depth.secchi <- ggplot(ME.secchi.summary, aes(x = depth, y = secchi.mean)) +
  geom_point() +
  labs(x = "Mean Depth (m)", y = "Mean Secchi Depth (m)")
print(ME.depth.secchi)
```




```
#scatter plot of mean area and mean secchi depth; log transformed area because all datapoints were to t
ggplot(ME.secchi.summary, aes(x = log(area), y = secchi.mean)) +
  geom_point() +
  labs(x = expression("Log of Lake Area (m"^2*")"), y = "Mean Secchi Depth (m)")
```




```r
#running regression on the depth and secchi depth variables
linearMod <- lm(data = ME.secchi.summary, secchi.mean ~ depth + log(area))
print(linearMod)
```

```
##
## Call:
## lm(formula = secchi.mean ~ depth + log(area), data = ME.secchi.summary)
##
## Coefficients:
## (Intercept)        depth    log(area)
##      3.1465       0.1447      -0.0978
```

```r
summary(linearMod)
```

```
##
## Call:
## lm(formula = secchi.mean ~ depth + log(area), data = ME.secchi.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   <2e-16 ***
## depth        0.144728   0.008422  17.184   <2e-16 ***
## log(area)   -0.097802   0.052920  -1.848   0.0651 .
## ---
```

```
## 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: < 2.2e-16
```

From the analysis above, as lake area increases, the average secchi depth decreases. As lake depth increases, the average secchi depth also increases. Between lake depth and area, lake depth seems to be the stronger determinent of secchi depth. This is evidenced in the scatter plots made, where area needed to be log transformed to show a relationship between secchi depth and area. Also, the regression analysis showed that depth was statistically significant in predicting secchi depth, whereas log(area) was not statistically significant.

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.FL <- filter(waterfeatures, STATE == "FL")

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

#filtering for just FL in HUC6 dataset
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")) #check projection

#checking projection of the data
st_crs(HUC6.FL) #already in coordinate reference system 4269
```

```
## Coordinate Reference System:
##   EPSG: 4269
##   proj4string: "+proj=longlat +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +no_defs"
```

```
st_crs(waterfeatures.FL) #does not have a coordinate reference system
```

```
## Coordinate Reference System: NA
```

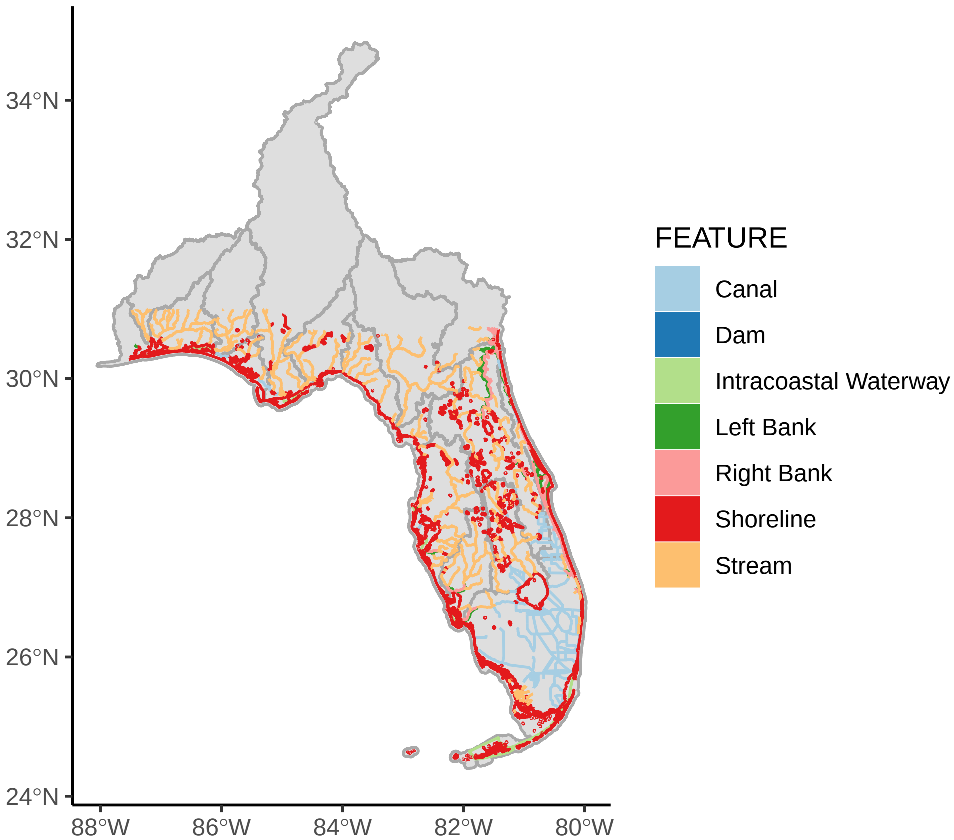
```
#updating coordinate reference system for waterfeatures.FL
waterfeatures.FL <- st_set_crs(waterfeatures.FL, 4269)
st_crs(waterfeatures.FL)
```

```
## Coordinate Reference System:
##   EPSG: 4269
##   proj4string: "+proj=longlat +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +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.

```
#map of watershed boundaries in FL
 ggplot() +
  geom_sf(data = HUC6.FL, fill = "gray", color = "darkgray", alpha = 0.5) +
  geom_sf(data = waterfeatures.FL, aes(color = FEATURE, fill = FEATURE)) +
  scale_fill_brewer(palette = "Paired") +
  scale_color_brewer(palette = "Paired")
```



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

    The dominant water features in Florida are shoreline, streams, canals, and right bank. This is similar to some of the dominant water features in NC. Similarly, FL and NC both have streams, shorelines, and canals as dominant water features. Comparatively, NC has intracoastal highway and left bank as dominant water features that FL does not have, but FL does have right bank as a dominant water feature which NC does not.

Reflection

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

    Summary points about mapping from this lesson is that the size of a dot in a graph can help assess whether there are variables that predict the dependent variable better. Another summary point is that with mapping, you will always want to make sure you have the correct coordinate system assigned to the layer. If the incorrect coordinate system is selected, then the projection of the map will be incorrect.

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

For the data depicting the size of the dot based off of the dependent variable, the plot showing the two maps in Maine support this conclusion. For the coordinate system, the code where I checked for the coordinate system for the water features layer and then corrected the coordinate system because it was not projected correctly.

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

    Yes, hands-on data analysis did impact my learning about mapping compared to theory based. I found that by doing the hands-on analysis, I was better able to understand the trends and patterns, that may not have been as evident without doing the analysis.

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

    Real-world data was what I expected compared to the theory.