Assignment 8: Mapping

Gabi Richichi

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] "/Users/gabriellerichichi/Documents/5th year @ Duke/Stats/Hydrologic_Data_Analysis/Assignments"
```

```
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------------------------------------

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

## -- 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
```

```
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
```

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

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

```
library(maps)
```

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

```
library(viridis)
```

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

```
theme_set(theme_classic())

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

```
## Reading layer `hydrogl020' from data source `/Users/gabriellerichichi/Documents/5th year @ Duke/Stats
## 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
```

```
class(waterfeatures)
```

```
## [1] "sf"         "data.frame"
```

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

```
## Reading layer `WBDHU6' from data source `/Users/gabriellerichichi/Documents/5th year @ Duke/Stats/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 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +no_defs
```

```
LAGOSdata <- lagosne_load()
```

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

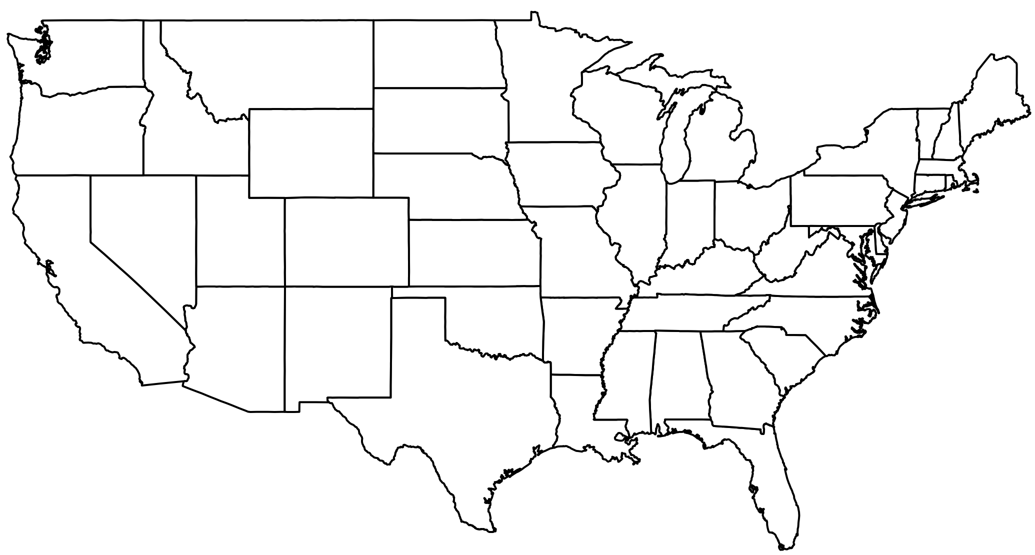
Mapping water quality in lakes

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.

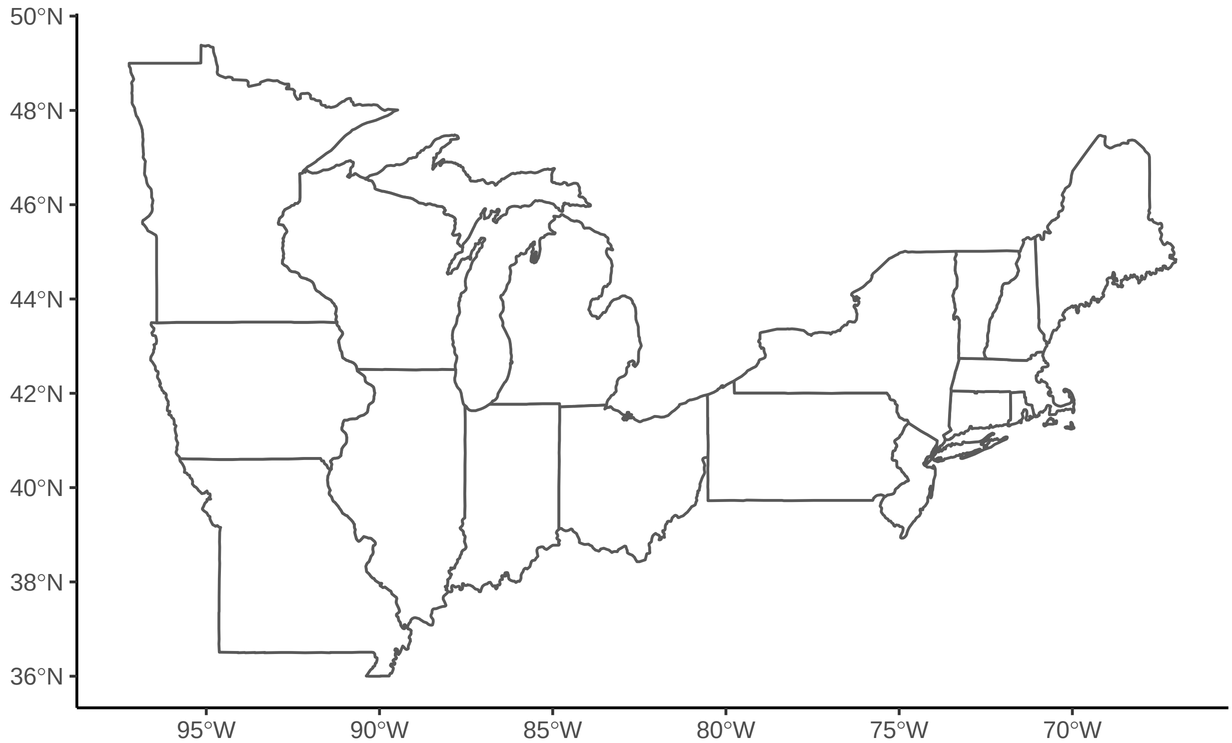
5. Filter the states and secchi depth datasets so that they contain Maine only. For the secchi depth dataset, create a summary dataset with just the mean secchi depth.

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



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

```r
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
#MAINE data

maine <- LAGOScombined %>%
  filter(LAGOScombined$state == "ME")
```

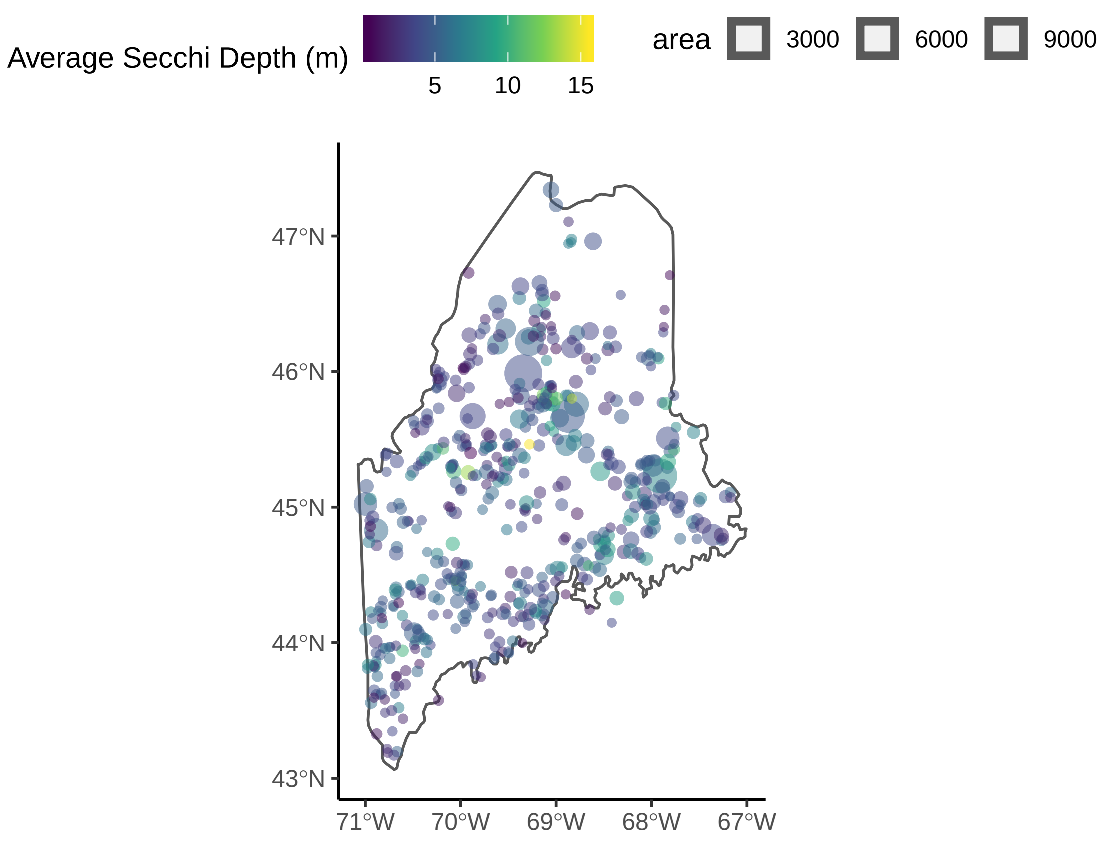
```r
states.subset.maine <- filter(states, ID %in%
                                c("maine"))

secchi.summary.maine <- maine %>%
  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()

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

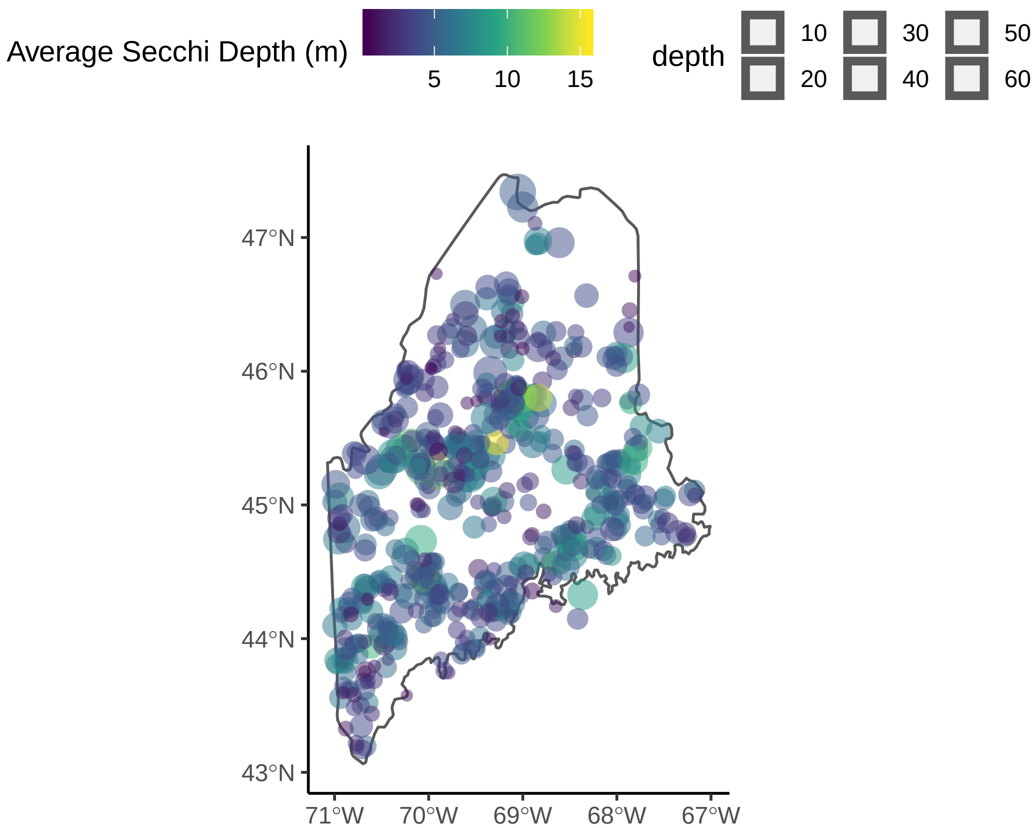
6. Create a plot of mean secchi depth for lakes in Maine, with mean secchi depth designated as color and the lake area as the size of the dot. Remember that you are using size in the aesthetics and should remove the size = 1 from the other part of the code. Adjust the transparency of points as needed.

```r
Secchiplotmaine <- ggplot() +
  geom_sf(data = states.subset.maine, fill = "white") +
  geom_sf(data = secchi.spatial.maine, aes(color = secchi.mean, size = area),
          alpha = 0.5) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi Depth (m)") +
  theme(legend.position = "top")
print(Secchiplotmaine)
```



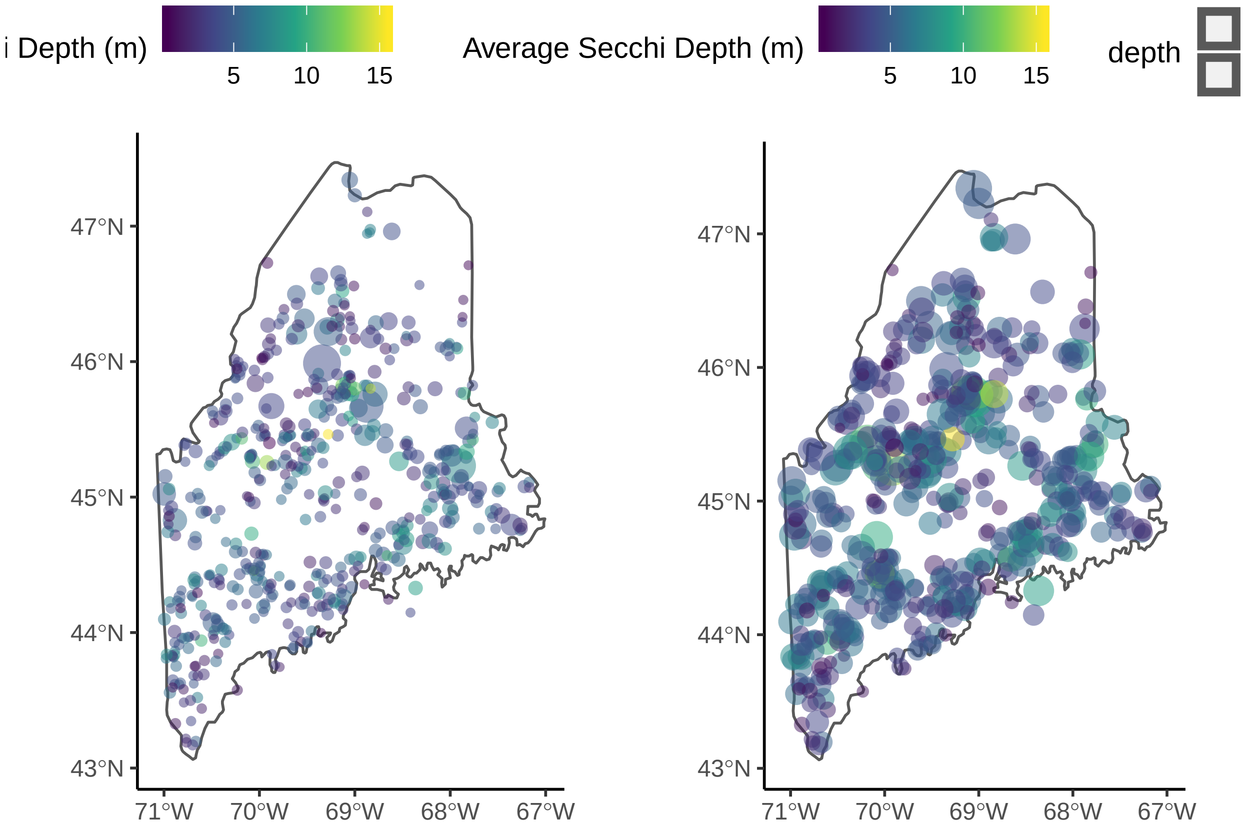
7. Create a second plot, but this time use maximum depth of the lake as the size of the dot.

```
Secchiplotdepth <- ggplot() +
  geom_sf(data = states.subset.maine, fill = "white") +
  geom_sf(data = secchi.spatial.maine, aes(color = secchi.mean, size = depth),
          alpha = 0.5) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi Depth (m)") +
  theme(legend.position = "top")
print(Secchiplotdepth)
```



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).

```
giantplot <- plot_grid(Secchiplotmaine, Secchiplotdepth, axis = c("l"))
print(giantplot)
```



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

I would make the legend smaller so that all of the text could be read. I would also align it better so both legends don't overlap each other.

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

```
DepthTest <- lm(data = secchi.spatial.maine, log(depth) ~ secchi.mean)
summary(DepthTest)
```

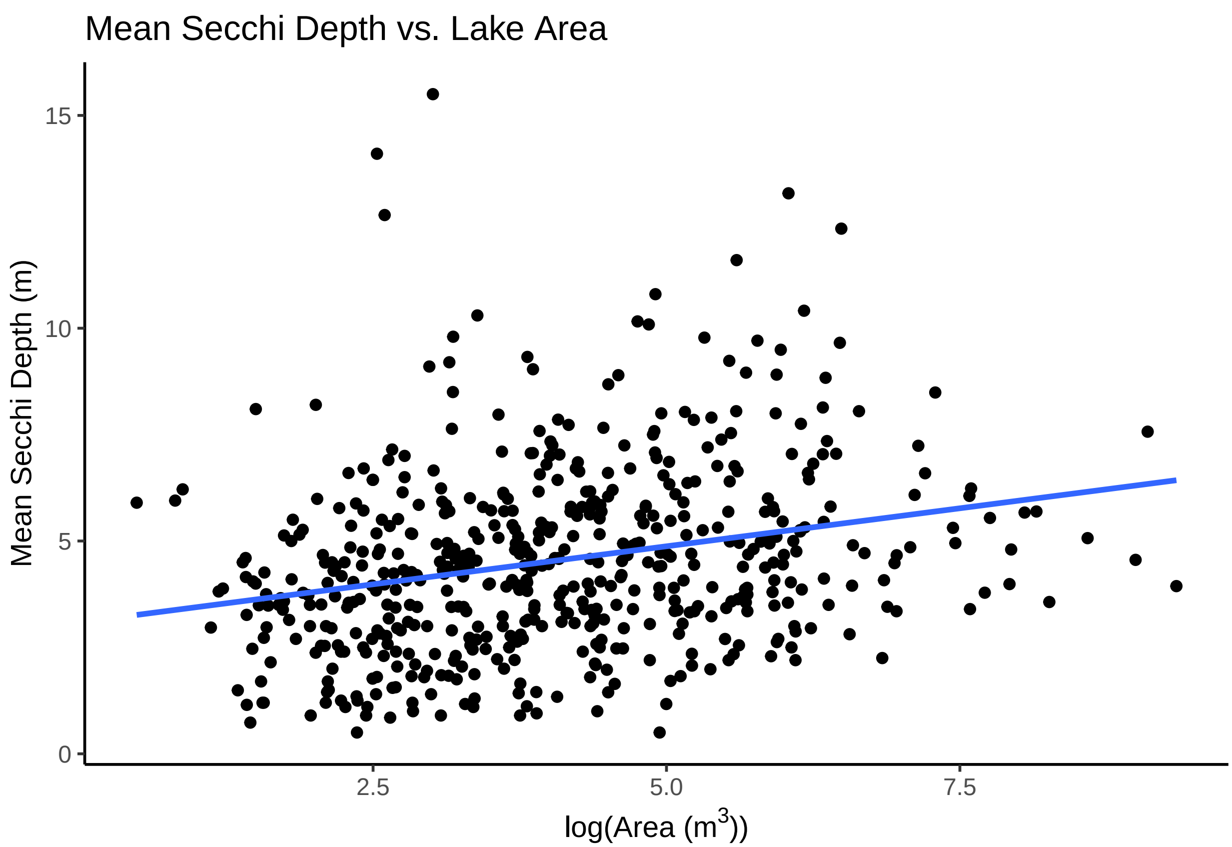
```
##
## Call:
## lm(formula = log(depth) ~ secchi.mean, data = secchi.spatial.maine)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.06070 -0.41905 -0.03564  0.37903  1.74731
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.10120    0.05598   19.67   <2e-16 ***
## secchi.mean  0.24722    0.01112   22.23   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.568 on 547 degrees of freedom
## Multiple R-squared:  0.4747, Adjusted R-squared:  0.4738
## F-statistic: 494.4 on 1 and 547 DF,  p-value: < 2.2e-16
```

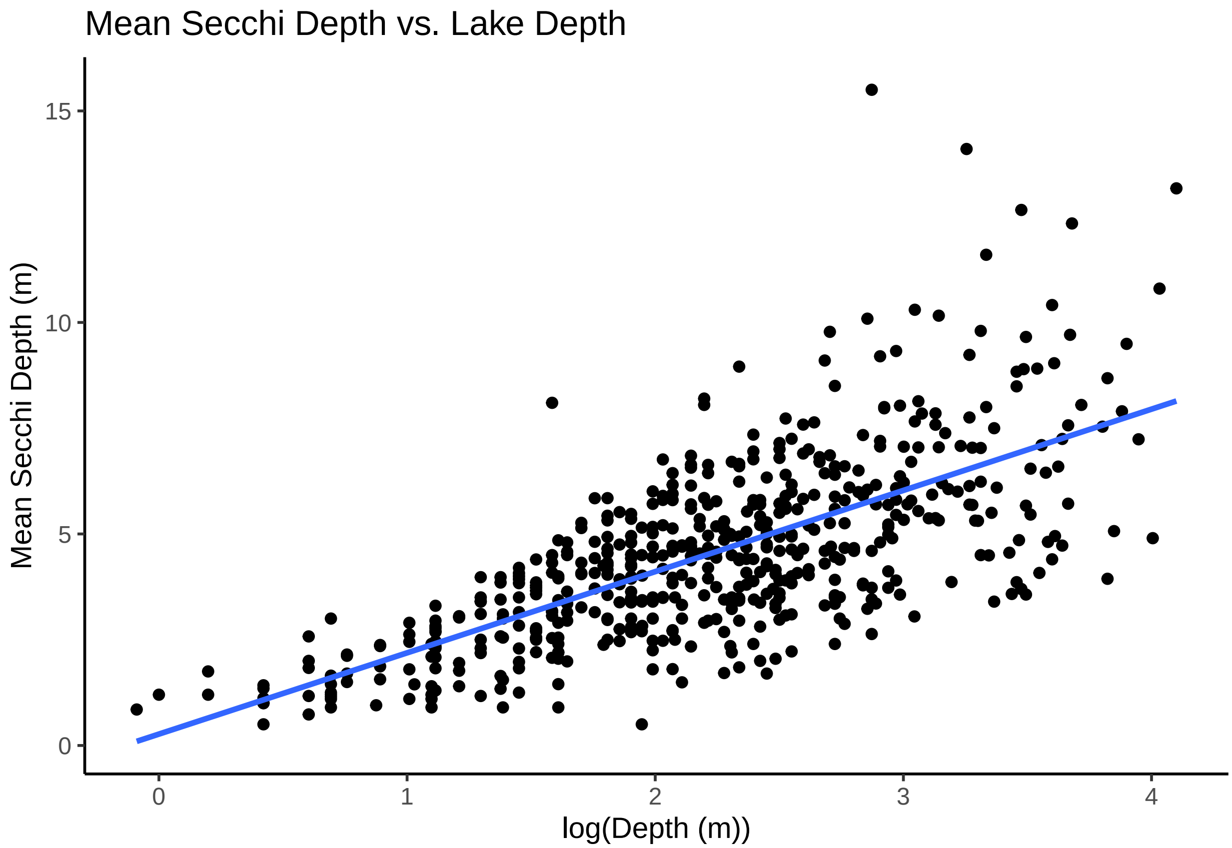
```r
AreaTest <- lm(data = secchi.spatial.maine, log(area) ~ secchi.mean)
summary(AreaTest)
```

```
##
## Call:
## lm(formula = log(area) ~ secchi.mean, data = secchi.spatial.maine)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8252 -1.1238 -0.1443  0.9638  5.4036
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.19817    0.15107  21.170  < 2e-16 ***
## secchi.mean  0.18865    0.03001   6.287 6.62e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.533 on 547 degrees of freedom
## Multiple R-squared:  0.06739,    Adjusted R-squared:  0.06569
## F-statistic: 39.53 on 1 and 547 DF,  p-value: 6.622e-10
```

```r
scatterplotArea <- ggplot(secchi.spatial.maine, aes(x = log(area), y = secchi.mean)) +
  geom_point() +
  labs(x = expression("log(Area (m"^3*"))"), y = "Mean Secchi Depth (m)") +
  ggtitle("Mean Secchi Depth vs. Lake Area") +
  geom_smooth(se = FALSE, method = lm)
print(scatterplotArea)
```




```r
scatterplotDepth <- ggplot(secchi.spatial.maine, aes(x = log(depth), y = secchi.mean)) +
  geom_point() +
  labs(x = "log(Depth (m))", y = "Mean Secchi Depth (m)") +
  ggtitle("Mean Secchi Depth vs. Lake Depth") +
  geom_smooth(se = FALSE, method = lm)
print(scatterplotDepth)
```



There seems to be a stronger relationship between lake depth and mean secchi depth than between lake area and mean secchi depth. The relationship between lake depth and mean secchi depth is a positive one. As lake depth increases so does mean secchi depth. The R-squared value of this log-transformed linear regression is 0.47, whereas the R-squared value of the log-transformed linear regression for area is 0.066. Therefore, there is a closer correlation between lake depth and secchi depth than between lake area and secchi depth.

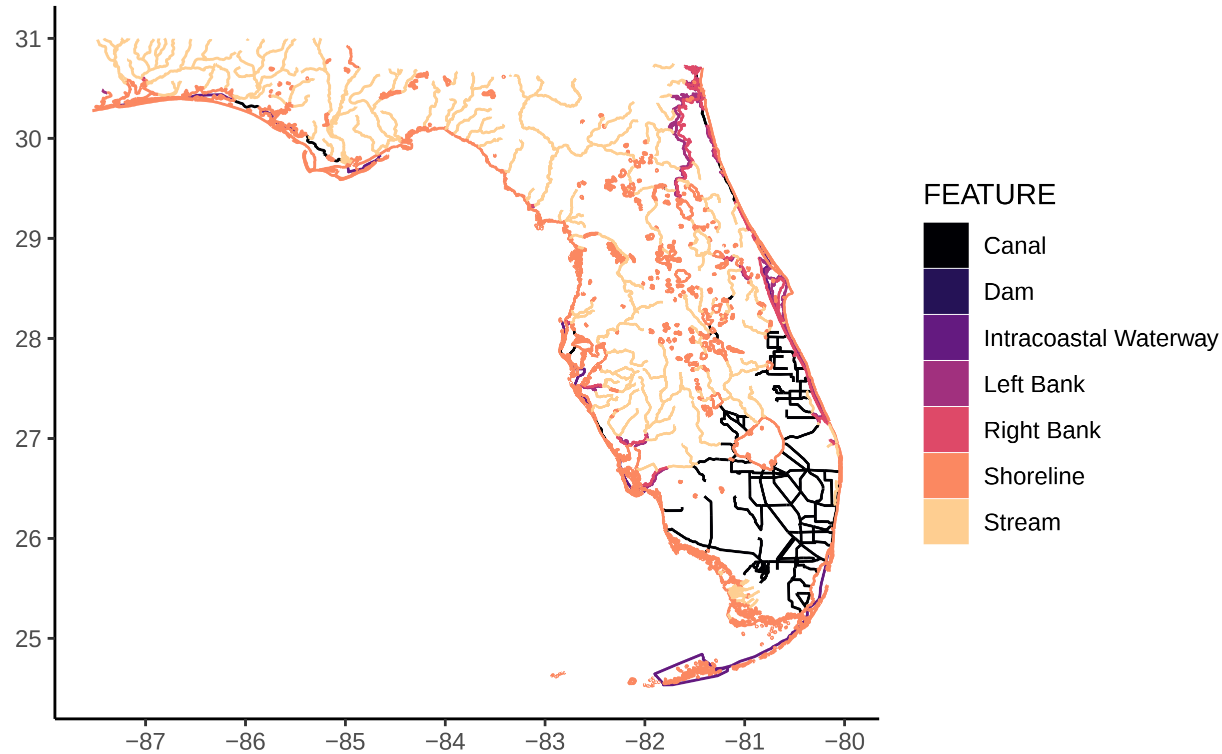
Mapping water features and watershed boundaries

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")

# Filter further
waterfeatures <- filter(waterfeatures, FEATURE != "Apparent Limit" & FEATURE != "Closure Line")

Waterfeaturesplot <-
ggplot(waterfeatures) +
  geom_sf(aes(fill = FEATURE, color = FEATURE)) +
  scale_color_viridis_d(option = "magma", end = 0.9) +
  scale_fill_viridis_d(option = "magma", end = 0.9)
print(Waterfeaturesplot)
```




```r
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
```

```r
HUC6.FL <- HUC6 %>%
  filter(States %in% c("AL,FL", "AL,FL,GA", "FL,GA", "FL"))

st_crs(waterfeatures)
```

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

```r
st_crs(HUC6.FL)
```

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

```r
waterfeatures <- st_set_crs(waterfeatures, 4269)
st_crs(waterfeatures)
```

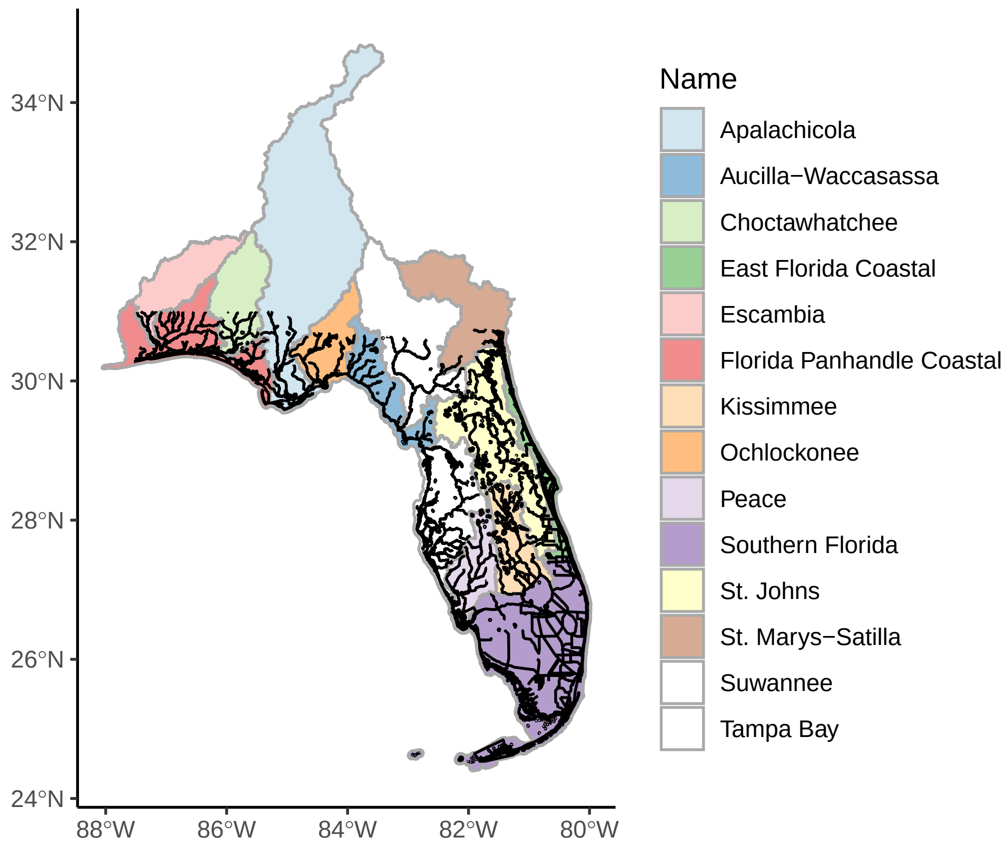
```
## Coordinate Reference System:
##   EPSG: 4269
##   proj4string: "+proj=longlat +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +no_defs"
```

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.

```r
FLlayers <- ggplot() +
  geom_sf(data = HUC6.FL, aes(fill = Name), color = "darkgray", alpha = 0.5) +
  geom_sf(data = waterfeatures) +
  scale_fill_brewer(palette = "Paired")
print(FLlayers)
```

```
## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum for palette Paired is 12
## Returning the palette you asked for with that many colors
```



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 coastally-located watersheds, such as bays and estuaries. North Carolina had more inland lakes and rivers, in addition to coastal waterbodies.

Reflection

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

    1. The LAGOS database has information on all the states in the US with regard to lake information. The geographical information that this database provides also aids in mapping.
    2. The HUC6 database contains watershed data that can be mapped.
    3. Color palettes can really aid in data visualization, in particular with mapping and mapping watersheds.

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

The watershed mapping exercise helped me to understand the role data visualization can have on conveying information. It was really enlightening to see all the watersheds outlined and colored. I felt like I was learning new information I hadn't already known about the environment of Florida.

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

    Yes, for sure. Hand-on learning always helps.

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

    I expected Florida to have watersheds near its coasts. I also expected secchi depth and lake depth to be somewhat related.