Assignment 8: Mapping

Walker Grimshaw

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.

```
knitr::opts_chunk$set(warning = FALSE)

getwd()
```

```
## [1] "C:/Users/walke/OneDrive/Documents/Duke/Courses/Fall_2019/Hydrologic_Data_Analysis/Assignments"
```

```
packages <- c("tidyverse",
              "cowplot",
              "LAGOSNE",
              "sf",
              "lubridate",
              "maps",
              "viridis")
invisible(lapply(packages, library, character.only = TRUE))

theme_set(theme_bw())
```

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 into R session
LAGOSdata <- lagosne_load()

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

# Join data frames 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")
```

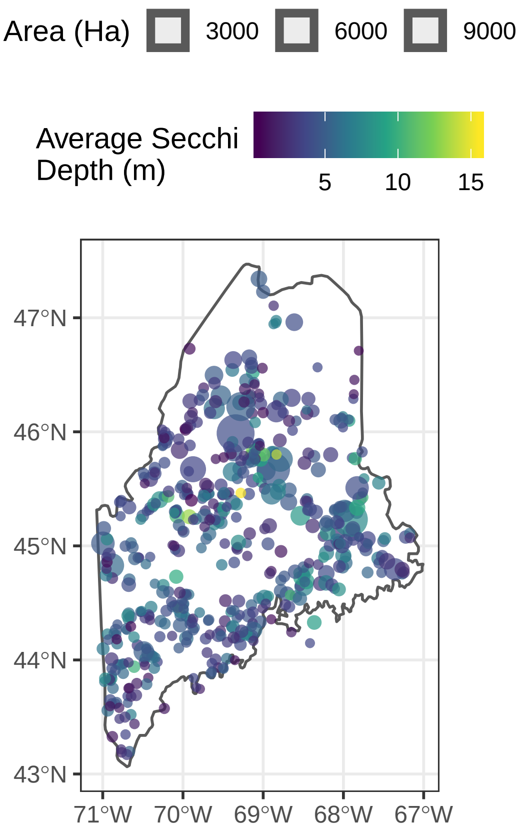
```
## Joining, by = "state_zoneid"
```

```r
# secchi for maine
secchiMaine <- LAGOScombined %>%
  filter(state == "ME") %>%
  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()

# Make secchi data for Maine a spatial object
secchiMaine_spatial <- st_as_sf(secchiMaine, coords = c("long", "lat"), crs = 4326)

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

# state outline of Maine
Maine <- filter(states, 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.

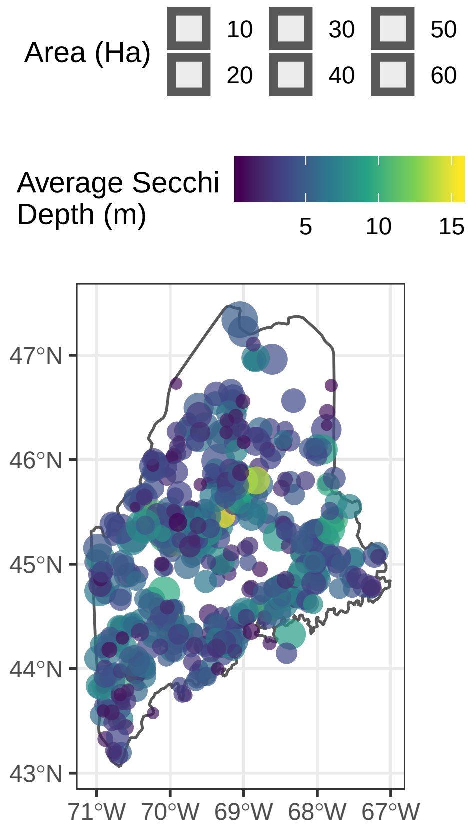
```r
# Secchi vs Area Map
SecchiArea <- ggplot() +
  geom_sf(data = Maine, fill = "white") +
```

```r
  geom_sf(data = secchiMaine_spatial, aes(color = secchi.mean, size = area),
          alpha = 0.7) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi\nDepth (m)",
       size = "Area (Ha)") +
  theme(legend.position = "top", legend.box = "vertical")
print(SecchiArea)
```



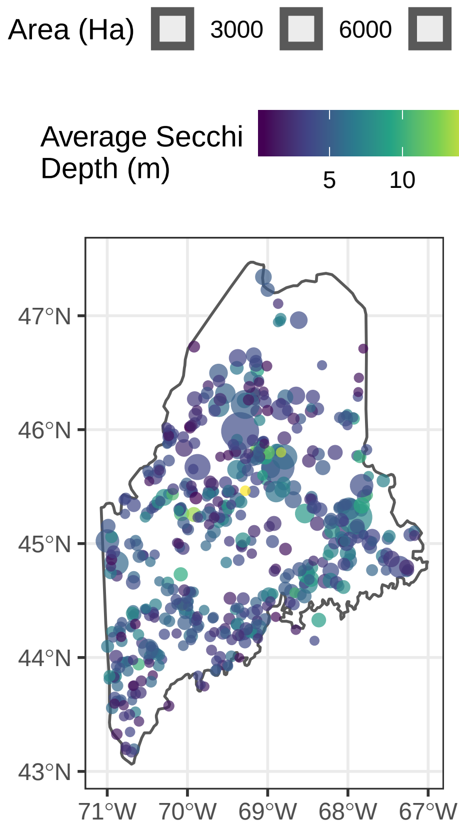
7. Create a second plot, but this time use maximum depth of the lake as the size of the dot.

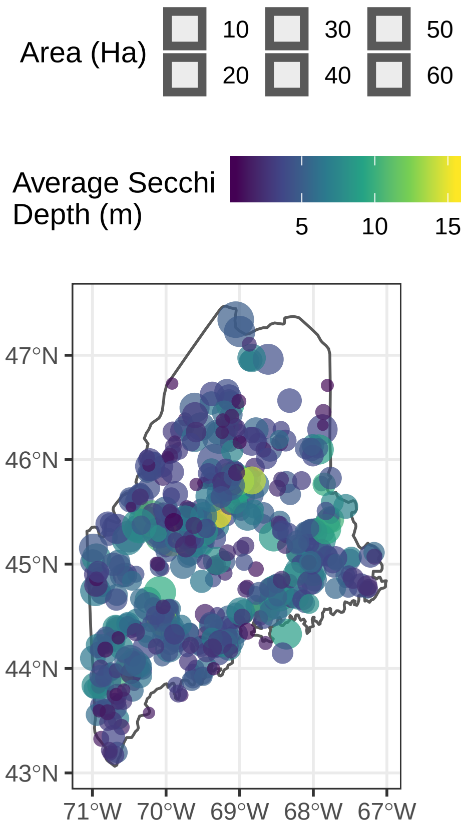
```r
# Secchi vs Depth Map
SecchiDepth <- ggplot() +
  geom_sf(data = Maine, fill = "white") +
  geom_sf(data = secchiMaine_spatial, aes(color = secchi.mean, size = depth),
          alpha = 0.7) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi\nDepth (m)",
       size = "Area (Ha)") +
  theme(legend.position = "top", legend.box = "vertical")
print(SecchiDepth)
```



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).

```
plot_grid(SecchiArea, SecchiDepth, nrow = 1)
```





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

> The legend does not show how the size of the points correspond to either lake area or maximum lake depth. Though we know larger dots correspond to larger or deeper lakes, this is not reflected in the legend. Both the direction of this relationship and ideally the actual dot sizes should be in the legend. The majority of secchi depths are also below approximately 7 meters. Rather than having a linear color-secchi depth relationship, the visualization may be more effective if the legend had a geometric relationship.

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)

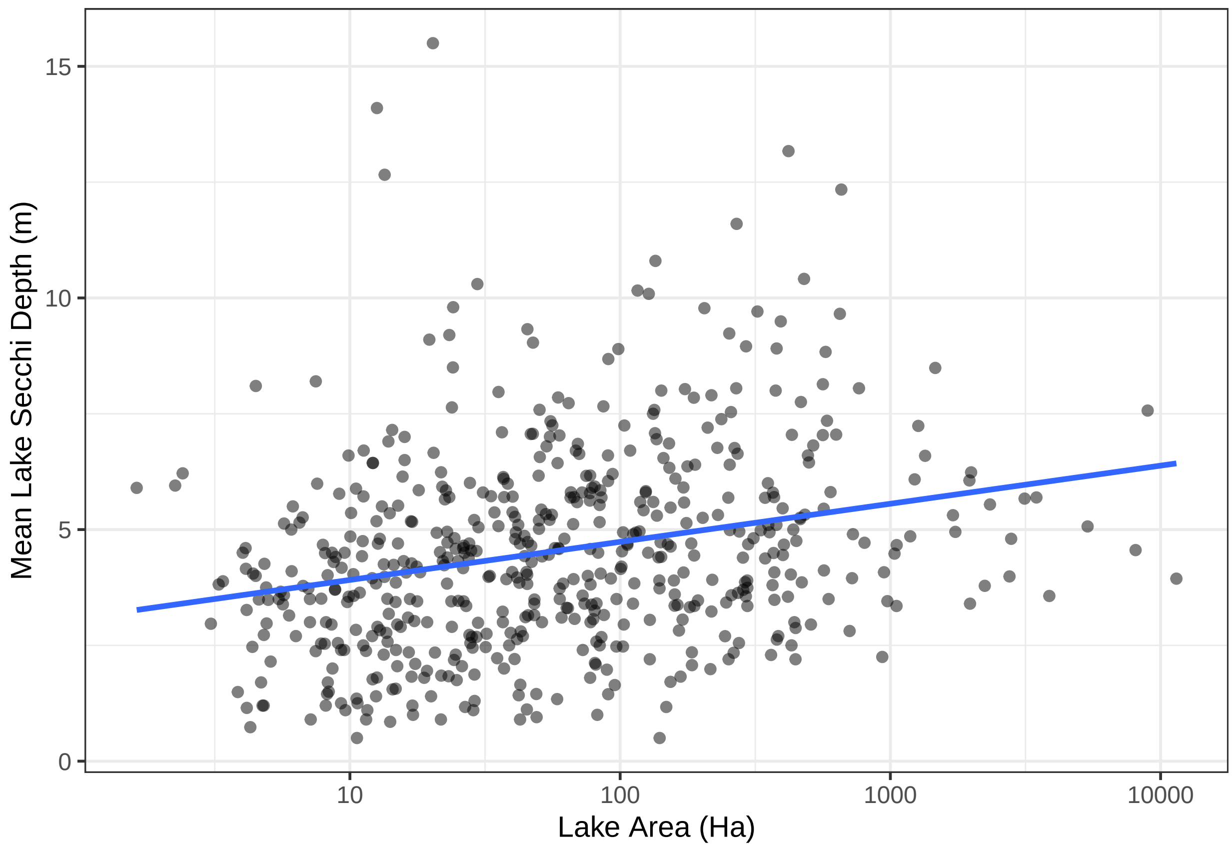
Note: consider log-transforming a predictor variable if appropriate

```
## Secchi depth vs Area lm and plot, log transform area
lm.SecchiArea <- lm(data = secchiMaine, secchi.mean ~ log10(area))

summary(lm.SecchiArea)
```

```
##
## Call:
## lm(formula = secchi.mean ~ log10(area), data = secchiMaine)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.355 -1.492 -0.167  1.145 11.335
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0898     0.2473  12.492  < 2e-16 ***
## log10(area)   0.8226     0.1308   6.287 6.62e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.109 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
```

```
ggplot(secchiMaine, aes(x = area, y = secchi.mean)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", se = F) +
  scale_x_log10() +
  labs(x = "Lake Area (Ha)", y = "Mean Lake Secchi Depth (m)")
```

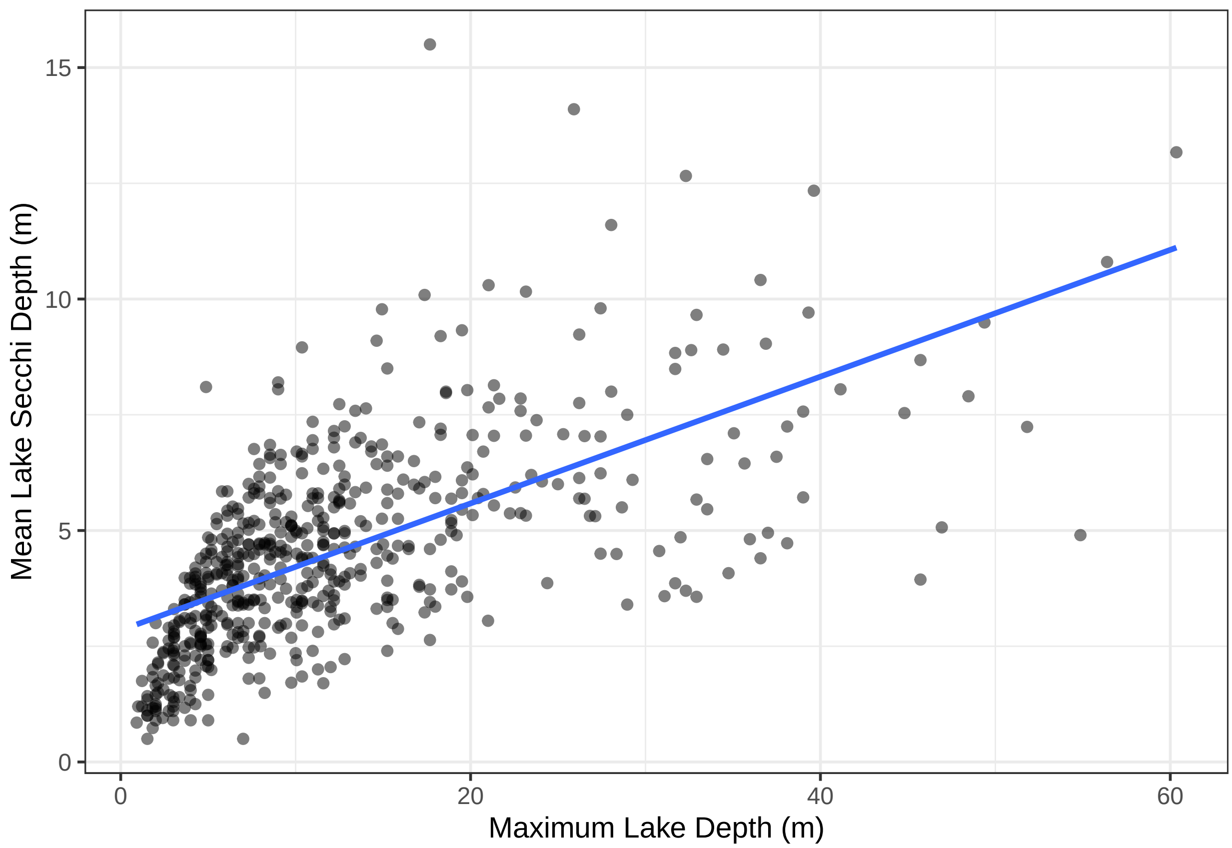



```
## Secchi depth vs Max depth lm and plot
lm.SecchiDepth <- lm(data = secchiMaine, secchi.mean ~ depth)

summary(lm.SecchiDepth)
```

```
##
## Call:
## lm(formula = secchi.mean ~ depth, data = secchiMaine)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -5.4592 -1.1484 -0.1410  0.9497 10.2329
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.846212   0.115959   24.55   <2e-16 ***
## depth       0.136939   0.007308   18.74   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.705 on 547 degrees of freedom
## Multiple R-squared:  0.3909, Adjusted R-squared:  0.3898
## F-statistic: 351.1 on 1 and 547 DF,  p-value: < 2.2e-16
```

```r
ggplot(secchiMaine, aes(x = depth, y = secchi.mean)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", se = F) +
  labs(x = "Maximum Lake Depth (m)", y = "Mean Lake Secchi Depth (m)")
```



In the maps, secchi depth appears to have a stronger relationship with maximum lake depth than lake area. Upon creating linear models of each relationship, both lake area and maximum lake depth have statistically significant impact on mean lake secchi depth. A direct comparison between the magnitude of the relationships is useless because there is a much greater range of lake area than lake depth. However, the adjusted R^2 of the area relationship, 0.066, is much smaller than the adjusted R^2 of the depth relationship, 0.39, indicating maximum lake depth is a better predictor of mean lake 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.

```
## Read in national water features data and filter for Florida Data
waterfeatures <- st_read("../Data/Raw/hydrogl020.dbf")
```

```
## Reading layer `hydrogl020' from data source `C:\Users\walke\OneDrive\Documents\Duke\Courses\Fall_201
## 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
```

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

## Read in HUC6 watershed boundaries
HUC6 <- st_read("../Data/Raw/Watersheds_Spatial/WBDHU6.dbf")
```

```
## Reading layer `WBDHU6' from data source `C:\Users\walke\OneDrive\Documents\Duke\Courses\Fall_2019\Hy
## 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
```

```
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
```

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

## Check coordinate systems of two datasets
st_crs(FLwater) ## need to set projection
```

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

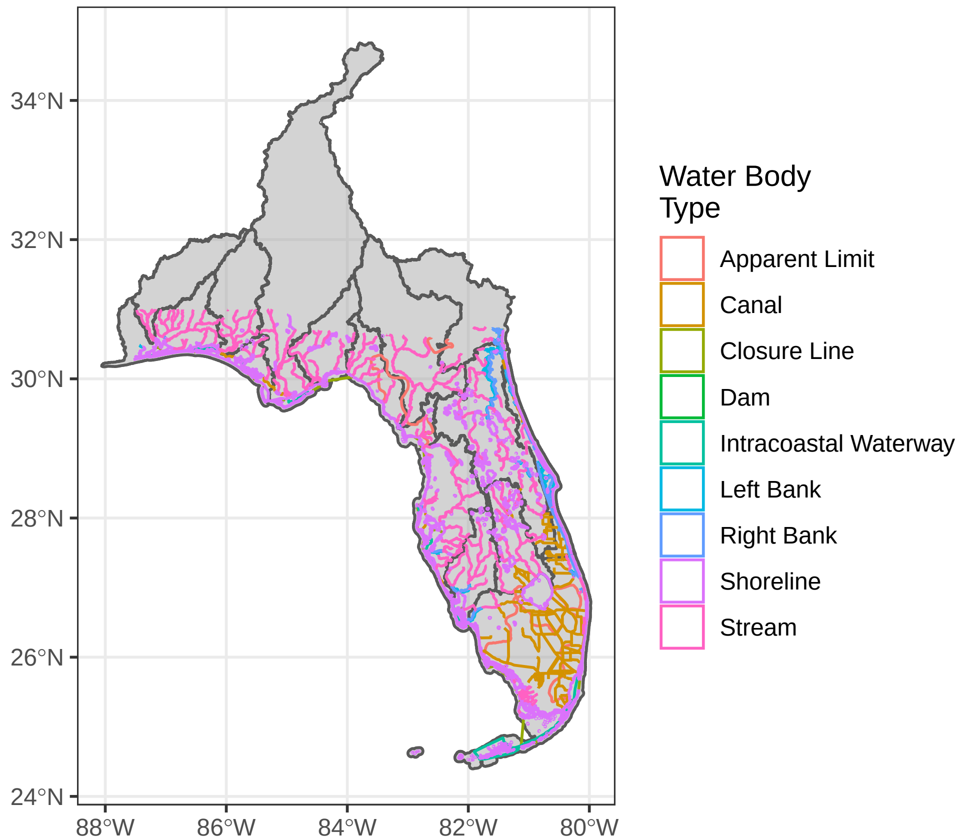
```
st_crs(HUC6.FL)
```

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

```
## define projection for water features
FLwater <- st_set_crs(FLwater, 4269)
```

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
ggplot() +
  ## first draw watersheds
  geom_sf(data = HUC6.FL, fill = "darkgray", alpha = 0.5) +
  geom_sf(data = FLwater, aes(color = FEATURE), fill = NA) +
  scale_fill_brewer(palette = "Paired") +
  labs(color = "Water Body\nType")
```



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

    The primary water feature types appear to be shoreline, stream, and canal mostly in Southern Florida.

Reflection

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

    While viewing the spatial distribution of parameters may help for some qualitative measurements, it may also be very difficult to see relationships among other parameters when showing geography as well. However, maps may help the audience place scientific concepts in a real world context more easily.

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

    The Maine example shows the difficulty of drawing statistical relationships from a map, though the apparent relationship between maximum depth and secchi depth was verified by a linear model.

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

    As was shown through our discussions of different projections, it is difficult to explain many spatial or mapping concepts without hands on practice, so it was much better to have practice as opposed to just theory.

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

    My expectations were met by the real world data for this assignment and corresponding lessons.