

Assignment 8: Mapping

Keqi He

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.pdf”) 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/kh382/Documents/Hydrologic_Data_Analysis"
```

```
#install.packages("tidyverse")
#install.packages("lubridate")
#install.packages("cowplot")
#install.packages("LAGOSNE")
#install.packages("sf")
#install.packages("maps")
#install.packages("viridis")
```

```
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   1.0.0    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 3.6.1, GDAL 2.2.3, PROJ 4.9.3
library(maps)

##
## Attaching package: 'maps'
## The following object is masked from 'package:purrr':
##
##      map
library(viridis)

## Loading required package: viridisLite
theme_set(theme_classic())

load(file = "./Data/Raw/LAGOSdata.rda")
waterfeatures <- st_read("./Data/Raw/hydrogl020.dbf")

## Reading layer `hydrogl020' from data source `C:\Users\kh382\Documents\Hydrologic_Data_Analysis\Data\
## 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\kh382\Documents\Hydrologic_Data_Analysis\Data\Raw\
## 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

```

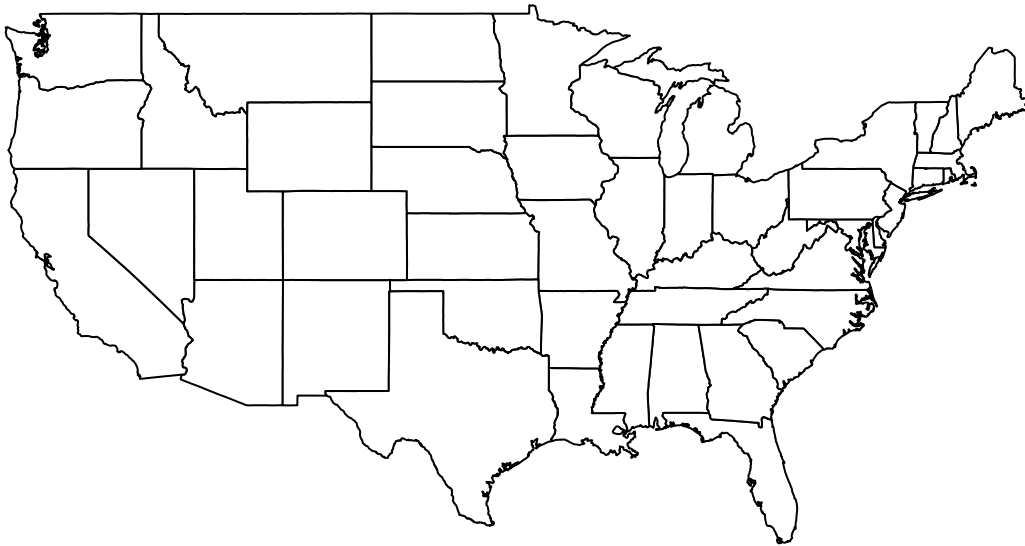
```
## epsg (SRID):      4269
## proj4string:      +proj=longlat +datum=NAD83 +no_defs
```

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.

```
states <- st_as_sf(map(database = "state", plot = TRUE, fill = TRUE, col = "white"))
```



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

```
# 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) %>% # R will choose which column to join by
  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"
secchi.maine <- filter(LAGOScombined, state %in%
  c("ME"))

secchi.summary.maine <- secchi.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)



```

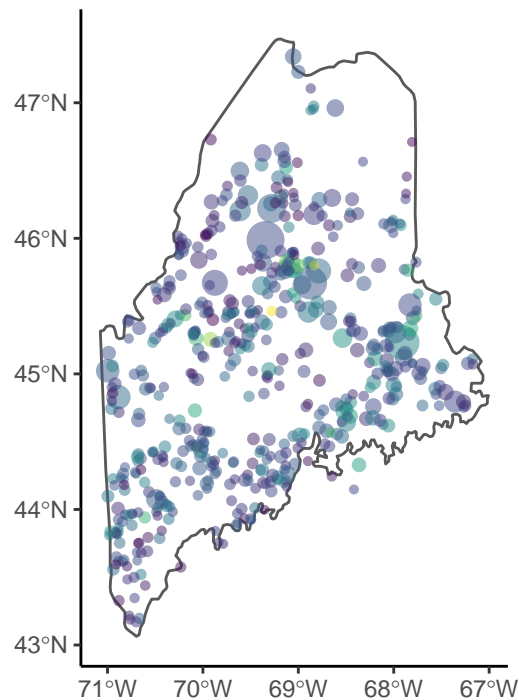
6. Create a plot of mean secchi depth for lakes in Maine, with mean secchi depth designated as color and the lake area as the size of the dot. Remember that you are using size in the aesthetics and should remove the size = 1 from the other part of the code. Adjust the transparency of points as needed.

```

Secchiplot.maine2 <- ggplot() +
  geom_sf(data = state.maine, fill = "white") +
  geom_sf(data = secchi.spatial.maine, aes(size = area, color = secchi.mean),
    alpha = 0.5) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi Depth (m)") +
  theme(legend.position = "top")
print(Secchiplot.maine2)

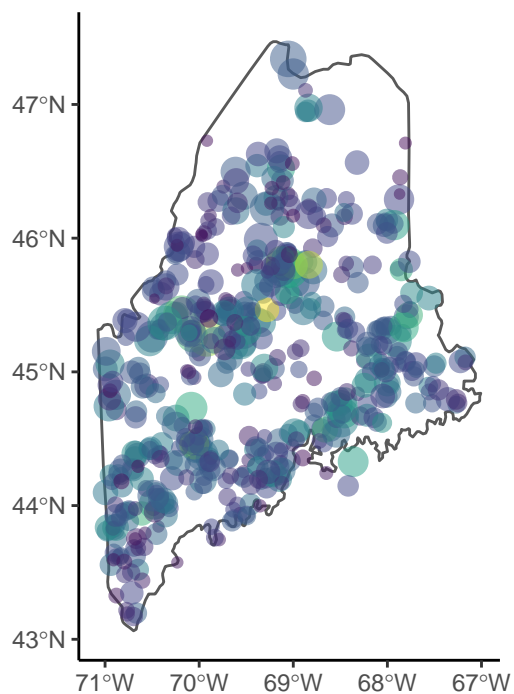
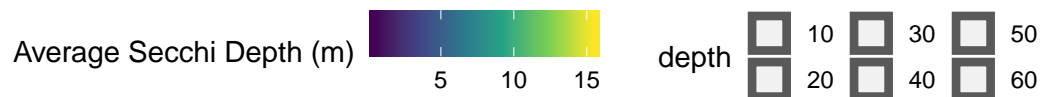
```

Average Secchi Depth (m)  area 



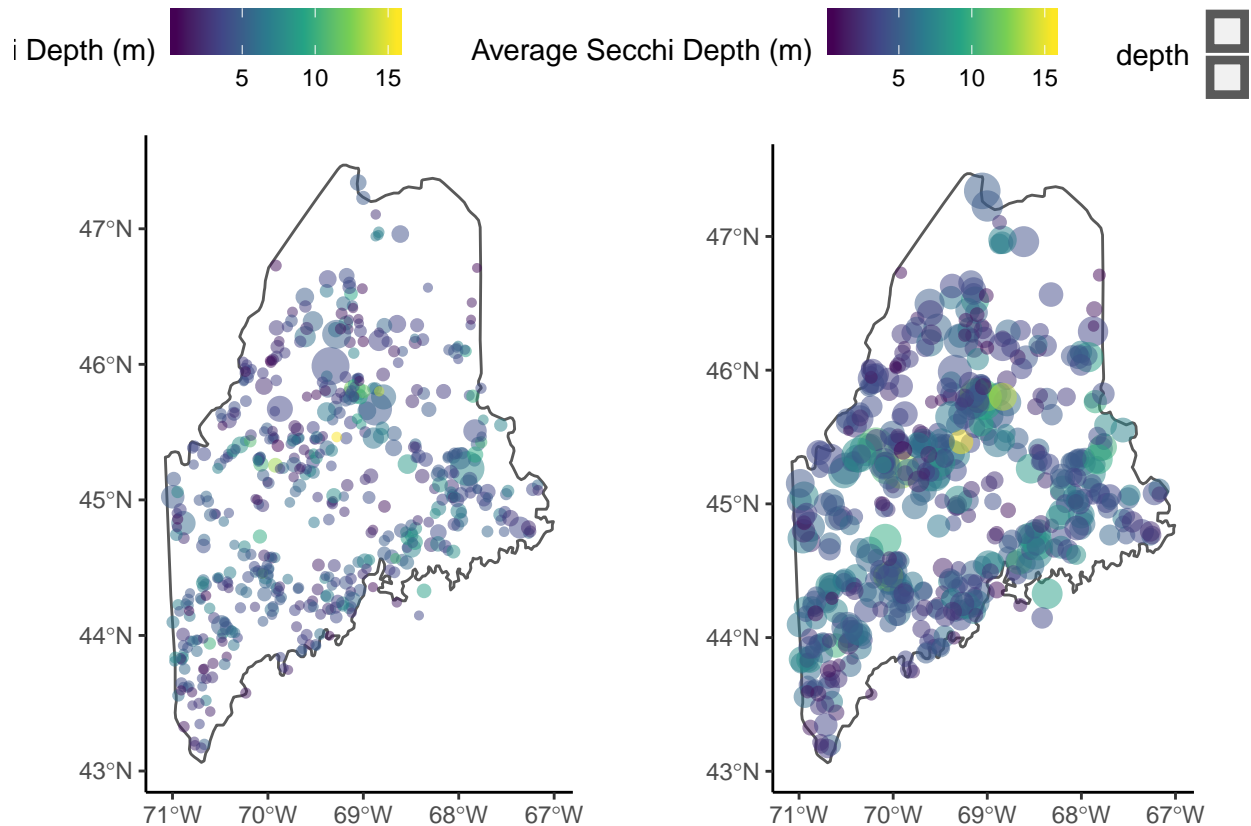
7. Create a second plot, but this time use maximum depth of the lake as the size of the dot.

```
Secchiplot.maine3 <- ggplot() +
  geom_sf(data = state.maine, fill = "white") +
  geom_sf(data = secchi.spatial.maine, aes(size = depth, color = secchi.mean),
    alpha = 0.5) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi Depth (m)") +
  theme(legend.position = "top")
print(Secchiplot.maine3)
```



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(Secchiplot.maine2,Secchiplot.maine3)
```



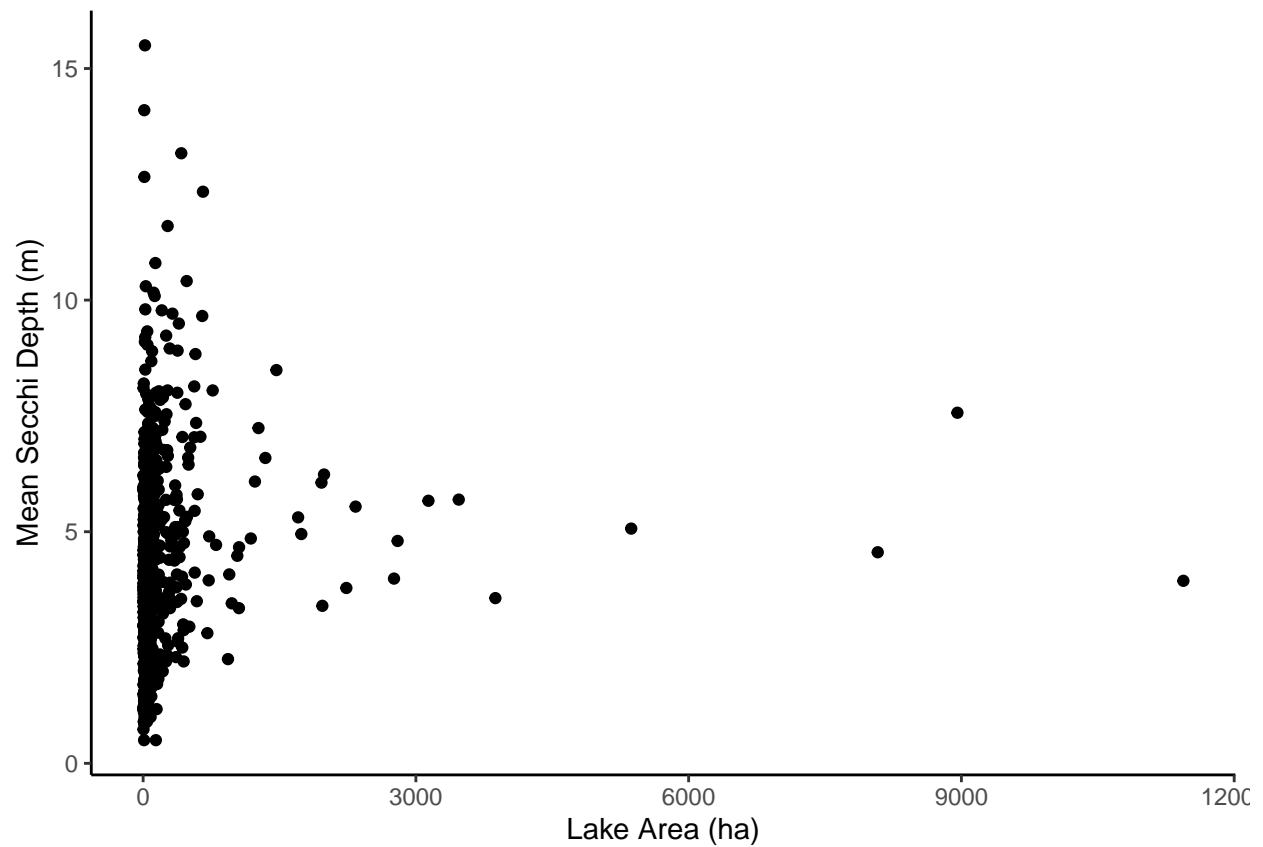
What would you change about the legend to make it a more effective visualization?

I will change the type of legend of the lake area and maximum depth to a series of concentric circles if it's possible. Different radii represent different lake areas or maximum depth. Larger radii mean bigger lake areas or deeper maximum depth. Dots that represent same lake areas or maximum depth will have same radii.

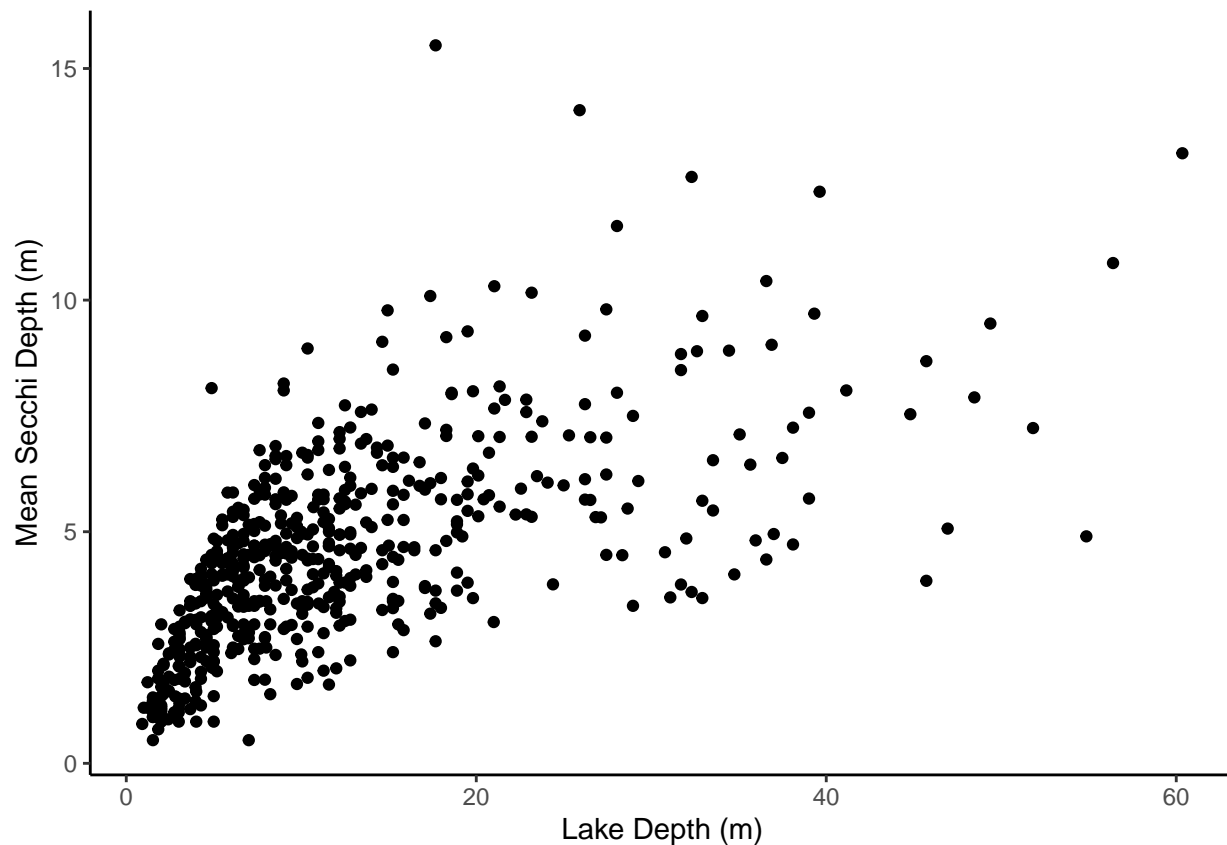
9. What relationships do you see between secchi depth, lake area, and lake depth? Which of the two lake variables seems to be a stronger determinant of secchi depth? (make a scatterplot and run a regression to test this)

Note: consider log-transforming a predictor variable if appropriate

```
secchidepth.lakeareaplot <-
  ggplot(secchi.spatial.maine, aes(x = area, y = secchi.mean)) +
  geom_point() +
  labs(x = "Lake Area (ha)", y = "Mean Secchi Depth (m)")
print(secchidepth.lakeareaplot)
```



```
secchidepth.lakedepthplot <-  
  ggplot(secchi.spatial.maine, aes(x = depth, y = secchi.mean)) +  
  geom_point() +  
  labs(x = "Lake Depth (m)", y = "Mean Secchi Depth (m)")  
print(secchidepth.lakedepthplot)
```

```
secchidepth.lakearea <- lm(data = secchi.spatial.maine, secchi.mean ~ area)
summary(secchidepth.lakearea)
```

```
##
## Call:
## lm(formula = secchi.mean ~ area, data = secchi.spatial.maine)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.0145 -1.4949 -0.1905  1.1919 11.0120
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.4835331  0.0968202  46.308  <2e-16 ***
## area          0.0002215  0.0001109   1.998   0.0462 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.176 on 547 degrees of freedom
## Multiple R-squared:  0.007247, Adjusted R-squared:  0.005432
## F-statistic: 3.993 on 1 and 547 DF, p-value: 0.04618
```

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

```
##
## Call:
```

```
## lm(formula = secchi.mean ~ depth, data = secchi.spatial.maine)
##
## 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
```

There may be linear relationships between secchi depth verse lake area and secchi depth verse lake depth. Lake depth seems to be a stronger determinant of secchi depth for its p-value is much smaller.

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.

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

```
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(waterfeatures.FL)
```

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

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

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

```
## EPSG: 4269
```

```
## proj4string: "+proj=longlat +datum=NAD83 +no_defs"
```

```
waterfeatures.FL <- waterfeatures.FL %>% st_set_crs(st_crs(HUC6.FL))
st_crs(waterfeatures.FL)
```

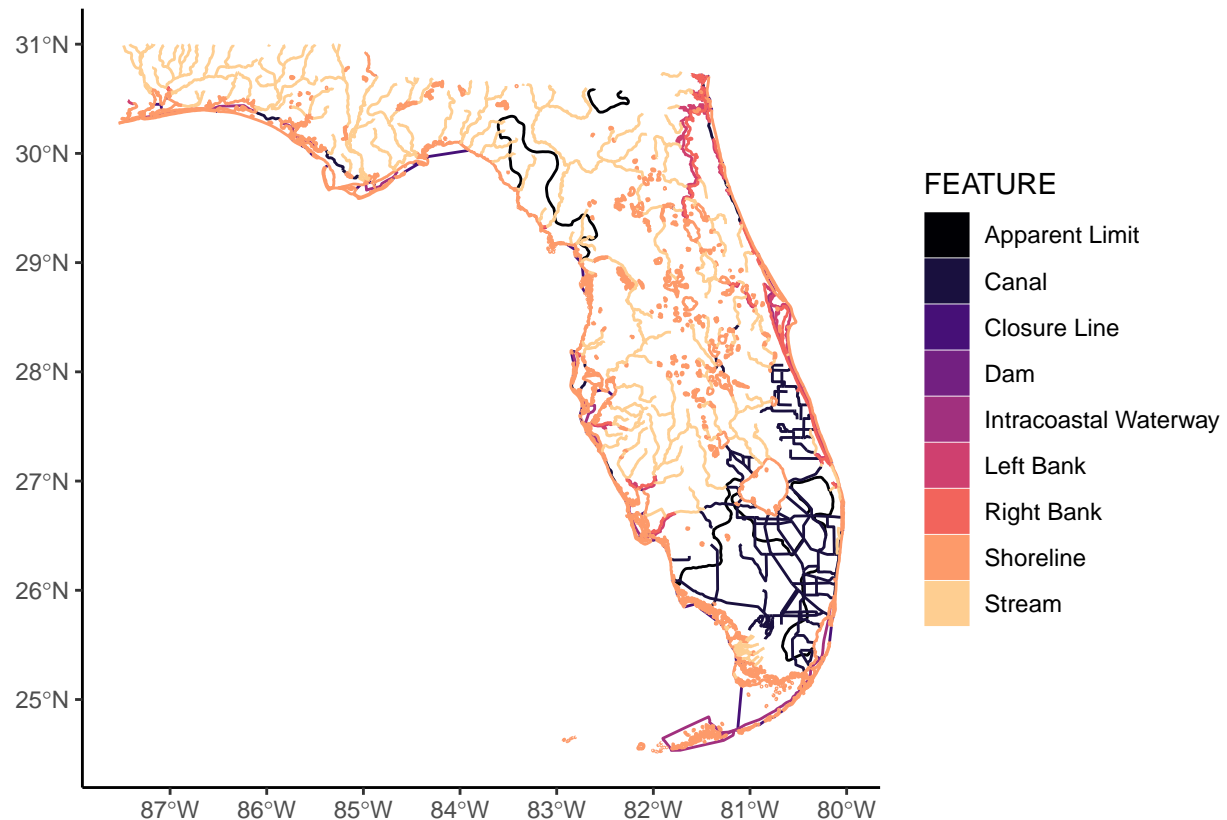
```
## Coordinate Reference System:
```

```
## EPSG: 4269
```

```
## proj4string: "+proj=longlat +datum=NAD83 +no_defs"
```

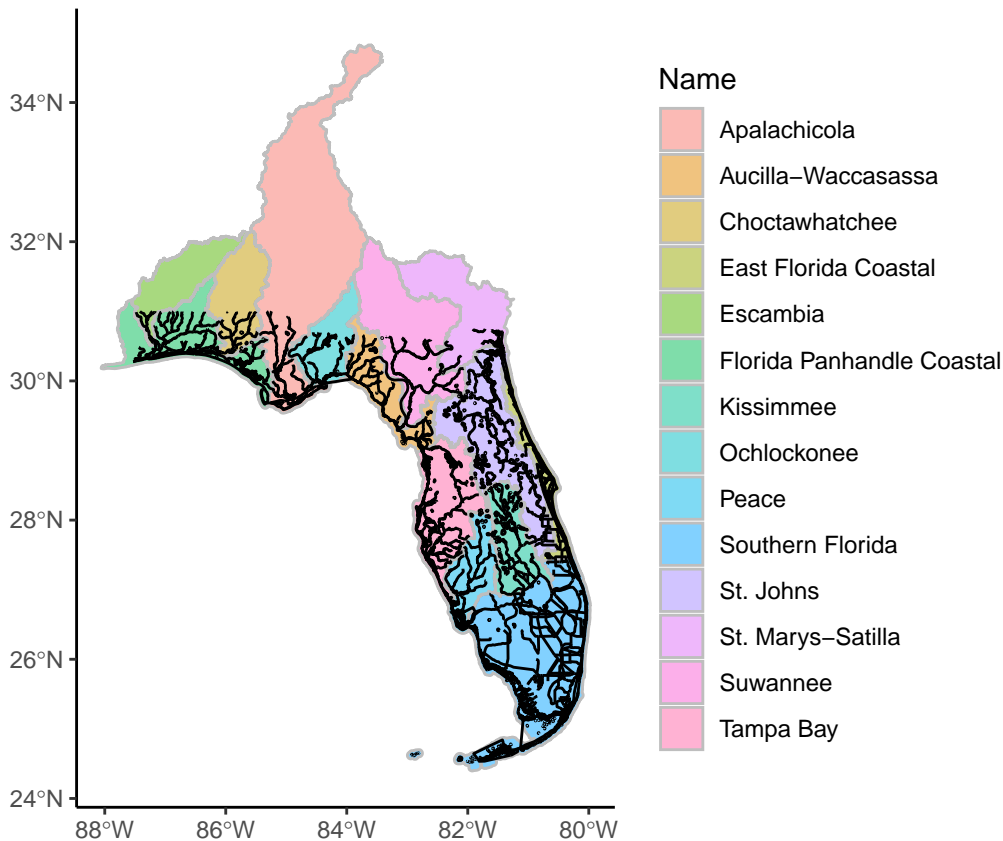
```
Waterfeaturesplot.FL <-
  ggplot(waterfeatures.FL) +
```

```
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.FL)
```



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, aes(fill = Name), color = "gray", alpha = 0.5) +
  geom_sf(data = waterfeatures.FL)
print(FLlayers)
```



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

Stream and apparent limit are the dominant water features in Florida. There is high density but almost even distribution of streams in North Carolina. However, northern part of Florida is mainly occupied by streams while southern part of Florida is mainly occupied by apparent limit. In addition, stream and apparent limit in Florida is not evenly distributed.

Reflection

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

The geometry of a given spatial data point consists of coordinates in 2-, 3-, or 4-dimensional space. These dimensions are: longitude, latitude, altitude (optional), measurement (optional). A feature (object) most often falls into one of these three categories: point, line and polygon.

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

Data used includes the lagos database, the USA rivers water features shape data, and the HUC6 watershed shape data. `sf` ("simple features"), `st_as_sf`, `ggplot`, `geom_sf` are used to draw those conclusions.

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 can help me understand the theory and master some useful tools of time series analysis better. It's amazing to draw nice figures by myself and I will remember it better.

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

In mapping, the real-world data is quite in line with my expectations from theory.