Assignment 8: Mapping

Yixin Wen

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/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Assignments
```

```
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
```

```
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())

load(file = "/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Data/L
waterfeatures <- st_read("/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analysis/Hydrologic_Data_An
```

```
## Reading layer `hydrogl020' from data source `/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analy
## 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("/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Da
```

```
## Reading layer `WBDHU6' from data source `/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analysis/
## 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
```

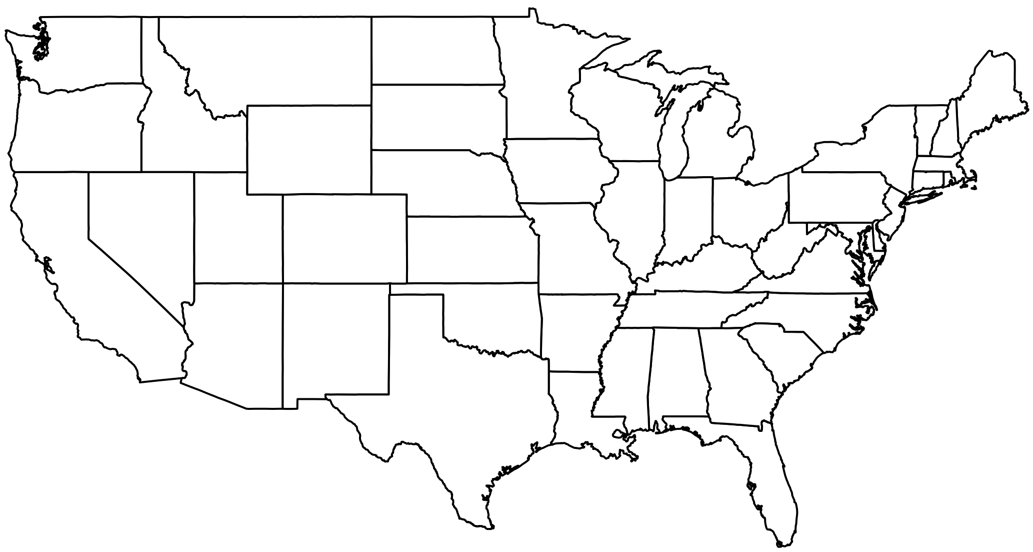
Mapping water quality in lakes

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.

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



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

LAGOSlocus <- LAGOSdata$locus
LAGOSstate <- LAGOSdata$state
LAGOSnutrient <- LAGOSdata$epi_nutr
LAGOSlimno <- LAGOSdata$lakes_limno

Maine_combined <-
  left_join(LAGOSnutrient, LAGOSlocus) %>%
  left_join(., LAGOSlimno) %>%
  left_join(., LAGOSstate) %>%
  filter(state == "ME") %>%
  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"
```

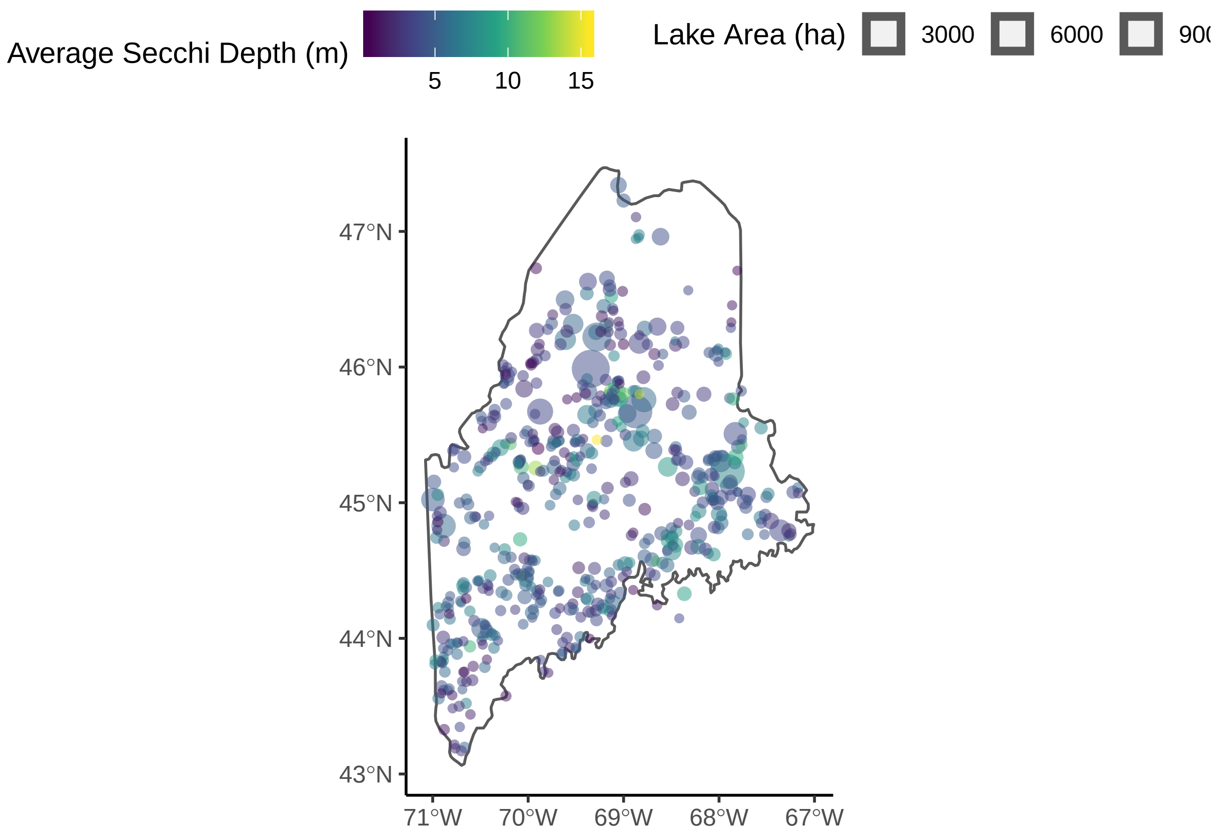
```
secchi_maine.summary <- Maine_combined %>%
  group_by(lagoslakeid) %>%
  summarise(secchi.mean = mean(secchi),
            area = mean(lake_area_ha),
            depth = mean(maxdepth),
            lat = mean(nhd_lat),
            long = mean(nhd_long)) %>%
  drop_na()

secchi_maine_depth.summary <- secchi_maine.summary %>%
  select("lagoslakeid", "secchi.mean")
```

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.

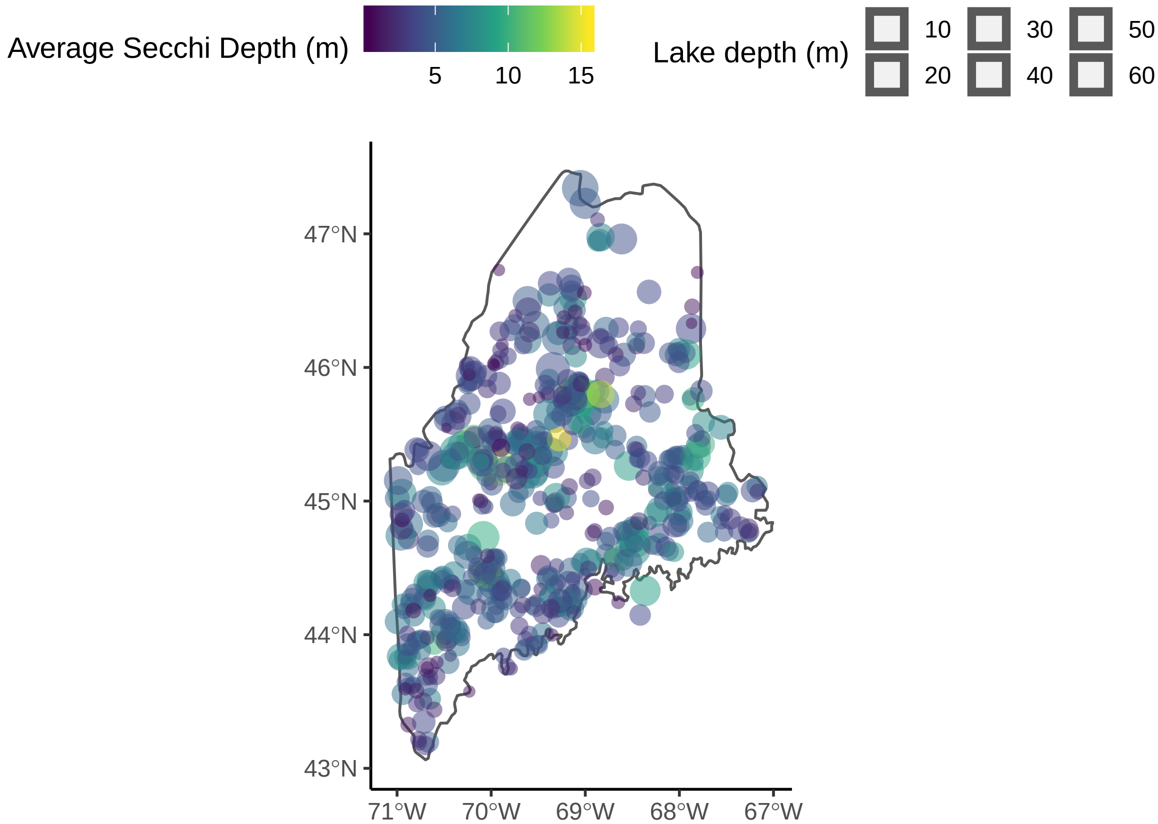
```
secchi_maine.spatial <- st_as_sf(secchi_maine.summary, coords = c("long", "lat"), crs = 4326)

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



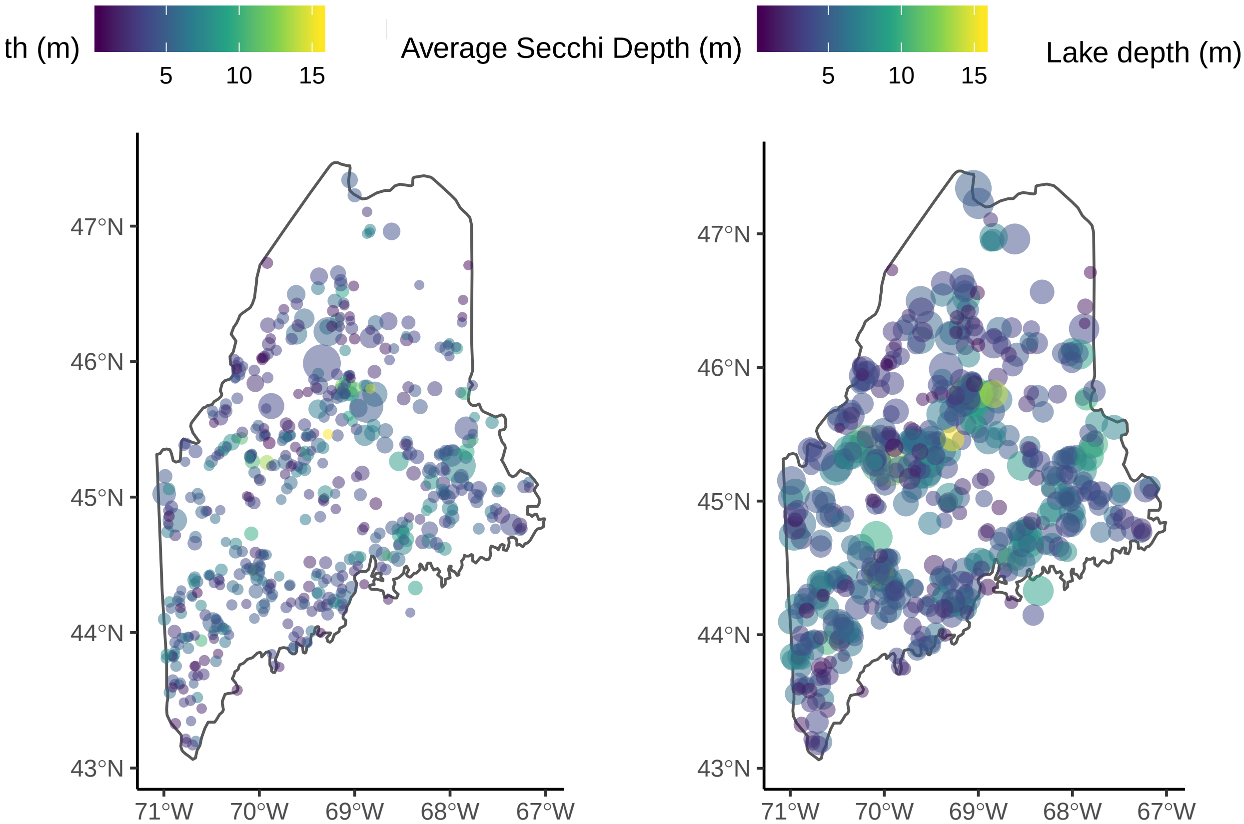
7. Create a second plot, but this time use maximum depth of the lake as the size of the dot.

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

```
plot_grid(Secchi_maine_area_plot,Secchi_maine_depth_plot)
```



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

> I will change the legends of lake area and lake depth, since they represent different variables but right now they have the same legend signature. Using different legend may make this plot a better visualization.

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_area_model <- lm(data = secchi_maine.summary, log(secchi.mean) ~ log(area))
summary(secchi_area_model)
```

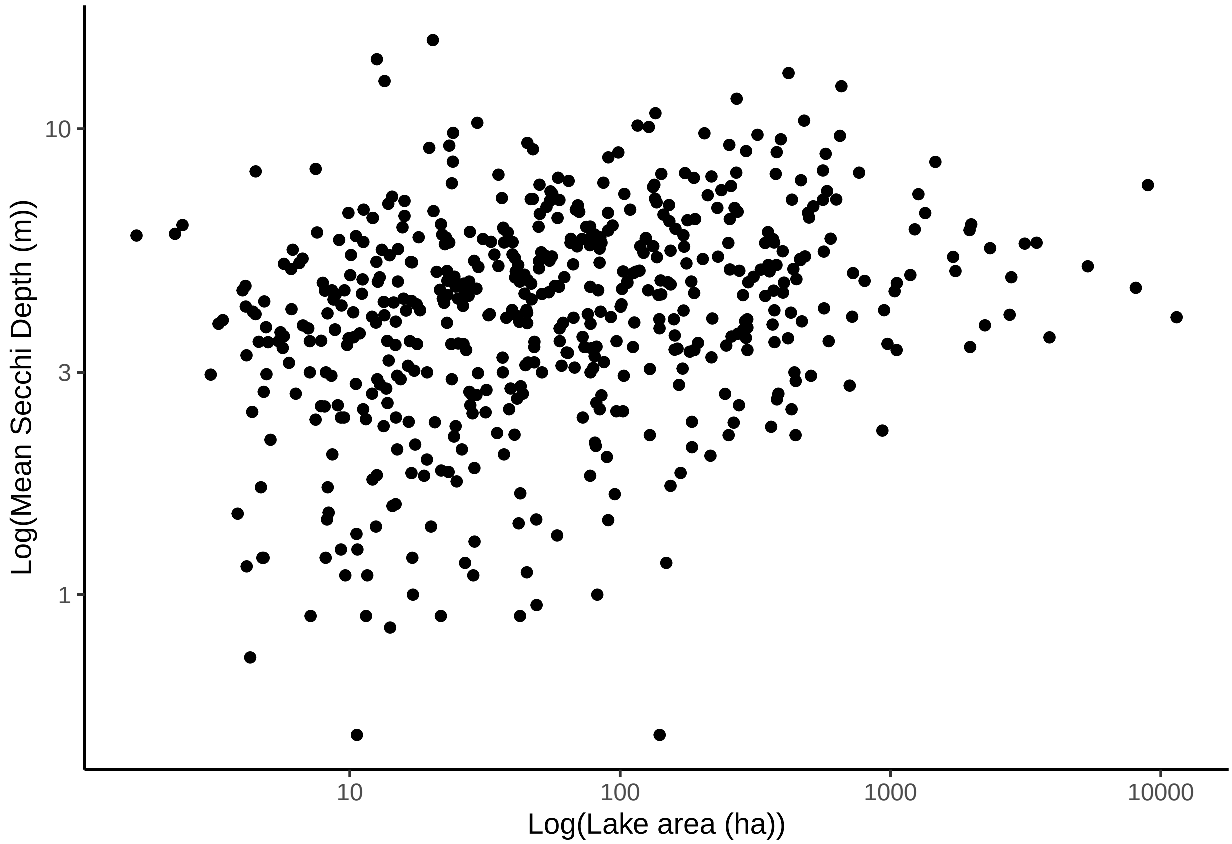
```
##
## Call:
## lm(formula = log(secchi.mean) ~ log(area), data = secchi_maine.summary)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.16467 -0.27492  0.07398  0.34061  1.45290
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.00193    0.05991  16.724  < 2e-16 ***
## log(area)    0.09503    0.01376   6.905  1.4e-11 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.511 on 547 degrees of freedom
## Multiple R-squared:  0.08017,    Adjusted R-squared:  0.07848
## F-statistic: 47.67 on 1 and 547 DF,  p-value: 1.402e-11
```

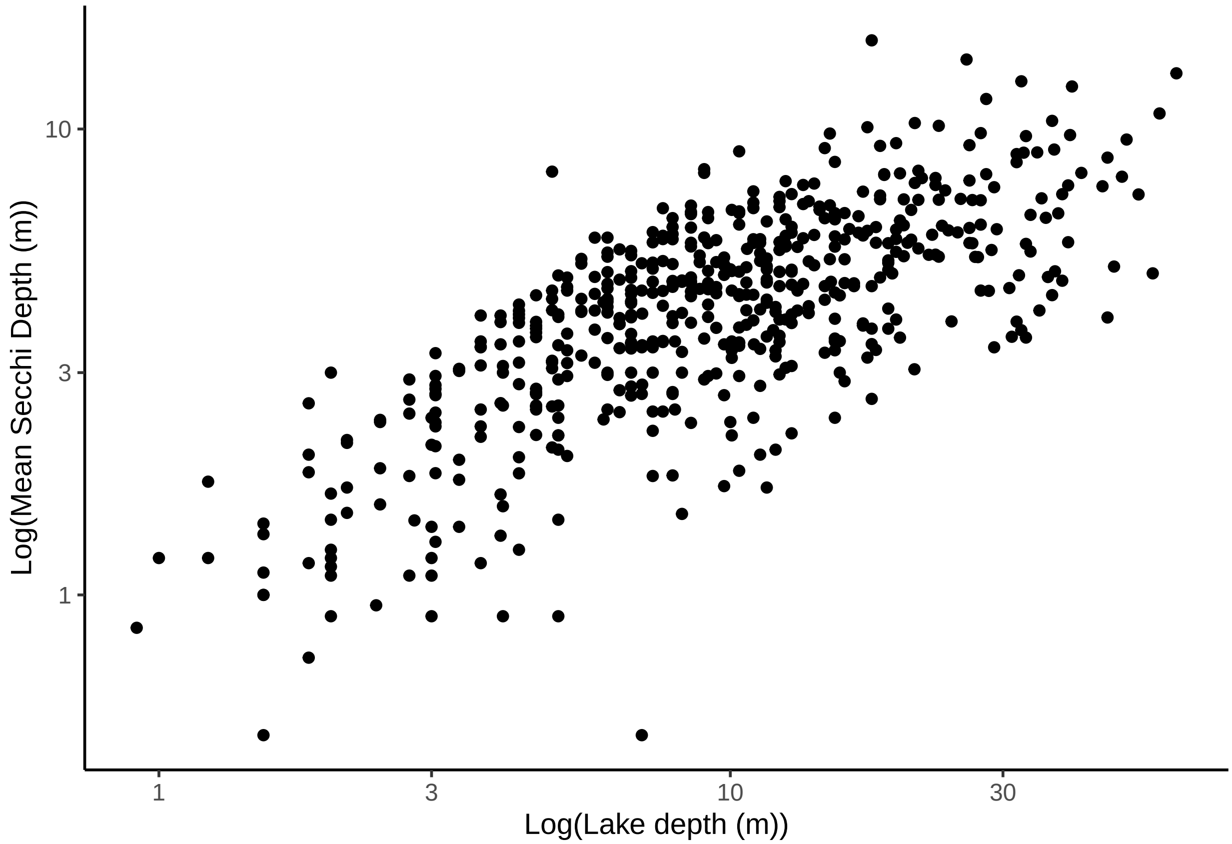
```r
secchi_depth_model <- lm(data = secchi_maine.summary, log(secchi.mean)~ log(depth))
summary(secchi_depth_model)
```

```
##
## Call:
## lm(formula = log(secchi.mean) ~ log(depth), data = secchi_maine.summary)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.94362 -0.20456  0.04784  0.25361  1.03341
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.29067    0.04713   6.167 1.35e-09 ***
## log(depth)   0.49324    0.02000  24.663  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3666 on 547 degrees of freedom
## Multiple R-squared:  0.5265, Adjusted R-squared:  0.5256
## F-statistic: 608.3 on 1 and 547 DF,  p-value: < 2.2e-16
```

```r
secchi_depthVSlake_area.plot <-
  ggplot(secchi_maine.summary, aes(x= area, y = secchi.mean))+
  geom_point()+
  scale_x_log10() + scale_y_log10()+
  labs(x = "Log(Lake area (ha))", y = "Log(Mean Secchi Depth (m))")
print(secchi_depthVSlake_area.plot)
```




```
secchi_depthVSlake_depth.plot <-
  ggplot(secchi_maine.summary, aes(x= depth, y = secchi.mean))+
  geom_point()+
  scale_x_log10() + scale_y_log10()+
  labs(x = "Log(Lake depth (m))", y = "Log(Mean Secchi Depth (m))")
print(secchi_depthVSlake_depth.plot)
```



The relationship between secchi depth and lake area is not obvious in log linear regression, the multiple R-squared is only around 8%. The relationship between secchi depth and lake depth is stronger in log linear regression, the multiple R-squared is 52.65%. Thus, lake depth is a stronger determinant of 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.

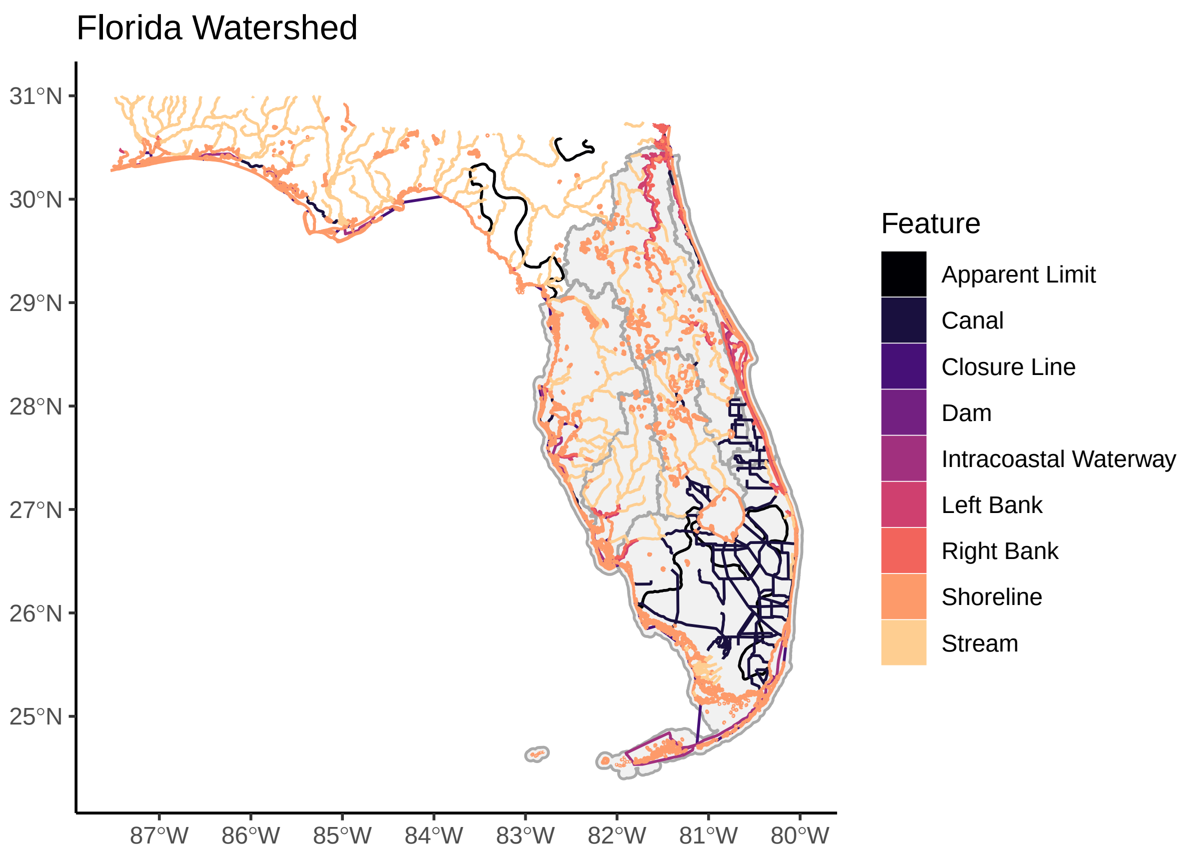
```
waterfeatures.FL <- filter(waterfeatures, STATE == "FL")
HUC6.FL <- HUC6 %>%
  filter(States %in% "FL")

waterfeatures.FL <- st_set_crs(waterfeatures.FL, 4269)
HUC6.FL <- st_set_crs(HUC6.FL,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.

```
FLlayers <- ggplot() +
  geom_sf(data = HUC6.FL, color = "darkgray", alpha = 0.5) +
  geom_sf(data = waterfeatures.FL, aes(fill = FEATURE, color = FEATURE)) +
  labs(title = "Florida Watershed",
       color = "Feature", fill = "Feature")+
```

```
  scale_color_viridis_d(option = "magma", end = 0.9) +
  scale_fill_viridis_d(option = "magma", end = 0.9)
print(FLlayers)
```



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

    Stream and shoreline are the dominant water features in Florida. There are much more shorelines in Florida compared to NC. In Florida, streams are seldom in the south part, they are distributed in central or north part of the Florida. In NC, streams are mostly distributed in central and west part, only a small part of east does not have streams.

Reflection

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

    1.projections: can be unique to specific area, 2."sdt_as_sf" can be used in R to map 3. "st_set_crs" can be used to set projections for different dataset.

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

    the florida watershed map

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

    Hands-on data analysis can let me explore the theory on my own, and it can help me understand it better from examples.

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

The real-world map sometimes does not have the correct display since the data may not be correct.