Assignment 8: Mapping

Felipe Raby Amadori

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.

```
# 1. Verify your working directory is set to the R project file
getwd()
```

```
## [1] "C:/Users/Felipe/OneDrive - Duke University/1. DUKE/Ramos 3 Semestre/Hydrologic_Data_Analysis"
```

```
# 2. Load the tidyverse, lubridate, cowplot, LAGOSNE, sf, maps, and viridis packages

packages <- c(
  "ggplot2",
  "tidyverse",
  "lubridate",
  "cowplot",
  "LAGOSNE",
  "sf",
  "maps",
  "viridis"
  )
invisible(
  suppressPackageStartupMessages(
    lapply(packages, library, character.only = TRUE)
    )
  )

# 3. Set your ggplot theme (can be theme_classic or something else)
felipe_theme <- theme_light(base_size = 12) +
  theme(axis.text = element_text(color = "grey8"),
```

```r
        legend.position = "right", plot.title = element_text(hjust = 0.5))
theme_set(felipe_theme)

# 4. Load the lagos database, the USA rivers water features shape file, and the HUC6
# watershed shape file.

LAGOSdata <- lagosne_load()
```

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

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

```
## Reading layer `WBDHU6' from data source `C:\Users\Felipe\OneDrive - Duke University\1. DUKE\Ramos 3 S
## 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
```

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

```
## Reading layer `hydrogl020' from data source `C:\Users\Felipe\OneDrive - Duke University\1. DUKE\Ramos
## 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
```

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

# filter only states that are included in the LAGOSNE database
states.subset <- filter(states, ID %in%
                        c("maine"))

# load LAGOSNE data frames
```

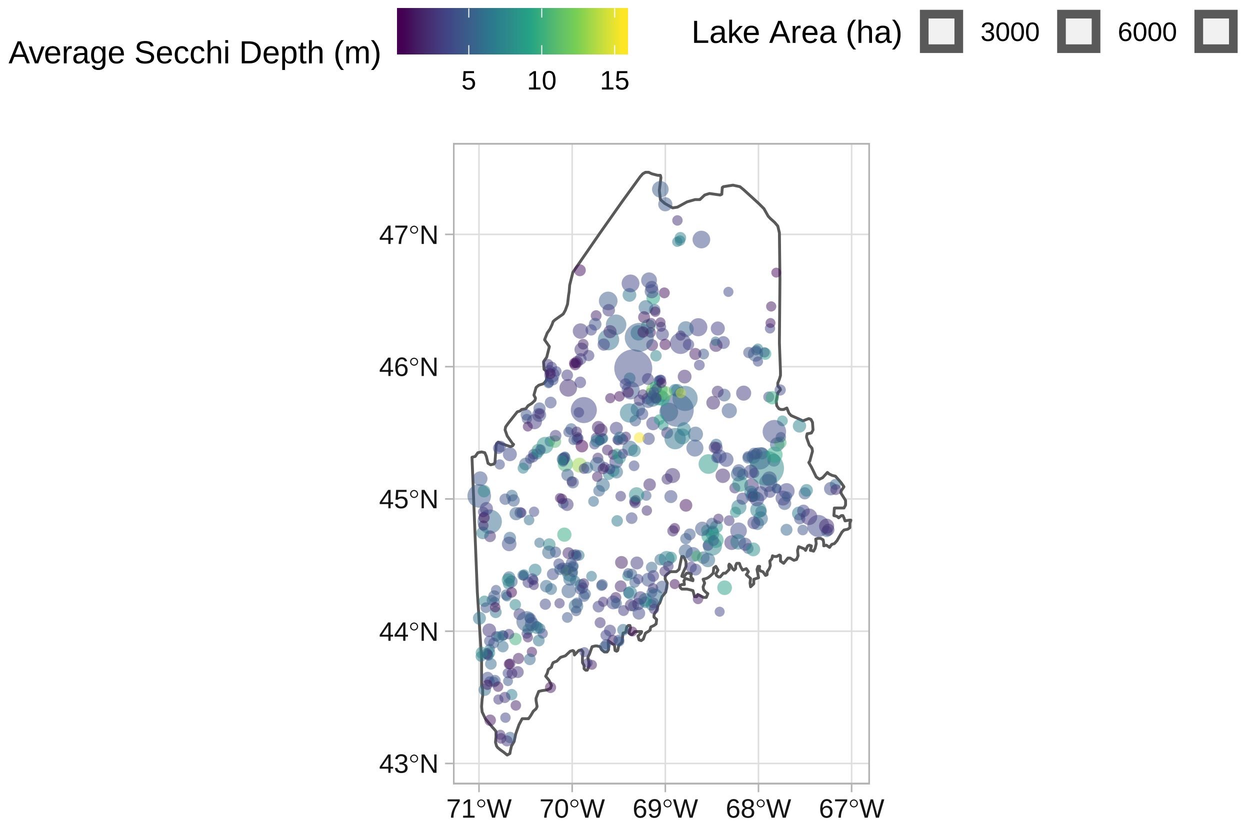
```r
LAGOSlocus <- LAGOSdata$locus
LAGOSstate <- LAGOSdata$state
LAGOSnutrient <- LAGOSdata$epi_nutr
LAGOSlimno <- LAGOSdata$lakes_limno

# Create a data frame to visualize secchi depth
LAGOScombined <-
  left_join(LAGOSnutrient, LAGOSlocus) %>% #locus gives you location
  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
LAGOScombined.maine <- filter(LAGOScombined, state == "ME")

# create a summary dataset with just the mean secchi depth
secchi.summary <- LAGOScombined.maine %>%
  group_by(lagoslakeid) %>%
  summarise(secchi.mean = mean(secchi)) %>%
  drop_na()
```

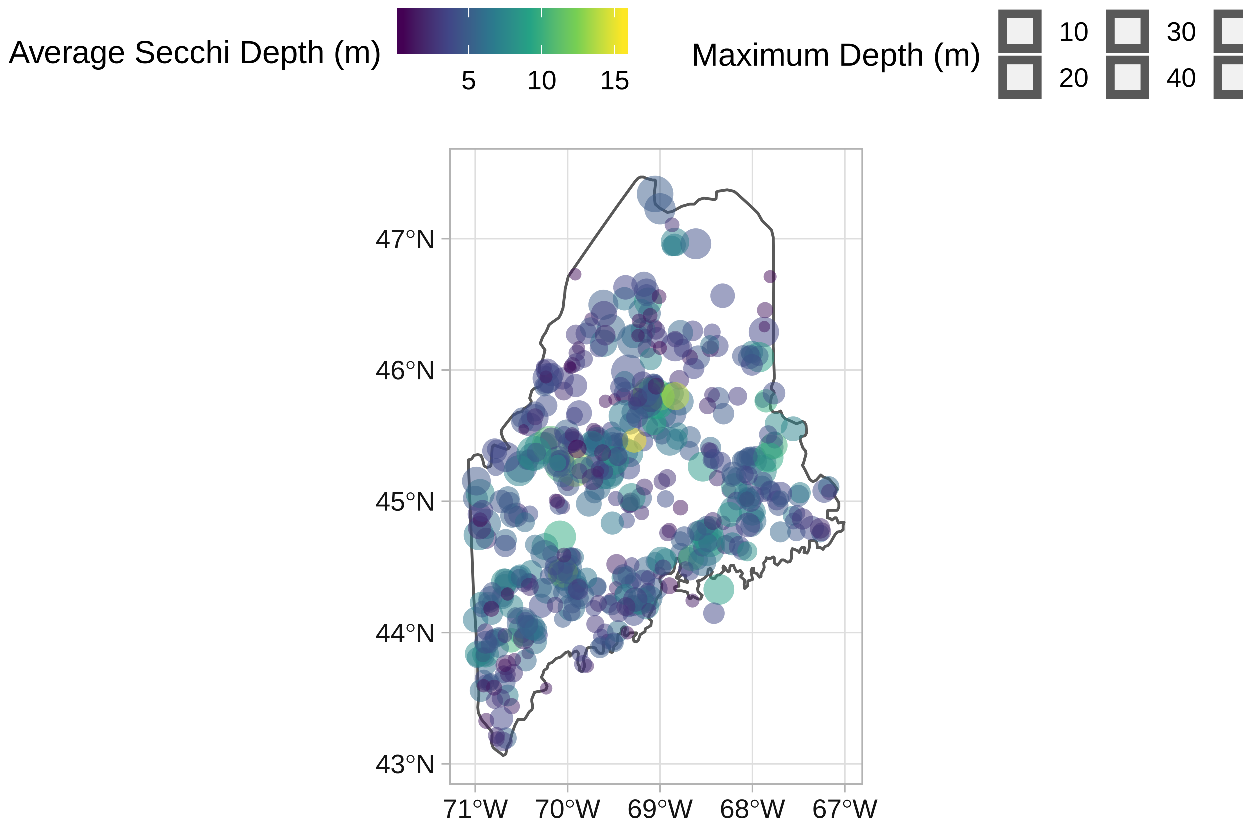
6. Create a plot of mean secchi depth for lakes in Maine, with mean secchi depth designated as color and the lake area as the size of the dot. Remember that you are using size in the aesthetics and should remove the size = 1 from the other part of the code. Adjust the transparency of points as needed.

```r
secchi.summary.plot <- LAGOScombined.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.sf <- st_as_sf(secchi.summary.plot, coords = c("long", "lat"), crs = 4326)

Secchiplot <- ggplot() +
  geom_sf(data = states.subset, fill = "white") +
  geom_sf(data = secchi.sf, 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(Secchiplot)
```



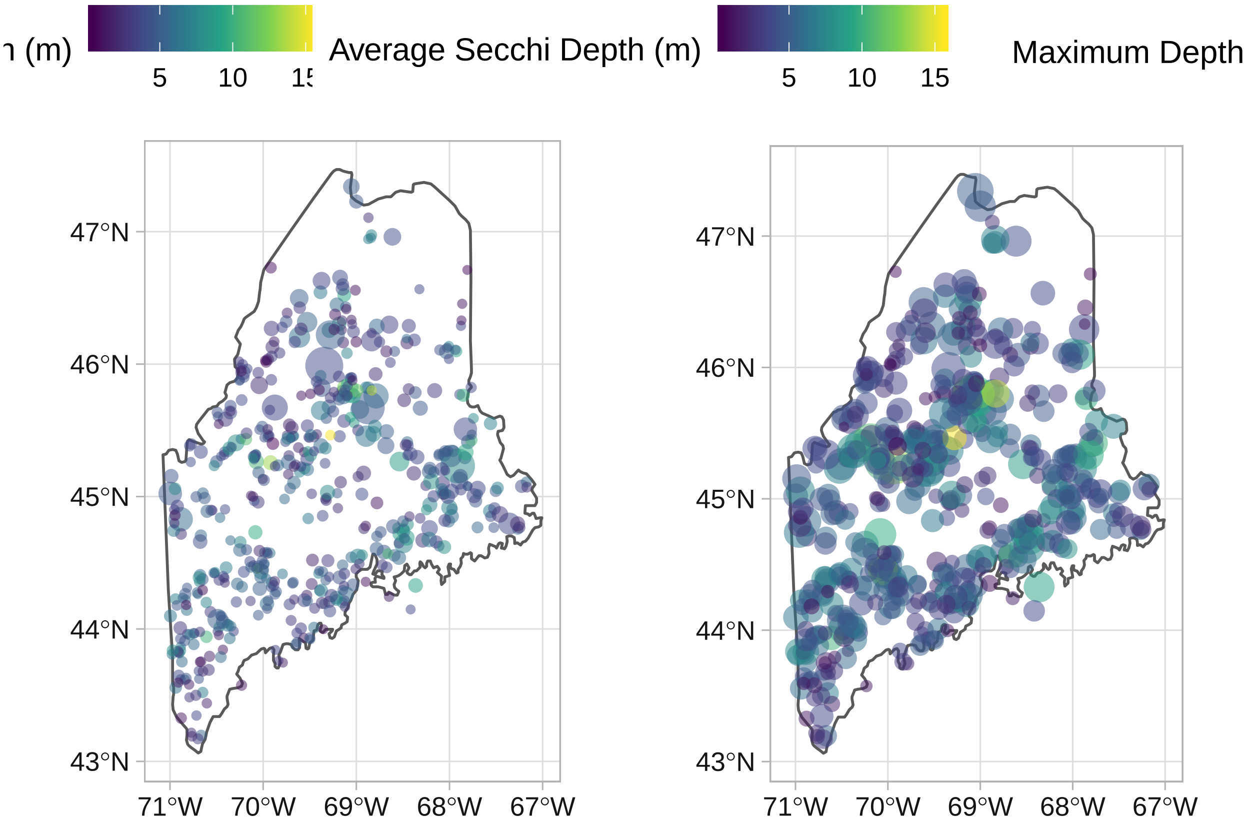
7. Create a second plot, but this time use maximum depth of the lake as the size of the dot.

```
Secchiplot2 <- ggplot() +
  geom_sf(data = states.subset, fill = "white") +
  geom_sf(data = secchi.sf, aes(color = secchi.mean, size = depth),
          alpha = 0.5) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi Depth (m)", size = "Maximum Depth (m)") +
  theme(legend.position = "top")
print(Secchiplot2)
```



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, Secchiplot2)
```



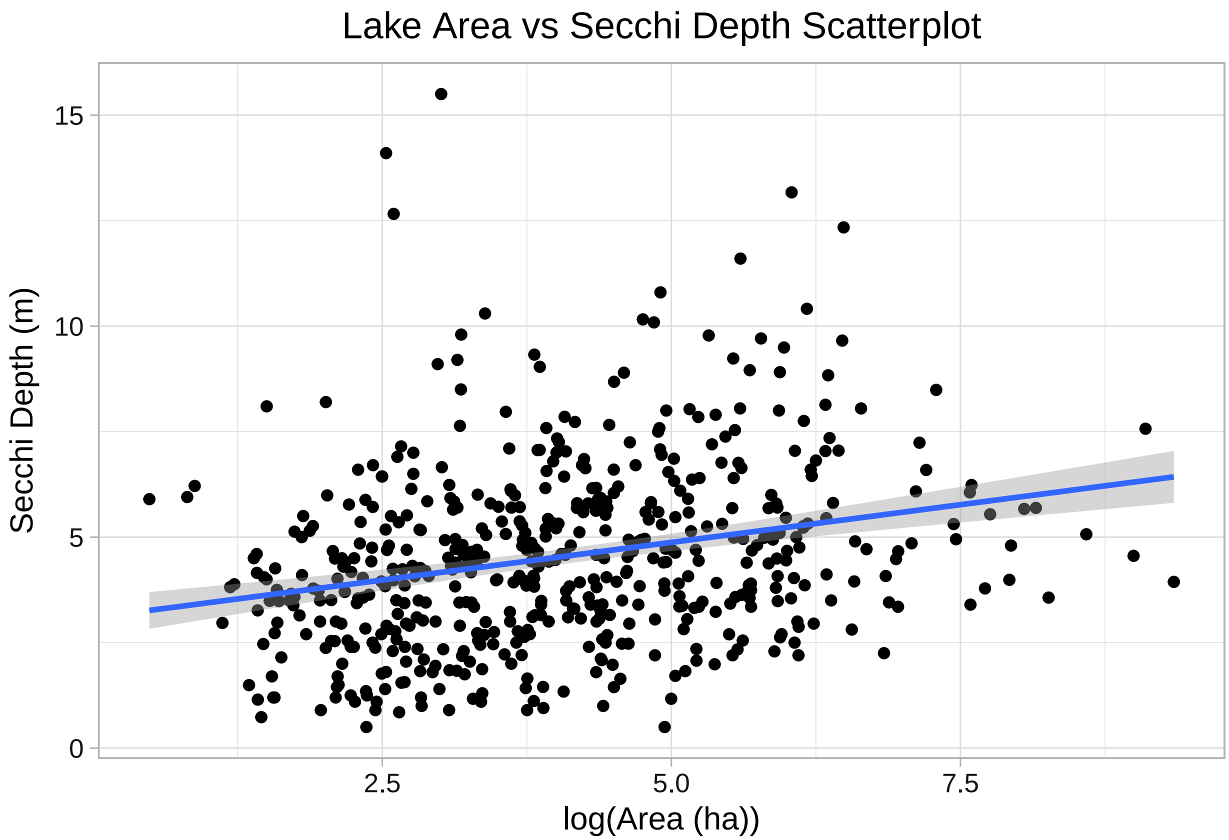
What would you change about the legend to make it a more effective visualization?

> First of all I would made them fit into the width of the page, so It can be seen all the information that the legend is suppose to provide. Second, I would try to create a legend that explains what the actual diameters of the points mean in terms of Lake Area for one plot and Maximum Depth for the other one instead of the squares that are been shown right now. This squares do not communicate anything to the reader except maybe information about the scale of the values (in the case of Lake Area it does not even do that, showing only values of thousands of hectares when many lakes have areas below 100 hectares). You could also use only one Average Secchi Depth legend after making sure that the scales in both plots for Secchi Depth is equivalent.

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

```
ggplot(secchi.summary.plot, aes(x = log(area), y = secchi.mean)) +
  geom_point() +
  geom_smooth(method=lm) +
  xlab(expression("log(Area (ha))")) +
  ylab(expression("Secchi Depth (m)")) +
  ggtitle("Lake Area vs Secchi Depth Scatterplot") +
  theme(plot.title = element_text(hjust = 0.5))
```



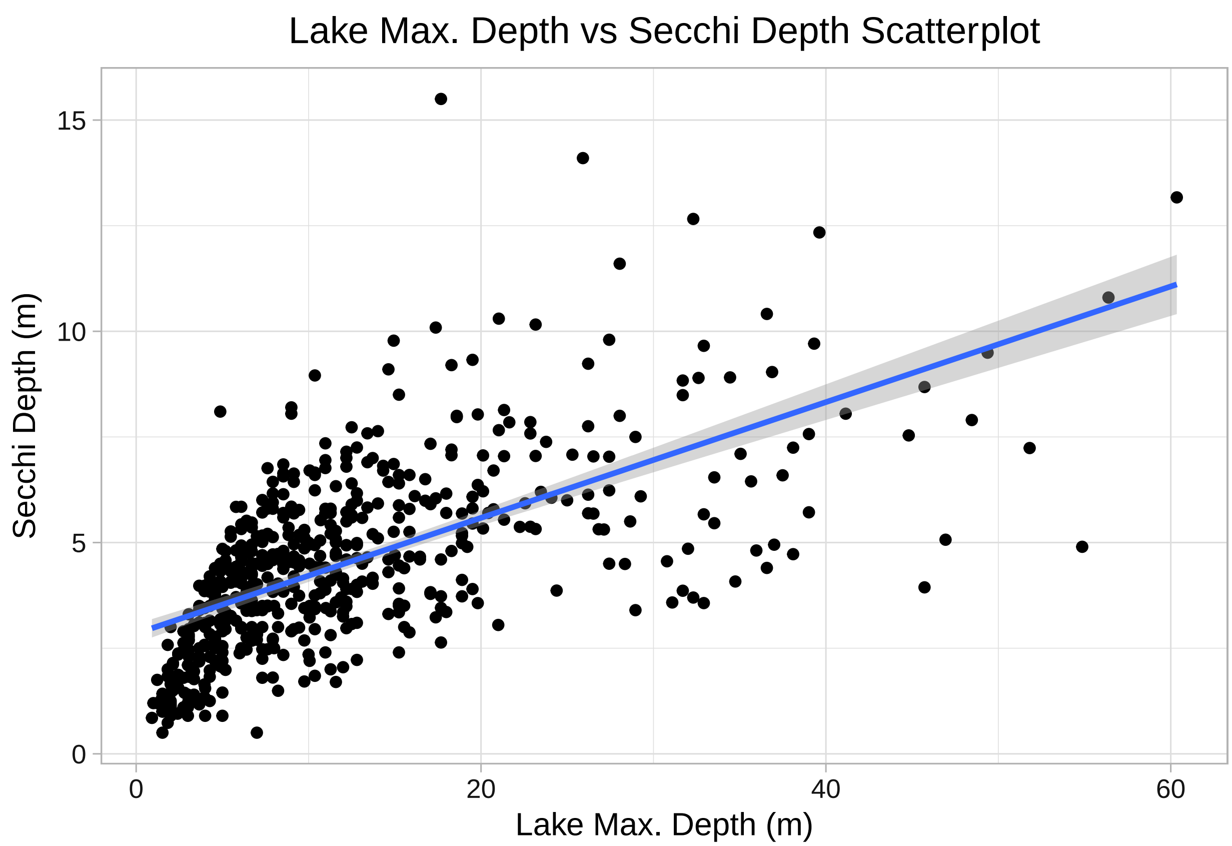

```r
lmAreaSecchi <- lm(log(area) ~ secchi.mean, data = secchi.summary.plot)
summary(lmAreaSecchi)
```

```
##
## Call:
## lm(formula = log(area) ~ secchi.mean, data = secchi.summary.plot)
##
## 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
```

For the Lake Area variable we performed a log transformation because the tail was too long and most of the values were concentrated in the smaller area values. In the Lake Area vs Secchi Depth plot it can be seen that it seems to exists a positive correlation between log(Area) and Secchi Depth, but with an important variability.

The linear regression confirms these observations. The null hypothesis that there is no effect of Lake Area on Secchi Depth is rejected with a 5% level of significance (F-statistic: 39.53 on 1 and 547 DF, p-value: 6.622e-10 with a positive slope of the regression line = 0.18865); nevertheless, the linear model explains only 6.5% of the variance of the response variable.

```
ggplot(secchi.summary.plot, aes(x = depth, y = secchi.mean)) +
  geom_point() +
  geom_smooth(method=lm) +
  xlab(expression("Lake Max. Depth (m)")) +
  ylab(expression("Secchi Depth (m)")) +
  ggtitle("Lake Max. Depth vs Secchi Depth Scatterplot") +
  theme(plot.title = element_text(hjust = 0.5))
```




```
lmDepthSecchi <- lm(depth ~ secchi.mean, data = secchi.summary.plot)
summary(lmDepthSecchi)
```

```
##
## Call:
## lm(formula = depth ~ secchi.mean, data = secchi.summary.plot)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.971  -4.605  -1.679   2.311  41.476
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.6005     0.7671  -0.783    0.434
```

```
## secchi.mean   2.8548     0.1524  18.738   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.784 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
```

In the Lake Maximum Depth vs Secchi Depth plot it can be seen that there is a clear positive correlation between Lake Max. Depth and Secchi Depth, and with less variability than Lake Area.

The linear regression confirms these observations. The null hypothesis that there is no effect of Lake Maximum Depth on Secchi Depth is rejected with a 5% level of significance (F-statistic: 351.1 on 1 and 547 DF, p-value: $<$ 2.2e-16 with a positive slope of the regression line = 2.8548); moreover, the linear model explains 39% of the variance of the response variable.

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.

```
HUC6.FL <- HUC6.WS %>%
  filter(grepl("FL", States))

#To chech that it worked
summary(HUC6.FL$States)
```

```
##       AL    AL,FL AL,FL,GA AL,GA,TN AL,LA,MS    AL,MS       FL    FL,GA
##        0        3        1        0        0        0        6        4
##       GA GA,NC,SC    LA,MS       NC    NC,SC NC,SC,VA    NC,VA       SC
##        0        0        0        0        0        0        0        0
```

```
#it did

# Filter for Florida
USAwaterfeatures <- filter(USAwaterfeatures, STATE == "FL")

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

st_crs(USAwaterfeatures)
```

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

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

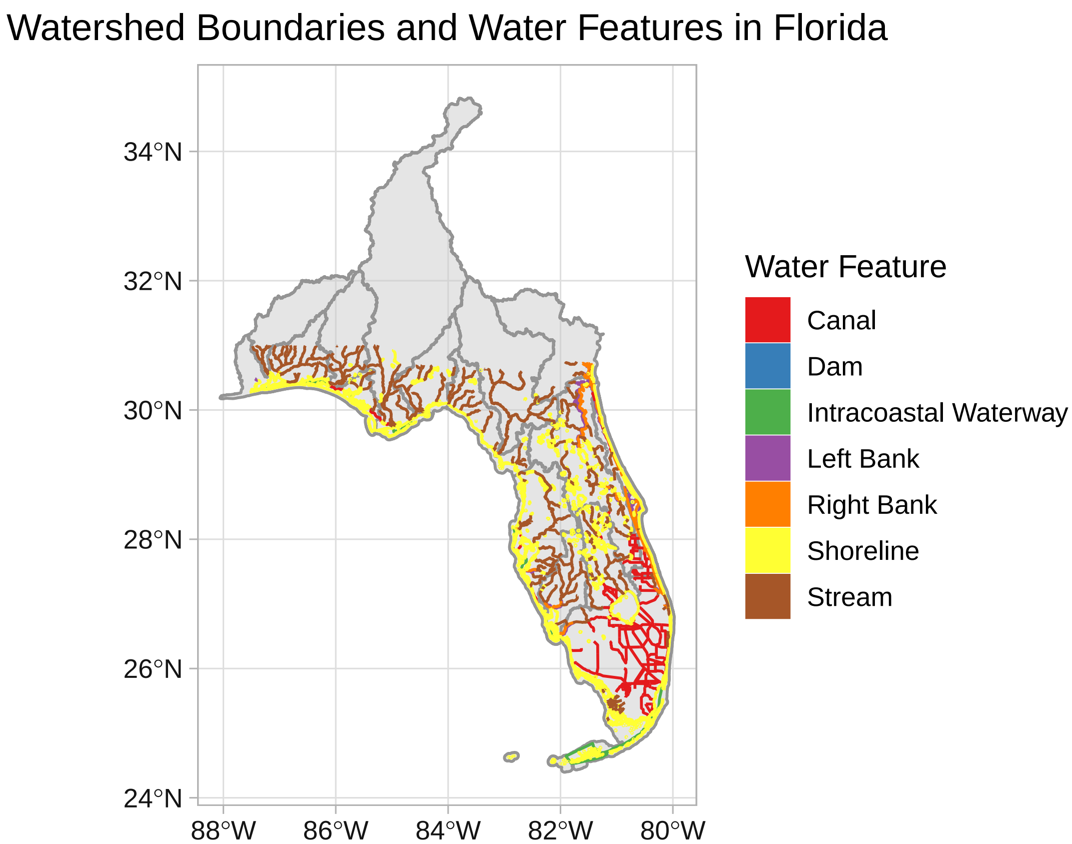
```
## Coordinate Reference System:
##   EPSG: 4269
##   proj4string: "+proj=longlat +datum=NAD83 +no_defs"
```

```
# USAwaterfeatures does not have coordinate reference information. HUC6 coordinate
# reference information is NAD83 ESPG: 4269.

USAwaterfeatures <- st_set_crs(USAwaterfeatures, 4269)
st_crs(USAwaterfeatures)
```

```
## 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.

```
FLlayers <- ggplot() +
  geom_sf(data = HUC6.FL, color = "gray58", fill = "gray80", alpha = 0.5) +
  geom_sf(data = USAwaterfeatures, aes(fill = FEATURE, color = FEATURE)) +
  scale_fill_brewer(palette = "Set1") +
  scale_color_brewer(palette = "Set1") +
  labs(color = 'Water Feature', fill = 'Water Feature') +
  ggtitle("Watershed Boundaries and Water Features in Florida") +
  theme(plot.title = element_text(hjust = 0.5))
print(FLlayers)
```



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

    In most of the inland part of the state the most dominant features are streams and lakes; nevertheless, in the south part canals are clearly more dominant than streams. Banks can be appreciated all over both coastal lines. The extreme south part of the states show some intracoastal waterways. Meanwhile, North Carolina is almost everywhere dominated by streams and by man made reservoirs built in the stream's riverbeds. There are some canals near the coast but are far less dominant than canals in south Florida.

Reflection

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

    1. Lake Area can have a positive correlation with Secchi Depth
    2. Lake Maximum Depth can have a positive correlation with Secchi Depth, greater than Lake Area and explaining more variance.
    3. Dominant water features differ among states.

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

    1. Log(Lake Area) and Secchi Depth scatterplot and linear model.
    2. Lake Maximum Depth and Secchi Depth scatterplot and linear model.
    3. Maps of water features created for Florida and NC (this last one in class).

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

    It did. Playing with data is the best way to understand concepts and later have them available to apply them in other settings.

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

    I did not have much expectations from theory in this area.