Assignment 8: Mapping

Jake Greif

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/jakegreif/Duke/Fall_2019/Hydrologic_Data_Analysis"
```

```
library(tidyverse)
```

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

## 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_co
## 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())

LAGOSdata <- lagosne_load()
```

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

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

```
## Reading layer `hydrogl020' from data source `/Users/jakegreif/Duke/Fall_2019/Hydrologic_Data_Analysis
## 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 `/Users/jakegreif/Duke/Fall_2019/Hydrologic_Data_Analysis/Da
## 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:

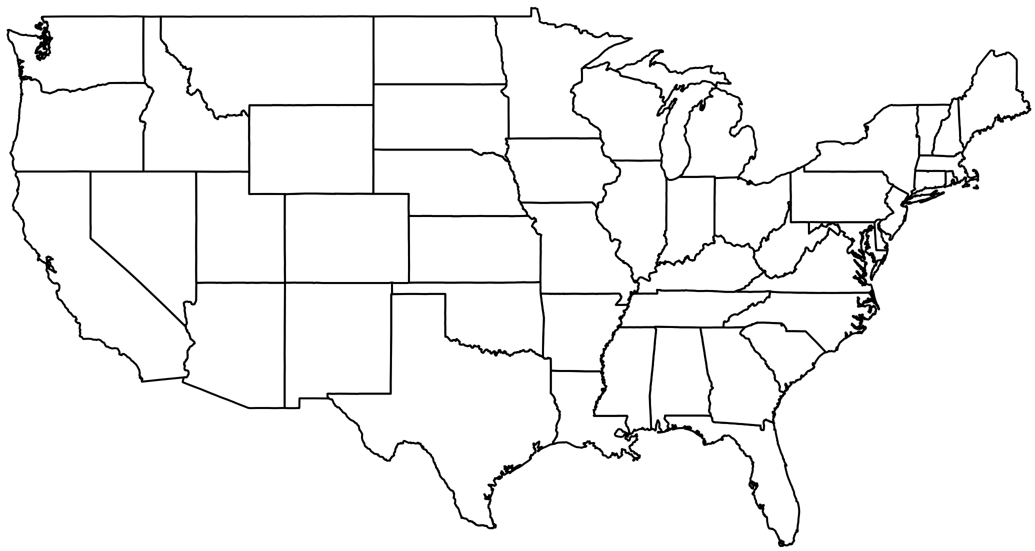
- Make sure all your wrangling is done in this document (this includes basic wrangling of the LAGOS database)
- In your cowplot, do not adjust the legend items (even though they look ugly). Rather, reflect on how you would improve them with additional coding.
- For item 9, **do** run a regression on 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.

```
# 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"
```

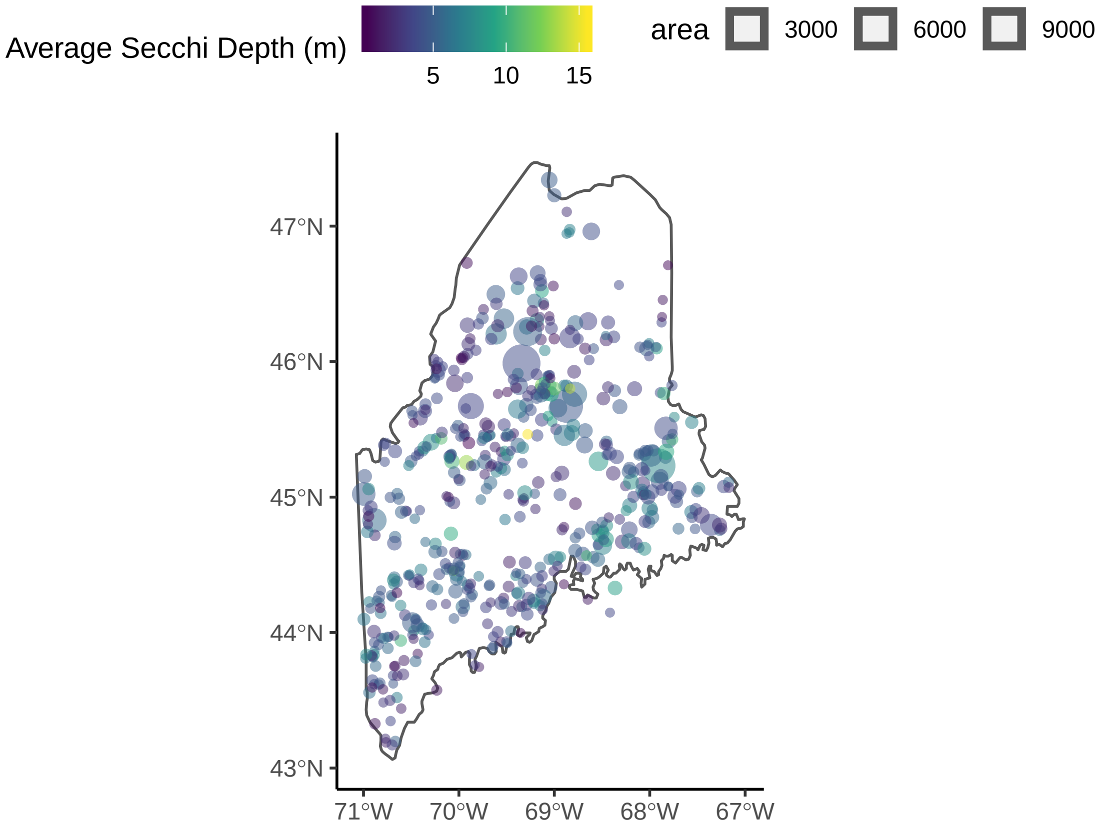
```
# Isolate Maine data
states <- st_as_sf(map(database = "state", plot = TRUE,
                       fill = TRUE, col = "white"))
```



```r
Maine <- filter(states, ID == "maine")

Maine.dat <- filter(LAGOScombined, 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()

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

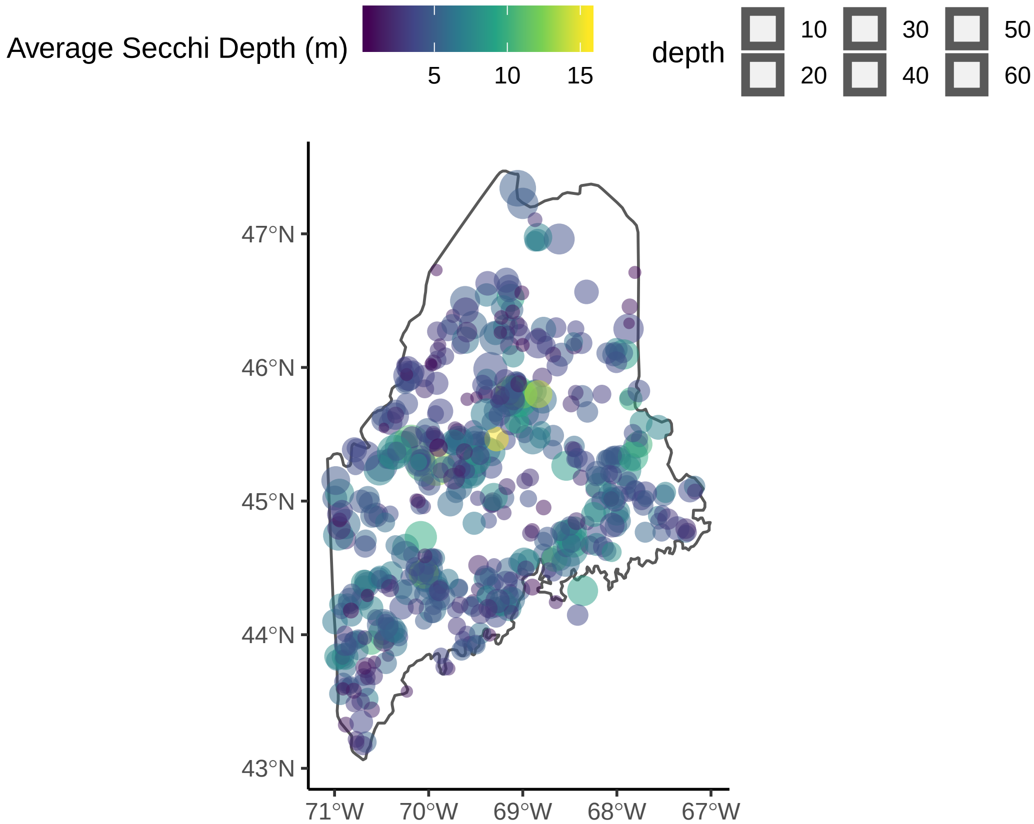
6. Create a plot of mean secchi depth for lakes in Maine, with mean secchi depth designated as color and the lake area as the size of the dot. Remember that you are using size in the aesthetics and should remove the size = 1 from the other part of the code. Adjust the transparency of points as needed.

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



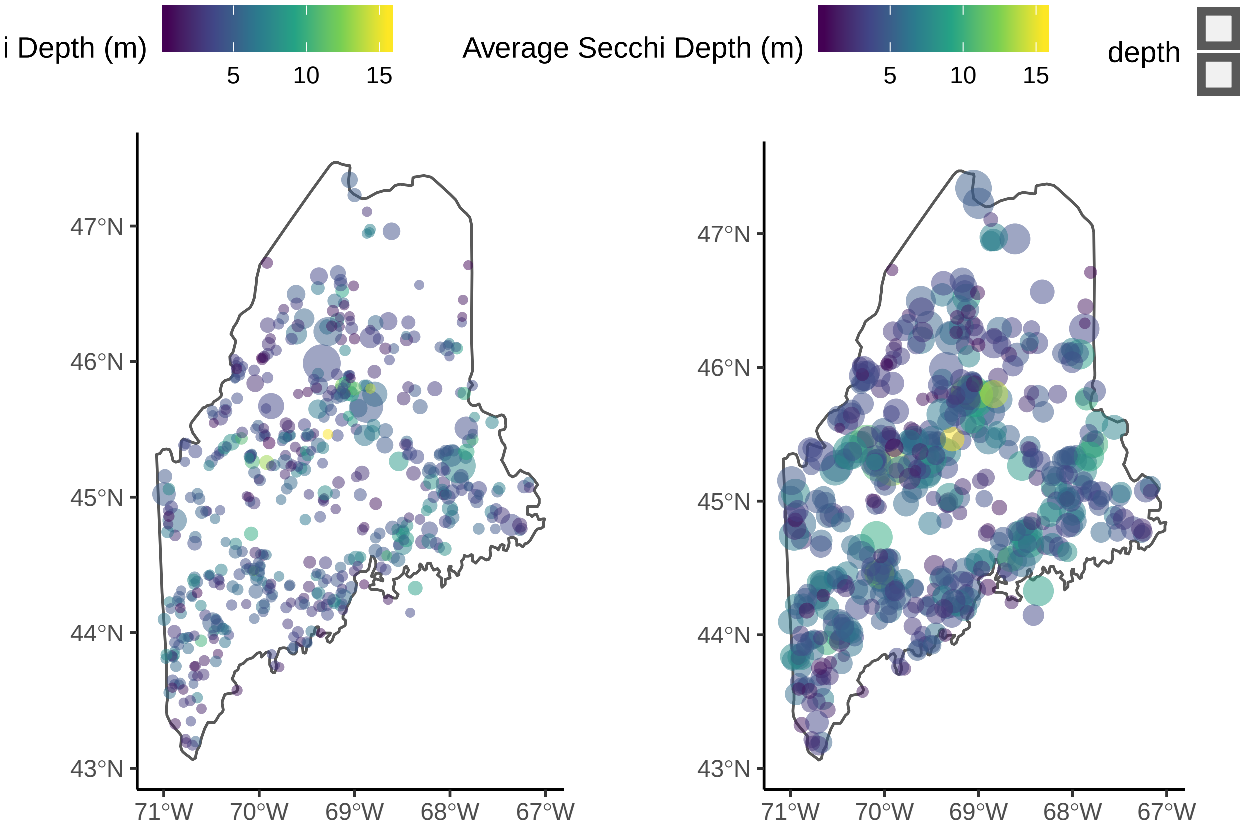
7. Create a second plot, but this time use maximum depth of the lake as the size of the dot.

```
Maine.plot2 <- ggplot() +
  geom_sf(data = Maine, fill = "white") +
  geom_sf(data = Maine.spatial, aes(color = secchi.mean, size = depth),
          alpha = 0.5) +
  scale_color_viridis_c() +
  labs(color = "Average Secchi Depth (m)") +
  theme(legend.position = "top")

print(Maine.plot2)
```



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(Maine.plot, Maine.plot2)
```



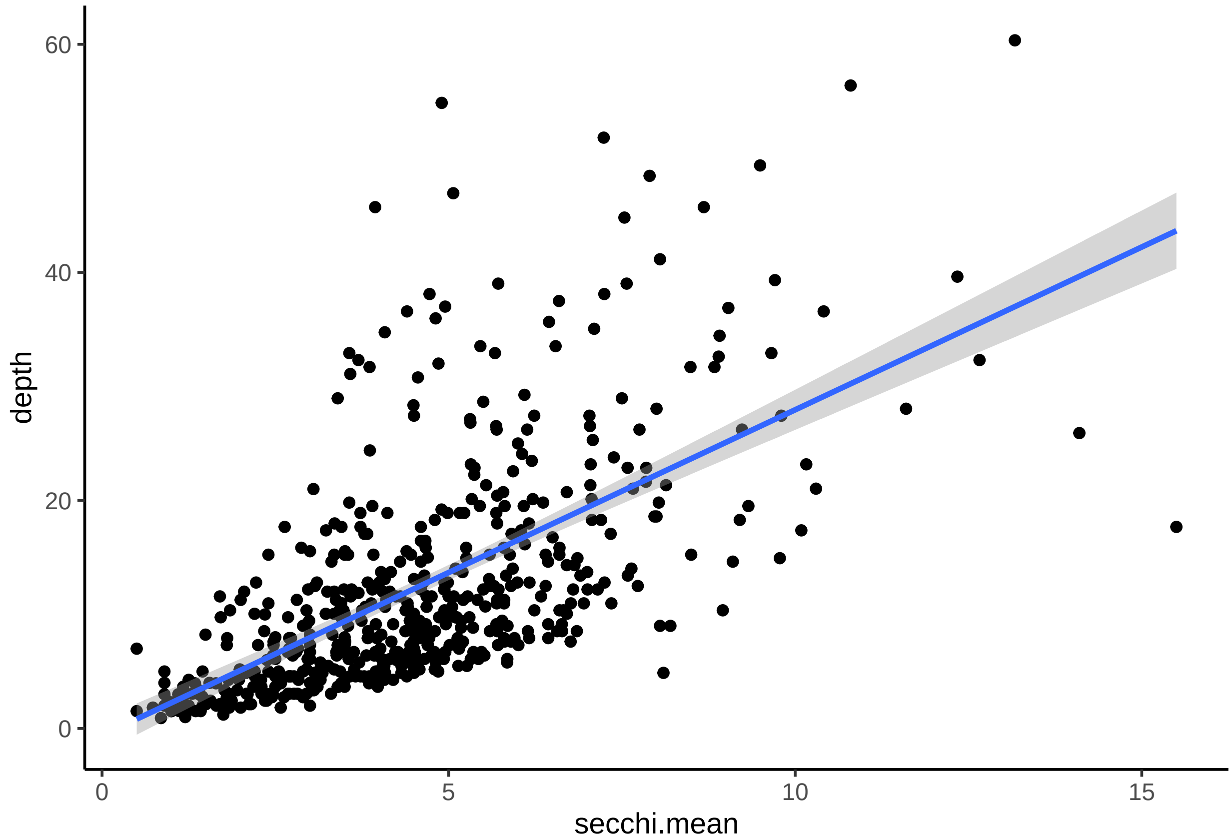
What would you change about the legend to make it a more effective visualization?

> I would remove one of the Average secchi depth scales as well as both the depth and lake size components of the legend. In order to differentiate between the maps, I would include what the dot size indicates in the title for each plot.

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

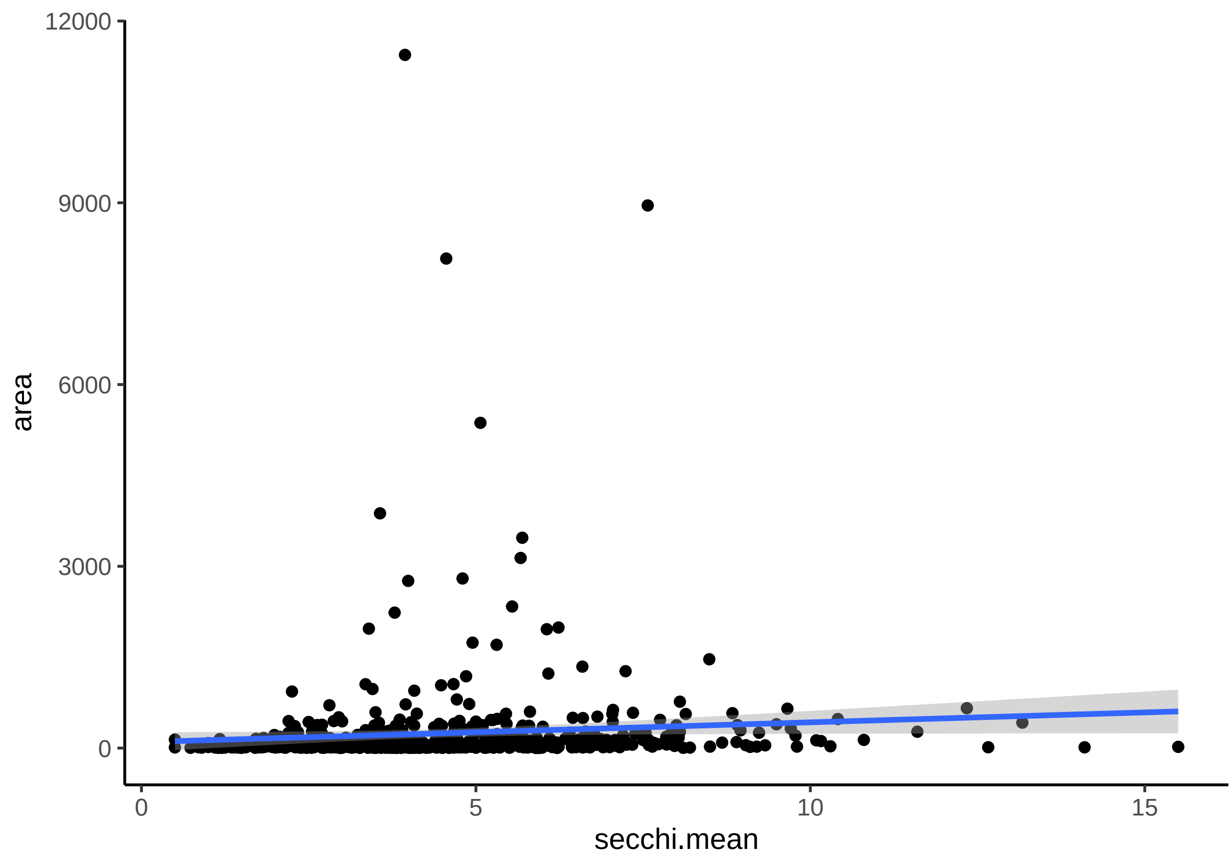
```
Maine.depth.plot <- ggplot(Maine.dat, aes(x = secchi.mean, y = depth)) +
  geom_point() +
  geom_smooth(method = 'lm')
print(Maine.depth.plot)
```




```r
Maine.depth.model <- lm(data = Maine.dat, depth ~ secchi.mean)
summary(Maine.depth.model)
```

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

```r
Maine.area.plot <- ggplot(Maine.dat, aes(x = secchi.mean, y = area)) +
  geom_point() +
  geom_smooth(method = 'lm')
print(Maine.area.plot)
```




```
Maine.area.model <- lm(data = Maine.dat, area ~ secchi.mean)
summary(Maine.area.model)
```

```
##
## Call:
## lm(formula = area ~ secchi.mean, data = Maine.dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -584.7  -217.6  -159.2   -79.9 11214.5
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    97.90      82.43   1.188   0.2355
## secchi.mean    32.72      16.37   1.998   0.0462 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 836.4 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
```

Secchi depth has a significant, positive correlation with both lake area and depth. Though, lake depth is a stronger determinant of secchi depth than the lake area.

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% c("AL,FL", "AL,FL,GA", "FL", "FL,GA"))

st_crs(waterfeatures.FL)
```

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

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

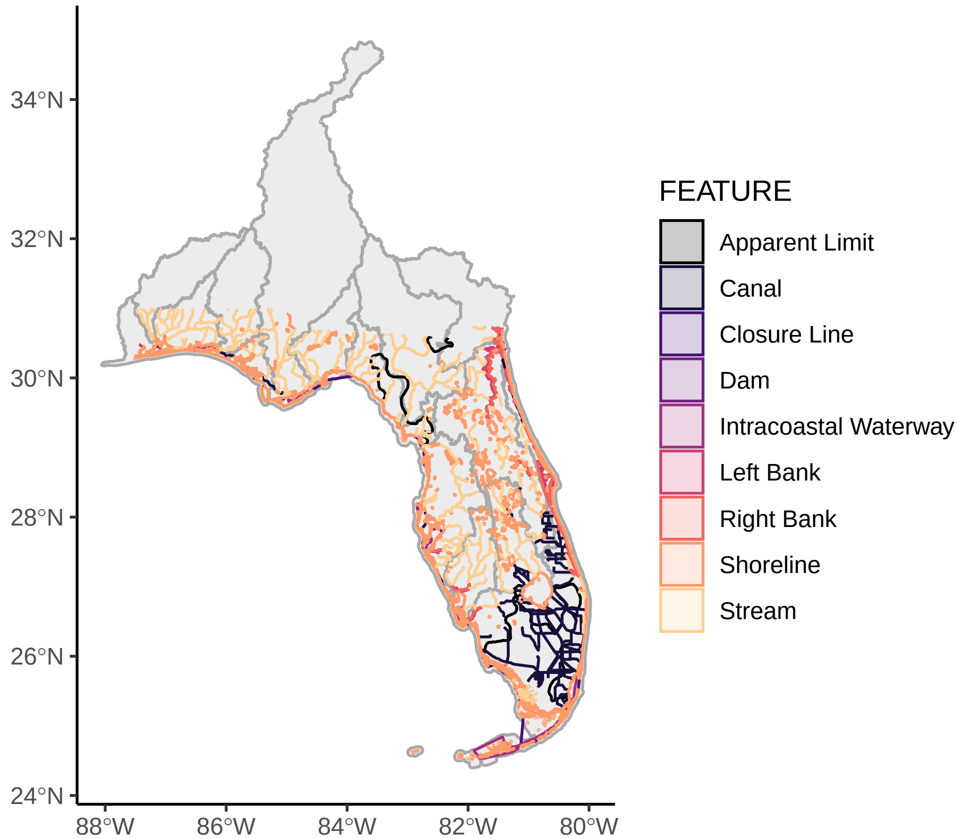
```
## Coordinate Reference System:
##   EPSG: 4269
##   proj4string: "+proj=longlat +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +no_defs"
```

```
waterfeatures.FL <- st_set_crs(waterfeatures.FL, 4269)
st_crs(waterfeatures.FL)
```

```
## 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.

```
FL.map <- ggplot() +
  geom_sf(data = HUC6.FL, color = "darkgray", alpha = 0.7) +
  geom_sf(data = waterfeatures.FL, aes(fill = FEATURE, color = FEATURE), alpha = 0.2) +
    scale_color_viridis_d(option = "magma", end = 0.9) +
  scale_fill_viridis_d(option = "magma", end = 0.9)
print(FL.map)
```



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 shorelines followed by streams, though there is a cluster of canals in the southern 1/3rd of the state. North Carolina was heavily dominated by streams, whereas Florida's border is nearly all coastline. Moreover, Florida seems to have a less extreme distribution of features compared to NC.

Reflection

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

    R is a great open source alternative to GIS software when performing basic spatial analysis. Representing data using maps is a good way of communicating information, particularly to the general public.

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

    My conclusions are based on class discussion and experience making maps prior to this class.

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

    Yes. This section was primarily about the code, so it was helpful to try creating the code myself.

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

    Regarding the data, I was not surprised to find that lake depth is a good indicator of average secchi depth.