Assignment 8: Mapping

Ethan Ready

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/ethanready/Hydrologic_Data_Analysis/Assignments"
```

```
pack<-c("tidyverse", "lubridate", "cowplot", "LAGOSNE", "sf", "maps", "viridis")

invisible(lapply(pack, library, character.only = TRUE))
```

```
## -- 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   0.8.3     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()

##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date

##
## ******************************************************

## 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

## Linking to GEOS 3.7.2, GDAL 2.4.2, PROJ 5.2.0

##
## Attaching package: 'maps'

## The following object is masked from 'package:purrr':
##
##     map

## Loading required package: viridisLite
```

```
theme_set(theme_classic())

# I know you're not supposed to go over 80 characters in a line but this is the only way that I can get

load(
  file = "/Users/ethanready/Hydrologic_Data_Analysis/Data/Raw/LAGOSdata.rda")

USA.waterfeatures <- st_read(
  "/Users/ethanready/Hydrologic_Data_Analysis/Data/Raw/hydrogl020.dbf")
```

```
## Reading layer `hydrogl020' from data source `/Users/ethanready/Hydrologic_Data_Analysis/Data/Raw/hyd
## 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
```

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

```
## Reading layer `WBDHU6' from data source `/Users/ethanready/Hydrologic_Data_Analysis/Data/Raw/Watershe
## 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.

```r
LAGOScomb <-
  left_join(LAGOSdata$epi_nutr, LAGOSdata$locus) %>%
  left_join(., LAGOSdata$lakes_limno) %>%
  left_join(., LAGOSdata$state) %>%
  filter(!is.na(state)) %>%
  select(secchi, lagoslakeid, state, lake_area_ha, maxdepth, nhd_lat, nhd_long)
```

```
## Joining, by = "lagoslakeid"
```

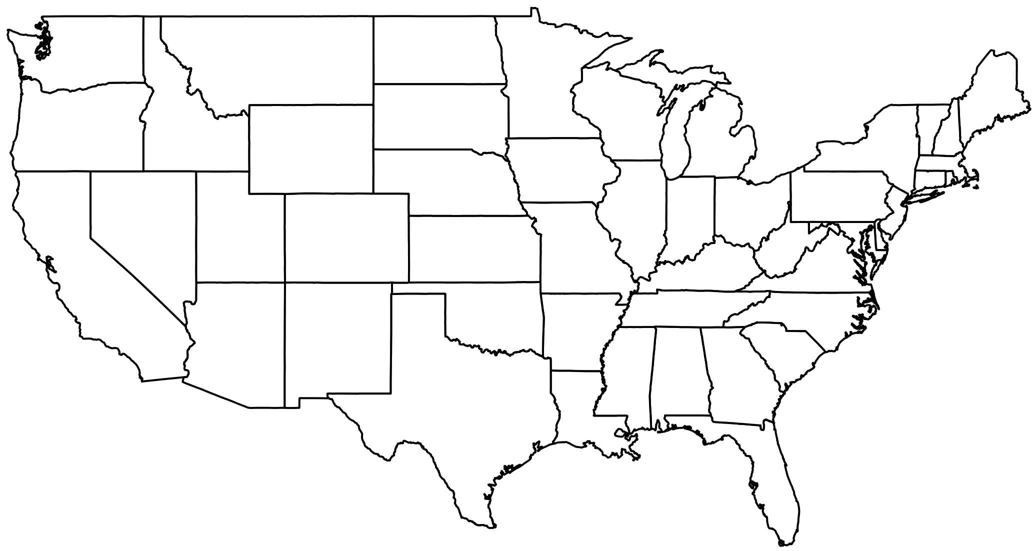
```
## Joining, by = c("lagoslakeid", "nhdid", "nhd_lat", "nhd_long")
```

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

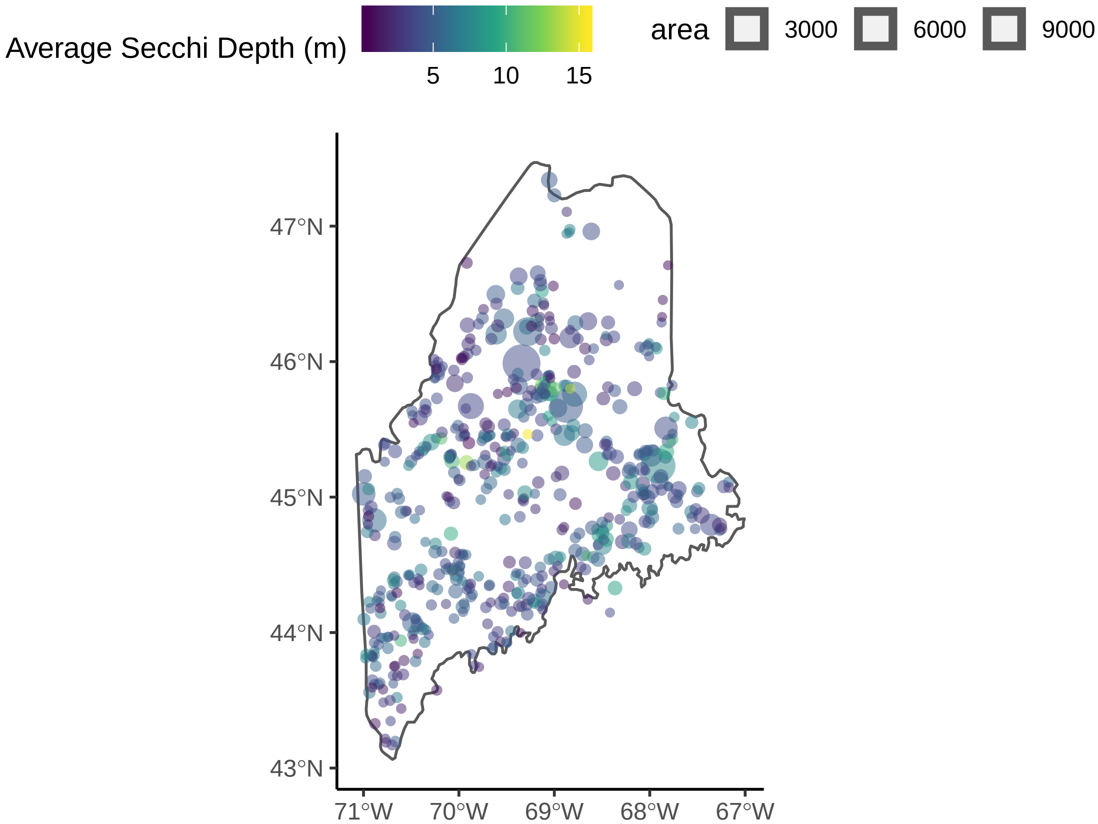
```r
LAGOS_secchi <- LAGOScomb %>%
  filter(state == "ME")%>%
  group_by(lagoslakeid)%>%
  summarise(secchi.mean = mean(secchi),
            area = mean(lake_area_ha),
            depth = mean(maxdepth),
            lat = mean(nhd_lat),
            long = mean(nhd_long))%>%
  drop_na()
```

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

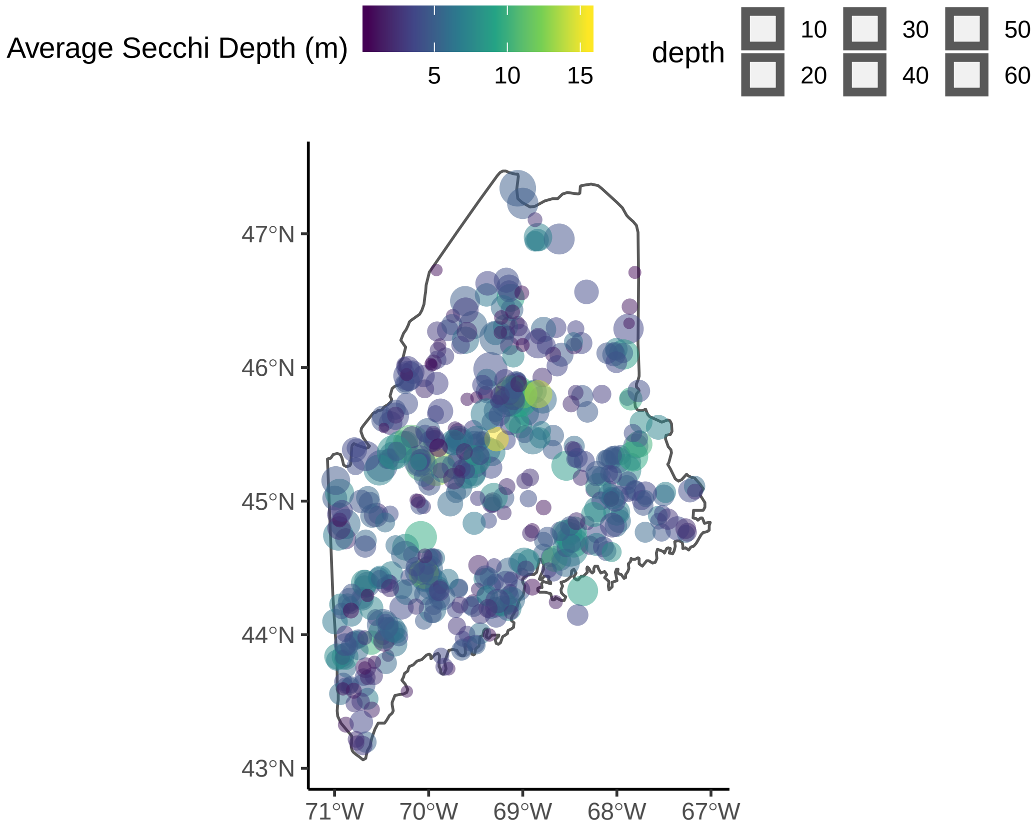


```r
# filter only states that are included in the LAGOSNE database
maine <- filter(states, ID == "maine")


secchi.spatial <- st_as_sf(LAGOS_secchi, coords = c("long", "lat"), crs = 4326)

Secchiplot_area <- ggplot() +
  geom_sf(data = maine, fill = "white") +
  geom_sf(data = secchi.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(Secchiplot_area)
```



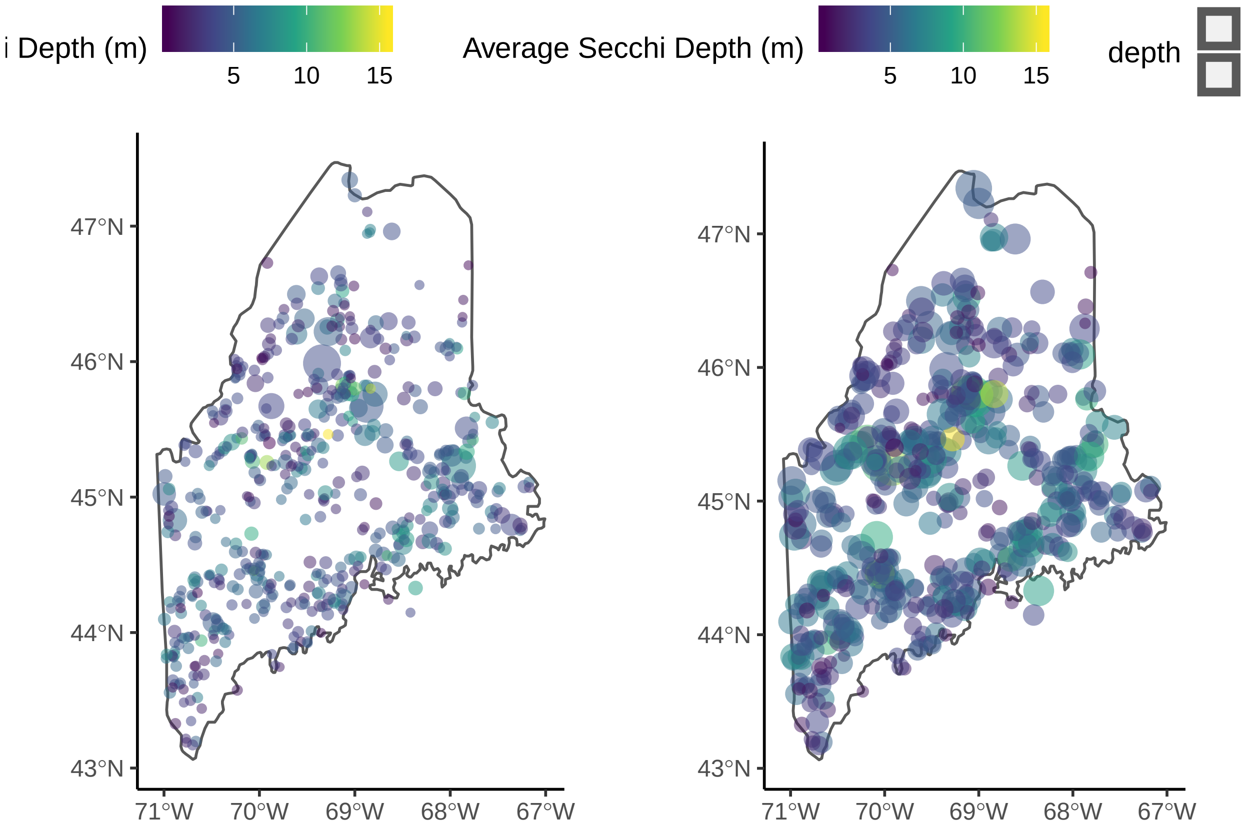
7. Create a second plot, but this time use maximum depth of the lake as the size of the dot.

```
Secchiplot_depth <- ggplot() +
  geom_sf(data = maine, fill = "white") +
  geom_sf(data = secchi.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(Secchiplot_depth)
```



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_area, Secchiplot_depth)
```



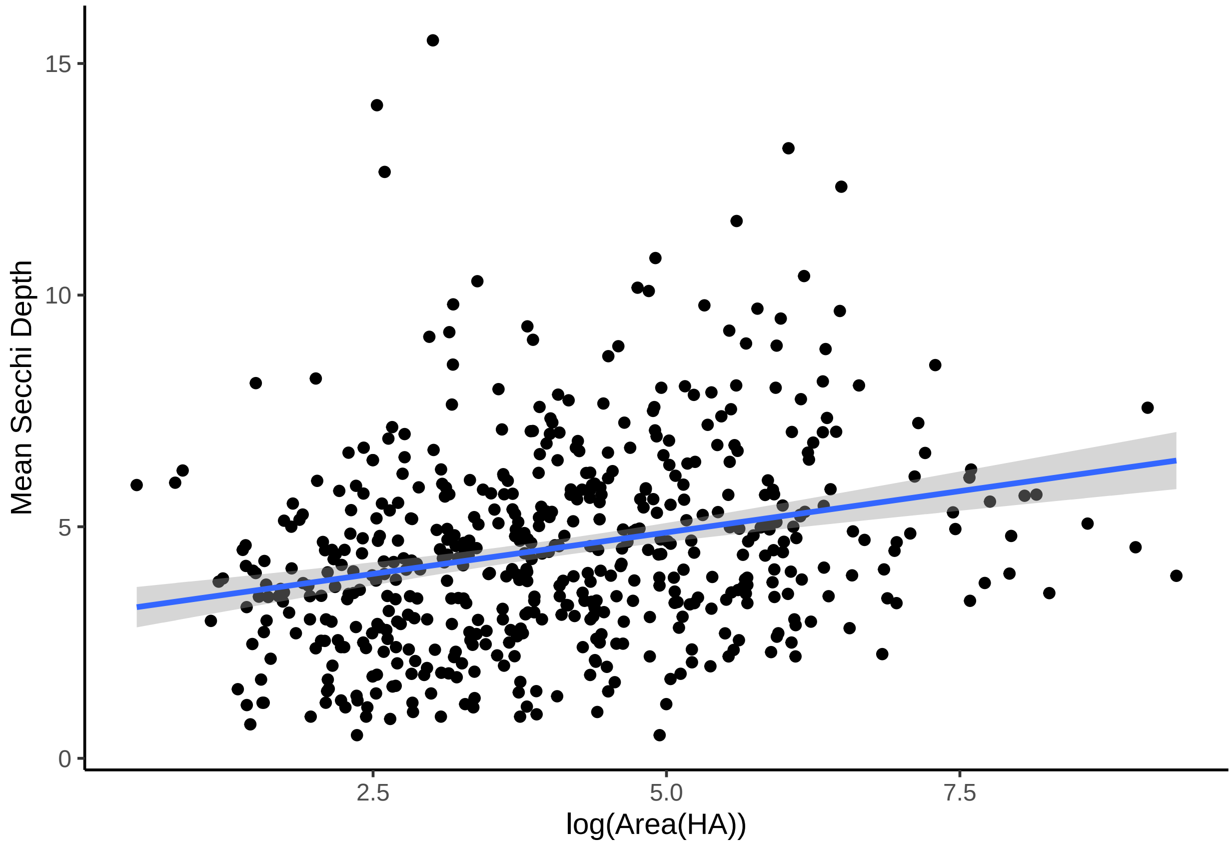
What would you change about the legend to make it a more effective visualization?

> I would put units on the depth and area legends

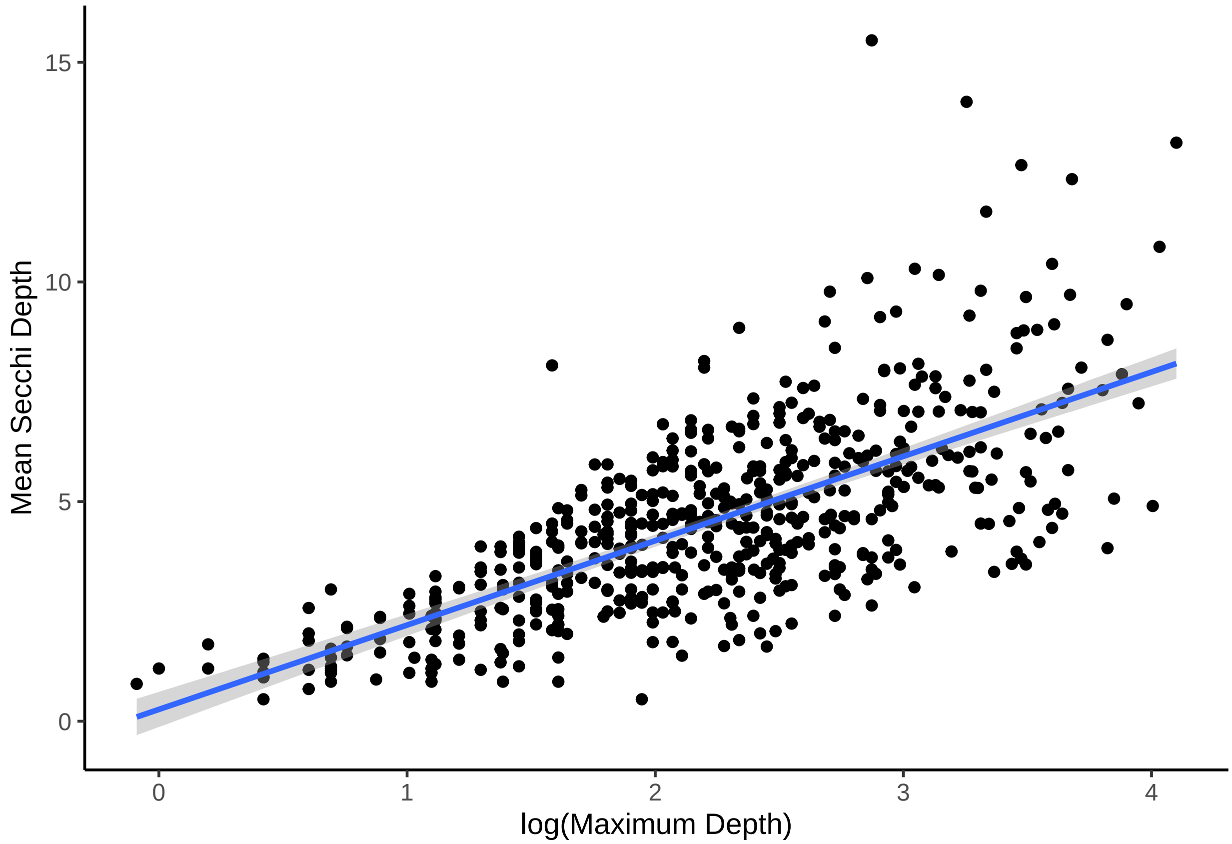
9. What relationships do you see between secchi depth, lake area, and lake depth? Which of the two lake variables seems to be a stronger determinant of secchi depth? (make a scatterplot and run a regression to test this)

Note: consider log-transforming a predictor variable if appropriate

```
Area_linreg_plot <- ggplot(LAGOS_secchi, aes(x = log(area), y = secchi.mean))+
  geom_point()+
  geom_smooth(method = 'lm', formula = y~x)+
  labs(x="log(Area(HA))", y="Mean Secchi Depth")

Depth_linreg_plot <- ggplot(LAGOS_secchi, aes(x = log(depth), y = secchi.mean))+
  geom_point()+
  geom_smooth(method = 'lm', formula = y~x)+
  labs(x="log(Maximum Depth)", y="Mean Secchi Depth")

print(Area_linreg_plot)
```




```r
print(Depth_linreg_plot)
```




```r
lm.depth<-lm(secchi.mean~log(depth), LAGOS_secchi)
lm.area<-lm(secchi.mean~log(area), LAGOS_secchi)

summary(lm.depth)
```

```
##
## Call:
## lm(formula = secchi.mean ~ log(depth), data = LAGOS_secchi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6696 -0.9933 -0.0242  0.8171  9.7151
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.26897    0.20354   1.321    0.187
## log(depth)   1.92036    0.08637  22.235   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.583 on 547 degrees of freedom
## Multiple R-squared:  0.4747, Adjusted R-squared:  0.4738
## F-statistic: 494.4 on 1 and 547 DF,  p-value: < 2.2e-16
```

```
summary(lm.area)
```

```
##
## Call:
## lm(formula = secchi.mean ~ log(area), data = LAGOS_secchi)
##
## 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)  3.08979    0.24734  12.492  < 2e-16 ***
## log(area)    0.35723    0.05682   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
```

> The maximum depth has a much clearer impact on secchi depth. Both area and depth appear to have slight impacts on secci depth, but the coeffecient for the predictor variable is higher and has a lower p-value for the depth regression.

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.

```
USA.waterfeatures.FL<- USA.waterfeatures %>%
  filter(STATE == "FL")

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

st_crs(USA.waterfeatures.FL)
```

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

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

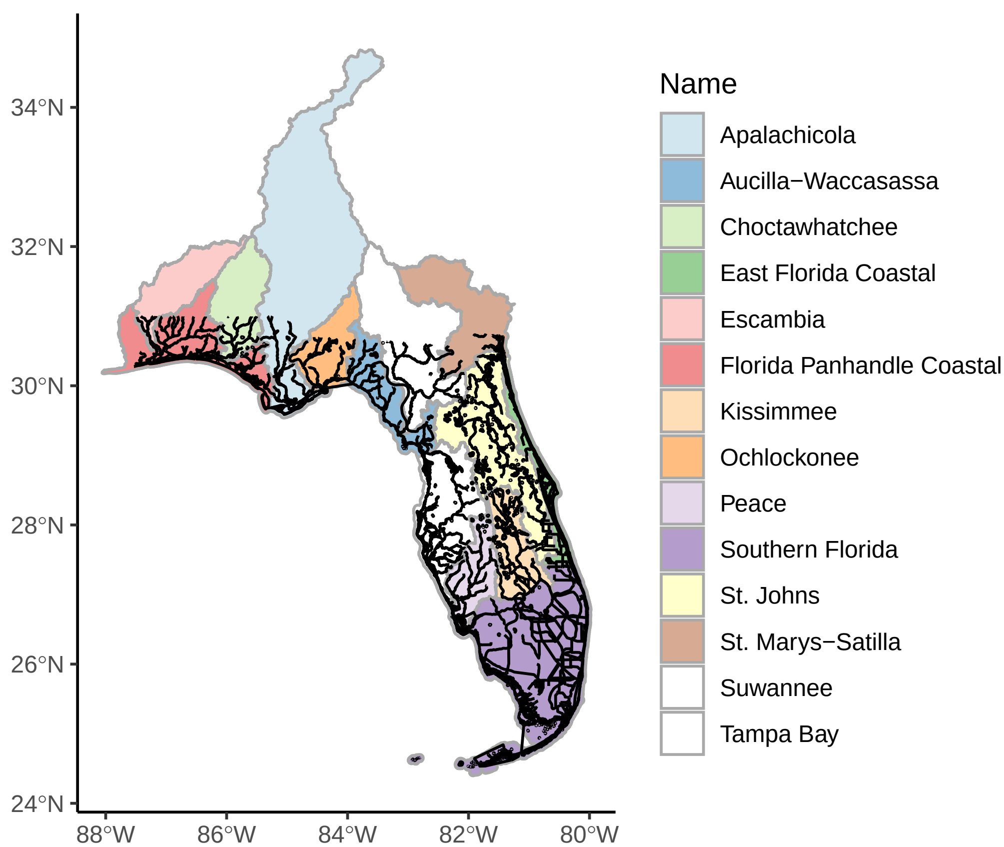
```
## Coordinate Reference System:
##   EPSG: 4269
##   proj4string: "+proj=longlat +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +no_defs"
```

```
USA.waterfeatures.FL <- st_set_crs(USA.waterfeatures.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.

```
FL.layers <- ggplot() +
  geom_sf(data = HUC6.FL, aes(fill = Name), color = "darkgray", alpha = 0.5) +
  geom_sf(data = USA.waterfeatures.FL) +
  scale_fill_brewer(palette = "Paired")
print(FL.layers)
```

```
## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum for palette Paired is 12
## Returning the palette you asked for with that many colors
```



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

    The dominant water features seem to be coastal wetlands, as the whole coast has lots of creeks. Especially South Florida, which seems to be less of a river basin and more of one giant wetland. North Carolina is different in that it has some coastal wetlands but also lots of big traditional rivers and mountain streams.

Reflection

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

1) It's possible to create geographic data modelling in R. 2) Lots of data is better visualized through geogrphic representation. 3) There's data out there made for geographic modelling.

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

    The river basins maps and secchi depth maps

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

    Yes, I saw how easy it is to create maps

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

    I though more lakes in maine would be clear, especially ones in wilderness areas