Final_Wrangling_Doc

BGRR

11/11/2019

```
getwd()
```

```
## [1] "C:/Users/Felipe/OneDrive - Duke University/1. DUKE/Ramos 3 Semestre/722 Hydro Data/HDA_Project_
```

```
#load packages
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   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()
```

```
library(lubridate)
```

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

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

```
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(mapview)

#load LAGOS data
LAGOSdata <- lagosne_load()
```

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

```
# creating specific lagos files
LAGOSstate <- LAGOSdata$state
LAGOSlocus <- LAGOSdata$locus
LAGOSnutrient <- LAGOSdata$epi_nutr
```

```
LAGOSiwslulc <- LAGOSdata$iws.lulc
LAGOSiws <- LAGOSdata$iws

theme_set(theme_classic())
```

```
#checking each state for nutrient data

#State 14: Minnesota
LAGOSlocus.MN <- LAGOSlocus %>% filter (state_zoneid == "State_14")
LAGOSnutrient.MN <- LAGOSlocus.MN %>%
  left_join(LAGOSnutrient, by = "lagoslakeid")
```

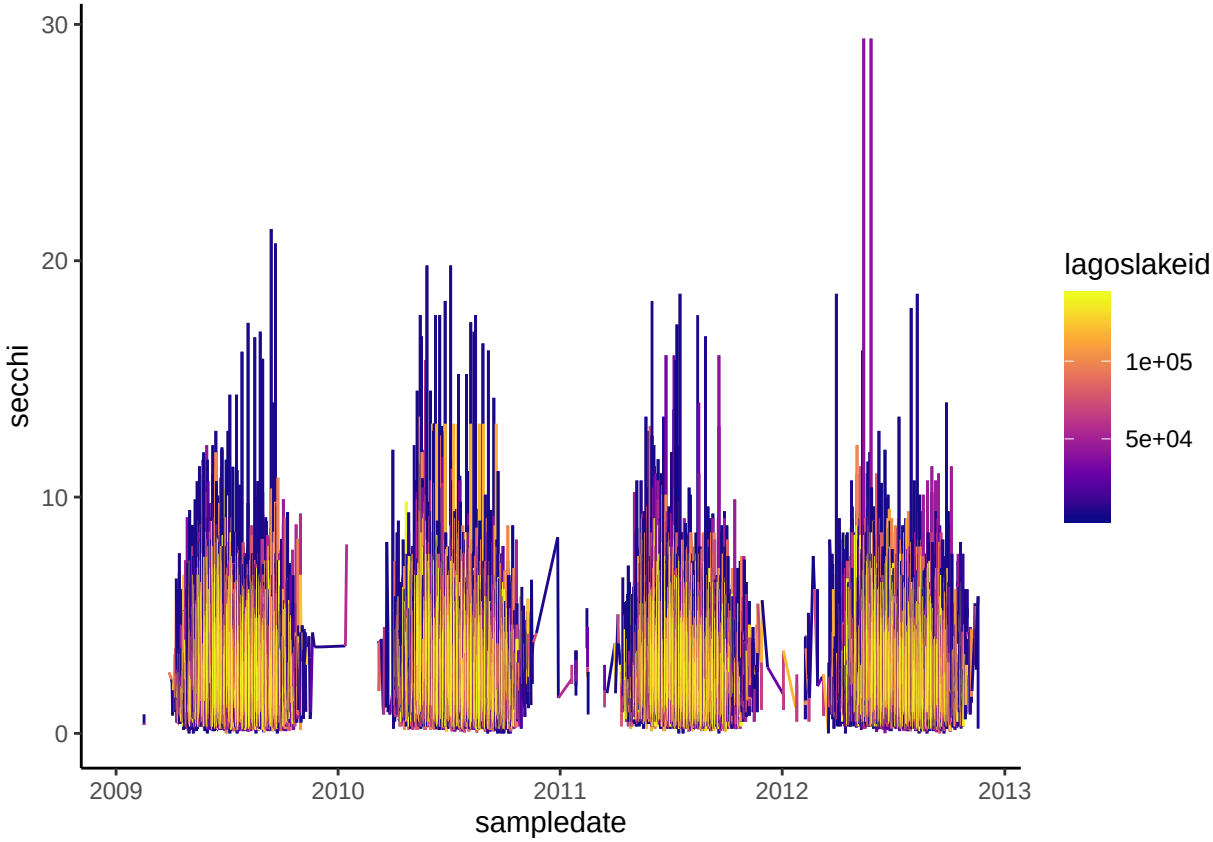
```
#getting the area of the iws
LAGOSiws.area <- select(LAGOSiws, "lagoslakeid", "iws_ha")


#joining MN locus with iwslulc and adding the area of IWS
LAGOSiws.MN <- LAGOSlocus.MN %>%
  left_join(LAGOSiwslulc, by = "lagoslakeid") %>%
  left_join(LAGOSiws.area, by = "lagoslakeid")

##selecting 2011 lulc
LAGOSiws2011.MN <- LAGOSiws.MN %>%
  select(lagoslakeid,state_zoneid, lake_area_ha, iws_ha,
         iws_nlcd2011_pct_11, iws_nlcd2011_pct_21, iws_nlcd2011_pct_22,
         iws_nlcd2011_pct_23, iws_nlcd2011_pct_24, iws_nlcd2011_pct_31,
         iws_nlcd2011_pct_41, iws_nlcd2011_pct_42, iws_nlcd2011_pct_43,
         iws_nlcd2011_pct_52, iws_nlcd2011_pct_71, iws_nlcd2011_pct_81,
         iws_nlcd2011_pct_82, iws_nlcd2011_pct_90, iws_nlcd2011_pct_95)
```

```
#filtering state nutrient data #FRA. We can expand this range
LAGOSnutrient.MN.skinny <- LAGOSnutrient.MN %>%
  filter(sampledate > "2008-12-31" & sampledate < "2015-01-01")

# sapply(LAGOSnutrient.MN.skinny, summary)
```

```
##Joining iws.lulc and nutrient
LAGOSiws.nutrient.2011.MN <-  left_join(LAGOSnutrient.MN.skinny,
                                        LAGOSiws2011.MN, by =
                                          c("lagoslakeid", "lake_area_ha"))

LAGOSiws.nutrient.2011.MN <- LAGOSiws.nutrient.2011.MN %>%
  select(lagoslakeid, nhd_lat, nhd_long, lake_area_ha, lake_perim_meters,
         iws_zoneid, iws_ha, state_zoneid.x, elevation_m, sampledate, chla,secchi,
         iws_nlcd2011_pct_11, iws_nlcd2011_pct_21, iws_nlcd2011_pct_22,
         iws_nlcd2011_pct_23, iws_nlcd2011_pct_24, iws_nlcd2011_pct_31,
         iws_nlcd2011_pct_41, iws_nlcd2011_pct_42, iws_nlcd2011_pct_43,
         iws_nlcd2011_pct_52, iws_nlcd2011_pct_71, iws_nlcd2011_pct_81,
         iws_nlcd2011_pct_82, iws_nlcd2011_pct_90, iws_nlcd2011_pct_95)

# sapply(LAGOSiws.nutrient.2011.MN, summary)
```

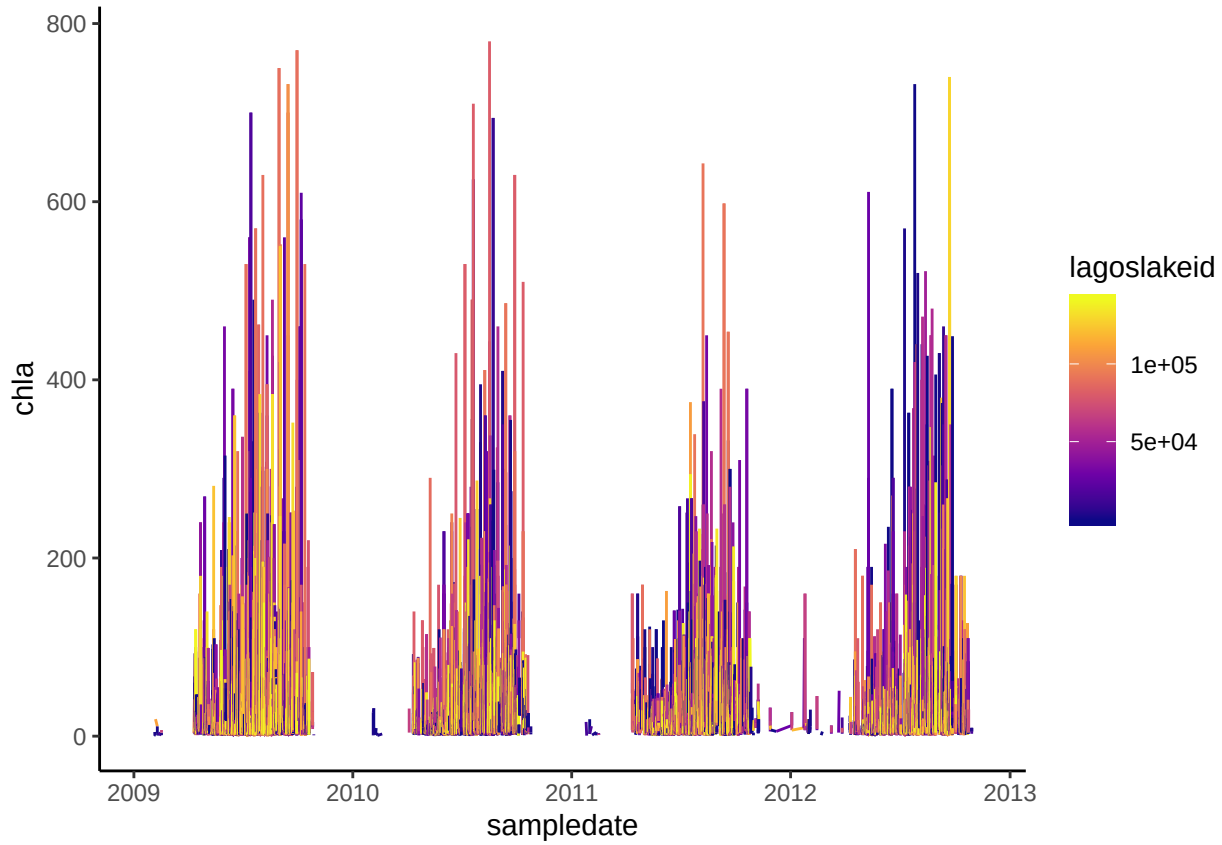
```
# Looking at the data

ggplot(LAGOSiws.nutrient.2011.MN, aes(x = sampledate, y = secchi, color = lagoslakeid)) +
```

```
  scale_color_viridis_c(option = "plasma") +
  geom_line()
```




```
ggplot(LAGOSiws.nutrient.2011.MN, aes(x = sampledate, y = chla, color = lagoslakeid)) +
  scale_color_viridis_c(option = "plasma") +
  geom_line()
```

```
## Warning: Removed 10 rows containing missing values (geom_path).
```




```
LAGOS.MN.processed <- LAGOSiws.nutrient.2011.MN %>%
  mutate(Water.pct = iws_nlcd2011_pct_11,
         Urban.pct = iws_nlcd2011_pct_21 + iws_nlcd2011_pct_22 +
           iws_nlcd2011_pct_23 + iws_nlcd2011_pct_24,
         Undeveloped.pct = iws_nlcd2011_pct_31 + iws_nlcd2011_pct_41 +
           iws_nlcd2011_pct_42 + iws_nlcd2011_pct_43 +
           iws_nlcd2011_pct_52 + iws_nlcd2011_pct_90 + iws_nlcd2011_pct_95,
         Ag.pct =  iws_nlcd2011_pct_81 + iws_nlcd2011_pct_82 + iws_nlcd2011_pct_71) %>%
  select(-c(iws_nlcd2011_pct_11, iws_nlcd2011_pct_21, iws_nlcd2011_pct_22,
            iws_nlcd2011_pct_23, iws_nlcd2011_pct_24, iws_nlcd2011_pct_31,
            iws_nlcd2011_pct_41, iws_nlcd2011_pct_42, iws_nlcd2011_pct_43,
            iws_nlcd2011_pct_52, iws_nlcd2011_pct_71, iws_nlcd2011_pct_81,
            iws_nlcd2011_pct_82, iws_nlcd2011_pct_90, iws_nlcd2011_pct_95)) %>%
  na.omit() %>%
  mutate(LakeIWS.Ratio = lake_area_ha/iws_ha) %>%
  mutate(DOY = yday(sampledate))

# sapply(LAGOS.MN.processed, summary)
```

```
#We are creating growing seasons; early, prime, late. They will be based off of water temperature and d

LAGOS.MN.processed$EarlyTrue <- LAGOS.MN.processed$DOY < 136 #Before May 15
LAGOS.MN.processed$PrimeTrue <- LAGOS.MN.processed$DOY >=136 & LAGOS.MN.processed$DOY <= 273 #May 15 to
LAGOS.MN.processed$LateTrue <- LAGOS.MN.processed$DOY > 273  #October 1 and later

LAGOS.MN.processed$EarlyTrue <- ifelse(LAGOS.MN.processed$EarlyTrue == TRUE, "Early", "No")
```

```
LAGOS.MN.processed$PrimeTrue <- ifelse(LAGOS.MN.processed$PrimeTrue == TRUE, "Prime", "No")
LAGOS.MN.processed$LateTrue <- ifelse(LAGOS.MN.processed$LateTrue == TRUE, "Late", "No")

LAGOS.MN.processed$Season <- LAGOS.MN.processed$EarlyTrue == "Early"

LAGOS.MN.processed$Season[LAGOS.MN.processed$EarlyTrue == "Early"] <- "Early"
LAGOS.MN.processed$Season[LAGOS.MN.processed$PrimeTrue == "Prime"] <- "Prime"
LAGOS.MN.processed$Season[LAGOS.MN.processed$LateTrue == "Late"] <- "Late"

LAGOS.MN.processed  <- LAGOS.MN.processed %>%
  select(-c(EarlyTrue, PrimeTrue, LateTrue))
```

```
LAGOS.MN.processed.sf <- st_as_sf(LAGOS.MN.processed, coords = c("nhd_long", "nhd_lat"), crs = 4326)

LAGOS.MN.processed.UTM.sf <- st_transform(LAGOS.MN.processed.sf, crs=26917)

MN.Ecoregions.sf <- st_read('./Data/Raw/mn_eco_l3.shp')
```

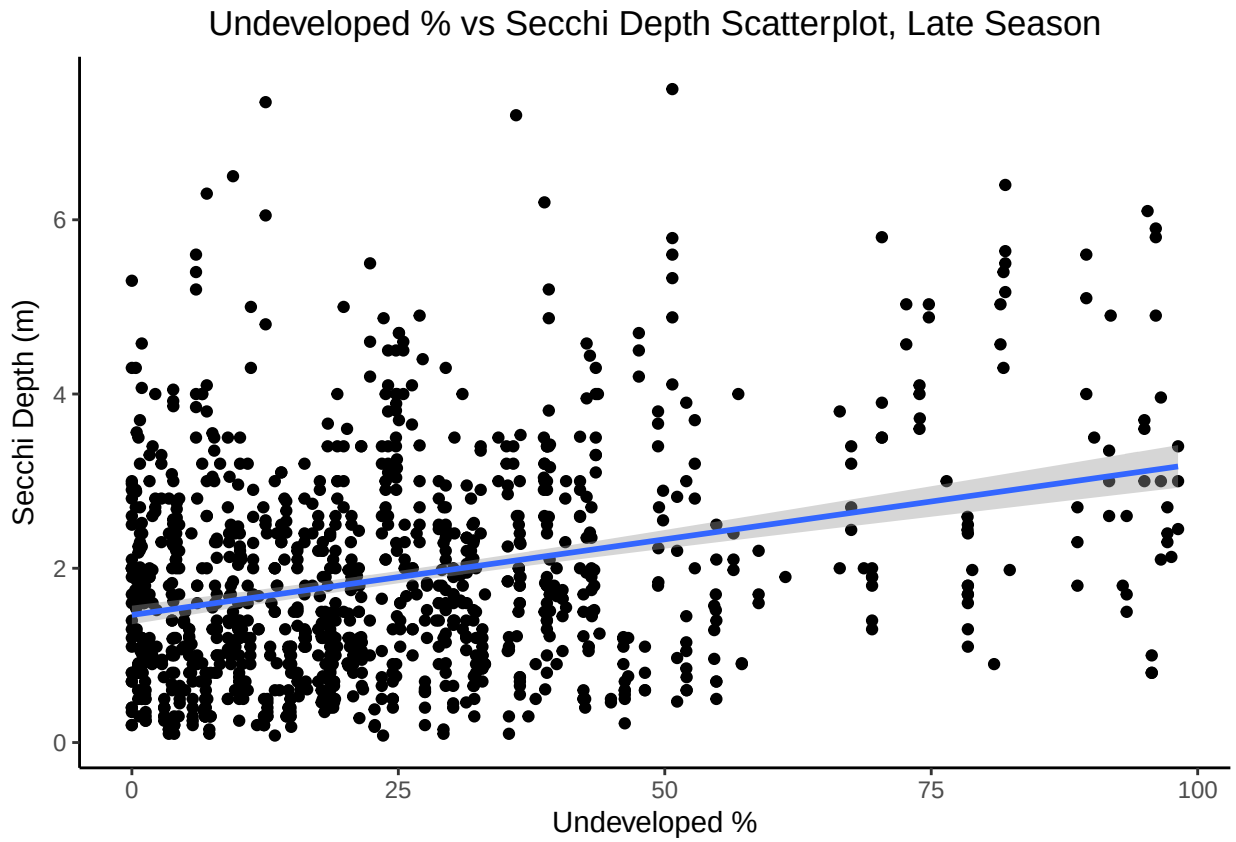
```
## Reading layer `mn_eco_l3' from data source `C:\Users\Felipe\OneDrive - Duke University\1. DUKE\Ramos
## Simple feature collection with 7 features and 13 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -91854.57 ymin: 2278542 xmax: 489296.4 ymax: 2930681
## epsg (SRID):    NA
## proj4string:    +proj=aea +lat_1=29.5 +lat_2=45.5 +lat_0=23 +lon_0=-96 +x_0=0 +y_0=0 +datum=NAD83 +u
```

```
#Selecting level 3 ecoregions names
MN.Ecoregions.sf <- select(MN.Ecoregions.sf, US_L3NAME)

#mapview(MN.Ecoregions.sf)

MN.Ecoregions.UTM.sf <- st_transform(MN.Ecoregions.sf, crs=26917)

LAGOS.MN.processed.sf <- st_join(LAGOS.MN.processed.UTM.sf, MN.Ecoregions.UTM.sf)
```

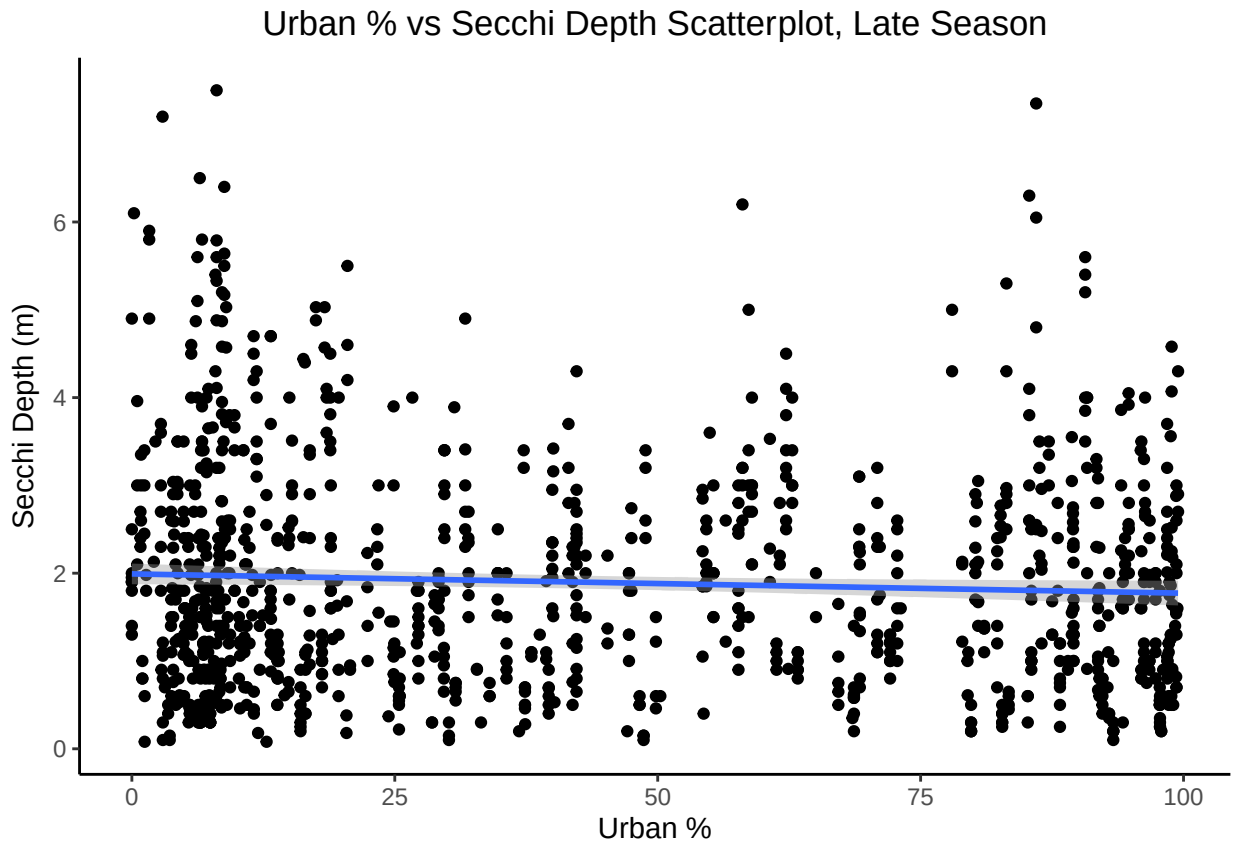
```
#Creating sf seasons files
LAGOS.MN.processed.Early.sf <- filter(LAGOS.MN.processed.sf, Season == "Early")
LAGOS.MN.processed.Prime.sf <- filter(LAGOS.MN.processed.sf, Season == "Prime")
LAGOS.MN.processed.Late.sf <- filter(LAGOS.MN.processed.sf, Season == "Late")

#Creating regular season files (doesn't have ecoregions column)
LAGOS.MN.processed.Early <- filter(LAGOS.MN.processed, Season == "Early")
LAGOS.MN.processed.Prime <- filter(LAGOS.MN.processed, Season == "Prime")
LAGOS.MN.processed.Late <- filter(LAGOS.MN.processed, Season == "Late")

# summary(LAGOS.MN.processed.Early.sf$chla)
# summary(LAGOS.MN.processed.Prime.sf$chla)
# summary(LAGOS.MN.processed.Late.sf$chla)
#
# summary(LAGOS.MN.processed.Early.sf$secchi)
# summary(LAGOS.MN.processed.Prime.sf$secchi)
# summary(LAGOS.MN.processed.Late.sf$secchi)
```

```
ggplot(LAGOS.MN.processed.Late.sf, aes(x = Undeveloped.pct, y = secchi)) +
geom_point() +
```

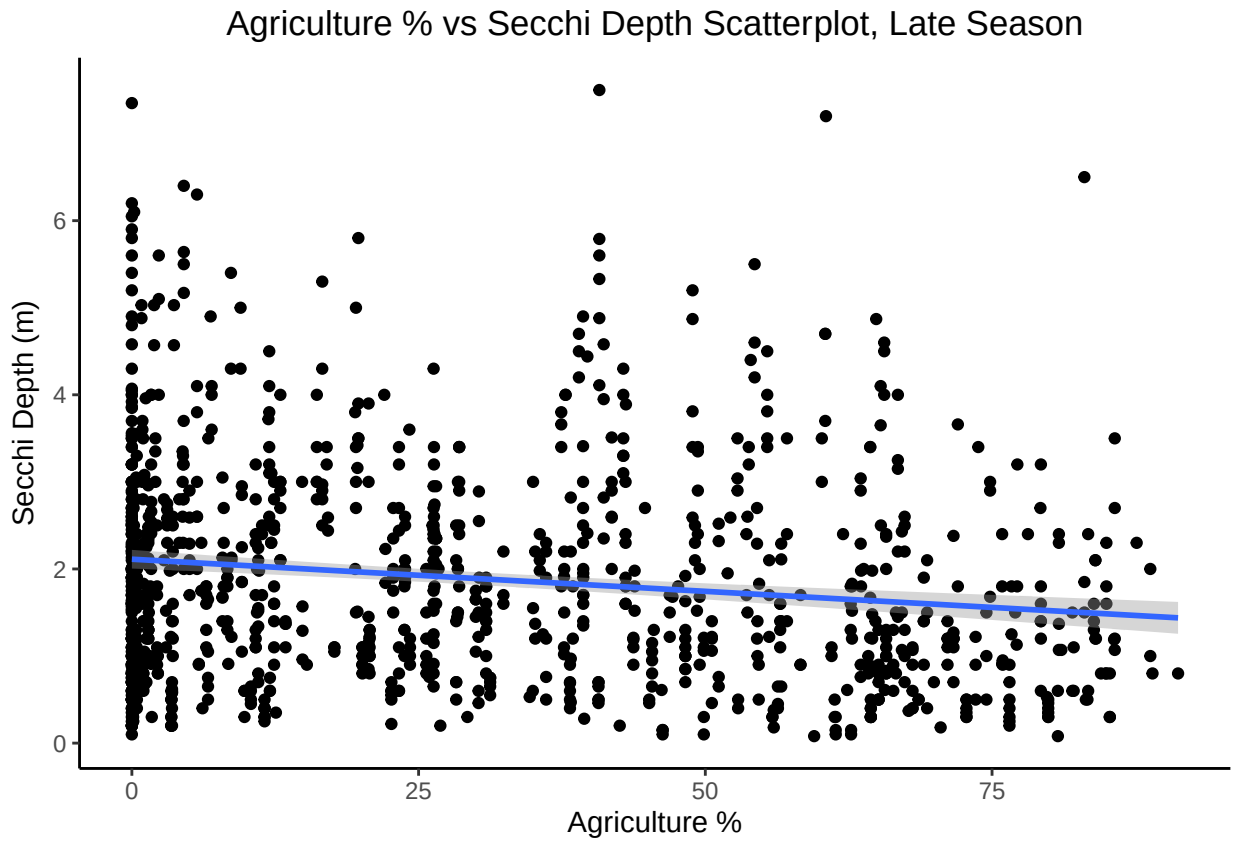
```
geom_smooth(method=lm) +
xlab(expression("Undeveloped %")) +
ylab(expression("Secchi Depth (m)")) +
ggtitle("Undeveloped % vs Secchi Depth Scatterplot, Late Season") +
theme(plot.title = element_text(hjust = 0.5))
```



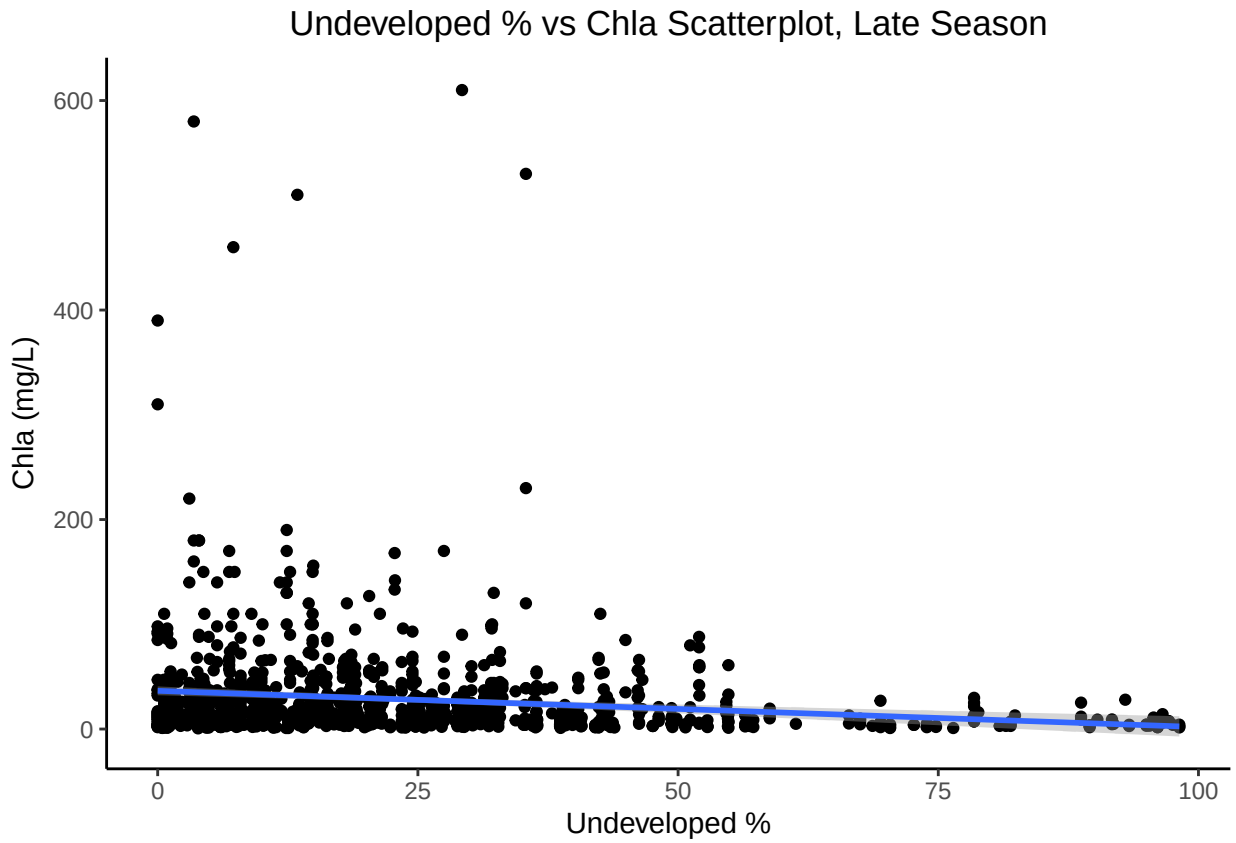

```
ggplot(LAGOS.MN.processed.Late.sf, aes(x = Urban.pct, y = secchi)) +
geom_point() +
geom_smooth(method=lm) +
xlab(expression("Urban %")) +
ylab(expression("Secchi Depth (m)")) +
ggtitle("Urban % vs Secchi Depth Scatterplot, Late Season") +
theme(plot.title = element_text(hjust = 0.5))
```



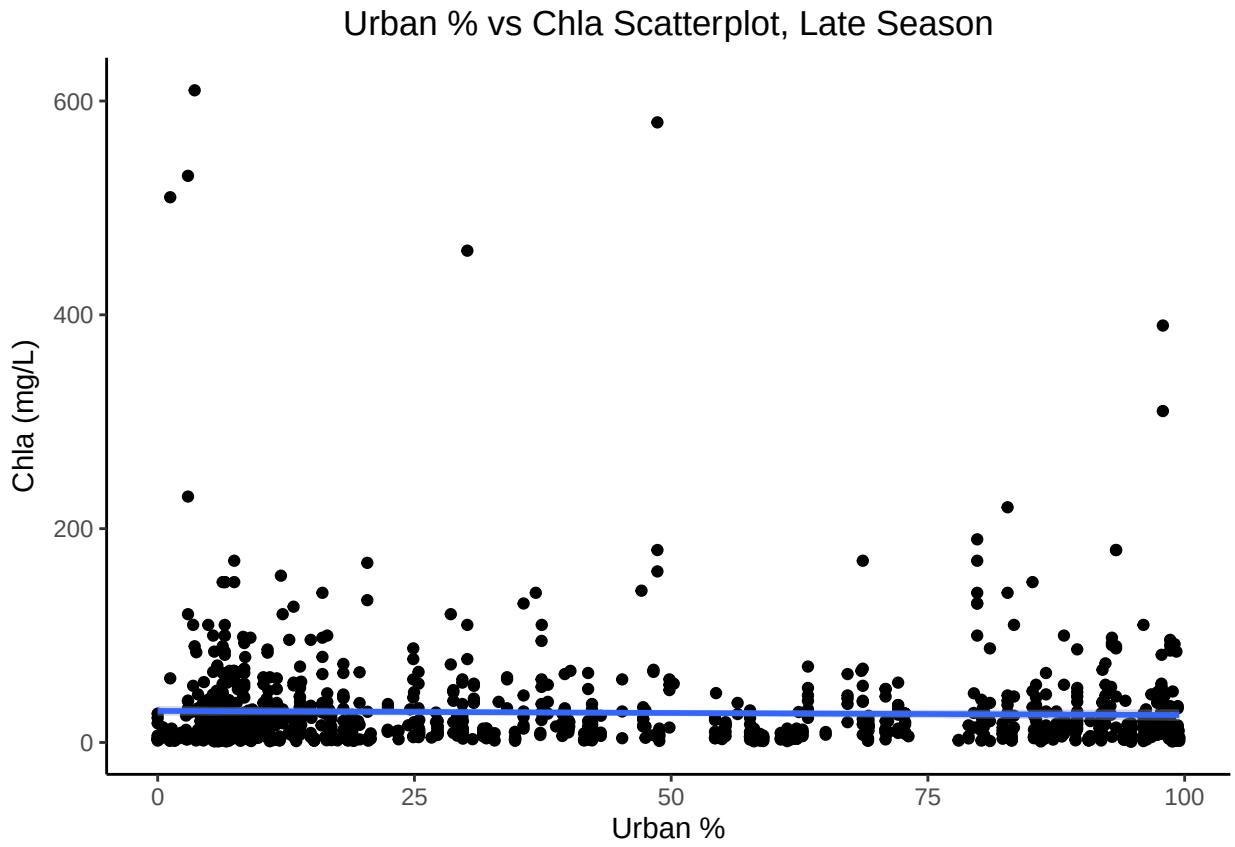

```
ggplot(LAGOS.MN.processed.Late.sf, aes(x = Ag.pct, y = secchi)) +
geom_point() +
geom_smooth(method=lm) +
xlab(expression("Agriculture %")) +
ylab(expression("Secchi Depth (m)")) +
ggtitle("Agriculture % vs Secchi Depth Scatterplot, Late Season") +
theme(plot.title = element_text(hjust = 0.5))
```



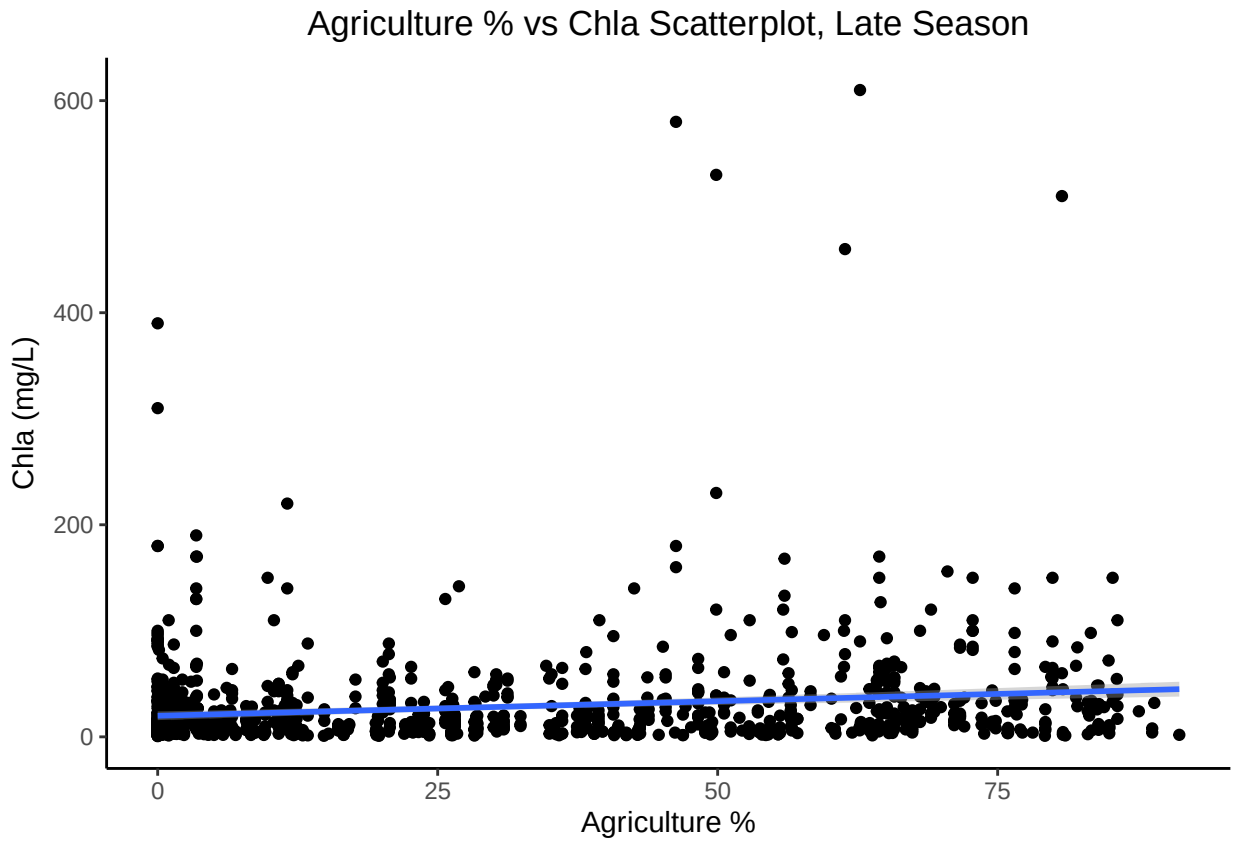

```
ggplot(LAGOS.MN.processed.Late.sf, aes(x = Undeveloped.pct, y = chla)) +
geom_point() +
geom_smooth(method=lm) +
xlab(expression("Undeveloped %")) +
ylab(expression("Chla (mg/L)")) +
ggtitle("Undeveloped % vs Chla Scatterplot, Late Season") +
theme(plot.title = element_text(hjust = 0.5))
```




```
ggplot(LAGOS.MN.processed.Late.sf, aes(x = Urban.pct, y = chla)) +
geom_point() +
geom_smooth(method=lm) +
xlab(expression("Urban %")) +
ylab(expression("Chla (mg/L)")) +
ggtitle("Urban % vs Chla Scatterplot, Late Season") +
theme(plot.title = element_text(hjust = 0.5))
```

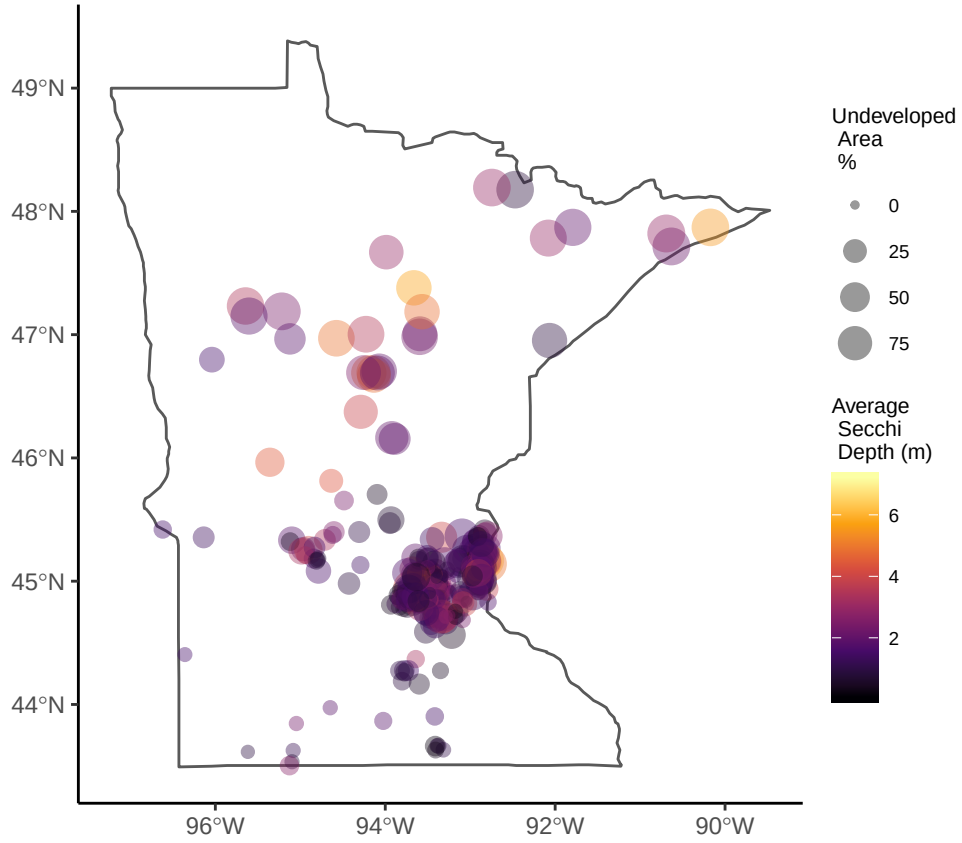



```
ggplot(LAGOS.MN.processed.Late.sf, aes(x = Ag.pct, y = chla)) +
geom_point() +
geom_smooth(method=lm) +
xlab(expression("Agriculture %")) +
ylab(expression("Chla (mg/L)")) +
ggtitle("Agriculture % vs Chla Scatterplot, Late Season") +
theme(plot.title = element_text(hjust = 0.5))
```



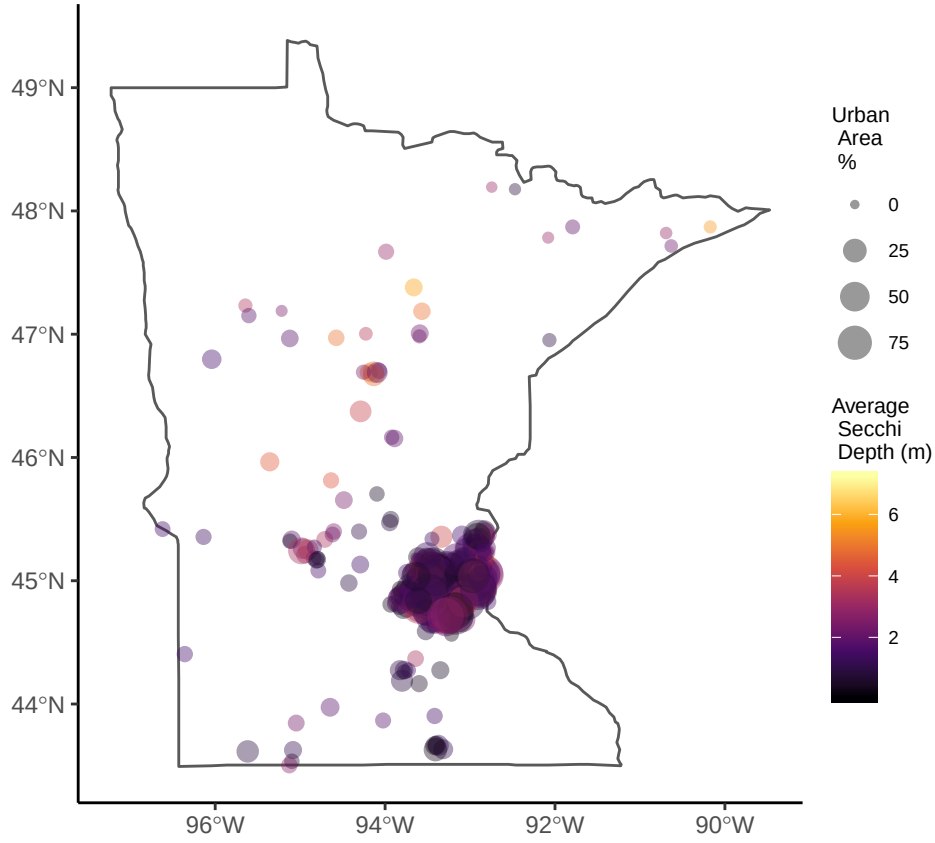

```
LAGOS.MN.Summary.Late <- LAGOS.MN.processed.Late %>%
group_by(lagoslakeid) %>%
summarise(secchi.mean = mean(secchi),
          chla.mean = mean(chla),
          lake.area = mean(lake_area_ha),
          iws.area = mean(iws_ha),
          LakeIWS.Ratio = mean(LakeIWS.Ratio),
          Water.pct = mean(Water.pct),
          Urban.pct = mean(Urban.pct),
          Undeveloped.pct = mean(Undeveloped.pct),
          Ag.pct = mean(Ag.pct),
          Lat = mean(nhd_lat),
          Long = mean(nhd_long)
          ) %>%
  drop_na()

#SF file with the summary
LAGOS.MN.Summary.Late.sf <- st_as_sf(LAGOS.MN.Summary.Late, coords = c("Long", "Lat"), crs = 4326)


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

# filter MN
states.MN <- filter(states, ID %in%
```

```
c("minnesota"))

secchiplot1.Late.MN <- ggplot() +
  geom_sf(data = states.MN, fill = "white") +
  geom_sf(data = LAGOS.MN.Summary.Late.sf,
          aes(size = Undeveloped.pct, color = secchi.mean),
          alpha = 0.4, show.legend = "point") +
  scale_color_viridis_c(option = "inferno") +
  labs(color = "Average \n Secchi \n Depth (m)",
       size = "Undeveloped \n Area \n %") +
  theme(legend.position = "right", legend.text = element_text(size = 7),
        legend.title = element_text(size = 8),
        legend.margin = margin(0,0,0,0))
print(secchiplot1.Late.MN)
```

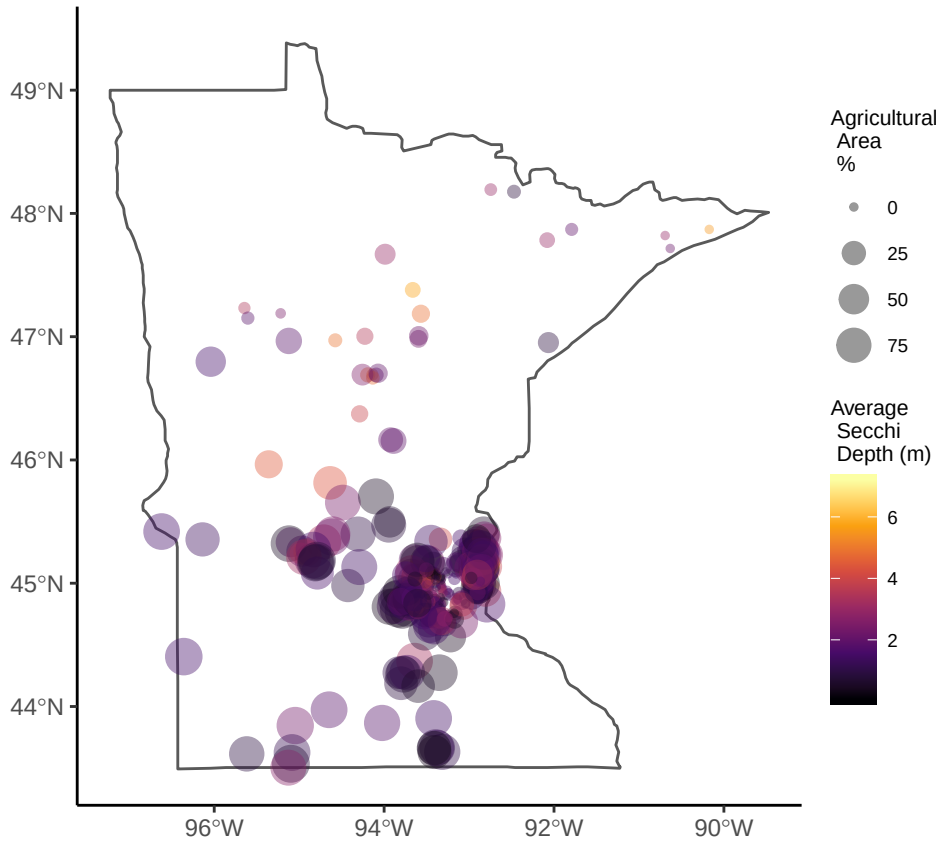



```
secchiplot2.Late.MN <- ggplot() +
  geom_sf(data = states.MN, fill = "white") +
  geom_sf(data = LAGOS.MN.Summary.Late.sf,
          aes(size = Urban.pct, color = secchi.mean),
          alpha = 0.4, show.legend = "point") +
  scale_color_viridis_c(option = "inferno") +
  labs(color = "Average \n Secchi \n Depth (m)",
       size = "Urban \n Area \n %") +
  theme(legend.position = "right", legend.text = element_text(size = 7),
        legend.title = element_text(size = 8),
        legend.margin = margin(0,0,0,0))
```

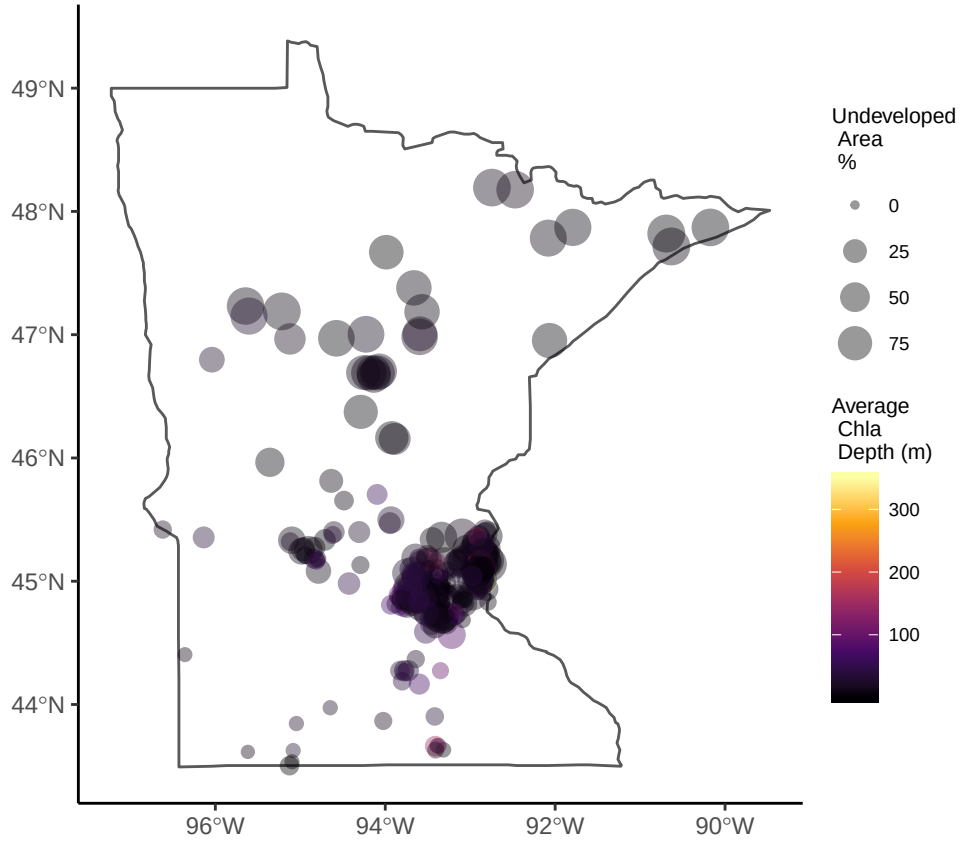
```
print(secchiplot2.Late.MN)
```



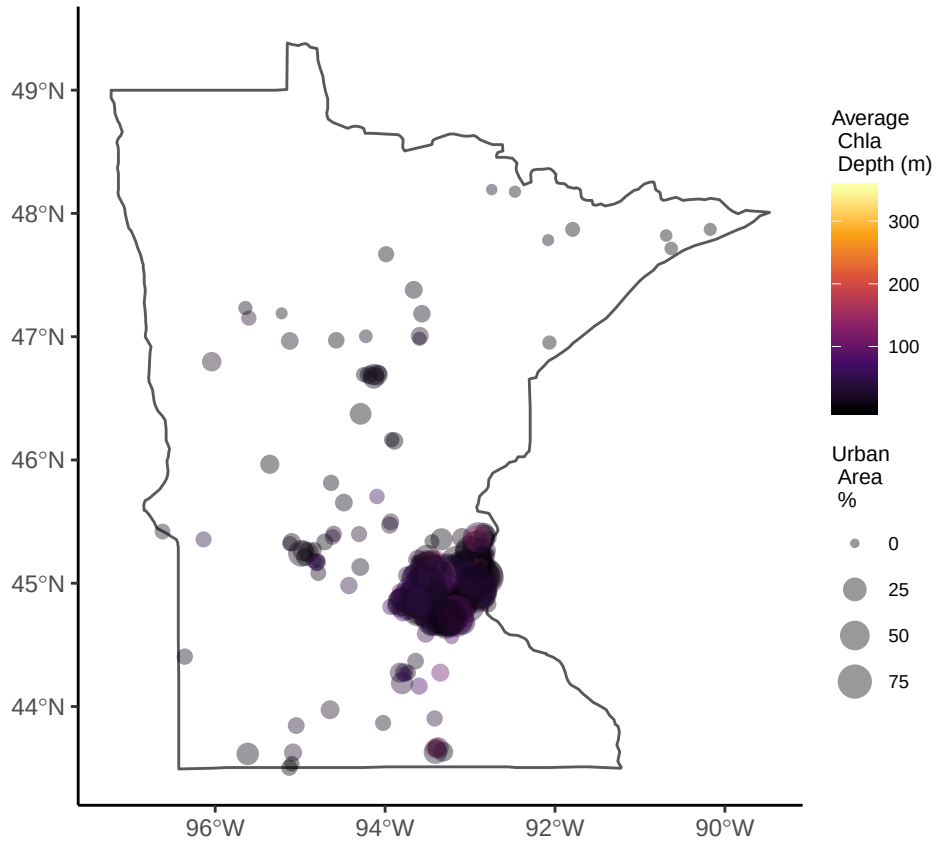

```
secchiplot3.Late.MN <- ggplot() +
  geom_sf(data = states.MN, fill = "white") +
  geom_sf(data = LAGOS.MN.Summary.Late.sf,
          aes(size = Ag.pct, color = secchi.mean),
          alpha = 0.4, show.legend = "point") +
  scale_color_viridis_c(option = "inferno") +
  labs(color = "Average \n Secchi \n Depth (m)",
       size = "Agricultural \n Area \n %") +
  theme(legend.position = "right", legend.text = element_text(size = 7),
        legend.title = element_text(size = 8),
        legend.margin = margin(0,0,0,0))
print(secchiplot3.Late.MN)
```



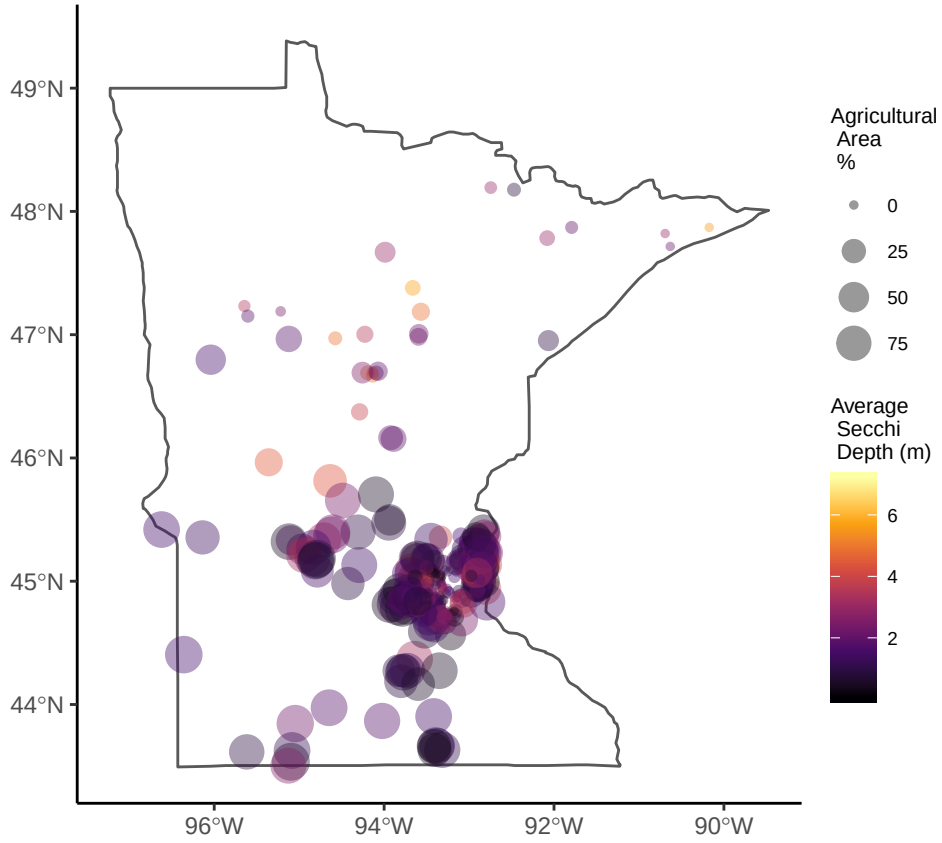

```
chlaplot1.Late.MN <- ggplot() +
  geom_sf(data = states.MN, fill = "white") +
  geom_sf(data = LAGOS.MN.Summary.Late.sf,
          aes(size = Undeveloped.pct, color = chla.mean),
          alpha = 0.4, show.legend = "point") +
  scale_color_viridis_c(option = "inferno") +
  labs(color = "Average \n Chla \n Depth (m)",
       size = "Undeveloped \n Area \n %") +
  theme(legend.position = "right", legend.text = element_text(size = 7),
        legend.title = element_text(size = 8),
        legend.margin = margin(0,0,0,0))
print(chlaplot1.Late.MN)
```




```
chlaplot2.Late.MN <- ggplot() +
  geom_sf(data = states.MN, fill = "white") +
  geom_sf(data = LAGOS.MN.Summary.Late.sf,
          aes(size = Urban.pct, color = chla.mean),
          alpha = 0.4, show.legend = "point") +
  scale_color_viridis_c(option = "inferno") +
  labs(color = "Average \n Chla \n Depth (m)",
       size = "Urban \n Area \n %") +
  theme(legend.position = "right", legend.text = element_text(size = 7),
        legend.title = element_text(size = 8),
        legend.margin = margin(0,0,0,0))
print(chlaplot2.Late.MN)
```




```
chlaplot3.Late.MN <- ggplot() +
  geom_sf(data = states.MN, fill = "white") +
  geom_sf(data = LAGOS.MN.Summary.Late.sf,
          aes(size = Ag.pct, color = chla.mean),
          alpha = 0.4, show.legend = "point") +
  scale_color_viridis_c(option = "inferno") +
  labs(color = "Average \n Chla \n Depth (m)",
       size = "Agricultural \n Area \n %") +
  theme(legend.position = "right", legend.text = element_text(size = 7),
        legend.title = element_text(size = 8),
        legend.margin = margin(0,0,0,0))
print(secchiplot3.Late.MN)
```




```
st_geometry(LAGOS.MN.processed.Early.sf) <- NULL
write.csv(x =  LAGOS.MN.processed.Early.sf, row.names = FALSE,
          file =  "./data/processed/LAGOS.MN.processed.Early.csv")

st_geometry(LAGOS.MN.processed.Prime.sf) <- NULL
write.csv(x =  LAGOS.MN.processed.Prime.sf, row.names = FALSE,
          file =  "./data/processed/LAGOS.MN.processed.Prime.csv")

st_geometry(LAGOS.MN.processed.Late.sf) <- NULL
write.csv(x =  LAGOS.MN.processed.Late.sf, row.names = FALSE,
          file =  "./data/processed/LAGOS.MN.processed.Late.csv")
```