Data Exploration: Environmental Statistics Final Project

Cecilia Martinez

DUE: 11/8 11:59pm

```r
library(tigris)
library(sf)
library(spmodel)
library(spatstat)
library(ggplot2)
library(ggmap)
library(readr)
library(tidyr)
library(dplyr)
library(scales)
```

I. Data Description and Wrangling

In this exploratory analysis, I examine potential connections between specific climate variables and the annual growth of Whitebark Pine (*Pinus albicaulis*) in Wyoming, as inferred from tree rings. Additionally, I aim to investigate whether spatial dependence exists in growth responses before and after accounting for some climate variables

The data used in this analysis comprise three main components: tree ring time series data, forest inventory data that have location data and can be linked to the tree ring time series data **(FIA Data Mart)**, and long-term climate data obtained at the plot level. Each permanent sample plot, rather than each individual tree, has associated latitude and longitude coordinates, meaning the spatial resolution of the climate data is plot-based rather than individual tree-based. Climate data were sourced from the **PRISM Climate Group** at Oregon State University at a 4km resolution, capturing monthly climate values since 1895 and climate normals, MAT and MAP, for each associated plot location.

```r
# data loading and prep
wbp_clim_all <- read_csv("wbp_all_climate_data_all.csv")
wbp_rw_all <- read_csv("wbp_rw_all.csv")
wbp_meta_all <- read_csv("wbp_meta_all.csv")

# filter relevant attributed from FIA datamart tables
wbp_meta_st <- wbp_meta_all %>%
  dplyr::select(TRE_CN, PLOT_CN, STATECD, LAT, LON)

# make plot identifiers characters instead of numeric
wbp_clim_all$PLOT_CN <- as.character(wbp_clim_all$PLOT_CN)
wbp_meta_st$PLOT_CN <- as.character(wbp_meta_st$PLOT_CN)
wbp_rw_all$PLOT_CN <- as.character(wbp_rw_all$PLOT_CN)

# filter all the data sets to just include the state of wyoming
```

```r
wbp_rw_all <- left_join(wbp_rw_all, wbp_meta_st, by = c("TRE_CN", "PLOT_CN"))
wbp_wy <- wbp_rw_all %>% filter(STATECD == 56)
wbp_clim_wy <- wbp_clim_all %>% filter(PLOT_CN %in% unique(wbp_wy$PLOT_CN))
```

II. Visualizing and Mapping the Data

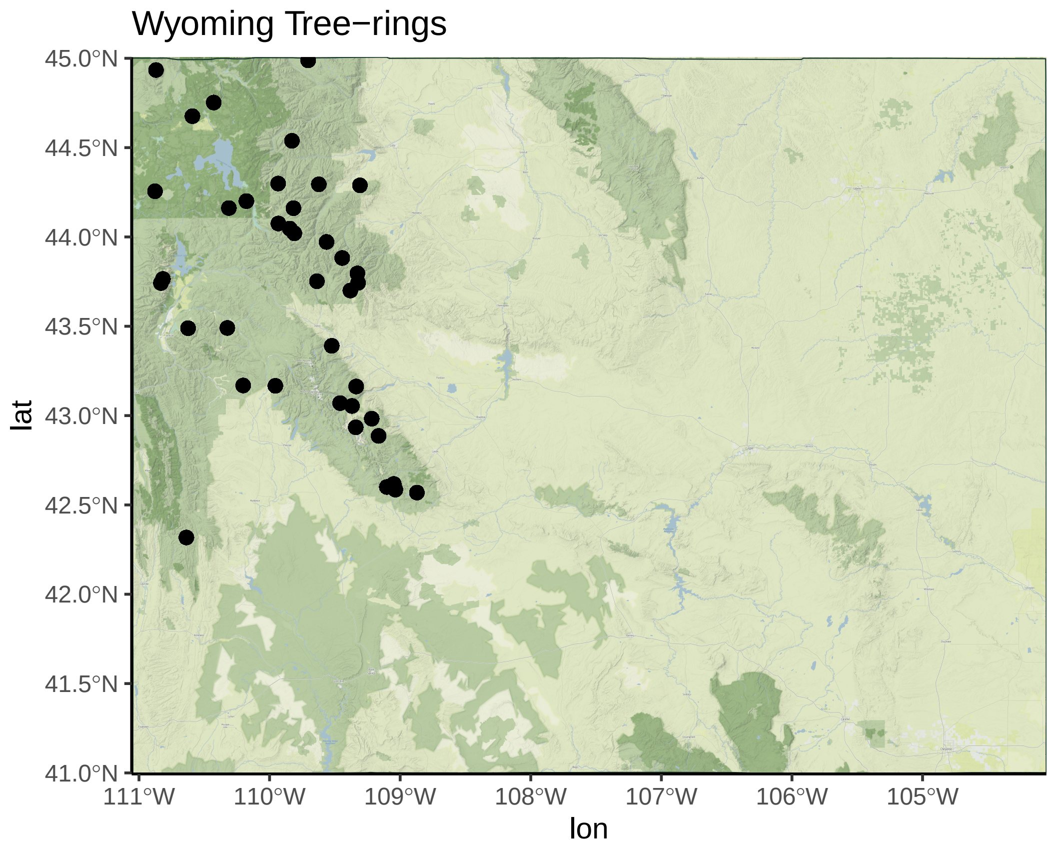
Mapping the data offers a spatial overview of tree ring locations, mean annual temperature (MAT), and mean annual precipitation (MAP) across Wyoming. There are a total of 50 plotted for this preliminary analysis.

The first map displays tree ring data points against a terrain basemap, showing that these trees are located in high-elevation mountainous regions (which we would expect from what we know about the ecology of these trees). The second map depicts the spatial distribution of MAT across trees, suggesting potential clustering across the landscape. Lastly, the third map illustrates MAP across plots, which looks like there could be some spatial dependence. It appears that the spatial dependence might be more pronounced for MAP than for MAT, potentially due to differences in scaling. To address this issue we could standardize MAT and MAP to directly compare.

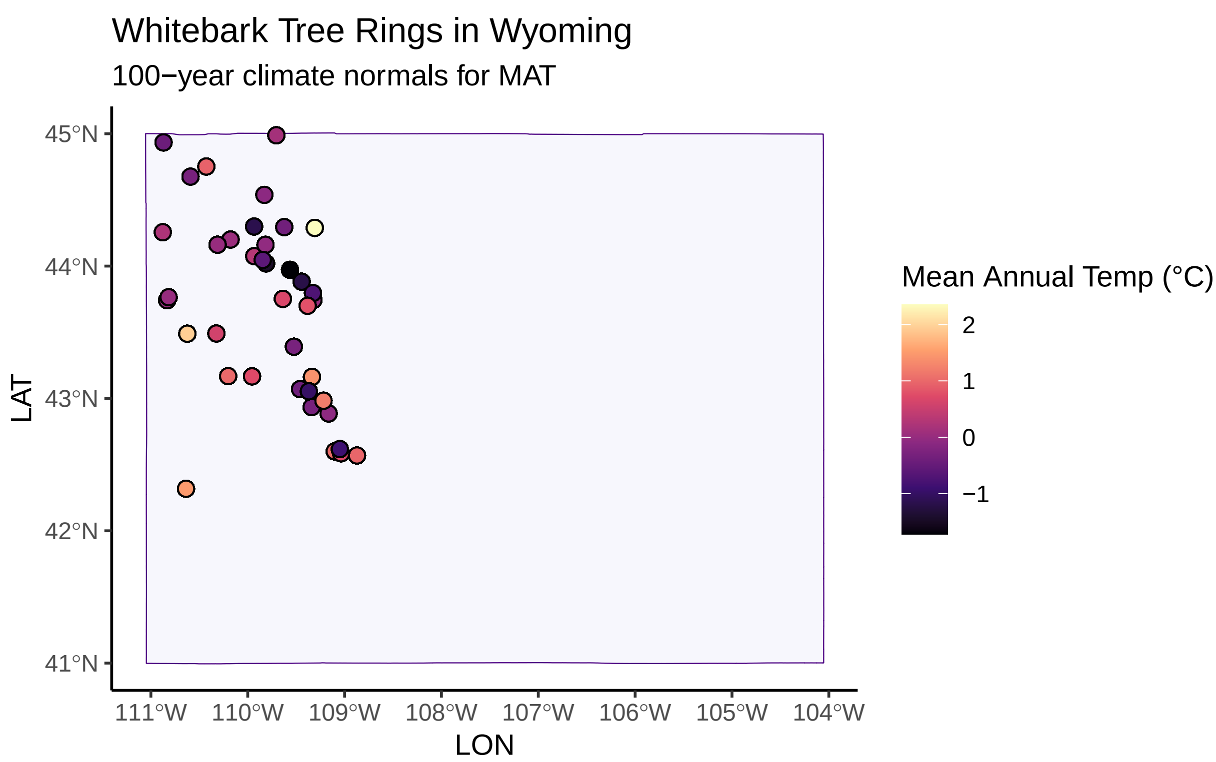
```r
us_states <- states()
```

```
##   |                                                                          |
```

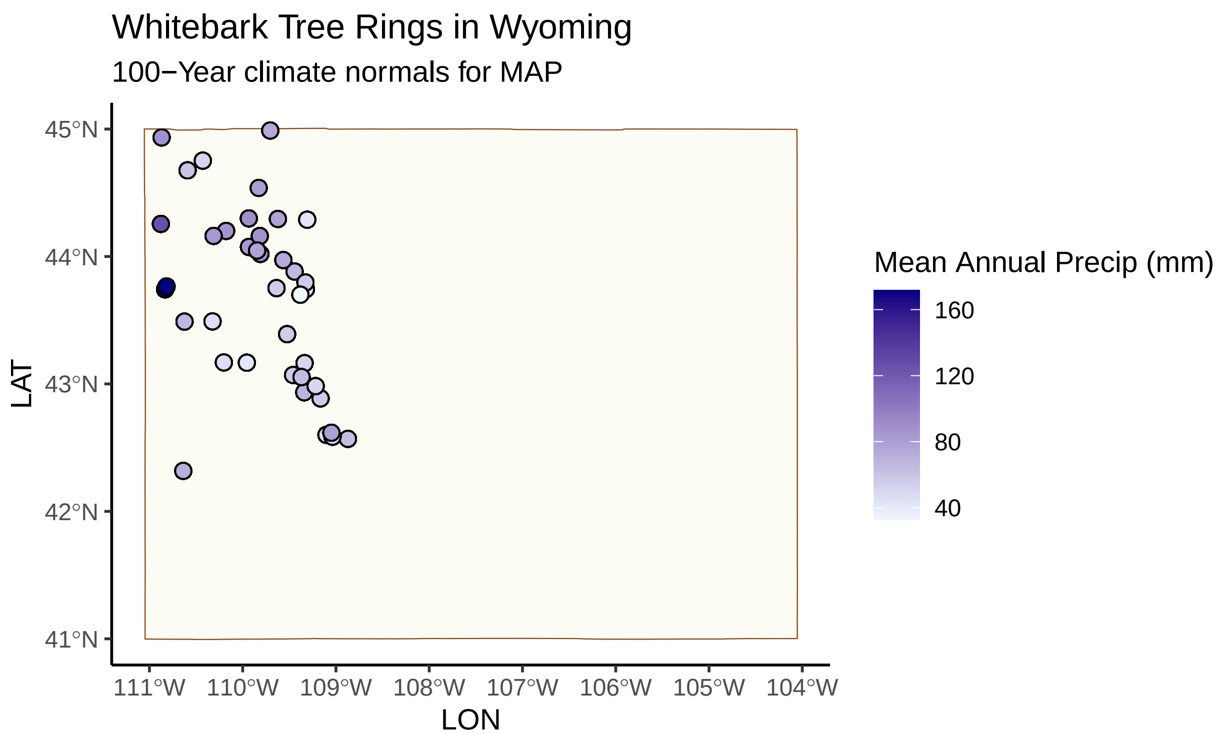
```r
wyoming <- us_states[us_states$NAME == "Wyoming", ] %>%
  dplyr::select(geometry)

wyoming_bbox <- c(left = -111.056, bottom = 40.9942, right = -104.052, top = 45.0059)
wyoming_map <- get_stadiamap(bbox = wyoming_bbox, maptype = "stamen_terrain", zoom = 10)

# plot tree riunig data with terrain baselayer
ggmap(wyoming_map) +
  geom_sf(data = wyoming, fill = "#DDF2D1", color = "#18392B", alpha = 0.2, inherit.aes = FALSE) +
  geom_point(aes(x = LON, y = LAT), data = wbp_wy, shape = 21, fill = "black", size = 2) +
  ggtitle("Wyoming Tree-rings") +
  theme_classic()
```




```r
#transforming crs for wyoming state for better plotting
wyoming_sf <- st_transform(wyoming, crs = 4326)

#extract climate normals from climate df
climate_normals <- wbp_clim_wy %>%
  select(PLOT_CN, meantemp, precip) %>%
  distinct()
#join the climate data to the wbp dataframe
wbp_wy_clim <- wbp_wy %>%
  left_join(climate_normals, by = "PLOT_CN")

#plot with points colored by 100 year climate normals (MAT)
ggplot(data = wyoming_sf) +
  geom_sf(fill = "#E6E6FA", color = "#4B0082", alpha = 0.3) +
  geom_point(
    data = wbp_wy_clim,
    aes(x = LON, y = LAT, fill = meantemp),
    shape = 21, size = 2.5
  ) +
  scale_fill_viridis_c(name = "Mean Annual Temp (°C)", option = "magma", direction = 1) +
  labs(title = "Whitebark Tree Rings in Wyoming", subtitle = "100-year climate normals for MAT") +
  theme_classic() +
  coord_sf()
```




```r
#plot with points colored by 100 year climate normals (MAP)
ggplot(data = wyoming_sf) +
  geom_sf(fill = "#f5f5dc", color = "#8b4513", alpha = 0.3) +
  geom_point(
    data = wbp_wy_clim,
    aes(x = LON, y = LAT, fill = precip),
    shape = 21, size = 2.5
  ) + # Points colored by precipitation
  scale_fill_gradient(
    name = "Mean Annual Precip (mm)",
    low = "#f0f8ff",
    high = "#000080"
  ) +
  labs(title = "Whitebark Tree Rings in Wyoming", subtitle = "100-Year climate normals for MAP") +
  theme_classic() +
  coord_sf()
```



III. Modeling Tree Growth and Exploring Residuals for Spatial Dependence

```r
#let's scale cliamte variables first to compare effect sizes of model outputs
wbp_wy_clim <-  wbp_wy_clim %>%
  mutate(
    meantemp_scaled = as.numeric(scale(meantemp)),
    precip_scaled = as.numeric(scale(precip))
  )
```

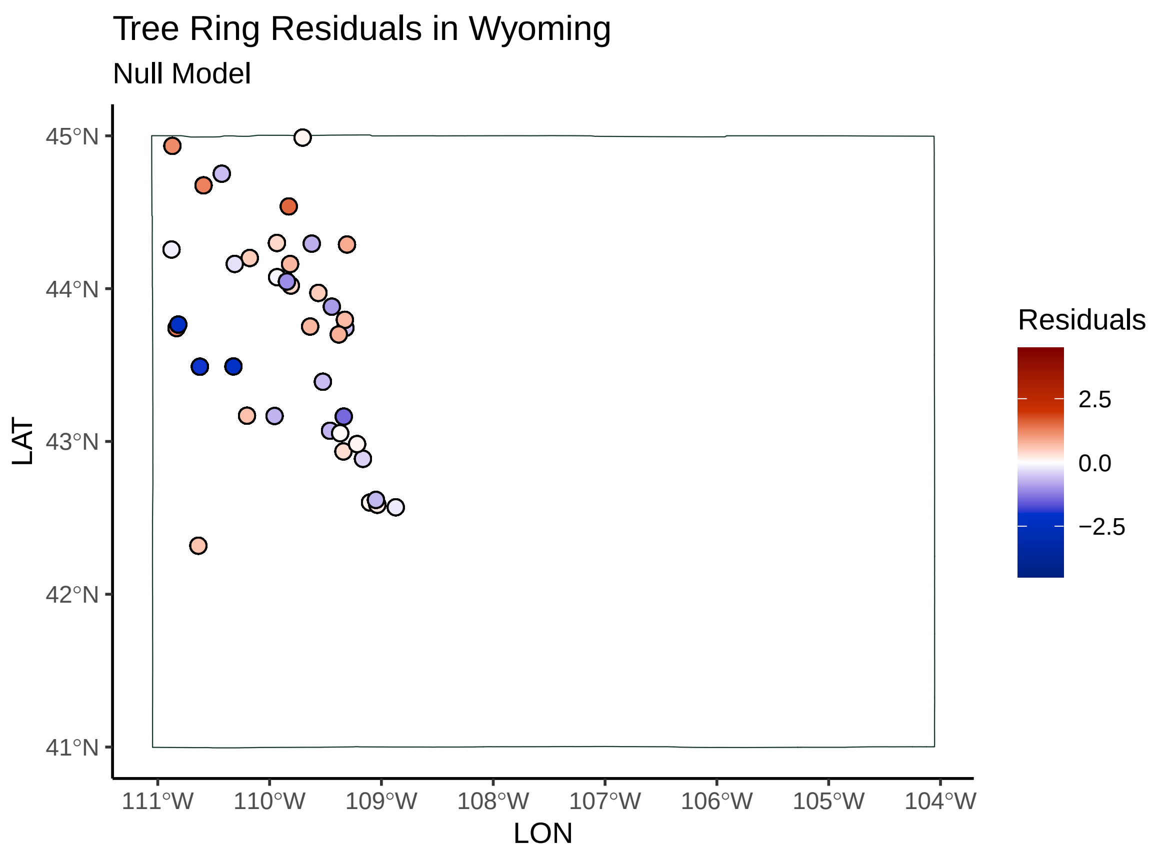
A.Null Model (No Covariates) We'll first apply a null model to annual tree-ring growth (log-transformed) to examine residual spatial patterns without the influence of covariates. This model serves as a baseline to identify any inherent spatial dependence.

```r
#okay explore model residuals for tree growth without any covariates
mod_01 <- lm(log(RW + 0.001) ~ 1, data = wbp_wy_clim)
summary(mod_01)
```

```
##
## Call:
## lm(formula = log(RW + 0.001) ~ 1, data = wbp_wy_clim)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -6.3738 -0.6024  0.1047  0.6392  4.4938
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.53393    0.01076  -49.63   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.835 on 6025 degrees of freedom
```

```r
wbp_wy_clim_sf <- st_as_sf(wbp_wy_clim, coords = c("LON", "LAT"), crs = 4326)
wbp_wy_clim$residuals <- mod_01$residuals

#make custom color palettes
custom_gradient <- scale_fill_gradientn(
  colors = c("#002080", "#0033CC", "#FFFFFF", "#CC3300", "#800000"),
  values = rescale(c(-4.5, -2, 0, 2, 4.5)),
  limits = c(-4.5, 4.5),
  name = "Residuals"
)

#map residuals
ggplot(data = wyoming_sf) +
  geom_sf(fill = "white", color = "#18392B") +
  geom_point(
    data = wbp_wy_clim,
    aes(x = LON, y = LAT, fill = residuals), # Use fill for shape 21
    shape = 21, size = 2.5
  ) +
  custom_gradient +
  labs(title = "Tree Ring Residuals in Wyoming", subtitle = "Null Model") +
  theme_classic() +
  coord_sf()
```



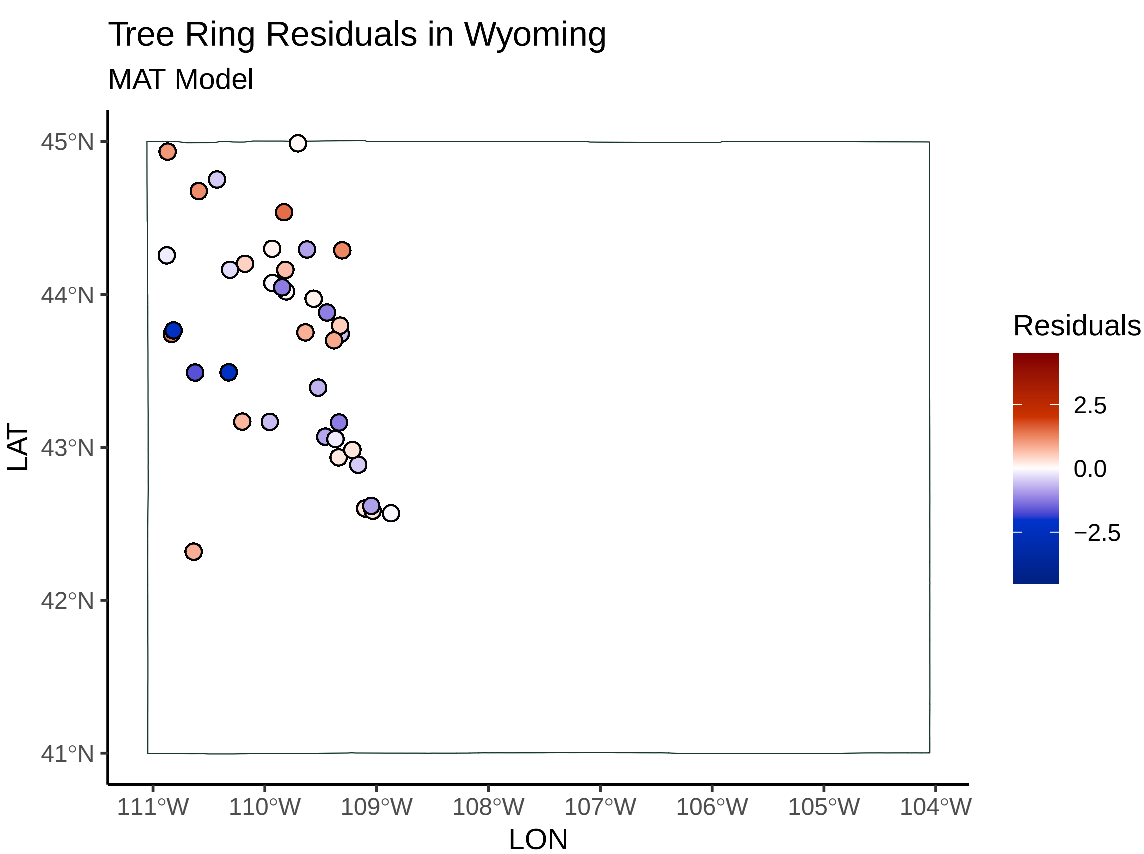
B.Model with MAT (Mean Annual Temperature) Next, MAT is introduced as a predictor variable to determine if it explains any spatial structure in tree growth. Residuals from this model allow us to see if accounting for MAT reduces clustering patterns observed in the null model residuals.

```
#okay now explore model residuals for tree growth with MAT as a covariate
mod_02 <- lm(log(RW + 0.001) ~ meantemp, data = wbp_wy_clim)
summary(mod_02)
```

```
##
## Call:
## lm(formula = log(RW + 0.001) ~ meantemp, data = wbp_wy_clim)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4280 -0.5420  0.0421  0.5758  4.6270
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.47959    0.01102  -43.51   <2e-16 ***
## meantemp    -0.18384    0.01112  -16.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8168 on 6024 degrees of freedom
## Multiple R-squared:  0.04341,    Adjusted R-squared:  0.04325
```

```
## F-statistic: 273.4 on 1 and 6024 DF,  p-value: < 2.2e-16
```

```
wbp_wy_clim_sf <- st_as_sf(wbp_wy_clim, coords = c("LON", "LAT"), crs = 4326)
wbp_wy_clim$residuals_2 <- mod_02$residuals

#make custom color palettes
custom_gradient <- scale_fill_gradientn(
  colors = c("#002080", "#0033CC", "#FFFFFF", "#CC3300", "#800000"),
  values = rescale(c(-4.5, -2, 0, 2, 4.5)),
  limits = c(-4.5, 4.5),
  name = "Residuals"
)

#map residuals
ggplot(data = wyoming_sf) +
  geom_sf(fill = "white", color = "#18392B") +
  geom_point(
    data = wbp_wy_clim,
    aes(x = LON, y = LAT, fill = residuals_2), # Use fill for shape 21
    shape = 21, size = 2.5
  ) +
  custom_gradient +
  labs(title = "Tree Ring Residuals in Wyoming", subtitle = "MAT Model") +
  theme_classic() +
  coord_sf()
```

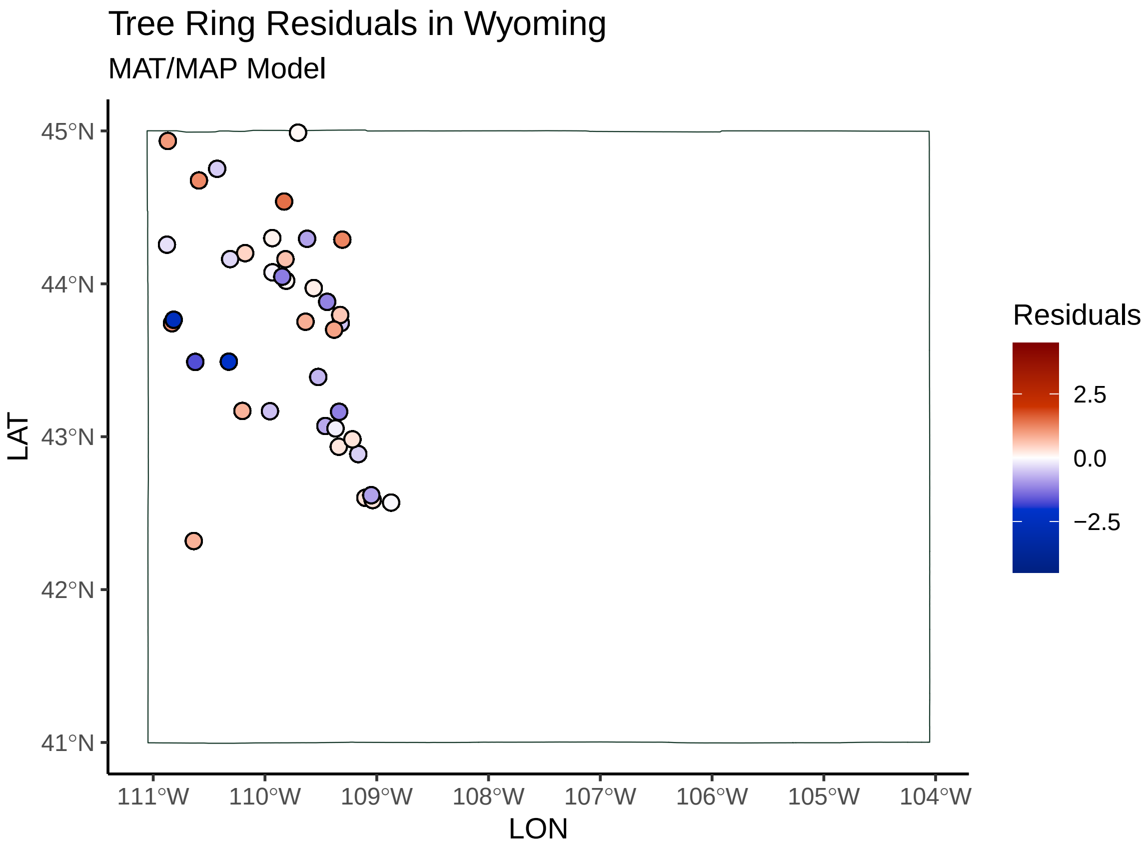


C. Model with MAT and MAP (Mean Annual Precipitation) Finally, I'm going to fit a model incorporating both MAT and MAP.

```
#okay now explore model residuals for tree growth with MAT as a covariate
mod_03 <- lm(log(RW + 0.001) ~ meantemp_scaled + precip_scaled, data = wbp_wy_clim)
summary(mod_03)
```

```
##
## Call:
## lm(formula = log(RW + 0.001) ~ meantemp_scaled + precip_scaled,
##     data = wbp_wy_clim)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6219 -0.5483  0.0443  0.5673  4.6535
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.53393    0.01050 -50.870   <2e-16 ***
## meantemp_scaled -0.15440    0.01107 -13.943   <2e-16 ***
## precip_scaled    0.06145    0.01107   5.549    3e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8148 on 6023 degrees of freedom
## Multiple R-squared:  0.04827,    Adjusted R-squared:  0.04796
## F-statistic: 152.7 on 2 and 6023 DF,  p-value: < 2.2e-16
```

```
wbp_wy_clim_sf <- st_as_sf(wbp_wy_clim, coords = c("LON", "LAT"), crs = 4326)
wbp_wy_clim$residuals_3 <- mod_03$residuals

#make custom color palettes
custom_gradient <- scale_fill_gradientn(
  colors = c("#002080", "#0033CC", "#FFFFFF", "#CC3300", "#800000"),
  values = rescale(c(-4.5, -2, 0, 2, 4.5)),
  limits = c(-4.5, 4.5),
  name = "Residuals"
)

#map residuals
ggplot(data = wyoming_sf) +
  geom_sf(fill = "white", color = "#18392B") +
  geom_point(
    data = wbp_wy_clim,
    aes(x = LON, y = LAT, fill = residuals_3), # Use fill for shape 21
    shape = 21, size = 2.5
  ) +
  custom_gradient +
  labs(title = "Tree Ring Residuals in Wyoming", subtitle = "MAT/MAP Model") +
  theme_classic() +
  coord_sf()
```



IV. Dataset Description (combined)

A (very) preliminary analysis of tree ring data for Whitebark Pine in Wyoming reveals possible clustering patterns and spatial dependence related to climate variables such as mean annual temperature (MAT) and mean annual precipitation (MAP). We mapped MAT and MAP across plot locations, where it appears that there may be spatial dependence between climate variables and annual tree growth across the series of a tree. These findings suggest that more thorough exploration of climate variables needs to be explored. Initial plots of MAT and MAP show that climate normals exhibit spatial variability across Wyoming, so I thought that models with these variables would have higher R^2 values.

V. Challenges

Generally, I would say that it is hard to tell if there is spatial dependence in the residuals in any of the plots, I might need to figure out a way to include more data points to actually determine if there is a pattern. I evaluated the residuals from a model with no covariates, a model with MAT as a covariate, and a model with both MAT and MAP as covariates to see if they have any impact on removing the clustering of the residuals (especially of the dark blue residuals). Something that I have not included in these models, which model raw annual ring width, is the size effect, which is well established in the dendrochrnology literature to play a significant role driving annual raw ring width increment. A common practice to get rid of the size effect is to "detrend" the tree-ring data in order to get rid of this size effect.