Section 6: Quantifying the Impact of Weather on Crop Yields (Continued)

Shuo Yu

2024-02-26

```
rm(list = ls())

library(tidyverse)
library(fixest)

setwd("C:/Users/shuoy/Dropbox/161/Sections/Section6")
```

Data

```
full_df <- read.csv("YieldWeather.csv", header = TRUE)
# Convert GEOID to 5-digit character format
full_df <- full_df %>%
  mutate(GEOID = sprintf("%05d", GEOID))

wp245 <- read.csv("WeatherProjection245.csv", header = TRUE)
wp245 <- wp245 %>%
  mutate(GEOID = sprintf("%05d", GEOID)) %>%
  mutate(Year = Year-1980)
wp585 <- read.csv("WeatherProjection585.csv", header = TRUE)
wp585 <- wp585 %>%
  mutate(GEOID = sprintf("%05d", GEOID)) %>%
  mutate(Year = Year-1980)
```

Let's use corn yields as an example. You can follow similar steps and adjust the code to obtain results for soybeans.

```
# Drops rows with missing values (na.omit())
corn_df <- full_df %>%
  select(-SoyYield) %>%
  na.omit()

# Filters counties that have at least 21 observations to ensure major corn growing counties w
corn_df <- corn_df %>%
  group_by(GEOID) %>%
  filter(n() > 20) %>%
  ungroup()
```

Panel Regression Analysis by Region

To assess the impact of weather on corn yield, we conduct a panel regression separately for counties east and west of the 100th meridian line, a historically significant climatic divide in the United States. The regression model is specified as follows:

$$CornYield_{it} = \beta_0 + \beta_1 GDD_{it} + \beta_2 EDD_{it} + \beta_3 PPT_{it} + \beta_4 PPT_{it}^2 + \alpha_i + \delta_t + \gamma_i Year_{it} + \varepsilon_{it}$$

- Outcome variable: $CornYield_{it}$ – Corn yield in county i and year t
- Independent variables:
  - GDD_{it} – Growing degree days during the growing season (April to September)
  - EDD_{it} – Extreme degree days during the growing season
  - PPT_{it} – Cumulative precipitation during the growing season
- **County Fixed Effects** (α_i): Removes the impact of factors that differ across counties but stay the same over time (e.g., soil quality, local policies).
- **Year Fixed Effects** (δ_t): Controls for yearly shocks affecting all counties (e.g., nationwide economic conditions, federal policies).
- **County-Specific Time Trends** (γ_i): Allows each county to have its own trend over time (e.g., gradual technology adoption in some regions).
- ε_{it} represents the error term, capturing unobserved factors affecting corn yield.

Spatial Heterogeneity

```r
# Summary statistics based on whether the county is located east or west of the 100th meridia
# East of the 100th meridian line
corn_df %>%
  filter(east_dummy==1) %>%
  summary()
```

```
      Year         CornYield        GEOID               NAME          
 Min.   : 1.00   Min.   :  0.0   Length:69146       Length:69146      
 1st Qu.:10.00   1st Qu.: 88.9   Class :character   Class :character  
 Median :20.00   Median :115.0   Mode  :character   Mode  :character  
 Mean   :20.75   Mean   :117.5                                        
 3rd Qu.:31.00   3rd Qu.:146.0                                        
 Max.   :43.00   Max.   :246.7                                        
      edd               gdd              ppt               tavg      
 Min.   :   0.00   Min.   : 926.4   Min.   :  75.63   Min.   :11.91  
 1st Qu.:  44.17   1st Qu.:1826.6   1st Qu.: 489.43   1st Qu.:17.72  
 Median : 110.80   Median :2184.1   Median : 588.36   Median :19.88  
 Mean   : 156.37   Mean   :2216.3   Mean   : 603.41   Mean   :19.99  
 3rd Qu.: 227.83   3rd Qu.:2562.8   3rd Qu.: 701.54   3rd Qu.:22.06  
 Max.   :1256.27   Max.   :3781.0   Max.   :1705.49   Max.   :29.42  
      tmax         east_dummy Ag.District       
 Min.   :17.77   Min.   :1    Length:69146      
 1st Qu.:23.86   1st Qu.:1    Class :character  
 Median :26.05   Median :1    Mode  :character  
 Mean   :26.13   Mean   :1                      
 3rd Qu.:28.24   3rd Qu.:1                      
 Max.   :36.78   Max.   :1                      
```

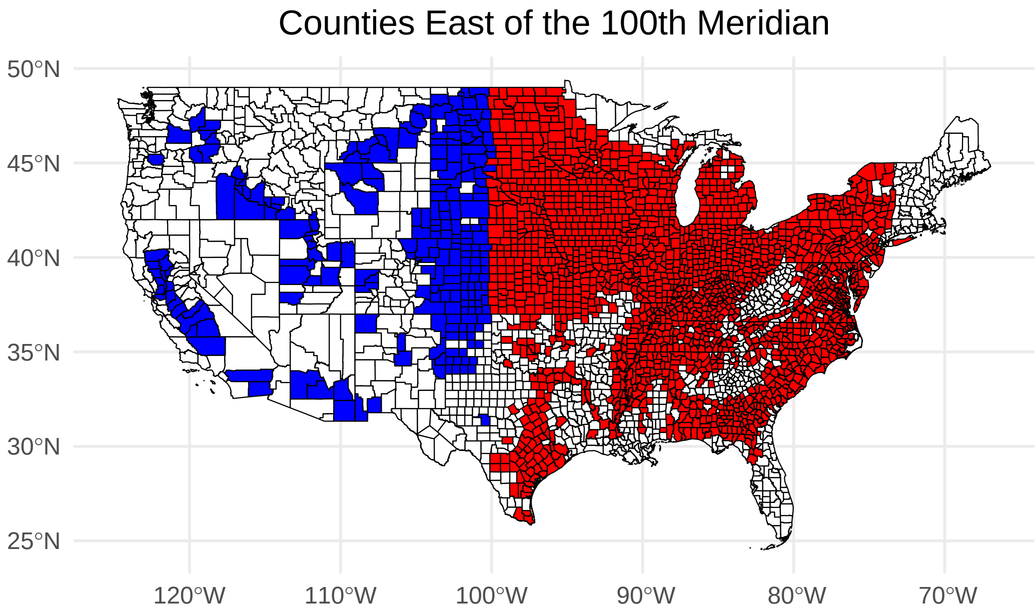
```r
# West of the 100th meridian line
corn_df %>%
  filter(east_dummy==0) %>%
  summary()
```

```
      Year         CornYield        GEOID               NAME          
 Min.   : 1.00   Min.   :  0.0   Length:7556        Length:7556       
 1st Qu.: 9.00   1st Qu.:101.6   Class :character   Class :character  
 Median :18.00   Median :133.2   Mode  :character   Mode  :character  
 Mean   :18.85   Mean   :129.2                                        
 3rd Qu.:27.00   3rd Qu.:159.9                                        
 Max.   :43.00   Max.   :277.1                                        
```

```
      edd              gdd             ppt             tavg
 Min.   :   6.473   Min.   : 974.5   Min.   :  0.8864   Min.   :12.71
 1st Qu.: 138.062   1st Qu.:1623.9   1st Qu.:173.7575   1st Qu.:16.45
 Median : 228.916   Median :1875.9   Median :287.8665   Median :18.01
 Mean   : 260.393   Mean   :1944.5   Mean   :277.7861   Mean   :18.38
 3rd Qu.: 341.720   3rd Qu.:2258.2   3rd Qu.:375.3860   3rd Qu.:20.28
 Max.   :1558.209   Max.   :3669.0   Max.   :743.3722   Max.   :30.17
      tmax         east_dummy Ag.District
 Min.   :19.33   Min.   :0    Length:7556
 1st Qu.:24.37   1st Qu.:0    Class :character
 Median :26.27   Median :0    Mode  :character
 Mean   :26.42   Mean   :0
 3rd Qu.:28.42   3rd Qu.:0
 Max.   :38.40   Max.   :0
```

Map of Counties Included in the Regression East of the 100th Meridian

```
  |
  |                                                                      |   0%
  |
  |                                                                      |   1%
  |
  |=                                                                     |   1%
  |
  |=                                                                     |   2%
  |
  |==                                                                    |   2%
  |
  |==                                                                    |   3%
  |
  |===                                                                   |   4%
  |
  |====                                                                  |   6%
  |
  |=====                                                                 |   7%
  |
  |=====                                                                 |   8%
  |
  |======                                                                |   8%
  |
  |======                                                                |   9%
```



East of 100th Meridian Line

```r
fe_east <- feols(CornYield ~ gdd + edd + ppt + I(ppt^2) |
                  Year + GEOID[Year],
                  data = corn_df[corn_df$east_dummy==1, ])

summary(fe_east)
```

```
OLS estimation, Dep. Var.: CornYield
Observations: 69,146
Fixed-effects: Year: 43,  GEOID: 1,798
Varying slopes: Year (GEOID): 1,798
Standard-errors: Clustered (Year)
           Estimate Std. Error  t value   Pr(>|t|)
gdd        0.021023   0.011999  1.75214 8.7049e-02 .
edd       -0.151737   0.016663 -9.10636 1.6998e-11 ***
ppt        0.076309   0.015895  4.80073 2.0277e-05 ***
I(ppt^2) -0.000058   0.000012 -4.97231 1.1647e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 16.8     Adj. R2: 0.817636
```

```
            Within R2: 0.174532
```

West of 100th Meridian Line

```r
fe_west <- feols(CornYield ~ gdd + edd + ppt + I(ppt^2) |
                   Year + GEOID[Year],
                   data = corn_df[corn_df$east_dummy==0, ])

summary(fe_west)
```

```
OLS estimation, Dep. Var.: CornYield
Observations: 7,556
Fixed-effects: Year: 43,  GEOID: 225
Varying slopes: Year (GEOID): 225
Standard-errors: Clustered (Year)
            Estimate Std. Error   t value   Pr(>|t|)
gdd       0.02930561   0.010085  2.905742 5.8240e-03 **
edd      -0.09001536   0.013799 -6.523438 7.0119e-08 ***
ppt       0.01992034   0.018889  1.054577 2.9765e-01
I(ppt^2)  0.00000989   0.000028  0.354048 7.2507e-01
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 17.8     Adj. R2: 0.830369
            Within R2: 0.061924
```

Predictions Based on Weather Projections

Climate Model and Future Weather Projections

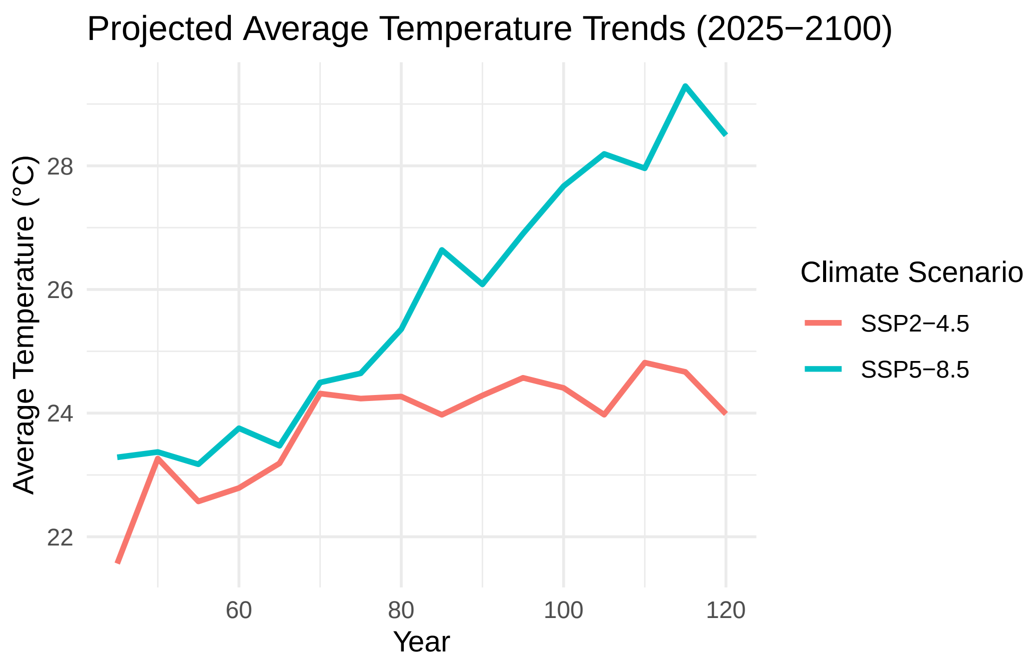
To project the impact of climate change on corn yields, we utilize future climate scenarios from the CanESM5 model (data source: Google Earth Engine), a widely recognized Earth system model that simulates future temperature and precipitation patterns.

We consider two Shared Socioeconomic Pathway (SSP) scenarios:

- SSP5-8.5 (High Emissions Scenario): Represents a future characterized by continued fossil fuel dependency, leading to substantial warming and increased climate extremes.
- SSP2-4.5 (Intermediate Emissions Scenario): Assumes moderate emissions reduction policies, resulting in more gradual warming trends.

For each county, we extract climate projections for the period 2025–2100, incorporating temperature and precipitation variables, and apply the estimated regression coefficients to predict future yield changes. First, let's visualize the trends in weather variables over time.

```r
# Visualizing the time trend of average temperature under different scenarios
# Compute yearly average temperature for each dataset
avg_temp_585 <- wp585 %>%
  group_by(Year) %>%
  summarise(Avg_Temperature = mean(tavg, na.rm = TRUE))

avg_temp_245 <- wp245 %>%
  group_by(Year) %>%
  summarise(Avg_Temperature = mean(tavg, na.rm = TRUE))

# Combine the datasets for plotting
avg_temp_585$Scenario <- "SSP5-8.5"
avg_temp_245$Scenario <- "SSP2-4.5"

combined_data <- bind_rows(avg_temp_585, avg_temp_245)

# Plot the time trend
ggplot(combined_data, aes(x = Year, y = Avg_Temperature, color = Scenario)) +
  geom_line(linewidth = 1) +
  labs(title = "Projected Average Temperature Trends (2025-2100)",
       x = "Year",
       y = "Average Temperature (°C)",
       color = "Climate Scenario") +
  theme_minimal()
```



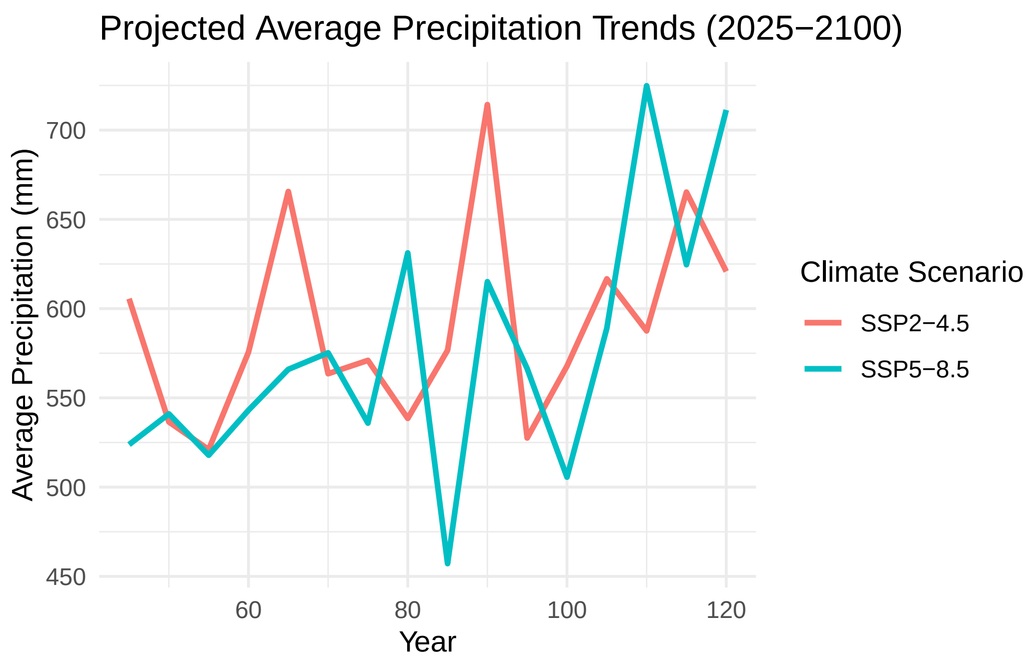
Question: Describe the trends you observe in the figure. What patterns do you notice, and how do they change over time?

The figure shows the projected average temperature trends from 2025 to 2100 under two climate scenarios: SSP2-4.5 (moderate emissions) and SSP5-8.5 (high emissions). The y-axis represents average temperature (°C), while the x-axis represents years. The SSP5-8.5 scenario (blue line) shows a sharper increase in temperature over time compared to the SSP2-4.5 scenario (red line), which exhibits a more gradual rise. This suggests that higher emissions lead to greater temperature increases by the end of the century.

```
# Compute yearly average precipitation for each dataset
avg_ppt_585 <- wp585 %>%
  group_by(Year) %>%
  summarise(Avg_Precipitation = mean(ppt, na.rm = TRUE))

avg_ppt_245 <- wp245 %>%
  group_by(Year) %>%
  summarise(Avg_Precipitation = mean(ppt, na.rm = TRUE))

# Combine the datasets for plotting
avg_ppt_585$Scenario <- "SSP5-8.5"
avg_ppt_245$Scenario <- "SSP2-4.5"

combined_data <- bind_rows(avg_ppt_585, avg_ppt_245)
```

```r
# Plot the time trend for precipitation
ggplot(combined_data, aes(x = Year, y = Avg_Precipitation, color = Scenario)) +
  geom_line(size = 1) +
  labs(title = "Projected Average Precipitation Trends (2025-2100)",
       x = "Year",
       y = "Average Precipitation (mm)",
       color = "Climate Scenario") +
  theme_minimal()
```

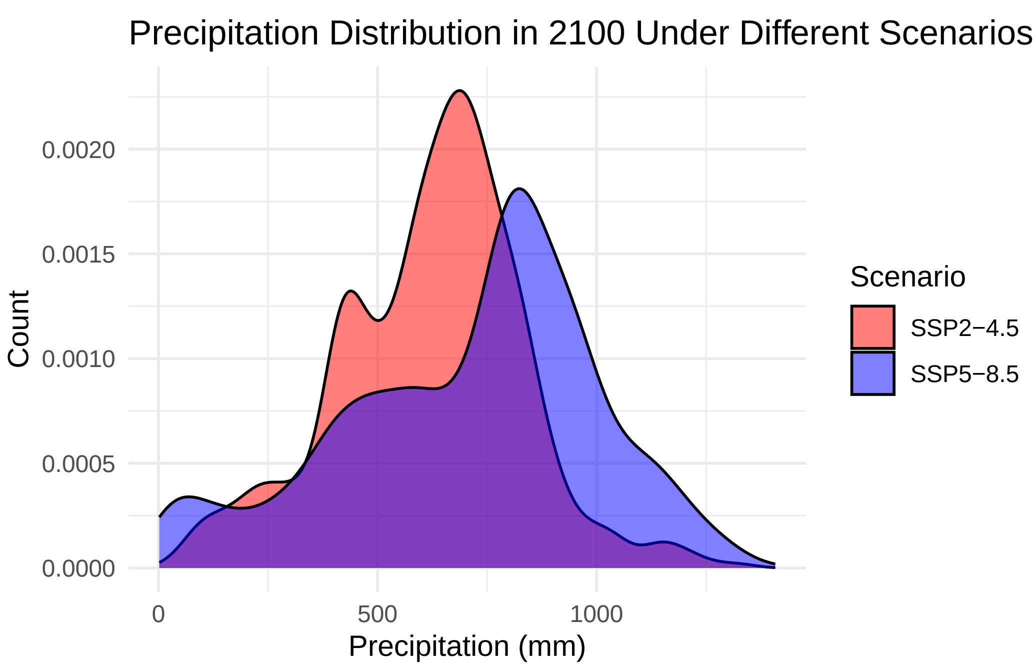
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```



Question: Describe the trends you observe in the figure. What patterns do you notice, and how do they change over time?

The figure shows the projected average precipitation trends from 2025 to 2100 under two climate scenarios. The y-axis represents average precipitation (mm), while the x-axis represents years. Both scenarios exhibit high interannual variability, with no clear long-term trend. While precipitation levels fluctuate significantly under both scenarios, SSP5-8.5 (blue line) appears to have slightly more extreme variations compared to SSP2-4.5 (red line).

```r
# Filter data for the year 2100
ppt_585_2100 <- wp585 %>%
  filter(Year == 120) %>%
  mutate(Scenario = "SSP5-8.5")

ppt_245_2100 <- wp245 %>%
  filter(Year == 120) %>%
  mutate(Scenario = "SSP2-4.5")

# Combine datasets for plotting
combined_ppt_2100 <- bind_rows(ppt_585_2100, ppt_245_2100)

# Plot histograms
ggplot(combined_ppt_2100, aes(x = ppt, fill = Scenario)) +
  geom_density(alpha = 0.5, color = "black") +
  scale_fill_manual(values = c("SSP5-8.5" = "blue", "SSP2-4.5" = "red")) +
  labs(title = "Precipitation Distribution in 2100 Under Different Scenarios",
       x = "Precipitation (mm)",
       y = "Count",
       fill = "Scenario") +
  theme_minimal()
```

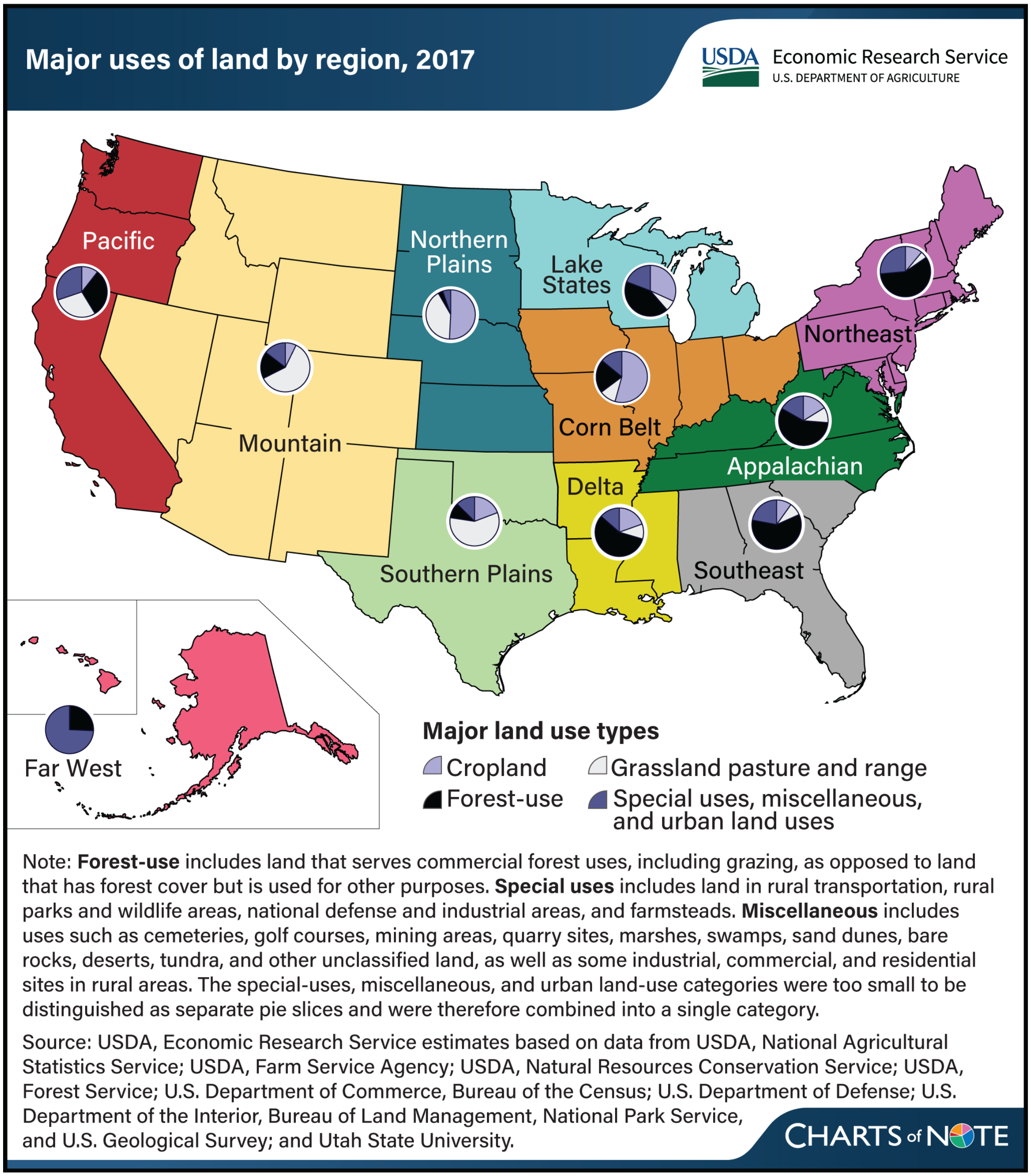


Question: What trends do you observe in the distributions? Compare the peaks, spread, and overall shape of the two scenarios. What does this suggest about the impact of different climate scenarios on precipitation patterns?

The figure displays the precipitation distribution in 2100 under two climate scenarios. The x-axis represents precipitation (mm), while the y-axis shows the density (normalized count). SSP2-4.5 exhibits a more concentrated distribution with a peak at lower precipitation values, whereas SSP5-8.5 shows a broader distribution with higher precipitation extremes, suggesting greater variability and more frequent extreme precipitation events under the high-emissions scenario.

USDA Farm Production Regions

The USDA's Farm Production Regions are traditional groupings of U.S. states, organized based on predominant farming activities, geographic features, and climatic conditions. These regions facilitate the analysis and reporting of agricultural data by grouping areas with similar agricultural practices.



For example, the **Northeast region** includes states such as Maine, New York, and Pennsylvania, known for diverse agricultural activities, including dairy farming, fruit production, and horticulture. The **Lake States**, comprising Michigan, Wisconsin, and Minnesota, are notable for dairy farming as well as corn and soybean production. The **Corn Belt**, which includes Iowa, Illinois, Indiana, Missouri, and Ohio, is the heart of U.S. grain production, primarily cultivating corn and soybeans. The **Northern Plains**, covering North Dakota, South Dakota, Nebraska, and Kansas, specializes in wheat, corn, and soybean farming, along with significant cattle ranching. The **Southern Plains**, consisting of Oklahoma and Texas, supports a diverse agricultural economy with wheat, cotton, and cattle production. The **Pacific region**, which includes Washington, Oregon, and California, features a highly varied agricultural landscape, producing fruits, vegetables, nuts, and wine.

Predictions

Extract Model Coefficients

```r
coeff_west <- coef(fe_west)
coeff_east <- coef(fe_east)

print(coeff_west)  # Check if coefficients exist
```

```
          gdd           edd           ppt      I(ppt^2)
 2.930561e-02 -9.001536e-02  1.992034e-02  9.885540e-06
```

```r
print(coeff_east)
```

```
          gdd           edd           ppt      I(ppt^2)
 2.102308e-02 -1.517374e-01  7.630875e-02 -5.779029e-05
```

```r
# Extract fixed effects from fe_east
fixef_east <- fixef(fe_east)
# Extract county fixed effects (GEOID), year fixed effects (Year), and county-year fixed effe
geo_fe_east <- fixef_east$GEOID
year_fe_east <- fixef_east$Year
geo_year_fe_east <- fixef_east$`GEOID[[Year]]`  # Extract full county-year fixed effects

# Extract fixed effects from fe_west
fixef_west <- fixef(fe_west)
geo_fe_west <- fixef_west$GEOID
year_fe_west <- fixef_west$Year
geo_year_fe_west <- fixef_west$`GEOID[[Year]]`
```

Compute Predictions

1. **Computing an Approximate Intercept**: The approximate intercept accounts for fixed effects (county-level, year-level, and county-year interactions) to provide a baseline for prediction.

```r
# Compute an Approximate Intercept
# Compute the mean of all fixed effects (county, year, and county-year)
approx_intercept_east <- mean(unlist(geo_fe_east), na.rm = TRUE) +
                         mean(unlist(year_fe_east), na.rm = TRUE) +
                         mean(unlist(geo_year_fe_east), na.rm = TRUE)*43
approx_intercept_west <- mean(unlist(geo_fe_west), na.rm = TRUE) +
                         mean(unlist(year_fe_west), na.rm = TRUE) +
                         mean(unlist(geo_year_fe_west), na.rm = TRUE)*43
```

2. **Computing Predictions for SSP2-4.5**

```r
# Compute predictions for west counties
wp245 <- wp245 %>%
  mutate(pred_yield_west = approx_intercept_west +
           coeff_west["gdd"] * gdd +
           coeff_west["edd"] * edd +
           coeff_west["ppt"] * ppt +
           coeff_west["I(ppt^2)"] * ppt^2)

# Compute predictions for east counties
wp245 <- wp245 %>%
  mutate(pred_yield_east = approx_intercept_east +
           coeff_east["gdd"] * gdd +
           coeff_east["edd"] * edd +
           coeff_east["ppt"] * ppt +
           coeff_east["I(ppt^2)"] * ppt^2)

# Assigning the final predicted yield
wp245 <- wp245 %>%
  mutate(pred_yield = ifelse(east_dummy == 0, pred_yield_west, pred_yield_east))
```

3. **Computing Predictions for SSP5-8.5**

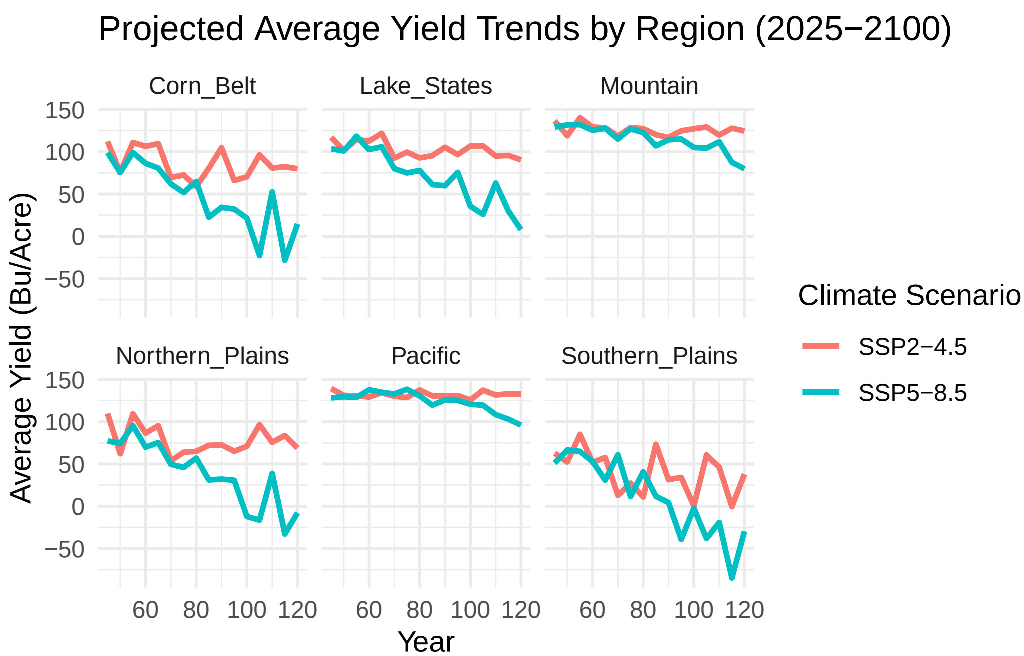
```r
# Compute predictions for west counties
wp585 <- wp585 %>%
  mutate(pred_yield_west = approx_intercept_west +
           coeff_west["gdd"] * gdd +
```

```
           coeff_west["edd"] * edd +
           coeff_west["ppt"] * ppt +
           coeff_west["I(ppt^2)"] * ppt^2)

# Compute predictions for east counties
wp585 <- wp585 %>%
  mutate(pred_yield_east = approx_intercept_east +
           coeff_east["gdd"] * gdd +
           coeff_east["edd"] * edd +
           coeff_east["ppt"] * ppt +
           coeff_east["I(ppt^2)"] * ppt^2)

# Assigning the final predicted yield
wp585 <- wp585 %>%
  mutate(pred_yield = ifelse(east_dummy == 0, pred_yield_west, pred_yield_east))
```

Predicted Trends for Different Regions

```
# List of regions to analyze
regions <- c("Lake_States", "Pacific", "Mountain", "Southern_Plains", "Corn_Belt", "Northern_

# Compute average yield trends for each scenario and region
avg_yield_585 <- wp585 %>%
  filter(district %in% regions) %>%
  group_by(Year, district) %>%
  summarise(Avg_Yield = mean(pred_yield, na.rm = TRUE), .groups = "drop") %>%
  mutate(Scenario = "SSP5-8.5")

avg_yield_245 <- wp245 %>%
  filter(district %in% regions) %>%
  group_by(Year, district) %>%
  summarise(Avg_Yield = mean(pred_yield, na.rm = TRUE), .groups = "drop") %>%
  mutate(Scenario = "SSP2-4.5")

# Combine both datasets
combined_data <- bind_rows(avg_yield_585, avg_yield_245)

# Find global min and max for y-axis
y_min <- min(combined_data$Avg_Yield, na.rm = TRUE)
y_max <- max(combined_data$Avg_Yield, na.rm = TRUE)
```

```r
# Plot with fixed y-axis
ggplot(combined_data, aes(x = Year, y = Avg_Yield, color = Scenario)) +
  geom_line(linewidth = 1) +
  facet_wrap(~ district, scales = "fixed") +  # Fix y-axis across regions
  ylim(y_min, y_max) +  # Set y-axis limits manually
  labs(title = "Projected Average Yield Trends by Region (2025-2100)",
       x = "Year",
       y = "Average Yield (Bu/Acre)",
       color = "Climate Scenario") +
  theme_minimal()
```



Questions:

1. Examine the projected average yield trends for different regions under the two climate scenarios (SSP2-4.5 and SSP5-8.5) from 2025 to 2100. What patterns do you observe in yield trends across regions? How do the trends differ between the two scenarios?

Yields decrease across all regions, with steeper declines under SSP5-8.5 (blue) compared to SSP2-4.5 (red). SSP2-4.5: More moderate declines, suggesting some adaptation potential. SSP5-8.5: Stronger yield losses, particularly in heat-vulnerable southern regions.

2. Which regions appear to be most negatively impacted under SSP5-8.5, and what might be driving these differences?

- Corn Belt & Northern Plains: Sharp declines, especially under SSP5-8.5.
- Lake States: Gradual decline; divergence between scenarios increases over time.
- Mountain & Pacific: More stable trends, less sensitivity.
- Southern Plains: Highly volatile with sharp drops in both scenarios.

Western regions are less sensitive due to irrigation buffering climate extremes, while southern areas face greater heat damage, leading to sharper declines. Climate mitigation efforts (SSP2-4.5) help reduce severe losses. This assumes no significant land-use changes. In reality, farmers may adapt, leading to more crop production in the north and less in the south, potentially mitigating losses over time.

References

Schlenker, Wolfram, and Michael J. Roberts. "Nonlinear effects of weather on corn yields." *Review of agricultural economics* 28, no. 3 (2006): 391-398. link

Schlenker, Wolfram, and Michael J. Roberts. "Nonlinear temperature effects indicate severe damages to US crop yields under climate change." *Proceedings of the National Academy of sciences* 106, no. 37 (2009): 15594-15598. link