Final Project

Ben Kelley

2025-04-26

The Problem

https://coloradosun.com/2023/11/19/colorado-river-lake-mead-lake-powell-dry/

https://www.naco.org/news/lower-basin-states-strike-agreement-preserve-water-supply-colorado-river-basin

In 2023, the federal government intervened in a dispute between Nevada, Arizona, and California over usage of the Colorado River. The Colorado is a crucial source of water for all three states and over the past twenty years, the river's water flow has decreased dramatically. Consequently, federal and state reservoirs in all three states are running dangerously low. In the early 2020s, the situation became dire enough to require a dramatic remedy in the form of reduced usage of water from the river. The three states, whose adjoining regions all depend on the river, were unable to reach an agreement on the distribution of burdens resulting from a reduction in water usage. Ultimately, federal intervention forced an agreement which preserved the water reservoirs.

Many studies today demonstrate that climate change is likely to make water scarcity a more common phenomenon in the future. That means that situations like what we saw in 2023 are likely to appear again. However, it may be possible to avoid such drastic measures if we anticipate the water shortage and work earlier to prevent it.

This project is aimed at informing legislators at state levels or regulators at state or federal levels about where potential water shortages may occur. I will take data on rainfall in every county in the continental United States from the National Oceanic and Atmospheric Association (NOAA) and try to assess whether there are any regions experiencing trends similar to this region at the intersection of Nevada, Arizona, and California to identify "red flags" in existing climate data. This will help regulators to make smart decisions now in order to avoid future burdens.

Data Loading

Precipitation: https://www.ncei.noaa.gov/pub/data/cirs/climdiv/climdiv-norm-pcpncy-v1.0.0-20250404

CLIMDIV: https://www.ncei.noaa.gov/pub/data/cirs/climdiv/county-to-climdivs.txt

Region names: https://www.ncei.noaa.gov/access/monitoring/reference-maps/conus-climate-divisions

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.3.3
```

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.3.3
```

```r
# Reading Function

weather_read <- function(new_url, data_name){

url <- new_url
temp_file <- tempfile()
download.file(url, temp_file, mode = "wb")

widths <- c(11, rep(7, 12))

raw_data <- read_fwf(
  temp_file,
  fwf_widths(widths),
  col_types = cols(.default = "c")
)

colnames(raw_data) <- c("ID", month.abb)

processed_data <- raw_data %>%
  mutate(
    county_code     = substr(ID, 1, 5),
    data_type = substr(ID, 6, 7),
    year      = substr(ID, 8, 11)
  ) %>%
  select(-ID) %>%
  mutate(across(all_of(month.abb), ~ as.numeric(.x)))

long_data <- processed_data %>%
  pivot_longer(cols = all_of(month.abb), names_to = "month", values_to = data_name)

return(long_data)
```

```
}

# Precipitation

precip <- weather_read(new_url = "https://www.ncei.noaa.gov/pub/data/cirs/climdiv/climdiv-pcpncy-v1.0.0

climate <- precip

climate <- climate[, -2]

# Climate Divisions
url <- "https://www.ncei.noaa.gov/pub/data/cirs/climdiv/county-to-climdivs.txt"

# Read file, header included
county_to_climdiv <- read_table(url, skip = 2)
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   POSTAL_FIPS_ID = col_character(),
##   NCDC_FIPS_ID = col_character(),
##   CLIMDIV_ID = col_character()
## )

## Warning: 1 parsing failure.
##  row col  expected    actual                                                                        fil
## 3108  -- 3 columns 1 columns 'https://www.ncei.noaa.gov/pub/data/cirs/climdiv/county-to-climdivs.txt
```

```
county_to_climdiv <- county_to_climdiv[complete.cases(county_to_climdiv), ]

county_to_climdiv <- county_to_climdiv[, c(2, 3)]

colnames(county_to_climdiv)[1] <- "county_code"

climate <- climate |> left_join(county_to_climdiv, by = "county_code")

colnames(climate)[5] <- "climdiv"

climate <- climate %>%
  group_by(climdiv, year, month) %>%
  summarize(mean_precip = mean(precip_in, na.rm = TRUE), .groups = "drop")

month_order <- month.abb

# Sort data by year and month
climate <- climate %>%
  mutate(month = factor(month, levels = month_order)) %>%  # Ensure month order is correct
  arrange(climdiv, year, month)  # Sort by division, year, and month

regions <- read.csv("~/Machine Learning/CPP Regions.csv")

climate$climdiv <- as.integer(climate$climdiv)
```

```r
colnames(regions)[1] <- "climdiv"

climate <- climate |> left_join(regions, by = "climdiv")

climate <- climate |> mutate(Region = paste(State, Name, sep = ", "))
```

Data Exploration

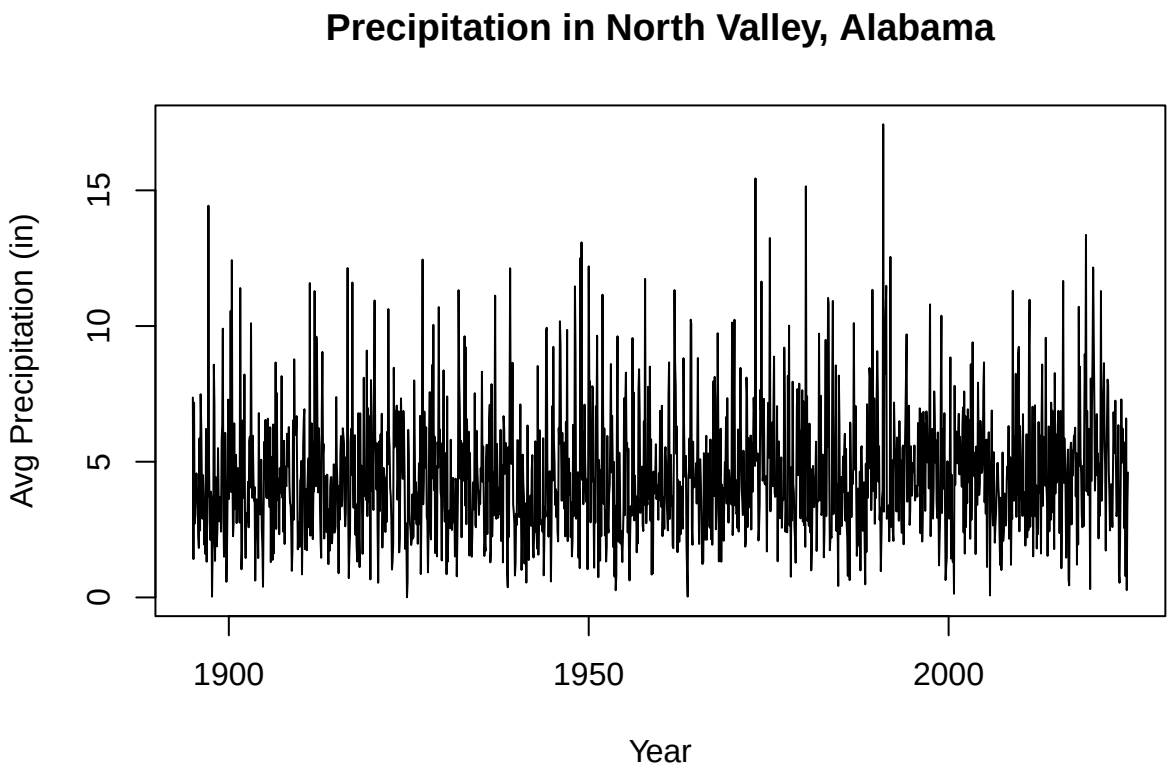
```r
library(zoo)
```

```
## Warning: package 'zoo' was built under R version 4.3.3
```

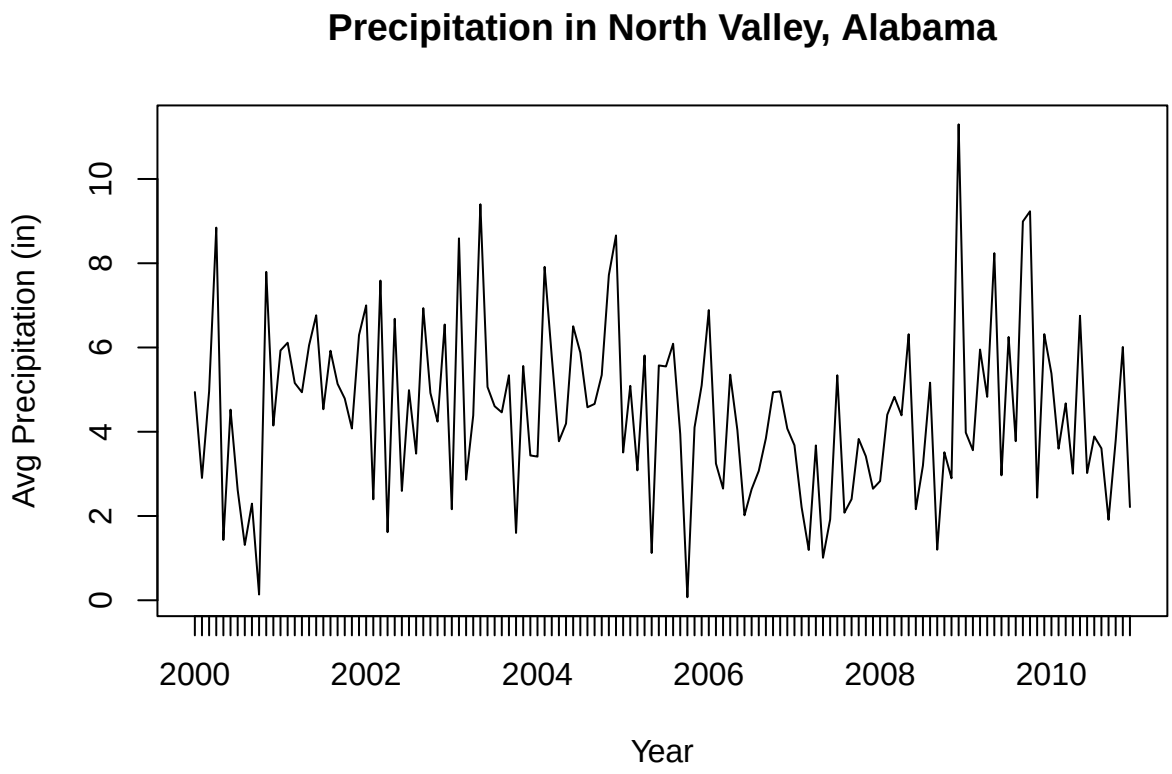
```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

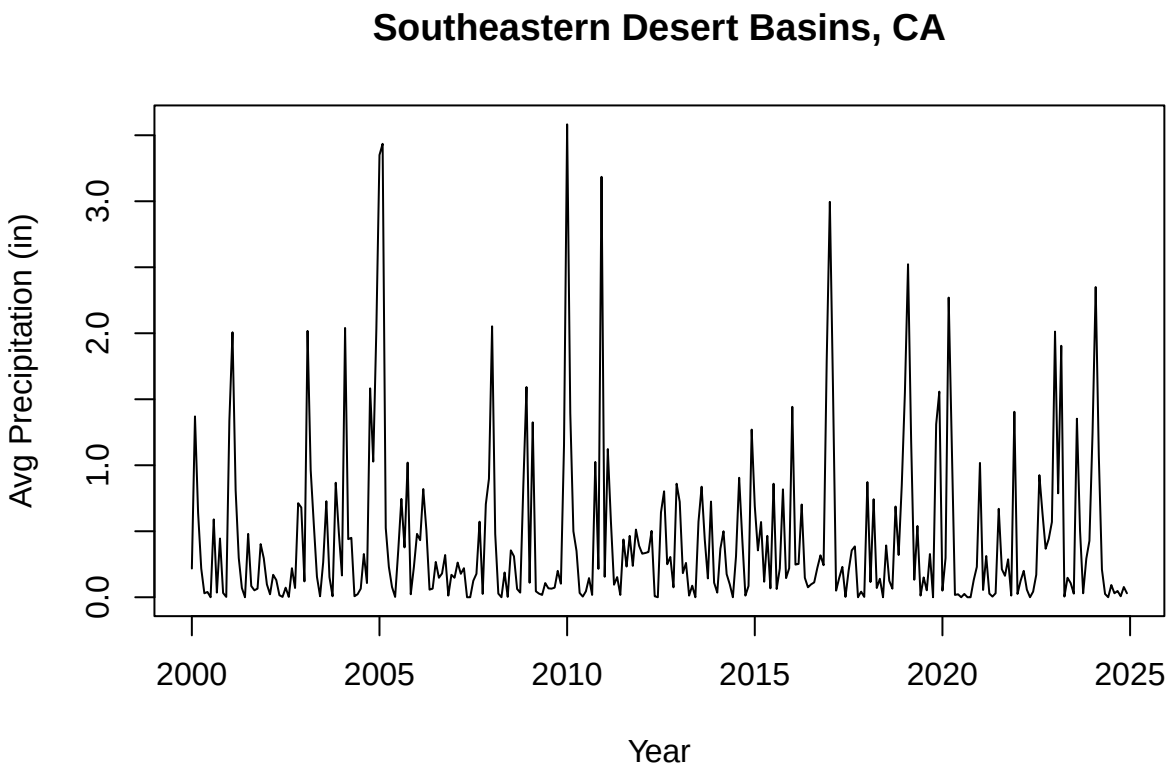
```r
climate <- climate %>%
  mutate(
    year_month = as.yearmon(paste(year, month), format = "%Y %b")
  )

climate <- climate[!climate$year == 2025, ]

div101 <- climate |> filter(climdiv == 101)

plot(div101$year_month, div101$mean_precip,
     xlab = "Year", ylab = "Avg Precipitation (in)", main = "Precipitation in North Valley, Alabama", ty
```




```
div101_recent <- climate |> filter(climdiv == 101, year >= 2000, year <= 2010)

plot(div101_recent$year_month, div101_recent$mean_precip,
     xlab = "Year", ylab = "Avg Precipitation (in)", main = "Precipitation in North Valley, Alabama", ty
```




```
div407_recent <- climate |> filter(climdiv == 407, year >= 2000)

plot(div407_recent$year_month, div407_recent$mean_precip, xlab = "Year", ylab = "Avg Precipitation (in)"
```



These graphs show the nature of the data. The structure of the data is a very long time series with about 130 * 12 entries for every climate region, as defined by the NOAA. The amount of noise and data will require me to limit the analysis to the most recent 25 years. That timeframe is ideal regardless since the issues in the Southern portion of the Colorado River began in earnest around 2005 and using data from a century beforehand would likely obscure the recent trends that are of concern here.

Modeling

I am here interested in analyzing trends in the data and seeing which regions are experiencing similar changes in precipitation, rather than seeing which regions are climatically similar by precipitation. Therefore, I will scale the data across regions in order to make trends the point of analysis. Then, I will pursue hierarchical clustering. Hierarchical clustering is a form of unsupervised learning that allows the computer to make mathematical judgments about what number of clusters is appropriate, since I do not have intuition about this.

```
precip <- climate[, c(1, 4:8)]

dfmod <- precip %>%
  pivot_wider(
    names_from = year_month,
    values_from = mean_precip
  )

dfmod <- dfmod[-338, ]
```

```r
library(dplyr)
library(tidyr)

df_long <- dfmod %>%
  pivot_longer(
    cols = -c(Region, climdiv, Name, State),
    names_to = "year_month",
    values_to = "precip"
  ) %>%
  mutate(
    year = sub(".* ", "", year_month)
  ) %>%
  group_by(Region, year) %>%
  summarise(
    annual_precip = sum(precip, na.rm = TRUE),
    .groups = "drop"
  )

df_yearly <- df_long %>%
  pivot_wider(
    names_from = year,
    values_from = annual_precip
  )

df_yearly <- df_yearly[, c(1, 107:131)]

library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Warning: package 'stringr' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v purrr     1.0.2
## v ggplot2   3.5.1     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
mat <- df_yearly %>%
  column_to_rownames("Region") %>%
  as.matrix()

mat_scaled <- scale(mat, center = TRUE, scale = TRUE)

library(pvclust)
```

```
## Warning: package 'pvclust' was built under R version 4.3.3
```

```r
set.seed(1001)

fit <- pvclust(t(mat_scaled), method.hclust = "ward.D2", method.dist = "euclidean")
```

```
## Bootstrap (r = 0.48)... Done.
## Bootstrap (r = 0.6)... Done.
## Bootstrap (r = 0.68)... Done.
## Bootstrap (r = 0.8)... Done.
## Bootstrap (r = 0.88)... Done.
## Bootstrap (r = 1.0)... Done.
## Bootstrap (r = 1.08)... Done.
## Bootstrap (r = 1.2)... Done.
## Bootstrap (r = 1.28)... Done.
## Bootstrap (r = 1.4)... Done.
```

```r
clusters <- pvpick(fit, alpha = 0.95, pv = "au", type = "geq")$clusters

clusters
```

```
## [[1]]
## [1] "North Dakota, Northwest"    "North Dakota, West Central"
##
## [[2]]
## [1] "Maryland, Appalachian Mountains" "West Virginia, Northeastern"
##
## [[3]]
## [1] "Utah, North Central"      "Utah, Northern Mountains"
##
## [[4]]
## [1] "Montana, Central"       "Montana, South Central"
##
## [[5]]
## [1] "Washington, Central Basin"     "Washington, Okanogan Big Bend"
##
## [[6]]
## [1] "Michigan, South Central Lower" "Michigan, Southwest Lower"
##
## [[7]]
## [1] "Nebraska, Central"       "Nebraska, South Central"
##
## [[8]]
## [1] "Iowa, Northeast"      "Wisconsin, Southwest"
##
## [[9]]
## [1] "Idaho, North Central Canyons"      "Washington, Plouse Blue Mountains"
##
## [[10]]
## [1] "Illinois, Northeast" "Illinois, Northwest"
##
## [[11]]
## [1] "Maine, Coastal"        "Maine, South Interior"
##
## [[12]]
```

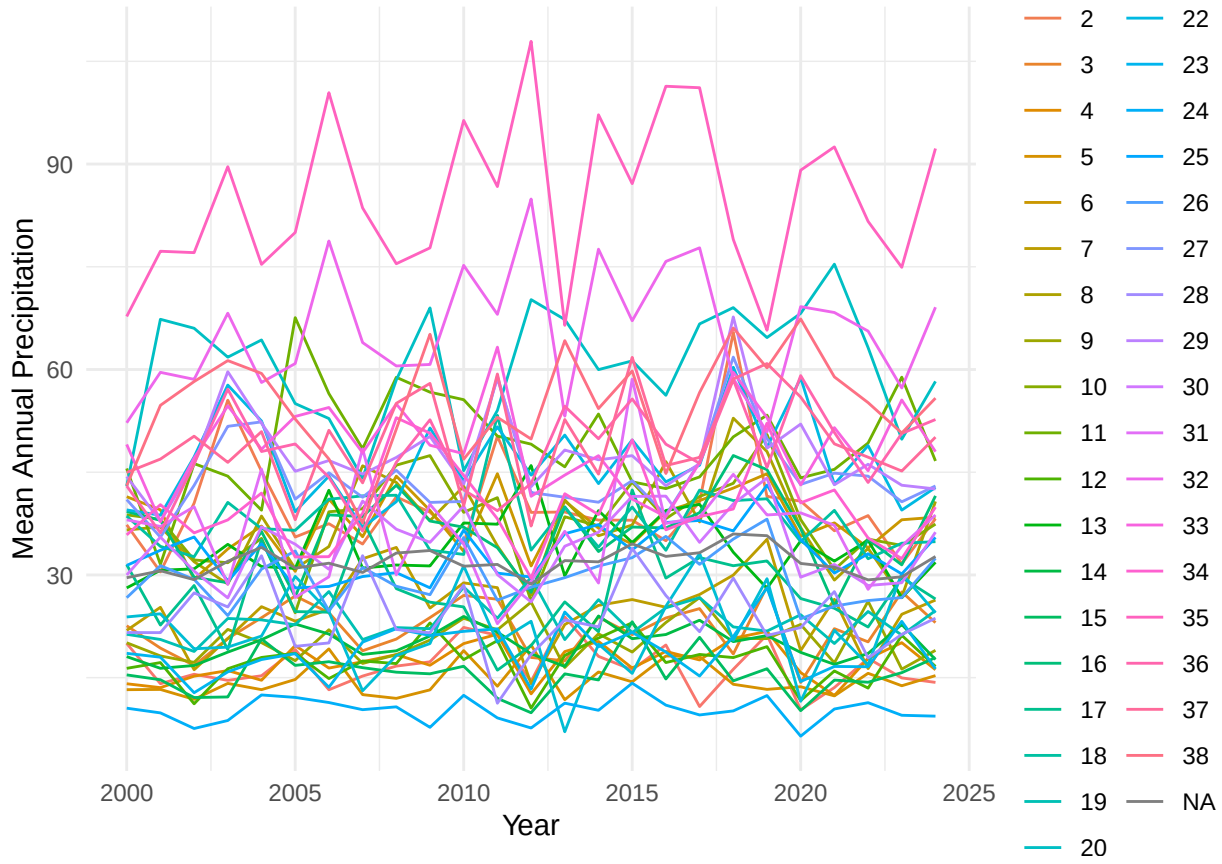
```
## [1] "Colorado, Kansas Drainage" "Colorado, Platte Drainage"
## [3] "Nebraska, Panhandle"
##
## [[13]]
## [1] "Idaho, Panhandle"                "Washington, East Slope Cascades"
##
## [[14]]
## [1] "Idaho, Eastern Highlands"    "Idaho, Northeastern Valleys"
## [3] "Montana, Southwestern"
##
## [[15]]
## [1] "Colorado, Arkansas Drainage"     "New Mexico, Northeastern Plains"
## [3] "New Mexico, Northern Mountains"
##
## [[16]]
## [1] "Wisconsin, South Central" "Wisconsin, Southeast"
##
## [[17]]
## [1] "Oklahoma, Southwest"    "Oklahoma, West Central"
##
## [[18]]
## [1] "Ohio, North Central" "Ohio, Northwest"
##
## [[19]]
## [1] "Oregon, North Central"    "Oregon, Northeast"
## [3] "Washington, Northeastern"
##
## [[20]]
## [1] "Mississippi, South Central" "Mississippi, Southeast"
##
## [[21]]
## [1] "California, Central Coast Drainage"
## [2] "California, Northeast Interior Drainage"
## [3] "California, San Joaquin Drainage"
##
## [[22]]
## [1] "Virginia, Southwestern Mountain" "West Virginia, Southern"
## [3] "West Virginia, Southwestern"
##
## [[23]]
## [1] "North Dakota, South Central" "South Dakota, Black Hills"
## [3] "South Dakota, Central"       "South Dakota, North Central"
## [5] "South Dakota, Southwest"
##
## [[24]]
## [1] "Arizona, Northeast"               "Arizona, Northwest"
## [3] "Arizona, South Central"           "New Mexico, Central Valley"
## [5] "New Mexico, Northwestern Plateau" "New Mexico, Southern Desert"
## [7] "Utah, Southeast"
##
## [[25]]
## [1] "Michigan, East Upper"     "Michigan, West Upper"
## [3] "Wisconsin, Central"       "Wisconsin, East Central"
## [5] "Wisconsin, North Central" "Wisconsin, Northeast"
```

```
## [7] "Wisconsin, Northwest"
##
## [[26]]
## [1] "Iowa, Northwest"         "Minnesota, Central"
## [3] "Minnesota, East Central" "Minnesota, Northeast"
## [5] "Minnesota, Southwest"
##
## [[27]]
##  [1] "Ohio, South Central"
##  [2] "Ohio, Southeast"
##  [3] "Ohio, Southwest"
##  [4] "Pennsylvania, Central Mountains"
##  [5] "Pennsylvania, Lower Susquehanna"
##  [6] "Pennsylvania, Middle Susquehanna"
##  [7] "Pennsylvania, Northwest Plateau"
##  [8] "Pennsylvania, South Central Mountains"
##  [9] "Pennsylvania, Southwest Plateau"
## [10] "Pennsylvania, Upper Susquehanna"
## [11] "West Virginia, Northwestern"
##
## [[28]]
## [1] "Texas, Edwards Plateau"    "Texas, Low Rolling Plains"
## [3] "Texas, Lower Valley"       "Texas, South"
##
## [[29]]
## [1] "Delaware, Northern"                  "Maryland, Allegheny Plateau"
## [3] "Maryland, North Central"             "New Jersey, Southern"
## [5] "Pennsylvania, Southeastern Piedmont" "West Virginia, Central"
## [7] "West Virginia, North Central"
##
## [[30]]
## [1] "Iowa, North Central"      "Iowa, Southwest"
## [3] "Iowa, West Central"       "Kansas, East Central"
## [5] "Kansas, Northeast"        "Minnesota, South Central"
## [7] "Minnesota, Southeast"     "Wisconsin, West Central"
##
## [[31]]
## [1] "Oklahoma, Central"       "Oklahoma, South Central"
## [3] "Texas, North Central"    "Texas, South Central"
##
## [[32]]
## [1] "Oregon, Willamette Valley"
## [2] "Washington, East Olympic Cascade Foothills"
##
## [[33]]
##  [1] "Connecticut, Central"
##  [2] "Connecticut, Northwest"
##  [3] "Massachusetts, Central"
##  [4] "Massachusetts, Coastal"
##  [5] "Massachusetts, West"
##  [6] "New Hampshire, South"
##  [7] "New Jersey, Northern"
##  [8] "New York, Coastal"
##  [9] "New York, Eastern Plateau"
```

```
## [10] "New York, Hudson Valley"
## [11] "New York, Mohawk Valley"
## [12] "New York, Northern Plateau"
## [13] "Pennsylvania, East Central Mountains"
## [14] "Pennsylvania, Pocono Mountains"
## [15] "Rhode Island, Rhode Island"
## [16] "Vermont, Southeastern"
##
## [[34]]
## [1] "Illinois, Central"             "Illinois, West Central"
## [3] "Illinois, West Southwest"      "Kansas, Southeast"
## [5] "Missouri, Northeast Prairie"   "Missouri, Northwest Prairie"
## [7] "Missouri, West Central Plains" "Oklahoma, Northeast"
##
## [[35]]
## [1] "Oregon, Coastal Area"               "Washington, Cascade Mountains West"
## [3] "Washington, West Olympic Coast"
##
## [[36]]
##  [1] "Georgia, Central"
##  [2] "Georgia, East Central"
##  [3] "Georgia, South Central"
##  [4] "Georgia, Southeast"
##  [5] "Georgia, Southwest"
##  [6] "North Carolina, Central Coastal Plain"
##  [7] "North Carolina, Northern Coastal Plain"
##  [8] "North Carolina, Southern Coastal Plain"
##  [9] "South Carolina, Central"
## [10] "South Carolina, North Central"
## [11] "South Carolina, Northeast"
## [12] "South Carolina, Southern"
## [13] "South Carolina, West Central"
## [14] "Virginia, Tidewater"
##
## [[37]]
##  [1] "Arkansas, Central"        "Arkansas, East Central"
##  [3] "Arkansas, North Central"  "Arkansas, Northeast"
##  [5] "Arkansas, Northwest"      "Arkansas, West Central"
##  [7] "Illinois, East Southeast" "Illinois, Southeast"
##  [9] "Illinois, Southwest"      "Indiana, South Central"
## [11] "Indiana, Southeast"       "Indiana, Southwest"
## [13] "Kentucky, Blue Grass"     "Kentucky, Central"
## [15] "Kentucky, Eastern"        "Kentucky, Western"
## [17] "Missouri, Bootheel"       "Missouri, East Ozarks"
## [19] "Missouri, West Ozarks"    "Oklahoma, East Central"
## [21] "Oklahoma, Southeast"
##
## [[38]]
##  [1] "Alabama, Appalachian Mountain"
##  [2] "Alabama, Coastal Plain"
##  [3] "Alabama, Eastern Valley"
##  [4] "Alabama, North Valley"
##  [5] "Alabama, Piedmont Plateau"
##  [6] "Alabama, Prairie"
```

```
##  [7] "Alabama, Upper Plains"
##  [8] "Florida, Everglades and Southwest Co"
##  [9] "Florida, North"
## [10] "Florida, North Central"
## [11] "Florida, Northwest"
## [12] "Florida, South Central"
## [13] "Georgia, North Central"
## [14] "Georgia, Northeast"
## [15] "Georgia, Northwest"
## [16] "Georgia, West Central"
## [17] "Mississippi, Central"
## [18] "Mississippi, East Central"
## [19] "Mississippi, North Central"
## [20] "Mississippi, Northeast"
## [21] "Mississippi, Upper Delta"
## [22] "North Carolina, Northern Mountains"
## [23] "North Carolina, Southern Mountains"
## [24] "South Carolina, Northwest"
## [25] "Tennessee, Cumberland Plateau"
## [26] "Tennessee, Eastern"
## [27] "Tennessee, Middle"
## [28] "Tennessee, Western"
```

```r
cluster_assignments <- tibble(
  Region = unlist(clusters),
  Cluster = rep(seq_along(clusters), times = sapply(clusters, length))
)

df_clustered <- df_long %>%
  left_join(cluster_assignments, by = "Region") %>%
  mutate(
    precip = as.numeric(annual_precip),      # <- fix here
    year = as.numeric(year)           # if plotting by year
  )

df_clustered <- df_clustered |>
  filter(year >= 2000)

# Plot average trend per cluster
df_clustered %>%
  group_by(Cluster, year) %>%
  summarise(mean_precip = mean(precip, na.rm = TRUE), .groups = "drop") %>%
  ggplot(aes(x = as.numeric(year), y = mean_precip, color = as.factor(Cluster))) +
  geom_line() +
  labs(x = "Year", y = "Mean Annual Precipitation", color = "Cluster") +
  theme_minimal()
```

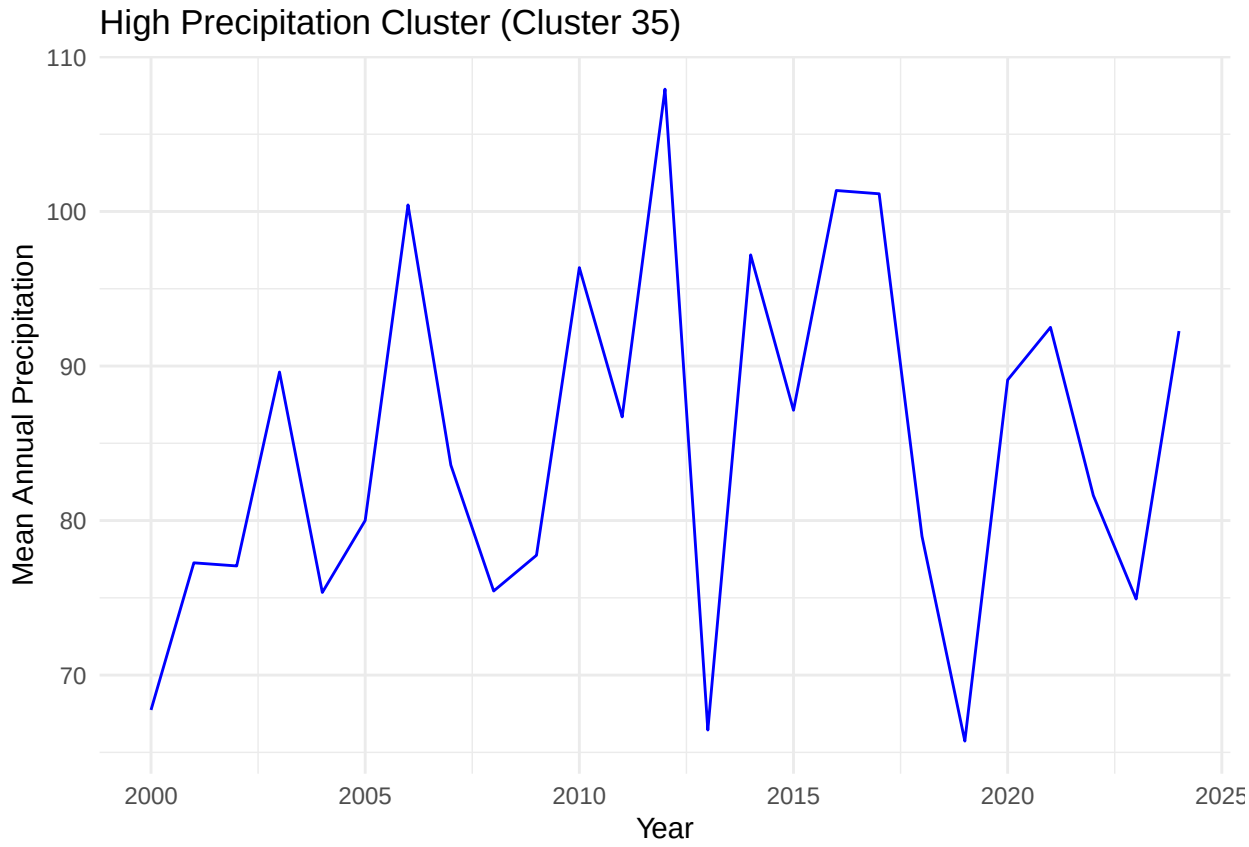



```
df_clustered %>%
  group_by(Cluster) %>%
  summarise(overall_mean = mean(precip, na.rm = TRUE)) %>%
  arrange(desc(overall_mean))
```

```
## # A tibble: 39 x 2
##    Cluster overall_mean
##      <int>        <dbl>
##  1      35         84.9
##  2      32         65.6
##  3      20         60.5
##  4      38         55.0
##  5      37         49.6
##  6      11         49.1
##  7      33         48.9
##  8      36         48.0
##  9      29         47.8
## 10      22         46.5
## # i 29 more rows
```

```
df_clustered %>%
  filter(Cluster == 35) %>%
  group_by(Cluster, year) %>%
  summarise(mean_precip = mean(precip, na.rm = TRUE), .groups = "drop") %>%
  ggplot(aes(x = as.numeric(year), y = mean_precip)) +
```

```
  geom_line(color = "blue") +
  labs(x = "Year", y = "Mean Annual Precipitation", title = "High Precipitation Cluster (Cluster 35)") +
  theme_minimal()
```




```
clusters
```

```
## [[1]]
## [1] "North Dakota, Northwest"    "North Dakota, West Central"
##
## [[2]]
## [1] "Maryland, Appalachian Mountains" "West Virginia, Northeastern"
##
## [[3]]
## [1] "Utah, North Central"      "Utah, Northern Mountains"
##
## [[4]]
## [1] "Montana, Central"       "Montana, South Central"
##
## [[5]]
## [1] "Washington, Central Basin"     "Washington, Okanogan Big Bend"
##
## [[6]]
## [1] "Michigan, South Central Lower" "Michigan, Southwest Lower"
##
## [[7]]
```

```
## [1] "Nebraska, Central"       "Nebraska, South Central"
##
## [[8]]
## [1] "Iowa, Northeast"      "Wisconsin, Southwest"
##
## [[9]]
## [1] "Idaho, North Central Canyons"      "Washington, Plouse Blue Mountains"
##
## [[10]]
## [1] "Illinois, Northeast" "Illinois, Northwest"
##
## [[11]]
## [1] "Maine, Coastal"        "Maine, South Interior"
##
## [[12]]
## [1] "Colorado, Kansas Drainage" "Colorado, Platte Drainage"
## [3] "Nebraska, Panhandle"
##
## [[13]]
## [1] "Idaho, Panhandle"                "Washington, East Slope Cascades"
##
## [[14]]
## [1] "Idaho, Eastern Highlands"    "Idaho, Northeastern Valleys"
## [3] "Montana, Southwestern"
##
## [[15]]
## [1] "Colorado, Arkansas Drainage"     "New Mexico, Northeastern Plains"
## [3] "New Mexico, Northern Mountains"
##
## [[16]]
## [1] "Wisconsin, South Central" "Wisconsin, Southeast"
##
## [[17]]
## [1] "Oklahoma, Southwest"    "Oklahoma, West Central"
##
## [[18]]
## [1] "Ohio, North Central" "Ohio, Northwest"
##
## [[19]]
## [1] "Oregon, North Central"    "Oregon, Northeast"
## [3] "Washington, Northeastern"
##
## [[20]]
## [1] "Mississippi, South Central" "Mississippi, Southeast"
##
## [[21]]
## [1] "California, Central Coast Drainage"
## [2] "California, Northeast Interior Drainage"
## [3] "California, San Joaquin Drainage"
##
## [[22]]
## [1] "Virginia, Southwestern Mountain" "West Virginia, Southern"
## [3] "West Virginia, Southwestern"
##
```

```
## [[23]]
## [1] "North Dakota, South Central" "South Dakota, Black Hills"
## [3] "South Dakota, Central"       "South Dakota, North Central"
## [5] "South Dakota, Southwest"
##
## [[24]]
## [1] "Arizona, Northeast"               "Arizona, Northwest"
## [3] "Arizona, South Central"           "New Mexico, Central Valley"
## [5] "New Mexico, Northwestern Plateau" "New Mexico, Southern Desert"
## [7] "Utah, Southeast"
##
## [[25]]
## [1] "Michigan, East Upper"     "Michigan, West Upper"
## [3] "Wisconsin, Central"       "Wisconsin, East Central"
## [5] "Wisconsin, North Central" "Wisconsin, Northeast"
## [7] "Wisconsin, Northwest"
##
## [[26]]
## [1] "Iowa, Northwest"         "Minnesota, Central"
## [3] "Minnesota, East Central" "Minnesota, Northeast"
## [5] "Minnesota, Southwest"
##
## [[27]]
##  [1] "Ohio, South Central"
##  [2] "Ohio, Southeast"
##  [3] "Ohio, Southwest"
##  [4] "Pennsylvania, Central Mountains"
##  [5] "Pennsylvania, Lower Susquehanna"
##  [6] "Pennsylvania, Middle Susquehanna"
##  [7] "Pennsylvania, Northwest Plateau"
##  [8] "Pennsylvania, South Central Mountains"
##  [9] "Pennsylvania, Southwest Plateau"
## [10] "Pennsylvania, Upper Susquehanna"
## [11] "West Virginia, Northwestern"
##
## [[28]]
## [1] "Texas, Edwards Plateau"    "Texas, Low Rolling Plains"
## [3] "Texas, Lower Valley"       "Texas, South"
##
## [[29]]
## [1] "Delaware, Northern"                   "Maryland, Allegheny Plateau"
## [3] "Maryland, North Central"              "New Jersey, Southern"
## [5] "Pennsylvania, Southeastern Piedmont" "West Virginia, Central"
## [7] "West Virginia, North Central"
##
## [[30]]
## [1] "Iowa, North Central"      "Iowa, Southwest"
## [3] "Iowa, West Central"       "Kansas, East Central"
## [5] "Kansas, Northeast"        "Minnesota, South Central"
## [7] "Minnesota, Southeast"     "Wisconsin, West Central"
##
## [[31]]
## [1] "Oklahoma, Central"       "Oklahoma, South Central"
## [3] "Texas, North Central"    "Texas, South Central"
```

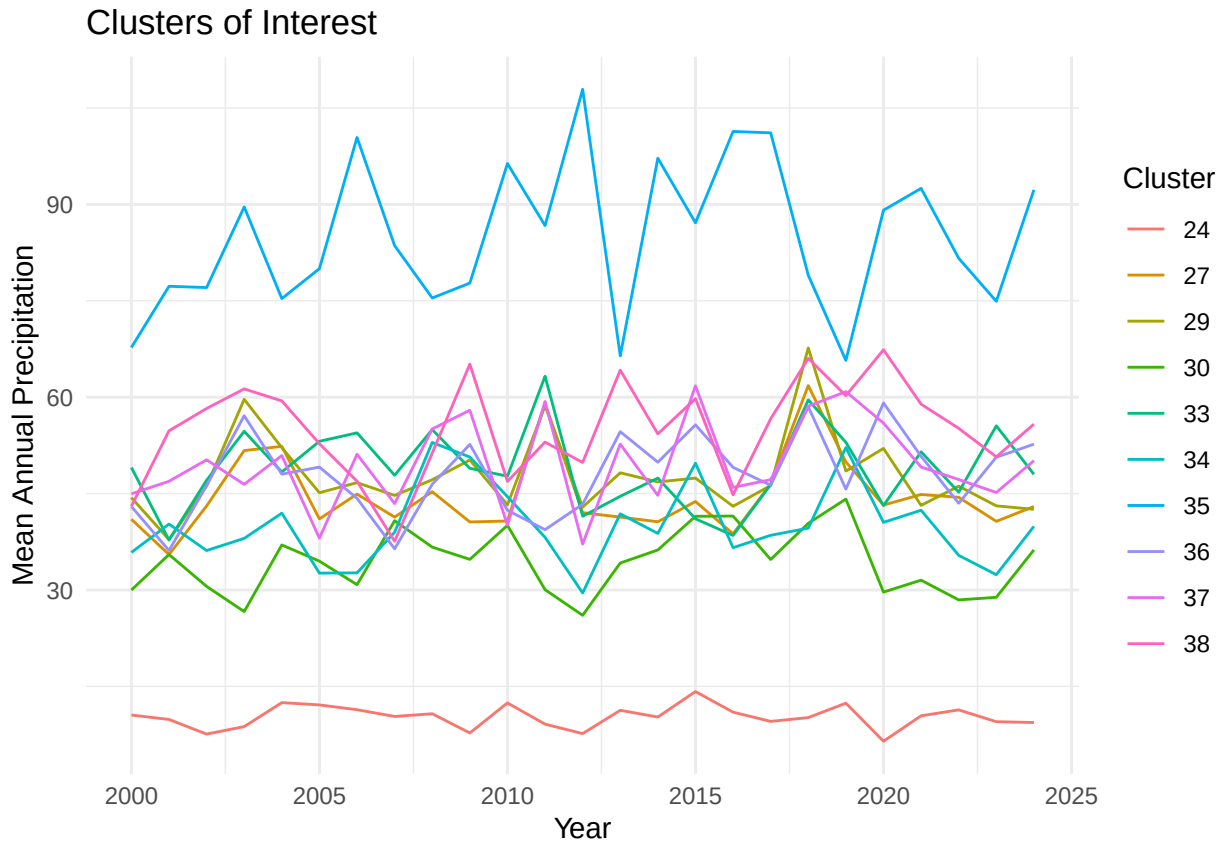
```
##
## [[32]]
## [1] "Oregon, Willamette Valley"
## [2] "Washington, East Olympic Cascade Foothills"
##
## [[33]]
##  [1] "Connecticut, Central"
##  [2] "Connecticut, Northwest"
##  [3] "Massachusetts, Central"
##  [4] "Massachusetts, Coastal"
##  [5] "Massachusetts, West"
##  [6] "New Hampshire, South"
##  [7] "New Jersey, Northern"
##  [8] "New York, Coastal"
##  [9] "New York, Eastern Plateau"
## [10] "New York, Hudson Valley"
## [11] "New York, Mohawk Valley"
## [12] "New York, Northern Plateau"
## [13] "Pennsylvania, East Central Mountains"
## [14] "Pennsylvania, Pocono Mountains"
## [15] "Rhode Island, Rhode Island"
## [16] "Vermont, Southeastern"
##
## [[34]]
## [1] "Illinois, Central"             "Illinois, West Central"
## [3] "Illinois, West Southwest"      "Kansas, Southeast"
## [5] "Missouri, Northeast Prairie"   "Missouri, Northwest Prairie"
## [7] "Missouri, West Central Plains" "Oklahoma, Northeast"
##
## [[35]]
## [1] "Oregon, Coastal Area"               "Washington, Cascade Mountains West"
## [3] "Washington, West Olympic Coast"
##
## [[36]]
##  [1] "Georgia, Central"
##  [2] "Georgia, East Central"
##  [3] "Georgia, South Central"
##  [4] "Georgia, Southeast"
##  [5] "Georgia, Southwest"
##  [6] "North Carolina, Central Coastal Plain"
##  [7] "North Carolina, Northern Coastal Plain"
##  [8] "North Carolina, Southern Coastal Plain"
##  [9] "South Carolina, Central"
## [10] "South Carolina, North Central"
## [11] "South Carolina, Northeast"
## [12] "South Carolina, Southern"
## [13] "South Carolina, West Central"
## [14] "Virginia, Tidewater"
##
## [[37]]
##  [1] "Arkansas, Central"        "Arkansas, East Central"
##  [3] "Arkansas, North Central"  "Arkansas, Northeast"
##  [5] "Arkansas, Northwest"      "Arkansas, West Central"
##  [7] "Illinois, East Southeast" "Illinois, Southeast"
```

```
##  [9] "Illinois, Southwest"      "Indiana, South Central"
## [11] "Indiana, Southeast"       "Indiana, Southwest"
## [13] "Kentucky, Blue Grass"     "Kentucky, Central"
## [15] "Kentucky, Eastern"        "Kentucky, Western"
## [17] "Missouri, Bootheel"       "Missouri, East Ozarks"
## [19] "Missouri, West Ozarks"    "Oklahoma, East Central"
## [21] "Oklahoma, Southeast"
##
## [[38]]
##  [1] "Alabama, Appalachian Mountain"
##  [2] "Alabama, Coastal Plain"
##  [3] "Alabama, Eastern Valley"
##  [4] "Alabama, North Valley"
##  [5] "Alabama, Piedmont Plateau"
##  [6] "Alabama, Prairie"
##  [7] "Alabama, Upper Plains"
##  [8] "Florida, Everglades and Southwest Co"
##  [9] "Florida, North"
## [10] "Florida, North Central"
## [11] "Florida, Northwest"
## [12] "Florida, South Central"
## [13] "Georgia, North Central"
## [14] "Georgia, Northeast"
## [15] "Georgia, Northwest"
## [16] "Georgia, West Central"
## [17] "Mississippi, Central"
## [18] "Mississippi, East Central"
## [19] "Mississippi, North Central"
## [20] "Mississippi, Northeast"
## [21] "Mississippi, Upper Delta"
## [22] "North Carolina, Northern Mountains"
## [23] "North Carolina, Southern Mountains"
## [24] "South Carolina, Northwest"
## [25] "Tennessee, Cumberland Plateau"
## [26] "Tennessee, Eastern"
## [27] "Tennessee, Middle"
## [28] "Tennessee, Western"
```

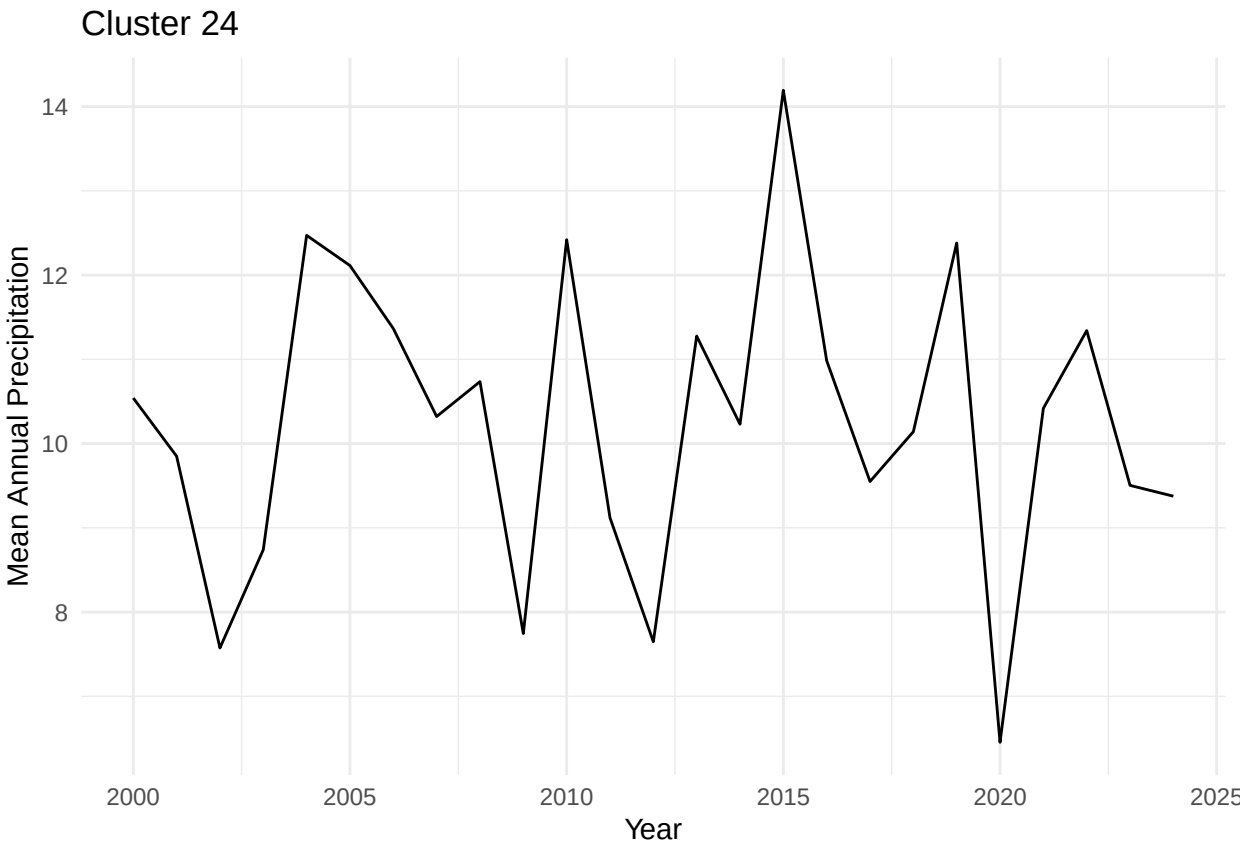
The model produced many clusters. There are a number of them with only one or two regions. It is difficult to identify any particularly notable differences here, with the exception of cluster 35, which I have isolated in the second plot, which contains three regions in the Pacific Northwest - a particularly rainy part of the country. In order to understand these results better, I will isolate a few clusters of interest. I will choose clusters with many regions, which presumably show general trends instead of outliers, and clusters that contain one of the areas in the drought that prompted the question for this project.

```
int_clus <- c(39, 38, 37, 36, 35, 34, 33, 30, 29, 27, 24)
#AZ, NW is in cluster 24

df_clustered %>%
  filter(Cluster %in% int_clus) |>
  group_by(Cluster, year) %>%
  summarise(mean_precip = mean(precip, na.rm = TRUE), .groups = "drop") %>%
  ggplot(aes(x = as.numeric(year), y = mean_precip, color = as.factor(Cluster))) +
  geom_line() +
```

```
  labs(x = "Year", y = "Mean Annual Precipitation", color = "Cluster", title = "Clusters of Interest") +
  theme_minimal()
```




```
df_clustered %>%
  filter(Cluster == 24) |>
  group_by(Cluster, year) %>%
  summarise(mean_precip = mean(precip, na.rm = TRUE), .groups = "drop") %>%
  ggplot(aes(x = as.numeric(year), y = mean_precip)) +
  geom_line() +
  labs(x = "Year", y = "Mean Annual Precipitation", title = "Cluster 24") +
  theme_minimal()
```



The graph of the clusters of interest shows us a couple of helpful things. It contains the largest clusters and two clusters with atypical rainfall, which are also the two clusters we chose because of other information (the second graph helps to make positive identification of the especially low rainfall cluster in the clusters of interest as cluster 24, the cluster containing the drought region). The large clusters with more general trends all appear somewhat similar. Certain of those clusters are higher and lower, but they all stay within a reasonable range and display a lot of similar peaks and dips. The Pacific Northwest cluster (35) is far above the rest. The drought cluster (24, containing Northwest Arizona) is far below the others. Returning to the graph with all of the clusters, it is clear that cluster 24 is still noticeably lower than any other cluster, even with all of our small clusters on the plot. It does not, however, appear to be an outlier in that graph.

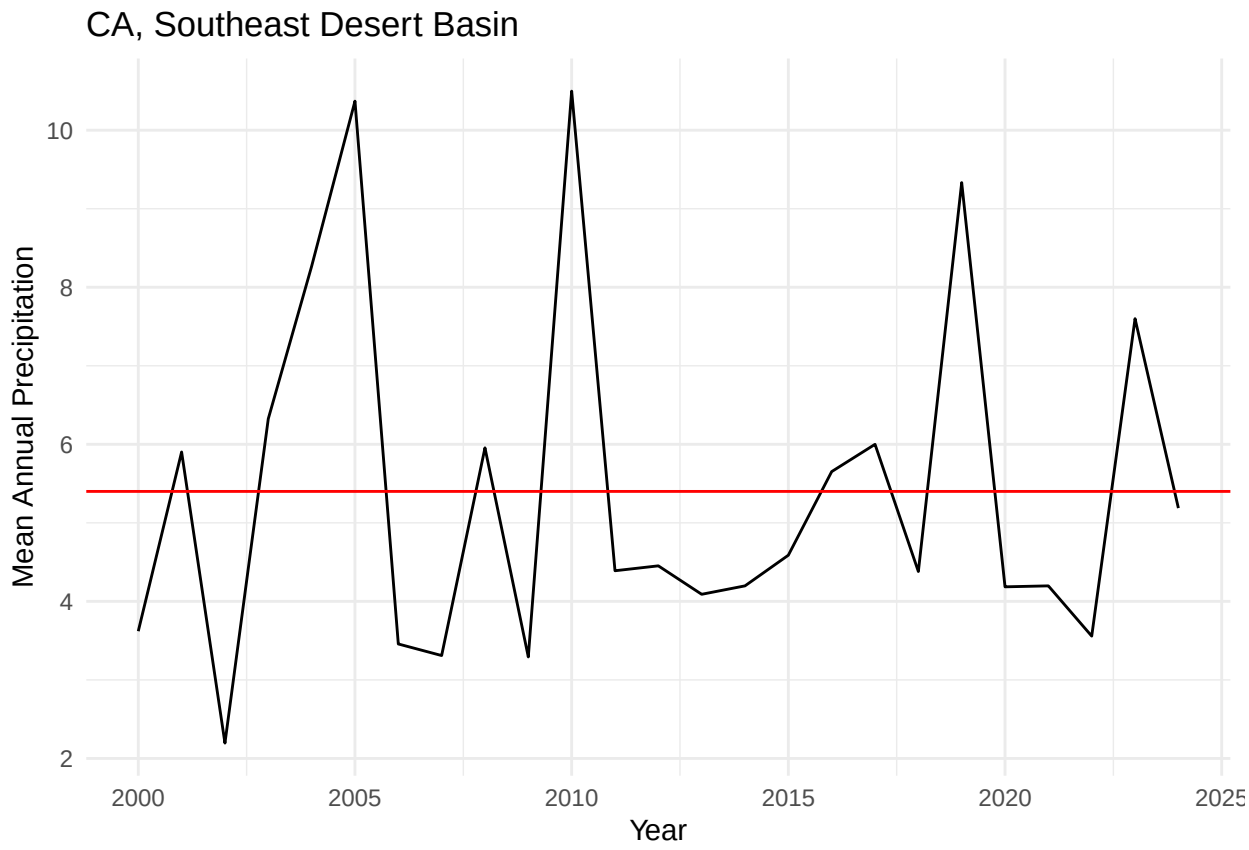
Conclusion

```
clusters[[24]]
```

```
## [1] "Arizona, Northeast"                "Arizona, Northwest"
## [3] "Arizona, South Central"            "New Mexico, Central Valley"
## [5] "New Mexico, Northwestern Plateau" "New Mexico, Southern Desert"
## [7] "Utah, Southeast"
```

```
region_mean <- df_clustered %>%
  filter(Region == "California, Southeast Desert Basin") %>%
  summarise(mean_precip = mean(precip, na.rm = TRUE)) %>%
  pull(mean_precip)
```

```r
df_clustered %>%
  filter(Region == "California, Southeast Desert Basin") |>
  ggplot(aes(x = as.numeric(year), y = precip)) +
  geom_line() +
  geom_hline(yintercept = region_mean, col = "red") +
  labs(x = "Year", y = "Mean Annual Precipitation", title = "CA, Southeast Desert Basin") +
  theme_minimal()
```



This analysis demonstrates that there are some regions to be concerned about because of changes in their precipitation. Northwest Arizona, one of the drought sites for the prompting question is in a cluster with, on average, considerably lower precipitation than anywhere else - and that level is declining. Examining that cluster reveals, unsurprisingly, that it is comprised of nearby regions in Arizona, New Mexico, and Utah. These are all hot, dry areas, in general. However, measures to conserve water have not, to my knowledge, needed to be taken in New Mexico or Utah yet. Regulators in these areas should more closely examine their data and prepare to take measures to conserve water.

However, there is also reason to think that the analysis demonstrates that we do not need to be overly concerned about the trends in this cluster and other clusters that dip to similar levels. Both of the drought regions of interest in California and Nevada were not assigned a cluster at all - marking them as outlier patterns. An examination of the trend for Southeast Desert Basins in California, a region afflicted by the drought that prompted the initial question of this project, reveals that it receives far less precipitation than the cluster containing the Arizona region. The average value of the CA region is lower than the minimum value of the AZ cluster. This gives us reason to think that the necessity of federal intervention in water negotiations was truly an outlier situation - one that we do not have cause to think will happen again in the near future.