DATA 481 02_EDA

Abisai Lujan

2025-02-17

Read the data

Data sourced from Github repository rfordatascience/tidytuesday
(https://github.com/rfordatascience/tidytuesday/blob/main/data/2021/2021-07-20/readme.md)

`Data Reference:`
"The U.S. Drought Monitor is jointly produced by the National Drought Mitigation Center at the University of Nebraska-Lincoln, the United States Department of Agriculture, and the National Oceanic and Atmospheric Administration. Map courtesy of NDMC."

This data contains weekly recorded (mapped) drought levels, from 2001 to 2021, across the 50 U.S. states, U.S. Capital (DC), and Puerto Rico (PR).

Included in these records are categorical levels of drought, as well as statistics of the land and populations affected within each drought category.

Schema

`map_date`: Date map released

`state_abb`: State abbreviation

`valid_start`: Start date of weekly data

`valid_end`: End date of weekly data

`stat_fmt`: Statistic format (2 for "categorical", 1 for "cumulative" ")

`drought_lvl`: Drought level (None, DO, D1, D2, D3, D4) which corresponds to no drought, abnormally dry, moderate drought, severe drought, extreme drought or exceptional drought.

`area_pct`: Percent of state currently in that drought category

`area_total`: Total land area (sq miles) of state currently in that drought category

`pop_pct`: Population percent of total state population in that drought category

`pop_total`: Population total of that state in that drought category

Preview of dataset

```r
drought <- read.csv("RData/Drought_data/drought.csv")
head(drought)
```

```
##   map_date state_abb valid_start  valid_end stat_fmt drought_lvl area_pct
## 1 20210713        AK  2021-07-13 2021-07-19        2        None    74.35
## 2 20210713        AK  2021-07-13 2021-07-19        2          D0    25.65
## 3 20210713        AK  2021-07-13 2021-07-19        2          D1     0.00
## 4 20210713        AK  2021-07-13 2021-07-19        2          D2     0.00
## 5 20210713        AK  2021-07-13 2021-07-19        2          D3     0.00
## 6 20210713        AK  2021-07-13 2021-07-19        2          D4     0.00
##   area_total pop_pct pop_total
## 1   433133.2   33.91  240644.2
## 2   149435.1   66.09  468985.8
## 3        0.0    0.00       0.0
## 4        0.0    0.00       0.0
## 5        0.0    0.00       0.0
## 6        0.0    0.00       0.0
```

Dimensions of dataset

```r
#size of data frame:
cat("The drought dataset has:", dim(drought)[1], "rows and", dim(drought)[2], "columns")
```

```
## The drought dataset has: 325728 rows and 10 columns
```

Check for any missing values

```r
sum(is.na(drought))
```

```
## [1] 0
```

There are no missing values within the dataset

Clean the Data
First, observe the types and names of the variables

```
str(drought)
```

```
## 'data.frame':    325728 obs. of  10 variables:
##  $ map_date   : int  20210713 20210713 20210713 20210713 20210713 20210713 20210706 20210706 2021070
##  $ state_abb  : chr  "AK" "AK" "AK" "AK" ...
##  $ valid_start: chr  "2021-07-13" "2021-07-13" "2021-07-13" "2021-07-13" ...
##  $ valid_end  : chr  "2021-07-19" "2021-07-19" "2021-07-19" "2021-07-19" ...
##  $ stat_fmt   : int  2 2 2 2 2 2 2 2 2 2 ...
##  $ drought_lvl: chr  "None" "D0" "D1" "D2" ...
##  $ area_pct   : num  74.3 25.6 0 0 0 ...
##  $ area_total : num  433133 149435 0 0 0 ...
##  $ pop_pct    : num  33.9 66.1 0 0 0 ...
##  $ pop_total  : num  240644 468986 0 0 0 ...
```

Generally, it is good practice to change date variables such as `map_date`, `valid_start` and `valid_end` dates to numeric types

```
drought$map_date <- as.numeric(gsub("-", "", drought$map_date))
drought$valid_start <- as.numeric(gsub("-", "", drought$valid_start))
drought$valid_end <- as.numeric(gsub("-","", drought$valid_end))
```

```
##  num [1:325728] 20210713 20210713 20210713 20210713 20210713 ...
```

```
##  num [1:325728] 20210713 20210713 20210713 20210713 20210713 ...
```

```
##  num [1:325728] 20210719 20210719 20210719 20210719 20210719 ...
```

— *Remove redundant columns* —

Out of curiosity, I'd like to investigate `stat_fmt` since `drought_lvl`

```
cat("Unique values in stat_fmt: ",unique(drought$stat_fmt))
```

```
## Unique values in stat_fmt:  2
```

The only unique value within `stat_fmt` is 2, let's observe how many rows hold values of 2 and 1

```
nrows_sf1 <- nrow(drought[drought$stat_fmt == 1,])
nrows_sf2 <- nrow(drought[drought$stat_fmt == 2,])
cat("Number of rows where stat_fmt = 1:", nrows_sf1,
    "; Number of rows where stat_fmt = 2:",nrows_sf2,"\n")
```

```
## Number of rows where stat_fmt = 1: 0 ; Number of rows where stat_fmt = 2: 325728
```

```
## Number of rows in drought.csv:  325728
```

As the `stat_fmt` column does not provide any useful information I will remove it from the dataset.

```r
drought <- drought %>%
  select(-c("stat_fmt"))
names(drought)
```

```
## [1] "map_date"    "state_abb"   "valid_start" "valid_end"   "drought_lvl"
## [6] "area_pct"    "area_total"  "pop_pct"     "pop_total"
```

— Reorganize dataset —

Reorder `map_date` in ascending value for each state; just my personal preference

```
drought_ordered <- drought[order(drought$state_abb, drought$valid_start),]
head(drought_ordered)
```

```
##      map_date state_abb valid_start valid_end drought_lvl area_pct area_total
## 6259 20010717        AK    20010717  20010723        None    97.92  570476.56
## 6260 20010717        AK    20010717  20010723          D0     2.08   12091.73
## 6261 20010717        AK    20010717  20010723          D1     0.00       0.00
## 6262 20010717        AK    20010717  20010723          D2     0.00       0.00
## 6263 20010717        AK    20010717  20010723          D3     0.00       0.00
## 6264 20010717        AK    20010717  20010723          D4     0.00       0.00
##      pop_pct pop_total
## 6259   92.71 657899.35
## 6260    7.29  51730.64
## 6261    0.00      0.00
## 6262    0.00      0.00
## 6263    0.00      0.00
## 6264    0.00      0.00
```

We may observe that there is a lot of redundancy in rows per date. As each state has 6 rows for the single week. The 6 separate rows are the statistics for each drought category.

– Split the Data —

To prepare for smoother data visualization, I will split the organized dataset for different purposes. The goal is to have separate datasets for Area Percentage, Area Total, Population Percentage, and Population Total.

Each new dataset will each have a more organized view of each respective quantity;

That is the Area Percentage dataset will have columns for each drought level, which will hold their respective `area_pct` statistics. The Area Total dataset will also have columns for each drought level, which will hold their respective `area_total` statistics, and so forth.

This way, it will be easier to reference the different data sets for different visualizations. We may visualize trends among drought levels for separate statistics.

To prepare, I will first split the organized dataset by drought level

Drought level None. Rename variable names to avoid confusion when joining later.

```
drought_none <- drought_ordered %>%
  filter(drought_lvl=="None") %>%
  rename(None = drought_lvl,
         None.area_pct = area_pct,
         None.area_total = area_total,
         None.pop_pct = pop_pct,
         None.pop_total = pop_total)

head(drought_none)
```

```
##   map_date state_abb valid_start valid_end None None.area_pct None.area_total
## 1 20010717        AK    20010717  20010723 None         97.92        570476.6
## 2 20010724        AK    20010724  20010730 None         97.92        570476.6
## 3 20010731        AK    20010731  20010806 None        100.00        582568.3
## 4 20010807        AK    20010807  20010813 None        100.00        582568.3
## 5 20010814        AK    20010814  20010820 None        100.00        582568.3
## 6 20010821        AK    20010821  20010827 None        100.00        582568.3
##   None.pop_pct None.pop_total
## 1        92.71       657899.4
## 2        92.71       657899.4
## 3       100.00       709630.0
## 4       100.00       709630.0
## 5       100.00       709630.0
## 6       100.00       709630.0
```

Drought level D0. Rename variable names to avoid confusion when joining later.

```
drought_D0 <- drought_ordered %>%
  filter(drought_lvl == "D0") %>%
  rename(D0 = drought_lvl,
         D0.area_pct = area_pct,
         D0.area_total = area_total,
         D0.pop_pct = pop_pct,
         D0.pop_total = pop_total)
head(drought_D0)
```

```
##   map_date state_abb valid_start valid_end D0 D0.area_pct D0.area_total
## 1 20010717        AK    20010717  20010723 D0        2.08      12091.73
## 2 20010724        AK    20010724  20010730 D0        2.08      12091.73
## 3 20010731        AK    20010731  20010806 D0        0.00          0.00
## 4 20010807        AK    20010807  20010813 D0        0.00          0.00
## 5 20010814        AK    20010814  20010820 D0        0.00          0.00
## 6 20010821        AK    20010821  20010827 D0        0.00          0.00
##   D0.pop_pct D0.pop_total
## 1       7.29     51730.64
## 2       7.29     51730.64
## 3       0.00         0.00
## 4       0.00         0.00
## 5       0.00         0.00
## 6       0.00         0.00
```

Drought level D1. Rename variable names to avoid confusion when joining later.

```
drought_D1 <- drought_ordered %>%
  filter(drought_lvl == "D1") %>%
    rename(D1 = drought_lvl,
         D1.area_pct = area_pct,
         D1.area_total = area_total,
         D1.pop_pct = pop_pct,
         D1.pop_total = pop_total)
head(drought_D1)
```

```
##   map_date state_abb valid_start valid_end D1 D1.area_pct D1.area_total
## 1 20010717        AK    20010717  20010723 D1           0             0
## 2 20010724        AK    20010724  20010730 D1           0             0
## 3 20010731        AK    20010731  20010806 D1           0             0
## 4 20010807        AK    20010807  20010813 D1           0             0
## 5 20010814        AK    20010814  20010820 D1           0             0
## 6 20010821        AK    20010821  20010827 D1           0             0
##   D1.pop_pct D1.pop_total
## 1          0            0
## 2          0            0
## 3          0            0
## 4          0            0
## 5          0            0
## 6          0            0
```

Drought level D2. Rename variable names to avoid confusion when joining later.

```
drought_D2 <- drought_ordered %>%
  filter(drought_lvl == "D2") %>%
    rename(D2 = drought_lvl,
         D2.area_pct = area_pct,
         D2.area_total = area_total,
         D2.pop_pct = pop_pct,
         D2.pop_total = pop_total)
head(drought_D2)
```

```
##   map_date state_abb valid_start valid_end D2 D2.area_pct D2.area_total
## 1 20010717        AK    20010717  20010723 D2           0             0
## 2 20010724        AK    20010724  20010730 D2           0             0
## 3 20010731        AK    20010731  20010806 D2           0             0
## 4 20010807        AK    20010807  20010813 D2           0             0
## 5 20010814        AK    20010814  20010820 D2           0             0
## 6 20010821        AK    20010821  20010827 D2           0             0
##   D2.pop_pct D2.pop_total
## 1          0            0
## 2          0            0
## 3          0            0
## 4          0            0
## 5          0            0
## 6          0            0
```

Drought level D3. Rename variable names to avoid confusion when joining later.

```
drought_D3 <- drought_ordered %>%
  filter(drought_lvl == "D3") %>%
    rename(D3 = drought_lvl,
         D3.area_pct = area_pct,
         D3.area_total = area_total,
         D3.pop_pct = pop_pct,
         D3.pop_total = pop_total)
head(drought_D3)
```

```
##   map_date state_abb valid_start valid_end D3 D3.area_pct D3.area_total
## 1 20010717        AK    20010717  20010723 D3           0             0
## 2 20010724        AK    20010724  20010730 D3           0             0
## 3 20010731        AK    20010731  20010806 D3           0             0
## 4 20010807        AK    20010807  20010813 D3           0             0
## 5 20010814        AK    20010814  20010820 D3           0             0
## 6 20010821        AK    20010821  20010827 D3           0             0
##   D3.pop_pct D3.pop_total
## 1          0            0
## 2          0            0
## 3          0            0
## 4          0            0
## 5          0            0
## 6          0            0
```

Drought level D4. Rename variable names to avoid confusion when joining later.

```
drought_D4 <- drought_ordered %>%
  filter(drought_lvl == "D4") %>%
    rename(D4 = drought_lvl,
         D4.area_pct = area_pct,
         D4.area_total = area_total,
         D4.pop_pct = pop_pct,
         D4.pop_total = pop_total)
head(drought_D4)
```

```
##   map_date state_abb valid_start valid_end D4 D4.area_pct D4.area_total
## 1 20010717        AK    20010717  20010723 D4           0             0
## 2 20010724        AK    20010724  20010730 D4           0             0
## 3 20010731        AK    20010731  20010806 D4           0             0
## 4 20010807        AK    20010807  20010813 D4           0             0
## 5 20010814        AK    20010814  20010820 D4           0             0
## 6 20010821        AK    20010821  20010827 D4           0             0
##   D4.pop_pct D4.pop_total
## 1          0            0
## 2          0            0
## 3          0            0
## 4          0            0
## 5          0            0
## 6          0            0
```

— Create the new split datasets —

To avoid repeatedly creating new empty datasets, I will create an empty dataframe that will act as a base for easy data insertion.

```r
#dimensions
numcols <- ncol(drought_none) + 1
numrows <- nrow(drought_none)

#create a matrix with placeholder columns and convert to a data frame
empty_template <- data.frame(
  matrix(ncol = numcols,
         nrow = numrows))
```

The only difference between the new datasets will be their respective statistic values for each drought level. Therefore I will go ahead and rename the *constant* placeholder columns.

```r
template_cnames <- c("map_date", "state_abb", "valid_start", "valid_end",
                              "None", "D0", "D1", "D2", "D3", "D4")
colnames(empty_template) <- template_cnames
```

`drought_area_pct`: new dataset which will hold `area_pct` statistics from the separate drought level datasets.

```r
drought_area_pct <- empty_template %>%
  #insert correct data
  mutate(
  map_date = drought_none$map_date,
  state_abb = drought_none$state_abb,
  valid_start = drought_none$valid_start,
  valid_end = drought_none$valid_end,
  None = drought_none$None.area_pct,
  D0 = drought_D0$D0.area_pct,
  D1 = drought_D1$D1.area_pct,
  D2 = drought_D2$D2.area_pct,
  D3 = drought_D3$D3.area_pct,
  D4 = drought_D4$D4.area_pct
  )

head(drought_area_pct)
```

```
##   map_date state_abb valid_start valid_end   None   D0 D1 D2 D3 D4
## 1 20010717        AK    20010717  20010723  97.92 2.08  0  0  0  0
## 2 20010724        AK    20010724  20010730  97.92 2.08  0  0  0  0
## 3 20010731        AK    20010731  20010806 100.00 0.00  0  0  0  0
## 4 20010807        AK    20010807  20010813 100.00 0.00  0  0  0  0
## 5 20010814        AK    20010814  20010820 100.00 0.00  0  0  0  0
## 6 20010821        AK    20010821  20010827 100.00 0.00  0  0  0  0
```

`drought_area_total`: new dataset which will hold `area_total` statistics from the separate drought level datasets.

```
drought_area_total <- empty_template %>%
  #insert correct data
  mutate(
  map_date = drought_D0$map_date,
  state_abb = drought_D0$state_abb,
  valid_start = drought_D0$valid_start,
  valid_end = drought_D0$valid_end,
  None = drought_none$None.area_total,
  D0 = drought_D0$D0.area_total,
  D1 = drought_D1$D1.area_total,
  D2 = drought_D2$D2.area_total,
  D3 = drought_D3$D3.area_total,
  D4 = drought_D4$D4.area_total
  )

head(drought_area_total)
```

```
##   map_date state_abb valid_start valid_end     None       D0 D1 D2 D3 D4
## 1 20010717        AK    20010717  20010723 570476.6 12091.73  0  0  0  0
## 2 20010724        AK    20010724  20010730 570476.6 12091.73  0  0  0  0
## 3 20010731        AK    20010731  20010806 582568.3     0.00  0  0  0  0
## 4 20010807        AK    20010807  20010813 582568.3     0.00  0  0  0  0
## 5 20010814        AK    20010814  20010820 582568.3     0.00  0  0  0  0
## 6 20010821        AK    20010821  20010827 582568.3     0.00  0  0  0  0
```

`drought_pop_pct`: new dataset which will hold `pop_pct` statistics from the separate drought level datasets.

```
drought_pop_pct <- empty_template %>%
  #insert correct data
  mutate(
  map_date = drought_D1$map_date,
  state_abb = drought_D1$state_abb,
  valid_start = drought_D1$valid_start,
  valid_end = drought_D1$valid_end,
  None = drought_none$None.pop_pct,
  D0 = drought_D0$D0.pop_pct,
  D1 = drought_D1$D1.pop_pct,
  D2 = drought_D2$D2.pop_pct,
  D3 = drought_D3$D3.pop_pct,
  D4 = drought_D4$D4.pop_pct
  )

head(drought_pop_pct)
```

```
##   map_date state_abb valid_start valid_end   None   D0 D1 D2 D3 D4
## 1 20010717        AK    20010717  20010723  92.71 7.29  0  0  0  0
## 2 20010724        AK    20010724  20010730  92.71 7.29  0  0  0  0
## 3 20010731        AK    20010731  20010806 100.00 0.00  0  0  0  0
## 4 20010807        AK    20010807  20010813 100.00 0.00  0  0  0  0
## 5 20010814        AK    20010814  20010820 100.00 0.00  0  0  0  0
## 6 20010821        AK    20010821  20010827 100.00 0.00  0  0  0  0
```

`drought_pop_total`: new dataset which will hold `pop_total` statistics from the separate drought level datasets.

```
drought_pop_total <- empty_template %>%
  #insert correct data
  mutate(
  map_date = drought_D2$map_date,
  state_abb = drought_D2$state_abb,
  valid_start = drought_D2$valid_start,
  valid_end = drought_D2$valid_end,
  None = drought_none$None.pop_pct,
  D0 = drought_D0$D0.pop_total,
  D1 = drought_D1$D1.pop_total,
  D2 = drought_D2$D2.pop_total,
  D3 = drought_D3$D3.pop_total,
  D4 = drought_D4$D4.pop_total
  )
```

```
head(drought_pop_total)
```

```
##   map_date state_abb valid_start valid_end   None       D0 D1 D2 D3 D4
## 1 20010717        AK    20010717  20010723  92.71 51730.64  0  0  0  0
## 2 20010724        AK    20010724  20010730  92.71 51730.64  0  0  0  0
## 3 20010731        AK    20010731  20010806 100.00     0.00  0  0  0  0
## 4 20010807        AK    20010807  20010813 100.00     0.00  0  0  0  0
## 5 20010814        AK    20010814  20010820 100.00     0.00  0  0  0  0
## 6 20010821        AK    20010821  20010827 100.00     0.00  0  0  0  0
```

Exploratory Data Analysis

Lets see how the total state populations have been affected by drought

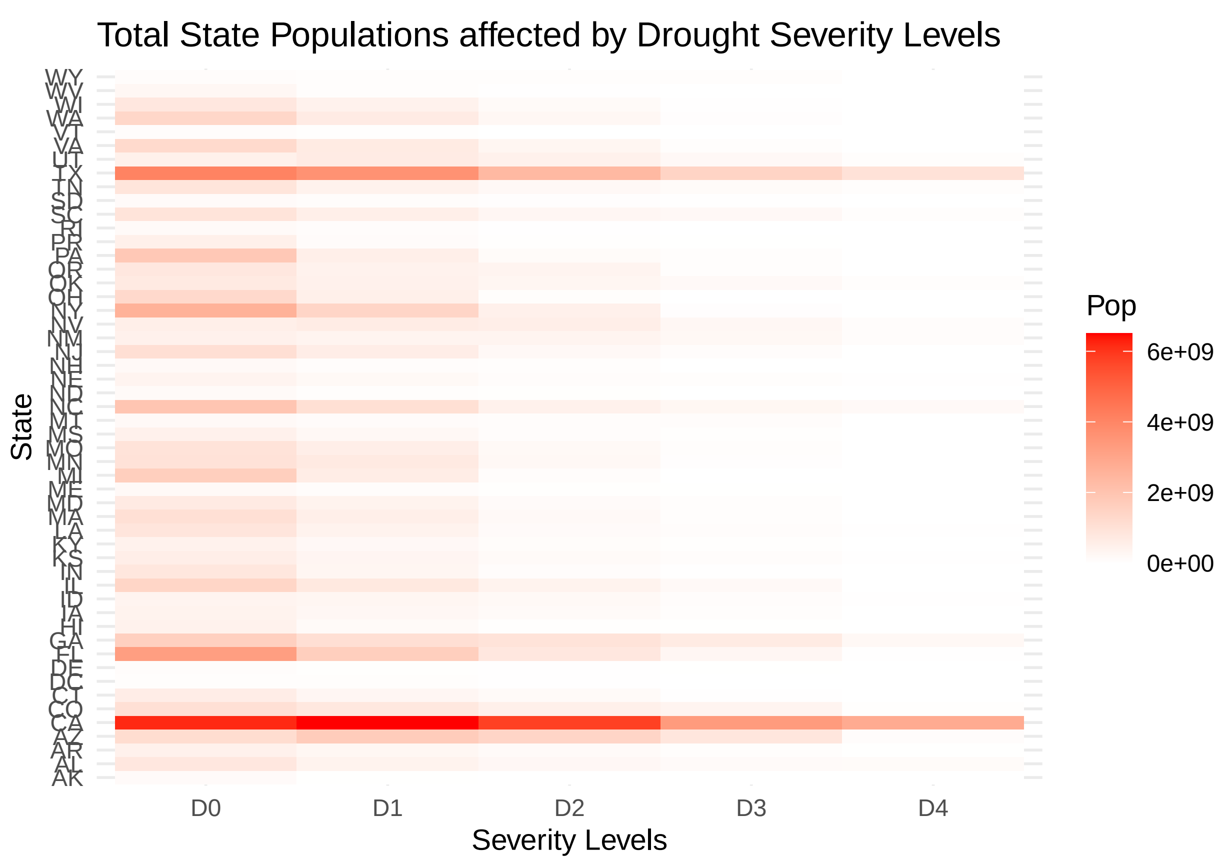
Visualization: Heat map for total populations affected by drought, summed by state from `drought_pop_total`

```
drought_categories <- c("D0", "D1", "D2", "D3", "D4")
```

```
state_summary_pt_sum <- drought_pop_total %>%
  group_by(state_abb) %>%
  summarise(across(all_of(drought_categories),
                   sum, na.rm = TRUE))
```

```
state_summary_long_pt_sum <- state_summary_pt_sum %>%
  pivot_longer(cols = all_of(drought_categories),
               names_to = "Category", values_to = "Pop")
```

```
ggplot(state_summary_long_pt_sum, aes(x = Category, y = state_abb, fill = Pop)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "red") +
  labs(title = "Total State Populations affected by Drought Severity Levels",
       x = "Severity Levels",
       y = "State") +
  theme_minimal()
```



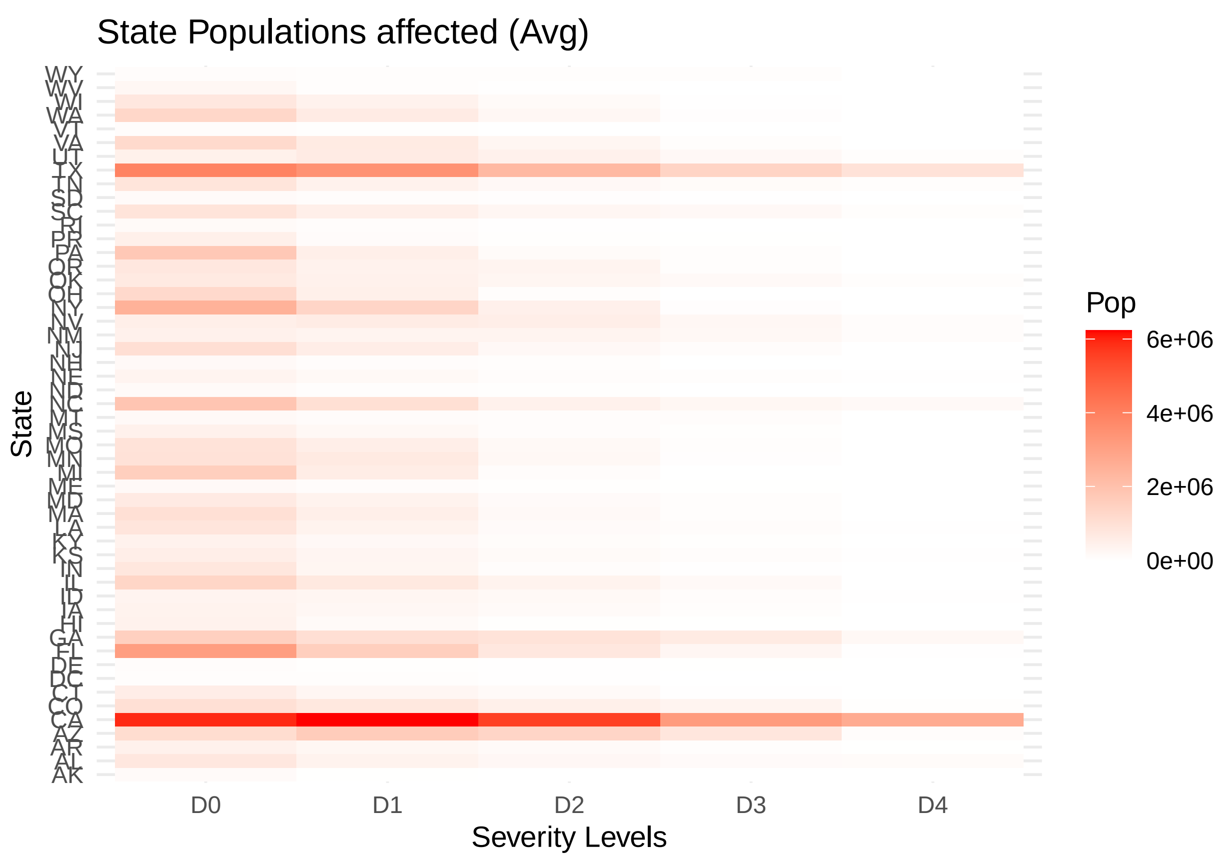
As the heat map shows, it is not very surprising that the populations of California and Texas have been reported to be highly affected by drought. California is highly populated. Both California and Texas are among hot, dry weather conditions that often receive droughts.

Visualization: heatmap for average drought scales summed by state from `drought_pop_total`

```
state_summary_pt_avg <- drought_pop_total %>%
  group_by(state_abb) %>%
  summarise(across(all_of(drought_categories),
                   mean, na.rm = TRUE))
```

```
state_summary_long_pt_avg <- state_summary_pt_avg %>%
  pivot_longer(cols = all_of(drought_categories),
               names_to = "Category", values_to = "Pop")

ggplot(state_summary_long_pt_avg, aes(x = Category, y = state_abb, fill = Pop)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "red") +
  labs(title = "State Populations affected (Avg)",
       x = "Severity Levels",
       y = "State") +
  theme_minimal()
```

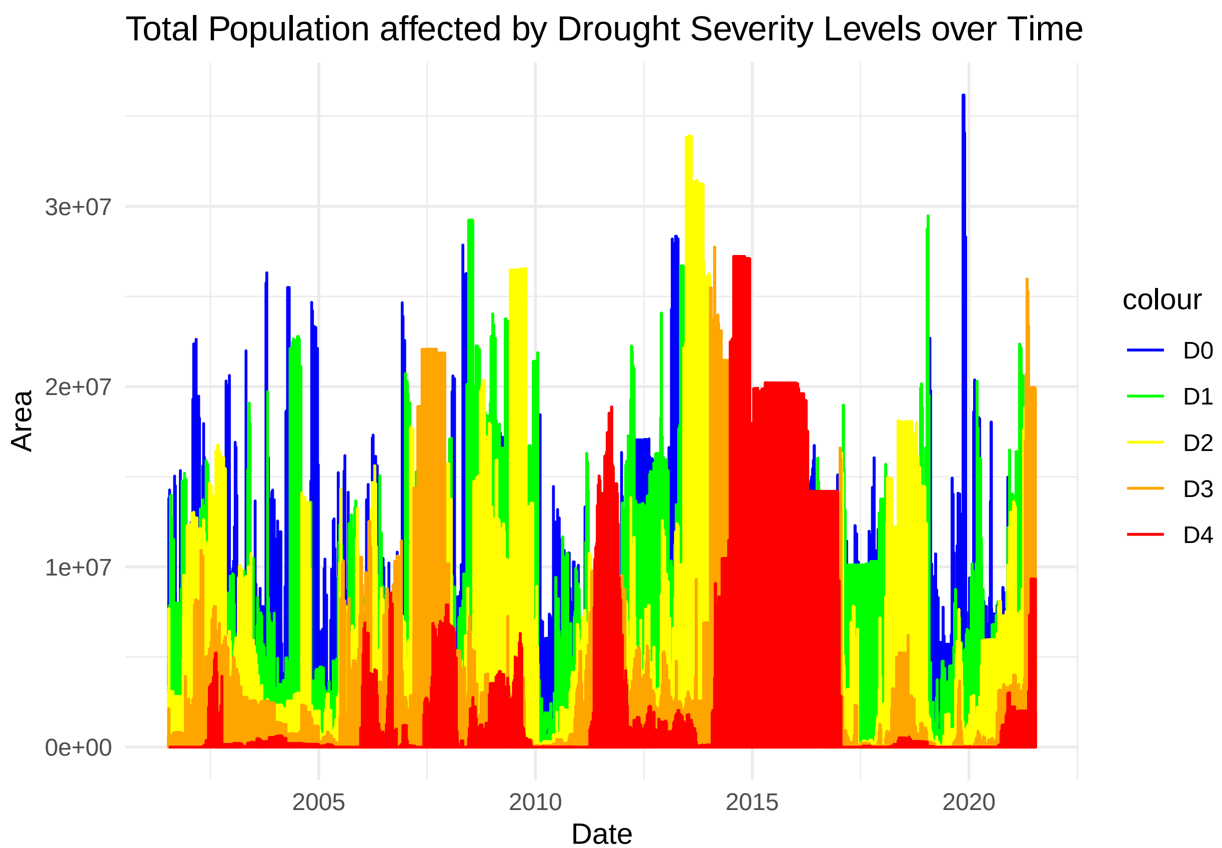


Similar to the total populations affected, the same states yield higher averages than the rest. Although the scale is lower overall, it still makes sense that higher populations will have higher averages.

Visualization: Total state populations affected by each drought severity level over time

```
drought_pop_total$map_date <- as.Date(as.character(drought_pop_total$map_date), format="%Y%m%d")
```

```
ggplot(drought_pop_total, aes(x=map_date)) +
  geom_line(aes(y=D0, color="D0")) +
  geom_line(aes(y=D1, color="D1")) +
  geom_line(aes(y=D2, color="D2")) +
  geom_line(aes(y=D3, color="D3")) +
  geom_line(aes(y=D4, color="D4")) +
  labs(title="Total Population affected by Drought Severity Levels over Time",
       x="Date",
       y = "Area") +
scale_color_manual(values = c("blue", "green", "yellow", "orange", "red")) + theme_minimal()
```



Focus on drought severity levels' affect on land area over time

First lets look at the total land area affected by each drought severity level over time

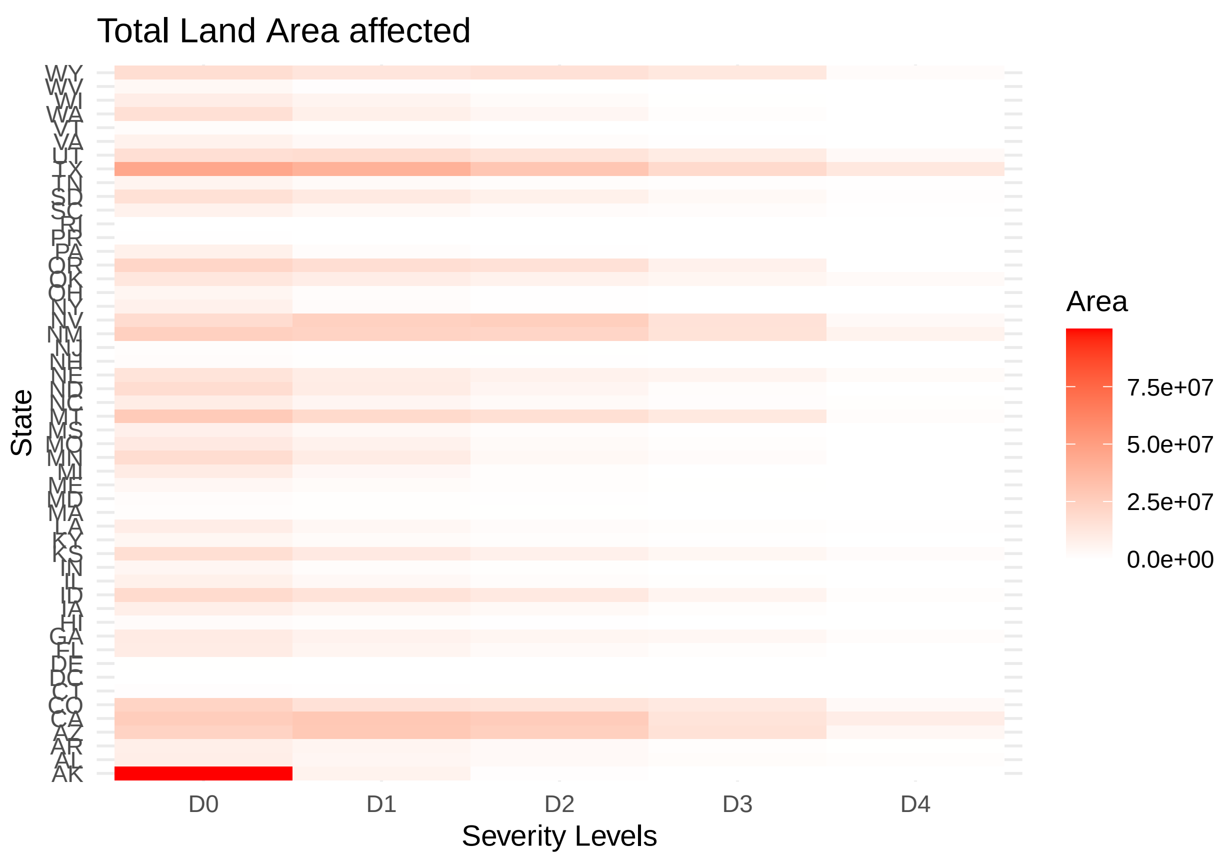
```
drought_categories <- c("D0", "D1", "D2", "D3", "D4")
```

```
state_summary_at_sum <- drought_area_total %>%
  group_by(state_abb) %>%
  summarise(across(all_of(drought_categories), sum, na.rm = TRUE))
```

Visualization: heatmap of total land area by state affected by each drought category, from `drought_area_total`

```
state_summary_long_at_sum <- state_summary_at_sum %>%
pivot_longer(cols = all_of(drought_categories), names_to = "Category", values_to = "Area")

ggplot(state_summary_long_at_sum, aes(x = Category, y = state_abb, fill = Area)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "red") +
  labs(title = "Total Land Area affected",
       x = "Severity Levels",
       y = "State") +
  theme_minimal()
```



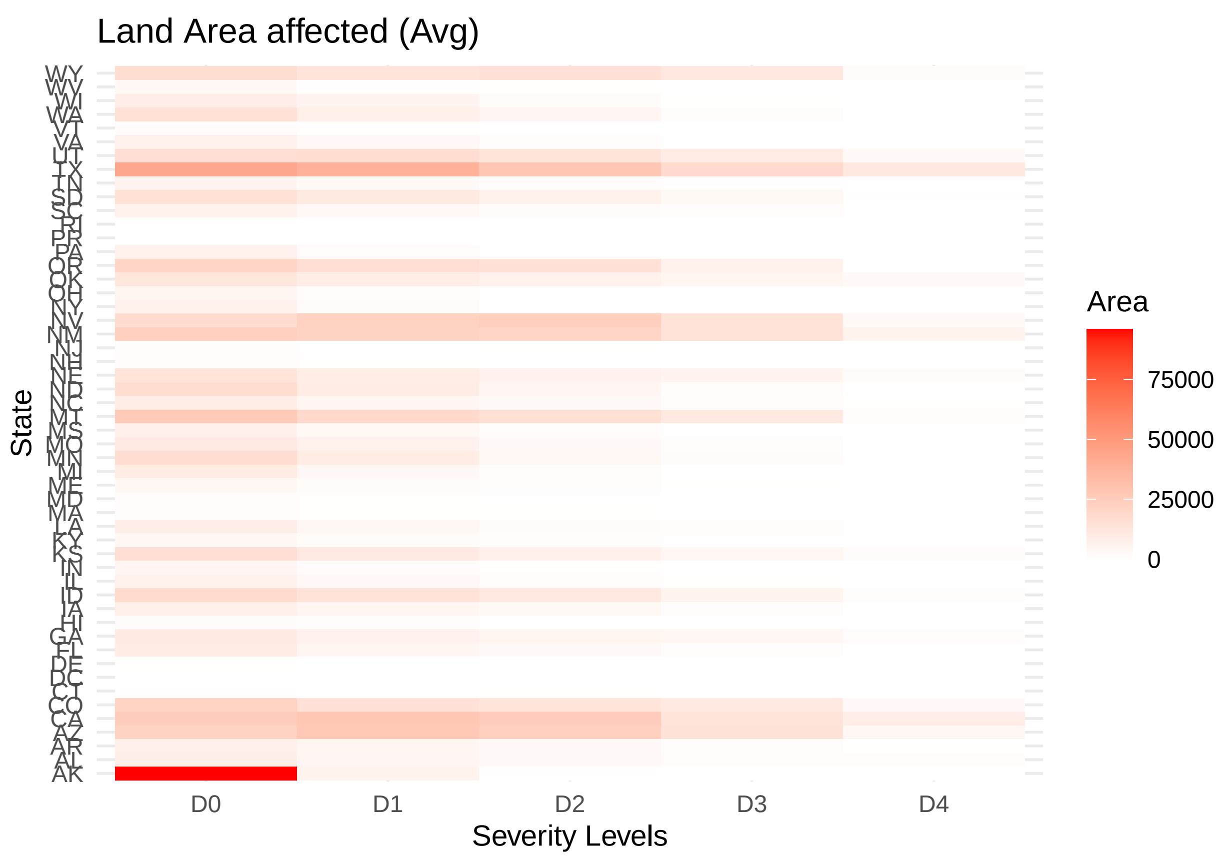
From `drought_area_total` let's observe a State-level summary of the average land area has been affected by each drought severity level over time

```
state_summary_at_avg <- drought_area_total %>%
  group_by(state_abb) %>%
  summarise(across(all_of(drought_categories), mean, na.rm = TRUE))
```

Visualization: heatmap for average of total land area by state affected by each drought category, from `drought_area_total`

```
state_summary_long_at_avg <- state_summary_at_avg %>%
pivot_longer(cols = all_of(drought_categories), names_to = "Category", values_to = "Area")

ggplot(state_summary_long_at_avg, aes(x = Category, y = state_abb, fill = Area)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "red") +
  labs(title = "Land Area affected (Avg)",
       x = "Severity Levels",
       y = "State") +
  theme_minimal()
```



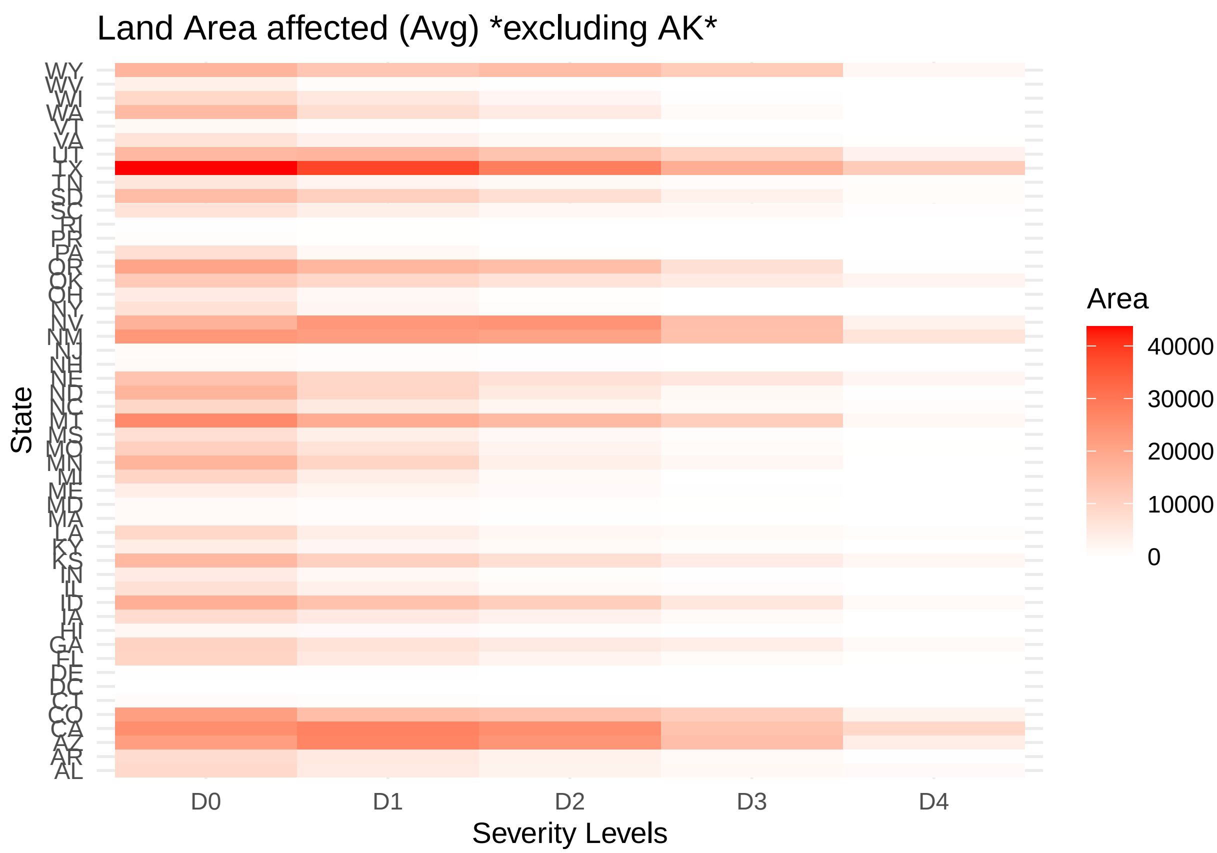
As we can see, Arkansas' total land area mostly experiences drought at the milder D0 category. States such Texas, Nevada, New Mexico, Colorado, California, and Arizona have a fairly even, extensive spread of land that has been affected up to drought levels of D3. Some even reach D4. This may be likely a sign of a steady increase in climate change and its effects increase yearly.

Lets get a clearer look by removing AK (since it mostly scaled in the D0 category)

```r
state_at_no_AK <- state_summary_at_avg %>%
  filter(state_abb != "AK")
```

```r
state_summary_long_at_no_AK <- state_at_no_AK %>%
pivot_longer(cols = all_of(drought_categories), names_to = "Category", values_to = "Area")

ggplot(state_summary_long_at_no_AK, aes(x = Category, y = state_abb, fill = Area)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "red") +
  labs(title = "Land Area affected (Avg) *excluding AK*",
       x = "Severity Levels",
       y = "State") +
  theme_minimal()
```

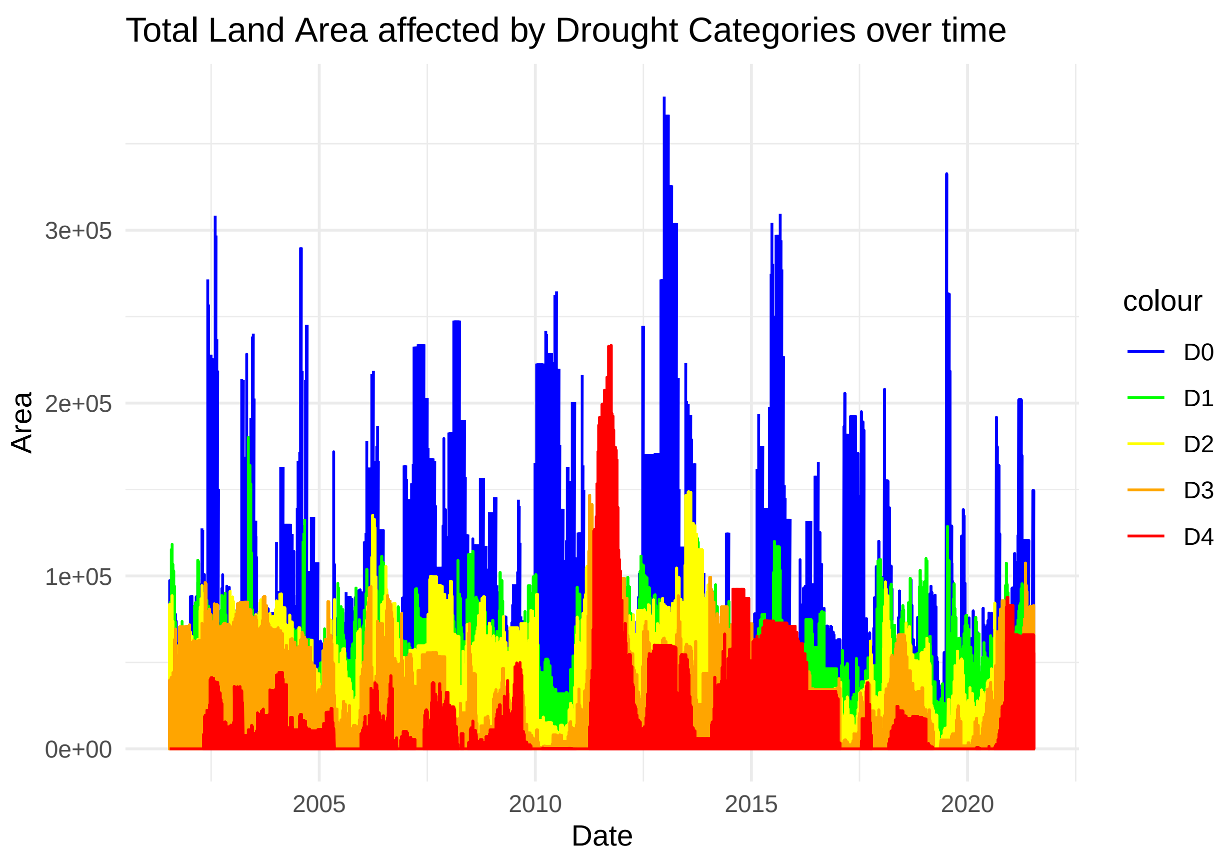


By removing AK we can see a clearer picture of the average land area drought severity level over time. Note that the same states as before are among those who experienced higher drought severity levels over time.

Visualization: total land area affected by the drought severity levels over time

```
drought_area_total$map_date <- as.Date(as.character(drought_area_total$map_date), format="%Y%m%d")
```

```
ggplot(drought_area_total, aes(x=map_date)) +
  geom_line(aes(y=D0, color="D0")) +
  geom_line(aes(y=D1, color="D1")) +
  geom_line(aes(y=D2, color="D2")) +
  geom_line(aes(y=D3, color="D3")) +
  geom_line(aes(y=D4, color="D4")) +
  labs(title="Total Land Area affected by Drought Categories over time",
       x="Date",
       y = "Area") +
  scale_x_date(limits = c(as.Date("2001-07-10"),
                          as.Date("2021-07-30")),
               date_labels = "%Y") +
  scale_color_manual(values = c("blue", "green", "yellow", "orange", "red")) +
  theme_minimal()
```



We may observe there still steady spikes in D0-level droughts, as well as higher amounts of land affected by D4-level droughts. The D4-level droughts's effects more land over time and have much longer durations prior to ~2011. Among each drought category, there is a notable dip around the late 2010s, which is likely when the effects of icecaps and glaciers melting cooled the weather and alleviated land area affected by drought. Then we see a spike and increase in duration again entering 2021, which may represent the cyclical nature of regular seasonal changes or even that of climate change.