01_make_data_.Rmd

Abisai Lujan

2025-01-14

```
library(tidyverse)
```

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

```
library(lubridate)
library(tidytuesdayR)
```

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

Drought data from Github repository on ~ 2 decades worth of weekly recorded drought data across the US and territory Puerto Rico.

```
drought1 <- read.csv("../RData/Drought_data/drought.csv")
head(drought1)
```

```
##   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
```

```
dim(drought1)
```

```
## [1] 325728     10
```

The dataset has 325728 and 10 columns. Information includes drought levels and their effects on different states' land area and population on a weekly basis.

The github repository also split the drought data into different sub data sets for percent of land area affected, total land area affected, percent of population affected, and total population affected. We will manually split the data ourselves, but will load the Github data for reference.

Load drought area percent data

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

```
##    MapDate StateAbbreviation  None    D0 D1 D2 D3 D4 ValidStart   ValidEnd
## 1 20210713                AK 74.35 25.65  0  0  0  0 2021-07-13 2021-07-19
## 2 20210706                AK 74.35 25.65  0  0  0  0 2021-07-06 2021-07-12
## 3 20210629                AK 85.92 14.08  0  0  0  0 2021-06-29 2021-07-05
## 4 20210622                AK 85.92 14.08  0  0  0  0 2021-06-22 2021-06-28
## 5 20210615                AK 85.92 14.08  0  0  0  0 2021-06-15 2021-06-21
## 6 20210608                AK 85.92 14.08  0  0  0  0 2021-06-08 2021-06-14
##   StatisticFormatID
## 1                 2
## 2                 2
## 3                 2
## 4                 2
## 5                 2
## 6                 2
```

```
str(drought_area_pct)
```

```
## 'data.frame':    54288 obs. of  11 variables:
##  $ MapDate          : int  20210713 20210706 20210629 20210622 20210615 20210608 20210601 20210525 20
##  $ StateAbbreviation: chr  "AK" "AK" "AK" "AK" ...
##  $ None             : num  74.3 74.3 85.9 85.9 85.9 ...
##  $ D0               : num  25.6 25.6 14.1 14.1 14.1 ...
##  $ D1               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ D2               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ D3               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ D4               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ ValidStart       : chr  "2021-07-13" "2021-07-06" "2021-06-29" "2021-06-22" ...
##  $ ValidEnd         : chr  "2021-07-19" "2021-07-12" "2021-07-05" "2021-06-28" ...
##  $ StatisticFormatID: int  2 2 2 2 2 2 2 2 2 2 ...
```

Load area total data

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

```
##    MapDate StateAbbreviation       None         D0   D1   D2   D3   D4
## 1 20210713                AK 433,133.18 149,435.11 0.00 0.00 0.00 0.00
## 2 20210706                AK 433,133.18 149,435.11 0.00 0.00 0.00 0.00
## 3 20210629                AK 500,539.66  82,028.63 0.00 0.00 0.00 0.00
## 4 20210622                AK 500,539.66  82,028.63 0.00 0.00 0.00 0.00
## 5 20210615                AK 500,539.66  82,028.63 0.00 0.00 0.00 0.00
## 6 20210608                AK 500,539.66  82,028.63 0.00 0.00 0.00 0.00
##   ValidStart   ValidEnd StatisticFormatID
## 1 2021-07-13 2021-07-19                2
## 2 2021-07-06 2021-07-12                2
## 3 2021-06-29 2021-07-05                2
## 4 2021-06-22 2021-06-28                2
## 5 2021-06-15 2021-06-21                2
## 6 2021-06-08 2021-06-14                2
```

Load drought population percent

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

```
##    MapDate StateAbbreviation  None    D0 D1 D2 D3 D4 ValidStart   ValidEnd
## 1 20210713                AK 33.91 66.09  0  0  0  0 2021-07-13 2021-07-19
## 2 20210706                AK 33.91 66.09  0  0  0  0 2021-07-06 2021-07-12
## 3 20210629                AK 98.96  1.04  0  0  0  0 2021-06-29 2021-07-05
## 4 20210622                AK 98.96  1.04  0  0  0  0 2021-06-22 2021-06-28
## 5 20210615                AK 98.96  1.04  0  0  0  0 2021-06-15 2021-06-21
## 6 20210608                AK 98.96  1.04  0  0  0  0 2021-06-08 2021-06-14
##   StatisticFormatID
## 1                 2
## 2                 2
## 3                 2
## 4                 2
## 5                 2
## 6                 2
```

Load drought population total data

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

```
##    MapDate StateAbbreviation       None         D0   D1   D2   D3   D4
## 1 20210713                AK 240,644.16 468,985.84 0.00 0.00 0.00 0.00
## 2 20210706                AK 240,644.16 468,985.84 0.00 0.00 0.00 0.00
## 3 20210629                AK 702,217.65   7,412.34 0.00 0.00 0.00 0.00
## 4 20210622                AK 702,217.65   7,412.34 0.00 0.00 0.00 0.00
## 5 20210615                AK 702,217.65   7,412.34 0.00 0.00 0.00 0.00
## 6 20210608                AK 702,217.65   7,412.34 0.00 0.00 0.00 0.00
##   ValidStart   ValidEnd StatisticFormatID
## 1 2021-07-13 2021-07-19                2
## 2 2021-07-06 2021-07-12                2
## 3 2021-06-29 2021-07-05                2
## 4 2021-06-22 2021-06-28                2
## 5 2021-06-15 2021-06-21                2
## 6 2021-06-08 2021-06-14                2
```

Our question of interest is if we can identify trends in how different drought levels affect land area and population over time. Our killer graph could be a time series of drought conditions on area and population; also, the killer graph could break the land down by region and identify geographical trends.