01_make_data_.Rmd

Abisai Lujan

2025-01-14

```
library(tidyverse)
```

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

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- 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
```

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

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

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 ...
```

Change `ValidStart` and `ValidEnd` dates to numeric

```
drought_area_pct$ValidStart=as.numeric(gsub("-","",drought_area_pct$ValidStart))
drought_area_pct$ValidEnd=as.numeric(gsub("-","", drought_area_pct$ValidEnd))
```

```
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   20210713 20210719
## 2 20210706                AK 74.35 25.65  0  0  0  0   20210706 20210712
## 3 20210629                AK 85.92 14.08  0  0  0  0   20210629 20210705
## 4 20210622                AK 85.92 14.08  0  0  0  0   20210622 20210628
## 5 20210615                AK 85.92 14.08  0  0  0  0   20210615 20210621
## 6 20210608                AK 85.92 14.08  0  0  0  0   20210608 20210614
##   StatisticFormatID
## 1                 2
## 2                 2
## 3                 2
## 4                 2
## 5                 2
## 6                 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
```

Change `ValidStart` and `ValidEnd` to numeric

```
drought_area_total$ValidStart=as.numeric(gsub("-","",drought_area_total$ValidStart))
drought_area_total$ValidEnd=as.numeric(gsub("-","", drought_area_total$ValidEnd))
```

```
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   20210713 20210719                2
## 2   20210706 20210712                2
## 3   20210629 20210705                2
## 4   20210622 20210628                2
## 5   20210615 20210621                2
## 6   20210608 20210614                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
```

Change `ValidStart` and `ValidEnd` dates to numeric

```
drought_pop_pct$ValidStart=as.numeric(gsub("-","",drought_pop_pct$ValidStart))
drought_pop_pct$ValidEnd=as.numeric(gsub("-","",drought_pop_pct$ValidEnd))
```

```
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   20210713 20210719
## 2 20210706                AK 33.91 66.09  0  0  0  0   20210706 20210712
## 3 20210629                AK 98.96  1.04  0  0  0  0   20210629 20210705
## 4 20210622                AK 98.96  1.04  0  0  0  0   20210622 20210628
## 5 20210615                AK 98.96  1.04  0  0  0  0   20210615 20210621
## 6 20210608                AK 98.96  1.04  0  0  0  0   20210608 20210614
##   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
```

Change `ValidStart` and `ValidEnd` dates to numeric

```
drought_pop_total$ValidStart=as.numeric(gsub("-","",drought_pop_total$ValidStart))
drought_pop_total$ValidEnd=as.numeric(gsub("-","",drought_pop_total$ValidEnd))
```

```
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   20210713 20210719                 2
## 2   20210706 20210712                 2
## 3   20210629 20210705                 2
## 4   20210622 20210628                 2
## 5   20210615 20210621                 2
## 6   20210608 20210614                 2
```

"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."

Data set on U.S. Droughts from TidyTuesday Github data. Source: U.S. Drought Monitor which describes drought levels from 2001 to 2021 among each state.

Project: Analyzing drought conditions across the U.S. over ~20 years. State