

State Cancer Profiles Data Prep and a Linked Micromap

Brandon Tao

03/25/2019

0. Setup

```
library(tidyverse)

## -- Attaching packages -----
## ----- tidyverse 1.2.1 -----

## v ggplot2 3.1.0      v purrr 0.3.0
## v tibble 2.0.1       v dplyr 0.8.0.1
## v tidyr 0.8.2        v stringr 1.4.0
## v readr 1.3.1        v forcats 0.4.0

## Warning: package 'ggplot2' was built under R version 3.5.3
## Warning: package 'tibble' was built under R version 3.5.3

## -- Conflicts -----
## ----- tidyverse_conflicts() -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(micromapST)

## Warning: package 'micromapST' was built under R version 3.5.3
```

1. Look at the csv file, note issues and address them in R.

```
dat <- read_csv(file = 'rtcancer.csv',
  skip = 9, na = c(' ', "NA", "***", "-"), n_max = 53)

## Warning: Duplicated column names deduplicated: 'Lower 95% Confidence
## Interval' => 'Lower 95% Confidence Interval_1' [12], 'Upper 95%
## Confidence
## Interval' => 'Upper 95% Confidence Interval_1' [13]

## Parsed with column specification:
## cols(
##   States = col_character(),
##   FIPS = col_character(),
##   `Priority Index(1) 1=highest 9=lowest` = col_double(),
##   `Recent Trend(2)` = col_character(),
##   `State Death Rate Compared to US Rate` = col_character(),
##   `Average Annual Count` = col_double(),
##   `Age-Adjusted Death Rate - deaths per 100,000` = col_double(),
##   `Lower 95% Confidence Interval` = col_double(),
##   `Upper 95% Confidence Interval` = col_double(),
```

```

## `Rate Ratio(3) State to US` = col_double(),
## `Recent 5-Year Trend (2) in Death Rates` = col_double(),
## `Lower 95% Confidence Interval_1` = col_double(),
## `Upper 95% Confidence Interval_1` = col_double()
## )

dat2 <- as.data.frame(dat[, -c(2:6, 10)]) # Remove unused columns
colnames(dat2) # See the columns names left

## [1] "States"
## [2] "Age-Adjusted Death Rate - deaths per 100,000"
## [3] "Lower 95% Confidence Interval"
## [4] "Upper 95% Confidence Interval"
## [5] "Recent 5-Year Trend (2) in Death Rates"
## [6] "Lower 95% Confidence Interval_1"
## [7] "Upper 95% Confidence Interval_1"

dat2[, 1]

## [1] "United States" "District of Columbia" "Hawaii (8)"
## [4] "Louisiana" "Mississippi" "New Mexico"
## [7] "Rhode Island" "Texas" "Alabama"
## [10] "Alaska" "Arizona" "Arkansas"
## [13] "Connecticut" "Delaware" "Florida"
## [16] "Georgia" "Illinois" "Kentucky"
## [19] "Maryland" "Massachusetts" "Missouri"
## [22] "New Jersey" "New York" "North Carolina"
## [25] "Ohio" "Oklahoma" "Oregon"
## [28] "Pennsylvania" "South Carolina" "Tennessee"
## [31] "Virginia" "Washington" "Colorado"
## [34] "Idaho" "Indiana" "Iowa"
## [37] "Kansas" "Maine" "Michigan"
## [40] "Minnesota" "Montana" "Nebraska"
## [43] "Nevada" "New Hampshire" "South Dakota"
## [46] "Utah" "Vermont" "West Virginia"
## [49] "Wisconsin" "California" "Puerto Rico (8)"
## [52] "North Dakota" "Wyoming"

USrate <- dat2[1, 2] # Save for use as a reference value
UStrend <- dat2[1, 5] # Save for use as a reference value
dat2[3, 1] <- "Hawaii" # Remove the code (8) after Hawaii
dat3 <- dat2[-c(1, 51), ] # Omit US and Puerto Rico rows

```

2. Make a panelDescription dataframe.

```

panelDesc <- data.frame(
  type = c('mapcum', 'id', 'dotconf', 'dotconf'),
  col1 = c(NA, NA, 2, 5),
  col2 = c(NA, NA, 3, 6),
  col3 = c(NA, NA, 4, 7),
  lab1 = c(NA, NA, 'Age-Adjusted', '5-Year Average Annual'),
  lab2 = c(NA, NA, 'Death Rate', 'Percent Change'),

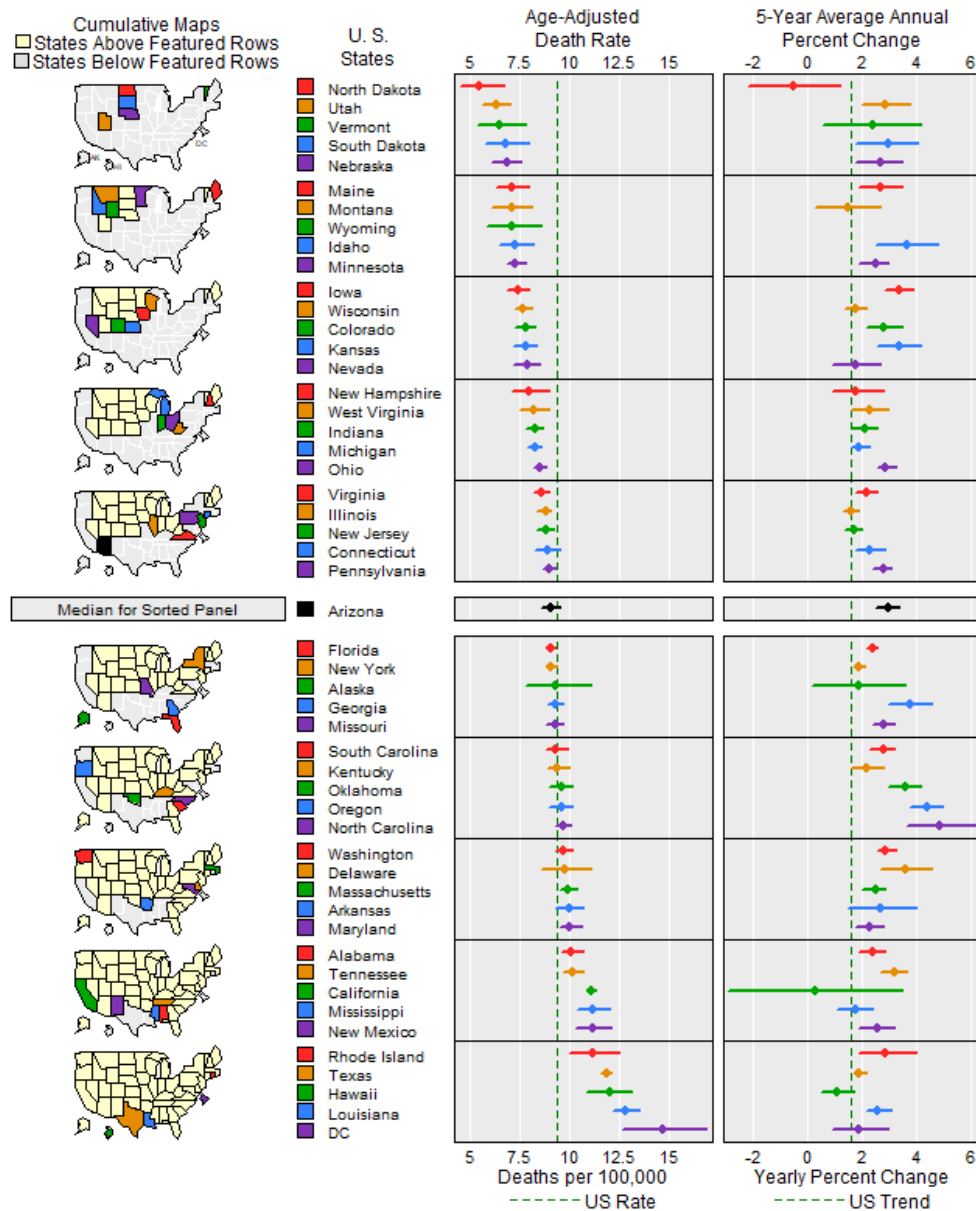
```

```
lab3 = c(NA,NA,'Deaths per 100,000','Yearly Percent Change'),  
refVals = c(NA,NA,USrate,USTrend),  
refTexts = c(NA,NA,'US Rate','US Trend')  
)
```

3. Produce the Status sorted plot.

```
micromapST(dat3,panelDesc,  
  rowNames = 'full', rowNamesCol = 1,  
  title = c('Liver and Bile Duct Cancer',  
    '2015 Male Mortality Rates'),  
  plotNames = 'full',  
  sortVar = 2)
```

Liver and Bile Duct Cancer 2015 Male Mortality Rates



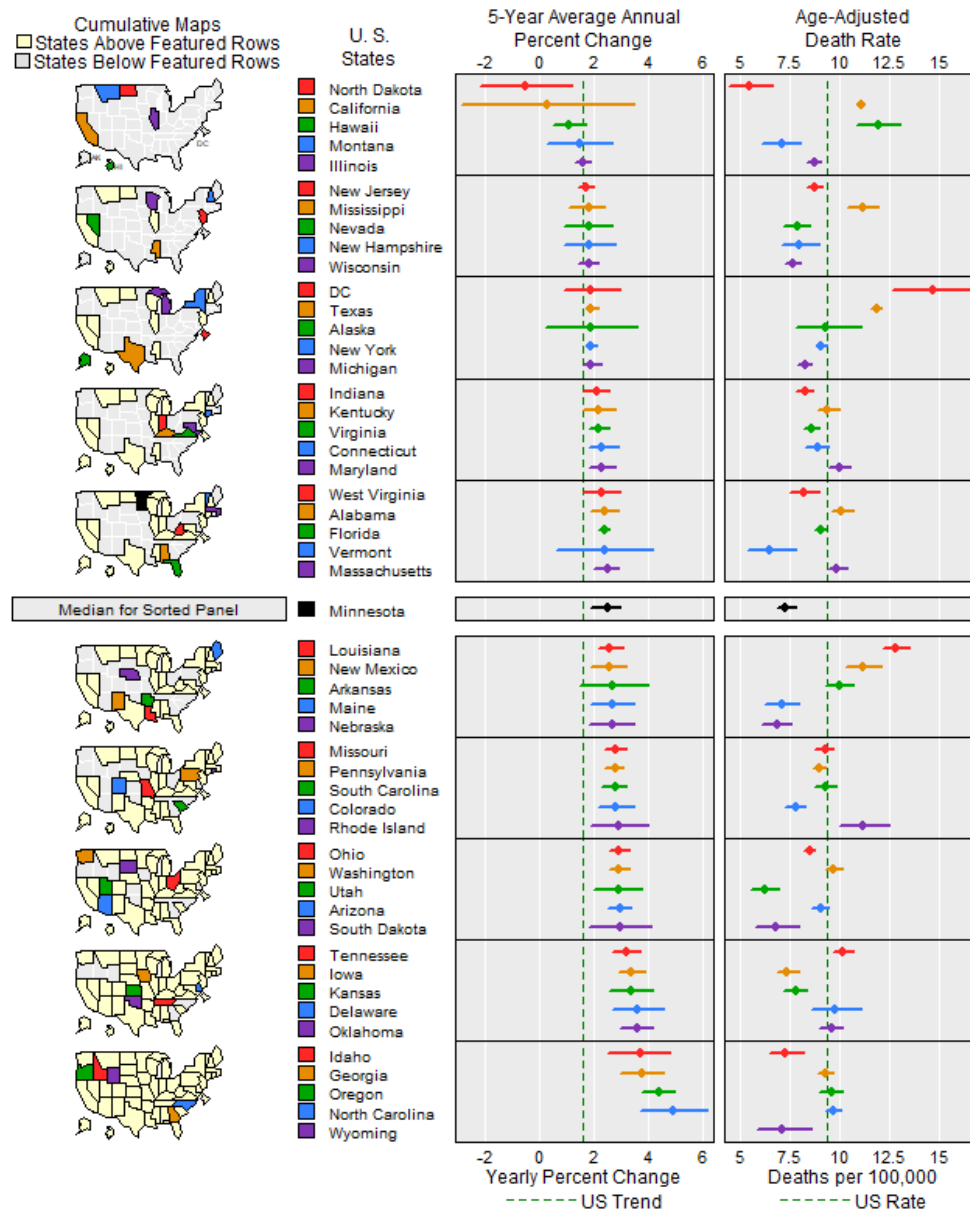
```
## [1] "micromapST Ends"
```

There is one big northern cluster on the left has low mortality rate, and another big cluster down the country's deep south that suffers the most from this type of cancer.

4. Produce a Trend sorted plot with the trend column first.

```
micromapST(dat3,panelDesc[c(1,2,4,3)],,
  rowNames = 'full', rowNamesCol = 1,
  title = c('Liver and Bile Duct Cancer',
    '2015 Male Mortality Rates'),
  plotNames = 'full',
  sortVar = 5)
```

Liver and Bile Duct Cancer 2015 Male Mortality Rates



```
## [1] "micromapST Ends"
```

The States with missing data sort to the bottom.

The mortality rate of liver and bile duct cancer for U.S. Male is overall increasing, however there's not obvious pattern on location distribution.