state_regression.R

wanchuangzhu

2020-06-22

```
#####################################################prepare the data
us=read.csv('../raw-data/us_result.csv')

library(dplyr)
```

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

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

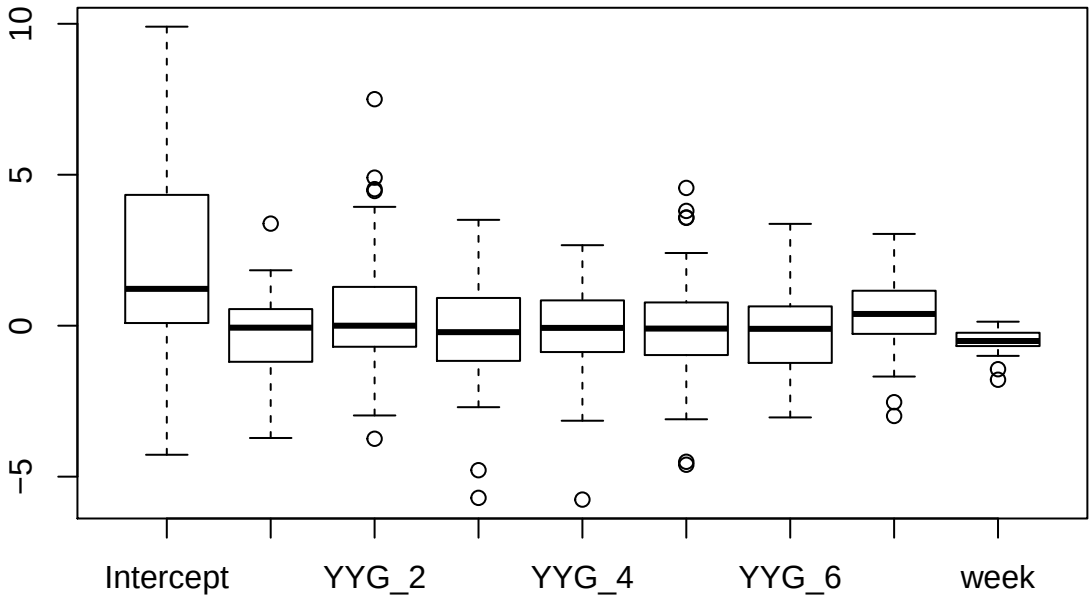
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
models= c("YYG") # YYG and IHME predict USA and also states, Geneva only predicts USA
ahead=7 # decide how many days are included in the regression
us=filter(us,model_name %in% models,location_short != "USA",lookahead<=ahead)

us=group_by(us,target_end_date,forecast_date,model_name,gt_source,location_short)
us=summarise(us,expected_value=(expected_value),gt=(gt),lookahead=(lookahead))




us_wide=pivot_wider(us,names_from = c("model_name"), values_from = expected_value) %>% filter(.,as.Date

#for(i in 1:nrow(us_wide)){
#  temp=filter(us_wide,forecast_date==us_wide$forecast_date[i],target_end_date==us_wide$target_end_date
#  us_wide[i,models[1]]=mean(as.matrix(temp[,models[1]]),na.rm = T)
#  us_wide[i,models[2]]=mean(as.matrix(temp[,models[2]]),na.rm = T)

#}

us_wide=ungroup(filter(us_wide,gt_source=="JHU")) %>% dplyr::select(.,-forecast_date,-gt_source)

us_wide=pivot_wider(us_wide,names_from = lookahead,values_from = "YYG",names_prefix = "YYG_")
us_wide=drop_na(us_wide) %>% mutate(.,week=as.numeric(weekdays(as.Date(target_end_date)) %in% c("Saturda
us_wide$target_end_date=as.Date(us_wide$target_end_date)
us_wide=us_wide[order(us_wide$target_end_date),]
## take the data into log-scale
us_wide_log=(us_wide)
```

```r
us_wide_log[,c(3:10)]=log(us_wide_log[,c(3:10)]+1) # deal with value of zero's


train.num=36 # split training and prediction set

## regression state-wise
states=as.character(unique(us_wide_log$location_short))
coef.mat=matrix(NA,nrow = length(states),ncol=ncol(us_wide_log)-2)
colnames(coef.mat)=c("Intercept",names(us_wide_log)[4:11])
for(i in 1:length(states)){
  datatemp=filter(us_wide_log,location_short==states[i])
  model=lm(gt~. ,data=datatemp[1:train.num,-c(1,2)])
  coef.mat[i,]=model$coefficients
}

coef.mat=as_tibble(coef.mat) %>% mutate(.,state=states)

boxplot(coef.mat[,1:9])
coef.mat
```

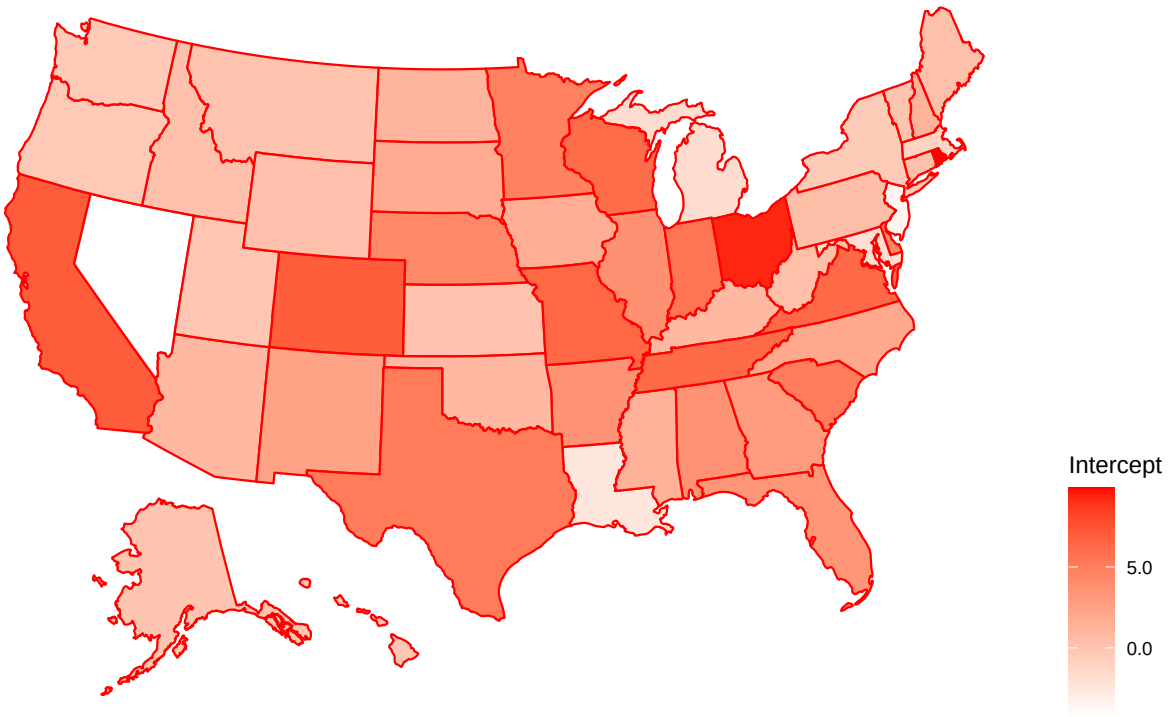
```
## # A tibble: 51 x 10
##    Intercept   YYG_1  YYG_2  YYG_3     YYG_4   YYG_5  YYG_6     YYG_7    week
##        <dbl>   <dbl>  <dbl>  <dbl>     <dbl>   <dbl>  <dbl>     <dbl>   <dbl>
##  1    0.0385 -0.0556 NA     NA     -3.66e-16 -0.0556 NA     -4.57e-17 -0.0385
##  2    3.63    1.39   -1.51  -0.665 -9.99e- 1  4.56   -3.04  -1.50e- 1 -0.905
##  3    3.57   -0.892   0.572 -0.795  1.91e- 1  0.648   0.549 -2.53e+ 0  0.109
##  4    0.865  -0.234   2.80   1.11  -1.00e+ 0 -4.51    2.24   3.89e- 1 -1.79
##  5    7.01    1.84   -2.00   0.270  1.46e+ 0 -3.10    1.38  -4.43e- 1 -0.726
##  6    6.91    1.44   -2.97   3.51   2.45e- 1 -4.61   -1.70   3.04e+ 0 -1.44
##  7    0.412  -1.44    4.52  -1.78  -1.36e+ 0  1.26   -2.42   2.15e+ 0 -0.762
##  8    1.03    1.30    1.29  -2.27   2.67e+ 0 -2.19   -0.181 -1.08e- 1 -0.0758
##  9    4.37   -0.171  -0.486  0.683  9.44e- 1 -1.97   -1.23   1.16e+ 0 -0.350
## 10    3.36   -0.0743  0.335 -1.54   6.93e- 1 -0.630   0.306  1.06e+ 0 -0.941
## # ... with 41 more rows, and 1 more variable: state <chr>
```

```r
library(usmap)
library(ggplot2)
```

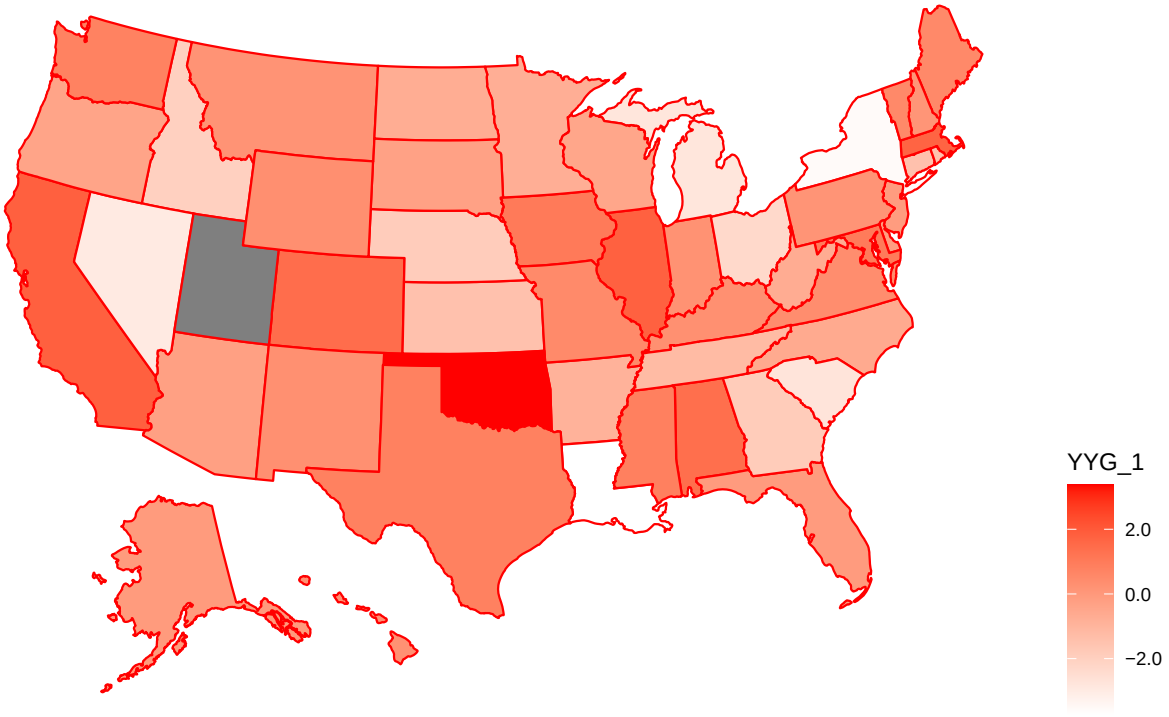



```
library(maps)

for(i in 1:9){
  p=plot_usmap(data = coef.mat, values = colnames(coef.mat)[i], color = "red") +
    scale_fill_continuous(low='white', high= 'red',name = colnames(coef.mat)[i], label = scales::comma)
    theme(legend.position = "right")
  print(p)
}
```

```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.

## Warning: Use of `map_df$y` is discouraged. Use `y` instead.

## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```

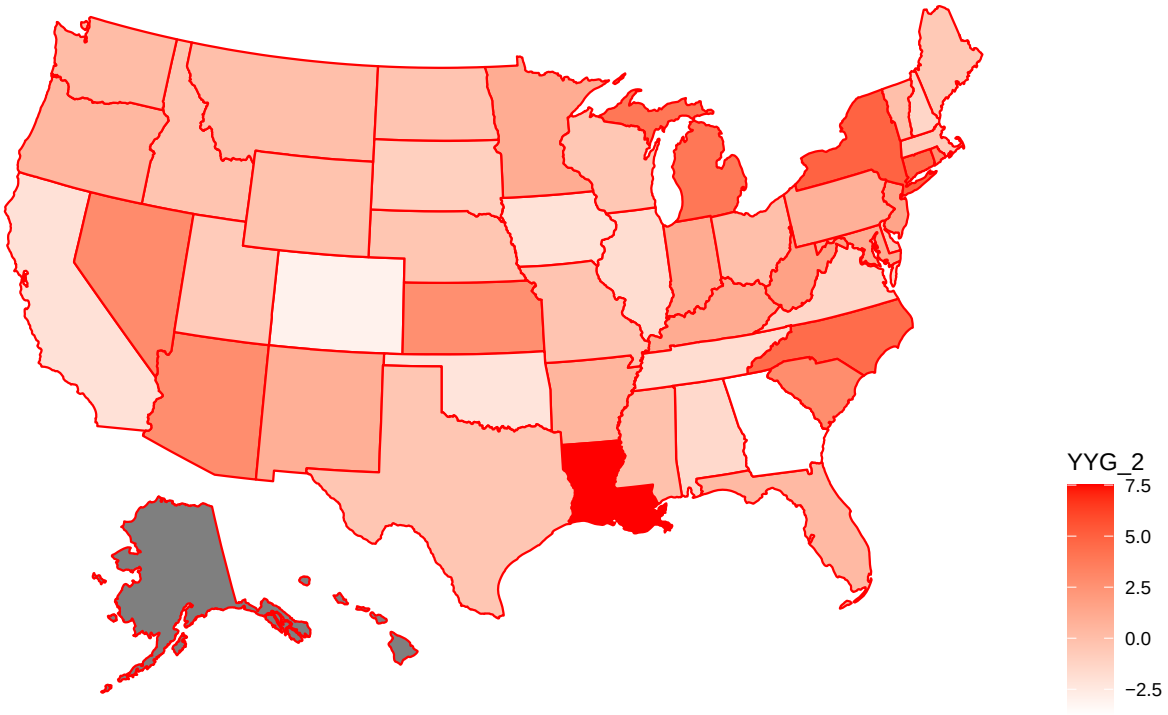



```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.
## Warning: Use of `map_df$y` is discouraged. Use `y` instead.
## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```

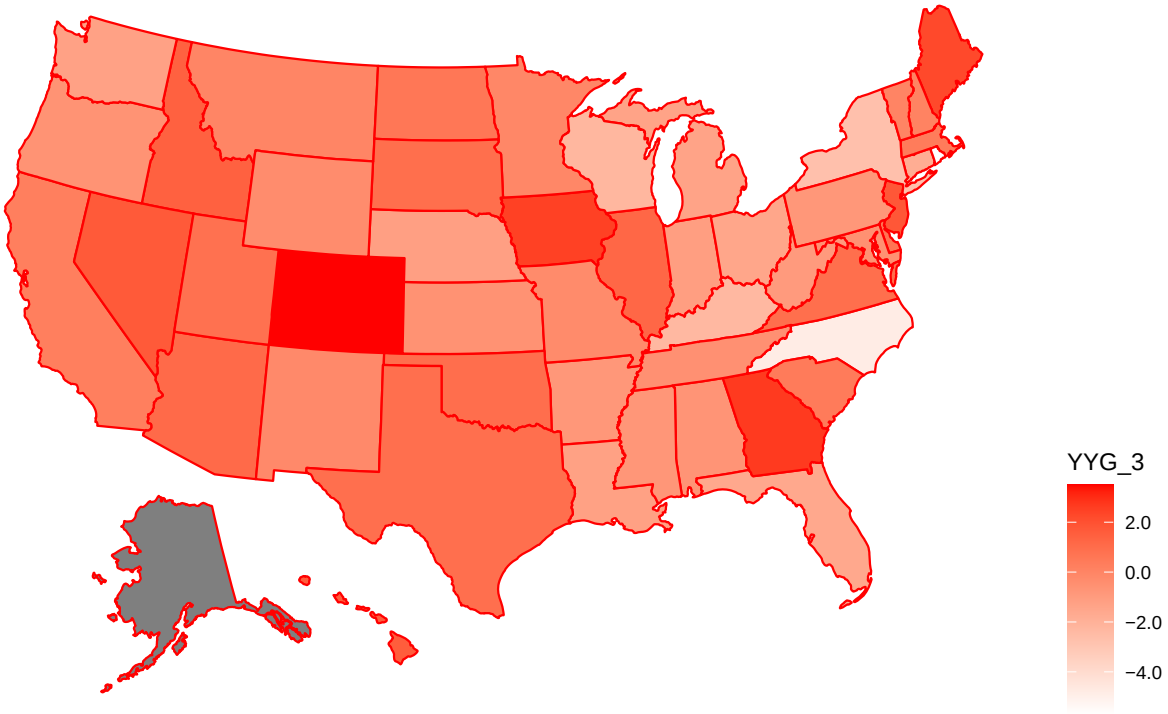



```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.
## Warning: Use of `map_df$y` is discouraged. Use `y` instead.
```

```
## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```

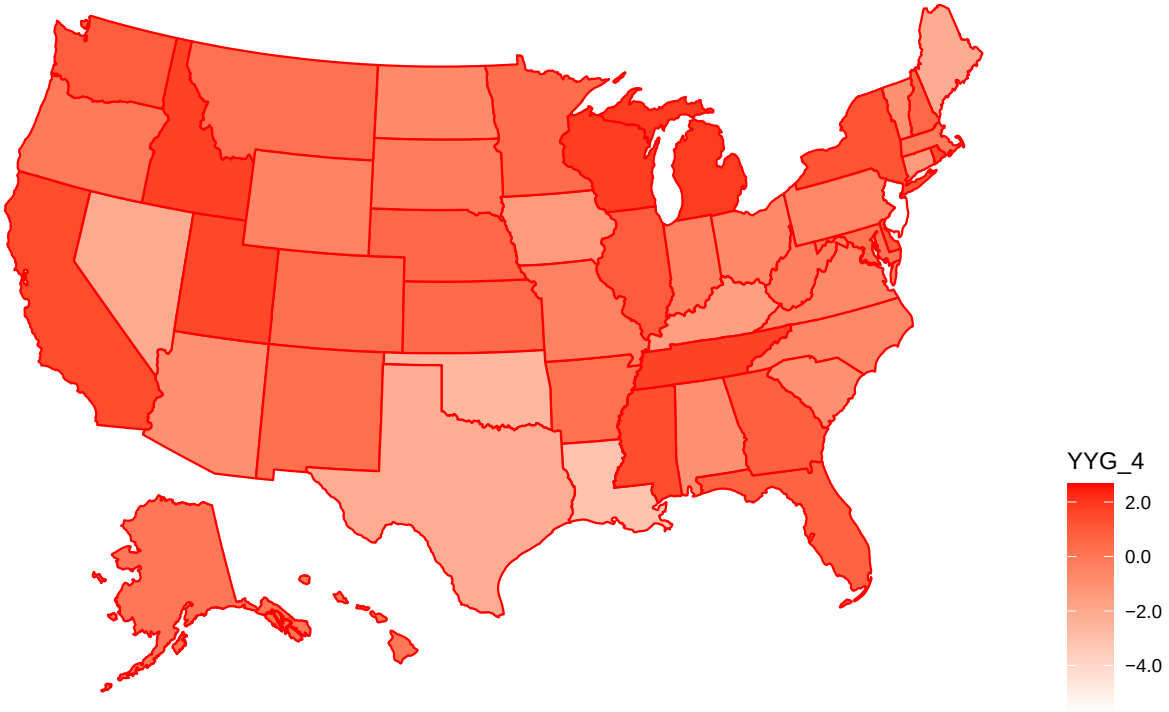



```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.
## Warning: Use of `map_df$y` is discouraged. Use `y` instead.
## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```

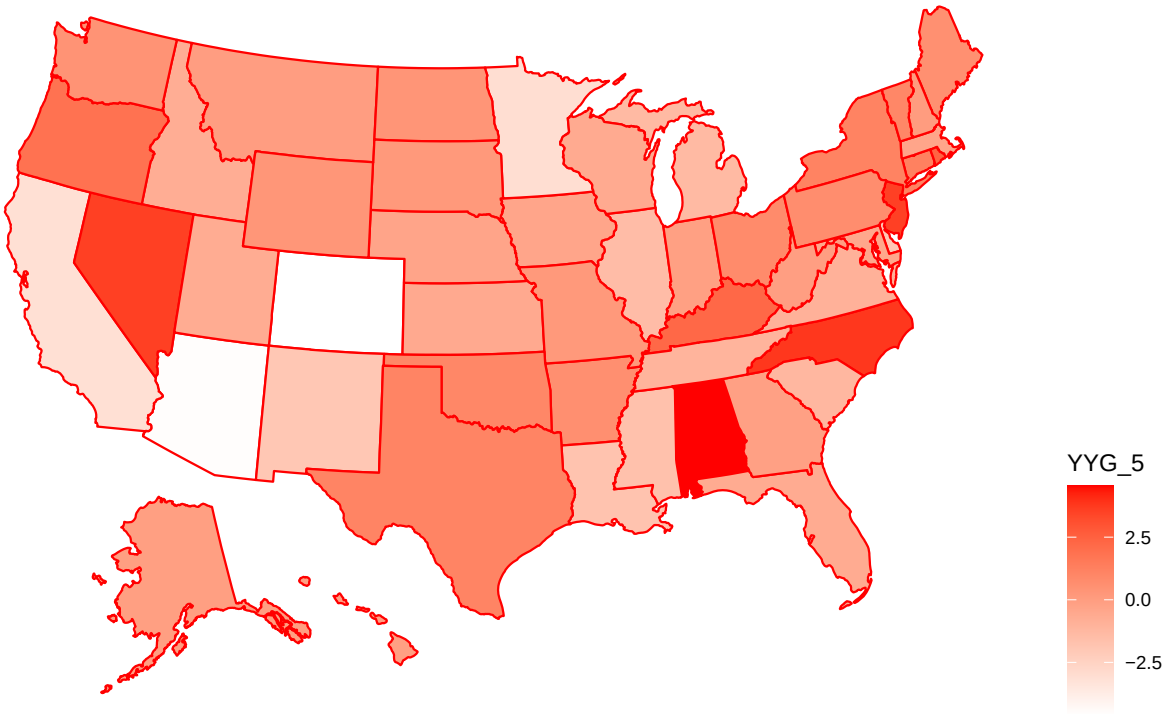



```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.
```

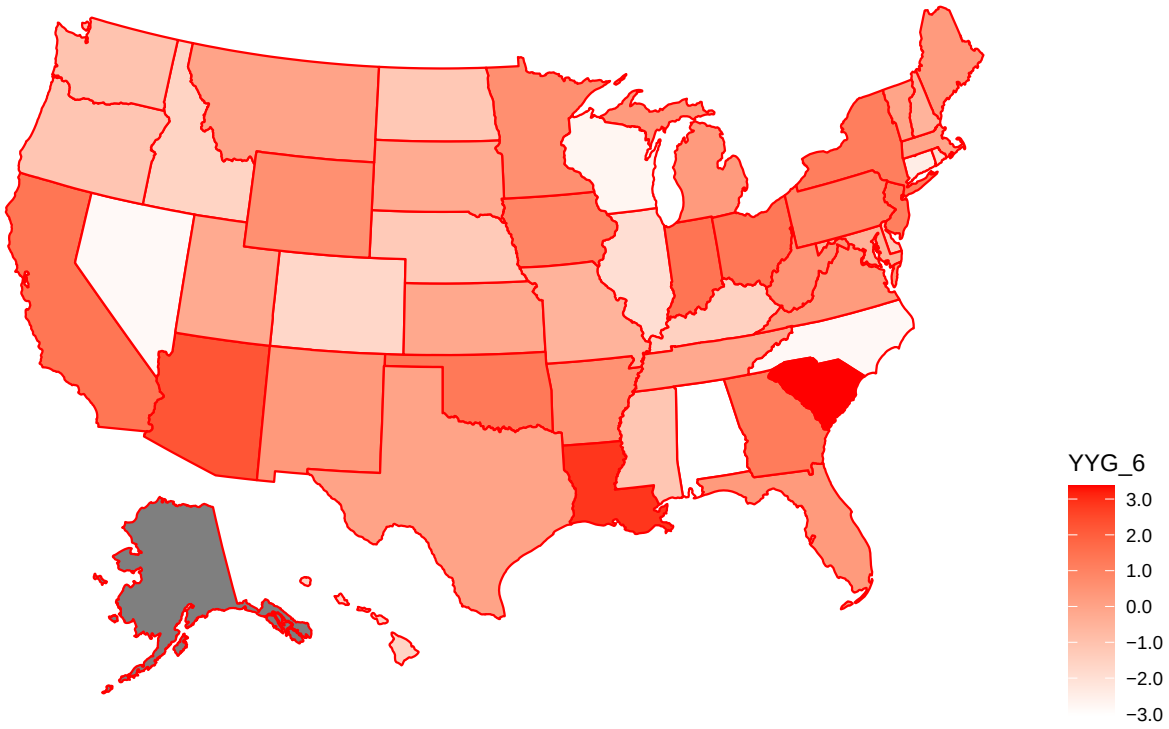
```
## Warning: Use of `map_df$y` is discouraged. Use `y` instead.
## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```



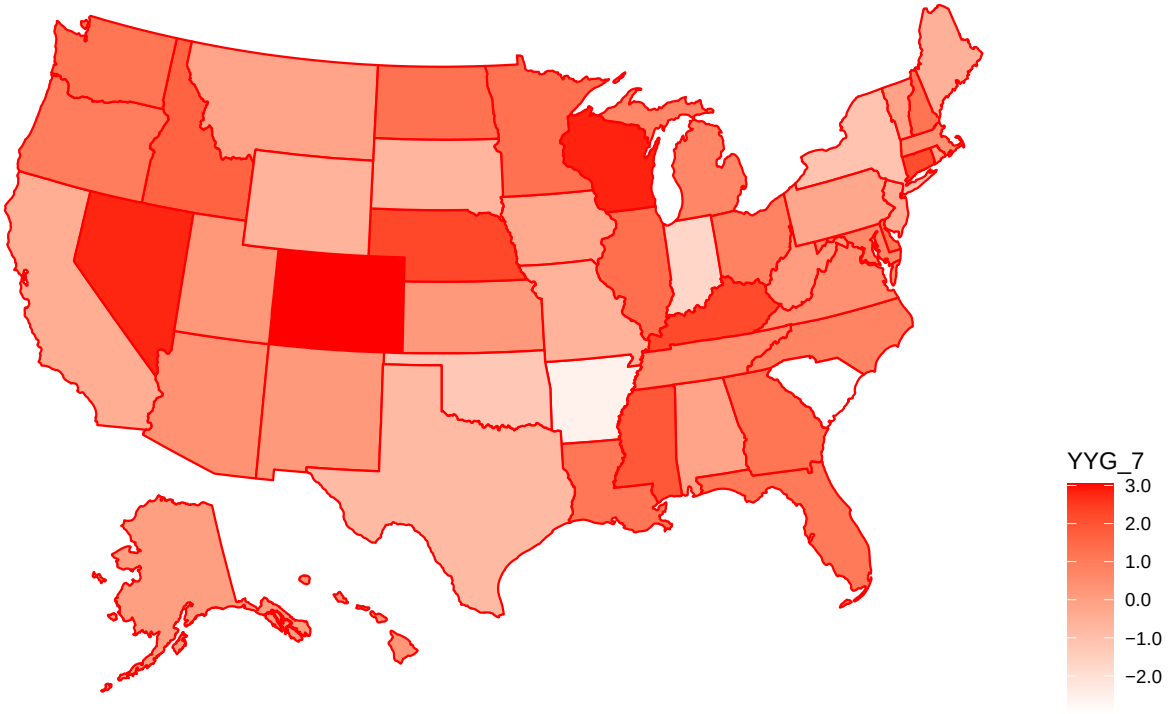

```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.
## Warning: Use of `map_df$y` is discouraged. Use `y` instead.
## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```



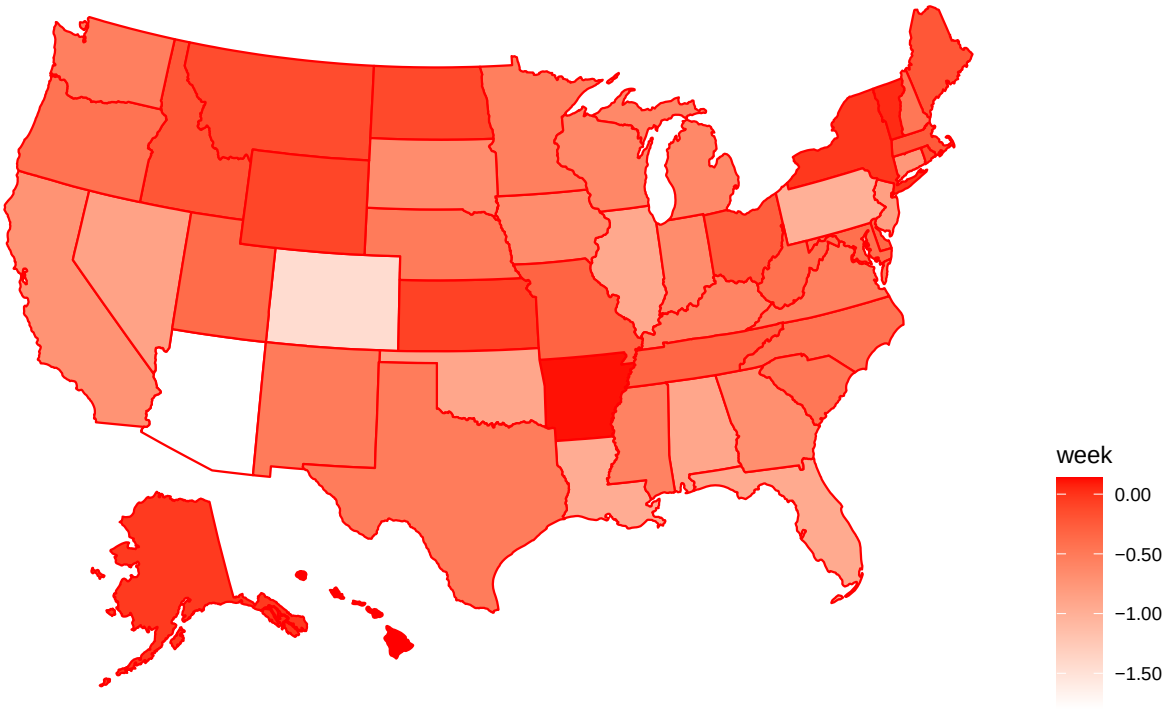
```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.
## Warning: Use of `map_df$y` is discouraged. Use `y` instead.
## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```




```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.
## Warning: Use of `map_df$y` is discouraged. Use `y` instead.
## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```




```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.
## Warning: Use of `map_df$y` is discouraged. Use `y` instead.
## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```




```
# https://remiller1450.github.io/s230s19/Intro_maps.html
```