

12/13/22 TidyTuesday

Emi

2022-12-14

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.7      v dplyr  1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

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

library(ggplot2)
library(dplyr)
library(maps)

##
## Attaching package: 'maps'

## The following object is masked from 'package:purrr':
##
##      map
```

Downloading Datasets

```
tuesdata <- tidyttuesdayR::tt_load('2022-12-13')

## Only 10 Github queries remaining until 2022-12-21 01:40:29 AM PST.
## Only 10 Github queries remaining until 2022-12-21 01:40:29 AM PST.
## Only 10 Github queries remaining until 2022-12-21 01:40:29 AM PST.
## Only 10 Github queries remaining until 2022-12-21 01:40:29 AM PST.

## Only 10 Github queries remaining until 2022-12-21 01:40:30 AM PST.

## --- Compiling #TidyTuesday Information for 2022-12-13 ----
```

```

## Only 9 Github queries remaining until 2022-12-21 01:40:30 AM PST.

## --- There are 2 files available ---

## Only 8 Github queries remaining until 2022-12-21 01:40:30 AM PST.

## --- Starting Download ---

## Only 8 Github queries remaining until 2022-12-21 01:40:30 AM PST.

## Downloading file 1 of 2: 'state_retail.csv'

## Only 7 Github queries remaining until 2022-12-21 01:40:29 AM PST.

## Downloading file 2 of 2: 'coverage_codes.csv'

## Only 6 Github queries remaining until 2022-12-21 01:40:30 AM PST.

## --- Download complete ---

tuesdata <- tidyTuesdayR::tt_load(2022, week = 50)

## Only 5 Github queries remaining until 2022-12-21 01:40:29 AM PST.

## Only 5 Github queries remaining until 2022-12-21 01:40:29 AM PST.

## --- Compiling #TidyTuesday Information for 2022-12-13 ----

## Only 5 Github queries remaining until 2022-12-21 01:40:29 AM PST.

## --- There are 2 files available ---

## Only 4 Github queries remaining until 2022-12-21 01:40:29 AM PST.

## --- Starting Download ---

## Only 4 Github queries remaining until 2022-12-21 01:40:29 AM PST.

## Downloading file 1 of 2: 'state_retail.csv'

## Only 3 Github queries remaining until 2022-12-21 01:40:30 AM PST.

## Downloading file 2 of 2: 'coverage_codes.csv'

## Only 2 Github queries remaining until 2022-12-21 01:40:30 AM PST.

## --- Download complete ---

```

```
state_retail <- tuesdata$state_retail
coverage_codes <- tuesdata$coverage_codes
```

Monthly State Retail Sales: This data comes from the United States Census Bureau's Monthly State Retail Sales - Census Bureau's new experimental data product featuring modeled state-level retail sales

EDA

```
glimpse(state_retail)
```

```
## Rows: 27,456
## Columns: 9
## $ fips      <chr> "00", "00", "00", "00", "00", "00", "00", "00", "00", "0~
## $ state_abbr <chr> "USA", "USA", "USA", "USA", "USA", "USA", "USA", "USA", ~
## $ naics      <dbl> 441, 441, 441, 441, 441, 441, 441, 441, 441, 441, 4~
## $ subsector  <chr> "Motor vehicle and parts dealers", "Motor vehicle and pa~
## $ year       <dbl> 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019, 20~
## $ month      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 4, 5, 6,~
## $ change_yoy <chr> "0.0", "0.4", "0.9", "4.3", "2.6", "0.0", "5.3", "5.5", ~
## $ change_yoy_se <chr> "0.6", "0.6", "0.6", "0.6", "0.6", "0.6", "0.6", "0.6", ~
## $ coverage_code <chr> "A", "A", "A", "A", "A", "A", "A", "A", "A", "A", "A", "~
```

```
state_retail %>%
  group_by(state_abbr)
```

```
## # A tibble: 27,456 x 9
## # Groups:   state_abbr [52]
##   fips state_abbr naics subsector      year month change_yoy change_yoy_se
##   <chr> <chr>      <dbl> <chr>      <dbl> <dbl> <chr>      <chr>
## 1 00 USA      441 Motor vehicle an~ 2019 1 0.0 0.6
## 2 00 USA      441 Motor vehicle an~ 2019 2 0.4 0.6
## 3 00 USA      441 Motor vehicle an~ 2019 3 0.9 0.6
## 4 00 USA      441 Motor vehicle an~ 2019 4 4.3 0.6
## 5 00 USA      441 Motor vehicle an~ 2019 5 2.6 0.6
## 6 00 USA      441 Motor vehicle an~ 2019 6 0.0 0.6
## 7 00 USA      441 Motor vehicle an~ 2019 7 5.3 0.6
## 8 00 USA      441 Motor vehicle an~ 2019 8 5.5 0.6
## 9 00 USA      441 Motor vehicle an~ 2019 9 3.3 0.6
## 10 00 USA      441 Motor vehicle an~ 2019 10 4.0 0.6
## # ... with 27,446 more rows, and 1 more variable: coverage_code <chr>
```

```
table(state_retail$state_abbr)
```

```
##
## AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA
## 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528
## MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI
## 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528 528
## SC SD TN TX USA UT VA VT WA WI WV WY
## 528 528 528 528 528 528 528 528 528 528 528 528
```

```
table(state_retail$month)
```

```
##
##      1      2      3      4      5      6      7      8      9     10     11     12
## 2496 2496 2496 2496 2496 2496 2496 2496 1872 1872 1872 1872
```

```
table(state_retail$subsector)
```

```
##
## Building Materials and Supplies Dealers      Clothing and Clothing Accessories
##                                     2288                                     2288
##           Electronics and Appliances                                     Food and Beverage
##                                     2288                                     2288
##           Furniture and Home Furnishing      Gasoline Stations
##                                     2288                                     2288
##           General Merchandise      Health and Personal Care
##                                     2288                                     2288
##           Miscellaneous Store Retailers      Motor vehicle and parts dealers
##                                     2288                                     2288
##           Sporting Goods and Hobby      total
##                                     2288                                     2288
```

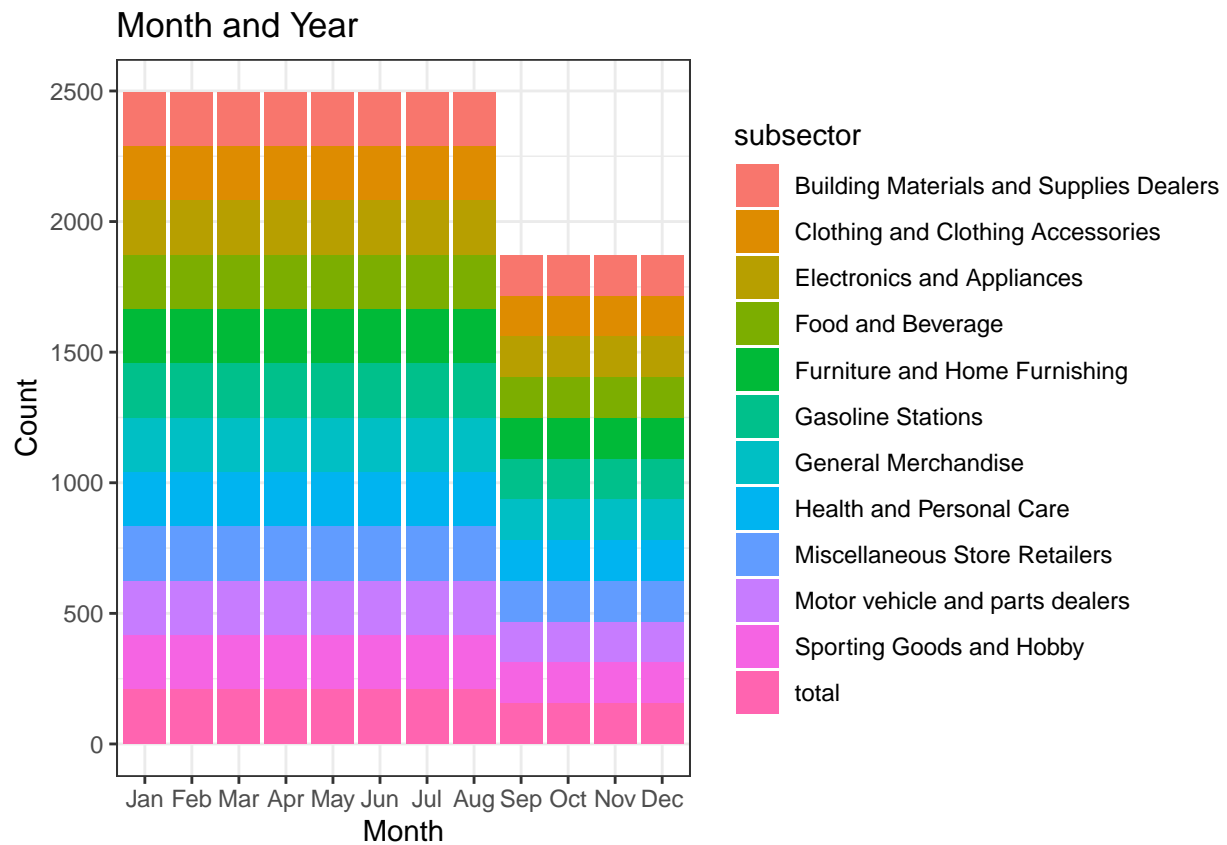
```
table(state_retail$year)
```

```
##
## 2019 2020 2021 2022
## 7488 7488 7488 4992
```

```
coverage_codes[2]
```

```
## # A tibble: 5 x 1
##   coverage
##   <chr>
## 1 non-imputed coverage is less than 10% of the state/NAICS total.
## 2 non-imputed coverage is greater than or equal to 10% and less than 25% of the~
## 3 non-imputed coverage is greater than or equal to 25% and less than 50% of the~
## 4 non-imputed coverage is greater than or equal to 50% of the state/NAICS total.
## 5 Suppressed due to data quality concerns
```

```
ggplot(data = state_retail,
       aes(x = month,
           fill = subsector)) +
  geom_bar() +
  theme_bw() +
  scale_x_discrete("Month", limits = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")) +
  labs(title = "Month and Year", y = "Count")
```

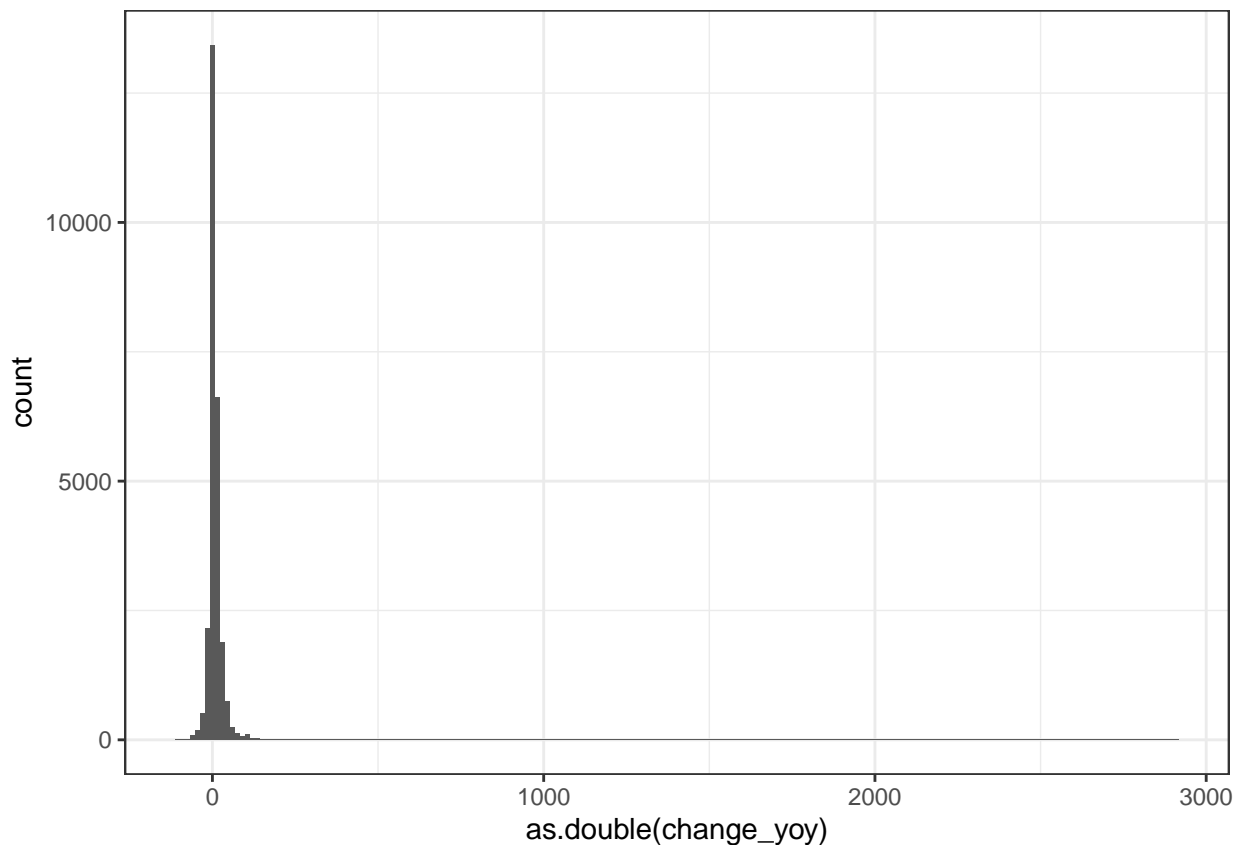


```
ggplot(data = state_retail,
       aes(x = as.double(change_yoy))) +
  geom_histogram(binwidth = 15)+
  theme_bw()
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
## Warning: Removed 1163 rows containing non-finite values (stat_bin).
```



```

summarize(state_retail,
  mean(is.na(as.double(change_yoy))),
  median(is.na(as.double(change_yoy))),
  min(is.na(as.double(change_yoy))),
  max(is.na(as.double(change_yoy)))
)

## Warning in mean(is.na(as.double(change_yoy))): NAs introduced by coercion
## Warning in median(is.na(as.double(change_yoy))): NAs introduced by coercion
## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion
## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## # A tibble: 1 x 4
##   'mean(is.na(as.double(change_yoy)))' 'median(is.na(as.double(change_yoy)))' 'min(is.na(as.double(change_yoy)))' 'max(is.na(as.double(change_yoy)))'
##   <dbl>                                <dbl>                                <int>                                <int>
## 1      0.0424                          0                                  0                                  1

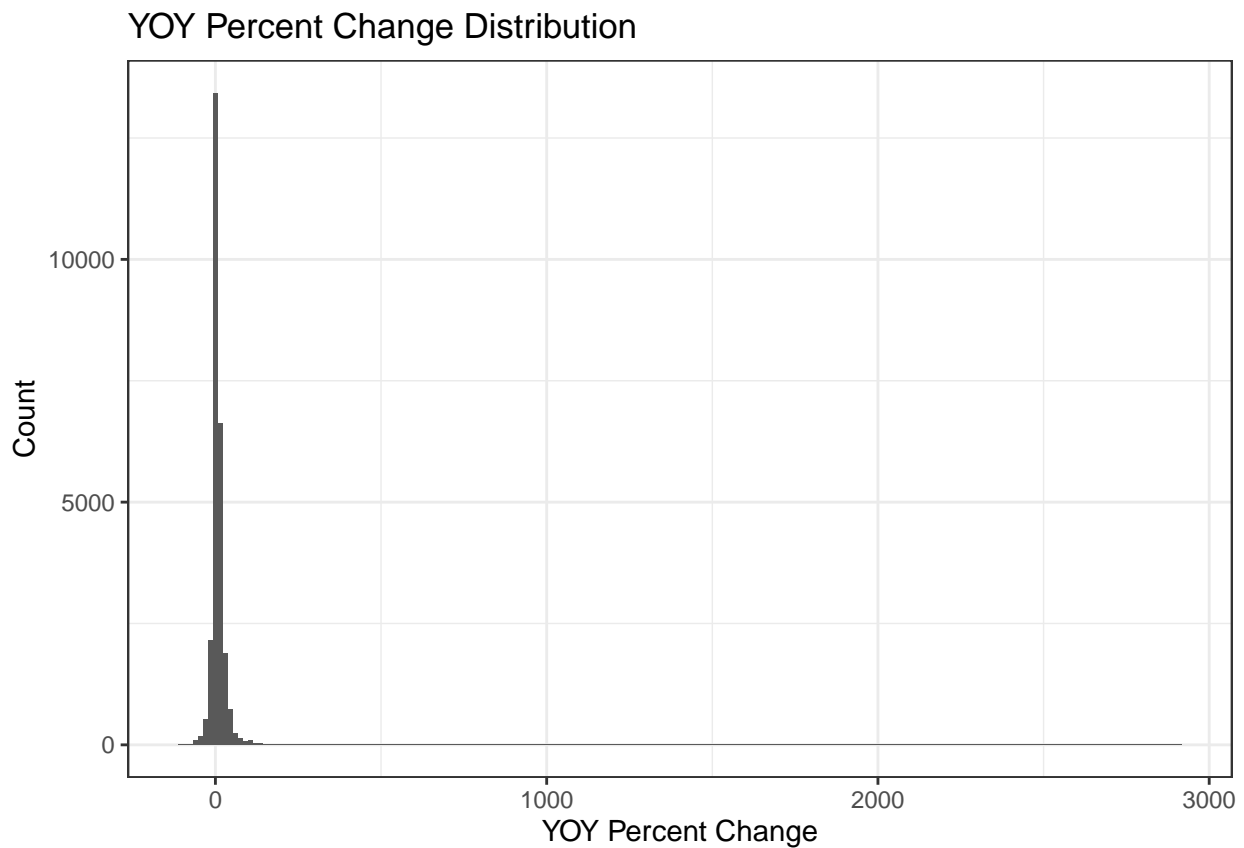
df <- state_retail %>%
  filter(change_yoy != "S") %>%
  mutate(change_yoy = as.numeric(change_yoy))

```

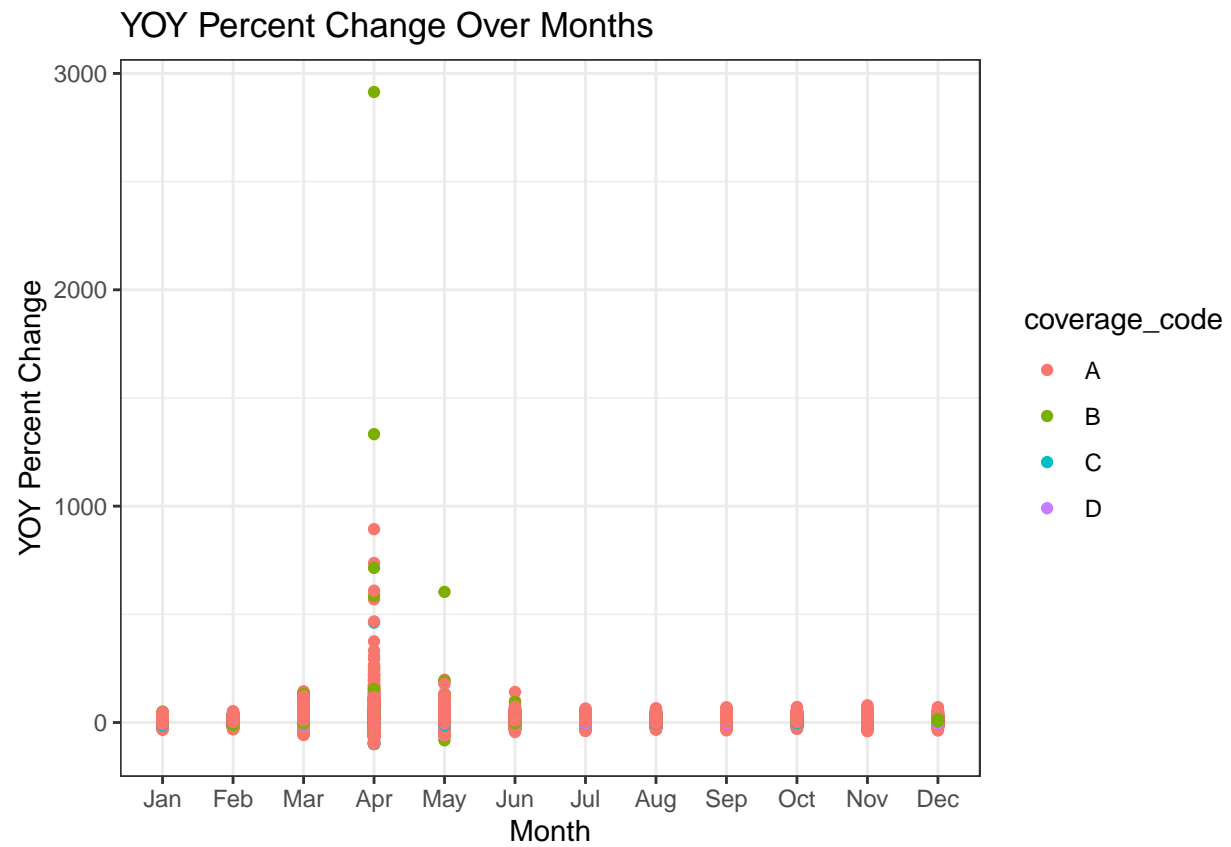
```
summary(df$change_yoy)
```

```
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
## -99.300  -1.000    4.300    7.632   12.600 2914.100
```

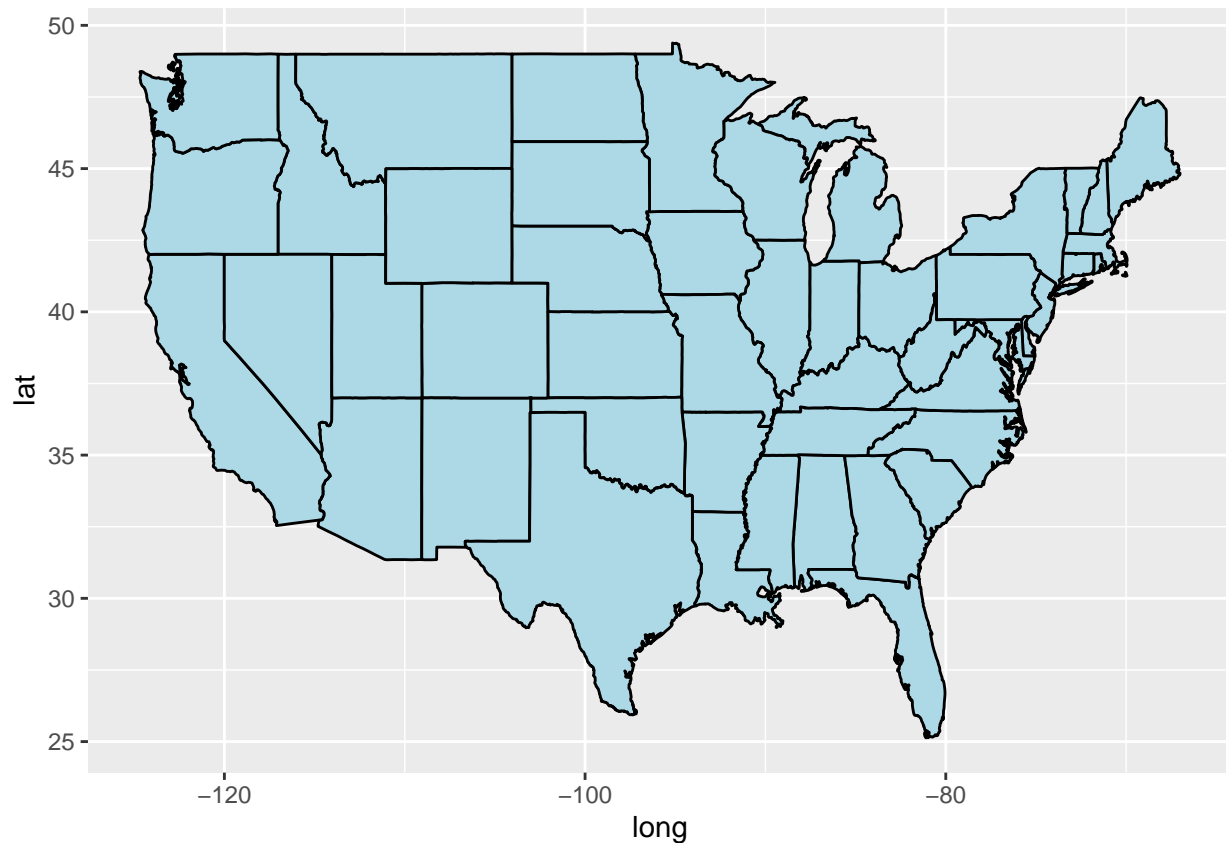
```
ggplot(data = df,
       aes(x = change_yoy)) +
  geom_histogram(binwidth = 15)+
  labs(title = "YOY Percent Change Distribution", x = "YOY Percent Change", y = "Count") +
  theme_bw()
```



```
ggplot(data = df,
       aes(x = month,
           y = change_yoy,
           color = coverage_code)) +
  geom_point()+
  scale_x_discrete("Month", limits = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov")) +
  labs(title = "YOY Percent Change Over Months", y = "YOY Percent Change") +
  theme_bw()
```



```
states <- map_data("state")
ggplot() +
  geom_polygon(data = states, aes(x = long, y = lat, group = group),
    color = "black", fill = "lightblue")
```

```
df2 <- df %>%
  filter(year == 2020 & month == 9 & subsector == "total")
head(df2)
```

```
## # A tibble: 6 x 9
##   fips state_abbr naics subsector year month change_yoy change_yoy_se
##   <chr> <chr>      <dbl> <chr>      <dbl> <dbl>      <dbl> <chr>
## 1 00    USA          NA total      2020  9         6  0.3
## 2 01    AL          NA total      2020  9        7.5 1.7
## 3 02    AK          NA total      2020  9        1.5 4.3
## 4 04    AZ          NA total      2020  9        1.4 2.8
## 5 05    AR          NA total      2020  9        3.7 2.2
## 6 06    CA          NA total      2020  9        2.9 0.9
## # ... with 1 more variable: coverage_code <chr>
```

```
usmap::plot_usmap(data = df2, values = "change_yoy", color = "grey", size = 0.25) +
  labs(title = "YOY Percentage Change")
```

```
## Warning: Ignoring unknown parameters: linewidth
```

YOY Percentage Change

