

Statistics Experience: TidyTuesday 4/8/25

Garrett Spooner

2025-04-07

```
library(tidyverse)
```

```
-- 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  1.9.3      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
```

```
library(dplyr)
```

```
care_state <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday')
```

```
Rows: 1232 Columns: 8
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
chr  (5): state, condition, measure_id, measure_name, footnote
```

```
dbl  (1): score
```

```
date (2): start_date, end_date
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
filtered_data <- care_state |>
  filter(measure_id == "OP_18c") |>
  select(state, measure_name, score)
```

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

Attaching package: 'maps'

The following object is masked from 'package:purrr':

map

```
library(patchwork)
```

```
filtered_data <- filtered_data |>
  mutate(state = tolower(state.name[match(state, state.abb)])) |>
  filter(!is.na(state))
```

```
states_map <- map_data("state")

northeast_states <- c("connecticut", "maine", "massachusetts", "new hampshire",
  "rhode island", "vermont", "new jersey", "new york", "pennsylvania")

southeast_states <- c("delaware", "florida", "georgia", "maryland", "north carolina",
  "south carolina", "virginia", "west virginia", "kentucky",
  "tennessee", "alabama", "mississippi", "arkansas", "louisiana")

map_data_joined <- states_map |>
  left_join(filtered_data, by = c("region" = "state"))
```

```
northeast_map <- map_data_joined |>
  filter(region %in% northeast_states)

shared_limits <- c(175, 450)

NE <- ggplot(northeast_map, aes(long, lat, group = group, fill = score)) +
  geom_polygon(color = "white") +
```

```

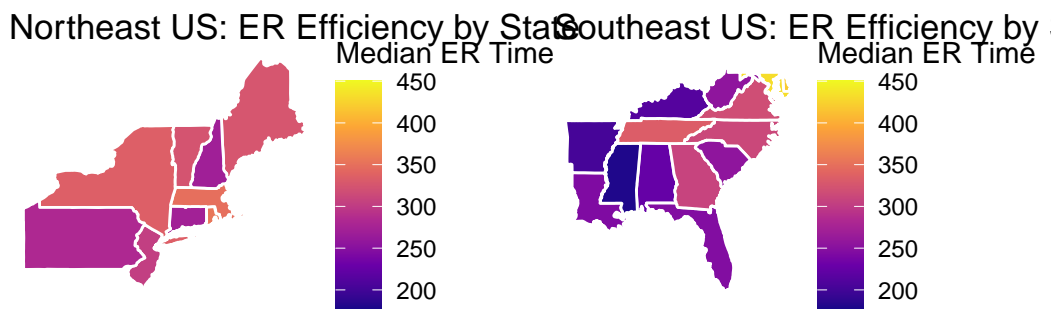
coord_fixed(1.3) +
scale_fill_viridis_c(option = "C", limits = shared_limits) +
labs(title = "Northeast US: ER Efficiency by State", fill = "Median ER Time") +
theme_void()

southeast_map <- map_data_joined %>%
  filter(region %in% southeast_states)

SE <- ggplot(southeast_map, aes(long, lat, group = group, fill = score)) +
  geom_polygon(color = "white") +
  coord_fixed(1.3) +
  scale_fill_viridis_c(option = "C", limits = shared_limits) +
  labs(title = "Southeast US: ER Efficiency by State", fill = "Median ER Time") +
  theme_void()

NE + SE

```



```
#Used ChatGPT to understand maps package
```

```

filtered_data_NE <- filtered_data |>
  filter(state %in% northeast_states)

```

```
filtered_data_SE <- filtered_data |>  
  filter(state %in% southeast_states)
```

```
mean(filtered_data_SE$score)
```

```
[1] 281.8571
```

```
mean(filtered_data_NE$score)
```

```
[1] 312.7778
```