Treatment Mapping

Load Libraries

```
library(tidyverse)
library(here)
library(arrow)
library(sf)
library(urbnmapr)
library(naniar)
library(janitor)
library(ggiraph)
options(scipen = 99)
```

Today’s Data

The data we will analyze today is SAMHSA’s TEDS-D Dataset. The metadata can be found here

Reading in feather files with `arrow`

```
teds_d <- read_parquet(here("data/teds_d_lecture.parquet"))
```

Clean names

```
teds_d <- teds_d %>%
  clean_names()
```

Selecting for relevant columns for today’s class

- State
- Frequency of use at discharge
- Treatment Service
- Length of Stay
- Reason for Discharge

```
 teds_d_select <- teds_d %>%
  select(freq1_d, stfips, services_d, los, reason)
```

```
write_parquet(teds_d_select, here("data/teds_d_lecture.parquet"))
```

```
teds_d_select <- read_parquet(here("data/teds_d_lecture.parquet"))
```

NA Analysis

How does the documentation label missing data?

```
teds_d_select[teds_d_select == "-9"] <- NA
```

```
miss_var_summary(teds_d_select)
```

```
# A tibble: 5 × 3
  variable    n_miss  pct_miss
  <chr>        <int>     <num>
1 freq1_d    7263891 51.8
2 services_d 4715728 33.6
3 reason         140  0.000997
4 los             18  0.000128
5 stfips           0  0
```

Variable Re-coding

Frequency of Use at Discharge

```
teds_d_select$freq1_d <- as.character(teds_d_select$freq1_d)

teds_d_select$freq1_d[teds_d_select$freq1_d == "1"] <- "no use"

teds_d_select$freq1_d[teds_d_select$freq1_d == "2"] <- "some use"

teds_d_select$freq1_d[teds_d_select$freq1_d == "3"] <- "daily use"

teds_d_select$freq1_d[is.na(teds_d_select$freq1_d)] <- "unknown"
```

Services

```
teds_d_select$services_d <- as.character(teds_d_select$services_d)


teds_d_select$services_d[teds_d_select$services_d == "1"] <- "Detox, 24-hour, hos

teds_d_select$services_d[teds_d_select$services_d == "2"] <- "Detox, 24-hour, fre

teds_d_select$services_d[teds_d_select$services_d == "3"] <- "Rehab/residential,

teds_d_select$services_d[teds_d_select$services_d == "4"] <- "Rehab/residential,

teds_d_select$services_d[teds_d_select$services_d == "5"] <- "Rehab/residential,

teds_d_select$services_d[teds_d_select$services_d == "6"] <- "Ambulatory, intensi

teds_d_select$services_d[teds_d_select$services_d == "7"] <- "Ambulatory, non-int
```

```
teds_d_select$services_d[teds_d_select$services_d == "8"] <- "Ambulatory, detoxif

teds_d_select$services_d[is.na(teds_d_select$services_d)] <- "unknown"
```

Reason

```
teds_d_select$reason <- as.character(teds_d_select$reason)

teds_d_select$reason[teds_d_select$reason == "1"] <- "completed"

teds_d_select$reason[teds_d_select$reason == "2"] <- "dropped out"

teds_d_select$reason[teds_d_select$reason == "3"] <- "terminated by facility"

teds_d_select$reason[teds_d_select$reason == "4"] <- "transfered"

teds_d_select$reason[teds_d_select$reason == "5"] <- "incarcerated"

teds_d_select$reason[teds_d_select$reason == "6"] <- "death"

teds_d_select$reason[teds_d_select$reason == "7"] <- "other"
```

Mapping

We want to map the percentage of complete treatments by state

First, let’s calculate the percentage of completed treatments by state

```
percent_completed_by_state <- teds_d_select %>%
  group_by(stfips) %>%
  summarize(
    total_cases = n(),
    completed_cases = sum(reason == "completed", na.rm = TRUE)
  ) %>%
  mutate(percentage_completed = (completed_cases / total_cases) * 100)
```

Next, let’s bring in some mapping data

```
states_map <- get_urbn_map(map = "states", sf = TRUE)
```

What do we notice that’s different between the teds-d stfips column and the states_map stfips column?

```
percent_completed_by_state$stfips_recode <- sprintf('%02d', percent_completed_by_
```

```
colnames(percent_completed_by_state)[colnames(percent_completed_by_state) == "stf
```

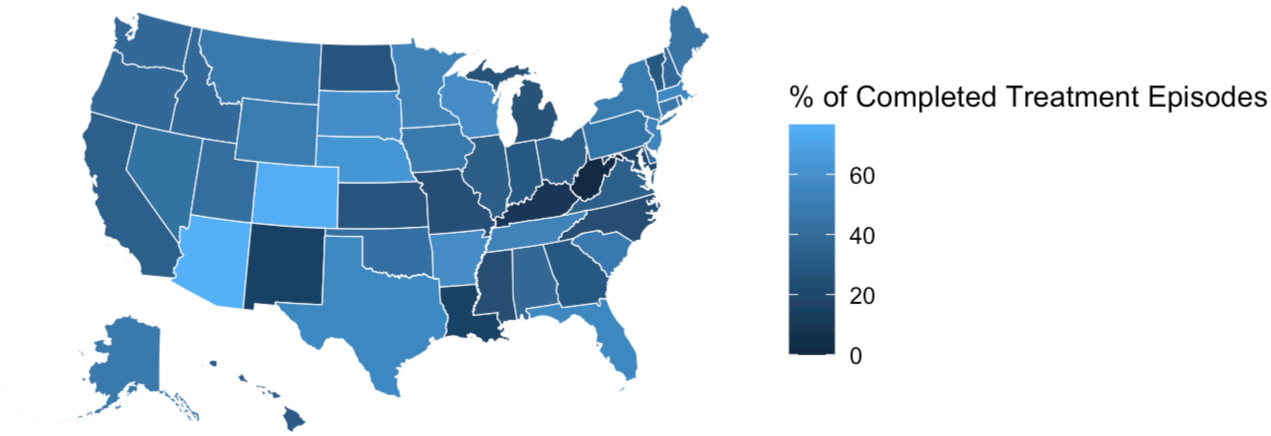
Joining data

```
percent_completed_by_state_map <- full_join(percent_completed_by_state,
                          states_map,
                          by = "state_fips")
```

```
old-style crs object detected; please recreate object with a recent sf::st_crs()
```

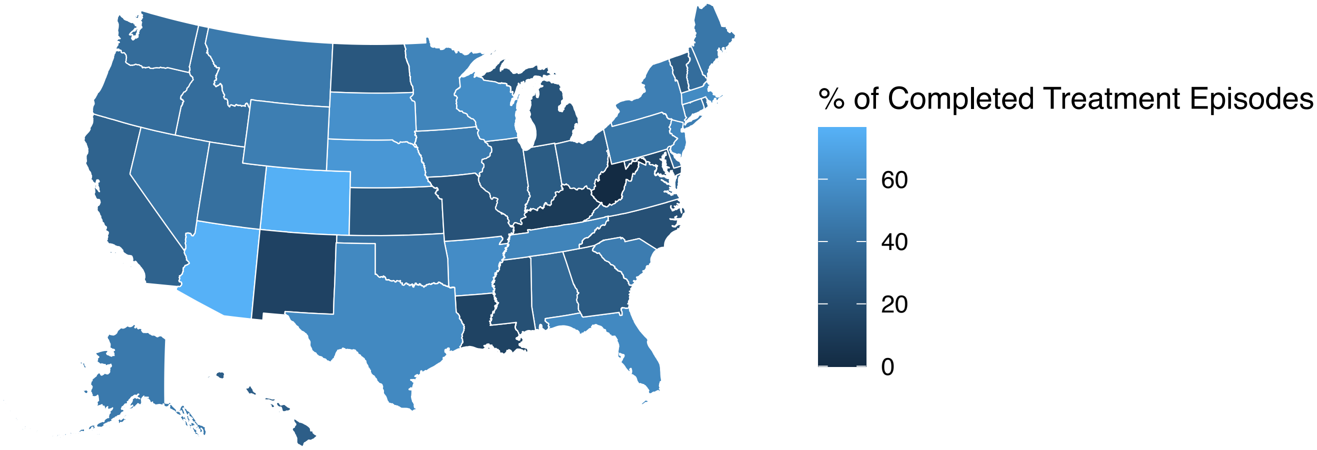
Plotting Map

```
ggplot(percent_completed_by_state_map) +
  geom_sf(percent_completed_by_state_map,
          mapping = aes(geometry = geometry, fill = percentage_completed),
          color = "#ffffff", size = 0.25) +
  labs(fill = "% of Completed Treatment Episodes") +
   coord_sf(datum = NA)+
   theme_minimal()
```



Making interactive with `ggiprah`

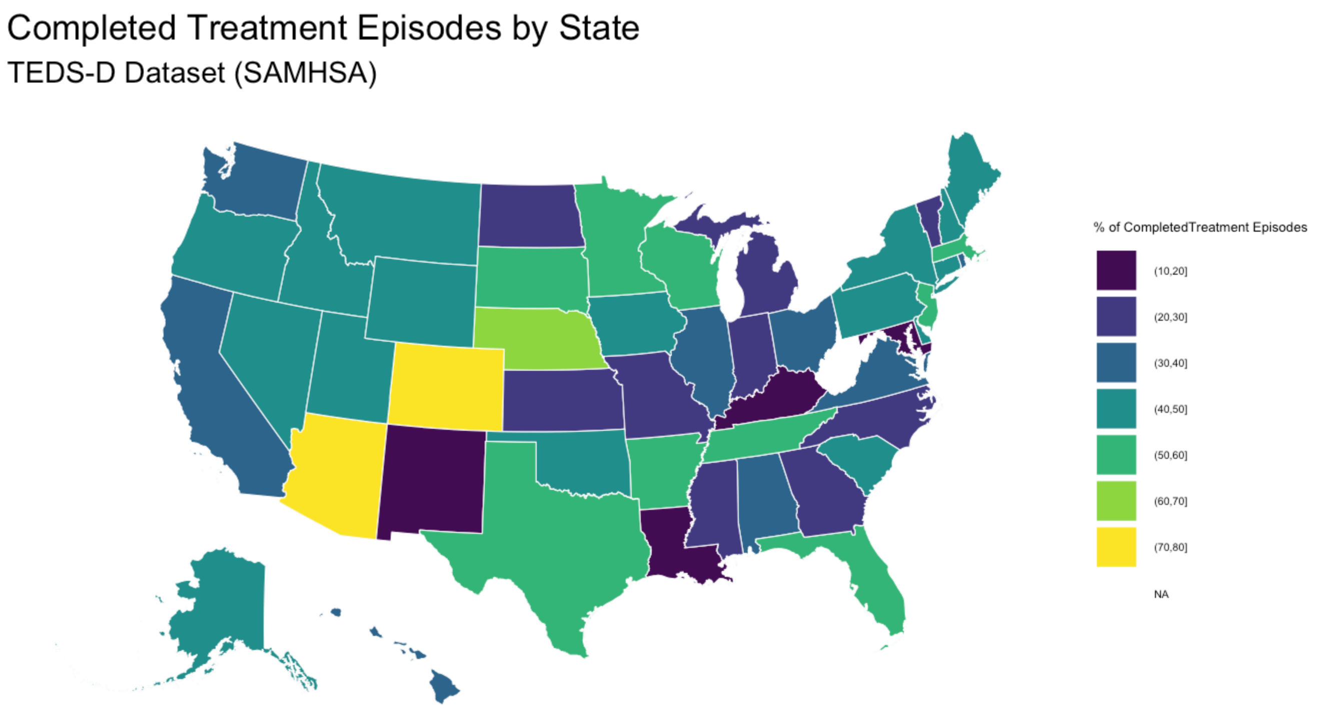
```
interactive_completed_treatment_map <- ggplot(percent_completed_by_state_map) +
  geom_sf_interactive(
    mapping = aes(
      geometry = geometry,
      fill = percentage_completed,
      tooltip = paste("State FIPS:", stfips, "State Name:", state_name, "<br>Comp
    ),
    color = "#ffffff",
    size = 0.25
  ) +
  labs(fill = "% of Completed Treatment Episodes") +
  coord_sf(datum = NA) +
  theme_minimal()

# Use `girafe` to render the interactive plot
girafe(ggobj = interactive_completed_treatment_map)
```



Round & Add state name to tooltip

Adding color bins

```
percent_completed_by_state_map <- percent_completed_by_state_map %>%
  mutate(percentage_bin = cut(percentage_completed, breaks=c(0, 10,20,30,40,50, 6

ggplot(percent_completed_by_state_map) +
  geom_sf(mapping = aes(geometry = geometry, fill = percentage_bin),
          color = "#ffffff", size = 0.25) +
  labs(fill = "% of CompletedTreatment Episodes",
      title = "Completed Treatment Episodes by State",
      subtitle = "TEDS-D Dataset (SAMHSA)") +
  scale_fill_viridis_d(option = "D") +
  coord_sf(datum = NA) +
  theme_minimal() +
  theme(
    panel.background = element_blank(),
    axis.ticks = element_blank(),
    axis.text.x = element_blank(),
    axis.text.y = element_blank(),
    legend.text = element_text(size = 4),
    legend.title = element_text(size = 5),
    strip.text = element_text(size = 4)
  )
```



Assignment

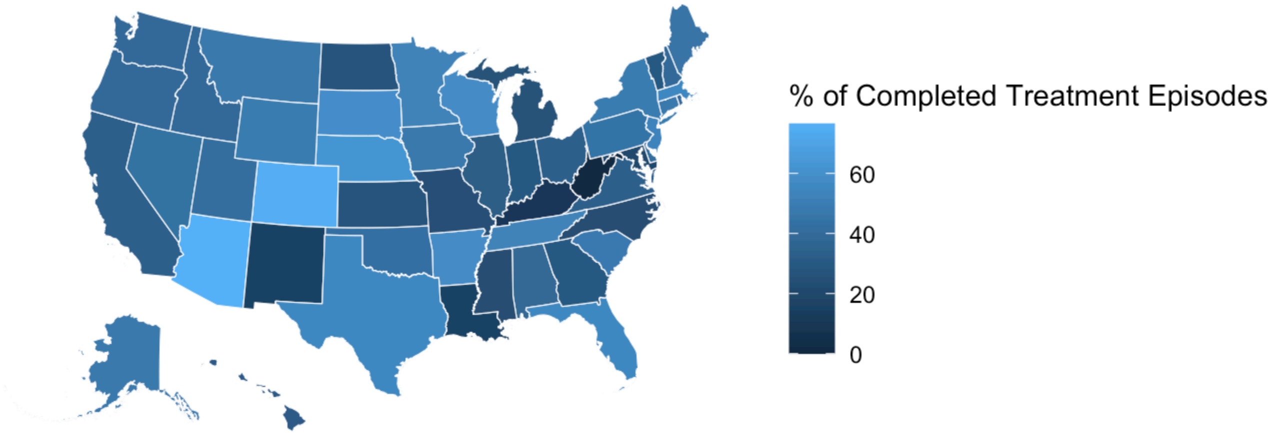
1. Make an interactive map with `ggiraph` showing the percentage of completed treatments that end with no use at discharge (freq1_d)

   –> no use at discharge / completed treatments

```
percent_completed_by_states_map <- full_join(percent_completed_by_state,
                          states_map,
                          by = "state_fips")
```

```
old-style crs object detected; please recreate object with a recent sf::st_crs()
```

```
ggplot(percent_completed_by_state_map) +
  geom_sf(percent_completed_by_state_map,
          mapping = aes(geometry = geometry, fill = percentage_completed),
          color = "#ffffff", size = 0.25) +
  labs(fill = "% of Completed Treatment Episodes") +
   coord_sf(datum = NA)+
   theme_minimal()
```



```
percent_no_use_discharge <- teds_d_select %>%
  group_by(stfips) %>%
  summarize(
    total_cases = n(),
    completed_cases = sum(reason == "completed", na.rm = TRUE),
    completed_no_use_discharge = sum(freq1_d == "no use" & reason == "completed",
  ) %>%
  mutate(percentage_no_use = (completed_no_use_discharge / completed_cases) * 100
```

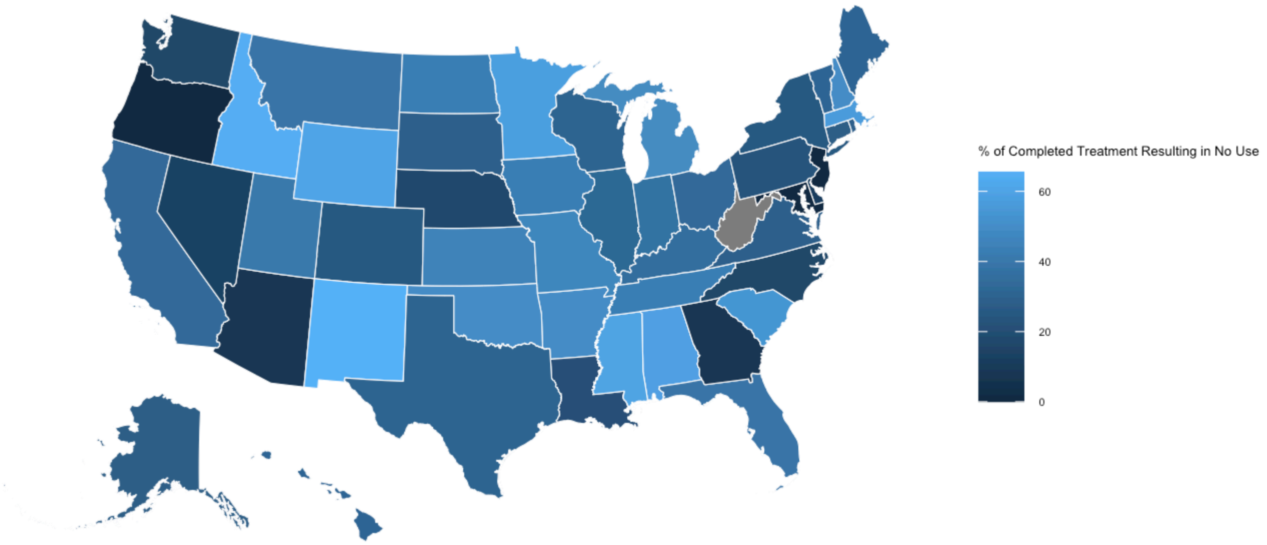
```
percent_no_use_discharge$stfips_recode <- sprintf('%02d',percent_no_use_discharge
```

```
colnames(percent_no_use_discharge)[colnames(percent_no_use_discharge) == "stfips_
```

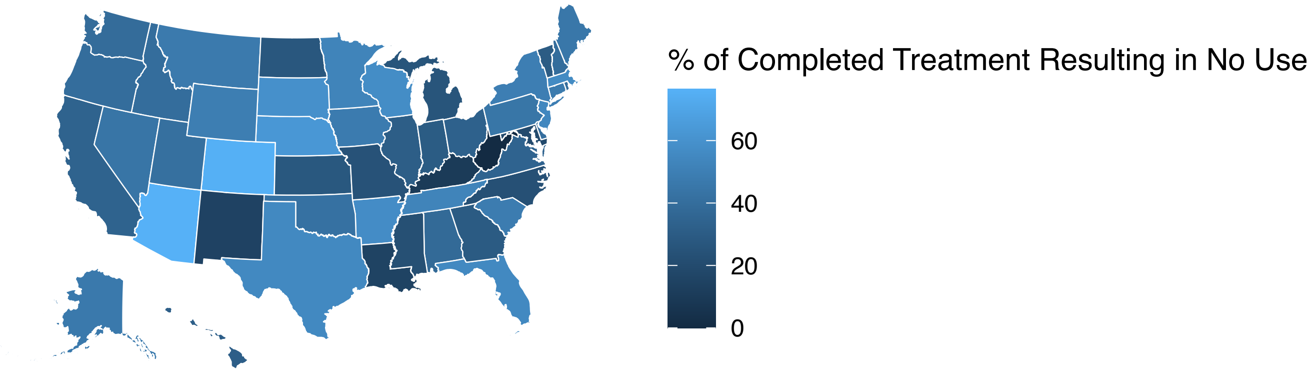
```
percent_no_use_discharge_map <- full_join(percent_no_use_discharge,
                          states_map,
                          by = "state_fips")
```

```
old-style crs object detected; please recreate object with a recent sf::st_crs()
```

```
interactive_no_use_discharge_map <- ggplot(percent_no_use_discharge_map) +
  geom_sf_interactive(
    mapping = aes(
      geometry = geometry,
      fill = percentage_no_use,
      tooltip = paste("State FIPS:", stfips, "<br>Completed with no use:", round(
    ),
    color = "#ffffff",
    size = 0.1
  ) +
  labs(fill = "% of Completed Treatment Resulting in No Use",
       title = "Completed Treatments that end with No Use at Discharge") +
  coord_sf(datum = NA) +
  theme_minimal() +
  theme(
    panel.background = element_blank(),
    legend.text = element_text(size = 4),
    legend.title = element_text(size = 5)
  )

interactive_no_use_discharge_map
```


Completed Treatments that end with No Use at Discharge

```
interactive_no_use_discharge_map <- ggplot(percent_completed_by_state_map) +
  geom_sf_interactive(
    mapping = aes(
      geometry = geometry,
      fill = percentage_completed,
      tooltip = paste("State FIPS:", stfips, "State Name:", state_name, "<br>Comp
    ),
    color = "#ffffff",
    size = 0.25
  ) +
  labs(fill = "% of Completed Treatment Resulting in No Use",
       title = "Completed Treatments that end with No Use at Discharge") +
  coord_sf(datum = NA) +
  theme_minimal()

# Use `girafe` to render the interactive plot
girafe(ggobj = interactive_no_use_discharge_map)
```



Completed Treatments that end with No Use at Discharge

1. How does the **percentage of treatments being <u>completed</u>** & **percentage of treatments ending with no use <u>vary by the service</u>** (completed_cases) and **<u>length of stay</u>**. (services_d) Create at least 3 visualizations to try and answer this question
    1. group_by service or LOS, etc.

```
summary(as.factor(teds_d_select$services_d))
```

```
                   Ambulatory, detoxification 
                                        81356 
             Ambulatory, intensive outpatient 
                                      1255393 
         Ambulatory, non-intensive outpatient 
                                      4564225 
    Detox, 24-hour, free-standing residential 
                                      1449619 
           Detox, 24-hour, hospital inpatient 
                                       242358 
       Rehab/residential, hospital (non-detox) 
                                        23570 
Rehab/residential, long term (more than 30 days) 
                                       721832 
Rehab/residential, short term (30 days or fewer) 
                                       981477 
```

```
unknown
4715728
```

```r
summary(as.factor(teds_d_select$los))
```

```
      1       2       3       4       5       6       7       8       9      10 
1580286  599892  550056  535766  475859  319425  268916  195634  133531  121493 
     11      12      13      14      15      16      17      18      19      20 
  99963   95348  130989  180950  151021   96033   84998   76844   76950  100876 
     21      22      23      24      25      26      27      28      29      30 
 150730  118777   83618   77527   71309   72548  111857  217478  155176  125686 
     31      32      33      34      35      36      37    NA's 
 939137  722560 1213096  983155 1132884 1263605  721567      18 
```

```r
percent_completed_w_services <- teds_d_select %>%
  group_by(stfips, services_d) %>%
  summarize(
    total_cases = n(),
    completed_cases = sum(reason == "completed", na.rm = TRUE)
  ) %>%
  mutate(percent_complete_by_service = (completed_cases / total_cases) * 100)
```

```
`summarise()` has grouped output by 'stfips'. You can override using the
`.groups` argument.
```

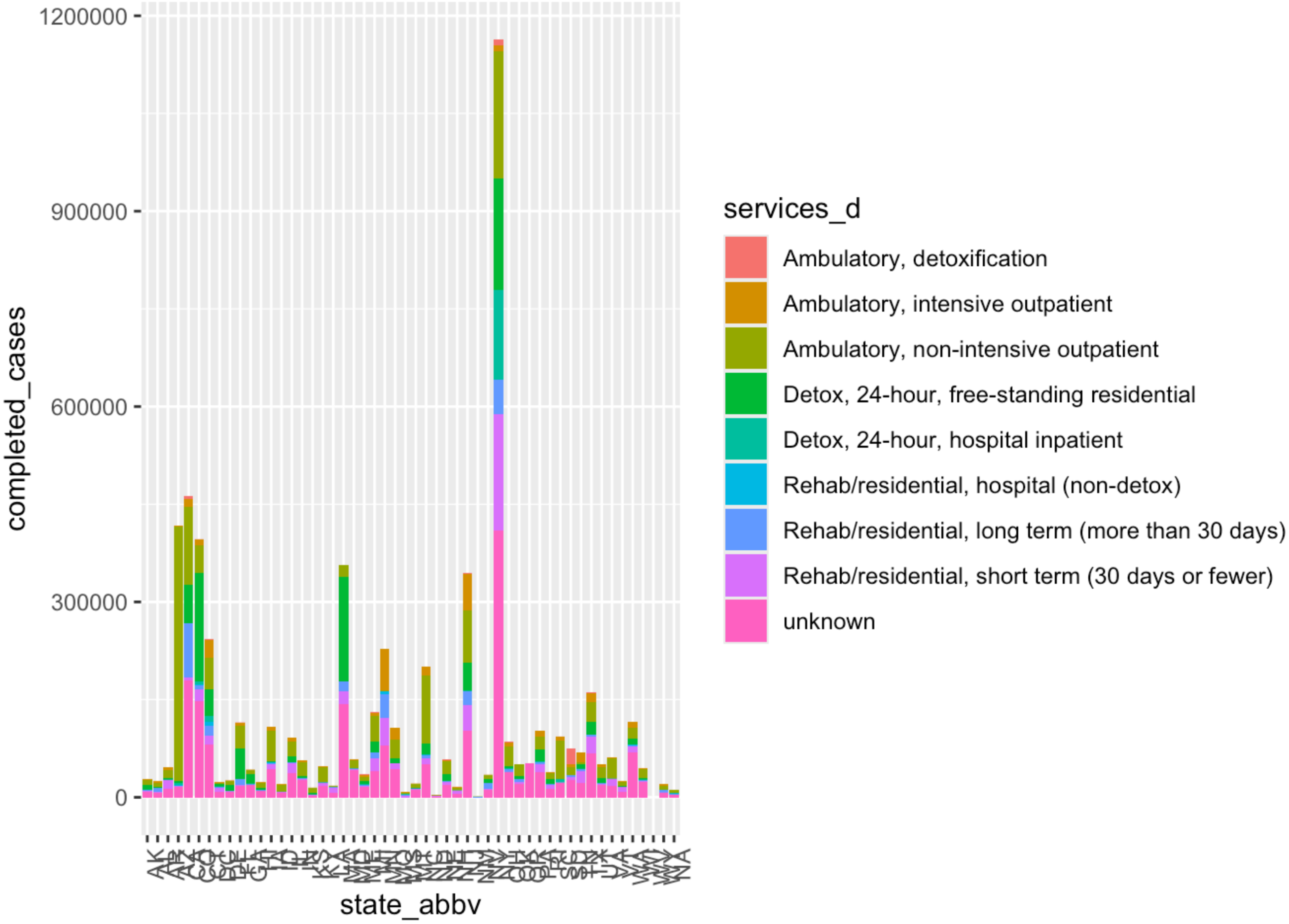
```r
nlevels(as.factor(percent_completed_w_services$stfips)) # checking to see that al
```

```
[1] 52
```

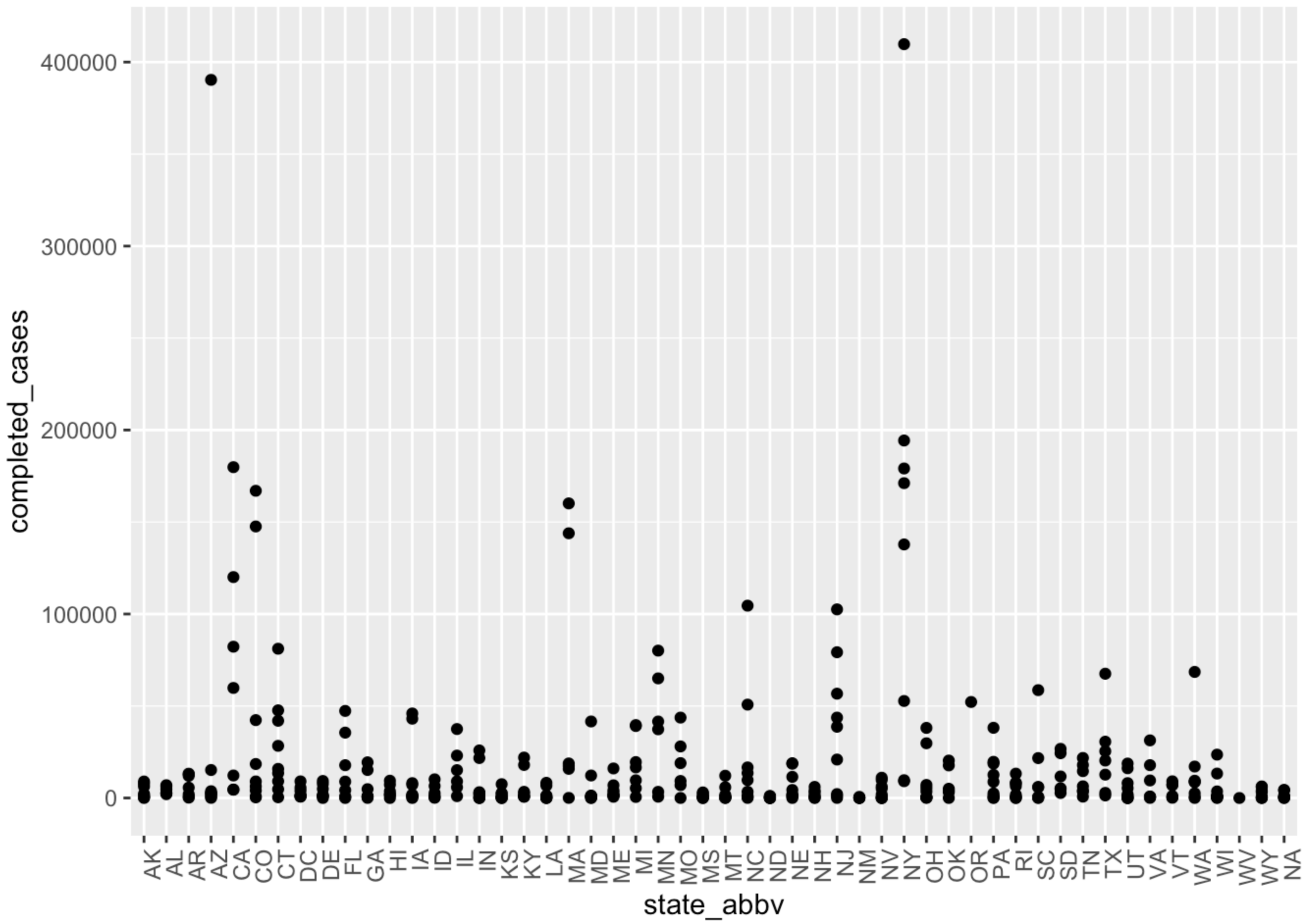
```r
percent_completed_w_services$state_fips <- sprintf('%02d', percent_completed_w_se

percent_completed_w_services_map <- full_join(percent_completed_w_services,
                                   states_map,
                                   by = "state_fips")
```

```
old-style crs object detected; please recreate object with a recent sf::st_crs()
```

```r
ggplot(percent_completed_w_services_map, aes(x = state_abbv, y = completed_cases,
  geom_bar(stat = "identity") +
   theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

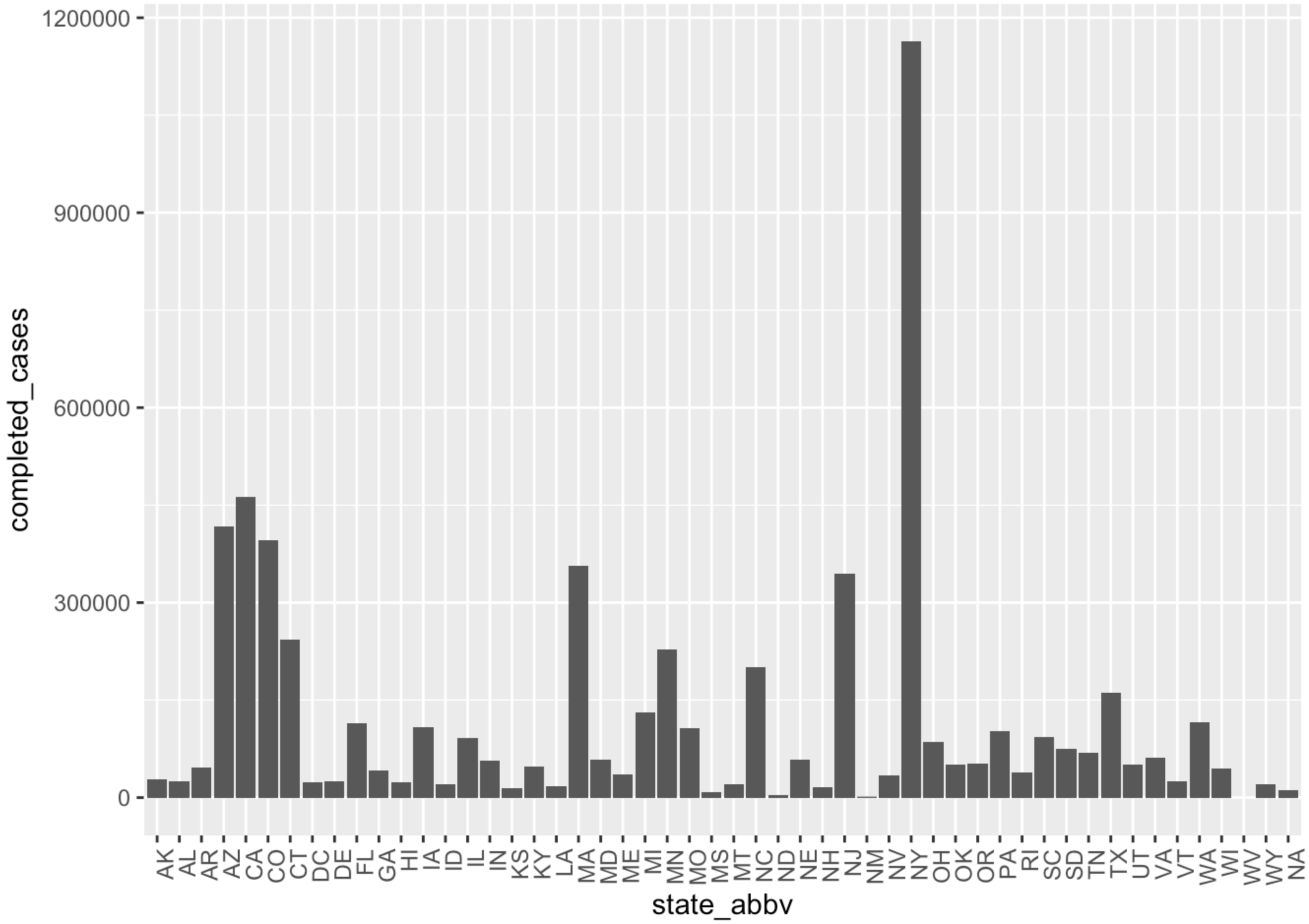


```
ggplot(percent_completed_w_services_map, aes(x = state_abbv, y = completed_cases)
  geom_point(stat = "identity") +
   theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

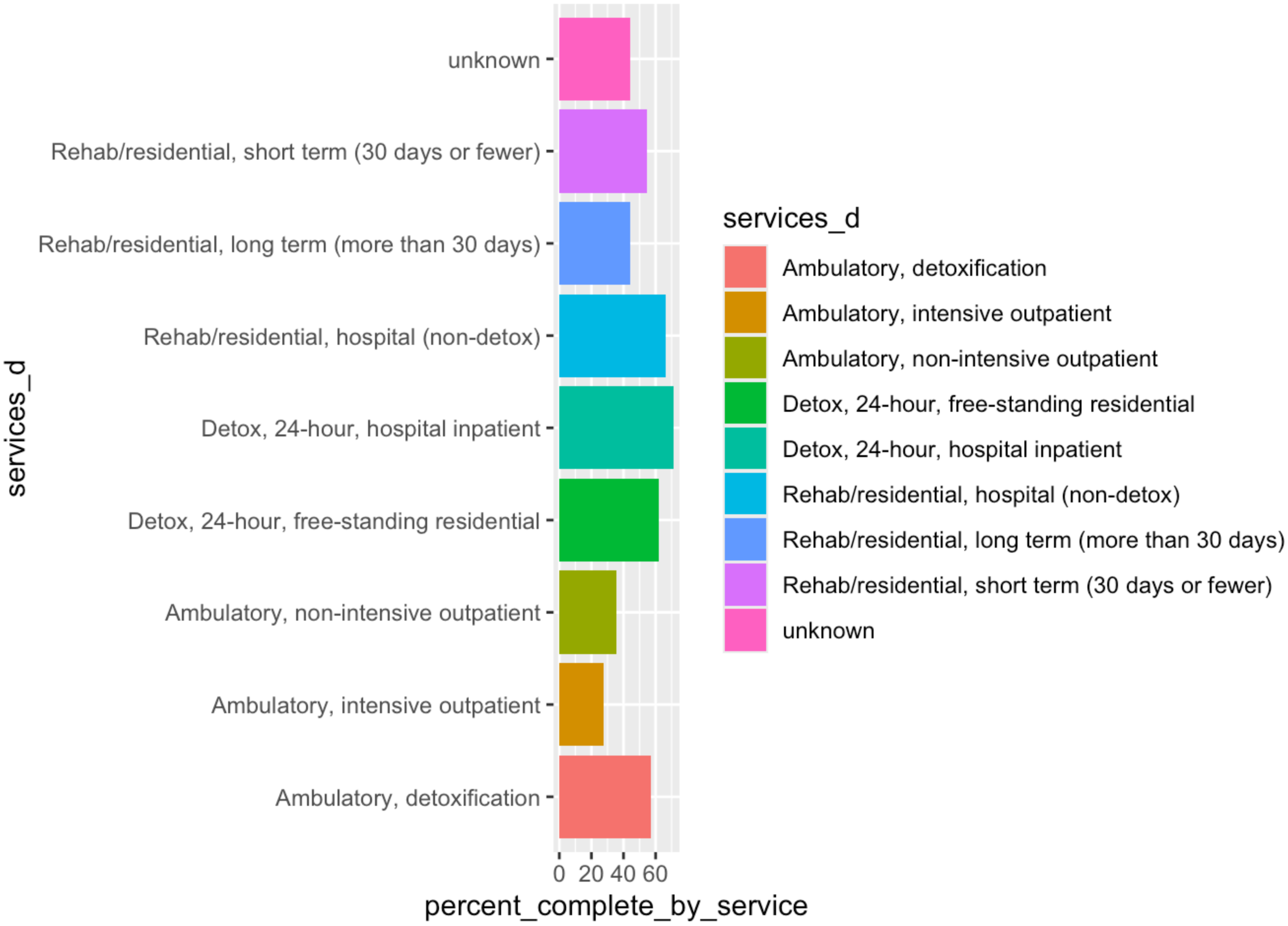


```
ggplot(percent_completed_w_services_map, aes(x = state_abbv, y = completed_cases)
  geom_histogram(stat = "identity") +
   theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

```
Warning in geom_histogram(stat = "identity"): Ignoring unknown parameters:
`binwidth`, `bins`, and `pad`
```



```
percent_completed_w_services <- teds_d_select %>%
  group_by(services_d) %>%
  summarize(
    total_cases = n(),
    completed_cases = sum(reason == "completed", na.rm = TRUE)
  ) %>%
  mutate(percent_complete_by_service = (completed_cases / total_cases) * 100)

ggplot(percent_completed_w_services, aes(x = percent_complete_by_service, y = ser
  geom_bar(stat = "identity")
```


unknown
Rehab/residential, short term (30 days or fewer)
Rehab/residential, long term (more than 30 days)
Rehab/residential, hospital (non-detox)
Detox, 24-hour, hospital inpatient
Detox, 24-hour, free-standing residential
Ambulatory, non-intensive outpatient
Ambulatory, intensive outpatient
Ambulatory, detoxification
services_d
0
20
40
60
percent_complete_by_service
services_d
Ambulatory, detoxification
Ambulatory, intensive outpatient
Ambulatory, non-intensive outpatient
Detox, 24-hour, free-standing residential
Detox, 24-hour, hospital inpatient
Rehab/residential, hospital (non-detox)
Rehab/residential, long term (more than 30 days)
Rehab/residential, short term (30 days or fewer)
unknown