Treatment Mapping

Load Libraries

```r
library(plotly)
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`

```r
#teds_d <- read_parquet(here("data/tedsD_2012_2020.parquet"))
```

Clean names

```r
# 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 x 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, hospital inpati

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

teds_d_select$services_d[teds_d_select$services_d == "3"] <- "Rehab/residential, hospital (no

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

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

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

teds_d_select$services_d[teds_d_select$services_d == "7"] <- "Ambulatory, non-intensive outpa

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

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"
```

```r
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

```r
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

```r
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?

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

colnames(percent_completed_by_state)[colnames(percent_completed_by_state) == "stfips_recode"]
```

Joining data

```r
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()
```

```
Warning in CPL_crs_from_input(x): GDAL Message 1: CRS EPSG:2163 is deprecated.
Its non-deprecated replacement EPSG:9311 will be used instead. To use the
original CRS, set the OSR_USE_NON_DEPRECATED configuration option to NO.
```

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()
```
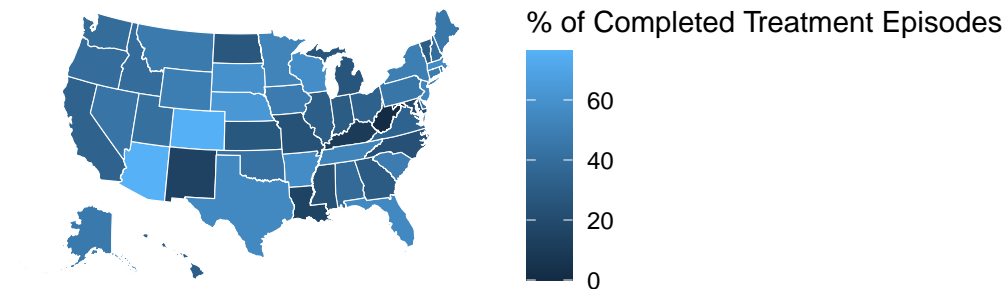

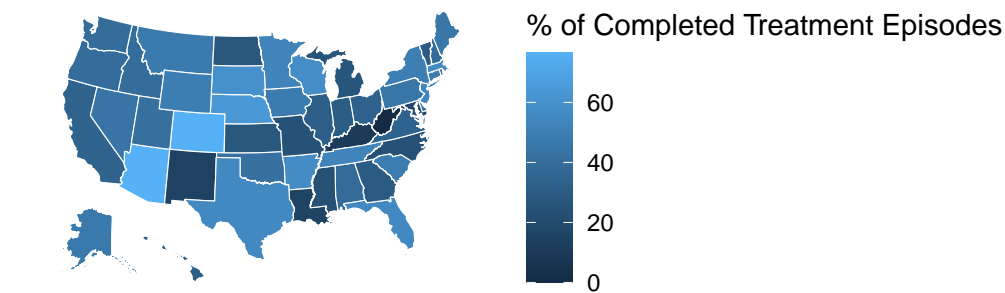Making interactive with ggiprah

```
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, "<br>Completed:", percentage_completed, "%")
    ),
    color = "#ffffff",
    size = 0.25
  ) +
  labs(fill = "% of Completed Treatment Episodes") +
  coord_sf(datum = NA) +
  theme_minimal()

interactive_completed_treatment_map
```



```
# 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, 60, 70, 80)))
```

```r
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)
  )
```

Completed Treatment Episodes by State
TEDS−D Dataset (SAMHSA)
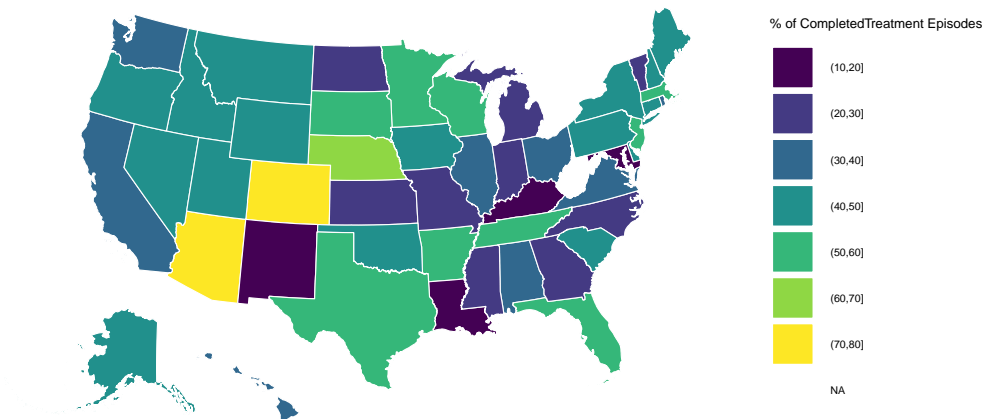
Assignment

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

```
unique(teds_d_select$services_d)
```

```
[1] "unknown"
[2] "Ambulatory, non-intensive outpatient"
[3] "Ambulatory, intensive outpatient"
[4] "Rehab/residential, long term (more than 30 days)"
[5] "Ambulatory, detoxification"
[6] "Detox, 24-hour, free-standing residential"
[7] "Detox, 24-hour, hospital inpatient"
[8] "Rehab/residential, short term (30 days or fewer)"
[9] "Rehab/residential, hospital (non-detox)"
```

```
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", na.rm = TRU
  ) %>%
  mutate(percentage_no_use = (completed_no_use_discharge / completed_cases) * 100)
```

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

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

```
colnames(percent_no_use_discharge)[colnames(percent_no_use_discharge) == "stfips_recode"] <-
```

```
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(percentage_
    ),
    color = "#ffffff",
    size = 0.1
  ) +
  labs(fill = "% of Completed Treatment Resulting in No Use",
       title = "Completed Treatment Episodes Resulting in No Use by State",
       subtitle = "TEDS-D Dataset (SAMHSA)") +
  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 Treatment Episodes Resulting in No Use by State
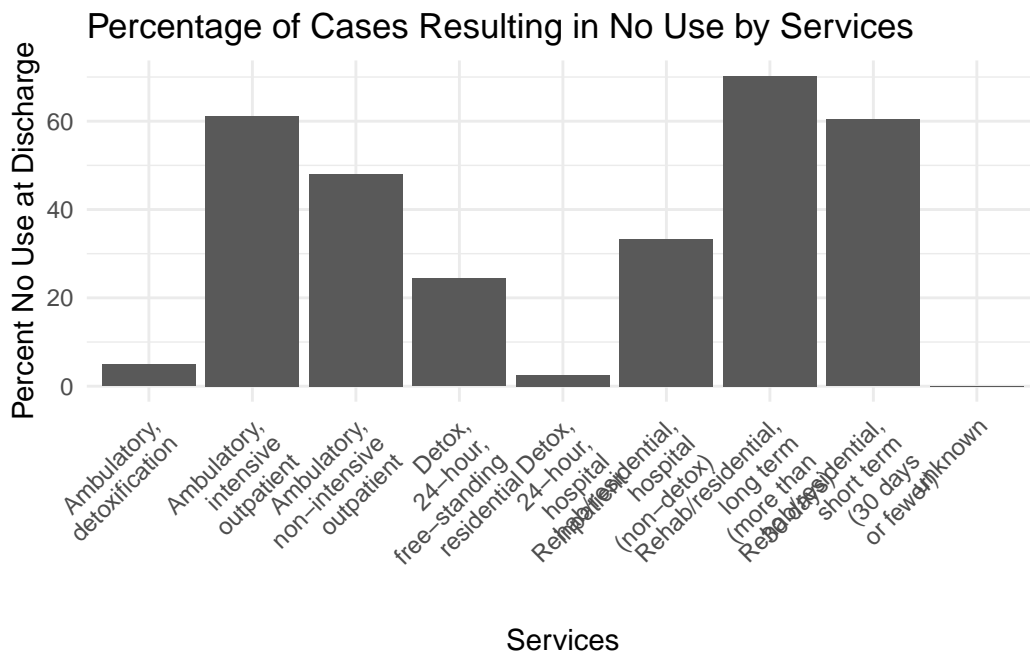TEDS−D Dataset (SAMHSA)
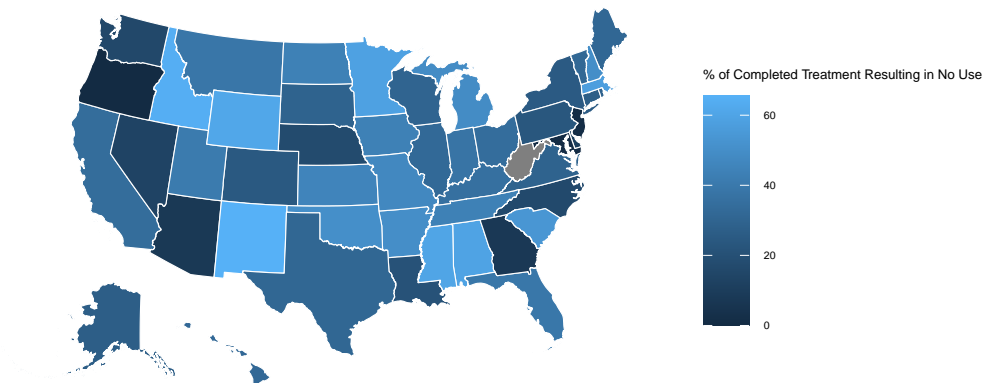
```
# Use `girafe` to render the interactive plot
#girafe(ggobj = interactive_no_use_discharge_map)
```

1. How does the percentage of treatments being completed & percentage of treatments ending with no use vary by the service and length of stay. Create at least 3 visualizations to try and answer this question

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

```
ggplot(data = no_use_by_services, aes(x = services_d, y = percentage_no_use)) +
  geom_bar(stat = "identity") +
  labs(title = "Percentage of Cases Resulting in No Use by Services", x = "Services", y = "Pe
  theme_minimal() +
  scale_x_discrete(labels = function(x) str_wrap(x, width = 10)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Percentage of Cases Resulting in No Use by Services

```
ggplot(data = no_use_by_services, aes(x = services_d, y = percentage_completed)) +
  geom_bar(stat = "identity") +
  labs(title = "Percentage of Cases Completed by Services", x = "Services", y = "Percent of (
  theme_minimal() +
  scale_x_discrete(labels = function(x) str_wrap(x, width = 10)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Percentage of Cases Completed by Services
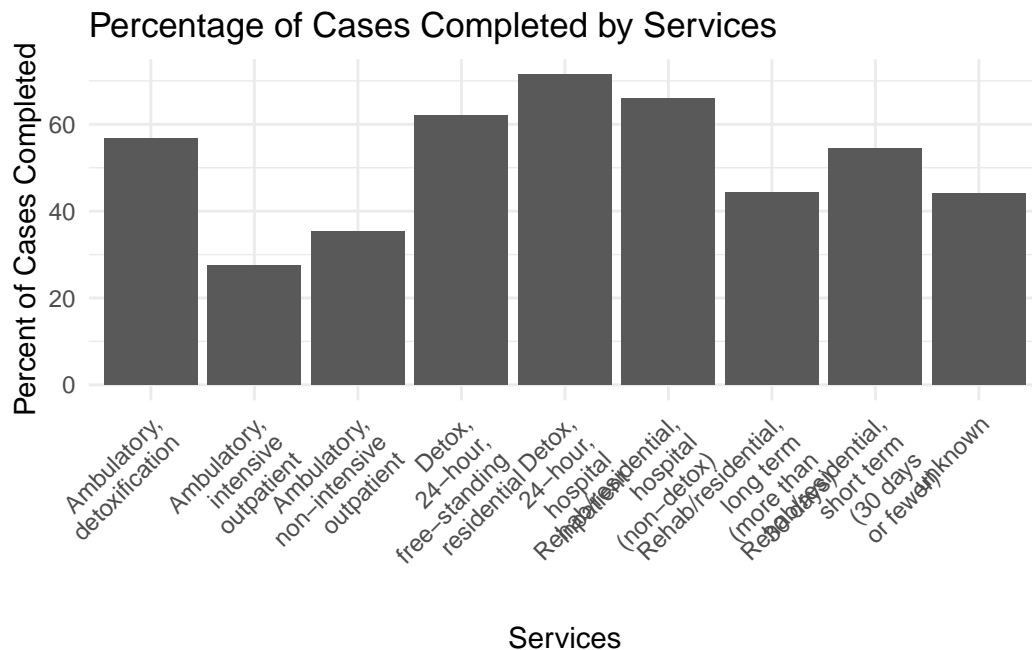
```
# Reshape the data to long format
no_use_by_services_long <- no_use_by_services %>%
  pivot_longer(cols = c(percentage_no_use, percentage_completed),
               names_to = "Outcome",
               values_to = "Percentage")

# Create the combined plot
combined_plot_services <- ggplot(data = no_use_by_services_long, aes(x = services_d, y = Perc
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Percentage of Cases by Services",
       x = "Services",
       y = "Percentage") +
  theme_minimal() +
  scale_x_discrete(labels = function(x) str_wrap(x, width = 10)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
```
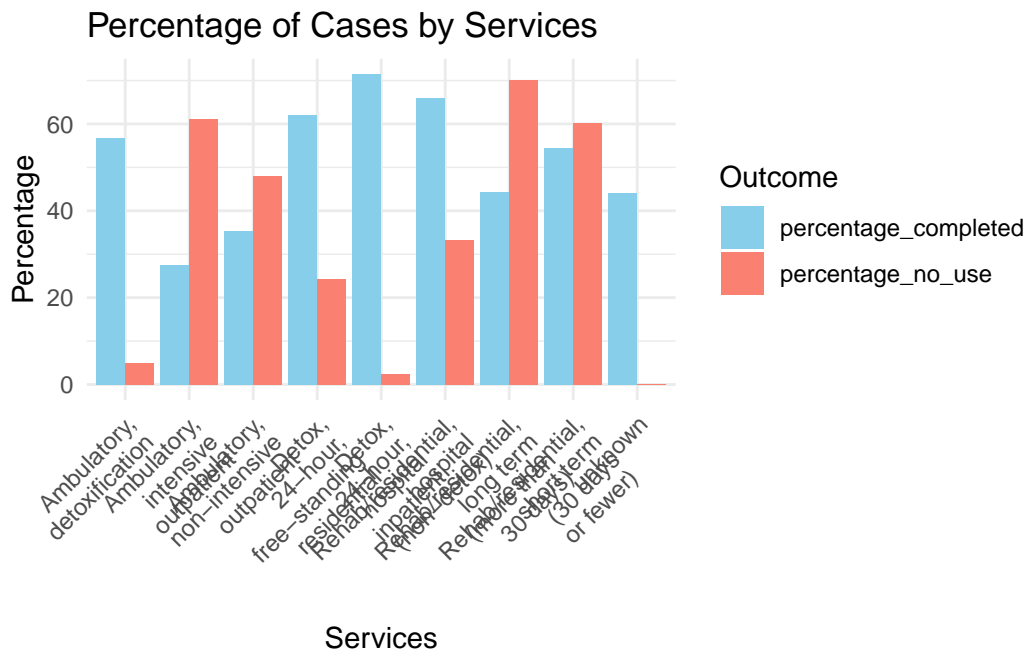
```
  scale_fill_manual(values = c("skyblue", "salmon")) +   # Customize bar colors
  scale_color_manual(values = c("blue", "red"))          # Customize line colors

# Display the plot
print(combined_plot_services)
```

Percentage of Cases by Services


```
no_use_by_los <- teds_d_select %>%
  group_by(los) %>%
  summarize(
    total_cases = n(),
    completed_cases = sum(reason == "completed", na.rm = TRUE),
    completed_no_use_discharge = sum(freq1_d == "no use" & reason == "completed", na.rm = TRU
  ) %>%
  mutate(percentage_no_use = (completed_no_use_discharge / completed_cases) * 100) %>%
  mutate(percentage_completed = (completed_cases / total_cases) * 100)
```
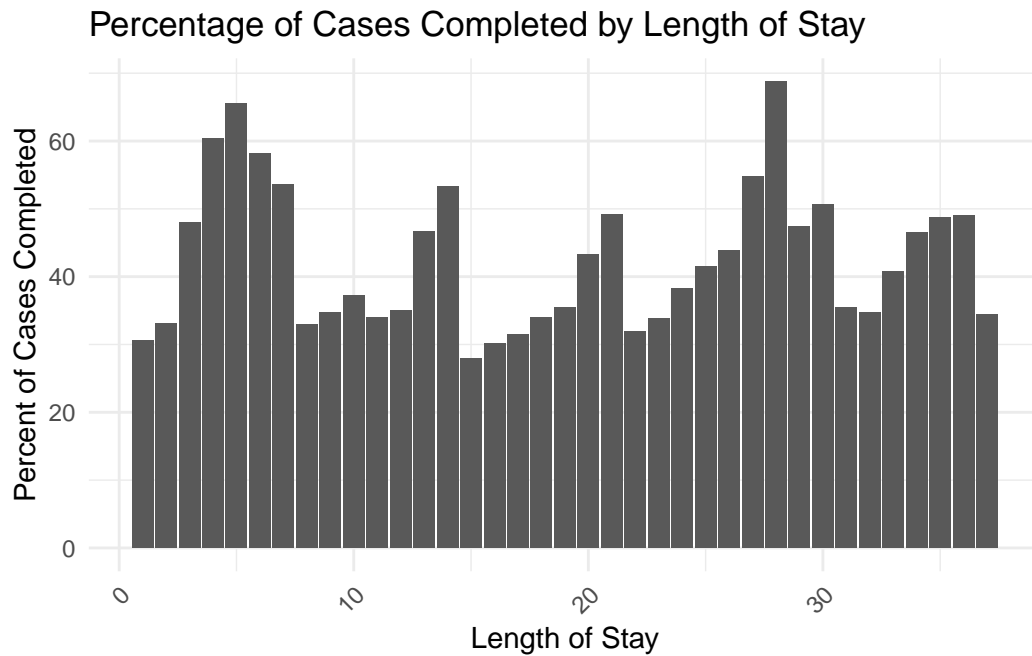
```
ggplot(data = no_use_by_los, aes(x = los, y = percentage_completed)) +
  geom_bar(stat = "identity") +
  labs(title = "Percentage of Cases Completed by Length of Stay", x = "Length of Stay", y = "
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
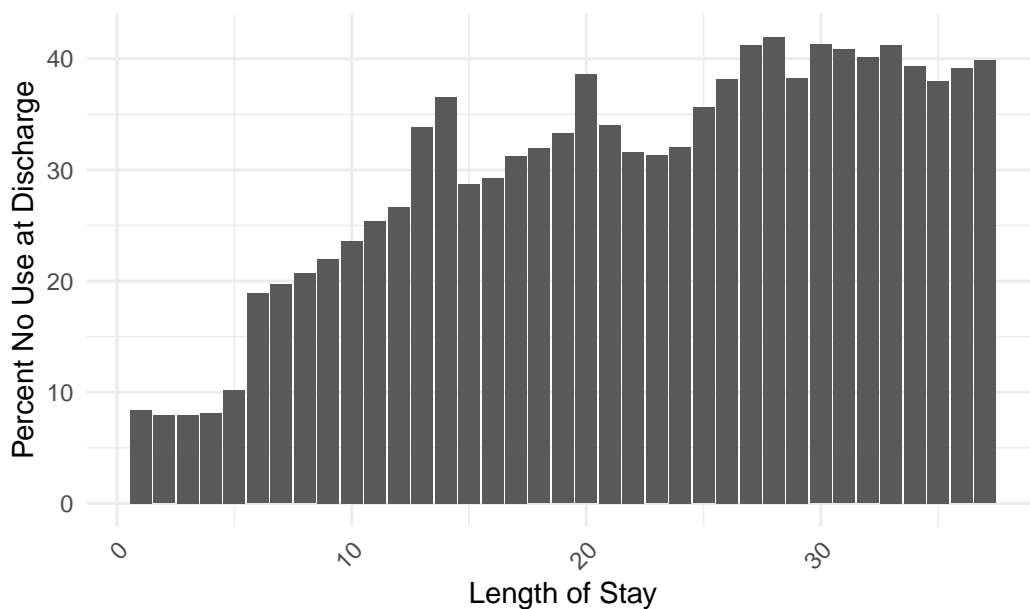
```
Warning: Removed 1 row containing missing values or values outside the scale range
(`geom_bar()`).
```

Percentage of Cases Completed by Length of Stay


```
ggplot(data = no_use_by_los, aes(x = los, y = percentage_no_use)) +
  geom_bar(stat = "identity") +
  labs(title = "Percentage of Cases Resulting in No Use by Length of Stay", x = "Length of St
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
Warning: Removed 1 row containing missing values or values outside the scale range
(`geom_bar()`).
```

Percentage of Cases Resulting in No Use by Length of Stay


```r
# Reshape the data to long format
no_use_by_los_long <- no_use_by_los %>%
  pivot_longer(cols = c(percentage_no_use, percentage_completed),
               names_to = "Outcome",
               values_to = "Percentage")

# Create the combined plot
combined_plot_los <- ggplot(data = no_use_by_los_long, aes(x = los, y = Percentage, fill = Ou
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Percentage of Cases by Length of Stay",
       x = "Length of Stay",
       y = "Percentage") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_manual(values = c("skyblue", "salmon")) +   # Customize bar colors
  scale_color_manual(values = c("blue", "red"))          # Customize line colors

# Display the plot
print(combined_plot_los)
```

Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_bar()`).

Percentage of Cases by Length of Stay