Titanic

2025-03-24

Titanic Dataset Analysis

```
crude_summary = full %>%
  select(PassengerId, Survived) %>%
  group_by(Survived) %>%
  summarise(n = n()) %>%
  mutate(freq = n / sum(n))
crude_summary
```

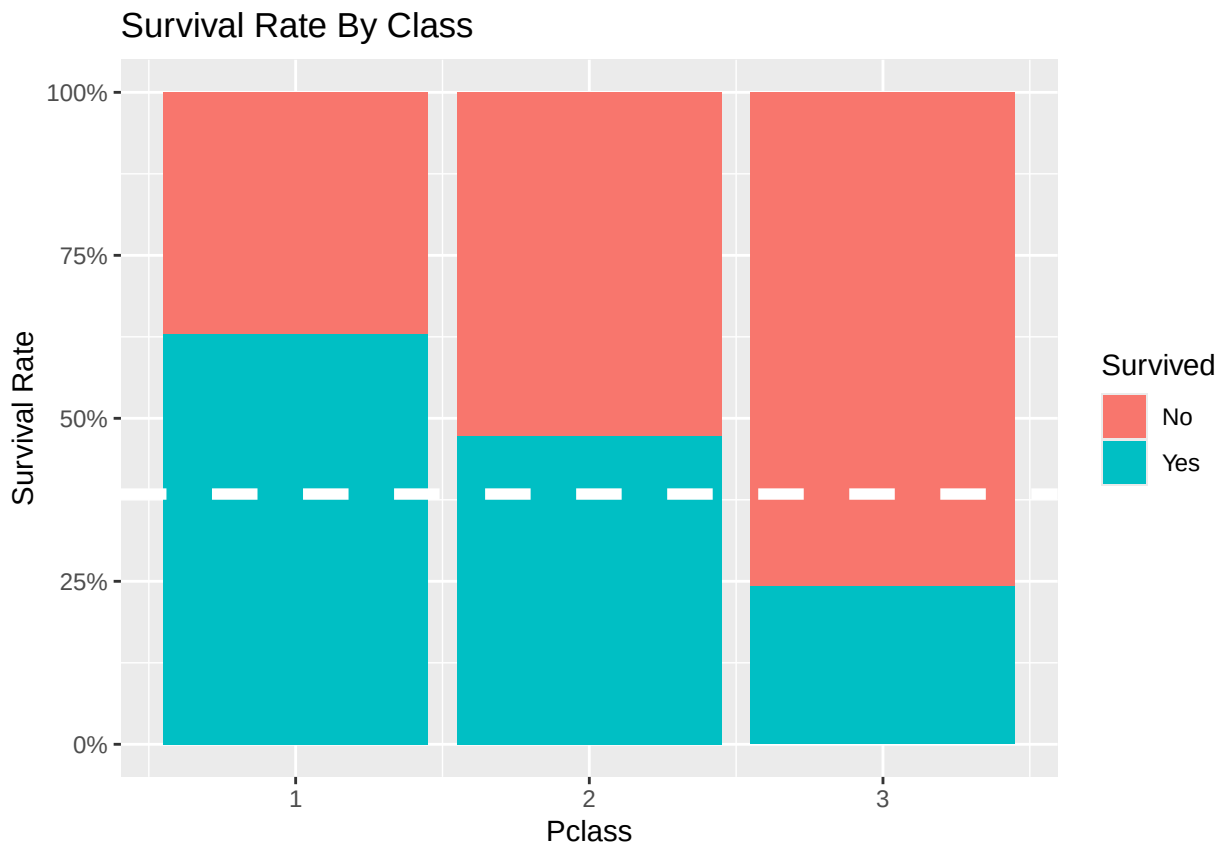
```
## # A tibble: 2 x 3
##   Survived     n  freq
##   <chr>    <int> <dbl>
## 1 No         549 0.616
## 2 Yes        342 0.384
```

```
crude_survrate = crude_summary$freq[crude_summary$Survived == "Yes"]
crude_survrate
```

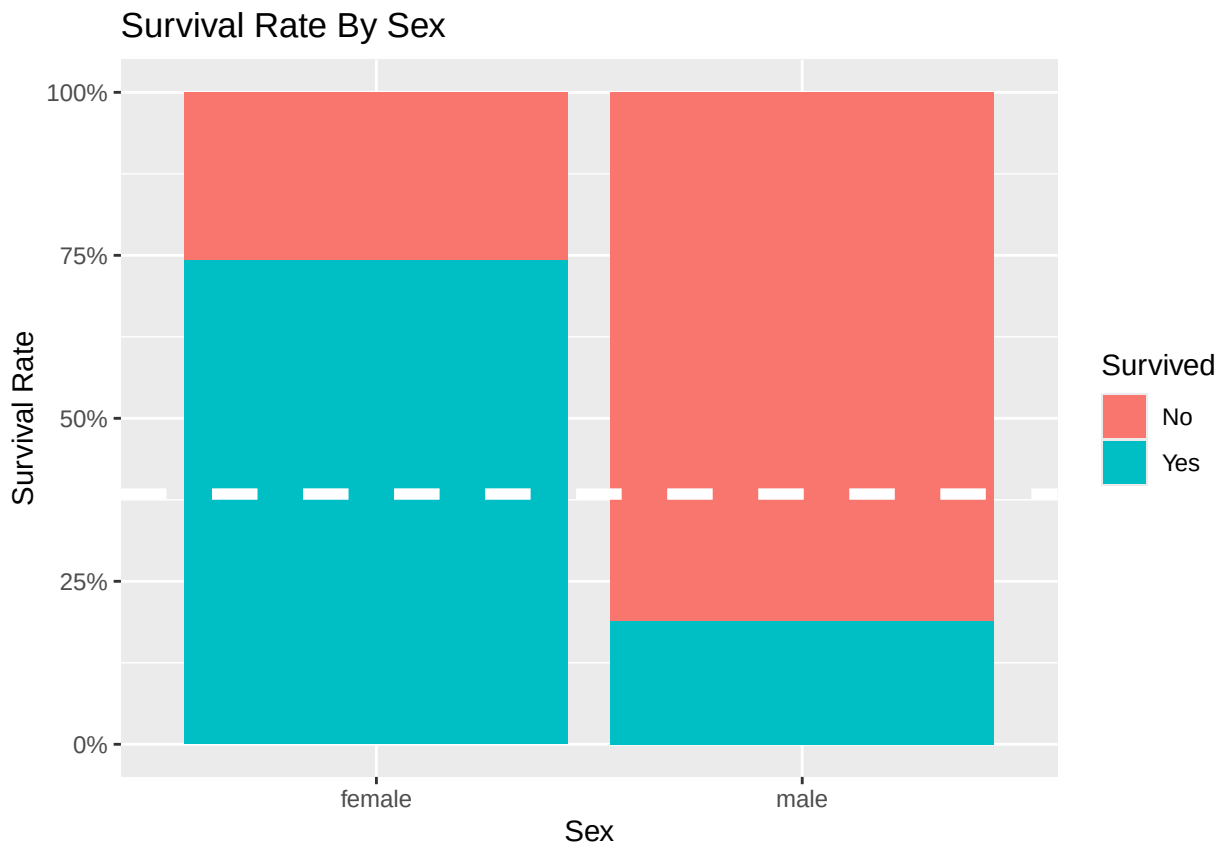
```
## [1] 0.3838384
```

```
ggplot(full, aes(Pclass, fill = Survived)) + geom_bar(position = "Fill") + scale_y_continuous(label = p
  ggtitle("Survival Rate By Class") + geom_hline(yintercept = crude_survrate, col = "white", size = 2, l
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```




```
ggplot(full, aes(Sex, fill = Survived)) + geom_bar(position = "Fill") + scale_y_continuous(label = perc
  ggtitle("Survival Rate By Sex") + geom_hline(yintercept = crude_survrate, col = "white", size = 2, lty
```



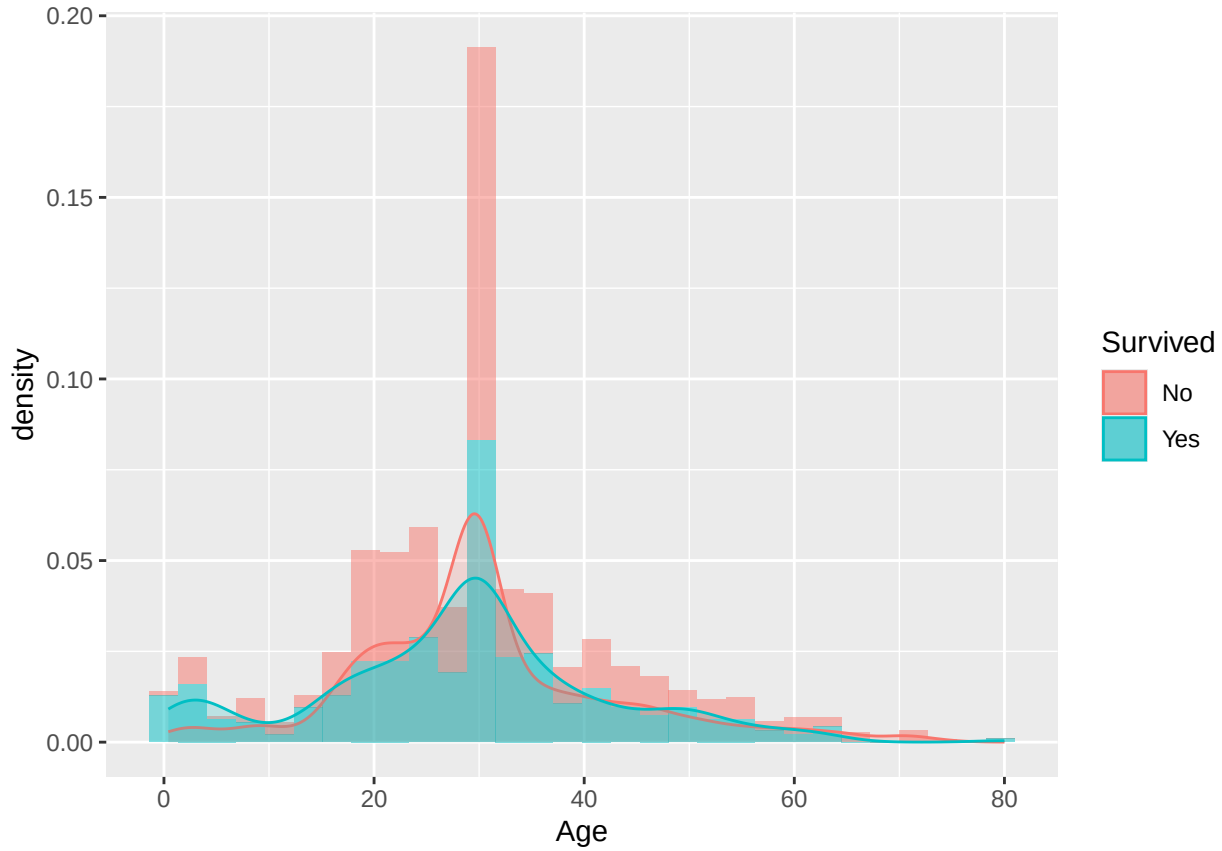

```
tbl_age = full %>%
  select(Age, Survived) %>%
  group_by(Survived) %>%
  summarise(mean.age = mean(Age, na.rm = T))
tbl_age
```

```
## # A tibble: 2 x 2
##   Survived mean.age
##   <chr>       <dbl>
## 1 No           30.4
## 2 Yes          28.5
```

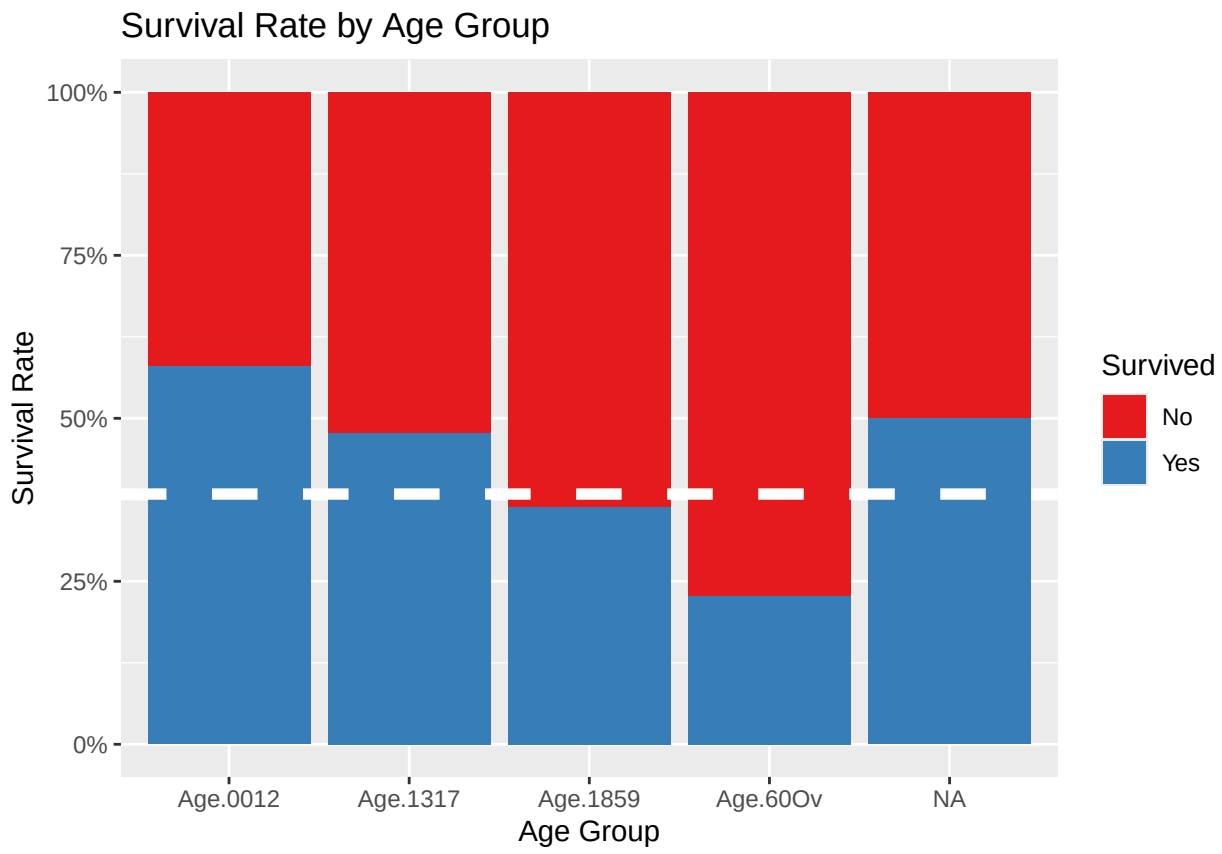
```
ggplot(full, aes(Age, fill = Survived)) + geom_histogram(aes(y = ..density..), alpha = 0.5) +
  geom_density(alpha = 0.2, aes(color = Survived))
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

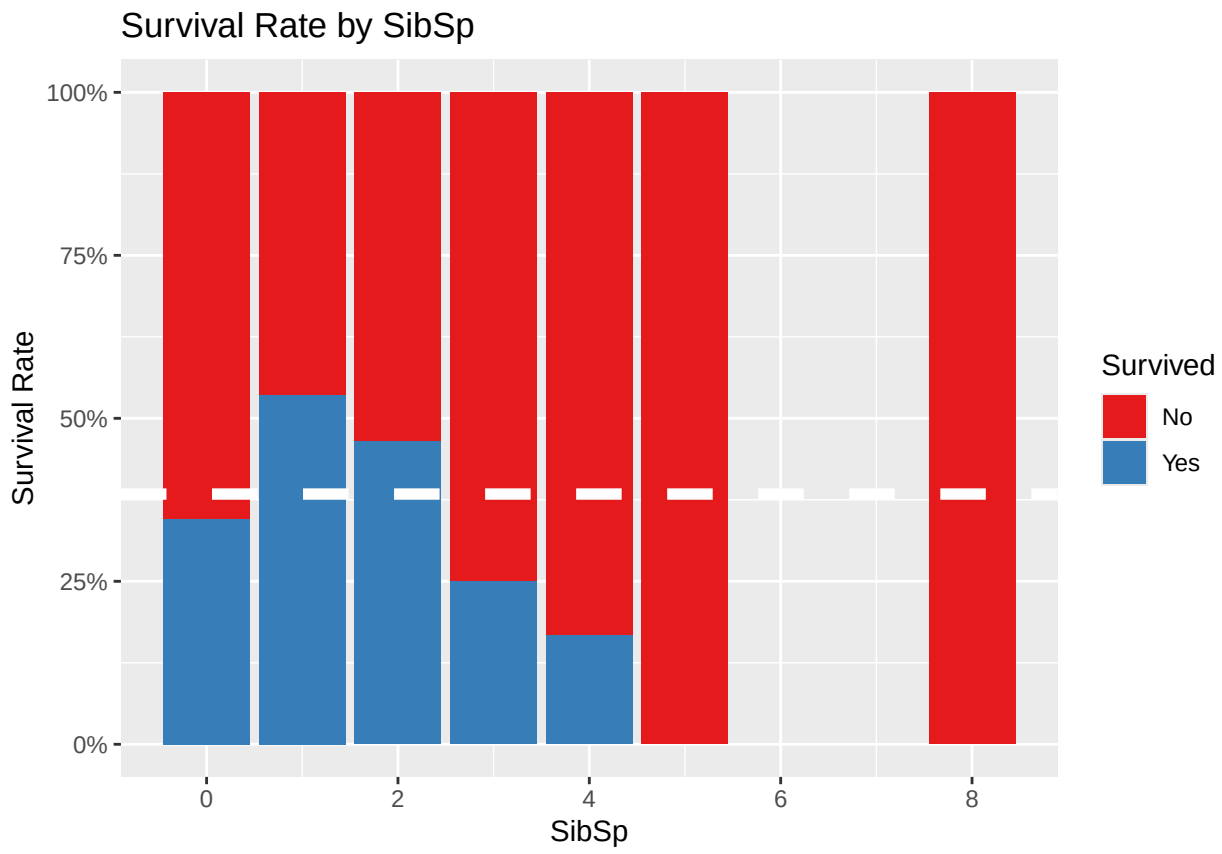
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



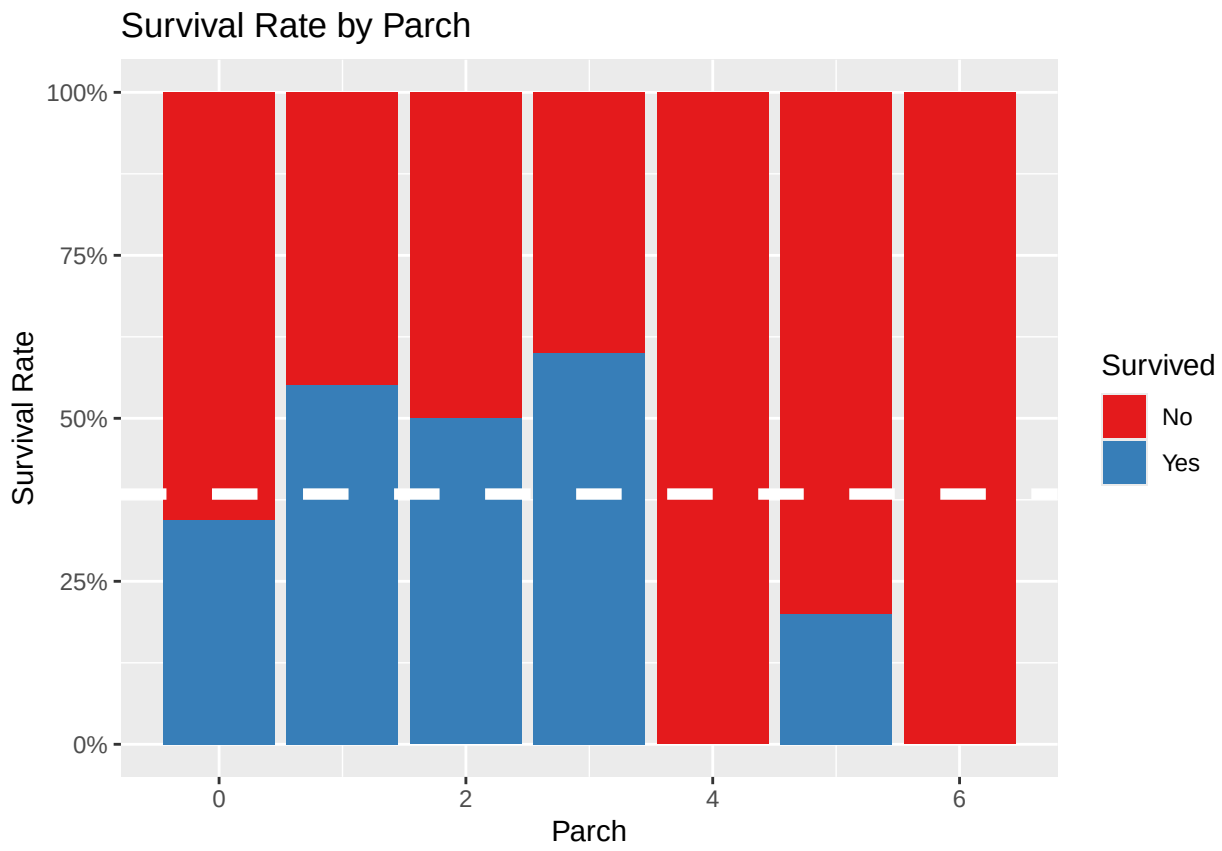

```
ggplot(full %>% filter( !is.na(Age)),
       aes(`Age Group`, fill = Survived)) + geom_bar(position = 'fill') +
  scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(labels = percent) +
  ylab("Survival Rate") +
  geom_hline(yintercept = crude_survrate, col = "white", lty = 2, size = 2) +
  ggtitle("Survival Rate by Age Group")
```



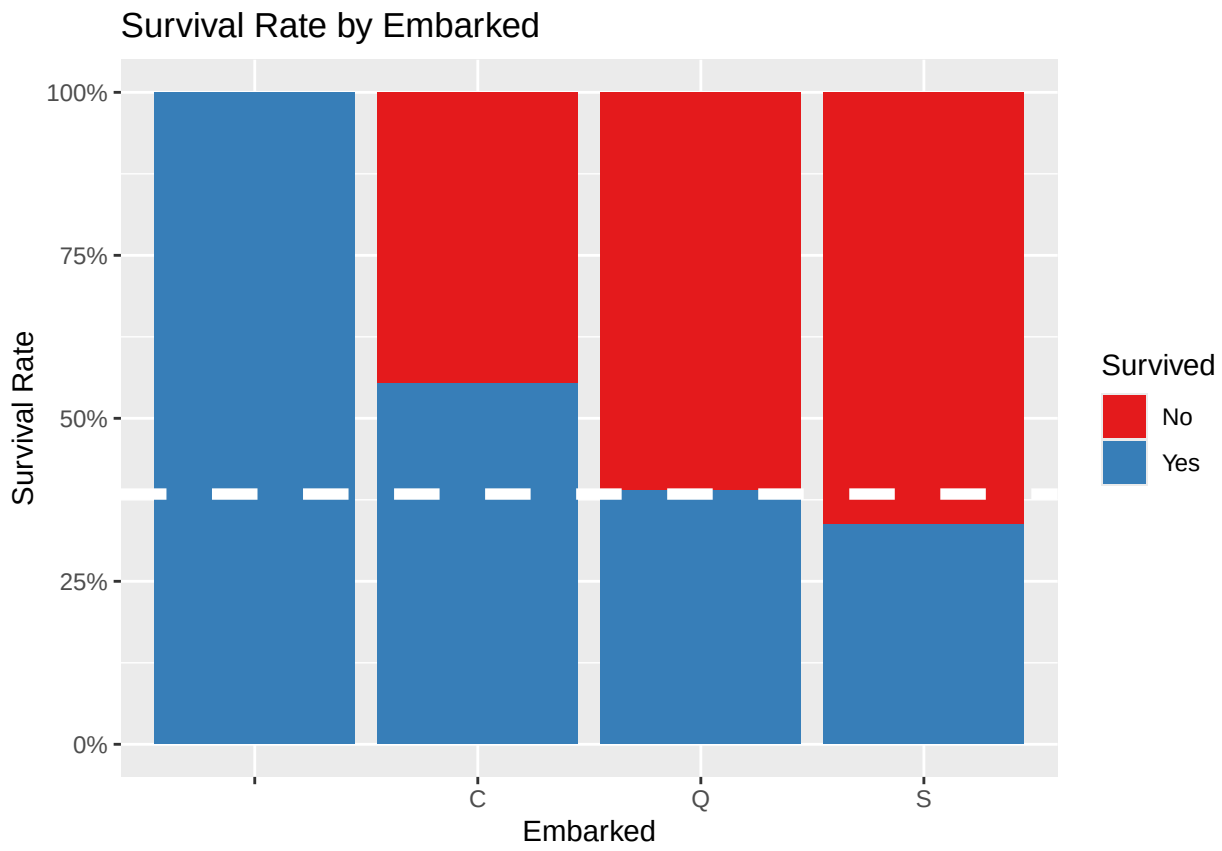

```
ggplot(full,
       aes(SibSp, fill = Survived)) + geom_bar(position = 'fill') +
  scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(labels = percent) +
  ylab("Survival Rate") +
  geom_hline(yintercept = crude_survrate, col = "white", lty = 2, size = 2) +
  ggtitle("Survival Rate by SibSp")
```



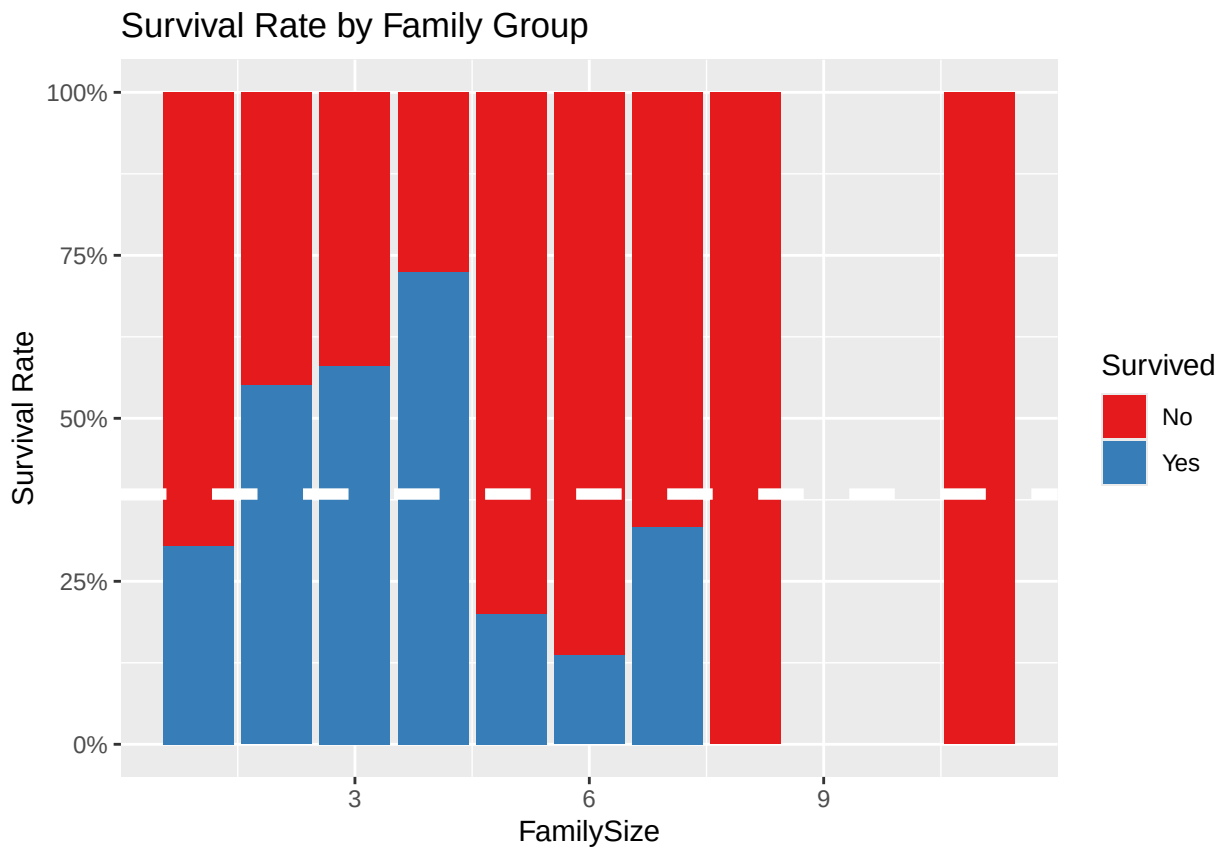

```
ggplot(full,
       aes(Parch, fill = Survived)) + geom_bar(position = 'fill') +
  scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(labels = percent) +
  ylab("Survival Rate") +
  geom_hline(yintercept = crude_survrate, col = "white", lty = 2, size = 2) +
  ggtitle("Survival Rate by Parch")
```



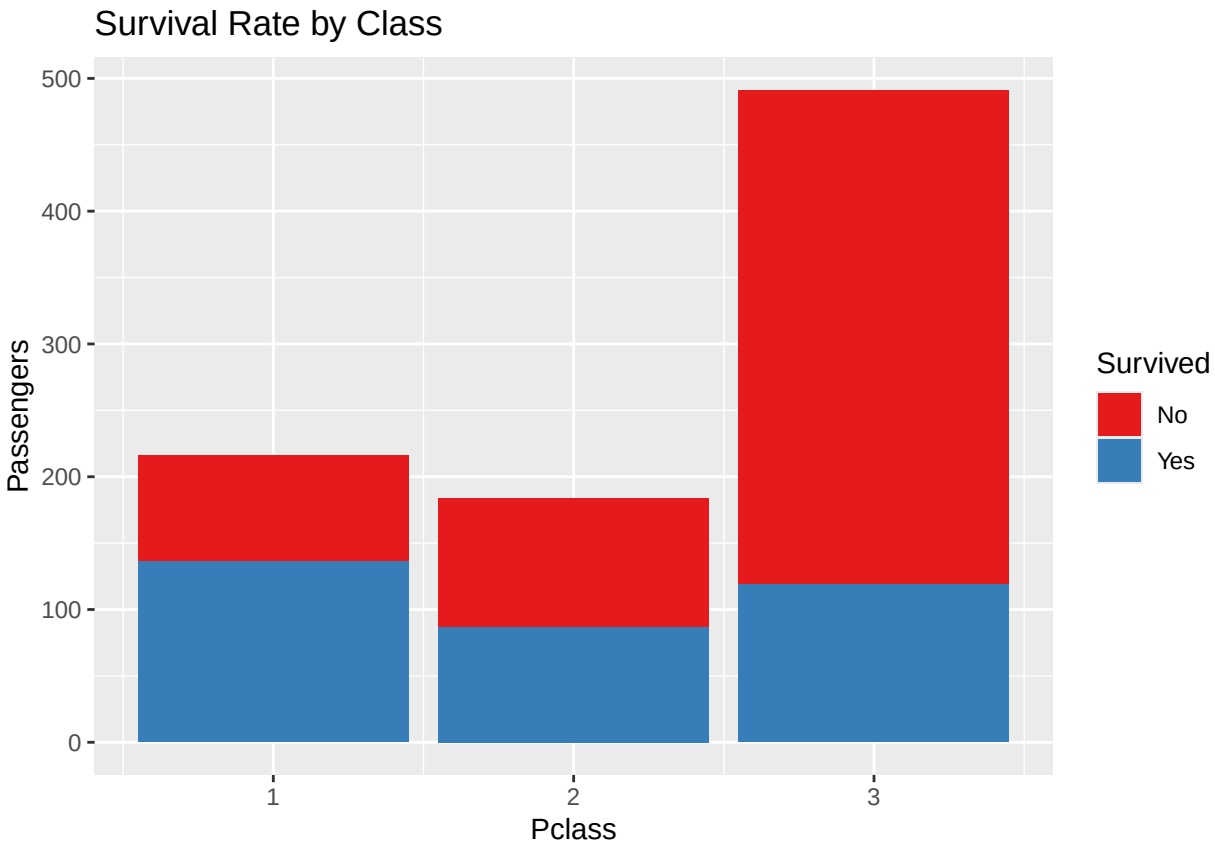

```
ggplot(full,
       aes(Embarked, fill = Survived)) + geom_bar(position = 'fill') +
  scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(labels = percent) +
  ylab("Survival Rate") +
  geom_hline(yintercept = crude_survrate, col = "white", lty = 2, size = 2) +
  ggtitle("Survival Rate by Embarked")
```



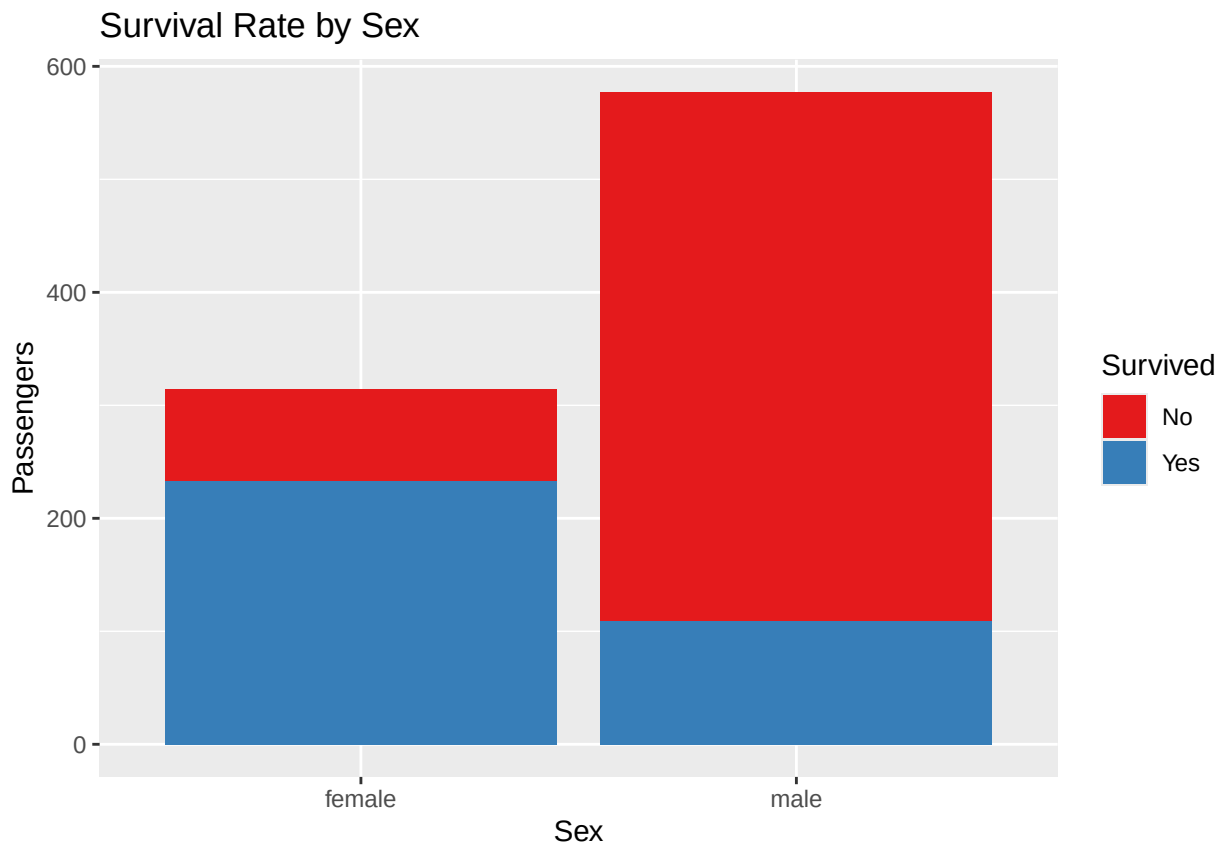

```
ggplot(full %>% na.omit,
       aes(`FamilySize`, fill = Survived)) + geom_bar(position = 'fill') +
  scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(labels = percent) +
  ylab("Survival Rate") +
  geom_hline(yintercept = crude_survrate, col = "white", lty = 2, size = 2) +
  ggtitle("Survival Rate by Family Group")
```




```
ggplot(full,
       aes(Pclass, fill = Survived)) + geom_bar(position = 'stack') +
  scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(labels = comma) +
  ylab("Passengers") +
  ggtitle("Survival Rate by Class")
```

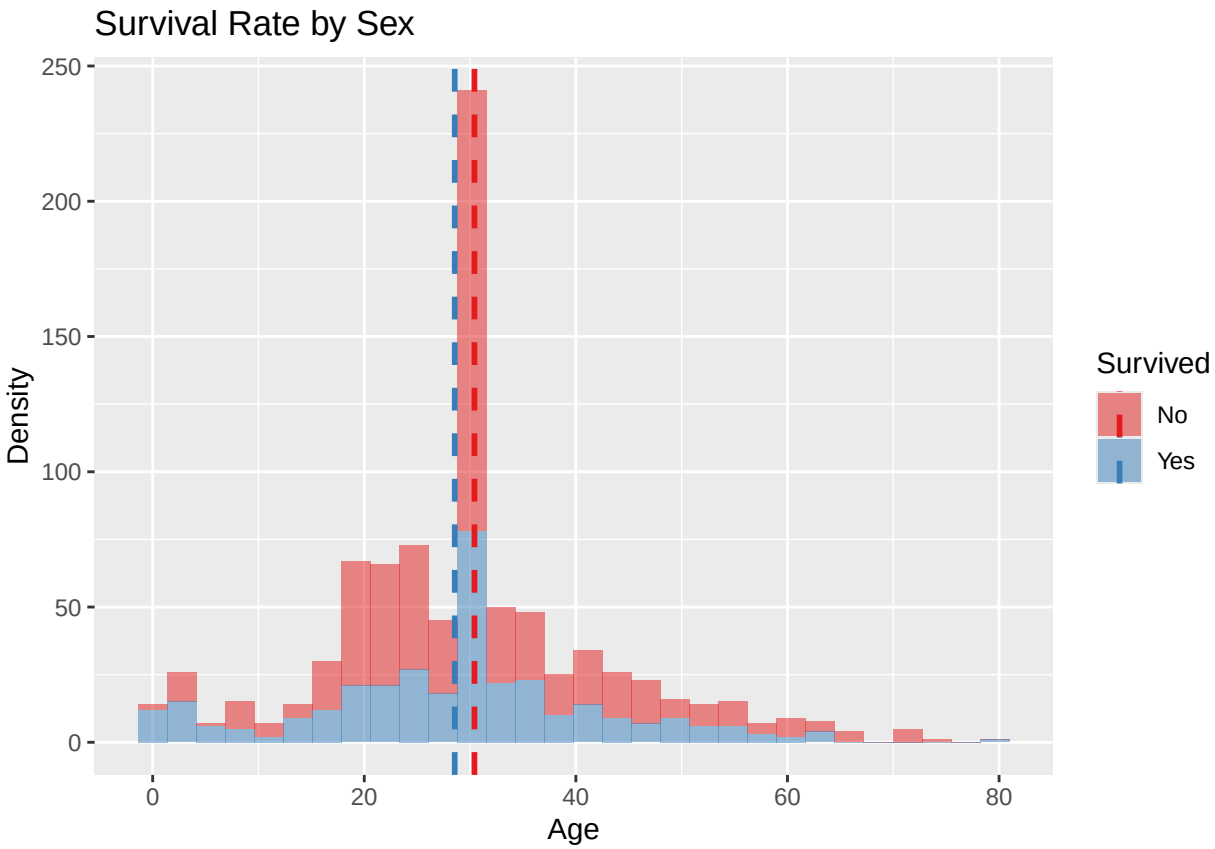



```
ggplot(full,
       aes(Sex, fill = Survived)) + geom_bar(position = 'stack') +
  scale_fill_brewer(palette = "Set1") +
  scale_y_continuous(labels = comma) +
  ylab("Passengers") +
  ggtitle("Survival Rate by Sex")
```




```
ggplot(full,
       aes(Age, fill = Survived)) + geom_histogram(aes(y = ..count..), alpha = 0.5) +
  geom_vline(data = tbl_age, aes(xintercept = mean.age, colour = Survived), lty=2, size=1) +
  scale_fill_brewer(palette = "Set1") +
  scale_colour_brewer(palette = "Set1") +
  scale_y_continuous(labels = comma) +
  ylab("Density") +
  ggtitle("Survival Rate by Sex")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Survival Rate by Sex
Density
250
200
150
100
50
0
0
20
40
60
80
Age
Survived
No
Yes