Lab 12 - Statistics, Coordinates, Facets, and Themes

William Bernard

November 14, 2017

Complete the following exercises below. Knit together the PDF document and commit both the Lab 12 RMDfile and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Loading required package: lattice

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

##
## Attaching package: 'memisc'

## The following objects are masked from 'package:dplyr':
##
##     collect, recode, rename

## The following objects are masked from 'package:stats':
##
##     contr.sum, contr.treatment, contrasts

## The following object is masked from 'package:base':
##
##     as.array
```

1. Choose one or more graphics you created for Lab 11 and either experiment with the underlying statistical layer if it already has one (i.e. if you made a histogram experiement with different bin widths) or add a separate statistical layer to your plot (i.e. a smoothing curve). Choose something you think will offer meaningful insight and describe why you made the choice you did. What additional information does this provide viewers of your graphic?
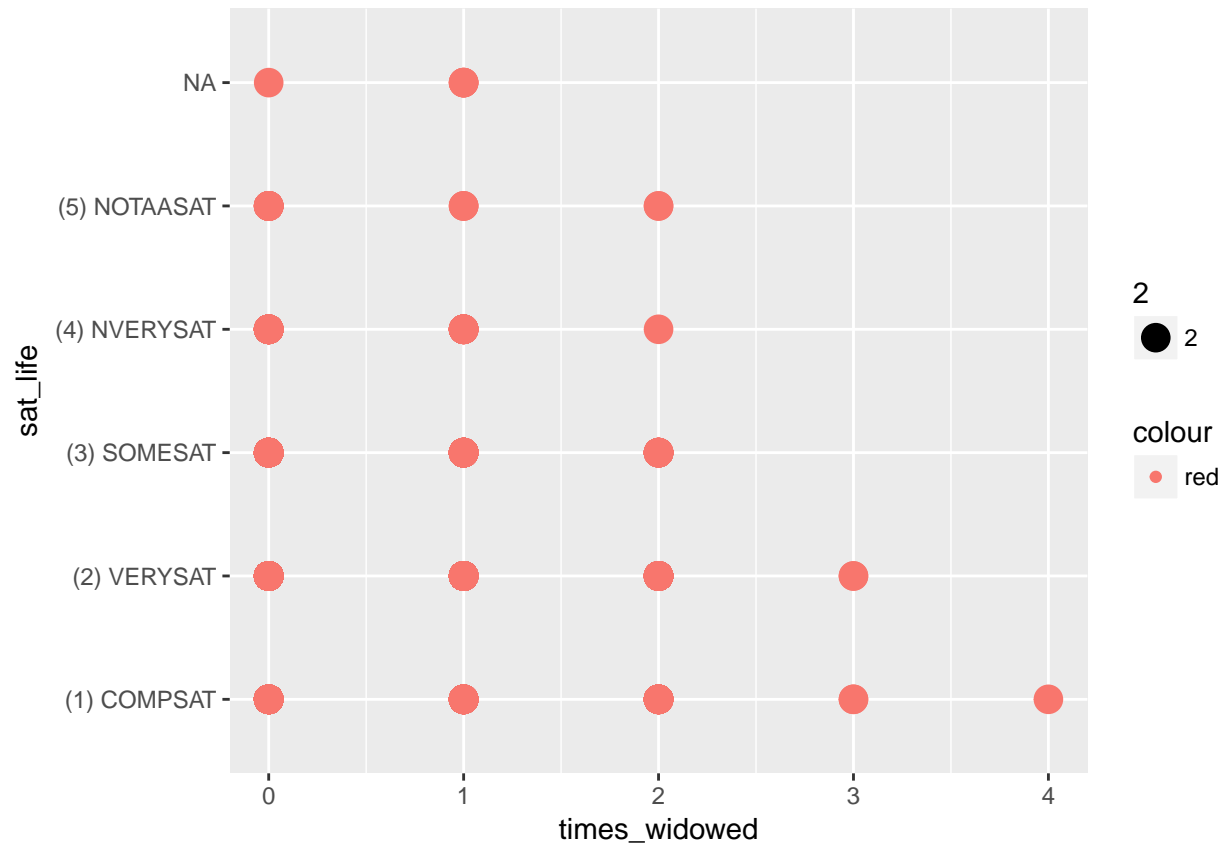
```
ggplot(data = changing_lives_subset,
       aes(x = times_widowed, y = sat_life, col = "red",)) +
         geom_point()
```

```
## Warning: Removed 2795 rows containing missing values (geom_point).
```
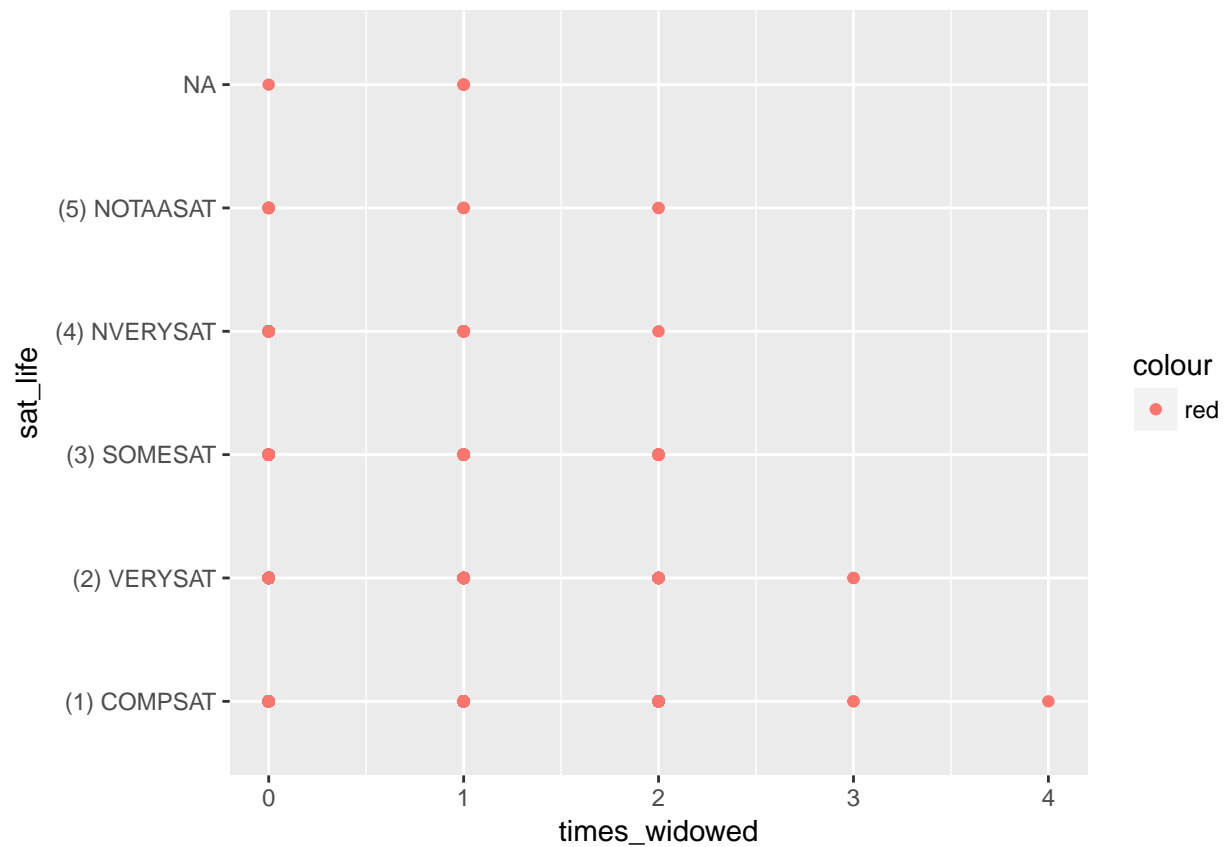
```
ggplot(data = changing_lives_subset,
       aes(x = times_widowed, y = sat_life, col = "red", size = 2)) +
    geom_point()
```

Warning: Removed 2795 rows containing missing values (geom_point).

```
ggplot(data = changing_lives_subset,
       aes(x = times_widowed, y = sat_life, col = "red")) +
          geom_point(position = "jitter") +
  geom_boxplot(colour = "black")
```

Warning: Removed 2795 rows containing non-finite values (stat_boxplot).

Warning: position_dodge requires non-overlapping x intervals

## Warning: Removed 2795 rows containing missing values (geom_point).
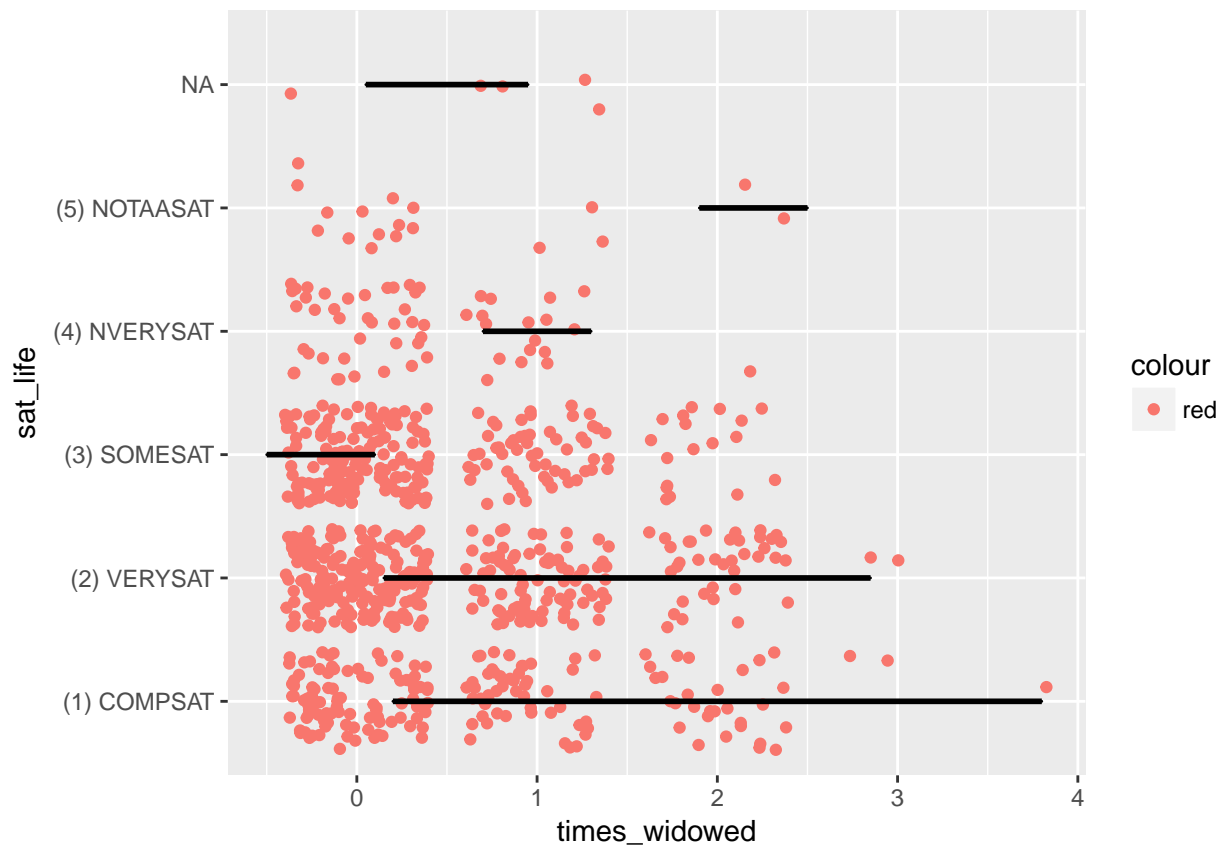
I thought adding a geom_boxplot would add to the visualization. it is able to show the distribution of the points as a whole more effectivly, this is espeically imprtant considering a jitter effect was used to make the plot viewable in the first place.

2. With the same or a different plot created in Lab 11, experiment with zooming in on specific areas of your graphic and changing the aspect ratio. Are their any benefits/drawbacks with either or both of these approaches for the visualizations you've created? What are they?

```
ggplot(data = changing_lives_subset,
       aes(x = times_widowed, y = sat_life, col = "red")) +
         geom_point(position = "jitter") +
  geom_boxplot(colour = "black") +
  coord_cartesian(xlim = 1)
```

Warning: Removed 2795 rows containing non-finite values (stat_boxplot).

Warning: position_dodge requires non-overlapping x intervals

## Warning: Removed 2795 rows containing missing values (geom_point).
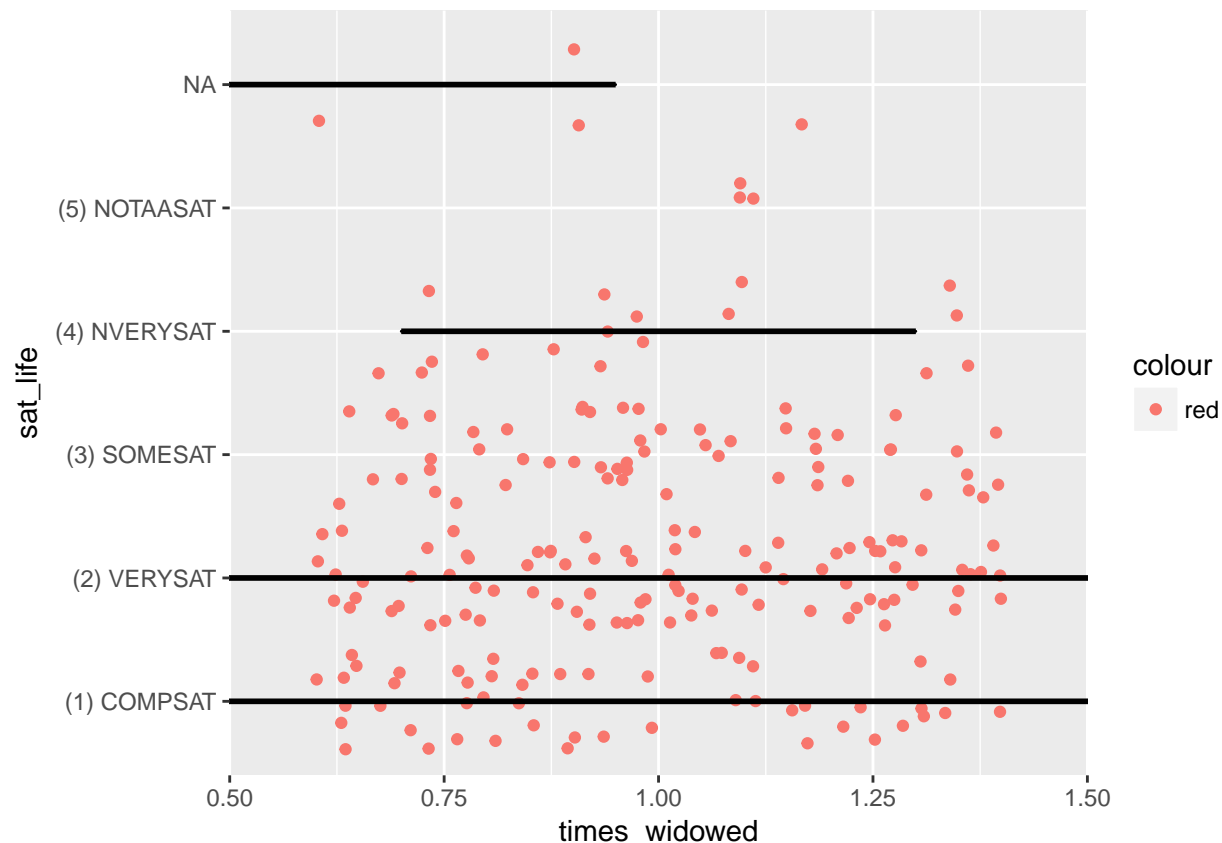
With this plot, I have zoomed into only the x = 1 variable. With this new plot we are only looking at the data points pertaining to individuals who have been widowed once. This allows us to more closely and accurately look at the data of a specific subset of the data. stretching the data would by changing the aspect ratio would not be helpful in this case because each x value is descrete.

3. Try facetting a plot you have made by another categorical variable in your data (this can even be as simple as Male/Female). What is the difference between `facet_wrap()` and `facet_grid()`? How might facetting be useful in data visualization?

```
ggplot(data = changing_lives_subset,
       aes(x = times_widowed, y = sat_life, col = "red")) +
        geom_point(position = "jitter") +
  facet_grid(. ~sex_r)
```

```
## Warning: Removed 2795 rows containing missing values (geom_point).
```
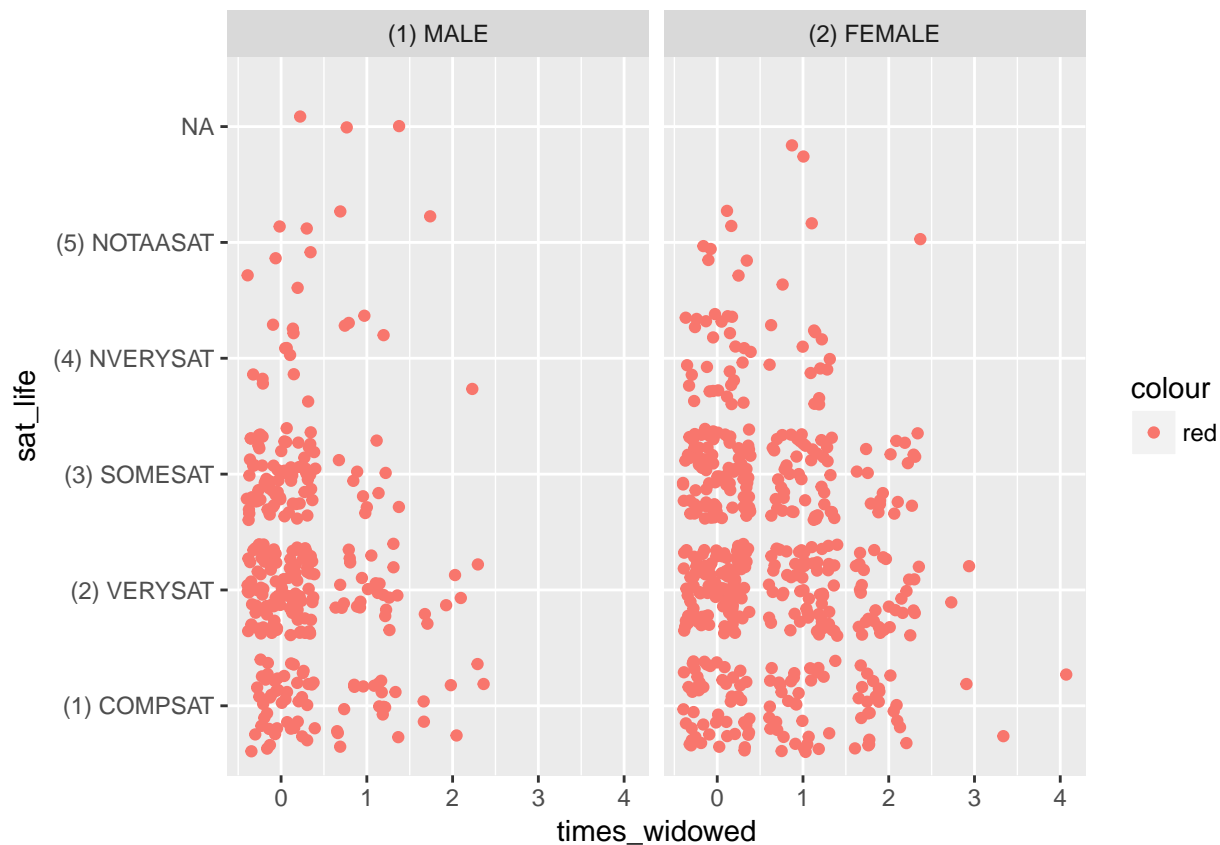
This is super helpful for data visualization. It allows you to look at the data in the context of another variable and orginize it in a way that lets you easily compare the two.

4. Use the `theme()` layer to change the appearance of a plot of your choice including the

- plot, axes, and legend titles
- axes tick marks
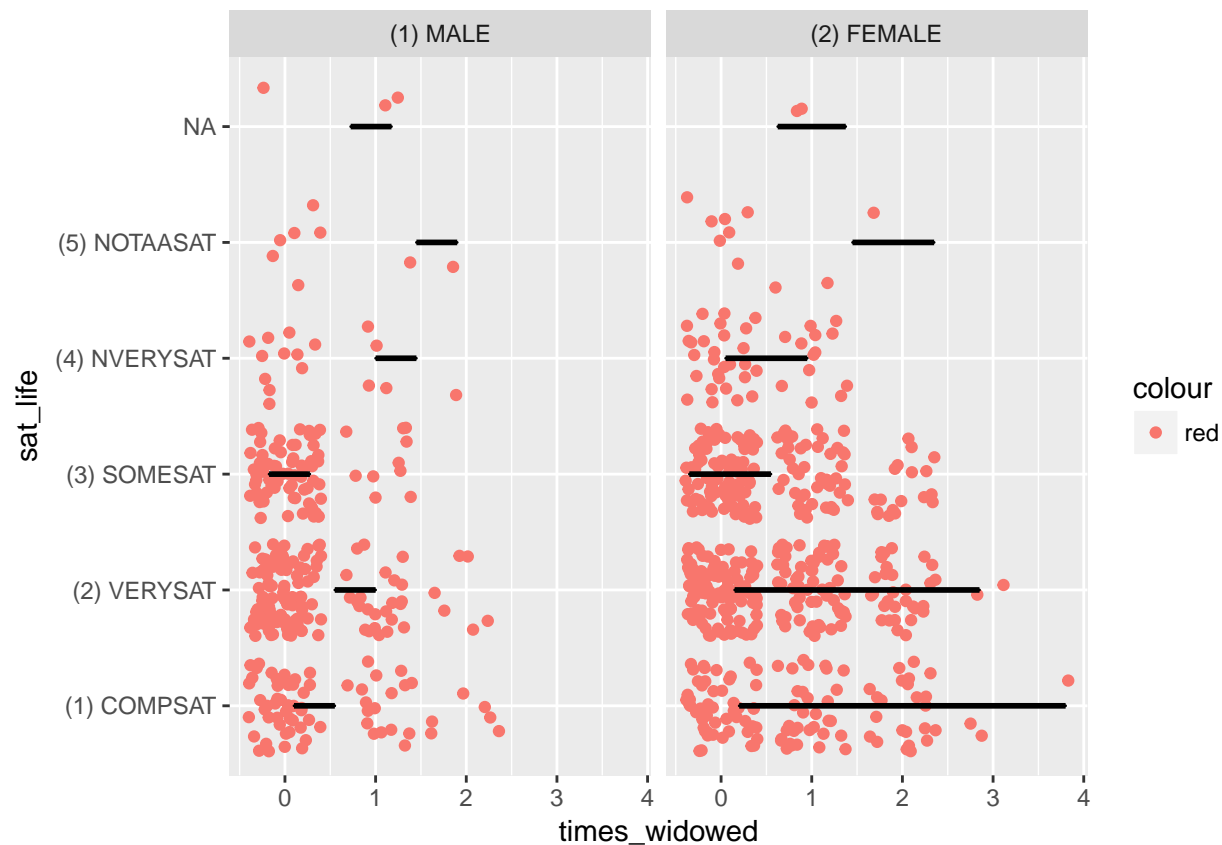- text size
- legend position

```
ggplot(data = changing_lives_subset,
       aes(x = times_widowed, y = sat_life, col = "red")) +
         geom_point(position = "jitter") +
  geom_boxplot(colour = "black") +
  facet_grid(. ~sex_r)
```

```
## Warning: Removed 2795 rows containing non-finite values (stat_boxplot).

## Warning: position_dodge requires non-overlapping x intervals

## Warning: position_dodge requires non-overlapping x intervals

## Warning: Removed 2795 rows containing missing values (geom_point).
```

```
theme_bernard <- theme(
  panel.background = element_blank(),
  legend.background = element_blank(),
  axis.title = element_text(colour = "red", size = 10),
  axis.text = element_text(colour = "black", size = 5),
  legend.position = "bottom"


)


ggplot(data = changing_lives_subset,
       aes(x = times_widowed, y = sat_life, col = "red")) +
         geom_point(position = "jitter") +
  geom_boxplot(colour = "black") +
  facet_grid(. ~sex_r) +
  theme_bernard
```
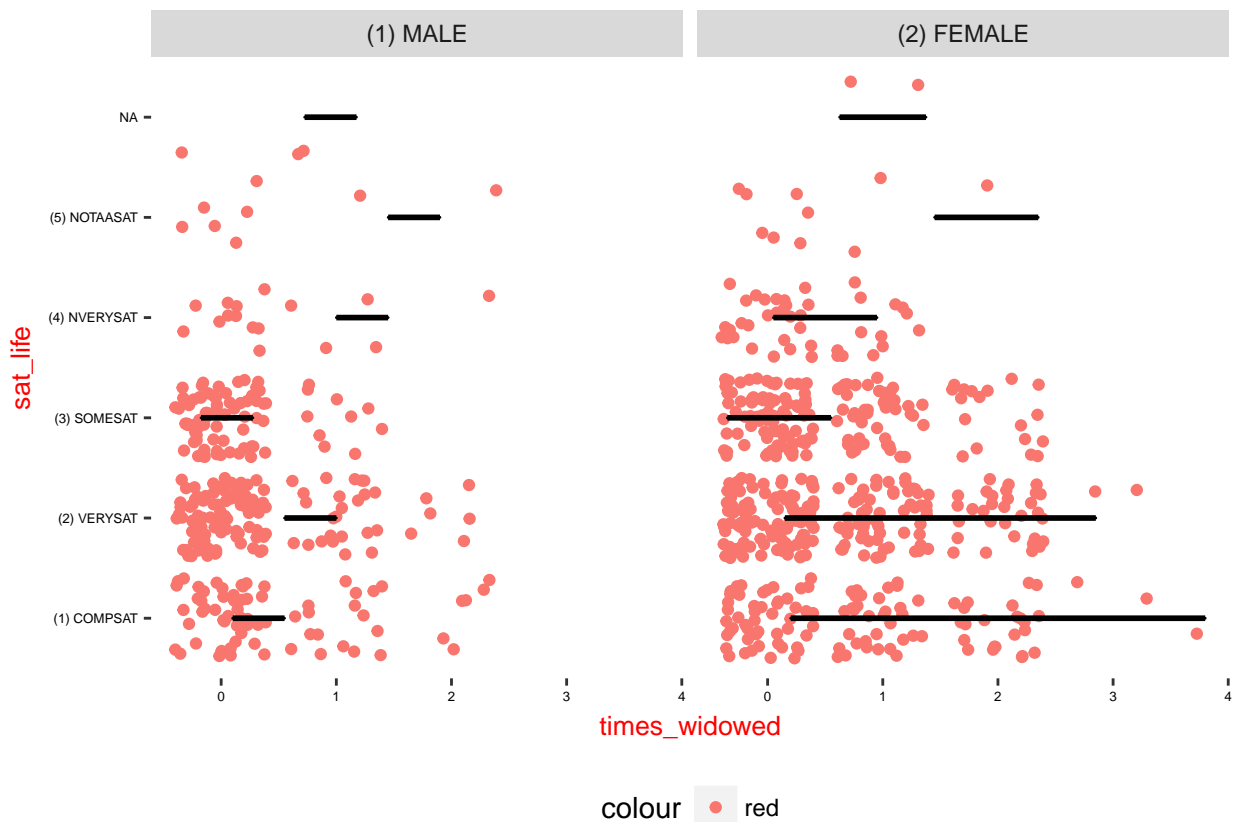
```
## Warning: Removed 2795 rows containing non-finite values (stat_boxplot).

## Warning: position_dodge requires non-overlapping x intervals

## Warning: position_dodge requires non-overlapping x intervals

## Warning: Removed 2795 rows containing missing values (geom_point).
```

5. Create three versions of a graphic of your choice using different built-in themes or a theme created from **ggthemes**. Which ones do you think are best for presenting in an academic journal? A poster session? What are the qualities of the themes that you choice that you think make them more appropriate for presentation?

```
library(ggthemes)

plot_one <- ggplot(data = changing_lives_subset,
      aes(x = times_widowed, y = sat_life, col = "red")) +
        geom_point(position = "jitter") +
  geom_boxplot(colour = "black") +
  facet_grid(. ~sex_r)

plot_one + theme_stata()
```
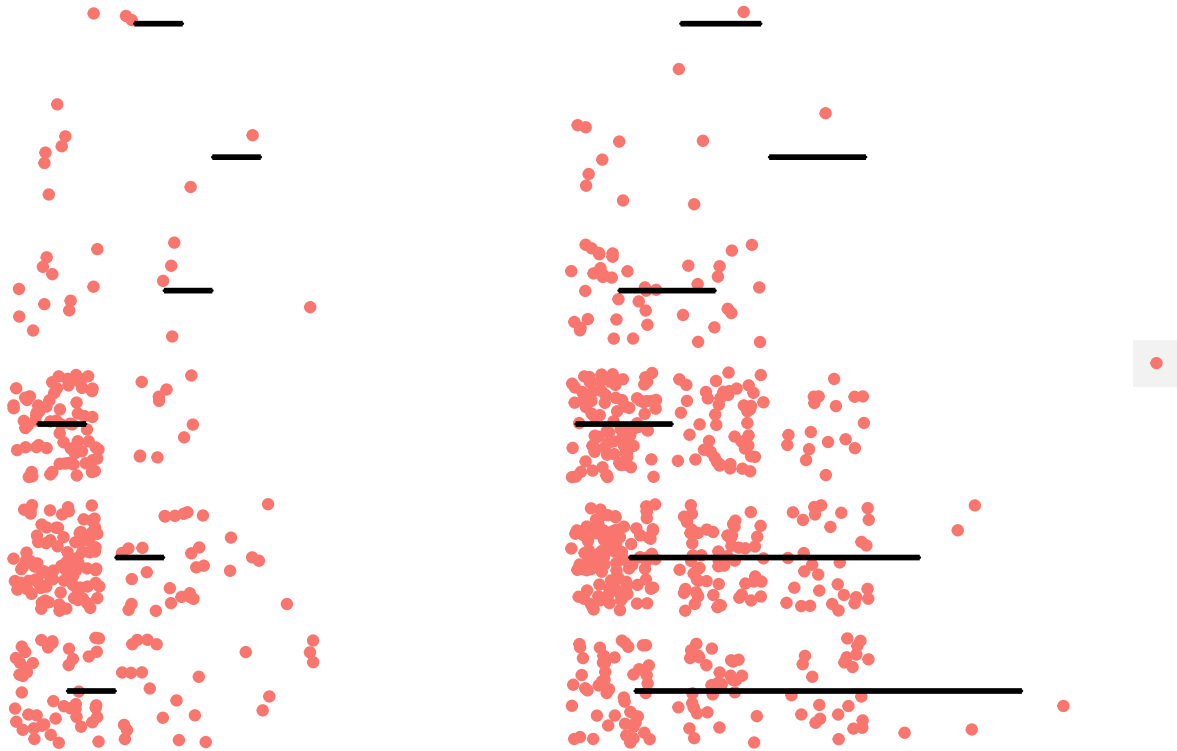
Warning: Removed 2795 rows containing non-finite values (stat_boxplot).

Warning: position_dodge requires non-overlapping x intervals

Warning: position_dodge requires non-overlapping x intervals

Warning: Removed 2795 rows containing missing values (geom_point).
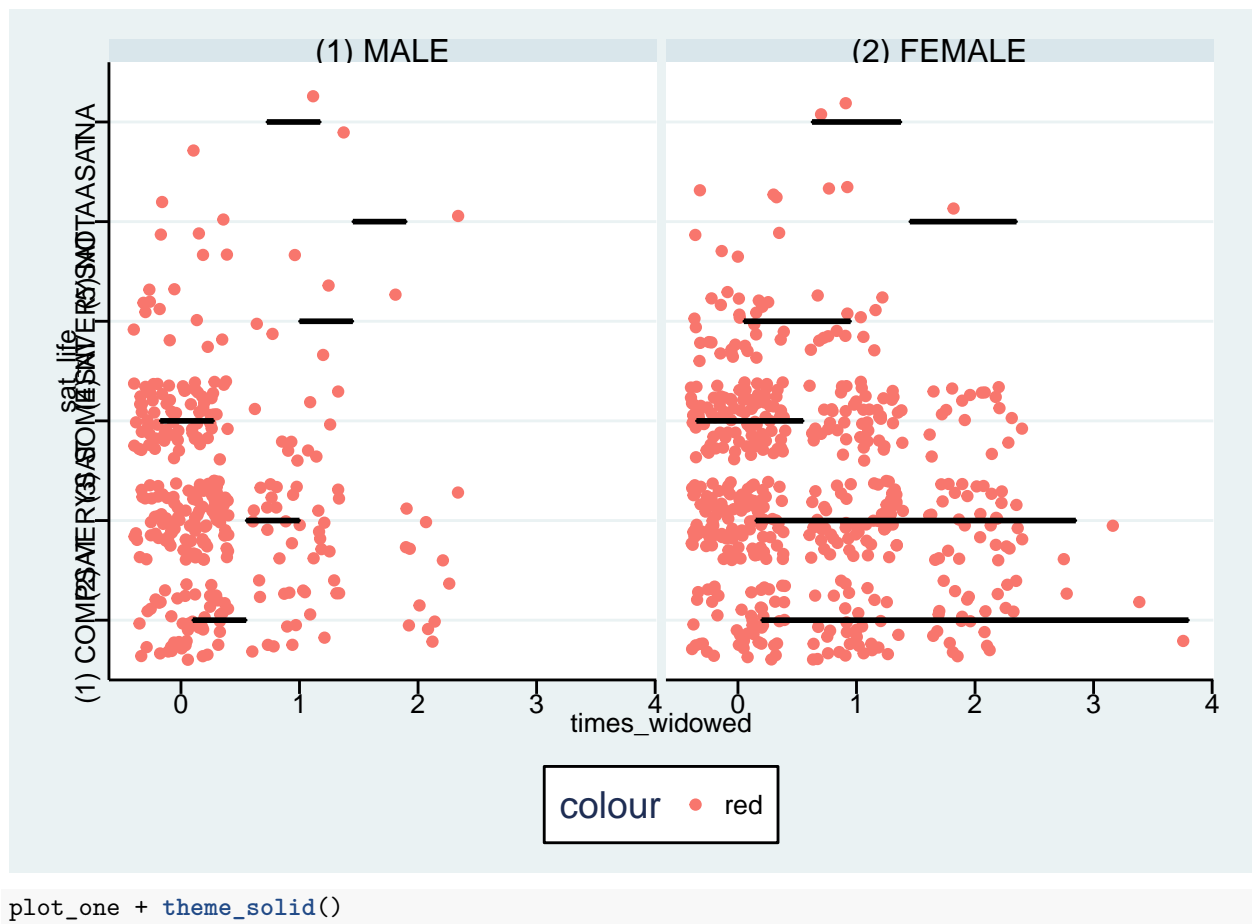
```
plot_one + theme_solid()
```

Warning: Removed 2795 rows containing non-finite values (stat_boxplot).

Warning: position_dodge requires non-overlapping x intervals

Warning: position_dodge requires non-overlapping x intervals

Warning: Removed 2795 rows containing missing values (geom_point).

```
plot_one + theme_tufte()
```

Warning: Removed 2795 rows containing non-finite values (stat_boxplot).

Warning: position_dodge requires non-overlapping x intervals

Warning: position_dodge requires non-overlapping x intervals

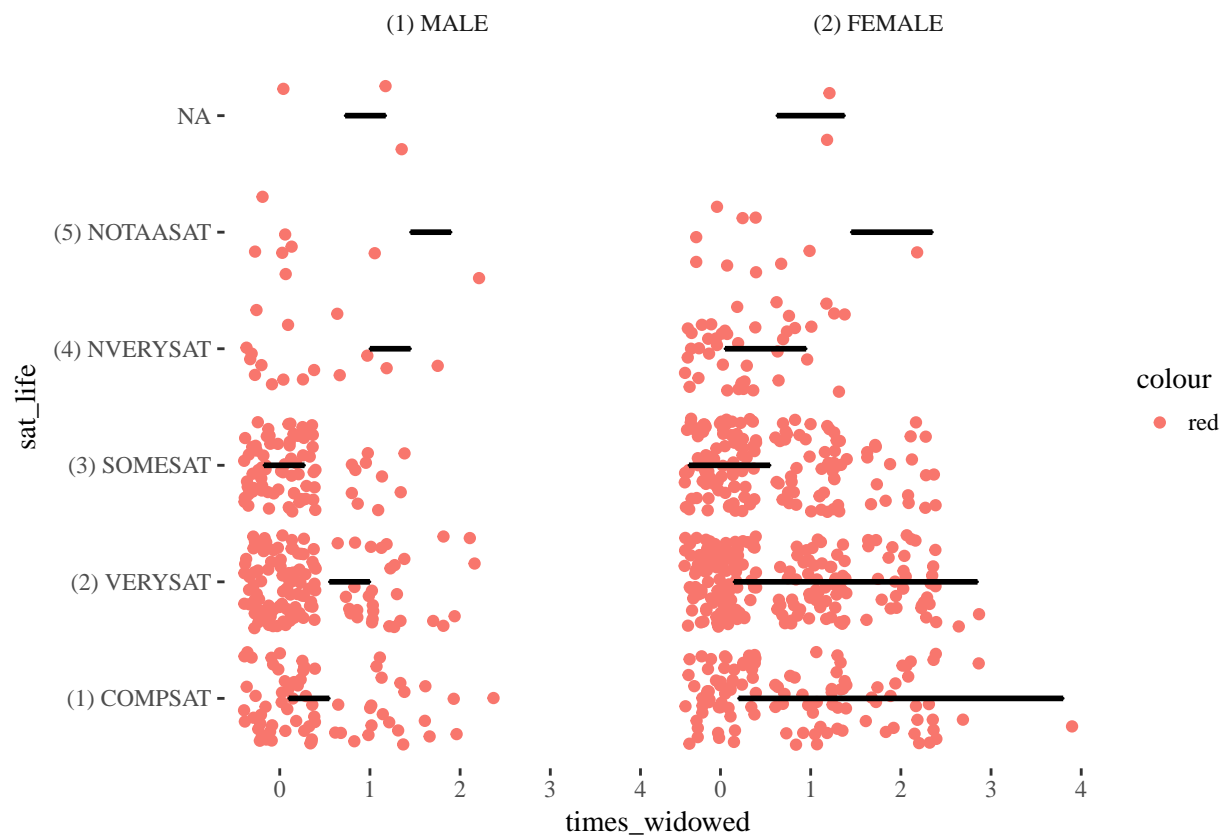## Warning: Removed 2795 rows containing missing values (geom_point).

Out of the three I personally enjoy theme_tufte the most. Primarily because it isnt oddly formatted because of the conversion(in the case of theme_stata), and because there are actual descriptors(unlike with theme_solid). The font is visually appealing and the theme itself is minimalist, it would most likely work well for the poster project.