

Lab 12 - Statistics, Coordinates, Facets, and Themes

Snow Christensen

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.

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 experiment 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?

I changed the position of the bars to be stacked instead of dodged. This can help make a comparison between the two TRAIT variables that is sometimes difficult especially when the bars are very short and they are side by side. The `geom_bar()` function already has a statistical layer so I didn't need to change or add much.

```
#install.packages("magrittr")
library(magrittr)
```

```
#install.packages("dplyr")
library("dplyr")
```

```
##
```

```
## 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
```

```
#install.packages("tidyverse")
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
```

```
## Loading tidyverse: tibble
```

```
## Loading tidyverse: tidyr
```

```
## Loading tidyverse: readr
```

```
## Loading tidyverse: purrr
```

```
## Conflicts with tidy packages -----
```

```
## filter(): dplyr, stats
```

```
## lag(): dplyr, stats
```

```
#install.packages("haven")
library(haven)
```

```
#install.packages("tidyr")
library("tidyr")
```

```
#load full dataset
```

```

data <- read_por("~/Projects/sexual_violence_college/ICPSR_03212/DS0001/03212-0001-Data.por")

#data subset
data_ID <- data %>%
  dplyr::select(CODENUM, RACE, NOASSERT, PLEASE, PRESSSI, AUTHSI)

#convert columns to numerics
data_ID$CODENUM <- as.numeric(data_ID$CODENUM)

## Warning: NAs introduced by coercion
data_ID$RACE <- as.numeric(data_ID$RACE)

data_ID$NOASSERT<- as.numeric(data_ID$NOASSERT)

data_ID$PLEASE <- as.numeric(data_ID$PLEASE)

data_ID$AUTHSI <- as.numeric(data_ID$AUTHSI)

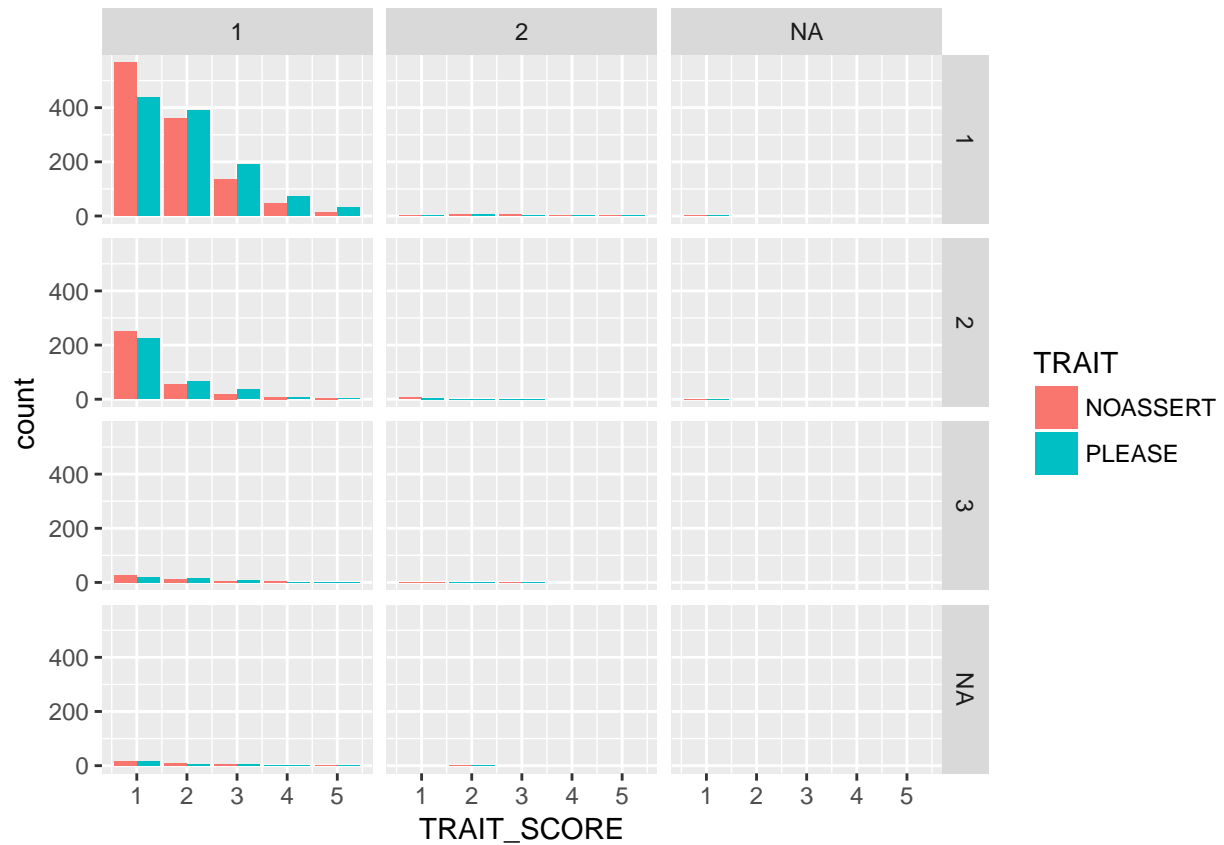
data_ID$PRESSSI <- as.numeric(data_ID$PRESSSI)

#gather PLEASE and NOASSERT into TRAIT_SCORE for visualization
data.gather <- data_ID %>%
  gather(key = "TRAIT", value = "TRAIT_SCORE", ... = -CODENUM, -AUTHSI, -PRESSSI, -RACE)

data_ID$RACE[data_ID$RACE == "1"] = "Black"
data_ID$RACE[data_ID$RACE == "2"] = "White"
data_ID$RACE[data_ID$RACE == "3"] = "Other"
data_ID$AUTHSI[data_ID$AUTHSI == "0"] = "No Response"
data_ID$AUTHSI[data_ID$AUTHSI == "1"] = "Never"
data_ID$AUTHSI[data_ID$AUTHSI == "2"] = "At least once"
data_ID$PRESSSI[data_ID$PRESSSI == "0"] = "No Response"
data_ID$PRESSSI[data_ID$PRESSSI == "1"] = "Never"
data_ID$PRESSSI[data_ID$PRESSSI == "2"] = "At least once"

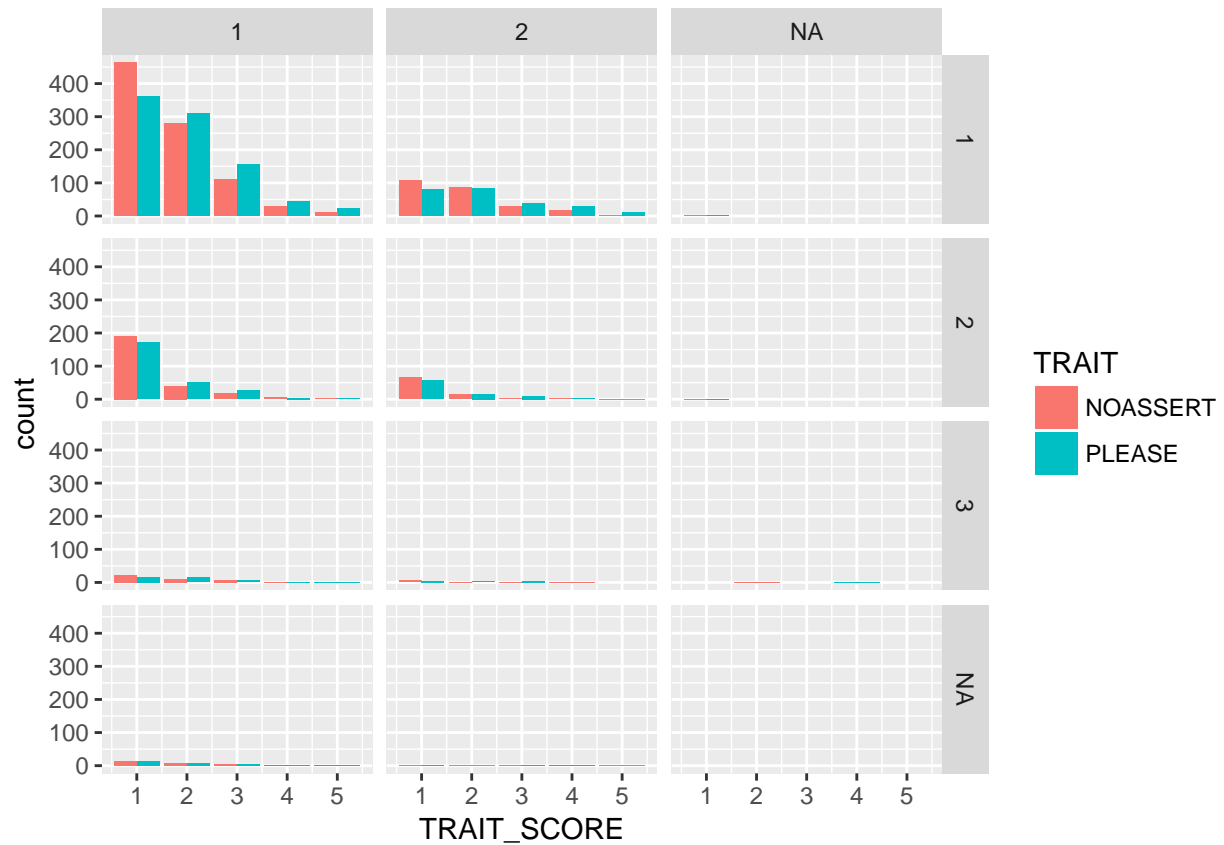
#plot TRAIT_SCORE facettted by RACE and AUTHSI
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge", na.rm = TRUE, width = ) +
  facet_grid(RACE~AUTHSI)

```

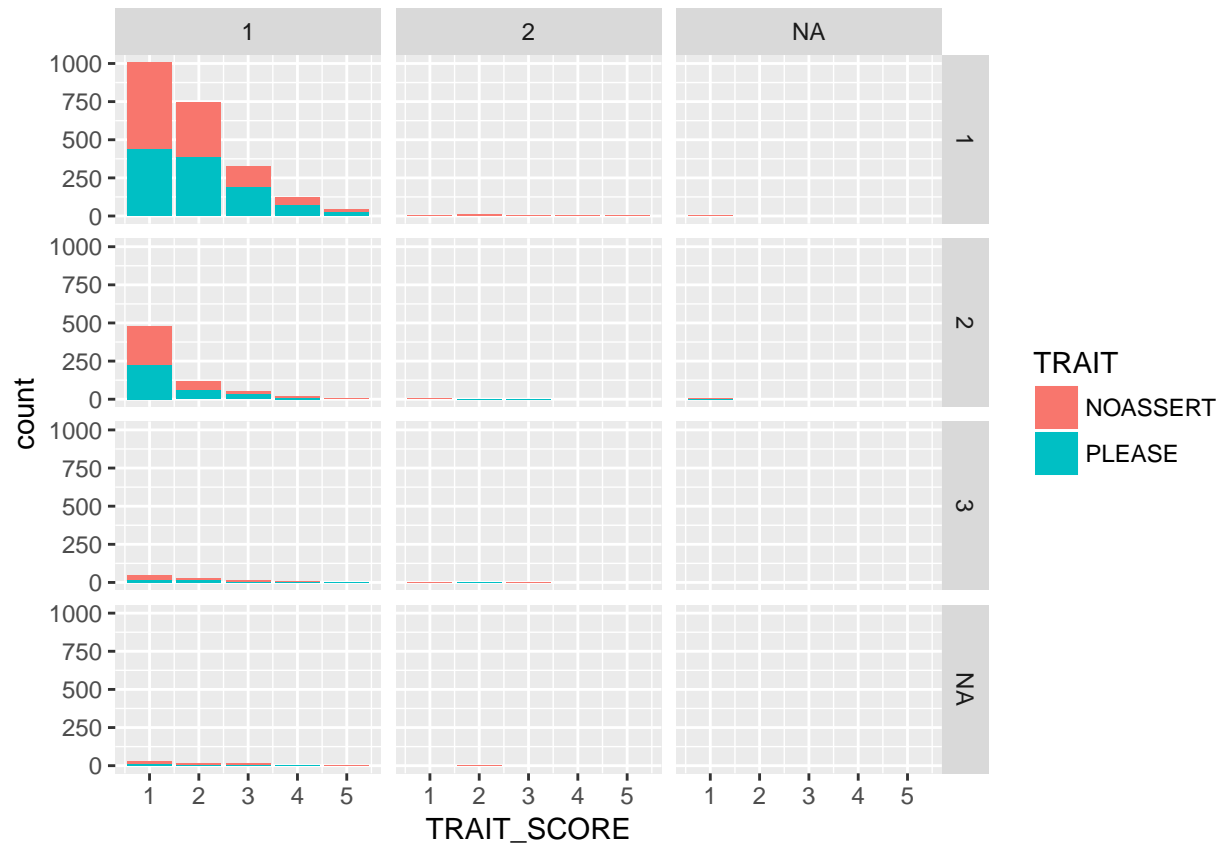


```
#plot TRAIT_SCORE facettted by RACE and PRESSSI
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge") +
  facet_grid(RACE~PRESSSI)
```

```
## Warning: Removed 20 rows containing non-finite values (stat_count).
```

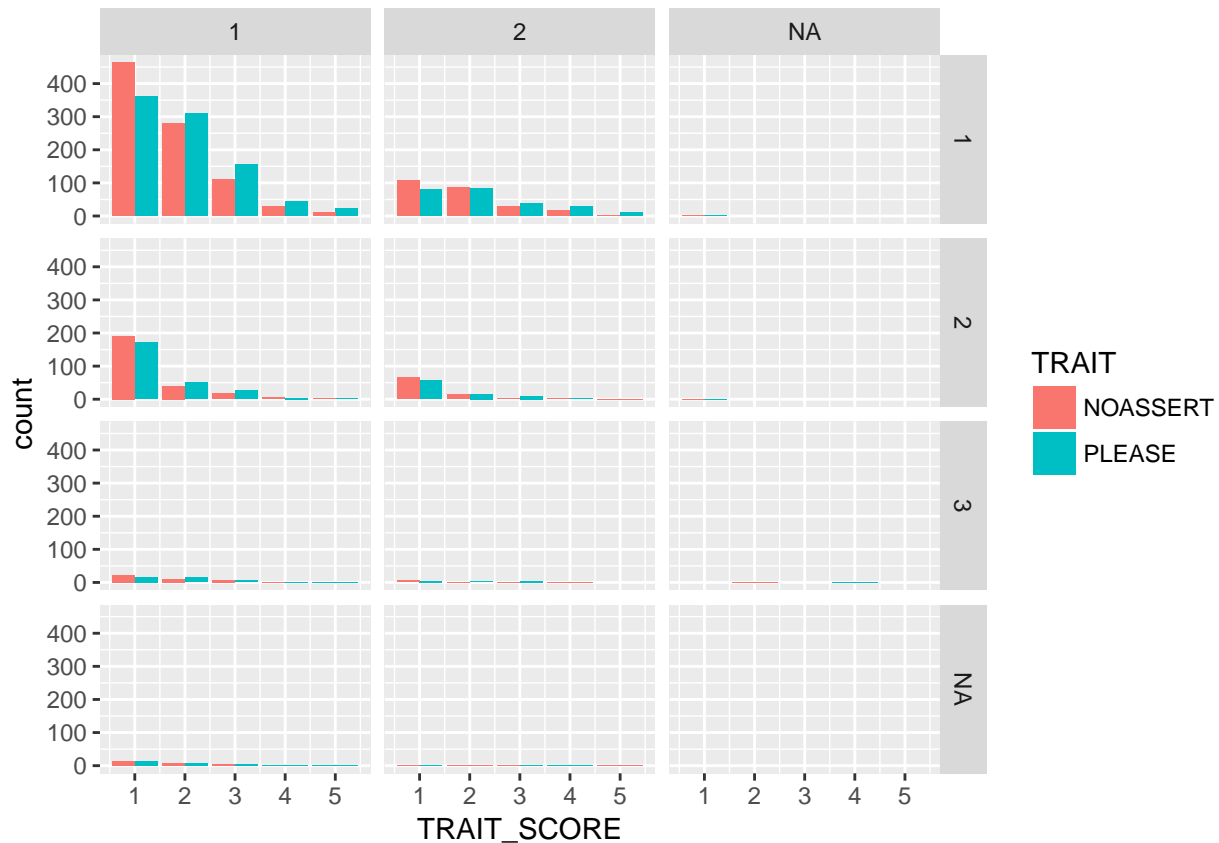


```
#change AUTHSI plot to be stacked
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "stack", na.rm = TRUE) +
  facet_grid(RACE~AUTHSI)
```



```
#change PRESSSI plot to be stacked
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge") +
  facet_grid(RACE~PRESSSI)
```

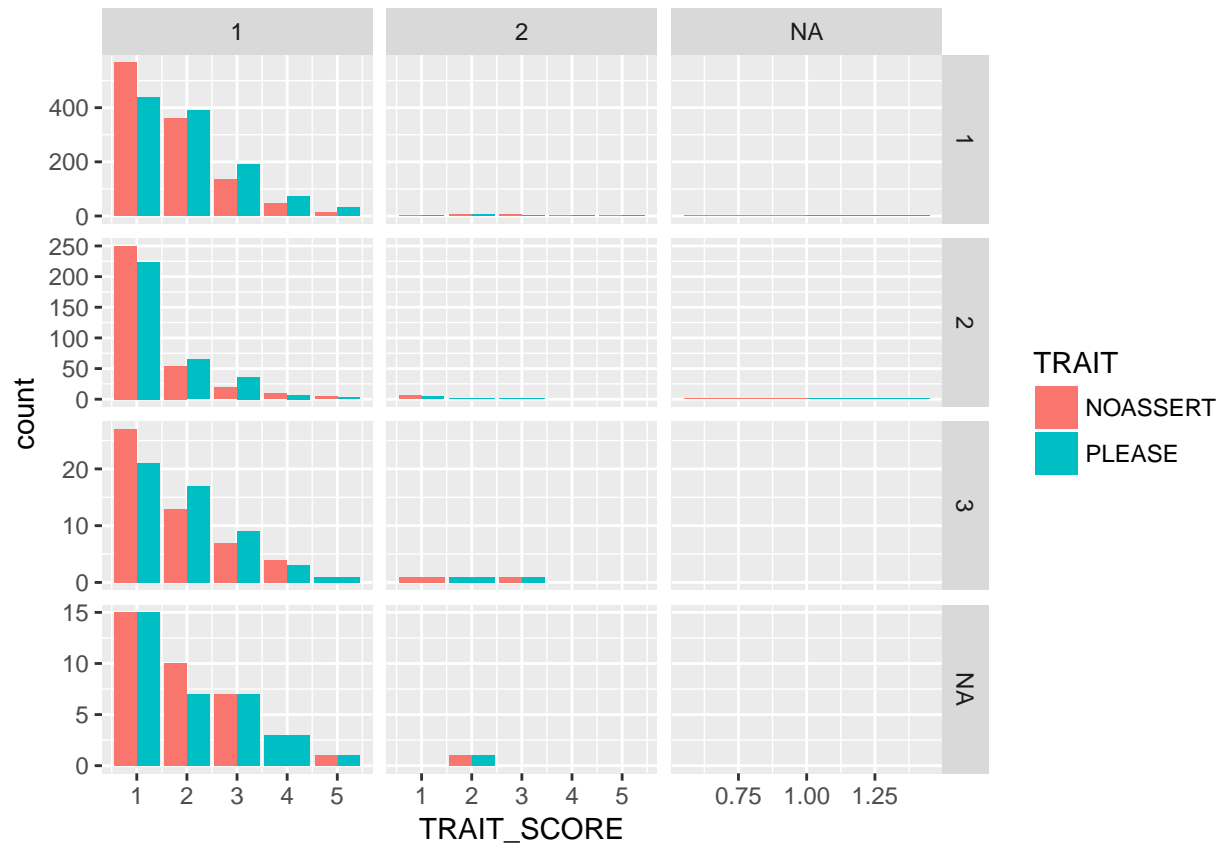
```
## Warning: Removed 20 rows containing non-finite values (stat_count).
```



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 there any benefits/drawbacks with either or both of these approaches for the visualizations you've created? What are they?

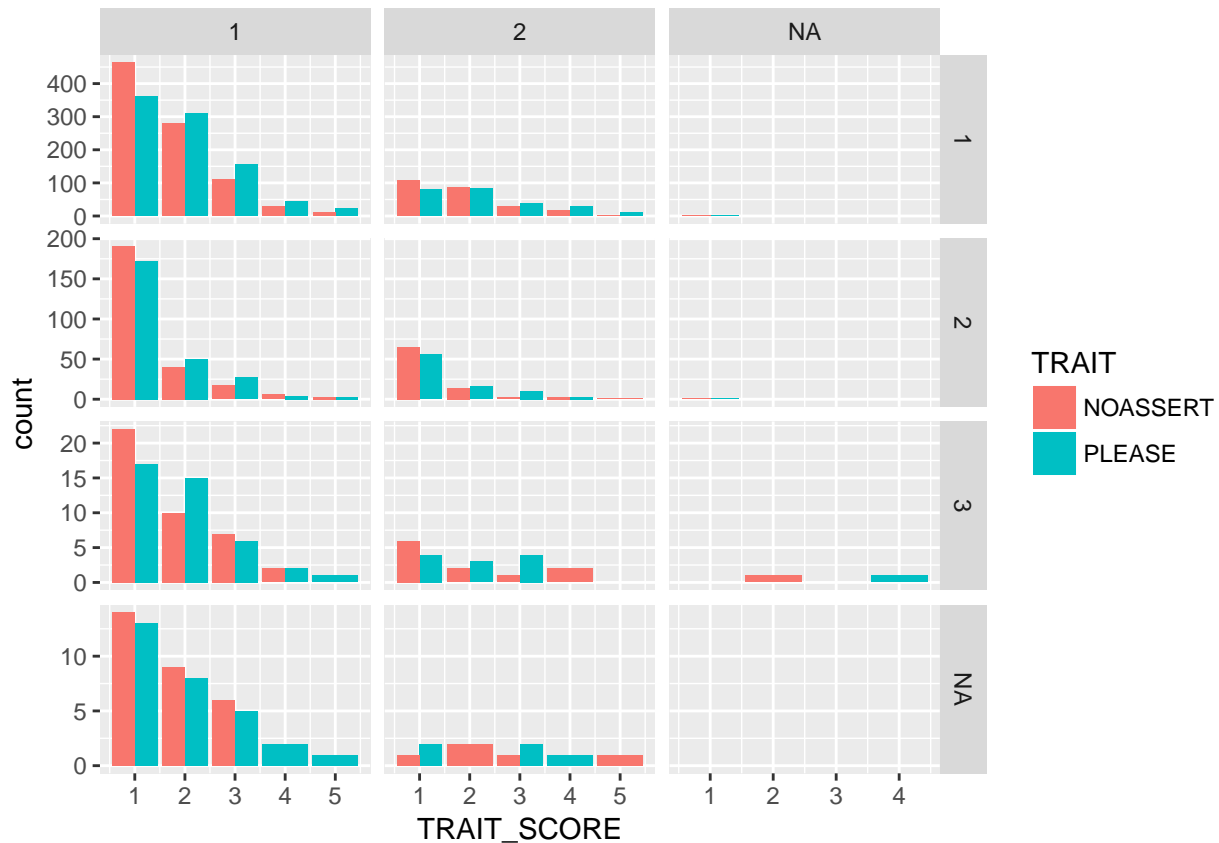
I tried scaling differently depending on the facet by using `scales = "free"`. The benefits of this is that it makes comparing between the horizontal facets easier, however, if you are not paying close attention it could be misleading when comparing vertically because the scales are so different along the y axes.

```
#change AUTHSI plot to scale differently based on the facet
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge", na.rm = TRUE) +
  facet_grid(RACE~AUTHSI, scales = "free")
```



```
#change PRESSSI plot to scale differently based on the facet
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge") +
  facet_grid(RACE~PRESSSI, scales = "free")
```

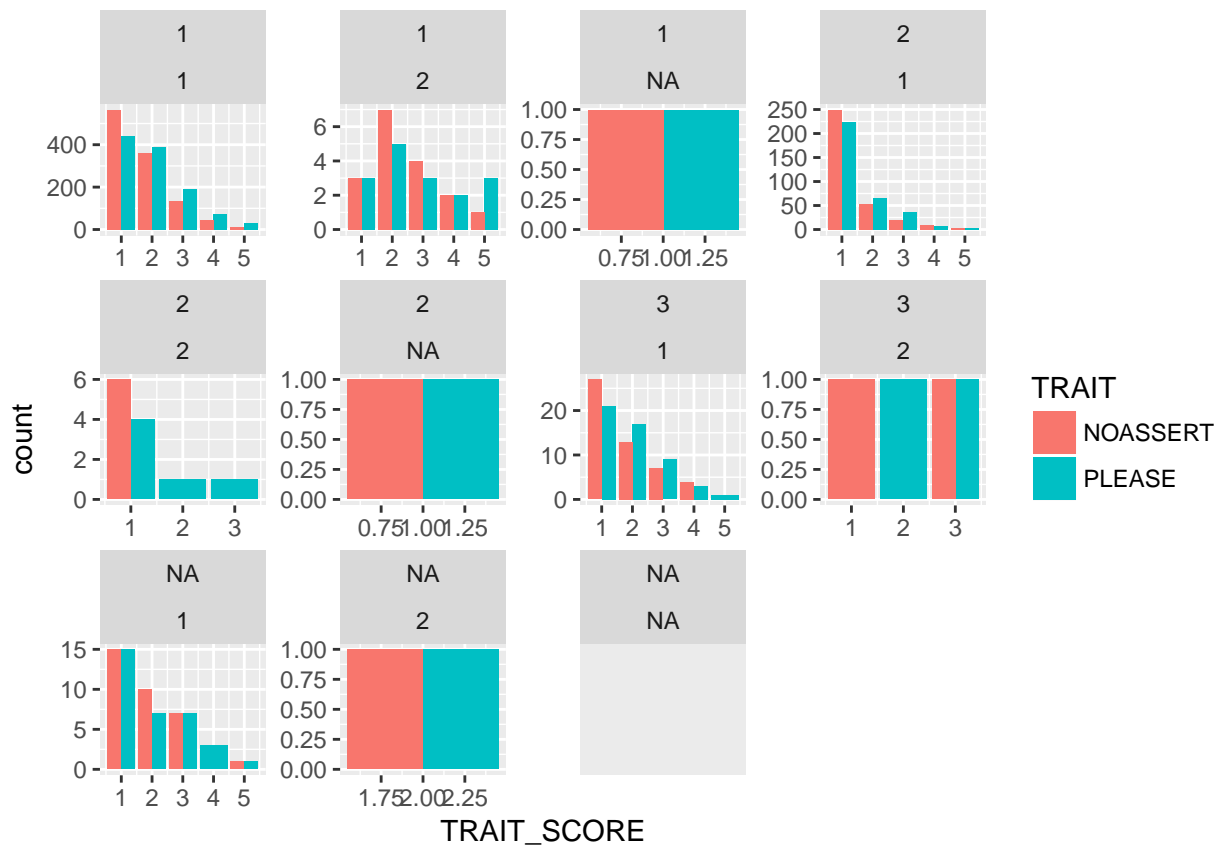
```
## Warning: Removed 20 rows containing non-finite values (stat_count).
```



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?

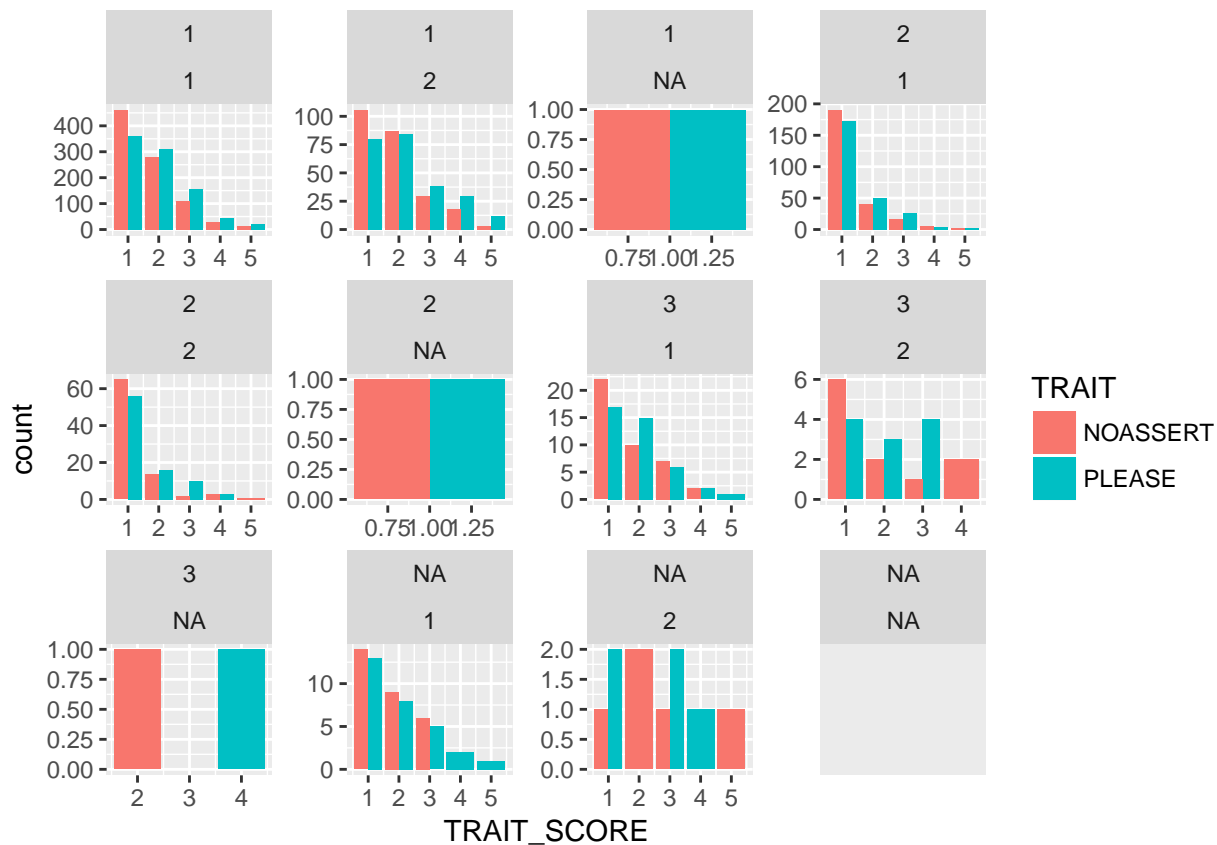
I have already faceted by RACE, AUTHSI, AND PRESSSI using `face_grid()`. For this question I tried using `facet_wrap()` to visualize the difference. Right now `facet_wrap` seems more confusing because I haven't been able to label the different facet options, but I think that once they are labelled this could be a potentially useful way to visualize the differences within each facet without being confused by the different scales on one axis.

```
#change AUTHSI plot using facet_wrap()
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge", na.rm = TRUE) +
  facet_wrap(RACE~AUTHSI, scales = "free")
```

```
#change PRESSSI plot using facet_wrap()
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge") +
  facet_wrap(RACE~PRESSSI, scales = "free")
```

```
## Warning: Removed 20 rows containing non-finite values (stat_count).
```

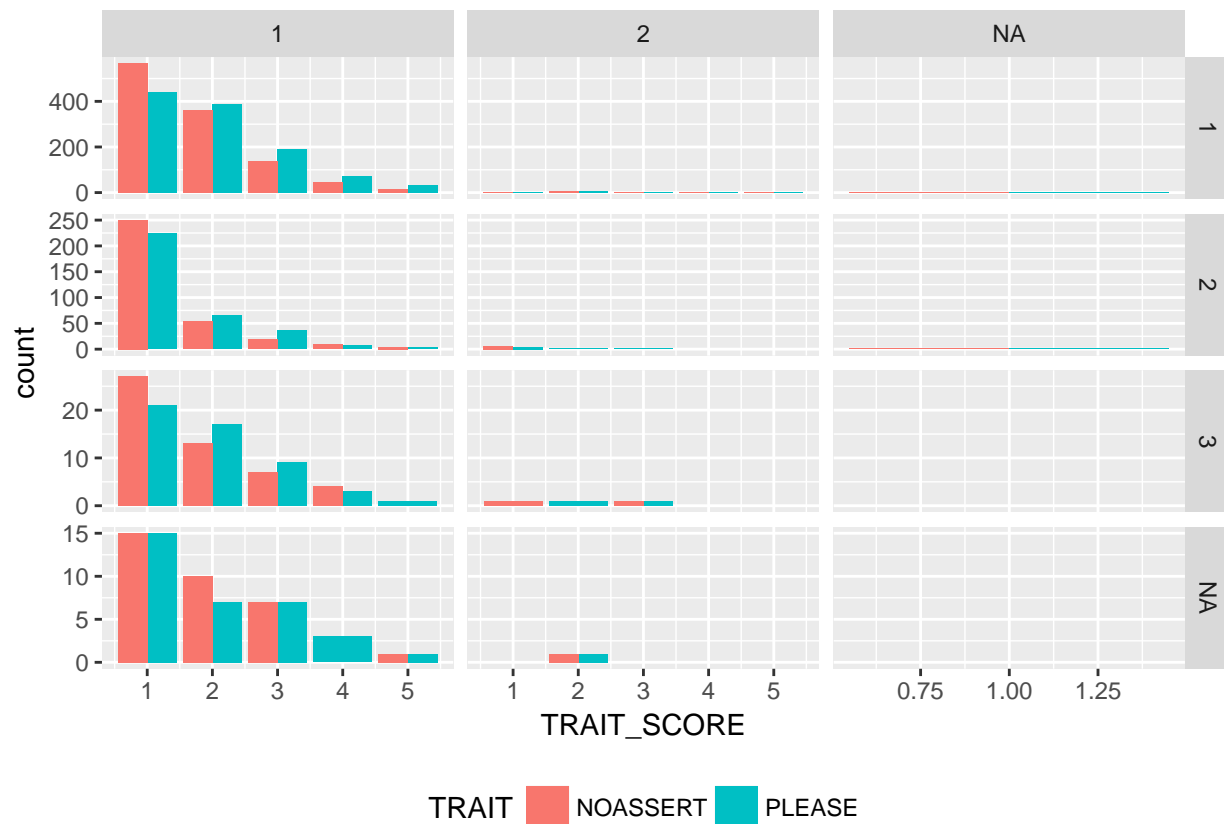


4. Use the `theme()` layer to change the appearance of a plot of your choice including the

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

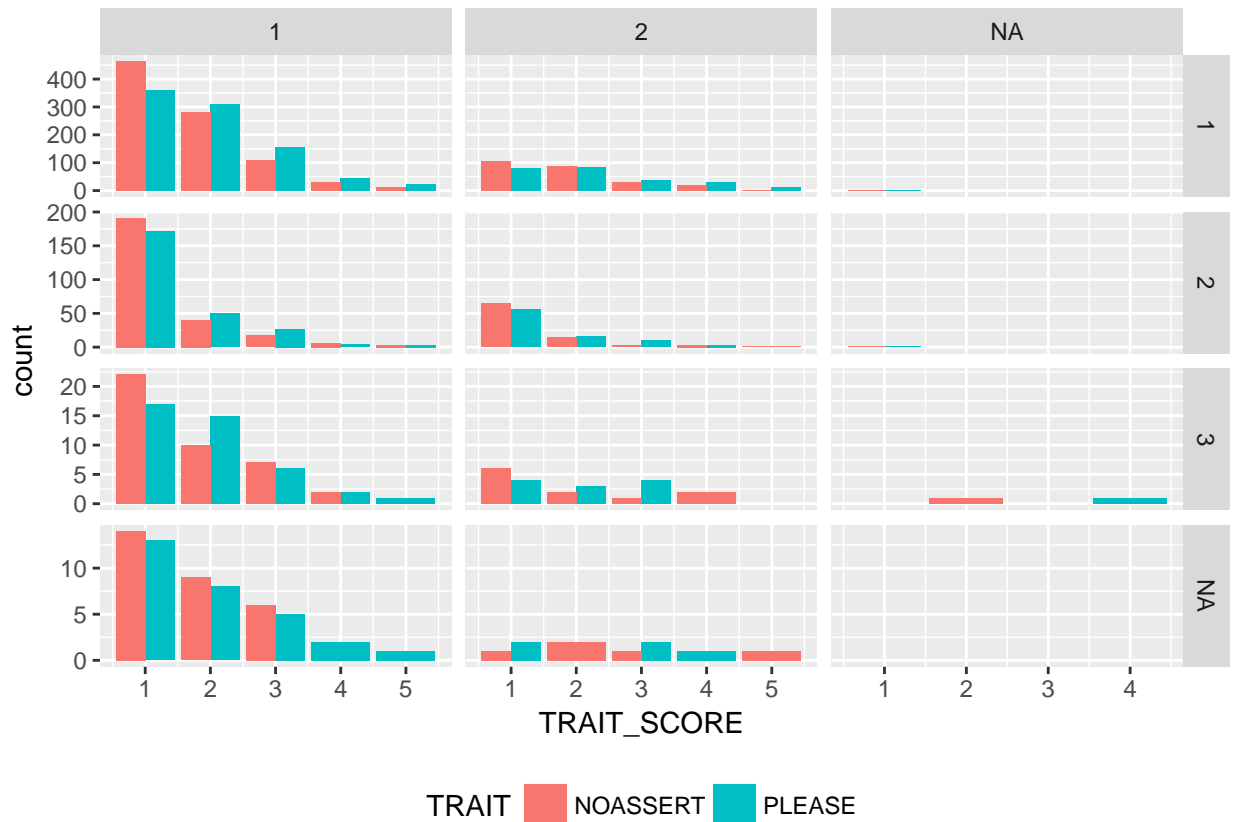
```
library("ggplot2")

#change AUTHSI plot through theme()
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge", na.rm = TRUE) +
  facet_grid(RACE~AUTHSI, scales = "free") +
  theme(plot.title = element_text("Traits 'NOASSERT' and 'PLEASE' when accounting for RACE and AUTHSI"))
```



```
#change PRESSSI plot through theme()
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge") +
  facet_grid(RACE~PRESSSI, scales = "free") +
  theme(plot.title = element_text("Traits 'NOASSERT' and 'PLEASE' when accounting for RACE and PRESSSI"))
```

```
## Warning: Removed 20 rows containing non-finite values (stat_count).
```



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 choose that you think make them more appropriate for presentation?

I like the background of the first one but the layout of the second one better. The layouts of the first and third themes are a little confusing for a presentation in my opinion. I want a theme that is visually pleasing in terms of colors but also easy to understand at first glance for a poster session and the second solarized palette would be the closest to fitting that description. The Economist theme and the Few themes might be better for an academic journal.

```
install.packages("ggthemes")
```

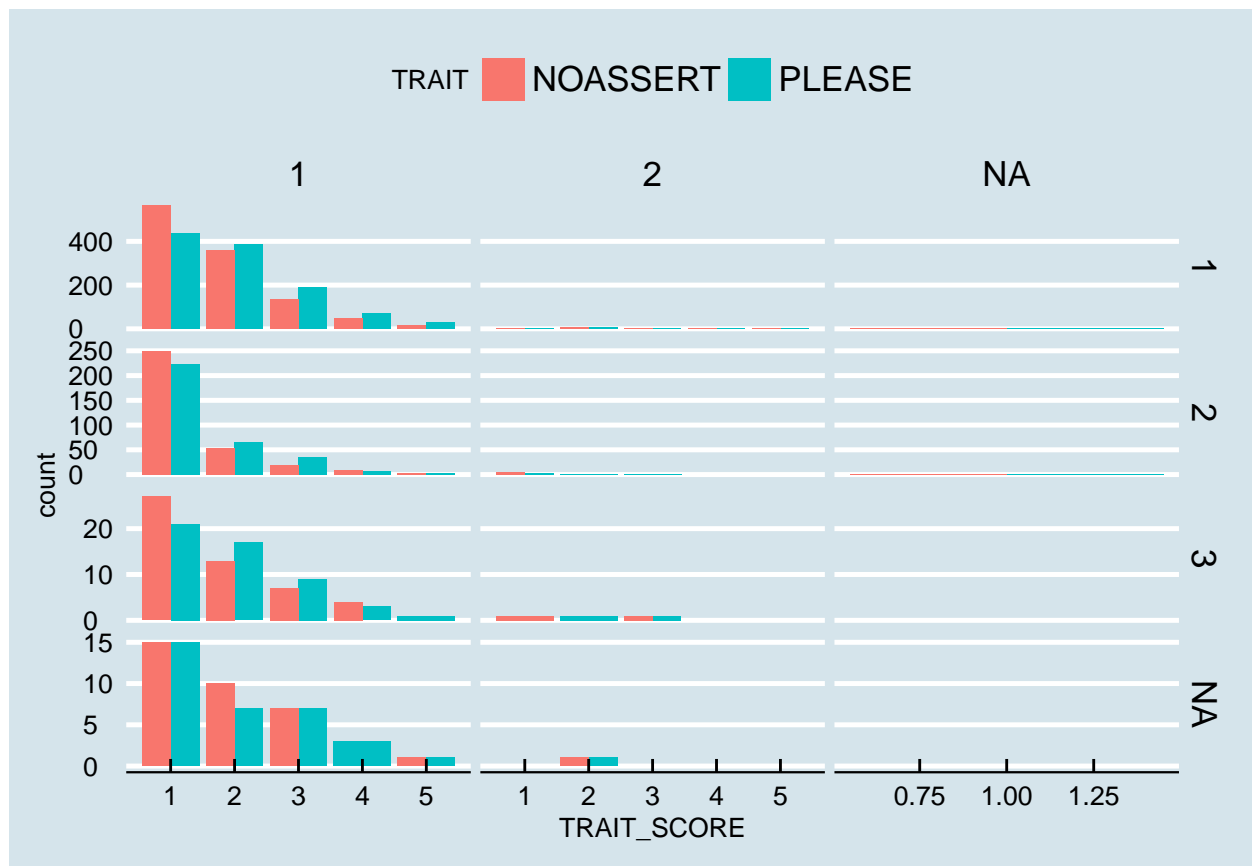
```
## Installing package into 'C:/Users/snowbc/Documents/Projects/sexual_violence_college/packrat/lib/x86_64-w64-mingw32'
## (as 'lib' is unspecified)
```

```
## package 'ggthemes' successfully unpacked and MD5 sums checked
##
```

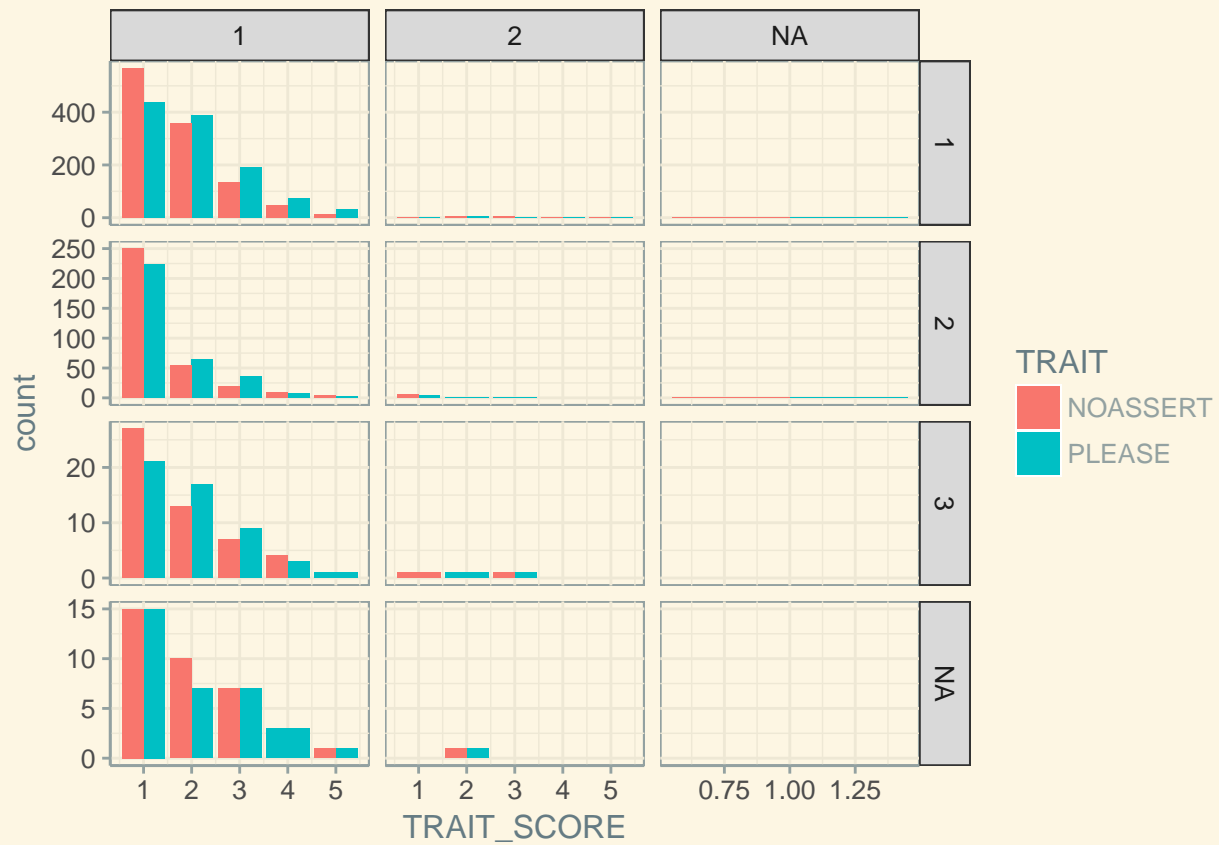
```
## The downloaded binary packages are in
## C:\Temp\31\RtmpK087vT\downloaded_packages
```

```
library("ggthemes")
```

```
#change AUTHSI plot to theme based on The Economist
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge", na.rm = TRUE) +
  facet_grid(RACE~AUTHSI, scales = "free") +
  theme_economist()
```



```
#change AUTHSI plot to theme using the solarized color palette
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge", na.rm = TRUE) +
  facet_grid(RACE~AUTHSI, scales = "free") +
  theme_solarized()
```



```
#change AUTHSI plot to theme based on Stephen Few's rules
ggplot(data.gather, aes(x = TRAIT_SCORE, fill = TRAIT)) +
  geom_bar(position = "dodge", na.rm = TRUE) +
  facet_grid(RACE~AUTHSI, scales = "free") +
  theme_few()
```

