

Data visualization in R

Laura Lambert

James Madison University

Introductions

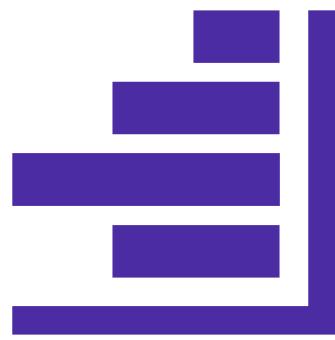
- Doctoral candidate in James Madison University's Assessment & Measurement program
- Full-time A&P faculty in Integrated Science and Technology (also at JMU)
- Data visualization enthusiast
- Believer in access and inclusion

Outline



Contextualize this in the Assessment Cycle

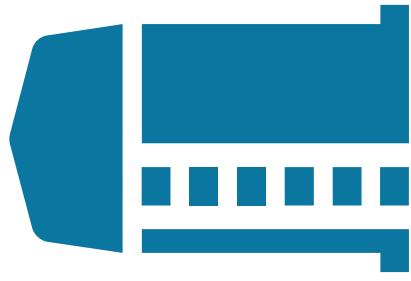
Outline



Data visualization best practices



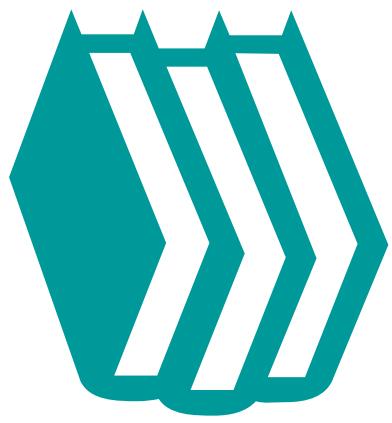
Outline



In-depth look at ggplot



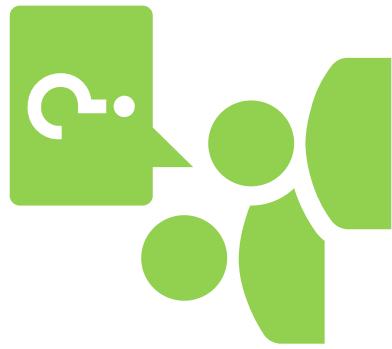
Outline



Show some open-source resources



Outline



Time for questions



Outline

Let's get started!



Context

Analyzing Data,
Reporting Results,
and
Maintaining
Information

Fidelity

Information

What is “ggplot”?

- Package in R for graphing data
- Utilizes the “grammar of graphics”
 - Breaks graphics down into components
- Builds charts through the stacking of layers
 - Not limited to pre-defined graphs
- Variables mapped to aesthetics
- Integrates with `tidyverse` environment

R Graph Gallery

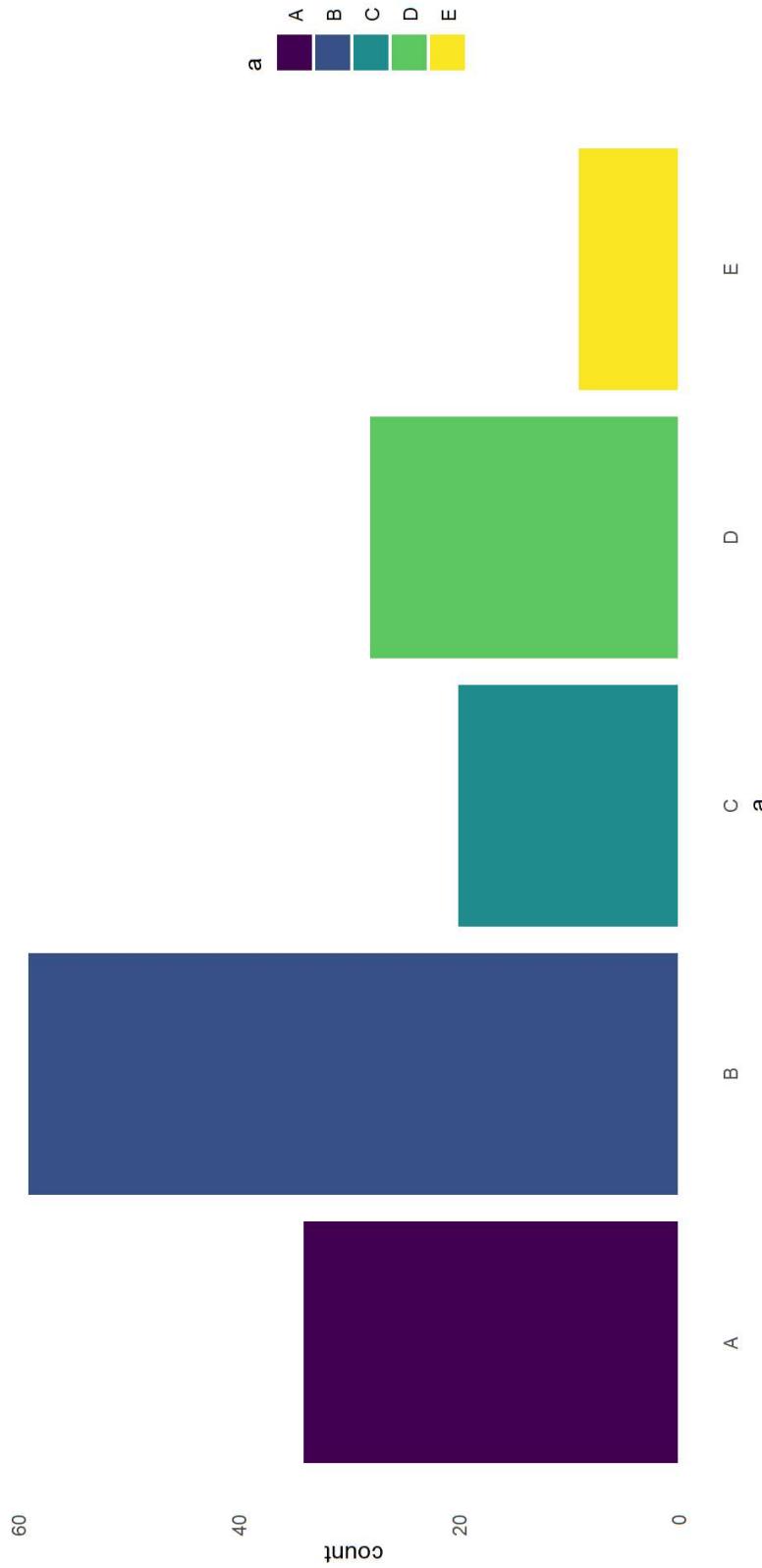
<https://r-graph-gallery.com/>

Data Visualization Best Practices

- Be clear
- Who's story are we telling?
- Avoid clutter
- Avoid misleading or confusing elements
- Color considerations (accessibility, consistency)
- NO PIE CHARTS!

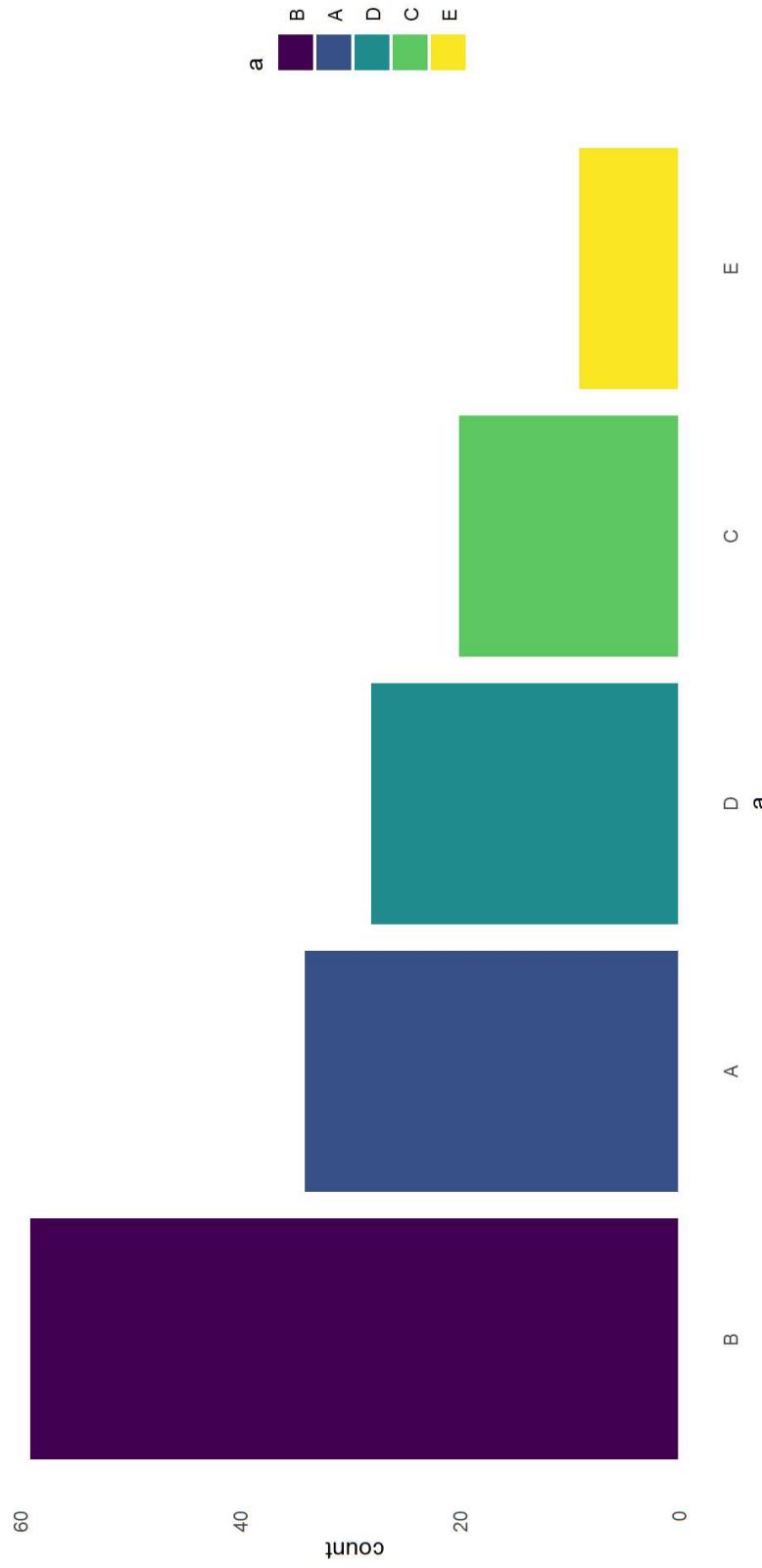
Ordered bars

Having bars in a bar plot be ascending or descending aids in interpretation and comparison

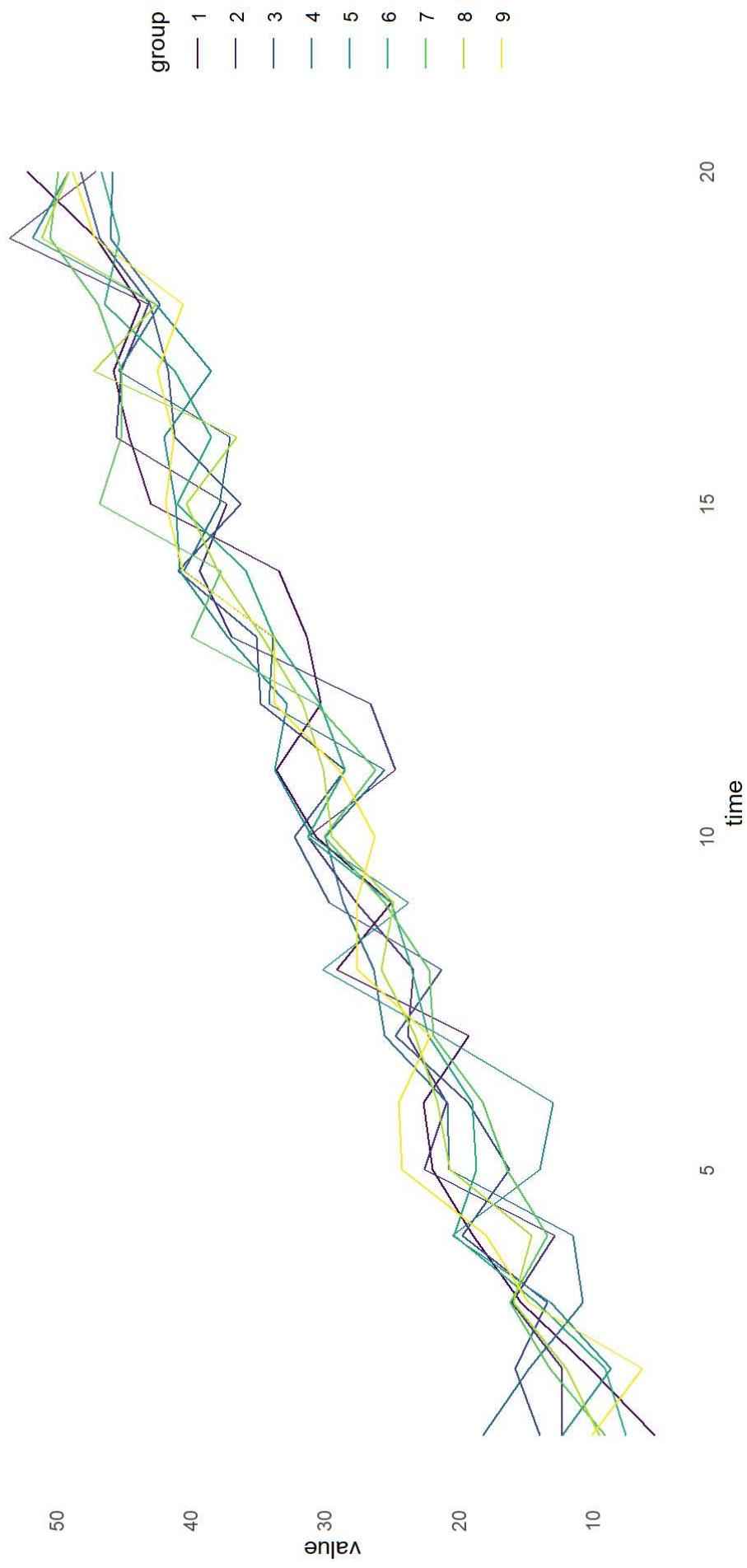


Ordered bars

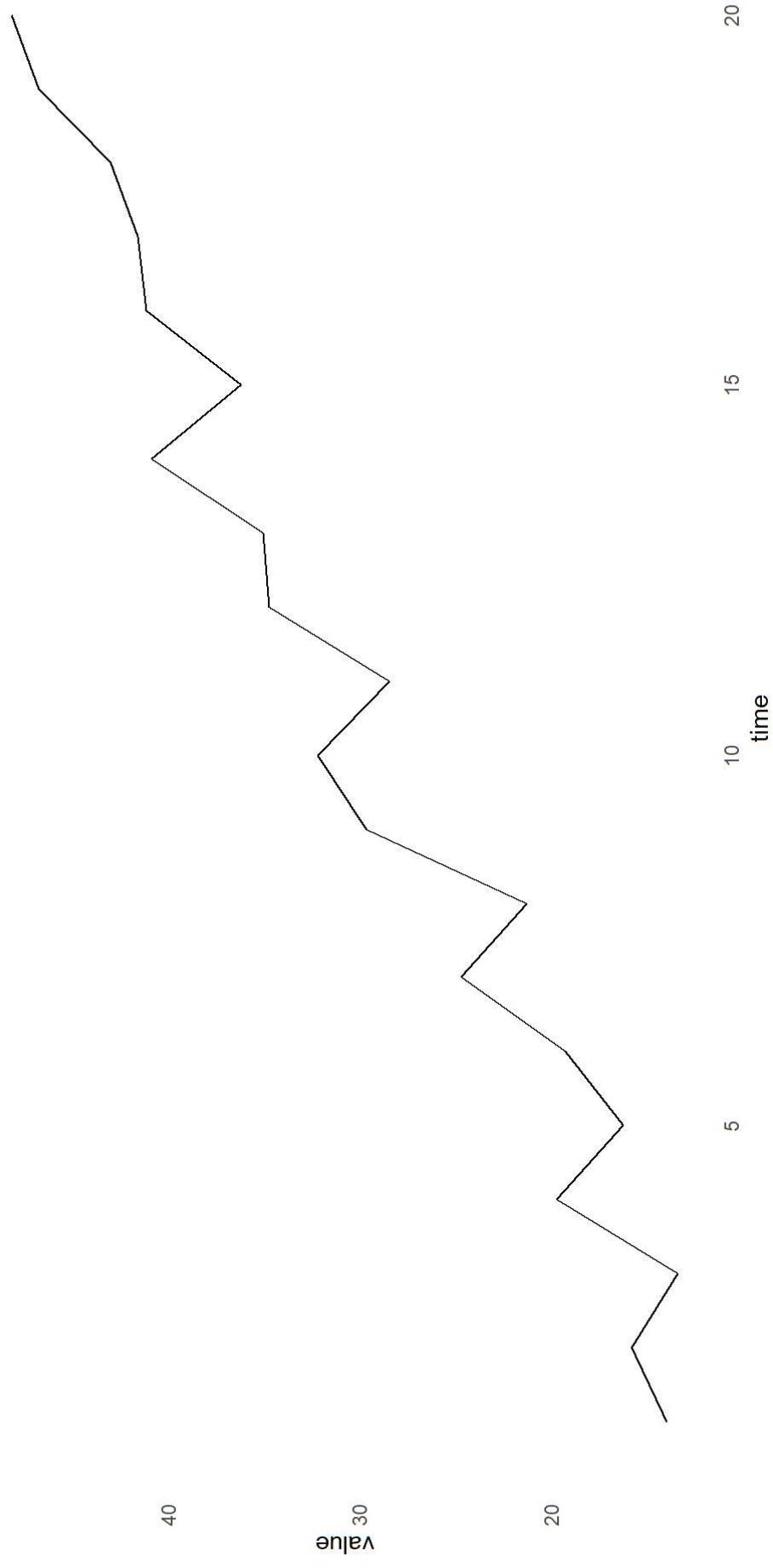
Having bars in a bar plot be ascending or descending aids in interpretation and comparison



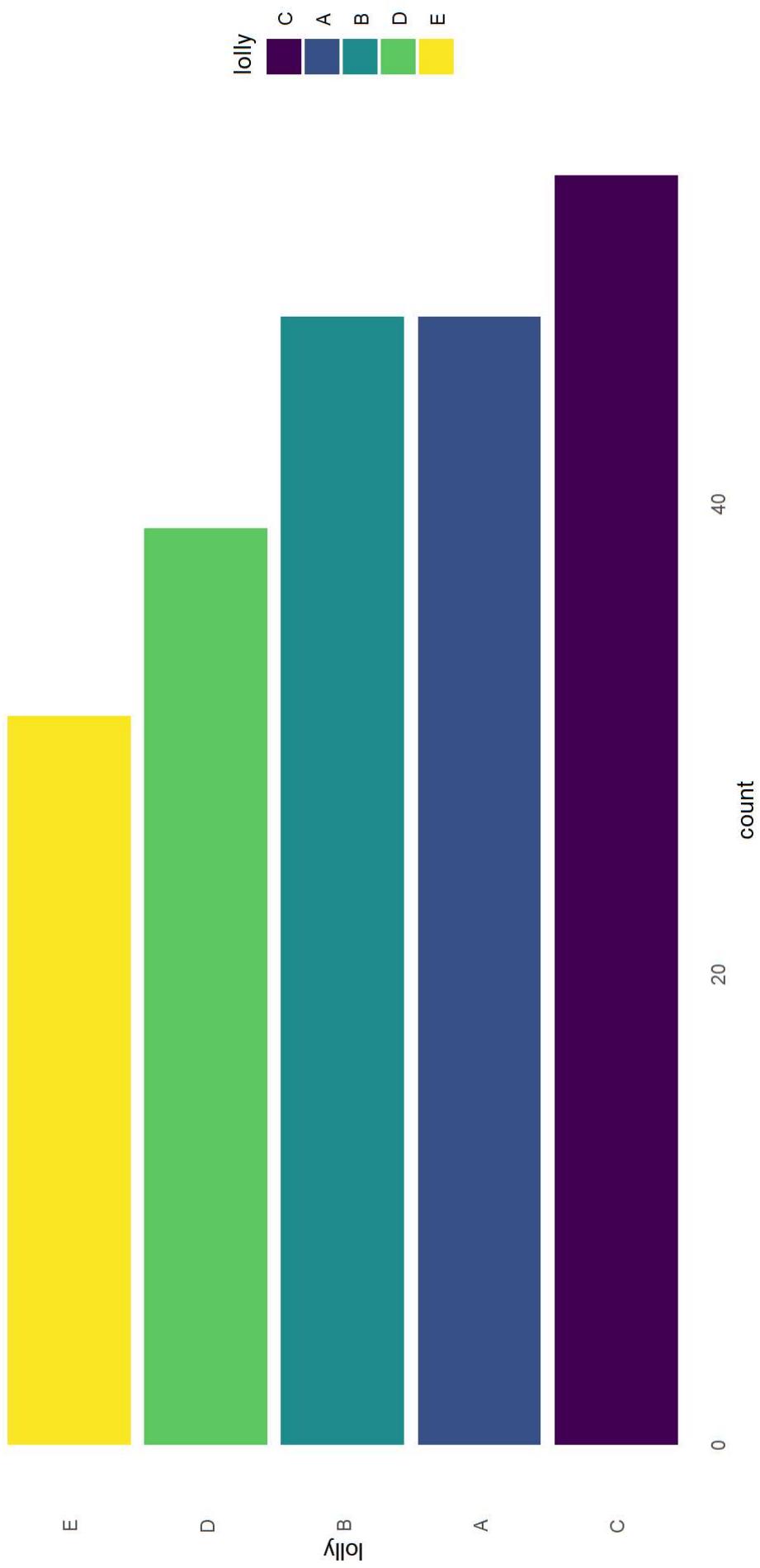
Avoid spaghetti charts



Avoid spaghetti charts



Lollipop charts



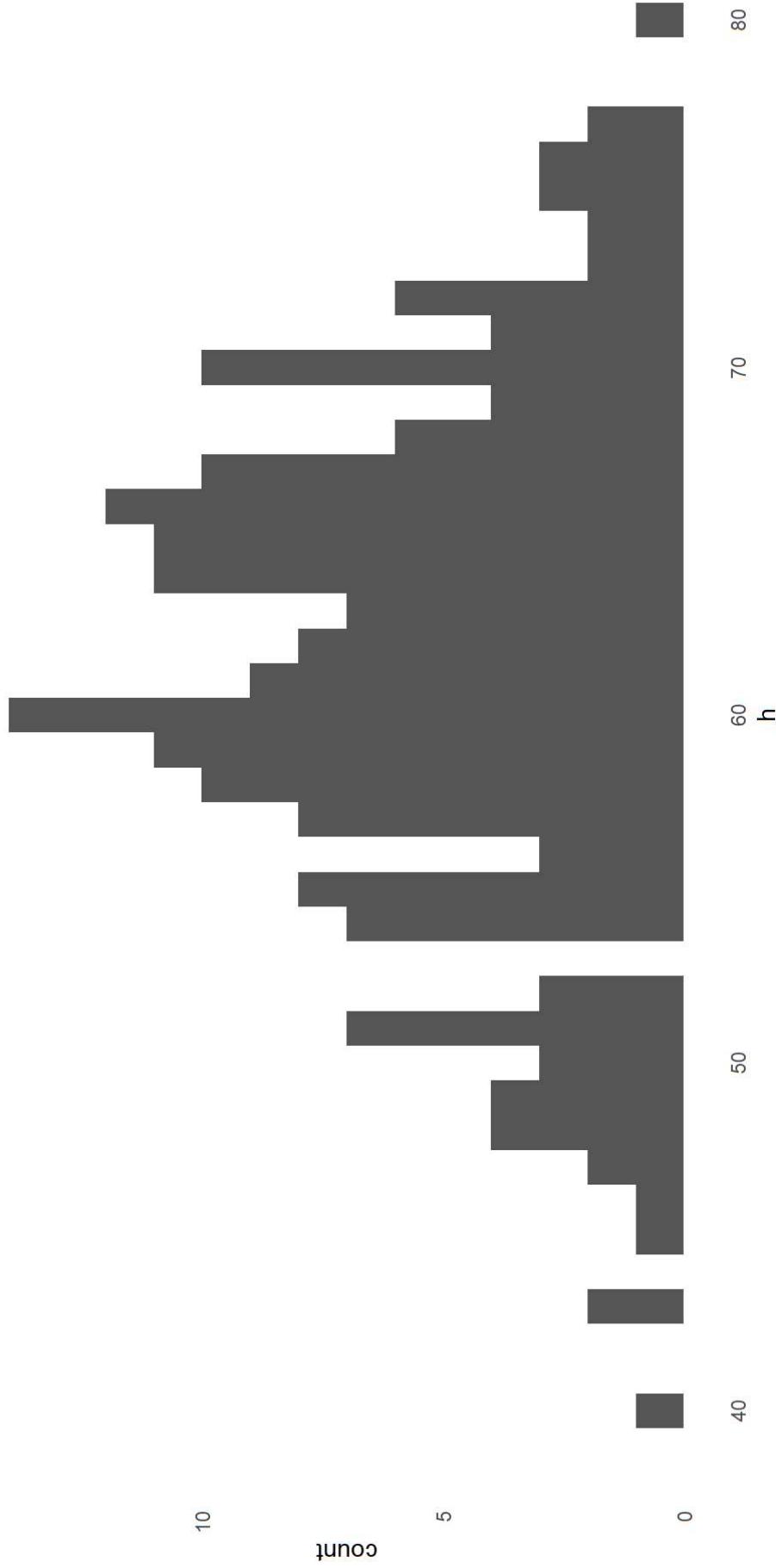
Lollipop charts



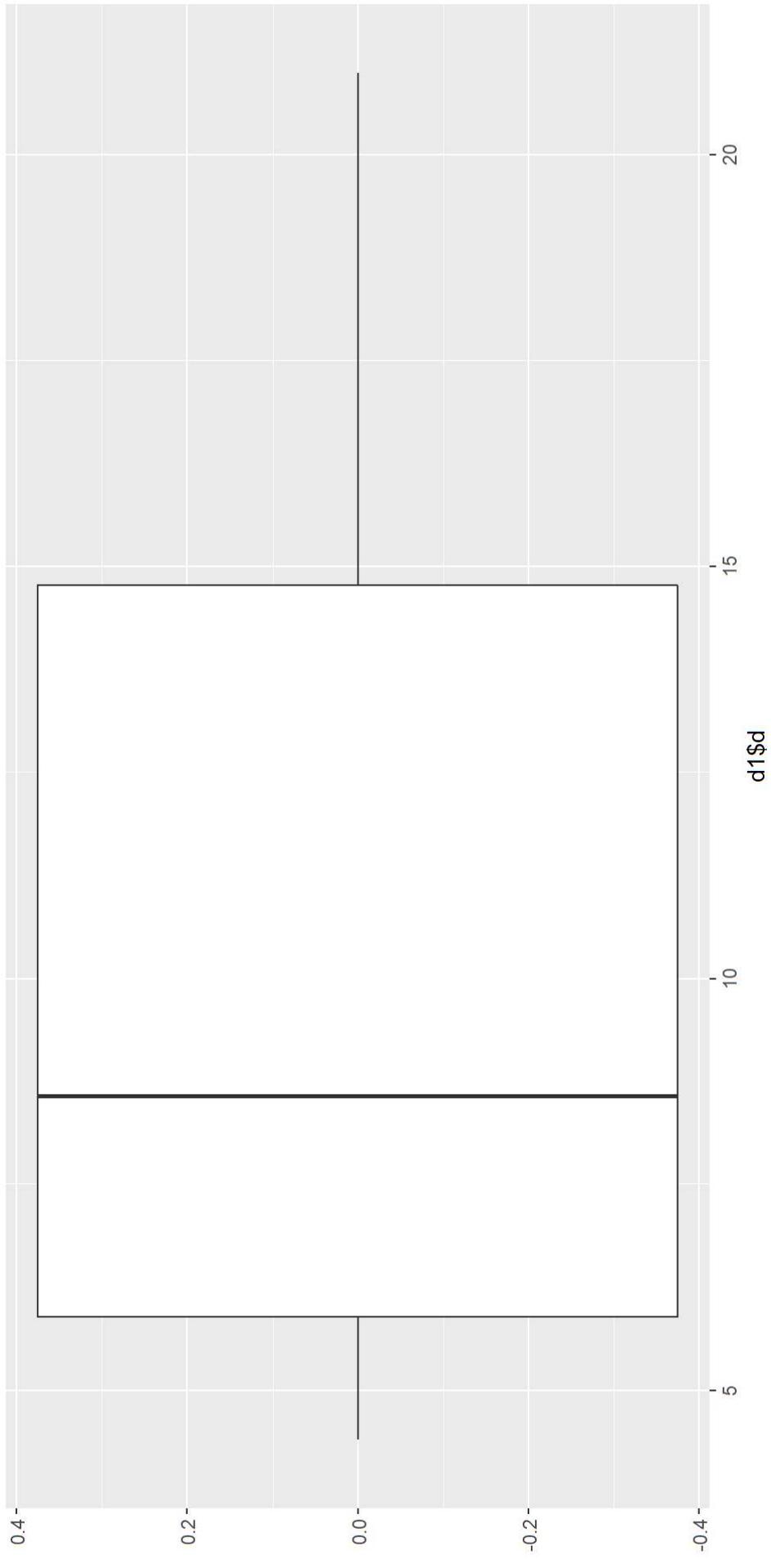
Bin Sizes



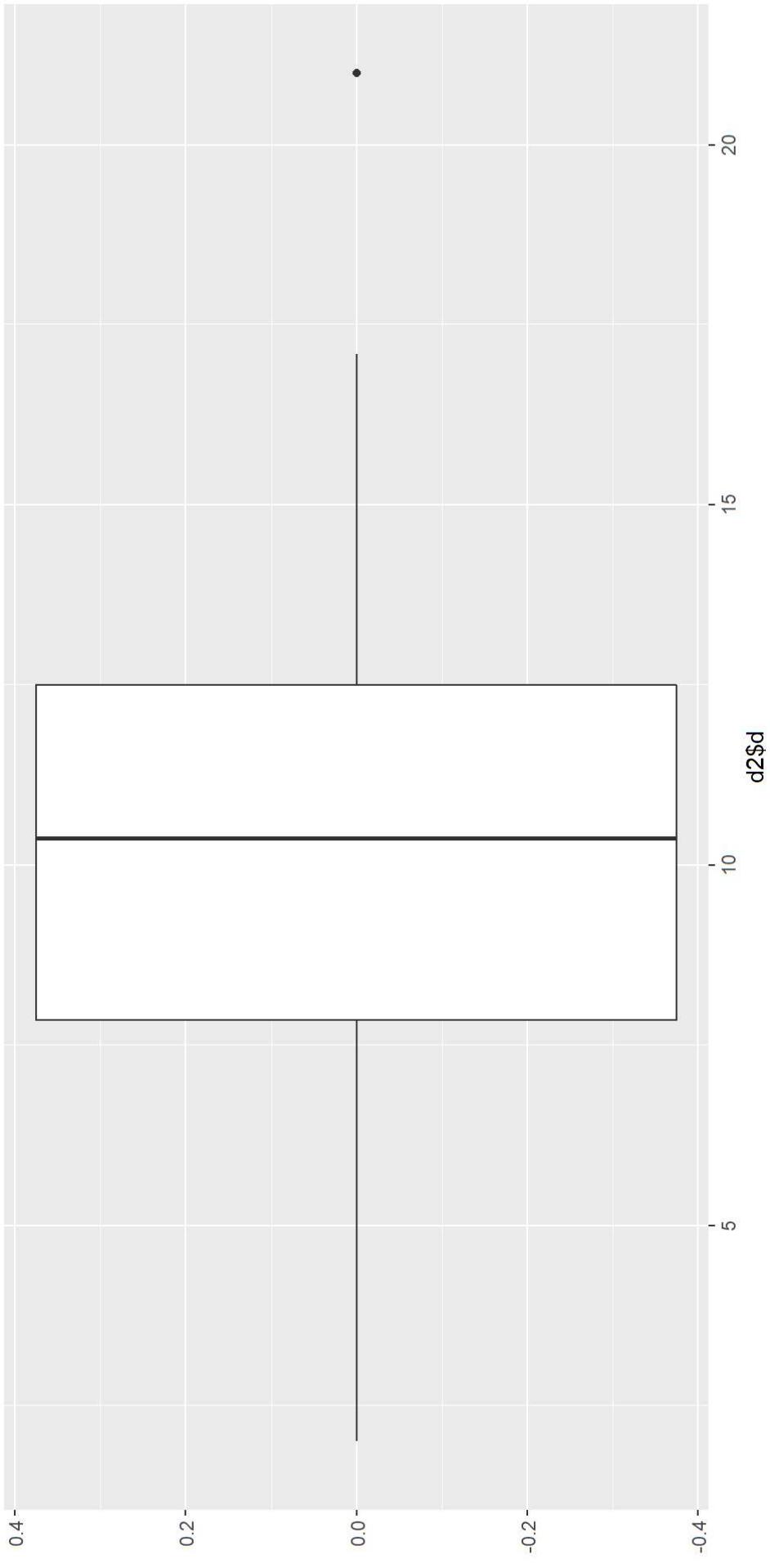
Bin Sizes



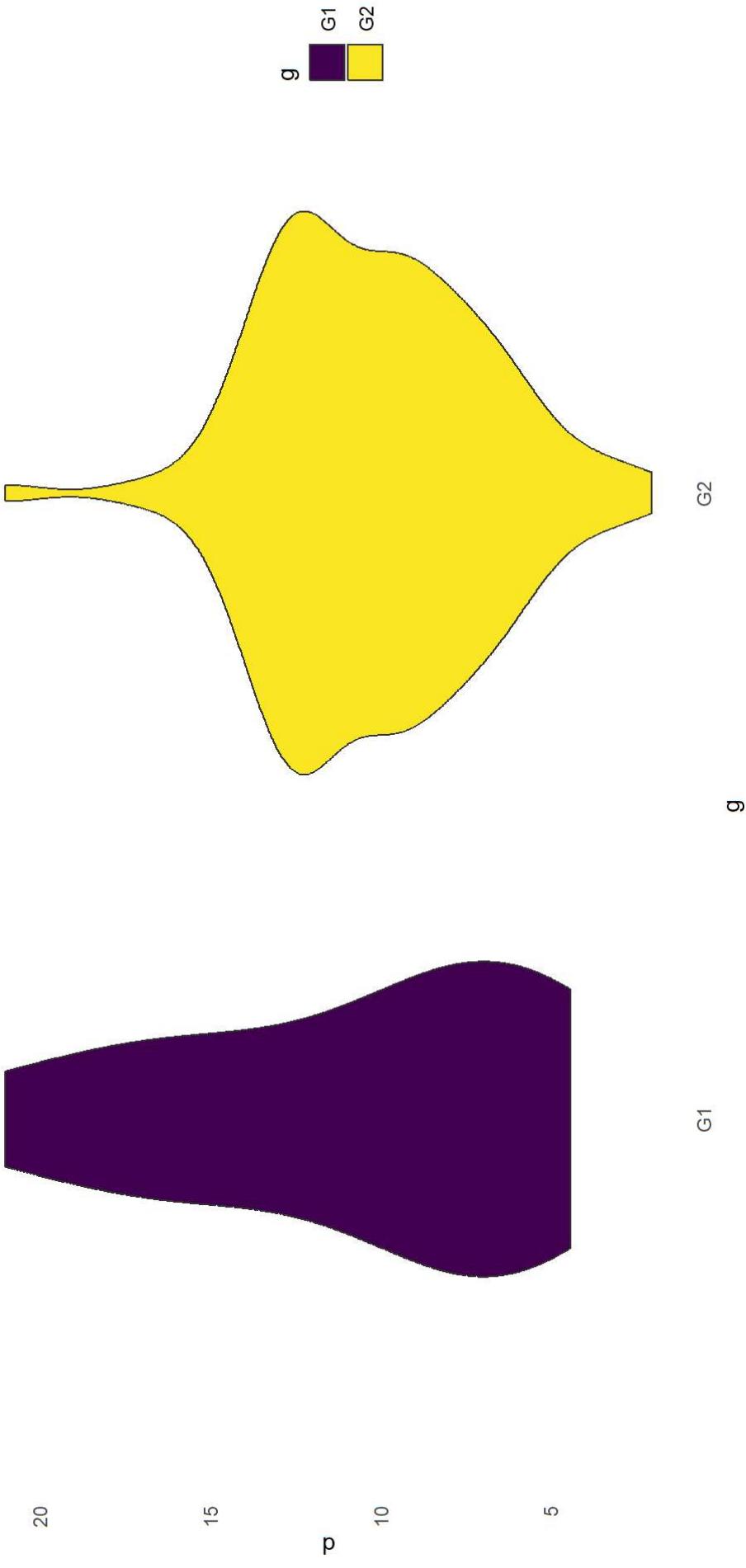
Boxplot considerations



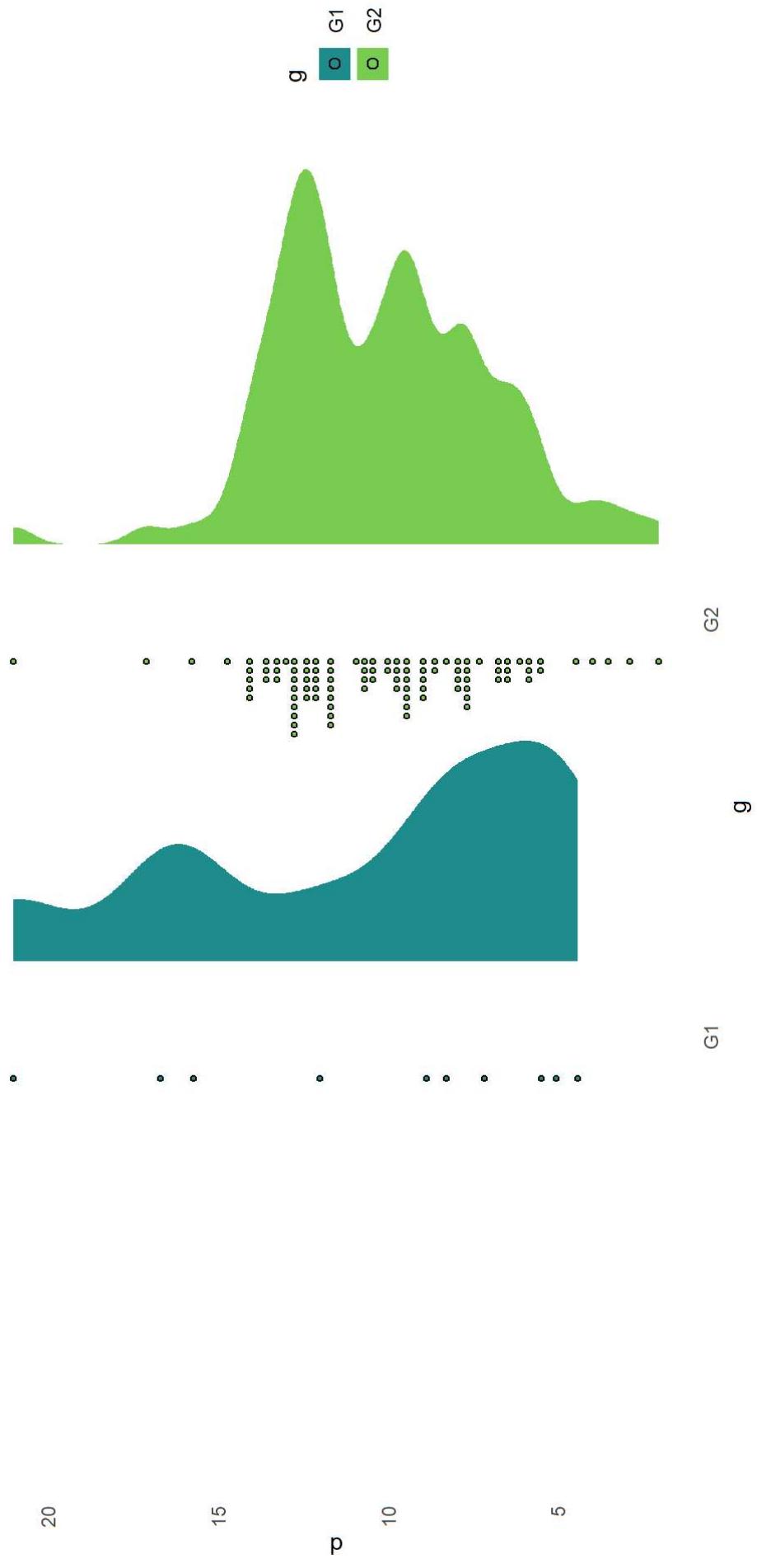
Boxplot considerations



Alternative

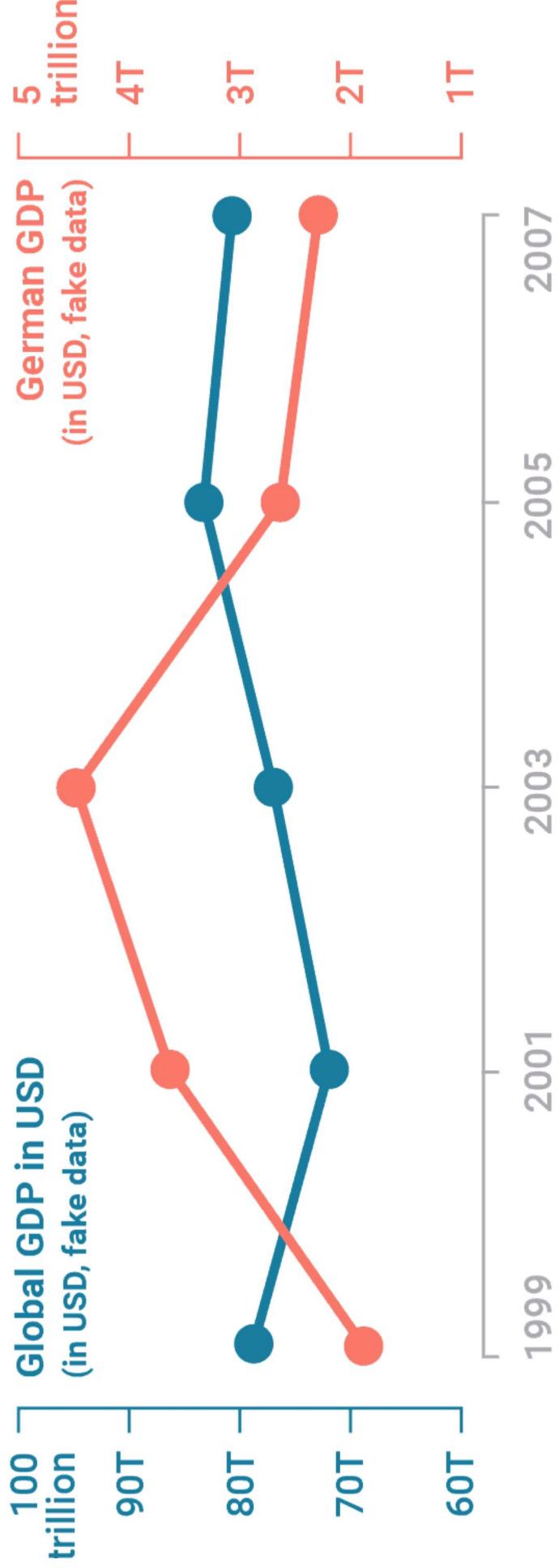


Alternative

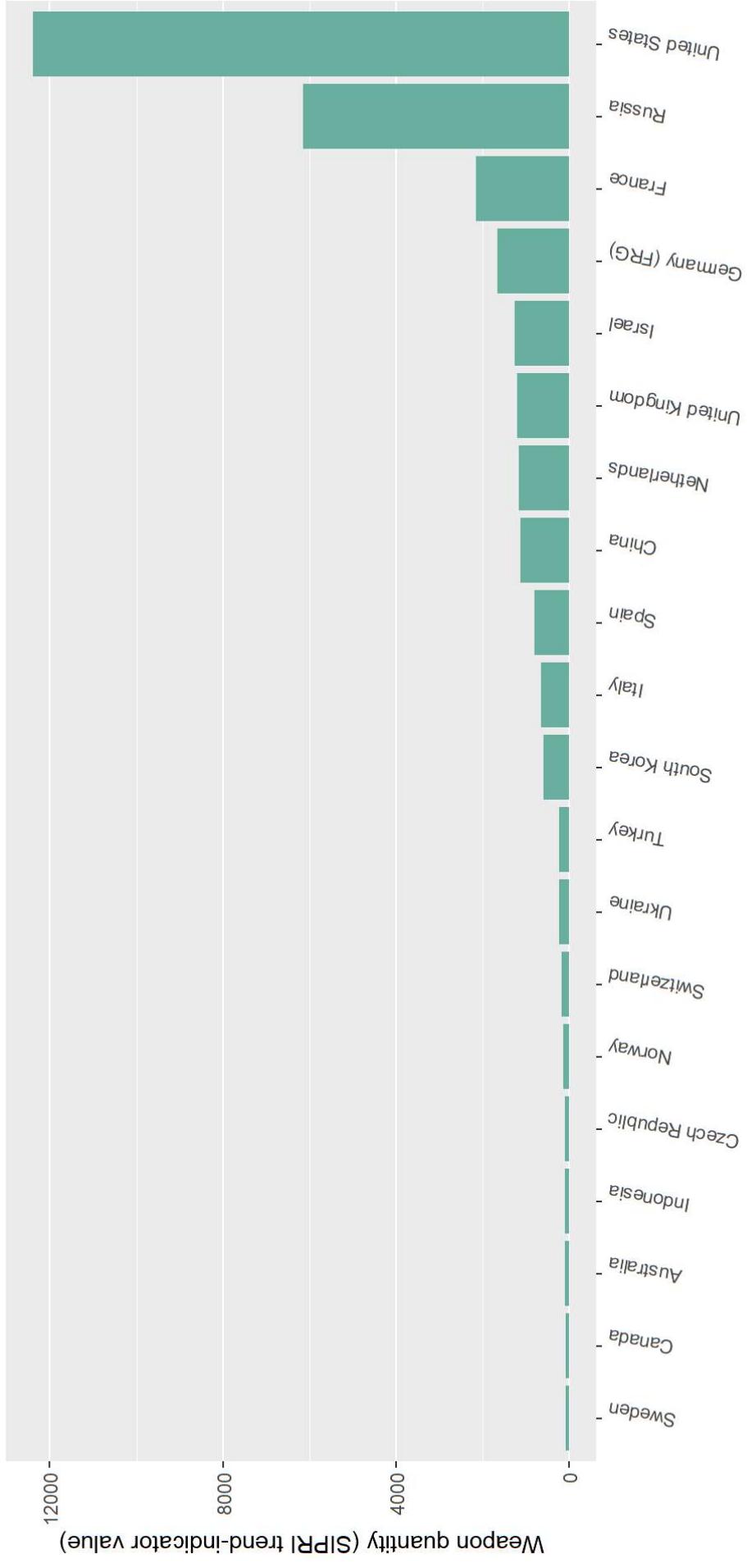


Dual Axis

Just....don't.

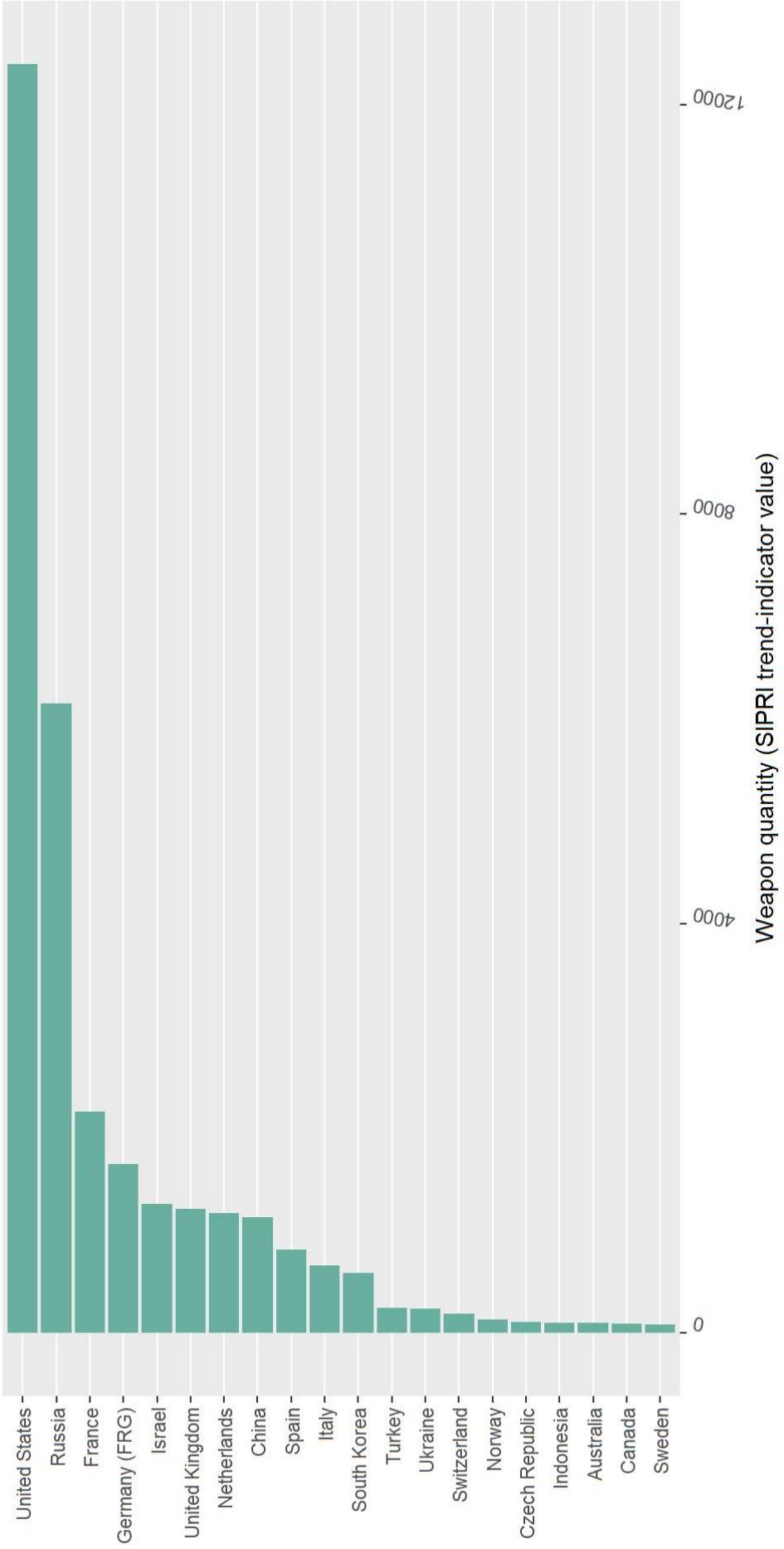


Axis Labels



(example shamelessly lifted from https://www.data-to-viz.com/caveat/hard_label.html)

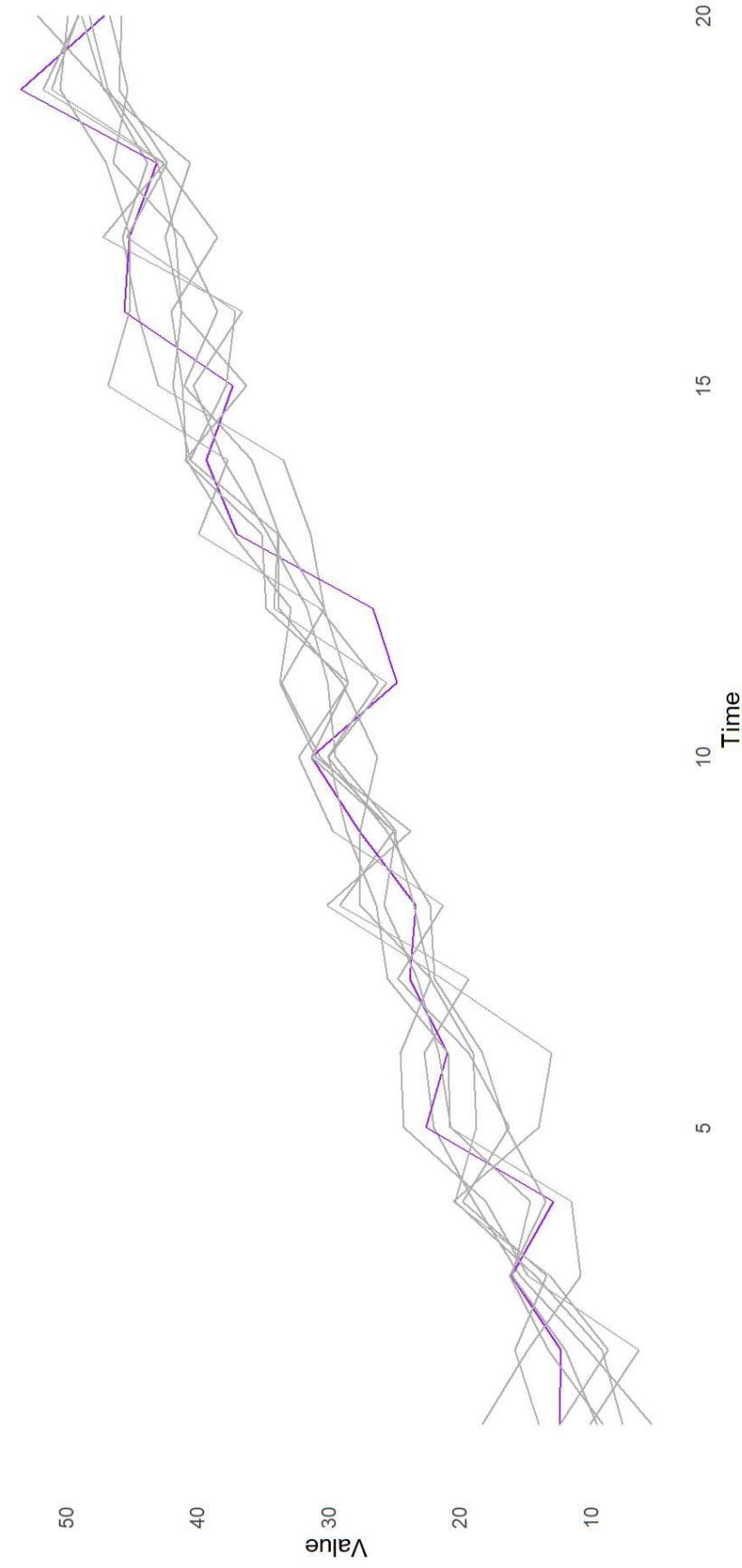
AxIs Labels



Color Considerations

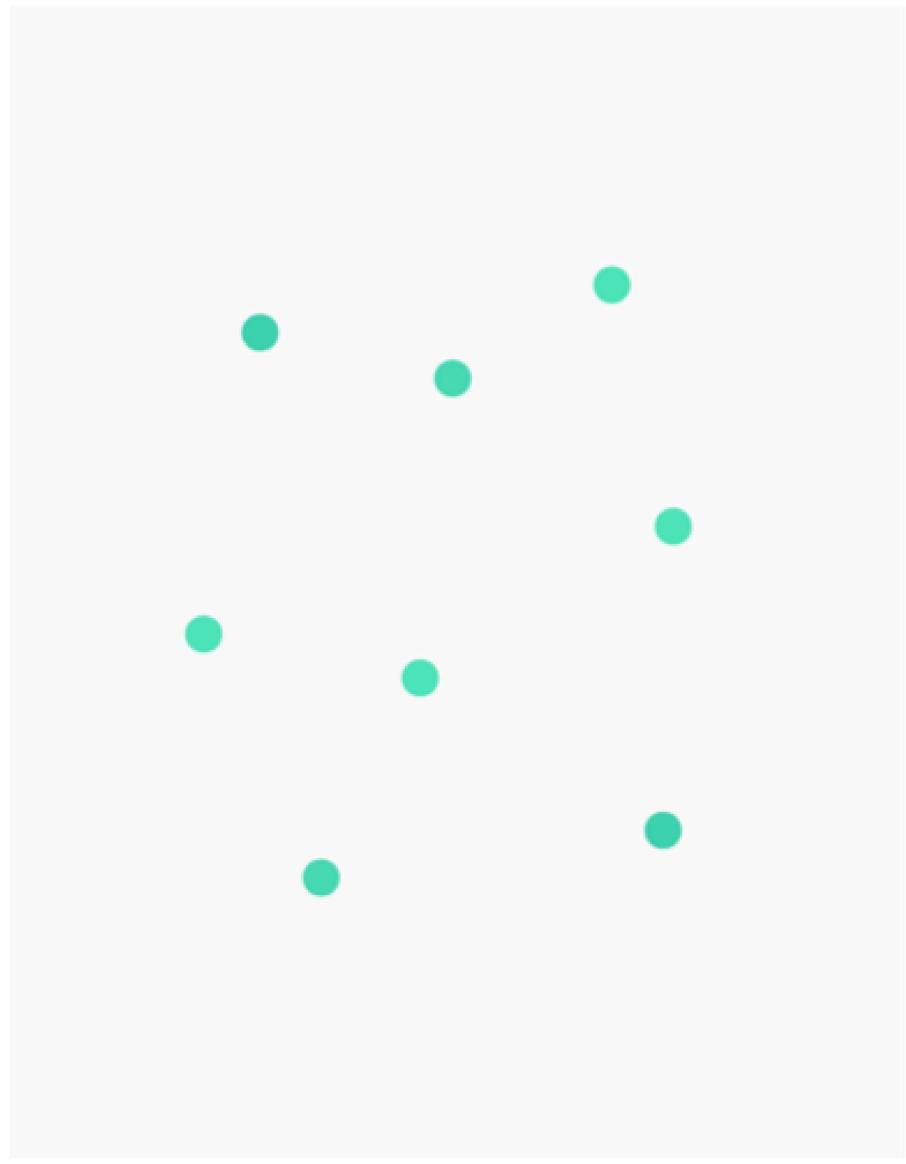
Color as a way to highlight particular data

Highlighted Line with Other Groups in Grey



Color Considerations

Contrast is important



Color Considerations

Contrast is important - so is proximity

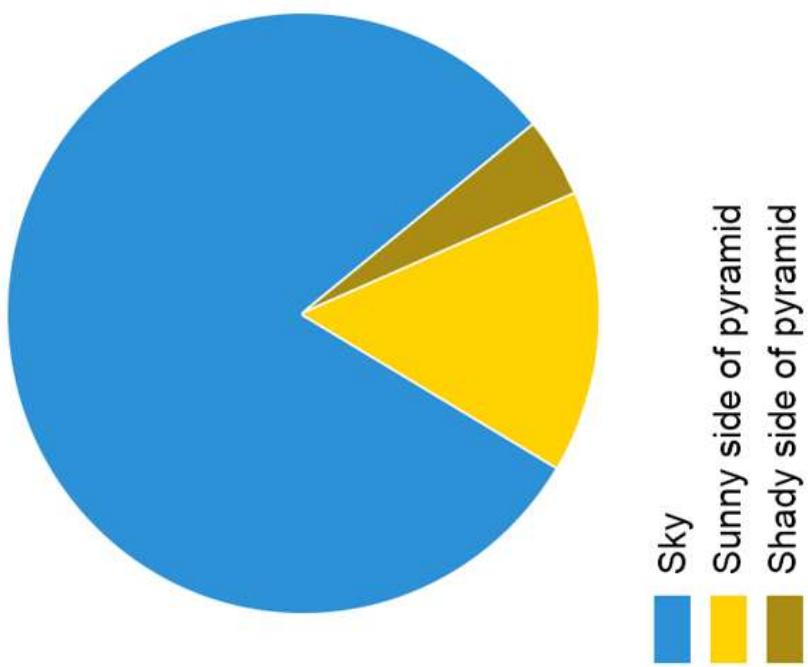


Color Considerations

Rainbow may be pretty, but not accessible (more on this in a bit)

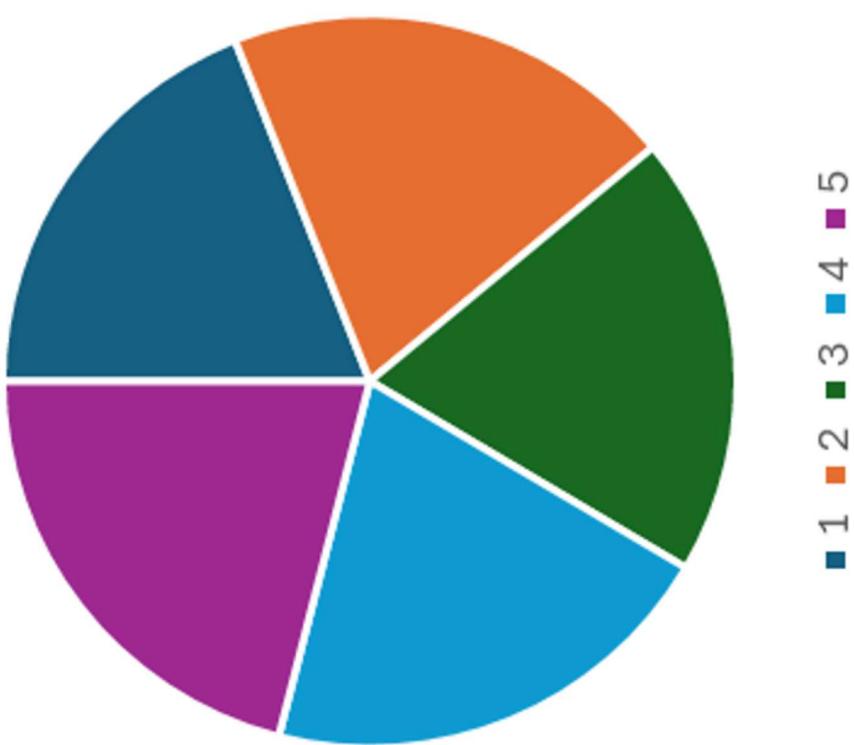
NO PIE CHARTS!

Pyramid Pie Chart



NO PIE CHARTS!

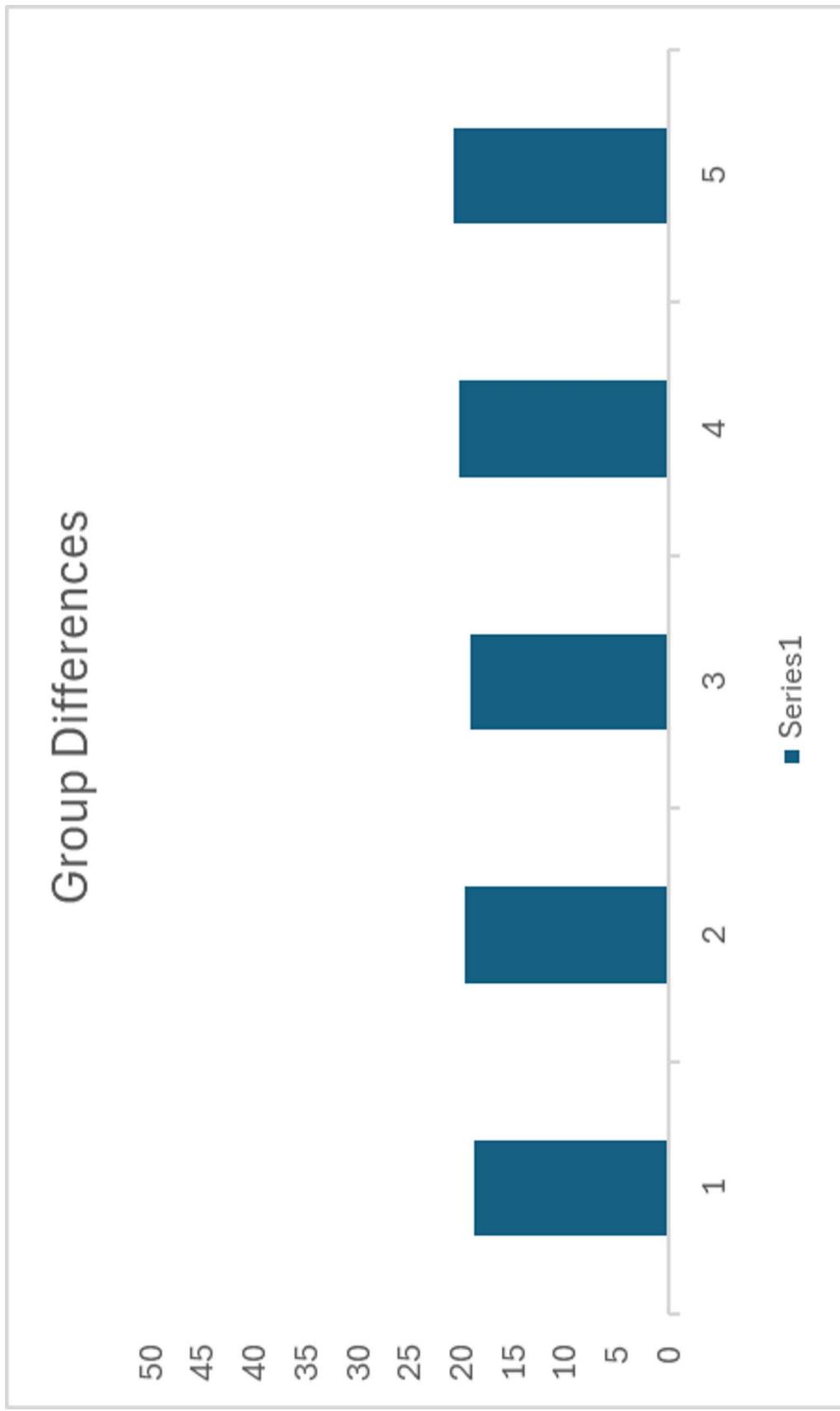
Group Differences



...
...

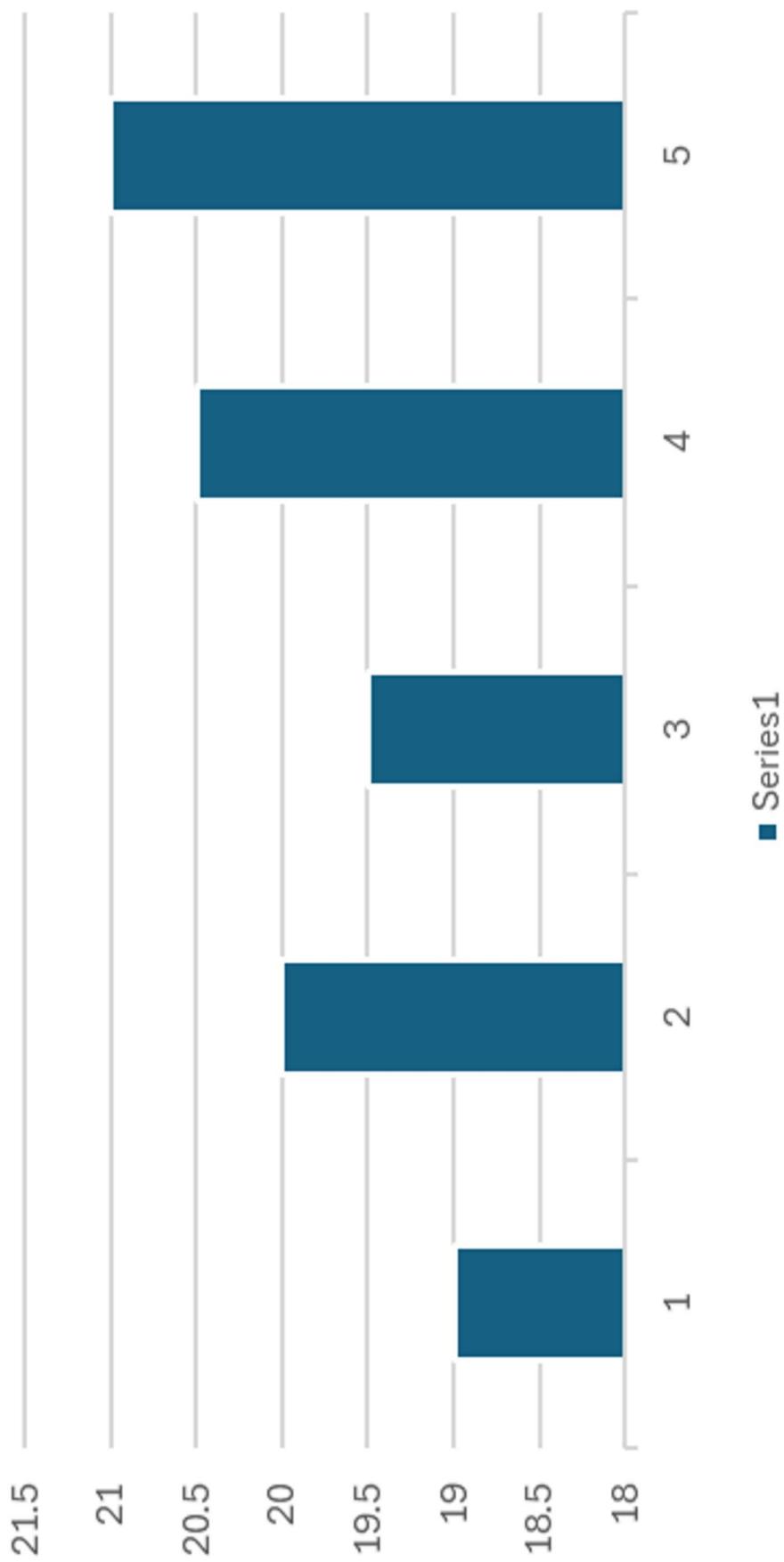
{.notes} humans are pretty bad at telling angles. Can you see what, if any, of the groups are different here? ...

NO PIE CHARTS!



NO PIE CHARTS!

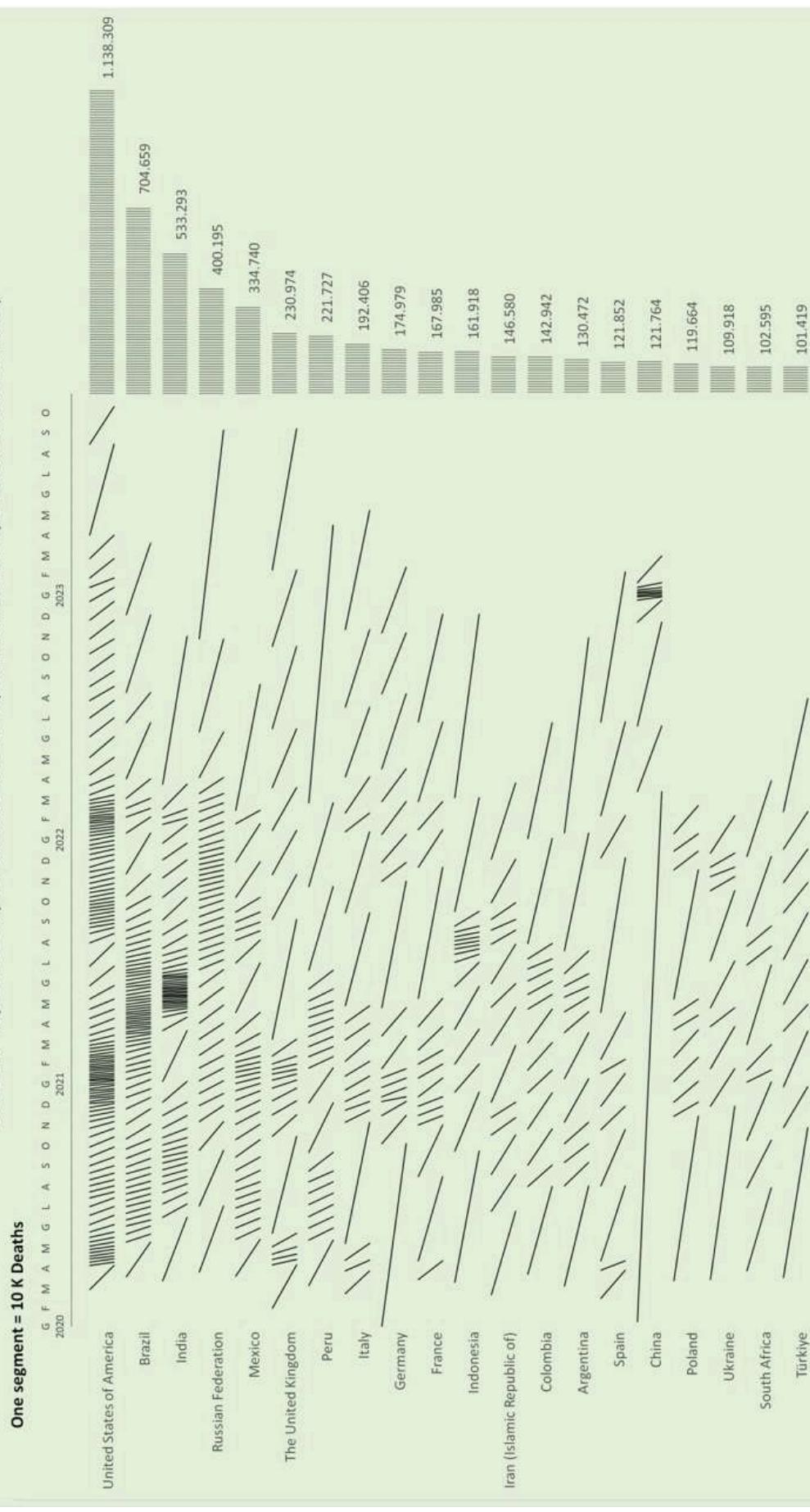
Group Differences



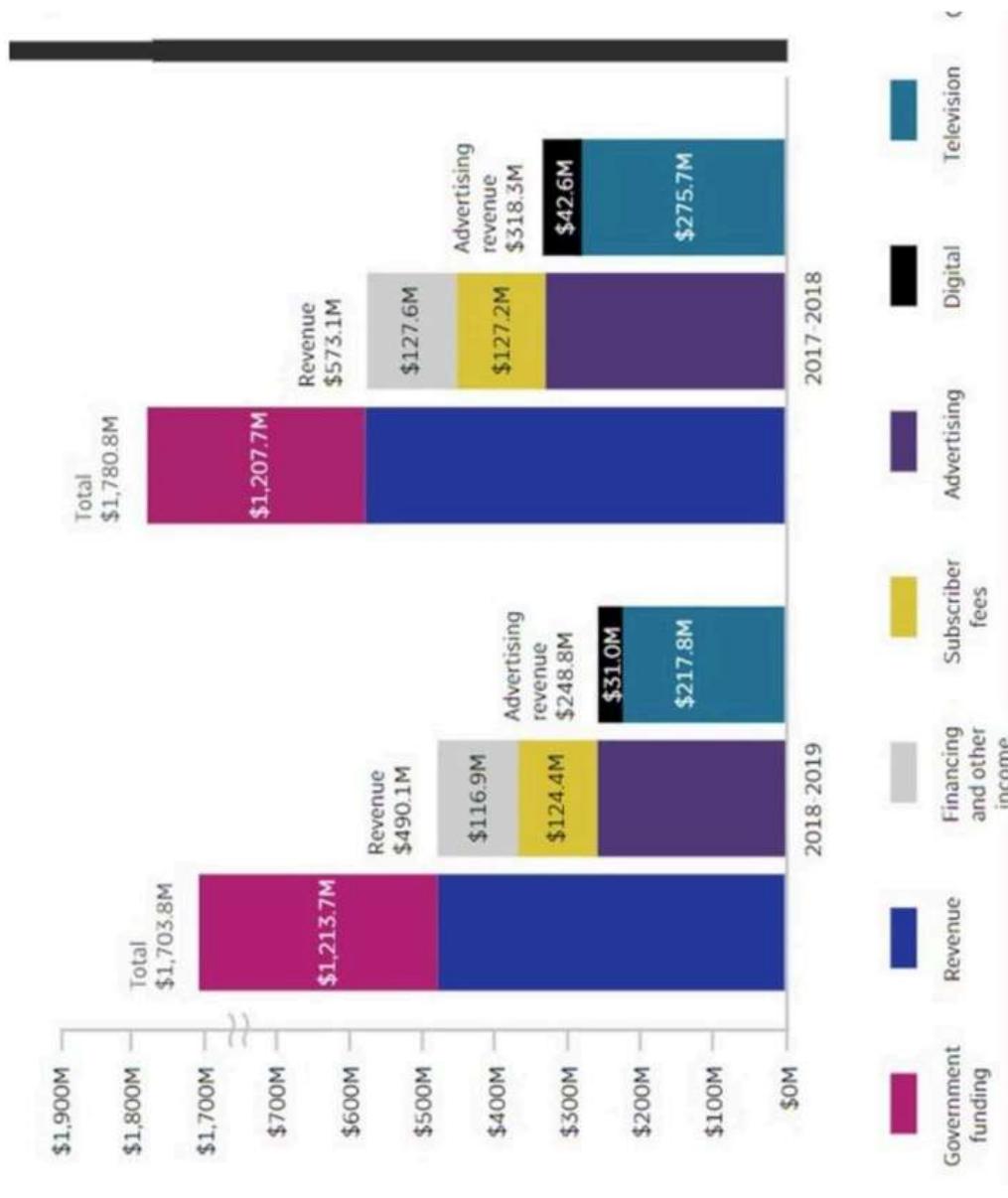
{.notes} Can you tell differences now? This also brings us to our next point - honesty in visualizations :::

Creativity

COVID 19 - Top 20 Country for Cumulative Deaths (data from 5 January 2020 to 30 October 2023)

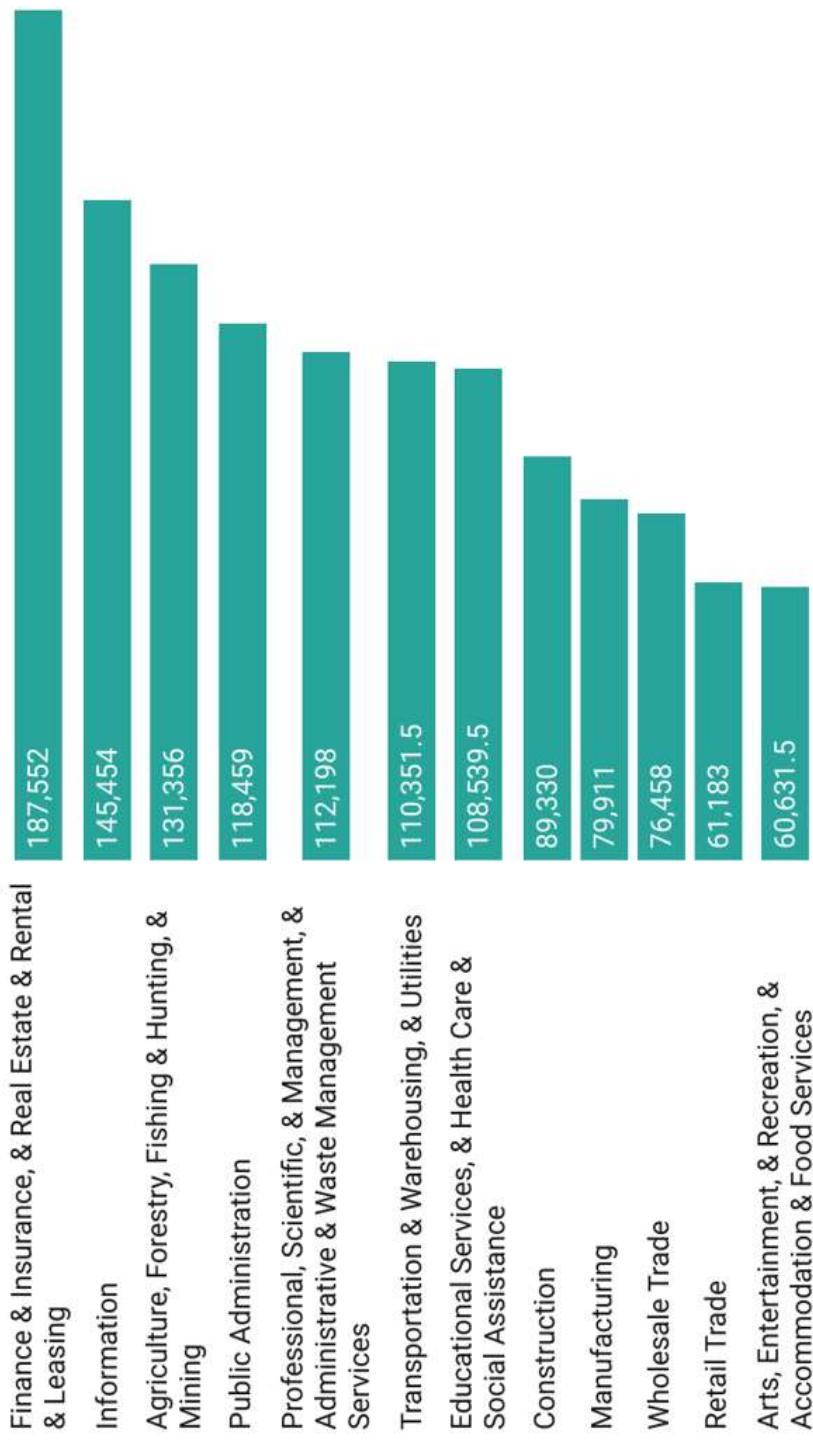


Honesty in Visualizations



Who's Story is Being Told?

Median earnings by industry

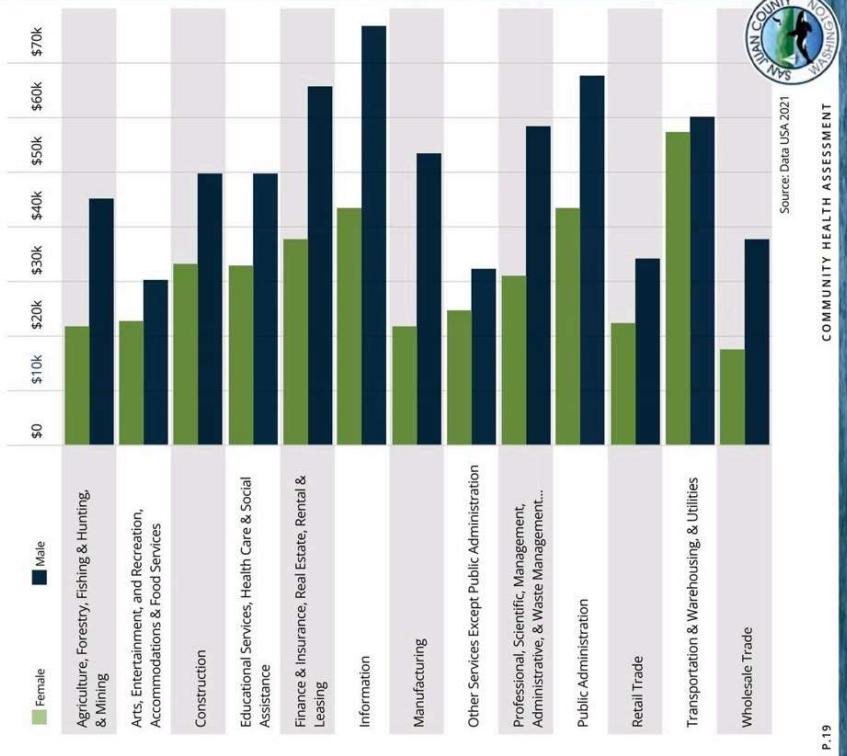


Created with Datawrapper

Who's Story is Being Told?

MEDIAN EARNINGS BY OCCUPATION AND GENDER

Calculating the median means that any very high or very low incomes are not included, like they would be in an average. Instead, the median represents the middle of the most common or most frequent incomes in the county. Earnings by men in San Juan County are higher in every industry than women's earnings in the same industry.



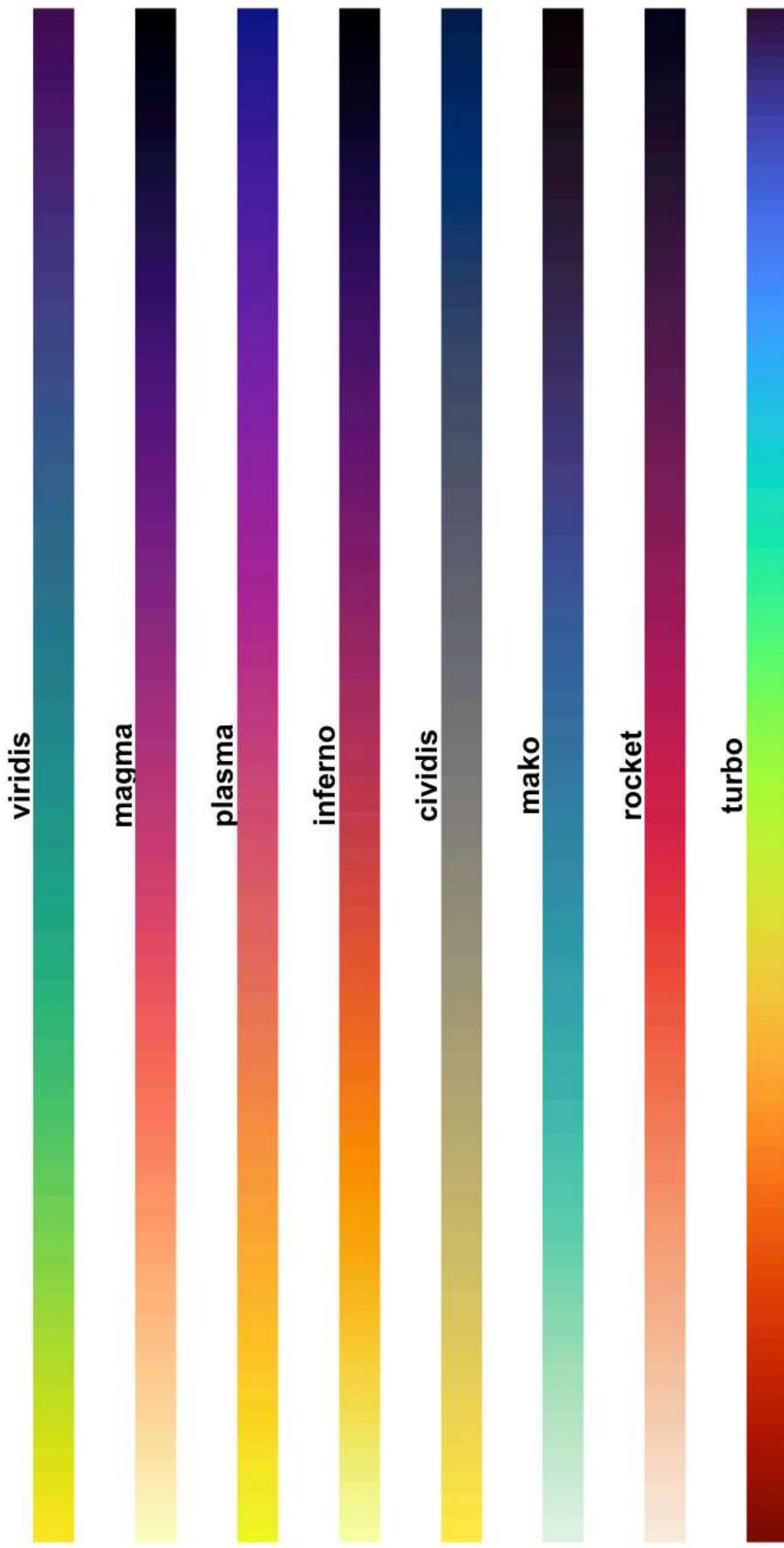
Accessibility Considerations

- Colorblindness is one consideration
 - Multiple types
- Accessibility to color printing another
- Visual contrast
 - Other visual considerations
- Multiple ways to represent groups (color and shape)

Colorblind checkers

- <https://www.Color-blindness.com>
- <https://www.Venngage.com/tools/color-blind-simulator#simulator>

viridis Theme

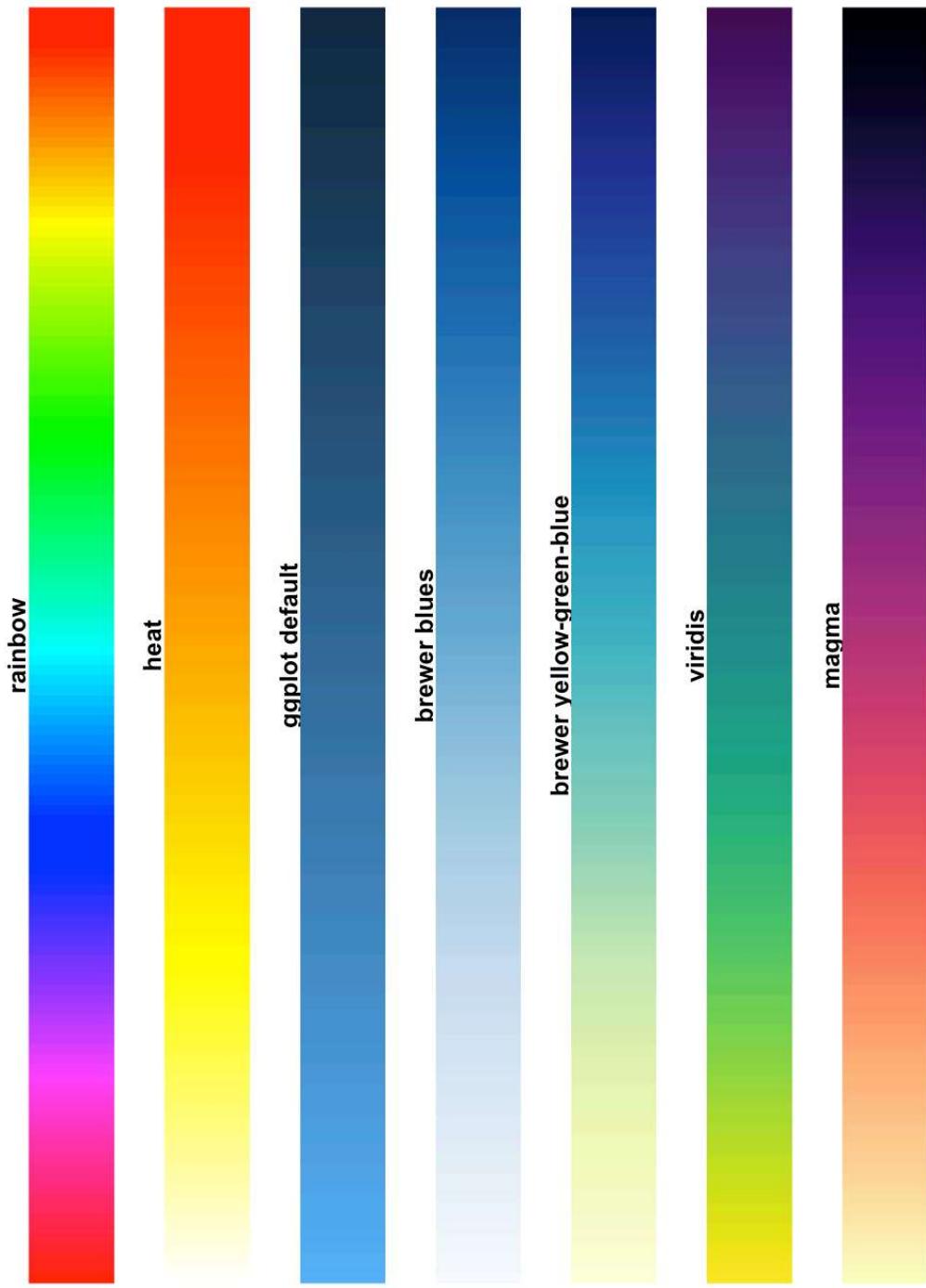


(<https://cran.r-project.org>)

project.org/web/packages/viridis/vignettes/intro-to-viridis.html) ## **viridis** Theme

```
1 #Install the package  
2 install.packages("viridis")  
  
3  
4 #Call the package  
5 library(viridis)
```

viridis Theme - A comparison



viridis Theme - A comparison

Green-Blind (Deuteranopia)

viridis Theme - A comparison

Red-Blind (Protanopia)

viridis Theme - A comparison

Blue-Blind (Tritanopia)

viridis Theme - A comparison

Desaturated

Let's Dive into `ggplot`

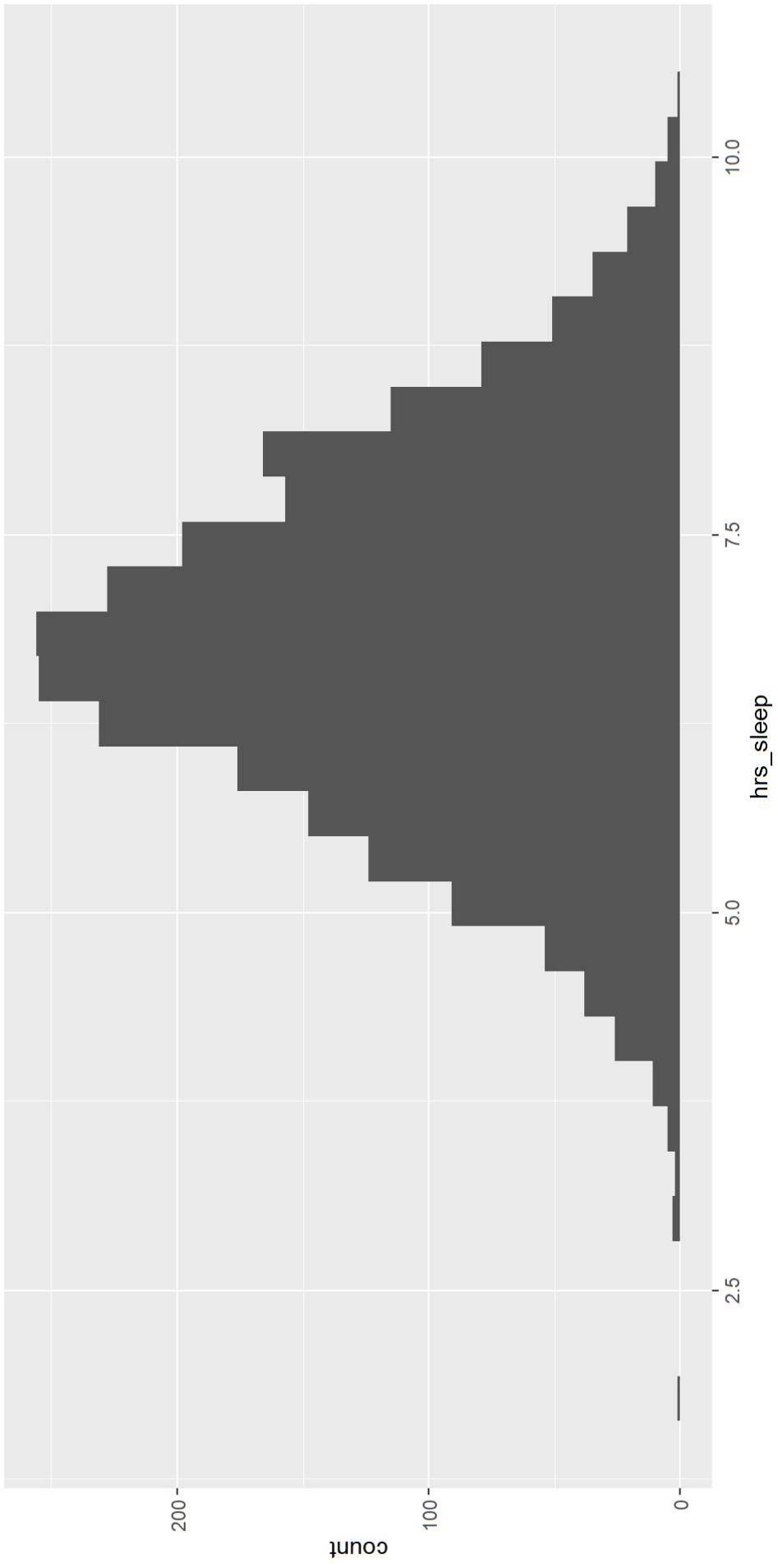
ggplot Structure

- Define data dataframe)
- Choose graph structure (geom)
- Specify variables to use (aesthetics)
- Provide context (titles and labels)
- Make it pretty (themes and colors)

A Basic Graph

```
1 ggplot(data = g) +  
2 geom_histogram(aes(hrs_sleep))
```

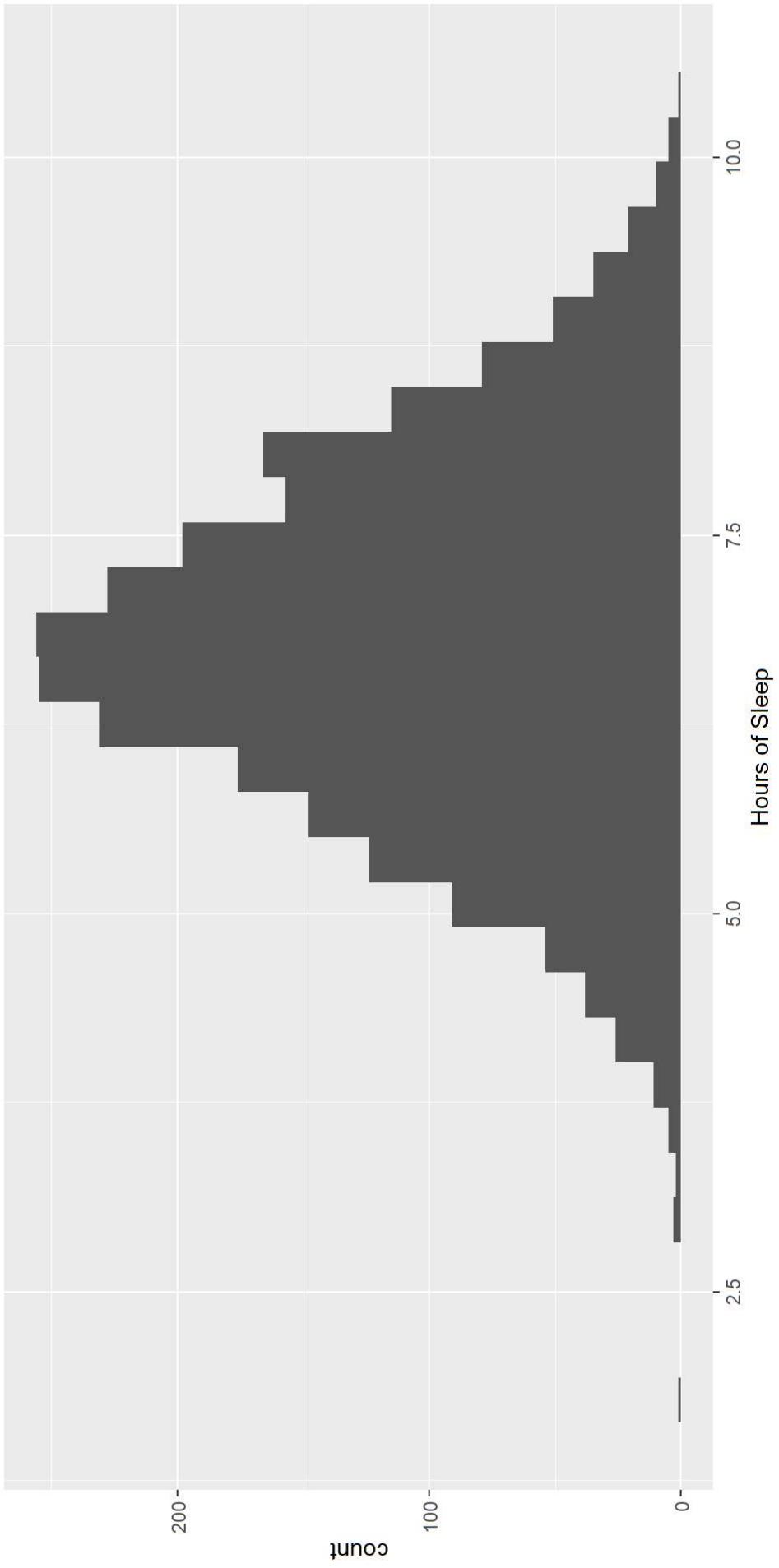
A Basic Graph



Axis Labels

```
1 ggplot(data = g) +  
2   geom_histogram(aes(hrs_sleep)) +  
3   labs(x = "Hours of Sleep")
```

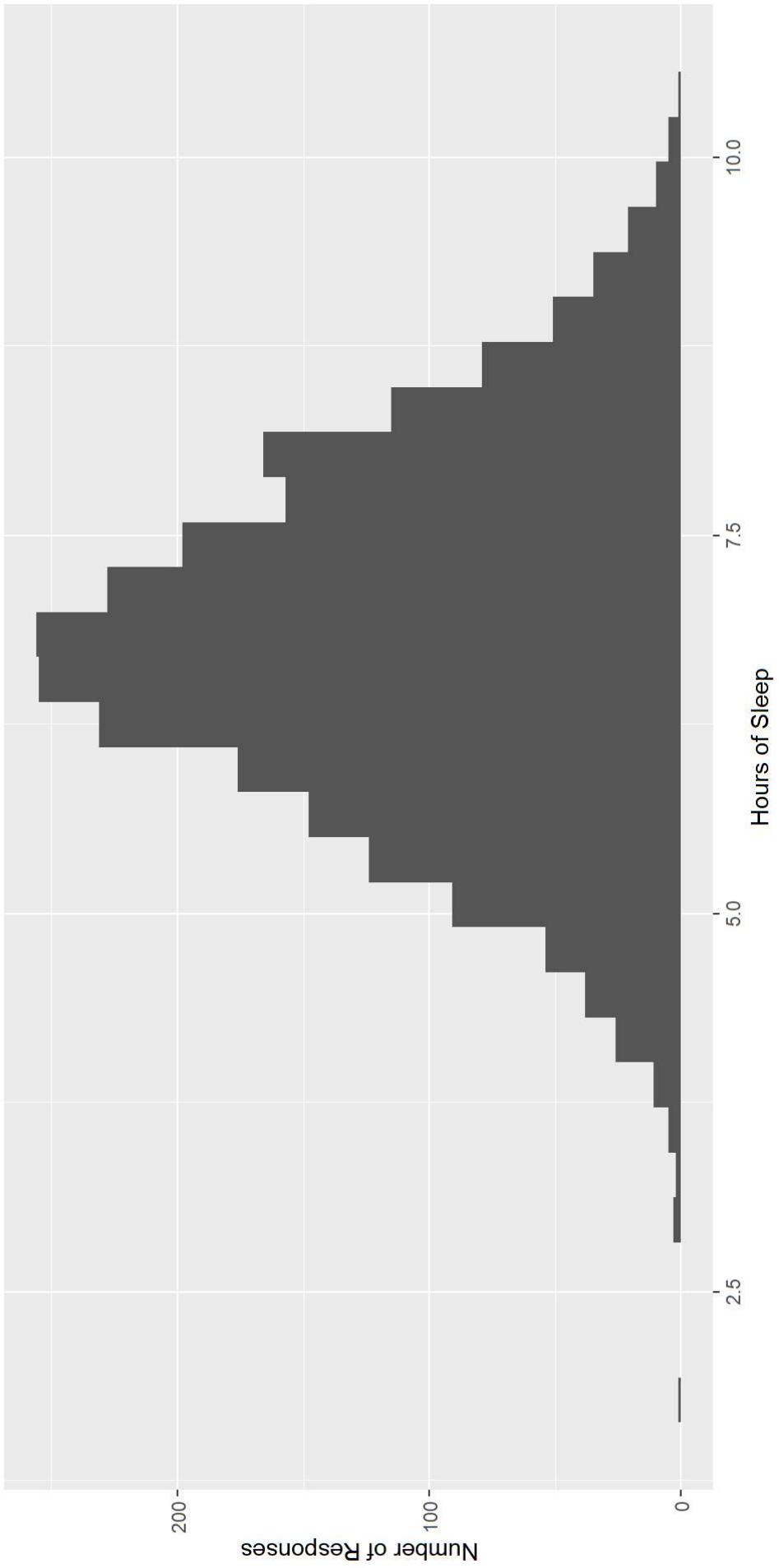
Axis Labels



Axis Labels

```
1 ggplot(data = g) +  
2   geom_histogram(aes(hrs_sleep)) +  
3   labs(x = "Hours of Sleep",  
4         y = "Number of Responses")
```

Axis Labels

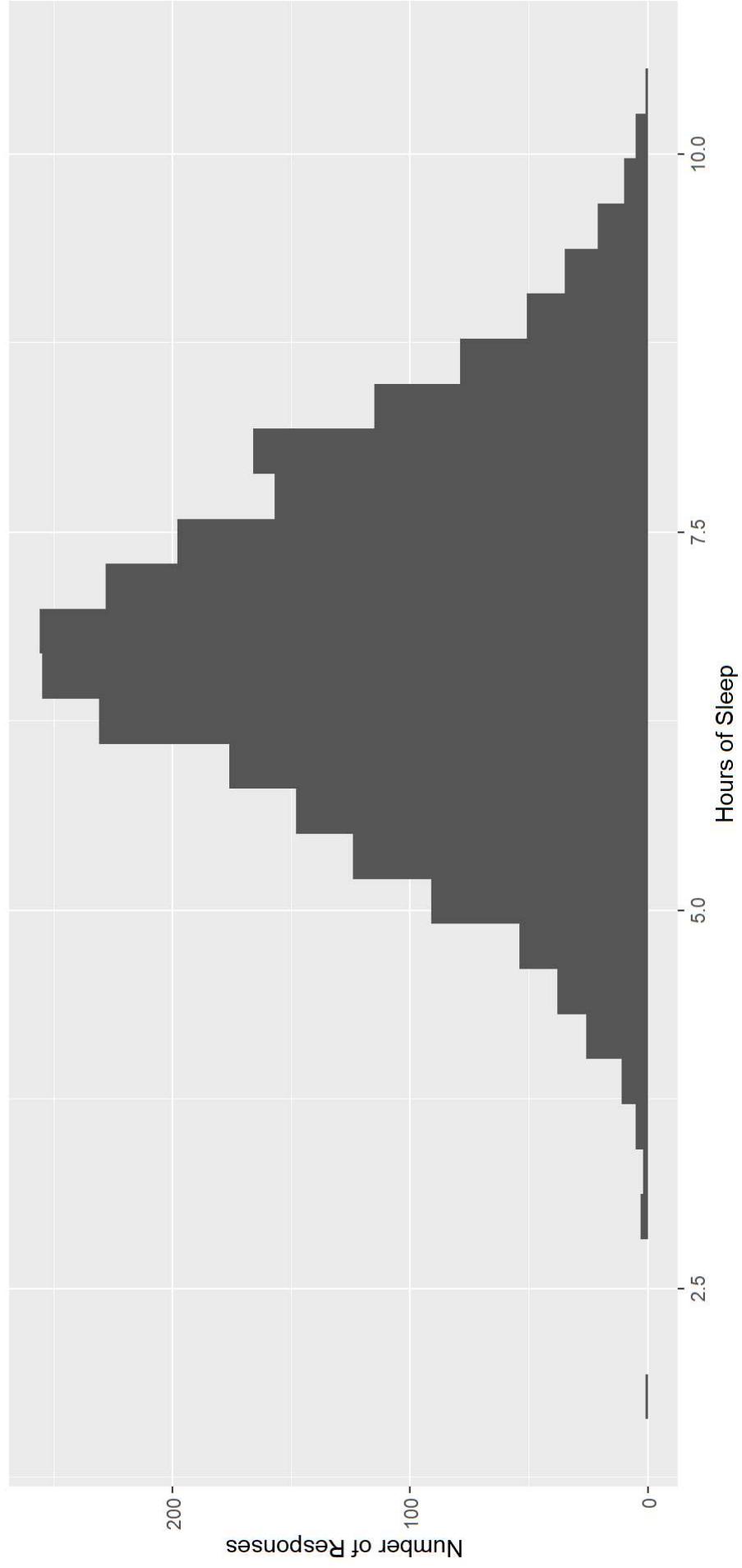


Adding a Title

```
1 ggplot(data = g) +  
2   geom_histogram(aes(hrs_sleep)) +  
3   labs(x = "Hours of Sleep",  
4         Y = "Number of Responses",  
5         title = "Average Hours of Sleep for College Students")
```

Adding a Title

Average Hours of Sleep for College Students

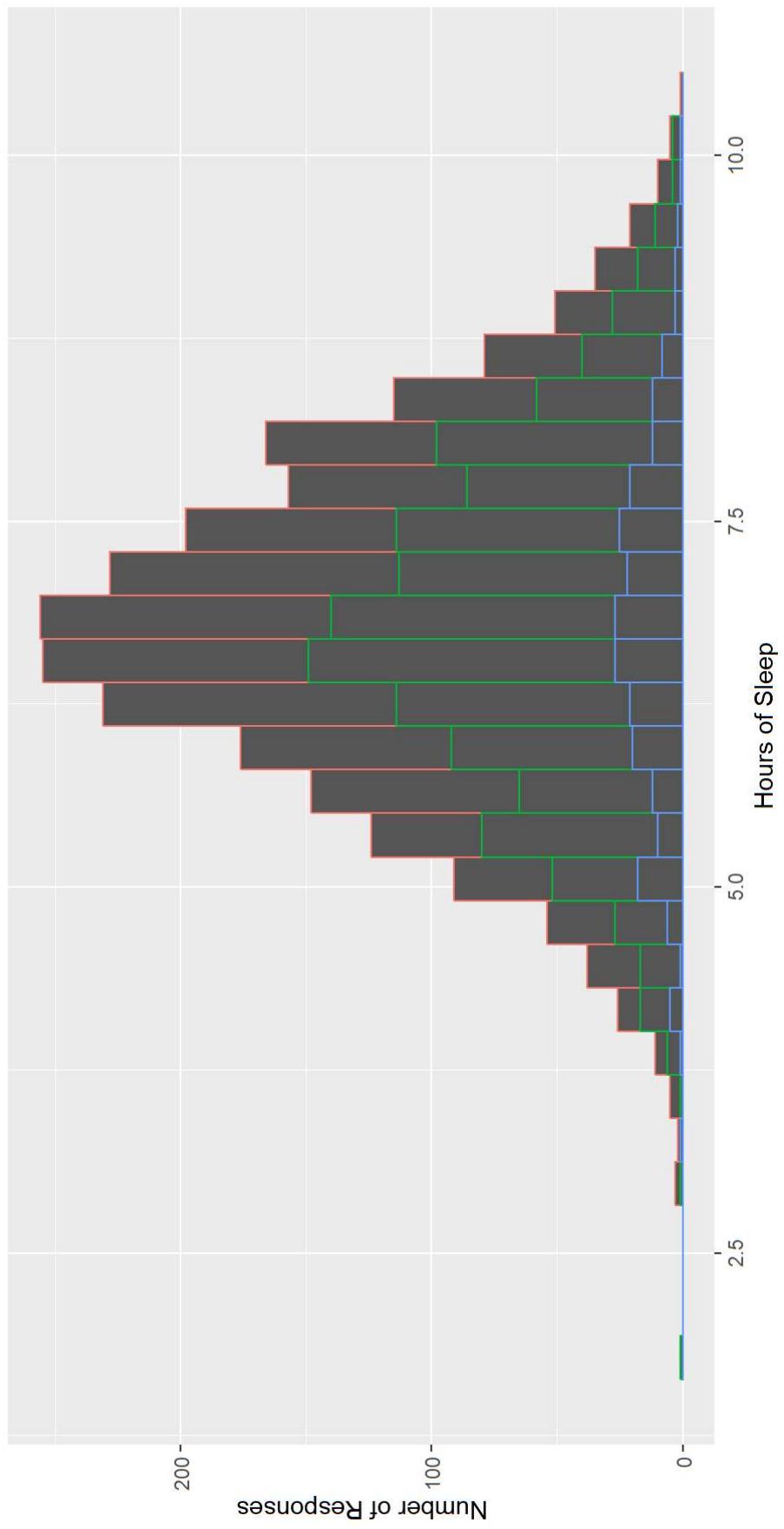


Color by Group

```
1 ggplot(data = g) +
2   geom_histogram(aes(x = hrs_sleep, group = gender, color = gender)) +
3   labs(x = "Hours of Sleep",
4        Y = "Number of Responses",
5        title = "Average Hours of Sleep for College Students")
```

Color by Group

Average Hours of Sleep for College Students

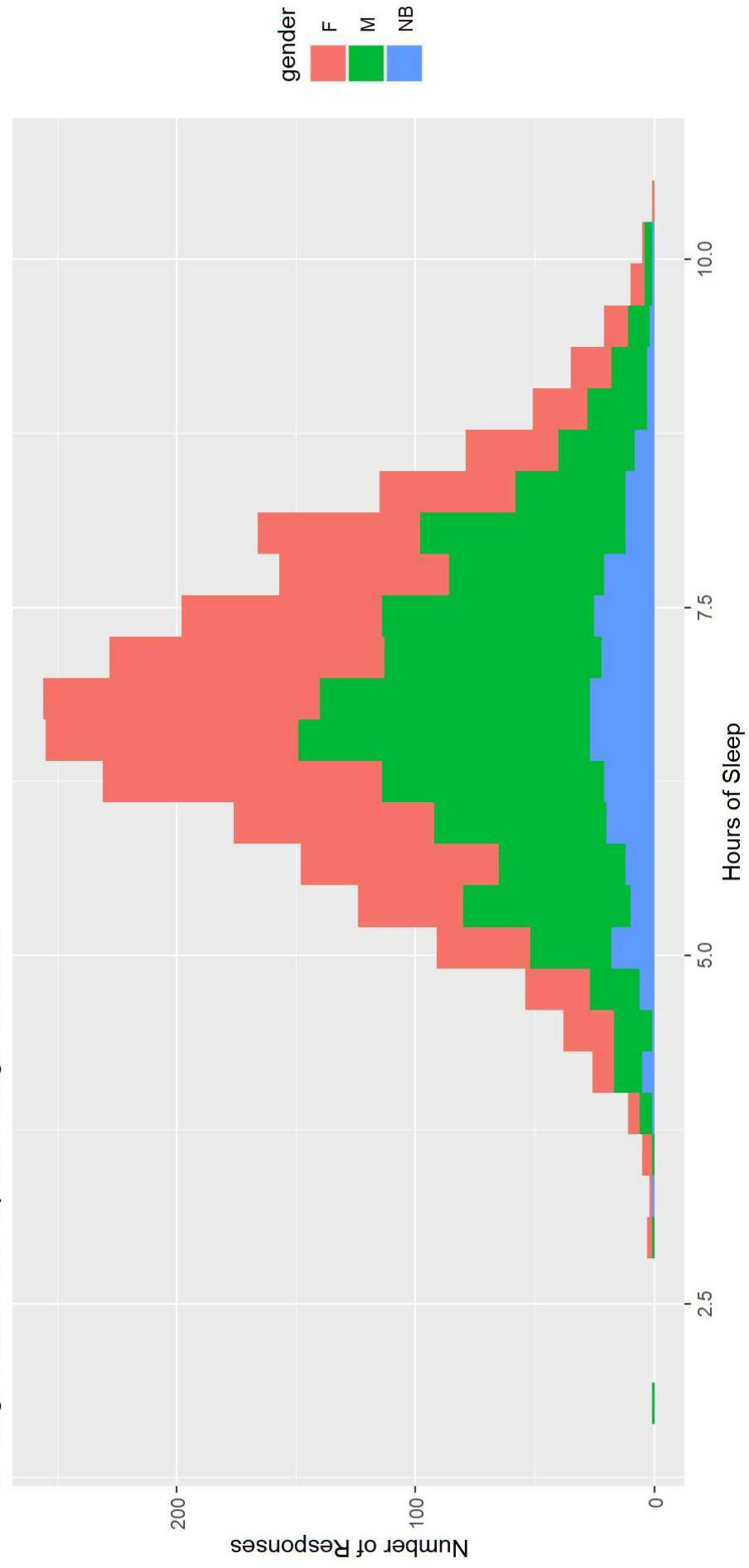


Color by Group

```
1 ggplot(data = g) +
2   geom_histogram(aes(x = hrs_sleep, group = gender, fill = gender)) +
3   labs(x = "Hours of Sleep",
4        Y = "Number of Responses",
5        title = "Average Hours of Sleep for College Students")
```

Color by Group

Average Hours of Sleep for College Students

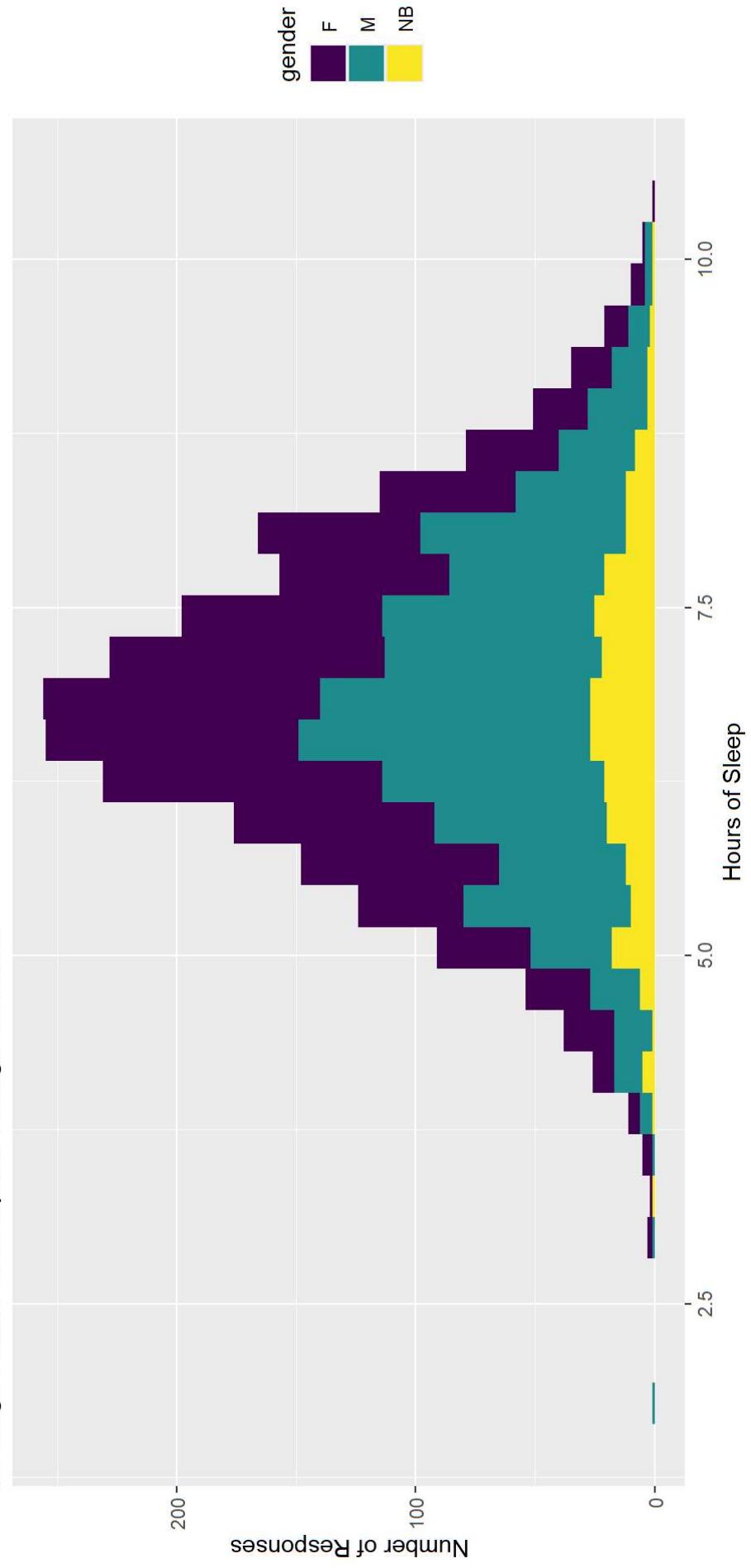


Color by Group

```
1 ggplot(data = g) +
2   geom_histogram(aes(x = hrs_sleep, group = gender, fill = gender)) +
3   labs(x = "Hours of Sleep",
4        Y = "Number of Responses",
5        title = "Average Hours of Sleep for College Students") +
6   scale_fill_viridis(discrete = TRUE)
```

Color by Group

Average Hours of Sleep for College Students

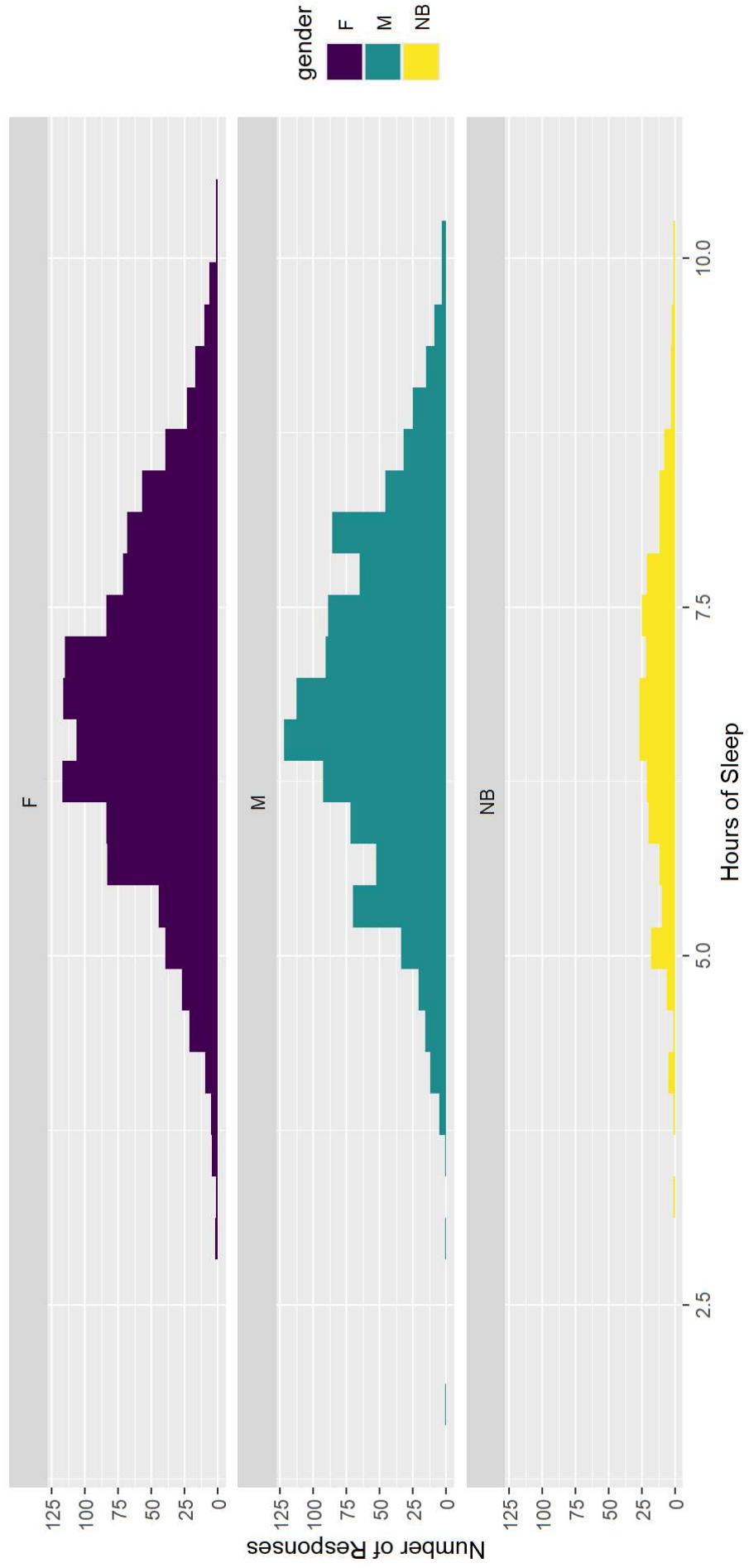


Color by Group

```
1 ggplot(data = g) +
2   geom_histogram(aes(x = hrs_sleep, fill = gender)) +
3     labs(x = "Hours of Sleep",
4       Y = "Number of Responses",
5       title = "Average Hours of Sleep for College Students") +
6     facet_wrap(~ gender, ncol = 1) +
7     scale_fill_viridis(discrete = TRUE)
```

Color by Group

Average Hours of Sleep for College Students

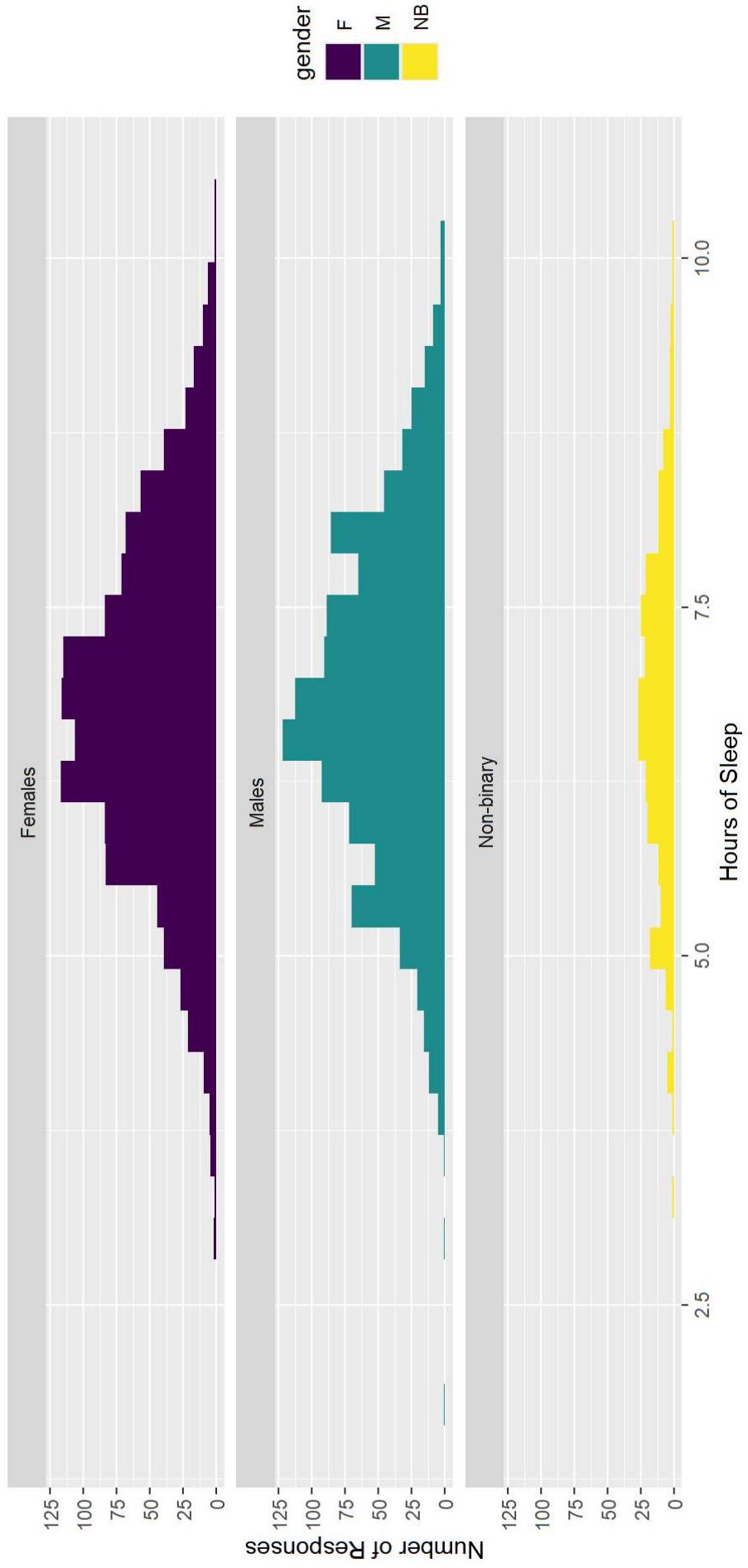


Color by Group

```
1 custom_labels <- c("M" = "Males", "F" = "Females", "NB" = "Non-binary")
2
3 ggplot(data = g) +
4   geom_histogram(aes(x = hrs_sleep, fill = gender)) +
5   labs(x = "Hours of Sleep",
6        y = "Number of Responses",
7        title = "Average Hours of Sleep for College Students") +
8   facet_wrap(~ gender, ncol = 1, labeller = as_labeller(custom_labels)) +
9   scale_fill_viridis(discrete = TRUE)
```

Color by Group

Average Hours of Sleep for College Students

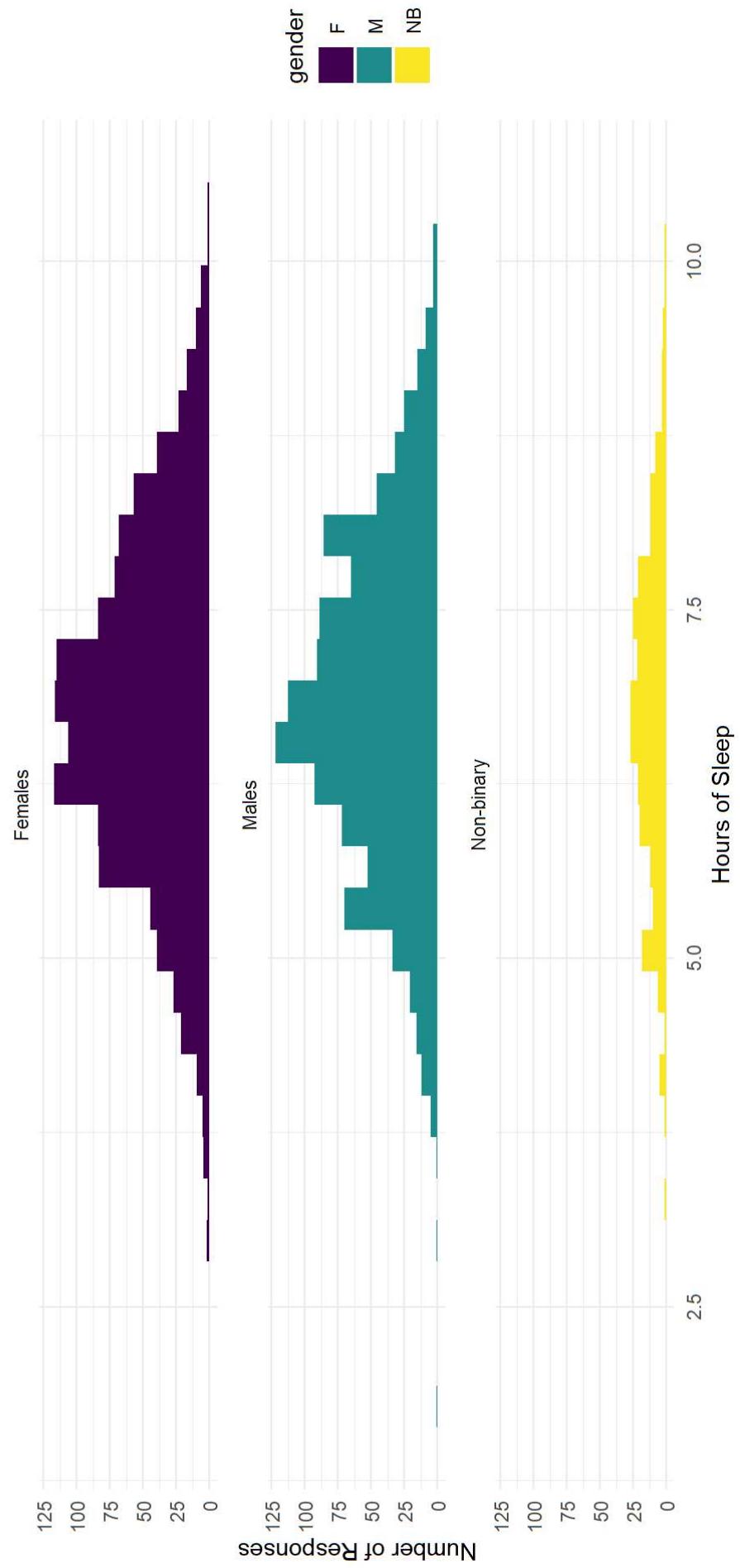


Modifying the Theme

```
1 custom_labels <- c("M" = "Males", "F" = "Females", "NB" = "Non-binary")  
2  
3 ggplot(data = g) +  
4   geom_histogram(aes(x = hrs_sleep, fill = gender)) +  
5   labs(x = "Hours of Sleep",  
6       y = "Number of Responses",  
7       title = "Average Hours of Sleep for College Students") +  
8   facet_wrap(~ gender, ncol = 1, labeller = as_labeller(custom_labels)) +  
9   scale_fill_viridis(discrete = TRUE) +  
10  theme_minimal()
```

Modifying the Theme

Average Hours of Sleep for College Students

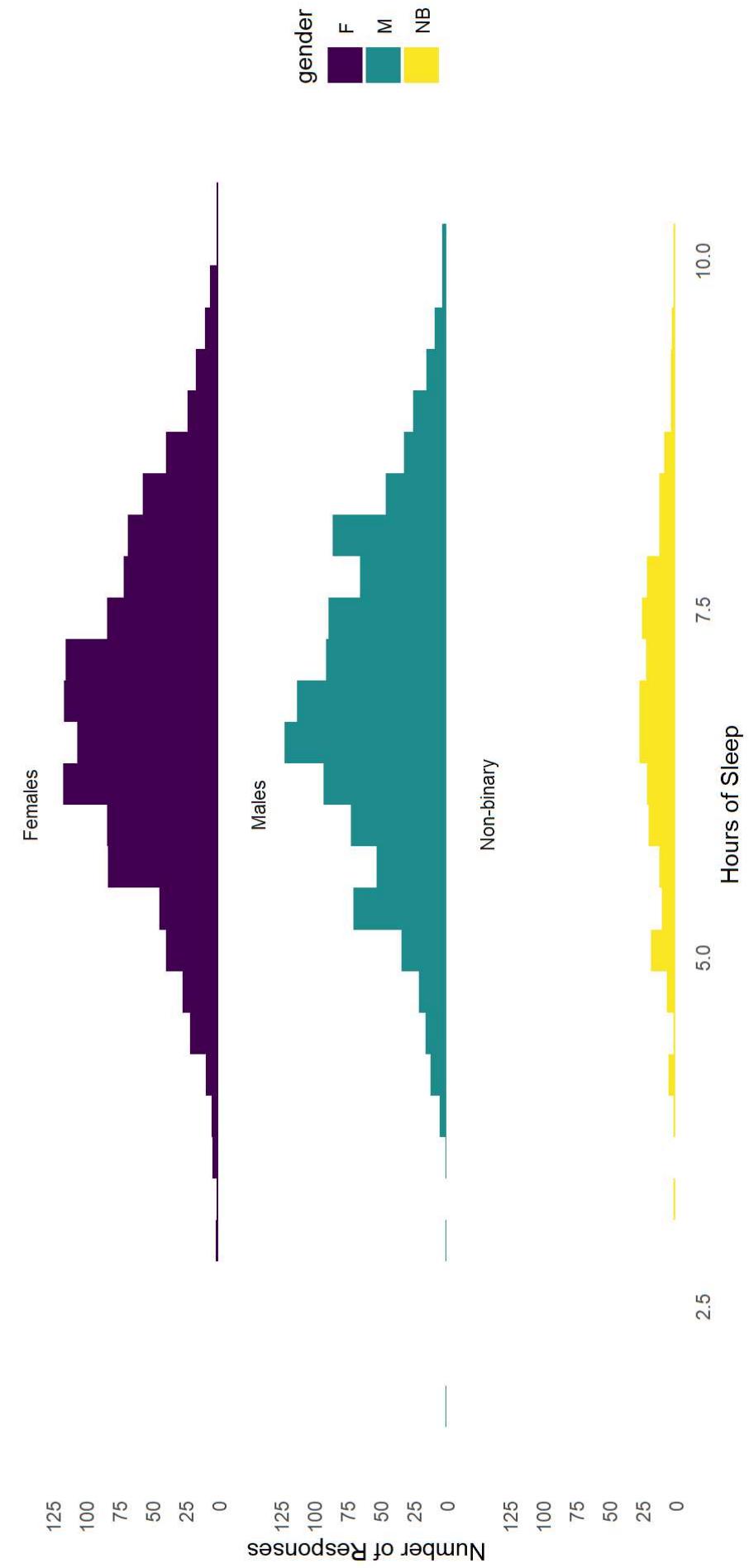


Removing Gridlines

```
1 custom_labels <- c("M" = "Males", "F" = "Females", "NB" = "Non-binary")  
2  
3 ggplot(data = g) +  
4   geom_histogram(aes(x = hrs_sleep, fill = gender)) +  
5   labs(x = "Hours of Sleep",  
6        y = "Number of Responses",  
7        title = "Average Hours of Sleep for College Students") +  
8   facet_wrap(~ gender, ncol = 1, labeller = as_labeller(custom_labels)) +  
9   scale_fill_viridis(discrete = TRUE) +  
10  theme_minimal() +  
11  theme(panel.grid = element_blank())
```

Removing the Gridlines

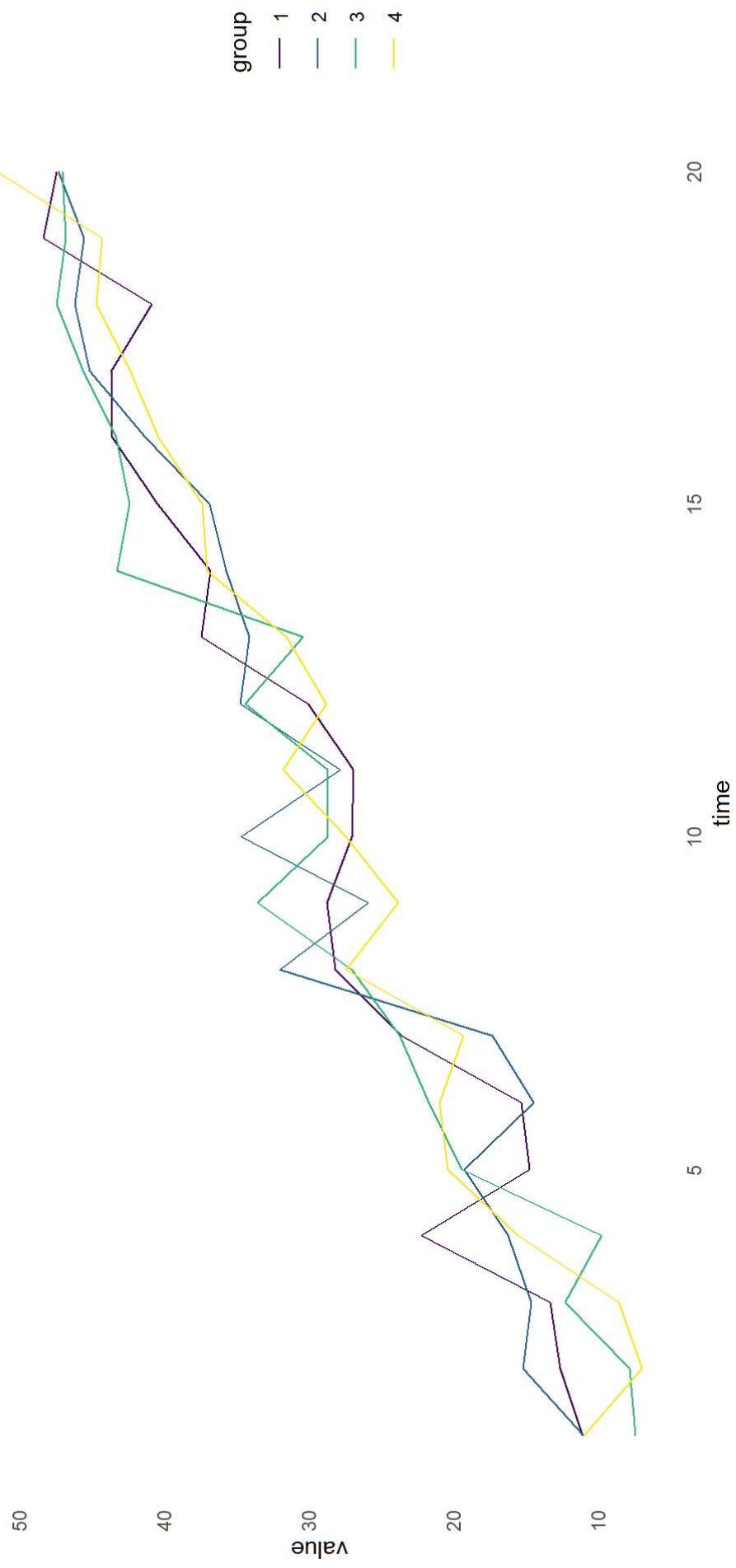
Average Hours of Sleep for College Students



Going to a Different Graph

```
1 ggplot(df_shape, aes(x = time, y = value, color = group) ) +  
2   geom_line() +  
3   scale_color_viridis(discrete = TRUE) +  
4   theme_minimal() +  
5   theme(panel.grid = element_blank() )
```

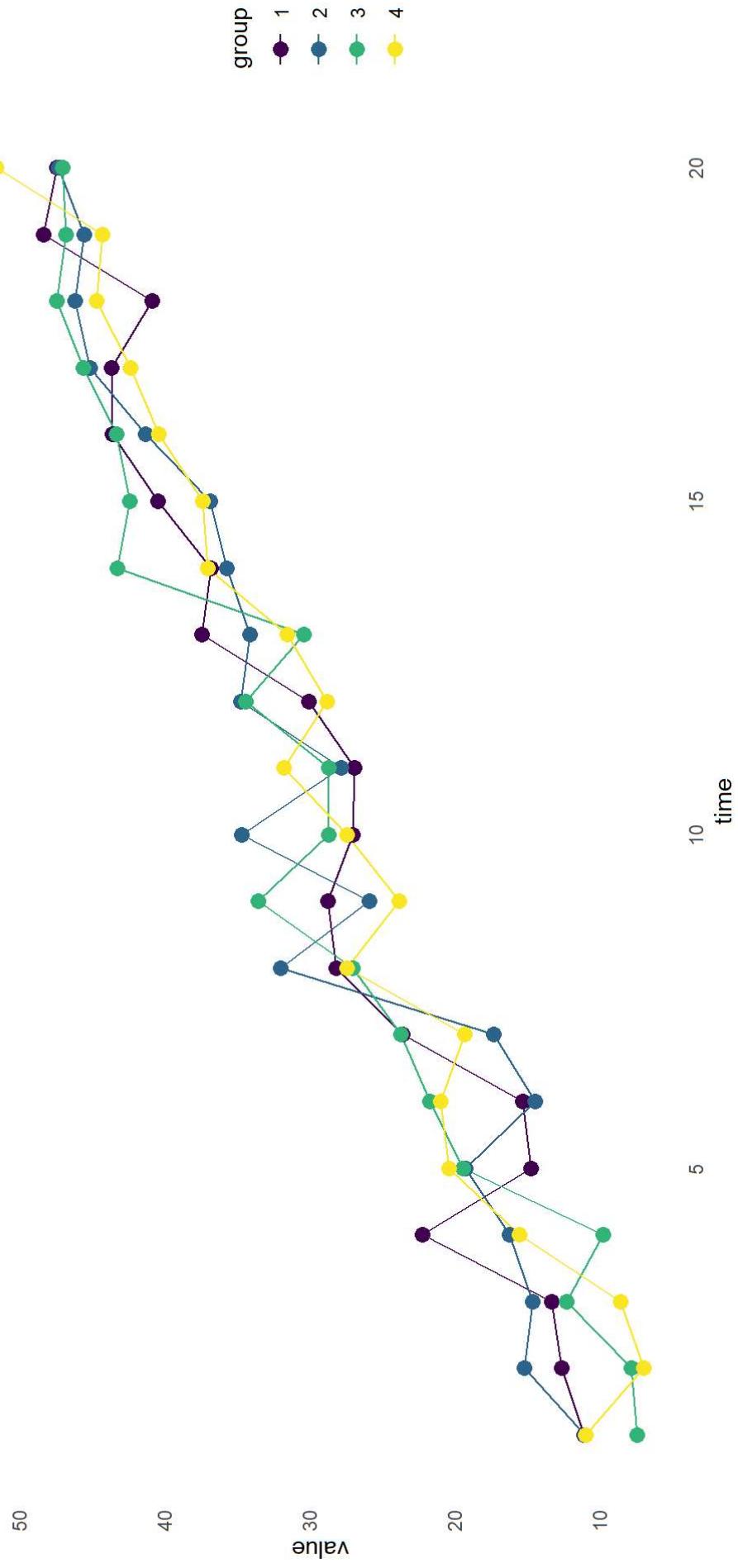
Going to a Different Graph



Multiple Layers

```
1 ggplot(df_shape, aes(x = time, y = value, color = group) ) +  
2   geom_line() +  
3   geom_point(size = 3) +  
4   scale_color_viridis(discrete = TRUE) +  
5   theme_minimal() +  
6   theme(panel.grid = element_blank() )
```

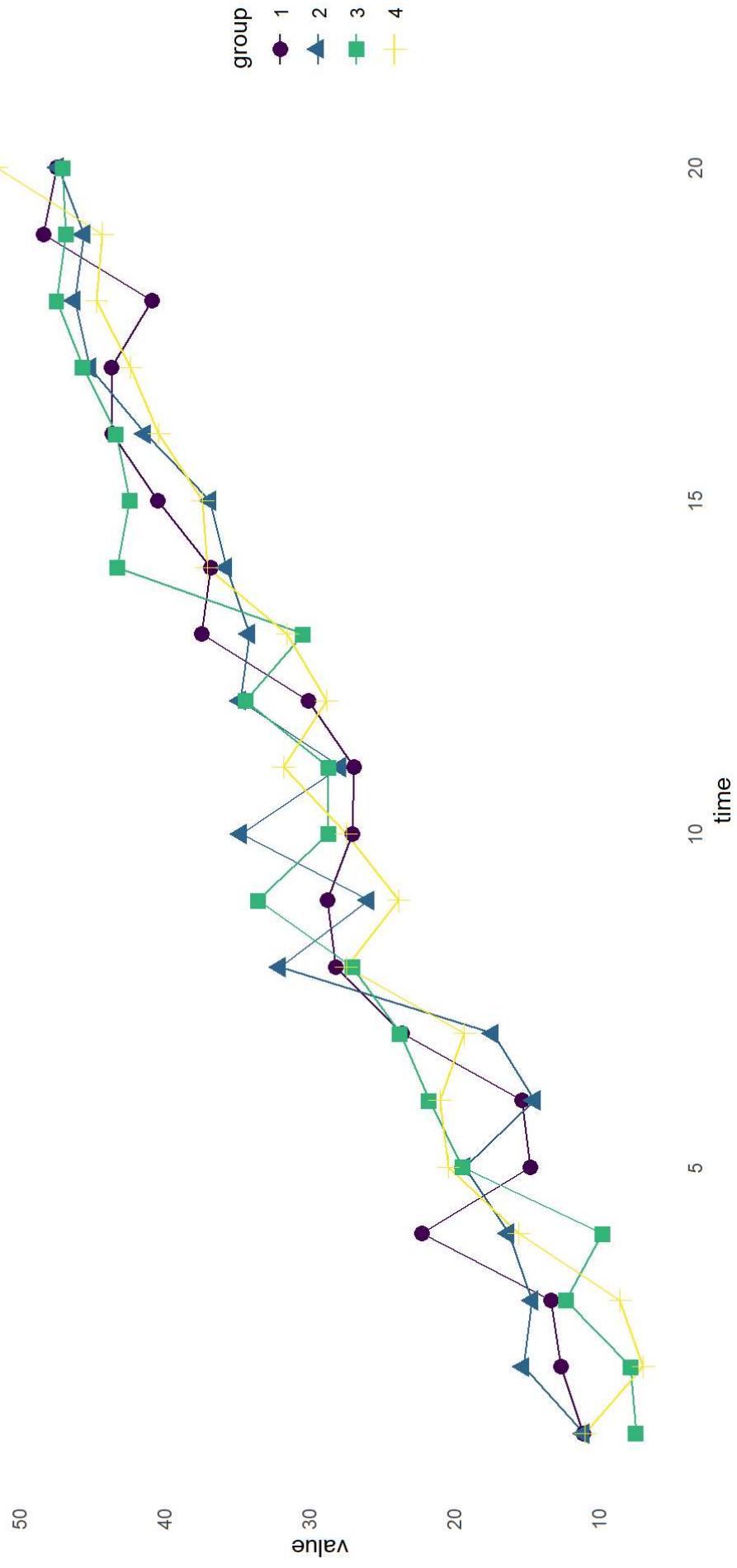
Multiple Layers



Adding shapes

```
1 ggplot(df_shape, aes(x = time, y = value, color = group, group = group)) +  
2   geom_line() +  
3   geom_point(aes(shape = group), size = 3) +  
4   scale_color_viridis(discrete = TRUE) +  
5   theme_minimal() +  
6   theme(panel.grid = element_blank())
```

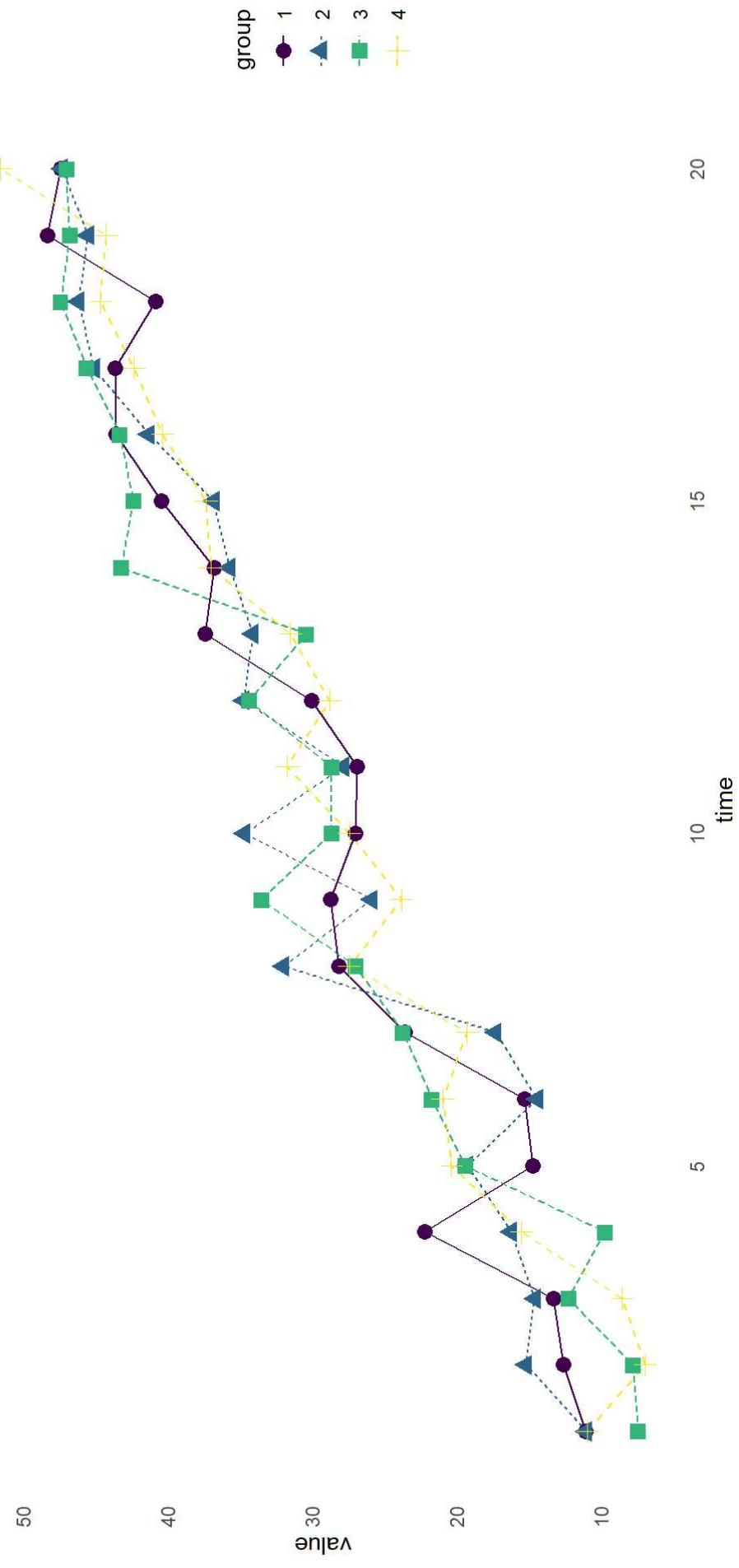
Adding Shapes



Line Types

```
1 ggplot(df_shape, aes(x = time, y = value, color = group, lin  
2 geom_line() +  
3 geom_point(aes(shape = group), size = 3) +  
4 scale_color_viridis(discrete = TRUE) +  
5 theme_minimal() +  
6 theme(panel.grid = element_blank())
```

Line Types



Resources

- Google
- R-Graph gallery (<https://r-graph-gallery.com>)
- R Graphics Cookbook (<https://r-graphics.org/>)

Time for Questions

Laura Lambert laycoclajmu.edu

