Applied Statistics for Data Scientists with R

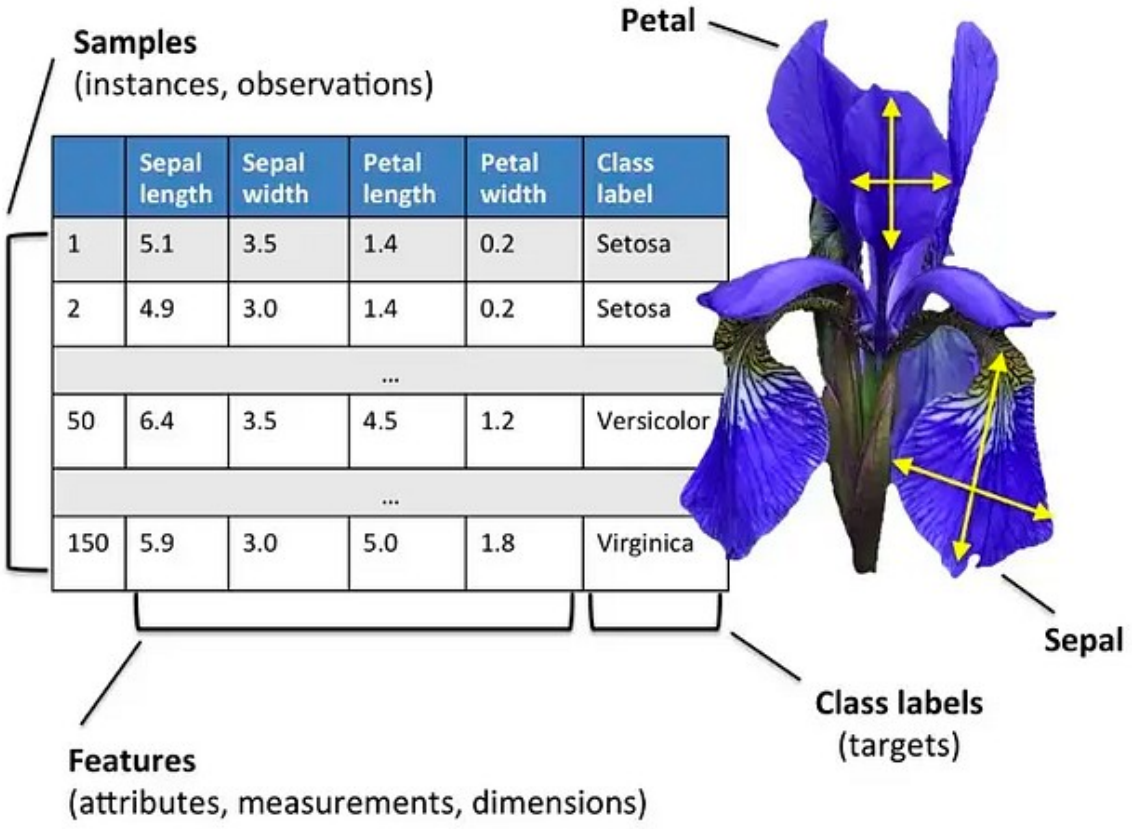
Class 10: Data Visualization with ggplot2

Learning Objective

1. Create publication ready reproducible plots
2. Helper packages and Add-ins for ggplot2

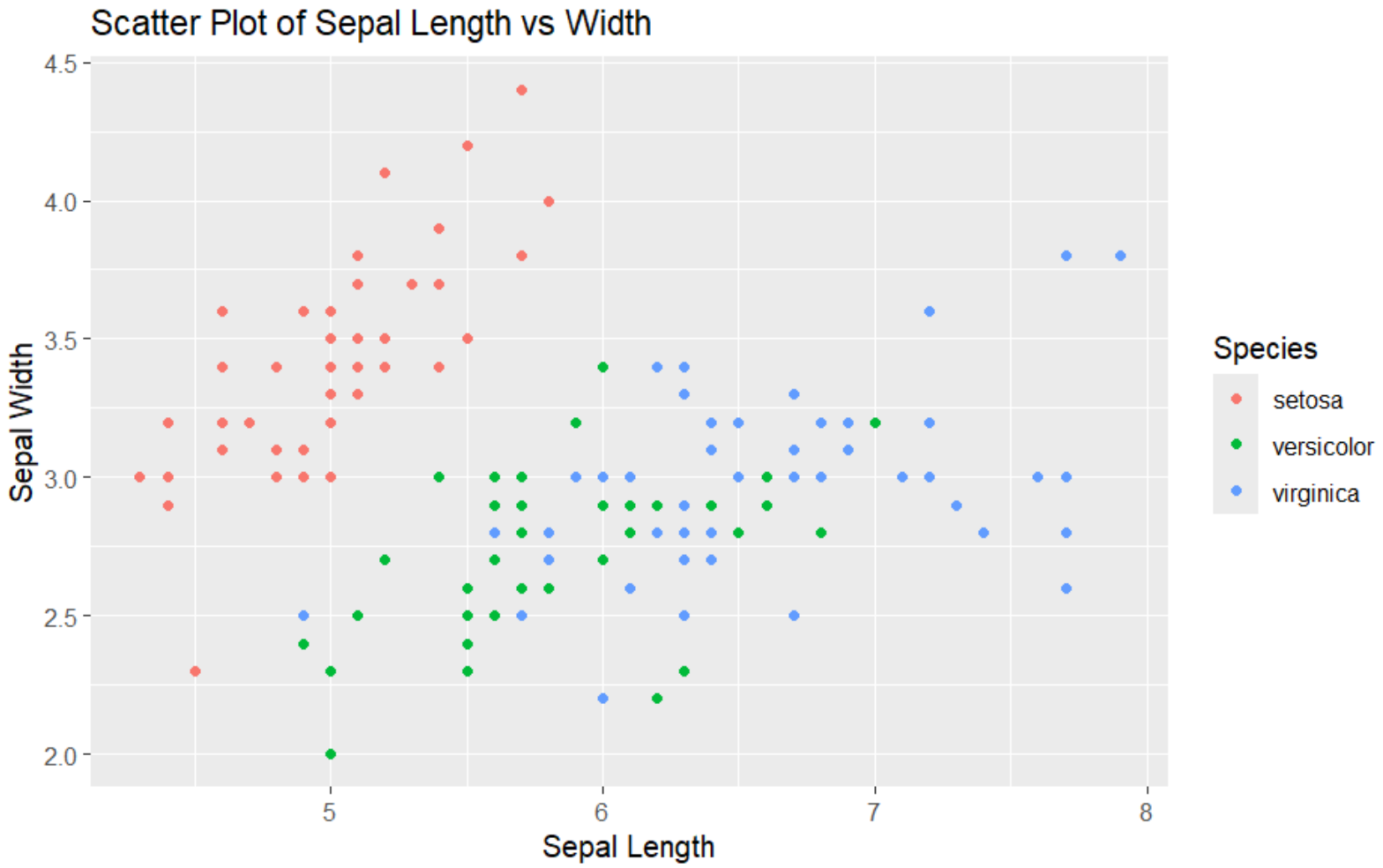
Data Description

- In this session we will be using the iris dataset

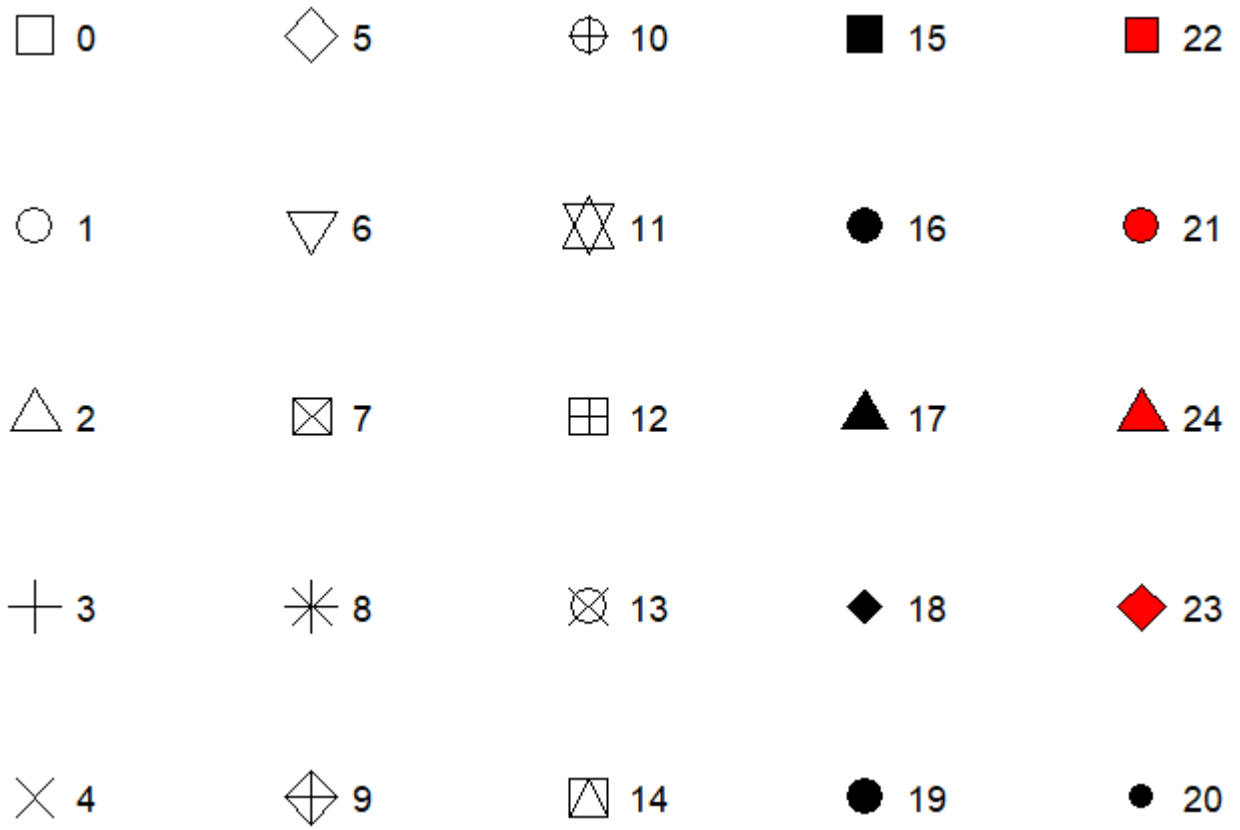


	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

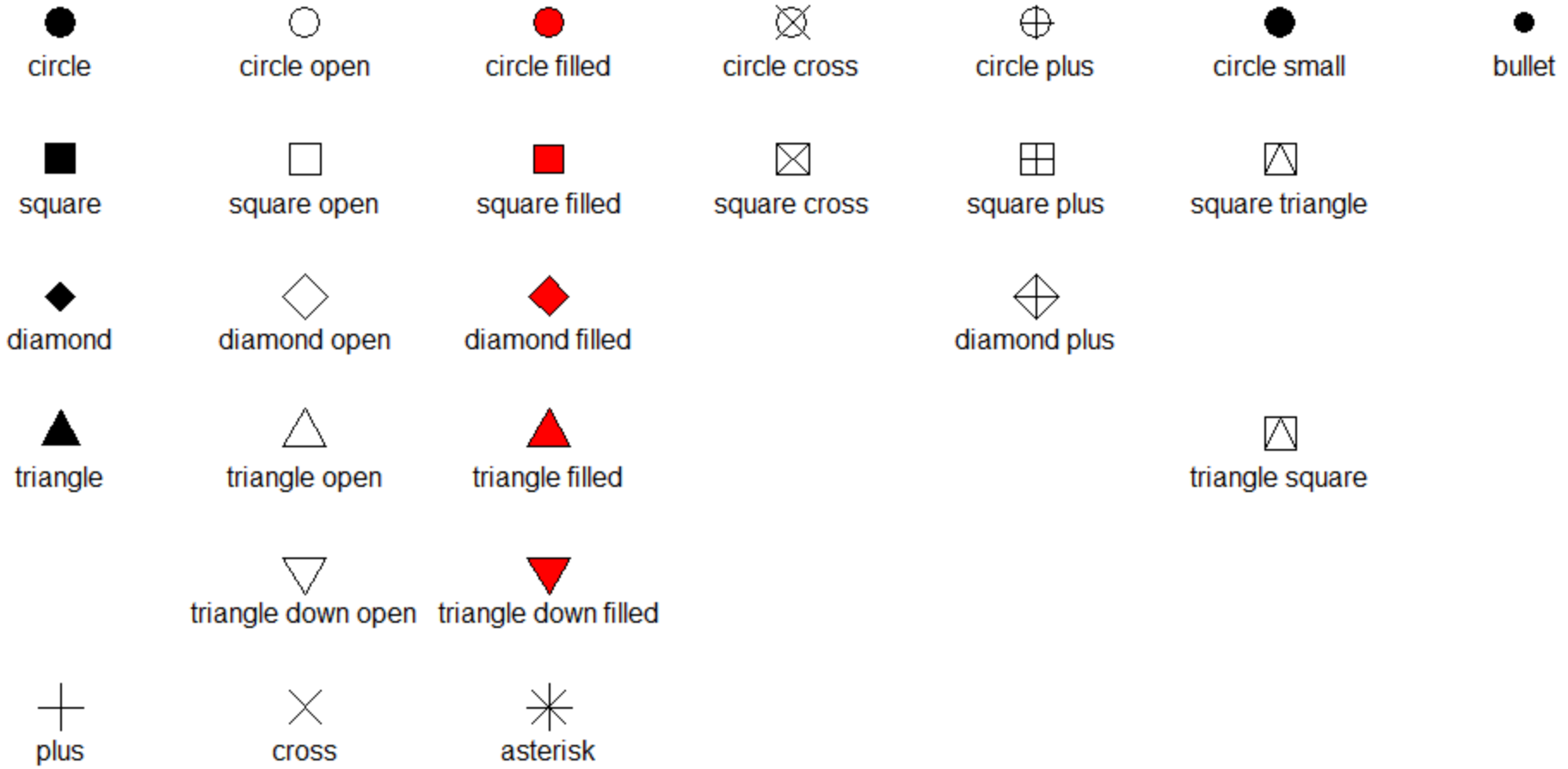
Scatter Plot



Shapes

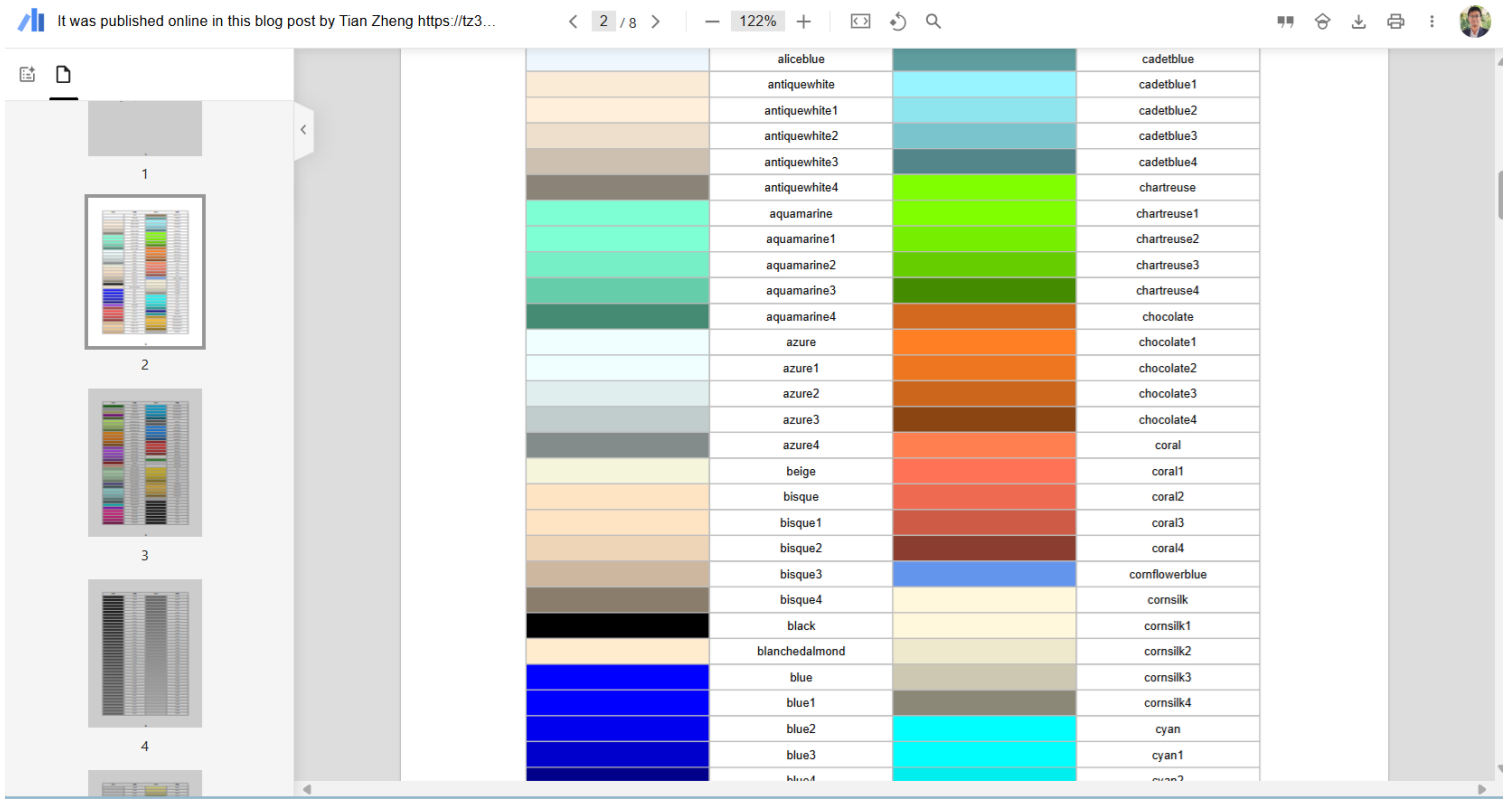


Shapes

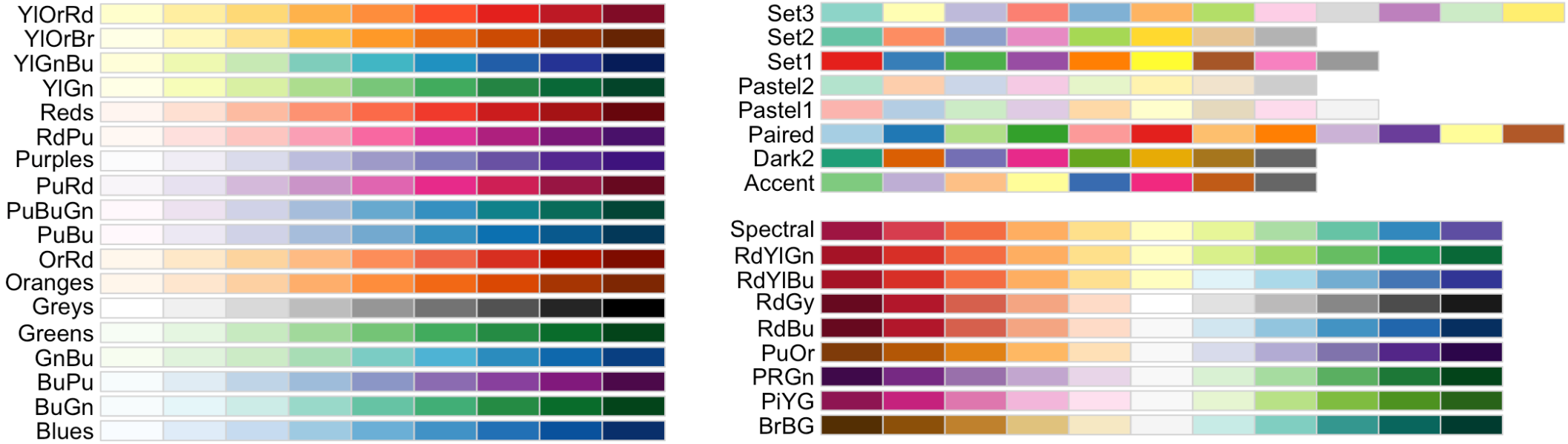

circle
circle open
circle filled
circle cross
circle plus
circle small
bullet
square
square open
square filled
square cross
square plus
square triangle
diamond
diamond open
diamond filled
diamond plus
triangle
triangle open
triangle filled
triangle square
triangle down open
triangle down filled
plus
cross
asterisk

Colors

- stat.columbia.edu/~tzheng/files/Rcolor.pdf



Palettes


YlOrRd
YlOrBr
YlGnBu
YlGn
Reds
RdPu
Purples
PuRd
PuBuGn
PuBu
OrRd
Oranges
Greys
Greens
GnBu
BuPu
BuGn
Blues
Set3
Set2
Set1
Pastel2
Pastel1
Paired
Dark2
Accent
Spectral
RdYlGn
RdYlBu
RdGy
RdBu
PuOr
PRGn
PiYG
BrBG

Themes

- https://r-graph-gallery.com/ggplot2-package.html

→ ggplot2

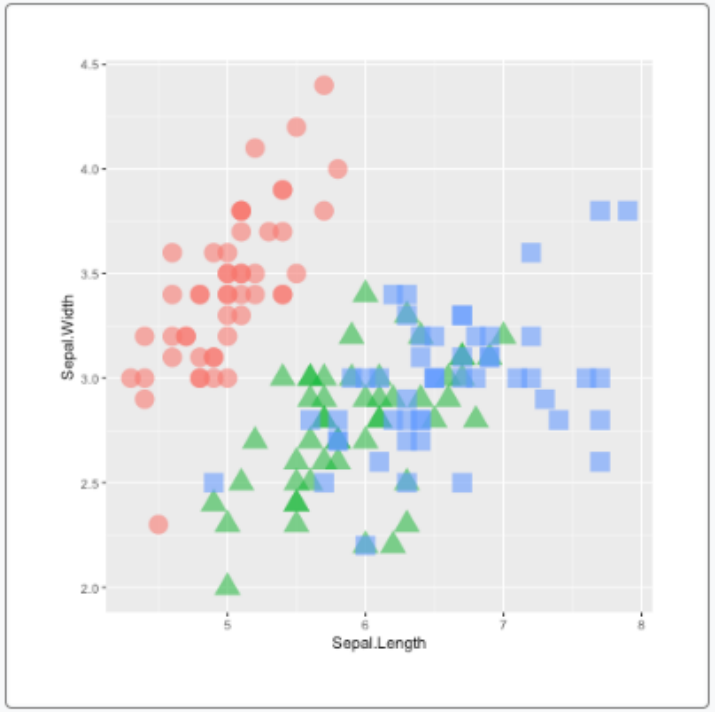
`default` `theme_bw()` `theme_minimal()`
`theme_classic()` `theme_gray()`

→ ggthemes

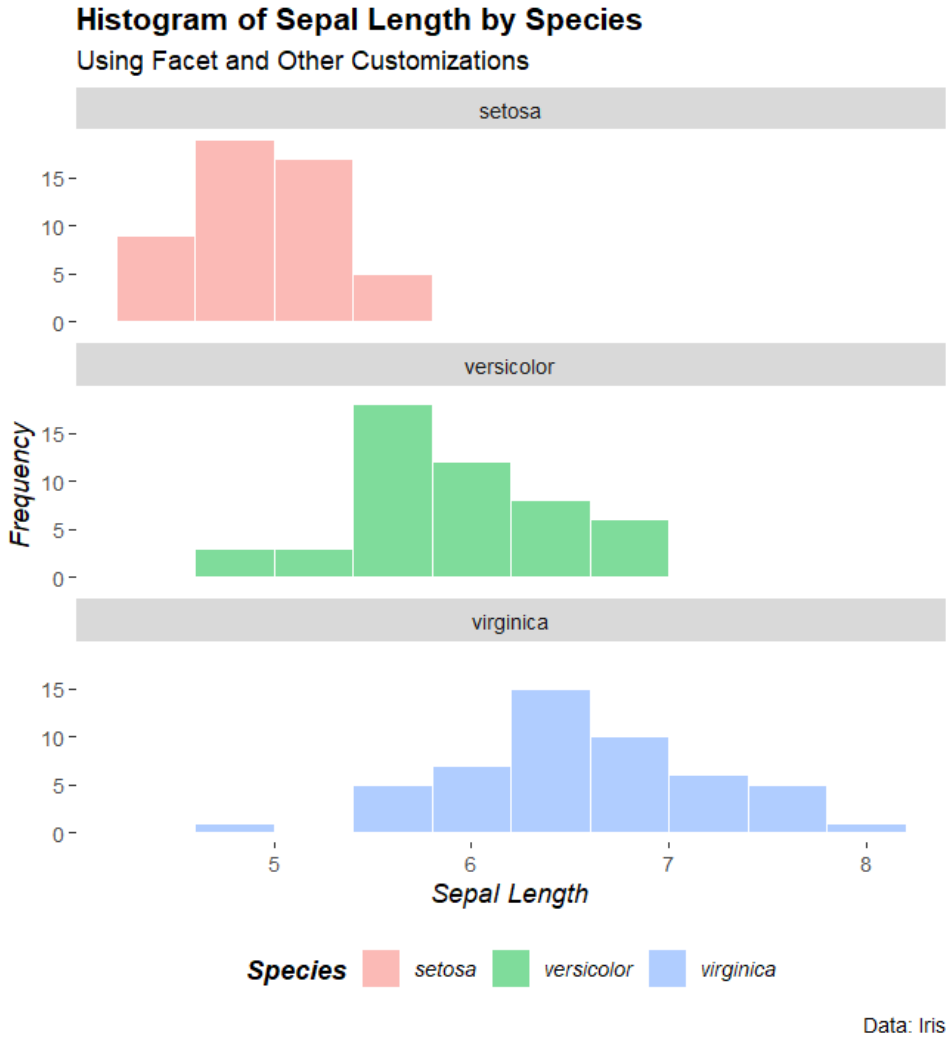
`theme_excel()` `theme_economist()`
`theme_fivethirtyeight()` `theme_tufte()`
`theme_gdocs()` `theme_wsj()` `theme_calc()`
`theme_hc()`

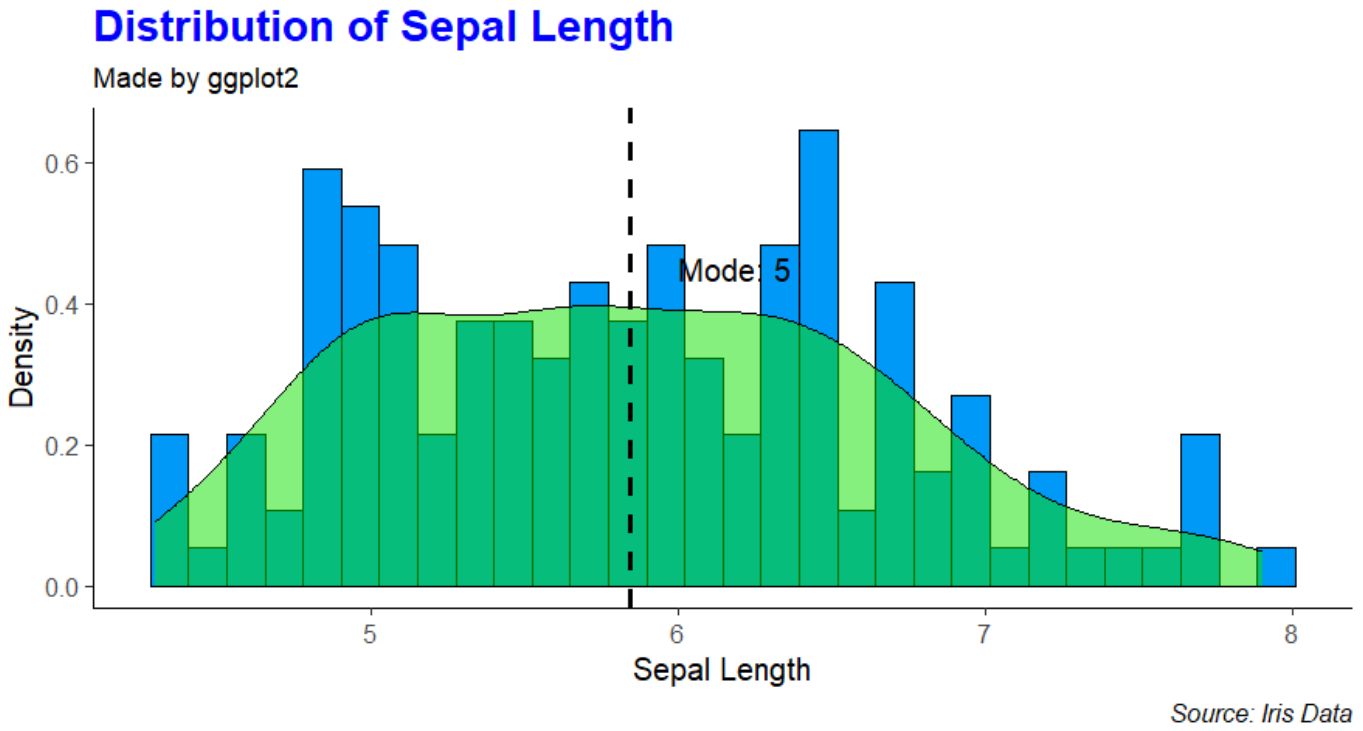
→ other

`theme_article()` `theme_pubclean()`
`theme_bigstatsr()` `theme_ipsum()`

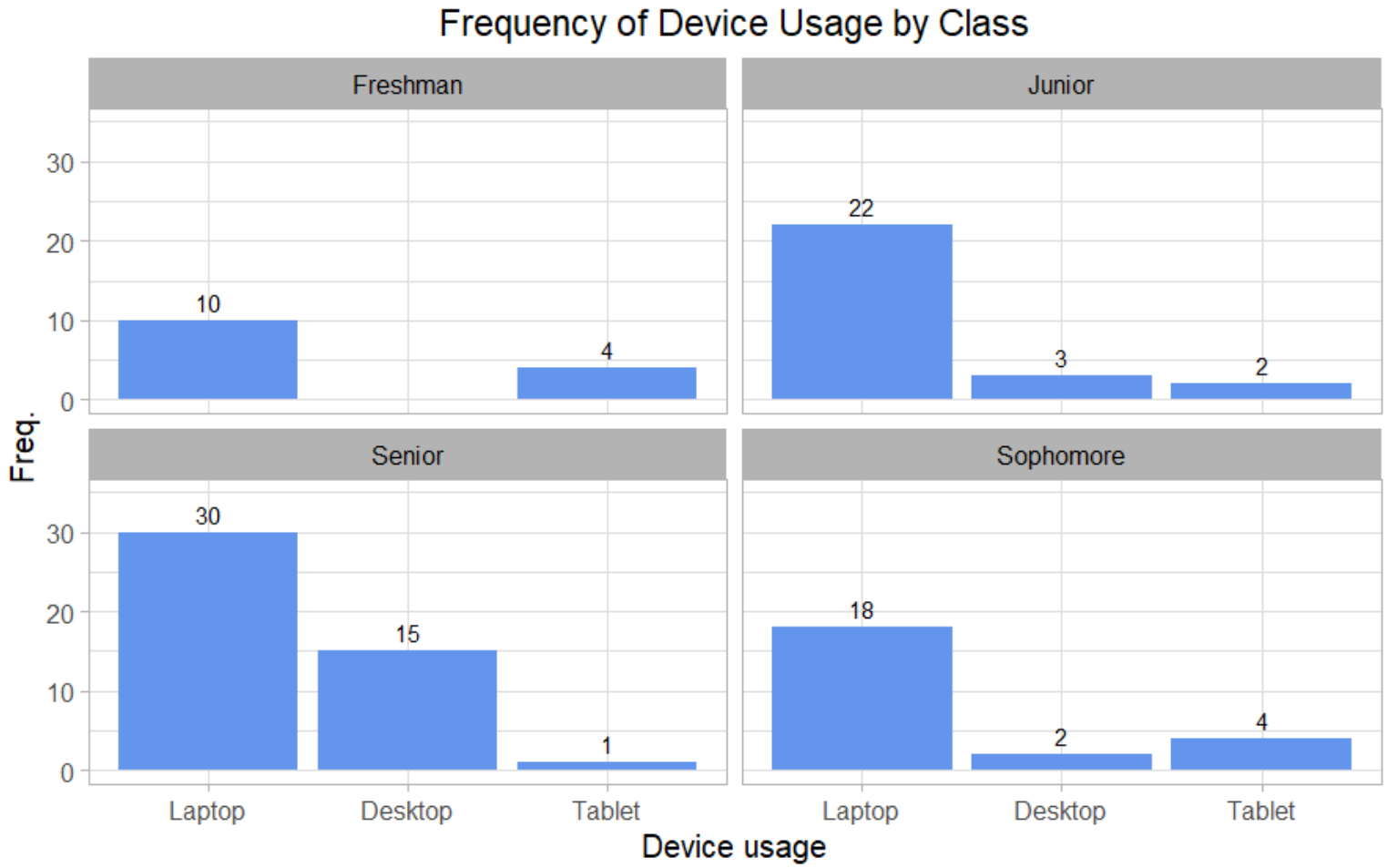


Histogram and Density Plot

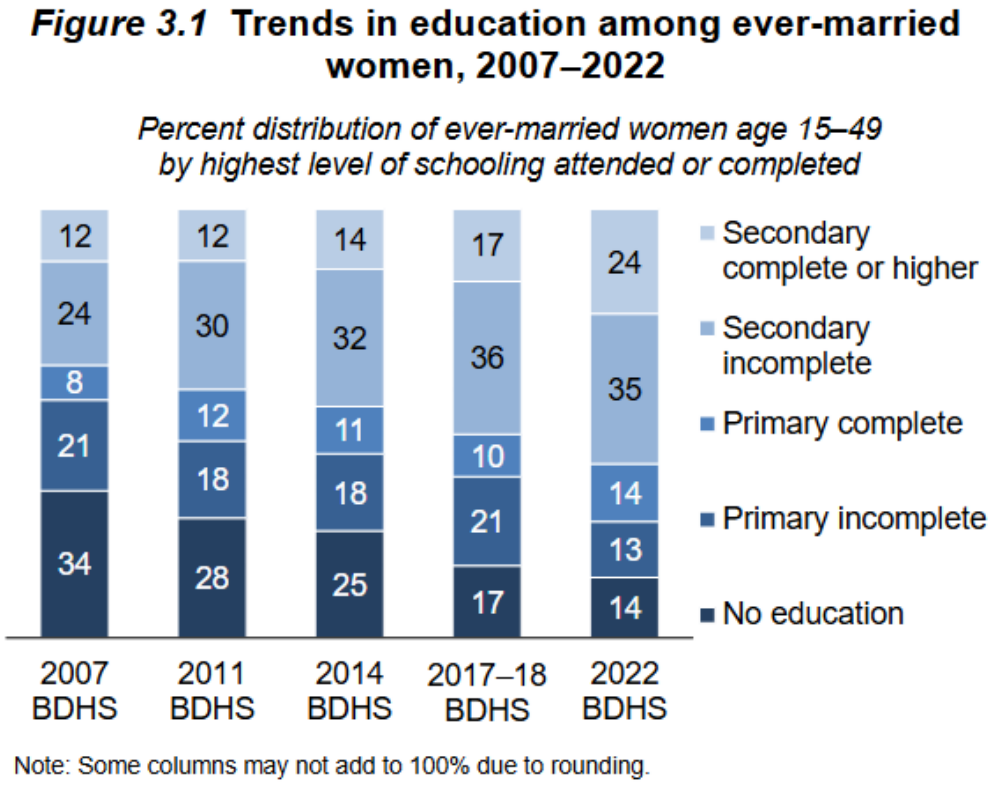


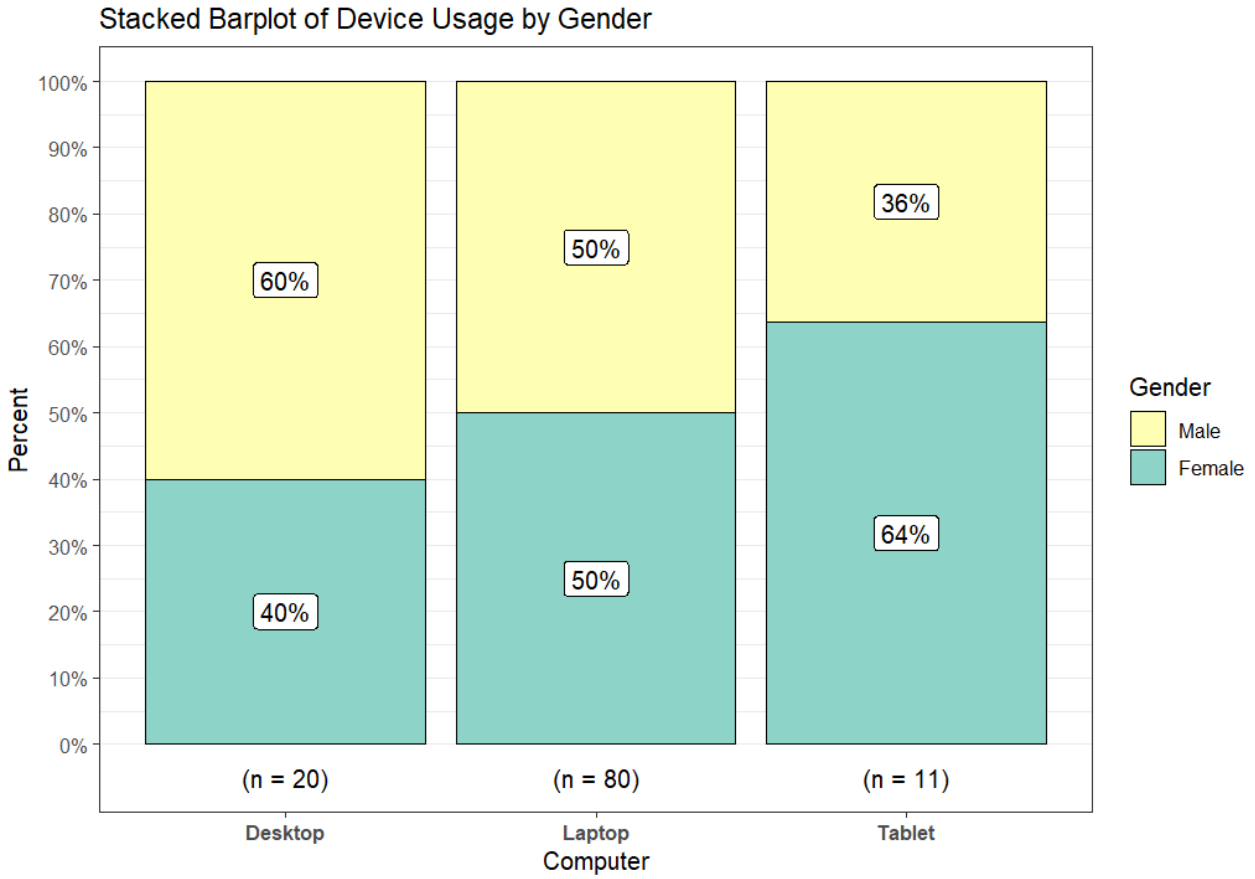


Bar Plot / Column Plot

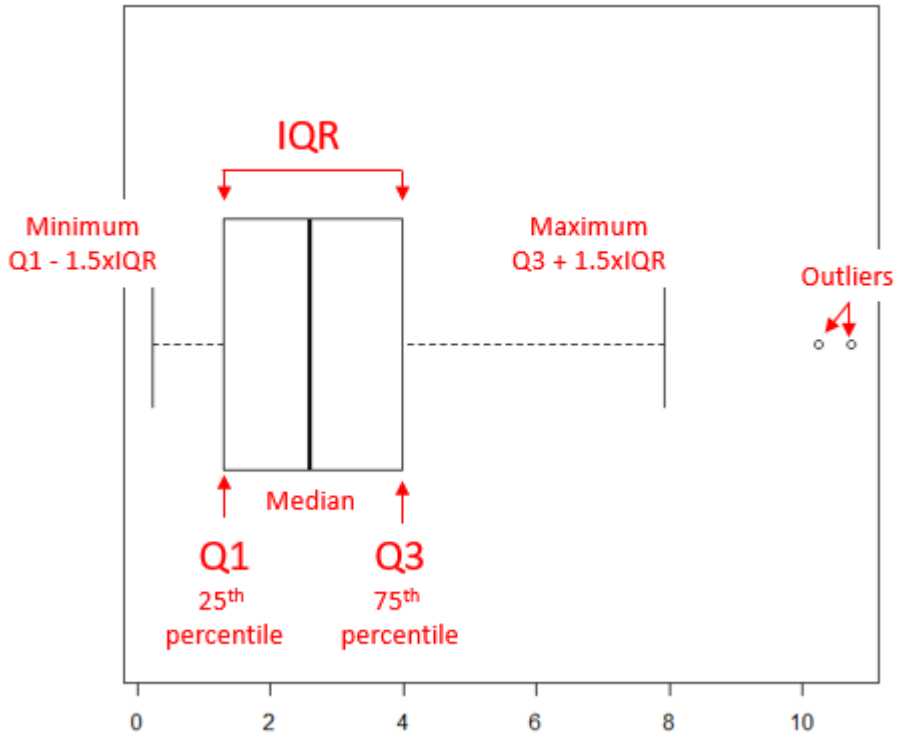


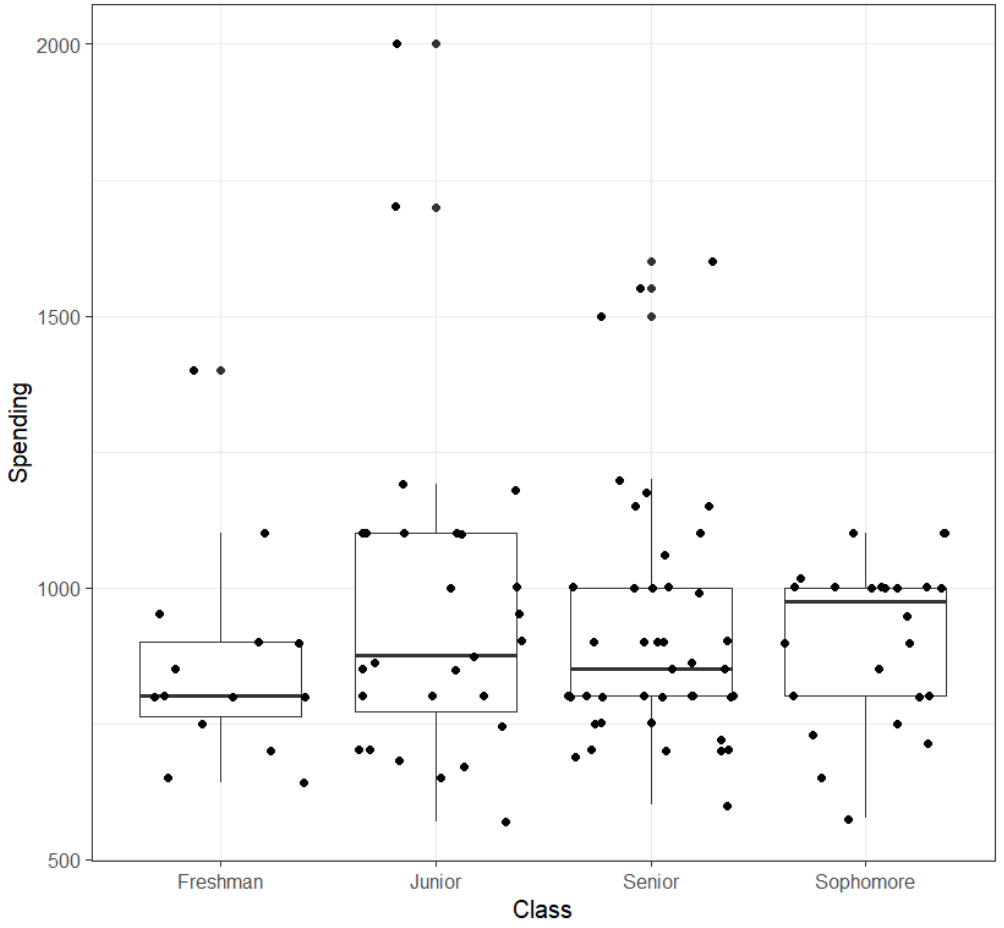
Stacked Bar Plot





Box Plot


IQR
Minimum
Q1 - 1.5xIQR
Maximum
Q3 + 1.5xIQR
Outliers
Median
Q1
25th
percentile
Q3
75th
percentile
0
2
4
6
8
10


2000
1500
1000
500
Spending
Freshman
Junior
Senior
Sophomore
Class

Line Type

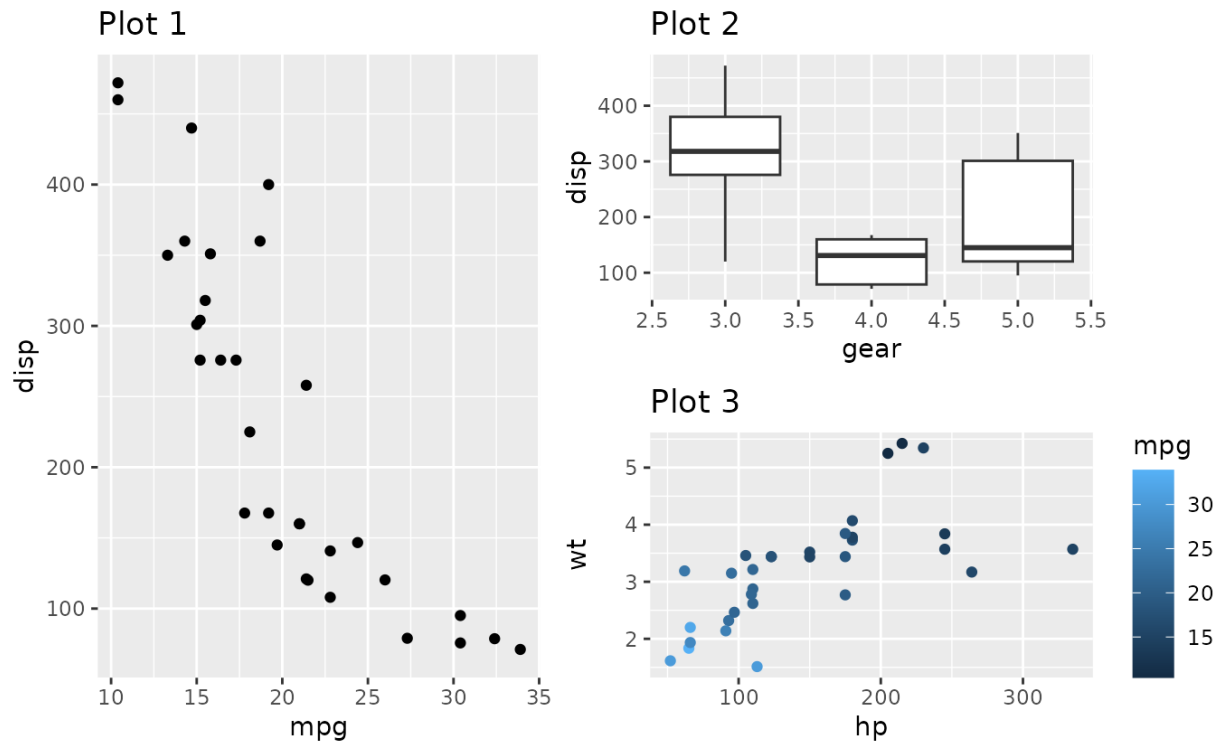

solid
dashed
dotted
dotdash
longdash
twodash
11
18
1f
81
88
8f
f1
f8
ff

Theme

- `ggThemeAssist` add-ins for easy experimentation with different themes

Combining Multiple Plots

- patchwork
- gridExtra::grid.arrange()



Saving plots

- ggsave() function (You may require svglite package to save into svg file)
- sjPlot::save_plot()