Data visualization with ggplot2

CAROLINA M. SINISCALCHI

DATA SCIENCE COORDINATOR

The Grammar of Graphics

– A way of thinking about how graphs are constructed

– In language grammar, sentences have different structures or components that are layered together
  – Nouns, verbs, pronouns, conjunctions, propositions, etc

– This is the same idea behind the grammar of graphs, a series of structures that together form the graph

– This creates a highly customizable and powerful system

The Grammar of Graphics

– We will use the grammar of graphs implemented in the package ggplot2

– This package is part of a group of packages known as tidyverse

The Grammar of Graphics

–   Ggplot2 graphs have 3 required layers (or structures) for any type of plot

    –   **Data**: a data frame (table) that contain the observations and variables we want to plot
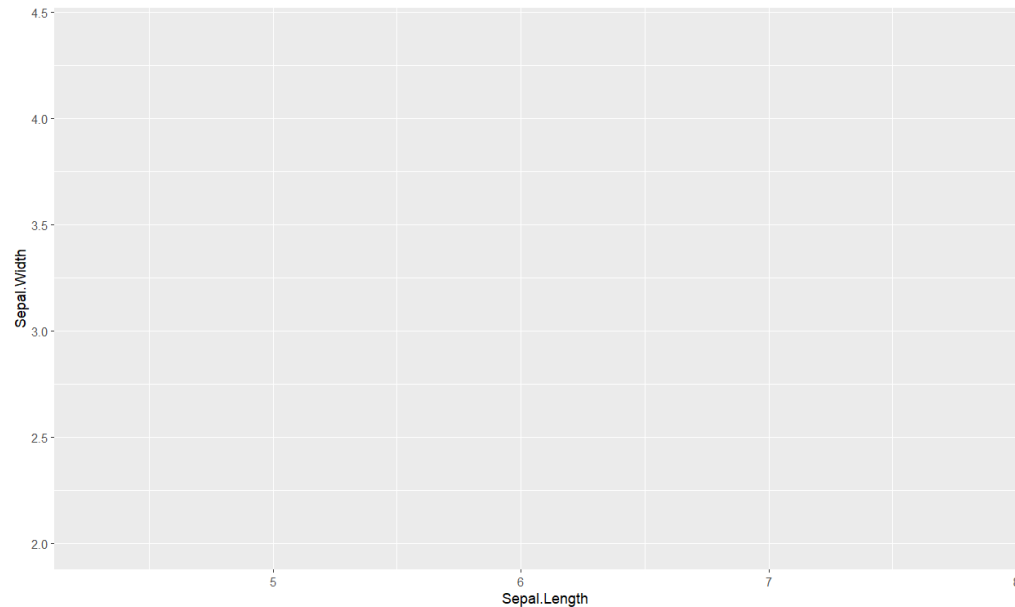
	A	B	C	D	E	F
1	ID	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
2	1	5.1	3.5	1.4	0.2	setosa
3	2	4.9	3	1.4	0.2	setosa
4	3	4.7	3.2	1.3	0.2	setosa
5	4	4.6	3.1	1.5	0.2	setosa
6	5	5	3.6	1.4	0.2	setosa
7	6	5.4	3.9	1.7	0.4	setosa
8	7	4.6	3.4	1.4	0.3	setosa
9	8	5	3.4	1.5	0.2	setosa
10	9	4.4	2.9	1.4	0.2	setosa
11	10	4.9	3.1	1.5	0.1	setosa
12	11	5.4	3.7	1.5	0.2	setosa

ggplot(iris,

The Grammar of Graphics

– Ggplot2 graphs have 3 required layers (or structures) for any type of plot

  – Data: a data frame (table) that contain the observations and variables we want to plot

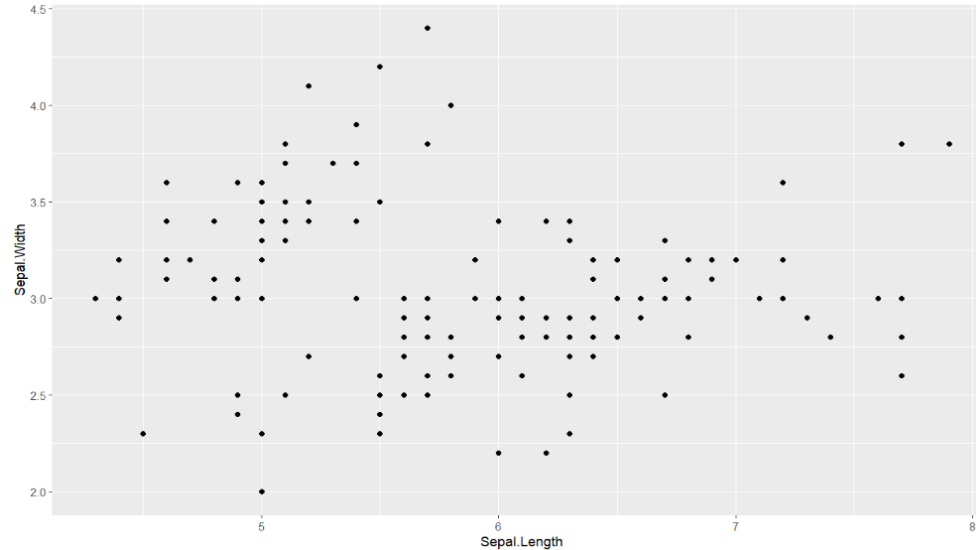 – **Aesthetic:** the mapping of our variables of interest (the x and y axis, for example)

ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width))

The Grammar of Graphics

– Ggplot2 graphs have 3 required layers (or structures) for any type of plot

  – Data: a data frame (table) that contain the observations and variables we want to plot

  – Aesthetic: the mapping of our variables of interest (the x and y axis, for example)

  – **Geometry:** the type or shape of the visual elements of the graph (for example, bars or points)

ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width)) +
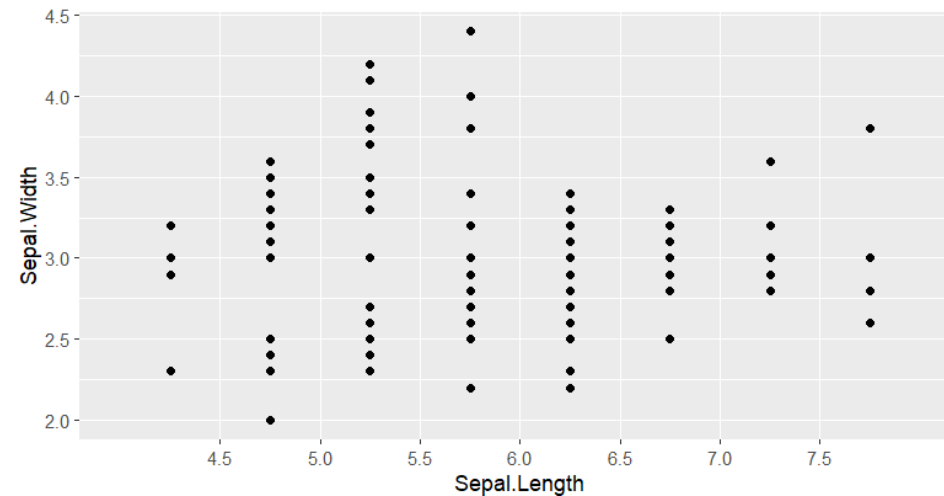geom_point( )

The Grammar of Graphics

- Ggplot2 graphs have 3 required layers (or structures) for any type of plot

  - Data: a data frame (table) that contain the observations and variables we want to plot
  - Aesthetic: the mapping of our variables of interest (the x and y axis, for example)
  - Geometry: the type or shape of the visual elements of the graph (for example, bars or points)

ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width))+
geom_point( )

The Grammar of Graphics

– There are other elements that can be added to make modifications to how data or other elements appear in the plot

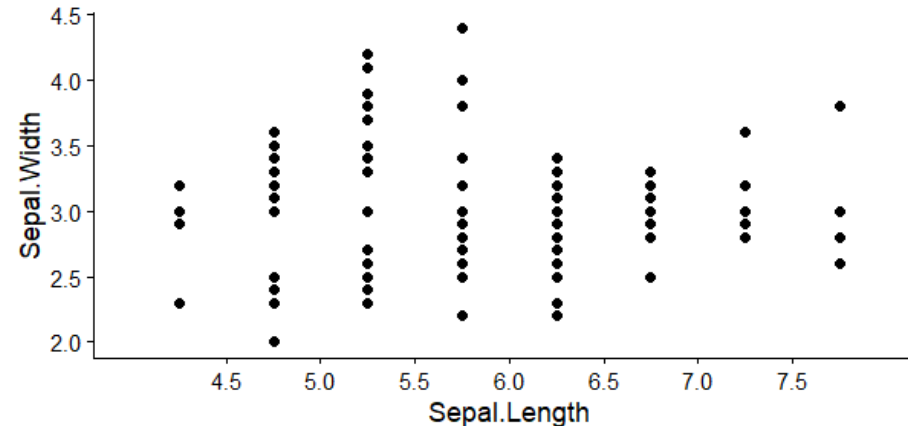 – Scales: can be used to adjust axis coordinates, add color to geometry structures, etc.

ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width))+
geom_point( ) +
scale_x_binned( )

The Grammar of Graphics

– There are other elements that can be added to make modifications to how data or other elements appear in the plot

- Scales: can be used to adjust axis coordinates, add color to geometry structures, etc.
- Themes: change and control non-data display elements

ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width))+
        geom_point( ) +
            scale_x_binned( ) +
            theme_classic( )

Workshop

– Here we will use RStudio for interfacing with R

– We will use the "iris" dataset included in the ggplot package. It has four **continuous** variables and one **categorical** variable.

– We will see how to make scatter plots, histograms and box plots.

– We will also learn how to change different aspects of our plots.

– Link for the tutorial: https://github.com/carol-siniscalchi/public_tutorials/blob/main/ggplot_dataviz/20250923_data_viz.R