Creating data visualizations using ggplot

Introduction to Data Analysis and Visualization

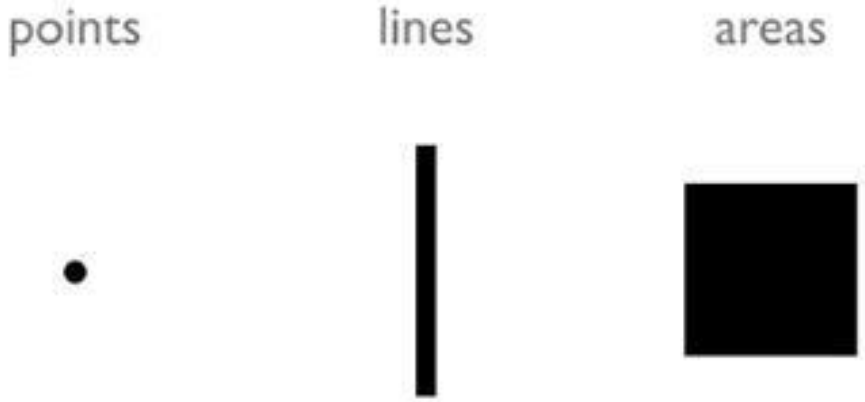
Jan 11, 2022

Lesson Objectives

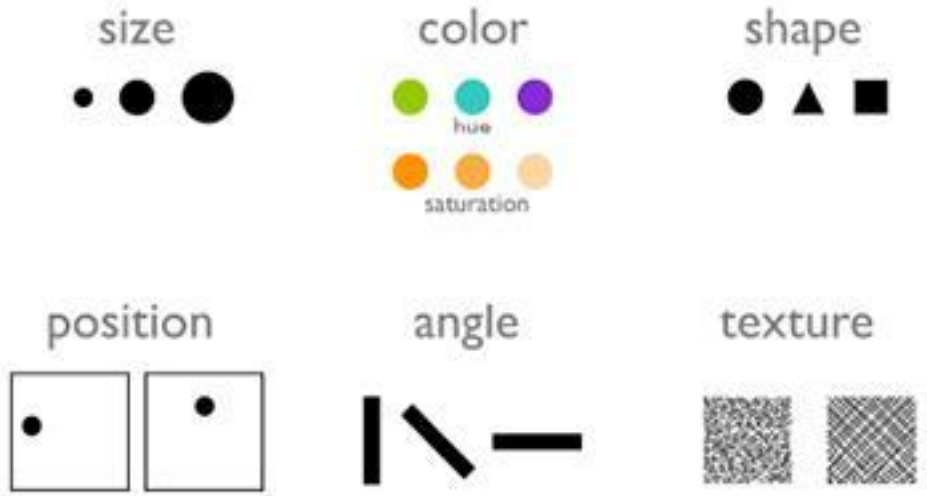
- **By the end of this lesson students should be able to:**
  - Produce a data visualization if given specified geometric primitives and visual channels using the R package ggplot2
  - Locate resources for troubleshooting the syntax of the ggplot2 package

Review of data encodings

- **Geometric primitives** are the simplest graphical markings
- 1D data visualizations use



- **Visual channels** are attributes that describe the appearance of graphical markings



ggplot2: an R package

ggplot2 is based on the *grammar of graphics*, the idea that you can build every graph from the same components: a data set, a coordinate system, and **geoms**—visual marks that represent data points.

To display values, map variables in the data to visual properties of the geom (**aesthetics**) like size, color, and x and y locations.

ggplot2: an R package

ggplot2 is based on the *grammar of graphics*, the idea that you can build every graph from the same components: a data set, a coordinate system, and **geoms**—visual marks that represent data points.

geometric primitives

To display values, map variables in the data to visual properties of the geom (**aesthetics**) like size, color, and x and y locations.

visual channels

ggplot: an R function

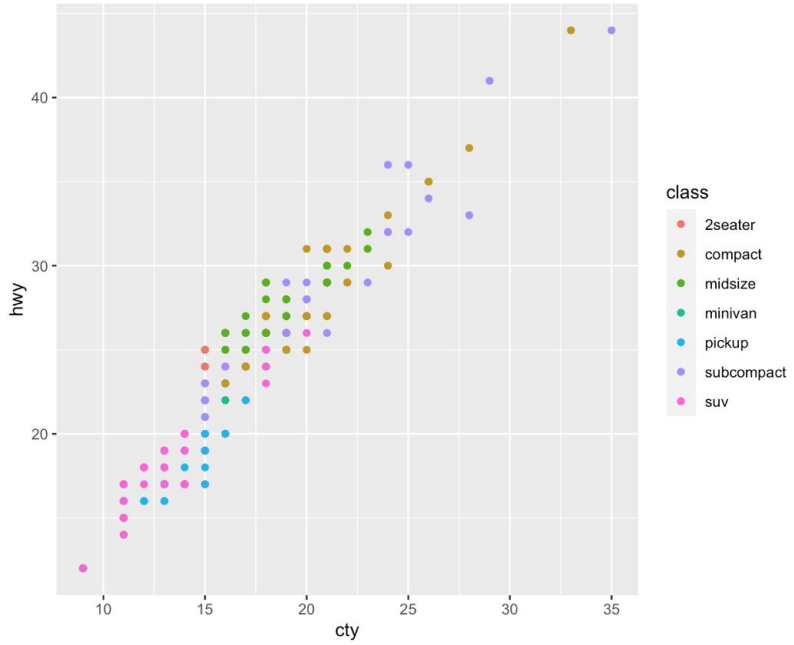
```
ggplot(data = <DATA>) +
<GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

geometric primitives

visual channels

Example

cty	hwy	class
18	29	compact
20	25	compact
15	25	2seater
12	10	suv



- *Data:* `mpg` dataframe
  - miles per gallon in the city and miles per gallon on the highway for cars of different classes

- *Geometric primitive:* points
- *Visual channels:* x- and y- positions and color (hue)

```
> library(ggplot2)
> ggplot(data = mpg) +
geom_point(mapping = aes(x = cty, y = hwy, color = class))
```

Active Learning Exercise

Instructions:

1. Skim the "Data visualization with ggplot2 : : CHEAT SHEET”
    1. https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf
2. In groups, discuss
    1. your understanding of how the ggplot function is structured and
    2. which sections or code segments in the cheatsheet seem most useful in general
3. Create concept maps to better understand how to utilize the cheat sheet