

Scatter Plots for Exploratory Analysis: Takeaways



by Dataquest Labs, Inc. - All rights reserved © 2020

Syntax

- Creating a scatter plot using ggplot2:

```
ggplot(data = Example_Data,  
       aes(x = Variable_1, y = Variable_2)) +  
  geom_point()
```

- Specifying axes ranges:

```
ggplot(data = reviews_2,  
       aes(x = Fandango, y = Rotten_Tomatoes)) +  
  xlim(1,5) +  
  ylim(1,5)
```

- Changing the transparency of the points:

```
ggplot(data = reviews_2,  
       aes(x = Fandango, y = Rotten_Tomatoes)) +  
  geom_point(alpha = 0.3) +  
  xlim(1,5) +  
  ylim(1,5)
```

- Altering the plot background:

```
ggplot(data = reviews_2,  
       aes(x = Fandango, y = Rotten_Tomatoes)) +  
  geom_point(alpha = 0.3) +  
  xlim(1,5) +  
  ylim(1,5) +  
  theme(panel.background = element_rect(fill = "white"))
```

- Using a functional to create multiple scatter plots:

```
create_scatter = function(x, y) {  
  ggplot(data = reviews_2) +  
    aes_string(x = x, y = y) +  
    geom_point(alpha = 0.3) +  
    xlim(1,5) +  
    ylim(1,5) +  
    theme(panel.background = element_rect(fill = "white"))  
}  
  
x_var <- names(reviews_2)[5]  
y_var <- names(reviews_2)[2:4]  
  
ratings_comparisons <- map2(x_var, y_var, create_scatter)
```

Concepts

- Scatter plots represent data using points that have the value of one variable determining the position on the x-axis and the value of the other variable determining the position on the y-axis.
- Types of relationships between variables:
 - Variables are said to have a positive relationship if both variables increase in value while the other one increases in value.
 - Variables are said to have a negative relationship if one variable increases in value while the other one decreases in value.
 - Variables are said to have a strong relationship if points are clustered together as opposed to being spread out.
 - Variables are said to have a weak relationship if points are spread out as opposed to being clustered together.
 - Variables are said to have no relationship if it appears the points are arranged in a shapeless cloud.

- Using `aes_string()` allows you to pass vectors of variable names into your function for efficiency.
- `gridExtra` is a package useful for organizing plots created using `ggplot2` in a grid on a single page.

Resources

- [Information page for gridExtra](#)
- [Scatter plots quick start guide](#)



Takeaways by Dataquest Labs, Inc. - All rights reserved © 2020