Getting started w/ggplot2: an excellent data visualization package

William Christiansen, Ph.D. - Mount St. Mary's University

March 4, 2024

Contents

1 Introduction **2**

2 Getting Started with ggplot2 **2**
2.1 Install 2
2.2 Try This First 2

3 Data Transformation **2**
3.1 Transforming Variables 2
3.2 Working with the QoG Dataset 2

4 Advanced ggplot2 **3**
4.1 Faceting 3
4.2 Themes 3
4.3 Specifying Confidence Intervals 3
4.4 Creating Fitted Lines with Linear Models 3

5 Boxplots and Their Interpretation **4**

6 Conclusion **4**

7 References **4**

1 Introduction

The ggplot2 package in R is a super powerful and adaptable approach to generating an array of publication quality data visualizations. This document explores the fundamentals of ggplot2, data transformation techniques, and practical examples using the Quality of Governance (QoG) dataset.

2 Getting Started with ggplot2

ggplot2, based on The Grammar of Graphics, allows for building plots iteratively by adding layers of annotations.

2.1 Install

To install ggplot2, run the following command in R:

```
install.packages("ggplot2")
```

2.2 Try This First

Creating a scatter plot with ggplot2 involves specifying a dataset and mapping aesthetic properties (like x and y axes) to variables.

```
library(ggplot2)
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

3 Data Transformation

Visualization may require data transformation to improve readability or focus on specific aspects of the data.

3.1 Transforming Variables

The QoG dataset, with its diverse range of variables, often requires transformation for meaningful analysis.

```
qog_data <- read.csv("qog.csv")
qog_data$region <- as.factor(qog_data$region)
```

3.2 Working with the QoG Dataset

Plotting a variable from the QoG dataset after transforming it into a factor:

```
ggplot(qog_data, aes(x = region, y = hdi)) +
  geom_point()
```

For detailed documentation on the QoG dataset, refer to the QoG codebook (http://www.qogdata.pol.gu.se/data/).

4 Advanced ggplot2

Beyond basic plots, ggplot2 supports facets, themes, and statistical transformations for more complex visualizations.

4.1 Faceting

Faceting creates multiple plots based on a factor variable.

```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point() +
  facet_wrap(~class, nrow = 2)
```

4.2 Themes

Changing the plot theme affects the visual style of your graph.

```
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point() +
  theme_minimal()
```

4.3 Specifying Confidence Intervals

Confidence intervals can be added to scatter plots to estimate uncertainty around the mean of a variable. In ggplot2, this can be achieved via smooth. The 'method' argument specifies the type of model or method, and 'level' controls the confidence level.

For a 95% confidence interval using a linear model:

```
ggplot(data = qog_data, aes(x = some_independent_variable, y
    = some_dependent_variable)) +
  geom_point() +
  geom_smooth(method = "lm", level = 0.95)
```

This command overlays the scatter plot with a fitted line and shaded confidence intervals representing the uncertainty around the line.

4.4 Creating Fitted Lines with Linear Models

To further emphasize the relationship between two variables, ggplot2 allows the addition of fitted lines using linear models. This is helpful for correlational analysis.

```
ggplot(data = qog_data, aes(x = some_independent_variable, y
    = some_dependent_variable)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```

The 'se' argument, when set to FALSE, removes the shading, focusing on the fitted linear model.

5 Boxplots and Their Interpretation

Boxplots are a great way to visualize the spread of a variable. They provide insights into the central tendency, dispersion, and outliers within the dataset.

Creating a simple boxplot in ggplot2:

```
ggplot(data = qog_data, aes(x = factor_variable, y = numeric
    _variable)) +
  geom_boxplot()
```

This boxplot groups data by the factor variable, displaying the distribution of the numeric variable within each group.

6 Conclusion

This guide has introduced ggplot2's core functionalities, from basic plots to advanced visualizations including confidence intervals, fitted lines with linear models, and boxplots. For comprehensive learning, the reader is encouraged to explore the ggplot2 documentation and engage with community forums and tutorials. This package has wonderful documentation and support. There are also tons of other packages that integrate well with ggplot2.

7 References

- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- ggplot2 Documentation.
- Quality of Governance Institute - Data and Codebook.