Sample Plots

Taylor Mackay

Contents

Introduction **3**
Useful Resources 3

Univariate Plots **4**
Density Plot 4
Histograms with Grid Arrange 5

Two-Way Plots **6**
Scatter Plot 6
Line Plot with Outcome Grouped by Factor Variable 7

Introduction

This file contains examples of basic plots created using the `ggplot2` package in `R` and the corresponding code required to create each plot. All examples below require loading `ggplot2`– any other required packages are noted as needed in the included code.

NOTE: The specific style of the plots below is specified by using `theme_bcg` in addition to the other plot options. This calls the code below in order to specify the plot style, font type and size, and center plot titles.

```
# Setting options for plot formatting, including font type + size, and title
# alignment, using `minimal` theme

theme_bcg <- theme_minimal(base_size = 9, base_family = "Palatino") +
  theme(plot.title = element_text(hjust = 0.5))
```

Useful Resources

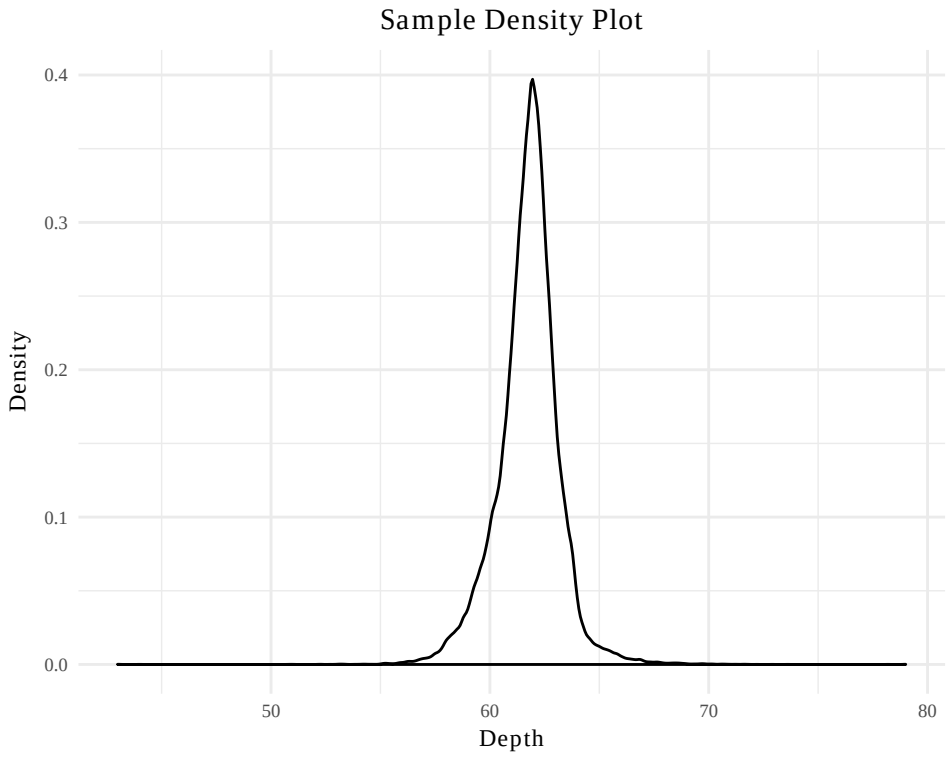
Useful websites with more information on `R` and `ggplot2` (click bulleted items for link to URL).

- RStudio `ggplot2` Cheatsheet
  - Two page PDF cheat sheet covering the basics of the `ggplot2` package
- Gallery of `ggplot2` Examples
  - 50 different examples of plots, covering a range of plot types and customizations to things like legends and annotations
- `R` Datasets Package
  - A list of the sample datasets available with `R` that are used in this document. Includes a detailed description of all variables in each dataset.

Univariate Plots

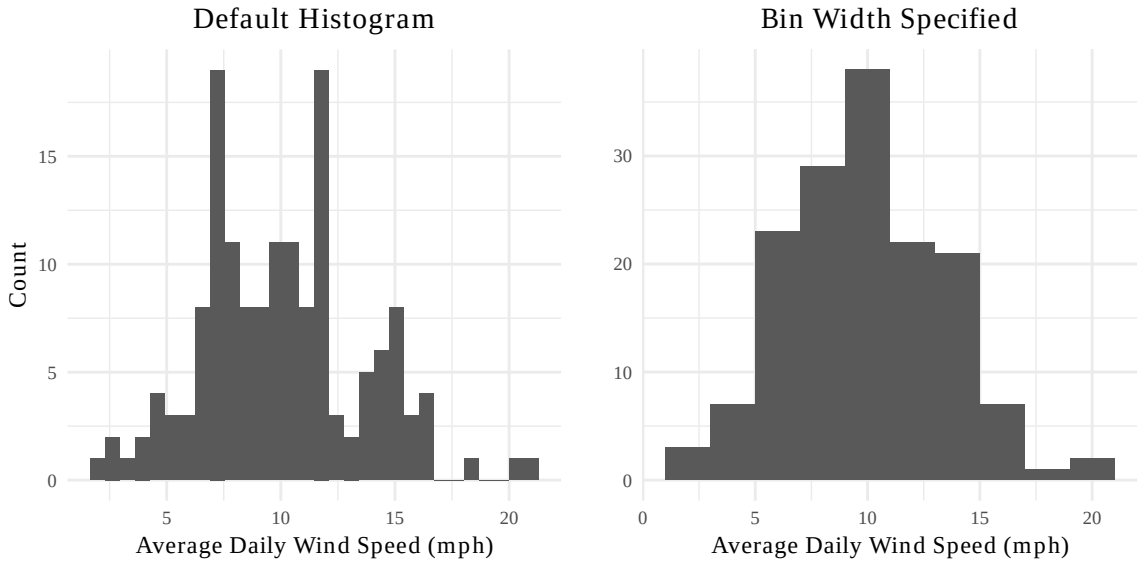
Density Plot

```
# Diamonds samples dataset has prices and other attributes of ~54,000 diamonds.

data(diamonds)

# Generating density plot

ggplot(data = diamonds, aes(x = depth)) + geom_density() +
  labs(title = "Sample Density Plot",
       y = "Density",
       x = "Depth") +
  theme_bcg
```



Histograms with Grid Arrange

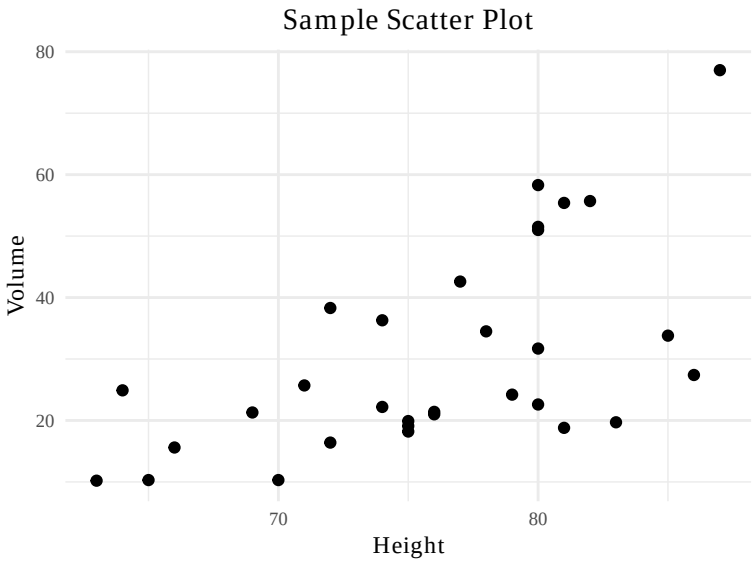
```
# `gridExtra` allows you to print multiple plots together

library(gridExtra)

# Airquality sample dataset has measurements of temperature, windspeed, and
# daily air quality in New York from May to September, 1973.

data("airquality")

# Default Histogram

p.1 <- ggplot(data = airquality) + geom_histogram(aes(x = Wind)) +
  labs(title = "Default Histogram",
       y = "Count",
       x = "Average Daily Wind Speed (mph)") +
  theme_bcg

p.2 <- ggplot(data = airquality) + geom_histogram(aes(x = Wind), binwidth = 2) +
  labs(title = "Bin Width Specified",
       y = "",
       x = "Average Daily Wind Speed (mph)") +
  theme_bcg

# Using `grid.arrange` to print both plots side by side (by setting nrow = 1)

grid.arrange(p.1, p.2, nrow = 1)
```



Two-Way Plots

Scatter Plot

```
# Trees sample dataset has measurements of the height, weight, and length of
# 31 observations.

data(trees)

# Scatter Plot

ggplot(data = trees) + geom_point(aes(x = Height, y = Volume)) +
  labs(title = "Sample Scatter Plot",
       y = "Volume",
       x = "Height") +
  theme_bcg
```



Line Plot with Outcome Grouped by Factor Variable

```
# Orange sample data set has 7 measurements of age and circumference for 5
# different oranges (total of 35 observations)

data(Orange)

# Start by creating a observation count by ID variable using `dplyr`. Note that
# data needs to be in *long* form.

library(dplyr)

df <- group_by(Orange, Tree) %>%
  mutate(count = row_number())

# Creating re-ordered `tree` factor variable

df$Tree <- factor(df$Tree, levels = c(1,2,3,4,5))

# Line Plot-- notice options for setting x-axis ticks + legend label

ggplot(data = df) + geom_line(aes(x = count, y = circumference,
                                  color = Tree)) +
  labs(title = "Sample Line Plot with Factor Groupings",
       y = "Circumference (mm)", x = "Observation",
       color = "Tree") +
  scale_x_continuous(breaks=seq(1, 7, 1)) +
  theme_bcg
```