Data Visualization Analysis

Chen

2024-01-01

0.1 Introduction

This tutorial is designed to help you learn data visualization analysis by providing simple and useful information in a way that is easy to follow and understand.

0.2 Preparation

In order to draw a chart, we need to include the required packages for visualization and dataset. For example, `ggplot2` package is for drawing charts, `gcookbook` is for using `pg_mean` dataset, and `palmerpenguins` provides the `penguins` dataset.

```r
library(ggplot2)
library(gcookbook)
library(dplyr)          # Provides %>% operator
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
library(palmerpenguins) # Provides penguins dataset
```

```
Warning: package 'palmerpenguins' was built under R version 4.4.2
```

0.3 Bar chart

In this section, we will draw a bar chart using `pg_mean` dataset. The dataset has two columns: `group`, `weight`.

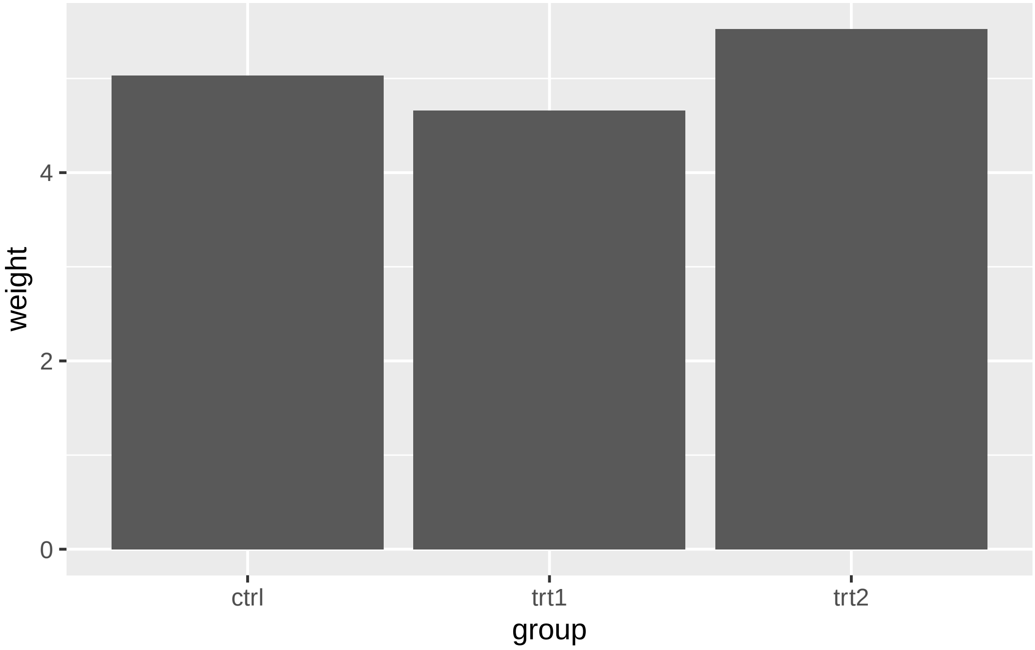
```
pg_mean
```

```
  group weight
1  ctrl  5.032
2  trt1  4.661
3  trt2  5.526
```

This dataset compares the weight across three groups:

- `ctrl`: Control group (baseline, weight = 5.032).
- `trt1`: Treatment 1 group (weight = 4.661).
- `trt2`: Treatment 2 group (weight = 5.526).

```
ggplot(pg_mean, aes(x = group, y = weight)) +
  geom_col()
```



It initializes a ggplot with the dataset `pg_mean`.

`aes(x = group, y = weight)` specifies the aesthetics:

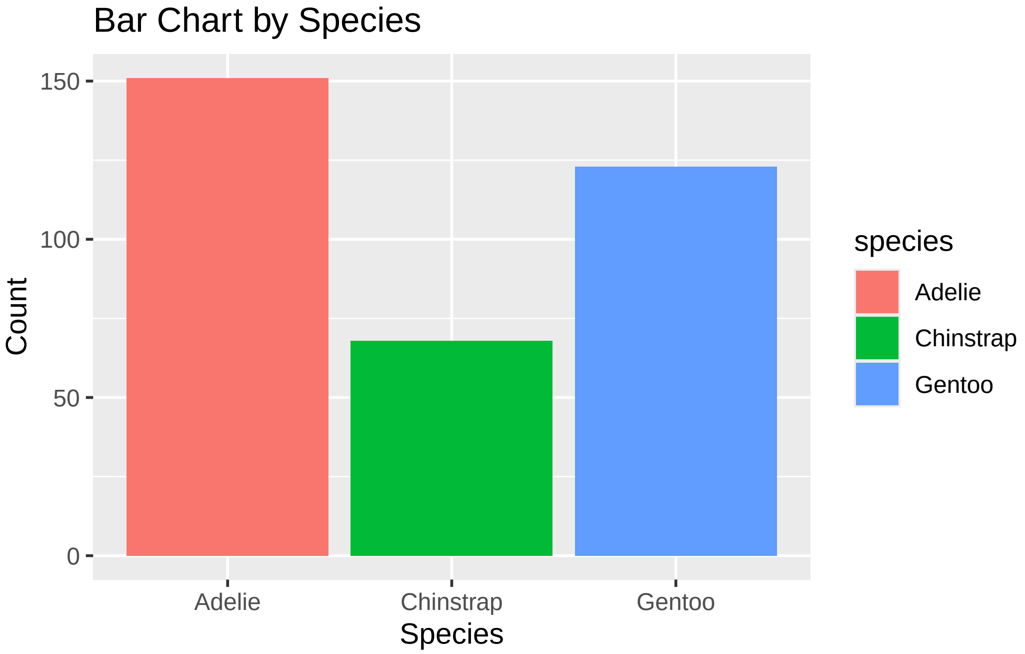
- `x = group`: Assign the `group` variable to the x-axis (categorical data, such as `ctrl`, `trt1`, `trt2`).
- `y = weight`: Assign the `weight` variable to the y-axis (numerical data).

`geom_col()`:

- Adds a column geometry to the plot.
- `geom_col()` creates bars where the height of each bar corresponds to the value of weight for each group.

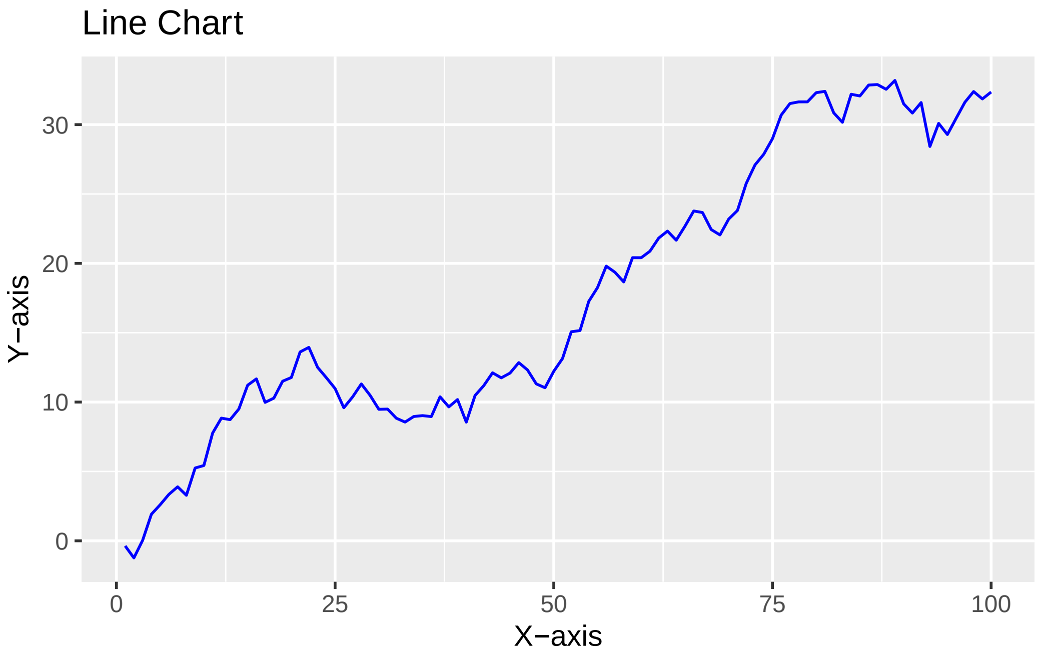
0.4 Bar Chart with Color

```r
# Sample data
data <- penguins %>% filter(!is.na(body_mass_g))

# Bar chart by species
p <- ggplot(data, aes(x = species, fill = species)) +
  geom_bar() +
  labs(title = "Bar Chart by Species", x = "Species", y = "Count")

print(p)
```



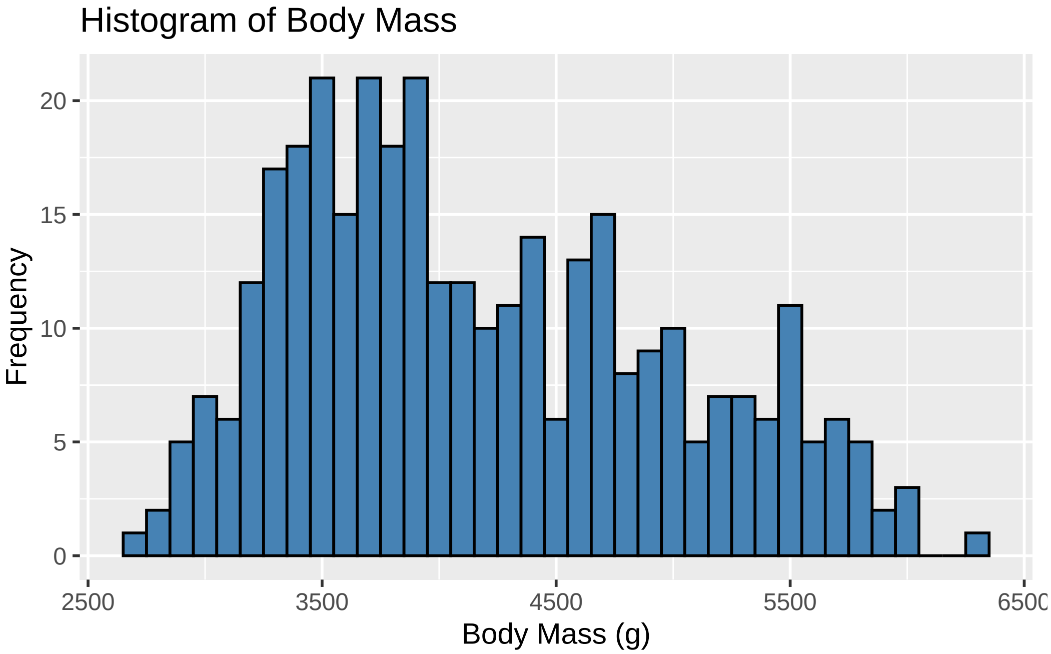
0.5 Line Chart

```
# Sample data
line_data <- tibble(
  x = 1:100,
  y = cumsum(rnorm(100))
)

# Line chart
p <- ggplot(line_data, aes(x = x, y = y)) +
  geom_line(color = "blue") +
  labs(title = "Line Chart", x = "X-axis", y = "Y-axis")

print(p)
```



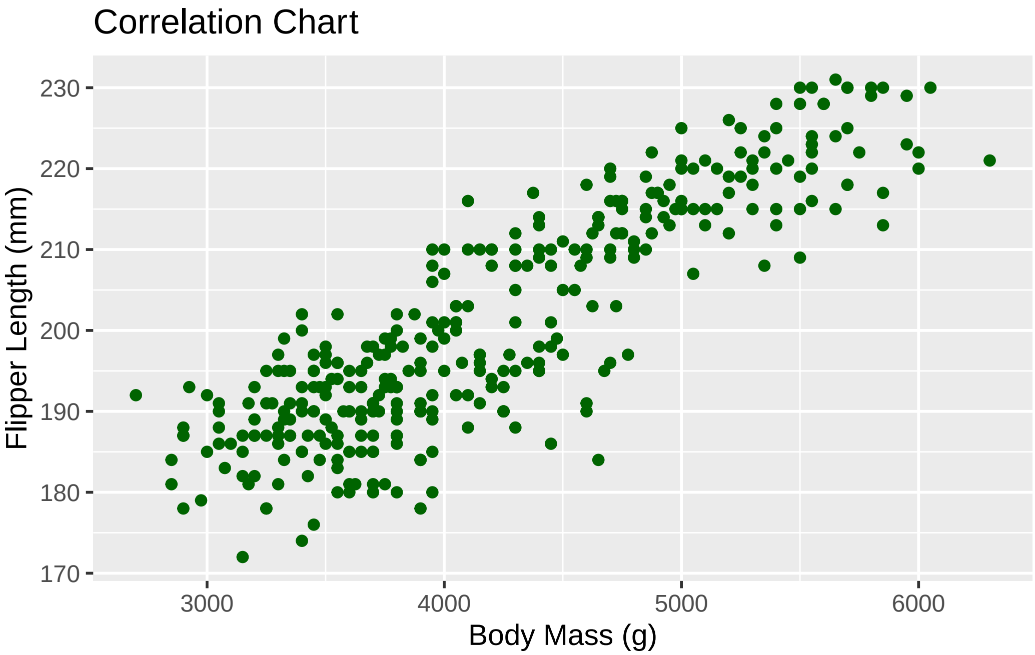
0.6 Histogram

```
# Histogram of penguin body mass
p <- ggplot(data, aes(x = body_mass_g)) +
  geom_histogram(binwidth = 100, fill = "steelblue", color = "black") +
  labs(title = "Histogram of Body Mass", x = "Body Mass (g)", y = "Frequency")

print(p)
```



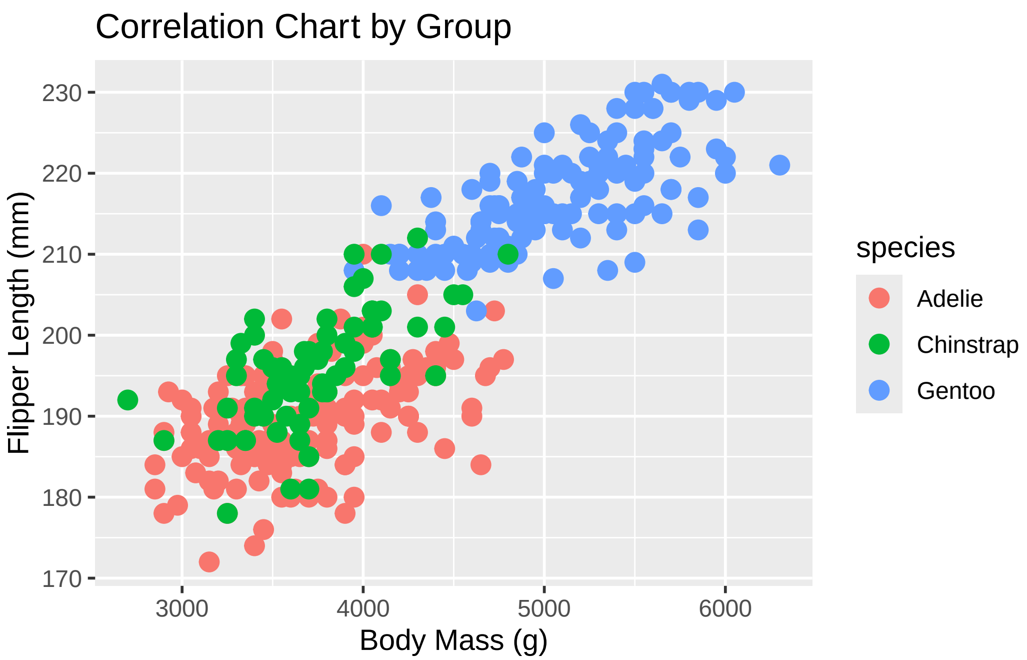
0.7 Correlation Chart

```
# Scatter plot of body mass vs. flipper length
p <- ggplot(data, aes(x = body_mass_g, y = flipper_length_mm)) +
  geom_point(color = "darkgreen") +
  labs(title = "Correlation Chart", x = "Body Mass (g)", y = "Flipper Length (mm)")

print(p)
```



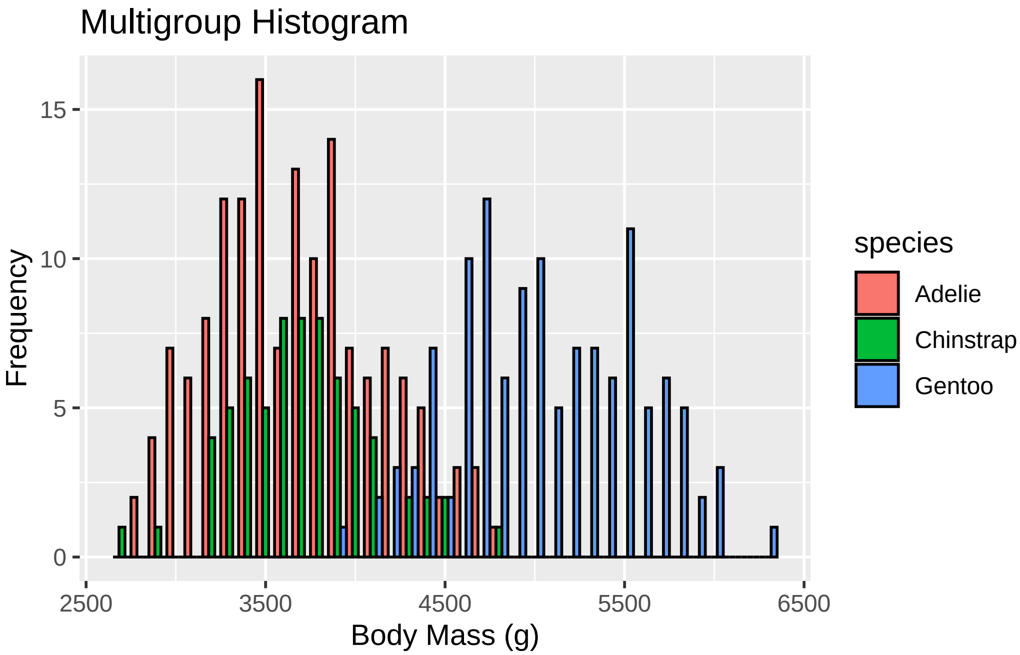
0.8 Correlation Chart: Color by Group

```
# Scatter plot with color by species
p <- ggplot(data, aes(x = body_mass_g, y = flipper_length_mm, color = species)) +
  geom_point(size = 3) +
  labs(title = "Correlation Chart by Group", x = "Body Mass (g)", y = "Flipper Length (mm)")

print(p)
```



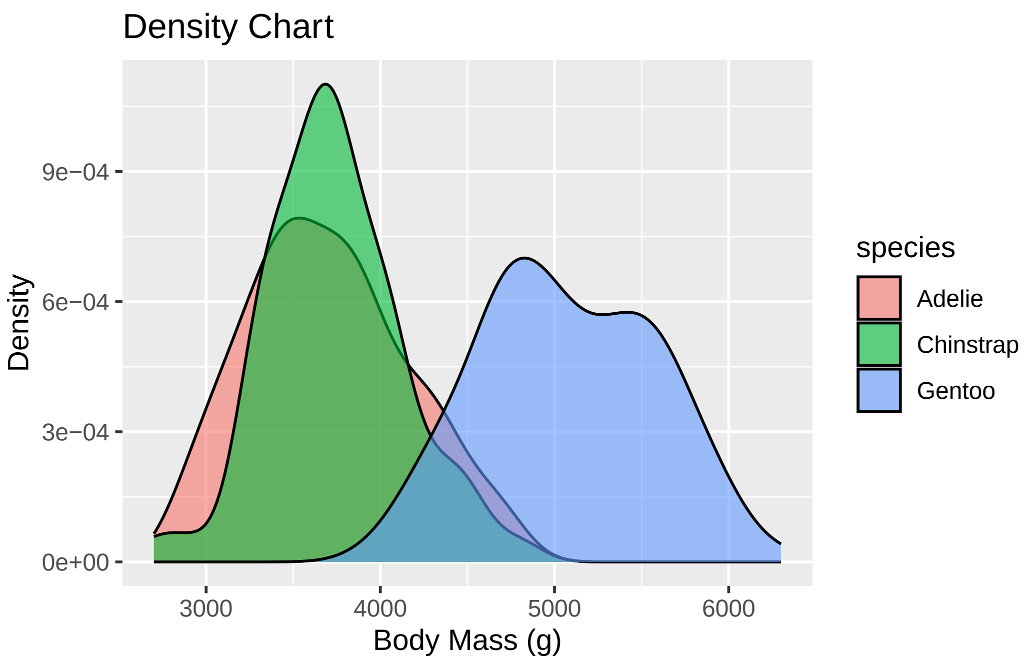
0.9 Multigroup Histogram

```
# Histogram of body mass by species
p <- ggplot(data, aes(x = body_mass_g, fill = species)) +
  geom_histogram(binwidth = 100, position = "dodge", color = "black") +
  labs(title = "Multigroup Histogram", x = "Body Mass (g)", y = "Frequency")

print(p)
```



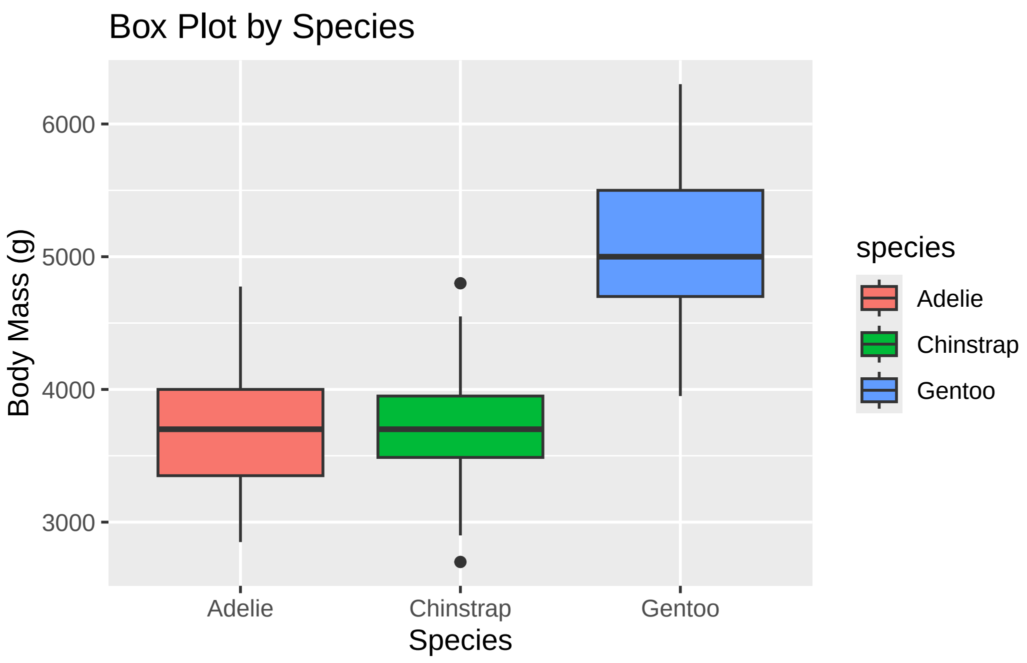
0.10 Density Chart

```r
# Density chart of body mass
p <- ggplot(data, aes(x = body_mass_g, fill = species)) +
  geom_density(alpha = 0.6) +
  labs(title = "Density Chart", x = "Body Mass (g)", y = "Density")

print(p)
```



0.11 Box Plot

```
# Box plot of body mass by species
p <- ggplot(data, aes(x = species, y = body_mass_g, fill = species)) +
  geom_boxplot() +
  labs(title = "Box Plot by Species", x = "Species", y = "Body Mass (g)")

print(p)
```


Box Plot by Species
Body Mass (g)
6000
5000
4000
3000
Adelie
Chinstrap
Gentoo
Species
species
Adelie
Chinstrap
Gentoo