Sample Document

Data Science with R

Welcome to this Sample Document

Solving quadratic equations

Consider equations of the form $ax^2 + bx + c = 0$. This is a quadratic equation; changing the values of a, b, and c, changes the shape of the quadratic. There are two solutions to a quadratic equation:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Let's draw a quadratic curve for different values of $a, b,$ and c.

```r
a <- 10
b <- 3
c <- -5
x <- seq(-2, 2, length = 1e3)
fx <- a*x^2 + b*x + c
plot(x, fx, type = 'l', xlab = "x", ylab = "Quadratic function")
# adding line for solution
abline(h = 0, lty = 2, col = "red")
```
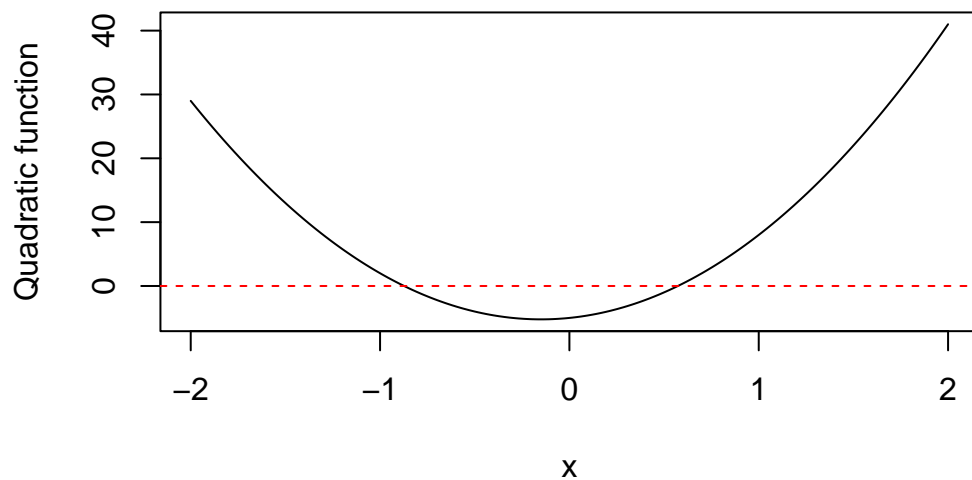
In the above, we can also change the way the plot appears. I will rerun the above code, without displaying the code, and only producing the plot of figure height "4" units and figure width "4" units
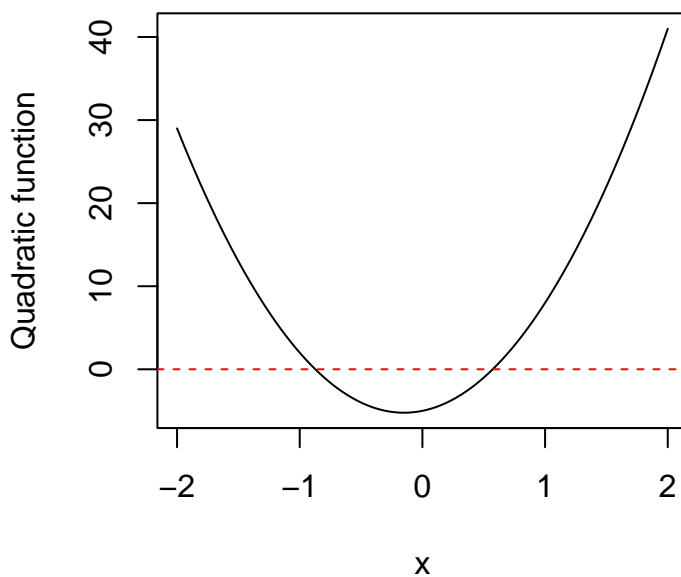

Figure 1: A quadratic function

Studying the Cricket dataset

We will study the cricket dataset more closely now. The `battingbowling.csv` file in your repository contains information regarding various potential all-rounders in men's ODI cricket. The players' batting average, bowling average, and team is present in the dataset.

We will make a few visualizations. But first, we will need a few libraries.

```r
library(ggplot2)
library(tibble)
```

Let's load the data the data and convert it to a tibble.

```r
cric <- read.csv("battingbowling.csv")
cric <- as_tibble(cric)
cric
```

```
# A tibble: 78 x 4
   X               Bowling Batting Team
   <chr>             <dbl>   <dbl> <chr>
 1 Khalil Ahmed       31      4.5  India
 2 Jasprit Bumrah     24.4   13.8  India
 3 Y Chahal           28.8    8.16 India
 4 R Jadeja           36.6   31.9  India
 5 K Jadhav           37.8   42.1  India
 6 V Kohli            66.2   59.3  India
 7 K Yadav            26.2   13.1  India
 8 B Kumar            34.6   14.2  India
 9 M Shami            25.4    3.73 India
10 H Pandya           40.6   29.9  India
# ... with 68 more rows
```
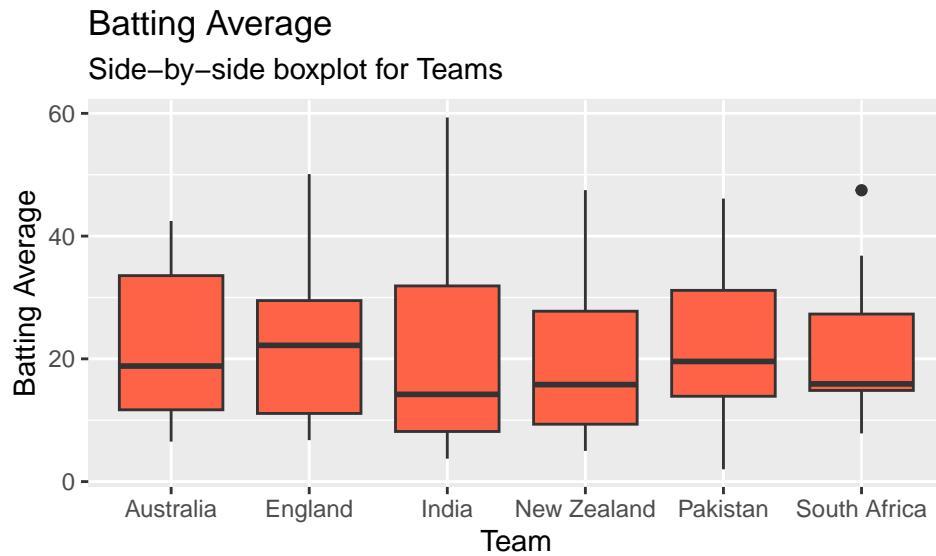
Let's create a side-by-side boxplot of the Batting averages:

```r
p <- ggplot(cric, aes(x = Team,y = Batting))
p + geom_boxplot(fill = "tomato")+
  labs(title = "Batting Average",
       subtitle = "Side-by-side boxplot for Teams",
       y = "Batting Average")
```
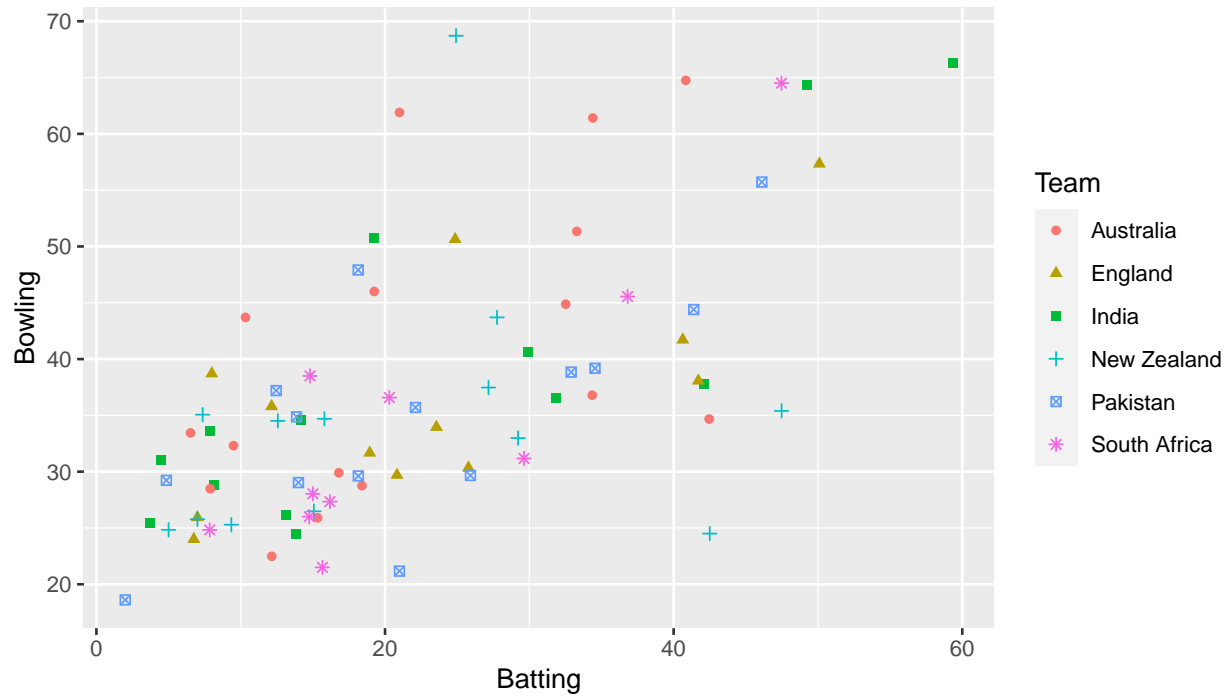
Batting Average
Side–by–side boxplot for Teams

We can do the same thing for the Bowling average as well (notice that I am not showing the code for this)

## Bowling Average
Side–by–side boxplot for Teams
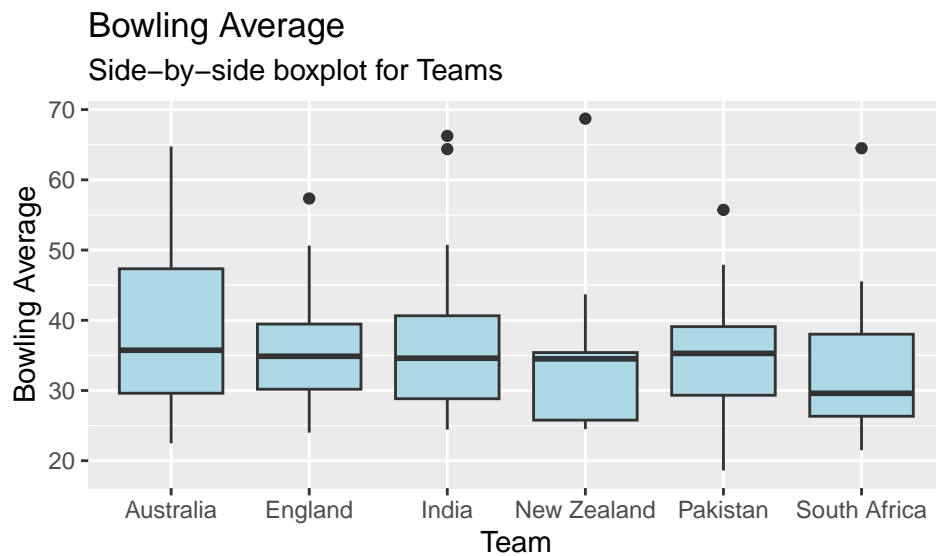
Next, we make a scatterplot of the Batting and Bowling average, colored by the Team the player is associated with:

```
p <- ggplot(cric, aes(x = Batting, y = Bowling))
p + geom_point(aes(color = Team, shape = Team))
```

Additionally, we want to write down, in text, the names of the players with batting average more than 35 and bowling average less than 40. Below is a plot for that.
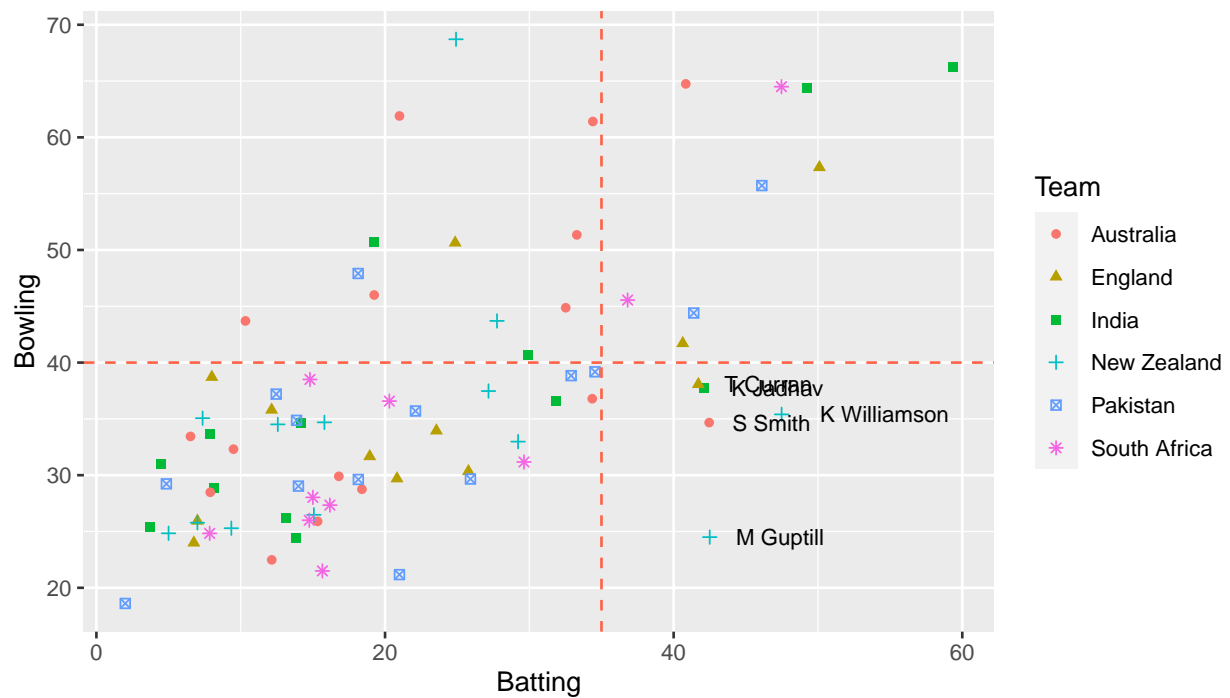
Some more mathematical equations

Let's practice some mathematical typing.

1. Consider a rotation matrix:

$$A = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

Our goal is to find the determinant of the above

$$\det(A) = \cos(\theta)\cos(\theta) - \sin(\theta)(-\sin(\theta))$$
$$= \sin^2(\theta) + \cos^2(\theta)$$
$$= 1.$$

2. Consider the series $A = \sum_{i=1}^{n} i$:

$$2A = \sum_{i=1}^{n} i + \sum_{i=1}^{n} i$$
$$= (1 + 2 + \cdots + n) + (1 + 2 + \cdots + n)$$
$$= (1 + n) + (2 + n - 1) + \ldots (n + 1)$$
$$= n \cdot (n + 1)$$
$$\Rightarrow A = \frac{n(n + 1)}{2}.$$