

# Assignment\_2\_DT

aditya venugopalan a1899824

2023-10-03

```
ashes_tidy<-gather(ashes, key="test_innings", value="details", 5:14)
ashes_tidy <- tibble(ashes_tidy)
ashes_tidy
```

```
## # A tibble: 260 x 6
##       X batter      team    role    test_innings  details
##   <int> <chr>      <chr>   <chr>   <chr>         <chr>
## 1     1 Anderson   England bowl    Test.1..Innings.1 Batting at number~
## 2     2 Bell      England batsman Test.1..Innings.1 Batting at number~
## 3     3 Clark     Australia bowler  Test.1..Innings.1 Batting at number~
## 4     4 Clarke    Australia batter  Test.1..Innings.1 Batting at number~
## 5     5 Collingwood England bat      Test.1..Innings.1 Batting at number~
## 6     6 Cook      England batter  Test.1..Innings.1 Batting at number~
## 7     7 Flintoff   England all rounder Test.1..Innings.1 Batting at number~
## 8     8 Gilchrist  Australia wicketkeeper Test.1..Innings.1 Batting at number~
## 9     9 Giles     England bowl    Test.1..Innings.1 Batting at number~
## 10    10 Harmison  England bolw    Test.1..Innings.1 Batting at number~
## # i 250 more rows
```

```
#using str_match()
head(ashes_tidy)
```

```
## # A tibble: 6 x 6
##       X batter      team    role    test_innings  details
##   <int> <chr>      <chr>   <chr>   <chr>         <chr>
## 1     1 Anderson   England bowl    Test.1..Innings.1 Batting at number 11, s~
## 2     2 Bell      England batsman Test.1..Innings.1 Batting at number 3, sc~
## 3     3 Clark     Australia bowler  Test.1..Innings.1 Batting at number 10, s~
## 4     4 Clarke    Australia batter  Test.1..Innings.1 Batting at number 6, sc~
## 5     5 Collingwood England bat      Test.1..Innings.1 Batting at number 4, sc~
## 6     6 Cook      England batter  Test.1..Innings.1 Batting at number 2, sc~
```

```
ashes_tidy <- ashes_tidy %>%
mutate(batting_order=str_match(details, "Batting at number (\\d+), scored")[,2], score=str_match(detail.
ashes_tidy
```

```
## # A tibble: 260 x 9
##       X batter      team    role    test_innings  details  batting_order  score  balls
##   <int> <chr>      <chr>   <chr>   <chr>         <chr>   <chr>         <chr> <chr>
## 1     1 Anderson   Engla~ bowl    Test.1..Inn~ Battin~ 11          2      8
```

```
## 2      2 Bell      Engla~ bats~ Test.1..Inn~ Battin~ 3      50      162
## 3      3 Clark     Austr~ bowl~ Test.1..Inn~ Battin~ 10     39      23
## 4      4 Clarke    Austr~ batt~ Test.1..Inn~ Battin~ 6      56      94
## 5      5 Collingwood Engla~ bat  Test.1..Inn~ Battin~ 4      5      13
## 6      6 Cook      Engla~ batt~ Test.1..Inn~ Battin~ 2      11      15
## 7      7 Flintoff  Engla~ all ~ Test.1..Inn~ Battin~ 6      0      3
## 8      8 Gilchrist Austr~ wick~ Test.1..Inn~ Battin~ 7      0      3
## 9      9 Giles     Engla~ bowl Test.1..Inn~ Battin~ 8      24      39
## 10     10 Harmison Engla~ bolw Test.1..Inn~ Battin~ 10     0      5
## # i 250 more rows
```

```
head(ashes_tidy)
```

```
## # A tibble: 6 x 9
##       X batter      team    role test_innings details batting_order score balls
##   <int> <chr>      <chr>  <chr> <chr>      <chr>  <chr>      <chr> <chr>
## 1      1 Anderson  England bowl Test.1..Inn~ Battin~ 11        2      8
## 2      2 Bell      England bats~ Test.1..Inn~ Battin~ 3        50     162
## 3      3 Clark     Austr~ bowl~ Test.1..Inn~ Battin~ 10       39     23
## 4      4 Clarke    Austr~ batt~ Test.1..Inn~ Battin~ 6        56     94
## 5      5 Collingwood England bat  Test.1..Inn~ Battin~ 4         5     13
## 6      6 Cook      England batt~ Test.1..Inn~ Battin~ 2        11     15
```

```
#question 1(b)
```

```
ashes_tidy$batting_order <- as.integer(ashes_tidy$batting_order)
ashes_tidy$score <- as.integer(ashes_tidy$score)
ashes_tidy$balls <- as.integer(ashes_tidy$balls)
ashes_tidy$role <- as_factor(ashes_tidy$role)
ashes_tidy$team <- as_factor(ashes_tidy$team)
ashes_tidy
```

```
## # A tibble: 260 x 9
##       X batter      team    role test_innings details batting_order score balls
##   <int> <chr>      <fct>  <fct> <chr>      <chr>      <int> <int> <int>
## 1      1 Anderson  Engla~ bowl Test.1..Inn~ Battin~ 11        2      8
## 2      2 Bell      Engla~ bats~ Test.1..Inn~ Battin~ 3        50     162
## 3      3 Clark     Austr~ bowl~ Test.1..Inn~ Battin~ 10       39     23
## 4      4 Clarke    Austr~ batt~ Test.1..Inn~ Battin~ 6        56     94
## 5      5 Collingwood Engla~ bat  Test.1..Inn~ Battin~ 4         5     13
## 6      6 Cook      Engla~ batt~ Test.1..Inn~ Battin~ 2        11     15
## 7      7 Flintoff  Engla~ all ~ Test.1..Inn~ Battin~ 6         0      3
## 8      8 Gilchrist Austr~ wick~ Test.1..Inn~ Battin~ 7         0      3
## 9      9 Giles     Engla~ bowl Test.1..Inn~ Battin~ 8        24     39
## 10     10 Harmison Engla~ bolw Test.1..Inn~ Battin~ 10        0      5
## # i 250 more rows
```

```
head(ashes_tidy)
```

```
## # A tibble: 6 x 9
##       X batter      team    role test_innings details batting_order score balls
##   <int> <chr>      <fct>  <fct> <chr>      <chr>      <int> <int> <int>
## 1      1 Anderson  England bowl Test.1..Inn~ Battin~ 11        2      8
```

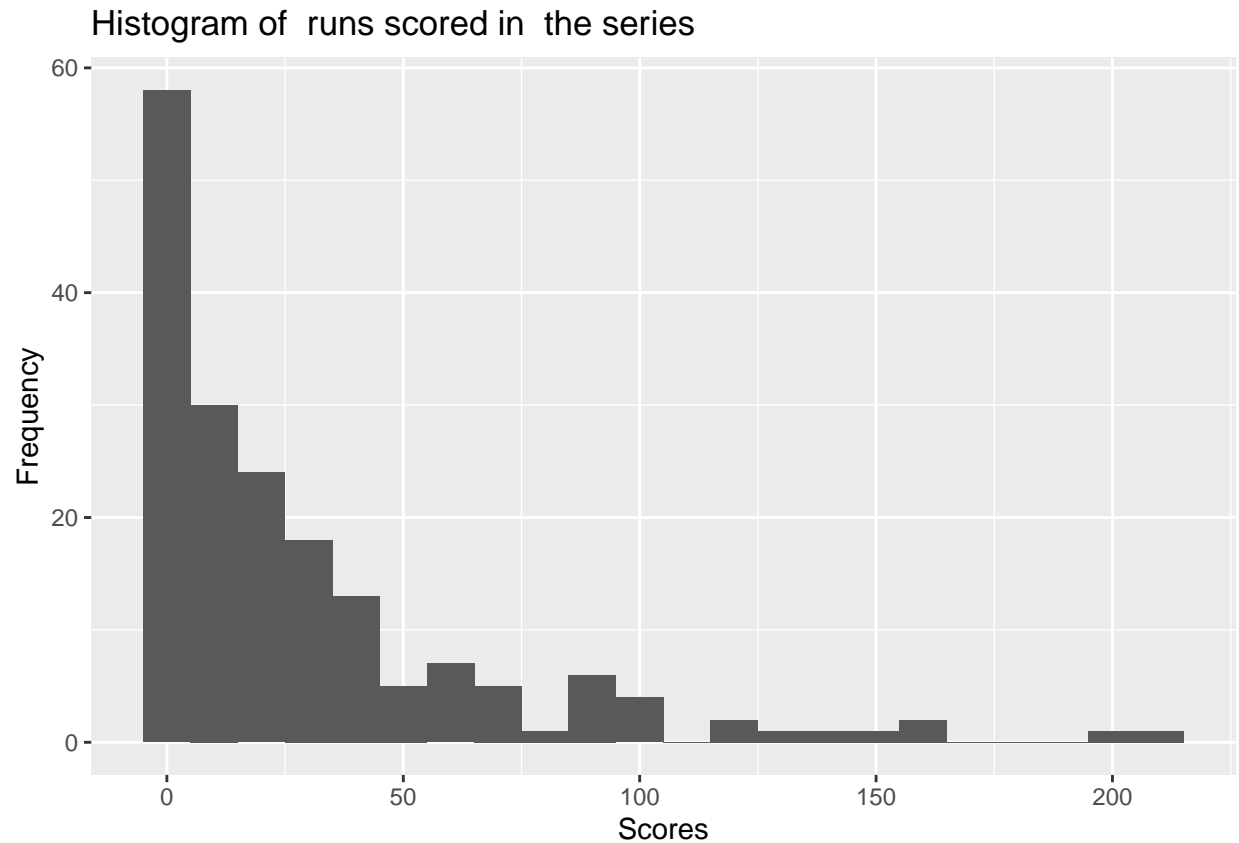
```
## 2      2 Bell      England bats~ Test.1..Inn~ Battin~      3      50      162
## 3      3 Clark      Austr~ bowl~ Test.1..Inn~ Battin~     10      39      23
## 4      4 Clarke      Austr~ batt~ Test.1..Inn~ Battin~      6      56      94
## 5      5 Collingwood England bat   Test.1..Inn~ Battin~      4       5      13
## 6      6 Cook      England batt~ Test.1..Inn~ Battin~      2      11      15
```

```
#1c
ashes_tidy$role <- ashes_tidy$role %>%
fct_recode(
all_rounder="all rounder",
bowler="bolw",
bowler="bowl",
batsman="batter",
batsman="bat")
ashes_tidy
```

```
## # A tibble: 260 x 9
##       X batter      team  role test_innings details batting_order score balls
##   <int> <chr>      <fct> <fct> <chr>      <chr>          <int> <int> <int>
## 1      1 Anderson  Engla~ bowl~ Test.1..Inn~ Battin~      11       2       8
## 2      2 Bell      Engla~ bats~ Test.1..Inn~ Battin~      3      50      162
## 3      3 Clark      Austr~ bowl~ Test.1..Inn~ Battin~     10      39      23
## 4      4 Clarke      Austr~ bats~ Test.1..Inn~ Battin~      6      56      94
## 5      5 Collingwood Engla~ bats~ Test.1..Inn~ Battin~      4       5      13
## 6      6 Cook      Engla~ bats~ Test.1..Inn~ Battin~      2      11      15
## 7      7 Flintoff  Engla~ all_~ Test.1..Inn~ Battin~      6       0       3
## 8      8 Gilchrist  Austr~ wick~ Test.1..Inn~ Battin~      7       0       3
## 9      9 Giles      Engla~ bowl~ Test.1..Inn~ Battin~      8      24      39
## 10     10 Harmison  Engla~ bowl~ Test.1..Inn~ Battin~     10       0       5
## # i 250 more rows
```

```
#2a
ggplot(data=ashes_tidy, aes(x=score))+
geom_histogram(binwidth = 10)+ labs(title = "Histogram of runs scored in the series", x = "Scores", y
```

```
## Warning: Removed 80 rows containing non-finite values ('stat_bin()').
```



#Q 2b shape:unimodal,right-skewed outliers:- between 190-220 spread :- range lies between 0 to 220  
location:- The peak is between 0 to 10

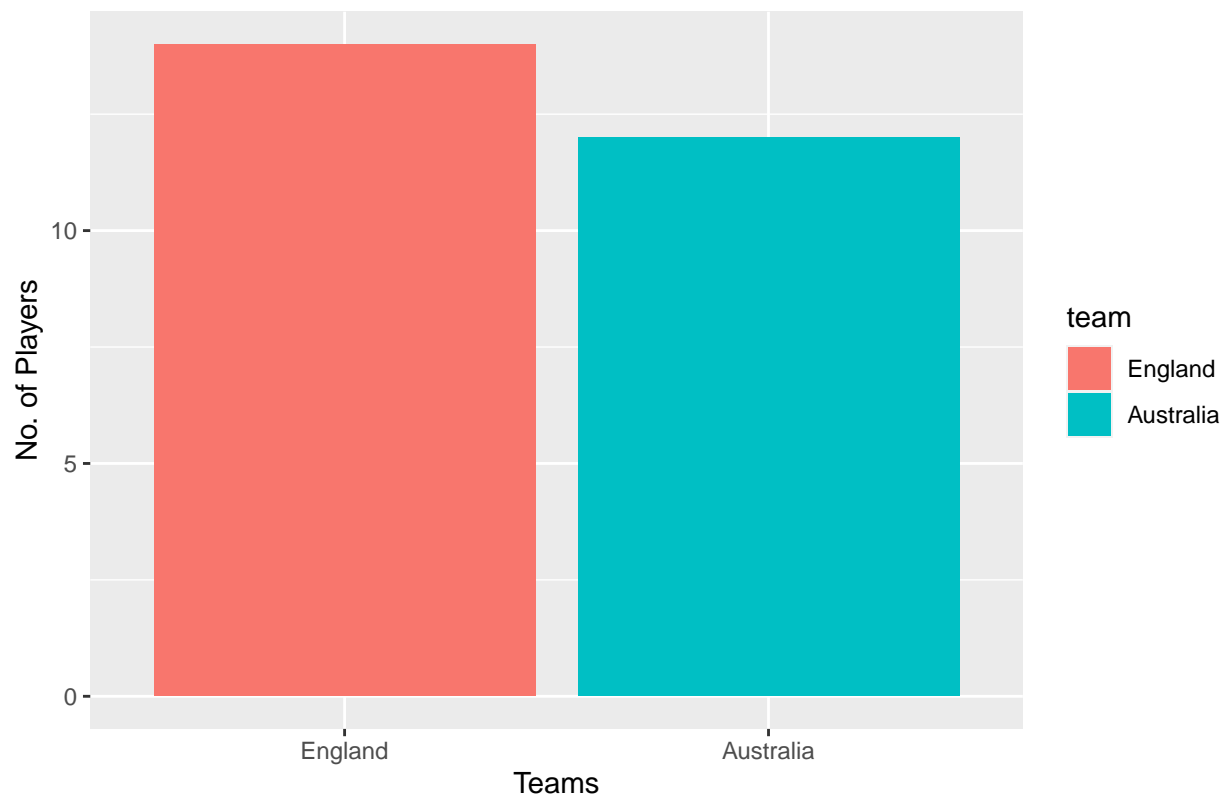
```
#Q2 c
sum_players <- ashes_tidy %>%
  distinct(batter,team) %>%
  group_by(team) %>%
  summarise(players=n())

sum_players
```

```
## # A tibble: 2 x 2
##   team      players
##   <fct>      <int>
## 1 England         14
## 2 Australia        12
```

```
ggplot(sum_players,aes(x=team,y=players, fill = team)) +
  geom_bar(stat="identity") +
  labs(title = "Bar Chart representing different Teams Participating in the Series", x = "Teams", y = "N
```

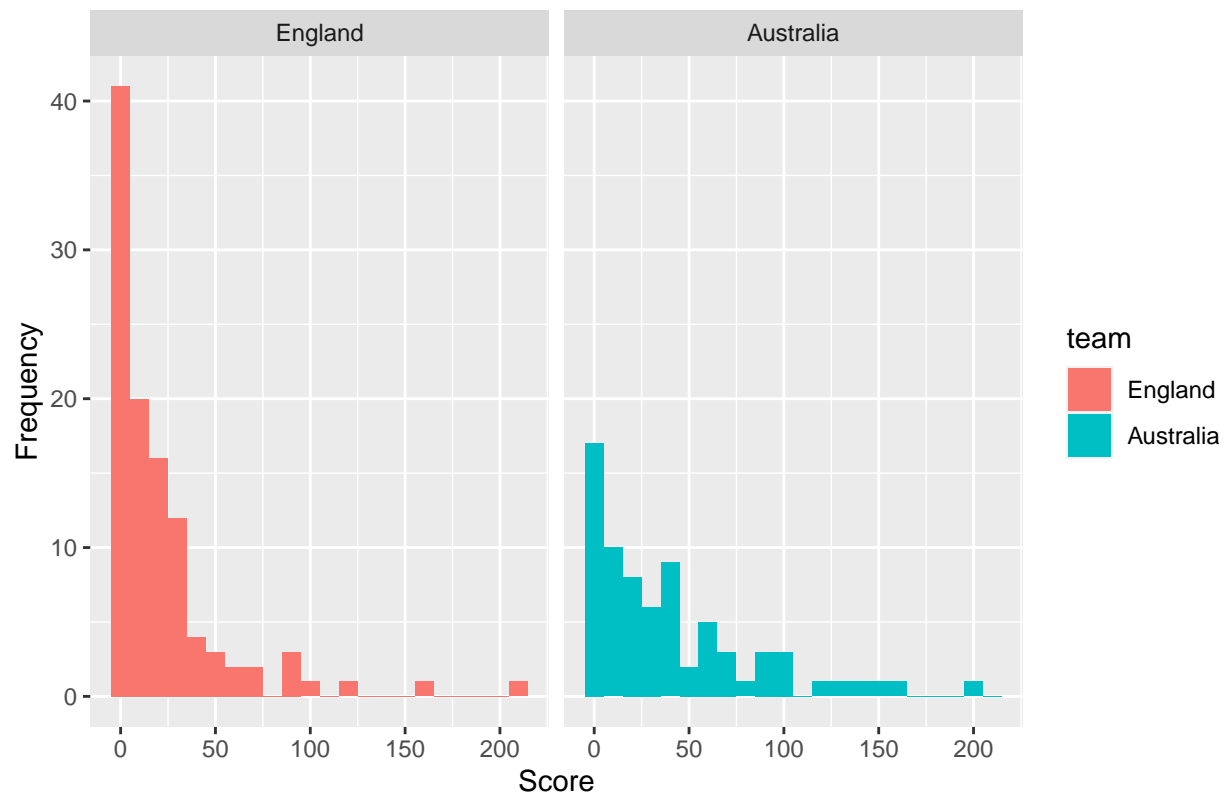
Bar Chart representing different Teams Participating in the Series



```
#Q3 a
ggplot(ashes_tidy, aes(x = score, fill = team)) +
  geom_histogram(position = "identity", binwidth = 10) +
  labs(title = "Histogram of Runs Scored in the Ashes Series (by Team Faceted)",
    x = "Score",
    y = "Frequency") +
  facet_wrap(~team)
```

```
## Warning: Removed 80 rows containing non-finite values ('stat_bin()').
```

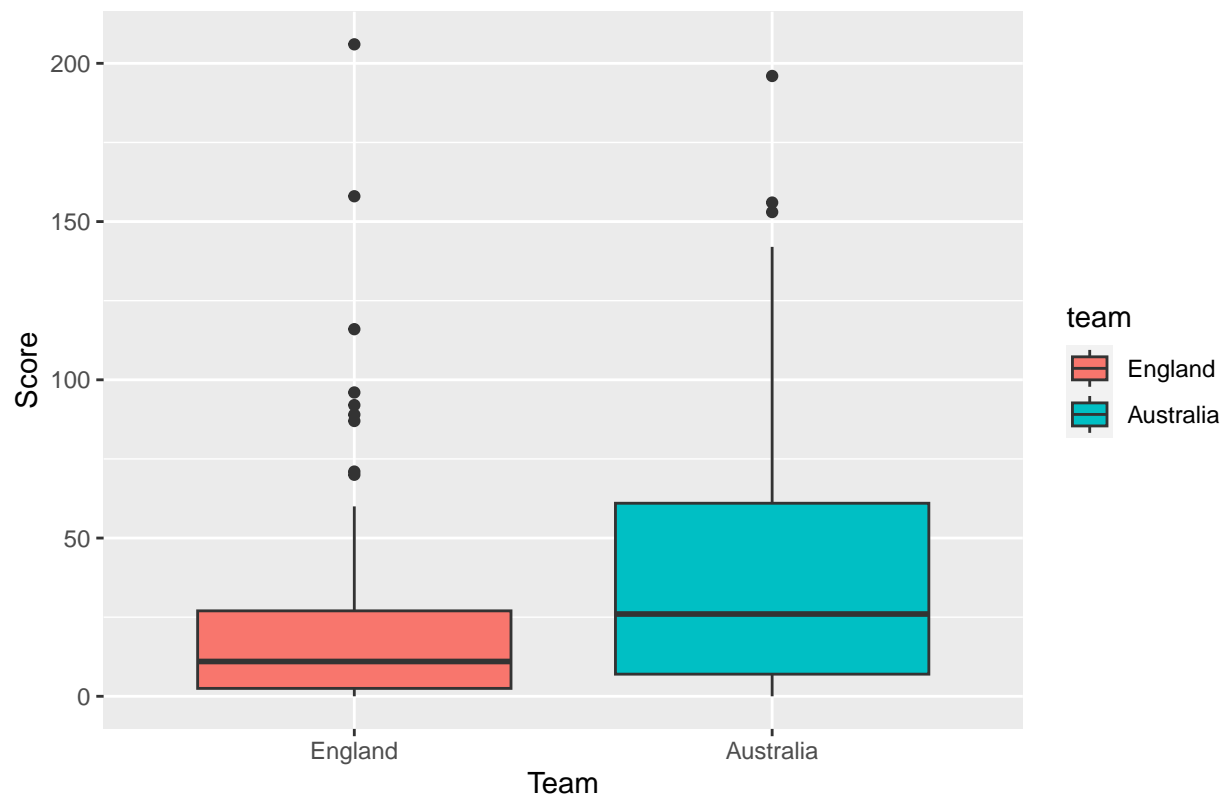
Histogram of Runs Scored in the Ashes Series (by Team Faceted)



```
# Q3 b
ggplot(ashes_tidy, aes(x = team, y = score, fill = team)) +
  geom_boxplot() +
  labs(title = "Boxplots of Runs Scored during the Ashes Series by team",
x = "Team",
y = "Score")
```

```
## Warning: Removed 80 rows containing non-finite values ('stat_boxplot()').
```

Boxplots of Runs Scored during the Ashes Series by team

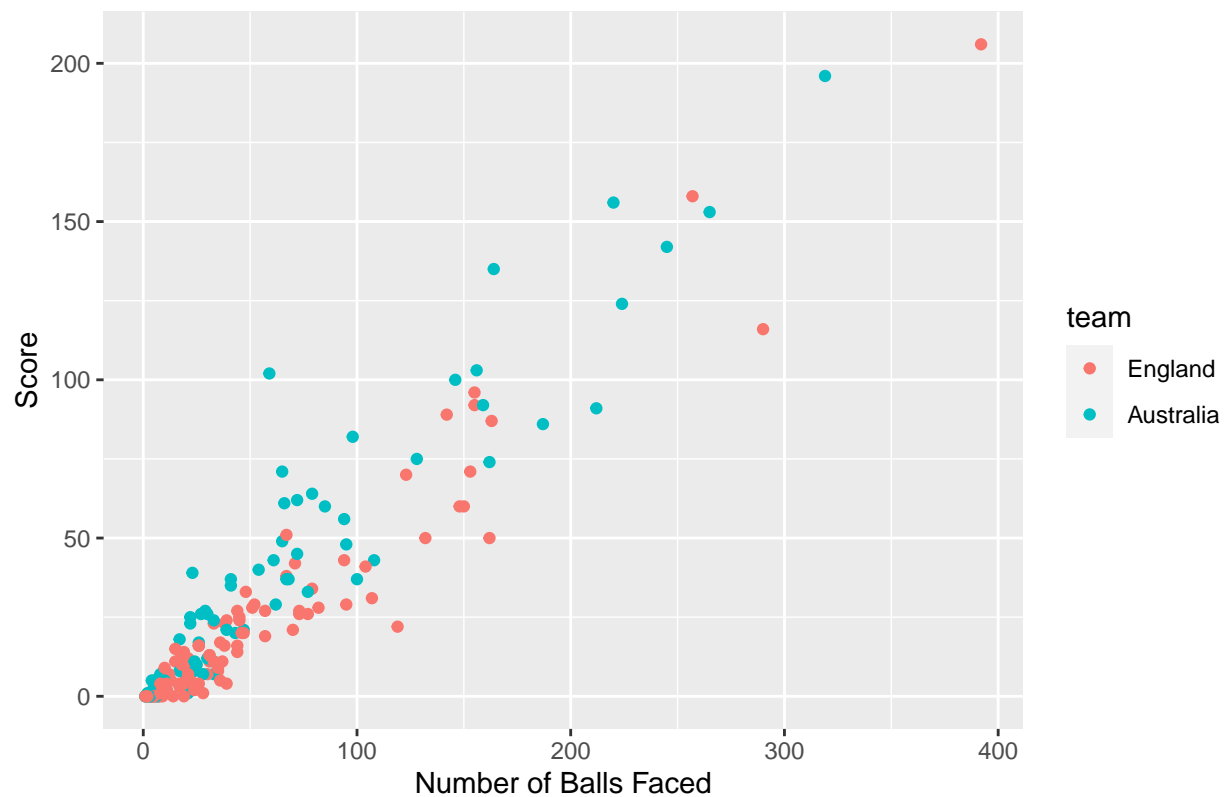


Shape: The distribution of scores in Australia seems to be right - skewed . The amount of low scores are more as compared to less high scores. In the case of box plot , the box is slightly right-skewed distribution. Now talking about England , the distribution for England is also same as Australia , that is right skewed. Location:- Australia distribution according to the histogram and boxplot is around 20 to 30 runs Location :- Once again same conclusion can be made , that is the center of the distribution for England is also around 20 to 30 runs Spread: Australia distribution has a wider spread as shown in histogram , displaying a high range of scores spread across. England on the other hand also have a wide spread , which is evident by observing both of the plots. This means just like Australia , England too has variety of scores. Outliers: Both Australia and England are displaying outliers in high scores which can be seen when referring the box plots . Conclusion :- According to my Analysis , what i have discovered is that both teams have very much similar variability in scores , but England seems to have a bit lower variability then Australia. Another conclusion that i was able to make was both the teams have a right-skewed distribution accompanied by wide spread of sources and outliers.

```
# Q 4A
ggplot(ashes_tidy, aes(x = balls, y = score, color = team)) +
  geom_point() +
  labs(title = "Scatterplot of Runs Scored vs Number of Balls Faced",
       x = "Number of Balls Faced",
       y = "Score")
```

```
## Warning: Removed 80 rows containing missing values ('geom_point()').
```

### Scatterplot of Runs Scored vs Number of Balls Faced



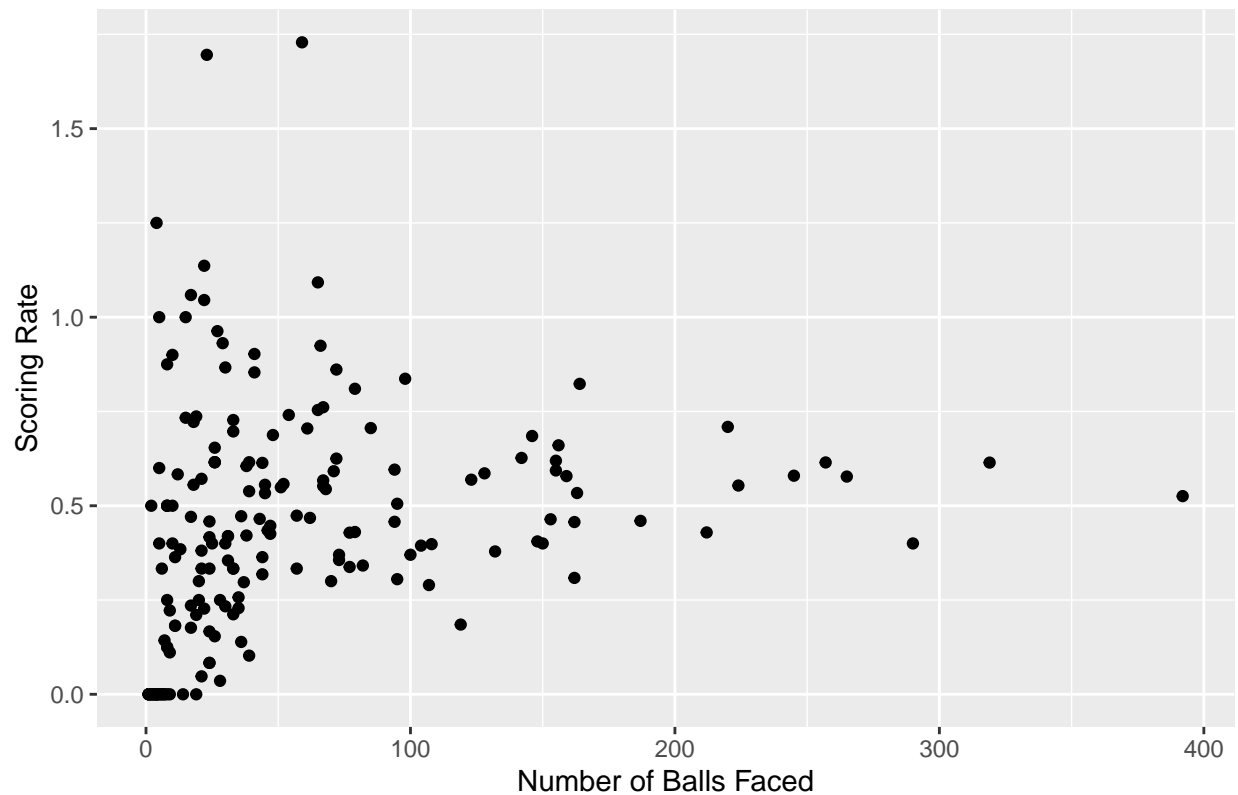
Q4 B The conclusions that can be made after observing the scatterplot is , a positive relationship can be seen between score and the number of balls the batsman faced. second conclusion that i was able to identify was , if we ignore few outliers in the case of Australia , it is highly possible for the players had faced more balls in order to score more runs

```
ashes_tidy <- ashes_tidy %>%
mutate(scoring_rate = score / balls)
ggplot(ashes_tidy, aes(x = balls, y = scoring_rate)) +
geom_point() +
labs(title = "Scatterplot of Scoring Rate VS The number of Balls Faced",
x = "Number of Balls Faced",
y = "Scoring Rate")
```

```
## Warning: Removed 80 rows containing missing values ('geom_point()').
```



Scatterplot of Scoring Rate VS The number of Balls Faced



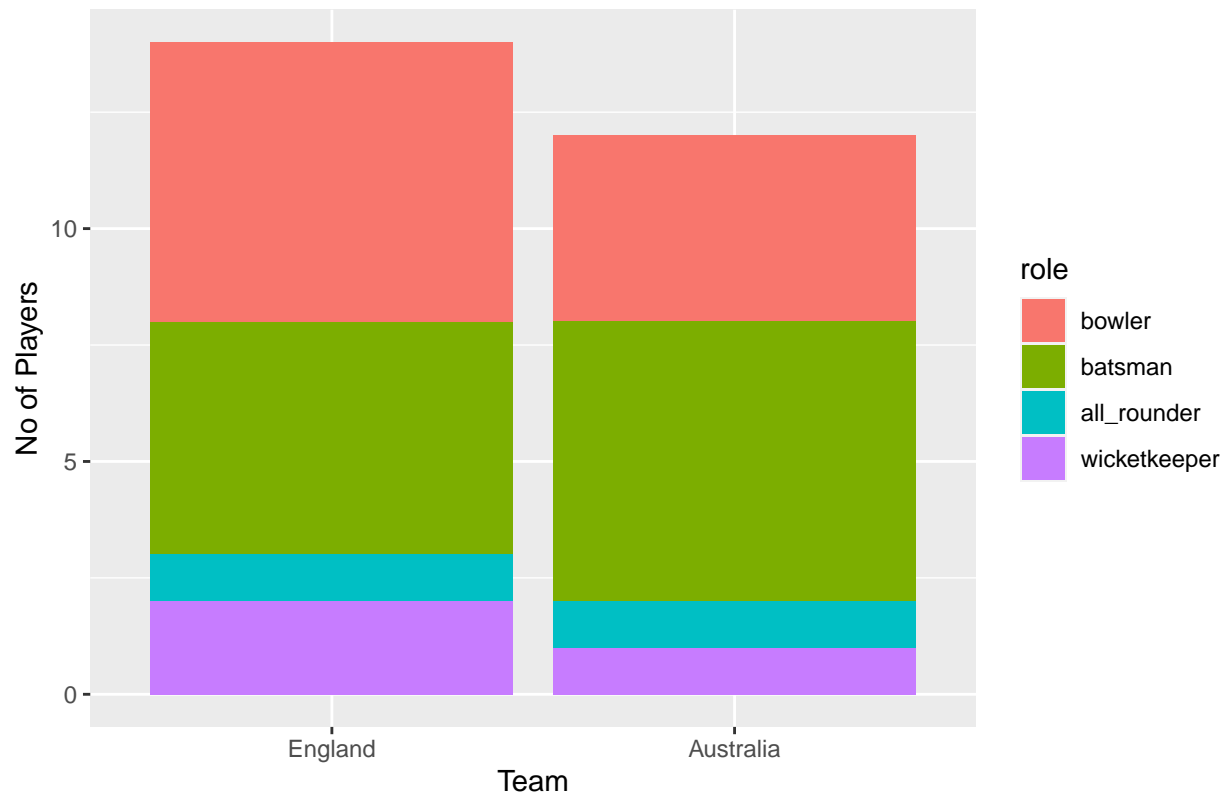
#Q4 d The scatterplot shown above shows us a mixed relationship between the scoring rate on y-axis and the number of balls faced by the batsman on the x-axis. If observed seriously, we can also observe there is no clear trend and the players who were facing more balls are not scoring more runs quickly and the players who are facing less balls aren't scoring slowly .

```
player_details <- ashes_tidy %>%
distinct(batter, team, role) %>%
group_by(team, role) %>%
summarise(players= n())
```

## 'summarise()' has grouped output by 'team'. You can override using the  
## '.groups' argument.

```
ggplot(player_details, aes(x = team, y = players, fill = role)) +
geom_bar(stat = "identity") +
labs(title = "The number of Players per Team by their role",
x = "Team",
y = "No of Players")
```

The number of Players per Team by their role



```
contingency_table <- ashes_tidy$role %>%
table(ashes_tidy$team) %>%
prop.table()
contingency_table
```

```
##
## .           England Australia
##  bowler      0.23076923 0.15384615
##  batsman     0.19230769 0.23076923
##  all_rounder 0.03846154 0.03846154
##  wicketkeeper 0.07692308 0.03846154
```

With the help of the above figure, It is very obvious that Australia mainly consist of batsman , but when it comes to all rounder , both the teams share equal proportions.