Assignemgent 1

A1844790 - Dylan Rohan

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.5 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.0.2 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(dplyr)  
library(stringr)  
library(rmarkdown)  
  
ashes <- read\_csv("C:\\Users\\rohad\\OneDrive\\Documents\\Data science\\Data Taming, modelling and Vizalization\_RStudio\\a1\\a1\\ashes.csv")

## Rows: 27 Columns: 13

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (13): batter, team, role, Test 1, Innings 1, Test 1, Innings 2, Test 2, ...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#double slashes for windows directory  
knitr::kable(head(ashes), caption = "Checking the table, currently a tibble of 27 x 13.")

Checking the table, currently a tibble of 27 x 13.

| batter | team | role | Test 1, Innings 1 | Test 1, Innings 2 | Test 2, Innings 1 | Test 2, Innings 2 | Test 3, Innings 1 | Test 3, Innings 2 | Test 4, Innings 1 | Test 4, Innings 2 | Test 5, Innings 1 | Test 5, Innings 2 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ali | England | allrounder | Batting at number 6, scored 38 runs from 102 balls including 2 fours and 1 sixes. | Batting at number 6, scored 40 runs from 64 balls including 6 fours and 0 sixes. | Batting at number 6, scored 25 runs from 57 balls including 2 fours and 0 sixes. | Batting at number 7, scored 2 runs from 20 balls including 0 fours and 0 sixes. | Batting at number 7, scored 0 runs from 2 balls including 0 fours and 0 sixes. | Batting at number 7, scored 11 runs from 56 balls including 2 fours and 0 sixes. | Batting at number 7, scored 20 runs from 14 balls including 2 fours and 1 sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number 7, scored 30 runs from 58 balls including 2 fours and 0 sixes. | Batting at number 7, scored 13 runs from 43 balls including 1 fours and 0 sixes. |
| Anderson | English | bowl | Batting at number 11, scored 5 runs from 9 balls including 1 fours and 0 sixes. | Batting at number 11, scored 0 runs from 1 balls including 0 fours and 0 sixes. | Batting at number 11, scored 0 runs from 3 balls including 0 fours and 0 sixes. | Batting at number 11, scored 0 runs from 0 balls including 0 fours and 0 sixes. | Batting at number 11, scored 0 runs from 7 balls including 0 fours and 0 sixes. | Batting at number 11, scored 1 runs from 7 balls including 0 fours and 0 sixes. | Batting at number 11, scored 0 runs from 16 balls including 0 fours and 0 sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number 11, scored 0 runs from 3 balls including 0 fours and 0 sixes. | Batting at number 11, scored 2 runs from 23 balls including 0 fours and 0 sixes. |
| Bairstow | England | wicketkeeper | Batting at number 7, scored 9 runs from 24 balls including 1 fours and 0 sixes. | Batting at number 7, scored 42 runs from 75 balls including 2 fours and 1 sixes. | Batting at number 7, scored 21 runs from 50 balls including 2 fours and 0 sixes. | Batting at number 8, scored 36 runs from 57 balls including 5 fours and 0 sixes. | Batting at number 6, scored 119 runs from 215 balls including 18 fours and 0 sixes. | Batting at number 6, scored 14 runs from 26 balls including 3 fours and 0 sixes. | Batting at number 6, scored 22 runs from 39 balls including 3 fours and 0 sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number 6, scored 5 runs from 7 balls including 1 fours and 0 sixes. | Batting at number 6, scored 38 runs from 143 balls including 4 fours and 0 sixes. |
| Ball | England | bowl | Batting at number 10, scored 14 runs from 11 balls including 3 fours and 0 sixes. | Batting at number 10, scored 1 runs from 5 balls including 0 fours and 0 sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. |
| Bancroft | Australia | bat | Batting at number 1, scored 5 runs from 19 balls including 0 fours and 0 sixes. | Batting at number 1, scored 82 runs from 182 balls including 10 fours and 1 sixes. | Batting at number 1, scored 10 runs from 41 balls including 0 fours and 0 sixes. | Batting at number 1, scored 4 runs from 8 balls including 1 fours and 0 sixes. | Batting at number 1, scored 25 runs from 55 balls including 3 fours and 0 sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number 1, scored 26 runs from 95 balls including 2 fours and 0 sixes. | Batting at number 1, scored 27 runs from 42 balls including 4 fours and 0 sixes. | Batting at number 1, scored 0 runs from 7 balls including 0 fours and 0 sixes. | Batting at number NA, scored NA including NA fours and NA sixes. |
| Bird | Australia | bowl | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number 9, scored 4 runs from 6 balls including 1 fours and 0 sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. | Batting at number NA, scored NA including NA fours and NA sixes. |

unique(ashes$team)

## [1] "England" "English" "Australia"

#need to correct variable English to be England  
unique(ashes$role)

## [1] "allrounder" "bowl" "wicketkeeper" "bat" "bowler"   
## [6] "batting" "batsman" "all rounder" "all-rounder"

#many duplicates under alternate variable names, eg. bat, batsman, batting

#### 1.1

colnames(ashes)

## [1] "batter" "team" "role"   
## [4] "Test 1, Innings 1" "Test 1, Innings 2" "Test 2, Innings 1"  
## [7] "Test 2, Innings 2" "Test 3, Innings 1" "Test 3, Innings 2"  
## [10] "Test 4, Innings 1" "Test 4, Innings 2" "Test 5, Innings 1"  
## [13] "Test 5, Innings 2"

ashes\_longform <- gather(ashes, key = "innings", value = "description", "Test 1, Innings 1" : "Test 5, Innings 2")  
ashes\_longform

## # A tibble: 270 x 5  
## batter team role innings description   
## <chr> <chr> <chr> <chr> <chr>   
## 1 Ali England allrounder Test 1, Innings 1 Batting at number 6, score~  
## 2 Anderson English bowl Test 1, Innings 1 Batting at number 11, scor~  
## 3 Bairstow England wicketkeeper Test 1, Innings 1 Batting at number 7, score~  
## 4 Ball England bowl Test 1, Innings 1 Batting at number 10, scor~  
## 5 Bancroft Australia bat Test 1, Innings 1 Batting at number 1, score~  
## 6 Bird Australia bowl Test 1, Innings 1 Batting at number NA, scor~  
## 7 Broad England bowler Test 1, Innings 1 Batting at number 9, score~  
## 8 Cook England bat Test 1, Innings 1 Batting at number 1, score~  
## 9 Crane England bowl Test 1, Innings 1 Batting at number NA, scor~  
## 10 Cummins Australia bowl Test 1, Innings 1 Batting at number 9, score~  
## # ... with 260 more rows

#tibble now in long form, 270 x 5  
ashes\_innings\_first <- ashes\_longform[c(4, 1, 2, 3, 5)]  
knitr::kable(head(ashes\_innings\_first), caption = "Table now in long form with subject first.")

Table now in long form with subject first.

| innings | batter | team | role | description |
| --- | --- | --- | --- | --- |
| Test 1, Innings 1 | Ali | England | allrounder | Batting at number 6, scored 38 runs from 102 balls including 2 fours and 1 sixes. |
| Test 1, Innings 1 | Anderson | English | bowl | Batting at number 11, scored 5 runs from 9 balls including 1 fours and 0 sixes. |
| Test 1, Innings 1 | Bairstow | England | wicketkeeper | Batting at number 7, scored 9 runs from 24 balls including 1 fours and 0 sixes. |
| Test 1, Innings 1 | Ball | England | bowl | Batting at number 10, scored 14 runs from 11 balls including 3 fours and 0 sixes. |
| Test 1, Innings 1 | Bancroft | Australia | bat | Batting at number 1, scored 5 runs from 19 balls including 0 fours and 0 sixes. |
| Test 1, Innings 1 | Bird | Australia | bowl | Batting at number NA, scored NA including NA fours and NA sixes. |

#### 1.2

order <- str\_match(ashes\_innings\_first$description, "Batting at number ..")  
with\_order <- cbind(ashes\_innings\_first, order)  
#Order now has its own column, values are strings  
runs <- str\_match(with\_order$description, "scored ....")  
with\_runs <- cbind(with\_order, runs)  
#runs now has its own column, values are strings  
no.\_of\_balls <- str\_match(with\_runs$description, "from ....")  
all\_columns<- cbind(with\_runs, no.\_of\_balls)  
#no. of balls now has its own column, values are  
batting\_order <- str\_replace\_all(all\_columns$order, "[^0-9.-]", "")  
runs\_ <- str\_replace\_all(all\_columns$runs, "[^0-9.-]", "")  
balls\_ <- str\_replace\_all(all\_columns$no.\_of\_balls, "[^0-9.-]", "")   
#Taking numerical values from strings  
order1 <- tibble(batting\_order)  
runs1 <- tibble(runs\_)  
balls1 <- tibble(balls\_)  
#making data frames from those values  
a1\_o <- cbind(ashes\_innings\_first, order1)  
a1\_o\_r <- cbind(a1\_o, runs1)  
a1\_o\_r\_b<- cbind(a1\_o\_r, balls1)  
#Order same, so binding columns  
a1o\_r\_b <- a1\_o\_r\_b$description <- NULL  
a1\_o\_r\_b <- a1\_o\_r\_b %>%  
 mutate\_all(na\_if, "")  
#removing description column  
knitr::kable(head(a1\_o\_r\_b), caption = "Now a table of 270 x 7 (removed description, but it's still accessible in 'ashes\_innings\_first'")

Now a table of 270 x 7 (removed description, but it’s still accessible in ‘ashes\_innings\_first’

| innings | batter | team | role | batting\_order | runs\_ | balls\_ |
| --- | --- | --- | --- | --- | --- | --- |
| Test 1, Innings 1 | Ali | England | allrounder | 6 | 38 | 102 |
| Test 1, Innings 1 | Anderson | English | bowl | 11 | 5 | 9 |
| Test 1, Innings 1 | Bairstow | England | wicketkeeper | 7 | 9 | 24 |
| Test 1, Innings 1 | Ball | England | bowl | 10 | 14 | 11 |
| Test 1, Innings 1 | Bancroft | Australia | bat | 1 | 5 | 19 |
| Test 1, Innings 1 | Bird | Australia | bowl | NA | NA | NA |

#\_\_\_\_\_\_\_\_\_\_\_Alternative method\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_#  
  
trial <- ashes\_innings\_first %>%  
 mutate("runs\_"=str\_match(description,"from ....") , "batting\_order" = str\_match(description, "Batting at number .."), "balls\_" = str\_match(description, "scored ...."))  
#description string broken into appropriate columns  
trial <- trial %>%  
 mutate("runs\_" = str\_replace\_all(trial$runs\_, "[^0-9.-]",""), "balls\_"=str\_replace\_all(trial$balls\_, "[^0-9.-]",""), "batting\_order"=str\_replace\_all(trial$batting\_order, "[^0-9.-]",""))  
trial <- mutate\_all(trial, na\_if, "")  
#Left the description column in here, but all is right with the world  
knitr::kable(head(trial), caption = "Alternate method to get the same answer.")

Alternate method to get the same answer.

| innings | batter | team | role | description | runs\_ | batting\_order | balls\_ |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Test 1, Innings 1 | Ali | England | allrounder | Batting at number 6, scored 38 runs from 102 balls including 2 fours and 1 sixes. | 102 | 6 | 38 |
| Test 1, Innings 1 | Anderson | English | bowl | Batting at number 11, scored 5 runs from 9 balls including 1 fours and 0 sixes. | 9 | 11 | 5 |
| Test 1, Innings 1 | Bairstow | England | wicketkeeper | Batting at number 7, scored 9 runs from 24 balls including 1 fours and 0 sixes. | 24 | 7 | 9 |
| Test 1, Innings 1 | Ball | England | bowl | Batting at number 10, scored 14 runs from 11 balls including 3 fours and 0 sixes. | 11 | 10 | 14 |
| Test 1, Innings 1 | Bancroft | Australia | bat | Batting at number 1, scored 5 runs from 19 balls including 0 fours and 0 sixes. | 19 | 1 | 5 |
| Test 1, Innings 1 | Bird | Australia | bowl | Batting at number NA, scored NA including NA fours and NA sixes. | NA | NA | NA |

# \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_#

#### 1.3

ashes\_tibble <- as\_tibble(a1\_o\_r\_b)  
#making a data frame from a1\_o\_r\_b to set value type  
ashes\_tibble$batting\_order <- as.factor(ashes\_tibble$batting\_order)  
#low level ordinal, label = factor  
ashes\_tibble$runs\_ <- as.integer(ashes\_tibble$runs\_)  
ashes\_tibble$balls\_ <- as.integer(ashes\_tibble$balls\_)  
#countable, discrete = integer  
ashes\_tibble$innings <- as.factor(ashes\_tibble$innings)  
#innings total = 10 (low level and ordinal), is a label/name = factors  
ashes\_tibble <- rename(ashes\_tibble,"player"="batter")  
ashes\_tibble$player <- as.character(ashes\_tibble$player)  
#player makes more sense as a variable name. The teams have several people that could take the position, categorical variable = character. Stores string information which might be useful in future analysis.  
ashes\_tibble$team <- as.factor(ashes\_tibble$team)  
ashes\_tibble$role <- as.factor(ashes\_tibble$role)  
#both low value labels, so factors   
#demonstrating the value types have been set:  
ashes\_tibble

## # A tibble: 270 x 7  
## innings player team role batting\_order runs\_ balls\_  
## <fct> <chr> <fct> <fct> <fct> <int> <int>  
## 1 Test 1, Innings 1 Ali England allrounder 6 38 102  
## 2 Test 1, Innings 1 Anderson English bowl 11 5 9  
## 3 Test 1, Innings 1 Bairstow England wicketkeeper 7 9 24  
## 4 Test 1, Innings 1 Ball England bowl 10 14 11  
## 5 Test 1, Innings 1 Bancroft Australia bat 1 5 19  
## 6 Test 1, Innings 1 Bird Australia bowl <NA> NA NA  
## 7 Test 1, Innings 1 Broad England bowler 9 20 32  
## 8 Test 1, Innings 1 Cook England bat 1 2 10  
## 9 Test 1, Innings 1 Crane England bowl <NA> NA NA  
## 10 Test 1, Innings 1 Cummins Australia bowl 9 42 120  
## # ... with 260 more rows

#### 1.4

unique(ashes\_tibble$player)

## [1] "Ali" "Anderson" "Bairstow" "Ball" "Bancroft" "Bird"   
## [7] "Broad" "Cook" "Crane" "Cummins" "Curran" "Handscomb"  
## [13] "Hazlewood" "Khawaja" "Lyon" "Malan" "MMarsh" "Overton"   
## [19] "Paine" "Root" "SMarsh" "Smith" "Starc" "Stoneman"   
## [25] "Vince" "Warner" "Woakes"

summary(unique(ashes\_tibble$innings))

## Test 1, Innings 1 Test 1, Innings 2 Test 2, Innings 1 Test 2, Innings 2   
## 1 1 1 1   
## Test 3, Innings 1 Test 3, Innings 2 Test 4, Innings 1 Test 4, Innings 2   
## 1 1 1 1   
## Test 5, Innings 1 Test 5, Innings 2   
## 1 1

summary(unique(ashes\_tibble$team))

## Australia England English   
## 1 1 1

unique(ashes\_tibble$team)

## [1] England English Australia  
## Levels: Australia England English

unique(ashes\_tibble$role)

## [1] allrounder bowl wicketkeeper bat bowler   
## [6] batting batsman all rounder all-rounder   
## 9 Levels: all-rounder all rounder allrounder bat batsman batting ... wicketkeeper

#English to England, unify roles  
ashes\_corrected\_ <- ashes\_tibble %>%  
 mutate(team = fct\_recode(team, "England" = "English"))%>%  
 mutate(role = fct\_recode(role, "all-rounder" = "allrounder", "all-rounder"="all rounder", "batsman"="batting", "batsman"="bat", "bowler"="bowl"))  
ac <- ashes\_corrected\_  
knitr::kable(head(ac), caption = "Table demonstrating the data is now clean and tame")

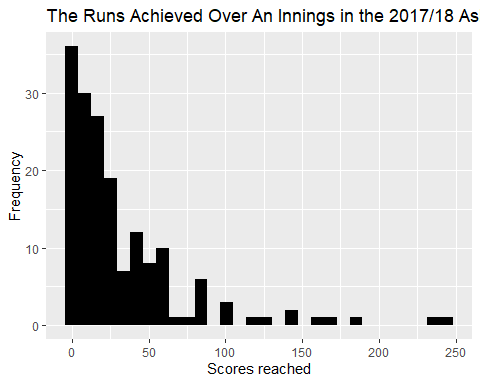
Table demonstrating the data is now clean and tame

| innings | player | team | role | batting\_order | runs\_ | balls\_ |
| --- | --- | --- | --- | --- | --- | --- |
| Test 1, Innings 1 | Ali | England | all-rounder | 6 | 38 | 102 |
| Test 1, Innings 1 | Anderson | England | bowler | 11 | 5 | 9 |
| Test 1, Innings 1 | Bairstow | England | wicketkeeper | 7 | 9 | 24 |
| Test 1, Innings 1 | Ball | England | bowler | 10 | 14 | 11 |
| Test 1, Innings 1 | Bancroft | Australia | batsman | 1 | 5 | 19 |
| Test 1, Innings 1 | Bird | Australia | bowler | NA | NA | NA |

#### 2.1

#Histogram default below, bin of 30  
ggplot(ac)+geom\_histogram(aes(x=runs\_, ), fill= "black", na.rm=TRUE) +   
 ggtitle("The Runs Achieved Over An Innings in the 2017/18 Ashes Series")+  
 labs(x= "Scores reached", y ="Frequency")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#ac$runs\_ %>%  
# unique()  
#cool find -> 70 unique values excluding NA, bin of 70 width = 1 for a bar chart as below  
#ggplot(ac)+geom\_histogram(mapping = aes(x=runs\_), na.rm=TRUE, bins=70, binwidth = 1)+  
# ggtitle("Total runs acheieved")+labs(x= "Total runs")

#### 2.2

summary(ac$runs\_, na.rm = TRUE)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00 6.00 18.00 32.09 41.00 244.00 101

range(ac$runs\_, na.rm = TRUE, finite= TRUE)

## [1] 0 244

sd(ac$runs\_, na.rm = TRUE)

## [1] 41.30805

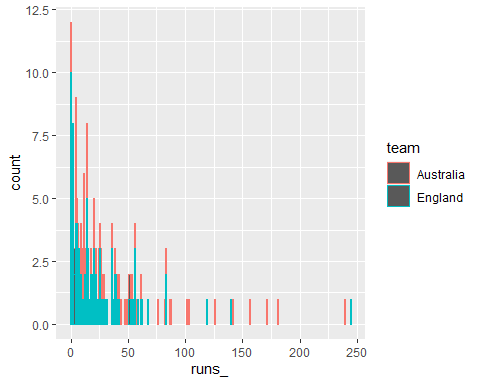
table(ac$runs\_)

##   
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19   
## 12 4 8 3 9 5 4 3 2 4 3 6 3 2 8 3 1 3 2 2   
## 20 21 22 23 24 25 26 27 28 29 30 31 36 37 38 39 40 41 42 44   
## 5 1 3 1 3 4 3 2 1 2 1 1 4 1 3 2 2 2 2 1   
## 47 49 50 51 53 54 55 56 57 58 61 62 67 76 82 83 86 87 101 102   
## 1 1 1 2 2 1 1 4 1 1 2 1 1 1 1 3 1 1 1 1   
## 103 119 126 140 141 156 171 181 239 244   
## 1 1 1 1 1 1 1 1 1 1

#### 2.3

ggplot(ac, aes(x= runs\_, col=team))+geom\_bar()

## Warning: Removed 101 rows containing non-finite values (stat\_count).



#^this maps every players innings, we need to combine player scores across the innings  
indiv\_runs <- ac%>%  
 group\_by(player) %>%  
 summarise(team,role,runs\_in\_series = sum(runs\_, na.rm=TRUE))%>%  
 unique()

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

knitr::kable(head(indiv\_runs), caption = "Demonstrating the scores have been totalled for each player")

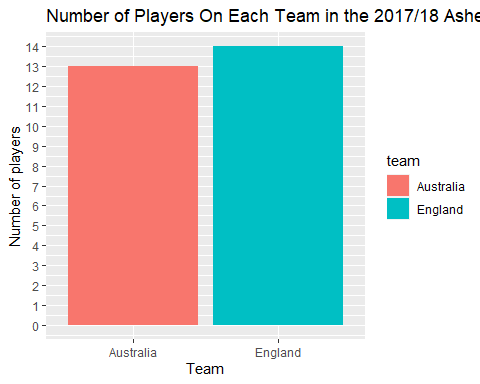
Demonstrating the scores have been totalled for each player

| player | team | role | runs\_in\_series |
| --- | --- | --- | --- |
| Ali | England | all-rounder | 179 |
| Anderson | England | bowler | 8 |
| Bairstow | England | wicketkeeper | 306 |
| Ball | England | bowler | 15 |
| Bancroft | Australia | batsman | 179 |
| Bird | Australia | bowler | 4 |

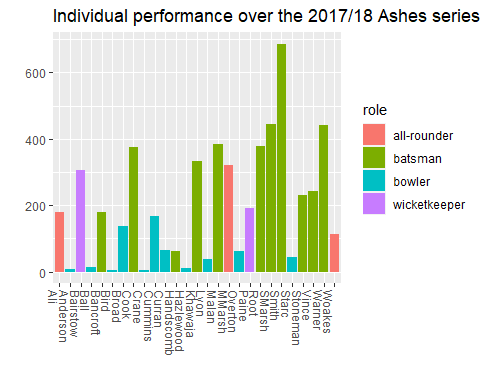
unique(ac$player)

## [1] "Ali" "Anderson" "Bairstow" "Ball" "Bancroft" "Bird"   
## [7] "Broad" "Cook" "Crane" "Cummins" "Curran" "Handscomb"  
## [13] "Hazlewood" "Khawaja" "Lyon" "Malan" "MMarsh" "Overton"   
## [19] "Paine" "Root" "SMarsh" "Smith" "Starc" "Stoneman"   
## [25] "Vince" "Warner" "Woakes"

#all players accounted for  
ggplot(indiv\_runs, aes(x=team, fill=team))+  
 geom\_bar()+ggtitle("Number of Players On Each Team in the 2017/18 Ashes Series")+  
 scale\_y\_continuous(breaks = seq(0, 20, by = 1))+  
 labs(x = "Team", y= "Number of players")



#players per team^   
  
 #\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
#What I thought question 2.3 wanted  
indiv\_runs %>%  
 ggplot(aes(x=player, y=runs\_in\_series, fill=role))+  
 geom\_bar(stat="identity")+  
 ggtitle("Individual performance over the 2017/18 Ashes series")+  
 labs(x = "", y= "")+  
 theme(axis.text.x= element\_text(angle =-90, hjust = 0))



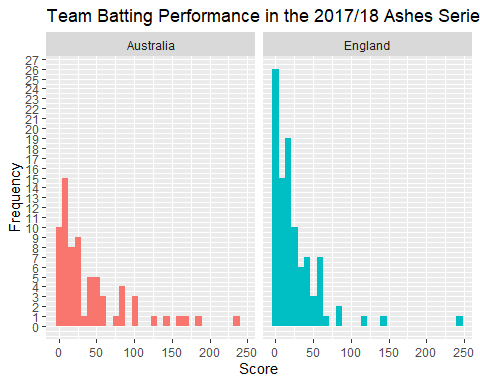
#score per player  
 #\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#### 3.1

ac %>%  
 ggplot(aes(x=runs\_, fill=team))+  
 geom\_histogram(show.legend = FALSE)+  
 scale\_y\_continuous(breaks = seq(0, 30, by = 1))+  
 facet\_wrap(~team)+  
 ggtitle("Team Batting Performance in the 2017/18 Ashes Series")+  
 labs(x = "Score", y= "Frequency")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

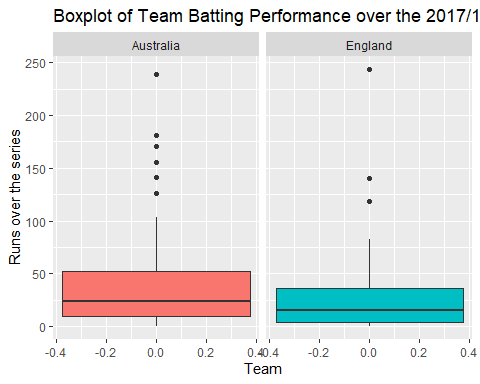
## Warning: Removed 101 rows containing non-finite values (stat\_bin).



#### 3.2

ac %>%  
 ggplot(aes(y=runs\_, fill=team))+  
 geom\_boxplot(show.legend = FALSE)+  
 facet\_grid(~team)+  
 ggtitle("Boxplot of Team Batting Performance over the 2017/18 Ashes Series")+  
 labs(x = "Team", y="Runs over the series")

## Warning: Removed 101 rows containing non-finite values (stat\_boxplot).



#### 3.3

#ENGLISH INDIVIDUALS   
england\_players <- ac[ac$team != "Australia", ]  
summary(england\_players$runs\_, na.rm =TRUE)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00 4.00 15.00 25.28 36.00 244.00 41

sd(england\_players$runs\_, na.rm = TRUE)

## [1] 33.61336

england\_players %>%  
 arrange(runs\_)

## # A tibble: 140 x 7  
## innings player team role batting\_order runs\_ balls\_  
## <fct> <chr> <fct> <fct> <fct> <int> <int>  
## 1 Test 1, Innings 1 Woakes England all-rounder 8 0 4  
## 2 Test 1, Innings 2 Anderson England bowler 11 0 1  
## 3 Test 2, Innings 1 Anderson England bowler 11 0 3  
## 4 Test 2, Innings 2 Anderson England bowler 11 0 0  
## 5 Test 3, Innings 1 Ali England all-rounder 7 0 2  
## 6 Test 3, Innings 1 Anderson England bowler 11 0 7  
## 7 Test 3, Innings 2 Broad England bowler 10 0 2  
## 8 Test 4, Innings 1 Anderson England bowler 11 0 16  
## 9 Test 5, Innings 1 Anderson England bowler 11 0 3  
## 10 Test 5, Innings 2 Stoneman England batsman 2 0 9  
## # ... with 130 more rows

#England's statistics  
  
  
#AUSTRALIAN INDIVIDUALS  
aus\_players <- ac[ac$team != "England",]  
summary(aus\_players$runs\_, na.rm= TRUE)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.00 10.00 24.00 41.71 52.50 239.00 60

sd(aus\_players$runs\_, na.rm = TRUE)

## [1] 48.88174

aus\_players %>%  
 arrange(runs\_)

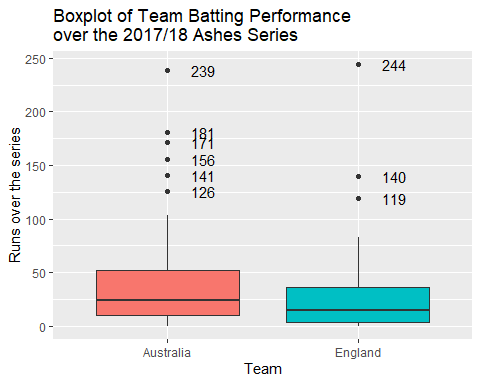
## # A tibble: 130 x 7  
## innings player team role batting\_order runs\_ balls\_  
## <fct> <chr> <fct> <fct> <fct> <int> <int>  
## 1 Test 4, Innings 1 Lyon Australia bowler 11 0 10  
## 2 Test 5, Innings 1 Bancroft Australia batsman 1 0 7  
## 3 Test 3, Innings 1 Starc Australia bowler 8 1 3  
## 4 Test 4, Innings 1 Hazlewood Australia bowler 10 1 12  
## 5 Test 2, Innings 2 Hazlewood Australia bowler 11 3 7  
## 6 Test 2, Innings 2 Bancroft Australia batsman 1 4 8  
## 7 Test 3, Innings 1 Lyon Australia bowler 10 4 3  
## 8 Test 4, Innings 1 Bird Australia bowler 9 4 6  
## 9 Test 4, Innings 1 Cummins Australia bowler 8 4 18  
## 10 Test 4, Innings 2 SMarsh Australia batsman 5 4 22  
## # ... with 120 more rows

#for outliers  
ggplot(ac, aes(x = team, y = runs\_, fill =team)) +   
 geom\_boxplot(show.legend = FALSE) +  
 stat\_summary(  
 aes(label = round(stat(y), 1)),  
 geom = "text",   
 fun.y = function(y) { o <- boxplot.stats(y)$out; if(length(o) == 0) NA else o },  
 hjust = -1)+  
 ggtitle("Boxplot of Team Batting Performance \nover the 2017/18 Ashes Series")+  
 labs(x = "Team", y="Runs over the series")

## Warning: `fun.y` is deprecated. Use `fun` instead.

## Warning: Removed 101 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 101 rows containing non-finite values (stat\_summary).



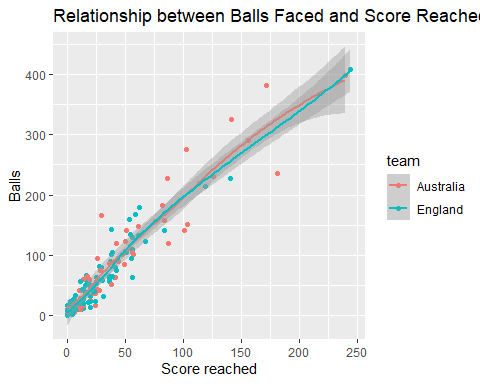
#### 4.1

ggplot(ac, aes( x = runs\_, y= balls\_, col=team))+  
 geom\_point()+  
 geom\_smooth()+  
 ggtitle("Relationship between Balls Faced and Score Reached in the 2017/18 Ashes Series")+  
 labs(x = "Score reached", y="Balls")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 101 rows containing non-finite values (stat\_smooth).

## Warning: Removed 101 rows containing missing values (geom\_point).



#### 4.3

scoring\_rate\_tibble <- ac %>%  
 mutate(scoring\_rates = runs\_/balls\_)  
knitr::kable(head(scoring\_rate\_tibble), caption = "Introduced a scoring rate column.")

Introduced a scoring rate column.

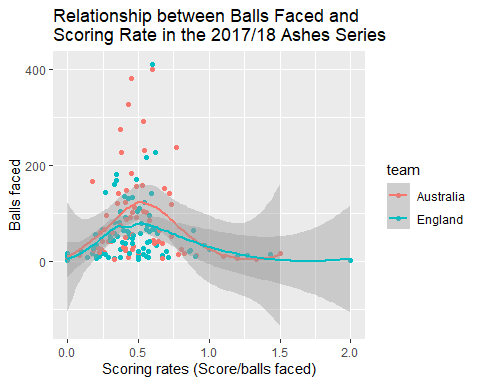
| innings | player | team | role | batting\_order | runs\_ | balls\_ | scoring\_rates |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Test 1, Innings 1 | Ali | England | all-rounder | 6 | 38 | 102 | 0.3725490 |
| Test 1, Innings 1 | Anderson | England | bowler | 11 | 5 | 9 | 0.5555556 |
| Test 1, Innings 1 | Bairstow | England | wicketkeeper | 7 | 9 | 24 | 0.3750000 |
| Test 1, Innings 1 | Ball | England | bowler | 10 | 14 | 11 | 1.2727273 |
| Test 1, Innings 1 | Bancroft | Australia | batsman | 1 | 5 | 19 | 0.2631579 |
| Test 1, Innings 1 | Bird | Australia | bowler | NA | NA | NA | NA |

ggplot(scoring\_rate\_tibble, aes( x = scoring\_rates, y= balls\_, col=team))+  
 geom\_point()+  
 geom\_smooth()+  
 ggtitle("Relationship between Balls Faced and \nScoring Rate in the 2017/18 Ashes Series")+  
 labs(x = "Scoring rates (Score/balls faced)", y="Balls faced")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

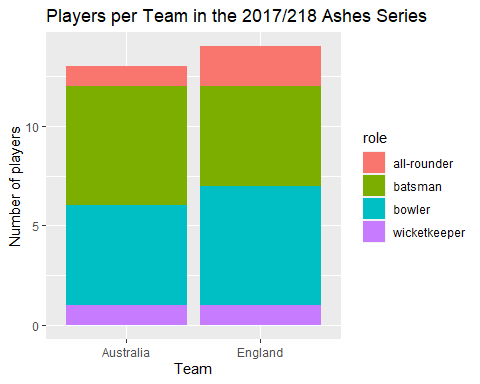
## Warning: Removed 102 rows containing non-finite values (stat\_smooth).

## Warning: Removed 102 rows containing missing values (geom\_point).



#### 5.1

ggplot(indiv\_runs, aes(x=team, fill=role))+  
 geom\_bar()+  
 ggtitle("Players per Team in the 2017/218 Ashes Series")+  
 labs(x = "Team", y= "Number of players")



#### 5.2

con\_table <- indiv\_runs %>%  
 group\_by(role) %>%  
 summarise(team, role, player)

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

#keeps 27 subjects and all variables required  
con\_table <- con\_table %>%  
 count(team, role)%>%  
 spread(key = "team", value = n)  
#gives a table showing the total players in each roler per team  
ct <- mutate(con\_table, total = sum(Australia+England))  
#adds a column for row totals  
contingency\_table <- ct%>%  
 mutate(Aus=Australia/total, Eng= England/total)  
#adds a column indicating the proportion of each  
contingency\_table <- contingency\_table %>%  
 mutate(Australia = NULL, England =NULL, total=NULL)  
#removes unnecessary columns to reveal the...  
knitr::kable(head(contingency\_table), caption = "Contingency table describing the proportion of roles found in each team.")

Contingency table describing the proportion of roles found in each team.

| role | Aus | Eng |
| --- | --- | --- |
| all-rounder | 0.3333333 | 0.6666667 |
| batsman | 0.5454545 | 0.4545455 |
| bowler | 0.4545455 | 0.5454545 |
| wicketkeeper | 0.5000000 | 0.5000000 |

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_Alternate method\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_#  
install.packages("gmodels", repo = "https://cran.rstudio.com/bin/windows/Rtools/")

## Installing package into 'C:/Users/rohad/OneDrive/Documents/R/win-library/4.1'  
## (as 'lib' is unspecified)

## Warning: unable to access index for repository https://cran.rstudio.com/bin/windows/Rtools/src/contrib:  
## cannot open URL 'https://cran.rstudio.com/bin/windows/Rtools/src/contrib/PACKAGES'

## Warning: package 'gmodels' is not available for this version of R  
##   
## A version of this package for your version of R might be available elsewhere,  
## see the ideas at  
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages

## Warning: unable to access index for repository https://cran.rstudio.com/bin/windows/Rtools/bin/windows/contrib/4.1:  
## cannot open URL 'https://cran.rstudio.com/bin/windows/Rtools/bin/windows/contrib/4.1/PACKAGES'

library(gmodels)  
CrossTable(indiv\_runs$role, indiv\_runs$team)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Chi-square contribution |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 27   
##   
##   
## | indiv\_runs$team   
## indiv\_runs$role | Australia | England | Row Total |   
## ----------------|-----------|-----------|-----------|  
## all-rounder | 1 | 2 | 3 |   
## | 0.137 | 0.127 | |   
## | 0.333 | 0.667 | 0.111 |   
## | 0.077 | 0.143 | |   
## | 0.037 | 0.074 | |   
## ----------------|-----------|-----------|-----------|  
## batsman | 6 | 5 | 11 |   
## | 0.093 | 0.087 | |   
## | 0.545 | 0.455 | 0.407 |   
## | 0.462 | 0.357 | |   
## | 0.222 | 0.185 | |   
## ----------------|-----------|-----------|-----------|  
## bowler | 5 | 6 | 11 |   
## | 0.017 | 0.015 | |   
## | 0.455 | 0.545 | 0.407 |   
## | 0.385 | 0.429 | |   
## | 0.185 | 0.222 | |   
## ----------------|-----------|-----------|-----------|  
## wicketkeeper | 1 | 1 | 2 |   
## | 0.001 | 0.001 | |   
## | 0.500 | 0.500 | 0.074 |   
## | 0.077 | 0.071 | |   
## | 0.037 | 0.037 | |   
## ----------------|-----------|-----------|-----------|  
## Column Total | 13 | 14 | 27 |   
## | 0.481 | 0.519 | |   
## ----------------|-----------|-----------|-----------|  
##   
##

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_#

citation()

##   
## To cite R in publications use:  
##   
## R Core Team (2021). R: A language and environment for statistical  
## computing. R Foundation for Statistical Computing, Vienna, Austria.  
## URL https://www.R-project.org/.  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {R: A Language and Environment for Statistical Computing},  
## author = {{R Core Team}},  
## organization = {R Foundation for Statistical Computing},  
## address = {Vienna, Austria},  
## year = {2021},  
## url = {https://www.R-project.org/},  
## }  
##   
## We have invested a lot of time and effort in creating R, please cite it  
## when using it for data analysis. See also 'citation("pkgname")' for  
## citing R packages.

citation("tidyverse")

##   
## Wickham et al., (2019). Welcome to the tidyverse. Journal of Open  
## Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Article{,  
## title = {Welcome to the {tidyverse}},  
## author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henry and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bache and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani},  
## year = {2019},  
## journal = {Journal of Open Source Software},  
## volume = {4},  
## number = {43},  
## pages = {1686},  
## doi = {10.21105/joss.01686},  
## }

citation("dplyr")

##   
## To cite package 'dplyr' in publications use:  
##   
## Hadley Wickham, Romain François, Lionel Henry and Kirill Müller  
## (2021). dplyr: A Grammar of Data Manipulation. R package version  
## 1.0.7. https://CRAN.R-project.org/package=dplyr  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {dplyr: A Grammar of Data Manipulation},  
## author = {Hadley Wickham and Romain François and Lionel Henry and Kirill Müller},  
## year = {2021},  
## note = {R package version 1.0.7},  
## url = {https://CRAN.R-project.org/package=dplyr},  
## }

citation("stringr")

##   
## To cite package 'stringr' in publications use:  
##   
## Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for  
## Common String Operations. R package version 1.4.0.  
## https://CRAN.R-project.org/package=stringr  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {stringr: Simple, Consistent Wrappers for Common String Operations},  
## author = {Hadley Wickham},  
## year = {2019},  
## note = {R package version 1.4.0},  
## url = {https://CRAN.R-project.org/package=stringr},  
## }

citation("gmodels")

##   
## To cite package 'gmodels' in publications use:  
##   
## Gregory R. Warnes, Ben Bolker, Thomas Lumley, Randall C Johnson.  
## Contributions from Randall C. Johnson are Copyright SAIC-Frederick,  
## Inc. Funded by the Intramural Research Program, of the NIH, National  
## Cancer Institute and Center for Cancer Research under NCI Contract  
## NO1-CO-12400. (2018). gmodels: Various R Programming Tools for Model  
## Fitting. R package version 2.18.1.  
## https://CRAN.R-project.org/package=gmodels  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {gmodels: Various R Programming Tools for Model Fitting},  
## author = {Gregory R. Warnes and Ben Bolker and Thomas Lumley and Randall C Johnson. Contributions from Randall C. Johnson are Copyright SAIC-Frederick and Inc. Funded by the Intramural Research Program and of the NIH and National Cancer Institute and Center for Cancer Research under NCI Contract NO1-CO-12400.},  
## year = {2018},  
## note = {R package version 2.18.1},  
## url = {https://CRAN.R-project.org/package=gmodels},  
## }  
##   
## ATTENTION: This citation information has been auto-generated from the  
## package DESCRIPTION file and may need manual editing, see  
## 'help("citation")'.