Assessment 1 – Data set

A1844790 - Dylan Rohan

04/11/2021

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  
ashes

## # A tibble: 27 x 13  
## batter team role `Test 1, Innings~ `Test 1, Inning~ `Test 2, Inning~  
## <chr> <chr> <chr> <chr> <chr> <chr>   
## 1 Ali England allro~ Batting at numbe~ Batting at numb~ Batting at numb~  
## 2 Anderson English bowl Batting at numbe~ Batting at numb~ Batting at numb~  
## 3 Bairstow England wicke~ Batting at numbe~ Batting at numb~ Batting at numb~  
## 4 Ball England bowl Batting at numbe~ Batting at numb~ Batting at numb~  
## 5 Bancroft Australia bat Batting at numbe~ Batting at numb~ Batting at numb~  
## 6 Bird Australia bowl Batting at numbe~ Batting at numb~ Batting at numb~  
## 7 Broad England bowler Batting at numbe~ Batting at numb~ Batting at numb~  
## 8 Cook England bat Batting at numbe~ Batting at numb~ Batting at numb~  
## 9 Crane England bowl Batting at numbe~ Batting at numb~ Batting at numb~  
## 10 Cummins Australia bowl Batting at numbe~ Batting at numb~ Batting at numb~  
## # ... with 17 more rows, and 7 more variables: Test 2, Innings 2 <chr>,  
## # Test 3, Innings 1 <chr>, Test 3, Innings 2 <chr>, Test 4, Innings 1 <chr>,  
## # Test 4, Innings 2 <chr>, Test 5, Innings 1 <chr>, Test 5, Innings 2 <chr>

#Checking the table, currently a tibble of 27 x 13  
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))

| 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. |