hw4.1_estes

Andrew Estes

11/14/2021

Question 1

The package dplyr in tidyverse performs query functions much like with SQL RDBMS systems. Read into tibbles the files baseball01.csv and baseball02.csv. Make three new "rectangle" datasets that demonstrate all of the following: inner join, left join, and full join. In your RMarkdown file, discuss how the three resulting tibbles are different.

When you have finished, upload a pdf of your RMarkdown document.

```
#this code shows the libraries/packages that were used for this exercise
suppressPackageStartupMessages(library(tidyverse))
suppressPackageStartupMessages(library(jsonlite))
suppressPackageStartupMessages(library(ggplot2))
suppressPackageStartupMessages(library(knitr))
suppressPackageStartupMessages(library(readxl))

#this code downloads the data into a tibble/dataframe
tib1 <- read_excel("C:/Users/andre/OneDrive/Desktop/PDAT 611/baseball01.xlsx")
tib2 <- read_excel("C:/Users/andre/OneDrive/Desktop/PDAT 611/baseball02.xlsx")

#this code creates the joins, altering the data as requested
inner <- inner_join(tib1, tib2, by = "Name")
left <- left_join(tib1, tib2, by = "Name")
full <- full_join(tib1, tib2, by = "Name")

#finally, we can view what happens with an inner joint
kable(inner)
```

The inner join shows the only two players who have an observation for each column header.

```
#and what the left join does
kable(left)
```

The left join adds "color" and "position" columns to the existing tibble from the baseball01 file.

```
#finally, the full join
kable(full)
```

The full join adds all datapoints across all datasets. This results in additional two rows in the player variable not originally shown in the first tibble because they had no observations under the three columns in tibble 1 (hits, dingers, rbi').