

Lesson 3 - Intro to R tidyverse

Kevin Cummiskey

January 14, 2020

Review

Last class, we discussed the five tidyverse verbs (filter, select, mutate, summarize, arrange). Use the Pitching database in the Lahman package to generate the table (Pitchers with 350 wins or more) on page 65 of the Marchi text. Note the Pitching database contains one row for each pitcher for each season or stint (if the pitcher played for more than one team in a season). For this problem, you can use playerID instead of full name in the table. Write your code below.

```
Pitching %>%
  group_by(playerID) %>%
  summarize(W = sum(W, na.rm = TRUE),
            L = sum(L, na.rm = TRUE),
            SO = sum(SO, na.rm = TRUE),
            BB = sum(BB, na.rm = TRUE)) %>%
  filter(W > 349) -> Wins.350
```

How could you add the player's full name? (hint: use the Master database.)

```
Wins.350 %>%
  left_join(Master %>% select(playerID, nameFirst, nameLast),
            by = "playerID") %>%
  mutate(player = paste(nameFirst, nameLast)) %>%
  select(player, W, L, SO, BB) -> Wins.350
```

Pitching

There has been a lot of discussion recently about the evolving use of relief pitchers. Historically, starting pitchers were expected to go deep into the game and they frequently pitched the whole game. In the last several years, teams have begun employing relief pitchers in highly specialized roles. For a nice discussion on this, see Nate Silver's article "Relievers Have Broken Baseball. We Have A Plan To Fix It."

(<https://fivethirtyeight.com/features/relievers-have-broken-baseball-we-have-a-plan-to-fix-it/>)

1. What are some questions we would want to investigate to better understand how the use of relief pitchers has changed over time?

Complete Games

For each season from 1970 to 2018, calculate the fraction of games each year that were complete games (hint: Teams database). Discuss two interesting results.

Does the trend differ between the American and National Leagues?

Batters Faced

We might also want to look at the typical number of batters a pitcher faces. For each season from 1970 to 2018, calculate the average number of batters faced per game (hint: Pitching database). Discuss two interesting results.

Ideally, we would want to perform the analysis in the last question separately for relief and starting pitchers. However, the Pitching database does not identify pitchers as starters or relievers. Propose a way to identify starting pitchers in the Pitching database.

Using your method, identify pitchers as relievers or starters and repeat your 'batter's faced' analysis separately for relievers and starters.