

Lab 3

Lab Overview

For this question, a subset of the tables contained in the History of Baseball database are available. Additional details are available here: <https://www.kaggle.com/seanlahman/the-history-of-baseball>. The following tables will be used for these questions:

- player
- all_star
- salary

```
library(readr)
library(dplyr)
library(knitr)
library(tidyr)
player <- read_csv("http://math.montana.edu/ahoegh/teaching/stat408/datasets/player.csv")
all_star <- read_csv("http://math.montana.edu/ahoegh/teaching/stat408/datasets/all_star.csv")
salary <- read_csv("http://math.montana.edu/ahoegh/teaching/stat408/datasets/salary.csv")
```

1. (2 points)

How many players were born in Montana.

2. (4 points)

Print a table that contains each player born in Montana. The table should contain the player_id as well as given name and their total salary across all years. If salary is not available (pre-1985), include the player but have an NA for salary.

3. (2 points)

Create a thin dataset for that contains the yearly salaries of David Ortiz, Derek Jeter, and Troy Tulowitzki.

4. (4 points)

Create a wide dataset for that contains the salaries of David Ortiz, Derek Jeter, and Troy Tulowitzki.

5. (4 points)

Which player(s) made the most appearances as an all star representing the National League (NL)?