Homework 7

Carmen Canedo

27 November 2020

Loading Libraries

Keys

Primary keys uniquely identify an observation within its own table. *Foreign keys* uniquely identify an observation in another table.

To be sure the identification is unique, count the primary keys and see if any are greater than 1.

There is not a primary key in the flights tibble because the values repeat. We need to add a *surrogate key* to flights.

Exercise 1 Identify the primary keys in three datasets.

Be sure to show that you have the primary key by showing there are no duplicate entries.

a. Lahman Batting Primary Key

```
## [1] playerID stint    yearID   n
## <0 rows> (or 0-length row.names)
```

The primary key for `Lahman::Batting` is `playerID`, `stint`, and `yearID`.

b. Babynames Primary key

The primary key for `babynames` is `name`, `year`, and `sex`.

c. Atmos Primary Key

```
## # A tibble: 0 x 5
## # ... with 5 variables: lat <dbl>, long <dbl>, year <int>, month <int>, n <int>
```

the primary key for `atmos` is `lat`, `long`, `year`, `month`.

Exercise 2 What is the relationship between the Batting, Master, and Salaries tables in the Lahman package? What are the keys for ech dataset and how do they relate to each other?

```
## [1] playerID n
## <0 rows> (or 0-length row.names)
```

```
## [1] playerID teamID   yearID   n
## <0 rows> (or 0-length row.names)
```

As we know from above, the primary key for `Batting` is `playerID`, `stint`, and `yearID`. The primary key for `Master` is `playerID`, and for `Salaries` it is `playerID`, `teamID`, and `yearID`.

`playerID` in `Master` has a 1 to many relationship in Salaries and Batting.