STAT 33A/B Lab Workbook Wk 14

Won Shil Park (3033452021)

Apr 29, 2021

This workbook is due **Apr 29, 2021** by 9:00am PT, or if you attend lab the workbook is due **Apr 30, 2021** by 11:59pm.

- Knit and submit the generated PDF file on Gradescope.

This lab continues with the boat example.

```
load(url("http://www.stat.berkeley.edu/users/nolan/data/toyboat.rda"))
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

There are four data frames: sailor1, sailor2, boat, reservations. They are small enough that you can examine them by printing the entire data frame.

Exercise 1

Create a data frame of the boats rented by sailors who have high ratings (over 8). There should be one record per boat. The final data frame should include only the boat id and the number of times it was rented by a sailor with a high rating. (There should be 3 records in the output.)

```
sailors = union(sailor1, sailor2)
bestRenters = filter(sailors, rating > 8)

bestRenters = left_join(bestRenters, reservations, by = "sid")
bestRenters = group_by(bestRenters, bid)
bestRenters = summarize(bestRenters, n = n())

bestRenters
```

```
## # A tibble: 3 x 2
##     bid     n
##   <dbl> <int>
## 1   101     2
## 2   102     6
## 3   104     2
```

Exercise 2.

Create a data frame that contains one record for each boat reservaton. In addition to the sailor id, boat id, and day of the reservation, the data frame should contain the following information:

- sailor name
- boat name

```
reservations_with_sailors = right_join(select(sailors, sid, sname), reservations, by = "sid")

boats.sailors = left_join(reservations_with_sailors, select(boat, bid, bname), by = "bid")

boats.sailors
```

```
##    sid  sname bid day     bname
## 1   22 dustin 101  16 Interlake
## 2   22 dustin 104  18    Marine
## 3   22 dustin 101  21 Interlake
## 4   22 dustin 104  23    Marine
## 5   31 lubber 101  19 Interlake
## 6   31 lubber 101  20 Interlake
## 7   31 lubber 101  24 Interlake
## 8   58  rusty 102  17 Interlake
## 9   58  rusty 102  18 Interlake
## 10  58  rusty 104  20    Marine
## 11  58  rusty 102  21 Interlake
## 12  58  rusty 102  22 Interlake
## 13  58  rusty 102  23 Interlake
## 14  58  rusty 104  24    Marine
## 15  58  rusty 102  25 Interlake
## 16  28  yuppy 101  18 Interlake
## 17  28  yuppy 101  22 Interlake
## 18  44  guppy 103  18   Clipper
## 19  44  guppy 103  21   Clipper
## 20  44  guppy 103  23   Clipper
```

The resulting data frame should have 20 rows. Four of these rows appear below:

```
# sid   sname   bid day bname
# 31    lubber  101 19  Interlake
# 31    lubber  101 20  Interlake
# 58    rusty   104 20  Marine
# 58    rusty   102 21  Interlake
```

Exercise 3.

Create a data frame with columns: boat id, boat name, and counts of the number of times each boat has been rented. (No need to worry about boats that were not rented).

```
rentalCount = left_join(reservations, boat, by = "bid")
rentalCount = group_by(rentalCount, bid)
# first takes the first value from the vector
rentalCount = summarize(rentalCount, bname = first(bname), n = n())

rentalCount
```

```
## # A tibble: 4 x 3
##     bid bname         n
##   <dbl> <chr>     <int>
## 1   101 Interlake     7
## 2   102 Interlake     6
## 3   103 Clipper       3
## 4   104 Marine        4
```