

STAT 33A Lec Workbook Wk 14

Won Shil Park (3033452021)

Apr 26, 2021

This workbook is due **Apr 26, 2021** by 11:59pm PT.

- Knit and submit the generated PDF file on Gradescope.

```
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

Combine the two sailor data frames (i.e., `sailor1` and `sailor2`) into one data frame. Each sailor should appear only once.

```
union(sailor1, sailor2)
```

```
##   sid  sname rating age
## 1  22  dustin      7  45
## 2  31  lubber      8  55
## 3  58  rusty     10  35
## 4  28  yuppy      9  35
## 5  44  guppy      5  35
```

```
#full_join(sailor1, sailor2, by = "sid") NOT TO DO
```

Exercise 2.

Create a data frame that consists of the names of the sailors that are in `sailor1` and not in `sailor2`.

```
setdiff(sailor1, sailor2)
```

```
##   sid  sname rating age
## 1  22  dustin      7  45
```

```
setdiff(sailor2, sailor1)
```

```
##   sid sname rating age
## 1  28  yuppy      9  35
## 2  44  guppy      5  35
```

Exercise 3.

Create a data frame of the reservations of boat 103 that includes the sailor's name and id, as well as the boat id and the day of the reservation. (You should find that only one particular sailor reserved this boat several times.)

```
sailor_all = union(sailor1, sailor2)

sailor_all_name = select(sailor_all, sid, sname)

boat103R = filter(reservations, bid == 103)

left_join(boat103R, sailor_all_name, by = "sid")
```

```
##   sid bid day sname
## 1  44 103  18  guppy
## 2  44 103  21  guppy
## 3  44 103  23  guppy
```

```
s_a_n = select(union(sailor1, sailor2), sid, sname)
left_join(filter(reservations, bid == 103), s_a_n, by = "sid")
```

```
##   sid bid day sname
## 1  44 103  18  guppy
## 2  44 103  21  guppy
## 3  44 103  23  guppy
```

Exercise 4

Create a data frame that contains the sailor name, boat name, and the number of times the sailor has rented the boat. That is, there should be one record for each sailor - boat combination. Don't worry about the cases where the sailor didn't rent a particular boat. There should be 7 rows in the final data frame.

```
sailor_all_name = select(sailor_all, sid, sname)
boat_name = select(boat, bid, bname)

fav_rentals = left_join(reservations, boat_name, by = "bid")
fav_rentals = left_join(fav_rentals, sailor_all_name, by = "sid")
head(fav_rentals)
```

```
##   sid bid day    bname  sname
## 1  22 101  16 Interlake dustin
## 2  58 102  17 Interlake  rusty
## 3  58 102  18 Interlake  rusty
## 4  28 101  18 Interlake  yuppy
## 5  44 103  18  Clipper  guppy
## 6  22 104  18   Marine dustin
```

```
fav_rentals_grp = group_by(fav_rentals, bid, sid)

summarize(fav_rentals_grp, ct = n(), boat = first(bname), sailor = last(sname))
```

'summarise()' has grouped output by 'bid'. You can override using the '.groups' argument.

```
## # A tibble: 7 x 5
## # Groups:   bid [4]
##   bid   sid   ct boat      sailor
##   <dbl> <dbl> <int> <chr>    <chr>
## 1   101    22     2 Interlake dustin
## 2   101    28     2 Interlake yuppy
## 3   101    31     3 Interlake lubber
## 4   102    58     6 Interlake rusty
## 5   103    44     3 Clipper  guppy
## 6   104    22     2 Marine   dustin
## 7   104    58     2 Marine   rusty
```

Four of the rows are:

```
# 101  22  2  Interlake  dustin
# 101  28  2  Interlake  yuppy
# 103  44  3  Clipper    guppy
# 104  58  2  Marine     rusty
```