R-intro - Session 4 (part 1 - dplyr) exercise solutions

João Gonçalves

31 de Julho de 2018

Contents

Objectives of session 4 (*dplyr*)

- A brief tour to the `tidyverse`
- Data manipulation with the `dplyr` package
  - Main `dplyr` verbs
  - The pipe `%>%` operator

The *dplyr* package

```
library(dplyr)
```

Selecting columns using `select()`

QUICK EXERCISE 1

Using the `select` verb from the `dplyr` package and the `airquality` dataset, answer to the following questions:

1a.

```
select(airquality, 1:3)
```

```
##     Ozone Solar.R Wind
## 1      41     190  7.4
## 2      36     118  8.0
```

```
## 3      12     149 12.6
##  [ reached getOption("max.print") -- omitted 150 rows ]
```

1b.

```
select(airquality, -Ozone, -Temp)
```

```
##     Solar.R Wind Month Day
## 1       190  7.4     5   1
## 2       118  8.0     5   2
##  [ reached getOption("max.print") -- omitted 151 rows ]
```

1c.

```
select(airquality, Solar.R:Day)
```

```
##     Solar.R Wind Temp Month Day
## 1       190  7.4   67     5   1
## 2       118  8.0   72     5   2
##  [ reached getOption("max.print") -- omitted 151 rows ]
```

Selecting rows with `filter()`

QUICK EXERCISE 2

2a.

```
filter(iris, Petal.Length > 5)
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1           6.0         2.7          5.1         1.6 versicolor
## 2           6.3         3.3          6.0         2.5  virginica
##  [ reached getOption("max.print") -- omitted 40 rows ]
```

2b.

```
filter(iris, Petal.Length >= 3 & (Petal.Width < 1.2 | Sepal.Width < 2.8))
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1           5.5         2.3          4.0         1.3 versicolor
## 2           4.9         2.4          3.3         1.0 versicolor
##  [ reached getOption("max.print") -- omitted 30 rows ]
```

2c.

```
filter(iris, Species == "versicolor")
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1           7.0         3.2          4.7         1.4 versicolor
## 2           6.4         3.2          4.5         1.5 versicolor
##  [ reached getOption("max.print") -- omitted 48 rows ]
```

```
# OR
filter(iris, Species %in% "versicolor")
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1           7.0         3.2          4.7         1.4 versicolor
## 2           6.4         3.2          4.5         1.5 versicolor
##  [ reached getOption("max.print") -- omitted 48 rows ]
```

The pipe operator: %>%

Sorting rows with arrange()

QUICK EXERCISE 3

Using the arrange verb from the dplyr package and the iris dataset, answer to the following question:

3a.

```
iris %>% arrange(desc(Petal.Length), Sepal.Length)
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1            7.7         2.6          6.9         2.3  virginica
## 2            7.7         3.8          6.7         2.2  virginica
##  [ reached getOption("max.print") -- omitted 148 rows ]
```

Adding new columns with mutate()

QUICK EXERCISE 4

4a.

```
trees %>% mutate(Girth_cm = Girth * 2.54,
                 Height_mt = Height / 3.2808399)
```

```
##    Girth Height Volume Girth_cm Height_mt
## 1    8.3     70   10.3   21.082   21.3360
## 2    8.6     65   10.3   21.844   19.8120
##  [ reached getOption("max.print") -- omitted 29 rows ]
```

4b.

```
trees %>%
  mutate(hvr = Height / Volume) %>%
  filter(hvr >= 3, hvr < 5) %>%
  nrow
```

```
## [1] 14
```

Creating data summaries using summarise() and group_by()

QUICK EXERCISE 5

5a.

```
PlantGrowth %>%
  group_by(group) %>%
  summarise(min = min(weight),
            max = max(weight),
            q25 = quantile(weight, probs=0.25),
            q50 = quantile(weight, probs=0.50),
            q75 = quantile(weight, probs=0.75),
            Nobs   = n())
```

```
## # A tibble: 3 x 7
##   group   min   max   q25   q50   q75  Nobs
##   <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <int>
```

```
## 1 ctrl   4.17  6.11  4.55  5.15  5.29    10
## 2 trt1   3.59  6.03  4.21  4.55  4.87    10
## 3 trt2   4.92  6.31  5.27  5.44  5.74    10
```