Examine, Subset, and Filter Data: Week 2 Practice

# Examine, Subset and Filter Data

### Week 2 Practice

We’ll use the built-in dataset ToothGrowth for this practice.

## Examine Data

Look at the head() of ToothGrowth

head(ToothGrowth)

## len supp dose  
## 1 4.2 VC 0.5  
## 2 11.5 VC 0.5  
## 3 7.3 VC 0.5  
## 4 5.8 VC 0.5  
## 5 6.4 VC 0.5  
## 6 10.0 VC 0.5

And the tail()

tail(ToothGrowth)

## len supp dose  
## 55 24.8 OJ 2  
## 56 30.9 OJ 2  
## 57 26.4 OJ 2  
## 58 27.3 OJ 2  
## 59 29.4 OJ 2  
## 60 23.0 OJ 2

What are the dimensions of ToothGrowth?

dim(ToothGrowth)

## [1] 60 3

What is ToothGrowth’s structure?

str(ToothGrowth)

## 'data.frame': 60 obs. of 3 variables:  
## $ len : num 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...  
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...  
## $ dose: num 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...

## Subset Data

Select the first column only using square brackets.

ToothGrowth[,1]

## [1] 4.2 11.5 7.3 5.8 6.4 10.0 11.2 11.2 5.2 7.0 16.5 16.5 15.2 17.3  
## [15] 22.5 17.3 13.6 14.5 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5  
## [29] 23.3 29.5 15.2 21.5 17.6 9.7 14.5 10.0 8.2 9.4 16.5 9.7 19.7 23.3  
## [43] 23.6 26.4 20.0 25.2 25.8 21.2 14.5 27.3 25.5 26.4 22.4 24.5 24.8 30.9  
## [57] 26.4 27.3 29.4 23.0

Select the 42nd row using square brackets.

ToothGrowth[42,]

## len supp dose  
## 42 23.3 OJ 1

Select rows 20 to 40 and columns 1 and 3 using square brackets.

ToothGrowth[20:40, c(1,3)]

## len dose  
## 20 15.5 1.0  
## 21 23.6 2.0  
## 22 18.5 2.0  
## 23 33.9 2.0  
## 24 25.5 2.0  
## 25 26.4 2.0  
## 26 32.5 2.0  
## 27 26.7 2.0  
## 28 21.5 2.0  
## 29 23.3 2.0  
## 30 29.5 2.0  
## 31 15.2 0.5  
## 32 21.5 0.5  
## 33 17.6 0.5  
## 34 9.7 0.5  
## 35 14.5 0.5  
## 36 10.0 0.5  
## 37 8.2 0.5  
## 38 9.4 0.5  
## 39 16.5 0.5  
## 40 9.7 0.5

Select the supp column using a dollar sign

ToothGrowth$supp

## [1] VC VC VC VC VC VC VC VC VC VC VC VC VC VC VC VC VC VC VC VC VC VC VC  
## [24] VC VC VC VC VC VC VC OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ  
## [47] OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ OJ  
## Levels: OJ VC

## Filter/Select

Filter ToothGrowth for a dose greater than 1

filter(ToothGrowth, dose > 1)

## len supp dose  
## 1 23.6 VC 2  
## 2 18.5 VC 2  
## 3 33.9 VC 2  
## 4 25.5 VC 2  
## 5 26.4 VC 2  
## 6 32.5 VC 2  
## 7 26.7 VC 2  
## 8 21.5 VC 2  
## 9 23.3 VC 2  
## 10 29.5 VC 2  
## 11 25.5 OJ 2  
## 12 26.4 OJ 2  
## 13 22.4 OJ 2  
## 14 24.5 OJ 2  
## 15 24.8 OJ 2  
## 16 30.9 OJ 2  
## 17 26.4 OJ 2  
## 18 27.3 OJ 2  
## 19 29.4 OJ 2  
## 20 23.0 OJ 2

Filter ToothGrowth for a supplements that are NOT “VC” and length less than or equal to ten

filter(ToothGrowth, supp != 'VC', len < 10)

## len supp dose  
## 1 9.7 OJ 0.5  
## 2 8.2 OJ 0.5  
## 3 9.4 OJ 0.5  
## 4 9.7 OJ 0.5

Select all the columns, BUT the dose columns

select(ToothGrowth, len, supp)

## len supp  
## 1 4.2 VC  
## 2 11.5 VC  
## 3 7.3 VC  
## 4 5.8 VC  
## 5 6.4 VC  
## 6 10.0 VC  
## 7 11.2 VC  
## 8 11.2 VC  
## 9 5.2 VC  
## 10 7.0 VC  
## 11 16.5 VC  
## 12 16.5 VC  
## 13 15.2 VC  
## 14 17.3 VC  
## 15 22.5 VC  
## 16 17.3 VC  
## 17 13.6 VC  
## 18 14.5 VC  
## 19 18.8 VC  
## 20 15.5 VC  
## 21 23.6 VC  
## 22 18.5 VC  
## 23 33.9 VC  
## 24 25.5 VC  
## 25 26.4 VC  
## 26 32.5 VC  
## 27 26.7 VC  
## 28 21.5 VC  
## 29 23.3 VC  
## 30 29.5 VC  
## 31 15.2 OJ  
## 32 21.5 OJ  
## 33 17.6 OJ  
## 34 9.7 OJ  
## 35 14.5 OJ  
## 36 10.0 OJ  
## 37 8.2 OJ  
## 38 9.4 OJ  
## 39 16.5 OJ  
## 40 9.7 OJ  
## 41 19.7 OJ  
## 42 23.3 OJ  
## 43 23.6 OJ  
## 44 26.4 OJ  
## 45 20.0 OJ  
## 46 25.2 OJ  
## 47 25.8 OJ  
## 48 21.2 OJ  
## 49 14.5 OJ  
## 50 27.3 OJ  
## 51 25.5 OJ  
## 52 26.4 OJ  
## 53 22.4 OJ  
## 54 24.5 OJ  
## 55 24.8 OJ  
## 56 30.9 OJ  
## 57 26.4 OJ  
## 58 27.3 OJ  
## 59 29.4 OJ  
## 60 23.0 OJ

### OR  
  
select(ToothGrowth, -dose)

## len supp  
## 1 4.2 VC  
## 2 11.5 VC  
## 3 7.3 VC  
## 4 5.8 VC  
## 5 6.4 VC  
## 6 10.0 VC  
## 7 11.2 VC  
## 8 11.2 VC  
## 9 5.2 VC  
## 10 7.0 VC  
## 11 16.5 VC  
## 12 16.5 VC  
## 13 15.2 VC  
## 14 17.3 VC  
## 15 22.5 VC  
## 16 17.3 VC  
## 17 13.6 VC  
## 18 14.5 VC  
## 19 18.8 VC  
## 20 15.5 VC  
## 21 23.6 VC  
## 22 18.5 VC  
## 23 33.9 VC  
## 24 25.5 VC  
## 25 26.4 VC  
## 26 32.5 VC  
## 27 26.7 VC  
## 28 21.5 VC  
## 29 23.3 VC  
## 30 29.5 VC  
## 31 15.2 OJ  
## 32 21.5 OJ  
## 33 17.6 OJ  
## 34 9.7 OJ  
## 35 14.5 OJ  
## 36 10.0 OJ  
## 37 8.2 OJ  
## 38 9.4 OJ  
## 39 16.5 OJ  
## 40 9.7 OJ  
## 41 19.7 OJ  
## 42 23.3 OJ  
## 43 23.6 OJ  
## 44 26.4 OJ  
## 45 20.0 OJ  
## 46 25.2 OJ  
## 47 25.8 OJ  
## 48 21.2 OJ  
## 49 14.5 OJ  
## 50 27.3 OJ  
## 51 25.5 OJ  
## 52 26.4 OJ  
## 53 22.4 OJ  
## 54 24.5 OJ  
## 55 24.8 OJ  
## 56 30.9 OJ  
## 57 26.4 OJ  
## 58 27.3 OJ  
## 59 29.4 OJ  
## 60 23.0 OJ

## Anything Goes Challenge Questions

Return only VC supplements, with length less than or equal to 10 and greater than or equal to 25.

ToothGrowth[(ToothGrowth$supp == 'VC' & (ToothGrowth$len <= 10 | ToothGrowth$len >= 25)),]

## len supp dose  
## 1 4.2 VC 0.5  
## 3 7.3 VC 0.5  
## 4 5.8 VC 0.5  
## 5 6.4 VC 0.5  
## 6 10.0 VC 0.5  
## 9 5.2 VC 0.5  
## 10 7.0 VC 0.5  
## 23 33.9 VC 2.0  
## 24 25.5 VC 2.0  
## 25 26.4 VC 2.0  
## 26 32.5 VC 2.0  
## 27 26.7 VC 2.0  
## 30 29.5 VC 2.0

### OR  
  
filter(ToothGrowth, supp == 'VC', len <= 10 | len >= 25)

## len supp dose  
## 1 4.2 VC 0.5  
## 2 7.3 VC 0.5  
## 3 5.8 VC 0.5  
## 4 6.4 VC 0.5  
## 5 10.0 VC 0.5  
## 6 5.2 VC 0.5  
## 7 7.0 VC 0.5  
## 8 33.9 VC 2.0  
## 9 25.5 VC 2.0  
## 10 26.4 VC 2.0  
## 11 32.5 VC 2.0  
## 12 26.7 VC 2.0  
## 13 29.5 VC 2.0

Return length less than 10 or dose equals 2 and drop the supp column

ToothGrowth[ToothGrowth$len < 10 | ToothGrowth$dose == 2, c(1, 3)]

## len dose  
## 1 4.2 0.5  
## 3 7.3 0.5  
## 4 5.8 0.5  
## 5 6.4 0.5  
## 9 5.2 0.5  
## 10 7.0 0.5  
## 21 23.6 2.0  
## 22 18.5 2.0  
## 23 33.9 2.0  
## 24 25.5 2.0  
## 25 26.4 2.0  
## 26 32.5 2.0  
## 27 26.7 2.0  
## 28 21.5 2.0  
## 29 23.3 2.0  
## 30 29.5 2.0  
## 34 9.7 0.5  
## 37 8.2 0.5  
## 38 9.4 0.5  
## 40 9.7 0.5  
## 51 25.5 2.0  
## 52 26.4 2.0  
## 53 22.4 2.0  
## 54 24.5 2.0  
## 55 24.8 2.0  
## 56 30.9 2.0  
## 57 26.4 2.0  
## 58 27.3 2.0  
## 59 29.4 2.0  
## 60 23.0 2.0

### OR  
  
select(filter(ToothGrowth, len < 10 | dose == 2), -supp)

## len dose  
## 1 4.2 0.5  
## 2 7.3 0.5  
## 3 5.8 0.5  
## 4 6.4 0.5  
## 5 5.2 0.5  
## 6 7.0 0.5  
## 7 23.6 2.0  
## 8 18.5 2.0  
## 9 33.9 2.0  
## 10 25.5 2.0  
## 11 26.4 2.0  
## 12 32.5 2.0  
## 13 26.7 2.0  
## 14 21.5 2.0  
## 15 23.3 2.0  
## 16 29.5 2.0  
## 17 9.7 0.5  
## 18 8.2 0.5  
## 19 9.4 0.5  
## 20 9.7 0.5  
## 21 25.5 2.0  
## 22 26.4 2.0  
## 23 22.4 2.0  
## 24 24.5 2.0  
## 25 24.8 2.0  
## 26 30.9 2.0  
## 27 26.4 2.0  
## 28 27.3 2.0  
## 29 29.4 2.0  
## 30 23.0 2.0

### OR  
  
filter(ToothGrowth, len < 10 | dose == 2) %>% select(-supp)

## len dose  
## 1 4.2 0.5  
## 2 7.3 0.5  
## 3 5.8 0.5  
## 4 6.4 0.5  
## 5 5.2 0.5  
## 6 7.0 0.5  
## 7 23.6 2.0  
## 8 18.5 2.0  
## 9 33.9 2.0  
## 10 25.5 2.0  
## 11 26.4 2.0  
## 12 32.5 2.0  
## 13 26.7 2.0  
## 14 21.5 2.0  
## 15 23.3 2.0  
## 16 29.5 2.0  
## 17 9.7 0.5  
## 18 8.2 0.5  
## 19 9.4 0.5  
## 20 9.7 0.5  
## 21 25.5 2.0  
## 22 26.4 2.0  
## 23 22.4 2.0  
## 24 24.5 2.0  
## 25 24.8 2.0  
## 26 30.9 2.0  
## 27 26.4 2.0  
## 28 27.3 2.0  
## 29 29.4 2.0  
## 30 23.0 2.0