Cereales

Diana Marcela Beltrán Pedroza

2025-09-30

```
# install.packages("readxl")
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.4.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.3
```

```
##
## Adjuntando el paquete: 'dplyr'
```

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

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

```
cereal <- read_excel("cereal.xlsx")
head(cereal)
```

```
## # A tibble: 6 x 10
##   name          company serving calories   fat sodium carbs fiber sugars protein
##   <chr>         <chr>     <dbl>    <dbl> <dbl>  <dbl> <dbl> <dbl>  <dbl>   <dbl>
## 1 AppleJacks    K          1         117   0.6    143    27   0.5   15       1
## 2 Boo Berry     G          1         118   0.8    211    27   0.1   14       1
## 3 Cap'n Crunch  Q          0.75      144   2.1    269    31   1.1   16       1.3
## 4 Cinnamon Toa~ G          0.75      169   4.4    408    32   1.7   13.3     2.7
## 5 Cocoa Blasts  Q          1         130   1.2    135    29   0.8   16       1
## 6 Cocoa Puffs   G          1         117   1      171    26   0.8   14       1
```

Q1 Extract rows 1 to 10 from the cereal data frame. (Extraiga las filas 1 a 10 del marco de datos cereal.)

```
cereal[1:10, ]
```

```
## # A tibble: 10 x 10
##    name         company serving calories   fat sodium carbs fiber sugars protein
##    <chr>        <chr>     <dbl>    <dbl> <dbl>  <dbl> <dbl> <dbl>  <dbl>   <dbl>
##  1 AppleJacks   K          1         117   0.6    143    27   0.5   15       1
##  2 Boo Berry    G          1         118   0.8    211    27   0.1   14       1
##  3 Cap'n Crunch Q          0.75      144   2.1    269    31   1.1   16       1.3
##  4 Cinnamon To~ G          0.75      169   4.4    408    32   1.7   13.3     2.7
##  5 Cocoa Blasts Q          1         130   1.2    135    29   0.8   16       1
##  6 Cocoa Puffs  G          1         117   1      171    26   0.8   14       1
##  7 Cookie Crisp G          1         117   0.9    178    26   0.5   13       1
##  8 Corn Flakes  K          1         101   0.1    202    24   0.8    3       2
##  9 Corn Pops    K          1         117   0.2    120    28   0.3   15       1
## 10 Crispix      K          1         113   0.3    229    26   0.1    3       2
```

Q2

```
# data frane

Kelloggs <- cereal[cereal$company == "K",]
head(Kelloggs)
```

```
## # A tibble: 6 x 10
##   name          company serving calories   fat sodium carbs fiber sugars protein
##   <chr>         <chr>     <dbl>    <dbl> <dbl>  <dbl> <dbl> <dbl>  <dbl>   <dbl>
## 1 AppleJacks    K             1      117   0.6    143    27   0.5     15       1
## 2 Corn Flakes   K             1      101   0.1    202    24   0.8      3       2
## 3 Corn Pops     K             1      117   0.2    120    28   0.3     15       1
## 4 Crispix       K             1      113   0.3    229    26   0.1      3       2
## 5 Froot Loops   K             1      118   0.9    150    26   0.8     12       2
## 6 Frosted Mini~ K             1      175   0.8      5    41   5       10       5
```

Q3

```
cereal[cereal$sugars > 10, "name"]
```

```
## # A tibble: 17 x 1
##    name
##    <chr>
##  1 AppleJacks
##  2 Boo Berry
##  3 Cap'n Crunch
##  4 Cinnamon Toast Crunch
##  5 Cocoa Blasts
##  6 Cocoa Puffs
##  7 Cookie Crisp
##  8 Corn Pops
##  9 Froot Loops
## 10 Golden Grahams
## 11 Honey Nut Clusters
## 12 Honey Nut Heaven
## 13 Lucky Charms
## 14 Raisin Bran
```

```
## 15 Reese's Puffs
## 16 Rice Krispie Treats
## 17 Smart Start
```

Q4

```
mean(cereal$calories)
```

```
## [1] 133.8333
```

Q5

```
subset(cereal, fiber >= 2)
```

```
## # A tibble: 10 x 10
##    name         company serving calories   fat sodium carbs fiber sugars protein
##    <chr>        <chr>     <dbl>    <dbl> <dbl>  <dbl> <dbl> <dbl>  <dbl>   <dbl>
##  1 Crunchy Bran Q          0.75      120   1.3    309    31   6.4    8       1.3
##  2 Frosted Min~ K          1         175   0.8      5    41   5     10       5
##  3 Honey Nut C~ G          1         214   2.7    249    46   2.8   17       4
##  4 Honey Nut H~ Q          1         192   3.7    216    38   3.5   13       4
##  5 Life         Q          0.75      160   1.9    219    33   2.7    8       4
##  6 Multi-Grain~ G          1         108   1.2    201    24   2.8    6       2
##  7 Raisin Bran  K          1         195   1.6    362    47   7.3   20       5
##  8 Smart Start  K          1         182   0.7    275    43   2.8   14       4
##  9 Total        G          0.75      129   0.9    256    31   3.7    6.7     4
## 10 Wheaties     G          1         107   1      218    24   3      4       3
```

New practice

```
EquationCitations <- read.csv("EquationCitations.csv")
head(EquationCitations)
```

```
##   rownames journal            authors volume startpage pages equations
## 1        1   AmNat     Prodohl et al.    151         7    13         0
## 2        2   AmNat           Mauricio    151        20     9         2
## 3        3   AmNat      Thrall et al.    151        29    17        15
## 4        4   AmNat      Watson et al.    151        46    13         7
## 5        5   AmNat     Klein & Nelson    151        59     9         0
## 6        6   AmNat Blackburn & Gaston    151        68    16         2
##   mainequations appequations cites selfcites othercites theocites nontheocites
## 1             0            0    37         4         33         8           22
## 2             2            0   178         5        173        19          150
## 3            15            0    25        10         15         8            6
## 4             7            0   127        11        116        17           98
## 5             0            0    43         8         35         5           28
## 6             2            0    91         7         84        21           61
```

Q1 On average, how many equations per article are published in each journal? (Hint: use group_by() and summarise())

```r
EquationCitations %>%
  group_by(journal)%>%
  summarize(promedio = mean(equations))
```

```
## # A tibble: 3 x 2
##   journal   promedio
##   <chr>        <dbl>
## 1 AmNat         9.04
## 2 Evolution     2.91
## 3 ProcB         2.69
```

Q2 For each journal, calculate the average proportion of theoretical citations relative to total citations. (Hint: create a new column with mutate(theo_ratio = theocites / cites), then group_by() and summarise()) ## Para cada revista, calcule la proporción promedio de citas teóricas respecto al total de citas. (Sugerencia: cree una nueva columna con mutate(theo_ratio = theocites / cites), luego group_by() y summarise())

```r
EquationCitations %>%
  mutate(theo_ratio = theocites / cites) %>%
  group_by(journal) %>%
  summarise(citaspromedio = mean(theo_ratio))
```

```
## # A tibble: 3 x 2
##   journal   citaspromedio
##   <chr>             <dbl>
## 1 AmNat             0.253
## 2 Evolution         0.138
## 3 ProcB             0.176
```