

# R\_Activity\_Assignment\_2

Dahee Ahn

2024-09-06

## R Markdown

```
rm(list=ls())
```

Question 1. Load in the breakfast data set.  
Remember to check the file type so that you can identify the right command for loading in the data.

```
breakfast_df <- read.table("C:/Users/chemk/OneDrive/Desktop/Classes/ENT6707_DataAnalysis/week3/breakfast.txt", header = TRUE)
summary(breakfast_df)
```

```
##      Name      Company      Serving      Calories
## Length:30      Length:30      Min.   :0.7500      Min.   : 80.0
## Class :character Class :character 1st Qu.:0.8125      1st Qu.:113.2
## Mode  :character Mode  :character  Median :1.0000      Median :118.0
##                                     Mean   :0.9683      Mean   :133.8
##                                     3rd Qu.:1.0000      3rd Qu.:160.0
##                                     Max.   :1.5000      Max.   :214.0
##      Fat      Sodium      Carbs      Fiber
## Min.   :0.100      Min.   : 5.0      Min.   :17.00      Min.   :0.000
## 1st Qu.:0.625      1st Qu.:183.8      1st Qu.:25.00      1st Qu.:0.575
## Median :0.950      Median :217.0      Median :27.50      Median :0.950
## Mean   :1.283      Mean   :220.2      Mean   :29.67      Mean   :1.797
## 3rd Qu.:1.525      3rd Qu.:251.2      3rd Qu.:32.75      3rd Qu.:2.800
## Max.   :4.400      Max.   :408.0      Max.   :47.00      Max.   :7.300
##      Sugars      Protein
## Min.   : 1.60      Min.   :1.00
## 1st Qu.: 4.50      1st Qu.:1.30
## Median :12.50      Median :2.00
## Mean   :10.42      Mean   :2.48
## 3rd Qu.:14.53      3rd Qu.:3.75
## Max.   :20.00      Max.   :7.00
```

```
nrow(breakfast_df)
```

```
## [1] 30
```

```
head(breakfast_df)
```

```
##           Name Company Serving Calories Fat Sodium Carbs Fiber Sugars
## 1      AppleJacks      K   1.00    117 0.6   143   27   0.5   15.0
## 2        Boo Berry      G   1.00    118 0.8   211   27   0.1   14.0
## 3    Cap'n Crunch      Q   0.75    144 2.1   269   31   1.1   16.0
## 4 Cinnamon Toast Crunch  G   0.75    169 4.4   408   32   1.7   13.3
## 5      Cocoa Blasts      Q   1.00    130 1.2   135   29   0.8   16.0
## 6      Cocoa Puffs      G   1.00    117 1.0   171   26   0.8   14.0
## Protein
## 1      1.0
## 2      1.0
## 3      1.3
## 4      2.7
## 5      1.0
## 6      1.0
```

```
tail(breakfast_df)
```

```
##           Name Company Serving Calories Fat Sodium Carbs Fiber Sugars
## 25      Rice Chex      G   1.25     94 0.2   234   22   0.1    1.6
## 26 Rice Krispie Treats  K   0.75    160 1.7   252   35   0.0   12.0
## 27    Smart Start      K   1.00    182 0.7   275   43   2.8   14.0
## 28      Special K      K   1.00    117 0.4   224   22   0.8    4.0
## 29          Total      G   0.75    129 0.9   256   31   3.7    6.7
## 30      Wheaties      G   1.00    107 1.0   218   24   3.0    4.0
## Protein
## 25      1.6
## 26      1.3
## 27      4.0
## 28      7.0
## 29      4.0
## 30      3.0
```

```
View(breakfast_df)
```

Question 2. Use a single command to count the number of columns in the breakfast data.

```
breakfast_df$Calories <- as.numeric(breakfast_df$Calories)
str(breakfast_df)
```

```
## 'data.frame': 30 obs. of 10 variables:
## $ Name : chr "AppleJacks" "Boo Berry" "Cap'n Crunch" "Cinnamon Toast Crunch" ...
## $ Company : chr "K" "G" "Q" "G" ...
## $ Serving : num 1 1 0.75 0.75 1 1 1 1 1 1 ...
## $ Calories: num 117 118 144 169 130 117 117 101 117 113 ...
## $ Fat : num 0.6 0.8 2.1 4.4 1.2 1 0.9 0.1 0.2 0.3 ...
## $ Sodium : int 143 211 269 408 135 171 178 202 120 229 ...
## $ Carbs : int 27 27 31 32 29 26 26 24 28 26 ...
## $ Fiber : num 0.5 0.1 1.1 1.7 0.8 0.8 0.5 0.8 0.3 0.1 ...
## $ Sugars : num 15 14 16 13.3 16 14 13 3 15 3 ...
## $ Protein : num 1 1 1.3 2.7 1 1 1 2 1 2 ...
```

```
ncol(breakfast_df)
```

```
## [1] 10
```

Question 3. Use base R (i.e., no external packages should be used for this question) to create a subset of the data that has only observations with values of Calories greater than 130.

```
breakfast_df[which(breakfast_df$Calories>130),]
```

##	Name	Company	Serving	Calories	Fat	Sodium	Carbs	Fiber	Sugars
## 3	Cap'n Crunch	Q	0.75	144	2.1	269	31	1.1	16.0
## 4	Cinnamon Toast Crunch	G	0.75	169	4.4	408	32	1.7	13.3
## 13	Frosted Mini-Wheats	K	1.00	175	0.8	5	41	5.0	10.0
## 14	Golden Grahams	G	0.75	149	1.3	359	33	1.3	14.7
## 15	Honey Nut Clusters	G	1.00	214	2.7	249	46	2.8	17.0
## 16	Honey Nut Heaven	Q	1.00	192	3.7	216	38	3.5	13.0
## 19	Life	Q	0.75	160	1.9	219	33	2.7	8.0
## 23	Raisin Bran	K	1.00	195	1.6	362	47	7.3	20.0
## 24	Reese's Puffs	G	0.75	171	3.9	223	31	0.0	16.0
## 26	Rice Krispie Treats	K	0.75	160	1.7	252	35	0.0	12.0
## 27	Smart Start	K	1.00	182	0.7	275	43	2.8	14.0
##	Protein								
## 3	1.3								
## 4	2.7								
## 13	5.0								
## 14	2.7								
## 15	4.0								
## 16	4.0								
## 19	4.0								
## 23	5.0								
## 24	2.7								
## 26	1.3								
## 27	4.0								

Question 4. Use the pipe operators in tidyverse to create a subset of the data that has only observations with values of Calories greater than 130.

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2    3.5.1      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1
## ✓ purrr      1.0.2
## — Conflicts ————— tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
breakfast_df_subset_1 <- breakfast_df %>% filter(Calories > 130)
summary(breakfast_df_subset_1)
```

```
##      Name      Company      Serving      Calories
## Length:11      Length:11      Min.    :0.7500  Min.    :144.0
## Class :character Class :character 1st Qu.:0.7500 1st Qu.:160.0
## Mode  :character Mode  :character Median :0.7500 Median :171.0
##                                     Mean  :0.8636 Mean  :173.7
##                                     3rd Qu.:1.0000 3rd Qu.:187.0
##                                     Max.  :1.0000 Max.  :214.0
##      Fat      Sodium      Carbs      Fiber      Sugars
## Min.    :0.700  Min.    : 5.0  Min.    :31.00  Min.    :0.000  Min.    : 8.0
## 1st Qu.:1.450  1st Qu.:221.0 1st Qu.:32.50  1st Qu.:1.200  1st Qu.:12.5
## Median :1.900  Median :252.0 Median :35.00  Median :2.700  Median :14.0
## Mean   :2.255  Mean   :257.9 Mean   :37.27  Mean   :2.564  Mean   :14.0
## 3rd Qu.:3.200  3rd Qu.:317.0 3rd Qu.:42.00  3rd Qu.:3.150  3rd Qu.:16.0
## Max.   :4.400  Max.   :408.0 Max.   :47.00  Max.   :7.300  Max.   :20.0
##      Protein
## Min.    :1.300
## 1st Qu.:2.700
## Median :4.000
## Mean   :3.336
## 3rd Qu.:4.000
## Max.   :5.000
```

Question 5. Use the pipe operators in tidyverse to calculate the mean Sugars by Company.

```
breakfast_df %>% group_by(Company) %>% summarise(MeanS=mean(Sugars),) %>% arrange(MeanS)
```

```
## # A tibble: 3 × 2
##   Company Means
##   <chr>   <dbl>
## 1 K      10.2
## 2 G      10.4
## 3 Q      10.8
```