**Table of Contents**

Week 2 Exercises

## Sue Susman MEd, BSN, RN

### March 28, 2023

[Exercise 1 1](#_bookmark0)

[Exercise 2 2](#_bookmark1)

[Exercise 3 2](#_bookmark2)

[Exercise 4 3](#_bookmark3)

[Exercise 5 4](#_bookmark4)

[Exercise 6 4](#_bookmark5)

[Exercise 7 4](#_bookmark6)

[Exercise 8 5](#_bookmark7)

[Exercise 9 6](#_bookmark8)

[Exercise 10 7](#_bookmark9)

# Exercise 1

### Read the sales\_pipe.txt file into an R data frame as sales.

*# Your code here*

sales <- read.delim("sales\_pipe.txt", sep = "|") head(sales)

## O..Row.ID Order.ID Order.Date Ship.Date Ship.Mode

## 1 1 CA-2016-152156 11/8/2016 November 11 2016 Second Class

## 2 2 CA-2016-152156 11/8/2016 November 11 2016 Second Class

## 3 3 CA-2016-138688 6/12/2016 June 16 2016 Second Class

## 4 4 US-2015-108966 10/11/2015 October 18 2015 Standard Class

## 5 5 US-2015-108966 10/11/2015 October 18 2015 Standard Class

## 6 6 CA-2014-115812 6/9/2014 June 14 2014 Standard Class ## Customer.ID Customer.Name Segment Country City

## 1 CG-12520 Claire Gute Consumer United States Henderson ## 2 CG-12520 Claire Gute Consumer United States Henderson ## 3 DV-13045 Darrin Van Huff Corporate United States Los Angeles

## 4 SO-20335 Sean O'Donnell Consumer United States Fort Lauderdale ## 5 SO-20335 Sean O'Donnell Consumer United States Fort Lauderdale ## 6 BH-11710 Brosina Hoffman Consumer United States Los Angeles ## State Postal.Code Region Product.ID Category Sub.Category

## 1 Kentucky 42420 South FUR-BO-10001798 Furniture Bookcases ## 2 Kentucky 42420 South FUR-CH-10000454 Furniture Chairs ## 3 California 90036 West OFF-LA-10000240 Office Supplies Labels ## 4 Florida 33311 South FUR-TA-10000577 Furniture Tables

## 5 Florida 33311 South OFF-ST-10000760 Office Supplies Storage ## 6 California 90032 West FUR-FU-10001487 Furniture Furnishings ## Product.Name Sales

## 1 Bush Somerset Collection Bookcase 261.9600

## 2 Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back 731.9400 ## 3 Self-Adhesive Address Labels for Typewriters by Universal 14.6200

## 4 Bretford CR4500 Series Slim Rectangular Table 957.5775 ## 5 Eldon Fold 'N Roll Cart System 22.3680

## 6 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.8600 ## Quantity Discount Profit

## 1 2 0.00 41.9136

|  |  |  |
| --- | --- | --- |
| ## 2 | 3 | 0.00 219.5820 |
| ## 3 | 2 | 0.00 6.8714 |
| ## 4 | 5 | 0.45 -383.0310 |
| ## 5 | 2 | 0.20 2.5164 |
| ## 6 | 7 | 0.00 14.1694 |

# Exercise 2

### You can extract a vector of columns names from a data frame using the colnames() function. Notice the first column has some odd characters. Change the column name for the FIRST column in the sales date frame to Row.ID.

## Note: You will need to assign the first element of colnames to a single character.

*# Your code here*

colnames(sales)[1] <- "Row.ID"

# Exercise 3

### Convert both Order.ID and Order.Date to date vectors within the sales data frame. What is the number of days between the most recent order and the oldest order? How many years is that? How many weeks?

## Note: Use lubridate

*# Your code here*

library(lubridate)

## Attaching package: 'lubridate'

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

***#Convert character to date format*** sales$Order.ID <- mdy(sales$Order.ID) class(sales$Order.ID)

## [1] "Date"

sales$Order.Date <- mdy(sales$Order.Date) class(sales$Order.Date)

## [1] "Date"

start\_date <- min(sales$Order.Date) start\_date

## [1] "2014-01-03"

end\_date <- max(sales$Order.Date) end\_date

## [1] "2017-12-30"

start\_date <- as.Date("2014-01-03") end\_date <- as.Date("2017-12-30")

***#Total number of days***

num\_days <- as.numeric(difftime(end\_date, start\_date, units = "days")) num\_days\_rounded <- round(num\_days, 2) print(paste0(num\_days\_rounded, " days"))

## [1] "1457 days"

***#Total number of weeks***

num\_days <- as.numeric(difftime(end\_date, start\_date, units = "weeks")) num\_days\_rounded <- round(num\_days, 2) print(paste0(num\_days\_rounded, " weeks"))

## [1] "208.14 weeks"

***# Total number of Years***

num\_days <- as.numeric(difftime(end\_date, start\_date, units = "weeks"))/52.25 num\_days\_rounded <- round(num\_days, 2)

print(paste0(num\_days\_rounded, " years"))

## [1] "3.98 years"

# Exercise 4

### What is the average number of days it takes to ship an order?

*# Your code here*

library(lubridate)

sales$Ship.Date <- mdy(sales$Ship.Date) class(sales$Ship.Date)

## [1] "Date"

***# assuming the sales data is stored in a data frame called 'sales'***

***# convert the Order.Date and Ship.Date columns to Date objects*** sales$Order.Date <- as.Date(sales$Order.Date, format = "%m/%d/%Y") sales$Ship.Date <- as.Date(sales$Ship.Date, format = "%m/%d/%Y")

***# calculate the number of days it takes to ship each order***

days\_to\_ship <- as.numeric(difftime(sales$Ship.Date, sales$Order.Date, units = "days"))

***# calculate the average number of days it takes to ship an order***

avg\_days\_to\_ship <- mean(days\_to\_ship, na.rm = TRUE)

***# print the result***

print(paste0("The average number of days it takes to ship an order is ", round(avg\_days\_to\_ship, 2), " days."))

## [1] "The average number of days it takes to ship an order is 3.91 days."

# Exercise 5

### How many customers have the first name Bill? You will need to split the customer name into first and last name segments and then use a regular expression to match the first name bill. Use the length() function to determine the number of customers with the first name Bill in the sales data.

*# Your code here*

library(stringr)

***# Split the "Customer.Name" column into first and last name segments*** names <- str\_split(sales$Customer.Name, pattern = " ", simplify = TRUE) head(names)

## [,1] [,2] [,3]

## [1,] "Claire" "Gute" ""

## [2,] "Claire" "Gute" ""

## [3,] "Darrin" "Van" "Huff"

## [4,] "Sean" "O'Donnell" ""

## [5,] "Sean" "O'Donnell" ""

## [6,] "Brosina" "Hoffman" ""

***# Extract the first names*** first\_names <- names[, 1] head(first\_names)

## [1] "Claire" "Claire" "Darrin" "Sean" "Sean" "Brosina"

***# Count the number of customers with the first name "Bill"*** num\_bills <- length(grep("^Bill$", first\_names, ignore.case = TRUE)) head(num\_bills)

## [1] 37

# Exercise 6

How many mentions of the word ‘table’ are there in the Product.Name column? **Note you can do this in one line of code**

*# Your code here*

num\_tables <- sum(grepl("table", sales$Product.Name, ignore.case = TRUE)) head(num\_tables)

## [1] 371

# Exercise 7

### Create a table of counts for each state in the sales data. The counts table should be ordered alphabetically from A to Z.

*# Your code here*

state\_counts <- sort(table(sales$State)) head(state\_counts)

##

## District of Columbia

1

Maine 4

1

Wyoming

2

Montana

##

## ##

West Virginia

4

North Dakota

7

# Exercise 8

### Create an alphabetically ordered barplot for each sales Category in the State of Texas.

*# Your code here*

tx\_sales <- subset(sales, State == "Texas") head(tx\_sales)

## Row.ID Order.ID Order.Date Ship.Date Ship.Mode Customer.ID ## 15 15 <NA> 2015-11-22 2015-11-26 Standard Class HP-14815

## 16 16 <NA> 2015-11-22 2015-11-26 Standard Class HP-14815

## 78 78 <NA> 2017-12-09 2017-12-11 First Class KB-16600

## 79 79 <NA> 2014-11-26 2014-12-01 Second Class JE-15745

## 89 89 <NA> 2016-04-05 2016-04-10 Second Class GM-14455

## 345 345 <NA> 2015-12-20 2015-12-24 Standard Class CS-12130

## Customer.Name Segment Country City State Postal.Code

## 15 Harold Pawlan Home Office United States Fort Worth Texas 76106 ## 16 Harold Pawlan Home Office United States Fort Worth Texas 76106 ## 78 Ken Brennan Corporate United States Houston Texas 77041 ## 79 Joel Eaton Consumer United States Houston Texas 77070

## 89 Gary Mitchum Home Office United States Houston Texas 77095 ## 345 Chad Sievert Consumer United States Austin Texas 78745 ## Region Product.ID Category Sub.Category

## 15 Central OFF-AP-10002311 Office Supplies Appliances ## 16 Central OFF-BI-10000756 Office Supplies Binders ## 78 Central OFF-ST-10000615 Office Supplies Storage ## 79 Central FUR-FU-10003194 Furniture Furnishings ## 89 Central OFF-ST-10003442 Office Supplies Storage ## 345 Central OFF-ST-10000107 Office Supplies Storage ## Product.Name

## 15 Holmes Replacement Filter for HEPA Air Cleaner, Very Large Room, HEPA Filter ## 16 Storex DuraTech Recycled Plastic Frosted Binders

## 78 SimpliFile Personal File, Black Granite, 15w x 6-15/16d x 11-1/4h ## 79 Eldon Expressions Desk Accessory, Wood Pencil Holder, Oak ## 89 Eldon Portable Mobile Manager

## 345 Fellowes Super Stor/Drawer ## Sales Quantity Discount Profit

## 15 68.810 5 0.8 -123.8580

|  |  |  |  |
| --- | --- | --- | --- |
| ## 16 | 2.544 | 3 | 0.8 -3.8160 |
| ## 78 | 27.240 | 3 | 0.2 2.7240 |
| ## 79 | 19.300 | 5 | 0.6 -14.4750 |
| ## 89 | 158.368 | 7 | 0.2 13.8572 |
| ## 345 88.800 | | 4 | 0.2 -2.2200 |

tx\_category\_counts <- sort(table(tx\_sales$Category))

head(tx\_category\_counts)

##

## ##

Furniture

86

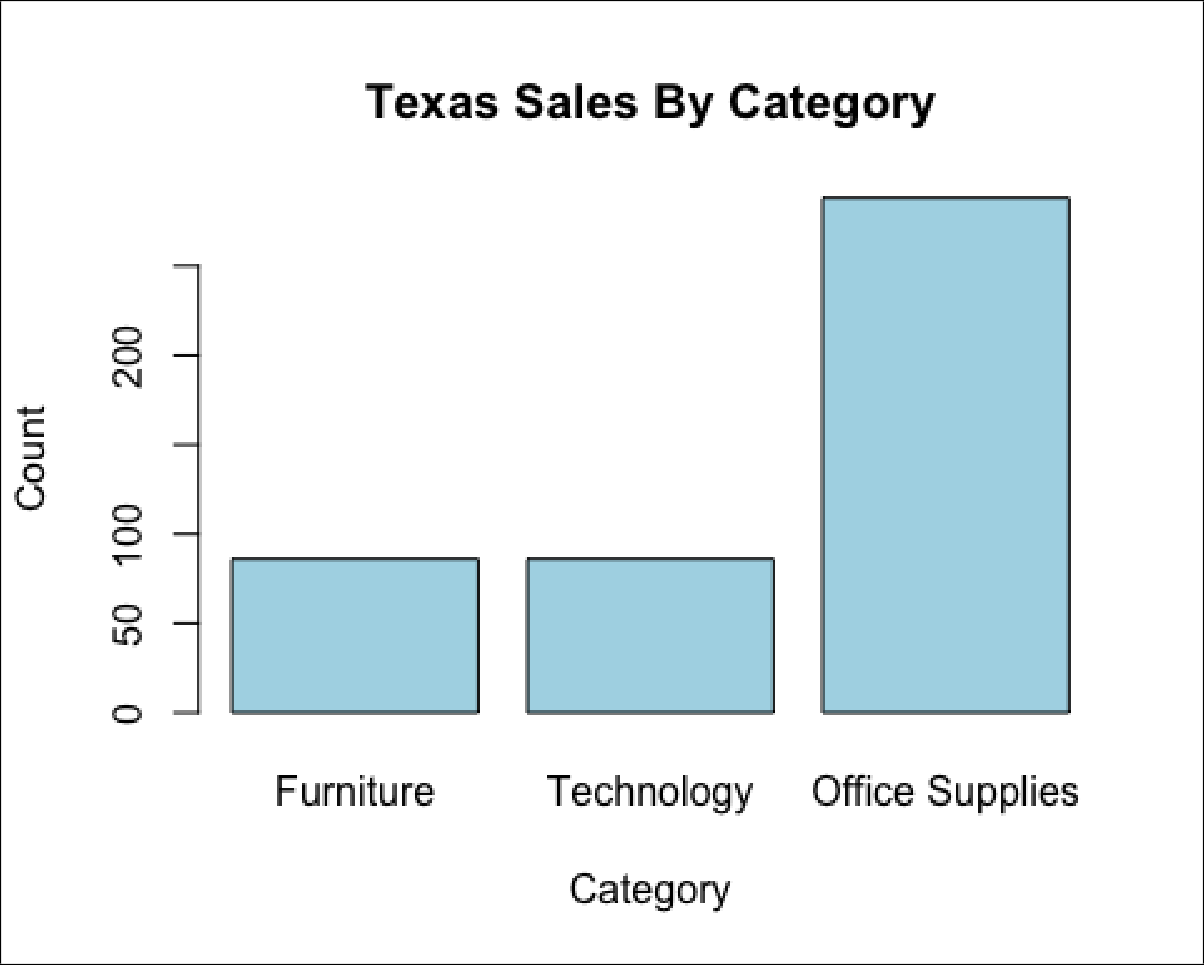
Technology Office Supplies

86

288

barplot(tx\_category\_counts, main = c("Texas Sales By Category"), xlab = "Category", ylab = "Count", col =

"lightblue")



# Exercise 9

Find the average profit by region. **Note: You will need to use the aggregate() function to do this. To understand how the function works type ?aggregate in the console.**

*# Your code here*

region\_profit <- aggregate(Profit ~ Region, data = sales, FUN = mean) head(region\_profit)

## Region Profit

## 1 Central 20.46822

## 2 East 29.91937

## 3 South 11.27720

## 4 West 32.77000

***## Region Profit***

***## 1 Central 17.09271***

***## 2 East 32.13581***

***## 3 South 28.85767***

***## 4 West 33.84903***

# Exercise 10

### Find the average profit by order year.

## Note: You will need to use the aggregate() function to do this. To understand how the function works type ?aggregate in the console.

*# Your code here*

year\_profit <- aggregate(Profit ~ as.numeric(format(as.Date(Order.Date), "%Y")), data = sales, FUN = mean) head(year\_profit)

## as.numeric(format(as.Date(Order.Date), "%Y")) Profit

|  |  |
| --- | --- |
| ## 1 | 2014 32.24582 |
| ## 2 | 2015 21.58676 |
| ## 3 | 2016 30.10960 |
| ## 4 | 2017 21.31825 |
| ***## as.numeric(format(as.Date(Order.Date), "%Y")) Profit*** | |
| ***## 1*** | ***2014 24.85899*** |
| ***## 2*** | ***2015 29.31427*** |
| ***## 3*** | ***2016 31.61777*** |
| ***## 4*** | ***2017 28.21234*** |