

W1

Alan

2025-01-14

```
library(tinytex)
library(magrittr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2    3.5.1      v tibble     3.2.1
## v lubridate  1.9.4      v tidyr      1.3.1
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x tidyr::extract() masks magrittr::extract()
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag() masks stats::lag()
```

```
## x purrr::set_names() masks magrittr::set_names()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
df <- read.csv("superstore.csv", stringsAsFactors = T)
```

```
head(df)
```

```
##   Row.ID      Order.ID Order.Date Ship.Date      Ship.Mode Customer.ID
## 1      1 CA-2017-114412  15-04-17  20-04-17 Standard Class    AA-10480
## 2      2 US-2017-156909  16-07-17  18-07-17  Second Class    SF-20065
## 3      3 CA-2017-107727  19-10-17  23-10-17  Second Class    MA-17560
## 4      4 CA-2017-120999  10-09-17  15-09-17 Standard Class    LC-16930
## 5      5 CA-2017-139619  19-09-17  23-09-17 Standard Class    ES-14080
## 6      6 CA-2017-114440  14-09-17  17-09-17  Second Class    TB-21520
```

```
##      Customer.Name      State Postal.Code Region      Product.ID
## 1  Andrew Allen North Carolina      28027  South OFF-PA-10002365
## 2 Sandra Flanagan  Pennsylvania      19140   East FUR-CH-10002774
## 3   Matt Abelman      Texas      77095 Central OFF-PA-10000249
## 4 Linda Cazamias      Illinois      60540 Central TEC-PH-10004093
## 5   Erin Smith      Florida      32935  South OFF-ST-10003282
## 6 Tracy Blumstein      Michigan      49201 Central OFF-PA-10004675
```

```
##      Category Sub.Category
## 1 Office Supplies      Paper
## 2      Furniture      Chairs
## 3 Office Supplies      Paper
## 4      Technology      Phones
## 5 Office Supplies      Storage
## 6 Office Supplies      Paper
```

```
##                                     Product.Name
## 1                                     Xerox 1967
```

```
## 2                                Global Deluxe Stacking Chair, Gray
## 3                                Easy-staple paper
## 4                                Panasonic Kx-TS550
## 5 Advantus 10-Drawer Portable Organizer, Chrome Metal Frame, Smoke Drawers
## 6 Telephone Message Books with Fax/Mobile Section, 5 1/2" x 3 3/16"
## Sales Quantity Discount Profit
## 1 15.552      3      0.2 5.4432
## 2 71.372      2      0.3 -1.0196
## 3 29.472      3      0.2 9.9468
## 4 147.168     4      0.2 16.5564
## 5 95.616      2      0.2 9.5616
## 6 19.050      3      0.0 8.7630
```

```
df1 <- df %>% select(State, Region, Sales)
tail(df1, 3)
```

```
## State Region Sales
## 3268 California West 258.576
## 3269 California West 29.600
## 3270 California West 243.160
```

```
library(tidyverse)
```

```
df2 <- df %>% filter(Region == "Central") %>% group_by(Ship.Mode) %>% summarise(n = n(), AverageProfit =
str(df2)
```

```
## tibble [4 x 4] (S3: tbl_df/tbl/data.frame)
## $ Ship.Mode : Factor w/ 4 levels "First Class",...: 1 2 3 4
## $ n : int [1:4] 127 53 139 442
## $ AverageProfit: num [1:4] -1.7 20 19.2 9
## $ TotalSales : num [1:4] 23792 8153 32289 81964
```

```
df2
```

```
## # A tibble: 4 x 4
## Ship.Mode n AverageProfit TotalSales
## <fct> <int> <dbl> <dbl>
## 1 First Class 127 -1.70 23792.
## 2 Same Day 53 20.0 8153.
## 3 Second Class 139 19.2 32289.
## 4 Standard Class 442 9.00 81964.
```

```
states <- df %>% group_by(State)
states %>% summarise(MedianProfit = median(Profit)) %>% arrange(desc(MedianProfit))
```

```
## # A tibble: 47 x 2
## State MedianProfit
## <fct> <dbl>
## 1 Oklahoma 60.6
## 2 Nevada 47.9
## 3 Mississippi 43.4
## 4 Vermont 37.6
## 5 Delaware 33.3
## 6 Michigan 33.1
## 7 Louisiana 28.3
## 8 New Hampshire 26.6
## 9 Georgia 24.0
```

```
## 10 South Dakota          23.8  
## # i 37 more rows
```

```
negative_profit <- df %>% filter(Profit < 0)  
average_sales_of_negative_profit_sales <- negative_profit %>% select(Sales)  
average_sales_of_negative_profit_average_sales <- sum(average_sales_of_negative_profit_sales) / length(average_sales_of_negative_profit_average_sales)
```

```
## [1] 145928.2
```