

# Lab 4

2023-06-08

## Question 1

```
#1a.  
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —  
## ✓ dplyr      1.1.2      ✓ readr      2.1.4  
## ✓ forcats    1.0.0      ✓ stringr    1.5.0  
## ✓ ggplot2    3.4.2      ✓ tibble     3.2.1  
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0  
## ✓ purrr      1.0.1  
## — Conflicts — tidyverse_conflicts() —  
## X dplyr::filter() masks stats::filter()  
## X dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#1b.  
super <- read.csv("Superstores.csv")  
  
#1c.  
total_sales <- super %>% filter(Category == "Office Supplies") %>% summarise(total_sales = sum(Sales))  
total_sales
```

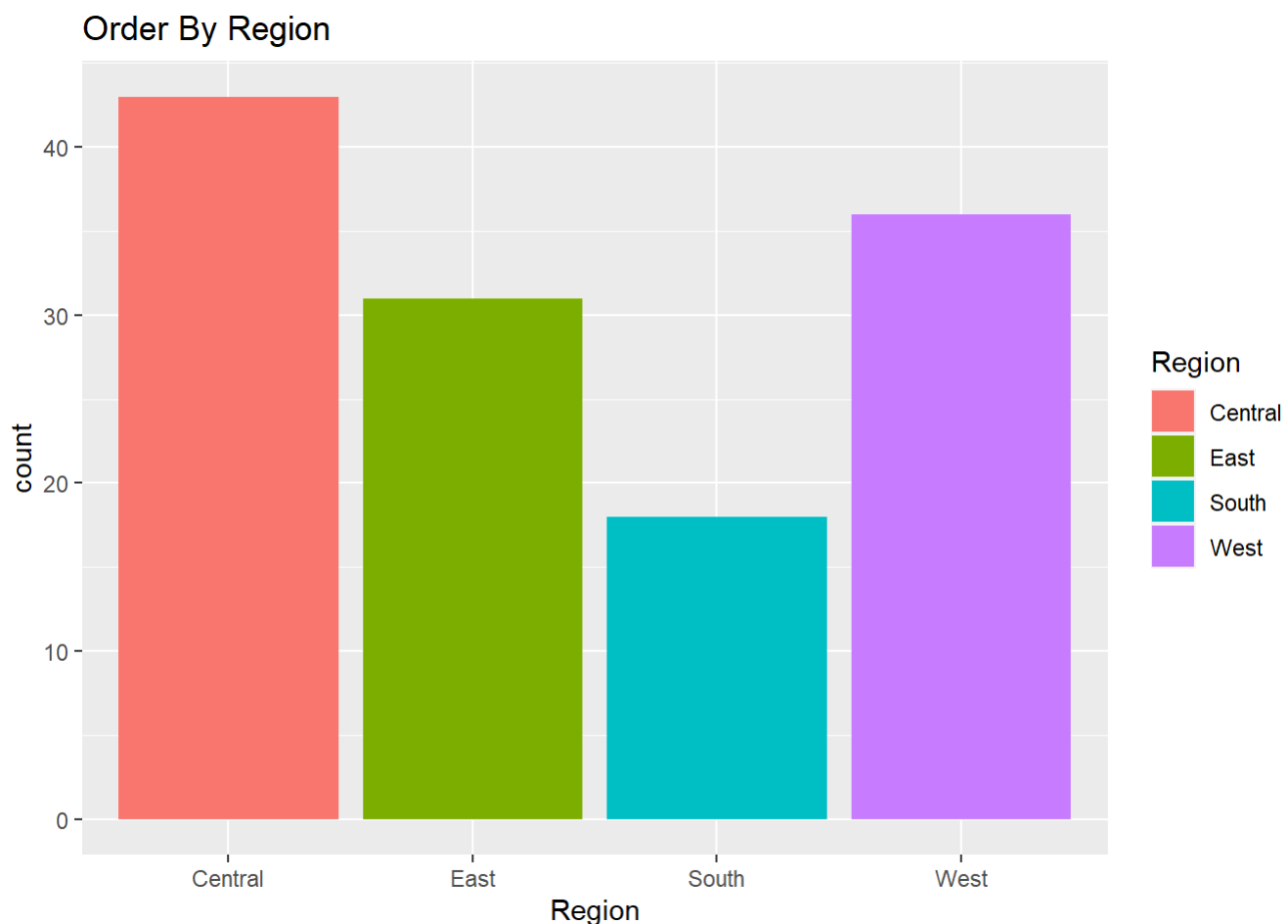
```
##   total_sales  
## 1      6378.938
```

```
#1d.  
superCali <- super %>% filter(State == "California")  
head(superCali)
```

```
## Row.ID Order.ID Order.Date Ship.Date Ship.Mode Customer.ID
## 1 3 CA-2016-138688 6/12/2016 6/16/2016 Second Class DV-13045
## 2 6 CA-2014-115812 6/9/2014 6/14/2014 Standard Class BH-11710
## 3 7 CA-2014-115812 6/9/2014 6/14/2014 Standard Class BH-11710
## 4 8 CA-2014-115812 6/9/2014 6/14/2014 Standard Class BH-11710
## 5 9 CA-2014-115812 6/9/2014 6/14/2014 Standard Class BH-11710
## 6 10 CA-2014-115812 6/9/2014 6/14/2014 Standard Class BH-11710
## Customer.Name Segment Country City State Postal.Code
## 1 Darrin Van Huff Corporate United States Los Angeles California 90036
## 2 Brosina Hoffman Consumer United States Los Angeles California 90032
## 3 Brosina Hoffman Consumer United States Los Angeles California 90032
## 4 Brosina Hoffman Consumer United States Los Angeles California 90032
## 5 Brosina Hoffman Consumer United States Los Angeles California 90032
## 6 Brosina Hoffman Consumer United States Los Angeles California 90032
## Region Product.ID Category Sub.Category
## 1 West OFF-LA-1000240 Office Supplies Labels
## 2 West FUR-FU-10001487 Furniture Furnishings
## 3 West OFF-AR-10002833 Office Supplies Art
## 4 West TEC-PH-10002275 Technology Phones
## 5 West OFF-BI-10003910 Office Supplies Binders
## 6 West OFF-AP-10002892 Office Supplies Appliances
## Product.Name Sales
## 1 Self-Adhesive Address Labels for Typewriters by Universal 14.620
## 2 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.860
## 3 Newell 322 7.280
## 4 Mitel 5320 IP Phone VoIP phone 907.152
## 5 DXL Angle-View Binders with Locking Rings by Samsill 18.504
## 6 Belkin F5C206VTEL 6 Outlet Surge 114.900
## Quantity Discount Profit
## 1 2 0.0 6.8714
## 2 4 0.0 14.1694
## 3 4 0.0 1.9656
## 4 4 0.2 90.7152
## 5 3 0.2 5.7825
## 6 5 0.0 34.4700
```

*#1e. You can use your creativity to produce different layouts.*

```
ggplot(super, aes(x = Region, fill = Region)) + geom_bar() + labs(title = "Order By Region")
```



## Question 2

```
library(palmerpenguins)
#2a.
```

```
head(penguins)
```

```
## # A tibble: 6 × 8
##   species island   bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##   <fct>   <fct>         <dbl>         <dbl>         <int>         <int>
## 1 Adelie  Torgersen      39.1          18.7           181          3750
## 2 Adelie  Torgersen      39.5          17.4           186          3800
## 3 Adelie  Torgersen      40.3          18            195          3250
## 4 Adelie  Torgersen      NA            NA             NA            NA
## 5 Adelie  Torgersen      36.7          19.3           193          3450
## 6 Adelie  Torgersen      39.3          20.6           190          3650
## # i 2 more variables: sex <fct>, year <int>
```

#2b.

# specify the width of the bins to suit your need

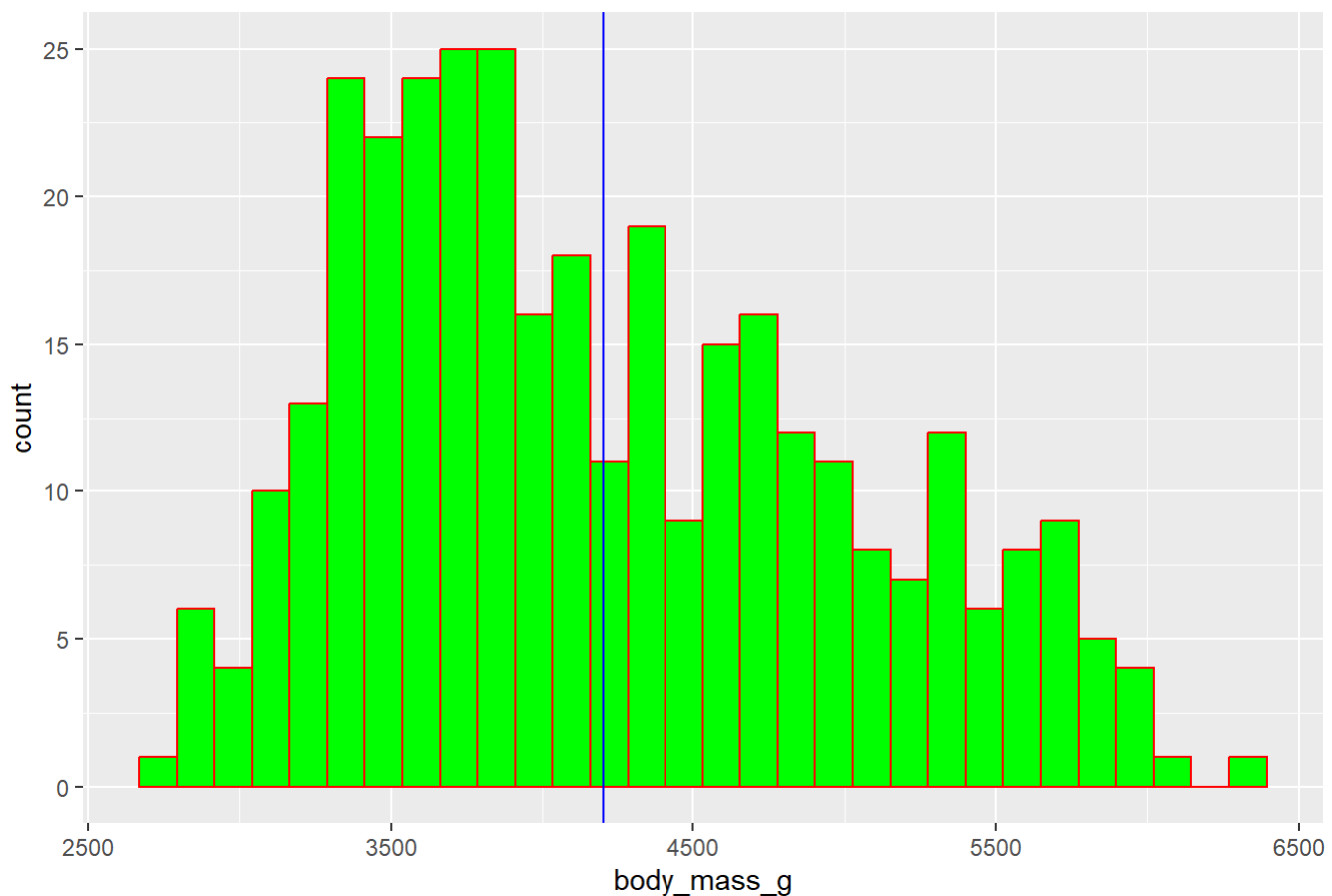
average\_body\_mass &lt;- mean(penguins\$body\_mass\_g, na.rm = TRUE)

```
ggplot(penguins, aes(x = body_mass_g)) + geom_histogram(fill = "green", color = "red") + geom_vline(xintercept = average_body_mass, color = "blue", linetype = "solid") + labs(title = "Penguin Body Mass")
```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 2 rows containing non-finite values (`stat\_bin()`).

Penguin Body Mass

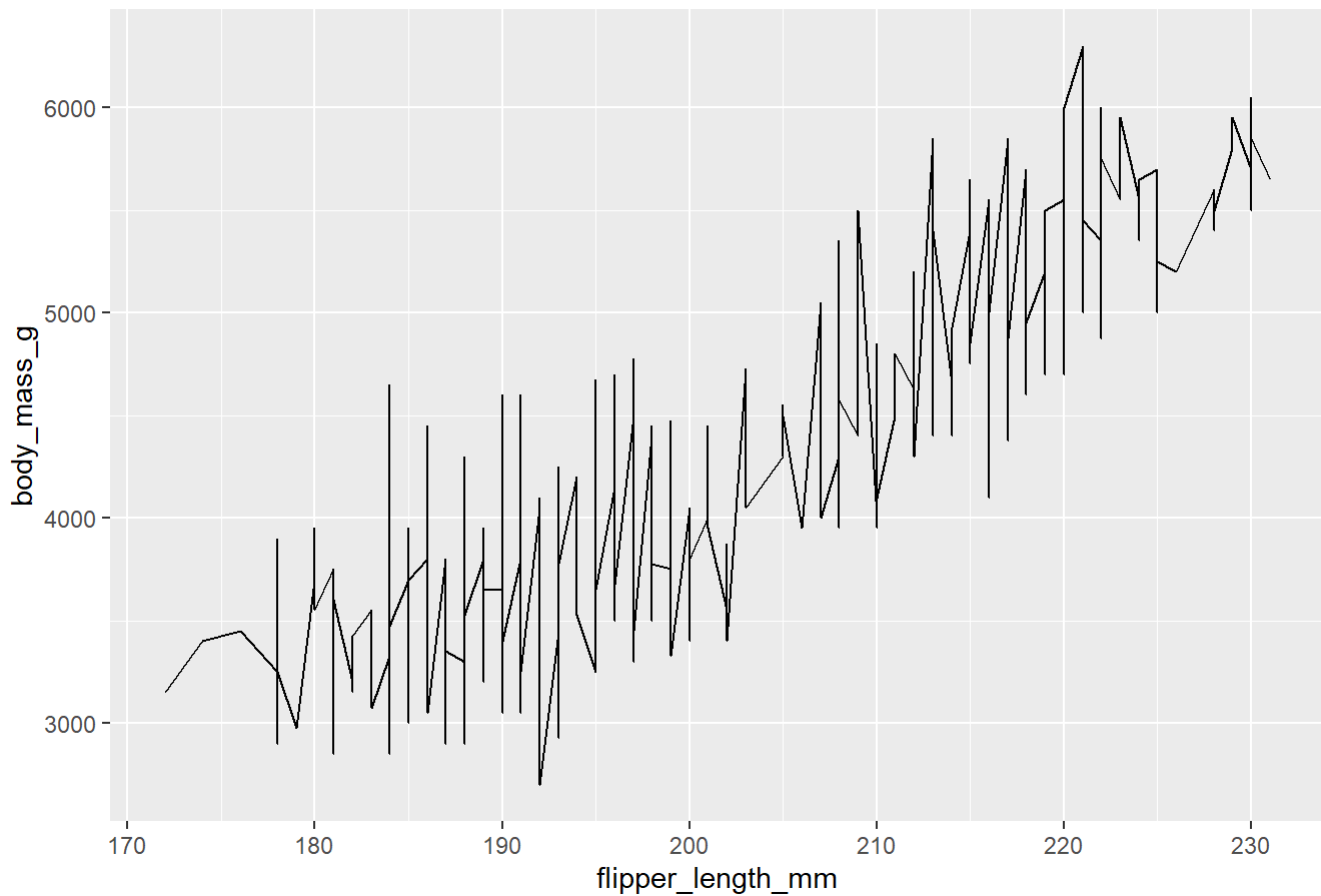


## Question 3

```
ggplot(penguins, aes(x = flipper_length_mm, y = body_mass_g)) + geom_line() + labs(title = "Penguin flipper Length vs. Body Mass")
```

## Warning: Removed 2 rows containing missing values (`geom\_line()`).

### Penguin flipper Length vs. Body Mass



What do you think of the plot? Do you think a line plot is appropriate? Why or why not? yes its appropriate because it generally shpows the bigger the penguin the longer the flipper.