

Stat123 Lab5 O'Toole

Evan O'Toole

```
#install.packages("palmerpenguins")
```

QUESTION #1

```
library(palmerpenguins)
```

Warning: package 'palmerpenguins' was built under R version 4.3.3

```
library(ggplot2)
```

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
#a)
```

```
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	375
2	Adelie	Torgersen	39.5	17.4	186	380
3	Adelie	Torgersen	40.3	18	195	325
4	Adelie	Torgersen	NA	NA	NA	NA
5	Adelie	Torgersen	36.7	19.3	193	345
6	Adelie	Torgersen	39.3	20.6	190	365

```
# ⓘ 2 more variables: sex <fct>, year <int>
```

```
dim(penguins)
```

```

[1] 344    8

str(penguins)

tibble [344 × 8] (S3: tbl_df/tbl/data.frame)
 $ species      : Factor w/ 3 levels "Adelie","Chinstrap",...: 1 1 1 1 1 1 1 1 1 1 1 1 ...
 $ island       : Factor w/ 3 levels "Biscoe","Dream",...: 3 3 3 3 3 3 3 3 3 3 3 3 ...
 $ bill_length_mm : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
 $ bill_depth_mm  : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20 .2 ...
 $ flipper_length_mm: int [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
 $ body_mass_g    : int [1:344] 3750 3800 3250 NA 3450 3650 3625 4675 3475 4250 ...
 $ sex           : Factor w/ 2 levels "female","male": 2 1 1 NA 1 2 1 2 NA NA ...
 $ year          : int [1:344] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...

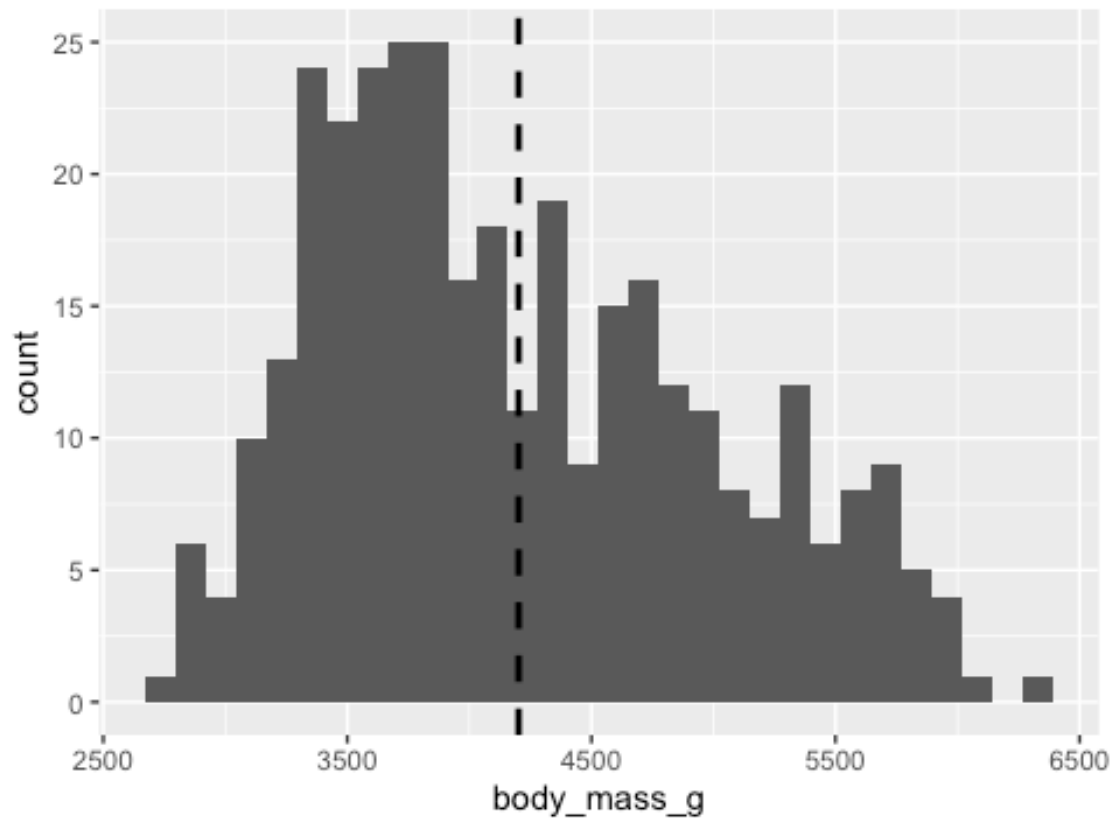
#b)
ggplot(data = penguins, aes(x = body_mass_g)) + geom_histogram() + geom_vline(
  xintercept = mean(penguins$body_mass_g, na.rm = TRUE), linetype = "dashed",
  linewidth = 1.0) + ggtitle("Histogram of penguins weights and the mean")

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Warning: Removed 2 rows containing non-finite outside the scale range
(`stat_bin()`).

```

Histogram of penguins weights and the mean



```
#c)
#It seems to be a bit right skewed and not very symmetric. so use median and
median(penguins$body_mass_g, na.rm = TRUE)

[1] 4050

quantile(penguins$body_mass_g, na.rm = TRUE)

 0%  25%  50%  75% 100%
2700 3550 4050 4750 6300
```

QUESTION #2

```
#a)
#head(iris)
g <- iris |>
  filter(Sepal.Length > 4.6 & Petal.Width > 0.5)
head(g)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.0	3.5	1.6	0.6	setosa
2	7.0	3.2	4.7	1.4	versicolor
3	6.4	3.2	4.5	1.5	versicolor
4	6.9	3.1	4.9	1.5	versicolor
5	5.5	2.3	4.0	1.3	versicolor
6	6.5	2.8	4.6	1.5	versicolor

```
#b)
i <- iris |> arrange(Sepal.Width)
head(i)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.0	2.0	3.5	1.0	versicolor
2	6.0	2.2	4.0	1.0	versicolor
3	6.2	2.2	4.5	1.5	versicolor
4	6.0	2.2	5.0	1.5	virginica
5	4.5	2.3	1.3	0.3	setosa
6	5.5	2.3	4.0	1.3	versicolor

```
#c)
i <- iris |> mutate(proportion = Sepal.Length/Sepal.Width)
head(i)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	proportion
1	5.1	3.5	1.4	0.2	setosa	1.457143
2	4.9	3.0	1.4	0.2	setosa	1.633333
3	4.7	3.2	1.3	0.2	setosa	1.468750
4	4.6	3.1	1.5	0.2	setosa	1.483871
5	5.0	3.6	1.4	0.2	setosa	1.388889
6	5.4	3.9	1.7	0.4	setosa	1.384615

```
#d)
a <- iris |> select(Sepal.Length, Sepal.Width, Petal.Length)
head(a)
```

	Sepal.Length	Sepal.Width	Petal.Length
1	5.1	3.5	1.4
2	4.9	3.0	1.4
3	4.7	3.2	1.3
4	4.6	3.1	1.5
5	5.0	3.6	1.4
6	5.4	3.9	1.7

```
#e)
np <- iris |> select(!Petal.Width)
head(np)
```

	Sepal.Length	Sepal.Width	Petal.Length	Species
1	5.1	3.5	1.4	setosa
2	4.9	3.0	1.4	setosa
3	4.7	3.2	1.3	setosa
4	4.6	3.1	1.5	setosa
5	5.0	3.6	1.4	setosa
6	5.4	3.9	1.7	setosa

```
#f)
iris |>
  select(Sepal.Width, Sepal.Length) |>
  head()
```

	Sepal.Width	Sepal.Length
1	3.5	5.1
2	3.0	4.9
3	3.2	4.7
4	3.1	4.6
5	3.6	5.0
6	3.9	5.4

```
#g)
i <- iris |> select(Sepal.Length, Sepal.Width, Petal.Length) |> arrange(Sepal.Length) |> arrange(Sepal.Width)
head(i)
```

	Sepal.Length	Sepal.Width	Petal.Length
1	5.0	2.0	3.5
2	6.0	2.2	4.0
3	6.0	2.2	5.0
4	6.2	2.2	4.5
5	4.5	2.3	1.3
6	5.0	2.3	3.3

```
#h)
iris |> summarize(avg_slength = mean(Sepal.Length))
```

	avg_slength
1	5.843333

```
#i)
iris |> group_by(Species) |> summarize(avg_SL = mean(Sepal.Length, na.rm = TRUE), max_SL = max(Sepal.Length, na.rm = TRUE), min_SL = min(Sepal.Length, na.rm = TRUE))
```

```
# A tibble: 3 × 4
  Species    avg_SL max_SL min_SL
  <fct>    <dbl> <dbl> <dbl>
```

1	setosa	5.01	5.8	4.3
2	versicolor	5.94	7	4.9
3	virginica	6.59	7.9	4.9