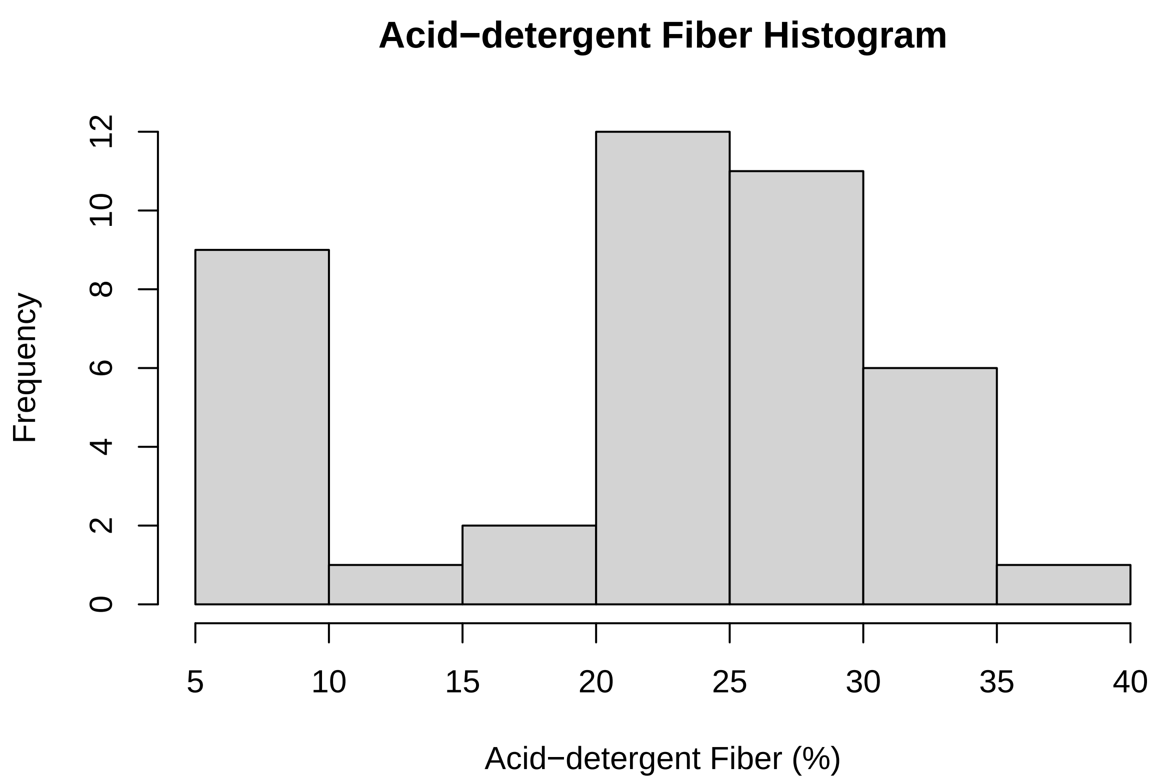
# lab1

Ayimen H.

2025-01-16

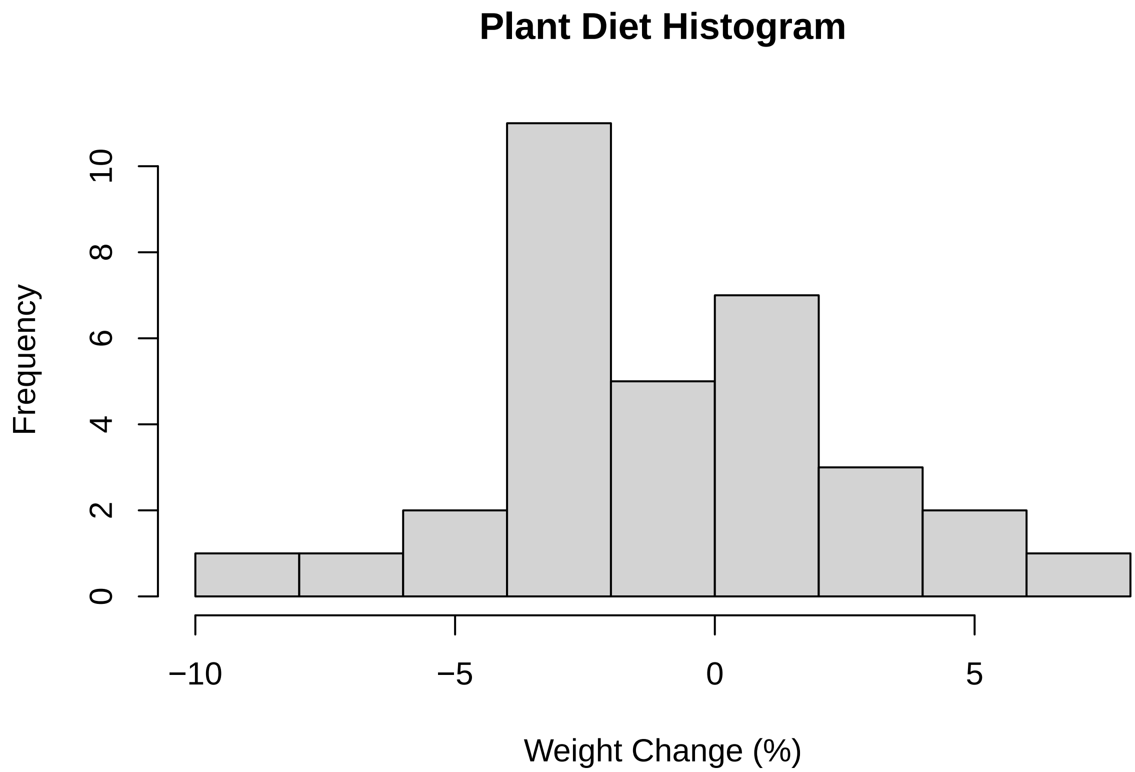
```
library(readxl)

Snowgeese <- read_excel("Snowgeese.xls", col_names = c("trial_number", "diet", "weight_change", "digest

acid_detergent_fiber <- Snowgeese$acid_detergent_fiber

hist(acid_detergent_fiber, breaks = 10, main = "Acid-detergent Fiber Histogram", xlab = "Acid-detergent
```




#1.a The histogram is likely bimodal because the geese under 10 have an extremely high digestion efficiency when comparing to the rest of the geese. A high digestion efficiency seems to lead to a low acid-detergent fiber percentage.

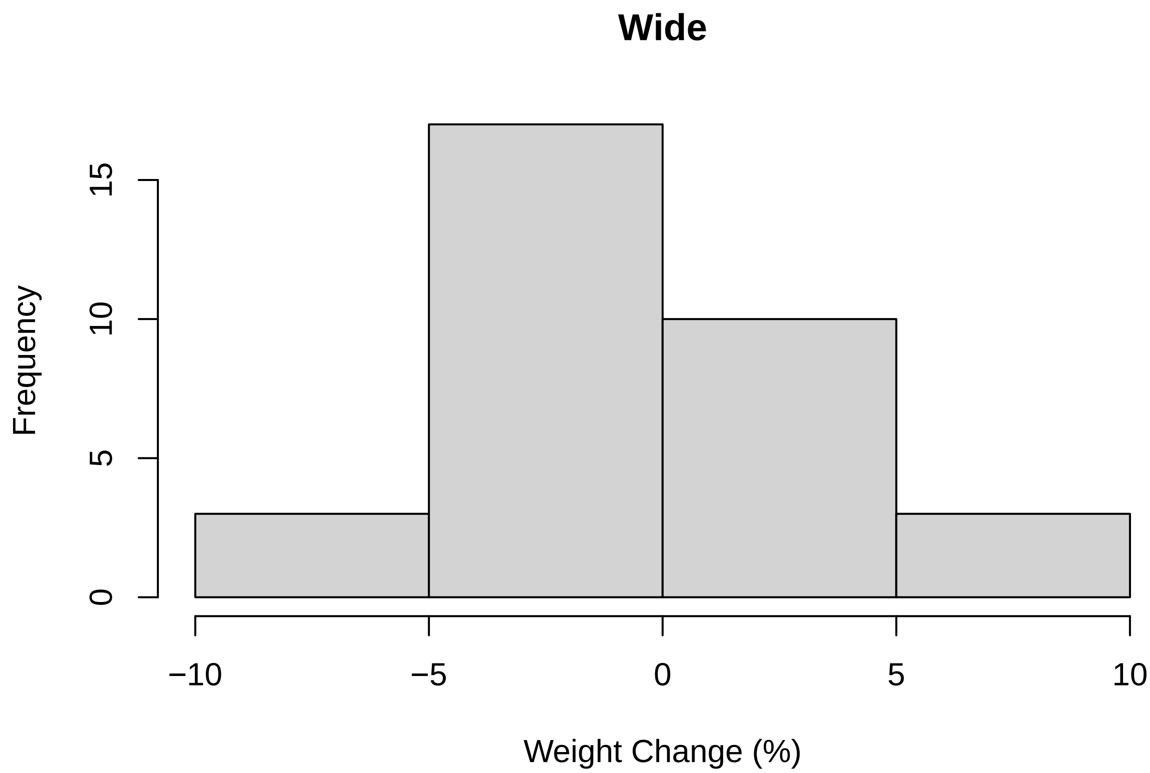
```
library(readxl)

Snowgeese <- read_excel("Snowgeese.xls", col_names = c("trial_number", "diet", "weight_change", "digest
```

```r
plant_diet <- Snowgeese[Snowgeese$diet == "Plants", ]

hist(plant_diet$weight_change, breaks = 10, main = "Plant Diet Histogram", xlab = "Weight Change (%)")
```

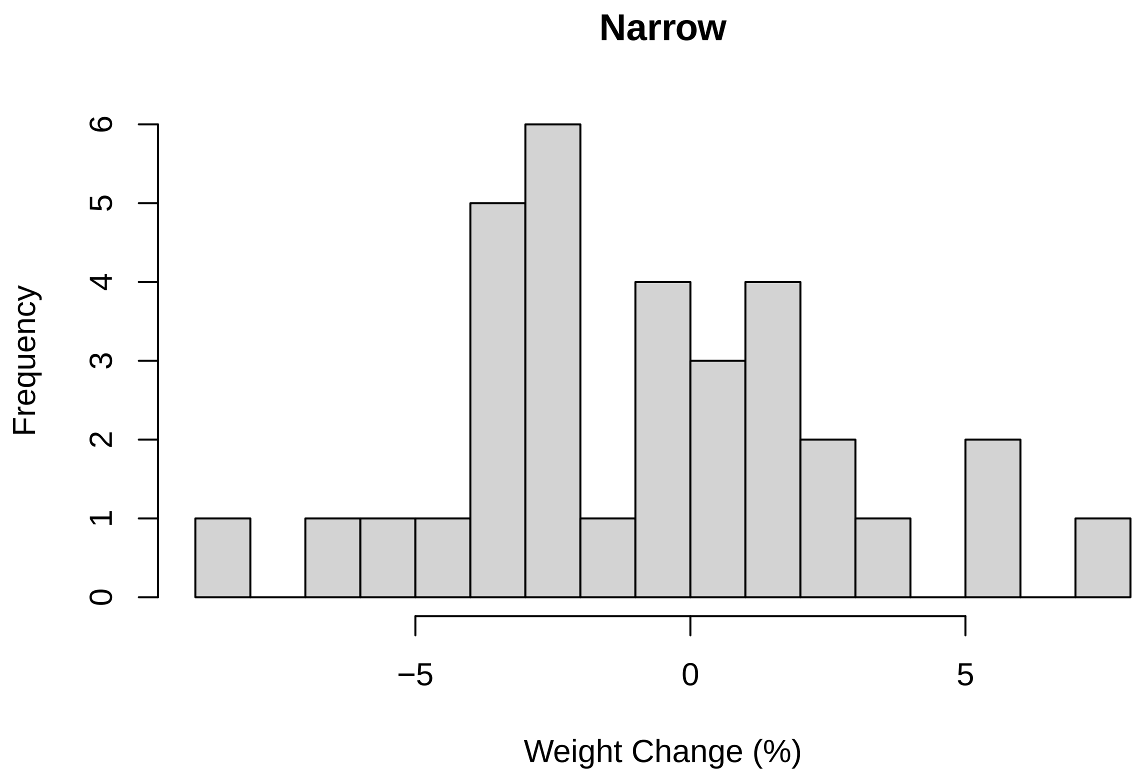



#1.b The histogram appears to be bimodal. There is a relative equality in weight gain and lose for the geese on a plants diet, leaning towards the loss side as a majority of the geese in the lose category were around -3.0.

```r
library(readxl)

Snowgeese <- read_excel("Snowgeese.xls", col_names = c("trial_number", "diet", "weight_change", "digest

plant_diet <- Snowgeese[Snowgeese$diet == "Plants", ]

hist(plant_diet$weight_change, breaks = 5, main = "Wide", xlab = "Weight Change (%)")
```

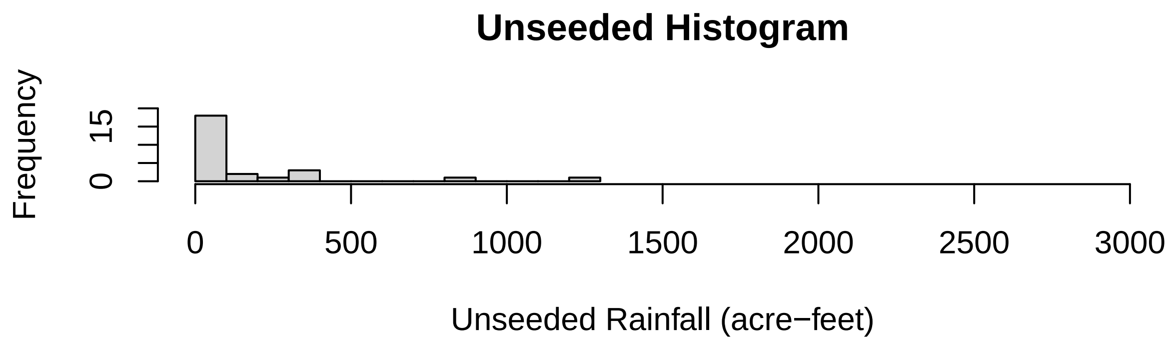


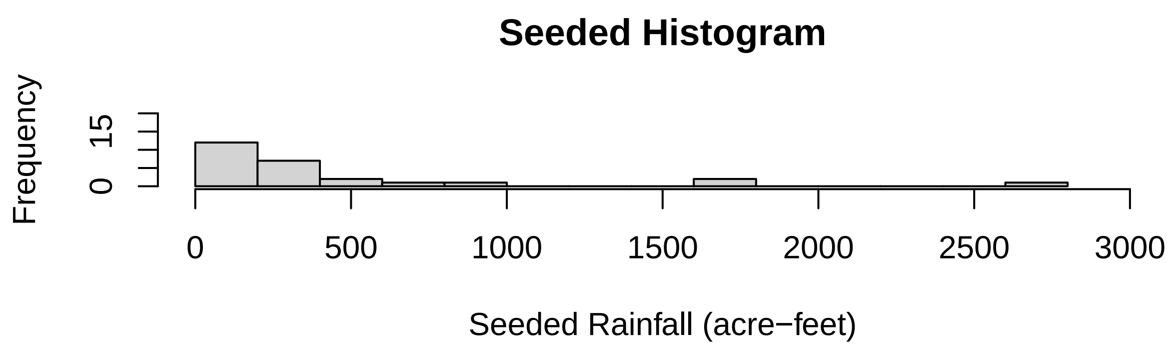

```
hist(plant_diet$weight_change, breaks = 20, main = "Narrow", xlab = "Weight Change (%)")
```




#1.c Surprisingly I think the smallest width histogram is actually the most useful as a summary. I think this because it shows the relatively even split in weight loss and gain, while also making it clear the loss was more prevalent. Also the narrow one had gaps in the histogram that made visually comparing values more difficult.

```r
library(readxl)

clouds <- read_excel("clouds.xlsx")

Seeded <- clouds[clouds$Treatment == "Seeded", ]
Unseeded <- clouds[clouds$Treatment == "Unseeded", ]

par(mfrow = c(2, 1))

hist(Unseeded$Rainfall, breaks = 10, main = "Unseeded Histogram", xlab = "Unseeded Rainfall (acre-feet)

hist(Seeded$Rainfall, breaks = 10, main = "Seeded Histogram", xlab = "Seeded Rainfall (acre-feet)", xli
```

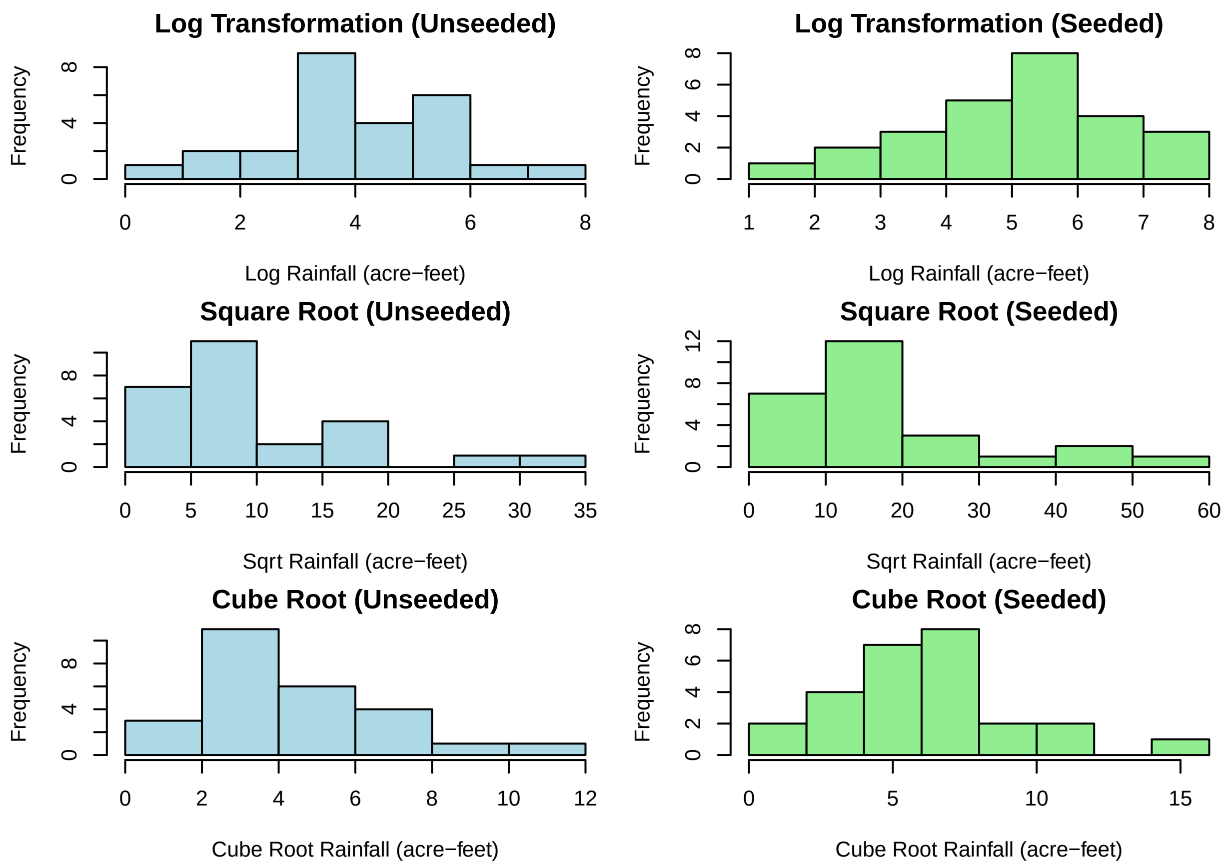



#2.a Based on the two histograms it seems like the seeded clouds produced more rainfall. There are a couple seeded clouds above 1500 acre-feet and all the unseeded fall under the 1500 mark. So I believe it was effective in increasing rainfall.

```r
library(readxl)

clouds <- read_excel("clouds.xlsx")

clouds$log_Rainfall <- log(clouds$Rainfall + 1)
clouds$sqrt_Rainfall <- sqrt(clouds$Rainfall)
clouds$cube_root_Rainfall <- clouds$Rainfall^(1/3)

par(mfrow = c(3, 2), mar = c(4, 4, 2, 1))

# Log transformation
hist(clouds$log_Rainfall[clouds$Treatment == "Unseeded"],
     main = "Log Transformation (Unseeded)", xlab = "Log Rainfall (acre-feet)",
     col = "lightblue")
hist(clouds$log_Rainfall[clouds$Treatment == "Seeded"],
     main = "Log Transformation (Seeded)", xlab = "Log Rainfall (acre-feet)",
     col = "lightgreen")

# Square root transformation
hist(clouds$sqrt_Rainfall[clouds$Treatment == "Unseeded"],
     main = "Square Root (Unseeded)", xlab = "Sqrt Rainfall (acre-feet)",
     col = "lightblue")
hist(clouds$sqrt_Rainfall[clouds$Treatment == "Seeded"],
```

```
     main = "Square Root (Seeded)", xlab = "Sqrt Rainfall (acre-feet)",
     col = "lightgreen")

# Cube root transformation
hist(clouds$cube_root_Rainfall[clouds$Treatment == "Unseeded"],
     main = "Cube Root (Unseeded)", xlab = "Cube Root Rainfall (acre-feet)",
     col = "lightblue")
hist(clouds$cube_root_Rainfall[clouds$Treatment == "Seeded"],
     main = "Cube Root (Seeded)", xlab = "Cube Root Rainfall (acre-feet)",
     col = "lightgreen")
```




#2.b I chose square rooting the values, log and finally cube root for my transformations. The most symmetric histogram came from doing a log transformation on the data set especially on the seeded dataset.