penguins

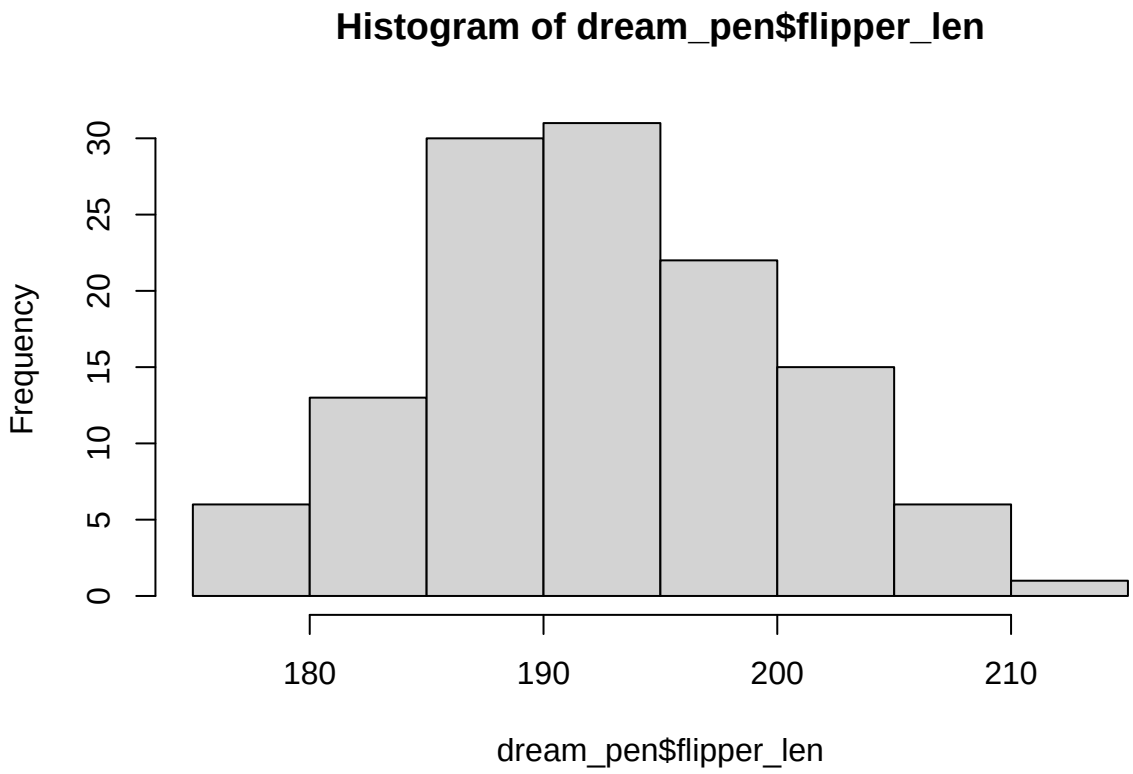
2025-11-18

Notes Example

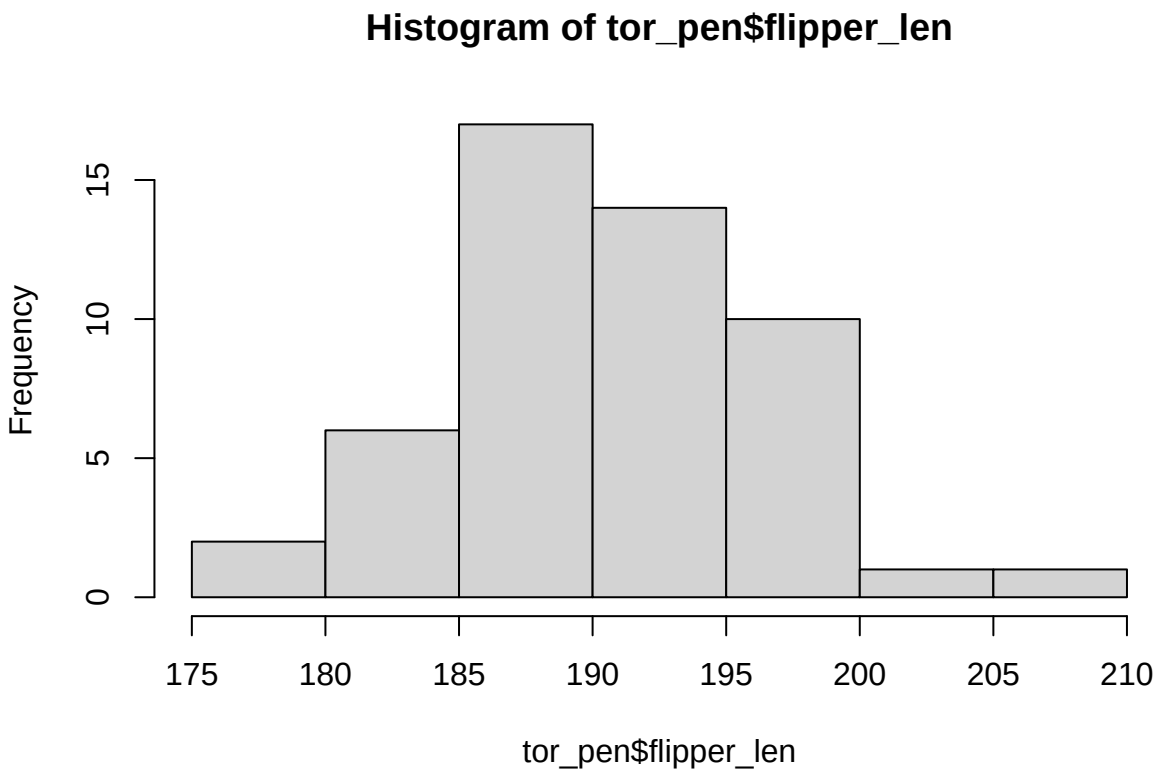
```
data("penguins")


dream_pen <- subset(penguins, island == 'Dream')
tor_pen <- subset(penguins, island == 'Torgersen')

hist(dream_pen$flipper_len)
```




```
hist(tor_pen$flipper_len)
```




```
t.test(dream_pen$flipper_len,
       tor_pen$flipper_len,
       alternative = 'two.sided',
       mu = 0,
       conf.level = .95)
```

```
##
##  Welch Two Sample t-test
##
## data:  dream_pen$flipper_len and tor_pen$flipper_len
## t = 1.7016, df = 111.37, p-value = 0.09162
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3086337  4.0616381
## sample estimates:
## mean of x mean of y
##  193.0726  191.1961
```

Example to try on your own

We are ignoring tension for the moment if you are wondering.

```r
data("warpbreaks")

yarn_a <- subset(warpbreaks, wool == 'A')
yarn_b <- subset(warpbreaks, wool == 'B')
```

1) State you null and alternative hypothesis

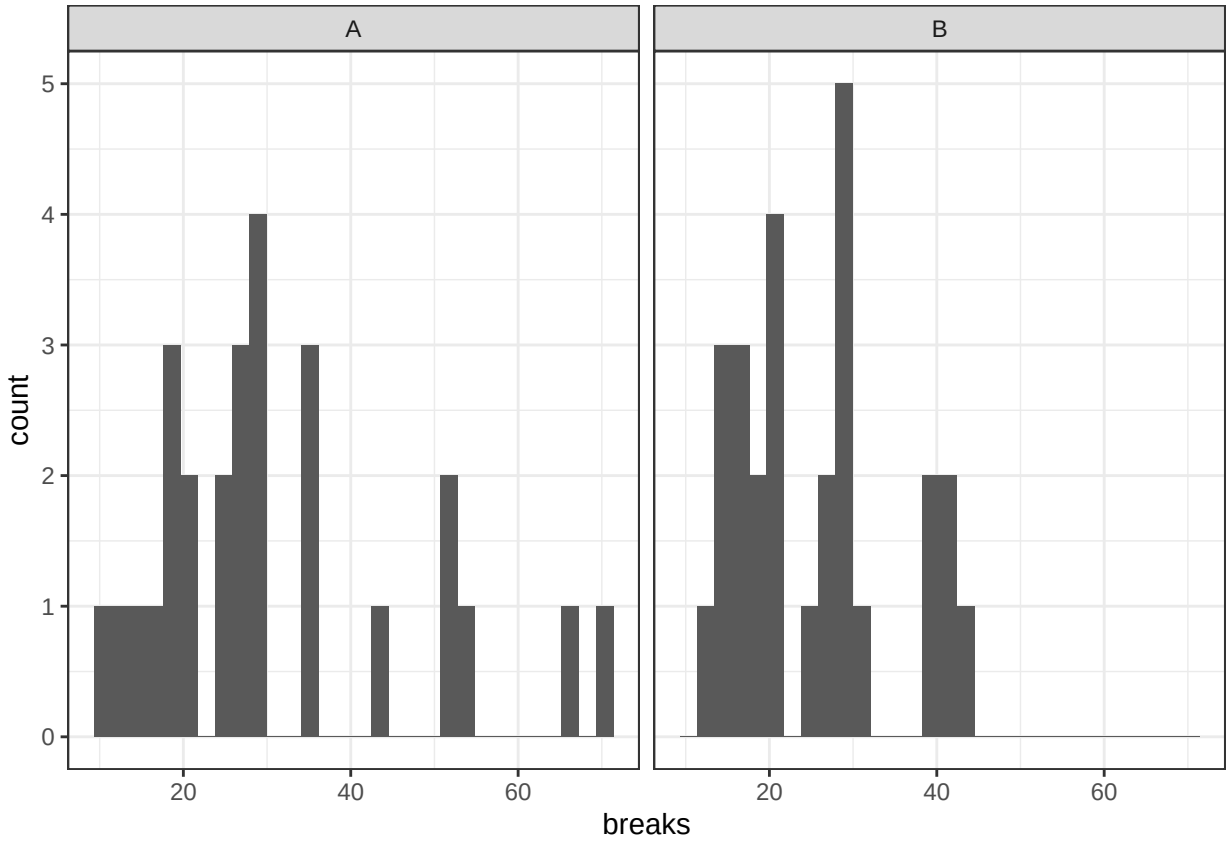
H_0: mu_A - mu_B = 0 H_A: mu_A - mu_B $\neq$ 0 (not equal to)

2) Visualize your data

```r
library(ggplot2)

ggplot(warpbreaks,
       aes(breaks)) +
  geom_histogram() +
  facet_grid(~wool) +
  theme_bw()
```

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



3) Check your assumptions (random, independent and identically distributed, large n)

i) Random: Sure
ii) IID: Yes to both of these as there isn't a reason why one bundle of yarn would affect another. Also, there isn't an a priori belief some yarn might be different so there isn't lurking variables so they should be identically distributed
iii) Large n or the pop is normal: Given the sample distributions above it looks like the data is approximately normal for both groups; at least normal enough that the sample size (n = 27 for both groups) should kick in

4) Regardless if we fail our assumptions, let's continue. Using R, calculate the test-statistic and the p-value

```
t.test(yarn_a$breaks,
       yarn_b$breaks,
       conf.level = .95)
```

```
##
##  Welch Two Sample t-test
##
## data:  yarn_a$breaks and yarn_b$breaks
## t = 1.6335, df = 42.006, p-value = 0.1098
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.360096 12.915652
## sample estimates:
## mean of x mean of y
##  31.03704  25.25926
```

5) Give a decision for your test

We have little to no evidence that the true difference in the mean number of breaks is different between the two yarns.

6) State and interpret a 95% confidence interval for the difference in their means

From the R output above: (-1.36, 12.92) so....

We are 95% confident the true difference in mean breaks between yarn A and B is between -1.36 and 12.92 breaksw.