

# Penguin flipper lengths anova

*Sean Denny*

*11/7/2018*

1. Load packages and read in the penguins.csv data

Packages: tidyverse, car, onewaytests

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.2.0
```

```
## v ggplot2 3.0.0    v purrr  0.2.5
## v tibble  1.4.2    v dplyr  0.7.7
## v tidyr   0.8.1    v stringr 1.3.1
## v readr   1.1.1    v forcats 0.3.0
```

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

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

```
library(car)
```

```
## Loading required package: carData
```

```
##
```

```
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      recode
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
##      some
```

```
library(onewaytests)
```

```
penguins <- read_csv("penguins.csv")
```

```
## Parsed with column specification:
```

```
## cols(
```

```
##   studyName = col_character(),
```

```
##   `Sample Number` = col_integer(),
```

```
##   Species = col_character(),
```

```
##   Region = col_character(),
```

```
##   `Culmen Length (mm)` = col_double(),
```

```
##   `Culmen Depth (mm)` = col_double(),
```

```
##   `Flipper Length (mm)` = col_integer(),
```

```
##   `Body Mass (g)` = col_integer(),
```

```
##   Sex = col_character()
```

```
## )
```

Our question: Is there a significant difference in mean flipper lengths for female chinstrap, gentoo, and adelic penguins near Palmer Archipelago?

Maybe I can use a one-way ANOVA... but there are assumptions tied to that: -Independent observations  
-Equal variance (Leven's test) -Normality (histograms/qplots)

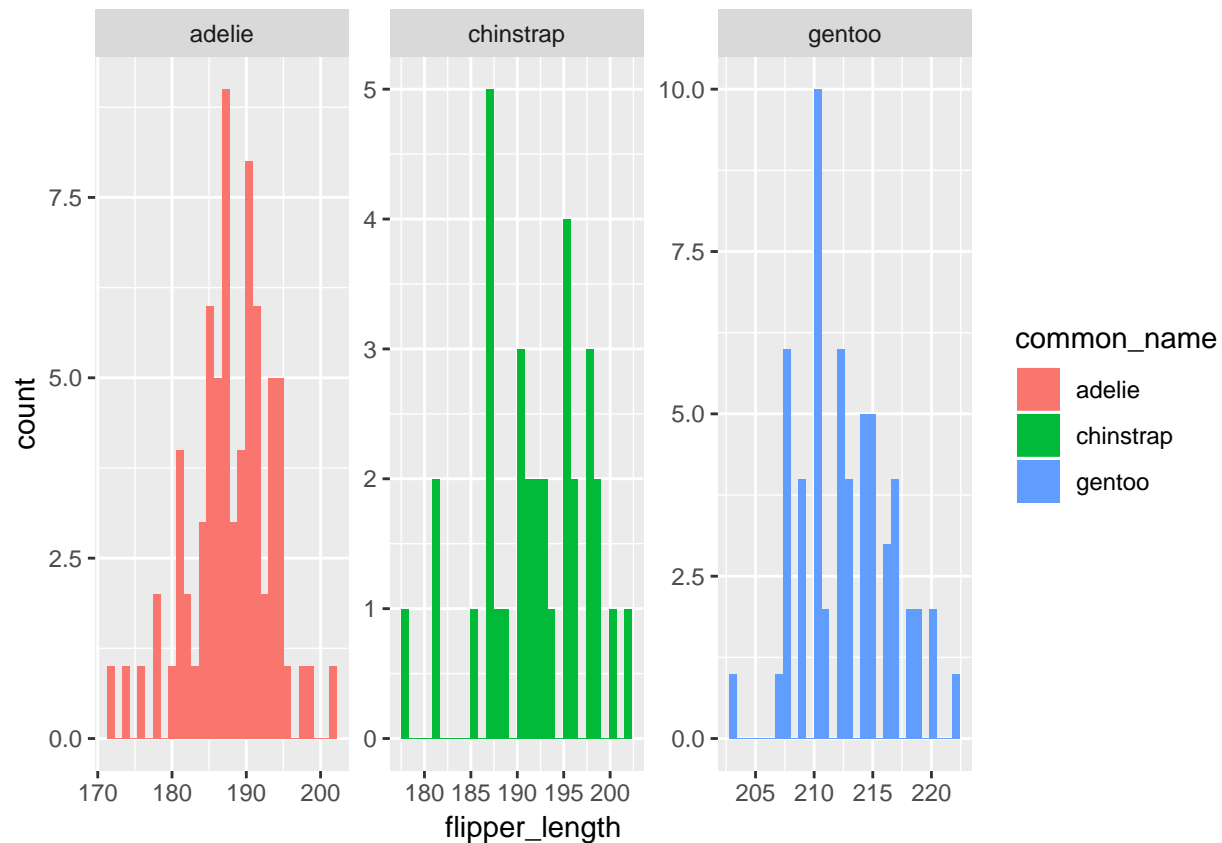
```
counts <- with(penguins, table(Species, Sex))

penguins2 <- penguins %>%
  rename(flipper_length = `Flipper Length (mm)`) %>%
  mutate(
    common_name =
      case_when(
        Species == "Chinstrap penguin (Pygoscelis antarctica)" ~ "chinstrap",
        Species == "Gentoo penguin (Pygoscelis papua)" ~ "gentoo",
        Species == "Adelie Penguin (Pygoscelis adeliae)" ~ "adelie"
      )
  ) %>%
  select(-Species) %>%
  filter(Sex == "FEMALE")
```

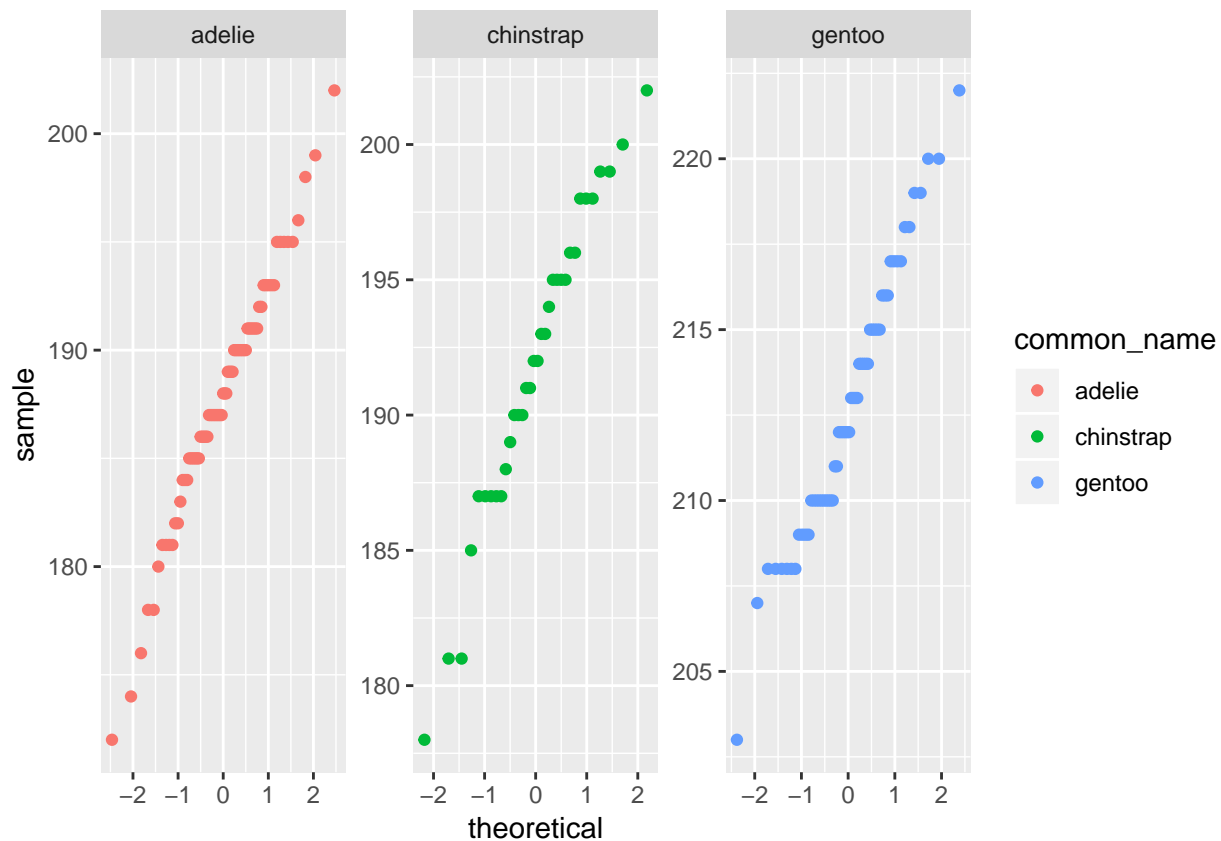
Exploratory graphs and Leven's test for equal variances

```
hists <- ggplot(penguins2, aes(x = flipper_length)) +
  geom_histogram(aes(fill = common_name)) +
  facet_wrap(~common_name, scale = "free")
hists
```

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



```
qqs <- ggplot(penguins2, aes(sample = flipper_length)) +
  geom_qq(aes(color = common_name)) +
  facet_wrap(~common_name, scale = "free")
qqs
```



*# Based on qq-plots and histograms, observations for each group appear normally distributed. Also,  $n > 30$ .*

Evaluate variances and perform a Leven's Test for equal variances.

*# If the largest sample variance is  $< 4\times$  greater than the smallest sample variance, then these are "close".*

```
variances <- penguins2 %>%
  group_by(common_name) %>%
  summarize(
    variance = var(flipper_length)
  )
#variances
```

*# The largest variance (33.1) is definitely not more than  $4\times$  greater than the smaller variance (15.2).*

```
# Leven's Test
# levenTest function in the 'car' package
```

```
levene_penguin <- leveneTest(flipper_length ~ common_name, data = penguins2)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
#levene_penguin
```

```
#H0: No difference in variances (variances are equal)
#HA: Variances are NOT equal
```

```
#There is no significant differences in variances across species.
```

One-way ANOVA for penguin flipper lengths by species (common\_name)

*#H0: Means across all groups are equal*

*#HA: At least two group means differ significantly*

```
penguin_aov <- aov(flipper_length ~ common_name, data = penguins2)
penguin_aov
```

```
## Call:
```

```
##   aov(formula = flipper_length ~ common_name, data = penguins2)
```

```
##
```

```
## Terms:
```

```
##               common_name Residuals
```

```
## Sum of Squares    21415.629  4212.553
```

```
## Deg. of Freedom         2       162
```

```
##
```

```
## Residual standard error: 5.099354
```

```
## Estimated effects may be unbalanced
```

```
summary(penguin_aov)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
```

```
## common_name  2  21416   10708   411.8 <2e-16 ***
```

```
## Residuals  162   4213     26
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*#Post-hoc testing by Tukey's HSD*

*#H0: No significant difference in pairwise means*

*#HA: There is a significant difference in pairwise means*

```
penguin_post_hoc <- TukeyHSD(penguin_aov)
penguin_post_hoc
```

```
##   Tukey multiple comparisons of means
```

```
##     95% family-wise confidence level
```

```
##
```

```
## Fit: aov(formula = flipper_length ~ common_name, data = penguins2)
```

```
##
```

```
## $common_name
```

```
##              diff          lwr          upr          p adj
```

```
## chinstrap-adelie  3.940774  1.436275  6.445272  0.0007922
```

```
## gentoo-adelie    24.912376 22.790647 27.034105 0.0000000
```

```
## gentoo-chinstrap 20.971602 18.366227 23.576977 0.0000000
```

Mean flipper lengths of female penguins differed significantly for all species studied (one-way anova;  $F(df) =$  statistic,  $p < 0.001$ ,  $\alpha = 0.05$  with post-hoc Tukey's HSD,  $\alpha = 0.05$ ).