Lab5

key

For this lab we will be starting to think about analyzing our airplane data. Clean the `Airplane` dataset and recreate a figure similar to lab 2.

```
airplane <- read_csv("https://raw.githubusercontent.com/stat441/Labs/main/airplane_clean.csv")

airplane_wide <- airplane %>%
  mutate(value = feet_dec) %>%
  dplyr::select(-feet_dec) %>%
  pivot_wider(names_from = name, values_from = value)
```

Data Visualization

```
airplane %>%
  ggplot(aes(y = feet_dec, x = name, label = id)) +
  geom_violinhalf() +
  geom_boxplot(width=0.1) +
  geom_text(position = position_jitter(seed = 1)) +
  theme_minimal() +
  ylab('Distance traveled (feet)') +
  xlab('Airplane Type') +
  ggtitle('Airplane Distance from STAT441/541 Experiment')
```

1. Write out the statistical model suggested implied by the following code.

```
lm(feet_dec ~ name, data = airplane) %>% display()
```

```
## lm(formula = feet_dec ~ name, data = airplane)
##             coef.est coef.se
## (Intercept) 13.30     1.07
## nameGlider  -5.22     1.51
## ---
## n = 40, k = 2
## residual sd = 4.78, R-Squared = 0.24
```

$$y = \beta_0 + \beta_1 x_{glider} + \epsilon \tag{1}$$

where y is the distance a plane traveled, β_0, or (`(Intercept)`), is the expected distance for the reference group `dart`, β_1, or (`nameGlider`), is the expected difference between the `dart` and the `glider`, x_{glider} is a binary dummy variable indicating the observation was a glider, and $\epsilon \sim N(0, \sigma^2)$ is an error term.

2. Write out the statistical model suggested implied by the following code.

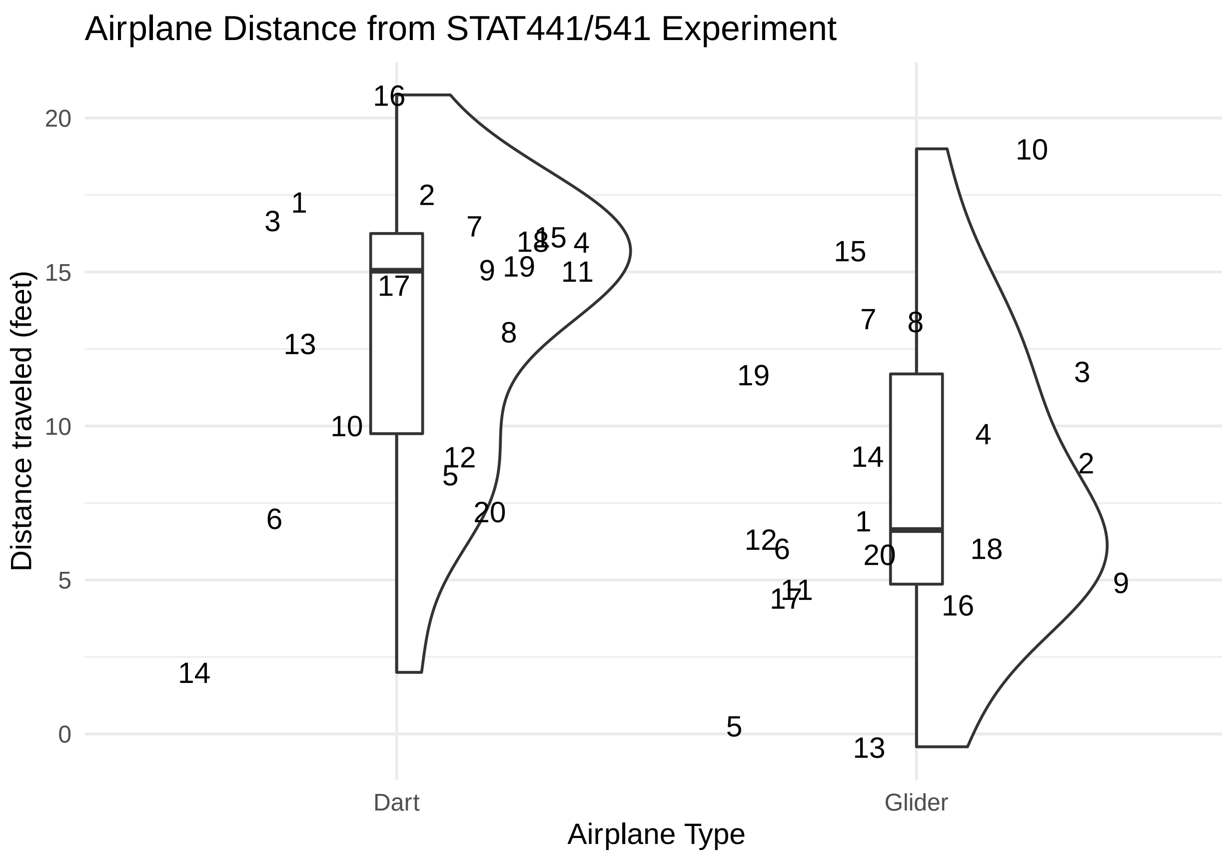

Figure 1: Distance traveled for paper airplane in STAT 441 / 541 experiment. Numbers represent unique paper airplane maker/throwers.

```r
lm(feet_dec ~ name - 1, data = airplane) %>% display()
```

```
## lm(formula = feet_dec ~ name - 1, data = airplane)
##            coef.est coef.se
## nameDart   13.30     1.07
## nameGlider  8.08     1.07
## ---
## n = 40, k = 2
## residual sd = 4.78, R-Squared = 0.85
```

$$y = \beta_0 x_{dart} + \beta_1 x_{glider} + \epsilon \tag{2}$$

where y is the distance a plane traveled, β_0, or (`nameDart`), is the expected distance for `dart`, β_1, or (`nameGlider`), is the expected distance for the `glider`, x_{dart} and x_{glider} are a binary dummy variables, and $\epsilon \sim N(0, \sigma^2)$ is an error term.

3. Do the statistical models in Q1 and Q2 account for the blocking structure of our designed experiment? If not, write out pseudo-code to include this factor using the reference case specification of Q1.

```r
summary(lm(feet_dec ~ name + factor(id), data = airplane))
```

```
##
## Call:
## lm(formula = feet_dec ~ name + factor(id), data = airplane)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.109 -2.367  0.000  2.367  7.109
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   14.6894     3.2993   4.452 0.000273 ***
## nameGlider    -5.2187     1.4399  -3.624 0.001806 **
## factor(id)2    1.0700     4.5535   0.235 0.816736
## factor(id)3    2.1283     4.5535   0.467 0.645527
## factor(id)4    0.7950     4.5535   0.175 0.863249
## factor(id)5   -7.7550     4.5535  -1.703 0.104857
## factor(id)6   -5.5800     4.5535  -1.225 0.235389
## factor(id)7    2.9200     4.5535   0.641 0.529013
## factor(id)8    1.1700     4.5535   0.257 0.799984
## factor(id)9   -2.0800     4.5535  -0.457 0.653002
## factor(id)10   2.4200     4.5535   0.531 0.601261
## factor(id)11  -2.2259     4.5535  -0.489 0.630564
## factor(id)12  -4.4135     4.5535  -0.969 0.344598
## factor(id)13  -5.9547     4.5535  -1.308 0.206576
## factor(id)14  -6.5800     4.5535  -1.445 0.164740
## factor(id)15   3.8367     4.5535   0.843 0.409950
## factor(id)16   0.3783     4.5535   0.083 0.934652
## factor(id)17  -2.5800     4.5535  -0.567 0.577622
## factor(id)18  -1.0800     4.5535  -0.237 0.815056
## factor(id)19   1.3400     4.5535   0.294 0.771737
## factor(id)20  -5.5800     4.5535  -1.225 0.235389
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.554 on 19 degrees of freedom
## Multiple R-squared:  0.6551, Adjusted R-squared:  0.292
## F-statistic: 1.804 on 20 and 19 DF,  p-value: 0.1021
```

4. Analyze the data using a paired t-test

```
t.test(x = airplane_wide$Dart, y = airplane_wide$Glider, paired = T)
```

```
##
##  Paired t-test
##
## data:  airplane_wide$Dart and airplane_wide$Glider
## t = 3.6242, df = 19, p-value = 0.001806
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.204885 8.232585
## sample estimates:
## mean of the differences
##                5.218735
```

5. Analyze the data using a t-test on the difference (Dart - Glider) for each participant

```
airplane_wide <- airplane_wide %>% mutate(diff = Dart - Glider)
t.test(airplane_wide$diff)
```

```
##
##  One Sample t-test
##
## data:  airplane_wide$diff
## t = 3.6242, df = 19, p-value = 0.001806
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.204885 8.232585
## sample estimates:
## mean of x
##  5.218735
```

6. Which of the analyses Q1 - Q5 provide the same inferences from the experiment?

Q3 - Q4 - Q5 all provide the same inferences. The include a variable (and associated test) to determine if there is an expected distance between average flight distance of the two airplanes. A paired t-test is a special case of an ANOVA model with a blocking factor. Similarly, looking just at the difference between the distance for each participant includes the same test.