Paper Airplane Lab

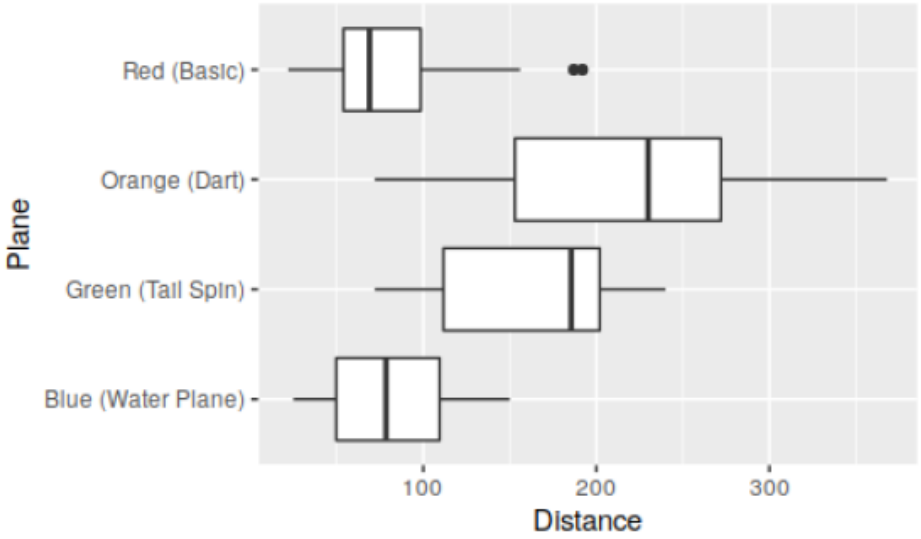
Max Kurzweil

4/16/2025

```
library(readr)
library(mosaic)
library(ggformula)

Airplane <- read_csv("Airplane.csv")
## ## Rows: 78 Columns: 2
## ── Column specification
──────────────────────────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Plane
## dbl (1): Distance
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

head(Airplane)

## # A tibble: 6 × 2
##   Plane                Distance
##   <chr>                   <dbl>
## 1 Red (Basic)                58
## 2 Green (Tail Spin)          86
## 3 Orange (Dart)             256
## 4 Blue (Water Plane)         74
## 5 Red (Basic)                48
## 6 Red (Basic)               109
```

Ho: mu(red) = mu(blue) = mu (orange) = mu(green) Ha: Not All the planes flew the same

```
gf_boxplot(Distance ~ Plane, data = Airplane)
```



The general trend I see is that the plane that flew the farthest was the Orange (Dart). The most significant difference according to this graph is between the Orange (Dart) airplane and the Red (Basic) airplane. The pair of planes that I do not expect to see a significant difference between is the Green (Tail Spin) and the Orange (Dart) since the median flight distances for both appear relatively close (around 185 and 230).

```
favstats(Distance ~ Plane, data = Airplane)

##                Plane min    Q1 median    Q3 max     mean       sd  n missing
## 1   Blue (Water Plane)  25 49.75   78.5 109.5 150  80.5000 37.81262 18       0
## 2    Green (Tail Spin)  72 111.75  185.5 202.0 240 166.1818 53.42341 22       0
## 3      Orange (Dart)  72 153.00  230.0 272.0 368 219.6111 86.07735 18       0
## 4        Red (Basic)  22 54.00   69.0  98.5 192  83.6000 48.38921 20
```

Using the rule of thumb, we check if 2 × smallest SD > Largest SD. The smallest standard deviation is 37.81, so 2 × 37.81 = 75.62 , which is not greater than the largest standard deviation of 86.08.

One significant outlier is visible in the Red (Basic) group, as indicated by the point at approximately 192 in the boxplot. This suggests we should be cautious when comparing means involving this group.

```
model <- aov(Distance ~ Plane, data = Airplane)
summary(model)

##              Df Sum Sq      Mean Sq       F value Pr(>F)
## Plane         3 256044        85348          24.8 3.28e-11 ***
## Residuals   74 254689          3442
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

256044/256044+254689 = 0.501

In terms of R^2, we can say that approximately 50.1% of the variability in distance is explained by the plane type.

Yes, we would reject Ho and confidently say that there is strong statistical evidence that not all planes fly the same distance on average ($F(3,74)=24.8$, $p<0.001$).

```
TukeyHSD(model)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Distance ~ Plane, data = Airplane)
##
## $Plane
##                                          diff         lwr         upr     p adj
## Green (Tail Spin)-Blue (Water Plane)    85.68    36.67457   134.68907 0.0001010
## Orange (Dart)-Blue (Water Plane)       139.11    87.71187   190.51035 0.0000000
## Red (Basic)-Blue (Water Plane)           3.10   -46.99778    53.19778 0.9984551
## Orange (Dart)-Green (Tail Spin)         53.43     4.42204   102.43655 0.0271839
## Red (Basic)-Green (Tail Spin)          -82.58  -130.22231   -34.94132 0.0001167
## Red (Basic)-Orange (Dart)             -136.01  -186.10889   -85.91333 0.0000000
```

Based on the multiple comparisons test, the only pair that does not show a significant difference is Red (Basic)-Blue (Water Plane) ($p=0.998$). All other pairwise comparisons are statistically significant at $\alpha=0.05$. The largest significant difference is between Red (Basic) and Orange (Dart), with a mean difference of 136.01 feet ($p<0.001$).