```
library("dplyr")
library("rcompanion")
library("car")
library("ggplot2")
library("IDPmisc")

# See if the number of people who cross differs by the way they cross

# Tes ng Assump ons

plotNormalHistogram(BorderCrossing$Value)

## That is not normally distributed at all - posi vely skewed

### Square root

BorderCrossing$ValueSQRT <- sqrt(BorderCrossing$Value)

plotNormalHistogram(BorderCrossing$ValueSQRT)

### Log

BorderCrossing$ValueLOG <- log(BorderCrossing$Value)

BorderCrossing2 <- NaRV.omit(BorderCrossing)

plotNormalHistogram(BorderCrossing2$ValueLOG)

#### Log looks great, go with that

## Bartle 's Test for homogeneity of variance

bartle .test(ValueLOG ~ Measure, data=BorderCrossing2)

### If it's significant, that's bad, you have violated homogeneity (like we have here)

## Sample size - we have 355,000 something rows, so good there

# Run the test, use the Welch's test because we violated homogeneity of variance

ANOVA <- lm(ValueLOG ~ Measure, data=BorderCrossing2)
Anova(ANOVA, Type="II", white.adjust=TRUE)

## It is significant, meaning that the number of people does vary based on their method of travel across the
border

# Post hocs
```

```
pairwise.t.test(BorderCrossing$Value, BorderCrossing$Measure, p.adjust="bonferroni", pool.sd = FALSE)

## Every single group differs from every single other group. That is so not helpful!

# Look at the means

crossingMeans <- BorderCrossing %>% group_by(Measure) %>% summarize(Mean = mean(Value)) %>%
arrange(desc(Mean))

# The fewest people come by trains, the most come by car.  This needs more wrangling to get the best picture of
all, though!
```