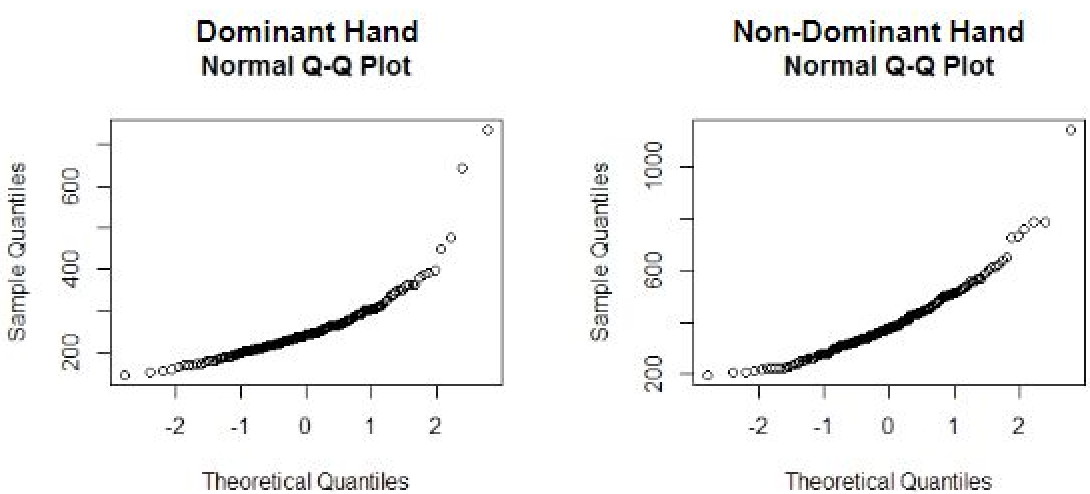
Homework 3 || Hand Results

Assessing Normality

To assess normality, I used the Shapiro test and made a few qq plots. For the Shapiro test, I got results both for the dominant hand, and the non-dominant hand, with results shown below:

```
> by(df$Time, df$Hand, shapiro.test)
-----------------------------------------------------
df$Hand: dominantHand
      Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.79619, p-value = 1.423e-14
-----------------------------------------------------
df$Hand: nonDominantHand
      Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.90627, p-value = 2.814e-09
-----------------------------------------------------
```

The important thing to note from these results is the p-values (bolded) returned by this test. As you can see from the results, both the p-values are below 0.05. This indicates that it is not likely that the data displays normal behavior (that is, that the shape of the data does not resemble a Gaussian curve). In addition to the shapiro test, I also created a few QQ plots, shown below:



These QQ-plots reaffirm that the data is not normal in nature. A QQ-plot that displays normal data is very linear, and these QQ-plots are more sloped than linear.

Therefore, the assumption that the data is normal is a false assumption. This is an indicator that the data is non-parametric.

Assessing Homogeneity of Variance

To assess homogeneity of variance, I used the Fligner-Killeen test; this test is very robust because it does not require that the data be normally distributed. This is ideal for our situation since, as we proved in the section above, our data does not appear to be normally distributed. The results of the Fligner-Killeen test is shown below:

```
> fligner.test(df$Time ~ df$Hand, data = df)
-----------------------------------------------------
      Fligner-Killeen test of homogeneity of variances
data:  df$Time by df$Hand
Fligner-Killeen:med chi-squared = 50.723, df = 1, p-value = 1.064e-12
-----------------------------------------------------
```

Again, the important thing to note from these results is the p-value (bolded) returned by this test. As you can see from the results, the p-values is, again, well below 0.05. This indicates that it is not likely that the data displays homogeneity of variance.

Therefore, the assumption that the data has homogeneity of variance is a false assumption. This is an indicator that the data is non-parametric.

Assessing Parametric-ness

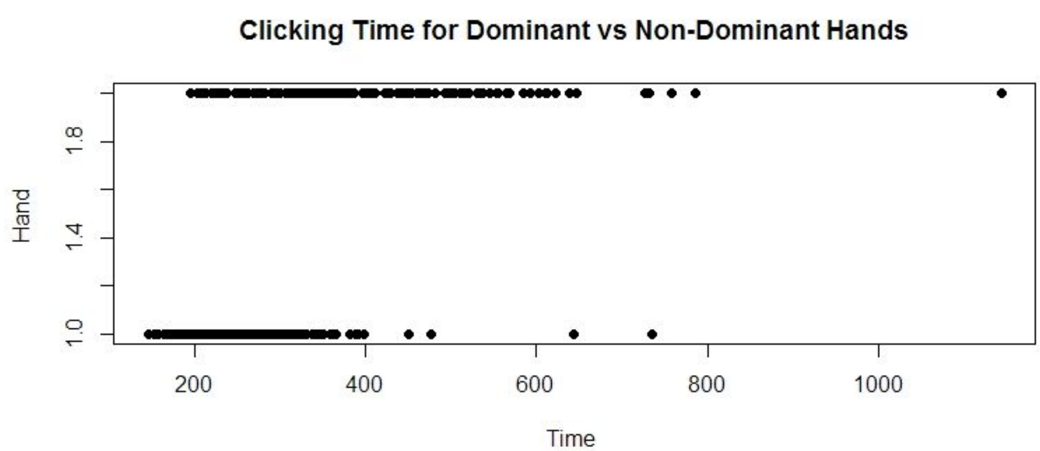
Since it appears that our data has neither a normal distribution nor homogeneity of variance, it seems like we can safely say that our data is non-parametric. This makes sense since our sample size was very small and we only had one participant who took the survey several times. One person is not very representative of the millions of people in a population. If this study had more than 10 participants and each participant was actually a different person, and we, for example, sourced our participants from Mechanical Turk online to get a nice variety of participants, it is far more likely that we would have gotten parametric results.

Assessing Correlation

To assess correlation, I used the Kendall's tau test, which is designed to assess the correlation of non-parametric data. These tests are ideal for this situation because our results were non-parametric. Additionally, since our data is ordinal (aka one axis is broken up into discrete categories) this means that there are lots of "ties", or spots where different data points have the same value. The Kendall test handles this very well, whereas other tests, such as the Spearman test, do not. The results of the Kendall's tau test are shown below:

```
> cor.test(df$Time, as.numeric(df$Hand), method = "kendall")
-----------------------------------------------------
      Kendall's rank correlation tau
data:  df$Time and as.numeric(df$Hand)
z = 12.124, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.5239048
-----------------------------------------------------
```

The important thing to note from these results is the p-value (bolded) and the tau value (also bolded) returned by this test. As you can see from the results, the tau value is positive and has a fairly high magnitude (the magnitude of this variable ranges from [0, 1]). The high magnitude indicates that there is a monotonic relationship between these two variables, and the positive sign indicates that the relationship is increasing. The p-value returned by this test is well below 0.05. This indicates that it is very likely that this relationship was caused by the dependant variable (which hand was used) rather than by chance.



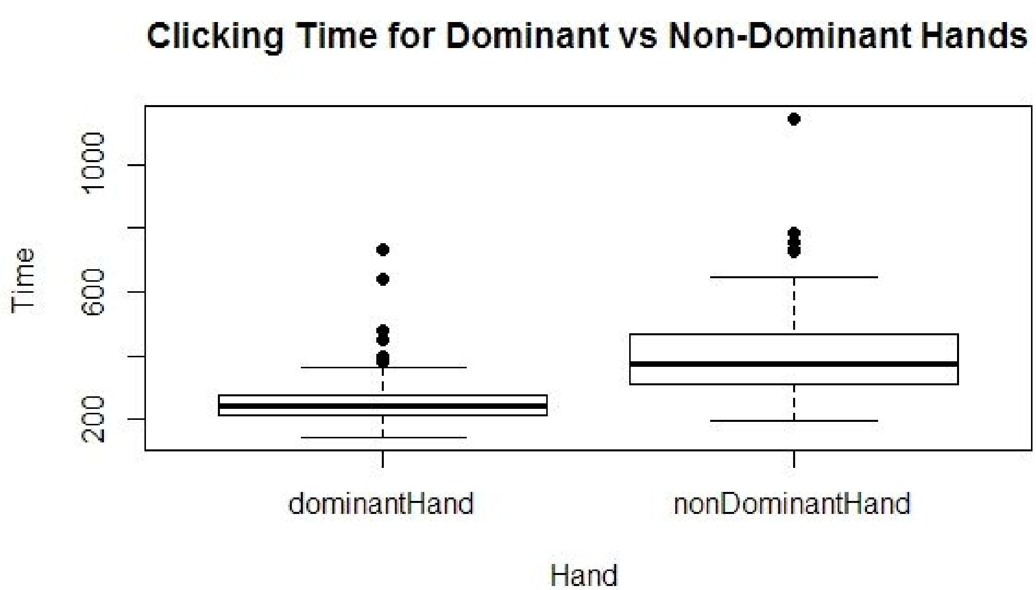
As you can see from the scatterplot shown at the bottom of the previous page, where the dominant hand is represented by 1 on the y axis and the non-dominant hand is represented as 2 on the y axis, it does look like the clicking time is, on average, higher when you use the non-dominant hand. This confirms the results of Kendall's Tau test. Therefore, we can say with some degree of certainty that there is a relationship between what hand you use and the amount of time it takes to click on objects.

Assessing Difference Between Means

To assess the difference between means, I used ANOVA, specifically the Kruskal Wallis Test One Way Anova by Ranks, which allows for non-parametric data, making it a good fit for our situation. The results of the test are shown below:

```
> kruskal.test(df$Time ~ df$Hand)
-----------------------------------------------------
      Kruskal-Wallis rank sum test
data:  df$Time by df$Hand
Kruskal-Wallis chi-squared = 146.99, df = 1, p-value < 2.2e-16
-----------------------------------------------------
```

The important thing to note from these results is the p-value (bolded). As you can see from the results, the p-values is well below 0.05. This indicates that it is very likely that the difference in means was caused by the dependant variable (which hand was used) rather than by chance, hence disproving the null hypothesis. The box plot below is a nice visualization of these results.



Homework 3 || Fitts Law Results

Assessing Normality

To assess normality, I used the Shapiro test and made a few qq plots. For the this test, I got results both for each category of target size and for each category of target distance; I will discuss both separately. The results for size are shown below:

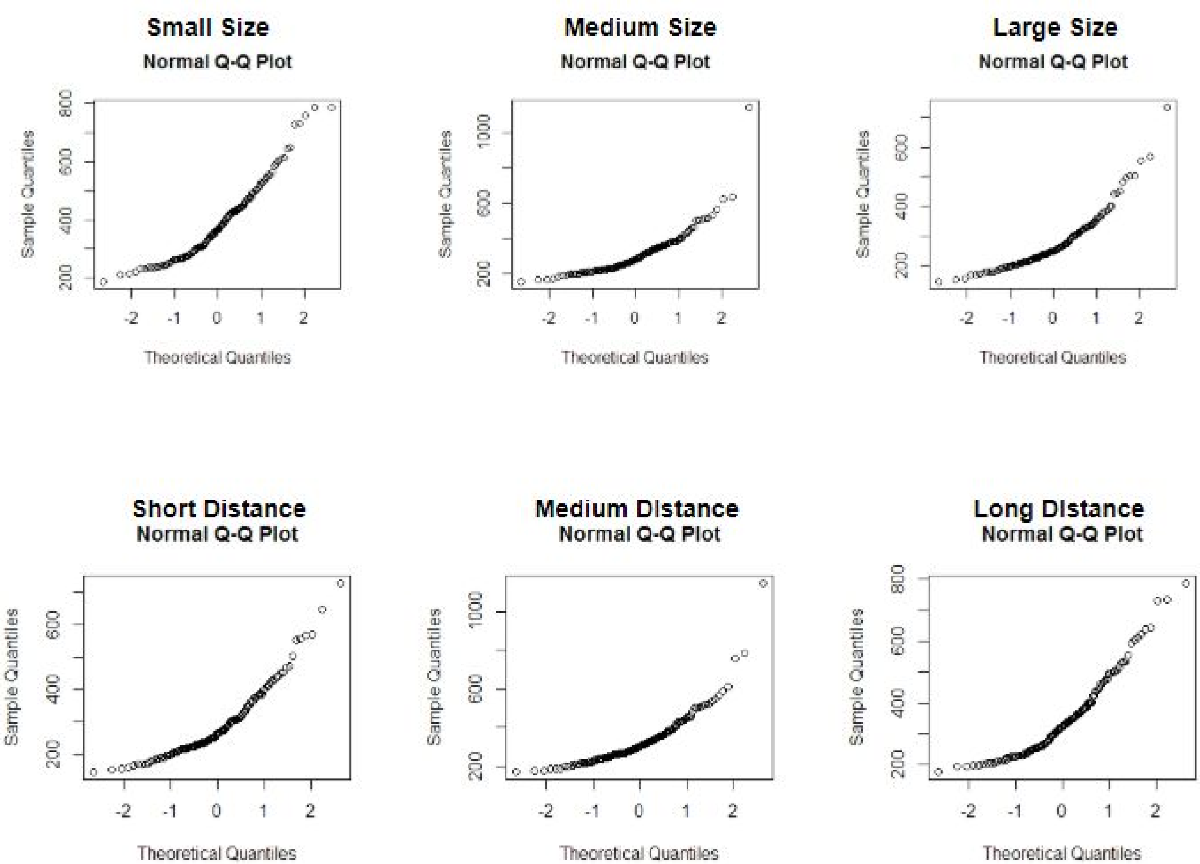
```
> by(df$Time, df$Size, shapiro.test)
-----------------------------------------------------
df$Size: 30
      Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.92783, p-value = 7.068e-06
-----------------------------------------------------
df$Size: 45
      Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.78241, p-value = 4.733e-12
-----------------------------------------------------
df$Size: 90
      Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.8608, p-value = 3.057e-09
-----------------------------------------------------
```

The important thing to note from these results is the p-values (bolded) returned by this test. As you can see from the results, all of the p-values are below 0.05. This indicates that it is not likely that the time-vs-size data displays normal behavior (that is, that the shape of the data does not resemble a Gaussian curve). The results for distance are shown below:

```
> by(df$Time, df$Distance, shapiro.test)
-----------------------------------------------------
df$Distance: 330
      Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.88924, p-value = 5.739e-08
-----------------------------------------------------
```

```
----------------------------------------------------
df$Distance: 455
      Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.81133, p-value = 4.156e-11
----------------------------------------------------
df$Distance: 770
      Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.90577, p-value = 3.919e-07
----------------------------------------------------
```

The important thing to note from these results is the p-values (bolded) returned by this test. As you can see from the results, all of the p-values are below 0.05. This indicates that it is not likely that the time-vs-distance data displays normal behavior (that is, that the shape of the data does not resemble a Gaussian curve). I also created a few QQ plots, shown below:



These QQ-plots reaffirm that the data is not normal in nature. A QQ-plot that displays normal data is very linear, and these QQ-plots are more sloped than linear.

Therefore, the assumption that the size and distance data is normal is a false assumption. This is an indicator that the data is non-parametric.

Assessing Homogeneity of Variance

To assess homogeneity of variance, I used the Fligner-Killeen test; this test is very robust because it does not require that the data be normally distributed. This is ideal for our situation since, as we proved in the section above, our data does not appear to be normally distributed. The results of the Fligner-Killeen test for the size attribute is shown below:

```
> fligner.test(df$Time ~ df$Size, data = df)
-----------------------------------------------------
      Fligner-Killeen test of homogeneity of variances
data:  df$Time by df$Size
Fligner-Killeen:med chi-squared = 21.681, df = 2, p-value = 1.959e-05
-----------------------------------------------------
```

Again, the important thing to note from these results is the p-value (bolded) returned by this test. As you can see from the results, the p-values is, again, well below 0.05. This indicates that it is not likely that the time-vs-size data displays homogeneity of variance. The results of the Fligner-Killeen test for the distance attribute is shown below:

```
> fligner.test(df$Time ~ df$Distance, data = df)
-----------------------------------------------------
      Fligner-Killeen test of homogeneity of variances
data:  df$Time by df$Distance
Fligner-Killeen:med chi-squared = 4.7705, df = 2, p-value = 0.09207
-----------------------------------------------------
```

As you can see from these results, the p-values is actually above 0.05, although not by much. This indicates that the time-vs-distance data displays homogeneity of variance.

Therefore, the assumption that the data has homogeneity of variance is a false assumption for the size attribute, but a true assumption for the distance attribute. This is an indicator that the size-vs-time data is non-parametric, but that the distance-vs-time data may be parametric.

Assessing Parametric-ness

Since it appears that neither kinds of data (size and distance) have normal distribution, this means our data must not be parametric. The non-homogeneity of variance of the size data supports this view. While the distance data does show homogeneity of variance, it does not matter since breaking normality is enough to disprove the parametric assumption.

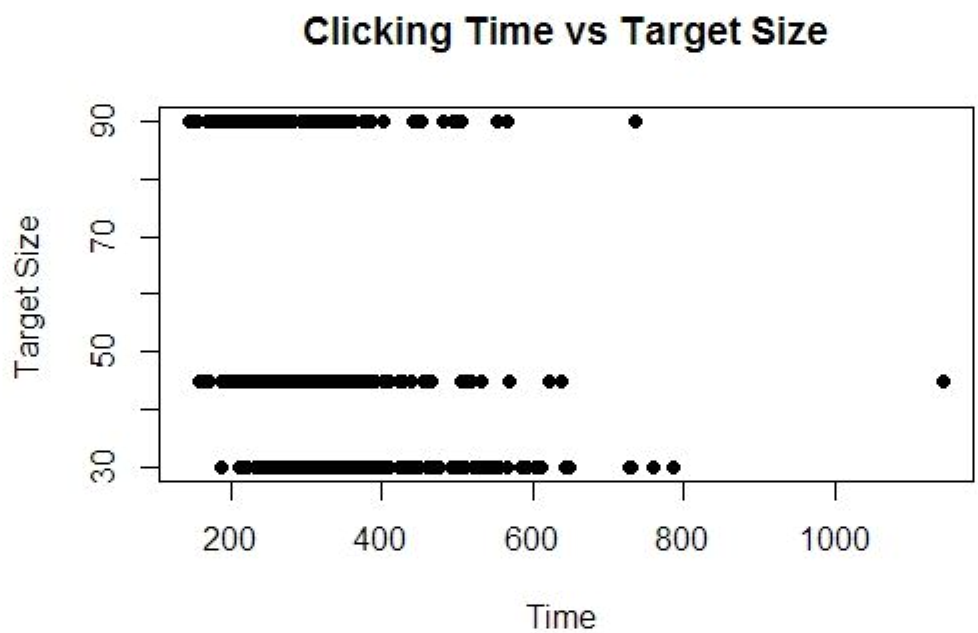
This makes sense since our sample size was very small and we only had one participant who took the survey several times. One person is not very representative of the millions of people in a population. If this study had more than 10 participants and each participant was actually a different person, and we, for example, sourced our participants from Mechanical Turk online to get a nice variety of participants, it is more likely that we would have gotten parametric results.

Assessing Correlation

To assess correlation, I used the Kendall's tau test, which is designed to assess the correlation of non-parametric data. These tests are ideal for this situation because our results were non-parametric. Additionally, since our data is ordinal (aka one axis is broken up into discrete categories) this means that there are lots of "ties", or spots where different data points have the same value. The Kendall test handles this very well, whereas other tests, such as the Spearman test, do not. The results of the Kendall's tau test on size are shown below:

```
> cor.test(df$Time, as.numeric(df$Size), method = "kendall")
-----------------------------------------------------
      Kendall's rank correlation tau
data:  df$Time and as.numeric(df$Size)
z = -7.7035, p-value = 1.324e-14
alternative hypothesis: true tau is not equal to 0
sample estimates:
       tau
-0.3138989
-----------------------------------------------------
```

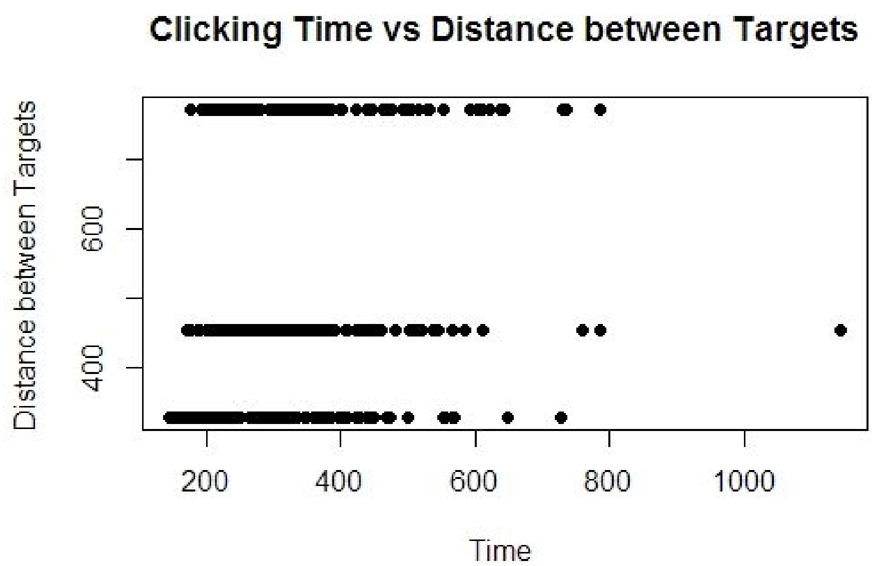
The important thing to note from these results is the p-value (bolded) and the tau value (also bolded) returned by this test. As you can see from the results, the tau value is negative and has some small magnitude (the magnitude of this variable ranges from [0, 1]). The magnitude indicates that there is a slight monotonic relationship between these two variables, and the negative sign indicates that the relationship is decreasing. The p-value returned by this test is well below 0.05. This indicates that it is very likely that this relationship was caused by the dependant variable (size of the target) rather than by chance.



As you can see from the scatterplot shown above, it does look like the clicking time is slightly faster when the target size is bigger, and gets slower as the size of the target decreases, a negative relationship. This confirms the results of Kendall's Tau test. Therefore, we can say with some degree of certainty that there is a slight relationship between the amount of time it takes to click on targets and the size of the target. The results of the Kendall's tau test on distance are shown below:

```
> cor.test(df$Time, as.numeric(df$Distance), method = "kendall")
-----------------------------------------------------
      Kendall's rank correlation tau
data:  df$Time and as.numeric(df$Distance)
z = 3.8555, p-value = 0.0001155
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.1571011
-----------------------------------------------------
```

The important thing to note from these results is the p-value (bolded) and the tau value (also bolded) returned by this test. As you can see from the results, the tau value is positive and has some small magnitude (the magnitude of this variable ranges from [0, 1]). The magnitude indicates that there is a very slight monotonic relationship between these two variables, and the positive sign indicates that the relationship is increasing. The p-value returned by this test is well below 0.05. This indicates that it is very likely that this relationship was caused by the dependant variable (size of the target) rather than by chance.



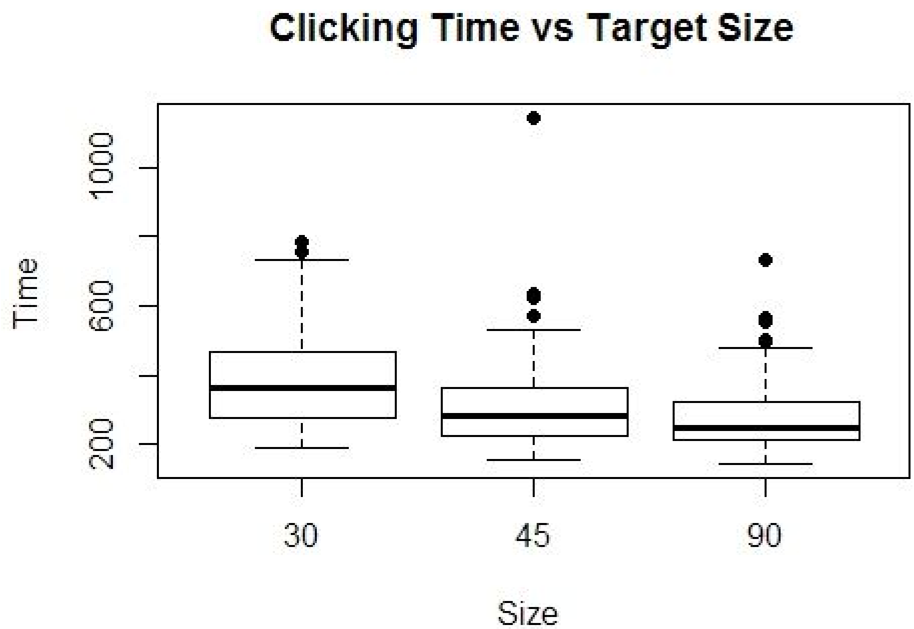
As you can see from the scatterplot shown above, it does look like the clicking time is, on average, very slightly faster when the distance between the targets is smaller, and that time increases as the distance between the targets increases. This confirms the results of Kendall's Tau test. Therefore, we can say with some degree of certainty that there is some small relationship between what hand you use and the amount of time it takes to click on objects.

Assessing Difference Between Means

To assess the difference between means, I used ANOVA, specifically the Kruskal Wallis Test One Way Anova by Ranks, which allows for non-parametric data, making it a good fit for our situation. The results of the size-vs-time test are shown below:

```
> kruskal.test(df$Time ~ df$Size)
-----------------------------------------------------
      Kruskal-Wallis rank sum test
data:  df$Time by df$Size
Kruskal-Wallis chi-squared = 60.09, df = 2, p-value = 8.948e-14
-----------------------------------------------------
```

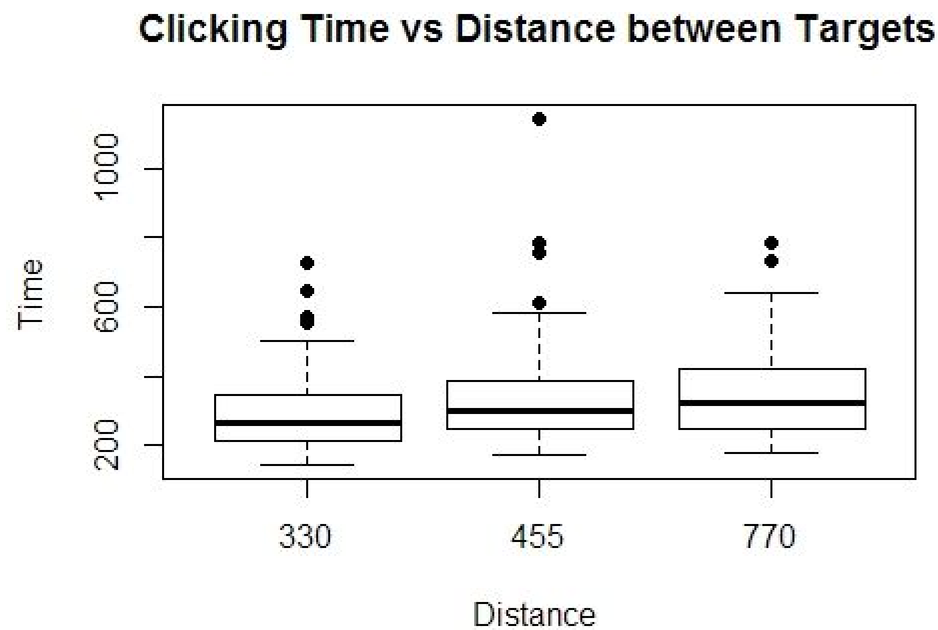
The important thing to note from these results is the p-value (bolded). As you can see from the results, the p-values is well below 0.05. This indicates that it is very likely that the difference in means was caused by the dependant variable (the size of the target) rather than by chance, hence disproving the null hypothesis. The box plot below is a nice visualization of these results.



Notice how the means in the graph above are markedly different, and have a decreasing relationship. The results of the distance-vs-time test are shown below:

```
> kruskal.test(df$Time ~ df$Distance)
--------------------------------------------------
      Kruskal-Wallis rank sum test
data:  df$Time by df$Distance
Kruskal-Wallis chi-squared = 16.457, df = 2, p-value = 0.000267
--------------------------------------------------
```

The important thing to note from these results is the p-value (bolded). As you can see from the results, the p-values is well below 0.05. This indicates that it is very likely that the difference in means was caused by the dependant variable (the distance between targets) rather than by chance, hence disproving the null hypothesis. The box plot at the top of the following page is a nice visualization of these results. Note how the means are markedly different, and have an increasing relationship.


Clicking Time vs Distance between Targets
Time
200
600
1000
330
455
770
Distance