

Distributed Server Project

Matthew Angle, Josh Wood

Overview

The purpose of this project was to learn how to use CORBA to set up a distributed system across servers and measure the difference in transfer time from one server to another. The tools we used for this experiment were provided but included a Java implemented server and client. CORBA stands for “Common Object Request Broker Architecture”, CORBA provides “vendor-independent architecture and infrastructure that computer applications use to work together over networks”.^[1] The Java program we were provided requires us to first setup a ORBD Daemon which the Java program will then use to establish a connection between the server and the client.

Experiment Conditions

Clipper is a server located at Shippensburg University. Clipper is an Ubuntu 18.04.4 server with Linux Kernel 4.15.0-88-generic, it has an Intel Xeon 2680 v3 and has 20 CPU cores and 24 GB's of ram, which have been assigned to the virtual machine running Clipper. The ORBD daemon was started on CLipper and our team used ports 1054 and 1055. Java 8 was used to build and run the server and the client.

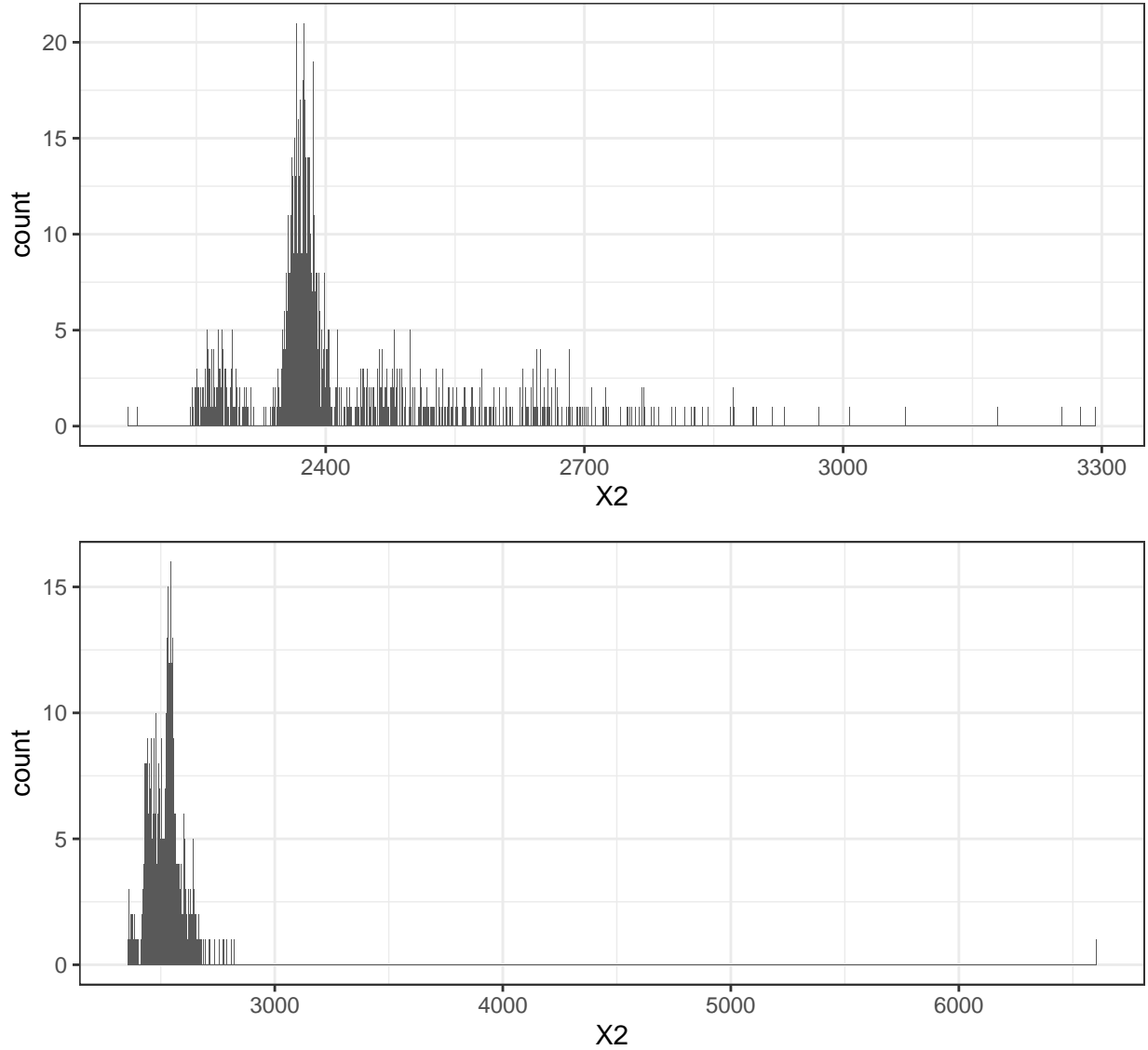
The German server is a remote virtual server hosted in Germany and uses an Intel Xeon e5-2660 v4 but is assigned one virtual core and 1 GB of ram. The server runs Ubuntu 18.04.5 with Linux Kernel 4.15.0-122-generic x86_64. The ORBD daemon was started on CLipper and our team used ports 1054 and 1055. Java 8 was used to build and run the server and the client.

The Spanish server is also a remote virtual server hosted in Spain instead of Germany. It uses an Intel Xeon E5-2683 v3 but is also assigned one virtual core and 1 GB of ram. The server runs Ubuntu 18.04.5 with Linux Kernel 4.15.0-122-generic x86_64. The ORBD daemon was started on CLipper and our team used ports 1054 and 1055. Java 8 was used to build and run the server and the client.

First Experiment

We initially set up our experiment using a bash script that would start and stop the ORBD Daemon and the Java Server, then run the client and measure the time the client took to complete the file transfer. In our experiment we used a text document full of random characters that was 1 Mb in size, we stored that on our server in the appropriate folder in the Java Server build path and then transferred that file to our client. We ran this bash script 1000 times on the German server and the Spanish server and stored the results in a csv.

First Results



Second Experiment

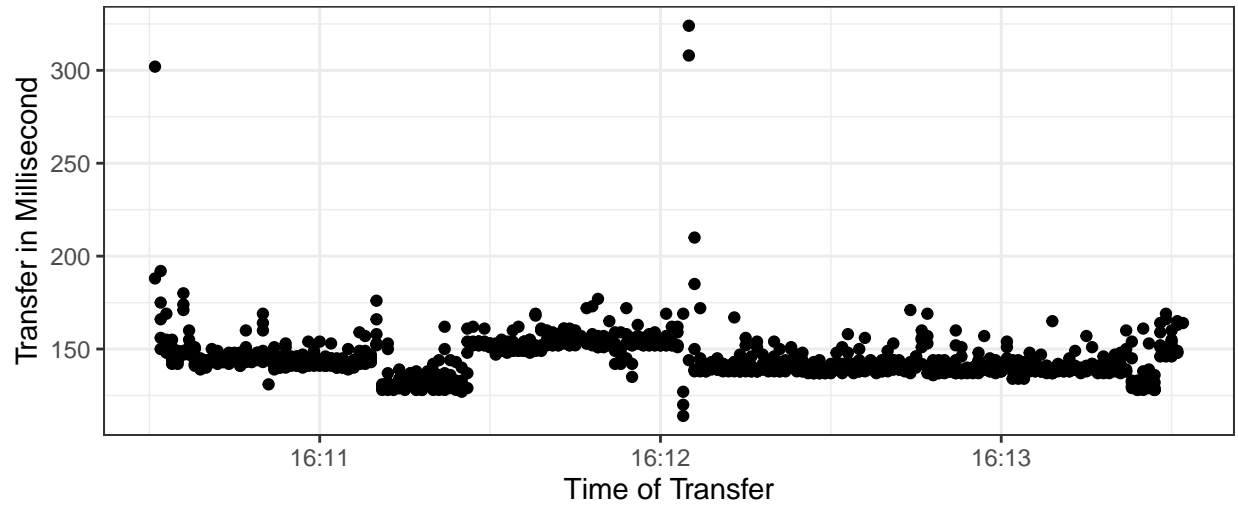
After some consideration we became concerned that the differences in compile time on the each server may have influenced our results and rendered them void.

We set up a second experiment where we transferred in the same manner a 2 MB text file, but instead of measuring total time to compile and transfer we only measured the time that it took to transfer using a Java method to measure the start and stop time of the transfer file method. We nested this method in a loop and ran it 1000 times on Clipper, the German and the Spanish server. We ran these tests 3 times each for each server and stored the results in corresponding csv files. The Clipper test was ran as a baseline to see best case scenarios for the transfer. The file size chose was intended to be large enough to be significant but not too large to be overwhelming for the remote servers. We ran the tests 3 times on each server to ensure we could have a large enough sample size to make significant analysis.

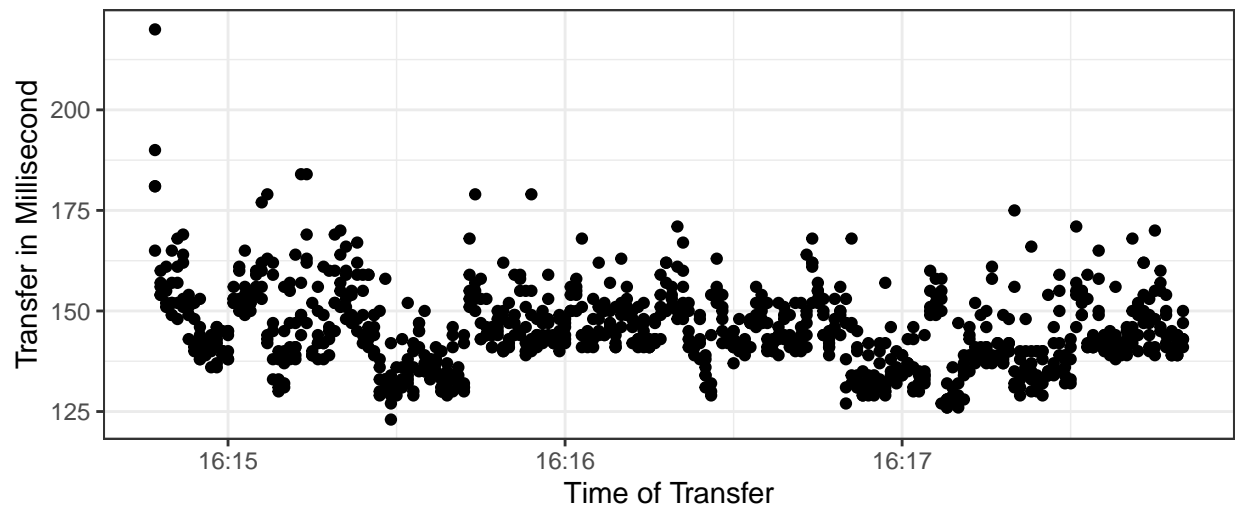
Clipper Server Results

We started analysis by looking at a point plot of the data over time.

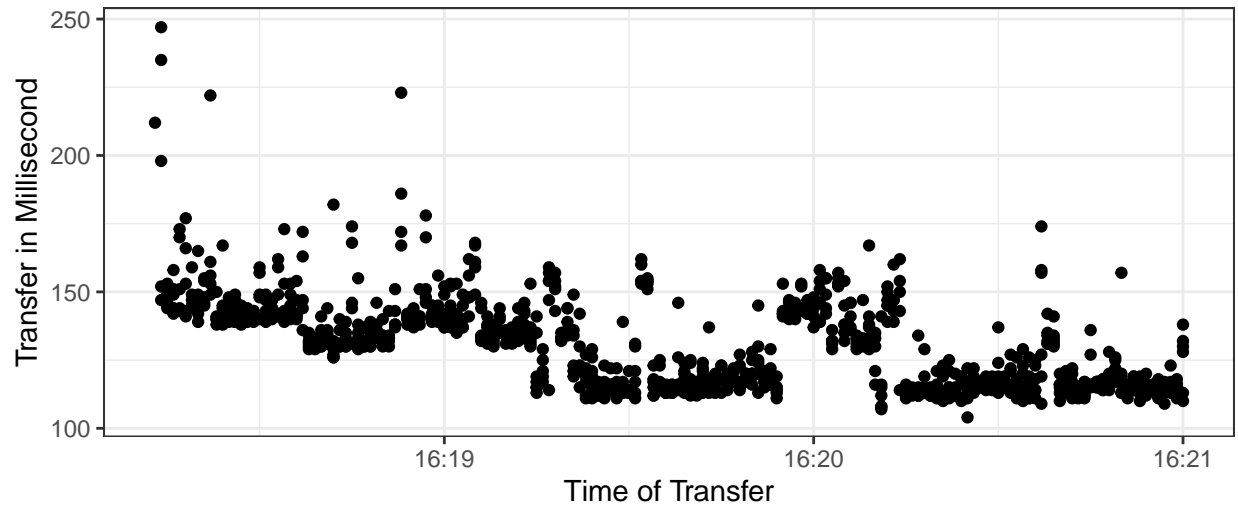
Clipper Results 1



Clipper Results 2

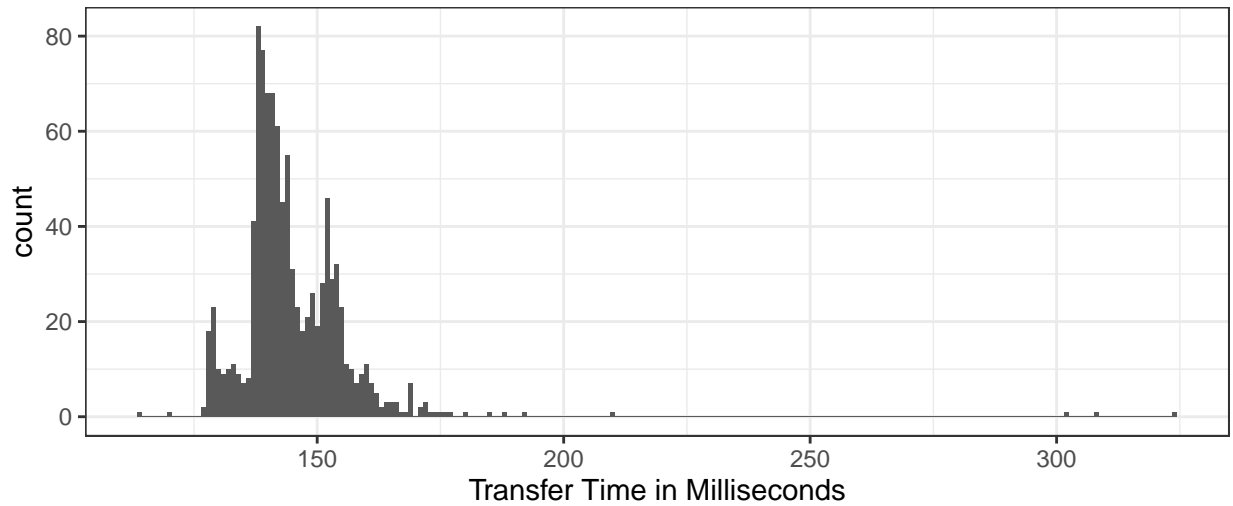


Clipper Results 3

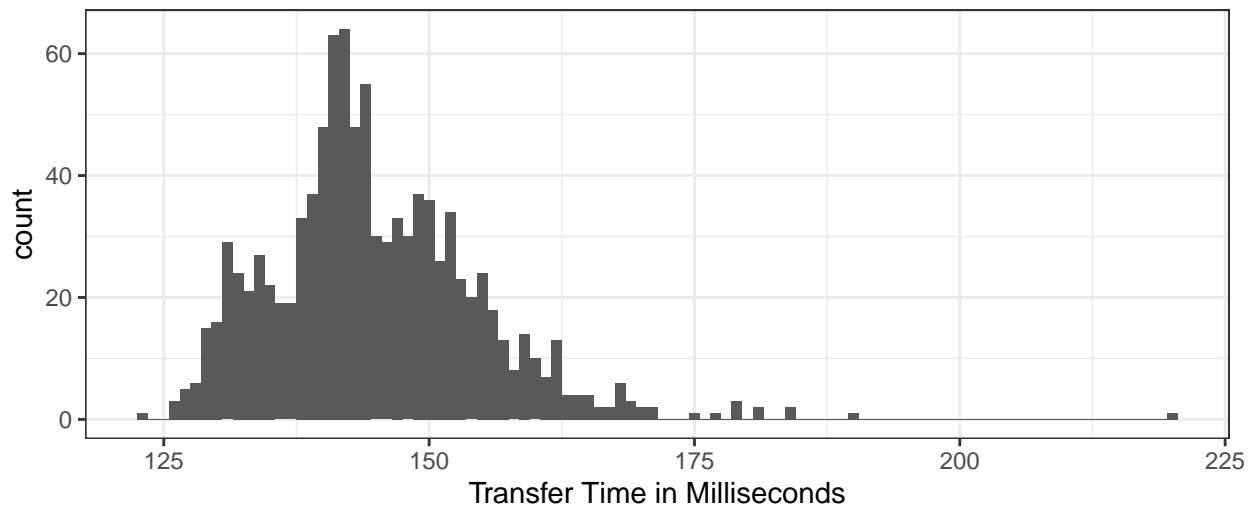


The time series data seems to show a mostly flat line across the first two tests with a few outliers except for the third plot which seems to stepped into two separate lines. To get a better look at these and to see if the data was normally distributed we made histograms using the data.

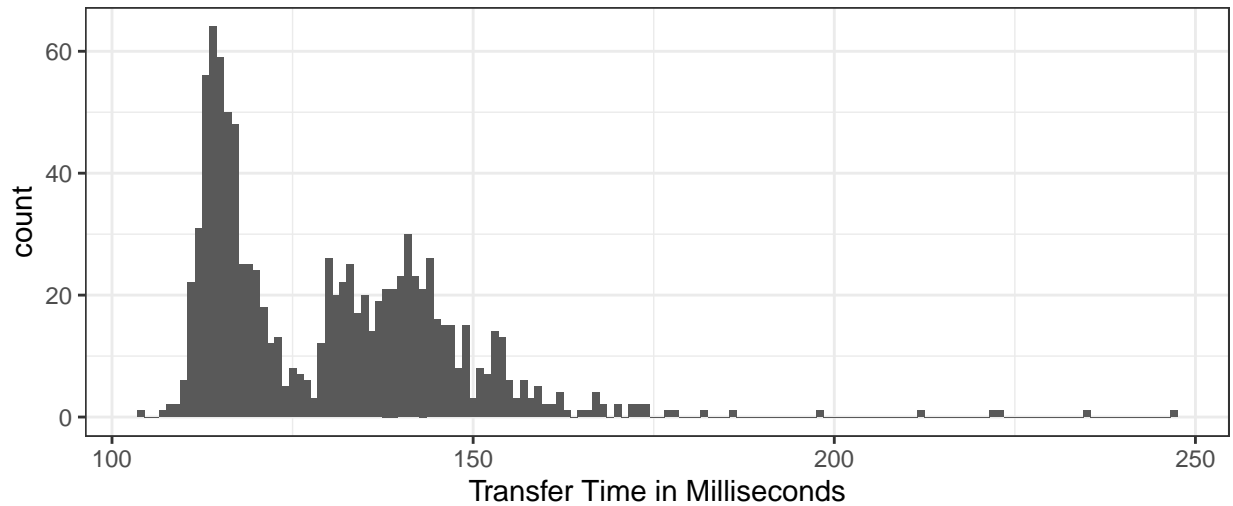
Clipper Transfer Time Histogram 1



Clipper Transfer Time Histogram 2

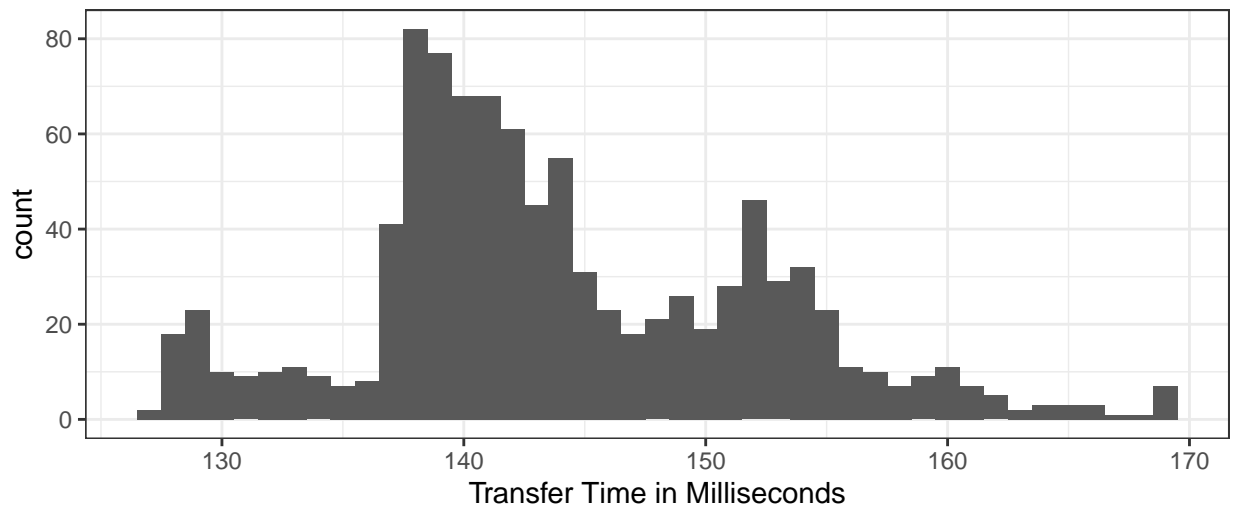


Clipper Transfer Time Histogram 3

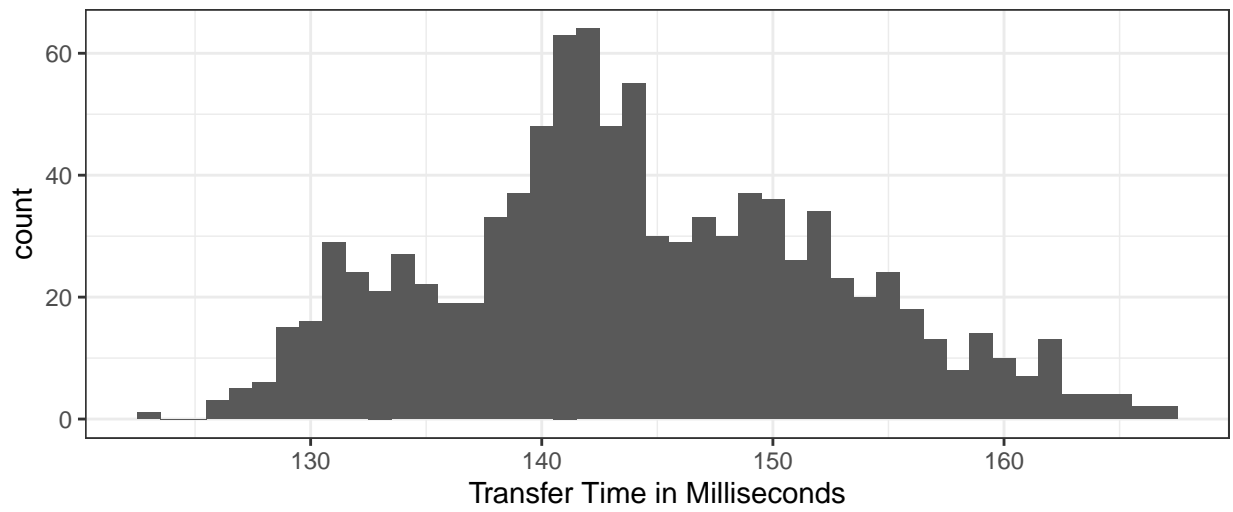


It became evident that we did have a few outliers and likely a bimodal distribution in the third set of data. To account for the outliers we used the boxplot function contained in the ggstatsplot package to remove outliers and clean up the data. The boxplot function can be used to identify outliers “which lie beyond the extremes of the whiskers”. [2] After removing the outliers, the following were our resulting histograms.

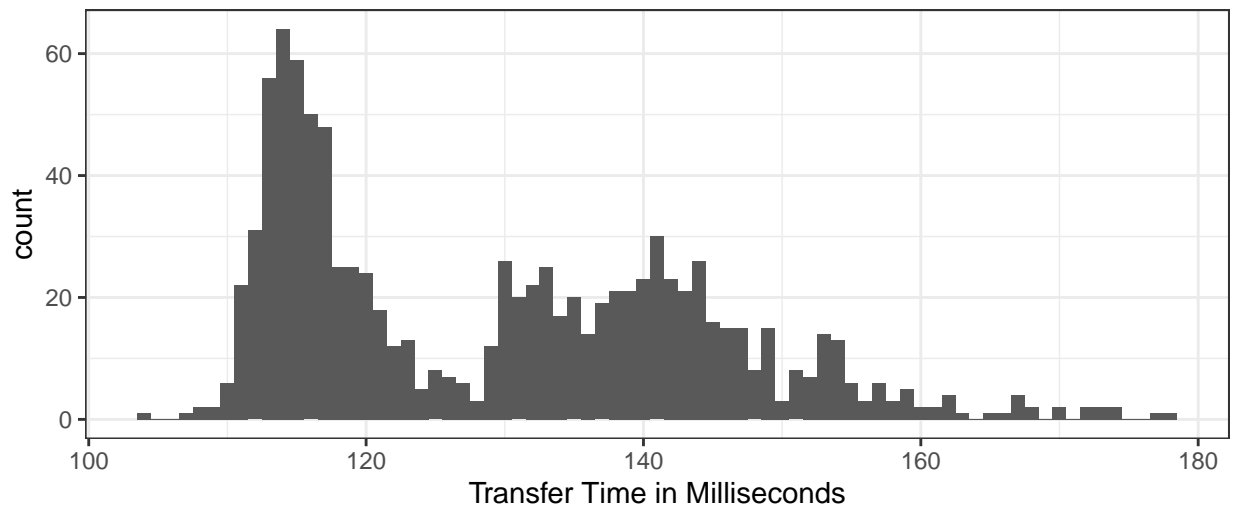
Clipper Transfer Time Histogram 1



Clipper Transfer Time Histogram 2

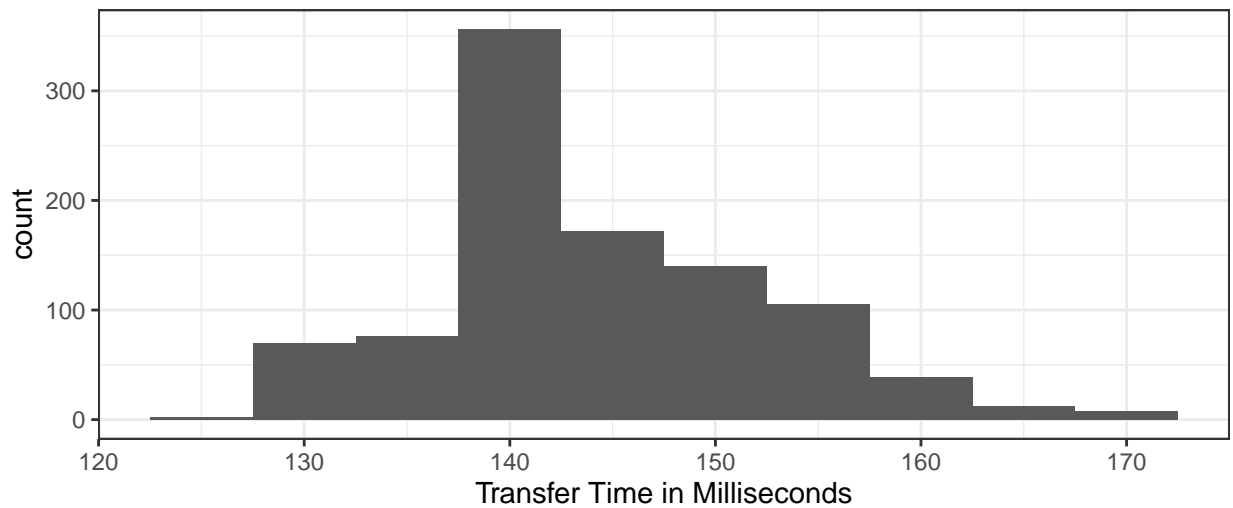


Clipper Transfer Time Histogram 3

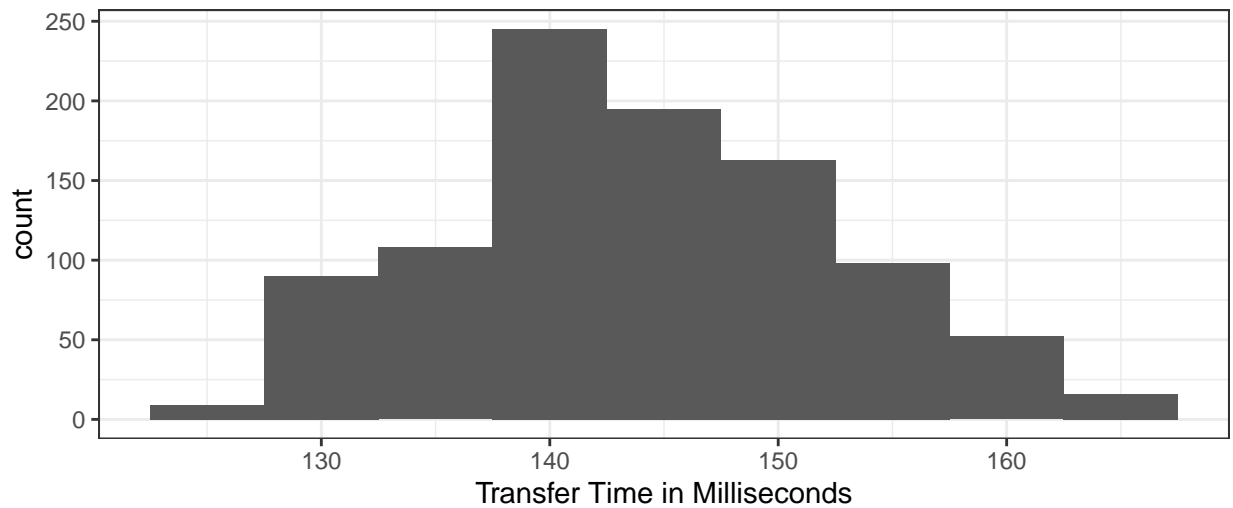


As can be seen in the resulting histograms the second test on Clipper seems to approach a normal distribution. The first and third test both seem to be bimodal at first. Upon further examination we found that if we changed the bin width for the histograms we can see that with an adjusted bin width of 5, the first test on Clipper also seems to be normally distributed. The third test is certainly bimodal.

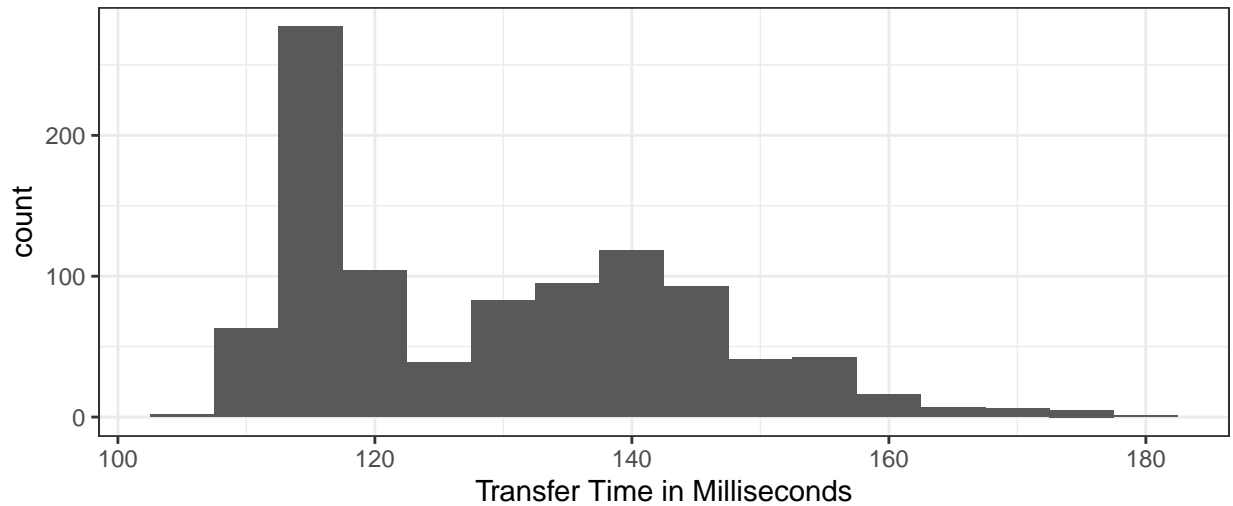
Clipper Transfer Time Histogram 1



Clipper Transfer Time Histogram 2



Clipper Transfer Time Histogram 3



After we removed the outliers and confirmed that the first two data sets were normally distributed we needed ensure that the first and second test on clipper accurately represent the same underlying distribution. We chose a t-test to compare the first and the second dataset, because a t-test can be used to determine if there is a significant difference between the means of two groups. First we'll assume a null hypothesis which is that Test 1 = Test 2 , Test 2 = Test 3, and Test 1 = Test 3, that is the data from three separate tests are = and there is no statistical difference between them. Our Alternate Hypothesis will be that Test 1 != Test 2 , Test 2 != Test 3, and Test 1 != Test 3, or that is there is a statistically significant difference between the three sets of data.

```
##
##  Welch Two Sample t-test
##
## data:  javaResults01$X2 and javaResults02$X2
## t = 0.2201, df = 1861, p-value = 0.8258
```



```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.901823  1.129823
## sample estimates:
## mean of x mean of y
##  144.973  144.859
```

```
##
## Welch Two Sample t-test
##
## data:  javaResults02$X2 and javaResults03$X2
## t = 23.973, df = 1619.8, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  13.53768 15.95032
## sample estimates:
## mean of x mean of y
##  144.859  130.115
```

```
##
## Welch Two Sample t-test
##
## data:  javaResults01$X2 and javaResults03$X2
## t = 22.123, df = 1885.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  13.54084 16.17516
## sample estimates:
## mean of x mean of y
##  144.973  130.115
```

As can be seen in the results the mean of the first test was 144.973 and the mean of the second test was 144.859. The p-value was .8258 indicating the underlying distributions are similar and we can not reject a null hypothesis in this instance. Our second t-test has an extremely low p-value ($2.2e-16$) and shows a sample mean of 144.859 for Test 1 and 130.115 for Test 2. This low p-value indicates that we should reject the null hypothesis. There is a statistically significant difference between our results in Test 2 and our results in Test 3. Likewise when comparing Test 1 and Test 3 we get similar results with a p-value of $2.2e-16$. With all that in mind we can conclude that test's 1 and 2 belong to the same underlying distribution and will be ok to use when comparing the performance of other servers. As test 3 was bimodal and did not fit with Test 1 and Test 2 we will not use Test 3 as a baseline or include it in further analysis.

```
##  0%  25%  50%  75% 100%
## 114  139  143  151  324

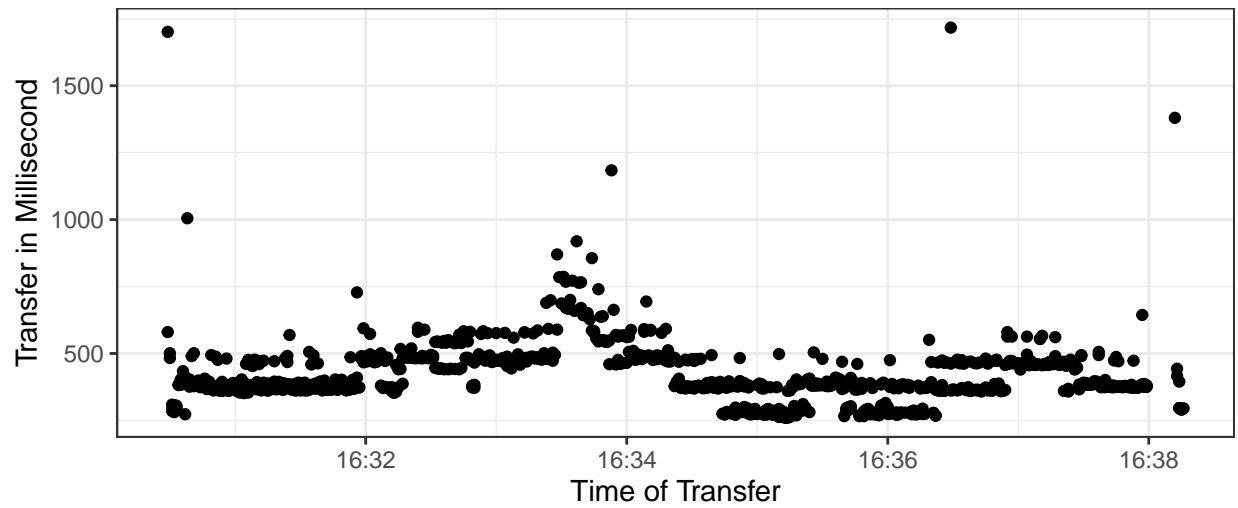
## Max = 324
## Min = 114
## Mean = 144.916
```

So our final results for best case scenario transfer time is a mean time of

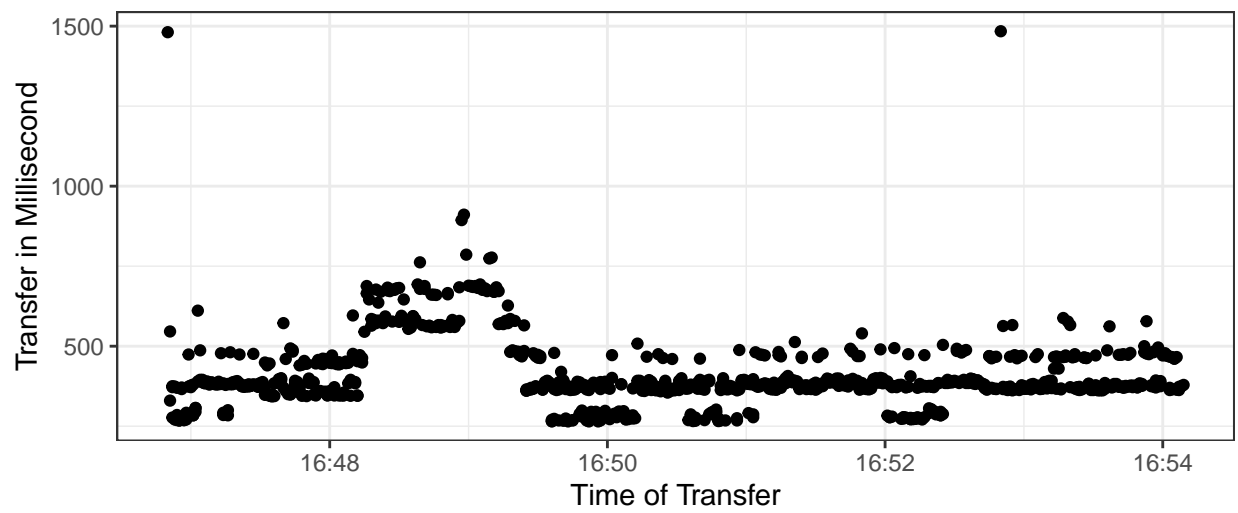
German Server Results

We reran the tests on the German server

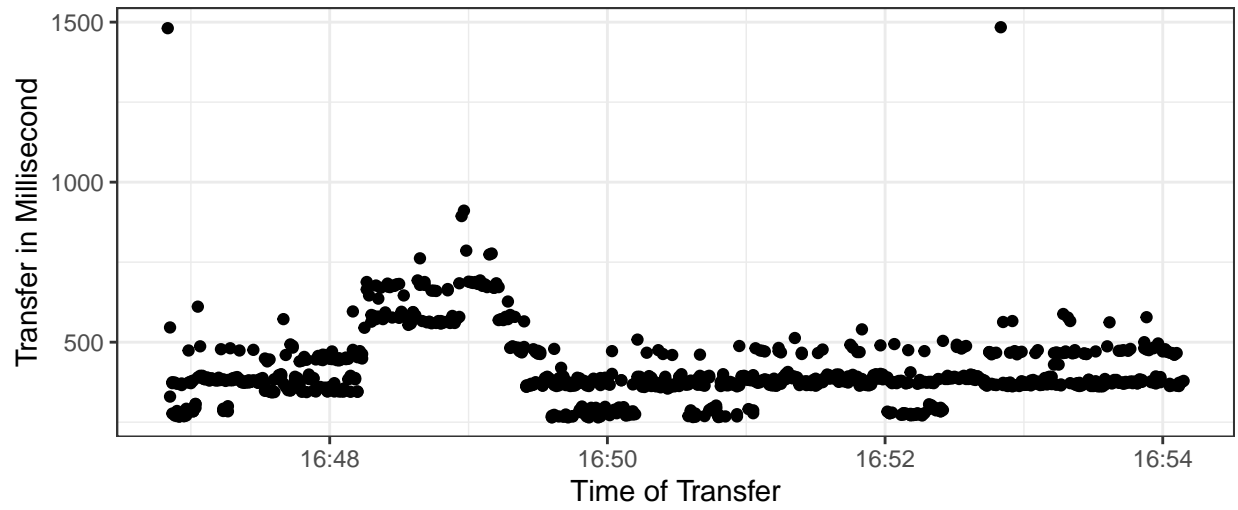
Germany Results 1



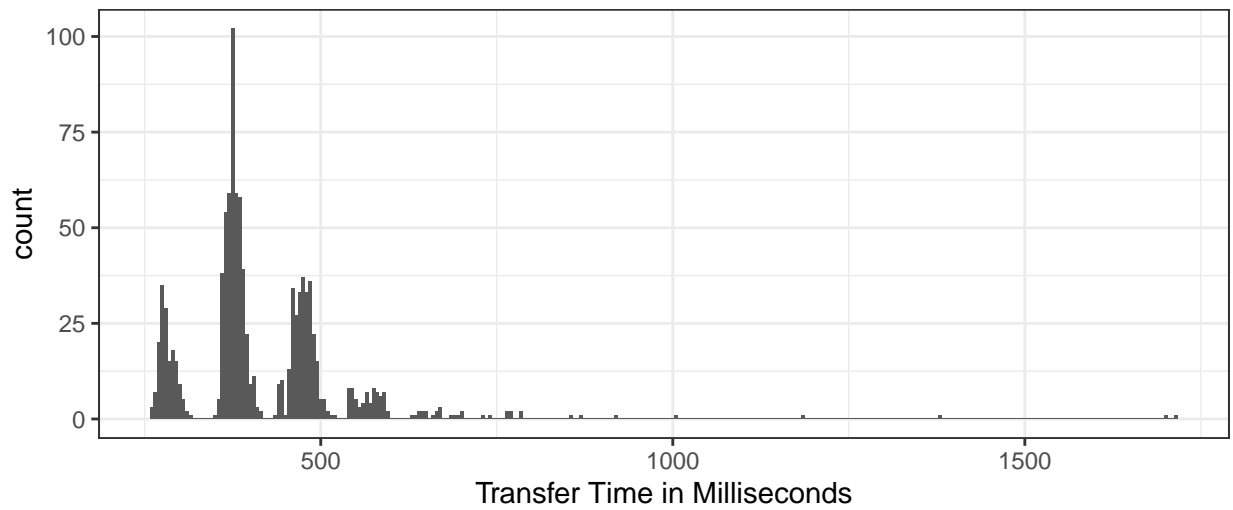
Germany Results 2



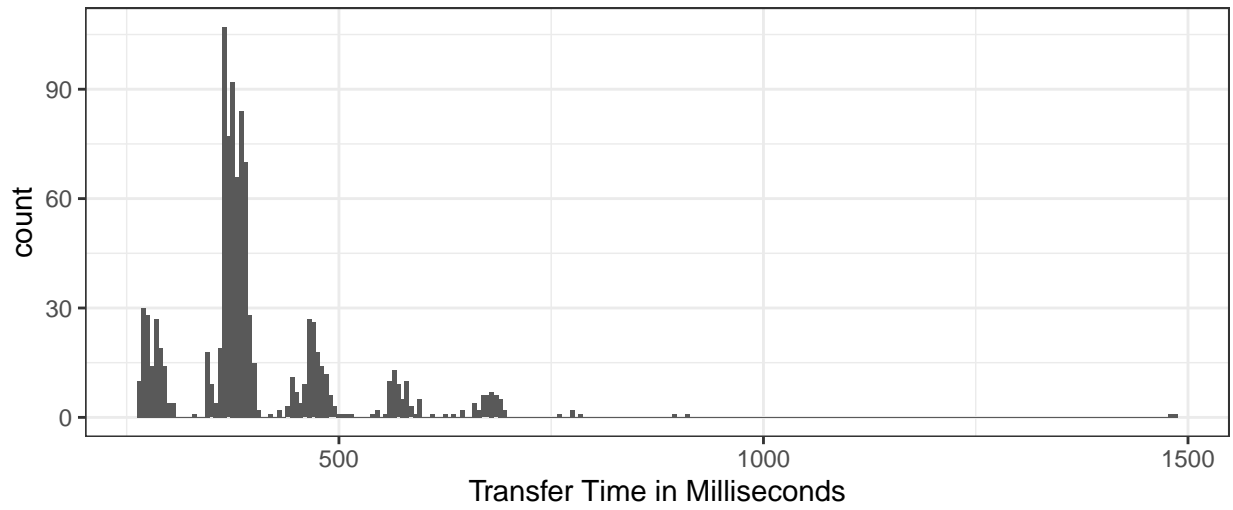
Germany Results 3



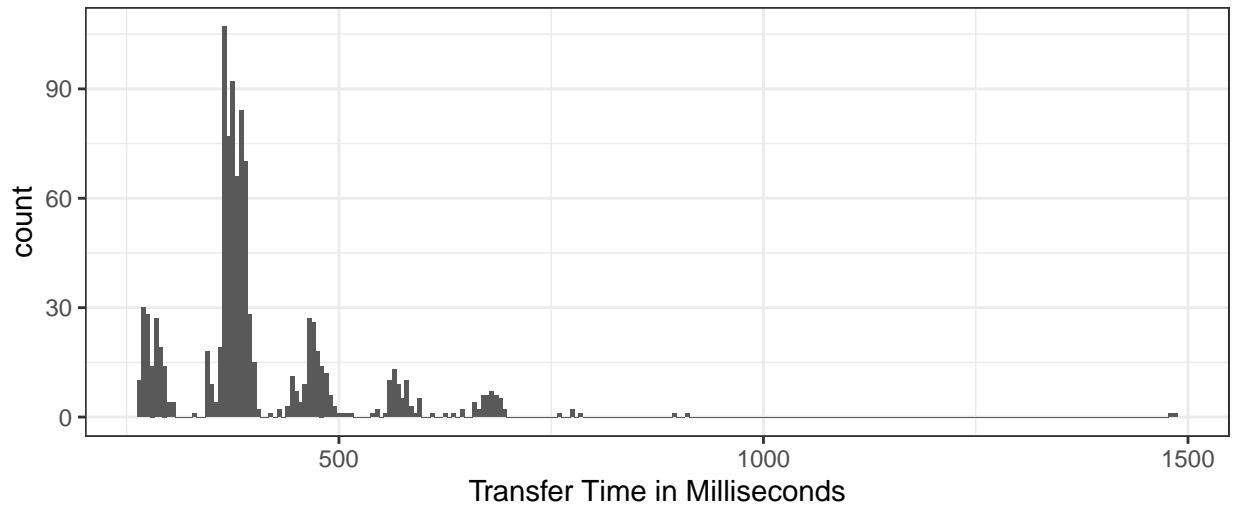
German Transfer Time Histogram 1



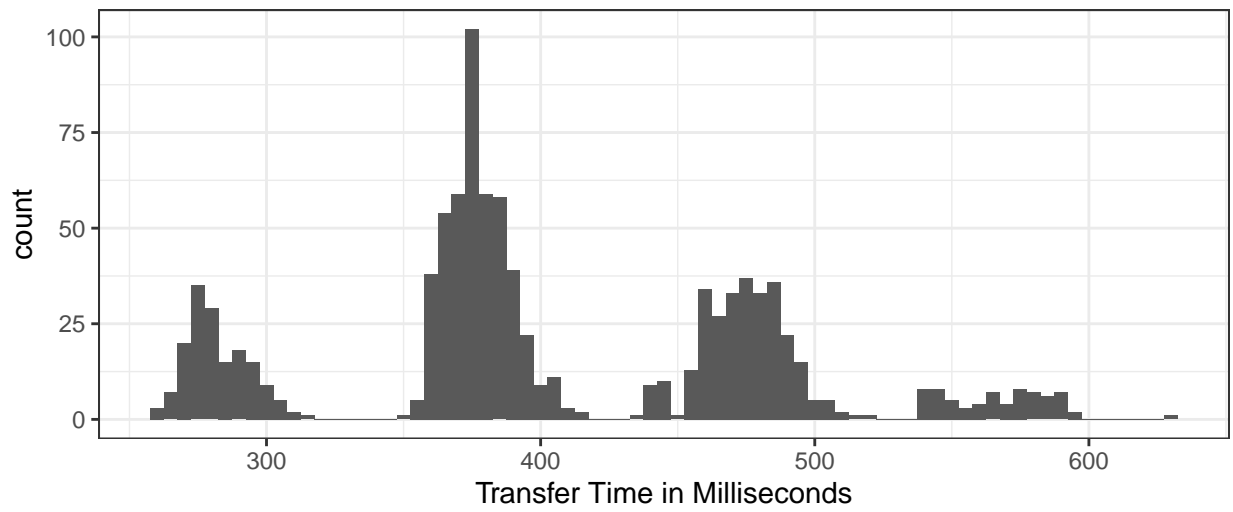
German Transfer Time Histogram 2



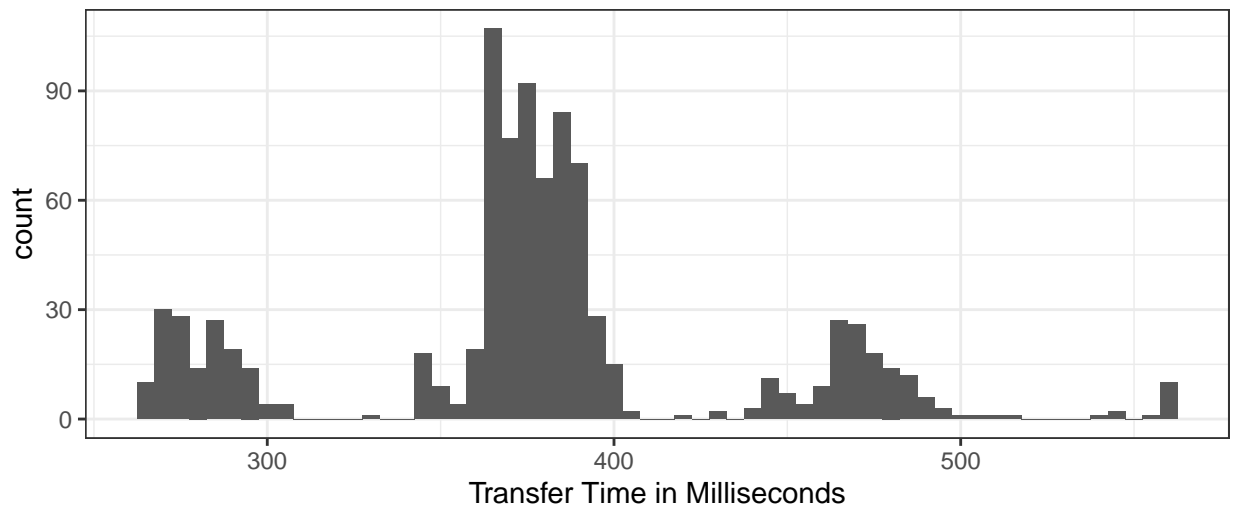
German Transfer Time Histogram 3



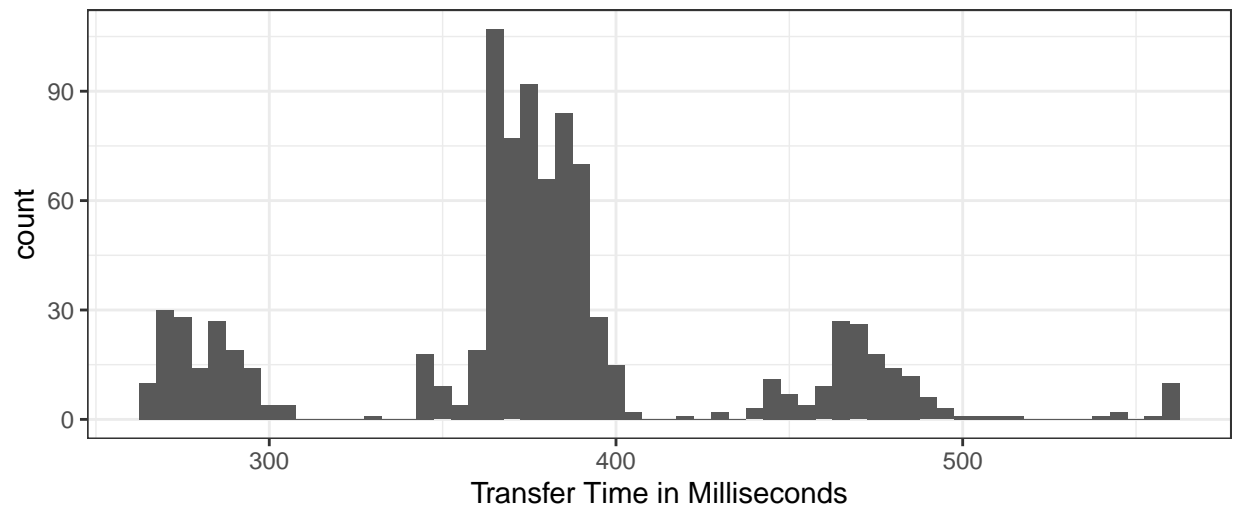
Germany Transfer Time Histogram 1



Germany Transfer Time Histogram 2

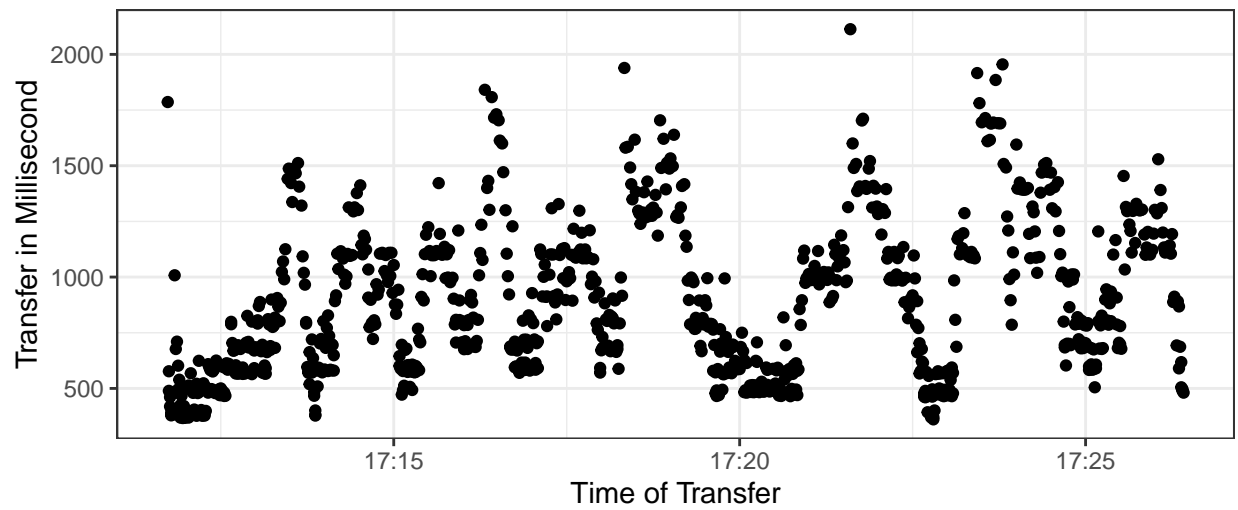


Germany Transfer Time Histogram 3

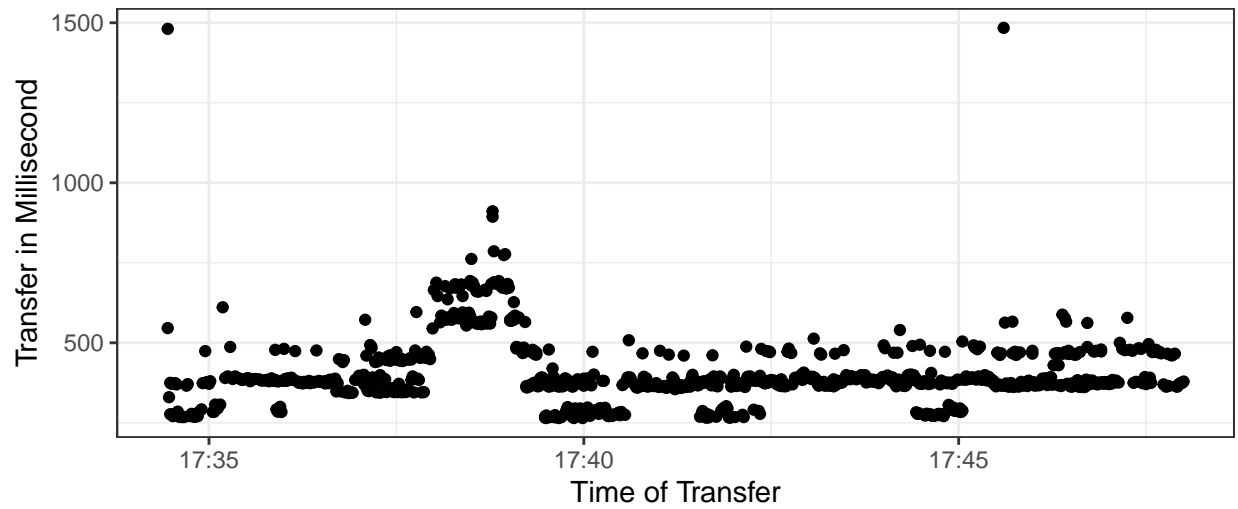


Spanish Server Results

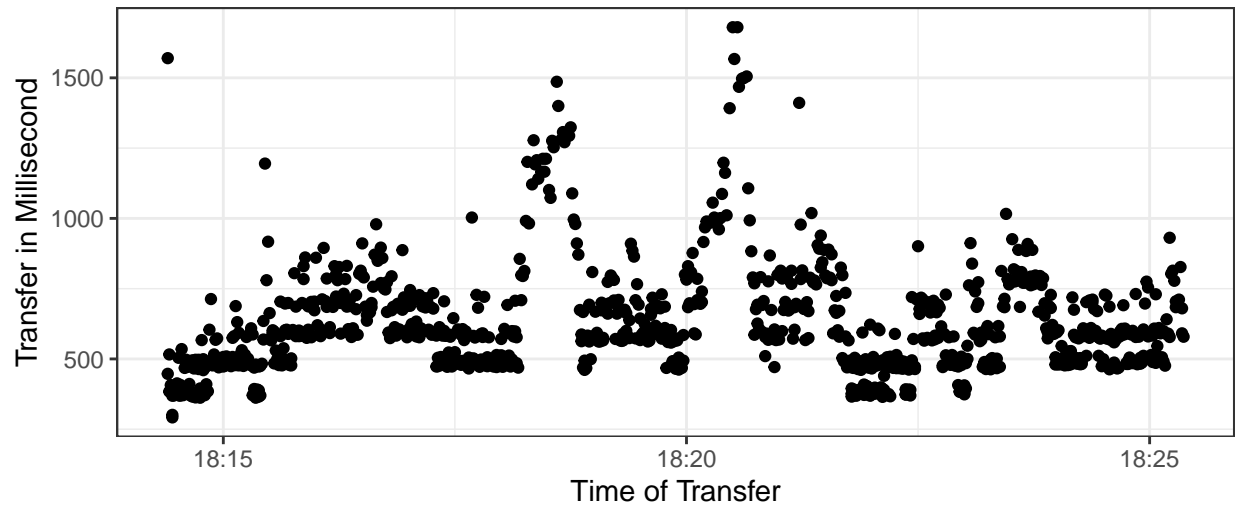
Spain Results 1



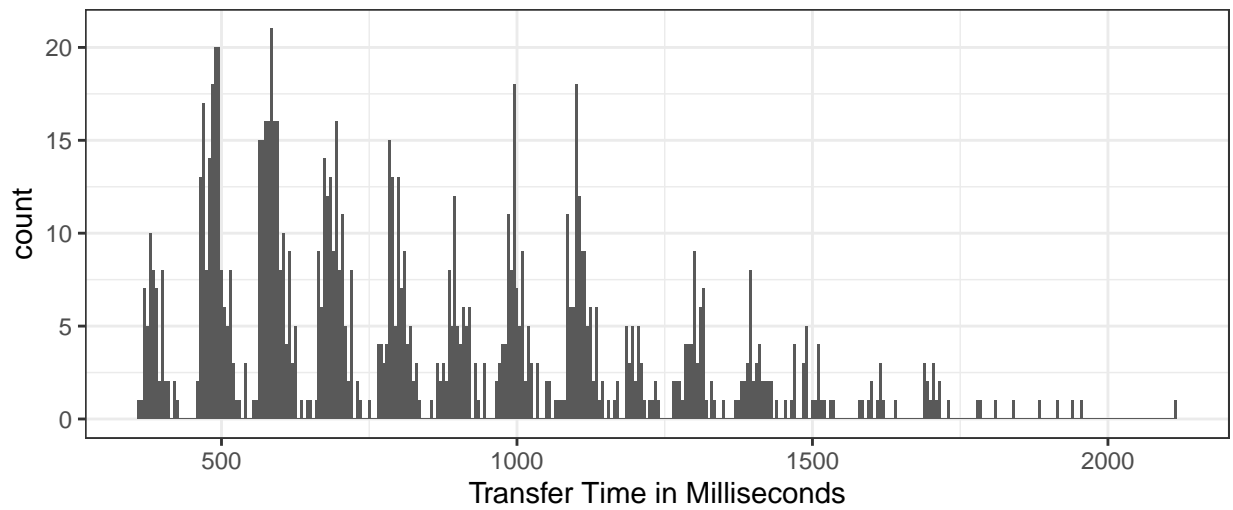
Spain Results 2



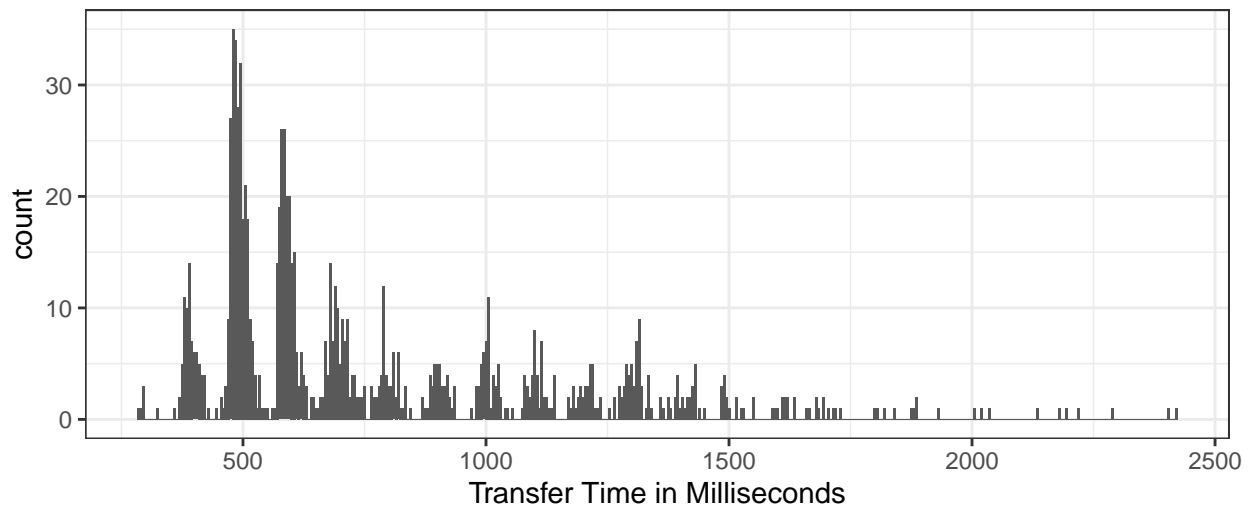
Spain Results 3



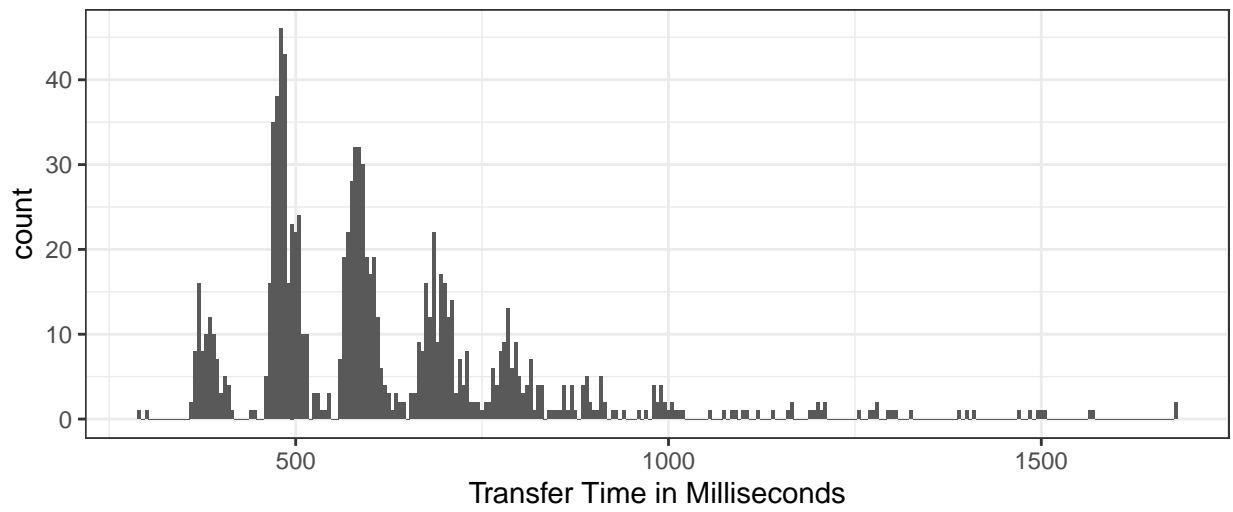
Spain Transfer Time Histogram 1



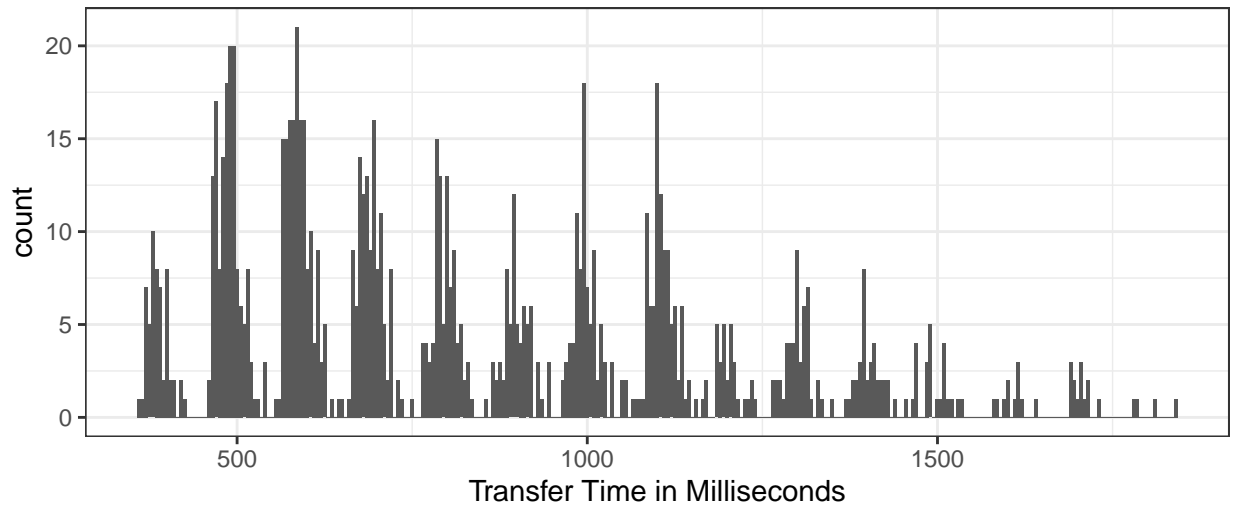
Spain Transfer Time Histogram 2



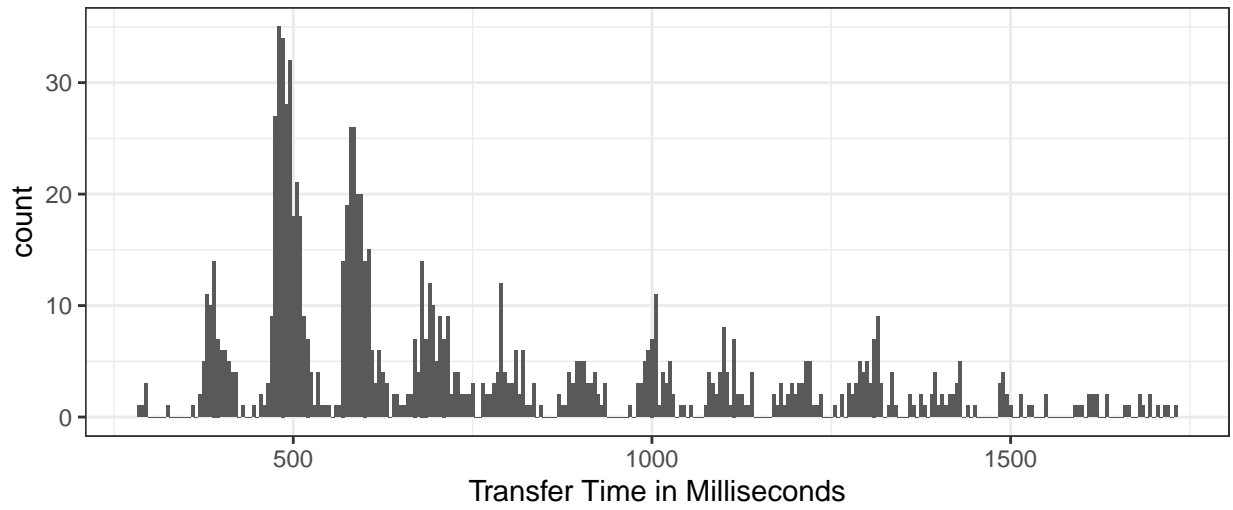
Spain Transfer Time Histogram 3



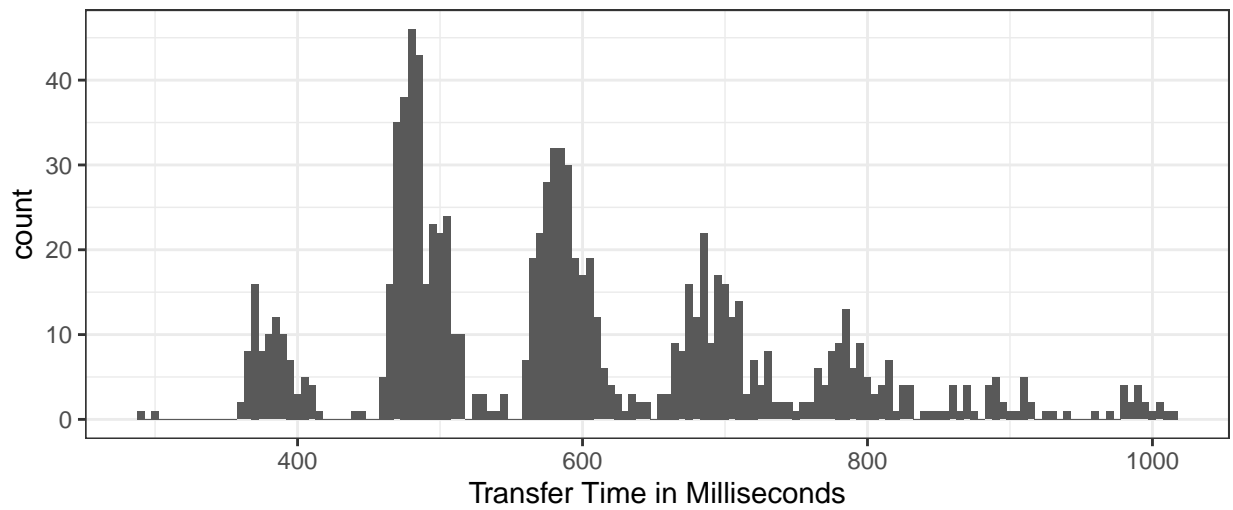
Spain Transfer Time Histogram 1



Spain Transfer Time Histogram 2



Spain Transfer Time Histogram 3



Hypothesis Testing

Conclusion

[1] "CORBA FAQ." [Online]. Available: <https://corba.org/faq.htm>.

[2] "graphics." [Online]. Available: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/boxplot>.