IE 324 Term Project

2022 Spring

First Round Submission

Simge Çınar, 21901465

Sena Ceren Göktaş, 21704202

Yağız Yaman, 21702889

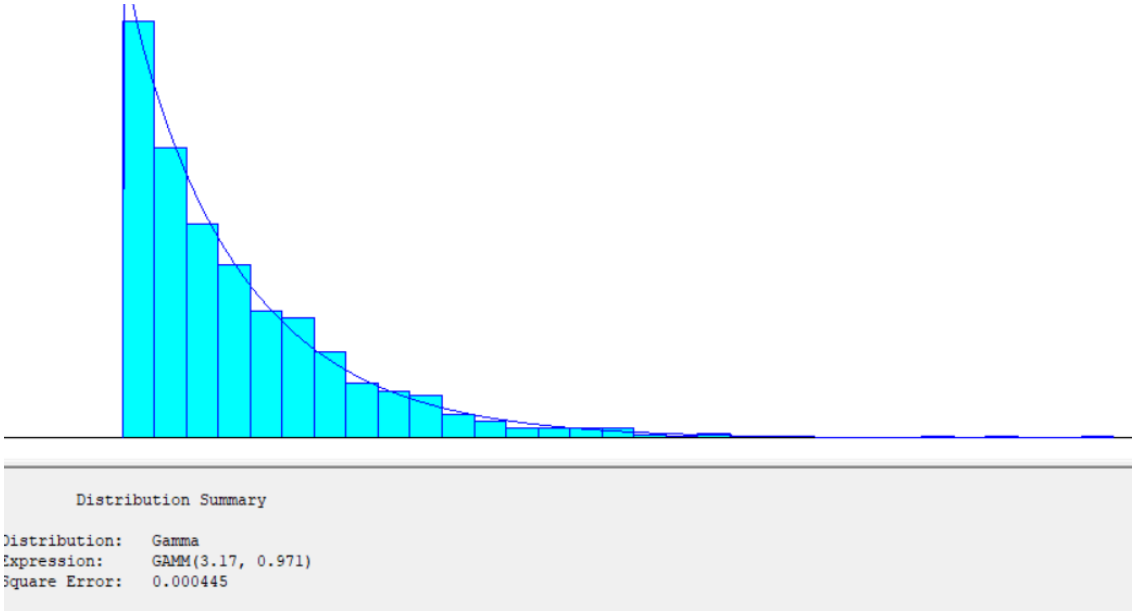
08.04.2022

Data Analysis

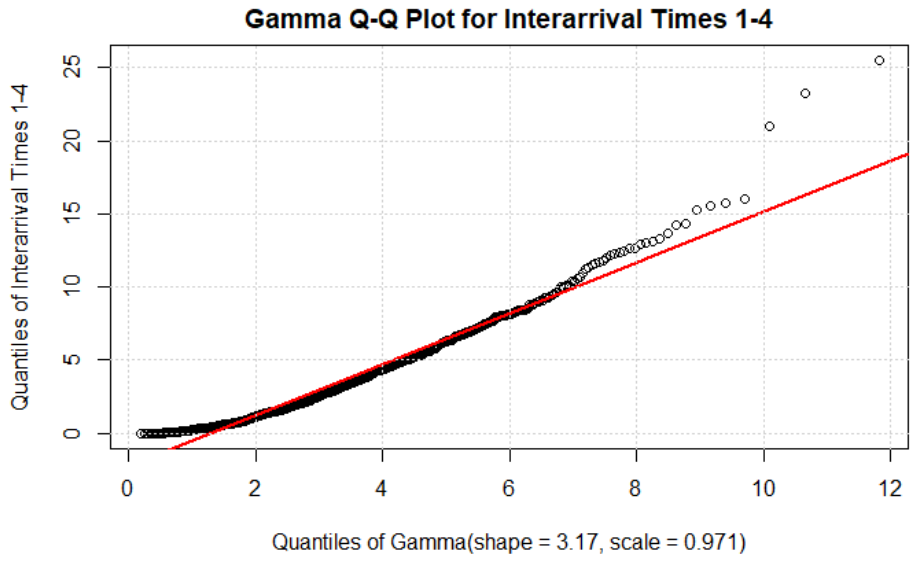
Firstly, we use the Input Analyzer Tool from Arena to see the histogram of the interarrival times for each region.

Interarrival Times for Each Region:

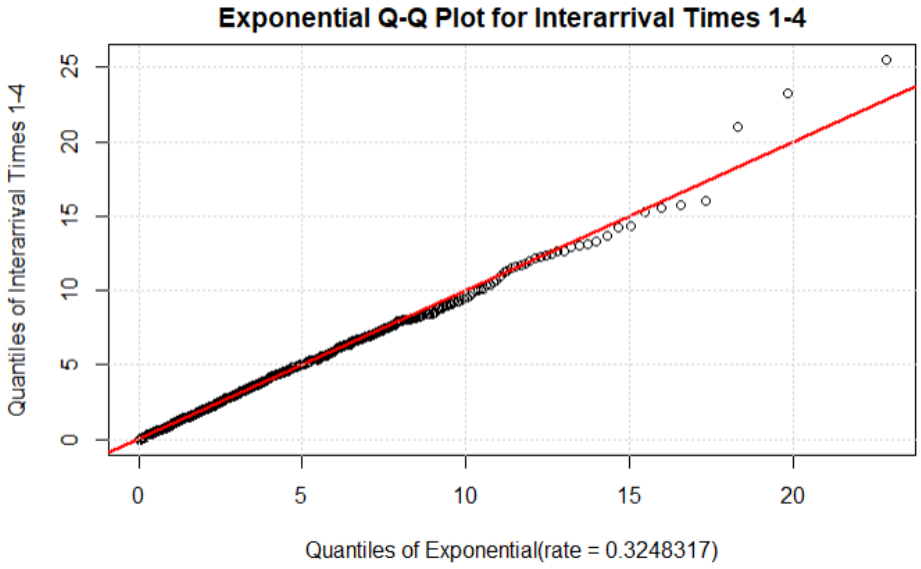
- **Ferry Station (1-4):**



The interarrival times between 1-4 best fitted to the gamma distribution with parameters (3.17, 0.971) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.000445. However, when we graph it as a Q-Q plot using RStudio, the following plot obtained:

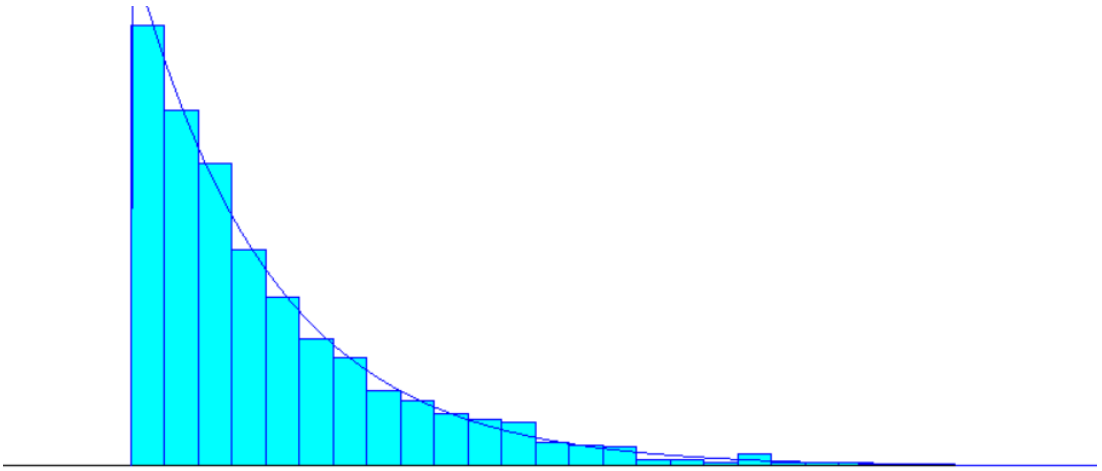


This Q-Q plot looked like a poor fit. Therefore, we decided to try the exponential distribution with estimated parameter 0.324831729 since it is the next choice with the second smallest square error which is 0.000510. (The Erlang distribution also has the same error value, however, since its shape parameter is 1, it also indicates the exponential distribution. Only trying the exponential distribution is enough in this case.)

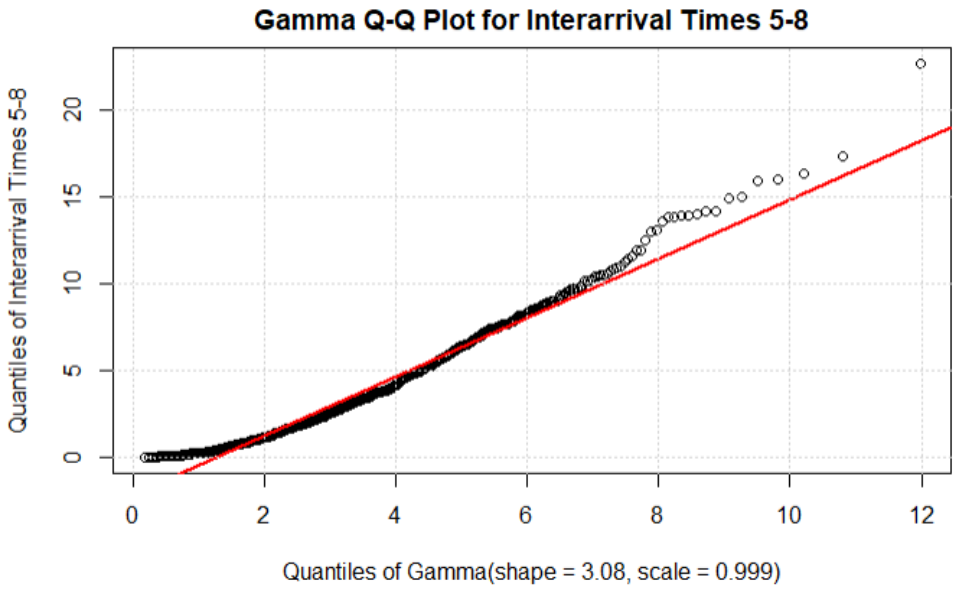


It fits better compared to the Gamma distribution. Since the difference between the square errors are very small and due to the interpretation of the Q-Q plots, we decided that exponential would be a better fit for the interarrival times of Ferry Station 1-4.

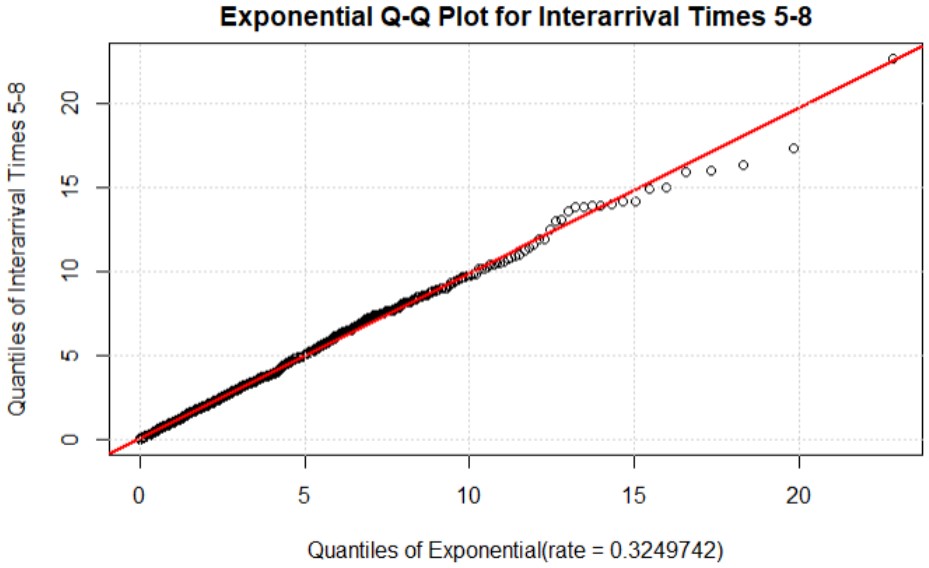
- **Ferry Station (5-8):**



The interarrival times between 5-8 best fitted to again the gamma distribution with parameters (3.08, 0.999). It has the square error 0.000243. Again, we used R to check the Q-Q plots of this distribution.

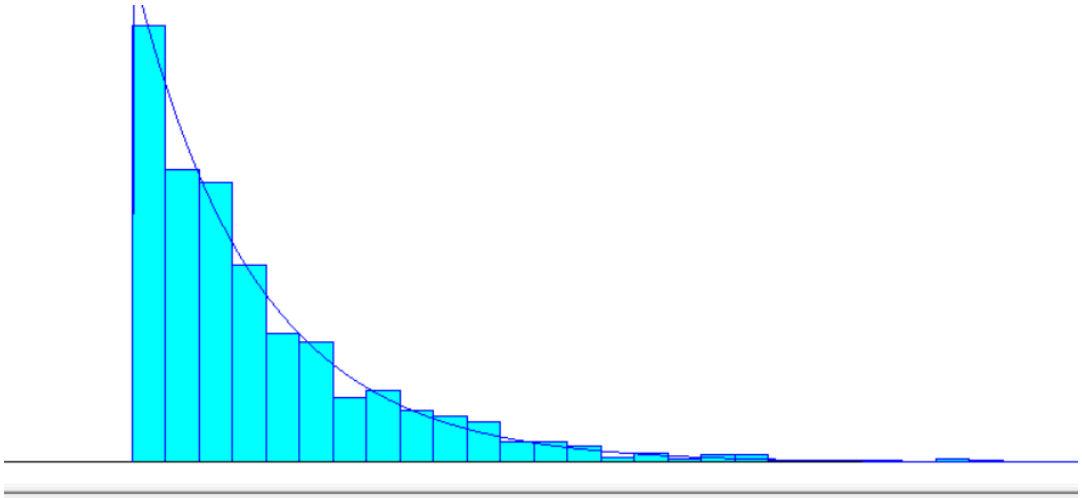


Let us look for alternatives which improve the Q-Q plot. The same thought process applies here as well. We looked at the Q-Q plot of the exponential distribution since there is a little difference between their square error values (0.000263).



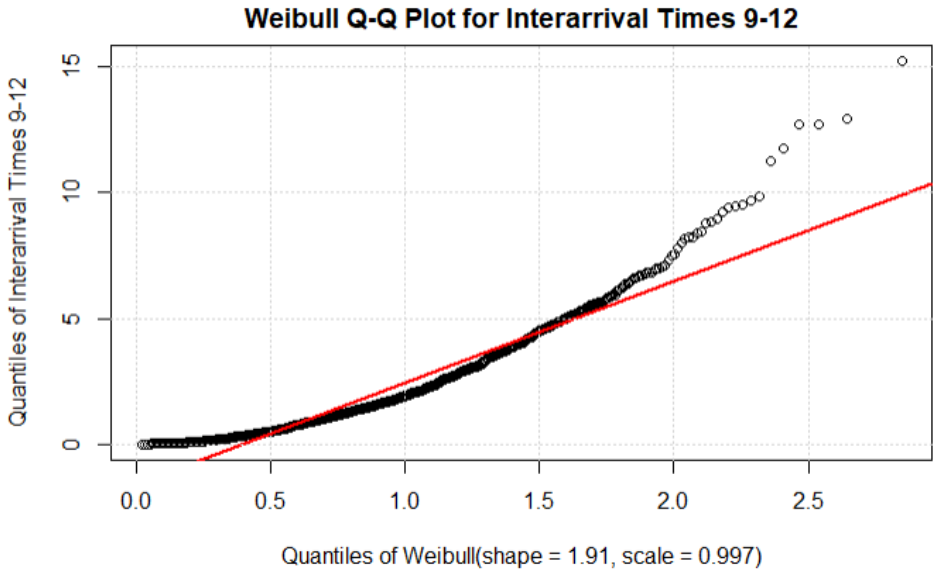
It fits better compared to the Gamma distribution. Since the difference between the square errors are very small and due to the interpretation of the Q-Q plots, we decided that exponential would be a better fit for the interarrival times of Ferry Station 5-8, hence the same reasoning.

- **Ferry Station (9-12):**

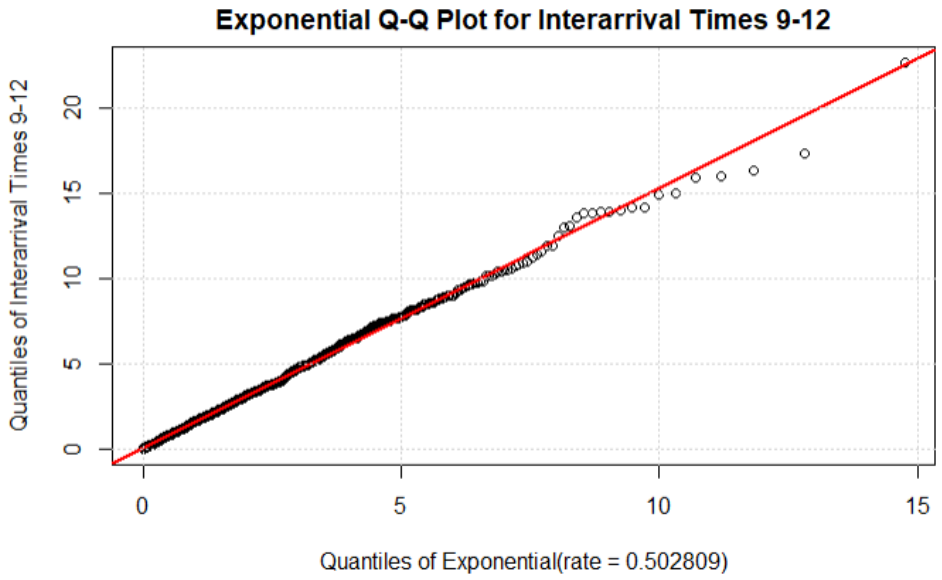


```
       Distribution Summary

Distribution:   Weibull
Expression:     WEIB(1.91, 0.997)
Square Error:   0.001017
```

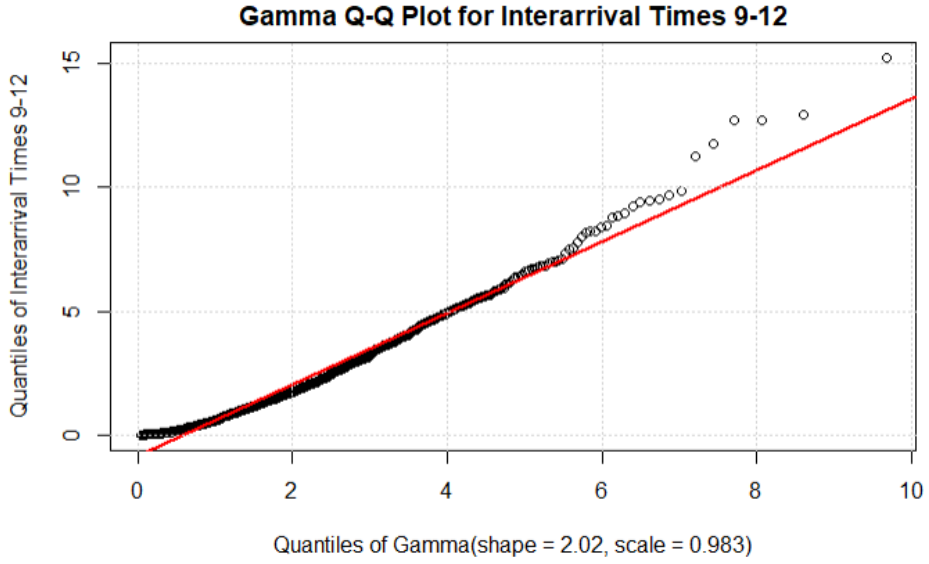
The interarrival times between 9-12 best fitted to the Weibull distribution with parameters (1.91, 0.997) in Input Analyzer. It has the square error 0.001017.



This Q-Q plot does not seem to fit well. Therefore, we looked at the next smallest least square distribution which is the Exponential distribution with square error 0.001025 and estimated lambda 0.502809.

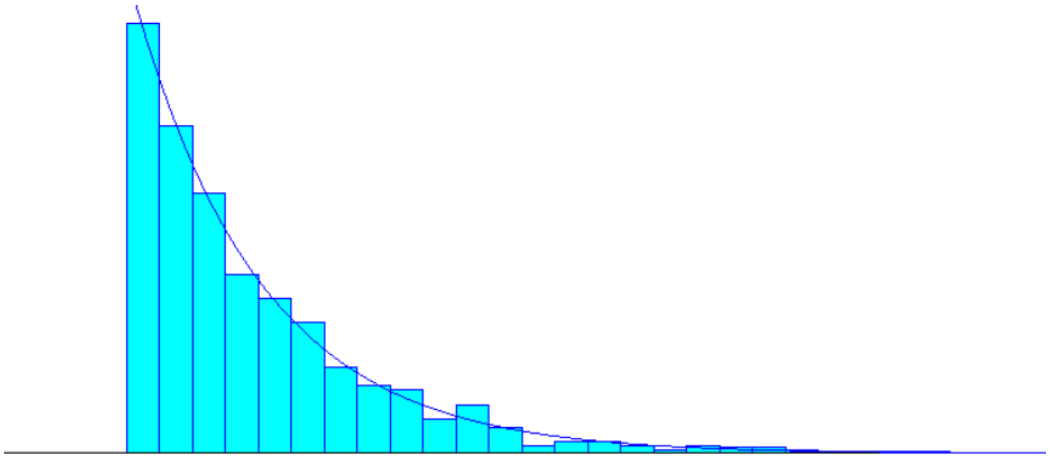


It can be seen that the fit is improved in the new exponential Q-Q plot compared to the Weibull distribution. The next smallest least square is from the Gamma distribution with square error 0.001032. Let us check that to see if we can improve it further.



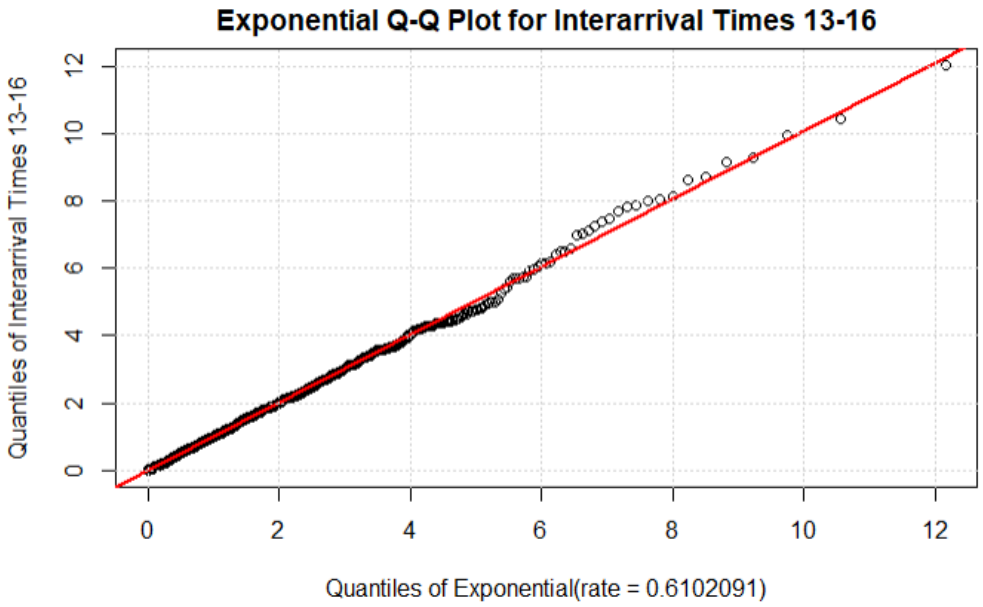
It seems like among three distributions we have investigated, the most suitable one is Exponential Distribution with parameter 0.502809 and square error 0.001025.

- **Ferry Station (13-16):**

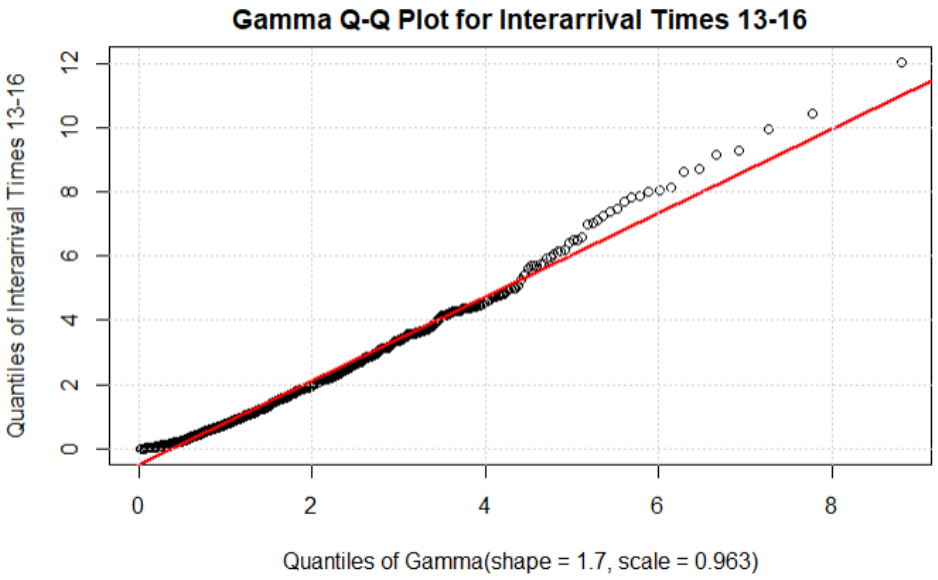


```
       Distribution Summary

Distribution:   Exponential
Expression:     EXPO(1.64)
Square Error:   0.000343
```

The interarrival times between 13-16 best fitted to the Exponential distribution with parameter (1.64) according to the Input Analyzer. It has the square error 0.000343. Let us investigate the Q-Q plot.

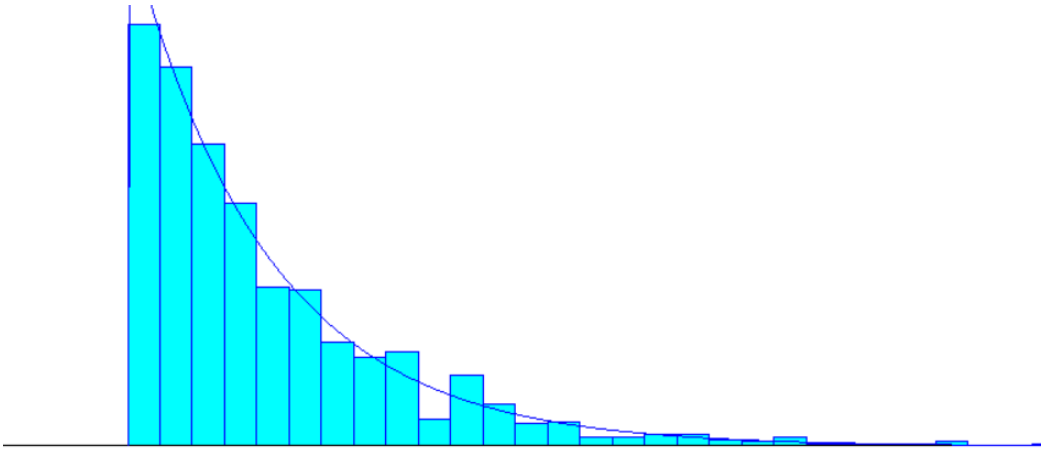


It seems like it fits quite well. However, let us investigate whether we can improve this plot more by looking at the next smallest square error distribution which is Gamma(1.7, 0.963) with square error 0.000349.



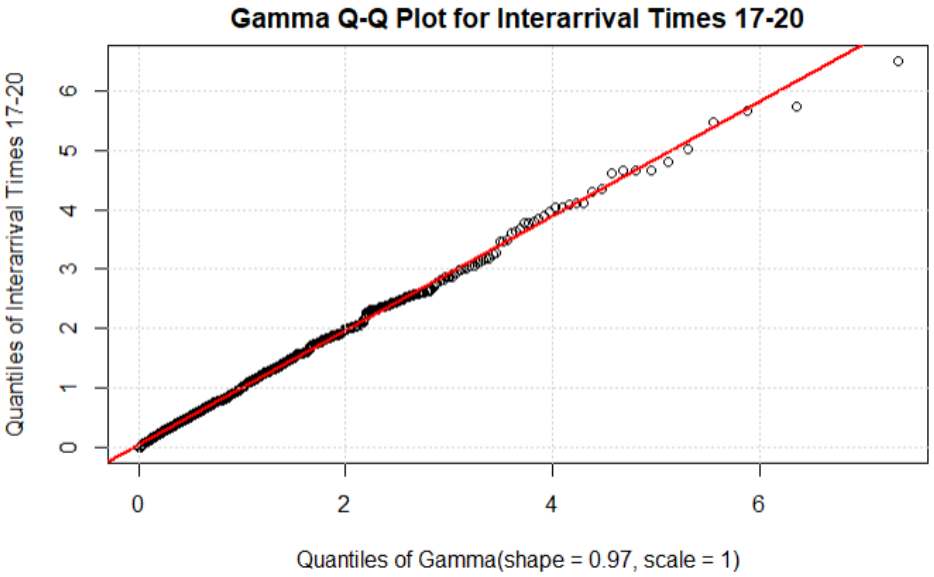
It looks like the Exponential Distribution fits slightly better than the Gamma Distribution. Therefore, we stay with the decision of exponential distribution with square error 0.000343 and parameter $\lambda = \frac{1}{\bar{x}}$ which is 0.6102091.

- **Ferry Station (17-20):**

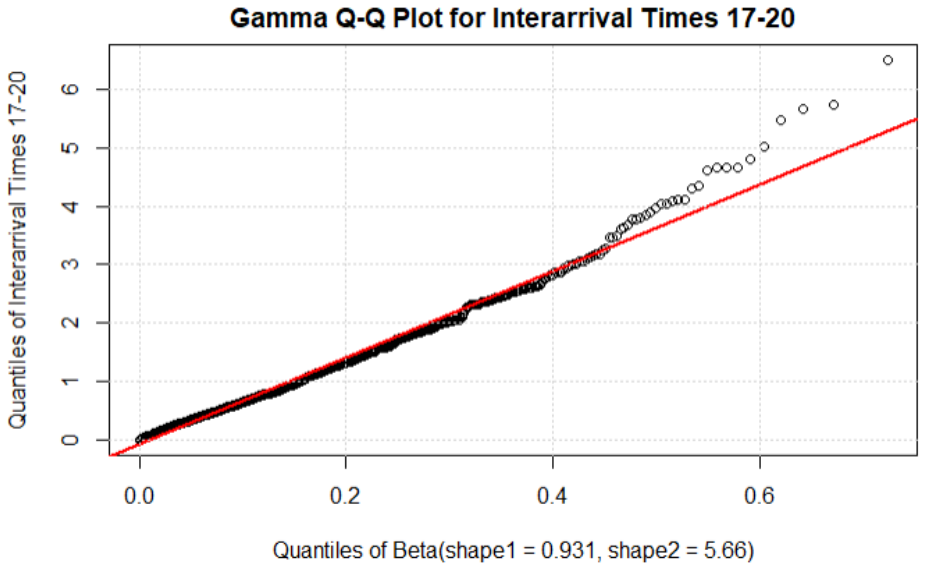


```
       Distribution Summary

Distribution:   Gamma
Expression:     GAMM(0.97, 1)
Square Error:   0.000977
```

The interarrival times between 17-20 best fitted to the Gamma distribution with parameters (0.97, 1) according to the Input Analyzer. It has the square error 0.000977. Let us investigate the Q-Q plot of this data.

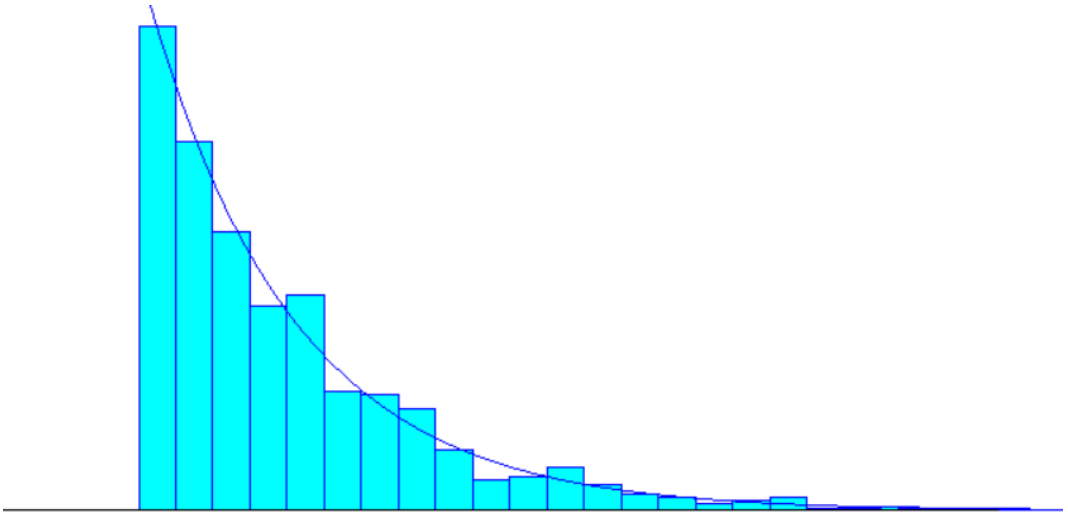


It seems to be fitting quite well. Again, let us look for improvement in the plot by investigating other distributions as well. The next distribution with the smallest square error is Exponential with error 0.001063. Yet, the graph looked almost the same since the shape of the Gamma Distribution is so close to 1 (0.97). Therefore, we tried the next distribution which is Beta(0.931, 5.66).



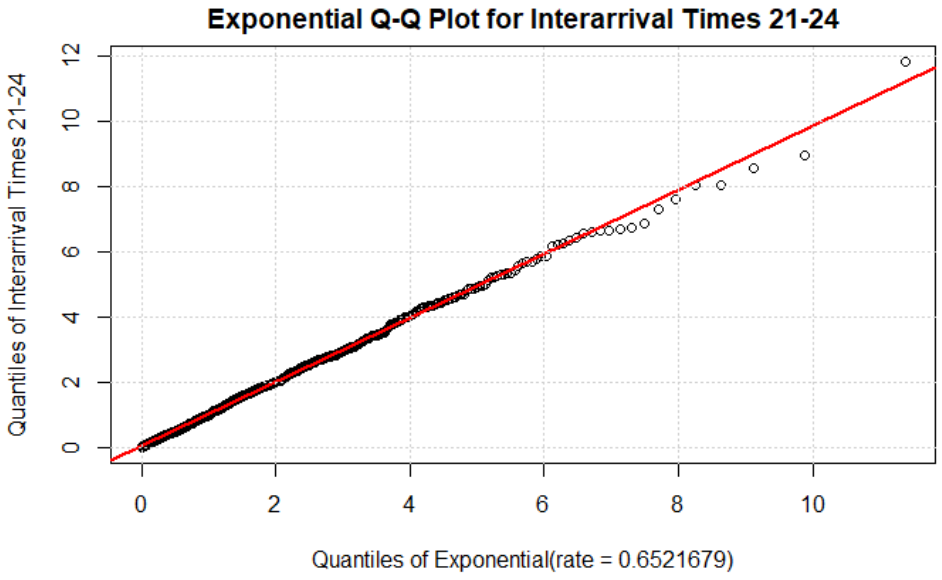
It seems like among the distributions we have tried, with the smallest square error and best fitted one is the Gamma(0.97, 1) with square error 0.000977.

- **Ferry Station (21-24):**

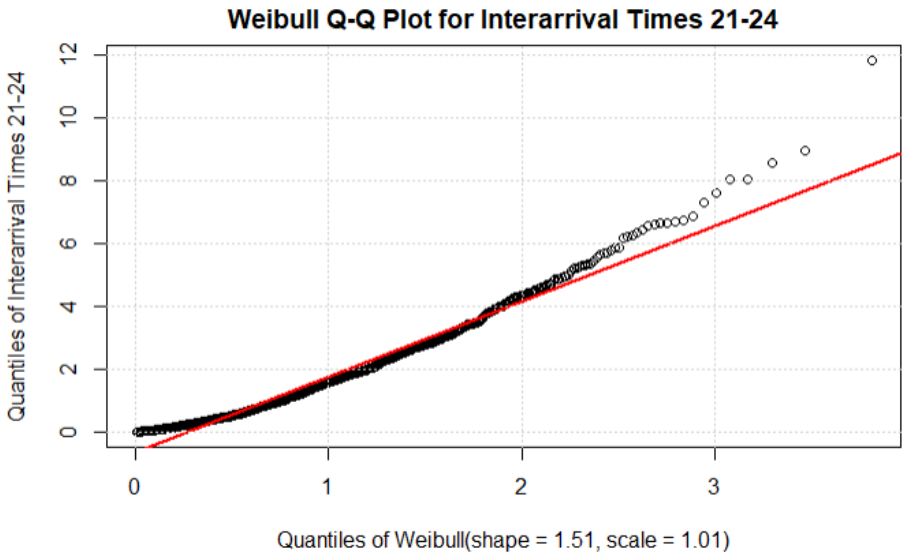


```
      Distribution Summary

Distribution:   Exponential
Expression:     EXPO(1.53)
Square Error:   0.000797
```

The interarrival times between 21-24 best fitted to the Exponential distribution with parameters (1.53) according to the Input Analyzer. It has the square error 0.000797. Let us investigate the Q-Q plot of this data.

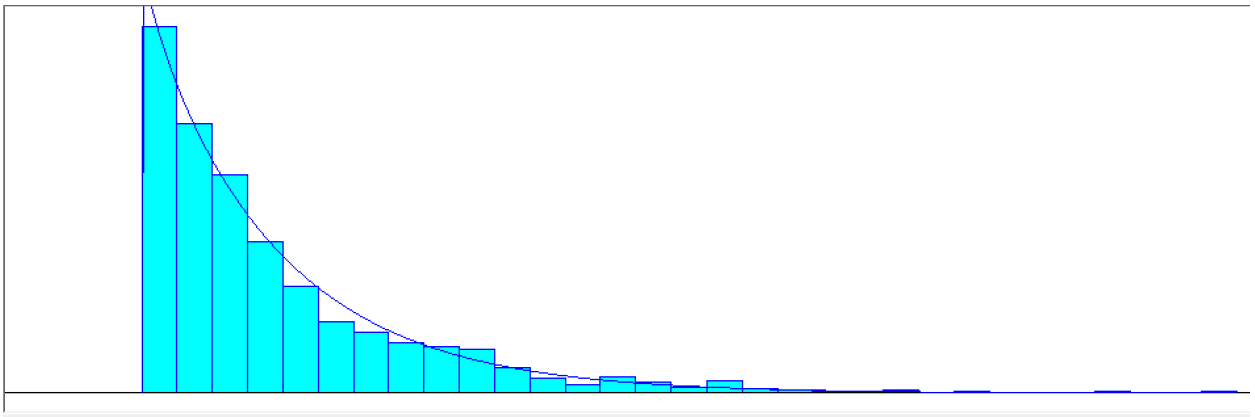


It seems to be fitting quite well. Again, let us look for improvement in the plot by investigating other distributions as well. The next distribution with the smallest square error is Weibull(1.51, 1.01) with error 0.000841.



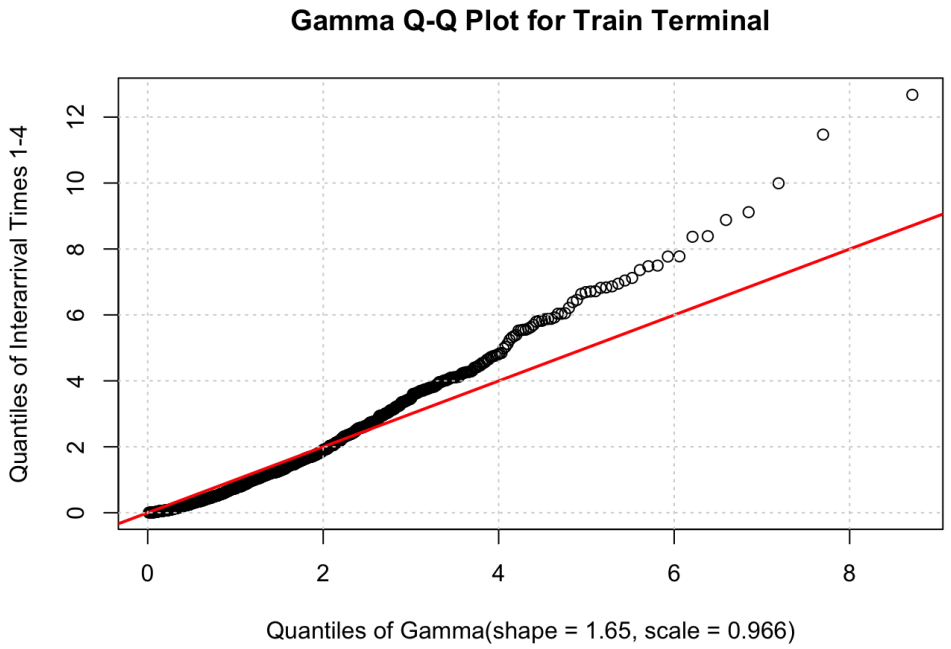
However, it seems like the Exponential Distribution is fitting better than the Weibull distribution. Therefore, as Input Analyzer suggested, this data is more like an Exponential Distribution with rate parameter 0.6521679 and square error 0.000797.

- **Train Terminal (1-4):**

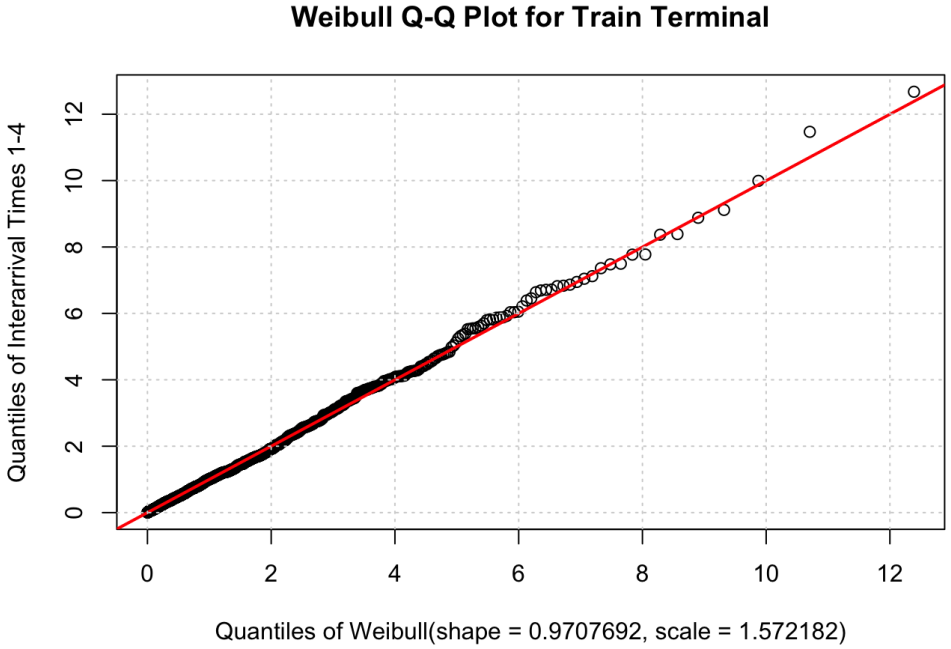


```
      Distribution Summary

Distribution:   Gamma
Expression:     GAMM(1.65, 0.966)
Square Error:   0.000607
```

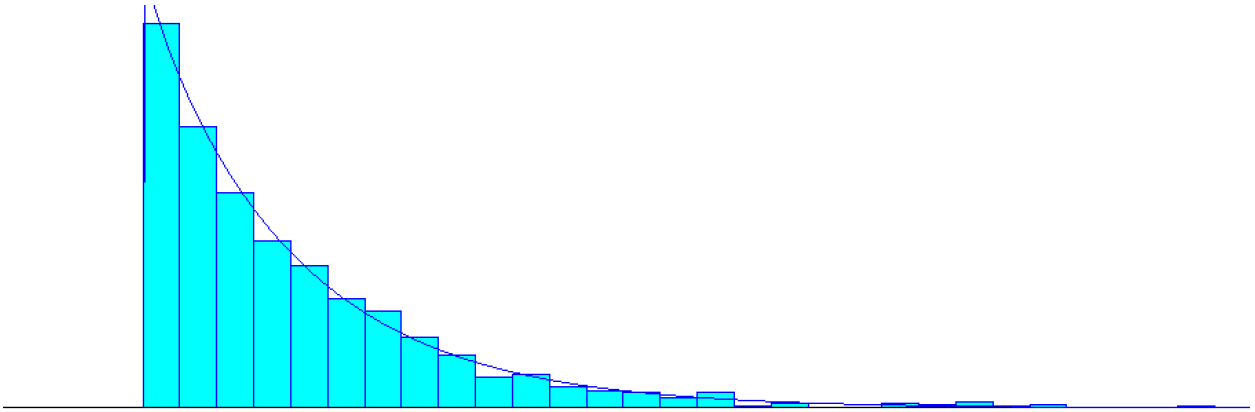
The interarrival times between 1-4 best fitted to the gamma distribution with parameters (1.65, 0.966) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.000607. However, when we graph it as a Q-Q plot using RStudio, the following plot obtained:



This Q-Q plot looks like a poor fit. Therefore, we decided to try the exponential and weibull distribution since our histogram is right skewed. Both of them look like a good fit so we choose the one with least square error. The most suitable one is the weibull distribution with square error 0.000623. (Note that parameters are estimated using MLE)

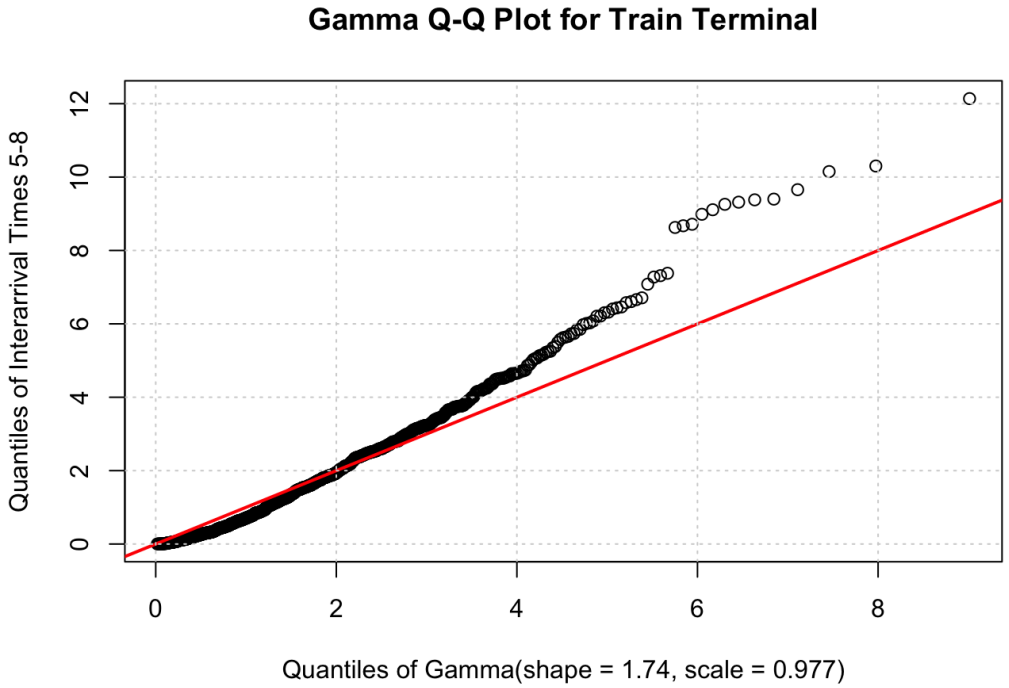


- **Train Terminal (5-8):**



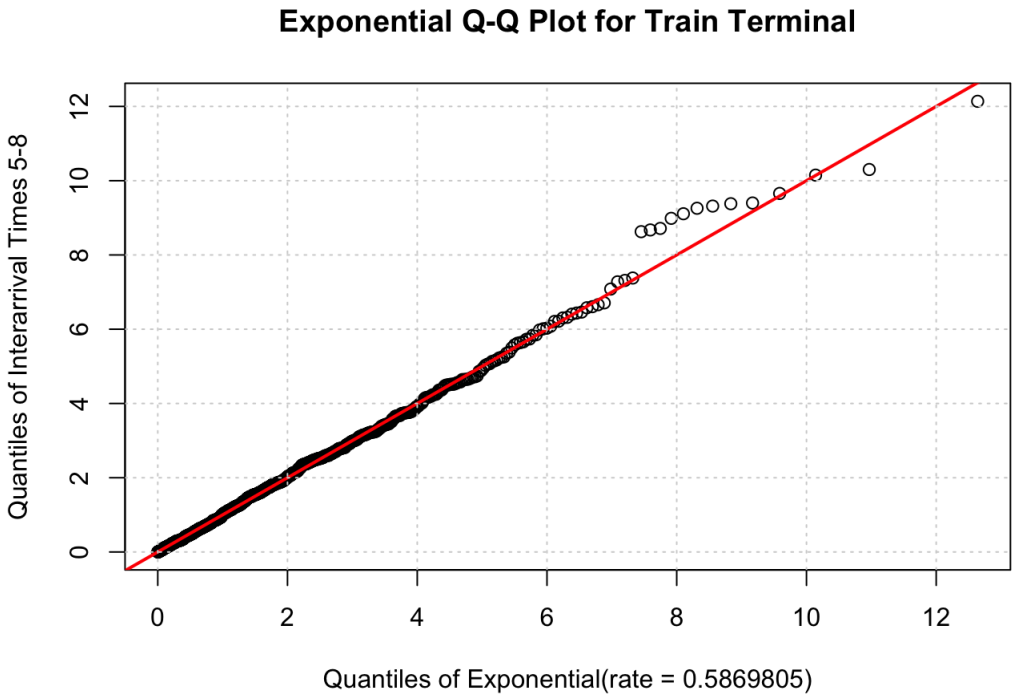
```
       Distribution Summary

Distribution:   Gamma
Expression:     GAMM(1.74, 0.977)
Square Error:   0.000296
```

The interarrival times between 5-8 best fitted to the gamma distribution with parameters (1.74, 0.977) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.000296. However, when we graph it as a Q-Q plot using RStudio, the following plot obtained:

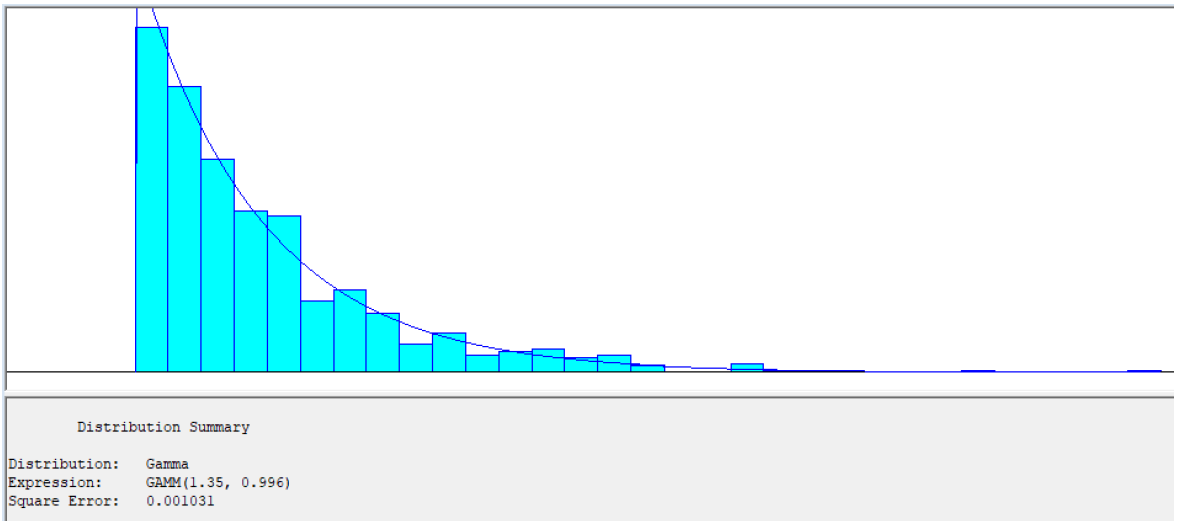


This Q-Q plot looks like a poor fit. Therefore, we decided to try the exponential and weibull distribution since our histogram is right skewed. Both of them look like a good fit so we choose

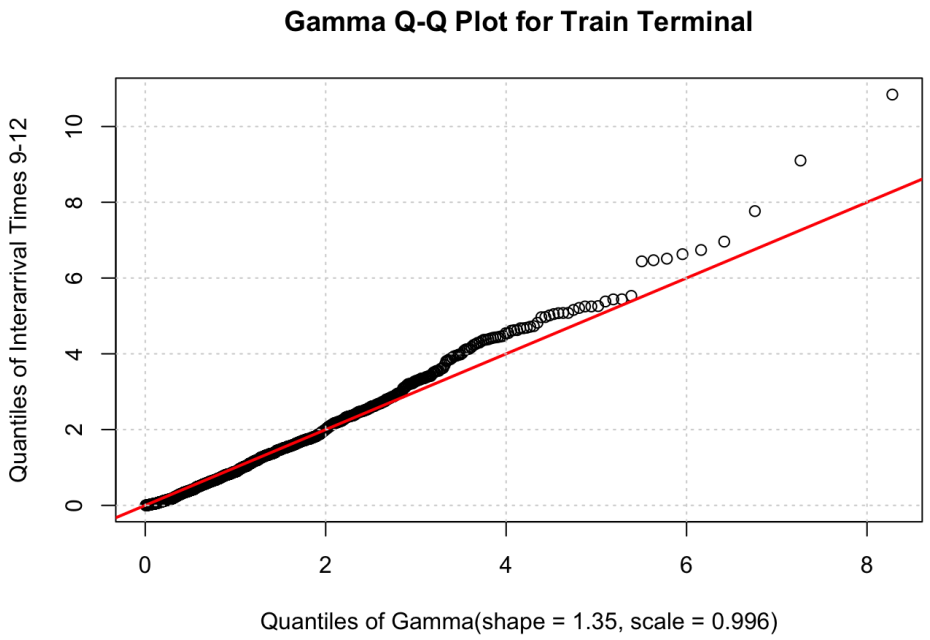
the one with least square error. The most suitable one is the exponential distribution with square error 0.000326. (Note that parameters are estimated using MLE)



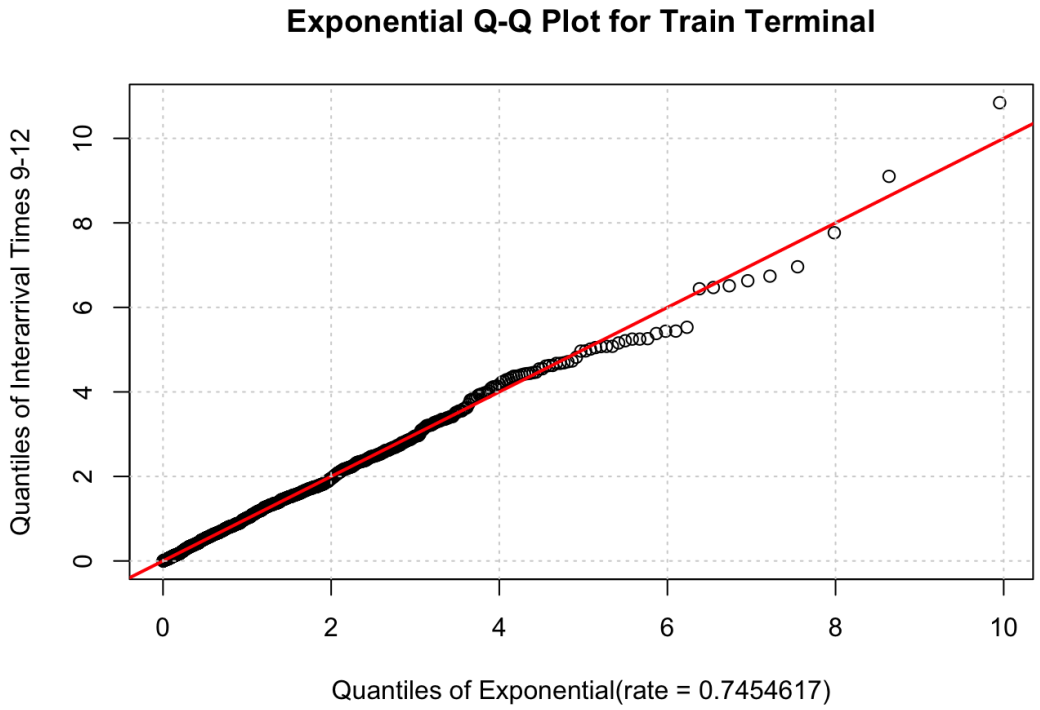
- **Train Terminal (9-12):**



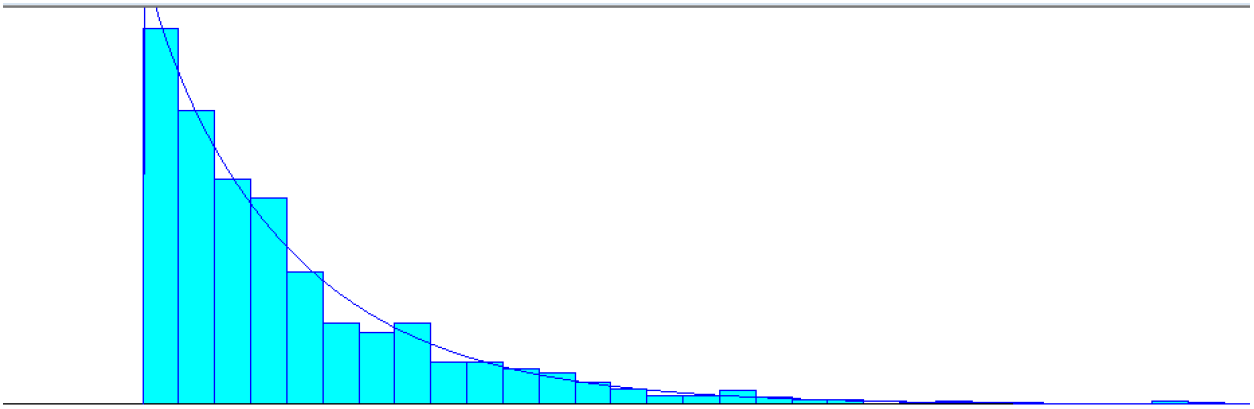
The interarrival times between 9-12 best fitted to the gamma distribution with parameters (1.35, 0.996) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.001031. To check whether it's a good fit we graph it as a Q-Q plot using RStudio, the following plot obtained:



This Q-Q plot doesn't look bad but let's investigate whether we can find a better fit. We decided to try the exponential and weibull distribution since our histogram is right skewed. Q-Q plot of both of them look like a better fit so we choose the one with least square error. The most suitable one is the exponential distribution with square error 0.001072. The difference between square errors for exponential and gamma is very small which is 0.000041 so it's okay to choose exponential instead of gamma. (Note that parameters are estimated using MLE)

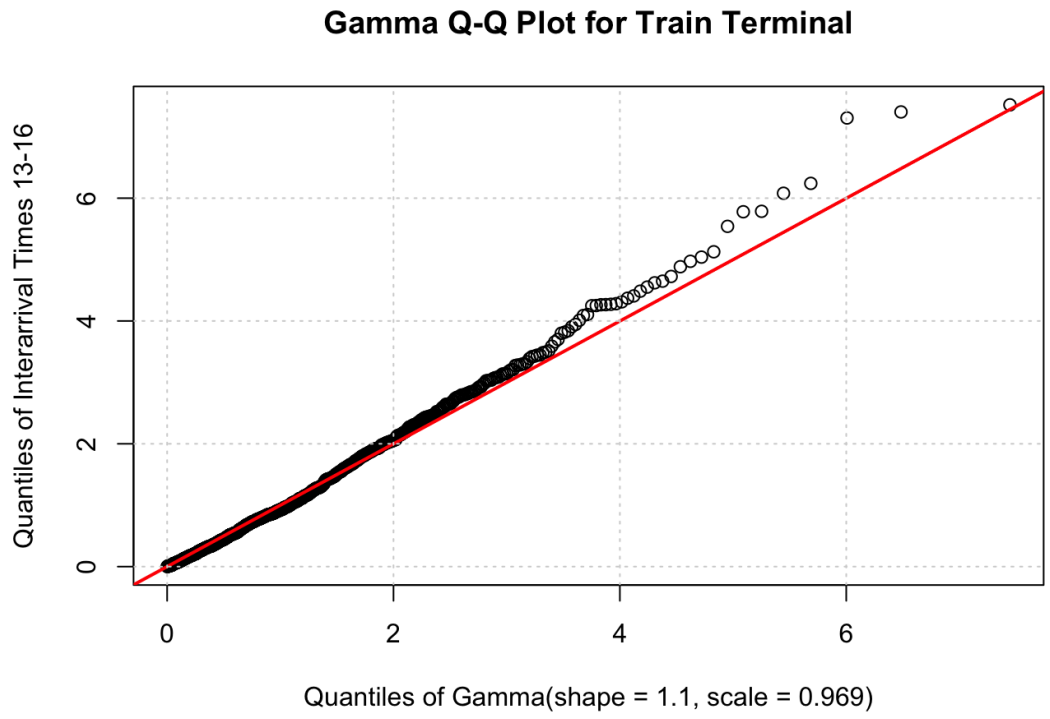


- **Train Terminal (13-16):**



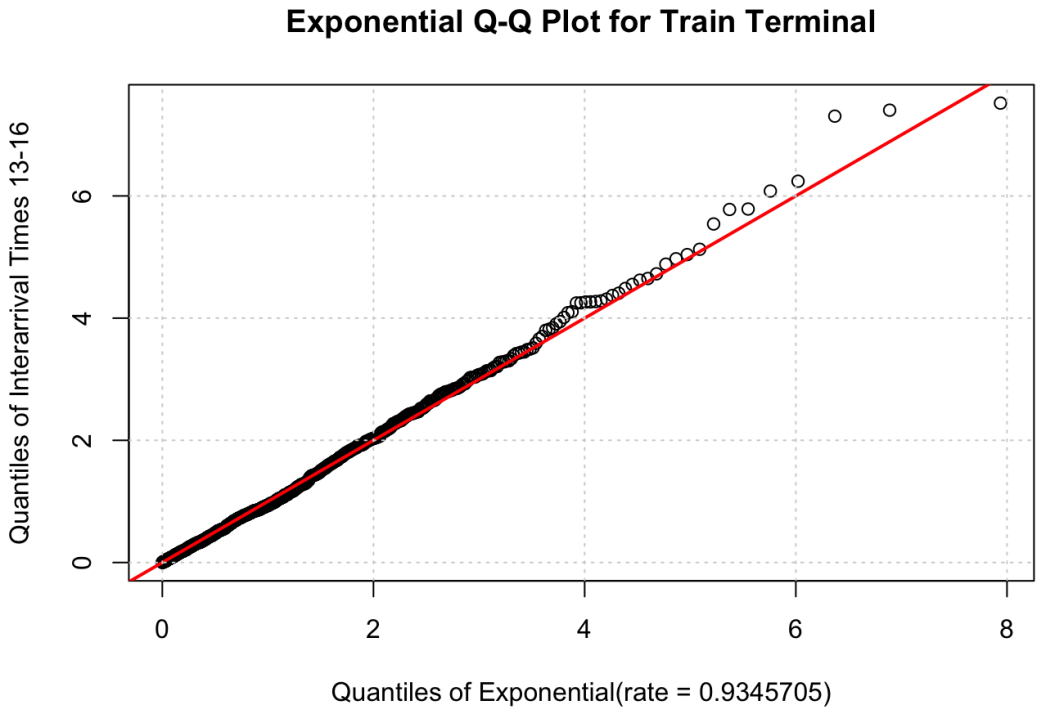
```
       Distribution Summary

Distribution:   Gamma
Expression:     GAMM(1.1, 0.969)
Square Error:   0.000783
```

The interarrival times between 13-16 best fitted to the gamma distribution with parameters (1.1, 0.969) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.000783. However, when we graph it as a Q-Q plot using RStudio, the following plot obtained:

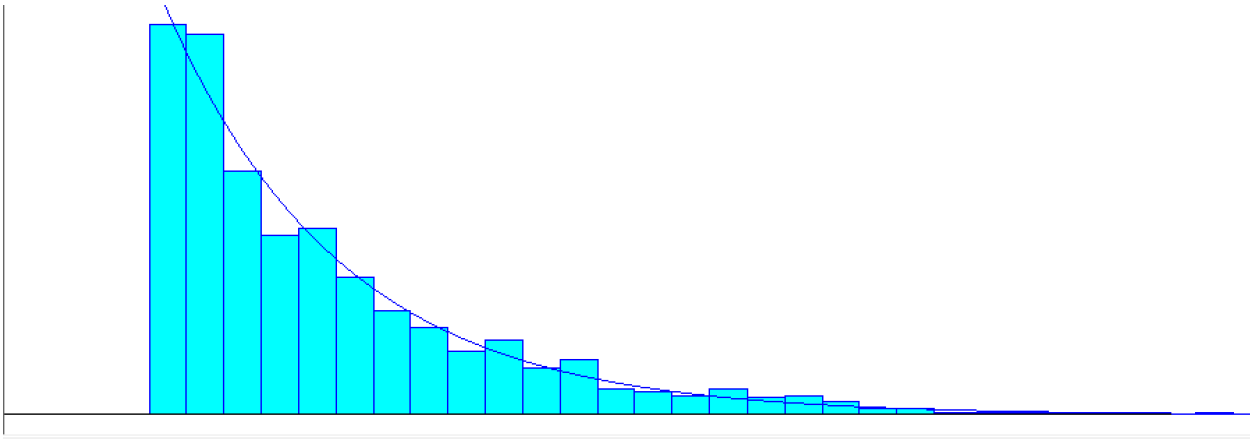


This Q-Q plot doesn't look bad but let's investigate whether we can find a better fit. We decided to try the exponential and weibull distribution since our histogram is right skewed. Q-Q plot of

both of them look like a better fit so we choose the one with least square error. The most suitable one is the exponential distribution with square error 0.000783. Also the square error is the same for exponential and gamma so it's okay to choose exponential instead of gamma. (Note that parameters are estimated using MLE)



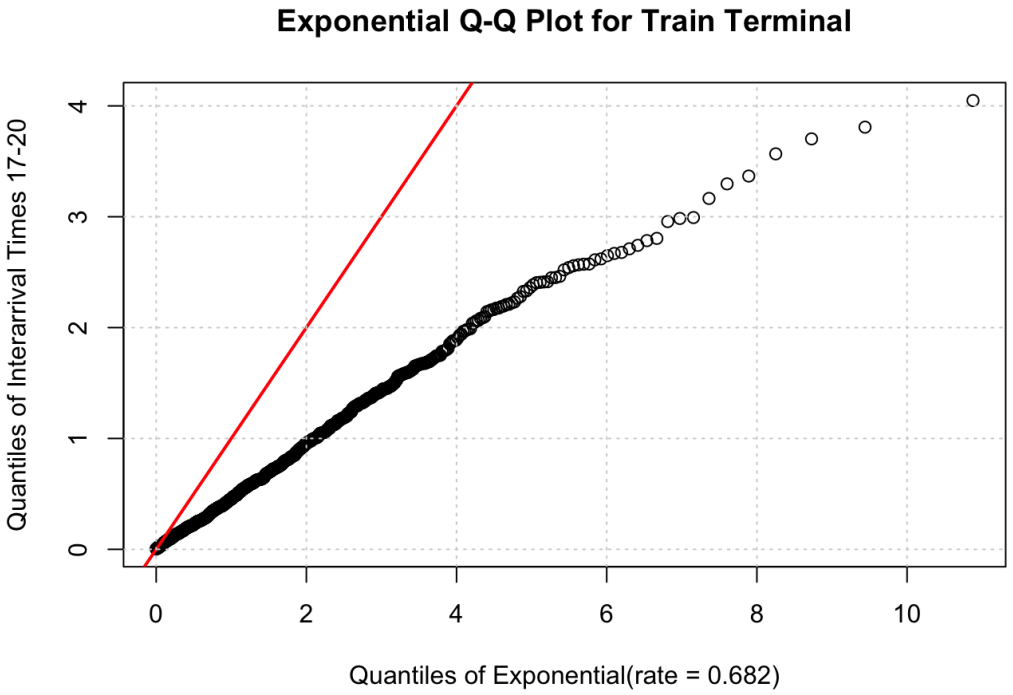
- **Train Terminal (17-20):**



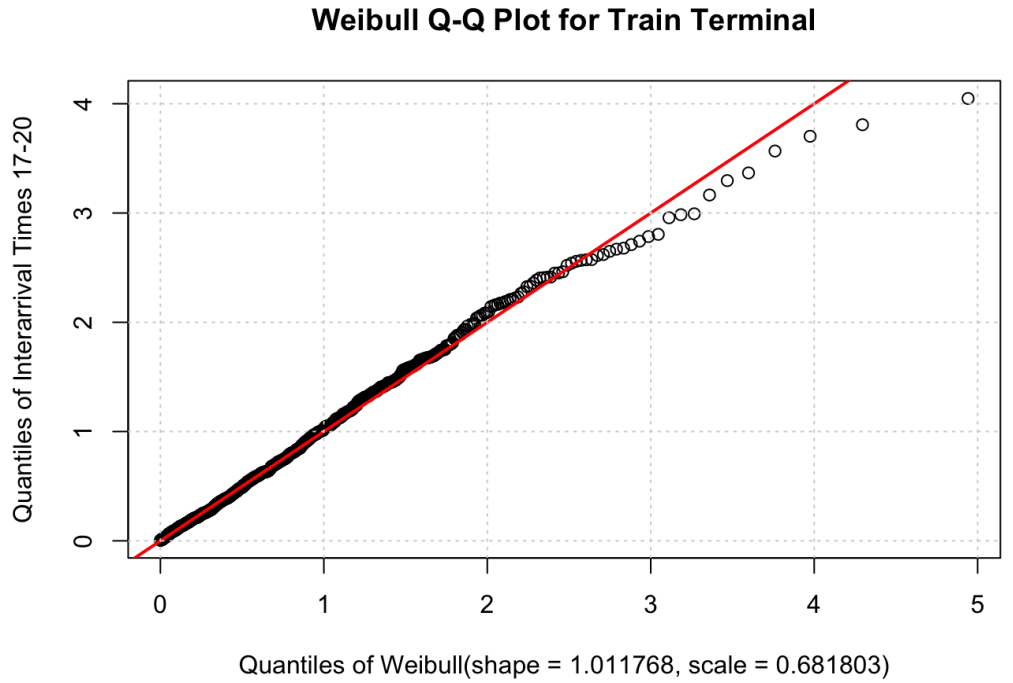
```
       Distribution Summary

Distribution:   Exponential
Expression:     EXPO(0.682)
Square Error:   0.001306
```

The interarrival times between 17-20 best fitted to the exponential distribution with parameter (0.682) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous

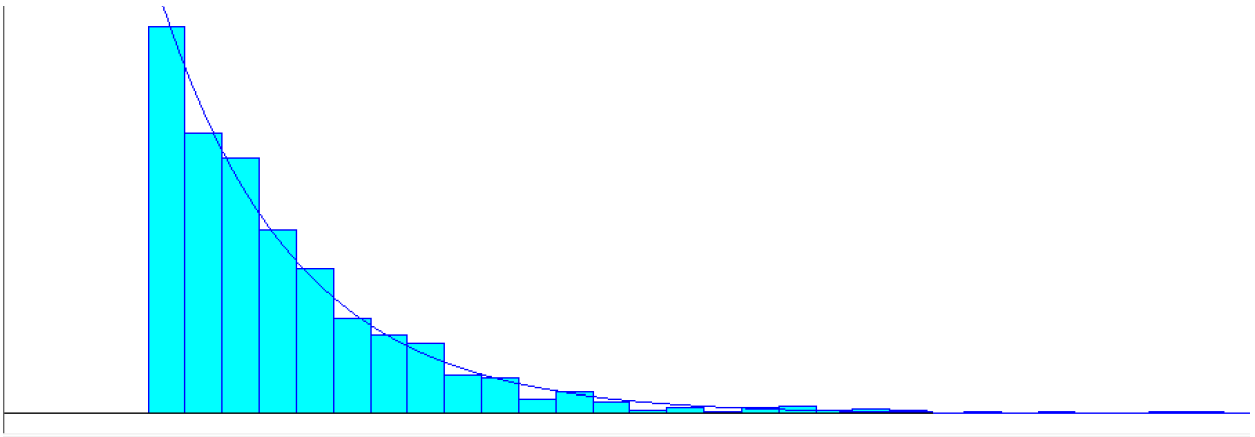
distributions, which is 0.001306. However, when we graph it as a Q-Q plot using RStudio, the following plot obtained:



This Q-Q plot looks like a very poor fit. Therefore, we decided to try the gamma and weibull distribution since our histogram is right skewed. Both of them look like a good fit so we choose the one with least square error. The most suitable one is the weibull distribution with square error 0.001307. We improved our fit increasing error square only 0.000001.

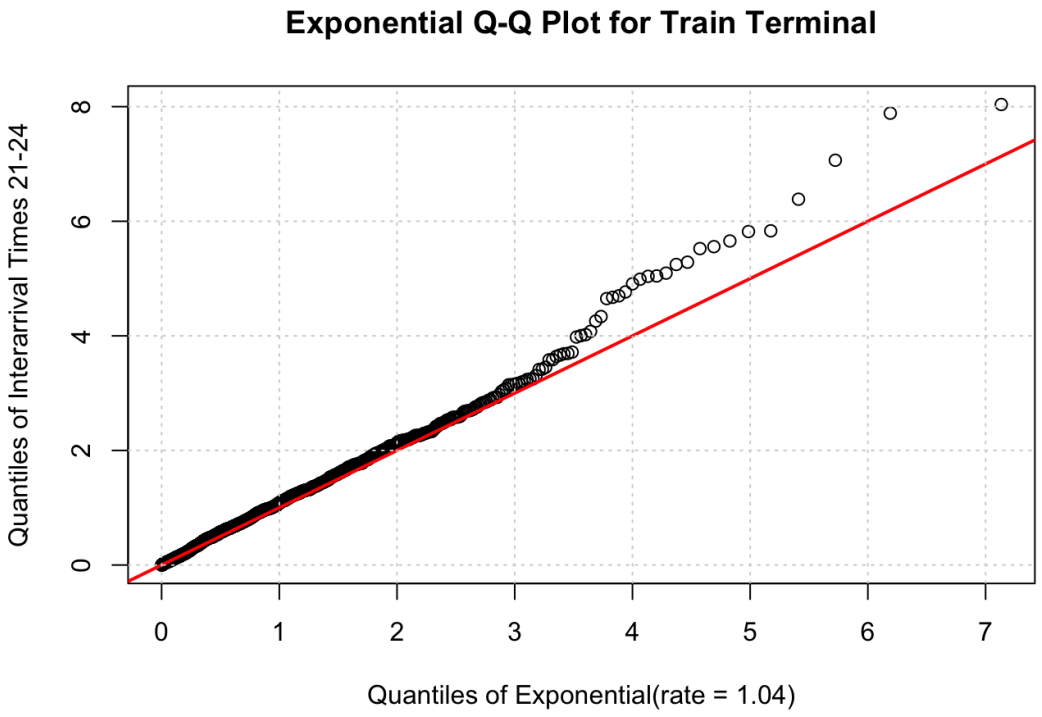


- **Train Terminal (21-24):**



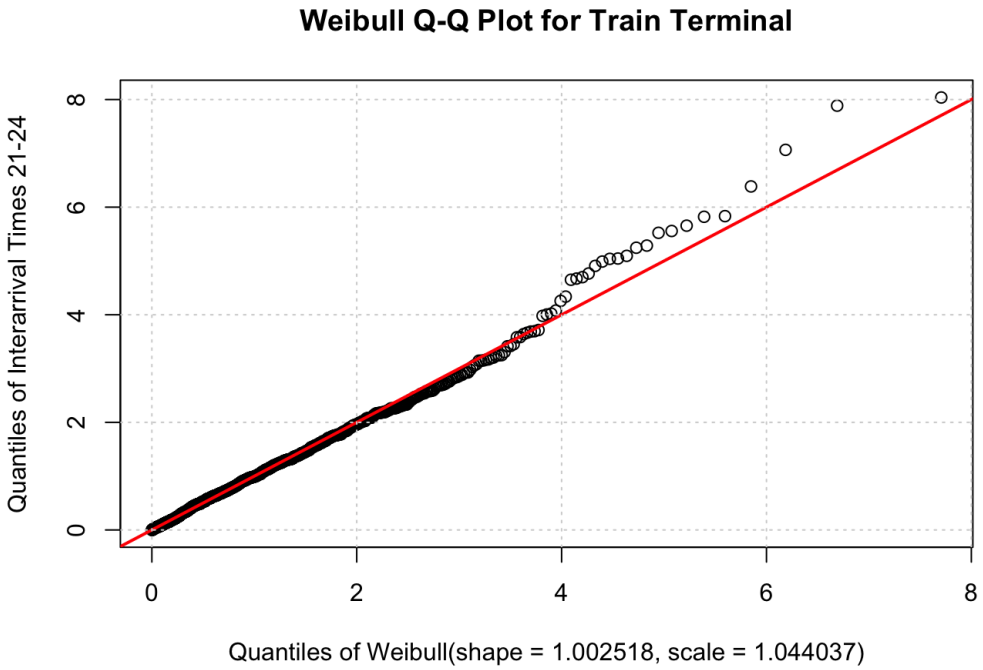
```
       Distribution Summary

Distribution:   Exponential
Expression:     EXPO(1.04)
Square Error:   0.000706
```

The interarrival times between 21-24 best fitted to the exponential distribution with parameter (1.04) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.000706. However, when we graph it as a Q-Q plot using RStudio, the following plot obtained:

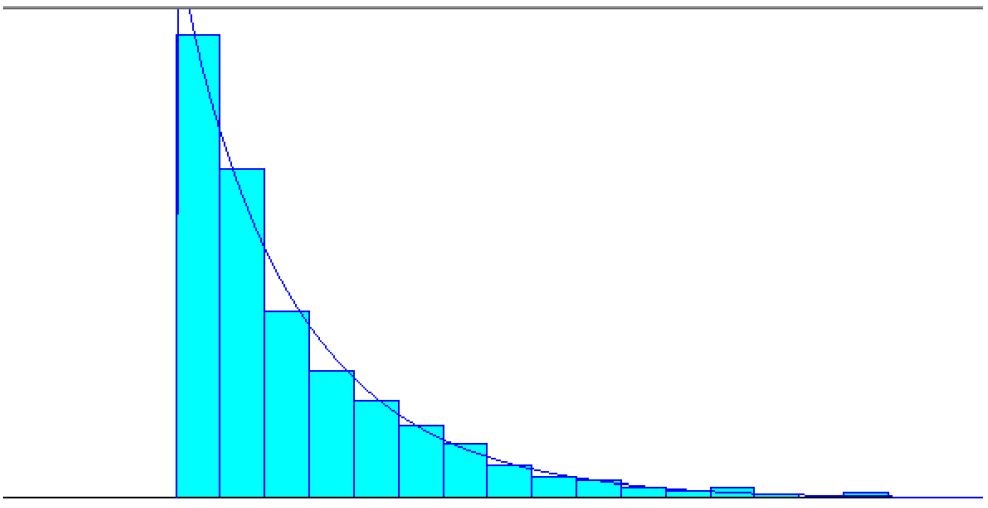


This Q-Q plot looks like a poor fit. Therefore, we decided to try the gamma and weibull distribution since our histogram is right skewed. Both of them look like a good fit so we choose

the one with least square error. The most suitable one is the weibull distribution with square error 0.000954.



- **Other Locations (1-4):**

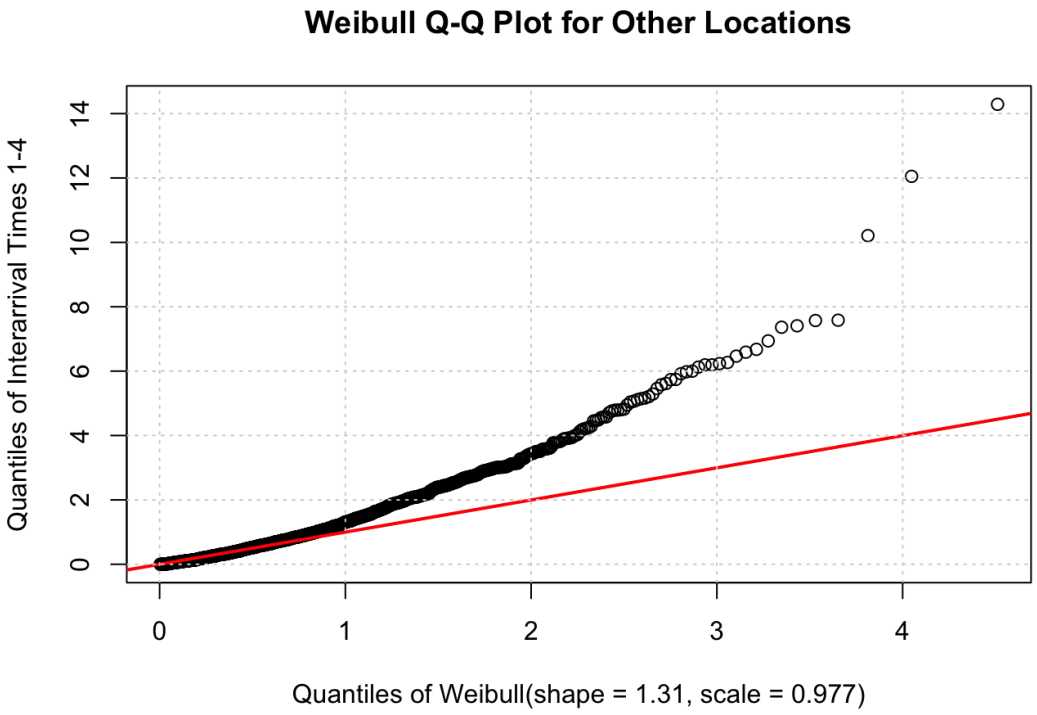


Distribution Summary

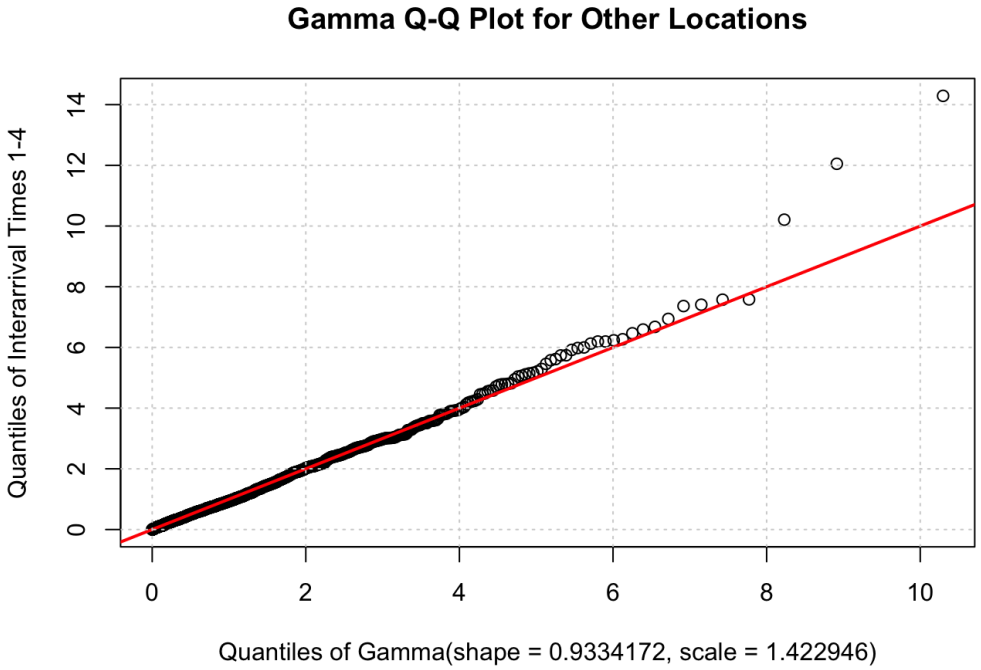
Distribution: Weibull
Expression: WEIB(1.31, 0.977)
Square Error: 0.000782

The interarrival times between 1-4 best fitted to the weibull distribution with parameters (1.31, 0.977) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous

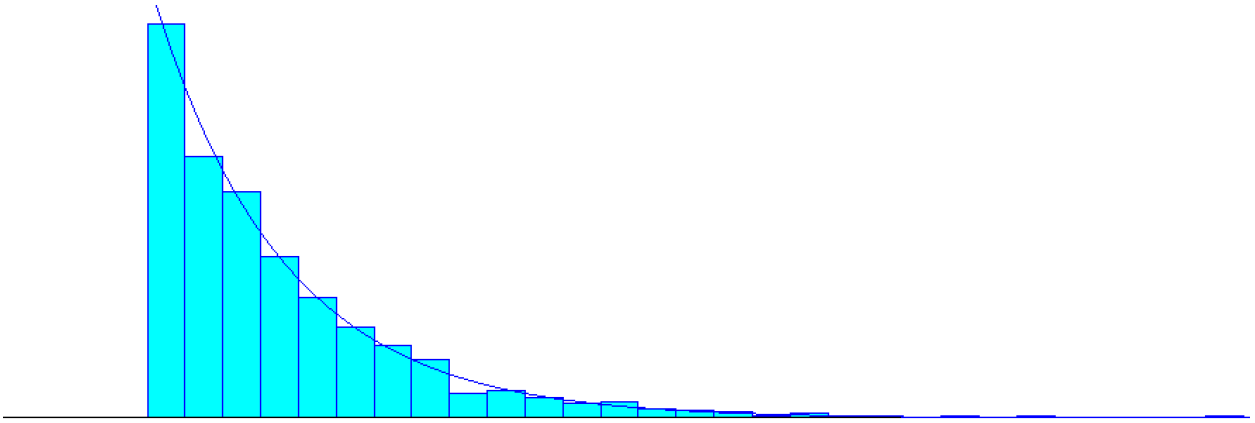
distributions, which is 0.000782. However, when we graph it as a Q-Q plot using RStudio, the following plot obtained:



This Q-Q plot looks like a poor fit. Therefore, we decided to try the gamma and exponential distribution since our histogram is right skewed. Both of them look like a good fit so we choose the one with least square error. The most suitable one is the gamma distribution with square error 0.000809.

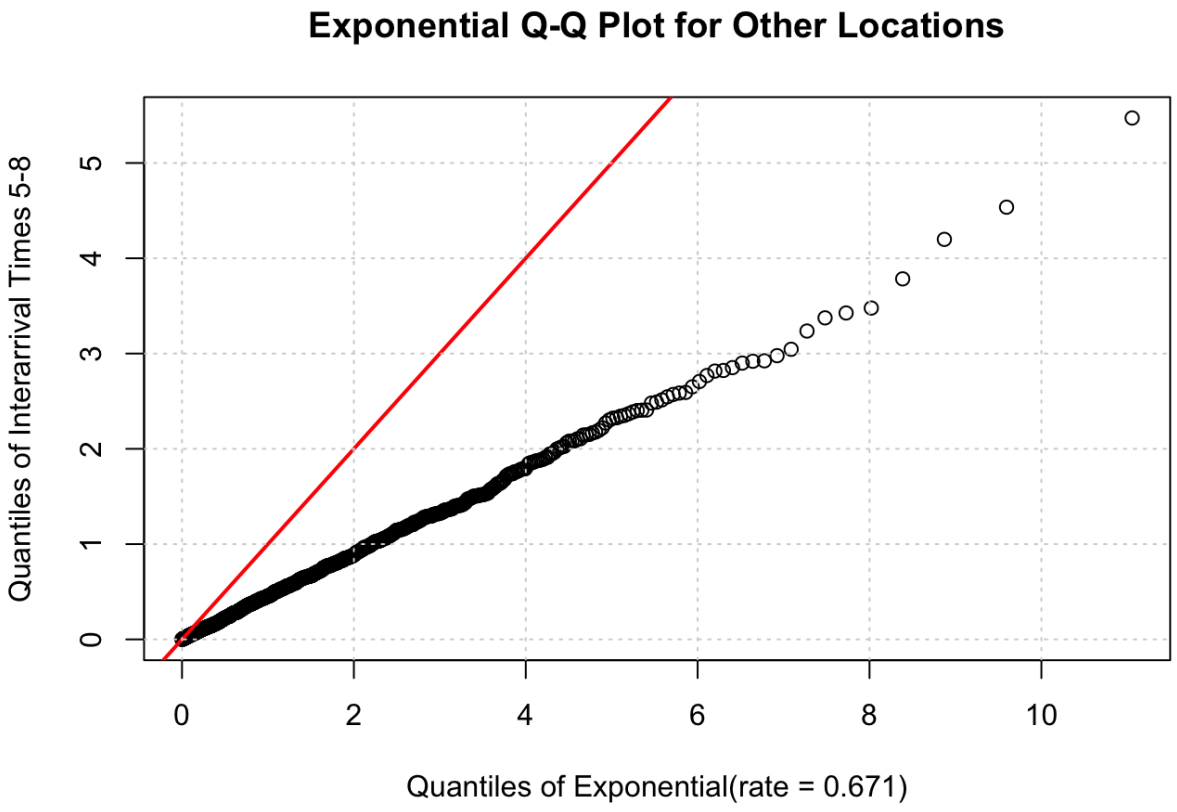


- **Other Locations (5-8):**

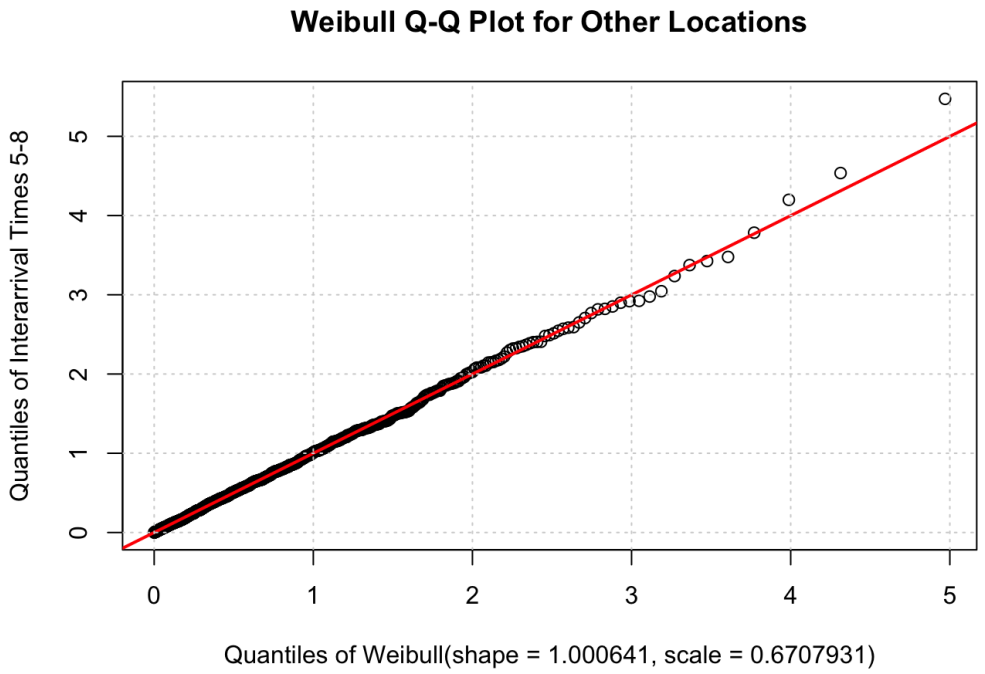


```
       Distribution Summary

Distribution:   Exponential
Expression:     EXPO(0.671)
Square Error:   0.000493
```

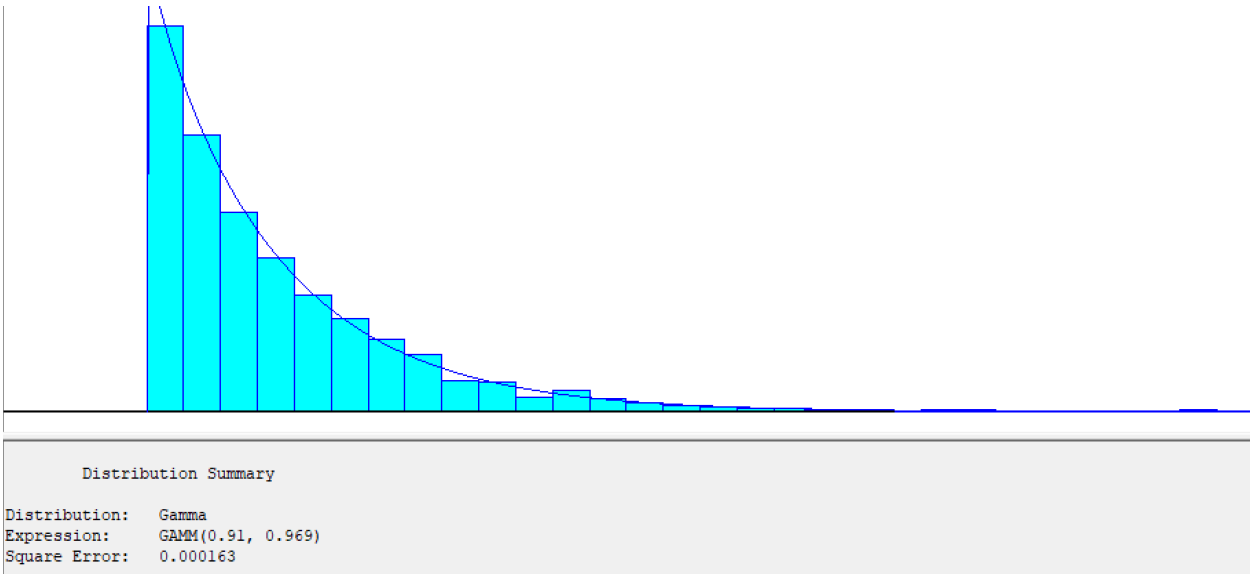
The interarrival times between 5-8 best fitted to the exponential distribution with parameter (0.671) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.000493. However, when we graph it as a Q-Q plot using RStudio, the following plot obtained:



This Q-Q plot looks like a very poor fit. Therefore, we decided to try the gamma and weibull distribution since our histogram is right skewed. Both of them look like a good fit so we choose the one with least square error. The most suitable one is the weibull distribution with square error 0.000526.

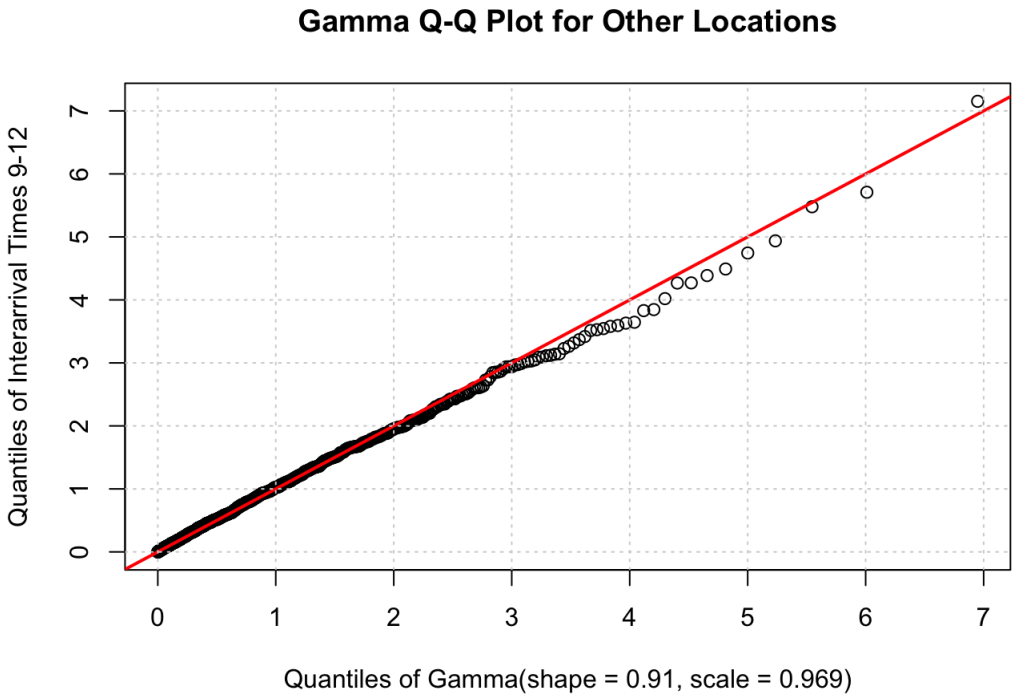


- **Other Locations (9-12):**

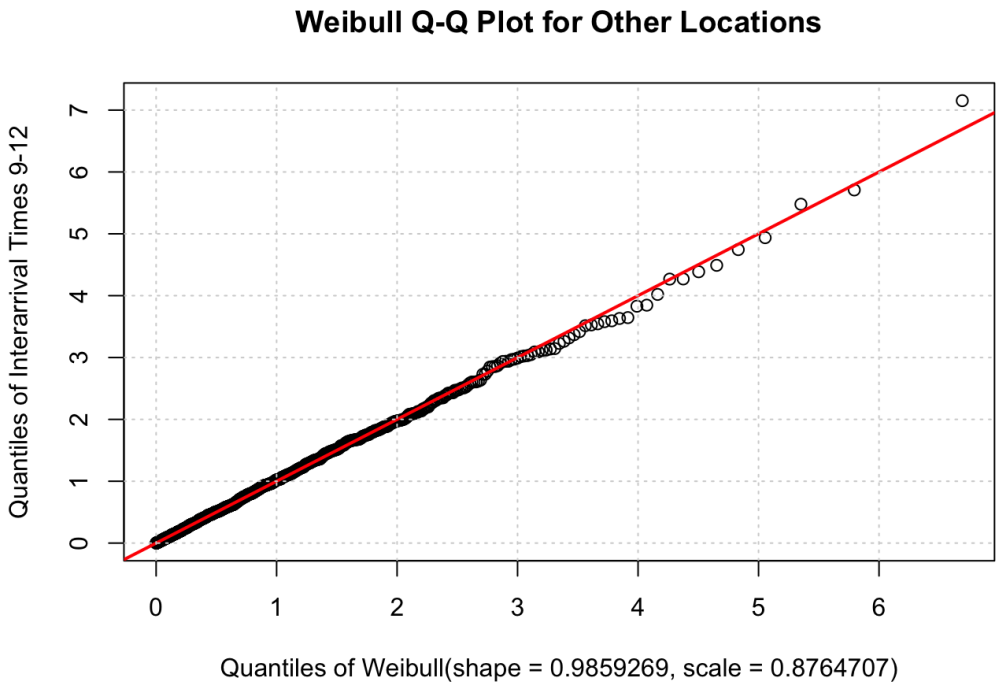


The interarrival times between 9-12 best fitted to the gamma distribution with parameters (0.91, 0.969) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous

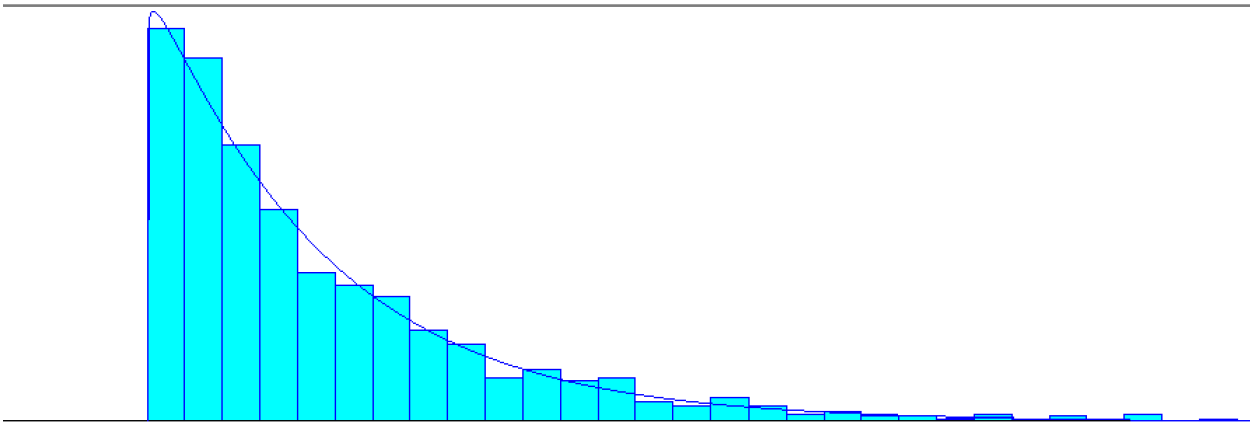
distributions, which is 0.000163. To check whether it's a good fit we graph it as a Q-Q plot using RStudio, the following plot obtained:



This Q-Q plot doesn't look bad but let's investigate whether we can find a better fit. We decided to try the exponential and weibull distribution since our histogram is right skewed. Q-Q plot of both of them look like a better fit so we choose the one with least square error. The most suitable one is the weibull distribution with square error 0.000211. The difference between square errors for weibull and gamma is very small which is 0.000048 but our initial fit was really good so there's no need to increase square error. We choose gamma distribution. (Note that parameters are estimated using MLE)

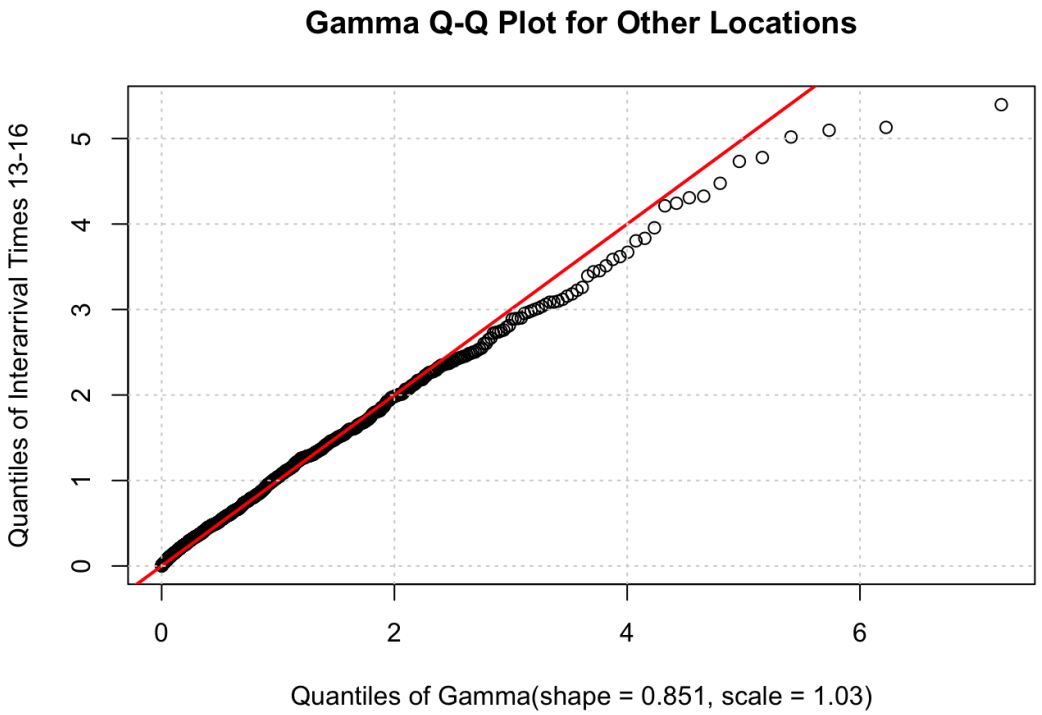


- **Other Locations (13-16):**

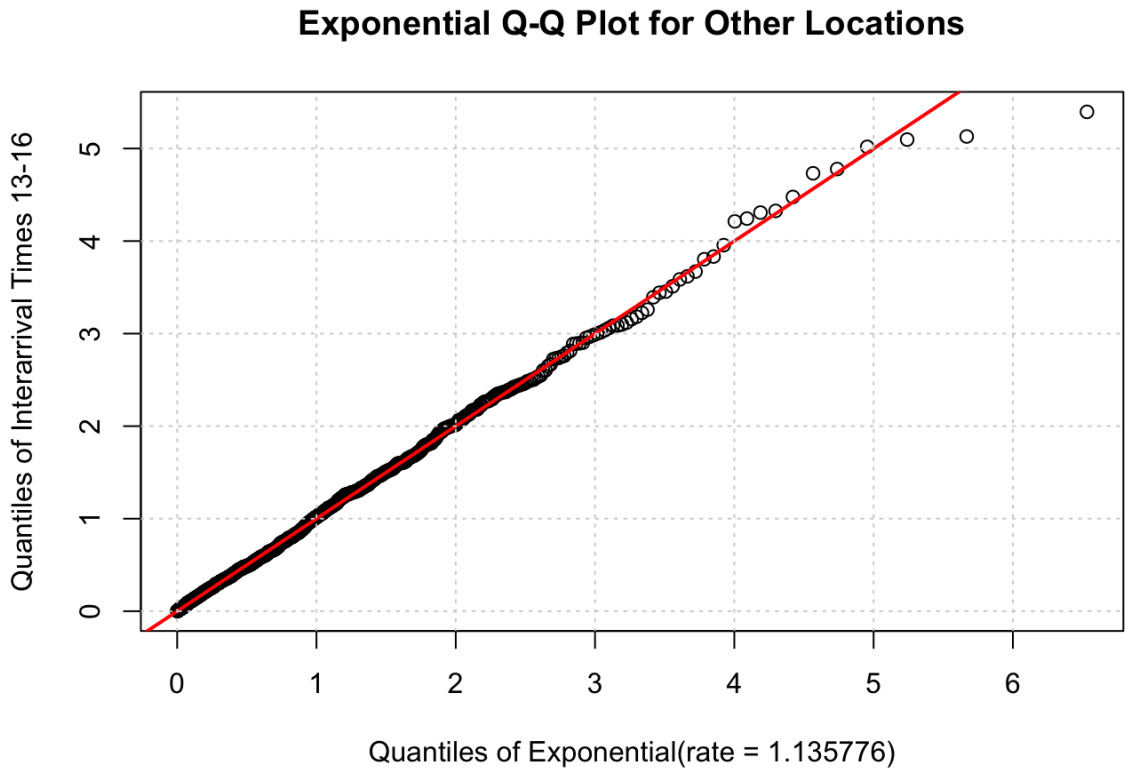


```
      Distribution Summary

Distribution:   Gamma
Expression:     GAMM(0.851, 1.03)
Square Error:   0.000644
```

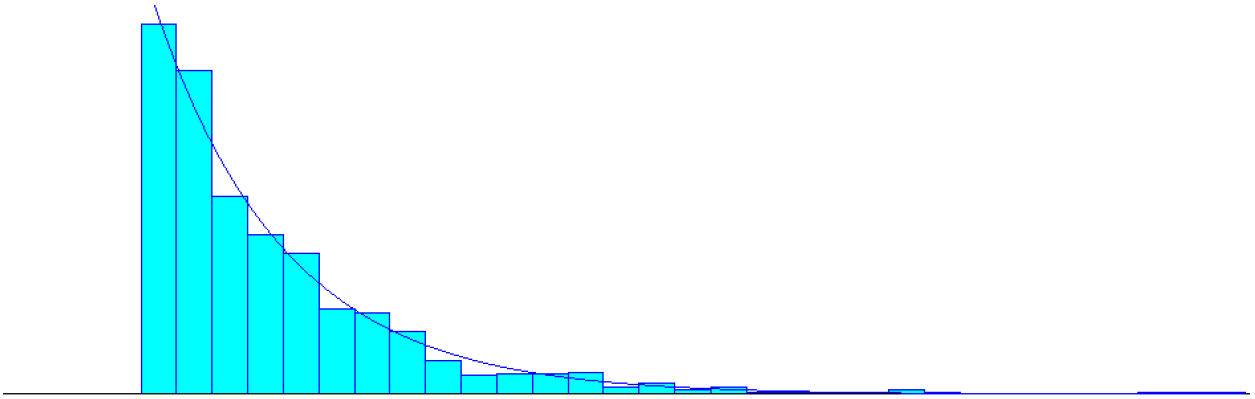
The interarrival times between 13-16 best fitted to the gamma distribution with parameters (0.851, 1.03) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.000644. However, when we graph it as a Q-Q plot using RStudio, the following plot obtained:



This Q-Q plot looks like a poor fit. Therefore, we decided to try the exponential and weibull distribution since our histogram is right skewed. Both of them look like a good fit so we choose the one with least square error. The most suitable one is the exponential distribution with square error 0.000649.

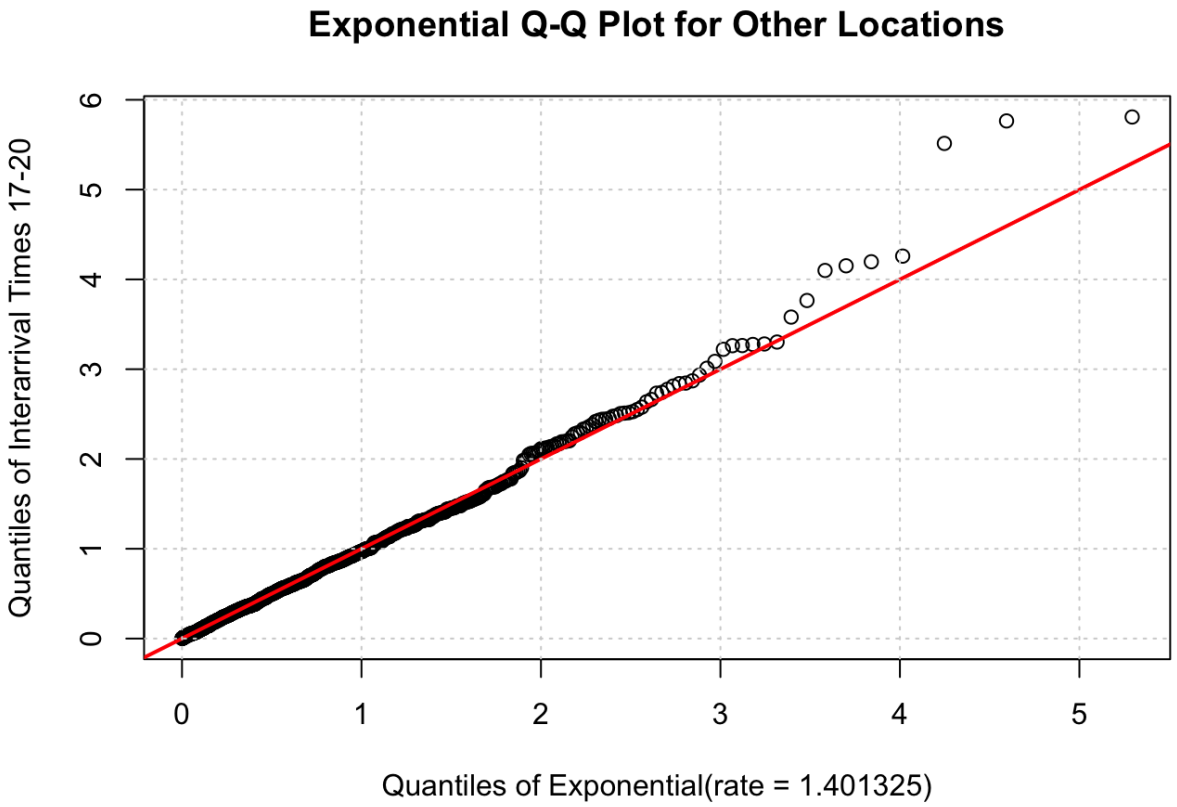


- **Other Locations (17-20):**

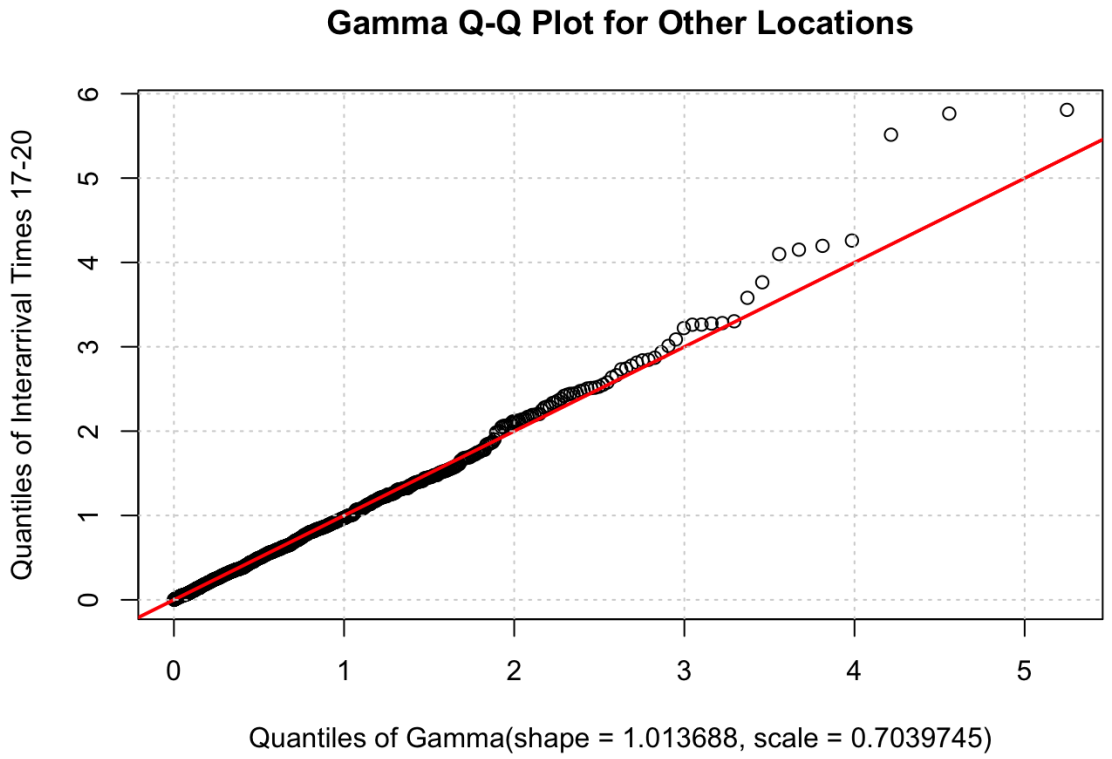


```
       Distribution Summary

Distribution:   Exponential
Expression:     EXPO(0.714)
Square Error:   0.001068
```

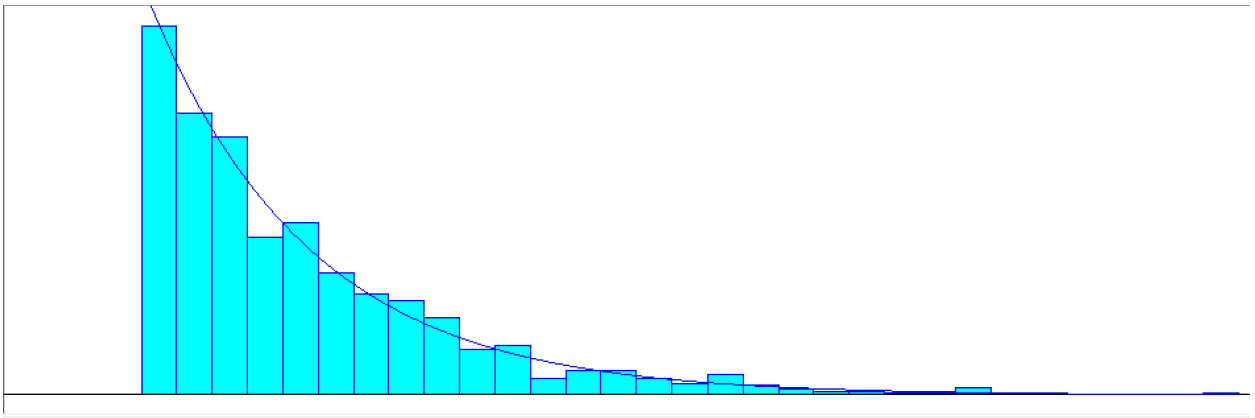
The interarrival times between 17-20 best fitted to the exponential distribution with parameter (0.714) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.001068. To check whether it's a good fit we graph it as a Q-Q plot using RStudio, the following plot obtained:



This Q-Q plot doesn't look bad but let's investigate whether we can find a better fit. We decided to try the gamma and weibull distribution since our histogram is right skewed. Q-Q plot of both of them look like a better fit so we choose the one with least square error. The most suitable one is the gamma distribution with square error 0.001171. Since there isn't much improvement in fit, there's no need to increase square error. We choose exponential distribution.

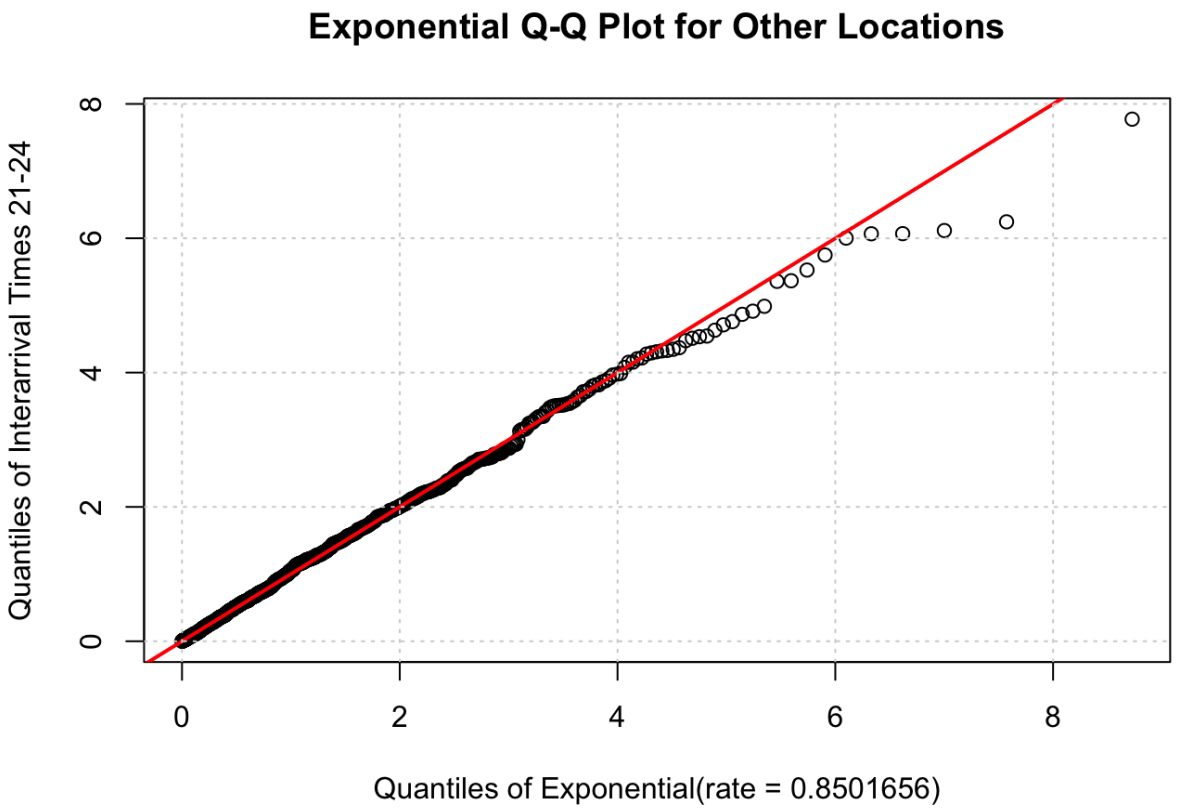


- **Other Locations (21-24):**

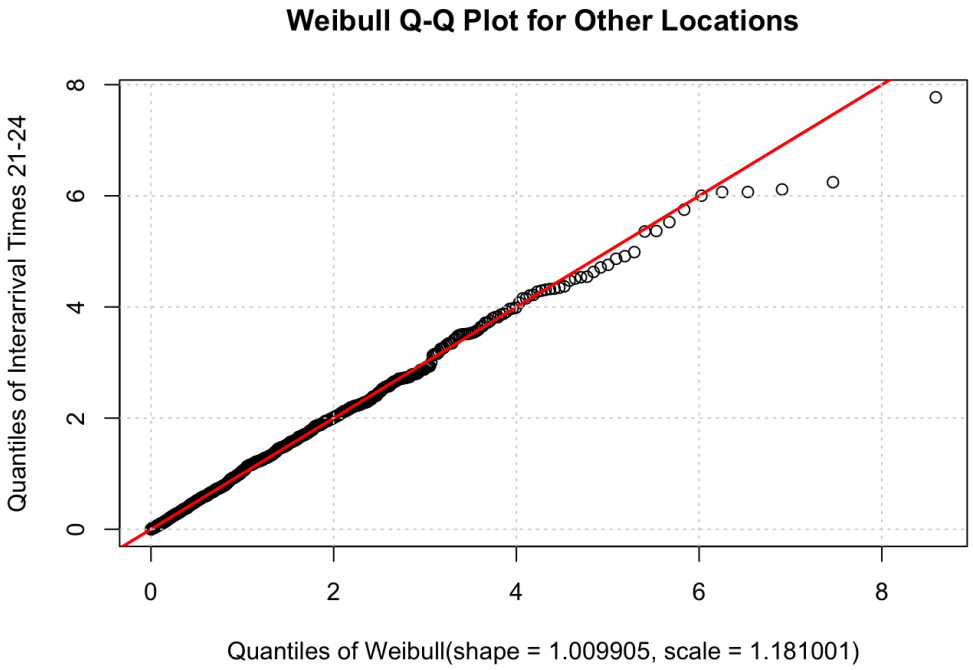


```
       Distribution Summary

Distribution:   Exponential
Expression:     EXPO(1.18)
Square Error:   0.000848
```

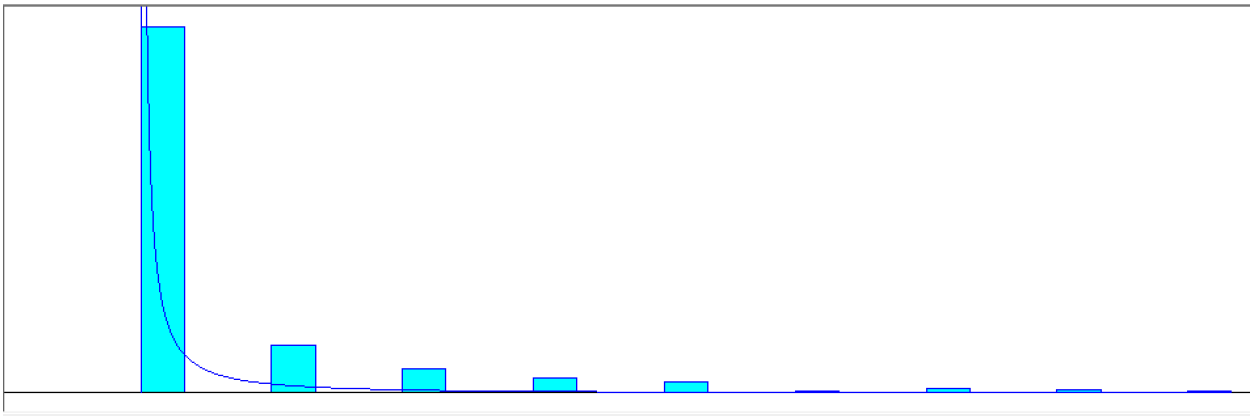
The interarrival times between 21-24 best fitted to the exponential distribution with parameter (1.18) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried continuous distributions, which is 0.000848. To check whether it's a good fit we graph it as a Q-Q plot using RStudio, the following plot obtained:



This Q-Q plot doesn't look bad but let's investigate whether we can find a better fit. We decided to try the gamma and weibull distribution since our histogram is right skewed. Q-Q plot of both of them look like a better fit so we choose the one with least square error. The most suitable one is the weibull distribution with square error 0.000964. Since there isn't much improvement in the fit, there's no need to increase square error. We choose exponential distribution.

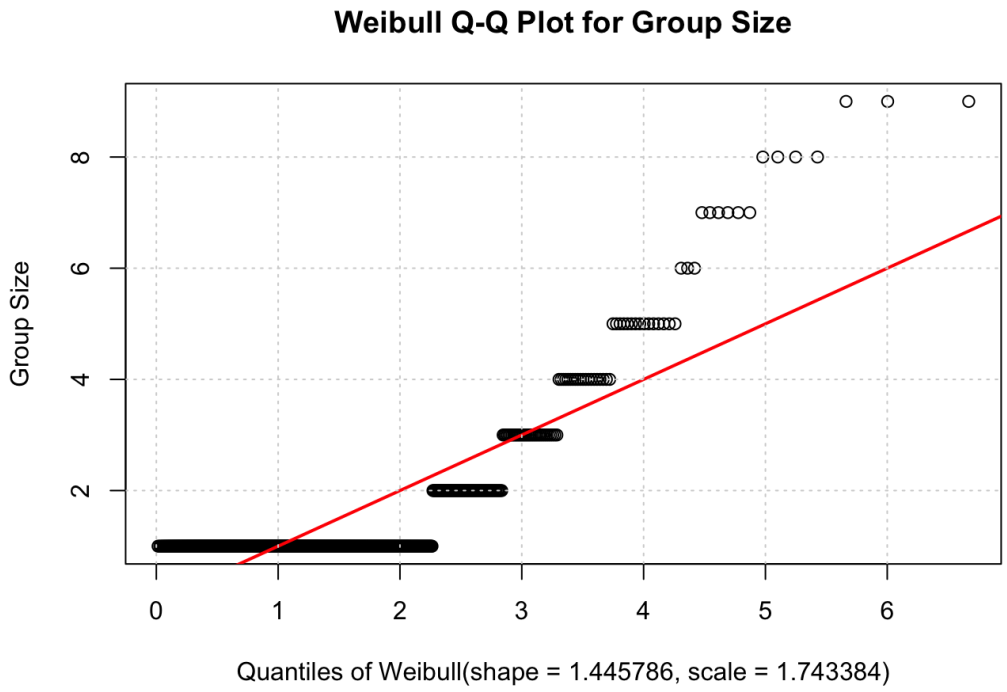


Party Size:



```
      Distribution Summary

Distribution:   Weibull
Expression:     0.999 + WEIB(0.0341, 0.274)
Square Error:   0.019481
```

The group size fitted to the weibull with expression 0.999 + weibull(0.0341, 0.274) which is according to the outputs of Input Analyzer with the Distribution Summary and Histogram Summary above. It has the least square error among other tried distributions, which is 0.019481. Let's see the Q-Q plot using RStudio:



We can see that it's a poor fit but first of all we should note that our data set is discrete, therefore we decided that empirical distribution is better.