sample route4

Mingze Li 300137754

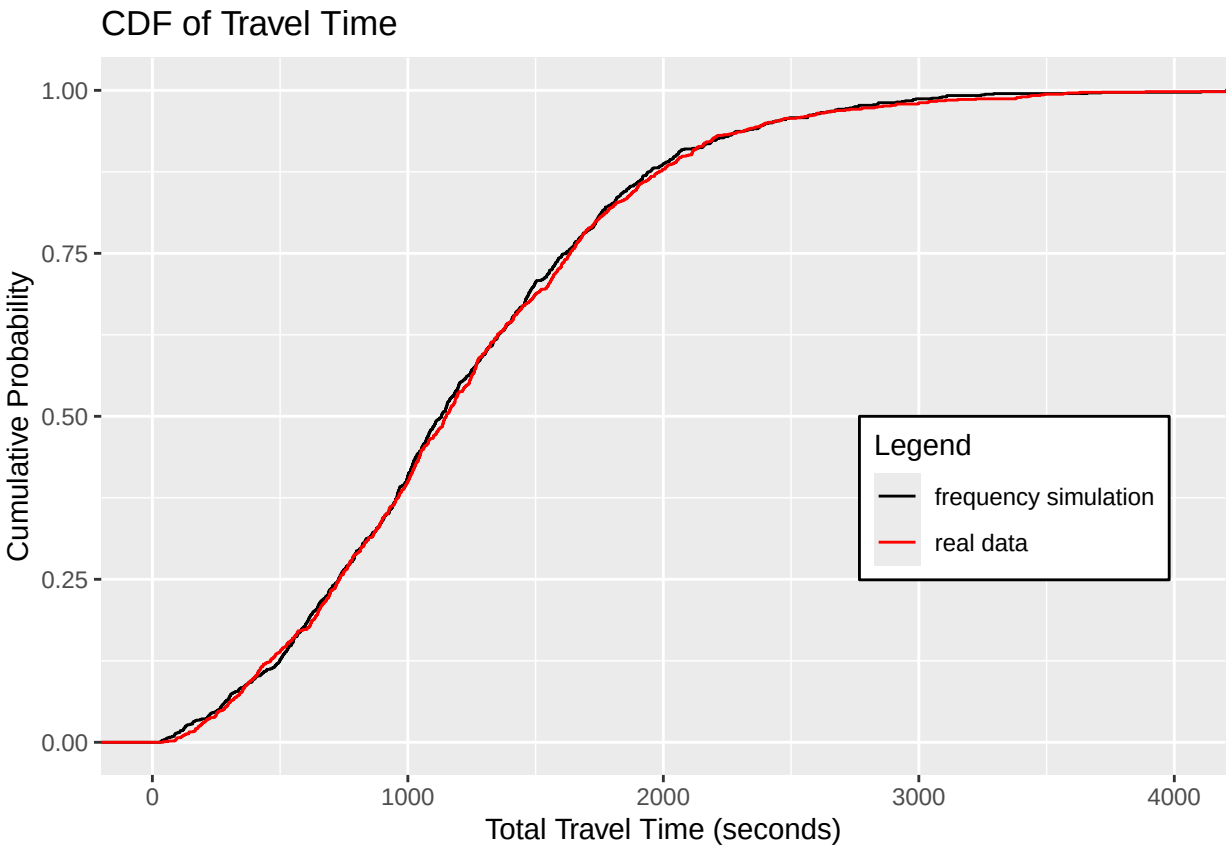
2025-03-15

```
#source('traveltimeCLTfunctions.R')
library(traveltimeCLT)
library(data.table)
```

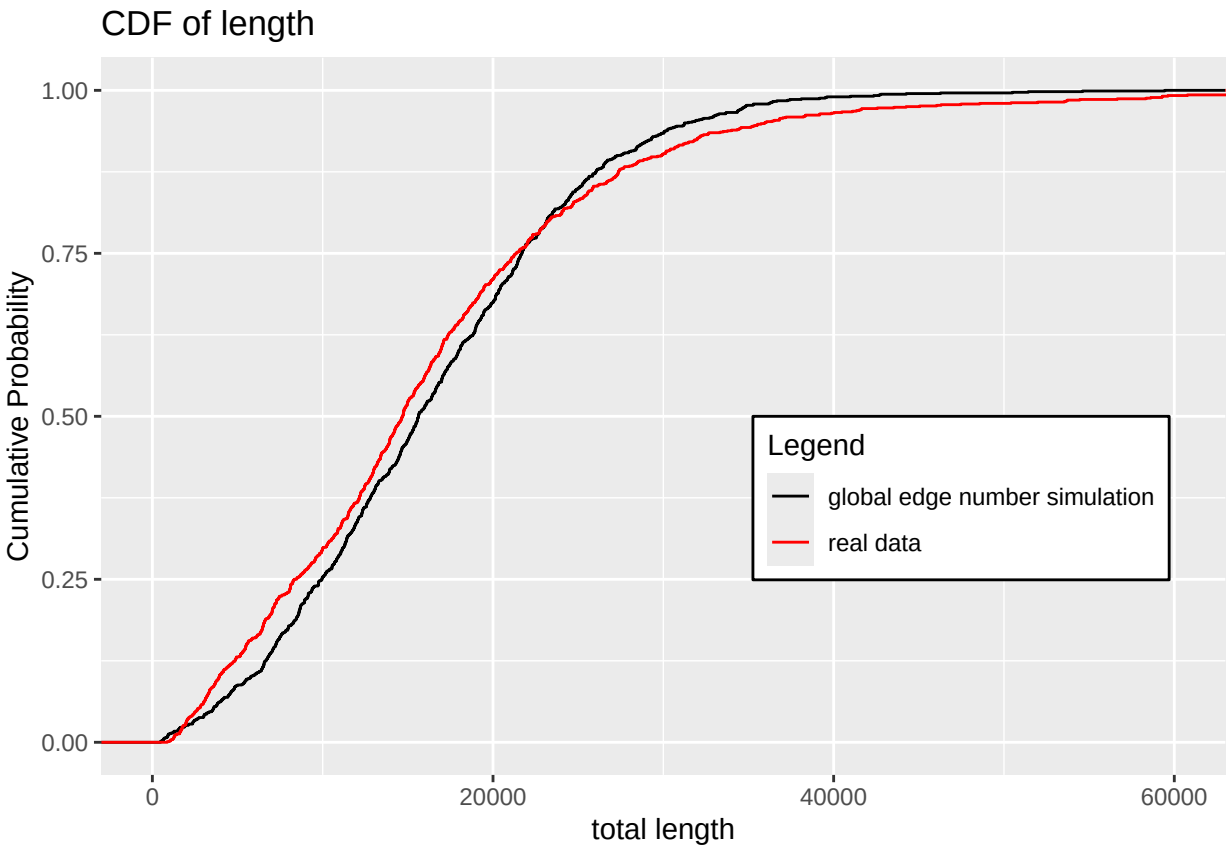
```
## Warning: package 'data.table' was built under R version 4.3.3
```

```
trips <- fread('data/trips.csv')
id <- sample(unique(trips$trip),1000)
train = trips[!trips$trip %in% id,]
test =  trips[trips$trip %in% id,]
timeBin_x_edge <- get_timeBin_x_edges(train)
```

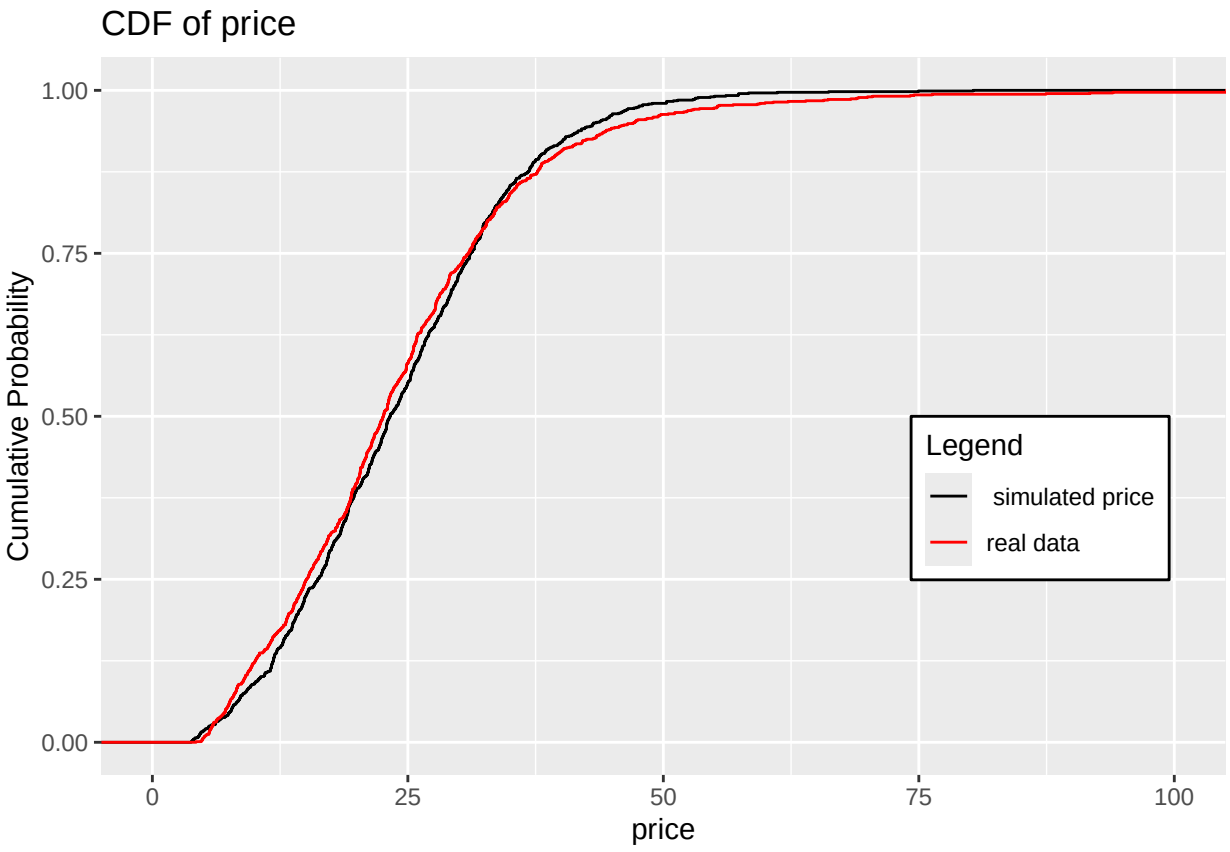
```
#sample_route(id,trips,1001)
sample <- sample_route(id,trips,1001,timeBin_x_edge)
pressure_test <- sample_route(id,trips,1001,timeBin_x_edge,abuse_ratio = 0.4,abuse_multiplier = 1.1)
plot_CDF_compare(sample[[2]]$real_time,sample[[1]]$dependent_time,"frequency simulation")
```




```
plot_CDF_compare(sample[[2]]$real_length,sample[[1]]$simulated_length,"global edge number simulation","
```




```
real_price=price(sample[[2]]$real_time,sample[[2]]$real_length)[,1]
simulated_price<-data.table(simulated_price=price(sample[[1]]$dependent_time,sample[[1]]$simulated_leng
plot_CDF_compare(real_price,simulated_price$simulated_price," simulated price","price","CDF of price",1
```




```
names(train)[c(2,3,5,7,8)]=c("tripID","entry_time","duration_secs","distance_meters","linkID")
train$speed=exp(train$logspeed)
train$timeBin=time_bins_readable(train$entry_time)
fit <- traveltimeCLT(train)
```

```
## Warning in traveltimeCLT(train): 4 trips have less than 1 observation, and will
## not be used to estimate autocorrelations, or residual variance parameters
```

```
test = trips[trips$trip %in% id,]
names(test)[c(2,3,5,7,8)]=c("tripID","entry_time","1","distance_meters","linkID")
names(pressure_test[[3]])=c("tripID","linkID","entry_time","distance_meters")
p=predict(fit, test)
pressure_p=predict(fit,pressure_test[[3]])
fit2 <- traveltimeCLT(train, model = 'population')
p2=predict(fit2, test)
pressure_p2=predict(fit2,pressure_test[[3]])
```

```
start_times <- test[, .(start_time = entry_time[1]), by = tripID]
pressure_start_time <- (pressure_test[[3]][, .(start_time = entry_time[1]), by = tripID][,2])
pressure_start_time <-pressure_start_time[["start_time"]]
R1=request_R(p,start_times$start_time-300,start_times$start_time,sample[[2]]$real_length,1,risk_free=0)
R2=request_R(p2,start_times$start_time-300,start_times$start_time,sample[[2]]$real_length,1,risk_free=0)
pressure_R1=request_R(pressure_p,pressure_start_time-300,pressure_start_time,pressure_test[[1]]$simulate
pressure_R2=request_R(pressure_p2,pressure_start_time-300,pressure_start_time,pressure_test[[1]]$simula
all(R1==R2)
```

```
## [1] FALSE
```

```
all(p==p2)
```

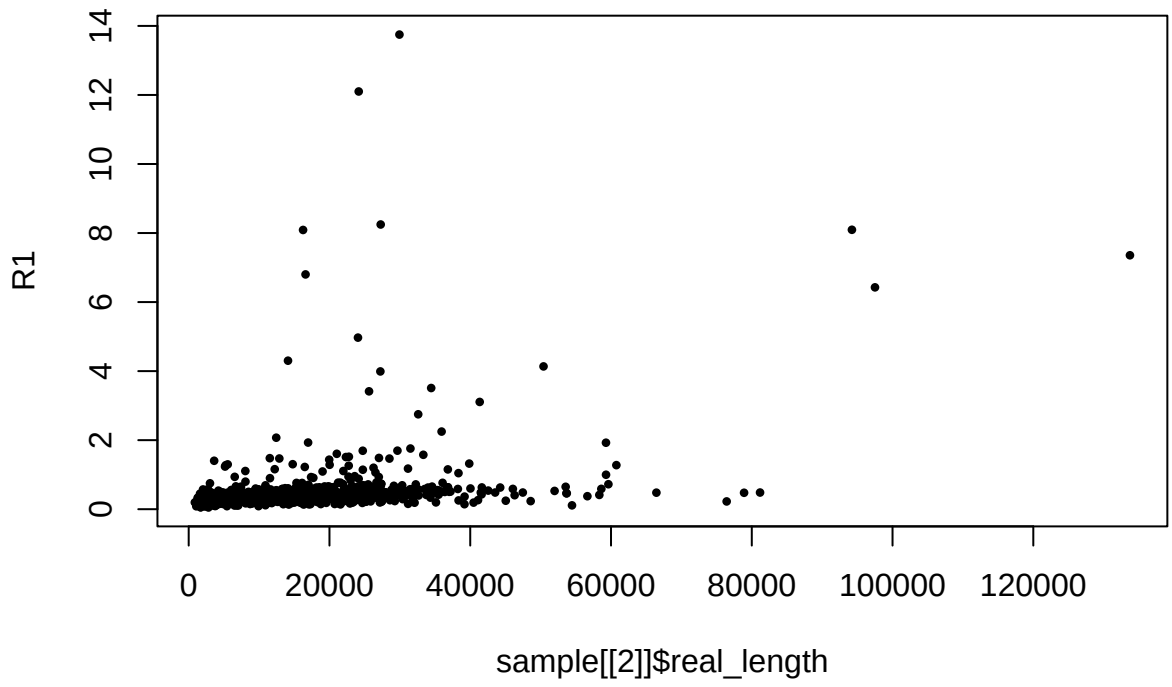
```
## [1] FALSE
```

```
all(R1>0)
```

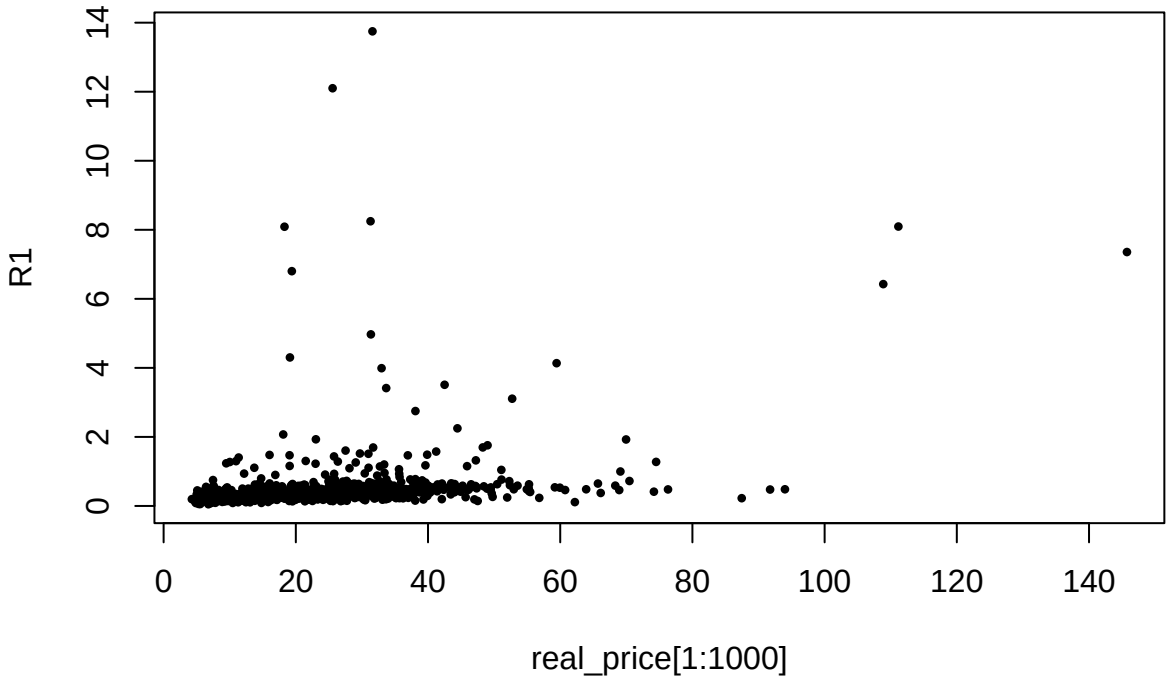
```
## [1] TRUE
```

```
#all(R1==R2,na.rm = T)
#all(p==p2,na.rm = T)
#which(is.na(p$variance)==T)
```

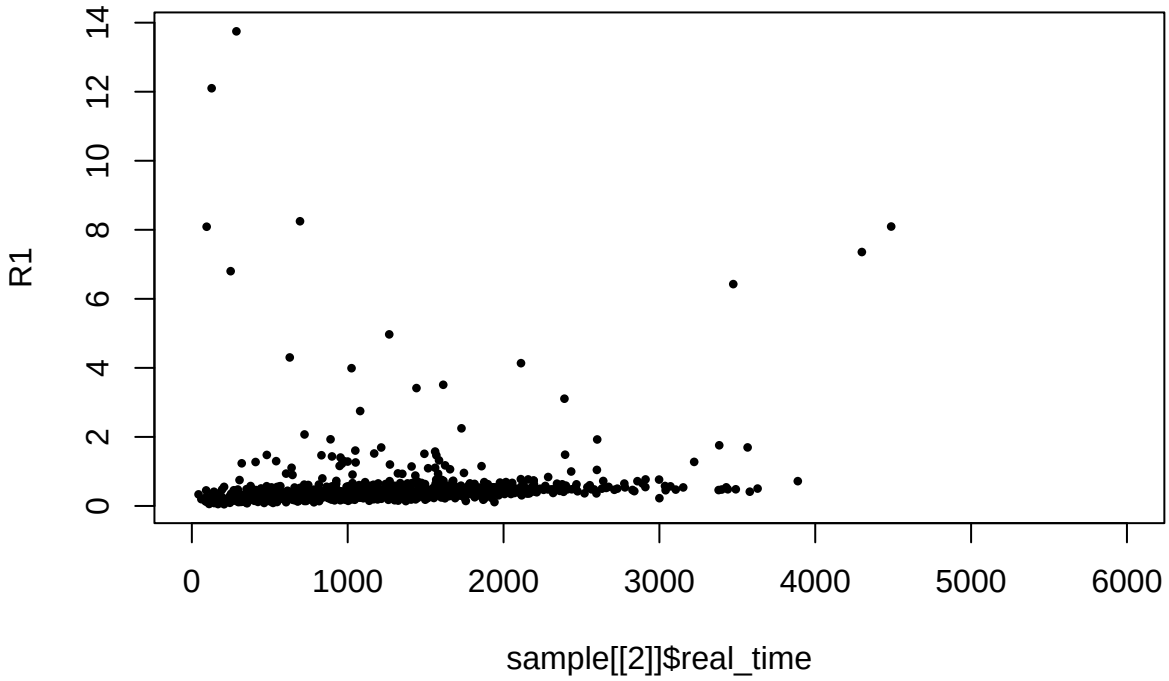
```
plot(sample[[2]]$real_length,R1,pch = 16,cex = 0.6)
```



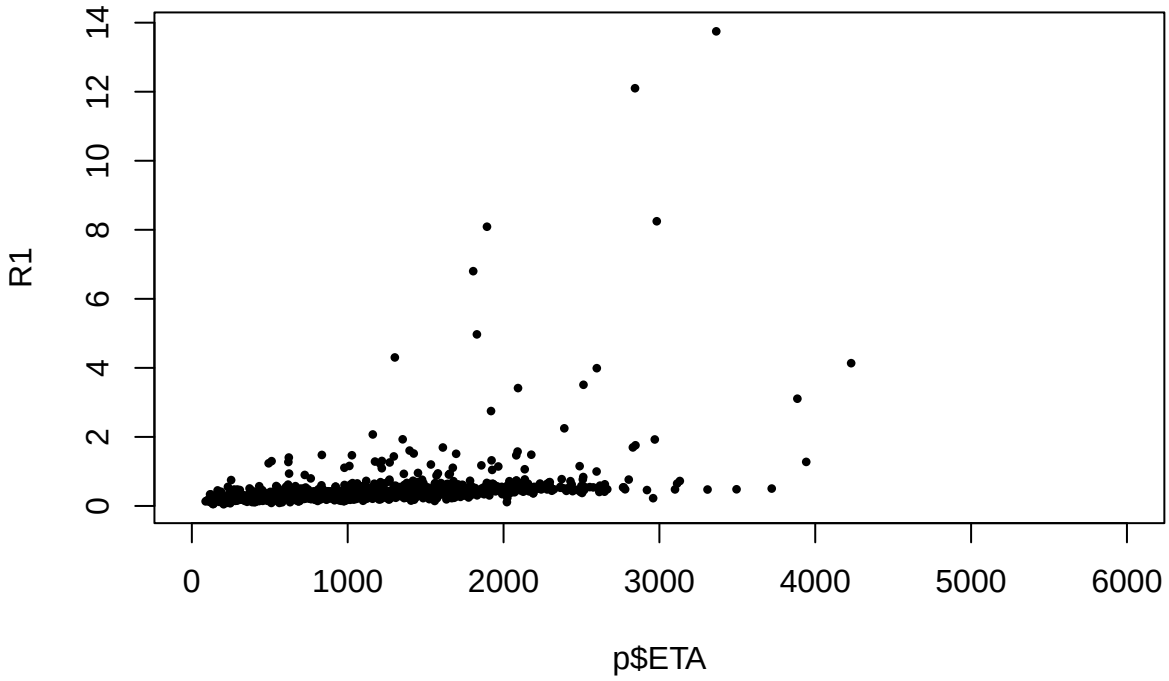

```
plot(real_price[1:1000],R1,pch = 16,cex = 0.6)
```



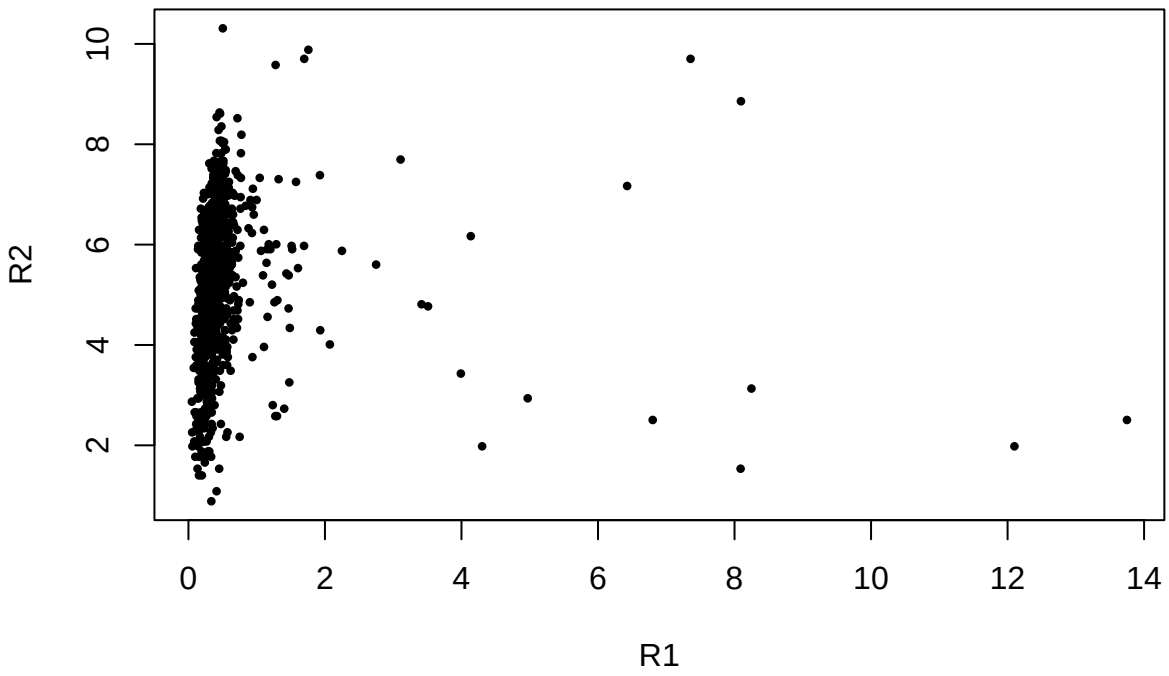

```
plot(sample[[2]]$real_time,R1,xlim = c(0, 6000),pch = 16,cex = 0.6)
```



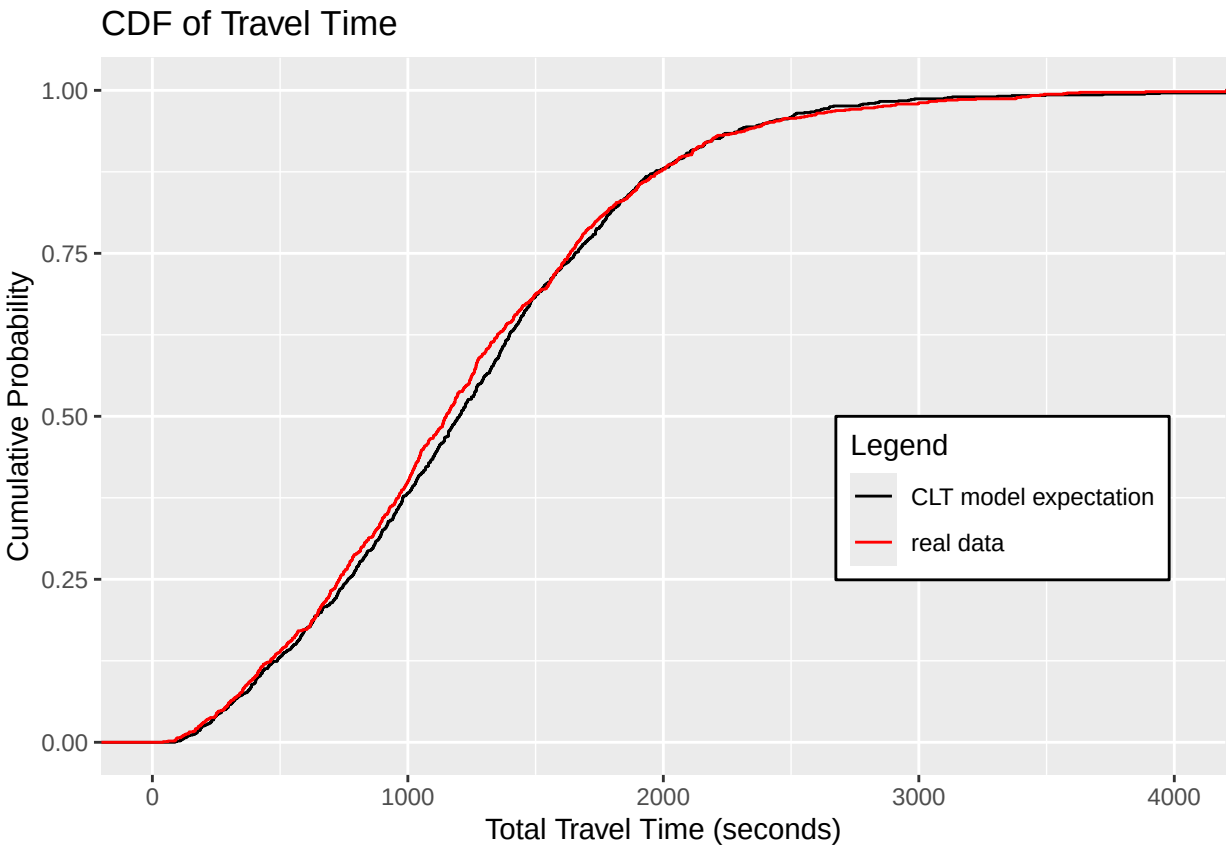

```
plot(p$ETA,R1,xlim = c(0, 6000),pch = 16,cex = 0.6)
```



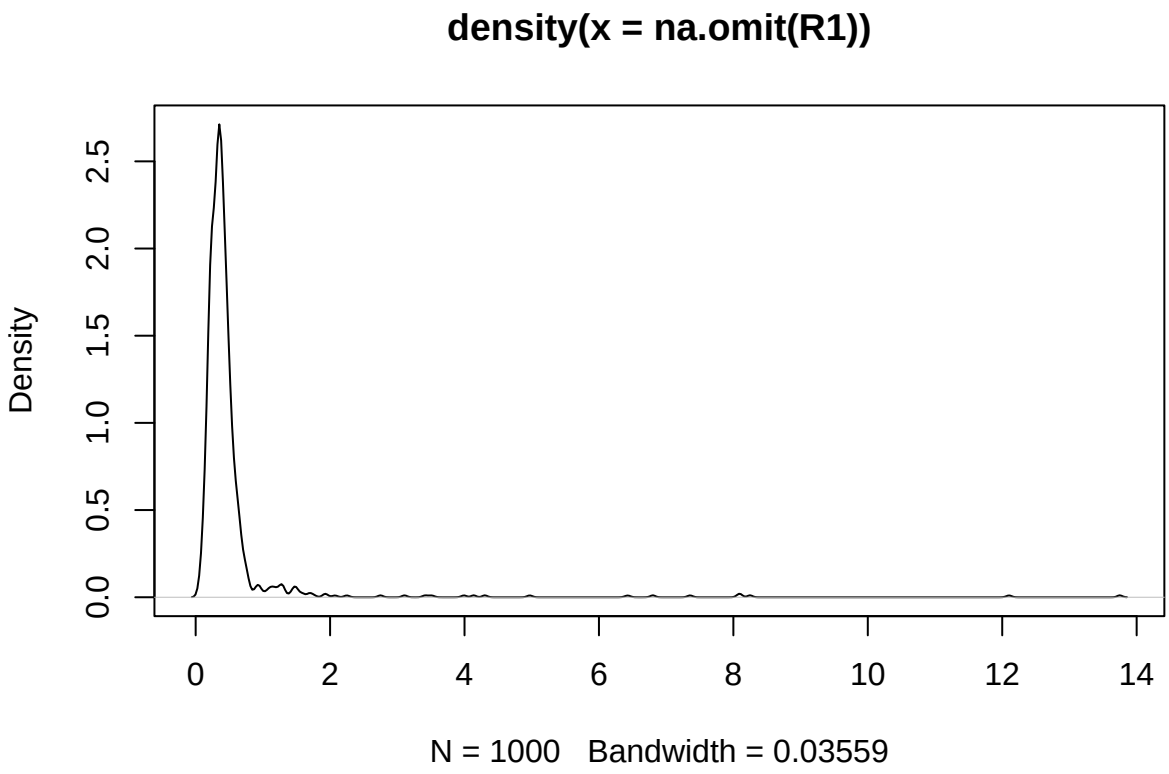

```
plot(R1,R2,pch = 16,cex = 0.6)
```



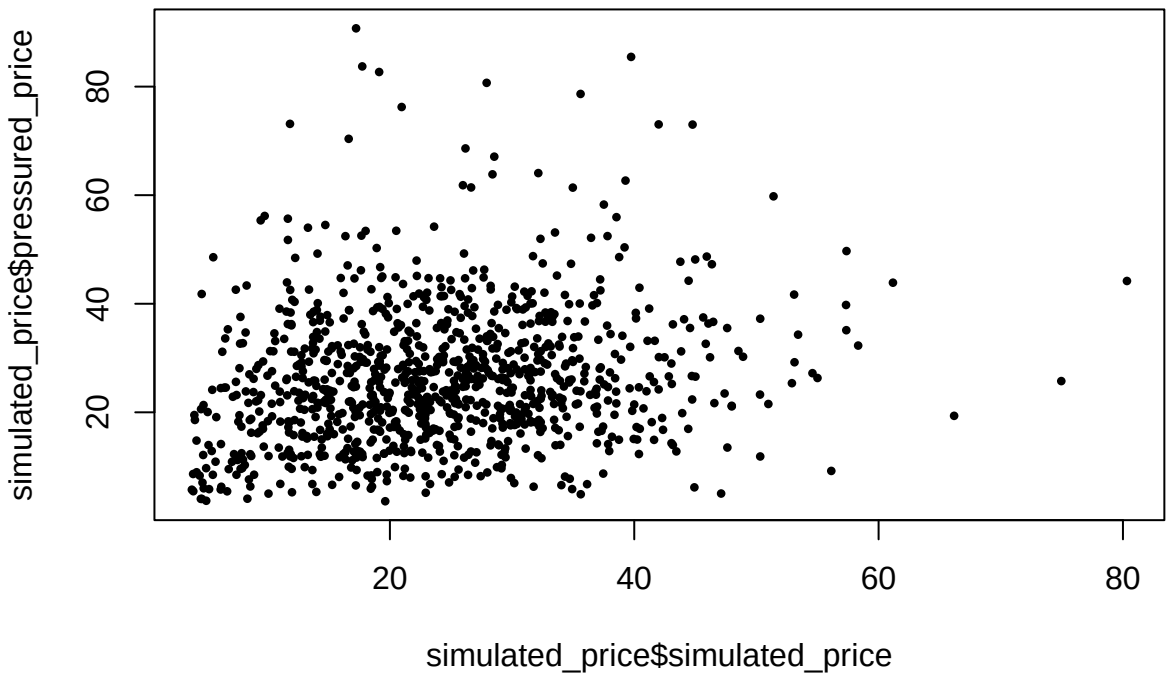

```
plot_CDF_compare(sample[[2]]$real_time,p$ETA,"CLT model expectation")
```



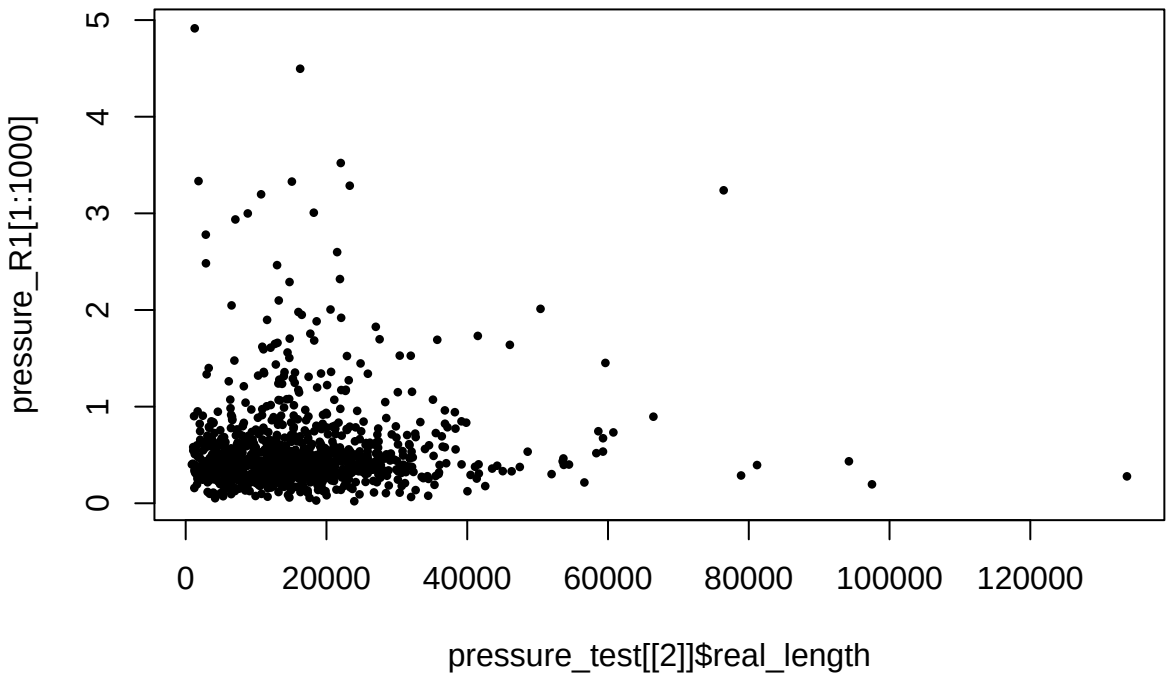

```
plot(density(na.omit(R1)))
```



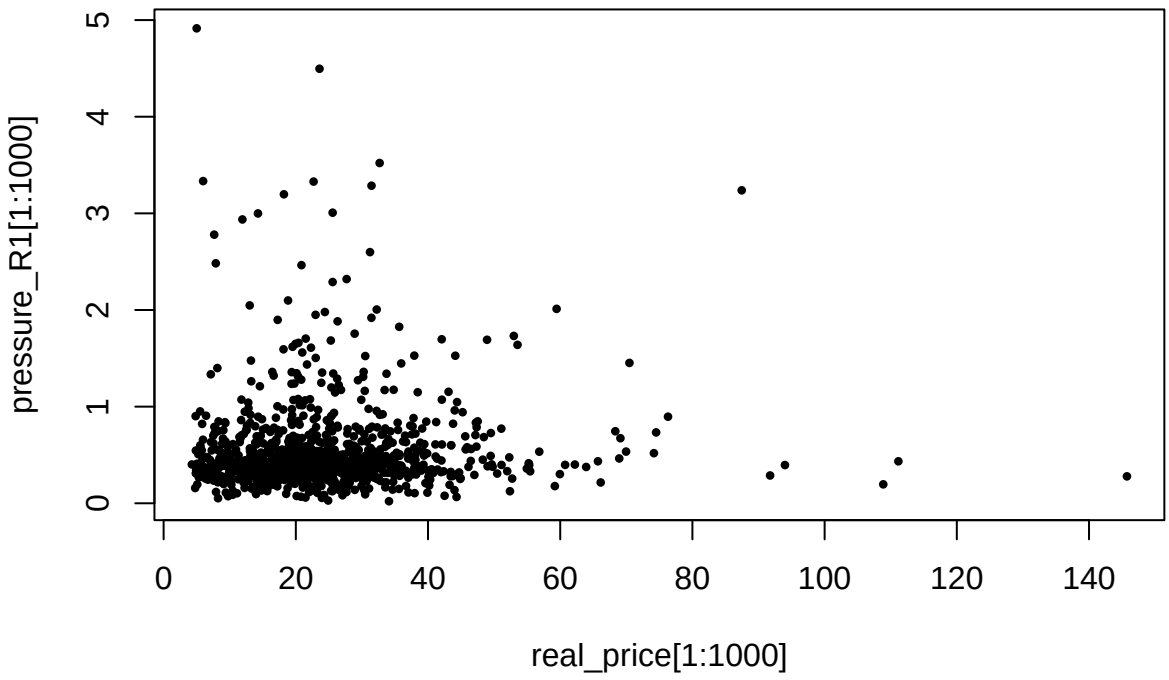

```
plot(simulated_price$simulated_price,simulated_price$pressured_price,pch = 16,cex = 0.6)
```



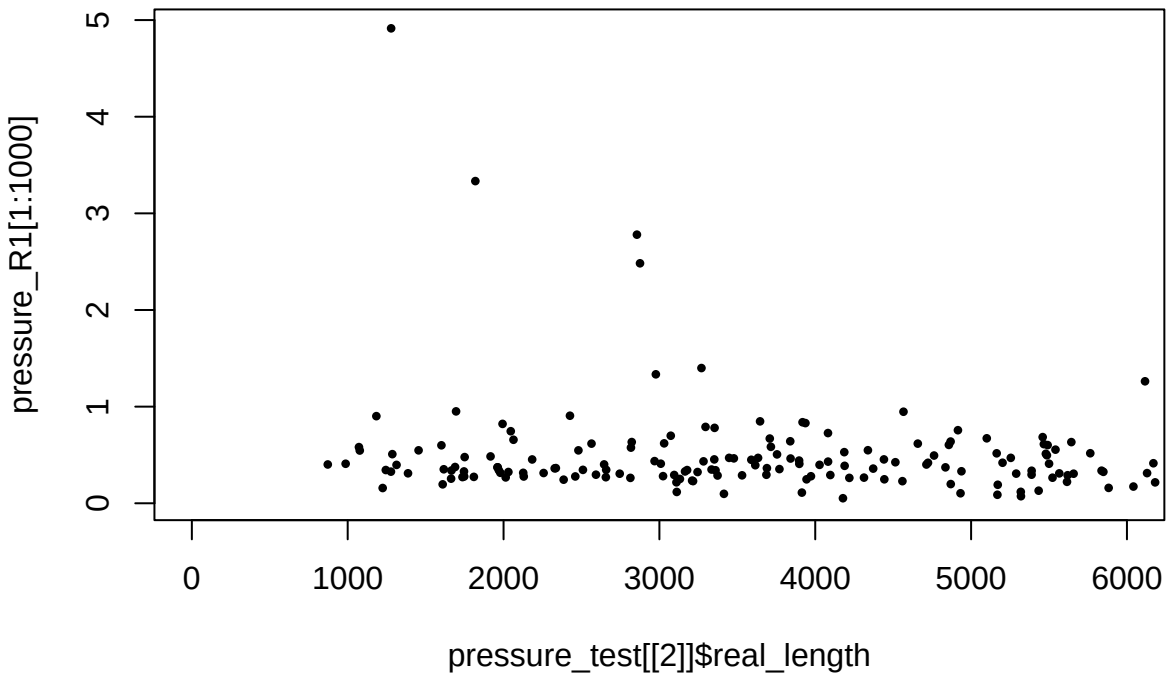

```
plot(pressure_test[[2]]$real_length,pressure_R1[1:1000],pch = 16,cex = 0.6)
```



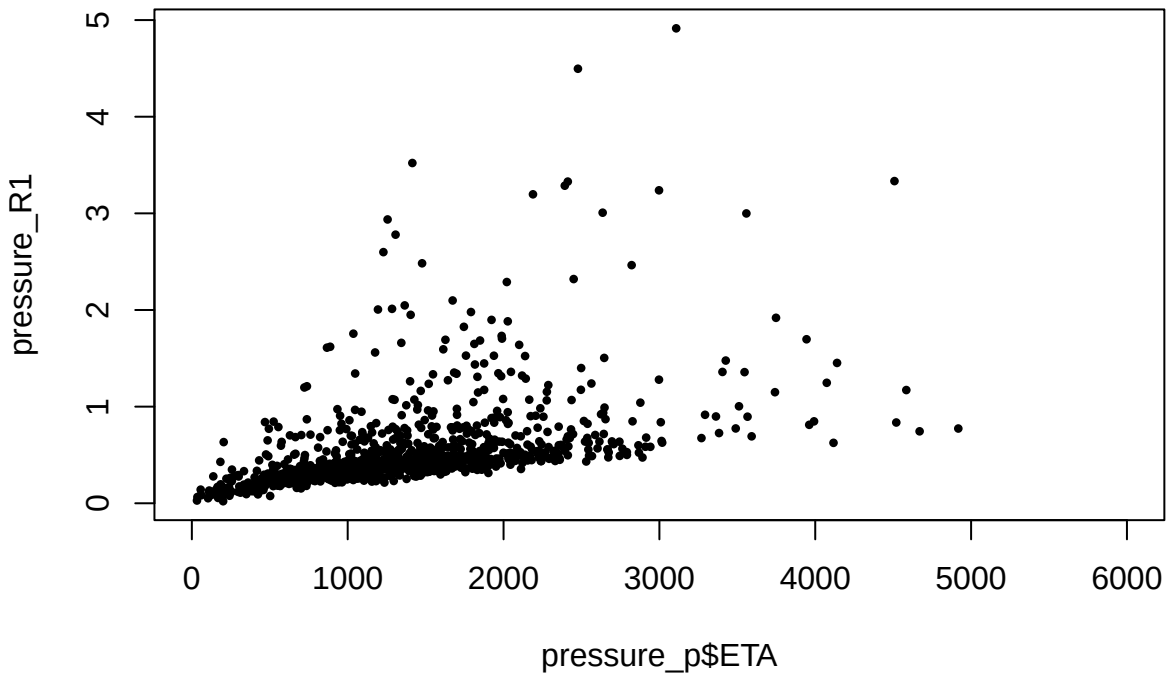

```
plot(real_price[1:1000],pressure_R1[1:1000],pch = 16,cex = 0.6)
```



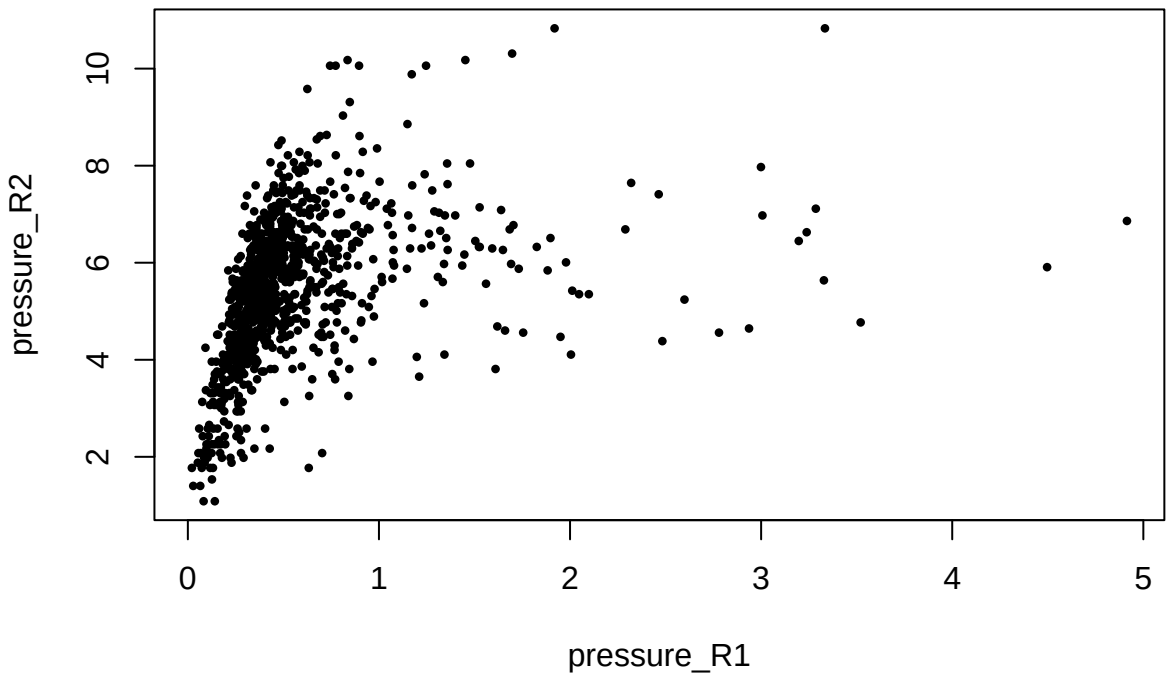

```
plot(pressure_test[[2]]$real_length,pressure_R1[1:1000],xlim = c(0, 6000),pch = 16,cex = 0.6)
```



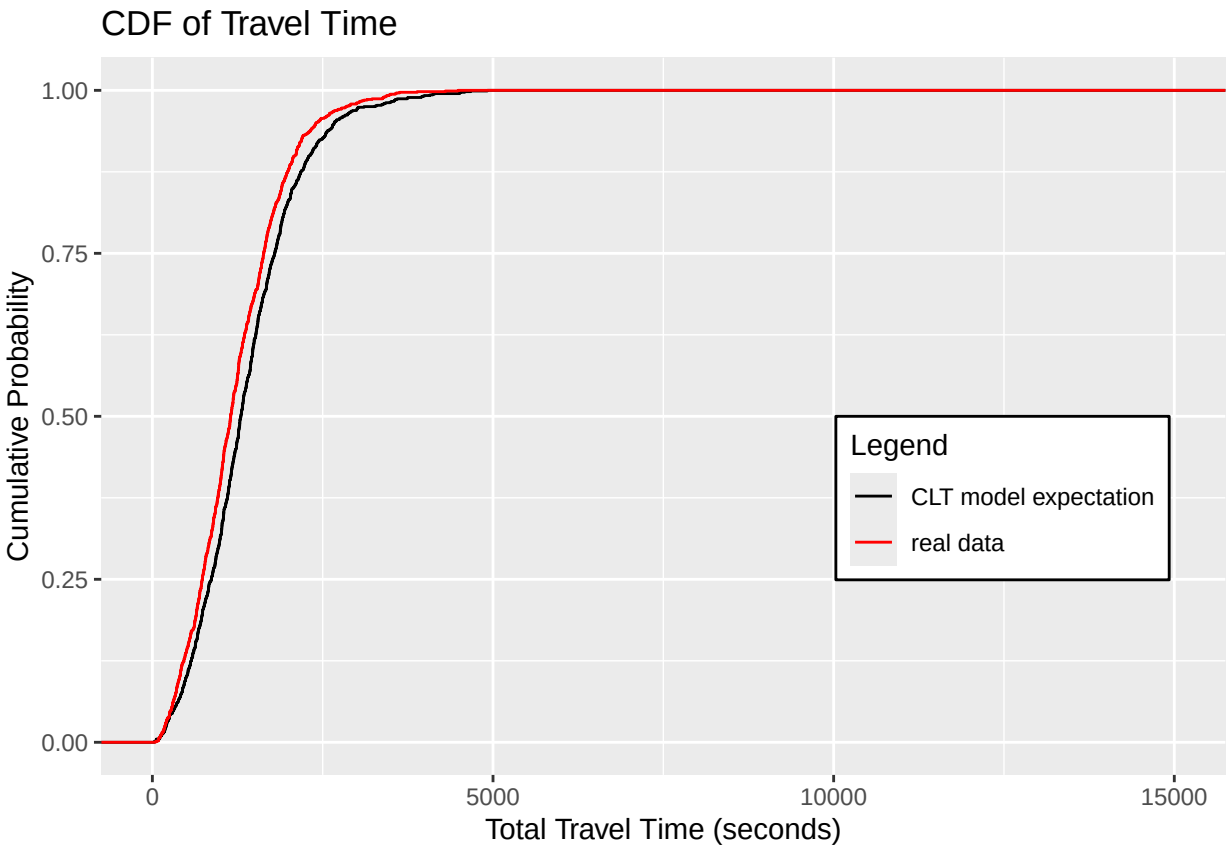

```
plot(pressure_p$ETA,pressure_R1,xlim = c(0, 6000),pch = 16,cex = 0.6)
```



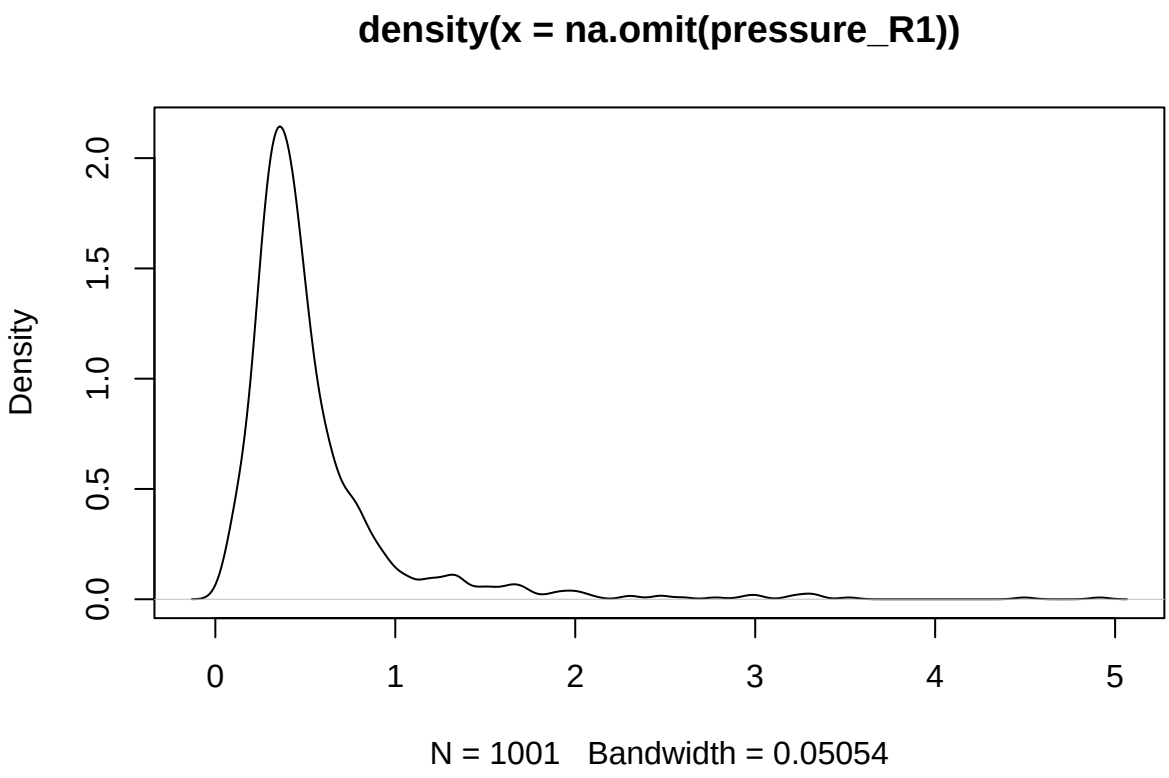

```
plot(pressure_R1,pressure_R2,pch = 16,cex = 0.6)
```




```
plot_CDF_compare(sample[[2]]$real_time,pressure_p$ETA,x_max = 15000,"CLT model expectation")
```




```
plot(density(na.omit(pressure_R1)))
```


density(x = na.omit(pressure_R1))
Density
0.0
0.5
1.0
1.5
2.0
0
1
2
3
4
5
N = 1001 Bandwidth = 0.05054