**Task1**

colnames(flowdata)<-c("src","dst","proto","valid","sport","dport","pkt","bytes","flows","start","end")

index<-sample((1:dim(flowdata)[1]),1000)

ggparcoord(flowdata[index,])

www<-flowdata[flowdata$sport == 80,]

ggparcoord(www)

ggplot(data=flowdata[index,]) + geom\_point(mapping = aes(x=bytes, y=pkt)) + scale\_y\_log10() + scale\_x\_log10()

ggplot(data=flowdata[index,]) + geom\_point(mapping = aes(x=bytes, y=pkt)) + scale\_y\_log10() + scale\_x\_log10() + geom\_point(data=www, colour='red',mapping=aes(x=bytes,y=pkt))

> ps<-vector("double",100000)

> for(i in seq(1,100000)){ps[i]<-(flowdata[i,8]/flowdata[i,7])}

> summary(ps)

Min. 1st Qu. Median Mean 3rd Qu. Max.

28.00 32.00 32.00 42.95 32.00 1500.00

> throughput<-flowdata$bytes/(flowdata$end-flowdata$start)

> view(throughput)

> INF = which(throughput==Inf)

> throughput0 = throughput[-INF]

> View(throughput0)

> summary(throughput0)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1 32 87 1128592 1272 198180864

m<-vector("double",100)

> for(i in seq(1,100)){m[i]<-mean(flowdata[sample(1:dim(flowdata)[1],100),8])}

> summary(m)

Min. 1st Qu. Median Mean 3rd Qu. Max.

33.86 51.51 108.86 1992.43 252.38 87315.90

**Task2**

hist(round(sampling[,]),col="blue", xlab = "inter-arrival times", main = "Histogram of inter-arrival times")

summary(sampling)

V1

Min. : 1.39

1st Qu.: 4731.55

Median : 11503.40

Mean : 16432.25

3rd Qu.: 22410.16

Max. :181717.71

> index<-sample((1:dim(sampling)[1]),5000)

> hist(round(sampling[index,]),col="blue", xlab = "inter-arrival times", main = "Histogram of inter-arrival times of 5000 samples")

> summary(sampling[index,])

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.89 4743.42 11395.47 16008.87 21781.81 147635.67

> m1<-vector("double",10000)

> for(i in seq(1,10000)){m1[i]<-mean(sampling[sample(1:dim(sampling)[1],5),])}

> hist(round(m1),col="blue", xlab = "inter-arrival times", main = "Histogram of mean inter-arrival times of 10000 ramdom samples of size(5)")

> summary(m1)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1019 11090 15294 16434 20644 57014

> qqplot(m1,sampling[,], ylab = "normal distribution", main="Q-Q plot of 10000 ramdom samples of size(5) against normal distribution", col="blue")

> sd(m1)

[1] 7333.584

> m2<-vector("double",10000)

> for(i in seq(1,10000)){m2[i]<-mean(sampling[sample(1:dim(sampling)[1],10),])}

> hist(round(m2),col="blue", xlab = "inter-arrival times", main = "Histogram of mean inter-arrival times of 10000 ramdom samples of size(10)")

> summary(m2)

Min. 1st Qu. Median Mean 3rd Qu. Max.

3628 12648 15890 16432 19669 41784

> qqplot(m2,sampling[,], ylab = "normal distribution", main="Q-Q plot of 10000 ramdom samples of size(10) against normal distribution", col="blue")> sd(m1)

> sd(m2)

[1] 5255.351

> m3<-vector("double",10000)

> for(i in seq(1,10000)){m3[i]<-mean(sampling[sample(1:dim(sampling)[1],100),])}

> hist(round(m3),col="blue", xlab = "inter-arrival times", main = "Histogram of mean inter-arrival times of 10000 ramdom samples of size(100)")

> summary(m3)

Min. 1st Qu. Median Mean 3rd Qu. Max.

10206 15242 16361 16416 17505 23624

> qqplot(m3,sampling[,], ylab = "normal distribution", main="Q-Q plot of 10000 ramdom samples of size(100) against normal distribution", col="blue")

> sd(m3)

[1] 1673.198

**Task3**

"norm"，"lnorm"，"pois"，"exp"， "gamma"，"nbinom"，"geom"，"beta"，"unif"和"logis"

> descdist(distr\_a[,1], discrete = FALSE, boot = NULL, method = "unbiased",

+ graph = TRUE, obs.col = "darkblue", obs.pch = 16, boot.col = "orange")

summary statistics

------

min: 8.334182e-13 max: 24.78845

median: 0.2364361

mean: 1.146677

estimated sd: 2.220498

estimated skewness: 3.89215

estimated kurtosis: 24.37036

> summary(distr\_a)

val

Min. : 0.00000

1st Qu.: 0.01804

Median : 0.23644

Mean : 1.14668

3rd Qu.: 1.24704

Max. :24.78845

> fit.gamma<-fitdist(distr\_a[,1],"gamma")

> plot(fit.gamma)

> fitdist(distr\_a[,1],"gamma")

Fitting of the distribution ' gamma ' by maximum likelihood

Parameters:

estimate Std. Error

shape 0.2759024 0.003917206

rate 0.2406022 0.006785372

> descdist(distr\_b[,1], discrete = FALSE, boot = NULL, method = "unbiased",

+ graph = TRUE, obs.col = "darkblue", obs.pch = 16, boot.col = "orange")

summary statistics

------

min: 0 max: 16

median: 6

mean: 6.052381

estimated sd: 2.543405

estimated skewness: 0.4604355

estimated kurtosis: 3.245417

> summary(distr\_b)

V1

Min. : 0.000

1st Qu.: 4.000

Median : 6.000

Mean : 6.052

3rd Qu.: 8.000

Max. :16.000

> fit.norm<-fitdist(distr\_b[,1],"norm")

> plot(fit.norm)

> fit.gamma<-fitdist(distr\_b[,1],"gamma","mme")

> plot(fit.gamma)

> fitdist(distr\_b[,1],"gamma", 'mme')

Fitting of the distribution ' gamma ' by matching moments

Parameters:

estimate

shape 5.6653714

rate 0.9360566

> fitdist(distr\_b[,1],"norm","mme")

Fitting of the distribution ' norm ' by matching moments

Parameters:

estimate

mean 6.052381

sd 2.542799

> fitdist(distr\_b[,1],"gamma","mme")

Fitting of the distribution ' gamma ' by matching moments

Parameters:

estimate

shape 5.6653714

rate 0.9360566

> descdist(distr\_c[,1], discrete = FALSE, boot = NULL, method = "unbiased",

+ graph = TRUE, obs.col = "darkblue", obs.pch = 16, boot.col = "orange")

summary statistics

------

min: 0.005968803 max: 3.199112e+12

median: 1498015

mean: 6785925289

estimated sd: 88664087008

estimated skewness: 23.37977

estimated kurtosis: 680.6618

> summary(distr\_c)

V1

Min. :0.000e+00

1st Qu.:4.222e+04

Median :1.498e+06

Mean :6.786e+09

3rd Qu.:4.362e+07

Max. :3.199e+12

> descdist(distr\_c[,1], discrete = FALSE, boot = NULL, method = "unbiased",

+ graph = TRUE, obs.col = "darkblue", obs.pch = 16, boot.col = "orange")

summary statistics

------

min: 0.005968803 max: 3.199112e+12

median: 1498015

mean: 6785925289

estimated sd: 88664087008

estimated skewness: 23.37977

estimated kurtosis: 680.6618

> fit.gamma<-fitdist(distr\_c[,1],"gamma","mme")

> plot(fit.gamma)

> fitdist(distr\_c[,1],"gamma", 'mme')

Fitting of the distribution ' gamma ' by matching moments

Parameters:

estimate

shape 5.859561e-03

rate 8.634874e-13

**Task4**

summary(flows)

packets bytes

Min. : 1.000 Min. : 32

1st Qu.: 1.000 1st Qu.: 40

Median : 1.000 Median : 40

Mean : 7.761 Mean : 6016

3rd Qu.: 1.000 3rd Qu.: 88

Max. :27037.000 Max. :39563775

plot(log10(flows[,1]), xlab = "flows", ylab = "number of packets") + title(main = "plot of packets in flows in logarithmic scale")

integer(0)

> plot(log10(flows[,2]), xlab = "flows", ylab = "number of bytes") + title(main = "plot of bytes in flows in logarithmic scale")

integer(0)

> m<-vector("double",20000)

> for(i in seq(1,20000)){m[i]<-mean(flows[1:i,2])}

> plot(m, xlab = "amount of flows passed", ylab = "mean values of bytes", type = "l", lwd = 2) + title(main = "plot of mean value of bytes as flows passed")

integer(0)

> md<-vector("double",20000)

> for(i in seq(1,20000)){md[i]<-median(flows[1:i,2])}

> plot(md, xlab = "amount of flows passed", ylab = "median values of bytes", type = "l", lwd = 2) + title(main = "plot of median value of bytes as flows passed")

integer(0)

> summary(md)

Min. 1st Qu. Median Mean 3rd Qu. Max.

40 40 40 40 40 44

> summary(m)

Min. 1st Qu. Median Mean 3rd Qu. Max.

40.8 7120.2 8691.4 13502.6 11672.0 247477.1