

Exam_Question_11

1.

```
setwd("/Users/wangka/Desktop")
exon <- read.delim("~/Desktop/question10_exon_nostrand.bed", header=FALSE)
exon$length <- NA
names(exon)[1] <- "chromosome"
names(exon)[2] <- "database"
names(exon)[3] <- "type"
names(exon)[4] <- "start"
names(exon)[5] <- "stop"
exon$length <- exon$stop - exon$start
exon.chr1 <-subset (exon, chromosome == 1)
exon.chr2 <-subset (exon, chromosome == 2)
exon.chr4 <-subset (exon, chromosome == 3)
exon.chr3 <-subset (exon, chromosome == 3)
exon.chr4 <-subset (exon, chromosome == 4)
exon.chr5 <-subset (exon, chromosome == 5)
exon.chr6 <-subset (exon, chromosome == 6)
exon.chr7 <-subset (exon, chromosome == 7)
exon.chr8 <-subset (exon, chromosome == 8)
exon.chr9 <-subset (exon, chromosome == 9)
exon.chr10 <-subset (exon, chromosome == 10)
exon.chr11 <-subset (exon, chromosome == 11)
exon.chr12 <-subset (exon, chromosome == 12)
exon.chr13 <-subset (exon, chromosome == 13)
exon.chr14 <-subset (exon, chromosome == 14)
exon.chr15 <-subset (exon, chromosome == 15)
exon.chr16 <-subset (exon, chromosome == 16)
exon.chr17 <-subset (exon, chromosome == 17)
exon.chr18 <-subset (exon, chromosome == 18)
exon.chr19 <-subset (exon, chromosome == 19)
exon.chr20 <-subset (exon, chromosome == 20)
exon.chr21 <-subset (exon, chromosome == 21)
exon.chr22 <-subset (exon, chromosome == 22)
exon.chrX <-subset (exon, chromosome == "X")
exon.chrY <-subset (exon, chromosome == "Y")
e1<-mean(exon.chr1$length)
e2<-mean(exon.chr2$length)
e3<-mean(exon.chr3$length)
e4<-mean(exon.chr4$length)
e5<-mean(exon.chr5$length)
e6<-mean(exon.chr6$length)
e7<-mean(exon.chr7$length)
e8<-mean(exon.chr8$length)
e9<-mean(exon.chr9$length)
e10<-mean(exon.chr10$length)
e11<-mean(exon.chr11$length)
e12<-mean(exon.chr12$length)
e13<-mean(exon.chr13$length)
e14<-mean(exon.chr14$length)
e15<-mean(exon.chr15$length)
```

```

e16<-mean(exon.chr16$length)
e17<-mean(exon.chr17$length)
e18<-mean(exon.chr18$length)
e19<-mean(exon.chr19$length)
e20<-mean(exon.chr20$length)
e21<-mean(exon.chr21$length)
e22<-mean(exon.chr22$length)
eX<-mean(exon.chrX$length)
eY<-mean(exon.chrY$length)
std <- function(x) sd(x)/sqrt(length(x))
ee1 <- std(exon.chr1$length)
ee2 <- std(exon.chr2$length)
ee3 <- std(exon.chr3$length)
ee4 <- std(exon.chr4$length)
ee5 <- std(exon.chr5$length)
ee6 <- std(exon.chr6$length)
ee7 <- std(exon.chr7$length)
ee8 <- std(exon.chr8$length)
ee9 <- std(exon.chr9$length)
ee10 <- std(exon.chr10$length)
ee11 <- std(exon.chr11$length)
ee12 <- std(exon.chr12$length)
ee13 <- std(exon.chr13$length)
ee14 <- std(exon.chr14$length)
ee15 <- std(exon.chr15$length)
ee16 <- std(exon.chr16$length)
ee17 <- std(exon.chr17$length)
ee18 <- std(exon.chr18$length)
ee19 <- std(exon.chr19$length)
ee20 <- std(exon.chr20$length)
ee21 <- std(exon.chr21$length)
ee22 <- std(exon.chr22$length)
eeX <- std(exon.chrX$length)
eeY <- std(exon.chrY$length)
exon_answer <- matrix (c(e1, e2, e3, e4, e5, e6, e7, e8, e9, e10, e11, e12, e13, e14, e15, e16, e17, e18, e19, e20, e21, e22, eX, eY), nrow=24, byrow=TRUE)

```

```

## Warning in matrix(c(e1, e2, e3, e4, e5, e6, e7, e8, e9, e10, e11, e12,
## e13, : data length [47] is not a sub-multiple or multiple of the number of
## rows [24]

```

```

View(exon_answer)
colnames(exon_answer) <- c('mean', 'SE')
View(exon_answer)
row.names(exon_answer) <- c('Chromosome 1', 'Chromosome 2', 'Chromosom3', 'Chromosome 4', 'Chromosome 5')
write.table(exon_answer, "/Users/wangka/question11_exon.txt", sep="\t")
print(exon_answer, type="html")

```

```

##           mean      SE
## Chromosome 1 171.806803 1.650216
## Chromosome 2 170.235053 1.585111
## Chromosom3   167.444280 2.076310
## Chromosome 4 180.710319 2.148415

```

```
## Chromosome 5 179.452784 1.756959
## Chromosome 6 175.642115 1.922756
## Chromosome 7 175.297981 2.188355
## Chromosome 8 172.356524 2.012749
## Chromosome 9 174.121962 1.719529
## Chromosome 10 172.484256 2.043632
## Chromosome 11 178.048514 1.506568
## Chromosome 12 166.453402 3.595995
## Chromosome 13 176.768957 2.670241
## Chromosome 14 177.292866 2.128271
## Chromosome 15 172.853020 1.868135
## Chromosome 16 172.862047 1.624757
## Chromosome 17 171.877360 2.454048
## Chromosome 18 174.178097 2.598789
## Chromosome 19 196.754024 2.233778
## Chromosome 20 168.643163 3.111300
## Chromosome 21 175.685289 2.924695
## Chromosome 22 193.392367 2.600181
## Chromosome X 191.061722 4.477689
## Chromosome Y 1.167457 171.806803
```

```
setwd("/Users/wangka/Desktop")
transcript <- read.delim("~/Desktop/question10_transcript_nostrand.bed", header=FALSE)
transcript$length <- NA
names(transcript)[1]<-"chromosome"
names(transcript)[2] <- "database"
names(transcript)[3] <- "type"
names(transcript)[4] <- "start"
names(transcript)[5] <- "stop"
transcript$length <- transcript$stop - transcript$start
transcript.chr1 <-subset (transcript, chromosome == 1)
transcript.chr2 <-subset (transcript, chromosome == 2)
transcript.chr4 <-subset (transcript, chromosome == 3)
transcript.chr3 <-subset (transcript, chromosome == 3)
transcript.chr4 <-subset (transcript, chromosome == 4)
transcript.chr5 <-subset (transcript, chromosome == 5)
transcript.chr6 <-subset (transcript, chromosome == 6)
transcript.chr7 <-subset (transcript, chromosome == 7)
transcript.chr8 <-subset (transcript, chromosome == 8)
transcript.chr9 <-subset (transcript, chromosome == 9)
transcript.chr10 <-subset (transcript, chromosome == 10)
transcript.chr11 <-subset (transcript, chromosome == 11)
transcript.chr12 <-subset (transcript, chromosome == 12)
transcript.chr13 <-subset (transcript, chromosome == 13)
transcript.chr14 <-subset (transcript, chromosome == 14)
transcript.chr15 <-subset (transcript, chromosome == 15)
transcript.chr16 <-subset (transcript, chromosome == 16)
transcript.chr17 <-subset (transcript, chromosome == 17)
transcript.chr18 <-subset (transcript, chromosome == 18)
transcript.chr19 <-subset (transcript, chromosome == 19)
transcript.chr20 <-subset (transcript, chromosome == 20)
transcript.chr21 <-subset (transcript, chromosome == 21)
transcript.chr22 <-subset (transcript, chromosome == 22)
transcript.chrX <-subset (transcript, chromosome == "X")
```

```

transcript.chrY <-subset (transcript, chromosome == "Y")
t1<-mean(transcript.chr1$length)
t2<-mean(transcript.chr2$length)
t3<-mean(transcript.chr3$length)
t4<-mean(transcript.chr4$length)
t5<-mean(transcript.chr5$length)
t6<-mean(transcript.chr6$length)
t7<-mean(transcript.chr7$length)
t8<-mean(transcript.chr8$length)
t9<-mean(transcript.chr9$length)
t10<-mean(transcript.chr10$length)
t11<-mean(transcript.chr11$length)
t12<-mean(transcript.chr12$length)
t13<-mean(transcript.chr13$length)
t14<-mean(transcript.chr14$length)
t15<-mean(transcript.chr15$length)
t16<-mean(transcript.chr16$length)
t17<-mean(transcript.chr17$length)
t18<-mean(transcript.chr18$length)
t19<-mean(transcript.chr19$length)
t20<-mean(transcript.chr20$length)
t21<-mean(transcript.chr21$length)
t22<-mean(transcript.chr22$length)
tX<-mean(transcript.chrX$length)
tY<-mean(transcript.chrY$length)
std <- function(x) sd(x)/sqrt(length(x))
tt1 <- std(transcript.chr1$length)
tt2 <- std(transcript.chr2$length)
tt3 <- std(transcript.chr3$length)
tt4 <- std(transcript.chr4$length)
tt5 <- std(transcript.chr5$length)
tt6 <- std(transcript.chr6$length)
tt7 <- std(transcript.chr7$length)
tt8 <- std(transcript.chr8$length)
tt9 <- std(transcript.chr9$length)
tt10 <- std(transcript.chr10$length)
tt11 <- std(transcript.chr11$length)
tt12 <- std(transcript.chr12$length)
tt13 <- std(transcript.chr13$length)
tt14 <- std(transcript.chr14$length)
tt15 <- std(transcript.chr15$length)
tt16 <- std(transcript.chr16$length)
tt17 <- std(transcript.chr17$length)
tt18 <- std(transcript.chr18$length)
tt19 <- std(transcript.chr19$length)
tt20 <- std(transcript.chr20$length)
tt21 <- std(transcript.chr21$length)
tt22 <- std(transcript.chr22$length)
ttX <- std(transcript.chrX$length)
ttY <- std(transcript.chrY$length)
transcript_answer <- matrix (c(t1, t2, t3, t4, t5, t6, t7, t8, t9, t10, t11, t12, t13, t14, t15, t16, t

```

```
## Warning in matrix(c(t1, t2, t3, t4, t5, t6, t7, t8, t9, t10, t11, t12,
```

```
## t13, : data length [47] is not a sub-multiple or multiple of the number of
## rows [24]
```

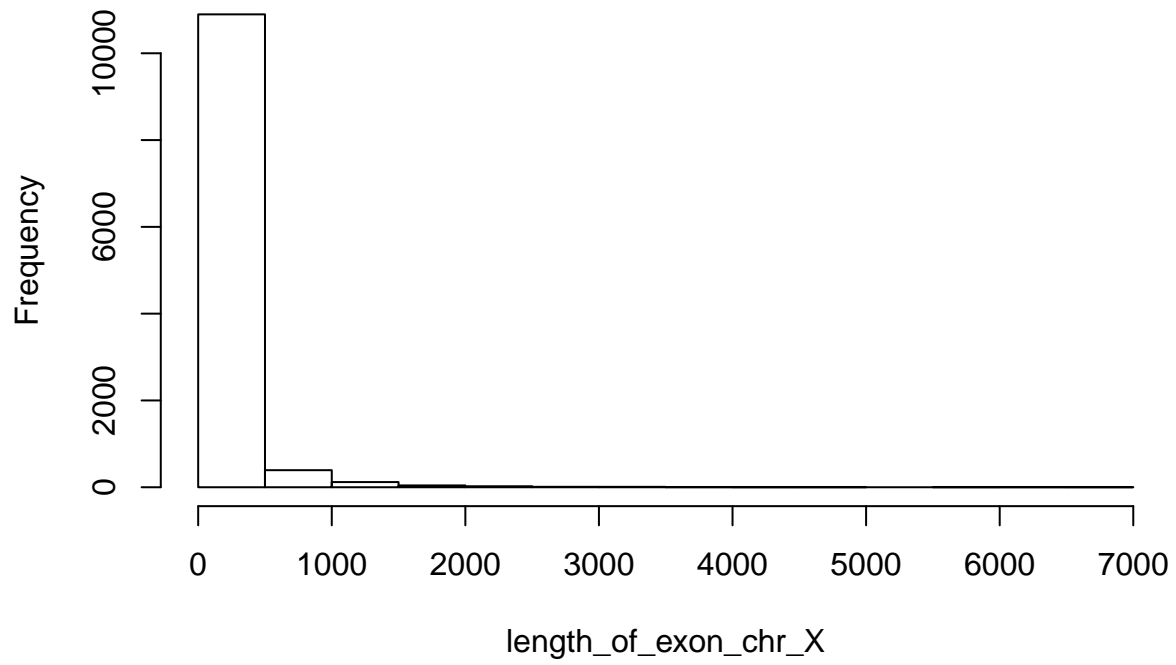
```
View(transcript_answer)
colnames(transcript_answer) <- c('mean', 'SE')
View(transcript_answer)
row.names(transcript_answer) <- c('Chromosome 1', 'Chromosome 2', 'Chromosom3', 'Chromosome 4', 'Chromo
write.table(transcript_answer, "/Users/wangka/question11_transcript.txt", sep="\t")
print(transcript_answer, type="html")
```

```
##           mean      SE
## Chromosome 1 38640.7541 1031.0711
## Chromosome 2 45415.9958 1299.0419
## Chromosom3   50669.6469 1482.6656
## Chromosome 4 51478.5129 1304.0569
## Chromosome 5 48089.7041 1220.2318
## Chromosome 6 43271.9013 1163.9141
## Chromosome 7 39159.9787 1317.8275
## Chromosome 8 44538.8063 1306.4864
## Chromosome 9 42599.1699 1110.7922
## Chromosome 10 37649.0076 1203.5705
## Chromosome 11 37019.3280 1335.9496
## Chromosome 12 44291.2077 1770.8841
## Chromosome 13 46892.9970 1707.7169
## Chromosome 14 43523.4184 1336.0761
## Chromosome 15 40051.7286  957.2974
## Chromosome 16 29383.6975  970.2432
## Chromosome 17 30397.3793 1690.4753
## Chromosome 18 44128.2194  697.7406
## Chromosome 19 22339.4047 1466.0287
## Chromosome 20 36547.0123 2187.2548
## Chromosome 21 39109.1294 1266.4161
## Chromosome 22 49383.0133 1804.1818
## Chromosome X 41082.2401 4011.7817
## Chromosome Y   850.2083 38640.7541
```

2.

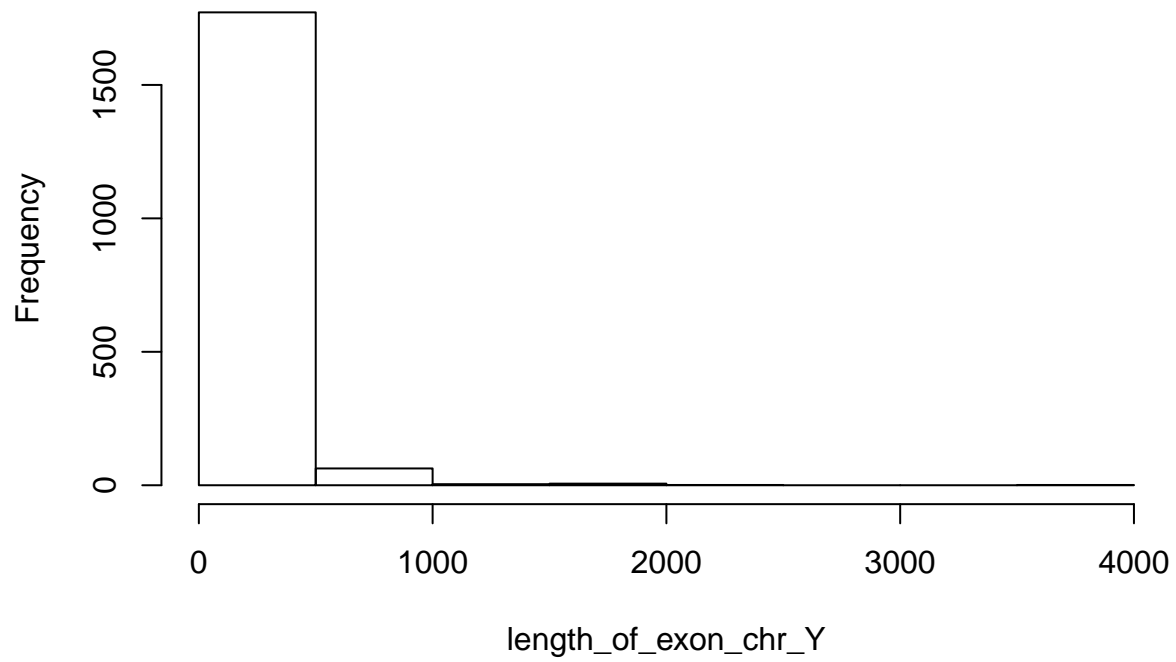
```
length_of_exon_chr_X = exon.chrX$length
hist(length_of_exon_chr_X)
```

Histogram of length_of_exon_chr_X



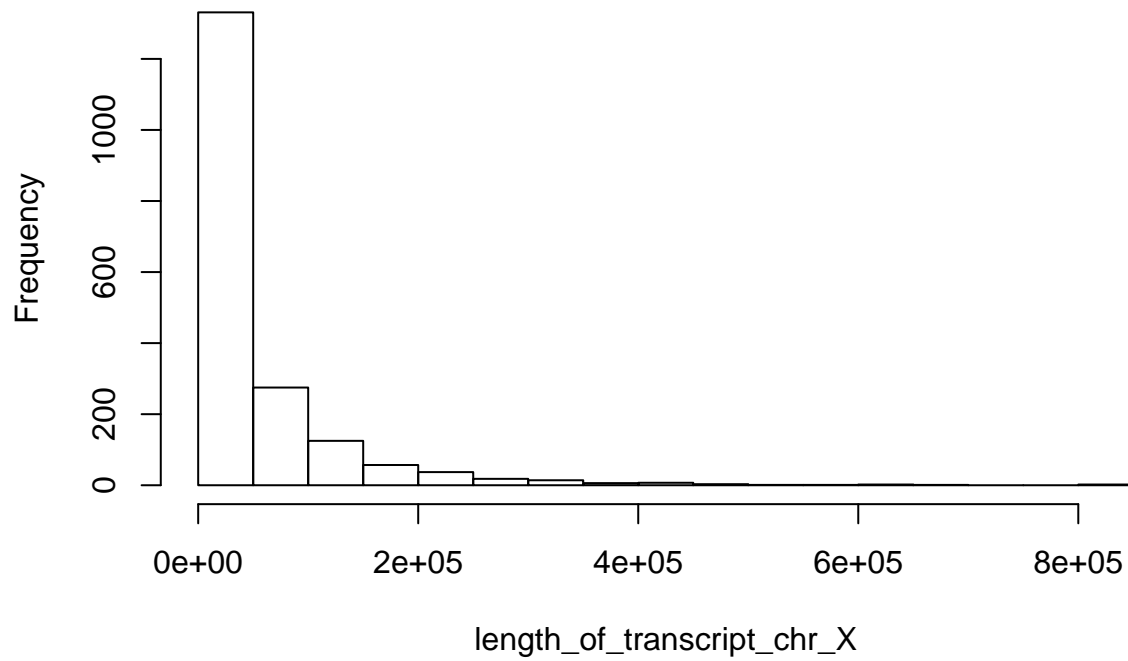
```
length_of_exon_chr_Y = exon.chrY$length  
hist(length_of_exon_chr_Y)
```

Histogram of length_of_exon_chr_Y



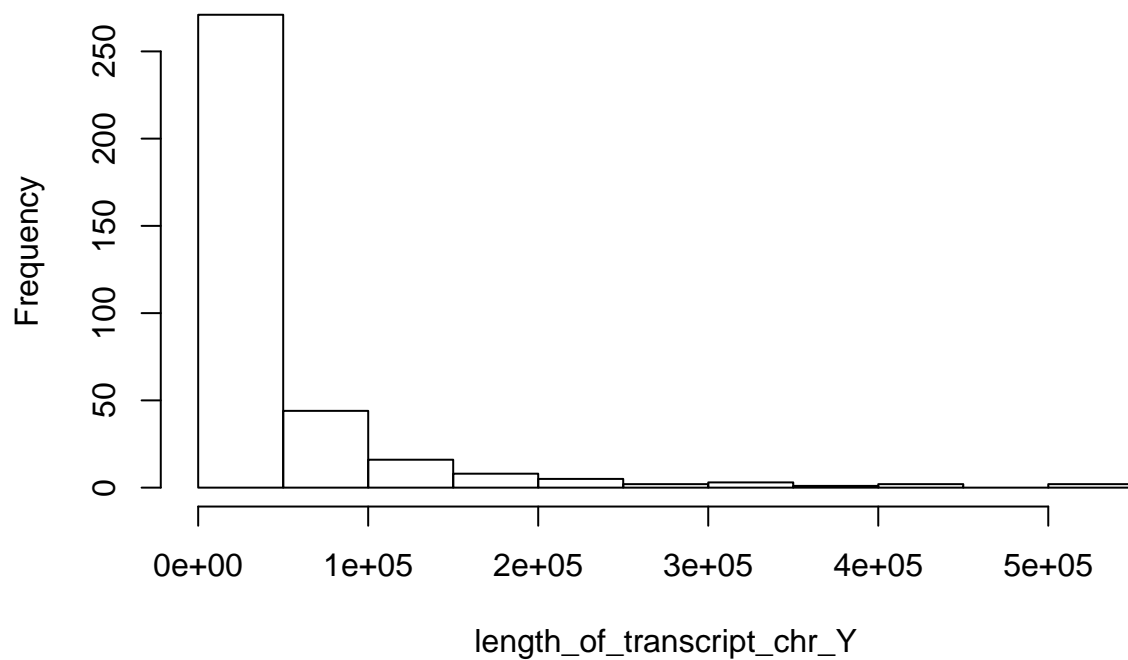
```
length_of_transcript_chr_X = transcript.chrX$length  
hist(length_of_transcript_chr_X)
```

Histogram of length_of_transcript_chr_X



```
length_of_transcript_chr_Y = transcript.chrY$length  
hist(length_of_transcript_chr_Y)
```

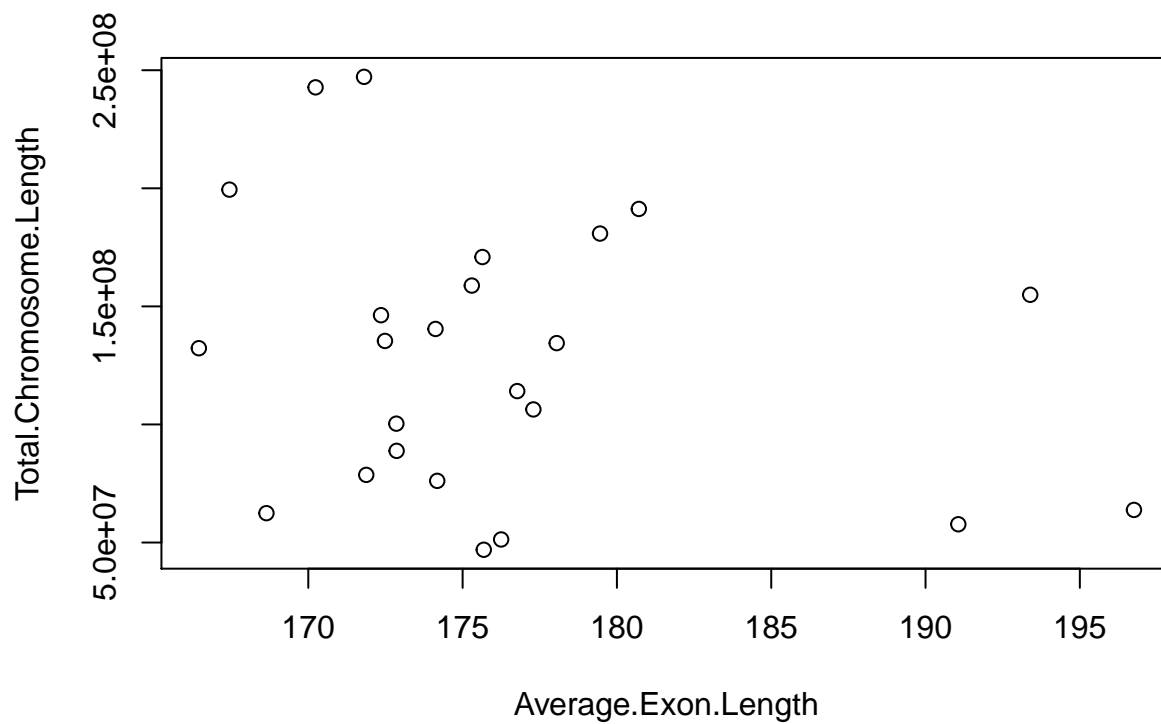
Histogram of length_of_transcript_chr_Y



3. No, there isn't a relationship between total chromosome length and average exon length across chromo-

somes.

```
exon_length_mean <- matrix(c(171.8068, 170.2351, 167.4443, 180.7103, 179.4528, 175.6421, 175.2980, 172.1),  
chromosome_length_mean <-matrix (c(247199719, 242751149, 199446827, 191263063, 180837866, 170896993, 15),  
Exon_Length_To_Chromosome_Length <- data.frame("Average Exon Length"=exon_length_mean, "Total Chromosome  
plot(Exon_Length_To_Chromosome_Length)
```



```
cor(exon_length_mean, chromosome_length_mean)
```

```
##           [,1]  
## [1,] -0.2400167
```