```
setwd("/Users/Eric/Desktop")
#Set my working directory to a location where I would have easy and fast access to my
files.

exon <- read.delim ("~/Desktop/10C_Exon.bed", header=FALSE)
#Uploaded my exon bed file to R.

exon$length <- NA
names(exon)[1] <- "chromosome"
names(exon)[2] <- "Database"
names(exon)[3] <- "Type"
names(exon)[4] <- "Start"
names(exon)[5] <- "Stop"
#Changing the column names within the exon data frame.

exon$length <- exon$Stop - exon$Start
#Creating a variable for exon length.

exon.chr1 <- subset(exon, chromosome == 1)
exon.chr2 <- subset(exon, chromosome == 2)
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")
#Subset every exon of every chromosome.

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)
```

```r
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)
#Calculated the mean length for the exons of every chromosome.


SE <- function(x) sd(x)/sqrt(length(x))
#Created a function for calculating standard error.

ee1 <- SE(exon.chr1$length)
ee2 <- SE(exon.chr2$length)
ee3 <- SE(exon.chr3$length)
ee4 <- SE(exon.chr4$length)
ee5 <- SE(exon.chr5$length)
ee6 <- SE(exon.chr6$length)
ee7 <- SE(exon.chr7$length)
ee8 <- SE(exon.chr8$length)
ee9 <- SE(exon.chr9$length)
ee10 <- SE(exon.chr10$length)
ee11 <- SE(exon.chr11$length)
ee12 <- SE(exon.chr12$length)
ee13 <- SE(exon.chr13$length)
ee14 <- SE(exon.chr14$length)
ee15 <- SE(exon.chr15$length)
ee16 <- SE(exon.chr16$length)
ee17 <- SE(exon.chr17$length)
ee18 <- SE(exon.chr18$length)
ee19 <- SE(exon.chr19$length)
ee20 <- SE(exon.chr2$length)
ee21 <- SE(exon.chr21$length)
ee22 <- SE(exon.chr22$length)
eeX <- SE(exon.chrX$length)
eeY <- SE(exon.chrY$length)
#Calculated the standard error for the exon lengths of every chromosome in the data f
rame.
```

```
Answer_Exon <- 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, ee1, ee2, ee3, ee4, ee5, ee6, ee7, ee
8, ee9, ee10, ee11, ee12, ee13, ee14, ee15, ee16, ee17, ee18, ee19, ee20, ee21, ee22,
eeX, eeY), ncol=2)
#Created a table with the mean and standard error data for the exons of every chromos
omes.
```

```
colnames(Answer_Exon) <- c('Mean Exon Length', 'SE for Exon Length')
#Changed the column names for the exon data frame.

row.names(Answer_Exon) <- c('Chromosome 1', 'Chromosome 2', 'Chromosome 3', 'Chromoso
me 4', 'Chromosome 5', 'Chromosome 6', 'Chromosome 7', 'Chromosome 8', 'Chromosome 9'
, 'Chromosome 10', 'Chromosome 11', 'Chromosome 12', 'Chromosome 13', 'Chromosome 14'
, 'Chromosome 15', 'Chromosome 16', 'Chromosome 17', 'Chromosome 18', 'Chromosome 19'
, 'Chromosome 20', 'Chromosome 21', 'Chromosome 22', 'Chromosome X', 'Chromosome Y')
#Changed the row names for the exon data frame.

write.table(Answer_Exon, "/Users/Eric/Desktop/11_Exon.txt", sep="\t")

print(Answer_Exon, type="html")
```

```
##               Mean Exon Length SE for Exon Length
## Chromosome 1          171.8068           1.167457
## Chromosome 2          170.2351           1.650216
## Chromosome 3          167.4443           1.585111
## Chromosome 4          180.7103           2.076310
## Chromosome 5          179.4528           2.148415
## Chromosome 6          175.6421           1.756959
## Chromosome 7          175.2980           1.922756
## Chromosome 8          172.3565           2.188355
## Chromosome 9          174.1220           2.012749
## Chromosome 10         172.4843           1.719529
## Chromosome 11         178.0485           2.043632
## Chromosome 12         166.4534           1.506568
## Chromosome 13         176.7690           3.595995
## Chromosome 14         177.2929           2.670241
## Chromosome 15         172.8530           2.128271
## Chromosome 16         172.8620           1.868135
## Chromosome 17         171.8774           1.624757
## Chromosome 18         174.1781           2.454048
## Chromosome 19         196.7540           2.598789
## Chromosome 20         168.6432           1.650216
## Chromosome 21         175.6853           3.111300
## Chromosome 22         176.2492           2.924695
## Chromosome X          193.3924           2.600181
## Chromosome Y          191.0617           4.477689
```

```
setwd("/Users/Eric/Desktop")
```

```r
#Set my working directory to a location where I would have easy and fast access to my
files.

transcript <- read.delim("~/Desktop/10C_Transcript.bed", header=FALSE)
#Uploaded the transcript bed file to R.

transcript$length <- NA
names(transcript)[1]<-"Chromosome"
names(transcript)[2] <- "Database"
names(transcript)[3] <- "Type"
names(transcript)[4] <- "Start"
names(transcript)[5] <- "Stop"
#Changed the names of the columns for the transcript data drame.

transcript$length <- transcript$Stop - transcript$Start
#Created a variable for transcript length.

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")
#Subset the transcripts for every chromosome.

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)
```

```r
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)
#Calculated the mean length for the transcripts of every chromosome.


SE <- function(x) sd(x)/sqrt(length(x))
#Created a function for standard error.

tt1 <- SE(transcript.chr1$length)
tt2 <- SE(transcript.chr2$length)
tt3 <- SE(transcript.chr3$length)
tt4 <- SE(transcript.chr4$length)
tt5 <- SE(transcript.chr5$length)
tt6 <- SE(transcript.chr6$length)
tt7 <- SE(transcript.chr7$length)
tt8 <- SE(transcript.chr8$length)
tt9 <- SE(transcript.chr9$length)
tt10<- SE(transcript.chr10$length)
tt11 <- SE(transcript.chr11$length)
tt12 <- SE(transcript.chr12$length)
tt13 <- SE(transcript.chr13$length)
tt14 <- SE(transcript.chr14$length)
tt15 <- SE(transcript.chr15$length)
tt16 <- SE(transcript.chr16$length)
tt17 <- SE(transcript.chr17$length)
tt18 <- SE(transcript.chr18$length)
tt19 <- SE(transcript.chr19$length)
tt20 <- SE(transcript.chr20$length)
tt21 <- SE(transcript.chr21$length)
tt22 <- SE(transcript.chr22$length)
ttX <- SE(transcript.chrX$length)
ttY <- SE(transcript.chrY$length)
#Calculated the standard error of the transcript length for every chromosome.

Answer_Transcript <- matrix(c(t1, t2, t3, t4, t5, t6, t7, t8, t9,t10, t11, t12, t13,
```

```
t14, t15, t16, t17, t18, t19, t20, t21, t22, tX, tY, tt1, tt2, tt3, tt4, tt5, tt6, tt
7, tt8, tt9, tt10, tt11, tt12, tt13, tt14, tt15, tt16, tt17, tt18, tt19, tt20, tt21,
tt22, ttX, ttY), ncol=2)
#Created a matrix for the mean and standard error for the transcript length for every
chromosome in the data frame.
```

```
colnames(Answer_Transcript) <- c('Mean Transcript Length', 'SE for Transcript Length'
)
#Changed the column names for the transcript data frame.

row.names(Answer_Transcript) <- c('Chromosome 1', 'Chromosome 2', 'Chromosom3', 'Chro
mosome 4', 'Chromosome 5', 'Chromosome 6', 'Chromosome 7', 'Chromosome 8', 'Chromosom
e 9', 'Chromosome 10', 'Chromosome 11', 'Chromosome 12', 'Chromosome 13', 'Chromosome
14', 'Chromosome 15', 'Chromosome 16', 'Chromosome 17', 'Chromosome 18', 'Chromosome
19', 'Chromosome 20', 'Chromosome 21', 'Chromosome 22', 'Chromosome X', 'Chromosome Y
')
#Changed the row names for the transcript data frame.

write.table(Answer_Transcript, "/Users/Eric/11_Transcript.txt", sep="\t")

print(Answer_Transcript, type="html")
```

```
##              Mean Transcript Length SE for Transcript Length
## Chromosome 1            38640.75                 850.2083
## Chromosome 2            45416.00                1031.0711
## Chromosom3              50669.65                1299.0419
## Chromosome 4            51478.51                1482.6656
## Chromosome 5            48089.70                1304.0569
## Chromosome 6            43271.90                1220.2318
## Chromosome 7            39159.98                1163.9141
## Chromosome 8            44538.81                1317.8275
## Chromosome 9            42599.17                1306.4864
## Chromosome 10           37649.01                1110.7922
## Chromosome 11           37019.33                1203.5705
## Chromosome 12           44291.21                1335.9496
## Chromosome 13           46893.00                1770.8841
## Chromosome 14           43523.42                1707.7169
## Chromosome 15           40051.73                1336.0761
## Chromosome 16           29383.70                 957.2974
## Chromosome 17           30397.38                 970.2432
## Chromosome 18           44128.22                1690.4753
## Chromosome 19           22339.40                 697.7406
## Chromosome 20           36547.01                1466.0287
## Chromosome 21           39109.13                2187.2548
## Chromosome 22           26710.05                1266.4161
## Chromosome X            49383.01                1804.1818
## Chromosome Y            41082.24                4011.7817
```
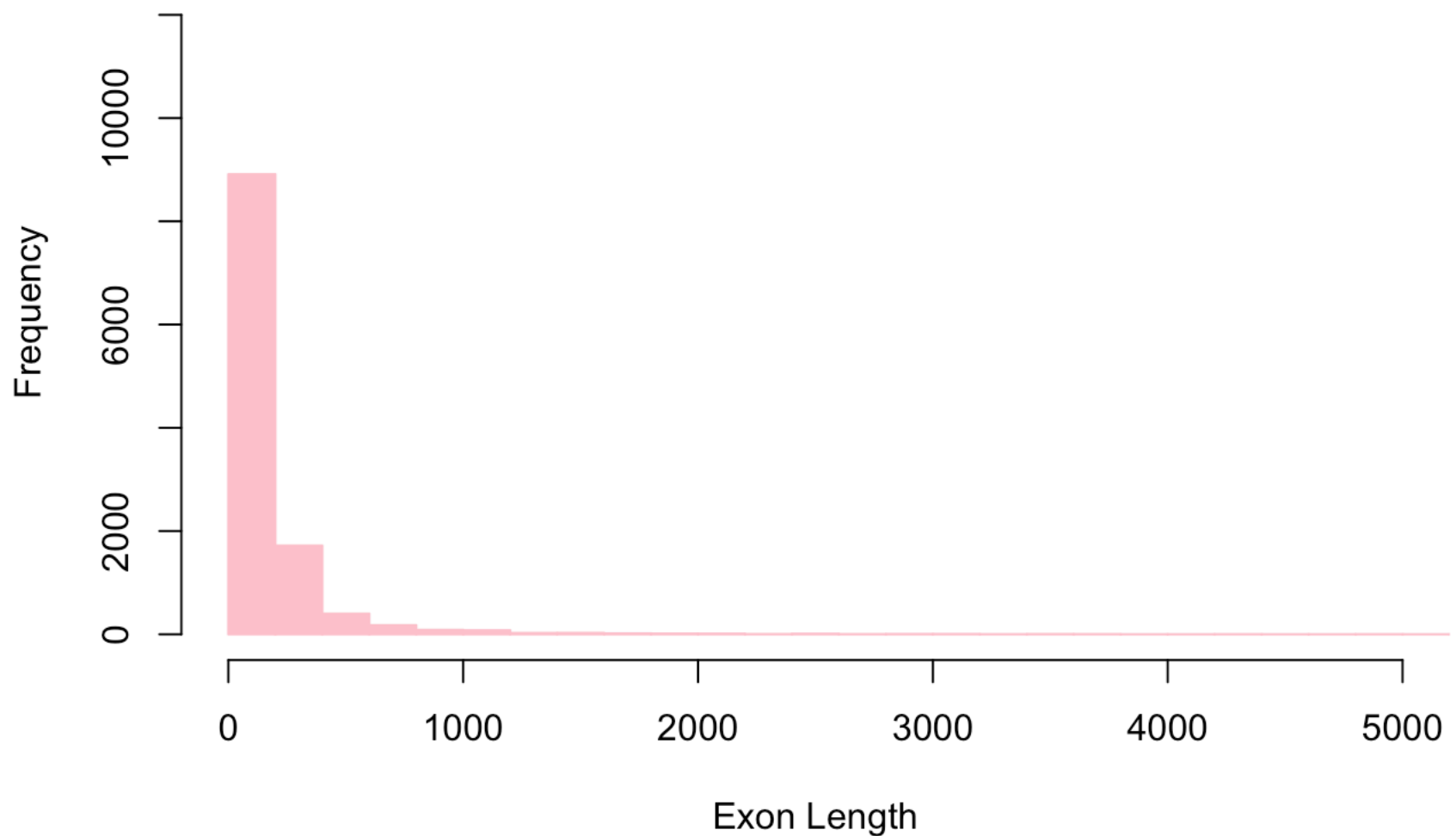
```
#Plotting the exon and transcript length for the X Chromosome.

Exon_length_of_chr_X = exon.chrX$length

hist(Exon_length_of_chr_X, xlab = "Exon Length", ylab = "Frequency", main = "Frequenc
y Histogram of Exon Lengths for the X Chromosome", xlim = c(0, 5000), ylim = c(0, 125
00), breaks=25, border="pink", col = "pink")
```


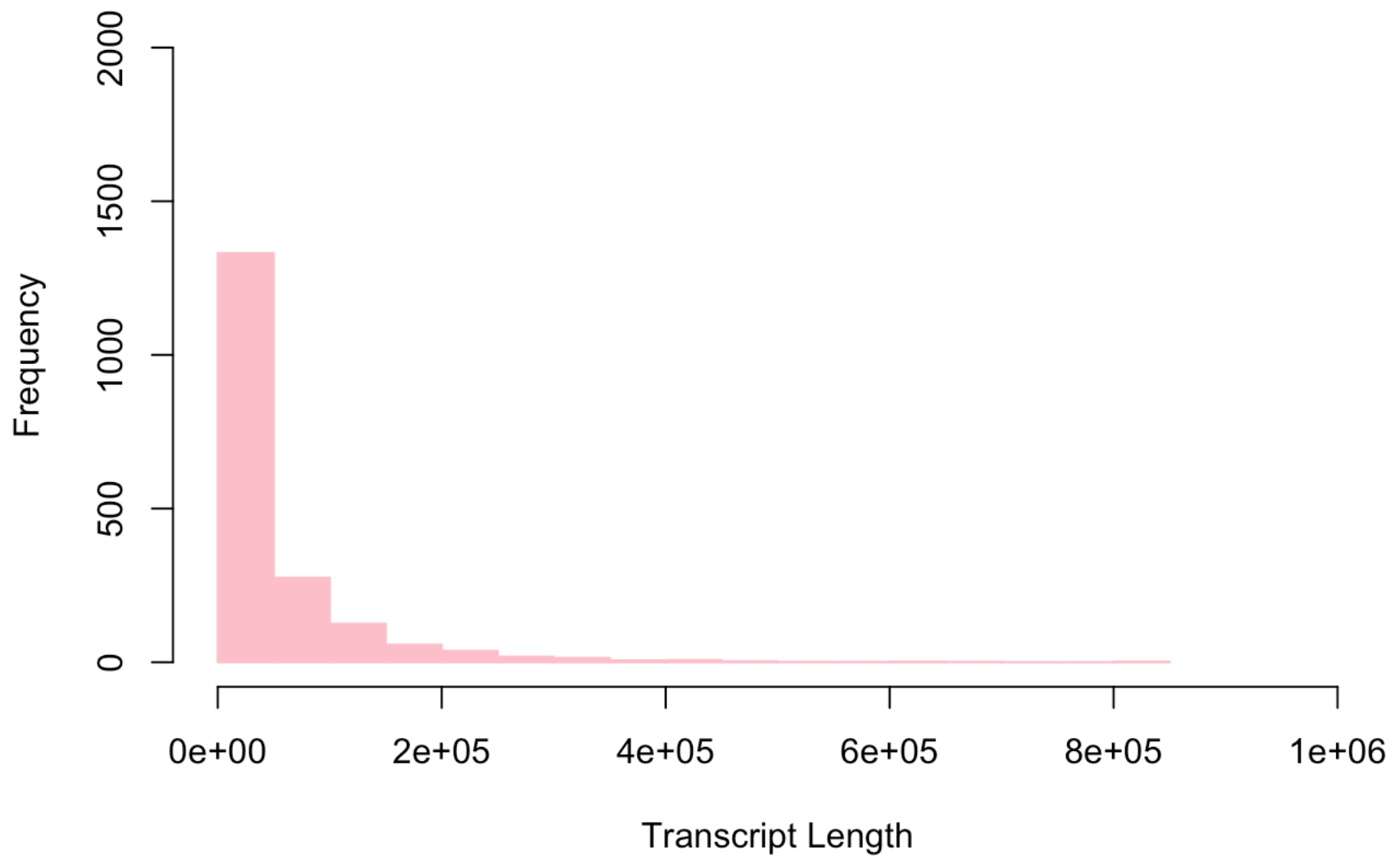Frequency Histogram of Exon Lengths for the X Chromosome

```
#Creating a histogram of exon length for the X chromosome.

Transcript_length_of_chr_X = transcript.chrX$length
hist(Transcript_length_of_chr_X, xlab = "Transcript Length", ylab = "Frequency", main
= "Frequency Histogram of Transcript Lengths for the X Chromosome", xlim = c(0, 10e+0
5), ylim = c(0, 2000), breaks=25, border="pink", col="pink")
```

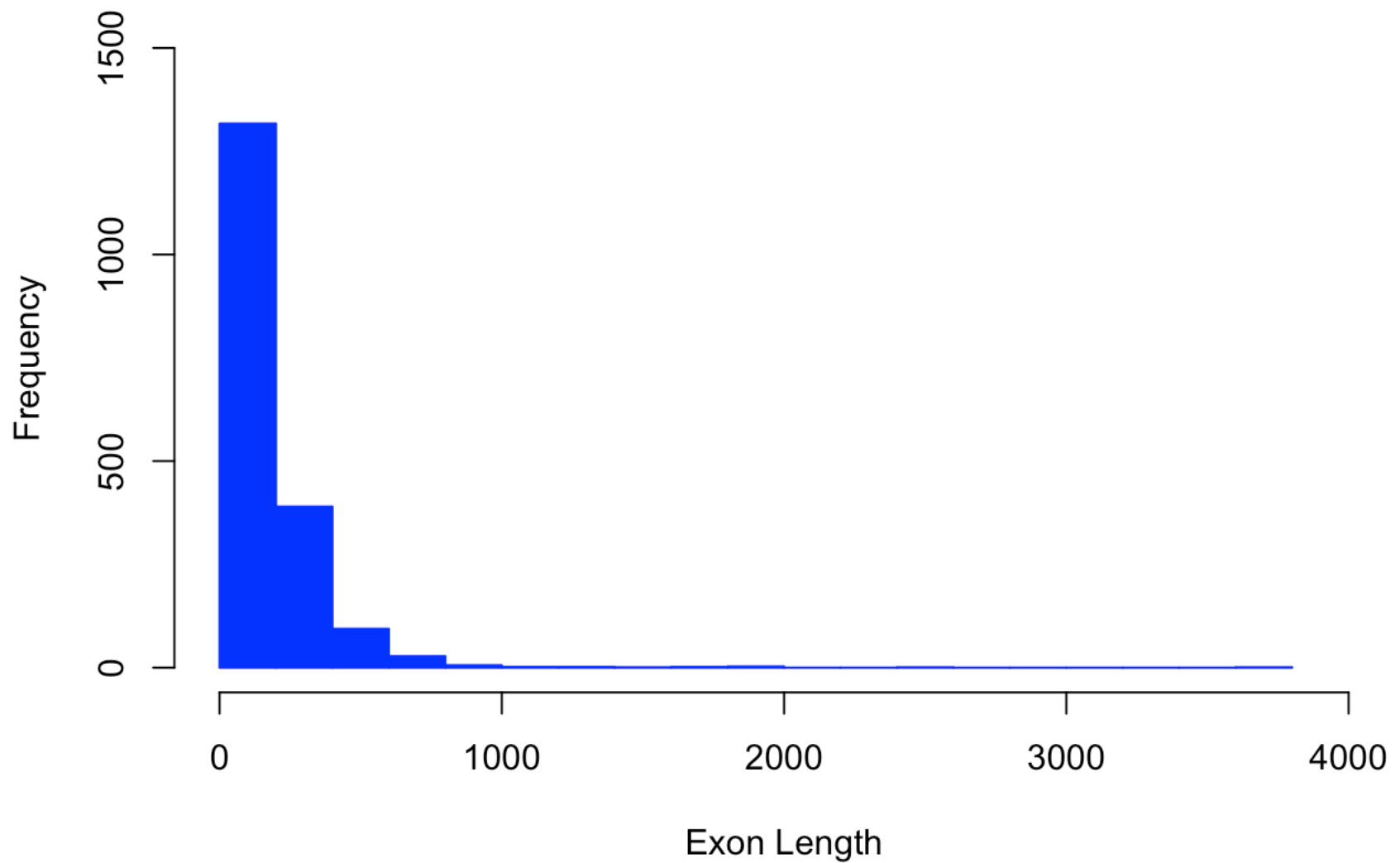# Frequency Histogram of Transcript Lengths for the X Chromosome


```
#Creating a histogram of transcript length for the X chromosome.

#Plotting the exon and transcript length for the Y Chromosome.

Exon_length_of_chr_Y = exon.chrY$length

hist(Exon_length_of_chr_Y, xlab = "Exon Length", ylab = "Frequency", main = "Frequenc
y Histogram of Exon Lengths for the Y Chromosome", xlim = c(0,4000), ylim = c(0,1500)
, breaks=25, border = "blue", col = "blue")
```

Frequency Histogram of Exon Lengths for the Y Chromosome


```
#Creating a histogram of exon length for the Y chromosome.

Transcript_length_of_chr_Y = transcript.chrY$length
hist(Transcript_length_of_chr_Y, xlab = "Transcript Length", ylab = "Frequency", main
= "Frequency Histogram of Transcript Lengths for the Y Chromosome", xlim = c(0,6e+05)
, ylim = c(0,250), breaks=25, border = "blue", col = "blue")
```

Frequency Histogram of Transcript Lengths for the Y Chromosome
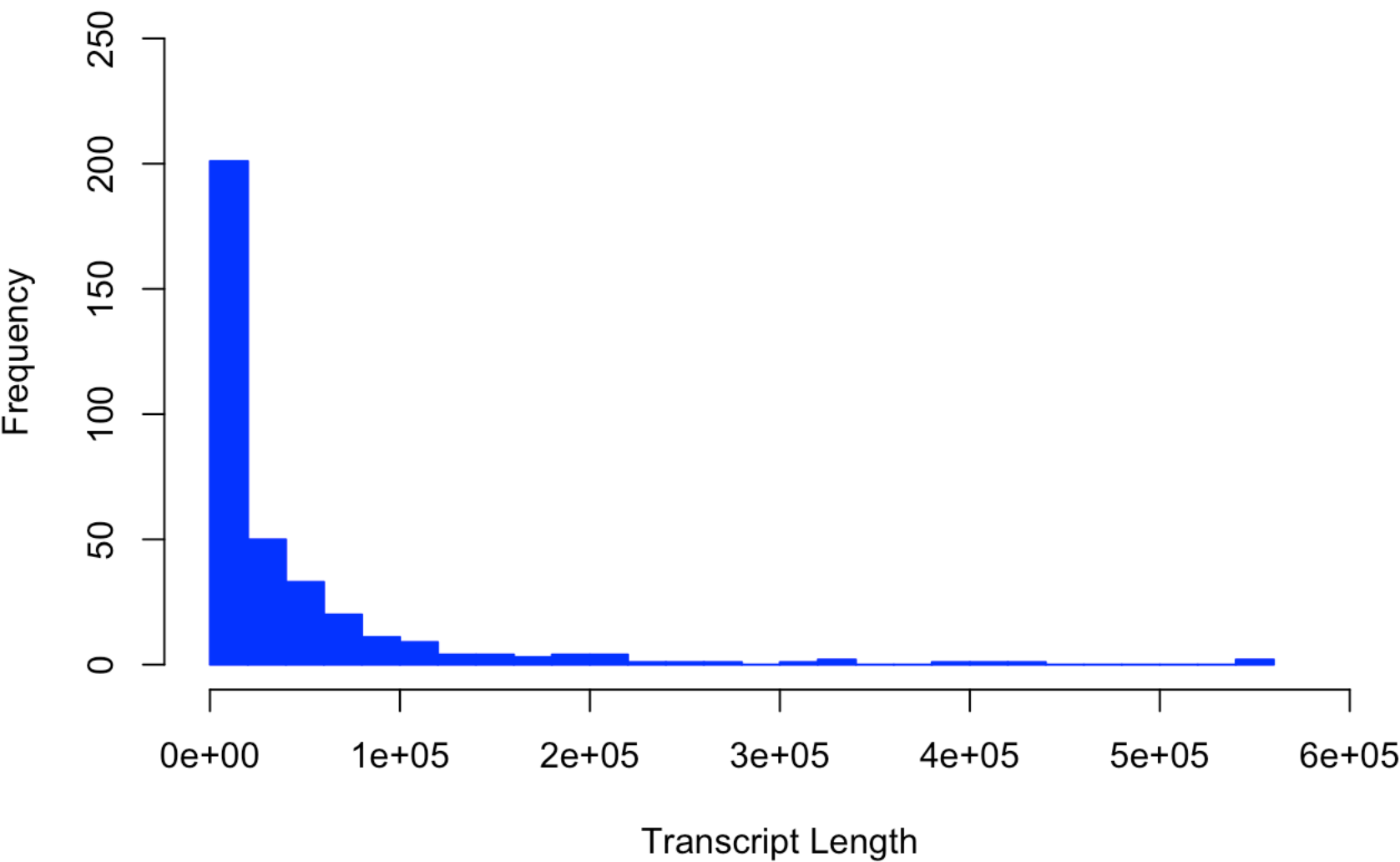
#Creating a histogram of transcript length for the Y chromosome.

```r
#Plotting Total Chromosome Length to Exon Mean Length Across the Chromosomes.

Mean_Exon_Length <- matrix(c(171.8068, 170.2351, 167.4443, 180.7103, 179.4528, 175.64
21, 175.2980, 172.3565, 174.1220, 172.4843, 178.0485, 166.4534, 176.7690, 177.2929, 1
72.8530, 172.8620, 171.8774, 174.1781, 196.7540, 168.6432, 175.6853, 176.2492, 193.39
24, 191.0617), ncol = 1)
#Matrix of the Mean Exon Length Data.

Mean_Total_Chromosome_Length <- matrix(c(247199719, 242751149, 199446827, 191263063,
180837866, 170896993, 158821424, 146274826, 140442298, 135374737, 134452384, 13228953
4, 114127980, 106360585, 100338915, 88822254, 78654742, 76117153, 63806651, 62435965,
46944323, 51304566, 154913754, 57741652), ncol=1)
#Matrix of the Mean Total Chromosome Length Data.

AverageExonLengthPerChromosome_TotalChromosomeLength <- data.frame("Average Exon Leng
th Per Chromosome" = Mean_Exon_Length, "Total Chromosome Length" = Mean_Total_Chromos
ome_Length)
#Created a data frame for the average exon length per chromosome compared to the tota
l chromsome length.

plot.default(AverageExonLengthPerChromosome_TotalChromosomeLength, type = "p", main =
"Average Exon Length Per Chromosome Compared To Total Chromosome Length", xlab = "Ave
rage Exon Length Per Chromosome", ylab = "Total Chromosome Length")
```
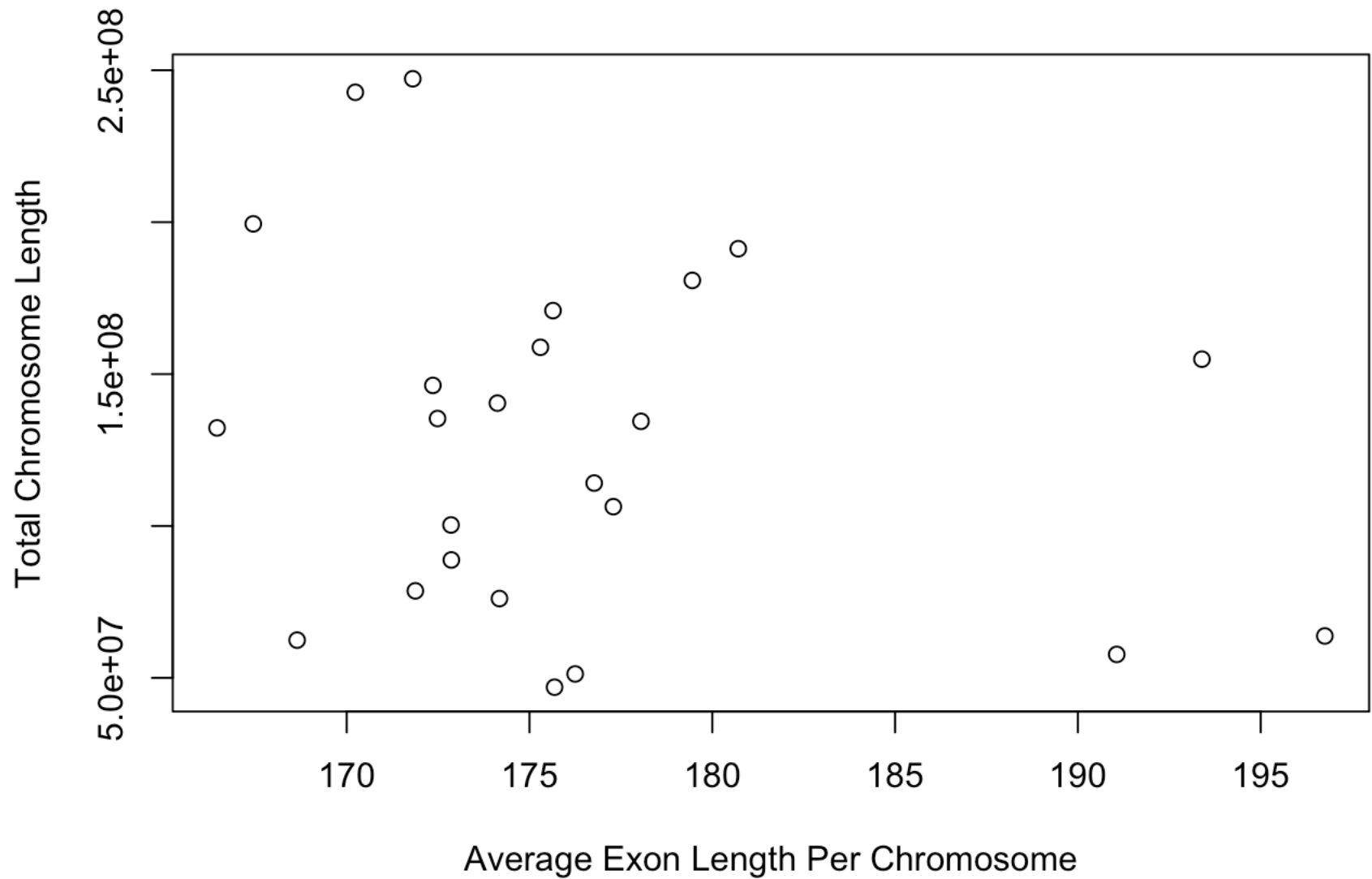
verage Exon Length Per Chromosome Compared To Total Chromosome L


```
#Plotted the average exon length per chromosome compared to the total chromosome leng
th.

cor(Mean_Exon_Length, Mean_Total_Chromosome_Length)
```

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

```
#Calculated the correlation coefficient between the two variables.

Linear_Model = lm(Mean_Total_Chromosome_Length~Mean_Exon_Length)
#Created a linear model for the mean length data.

summary(Linear_Model)
```

```
## 
## Call:
## lm(formula = Mean_Total_Chromosome_Length ~ Mean_Exon_Length)
## 
## Residuals:
##       Min        1Q    Median        3Q       Max
## -82613907 -44293556  -6738375  44658188 110565178
## 
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        450095532  277648877   1.621    0.119
## Mean_Exon_Length    -1824497    1573281  -1.160    0.259
## 
## Residual standard error: 57540000 on 22 degrees of freedom
## Multiple R-squared:  0.05761,    Adjusted R-squared:  0.01477
## F-statistic: 1.345 on 1 and 22 DF,  p-value: 0.2586
```

#Printed the summary statistics for the linear model of the mean length data.

##The statistical parameters, such as the correlation coefficient (-0.2400167) and the r-squared value (0.05761), indicate that there is not a definitive linear relationship between total chromosome length and average exon length across the chromosomes.

#Ensure reproducibility by documenting the software versions I used to create this analysis.
```
sessionInfo()
```

```
## R version 3.2.2 (2015-08-14)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: OS X 10.11.1 (El Capitan)
## 
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
## 
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
## 
## loaded via a namespace (and not attached):
##  [1] magrittr_1.5    tools_3.2.2     htmltools_0.2.6 yaml_2.1.13
##  [5] stringi_1.0-1   rmarkdown_0.8.1 knitr_1.11      stringr_1.0.0
##  [9] digest_0.6.8    evaluate_0.8
```