

Lab 2: Data Analysis - Part I

Armin Hatefi

Saturday, February 5, 2022

Contents

Read Data Sets	1
Line Graph	3
Bar Chart	4
Pie Chart	6
Histogram	8

This document shows you how to do basic statistical methods with RStudio.

Note: Throughout this lab, we type the commands in RStudio Editor and then run the command (i.e. send it to the console).

Read Data Sets

```
# Read data from txt

ex11Data <- read.table(file="ex11-09.txt", header=T, sep="\t")
ex11Data
```

```
##      State Percentage
## 1      AL      22.71557
## 2      AK      25.94752
## 3      AZ      23.61538
## 4      AR      22.62697
## 5      CA      24.78167
## 6      CO      24.51913
## 7      CT      20.85119
## 8      DE      22.10767
## 9      FL      21.28983
## 10     GA      23.79723
## 11     HI      23.83540
## 12     ID      23.68766
## 13     IL      24.01953
## 14     IN      22.89478
## 15     IA      22.47752
```

```
## 16    KS    23.59029
## 17    KY    22.67510
## 18    LA    24.14418
## 19    ME    19.90881
## 20    MD    22.61271
## 21    MA    23.05325
## 22    MI    22.18334
## 23    MN    23.04598
## 24    MS    23.51140
## 25    MO    22.71651
## 26    MT    22.23371
## 27    NE    23.49972
## 28    NV    23.07692
## 29    NH    20.51672
## 30    NJ    21.14477
## 31    NM    23.73992
## 32    NY    23.39148
## 33    NC    22.96682
## 34    ND    25.58502
## 35    OH    22.15741
## 36    OK    23.83306
## 37    OR    23.00792
## 38    PA    21.44120
## 39    RI    23.31113
## 40    SC    22.76786
## 41    SD    22.88557
## 42    TN    22.46179
## 43    TX    24.79961
## 44    UT    28.36257
## 45    VT    21.41707
## 46    VA    23.58090
## 47    WA    23.45396
## 48    WV    21.44432
## 49    WI    22.61905
## 50    WY    23.63977
```

#Reading Data from csv

```
ex14Data <- read.csv(file="eg14-01.csv",header = T, sep=",")
ex14Data
```

```
##      Distance Velocity
## 1      0.032      170
## 2      0.034      290
## 3      0.214     -130
## 4      0.263     -70
## 5      0.275    -185
## 6      0.275    -220
## 7      0.450     200
## 8      0.500     290
## 9      0.500     270
## 10     0.630     200
## 11     0.800     300
## 12     0.900     -30
## 13     0.900     650
## 14     0.900     150
```

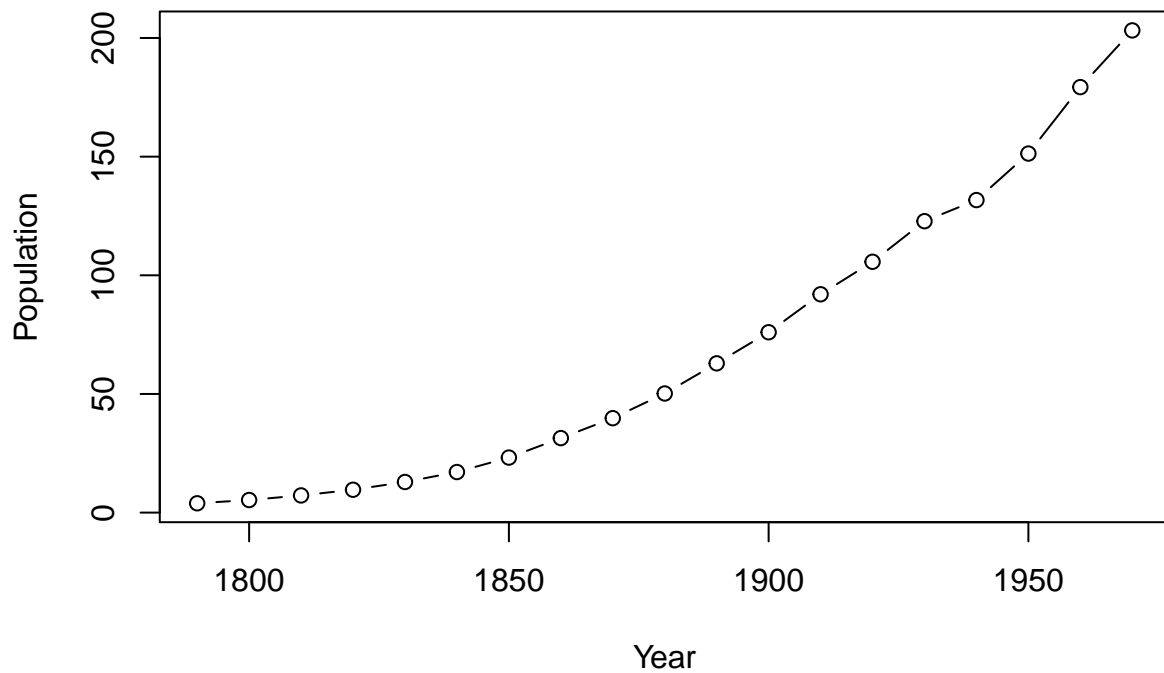
```
## 15    0.900    500
## 16    1.000    920
## 17    1.100    450
## 18    1.100    500
## 19    1.400    500
## 20    1.700    960
## 21    2.000    500
## 22    2.000    850
## 23    2.000    800
## 24    2.000   1090
```

Line Graph

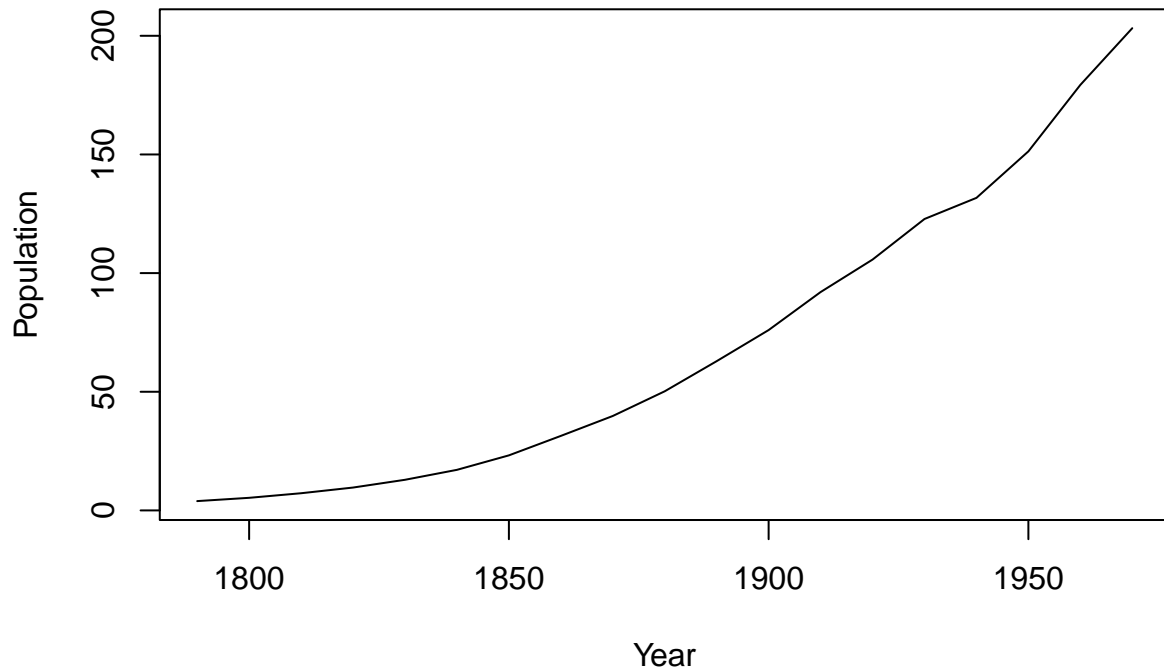
```
USdata <- read.table(file="UScensus.txt", sep="\t", header=T)
head(USdata)
```

```
##   Year Population
## 1 1790      3.93
## 2 1800      5.31
## 3 1810      7.24
## 4 1820      9.64
## 5 1830     12.90
## 6 1840     17.10
```

```
attach(USdata)
plot(Year, Population, type="b")
```



```
plot(Year,Population, type="l")
```



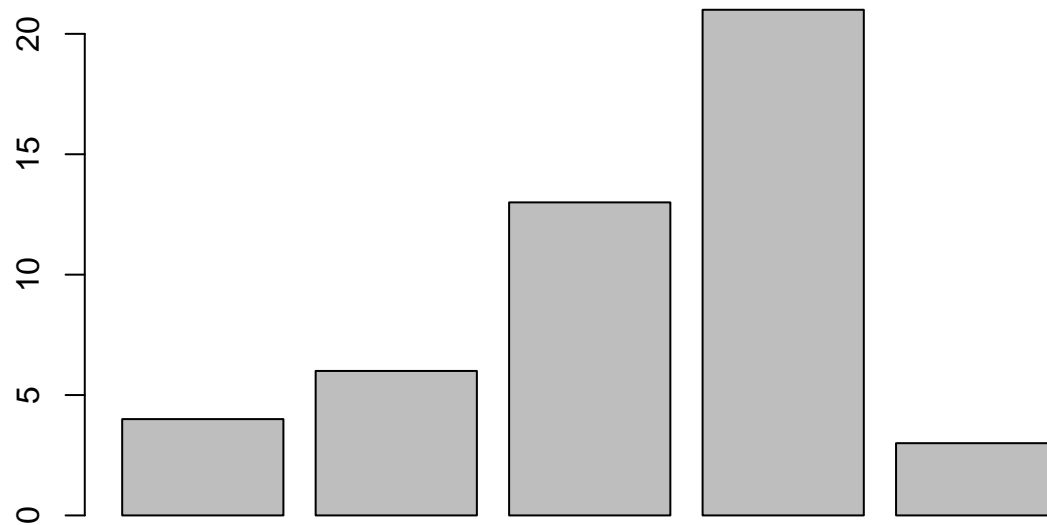
```
detach(USdata)
```

Bar Chart

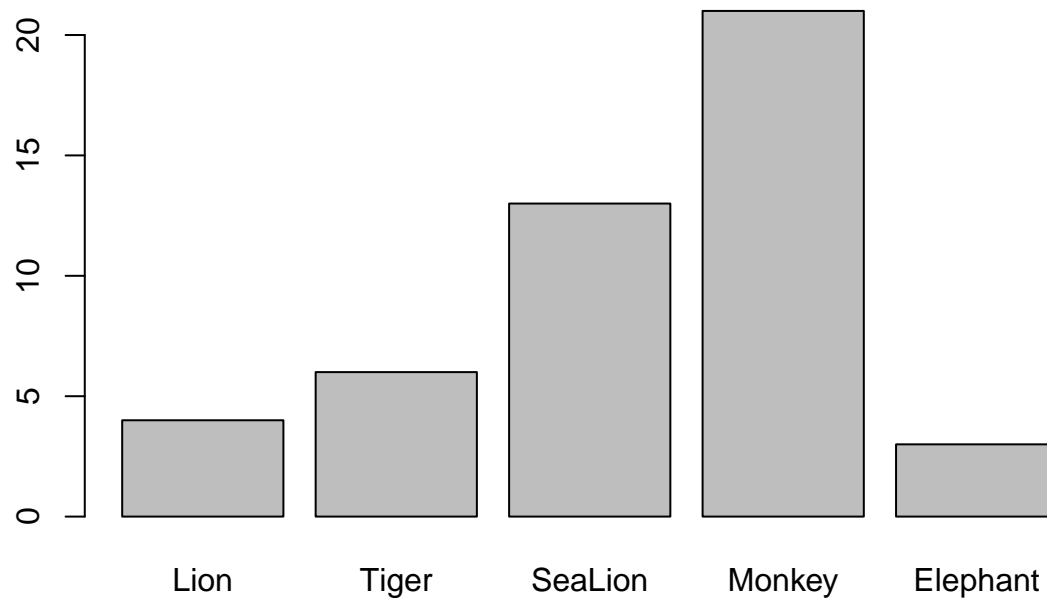
```
Zoo <- read.table(file="Zoo.txt",sep=" ",header=T)
head(Zoo)
```

```
##      Animal Count
## 1      Lion      4
## 2      Tiger      6
## 3    SeaLion     13
## 4     Monkey     21
## 5  Elephant      3
```

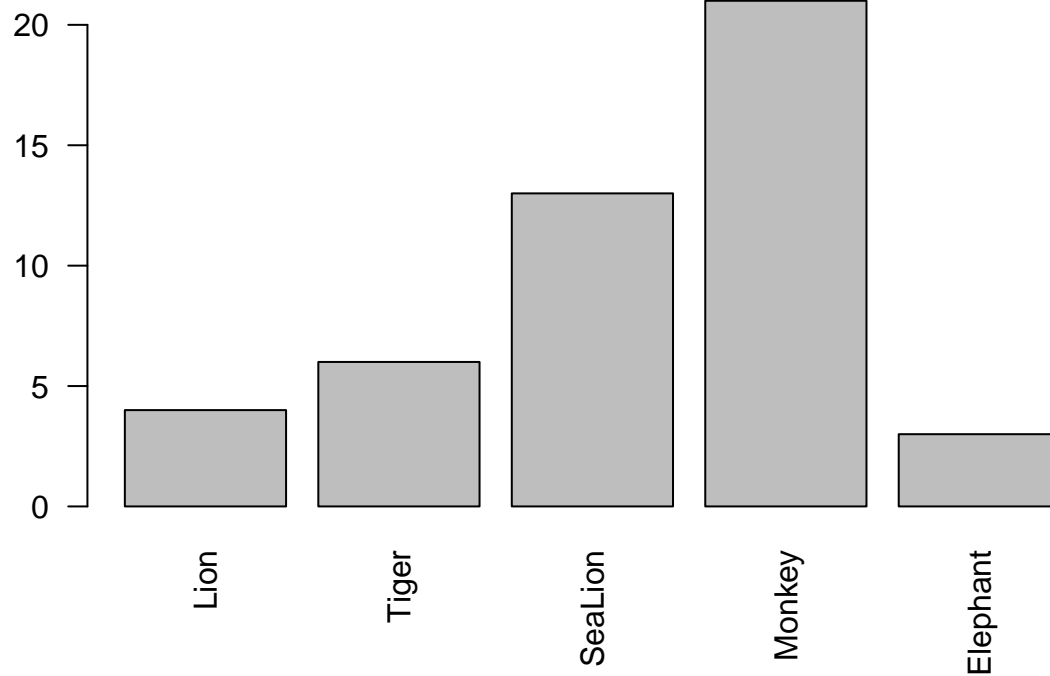
```
attach(Zoo)
barplot(Count)
```



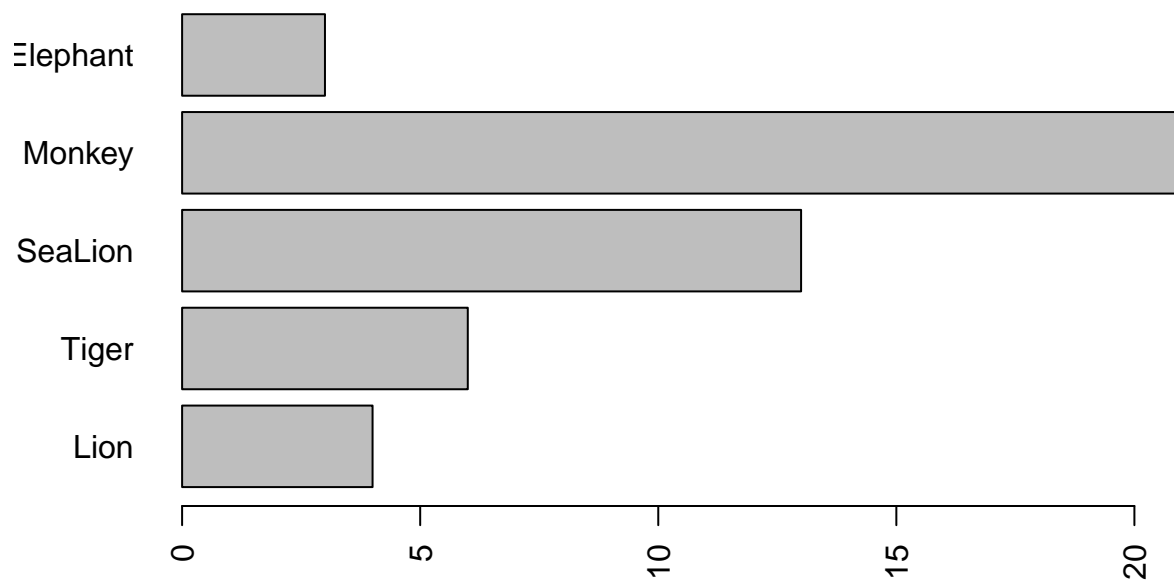
```
barplot(Count,names=Animal)
```



```
barplot(Count,names=Animal,las=2)
```



```
barplot(Count,names=Animal,las=2, horiz = TRUE)
```



```
detach(Zoo)
```

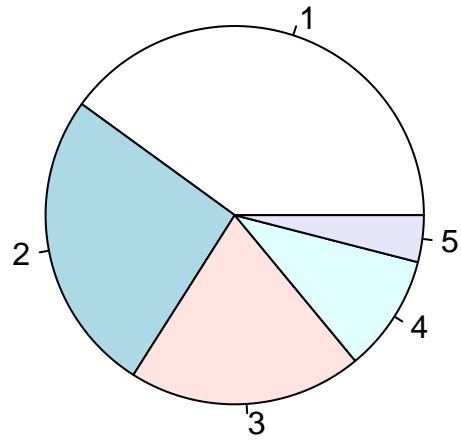
Pie Chart

```
myData <- read.csv(file="Exam1.csv",sep=" ",header=T)
head(myData)
```

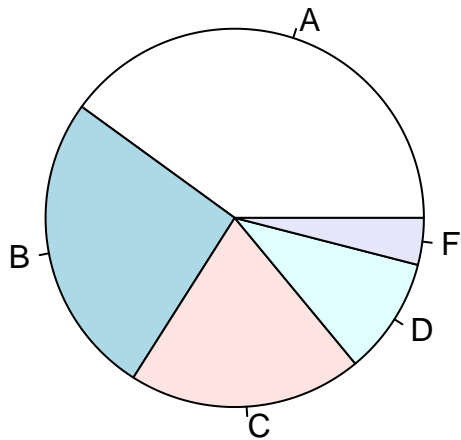
```
##   Grade Count
## 1      A    20
```

```
## 2    B    13
## 3    C    10
## 4    D     5
## 5    F     2
```

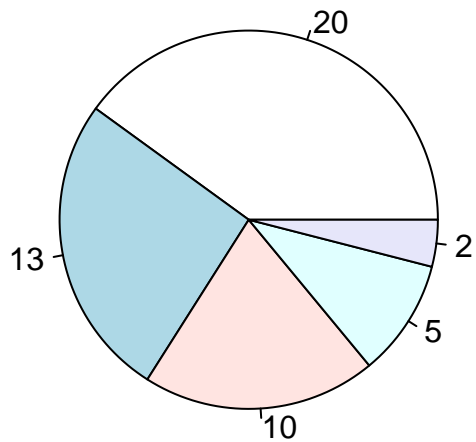
```
pie(myData$Count)
```



```
attach(myData)
pie(Count, labels=Grade)
```



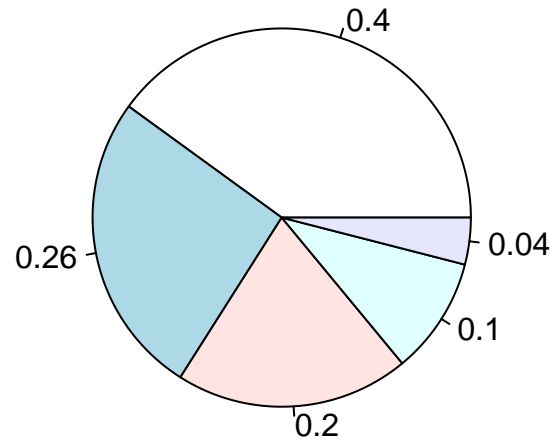
```
pie(Count, labels=Count)
```



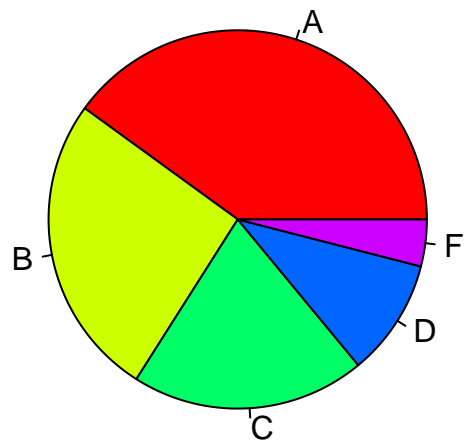
```
mylabel <- Count/sum(Count)
mylabel
```

```
## [1] 0.40 0.26 0.20 0.10 0.04
```

```
pie(Count, labels=mylabel)
```



```
pie(Count, col=rainbow(5), labels=Grade)
```

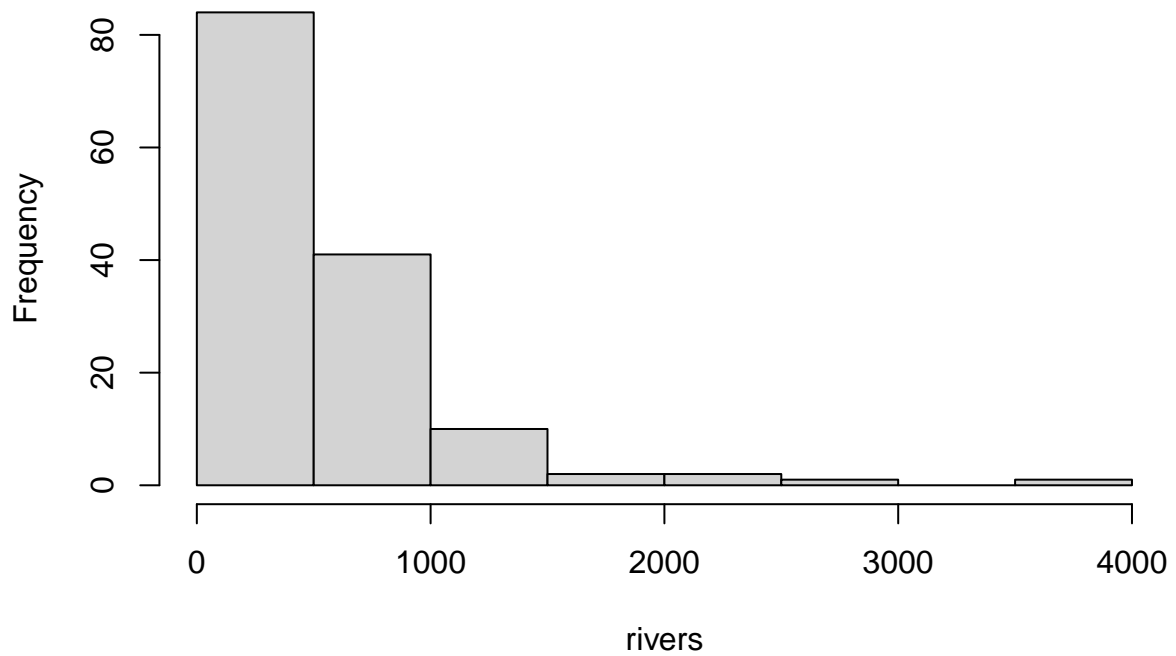


```
detach(myData)
```

Histogram

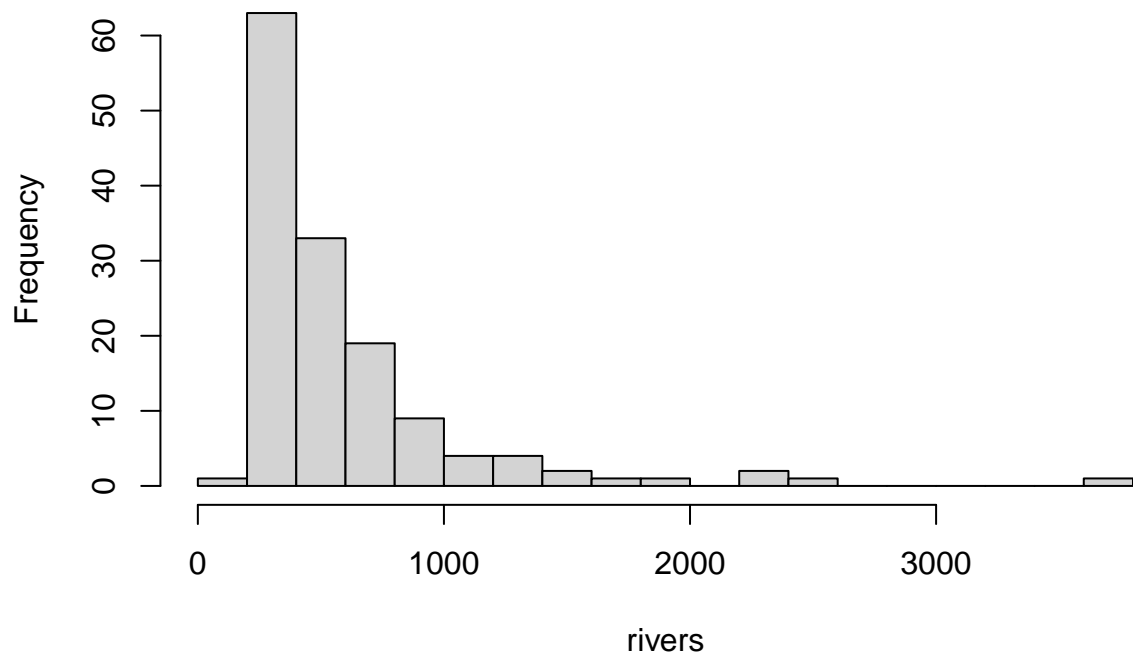
```
hist(rivers)
```


Histogram of rivers



```
hist(rivers,breaks=21)
```

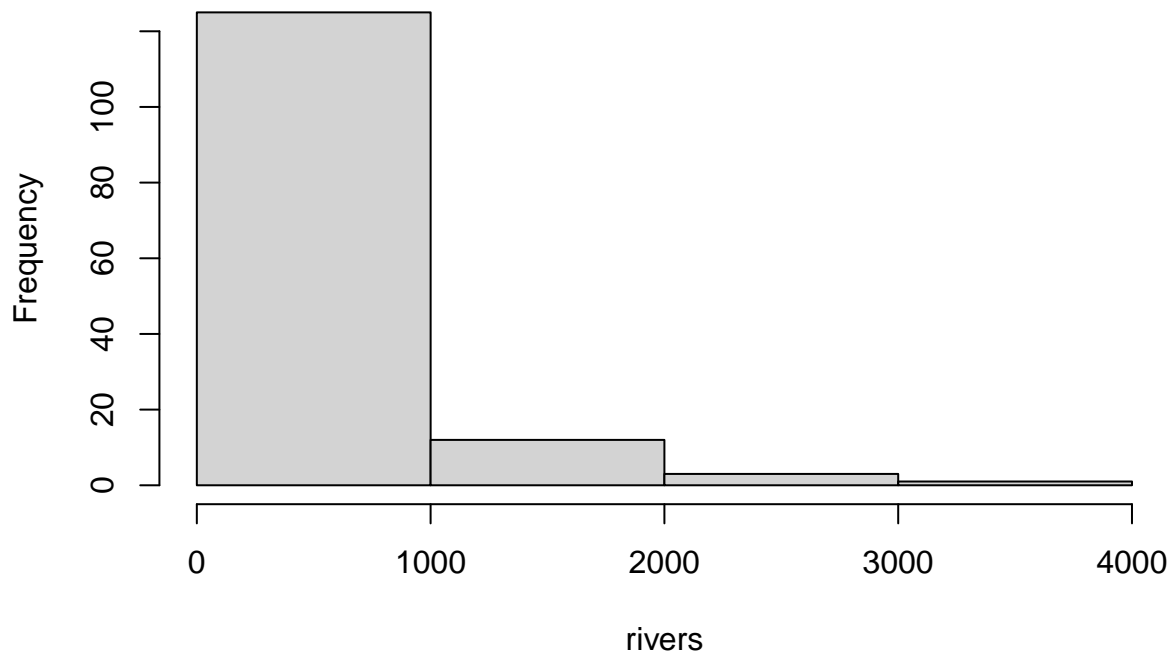
Histogram of rivers



```
bvector <- c(0,1000,2000,3000,4000)
```

```
hist(rivers,breaks=bvector)
```

Histogram of rivers

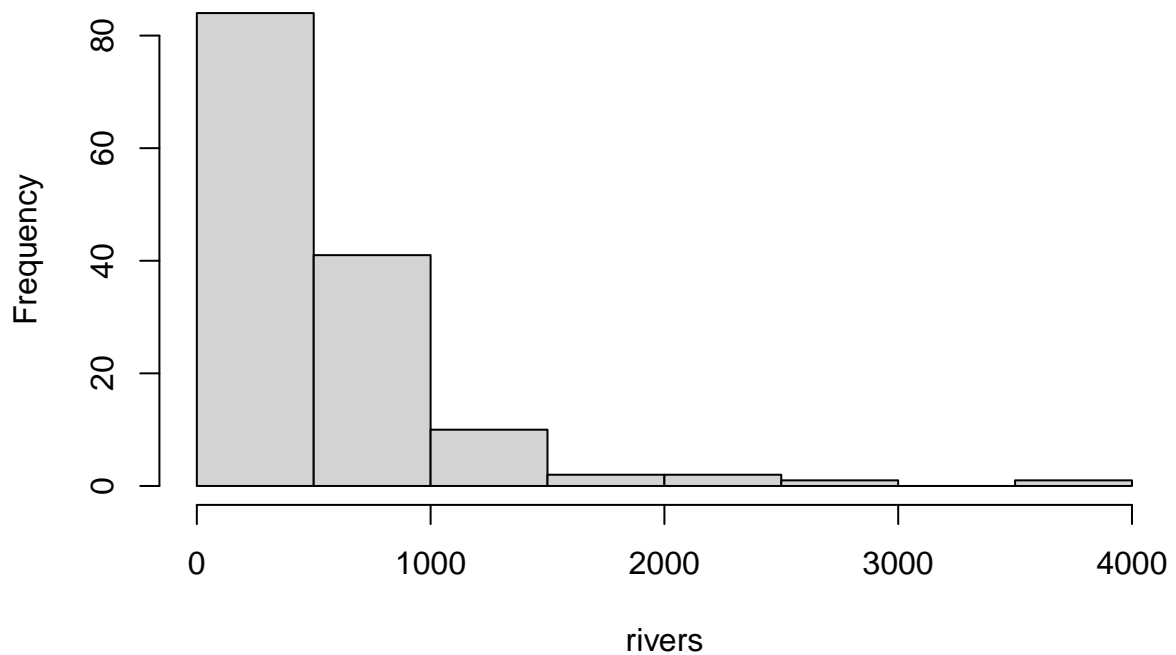


```
bvector2 <- seq(0,4000,by=500)
bvector2
```

```
## [1] 0 500 1000 1500 2000 2500 3000 3500 4000
```

```
hist(rivers,breaks=bvector2)
```

Histogram of rivers

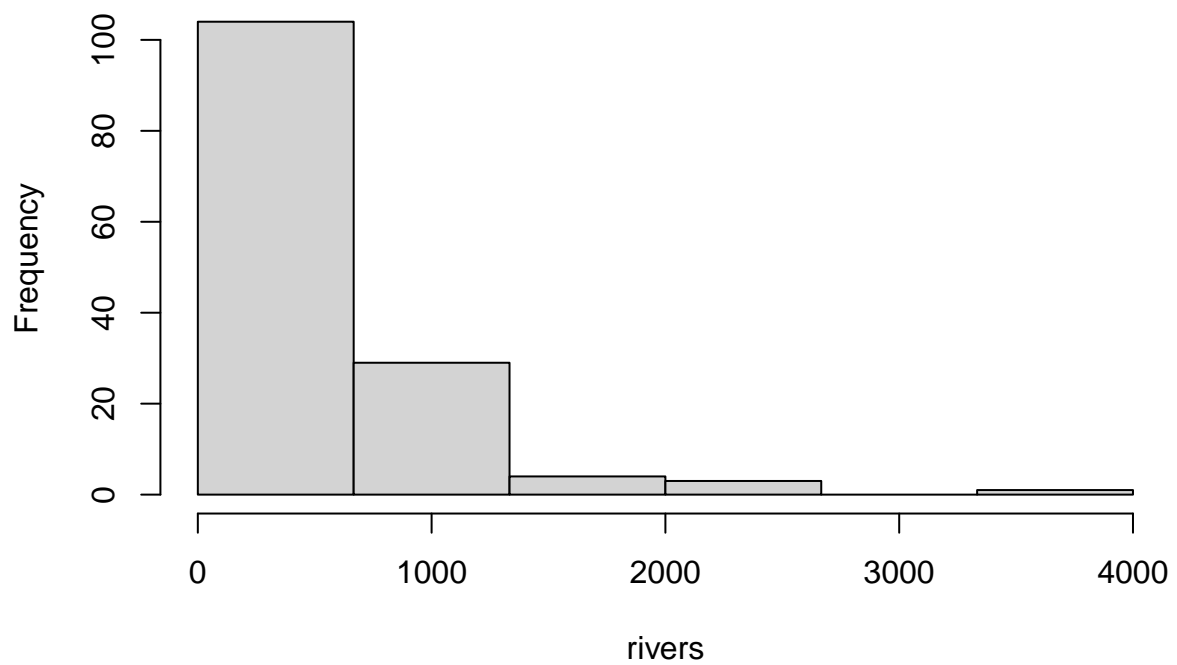


```
bvector3 <- seq(0,4000,length=7)
bvector3
```

```
## [1] 0.0000 666.6667 1333.3333 2000.0000 2666.6667 3333.3333 4000.0000
```

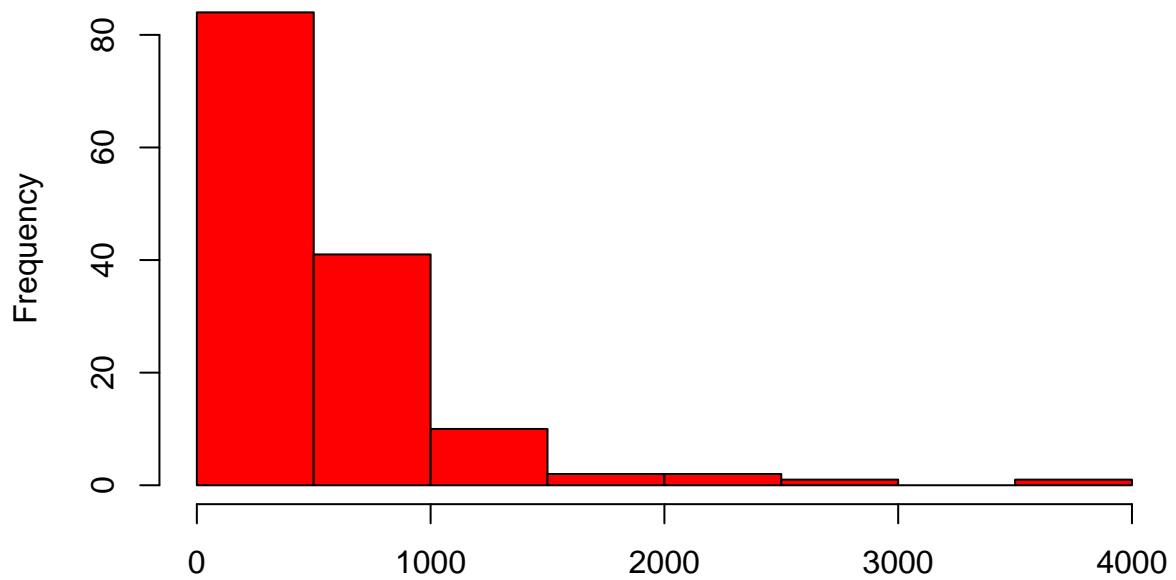
```
hist(rivers,breaks=bvector3)
```

Histogram of rivers



```
hist(rivers,breaks=bvector2,col="red", main="Distribution of Rivers",xlab="Rivers Data")
```

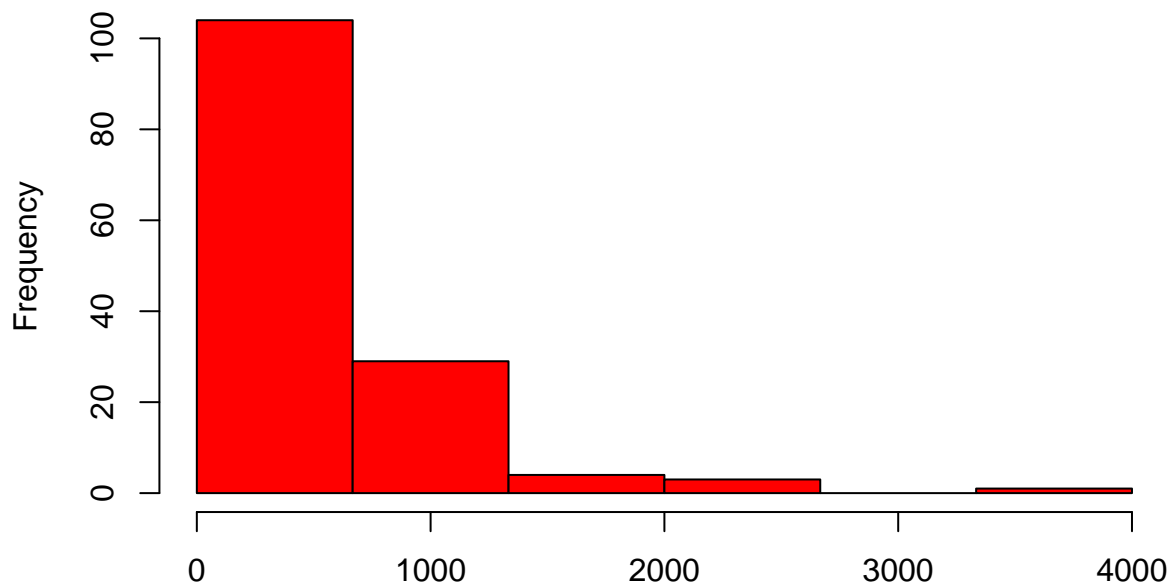
Distribution of Rivers



Rivers Data

```
hist(rivers,breaks=bvector3,col="red",freq=T)
```

Histogram of rivers



rivers

```
hist(rivers,breaks=bvector3,col="red",freq=FALSE)
```

Histogram of rivers

