Homework 2

View it online: http://acsweb.ucsd.edu/~djc035/Assignment2.html (http://acsweb.ucsd.edu/~djc035/Assignment2.html)

Objective

The objective of this study is to investigate the responses of the participants in the study with the intention of providing udeful information about the students to the designers of the new computer lab?

Scenario 1

Begin by providing an estimate for the fraction of students who played a video game in the week prior to the survey. Provide an interval estimate as well as a point estimate for this proportion.

Point and Interval Estimation

We are looking for the point estimate of the proportion of students who played a video game in the week prior to the survey. We created a new column called “binary_time” that is 1 if the student played more than 0 hours of a video game and 0 for students who played 0 hours.

```
#read in "videodata.txt" into data
data <- read.table("videodata.txt", header=TRUE)
data["binary_time"] <- ifelse(data$time>0, 1, 0)
head(data)
```

```
##   time like where freq busy educ sex age home math work own cdrom email grade
## 1  2.0    3     3    2    0    1   0  19    1    0   10   1     0     1     4
## 2  0.0    3     3    3    0    0   0  18    1    1    0   1     1     1     2
## 3  0.0    3     1    3    0    0   1  19    1    0    0   1     0     1     3
## 4  0.5    3     3    3    0    1   0  19    1    0    0   1     0     1     3
## 5  0.0    3     3    4    0    1   0  19    1    1    0   0     0     1     3
## 6  0.0    3     2    4    0    0   1  19    0    0   12   0     0     0     3
##   binary_time
## 1           1
## 2           0
## 3           0
## 4           1
## 5           0
## 6           0
```

The point estimation of the proportion of students who played a video game in the week prior to the survey is:

```
play.percentage <- mean(data$binary_time)
play.percentage
```

```
## [1] 0.3736264
```

Now we also want to have a confidence interval of this estimator. However, clearly the distribution of binary_time variable is not Normal, it is a Bernoulli random variable. We know our data were drawn from a population with size N=314. Hence, we first create a bootstrap population of this size by repeating every sample for 31491=3.45 times. Here, we'll just specify the parameter length.out to be 314.

```
set.seed(573929)
shuffle.ind=sample(1:nrow(data))
boot.population <- rep(data$binary_time[shuffle.ind], length.out = 314)
length(boot.population)
```

```
## [1] 314
```

Then we will choose n=91 samples from the Bootstrap population and call this a Bootstrap sample.

```
sample1 <- sample(boot.population, size = 91, replace = TRUE)
```

Continue this procedure until we have 400 Bootstrap samples.

```
B = 400 # the number of bootstrap samples we want
boot.sample <- array(dim = c(B, 91))
for (i in 1:B) {
  boot.sample[i, ] <- sample(boot.population, size = 91, replace = TRUE)
}
```

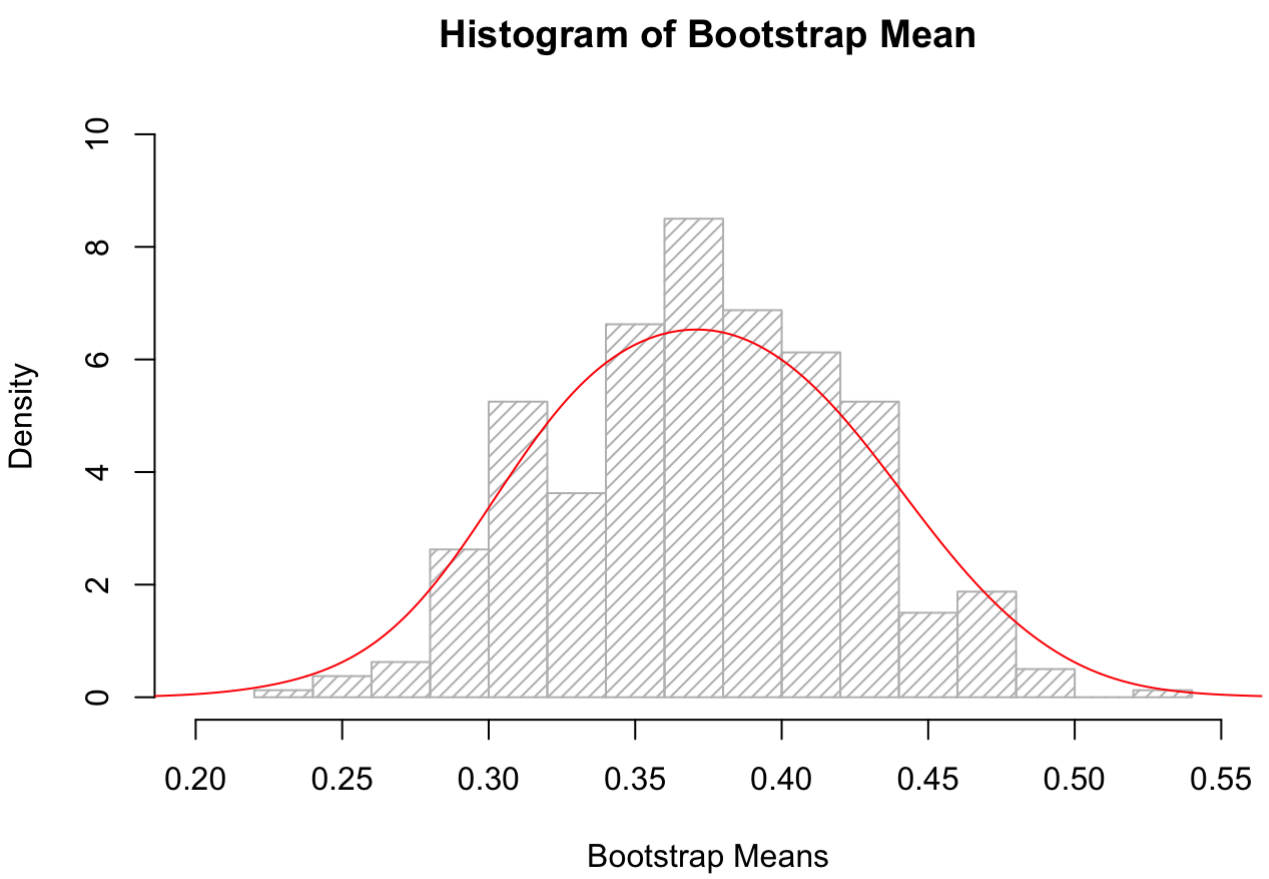
Then we can calculate the sample mean for each Bootstrap sample (i.e. each row of the Bootstrap sample matrix).

```
boot.mean <- apply(X = boot.sample, MARGIN = 1, FUN = mean)
head(boot.mean)
```

```
## [1] 0.3516484 0.2967033 0.3626374 0.3736264 0.3406593 0.4065934
```

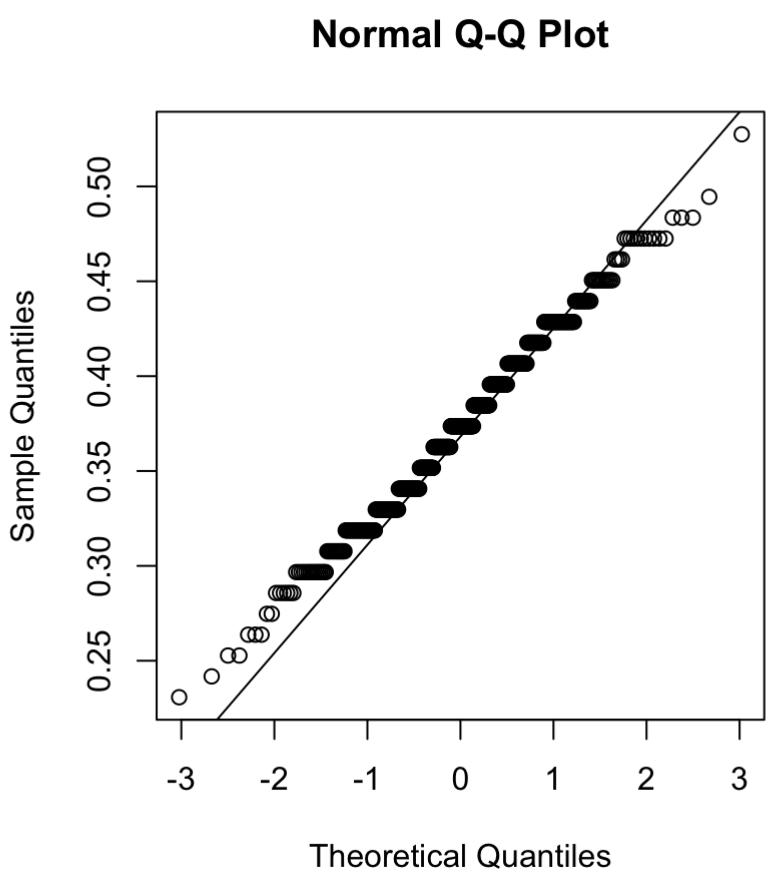
Let's see the histogram of these Bootstrap sample means.

```
hist(boot.mean, xlim=c(0.2,0.55), ylim=c(0,10), breaks = 20, probability = TRUE, density
= 20, col = 8, border = 8, main = "Histogram of Bootstrap Mean", xlab = "Bootstrap Mean
s")
lines(density(boot.mean, adjust = 2), col = 2)
```



Check Normality by Q-Q plot and Kolmogorov-Smirnov test.

```
par(pty = 's')
qqnorm(boot.mean)
qqline(boot.mean)
```




```
ks.test((boot.mean - mean(boot.mean))/sd(boot.mean), pnorm)
```

```
## Warning in ks.test((boot.mean - mean(boot.mean))/sd(boot.mean), pnorm): ties
## should not be present for the Kolmogorov-Smirnov test
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  (boot.mean - mean(boot.mean))/sd(boot.mean)
## D = 0.071014, p-value = 0.03539
## alternative hypothesis: two-sided
```

So we can accept that the sample mean follows a Normal distribution. Then we can construct 95% confidence intervals.

```
boot.sd <- sd(boot.mean)
play.percentage + c(-1, 1)*1.96*boot.sd
```

```
## [1] 0.2745089 0.4727438
```

Scenario 2

Check to see how the amount of time spent playing video games in the week prior to the survey compares to the reported frequency of play (daily, weekly, etc). How might the fact that there was an exam in the week prior to the survey affect your previous estimates and this comparison?

```
#read in "videodata.txt" into data
data <- read.table("videodata.txt", header=TRUE)
head(data)
```

```
##   time like where freq busy educ sex age home math work own cdrom email grade
## 1  2.0    3     3    2    0    1   0  19    1    0   10   1     0     1     4
## 2  0.0    3     3    3    0    0   0  18    1    1    0   1     1     1     2
## 3  0.0    3     1    3    0    0   1  19    1    0    0   1     0     1     3
## 4  0.5    3     3    3    0    1   0  19    1    0    0   1     0     1     3
## 5  0.0    3     3    4    0    1   0  19    1    1    0   0     0     1     3
## 6  0.0    3     2    4    0    0   1  19    0    0   12   0     0     0     3
```

Show the number of rows in data

```
nrow(data)
```

```
## [1] 91
```

```
#cleaning unknown data rows
data <- data[!(data$freq==99),]
```

Show the number of rows in data

```
nrow(data)
```

```
## [1] 78
```

Make four seperate data versions

```
#use which() function to select rows which contain observation of smokers
freq_1.ind <- which(data['freq'] == 1)
#we pass in the vector of indices and use setdiff() function to get the non-smokers
data.freq_1 <- data[freq_1.ind,]

freq_2.ind <- which(data['freq'] == 2)
data.freq_2 <- data[freq_2.ind,]

freq_3.ind <- which(data['freq'] == 3)
data.freq_3 <- data[freq_3.ind,]

freq_4.ind <- which(data['freq'] == 4)
data.freq_4 <- data[freq_4.ind,]

data.freq_1
```

```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 9     2    3     2    1    1    1   1  19    0    0    0   0     0     0     4
## 13    0    2     4    1    0    1   0  19    1    1    0   0     0     1     4
## 23    2    3     3    1    1    1   1  19    0    0    0   1     0     1     4
## 35    0    3     3    1    0    1   0  19    1    0   12   1     0     1     3
## 54    3    2     3    1    0    1   1  18    1    0    7   1     0     1     3
## 58    4    2    99    1    1    1   1  20    1    0    6   1     0     0     4
## 60   14    2    99    1    1    0   0  19    1    0    0   1     0     1     2
## 65   14    2     4    1    1    1   1  18    1    0   35   1     1     1     3
## 81    1    2     3    1    0    0   1  20    1    1    0   1     0     1     4
```

```
data.freq_2
```

```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 1   2.0    3     3    2    0    1   0  19    1    0   10   1     0     1     4
## 14  3.0    3     3    2    1    0   0  18    0    0    0   0     0     1     3
## 15  1.0    3     5    2    0    1   0  18    1    1   14   1     0     1     3
## 19  2.0    2     2    2    1    0   1  18    1    0    0   1     0     1     4
## 21  2.0    3     2    2    0    1   1  20    1    0   15   1     0     0     4
## 32  1.0    3     5    2    0    1   1  19    1    0   99   1     1     1     3
## 36  0.1    2     6    2    0    1   1  18    0    0    5   1     1     1     4
## 40  0.0    3     3    2    1    1   0  20    1    0   20   1     0     0     3
## 42  2.0    2     4    2    0    0   1  19    1    0    0   1     0     1     3
## 43  2.0    3     4    2    0    1   1  19    0    0   10   1     1     1     3
## 44  0.5    3     4    2    1    0   1  19    1    1   99   0     0     1     4
## 46  2.0    3     5    2    1    1   1  19    1    0   15   0     0     1     4
## 47  0.0    3     4    2    0    0   1  19    1    1    0   1     1     0     3
## 50  2.0    3     2    2    0    0   1  19    1    0    0   1     0     1     4
## 53  0.5    3     2    2    0    0   1  19    1    0   16   1     0     1     3
## 59 30.0    2    99    2    1    0   1  19    0    1    0   0     0     1     3
## 64  0.5    2     3    2    1    1   1  19    1    0   20   1     1     1     4
## 66  1.0    2     4    2    0    1   1  19    1    0   19   1     0     1     4
## 68  0.0    2     5    2    1    1   1  20    1    1   20   0     0     1     4
## 69  1.5    3     3    2    0    1   0  19    1    1    8   1     0     0     3
## 72  2.0    2    99    2    1   99   1  20    1    0   10   1     1     1     3
## 74  0.0    3     3    2    0    0   1  23    0    0    0   1     0     1     4
## 80  2.0    2     1    2    0    1   1  19    1    0   10   1     0     1     3
## 84  2.0    2     3    2    0    1   1  21    0    0   15   0     0     1     4
## 86  2.0    2     4    2    1    0   1  19    0    0    0   1     0     1     3
## 87  2.0    3     4    2    1    0   1  19    1    0    0   1    99     1     4
## 88  5.0    3     3    2    0    1   0  20    1    0   14   1     1     1     4
## 90  3.0    3     3    2    0    0   1  19    1    0    5   1     1     1     3
```

```
data.freq_3
```

```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 2   0.0    3     3    3    0    0   0  18    1    1    0   1     1     1     2
## 3   0.0    3     1    3    0    0   1  19    1    0    0   1     0     1     3
## 4   0.5    3     3    3    0    1   0  19    1    0    0   1     0     1     3
## 18  0.0    3     2    3    0    0   1  20    1    0    0   1     0     1     3
## 22  0.0    3     2    3    0    1   1  24    1    0   10   0     0     0     4
## 28  0.0    3     2    3    0    0   1  18    0    0   10   0     0     0     3
## 33  0.0    4     2    3    0    0   1  19    1    1    0   1     1     1     3
## 34  0.0    2     1    3    0    0   1  19    1    0   10   0     0     1     3
## 37  0.5    4     3    3    0    0   0  19    1    0    0   1     0     0     3
## 48  0.0    3     4    3    1    1   0  19    1    1    0   1     0     1     3
## 55  0.0    3     1    3    0    0   1  19    0    0   15   0     0     1     3
## 56  0.0    4     3    3    0    1   0  21    1    0    5   1     0     1     4
## 61  0.0    3     1    3    0    1   1  19    0    0    0   0     0     0     3
## 62  0.0    2    99    3    0    1   0  21    0    0   18   1     0     0     2
## 71  0.0    3     4    3    0    0   1  19    1    1    0   1     0     1     3
## 76  0.0    2     3    3    0    1   0  20    0    0    0   1     1     1     2
## 79  0.0    2     3    3    0    1   1  25    0    0   55   1     0     1     3
## 91  0.0    3     4    3    0    1   0  19    0    1    5   1     0     1     2
```

```
data.freq_4
```

```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 5     0    3     3    4    0    1   0  19    1    1    0   0     0     1     3
## 6     0    3     2    4    0    0   1  19    0    0   12   0     0     0     3
## 7     0    4     3    4    0    0   1  20    1    1   10   1     0     1     3
## 8     0    3     3    4    0    0   0  19    1    0   13   0     0     1     3
## 10    0    3     3    4    0    1   1  19    1    1    0   1     0     1     4
## 11    0    3     1    4    0    0   0  20    1    0    0   1     0     0     3
## 12    0    3     2    4    0    0   0  19    1    0    0   1     0     1     4
## 17    0    3     3    4    0    1   1  21    1    0    2   1     0     1     4
## 26    0    3     3    4    0   99   0  20    1    1    0   1     0     1     3
## 27    0    2     3    4    0    0   1  22    1    1    0   1     1     1     4
## 29    0    4     3    4    0    0   1  19    1    1    0   1     0     1     3
## 30    0    4     3    4    0    1   0  20    1    0    0   1     0     1     3
## 31    0    4     3    4    0    0   0  19    1    1    0   0     0     1     4
## 38    1    3     4    4   99    1   0  20    1    0    0   1     0     1     3
## 39    0    3     1    4    0    0   0  19    0    0    0   0     0     1     3
## 51    0    4    99    4    0   99   0  18    1    1    0   1     0     1     3
## 57    0    4     3    4    0    0   0  18    1    0    0   1     0     1     4
## 67    0    4     2    4    0    0   1  18    1    0    0   1     0     1     4
## 70    0    4     2    4    0    0   1  19    1    1    0   1     0     0     3
## 78    0    3     3    4    0    0   1  19    1    0   16   0     0     1     3
## 83    0    3     2    4    0    0   0  19    0    1   15   0    99     1     2
## 85    0    3     2    4    0    0   0  18    1    1   15   0    99     0     3
## 89    0    2     5    4    0    1   0  33    1    0   40   1     0     0     2
```

```
mean(data.freq_1$time)
```

```
## [1] 4.444444
```

```
mean(data.freq_2$time)
```

```
## [1] 2.539286
```

```
mean(data.freq_3$time)
```

```
## [1] 0.05555556
```

```
mean(data.freq_4$time)
```

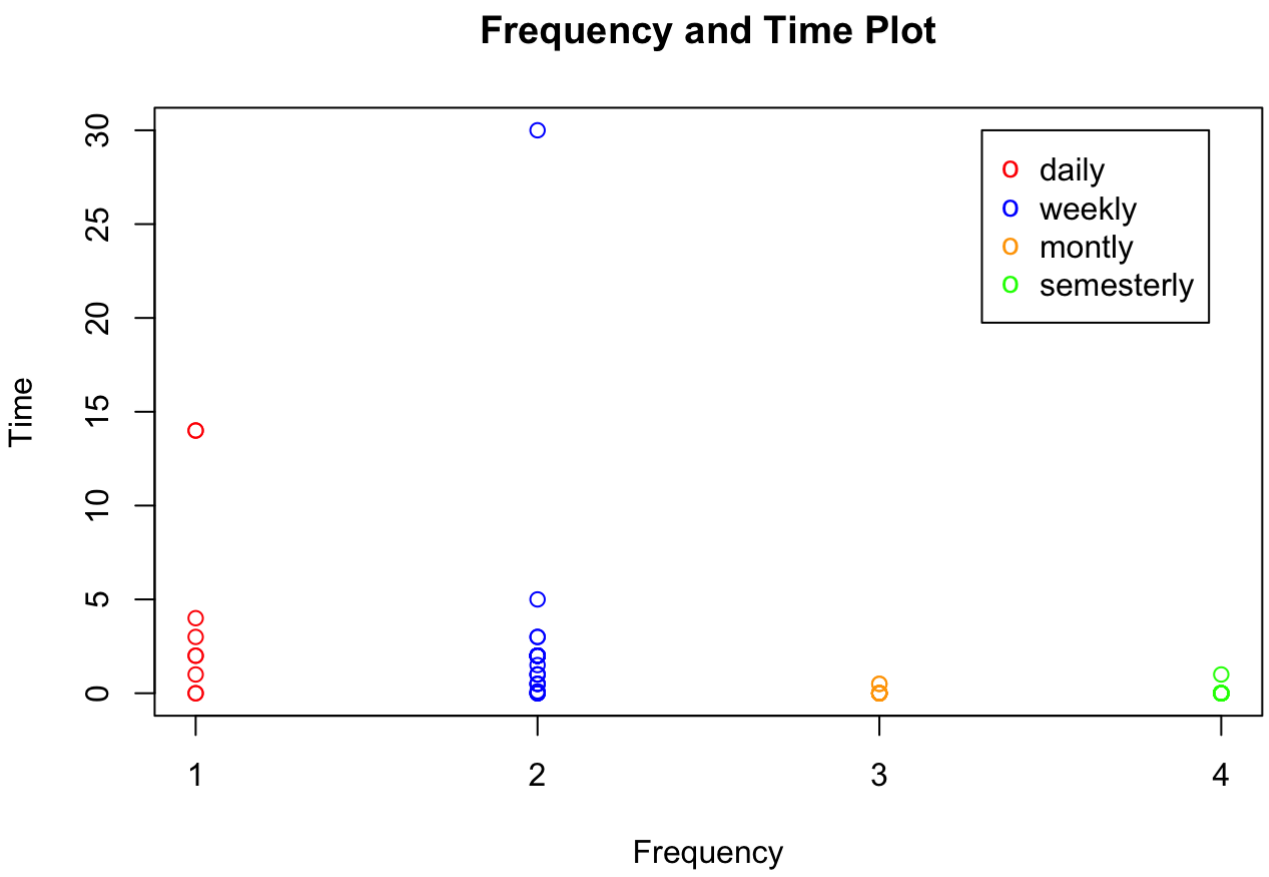
```
## [1] 0.04347826
```

Plot for frequency and time.

```
data$color[data$freq==1]="red"
data$color[data$freq==2]="blue"
data$color[data$freq==3]="orange"
data$color[data$freq==4]="green"
data$color
```

```
##  [1] "blue"   "orange" "orange" "orange" "green"  "green"  "green"  "green"
##  [9] "red"    "green"  "green"  "green"  "red"    "blue"   "blue"   "green"
## [17] "orange" "blue"   "blue"   "orange" "red"    "green"  "green"  "orange"
## [25] "green"  "green"  "green"  "blue"   "orange" "orange" "red"    "blue"
## [33] "orange" "green"  "green"  "blue"   "blue"   "blue"   "blue"   "blue"
## [41] "blue"   "orange" "blue"   "green"  "blue"   "red"    "orange" "orange"
## [49] "green"  "red"    "blue"   "red"    "orange" "orange" "blue"   "red"
## [57] "blue"   "green"  "blue"   "blue"   "green"  "orange" "blue"   "blue"
## [65] "orange" "green"  "orange" "blue"   "red"    "green"  "blue"   "green"
## [73] "blue"   "blue"   "blue"   "green"  "blue"   "orange"
```

```
plot(x = data$freq, y = data$time, col=data$color, axes=FALSE, main = "Frequency and Tim
e Plot", xlab = "Frequency", ylab = "Time")
axis(side=1, at=c(1, 2, 3, 4))
axis(side=2, at=seq(0, 30, by=5))
legend(3.3, 30, legend=c("daily","weekly","montly","semesterly"), col=c("red", "blue",
"orange", "green"), pch="o")
box()
```



Grouped bar chart for busy or not busy frequency and time

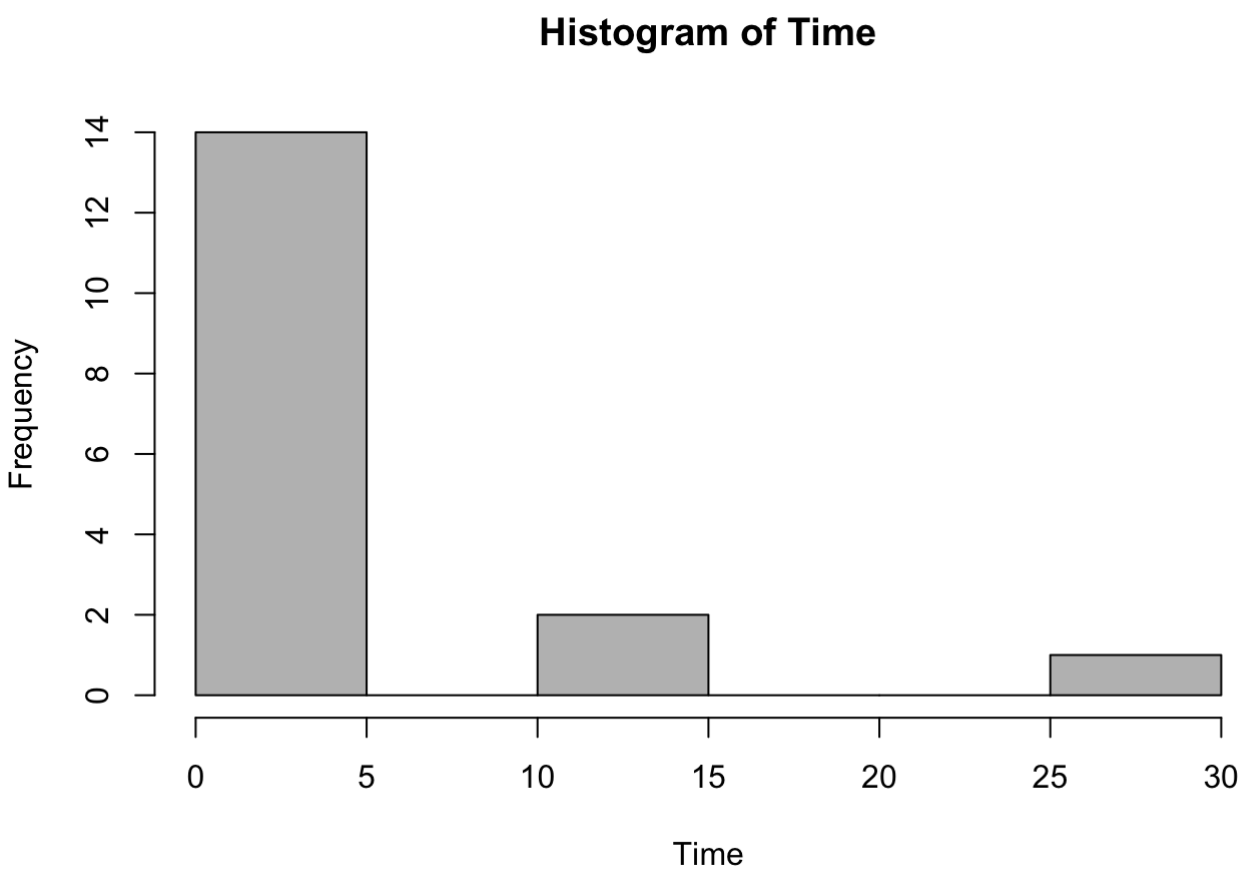
```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

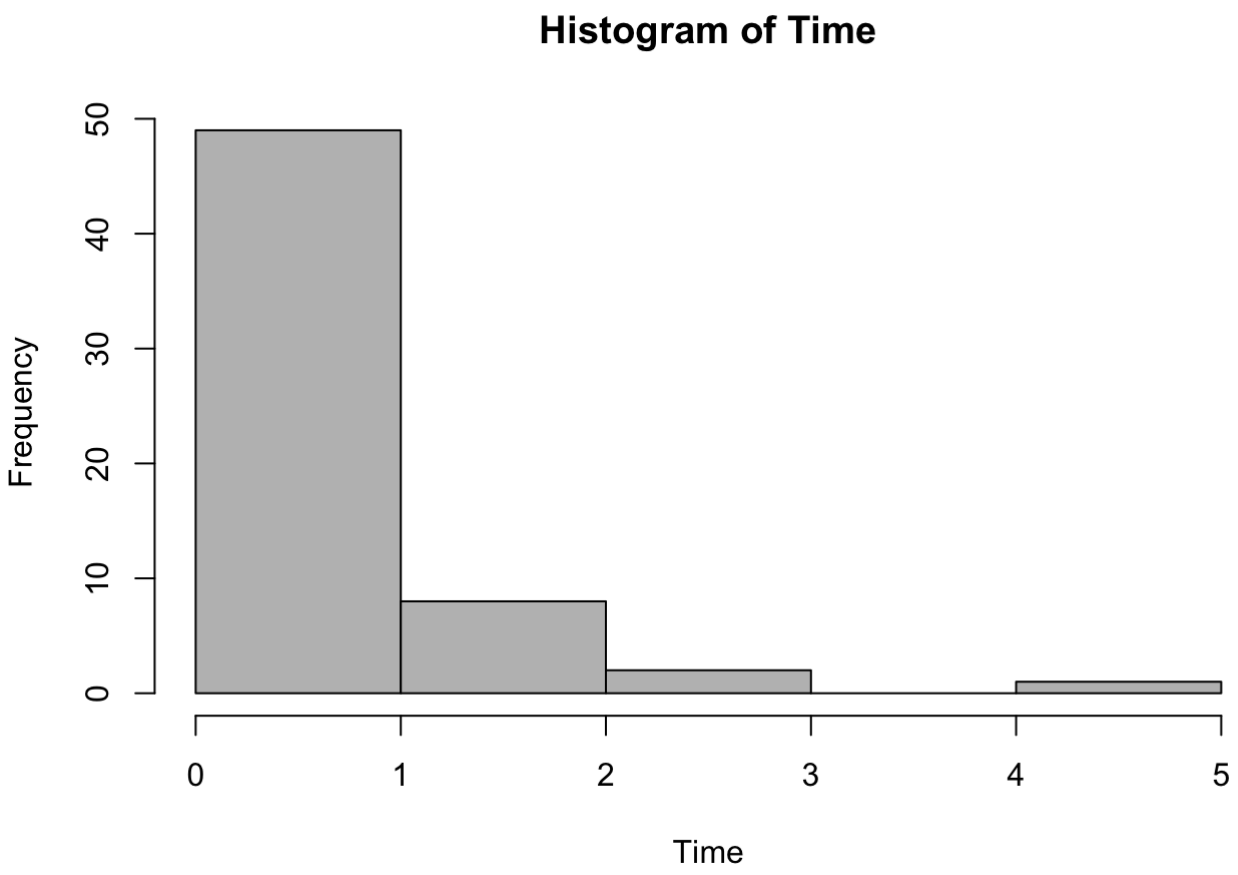
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

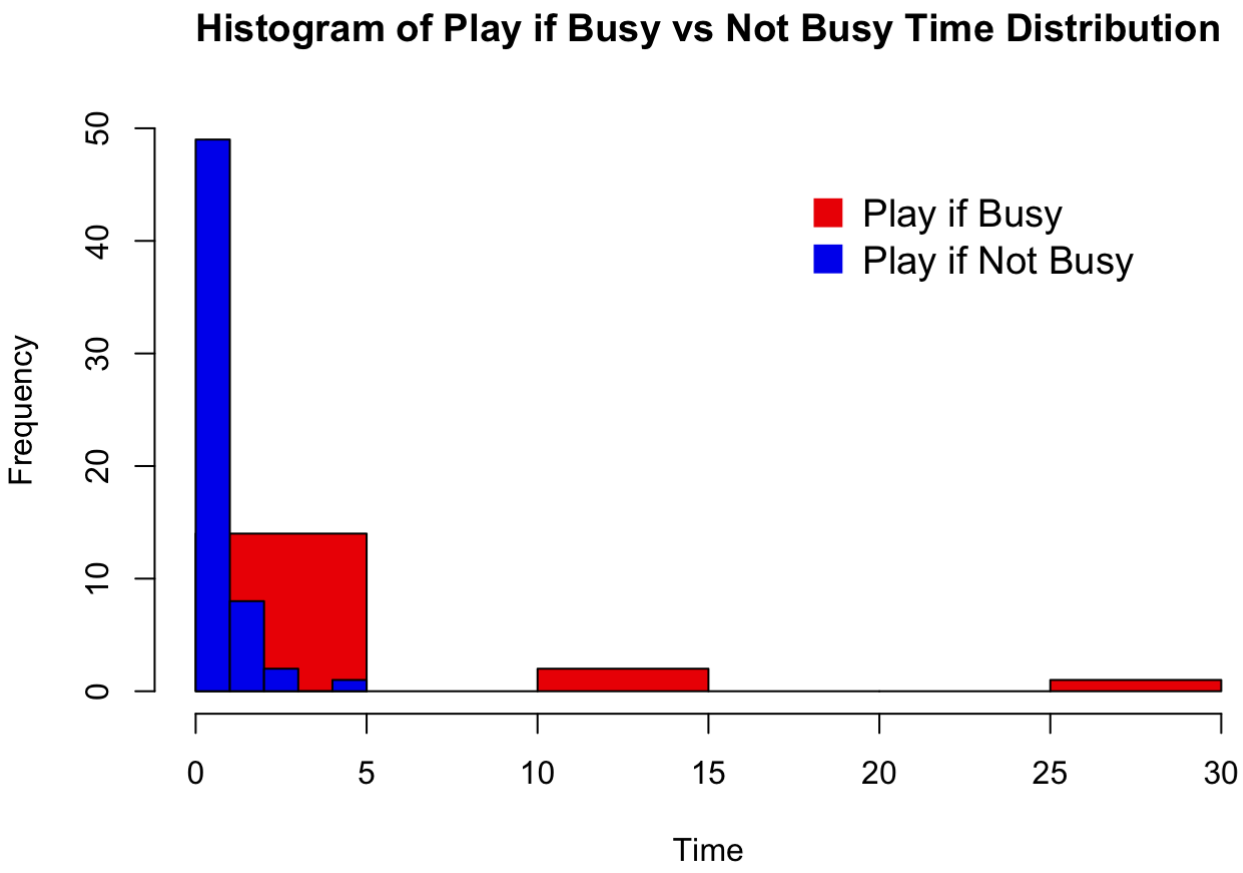
```
data <- read.table("videodata.txt", header=TRUE)
data <- data[!(data$freq==99 | data$busy==99),]

busy.ind <- which(data['busy'] == 1)
data.busy <- data[busy.ind,]

not_busy.ind <- which(data['busy'] == 0)
data.not_busy <- data[not_busy.ind,]

#mutate(data, busy = factor(busy, labels = c("busy", "not busy")), freq = factor(freq))

# Bar chart side by side
#ggplot(data, aes(x = freq, fill = busy)) + geom_bar(position=position_dodge()) + theme_
classic

#ggplot(data, aes(freq, )) + geom_bar(aes(fill = busy), position = "dodge")

c1 <- "red2"
c2 <- "blue2"

h_busy <- hist(data.busy$time, col=8, breaks=5, main = "Histogram of Time", xlab = "Tim
e")
```




```
h_not_busy <- hist(data.not_busy$time, col=8, breaks=5, main = "Histogram of Time", xlab
= "Time")
```




```
plot(h_busy, col = c1, ylim=c(0,50), main="Histogram of Play if Busy vs Not Busy Time Di
stribution", xlab="Time")
plot(h_not_busy, col = c2, add = TRUE)
legend("topright",
  legend = c("Play if Busy", "Play if Not Busy"),
  col = c(c1, c2),
  pch = c(15,15),
  bty = "n",
  pt.cex = 2,
  cex = 1.2,
  text.col = "black",
  horiz = F ,
  inset = c(0.1, 0.1))
```




```
data <- read.table("videodata.txt", header=TRUE)
data <- data[!(data$freq==99 | data$busy==99),]

busy.ind <- which(data['busy'] == 1)
data.busy <- data[busy.ind,]

not_busy.ind <- which(data['busy'] == 0)
data.not_busy <- data[not_busy.ind,]
nrow(data.busy)/nrow(data)
```

```
## [1] 0.2207792
```

```
nrow(data.not_busy)/nrow(data)
```

```
## [1] 0.7792208
```

Scenario 3

Consider making an interval estimate for the average amount of time spent playing video games in the week prior to the survey. Keep in mind the overall shape of the sample distribution. A simulation study may help determine the appropriateness of a interval estimate.

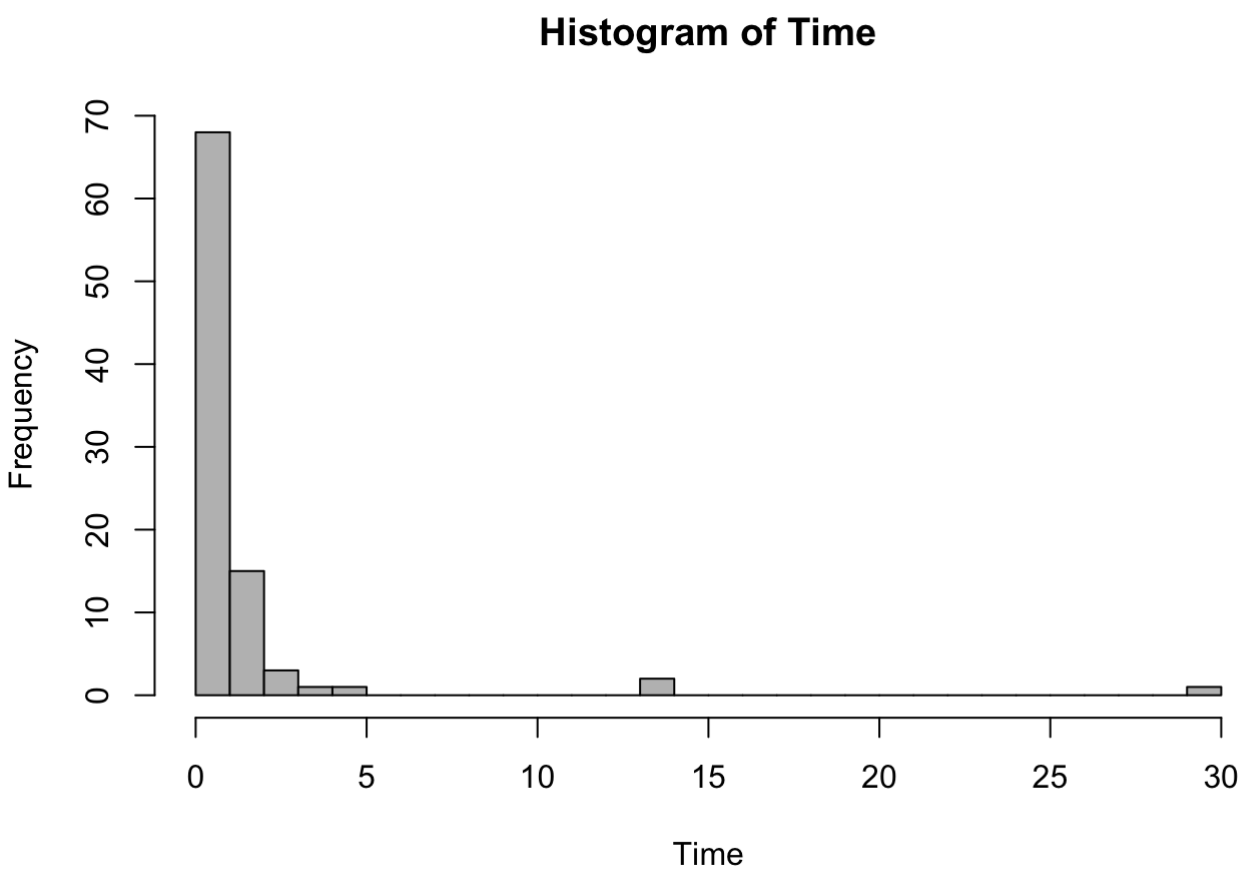
```
data <- read.table("videodata.txt", header=TRUE)
head(data)
```

```
##   time like where freq busy educ sex age home math work own cdrom email grade
## 1  2.0    3     3    2    0    1   0  19    1    0   10   1     0     1     4
## 2  0.0    3     3    3    0    0   0  18    1    1    0   1     1     1     2
## 3  0.0    3     1    3    0    0   1  19    1    0    0   1     0     1     3
## 4  0.5    3     3    3    0    1   0  19    1    0    0   1     0     1     3
## 5  0.0    3     3    4    0    1   0  19    1    1    0   0     0     1     3
## 6  0.0    3     2    4    0    0   1  19    0    0   12   0     0     0     3
```

```
time.mean <- mean(data$time)
time.mean
```

```
## [1] 1.242857
```

```
hist(data$time, col=8, breaks=30, main = "Histogram of Time", xlab = "Time")
```



```
#old seed set.seed(189289)
set.seed(573929)
shuffle.ind=sample(1:nrow(data))
boot.population <- rep(data$time[shuffle.ind], length.out = 314)
length(boot.population)
```

```
## [1] 314
```

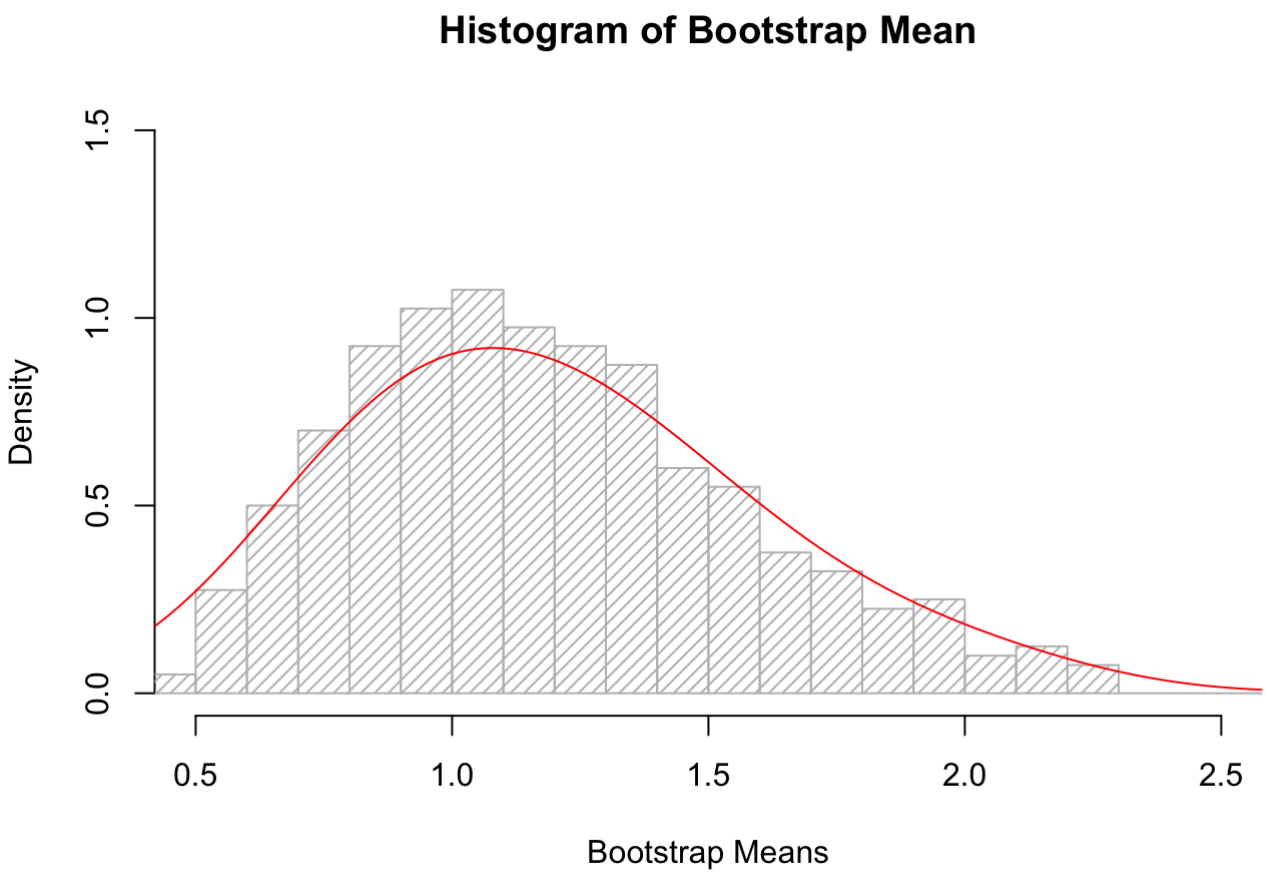
```
sample1 <- sample(boot.population, size = 91, replace = TRUE)
```

```
B = 400 # the number of bootstrap samples we want
boot.sample <- array(dim = c(B, 91))
for (i in 1:B) {
  boot.sample[i, ] <- sample(boot.population, size = 91, replace = TRUE)
}
```

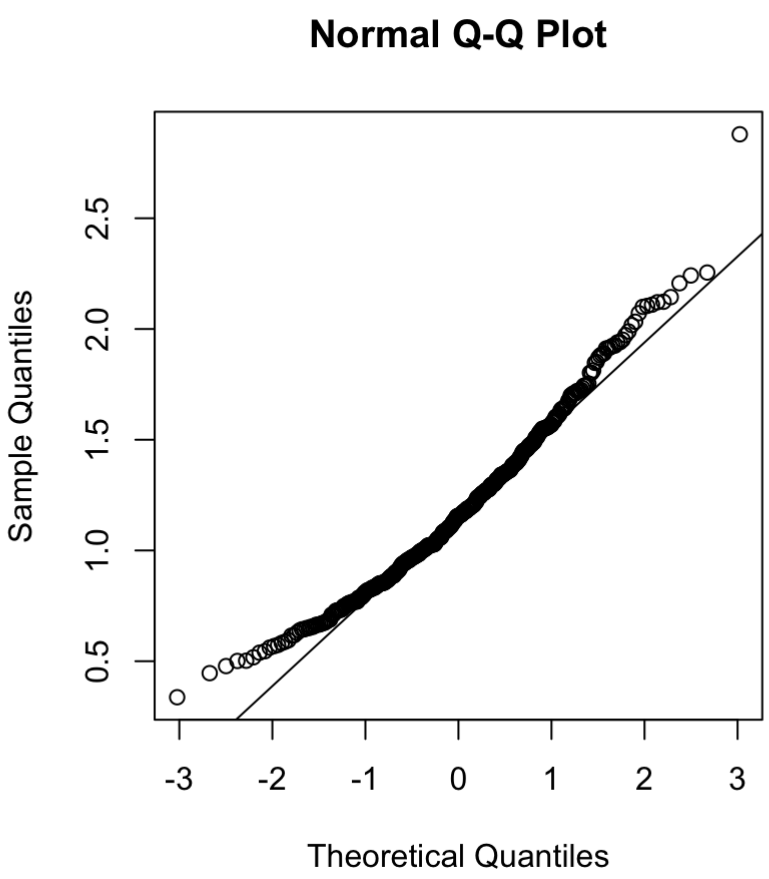
```
boot.mean <- apply(X = boot.sample, MARGIN = 1, FUN = mean)
head(boot.mean)
```

```
## [1] 1.0010989 0.5824176 0.9021978 0.6549451 1.1604396 1.2417582
```

```
hist(boot.mean, xlim=c(0.5,2.5), ylim=c(0,1.5), breaks = 20, probability = TRUE, density
= 20, col = 8, border = 8, main = "Histogram of Bootstrap Mean", xlab = "Bootstrap Mean
s")
lines(density(boot.mean, adjust = 2), col = 2)
```




```
par(pty = 's')
qqnorm(boot.mean)
qqline(boot.mean)
```




```
ks.test((boot.mean - mean(boot.mean))/sd(boot.mean), pnorm)
```

```
## Warning in ks.test((boot.mean - mean(boot.mean))/sd(boot.mean), pnorm): ties
## should not be present for the Kolmogorov-Smirnov test
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  (boot.mean - mean(boot.mean))/sd(boot.mean)
## D = 0.060836, p-value = 0.1035
## alternative hypothesis: two-sided
```

```
boot.sd <- sd(boot.mean)
time.mean + c(-1, 1)*1.96*boot.sd
```

```
## [1] 0.4812784 2.0044359
```

Scenario 4

Next consider the “attitude” questions. In general, do you think the students enjoy playing video games? if you had to make a short list of the most important reasons why studetns like/dislike video games, what would you put on the list? Don’t forget that those students who say that they have never played video games or do not at all like

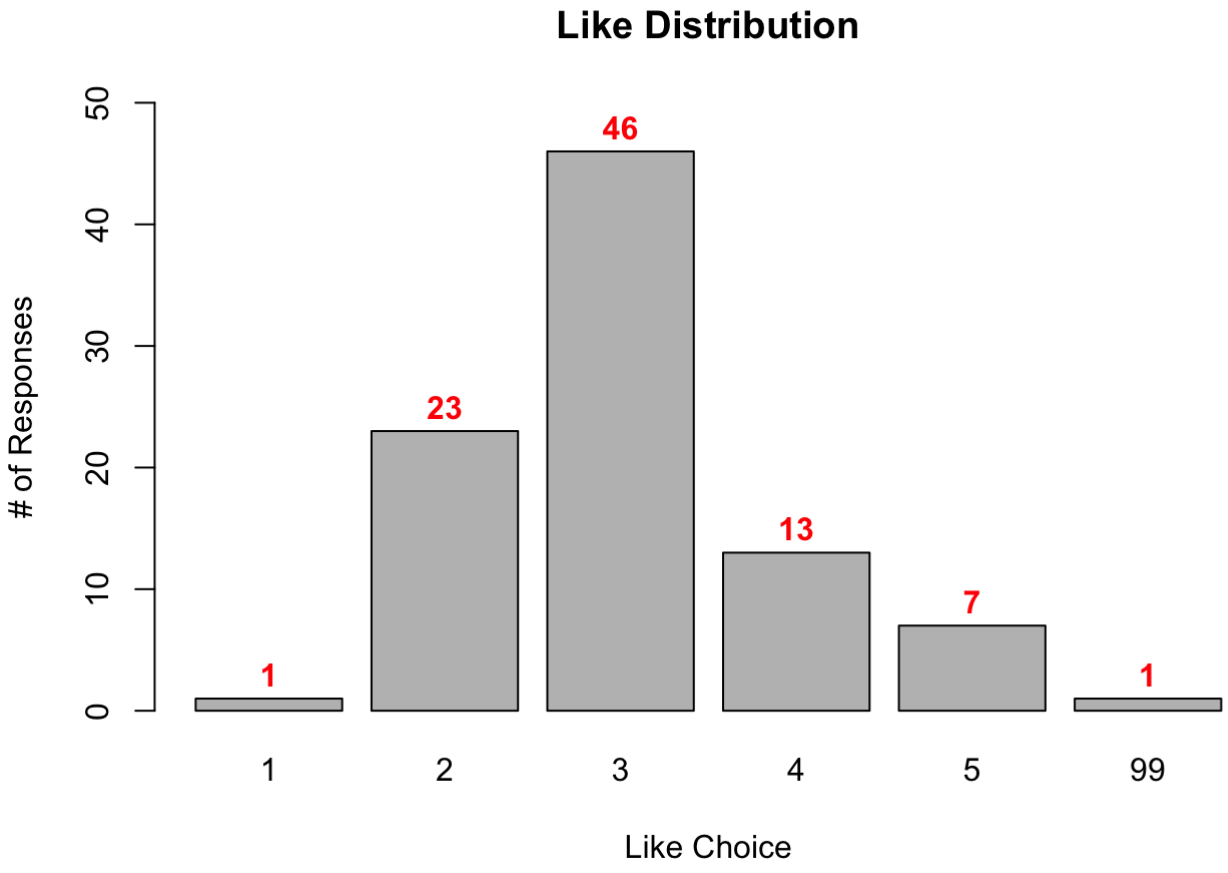
video games are asked to skip over some of these questions. So, there may be many nonrespondents to the questions as to whether they think video games are educational, where they play video games, etc.

Barplot for “like” distribution in the “videodata.txt” dataset

```
counts <- table(data$like)
counts
```

```
##
##  1  2  3  4  5 99
##  1 23 46 13  7  1
```

```
bp <- barplot(counts, main="Like Distribution", xlab="Like Choice", ylab="# of Response
s", ylim=c(0,50), col=8)
text(bp, counts + 2, counts, font=2, col=2)
```




```
data <- read.table("videodata.txt", header=TRUE)

like.ind <- which(!(data$like==1 | data$like==5 | data$like==99))
like <- data[!(data$like==1 | data$like==5 | data$like==99),]

like.ind
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 17 18 19 20 21 22 23 26 27 28
## [26] 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 50 51 53 54 55
## [51] 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 74 76 78 79 80 81 83 84
## [76] 85 86 87 88 89 90 91
```

Reading in the new “videoMultiple.txt” dataset

```
#read in "videoMultiple.txt" into data
data <- read.table("videoMultiple.txt", header=TRUE)
data.like <- data[like.ind,]
data = data.like
head(data)
```

```
##   action adv sim sport strategy relax coord challenge master bored other
## 1      0   0   0     0        1     1     0         1      1     0
## 2      0   1   0     0        1     0     0         0      0     1
## 3      1   0   0     1        1     1     0         0      0     0
## 4      0   0   0     0        1     0     0         1      0     0
## 5      0   0   0     0        1     1     0         1      1     0
## 6      1   0   0     1        1     1     0         0      1     1
##   graphic time frust lonely rules cost boring friends point other2
## 1       0    1     0      0     0    1      0       0     1
## 2       0    1     1      0     0    0      0       0     0
## 3       0    0     0      0     0    1      0       0     0
## 4       0    1     0      0     0    0      0       0     0
## 5       0    0     0      0     1    1      0       0     0
## 6       0    1     1      0     0    1      0       0     0
```

Only like = 2,3,4

```
counts <- table(like$like)
counts
```

```
##
##  2  3  4
## 23 46 13
```

```
bp <- barplot(counts, main="New Like Distribution", xlab="Like Choice", ylab="# of Respo
nses", ylim=c(0,50), col=8)
text(bp, counts + 2, counts, font=2, col=2)
```

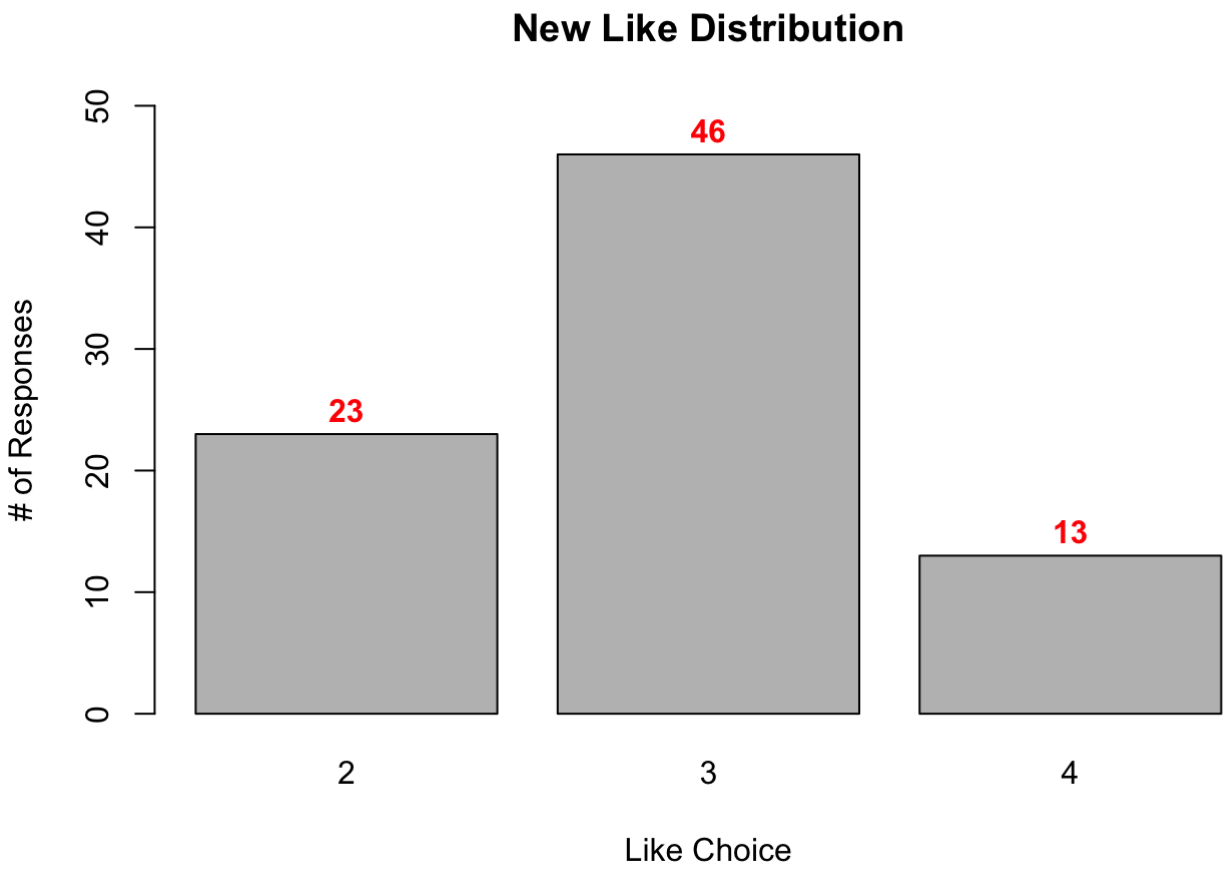


Table 1

action = Action

adv = Adventure

sim = Simulation

sport = Sports

strategy = Strategy

```
table1 <- data[,c("action", "adv", "sim", "sport", "strategy")]
head(table1)
```

```
##   action adv sim sport strategy
## 1      0   0   0     0        1
## 2      0   1   0     0        1
## 3      1   0   0     1        1
## 4      0   0   0     0        1
## 5      0   0   0     0        1
## 6      1   0   0     1        1
```

Table 2

graphic = Graphics/Realism

relax = Relaxation

coord = Eye/hand coordination

challenge = MentalChallenge

master = Felling of mastery

bored = Bored

other = other

```
table2 <- data[,c("graphic", "relax", "coord", "challenge", "master", "bored")]
head(table2)
```

```
##   graphic relax coord challenge master bored
## 1       0     1     0         1      1     0
## 2       0     0     0         0      0     1
## 3       0     1     0         0      0     0
## 4       0     0     0         1      0     0
## 5       0     1     0         1      1     0
## 6       0     1     0         0      1     1
```

```
graphic.ind <- which(table2["graphic"] == 1)
table2.graphic <- table2[graphic.ind,]

relax.ind <- which(table2["relax"] == 1)
table2.relax <- table2[relax.ind,]

coord.ind <- which(table2["coord"] == 1)
table2.coord <- table2[coord.ind,]

challenge.ind <- which(table2["challenge"] == 1)
table2.challenge <- table2[challenge.ind,]

master.ind <- which(table2["master"] == 1)
table2.master <- table2[master.ind,]

bored.ind <- which(table2["bored"] == 1)
table2.bored <- table2[bored.ind,]

like <- matrix(c(nrow(table2.graphic),nrow(table2.relax),nrow(table2.coord),nrow(table2.
challenge),nrow(table2.master),nrow(table2.bored)),ncol=6,byrow=TRUE)
colnames(like) <- c("graphic","relax","coord","challenge","master","bored")
like <- as.table(like)
like
```

```
##   graphic relax coord challenge master bored
## A      23    58     4        21     25    24
```

```
bp <- barplot(like, main="Like Game Reason Distribution", xlab="Categories", ylab="# of
 Responses", ylim=c(0,65), col=8)
text(bp, like + 2, like, font=2, col=2)
```

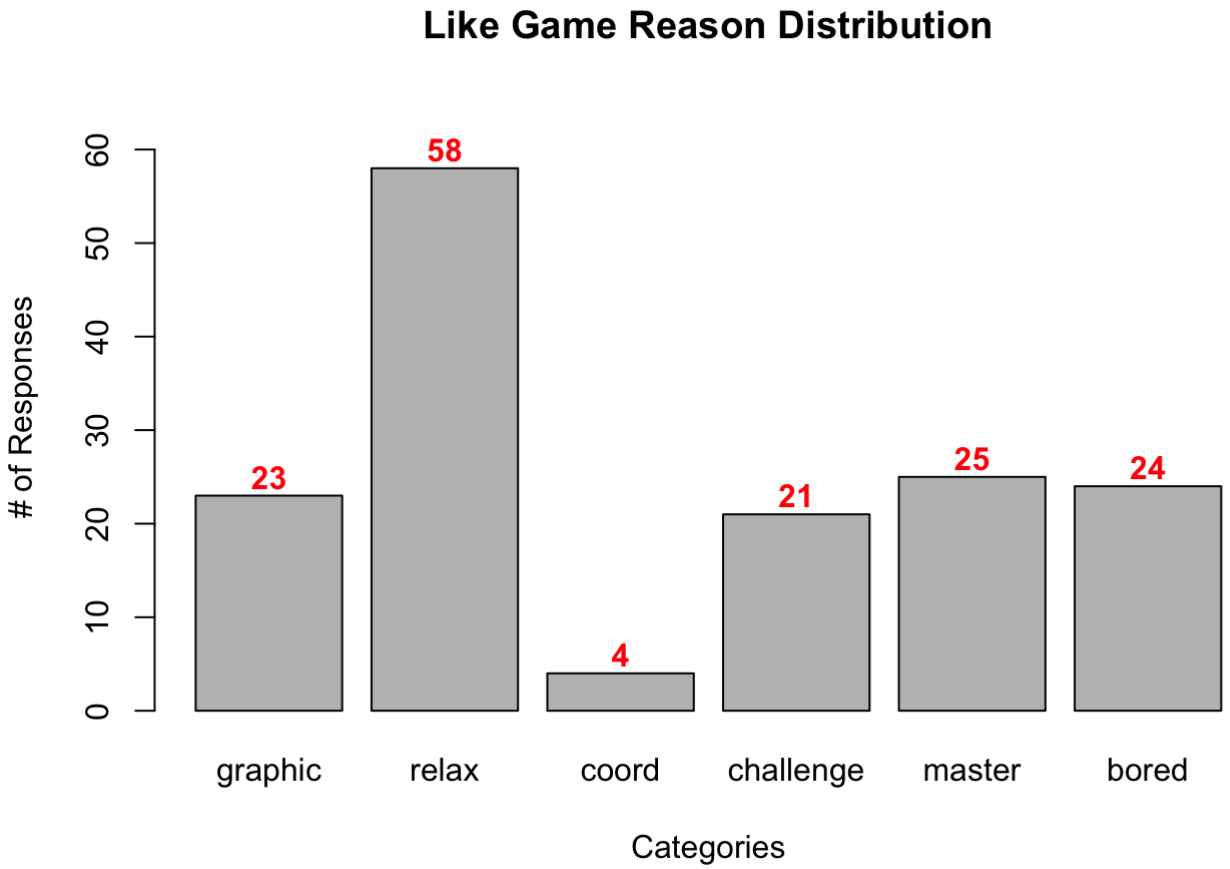


Table 3

time = Too much time

frust = Frustrating

lonely = Lonely

rules = Too many rules

cost = Costs too much

boring = Boring

friends = Friend's don't play

point = It is pointless

other2 = other

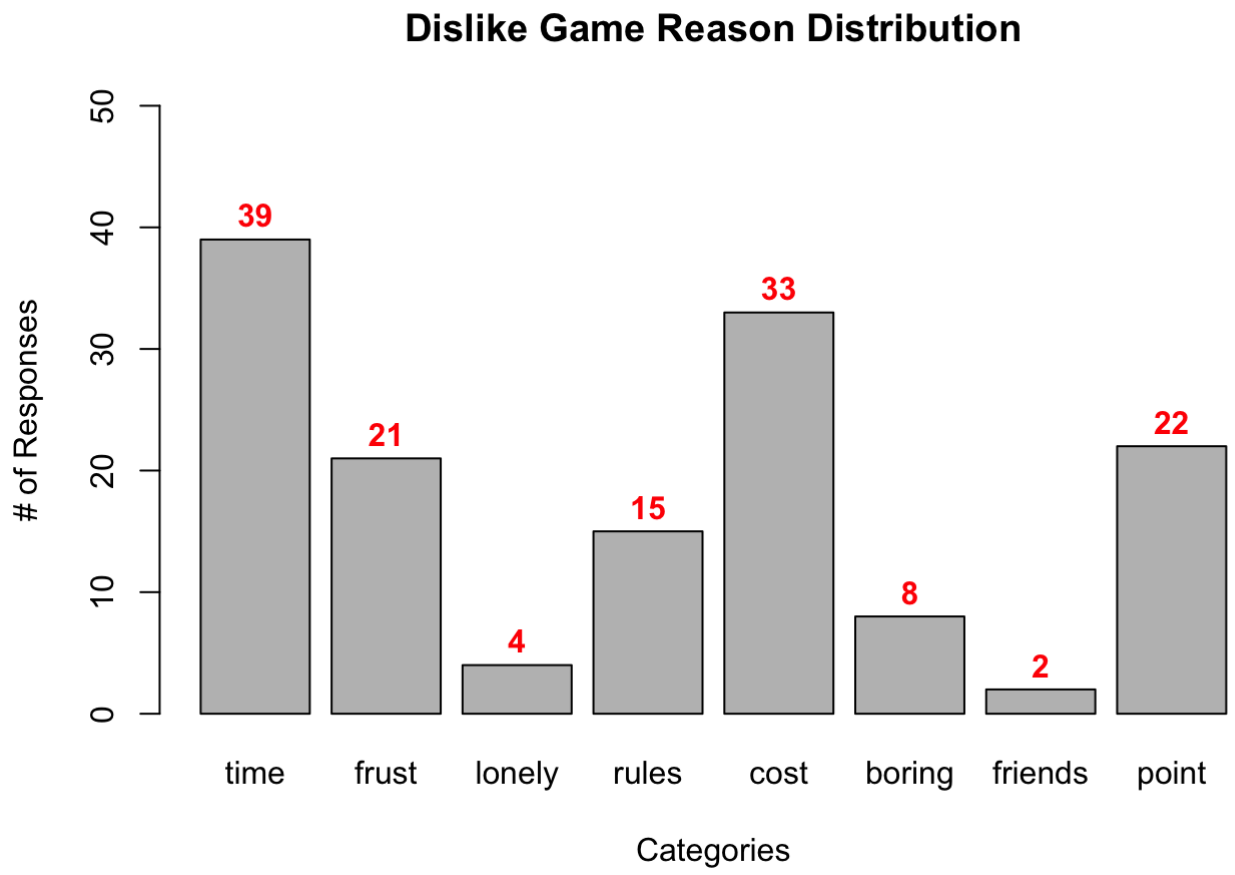
```
table3 <- data[,c("time", "frust", "lonely", "rules", "cost", "boring", "friends", "poin
t")]
head(table3)
```

```
##   time frust lonely rules cost boring friends point
## 1    1     0      0     0    1      0       0     1
## 2    1     1      0     0    0      0       0     0
## 3    0     0      0     0    1      0       0     0
## 4    1     0      0     0    0      0       0     0
## 5    0     0      0     1    1      0       0     0
## 6    1     1      0     0    1      0       0     0
```

```
time.ind <- which(table3["time"] == 1)
table3.time <- table3[time.ind,]

frust.ind <- which(table3["frust"] == 1)
table3.frust <- table3[frust.ind,]

lonely.ind <- which(table3["lonely"] == 1)
table3.lonely <- table3[lonely.ind,]

rules.ind <- which(table3["rules"] == 1)
table3.rules <- table3[rules.ind,]

cost.ind <- which(table3["cost"] == 1)
table3.cost <- table3[cost.ind,]

boring.ind <- which(table3["boring"] == 1)
table3.boring <- table3[boring.ind,]

friends.ind <- which(table3["friends"] == 1)
table3.friends <- table3[friends.ind,]

point.ind <- which(table3["point"] == 1)
table3.point <- table3[point.ind,]

dislike <- matrix(c(nrow(table3.time),nrow(table3.frust),nrow(table3.lonely),nrow(table
3.rules),nrow(table3.cost),nrow(table3.boring),nrow(table3.friends),nrow(table3.point)),
ncol=8,byrow=TRUE)
colnames(dislike) <- c("time", "frust", "lonely", "rules", "cost", "boring", "friends",
"point")
dislike <- as.table(dislike)
dislike
```

```
##   time frust lonely rules cost boring friends point
## A   39    21      4    15   33      8       2    22
```

```
bp <- barplot(dislike, main="Dislike Game Reason Distribution", xlab="Categories", ylab=
"# of Responses", ylim=c(0,50), col=8)
text(bp, dislike + 2, dislike, font=2, col=2)
```



Advanced Analysis and Scenario 5

```
data <- read.table("videodata.txt", header=TRUE)
```

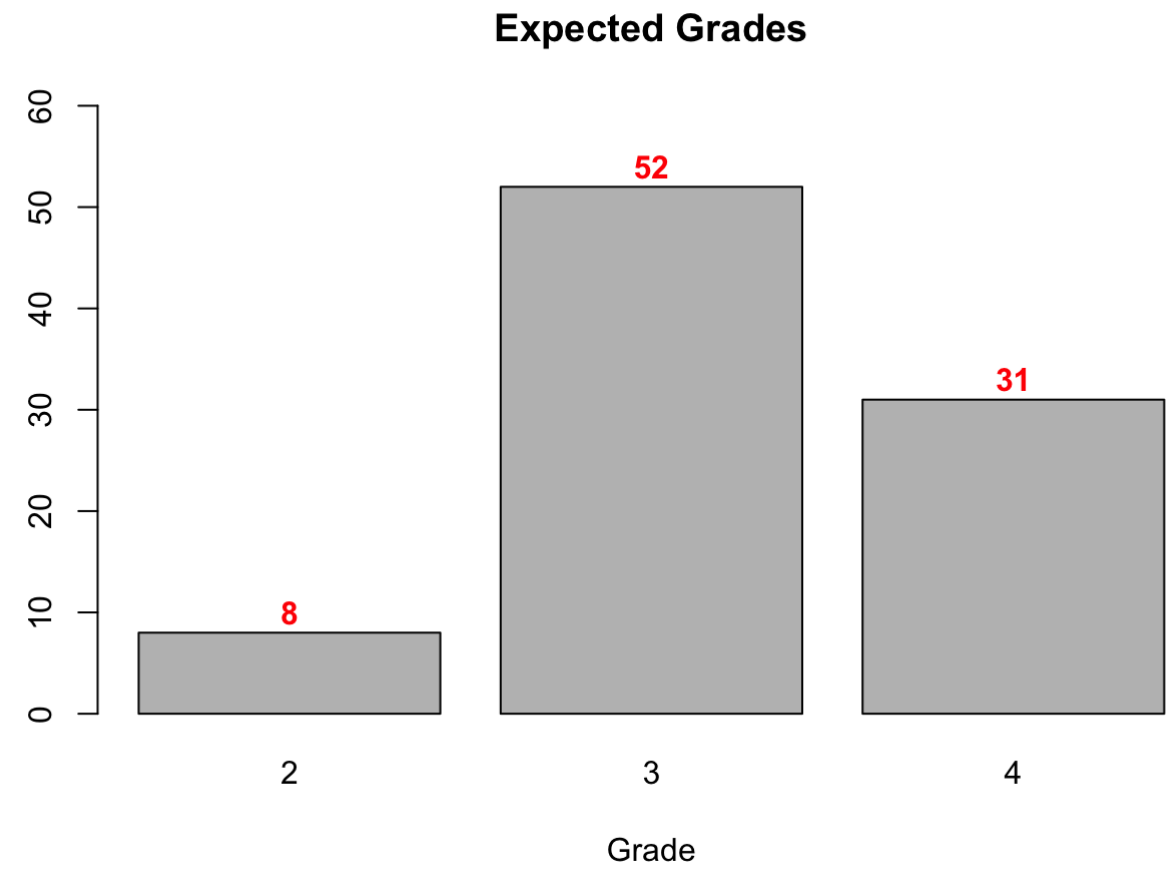
```
head(data)
```

```
##   time like where freq busy educ sex age home math work own cdrom email grade
## 1  2.0    3     3    2    0    1   0  19    1    0   10   1     0     1     4
## 2  0.0    3     3    3    0    0   0  18    1    1    0   1     1     1     2
## 3  0.0    3     1    3    0    0   1  19    1    0    0   1     0     1     3
## 4  0.5    3     3    3    0    1   0  19    1    0    0   1     0     1     3
## 5  0.0    3     3    4    0    1   0  19    1    1    0   0     0     1     3
## 6  0.0    3     2    4    0    0   1  19    0    0   12   0     0     0     3
```

```
fdata <- read.table("videoMultiple.txt",header=TRUE)
```

```
##   action adv sim sport strategy relax coord challenge master bored other
## 1      0   0   0     0        1     1     0         1      1     0
## 2      0   1   0     0        1     0     0         0      0     1
## 3      1   0   0     1        1     1     0         0      0     0
## 4      0   0   0     0        1     0     0         1      0     0
## 5      0   0   0     0        1     1     0         1      1     0
## 6      1   0   0     1        1     1     0         0      1     1
##   graphic time frust lonely rules cost boring friends point other2
## 1       0    1     0      0     0    1      0       0     1
## 2       0    1     1      0     0    0      0       0     0
## 3       0    0     0      0     0    1      0       0     0
## 4       0    1     0      0     0    0      0       0     0
## 5       0    0     0      0     1    1      0       0     0
## 6       0    1     1      0     0    1      0       0     0
```

```
counts <- table(data$grade)
bp <-barplot(counts, main="Expected Grades", xlab="Grade", ylim=c(0,60))
text(bp, counts + 2, counts, font=2, col=2)
```



Advanced Analysis

```
kruskal.test(action~point,data=fdata)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  action by point
## Kruskal-Wallis chi-squared = 5.119, df = 1, p-value = 0.02366
```

SCENARIO 5

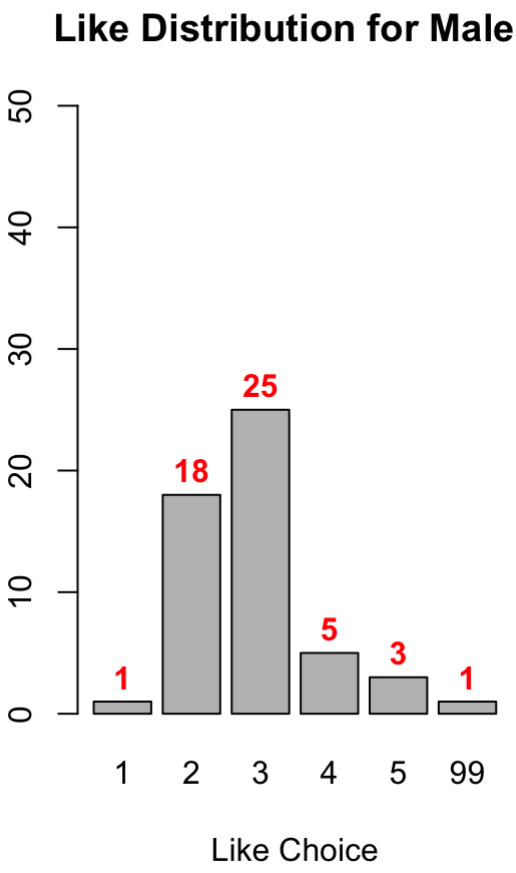
```
# data for males
male.ind <- which(data['sex'] == 1)
data.male <- data[male.ind,]
head(data.male)
```

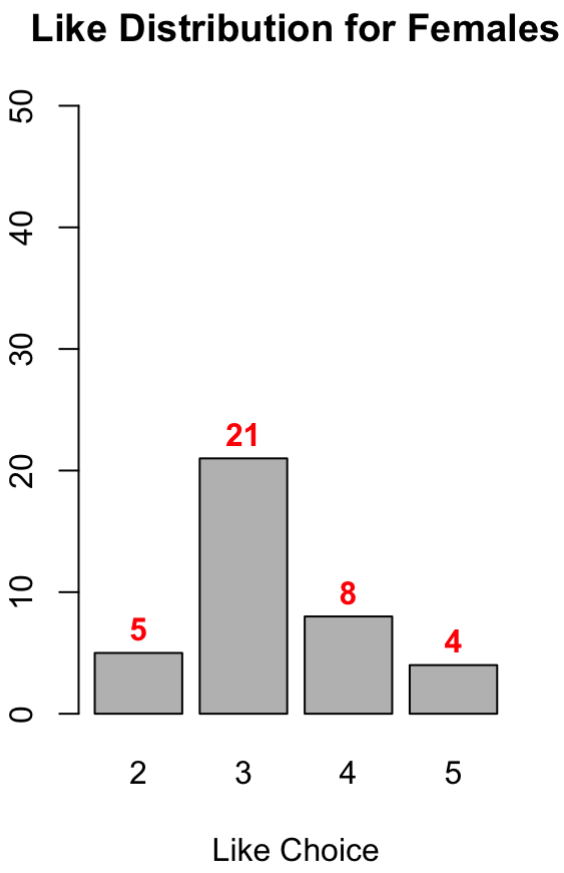
```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 3     0    3     1    3    0    0   1  19    1    0    0   1     0     1     3
## 6     0    3     2    4    0    0   1  19    0    0   12   0     0     0     3
## 7     0    4     3    4    0    0   1  20    1    1   10   1     0     1     3
## 9     2    3     2    1    1    1   1  19    0    0    0   0     0     0     4
## 10    0    3     3    4    0    1   1  19    1    1    0   1     0     1     4
## 16    0    5    99   99   99   99   1  19    1    0    0   1     0     1     3
```

```
# data for females
female.ind <- which(data['sex']==0)
data.female <- data[female.ind,]
head(data.female)
```

```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 1   2.0    3     3    2    0    1   0  19    1    0   10   1     0     1     4
## 2   0.0    3     3    3    0    0   0  18    1    1    0   1     1     1     2
## 4   0.5    3     3    3    0    1   0  19    1    0    0   1     0     1     3
## 5   0.0    3     3    4    0    1   0  19    1    1    0   0     0     1     3
## 8   0.0    3     3    4    0    0   0  19    1    0   13   0     0     1     3
## 11  0.0    3     1    4    0    0   0  20    1    0    0   1     0     0     3
```

```
par(mfrow = c(1:2))

counts <- table(data.male$like)
bp <-barplot(counts, main="Like Distribution for Males", xlab="Like Choice", ylim=c(0,50
))
text(bp, counts + 2, counts, font=2, col=2)

counts <- table(data.female$like)
bp <-barplot(counts, main="Like Distribution for Females", xlab="Like Choice", ylim=c(0,
50))
text(bp, counts + 2, counts, font=2, col=2)
```





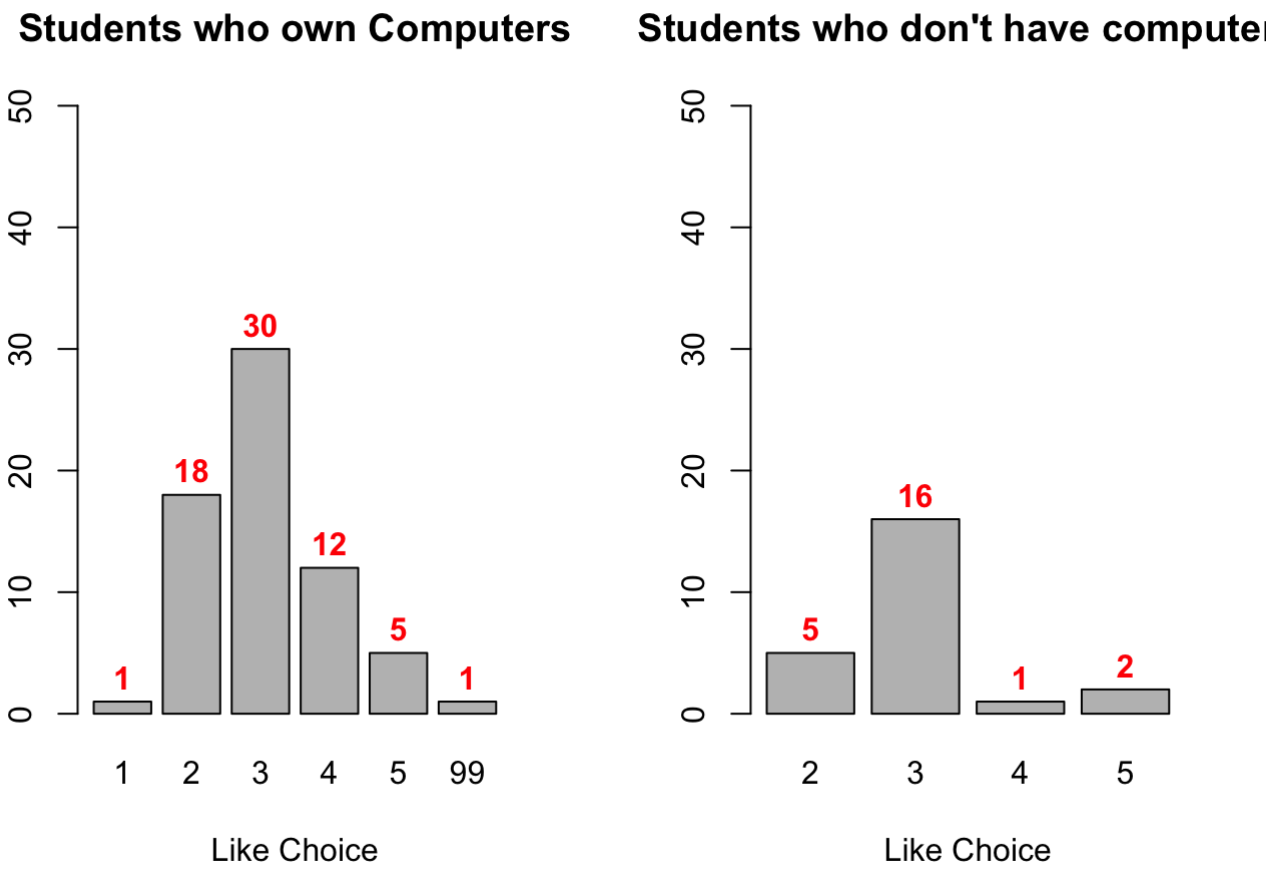

```
# data of students who do own a computer
owncom.ind <- which(data['own'] == 1)
data.owncom <- data[owncom.ind,]
head(data.owncom)
```

```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 1   2.0    3     3    2    0    1   0  19    1    0   10   1     0     1     4
## 2   0.0    3     3    3    0    0   0  18    1    1    0   1     1     1     2
## 3   0.0    3     1    3    0    0   1  19    1    0    0   1     0     1     3
## 4   0.5    3     3    3    0    1   0  19    1    0    0   1     0     1     3
## 7   0.0    4     3    4    0    0   1  20    1    1   10   1     0     1     3
## 10  0.0    3     3    4    0    1   1  19    1    1    0   1     0     1     4
```

```
# data of students who do not own a computer
nocom.ind <- which(data['own'] == 0)
data.nocom <- data[nocom.ind,]
head(data.nocom)
```

```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 5     0    3     3    4    0    1   0  19    1    1    0   0     0     1     3
## 6     0    3     2    4    0    0   1  19    0    0   12   0     0     0     3
## 8     0    3     3    4    0    0   0  19    1    0   13   0     0     1     3
## 9     2    3     2    1    1    1   1  19    0    0    0   0     0     0     4
## 13    0    2     4    1    0    1   0  19    1    1    0   0     0     1     4
## 14    3    3     3    2    1    0   0  18    0    0    0   0     0     1     3
```

```
par(mfrow = c(1:2))

counts <- table(data.owncom$like)
bp <-barplot(counts, main="Students who own Computers", xlab="Like Choice", ylim=c(0,50
))
text(bp, counts + 2, counts, font=2, col=2)

counts <- table(data.nocom$like)
bp <-barplot(counts, main="Students who don't have computers", xlab="Like Choice", ylim=
c(0,50))
text(bp, counts + 2, counts, font=2, col=2)
```



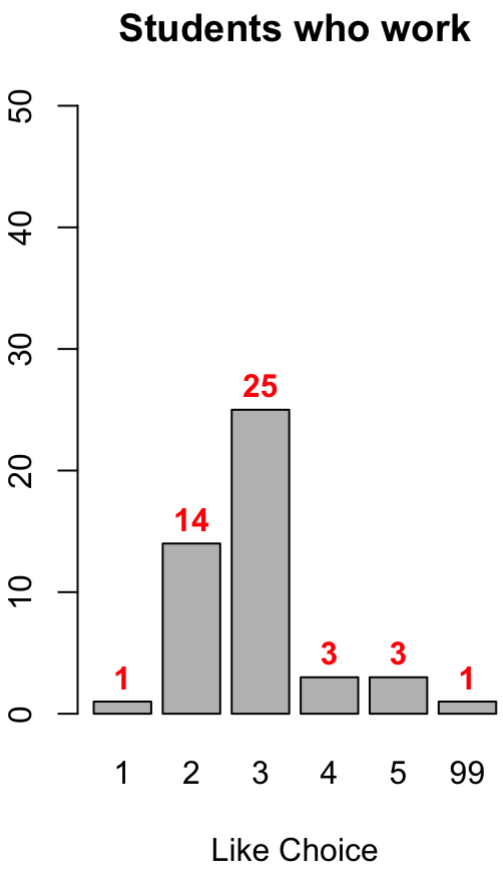

```
# data of students who do work
work.ind <- which(data['work'] > 0)
data.work <- data[work.ind,]
head(data.work)
```

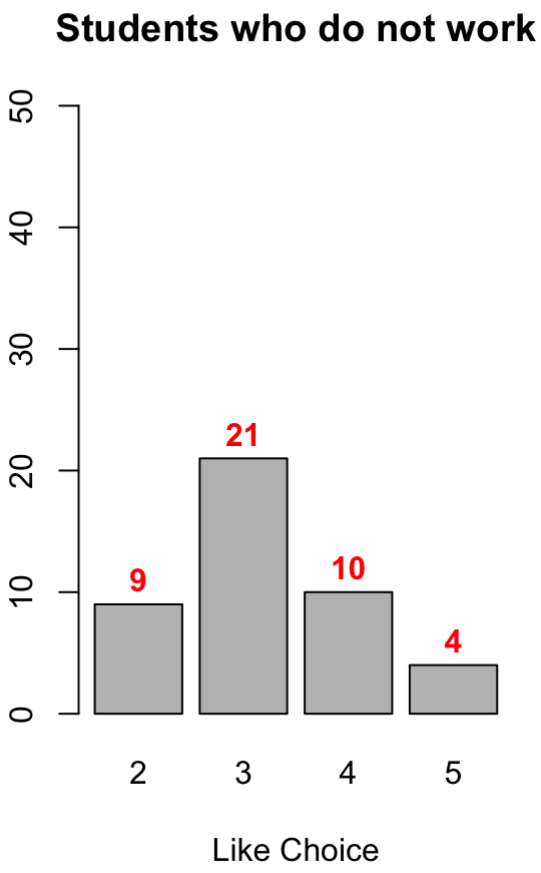
```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 1     2    3     3    2    0    1   0  19    1    0   10   1     0     1     4
## 6     0    3     2    4    0    0   1  19    0    0   12   0     0     0     3
## 7     0    4     3    4    0    0   1  20    1    1   10   1     0     1     3
## 8     0    3     3    4    0    0   0  19    1    0   13   0     0     1     3
## 15    1    3     5    2    0    1   0  18    1    1   14   1     0     1     3
## 17    0    3     3    4    0    1   1  21    1    0    2   1     0     1     4
```

```
# data of students that do not work
nowork.ind <- which(data['work'] == 0)
data.nowork <- data[nowork.ind,]
head(data.nowork)
```

```
##    time like where freq busy educ sex age home math work own cdrom email grade
## 2   0.0    3     3    3    0    0   0  18    1    1    0   1     1     1     2
## 3   0.0    3     1    3    0    0   1  19    1    0    0   1     0     1     3
## 4   0.5    3     3    3    0    1   0  19    1    0    0   1     0     1     3
## 5   0.0    3     3    4    0    1   0  19    1    1    0   0     0     1     3
## 9   2.0    3     2    1    1    1   1  19    0    0    0   0     0     0     4
## 10  0.0    3     3    4    0    1   1  19    1    1    0   1     0     1     4
```

```
par(mfrow = c(1:2))

counts <- table(data.work$like)
bp <-barplot(counts, main="Students who work", xlab="Like Choice", ylim=c(0,50))
text(bp, counts + 2, counts, font=2, col=2)

counts <- table(data.nowork$like)
bp <-barplot(counts, main="Students who do not work", xlab="Like Choice", ylim=c(0,50))
text(bp, counts + 2, counts, font=2, col=2)
```


Students who work
0
10
20
30
40
50
1
14
25
3
3
1
1
2
3
4
5
99
Like Choice


Students who do not work
0
10
20
30
40
50
9
21
10
4
2
3
4
5
Like Choice