

Homework: Data Visualization

This homework sheet will test your knowledge on visualization data using R.

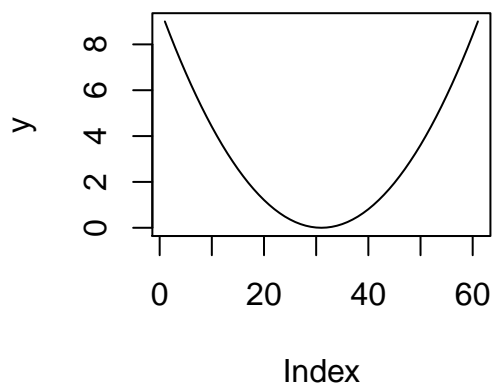
5

1

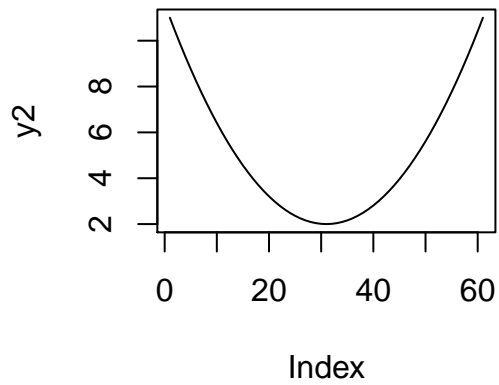
- a) Generate a sequence from -3 to 3 by 0.1 steps. Plot the function $y = x^2$ with x being the previously generated sequence. What does the function look like for $y = 2 + x^2$ and $y = 5 - x^2$?

Solution:

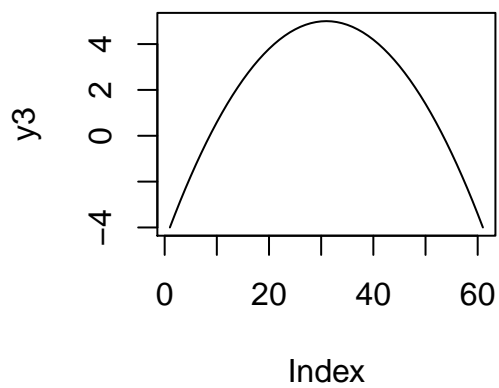
```
x <- (-30:30)/10  
y <- x^2  
plot(y, type = "l")
```



```
y2 <- 2 + x^2  
plot(y2, type = "l")
```



```
y3 <- 5 - x^2
plot(y3, type = "l")
```



1

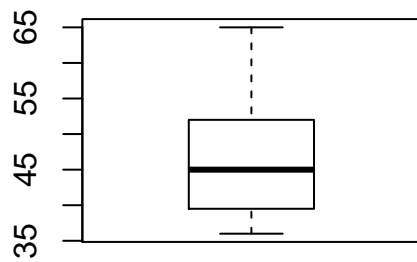
- b)** The following table gives the murders in New Jersey by date of the week:

Sun	Mon	Tue	Wed	Thu	Fri	Sat
53	42	51	45	36	37	65

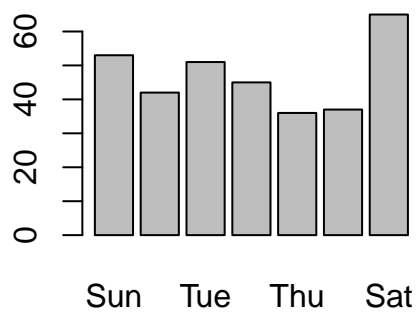
Calculate the relative and absolute frequencies, plot and interpret the results in a boxplot, a bar and a pie chart.

Solution:

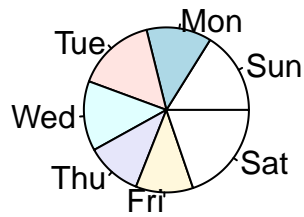
```
murder <- c(53, 42, 51, 45, 36, 37, 65)
names(murder) <- c("Sun", "Mon", "Tue", "Wed", "Thu", "Fri", "Sat")
freq.rel <- murder/sum(murder)
boxplot(murder)
```



```
barplot(murder)
```



```
pie(murder)
```



1

- c) The age of all U. S. presidents on inauguration day is given in the file `us-president.csv`. Study the data with a histogram, a boxplot and descriptive statistics.

Solution:

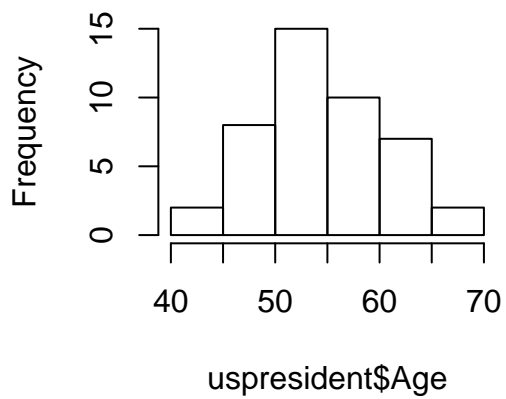
```
uspresident <- as.data.frame(read.csv("us-president.csv",
                                       header=TRUE, sep=","))

head(uspresident)

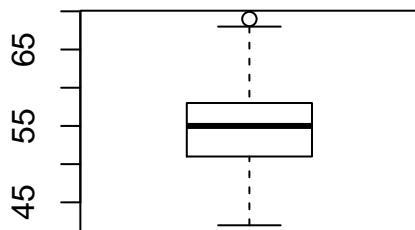
##      President Age
## 1 Washington  57
## 2   J.Adams   61
## 3 Jefferson  57
## 4   Madison  58
## 5   Monroe  58
## 6 J.Q.Adams  57

hist(uspresident$Age)
```

Histogram of uspresident\$Ag



```
boxplot(uspresident$Age)
```



```
attach(uspresident)
```

```
summary(Age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      42.0   51.0   55.0   54.9   58.0   69.0
```

```
mean(Age)
```

```
## [1] 54.89
```

```
sd(Age)
```

```
## [1] 6.218
```

```
var(Age)

## [1] 38.66

IQR(Age)

## [1] 7

range(Age)

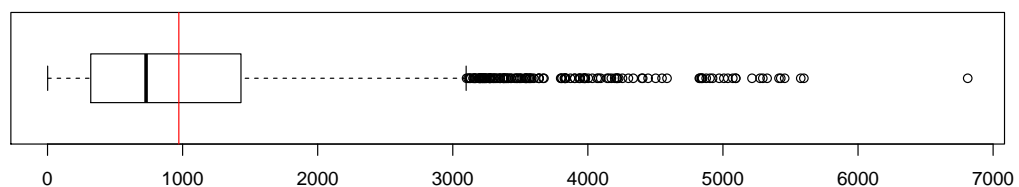
## [1] 42 69
```

1

- d) In the file `court_decisions.csv`, the duration of proceedings (column: `durationrev`) in Austria's Higher Administrative Court is given. Do the proceedings take longer now than 20 years ago when their average was 1 year and 3 month?

Solution:

```
decisions <- as.data.frame(read.csv("court_decisions.csv",
                                   header=TRUE, sep=","))
d <- decisions[decisions$durationrev != -9999, ]
boxplot(d$durationrev, horizontal=TRUE)
abline(v=972, col="red")
```



```
mean(d$durationrev)

## [1] 1016
```

1

- e) Describe the distribution of strokes needed in the first two rounds of the US-Masters in August 2009 (the columns `R1` and `R2` in the file `augusta.csv` store the results of the first two rounds).

Solution:

```
augusta <- as.data.frame(read.csv("augusta2009.csv",
                                header=TRUE, sep=","))

head(augusta)

##      Pos      Player Tot Thru R1 R2 R3 R4 Strokes
## 1    W   Angel Cabrera  -12   F 68 68 69 71     276
## 2   T2   Chad Campbell  -12   F 65 70 72 69     276
## 3   T2    Kenny Perry  -12   F 68 67 70 71     276
## 4    4 Shingo Katayama  -10   F 67 73 70 68     278
## 5    5  Phil Mickelson   -9   F 73 68 71 67     279
## 6   T6   Steve Flesch   -8   F 71 74 68 67     280

summary(augusta$R1)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      65.0    70.0    72.0    72.2    74.0    80.0

sd(augusta$R1)

## [1] 3.166

var(augusta$R1)

## [1] 10.02

range(augusta$R1)

## [1] 65 80

IQR(augusta$R1)

## [1] 4

summary(augusta$R2)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      65.0    71.0    74.0    73.7    76.0    84.0

sd(augusta$R2)

## [1] 3.683

var(augusta$R2)

## [1] 13.56

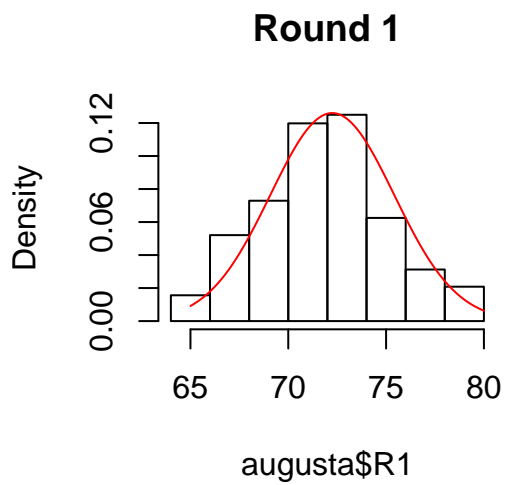
range(augusta$R2)

## [1] 65 84

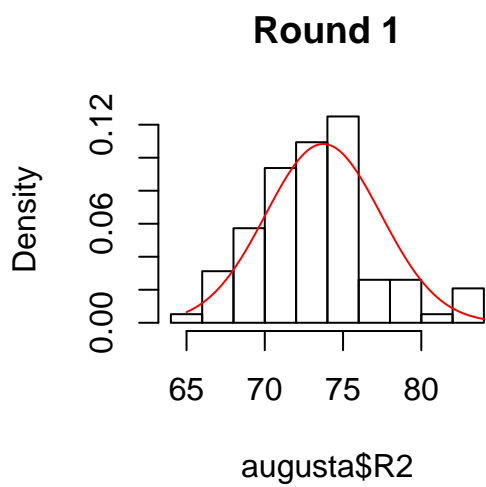
IQR(augusta$R2)

## [1] 5

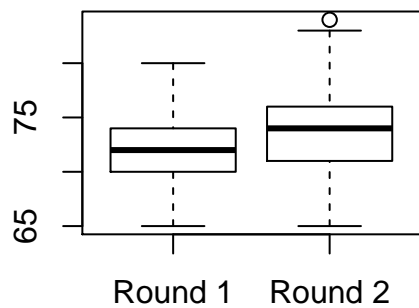
hist(augusta$R1, freq = FALSE, main="Round 1")
xx <- seq(min(augusta$R1), max(augusta$R1), 0.01)
lines(xx, dnorm(xx, mean=mean(augusta$R1), sd=sd(augusta$R1)), col="red")
```



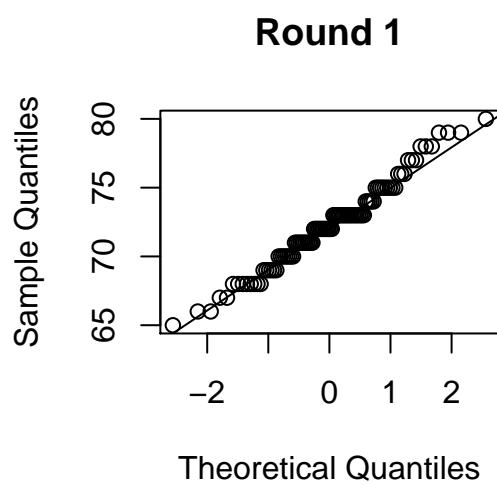
```
hist(augusta$R2, freq = FALSE, main="Round 1")
xx <- seq(min(augusta$R2), max(augusta$R2), 0.01)
lines(xx, dnorm(xx, mean=mean(augusta$R2), sd=sd(augusta$R2)), col="red")
```



```
boxplot(augusta$R1, augusta$R2, names=c("Round 1", "Round 2"))
```

```
qqnorm(augusta$R1, main = "Round 1")  
qqline(augusta$R1)
```



```
qqnorm(augusta$R2, main = "Round 2")  
qqline(augusta$R2)
```

