STAT 515 hw 1

Import data in R, basics of sets, basic probability

Download the free version of Rstudio [here](). Attach a sheet with the R plots and R code printed on it. You may write out your other answers by hand if you want. Just try to make it easy to grade!

1. Download the survey data from the first day of class from the course website. Then read it into R with this command (you will have to change the path to the file depending on where you save it):

```
data <- read.csv(file = "01_survey_results_fa_2023.csv")
```

(a) Use the command

```
table(data$bm, data$ie)
```

to make a table breaking down the beach/mountains and extrovert/introvert responses of the class. Give the table with row and column totals, as below.

	extrovert	introvert	total
beach			
mountains			
total			

```
data <- read.csv(file = "01_survey_results_fa_2023.csv")
table(data$bm, data$ie)
##
##      e  i
##   b 10 11
##   m  6  9
```

The table should be

	extrovert	introvert	total
beach	10	11	21
mountains	6	9	15
total	16	20	36

(b) Let $X_1, \dots, X_n$ represent the key weight measurements. Use the R functions `mean()`, `sd()`, and `summary()` to collect some summary statistics. Note that one value is `<NA>`, which means missing. To compute the mean ignoring the missing value use the command

```
mean(data$weight_keys, na.rm = TRUE)
```

You can add the same `na.rm = TRUE` option to the other functions. Report the following:

i. The sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$.

ii. The sample standard deviation $S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$

iii. The minimum, maximum, and median of $X_1, \dots, X_n$, as well as the number of missing values.

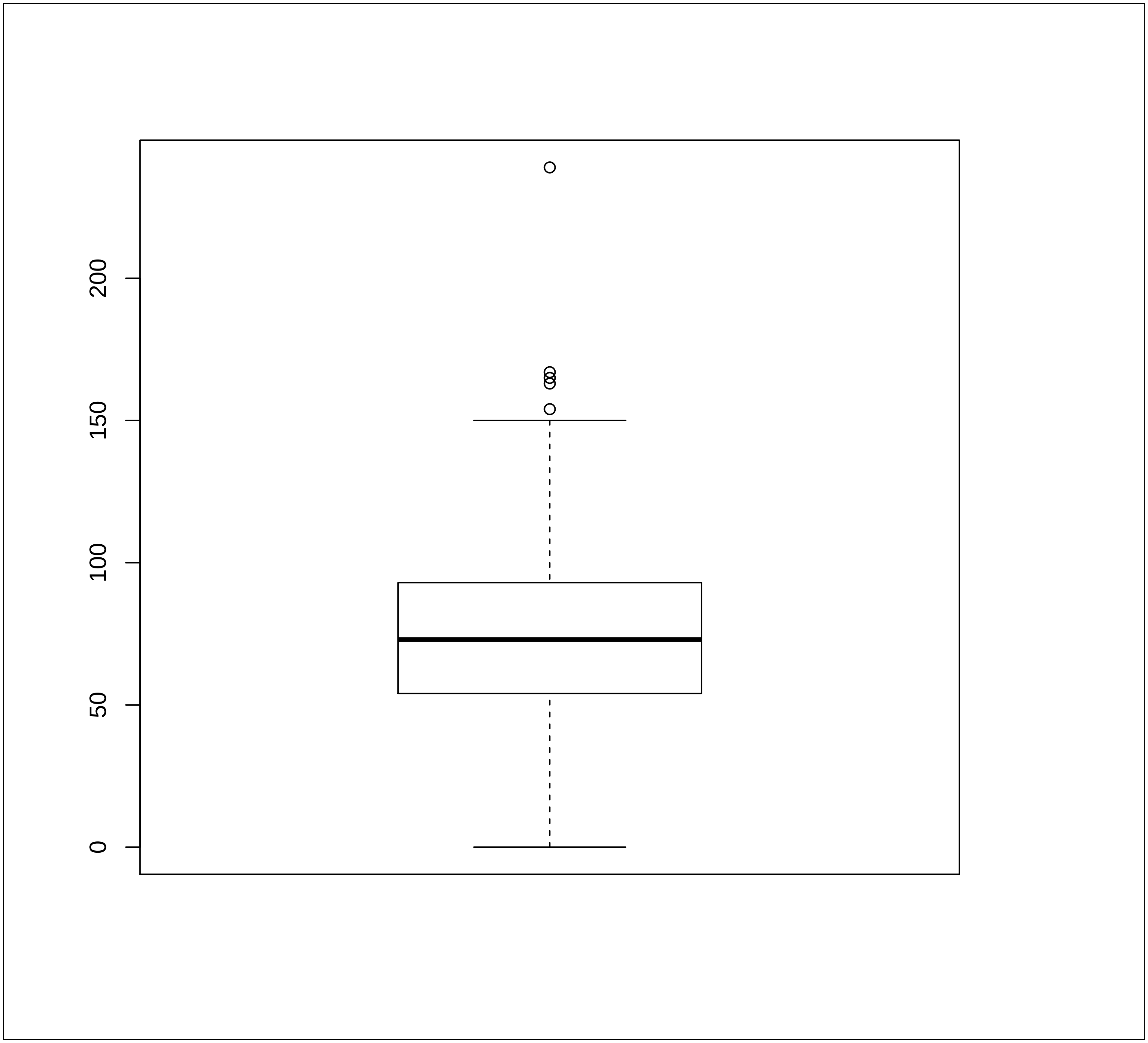
We have

```r
mean(data$weight_keys, na.rm = TRUE)
## [1] 80.13514
sd(data$weight_keys, na.rm = TRUE)
## [1] 51.75431
summary(data$weight_keys, na.rm = TRUE)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   54.00   73.00   80.14   93.00  239.00       1
```

(c) For the key weights, make the following plots:

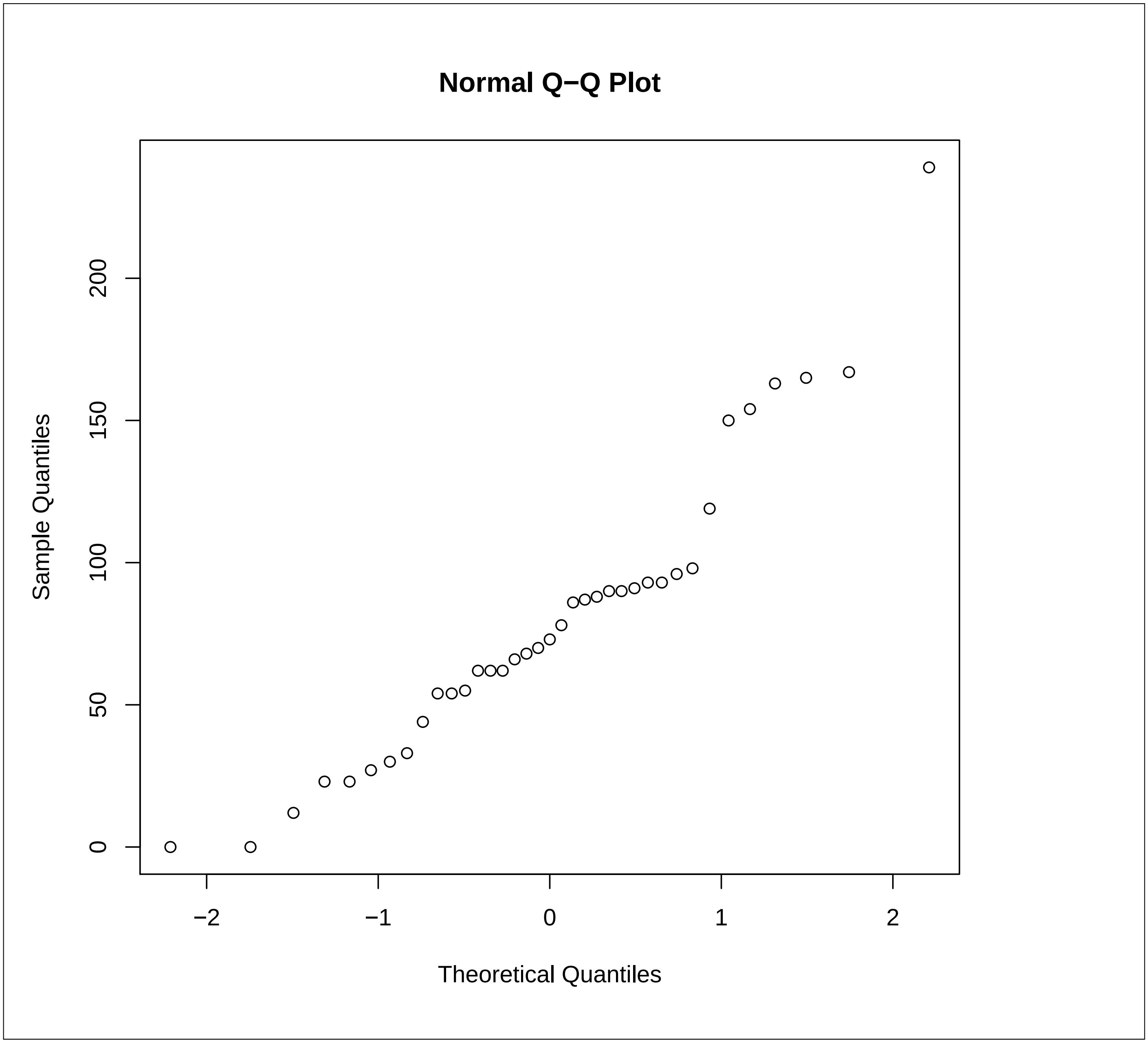
i. A boxplot using the `boxplot()` function. Ignore any missing values.

```r
boxplot(data$weight_keys)
```



ii. A plot called a Normal quantile-quantile plot using `qqnorm()` function. We will discuss later in the semester how to interpret this plot.

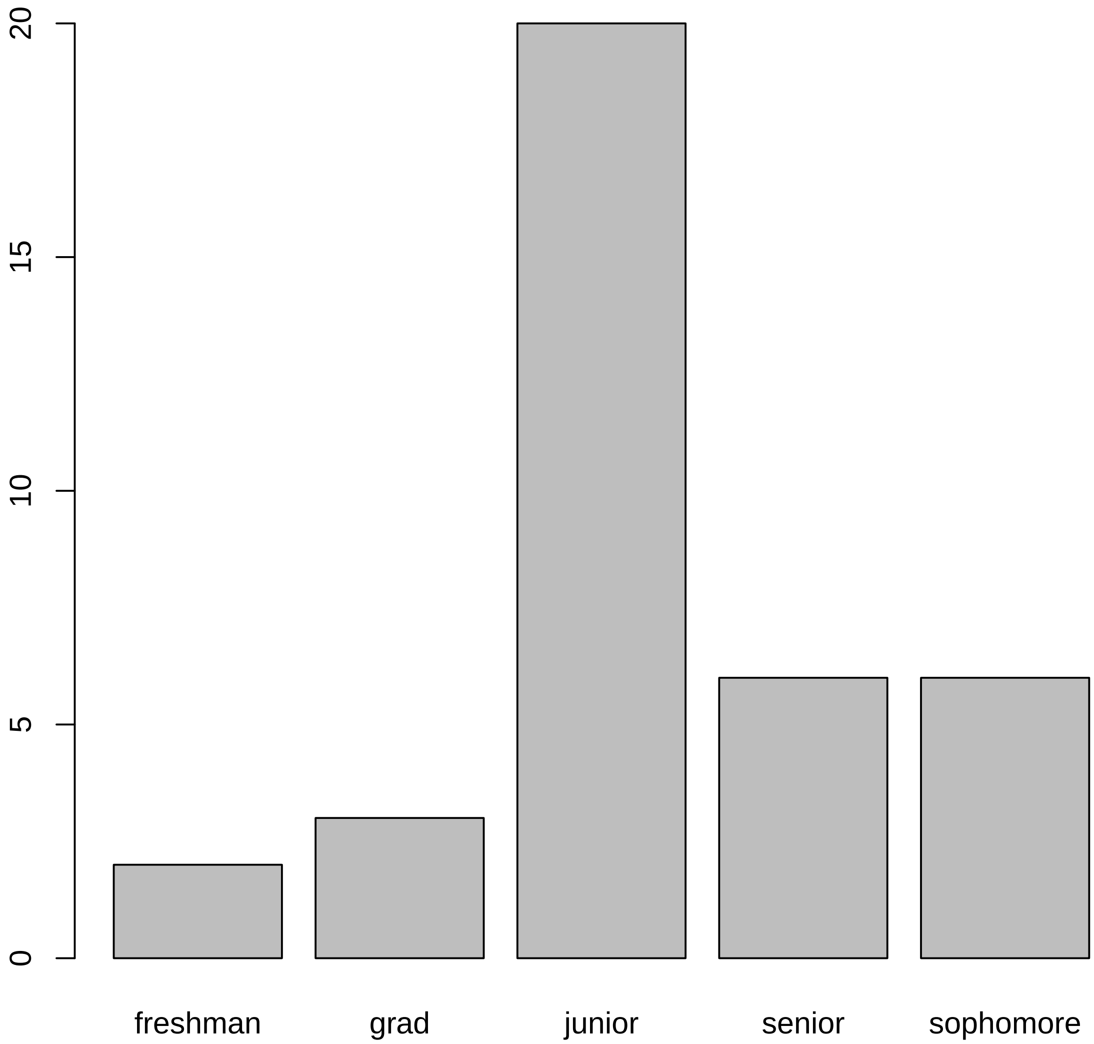
```
qqnorm(data$weight_keys)
```



(d) Make a barplot showing the number of students in each year: freshman, sophomore, etc. Use the command

```
barplot(table(data$year))
```

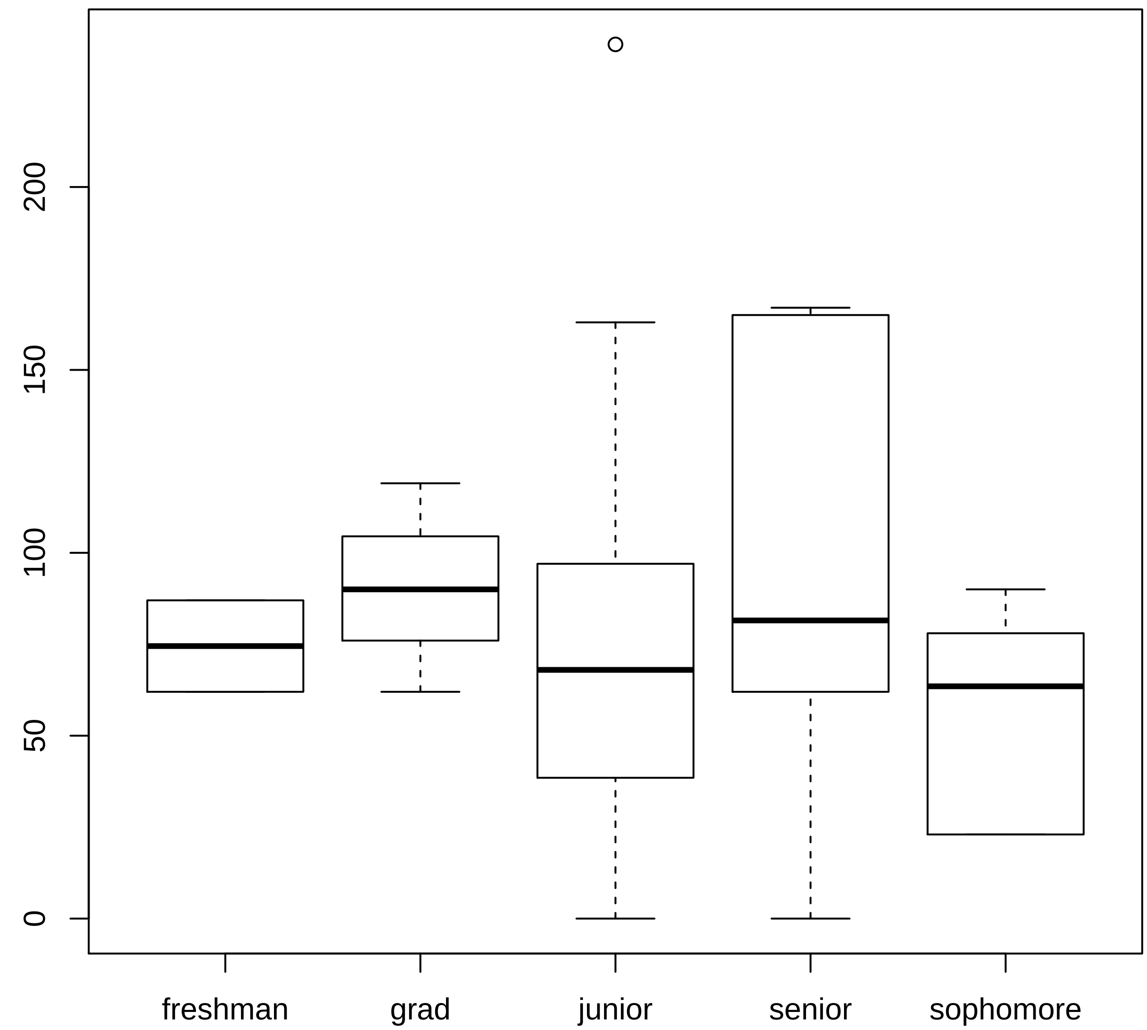
```
barplot(table(data$year))
```



(e) Make boxplots comparing the weights of keys versus the year of students. Use the command

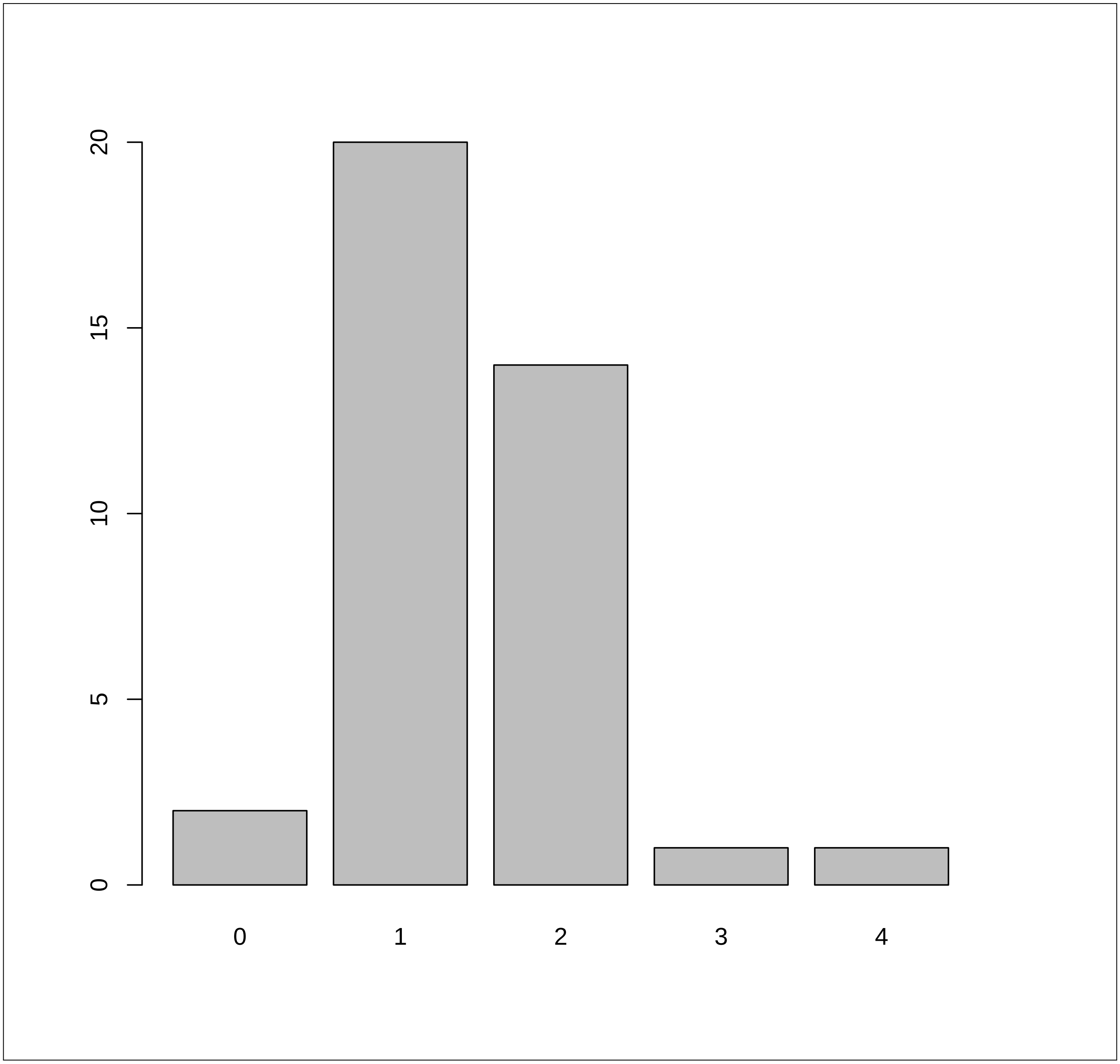
```
boxplot(data$weight_keys~data$year)
```

```
boxplot(data$weight_keys~data$year)
```



(f) Make a bar plot for the number of siblings of the students in the class.

```
barplot(table(data$sibs))
```



(g) Give the top three apps and the number of students using each.

```r
table(data$app)
##
##            CNN              X     blackboard clash_of_clans     delta_chat
##              1              3              1              1              1
##        discord  google_sheets      instagram       messages        netflix
##              1              1              8              1              1
##      pinterest         safari       snapchat        spotify         tiktok
##              1              1              5              3              2
##    yahoo_email        youtube
##              1              4
```

The above shows that 8 use Instagram, 5 use Snapchat, and 4 use YouTube a lot.

2. Consider rolling two dice and let

$$\begin{aligned} A &= \text{both rolls are at least 3} \\ B &= \text{both rolls are 3 or less} \\ C &= \text{the sum of the rolls is 10 or more} \\ D &= \text{the absolute value of the difference between the rolls is at most 1.} \end{aligned}$$

Give the following probabilities:

(a) $P(A)$

The possible outcomes are

$$\mathcal{S} = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\},$$

so we obtain $P(A) = 16/36$.

(b) $P(B)$

We have $P(B) = 9/36$.

(c) $P(C)$

We have $P(C) = 6/36$.

(d) $P(D)$

We have $P(D) = 16/36$.

(e) $P(A \cup B)$

We have $P(A \cup B) = 24/36 = 2/3$.

(f) $P(A \cap B)$

We have $P(A \cap B) = 1/36$.

(g) $P(A \cap B^c)$

We have $P(A \cap B^c) = 15/36$.

(h) $P((A \cap B)^c)$

We have $P((A \cap B)^c) = 35/36$.

(i) $P(A^c \cup B^c)$

We have $P(A^c \cup B^c) = 35/36$, by De Morgan's Laws.

(j) $P((A \cup B)^c)$

We have $P((A \cup B)^c) = 12/36$.

(k) $P(A^c \cap B^c)$

We have $P(A^c \cap B^c) = 12/36$, by De Morgan's Laws.

(l) $P(C \cap D)$

We have $P(C \cap D) = 4/36$.

(m) $P(C \cup D^c)$

We have $P(C \cap D^c) = 24/36 = 2/3$.

Hint: Begin by listing all possible outcomes of rolling two dice, i.e. the sample space.

3. Consider a bag of marbles, 19 of which are green, 25 of which are blue, and 6 of which are red. Moreover, suppose 9 of the green marbles are opaque, 5 of the blue marbles are opaque, and 3 of the red marbles are opaque, and the rest of the marbles are transparent.

(a) Suppose you draw one marble from the bag. Give the probability that you draw

i. a red marble.

6/50

ii. a transparent green marble.

10/50

iii. an opaque marble.

17/50

iv. a marble that is either blue or opaque or both.

37/50

(b) Suppose you remove all the opaque marbles from the bag and then draw one marble. Give the probability that you draw

i. a green marble.

10/33

ii. a red or a blue marble.

23/33

4. Suppose you draw 1 athlete at random from a group of 100 athletes such that: 30 swim; 44 run; 9 swim and run; 5 swim, bike, and run; 11 swim and bike; 10 bike and run but do not swim; and 35 only bike. Let S, B, and R denote the events that the athlete you draw swims, bikes, and runs, respectively. Give the following probabilities:

(a) $P(S \cup R)$

We have $P(S \cup R) = 30/100 + 44/100 - 9/100 = 65/100$.

(b) $P(S \cap R^c)$

We have $P(S \cap R^c) = 30/100 - 9/100 = 21/100$.

(c) $P(B)$

We have $P(B) = 35/100 + 10/100 + 5/100 + (11 - 5)/100 = 56/100$.

(d) $P(S \cup B)$

We have $P(S \cup B) = 30/100 + 56/100 - 11/100 = 75/100$.

(e) $P((S \cap R) \cap B^c)$

We have $P((S \cap R) \cap B^c) = 4/100 = 1/25$.

(f) $P(S^c \cup R^c)$

We have $P(S^c \cup R^c) = P((S \cap R)^c) = 1 - P(S \cap R) = 1 - 9/100 = 91/100$.

(g) $P((R \cap B) \cup (R \cap B^c))$

We have $P((R \cap B) \cup (R \cap B^c)) = P(R) = 44/100$.