Q1

```
> #Q1
> days.fav <- favstats(~Days, data = quine)
> days.fav$mean
[1] 16.4589
> days.fav$median
[1] 11
> days.fav$sd
[1] 16.25322
```

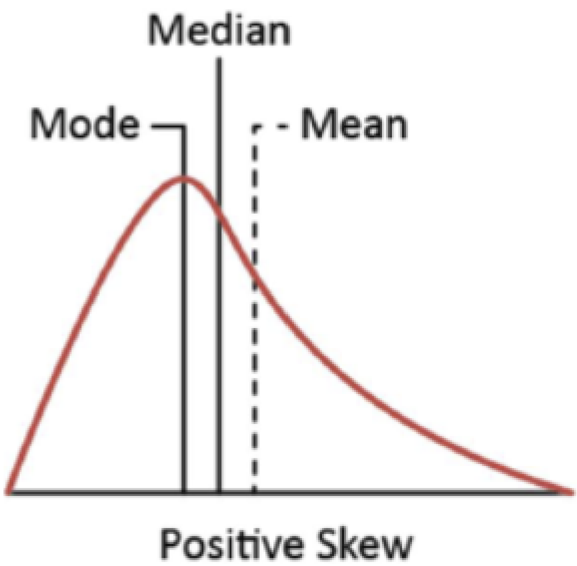
Q2

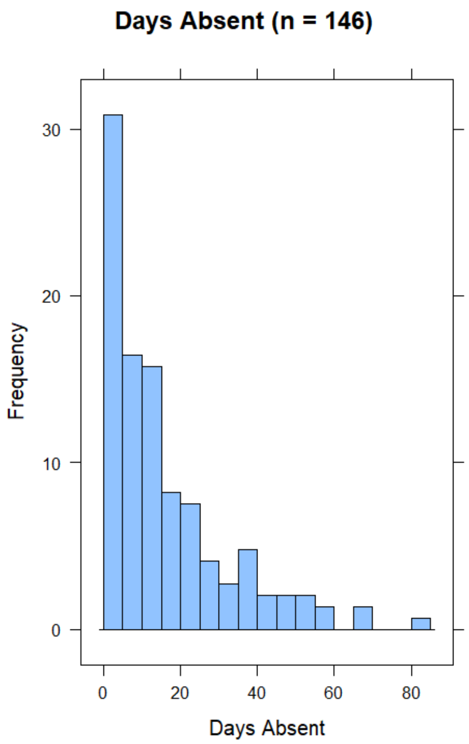
```
> #Q2
> #Sk = 3(bar X - Q2)/s
> days.barX <- mean(~Days, data = quine)
> days.Q2 <- median(~Days, data = quine)
> days.s <- sd(~Days, data = quine)
>
> days.Sk <- (3*(days.barX - days.Q2))/days.s
> days.Sk
[1] 1.007598
>
> Sk.direction <- if(days.Sk > 0.5){
+   "Skewed right"
+ }else if(days.Sk < -0.5){
+   "Skewed left"
+ }else{
+   "Symmetric"
+ }
> Sk.direction
[1] "Skewed right"
```

Q3

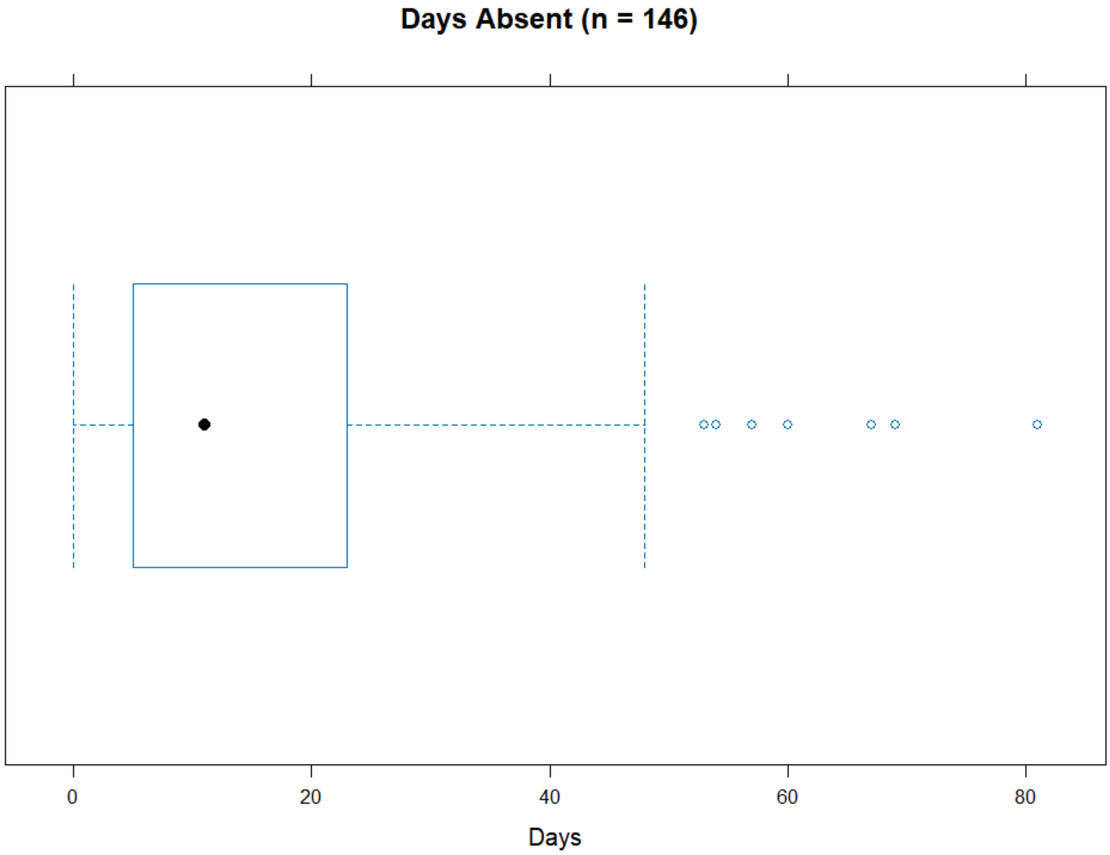
```
#Q3
range(quine$Days)
upper.limit <- seq(0, 85, by = 5)
histogram(
  ~Days,
  data = quine,
  breaks = upper.limit,
  xlab = "Days Absent",
  ylab = "Frequency",
  main = paste0(
          "Days Absent (n = ",
          sum(!is.na(quine$Days)),
          ")"
        )
)
```

The histogram really matches the Sk value, because it looks just like the textbook's right-skewed example (maybe)





Q4




```
#Q4
bwplot(
  ~Days,
  data = quine,
  xlab = "Days",
  main = paste0(
    "Days Absent (n = ",
    sum(!is.na(quine$Days)),
    ")"
  )
)
```

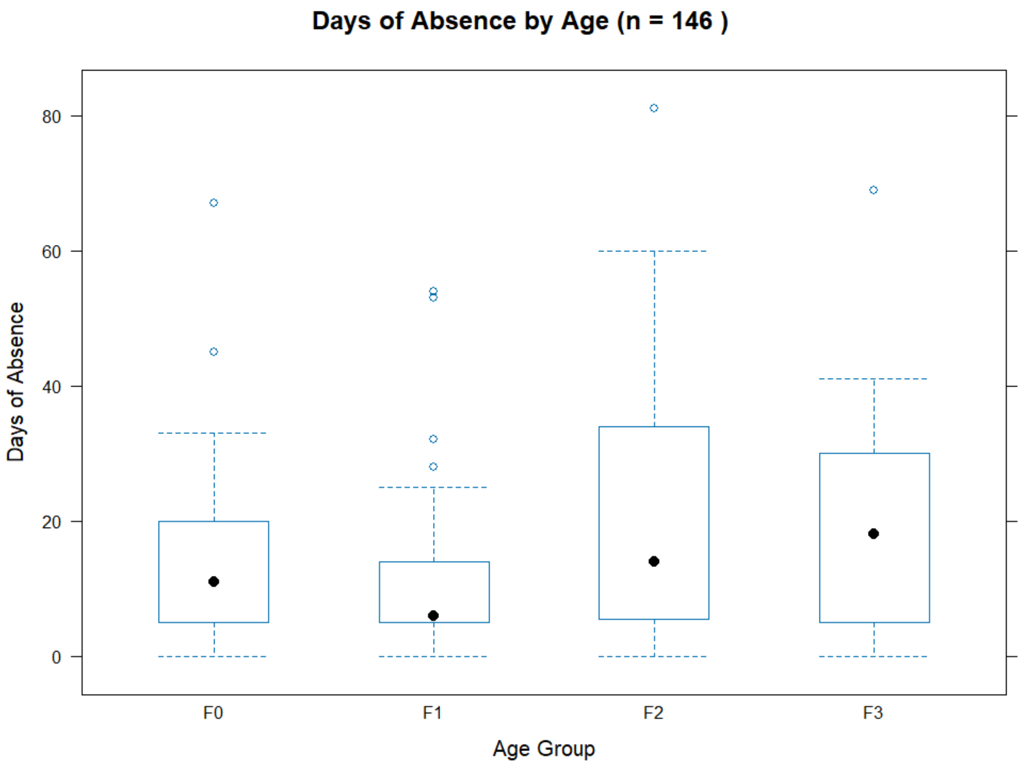
Q5

```
paste0("The middle 50% of students were absent between ",
       days.fav$Q1,
       " and ",
       days.fav$Q3,
       " days."
      )
```

Q6

```
#Q6
#lf = Q1 - 1.5 * IQR
#uf = Q3 + 1.5 * IQR
days.IQR <- days.fav$Q3 - days.fav$Q1
days.lower.fence <- days.fav$Q1 - 1.5 * days.IQR
days.upper.fence <- days.fav$Q3 + 1.5 * days.IQR
sum(quine$Days < days.lower.fence | quine$Days > days.upper.fence)
```

Q7




```
#Q7
bwplot(Days ~ Age, data = quine,
       xlab = "Age Group",
       ylab = "Days Absent",
       main = paste(
              "Days Absent by Age (n =",
              sum(!is.na(quine$Days)),
              ")"
              )
       )

favstats(Days ~ Age, data = quine)
```

```
> favstats(Days ~ Age, data = quine)
  Age min   Q1 median Q3 max     mean       sd  n missing
1  F0   0 5.00     11 20  67 14.85185 14.79528 27       0
2  F1   0 5.00      6 14  54 11.15217 11.64086 46       0
3  F2   0 5.75     14 33  81 21.05000 20.13665 40       0
4  F3   0 5.00     18 30  69 19.60606 15.97447 33       0
```

1. The F1 group had the fewest absences, since both its median and mean were very low
2. The F1 group showed the most consistent attendance, as it had the smallest standard deviation and the smallest IQR

3. The F2 group showed the most dispersed attendance pattern, as it had the largest standard deviation and the largest IQR

Q8

```
> #8
> favstats(Days ~ Age + Sex, data = quine)
  Age.Sex min    Q1 median    Q3 max     mean       sd  n missing
1    F0.F   3 10.25   15.5 24.75  45 18.70000 13.30873 10       0
2    F1.F   0  5.00    7.0 15.50  54 12.96875 13.17986 32       0
3    F2.F   0  2.50   10.0 17.00  81 18.42105 23.10199 19       0
4    F3.F   0  3.00   10.0 21.50  40 14.00000 12.81926 19       0
5    F0.M   0  5.00   11.0 14.00  67 12.58824 15.53648 17       0
6    F1.M   0  4.25    5.5 10.00  17  7.00000  5.30602 14       0
7    F2.M   0  8.00   17.0 36.00  57 23.42857 17.25854 21       0
8    F3.M   0 16.25   27.5 35.50  69 27.21429 17.09781 14       0
```

The male F1 group is the most consistent, because their standard deviation is significantly lower than that of the other groups. Similarly, the female F2 group is the least consistent

Q9

```
> #9
> q.Lrn <- quantile(Days ~ Lrn, probs = seq(0.2, 0.8, 0.2) , data = quine)
> q.Sex <- quantile(Days ~ Sex, probs = seq(0.2, 0.8, 0.2) , data = quine)
> q.Eth <- quantile(Days ~ Eth, probs = seq(0.2, 0.8, 0.2) , data = quine)
>
> q.Lrn
  Lrn 20% 40%  60% 80%
1  AL   5 8.8 16.0  27
2  SL   5 6.0 13.2  28
> q.Sex
  Sex 20%  40% 60%  80%
1   F   5  6.6  13 23.2
2   M   5 10.0  16 30.0
> q.Eth
  Eth 20% 40%  60%  80%
1   A   6  13 20.0 36.8
2   N   3   5 10.6 19.6
>
> q.Lrn.diff <- range(abs(q.Lrn[1, -1] - q.Lrn[2, -1]))
> q.Sex.diff <- range(abs(q.Sex[1, -1] - q.Sex[2, -1]))
> q.Eth.diff <- range(abs(q.Eth[1, -1] - q.Eth[2, -1]))
>
> q.Lrn.diff
[1] 0.0 2.8
> q.Sex.diff
[1] 0.0 6.8
> q.Eth.diff
[1]  3.0 17.2
```

It looks like Eth makes the greatest difference

Q10

```
> male.F0 <- filter(quine, Sex == "M" & Age == "F0")
> cbind(round(100*percent_rank(male.F0$Days), digits = 2))
        [,1]
 [1,]  12.50
 [2,]  50.00
 [3,]  75.00
 [4,]  25.00
 [5,]  25.00
 [6,]  68.75
 [7,]  87.50
 [8,]  93.75
 [9,]  37.50
[10,]  81.25
[11,] 100.00
[12,]   0.00
[13,]   0.00
[14,]  12.50
[15,]  43.75
[16,]  50.00
[17,]  62.50
```

Q11

```
> #Q11
> z_scores <- scale(quine$Days)
>
> pct_1sd <- mean(abs(z_scores) >= 1) * 100
> pct_2sd <- mean(abs(z_scores) >= 2) * 100
> pct_3sd <- mean(abs(z_scores) >= 3) * 100
>
> pct_1sd
[1] 21.91781
> pct_2sd
[1] 5.479452
> pct_3sd
[1] 2.054795
```

Q12

It satisfies Chebyshev's Theorem, since all the proportions outside ±2 and ±3 standard deviations from the mean are less than the theorem's upper bounds. However, the results do not satisfy the Empirical Rule, because the proportions outside ±1, ±2, and ±3 standard deviations differ significantly from the normal distribution values of 32%, 5%, and 0.3%. This indicates that the data distribution is not normal, but rather skewed and contains outliers