MA2611 Lab 1

Jacob Nguyen

2022-09-01

1. Create and store a vector of the data and answer the following questions:

```
x = c(2413, 20310, 12637, 2753, 14505, 14440, 2379, 447, 345, 4784, 13803, 12668,
1235, 1257, 1671, 4041, 4145, 535, 5270, 3360, 3489, 1979, 2302, 807, 1772, 12807,
5427, 13147, 6288, 1803, 13167, 5343, 6684, 3508, 1549, 4975, 11249, 3213, 811, 3560,
7244, 6643, 8571, 13534, 4395, 5729, 14417, 4863, 1951, 13809)
```

Vector of Highest Points (in ft) of Each US State:

```
x
```

```
##  [1]  2413 20310 12637  2753 14505 14440  2379   447   345  4784 13803 12668
## [13]  1235  1257  1671  4041  4145   535  5270  3360  3489  1979  2302   807
## [25]  1772 12807  5427 13147  6288  1803 13167  5343  6684  3508  1549  4975
## [37] 11249  3213   811  3560  7244  6643  8571 13534  4395  5729 14417  4863
## [49]  1951 13809
```

a. What is the height of the highest point in the US? And the lowest point in the US?

Height of Highest Point in the US:

```
max(x)
```

```
## [1] 20310
```

Height of Lowest Point in US:

```
min(x)
```

```
## [1] 345
```

b. Mt. Greylock, at 3489 ft, is the highest point in Massachusetts. Where does Massachusetts rank among US states for the highest point in the country?

```
sort(x, decreasing=TRUE)
```

```
##  [1] 20310 14505 14440 14417 13809 13803 13534 13167 13147 12807 12668 12637
## [13] 11249  8571  7244  6684  6643  6288  5729  5427  5343  5270  4975  4863
## [25]  4784  4395  4145  4041  3560  3508  3489  3360  3213  2753  2413  2379
## [37]  2302  1979  1951  1803  1772  1671  1549  1257  1235   811   807   535
## [49]   447   345
```

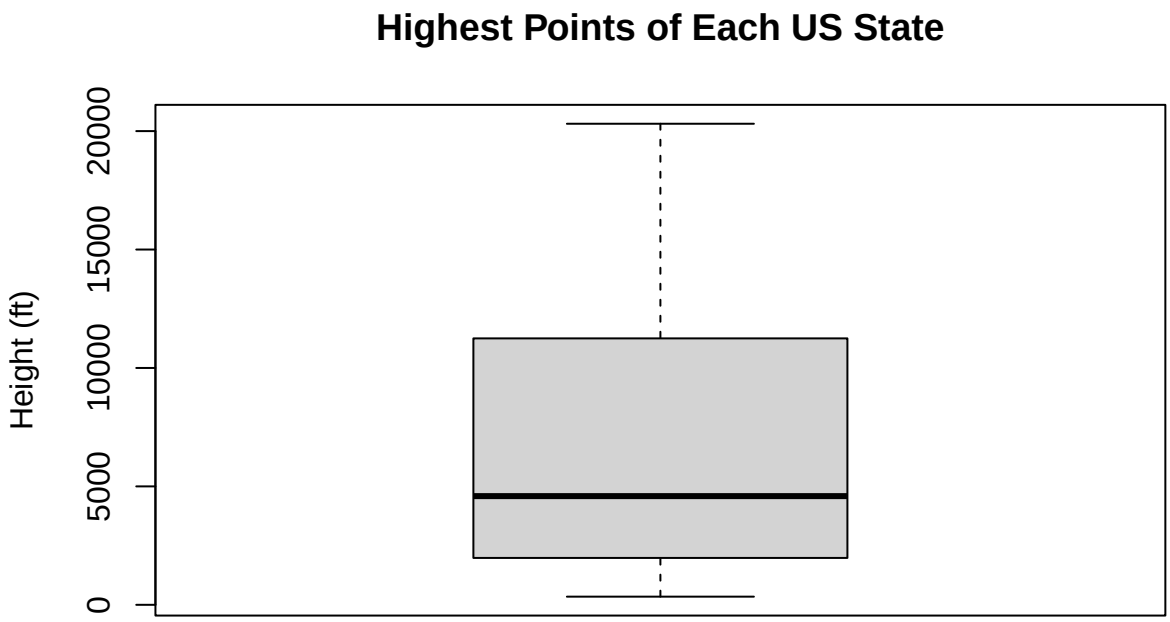
```
match(3489, x)
```

```
## [1] 21
```

Massachusetts ranks 21st among US States for the highest point in the country.

c. Create a boxplot of the data set. Add an appropriate label to the y-axis plus a title on the plot. Provide an interpretation of the boxplot. What does this tell you about the highest points in the US?

```
boxplot(x, main = "Highest Points of Each US State", ylab = "Height (ft)")
```



This data set has a 1st quartile value of 2060 ft, a median of 4590 ft, and a 3rd quartile value of 10580 ft. Given this, the IQR = 8520 ft. It has a maximum value of 20310 ft, and a minimum of 345 ft. This boxplot tells us that the highest points in the US are within the upper adjacent value, and lower adjacent value meaning that there are no outliers. We can conclude that this data has a positive skew since the 3rd quartile to the median is greater than the 1st quartile to the median. A greater majority of the data points lay below the median since it is closer to the 1st quartile.

2. Execute the following steps in order:

a. Create and store a vector with a sequence of values from 5 to -11 that progresses in steps of 0.3 without typing out each individual element

```
y = seq(from=5,to=-11, by = -0.3)
y
```

```
##  [1]   5.0   4.7   4.4   4.1   3.8   3.5   3.2   2.9   2.6   2.3   2.0   1.7
## [13]   1.4   1.1   0.8   0.5   0.2  -0.1  -0.4  -0.7  -1.0  -1.3  -1.6  -1.9
## [25]  -2.2  -2.5  -2.8  -3.1  -3.4  -3.7  -4.0  -4.3  -4.6  -4.9  -5.2  -5.5
## [37]  -5.8  -6.1  -6.4  -6.7  -7.0  -7.3  -7.6  -7.9  -8.2  -8.5  -8.8  -9.1
## [49]  -9.4  -9.7 -10.0 -10.3 -10.6 -10.9
```

b. Overwrite the part (a) vector using the same sequence with the order reversed

```
y = sort(y,decreasing=FALSE)
y
```

```
##  [1] -10.9 -10.6 -10.3 -10.0  -9.7  -9.4  -9.1  -8.8  -8.5  -8.2  -7.9  -7.6
## [13]  -7.3  -7.0  -6.7  -6.4  -6.1  -5.8  -5.5  -5.2  -4.9  -4.6  -4.3  -4.0
## [25]  -3.7  -3.4  -3.1  -2.8  -2.5  -2.2  -1.9  -1.6  -1.3  -1.0  -0.7  -0.4
## [37]  -0.1   0.2   0.5   0.8   1.1   1.4   1.7   2.0   2.3   2.6   2.9   3.2
## [49]   3.5   3.8   4.1   4.4   4.7   5.0
```

c. Create and store an index vector containing the first, middle, and last elements from the vector in part (b)

```
index = c(y[1],y[length(y)/2],y[length(y)])
index
```

```
## [1] -10.9  -3.1   5.0
```

3. Execute the following steps in order:

a. Create and store a vector that contains the following: - a sequence of integers from 6 to 12 - a threefold repetition of the value 5.3 - the number -3 - a twofold repetition of the numbers 1.2, 3.4, and 5.6

```
v = c(seq(from=6, to=12), rep(5.3, times=1, each=3), -3, rep(c(1.2, 3.4, 5.6), times=1, each=2))
v
```

```
##  [1]  6.0  7.0  8.0  9.0 10.0 11.0 12.0  5.3  5.3  5.3 -3.0  1.2  1.2  3.4  3.4
## [16]  5.6  5.6
```

b. Confirm the length of the vector in part (a) is 17

```
length(v)
```

```
## [1] 17
```

4. Create a data frame using data collected in your day-to-day life. It can be any data you like, but the data frame must contain at least 10 rows and 3 columns of data.

```
redsoxstats=data.frame(Name=c("Devers","Bogaerts","Verdugo","Martinez","Story","Arroyo",
"Pham","Refsnyder","Hernandez","Dalbec","Plawecki","McGuire"), Position=c("3B","SS","RF",
"DH","2B","UTIL","LF","RF","CF","1B","C", "C"),
AVG=(c(.288,.310,.287,.271,.233,.289,.282,.310,.219,.211,.216,.370)),
OBP=c(.344,.378,.332,.339,.299,.333,.325,.388,.277,.282,.282,.386),
OPS=c(.877,.839,.742,.767,.727,.760,.788,.879,.628,.644,.566,.849))
```

2022 Red Sox Stats (as of 9/2/22)

```
redsoxstats
```

```
##          Name Position   AVG   OBP   OPS
## 1     Devers       3B 0.288 0.344 0.877
## 2   Bogaerts       SS 0.310 0.378 0.839
## 3    Verdugo       RF 0.287 0.332 0.742
## 4   Martinez       DH 0.271 0.339 0.767
## 5      Story       2B 0.233 0.299 0.727
## 6     Arroyo     UTIL 0.289 0.333 0.760
## 7       Pham       LF 0.282 0.325 0.788
## 8  Refsnyder       RF 0.310 0.388 0.879
## 9  Hernandez       CF 0.219 0.277 0.628
## 10    Dalbec       1B 0.211 0.282 0.644
## 11  Plawecki        C 0.216 0.282 0.566
## 12   McGuire        C 0.370 0.386 0.849
```