



This is a graded discussion: 10 points possible

due -

Summarization: Let's draw a boxplot

6 21

As explained, one of the main ways to effectively visualize large 1D data is 'visually summarizing' the dataset by creating a box plot. It shows where the median is located, what is the inter-quartile range, what's the range where most data points are located, and finally outliers.

You may have been drawing many boxplots already. However, here I'd like to ask you to draw two boxplots based on the following two tables **by hand**. I think that's the best way to really understand how to draw boxplots. If you use a software, it takes care of, and hides, all the small details! So for each of these two tables, draw a boxplot by calculating all the location of the boxes and whiskers by hand. Report where the median, two ends of the box, and positions of the whiskers.

Feel free to share a photo of your boxplot but it's optional. Share what was tricky, if any, and how you should calculate the location of median, box, whiskers, etc. If you don't have any issues, then I'd say you **understand** boxplots!

14	-1
14	3
15	3
16	4
16	15
18	16
18	16
19	17
19	23
21	24
22	24
25	25
25	35
29	36
30	37
30	46

Search entries or author

Unread



✓ Subscribed

← Reply



Prem Amal (<https://iu.instructure.com/courses/2165942/users/6684842>)

Monday



Dataset 1

Minimum - 14

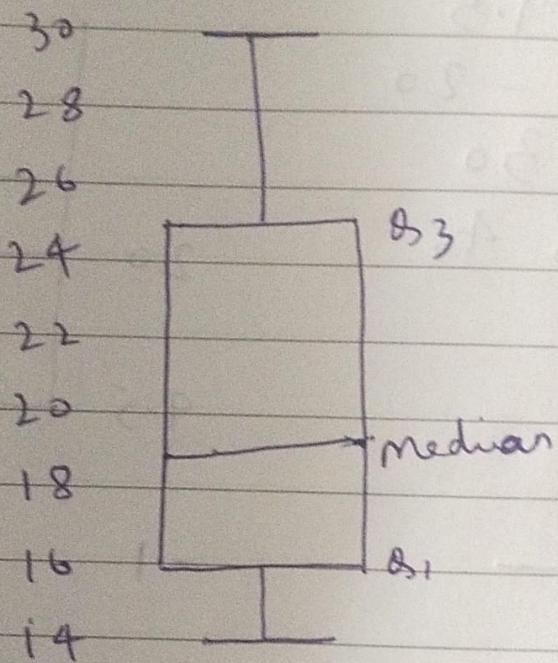
Q1 16

Median 19

Q3 25

Maximum 30

Box Plot for Dataset 1



Dataset 2

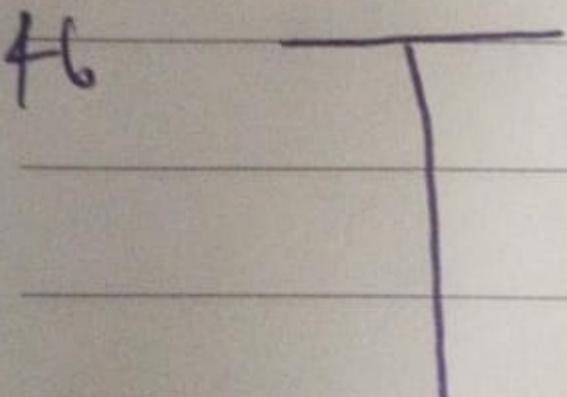
Minimum -1

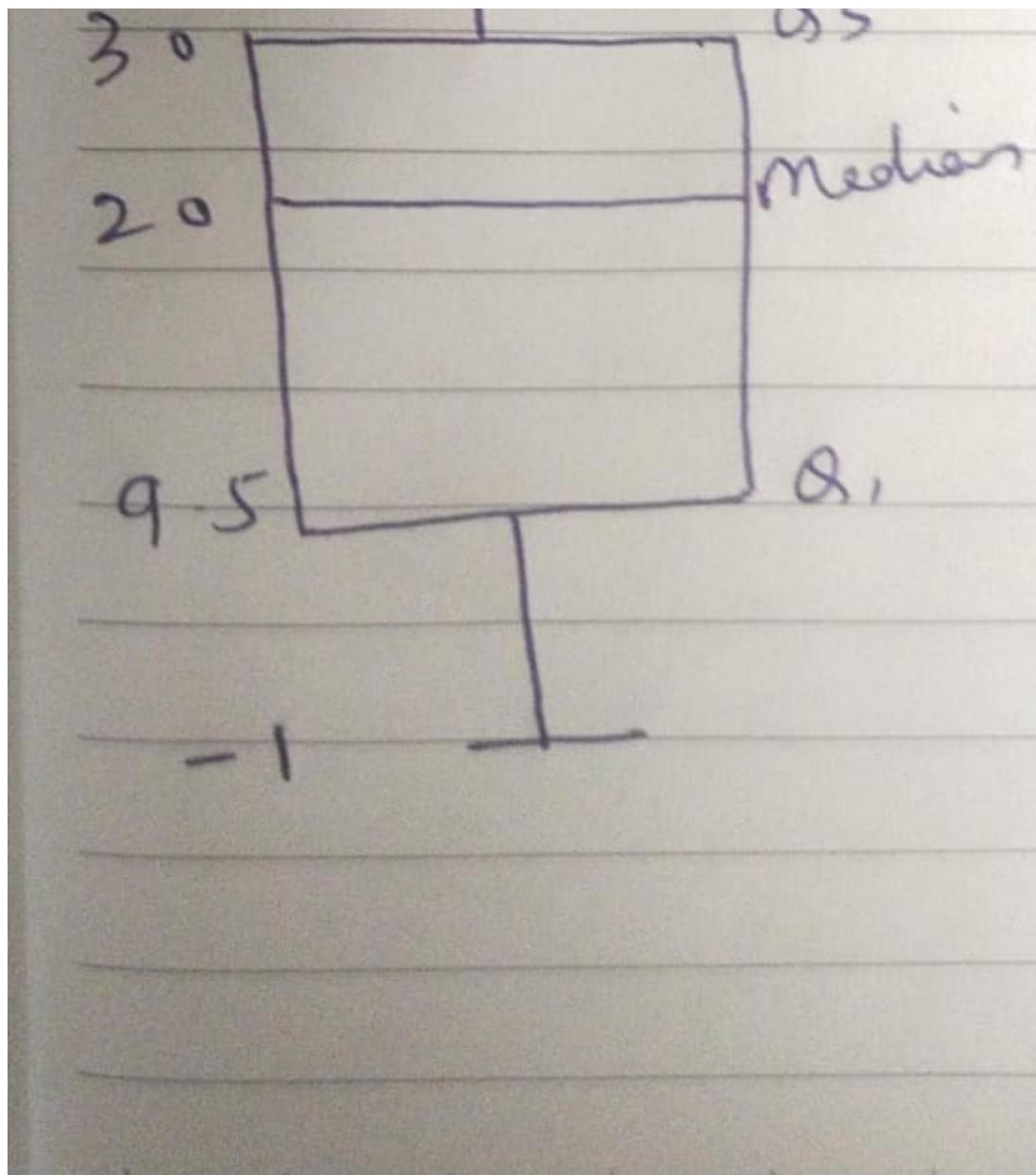
Q1 - 9.5

Median - 20

Q3 - 30

Maximum 46





Minimum is the minimum value of dataset. Box plot starts from there, the first horizontal line. Vertical line starts from there till the start of box which is the Q1. Then End of Box is Q3. The end of box plot is the maximum value of dataset. The middle line in the box plot is the median of the dataset.

Arrange the dataset in ascending order, from the smallest value to the largest value. If the dataset has an odd number of values, the median is the middle value. If the dataset has an even number of values, the median is the average of the two middle values. To find Q1 (the 25th percentile), count the number of data points below the median and calculate the median of that subset. To find Q3 (the 75th percentile), count the number of data points above the median and calculate the median of that subset. Q3 is the median of the values above the median.

https://

Dustin Cole (<https://iu.instructure.com/courses/2165942/users/6701715>)

Monday

3

It was not fun typing in the numbers! Sorting by hand would be even worse.

[Reply](#)

Thomas Jablenski (<https://iu.instructure.com/courses/2165942/users/6701599>)

Tuesday

5

There were no real issues while creating and getting the locations for the box plot. The median is the value in the middle of the dataset or if the dataset is even then it is the average between the two. To find the Q1, you take the 'center' value of everything less than the median and for Q3 you would do the same but to the values greater than the median. If the dataset contains an even amount of records you take the average of the 2 values closest to the 'center'. The whiskers are calculated by $Q1 - 1.5 * IQR$ and $Q3 + 1.5 * IQR$.

Edited by Thomas Jablenski (<https://iu.instructure.com/courses/2165942/users/6701599>) on Sep 19 at 1:49am

[Reply](#)

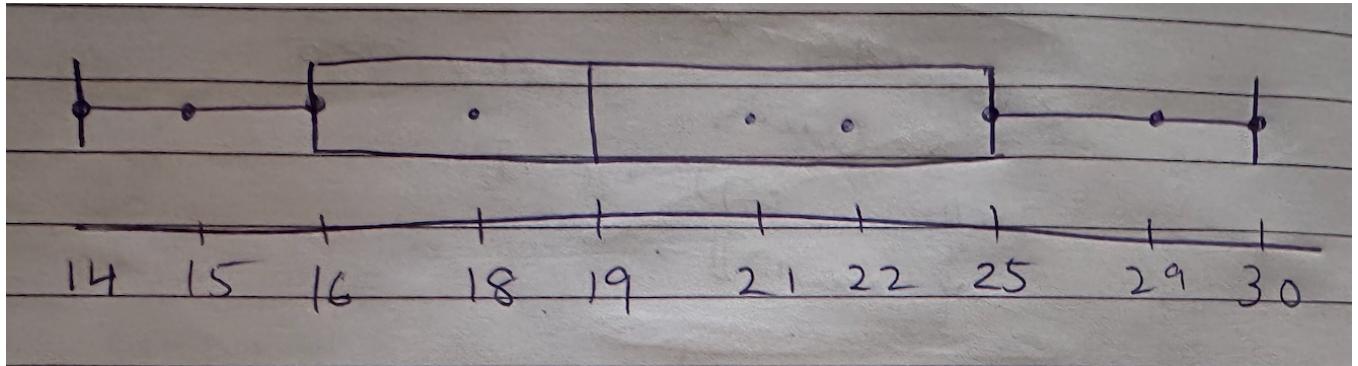
Sneha Satish (<https://iu.instructure.com/courses/2165942/users/6679606>)

Tuesday

1

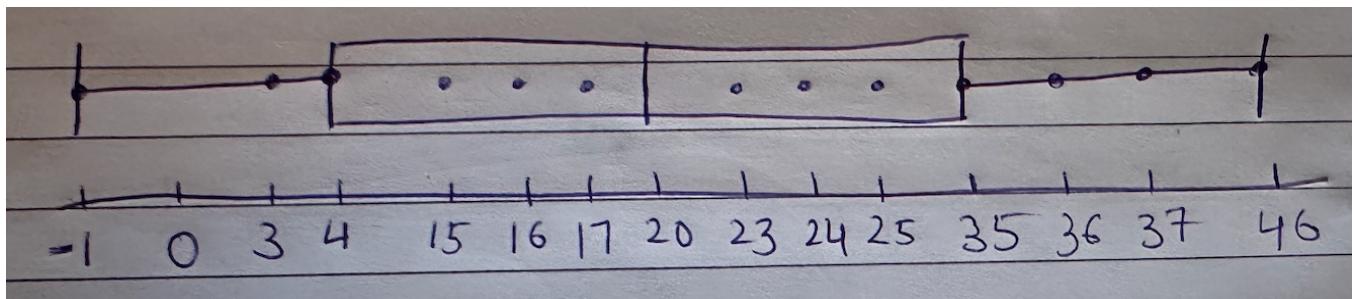
For the first table,

- Minimum is 14
- First quartile (Q1) is the median of the lower half (14, 14, 15, 16, 16, 18, 18), so the middle is = 16
- Median (Q2): 19
- Third Quartile (Q3): The median of the upper half of the data (21, 22, 25, 25, 29, 30, 30), so the middle is = 25
- Maximum is 30



For the second table,

- Minimum is -1
- First quartile (Q1) is the median of the lower half (-1 + 3 + 3 + 4 + 15 + 16 + 16), so the middle is = 4
- Median (Q2): $(17 + 23) / 2 = 20$
- Third Quartile (Q3): The median of the upper half of the data (24, 24, 25, 35, 36, 37, 46), so the middle is = 35
- Maximum is 46



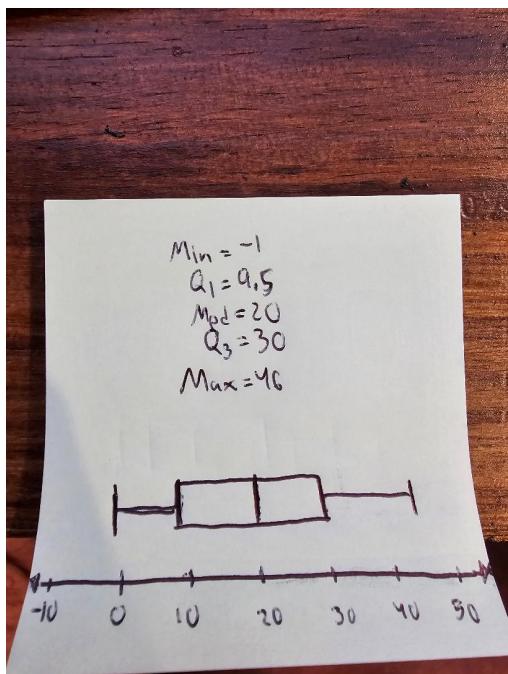
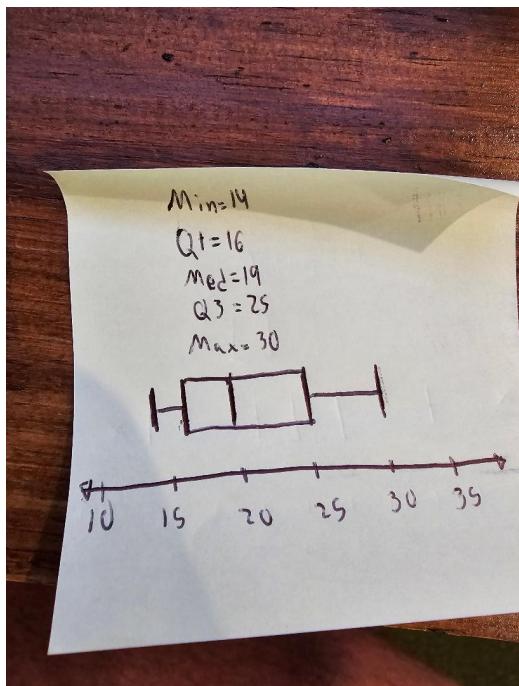
I didn't find any issues doing this as such

Edited by [Sneha Satish](https://iu.instructure.com/courses/2165942/users/6679606) (<https://iu.instructure.com/courses/2165942/users/6679606>) on Sep 19 at 3:37pm

← [Reply](#)

The median for this data set of 32 numbers would be the average of numbers 16 and 17 when ordered, which is the average of 19 and 19 (so 19). The lower quartile is the average of the 8th and 9th numbers, while the upper quartile is the average of the 24th and 25th, giving 15.5 and 25 respectively. The whiskers were set as 3 and 37, as these were the minimum and maximum values within the limits calculated using the IQR and the q1/q3 values. The values of -1 and 46 were outliers.

Edit: I accidentally combined the two tables into 1 and only created one box plot. Updated for two tables shown below.



Reply 

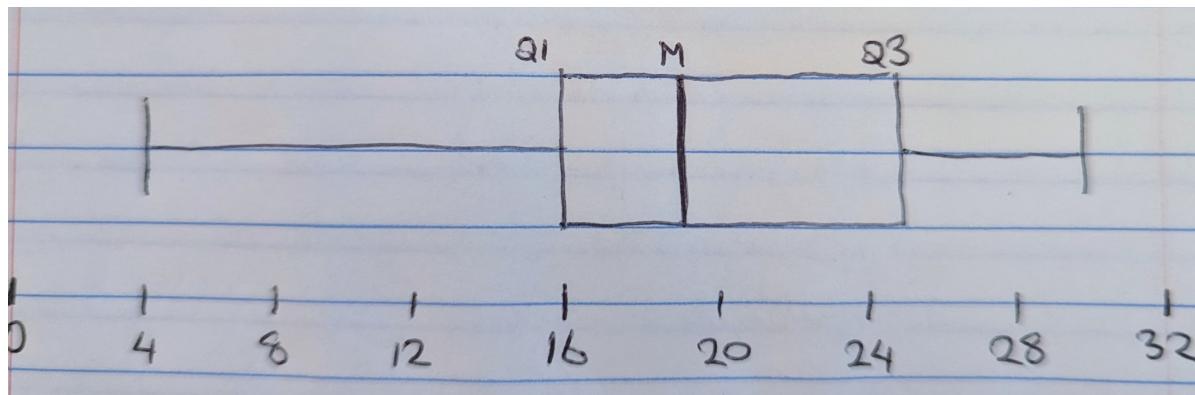
<https://>

[Gary Croke](https://iu.instructure.com/courses/2165942/users/6706306)

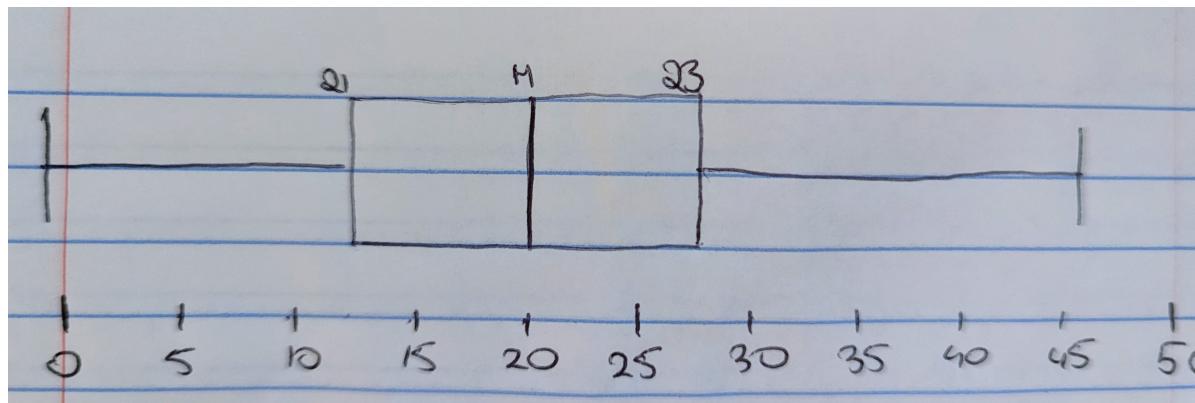
Wednesday

⋮

Set 1: Median = 19, Q1 = 16, Q3 = 25, IQR = 9, Whiskers 4, 30



Set 2: Median = 20, Q1 = 12.25, Q3 = 27.5, IQR = 15.25, Whiskers -1, 46



Whiskers chosen as lowest point within $Q1 - 1.5 \text{ IQR}$, and highest point within $Q3 + 1.5 \text{ IQR}$.

Reply 

<https://>

[Akash Patil](https://iu.instructure.com/courses/2165942/users/6699404)

Thursday

⋮

$$1) Q_2 = \frac{19+19}{2} = 19 \quad \text{Min} = 14$$

$$Q_1 = \frac{16+16}{2} = 16 \quad \text{Max} = 30$$

$$Q_3 = \frac{25+25}{2} = 25 \quad \text{IQR} = Q_3 - Q_1 = 25 - 16$$

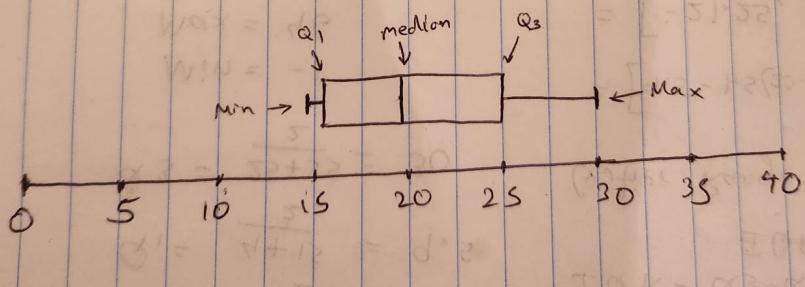
$$= 9$$

$$\text{Outlier range} = [Q_1 - 1.5(\text{IQR}), Q_3 + 1.5(\text{IQR})]$$

$$= [16 - 1.5(9), 25 + 1.5(9)]$$

$$= [\underline{\underline{6.5}}, 38.5]$$

All values inside outlier range



$$2) Q_2 = \frac{17+23}{2} = 20$$

$$Q_1 = \frac{4+15}{2} = 9.5$$

$$Q_3 = \frac{25+35}{2} = 30$$

Min = -1

Max = 46

$$\text{IQR} = Q_3 - Q_1 = 30 - 9.5$$

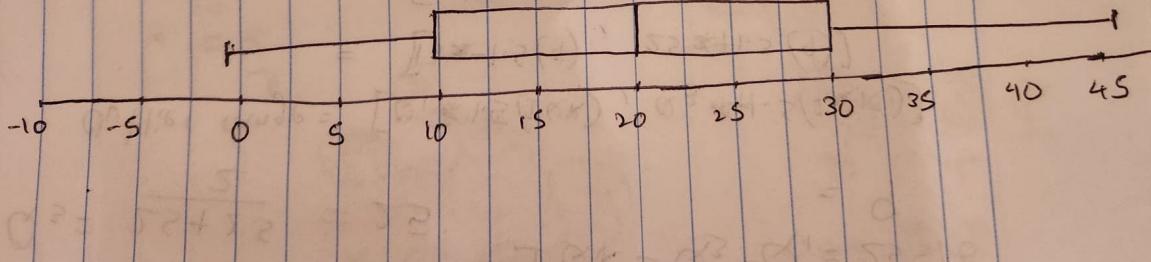
$$\text{IQR} = 20.5$$

$$\text{Outlier Range} = [Q_1 - 1.5 \times (\text{IQR}), Q_3 + 1.5 \times (\text{IQR})]$$

$$= [9.5 - 1.5 \times (20.5), 30 + 1.5 \times (20.5)]$$

$$= [-21.25, 60.75]$$

All values within outlier range



There was nothing particularly tricky.

[Reply](#) [Like](#)

O

(https://)

Shruthi Senthilmani (<https://iu.instructure.com/courses/2165942/users/6688148>)

Thursday

⋮

$$19 \text{ median } Q_2 = \frac{19+19}{2} = 19$$

$$Q_1 = \frac{n + (\frac{n}{2} + 1)}{2} = \frac{4^m + 5^m}{2} = \frac{16+16}{2} = 16$$

$Q_1 = 16$, similarly

$$Q_3 = \frac{25+25}{2} = 25$$

$$IQR = Q_3 - Q_1 = 25 - 16 = 9$$

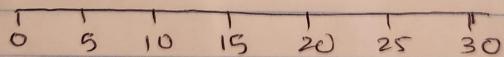
To identify outliers, it should be in the range $[Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)]$

$$\text{outlier range} = [16 - 1.5 \times 9, 25 + 1.5 \times 9] \\ = [16 - 13.5, 25 + 13.5]$$

$$\text{outlier range} = [2.5, 38.5]$$

Therefore, all values are inside this range (no outliers).

$$\min = 14, \max = 30$$



$$2) Q_2 (\text{median}) = \frac{17+23}{2} = \frac{42}{2} = 20$$

$$Q_1 = \frac{4+15}{2} = \frac{19}{2} = 9.5$$

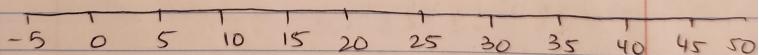
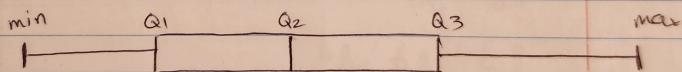
$$Q_3 = \frac{25+35}{2} = \frac{60}{2} = 30$$

$$IQR = Q_3 - Q_1 = 30 - 9.5 = 20.5$$

$$\text{Outlier range} = [Q_1 - 1.5 IQR, Q_3 + 1.5 IQR] \\ = [9.5 - 1.5(20.5), 30 + 1.5(20.5)] \\ = [9.5 - 30.75, 30 + 30.75] \\ = [-21.25, 60.75]$$

Therefore all data points are inside the range (no outliers)

$$\min = -1, \max = 46$$



I did not find anything tricky to understand. It was a pretty straightforward method to calculate Q1, Q2, Q3, and the outlier ranges for both distributions.

[Reply](#) 



[Onur Tekiner \(<https://iu.instructure.com/courses/2165942/users/6758180>\)](#)

Thursday

⋮

I was sure about my knowledge, but I forgot to subtract Q1 from Q3 to calculate IQR. Thanks for sharing this!

[Reply](#) 



[Onur Tekiner \(<https://iu.instructure.com/courses/2165942/users/6758180>\)](#)

Thursday

⋮

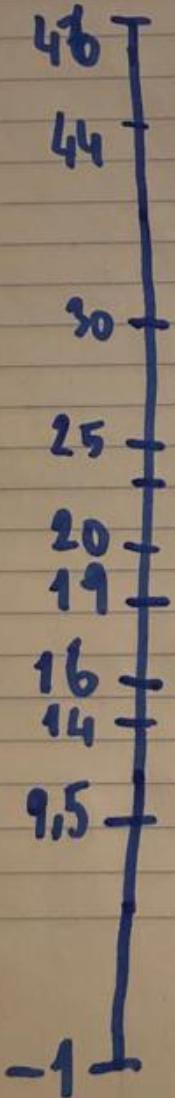
I didn't have any issues. The first plot was quite an easy first plot. The second plots have outliers. Here is my drawing.

$$1) M: 19$$

$$Q_1 = 16$$

$$Q_3 = 25$$

$$8 \cancel{16} \cdot \frac{3}{2} = 24$$



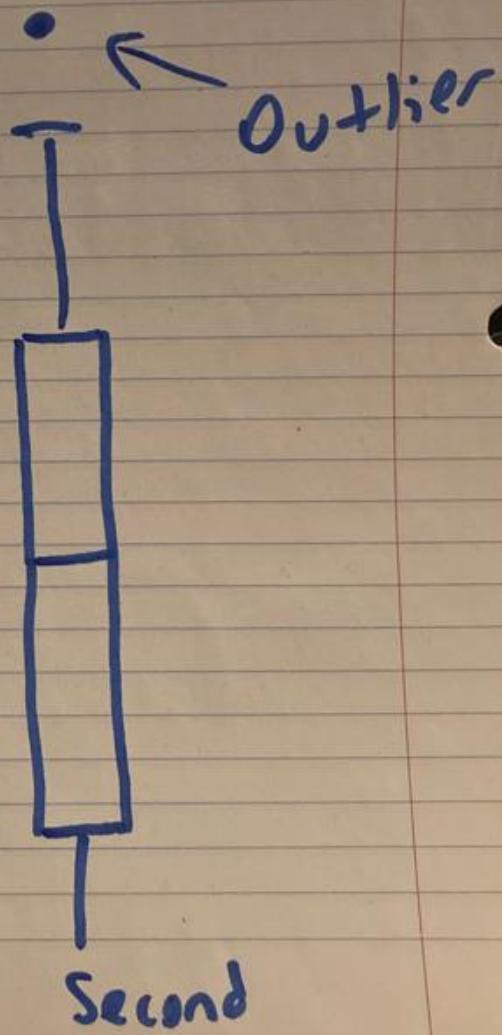
First

$$2) M: 20$$

$$Q_1 = \frac{4+15}{2} = 9.5$$

$$Q_3 = 30$$

$$9.5 \cdot \frac{3}{2} = 14.25$$



Second

O



[Sangzun Park \(https://iu.instructure.com/courses/2165942/users/6703376\)](https://iu.instructure.com/courses/2165942/users/6703376)

Yesterday

⋮

The median values, which are the center lines of the boxes, were 19 and 25, respectively. And the whisker positions representing the maximum and minimum values were 14/30 and -1/46, respectively. There were no outliers. The lines in the boxes representing the first and third quartiles were 16/25 and 9.5/30, respectively. However, in the case of the second data, the value calculated in Excel was 6.75/32.5. I also checked the method of discarding the decimal point, but this was also a different value.

[Reply](#)

O



[Ao Zhang \(https://iu.instructure.com/courses/2165942/users/6703098\)](https://iu.instructure.com/courses/2165942/users/6703098)

Yesterday

⋮

Since there are 16 numbers, the median is the average of the eighth number and the ninth number. Median is 19 for the first group. Similarly, median is 20 for the second group. The upper side of box is the 75th percentile, also known as upper quartile, which is the median of the upper half of dataset. For the first dataset, the upper quartile value is 25. Similarly, for the second dataset, the upper quartile value is 30. And the lower side of box is 25th percentile, also known as lower quartile, which is the median of the lower half of datasets. For the first dataset, the lower quartile is 16, and for the second dataset, the lower quartile is 9.5.

The whiskers are the 2 lines from lower quartile to the minimum value and from upper quartile to the maximum value.

[Reply](#)

O



[Mukul Gharpure \(https://iu.instructure.com/courses/2165942/users/6678592\)](https://iu.instructure.com/courses/2165942/users/6678592)

Yesterday

⋮

The challenges I faced were correctly calculating the positions of quartiles and whiskers and accurately identifying outliers. But, once we understand the calculations and the significance of each component, I think creating a boxplot become a straight process.

[Reply](#)

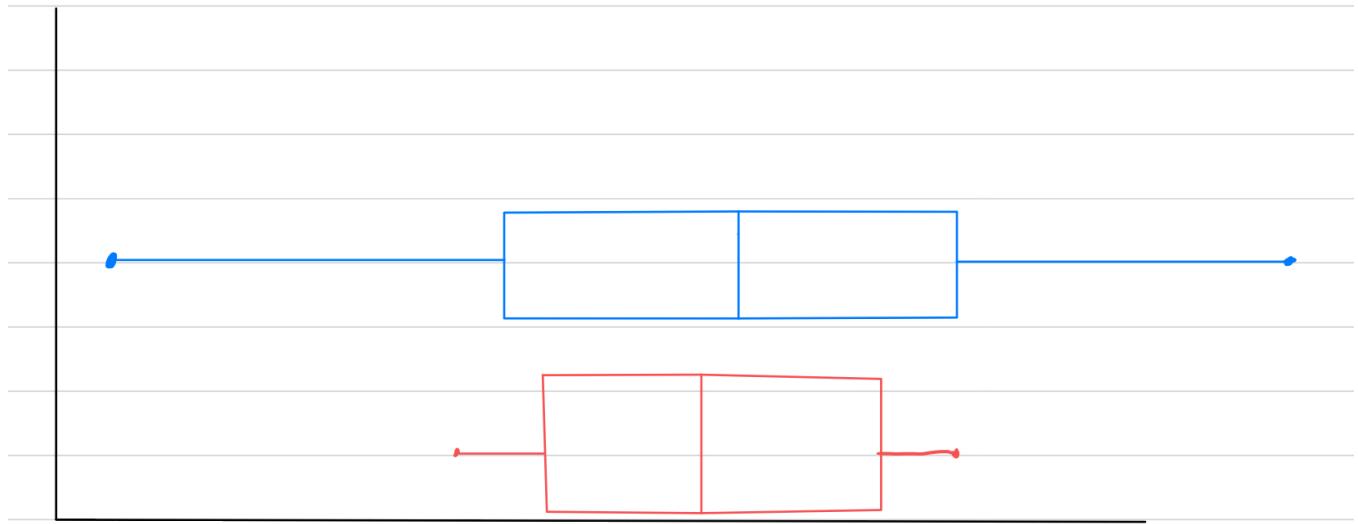


[Andrea Chung](https://iu.instructure.com/courses/2165942/users/6443321)

Yesterday

⋮

Through box plots, I was able to see clearly on which series has the largest distribution of the data points. I drew the box first by indicating median of each table series. Then, I roughly sketched box according to Q1 and Q3 of the data. Then, the whiskers are drawn according to the minimum and maximum data points to show the outliers. This was the rough sketch I got from the two tables above. The blue is the table on the left and red is table on the right.



← [Reply](#)



[Yumeng Liang](https://iu.instructure.com/courses/2165942/users/6587577)

9:56am

⋮

After I draw the boxplot, I think there are several things that can be tricky. Identifying outliers can be tricky, often relying on the $1.5 \times \text{IQR}$ rule. Handling skewed data distributions requires careful placement of whiskers. Quartiles (Q1 and Q3) must be accurately calculated, and whisker lengths determined based on chosen criteria. Understanding the nuances of boxplot construction, including box width, notation, and outlier representation, is essential for creating informative visualizations of data.

[Reply](#)



[Simon Driver \(https://iu.instructure.com/courses/2165942/users/6818242\)](https://iu.instructure.com/courses/2165942/users/6818242)

1:11pm

⋮

For the first table, I calculated the median as 19, and the 1st and 3rd quantiles as 16 and 25 respectively. Hence the two ends of my box were at 16 and 25, and the whiskers were at 14 and 30.

For the second table, I calculated the median as 20, and the 1st and 3rd quantiles as 9.5 and 30 respectively. This meant the the edges of my box were at 9.5 and 30, and my whiskers were at 46 and -3.

For each one, since there were 16 numbers, I calculated the median as the average of numbers 8 and 9. The 1st and 3rd quartiles I found by finding the median of the top half and the bottom half of the list, respectively. The whiskers extended to the farthest outliers then (the minimum and maximum values of the set). THe box captured the area from Q1 to Q3, with the median in the center.

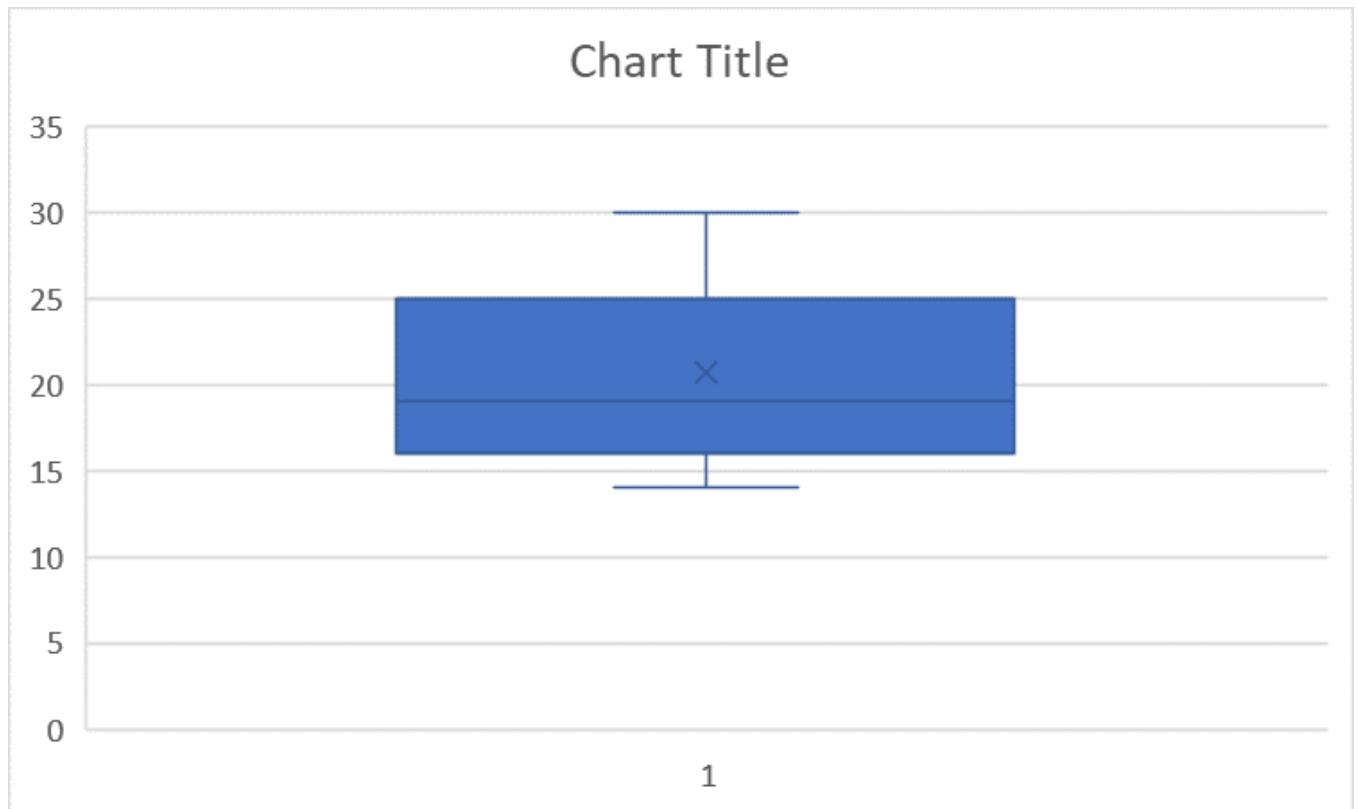
[Reply](#)



[Madhuri Patibandla \(she/her/hers\) \(https://iu.instructure.com/courses/2165942/users/6760559\)](https://iu.instructure.com/courses/2165942/users/6760559)

2:05pm

⋮



16Q1

25Q3

19Q2

Q2 : Median is sum of all the values / count of rows call as median.

Q1 : 25% of the first 7 Values comes as 16.

Q3 : 25% of Last 7 Values comes as 25

Second Plot :

12.25Q1

27.5Q3

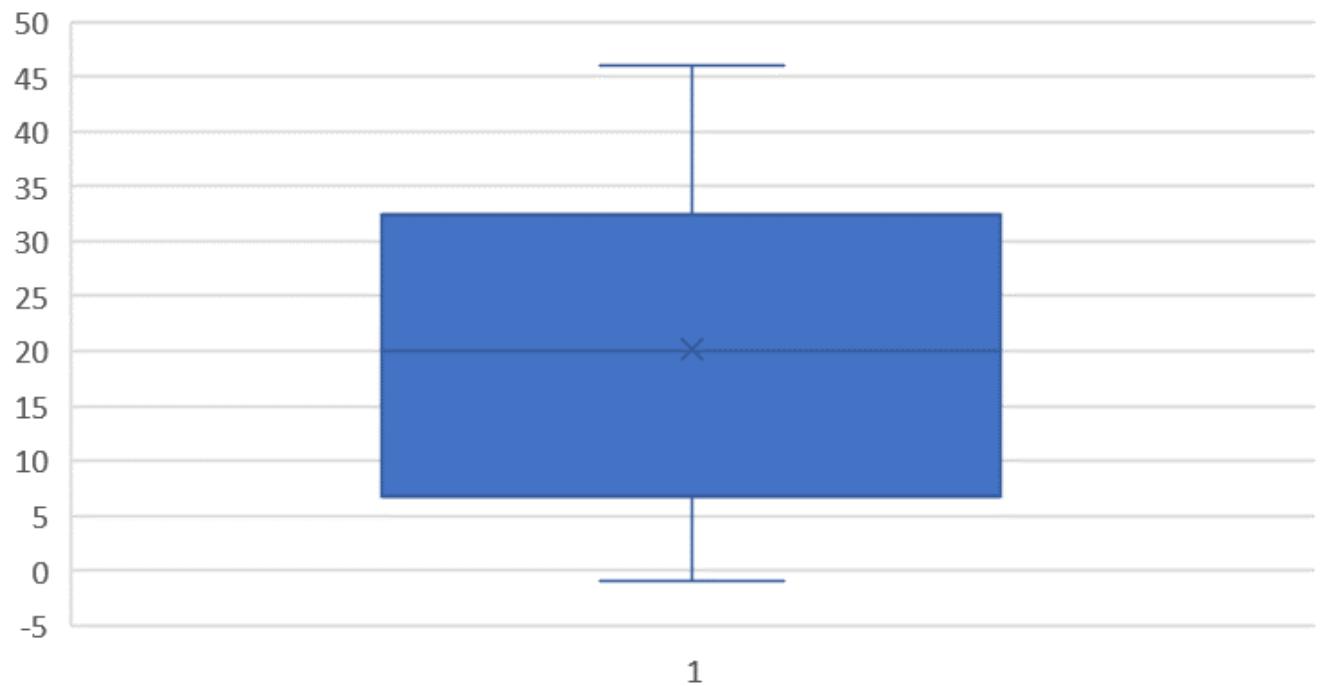
20Q2

Q2 : Median is sum of all the values / count of rows call as median as 20.

Q1 : 25% of the first 7 Values comes as 16.

Q3 : 25% of Last 7 Values comes as 25

Chart Title



↪ [Reply](#)



[Shreedeep Sadasivan Nair \(he/him/his\)](https://iu.instructure.com/courses/2165942/users/6813278) (https://iu.instructure.com/courses/2165942/users/6813278)

2:43pm

⋮

the tricky part is

↪ [Reply](#)



[Jash Shah](https://iu.instructure.com/courses/2165942/users/6684840) (https://iu.instructure.com/courses/2165942/users/6684840)

4:13pm

⋮

Dataset 1

$$\text{Min} = 14$$

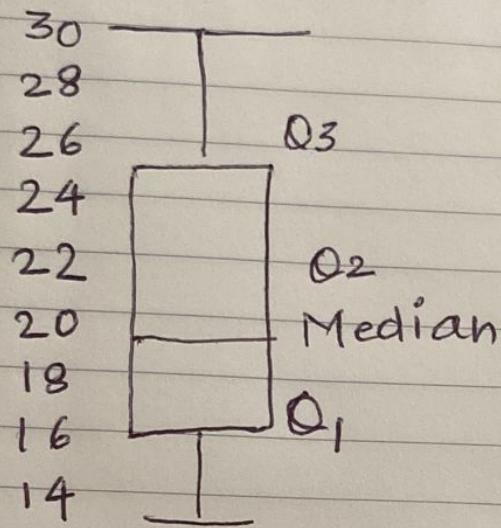
$Q_1 = \text{Median of first half } (14-19) = 16$

$$\text{Median} = 19 (Q_2)$$

$$Q_3 = 25$$

$$\text{Max} = 30$$

$$\text{IQR} = Q_3 - Q_1 = 25 - 16 = 9$$



Dataset 2

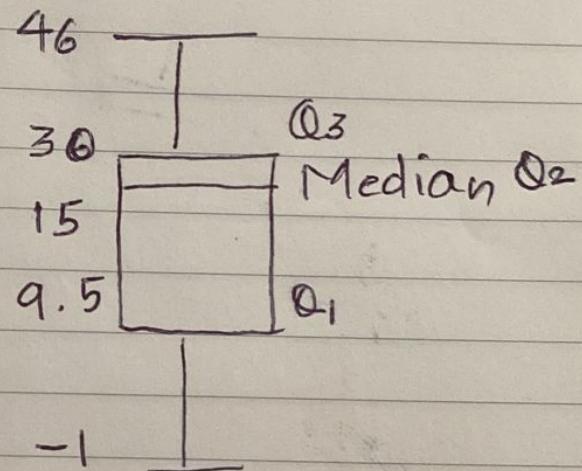
$$\text{Min} = -1$$

$$Q_1 = 9.5$$

$$\text{Median} = 20$$

$$Q_3 = 30$$

$$\text{Max} = 46$$



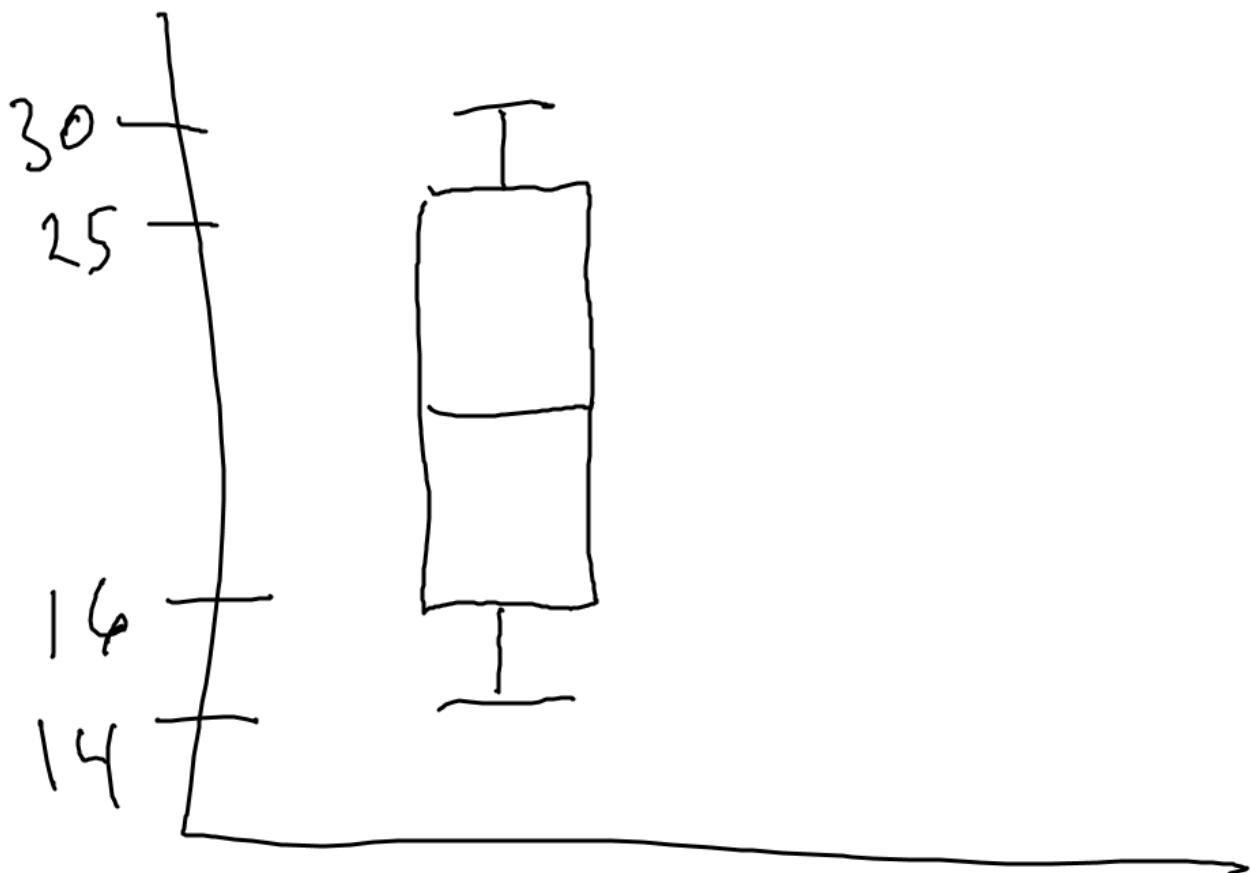


← [Reply](#)

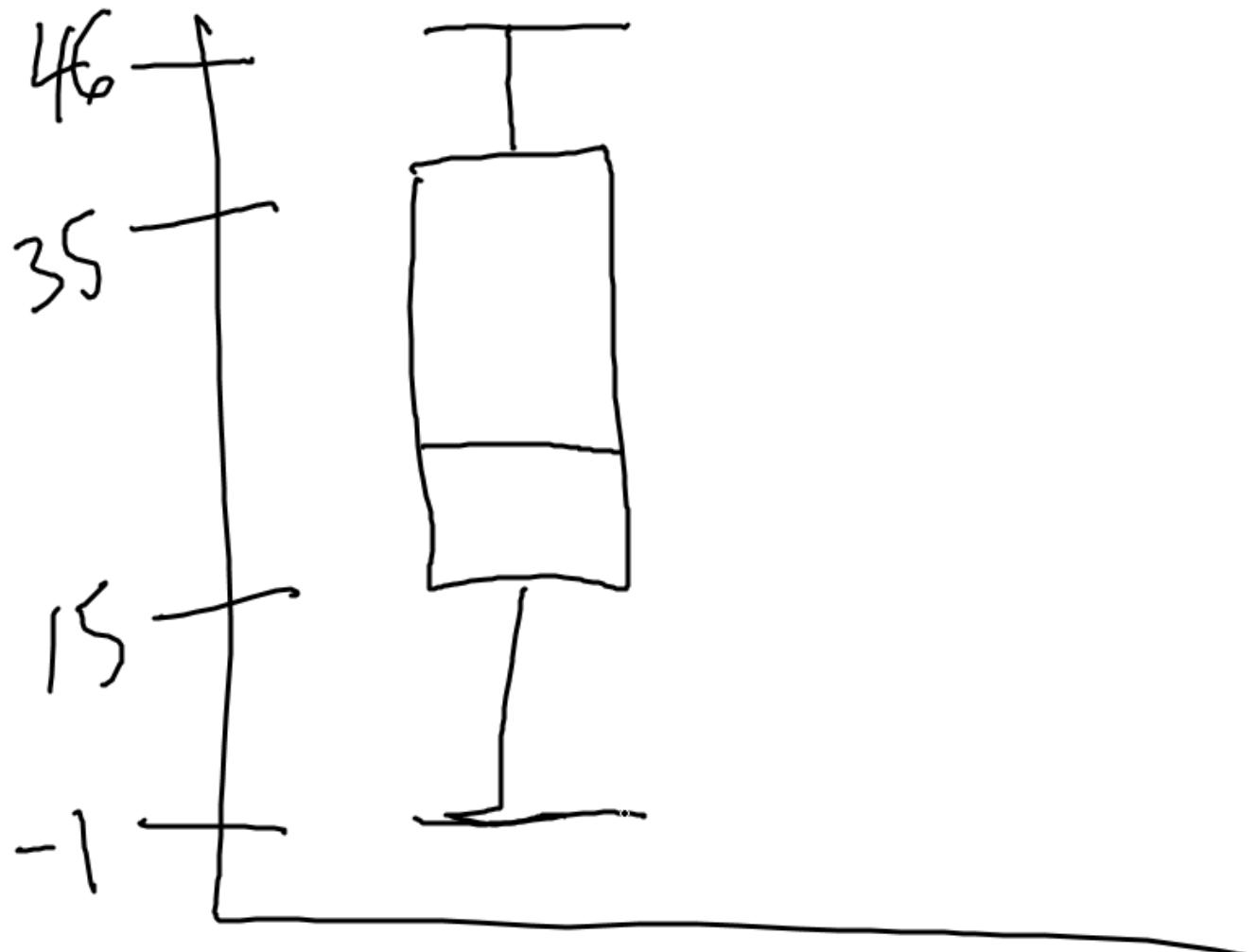


[Robert Perez \(he/him/his\) \(https://iu.instructure.com/courses/2165942/users/6701521\)](https://iu.instructure.com/courses/2165942/users/6701521)

5:03pm



For the first series, the plot was fairly simple. The minimum is 14, the maximum is 30 (the whiskers), and the middle 80% falls roughly between 16 and 25. I didn't actually do the math on the numbers; I just sort of estimated it in my head. However, the data appeared a tad left-skewed to me, so I drew the median in the lower half of the box, roughly around 19.



The second one was more challenging, as the data was both more spread out and more bunched in the middle. For the second, the whiskers are at -1 and 46. The box ranges from about 15-35, and the data in this case is again more left-skewed, so I drew the median line around 22.

Reply



O

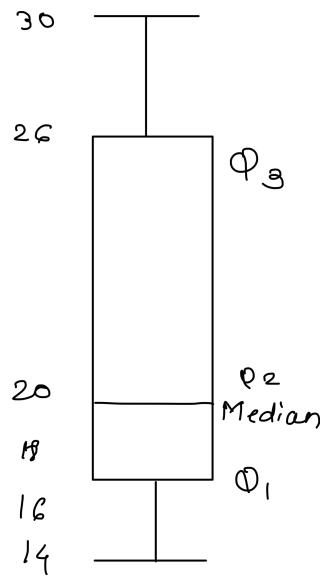


[Aditya Sanjay Mhaske](https://iu.instructure.com/courses/2165942/users/6692144) (<https://iu.instructure.com/courses/2165942/users/6692144>)

5:16pm

⋮

Dataset 1



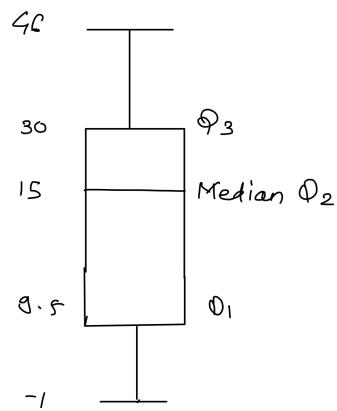
Min = 14

Q1 = 16

Median Q2 = 19

Max = 30

Dataset 2



Min = -1

Q1 = 9.5

Median = 20

Q3 = 30

Max = 46

Edited by [Aditya Sanjay Mhaske](https://iu.instructure.com/about/6692144) (<https://iu.instructure.com/about/6692144>) on Sep 23 at 6:21pm

[Reply](#) [Like](#)