Sneak Peak!

Unit 2 and 3 Content

LCQ: Frequency Tables with Calc Help

GOAL: Enter and sort data!

1. Enter data in L_1: STAT → 1:Edit
   12, 34, 50, 10, 11, 7, 8, 29, 30, 44, 21, 32, 30, 4, 28, 43, 12, 22, 38, 49

2. Sort data: STAT → 2:SortA(
   a) Then find L_1 (2ND → 1). This sorts your list from smallest to largest
   b) Then can go back into L_1 to view the sorted data set

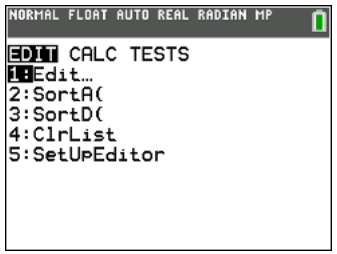
Next Goal: Construct a Frequency Table!

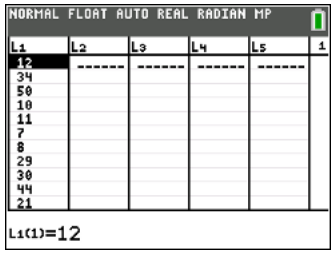
Bin	Frequency	Relative Frequency
0-10		
10-20		
20-30		
30-40		
40-50		
50-60		
Total:		

LCQ: Frequency Tables with Calc Help

GOAL: Enter and sort data!

1. Enter data in L_1: STAT → 1:Edit
   12, 34, 50, 10, 11, 7, 8, 29, 30, 44, 21, 32, 30, 4, 28, 43, 12, 22, 38, 49

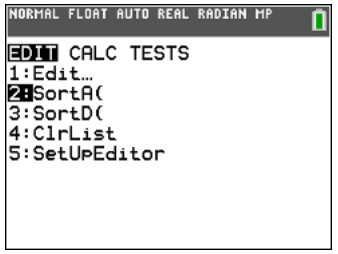


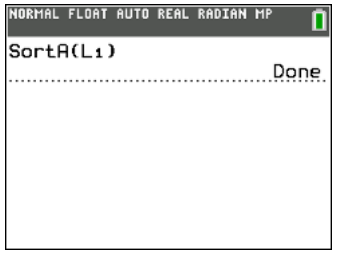


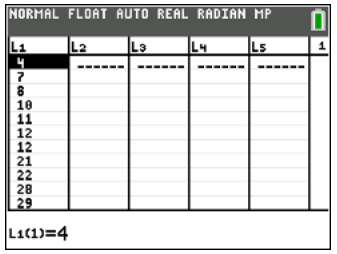
2. Sort data: STAT → 2:SortA(
   a) Then find L_1 (2ND → 1). This sorts your list from smallest to largest
   b) Then can go back into L_1 to view the sorted data set

Next Goal: Construct a Frequency Table!

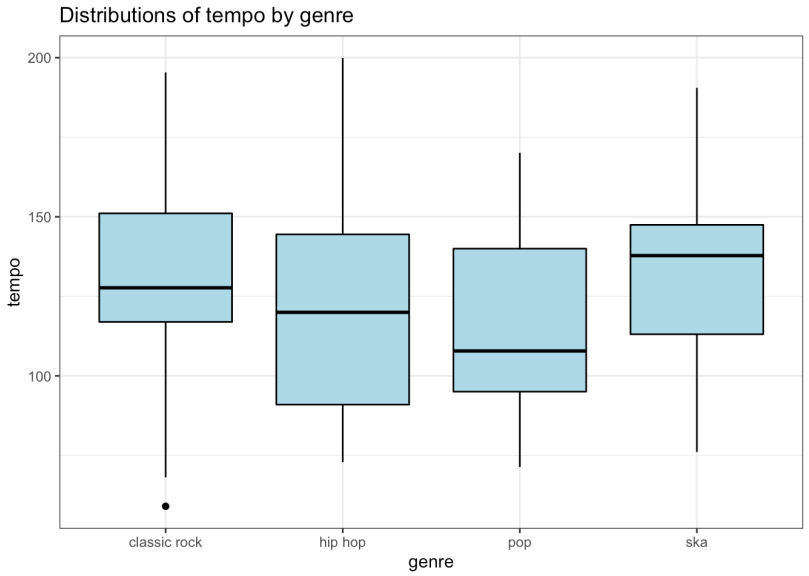
Bin	Frequency	Relative Frequency
0-10	*3*	*0.15*
10-20	*4*	*0.20*
20-30	*4*	*0.20*
30-40	*5*	*0.25*
40-50	*3*	*0.15*
50-60	*1*	*0.05*
Total:	*20*	*1*







Boxplots



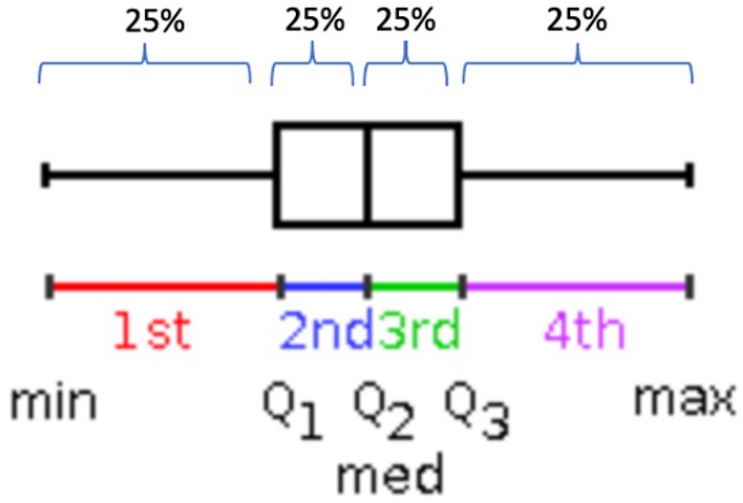
Boxplot (box and whisker)

- Great display for **quantitative** data.
- Shows summary information, not raw data.
  - Also shows outliers.
- Splits data into quarters, called Quartiles.
- Can see **shape**, but NOT **modality**!

How to Construct

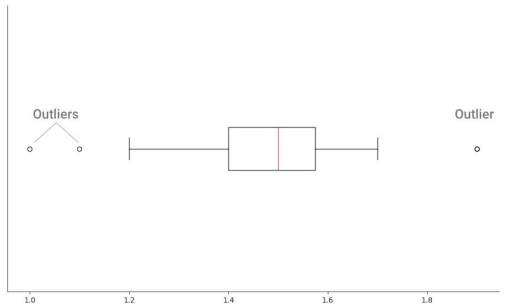
1,1,2,2,3,(Q1)3,4,4,5,5,(M=Q2)6,6,7,7,8,(Q3)8,9,9,10,10

- Calculate the 5-number summary (min, Q1, **median**, Q3, max) and draw the shape.
- Slight caveat when there is outliers…



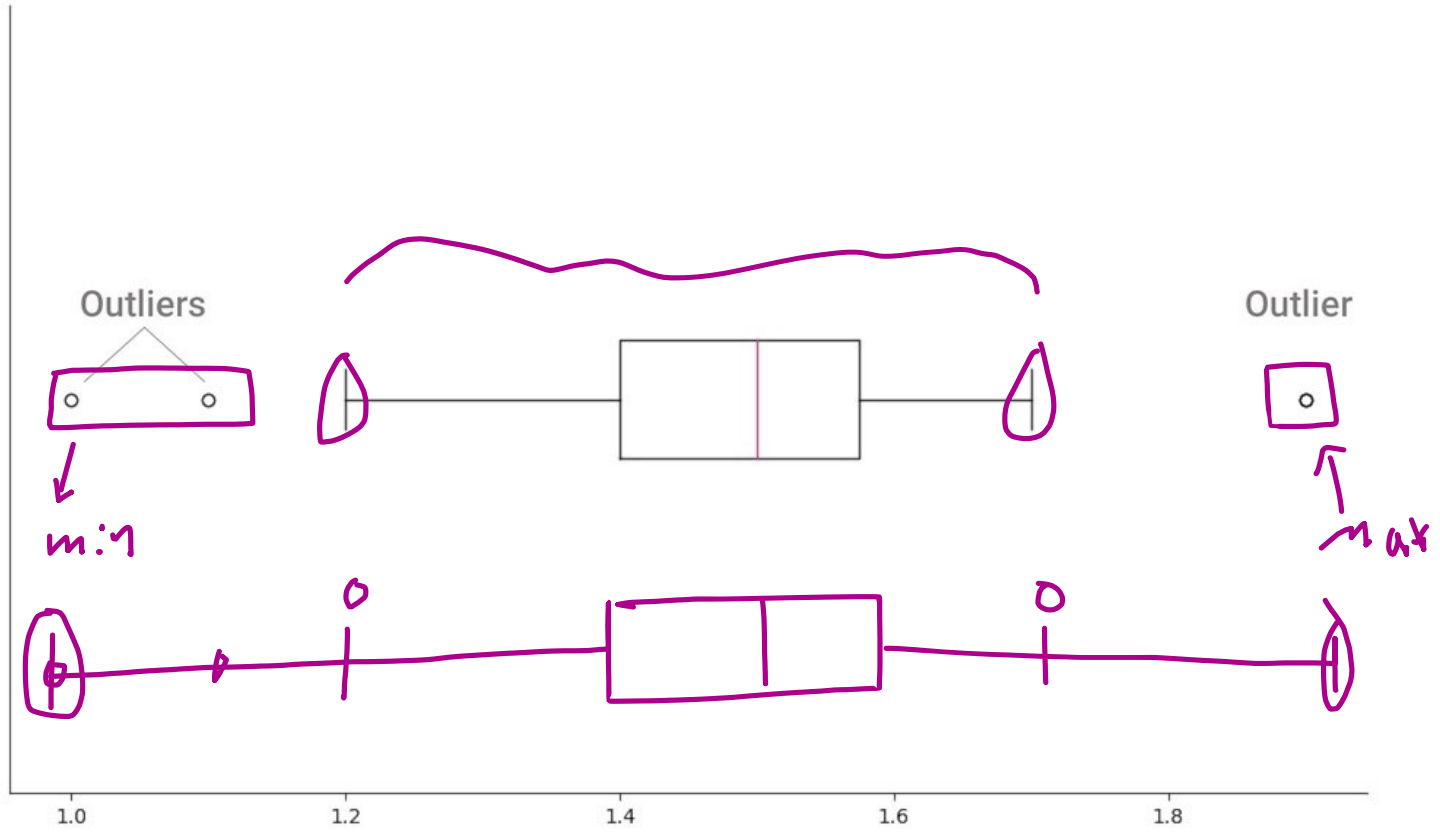
Two Types of Boxplots

- Both report *5-number summary* information.
- Only difference is if **outliers** are drawn separately.
  - **Outliers** are drawn with dots.
  - **Fences** are drawn at the next values that are NOT considered outliers!
  - We aren't going to worry about how to we classify a point as an **outlier,** just eyeball from a histogram**!**



https://towardsdatascience.com/create-and-customize-boxplots-with-pythons-matplotlib-to-get-lots-of-insights-from-your-data-d561c9883643

Two Types of Boxplots



"Advanced (modified) version" where outliers are separated

"Regular" version

How outliers look on histograms → few points separated from the majority of the data

Calculator Fun!!!
Boxplot

GOAL: Make a Boxplot!

1. Enter raw data (20 observations):
   - 38, 33, 5, 5, 47, 29, 24, 42, 3, 18, 30, 46, 25, 44, 40, 42, 39, 44, 29, 13
2. Set STAT PLOT: Boxplot.
   - Can do with outliers (the one with two dots) or without outliers.
3. ZOOM → 9:ZoomStat
   - This automatically zooms to whatever data range the stat plot requires
4. Graph and Trace!

This automatically computes the 5-number summary: Min, Q1, Med, Q3, Max in order to make the boxplot.

- *** (Will learn what this is next week) Can confirm the results by doing STAT → (over one to CALC) → 1-Var Stats → Calculate → (scroll down)

Calculator Fun!!! Boxplot

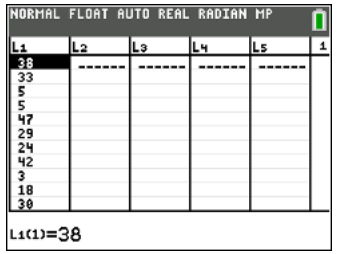
GOAL: Make a Boxplot!

1. Enter raw data (20 observations):
   - 38, 33, 5, 5, 47, 29, 24, 42, 3, 18, 30, 46, 25, 44, 40, 42, 39, 44, 29, 13
2. Set STAT PLOT: Boxplot.
   - Can do with outliers (the one with two dots) or without outliers.
   - ** Can leave the option with outliers on and just use that from now on!
3. ZOOM → 9:ZoomStat
   - This automatically zooms to whatever data range the stat plot requires
   - Make sure to redo this step each time you enter new data and want to draw a new boxplot
4. Graph and Trace!

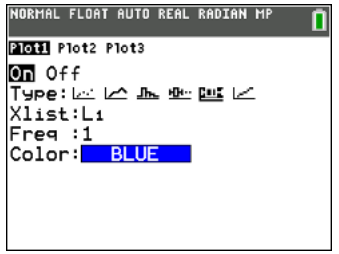
This automatically computes the 5-number summary: Min, Q1, Med, Q3, Max in order to make the boxplot.

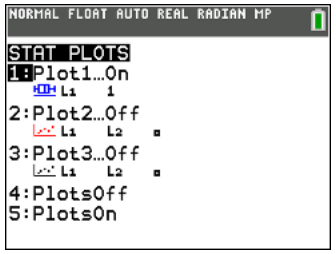
- *** (Will learn what this is next week) Can confirm the results by doing STAT → (over one to CALC) → 1-Var Stats → Calculate → (scroll down)

1)

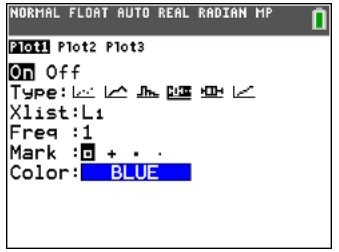


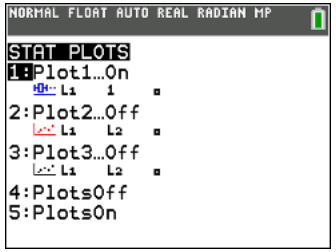
2)





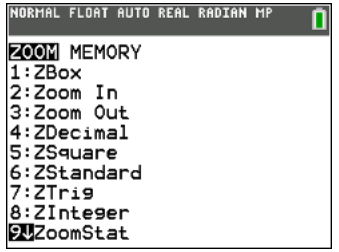
Regular boxplot



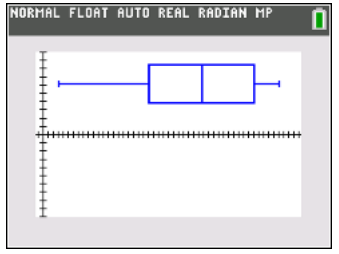


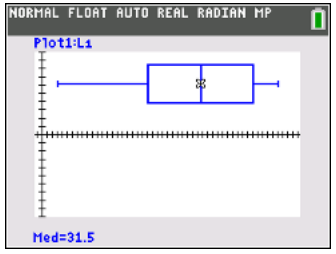
Modified boxplot

3)



4)





There were no outliers in this dataset, so the regular and modified boxplots look the same

Trace lets you arrow over to each number in the boxplot and displays the value

Calculator Fun!!! Boxplot

Additional Demonstration

- Lets add a new data point to see if it is classified as an **outlier**!

1) Add additional data point = 100 in L_1

2) Setting the STAT PLOT for the boxplot with outliers should already be done
   - We can leave this set so it is like our “default” option now

3) Now we will reset the Zoom
   - Remember we have to do this each time we make a new boxplot

4) Graph and Trace again!

Calculator Fun!!! Boxplot

Additional Demonstration

- Lets add a new data point to see if it is classified as an **outlier**!

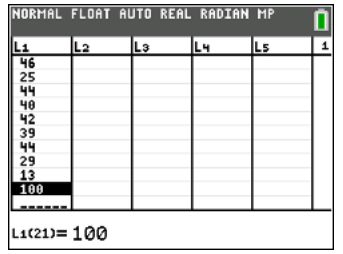
1) Add additional data point = 100 in L_1

2) Setting the STAT PLOT for the boxplot with outliers should already be done
   - We can leave this set so it is like our "default" option now

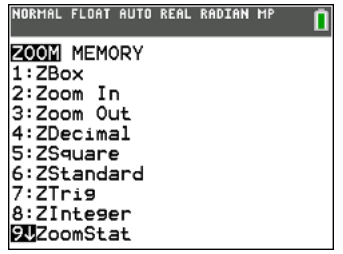
3) Now we will reset the Zoom
   - Remember we have to do this each time we make a new boxplot

4) Graph and Trace again!

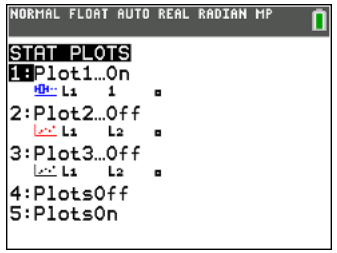
1)



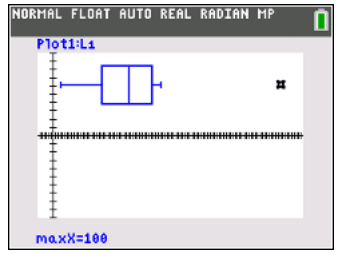
2)

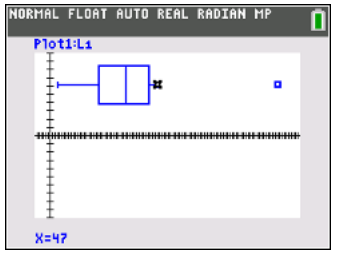


3)



4)



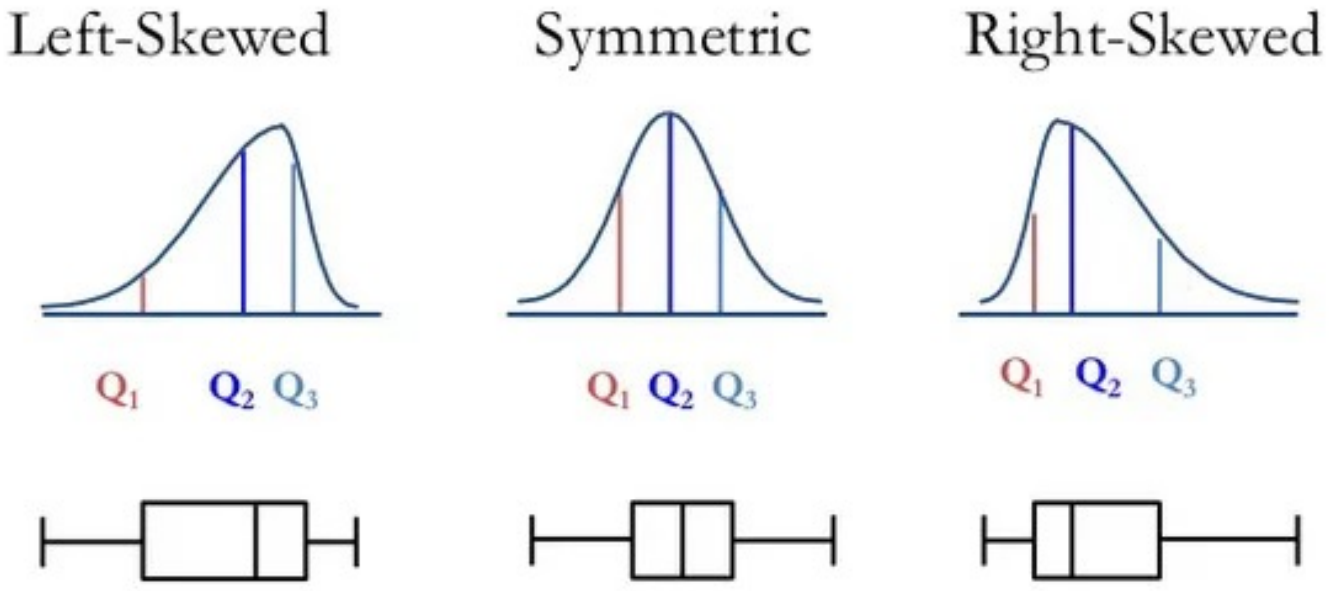


We see now that there is a point (100) that is drawn separately as an outlier!

And the previous maximum (47) is drawn as the right fence, indicating that it is the largest value NOT considered an outlier

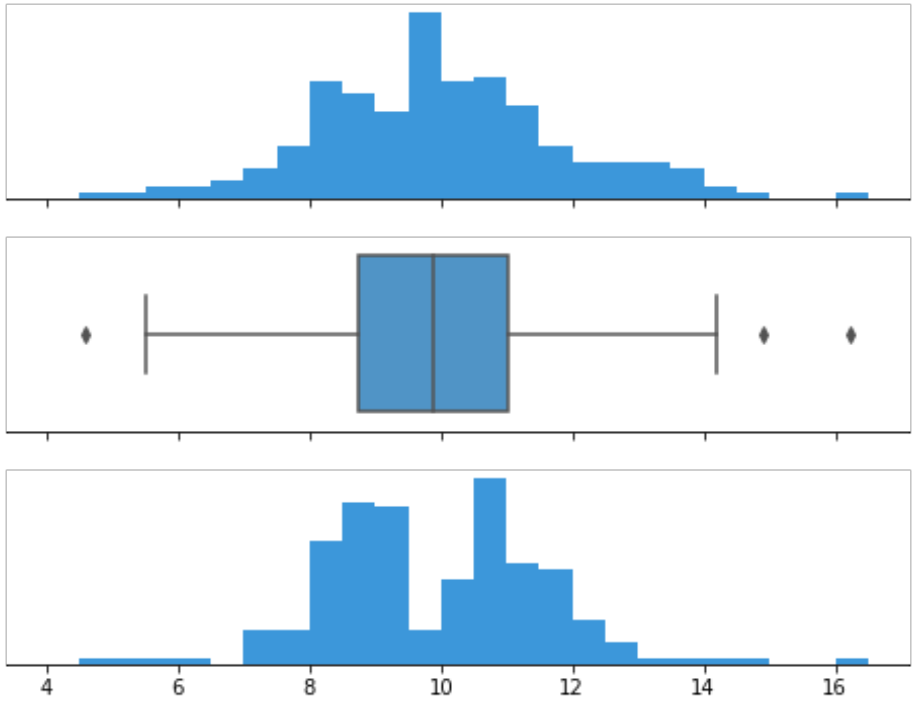
Other Visual Tools: Boxplots and Shape and Modality

- CAN tell **shape** from a boxplot.



https://www.simplypsychology.org/boxplots.html

- CANNOT tell **modality** from a boxplot.

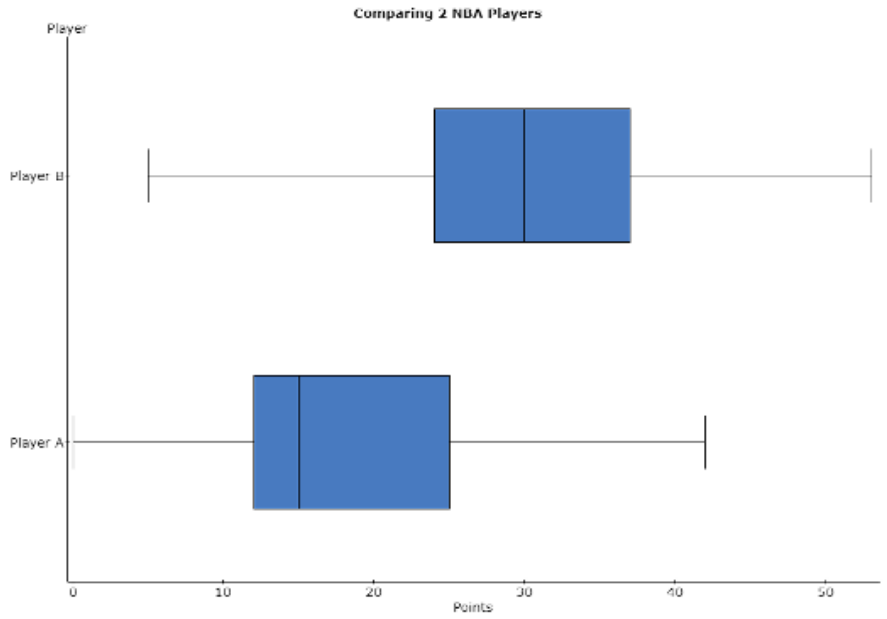


https://chartio.com/learn/charts/box-plot-complete-guide/

LCQ: Measures of Shape

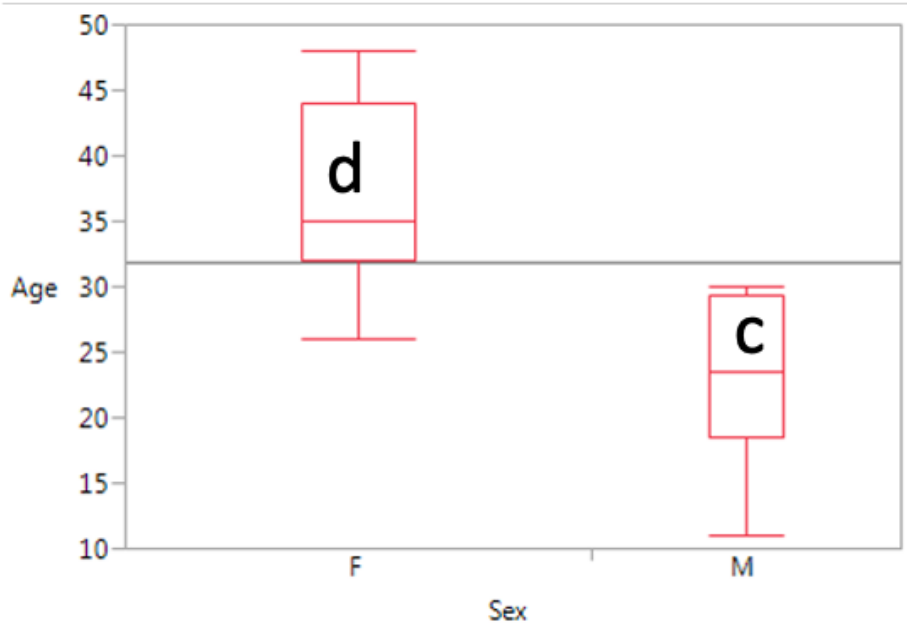
Describe the SHAPE and MODALITY of each boxplot

a)
b)
c)
d)



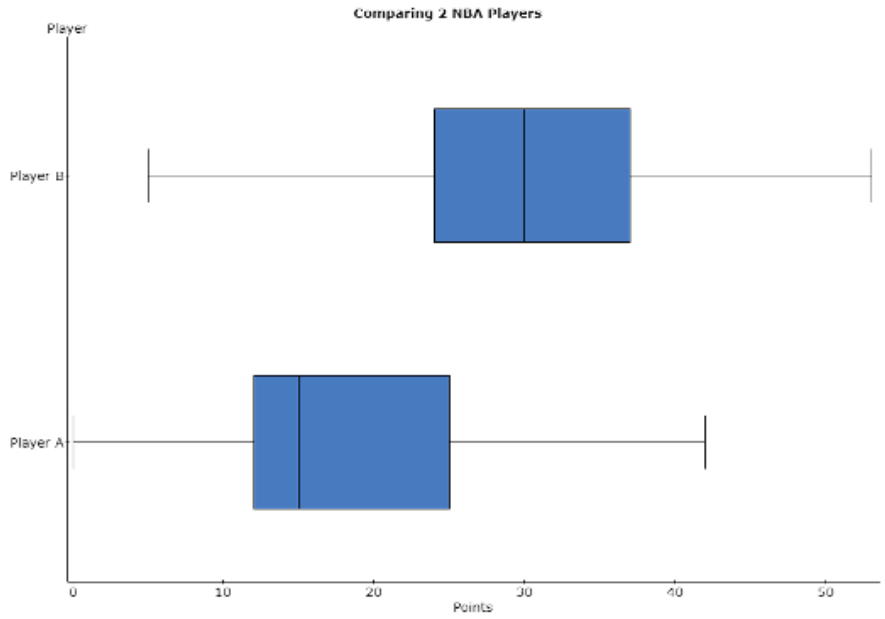
a

b



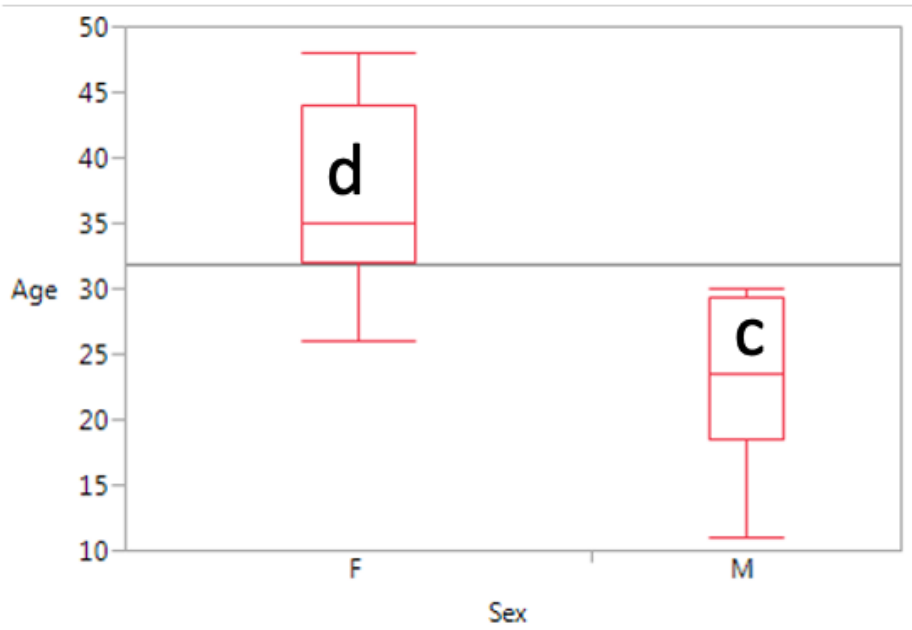
LCQ: Measures of Shape

Describe the SHAPE and MODALITY of each boxplot



a

b



a) Roughly Symmetric
b) Right skewed
c) Left skewed → make sure to rotate the image correctly if that's how you are going to look at it! Lower numbers need to be on the left and then increasing to the right. Then look at the shape
d) Skewed right

PROBLEM SESSION!!!!!!!!!!

Problem #21

The five-number summary for the total revenue (in $M) of the top 100 movies of 2012 looks like this:

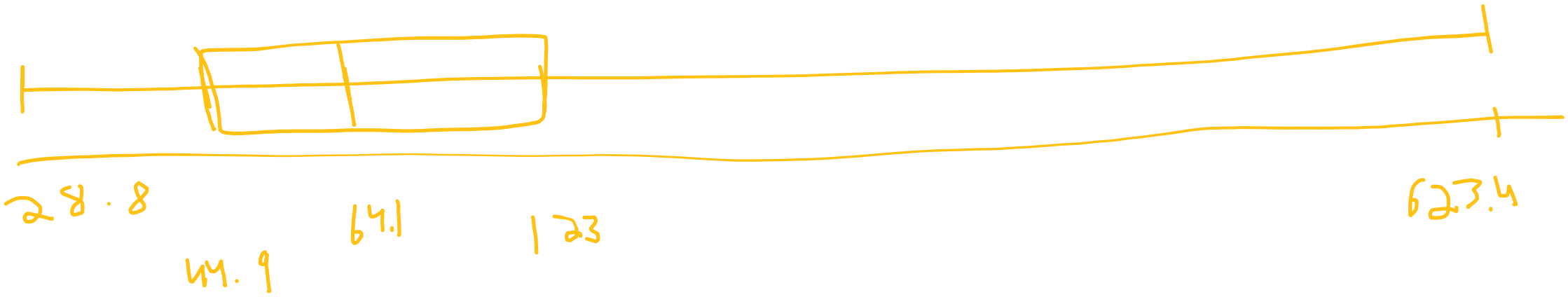
Min	Q1	Med	Q3	Max
28.8	44.9	64.1	123.0	623.4

- Sketch a boxplot and find the IQR and the Range

Problem #21 Solution

Min	Q1	Med	Q3	Max
28.8	44.9	64.1	123.0	623.4

- IQR = Q3 – Q1 = 123 – 44.9 = 78.1
- Range = Max – Min = 623.4 – 28.8 = 594.6



(as much to scale as I could…)

Problem #13

As part of the marketing team at an Internet music site, you want to understand who your customers are. You send out a survey to 25 customers asking for demographic information. One of the variables is the customer's age :

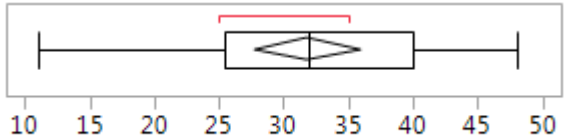
20	38	35	30	22	34	44	44	29	35	30	26	48
30	25	32	22	42	14	32	29	48	44	11	32	

a) Construct a boxplot.

b) Does the boxplot nominate any outliers?

Problem #13 Solution

a) Boxplot



b) No outliers are present

Problem #17

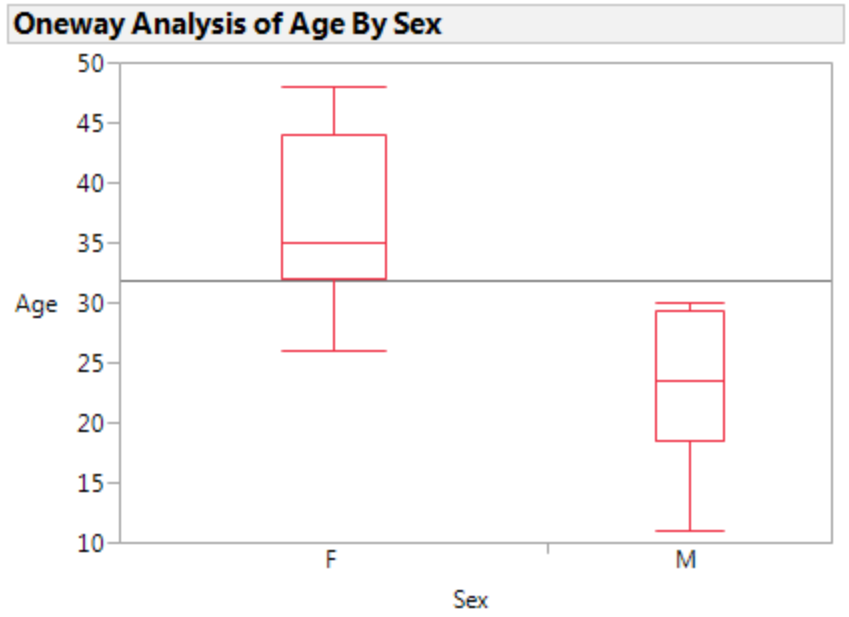
The survey from Exercise 13 also asked the customers to say whether they were male or female.

Age	Sex	Age	Sex	Age	Sex	Age	Sex	Age	Sex	Age	Sex	Age	Sex
20	M	35	F	22	M	44	F	29	M	30	M	48	F
30	F	32	F	42	F	32	F	48	F	11	M	44	F
38	F	30	M	34	F	44	F	35	F	26	F	32	F
25	M	22	M	14	M	29	M						

Construct boxplots to compare the ages of men and women and write a sentence summarizing what you find.

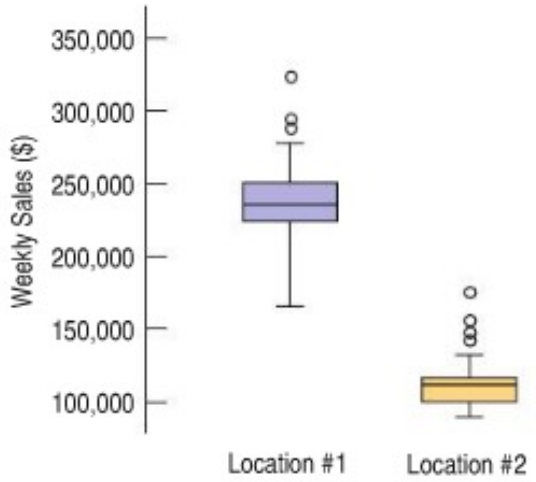
Problem #17 Solution

The females customers are much older than the male customers; more than 75% of the women are older than all of the all the men. Women have a median age of 35 compared to a median age of 23 for the men. No unusual values in either group. The range of ages for women is about 22 and the range for men is about 20.



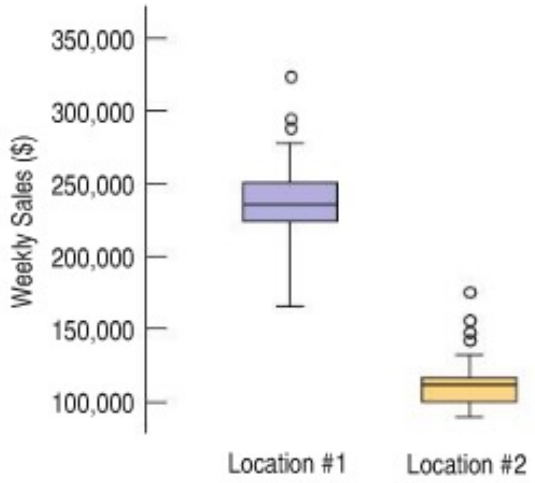
Problem #19

Here are boxplots of the weekly sales (in $US) over a two-year period for a regional food store for 2 locations. Location #1 is a metropolitan area that is known to be residential where shoppers walk to the store. Location #2 is a suburban area where shoppers drive to the store. Assume that the two towns have similar populations and that the two stores are similar in square footage. Write a brief report discussing what these data show.



Problem #19 Solution

Sales in Location #1 were higher than sales in Location #2 in nearly every week. However, there was much more variability in Location #1. Both distributions have outliers. The distribution for location #1 is roughly symmetric and for location #2 is right skewed. The company might want to compare other stores in locations like these to see if this phenomenon holds true for other locations.



Problem #53

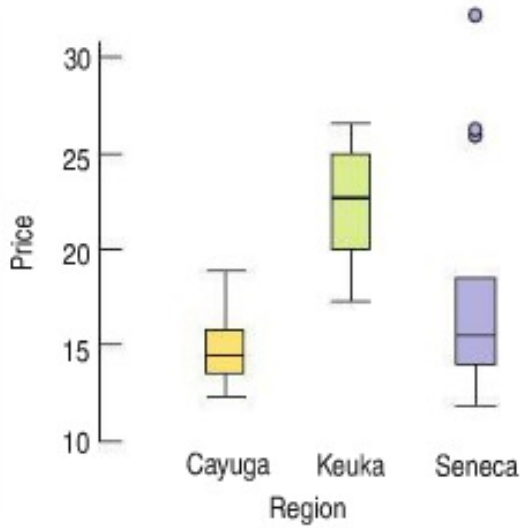
The boxplots display bottle prices (in dollars) of dry Riesling wines produced b vineyards along three of the Finger Lakes in upstate New York.

a) Which lake region produced the mos expensive wine?

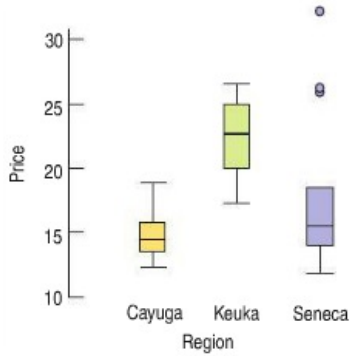
b) Which lake region produced the cheapest wine?

c) In which region were the wines generally more expensive?

d) Write a few sentences describing these prices.



Problem #53 Solution



a) The most expensive wine is produced in Seneca.
b) Cayuga and Seneca produce less expensive wines, with Cayuga producing the cheapest wines. Seneca has several high outliers.
c) The most expensive wines are typically produced in Keuka.
d) Cayuga Lake vineyards and Seneca Lake vineyards have approximately the same average price of about $15 per bottle, which a typical Keuka Lake vineyard has a price near $22. Keuka Lake vineyards have consistently high prices between about $18-$26 a bottle. Cayuga Lake vineyards have prices from $13-$19, and Seneca Lake vineyards have highly variable prices from $12 to over $30