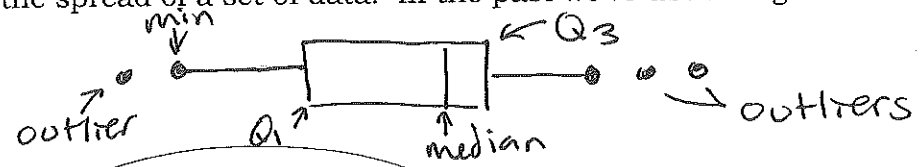
Box Plot Aka Box and Whisker

A box plot is a good way to convey the spread of a set of data. In the past we've used range to analyze spread.

Range = maximum – minimum



A box plot is a data display based on the 5-number summary. The 5-number summary is comprised of the:

5

minimum:	the lowest value of the data set
first (lower) quartile (Q_1):	the median of the numbers below the location of the median
second (middle) quartile (Q_2):	the median
third (upper) quartile (Q_3):	the median of the numbers above the median
maximum:	the highest value of the data set

To create a box plot, you first need to determine the value(s) of any outliers. To do this:

1. Find the value of $Q_3 - Q_1$, the interquartile range or IQR.
2. Add 1.5 x IQR to the 3rd quartile. Any number larger than $Q_3 + 1.5(IQR)$ is an outlier.
3. Subtract 1.5 x IQR from the 1st quartile. Any number smaller than $Q_1 - 1.5(IQR)$ is an outlier.

To construct a box plot:

1. Draw a number line including the maximum and minimum data values.
2. Put a point at each outlier value.
3. Draw a rectangle with opposite sides at Q_1 and Q_3.
4. In the box, draw a segment parallel to the sides of the rectangle at the median.
5. Draw segments from the midpoints of the sides of the rectangle to the lowest and highest values that are not outliers. If there are no outliers, use the maximum and minimum.

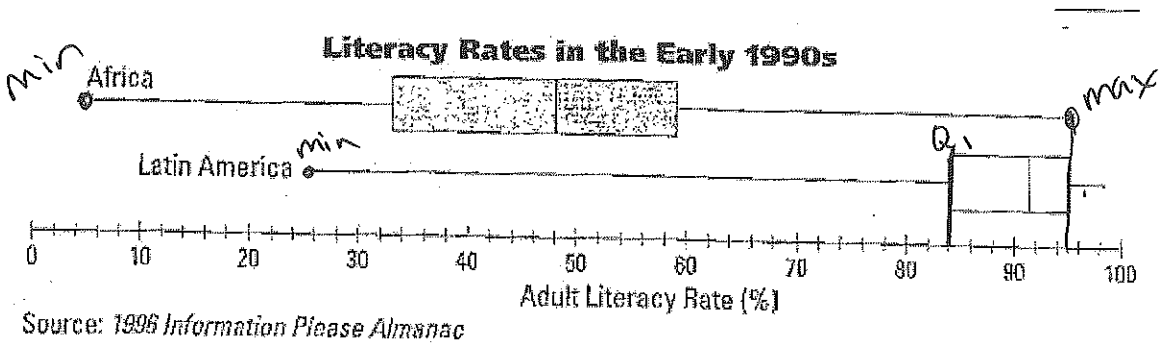
Categorical / Quantitative

Advantages: shows the 5-point summary, highlights outliers, easy to compare

Disadvantages: hard to construct, not visually appealing

Example 1
Use the height data we collected from class to create a box plot.

Example 2
Use the box plots to answer the following questions.



Source: *1996 Information Please Almanac*

a. Estimate the range of literacy rates in the two regions. 98-26=

95-5=
Africa 90 Latin America 72

b. In Latin America, what percent of literacy rates are below 84%?

25%.

c. In which part of the world, Africa or Latin America, are literacy rates generally higher? How do you know?

Latin America because their measures of center (5-point summary) is further to the right

d. Write a brief comparison of literacy rates in these two regions. Mention median and quartile values. be specific

Median for L.A. is 91%. which higher than Africa's 48%. L.A.'s Q_1 is more literate than Africa's Q_3.

Box Plot Practice Name: _____

1. Refer to the table below, which gives the mean length in days of a hospital stay by state in 1993.

State	Mean Stay	State	Mean Stay	State	Mean Stay
MT	18.7	WY	7.5	AR	6.7
MI	13.2	RI	7.3	DE	6.7
ND	13.0	VT	7.3	GA	6.7
PA	13.0	CO	7.2	OK	6.7
SD	12.4	IL	7.2	SC	6.7
NJ	11.6	NC	7.2	ID	6.6
MA	10.7	OH	7.2	LA	6.5
MN	10.3	AL	7.1	NV	6.4
NY	10.2	WV	7.1	UT	6.3
NE	9.7	FL	7.0	TX	6.2
CT	8.6	MD	7.0	CA	6.1
HI	8.5	MS	7.0	AK	5.9
WI	8.2	VA	7.0	NM	5.9
ME	7.8	IN	6.8	AZ	5.8
IA	7.6	KY	6.8	OR	5.8
KS	7.6	NH	6.8	WA	5.8
MO	7.5	TN	6.8		

① median ② Q_1, Q_3 ③ IQR ④ outliers

outliers

median

Q_1

Q_3

a. Give the five-number summary for the data set and make a box plot to display the data.

min : 5.8

Q_1 : 6.7

median : 7.1

Q_3 : 8.2

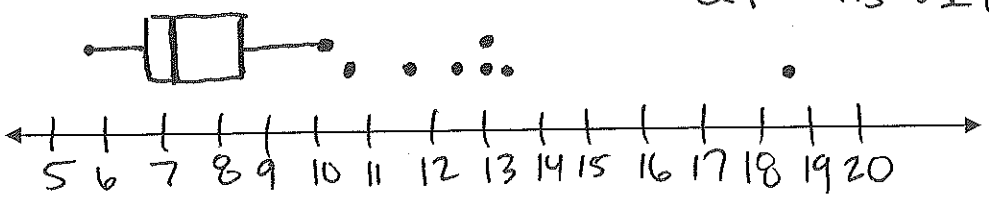
max: 18.7 (use 10.3)

highest # under 10.45

$IQR = Q_3 - Q_1 = 8.2 - 6.7 = 1.5$

$Q_3 + 1.5 \cdot IQR = 8.2 + 1.5 \cdot 1.5 = 10.45$

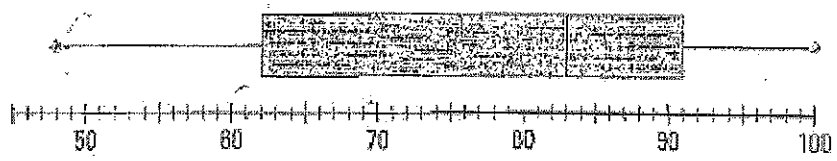
$Q_1 - 1.5 \cdot IQR = 6.7 - 1.5 \cdot 1.5 = 4.45$ (none)



b. What percent of hospital stays are between 6.7 and 8.2 days?

50%

2. Refer to the following box plot of student test scores on the January final last year.



a. What is the median score?

83

b. What is the interquartile range?

IQR = 91 − 62 = 29

c. What percent of the students scored between 62 and 91?

50%.

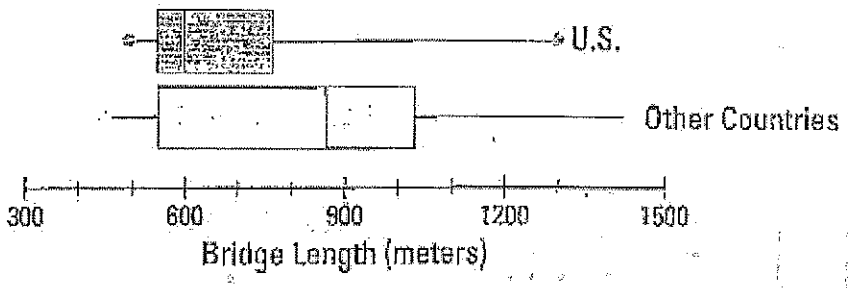
d. What is the interval of scores of students who ranked below the lower quartile?

48 → 62

e. Between which two quartiles is there the greatest spread?

Q_1 and the median

3. Use the box plots below of the lengths in meters of bridges in the U.S. and other countries in 1996. The 20 longest U.S. bridges are compared with the 20 longest bridges outside the U.S.



a. Each of the ten longest bridges in the U.S. is longer than _____ meters.

b. True or false. The shortest of the 40 bridges in the data set is in the U.S.