1.3 Day 2

- ✓ IDENTIFY outliers using the 1.5 × IQR rule.
- ✓ MAKE and INTERPRET boxplots of quantitative data.
- ✓ Use boxplots and numerical summaries to COMPARE distributions of quantitative data.

Informal Hands on Activity to get feel of things

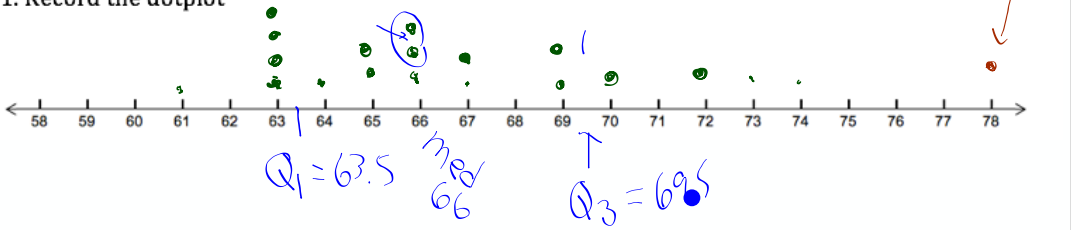
We'll formalize after

Lesson 1.3: Where Do I Stand?

How does my height compare with other AP Stats students?

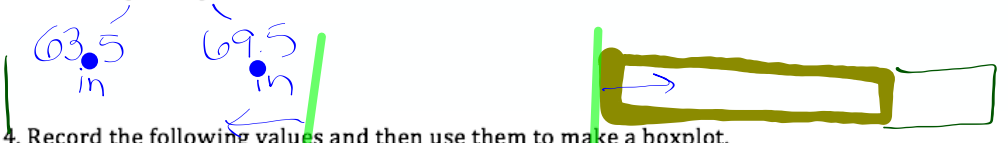
In pairs, measure each other's height, rounded to the nearest inch.
Record your height on the dotplot at the front of the room (females use red, males use green).
Make a line at the front of the room, shortest to tallest.

1. Record the dotplot



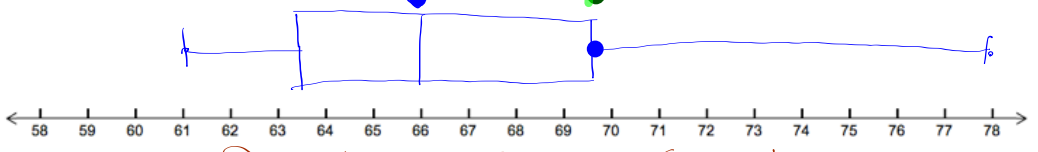
2. What is the median height? 66 inches

3. What is Q_1 and Q_3?



4. Record the following values and then use them to make a boxplot.

Minimum: 61 Q_1: 63.5 Median: 66 Q_3: 69.5 Maximum: 78



IQR = 69.5 − 63.5 = 6 inches

4. The **interquartile range** (or IQR) is defined as $Q_3 - Q_1$. Find the IQR. Where do you see the IQR in the boxplot?

5. An **outlier** is a data value that is way too small or way too big (using the rules below). Are there any outliers? Show your work.

Way too small $< Q_1 - 1.5IQR$ $63.5 - 1.5(6) = 54.5$ in

Way too big $> Q_3 + 1.5IQR$ $69.5 + 1.5(6) = 78.5$

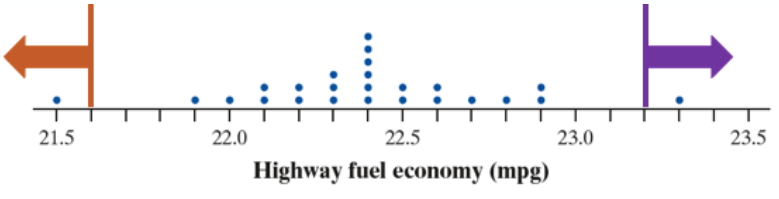
Outliers

Why look for outliers?

1. They might be inaccurate data values.
2. They can indicate a remarkable occurrence.
3. They can heavily influence the values of some summary statistics, like the mean, range, and standard deviation.

How do we find the cut off values?

Q_1 = 22.2 mpg
Q_3 = 22.6 mpg
IQR = 0.4 mpg



Although there are several rules for outliers, one of the most common rules is the 1.5 × IQR rule.

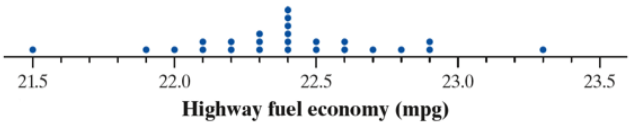
HOW TO IDENTIFY OUTLIERS: THE 1.5 × *IQR* RULE

Call an observation an outlier if it falls more than $1.5 \times IQR$ above the third quartile or below the first quartile. That is,

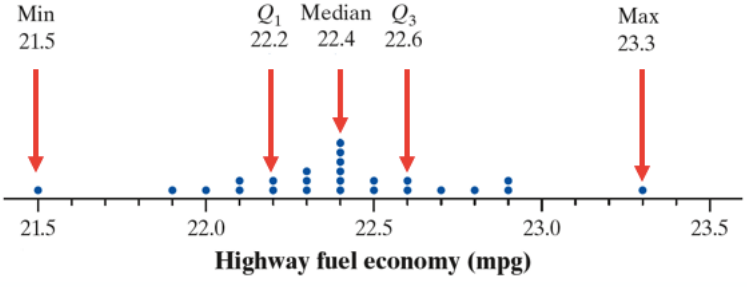
Low outliers $< Q_1 - 1.5 \times IQR$ High outliers $> Q_3 + 1.5 \times IQR$

Watch

The **five-number summary** of a distribution of quantitative data consists of the minimum, the first quartile Q_1, the median, the third quartile Q_3, and the maximum.

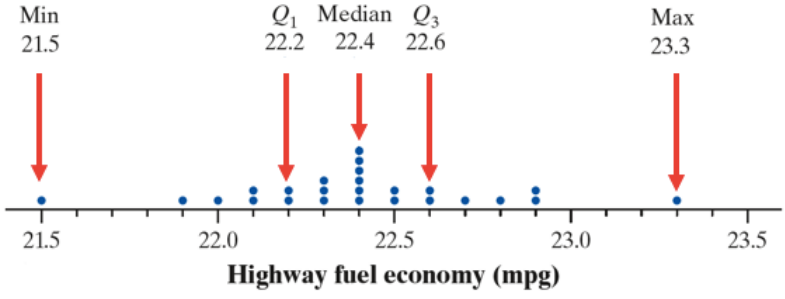


The **five-number summary** of a distribution of quantitative data consists of the minimum, the first quartile Q_1, the median, the third quartile Q_3, and the maximum.



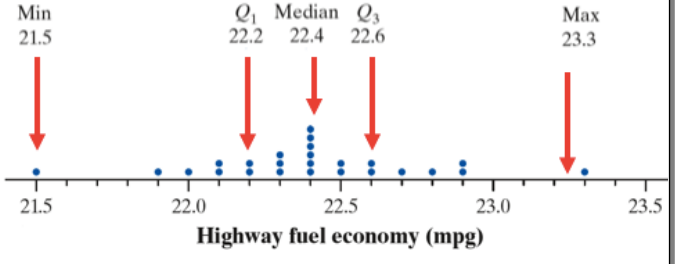
The **five-number summary** of a distribution of quantitative data consists of the minimum, the first quartile Q_1, the median, the third quartile Q_3, and the maximum.

A **boxplot** is a visual representation of the five-number summary.



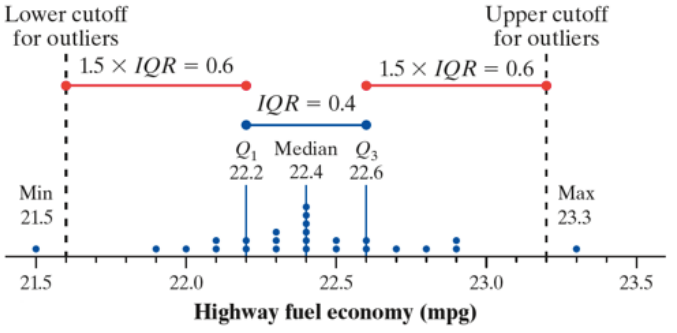
How to Make a Boxplot

- Find the five-number summary.
- Identify outliers using the 1.5 × IQR rule.
- Draw and label the horizontal axis.
- Scale the axis.
- Draw a box.
- Mark the median.
- Draw whiskers.



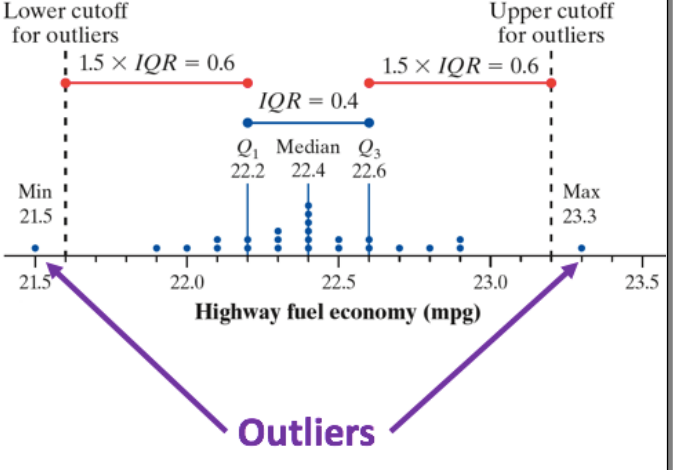
How to Make a Boxplot

- Find the five-number summary.
- Identify outliers using the 1.5 × IQR rule.
- Draw and label the horizontal axis.
- Scale the axis.
- Draw a box.
- Mark the median.
- Draw whiskers.



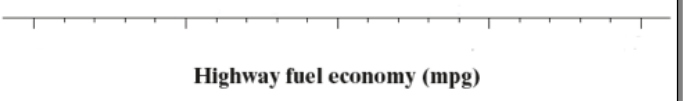
How to Make a Boxplot

- Find the five-number summary.
- Identify outliers using the 1.5 × IQR rule.
- Draw and label the horizontal axis.
- Scale the axis.
- Draw a box.
- Mark the median.
- Draw whiskers.



How to Make a Boxplot

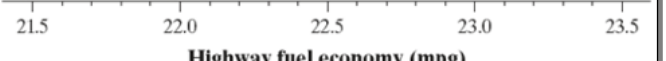
- Find the five-number summary.
- Identify outliers using the 1.5 × IQR rule.
- **Draw and label the horizontal axis.**
- Scale the axis.
- Draw a box.
- Mark the median.
- Draw whiskers.



Highway fuel economy (mpg)

How to Make a Boxplot

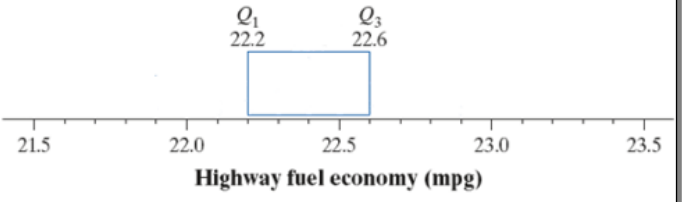
- Find the five-number summary.
- Identify outliers using the 1.5 × IQR rule.
- Draw and label the horizontal axis.
- **Scale the axis.**
- Draw a box.
- Mark the median.
- Draw whiskers.



Highway fuel economy (mpg)

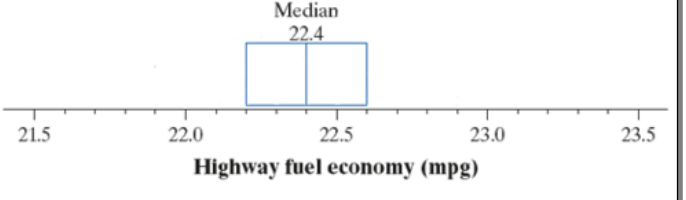
How to Make a Boxplot

- Find the five-number summary.
- Identify outliers using the 1.5 × IQR rule.
- Draw and label the horizontal axis.
- Scale the axis.
- **Draw a box.**
- Mark the median.
- Draw whiskers.



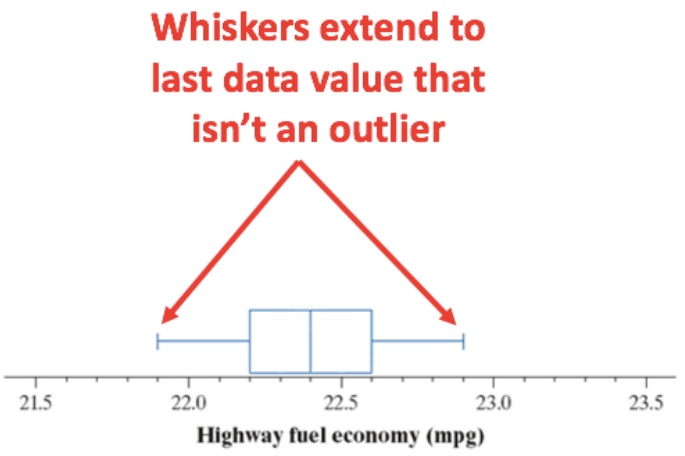
How to Make a Boxplot

- Find the five-number summary.
- Identify outliers using the 1.5 × IQR rule.
- Draw and label the horizontal axis.
- Scale the axis.
- Draw a box.
- **Mark the median.**
- Draw whiskers.



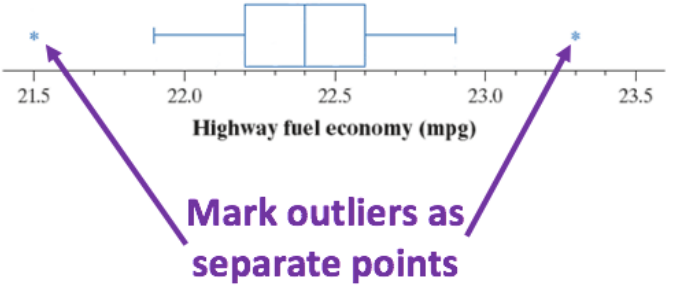
How to Make a Boxplot

- Find the five-number summary.
- Identify outliers using the 1.5 × IQR rule.
- Draw and label the horizontal axis.
- Scale the axis.
- Draw a box.
- Mark the median.
- Draw whiskers.



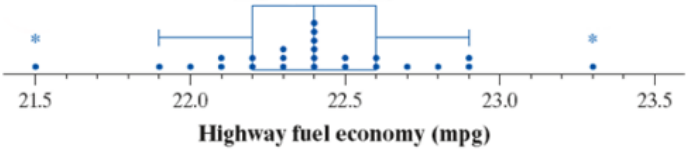
How to Make a Boxplot

- Find the five-number summary.
- Identify outliers using the 1.5 × IQR rule.
- Draw and label the horizontal axis.
- Scale the axis.
- Draw a box.
- Mark the median.
- Draw whiskers.



CAUTION:

- Boxplots do not display each individual value in a distribution.
- Boxplots don't show gaps, clusters, or peaks.

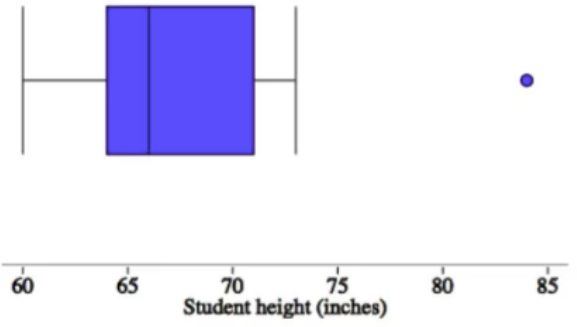


Enter the height data in L1

Make a box plot that shows outliers

Modified boxplot vs boxplot

Many students will learn in middle school that a "modified box" plot will show outliers as a special character (star or * or a blue dot)

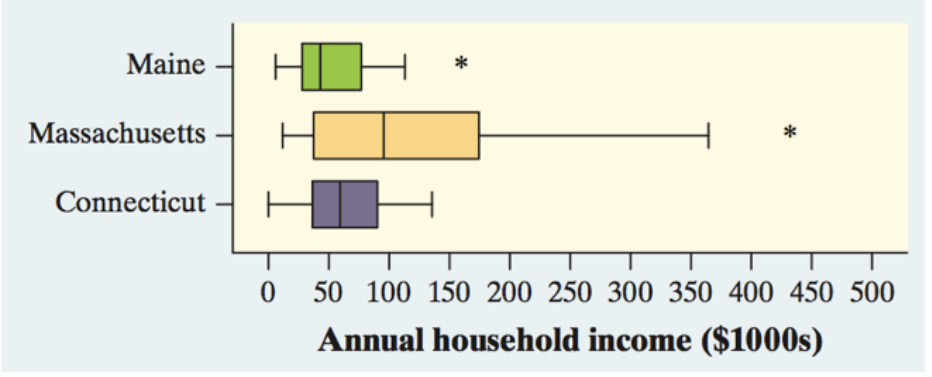


When John Tukey invented the boxplot in 1970, he used the above version (with outliers shown as distinct points). So, in reality, this is not a "modified" box plot...it is the "OG boxplot". Any boxplot students encounter in AP Statistics will be in this form. In other words, if students see a boxplot that does not show a distinct outlier, they can assume there is no outlier in the data set (according to the "way too small, way too large" rule above).

Comparing Distributions with Box Plots

Check Your Understanding:

The following boxplots show the total income of 40 randomly chosen households each from Connecticut, Maine, and Massachusetts, based on U.S. Census data from the American Community Survey. Compare the distributions of annual incomes in the three states.



shape:
outliers:
center:
variability:

shape:
outliers:
center:
variability:

COMPARE! DON'T LIST!

"less than"
"greater than"
"similar to"

Compare shape.
Compare outliers.
Compare center.
Compare variability.

+ Context

Annual household income ($1000s)

context is now established

The shape of the distribution of income for Connecticute is roughly symmetric while Maine incomes are slightly skewed right and Massachusetts incomes have a stronger right skew (note: hard to tell total shape with boxplot)

OCV + context

The center is highest for Massachusetts, followed by Connecticut, then Maine with the lowest.

The variability is highest for Massachusetts, with Maine & Connecticut having similar variability

Maine & Massachusetts each have a high outlier; while Connecticut has none.

LCQ

Assignment

1.3....109, 111, 113, 115, 123-126