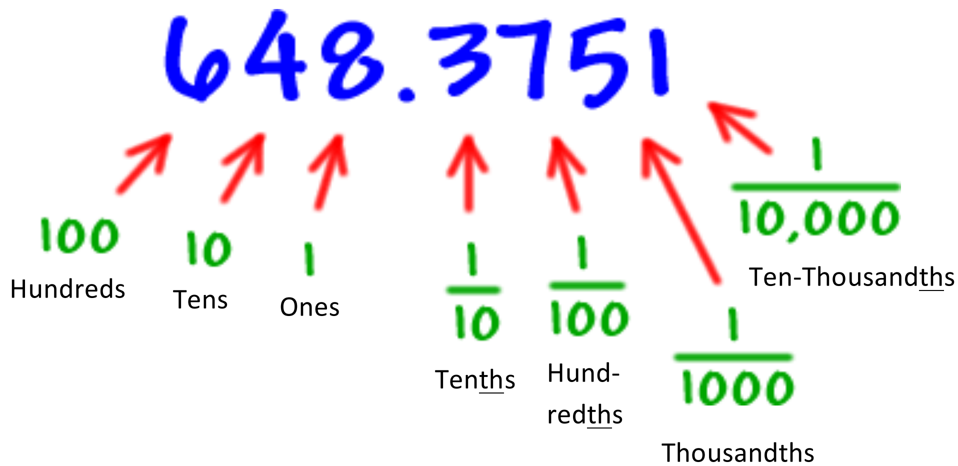
Part 1: Rounding

Rounding allows us to take numbers with many, many digits and re-write them in a way that makes them easier to understand and digest. Since statistics is all about *communicating* data and information, rounding is essential!

Rounding is an art that requires good intuition and number sense. If you don't round enough, your numbers will be hard to interpret. But if you round too much, you are sacrificing the accuracy of your information. Your job as a statistician is to round in such a way that creates a good balance between the two.

Just always remember: 5 and above round up, 4 and below round down!



EXAMPLES:

1. Round 29.4319 to the nearest hundredth

*"Hundredths" is the second decimal place, which gives 29.43. The next number is 1, so we round down; the final answer stays **29.43***

2. Round 4.39986 to the nearest ten thousandth

*"Thousandths" is the third decimal place, which gives 4.399. However, since next number is 8, we round up; the final answer becomes **4.400** (the next number up from 399 is 400)*

3. Round .037292319586

*This one is a judgment call. In AP Stat, we usually round to 2, 3, or 4 decimal places, depending on the situation. You should round <u>at least</u> to the nearest hundredth (**.04**), but you can do to the nearest thousandth (**.037**), or nearest ten-thousandth (**.0373**). All of these are technically correct! If you went further than this (such as .037293), you aren't wrong, but too many digits makes the number hard to contextualize and interpret – I would advise against it.*

Practice Problems – Check the answers to the odd-numbered ones in the back of the packet!

Round the following to the given decimal point.

1. Round 12.842 to the nearest tenth	2. Round 0.4892745 to the nearest hundredth	3. Round 0.0342119 to the nearest thousandth	4. Round 0.06049822 to the nearest ten thousandth

Round the following. Use your best judgment.

5. Round 25.6895234	6. Round 0.033231532	7. Round 0.00279625	8. Round 0.63636363…

Part 2: Fractions, Decimals, and Percentages

In Statistics, we often deal with ***proportions*** – an amount out of a total. Proportions can be expressed as fractions, decimals, or percentages (which are just proportions out of 100).

Converting between forms of proportions

Fraction → Decimal: This is easy: just *divide* your numerator and denominator! Be aware that you may have to *round* the answer

Examples:

1. Convert $\frac{11}{14}$ to a decimal

   *11÷14 = 0.7857142857…you must round this! It rounds nicely at 4 decimal places, or **0.7857.***
   *You could also do **0.79** or **0.786** if you would like.*

Percentage → Decimal: In order to use a percentage in an equation or a calculation, you MUST convert it into a decimal! Since percentages are always out of 100, just divide the percentage by 100!
**A shortcut to this is moving the decimal two places to the left!*

Examples:

1. Convert 35% to a decimal

   *35 ÷ 100 = **0.35.** Likewise, moving the decimal point 2 places to the left means **35.0 → .350,** or .35*

2. Convert 6.85% to a decimal

   *6.85 ÷ 100 = **0.0685.** Notice a ZERO between the decimal point and the 6 (you need to insert a zero in order to move 2 decimal places to the left)*

Decimal → Percentage: To turn a decimal into a percentage, simply do the opposite of the percentage-to-decimal procedure. You can *multiply* the decimal by 100, or move the decimal two places to the *right*.

Examples:

1. Convert 0.02 to a percentage

   *0.02 • 100 = **2%.** Likewise, moving the decimal point 2 places to the right means **0.02 → 2.0**, or 2*

2. Convert $\frac{5}{12}$ to a percentage

   *5÷12 = 0.4166666...let's round to 0.4167. Then, 0.4167 • 100 (or going 2 decimal places to the right) = **41.67%.***

Using Percentages

*To find a percentage of a number (such as 25% of 84), **multiply** the percentage (*as a decimal*) by the number

Examples:

1. Find 25% of 72

   *25% is 0.25, and 0.25•72 = **18***

2. Find 3.9% of 749

   *3.9% is 0.039, and 0.039•749 = **29.211***
   NOTE: If you need a whole number, round to **29**

Practice Problems – *Check the answers to the odd-numbered ones in the back of the packet!*

1. Convert $\frac{13}{3}$ to a decimal	2. Convert $\frac{41}{563}$ to a decimal	3. Convert 70% to a decimal	4. Convert 8% to a decimal
5. Give 22.45% as a decimal	6. Give 100% as a decimal	7. Convert 0.672 to a percentage	8. Convert 0.0052 to a percentage
9. Convert $\frac{4}{25}$ to a percentage	10. Convert $\frac{11}{285}$ to a %	11. What is 17.2% of 89?	12. What is 3% of 446?

Part 3: Summary Statistics – Center and Spread

A *statistic* is a number that gives information about a set of data. Common examples include mean, median, mode (which we won't worry about in AP Stat), range, standard deviation, and more!

SYMBOLOGY
In statistics, we use a variety of *symbols* to represent statistics. Sometimes, the symbol used depends on whether we are talking about a **population** or a **sample** (select members of a given population)

	Mean	Standard Deviation	Median	Number of data points you have
Population	$\boldsymbol{\mu}$ ("mu")	$\boldsymbol{\sigma}$ ("sigma")	*No symbol* (Often abbreviated "**Med**.")	$\boldsymbol{n}$
Sample	$\overline{\boldsymbol{x}}$ ("x-bar")	$\boldsymbol{s}$		

Measures of CENTER
The *center* of a data set lets us understand the "average" or "typical" value of a number in that data set. There are two main measures of center: mean and median.

MEAN

Add up all data points, then divide by the number of data points.

- $\boldsymbol{\mu}$ **(or** $\overline{\boldsymbol{x}}$**)** $= \frac{\sum x}{n}$ (The Σ symbol means "Sum")
- "Sum of data points over the number of data points"

Example 1: *Science grades of a sample of 15 juniors:*
91, 87, 66, 74, 85, 98, 43, 88, 77, 62, 83, 91, 89, 52, 100

This is a SAMPLE, so $\overline{x} = \frac{\sum x}{n} = \frac{1186}{15} = \mathbf{79.07}$

Example 2: *Heights of all 6 people in a family (inches):*
47, 58, 61, 65, 68, 70

This is the POPULATION, so $\mu = \frac{\sum x}{n} = \frac{369}{6} = \mathbf{61.5}$

MEDIAN

The *middle number* of the data set, assuming that the data points are **in order** (smallest to largest)

- If there are 2 numbers in the middle, find the *mean* of those two numbers!
- A nice **trick** for finding the <u>*position*</u> of the median is to use $\frac{n+1}{2}$

Example 1: *Science grades of a sample of 15 juniors:*
91, 87, 66, 74, 85, 98, 43, 88, 77, 62, 83, 91, 89, 52, 100

$\frac{n+1}{2} = \frac{15+1}{2} = \mathbf{8.}$ Median is the **8th** number (IN ORDER)

43, 52, 62, 66, 74, 77, 83, **85**, 87, 88, 89, 91, 91, 98, 100

Example 2: *Heights of all 6 people in a family (inches):*
47, 58, 61, 65, 68, 70

$\frac{n+1}{2} = \frac{6+1}{2} = \mathbf{3.5.}$ Median is **between the 3rd & 4th** number

47, 58, 61, 65, 68, 70; Average = $\frac{61+65}{2} = \mathbf{63}$

Measures of SPREAD
The *spread* of a data set tells us whether the data points are far apart or clustered together. The most important measure of spread is **standard deviation,** which is the <u>**typical distance of the data points from the mean**</u>. Other measures of spread, such as Range and IQR, will be discussed in Part 5 of the summer work.

The formulas for Standard Deviation are as follows. Note that they are slightly different for a population and a sample (the sample one will be slightly larger to account for the fact that the sample doesn't include all members of a population)

Population: $\boldsymbol{\sigma} = \sqrt{\frac{\Sigma(x_i-\mu)^2}{n}}$ Sample: $\boldsymbol{s} = \sqrt{\frac{\Sigma(x_i-\overline{x})^2}{n-1}}$

You will NOT have to calculate Standard Deviation by hand in this course!

What you *will* have to do, however, is be able to *interpret* and *compare* the Standard Deviations of different data sets:

- ***Larger*** **Standard Deviation:** The data is more spread out (points are typically further from the mean)
- ***Smaller*** **Standard Deviation:** The data is closer together (points are typically closer to the mean)

<u>**Example:**</u>

Data Set 1: 1, 2, 3, 17, 18, 19; $\mu = 10$, $\sigma = 8.04$ **Data Set 2:** 7, 8, 9, 11, 12, 13; $\mu = 10$, $\sigma = 2.16$

Notice how Data Set 1 is more spread out, while Data Set 2 is closer together. This is reflected in the fact that Set 1's Standard Deviation (8.04) is higher than Set 2's Standard Deviation (2.16)

Practice Problems – *Check the answers to the odd-numbered ones in the back of the packet!*

1. Find the mean and median of the following data set. **Show work** when appropriate!

Teaching experience of all Uni history teachers (*n* = 15): 1, 3, 3, 3, 4, 4, 5, 5, 5, 6, 7, 7, 18, 23, 26

Symbol for mean: ________ *Value of mean:* ________ *Position of Median:* ________ *Value of Median:* _______

2. Find the mean and median of the following data set. **Show work** when appropriate!

Weights of 8 randomly-selected chickens on a farm (in pounds): 5.4, 5.7, 6.2, 6.9, 7.2, 7.2, 8.1, 9.0

Symbol for mean: ________ *Value of mean:* ________ *Position of Median:* ________ *Value of Median:* _______

3. Find the mean and median of the following data set. **Show work** when appropriate!

Temperature readings on all thermostats in an office building: 71, 72, 72, 74, 68, 74, 71, 72, 69, 76

Symbol for mean: ________ *Value of mean:* ________ *Position of Median:* ________ *Value of Median:* _______

4. List the following data sets in order from *least spread out* to *most spread out*. Then, **write** 1-2 sentences explaining how you could tell.

Uni History Teachers: $\sigma = 7.02$ **Chickens:** $s = 1.21$ **Thermostats:** $\sigma = 2.26$

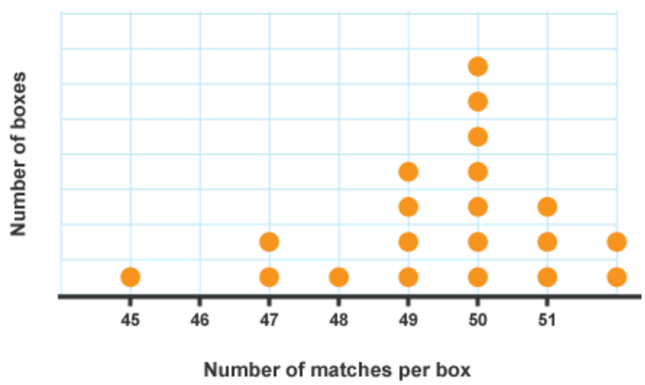
Part 4: Graphing Data – Dotplots and Stem-and-Leaf Plots

Statistics such as mean, median, and standard deviation are very useful in summarizing data and giving overall trends. But they don't tell the full story. By making a *graph* of the data, we can go *beyond* the numbers and see *shapes* and *patterns* in the data. Shown below are two common ways in which to graph data

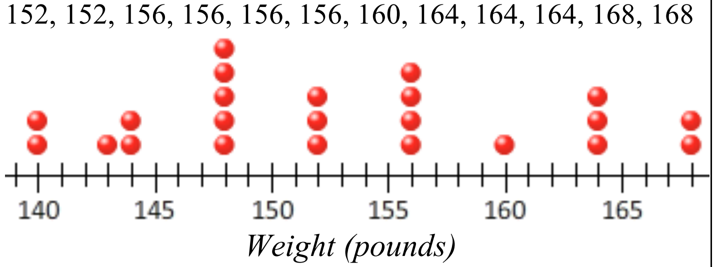
Dotplots

- Make an AXIS on the bottom (you can go by 1s, 2s, 5s, 10s…whatever makes sense for the data!)
- Put one dot for each data point on the axis. If there is more than one data point for a given value, *stack* the dots!

Example: *Number of matches in 20 randomly-selected boxes.*
45, 47, 47, 48, 49, 49, 49, 49, 50, 50, 50, 50, 50, 50, 50, 51, 51, 51, 52, 52



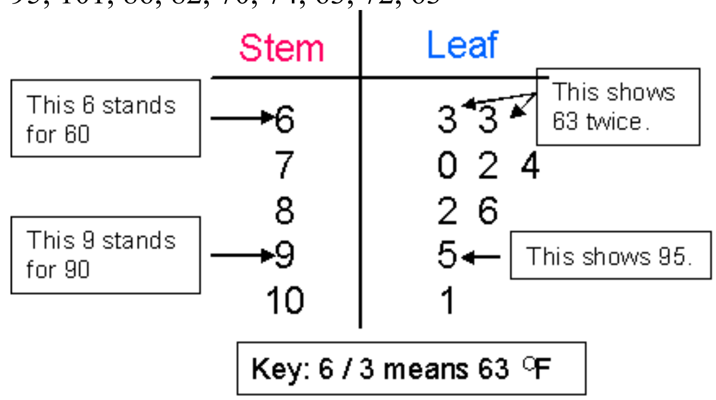
Example: *Weights of players on a high school baseball team*
140, 140, 143, 144, 144, 148, 148, 148, 148, 148, 152, 152, 152, 156, 156, 156, 156, 160, 164, 164, 164, 168, 168



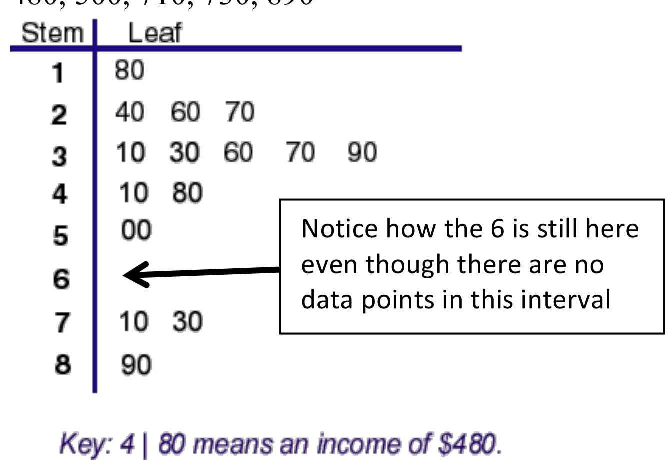
Stem-and-Leaf Plots (also called Stemplots)

- Use a KEY to determine what the stems and leaves are worth
- DO NOT SKIP STEMS. If there are no data points for that stem, just keep the stem there and put no leaves after it. Skipping the stem will alter what the stemplot looks like.

Example: *Temperatures at OU football games, 2009*
95, 101, 86, 82, 70, 74, 63, 72, 63



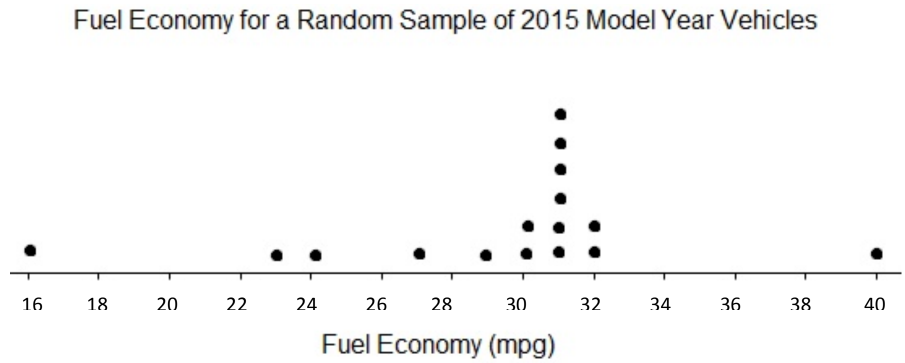
Example: *Gross National Product (per capita) of West African countries*
180, 240, 260, 270, 310, 330, 360, 370, 390, 410, 480, 500, 710, 730, 890



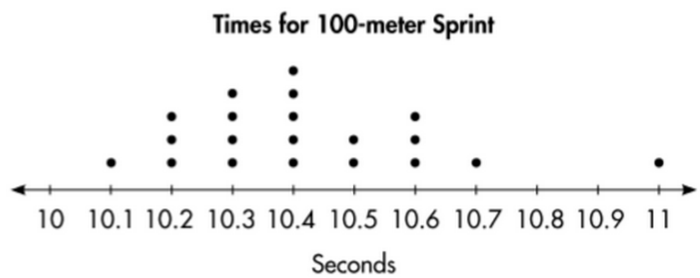
Practice Problems are on next page!

Practice Problems – *Check the answers to the odd-numbered ones in the back of the packet!*

1. List all the data points in the dotplot:



2. Using the dotplot shown, find the **mean** and **median** 100-meter sprint time



3. The following stemplot shows the final exam scores of a class with 10 students. List the scores of each student:

stem	leaf
6	9
7	
8	7 8 8 9
9	0 6 7 7
10	0

Key: 6|8 means 68

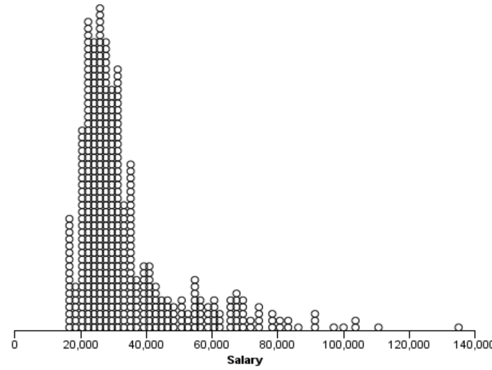
4. Find the **mean** and **median** of the data shown in the stemplot. ***NOTE:*** *Look at the key carefully!*

Stem	Leaf
2	0 2 3 6
3	2 3 5 6 7
4	6 8 9
5	4 7
6	2
7	3

KEY: 4|6 = 4.6

Part 5: Graphing Data – Box and Whisker Plots

We can see that dotplots and stemplots are effective ways of graphing and analyzing data sets. But what if a data set had hundreds, or even *thousands,* of data points? Take the dotplot shown at the right, for instance – analyzing each and every dot would take *forever!*

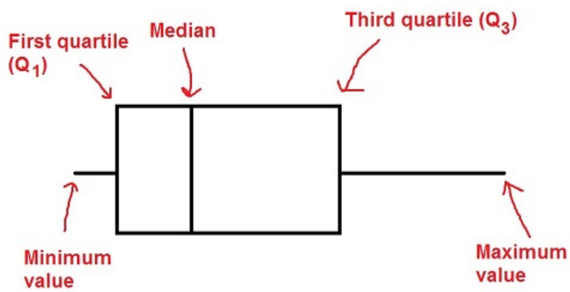


This is why it is important to be able to use *summary statistics* (such as mean, median, and standard deviation) to analyze our data. But in doing so, we lose our ability to *graph* the data and *see* what's going on.

Or do we?

This is the beauty of **Box and Whisker Plots** (or "Boxplots" for short).
A box and whisker plot is a graphical representation of just 5 numbers in a data set:

- **Minimum (Min):** Smallest value in the data set
- **Lower Quartile (Q1):** Midpoint between the Minimum and the Median
- **Median (Med):** Midpoint of the entire data set (as was discussed in Part 3 of the Summer Work)
- **Upper Quartile (Q3):** Midpoint between the Median and the Maximum
- **Maximum (Max):** Largest value in the data set



NOTE: The Mean and Standard Deviation of a data set are **NOT** included in a boxplot. **DO NOT** try to include them or talk about them!

It is important to note that **no matter how far apart or close together these 5 numbers are, ¼ (25%) of the data is in each of the 4 sections of the boxplot** (lower whisker, lower half of the box, upper half of the box, and upper whisker)

Boxplots also allow us to measure **Spread** in 2 ways:

- **Range:** Difference between the extremes of the data (Maximum minus Minimum)
- **Interquartile Range (IQR):** Difference between the two *quartiles* (Q3 – Q1)

Example 1: Points scored by Russell Westbrook, 2016 NBA Playoffs

14, 14, 16, 19, 19, 24, 25, 26, 27, 28, 28, 29, 30, 31, 31, 35, 36, 36

**Finding MEDIAN:* (18 games. $\frac{18+1}{2}$ = 9.5, so average the 9^{th} and 10^{th} numbers)

14, 14, 16, 19, 19, 24, 25, 26, **27, 28**, 28, 29, 30, 31, 31, 35, 36, 36

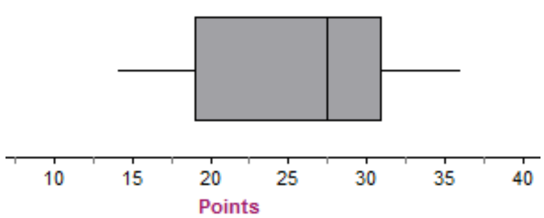
Median = 27.5

**Finding QUARTILES:* There are 18 data points, so *split* the data set in half! *Find the middle of <u>each half!</u>*

Q1 — 14, 14, 16, 19, **19**, 24, 25, 26, 27

Q3 — 28, 28, 29, 30, **31**, 31, 35, 36, 36

***BOXPLOT:** Min = 14, Q1 = 19, Med = 27.5, Q3 = 31, Max = 36



RANGE: 36 – 14 = **<u>22</u>**

IQR: 31 – 19 = **<u>12</u>**

Example 2: Ages of 9 employees in an office

37, 24, 51, 46, 62, 28, 35, 49, 55

Finding MEDIAN:* (9 people. $\frac{9+1}{2}$ = 5, it's the 5th number. **Remember that numbers must be <u>in order</u>!)

24, 28, 35, 37, **46,** 49 51, 55, 62

Median = 46

Finding QUARTILES:* There are 9 data points, so *split* the data set in half! ***NOTE:* *If there is one number that serves as the median, as with this data set,* ***it is not included in <u>either</u> half!***

Q1		*Q3*
24, **28, 35**, 37	46 (not included)	49, **51, 55,** 62
Average: 31.5		*Average: 53*

***BOXPLOT:** Min = 24, Q1 = 31.5, Med = 46, Q3 = 53, Max = 62

RANGE: 62 – 24 = **<u>38</u>**

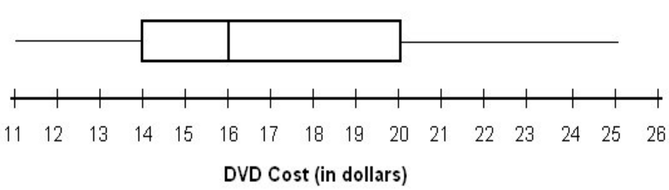
IQR: 53 – 31.5 = **<u>21.5</u>**

Practice Problems – *Check the answers to the odd-numbered ones in the back of the packet!*

1. Construct a box and whisker plot for the following data set. **Be sure to include an <u>axis</u>** like the ones in the examples!

17, 21, 24, 26, 31, 33, 36, 37, 41, 48

2. Analyze the following boxplot:



Minimum: Range:

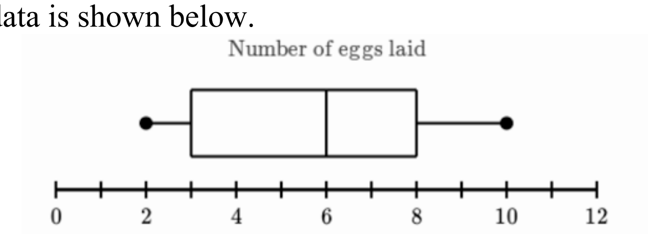
Q1: IQR:

Median:

Q3:

Maximum:

3. A farmer has 168 laying hens. He recorded how many eggs each hen laid in one week. A boxplot of the data is shown below.



HINT: *Remember that each of the 4 sections of the boxplot has ¼, or 25%, of the data points.*

Find the number of laying hens, out of 168, that laid…

A. More than 8 eggs _____ B. Fewer than 6 eggs ______

C. Between 3 & 8 eggs ______ D. Between 2 & 8 ______

Selected Answers

NOTE: *All appropriate work must be shown in order to earn full credit on the summer assignment!*

Part 1: Rounding

1) 12.8	3) 0.034	5) 25.7 *or* 25.69 *or* 25. 690 *or* 25.6895	7) 0.0028

Part 2: Fractions, Decimals, and Percentages

NOTE: *Your answers may be slightly different due to rounding. That's fine, as long as you rounded correctly*

1) 4.33	3) 0.7	5) 0.2245	7) 67.2%	9) 16%	11) 15.308

Part 3: Summary Statistics – Center and Spread

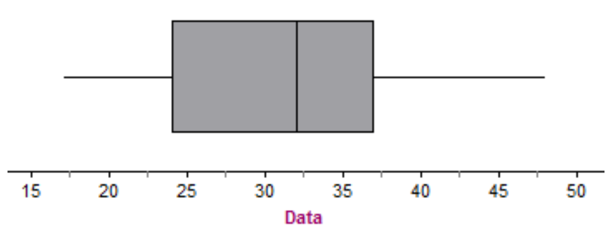
1) Symbol of Mean: μ Value of Mean: 8 years Position of Median: 8^{th} Value of Median: 5 years	3) Symbol of Mean: μ Value of Mean: 71.9 degrees Position of Median: 5^{th} & 6^{th} Value of Median: 72 degrees

Part 4: Graphing Data – Dotplots and Stem-and-Leaf Plots

1. 16, 23, 24, 27, 29, 30, 30, 31, 31, 31, 31, 31, 31, 32, 32, 40
3. 69, 87, 88, 88, 89, 90, 96, 97, 97, 100

Part 5: Graphing Data – Box-and-Whisker Plots

1. Min = 17, Q1 = 24, Med = 32, Q3 = 37, Max = 48



3. A. 42 chickens C. 84 chickens