Name:_____

Types of Data

Univariate Data

Single-variable data where we're only observing one aspect of something at a time. With single-variable data, we can put all our observations into a list of numbers.

Qualitative

Data seen as categories sometimes known as categorical data.

Quantitative

Data that describes information that can be counted or measured.

Discrete

Data that has an exact numerical value; counted; seen in ***bar charts***

Continuous

Data that can be measured; takes on any value within a range; seen in ***histograms***

Bar Chart or Histogram?

Classify each of the following quantitative data types as either discrete or continuous.

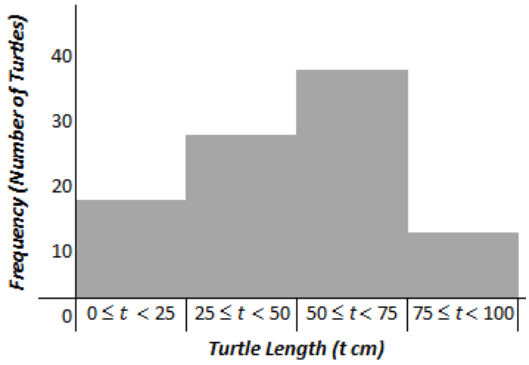
1. The number of sea turtles tagged by researchers.
   Discrete
2. The length of the sea turtle.
   Continuous
3. The time taken to catch, tag, and release the sea turtle.
   Continuous
4. The number of biologists working to tag the sea turtle.
   Discrete
5. The number of sea turtle nests along the Atlantic Coast in 2014.
   Discrete
6. The time sea turtle eggs incubate before hatching.
   Continuous

An easy way to look for patterns in a large set of data is to create a frequency table, bar chart, or histogram.

A researcher catches 100 sea turtles. The turtles are measured and then released. The lengths, t *cm*, of these turtles are shown in the frequency table.

Length (t *cm*)	$0 \le t < 25$	$25 \le t < 50$	$50 \le t < 75$	$75 \le t < 100$
Frequency (Number of Turtles)	17	28	40	15

Determine whether the data should be displayed in a bar chart or a histogram and create the appropriate display. *A histogram should be created since length is a continuous variable.*



Example adapted from International Baccalaureate released test items.

Creating a Box Plot

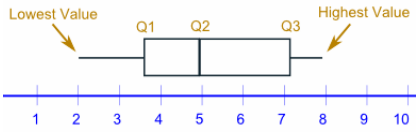
What are they and why are they useful?	How do I make one?	What about outliers?

A box plot is a type of graph used to represent univariate data.

They are useful because they show:

- _*Center*_____
- _*Spread*_____
- _*Distribution*_____
- _*Outliers*_____

A box plot is a graph of a data set along a number line where the box represents the middle 50% of data and the whiskers extend to the maximum and minimum values to represents the other 50% of the data.



Step 1:
Find the 5-Number Summary for the data: the minimum, maximum, median (Q2), first quartile (Q1), and third quartile (Q3).

Step 2:
Construct a consistent scale with values that include the minimum and maximum.

Step 3:
Construct a box (rectangle) extending from Q1 to Q3 and draw a vertical line in the box at the median value (Q2).

Step 4:
Draw lines extending outward to the minimum and maximum values.

An outlier is a value that is located _ *very far away* _____ from almost all of the other values.

Outliers can have a dramatic effect on the *mean*, *standard deviation*, and *distribution* of a data set.

Mathematically, an outlier is a value that is:

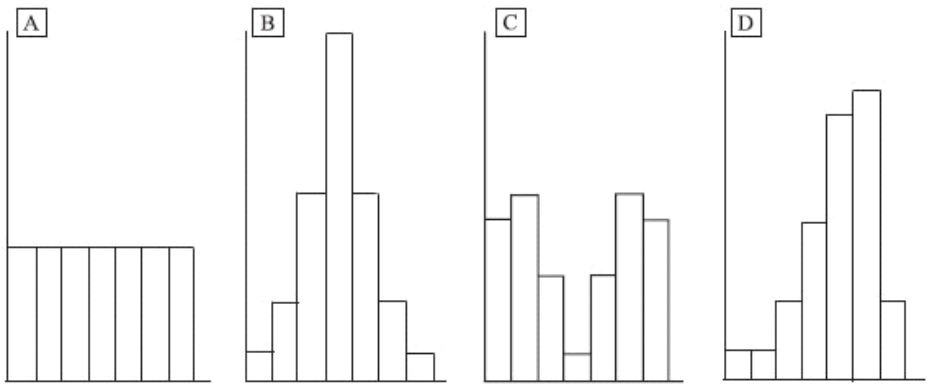
- Above Q3 (or below Q1) by an amount greater than __$1.5 \cdot IQR$_____

The IQR is the Interquartile Range, or *difference between Q3 and Q1*.

The box plot is modified by extending the whiskers to the minimum or maximum value that is not an outlier.

Example

The four populations A, B, C and D are the same size and have the same range. Frequency histograms for the four populations are given below.



Each of the three box and whisker plots below corresponds to one of the four populations. Write the letter of the correct population above each plot.

Example adapted from International Baccalaureate released test items

Data Analysis
Sea Turtle Nesting Data

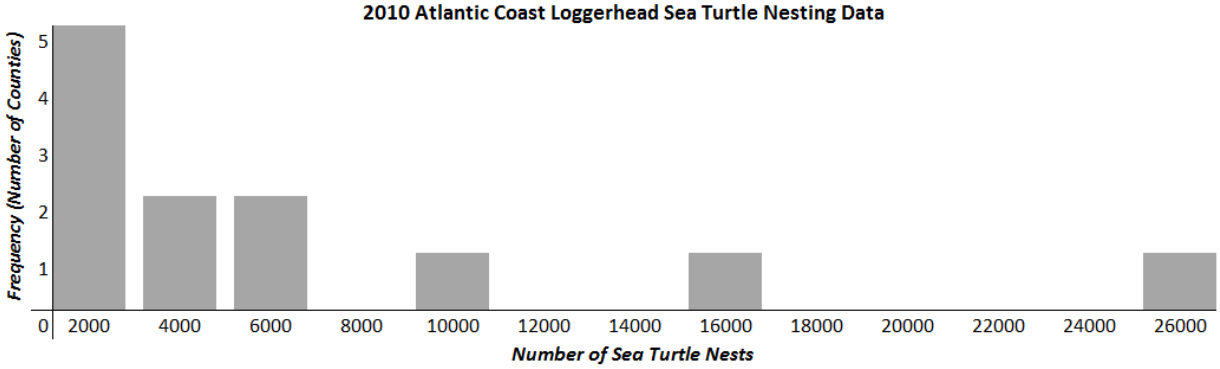
"The Statewide Nesting Beach Survey (SNBS) program was initiated in 1979 under a cooperative agreement between the Florida Fish and Wildlife Conservation Commission (FWC) and the U.S. Fish and Wildlife Service. Its purpose is to document the total distribution, seasonality and abundance of sea turtle nesting in Florida. Three species of sea turtles, the loggerhead (Caretta caretta), the green turtle (Chelonia mydas), and the leatherback (Dermochelys coriacea), nest regularly on Florida's beaches. Two other species, the hawksbill (Eretmochelys imbricata) and Kemp's ridley (Lepidochelys kempii), nest infrequently. All five species are listed as either threatened or endangered under the Endangered Species Act."

-Florida Fish and Wildlife Conservation Commission

Small Group Task

Objective: Classify, organize, represent, and analyze a set of univariate quantitative data.
Materials: Notebook paper, poster paper (or use the back of this sheet), markers/colored pencils, rulers, stapler/tape, calculators

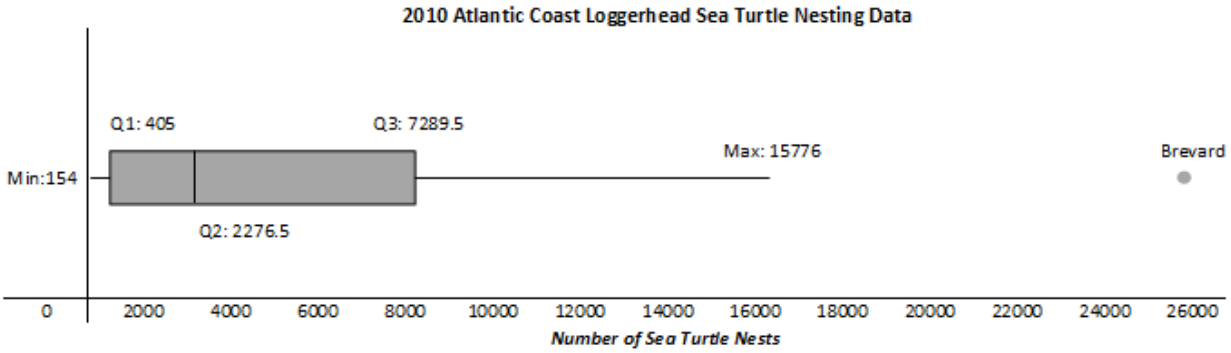
1. Determine whether the given sea turtle nesting data on the following page is discrete or continuous. *Be prepared to defend your answer!*
   The sea turtle data is discrete because it is an observed or counted data set.
2. Based on your decision from #1 above, create either a bar chart or histogram to represent the Atlantic Coast nesting data for the year your group was assigned.
   A bar chart should be created, as histograms are only appropriate for continuous data. Since the data is discrete, the bars on the graph should not touch, this emphasizes the fact that the data is countable and only whole number values.
   Horizontal Axis – Divides the number of seat turtle nests into subgroups or frequency groups, such as 0 to 500 nests. Below the frequency groups range by 2000, i.e. 0 to 2000 nests, 2001 to 4000 nests, 4001 to 6000 nests, etc…
   Vertical Axis – The scale for frequency or number of counties with the given number of sea turtle nests along the Atlantic coast. Below, it shows 5 counties with a sea turtle nest count from 0 to 2000, 2 counties with a count from 2001 to 4000, 2 counties with a count from 4001 to 6000, etc…
   Sample Bar Chart for 2010 Data:



3. Next, find the 5-number summary for your group's data. *Remember to identify any outliers!*

Year	*Minimum*	*Maximum*	*Q1*	*Q2 (Median)*	*Q3*	*Outliers*
2010	*154*	*25742*	*405*	*2276.5*	*7289.5*	*Brevard*
2011	*146*	*22893*	*382*	*2052*	*6619*	*Brevard*
2012	*187*	*33799*	*530.5*	*3084.5*	*8585*	*Brevard & Palm Beach*
2013	*184*	*24630*	*471*	*2367.5*	*7136.5*	*Brevard*
2014	*114*	*24951*	*423*	*2260.5*	*8122.5*	*Brevard & Palm Beach*

4. Create a box plot to represent your summary. Label any outliers with the county name.
Box plots may vary depending on the scale chosen.
Sample Box Plot for 2010 Data:



5. Display your group's graphs on a poster that can easily be seen by the entire class and attach all of the work on the back of the poster.
6. When your group is finished, hang the poster for the class to see.
7. Finally, after all posters are displayed, discuss the following questions within your group and come to a consensus for the answers you will share with the class.

Questions for Small Group Discussion

1. Describe any yearly trends or fluctuations in the nesting data that you observe for the Atlantic Coast. What do you think may have caused these trends/fluctuations?
Yearly trends seem relatively consistent with the same two counties as outliers. The range of the upper 25% fluctuates year to year. There are reasons for more dramatic trends/fluctuations when looking statewide that include: Inconsistent survey efforts and no standardized approach. Index beaches survey started to measure population trends. See http://myfwc.com/research/wildlife/sea-turtles/nesting/loggerhead-trends/

2. What counties are the outliers each year? Explain why, mathematically, these counties are considered to be outliers.

The outliers have been identified using the $Q1 - 1.5 \bullet IQR$ *or* $Q3 + 1.5 \bullet IQR$ *rule.*

2010: Brevard
2011: Brevard
2012: Brevard & Palm Beach
2013:Brevard
2014: Brevard & Palm Beach

3. Would you consider these outliers to be a part of the sample worth discarding or worth investigating further? Explain your reasoning.
   These counties are outliers because they have a high sea turtle nesting population. They should be studied further, not discarded. There may be external factors for these locations that can be replicated elsewhere to help sustain the sea turtle population.

4. What factors do you think exist that could make Brevard, St. Lucie, and Palm Beach counties have the highest sea turtle nesting densities in the entire state? Justify your reasoning.
   Sample answers may include: Length of beach surveyed, Population/development of beaches, Average temperature, Rainfall, Geographical location

   Lead discussion to determine factors that can be analyzed using bivariate analysis in order to connect to Module 2: Linear Regression.

Statewide Sea Turtle Nesting Data

		Number of Loggerhead Nests Statewide by County				
	County	**2010**	**2011**	**2012**	**2013**	**2014**
Atlantic Coast	Nassau	199	146	208	184	114
	Duval	154	152	187	186	119
	St. Johns	825	597	651	675	446
	Flagler	458	371	563	458	400
	Volusia	2270	1978	2885	2279	1643
	Brevard	25742	22893	33799	24630	23457
	Indian River	5147	4523	6729	5101	4482
	St. Lucie	5459	5763	5840	5775	5440
	Martin	9120	7475	10441	8498	10805
	Palm Beach	15776	15282	22192	16986	24951
	Broward	2283	2126	3284	2456	2878
	Miami-Dade	352	393	498	484	485
Gulf West Coast	Monroe	254	159	358	311	600
	Collier	778	757	1250	1091	1376
	Lee	750	961	1316	1315	1509
	Charlotte	527	713	1094	909	1323
	Sarasota	2517	2941	4695	4185	4884
	Manatee	274	280	634	690	539
	Hillsborough	29	54	61	79	47
	Pinellas	153	159	316	385	363
Gulf Panhandle	Franklin	307	387	628	665	415
	Gulf	187	251	561	292	328
	Bay	77	76	143	125	105
	Walton	36	44	118	67	60
	Okaloosa	9	31	55	56	34
	Santa Rosa	5	12	17	21	12
	Escambia	21	85	79	72	55
		73709	68609	98602	77975	86870