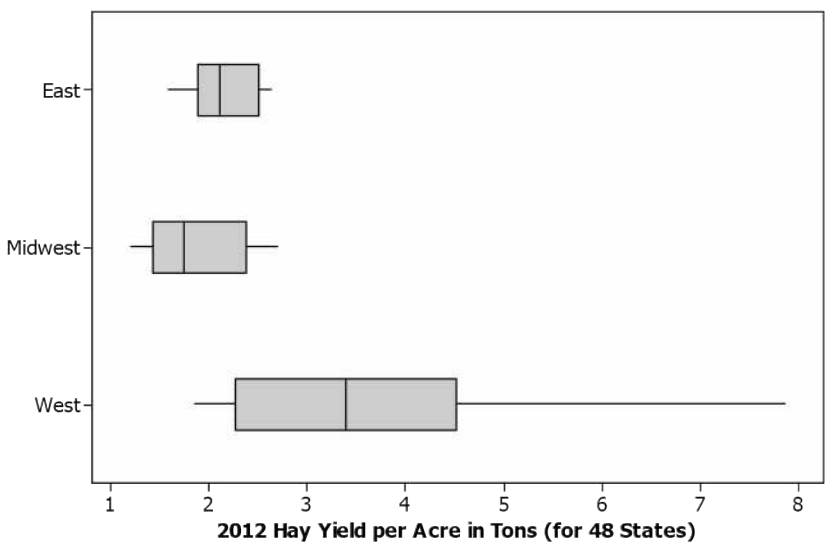
Name _____ Date_____

Comparing Data Distributions

According to the *United States Department of Agriculture National Agricultural Statistics Service Crop Production 2012 Summary*, in the contiguous 48 United States, there was a great deal of variability among states in terms of hay yield per acre. Do some regions of the United States generally have a higher hay yield per acre than other regions? The following box plots show the distribution of hay yield per acre (in tons) for 22 eastern states, 14 mid-western states, and 12 western states in 2012.

(From the *United States Department of Agriculture National Agricultural Statistics Service Crop Production 2012 Summary*, ISSN: 1057-7823, p. 75, accessed May 5, 2013 available at http://usda01.library.cornell.edu/usda/current/CropProdSu/CropProdSu-01-11-2013.pdf.)



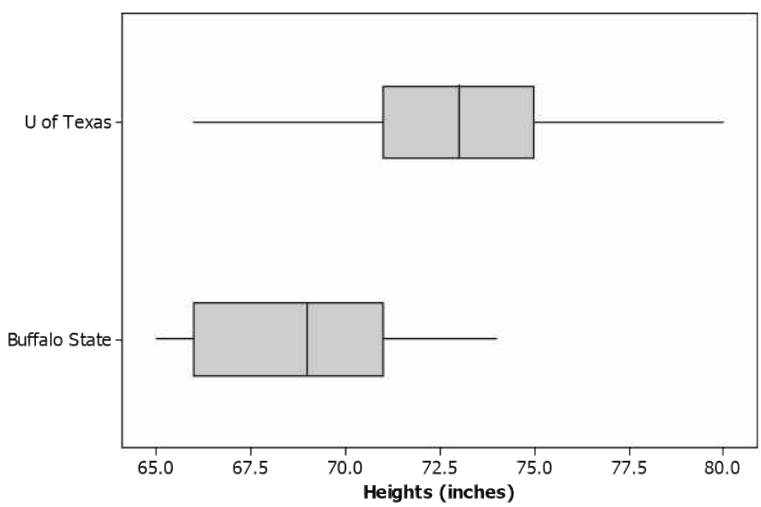
1. Which of the three regions' data sets has the least variability? Which has the greatest variability? To explain how you chose your answers, write a sentence or two that supports your choices by comparing relevant summary measures (i.e., median, IQR, etc.) or graphical attributes (i.e., shape, variability, etc.) from the three groups.

2. True or False: The Western state with the smallest hay yield per acre has a higher hay yield per acre than at least half of the Midwestern states. Explain how you know this is true or how this is false.

3. Which region typically has states with the largest hay yield per acre? To explain how you chose your answer, write a sentence or two that supports your choice by comparing relevant summary measures or graphical attributes from the three groups.

1. College athletic programs are separated into divisions based on school size, available athletic scholarships, and other factors. A researcher is curious to know if members of swimming and diving programs in Division I (schools that offer athletic scholarships and tend to have large enrollment) are generally taller than the swimmers and divers in Division III programs (schools that do not offer athletic scholarships and tend to have smaller enrollment). To begin the investigation, the researcher creates side-by-side box plots for the heights (in inches) of members of the 2012–2013 University of Texas Men's Swimming and Diving Team (a Division I program) and the heights (in inches) of members of the 2012–2013 Buffalo State College Men's Swimming and Diving Team (a Division III program).

   (From http://www.texassports.com/sports/m-swim/mtt/tex-m-swim-mtt.html accessed April 30, 2013, all 41 member heights listed, and http://www.buffalostateathletics.com/roster.aspx?path=mswim& accessed May 15, 2013, 11 members on roster; only 10 heights were listed)



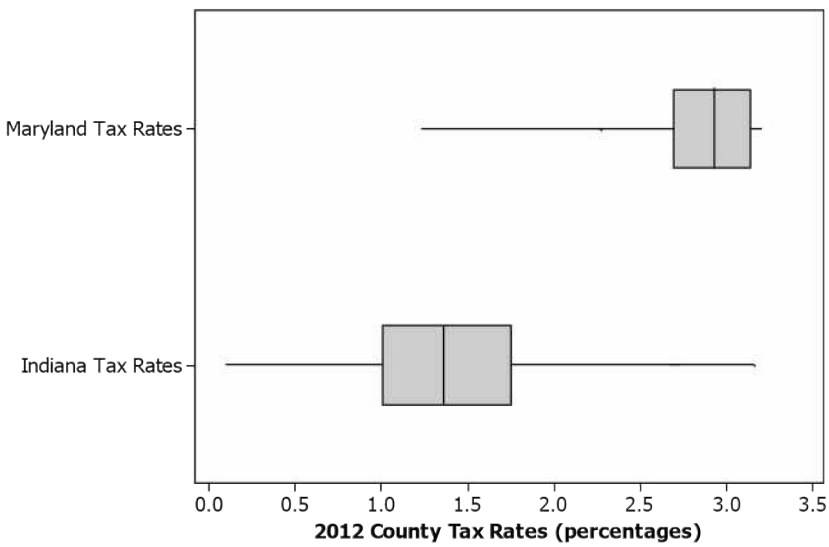
   a. Which data set has the smaller range?

   b. True or False: A team member of median height on the University of Texas team would be taller than a team member of median height on the Buffalo State College team.

   c. To be thorough, the researcher will examine many other college's sports programs to further investigate her claim that members of swimming and diving programs in Division I are generally taller than the swimmers and divers in Division III. But given the graph above, in this initial stage of her research, do you think that her claim might be valid? Carefully support your answer using comparative summary measures or graphical attributes.

2. Different states use different methods for determining a person's income tax. However, Maryland and Indiana both have systems where a person pays a different income tax rate based on the county in which he/she lives. Box plots summarizing the 24 different county tax rates for Maryland's 23 counties and Baltimore City (taxed like a county in this case) and the resident tax rates for 91 counties in Indiana in 2012 are shown below.

(From http://taxes.marylandtaxes.com/Individual_Taxes/Individual_Tax_Types/Income_Tax/Tax_Information/Tax_Rates/Local_and_County_Tax_Rates.shtml accessed May 5, 2013 and www.in.gov/dor/files/12-county-rates.pdf accessed May 16, 2013)



a. True or False: At least one Indiana county income tax rate is higher than the median county income tax rate in Maryland. Explain how you know.

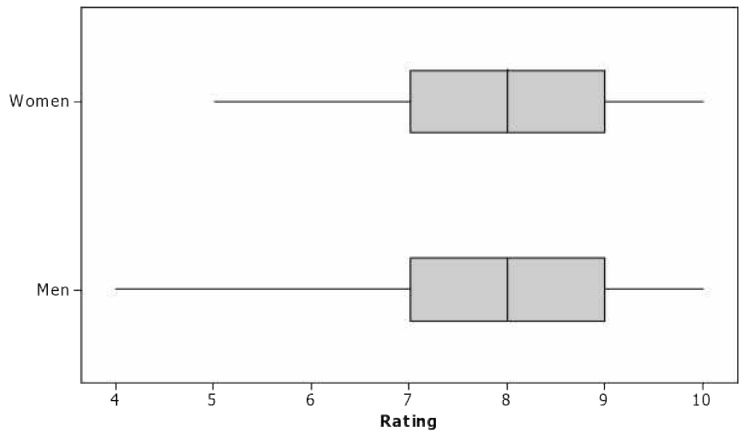
b. True or False: The 24 Maryland county income tax rates have less variability than the 91 Indiana county income tax rates. Explain how you know.

c. Which state appears to have typically lower county income tax rates? Explain.

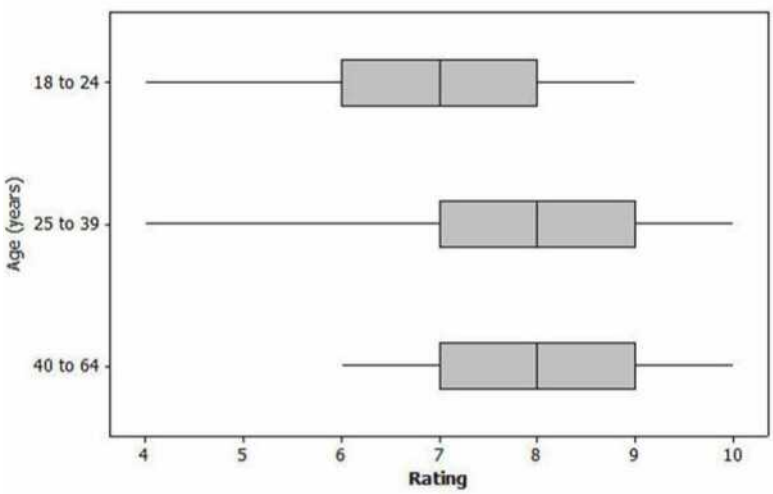
3. Many movie studios rely heavily on customer data in test markets to determine how a film will be marketed and distributed. Recently, previews of a soon to be released film were shown to 300 people. Each person was asked to rate the movie on a scale of 0 to 10, with 10 representing "best movie I've ever seen" and 0 representing "worst movie I've ever seen."

   Below are some side-by-side box plots that summarize the ratings based on certain demographic characteristics.

   For 150 women and 150 men:



   For 3 distinct age groups:

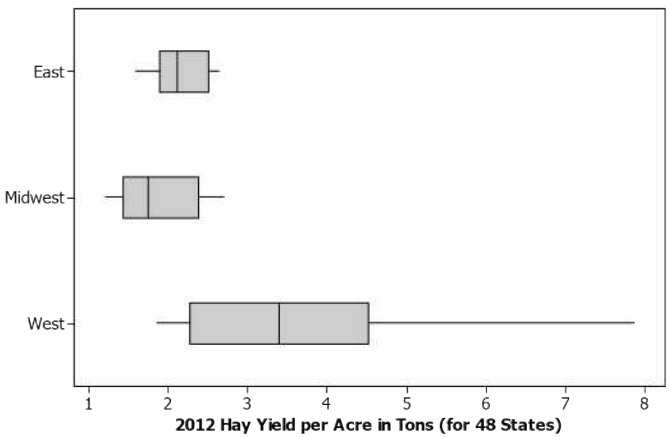


   a. Generally, does it appear that the men and women rated the film in a similar manner or in a very different manner? Write a few sentences explaining your answer using comparative information about center and spread from the graph.

   b. Generally, it appears that the film typically received better ratings from the older members of the group. Write a few sentences using comparative measures of center and spread or graphical attributes to justify this claim.

According to the *United States Department of Agriculture National Agricultural Statistics Service Crop Production 2012 Summary*, in the contiguous 48 United States, there was a great deal of variability among states in terms of hay yield per acre. Do some regions of the United States generally have a higher hay yield per acre than other regions? The following box plots show the distribution of hay yield per acre (in tons) for 22 eastern states, 14 mid-western states, and 12 western states in 2012.

(From the *United States Department of Agriculture National Agricultural Statistics Service Crop Production 2012 Summary*, ISSN: 1057-7823, p. 75, accessed May 5, 2013 available at http://usda01.library.cornell.edu/usda/current/CropProdSu/CropProdSu-01-11-2013.pdf.)



1. Which of the three regions' data sets has the least variability? Which has the greatest variability? To explain how you chose your answers, write a sentence or two that supports your choices by comparing relevant summary measures (i.e., median, IQR, etc.) or graphical attributes (i.e., shape, variability, etc.) from the three groups.

   The East data set has the least variability as it has the smallest range and the smallest IQR. The West data set has the greatest variability as it has the largest range and the largest IQR.

2. True or False: The Western state with the smallest hay yield per acre has a higher hay yield per acre than at least half of the Midwestern states. Explain how you know this is true or how this is false.

   This is true; the minimum value of the West data set is higher than the median value of the Midwest data set. Therefore, this minimum value for the West must be higher than at least half of the Midwestern states' values.

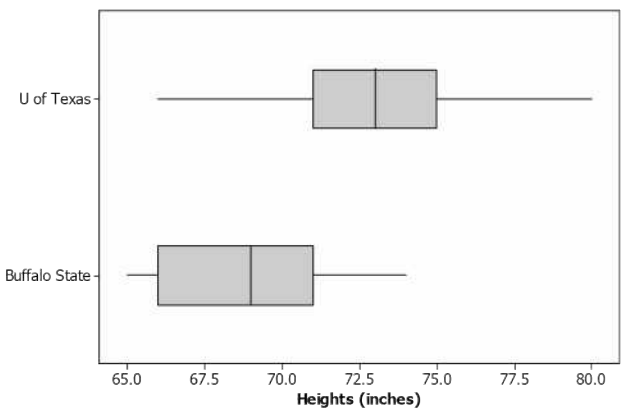
3. Which region typically has states with the largest hay yield per acre? To explain how you chose your answer, write a sentence or two that supports your choice by comparing relevant summary measures or graphical attributes from the three groups.

   The West typically has states with the largest hay yield per acre. Over half of the Western states have hay yields that are higher than any yield in either of the other two regions. Also, some Western yields are up to two or three times the largest Eastern and Midwestern yields.

Before students begin the problem set, consider providing them time to work on their projects. If students have not collected data, then provide assistance in completing that process. If students have collected data, then provide them time to develop numerical or graphical summaries of the data (dot plots, box plots, or histograms). Assign only one or two of the problems in the problem set if completion of the project needs to be addressed.

1. **College athletic programs are separated into divisions based on school size, available athletic scholarships, and other factors. A researcher is curious to know if members of swimming and diving programs in Division I (schools that offer athletic scholarships and tend to have large enrollment) are generally taller than the swimmers and divers in Division III programs (schools that do not offer athletic scholarships and tend to have smaller enrollment). To begin the investigation, the researcher creates side-by-side box plots for the heights (in inches) of members of the 2012–2013 University of Texas Men's Swimming and Diving Team (a Division I program) and the heights (in inches) of members of the 2012–2013 Buffalo State College Men's Swimming and Diving Team (a Division III program).**

   (From http://www.texassports.com/sports/m-swim/mtt/tex-m-swim-mtt.html accessed April 30, 2013, all 41 member heights listed, and http://www.buffalostateathletics.com/roster.aspx?path=mswim& accessed May 15, 2013, 11 members on roster; only 10 heights were listed)



   a. **Which data set has the smaller range?**

      Buffalo State

   b. **True or False: A team member of median height on the University of Texas team would be taller than a team member of median height on the Buffalo State College team.**

      True.

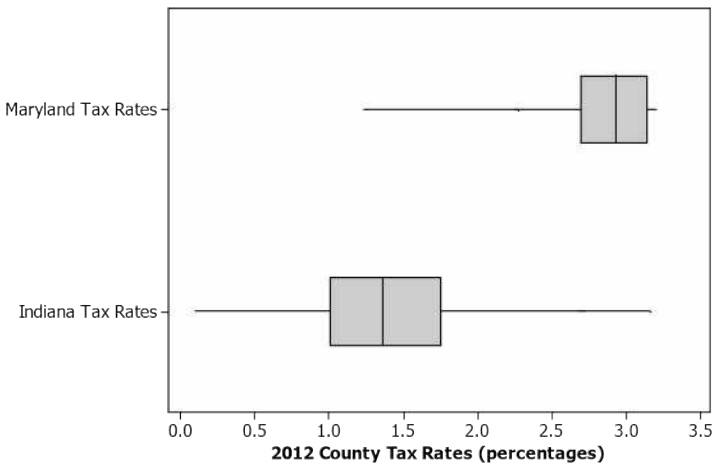
   c. **To be thorough, the researcher will examine many other college's sports programs to further investigate her claim that members of swimming and diving programs in Division I are generally taller than the swimmers and divers in Division III. But given the graph above, in this initial stage of her research, do you think that her claim might be valid? Carefully support your answer using comparative summary measures or graphical attributes.**

      Yes, a large portion of the University of Texas distribution is higher than the maximum value of the Buffalo State distribution. The median value for the University of Texas appears to be 4 inches higher than the median value of the Buffalo State distribution.

2. Different states use different methods for determining a person's income tax. However, Maryland and Indiana both have systems where a person pays a different income tax rate based on the county in which he/she lives. Box plots summarizing the 24 different county tax rates for Maryland's 23 counties and Baltimore City (taxed like a county in this case) and the resident tax rates for 91 counties in Indiana in 2012 are shown below.

(From http://taxes.marylandtaxes.com/Individual_Taxes/Individual_Tax_Types/Income_Tax/Tax_Information/Tax_Rates/Local_and_County_Tax_Rates.shtml accessed May 5, 2013 and www.in.gov/dor/files/12-county-rates.pdf accessed May 16, 2013)

a. True or False: At least one Indiana county income tax rate is higher than the median county income tax rate



in Maryland. Explain how you know.

True. The median tax rate for Maryland appears to be a little under 3%, and the maximum tax rate in Indiana is over 3%.

b. True or False: The 24 Maryland county income tax rates have less variability than the 91 Indiana county income tax rates. Explain how you know.

True. The tax rates in Maryland are more compact than for Indiana. Maryland has a smaller range and IQR compared to Indiana.

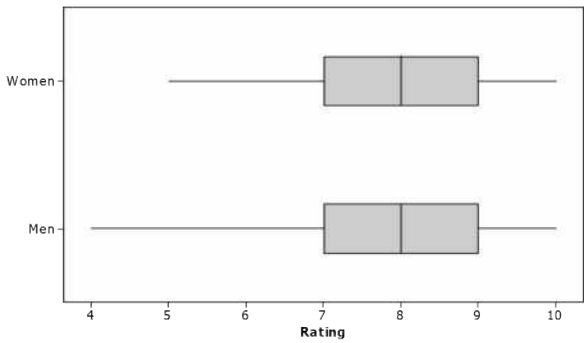
c. Which state appears to have typically lower county income tax rates? Explain.

Indiana counties typically appear to have lower county income tax rates. The median Indiana tax rate is much lower than the median Maryland tax rate, and a large part of the Indiana distribution is lower than the minimum value of the Maryland distribution.

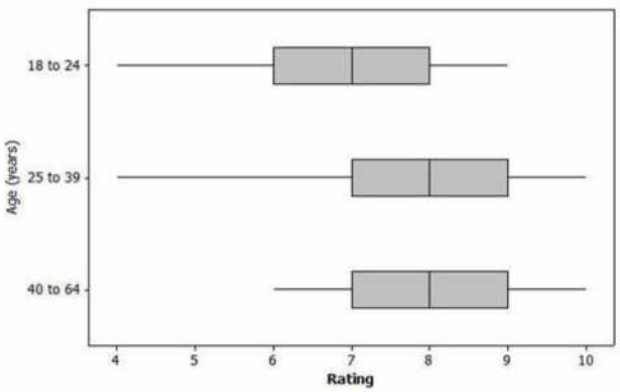
3. Many movie studios rely heavily on customer data in test markets to determine how a film will be marketed and distributed. Recently, previews of a soon to be released film were shown to 300 people. Each person was asked to rate the movie on a scale of 0 to 10, with 10 representing "best movie I've ever seen" and 0 representing "worst movie I've ever seen."

Below are some side-by-side box plots that summarize the ratings based on certain demographic characteristics.

For 150 women and 150 men:



For 3 distinct age groups:



Note: Students may be more likely to provide comparative values for this question given the discrete, integer nature of the data.

a. Generally, does it appear that the men and women rated the film in a similar manner or in a very different manner? Write a few sentences explaining your answer using comparative information about center and spread from the graph.

It appears that the men and women rated the film in a very similar manner: same quartile values, same medians, and same maximums. The only difference is that the minimum rating from men was slightly lower than the minimum rating from women.

b. Generally, it appears that the film typically received better ratings from the older members of the group. Write a few sentences using comparative measures of center and spread or graphical attributes to justify this claim.

For the two oldest age groups, the Q1, median, Q3, and maximum values are all higher than the $18-24$ counterparts. In fact the Q1 value for each of these two older groups equals the median rating of the youngest group, and the median value for each of these two older groups equals the Q3 rating of the youngest group. Additionally, while the two oldest groups have similar distributions, the minimum score of the oldest group was much higher than the minimum value of the $25-39$ group. This means that none of the $40-64$ respondents rated the movie with a score as low as a 4 (as was the case in the $25-39$ age group).