

Name (Last, First):

Student ID:

Assignment 3

METCS544A3A4_F2024

Instructions:

1. **This assignment has no specific R programming questions, but you're encouraged to use R to plot and graph the data and calculate relevant statistic summaries.**
2. For answering programming questions, please use Adobe Acrobat to edit the pdf file in two steps **[See Appendix: Example Question and Answer]**:
 - a. Copy and paste your R code as text in the box provided (so that your teaching team can run your code);
 - b. Screenshot your R console outputs, save them as a .PNG image file, and paste/insert them in the box provided.
 - c. Show all work - credit will not be given for code without showing the code in action by including the screenshot of R console outputs.
3. To answer non-programming questions, please type or handwrite your final answers clearly in the boxes. Show all work - credit will not be given for numerical solutions that appear without explanation in the space above the boxes.
4. **[Total 93 pts = 21 + 48 + 24 Extra Credit pts]**

Grading Rubric

Each question is worth 3 points and will be graded as follows:

3 points: Correct answer with work shown

2 points: Incorrect answer but attempt shows some understanding (work shown)

1 point: Incorrect answer but an attempt was made (work shown), or **correct answer without explanation (work not shown)**

0 points: Left blank or made little to no effort/work not shown

Reflective Journal [3 pts]


(Copy and paste the link to your live Google doc in the box below)

Part I: Quantitative Data [21 pts]

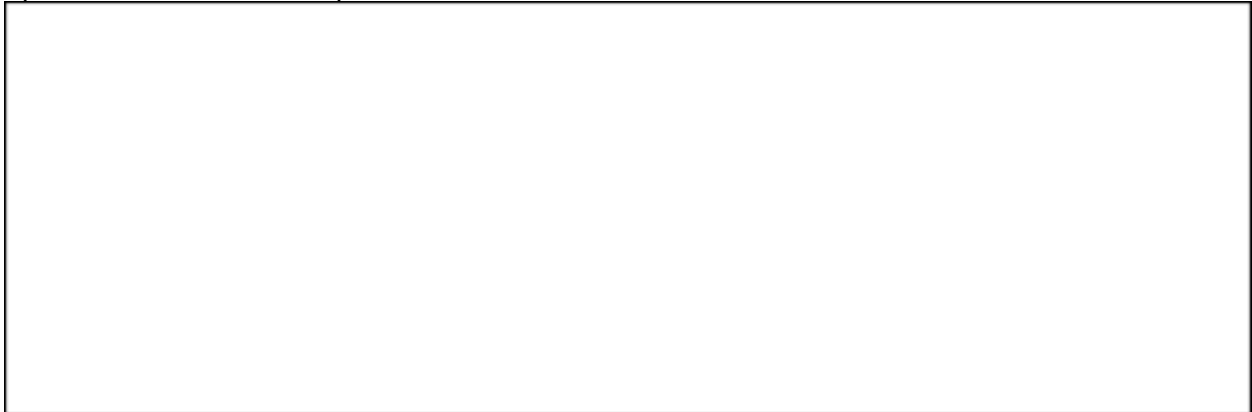
1) The following are the temperature highs for an Illinois town in August 2022.

78	75	83	85	87	91	78	91	85	84	80	79	83
85	86	80	88	94	89	86	88	84	81	85	83	80
79	87	88	91	95								

a) Create a histogram of this data (include a table of bins and counts).



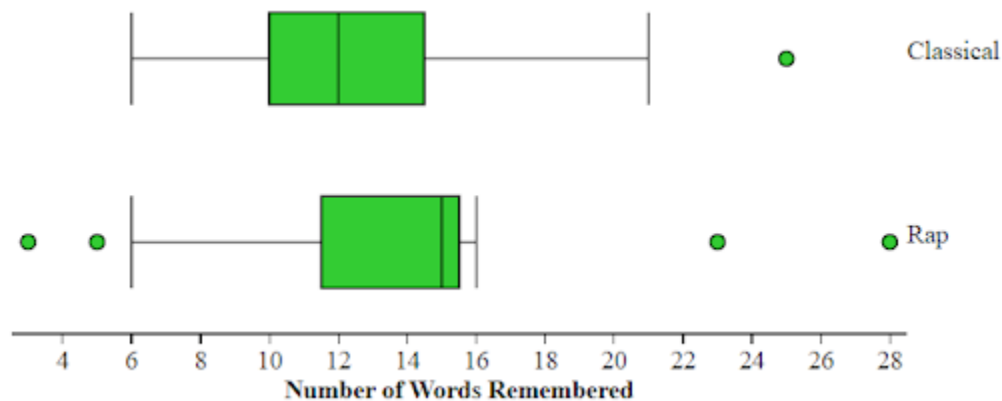
b) Create a stem and leaf plot for this data.



c) Which graph do you think displays the data the best? Why?



2) Amy and Bob want to know if listening to different types of music while studying will help you remember the material better. They randomly assigned a group of students to two different groups: one group studied a list of words while listening to classical music while the second group studied the same list of words while listening to rap music. There was a total of 30 words on the list and each group studied the list while listening to the music for 5 minutes. They were then asked to write down as many words as they could remember. The data is displayed in the boxplots below.



a) Approximate the interquartile range for each set of data. Why is this the appropriate measure of spread to use for these two data sets? (3 pts)

Answer:

b) Write **three** sentences comparing the number of words remembered between each of the two groups. (9 pts)

Answer:

Part II. Summary Statistics and Boxplots [48 pts = 16 x 3 pts]

1) Below is a list of calories and cholesterol amounts in 4 randomly selected menu items from 4 different fast food companies. [The data is accessible via a .csv file saved under the course shared folder "Course Contents" as "calories_and_cholesterol_amounts.csv"]

<i>Company</i>	<i>Menu Item</i>	<i>Calories</i>	<i>Cholesterol (mg)</i>
McDonald's	Bacon, Egg, & Cheese Biscuit	460	215
McDonald's	Big Mac®	590	85
McDonald's	Filet-O-Fish®	390	30
McDonald's	Cheeseburger	300	40
Burger King	Whopper® Sandwich with Cheese	740	115
Burger King	Cheeseburger	280	45
Burger King	Crispy Chicken Sandwich	670	60
Burger King	Bacon, Egg & Cheese Biscuit	400	170
Wendy's	Baconator®	960	155
Wendy's	Bacon Double Stack®	440	65
Wendy's	Classic Chicken Sandwich	490	75
Wendy's	Sausage, Egg, and Cheese Biscuit	580	285
Chick-fil-A	Chicken Biscuit	460	45
Chick-fil-A	Bacon, Egg, and Cheese Biscuit	420	180
Chick-fil-A	Grilled Chicken Sandwich	390	75
Chick-fil-A	Chicken Nuggets (8 count)	250	80

a) Using R, find the following summary statistics for CALORIES for each company. (6 pts = 2 x 3 pts)

Answer:

	Mean	Min	Q1	Med	Q3	Max	Std. Dev
McDonald's							
Burger King							
Wendy's							
Chick-fil-A							

Which company has the greatest calorie variability in the distribution?

Answer:

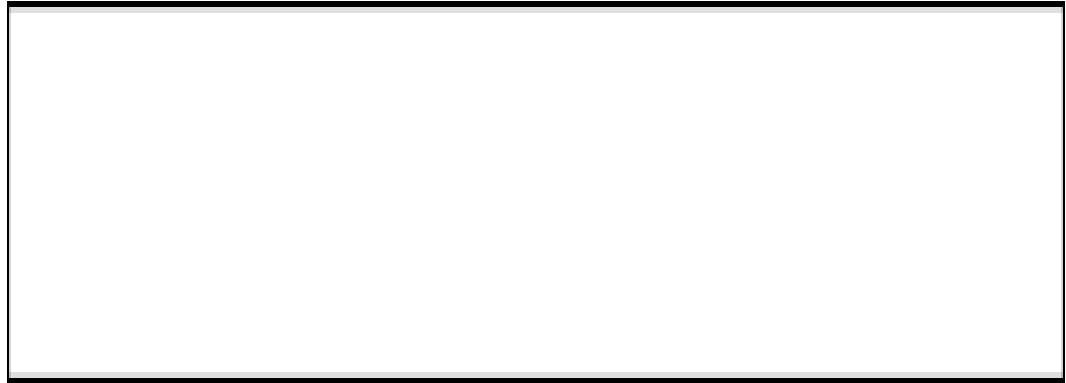
b) There are many methods for determining outliers. Two methods frequently used are:

Rule #1: An outlier is a value greater than $1.5 \times \text{IQR}$ above the third quartile or more than $1.5 \times \text{IQR}$ below the first quartile.

Rule #2: An outlier is a value located 2 or more standard deviations above, or below, the mean.

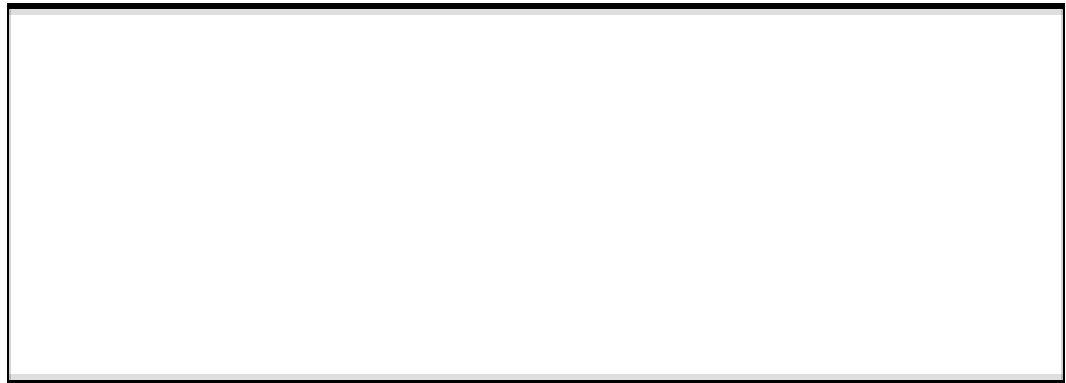
Using rule #1, are there any outliers in the Wendy's distribution? Show your work.

Answer:



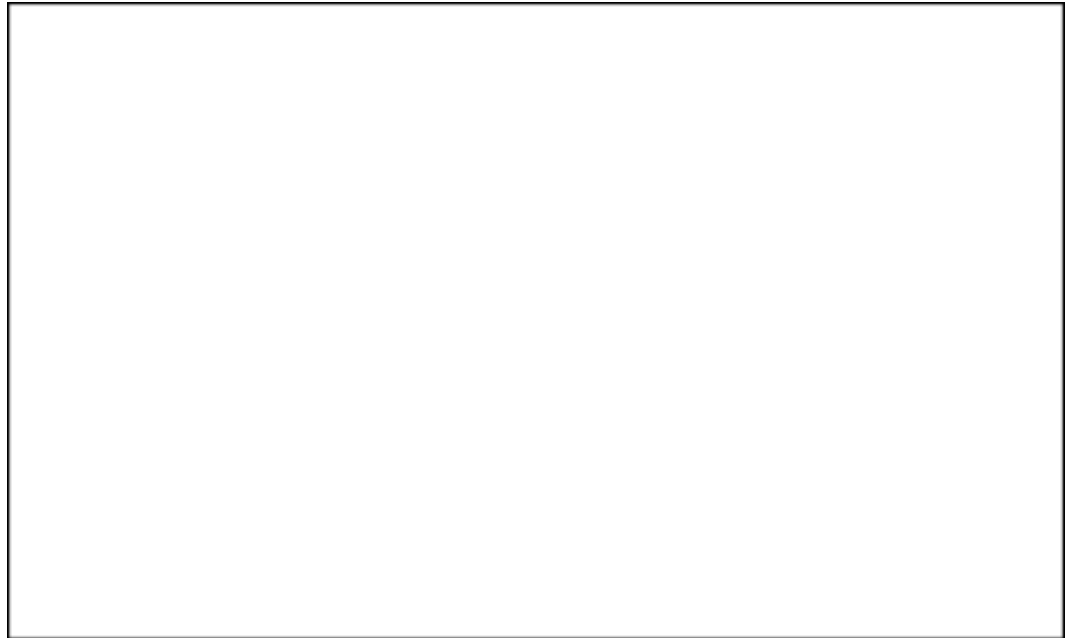
Using rule #2, are there any outliers in the Wendy's distribution? Show your work.

Answer:



c) Draw four modified boxplots comparing the calories for each of the four companies.

Answer:

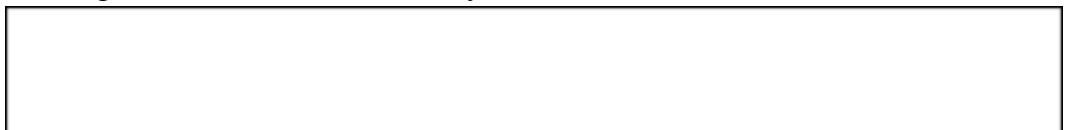


d) In your town, McDonald's and Burger King are on the Northside and Wendy's and Chick-fil-A are on the Southside. Using R, find the following summary statistics for CHOLESTEROL for the Northside and Southside fast food restaurants. **(6 pts = 2 x 3 pts)**

	Mean	Min	Q1	Med	Q3	Max	Std. Dev
Northside							
Southside							

Which region has the greatest cholesterol variability in the distribution?

Answer:



e) Using rule #1, are there any outliers in the Southside's distribution? Show your work.

Answer:

Using rule #2, are there any outliers in the Southside's distribution? Show your work.

Answer:

f) Remove the value of 285mg from the Southside cholesterol's data set. Use R to find the following values again:

	Mean	Min	Q1	Med	Q3	Max	Std. Dev
Southside							

What values changed the most? What values changed the least?

Answer

(Most):

Answer

(Least):

Part III Extra Credit Questions (21 pts)

Extra Credit Question (20pts)

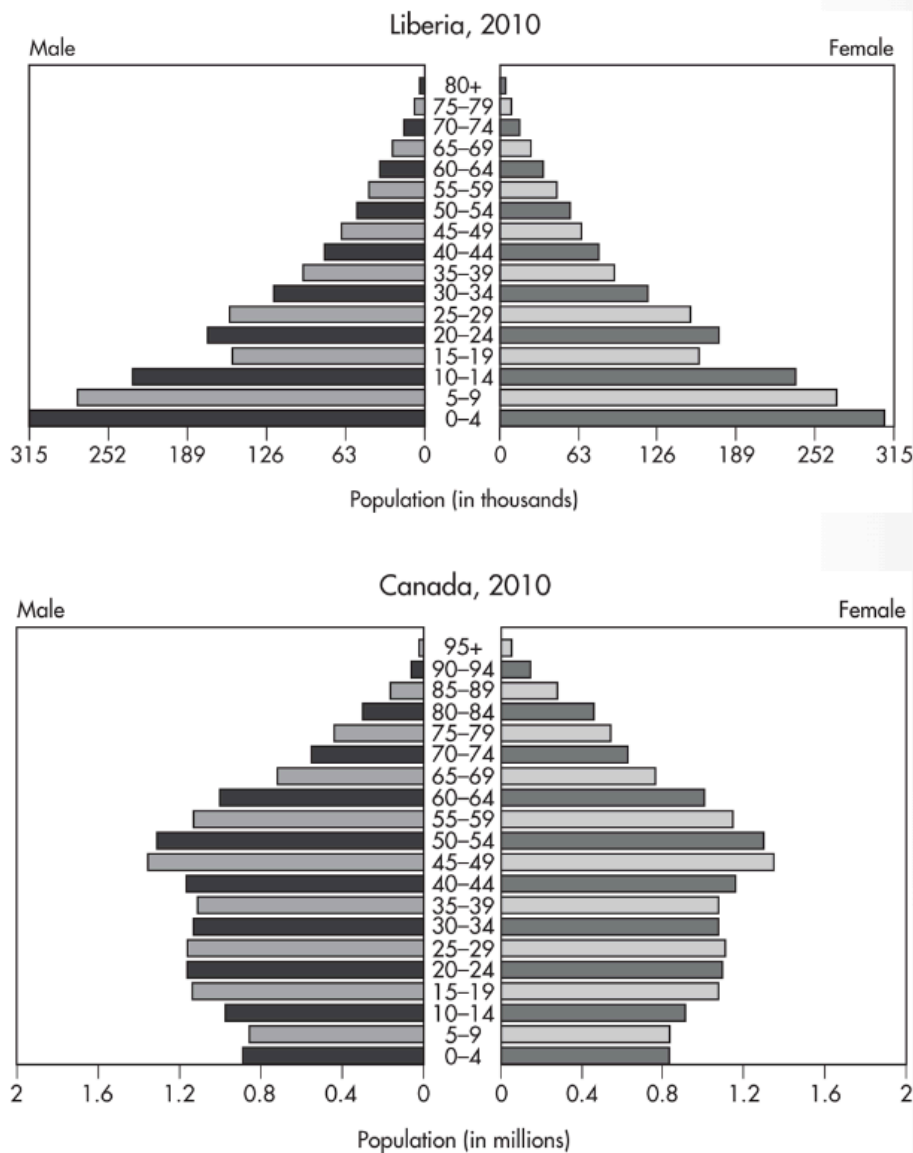
- (1) (6 pts) To analyze the social media behavior differences between boys and girls, Mr. P's Statistics class was asked to count the number of text messages that they sent over a three-day weekend. The following table summarizes the data:

	Values under Q_1	Q_1	Median	Q_3	Values over Q_3
Females	15, 43, 100	130	175	358	450, 573, 1098
Males	3, 59	72	183	273	293, 337

- Construct parallel boxplots of this set of data.
- Do the data indicate that females or males had the greater mean number of texts? Explain in detail (Shape, Outliers, Center, Spread; Conclusion).

Answer:

(2) (15 pts) Below are two population pyramids from the U.S. Census Bureau.



- The approximate median age of the Liberian population falls in which of these intervals: 0–4, 15–19, 30–34, 40–44? Explain.
- Explain why it is impossible to calculate the mean age of either population.
- Which country has more children younger than 10 years of age? Explain.
- Does the population pyramid indicate that Canadian men or Canadian women live longer? Explain.
- In 2010, Liberia had recently come out of a civil war with the extensive use of child soldiers. How is this visible in the population pyramid?

Answer:

A large, empty rectangular box with a thin black border, intended for the user to write their answer.

THE END