Homework 2

<div style="border:1px solid #000; display:inline-block; padding:10px; background:#1a3a6b; color:#fff;">Start Assignment</div>

Due Jan 25 by 11:59pm   **Points** 10   **Submitting** a file upload
File Types doc, docx, txt, and pdf   **Available** until Feb 1 at 11:59pm

Assignment Overview

The purpose of this assignment is to help you practice skills that are essential to your success in this course and within exploratory data analysis and visualization. For this assignment:

- You will need to download and use the bodyfat.csv file. This data set is modified from the Journal of Statistics Education and contains data on body fat measurements for a sample of 252 men. In the file, you will see the variables:

  - id: ID number
  - bodyfat: percentage body fat using Brozek's equation
  - density: (gm/cm^3)
  - age: years
  - weight: pounds
  - height: inches
  - adiposity: weight/height^2
  - neck, chest, abdomen, hip, though, knee, ankle, bicep, forearm, wrist: circumference in cm

- The American Council on Exercise uses the following body fat categorizations for men:

  - Essential fat: 2–5%
  - Athletes: 6–13%
  - Fitness: 14–17%
  - Average: 18–25%
  - Obese: 25+%

- To simplify our purposes, we will use the following categorizations:

  - Athlete: less than 14%
  - Average: 14–25%

○ Obese: greater than 25%

In a Word or PDF document, state the problem number (below) that you are answering. Provide the R code that you used **before** providing a graph/picture (if any) and answering the question. Questions 1 and 2 should just be R code. Keep all outliers (extreme cases).

1. Read in the bodyfat.csv data file and generate a variable "bodycat" to categorize body fat into the three categories above. Make sure all 252 observations are categorized into either athlete, average, or obese.
2. Using summarize to identify the four height quartiles, create a new variable "htcat" to categorize height into "short", "below average", "above average", and "tall".
3. Create a violin plot of weight separated by bodycat. Make sure your plots show up in some kind of order that makes sense. In complete sentences, summarize what the violin plots tell you. Are the weights evenly distributed within a range for all categories? Do athletes tend to be within a certain weight range? You may use summarize() to help you. Rough estimates are also okay.
4. Create a stem-and-leaf plot for weight. Be sure to find an appropriate scale for the data.
5.
    A. Create overlapping histograms of neck for the three body categories. For this exercise, do not use the default breaks. Use breaks that you think make sense. Remember to make sure that the first histogram is an appropriate window size so when you "add" the other graphs, those histograms aren't cut-off. Also, remember to use the same break widths for overlapped histograms.
    B. In the same window, add 3 density plots—1 for each body category. Do not use the default bandwidth. Use a bandwidth that you think makes sense. (Note, you will need to have used freq = F in your histograms.)
    C. In complete sentences, compare neck circumference across the three body categories using your histograms and density plots.
6. Repeat Question 5 for abdomen.

Optional: In preparation for your upcoming presentations, find a few quantitative data-driven articles or conference presentations. Skim through these for data visualization ideas on how you might want to approach the individual project (you'll have to find your own dataset). Here's **one covid-19 related article** ⇗ **(https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(20)30074-7/fulltext)** that uses boxplots and density plots.

Guidelines

1. Click **Start Assignment** to the upper right.
2. Upload your file (must be a .doc, .docx, .txt, or .pdf file type).
3. Click **Submit Assignment** to submit your work.

The Canvas Student Guide provides **instructions for submitting an online Assignment.** ↪
(https://community.canvaslms.com/videos/1121-assignments-submissions-students)

Grading Criteria

Use the rubric below as a checklist to ensure that you have successfully completed the assignment.

Homework 2

Criteria	Ratings		Pts
Categorization	**1 pts** **Full Marks** Correctly categorizes body fat as described in assignment (athlete, average, obese)	**0 pts** **No Marks**	1 pts
Height Quartiles	**1 pts** **Full Marks** Creates new variable htcat based on four height quartiles	**0 pts** **No Marks**	1 pts
Violin Plot	**2 pts** **Full Marks** Violin plot created in order of weight category, with detailed description of the graph	**0 pts** **No Marks**	2 pts
Stem-and-Leaf Plot	**1 pts** **Full Marks** Correctly creates a stem-and-leaf plot	**0 pts** **No Marks**	1 pts
Overlapping Histograms and Density Plots	**3 pts** **Full Marks** All three histograms overlapping and visible All histograms not cut off All three density plots overlapping histograms. Overlapping histograms generated for both neck and abdomen variables separately.	**0 pts** **No Marks**	3 pts
Overlapping Histograms and Density Plots Description	**2 pts** **Full Marks** Correct and detailed interpretation of the distribution of the neck variable. Correct and detailed interpretation of the distribution of the abdomen variable.	**0 pts** **No Marks**	2 pts

Total Points: 10