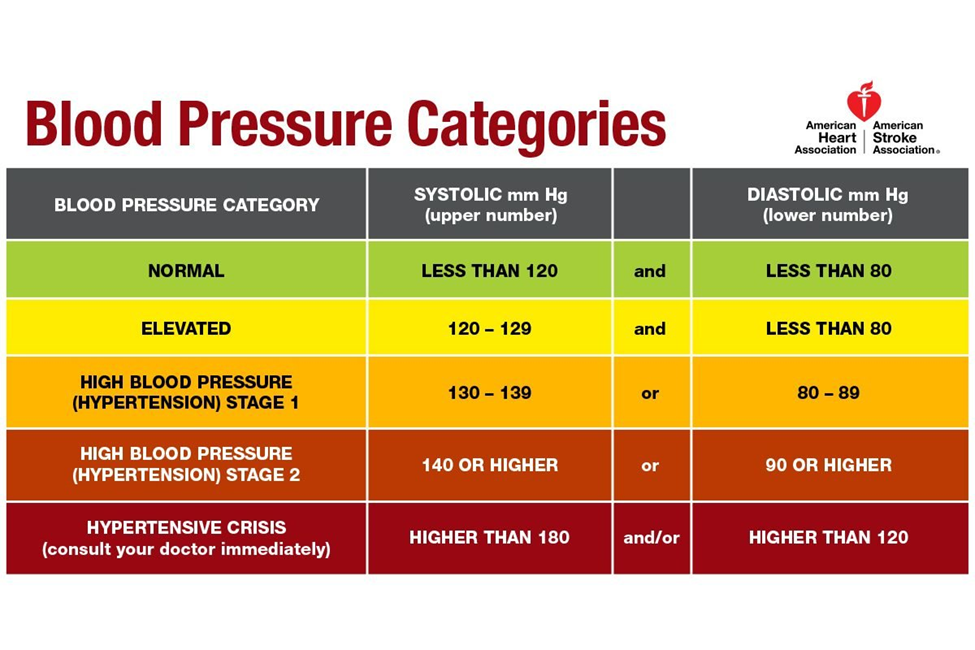
Project 1

Due Friday December 11, 2020 (**No extensions**)

* For this project you will use the SAS data file **analysis\_file,** you created for homework 4. You can download the file from Canvas.
* Include a word or pdf file with your answers
* Include your SAS program in a separate file and **called it project1.sas**

1. Read the SAS dataset **analysis\_file** for this project from CANVAS Project 1 Module. Describe the dataset: Use the procedures we have used in class to describe the dataset by doing the following:
   1. Use the contents procedure to see and report the name, variable type, and variable label, if available of the variables in the dataset.
   2. Report the
      1. number of people in the study,
      2. number of people by survival status,
      3. number of HF patients (Hint: create a binary variable called **HF\_status** that identifies HF patients by using **age\_at\_dx (Age at HF diagnosis)**: missing if patient does not have a HF diagnosis).
   3. Describe the **following 6 covariates in the study**: from sex, age, dbp, sbp, smoking and cholesterol in the order they appear in the dataset. Use the adequate way to describe the variables depending on whether they are discrete (binary, nominal) or continuous. Note: smoking is the daily number of cigarettes.
   4. Check if there are missing values in the variables and report how many per variable, if any.
2. Create The following variables: (**Variables should be numeric and should contain formats if the new variable is categorical**). Pay special attention to the possibility of missing values when creating these new variables and make sure they are not in a group where they are not supposed to be.
   1. BMI = , where kg is weight in kilograms and m is height in meters. (**NOTE: in this dataset weight is in pounds and height in inches so you have to convert pounds to kg and inches to meters first**).
   2. Define Blood\_pressure\_status as follows: (**NOTE: there are 2 variables systolic and diastolic that define this variable. Pay special attention to the “and” and “or”**).



**For Questions 3, 4 and 5 use plots and/or tables as you see fit and explain what you see. (Hint: which group is more likely to die/have HF)**

1. Explore the relationships between survival status and HF status. Do people with and without HF die at the same rate?
2. Explore the relationships between survival status and:
   1. Sex
   2. Age
   3. Blood Pressure Status (blood\_pressure\_status)
3. **Among those who died (HINT: Restrict to only those who died from the original dataset)**, show a table of the cause of death.
   1. Explore the relationship between cause of death and:
      1. Sex
      2. Blood\_pressure\_status
   2. Show the relationship between BMI and cholesterol levels (both continuous variables), by cause of death.