Homework 4

Due: Wednesday, July 6th (before 11:59 pm) on GauchoSpace.

Instructions: For the General Knowledge Questions, answer as succinctly as possible. For the Programming Assignment, show your code as well as a partial screenshot of your output.

General Knowledge Questions

1. Which option, to the REPORT procedure, instructs SAS to place a line underneath the column headers?
2. Which statement, in the DATA step, instructs SAS to keep the value(s) of calculated variable(s) for the next iteration of the step?
3. In the REPORT procedure, how are character values justified in the default report?
4. In which step can FIRST.by-variable and LAST.by-variables be created?
5. Which option, to the SYMBOL statement, specifies the shape of the plotting symbol?
6. What is the purpose of the STYLE= option to the ODS statement?
7. Name two procedures that can be used to create list reports.
8. Can LABEL and FORMAT statements be used in the GPLOT procedure?
9. The DROP= and KEEP= options for data sets in the SET statement are used to control the variables that are _____ the DATA step.
10. To create a summary report, the REPORT procedure must only contain _____ or _____ variables.
11. In the GCHART procedure, what type(s) of variable can be specified as a chart-variable?
12. In the DATA step, which statement is used to instruct SAS to immediately write a record to the output data set(s)?
13. By default, does the REPORT procedure use labels as column headers?
14. In the statement: `plot hours*classes;` is `hours` the vertical or horizontal variable?
15. Which procedure can be used to produce scatterplots?

Programming Assignment

1. The **`beach2019`** data set posted on GauchoSpace contains Adopt-A-Beach cleanup events on South Carolina's coast in 2019. Each observation is a cleanup event that was logged by volunteer participants.
   ~~a) Create a report that displays the data portion of the data set.~~
   ~~b) Create a vertical bar chart for the variable `beach`.~~
   c) Create a pie chart for the total number of items picked up at the each beach. Additionally,
      i. Format the total number of items with commas.
      ii. ~~Emphasize the slice for 'Folly Beach'.~~
      iii. ~~Use a solid pattern~~.
   d) ~~Create a scatterplot of **items** by **hours**. Only select observations in which the cleanup was at most 25 hours~~. Additionally,
      ~~i. Define the scale of the vertical axis (tick marks for every 100).~~

ii. ~~Label the axes as 'Items Cleaned Up' and 'Hours of Cleanup'.~~
iii. ~~Use green dot symbols.~~
iv. ~~Add a regression line.~~
v. ~~Add a regression equation.~~

2. Use the **`military2`** data set posted on GauchoSpace. This data set contains information for four branches of the military.
   a) ~~Using one DATA step, separate the **`military2`** data set into four individual data sets, each of which only contains observations for one branch of the military. Name the data sets as follows;~~ each data set should only contain the variables specified in parentheses:
      i. ~~**`Airforce`** (type code city state country airport awards)~~
      ii. ~~**`Army`** (code city state)~~
      iii. ~~**`Marines`** (city state country airport awards)~~
      iv. ~~**`Naval`** (city airport)~~
   b) ~~How many observations are in each of the data sets created in part A? How many observations are in the **`military2`** data set?~~
   c) ~~Use the **`military2`** data set to create new data sets (**`m1, m2`**) with a running total (**AwardsRT**) of the number of awards received. (Hint: use the variable `Awards`)~~
      i. ~~In the **`m1`** data set, use the `sum()` function; print out the data set~~
      ii. ~~In the **`m2`** data set, use the sum statement; print out the data set~~
   d) ~~Use the **`military2`** data set to create a new data set (**`m3`**) with **one** accumulated total (**TotalAwards**) for each group of the variable **Type**. Only keep the variables **Type** and **TotalAwards**. Print out the data set.~~

3. Use the **`cars`** data set posted on GauchoSpace.
   a) Create a *listing report* called *cars1.lst*. Use the REPORT procedure to create a list report displaying the **Type**, **EngineSize**, and **MSRP**. Additionally,
      i. ~~Display a title of 'Vehicle Prices: List Report'~~
      ii. ~~Start the page numbers at 5~~
      iii. ~~Set the column width for **Type** to be 12 and use a column header of 'Vehicle Type'~~
      iv. For the variable **Horsepower**: display 'Less than 150 hp' for values less than 150, 'Between 150 and 300 hp' for values between 150 and 300, and 'More than 300 hp' for values greater than 300
      v. ~~Set the column width for **Horsepower** to 25~~
      vi. ~~Format the **MSRP** variable with dollar signs, commas, and no decimal places~~
      vii. ~~Display 'Manufacturer's Suggested Retail Price' as the column header for **MSRP** and set the column width to 40~~

   b) Create a *listing report* called *cars2.lst.* Modify the code in part A to create a summary report that combines observations on **Type** and **Horsepower** values and displays the sum of **MSRP**. Additionally,
      i. ~~Display a title of 'Vehicle Prices: Summary Report'~~

ii. ~~Add a line below the column headers~~
iii. ~~Place a *blank* line below the column headers~~
iv. ~~Format the total MSRP with dollar signs, commas, and no decimal places~~
v. ~~Add a subtotal MSRP after each `Type` value – place a single line above and a double line below~~
vi. Add a grand total MSRP at the end of the report – place a double line below