Homework Manipulating Data

1. The SAS data set staff is in the STA5066 sub directory. Use this data set to create a SAS data set, work.increase
2. Create two new variables, Increase and NewSalary. Increase is the Salary increased by 10%. NewSalary is increase added to salary.
3. The data set work.increase should include only the variables: Employee\_ID, Salary, Increase, and NewSalary.
4. Formats for displaying commas should be stored for Salary, Increase, and NewSalary.
5. Include a PROC PRINT step in your program to display the data portion of the SAS data set work.increase.
6. The SAS data set customer is in the STA5066 subdirectory. Use this data set to create a SAS dataset work.birthday
7. Create three new variables
8. Bday2017 -- a SAS date in 2017. (Use the MDY function to extract month and day of birth from the variable Birth\_Date, and 2017 as a constant for year)
9. BdayDOW2017 is the day of week of Bday2017
10. Age2017 is the age of the customer in 2017. (Subtract Birth\_Date from Bday2017 and then divide by 365.25.
11. The data set should include only the variables: Customer\_Name, Birth\_Date, Bday2017, BdayDOW2017, and Age2017.
12. Bday2017should be formatted to look like a two-digit day, a three-letter month, and a four-digit year.
13. Age2017 should be formatted to display no digits after the decimal point.
14. Add a PROC PRINT step to display the 23 observations of the data set.
15. The data set supplier is in the STA5066 sub directory. Use this data set to create a new data set work.region
16. Create three new variables Discount, DiscountType and Region.
17. If country is equal to CA or US

Discount is 10%.

DiscountType is equal to Required

Region is equal to North America

1. If country is any other value

Discount is equal to 5%.

DiscountType is equal to Optional.

Region is equal to Not North America.

1. The data set work.region should include only the variables: Supplier\_Name, Country, Discount, DiscountType, and Region.
2. Use a PROC PRINT step to display the data in work.region
3. The SAS data set ORDERS is in the STA5066 sub directory. Use the data set to create a new data set work.ordertype
4. Create the new variable DayOfWeek, which is equal to the week day of Order\_Date.
5. Create a new variable Type, that is equal to

Catalog Sale if Order\_Type is equal to 1

Internet Sale if Order\_Type is equal to 2

Retail Sale if Order\_Type is equal to 3.

1. Create the new variable SaleAds, which is equal to

Mail if Order\_Type is equal to 1

Email if Order\_Type is equal to 2.

1. The data set work.ordertype should not include the variables Order\_Type, Employee\_ID, and Customer\_ID.
2. Use a PROC PRINT step to display the first 25 observations of the data in work.ordertype.
3. The data set nonsales is in the STA5066 directory. Use this data set to create a new data set work.gifts.
4. Create two new variables, Gift1 and Gift2, using a SELECT group with WHEN statements. (Google to find documentation of the select construct.)
5. If Gender is equal to F:

Gift1 is equal to Perfume.

Gift2 is equal to Cookware.

1. If Gender is equal to M:

Gift1 is equal to Cologne.

Gift2 is equal to Lawn Equipment.

1. If Gender is not equal to F or M:

Gift1 is equal to Coffee.

Gift2 is equal to Lawn Calendar.

1. work.gifts should include only the variables: Employee\_ID, First, Last, Gift1, and Gift2.
2. Use a PROC PRINT step to display the first 27 records of work.gifts
3. The data set staff is in the STA5066 sub directory. Use this data set to create a new data set work.increase
4. Use a where statement to select only the observations that have Emp\_Hire\_Date on or after July 1, 2006.
5. Create a new variable increase that is 10% of salary
6. Use the sum function to create a variable NewSalary that is the sum of salary and increase.
7. Use a subsetting IF statement to select only the observations that have an increase greater than 3000.
8. Include a comma10. format for the variables Salary, Increase, and NewSalary of work.increase
9. work.increase should contain only the variables Employee\_ID, Emp\_Hire\_Date, Salary, Increase, and NewSalary
10. Use a PROC PRINT step to display the data portion of work.increase
11. The data set employee\_donations is in the STA5066 subdirectory. Use this data set to create a new data set work.bigdonations.
12. Use the sum function to create the new variable Total, which is equal to sum of Qtr1, Qtr2, Qtr3, and Qtr4.
13. Create a new variable NoDonation, which is equal to the count of missing values in Qtr1, Qtr2, Qtr3, and Qtr4. Use the NMISS function. (google to find documentation on the nmiss function)
14. The data set should contain only observations meeting the following two conditions (Use an IF-THEN DELETE statement to eliminate the observations where the conditions are not met.):

Total values greater than or equal to 50

NoDonation values equal to 0.

1. Use a PROC PRINT step to display the data portion of work.bigdonations. Only print the variables: Employee\_ID, Qtr1, Qtr2, Qtr3, Qtr4, Total, and NoDonation.
2. The SAS data set adult is in the 5066/Nhanes3 sub directory. Use this data set to create a data set work.diabetes.
3. Include a keep option in the set statement to keep only the variables SEQN, DMARETHN, HSSEX, HSAGEIR, HAD1, HAD3, HAD4
4. Include a drop option in the data statement to drop HAD1, HAD3, and HAD4 on work.diabetes.
5. Drop any observations for which HAD1 is neither 1 nor 2.
6. Define a new variable diabetic that is 1, if had1 is 1 and 0 otherwise.
7. If the patient is female (hssex=2) and diabetic=1 and had4=2 then diabetic should be changed to 0.
8. Use a PROC FREQ step to obtain a one way frequency of the variable diabetic. The correct number of diabetics (diabetic=1) is 1509.