Instruction

- Please save your solution in **one single file** with the *.sas* extension.
- Please e-mail me (arthur.x.li@me.com) your assignment by the due date. You only need to send me your SAS code.
- Please do not send me the output and the log files.
- Please do not send me a zip file.
- When writing your homework, please follow the template on the last page of the assignment.

Problems from Chapter 7 & Chapter 10

This homework will be based on the materials from Chapters 7 & Chapter 10 of the textbook. I strongly recommend that you do all the exercises at the end of each chapter. However, you are only required to turn in one problem of your choice from each chapter. Since you can obtain the solution from the publisher's website, these two problems will be graded on completion only.

Problem 1

The *inventory.sas7dbat* data set contains the number of boxes of cookies for each person that is planning to sell:

	ID	QUANTITY
1	A	10
2	B	20
3	C	10
4	D	9

The *sales.sas7dbat* data set contains the number of boxes of cookies that some of the people have already sold:

	ID	SOLD
2	B	5
3	C	2

Create a data set that contains the number of cookies remaining by using match-merging. Note, you only need two SORT procedures and *one* DATA step to solve this problem. The final data should have three variables: ID, SOLD, and LEFT. The listing output should be look like the one below:

```
The SAS System

Obs    id    quantity    sold    left

 1     A        10         .      10
 2     B        20         5      15
 3     C        10         2       8
 4     D         9         .       9
```

Problem 2

You will use two data sets: *geocode.sas7bdat* and *households.sas7bdat*. These data sets were originally downloaded from the US Census Bureau. The description of these two data sets is listed below:

geocode.sas7bdat:

VARIABLE	TYPE	DESCRIPTION	EXAMPLE
GEOID	CHAR	9 - digit Geography Code	04000US34
STATE	CHAR	State Name	New Jersey

households.sas7bdat:

VARIABLE	TYPE	DESCRIPTION	EXAMPLE
GEOID	NUM	1- or 2-digit Geography Code	34
TOTHOUSE	NUM	Total Households	3064645
UNMARRIED	NUM	Total Unmarried-Partner Households	151318

The data set *geocode.sas7bdat* contains 51 observations and the data set households contains 52 observations. For this problem, you will need to create one single data set that contains the variables STATE, ID (in 1- or 2-digits), TOTHOUSE, and UNMARRIED, and only contains observations that occur from both data sets. Notice that the last two digits from the 9-digit geography code are the same as the 2-digit geography codes. When you combine these two data sets, be careful about the variable type. The first five observations of your final data set should look similar to the one below:

```
The SAS System

Obs    state                  id    tothouse    unmarried

  1    Alabama                 1     1737080       58537
  2    Alaska                  2      221600       16568
  3    Arizona                 4     1901327      118196
  4    Arkansas                5     1042696       40543
  5    California              6    11502870      683516
```

Problem 3

You will work with the *sbp.sas7bdat* dataset. Here are the first and last 10 observations of the data set.

Obs	id	visit	sbp
1	125F	1	122
2	13000M	1	.
3	13120M	1	116
4	13260M	1	122
5	13480M	1	.
6	13520M	1	132
7	13580M	1	116
8	1750F	1	120
9	2000F	1	302
10	21300F	1	120
...			
...			
45	x24950F	4	.
46	x24950F	5	122
47	x5925F	2	118
48	x8120M	2	394
49	x9380M	2	120
50	x9450F	2	118
51	x9500F	2	120
52	x9800M	2	124
53	x9800M	3	.
54	x9900M	2	122

This dataset contains SBP (Systolic Blood Pressure) measurements for each patient. Some patients were measured once and some were measured more than once. The description of each variable is described below:

VARIABLE	**DESCRIPTION**
ID	Patient ID
VISIT	The visiting time
SBP	Systolic blood pressure

The gender of each patient can be identified from the last field of the ID variable with 'M' for Male or 'F' for Female. The gender field can be in either upper or lower cases. Some of the IDs start with an 'x' and some don't. For example, 'x13260M' and '13260M' refer to the same person. Based on this dataset, create a new dataset that contains only one observation for each patient with the following variables:

- NEWID: is created by using the numerical fields of the ID variable. For example, if ID is 'x13260M' or '13260M', then NEWID is '13260'
- SEX: is created by using the last field of the ID variable, with values = M or F (all in upper cases)
- MAXSBP: is the maximum SBP for each patient. If an SBP is measured greater than 300, then the SBP value should be considered as invalid. In this case, the SBP is considered as the missing value. Some patients might only contain one missing value for this variable

- MAXVISIT: is the visiting time that corresponds to the maximum SBP when the patient was measured

For example, consider the following scenarios:

- If a patient was measured three times with
  - SBP = 120 when visit = 1
  - SBP = 142 when visit = 2
  - SBP = 132 when visit = 3

  then MAXSBP = 142 and MAXVISIT = 2
- If a patient was measured three times with
  - SBP = 120 when visit = 1
  - SBP = 305 when visit = 2
  - SBP = 132 when visit = 3

  then MAXSBP = 132 and MAXVISIT = 3
- If a patient was measured three times with
  - SBP = 120 when visit = 1
  - SBP = missing (.) when visit = 2
  - SBP = 132 when visit = 3,

  then MAXSBP = 132 and MAXVISIT = 3
- If a patient was measured once with SBP equals either missing (.) or a value greater than 300, then MAXSBP = . and MAXVISIT = 1
- If a patient was measured twice with
  - SBP = missing (.) when visit = 1
  - SBP = 320 when visit = 2

  then MAXSBP = . and MAXVISIT = 1

The Resulting data set looks like the one below:

Obs	newID	Sex	maxsbp	maxvisit
1	125	F	122	1
2	13000	M	.	1
3	13120	M	116	1
4	13260	M	122	1
5	13480	M	124	2
6	13520	M	132	1
7	13580	M	124	3
8	1750	F	120	1
9	2000	F	.	1
10	21300	F	120	1
11	21525	F	124	1
12	24950	F	130	2
13	4050	F	.	1
14	4250	F	.	1
15	4300	F	124	1
16	5200	F	116	1
17	5925	F	118	2
18	7200	F	.	1
19	8120	M	126	1
20	9020	M	120	1
21	9380	M	128	1
22	9450	F	118	2
23	9500	F	124	1
24	9775	F	122	1
25	9800	M	124	2
26	9900	M	122	2