

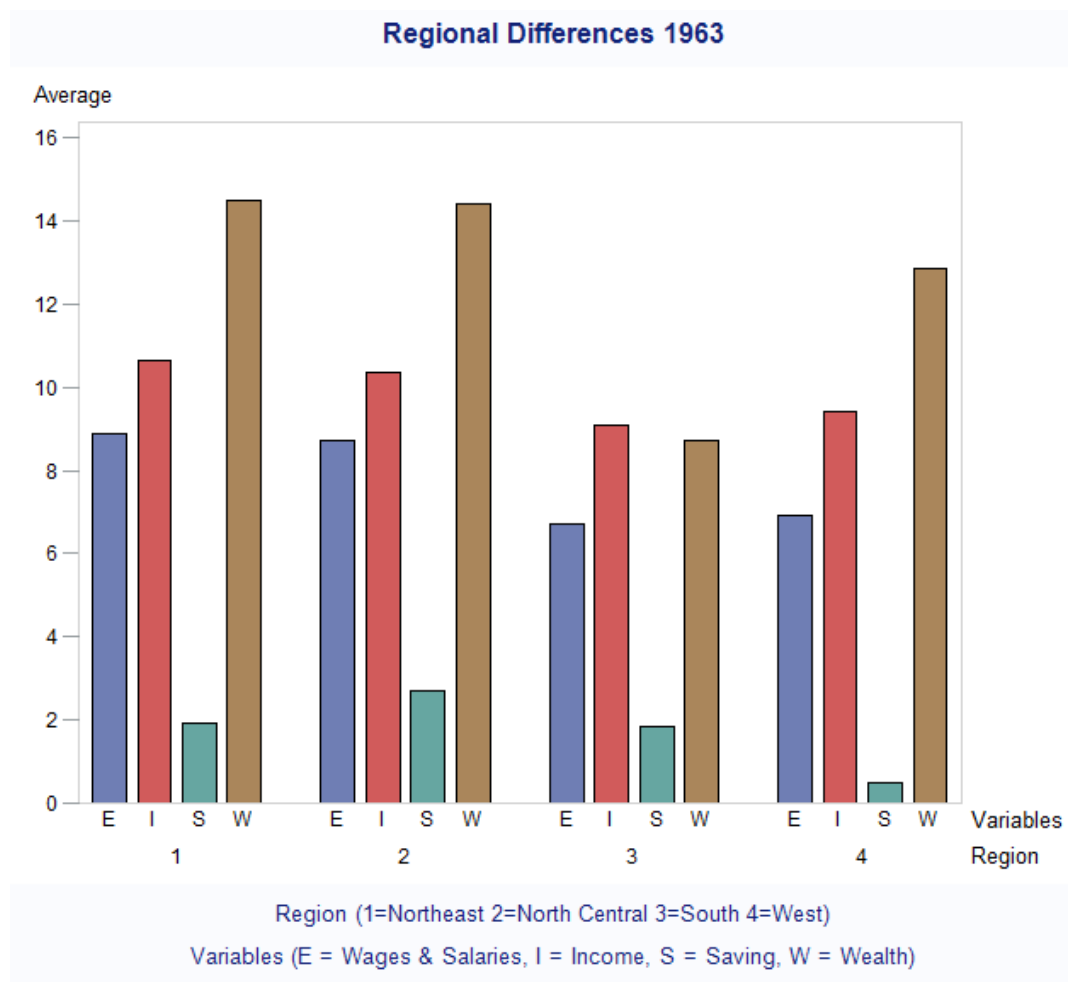
### Asgn 3

Please note: The assignment is due by 3/10 5:00 PM. We are expected to work alone on this assignment, and we must not receive or give help for any of the questions.

Please submit via Moodle a single SAS program file, named yourusername\_asgn3.sas that you wrote to answer the following three questions (note it should be a SAS program file not a project file). Please include brief comments that explain what you are doing so that a person not very familiar with SAS would be able to follow your program/work.

1. The data in P:\QAC\qac157\data\family63.sas7bdat represent a very small sample (100 observations extracted for demonstration purposes) from a couple of surveys on family finances and financial consumer characteristics carried out for the Board of Governors of the Federal Reserve System in 1963. Here are some of the variables: region (1=Northeast 2=North Central 3=South 4=West); E is the wage and salary earning of the head expressed in thousands of dollars; I is the family total income, in thousands; W is Family wealth, in thousands; S is the family saving (flow), in thousands.

Please write a SAS program to produce the following graph (this program should be the first part of the SAS file you submit via Moodle and should include all the necessary steps)



2. The stress lab – a psychology lab at WisdomU – needs your help. It seems that the research assistants did not follow the established data recording conventions and as a result it is not easy for them to move ahead with their analyses. The lab wants you to write a SAS program that will help them put all the data together. They have several hundred participants and for each participant they have measurements at 5 different times. They gave us data on only five participants in three different files that reflect the different ways that RAs have recorded the data. The following data files are in P:\QAC\qac157\STUDENTS\mkaparakis\Assignments\asn3

CardioA.txt is a tab delimited file for participants 1-4 and in addition to id and age it includes blood pressure data for the five different times blood pressure was measured (trials). The number at the end of the variable name shows the time period (trial) that the measure was taken. Note that age is missing for subjects 1 and 3, and the fourth measurement of blood pressure is missing for subject 2

CardioB.dat has pulse rate data on the same four participants. The number again in the variable name reflects the time period that the measurement was taken, but this time it is in the middle of the variable name.

CardioC.txt is an “unformatted” data file with the data for participant 5. Here each row represents measurements for the different time periods (variable t)

The fourth data file Cardio\_updates.xls includes data that were missing in CardioA.txt (Hint: update)

Please write a program (should be the second part of the file you submit) to process these files as needed so that at the end you have all participants in a single file (long format) like this

id	trial	age	bp	pu
1	1	40	115	54
1	2	40	86	87
1	3	40	129	93
1	4	40	105	81
1	5	40	127	92
2	1	30	123	92
2	2	30	136	88
2	3	30	107	125
2	4	30	111	87
2	5	30	120	58
3	1	16	124	105
3	2	16	122	97
3	3	16	101	128
3	4	16	109	57
3	5	16	112	68
4	1	23	105	52
4	2	23	115	79
4	3	23	121	71
4	4	23	129	106
4	5	23	137	39
5	1	18	116	70
5	2	18	128	64
5	3	18	112	52
5	4	18	125	68
5	5	18	111	59

3. The data set bp.dta in P:\QAC\qac157\students\mkaparakis\Assignments\asgn3 contains measurements of systolic blood pressure (bp1-bp5) for five individuals from five trials. We want to create a dummy variable to indicate whether a blood pressure measurement was high (greater than 130) in a particular trial. We can do it using a series of statements like  
hipb1=(bp1>=130);  
...

```
hipb5=(bp5>=130);
```

in our data step which could be the easiest way when you have a small number of trials, but we anticipate to have a very large number of trials and participants in our study. In anticipation of the large data file let us put our understanding of arrays and do loops to work in this small data set. So, Please use arrays and do loops to “automate” the creation of the dummy variables (i.e. create new variables that take the value one if bp”x”>=130 , zero otherwise).