STAT 541: Final Exam

1. Suppose age charts for the bass example in Chapter 16 had been provided in "long" format (see Age dataset at end of this test). Length and weight classes appear in a table after the Age dataset.

   (a) Use PROC TRANSPOSE to convert the data to array format so that it has the appearance shown below.

Length	Weight1	Weight2	Weight3	Weight4	Weight5
1	1	1	.	.	.
2	1	2	2	3	.
3	.	2	3	3	4
4	.	.	3	4	4
5	.	.	3	4	5

   (b) Assign ages to fish with the following lengths and weights using informats and a merge statement. Input the capture dates as listed using an INFORMAT statement.

Fish	Length	Weight	Capture Date
1	523	1340	8/27/2013
2	535	1297	8/27/2013
3	397	1020	8/31/2013
4	615	2115	8/31/2013

   (c) Suppose a fifth fish was added to the above data set on 9/2/2013 with weight=850 and length=510. Is an appropriate age class returned?

   (d) Return to the match merge with only 4 entries. Use a picture format for Age to generate a printed data set with the following format for Age:

ID	Age	Date	TL	WW	Length
1	3-year-old bass	08/27/13	523	3	3
2	2-year-old bass	08/27/13	535	2	3
3	1-year-old bass	08/31/13	397	2	1
4	4-year-old bass	08/31/13	615	4	4

2. Consider the champion tree data from Test 2 (see data set Tree at the end of this test).

   Suppose the researcher returned with more precise measuring tools and recorded additional measurements, including height (in feet) and crown width (average of two measures of the spread of the tree's top, taken at right angles to each other, in feet).

ID	Sector	Scientific Name	Common Name	cbh	Ht	Crown
QL1				.	123	87.5
FG1				.	108	.
FG3	Frenchman's Pond	Fagus grandifolia	American Beech	123	130	60.5
PT2	Muck Swamp East	Pinus taeda	Loblolly Pine	.	139	85

(a) Update the database using UPDATE and BY without necessarily trying to resolve all mistakes, if any. How well did this approach handle missing data in the input data set?

(b) Use a Full Join in PROC SQL to combine the data sets. Comment on problems in the output table; graduate students should resolve them.

3. The SAS data sets and text files on the website–**county29.txt**, **Lancaster.sas7bdat** (Lancaster); **county46.txt**, **York.sas7bdat** (York); **county28.txt**, **Kershaw.sas7bdat** (Kershaw); **county20.txt**, **Fairfield.sas7bdat** (Fairfield)–were extracted from the comma-delimited 2012 Traffic count data set. (*Note:* If you have trouble copying data sets or text files from the course website, all four pairs were created using commands similar to the commands below for Lancaster county. The code below assumes that traffic_count_data_2012.txt was imported into WORK as a comma-delimited text file with no row names):

```
PROC IMPORT OUT= WORK.Traffic
            DATAFILE= "d:\STAT 541\traffic_count_data_2012.txt"
            DBMS=DLM REPLACE;
     DELIMITER='2C'x;
     GETNAMES=NO;
     DATAROW=1;
RUN;
libname stat541 'd:\stat 541\final14\';
run;
data stat541.lancaster; set traffic (rename=(VAR1=county VAR2=site
 var3=roadtype  var4=roadnum  var6=aadt  var8=desc) drop=var5 var7);
 if county=29;
run;
proc contents data=stat541.lancaster; run;
PROC EXPORT DATA= STAT541.lancaster
            OUTFILE= "D:\STAT 541\Final14\county29.txt"
            DBMS=DLM REPLACE;
     DELIMITER='2C'x;
     PUTNAMES=NO;
RUN;
```

(a) Use a FILENAME statement, then INFILE and INPUT statements (Chapter 14) to successively add each of the comma-delimited data sets into a new SAS data set named Upper_Piedmont.

(b) Repeat the above step with INFILE and INPUT statements, but do not use FILENAME.

4. Suppose the four SAS data sets **Lancaster.sas7bdat**, **York.sas7bdat**, **Kershaw.sas7bdat** and **Fairfield.sas7bdat** above were the only files in a directory that was also a permanent SAS library called FINAL541. Write a macro that does not use the specific

data set names to append the data sets by the method of your choice (I used dictionary.tables to create a file of SAS data set names, then composed my macro).

Tree ID	Sector	SciName	Name	CBH
QL1	Frenchman's Pond	Quercus laurifolia	Laurel Oak	276
FG1	Frenchman's Pond	Fagus grandifolia	American Beech	121
FG2	Frenchman's Pond	Fagus grandifolia	American Beech	109
PT1	Muck Swamp East	Pinus taeda	Loblolly Pine	171
PT2	Muck Swamp East	Pinus taeda	Loblolly Pine	196
QL2	Muck Swamp East	Quercus lyrata	Overcup Oak	196
LS1	Muck Swamp East	Liquidambar styraciflua	Sweetgum	195

Table 1: Data Set Tree: Tree ID, Sector, Latin name, common name, and circumference at breast height for Champion Tree research project at Congaree National Park

Length	Age	WW
1	1	1
1	1	2
2	1	1
2	2	2
2	2	3
2	3	4
3	2	2
3	3	3
3	3	4
3	4	5
4	3	3
4	4	4
4	4	5
5	3	3
5	4	4
5	5	5

Table 2: Data Set Age: Length class, Age in year, and Weight class for aged striped bass

Weight	Weight Class	Length	Length Class
< 900	1	< 400	1
901-1300	2	401-500	2
1301-2000	3	501-600	3
2001-2500	4	601-675	4
> 2500	5	> 675	5

Table 3: Conversion table for length and weight classes