Outline

• What is SAS?
• Things you need to know before using SAS
• SAS user interface
• Using SAS to manage data
• Strengths and limitations of SAS
• Conclusions
What Is SAS?

• SAS is one of statistical software packages, just like Stata or SPSS.
• What does SAS do?
  – Data management
  – Data analysis
  – Ability to use graphs to present analysis results
• How does SAS differ from other statistical software?
  – Different user interface
  – Different data format
  – Different syntax rules
  – Can handle large data files
  – Can work on multiple data sets simultaneously
Things You Need to Know Before Using SAS

• SAS reads one observation at a time.

• There are two steps in using SAS to analyze data. The DATA step is for data management and the PROC step is for statistical analysis.

• The default setting of SAS is to construct and analyze data in a working directory. The content of this directory will be deleted when SAS is closed. Thus, if you want to keep your work permanently, you need to save it in another directory.

• Numeric variables are different from string variables.
Things You Need to Know Before Using SAS (Cont.)

- Make references to directories, files, variables
  - _NUMERIC_ specifies all numeric variables that are currently defined in the DATA step.
  - _CHARACTER_ specifies all string variables that are currently defined in the DATA step.
  - _ALL_ specifies all the numeric and string variables that are currently defined in the DATA step.
  - The colon “:” represents a wild card. For example, you can use “m:” to refer to two variables, make and mpg, simultaneously.
  - Two dashes “--” represents “from … to”. Thus, “make—mpg” represents make and mpg and every variables between them.
  - The statement “Variable A -numeric- variable B” represents all numeric variables from Variables A to B.
  - The statement “Variable A -character- Variable B” represents all the string variables from variables A and B.
Things You Need to Know Before Using SAS (Cont.)

- Know the differences among variables, variable values, variable format.

- Some important rules of writing SAS syntax
  - A SAS statement can start in any column.
  - A SAS statement can be in upper- or lower-case.
  - A SS statement can be on the same line as other statements.
  - A Ss statement can continue on the next line (as long as you don't split a word in two)
  - Every SAS statement ends with a semicolon.
• Rules of naming a variable
  – Names must start with a letter or an underscore (_).
  – Names can contain only letters, numerals, or underscores (_), but not other special characters, such as %, $, !, *, &, #, or @.
  – Names must be 32 characters or fewer in length.
  – Letters can be uppercase, lowercase, or a mixture of the case
  – The above rules applied to naming format
  – Similar rules applied to naming other things in SAS (e.g., informat, libref, and fileref), except that the length of name is 31 characters for a informat, 8 characters for a libref and a fileref must be 8 characters or fewer in length, and member names for versioned data sets must be 28 characters or fewer.
  – Don’t use the following names: _N_, _ERROR_, _CHARACTER_, _NUMERIC_, _ALL_, SASHELP, SasmSG, SASUSER, WORK, _NULL_, _DATA_, _LAST_, SASCAT, and SYS. SAS already uses these names for special purposes.

• Missing values is treated as the smallest possible value.

• Two ways of Adding comments in SAS command files
  – * this is a comment ;
  – This is a comment
SAS User Interface

• Three main windows
  – Explorer window for looking at the data
  – Editor window for writing a SAS command file
  – Log window for errors in the SAS program

• An additional window – the output window
  – The output window automatically pops up after you execute a SAS command file that produces SAS outputs

• The steps of using SAS
  – Using Editor window to write a SAS command file
  – Execute the command file
  – Check if there are the error messages in the log window
  – Check the output in the output window
Using SAS to Manage Data

• What are data?
• Use SAS to read in data
• Take a look at the data file
• Change the order of observations or variables
• Modify variables
• Add labels
• Create new variables
• Merge data
• Create a subset of data
## What Are Data?

### Raw data:

<table>
<thead>
<tr>
<th>make</th>
<th>price</th>
<th>mpg</th>
<th>rep78</th>
<th>headroom</th>
<th>trunk</th>
<th>weight</th>
<th>length</th>
<th>turn</th>
<th>displacement</th>
<th>gear_ratio</th>
<th>foreign</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC Concord</td>
<td>40999</td>
<td>22</td>
<td>3</td>
<td>2.5</td>
<td>11</td>
<td>2930</td>
<td>186</td>
<td>40</td>
<td>121</td>
<td>3.58</td>
<td>Domestic</td>
</tr>
<tr>
<td>AMC Pacer</td>
<td>47491</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>3350</td>
<td>173</td>
<td>40</td>
<td>258</td>
<td>2.53</td>
<td>Domestic</td>
</tr>
<tr>
<td>AMC Spirit</td>
<td>37992</td>
<td>22</td>
<td>3</td>
<td>3</td>
<td>12</td>
<td>2640</td>
<td>168</td>
<td>35</td>
<td>121</td>
<td>3.08</td>
<td>Domestic</td>
</tr>
<tr>
<td>Buick Century</td>
<td>48162</td>
<td>20</td>
<td>3</td>
<td>4.5</td>
<td>16</td>
<td>3250</td>
<td>196</td>
<td>40</td>
<td>196</td>
<td>2.93</td>
<td>Domestic</td>
</tr>
<tr>
<td>Buick Electra</td>
<td>78271</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>20</td>
<td>4080</td>
<td>222</td>
<td>43</td>
<td>350</td>
<td>2.41</td>
<td>Domestic</td>
</tr>
<tr>
<td>VW Dasher</td>
<td>71402</td>
<td>23</td>
<td>4</td>
<td>2.5</td>
<td>12</td>
<td>2160</td>
<td>172</td>
<td>36</td>
<td>97</td>
<td>3.74</td>
<td>Foreign</td>
</tr>
<tr>
<td>VW Diesel</td>
<td>53974</td>
<td>41</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>2040</td>
<td>155</td>
<td>35</td>
<td>90</td>
<td>3.78</td>
<td>Foreign</td>
</tr>
<tr>
<td>VW Rabbit</td>
<td>46972</td>
<td>25</td>
<td>4</td>
<td>3</td>
<td>15</td>
<td>1930</td>
<td>155</td>
<td>35</td>
<td>89</td>
<td>3.78</td>
<td>Foreign</td>
</tr>
<tr>
<td>VW Scirocco</td>
<td>68502</td>
<td>25</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>1990</td>
<td>156</td>
<td>36</td>
<td>97</td>
<td>3.78</td>
<td>Foreign</td>
</tr>
<tr>
<td>Volvo 260</td>
<td>119951</td>
<td>17</td>
<td>5</td>
<td>2.5</td>
<td>14</td>
<td>3170</td>
<td>193</td>
<td>37</td>
<td>163</td>
<td>2.98</td>
<td>Foreign</td>
</tr>
</tbody>
</table>

### Final data:

<table>
<thead>
<tr>
<th># of observation</th>
<th>make</th>
<th>price</th>
<th>mpg</th>
<th>rep78</th>
<th>headroom</th>
<th>trunk</th>
<th>weight</th>
<th>length</th>
<th>turn</th>
<th>displacement</th>
<th>gear_ratio</th>
<th>foreign</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AMC Concord</td>
<td>40999</td>
<td>22</td>
<td>3</td>
<td>2.5</td>
<td>11</td>
<td>2930</td>
<td>186</td>
<td>40</td>
<td>121</td>
<td>3.58</td>
<td>Domestic</td>
</tr>
<tr>
<td>2</td>
<td>AMC Pacer</td>
<td>47491</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>3350</td>
<td>173</td>
<td>40</td>
<td>258</td>
<td>2.53</td>
<td>Domestic</td>
</tr>
<tr>
<td>3</td>
<td>AMC Spirit</td>
<td>37992</td>
<td>22</td>
<td>3</td>
<td>3</td>
<td>12</td>
<td>2640</td>
<td>168</td>
<td>35</td>
<td>121</td>
<td>3.08</td>
<td>Domestic</td>
</tr>
<tr>
<td>4</td>
<td>Buick Century</td>
<td>48162</td>
<td>20</td>
<td>3</td>
<td>4.5</td>
<td>16</td>
<td>3250</td>
<td>196</td>
<td>40</td>
<td>196</td>
<td>2.93</td>
<td>Domestic</td>
</tr>
<tr>
<td>5</td>
<td>Buick Electra</td>
<td>78271</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>20</td>
<td>4080</td>
<td>222</td>
<td>43</td>
<td>350</td>
<td>2.41</td>
<td>Domestic</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>VW Dasher</td>
<td>71402</td>
<td>23</td>
<td>4</td>
<td>2.5</td>
<td>12</td>
<td>2160</td>
<td>172</td>
<td>36</td>
<td>97</td>
<td>3.74</td>
<td>Foreign</td>
</tr>
<tr>
<td>71</td>
<td>VW Diesel</td>
<td>53974</td>
<td>41</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>2040</td>
<td>155</td>
<td>35</td>
<td>90</td>
<td>3.78</td>
<td>Foreign</td>
</tr>
<tr>
<td>72</td>
<td>VW Rabbit</td>
<td>46972</td>
<td>25</td>
<td>4</td>
<td>3</td>
<td>15</td>
<td>1930</td>
<td>155</td>
<td>35</td>
<td>89</td>
<td>3.78</td>
<td>Foreign</td>
</tr>
<tr>
<td>73</td>
<td>VW Scirocco</td>
<td>68502</td>
<td>25</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>1990</td>
<td>156</td>
<td>36</td>
<td>97</td>
<td>3.78</td>
<td>Foreign</td>
</tr>
<tr>
<td>74</td>
<td>Volvo 260</td>
<td>119951</td>
<td>17</td>
<td>5</td>
<td>2.5</td>
<td>14</td>
<td>3170</td>
<td>193</td>
<td>37</td>
<td>163</td>
<td>2.98</td>
<td>Foreign</td>
</tr>
</tbody>
</table>
What Are Data? (Continued)

• The final data set looks just like an Excel table.

• Each column represents a variable, except the first column that I added to indicate the number of observations in the data.

• Each row represents an observation, except the first row that I added to indicate the name of each variable.

• The purpose of data management is to make a change or changes to this Excel table, for example,
  - You can change the value of a variable for some or all observations
  - You can change the name or attribute of a variable
  - You can add new variables, new observations, or both.
Using SAS to Read in Data

• If you have a SAS system file (i.e., auto.sas7bdat) stored in a directory (c:\temp\in) and you want to save it to another directory (c:\temp\out)

  LIBNAME in “c:\temp\in”;  
  LIBNAME out “c:\temp\out”;  
  DATA out.auto2;  
  SET “in.auto”;  
  RUN;

• If you have SAS export file (i.e. auto.xpt or “auto.exp) stored in a directory (c:\temp\in) and you want to save it to another directory (c:\temp\out)

  LIBNAME in xport “c:\temp\in\auto.xpt”;  
  LIBNAME out “c:\temp\out”;  
  DATA out.auto2;  
  SET in.auto;  
  RUN;
LIBNAME new c:\temp\out;
FILENAME datafile C:\temp\in\auto.dat;
DATA new.auto;
INFILE datafile LRECL=59;
INPUT make $ 1-17 price 18-22 mpg 23-24 rep78 25 headroom 26-28
      Trunk 29-30 weight 31-34 length 35-37 turn 38-39 displacement 40-42
      gear_ratio 43-58 foreign 59;
LABEL make = "Make and Model"
      price = "Price"
      mpg = "Mileage (mpg)"
      rep78 = "Repair Record 1978"
      headroom = "Headroom (in.)"
      trunk = "Trunk space (cu. ft.)"
      weight = "Weight (lbs.)"
      length = "Length (in.)"
      turn = "Turn Circle (ft.)"
      displacement = "Displacement (cu. in.)"
      gear_ratio = "Gear Ratio"
      foreign = "Car type" ;
RUN;
Use SAS to Read in Data (Continued)

The relation between the command file and the raw data file

12345678901234567890123456789012345678901234567890123456789
AMC Concord 40992232.5112930186401213.579999923706050

• if the data are in Excel, Stata, or SPSS format,
  – Use Stat/transfer software to translate the file into a SAS system file.

• If you need to input the data yourself,
  – Type data into an excel file, and use Stat/Transfer to transfer it into a SAS file
Take a Look at the Data

• Find the attribute of data
  PROC CONTENTS DATA = new.auto position;
  RUN;

• Take a look at the values of a variable
  – Summary statistics for numeric variable
    PROC MEANS DATA = new.auto;
    VAR price mpg;
    RUN;
  – Frequencies for both numeric and string variables
    PROC FREQ DATA = new.auto;
    TABLES make price mpg;
    RUN;

• Take a Look at the variable for some observations
  PROC PRINT DATA = new.auto (firstobs = 1 obs = 60);
  VAR make price mpg foreign;
  WHERE (mpg <=20 and foreign =0);
  RUN;
Take a Look at Data (Continued)

- **SAS operators**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Symbol</th>
<th>Text</th>
<th>Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>=</td>
<td>eq</td>
<td>Equal</td>
<td></td>
</tr>
<tr>
<td>^=</td>
<td>ne</td>
<td>Not equal</td>
<td></td>
</tr>
<tr>
<td>&gt;</td>
<td>gt</td>
<td>Greater than</td>
<td></td>
</tr>
<tr>
<td>&lt;</td>
<td>lt</td>
<td>Less than</td>
<td></td>
</tr>
<tr>
<td>&gt;=</td>
<td>ge</td>
<td>greater than or equal to</td>
<td></td>
</tr>
<tr>
<td>&lt;=</td>
<td>le</td>
<td>less than or equal to</td>
<td></td>
</tr>
</tbody>
</table>

**Logical**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&amp;</td>
<td>and</td>
<td>both</td>
<td></td>
</tr>
<tr>
<td></td>
<td>or</td>
<td>either</td>
<td></td>
</tr>
<tr>
<td>^</td>
<td>not</td>
<td>not true</td>
<td></td>
</tr>
</tbody>
</table>

**Arithmetic**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td></td>
<td>Addition</td>
</tr>
<tr>
<td>-</td>
<td></td>
<td>subtraction</td>
</tr>
<tr>
<td>*</td>
<td></td>
<td>Multiplication</td>
</tr>
<tr>
<td>/</td>
<td></td>
<td>Division</td>
</tr>
<tr>
<td>**</td>
<td></td>
<td>Exponentiation</td>
</tr>
</tbody>
</table>
• The order of priorities of SAS operators:
  – All comparison Operators are equal
  – All Logical operators are equal.
  – Within Arithmetic Operators (Exponentiation > Multiplication or Division > Addition or Subtraction.
  – Among these three types of operators, Arithmetic operators > Logical operators > Comparison Operators
  – If you don’t know how SAS will decide the order of operators, you can use parenthesis to make sure of it.
Change Orders of Observations or Variables

• The SAS command file to sort the observation

   PROC SORT DATA=new.auto OUT=new_auto_s;
   BY mpg;
   RUN;

• The SAS commands to check if the data had been sorted

   PROC PRINT DATA=new.auto;
   VAR mpg
   RUN;
   PROC PRINT DATA=new.auto_s;
   VAR mpg;
   RUN;
Change Orders of Observations or Variables (Cont.)

• The SAS command to change the order of variable:

```sas
DATA new.auto3;
RETAIN foreign make price mpg rep78
    headroom trunk weight length turn
    displacement gear_ratio;
SET new.auto;
RUN;
```

• The SAS command for checking if the order of variable has been changed

```sas
PROC CONTENTS DATA = new.auto3 position;
RUN;
```
Modify Variables

• Change the name of a variable

```sas
DATA new.auto2;
SET new.auto (rename=(mpg=mpg2 price=price2));
RUN;
```

```sas
DATA new.auto2;
SET new.auto;
RENAME mpg =mpg2 price=price2;
RUN;
```

• Change the value of a variable

```sas
DATA new.auto2;
SET new.auto;
repair = .;
IF (rep78=1) OR (rep78=2) THEN repair = 1;
IF (rep78=3) THEN repair = 2;
IF (rep78=4) OR (rep78=5) THEN repair = 3;
RUN;
PROC FREQ DATA=new.auto2;
TABLES rep78* repair;
RUN;
```
Modify Variables (Continued)

• Change the numeric variables to string variables and vice versa

DATA new.auto2;
SET new.auto;
s_mpg = put(mpg, $8.);
n_mpg = input(s_mpg, 8.0);
lable s_mpg = "mpg as a string variable"
lable n_mpg = "mpg as a numeric variable";
RUN;

• You can use the PROC CONTENTS command to check if a variable is a numeric or string variable

PROC CONTENTS DATA = new.auto2 position;
RUN;
Add Labels

• Three types of labels: data labels, variable labels, and value labels
  – Add labels to the data and variables
    DATA new.auto2 (LABEL = "new auto data");
    SET new.auto;
    LABEL rep78 = "Repair Record in 1978"
      mpg = "Miles Per Gallon"
    foreign= "foreign or domestic car";
    RUN;
  – Check the labels of the data and variables
    PROC CONTENTS DATA =new.auto2 POSITION;
    RUN;
Add Labels (Continued)

– Add value labels
PROC FORMAT;
VALUE forgnf 0="domestic" 1="foreign" ;
VALUE $makef "AMC" ="American Motors" "Buick" ="Buick (GM)"
   "Cad." ="Cadillac (GM)" "Chev." ="Chevrolet (GM)" "Datsun"
   ="Datsun (Nissan)";
RUN;

– Use value labels
PROC FREQ DATA=auto2;
FORMAT foreign forgnf. make $makef.;
TABLES foreign make;
RUN;
Create New Variables

• Some examples of creating new variables:

```sas
DATA new.auto2;
SET new.auto;
auto=1;
lag_mpg = lag(mpg);
dif_mpg = dif(mpg);
If  rep78 >=3 then dummy =1;
else if rep78 <3 and rep78 ne . then dummy =0;
else dummy =.;
dummy2 = dummy*2;
interact = foreign*price;
RUN;
```

• Check the accuracy of these new variables

```sas
PROC PRINT DATA = new.auto2;
VAR mpg lag_mpg dif_mpg rep78 dummy dummy2 foreign price interact;
RUN.
```
Merge Data

• Before you merge data files
  – How many data files do you want to merge them together?
  – Do these data sets have variables with the same name?
  – What is the ID variable used to merge these files?
  – Does the ID variable uniquely identify each observation in each data set? You can do one-to-one or one-to-many merge, but not many-to-many merge.

• Steps of merging data
  – Sort the first data file, based on the ID variable.
  – Sort the second data file, based on the ID variable.
  – Merge two data sets, with the use of the ID variable.
Merge Data (Continued)

• The example of one-to-one merge

<table>
<thead>
<tr>
<th>make</th>
<th>model</th>
<th>price</th>
<th>mpg</th>
<th>rep78</th>
<th>headroom</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC</td>
<td>Concord</td>
<td>4099</td>
<td>22</td>
<td>3</td>
<td>2.5</td>
</tr>
<tr>
<td>AMC</td>
<td>Pacer</td>
<td>4749</td>
<td>17</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>AMC</td>
<td>Spirit</td>
<td>3799</td>
<td>22</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Buick</td>
<td>Century</td>
<td>4816</td>
<td>20</td>
<td>3</td>
<td>4.5</td>
</tr>
<tr>
<td>Buick</td>
<td>Electra</td>
<td>7827</td>
<td>15</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Buick</td>
<td>LeSabre</td>
<td>5788</td>
<td>18</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Buick</td>
<td>Opel</td>
<td>4453</td>
<td>26</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Buick</td>
<td>Regal</td>
<td>5189</td>
<td>20</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Buick</td>
<td>Riviera</td>
<td>10372</td>
<td>16</td>
<td>3</td>
<td>3.5</td>
</tr>
<tr>
<td>Buick</td>
<td>Skylark</td>
<td>4082</td>
<td>19</td>
<td>3</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Table 4. The second sample data (i.e., data2)

<table>
<thead>
<tr>
<th>make</th>
<th>model</th>
<th>trunk</th>
<th>weight</th>
<th>length</th>
<th>turn</th>
<th>displacement</th>
<th>gear_ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buick</td>
<td>Opel</td>
<td>10</td>
<td>2230</td>
<td>170</td>
<td>34</td>
<td>304</td>
<td>2.87</td>
</tr>
<tr>
<td>Buick</td>
<td>Regal</td>
<td>16</td>
<td>3280</td>
<td>200</td>
<td>42</td>
<td>196</td>
<td>2.93</td>
</tr>
<tr>
<td>Buick</td>
<td>Riviera</td>
<td>17</td>
<td>3880</td>
<td>207</td>
<td>43</td>
<td>231</td>
<td>2.93</td>
</tr>
<tr>
<td>Buick</td>
<td>Skylark</td>
<td>13</td>
<td>3400</td>
<td>200</td>
<td>42</td>
<td>231</td>
<td>3.08</td>
</tr>
<tr>
<td>Cad.</td>
<td>Deville</td>
<td>20</td>
<td>4330</td>
<td>221</td>
<td>44</td>
<td>425</td>
<td>2.28</td>
</tr>
<tr>
<td>Cad.</td>
<td>Eldorado</td>
<td>16</td>
<td>3900</td>
<td>204</td>
<td>43</td>
<td>350</td>
<td>2.19</td>
</tr>
<tr>
<td>Cad.</td>
<td>Seville</td>
<td>13</td>
<td>4290</td>
<td>204</td>
<td>45</td>
<td>350</td>
<td>2.24</td>
</tr>
<tr>
<td>Chev.</td>
<td>Chevette</td>
<td>9</td>
<td>2110</td>
<td>163</td>
<td>34</td>
<td>231</td>
<td>2.93</td>
</tr>
<tr>
<td>Chev.</td>
<td>Impala</td>
<td>20</td>
<td>3690</td>
<td>212</td>
<td>43</td>
<td>250</td>
<td>2.56</td>
</tr>
<tr>
<td>Chev.</td>
<td>Malibu</td>
<td>17</td>
<td>3180</td>
<td>193</td>
<td>31</td>
<td>200</td>
<td>2.73</td>
</tr>
<tr>
<td>Chev.</td>
<td>Monte Carlo</td>
<td>16</td>
<td>3220</td>
<td>200</td>
<td>41</td>
<td>200</td>
<td>2.73</td>
</tr>
<tr>
<td>Chev.</td>
<td>Monza</td>
<td>7</td>
<td>2750</td>
<td>179</td>
<td>40</td>
<td>151</td>
<td>2.73</td>
</tr>
<tr>
<td>Chev.</td>
<td>Nova</td>
<td>13</td>
<td>3430</td>
<td>197</td>
<td>43</td>
<td>250</td>
<td>2.56</td>
</tr>
</tbody>
</table>
Merge Data (Continued)

Expected result of one-to-one merge

<table>
<thead>
<tr>
<th>make</th>
<th>model</th>
<th>price</th>
<th>mpg</th>
<th>rep78</th>
<th>headroom</th>
<th>trunk</th>
<th>weight</th>
<th>length</th>
<th>turn</th>
<th>displacement</th>
<th>gear_ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC</td>
<td>Concord</td>
<td>4099</td>
<td>22</td>
<td>3</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMC</td>
<td>Pacer</td>
<td>4749</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMC</td>
<td>Spirit</td>
<td>3799</td>
<td>22</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buick</td>
<td>Century</td>
<td>4816</td>
<td>20</td>
<td>3</td>
<td>4.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buick</td>
<td>Electra</td>
<td>7827</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buick</td>
<td>LeSabre</td>
<td>5788</td>
<td>18</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buick</td>
<td>Opel</td>
<td>4453</td>
<td>26</td>
<td>3</td>
<td></td>
<td>10</td>
<td>2230</td>
<td>170</td>
<td>34</td>
<td>304</td>
<td>2.87</td>
</tr>
<tr>
<td>Buick</td>
<td>Regal</td>
<td>5189</td>
<td>20</td>
<td>3</td>
<td>2</td>
<td>16</td>
<td>3280</td>
<td>200</td>
<td>42</td>
<td>196</td>
<td>2.93</td>
</tr>
<tr>
<td>Buick</td>
<td>Riviera</td>
<td>10372</td>
<td>16</td>
<td>3</td>
<td>3.5</td>
<td>17</td>
<td>3880</td>
<td>207</td>
<td>43</td>
<td>231</td>
<td>2.93</td>
</tr>
<tr>
<td>Buick</td>
<td>Skylark</td>
<td>4082</td>
<td>19</td>
<td>3</td>
<td>3.5</td>
<td>13</td>
<td>3400</td>
<td>200</td>
<td>42</td>
<td>231</td>
<td>3.08</td>
</tr>
<tr>
<td>Cad.</td>
<td>Deville</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>4330</td>
<td>221</td>
<td>44</td>
<td>425</td>
<td>2.28</td>
</tr>
<tr>
<td>Cad.</td>
<td>Eldorado</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>3900</td>
<td>204</td>
<td>43</td>
<td>350</td>
<td>2.19</td>
</tr>
<tr>
<td>Cad.</td>
<td>Seville</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>4290</td>
<td>204</td>
<td>45</td>
<td>350</td>
<td>2.24</td>
</tr>
<tr>
<td>Chev.</td>
<td>Corvette</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>2110</td>
<td>163</td>
<td>34</td>
<td>231</td>
<td>2.93</td>
</tr>
<tr>
<td>Chev.</td>
<td>Impala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>3690</td>
<td>212</td>
<td>43</td>
<td>250</td>
<td>2.56</td>
</tr>
<tr>
<td>Chev.</td>
<td>Malibu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td>3180</td>
<td>193</td>
<td>31</td>
<td>200</td>
<td>2.73</td>
</tr>
<tr>
<td>Chev.</td>
<td>Monte Carlo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>3220</td>
<td>200</td>
<td>41</td>
<td>200</td>
<td>2.73</td>
</tr>
<tr>
<td>Chev.</td>
<td>Monza</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td>3180</td>
<td>193</td>
<td>31</td>
<td>200</td>
<td>2.73</td>
</tr>
<tr>
<td>Chev.</td>
<td>Nova</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>3430</td>
<td>197</td>
<td>43</td>
<td>250</td>
<td>2.56</td>
</tr>
</tbody>
</table>
Merge Data (Continued)

• SAS commands for one-to-one merge.

PROC SORT DATA=new.data1;
BY make model;
RUN;

PROC SORT DATA=new.data2;
BY make model;
RUN;

DATA new.merged_data;
MERGE new.data1 (IN=data1) new.data2 (IN=data2);
BY make model;
data1 = data1;
data2 = data2;
RUN;
Merge Data (Continued)

• The example of one-to-many merge

<table>
<thead>
<tr>
<th>Table 6. The Make of the Car</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>make</td>
<td>foreign</td>
</tr>
<tr>
<td>AMC</td>
<td>Domestic</td>
</tr>
<tr>
<td>Buick</td>
<td>Domestic</td>
</tr>
<tr>
<td>Cad.</td>
<td>Domestic</td>
</tr>
<tr>
<td>Chev.</td>
<td>Domestic</td>
</tr>
<tr>
<td>Dodge</td>
<td>Domestic</td>
</tr>
<tr>
<td>Ford</td>
<td>Domestic</td>
</tr>
<tr>
<td>Linc.</td>
<td>Domestic</td>
</tr>
<tr>
<td>Merc</td>
<td>Domestic</td>
</tr>
<tr>
<td>Olds.</td>
<td>Domestic</td>
</tr>
<tr>
<td>Plym.</td>
<td>Domestic</td>
</tr>
<tr>
<td>Pont.</td>
<td>Domestic</td>
</tr>
<tr>
<td>Audi</td>
<td>Foreign</td>
</tr>
<tr>
<td>BMW</td>
<td>Foreign</td>
</tr>
<tr>
<td>Fiat</td>
<td>Foreign</td>
</tr>
<tr>
<td>Honda</td>
<td>Foreign</td>
</tr>
<tr>
<td>Mazda</td>
<td>Foreign</td>
</tr>
<tr>
<td>Peugeot</td>
<td>Foreign</td>
</tr>
<tr>
<td>Renault</td>
<td>Foreign</td>
</tr>
<tr>
<td>Subaru</td>
<td>Foreign</td>
</tr>
<tr>
<td>Toyota</td>
<td>Foreign</td>
</tr>
<tr>
<td>VW</td>
<td>Foreign</td>
</tr>
<tr>
<td>Volvo</td>
<td>Foreign</td>
</tr>
</tbody>
</table>
## Table 7. The Model of the Car

<table>
<thead>
<tr>
<th>make</th>
<th>model</th>
<th>price</th>
<th>mpg</th>
<th>rep78</th>
<th>headroom</th>
<th>trunk</th>
<th>weight</th>
<th>length</th>
<th>turn</th>
<th>displacement</th>
<th>gear_ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC</td>
<td>Concord</td>
<td>4099</td>
<td>22</td>
<td>3</td>
<td>2.5</td>
<td>11</td>
<td>2930</td>
<td>186</td>
<td>40</td>
<td>121</td>
<td>3.58</td>
</tr>
<tr>
<td>AMC</td>
<td>Pacer</td>
<td>4749</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>3350</td>
<td>173</td>
<td>40</td>
<td>258</td>
<td>2.53</td>
</tr>
<tr>
<td>AMC</td>
<td>Spirit</td>
<td>3799</td>
<td>22</td>
<td>3</td>
<td>3</td>
<td>12</td>
<td>2640</td>
<td>168</td>
<td>35</td>
<td>121</td>
<td>3.08</td>
</tr>
<tr>
<td>Buick</td>
<td>Century</td>
<td>4816</td>
<td>20</td>
<td>3</td>
<td>4.5</td>
<td>16</td>
<td>3250</td>
<td>196</td>
<td>40</td>
<td>196</td>
<td>2.93</td>
</tr>
<tr>
<td>Buick</td>
<td>Electra</td>
<td>7827</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>20</td>
<td>4080</td>
<td>222</td>
<td>43</td>
<td>350</td>
<td>2.41</td>
</tr>
<tr>
<td>Buick</td>
<td>LeSabre</td>
<td>5788</td>
<td>18</td>
<td>3</td>
<td>4</td>
<td>21</td>
<td>3670</td>
<td>218</td>
<td>43</td>
<td>231</td>
<td>2.73</td>
</tr>
<tr>
<td>VW</td>
<td>Dasher</td>
<td>7140</td>
<td>23</td>
<td>4</td>
<td>2.5</td>
<td>12</td>
<td>2160</td>
<td>172</td>
<td>36</td>
<td>97</td>
<td>3.74</td>
</tr>
<tr>
<td>VW</td>
<td>Diesel</td>
<td>5397</td>
<td>41</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>2040</td>
<td>155</td>
<td>35</td>
<td>90</td>
<td>3.78</td>
</tr>
<tr>
<td>VW</td>
<td>Rabbit</td>
<td>4697</td>
<td>25</td>
<td>4</td>
<td>3</td>
<td>15</td>
<td>1930</td>
<td>155</td>
<td>35</td>
<td>89</td>
<td>3.78</td>
</tr>
<tr>
<td>VW</td>
<td>Scirocco</td>
<td>6850</td>
<td>25</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>1990</td>
<td>156</td>
<td>36</td>
<td>97</td>
<td>3.78</td>
</tr>
<tr>
<td>Volvo</td>
<td></td>
<td>260</td>
<td>11995</td>
<td>5</td>
<td>2.5</td>
<td>14</td>
<td>3170</td>
<td>193</td>
<td>37</td>
<td>163</td>
<td>2.98</td>
</tr>
</tbody>
</table>
The Expected result of merging the data for the makes and models of the car

<table>
<thead>
<tr>
<th>make</th>
<th>model</th>
<th>price</th>
<th>mpg</th>
<th>rep78</th>
<th>headroom</th>
<th>trunk</th>
<th>weight</th>
<th>length</th>
<th>turn</th>
<th>isplacement</th>
<th>gear_ratio</th>
<th>foreign</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC</td>
<td>Concord</td>
<td>4099</td>
<td>22</td>
<td>3</td>
<td>2.5</td>
<td>11</td>
<td>2930</td>
<td>186</td>
<td>40</td>
<td>121</td>
<td>3.58</td>
<td>Domestic</td>
</tr>
<tr>
<td>AMC</td>
<td>Pacer</td>
<td>4749</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>3350</td>
<td>173</td>
<td>40</td>
<td>258</td>
<td>2.53</td>
<td>Domestic</td>
</tr>
<tr>
<td>AMC</td>
<td>Spirit</td>
<td>3799</td>
<td>22</td>
<td>3</td>
<td>12</td>
<td>2640</td>
<td>168</td>
<td>35</td>
<td>121</td>
<td>3.08</td>
<td>Domestic</td>
<td></td>
</tr>
<tr>
<td>Buick</td>
<td>Century</td>
<td>4816</td>
<td>20</td>
<td>3</td>
<td>4.5</td>
<td>16</td>
<td>3250</td>
<td>196</td>
<td>40</td>
<td>196</td>
<td>2.93</td>
<td>Domestic</td>
</tr>
<tr>
<td>Buick</td>
<td>Electra</td>
<td>7827</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>20</td>
<td>4080</td>
<td>222</td>
<td>43</td>
<td>350</td>
<td>2.41</td>
<td>Domestic</td>
</tr>
<tr>
<td>Buick</td>
<td>LeSabre</td>
<td>5788</td>
<td>18</td>
<td>3</td>
<td>4</td>
<td>21</td>
<td>3670</td>
<td>218</td>
<td>43</td>
<td>231</td>
<td>2.73</td>
<td>Domestic</td>
</tr>
<tr>
<td>Buick</td>
<td>Opel</td>
<td>4453</td>
<td>26</td>
<td>3</td>
<td>10</td>
<td>2230</td>
<td>170</td>
<td>34</td>
<td>304</td>
<td>2.87</td>
<td>Domestic</td>
<td></td>
</tr>
<tr>
<td>Buick</td>
<td>Regal</td>
<td>5189</td>
<td>20</td>
<td>3</td>
<td>2</td>
<td>16</td>
<td>3280</td>
<td>200</td>
<td>42</td>
<td>196</td>
<td>2.93</td>
<td>Domestic</td>
</tr>
<tr>
<td>Buick</td>
<td>Riviera</td>
<td>10372</td>
<td>16</td>
<td>3</td>
<td>3.5</td>
<td>17</td>
<td>3880</td>
<td>207</td>
<td>43</td>
<td>231</td>
<td>2.93</td>
<td>Domestic</td>
</tr>
<tr>
<td>Buick</td>
<td>Skylark</td>
<td>4082</td>
<td>19</td>
<td>3</td>
<td>3.5</td>
<td>13</td>
<td>3400</td>
<td>200</td>
<td>42</td>
<td>231</td>
<td>3.08</td>
<td>Domestic</td>
</tr>
<tr>
<td>Cad.</td>
<td>Deville</td>
<td>11385</td>
<td>14</td>
<td>3</td>
<td>4</td>
<td>20</td>
<td>4330</td>
<td>221</td>
<td>44</td>
<td>425</td>
<td>2.28</td>
<td>Domestic</td>
</tr>
<tr>
<td>Cad.</td>
<td>Eldorado</td>
<td>14500</td>
<td>14</td>
<td>2</td>
<td>3.5</td>
<td>16</td>
<td>3900</td>
<td>204</td>
<td>43</td>
<td>350</td>
<td>2.19</td>
<td>Domestic</td>
</tr>
<tr>
<td>Cad.</td>
<td>Seville</td>
<td>15906</td>
<td>21</td>
<td>3</td>
<td>3</td>
<td>13</td>
<td>4290</td>
<td>204</td>
<td>45</td>
<td>350</td>
<td>2.24</td>
<td>Domestic</td>
</tr>
<tr>
<td>VW</td>
<td>Dasher</td>
<td>7140</td>
<td>23</td>
<td>4</td>
<td>2.5</td>
<td>12</td>
<td>2160</td>
<td>172</td>
<td>36</td>
<td>97</td>
<td>3.74</td>
<td>Foreign</td>
</tr>
<tr>
<td>VW</td>
<td>Diesel</td>
<td>5397</td>
<td>41</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>2040</td>
<td>155</td>
<td>35</td>
<td>90</td>
<td>3.78</td>
<td>Foreign</td>
</tr>
<tr>
<td>VW</td>
<td>Rabbit</td>
<td>4697</td>
<td>25</td>
<td>4</td>
<td>3</td>
<td>15</td>
<td>1930</td>
<td>155</td>
<td>35</td>
<td>89</td>
<td>3.78</td>
<td>Foreign</td>
</tr>
<tr>
<td>VW</td>
<td>Scirocco</td>
<td>6850</td>
<td>25</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>1990</td>
<td>156</td>
<td>36</td>
<td>97</td>
<td>3.78</td>
<td>Foreign</td>
</tr>
<tr>
<td>Volvo</td>
<td>260</td>
<td>11995</td>
<td>17</td>
<td>5</td>
<td>2.5</td>
<td>14</td>
<td>3170</td>
<td>193</td>
<td>37</td>
<td>163</td>
<td>2.98</td>
<td>Foreign</td>
</tr>
</tbody>
</table>

Table 8. The Expected Data of the Make and Model of the Car
Merge Data (Continued)

- SAS commands for one-to-many merge.

```sas
PROC SORT DATA=new.auto_make;
BY make;
RUN;

PROC SORT DATA=new.auto_model;
BY make;
RUN;

DATA new.auto;
MERGE new.auto_make (IN=auto_make)
    new.auto_model (IN=auto_model);
    BY make;
    auto_make = auto_make;
    auto_model = auto_model;
RUN;
```
Create a Subset of Data

- Keep certain variables
  ```
  DATA new.auto2;
  SET new.auto;
  KEEP make mpg;
  RUN;
  ```

- Delete certain variables
  ```
  DATA new.auto2;
  SET new.auto;
  DROP make mpg;
  RUN;
  ```
Create a Subset of Data (Cont.)

• Keep certain respondents
  DATA new.auto2;
  SET new.auto;
  IF REP78 ^= . ;
  RUN;

• Delete respondents
  DATA new.auto2;
  SET new.auto;
  IF REP78 = . THEN DELETE;
  RUN;
Strengths and Weaknesses of SAS

• Strengths
  – SAS has the ability to handle large data sets
  – Can open multiple data sets at the same time
  – Separate the data step from the analysis step

• Weaknesses
  – It could be difficult to use SAS to change the data format from wide to long or vise versa
  – SAS is not based on open source codes, meaning that SAS does not quickly incorporate new analysis into it. For example, combining survey weights with event history analysis is available in Stata, but not available in SAS yet.
  – It still takes a lot of time to use SAS to create graphs.
  – The documentations of SAS are sometimes hard to understand
  – Rules may change from one version of SAS to another or from one platform to another
  – Needs to move back and forth between the data step and the analysis step
Conclusion

• SAS is a powerful tool for analyzing large data sets
• You can learn SAS by writing a few SAS command files first. For example, you could learn few key commands first and then explore various options of the commands.
• Be aware of all the rules in SAS programming.
• Beware of the logical flow in using SAS
  – You need to read a data before you can do anything about it.
  – You need to sort the data sets first before you can merge them together
  – You need to have data ready before analyzing them
• Several things we have not discussed, but are very important
  – The do loop
  – Array and Macro
  – Other ways of manipulate string variables.
  – Manipulate date variables
Where to Find Help

• Useful websites
  – SAS website
    (http://support.sas.com/documentation/cdl/en/lrdict/63026/HTML/default/viewer.htm#a000293668.htm)
  – UCLA (http://www.ats.ucla.edu/stat/SAS/)
  – University of Indiana
    (http://www.indiana.edu/~statmath/stat/sas/win/index.html)
• User group (http://support.sas.com/usergroups/index.html)
• CFDR programming support
  – Hsueh-Sheng Wu @ 372-3119 or wuh@bgsu.edu