Introduction to large-scale study data management using SAS

By
Jimmy Ardiansyah, PMP

October 2, 2009
Welcome to RFTS Website!

If hoping to be pregnant or you are less than 9 weeks pregnant, the Right from the Start (RFTS) study lets you contribute to medical knowledge and help other women like yourself.

- No medicines or pills to take
- No changes to daily lifestyle or existing health care routine

We will look at how a women's diet, health, behaviors, medical history, uterine fibroids and other factors affect her pregnancy.

About 5,400 women will contribute to our understanding of early pregnancy health.

Currently, we're recruiting women in the Tennessee in the area of Nashville, Chattanooga, Knoxville and Memphis. We appreciate all of the women in North Carolina and Texas who participated in the study.

You may be eligible if you:
- Are trying to get pregnant OR
- Are less than 9 weeks pregnant AND
- Are at least 18 years old

Want more information:
- Toll Free: 1.866.346.2684
- e-mail: rfts@vanderbilt.edu OR click here
- Download Brochure: click here

Latest News

RTFS Study Publications
If you would like to see the list of publications using RFTS study data. [click here]

Ultrasound Location
Are you wondering how to get to Ultrasound location? [click here]

Due Date Calculator
You can use our simple the Due Date Calculator to determine or to calculate your likely due date of your baby. The Due Date Calculator shows the likely conception date, the end of the first and the second trimester dates - and of course the due date. [click here]

Collaboration and Approval
RFTS is a Vanderbilt University study in collaboration with the University
Right From the Start (RFTS)

**Study Facts**

- Early Pregnancy Health Study
- Started on 1999
- Vanderbilt University study in collaboration with the University of North Carolina
- Approximately 1400 variables
- Approximately 40 tables
- Approximately 8500 participants from NC, TN and TX
- Database: Oracle, MySQL and Access
- Funded by the National Institute of Child Health and Human Development (NICHD)
Physical Design Schema

- Battelle (CATI Interview)
- Convert Flat Files using SAS
- Oracle Database
- Screening Apps
- Data Cleaning Using SAS
- Cleaned SAS Dataset
- SAS Dataset Audit
- Merging Process
- Final Common Dataset
- Vital Records
- Data Cleaning Using SAS
- Medical Records
- REDCap
- Data Cleaning Using SAS
Physical Design Schema

- Battelle (CATI Interview)
- Oracle Database
- Vital Records
- Medical Records
- Screening Apps
- REDCap

- Convert Flat Files using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS

- SAS Dataset Audit
- Cleaned SAS Dataset
- Merging Process
- Final Common Dataset
Physical Design Schema

- Battelle (CATI Interview)
- Oracle Database
- Screening Apps
- Vital Records
- Medical Records
- REDCap
- Convert Flat Files using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- SAS Dataset Audit
- Cleaned SAS Dataset
- Merging Process
- Final Common Dataset
Physical Design Schema

- Battelle (CATI Interview)
- Oracle Database
- Vital Records
- Medical Records
- Screening Apps
- REDCap
- Convert Flat Files using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- SAS Dataset Audit
- Cleaned SAS Dataset
- Merging Process
- Final Common Dataset
Physical Design Schema

Battelle (CATI Interview) → Convert Flat Files using SAS

Oracle Database → Screening Apps

Vital Records → Data Cleaning Using SAS

Medical Records → REDCap

Data Cleaning Using SAS

Final Common Dataset

SAS Dataset Audit

Cleaned SAS Dataset

Merging Process
Physical Design Schema

Battelle (CATI Interview)

Oracle Database

Vital Records

Medical Records

Screening Apps

Data Cleaning Using SAS

Data Cleaning Using SAS

Data Cleaning Using SAS

Convert Flat Files using SAS

SAS Dataset Audit

Cleaned SAS Dataset

Merging Process

Final Common Dataset
Physical Design Schema

- Battelle (CATI Interview)
- Oracle Database
- Screening Apps
- Vital Records
- Medical Records
- REDCap
- Convert Flat Files using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- SAS Dataset Audit
- Cleaned SAS Dataset
- Merging Process
- Final Common Dataset
Physical Design Schema

- Battelle (CATI Interview)
- Convert Flat Files using SAS
- Oracle Database
- Screening Apps
- Vital Records
- Medical Records
- REDCap
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- SAS Dataset Audit
- Cleaned SAS Dataset
- Merging Process
- Final Common Dataset
Physical Design Schema

Battelle (CATI Interview) → Convert Flat Files using SAS

Oracle Database

Screening Apps

Vital Records

Medical Records

Data Cleaning Using SAS

Data Cleaning Using SAS

Data Cleaning Using SAS

SAS Dataset Audit

Cleaned SAS Dataset

Merging Process

Final Common Dataset
Physical Design Schema

Battelle
(CATI Interview)

Oracle Database

Vital Records

Medical Records

Screening Apps

Convert Flat Files using SAS

Data Cleaning Using SAS

Data Cleaning Using SAS

Data Cleaning Using SAS

SAS Dataset Audit

Cleaned SAS Dataset

Merging Process

Final Common Dataset
Physical Design Schema

1. Battelle (CATI Interview)
2. Convert Flat Files using SAS
3. Oracle Database
4. Screening Apps
5. Vital Records
6. REDCap
7. Medical Records
8. Data Cleaning Using SAS
9. Data Cleaning Using SAS
10. Data Cleaning Using SAS
11. SAS Dataset Audit
12. Cleaned SAS Dataset
13. Merging Process
14. Final Common Dataset
Physical Design Schema

- Battelle (CATI Interview)
- Convert Flat Files using SAS
- Screening Apps
- Oracle Database
- Vital Records
- REDCap
- Medical Records
- Data Cleaning Using SAS
- Merging Process
- Final Common Dataset
- SAS Dataset Audit
- Cleaned SAS Dataset
Physical Design Schema

- Battelle (CATI Interview)
- Oracle Database
- Vital Records
- Medical Records
- Screening Apps
- REDCap
- Convert Flat Files using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- Data Cleaning Using SAS
- Cleaned SAS Dataset
- Merging Process
- Final Common Dataset
- SAS Dataset Audit

Data Cleaning Using SAS
Physical Design Schema

Battelle (CATI Interview) → Convert Flat Files using SAS → Screening Apps → Oracle Database

Oracle Database → Vital Records → REDCap

SAS Dataset Audit → Cleaned SAS Dataset → Merging Process → Final Common Dataset
Diary/Weekly Database Schema

- Web Application
- MySQL Database
- Transfers Into .CSV Files
- Convert into SAS Dataset
- Cleaned SAS Dataset
- SAS Dataset Audit
- Reports

MySQL Database

SAS Dataset Audit

Web Application

Convert into SAS Dataset

Transfers Into .CSV Files
# Introduction to SAS

## Layout for data file PATIENTS.TXT

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
<th>Starting Column</th>
<th>Length</th>
<th>Variable Type</th>
<th>Valid Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>PatNo</td>
<td>Patient Number</td>
<td>1</td>
<td>3</td>
<td>Character</td>
<td>Numeral Only</td>
</tr>
<tr>
<td>Gender</td>
<td>Gender</td>
<td>4</td>
<td>1</td>
<td>Character</td>
<td>M or F</td>
</tr>
<tr>
<td>Visit</td>
<td>Visit Date</td>
<td>5</td>
<td>10</td>
<td>MMDDYY10</td>
<td>Any valid date</td>
</tr>
<tr>
<td>HR</td>
<td>Heart Rate</td>
<td>15</td>
<td>3</td>
<td>Numeric</td>
<td>Between 40 and 100</td>
</tr>
<tr>
<td>SBP</td>
<td>Systolic Blood Pressure</td>
<td>18</td>
<td>3</td>
<td>Numeric</td>
<td>Between 80 and 200</td>
</tr>
<tr>
<td>DBP</td>
<td>Diastolic Blood Pressure</td>
<td>21</td>
<td>3</td>
<td>Numeric</td>
<td>Between 60 and 120</td>
</tr>
<tr>
<td>Dx</td>
<td>Diagnosis Code</td>
<td>24</td>
<td>3</td>
<td>Character</td>
<td>1 to 3 digit numeral</td>
</tr>
<tr>
<td>AE</td>
<td>Adverse Event</td>
<td>27</td>
<td>1</td>
<td>Character</td>
<td>0 or 1</td>
</tr>
</tbody>
</table>
Introduction to SAS

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>001M11/11/1998</td>
<td>88140</td>
<td>80</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>002F11/13/1998</td>
<td>84120</td>
<td>78</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>003X10/21/1998</td>
<td>68190100</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>004F01/01/1999</td>
<td>101200120</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>005M05/07/1998</td>
<td>68120</td>
<td>80</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>006 08/15/1999</td>
<td>72102</td>
<td>68</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>007M08/32/1998</td>
<td>88148102</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>008M11/11/1998</td>
<td>90190100</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>009F08/08/1998210</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>010M09/25/1999</td>
<td>86240180</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>011F10/19/1999</td>
<td>40120</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>012M13/13/1998</td>
<td>68300</td>
<td>20</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>013M10/12/98</td>
<td>60122</td>
<td>74</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>014F08/23/1999</td>
<td>74108</td>
<td>64</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>015M02/02/1999</td>
<td>22130</td>
<td>90</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>016F11/13/1998</td>
<td>84120</td>
<td>78</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>017M11/12/1999</td>
<td>58112</td>
<td>74</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>018F</td>
<td>82148</td>
<td>88</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>019F04/05/1999208</td>
<td>84</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>020M06/07/1999</td>
<td>58118</td>
<td>70</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>021M15/12/1999</td>
<td>60</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>022F09/99/9999</td>
<td>10</td>
<td>20</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>023M10/10/1999</td>
<td>48114</td>
<td>82</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>024F12/31/1998</td>
<td>22</td>
<td>34</td>
<td>78</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>025F11/09/199876</td>
<td>120</td>
<td>80</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>026M01/01/1999</td>
<td>74102</td>
<td>68</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>027FNOTAVAIL</td>
<td>NA</td>
<td>166106</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>028F03/28/1998</td>
<td>66150</td>
<td>90</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>029M05/15/1998</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>030F07/07/1999</td>
<td>82148</td>
<td>84</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Introduction to SAS

The image shows an Excel spreadsheet with data, which includes a column labeled "PatNo." The data contains information such as dates and numerical values, possibly related to patient records.
Introduction to SAS

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Gender</th>
<th>Date</th>
<th>Value1</th>
<th>Value2</th>
<th>Value3</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>M</td>
<td>M</td>
<td>11/11/1998</td>
<td>88140</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>002</td>
<td>F</td>
<td>F</td>
<td>11/11/1998</td>
<td>84120</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>003</td>
<td>X</td>
<td>M</td>
<td>21/10/1998</td>
<td>68190</td>
<td>100</td>
<td>31</td>
</tr>
<tr>
<td>004</td>
<td>F</td>
<td>F</td>
<td>01/01/1999</td>
<td>101200</td>
<td>12</td>
<td>51</td>
</tr>
<tr>
<td>005</td>
<td>M</td>
<td>M</td>
<td>07/07/1998</td>
<td>68120</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>006</td>
<td>M</td>
<td>M</td>
<td>06/15/1999</td>
<td>72102</td>
<td>68</td>
<td>61</td>
</tr>
<tr>
<td>007</td>
<td>M</td>
<td>M</td>
<td>32/03/1998</td>
<td>88148102</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>008</td>
<td>F</td>
<td>F</td>
<td>11/11/1998</td>
<td>90190</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>009</td>
<td>M</td>
<td>M</td>
<td>08/08/1998</td>
<td>210</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>010</td>
<td>M</td>
<td>M</td>
<td>25/09/1999</td>
<td>86240</td>
<td>80</td>
<td>41</td>
</tr>
<tr>
<td>011</td>
<td>F</td>
<td>F</td>
<td>19/09/1999</td>
<td>40120</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>012</td>
<td>M</td>
<td>M</td>
<td>13/12/1998</td>
<td>68300</td>
<td>20</td>
<td>41</td>
</tr>
<tr>
<td>013</td>
<td>M</td>
<td>M</td>
<td>12/12/1998</td>
<td>60122</td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td>014</td>
<td>M</td>
<td>M</td>
<td>23/02/1999</td>
<td>74108</td>
<td>64</td>
<td>1</td>
</tr>
<tr>
<td>015</td>
<td>M</td>
<td>M</td>
<td>02/02/1999</td>
<td>22130</td>
<td>90</td>
<td>1</td>
</tr>
<tr>
<td>016</td>
<td>F</td>
<td>F</td>
<td>13/12/1998</td>
<td>84120</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>017</td>
<td>M</td>
<td>M</td>
<td>12/12/1999</td>
<td>58112</td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td>018</td>
<td>F</td>
<td>F</td>
<td>04/05/1999</td>
<td>82148</td>
<td>88</td>
<td>31</td>
</tr>
<tr>
<td>019</td>
<td>M</td>
<td>M</td>
<td>07/07/1999</td>
<td>58118</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>020</td>
<td>M</td>
<td>M</td>
<td>12/12/1999</td>
<td>60</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>021</td>
<td>F</td>
<td>F</td>
<td>09/09/1999</td>
<td>90040</td>
<td>200</td>
<td>51</td>
</tr>
<tr>
<td>022</td>
<td>M</td>
<td>M</td>
<td>10/10/1999</td>
<td>48144</td>
<td>82</td>
<td>21</td>
</tr>
<tr>
<td>023</td>
<td>F</td>
<td>F</td>
<td>31/01/1998</td>
<td>23</td>
<td>34</td>
<td>75</td>
</tr>
<tr>
<td>024</td>
<td>F</td>
<td>F</td>
<td>09/09/1998</td>
<td>120</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>025</td>
<td>M</td>
<td>M</td>
<td>01/01/1999</td>
<td>74102</td>
<td>68</td>
<td>51</td>
</tr>
<tr>
<td>026</td>
<td>M</td>
<td>M</td>
<td>NA/10/1999</td>
<td>66106</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>027</td>
<td>F</td>
<td>F</td>
<td>28/03/1998</td>
<td>66150</td>
<td>90</td>
<td>30</td>
</tr>
<tr>
<td>028</td>
<td>M</td>
<td>M</td>
<td>15/05/1998</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>029</td>
<td>F</td>
<td>F</td>
<td>07/07/1999</td>
<td>82148</td>
<td>84</td>
<td>10</td>
</tr>
</tbody>
</table>
### Introduction to SAS

A SAS code snippet is shown, which appears to be reading data from a file named `patients.txt`. The data contains columns for Visit Date and another column, possibly for another metric such as age or another measure associated with the visit.

<table>
<thead>
<tr>
<th>Visit Date</th>
<th>Metric A</th>
<th>Metric B</th>
<th>Metric C</th>
</tr>
</thead>
<tbody>
<tr>
<td>001M1 1/11/1998</td>
<td>88140</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>002F11 1/13/1998</td>
<td>84120</td>
<td>78</td>
<td>X0</td>
</tr>
<tr>
<td>003X10 21/1998</td>
<td>68190</td>
<td>100</td>
<td>31</td>
</tr>
<tr>
<td>004F01 01/01/1998</td>
<td>10120</td>
<td>01</td>
<td>20</td>
</tr>
<tr>
<td>005M5 07/07/1998</td>
<td>68120</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>006 06/15/1998</td>
<td>72102</td>
<td>68</td>
<td>61</td>
</tr>
<tr>
<td>007M10 32/1998</td>
<td>88148</td>
<td>102</td>
<td>0</td>
</tr>
<tr>
<td>011M11 11/1998</td>
<td>90190</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>008F08 08/1998</td>
<td>210</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>009M09 25/1998</td>
<td>86240</td>
<td>180</td>
<td>41</td>
</tr>
<tr>
<td>010F10 19/1999</td>
<td>40120</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>011M13 13/1999</td>
<td>68200</td>
<td>20</td>
<td>41</td>
</tr>
<tr>
<td>012M10 12/98</td>
<td>60122</td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td>013208 23/1999</td>
<td>74108</td>
<td>64</td>
<td>1</td>
</tr>
<tr>
<td>014M12 02/1999</td>
<td>22130</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>015F11 13/1998</td>
<td>84120</td>
<td>78</td>
<td>X0</td>
</tr>
<tr>
<td>015M11 12/1999</td>
<td>58112</td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td>017F04 05/1999</td>
<td>82148</td>
<td>88</td>
<td>31</td>
</tr>
<tr>
<td>019M01 07/1999</td>
<td>58118</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>020F15 12/1999</td>
<td>60</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>321F109/99/1999</td>
<td>90040</td>
<td>02000</td>
<td>51</td>
</tr>
<tr>
<td>022F09/10/1999</td>
<td>48114</td>
<td>82</td>
<td>21</td>
</tr>
<tr>
<td>023F12/31/1998</td>
<td>2234</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>024F11/09/1999</td>
<td>76120</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>025M01/01/1999</td>
<td>74102</td>
<td>68</td>
<td>51</td>
</tr>
<tr>
<td>026F12/01/1999</td>
<td>NA</td>
<td>166106</td>
<td>70</td>
</tr>
<tr>
<td>028F03/28/1998</td>
<td>66150</td>
<td>90</td>
<td>30</td>
</tr>
<tr>
<td>029M05/15/1998</td>
<td></td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>006F107/07/1999</td>
<td>82148</td>
<td>84</td>
<td>10</td>
</tr>
</tbody>
</table>
Introduction to SAS

SAS Code to covert flat file into SAS Dataset

```sas
DATA patients;
  infile "C:\Documents and Settings\ardianj\My Documents\SampleFiles\patients.txt" truncover;

  input @1 Patno $3.
  @4 Gender $1.
  @5 visit mmddyy10.
  @13 HR 3.
  @18 SBP 3.
  @21 DBP 3.
  @24 Dx $3.
  @27 AE $1.;

  label Patno = "Patient Number"
  Gender = "Gender"
  Visit = "Visit Date"
  HR = "Heart Rate"
  SBP = "Systolic Blood Pressure"
  DBP = "Diastolic Blood Pressure"
  Dx = "Diagnosis Code"
  AE = "Adverse Event?";

  format visit mmddyy10.;

RUN;
```
Introduction to SAS

SAS Code to covert flat file into SAS Dataset

```sas
DATA patients;
    infile "C:\Documents and Settings\ardian\My Documents\SampleFiles\patients.txt" truncover;
    input @1 Patno $3.
    @4 Gender $1.
    @5 visit mmddyy10.
    @13 HR 3.
    @18 SBP 3.
    @21 DBP 3.
    @24 Dx $3.
    @27 AE $1.;

    label Patno = "Patient Number"
    Gender = "Gender"
    Visit = "Visit Date"
    HR = "Heart Rate"
    SBP = "Systolic Blood Pressure"
    DBP = "Diastolic Blood Pressure"
    Dx = "Diagnosis Code"
    AE = "Adverse Event??";

    format visit mmddyy10.;
RUN;
```
Introduction to SAS

SAS Code to covert flat file into SAS Dataset

```sas
DATA patients;
   infile "C:\Documents and Settings\ardianj\My Documents\SampleFiles\patients.txt" truncover;
   input @1 Patno $3.
       @4 Gender $1.
       @5 visit mm/dd/yyyy.
       @13 HR 3.
       @18 SBP 3.
       @21 DBP 3.
       @24 Dx $3.
       @27 AE $1.;
   label Patno  = "Patient Number"
      Gender = "Gender"
      Visit  = "Visit Date"
      HR     = "Heart Rate"
      SBP    = "Systolic Blood Pressure"
      DBP    = "Diastolic Blood Pressure"
      Dx     = "Diagnosis Code"
      AE     = "Adverse Event?";
   format visit mm/dd/yyyy.
RUN;
```
Introduction to SAS

SAS Code to covert flat file into SAS Dataset

```sas
DATA patients;
  infile "C:\Documents and Settings\ardianj\My Documents\SampleFiles\patients.txt" truncover;
  input @1 Patno $3.
  @4 Gender $1.
  @5 visit rmddyy10.
  @13 HR 3.
  @18 SBP 3.
  @21 DBP 3.
  @24 Dx  $3.
  @27 AE $1.;
  label Patno  = "Patient Number"
               Gender  = "Gender"
               Visit   = "Visit Date"
               HR      = "Heart Rate"
               SBP     = "Systolic Blood Pressure"
               DBP     = "Diastolic Blood Pressure"
               Dx      = "Diagnosis Code"
               AE      = "Adverse Event?";
  format visit rmddyy10.;
RUN;
```
## Introduction to SAS

### Conversions Result

<table>
<thead>
<tr>
<th>Patno</th>
<th>Gender</th>
<th>visit</th>
<th>Hr</th>
<th>SBP</th>
<th>DBP</th>
<th>Dx</th>
<th>NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>M</td>
<td>11/11/1998</td>
<td>88</td>
<td>140</td>
<td>80</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>002</td>
<td>F</td>
<td>11/13/1998</td>
<td>84</td>
<td>120</td>
<td>78</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>003</td>
<td>X</td>
<td>10/21/1998</td>
<td>68</td>
<td>190</td>
<td>100</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>004</td>
<td>F</td>
<td>01/01/1999</td>
<td>101</td>
<td>200</td>
<td>120</td>
<td>5</td>
<td>A</td>
</tr>
<tr>
<td>XX5</td>
<td>M</td>
<td>05/07/1998</td>
<td>68</td>
<td>120</td>
<td>80</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>006</td>
<td>M</td>
<td>06/15/1999</td>
<td>72</td>
<td>102</td>
<td>68</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>007</td>
<td>M</td>
<td>11/11/1998</td>
<td>88</td>
<td>148</td>
<td>102</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>008</td>
<td>F</td>
<td>08/08/1998</td>
<td>210</td>
<td></td>
<td></td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>009</td>
<td>M</td>
<td>03/25/1999</td>
<td>86</td>
<td>240</td>
<td>180</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>010</td>
<td>F</td>
<td>10/19/1999</td>
<td>40</td>
<td>120</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>011</td>
<td>M</td>
<td>11/11/1998</td>
<td>68</td>
<td>300</td>
<td>20</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>012</td>
<td>M</td>
<td>10/12/1998</td>
<td>60</td>
<td>122</td>
<td>74</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>013</td>
<td>F</td>
<td>08/23/1999</td>
<td>74</td>
<td>108</td>
<td>64</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>014</td>
<td>M</td>
<td>02/02/1999</td>
<td>68</td>
<td>130</td>
<td>90</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>002</td>
<td>F</td>
<td>11/13/1998</td>
<td>84</td>
<td>120</td>
<td>78</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>003</td>
<td>M</td>
<td>11/12/1999</td>
<td>58</td>
<td>112</td>
<td>74</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>015</td>
<td>F</td>
<td>08/18/1999</td>
<td>82</td>
<td>148</td>
<td>88</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>017</td>
<td>F</td>
<td>04/05/1999</td>
<td>208</td>
<td>84</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>019</td>
<td>M</td>
<td>06/07/1999</td>
<td>58</td>
<td>118</td>
<td>70</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>123</td>
<td>M</td>
<td>10/10/1999</td>
<td>50</td>
<td>60</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>321</td>
<td>F</td>
<td>09/09/1999</td>
<td>900</td>
<td>400</td>
<td>200</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>020</td>
<td>F</td>
<td>10/10/1999</td>
<td>10</td>
<td>20</td>
<td>8</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>022</td>
<td>M</td>
<td>10/10/1999</td>
<td>48</td>
<td>114</td>
<td>82</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>023</td>
<td>F</td>
<td>12/31/1998</td>
<td>22</td>
<td>34</td>
<td>78</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>024</td>
<td>F</td>
<td>11/09/1998</td>
<td>76</td>
<td>120</td>
<td>80</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>025</td>
<td>M</td>
<td>01/01/1999</td>
<td>74</td>
<td>102</td>
<td>68</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>027</td>
<td>F</td>
<td>01/01/1999</td>
<td>166</td>
<td>106</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>028</td>
<td>F</td>
<td>03/28/1999</td>
<td>66</td>
<td>150</td>
<td>90</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>029</td>
<td>M</td>
<td>05/15/1998</td>
<td>84</td>
<td>148</td>
<td>84</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Introduction to SAS

Using PROC FREQ to List All the Unique Values

```
proc freq data=patients;
   tables Gender Dx AE / nocum nopercnt;
run;
```

Title "Frequency Counts for Selected Character Variables";
**Introduction to SAS**

**PROC FREQ Output**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>12</td>
</tr>
<tr>
<td>M</td>
<td>14</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>f</td>
<td>2</td>
</tr>
</tbody>
</table>

Frequency Missing = 1

<table>
<thead>
<tr>
<th>Diagnosis Code</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>X</td>
<td>2</td>
</tr>
</tbody>
</table>

Frequency Missing = 8
Introduction to SAS

PROC PRINT to List Invalid Character Data

Title "Listing of invalid character values";
PROC PRINT data=patients;
  where Gender not in ('M' 'F' '') or
       notdigit (trim(Dx)) and not missing (Dx) or
       AE not in ('0' '1' ' ');
  id patno;
  var Gender Dx AE;
RUN;
**Introduction to SAS**

**PROC PRINT Output**

*Listing of invalid character values*

<table>
<thead>
<tr>
<th>Patno</th>
<th>Gender</th>
<th>Dx</th>
<th>AE</th>
</tr>
</thead>
<tbody>
<tr>
<td>002</td>
<td>F</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>003</td>
<td>X</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>004</td>
<td>F</td>
<td>5</td>
<td>A</td>
</tr>
<tr>
<td>010</td>
<td>f</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>013</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>002</td>
<td>F</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>023</td>
<td>f</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
Using WHERE Statement with PROC PRINT to list Out-of-Range Data

```
Title "Out-of-range value for numeric variable";
PROC PRINT data=patients;
   where (HR not between 40 and 100 and HR is not missing) or
         (SBP not between 80 and 200 and SBP is not missing) or
         (DBP not between 80 and 200 and DBP is not missing);
   id Patno;
   var HR SBP DBP;
RUN;
```
Introduction to SAS

Using WHERE Statement with PROC PRINT Output

Out-of-range value for numeric variable

15:38 Thursday,

<table>
<thead>
<tr>
<th>Patno</th>
<th>Hr</th>
<th>SBP</th>
<th>DBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>002</td>
<td>84</td>
<td>120</td>
<td>78</td>
</tr>
<tr>
<td>004</td>
<td>101</td>
<td>200</td>
<td>120</td>
</tr>
<tr>
<td>006</td>
<td>72</td>
<td>102</td>
<td>68</td>
</tr>
<tr>
<td>008</td>
<td>210</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>009</td>
<td>86</td>
<td>240</td>
<td>180</td>
</tr>
<tr>
<td>010</td>
<td>.</td>
<td>40</td>
<td>120</td>
</tr>
<tr>
<td>011</td>
<td>68</td>
<td>300</td>
<td>20</td>
</tr>
<tr>
<td>012</td>
<td>60</td>
<td>122</td>
<td>74</td>
</tr>
<tr>
<td>013</td>
<td>74</td>
<td>108</td>
<td>64</td>
</tr>
<tr>
<td>014</td>
<td>22</td>
<td>130</td>
<td>90</td>
</tr>
<tr>
<td>002</td>
<td>84</td>
<td>120</td>
<td>78</td>
</tr>
<tr>
<td>003</td>
<td>58</td>
<td>112</td>
<td>74</td>
</tr>
<tr>
<td>017</td>
<td>208</td>
<td>.</td>
<td>84</td>
</tr>
<tr>
<td>019</td>
<td>58</td>
<td>118</td>
<td>70</td>
</tr>
<tr>
<td>321</td>
<td>900</td>
<td>400</td>
<td>200</td>
</tr>
<tr>
<td>020</td>
<td>10</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>023</td>
<td>22</td>
<td>34</td>
<td>78</td>
</tr>
<tr>
<td>025</td>
<td>74</td>
<td>102</td>
<td>68</td>
</tr>
</tbody>
</table>
Introduction to SAS

PROC FORMAT to Check for Invalid Value

```
PROC FORMAT;
 value $gender 'F' , 'M' = 'Valid'
         ' ' = 'Missing'
         other = 'Miscoded';

 value $ae '0', '1' = 'Valid'
         ' ' = 'Missing'
         other = 'Miscoded';
RUN;
```

```
Title "Using Custom Formatting to Identify Invalid Values";

PROC FREQ data=patients;
   format Gender $gender.
   AE $ae.;
   tables Gender AE /nocum nopercent missing;
RUN;
```
PROC FORMAT to Check for Invalid Value

PROC FORMAT;
  value $gender 'F', 'M' = 'Valid'
     ' '    = 'Missing'
     other  = 'Miscoded';

  value $ae '0', '1' = 'Valid'
     ' '    = 'Missing'
     other  = 'Miscoded';
RUN;

Title "Using Custom Formatting to Identify Invalid Values";

PROC FREQ data=patients;
  format Gender $gender.
  AE     $ae.;
  tables Gender AE /nocum nopercent missing;
RUN;
Introduction to SAS

PROC Format Output

Using Custom Formatting to Identify Invalid Values

15:38 Thurs

The FREQ Procedure

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing</td>
<td>1</td>
</tr>
<tr>
<td>Miscoded</td>
<td>4</td>
</tr>
<tr>
<td>Valid</td>
<td>26</td>
</tr>
</tbody>
</table>

Adverse Event?

<table>
<thead>
<tr>
<th>AE</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing</td>
<td>1</td>
</tr>
<tr>
<td>Valid</td>
<td>29</td>
</tr>
<tr>
<td>Miscoded</td>
<td>1</td>
</tr>
</tbody>
</table>
Introduction to SAS

Using a User-Defined Format with INPUT function

```sas
PROC FORMAT:
  invalue $gender 'F', 'M' = 'Valid'
         other     = 'Error';

  invalue $ae '0', '1'    = 'Valid'
         other     = 'Error';
RUN;

Title "Using a User-Defined Format with INPUT Function";

DATA _NULL_
  file print;
  set patients;

  if input(Gender, $gender.) = 'Error' then
     put @1  " Error for gender for patient: " Patno "values is " Gender;

  if input(AE, $ae.) = 'Error' then
     put @1  " Error for AE for patient: " Patno "values is " AE;

RUN;
```
Using a User-Defined Format with INPUT function

PROC FORMAT:
  inval $gender 'F', 'M' = 'Valid'
     other = 'Error';

inval $ae '0', '1' = 'Valid'
     other = 'Error';

RUN;

Title "Using a User-Defined Format with INPUT Function";

DATA _NULL_
  file print;
  set patients;

  if input(Gender, $gender.) = 'Error' then
    put @1 " Error for gender for patient: " Patno "values is " Gender;

  if input(AE, $ae.) = 'Error' then
    put @1 " Error for AE for patient: " Patno "values is " AE;

RUN;
Introduction to SAS

Using a User-Defined Format Output

Using a User-Defined Format with INPUT Function

Error for gender for patient: 003 values is X
Error for AE for patient: 004 values is A
Error for gender for patient: 006 values is
Error for gender for patient: 010 values is f
Error for gender for patient: 013 values is 2
Error for AE for patient: 013 values is
Error for gender for patient: 023 values is f
Introduction to SAS

Using PROC MEANS to Detect Invalid and Missing

```
Title "Checking numeric variables in the patients dataset";
PROC MEANS data=patients n nmiss min max maxdec=3;
   var HR SBP DBP;
RUN;
```
Introduction to SAS

Using PROC MEANS Output

Checking numeric variables in the patients dataset
15:38 Thursday, October 1,

The MEANS Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
<th>N</th>
<th>Miss</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hr</td>
<td>Heart Rate</td>
<td>28</td>
<td>3</td>
<td>10.000</td>
<td>900.000</td>
</tr>
<tr>
<td>SBP</td>
<td>Systolic Blood Pressure</td>
<td>27</td>
<td>4</td>
<td>20.000</td>
<td>400.000</td>
</tr>
<tr>
<td>DBP</td>
<td>Diastolic Blood Pressure</td>
<td>28</td>
<td>3</td>
<td>8.000</td>
<td>200.000</td>
</tr>
</tbody>
</table>
Introduction to SAS

Using PROC UNIVARIATE to Look for Outliers

```sas
Title "Using PROC UNIVARIATE to look for outliers";
PROC univariate data=patients plot;
   id Patno;
   var HR SBP DBP;
RUN;
```
Introduction to SAS

Using PROC UNIVARIATE Output (partial)

Using PROC UNIVARIATE to look for outliers

The UNIVARIATE Procedure
Variable: Hr (Heart Rate)

Moments

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>28</td>
</tr>
<tr>
<td>Mean</td>
<td>107.392857</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>161.066436</td>
</tr>
<tr>
<td>Skewness</td>
<td>4.73965876</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>1023549</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>149.997347</td>
</tr>
<tr>
<td>Sum Weights</td>
<td>28</td>
</tr>
<tr>
<td>Sum Observations</td>
<td>3007</td>
</tr>
<tr>
<td>Variance</td>
<td>25948.8399</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>23.7861582</td>
</tr>
<tr>
<td>Corrected SS</td>
<td>700618.679</td>
</tr>
<tr>
<td>Std Error Mean</td>
<td>30.442475</td>
</tr>
</tbody>
</table>

Basic Statistical Measures

<table>
<thead>
<tr>
<th>Location</th>
<th>Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Std Deviation</td>
</tr>
<tr>
<td>Median</td>
<td>Variance</td>
</tr>
<tr>
<td>Mode</td>
<td>Range</td>
</tr>
<tr>
<td></td>
<td>Interquartile Range</td>
</tr>
<tr>
<td>107.3929</td>
<td>161.06644</td>
</tr>
<tr>
<td>74.0000</td>
<td>25349</td>
</tr>
<tr>
<td>68.0000</td>
<td>990.00000</td>
</tr>
<tr>
<td></td>
<td>27.0000000</td>
</tr>
</tbody>
</table>
Introduction to SAS

Reading Date in Data

```
DATA dates;
   INPUT name $ 1-4 $6 bday mmddyy11.;
RUN;
John 1 1 1960
Mary 07/11/1955
Joan 07-11-1955
Kate 12.12.1962
Mark 06081959
RUN;
```

```
PROC PRINT DATA=dates;
   FORMAT bday date9.;
RUN;
```

```
Obs  name     bday
1    John     01JAN1960
2    Mary     11JUL1955
3    Joan     11JUL1955
4    Kate     12NOV1962
5    Mark     08JUN1959
```
Introduction to SAS

Reading Date in Data

DATA dates;
  INPUT month 1-2 year 4-7 day 9-10;
  bday=MDY(month,day,year);
  Card;
  7 1948 11
  1 1960 1
  10 1970 15
  12 1971 10
; RUN;

PROC PRINT DATA=dates;
  FORMAT bday date9.;
RUN;

The SAS System 15:06 Wednesday, September 30,

<table>
<thead>
<tr>
<th>Obs</th>
<th>month</th>
<th>year</th>
<th>day</th>
<th>bday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>1948</td>
<td>11</td>
<td>11JUL1948</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1960</td>
<td>1</td>
<td>01JAN1960</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>1970</td>
<td>15</td>
<td>15OCT1970</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>1971</td>
<td>10</td>
<td>10DEC1971</td>
</tr>
</tbody>
</table>
Reading Date in Data

```sas
OPTIONS YEARCUTOFF=1900; /* sets the cutoff back to the default */

DATA dates;
  INPUT name $1-4 5. bday mddwdr11.;
  age2000 = (MDY(1,1,2000)-bday)/365.25;
  DATES111;
  John 1 1 1960
  Mary 07/11/1955
  Joan 07-11-1955
  Kate 11.12.1962
  Mark 06081959
; ran;

PROC PRINT DATA=dates;
  FORMAT bday dates9.
RUN;
```

Output - (Untitled)

```
Obs name     bday     age2000
--- ----- ----- -------
  1 John 01Jan1960 40.0000
  2 Mary 11JUL1955 44.4784
  3 Joan 11JUL1955 44.4784
  4 Kate 12Nov1962 37.1362
  5 Mark 06Jun1959 40.5667
```
Introduction to SAS

Using SAS function for Making Variable

```sas
/* DATA getdata;
   INPUT name $14. x test1 test2 test3;
   DATA Tfunc1;
   SET getdata;
   tiint = INT(test1); t2int = INT(test2);
   t1rnd = ROUND(test1); t2fr = floor(test2);
   tave = MEAN(test1, test2, test3);
   RUN;
*/
```

The SAS System

<table>
<thead>
<tr>
<th>Obs</th>
<th>test1</th>
<th>test2</th>
<th>test3</th>
<th>tiint</th>
<th>t2int</th>
<th>t1rnd</th>
<th>t2fr</th>
<th>tave</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86.5</td>
<td>84.35</td>
<td>81</td>
<td>86</td>
<td>84</td>
<td>87</td>
<td>84</td>
<td>84.0167</td>
</tr>
<tr>
<td>2</td>
<td>70.3</td>
<td>82.37</td>
<td>70</td>
<td>70</td>
<td>82</td>
<td>70</td>
<td>82</td>
<td>76.3350</td>
</tr>
<tr>
<td>3</td>
<td>92.1</td>
<td>84.81</td>
<td>97</td>
<td>92</td>
<td>92</td>
<td>92</td>
<td>92</td>
<td>94.5267</td>
</tr>
<tr>
<td>4</td>
<td>94.2</td>
<td>92.64</td>
<td>93</td>
<td>94</td>
<td>92</td>
<td>94</td>
<td>92</td>
<td>93.2800</td>
</tr>
<tr>
<td>5</td>
<td>79.7</td>
<td>79.07</td>
<td>72</td>
<td>79</td>
<td>79</td>
<td>80</td>
<td>79</td>
<td>76.9233</td>
</tr>
</tbody>
</table>
```
Convert SAS Dataset

From SAS to SPSS
PROC EXPORT data=mydata outfile= "C:\data\newdata3.sav";
run;

From SAS to STATA
PROC EXPORT data=mydata outfile= "C:\data\newdata3.dta";
run;
Introduction to PROC SQL

Why Learn PROC SQL

• Can retrieve information without having to learn SAS syntax
• Normally shorter and fewer statement that traditional SAS Code
• SQL often uses fewer computing resources that conventional DATA and PROC steps
• The knowledge learned is transferable to other SQL proprietary languages such as ANSI SQL, T-SQL, PL/SQL
Connecting to Oracle Database

- libname rftsv_conn
  oracle user=vanderbilt
  password=vanderbilt
  path='PTPROD.OAS.VANDERBILT'
  schema=rfts_prod_schema
  preserve_col_names=yes
  access=readonly;
A Simple PROC SQL

PROC SQL:
SELECT count(*)
FROM CDS.SCREENING ;
QUIT;

The SAS System 10:12 Tuesday, September 29,
Introduction to PROC SQL

Creating New Variables

- PROC SQL:
  ```sql
  SELECT name, substr(phone,1,3) LABEL='AREA CODE'
  FROM participants ;
  QUIT;
  ```

  ```
  1425  proc sql;
  1426  select name , substr(phone,1,3) label='area code'
  1427  from participants;
  1428  quit;
  NOTE: PROCEDURE SQL used (Total process time):
    real time       0.10 seconds
    cpu time        0.09 seconds
  ```

<table>
<thead>
<tr>
<th>name</th>
<th>area code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jardia</td>
<td>615</td>
</tr>
<tr>
<td>Brainey</td>
<td>479</td>
</tr>
<tr>
<td>Jacarr</td>
<td>404</td>
</tr>
</tbody>
</table>
Introduction to PROC SQL

The CALCULATED OPTION

- PROC SQL;
  SELECT name, (income * .05) as tax,
         (CALCULATED tax * .01) as rebate
  FROM participants;
  QUIT;

1443  proc sql;
1444  select name , (income *.05) as tax,
1445    calculated tax *.01 as rebate
1446  from participants;
1447  quit;
 NOTE: PROCEDURE SQL used (Total process time):
      real time        0.00 seconds
      cpu time         0.00 seconds

The SAS System  10:12 Tuesday,

<table>
<thead>
<tr>
<th>name</th>
<th>tax</th>
<th>rebate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jardia</td>
<td>5012.8</td>
<td>50.128</td>
</tr>
<tr>
<td>Braine</td>
<td>4500</td>
<td>45</td>
</tr>
<tr>
<td>jacarr</td>
<td>2600</td>
<td>26</td>
</tr>
</tbody>
</table>
Introduction to PROC SQL

Using CASE Logic inside SELECT Statement

PROC SQL:
SELECT name, 
CASE
  WHEN income < 60000 THEN 'LOW'
  WHEN income <= 90000 THEN 'AVG'
  WHEN income > 60000 THEN 'HIGH'
  ELSE 'VERY HIGH'
END as income_category
FROM participants;
QUIT;

<table>
<thead>
<tr>
<th>name</th>
<th>income_category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarda</td>
<td>HIGH</td>
</tr>
<tr>
<td>Brainy</td>
<td>AVG</td>
</tr>
<tr>
<td>jacarr</td>
<td>LOW</td>
</tr>
</tbody>
</table>
Introduction to PROC SQL

Creating New Table based on SELECT statement

- PROC SQL:
  CREATE TABLE tbl_income as
  SELECT name, CASE
  WHEN income <60000 THEN 'LOW'
  WHEN income <=90000 THEN 'AVG'
  WHEN income >60000 THEN 'HIGH'
  ELSE 'VERY HIGH'
  END as income_category
  FROM participants;
  QUIT;

NOTE: Table WORK.TBL_INCOME created, with 3 rows and 2 columns.

NOTE: PROCEDURE SQL used (Total process time):
  real time 0.01 seconds
cpu time 0.01 seconds
Using PROC DBLOAD

Upload New Table into Oracle Database

- PROC DBLOAD
  DBMS=oracle
  DATA=tbl_income append;
  USER=vanderbilt;
  PASSWORD=vanderbilt;
  PATH='ptprod.oas.vanderbilt';
  TABLE=tbl_participant_income;
  COMMIT=10000;
  ERRLIMIT=5;
  LOAD;
  RUN;
Combining Tables Horizontally using PROC SQL

**Inner Join**

<table>
<thead>
<tr>
<th>TBL_ONE</th>
<th>TBL_TWO</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
</tr>
</tbody>
</table>

**Outer Join**

- **Left Outer Join**
- **Right Outer Join**
- **Full Outer Join**
Combining Tables Horizontally using PROC SQL

**Inner Join**

<table>
<thead>
<tr>
<th>TBL_ONE</th>
<th>TBL_TWO</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
</tr>
<tr>
<td>3</td>
<td>y</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
</tr>
<tr>
<td>5</td>
<td>v</td>
</tr>
</tbody>
</table>

PROC SQL;
SELECT * FROM tbl_one, tbl_two
WHERE tbl_one.x = tbl_two.x;
QUIT;
Combining Tables Horizontally using PROC SQL

Inner Join

<table>
<thead>
<tr>
<th>TBL_ONE</th>
<th>TBL_TWO</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a_x</td>
</tr>
<tr>
<td>2</td>
<td>b_x</td>
</tr>
</tbody>
</table>

PROC SQL;
SELECT * FROM tbl_one, tbl_two
WHERE tbl_one.x = tbl_two.x;
QUIT;

The SAS System

<table>
<thead>
<tr>
<th>x</th>
<th>a</th>
<th>x</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>b</td>
<td>2</td>
<td>x</td>
</tr>
</tbody>
</table>
Combining Tables Horizontally using PROC SQL

**Left Outer Join**

```
PROC SQL;
SELECT * FROM tbl_one
LEFT JOIN tbl_two
on tbl_one.x = tbl_two.x;
QUIT;
```
Combining Tables Horizontally using PROC SQL

**Left Outer Join**

```
PROC SQL;
SELECT * FROM tbl_one
LEFT JOIN tbl_two on tbl_one.x = tbl_two.x;
QUIT;
```

<table>
<thead>
<tr>
<th>TBL_ONE</th>
<th>TBL_TWO</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
</tr>
<tr>
<td>3</td>
<td>y</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```
The SAS System
```
```
<table>
<thead>
<tr>
<th>x</th>
<th>a</th>
<th>x</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>b</td>
<td>2</td>
<td>x</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

A B

TBL_ONE

TBL_TWO
Combining Tables Horizontally using PROC SQL

Left Outer Join

```
PROC SQL;
SELECT * FROM tbl_one
LEFT JOIN tbl_two
on tbl_one.x = tbl_two.x;
QUIT;
```
Combining Tables Horizontally using PROC SQL

**Left Outer Join**

```
PROC SQL;
SELECT * FROM tbl_one
LEFT JOIN tbl_two
ON tbl_one.x = tbl_two.x;
QUIT;
```

<table>
<thead>
<tr>
<th>TBL_ONE</th>
<th>TBL_TWO</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
</tr>
<tr>
<td>3</td>
<td>y</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
</tr>
<tr>
<td>5</td>
<td>v</td>
</tr>
</tbody>
</table>

The SAS System

```
  x a x b
1 a
2 b 2 x
4 d
```
Combining Tables Horizontally using PROC SQL

**Left Outer Join**

PROC SQL;
SELECT * FROM tbl_one RIGHT JOIN tbl_two on tbl_one.x = tbl_two.x;
QUIT;
Combining Tables Horizontally using PROC SQL

**Left Outer Join**

```sql
PROC SQL;
SELECT * FROM tbl_one
RIGHT JOIN tbl_two
ON tbl_one.x = tbl_two.x;
QUIT;
```

<table>
<thead>
<tr>
<th>TBL_ONE</th>
<th>TBL_TWO</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
</tr>
<tr>
<td>3</td>
<td>y</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
</tr>
<tr>
<td>5</td>
<td>v</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x</th>
<th>a</th>
<th>x</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>b</td>
<td>2</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>v</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The SAS System
Combining Tables Horizontally using PROC SQL

**Full Outer Join**

```
PROC SQL;
SELECT * FROM tbl_one
FULL JOIN tbl_two
ON tbl_one.x = tbl_two.x;
QUIT;
```
Combining Tables Horizontally using PROC SQL

**Full Outer Join**

```
PROC SQL;
SELECT * FROM tbl_one FULL JOIN tbl_two ON tbl_one.x = tbl_two.x;
QUIT;
```

<table>
<thead>
<tr>
<th>TBL_ONE</th>
<th>TBL_TWO</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>1</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
</tr>
<tr>
<td>3</td>
<td>y</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
</tr>
<tr>
<td>5</td>
<td>v</td>
</tr>
</tbody>
</table>
Comparing PROC SQL and SAS

Left Outer Join

PROC SQL;
SELECT *
FROM tbl_one, tbl_two
where tbl_one.x = tbl_two.x
order by tbl_one.x;
QUIT;

PROC SORT DATA=tbl_one; BY x; RUN;

PROC SORT DATA=tbl_two; BY x; RUN;

DATA merged;
MERGE tbl_one tbl_two;
BY x;
RUN;

PROC SORT DATA=merged; BY x; RUN;