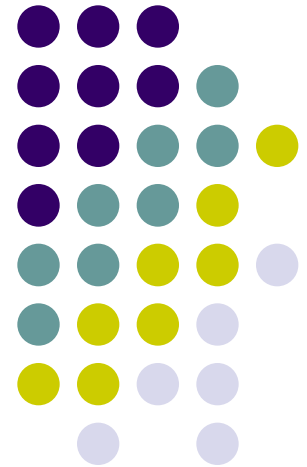


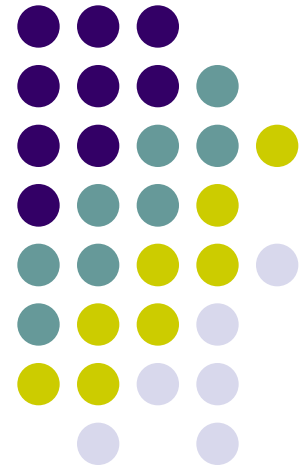
# Introduction to SAS Programming

Christina L. Ughrin  
Statistical Software Consulting  
Some notes pulled from SAS  
Programming I: Essentials Training



# SAS Datasets

Examining the structure of SAS  
Datasets



# SAS Data Sets

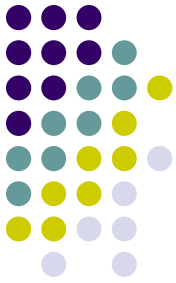


## Two Sections

Descriptor Section

Data Section

# Data Set Descriptor Section



Output - (Untitled) The SAS System 16:29 Monday, June 16, 2008 3

The CONTENTS Procedure

Data Set Name	WORK.TUTOR	Observations	30
Member Type	DATA	Variables	7
Engine	V9	Indexes	0
Created	Monday, June 16, 2008 04:31:22 PM	Observation Length	72
Last Modified	Monday, June 16, 2008 04:31:22 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_32		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information

Data Set Page Size	8192
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	113
Obs in First Data Page	30
Number of Data Set Repairs	0
File Name	C:\DOCUME~1\cughrin\LOCALS~1\Temp\SAS Temporary Files\_TD2980\tutor.sas7bdat
Release Created	9.0101M3
Host Created	XP_PRO

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
7	Date	Num	8	MMDYY10.	MMDYY10.
6	Degree	Char	1	\$1.	\$1.
2	Department	Char	36	\$36.	\$36.
1	ID	Num	8	BEST12.	BEST32.
3	Satisfaction	Num	8	BEST12.	BEST32.
5	Status	Char	3	\$3.	\$3.
4	Years	Num	8	BEST12.	BEST32.

# SAS Data Section



Output - (Untitled) The SAS System 7

Faculty ID	Department	Satisfaction with Department	Years Employed at University	Status	Degree	Date
1	Anthropology	5	20	T	D	02/12/2007
2	Family and Consumer Studies	4	15	NT	M	03/18/2007
3	Communication Studies	2	5	NT	M	05/05/2008
4	Speech Pathology and Audiology	5	5	PT	M	11/24/2007
5	Nursing	2	22	T	D	09/06/2007
6	English	5	27	T	D	04/25/2007
7	Nursing	3	13	NT	M	10/16/2007
8	Economics	2	8	NT	M	01/08/2008
9	History	4	5	PT	B	01/09/2007
10	Finance	3	7	T	D	01/10/2008
11	Mathematical Studies	5	16	T	D	06/11/2008
12	Accounting	4	18	T	D	05/18/2008
13	Psychology	2	9	T	D	02/13/2008
14	Economics	5	22	T	D	01/14/2008
15	Psychology	1	20	T	D	01/15/2007
16	Finance	2	5	TT	D	01/16/2008
17	Accounting	4	3	TT	D	01/01/2007
18	Biological Sciences	3	6	NT	M	01/31/2008
19	Psychology	5	24	NT	D	02/19/2008
20	Computer Science	4	16	T	D	02/20/2008
21	Philosophy	4	19	T	D	03/24/2008
22	History	5	6	PT	D	06/28/2007
23	Sociology	2	4	TT	D	01/22/2007
24	Physics	1	3	TT	D	01/24/2008
25	Sociology	1	5	TT	D	07/05/2007
26	Chemistry	5	10	NT	M	08/26/2007
27	Justice Studies	1	13	T	D	08/17/2007
28	Physics	5	4	PT	B	07/02/2007
29	Special Education	3	11	T	D	01/29/2008
30	Communication Studies	4	14	T	D	09/09/2007

N = 30



# Attributes of Variables

- Name
  - e.g. Status
- Type
  - Numeric or Character
  - e.g. Status in this example is character (T, TT, PT, or NTT) and Satisfaction is numeric (1 to 5).

# SAS Data Set Terminology



- *Variables* – columns in a SAS data set.
- *Observations* – rows in a SAS data set.
- *Numeric Data* – values that are treated as numeric and may include 8 bytes of floating storage for 16 to 17 significant digits.
- *Character Data* – non numeric data values such as letters, numbers, special characters, and blanks. May be stores with a length of 1 to 32, 767 bytes. One byte is equal to one character.

# SAS Data Set and Variable Name Criteria



- Can be 32 characters long.
- Can be uppercase, lowercase, or a mixture of the cases.
- Are not case sensitive
- Cannot start with number and cannot contain special characters or blanks.
- Must start with a letter or underscore.



# SAS Dates



- Dates are treated as special kind of numeric data.
  - They are the number of days since January 1<sup>st</sup>, 1960. January 1<sup>st</sup> 1960 is the 0 point. SAS dates can go back to 1582 (Gregorian Calendar) and forward to the year 2000.
  - Dates are displayed using a format. There are a number of different date formats supported by SAS.
- Time is scored as the number of seconds since midnight. SAS date time is the number of seconds since January 1<sup>st</sup>, 1960.



# Missing Data in SAS

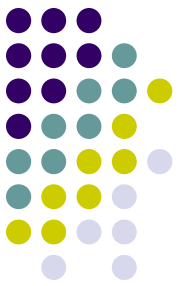
- Missing values are valid values.
  - For character data, missing values are displayed as blanks.
  - For numeric data, missing values are displayed as periods.

The SAS System 8

Faculty ID	Department	Satisfaction with Department	Years Employed at University	Status	Degree	Date
1	Anthropology	5	20		D	02/12/2007
2	Family and Consumer Studies	4	15	NT	M	03/18/2007
3	Communication Studies	2	5	NT	M	05/05/2008
4	Speech Pathology and Audiology	5	5	PT	M	11/24/2007
5	Nursing	2	22	T	D	09/06/2007
6	English	5	27	T	D	04/25/2007
7	Nursing	3	13	NT	M	10/16/2007
8	Economics	2	8	NT	M	01/08/2008
9	History	4	.	PT	B	01/09/2007
10	Finance	3	7	T	D	01/10/2008
11	Mathematical Studies	5	16	T	D	06/11/2008
12	Accounting	4	18	T	D	05/18/2008
13	Psychology	2	9	T	D	02/13/2008
14	Economics	5	22	T	D	01/14/2008

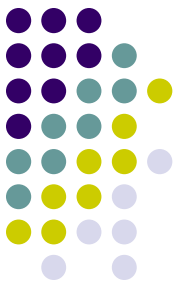


# SAS Syntax



# SAS Syntax

- Statements in SAS are like sentences. The punctuation though is a semicolon( ; )rather than a period ( . )
- Most Statements (but not all) start with an identifying key word (e.g. proc, data, label, options, format...)
- Statements are strung together into sections similar to paragraphs. These paragraphs in a Windows OS are ended with the word “run” and a semicolon.



# Example of SAS Syntax

```
proc print data=tutor NOOBS N label;
label ID= Faculty ID
      Department= Department
      Satisfaction= Satisfaction with Department
      Years= Years Employed at University
      Satus= Faculty Status
      Degree= Degree
      Date=Date;
options ls=100 nodate;
run;
|
```

# SAS Syntax Rules



- SAS statements are format free.
- One or more blanks or special characters are used to separate words.
- They can begin and end in any column.
- A single statement can span multiple lines.
- Several statements can be on the same line.

# Example of SAS Free Format



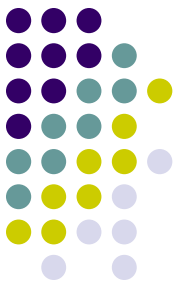
```
proc print data=tutor NOOBS N label;
label ID= Faculty ID Department= Department Satisfaction= Satisfaction with Department
Years= Years Employed at University Satus= Faculty Status Degree= Degree Date=Date;
options ls=100 nodate;
run;
|
```

Using the free-format Syntax

rules of SAS though can make it difficult for others (or you) to read your program. This is akin to

writing a page of text with little attention to line breaks. You may still have

Capital letters and periods, but where a sentence begins and ends may be a bit confusing

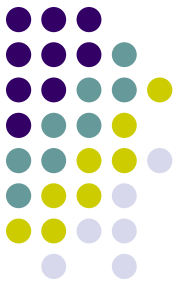


# Example of SAS Formatted

```
proc print data=tutor NOOBS N label;
label ID= Faculty ID
      Department= Department
      Satisfaction= Satisfaction with Department
      Years= Years Employed at University
      Satus= Faculty Status
      Degree= Degree
      Date=Date;
options ls=100 nodate;
run;
|
```

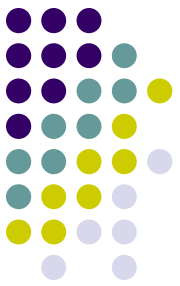
Using the free-format Syntax rules of SAS though can make it difficult for others (or you) to read your program. This is akin to writing a page of text with little attention to line breaks. You may still have capital letters and periods, but where a sentence begins and ends may be a bit confusing. Isn't this paragraph a bit easier to read?





# SAS Comments

- Type `/*` to begin a comment.
- Type your comment text.
- Type `*/` to end the comment.
- Or, type an `*` at the beginning of a line. Everything between the `*` and the `;` will be commented.
  - e.g. `*infile 'tutor.dat';`
- Alternatively, highlight the text that you would like to comment and use the keys `Ctrl /` to comment the line. To uncomment a line, highlight and use the `Ctrl Shift /` keys.



# SAS Comments

```
❏ proc print data=Tutormissing NOOBS N label;  
  label ID= Faculty ID  
        Department= Department  
        Satisfaction= Satisfaction with Department  
        Years= Years Employed at University  
        Satus= Faculty Status  
        Degree= Degree  
        Date=Date;  
/*options ls=100 nodate;*/  
run;
```



# SAS Windows

# SAS Windows



Explorer



Log



Editor

A screenshot of the SAS software interface. The Explorer window on the left shows 'Active Libraries' with icons for Gismaps, Maps, SasHELP, Sasuser, Train, and Work. The Log window at the top right displays the following text:

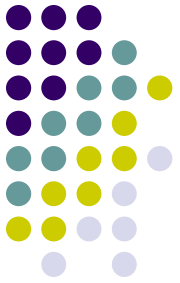
```
7 libname train '\\kent.edu\shares$\LMSData\Users\cughrin\My Documents\Software Issues\SAS\SAS  
7 | Tutorials';  
NOTE: Libref TRAIN was successfully assigned as follows:  
Engine: V9  
Physical Name: '\\kent.edu\shares$\LMSData\Users\cughrin\My Documents\Software Issues\SAS\SAS  
Tutorials'  
8 run;
```

The Editor window at the bottom right shows the following code:

```
libname train '\\kent.edu\shares$\LMSData\Users\cughrin\My Documents\Software Issues\SAS\SAS Tutorials';  
run;
```

The status bar at the bottom shows 'C:\Documents and Settings\cughrin'.

# Enhanced Editor Window



The screenshot displays the SAS Enhanced Editor interface. The top window, titled 'Output - (Untitled)', shows a table of data with the following columns: Department, Satisfaction with Department, Years at the University, Faculty Status, Degree, and Date. The bottom window, titled 'Editor - Untitled1 \*', contains the following SAS code:

```
proc print data=train.sastraining NOOBS label;
  label ID = Faculty ID
        Department= Department
        Satisfaction= Satisfaction with Department
        Years= Years at the University
        Status= Faculty Status
        Degree= Degree
        Date= Date;
  options ls=200 nodate;
run;
```

A red arrow points from the 'Enhanced Editor' label to the code editor window.

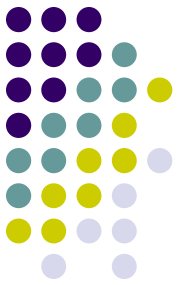
Department	Satisfaction with Department	Years at the University	Faculty Status	Degree	Date
Anthropology	5	20	T	D	02/12/2007
Family and Consumer Studies	4	15	NT	M	03/18/2007
Communication Studies	2	5	NT	M	05/05/2008
Speech Pathology and Audiology	5	5	PT	M	11/24/2007
Nursing	2	22	T	D	03/06/2007
English	5	27	T	D	04/25/2007
Nursing	3	13	NT	M	10/16/2007
Economics	2	8	NT	M	01/08/2008
History	4	5	PT	B	01/09/2007
Finance	3	7	T	D	01/10/2008
Mathematical Studies	5	16	T	D	06/11/2008
Accounting	4	18	T	D	05/18/2008
Psychology	2	9	T	D	02/13/2008
Economics	5	22	T	D	01/14/2008
Psychology	1	20	T	D	01/15/2007
Finance	2	5	TT	D	01/16/2008
Accounting	4	3	TT	D	01/01/2007
Biological Sciences	3	6	NT	M	01/31/2008
Psychology	5	24	NT	D	02/19/2008

Output

Enhanced Editor

- Your program script appears in this window.
- You can either bring it in from a file or type the program right into the window.
- Once the program is in the window, you can Click Submit (or the running guy).

# SAS Log



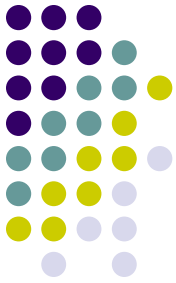
```
Log - (Untitled)
NOTE: Copyright (c) 2002-2008 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.2 (TS2M0)
      Licensed to KENT STATE UNIVERSITY FOR CAMPUS WIDE, Site 70005975.
NOTE: This session is executing on the XP_PRO platform.

NOTE: SAS initialization used:
      real time      1.57 seconds
      cpu time       0.85 seconds

1  libname COBA Demo 'E:\Trainings\JMP Training';
ERROR: The DEMO engine cannot be found.
ERROR: Error in the LIBNAME statement.
2  run'
```

- SAS Log provides a “blow by blow” account of the execution of your program. It includes how many observations were read and output, as well as, errors and notes.
- Note the errors in red.

# Output Window



The screenshot displays the SAS software interface. The main window is titled "Output - (Untitled)" and shows a table of data. Below the table is the "Editor - Untitled1\*" window containing SAS code. The table has the following columns: Faculty ID, Department, Satisfaction with Department, Years at the University, Faculty Status, Degree, and Date.

Faculty ID	Department	Satisfaction with Department	Years at the University	Faculty Status	Degree	Date
1	Anthropology	5	20	T	D	02/12/2007
2	Family and Consumer Studies	4	15	NT	M	03/18/2007
3	Communication Studies	2	5	NT	M	05/05/2008
4	Speech Pathology and Audiology	5	5	PT	M	11/24/2007
5	Nursing	2	22	T	D	09/06/2007
6	English	5	27	T	D	04/25/2007
7	Nursing	3	13	NT	M	10/16/2007
8	Economics	2	3	NT	M	01/05/2008
9	History	4	5	PT	B	01/09/2007
10	Finance	3	7	T	D	01/10/2008
11	Mathematical Studies	5	16	T	D	06/11/2008
12	Accounting	4	18	T	D	05/18/2008
13	Psychology	2	9	T	D	02/13/2008
14	Economics	5	22	T	D	01/14/2008
15	Psychology	1	20	T	D	01/15/2007
16	Finance	2	5	TT	D	01/15/2008
17	Accounting	4	3	TT	D	01/01/2007
18	Biological Sciences	3	6	NT	M	01/31/2008
19	Psychology	5	24	NT	D	02/19/2008

```
proc print data=train.sastraining NOOBS label;
  label ID = Faculty ID
        Department= Department
        Satisfaction= Satisfaction with Department
        Years= Years at the University
        Status= Faculty Status
        Degree= Degree
        Date= Date;
  options ls=200 nodate;
run;
```

# SAS Library

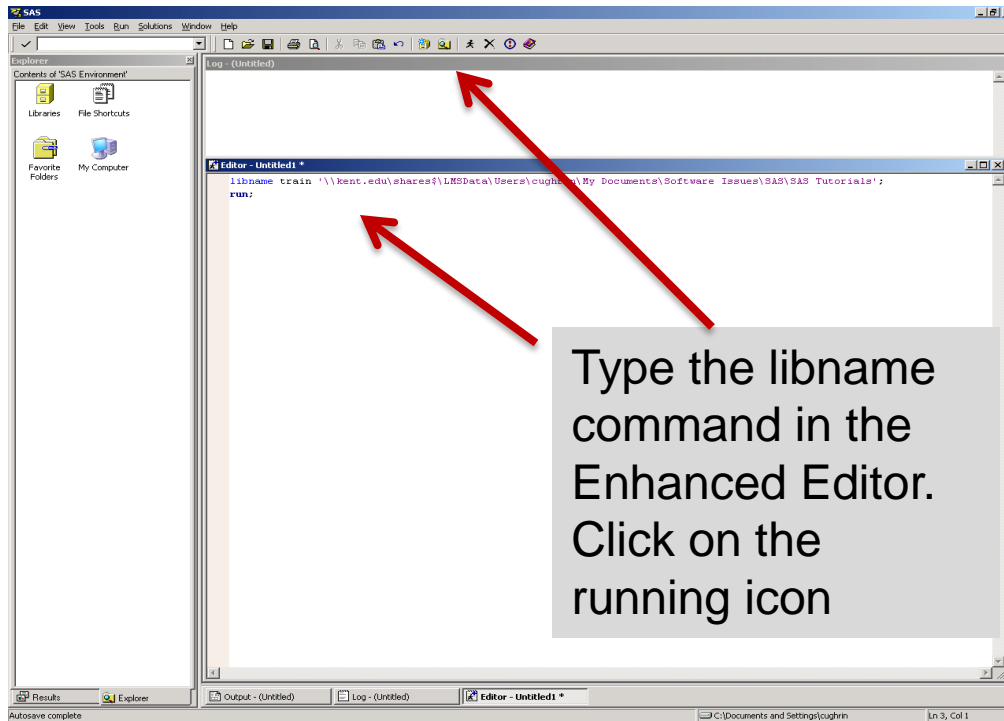


- SAS Data Libraries are like drawers in a filing cabinet. The SAS data sets are files within those drawers. Note the icons for the SAS library match that metaphor.
- In order to assign a “drawer”, you assign a library reference name (libref).
- There are two drawers already in your library: work (temporary) and sasuser (permanent).
- You can also create your own libraries (drawers) using the libname statement.



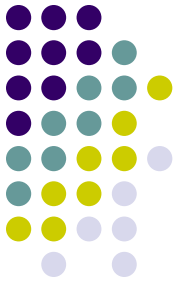


# Establishing the libname



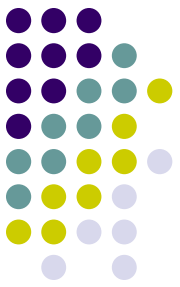
libname Tina 'E:\Trainings\JMP Training';  
run;

# Viewtable Window



The screenshot shows the SAS software interface. The main window is titled "VIEWTABLE: Written by SAS" and displays a data table with the following columns: ID, Department, Satisfaction, Years, Status, Degree, and Date. The table contains 30 rows of data. On the left, the Explorer window shows the "Contents of 'Train'" folder with sub-items "Sastraining..." and "Tutor". The bottom status bar shows the current directory as "C:\Documents and Settings\cughrin".

ID	Department	Satisfaction	Years	Status	Degree	Date
1	Anthropology	5	20	T	D	02/12/2007
2	Family and Consumer Studies	4	15	NT	M	03/18/2007
3	Communication Studies	2	5	NT	M	05/05/2008
4	Speech Pathology and Audiology	5	5	PT	M	11/24/2007
5	Nursing	2	22	T	D	09/06/2007
6	English	5	27	T	D	04/25/2007
7	Nursing	3	13	NT	M	10/16/2007
8	Economics	2	8	NT	M	01/08/2008
9	History	4	5	PT	B	01/03/2007
10	Finance	3	7	T	D	01/10/2008
11	Mathematical Studies	5	16	T	D	06/11/2008
12	Accounting	4	18	T	D	05/18/2008
13	Psychology	2	9	T	D	02/13/2008
14	Economics	5	22	T	D	01/14/2008
15	Psychology	1	20	T	D	01/15/2007
16	Finance	2	5	TT	D	01/16/2008
17	Accounting	4	3	TT	D	01/01/2007
18	Biological Sciences	3	6	NT	M	01/31/2008
19	Psychology	5	24	NT	D	02/19/2008
20	Computer Science	4	16	T	D	02/20/2008
21	Philosophy	4	19	T	D	03/24/2008
22	History	5	6	PT	D	06/28/2007
23	Sociology	2	4	TT	D	01/22/2007
24	Physics	1	3	TT	D	01/24/2008
25	Sociology	1	5	TT	D	07/05/2007
26	Chemistry	5	10	NT	M	08/26/2007
27	Justice Studies	1	13	T	D	08/17/2007
28	Physics	5	4	PT	B	07/02/2007
29	Special Education	3	11	T	D	01/29/2008
30	Communication Studies	4	14	T	D	09/09/2007



# Data Step Programming

- SAS data set can be created using another SAS data set as input or raw data
- To create a SAS data set using another SAS data set, the DATA and SET statements are used.
- To create a SAS data set from raw data, you use INFILE and INPUT statements.
- DATA and SET cannot be used for raw data and INFILE and INPUT cannot be used for existing SAS datasets.

# Reading a SAS Dataset



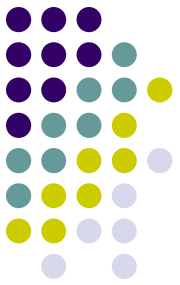
DATA (name of new SAS dataset)

    SET (name of existing SAS dataset)

    Additional statements

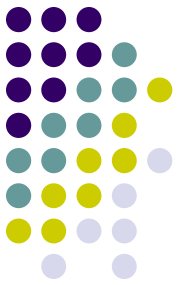
Run;

# Reading a SAS Dataset



```
Editor - Untitled1 *  
  
libname train 'H:\My Documents\Training Files -- Train';  
run;  
  
data train.newfile;  
    Set train.sastraining;  
    TotalYears= sum(years);  
run;
```

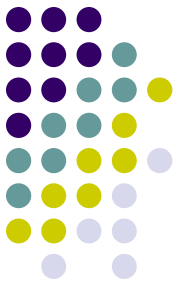
# Reading SAS Dataset



The screenshot displays the SAS software interface. The main window is titled "VIEWTABLE: Train.Newfile" and shows a table with the following columns: ID, Department, Satisfaction, Years, Status, Degree, Date, and TotalYears. The table contains 15 rows of data. On the left, the "Explorer" pane shows the contents of the "Train" folder, including files named "Dallasla", "Newfile", "Sastraining", and "Tutor".

ID	Department	Satisfaction	Years	Status	Degree	Date	TotalYears
1	Accounting	4	3	TT	D	01/01/2007	3
2	Accounting	4	18	T	D	05/18/2008	18
3	Anthropology	5	20	T	D	02/12/2007	20
4	Biological Sciences	3	6	NT	M	01/31/2008	6
5	Chemistry	5	10	NT	M	08/26/2007	10
6	Communication Studies	2	5	NT	M	05/05/2008	5
7	Communication Studies	4	14	T	D	09/09/2007	14
8	Computer Science	4	16	T	D	02/20/2008	16
9	Economics	2	8	NT	M	01/08/2008	8
10	Economics	5	22	T	D	01/14/2008	22
11	English	5	27	T	D	04/25/2007	27
12	Family and Consumer Studies	4	15	NT	M	03/18/2007	15
13	Finance	2	5	TT	D	01/16/2008	5
14	Finance	3	7	T	D	01/10/2008	7
15	History	4	5	NT	D	01/09/2007	5

# Reading Raw Data



The screenshot shows the SAS Editor window titled "SAS - [Editor - Untitled1 \*]". The menu bar includes File, Edit, View, Tools, Run, Solutions, Window, and Help. The toolbar contains icons for file operations and editing. The Explorer window on the left shows the contents of the 'Train' directory, including files named Dallasla, Newfile, Sastraining, and Tutor. The main editor window displays the following SAS code:

```
libname train 'H:\My Documents\Training Files -- Train';  
run;  
  
data train.DallasLA;  
  infile 'H:\My Documents\Training Files -- Train\DallasLA.txt';  
  input Flight $ 1-3  
         Date $ 4-11 Dest $ 12-14  
         FirstClass 15-17 Economy 18-20;  
run;
```

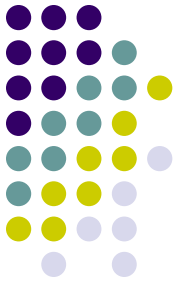
# Selecting Variables



- You can use a DROP or KEEP statement in a DATA step to control which variables are written to a new SAS data set.



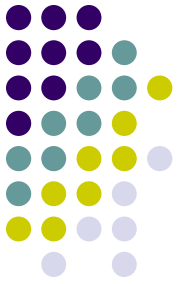
# Selecting Variables



```
libname train 'H:\My Documents\Training Files -- Train';  
run;  
data train.newfile;  
  Set train.sastraining;  
  TotalYears= sum(years);  
  Drop Years;  
run;
```

	ID	Department	Satisfaction	Status	Degree	Date	TotalYears
1	17	Accounting		4 TT	D	01/01/2007	3
2	12	Accounting		4 T	D	05/18/2008	18
3	1	Anthropology		5 T	D	02/12/2007	20
4	18	Biological Sciences		3 NT	M	01/31/2008	6
5	26	Chemistry		5 NT	M	08/26/2007	10
6	3	Communication Studies		2 NT	M	05/05/2008	5
7	30	Communication Studies		4 T	D	09/09/2007	16
8	20	Computer Science		4 T	D	02/20/2008	16
9	8	Economics		2 NT	M	01/08/2008	8
10	14	Economics		5 T	D	01/14/2008	22
11	6	English		5 T	D	04/25/2007	27
12	2	Family and Consumer Studies		4 NT	M	03/18/2007	15
13	16	Finance		2 TT	D	01/16/2008	5
14	10	Finance		3 T	D	01/10/2008	7
15	9	History		4 PT	B	01/09/2007	5
16	22	History		5 PT	D	06/28/2007	6
17	27	Justice Studies		1 T	D	08/17/2007	13
18	11	Mathematical Studies		5 T	D	06/11/2008	16
19	7	Nursing		3 NT	M	10/16/2007	13
20	5	Nursing		2 T	D	09/06/2007	22
21	21	Philosophy		4 T	D	03/24/2008	19
22	24	Physics		1 TT	D	01/24/2008	3
23	28	Physics		5 PT	B	07/02/2007	4
24	13	Psychology		2 T	D	02/13/2008	9
25	15	Psychology		1 T	D	01/15/2007	20
26	19	Psychology		5 NT	D	02/19/2008	24
27	23	Sociology		2 TT	D	01/22/2007	4
28	25	Sociology		1 TT	D	07/05/2007	5
29	29	Special Education		3 T	D	01/23/2008	11
30	4	Speech Pathology and Audiology		5 PT	M	11/24/2007	5

# Selecting Variables



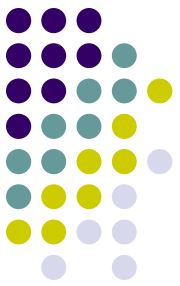
```
Editor - Untitled1 *  
  
libname train 'H:\My Documents\Training Files -- Train';  
run;  
data train.newfile;  
  Set train.sastraining;  
  TotalYears= sum(years);  
  Keep Department TotalYears;  
run;
```

VIEWTABLE: Train.Newfile

	ID	Department	Satisfaction	Status	Degree	Date	TotalYears
1	17	Accounting	4	TT	D	01/01/2007	3
2	12	Accounting	4	T	D	05/18/2008	18
3	1	Anthropology	5	T	D	02/12/2007	20
4	18	Biological Sciences	3	NT	M	01/31/2008	6
5	26	Chemistry	5	NT	M	08/26/2007	10
6	3	Communication Studies	2	NT	M	05/05/2008	5
7	30	Communication Studies	4	T	D	09/09/2007	14
8	20	Computer Science	4	T	D	02/20/2008	16
9	8	Economics	2	NT	M	01/08/2008	8
10	14	Economics	5	T	D	01/14/2008	22
11	6	English	5	T	D	04/25/2007	27
12	2	Family and Consumer Studies	4	NT	M	03/18/2007	15
13	16	Finance	2	TT	D	01/16/2008	5
14	10	Finance	3	T	D	01/10/2008	7
15	9	History	4	PT	B	01/09/2007	5
16	22	History	5	PT	D	06/28/2007	6
17	27	Justice Studies	1	T	D	08/17/2007	13
18	11	Mathematical Studies	5	T	D	06/11/2008	16

VIEWTABLE: Train.Newfile

	Department	TotalYears
1	Accounting	3
2	Accounting	18
3	Anthropology	20
4	Biological Sciences	6
5	Chemistry	10
6	Communication Studies	5
7	Communication Studies	14
8	Computer Science	16
9	Economics	8
10	Economics	22
11	English	27
12	Family and Consumer Studies	15
13	Finance	5
14	Finance	7
15	History	5
16	History	6
17	Justice Studies	13
18	Mathematical Studies	16



# Date Functions

- Create SAS date values
  - TODAY() – obtains the date value from the system clock
  - MDY(month,day,year) – uses numeric month, day, and year values to return the corresponding SAS date value.
- Extract information from SAS date values
  - YEAR (SAS-date) – extracts the year from a SAS date and returns a four-digit value for year
  - QTR (SAS-date) – extracts the quarter from a SAS date and returns a number from 1-4
  - MONTH (SAS-date) extracts the month from a SAS date and returns a number from 1 to 12
  - WEEKDAY (SAS-date) – extracts the day of the week and returns a number from 1 to 7

# Date Function – Weekday Function

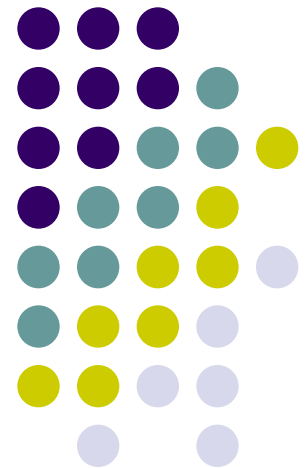


```
libname train 'H:\My Documents\Training Files -- Train';
run;
data train.newfile;
    Set train.sastraining;
    DayofWeek=weekday(Date);
run;
```

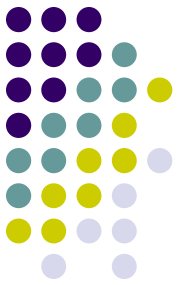
	ID	Department	Satisfaction	Years	Status	Degree	Date	DayofWeek
1	17	Accounting	4	3	TT	D	01/01/2007	2
2	12	Accounting	4	18	T	D	05/18/2008	1
3	1	Anthropology	5	20	T	D	02/12/2007	2
4	18	Biological Sciences	3	6	NT	M	01/31/2008	5
5	26	Chemistry	5	10	NT	M	08/26/2007	1
6	3	Communication Studies	2	5	NT	M	05/05/2008	2
7	30	Communication Studies	4	14	T	D	09/09/2007	1
8	20	Computer Science	4	16	T	D	02/20/2008	4
9	8	Economics	2	8	NT	M	01/08/2008	3
10	14	Economics	5	22	T	D	01/14/2008	2
11	6	English	5	27	T	D	04/25/2007	4
12	2	Family and Consumer Studies	4	15	NT	M	03/18/2007	1
13	16	Finance	2	5	TT	D	01/16/2008	4
14	10	Finance	3	7	T	D	01/10/2008	5
15	9	History	4	5	PT	B	01/09/2007	3
16	22	History	5	6	PT	D	06/28/2007	5
17	27	Justice Studies	1	13	T	D	08/17/2007	6
18	11	Mathematical Studies	5	16	T	D	06/11/2008	4
19	7	Nursing	3	13	NT	M	10/16/2007	3
20	5	Nursing	2	22	T	D	09/06/2007	5
21	21	Philosophy	4	19	T	D	03/24/2008	2
22	24	Physics	1	3	TT	D	01/24/2008	5
23	28	Physics	5	4	PT	B	07/02/2007	2
24	13	Psychology	2	9	T	D	02/13/2008	4
25	15	Psychology	1	20	T	D	01/15/2007	2
26	19	Psychology	5	24	NT	D	02/19/2008	3
27	23	Sociology	2	4	TT	D	01/22/2007	2
28	25	Sociology	1	5	TT	D	07/05/2007	5
29	29	Special Education	3	11	T	D	01/29/2008	3
30	4	Speech Pathology and Audiology	5	5	PT	M	11/24/2007	7

# Proc Univariate

## Proc Univariate

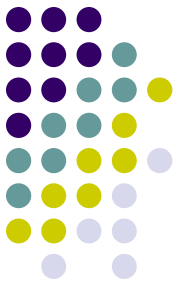


# Proc Univariate

A screenshot of the SAS software interface. The window title is "SAS - [Editor - Untitled1 \*]". The menu bar includes "File", "Edit", "View", "Tools", "Run", "Solutions", "Window", and "Help". The toolbar contains various icons for file operations and execution. On the left, a "Results" pane shows a tree view of the output, with "Univariate: The SAS System" expanded to show results for variables "ID", "Satisfaction", and "Years". Each variable has sub-items for "Moments", "Basic Measures of Location", "Tests For Location", "Quantiles", and "Extreme Observations". The main editor window on the right contains the following SAS code:

```
libname train 'H:\My Documents\Training Files -- Train';  
run;  
  
Proc Univariate data=train.Newfile;  
run;
```

# Proc Univariate



The SAS System 13:10 Monday, August 10

The UNIVARIATE Procedure  
Variable: ID

Moments

N	30	Sum Weights	30
Mean	15.5	Sum Observations	465
Std Deviation	8.80340843	Variance	77.5
Skewness	0	Kurtosis	-1.2
Uncorrected SS	9455	Corrected SS	2247.5
Coeff Variation	56.7961834	Std Error Mean	1.60727513

Basic Statistical Measures

Location		Variability	
Mean	15.50000	Std Deviation	8.80341
Median	15.50000	Variance	77.50000
Mode	.	Range	29.00000
		Interquartile Range	15.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 9.643651	Pr >  t  <.0001
Sign	M 15	Pr >=  M  <.0001
Signed Rank	S 232.5	Pr >=  S  <.0001

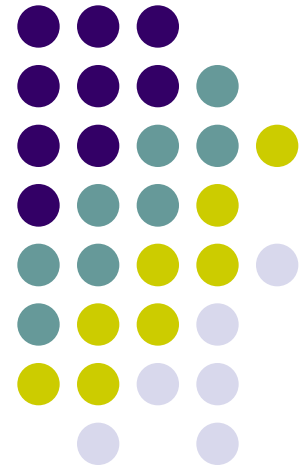
Quantiles (Definition 5)

Quantile	Estimate
100% Max	30.0
99%	30.0
95%	29.0

# Getting started with programming

---

Proc Print

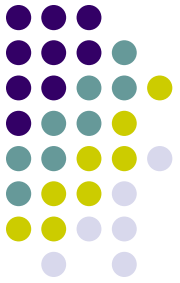




# Proc Print – Beginning Procedures



- Examining data using proc print procedure.
- Display particular variables of interest.
- Display particular observations.
- Display a list report with column totals.



# Default List Report

```
Proc print data=train.sastraining;  
Run;
```

Obs	ID	Department	Satisfaction	Years	Status	Degree	Date
1	1	Anthropology	5	20	T	D	02/12/2007
2	2	Family and Consumer Studies	4	15	NT	M	03/18/2007
3	3	Communication Studies	2	5	NT	M	05/05/2008
4	4	Speech Pathology and Audiology	5	5	PT	M	11/24/2007
5	5	Nursing	2	22	T	D	03/06/2007
6	6	English	5	27	T	D	04/25/2007
7	7	Nursing	3	13	NT	M	10/16/2007
8	8	Economics	2	8	NT	M	01/08/2008
9	9	History	4	5	PT	B	01/03/2007
10	10	Finance	3	7	T	D	01/10/2008
11	11	Mathematical Studies	5	16	T	D	06/11/2008
12	12	Accounting	4	18	T	D	05/10/2008
13	13	Psychology	2	9	T	D	02/13/2008
14	14	Economics	5	22	T	D	01/14/2008
15	15	Psychology	1	20	T	D	01/15/2007
16	16	Finance	2	5	TT	D	01/16/2008
17	17	Accounting	4	3	TT	D	01/01/2007
18	18	Biological Sciences	3	5	NT	M	01/31/2008
19	19	Psychology	5	24	NT	D	02/19/2008
20	20	Computer Science	4	16	T	D	02/20/2008
21	21	Philosophy	4	13	T	D	03/24/2008

# Printing Particular Variables



- Use the **VAR** statement which allows you to:
  - Select variables for your proc print
  - Define the order of the variables in the proc print.

```
Proc print  
  data=train.sastraining;  
  var ID Department  
      Satisfaction;  
Run;
```

A screenshot of the SAS software interface. The main window displays the output of a PROC PRINT statement, showing a table with three columns: ID, Department, and Satisfaction. The table contains 21 rows of data. Below the output, the SAS Editor window shows the code used to generate the output: PROC PRINT DATA=train.sastraining; VAR ID Department Satisfaction; RUN;.

ID	Department	Satisfaction
1	Anthropology	5
2	Family and Consumer Studies	4
3	Communication Studies	2
4	Speech Pathology and Audiology	5
5	Nursing	2
6	English	5
7	Nursing	3
8	Economics	2
9	History	4
10	Finance	3
11	Mathematical Studies	5
12	Accounting	4
13	Psychology	2
14	Economics	5
15	Psychology	1
16	Finance	2
17	Accounting	4
18	Biological Sciences	3
19	Psychology	5
20	Computer Science	4
21	Philosophy	4

# Suppressing Obs Column



The screenshot shows the SAS interface. The 'Output - (Untitled)' window displays the following table:

ID	Department	Satisfaction
1	Anthropology	5
2	Family and Consumer Studies	4
3	Communication Studies	2
4	Speech Pathology and Audiology	5
5	Nursing	2
6	English	5
7	Nursing	3
8	Economics	2
9	History	4
10	Finance	3
11	Mathematical Studies	5
12	Accounting	4
13	Psychology	2
14	Economics	5
15	Psychology	1
16	Finance	2
17	Accounting	4
18	Biological Sciences	3
19	Psychology	5
20	Computer Science	4
21	Philosophy	4

The 'Editor - (Untitled) \*' window contains the following SAS code:

```
proc print data=train.sastraining NOOBS;  
var ID Department Satisfaction;  
run;
```

- The NOOBS option suppresses the number of observations column that shows up on the left hand side of a proc print output.

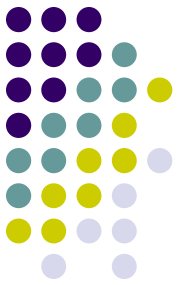
```
Proc print  
data=train.sastraining  
NOOBS;  
Run;
```

# Subsetting Data with the WHERE Statement



- Allows you to select particular observations based on criteria.
- Can be used with most SAS procedures (“IF” statements are generally used in the Data step though).
- Operands
  - Variables and Observations
- Operators
  - Comparisons
  - Logical,
  - Special
  - Functions

# Comparison Operators



Mnemonic	Symbol	Definition
EQ	=	equal to
NE	$\neq$ or $\sim =$	not equal to
GT	>	greater than
LT	<	less than
GE	$\geq$	greater than or equal to
LE	$\leq$	less than or equal to
IN		equal to one of a list

# Examples of WHERE Comparison Operators



## Proc print

```
data=train.sastraining  
NOOBS;  
where department=  
'Psychology';
```

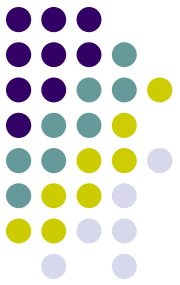
Run;

ID	Department	Satisfaction	Years	Status	Degree	Date
1	Anthropology	5	20	T	D	02/12/2007
2	Family and Consumer Studies	4	10	MT	R	03/10/2007
5	Nursing	2	22	T	D	09/06/2007
5	Dental	5	27	T	D	04/05/2008
7	Nursing	3	13	MT	R	10/15/2007
11	Mathematical Studies	5	18	T	D	05/15/2008
12	Accounting	4	19	T	D	05/10/2008
14	Economics	3	22	T	D	01/14/2008
15	Psychology	1	20	T	D	01/10/2008
19	Psychology	5	24	MT	D	02/19/2008
20	Computer Science	4	16	T	D	02/20/2008
21	Philosophy	4	19	T	D	03/24/2008
27	Special Education	3	11	T	D	01/20/2008
28	Special Education	1	11	T	D	01/20/2008
29	Communication Studies	4	14	T	D	05/05/2007

```
proc print data=train.sastraining  
NOOBS;  
where department=  
'Psychology';
```

run;

ID	Department	Satisfaction	Years	Status	Degree	Date
15	Psychology	2	9	T	D	02/13/2008
15	Psychology	1	20	T	D	01/15/2007
19	Psychology	5	24	MT	D	02/19/2008



# WHERE Logical Operators

- And (&) Used if both expressions are true, then the compound expression is true.
- OR (|) Used if either expression is true, then the compound expression is true.
- Not (^) Can be combined with other operators to reverse the logic of a comparison.



# Examples of WHERE Logical Operators



ID	Department	Sat In Factors	Years	Status	Degree	Date
10	Psychology	1	20	T	D	01/15/2007
15	Psychology	5	24	NT	D	02/15/2006

```
proc print data=train.sastraining NOOBS;
  where department="Psychology" and years>10;
run;
```

**proc print** data=train.sastraining NOOBS;  
 where department= 'Psychology' and  
 years>10;  
**run;**

ID	Department	Sat In Factors	Years	Status	Degree	Date
1	Anthropology	5	02	T	D	02/12/2007
10	Psychology	2	8	T	D	01/15/2006
15	Psychology	1	20	T	D	01/15/2007
18	Psychology	5	24	NT	D	02/15/2006

```
proc print data=train.sastraining NOOBS;
  where department="Psychology" or department="Anthropology";
run;
```

**proc print** data=train.sastraining NOOBS;  
 where department= 'Psychology' or  
 department='Anthropology';  
**run;**

# WHERE Special Operators



- BETWEEN-AND – Used to select observations in which the value of the variable falls within a range of values.
- CONTAINS ? – Used when one wants to select observations that include the specified substring.

# Examples of WHERE Special Operators



```
proc print data=train.sastraining NOOBS;  
  where years between 10 and 15;  
run;
```

The screenshot shows the SAS interface with a table of results. The table has columns for ID, Department, Status, Year, Station, Degree, and Date. The data is filtered to show only records where the year is between 10 and 15.

ID	Department	Status	Year	Station	Degree	Date
2	Family and Consumer Studies	4	10	MT	H	02/18/2007
2	Marketing	3	10	MT	H	10/12/2007
25	Chemistry	2	10	MT	H	02/12/2007
27	Education Studies	1	10	T	D	01/25/2008
28	General Education	3	11	T	D	01/25/2008
30	Communication Studies	3	14	T	D	02/09/2007

```
proc print data=train.sastraining NOOBS;  
  where Department ? 'Nurs';  
run;
```

The screenshot shows the SAS interface with a table of results. The table has columns for ID, Department, Status, Year, Station, Degree, and Date. The data is filtered to show only records where the department is 'Nurs'.

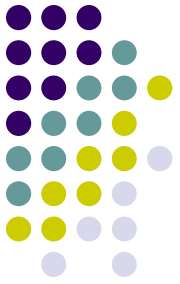
ID	Department	Status	Year	Station	Degree	Date
5	Nursing	2	22	T	D	02/02/2007
6	Nursing	2	15	MT	H	10/16/2007

# Column Totals



- Can provide a Total
- Can also provide subtotals if data is printed in groups.

# Example of Column Total



The screenshot displays the SAS interface with three windows: Results, Output - (Untitled), and Editor - Untitled1\*.

**Results Window:** Shows a printout of the data table.

**Output - (Untitled) Window:** Displays the following table:

ID	Department	Satisfaction	Years	Status	Degree	Date
1	Anthropology	5	20	T	D	02/12/2007
2	Family and Consumer Studies	4	15	NT	M	03/18/2007
3	Communication Studies	2	5	NT	M	05/05/2008
4	Speech Pathology and Audiology	5	5	PT	M	11/24/2007
5	Nursing	2	22	T	D	09/06/2007
6	English	5	27	T	D	04/25/2007
7	Nursing	3	13	NT	M	10/16/2007
8	Economics	2	8	NT	M	01/08/2008
9	History	4	5	PT	B	01/09/2007
10	Finance	3	7	T	D	01/10/2008
11	Mathematical Studies	5	16	T	D	06/11/2008
12	Accounting	4	18	T	D	05/18/2008
13	Psychology	2	9	T	D	02/13/2008
14	Economics	5	22	T	D	01/14/2008
15	Psychology	1	20	T	D	01/15/2007
16	Finance	2	5	TT	D	01/16/2008
17	Accounting	4	3	TT	D	01/01/2007
18	Biological Sciences	3	6	NT	M	01/31/2008
19	Psychology	5	24	NT	D	02/19/2008
20	Computer Science	4	16	T	D	02/20/2008
21	Philosophy	4	19	T	D	03/24/2008
22	History	5	6	PT	D	06/28/2007
23	Sociology	2	4	TT	D	01/22/2007
24	Physics	1	3	TT	D	01/24/2008
25	Sociology	1	5	TT	D	07/05/2007
26	Chemistry	5	10	NT	M	08/26/2007
27	Justice Studies	1	13	T	D	08/17/2007
28	Physics	5	4	PT	B	07/02/2007
29	Special Education	3	11	T	D	01/29/2008
30	Communication Studies	4	14	T	D	09/09/2007
		====				
			355			

**Editor - Untitled1\* Window:** Contains the following SAS code:

```
proc print data=train.sastraining NOOBS;
  sum years;
run;
```

**Bottom Status Bar:** NOTE: 3 Lines Submitted. C:\Documents and Settings\cughrin Ln 4, Col 1



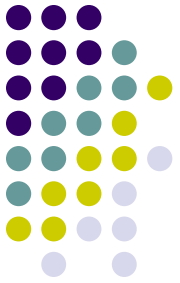
# Proc Sort



# Overview of Proc Sort

- Sorts (arranges) observations of the data set.
- Can create a new SAS data set containing rearranged observations.
- Can sort on more than one variable at a time.
- Sorts ascending (default) and descending.
- Does not provide printed output (that requires the proc print statements).
- Treats missing data as smallest possible value.

# Proc Sort Example

A screenshot of the SAS software interface. The main window displays the output of a PROC SORT and PROC PRINT statement. The output is a table with three columns: Department, Satisfaction, and Years. The data is sorted by Department. Below the table, the SAS code used to generate the output is shown in a separate window.

Department	Satisfaction	Years
Accounting	4	19
Accounting	4	3
Anthropology	5	20
Biological Sciences	3	6
Chemistry	2	5
Communication Studies	5	10
Communication Studies	4	14
Computer Science	4	16
Economics	2	8
Economics	5	22
English	5	27
Family and Consumer Studies	4	15
Finance	3	7
Finance	2	5
History	4	5
History	5	6
Justice Studies	1	13
Mathematical Studies	5	16
Nursing	2	22
Nursing	3	13
Philosophy	4	13
Physics	1	3
Physics	4	4
Psychology	2	9
Psychology	1	20
Psychology	5	24
Sociology	2	4
Sociology	1	5
Special Education	3	11
Speech Pathology and Audiology	5	5

```
proc sort data=train.sastraining;
  by Department;
run;
proc print data=train.sastraining NOOBS;
  var Department Satisfaction Years;
run;
```

```
proc sort data=train.sastraining;
  by Department;
run;
proc print data=train.sastraining NOOBS;
  var Department Satisfaction Years;
run;
```



# Printing Totals and Subtotals Proc Sort and Proc Print Example



```
proc sort data=train.sastraining;  
  by Department;  
run;  
proc print data=train.sastraining NOOBS;  
  by Department;  
  sum years;  
run;
```

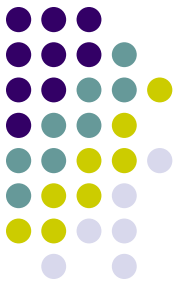
The screenshot shows the SAS interface with the following components:

- Output Window:** Displays the results of the PROC PRINT statement, showing a table with columns: ID, Satisfaction, Years, Status, Degree, and Date. The data is grouped by Department, with subtotals for each department.
- Editor Window:** Shows the SAS code used to generate the output: 

```
proc sort data=train.sastraining;  
  by Department;  
run;  
proc print data=train.sastraining NOOBS;  
  by Department;  
  sum years;  
run;
```

The SAS System					
----- Department=Accounting -----					
ID	Satisfaction	Years	Status	Degree	Date
17	4	3	TT	D	01/01/2007
12	4	18	T	D	05/18/2008
Department		21			
----- Department=Anthropology -----					
ID	Satisfaction	Years	Status	Degree	Date
1	5	20	T	D	02/12/2007
----- Department=Biological Sciences -----					
ID	Satisfaction	Years	Status	Degree	Date
18	3	6	NT	M	01/31/2008
----- Department=Chemistry -----					
ID	Satisfaction	Years	Status	Degree	Date
26	5	10	NT	M	08/26/2007
----- Department=Communication Studies -----					
ID	Satisfaction	Years	Status	Degree	Date
3	2	5	NT	M	05/05/2008
30	4	14	T	D	09/03/2007

# Page Breaks with Proc Sort and Proc Print



The screenshot shows the SAS software interface. The main window displays a table with the following data:

ID	Satisfaction	Years	Status	Degree	Date
12	4	2	TT	D	01/31/2007
12	4	16	T	D	02/10/2008
-----					
Department		21			

Below the table, the SAS code is visible in the editor window:

```
proc sort data=train.sastraining;  
  by Department;  
run;  
proc print data=train.sastraining NOOBS;  
  by Department;  
  Pageby Department;  
  sum years;  
run;
```

```
proc sort data=train.sastraining;  
  by Department;  
run;  
proc print data=train.sastraining NOOBS;  
  by Department;  
  Pageby Department;  
  sum years;  
run;
```