PSTAT 130

SAS BASE PROGRAMMING

- Lecture 1 -

Dr. Julie Swenson

SAS Background

- SAS was once an acronym for “Statistical Analysis System”
- Today, SAS is much more powerful
  - Original acronym is no longer used

SAS Components

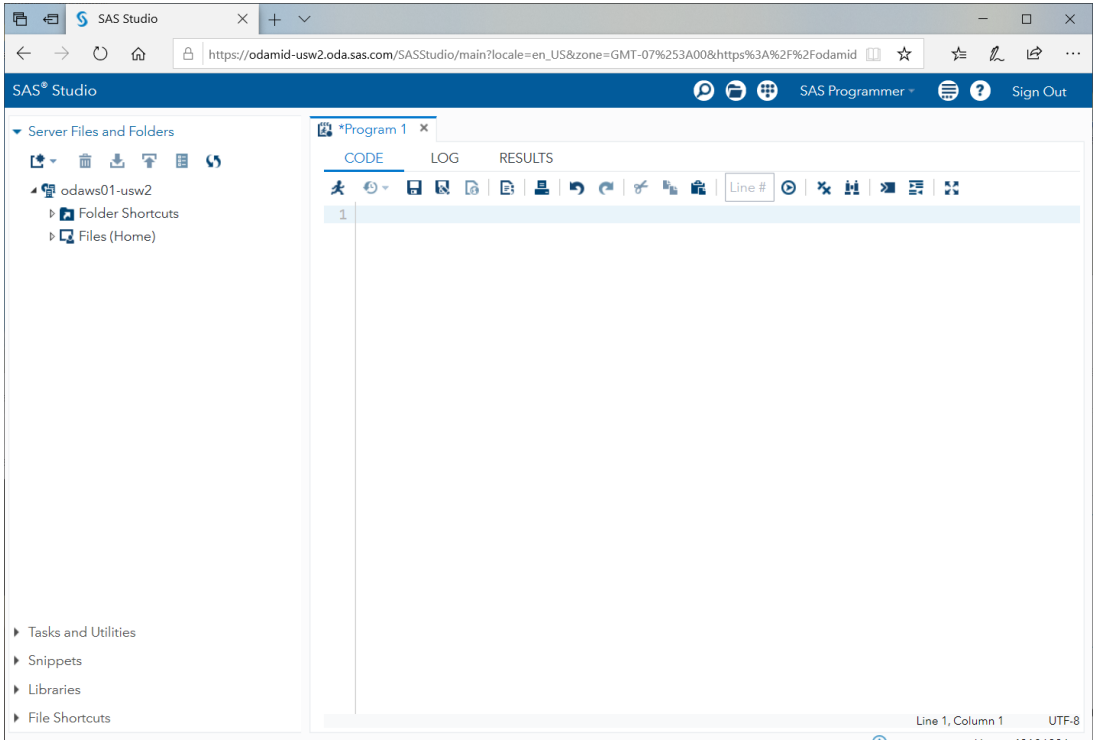
- This class
  - Base SAS – basic procedures and data management
- SAS has over 200 components, including
  - SAS/STAT – statistical analysis
  - SAS/GRAPH – high quality graphics & presentations
  - SAS/ACCESS – reads data directly from databases
  - SAS/ETS – econometrics and time series
  - SAS/INSIGHT – data mining
  - SAS/QC – quality control
  - SAS/PH – clinical trials

Objectives

- Open the program
- Navigate the environment
- Basic syntax
- First SAS program
- Datalines statement

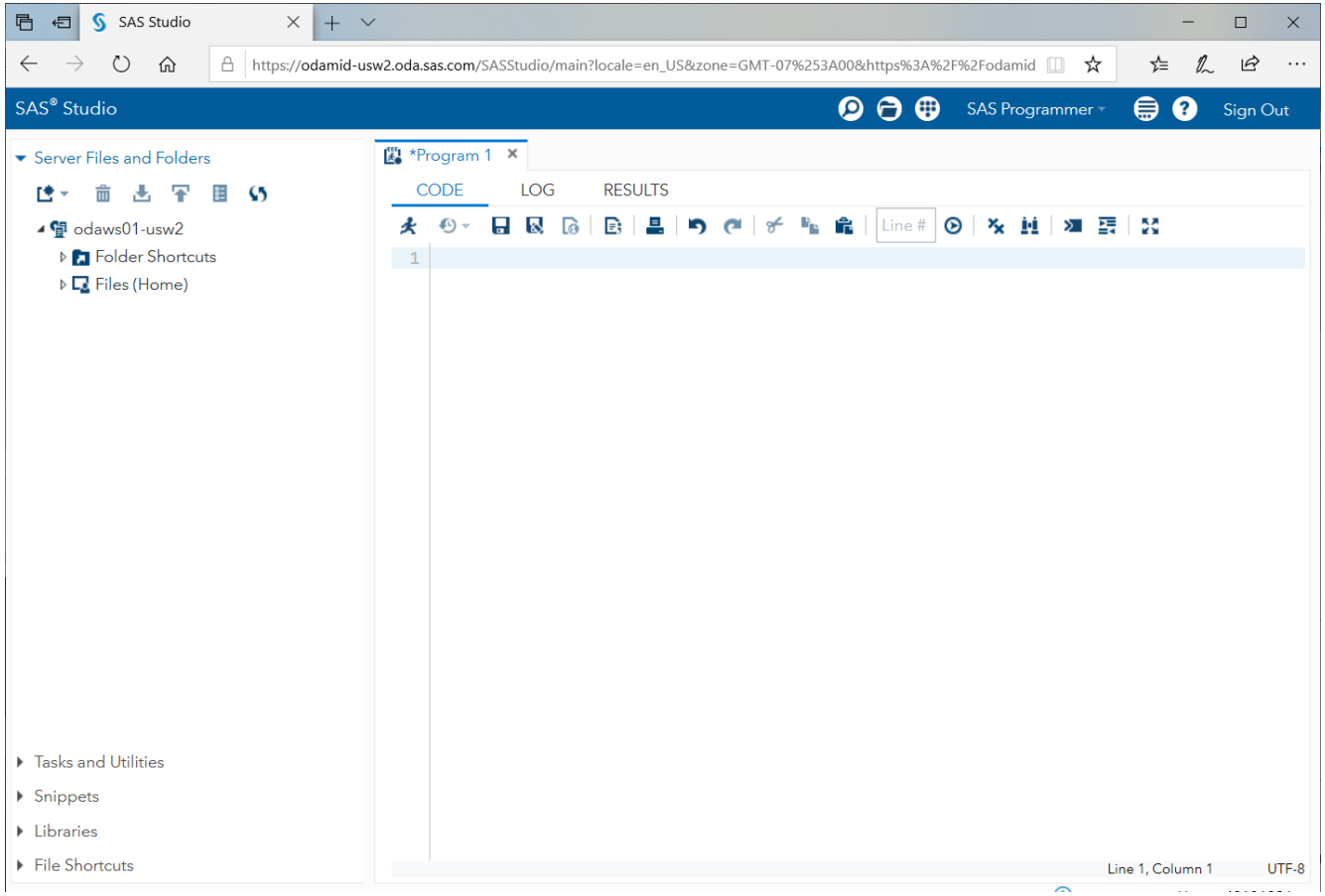
First Step

- Open SAS
  - SAS Studio through SAS OnDemand for Academics, *or*
  - SAS University Edition on your computer



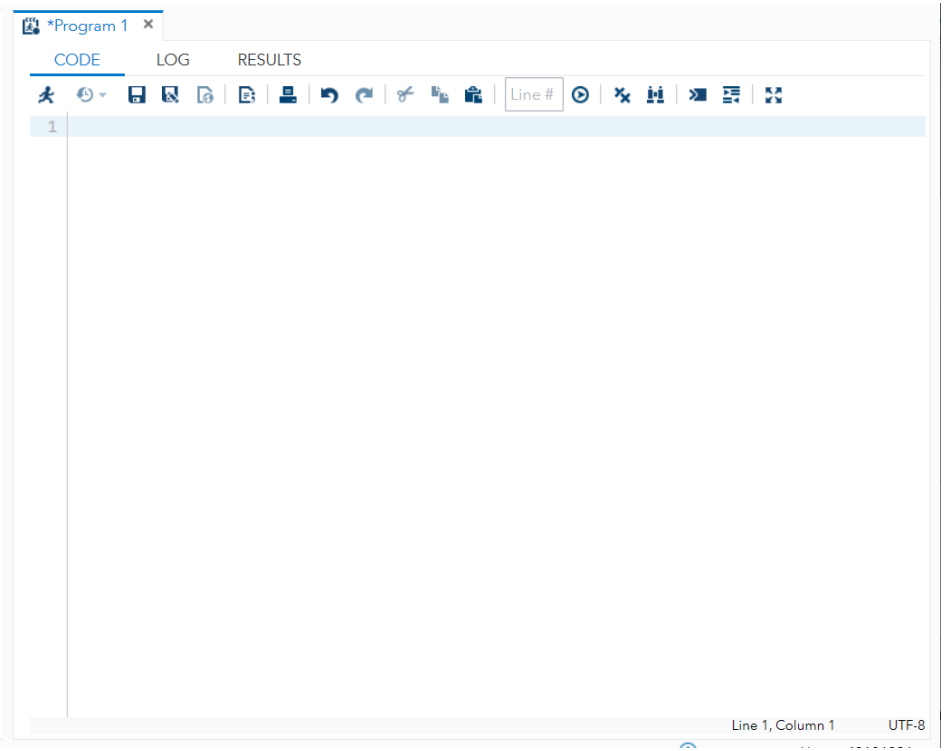
The Main Windows

- Code
- Log
- Results
- Output Data
- Explorer



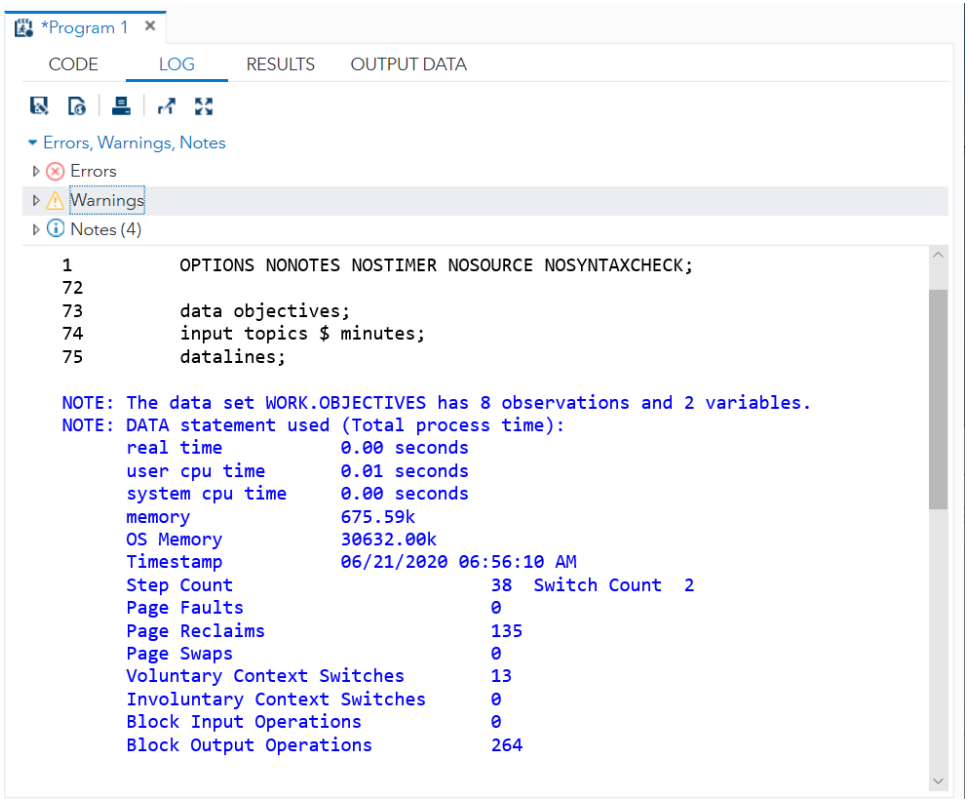
Code

- Edit, execute, and save SAS programs



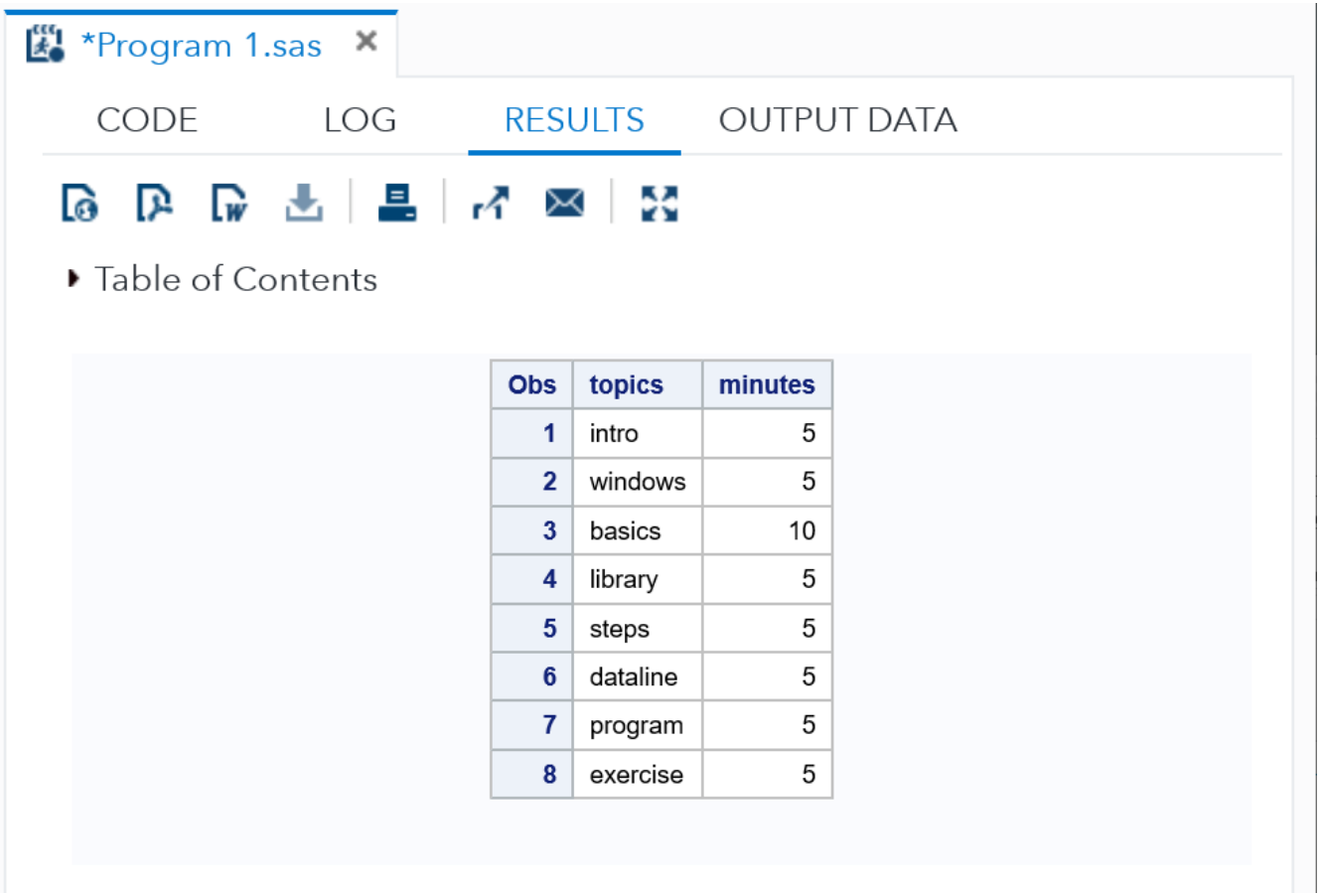
Log

- Displays status messages regarding the execution of SAS procedures



Results

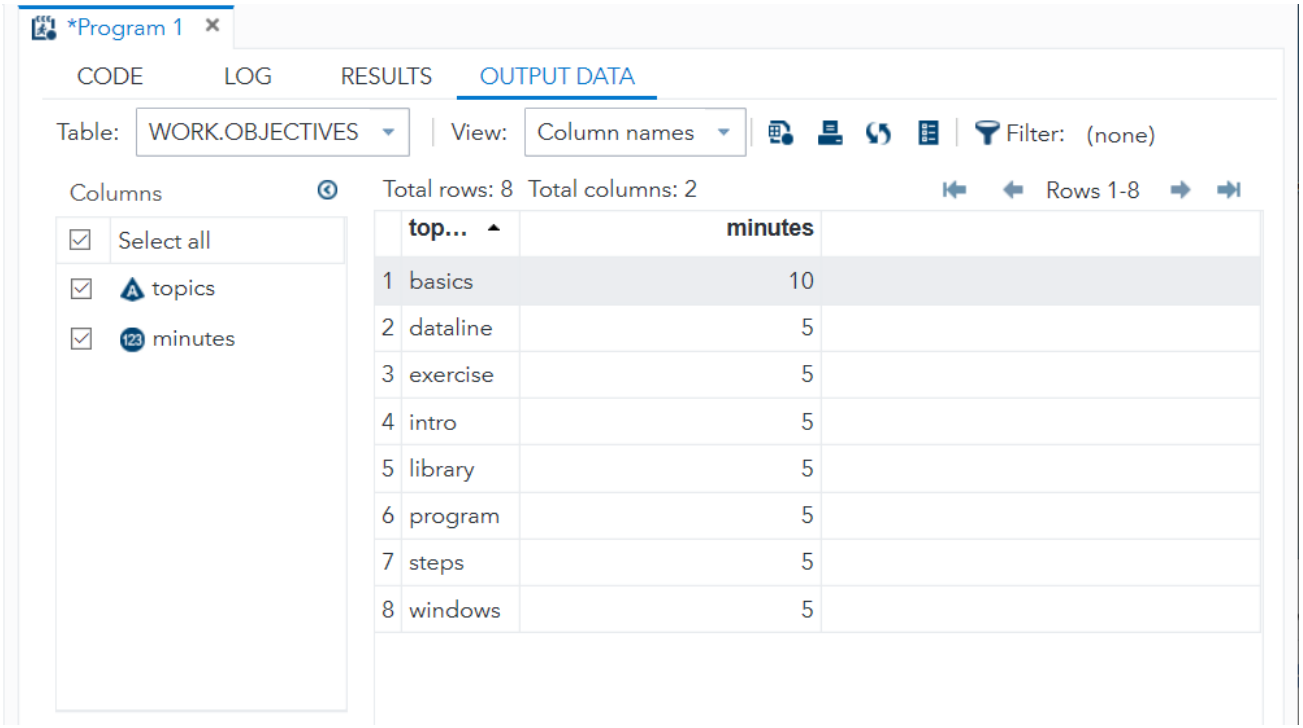
- Displays generated report(s) as a single html file



Obs	topics	minutes
1	intro	5
2	windows	5
3	basics	10
4	library	5
5	steps	5
6	dataline	5
7	program	5
8	exercise	5

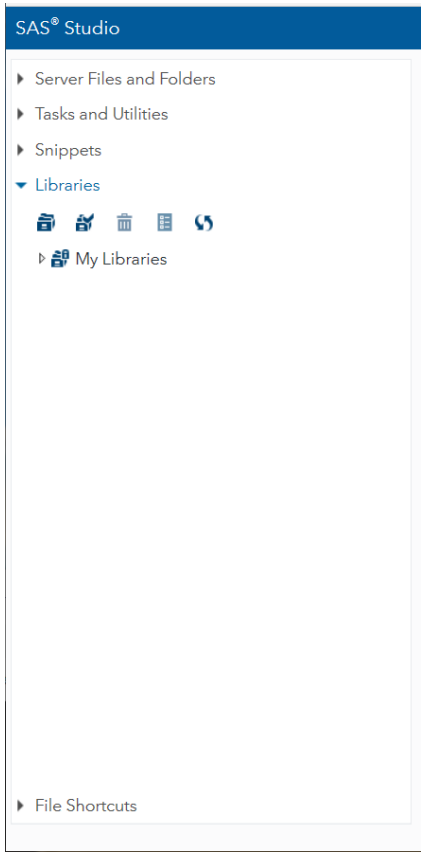
Output Data

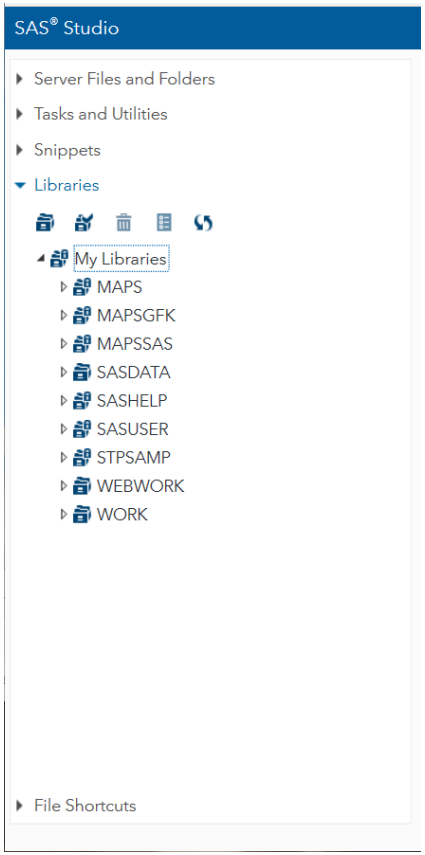
- Visible when displaying output SAS dataset(s)

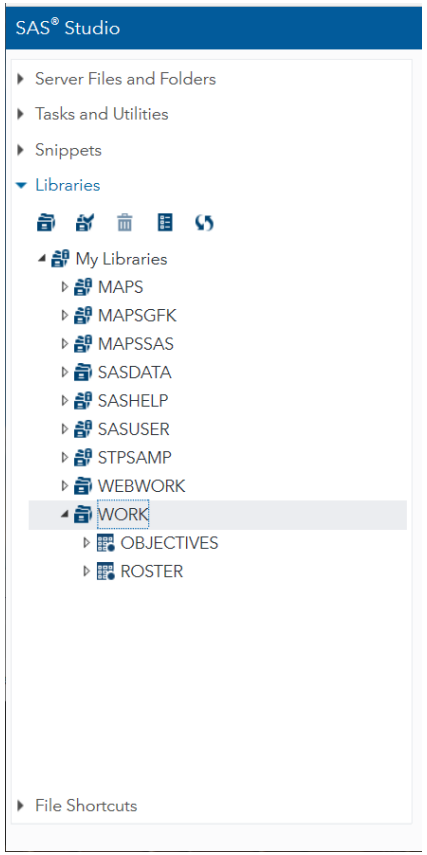


Explorer

- Navigate libraries, data sets, and other SAS objects







Getting Started

- How to create a new program?
  - F4
  - Right-click in the Explorer window: New → SAS Program

Getting Started

- How to submit/execute a program?
  - Submitting the entire program
    - F3
    - Click on the “running man” symbol in the toolbar

  - Submitting a portion of the program
    - Highlight the portion you’d like to submit
      - F3
      - <right click> → Run all or selected code (F3)

Getting Started

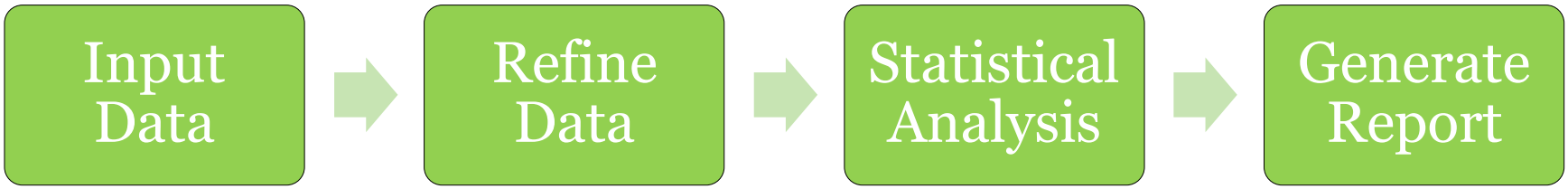
- How to save a program?
  - Click on the save button in the toolbar

  - Ctrl + S

File Extensions

- What are the file extensions for
  - A SAS program?
    - .sas
  - A SAS data set?
    - .sas7bdat
  - A SAS log file?
    - .log
  - A SAS report file?
    - .html (by default)
    - .lst (for a listing report)

The Basics: SAS Process



The Basics

- SAS statements
  - Always begin with a keyword
  - Always end with a semicolon (;)
  - Are free format
    - i.e. Can begin at any location and end at any location
      - Entire program can be written on one line, or many lines
    - EXCEPT when using the `datalines;` statement
- SAS
  - Is not case sensitive
    - i.e. `daTa nOtCaseSensitive;`
    - EXCEPT in the case of string comparisons

The Basics

- Names of SAS data sets and variables must
  - Be no longer than 32 characters
  - Begin with a letter or underscore
  - Contain only letters, numbers, or underscores (_)

The Basics

- Two types of variables
  - Character
    - Contains any value: letters, numbers, special characters, and blanks. Character values are stored with a length of 1 to 32,767 bytes. One byte equals one character.
  - Numeric
    - Stored as floating point numbers in 8 bytes of storage by default. Eight bytes of floating point storage provide space for 16 to 17 significant digits. You are not restricted to 8 digits.

The Basics

- Comments
  - What are they?
  - Why should we use them?
  - Single line: begin with an asterisk (*) and ends with a semicolon (;)
  - Multiple lines: begins with a /* and ends with a */

The Basics

- SAS has context-sensitive help.
  - Highlight a keyword, right click on it, click on “Syntax Help”

Libraries

- SAS file names contain 2 levels
  - Level 1: <library-name>
  - Level 2: <data-set-name>
    - i.e. <library-name>.<data-set-name>

- What does this mean?
  - SAS references folders called libraries when accessing SAS data sets
  - Libraries are simply pointers to folder locations on the disk drive
    - i.e. '/home/user'

Libraries

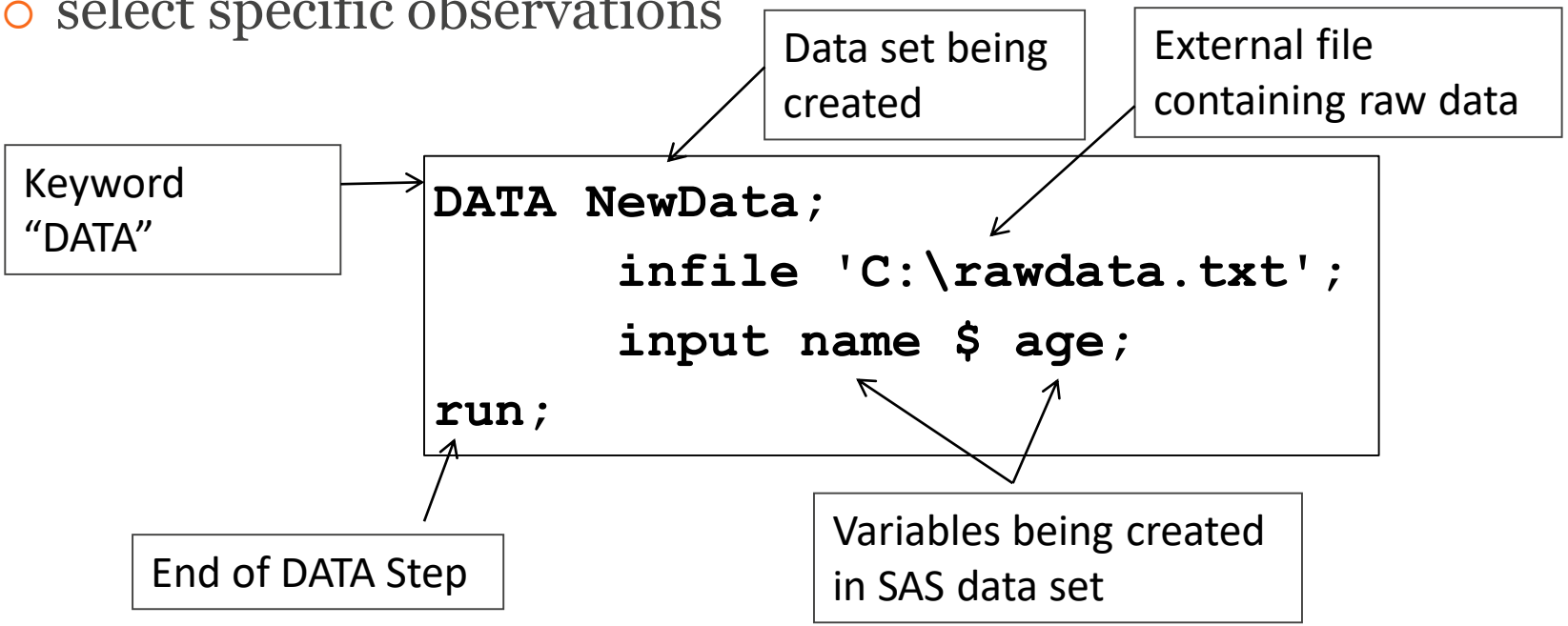
- SAS has a number of existing libraries including:
  - work
  - sashelp
- work
  - Is a temporary library
  - Is the default library
- sashelp
  - Is a permanent library
- New libraries must be assigned

Libraries

- To define a library, you need
  - The **libname** keyword
  - A user-defined libref name
  - A folder location
- Library references (libref) must
  - Start with a letter or an underscore
  - Be 8 characters or less
- General format:
  - `libname home '/home/user';`

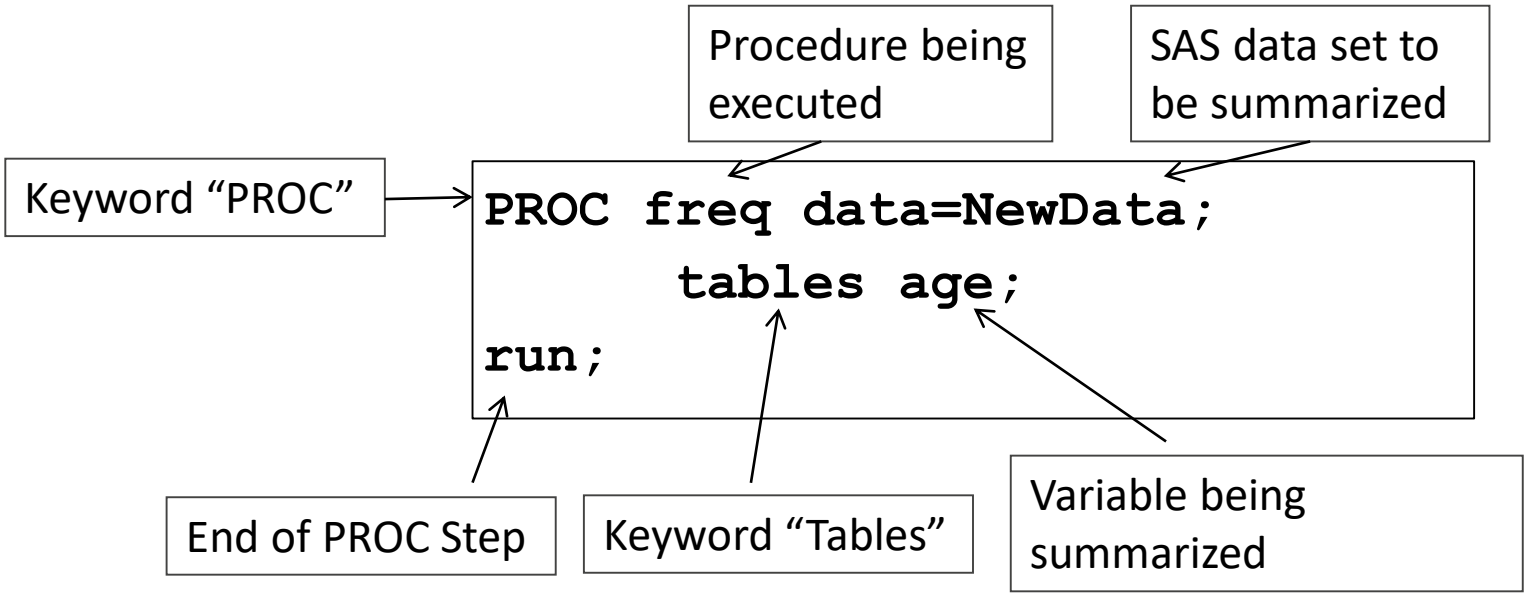
The DATA Step

- DATA Step: A set of statements that
  - read in a data file
  - assign variable names, labels, and formats
  - select specific observations



The PROC Step

- PROC Step: A set of statements that
  - perform "utility" operations on a data set
  - analyze data
  - output results or reports

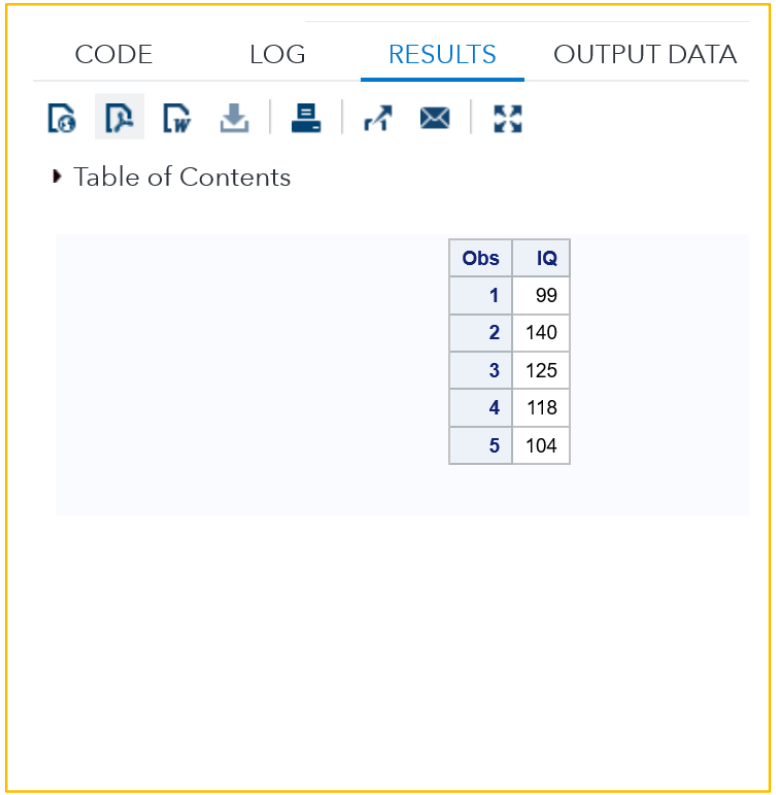


First SAS Program

```
DATA intelligence;
input IQ;
datalines;
99
140
125
118
104
;
run;

PROC print;
run;
```

i.e. work.intelligence



Obs	IQ
1	99
2	140
3	125
4	118
5	104

Datalines Statement

- Allows raw data to be placed within a SAS program
  - When is this used?
  - Why is this useful?

- Properties
  - Can only be used once in a DATA step
  - By default, each data line is a separate observation
  - Default delimiter is a blank space
  - Is the last statement in the DATA step

Datalines Statement

- Requires the use of an input statement
  - Input statement
    - Identifies the order of values in the data lines
    - Creates variable names
    - Assigns variable types
    - Assigns input values to corresponding variables
- Variables take one of two types
  - Character ($)
  - Numeric

Datalines Statement

- Structure:

```
datalines;
< data >
;
```

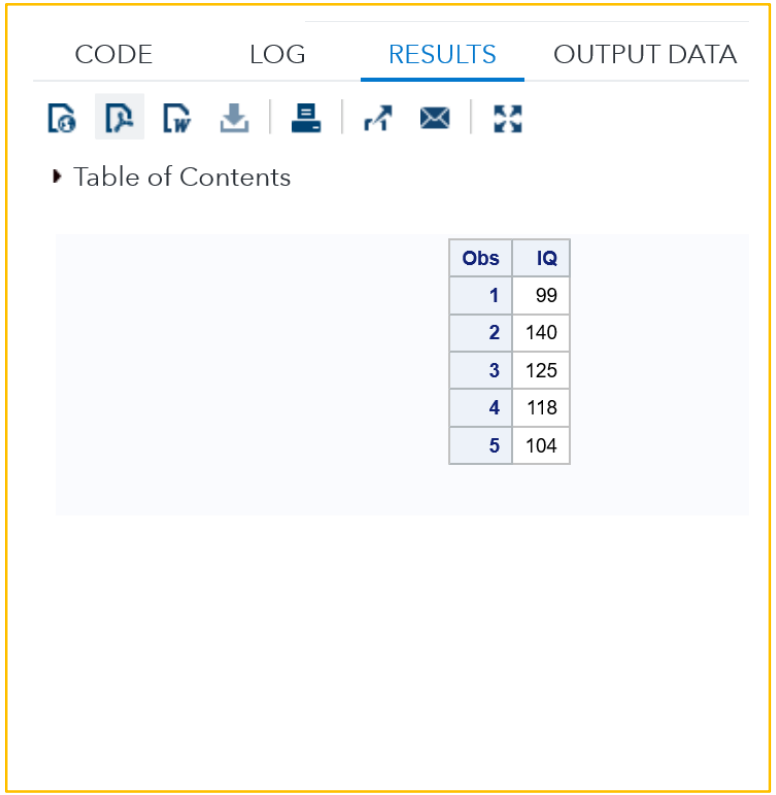
- Keyword `datalines;` precedes the data
- Data is listed
- The null statement (`;`) follows the last line of data
  - This indicates the end of the input data

Recall Our First SAS Program

```
DATA intelligence;
input IQ;
datalines;
99
140
125
118
104
;
run;

PROC print;
run;
```

i.e. work.intelligence



Obs	IQ
1	99
2	140
3	125
4	118
5	104

Class Exercise

- Create a dataset called HouseReps
- This data set should contain two variables
  - State
  - Reps
- Use the following data:

```
CA 53
NJ 12
NY 27
TX 36
WA 10
```