

Statistical Analysis System

SAS

# Brief History

Anthony J. Barr



Jim Goodnight



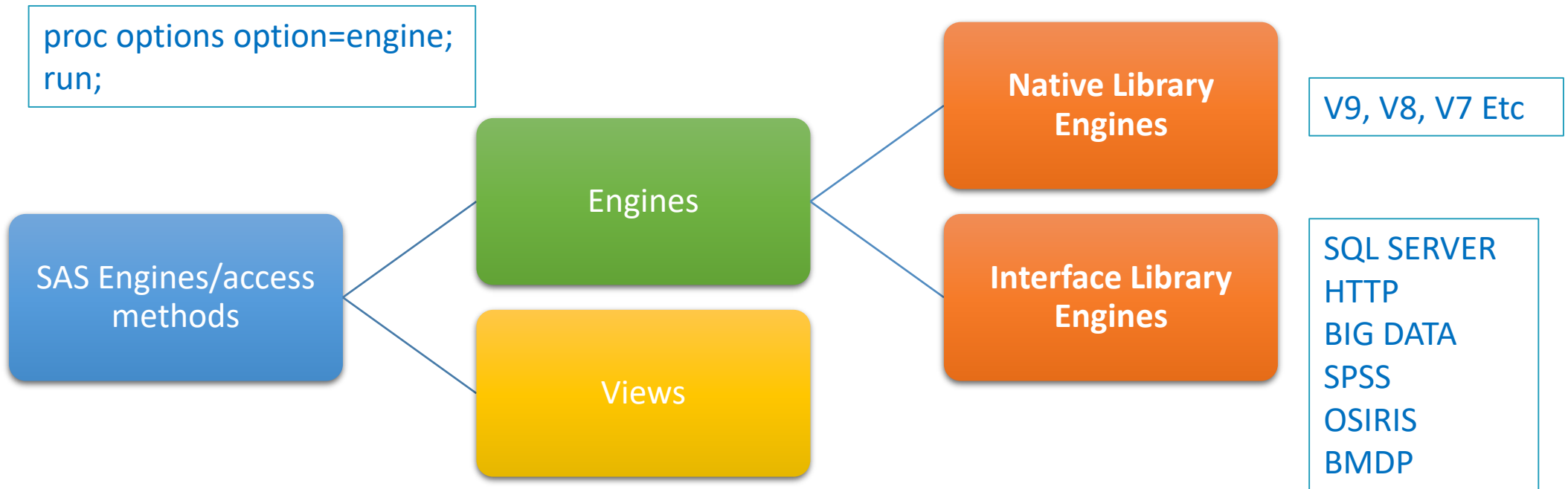
- **Sept 1962** - Began assistantship with North Carolina State University Computing Center. Program was written on IBM 1410 assembler.
- **June 1976** - SAS Institute, Inc. was incorporated. (Release of 1976 version of SAS.)  
[https://www.sas.com/en\\_us/home.html](https://www.sas.com/en_us/home.html)

# SAS ARCHITECTURE

- Multi Vendor Architecture (SAS run in almost all platforms)

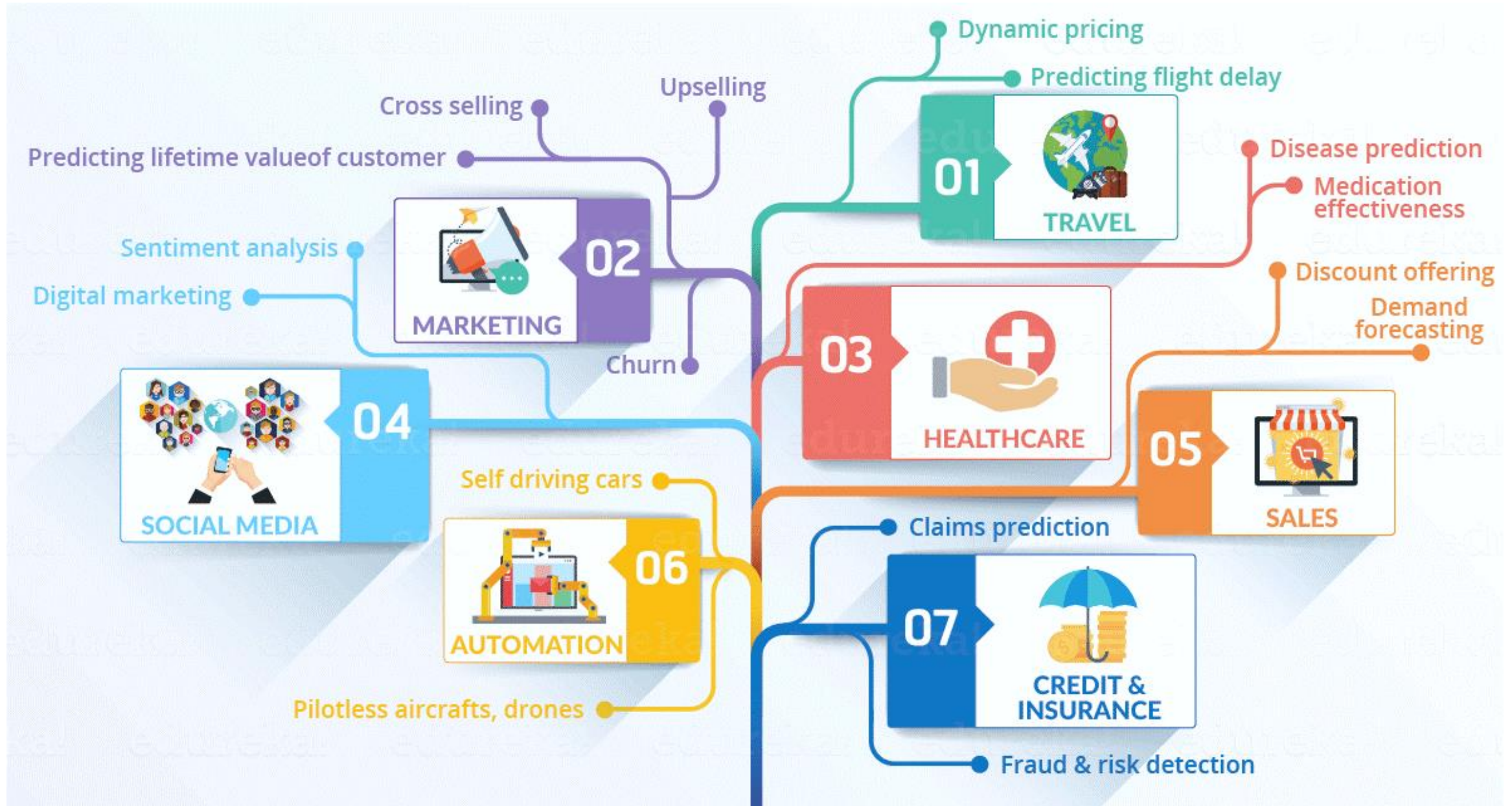


- Multi Engine Architecture.





# INTRODUCTION



# COMPONENTS OF SAS PROGRAMMING

## i. Base SAS

It is a basis for all SAS software. SAS has hardware agility and integrates into any computer environment. It is easy to learn and is not code intensive. It provides data ease in data readability and interpretation.

## ii. SAS/GRAPH

It a data visualization technique. It helps represent structured data into graphs by enhancing the process of interpretation.

## iii. SAS/STAT

It helps us perform various statistical analysis variance, regression, and psychometric analysis. Some statistical techniques like such as ANOVA procedures are specially strung in the SAS environment.

## iv. SAS/ETS

It is specially used for forecasting and for the time series analysis.

## v. SAS/IML

IML is called Interactive Matrix Language. It is used to translate mathematical formulas into an innovative program for matrix computation and optimization.

## vi. SAS/INSIGHTS and Enterprise Miner

SAS/INSIGHTS and Enterprise Miner are used for data mining.

# FEATURES OF SAS PROGRAMMING

**Strong Data  
Analysis Abilities**

**Flexible 4  
Generation  
Programming  
Language (4GL)**

**SAS Studio**

**Support for  
Various Types of  
Data Format**

**Management**

**Report Output  
Format**

**Data Encryption  
Algorithms**

# ADVANTAGES OF SAS

**Easy to learn**

**Ability to  
handle large  
database**

**Easy to debug**

**Tested  
algorithms**

**SAS Customer  
support**

**Data Security**

**SAS GUI**

**Nice Output**

**Huge Job  
Prospects**

# DISADVANTAGES OF SAS

**Cost**

**SAS is not open  
source**

**Lack of graphic  
representation**

**Difficult Text  
Mining**

**Difficult than R**

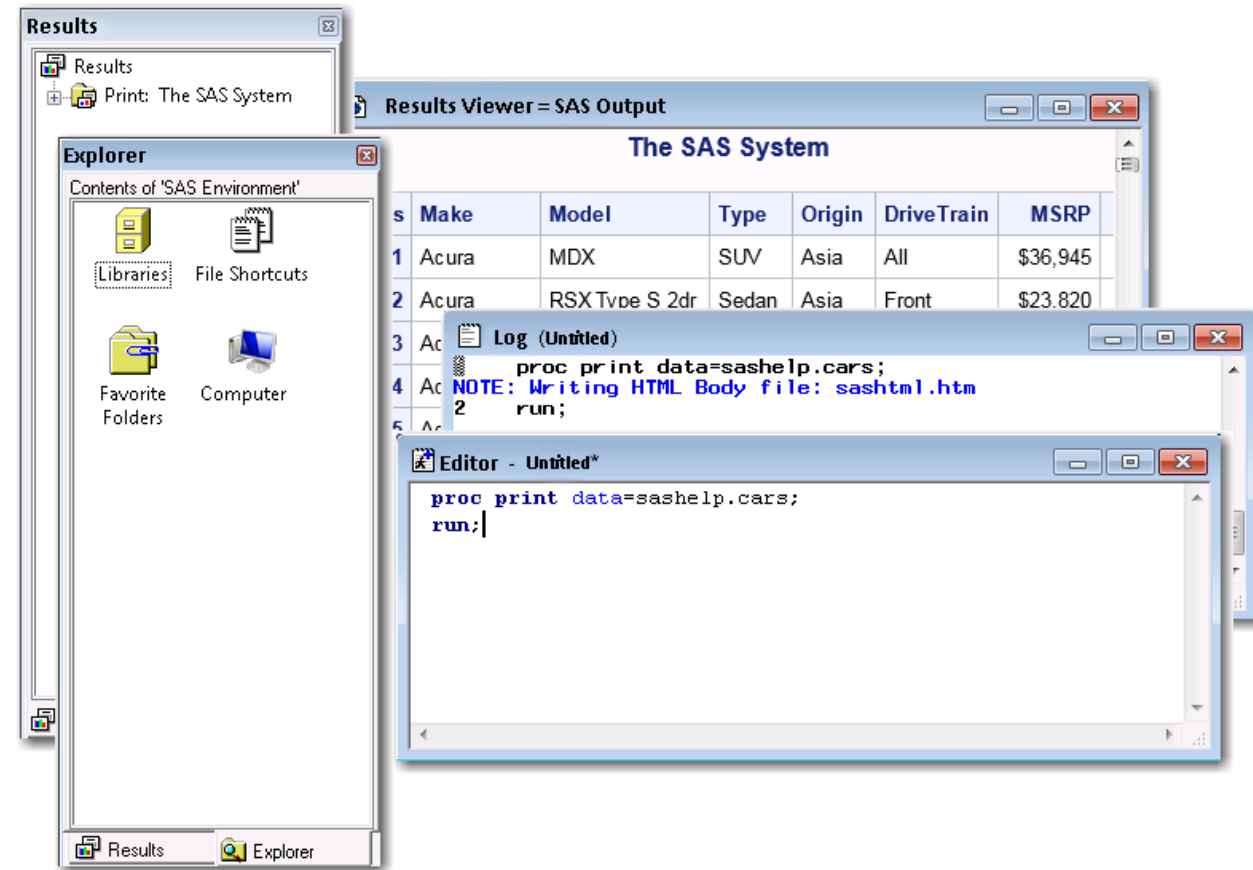
# DIFFERENT TYPES OF SAS SESSIONS

You can run SAS in any of several ways that might be available for your operating environment:

- SAS windowing environment
- Interactive line mode
- Non-Interactive mode
- Batch (or background) mode

In addition,

- SAS/ASSIST software provides a menu-driven system for creating and running your SAS programs.
- Object Server Mode





# SAS Editor Window (University Edition)

The screenshot displays the SAS Studio interface. The top navigation bar includes the 'SAS® Studio' logo, user profile icons, the role 'SAS Programmer', and a 'Sign Out' button. The left sidebar, titled 'Server Files and Folders', shows a tree structure with 'Folder Shortcuts', 'My Folders' (containing 'Crimes' and 'sasuser.v94'), and a list of files including 'Owners.csv', 'Pets.csv', 'ProceduresDetails.csv', and 'ProceduresHistory-1.csv'. The main workspace is divided into three tabs: 'CODE', 'LOG', and 'RESULTS'. The 'CODE' tab is active, showing a text editor with the following SAS code:

```
1  
2 Data Test;  
3 Set sashelp.class;  
4 run;  
5  
6 proc print data=test;  
7 run;
```

A red rectangular box highlights the code editor area, containing the text:

**Editor**  
➤ This is a text editor window to type, edit and submit SAS programs

The bottom status bar indicates the current position as 'Line 7, Column 5' in 'UTF-8' encoding, with links for 'Messages' and 'User: sasdemo'.

# SAS LOG Window

The screenshot displays the SAS Studio interface. The top navigation bar includes the SAS logo, the text 'SAS Studio', and several icons for user profile, workspace, and help. The left sidebar shows a tree view of 'Server Files and Folders' with a folder named 'sasuser.v94' containing several CSV files. The main window is titled '\*Program 1' and has tabs for 'CODE', 'LOG', 'RESULTS', and 'OUTPUT DATA'. The 'LOG' tab is active, showing a list of 'Errors, Warnings, Notes'. The log content includes SAS code snippets and system messages.

SAS Studio

Server Files and Folders

- Folder Shortcuts
- My Folders
  - Crimes
  - sasuser.v94
    - Owners.csv
    - Pets.csv
    - ProceduresDetails.csv
    - ProceduresHistory-1.csv
- Tasks and Utilities
- Snippets
- Libraries
- File Shortcuts

\*Program 1

CODE LOG RESULTS OUTPUT DATA

Errors, Warnings, Notes

- Errors
- Warnings
- Notes (5)

```
73      Data Test;  
74      Set sashelp.class;  
75      run;  
  
NOTE: There were 19 observations read from the data set SASHELP.CLASS.  
NOTE: The data set WORK.TEST has 19 observations and 5 variables.  
NOTE: DATA statement used (Total process time):  
      real time           0.03 seconds  
      cpu time            0.01 seconds  
  
76  
77      proc print data=test;  
78      run;  
  
NOTE: There were 19 observations read from the data set WORK.TEST.  
NOTE: PROCEDURE PRINT used (Total process time):  
      real time           0.18 seconds  
      cpu time            0.16 seconds
```

# SAS HTML Output

SAS® Studio

Server Files and Folders

Folder Shortcuts

My Folders

Crimes

sasuser.v94

Owners.csv

Pets.csv

ProceduresDetails.csv

ProceduresHistory-1.csv

Tasks and Utilities

Snippets

Libraries

File Shortcuts

\*Program 1

CODELOGRESULTSOUTPUT DATA






Table of Contents

Obs	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69.0	112.5
2	Alice	F	13	56.5	84.0
3	Barbara	F	13	65.3	98.0
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5
6	James	M	12	57.3	83.0
7	Jane	F	12	59.8	84.5
8	Janet	F	15	62.5	112.5
9	Jeffrey	M	13	62.5	84.0
10	John	M	12	59.0	99.5
11	Joyce	F	11	51.3	50.5
12	Judy	F	14	64.3	90.0
13	Louise	F	12	56.3	77.0
14	Mary	F	15	66.5	112.0
15	Philip	M	16	72.0	150.0
16	Robert	M	12	64.8	128.0
17	Ronald	M	15	67.0	133.0
18	Thomas	M	11	57.5	85.0
19	William	M	15	66.5	112.0

Messages: 1User: sasdemo

# SAS Output Data View

SAS® Studio

   SAS Programmer   Sign Out

► Server Files and Folders

► Tasks and Utilities

► Snippets

▼ Libraries



My Libraries

► SASHELP

► SASUSER

► WEBWORK

WORK

► TEST

► File Shortcuts

 \*Program 1 x

CODE

LOG

RESULTS

OUTPUT DATA

Table: WORK.TEST

View: Column names

    Filter: (none)

Columns



Total rows: 19 Total columns: 5

Rows 1-19

- ☒ Select all
- ☒ A Name
- ☒ A Sex
- ☒ 123 Age
- ☒ 123 Height
- ☒ 123 Weight

Property	Value
Label	
Name	
Length	
Type	
Format	

	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69	115
2	Alice	F	13	56.5	110
3	Barbara	F	13	65.3	115
4	Carol	F	14	62.8	110
5	Henry	M	14	63.5	115
6	James	M	12	57.3	110
7	Jane	F	12	59.8	110
8	Janet	F	15	62.5	110
9	Jeffrey	M	13	62.5	115
10	John	M	12	59	110
11	Joyce	F	11	51.3	110
12	Judy	F	14	64.3	110
13	Louise	F	12	56.3	110
14	Mary	F	15	66.5	110

```

Output - (Untitled)
The SAS System
22

The UNIVARIATE Procedure
Variable: y

Moments

N          30          Sum Weights          30
Mean       117.141514   Sum Observations   3514.24541
Std Deviation 181.359976   Variance           32891.4407
Skewness   2.06391409   Kurtosis           3.68025503
Uncorrected SS 1365515.81   Corrected SS       953851.781
Coeff Variation 154.821267   Std Error Mean     33.1116499

Basic Statistical Measures

Location                      Variability

Mean       117.1415   Std Deviation      181.35998
Median     31.0157   Variance           32891
Mode       .         Range              691.88626
                        Interquartile Range  151.49364

Tests for Location: Mu0=0

Test          -Statistic-          -----p Value-----

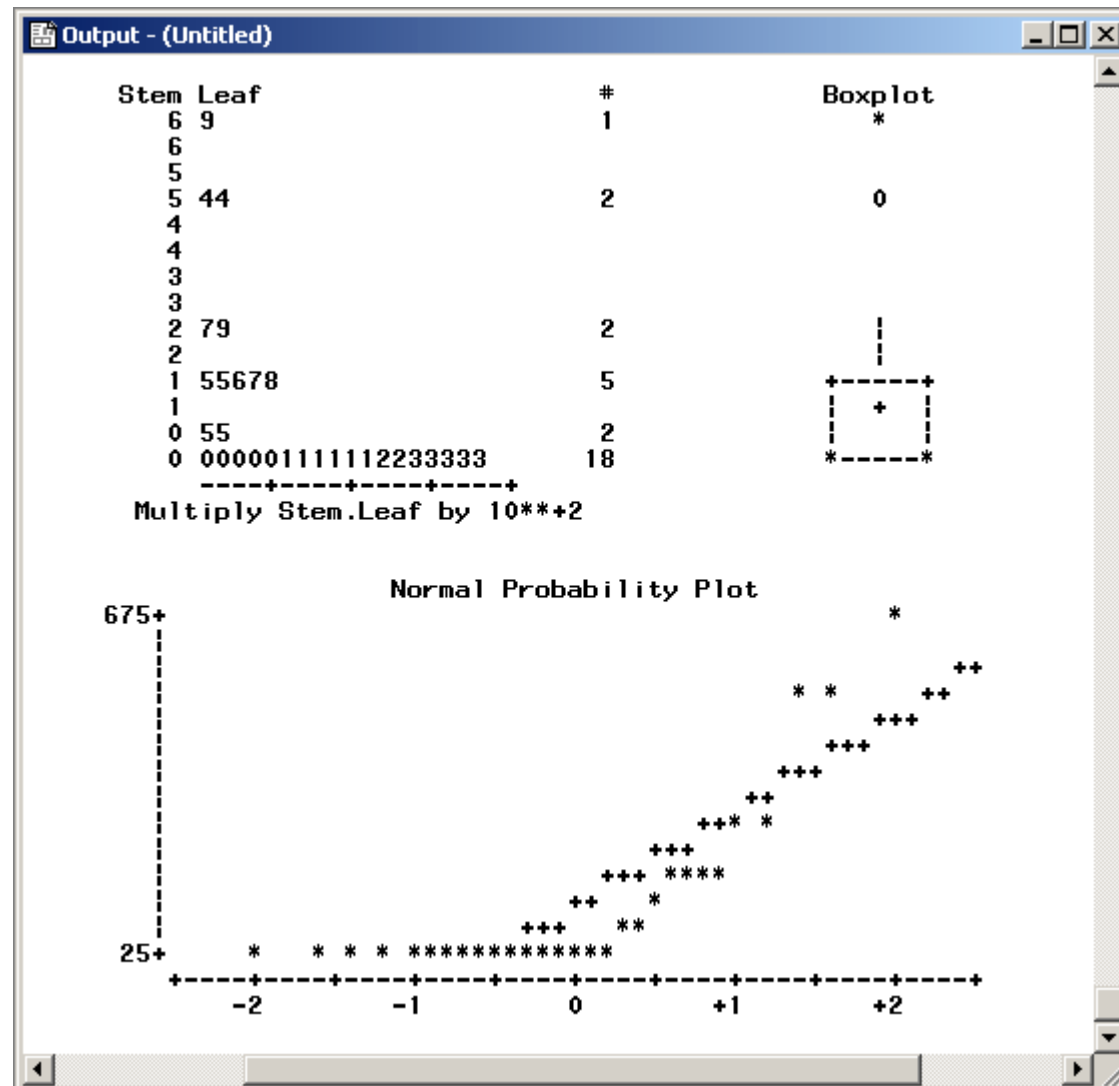
Student's t    t    3.537773    Pr > |t|    0.0014
Sign          M        15    Pr >= |M|    <.0001
Signed Rank    S    232.5    Pr >= |S|    <.0001

Quantiles (Definition 5)

Quantile      Estimate

100% Max      693.84212
99%           693.84212
95%           542.38483
90%           416.16411
75% Q3        160.51235
50% Median    31.01569
25% Q1        9.01870
10%           3.46781
5%            2.13982
1%            1.95586
0% Min        1.95586

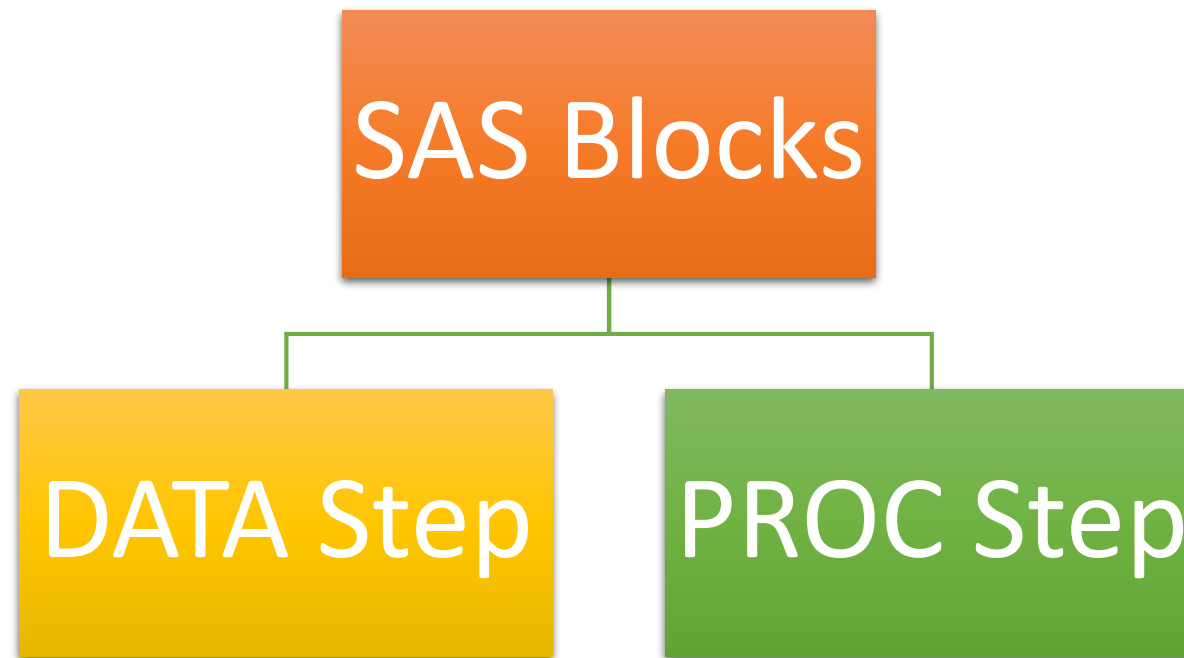
```





# Building Blocks of SAS Programming

- DATA helps to build a data set. The PROC refers to a Procedure. It processes the data.
- DATA steps are responsible to read data, Data Management Etc, on the other hand, PROC steps are used to perform utility functions, analysis of data, or print reports.



# SAS Block Diagram

When you submit a DATA step for execution, it is first compiled and then executed.

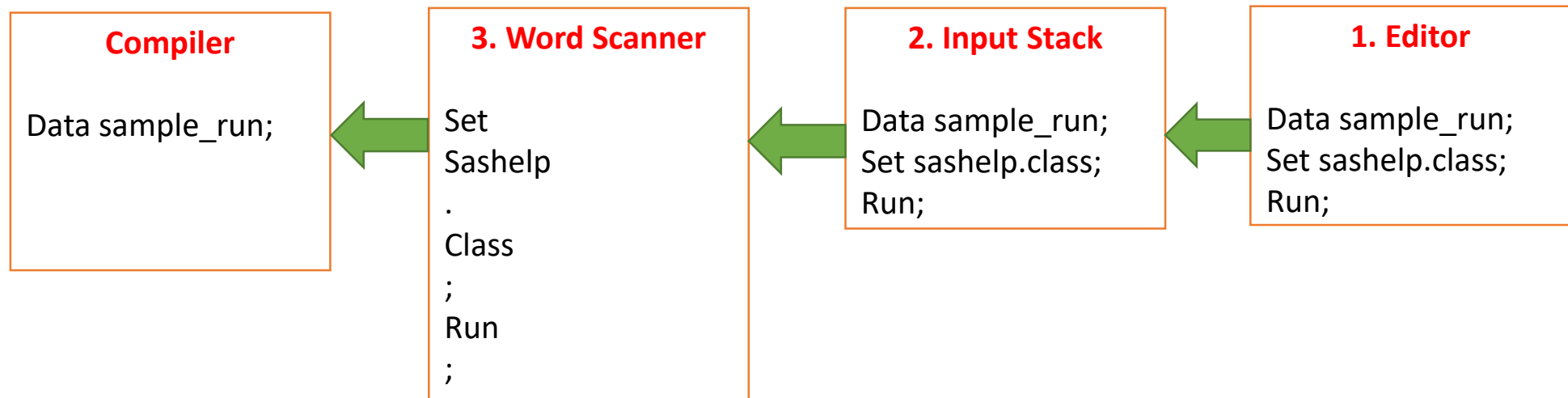
1. Compilation Phase
2. Execution Phase

## 1. Compilation Phase

- SAS checks the syntax of the SAS statements and compiles them, that is, automatically translates the statements into **machine code**.

During the compile phase, SAS creates the following three items:

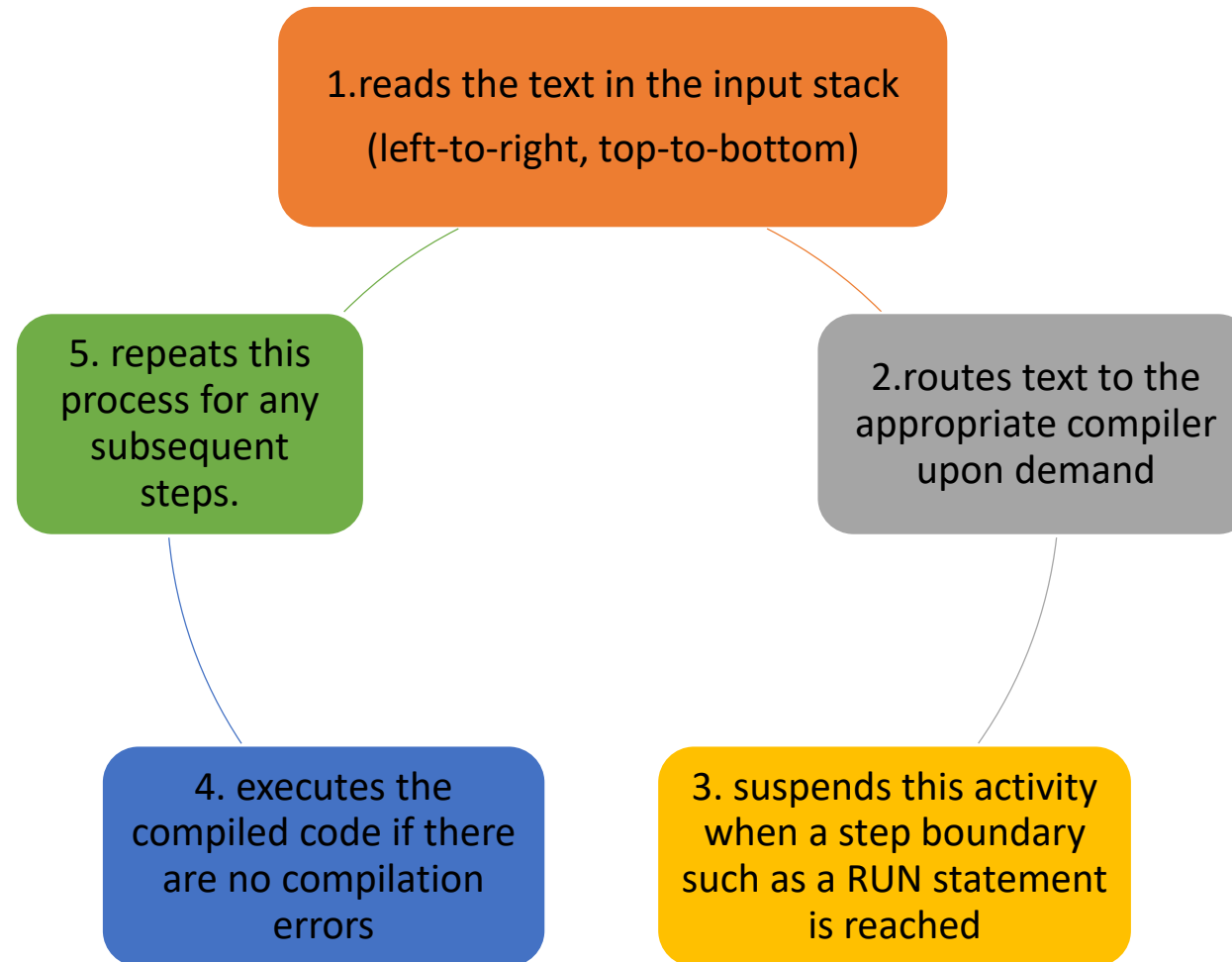
1. input buffer
2. program data vector (PDV)
3. descriptor information



# SAS Compilation Phase

## Input Stack

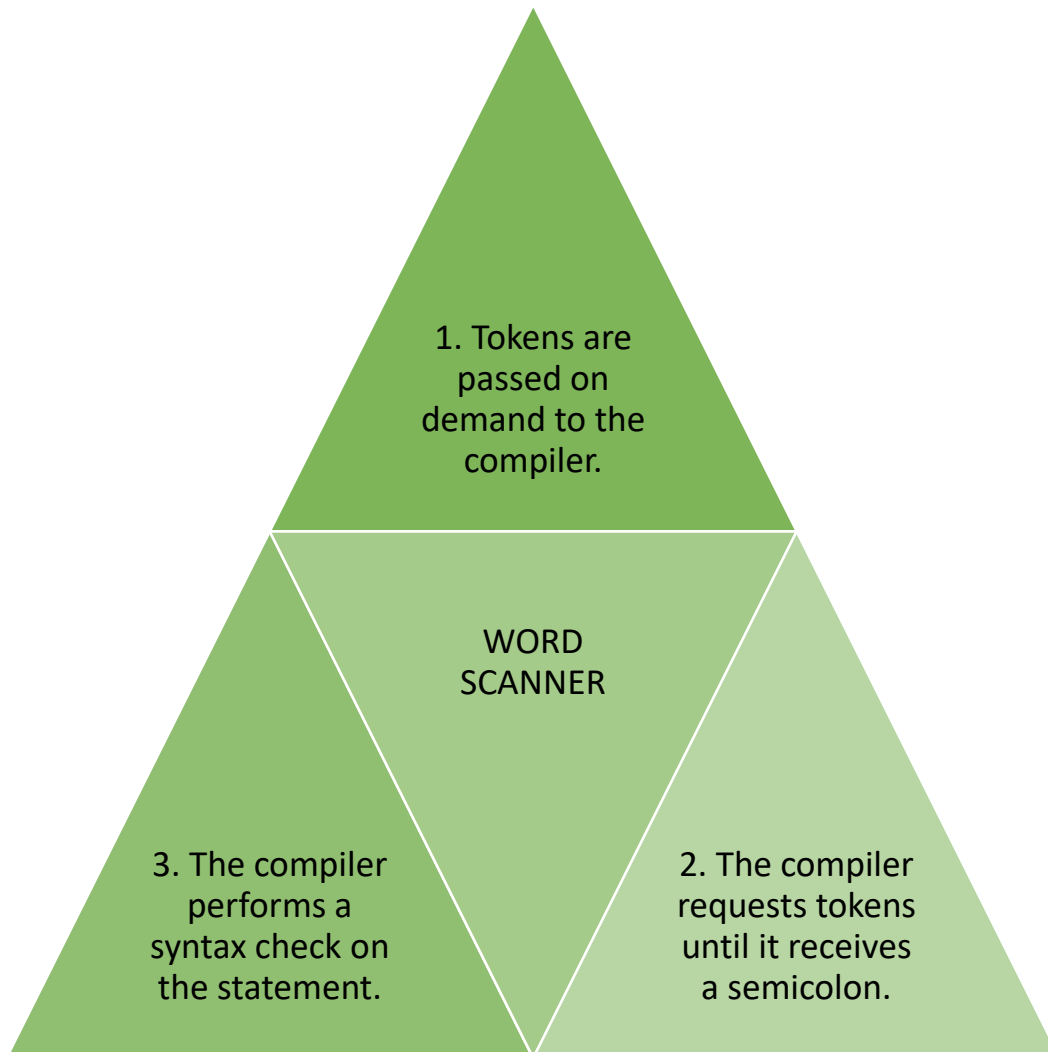
- When you submit a program, it goes to an area of memory called the input stack.



# SAS Compilation Phase

## Word Scanner

- Divides program text into fundamental units called tokens.



The word scanner recognizes four types of tokens:

**1. Literal token:**

- Examples: "Any text" 'Any text'.

**2. Number token:**

- Examples: 23 109 '01jan2002'd 5e8 42.7

**3. Name token:**

- Examples: infile \_n\_ item3 univariate dollar10.2

**4. Special token:**

- Examples: \* / + - \*\* ; \$ ( ) . & %

A token ends when the word scanner detects

- the beginning of another token
- a blank after a token.

The maximum length of any token is 32767 characters.

# SAS Compilation Phase

## Program Data Vector

- The program data vector is the area of memory where SAS builds a data set, one observation at a time. Like the term input buffer, the term program data vector refers to a logical concept.
- The program data vector contains two automatic variables that can be used for processing but
- which are not written to the data set as part of an observation.
  - `_N_` counts the number of times that the DATA step begins to execute.
  - `_ERROR_` signals the occurrence of an error that is caused by the data during execution.
    - The default value is 0, which means there is no error. When one or more errors occur, the value is set to 1.

<code>_N_</code>	<code>_ERROR_</code>	



# SAS Execution Phase

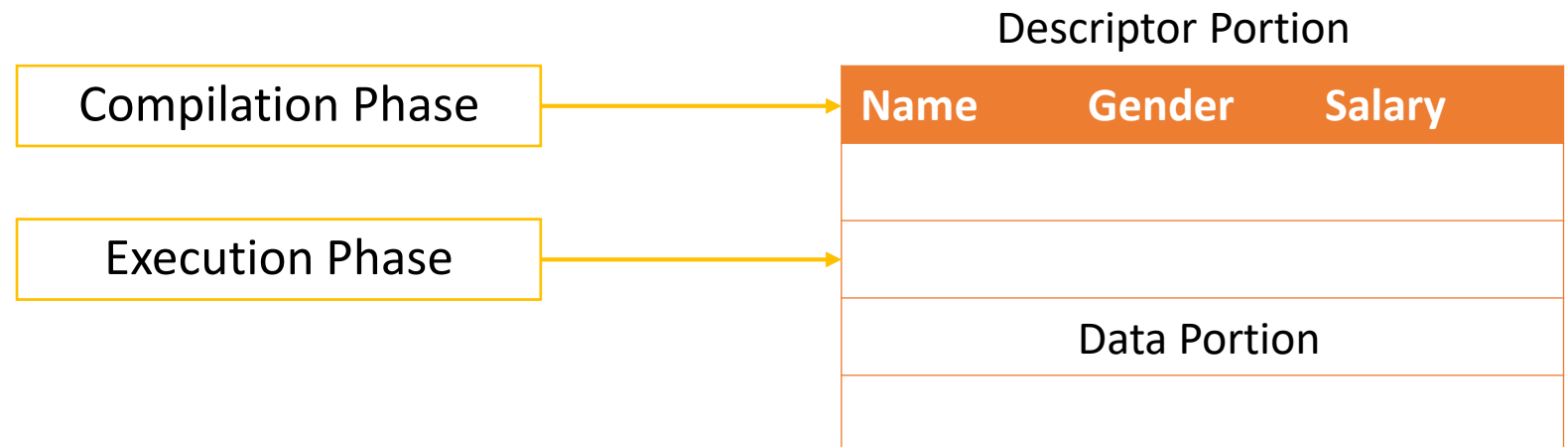
Name	Gender	Salary
Adam	M	100
Eve	F	100
Jack	M	200
Flossy	F	150

NULL Handling:

If a Numeric Data is NULL → Missing, Denoted by Period (.)

If a Character Data is NULL → BLANK

_N_	_ERROR_	Name	Gender	Salary
1	0			.



# SAS Execution Phase

Step-1

_N_	_ERROR_	Name	Gender	Salary
1	0			.

Name	Gender	Salary
Adam	M	100

Step-2

_N_	_ERROR_	Name	Gender	Salary
1	0	Adam	M	100

Step-3

_N_	_ERROR_	Name	Gender	Salary
2	0			.

Name	Gender	Salary
Adam	M	100
Eve	F	100

Step-4

_N_	_ERROR_	Name	Gender	Salary
2	0	Eve	F	100

Name	Gender	Salary
Adam	M	100
Eve	F	100
Jack	M	200
Flossy	F	150



Repeats till Data step completes

# BASIC RULES

## **SAS Statements**

- All the SAS statements end in a semicolon. This is a basic rule that differentiates a simple English statement from SAS statement.
- Not Case Sensitive
- Statements can continue on the next line.
- A statement can be on the same line as other statements.

## ➤ **Errors in SAS Programming**

- Errors in SAS occur mainly due to,
  - Missing semi-colon
  - Misspelled words.
  - Invalid variable names
  - Missing or invalid punctuation
  - Invalid options.
- We can see these errors in red color in the log window.