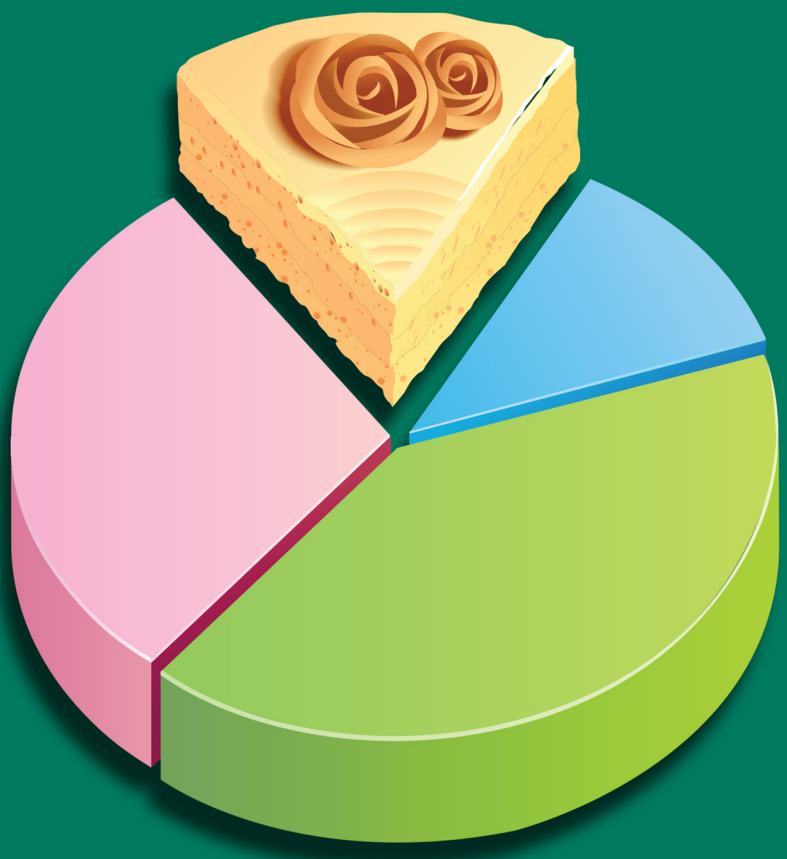
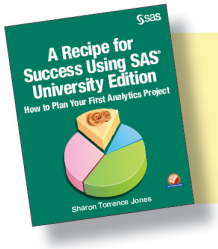
A Recipe for Success Using SAS® University Edition

How to Plan Your First Analytics Project



SAS® University Edition

Sharon Torrence Jones



From *A Recipe for Success Using SAS® University Edition.*
Full book available for purchase here.

Contents

About This Book ..... *v*
About the Author ..... *ix*
Acknowledgments ..... *xi*

Chapter 1 / The SAS Language **1**
Structure of the Book 2
Assembling Casseroles and Analyzing Data 2
SAS Software 5
Starting SAS University Edition 6
Windows and Commands in SAS University Edition 6
SAS Language 10
SAS Programs 10
SAS Statements 10
Layout of a SAS Program 13
SAS Data Sets 14
Additional Resources 22
Quick Tips 22

Chapter 2 / Planning and Executing a Successful Programming Project **23**
Successful Planning 23
Implementing a Project 33
Quick Tips 47

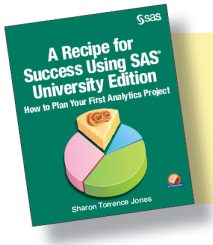
Chapter 3 / Basic Analytics **49**
Analytics and Statistics–The Essentials 50
What Are Analytics and Statistics? 51
Organizing Data 51
Tasks in SAS University Edition 54
Working with Statistics 59
Introduction to Variables and Graphs 59
Summary Statistics 66
Tests of Association 76
Quick Tips 86

Chapter 4 / Real-World Case Studies **87**
Introduction 87
Case Study Examples 90

Quick Tips 125

Index **127**

From *A Recipe for Success Using SAS® University Edition: How to Plan Your First Analytics Project* , by Sharon Torrence Jones. Copyright © 2016, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.



From *A Recipe for Success Using SAS® University Edition.*
Full book available for purchase here.

1

The SAS Language

Structure of the Book **2**

Assembling Casseroles and Analyzing Data **2**

SAS Software **5**

Starting SAS University Edition **6**

Windows and Commands in SAS University Edition **6**

Using the Windows in SAS University Edition 6

Navigation Pane 7

Work Area 8

SAS Language **10**

SAS Programs **10**

SAS Statements **10**

Layout of a SAS Program **13**

SAS Data Sets **14**

Useful Vocabulary 14

Variables and Observations 14

Browsing the Descriptor Portion 16

Understanding the Data Portion 17

Data Types 19

Missing Data 21

Additional Resources **22**

Quick Tips **22**

Structure of the Book

Do you hear the phrase, The Power of SAS®, but you really don't know what it means and what it could mean for you? Do you wonder how you can use SAS to make your life or job easier, but you do not know where to start? Well, you have come to the right place! This book is an easy guide on how to start using SAS and apply it to the real world. It is filled with helpful examples and real-life success stories of SAS users.

This book shows you:

- how to start using analytics
- how to use SAS to accomplish a goal
- how to effectively apply SAS to your community or job
- how users like you implemented SAS to solve their problems

You, too, will be able to harness The Power of SAS!

This book is broken down into easy-to-read chapters that introduce you to SAS vocabulary and structure, show you how to plan and execute a successful project, introduce you to simple statistics, and walk you through a few case studies to help inspire and motivate you to use analytics for a real project. At the end of each chapter, there are five quick tips to take away from the content learned. After reading this book, you will be ready to plan, create, and execute a project using SAS.

Assembling Casseroles and Analyzing Data

"More than just something to eat, more than just a dish on the table, casseroles serve a purpose; they elicit a response. They are not an ingredient; casseroles are a genre." Vivian Howard, A Chef's Life

I am a North Carolina girl, born and raised in Charlotte. I had a wonderful childhood split between "city" life and a more rural atmosphere, and we loved to eat! Food is a theme that runs through this book, and I often use it as a metaphor in the classroom. My grandparents and parents came from a generation that ate three meals a day at the kitchen table, which, in turn, is much how I grew up. My Papa always said that food brings fellowship, and he is correct. Food is something that we can all relate to and it gives us common ground!

Figure 1.1 *Green Bean Casserole*

"The goal of a casserole is to feed the hungry and the heartbroken; something warm and predicable hits the spot." Sheri Castle, Southern writer

Why are casseroles a genre? Put very simply, they are made to feed the hungry, and they include simple yet hearty ingredients. Casseroles take different elements to create one entity.

So what does this have to do with data? I think about data like I think about food. If you watch a cooking show, the host always follows a plan to create a dish. There is a beginning, middle, and end to the recipe process. Working with data has the same concept. When you prepare a casserole, you put all of the ingredients together in a pan and then you put it in the oven. When you work with data, your "ingredients" are variables and observations. When you create a data analytical program, you put all of the variables and observations together in the program and then you run it.

You probably have a few questions. What is data analytics? Why is data analytics important? These are a few of the questions that have probably driven you to explore SAS. The terms "data" and "analytics" are more common parts of the social vernacular, and you have likely heard the term "data-driven decisions" in different contexts. We are producing more data than ever before with our plugged-in lifestyles. As a result, questions arise about data and how it can be used. Understanding what to do with data is only half the problem, and even just starting with basic analytical practices can be a pivotal change.

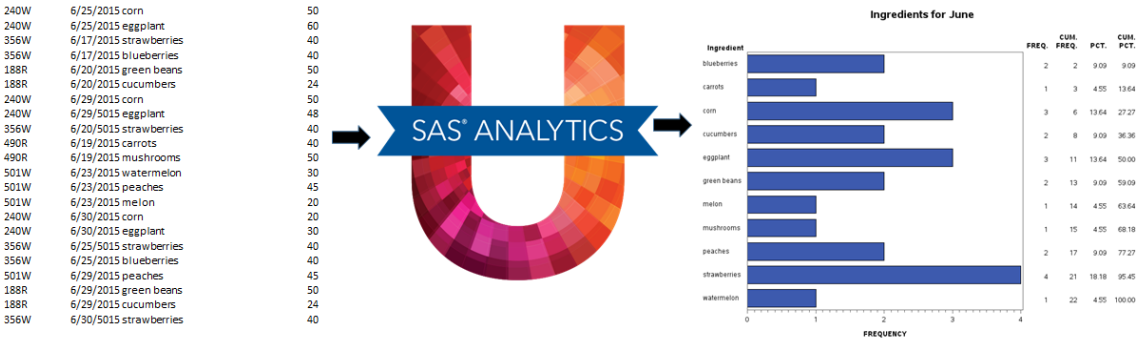
We can relate data analytics to a casserole because we understand the simplicity and the depth of a good casserole. Think about your favorite casserole. What image comes to mind? My mind immediately sees a green bean casserole. What are the key ingredients of your favorite casserole dish? Sheri Castle says that, *"If you make that casserole too "dolled" up, then you have missed the target. You may have hit another target, but your main target has*

been missed." When baking a casserole or analyzing data, the main target is to create a simple but pragmatic entity from key ingredients.

For good data analytics, we want many people to be able to use and comprehend the information that we compile. Like a casserole, we want simple ingredients to feed many people.

Figure 1.2 *Ingredients and Data*

Figure 1.3 *Data to Analysis*



Keep thinking about the concept of a casserole and let's look at data. Data is everywhere, shedding light on all aspects of life. Retailers know what's selling and who's buying. Polls test opinions on everything from political candidates to consumer goods. Doctors monitor their patients' vital signs and progress over time. Social networks register the interactions of millions. At sporting events, fans and coaches examine collected comprehensive statistics on their teams' performance. If something can be measured, then in all likelihood, a vast archive of data is already being compiled. Often data is unprocessed, waiting for someone to analyze it and discover new and valuable knowledge about the world.

So, exactly what is data analytics? Analytics can be defined in many different aspects. In short, analytics create insight. Simply put, someone wants an answer, so a question is asked. To properly phrase the question and find the answer, we look at what we know, how we know it, and what more could we know about it? With SAS, you have a tool that can refine those questions and return easy-to-understand results, and those results create insight and uncover knowledge.

Where do you get data? To start, you can do something as simple as writing down what you eat for lunch everyday and each item's calories. Or, you might be interested in the top television shows, music, or movies. If so, you can find this data on the Nielson site, download the Nielson ratings data, and upload the data to SAS. If you want to learn something about your community, you can work with your local government to research local event data or see how a non-profit delegates funding. Maybe you want to know how many Girl Scout Cookies your troop sold over the past two years, or how much popcorn the Boy Scouts sold. Or, maybe you just want to analyze your spending habits to help yourself start saving money. It is all about the data!

SAS Software

To start analyzing data and getting results, download SAS University Edition. SAS University Edition is free SAS software that can be used by anyone who wants to learn SAS! With this software, you can explore and solve important and stimulating problems. Gaining new perspective on the power of data and the insight provided through reliable analysis can open your eyes to many new possibilities and solutions to various problems. SAS University Edition provides the opportunity for you to expand your knowledge of statistics and quantitative methods by offering faster and easier access to the most up-to-date statistical methods. Writing and submitting code in SAS University Edition is easy. It has a powerful graphical user interface where there is a multitude of opportunities to run analyses.

The user interface for SAS University Edition, also known as SAS Studio, is a browser interface. It can be accessed from common Internet browsers including Safari, Firefox, Internet Explorer, and Chrome. The SAS University Edition program is robust and offers analytic capabilities from basic to advanced. Below is an overview of the key features of SAS University Edition. There are five significant features that help a user learn and use analytics with SAS. These include:

1. **Base SAS:** Make programming fast, easy, and graphical with the Base SAS programming language, ODS graphics, and reporting procedures.
2. **SAS/STAT:** Choose from a variety of statistical methods and techniques in the Tasks and Snippets sections.
3. **SAS/IML:** Use this matrix programming language for more specialized analyses and data exploration.
4. **SAS Studio:** Reduce programming time with autocomplete for hundreds of SAS statements and procedures and build in syntax.
5. **SAS/ACCESS:** Connect with data, no matter where it resides.

As you begin to work in SAS University Edition, you will find that you can use all or just a few of these features to help you find the answers that you need. In this book, we focus on understanding Base SAS and using Base SAS in SAS Studio.

Starting SAS University Edition

Now that you know why you should use SAS to learn analytics, let's get started with the software! The direct link to download SAS University Edition is:

http://www.sas.com/universityedition

For more information about how to get your software up and running and how to access resources like FAQs and video tutorials, visit the Getting Started page on the SAS support site at:

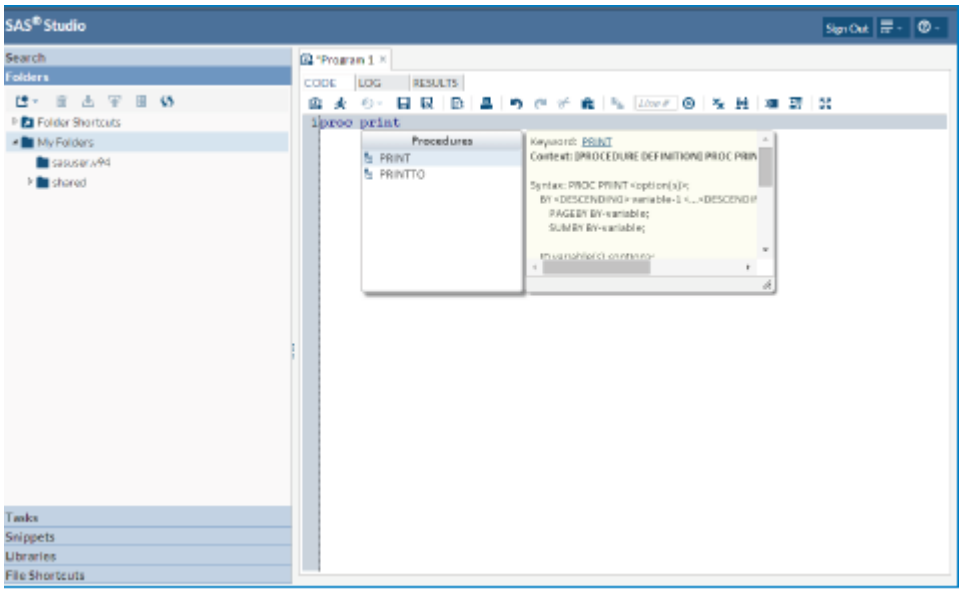
http://support.sas.com/software/products/university-edition

Windows and Commands in SAS University Edition

Using the Windows in SAS University Edition

Once you open the software, you'll see that SAS University Edition has two basic windows. The Navigation pane houses the files, folders, tasks, snippets, libraries, and file shortcuts. The Work area contains four tabs: Program, CODE, LOG, and RESULTS.

Figure 1.4 *SAS University Edition Snapshot*

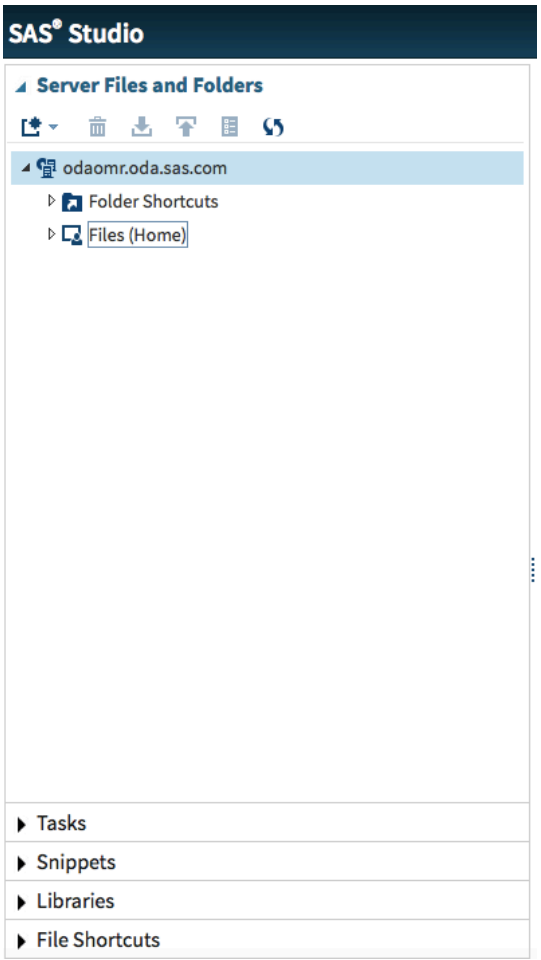


Navigation Pane

The Navigation pane gives you access to files, folders, tasks, snippets, libraries, and file shortcuts. Expand each item to show its contents.

- Server Files and Folders: Displays the content that is visible to SAS University Edition.
- Tasks: Contains menus for data manipulation, graphics, and statistical analysis.
- Snippets: Contains saved segments of SAS syntax or code that can be edited. You can also add your own snippets.
- Libraries: Contains a collection of SAS files. (SAS files are stored in libraries.)
- File Shortcuts: Provides shortcuts similar to what's in the Microsoft environment.

Figure 1.5 *Navigation Pane*



Work Area

The largest area of the screen is the Work area, where you draft a program or implement a task. The Work area has tabs that display different elements of your program or task.

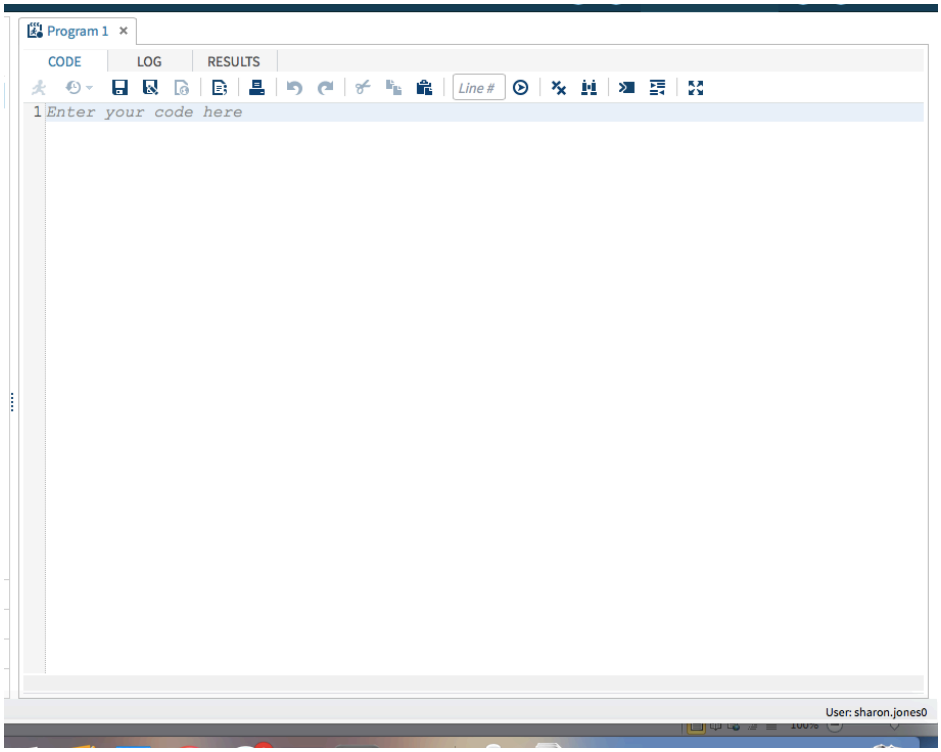
Program: This tab displays the active program.

CODE: This tab is where you type, edit, and submit your program.

LOG: This tab contains notes about your SAS session. After you submit your program, it shows any notes, errors, or warnings. These items are color-coded.

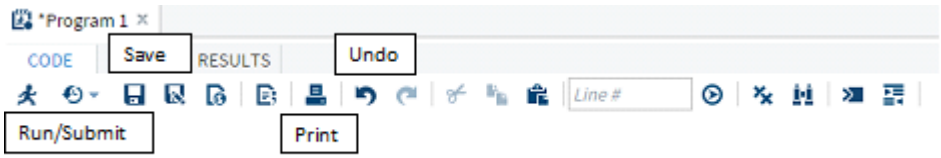
RESULTS: This tab shows your output and generates any printable results.

Figure 1.6 *Work Area*



Drop-down menus and icons on the toolbar enable you to perform a variety of tasks, including submitting your program; opening a program; saving files; printing, cutting, and pasting text; and undoing or redoing an action.

Figure 1.7 *SAS Studio Toolbar*



Now that you have an idea of the free software that you can download and use to do analytics, let's dig a little deeper into the SAS language and environment. The next few sections introduce you to the SAS language and its associated syntax.

You can find more resources about SAS University Edition through SAS free e-Learning at http://support.sas.com/edu/elearning. You can get started with SAS tutorials at http://support.sas.com/training/tutorial.

SAS Language

The SAS language is a computer programming language used for statistical analysis. SAS software can read data from common spreadsheets or databases. Or, you can create your own data in SAS and output it to tables and graphs. Most software applications are menu-driven or command-driven. This means that either a menu drives your coding or commands that you enter get results. SAS is actually neither. SAS uses statements to write a series of instructions called a program. The program communicates what you want to do and is written using SAS language syntax. If you want, you can do menu-driven coding in SAS University Edition. However, you have more flexibility and get more control over your code if you learn how to write your own program. Understanding how syntax is written and executed helps you better manipulate your data.

SAS Programs

A SAS program is a sequence of statements that execute one step at a time, statement by statement, independently of one another. Whereas other programming languages compile an entire program and then completely execute it, SAS reads each line of information separately. To better imagine how a SAS program works, think about your favorite restaurant. You enter the restaurant, take a seat, and then when the waiter or waitress arrives, you say what you would like to eat. When you state your order, think of the statements like a program:

```
I would like to place my order for my meal. I would like a glass of
water. I would like to order the chicken sandwich and chips. I would
also like a napkin.
```

Note that you first say what you want to do. You add subsequent statements that further detail your request. For example, in a program, you might want to know how much money was made in a quarter, or what was the percentage of individuals participating in a sport? Within a SAS program, SAS statements are what make the program run.

SAS Statements

SAS programs are made up of steps, and steps are made up of statements. Remember, you are learning a new language, and when you create a SAS program, as with any

programming language, there are rules to follow. Your SAS program is constructed of DATA and PROC steps. *DATA steps* are typically used to create SAS data sets. *PROC steps* (PROCedures) are used to process SAS data sets (or, in other words, to generate reports and sort data). The most important rule is:

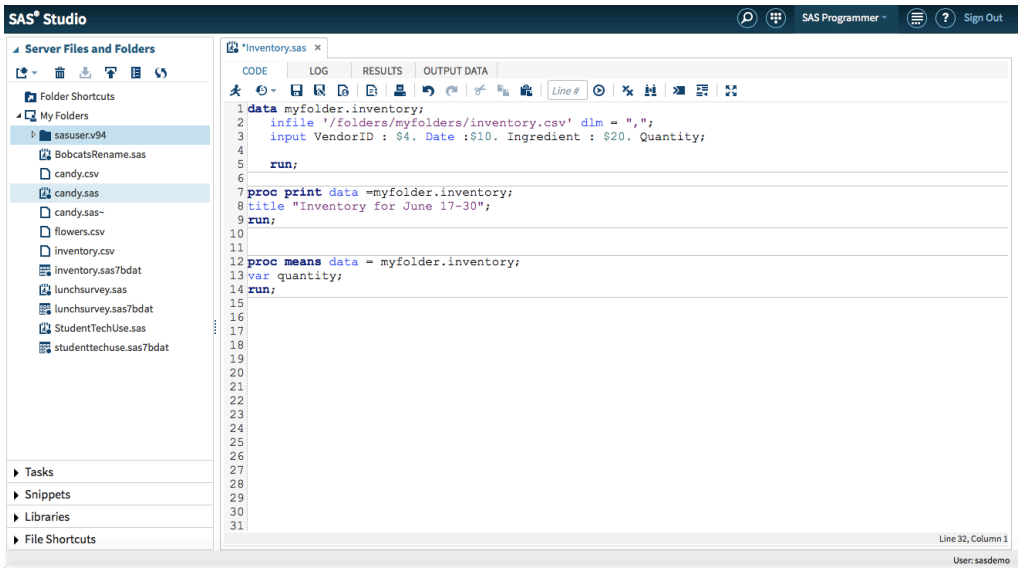
Every SAS statement ends with a semicolon.

Although this might seem simple, even the most experienced programmer will forget a semicolon! In the following programming example, PROC PRINT is used to display the data in a table format on the RESULTS tab. The data equals (data =) option references the data set that you would like to see.

```
PROC PRINT data = myfolder.inventory;
Run;
```

Figure 1.8 on page 11 shows a program in SAS Studio. Notice that each statement ends with a semicolon.

Figure 1.8 *SAS Coding Environment*



The DATA step portion creates a permanent or temporary SAS data set from the raw data or from another SAS data set. SAS procedures can be run only on SAS data sets. A temporary SAS data set is a data set that is created in the program and deleted when the SAS Studio session ends. A permanent SAS data set is useful once you have gotten your data in the format that you need and you expect to be working with it repeatedly.

As you are programming, you will be working on three main tabs of the graphical user interface:

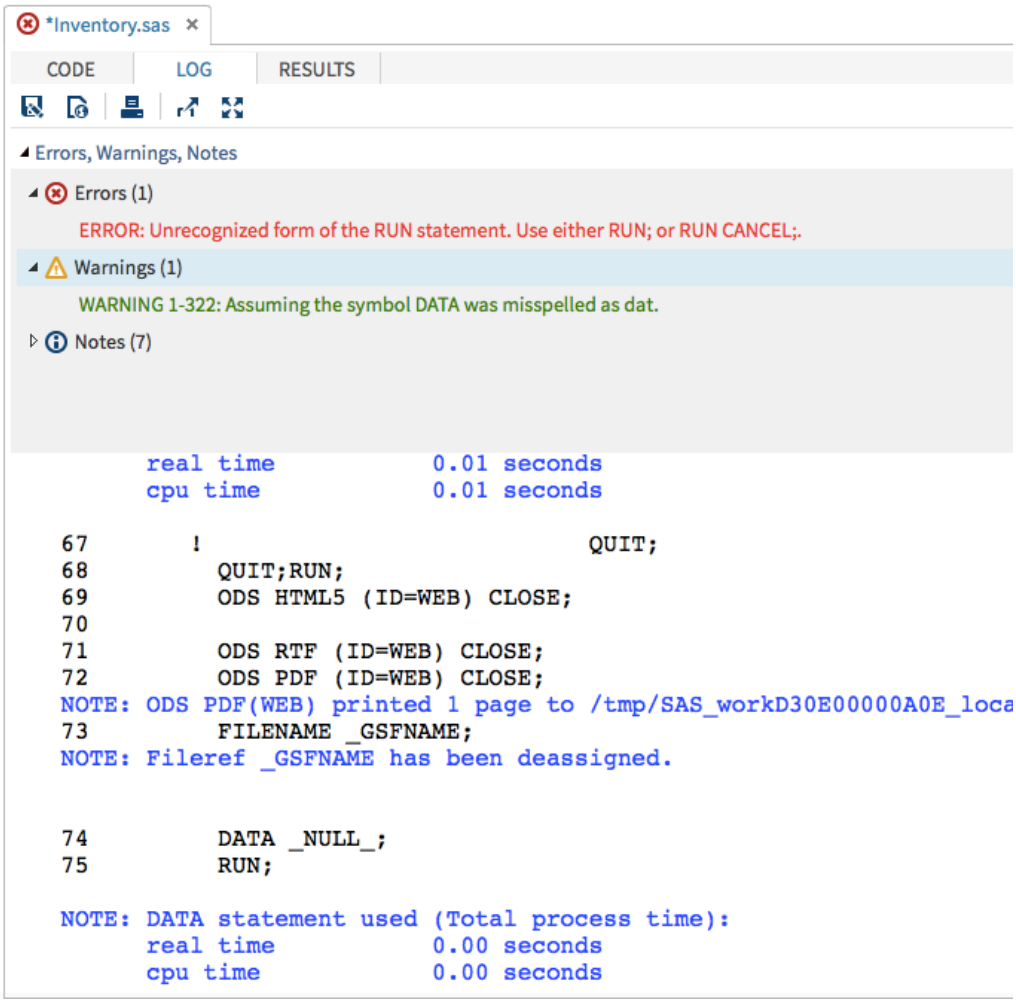
- CODE tab – where you write your code
- LOG tab – where a transcript of your code is shown

- RESULTS tab – where code results are shown

SAS has color-coding to help you spot errors as you are programming. There is a predominant blue color scheme. Dark blue is used for DATA and PROC steps. Light blue is used for statements within your steps. And, purple is used for quotations.

When you run your SAS program, the log shows the processing that occurred and also uses colors. The log uses three main colors: red, blue, and green. Errors are in red, warnings are in green, and notes are in blue. The color-coding in SAS helps you see whether there was a problem, where the problem occurred, and how to fix the problem.

Figure 1.9 *SAS Log Color-Coding*



Layout of a SAS Program

When creating a SAS program, the layout is flexible. It is good programming practice, however, to format your program in a neat-looking fashion so that it is easy to read and execute. Remember that a SAS program is made up of SAS statements and SAS statements analyze data. The SAS program is the shell that holds the statements. You should avoid writing code that is difficult to read. Here is an example of good layout:

```
Data myfolder.inventory;
    infile "/folders/myfolders/garden.dat";
    input vegetables $ fruit $;
Run;

PROC PRINT data=myfolder.inventory;
Run;
```

In the code, there are several SAS statements, each ending with a semicolon. In addition to the code, you can add comments. Comments make your program more understandable, and you can insert comments anywhere in the program that you like. SAS does not read comments as code. Comments are usually used to annotate your program, making it easier for someone to read and understand what you have done and why you have done it. The syntax for a comment is /* to initialize and */ to close. When you are programming in SAS, your comment code is color-coded green.

Here is an example of how to comment your code:

```
/*Creating a data set from the raw data file garden.dat*/

Data myfolder.inventory;
    infile "/folders/myfolders/garden.dat";
    input vegetables $ fruit $;
Run;

PROC PRINT data=myfolder.inventory; /*Printing the data in the created
    data set myfolder.inventory*/
Run;
```

TIP When you are first learning a programming language, it can be frustrating. It is much like learning a foreign language with all the rules and exceptions. The first time that you code a program, it will probably not run. That is okay! We all learn from our mistakes. Sometimes we can get caught up in the idea that it has to run perfectly the first time we write the code, but that is just not realistic. Many times it is because you forgot a semicolon. (Tradeoff: You get to join the Semicolon Club! I have been a member for many years!) When you are first learning the programming language, write simple programs. Do not try to tackle complicated multi-step programs until you

are comfortable. Always check your log after your program runs to confirm the results. As you get into that habit, you will increase your efficiency and begin to have more confidence in your coding. Remember, build a program piece by piece. This allows for easy-to-maintain code that is manageable should a problem arise.

SAS Data Sets

Useful Vocabulary

To run analysis on data that you have collected, the data must be in a format that SAS can read. Therefore, your data must be in the form of a SAS data set. A SAS data set is a specially structured file that contains data values. SAS data sets are processed by SAS procedures. SAS has a robust ability to work with and read almost any data. Once the observations or files are in SAS, the data is stored and can be opened whenever needed to process analyses and reports. A few vocabulary words that are associated with a data set include:

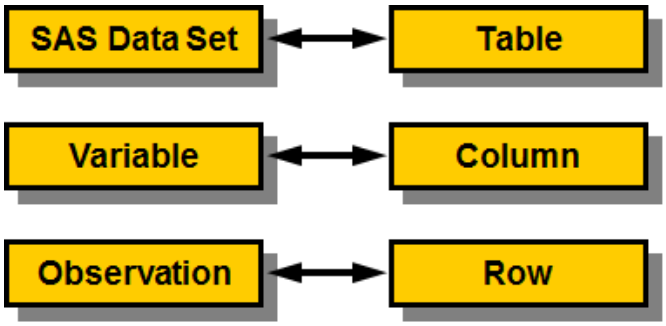
- variables
- observations
- descriptor portion
- data portion
- data types
  - numeric variables
  - character variables
- missing data

Variables and Observations

To better understand a SAS data set, you must understand SAS terminology. Data is the primary component of the data set, and every language has terminology that is associated with that data. In SAS, the terminology is:

- SAS data sets are also called tables.
- Observations are also called rows.
- Variables are also called columns.

Figure 1.10 *SAS Data Set Terminology*

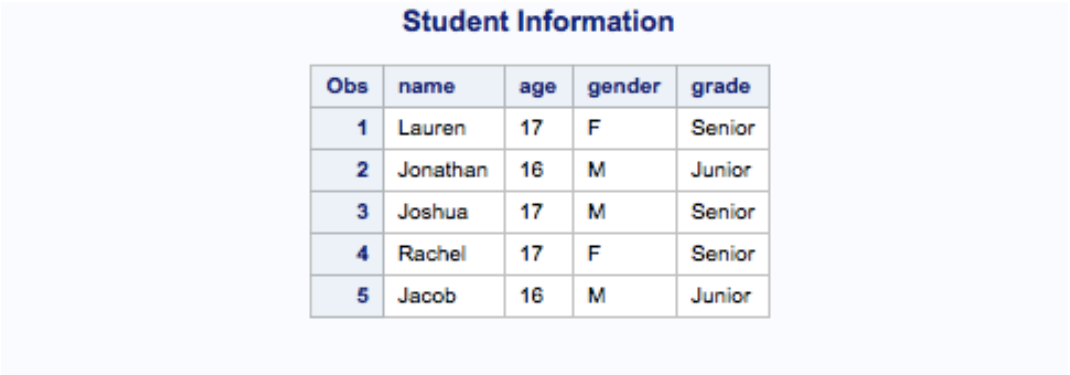


In Table 1.1 on page 15, you see how variables are columns and observations are rows. The abbreviation “Obs” stands for observations.

Table 1.1 *Student Data Set, Variables (Columns) and Observations (Rows)*

Obs	Name	Age	Gender	Grade
1	Lauren	17	F	Senior
2	Jonathan	16	M	Junior
3	Joshua	17	M	Senior
4	Rachel	17	F	Senior
5	Jacob	16	M	Junior

Figure 1.11 *Output in SAS*



Student Information

Obs	name	age	gender	grade
1	Lauren	17	F	Senior
2	Jonathan	16	M	Junior
3	Joshua	17	M	Senior
4	Rachel	17	F	Senior
5	Jacob	16	M	Junior

To further explore SAS terminology, let's look at the descriptor portion and the data portion of a SAS data set.

Browsing the Descriptor Portion

The descriptor portion of a SAS data set simply *describes* the data. It contains general information about the SAS data set such as the data set name and number of observations. This portion of the data set includes variable attributes, such as variable name, type (character or numeric), length, format, and label. You can see the descriptor portion of the data set using the CONTENTS procedure (PROC CONTENTS).

```
PROC CONTENTS data = myfolder.inventory;
Run;
```

Figure 1.12 *PROC CONTENTS*

The CONTENTS Procedure

Data Set Name	WORK.INVENTORY	Observations	22
Member Type	DATA	Variables	4
Engine	V9	Indexes	0
Created	11/02/2015 12:51:58	Observation Length	48
Last Modified	11/02/2015 12:51:58	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	2722
Obs in First Data Page	22
Number of Data Set Repairs	0
Filename	/saswork/SAS_work27450000F02F_odaws03-prod-us/SAS_work14060000F02F_odaws03-prod-us/inventory.sas7bdat
Release Created	9.0401M3
Host Created	Linux
Inode Number	1048593
Access Permission	rw-rw-r--
Owner Name	sharon.jones0
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

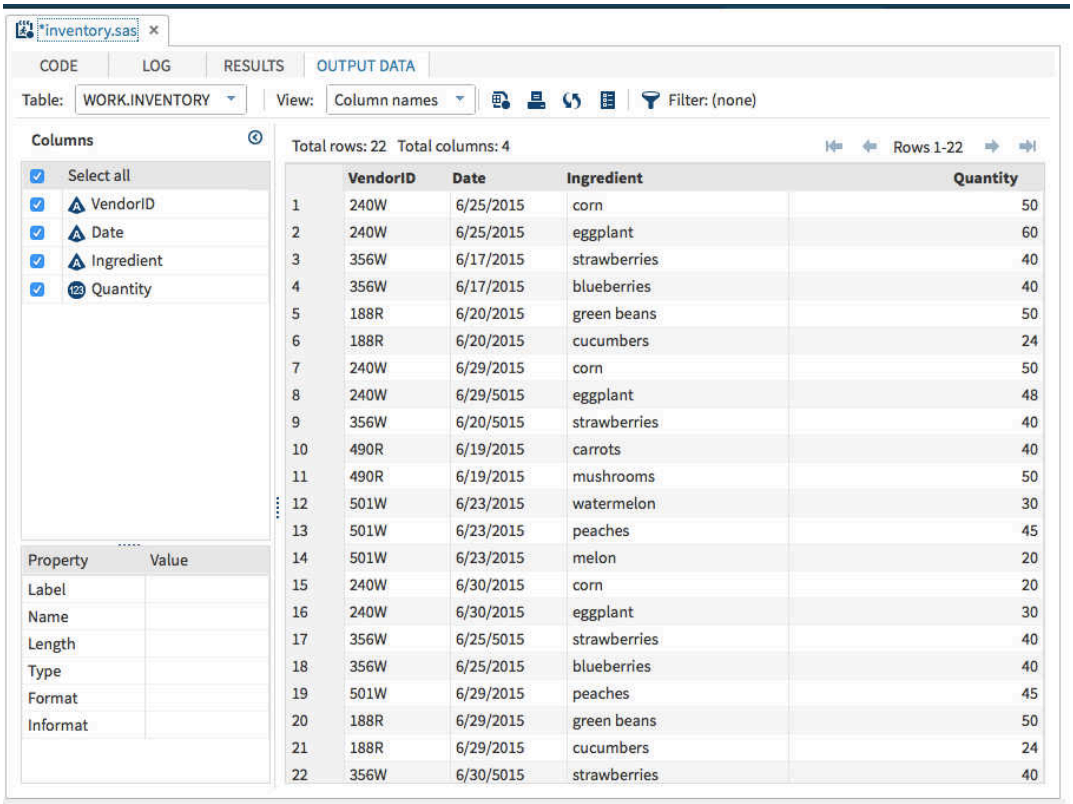
#	Variable	Type	Len
2	Date	Char	10
3	Ingredient	Char	20
4	Quantity	Num	8
1	VendorID	Char	4

The descriptor portion of your data set is extremely valuable as you become familiar with the data. It displays the variable attributes that help you better understand your data and how you would like to proceed with analyses.

Understanding the Data Portion

Once your data has become a SAS data set, you can see the listing of the observations in the Output Data window.

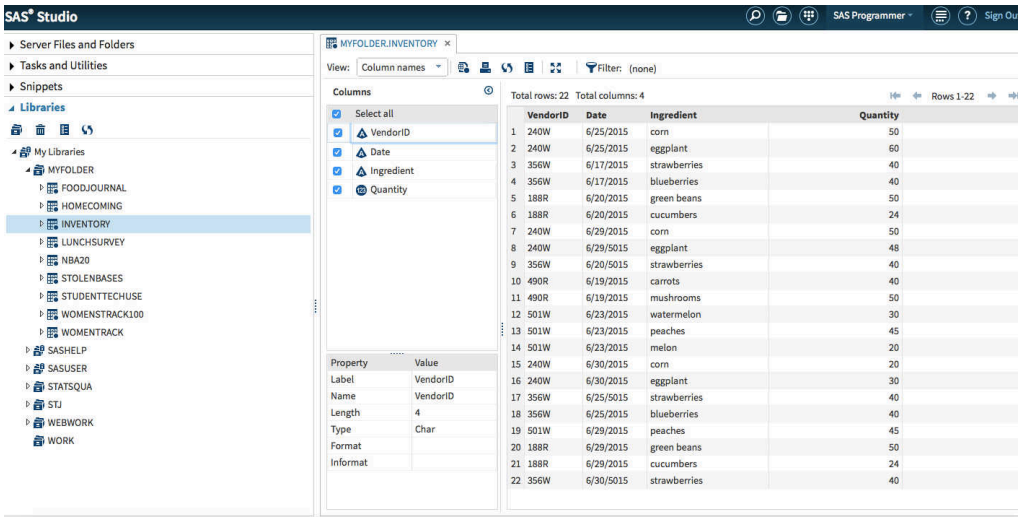
Figure 1.13 *Seeing Data in the Output Tab*



	VendorID	Date	Ingredient	Quantity
1	240W	6/25/2015	corn	50
2	240W	6/25/2015	eggplant	60
3	356W	6/17/2015	strawberries	40
4	356W	6/17/2015	blueberries	40
5	188R	6/20/2015	green beans	50
6	188R	6/20/2015	cucumbers	24
7	240W	6/29/2015	corn	50
8	240W	6/29/5015	eggplant	48
9	356W	6/20/5015	strawberries	40
10	490R	6/19/2015	carrots	40
11	490R	6/19/2015	mushrooms	50
12	501W	6/23/2015	watermelon	30
13	501W	6/23/2015	peaches	45
14	501W	6/23/2015	melon	20
15	240W	6/30/2015	corn	20
16	240W	6/30/2015	eggplant	30
17	356W	6/25/5015	strawberries	40
18	356W	6/25/2015	blueberries	40
19	501W	6/29/2015	peaches	45
20	188R	6/29/2015	green beans	50
21	188R	6/29/2015	cucumbers	24
22	356W	6/30/5015	strawberries	40

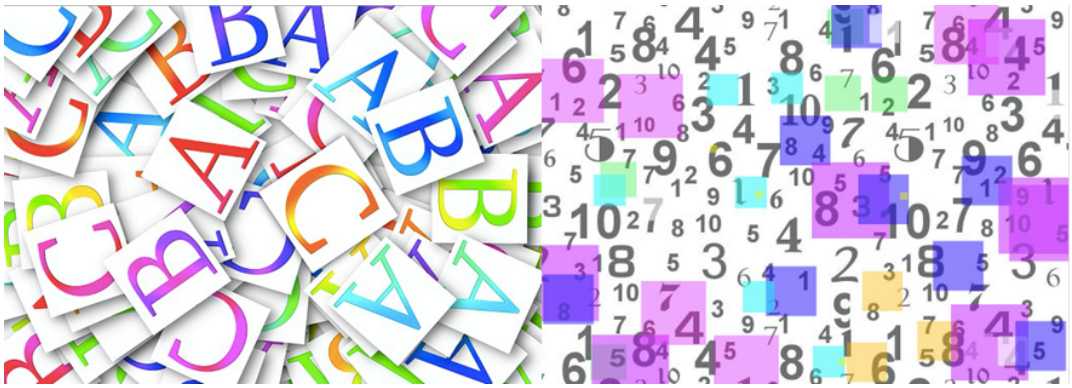
The Output Data tab holds the data. You can open the table and view the data in SAS. To view the table, click **Libraries** in the Navigation pane, click a folder, and then click a data set. Figure 1.14 on page 19 shows these steps. The data portion is the actual list data; the descriptor portion describes the data.

Figure 1.14 *Viewing the Data Table in SAS Studio*



Data Types

Figure 1.15 *Character and Numeric Variables*

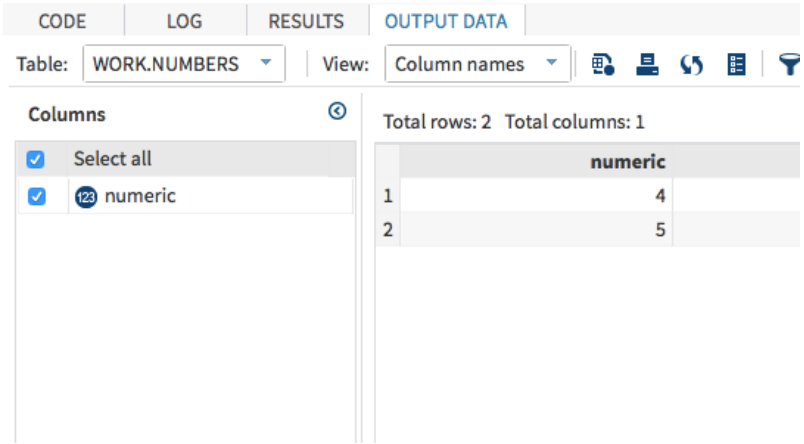


In SAS, there are two data types—numeric and character. Statisticians often refer to numeric variables as quantitative variables and character variables as qualitative variables. Numeric or quantitative variables are numbers that can be positive or negative and added or subtracted. In addition, they can be decimals. Character or qualitative variables contain any value including letters, numbers, special characters, and blanks.

A numeric variable can have a plus (+), minus (-), decimal (.), or E for scientific notation. These are all valid numeric options. The default storage or length of a numeric variable is 8

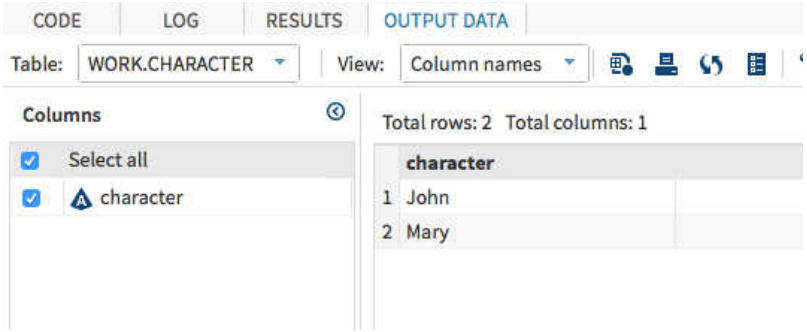
bytes, but it is not restricted to 8 digits. The value of a numeric variable is right-aligned when it is displayed in output.

Figure 1.16 *Right-Aligned Numeric or Quantitative Variables Output in SAS*



A character variable can have any value including letters, numbers, special characters, and blanks. If a variable contains letters or special characters, then it must be a character variable. If a variable contains only numbers, then it can be either numeric or character, depending on the use of the variable in the data set. Some numbers might make more sense as a character variable, like a phone number or employee ID, because you would not use mathematics on these variables. A character variable is left aligned.

Figure 1.17 *Left-Aligned Character or Qualitative Variables Output in SAS*



Variables are the columns of the data set and the pivotal part of the analysis. When naming a SAS variable, there are a few rules to keep in mind:

- A name must be 32 characters or fewer in length.
- A name must start with a letter or an underscore. It cannot begin with a numeral or contain special characters.

Names should be relevant and simple. When working with variables in your SAS data set, it is much easier to manipulate and analyze variables that have simple names. For example:

Favorite Color = Favcolor

Last Name = LastName

Simple relevant names make programming easier. Later on, you can label your variable to be more complex for a presentation.

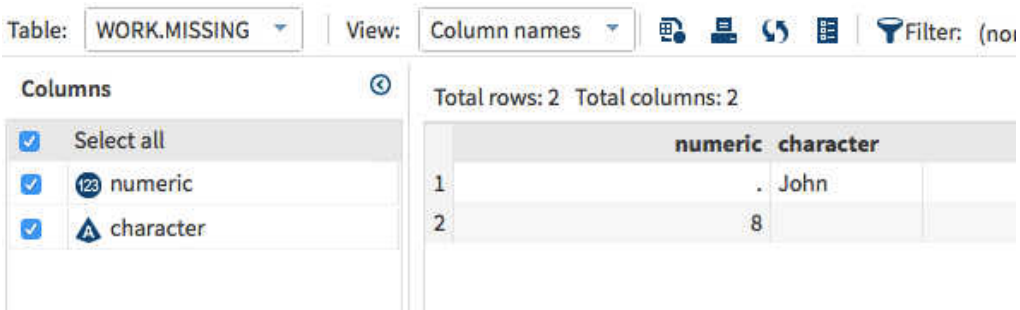
Missing Data

When working with data, it is often common to have missing values. Missing data can still be read in SAS, though. A missing value is noted with a period (.) for numeric data and with a blank space for character data. In Table 1.2 on page 21, the input data did not include an age for John, and there was no value for Name in the second row. The period in the first observation represents a missing numeric value, and the blank space in the second row represents a missing character value.

Table 1.2 *Missing Data in Numeric and Character Variables*

Obs	Age	Name
1	.	John
2	8	

Figure 1.18 *Missing Data Output in SAS*



Don’t worry, missing data is a normal part of any data project and it can be accounted for in the SAS system.

Additional Resources

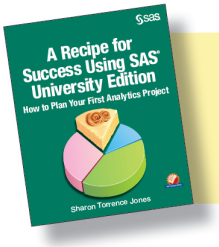
As you begin working in SAS University Edition, you might have questions or thoughts to share. SAS has an online community that can help answer those questions and connect you with other programmers! You can visit the SAS Analytics U site at http://communities.sas.com/t5/SAS-Analytics-U/bd-p/sas_analytics_u and post questions, review other questions asked, or just chat with other users. It is a great resource and is always available!

If you prefer a more hands-on option or a book, Ron Cody's *An Introduction to SAS University Edition* is very helpful.

Quick Tips

1. A SAS program contains one or more steps, and each step is a sequence of statements.
2. Every SAS statement ends with a semicolon.
3. Your SAS program is constructed of DATA and PROC steps.
4. In SAS, there are two data types or variable categories—numeric (quantitative), or character (qualitative).
5. Missing data can be read in SAS. A missing value is noted with a period (.) for numeric data and with a blank space for character data.

From *A Recipe for Success Using SAS® University Edition: How to Plan Your First Analytics Project* , by Sharon Torrence Jones. Copyright © 2016, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.



From *A Recipe for Success Using SAS® University Edition.*
Full book available for purchase here.

Index

A

A Chef's Life (Howard) 50
A Chef's Life (website) 101
A Chef's Life and North Carolina Agriculture case study 101
accessing data 35
adjusted R-square 84
An Introduction to SAS University Edition(Cody) 22
analysis of variance (ANOVA) 85
Analysis of Variance table 84
analysis variables 56
analytics
  about 50
  defined 86
  running 54
analyzing data 2, 46
ANOVA (analysis of variance) 85
ANOVA procedure 85
assigning variables 57
association, tests of 76

B

bar charts 62
Base SAS 5
Basic Statistical Measures table 70
blank space 21, 22
box plot 64

C

case studies
  A Chef's Life and North Carolina Agriculture 101
  about 87
  Color Flower Project 95
  examples of 90
  Food Journal Project 97
  Homecoming Project 90
  New Lunch Survey Project 109
  project setup 90
  Shuttle Stops and the Magnet Program 115
  StatSquad: Sports by the Numbers 124
  Student Technology Use 104
  The Analytics behind an NBA Name Change 118
  World Statistics Day 92
casserole analogy 2
Castle, Sheri 3, 23
character data types 19, 20, 21, 22
character variables 14, 60
Charlotte Bobcats 118
Charlotte Hornets 121
CLASS statement 80, 85, 86
classification variables 56
CLM (confidence limit for the mean) 73
CODE tab 6, 8, 11
Cody, Ron
  An Introduction to SAS University Edition 22
coefficient of variation 70
colon (:) 42
Color Flower Project case study 95
color-coding 12
Column data 41
columns 14
comma-separated values (CSV) file 91
comments 13, 32

Communicate step, in planning process 24, 32
confidence interval 73
confidence limit for the mean (CLM) 73
CONTENTS procedure 16, 28
continuous variables 56, 59
converting
  raw data files 43
  raw data into data sets using snippets 45
CORR procedure 28, 80, 86
correlations 80
CSV (comma-separated values) file 91

D

data
  accessing 35
  analyzing 2, 46
  getting 33
  organizing 51
  qualitative 54
  quantitative 53
data analytics 3, 4
data files, types of 41
data portion 14, 17
data sets, SAS 14, 35, 40, 45
DATA statement 40, 41, 102, 125
DATA steps 10, 11, 22, 42
data storytelling 125
DATA tab 56
data types
  about 14
  character 19, 20, 21, 22
  numeric 19, 21, 22
DATAFILE= option 45
DBMS statement 46
DBMS= option 45
decimal (.) notation 19
Delimited data 41
dependent variables 56, 61
Der, Geoff
  Essential Statistics Using SAS University Edition 86
descriptor portion 14, 16
Determine Applicable Code step, in planning process 24, 27
Develop and Brainstorm step, in planning process 24, 25
discrete variable 59
dollar sign ($) 42
drop-down menus 9

E

e-Learning 9
error-checking 12
Essential Statistics Using SAS University Edition (Der and Everitt) 86
Everitt, Brian S.
  Essential Statistics Using SAS University Edition 86
examples, of case studies 90
Extreme Observations table 71

F

File Shortcuts (Navigation pane) 7
Food Journal Project case study 97
Foormatted data 41
forward slash 46
FREQ procedure 28, 29, 47, 66, 74, 86, 99

G

Google form 34
graphs
  about 61
  bar charts 62
  box plot 64
  defined 86
  histograms 63
  scatterplot 65

H

Have Good Coding Habits step, in planning process 24, 31
Helmenstine, Anne Marie 95
histograms 63
Homecoming Project case study 90
Howard, Vivian 101
  A Chef's Life 50

I

icons 9
Identify Deliverables step, in planning process 24, 27
IF-THEN/ELSE statement 100
implementing projects 33
Import Data menu, converting and importing raw data files using 43
IMPORT procedure 35, 43, 45, 46, 47
IMPORT statement 46
Import XLSX File snippet 45
importing raw data files 43
independent variable 60
INFILE statement 40, 41, 42
INPUT statement 40, 41, 42
interquartile range (IQR) 71

K

Knight, Ben 101
kurtosis 70

L

language, SAS 10
layout, of SAS programs 13
libraries
  Navigation pane 7
  SAS 36
library reference (libref) 36
Location column, Basic Statistical Measures table 70
log 12
LOG tab 6, 8, 11

M

Make Supporting Plans step, in planning process 24, 27
MAX option, MEANS procedure 72
Maximum 73
mean 69, 70
MEANS procedure 28, 29, 47, 57, 66, 71, 77, 99
median 70
MIN option, MEANS procedure 72
Minimum 73
minus (-) notation 19
missing data 14, 21
mode 70
MODEL statement 85, 86
Moments table 69
Myfolders folder 37

N

N 73
N Miss 73
N option, MEANS procedure 72
Navigation pane 7, 54
New Lunch Survey Project case study 109
NMISS option, MEANS procedure 72
NOCUM option, FREQ procedure 75
nominal variable 60
North Carolina Agriculture Overview (website) 102
North Carolina Science Teachers Association PDI Conference 123
numeric data types 19, 21, 22
numeric variables 14

O

observations 14
One-Way ANOVA 85
online surveys 34
OPTIONS tab 57
ordinal variable 60
organizing data 51
OUT statement 46
OUTPUT tab 57, 58

P

p-value 82, 84
Parameter Estimates table 84
Pearson Correlation Coefficients table 81
period (.) 21, 22
Phillip O. Berry Academy of Technology 115
Plan a Schedule step, in planning process 24, 27
planning programming projects 23, 47
plus (+) notation 19
presentations, techniques for 47
PRINT procedure 11, 28, 33, 46, 47, 99
PROC statements 125
PROC steps 10, 22, 28
procedures
  ANOVA 85
  CONTENTS 16, 28
  CORR 28, 80, 86
  FREQ 28, 29, 47, 66, 74, 86, 99
  IMPORT 35, 43, 45, 46, 47
  MEANS 28, 29, 47, 57, 66, 71, 77, 99
  PRINT 11, 28, 33, 46, 47, 99
  REG 82
  REPORT 121
  SORT 28
  TTEST 77
  UNIVARIATE 66, 67, 71, 86
Program tab (Work area) 8
programming projects, planning 23, 47
programs, SAS 10
projects, implementing 33

Q

QRANGE option, MEANS procedure 72
qualitative data 54
qualitative variables 60
Quantiles table 71
quantitative data 53
quartile range 73

R

R-square 84
range 71, 73
RANGE option, MEANS procedure 72
raw data files
  about 35
  converting 43
  converting into data sets using snippets 45
  creating SAS data sets from 40
  importing 43
REG procedure 82
regression 82
rehearsing presentations 47
REPLACE statement 46
REPORT procedure 121
RESULTS tab 6, 8, 12
root mean square 84
rows 14
rule of three 47
RUN statement 46
running analytics 54

S

SAS Analytics U (website) 22
SAS data sets 14, 35, 40
SAS Global Forum 2015 123
SAS language 10
SAS libraries 36
SAS programs 10, 13
SAS statements 10, 13, 22

SAS Studio 5, 9
SAS University Edition
  about 5
  starting 6
  tasks in 54
SAS Work library 36
SAS/ACCESS 5
SAS/IML 5
SAS/STAT 5
scatterplot 65
scientific (E) notation 19
scope statement 27
selecting tasks 56
semicolon (;) 11, 13, 22
Server Files and Folders (Navigation pane) 7
Set Objectives and Goals step, in planning process 24, 26
shared folders, creating permament libraries from 38
Shuttle Stops and the Magnet Program case study 115
skewness 70
Snippets (Navigation pane) 7
snippets, converting raw data into data sets using 45
software, SAS 5
SORT procedure 28
standard deviation 69, 70
standard error of the mean 70, 84
statements
  CLASS 80, 85, 86
  DATA 40, 41, 102, 125
  DBMS 46
  IF-THEN/ELSE 100
  IMPORT 46
  INFILE 40, 41, 42
  INPUT 40, 41, 42
  MODEL 85, 86
  OUT 46
  PROC 125
  REPLACE 46
  RUN 46
  SAS 10, 13, 22
  scope 27
  TABLES 74
  VAR 73, 77, 80, 81
  WITH 81
statistics
  about 50
  summary 66
  working with 59
StatSquad: Sports by the Numbers case study 124
steps 22
storytelling 47, 48
Student Technology Use case study 104
sum of observations 70
sum of weights 70
summary statistics 66
support 6
syntax, for comments 13

T

t test 77
t value 84
tables 14, 17
TABLES statement 74
tabs
  OPTIONS 57
  OUTPUT 57, 58
tasks
  in SAS University Edition 54
  Navigation pane 7
  selecting 56
Tests for Location table 71
tests of association 76
The Analytics behind an NBA Name Change case study 118
toolbar 9
TTEST procedure 77

U

United States Department of Agriculture National Agricultural Statistics Service (website) 102
UNIVARIATE procedure 66, 67, 71, 86

V

VAR statement 73, 77, 80, 81
Variability column, Basic Statistical Measures table 70
variables
  about 14, 20, 59
  analysis 56
  assigning 57
  character 60
  classification 56
  continuous 56, 59
  dependent 56, 61
  discrete 59
  independent 60
  nominal 60
  ordinal 60
  qualitative 60
variance 70, 71
visual aids 47

W

websites
  A Chef's Life 101
  North Carolina Agriculture Overview 102
  SAS Analytics U 22
  SAS Global Forum 2015 poster 123
  SAS University Edition 6
  support 6
  United States Department of Agriculture National Agricultural Statistics Service 102
windows, in SAS University Edition 6
WITH statement 81
Work area, tabs in 6, 8
Work library 36
World Statistics Day case study 92

From *A Recipe for Success Using SAS® University Edition: How to Plan Your First Analytics Project* , by Sharon Torrence Jones. Copyright © 2016, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

About This Book

Purpose

This book is for the person that has an interest in analytics and knows about SAS®, but is not exactly sure how to get started, or, for the person who has an interest in data and wants to know more!

The book will take you on a journey through how to start using analytics, how to use SAS to accomplish a goal, how to apply SAS to your community and where it is effective, and provide real case studies of how users like you implemented SAS to solve their problems.

Is This Book for You?

This book is for anyone interested in understanding how to use SAS for real world applications. This book is great for students, in the classroom, and for the average person who wants to learn to use SAS, but has little to no experience in planning, creating, or executing a programming or analytics project.

Scope of This Book

The book will give users the ability to use SAS programs for their betterment. With data analysis at the forefront of business and more emphasis being placed on data-driven decisions, this book will help illustrate how you can use data to solve real world problems. This book will introduce you to using SAS and the power the software brings to data analytics and will provide a guide on using analytics. The book does not go into extensive detail of statistical concepts, but instead introduces them and provides an overview in the context of using SAS.

About the Examples

Software Used to Develop the Book's Content

If you are using SAS University Edition to access data and run your programs, then please check the SAS University Edition page to ensure that the software contains the product or products that you need to run the code: **http://support.sas.com/software/products/university-edition/index.html**

Example Code and Data

You can access the example code and data for this book by linking to its author page at **http://support.sas.com/publishing/authors**. Select the name of the author. Then, look for the cover thumbnail of this book, and select Example Code and Data to display the SAS programs that are included in this book.

If you are unable to access the code through the Web site, send e-mail to saspress@sas.com.

Additional Resources

Data is everywhere and has become such a large part of our lives. It can seem overwhelming, but SAS University Edition is a great platform to start exploring how data and analytics can change how you solve problems and make decisions. I was a programming novice when I began to learn SAS, and I was nervous about being able to completely understand analytics. But once I got started, I found that the syntax and logic were easy to learn and implement. I was able to create new knowledge by manipulating data that I was using in my everyday life! My hope is that as you move through this book, you will learn how to use SAS and apply to a real world project. We all have the power to transform data and create knowledge, and SAS University Edition can help you begin your journey.

To begin to explore a path in data science, there are educational resources for all ages and skill levels available for free on the web.

- Code.org **https://code.org/**
- MIT's Scratch project **https://scratch.mit.edu/**

- Code Academy **https://www.codeacademy.com/**
- Khan Academy's computer programming courses **https://www.khanacademy.org/computing/computer-programming**
- A SAS high school programming course **http://support.sas.com/learn/ap/hs/index.html**
- SAS University Edition **http://www.sas.com/en_us/software/university-edition.html**
- Online programming tutorials **https://support.sas.com/edu/schedules.html?ctry=us=2588**
- Online statistics tutorials **https://support.sas.com/edu/schedules.html?id=1320=US**

Use your passion as fuel and find the data science path that fits you!

Although this book illustrates many analyses regularly performed in businesses across industries, questions specific to your aims and issues may arise. To fully support you, SAS Institute and SAS Press offer you the following help resources:

- For questions about topics covered in this book, contact the author through SAS Press:
  - Send questions by email to saspress@sas.com; include the book title in your correspondence.
  - Submit feedback on the author's page at **http://support.sas.com/author_feedback**.
- For questions about topics in or beyond the scope of this book, post queries to the relevant SAS Support Communities at **https://communities.sas.com/welcome**.
- SAS Institute maintains a comprehensive website with up-to-date information. One page that is particularly useful to both the novice and the seasoned SAS user is its Knowledge Base. Search for relevant notes in the "Samples and SAS Notes" section of the Knowledge Base at **http://support.sas.com/resources**.
- ● Registered SAS users or their organizations can access SAS Customer Support at **http://support.sas.com**. Here you can pose specific questions to SAS Customer Support; under Support, click Submit a Problem. You will need to provide an email address to which replies can be sent, identify your organization, and provide a customer site number or license information. This information can be found in your SAS logs.

Keep in Touch

We look forward to hearing from you. We invite questions, comments, and concerns. If you want to contact us about a specific book, please include the book title in your correspondence.

To Contact the Author through SAS Press

By e-mail: saspress@sas.com

Via the Web: http://support.sas.com/author_feedback

SAS Books

For a complete list of books available through SAS, visit http://support.sas.com/bookstore.

Phone: 1-800-727-3228

Fax: 1-919-677-8166

E-mail: sasbook@sas.com

SAS Book Report

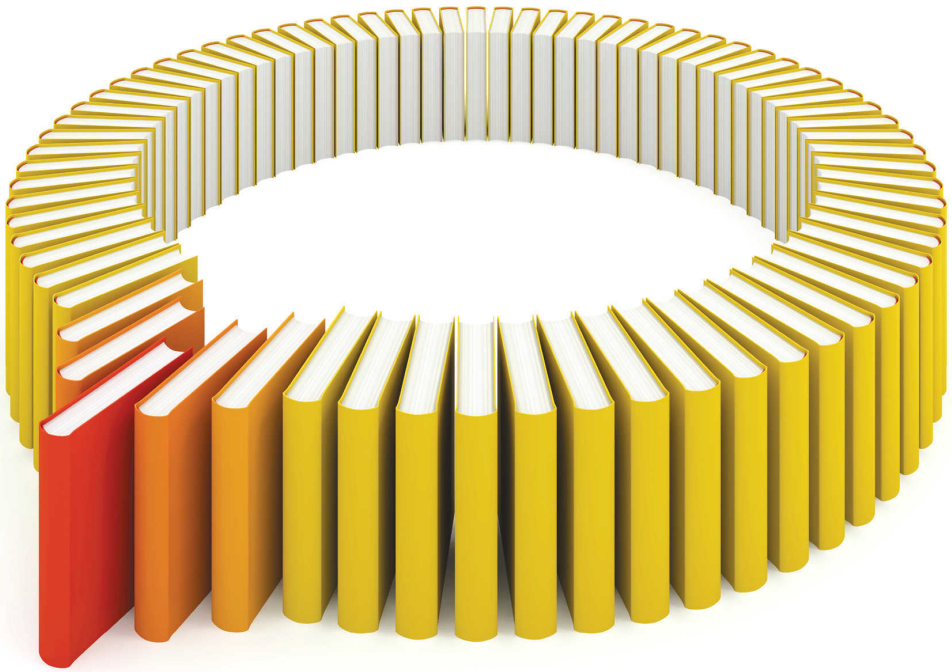
Receive up-to-date information about all new SAS publications via e-mail by subscribing to the SAS Book Report monthly eNewsletter. Visit http://support.sas.com/sbr.

Publish with SAS

SAS is recruiting authors! Are you interested in writing a book? Visit http://support.sas.com/saspress for more information.

About the Author

Sharon Jones Sharon Jones, Ed.D, is a faculty member at Central Piedmont Community college in Charlotte, NC, where she leads and continues to develop the SAS curriculum in the Continuing and Corporate Education department. Dr. Jones has been in education for 13 years as a Career and Technical Education teacher in the Charlotte Mecklenburg Schools, the Wake County Schools, and as an industry trainer. She has taught courses in computer programming, web design, ecommerce, computer science principles, and SAS programming. She has also presented and been published nationally and internationally on SAS and on integrating technology tools in the classroom.



Gain Greater Insight into Your SAS® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

support.sas.com/bookstore
for additional books and resources.

sas
THE POWER TO KNOW®

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. © 2013 SAS Institute Inc. All rights reserved. S107969US.0413