# Introduction to SAS

Sarah Bell

# Introductions

# Overview

- Day 1
  - Introduction to SAS
  - Making Changes to Data in SAS
- Day 2
  - Introduction to SAS Procedure
  - ODS Graphics Designer
  - Introduction to SAS EG

# Uses

- Access and manage data across multiple sources

- Generate reports and perform analyses

# Interfaces

- SAS Windowing Environment (SAS)
  - Provides a full programming interface
- SAS Enterprise Guide (SAS EG)
  - Provides a point-and-click interface with menus and wizards to create code

# Access at UI

- PC Installation
  - Requires purchase of SAS license
- Virtual Desktop
  - Provides access to a variety of programs through web-based system
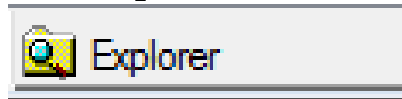  - Used on or off campus

# Starting the SAS System

- Off campus
  - [virtualdesktop.uiowa.edu](virtualdesktop.uiowa.edu)
  - Requires installation of Citrix Receiver software

- PC installed or on campus
  - Start → All Programs → SAS → SAS 9.3

# Interface Windows

- Enhanced Editor
- Log
- Output or Results Viewer
- Explorer
- Results

# Explorer

Explorer

- Provides easy access to SAS files and datasets
- Computer  provides access to all shared devices or drives
- Libraries  contains all libraries currently defined

THE UNIVERSITY OF IOWA

# Enhanced Editor

Editor - Untitled1
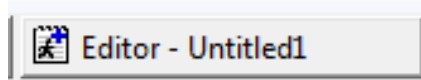
- Where you write your SAS programs
- A SAS program is a series of commands to:
  - Import and manipulate data
  - Generate reports and perform analyses
  - Output results

# Enhanced Editor

Editor - Untitled1

- Color coded to help you detect errors

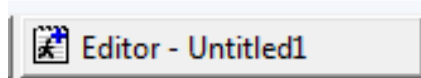| COLOR | COMMAND TYPE | EXAMPLE |
|-------|--------------|---------|
| **BOLD BLUE** | Major SAS commands | **DATA** |
| ROYAL BLUE | Sub commands, and recognized SAS words | INFILE<br>STUDENT |
| PURPLE | Words within quotes such as filenames or titles. | 'C:\My Documents\DATA.DAT' |
| **BOLD GREEN** | Numbers | 1-20 |
| GREEN | Commented out commands | *PLOT; |
| RED | Errors | TALBE |
| CALORIES | All user defined words such as variable names | CALORIES<br>RESDAT1 |

# Log

Log - (Untitled)
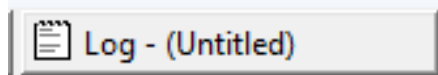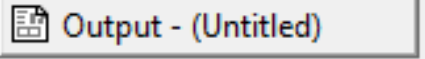
- Information pertaining to the program you've submitted is automatically displayed in the log

- Contains a list of:
  - Program commands and operations
  - Notes, warnings and errors

# Log
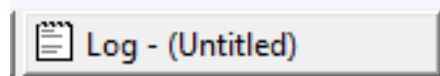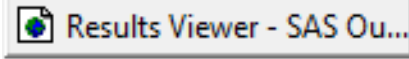
Log - (Untitled)

- Notes
  - Additional information
- Warnings
  - Program still executes but possibly not the way you expected
- Errors
  - Usually the result of a syntax or spelling error

THE UNIVERSITY OF IOWA

# Output or Results Viewer

Output - (Untitled)    Results Viewer - SAS Ou...

- When the SAS program executes without error, the results are displayed in the Output or Results Viewer
- The window the results will be displayed in will depend on the default setting

THE UNIVERSITY OF IOWA

# Output or Results Viewer

Output - (Untitled)　　　　　Results Viewer - SAS Ou...

- Output

The FREQ Procedure

| Status | Frequency | Percent |
|--------|-----------|---------|
| Alive | 3218 | 61.78 |
| Dead | 1991 | 38.22 |

- Results Viewer

The FREQ Procedure

| Status | Frequency | Percent |
|--------|-----------|---------|
| Alive | 3218 | 61.78 |
| Dead | 1991 | 38.22 |

THE UNIVERSITY OF IOWA

# Results

Results

- Provides table of contents for output

- Lists each procedure in outline form

- Can be expanded to show each part

THE UNIVERSITY OF IOWA
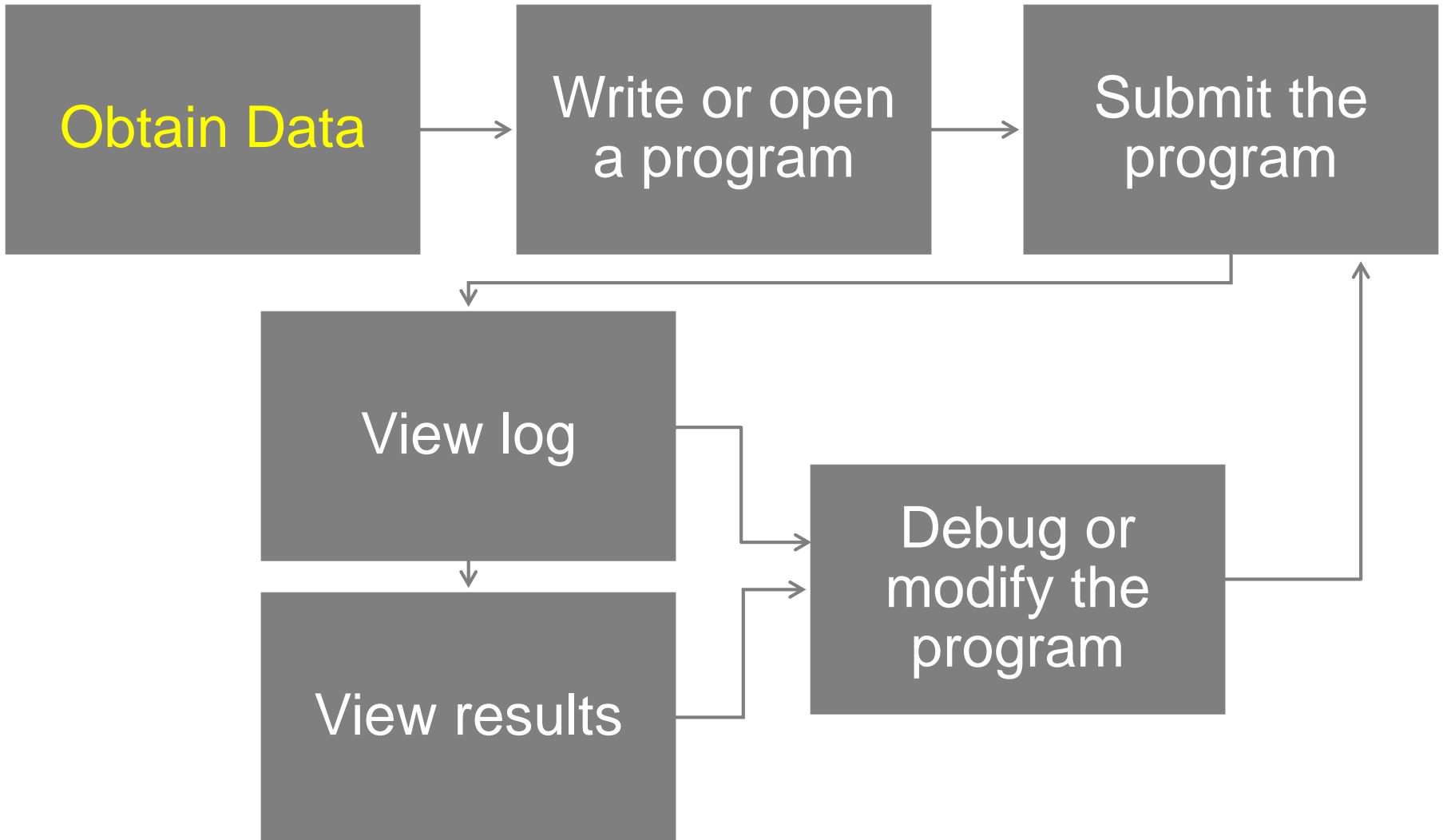
# Data Sources

- SAS only works with SAS datasets
  - Special proprietary format (.sas7dbat)

- Need to convert the dataset you have into a SAS dataset

# Converting to SAS dataset

- Can convert many types of data formats to SAS dataset
- Types
  - Excel
  - CSV
  - Access
  - Delimited
  - Text

# Data Layout

- Columns = Variables
- Rows = Observations

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Country | Life Expectancy in 2008 | Literacy Rate in 2008 | GNI per Capita | Is Developed |
| 2 | Albania | 76.63371 | 99 | 5323 | N |
| 3 | Angola | 47.03773 | 69.6 | 2829 | N |
| 4 | Argentina | 75.33398 | 97.7 | 13153 | N |
| 5 | Armenia | 73.53739 | 99.5 | 4048 | N |
| 6 | Bahrain | 75.91149 | 90.8 | 19748 | N |
| 7 | Bangladesh | 66.145 | 55 | 1998 | N |
| 8 | Belarus | 70.63288 | 99.7 | 8186 | N |
| 9 | Benin | 61.37798 | 40.8 | 1147 | N |
| 10 | Bosnia and Herzegovina | 75.10632 | 97.6 | 5827 | N |
| 11 | Botswana | 54.24102 | 83.3 | 10866 | N |
| 12 | Brunei Darussalam | 77.3641 | 95 | 24826 | N |

# Variable Name Rules

- 1-32 characters
- Must start with a letter or underscore
  - Subsequent characters can be letters, numbers, or underscores
- No blanks or special characters
- Can use any case letters
- Are not case sensitive

# Data Types

- Character
  - Can contain any character (letters, numbers, special characters, and blanks)
  - Range from 1-32,767 characters
- Numeric
  - Numbers (decimal point and minus sign)

# Dates

- Stored as numeric values



Calendar Date

| Jul 4 1776 | Jan 1 1959 | Jan 1 1960 | Jan 1 1961 | Jan 11 2202 |
|---|---|---|---|---|
| -67019 | -365 | 0 | 366 | 88399 |
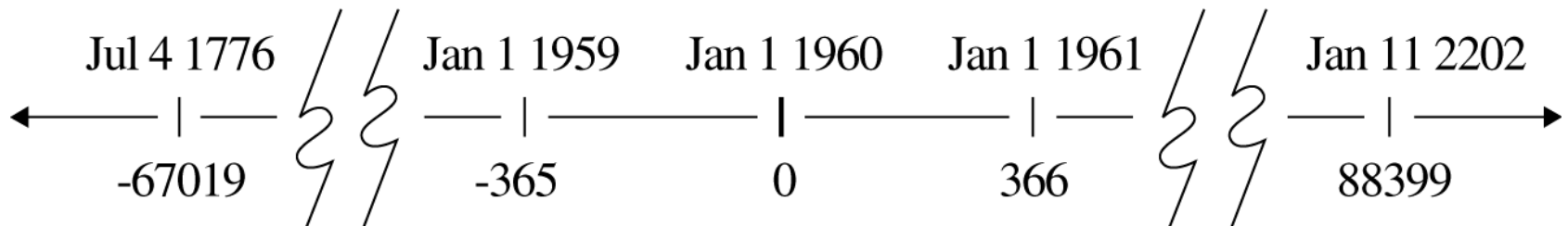
SAS Date Value

# Missing Data

- Character variables " "

- Numeric variables .

# SAS Help datasets

- Automatically installed datasets

- Used to illustrate different procedures

- Can be used to play around with

```
┌─────────────┐     ┌─────────────┐     ┌─────────────┐
│             │     │ Write or    │     │ Submit the  │
│ Obtain Data │ ──> │ open        │ ──> │ program     │
│             │     │ a program   │     │             │
└─────────────┘     └─────────────┘     └─────────────┘
                                                │
        ┌───────────────────────────────────────┘
        │
        v
┌─────────────┐
│             │
│ View log    │ ──────────┐
│             │           │
└─────────────┘           v
        │         ┌─────────────┐
        v         │ Debug or    │
┌─────────────┐   │ modify the  │
│             │   │ program     │
│ View results│──>│             │
│             │   └─────────────┘
└─────────────┘
```

Obtain Data → Write or open a program → Submit the program

View log → Debug or modify the program

View results → Debug or modify the program

THE UNIVERSITY OF IOWA
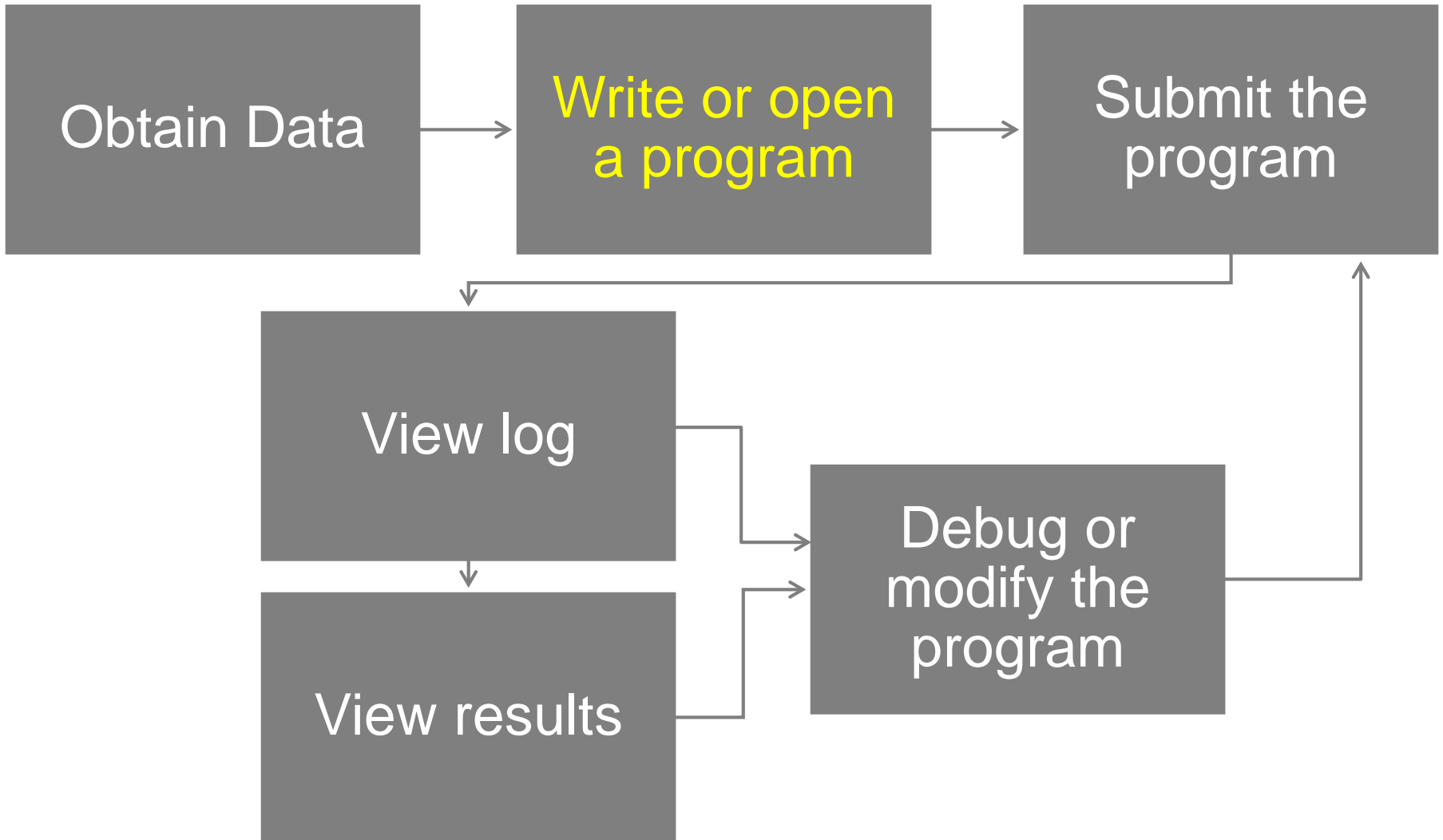
# Becoming a SAS Programmer

- SAS is best as a "write code then run" program
- To be proficient, you must learn how to write a program
  - Simple if you understand what is required

# Structural Components

- Every program typically has at least two parts:
  - DATA step
    - Reading data and variable manipulations
  - PROC step
    - Generates descriptive information and performs statistical analyses
- Must end with a RUN statement

THE UNIVERSITY OF IOWA

# DATA Step

- Used to read and modify data
  - Arithmetic calculations
  - Recoding variables
  - DO loops, IF-THEN/ELSE statements
  - Combine datasets by concatenation or match-merge operations
- Data steps execute line by line and observation by observation

# PROC Step

- Each procedure (PROC) has unique characteristics

- Basic PROC structure is similar to:

```
PROC _____ DATA= _____;
TITLE _____;
FOOTNOTE _____;
BY _____;
LABEL _____;
FORMAT _____;
RUN; <and/or> QUIT;
```

# PROC PRINT

```
PROC PRINT {options} DATA=filename{(OBS=100)};
   {TITLE 'Title of Output';}
   {BY variables;}
   {ID variables;}
    VAR variables;
   {SUM variables;}
   {SUMBY byvars;}
   {PAGEBY byvars;}
   {FORMAT variable numberfmt./$charfmt.;}
RUN;
```

# Rules for SAS Statements

- Begin and end in any column
- Must end with a semicolon (;)
- May consist of more than one line
- Multiple statements may appear on a single line
- One or more blanks should be placed between items
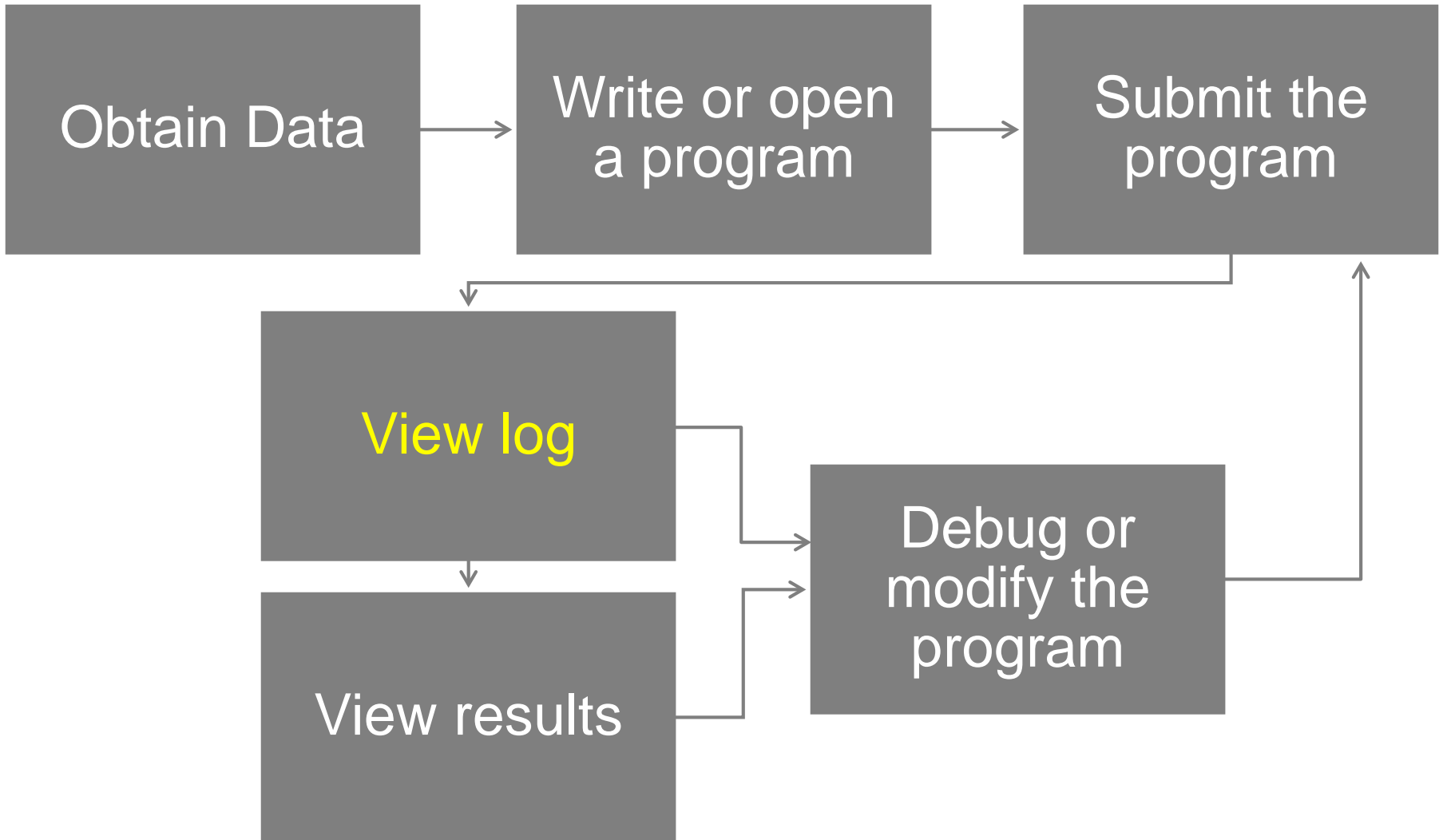- Unquoted items can be any case
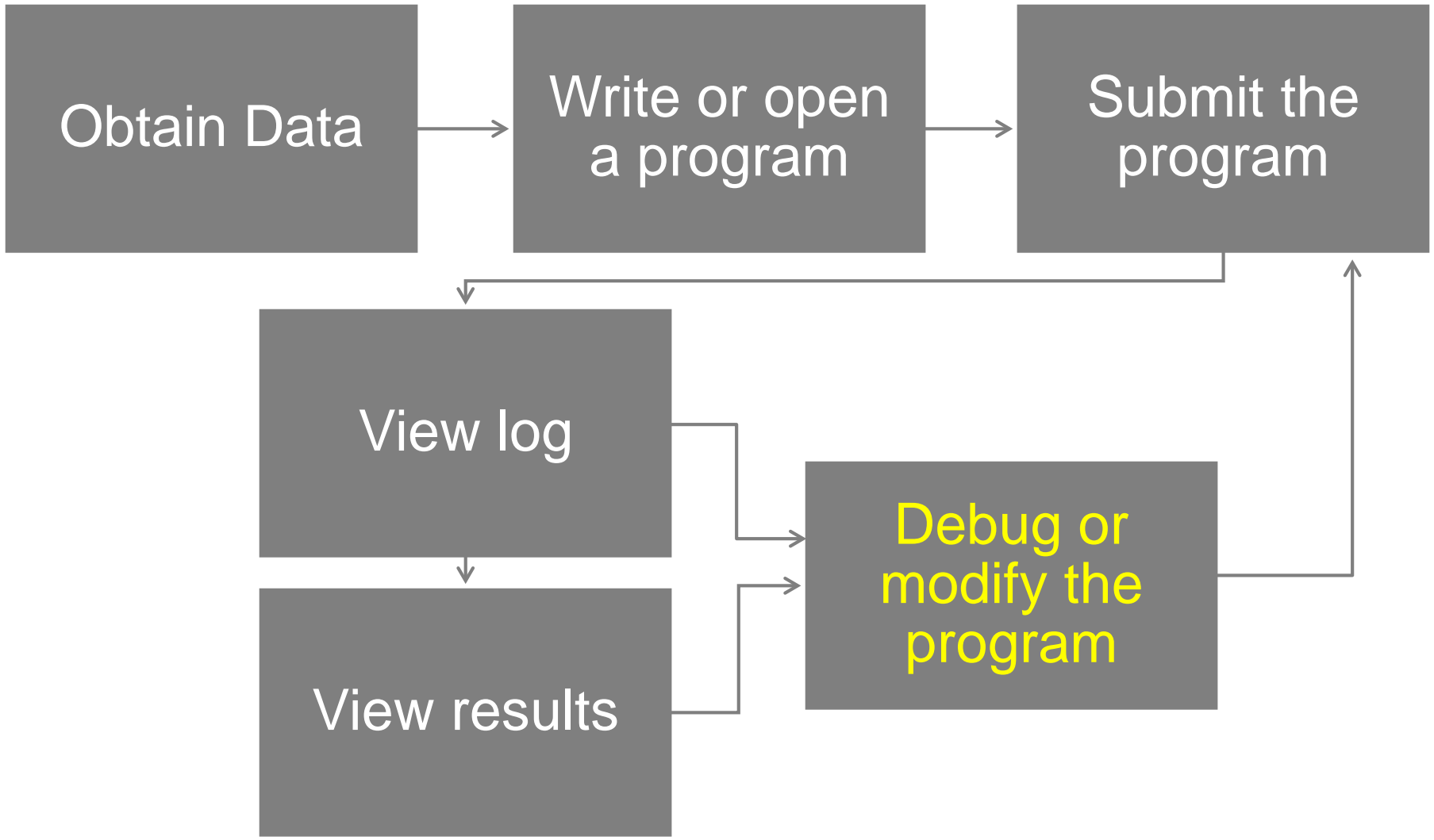
# Commenting Out

- Helpful to ignore parts of a program or add notes
- Two ways to comment out
  - Single line
    - * commented out ;
  - Multiple lines
    - /* commented out */
- Commented out text turns green
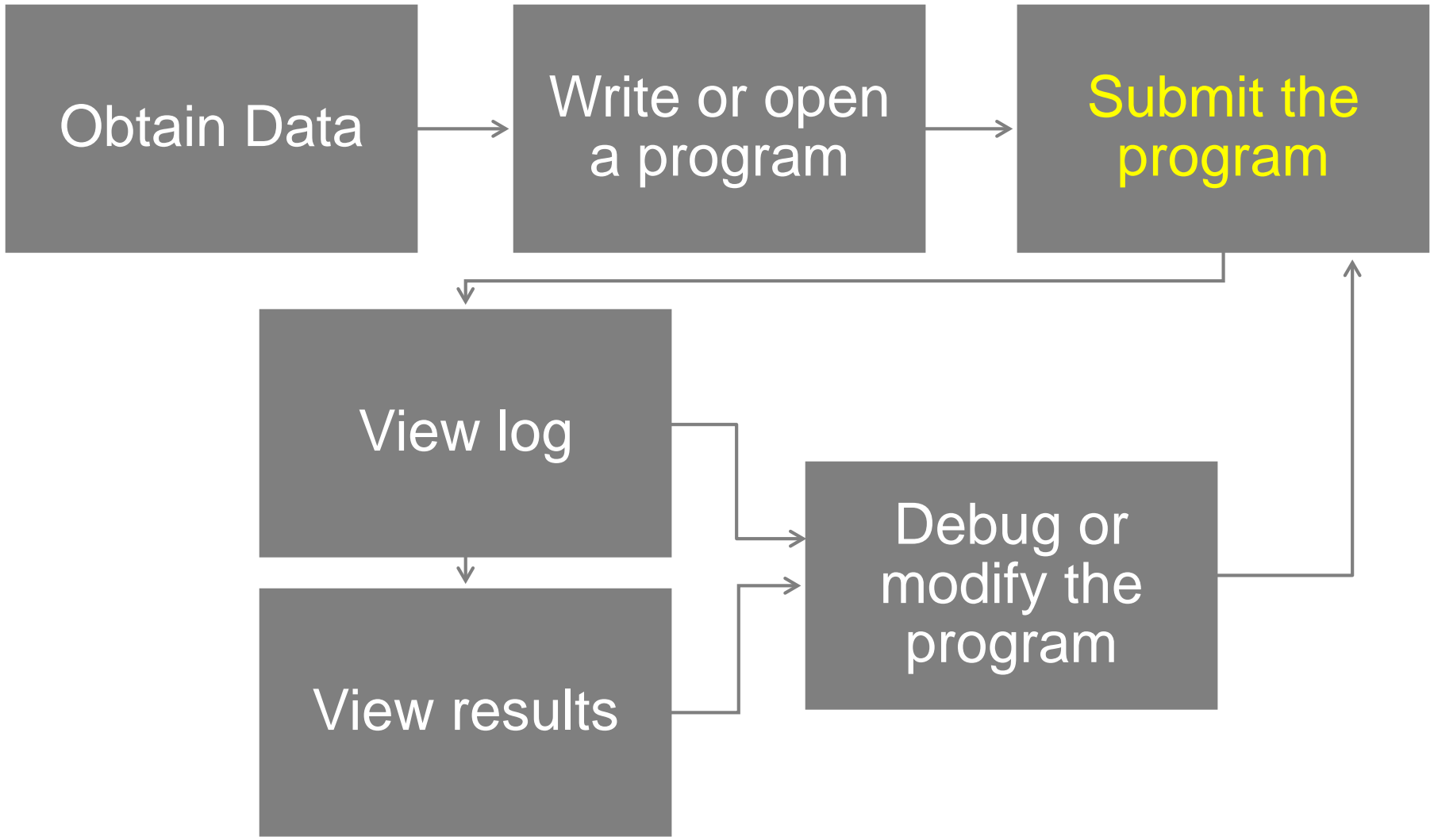
# Submitting the Program

- Can submit all or part of a program

- Click the "running man"

# Correcting Errors Checklist

- Read the Log
- Test each part of the program
- Test program using small datasets
- Be observant of the colors in your program

# Common Programming Errors

- No semicolon at the end of a statement
- Missing or mismatched quotation marks
- Misspellings
- Using the letter 'o' instead of number 0

# Correcting DATA Errors

- Data entry errors
  - Descriptive summaries
  - Create flags to alert you of errors


- SAS coding errors
  - Spot check data

# BREAK

# User-Defined Libraries

- Must submit a libname statement to create a library reference
- Is a pointer to folder on your computer where the data files are stored
- Short hand way of telling SAS where to look for SAS datasets

# Libname Statement

- General Format
  - `libname <name of library>` `"<folder location>";`
- Example
  - `libname class` `"H:\SASUsersGroup\2015.08-SSI\";`

# Libname Rules

- 1-8 characters

- Must start with a letter
  - Subsequent characters can be letters, numbers or an underscore

- No spaces

# SAS datasets

- Two parts
  - Descriptor
    - General information and variable properties
    - proc contents
  - Data
    - Data values
    - proc print

# Assignment Statements

- Basic method for adding to or modifying a SAS dataset
- Has the form
  - Variable=expression;
- Numeric constant
  - Year=2015;
- Character constant;
  - Study="Heart";
- Copy a variable
  - Newvariable=Oldvariable;

# Arithmetic Calculations

| Operation | Symbol | Example |
|---|---|---|
| Addition | + | CholestAdjust=Cholesterol+5; |
| Subtraction | - | SystAdjust=Systolic-10; |
| Multiplication | * | Heightm=Height*0.0254; |
| Division | / | BPRatio=SystAdjust/DiastAdjust |
| Exponentiation | ** | Heightm2=Heightm**2 |

# Order of Operation

- For numeric expression with >1 operator, SAS follows the order of operations

1. Parentheses

2. Exponents

3. Multiply or Divide

4. Add or Subtract

# Missing Values

- If you use a missing value in an arithmetic calculation, SAS sets the result of that calculation to missing