MODUL 6
SAS® STUDIO INTERFACE: EXPLORING DATA

THEME DESCRIPTION

Students could understand and can take advantage of the built-in SASHELP data set and various tasks that can be performed using simple point-and-click operations provided by SAS Studio in order to exploration the data for data preparation process.

WEEKLY LEARNING OUTCOMES (SUB-LESSONS)

CLO-4-Sub-CLO-6, Able to implement an integrated system of SAAS from accessing data across multiple sources to performing sophisticated analyses and delivering information -C3.

Through the following learning steps:

1. The DATA Step and the PROC Step
2. Exploring the Built-In Data Sets
3. Copying SAS Library data
4. Creating Filters: Adding. Listing and Removing
5. The Quiz Exercises

PRACTICUM SUPPORT

a. Windows Operating System

b. (any) Browser Application

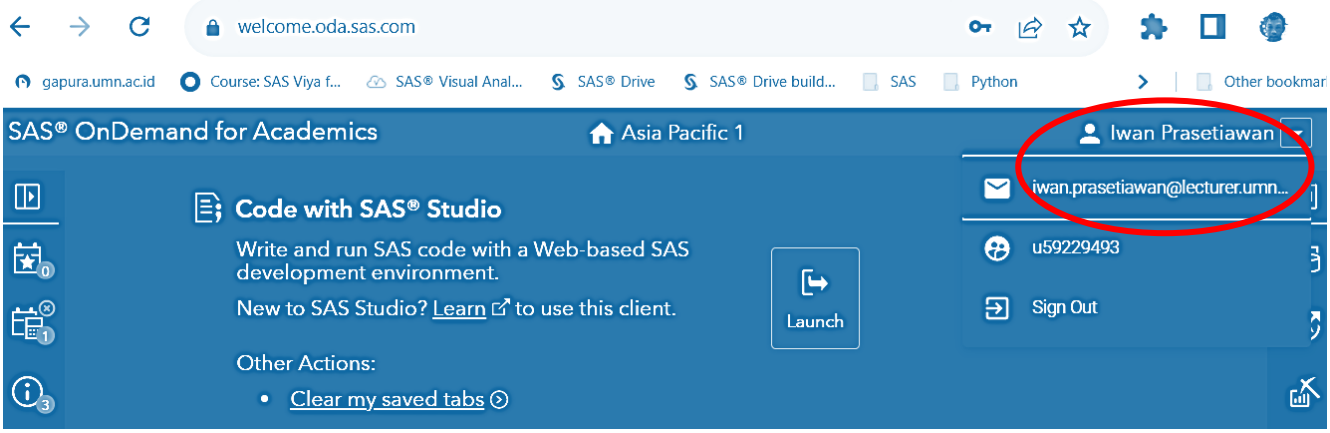
PRACTICUM STEPS

1. The DATA Step and the PROC Step

- a SAS program generally consists of a DATA step and a PROC step. Each step is a block of SAS statements with the leading keyword DATA or PROC.
- a SAS statement is a command executable by SAS. Statements in the DATA step are commands that establish or modify a SAS data set.
- Statements in the PROC step are commands that process or analyze a SAS data set.

Login to SAS Studio

Start, https://welcome.oda.sas.com



Click Launch to start SAS® Programming Language using SAS® Studio in the next 2. point below.

- The below example 1.1 contains a single DATA step and a single PROC step.

1.1 contains a single DATA step and a single PROC step

```
DATA new;                          /* Defines a data set called 'new' */
     INPUT id score;
DATALINES;
1  200
2  100
RUN;                               /* Executes the above DATA step statements */

PROC PRINT DATA=new;               /* Prints the data set 'new' */
RUN;                               /* Executes the above PROC PRINT statement */
```

- The DATA step starts with the keyword DATA and ends with the first RUN; statement.
- The PROC step starts with the keyword PROC and ends with the second RUN; statement.

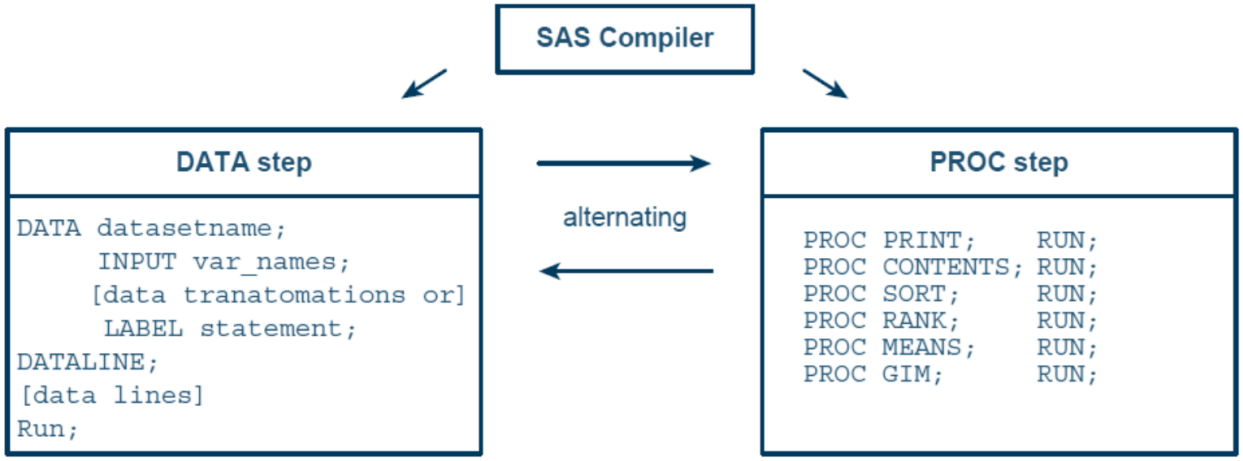
a. Create the "DATA and PROC" SAS program below:

Example 2.2 DATA and PROC in a SAS Program

```
DATA trans;
     INPUT test1 test2 test3;
     final=(test1+test2+test3)/3;
     LABEL test1='first test'
           test2='second test'
           test3='third test'
           final='overall score';
DATALINES;
60   80 99
50   87 65
100  99 98
RUN;

PROC PRINT DATA=trans;
RUN;
PROC MEANS;
RUN;
```

- See if you can distinguish those SAS statements that belong to the DATA step from those that belong to the PROC step.
- All SAS programs illustrated in this book are more complicated than the two shown above.
- Yet if you look carefully for the two keywords, DATA and PROC, you will nonetheless recognize them to be the building blocks in every program. In fact, the DATA step and the PROC step alternate and interweave throughout SAS programs.
- The flow from one step to the next is organized by a SAS compiler.
- Below diagram explains the alternation between a DATA step and a PROC step.
- All SAS statements
  - end with a semicolon (;),
  - can start in any column of a line,
  - are written in either upper- or lowercase, and

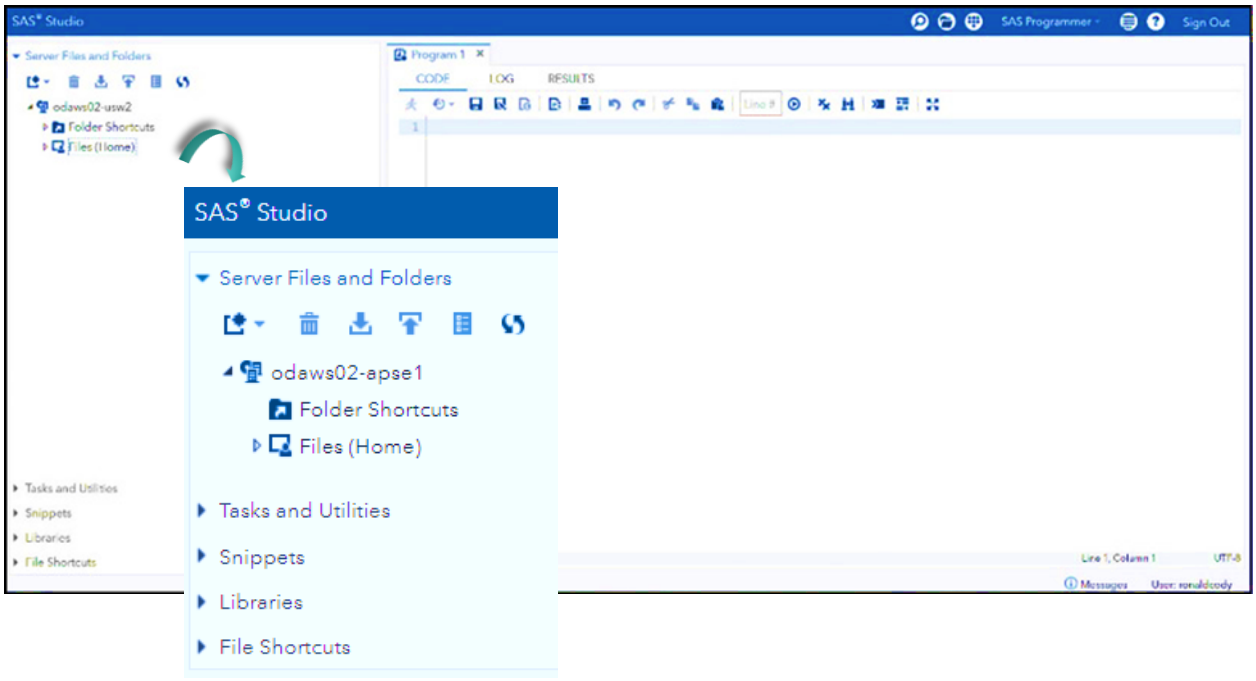
• may have any number of blank lines before or after



➲ **Save your program with the naming format file "IS-429 Week#6A DATAPROC NIM.sas**

2. Exploring the Built-In Data Sets

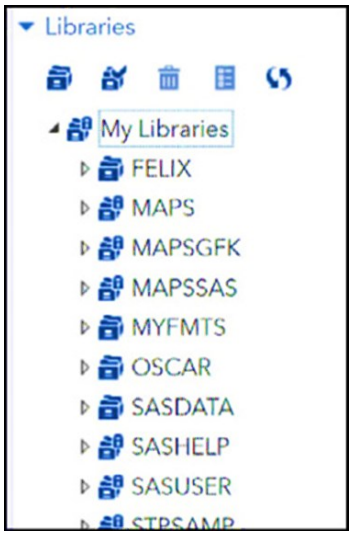
a. Get into SAS Studio, and on the left-hand pane you see the navigation pane—on the right, the work area. Here is a blow up of the navigation pane:



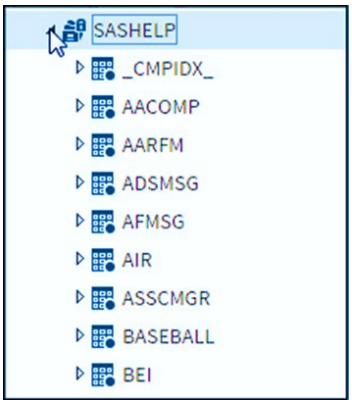
b. SAS libraries are where SAS stores SAS data sets.

c. SAS Studio ships with a number of data sets that you can play with. In the next chapter, you will see how to create your own SAS data sets from several different data sources and store them in a library of your own.

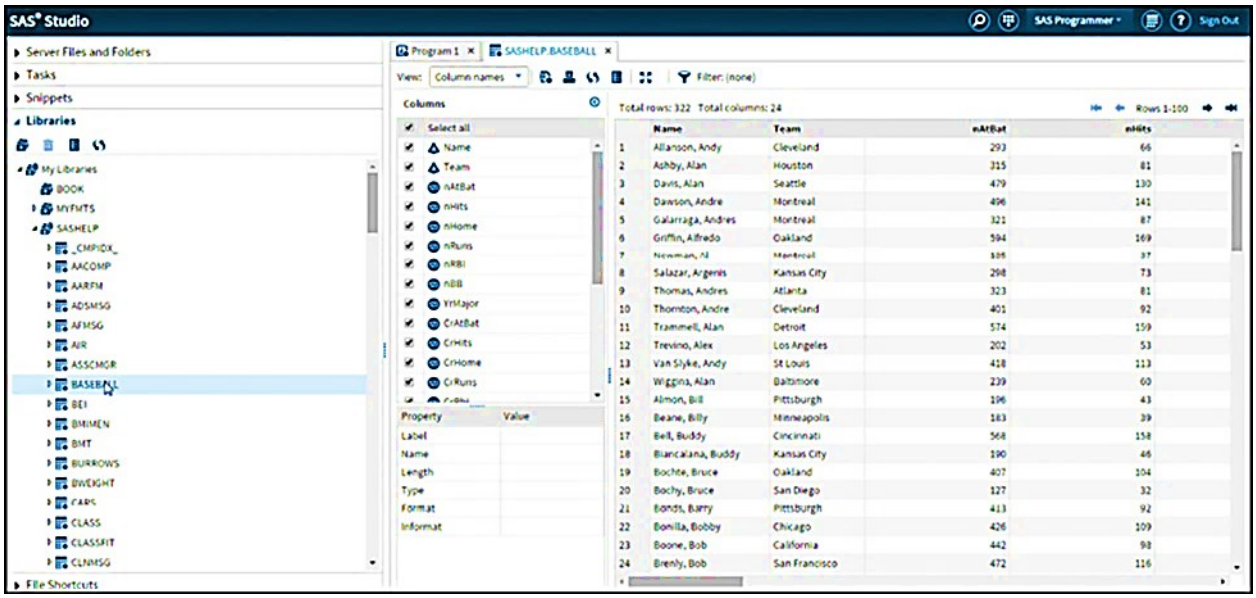
d. First click the libraries tab. Next click the My Libraries tab, you will see something like this:



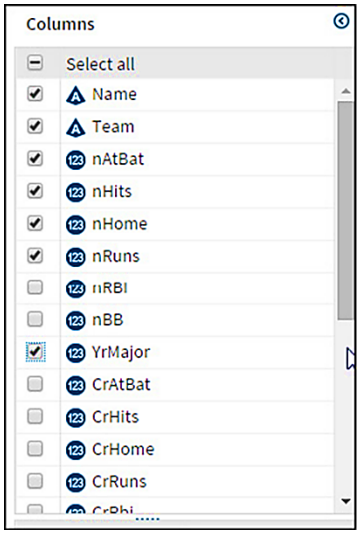
e. Some of the libraries that you see here were created by the author. Others, such as SASUSER and SASHELP, will always show up.
f. The SASHELP library is where SAS has stored all of the demonstration data sets.
g. Click the small triangle to the left of SASHELP to see a list of the SAS data sets stored there:



h. You can click the small triangle to the left of any of these data sets to see a list of variables.
i. As an alternative, double-click a data set of interest to display a list of variables and a partial listing of the data set.
j. For this example, let's double-click the BASEBALL data set. Here's what happens:



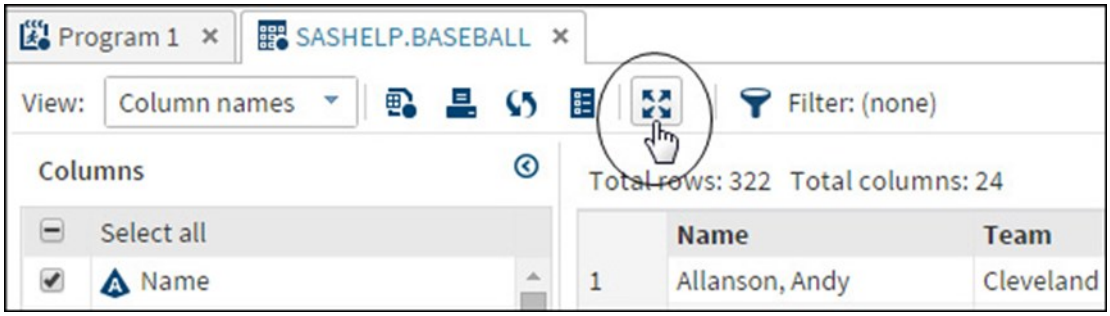
k. The middle pane shows a list of variables while the right pane shows a partial listing of the data. For a better view of the list of variables, the next figure shows an expanded view:



l. When the data set is displayed, all the variables are selected. You can click any variable to deselect it or, alternatively, uncheck Select All and then, holding down the Ctrl key, select the variables that you want to display.

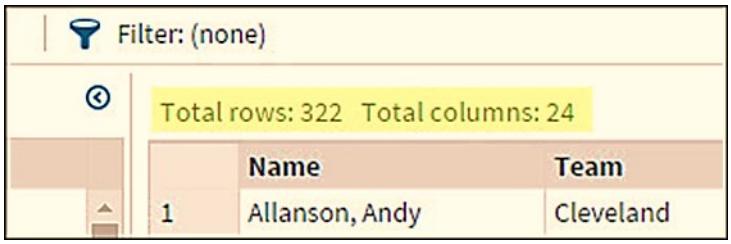
m. As you select or deselect variables, the work area changes to reflect these selections. Let's look at the work area with the variable selection shown in above figure

n. To enlarge the work area, click the Expand icon (shown below) to expand this area:



o. The work area is now enlarged and shown in Output **figure B**.

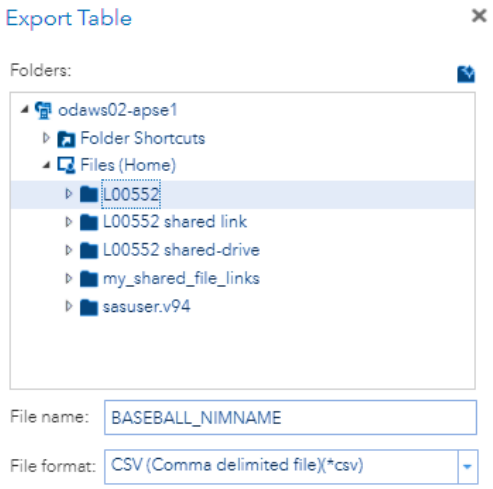
p. You can use the scroll bars to scroll left or right, up or down. At the top of the work area, you see the number of rows (observations) and columns (variables) in the data set. Here is an enlarged view.



You see that there are 322 rows and 24 columns. These numbers will change as you change your variable selections or create filters

3. Copying SAS Library Data

a. Right click to export baseball data to your own folder, it will produce raw data input as desire format to destination folder.



b. After the raw data has been successfully created in your folder, currently your baseball file is still in txt format it can't be read as database table.

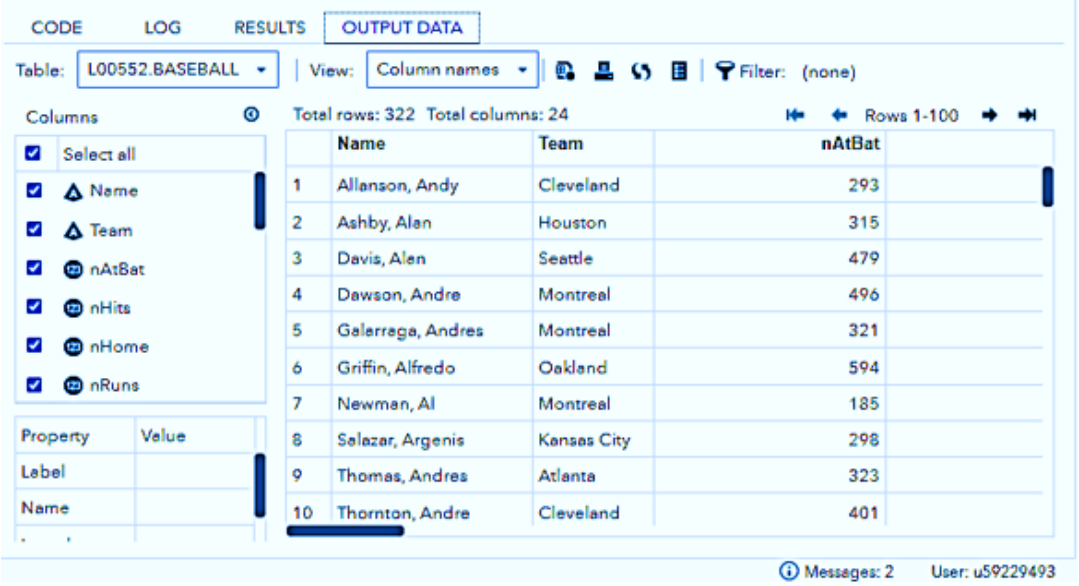
c. If you have a baseball table like in CAS SAHELP, you have to import the data first, continue with the import step into the library you have.

d. On the LOG tab under the work area, you can see the export program has been created automatically by the system. Copy the code into a new SAS program worksheet for you to change the output target as below:

```
/* Week#6B  */
/* Generated Code (IMPORT) */
/* Source File: BASEBALL_NIMNAME.csv */
/* Source Path: /home/u59229493/L00552 */
/* Code generated on: 2/26/22, 9:12 PM */

%web_drop_table(WORK.BASEBALL_NIMNAME);


FILENAME REFFILE '/home/u59229493/L00552/BASEBALL_NIMNAME.csv';

PROC IMPORT DATAFILE=REFFILE
   DBMS=CSV
   OUT=WORK.BASEBALL_NIMNAME;
   GETNAMES=YES;
RUN;

PROC CONTENTS DATA=WORK.BASEBALL_NIMNAME; RUN;


%web_open_table(WORK.BASEBALL_NIMNAME);
```

e. Run the program code and make sure that the output data from the export process is entered into your own library.

f. After the process is run there are no errors, check the number of records and the column is the same as the source data:



g. Now, use your own baseball to do some further data exploration exercises.
h. **Print it out** your baseball table that you already have with a display that matches the **output figure B in pdf format**

➲ **Save your code with the naming format of IS-429 Lab Week#6B Copying Data NIM.sas**

Sorting Your Data

i. If you place the cursor on a column heading, an arrow appears.
j. By clicking anywhere in the column heading (the column heading nAtBat stands for the number of times at bat), the values are sorted from low to high (an ascending sort).

nAtBat	nHits
293	66
315	81
479	130
496	141
321	87
594	169
185	37
298	73
323	81
401	92
574	159

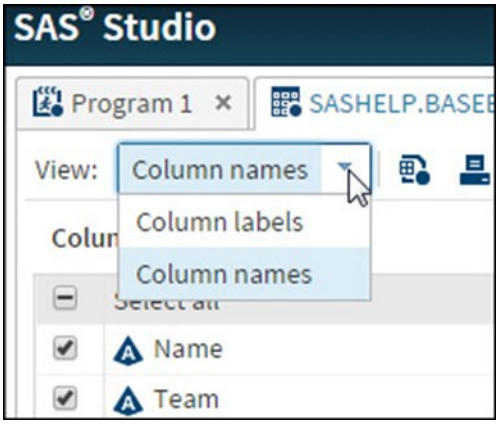
nAtBat ▲	nHits
127	32
127	32
138	31
143	39
151	41
155	44
155	41
160	39
161	36

k. You can click the triangle to the right of the column name to request a descending sort

Switching between Column Names and Column Labels

- SAS has been the premier data analysis language since computer variables were limited to only 8 characters (now **SAS variable names can be 32 characters long**).
- To fix the short variable name problem, long ago, SAS created labels that could be used to identify variables in addition to the variable names.

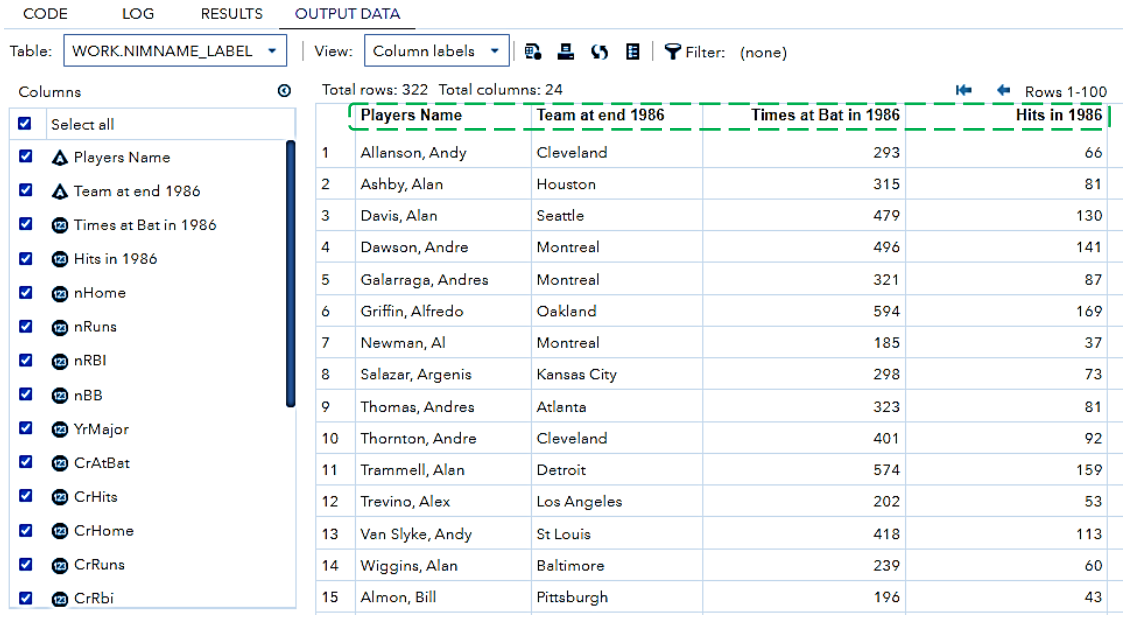
l. You can switch between variable names and variable labels by clicking on the View tab:



m. If you choose column labels, the column names are replaced by labels. (This only works, of course, if you creating the data set created labels for the variables as follows:

```
data work.NIMNAME_label;
    set work.BASEBALL_NIMNAME;

    label Name = 'Players Name'
    Team = 'Team at end 1986'
    nAtBat = 'Times at Bat in 1986'
    nHits ='Hits in 1986';
run;
```

n. Below figure shows the effect of switching to column labels for the BASEBALL data set (that did contain labels):



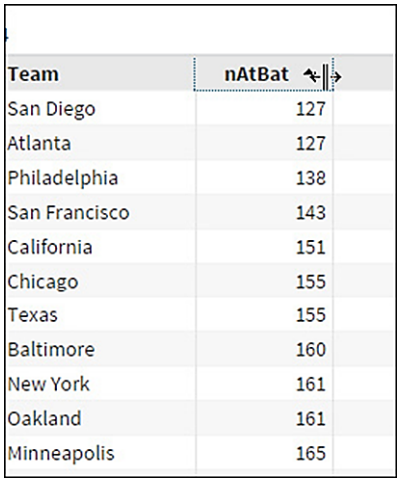
	Players Name	Team at end 1986	Times at Bat in 1986	Hits in 1986
1	Allanson, Andy	Cleveland	293	66
2	Ashby, Alan	Houston	315	81
3	Davis, Alan	Seattle	479	130
4	Dawson, Andre	Montreal	496	141
5	Galarraga, Andres	Montreal	321	87
6	Griffin, Alfredo	Oakland	594	169
7	Newman, Al	Montreal	185	37
8	Salazar, Argenis	Kansas City	298	73
9	Thomas, Andres	Atlanta	323	81
10	Thornton, Andre	Cleveland	401	92
11	Trammell, Alan	Detroit	574	159
12	Trevino, Alex	Los Angeles	202	53
13	Van Slyke, Andy	St Louis	418	113
14	Wiggins, Alan	Baltimore	239	60
15	Almon, Bill	Pittsburgh	196	43

➲ **Save your code with the naming format of IS-429 Lab Week#6C ColumnLabels NIM.sas**

Resizing Tables

o. By placing the cursor on the dividing line between columns, the pointer changes to double vertical lines, enabling you to then drag the border of the column left or right.

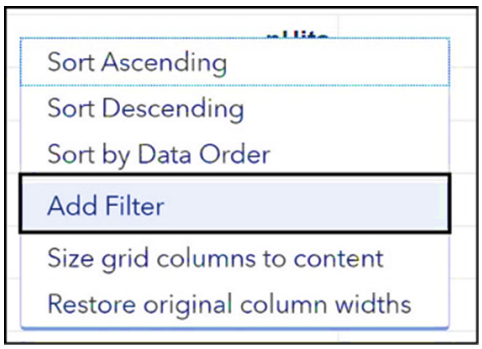
p. In below figure, the nAtBat column was resized (made smaller).



Team	nAtBat
San Diego	127
Atlanta	127
Philadelphia	138
San Francisco	143
California	151
Chicago	155
Texas	155
Baltimore	160
New York	161
Oakland	161
Minneapolis	165

4. Creating Filters: Adding. Listing and Removing

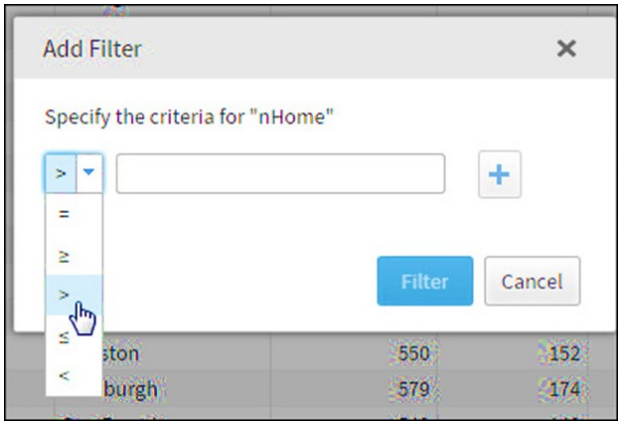
a. Right-clicking in any column will brings you up to the following menu



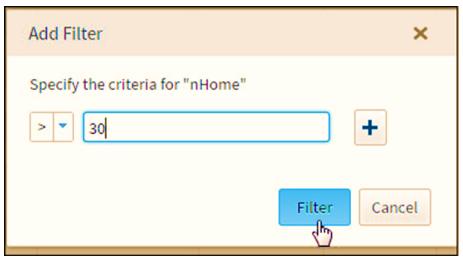
b. On the menu shows that you can perform ascending or descending sorts (the same as clicking a column heading). You can do things such as sort in data order, add a filter, or resize columns.

- Filters are very useful in exploring the data because they enable you to subset rows of the table by specifying criteria, such as displaying only rows where the number of home runs was 30 or more.
- The following screen shows how to create a filter.

c. Click Add Filter to display the screen shown next:



d. In the drop-down menu, select a logical operator. In this example, you are choosing greater than.



e. Next, enter a criterion for the variable that you have selected. If you choose a categorical variable like Team, then the filter dialog box gives you the list of values from which you can select. If you want to add more conditions, click the + (plus) sign.

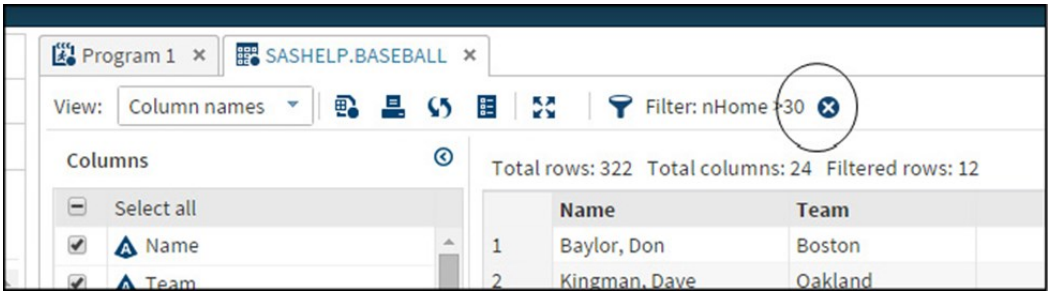
f. If you are finished, click Filter to complete your reques

Total rows: 322 Total columns: 24 Filtered rows: 11

	Name	Team	nAtBat	nHits	nHome
1	Barfield, Jesse	Toronto	589	170	40
2	Schmidt, Mike	Philadelphia	552	160	37
3	Kingman, Dave	Oakland	561	118	35
4	Gaetti, Gary	Minneapolis	596	171	34
5	Canseco, Jose	Oakland	600	144	33
6	Mattingly, Don	New York	677	238	31
7	Davis, Glenn	Houston	574	152	31
8	Baylor, Don	Boston	585	139	31
9	Bell, George	Toronto	641	198	31
10	Puckett, Kirby	Minneapolis	680	223	31
11	Parker, Dave	Cincinnati	637	174	31

g. In above figure, you see all rows in the BASEBALL data set where the player hit more than 30 home runs.

h. If you want to remove the filter, click the X next to the filter, as shown in figure below:



i. Create your baseball table display with a home-run record of at least 30 and a maximum of 80, according to the display on **output figure D**.

j. **Print it out the results** as shown on Output **figure D** by format name **"IS-429 Lab Week#6D Filtering Data.pdf"**.

Conclusion

you saw how to open one of the built-in SASHELP data sets and the various tasks that you can perform using simple point-and-click operations supplied by SAS Studio.

5. The Quiz Exercises

The following exercise is a lesson for you to get used to reading the very informative system log provided by SAS Studio.
Type the program code below in full, the scenario of the program is:

a. Make a list of BDA class students as many as 14 students
b. Calculating the BDA class average
c. Ranking the students based on their grades
d. Scored students

```
DATA roster;                                          /* Creates a data set 'roster' */

     INPUT name $ sex $ id $ stand pretest first second final; /* $ declares 'name' and 'id' */
                                                              /* as character variables     */
     composite = pretest + first + second + final;

     LABEL stand = 'academic standing in college';

DATALINES;
JOHN   m    1  1   9  31     45     .          /* A period (.) means a missing score */
DAN    m    2  4  18  46     53    54
LYNN   f    3  1   7  38     33    43
CATHY  f    4  2  12  34     50    32
JAMES  m    5  2  14  31     47    43
TIM    m    6  4  20  45     51    57
HOPE   f    7  4  17  34     46    50
TED    m    8  2  12  44     52    47
SASSY  f    9  4  18  50     57    56
NANCY  f   10  3  15  29     39    42
PAUL   m   11  3  15  24     48    49
LIN    m   12  4  18  48     54    54
TOM    m   13  4  21  48     52    42
BOB    m   14  1  11  32     41    40
RUN;

PROC PRINT; RUN;                                      /* Prints the raw data in 'roster' */

PROC MEANS DATA=roster; RUN;                          /* Computes averages */

PROC RANK DESCENDING OUT=temp;                        /* Ranks students by their 'composite' */
     VAR composite;
     RANKS rank;
RUN;

PROC SORT;                                            /* Sorts students by their 'rank' */
     BY rank;
RUN;

PROC PRINT DATA=temp;                                 /* Prints the ranked students */
RUN;
```

e. **Run the program and you should be able to answer the following questions**:
   1) What is the average score of the final variable?
   2) Who is Number 1 in the class according to the ranked data?
   3) What special note or warning is there in the Log window?

f. **answer the three questions above in the word file** of **"IS-429 Lab Week#6E Quiz Exercises NIM yourName.doc/docx"** and complete it with the evidence/screenshot if you have it.

➲ **Save your code with the format name of your file "IS-429 Week#6E NIM.sas"**

❖ **Finally**, today's practicum is over, **collect all your pdf, word file and all sas programs into "IS-429 Week#6 NIM yourName.zip** and submit immediately today to e-Learning IS-429 Practicum Week#6

RESULTS/ OUTPUT

1. The DATA Step and the PROC Step

CODE LOG RESULTS OUTPUT DATA

Table of Contents

A

Obs	test1	test2	test3	final
1	60	80	99	79.6667
2	50	87	65	67.3333
3	100	99	98	99.0000

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
test1	first test	3	70.0000000	26.4575131	50.0000000	100.0000000
test2	second test	3	88.6666667	9.6090235	80.0000000	99.0000000
test3	third test	3	87.3333333	19.3476958	65.0000000	99.0000000
final	overall score	3	82.0000000	15.9617599	67.3333333	99.0000000

2. Exploring the Built-In Data Sets

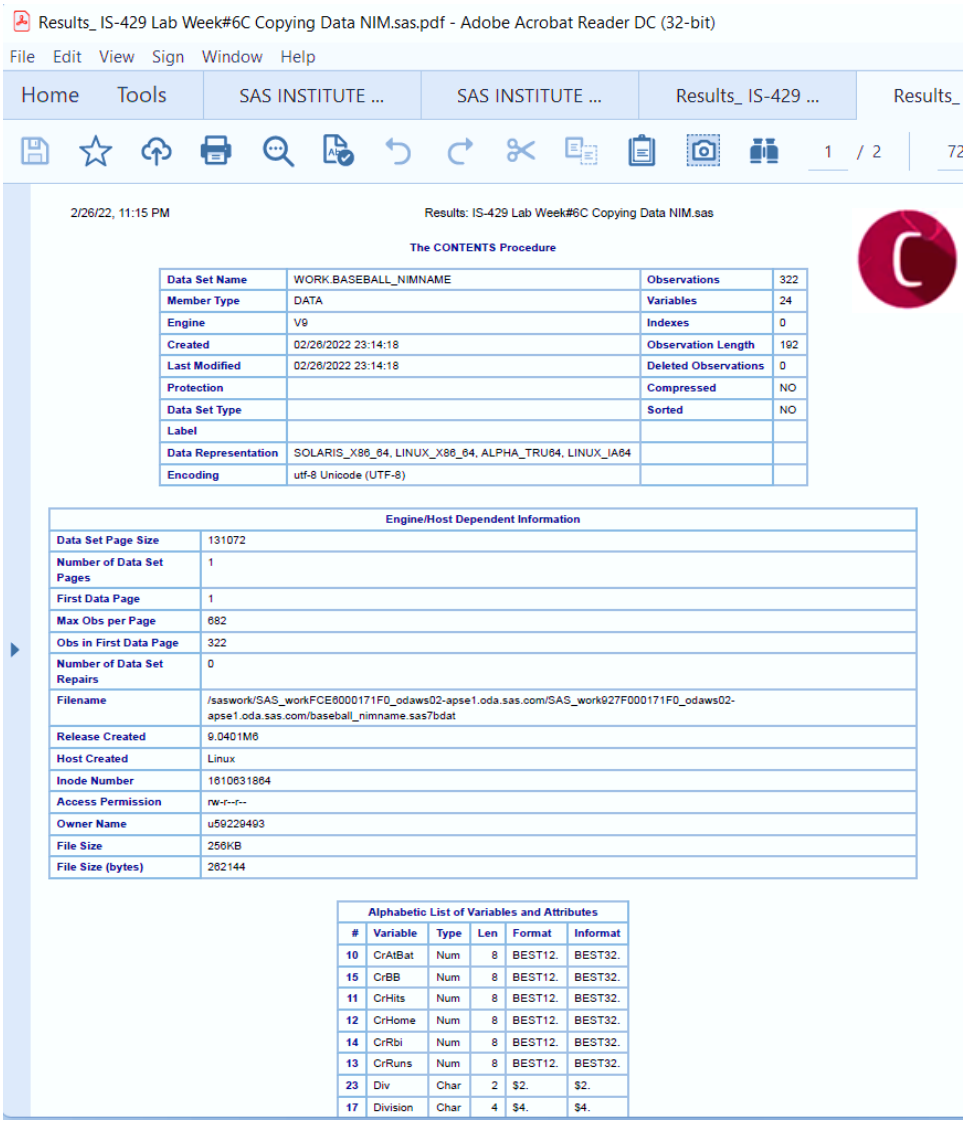
Total rows: 322 Total columns: 24 Rows 1-100

	Name	Team	nAtBat	nHits	nHome	nRuns
1	Aldrete, Mike	San Francisco	216	54	2	27
2	Allanson, Andy	Cleveland	293	60	1	30
3	Almon, Bill	Pittsburgh	196	43	7	29
4	Anderson, Dave	Los Angeles	216	53	1	31
5	Armas, Tony	Boston	425	112	11	40
6	Ashby, Alan	Houston	315	81	7	24
7	Backman, Wally	New York	387	124	1	67
8	Baines, Harold	Chicago	570	169	21	72
9	Baker, Dusty	Oakland	242	58	4	25
10	Balboni, Steve	Kansas City	512	117	29	54
11	Bando, Chris	Cleveland	254	68	2	28
12	Barfield, Jesse	Toronto	589	170	40	107
13	Barrett, Marty	Boston	625	179	4	94
14	Bass, Kevin	Houston	591	184	20	83
15	Baylor, Don	Boston	585	139	31	93
16	Beane, Billy	Minneapolis	183	39	3	20
17	Bell, Buddy	Cincinnati	568	158	20	89
18	Bell, George	Toronto	641	198	31	101
19	Belliard, Rafael	Pittsburgh	309	72	0	33
20	Beniquez, Juan	Baltimore	343	103	6	48
21	Bernazard, Tony	Cleveland	562	169	17	88
22	Biancalana, Buddy	Kansas City	190	46	2	24
23	Bilardello, Dann	Montreal	191	37	4	12
24	Bochte, Bruce	Oakland	407	104	6	57

B

RESULTS/ OUTPUT

3. Copying SAS Library data



Results_ IS-429 Lab Week#6C Copying Data NIM.sas.pdf - Adobe Acrobat Reader DC (32-bit)

File Edit View Sign Window Help

Home Tools SAS INSTITUTE ... SAS INSTITUTE ... Results_ IS-429 ... Results_

1 / 2 72

2/26/22, 11:15 PM Results: IS-429 Lab Week#6C Copying Data NIM.sas

The CONTENTS Procedure

Data Set Name	WORK.BASEBALL_NIMNAME	Observations	322
Member Type	DATA	Variables	24
Engine	V9	Indexes	0
Created	02/26/2022 23:14:18	Observation Length	192
Last Modified	02/26/2022 23:14:18	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)		

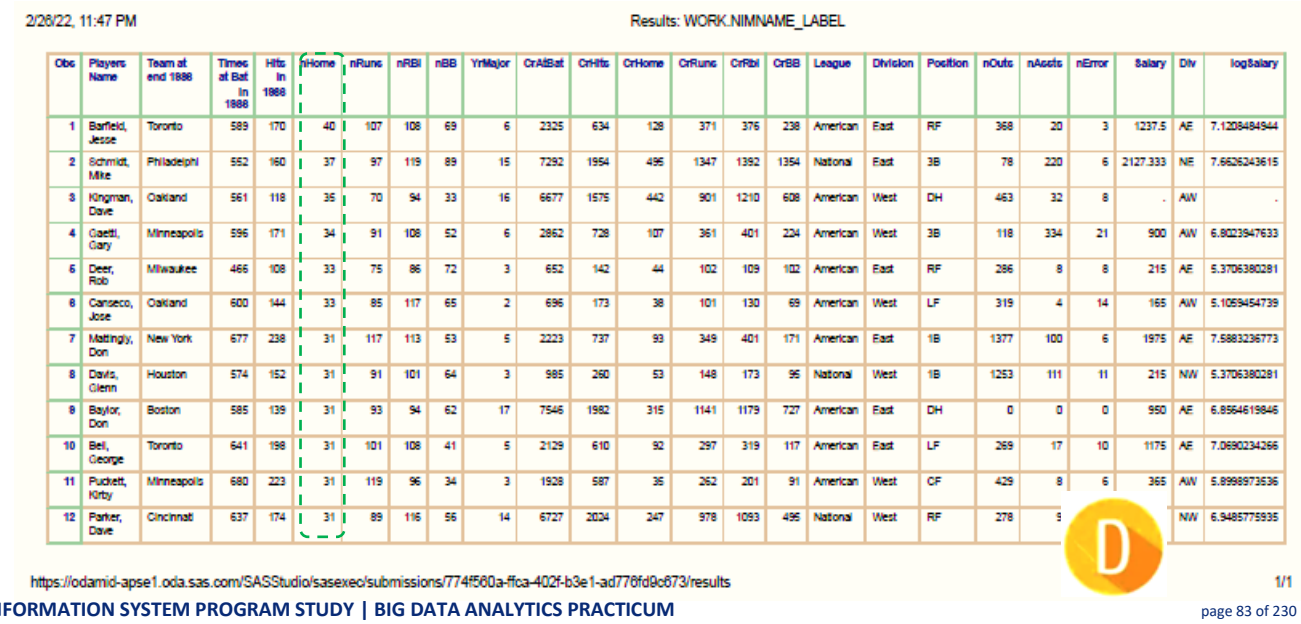
Engine/Host Dependent Information

Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	682
Obs in First Data Page	322
Number of Data Set Repairs	0
Filename	/saswork/SAS_workFCE6000171F0_odaws02-apse1.oda.sas.com/SAS_work927F000171F0_odaws02-apse1.oda.sas.com/baseball_nimname.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	1610631864
Access Permission	rw-r--r--
Owner Name	u59229493
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
10	CrAtBat	Num	8	BEST12.	BEST32.
15	CrBB	Num	8	BEST12.	BEST32.
11	CrHits	Num	8	BEST12.	BEST32.
12	CrHome	Num	8	BEST12.	BEST32.
14	CrRbi	Num	8	BEST12.	BEST32.
13	CrRuns	Num	8	BEST12.	BEST32.
23	Div	Char	2	$2.	$2.
17	Division	Char	4	$4.	$4.

4. Creating Filters: Adding. Listing and Removing

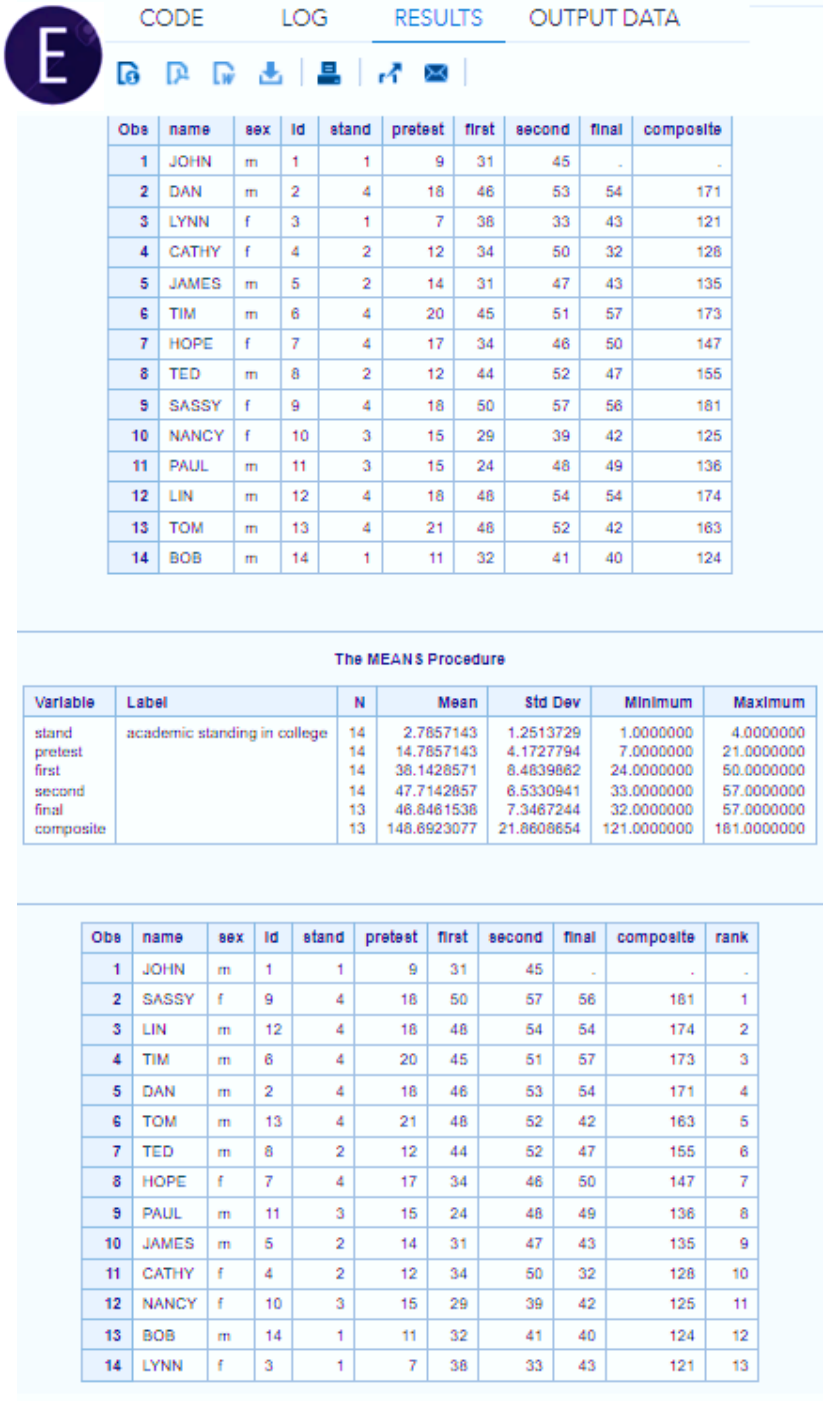


2/26/22, 11:47 PM Results: WORK.NIMNAME_LABEL

Obs	Players Name	Team at end 1886	Times at Bat in 1886	Hits in 1886	nHome	nRuns	nRBI	nBB	YrMajor	CrAtBat	CrHits	CrHome	CrRuns	CrRbi	CrBB	League	Division	Position	nOuts	nAssts	nError	Salary	Div	logSalary
1	Barfield, Jesse	Toronto	589	170	40	107	108	69	6	2325	634	128	371	376	238	American	East	RF	368	20	3	1237.5	AE	7.1208484944
2	Schmidt, Mike	Philadelphi	552	160	37	97	119	89	15	7292	1954	496	1347	1392	1354	National	East	3B	78	220	6	2127.333	NE	7.6626243615
3	Kingman, Dave	Oakland	561	118	35	70	94	33	16	6677	1575	442	901	1210	608	American	West	DH	463	32	8	.	AW	.
4	Gaetti, Gary	Minneapolis	596	171	34	91	108	52	6	2862	728	107	361	401	224	American	West	3B	118	334	21	900	AW	6.8023947633
5	Deer, Rob	Milwaukee	466	108	33	75	86	72	3	652	142	44	102	109	102	American	East	RF	286	8	8	215	AE	5.3706380281
6	Canseco, Jose	Oakland	600	144	33	85	117	65	2	696	173	38	101	130	69	American	West	LF	319	4	14	165	AW	5.1059454739
7	Mattingly, Don	New York	677	238	31	117	113	53	5	2223	737	93	349	401	171	American	East	1B	1377	100	6	1975	AE	7.5883236773
8	Davis, Glenn	Houston	574	152	31	91	101	64	3	985	260	53	148	173	95	National	West	1B	1253	111	11	215	NW	5.3706380281
9	Baylor, Don	Boston	585	139	31	93	94	62	17	7546	1982	315	1141	1179	727	American	East	DH	0	0	0	950	AE	6.8564619846
10	Bell, George	Toronto	641	198	31	101	108	41	5	2129	610	92	297	319	117	American	East	LF	269	17	10	1175	AE	7.0690234266
11	Puckett, Kirby	Minneapolis	680	223	31	119	96	34	3	1928	587	35	262	201	91	American	West	CF	429	8	6	365	AW	5.8998973536
12	Parker, Dave	Cincinnati	637	174	31	89	116	56	14	6727	2034	247	978	1093	495	National	West	RF	278	9			NW	6.9485775935

https://odamid-apse1.oda.sas.com/SASStudio/sasexec/submissions/774f560a-ffca-402f-b3e1-ad776fd9c673/results 1/1

5. The Quiz Exercises



CODE LOG RESULTS OUTPUT DATA

Obs	name	sex	id	stand	pretest	first	second	final	composite
1	JOHN	m	1	1	9	31	45	.	.
2	DAN	m	2	4	18	46	53	54	171
3	LYNN	f	3	1	7	38	33	43	121
4	CATHY	f	4	2	12	34	50	32	128
5	JAMES	m	5	2	14	31	47	43	135
6	TIM	m	6	4	20	45	51	57	173
7	HOPE	f	7	4	17	34	46	50	147
8	TED	m	8	2	12	44	52	47	155
9	SASSY	f	9	4	18	50	57	56	181
10	NANCY	f	10	3	15	29	39	42	125
11	PAUL	m	11	3	15	24	48	49	136
12	LIN	m	12	4	18	48	54	54	174
13	TOM	m	13	4	21	48	52	42	163
14	BOB	m	14	1	11	32	41	40	124

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
stand	academic standing in college	14	2.7857143	1.2513729	1.0000000	4.0000000
pretest		14	14.7857143	4.1727794	7.0000000	21.0000000
first		14	38.1428571	8.4839862	24.0000000	50.0000000
second		14	47.7142857	6.5330941	33.0000000	57.0000000
final		13	46.8461538	7.3467244	32.0000000	57.0000000
composite		13	148.6923077	21.8608654	121.0000000	181.0000000

Obs	name	sex	id	stand	pretest	first	second	final	composite	rank
1	JOHN	m	1	1	9	31	45	.	.	.
2	SASSY	f	9	4	18	50	57	56	181	1
3	LIN	m	12	4	18	48	54	54	174	2
4	TIM	m	6	4	20	45	51	57	173	3
5	DAN	m	2	4	18	46	53	54	171	4
6	TOM	m	13	4	21	48	52	42	163	5
7	TED	m	8	2	12	44	52	47	155	6
8	HOPE	f	7	4	17	34	46	50	147	7
9	PAUL	m	11	3	15	24	48	49	136	8
10	JAMES	m	5	2	14	31	47	43	135	9
11	CATHY	f	4	2	12	34	50	32	128	10
12	NANCY	f	10	3	15	29	39	42	125	11
13	BOB	m	14	1	11	32	41	40	124	12
14	LYNN	f	3	1	7	38	33	43	121	13

REFERENCE

1. SAS Institute Inc. 2020. SAS® Viya® Programming: Getting Started. Cary, NC: SAS Institute Inc.
2. David Pope. 2017. Big Data Analytics with SAS. Packt Publishing Ltd.Birmingham, UK.
3. SAS® Support | Documentation
4. Other additional references are excerpts from various Online Learning/websites.