CHAPTER 1

Inputting Raw Data

SOLUTION TO PROBLEM 1
DATA CLASS;
  INPUT F_NAME $ ID $ GENDER $ GPA HEIGHT WEIGHT;
DATA_LINES;
  Hector 123 M 3.5 59 155
  Nancy 328 F 3.7 52 99
  Edward 747 M 2.4 62 205
  Michelle 778 F 3.0 54 115
  Sampson 289 M 3.5 60 180
;
PROC PRINT DATA=CLASS;
  TITLE 'Listing of CLASS data set';
RUN;

SOLUTION TO PROBLEM 2
DATA CLASS;
  INPUT F_NAME $ ID $ GENDER $ GPA HEIGHT WEIGHT;
DATA_LINES;
  George 123 M 3.5 . 155
  328 F 3.7 52 99
  Edward 747 M . . 205
  Michelle 778 F 3.0 54 .
  Sampson 289 M 3.5 60 180
;
PROC PRINT DATA=CLASS;
  TITLE 'Listing of CLASS data set';
RUN;

SOLUTION TO PROBLEM 3
DATA CLASS;
  INFORMAT F_NAME $10.10;

/*----------------------*/
| Two alternative solutions are:
|    |    |    |
|    | LENGTH F_NAME $ 10; |
| or | INPUT F_NAME $10. ID $ GENDER $ GPA HEIGHT WEIGHT; |

--
DATA CLASS;
  INPUT F_NAME $ ID $ GENDER $ GPA HEIGHT WEIGHT;
DATA_LINES;
  Hector 123 M 3.5 59 155
  Nancy 328 F 3.7 52 99
  Edward 747 M 2.4 62 205
  Michelle 778 F 3.0 54 115
  Washington 289 M 3.5 60 180
;
PROC PRINT DATA=CLASS;
  TITLE 'Listing of CLASS data set';
RUN;

SOLUTION TO PROBLEM 4
DATA CLASS;

--
| The key here is to use the MISSOVER option which says that |
| if you reach the end of a data line and have not yet read |
| values for all your variables, set all the remaining values |
| to missing. |
--
The SAS Workbook Solutions

INFILE DATALINES MISSOVER;

| Try running this program without the MISSOVER option to see |
| what happens. |

INFILE DATALINES;

| Try running this program without the MISSOVER option to see |
| what happens. |

INPUT F_NAME $ ID $ GENDER $ GPA WEIGHT;
DATALINES;

George 123 M 3.5 155
328 F 3.7 52 99
Edward 747
Michelle 778 F 3.0 54
Sampson 289 M 3.5 60 180
;
PROC PRINT DATA=CLASS;
TITLE 'Listing of CLASS data set';
RUN;

SOLUTION TO PROBLEM 5

DATA CLASS;

| The key here is to use the DSD option which allows you to |
| read comma-delimited data, to treat 2 consecutive commas as |
| a missing value, and to remove the double quotes from quoted |
| strings. |

INFILE DATALINES DSD;

INPUT F_NAME $ ID $ GENDER $ GPA WEIGHT;
DATALINES;

George,123,M,3.5,,155
328,"F",3.7,52,99
"Edward",747,,,,
Michelle,778,F,3.0,54,,
Sampson,289,M,3.5,60,180
;
PROC PRINT DATA=CLASS;
TITLE 'Listing of CLASS data set';
RUN;

SOLUTION TO PROBLEM 6

DATA CLASS;

INPUT F_NAME $ ID $ GENDER $ GPA WEIGHT;
DATALINES;

George 123 M 3.5 59 155
Nancy 328 F 3.7 52 99
Edward 747 M 2.4 62 205
Michelle 778 F 3.0 54 115
Sampson 289 M 3.5 60 180
;
PROC PRINT DATA=CLASS;
TITLE 'Listing of CLASS data set';
RUN;
CHAPTER 1 Inputting Raw Data

--- SOLUTION TO PROBLEM 7 ---
DATA CLASS;
  /*-------------------------------------------------------------------*/
  | Use the PAD option to be sure that the SAS System will not |
  | try to read data from the next line. The need for this |
  | option will vary depending on which version of SAS Software |
  | you are running. It is a good idea to use the PAD option |
  | when reading fixed records from an external file. |
  /*-------------------------------------------------------------------*/
  INFILE DATALINES PAD;
  INPUT F_NAME $ 1-8 ID $ 13-15 GENDER $ 22 GPA 31-33 HEIGHT 39-40
  WEIGHT 49-51;
DATALINES;
  George 123 M 3.5 155
  328 F 3.7 52 99
  Edward 747
  Michelle 778 F 3.0 54
  Sampson 289 M 3.5 60 180
;
PROC PRINT DATA=CLASS;
  TITLE 'Listing of CLASS data set';
RUN;

--- SOLUTION TO PROBLEM 8 ---
DATA CLASS;
  /*-------------------------------------------------------------------*/
  | Use the PAD option to be sure that the SAS System will not |
  | try to read data from the next line. The need for this |
  | option will vary depending on which version of SAS Software |
  | you are running. It is a good idea to use the PAD option |
  | when reading fixed records from an external file. |
  /*-------------------------------------------------------------------*/
  INFILE DATALINES PAD;
  INPUT @1 F_NAME $8.
  @13 ID $3.
  @22 GENDER $1.
  @31 GPA 3.
  @39 HEIGHT 2.
  @49 WEIGHT 3.;
DATALINES;
  George 123 M 3.5 59 155
  Nancy 328 F 3.7 52 99
  Edward 747 M 2.4 62 205
  Michelle 778 F 3.0 54 115
  Sampson 289 M 3.5 60 180
;
PROC PRINT DATA=CLASS;
  TITLE 'Listing of CLASS data set';
RUN;

--- SOLUTION TO PROBLEM 9 ---
DATA SURVEY;
  /*-------------------------------------------------------------------*/
  | Use the PAD option to be sure that the SAS System will not |
  | try to read data from the next line. The need for this |
  | option will vary depending on which version of SAS Software |
  | you are running. It is a good idea to use the PAD option |
  | when reading fixed records from an external file. |
  /*-------------------------------------------------------------------*/
  INFILE DATALINES PAD;
  INPUT YEAR % 15-18 @1; *** Hold the line;
  IF YEAR = '1994' THEN
    INPUT ID $ 1-3
    GENDER $ 4
    PART $ 5
    VOTE $ 6
    NUM_TV 7-8;
  ELSE IF YEAR = '1995' THEN
    INPUT ID $ 1-3
    AGE 4-5
    GENDER $ 6
    PART $ 7
    VOTE $ 8
    NUM_TV 9-10;
The SAS Workbook Solutions

DATAINES:
0016553 1994
00923721 1995
010410 1994
00518994 1995
003064 1994
;
PROC PRINT;
TITLE 'Listing of SURVEY Data set';
RUN;

SOLUTION TO PROBLEM 10
DATA SURVEY;
   INPUT #1 SUBJECT $1-3
   $4 DOB MMDDYY8.
   $6 STATE $25-26
   $7 ZIP_CODE $40-44
   $2 NUMBER 5
   $A CAR1 $11-20
   $B CAR2 $21-30;
   /*---------------------------------------------*/
   | Alternative Code:
   |   | INPUT #1 #1 SUBJECT $3.
   |   | $4 DOB MMDDYY8.
   |   | $25 STATE $2.
   |   | $40 ZIP_CODE $5.
   |   | $2 #5 NUMBER 5.
   |   | $11 CAR1 $10.
   |   | $21 CAR2 $10.;
   /*---------------------------------------------*/
   FORMAT DOB MMDDYY8.;
   DATAINES:
   122120/21/46   NJ   08822
   123 2 Ford Oldsmobile
   23711/01/55 NY 11518
   237 1 Chevy
   PROC PRINT DATA=SURVEY;
   TITLE 'Listing of SURVEY data set';
   RUN;

SOLUTION TO PROBLEM 11
DATA TEMPER;
   INPUT TEMP #8;
   DATAINES:
   31 23 29 33 19 28
   33 39 43 44 28 27 29
   37 32 31 33 29
   ;
   PROC PRINT;
   TITLE 'Listing of TEMPER data set';
   RUN;

SOLUTION TO PROBLEM 12
DATA TEMP_DAY;
   INPUT DAY TEMP #8;
   DATAINES:
   5 21 6 23 7 29 8 33 9 19 10 28
   11 33 12 39 13 43 14 44 15 28 16 21 17 24 18 27 19 29
   20 37 21 32 22 31 23 33 24 29
   ;
PROC PRINT;
TITLE 'Listing of TEMP_DAY data set';
RUN;

SOLUTION TO PROBLEM 13
DATA RATS;
INPUT GROUP $ WEIGHT #;
DATALINES;
A 34 B 50 A 28 C 55
C 36 A 27 B 52 C 58 A 21 B 62
;
PROC PRINT;
TITLE 'Listing of RATS data set';
RUN;

SOLUTION TO PROBLEM 14
DATA RATS;
LENGTH GROUP $ 1;
RETAIN GROUP;
INPUT DUMMY $ #;
IF DUMMY IN ('A','B','C') THEN DO;
GROUP = DUMMY;
DELETE;
RETURN; *** RETURN not needed but OK to have;
END;
*** Wind up here only if a weight is read;
WEIGHT = INPUT (DUMMY,8.); *** Convert WEIGHT to numeric;
OUTPUT;
DROP DUMMY;
DATALINES;
A 34 B 58 52
62 C 55 A 27 21
C 36 58
;
PROC PRINT;
TITLE 'Listing of RATS data set';
RUN;

SOLUTION TO PROBLEM 15
DATA VARLIST;
INPUT &1 (01-05)/(2.)
&15 (DATE1-DATE3)(MMDDYY.)
&50 (X1-X3 Y1-Y3)(61.);
FORMAT DATE1-DATE3 MMDDYY.;
DATALINES;
1122334455 10/21/4611/13/4206/05/48 123456
9672347656 01/01/9501/02/9501/03/95 987654
;
PROC PRINT;
TITLE 'Listing of VARLIST data set';
RUN;

SOLUTION TO PROBLEM 16
DATA POINTER;
INPUT &1 (X1-X3)(2. +5)
&3 (Y1-Y3)(2. +5)
&5 (Z1-Z3)(63. +4);
DATALINES;
0102AA0304BBBB506CCC
2837ABC9676DEF8765GHI
;
The SAS Workbook Solutions

PROC PRINT;
TITLE 'Listing of POINTER data set';
RUN;