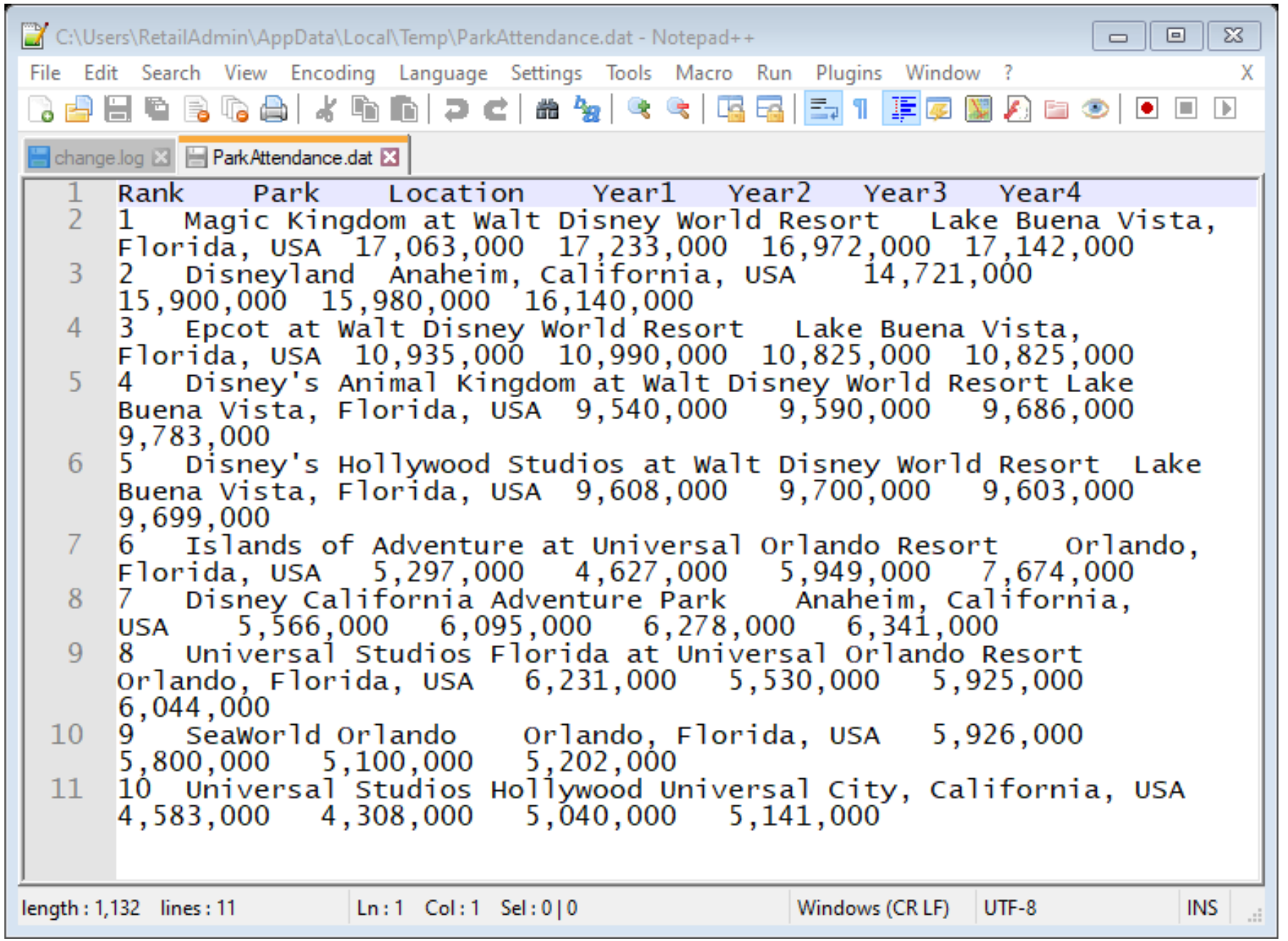
CTS2456C
Module 2 Lab

Practice Examples

Practice Exercise 1

- Annual attendance for the Top 10 amusement parks in North America is listed in the raw data file ParkAttendance.dat For each park, the data include the ranking, park name, location, and 4 years of attendance.
  a) Open the raw data file ParkAttendance.dat in a simple editor such as Wordpad and review. In a comment in your program, state the number of variables and observations.
  b) Use the IMPORT procedure to read the raw data file into SAS. View the log to verify that your data set has the same number of variables an observation as you stated in part a.
  c) Print the data set


C:\Users\RetailAdmin\AppData\Local\Temp\ParkAttendance.dat - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
change.log
ParkAttendance.dat
Rank Park Location Year1 Year2 Year3 Year4
1 Magic Kingdom at Walt Disney World Resort Lake Buena Vista, Florida, USA 17,063,000 17,233,000 16,972,000 17,142,000
2 Disneyland Anaheim, California, USA 14,721,000 15,900,000 15,980,000 16,140,000
3 Epcot at Walt Disney World Resort Lake Buena Vista, Florida, USA 10,935,000 10,990,000 10,825,000 10,825,000
4 Disney's Animal Kingdom at Walt Disney World Resort Lake Buena Vista, Florida, USA 9,540,000 9,590,000 9,686,000 9,783,000
5 Disney's Hollywood Studios at Walt Disney World Resort Lake Buena Vista, Florida, USA 9,608,000 9,700,000 9,603,000 9,699,000
6 Islands of Adventure at Universal Orlando Resort Orlando, Florida, USA 5,297,000 4,627,000 5,949,000 7,674,000
7 Disney California Adventure Park Anaheim, California, USA 5,566,000 6,095,000 6,278,000 6,341,000
8 Universal Studios Florida at Universal Orlando Resort Orlando, Florida, USA 6,231,000 5,530,000 5,925,000 6,044,000
9 SeaWorld Orlando Orlando, Florida, USA 5,926,000 5,800,000 5,100,000 5,202,000
10 Universal Studios Hollywood Universal City, California, USA 4,583,000 4,308,000 5,040,000 5,141,000
length : 1,132 lines : 11
Ln : 1 Col : 1 Sel : 0 | 0
Windows (CR LF)
UTF-8
INS

Practice Exercise 1 - SAS code

```
*Remember to run the libname program at the beginning of each session;

/*Practice Exercise 1
There are 7 variables and 10 observations in the ParkAttendance.dat file.
Use the IMPORT procedure to read the raw data file into SAS;
print the data set
*/
PROC IMPORT DATAFILE = '~/my_content/LSBP6Ex/Ch02/ParkAttendance.dat'
    OUT = pbdata.ParkAttendance
    DBMS = TAB REPLACE; *what happens if you use something besides TAB?;
RUN;

TITLE "Top 10 Amusement Park Attendance Data";

PROC PRINT DATA = pbdata.ParkAttendance; *hard to decipher error msgs;
RUN;
```

PROC IMPORT – the OUTPUT DATA tab

Table: PBDATA.PARKATTENDANCE | View: Column names | Filter: (none)

Columns

- Select all
- Rank
- Park
- Location
- Year1
- Year2
- Year3
- Year4

Total rows: 10 Total columns: 7 Rows 1-10

	Rank	Park	Location	
1	1	Magic Kingdom at Walt Disney World Resort	Lake Buena Vista, Florida, USA	17
2	2	Disneyland	Anaheim, California, USA	14
3	3	Epcot at Walt Disney World Resort	Lake Buena Vista, Florida, USA	10
4	4	Disney's Animal Kingdom at Walt Disney World Resort	Lake Buena Vista, Florida, USA	9
5	5	Disney's Hollywood Studios at Walt Disney World Resort	Lake Buena Vista, Florida, USA	9
6	6	Islands of Adventure at Universal Orlando Resort	Orlando, Florida, USA	5
7	7	Disney California Adventure Park	Anaheim, California, USA	5
8	8	Universal Studios Florida at Universal Orlando Resort	Orlando, Florida, USA	6
9	9	SeaWorld Orlando	Orlando, Florida, USA	5
10	10	Universal Studios Hollywood	Universal City, California, USA	4

PROC IMPORT – the OUTPUT DATA tab, con't.

Location	Year1	Year2	Year3	Year4
Lake Buena Vista, Florida, USA	17,063,000	17,233,000	16,972,000	17,142,000
Anaheim, California, USA	14,721,000	15,900,000	15,980,000	16,140,000
Lake Buena Vista, Florida, USA	10,935,000	10,990,000	10,825,000	10,825,000
Lake Buena Vista, Florida, USA	9,540,000	9,590,000	9,686,000	9,783,000
Lake Buena Vista, Florida, USA	9,608,000	9,700,000	9,603,000	9,699,000
Orlando, Florida, USA	5,297,000	4,627,000	5,949,000	7,674,000
Anaheim, California, USA	5,566,000	6,095,000	6,278,000	6,341,000
Orlando, Florida, USA	6,231,000	5,530,000	5,925,000	6,044,000
Orlando, Florida, USA	5,926,000	5,800,000	5,100,000	5,202,000
Universal City, California, USA	4,583,000	4,308,000	5,040,000	5,141,000

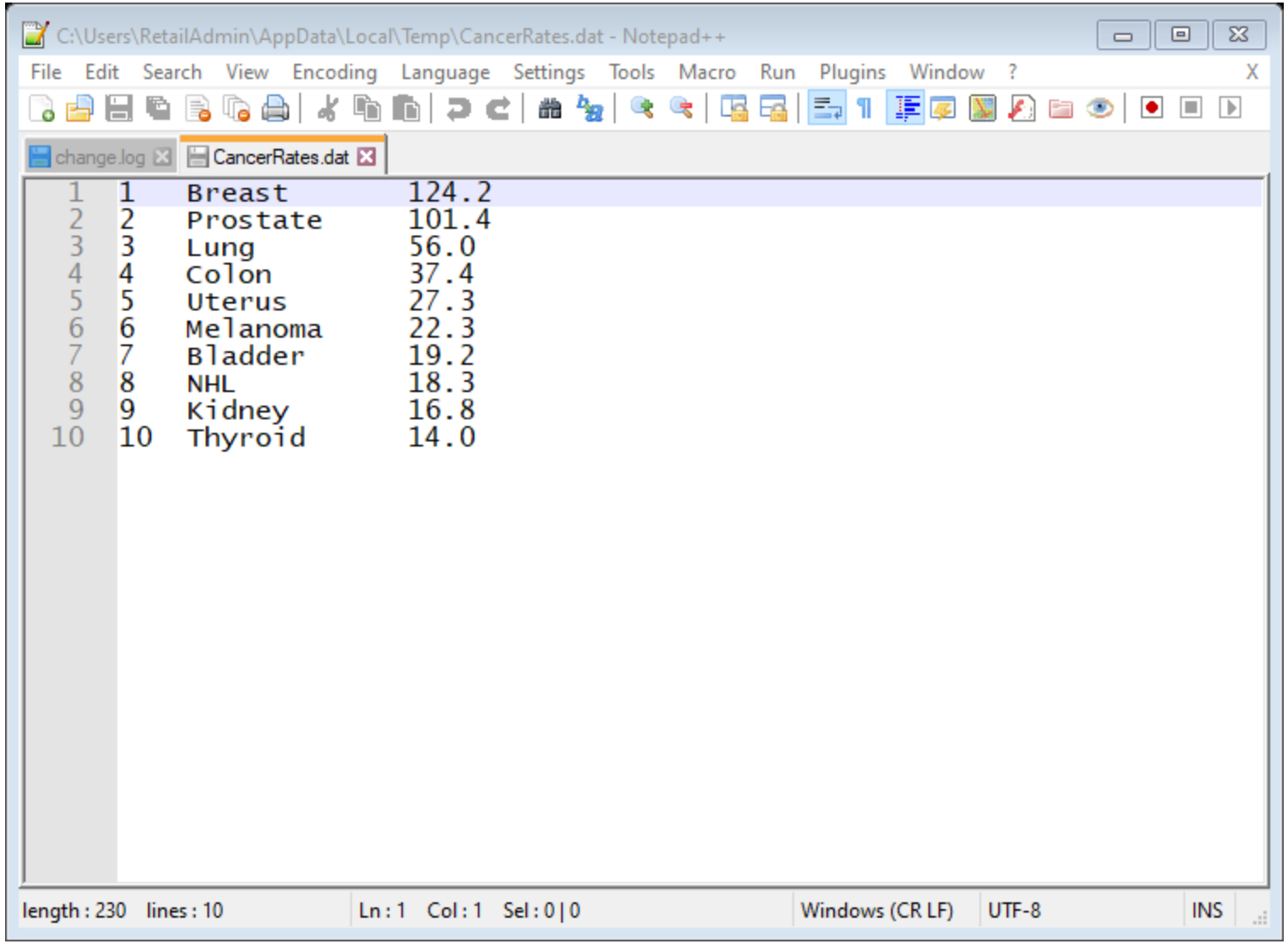
PROC PRINT – the RESULTS tab

Top 10 Amusement Park Attendance Data

Obs	Rank	Park	Location	Year1	Year2	Year3	Year4
1	1	Magic Kingdom at Walt Disney World Resort	Lake Buena Vista, Florida, USA	17,063,000	17,233,000	16,972,000	17,142,000
2	2	Disneyland	Anaheim, California, USA	14,721,000	15,900,000	15,980,000	16,140,000
3	3	Epcot at Walt Disney World Resort	Lake Buena Vista, Florida, USA	10,935,000	10,990,000	10,825,000	10,825,000
4	4	Disney's Animal Kingdom at Walt Disney World Resort	Lake Buena Vista, Florida, USA	9,540,000	9,590,000	9,686,000	9,783,000
5	5	Disney's Hollywood Studios at Walt Disney World Resort	Lake Buena Vista, Florida, USA	9,608,000	9,700,000	9,603,000	9,699,000
6	6	Islands of Adventure at Universal Orlando Resort	Orlando, Florida, USA	5,297,000	4,627,000	5,949,000	7,674,000
7	7	Disney California Adventure Park	Anaheim, California, USA	5,566,000	6,095,000	6,278,000	6,341,000
8	8	Universal Studios Florida at Universal Orlando Resort	Orlando, Florida, USA	6,231,000	5,530,000	5,925,000	6,044,000
9	9	SeaWorld Orlando	Orlando, Florida, USA	5,926,000	5,800,000	5,100,000	5,202,000
10	10	Universal Studios Hollywood	Universal City, California, USA	4,583,000	4,308,000	5,040,000	5,141,000

Practice Exercise 2

- The file CancerRates.dat contains data on the top 10 cancer sites in the US from the Centers for Disease Control and Prevention (CDC) website. These statistics are condensed across genders and races. The variables are ranking, cancer site, and incidence rate per 100,000 people.
- a. Open the raw data file CancerRates.dat in a simple editor. In a comment in your program, state the number of variables and observations.
- b. Read the raw data file into SAS. View the log to verify that your data set has the same number of variables and observations as you stated in part a.
- c. Print the data set.


C:\Users\RetailAdmin\AppData\Local\Temp\CancerRates.dat - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ? X
change.log
CancerRates.dat
1 1 Breast 124.2
2 2 Prostate 101.4
3 3 Lung 56.0
4 4 Colon 37.4
5 5 Uterus 27.3
6 6 Melanoma 22.3
7 7 Bladder 19.2
8 8 NHL 18.3
9 9 Kidney 16.8
10 10 Thyroid 14.0
length : 230 lines : 10
Ln : 1 Col : 1 Sel : 0 | 0
Windows (CR LF)
UTF-8
INS

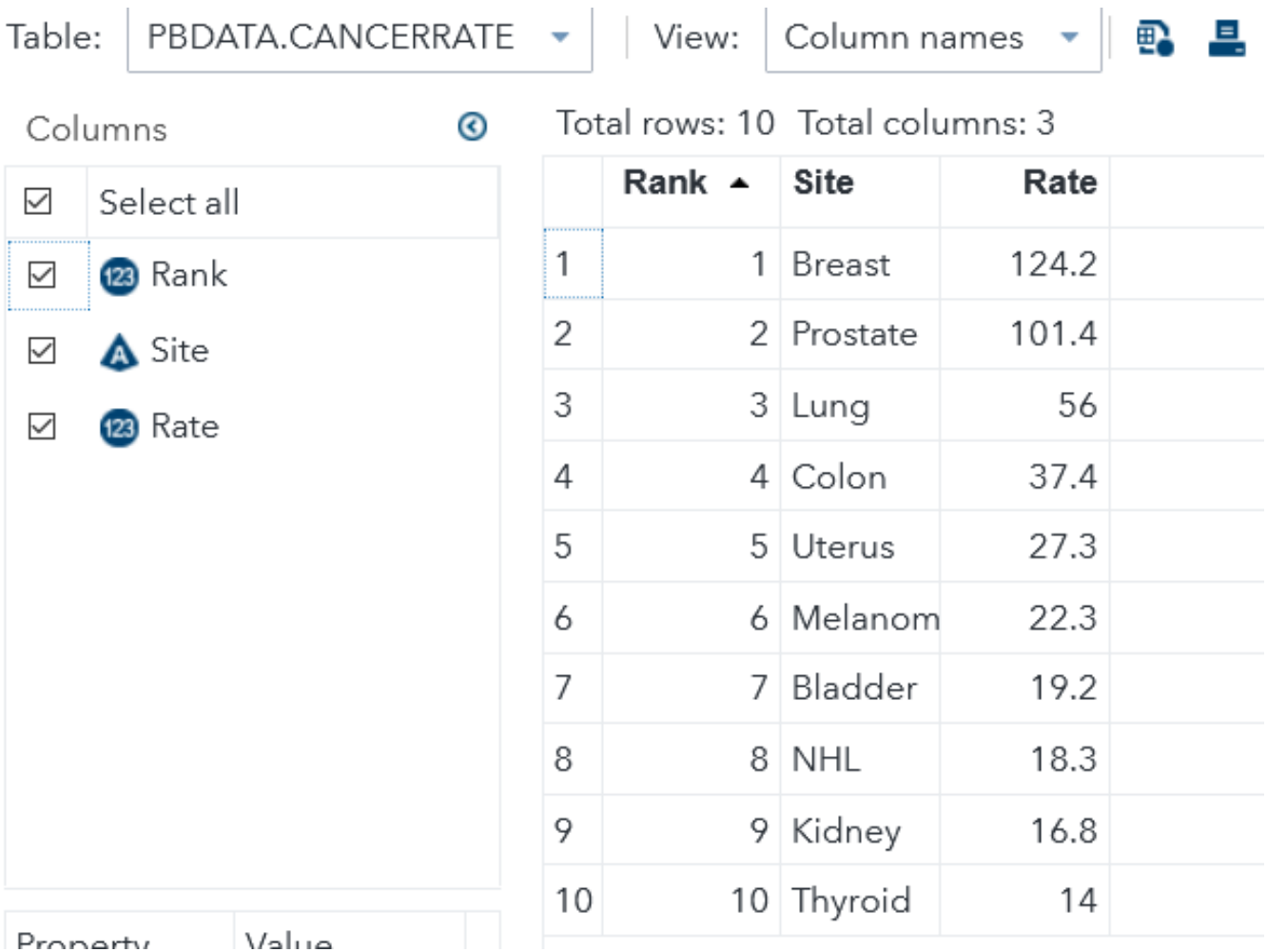
Practice Exercise 2 – SAS Code

```
/*Practice Exercise 2
  There are 3 variables and 10 observations, no column headings.
  Read the data in to SAS; then print the data set.
*/
DATA pbdata.CancerRate;
   INFILE '~/my_content/LSBP6Ex/Ch02/CancerRates.dat';
   INPUT Rank Site $ Rate;
RUN;

TITLE "Top 10 Cancer Sites in the Body";

PROC PRINT DATA = pbdata.CancerRate;
RUN;
```

DATA step – the OUTPUT DATA tab



	Rank	Site	Rate
1	1	Breast	124.2
2	2	Prostate	101.4
3	3	Lung	56
4	4	Colon	37.4
5	5	Uterus	27.3
6	6	Melanom	22.3
7	7	Bladder	19.2
8	8	NHL	18.3
9	9	Kidney	16.8
10	10	Thyroid	14

PROC PRINT – The RESULTS TAB

Top 10 Cancer Sites in the Body

Obs	Rank	Site	Rate
1	1	Breast	124.2
2	2	Prostate	101.4
3	3	Lung	56.0
4	4	Colon	37.4
5	5	Uterus	27.3
6	6	Melanoma	22.3
7	7	Bladder	19.2
8	8	NHL	18.3
9	9	Kidney	16.8
10	10	Thyroid	14.0

My PBDATA Library