You have some data that look as follows:

```sas
data study;
input
  id : $3.
  visit : mmddyy.
  chol
;,
format visit mmddyy10.;
datalines;
  001  10/15/2004      200
  002  10/15/2004      200
  003  10/15/2004      300
  004  10/15/2004      275
  005  10/15/2004      250
  002  11/10/2004      175
  002  11/10/2004      175
  002  11/10/2004      175
  002  11/10/2004      195
  004  11/13/2004      275
  003  11/14/2004      280
  004  12/14/2004      275
;
run;
```

Each record in the data file has an ID number; a date of visit; a cholesterol measurement. Some subjects have a single occurrence (IDs 001 and 005) while others have multiple occurrences. The following are some tasks that can be accomplished using FIRST. and LAST. variables. Use of these variables requires a SET statement + a BY statement, in this case...

```sas
set study;
by id;
```

Whenever you use a BY statement, SAS requires that the data set being used be sorted according to all the variables in the BY statement. Also, since you will be looking for first time you saw a subject, last time, etc., you also sort the data by VISIT within each ID...

```sas
proc sort data=study;
by id visit;
run;
proc print data=study;
run;
```

The data set now looks as follows...

```
Obs  id    visit    chol
  1   001    10/15/2004     200
  2   002    10/15/2004     200
  3   002    11/10/2004     175
  4   002    11/10/2004     175
  5   002    11/10/2004     175
  6   002    11/10/2004     195
  7   003    10/15/2004     300
  8   003    11/14/2004     280
  9   004    10/15/2004     275
 10   004    11/13/2004     275
 11   004    12/14/2004     275
 12   005    10/15/2004     250
```
One way to learn about FIRST. and LAST. variables is to print the data set showing the their values within each observation. Since FIRST. and LAST. variables ONLY EXIST FOR THE DURATION OF THE DATA STEP and are NOT ADDED TO THE DATA SET, you must create new variables that contain the values if the FIRST. and LAST. variables...

```sas
data fl;
set study;
by id;
first_id = first.id;
last_id  = last.id;
label
   first_id = 'first.id'
   last_id  = 'last.id'
;
run;

proc print data=fl label;
run;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>id</th>
<th>visit</th>
<th>chol</th>
<th>first_id</th>
<th>last_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>001</td>
<td>10/15/2004</td>
<td>200</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>002</td>
<td>10/15/2004</td>
<td>200</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>002</td>
<td>11/10/2004</td>
<td>175</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>002</td>
<td>11/10/2004</td>
<td>175</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>002</td>
<td>11/10/2004</td>
<td>175</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>002</td>
<td>11/10/2004</td>
<td>195</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>003</td>
<td>10/15/2004</td>
<td>300</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>003</td>
<td>11/14/2004</td>
<td>280</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>004</td>
<td>10/15/2004</td>
<td>275</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>004</td>
<td>11/13/2004</td>
<td>275</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>004</td>
<td>12/14/2004</td>
<td>275</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>005</td>
<td>10/15/2004</td>
<td>250</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The values of the FIRST. and LAST. variables are helpful in answering the questions about your data. For example, if you are looking for observations in which the variable FIRST.ID has a value of 1 (the first observation within each ID), you can use either...

```sas
if first.id;
```

or...

```sas
if first.id eq 1;
```

or...

```sas
if first.id ne 0;
```
#1 Create a new data set that contains one observation per ID --- the FIRST time each ID participated in your study.

look for observations where FIRST.ID has a value of 1

data study_f;
set study;
by id;
if first.id;
run;

FIRST VISIT

<table>
<thead>
<tr>
<th>Obs</th>
<th>id</th>
<th>visit</th>
<th>chol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>001</td>
<td>10/15/2004</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>002</td>
<td>10/15/2004</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>003</td>
<td>10/15/2004</td>
<td>300</td>
</tr>
<tr>
<td>4</td>
<td>004</td>
<td>10/15/2004</td>
<td>275</td>
</tr>
<tr>
<td>5</td>
<td>005</td>
<td>10/15/2004</td>
<td>250</td>
</tr>
</tbody>
</table>

#2 Create a new data set that contains one observation per ID --- the LAST time each ID participated in your study.

look for observations where LAST.ID has a value of 1

data study_l;
set study;
by id;
if last.id;
run;

LAST VISIT

<table>
<thead>
<tr>
<th>Obs</th>
<th>id</th>
<th>visit</th>
<th>chol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>001</td>
<td>10/15/2004</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>002</td>
<td>11/10/2004</td>
<td>195</td>
</tr>
<tr>
<td>3</td>
<td>003</td>
<td>11/14/2004</td>
<td>280</td>
</tr>
<tr>
<td>4</td>
<td>004</td>
<td>12/14/2004</td>
<td>275</td>
</tr>
<tr>
<td>5</td>
<td>005</td>
<td>10/15/2004</td>
<td>250</td>
</tr>
</tbody>
</table>

#3 Create a new data set that contains two observations per ID --- the FIRST and LAST time each ID participated in your study.

look for observations where FIRST.ID or LAST.ID has a value of 1

* first and last time you saw each ID;

data study_fl;
set study;
by id;
if first.id or last.id;
run;

FIRST AND LAST VISIT

<table>
<thead>
<tr>
<th>Obs</th>
<th>id</th>
<th>visit</th>
<th>chol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>001</td>
<td>10/15/2004</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>002</td>
<td>10/15/2004</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>002</td>
<td>11/10/2004</td>
<td>195</td>
</tr>
<tr>
<td>4</td>
<td>003</td>
<td>10/15/2004</td>
<td>300</td>
</tr>
<tr>
<td>5</td>
<td>003</td>
<td>11/14/2004</td>
<td>280</td>
</tr>
<tr>
<td>6</td>
<td>004</td>
<td>10/15/2004</td>
<td>275</td>
</tr>
<tr>
<td>7</td>
<td>004</td>
<td>12/14/2004</td>
<td>275</td>
</tr>
<tr>
<td>8</td>
<td>005</td>
<td>10/15/2004</td>
<td>250</td>
</tr>
</tbody>
</table>
#4 Create two data sets --- one with all subjects who only have ONE observation in the data set, one with subjects who have MULTIPLE observations in the data set.

Identify ONE observation subjects as those with both FIRST.ID and LAST.ID having the value 1; all others are MULTIPLE observation subjects.

```plaintext
data single multiple;
set study;
by id;
if first.id and last.id then output single;
else output multiple;
run;
```

**SINGLE VISIT**

<table>
<thead>
<tr>
<th>Obs</th>
<th>id</th>
<th>visit</th>
<th>chol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>001</td>
<td>10/15/2004</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>005</td>
<td>10/15/2004</td>
<td>250</td>
</tr>
</tbody>
</table>

**MULTIPLE VISITS**

<table>
<thead>
<tr>
<th>Obs</th>
<th>id</th>
<th>visit</th>
<th>chol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>002</td>
<td>10/15/2004</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>002</td>
<td>11/10/2004</td>
<td>175</td>
</tr>
<tr>
<td>3</td>
<td>002</td>
<td>11/10/2004</td>
<td>175</td>
</tr>
<tr>
<td>4</td>
<td>002</td>
<td>11/10/2004</td>
<td>175</td>
</tr>
<tr>
<td>5</td>
<td>002</td>
<td>11/10/2004</td>
<td>195</td>
</tr>
<tr>
<td>6</td>
<td>003</td>
<td>10/15/2004</td>
<td>300</td>
</tr>
<tr>
<td>7</td>
<td>003</td>
<td>11/14/2004</td>
<td>280</td>
</tr>
<tr>
<td>8</td>
<td>004</td>
<td>10/15/2004</td>
<td>275</td>
</tr>
<tr>
<td>9</td>
<td>004</td>
<td>11/13/2004</td>
<td>275</td>
</tr>
<tr>
<td>10</td>
<td>004</td>
<td>12/14/2004</td>
<td>275</td>
</tr>
</tbody>
</table>

#5 Create one data set from the original data set STUDY --- the FIRST time each ID participated in your study for only those subjects with multiple visits.

```plaintext
data study_fm;
set study;
by id;
if first.id and not last.id;
run;
```

**FIRST VISIT OF MULTIPLE VISITS**

<table>
<thead>
<tr>
<th>Obs</th>
<th>id</th>
<th>visit</th>
<th>chol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>002</td>
<td>10/15/2004</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>003</td>
<td>10/15/2004</td>
<td>300</td>
</tr>
<tr>
<td>3</td>
<td>004</td>
<td>10/15/2004</td>
<td>275</td>
</tr>
</tbody>
</table>

**NOTE:** You could have used the new data set MULTIPLE created in #4 and just specify...

```plaintext
if first.id;
```
There should not be any repeated dates with any ID — create a data set with repeated dates within any of the IDs.

Once again, it is helpful to know the values of the FIRST. and LAST. variables in the data step...

data fl;
set study;
by id visit;
first_id = first.id;
last_id  = last.id;
first_visit = first.visit;
last_visit  = last.visit;
label
  first_id = 'first.id'
  last_id  = 'last.id'
  first_visit = 'first.visit'
  last_visit  = 'last.visit'
;
run;

proc print data=fl label;
run;

How can you identify repeated dates within each ID...

data repeats;
set study;
by id visit;
if not (first.visit and last.visit);
run;

REPEATED VISITS WITHIN AN ID
Obs id visit chol first_id last_id first_visit last_visit
1 001 10/15/2004 200 1 1 1 1
2 002 10/15/2004 200 1 0 1 1
3 002 11/10/2004 175 0 0 1 0
4 002 11/10/2004 175 0 0 0 0
5 002 11/10/2004 175 0 0 0 0
6 002 11/10/2004 195 0 1 0 1
7 003 10/15/2004 300 1 0 1 1
8 003 11/14/2004 280 0 1 1 1
9 004 10/15/2004 275 1 0 1 1
10 004 11/13/2004 275 0 0 1 1
11 004 12/14/2004 275 0 1 1 1
12 005 10/15/2004 250 1 1 1 1
#6 As in #5, one common use of FIRST. and LAST. variables is to identify duplicate observations within a data set. For example, if you are working with the vital statistics death file, each observation contains a social security number (SSN). There should only be one observation per SSN --- no repeated 'deaths' or individuals with identical SSNs --- create one data set with observations with duplicate SSNs and another with unique SSNs.

* assume the data set with social security numbers is named DEATHS;
* assume the variable with the social security number is named SSN;

```
proc sort data=deaths;
by ssn;
run;

data duplicates unique;
set deaths;
by ssn;
if not (first.ssn and last.ssn) then output duplicates;
else output unique;
run;
```