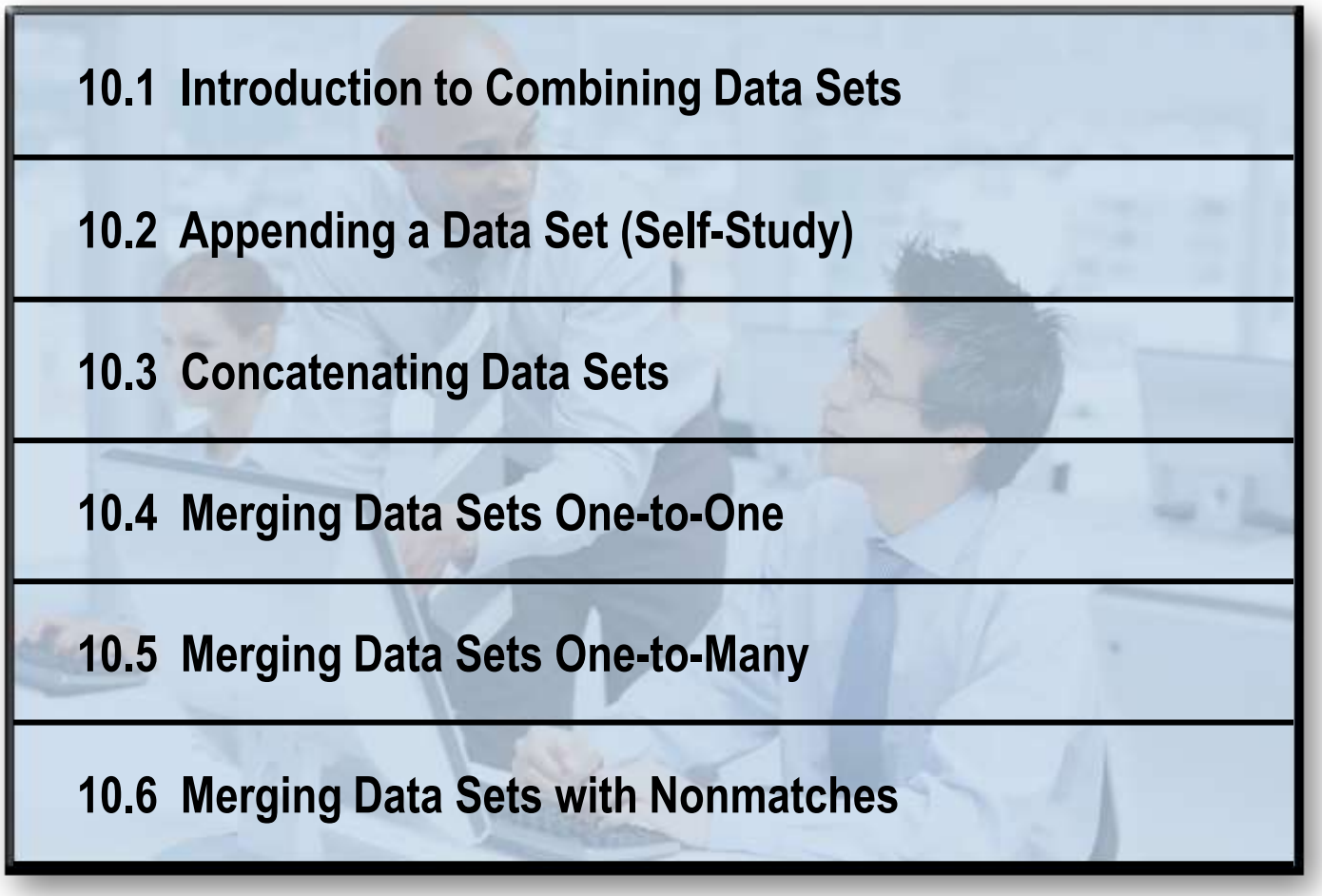
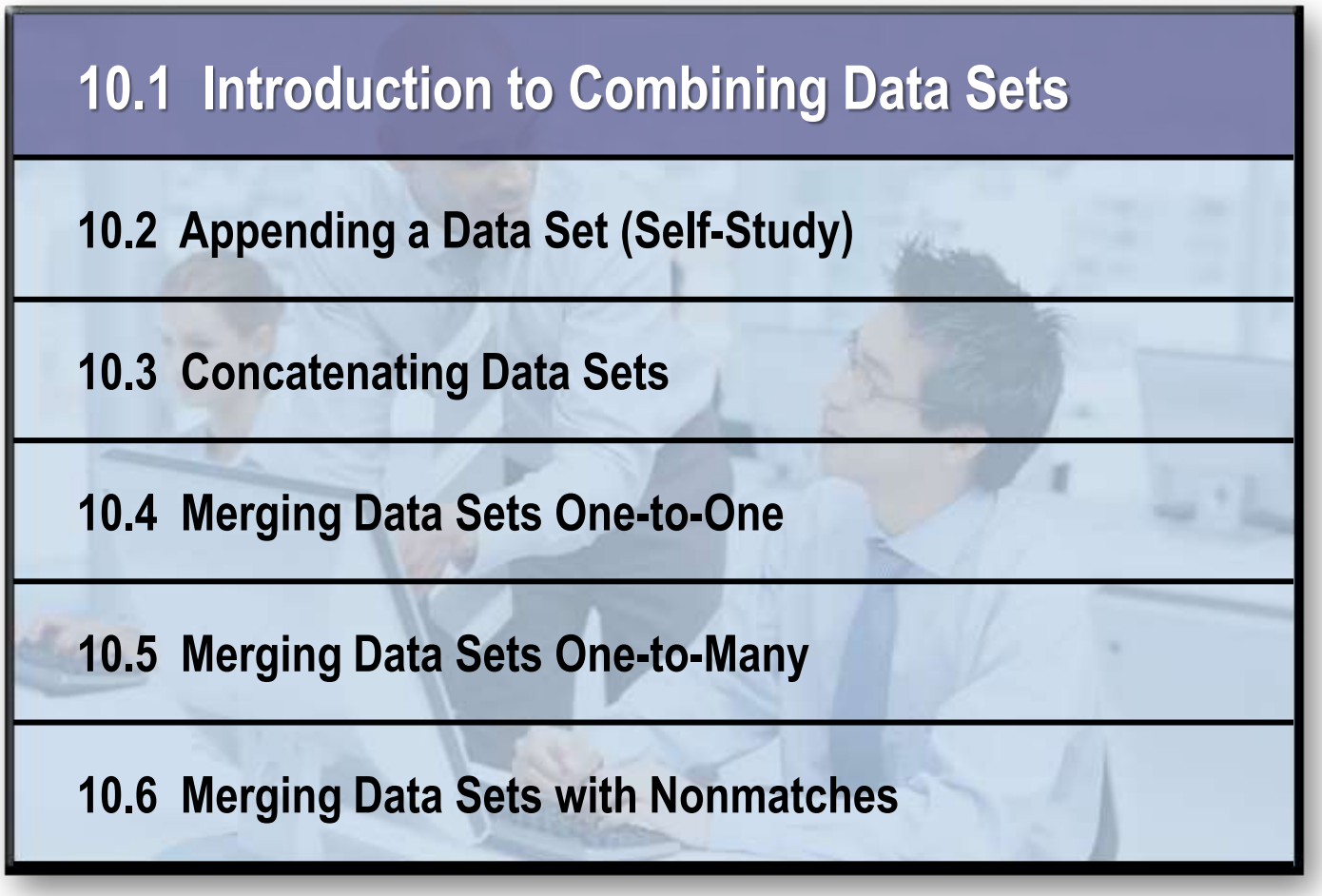
STAT604

Lesson SAS 11

P1-Chapter 10: Combining SAS Data Sets



Chapter 10: Combining SAS Data Sets

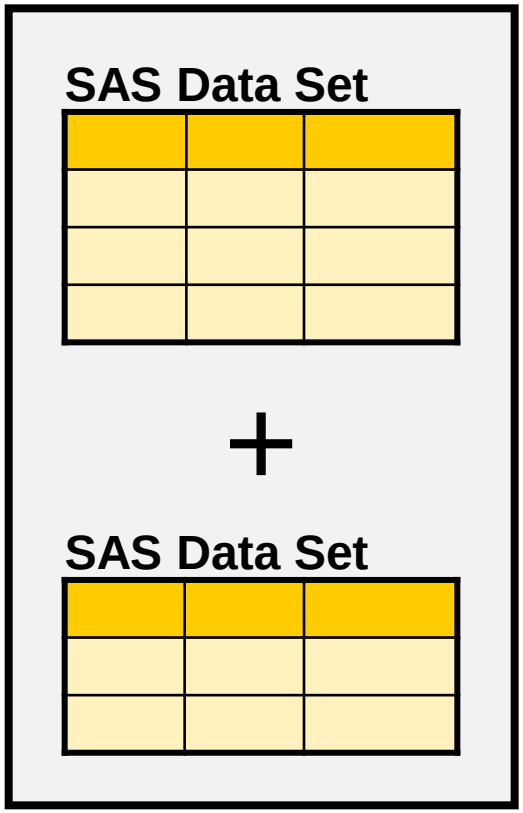


Objectives

- Define the methods for combining SAS data sets.

Appending and Concatenating

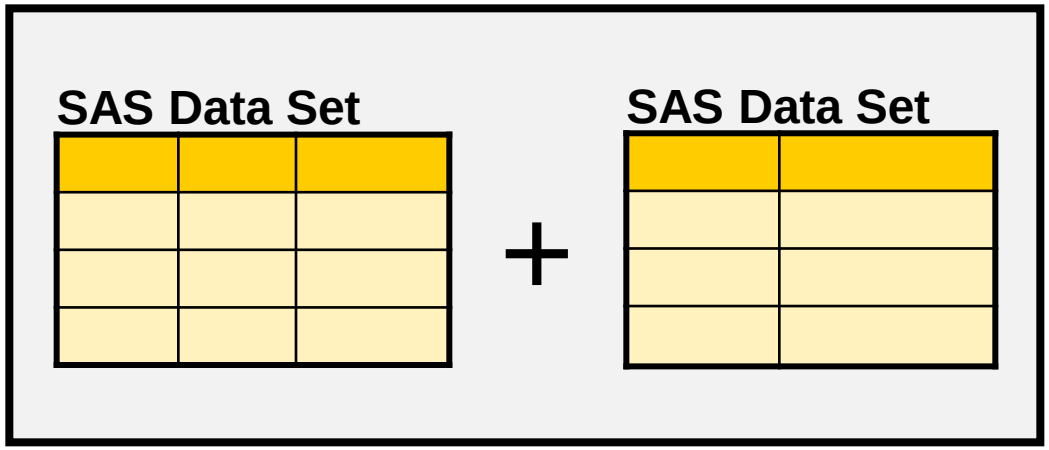
Appending and concatenating involves combining SAS data sets, one after the other, into a single SAS data set.



- *Appending* adds the observations in the second data set directly to the end of the original data set.
- *Concatenating* copies all observations from the first data set and then copies all observations from one or more successive data sets into a new data set.

Merging

Merging involves combining observations from two or more SAS data sets into a single observation in a new SAS data set.



Observations can be merged based on their positions in the original data sets or merged by one or more common variables.

Example: Appending a Data Set

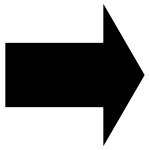
One data set is appended to a master data set.

Emps

First	Gender	HireYear
Stacey	F	2006
Gloria	F	2007
James	M	2007

Emps2008

First	Gender	HireYear
Brett	M	2008
Renee	F	2008



Emps

First	Gender	HireYear
Stacey	F	2006
Gloria	F	2007
James	M	2007
Brett	M	2008
Renee	F	2008

Example: Concatenating Data Sets

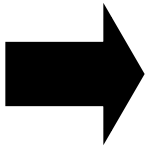
Two data sets are concatenated to create a new data set.

EmpsDK

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark

EmpsFR

First	Gender	Country
Pierre	M	France
Sophie	F	France



EmpsAll

First	Gender	Country
Lars	M	Denmark
Kari	F	Denmark
Jonas	M	Denmark
Pierre	M	France
Sophie	F	France

Example: Merging Data Sets

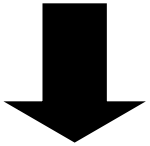
Two data sets are merged to create a new data set.

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneH

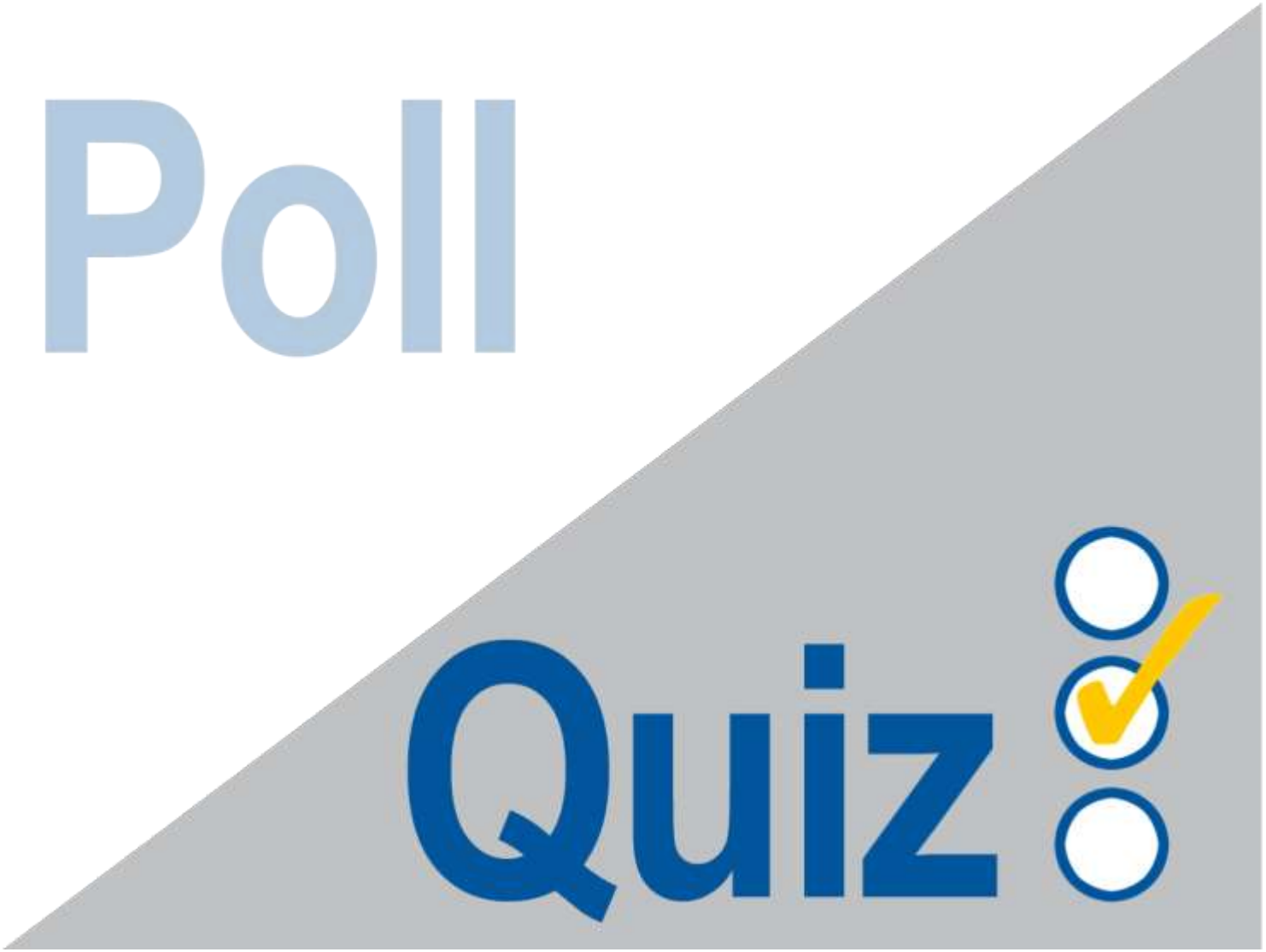
EmpID	Phone
121150	+61(2)5555-1793
121151	+61(2)5555-1849
121152	+61(2)5555-1665



EmpsAUH

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1793
Kylie	F	121151	+61(2)5555-1849
Birin	M	121152	+61(2)5555-1665

Poll



10.01 Quiz

Which method (appending, concatenating, or merging) should be used for the given business scenario?

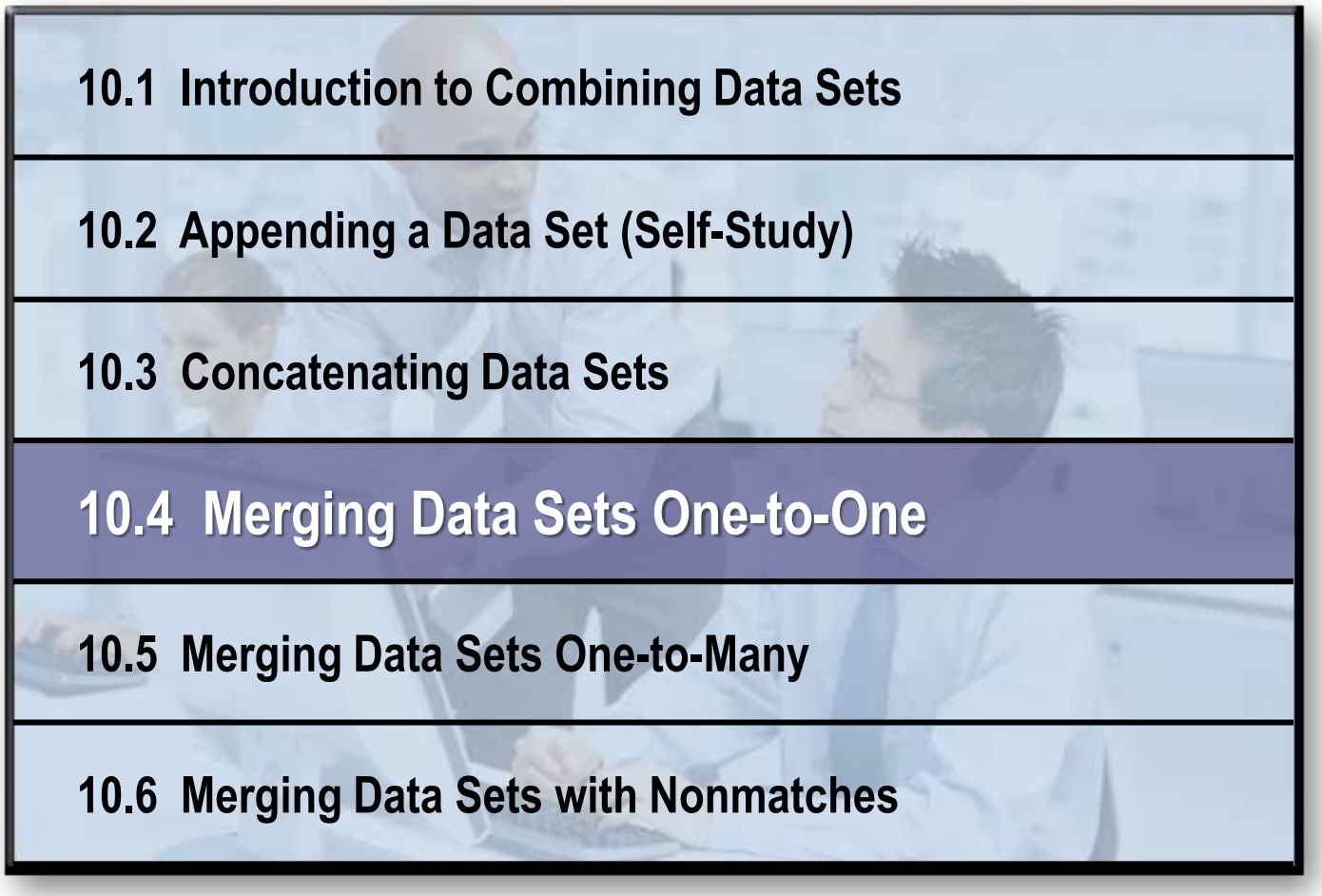
	Business Scenario	Method
1	The **JanSales**, **FebSales**, and **MarSales** data sets need to be combined to create the **Qtr1Sales** data set.	
2	The **Sales** data set needs to be combined with the **Target** data set by **month** to compare the sales data to the target data.	
3	The **OctSales** data set needs to be added to the **YTD** data set.	

10.01 Quiz – Correct Answer

Which method (appending, concatenating, or merging) should be used for the given business scenario?

	Business Scenario	Method
1	The **JanSales**, **FebSales**, and **MarSales** data sets need to be combined to create the **Qtr1Sales** data set.	**concatenating**
2	The **Sales** data set needs to be combined with the **Target** data set by **month** to compare the sales data to the target data.	**merging**
3	The **OctSales** data set needs to be added to the **YTD** data set.	**appending**

P1-Chapter 10: Combining SAS Data Sets

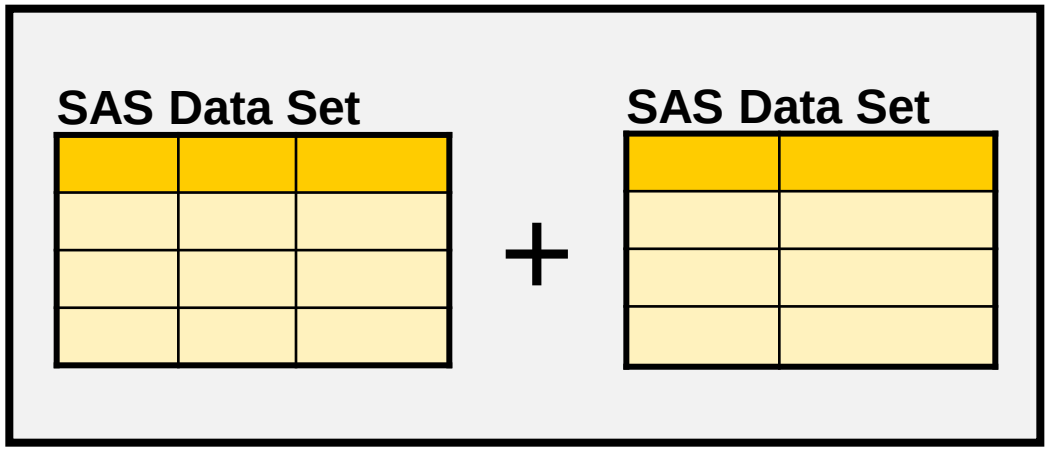


Objectives

- Define the different types of match-merging.
- Prepare data sets for merging using the SORT procedure.
- Merge SAS data sets one-to-one based on a common variable by using the MERGE and BY statements in a DATA step.

Merging

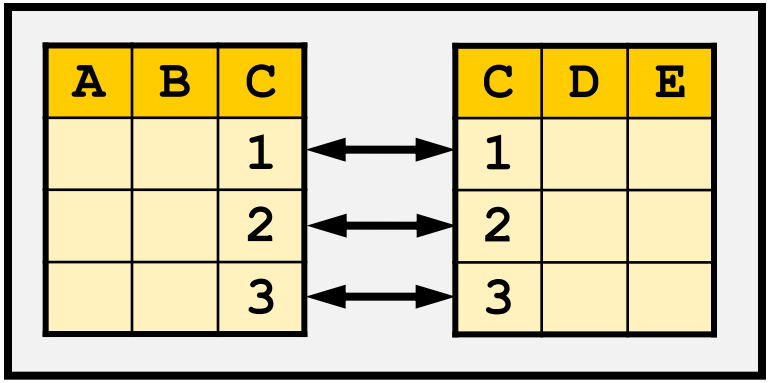
Merging involves combining observations from two or more SAS data sets into a single observation in a new SAS data set.

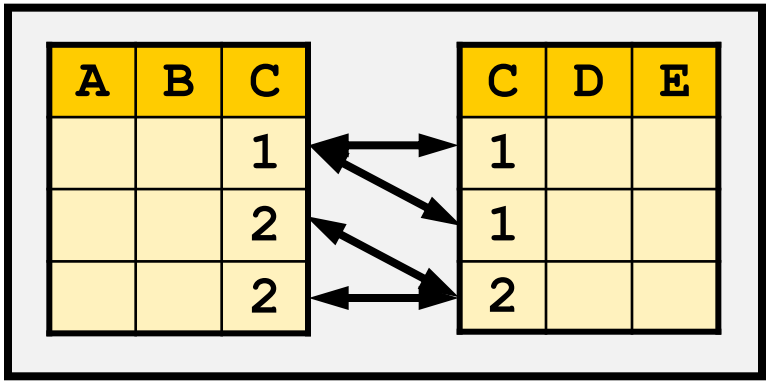


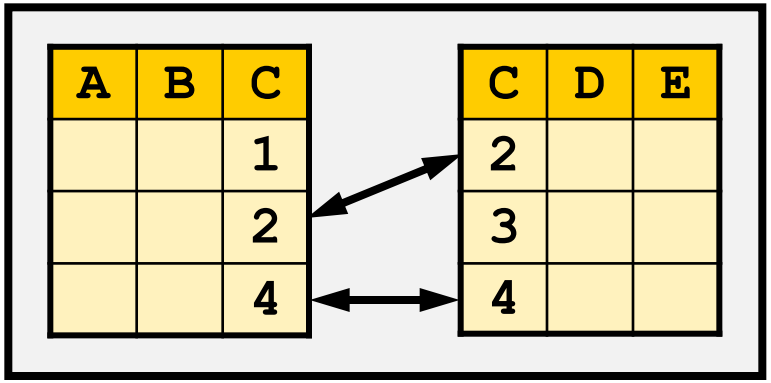
Observations can be merged based on their positions in the original data sets or merged by one or more common variables.

Match-Merging

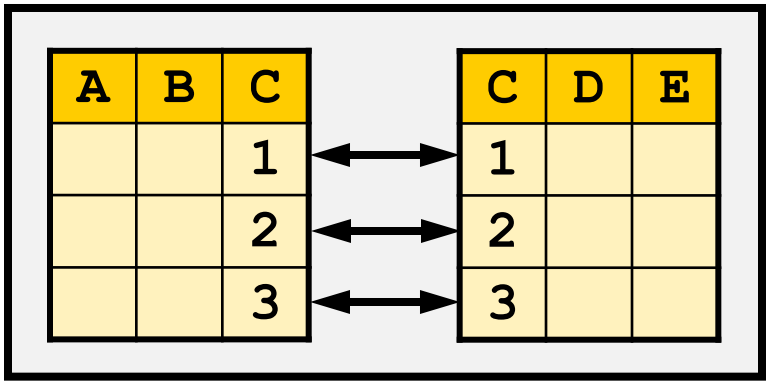
Match-merging combines observations from two or more SAS data sets into a single observation in a new data set based on the values of one or more common variables.





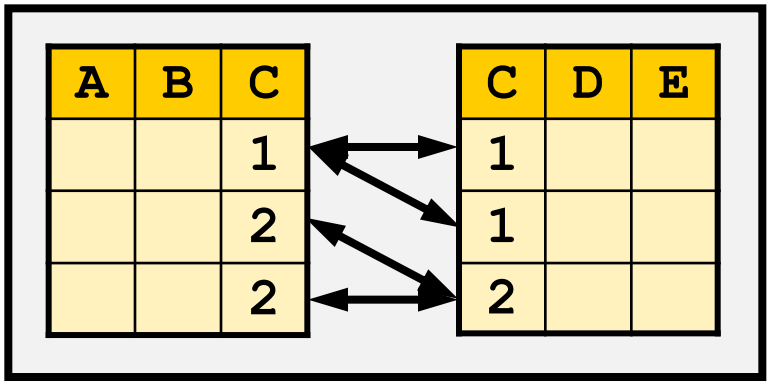


Match-Merging



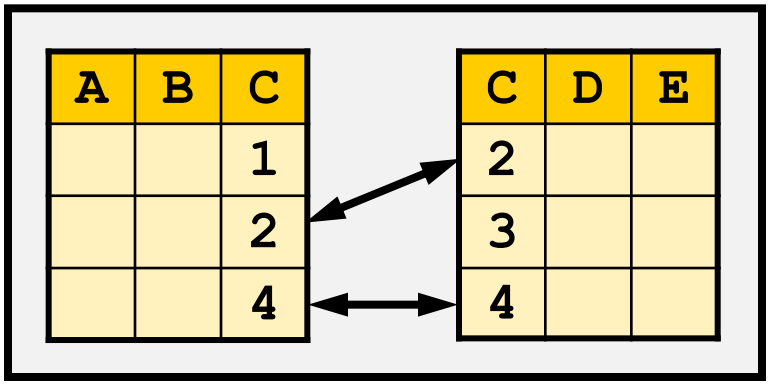
One-to-One

A single observation in one data set is related to one and only one observation from another data set based on the values of one or more selected variables.



One-to-Many or Many-to-One

A single observation in one data set is related to more than one observation from another data set based on the values of one or more selected variables and vice versa.



Nonmatches

At least one single observation in one data set is unrelated to any observation from another data set based on the values of one or more selected variables.

Match-Merging

In order to perform match-merging, the observations in each data set must be sorted by the one or more common variables that are being matched.

General form of the SORT procedure:

```
PROC SORT   DATA=input-SAS-data-set
                <OUT=output-SAS-data-set>;
      BY <DESCENDING> by-variable(s);
RUN;
```

The *SORT procedure* orders SAS data set observations by the values of one or more variables.

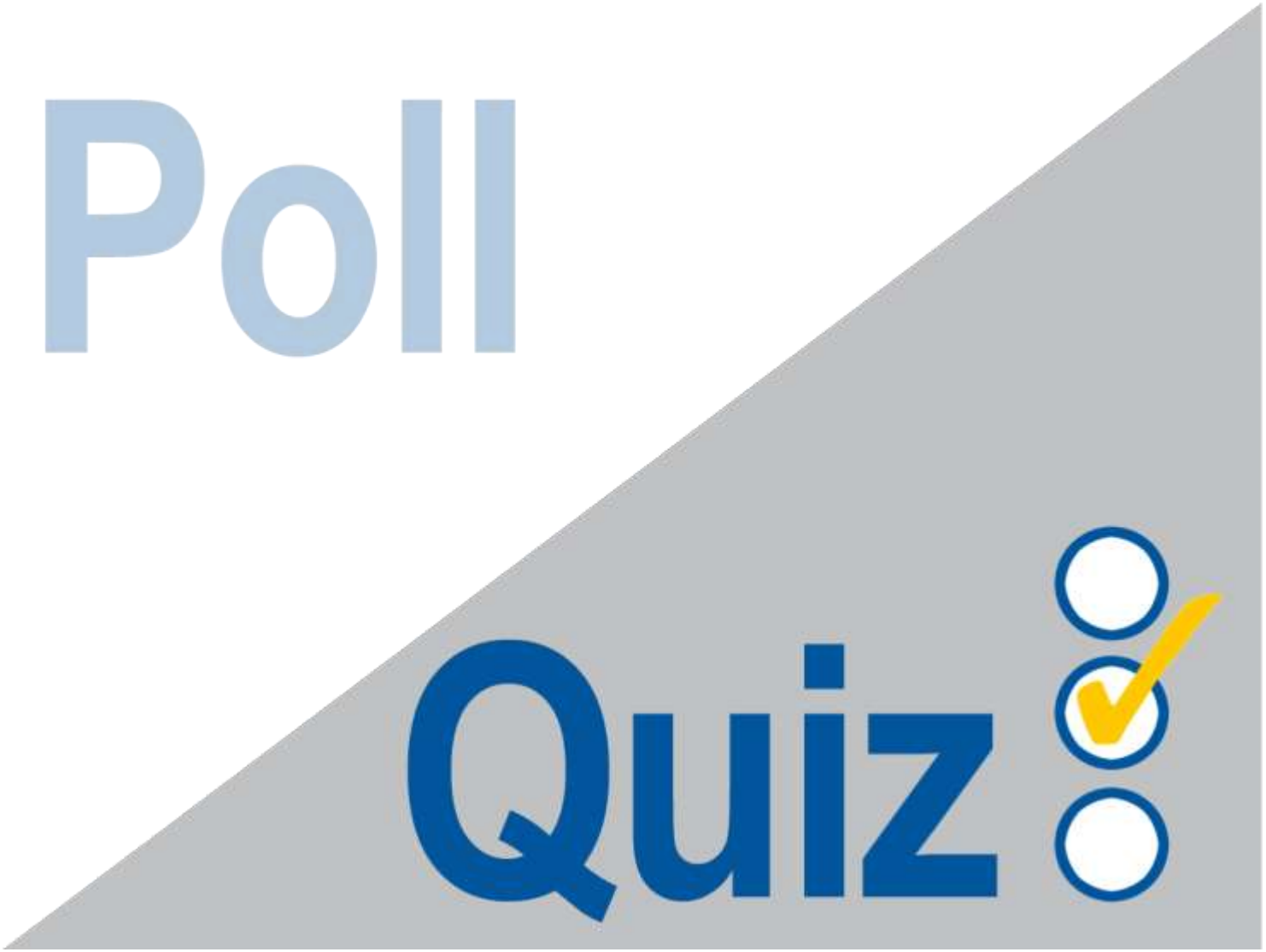
The SORT Procedure

```
PROC SORT  DATA=input-SAS-data-set
                <OUT=output-SAS-data-set>;
     BY <DESCENDING> by-variable(s);
RUN;
```

The SORT procedure

- rearranges the observations in a SAS data set
- either replaces the original data set or creates a new data set
- can sort on multiple variables
- can sort in ascending (default) or descending order
- does not generate printed output.

Poll



10.08 Quiz

Which step is sorting the observations in a SAS data set and overwriting the same SAS data set?

a.
```
proc sort data=work.EmpsAU
          out=work.sorted;
   by First;
run;
```

b.
```
proc sort data=work.EmpsAU
          out=orion.EmpsAU;
   by First;
run;
```

c.
```
proc sort data=work.EmpsAU;
   by First;
run;
```

10.08 Quiz – Correct Answer

Which step is sorting the observations in a SAS data set and overwriting the same SAS data set?

a.
```
proc sort data=work.EmpsAU
          out=work.sorted;
   by First;
run;
```

b.
```
proc sort data=work.EmpsAU
          out=orion.EmpsAU;
   by First;
run;
```

c.
```
proc sort data=work.EmpsAU;
   by First;
run;
```

The BY Statement

The *BY statement* specifies the sorting variables.

- PROC SORT first arranges the data set by the values in ascending order, by default, of the first BY variable.
- PROC SORT then arranges any observations that have the same value of the first BY variable by the values of the second BY variable in ascending order.
- This sorting continues for every specified BY variable.

The *DESCENDING option* reverses the sort order for the variable that immediately follows in the statement so that observations are sorted from the largest value to the smallest value.

The BY Statement

BY statement examples:

```
by Last First;
```

```
by descending Last First;
```

```
by Last descending First;
```

```
by descending Last descending First;
```

The MERGE and BY Statements

The *MERGE statement* in a DATA step joins observations from two or more SAS data sets into single observations.

```
DATA SAS-data-set;
      MERGE SAS-data-set1 SAS-data-set2 . . .;
      BY <DESCENDING> by-variable(s);
      <additional SAS statements>
RUN;
```

A *BY statement* after the MERGE statement performs a match-merge.

The MERGE and BY Statements

Requirements when two or more SAS data sets are specified in the MERGE statement:

- The variables in the BY statement must be common to all data sets.
- The data sets that are listed in the MERGE statement must be sorted in the order of the values of the variables that are listed in the BY statement.

One-to-One Merge

Merge **EmpsAU** and **PhoneH** by **EmpID** to create a new data set named **EmpsAUH**.

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneH

EmpID	Phone
121150	+61(2)5555-1793
121151	+61(2)5555-1849
121152	+61(2)5555-1665

The data sets are sorted by **EmpID**.

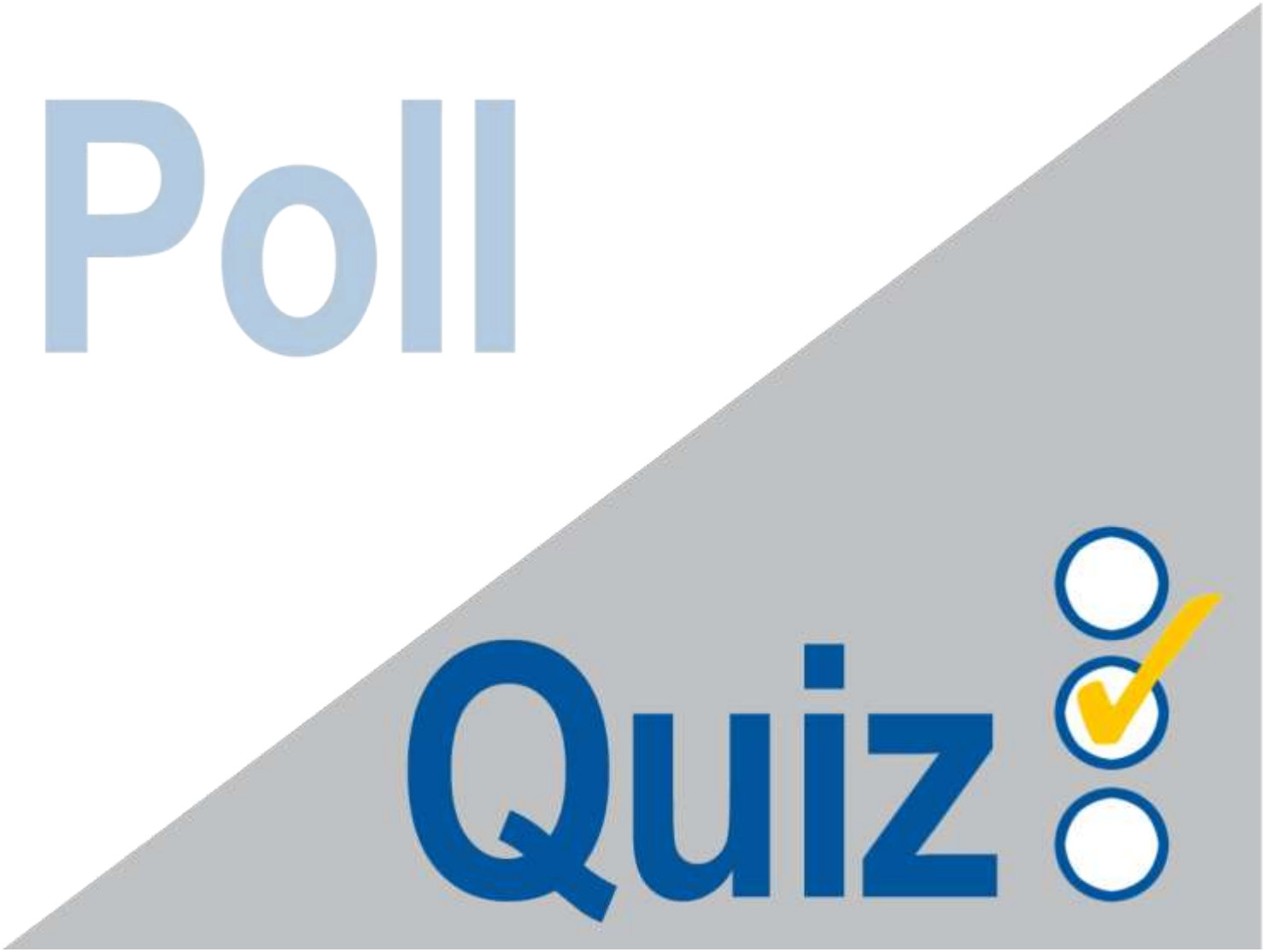
```
data EmpsAUH;
   merge EmpsAU PhoneH;
   by EmpID;
run;
```

Final Results

EmpsAUH

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1793
Kylie	F	121151	+61(2)5555-1849
Birin	M	121152	+61(2)5555-1665

Poll



10.10 Quiz

- Complete the program to match-merge the sorted SAS data sets referenced in the PROC SORT steps.

```
proc sort data=orion.employee_payroll
          out=work.payroll;
   by Employee_ID;
run;

proc sort data=orion.employee_addresses
          out=work.addresses;
   by Employee_ID;
run;

data work.payadd;
   merge

run;
```

10.10 Quiz – Correct Answer

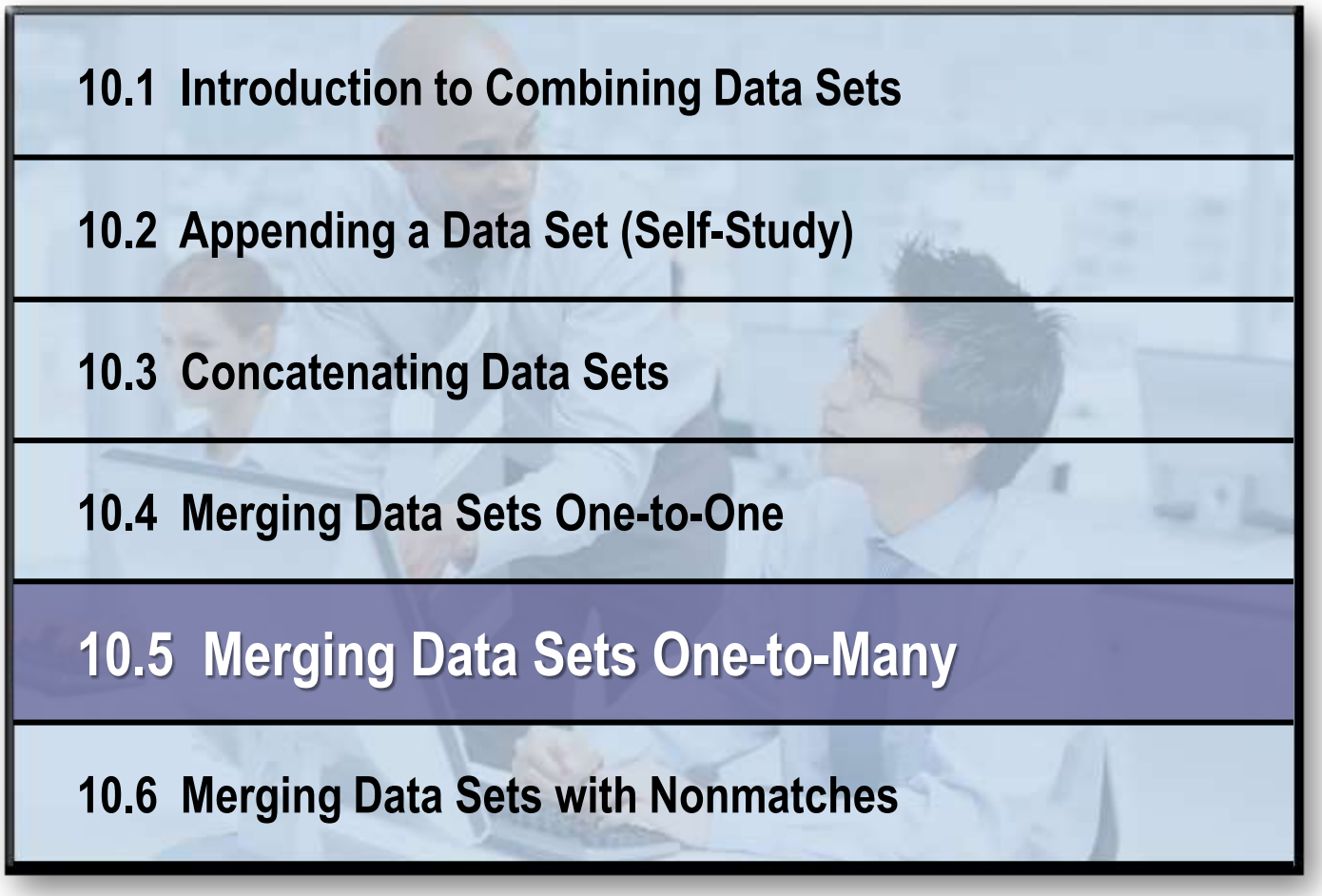
What are the modified, completed statements?

```
proc sort data=orion.employee_payroll
          out=work.payroll;
   by Employee_ID;
run;

proc sort data=orion.employee_addresses
          out=work.addresses;
   by Employee_ID;
run;

data work.payadd;
   merge work.payroll work.addresses;
   by Employee_ID;
run;
```

Chapter 10: Combining SAS Data Sets



Objectives

- Merge SAS data sets one-to-many based on a common variable by using the MERGE and BY statements in a DATA step.

One-to-Many Merge

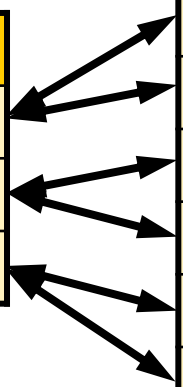
Merge **EmpsAU** and **PhoneHW** by **EmpID** to create a new data set named **EmpsAUHW**.

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666



```
data EmpsAUHW;
   merge EmpsAU PhoneHW;
   by EmpID;
run;
```

The data sets are sorted by **EmpID**.

p110d06

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU
    by EmpID;
run;
```

Initialize PDV

PDV

First	Gender	EmpID	Type	Phone
		.		

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Do the **EmpID**s match?

Yes

PDV

First	Gender	EmpID	Type	Phone
		.		

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Reads one observation from each matching data set

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
   merge EmpsAU PhoneHW;
   by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Do the **EmpID**s match?

No

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Is either **EmpID** the same as the **EmpID** currently in the PDV?

Yes

PDV

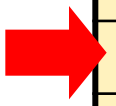
First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152



PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Reads the observation from the appropriate data set

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Work	+61(2)5555-1794

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
   merge EmpsAU PhoneHW;
   by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Work	+61(2)5555-1794

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Do the **EmpID**s match?

Yes

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Work	+61(2)5555-1794

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Is the **EmpID** the same as the **EmpID** currently in the PDV?

No

PDV

First	Gender	EmpID	Type	Phone
Togar	M	121150	Work	+61(2)5555-1794

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Reinitialize PDV

PDV

First	Gender	EmpID	Type	Phone
		.		

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Reads one observation from each matching data set

PDV

First	Gender	EmpID	Type	Phone
Kylie	F	121151	Home	+61(2)5555-1849

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
   merge EmpsAU PhoneHW;
   by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Kylie	F	121151	Home	+61(2)5555-1849

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Do the **EmpID**s match?

No

PDV

First	Gender	EmpID	Type	Phone
Kylie	F	121151	Home	+61(2)5555-1849

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Is either **EmpID** the same as the **EmpID** currently in the PDV?

Yes

PDV

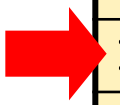
First	Gender	EmpID	Type	Phone
Kylie	F	121151	Home	+61(2)5555-1849

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152



PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
    merge EmpsAU PhoneHW;
    by EmpID;
run;
```

Reads the observation from the appropriate data set

PDV

First	Gender	EmpID	Type	Phone
Kylie	F	121151	Work	+61(2)5555-1850

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
   merge EmpsAU PhoneHW;
   by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Type	Phone
Kylie	F	121151	Work	+61(2)5555-1850

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneHW

EmpID	Type	Phone
121150	Home	+61(2)5555-1793
121150	Work	+61(2)5555-1794
121151	Home	+61(2)5555-1849
121151	Work	+61(2)5555-1850
121152	Home	+61(2)5555-1665
121152	Work	+61(2)5555-1666

```
data EmpsAUHW;
   merge EmpsAU Ph
   by EmpID;
run;
```

Continue until EOF on both data sets

PDV

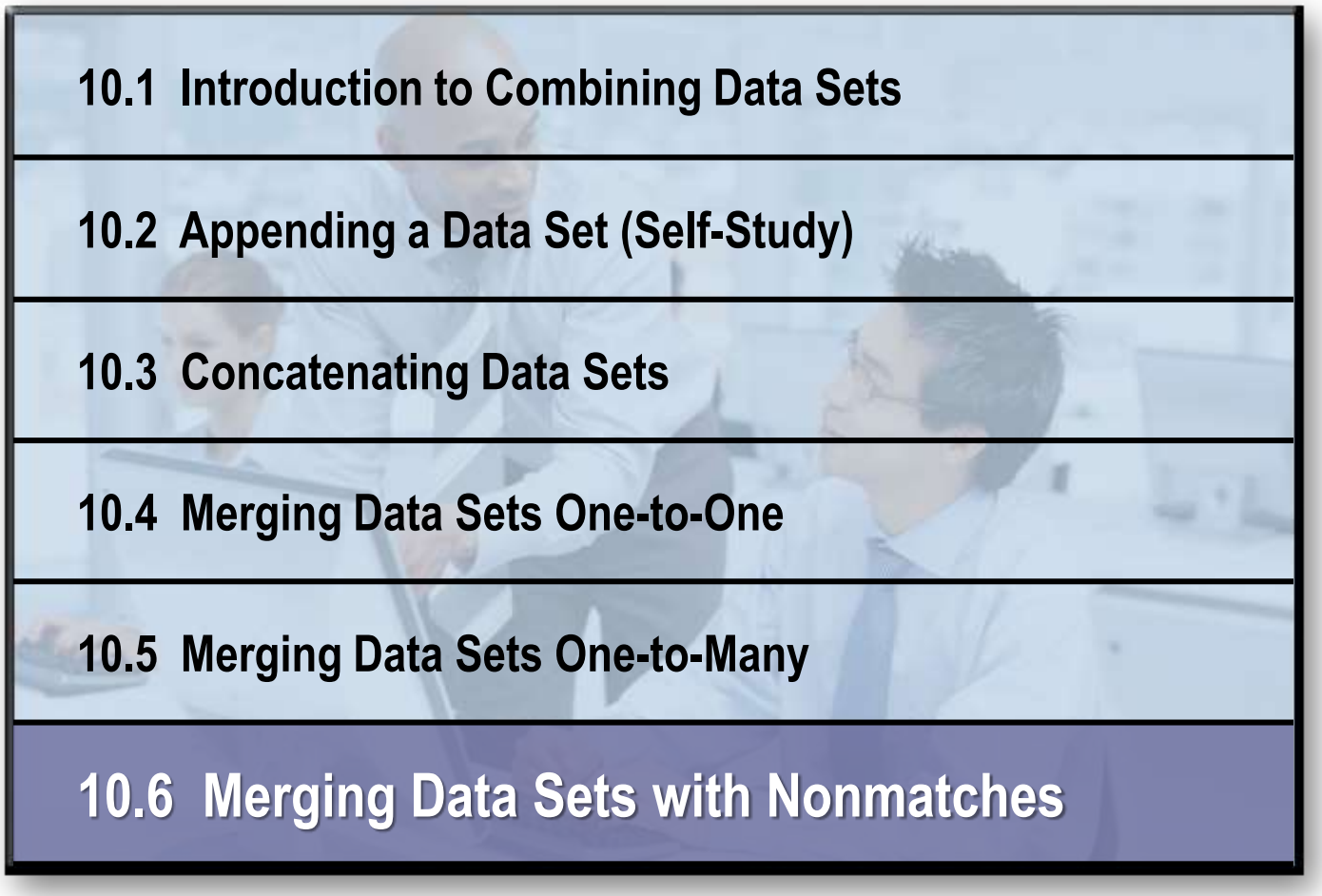
First	Gender	EmpID	Type	Phone
Kylie	F	121151	Work	+61(2)5555-1850

Final Results

EmpsAUHW

First	Gender	EmpID	Type	Phone
Togar	M	121150	Home	+61(2)5555-1793
Togar	M	121150	Work	+61(2)5555-1794
Kylie	F	121151	Home	+61(2)5555-1849
Kylie	F	121151	Work	+61(2)5555-1850
Birin	M	121152	Home	+61(2)5555-1665
Birin	M	121152	Work	+61(2)5555-1666

Chapter 10: Combining SAS Data Sets



Objectives

- Control the observations in the output data set by using the IN= option.
- Output observations to multiple data sets using the IN= option and the OUTPUT statement.
- Compare the results of a many-to-many merge based on using the DATA step or the SQL procedure.

Nonmatches Merge

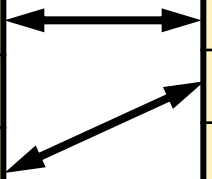
Merge **EmpsAU** and **PhoneC** by **EmpID** to create a new data set named **EmpsAUC**.

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348



The data sets are sorted by **EmpID**.

```
data EmpsAUC;
   merge EmpsAU PhoneC;
   by EmpID;
run;
```

p110d07

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsA
    by EmpID;
run;
```

Initialize PDV

PDV

First	Gender	EmpID	Phone
		.	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Do the **EmpID**s match?

Yes

PDV

First	Gender	EmpID	Phone
		.	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Reads one observation from each matching data set

PDV

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Do the **EmpID**s match?

No

PDV

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Is either **EmpID** the same as the **EmpID** currently in the PDV?

No

PDV

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Reinitialize PDV

PDV

First	Gender	EmpID	Phone
		.	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Which **EmpID** sequentially comes first?

121151

PDV

First	Gender	EmpID	Phone
		.	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Reads the observation from the **EmpID** that sequentially comes first

PDV

First	Gender	EmpID	Phone
Kylie	F	121151	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Phone
Kylie	F	121151	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Do the **EmpID**s match?

Yes

PDV

First	Gender	EmpID	Phone
Kylie	F	121151	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Is either **EmpID** the same as the **EmpID** currently in the PDV?

No

PDV

First	Gender	EmpID	Phone
Kylie	F	121151	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Reinitialize PDV

PDV

First	Gender	EmpID	Phone
		.	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Reads one observation from each matching data set

PDV

First	Gender	EmpID	Phone
Birin	M	121152	+61(2)5555-1667

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Phone
Birin	M	121152	+61(2)5555-1667

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

EOF

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Is the **EmpID** the same as the **EmpID** currently in the PDV?

No

PDV

First	Gender	EmpID	Phone
Birin	M	121152	+61(2)5555-1667

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
...n	M	121152

EOF

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Reinitialize PDV

PDV

First	Gender	EmpID	Phone
		.	

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
...n	M	121152

EOF

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Reads the observation from the appropriate data set

PDV

First	Gender	EmpID	Phone
		121153	+61(2)5555-1348

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
...n	M	121152

EOF

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

Implicit OUTPUT;
Implicit RETURN;

PDV

First	Gender	EmpID	Phone
		121153	+61(2)5555-1348

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
...n	M	121152

EOF

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
...53	+61(2)5555-1348

EOF

```
data EmpsAUC;
    merge EmpsAU PhoneC;
    by EmpID;
run;
```

PDV

First	Gender	EmpID	Phone
		121153	+61(2)5555-1348

Final Results

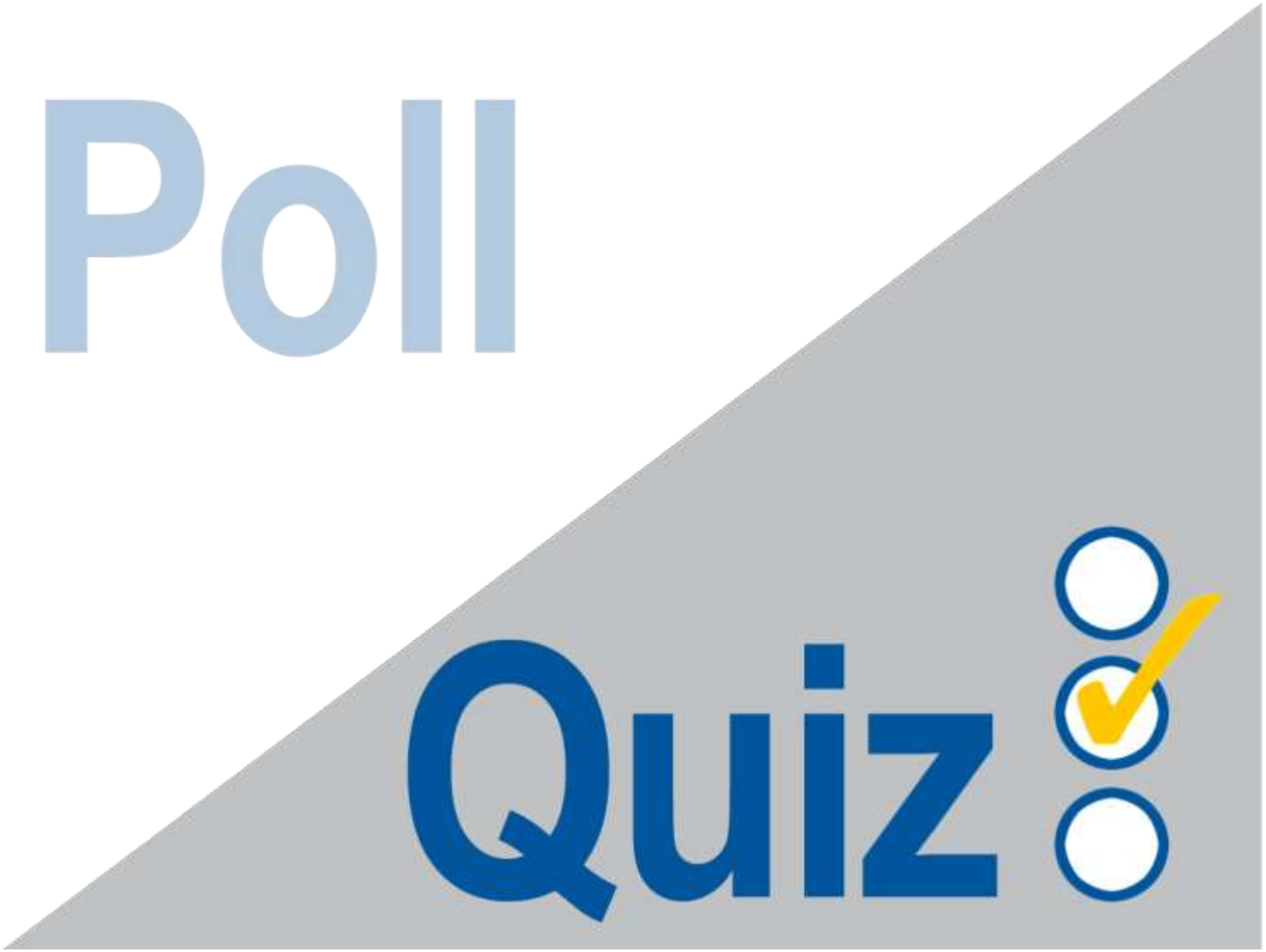
EmpsAUC

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795
Kylie	F	121151	
Birin	M	121152	+61(2)5555-1667
		121153	+61(2)5555-1348

The final results include matches and nonmatches.

- Matches are observations that contain data from both input data sets.
- Nonmatches are observations that contain data from only one input data set.

Poll



10.11 Quiz

How many observations in the final data set **EmpsAUC** are considered nonmatches?

a. 1
b. 2
c. 3
d. 4

EmpsAUC

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795
Kylie	F	121151	
Birin	M	121152	+61(2)5555-1667
		121153	+61(2)5555-1348

10.11 Quiz – Correct Answer

How many observations in the final data set **EmpsAUC** are considered nonmatches?

a. 1
b. 2
c. 3
d. 4

EmpsAUC

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795
Kylie	F	121151	
Birin	M	121152	+61(2)5555-1667
		121153	+61(2)5555-1348

The IN= Data Set Option

The *IN= data set option* creates a variable that indicates whether the data set contributed data to the current observation.

General form of the IN= data set option:

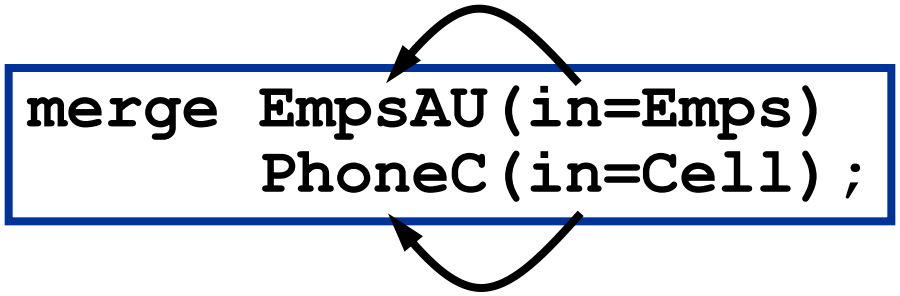
SAS-data-set (IN = *variable*)

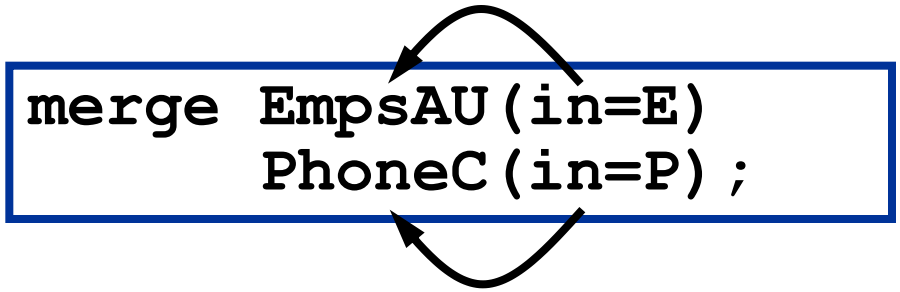
variable is a temporary numeric variable that has two possible values:

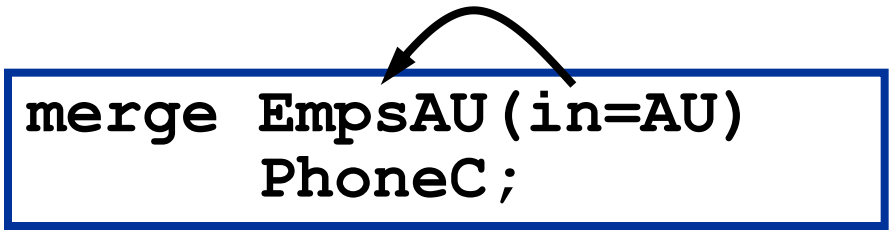
0	indicates that the data set did **not** contribute to the current observation.
1	indicates that the data set **did** contribute to the current observation.

The IN= Data Set Option

MERGE statement examples:







Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
   merge EmpsAU(in=Emps)
         PhoneC(in=Cell);
   by EmpID;
run;
```

PDV

First	Gender	EmpID	D	Emps	Phone	D	Cell
Togar	M	121150		1	+61(2)5555-1795		1

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
   merge EmpsAU(in=Emps)
         PhoneC(in=Cell);
   by EmpID;
run;
```

PDV

First	Gender	EmpID	D	Emps	Phone	D	Cell
Kylie	F	121151		1			0

...

Execution

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

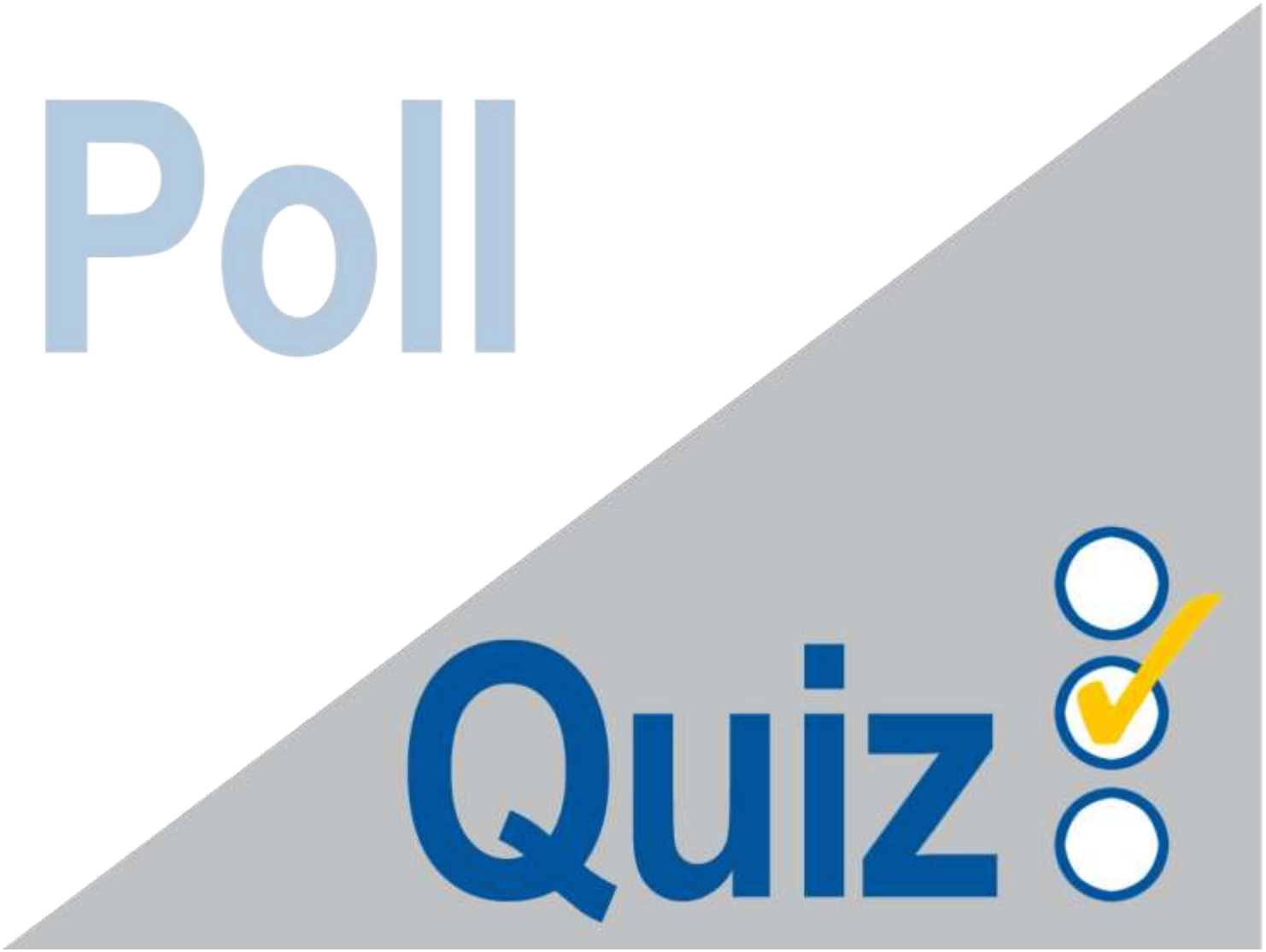
```
data EmpsAUC;
   merge EmpsAU(in=Emps)
         PhoneC(in=Cell);
   by EmpID;
run;
```

PDV

First	Gender	EmpID	D	Emps	Phone	D	Cell
Birin	M	121152		1	+61(2)5555-1667		1

...

Poll



10.12 Quiz

What are the values of **Emps** and **Cell**?

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
   merge EmpsAU(in=Emps)
         PhoneC(in=Cell);
   by EmpID;
run;
```

PDV

First	Gender	EmpID	D	Emps	Phone	D	Cell
		121153			+61(2)5555-1348		

10.12 Quiz – Correct Answer

What are the values of **Emps** and **Cell**?

EmpsAU

First	Gender	EmpID
Togar	M	121150
Kylie	F	121151
Birin	M	121152

PhoneC

EmpID	Phone
121150	+61(2)5555-1795
121152	+61(2)5555-1667
121153	+61(2)5555-1348

```
data EmpsAUC;
   merge EmpsAU(in=Emps)
         PhoneC(in=Cell);
   by EmpID;
run;
```

PDV

First	Gender	EmpID	D	Emps	Phone	D	Cell
		121153		0	+61(2)5555-1348		1

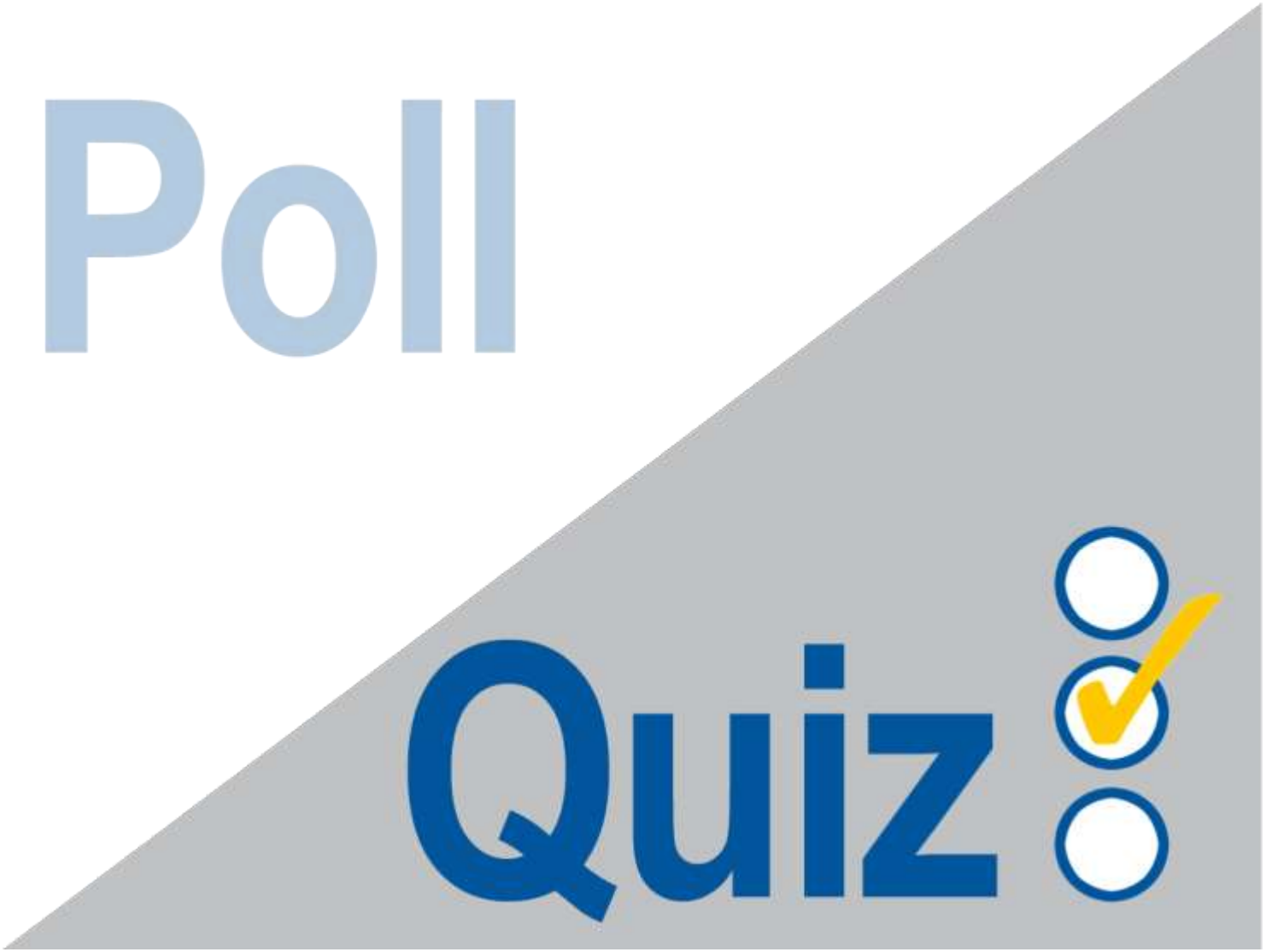
PDV Results

PDV

First	Gender	EmpID	Emps	Phone	Cell
Togar	M	121150	1	+61(2)5555-1795	1
Kylie	F	121151	1		0
Birin	M	121152	1	+61(2)5555-1667	1
		121153	0	+61(2)5555-1348	1

The variables created with the IN= data set option are only available during execution and are not written to the SAS data set.

Poll



10.13 Quiz

Which subsetting IF statement can be added to the DATA step to only output the matches?

a. `if Emps=1 and Cell=0;`

b. `if Emps=1 and Cell=1;`

c. `if Emps=1;`

d. `if Cell=0;`

PDV

First	Gender	EmpID	D Emps	Phone	D Cell
Togar	M	121150	1	+61(2)5555-1795	1
Kylie	F	121151	1		0
Birin	M	121152	1	+61(2)5555-1667	1
		121153	0	+61(2)5555-1348	1

10.13 Quiz – Correct Answer

Which subsetting IF statement can be added to the DATA step to only output the matches?

a. `if Emps=1 and Cell=0;`

b. `if Emps=1 and Cell=1;`

c. `if Emps=1;`

d. `if Cell=0;`

PDV

First	Gender	EmpID	D Emps	Phone	D Cell
Togar	M	121150	1	+61(2)5555-1795	1
Kylie	F	121151	1		0
Birin	M	121152	1	+61(2)5555-1667	1
		121153	0	+61(2)5555-1348	1

Matches Only

```
data EmpsAUC;
    merge EmpsAU(in=Emps)
          PhoneC(in=Cell);
    by EmpID;
    if Emps=1 and Cell=1;
run;
```

EmpsAUC

First	Gender	EmpID	Phone
Togar	M	121150	+61(2)5555-1795
Birin	M	121152	+61(2)5555-1667

Nonmatches from EmpsAU Only

```
data EmpsAUC;
   merge EmpsAU(in=Emps)
         PhoneC(in=Cell);
   by EmpID;
   if Emps=1 and Cell=0;
run;
```

EmpsAUC

First	Gender	EmpID	Phone
Kylie	F	121151	

Nonmatches from PhoneC Only

```
data EmpsAUC;
   merge EmpsAU(in=Emps)
         PhoneC(in=Cell);
   by EmpID;
   if Emps=0 and Cell=1;
run;
```

EmpsAUC

First	Gender	EmpID	Phone
		121153	+61(2)5555-1348

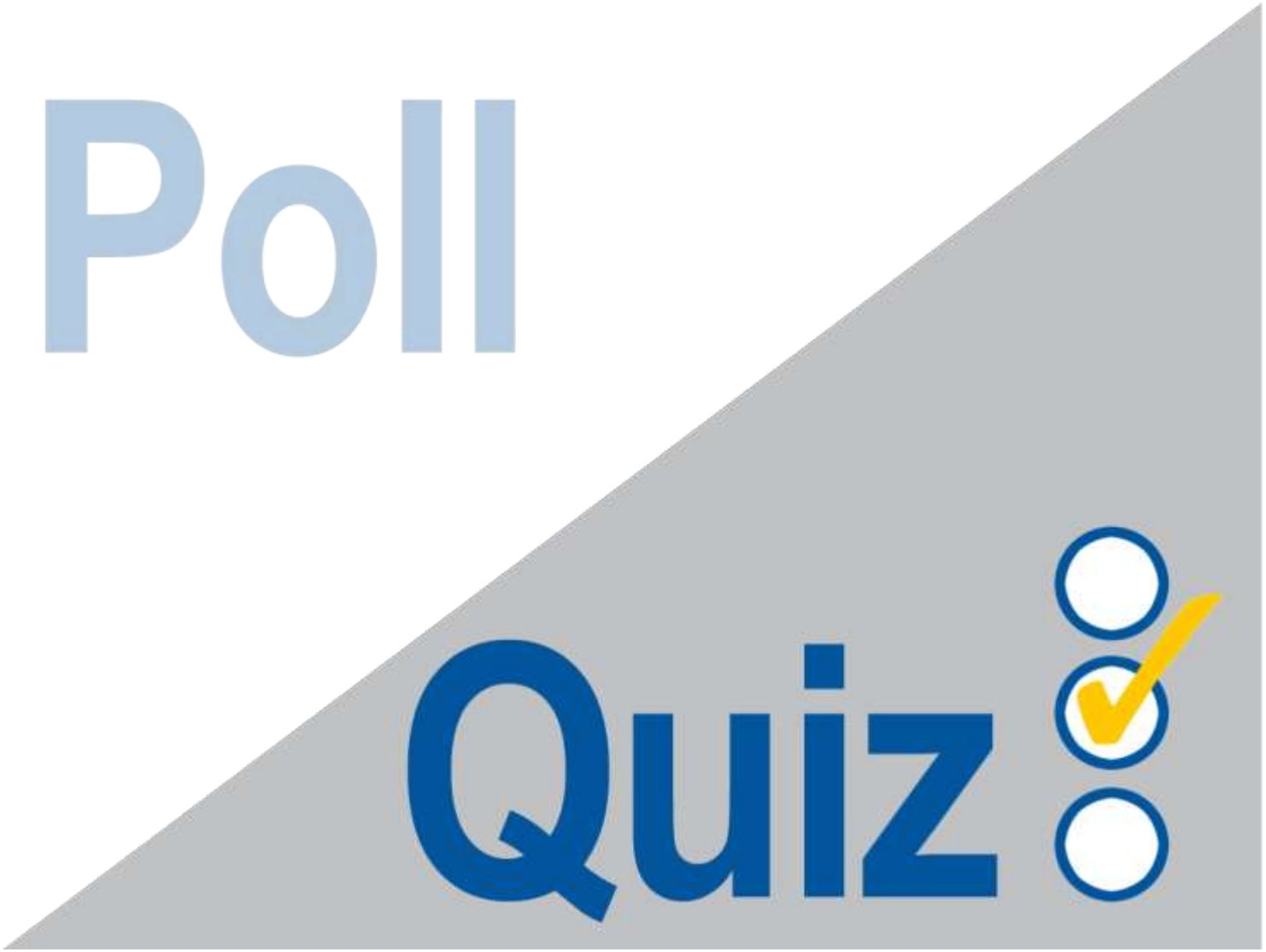
All Nonmatches

```
data EmpsAUC;
   merge EmpsAU(in=Emps)
         PhoneC(in=Cell);
   by EmpID;
   if Emps=0 or Cell=0;
run;
```

EmpsAUC

First	Gender	EmpID	Phone
Kylie	F	121151	
		121153	+61(2)5555-1348

Poll



10.14 Quiz

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
        OR
if A and not B;
```

X	Y	Z	W
2			60

10.14 Quiz – Correct Answer

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
        OR
if A and not B;
```

X	Y	Z	W
2			60

```
if A=0 and B=1;
        OR
if not A and B;
```

10.14 Quiz

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
        OR
if A and not B;
```

X	Y	Z	W
2			60

```
if A=0 and B=1;
        OR
if not A and B;
```

X	Y	Z	W
1	10	20	50
3	30	40	

10.14 Quiz – Correct Answer

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
      OR
if A and not B;
```

X	Y	Z	W
2			60

```
if A=0 and B=1;
      OR
if not A and B;
```

X	Y	Z	W
1	10	20	50
3	30	40	

```
if A=1;
   OR
if A;
```

10.14 Quiz

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
      OR
if A and not B;
```

X	Y	Z	W
2			60

```
if A=0 and B=1;
      OR
if not A and B;
```

X	Y	Z	W
1	10	20	50
3	30	40	

```
if A=1;
   OR
if A;
```

X	Y	Z	W
1	10	20	50
2			60

10.14 Quiz – Correct Answer

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
        OR
if A and not B;
```

X	Y	Z	W
2			60

```
if A=0 and B=1;
        OR
if not A and B;
```

X	Y	Z	W
1	10	20	50
3	30	40	

```
if A=1;
   OR
if A;
```

X	Y	Z	W
1	10	20	50
2			60

```
if B=1;
   OR
if B;
```

10.14 Quiz

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
        OR
if A and not B;
```

X	Y	Z	W
2			60

```
if A=0 and B=1;
        OR
if not A and B;
```

X	Y	Z	W
1	10	20	50
3	30	40	

```
if A=1;
   OR
if A;
```

X	Y	Z	W
1	10	20	50
2			60

```
if B=1;
   OR
if B;
```

X	Y	Z	W
1	10	20	50

10.14 Quiz – Correct Answer

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
   OR
if A and not B;
```

X	Y	Z	W
2			60

```
if A=0 and B=1;
   OR
if not A and B;
```

X	Y	Z	W
1	10	20	50
3	30	40	

```
if A=1;
   OR
if A;
```

X	Y	Z	W
1	10	20	50
2			60

```
if B=1;
   OR
if B;
```

X	Y	Z	W
1	10	20	50

```
if A=1 and B=1;
   OR
if A and B;
```

10.14 Quiz

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
        OR
if A and not B;
```

X	Y	Z	W
2			60

```
if A=0 and B=1;
        OR
if not A and B;
```

X	Y	Z	W
1	10	20	50
3	30	40	

```
if A=1;
   OR
if A;
```

X	Y	Z	W
1	10	20	50
2			60

```
if B=1;
   OR
if B;
```

X	Y	Z	W
1	10	20	50

```
if A=1 and B=1;
        OR
if A and B;
```

X	Y	Z	W
2			60
3	30	40	

10.14 Quiz – Correct Answer

Write an appropriate IF statement to create the desired data sets.

dataA

X	Y	Z
1	10	20
3	30	40

dataB

X	W
1	50
2	60

```
data new;
   merge dataA(in=A)
         dataB(in=B);
   by X;
run;
```

new

X	Y	Z	W
1	10	20	50
2			60
3	30	40	

Desired SAS Data Sets

X	Y	Z	W
3	30	40	

```
if A=1 and B=0;
   OR
if A and not B;
```

X	Y	Z	W
2			60

```
if A=0 and B=1;
   OR
if not A and B;
```

X	Y	Z	W
1	10	20	50
3	30	40	

```
if A=1;
   OR
if A;
```

X	Y	Z	W
1	10	20	50
2			60

```
if B=1;
   OR
if B;
```

X	Y	Z	W
1	10	20	50

```
if A=1 and B=1;
   OR
if A and B;
```

X	Y	Z	W
2			60
3	30	40	

```
if A=0 or B=0;
   OR
if not A or not B;
```

Merging Multiple Data Sets

The DATA statement can merge multiple input data sets as long as they all have a common variable.

payroll06

Obs	Employee_ID	Employee_Gender	Salary	Birth_Date	Employee_Hire_Date	Employee_Term_Date	Marital_Status	Dependents
1	120101	M	163040	6074	01JUL2003	.	S	0
2	120102	M	108255	3510	01JUN1989	.	O	2
3	120103	M	87975	-3996	01JAN1974	.	M	1
4	120104	F	46230	-2061	01JAN1981	.	M	1
5	120105	F	27110	5468	01MAY1999	.	S	0

payroll07

Obs	Employee_ID	Employee_Gender	Salary	Birth_Date	Employee_Hire_Date	Employee_Term_Date	Marital_Status	Dependents
1	120101	M	167931	6074	01JUL2003	.	S	0
2	120102	M	111503	3510	01JUN1989	.	O	2
3	120103	M	90614	-3996	01JAN1974	.	M	1
4	120104	F	47617	-2061	01JAN1981	.	M	1
5	120105	F	27923	5468	01MAY1999	.	S	0

payroll08

Obs	Employee_ID	Employee_Gender	Salary	Birth_Date	Employee_Hire_Date	Employee_Term_Date	Marital_Status	Dependents
1	120101	M	172969	6074	01JUL2003	.		.
2	120102	M	114848	3510	01JUN1989	.		.
3	120103	M	93332	-3996	01JAN1974	.		.
4	120104	F	49046	-2061	01JAN1981	.		.
5	120105	F	28761	5468	01MAY1999	.	S	0

Merging Multiple Data Sets

Note: The SAS log is extremely important in troubleshooting merges.

```
data payroll_hist;
        merge mylib.payroll06 mylib.payroll07
              mylib.payroll08(drop=marital_status);
        by employee_id;
run;
```

Common variables are overwritten from the right. Use UPDATE instead of MERGE to prevent missing data from overwriting existing data.