Data Types

Type		Example
Numeric	Integer	-1, 0, 1, 2, 56, 100001
	“Float”, “Real”	2.335, -6792.2938867546543687
Categorical	Binary	{0, 1}, {yes, no}, {alive, dead}
	Nominal	{cat, dog, mouse}, {red, blue, green, yellow}
	Ordinal	lowest, lower, middle, higher, highest

Select/subset/slice columns

SQL: SELECT X1 FROM TABLE1;

R: Table1$X1

ID	X1	X2
123A	1	A
1234	2	B
1235	3	C
1235B	4	D

Table1

SAS:

```
PROC PRINT DATA=TABLE1;
VAR X1;
RUN;
```

Select/subset/slice columns

SQL: SELECT X1 FROM TABLE1;

R: Table1$X1

ID	X1	X2
123A	1	A
1234	2	B
1235	3	C
1235B	4	D

Table1

X1
1
2
3
4

SAS:

```
PROC PRINT DATA=TABLE1;
VAR X1;
RUN;
```

Select/subset/slice rows

SQL: SELECT X1 FROM TABLE1 WHERE X1 >= 3;

R: Table1[Table1$X1 >= 3,]

ID	X1	X2
123A	1	A
1234	2	B
1235	3	C
1235B	4	D

Table1

SAS:

```
PROC PRINT DATA=TABLE1;
WHERE X1 >= 3;
RUN;
```

Select/subset/slice rows

SQL: SELECT X1 FROM TABLE1 WHERE X1 >= 3;

R: Table1[Table1$X1 >= 3,]

ID	X1	X2
123A	1	A
1234	2	B
1235	3	C
1235B	4	D

Table1

ID	X1	X2
1235	3	C
1235B	4	D

SAS:

```
PROC PRINT DATA=TABLE1;
WHERE X1 >= 3;
RUN;
```

Group by

SQL: SELECT ID, MEAN(X3) AS MEANX3 FROM TABLE2 GROUP BY ID;

R: summarise(group_by(Table1, ID), meanx3 = mean(X3))

ID	X3
123A	5
1234	6
1235	7
1235B	8
1234	9
1235	10
1235B	11

Table2

SAS:

```
PROC MEANS DATA=TABLE2 MEAN;
VAR X3;
BY ID;
RUN;
```

Group by

SQL: SELECT ID, MEAN(X3) AS MEANX3 FROM TABLE2 GROUP BY ID;

R: summarise(group_by(Table1, ID), MEANX3 = mean(X3))

ID	X3
123A	5
1234	6
1235	7
1235B	8
1234	9
1235	10
1235B	11

Table2

ID	MEANX3
123A	5
1234	7.5
1235	8.5
1235B	9.5

SAS:

```
PROC MEANS DATA=TABLE2 MEAN;
VAR X3;
BY ID;
RUN;
```

Left Join

SQL: SELECT * FROM TABLE1 LEFT JOIN TABLE2 ON TABLE1.ID = TABLE2.ID;

R: merge(x = Table2, y = Table1, by = 'ID', all.x = TRUE)

ID	X3
123A	5
1234	6
1235	7
1235B	8
1234	9
1235	10
1235B	11

Table2

+

ID	X1	X2
123A	1	A
1234	2	B
1235	3	C
1235B	4	D

Table1

SAS:

```
DATA TABLE2;
MERGE TABLE2 TABLE1;
BY ID;
RUN;
```

Left Join

SQL: SELECT * FROM TABLE1 LEFT JOIN TABLE2 ON TABLE1.ID = TABLE2.ID;

R: merge(x = Table2, y = Table1, by = 'ID', all.x = TRUE)

ID	X3
123A	5
1234	6
1235	7
1235B	8
1234	9
1235	10
1235B	11

Table2

+

ID	X1	X2
123A	1	A
1234	2	B
1235	3	C
1235B	4	D

Table1

ID	X3	X1	X2
1234	6	2	B
1234	9	2	B
1235	7	3	C
1235	10	3	C
1235B	8	4	D
1235B	11	4	D
123A	5	1	A

SAS:

```
DATA TABLE2;
MERGE TABLE2 TABLE1;
BY ID;
RUN;
```