Intro to SQL

Heide Jackson
heidej@umd.edu

University of Maryland Population Research Center

August 2019

High Level Things to Know

- SQL (sometimes pronounced SEQUEL or SQL) is a relational database language used for accessing and manipulating data.
- This presentation will focus on showing how SQL can be called in a SAS environment.

Why Use SQL with SAS

By default, manipulation of SAS data within columns can be clunky, limited and not intuitive.

SQL offers easy ways of manipulating data within columns.

SQL is also great for subsetting data and complex data merges.

Calling SQL in SAS

- SQL can be used to load in and create data sets just like a data statement

```
data savedata;
set loaddata;
run;
/*Same as*/
proc sql;
create table savedata as /*Make new data set*/
select * /*Keep all variables*/
from loaddata; /*The set statement piece*/
quit;
```

Manipulating Data

An Example Where SQL Clearly Beats Base SAS

```
/*How we could calculate number of hits per team
and number of players per team in base sas*/
proc sort data=baseball;
by team;
run;
data baseball2; set sashelp.baseball;
by team;
  if first.team then do;
    hitspteam=nhits;
    count_identifier=1;
  end;
  else do;
    hitspteam+nhits;
    count_identifier+1;
  end;
  if last.team then output;
run;
```

Manipulating Data

An Example Where SQL Clearly Beats Base SAS

```
proc sql;
create table baseballsql as
select * /*keep all variables*/,
sum(nhits) as hitspteam, /*Sums number of hits*/
count(team) as count_identifier /*Counts number on team*/
from sashelp.baseball
group by team;
/*Indicates that transformations on team level*/
quit;
```

Subsetting Data Using SQL

- SQL allows data to be easily subsetted via the where statement.
- The where statement can handle multiple conditions and read across data sets.

```
proc sql;
create table check as
select *
from sashelp.baseball
where nhits>40;
quit;
```

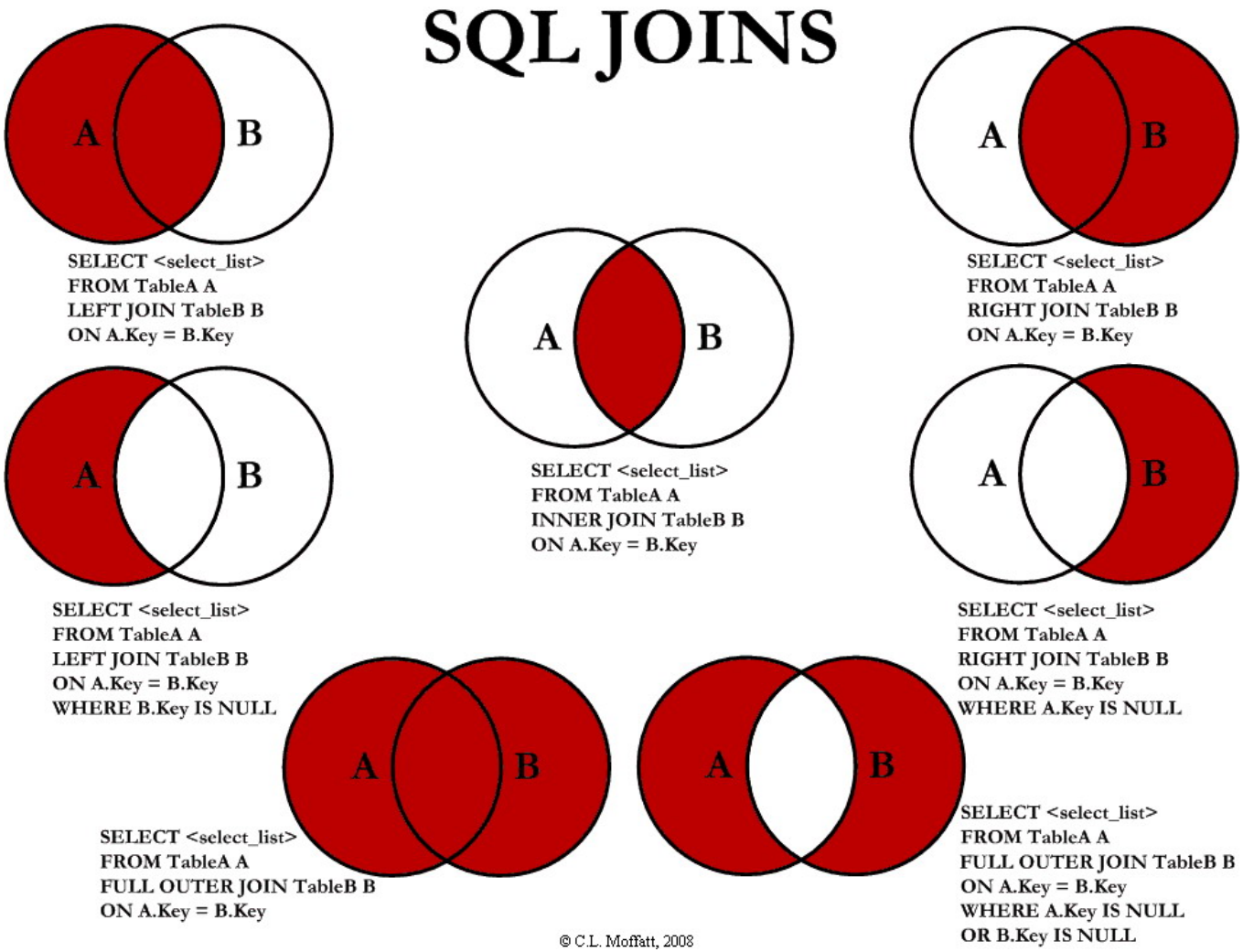
Merging Data Using SQL

SQL can also handle complex merges across multiple data sets.

This can be done in combination with other data manipulation.

```
proc sql;
create table merged as
/*Select can be used to specify subsets
from different data sources */
select a.* b.hitspteam
from sas.baseball as a, baseballsql as b
where a.team=b.team and nhits<100;
quit;
```

Other Types of Merges



source: `https://www.codeproject.com/Articles/33052/Visual-Representation-of-SQL-Joins`