

The US Temperature Record 6: Extracting annual averages using SQL

/ OCT 06, 2021

For our example I'm going to use Microsoft Access and the SQL language to make the create the yearly averages. This might also be very doable using a pivot table in Excel, but I don't know the intricacies of pivot table usage. In fact, pivot tables might be a great deal easier to use than Access/SQL, particularly if you can link the Excel data directly back to the .txt files, making the data refreshes each year very easy to do (well, you can do the same with Access, but I didn't).

In Access, we will now produce the annual averages for all stations in the U.S. using SQL, Structured Query Language. SQL is used to ask a database to produce specific data, and we will use the aggregate functions to get the averages over all stations.

It's actually more complex than that, because the thing we are aggregating is the yearly average,

```
([jan]+[feb]+[mar]+[apr]+[jun]+[jul]+[aug]+[sep]+[oct]+[nov]+[dec])/1200
```

Or as I designate the months columns in the Access database,

```
([M1]+[M2]+[M3]+[M4]+[M5]+[M6]+[M7]+[M8]+[M9]+[M10]+[M11]+[M12])/1200
```

Dividing by 1200 does two things: divides by 12 to get the annual average, and convert the temperature in centi-centigrade back to temperature in degrees centigrade (a "grade" being the difference in temperature between the freezing point and boiling point of water). So at this point we have a yearly average for each station for each year. This can be modified as

```
(([M1]+[M2]+[M3]+[M4]+[M5]+[M6]+[M7]+[M8]+[M9]+[M10]+[M11]+[M12])*9/5/100/12 + 32)
```

or

```
(([M1]+[M2]+[M3]+[M4]+[M5]+[M6]+[M7]+[M8]+[M9]+[M10]+[M11]+[M12])*0.0015 + 32)
```

to produce the temperatures in degrees Fahrenheit.

The aggregate function, `GROUP BY Year`, then averages all the years together to produce the single national average for all the annual averages.

The overall SQL statement is

```
SELECT t.Year, Avg(( [M1]+[M2]+[M3]+[M4]+[M5]+[M6]+[M7]+[M8]+[M9]+[M10]+[M11]+[M12] ) * 0.0015 + 32)
AS TavgRawAnnual
FROM [TAVG-RAW] AS t
GROUP BY t.Year;
```

and for precipitation,

```
SELECT t.Year, Avg(( [M1]+[M2]+[M3]+[M4]+[M5]+[M6]+[M7]+[M8]+[M9]+[M10]+[M11]+[M12] ) / 1200) AS
PRCPRawAnnual
FROM [PRCP-RAW] AS t
GROUP BY t.Year;
```

The name of the calculated column (here "TavgRawAnnual") and the table from which the data comes ("TAVG-RAW") need to be changes when using a different data set. I'm using an alias for the table name, "t," so that I don't need to change too much when I copy the SQL statement. This statement will produce a list of the nation-side average temperatures for each year. I create a query like this for each table I have. The data goes into an Excel sheet for graphing.

If you want the data for a particular month, the August highs each year, for example, you can simplify the query,

```
SELECT t.Year, Avg([M8] * 0.018 + 32) AS TmaxRawAugust
FROM [TMAX-RAW] AS t
GROUP BY t.Year;
```

Revision #1

Created 7 April 2024 19:11:46 by bruce

Updated 7 April 2024 19:12:34 by bruce