

Using MS Access for your research

12 steps (give or take a few) to giving up the excel habit

Steven Citron-Pousty

Steven.Citron-Pousty@yale.edu

Excel

- Good for calculations
- You are very familiar with it
- SOMETHING ELSE

But just because you know how to
use a hammer the whole world is
NOT a nail

Examples of problems

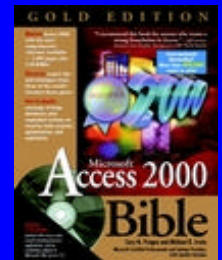
- Select all the people with brown hair, high blood pressure, and younger than 37
- Match all survey information with voting history for 400,000 people
- Show all the purchases on a given day and add on personal information for any people where we have the information
- More than 1 person needs to enter data at the same time
- You need to give certain people access to one view of the data and other people to a different view

Solution to these Problems

- The basic solution is to use an RDBMS (Relational Database Management System)
- Many examples – Oracle, DB2, PostgreSQL, MySQL, FileMaker, and MS Access
- We are going to cover using Access today

Why use Access

- Easy to install and run
- All over campus and the world
- Lots-o-reading with my favorite being the bible
- Good for people just getting started with DBs
- I am familiar with it



Example for today

- Invasive Species Database
- Person went around to different Herbariums (plant museums) and collected information on specimens from invasive species
- Analogous to collecting survey data on people in different schools OR patients at different clinical centers
- There is a data problem in one of the tables but otherwise there is a lot of relational structure
- Its in the c:\temp directory and called Example.mdb

Goals for today

- Teach you some DB terms and ideas
- Play with the existing DB for a bit
- Show you how to start creating your own DB
- Sum up

Golden rule: You must
ask me questions!!

Table

- Table – Like a spreadsheet yet not really a spreadsheet.
 - The data should be as atomic as possible/repeating fields – example Student/teacher
 - Each table can also be thought of as a person, place, or thing
 - Observation - same as row in a spreadsheet
 - No way to do calculations in a spreadsheet fashion, i.e. $=a1+b1$

What does it look like in Access

	species_id	genus	species	authority	common_nam	plant	date_entered	genusspeci	subspecies	plantsCode
+	5	Cabomba	caroliniana	A. Gray	Fanwort	<input checked="" type="checkbox"/>		Cabomcarol		
+	6	Ailanthus	altissima	(Mill.) Swin	Tree-of-heaven	<input checked="" type="checkbox"/>		Ailanaltis		
+	7	Najas	minor	Allioni	Eutrophic Wate	<input checked="" type="checkbox"/>		Najasminor		
+	8	Aegoposium	podagraria	L.	Goutweed	<input checked="" type="checkbox"/>		Aegoppodag		
+	9	Egeria	densa	Planchon	Brazilian water-	<input checked="" type="checkbox"/>		Egeridensa		
+	10	Hydrilla	verticillata	(L. f.) Royle	Hydrilla	<input checked="" type="checkbox"/>		Hydriverti		
+	11	Myriophyllum	heterophyllum	Michx.	Variable water-r	<input checked="" type="checkbox"/>		Myrioheter		
+	12	Myriophyllum	spicatum	L.	Eurasian water-	<input checked="" type="checkbox"/>		Myriospica		
+	13	Potamogeton	crispus	L.	Crispy-leaved P	<input checked="" type="checkbox"/>		Potamcrisp		
+	14	Rorippa	nasturium-aq	(L.) Hayek	Watercress	<input checked="" type="checkbox"/>		Roripnastu		
+	15	Trapa	natans	L.	Waterchestnut	<input checked="" type="checkbox"/>		Trapanatan		
+	16	Nymphoides	peltata	(Gmel.) Kunt	Yellow Floating	<input checked="" type="checkbox"/>		Nymphpelta		
+	19	Butomus	umbellatus	L.	Flowering Rush	<input checked="" type="checkbox"/>		Butomumbel		
+	21	Phragmites	australis	(Cav.) Trin.	Common Reed	<input checked="" type="checkbox"/>		Phragaustr		
+	22	Iris	pseudacorus	L.	Yellow Iris	<input checked="" type="checkbox"/>		Irispseud		
+	23	Vallisneria	javanica	(Houtt.) Decr	Japanese Knotw	<input checked="" type="checkbox"/>		Vallisneria		

Record: 1 of 62

Big Difference – Columns

- Access calls the columns fields we might call them variables
- They have unique names
- They have to be defined as a type (boolean, floating point, string...)
- You have to say if you are going to allow blanks

Lets start Access and take a look around

- Go to C:\temp
- Double click on example.mdb and you should see this:

Areas in a DB view

- Tables
- Queries
- Forms
- Wizards at the top of each area to help you with common tasks

Lets open the species table

Primary Keys

- Primary Key is a variable/attribute that uniquely identifies each row
- Can also be a combination of columns
- You may be tempted to use things like last name+first name – DON'T
- Autoincrements are a good idea

Foreign Key

- When two tables are related you need a way to show that they are related
- Foreign key is a primary key from another table in your table.
- It shows that the two tables are related and how one row in one table related to another table
- Look at specimen table

Make our own table

- Lets make a table for people that view a specimen
- Columns
 - Id column
 - Foreign key from Specimen
 - First name
 - Last name
 - Age
 - Weight
 - Date of birth
 - Novella
 - Female

Relationships between tables

- One to one – one row in one table goes to only one row in another table
- One to Many – one row in one table (parent) goes to multiple rows in another table (child)
 - Parent primary key is a foreign key in the child table
- Many to Many – not allowed in relational databases.
 - Solve by putting an intermediate table which has foreign keys from both the tables you are linking
 - Look at Collectors and specimens

Access Relationship view

- You must manually add in the PK and FK relationships
- Open relationships view



- Then you drag from parent and drop in child fields
- In our case we need to add the new field and then make the relationship
- Usually you do this after making all your tables and they all appear

Queries

- Now we have all this nice data how do we get out what we want
- You create a subset of the data set based upon criteria you specify
- You can do calculation in your queries
- Access makes this easy – cheesy
- You can save the queries and use them as tables
 - Add to a query
 - Edit data directly in the query
 - Update all values in a column according to criteria

Using Design View

- Click on Create query in design view
- Add the tables that have the information you want
- Drag columns to the data area
- * is code for all columns
- Let just add the specimen table and see what happens
 - And vs Or
 - Sorting
 - = vs Like and Wildcards

SQL

- This is the actual language used to query database tables
- I am not going to go into it today.
- Access lets you see and tweak the SQL if you want to.
- Lets peek at the SQL behind the query

Indices

- Helps your queries go faster
- Makes inserts and updates go slower
- If you know you are going to query on a column consistently (last name, SS#) then you might make an index
- Primary key fields are almost always indexed
- Set in your table view
- Add index to species in the species table

Joins

- Need matching columns
- Usually your primary key/foreign key relationship
- You can create joins in the query area
- You can also change the properties on the join

Lets try out a join

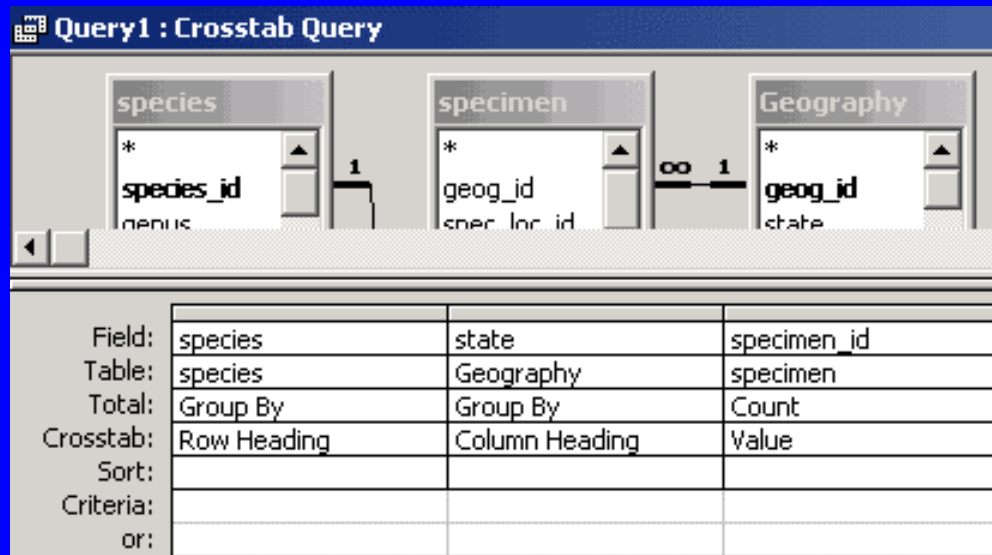
- We are going to use the table specimen and species in a new query
- I will show you some of the tweaks we can do on a join
- What happens if you remove the join?
 - Cartesian product

Calculations in queries

- You can change the names of columns
- You can have a column be a formula
- Example: `gen spp: [genus]+' '+[species]`

Crosstab query

- Lets you do a crosstab within the DB
- Count of the # of specimens per species by state
- New Query in design and add the tables
- Change the query type



Importing Data

- You can create a new table by importing
- You can cut and paste data
 - Works pretty well from excel
 - Queries can help make the data look “similar”
- You can also link to a table
 - Data isn't in the DB, its in the other file
 - Good allows other people to work on that data without needing access to your db
 - Bad need to have that file around when you use those tables
- You can import excel, dbase, lotus, and text files

Export

- Usually you will want to export from a query
- You can actually export the table
- You can also cut and paste
- Export to a lot of different formats
 - All the import formats
 - HTML
 - RTF

Not touched on

- Forms
- Reports
- Multi-user access
- Some useful wizards

Advanced Features

- Programming in Visual Basic
- Password protected databases
- ODBC connections to large databases or other files
- Replication

Resources

- Go this web page

Conclusion

- Access is good for
 - Beginners
 - Small to medium size DBs < 200mbs
 - 1 to 2 concurrent users
 - Windows only teams (for the most part)
 - Front ends to more complicated DBs