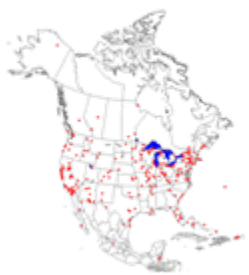




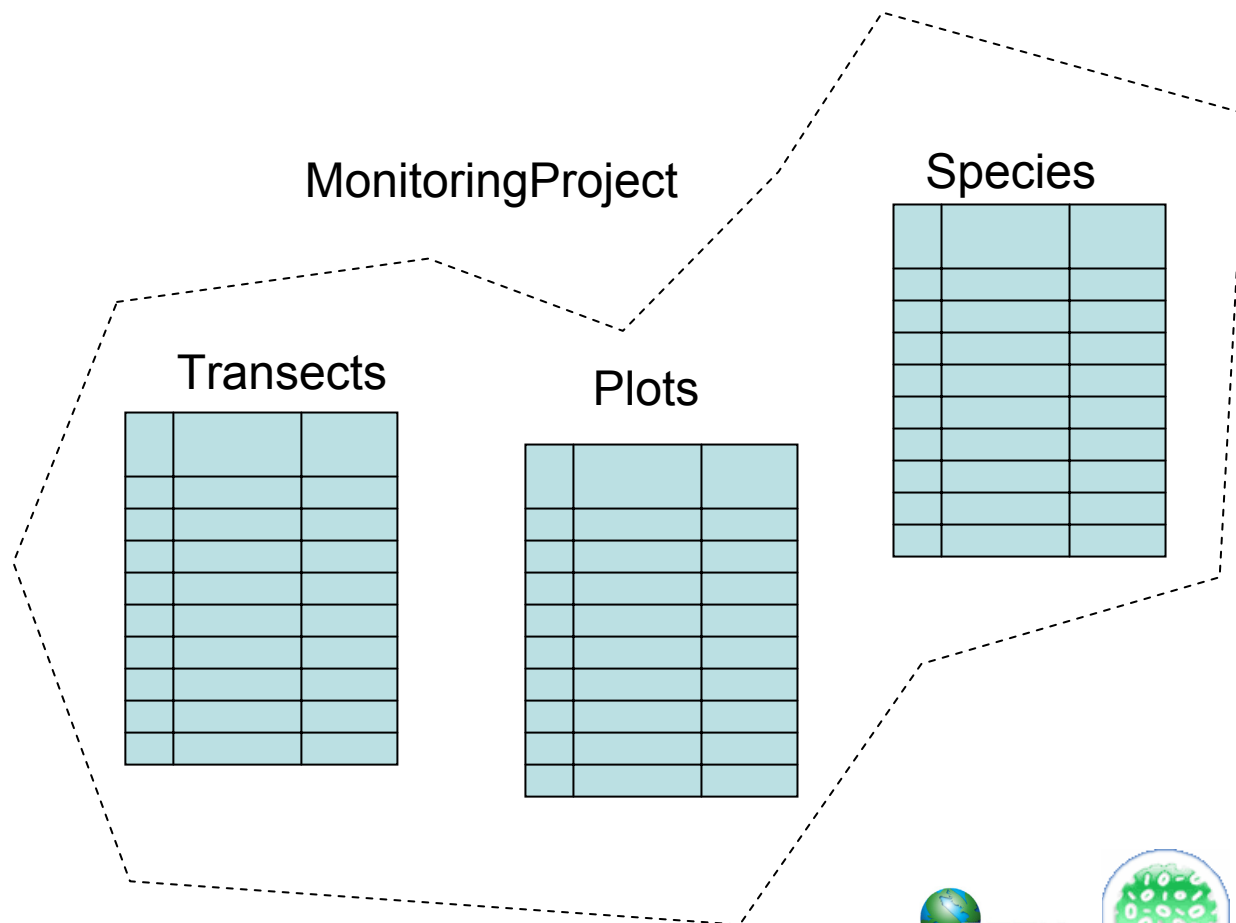
# Database Design

John Kim  
Field Station Programs  
San Diego State University



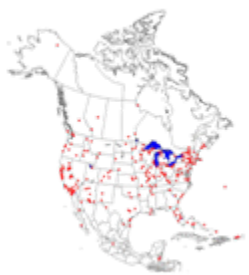


# What is a database?



- A set of tables
- Columns hold specific type of data
- Read/modify information



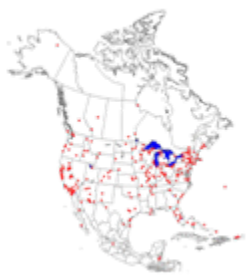


# What is a table?

Plots

Plot_id (integer)	Plot_type (text)	Lat (real no.)	Lon (real no.)	Notes (text)



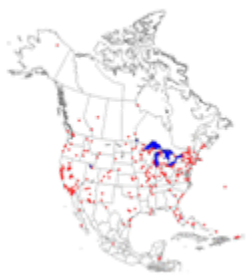


# Difference from Excel

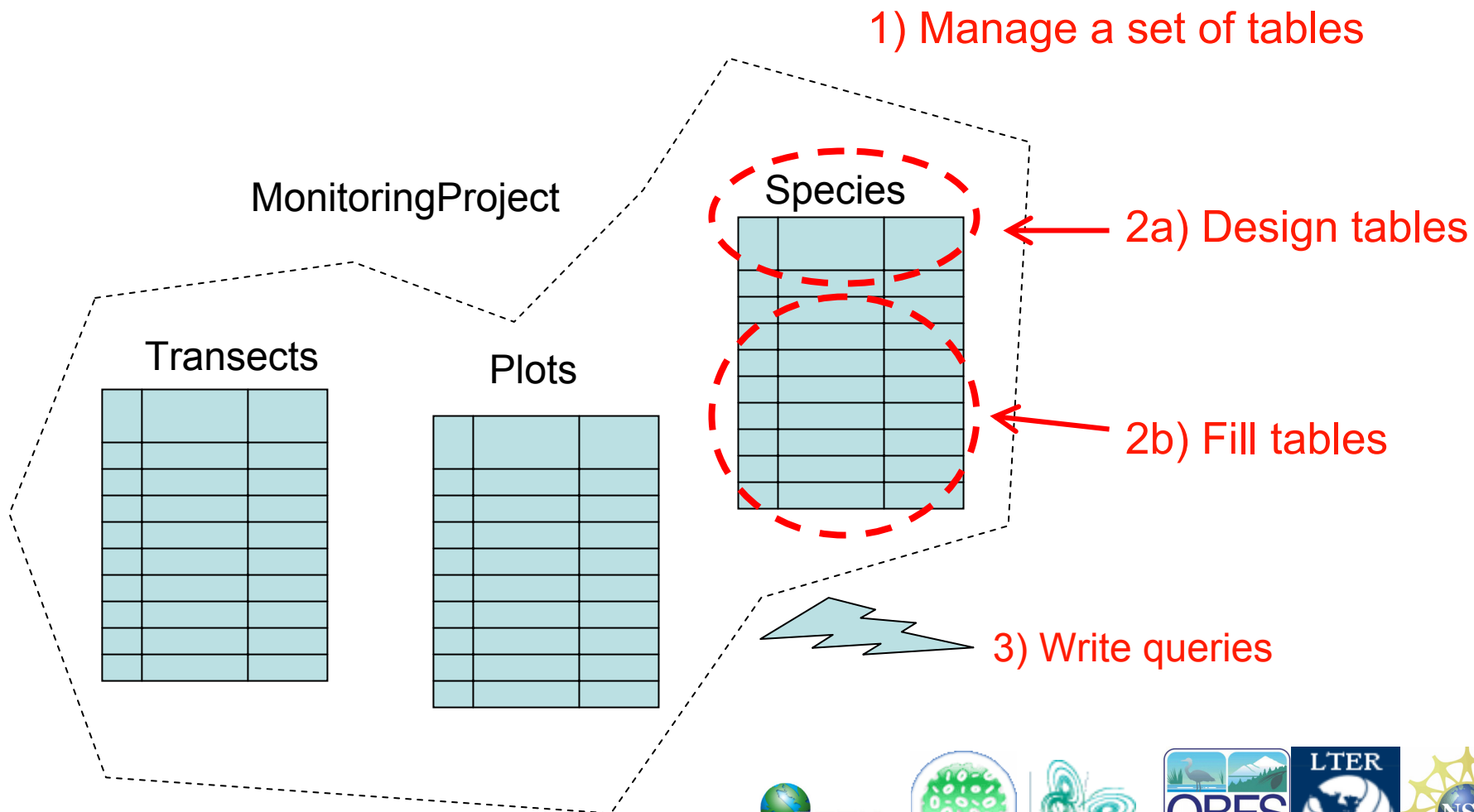
Species			
Species_id (integer)	Name (text)	Common_name (text)	Status (integer)

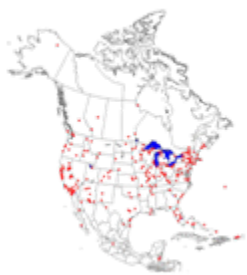
1. Tables are discrete
2. Each table has a name
3. Each column has a name
4. Each column has a type





# Essential Tasks





# What next?

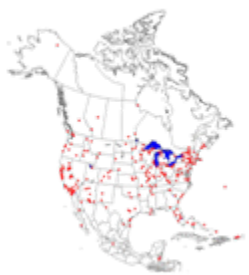
NEXT:

- I present strategy for dividing a table into a set of smaller tables

THEN:

- You design a database in groups
- We discuss your designs



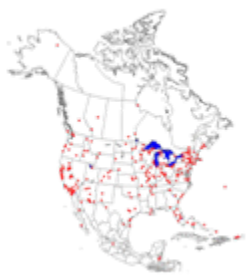


# Splitting up data into tables

3 main tasks:

- 1) Split up the data into tables.
- 2) Designate the type of data each column should hold.
- 3) Identify relations among the tables



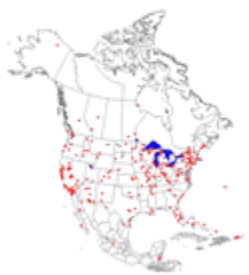


# Design Task 1: Splitting up data into tables

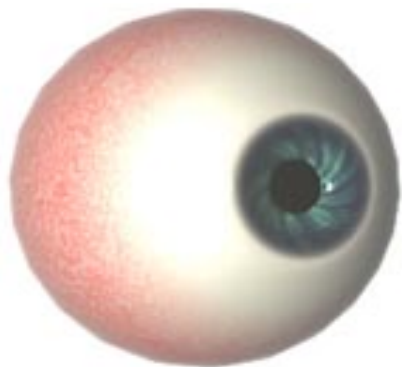
DATE	SITE	WEB	PLOT	QD	SPECIES	OBS	COVER	HT	COUNT	PHEN	COMMENTS
02/03/1999	FPC	1	E	1	ERPU8	1	0.5	4	13	V	NA
02/03/1999	FPC	1	E	1	ERPU8	2	0.1	2	16	V	NA
02/03/1999	FPC	1	E	1	GUSA2	1	0.01	4	2	V	NA
02/03/1999	FPC	1	E	1	GUSA2	2	0.1	5	1	V	NA
02/03/1999	FPC	1	E	1	GUSA2	3	0.5	12	1	V	NA
02/03/1999	FPC	1	E	1	LEFE	1	0.25	5	1	V	NA
02/03/1999	FPC	1	E	2	LATR2	1	7	36	2	V	Possible error
02/03/1999	FPC	1	E	2	LATR2	2	2	32	3	V	NA
02/03/1999	FPC	1	E	2	LATR2	3	8	61	1	V	NA
02/03/1999	FPC	1	E	2	LATR2	4	3	45	1	V	NA
02/03/1999	FPC	1	E	2	LATR2	5	2	24	1	V	NA
02/03/1999	FPC	1	E	2	ERPU8	1	0.25	3	3	V	NA
02/03/1999	FPC	1	E	2	ERPU8	2	0.05	2	11	V	NA

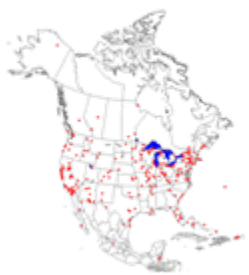






# Methods for splitting up data into tables

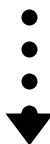




# Method 1 - Eyeballing (book example)

Books

TITLE	AUTHOR 1	AUTHOR 2	PUBLISHER	ISBN	QTY.
Ecology 101	Smith, A.B.	Gordon, D.A.	Univ. Press	4873895759	4324
Ecology for Dummies	Doe, J.		Wiley & Sons	0493802020	8998
Ecology and Politics	Kim, J.B.		McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Kim, J.B.		Univ. Press	2234849302	1



Books

TITLE	?	?
Ecology 101		
Ecology for Dummies		
Ecology and Politics		
Ecology and Modern Cinema		

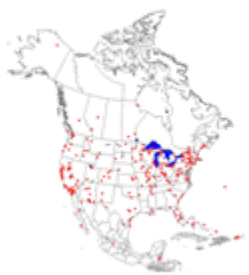
Authors

AUTHOR	?	?
Smith, A.B.		
Doe, J.		
Kim, J.B.		
Gordon, D.A.		

Publishers

PUBLISHERS	?	?
Univ. Press		
Wiley & Sons		
McGraw-Hill		

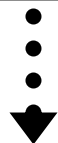




# Method 1 - Eyeballing (personnel example)

## Personnel

Last	First	M.I.	Institution	Sector	Position 1	Position 2
Smith	Ann	A	SDSU	Academic	P.I.	Community Liaison
Smith	Ann	Z	Acme Inc.	Private	Administrator	Field Technician
Kim	John	B	SDSU	Academic	P.I.	Data Manager



## Personnel

Last	First	M.I.	?	?
Smith	Ann	A		
Smith	Ann	Z		
Kim	John	B		

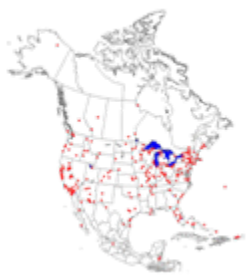
## Institutions

Institution	Sector	?	?
SDSU	Academic		
Acme Inc.	Private		
SDSU	Academic		

## Positions

Position	?	?
P.I.		
Administrator		
Community Liaison		
Field Technician		
Data Manager		





## Method 2 - Normalization



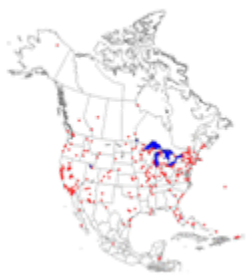
### Abbreviated Normalization Steps:

Step 1: Identify a **primary key** for each table.

Step 2: Eliminate duplicate columns.

Step 3: Eliminate duplicate rows.





Normalization Step 1: Identify a **primary key** for every table.

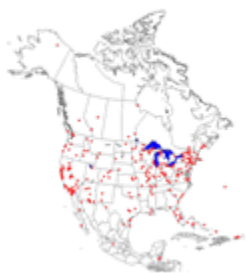
## Books

TITLE	AUTHOR 1	AUTHOR 2	PUBLISHER	ISBN	QTY.
Ecology 101	Smith, A.B.	Gordon, D.A.	Univ. Press	4873895759	4324
Ecology for Dummies	Doe, J.		Wiley & Sons	0493802020	8998
Ecology and Politics	Kim, J.B.		McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Kim, J.B.		Univ. Press	2234849302	1

## Personnel

id	Last	First	M.I.	Institution	Sector	Position 1	Position 2
0	Smith	Ann	A	SDSU	Academic	P.I.	Community Liaison
1	Smith	Ann	Z	Acme Inc.	Private	Administrator	Field Technician
2	Kim	John	B	SDSU	Academic	P.I.	Data Manager

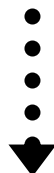




## Normalization Step 2: Eliminate duplicate columns (book example)

Books

TITLE	AUTHOR 1	AUTHOR 2	PUBLISHER	ISBN	QTY.
Ecology 101	Smith, A.B.	Gordon, D.A.	Univ. Press	4873895759	4324
Ecology for Dummies	Doe, J.		Wiley & Sons	0493802020	8998
Ecology and Politics	Kim, J.B.		McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Kim, J.B.		Univ. Press	2234849302	1



uh-oh!

Foreign Key

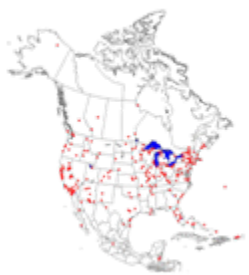
Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

Authors

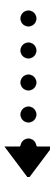
Id	ISBN	Author
0	4873895759	Smith, A.B.
1	4873895759	Gordon, D.A.
2	0493802020	Doe, J.
3	7482929292	Kim, J.B.
4	2234849302	Kim, J.B.





## Normalization Step 2: Eliminate duplicate columns (personnel example)

id	Last	First	M.I.	Institution	Sector	Position 1	Position 2
0	Smith	Jane	A	SDSU	Academic	P.I.	Community Liaison
1	Smith	Jane	Z	Acme Inc.	Private	Administrator	Field Technician
2	Kim	John	B	SDSU	Academic	P.I.	Data Manager



personnel

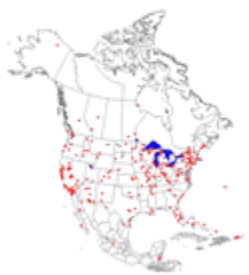
id	Last	First	M.I.	Institution	Sector
0	Smith	Ann	A	SDSU	Academic
1	Smith	Ann	Z	Acme Inc.	Private
2	Kim	John	B	SDSU	Academic

positions Foreign Key

id	person_id	Position
0	0	P.I.
1	0	Community Liaison
2	1	Administrator
3	1	Field Technician
4	2	P.I.
5	2	Data Manager







# Normalization Step 3: Eliminate duplicate rows (book example)

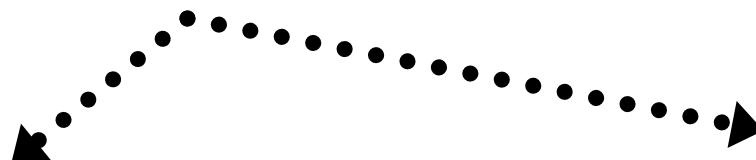
Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

Authors

Id	ISBN	Author
0	4873895759	Smith, A.B.
1	4873895759	Gordon, D.A.
2	0493802020	Doe, J.
3	7482929292	Kim, J.B.
4	2234849302	Kim, J.B.

Books



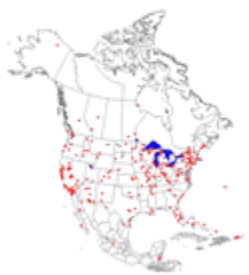
TITLE	PUBLISHER_id	ISBN	QTY.
Ecology 101	0	4873895759	4324
Ecology for Dummies	1	0493802020	8998
Ecology and Politics	2	7482929292	900
Ecology and Modern Cinema	0	2234849302	1

Publishers

Publisher_id	PUBLISHER
0	Univ. Press
1	Wiley & Sons
2	McGraw-Hill







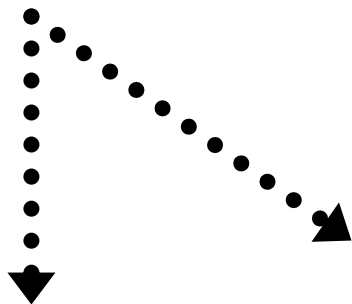
## Normalization Step 3: Eliminate duplicate rows (personnel example)

personnel

id	Last	First	M.I.	Institution	Sector
0	Smith	Ann	A	SDSU	Academic
1	Smith	Ann	Z	Acme Inc.	Private
2	Kim	John	B	SDSU	Academic

positions

id	person_id	Position
0	0	P.I.
1	0	Community Liaison
2	1	Administrator
3	1	Field Technician
4	2	P.I.
5	2	Data Manager



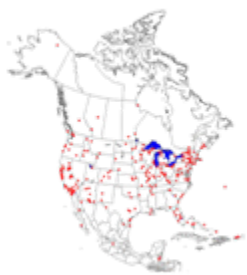
personnel

id	Last	First	M.I.	Institution_id
0	Smith	Ann	A	0
1	Smith	Ann	Z	1
2	Kim	John	B	0

institutions

Institution_id	Institution	Sector
0	SDSU	Academic
1	Acme Inc.	Private





## Method 3 - the Zen way.

The truth is attained by directly knowing.



Attachment to worldly logic leads only to suffering.



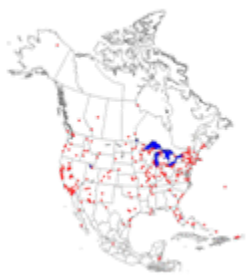


Task 2: Designate the type of data that each column should hold.

Available data types –

Numeric types:		Examples
integer	integers	2005
float	single-precision real numbers (up to 23 places)	3.1415926
double	double-precision real numbers (up to 53 places)	3.141592653 58979323846 26433832795 02884197169 39937510



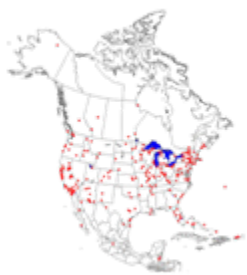


## Task 2: Designate the type of data that each column should hold.

Available data types (continued) –

Text types:		Examples
char char(n)	Characters	“pantroglodytes”
varchar(n)	variable length characters ( $< 255$ chars)	“could not determine species”
text	a large chunk of text	
enum	enumerated values	‘north’, ‘south’, ‘east’, ‘west’
set	a set of values	‘heard’, ‘seen’, ‘captured’





## Task 2: Designate the type of data that each column should hold.

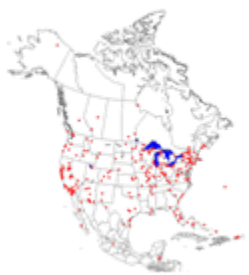
Available data types (continued) –

Date & Time:		
DATETIME	'YYYY-MM-DD HH:MM:SS'	'2005-01-05 15:07:43'
DATE	'YYYY-MM-DD'	'0000-00-00'
TIMESTAMP	Seconds passed since $t_0$	148939284758498
TIME	'HH:MM:SS'	'15:07:43'
YEAR	YYYY	2005

More on column types:

<http://dev.mysql.com/doc> -> Section 11 "Column Types"





Task 2: Designate the type of data that each column should hold.

## Books

<b>TITLE</b>	varchar(255)
<b>PUBLISHER_id</b>	integer
<b>ISBN</b>	varchar(10)
<b>QTY.</b>	integer

## Authors

<b>Id</b>	integer
<b>ISBN</b>	varchar(10)
<b>Author</b>	varchar(255)

## Publishers

<b>PUBLISHER_id</b>	integer
<b>PUBLISHER</b>	varchar(255)

## Personnel

<b>id</b>	integer
<b>Last</b>	varchar(40)
<b>First</b>	varchar(40)
<b>M.I.</b>	char(1)
<b>Institution_id</b>	integer

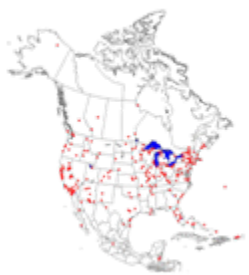
## institutions

<b>Institution_id</b>	integer
<b>Institution</b>	varchar(255)
<b>Sector</b>	enum('academic', 'industry', 'govt', 'NGO')

## Positions

<b>id</b>	integer
<b>person_id</b>	integer
<b>position</b>	varchar





## Task 3: Identify the **relations** among tables.

Books

TITLE	PUBLISHER_id	ISBN	QTY.
Ecology 101	0	4873895759	4324
Ecology for Dummies	1	0493802020	8998
Ecology and Politics	2	7482929292	900
Ecology and Modern Cinema	0	2234849302	1

### Relations Types:

**1 to 1 (1:1)**

**1 to many (1:n)**

**many to many (m:n)**

Authors

Id	ISBN	Author
0	4873895759	Smith, A.B.
1	4873895759	Gordon, D.A.
2	0493802020	Doe, J.
3	7482929292	Kim, J.B.
4	2234849302	Kim, J.B.

Publishers

PUBLISHER_id	PUBLISHER
0	Harcourt Brace
1	Wiley & Sons
2	McGraw-Hill

