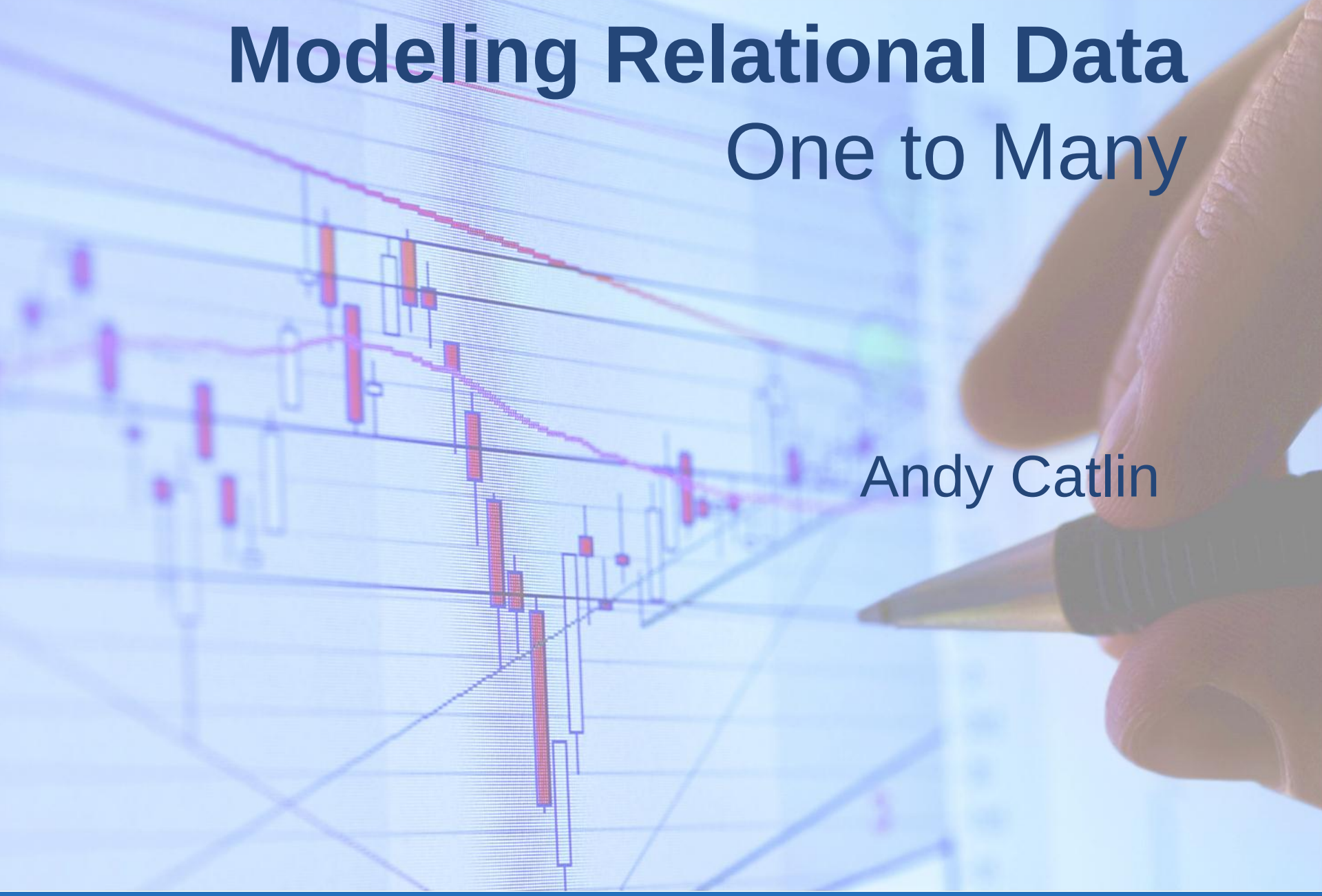
Modeling Relational Data

One to Many

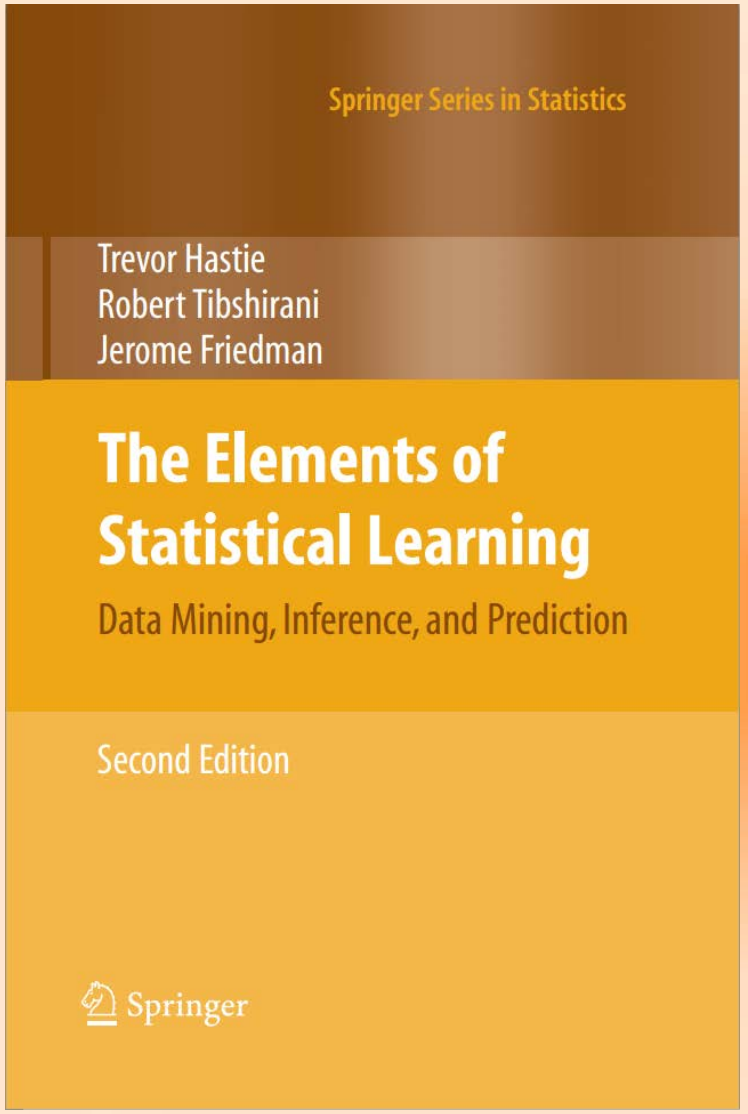


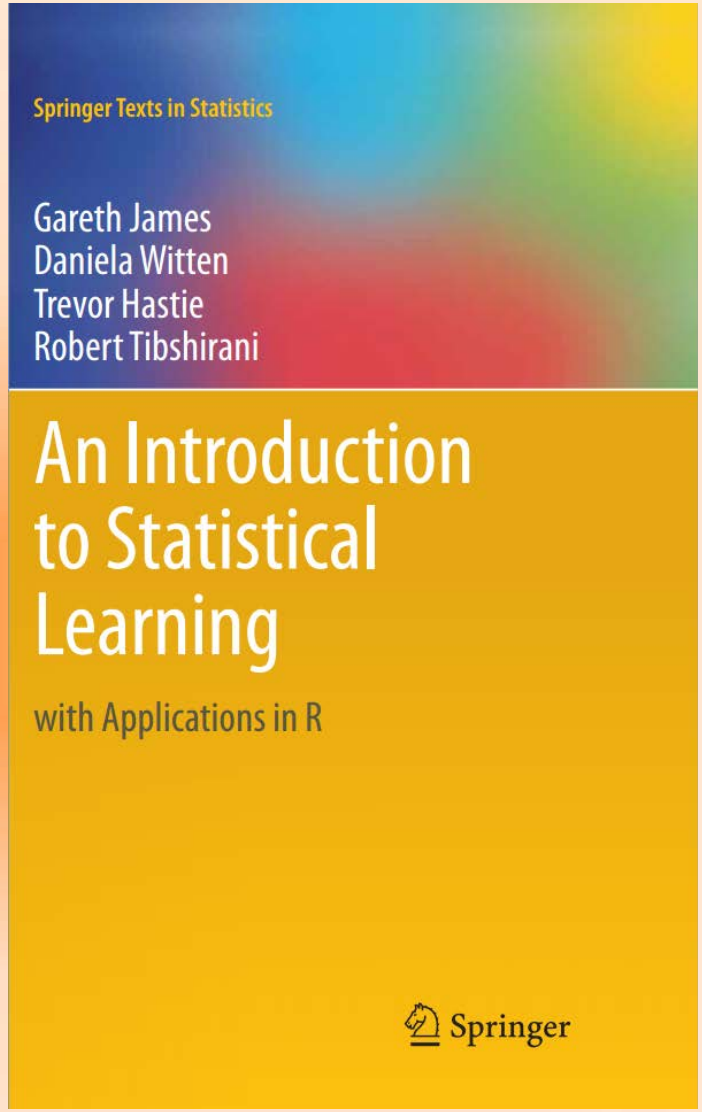
Andy Catlin

Modeling One-to-Many Relationships

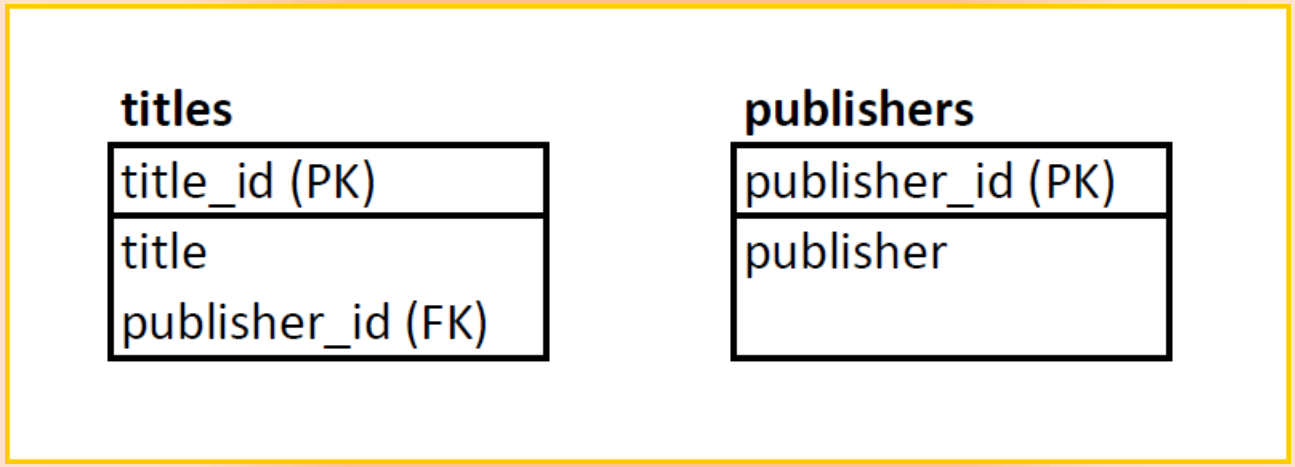
Different Types of JOINS

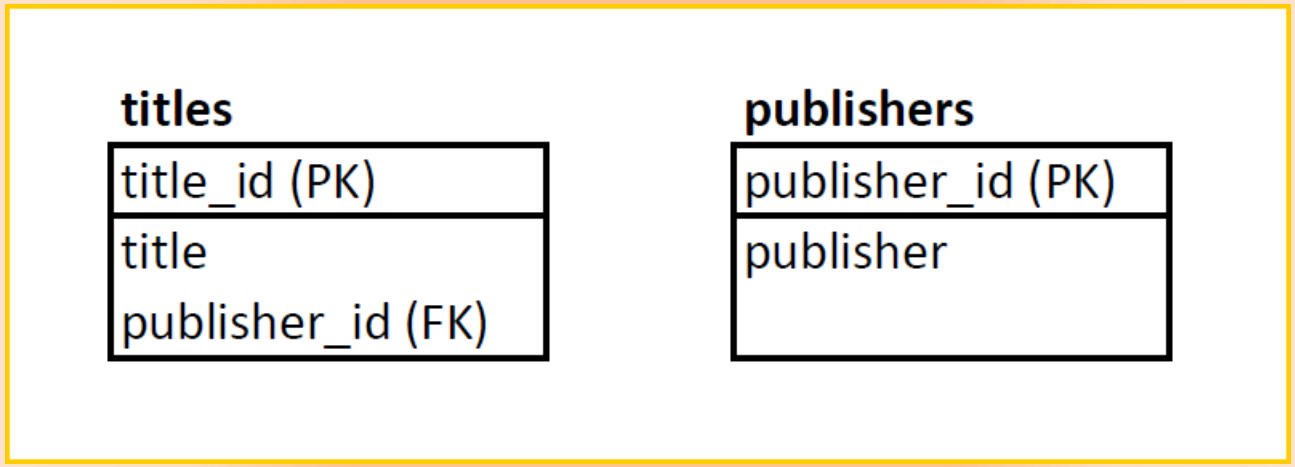
- Inner
- Left
- Right
- Full
- Cross


Springer Series in Statistics
Trevor Hastie
Robert Tibshirani
Jerome Friedman
The Elements of Statistical Learning
Data Mining, Inference, and Prediction
Second Edition
Springer


Springer Texts in Statistics
Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani
An Introduction to Statistical Learning
with Applications in R
Springer

Elements of Statistical Learning	Springer
Introduction to Statistical Learning	Springer


titles
title_id (PK)
title
publisher_id (FK)
publishers
publisher_id (PK)
publisher


titles
title_id (PK)
title
publisher_id (FK)
publishers
publisher_id (PK)
publisher

```
CREATE TABLE publishers
(
  publisher_id int PRIMARY KEY,
  publisher varchar(30) NOT NULL
);

CREATE TABLE titles
(
  title_id int PRIMARY KEY,
  title varchar(100) NOT NULL UNIQUE,
  publisher_id int NULL REFERENCES
          publishers(publisher_id)
);
```

```
INSERT INTO publishers (publisher_id, publisher)
       VALUES (1, 'Springer');
INSERT INTO publishers (publisher_id, publisher)
       VALUES (2, 'O''Reilly');

INSERT INTO titles (title_id, title, publisher_id)
       VALUES ( 1, 'Elements of Statistical Learning', 1);
INSERT INTO titles (title_id, title, publisher_id)
       VALUES ( 2, 'Introduction to Statistical Learning', 1);
INSERT INTO titles (title_id, title, publisher_id)
       VALUES ( 3, 'Easy Peasy Database Modeling', NULL);
```

```
INSERT INTO publishers (publisher_id, publisher)
       VALUES (1, 'Springer');
INSERT INTO publishers (publisher_id, publisher)
       VALUES (2, 'O''Reilly');

INSERT INTO titles (title_id, title, publisher_id)
       VALUES ( 1, 'Elements of Statistical Learning', 1);
INSERT INTO titles (title_id, title, publisher_id)
       VALUES ( 2, 'Introduction to Statistical Learning', 1);
INSERT INTO titles (title_id, title, publisher_id)
       VALUES ( 3, 'Easy Peasy Database Modeling', NULL);
```

```
INSERT INTO publishers (publisher_id, publisher)
       VALUES (1, 'Springer');
INSERT INTO publishers (publisher_id, publisher)
       VALUES (2, 'O''Reilly');

INSERT INTO titles (title_id, title, publisher_id)
       VALUES ( 1, 'Elements of Statistical Learning', 1);
INSERT INTO titles (title_id, title, publisher_id)
       VALUES ( 2, 'Introduction to Statistical Learning', 1);
INSERT INTO titles (title_id, title, publisher_id)
       VALUES ( 3, 'Easy Peasy Database Modeling', NULL);
```

```
SELECT t.title, p.publisher
  FROM titles t
       INNER JOIN publishers p
       ON t.publisher_id = p.publisher_id;
```

Elements of Statistical Learning	Springer
Introduction to Statistical Learning	Springer

```
SELECT t.title, p.publisher
  FROM titles t
       LEFT JOIN publishers p
       ON t.publisher_id = p.publisher_id;
```

Elements of Statistical Learning	Springer
Introduction to Statistical Learning	Springer
Easy Peasy Database Modeling	

```
SELECT t.title, p.publisher
  FROM titles t
       RIGHT JOIN publishers p
       ON t.publisher_id = p.publisher_id;
```

title character varying(100)	publisher character varying(30)
Elements of Statistical Learning	Springer
Introduction to Statistical Learning	Springer
	O'Reilly

```
SELECT t.title, p.publisher
  FROM titles t
       FULL JOIN publishers p
       ON t.publisher_id = p.publisher_id;
```

title character varying(100)	publisher character varying(30)
Elements of Statistical Learning	Springer
Introduction to Statistical Learning	Springer
Easy Peasy Database Modeling	
	O'Reilly

```
SELECT t.title, p.publisher
  FROM titles t
       CROSS JOIN publishers p;
```

title character varying(100)	publisher character varying(30)
Elements of Statistical Learning	Springer
Introduction to Statistical Learning	Springer
Easy Peasy Database Modeling	Springer
Elements of Statistical Learning	O'Reilly
Introduction to Statistical Learning	O'Reilly
Easy Peasy Database Modeling	O'Reilly

Once you understand what data each type of JOIN returns, the challenging part is working with your “business user” to get clear on what data should be included in the report.

Once you understand what data each type of JOIN returns, the challenging part is working with your “business user” to get clear on what data should be included in the report.

You may be combining information from many tables. Each of the JOIN relationship needs to be analyzed, so that the data that is actually returned is the data that is expected to be returned.

Once you understand what data each type of JOIN returns, the challenging part is working with your “business user” to get clear on what data should be included in the report.

You may be combining information from many tables. Each of the JOIN relationship needs to be analyzed, so that the data that is actually returned is the data that is expected to be returned.

This is especially important, since the source data for a business report may have tables with millions of rows. Here, incorrect or undesired results may not be immediately obvious.