

リレーショナルデータベースの 基本概念と応用

URL: <https://www.kkaneko.jp/de/db/index.html>

金子邦彦



- データ共有の問題点

- データはデータファイルにエンコードされる
- 他のユーザーはデータファイルを理解できますか？

- リレーショナル・データベース

- リレーショナル・データベースは、以下の標準規格である。
 - データ形式
 - データ操作 (問い合わせと更新)
 - データ形式の記述 方法
 - 制約の記述 方法

データ形式を記述する



リレーショナルデータベース

リレーショナルデータベース
はテーブルの集合

	id	product_name	type	cost	created_at
1	1	Fukuoka apple	apple	50	2013-11-28 22:53:50
2	2	Kumamoto orange L	orange	30	2013-11-28 22:53:50
3	3	Kumamoto orange M	orange	20	2013-11-28 22:53:50
4	4	Fukuoka melon	melon	{null}	2013-11-28 22:53:50

	name	score	student_name	created_at	updated_at
1	Database	80	KK	2013-11-28 22:47:33	{null}
2	Database	95	AA	2013-11-28 22:47:33	{null}
3	Database	80	LL	2013-11-28 22:47:33	{null}
4	Programming	85	KK	2013-11-28 22:47:33	{null}
5	Programming	75	LL	2013-11-28 22:47:33	{null}

product(id, product_name, type, cost, created_at)

スキーマ

score(name, score, student_name, created_at, updated_at)

スキーマ

スキーマの形式：テーブル名(属性名1, 属性名2, ...)

SQL によるテーブル定義



スキーマ: score(name, score, student_name, created_at, updated_at)

```
CREATE TABLE products (  
  id          INTEGER PRIMARY KEY AUTOINCREMENT NOT NULL,  
  product_name TEXT    UNIQUE NOT NULL,  
  type       TEXT     NOT NULL,  
  cost       REAL,  
  created_at DATETIME NOT NULL );
```

SQL によるテーブル定義

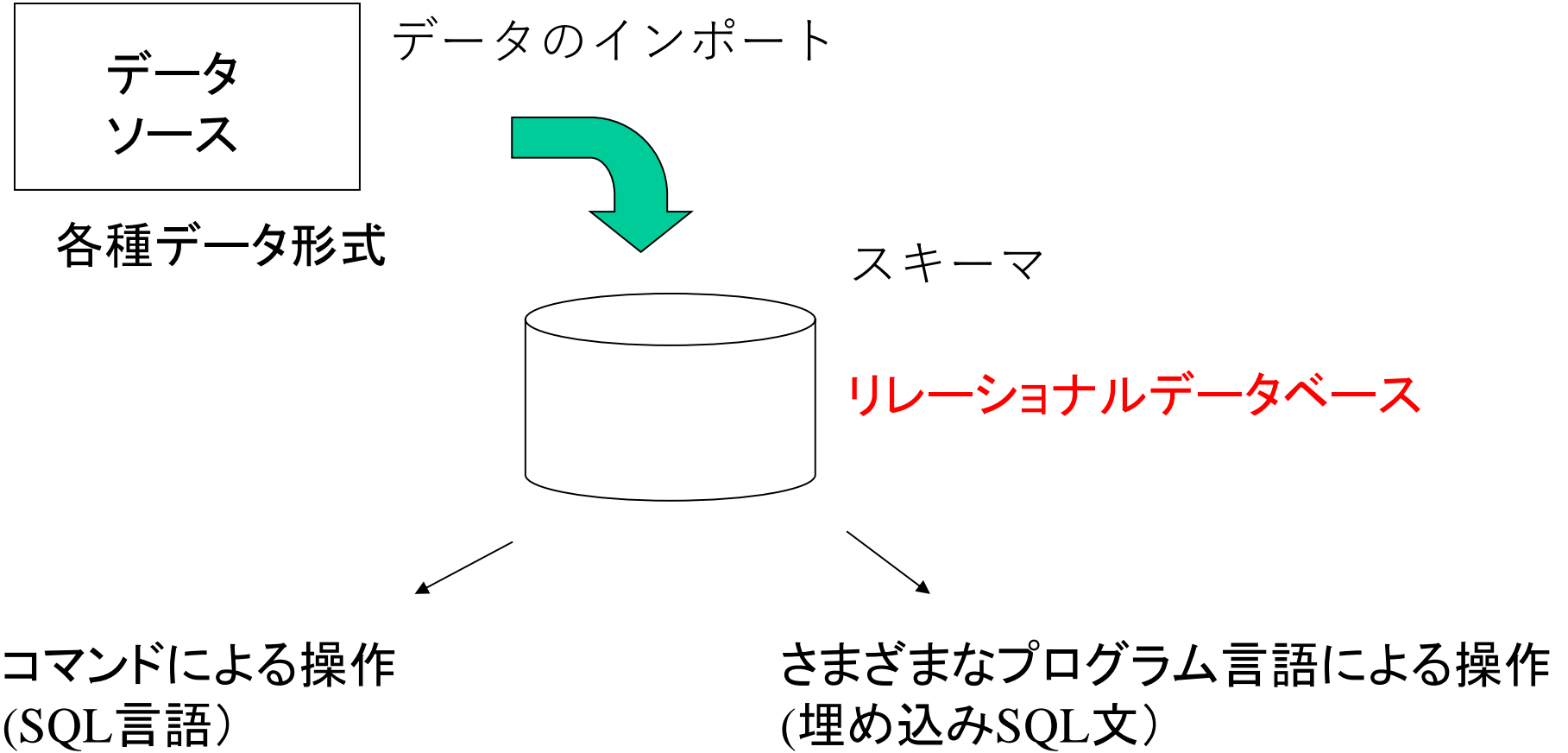
テーブル定義で使用される主なキーワード
INTEGER, REAL, TEXT, DATETIME,
NOT NULL, UNIQUE, PRIMARY KEY

テーブルと属性と行（レコード）



	id	product_name	type	cost	created_at
1	1	Fukuoka apple	apple	50	2013-11-28 22:53:50
2	2	Kumamoto orange L	orange	30	2013-11-28 22:53:50
3	3	Kumamoto orange M	orange	20	2013-11-28 22:53:50
4	4	Fukuoka melon	melon	{null}	2013-11-28 22:53:50

テーブルの本体は行の集合





```
cat >/tmp/a.$$sql <<-SQL
モード csv
.import /tmp/a.$$.csv quoted
SQL
#
tail -n +2 /tmp/Book1.csv >
/tmp/a.$$csv
cat /tmp/a.$$sql | sqlite3
/tmp/quotedb
```

データインポートのプログラム

スキーマのプログラム

```
cat >/tmp/a.$$sql <<-SQL
create table seq (
seq INTEGER PRIMARY KEY NOT NULL,
at datetime)
USDリアル、
GBPは本物だ、
EURリアル、
CADは本物だ、
CHFリアル、
SEKリアル、
DKKリアル、
NOKリアル、
豪ドルは本物だ、
NZDリアル、
ZARリアル、
BHDリアル、
IDR100リアル、
CNYリアル、
HKDリアル、
INRリアル、
MYRリアル、
PHPは本物だ、
SGDリアル、
KRW100リアル、
THBリアル、
KWDリアル、
SARは本物だ、
AEDリアル、
MXNリアル、
PGKは本物だ、
HUFリアル、
CZKリアル、
PLNリアル、
RUBリアル、
本物を試す、
301本物、
IDR1000リアル、
CNYbは本物だ、
MYRb本物、
KRW100b、
TWDリアル、
)
SQL
cat /tmp/a.$$sql | sqlite3
/tmp/quotedb
```

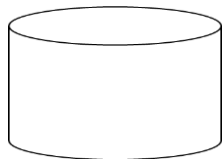
日付

	USD	GBP	EUR	CAD	CHF	SEK	DKK	NOK
2002/4/1	133.15	189.79	116.12	83.48	79.28	12.87	15.63	15.08
2002/4/2	133.2	191.78	117.18	83.38	80.15	13	15.77	15.24
2002/4/3	133.2	191.26	116.96	83.65	80.02	12.95	15.75	15.23
2002/4/4	133.1	191.13	117.15	83.72	80.18	12.93	15.77	15.33
2002/4/5	132.3	189.74	116.32	82.96	79.47	12.87	15.65	15.22

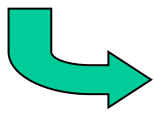
為替データ

データソース

seq	at	USD	GBP	EUR	CAD	CHF	SEK	DKK
1	2002/4/1	133.15	189.79	116.12	83.48	79.28	12.87	15.08
2	2002/4/2	133.2	191.78	117.18	83.38	80.15	13	15.24
3	2002/4/3	133.2	191.26	116.96	83.65	80.02	12.95	15.23
4	2002/4/4	133.1	191.13	117.15	83.72	80.18	12.93	15.33
5	2002/4/5	132.3	189.74	116.32	82.96	79.47	12.87	15.22
6	2002/4/8	131.55	188.54	115.69	82.8	79	12.82	15
7	2002/4/9	131.55	188.31	115.05	82.45	78.46	12.69	15
8	2002/4/10	130.7	187.84	115.16	81.93	78.47	12.63	15
9	2002/4/11	131.1	188.36	115.41	82.33	78.62	12.69	15
10	2002/4/12	131.55	189.06	115.99	82.83	79.2	12.76	15
11	2002/4/15	132.2	189.84	116.38	83.27	79.38	12.83	15
12	2002/4/16	131.85	189.49	116.01	83.15	79.05	12.74	15
13	2002/4/17	130.95	188.55	115.59	82.81	78.74	12.63	15
14	2002/4/18	130.6	188.97	116.31	82.93	79.19	12.73	15
15	2002/4/19	130.1	188.59	115.93	82.65	78.98	12.62	15
16	2002/4/22	130.3	188.62	115.99	82.83	79.13	12.66	15
17	2002/4/23	129.9	188.43	115.4	82.65	78.6	12.55	15
18	2002/4/24	130.25	188.58	115.68	82.79	78.87	12.6	15
19	2002/4/25	129.45	187.73	115.57	82.6	78.83	12.56	15
20	2002/4/26	128.75	187.43	115.62	82.25	79.04	12.54	15

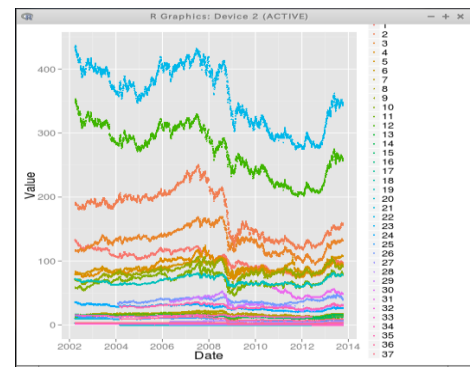


リレーショナルデータベース

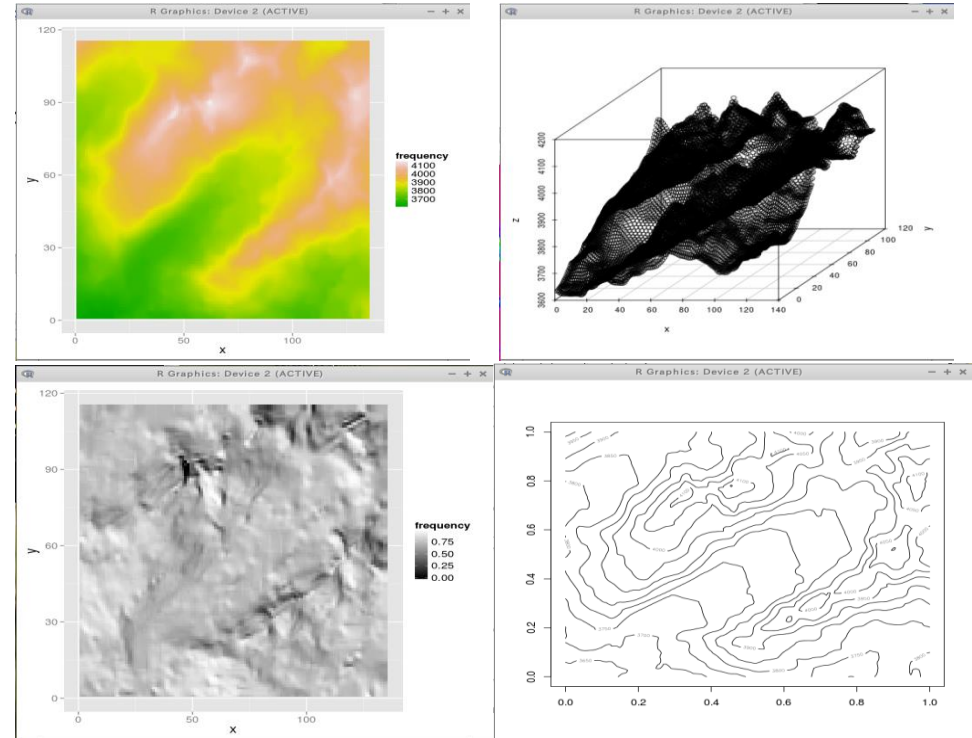
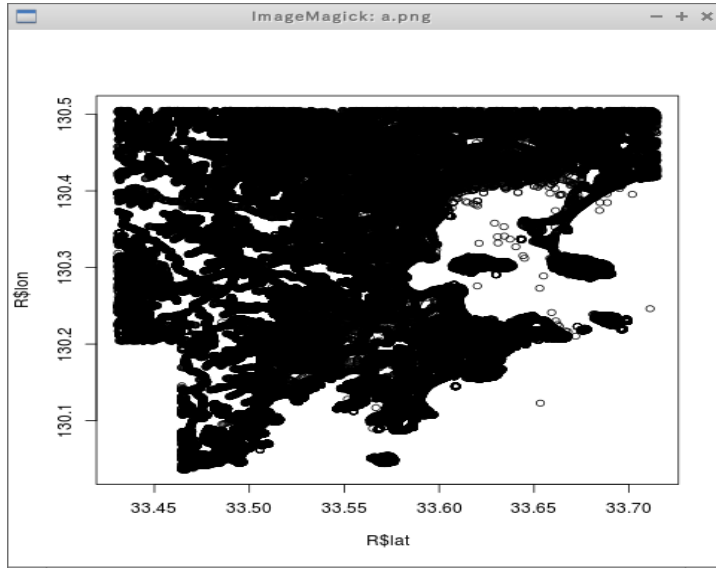


プロットプログラム

```
M <- table_to_melt(T, T$at, "%Y/%m/%d")
#
ggplot(M, aes(x=Date, y=Value, color=factor(AttrNum)))+
geom_point(size=1) ;
```



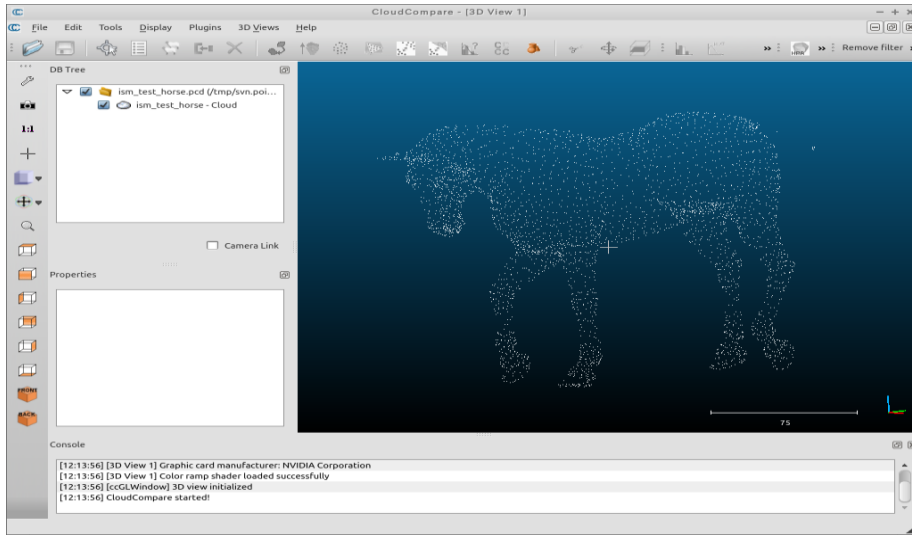
プロットの結果



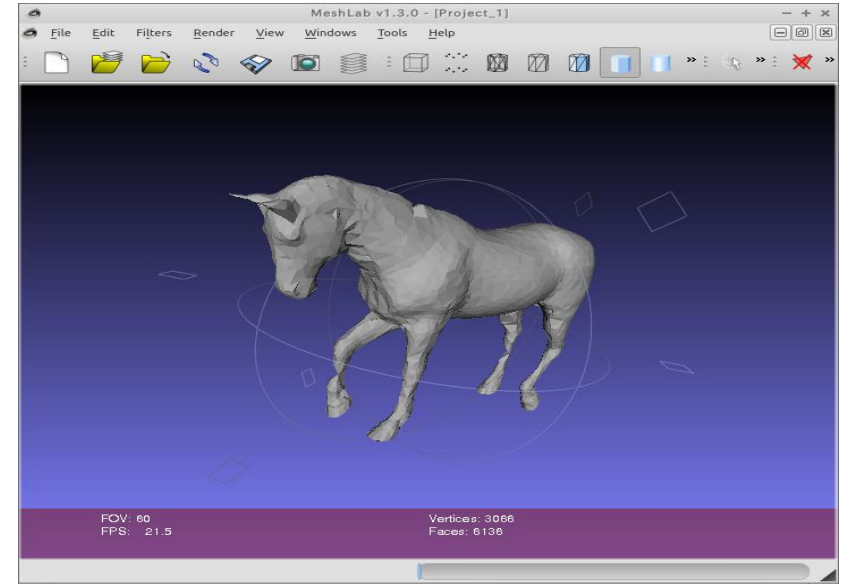
福岡市地図データ

A デジタル標高地図データ

リレーショナルデータベースを使用したプロット例



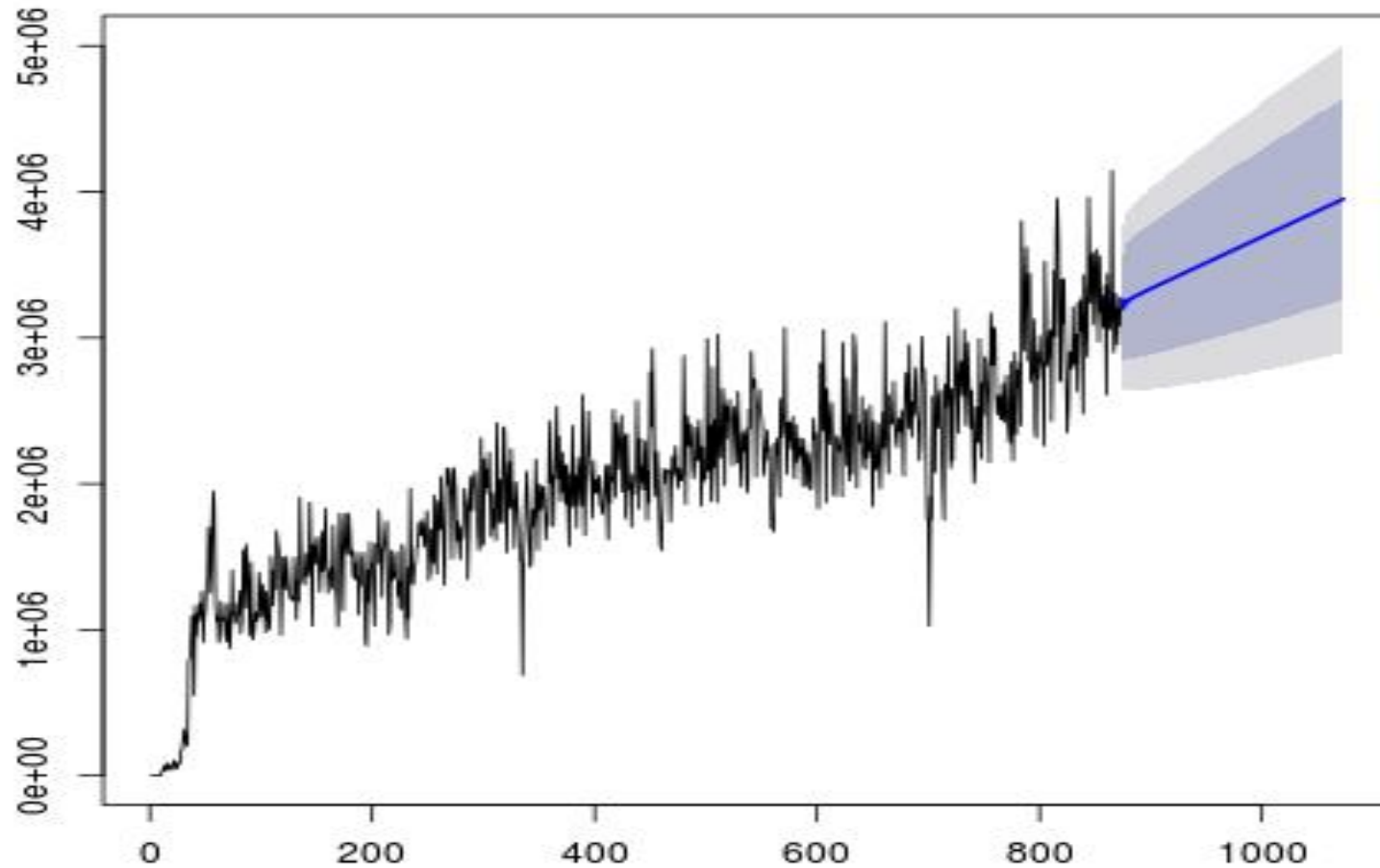
点群データ



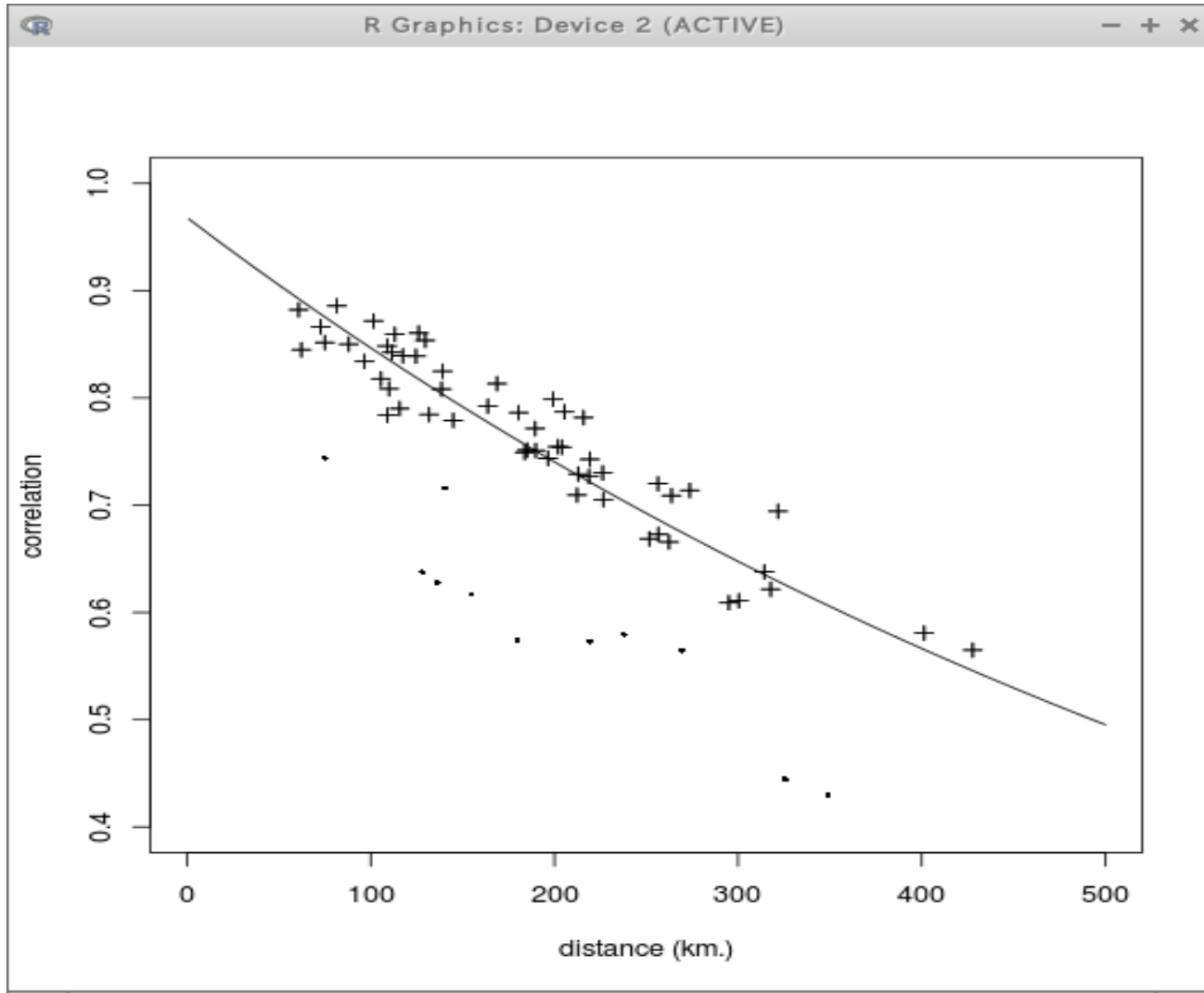
ポリゴンデータ

リレーショナルデータベースを使った三次元プロットの例

Forecasts from ARIMA(2,1,3) with drift



データによる将来予測の例 (ARIMAを使用)



トレンドの検出, 異常値の発見の例

まとめ



- リレーショナル・データベースは簡単
- データ形式と制約を簡単に記述できる
- リレーショナル・データベースは様々な種類のデータを扱うことができる
 - 空間的
 - 時間的
- データ分析手法にはすでに多くの種類がある