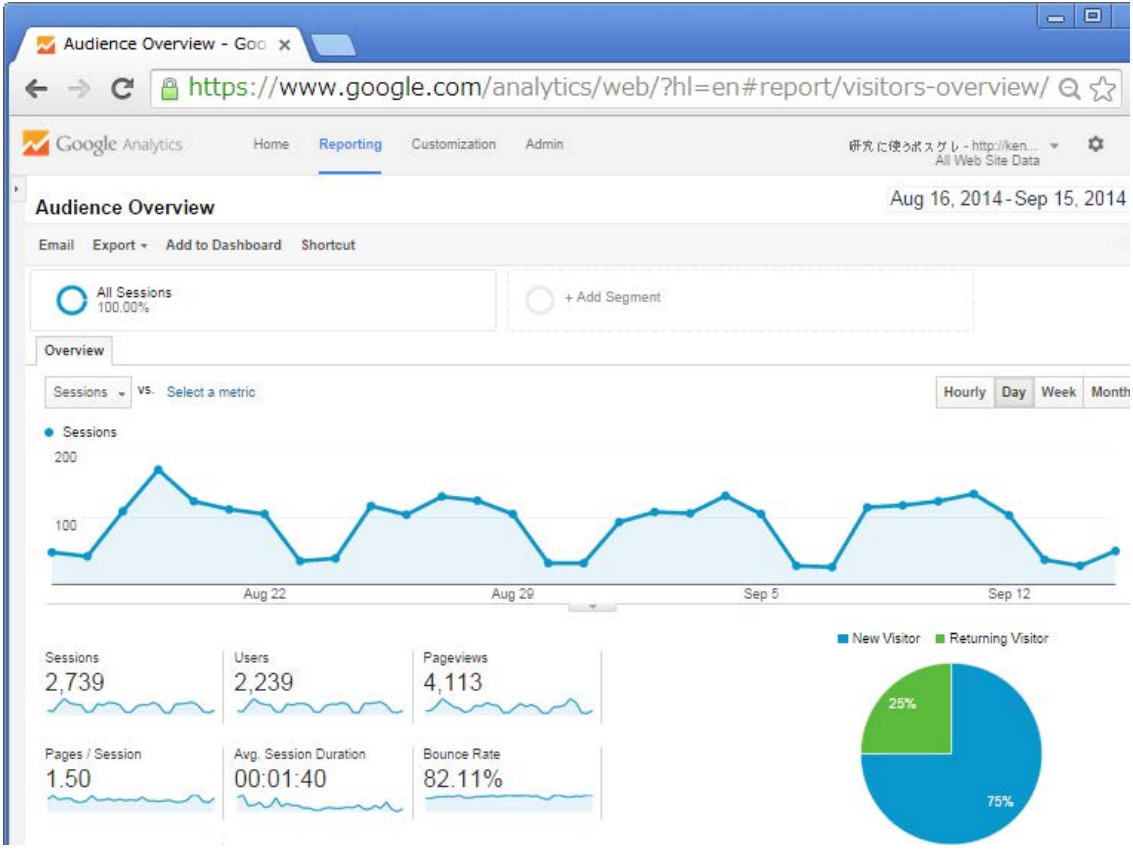
【使ってみた】

PostgreSQL 9.5 の新しい GROUP BY

＋　Google Analytics API の JSON を
PostgreSQL でテーブル化

2015/05/30　Satoshi Koda
http://kenpg.bitbucket.org/

今日のデータ：Google Analytics によるブログアクセス



Google の Analytics API の Core Reporting API（v3.0）

（クエリパラメータ例）

- dimensions=ga:date, ga:country, ga:city, ga:operatingSystem, ……
- metrics=ga:sessions
- start-date=2015-04-01&end-date=1daysAgo

（レスポンスの JSON 一部）

```
{ "columnHeaders": [ {
    "name": "ga:date",
    "columnType": "DIMENSION",
    "dataType": "STRING" }, ……
```

```
{ "rows": [ [
    "20150401",
    "Japan",
    "Saitama",
    "Windows",
    "Firefox",
    "36.0", ……
```

JSON データの処理

1. 一回のレスポンスの JSON を一行ずつテキストファイルに保存

```
{ "columnHeaders" : [ { "name" : "ga:date", "columnType" : ……
{ "columnHeaders" : [ { "name" : "ga:date", "columnType" : ……
{ "columnHeaders" : [ { "name" : "ga:date", "columnType" : ……
```

2. テキストファイルを pg_read_file 関数で読み込み、一時テーブル化

```
CREATE TABLE hoge AS
SELECT unnest (
  string_to_array( pg_read_file( ' ファイルパス ' ), E'¥n' )
) :: jsonb;
```

3. JSON のデータ部分（2 次元配列）をクエリで表形式に変換

（次頁）

JSON データの処理（続）

3. JSON のデータ部分（2 次元配列）を表形式に変換

```
CREATE TABLE fuga AS
SELECT
  ( ary->>0 ) :: date AS ymd, -- 添字が 0 から（配列型は 1 から）
  ( ary->>1 ) :: int AS h24,
  ary->>2 AS country,
  ary->>3 AS os,
  ary->>4 AS browser,
  ( ary->>5 ) :: int AS sessions ( , …… )
FROM (
  SELECT jsonb_array_elements(jsonb_dat -> 'rows') AS ary
  FROM hoge
) foo;
```

できたテーブル

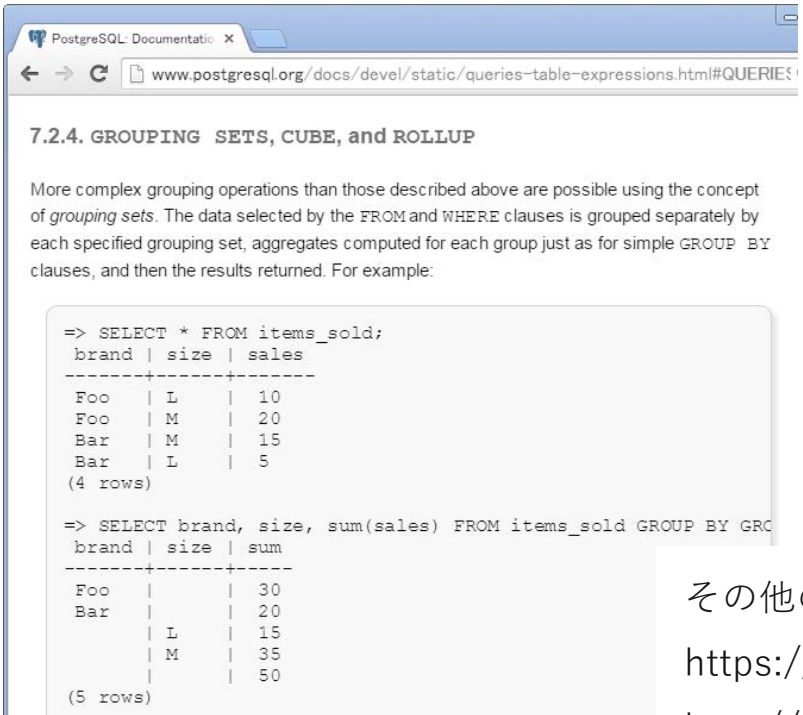
Output pane

Data Output | Explain | Messages | History

	data_id integer	ymd date	h24 inte	city text	os text	browser text	browser_ver text	page_title text	page_views integer
33866	2	2015-05-28	20	Chuo	Windows	Internet Exp	11.0	psycopg2 でリアルタ	1
33867	2	2015-05-28	20	Osaka	Windows	Firefox	31.0	コードをポスグレ内	1
33868	2	2015-05-28	20	Saitama	Macintosh	Chrome	43.0.2357.65	CentOS 6.6 で環境構	1
33869	2	2015-05-28	21	Kawasaki	Windows	Chrome	43.0.2357.81	pgAdmin クエリツー	1
33870	2	2015-05-28	21	Nagoya	Windows	Internet Exp	9.0	Excel データインポ	1
33871	2	2015-05-28	21	Osaka	Android	Chrome	42.0.2311.111	ポータブルなRとPost	1
33872	2	2015-05-28	21	Osaka	Windows	Chrome	43.0.2357.81	VBS で Excel データ	1
33873	2	2015-05-28	21	Osaka	Windows	Chrome	43.0.2357.81	psql + COPY コマン	1
33874	2	2015-05-28	22	Izumi	Linux	Firefox	37.0	CentOS 6.6 で環境構	1
33875	2	2015-05-28	22	Kawasaki	Windows	Internet Exp	11.0	R + Qt で簡易コンソ	1
33876	2	2015-05-28	22	Kochi	Macintosh	Firefox	38.0	リサンプリングを用	1
33877	2	2015-05-28	22	Osaka	Android	Chrome	42.0.2311.111	CentOS で PL/R をイ	1
33878	2	2015-05-28	22	Shibuya	iOS	Safari	8.0	CentOS 6.6 で環境構	1
33879	2	2015-05-28	22	Urayasu	Windows	Firefox	38.0	日付の妥当性を PL/	1
33880	2	2015-05-28	22	Yokohama	Windows	Chrome	43.0.2357.81	東芝 R732 : SSD 換	1
33881	2	2015-05-28	23	Chiyoda	Windows	Chrome	43.0.2357.81	pgAdmin クエリツー	1
33882	2	2015-05-28	23	Fukuoka	Macintosh	Safari	8.0.6	pgAdmin の便利機能	1
33883	2	2015-05-28	23	Fukuoka	Macintosh	Safari	8.0.6	pgAdmin：ショートカ	1
33884	2	2015-05-28	23	Osaka	iOS	Safari	8.0	横罫線だけの表を PI	1
33885	2	2015-05-28	23	San Diego	Windows	Firefox	38.0	CentOS 6.6 で環境構	1
33886	2	2015-05-28	23	Shinjuku	Windows	Firefox	38.0	紛わしい関数 readl	1
33887	2	2015-05-28	23	Yokohama	Macintosh	Chrome	43.0.2357.81	PostgreSQL 9.4 Win	1

【本題】PostgreSQL 9.5 の新しい GROUP BY 機能

参考：英語ドキュメント devel 版 → 7.2. Table Expressions → 7.2.4



GROUP BY

- GROUPING SETS (…)
- CUBE (…)
- ROLLUP (…)

その他の参考

https://wiki.postgresql.org/wiki/

http://www.depesz.com/ など

普通の GROUP BY

```
pg95plr=#
pg95plr=# -- 普通の GROUP BY  (列名でなく列番号を使用)
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tb1
pg95plr-# GROUP BY 1, 2
pg95plr-# HAVING sum(sessions) > 20
pg95plr-# ORDER BY 1, 2
pg95plr-# ;
    os     |      browser       |  sum
-----------+--------------------+-------
 Android   | Android Browser    |   144
 Android   | Chrome             |   279
 Android   | Firefox            |    78
 Android   | Opera              |    26
 iOS       | Chrome             |    35
 iOS       | Safari             |   471
 Linux     | Chrome             |   145
 Linux     | Firefox            |   490
 Macintosh | Chrome             |   950
 Macintosh | Firefox            |   278
 Macintosh | Safari             |   602
 Windows   | Chrome             | 10285
 Windows   | Firefox            |  4900
 Windows   | Internet Explorer  |  8489
 Windows   | Iron               |    62
 Windows   | Opera              |   276
(16 rows)
```

GROUPING SETS：GROUP BY を集計行に変える or 足す

```
pg95plr=#
pg95plr=# -- 普通の GROUP BY （列名でなく列番
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tbl
pg95plr-# GROUP BY 1, 2
pg95plr-# HAVING sum(sessions) > 20
pg95plr-# ORDER BY 1, 2
pg95plr-# ;
    os     |      browser       |  sum
-----------+--------------------+-------
 Android   | Android Browser    |   144
 Android   | Chrome             |   279
 Android   | Firefox            |    78
 Android   | Opera              |    26
 iOS       | Chrome             |    35
 iOS       | Safari             |   471
 Linux     | Chrome             |   145
 Linux     | Firefox            |   490
 Macintosh | Chrome             |   950
 Macintosh | Firefox            |   278
 Macintosh | Safari             |   602
 Windows   | Chrome             | 10285
 Windows   | Firefox            |  4900
 Windows   | Internet Explorer  |  8489
 Windows   | Iron               |    62
 Windows   | Opera              |   276
(16 rows)
```

```
pg95plr=#
pg95plr=# -- GROUPING SETS にすると
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tbl
pg95plr-# GROUP BY GROUPING SETS (1, 2)
pg95plr-# HAVING sum(sessions) > 20
pg95plr-# ORDER BY 1, 2
pg95plr-# ;
    os     |      browser       |  sum
-----------+--------------------+-------
 Android   |                    |   532
 iOS       |                    |   519
 Linux     |                    |   647
 Macintosh |                    |  1847
 Windows   |                    | 24054
           | Android Browser    |   144
           | Chrome             | 11700
           | Firefox            |  5752
           | Internet Explorer  |  8490
           | Iron               |    72
           | Opera              |   319
           | Safari             |  1080
(12 rows)

pg95plr=#
```

カッコ内での ( ) は、どの列も絞り込まない = 総計行

```
pg95plr=#
pg95plr=# -- GROUPING SETS にすると
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tb1
pg95plr-# GROUP BY GROUPING SETS (1, 2)
pg95plr-# HAVING sum(sessions) > 20
pg95plr-# ORDER BY 1, 2
pg95plr-# ;
    os     |      browser       |  sum
-----------+--------------------+-------
 Android   |                    |   532
 iOS       |                    |   519
 Linux     |                    |   647
 Macintosh |                    |  1847
 Windows   |                    | 24054
           | Android Browser    |   144
           | Chrome             | 11700
           | Firefox            |  5752
           | Internet Explorer  |  8490
           | Iron               |    72
           | Opera              |   319
           | Safari             |  1080
(12 rows)

pg95plr=#
```

```
pg95plr=#
pg95plr=# -- GROUPING SETS ( () ) は総計行
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tb1
pg95plr-# GROUP BY GROUPING SETS (1, 2, ()) --
pg95plr-# HAVING sum(sessions) > 20
pg95plr-# ORDER BY 1, 2
pg95plr-# ;
    os     |      browser       |  sum
-----------+--------------------+-------
 Android   |                    |   532
 iOS       |                    |   519
 Linux     |                    |   647
 Macintosh |                    |  1847
 Windows   |                    | 24054
           | Android Browser    |   144
           | Chrome             | 11700
           | Firefox            |  5752
           | Internet Explorer  |  8490
           | Iron               |    72
           | Opera              |   319
           | Safari             |  1080
           |                    | 27631
(13 rows)

pg95plr=#
```

総計行を足した出力が、GROUPING SETS で簡単になる

```
pg95plr=#
pg95plr=# -- クロス表の例
pg95plr=# SELECT os, coalesce(sum(sessions) FILTER (WHERE browser = 'Chrome'), 0) chrome,
pg95plr-#     coalesce(sum(sessions) FILTER (WHERE browser = 'Firefox'), 0) firefox,
pg95plr-#     coalesce(sum(sessions) FILTER (WHERE browser = 'Internet Explorer'), 0) ie,
pg95plr-#     coalesce(sum(sessions), 0) all_brosers
pg95plr-# FROM test.ga_tb1
pg95plr-# WHERE os IN ('Linux', 'Macintosh', 'Windows')
pg95plr-#     AND browser IN ('Chrome', 'Firefox', 'Internet Explorer')
pg95plr-# GROUP BY 1
pg95plr-# ;
    os     | chrome | firefox |  ie  | all_brosers
-----------+--------+---------+------+-------------
 Linux     |    145 |     490 |    0 |         635
 Macintosh |    950 |     278 |    0 |        1228
 Windows   |  10285 |    4900 | 8489 |       23674
(3 rows)

pg95plr=# -- ↑ これに総計行を足すのが、従来は面倒
```

総計行を足した出力が、GROUPING SETS で簡単になる

```
pg95plr=#
pg95plr=# -- 9.5からは総計行を足すのが簡単に
pg95plr=# SELECT os, coalesce(sum(sessions) FILTER (WHERE browser = 'Chrome'), 0) chrome,
pg95plr-#     coalesce(sum(sessions) FILTER (WHERE browser = 'Firefox'), 0) firefox,
pg95plr-#     coalesce(sum(sessions) FILTER (WHERE browser = 'Internet Explorer'), 0) ie,
pg95plr-#     coalesce(sum(sessions), 0) all_brosers
pg95plr-# FROM test.ga_tbl
pg95plr-# WHERE os IN ('Linux', 'Macintosh', 'Windows')
pg95plr-#     AND browser IN ('Chrome', 'Firefox', 'Internet Explorer')
pg95plr-# GROUP BY GROUPING SETS (1, ()) -- ここを変えるだけ
pg95plr-# ;
    os     | chrome | firefox |  ie  | all_brosers
-----------+--------+---------+------+-------------
 Linux     |    145 |     490 |    0 |         635
 Macintosh |    950 |     278 |    0 |        1228
 Windows   |  10285 |    4900 | 8489 |       23674
           |  11380 |    5668 | 8489 |       25537
(4 rows)
```

従来：クロス表に総計行を足すには UNION ALL など面倒

```
pg95plr=#
pg95plr=# -- 従来は、クロス表に総計行を足すには UNION ALL
pg95plr=# WITH cte AS (
pg95plr(#     SELECT os, coalesce(sum(sessions) FILTER (WHERE browser = 'Chrome'), 0) chrom
pg95plr(#         coalesce(sum(sessions) FILTER (WHERE browser = 'Firefox'), 0) firefox,
pg95plr(#         coalesce(sum(sessions) FILTER (WHERE browser = 'Internet Explorer'), 0) i
pg95plr(#         coalesce(sum(sessions), 0) all_browsers
pg95plr(#     FROM test.ga_tb1
pg95plr(#     WHERE os IN ('Linux', 'Macintosh', 'Windows')
pg95plr(#         AND browser IN ('Chrome', 'Firefox', 'Internet Explorer')
pg95plr(#     GROUP BY 1
pg95plr(# )
pg95plr-# SELECT * FROM cte
pg95plr-# UNION ALL
pg95plr-# SELECT NULL::text, sum(chrome), sum(firefox), sum(ie), sum(all_browsers)
pg95plr-# FROM cte
pg95plr-# ;
    os     | chrome | firefox |  ie  | all_browsers
-----------+--------+---------+------+--------------
 Linux     |    145 |     490 |    0 |          635
 Macintosh |    950 |     278 |    0 |         1228
 Windows   |  10285 |    4900 | 8489 |        23674
           |  11380 |    5668 | 8489 |        25537
(4 rows)
```

GROUPING SETS の例

```
pg95plr=#
pg95plr=# -- 総計行だけ足す例
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tbl
pg95plr-# GROUP BY GROUPING SETS ( (1, 2), () ) -- ( 1, 2, () ) とは別
pg95plr-# HAVING sum(sessions) > 20
pg95plr-# ORDER BY 1 NULLS FIRST, 2 NULLS FIRST
pg95plr-# ;
    os     |      browser       |  sum
-----------+--------------------+-------
           |                    | 27631
 Android   | Android Browser    |   144
 Android   | Chrome             |   279
 Android   | Firefox            |    78
 Android   | Opera              |    26
 iOS       | Chrome             |    35
 iOS       | Safari             |   471
 Linux     | Chrome             |   145
 Linux     | Firefox            |   490
 Macintosh | Chrome             |   950
 Macintosh | Firefox            |   278
 Macintosh | Safari             |   602
 Windows   | Chrome             | 10285
 Windows   | Firefox            |  4900
 Windows   | Internet Explorer  |  8489
 Windows   | Iron               |    62
 Windows   | Opera              |   276
(17 rows)
```

```
pg95plr=#
pg95plr=# -- GROUPING SETS ( () ) は総計行
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tbl
pg95plr-# GROUP BY GROUPING SETS (1, 2, ()) -- ここ
pg95plr-# HAVING sum(sessions) > 20
pg95plr-# ORDER BY 1, 2
pg95plr-# ;
    os     |      browser       |  sum
-----------+--------------------+-------
 Android   |                    |   532
 iOS       |                    |   519
 Linux     |                    |   647
 Macintosh |                    |  1847
 Windows   |                    | 24054
           | Android Browser    |   144
           | Chrome             | 11700
           | Firefox            |  5752
           | Internet Explorer  |  8490
           | Iron               |    72
           | Opera              |   319
           | Safari             |  1080
           |                    | 27631
(13 rows)
```

GROUPING SETS で、普通の GROUP BY ＋小計行

```
pg95plr=# -- 小計行も加えてみる
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tbl
pg95plr-# GROUP BY GROUPING SETS ( (1, 2), 1, () )
pg95plr-# HAVING sum(sessions) > 20
pg95plr-# ORDER BY 1, 2
pg95plr-# ;
```

os	browser	sum
Android	Android Browser	144
Android	Chrome	279
Android	Firefox	78
Android	Opera	26
Android		532
iOS	Chrome	35
iOS	Safari	471
iOS		519
Linux	Chrome	145
Linux	Firefox	490
Linux		647
Macintosh	Chrome	950
Macintosh	Firefox	278
Macintosh	Safari	602
Macintosh		1847
Windows	Chrome	10285
Windows	Firefox	4900
Windows	Internet Explorer	8489
Windows	Iron	62
Windows	Opera	276
Windows		24054
		27631

(22 rows)

GROUPING SETS でクロス表と同じデータ．ただし…

```
pg95plr=# -- クロス表と同じデータ、ただし存在しない組み合わせは出ない
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tbl
pg95plr-# WHERE os IN ('Linux', 'Macintosh', 'Windows')
pg95plr-#     AND browser IN ('Chrome', 'Firefox', 'Internet Explorer')
pg95plr-# GROUP BY GROUPING SETS ( (1, 2), 1, 2, () )
pg95plr-# ORDER BY 1 NULLS FIRST, 2 NULLS FIRST
pg95plr-# ;
    os     |      browser      |  sum
-----------+-------------------+-------
           |                   | 25537
           | Chrome            | 11380
           | Firefox           |  5668
           | Internet Explorer |  8489
 Linux     |                   |   635
 Linux     | Chrome            |   145
 Linux     | Firefox           |   490
 Macintosh |                   |  1228
 Macintosh | Chrome            |   950
 Macintosh | Firefox           |   278
 Windows   |                   | 23674
 Windows   | Chrome            | 10285
 Windows   | Firefox           |  4900
 Windows   | Internet Explorer |  8489
(14 rows)
```

GROUPING SETS でクロス表と同じデータ．ただし…

```
pg95plr=# -- クロス表と同じデータ、ただし存在しない組み合わせは出ない
pg95plr=# SELECT os, browser, sum(sessions)
pg95plr-# FROM test.ga_tbl
pg95plr-# WHERE os IN ('Linu
pg95plr-#     AND browser IN
pg95plr-# GROUP BY GROUPING
pg95plr-# ORDER BY 1 NULLS F
pg95plr-# ;
```

os	browser	sum
		25537
	Chrome	11380
	Firefox	5668
	Internet Explorer	8489
Linux		635
Linux	Chrome	145
Linux	Firefox	490
Macintosh		1228
Macintosh	Chrome	950
Macintosh	Firefox	278
Windows		23674
Windows	Chrome	10285
Windows	Firefox	4900
Windows	Internet Explorer	8489

(14 rows)

os	chrome	firefox	ie	all_browsers
Linux	145	490	0	635
Macintosh	950	278	0	1228
Windows	10285	4900	8489	23674
	11380	5668	8489	25537

CUBE と ROLLUP：ある種の GROUPING SETS の短縮形

＊ GROUP BY GROUPING SETS ((1, 2), 1, 2, ( ))

||

GROUP BY CUBE (1, 2) …… 各列で絞る or 絞らないの、全組み合わせ

＊ GROUP BY GROUPING SETS ((1, 2), 1, ( ))

||

GROUP BY ROLLUP (1, 2) …… 片方向に（ツリー状に）絞る組み合わせ

■ まとめ

- これまで別々に出していた総計・小計が、一回のクエリで済む
- 外部アプリなどからクエリして表やグラフ作る際は、役立つかも
  （割合の算出、グラフ描画範囲の確定などに総計・小計をよく使う）
- 元々ない組み合わせは、出ない（クエリ表のゼロ埋めはない）