

スタートアップゼミについて

スタートアップゼミとは？

研究室に新しく配属された皆さんを中心に，研究に向けた基礎を習得してもらうゼミ

テーマ

1. **都市形成史分析**：都市の成り立ちを読み解く（2回）
2. **交通行動モデル**：交通行動をモデル化する（3回）
3. **交通量配分モデル**：交通量をネットワーク上に配分する（3回）

- 毎週木曜日 8:30-10:30
- 計11回を予定（レクチャー×8，課題発表×3）
- 各回課題あり

シラバス（予定）

4/11 行動データ分析基礎

4/18 最短経路探索

4/25 離散選択モデル

GW休み

5/9 均衡配分

5/16 課題発表①

5/23 離散連続モデル

5/30 都市経済モデル

6/6 課題発表②

6/13 都市形成史分析1

6/20 都市形成史分析2

6/27 課題発表③

スタートアップゼミ#1

行動データ分析基礎

M1 飯塚卓哉

目次

- はじめに
- 基礎集計 - Excel
- 可視化と簡単な空間分析 - GIS

目標

行動データを見て，分析の下準備・簡単な分析ができるようになる

はじめに

何がしたいの？

羽藤研では人の行動データから色々なことを分析するというのをよくやります

= さまざまな人の選択行動（意思決定）の結果



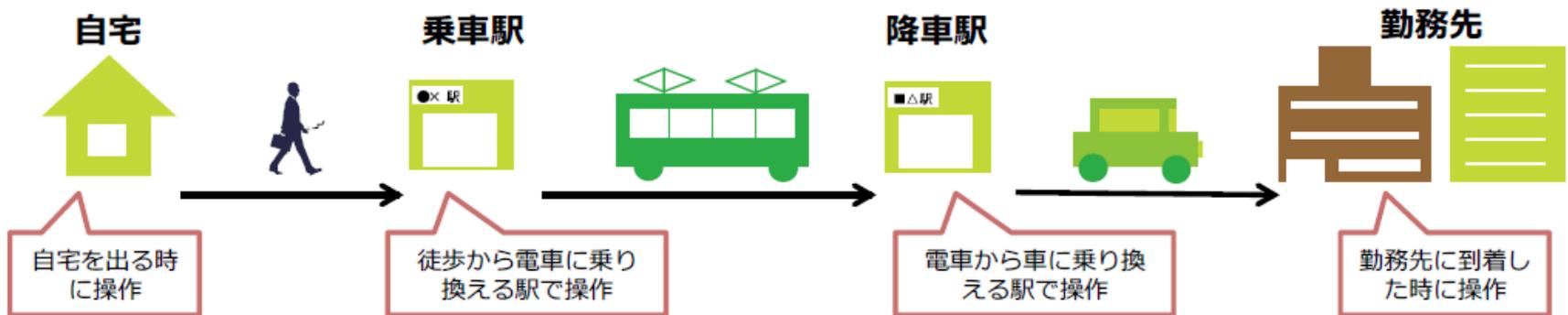
人の意思決定メカニズムをモデル化

行動モデル

- どこに行くのか
 - どの移動手段選ぶのか
 - どの道を通るのか
 - そこにどれくらいの時間いるのか
- etc...

行動データとは？

- パーソントリップ (PT) データ
 - **プローブパーソン (PP) データ** → GPS機能を搭載した携帯電話と移動通信機器と連動したWebダイアリーを用いて、モニタの移動活動記録と数秒間隔の位置情報を取得できる
 - Wi-fiデータ
 - Bluetooth
 - 加速度データ
 - 映像データ
 - アンケート調査データ
- などなど
- ✓ 大量, 詳細な移動データ
 - ✓ day-to-dayの行動記録 (同一個人の複数日にわたる行動履歴)



PPデータは大きく, **location data**, **trip data**の二つ

PPデータの説明

■ location data：各時点での位置情報のデータ（5～10秒間隔）

tripID, locationID, userID, 移動手段,
時刻, 位置座標, 測位モード

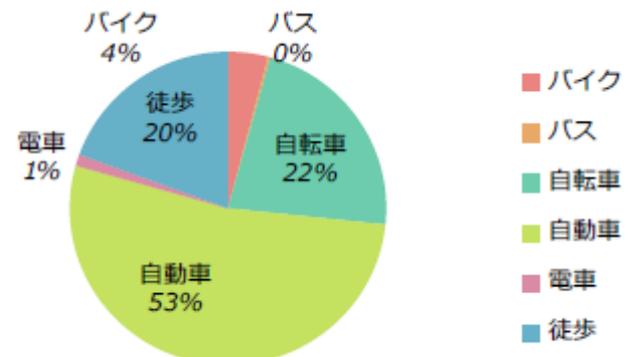


	A	B	C	D	E	F	G	H	I	J	K
1	トリップID	測位データID	移動手段	ユーザーコード	測位日時	緯度（日本測地系）	経度（日本測地系）	緯度（世界測地系）	経度（世界測地系）	測位モード	有効性
2	15375	2023091	バイク	ec029	2007/2/19 1:14	33.8348734	132.757748	33.83817494	132.7552057	4	1
3	15375	2023092	バイク	ec029	2007/2/19 1:15	33.8304471	132.768558	33.8337493	132.766015	4	1
4	15375	2023093	バイク	ec029	2007/2/19 1:15	33.8350557	132.7634616	33.83835732	132.7609188	4	1
5	15375	2023094	バイク	ec029	2007/2/19 1:15	33.8348359	132.7575173	33.83813744	132.754975	4	1
6	15375	2023095	バイク	ec029	2007/2/19 1:15	33.8305974	132.7684292	33.83389959	132.7658862	4	1
7	15375	2023096	バイク	ec029	2007/2/19 1:16	33.8348734	132.757748	33.83817494	132.7552057	4	1
8	15375	2023097	バイク	ec029	2007/2/19 1:16	33.8368153	132.7709617	33.84011686	132.7684182	1	1

PPデータの説明

trip data : トリップごとのデータ

tripID, userID, 移動目的, 移動手段,
出発・到着時刻, 出発地・到着地位置座標 etc...

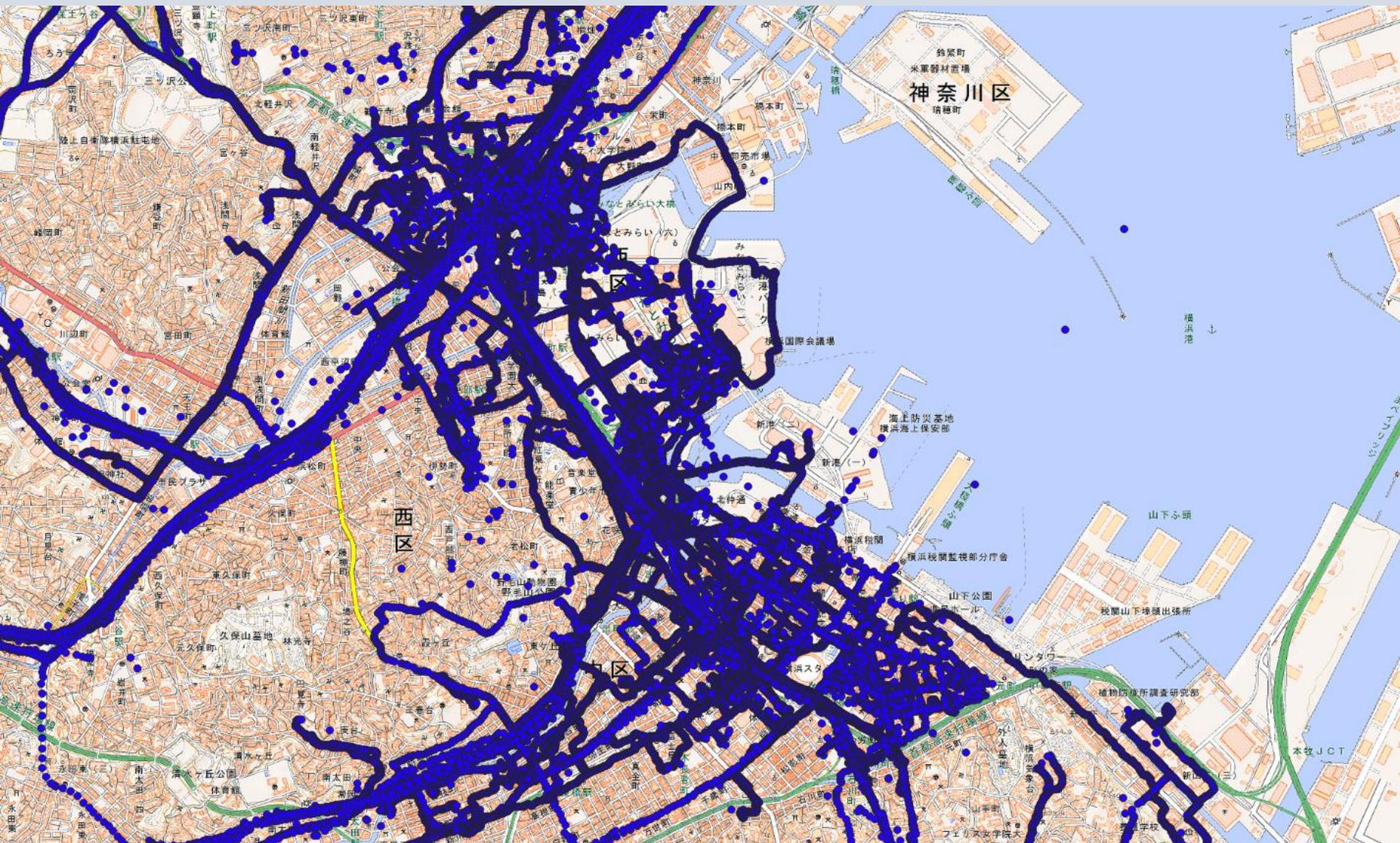


	A	B	C	D	E	F	G	H	I	J
1	トリップID	ユーザーコード	目的コード	目的	出発日付	到着日付	トリップ時間_秒	出発地コード	出発地	出発地ユーザー施設名
2	15521	ec003	410	食事	2007/2/19 18:03	2007/2/19 18:41	2314	10578	松山河川国道事務所	松山河川国道事務所
3	15556	ec003	200	帰宅	2007/2/19 21:46	2007/2/19 22:14	1648	10581	なが坂	なが坂
4	15595	ec003	100	出勤・登校	2007/2/20 7:51	2007/2/20 8:19	1635	10579	自宅	自宅
5	15880	ec003	0	--	2007/2/20 21:03	2007/2/20 21:03	0			
6	15882	ec003	100	出勤・登校	2007/2/21 7:46	2007/2/21 8:11	1459	10579	自宅	自宅
7	15944	ec003	310	業務	2007/2/21 11:21	2007/2/21 12:57	5770	10578	松山河川国道事務所	松山河川国道事務所
8	16008	ec003	300	帰社・帰校	2007/2/21 15:58	2007/2/21 17:04	3958	10580	今治商工会議所	今治商工会議所
9	16068	ec003	200	帰宅	2007/2/21 19:19	2007/2/21 19:40	1258	10578	松山河川国道事務所	松山河川国道事務所

「トリップ」が何かわからない人はこちら

<http://www.mlit.go.jp/crd/tosiko/pt.html> 国土交通省-PT調査とは？

可視化してみる……



どうやって分析するの？

分析の流れ

(ものすごく簡単な例)

• データの特性を知る

↓
分析を行う前に、そのデータはどのような情報を取得したものなのか、データから何が分かるのかを把握しておきましょう。

• 基礎集計を行い、傾向を探る

↓
様々な属性を掛け合わせて何と何に相関があるのか、その行動の要因となっているものは何か、仮説を立てます。

• モデルを構築し、推定を行う

↓
基礎集計から仮説がたったら、モデルを構築して因果関係を定量的に分析します。その選択行動に効いている要因とは何かを把握します。



色んなソフトを使います

基礎集計 - Excel

基礎集計とは

データ分析をする前にする下ごしらえ的なもの

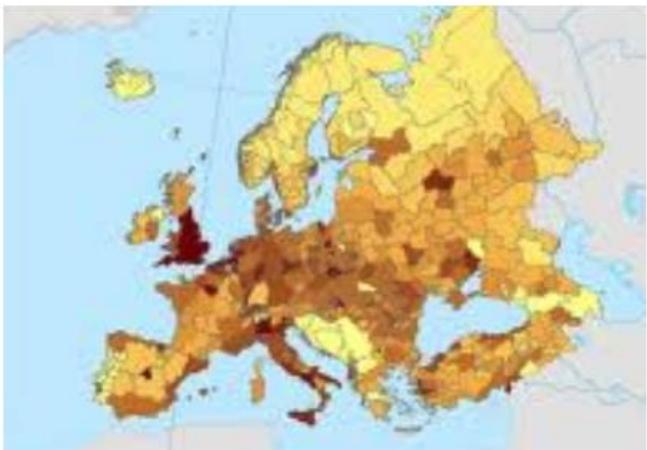
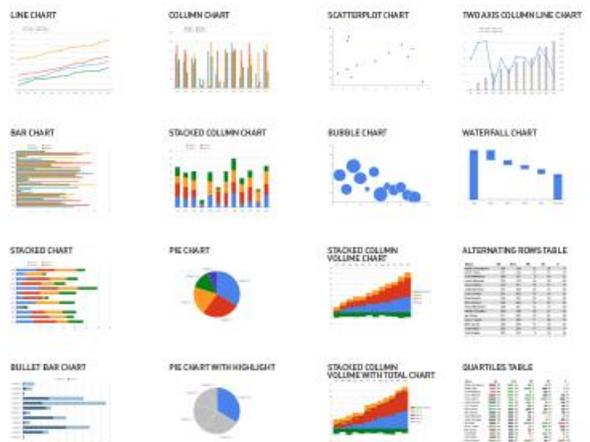
- 全体の傾向をみる
- データの分布をみる
- 異常値を発見する
- 欠損値を発見する
- モデルに入れる変数の検討をつける などなど...

集計に使うソフトウェア

① データ整理/正規化 **R/Python/Java/Excel...**

② データ集計

③ 可視化 **R/GIS/Excel/Google Earth...**



基礎集計の方法

■ データ整理/正規化

まず、自分の目的に合わせてデータを整理する必要があります。

▶ データクリーニング / 補正 / マーケット・セグメンテーション...

- トリップデータ：1行=1回の移動

80010	80010	自転車	a0452	33.84804551	132.7452547	33.83867924	132.7637083	2007/12/12 07:55:37	2007/12/12 08:09:13	816
80568	80568	バイク	a0452	33.83873823	132.7636493	33.84797582	132.7451259	2007/12/12 20:15:15	2007/12/12 20:25:38	623
80696	80696	バイク	a0452	33.84802941	132.7452654	33.83827687	132.7633971	2007/12/13 07:59:09	2007/12/13 08:11:28	739
81160	81160	バイク	a0452	33.83880802	132.7635795	33.8481152	132.7452279	2007/12/13 19:51:42	2007/12/13 20:13:03	1281
81292	81292	バイク	a0452	33.84806701	132.7453459	33.8388884	132.763145	2007/12/14 07:55:45	2007/12/14 08:06:26	641
81696	81696	タクシー	a0452	33.83913521	132.7650065	33.85430583	132.7845866	2007/12/14 17:54:21	2007/12/14 18:12:16	1075
81847	81847	タクシー	a0452	33.84004712	132.7711648	33.84906479	132.7445466	2007/12/14 23:37:43	2007/12/14 23:49:45	722
81938	81938	徒歩	a0452	33.84841567	132.7450348	33.85227482	132.7439179	2007/12/15 08:51:16	2007/12/15 09:11:08	1192
181939	181939	電車	a0452	33.85225658	132.7441217	33.83561827	132.7619284	2007/12/15 09:11:08	2007/12/15 09:19:36	508



- ツアーデータ：1行=1日の行動

1707	a0452	HWH	2	2007/12/13 08:11:28	2007/12/13 19:51:42	42014	バイク	バイク	33.83827687	132.7633971	出勤・登校
1708	a0452	HWOH	3	2007/12/14 08:06:26	2007/12/14 17:54:21	35275	バイク	タクシー	33.8388884	132.763145	出勤・登校
1709	a0452	HOOOHOH	7	2007/12/15 09:19:36	2007/12/15 11:11:24	6708	電車	バイク	33.83883482	132.7639658	null
1710	a0452	HOHOH	4	2007/12/16 16:24:51	2007/12/16 18:45:34	8443	自動車	自動車	33.83542843	132.7605057	買い物
1726	a0456	HWH	2	2007/12/11 08:21:54	2007/12/11 19:17:49	39355	自転車	自転車	33.84125952	132.7670664	出勤・登校
1727	a0456	HWH	2	2007/12/12 08:20:27	2007/12/12 17:46:39	33972	自転車	自転車	33.8414258	132.7667284	出勤・登校
1728	a0456	HWH	2	2007/12/13 08:19:20	2007/12/13 11:57:01	13061	自転車	自転車	33.84156527	132.7660472	出勤・登校
1729	a0456	HWH	2	2007/12/14 08:20:48	2007/12/14 21:59:30	49122	自転車	自転車	33.8415277	132.7677155	出勤・登校

基礎集計の方法

■ Excelでのクロス集計：ピボットテーブル

➤ [挿入]タブからピボットテーブルを選択 → データの範囲を選択してOK

ピボットテーブルの作成

分析するデータを選択してください。

テーブルまたは範囲を選択(S)

テーブル/範囲(D): trip!\$A\$1:\$CD\$7811

外部データソースを使用(M)

接続の選択(C)...

接続名:

このブックのデータ モデルを使用する(D)

ピボットテーブル レポートを配置する場所を選択してください。

新規ワークシート(N)

既存のワークシート(E)

場所(L):

複数のテーブルを分析するかどうかを選択

このデータをデータ モデルに追加する(M)

OK キャンセル

トリップID	ユーザー	目的地	出発地	出発地ユー	出発地施設	出発地施設	出発地施設	出発地施設	出発地施設	目的地	目的地	目的地ユー	目的地施設	目的地施設	目的地施設	目的地施設	目的地施設	目的地施設	目的地施設	到着地
15521	ec003		山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74										10581
15556	ec003		が坂	なが坂	190	飲食店・喫	33.83653	132.7724	33.83983	132.7698										10579
15595	ec003		宅	自宅	110	自宅	33.82894	132.7838	33.83225	132.7813										10578
15880	ec003		宅	自宅	110	自宅	33.82894	132.7838	33.83225	132.7813										10578
15882	ec003		山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74										10581
15944	ec003		山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74										10580
16008	ec003		治南工	今治南工	270	病院・医	34.06048	133.0002	34.06376	132.9976										10578
16068	ec003		山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74										10579
16135	ec003		宅	自宅	110	自宅	33.82894	132.7838	33.83225	132.7813										10578
16199	ec003		山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74										12072
16290	ec003		徳島道	西条出張	130	事務所・	33.90871	133.2058	33.91202	133.2032										10578
16406	ec003		宅	自宅	110	自宅	33.82894	132.7838	33.83225	132.7813										10578
16492	ec003		山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74										12073
16565	ec003		徳大工	学大	130	事務所・	33.84521	132.773	33.84851	132.7705										10578
16655	ec003	200	帰宅	#####	1466	10578	松山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74						10579
16982	ec003	400	買い物	#####	6916	10579	自宅	自宅	110	自宅	33.82894	132.7838	33.83225	132.7813						10579
17131	ec003	499	その他私用	#####	719	10579	自宅	自宅	110	自宅	33.82894	132.7838	33.83225	132.7813						12071
17147	ec003	400	買い物	#####	640	12071	実家	実家	140	個人宅	33.82268	132.7754	33.82599	132.7729						12074
17172	ec003	200	帰宅	#####	393	12075	いよ立花	立花駅	280	公共施設	33.8243	132.7757	33.82761	132.7732						12071
17227	ec003	410	食事	#####	6051	12071	実家	実家	140	個人宅	33.82268	132.7754	33.82599	132.7729						12071
17305	ec003	200	帰宅	#####	530	12071	実家	実家	140	個人宅	33.82268	132.7754	33.82599	132.7729						10579
17346	ec003	100	出勤・登	#####	1856	10579	自宅	自宅	110	自宅	33.82894	132.7838	33.83225	132.7813						10578
17611	ec003	200	帰宅	#####	1410	10578	松山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74						10579
17714	ec003	100	出勤・登	#####	986	10579	自宅	自宅	110	自宅	33.82894	132.7838	33.83225	132.7813						10578
17825	ec003	310	業務	#####	1120	10578	松山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74						12073
17831	ec003	300	帰社・帰	#####	1165	12073	愛媛大学	学大	130	事務所・	33.84521	132.773	33.84851	132.7705						10578
17868	ec003	310	業務	#####	1654	10578	松山河川	松山河川	120	勤務・通	33.81838	132.7425	33.82168	132.74						12076

基礎集計の方法

例) 移動手段と移動目的の関係が知りたい

行エリアに目的, 列エリアに移動手段, 値エリアにトリップIDを入れてデータの種類を「個数」とすると目的別移動手段分担率が出る

行ラベル	その他	タクシー	バイク	バス	自転車	自動車	船	待ち時間	電車	徒歩	飛行機	総計	
--	44	4	89	2	96	307	1	13	44	600			
その他私用		1	4	90	4	210	821	1	3	16	293	1	1444
帰社・帰校			3	20	30	66			11	85		215	
帰宅			15	232	7	470	1064	1	1	34	291	1	2116
業務			3	69		28	269	2	1	5	40		417
娯楽	1		2	14	3	31	137		1	89		2	280
出勤・登校			5	251	11	300	394		14	100			1075
食事			1	21	1	30	142		6	114			315
買い物				69	2	205	727		3	342			1348
総計	44	2	37	855	30	1400	3927	4	6	103	1398	4	7810

グラフにすると傾向がより分かりやすい。

(出勤・登校は二輪車が多いなど...)

可視化と簡単な空間分析 - GIS

(Geographic Information System)

■ ソフトウェアの紹介



ArcGIS

- 有料ソフト（ただし工学系の学生は無料で使用できる）
- 豊富なデータ分析機能，大規模データ処理にも強い
- 印刷やpdf出力がやりやすい

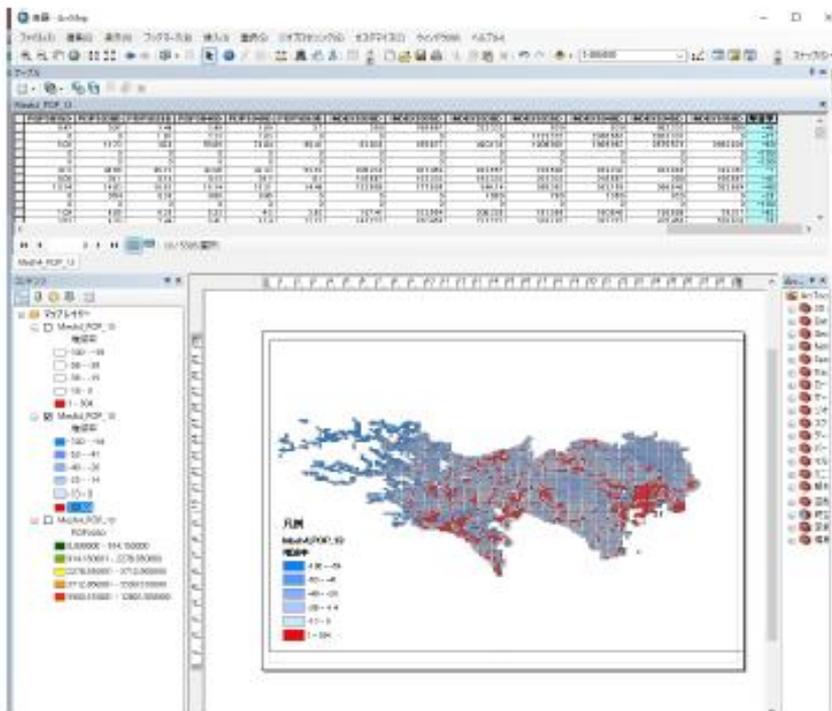


QGIS

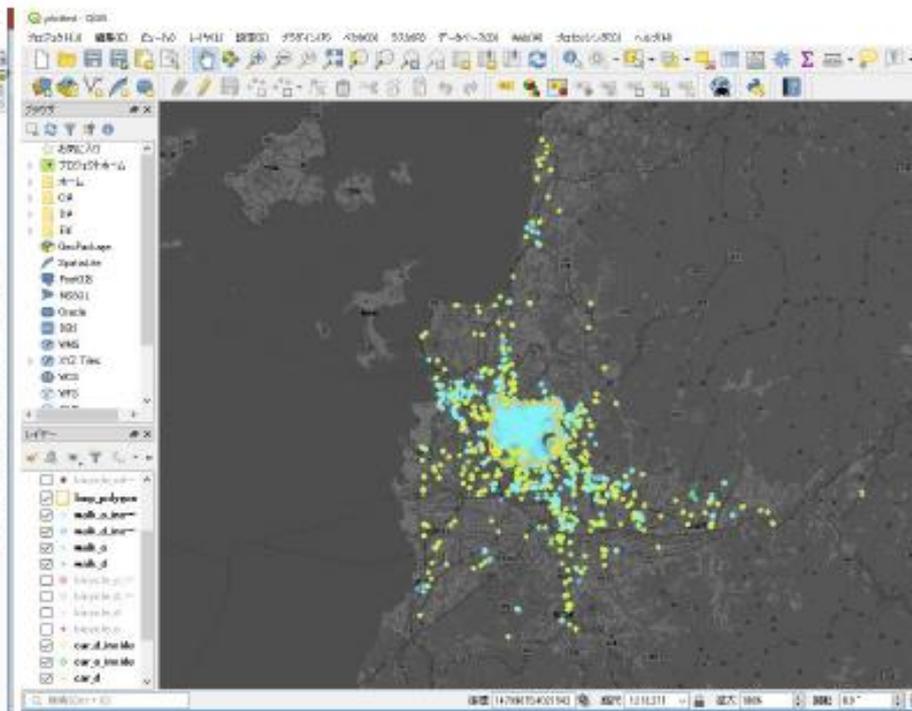
- 無料ソフト，オープンソース
- 直観的なUI，プラグインが豊富なので基本的な機能はArcGISと遜色ない
- フリーソフトなので動作が不安定になりがち

GISでできること

- 空間的に結びついた各種データを投影するためのツール



▲ ArcGIS



▲ QGIS

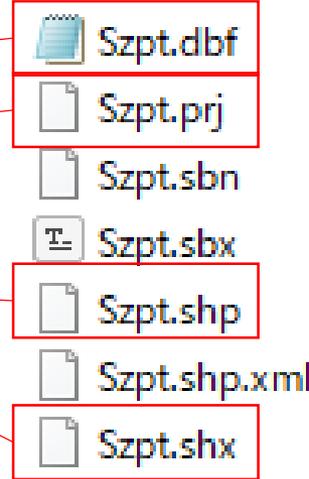
- 色々なデータを重ねて投影する (分析)
- GIS上で計算して結果を図次できる (描画)

GISの基礎1：使う・使えるデータ

① シェープファイル

図形情報と属性情報を持った地図データファイル

- .dbf 属性の情報が保存
- .prj 投影法の情報が保存
- .shp 図形の情報が保存
- .shx shpの図形とdbfの属性の対応関係が保存
- etc...

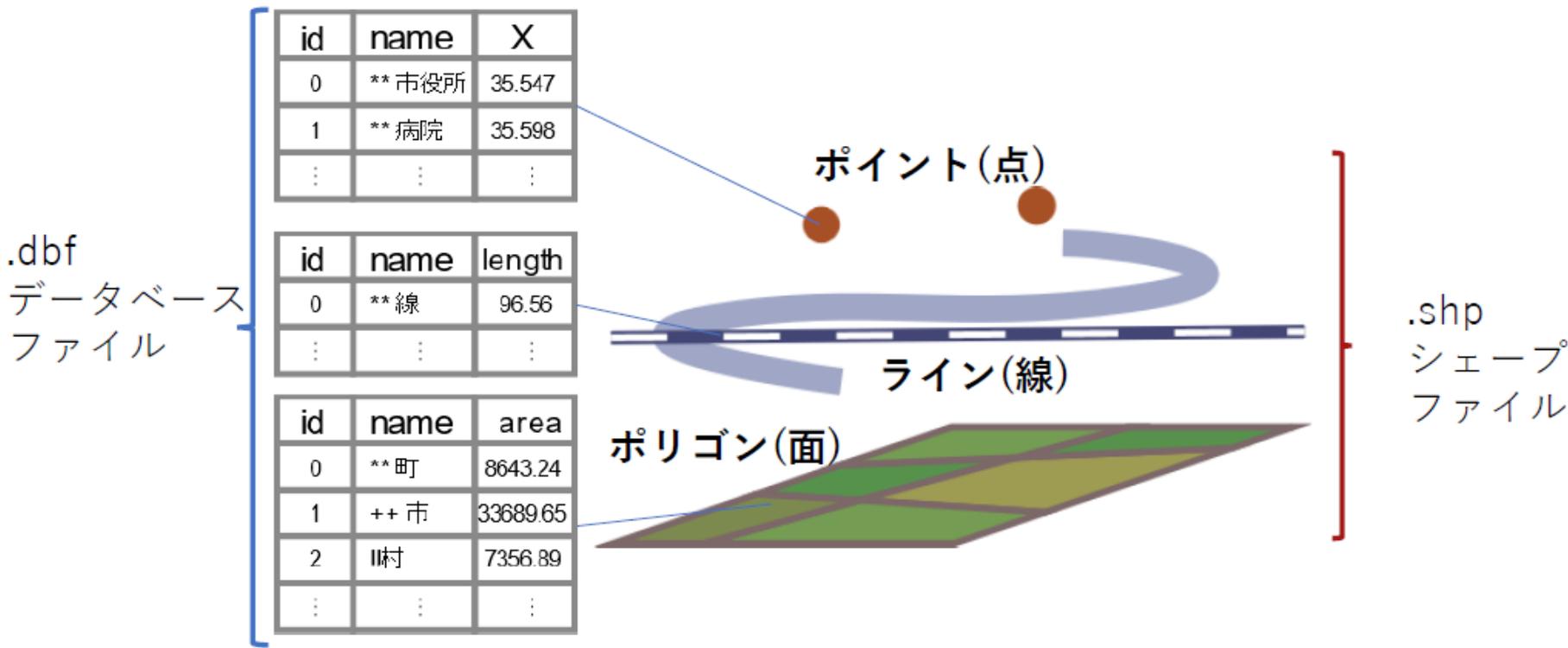


複数の役割を持ったこれらのデータセットで1セット

GISの基礎1：使う・使えるデータ

① シェープファイル

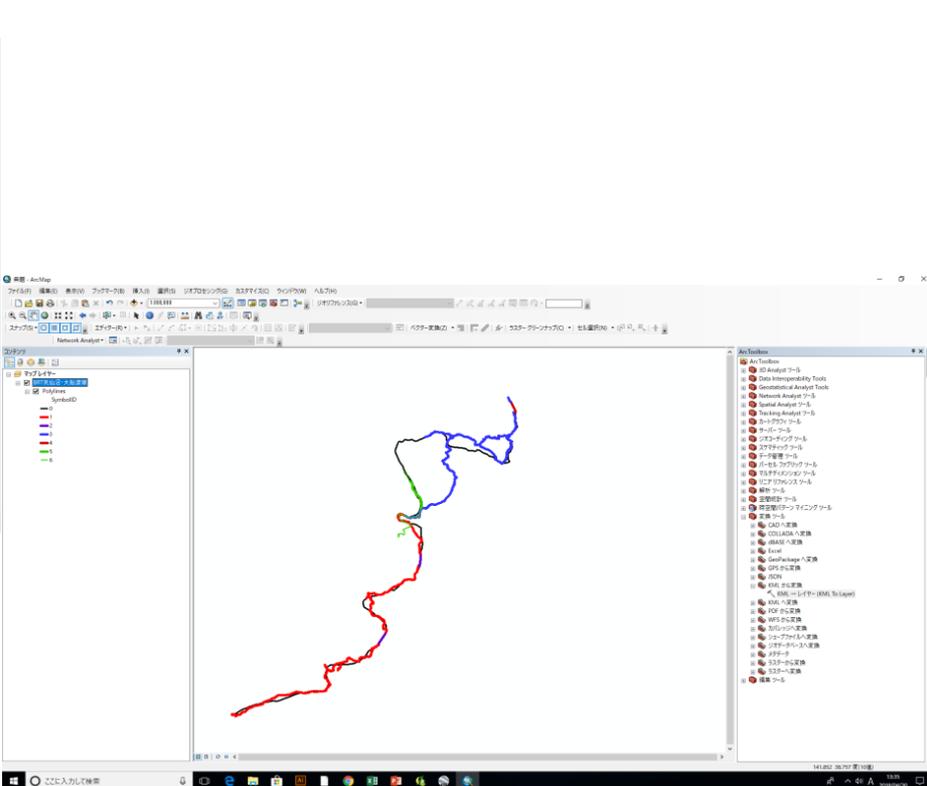
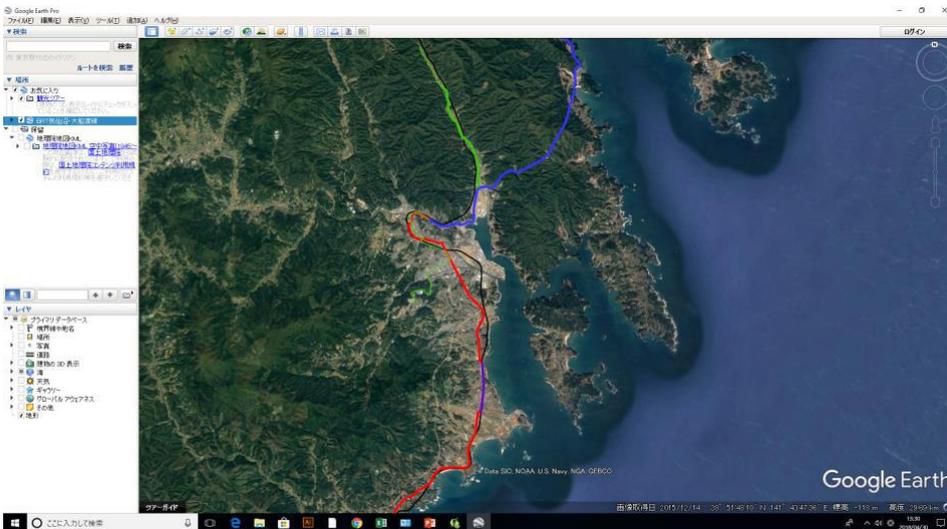
図形情報と属性情報を持った地図データファイル



GISの基礎1：使う・使えるデータ

③kmlファイル

図形情報と属性情報（関連コンテンツ）を持った地図データファイル
Google Earthなど一般の人でも良く使う形式



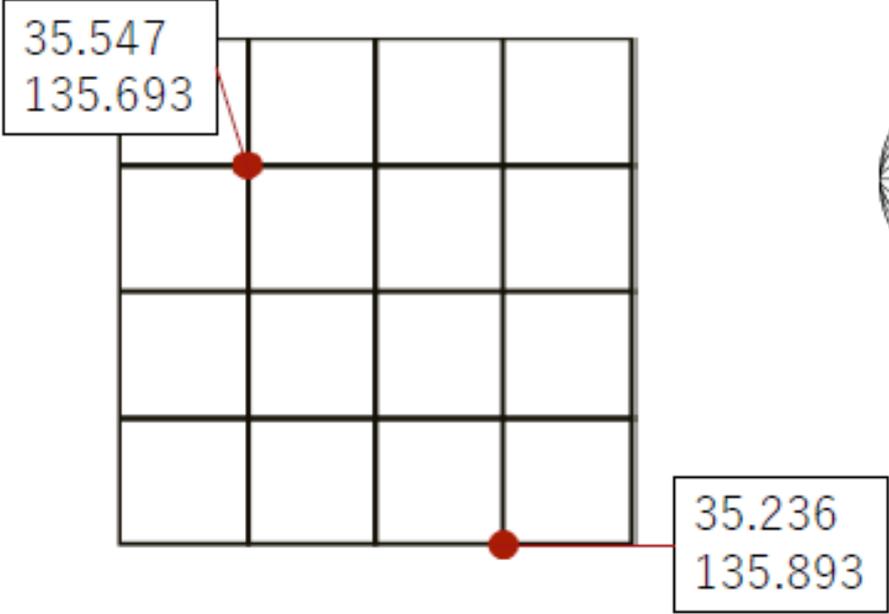
GISの基礎2：「測地系」と「座標系」

■ GISで扱う座標系

座標系：特定の位置を示すために座標を用いる際の座標表現についての取り決め

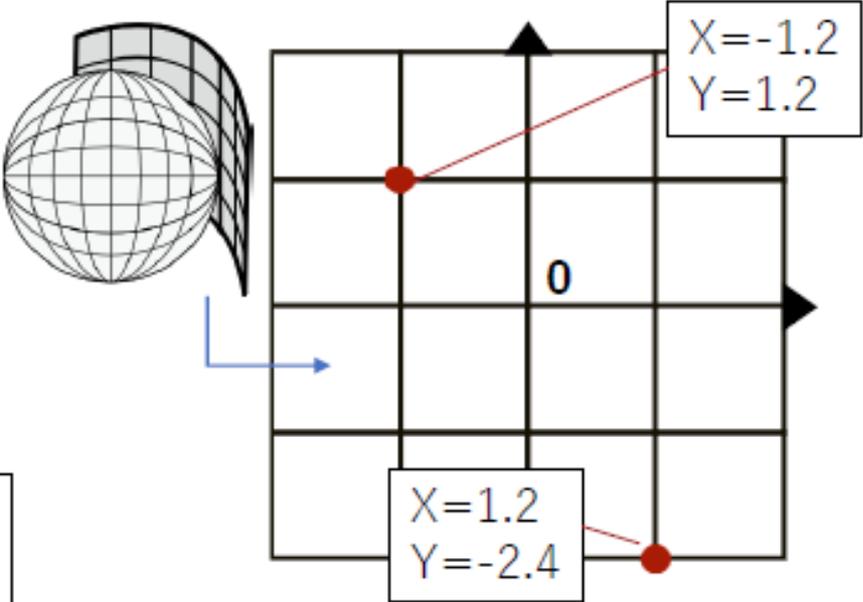
① 地理座標系

緯度経度で表現



② 投影座標系

三次元空間を二次元に“投影”し、XY座標で原点からの距離を表現

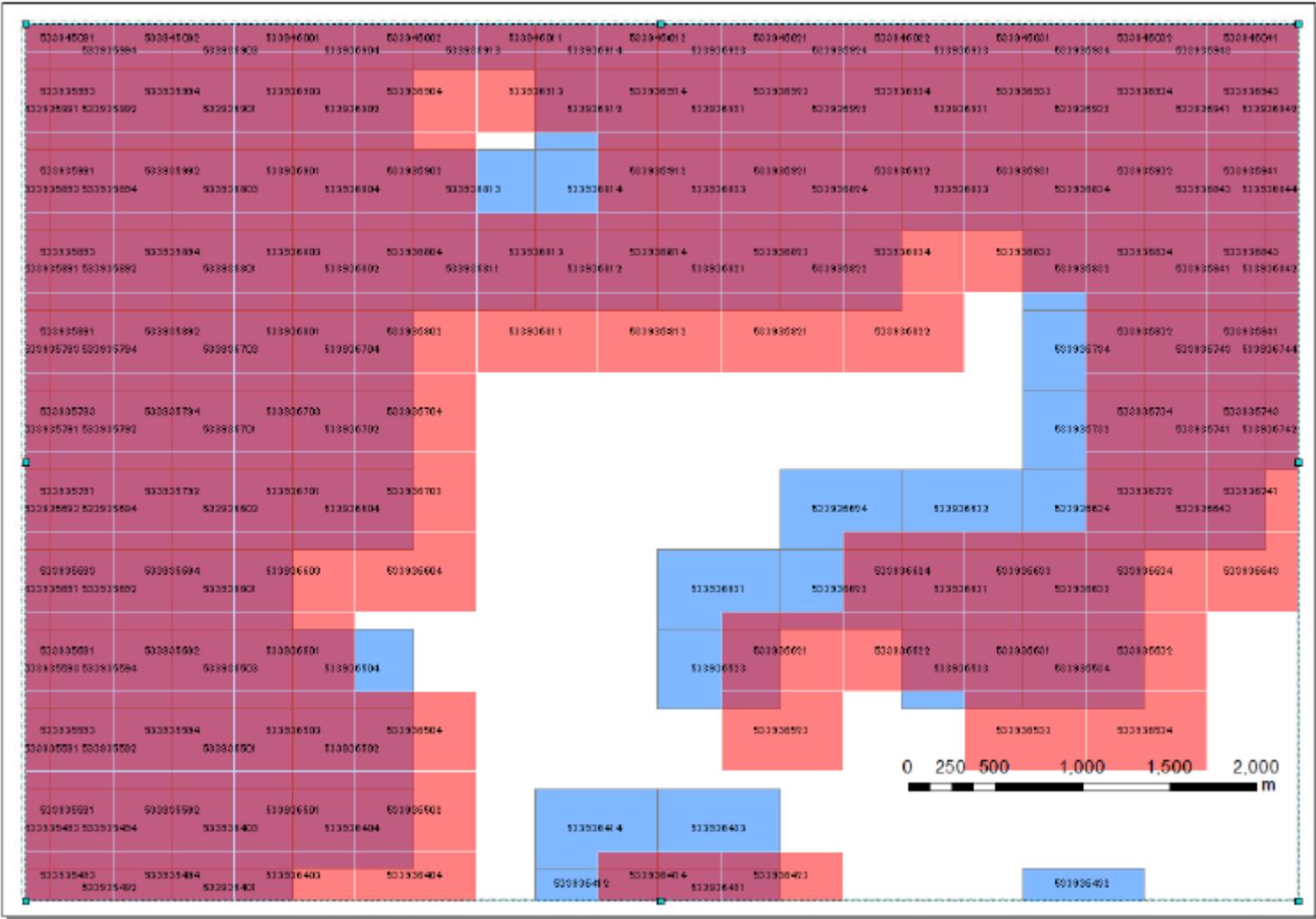


■ 測地系について

測地系：緯度経度の座標軸を使って、地図画面上の特定の位置を示す際の基準となる前提条件

- これが異なると、同じ位置でも座標が異なる
- GIS上で表示するとズレが生じる

GISの基礎2：「測地系」と「座標系」



赤：日本測地系(Tokyo)

青：世界測地系(WGS_1984)

■ 測地系について

測地系：緯度経度の座標軸を使って、地図画面上の特定の位置を示す際の基準となる前提条件

→これが異なると、同じ位置でも座標が異なる

→GIS上で表示するとズレが生じる

「世界測地系」を使おう

世界測地系：JGD2000(2011), WGS1984など

国土数値情報はJGD2000, GPSデータはWGS1984

GISを使うときはデータの測地系に気を配りましょう

ArcもQも「レイヤプロパティ」から確認できます

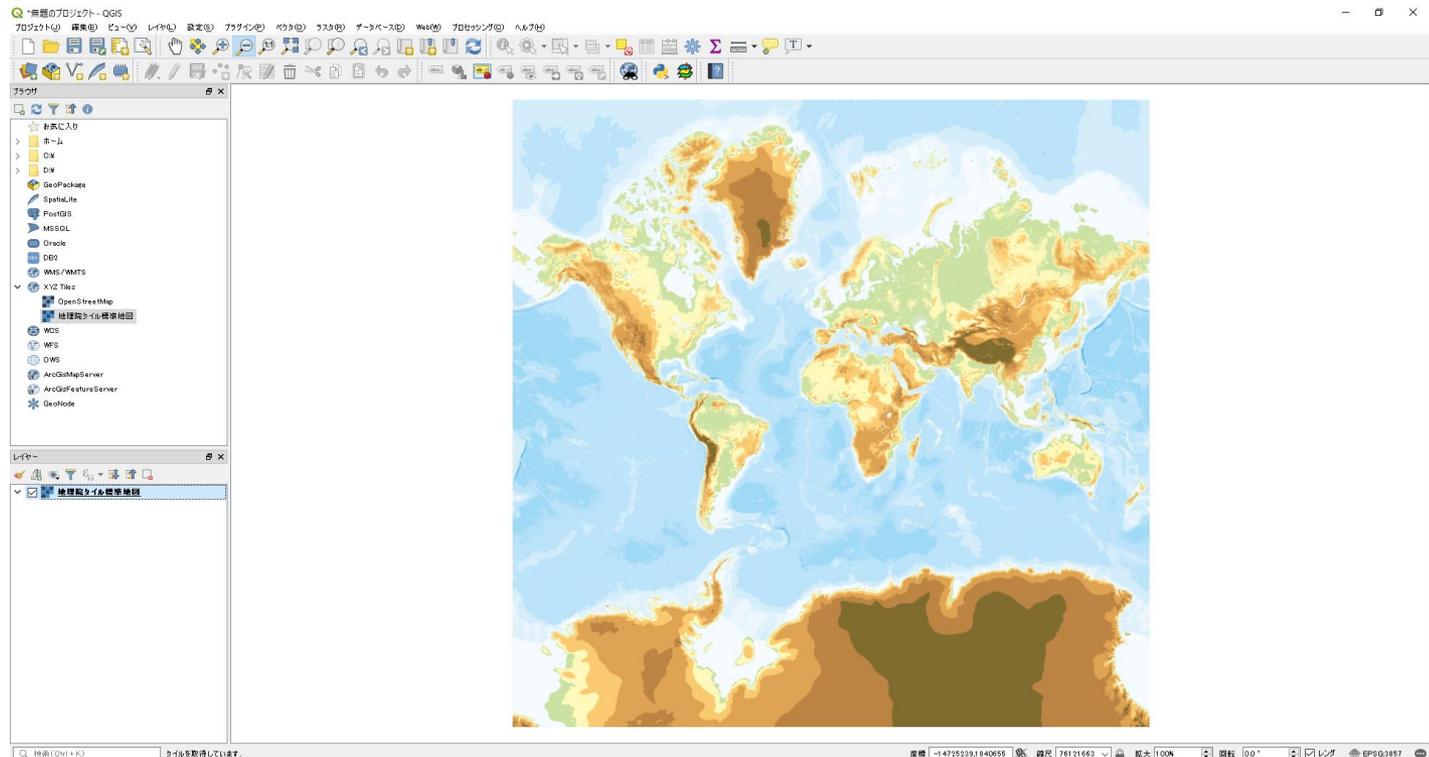
GIS活用例

trip.csv : 各トリップの出発地・到着地の緯度・経度は書いてある
→ どこからどこへ移動する人が多いのか知りたい

QGISを例にやってみる

■ ベースマップの表示

「ブラウザ」→「XYZ Tiles」→好きな地図をダブルクリック

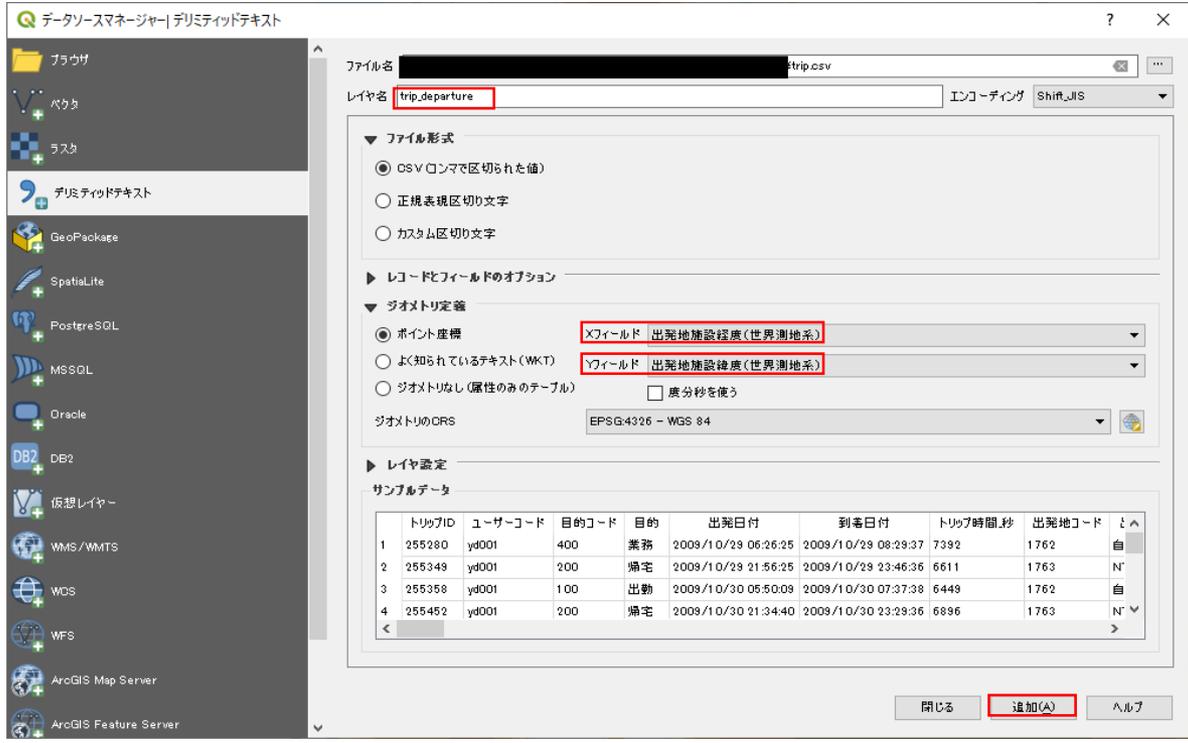


■ tripの起終点の可視化

「レイヤ」→「レイヤの追加」→「ディリミティッドテキストレイヤの追加」

追加したいファイル (trip.csv) のパスをファイル名に入力し、分かりやすいレイヤ名を付けます。

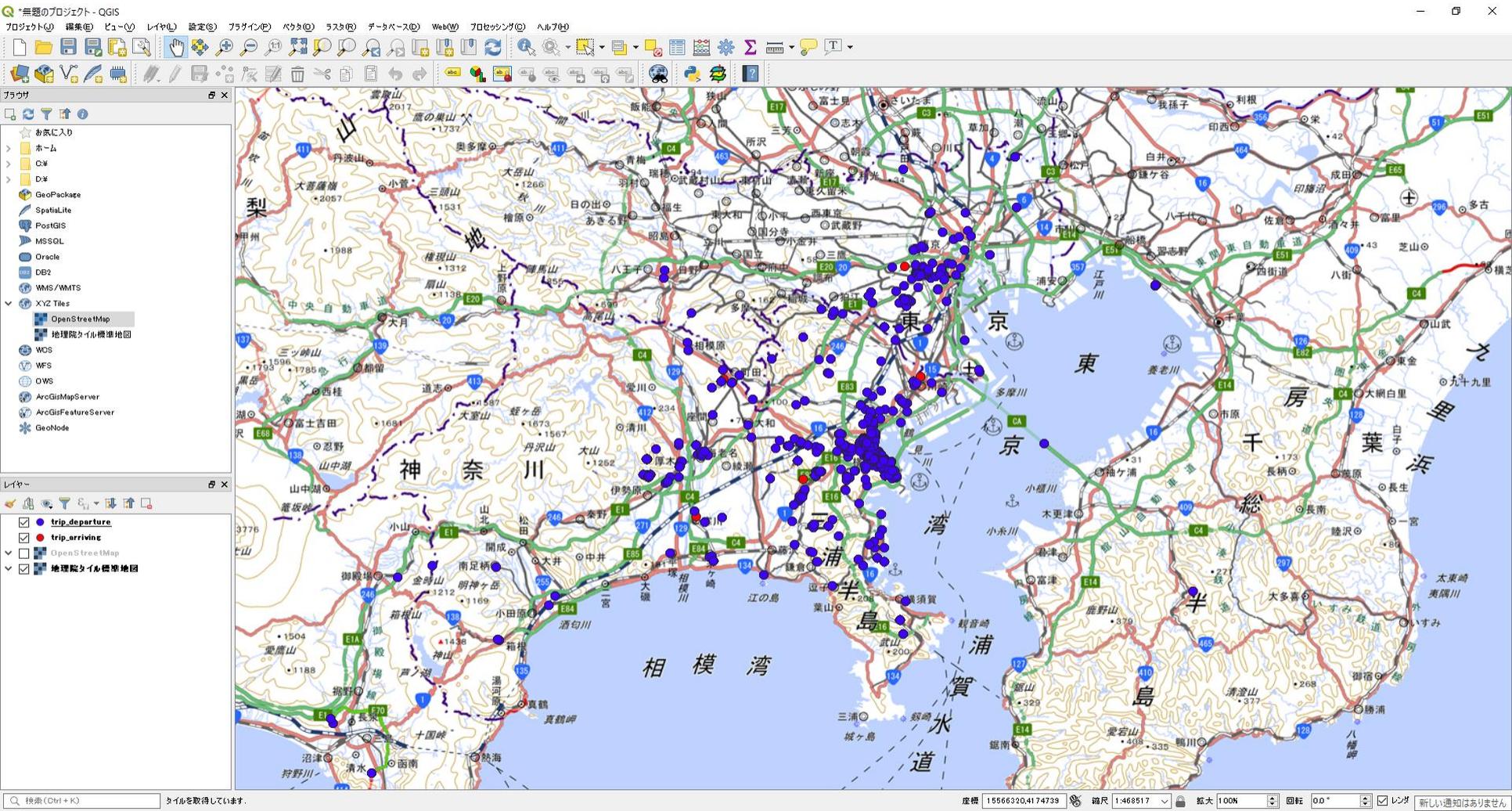
「ジオメトリ定義」の「Xフィールド」に出発地の経度の列を、「Yフィールド」に出発地の緯度の列を指定します。到着地を追加する場合も同様です。



←最後に追加

GIS活用例

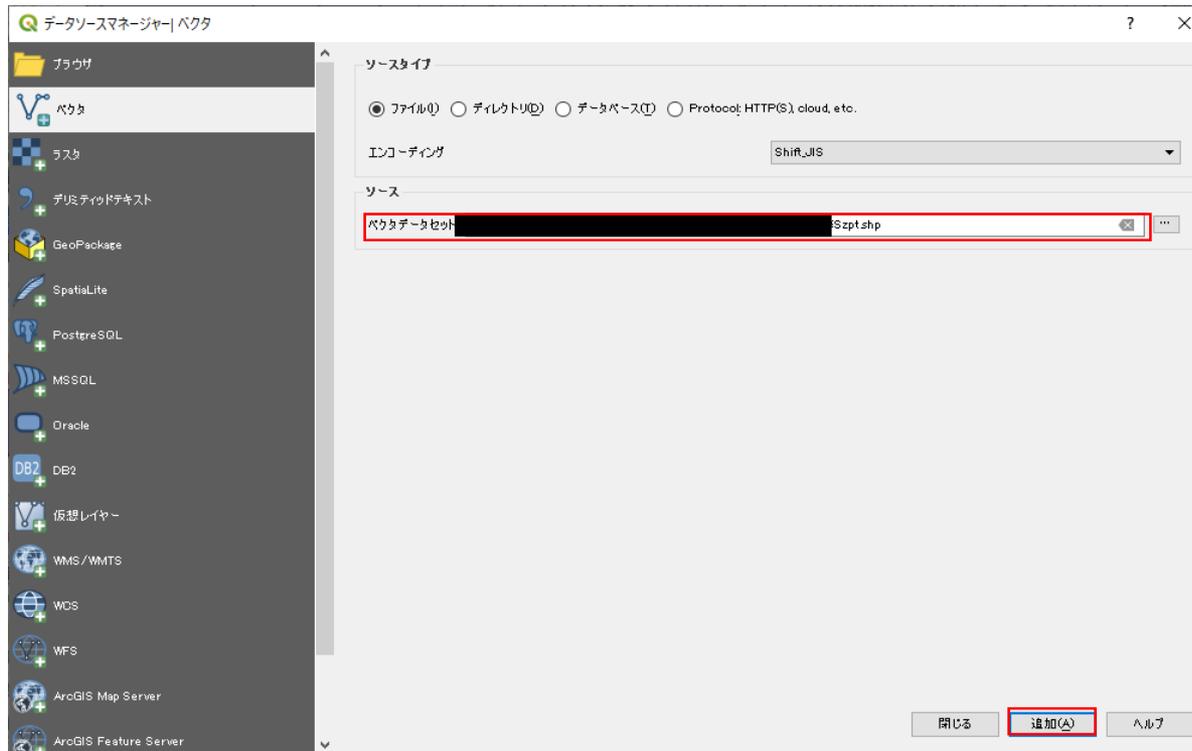
追加されました



■ PT小ゾーンのshpファイルの追加

「レイヤ」 → 「レイヤの追加」 → 「ベクタレイヤの追加」

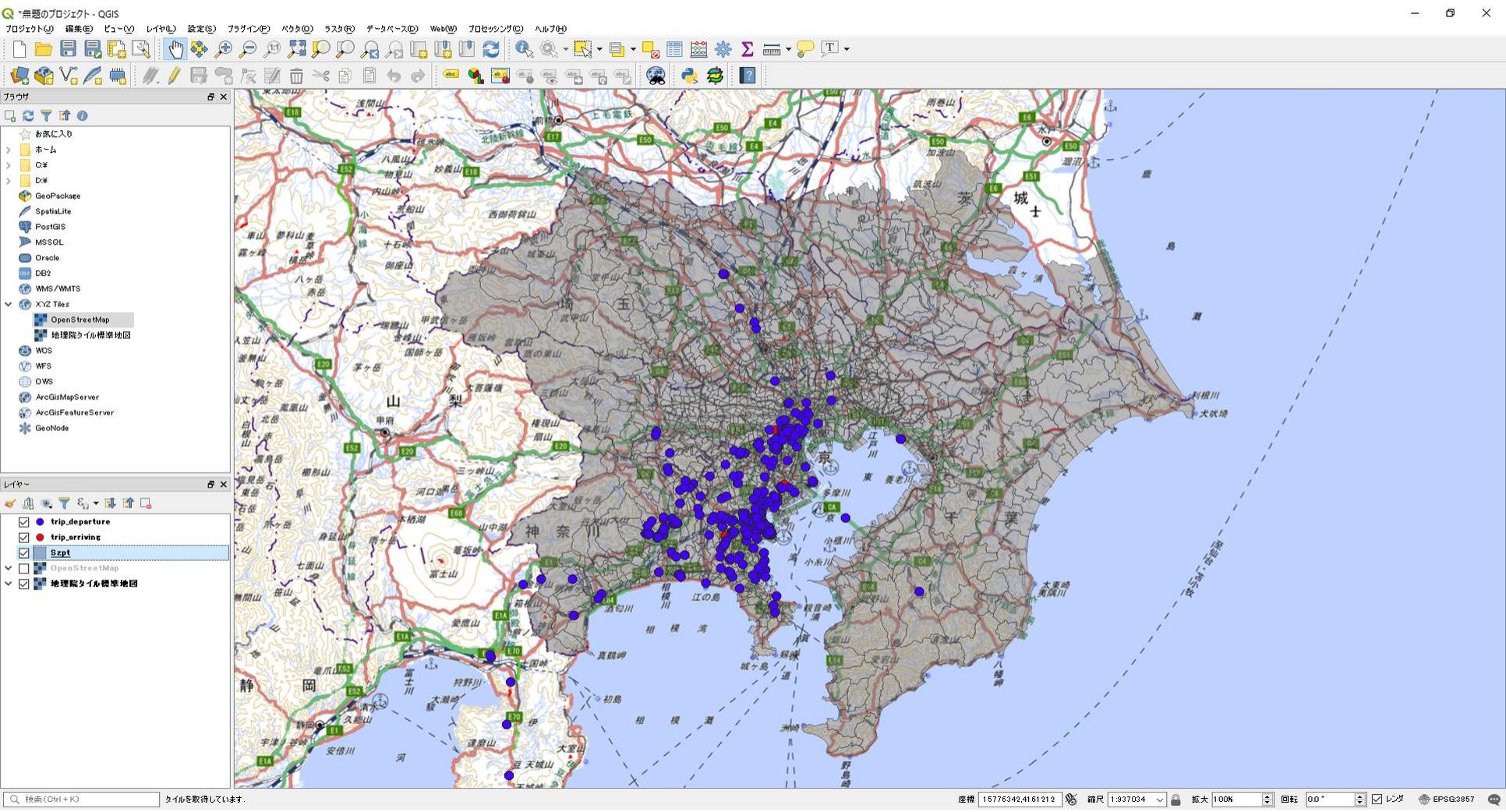
「ソース」 に小ゾーンのshpファイルのパスを指定します。



←最後に追加

GIS活用例

追加されました

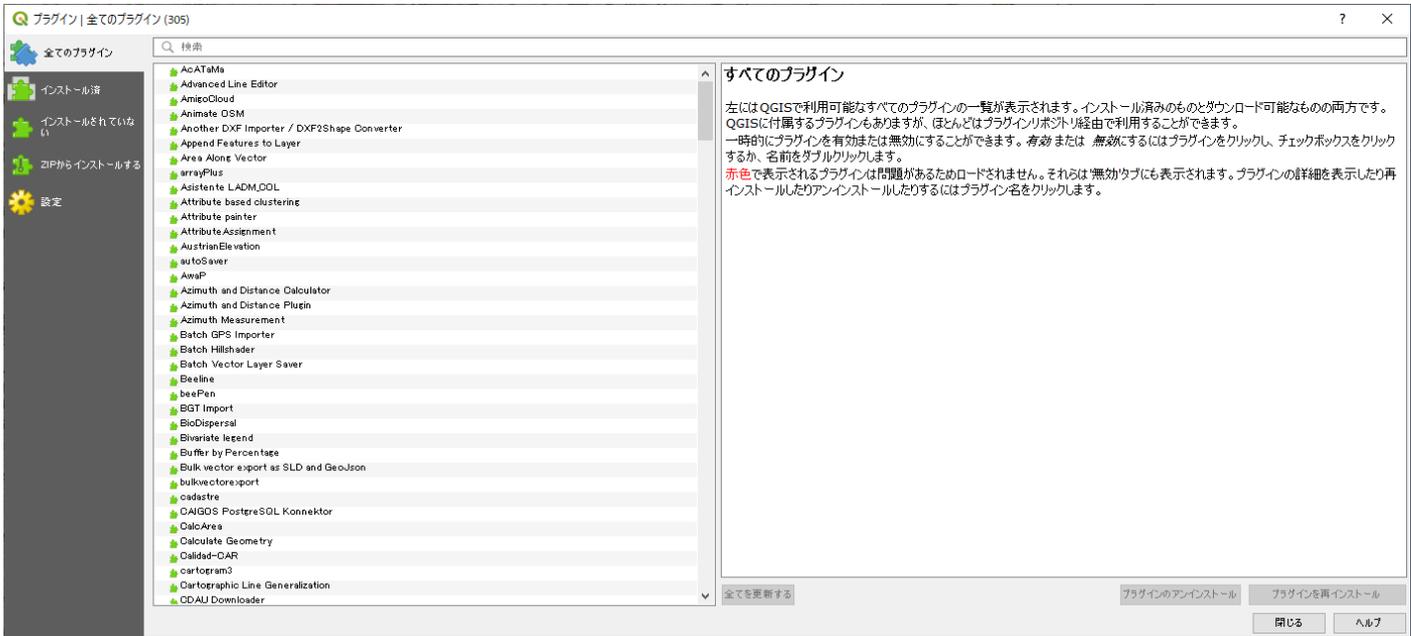


GIS活用例

「trip.csvの各起終点がどの小ゾーンに属しているかという情報をtrip.csvに付け足す」ということを行います。

「プラグイン」→「プラグインの管理とインストール」で「Point Sampling Tool」というものをインストールします。

インストールしたら「プラグイン」→「Analysis」→「Point Sampling Tool」で立ち上げます

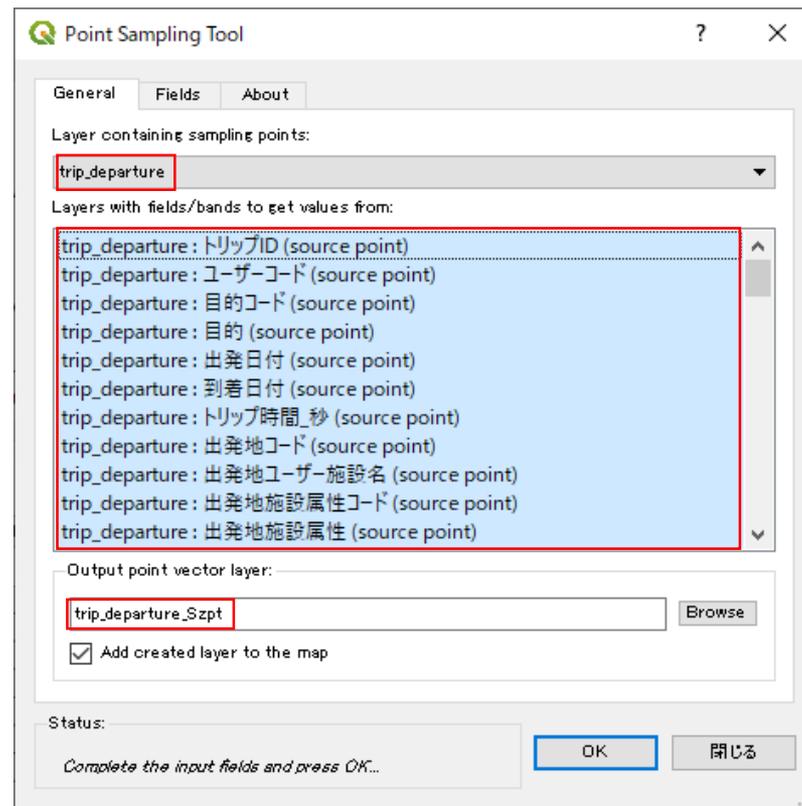


このように豊富なプラグインを用いて色々な分析ができます
もちろんArcでも同じことができます

GIS活用例

Point Sampling Tool ウィンドウが立ち上がったたら、「Layer containing sampling points」で使用する点データを、「Layers with fields/bands to get values from」で値を取得する属性をクリックします。ここでは、ポイントの情報と小ゾーンの番号等の属性情報の両方を選択するようにします。

あとは出力するファイル名を入力しOKをクリック。



GIS活用例

新しくできたポイントのshpファイルには小ゾーンの属性情報が追加されています。
ポイントの属性情報を見るにはレイヤ名の上で右クリックして「属性テーブルを開く」で確認
できます。

	KEY_CODE	AREA_NAME	KEY_ID	Kzone
1				
2	1211	506.6328522693...	12211	1221
3	1211	506.6328522693...	12211	1221
4	12110	118.6001372082...	12110	1211
5	1211	506.6328522693...	12211	1221
6	12110	118.6001372082...	12110	1211
7	1211	506.6328522693...	12211	1221
8	1211	506.6328522693...	12211	1221
9	1211	506.6328522693...	12211	1221
10	12110	118.6001372082...	12110	1211
11	12110	118.6001372082...	12110	1211
12	1211	506.6328522693...	12211	1221
13	12100	121.5447776116...	12100	1210
14	12110	118.6001372082...	12110	1211
15	1211	506.6328522693...	12211	1221
16	12110	118.6001372082...	12110	1211

末尾に小ゾーンの属性情報が追加されている

同じく右クリックして、「エクスポート」から「Save Features As...」で新しい属性情報を持ったshpファイルをcsvなどのファイル形式で保存することもできます。

locDataやtripDataに色々な情報を追加して基礎集計するとまたいろいろなことが見えてきて楽しいと思います。

■ PPデータを用いた行動の特徴の把握

- 松山（2007），横浜（2009）のPPデータ，PT調査の小ゾーンshpファイルを渡します。（これ以外のデータを使ってもいいです）
- 近年，都市は様々な変化・課題とともにあります。
 - 人口減少，高齢化
 - 都心と郊外
 - 東京と地方
 - リニア開通
 - オリンピックetc...
- PPデータを用いて基礎集計，空間分析を行い，松山・横浜における人々の行動の特徴，今後起こると考えられる課題などをまとめてください。
- はじめに述べた通り，基礎集計は行動モデル構築・推定の仮説を発見するためのものでもあるので，行動とその要因となっているものの因果関係を意識して集計の視点を考えてみてください。
- スライド枚数は自由