



特定事例に基づくAPI機能を使ったデータの使い方の紹介

Otani&Co., Inc.
Yuichiro Otani
yuichiro@otani.co

自己紹介

小谷 祐一郎 (Yuichiro Otani)

株式会社おたに (Otani&Co., Inc.)

- データ分析 (日経ビッグデータで記事も...)
- Rパッケージ (govStatJPN)
- 宅地建物取引士 (売買等も...)
- GEEEO (不動産価格の予測サービス)

本日の内容

GEEOとは

コアアセット


オープンデータとe-Stat

GEEOにおけるe-Statの使い方

e-Statを更に使いこなす

GEEOとは

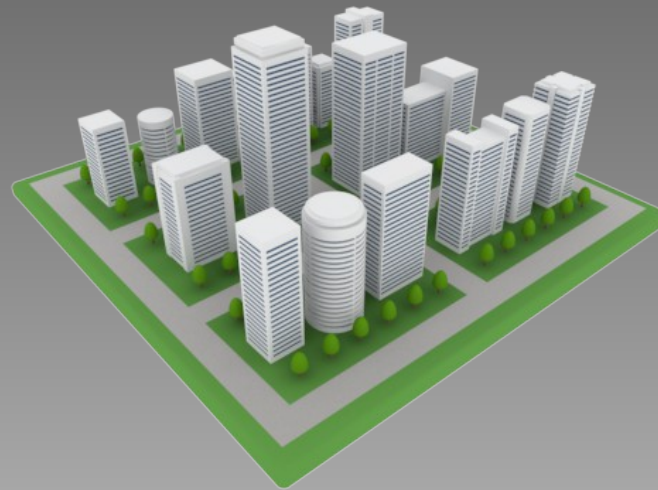


 GEEO.OTANI.C

ブラックジャックによろしく 佐藤秀峰 漫画 on web URL: <http://mangaonweb>

デモ

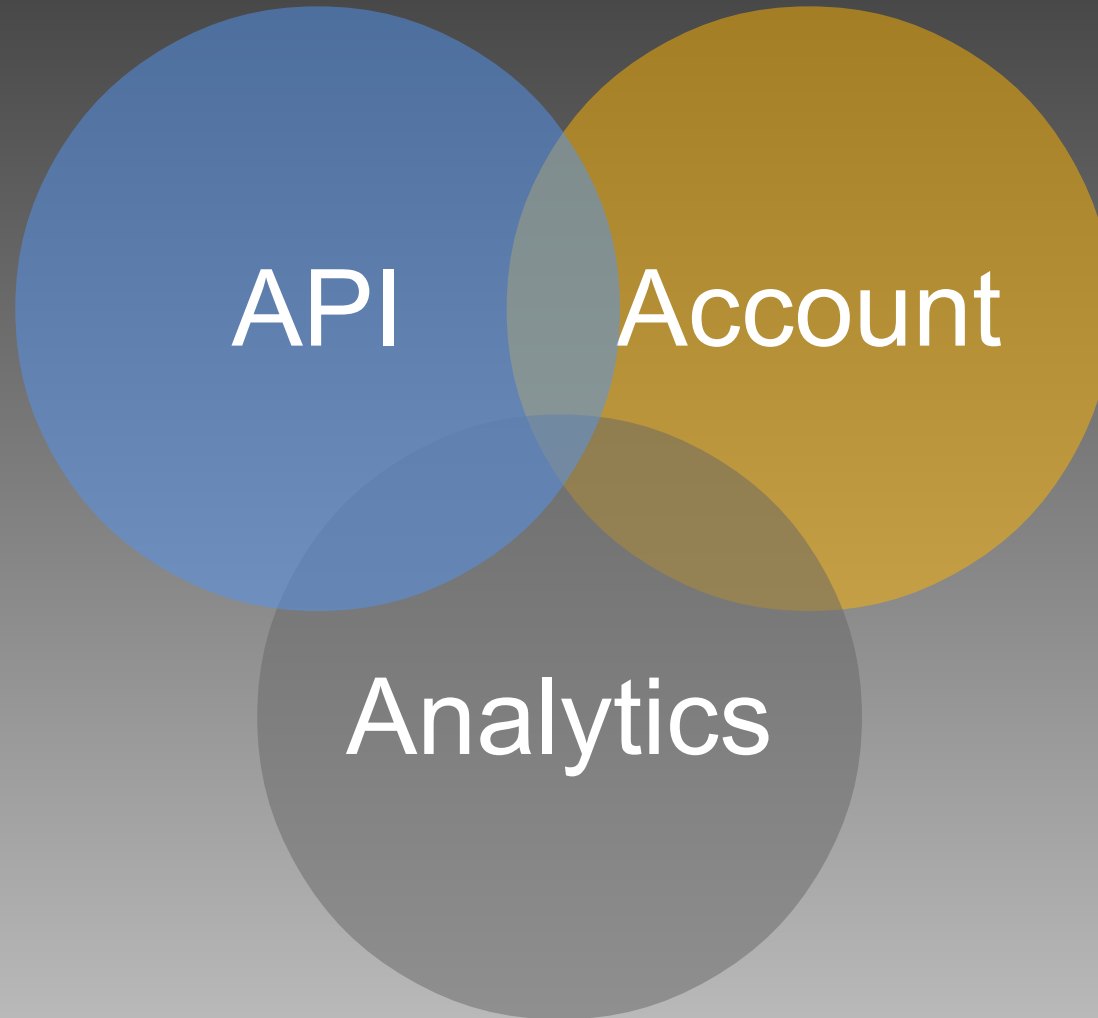
¥25,800,000 ↗



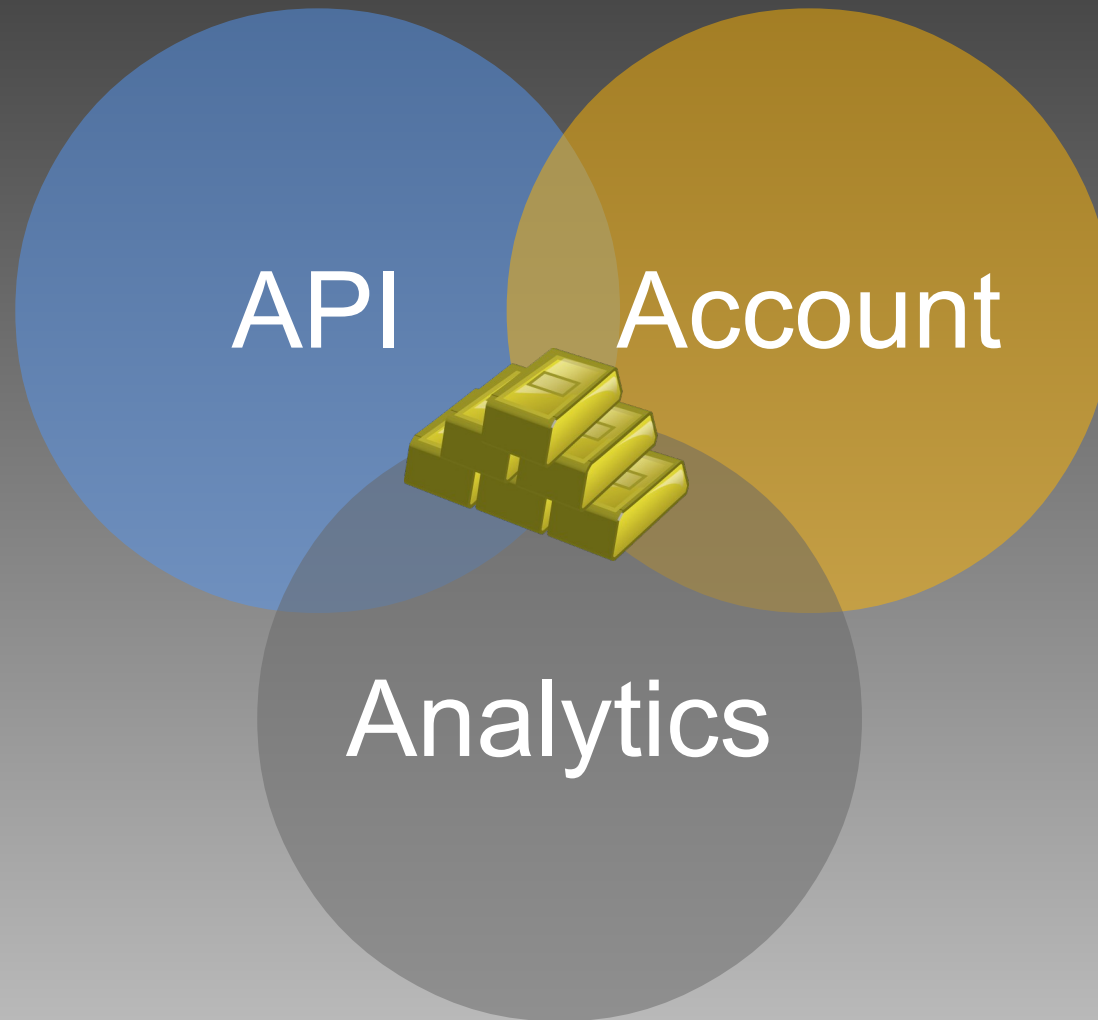
GEEOのこれまで



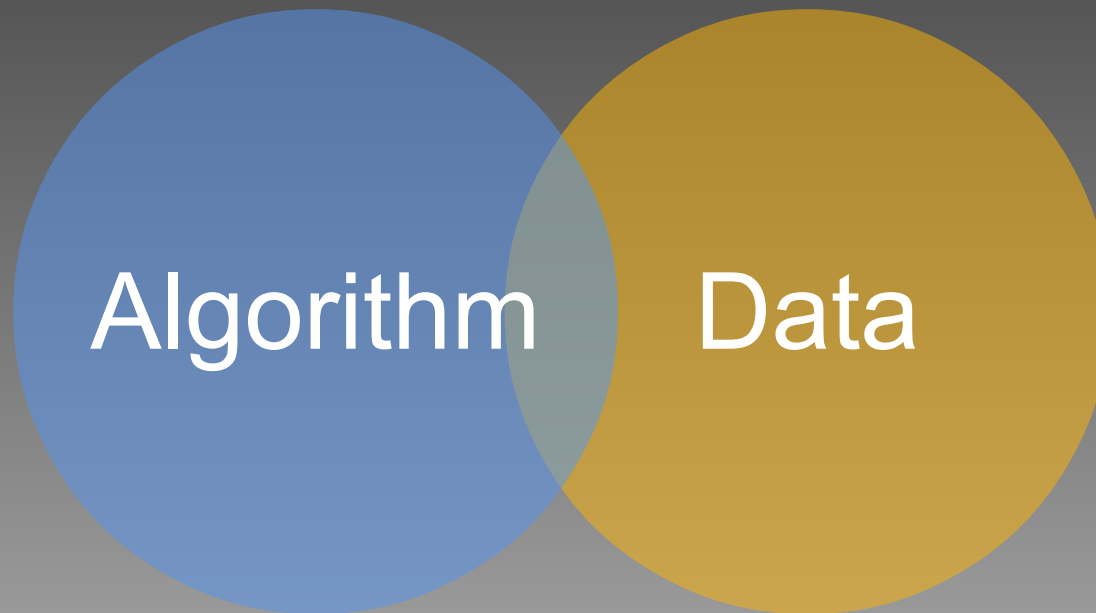
Businessについて



コアアセット



コアアセット



Statistical Algorithm

- 最高97%の精度(GEEO)

*自由度調整済みR二乗

$$\prod f(x, \theta)$$

- データに非依存
- 計算に国境なし
- 応用範囲が広い

応用範囲の例

- 発生の予測
 - 事件・事故の発生箇所の予測
 - 空家発生の予測
- 被害の予測
 - カラスによるゴミ被害等の予測
 - ネズミ・ハチ等の被害予測
- マーケットの予測
 - 飲食店等の出店による売上等の予測
 - 住宅単位での潜在顧客の指標化

Aggregated Data

- オープンデータ
 - 各省庁のデータ
 - 自治体のデータ
- ×
- 独自データ
 - 調査データ
 - 作成データ
 - オープンデータから作成

オープンデータとe-Stat



政府統計への期待と失望

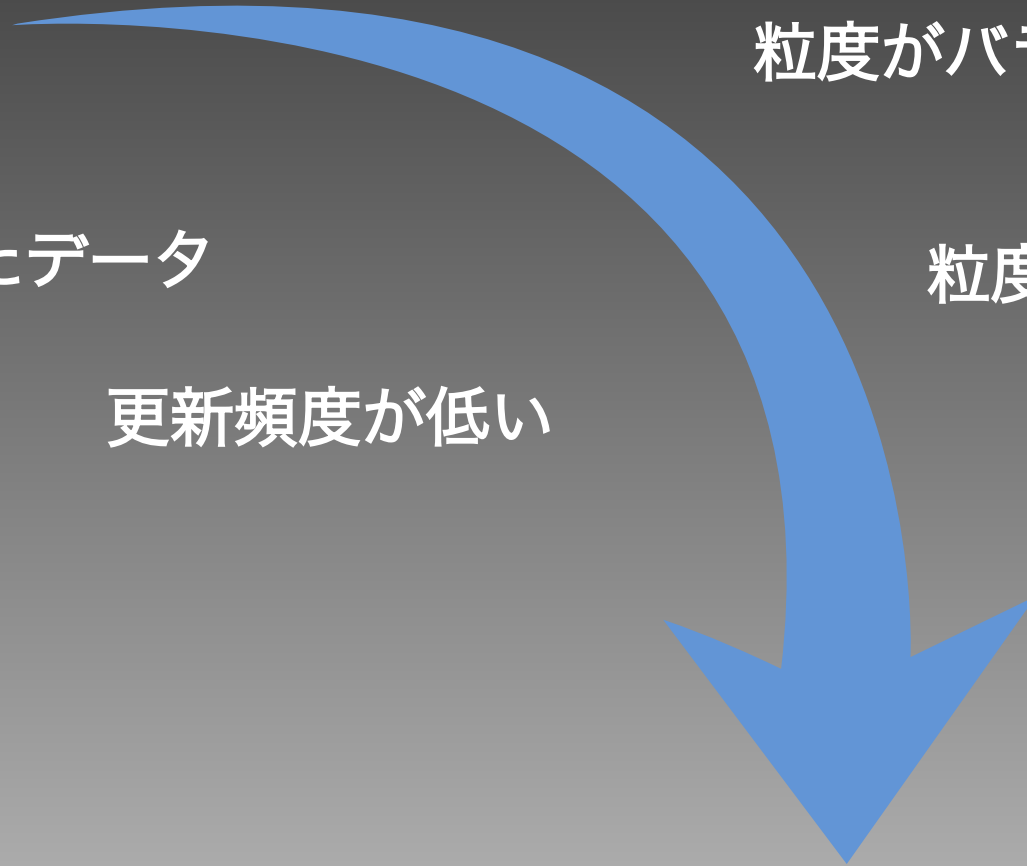
誰でも手に入る

粒度がバラバラ

集約されたデータ

粒度が荒い

更新頻度が低い



失望を希望へ変えるには

独自価値

値決めができる

使い方を決められる

企業努力

価値創造の基盤

原価0円

制約が緩い

オープンデータを使うために...



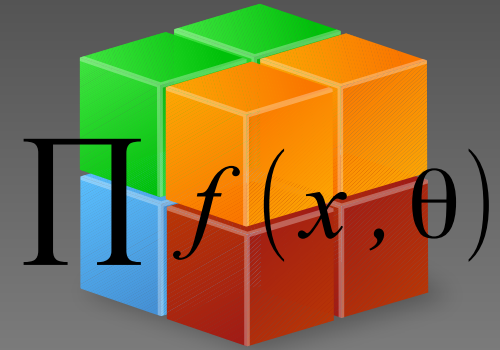
オープンデータを使ったビジネスのタイプ



Add-on



Aggregation

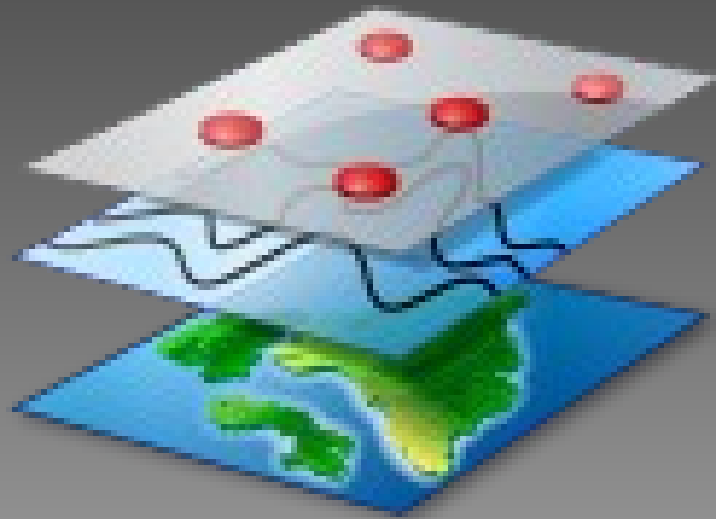


Algorithm

GEEOにおけるe-Statの使い方

- 境界表示
 - Add-on
- 付随情報
 - Add-on
- 分析用エンジン
 - Aggregation
 - Algorithm

境界データの表示(Add-On)



Downloadable

- 1948 files(GML)

JSON

- from GML to JSON

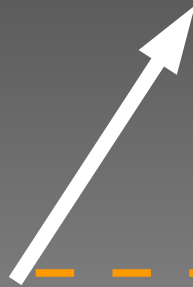
Cache

- Web Browser

境界データの表示(Add-On)

User

- 最初のリクエストでキャッシュ
- キャッシュデータを使用



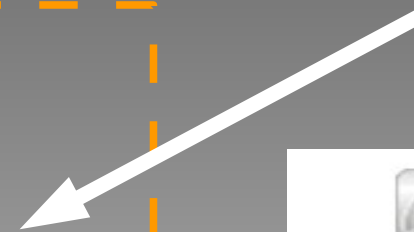
Web (JSON) サーバー

- レスポンス
- キャッシュ



計算サーバー

- ダウンロード
- GMLからJSON



付随情報(Add-On)



市川市鬼越1-2周辺

この地域に住む主世帯720世帯の内訳は、一戸建に351世帯、長屋建に21世帯、共同住宅に347世帯、共同住宅1・2階建に220世帯、共同住宅3～5階建に127世帯、共同住宅6～10階建に-世帯、共同住宅11階建以上に-世帯、その他に1世帯、となっています。最大は一戸建で全体の48.8%が居住しています。つまり、戸建を中心とした地域です。

市川市北方1丁目には747世帯1556人が居住しています。男性は769人、女性は787人です。1世帯あたり2.1人が住んでいます。夫婦2人のみ又は夫婦と小さい子供1人が居住できるやや広めの2LDKや3LDK等が適切な間取りと考えられます。但し、下記の人口情報も考慮してください。

総数、年齢「不詳」含む	1556人
総数0～4歳	37人
総数5～9歳	59人
総数10～14歳	48人
総数15～19歳	55人
総数20～24歳	60人

Request to API

- Each Request

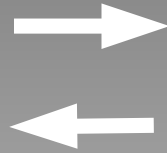
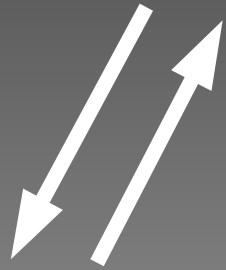
JSON to JSON

- API JSON to Town JSON

Cached Codes

- 市区町村コード

付随情報(Add-On)



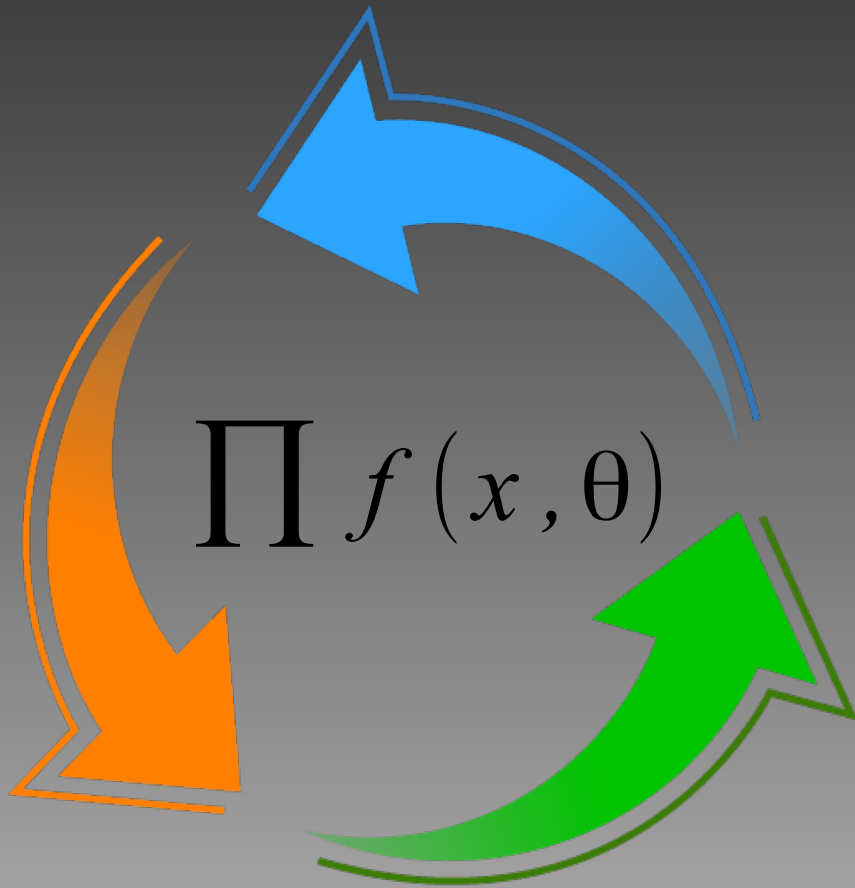
Request

- 逆ジオコーディング(丁目単位)
- 住所から市区町村コードを指定
- 統計IDと市区町村コードを指定してAPIにリクエスト

Response

- API JSONからデータを抽出
- 各データの計算と文章の作成
- JSONで結果を返す

分析用エンジン



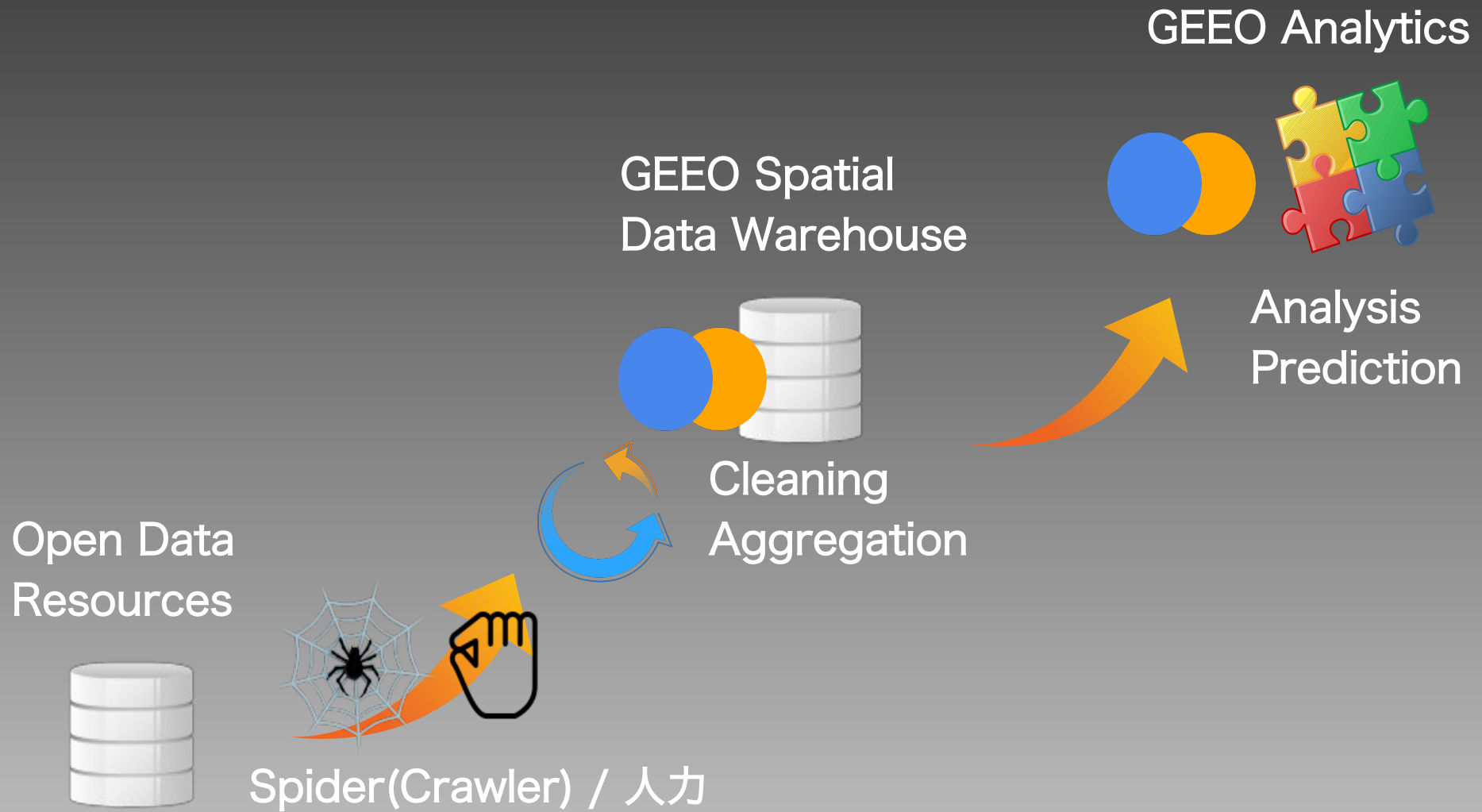
前処理

- 収集
- 統合

分析

- データの探索

全体構成



前処理(1) データの取得

Open Data
Resources



人力 (プログラムの実行)



GEEO Spatial
Data Warehouse



前処理(2) クレンジングとアグリゲーション

GEEO Spatial
Data Warehouse



A ↓ 三 ↓ 1 ↓
Z ↓ 三 ↓ 2 ↓

例：次の住所は全て統一

- 横浜市鶴見区鶴見中央5-7-2
- 神奈川県横浜市鶴見区鶴見中央5の7の2
- 鶴見区鶴見中央五丁目七番地2

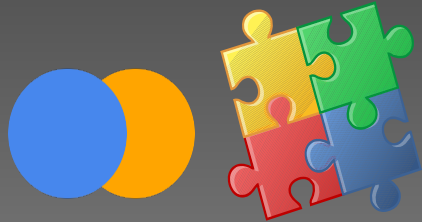


e-Stat内のデータで異なる事も...

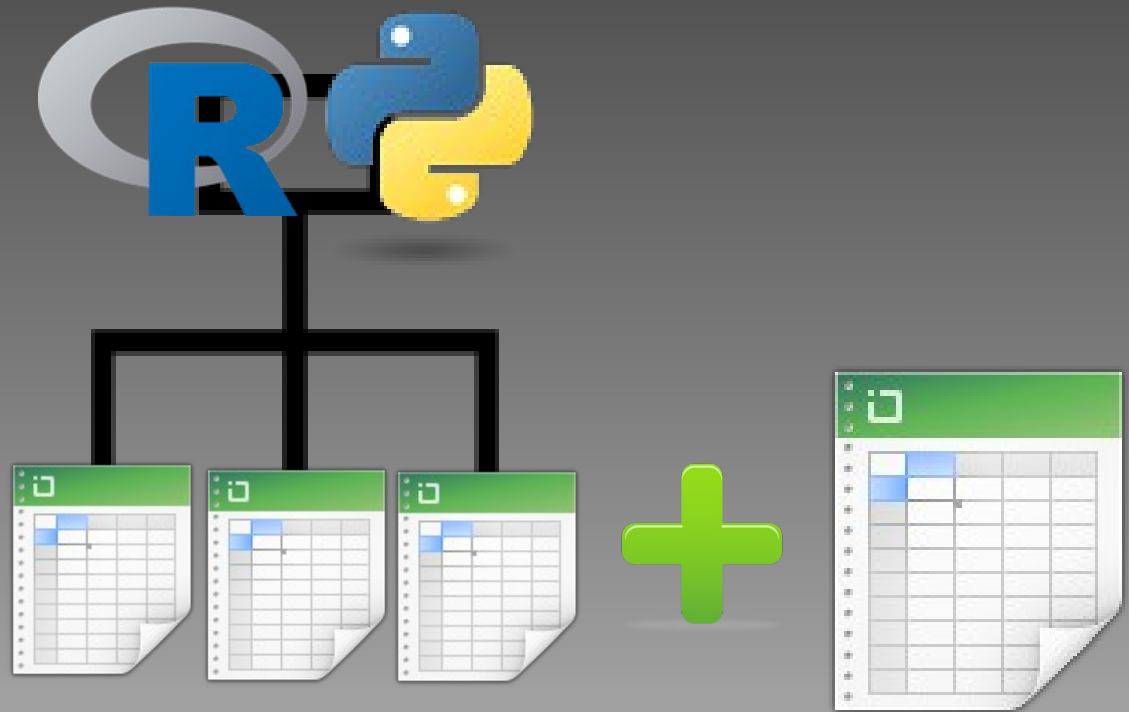


分析 データの探索

GEEO Analytics



Analysis
Prediction



e-StatをDatabaseとして使うために

- 正規化用プログラム
 - 表示用はJSON
 - 分析用はCSV
- 予め用意しておくもの
 - 市区町村コード
 - 座標と住所(ジオコーディング用)
 - 目的に応じたデータの形式
- キャッシュ化

e-Statを更に使いこなす

住宅単位での潜在顧客の指標化

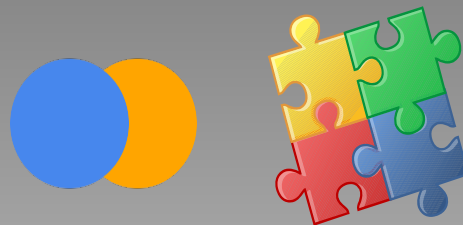
GEEO Spatial
Data Warehouse



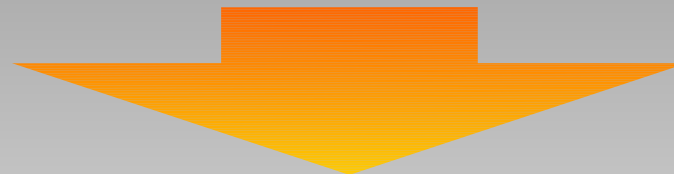
Open Data
Resources



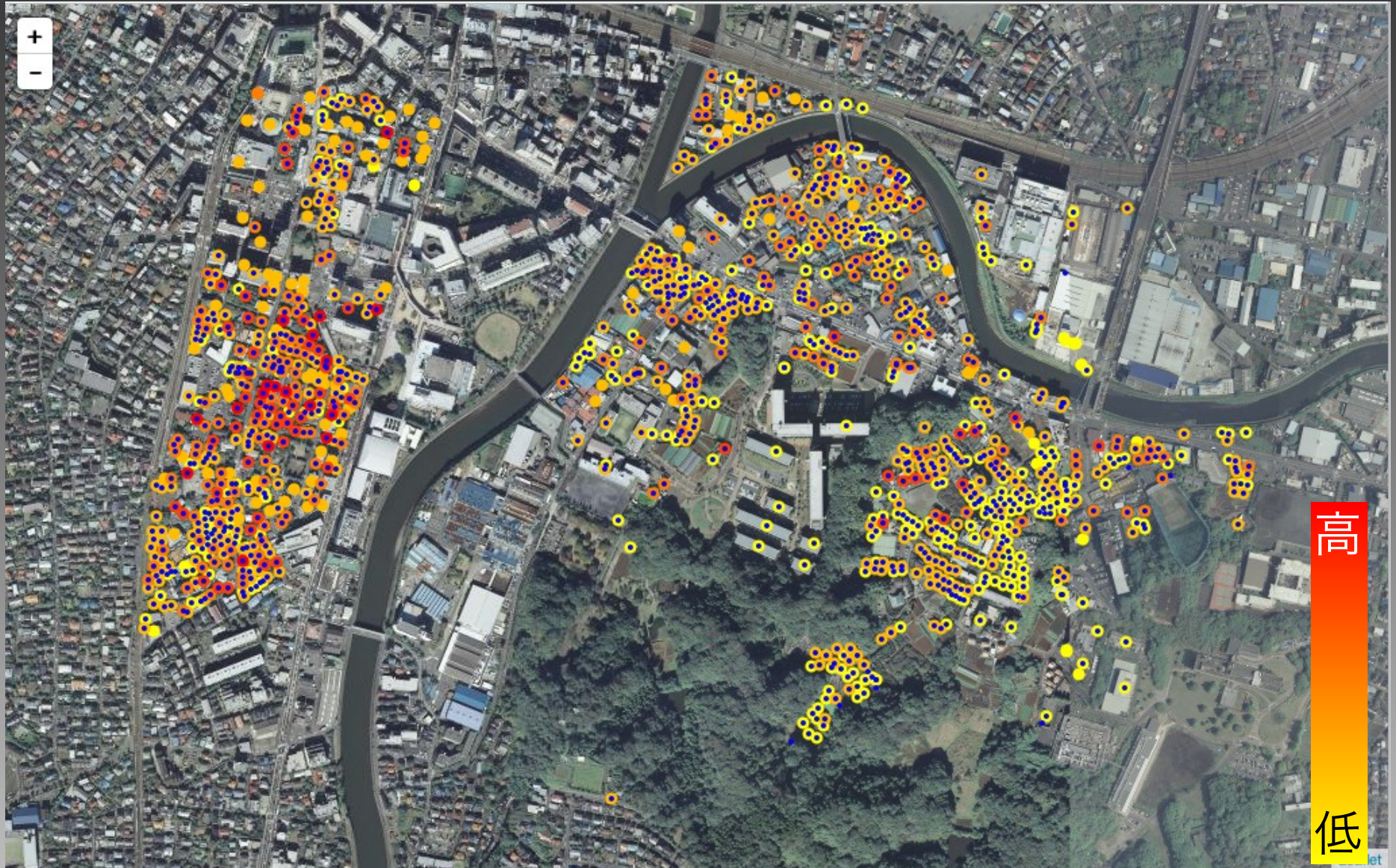
GEEO Analytics



Analysis & Prediction

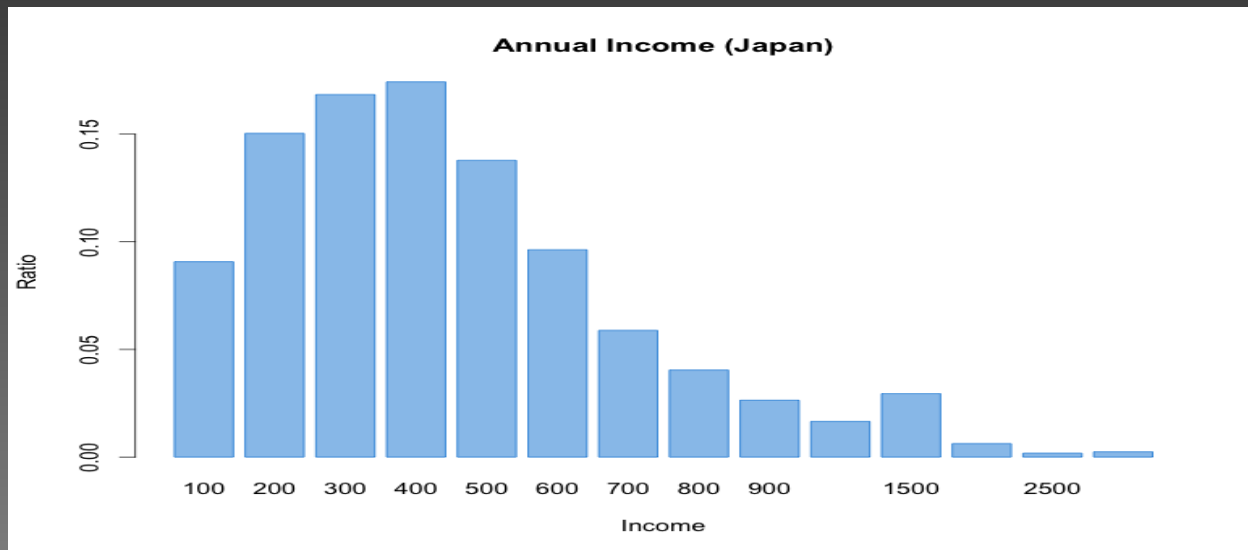


世帯年収の推計

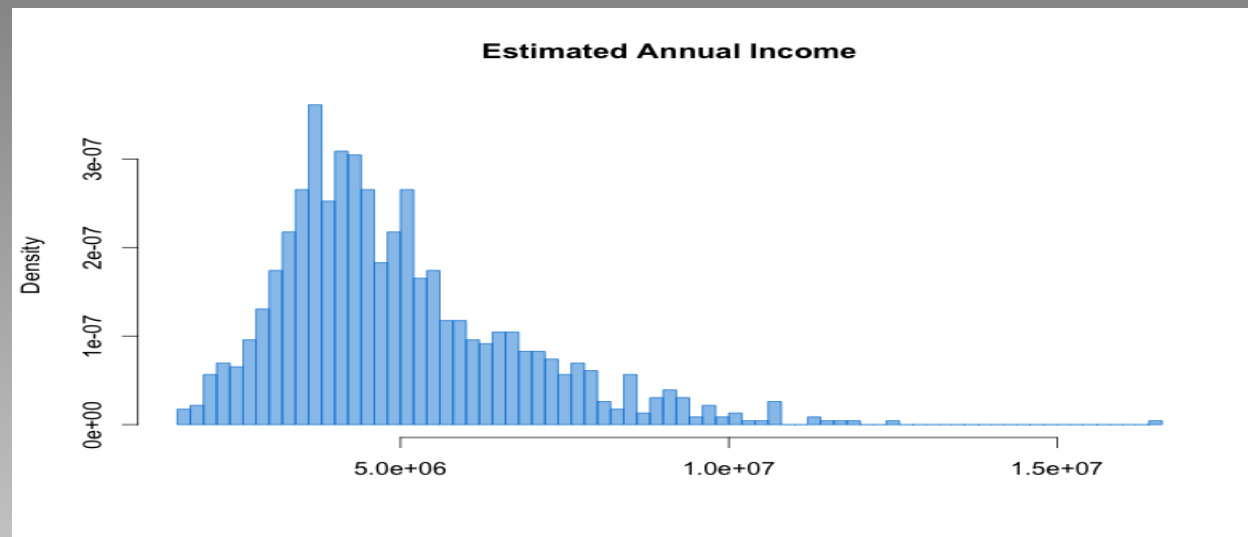


※ これはGEEOの不動産価格の予測値と家計調査等から予測した値です。

妥当性の検証



国税庁 民間給与実態調査
平成26年より作成



推定した世帯年収の
ヒストグラム

まとめ

- オープンデータから生み出せる価値
- GEEQOにおけるe-Statの使い方
- e-Statを更に使いこなす

連絡先 : yuichiro@otani.co (小谷 祐一郎)

最後に...

- 次回について
 - 12/18(応用編)
 - API機能の紹介
 - 中間アプリ等の紹介
 - ダウンスケールの方法等の紹介
 - Ustreamで配信予定

ご静聴ありがとうございました。