

2020年12月11日
@相模女子大学

市民参加により創出されるオープンデータ － その分析と活用

合同会社緑 I T 事務所
小池 隆

自己紹介

- ➡ 早稲田大学第一文学部哲学科心理学専修卒
- ➡ 準大手 S I 企業にてソフトウェア受託開発とソフトウェア工学の研究に従事
- ➡ 2014年に合同会社緑 I T 事務所設立
 - ➡ 「オープンソース&オープンデータによる、健全で豊かな I T 文化の創出」
 - ➡ 地理情報システム、データ分析・可視化、地域情報化
 - ➡ 最近では AI（機械学習）にも取り組む

オープンデータとは

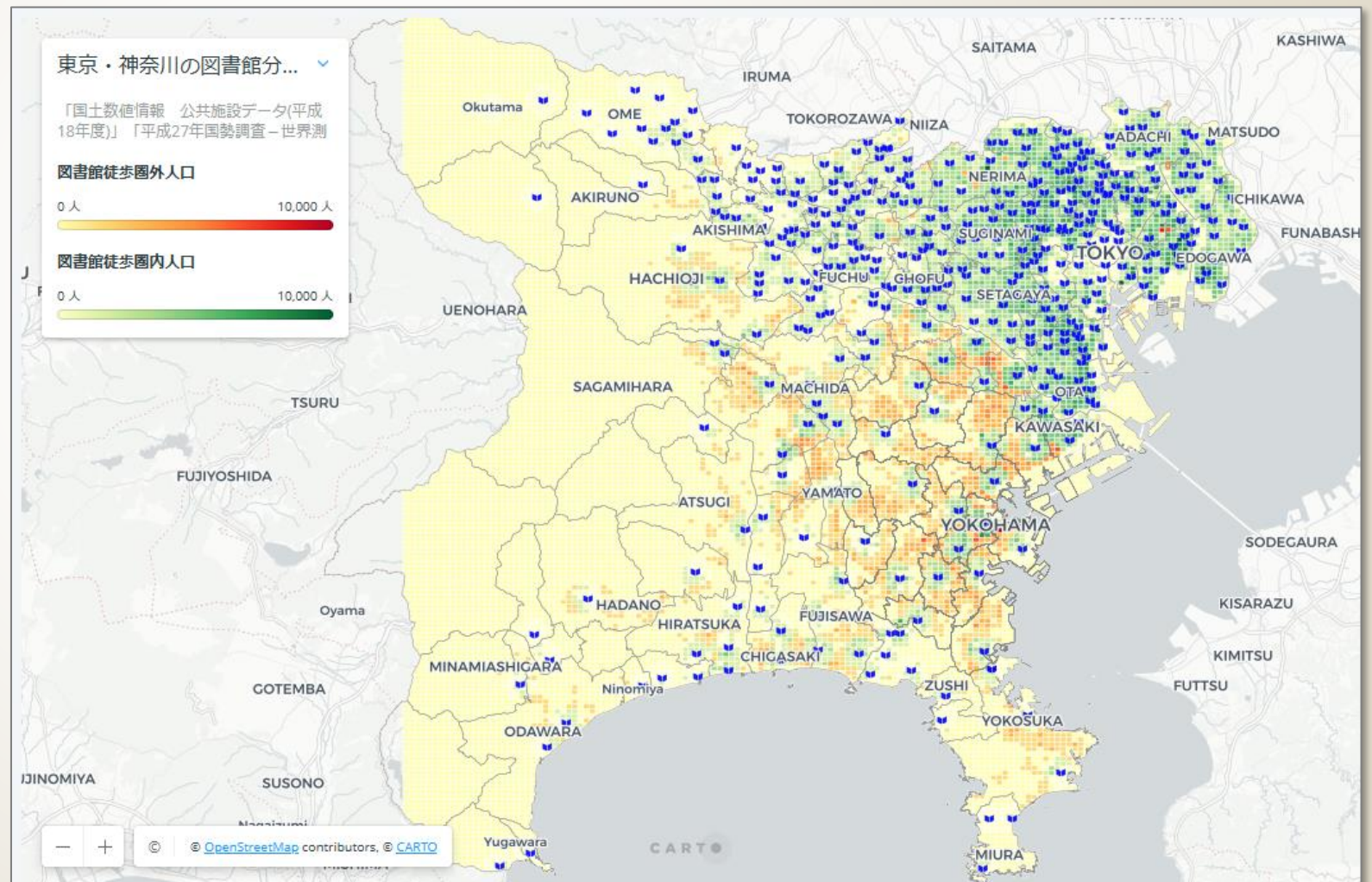
- ▶ 国、地方公共団体及び事業者が公開するデータ
 - ▶ 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
 - ▶ 機械判読に適したもの
 - ▶ 無償で利用できるもの
- ▶ オープンデータの意義・目的
 - ▶ 国民参加・官民協働の推進を通じた諸課題の解決、経済活性化
 - ▶ 行政の高度化・効率化
 - ▶ 透明性・信頼の向上

オープンデータのライセンス

- ▶ クリエイティブ・コモンズ（CC）ライセンスが主流
 - ▶ CC0
 - ▶ パブリックドメイン（全ての権利を放棄）
 - ▶ CC BY
 - ▶ 出典を記載すれば再配布可能
 - ▶ 改変可能、商用利用可能
 - ▶ その他（CC BY-xx）
 - ▶ ND（改変禁止）
 - ▶ NC（商用利用禁止）
 - ▶ SA（派生物を同一ライセンスで配布）
- } オープンデータの定義から外れる

オープンデータの分析例

東京・神奈川の図書館分布と人口

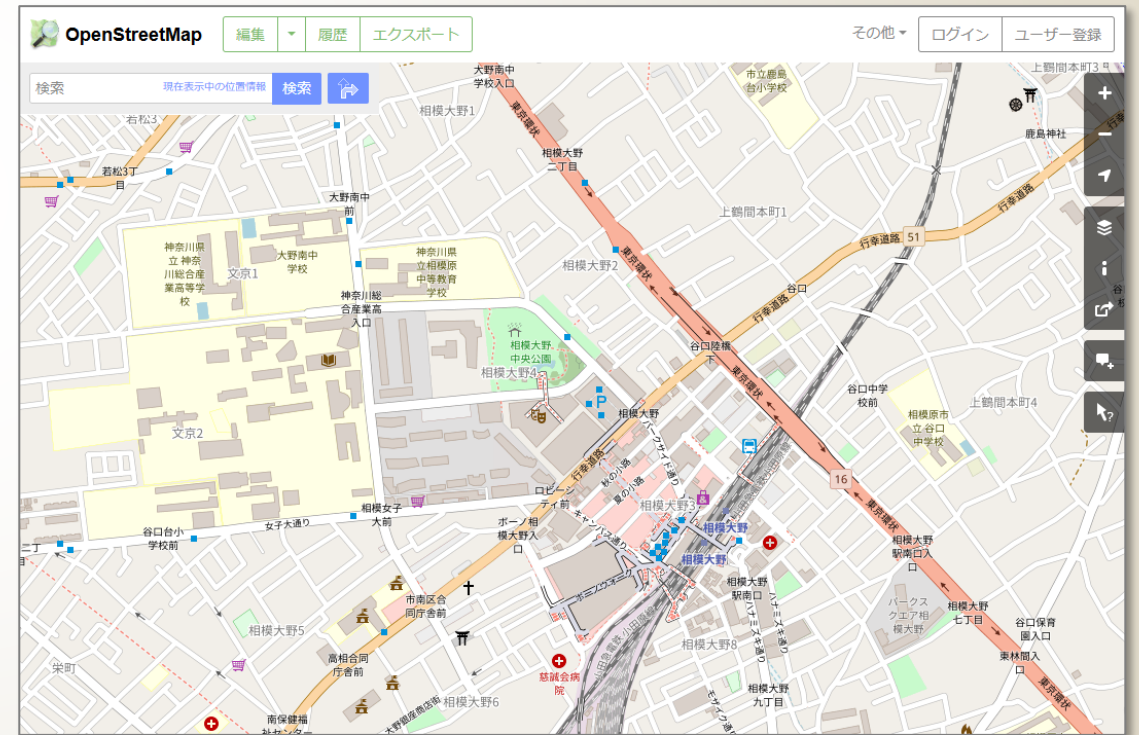


データ出典

- 国土数値情報
公共施設データ
- 平成27年国勢調査
500mメッシュ人口

市民参加により創られるオープンデータ

- ▶ ウィキペディア、OpenStreetMapなど
- ▶ 誰でも編集できるフリーな百科事典／地図



テキストはCC BY-SA
写真等はCC0、CC BY、CC BY-SA等

地図データはODbLライセンス
地図画像はCC BY-SA

オフラインでの活動

- ➡ ウィキペディアタウン
 - ➡ ウィキペディア記事執筆
- ➡ マッピングパーティー
 - ➡ OpenStreetMap地図編集

ウィキペディア街道プロジェクト



Build a Jekyll blog in minutes, without touching the command line.

About

ウィキペディア街道は、旧街道沿いの名所旧跡などの歴史・文化情報をインターネット上の百科事典「ウィキペディア」に掲載して世界に発信するプロジェクトです。

Code for Setagaya、オープン川崎、Code for Kanagawaの協力で運営しています。

<https://wikipedia-kaido.github.io/>

シチズンサイエンス

➡ iNaturalist

- ➡ 生物の観察記録を公開・共有
 - ➡ 種の同定にはプロの科学者も協力
- ➡ データのライセンスは参加者が選択
 - ➡ オープンデータとは限らない

The screenshot shows the iNaturalist website interface. At the top, there is a search bar and navigation links for '探索' (Explore), '自分の観察記録' (My observations), 'コミュニティ' (Community), and 'その他' (Other). Below this, the '観察記録' (Observations) section is active, displaying a grid of observation records for various plant species. The grid includes the following entries:

種名	学名	投稿日時	ライセンス
ヤツデ	<i>Fatsia japonica</i>	16日前	研究用
コウヤボウキ	<i>Pertya scandens</i>	17日前	研究用
チャノキ	<i>Camellia sinensis</i>	17日前	研究用
ミソソバ	<i>Persicaria thunbergii</i>	2か月前	研究用
マルバルコウ	<i>Ipomoea coccinea</i>	2か月前	研究用
アキニレ	<i>Ulmus parvifolia</i>	2か月前	カジュアル
ミスヒキ	<i>Persicaria filiformis</i>	2か月前	研究用
センニンソウ	<i>Clematis terniflora</i>	2か月前	研究用

シチズンサイエンス

▶ みんなで翻刻

- ▶ くずし字で書かれた古文書を活字化
 - ▶ AIくずし字認識がサポート
- ▶ 翻刻文データのライセンスはCC BY-SA

みんなde翻刻
MINNA DE HONKOKU

歴史資料を読み解く 歴史資料を解読するために お問い合わせ JOIN 参加する

みんなde翻刻
MINNA DE HONKOKU

仲間と、AIと協力して、
古文書解読に参加しよう！

JOIN 参加する

※現在新バージョンの試験運用中です。
安定稼働するまで旧バージョンも併せてご利用ください。

いいね! ツイート B!ブックマーク LINEで送る

月待ビンゴ プロジェクト

- ▶ Twitterを用いて「月待塔」の情報を収集・公開
- ▶ 月待塔とは
 - ▶ 特定の月齢の夜に信仰行事をおこなう講中が造立した塔

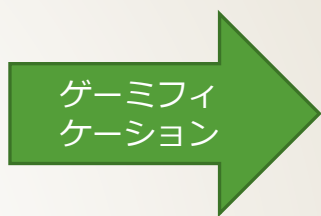
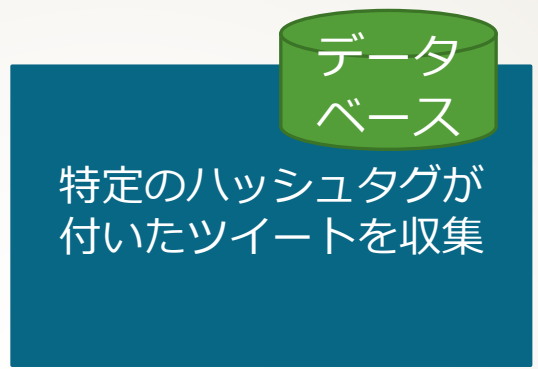
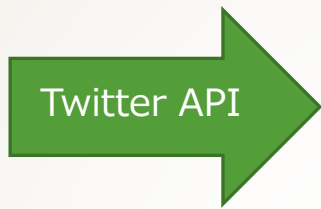


- ▶ 庚申塔、道祖神などと比べるとマイナー

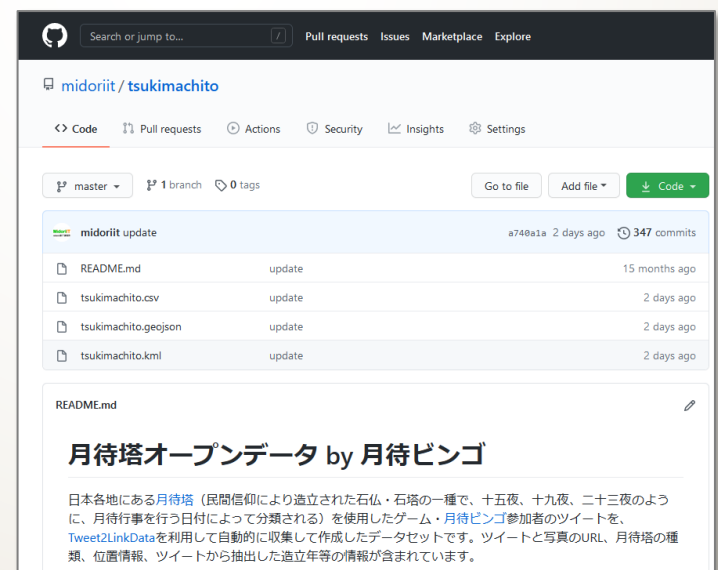
月待ビンゴの仕組み



ハッシュタグと
緯度経度を付けて
ツイート



ビンゴカード



<https://github.com/midoriit/tsukimachito>

月待塔オープンデータ

▶ データ項目

- ▶ ツイートした日付、ツイートしたTwitterアカウント名、ツイートのURL、添付画像のURL

- ▶ Twitter APIで取得

▶ 緯度経度

- ▶ 以前はTwitter APIで取得、現在はツイート本文から抽出

- ▶ 緯度経度から住所を逆ジオコーディング

▶ 月待塔の種類、造立年、所在地 (POI)

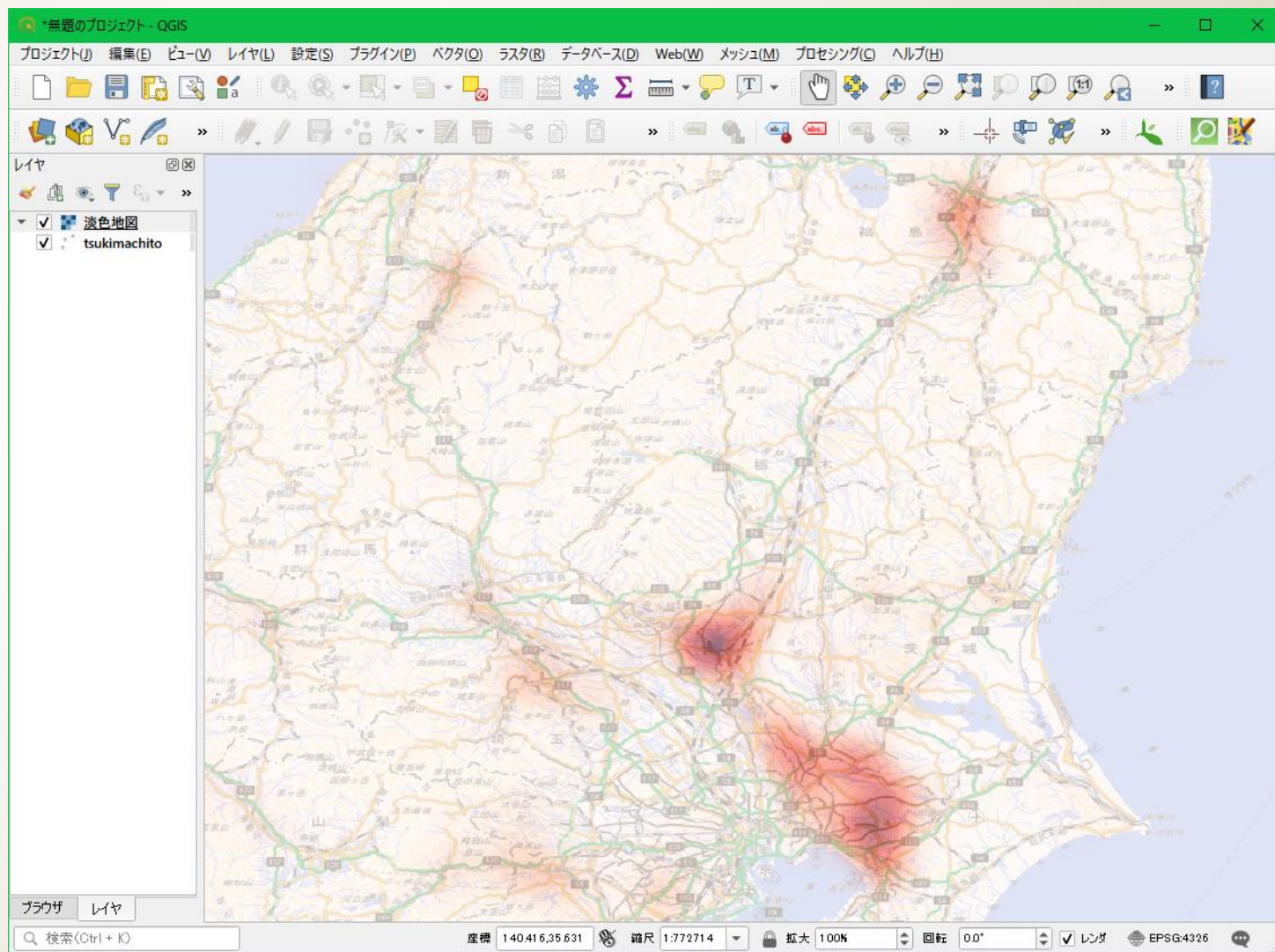
- ▶ ツイート本文から抽出

▶ CC0ライセンス (但しURLの[リンク先には著作権あり](#))

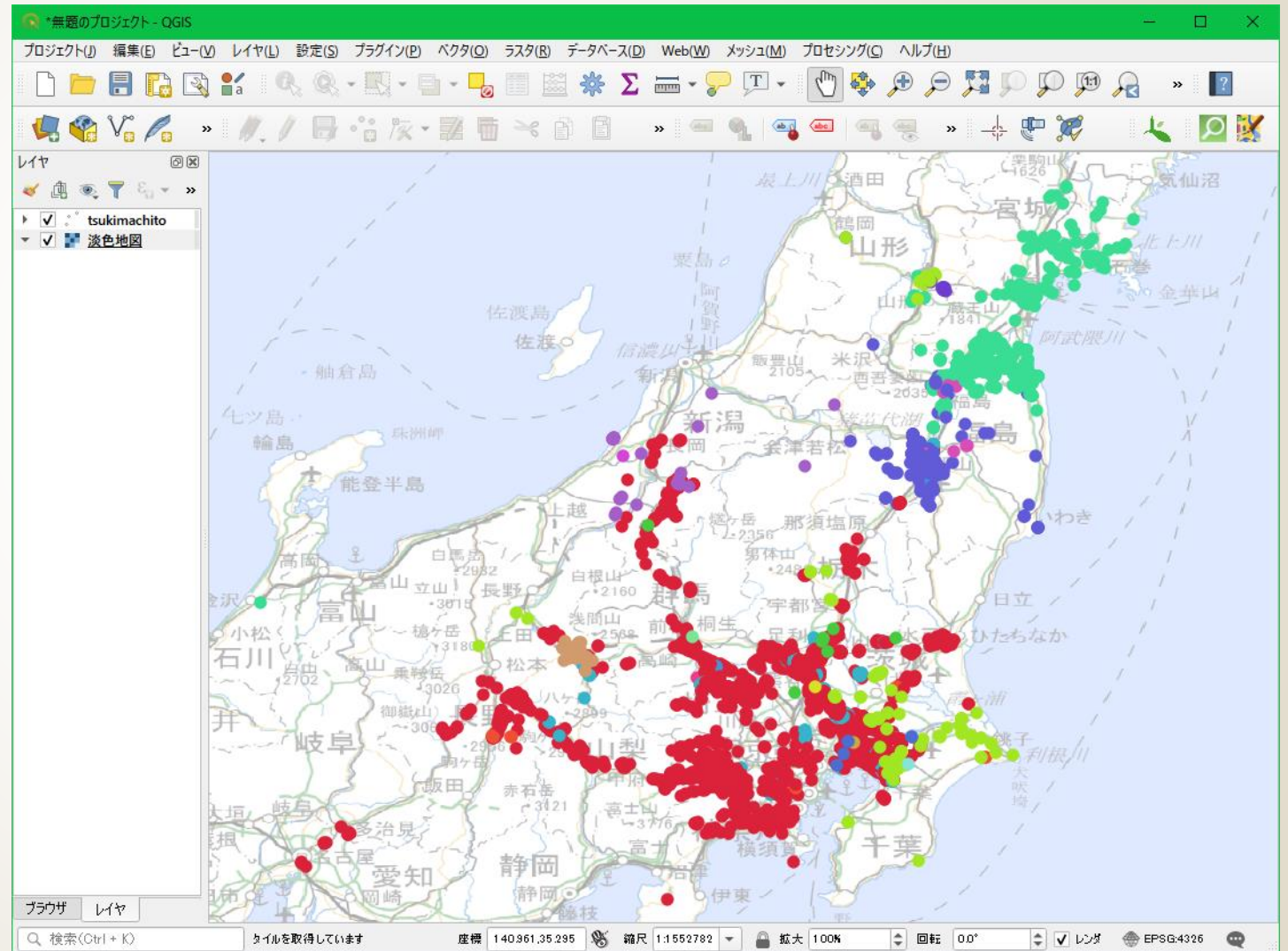
QGISを用いた月待塔オープンデータの分析

- ➡ 実際にQGISを使用してハンズオンを実施します
- ➡ 詳細な手順は別紙資料参照
 - ➡ GitHubからCSVファイルをダウンロード
 - ➡ CSVファイルをQGISに読み込む
 - ➡ 背景地図の表示
 - ➡ ヒートマップの作成
 - ➡ データの色分けによる可視化と分析
 - ➡ クロス集計表の作成
- ➡ 気づきを共有しましょう

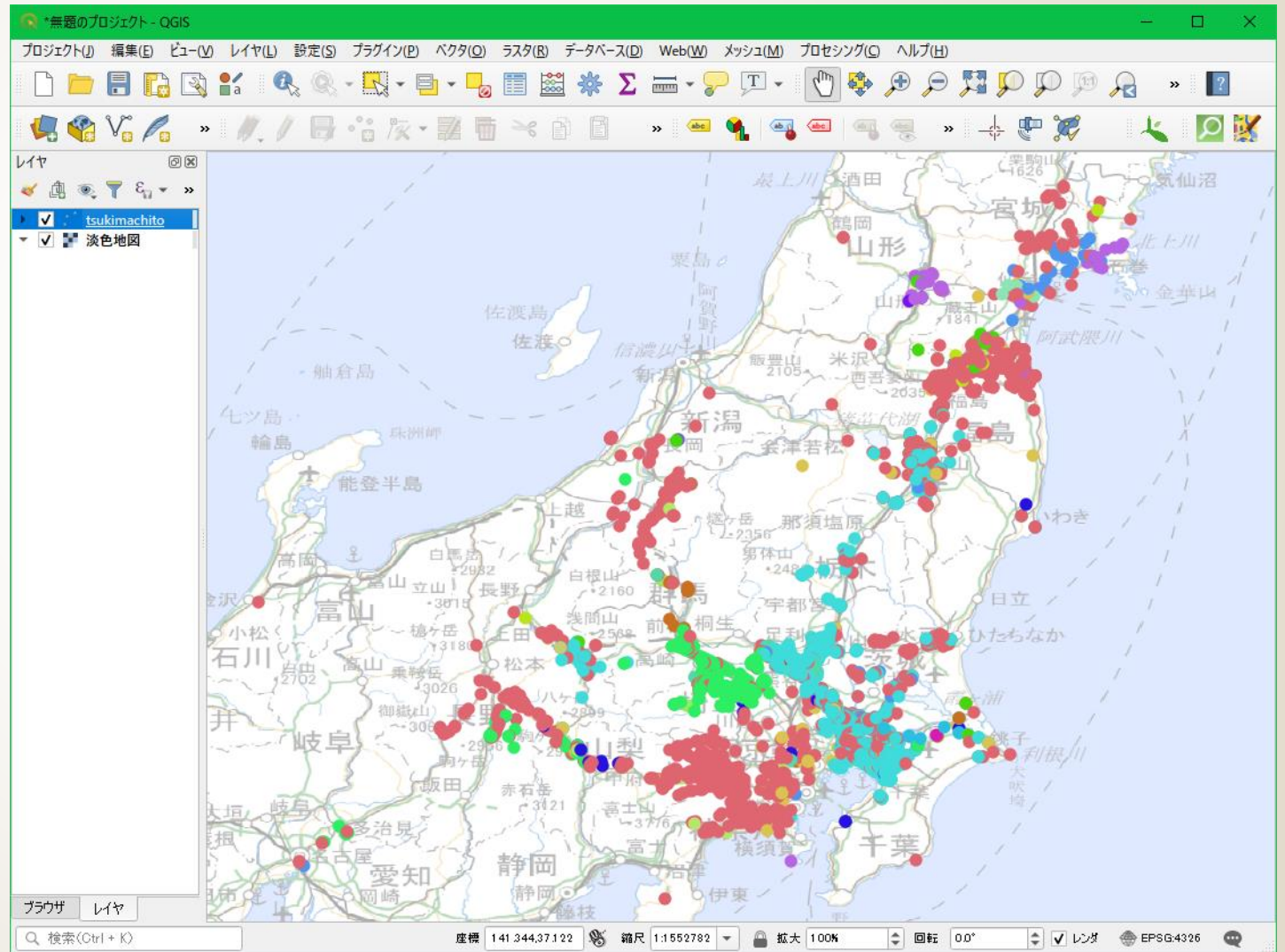
課題 1 : ヒートマップ



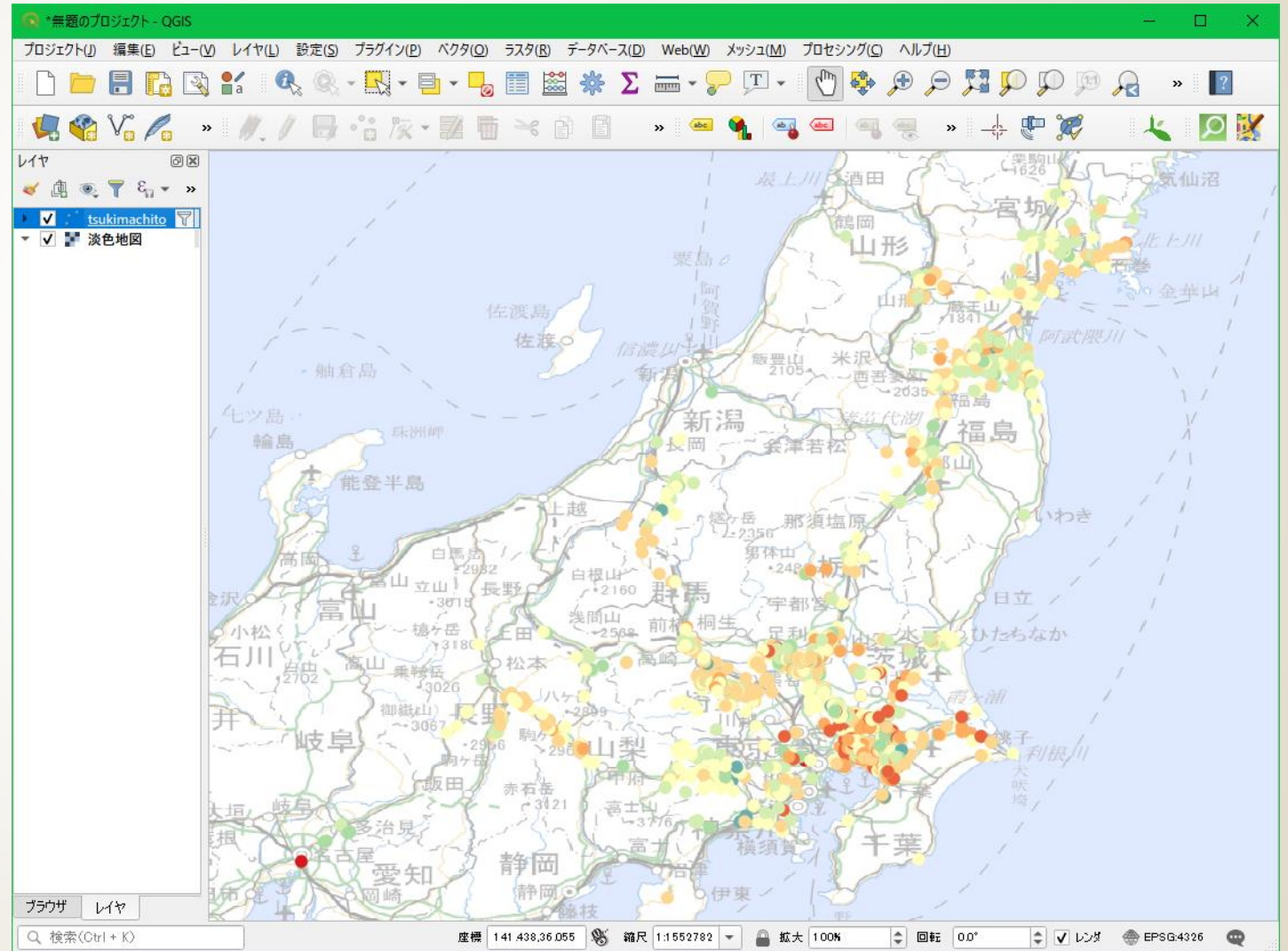
課題 2 : ツイートしたアカウント毎の色分け



課題 3 : 月待塔の種類による色分け



課題 4 : 月待塔の造立年による色分け



課題 5 : クロス集計表

The screenshot shows the Group Stats application interface. The main window displays a cross-tabulation table with 'prefecture' as the column headers and 'contributor' as the row headers. The data is summarized in the following table:

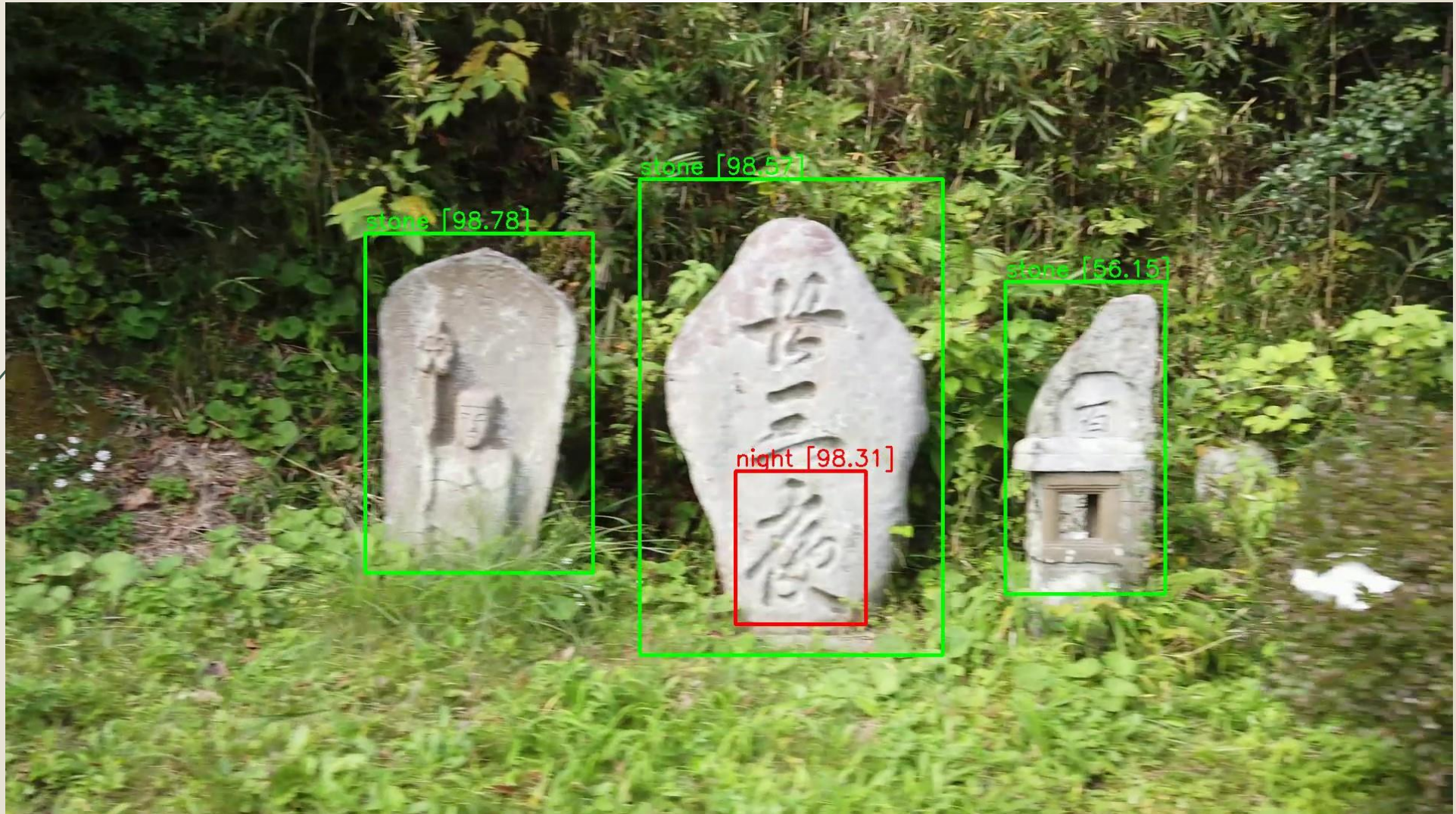
contributor	佐賀県	千葉県	埼玉県	宮城県	山形県	山梨県	岐阜県	岩手県	愛知県	新潟県	東京都	栃木県	石川県
HaNa87_8					14								
mamehito				224	8								
fragment1873										5			
higa4		1											
jadefoxrot			1							1		5	
nouveau_vent_tk		1											
nyr884		2											
sekinema			1										
strelitziaaugus					2			1					
tk01k	20	724	305			129	5		6	129	91	277	
tsmie22		4									2		
tsukimachi_low										20			
tsukimachito		75	13			4					2	15	

The control panel on the right shows the following configuration:

- Layers: tsukimachito
- Fields: prefecture, tweet_date, tweet_id, tweet_url, type, year_built, year_built_ce, average, count, max, median, min
- Filter: (empty)
- Columns: prefecture
- Rows: contributor, tweet_id
- Value: use NULL values
- Buttons: Calculate, Clear

Calculate... 100% | generate view...100% | done.

月待塔オープンデータの活用



AI（機械学習）の教師データとしての利用

- ▶ 物体検出の機械学習には大量の教師データが必要
 - ▶ マイナーな石造物の画像を大量に集めるのは困難
 - ▶ 月待塔オープンデータには画像URLが含まれているので、簡単なプログラムにより画像ファイルの一括入手が可能
- ▶ ツイートの添付画像は著作権保護の対象だが...

▶ 著作権法 第四十七条の五

電子計算機を用いた情報処理により新たな知見又は情報を創出することによつて著作物の利用の促進に資する次の各号に掲げる行為を行う者は、**公衆への提供又は提示が行われた著作物**について、当該各号に掲げる行為の目的上必要と認められる限度において、当該行為に付随して、いずれの方法によるかを問わず、**利用を行うことができる。**

二 電子計算機による**情報解析を行い、及びその結果を提供**すること

ご清聴ありがとうございました

midoriit.com