

ジオタグ付きツイートにおける 特徴的ハッシュタグの分析

尾崎研究室所属

学籍番号 5410047 見玉翔太

研究の背景

山口「でもどうやって？」
松岡「プロに聞いてみようぜ！」
～カキ農園～
長瀬「こんにちわー！」
#昔話にTOKIOが介入

返信 ツイート お気に入り登録 その他

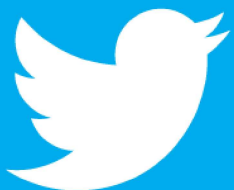
ツイート	お気に入り	
195	94	

12:22 - 2014年7月24日

#横浜

#理系あるある

#春アニメ



研究の背景

地域特有の
ハッシュタグの傾向がみたい！

横浜

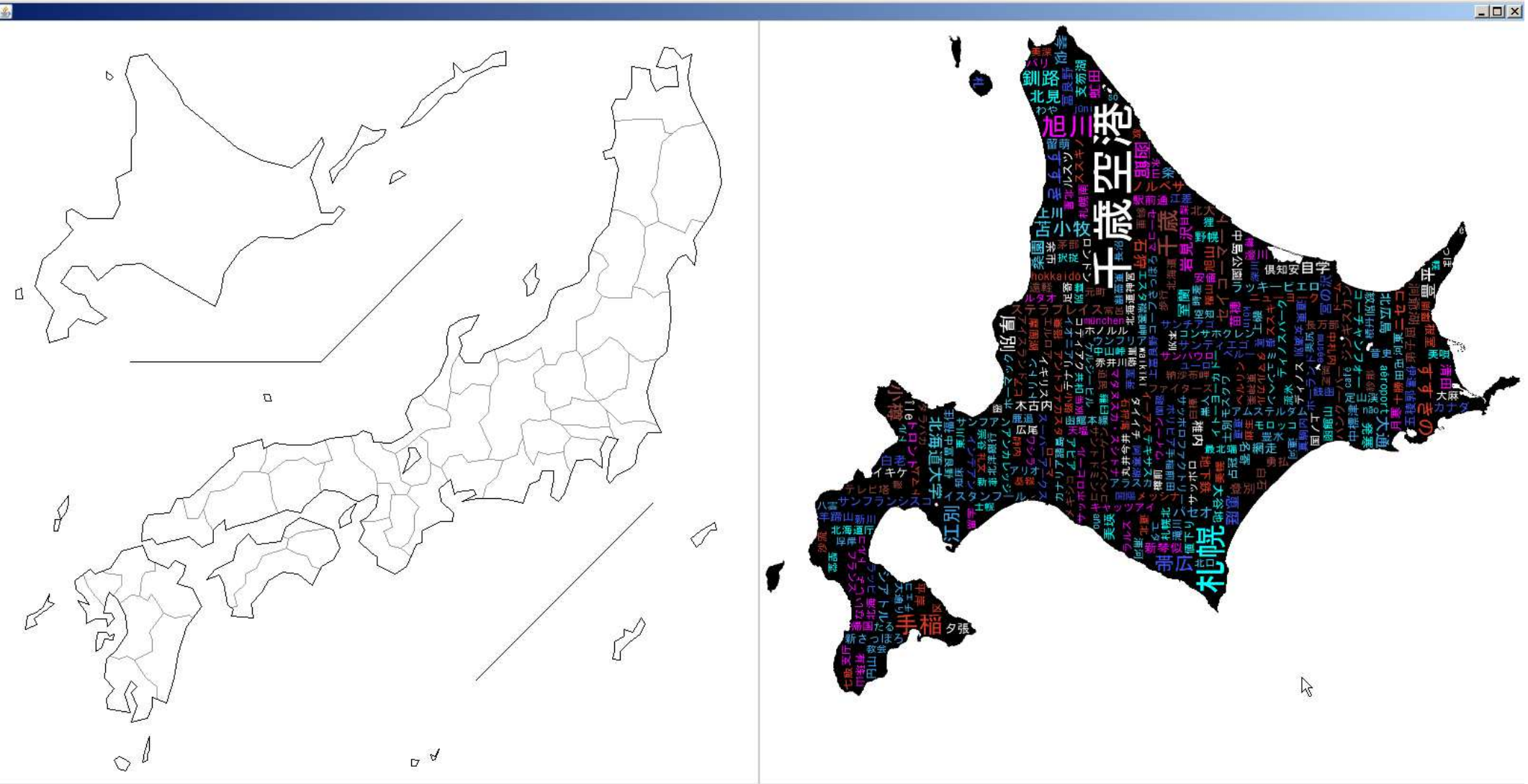
理系あるある

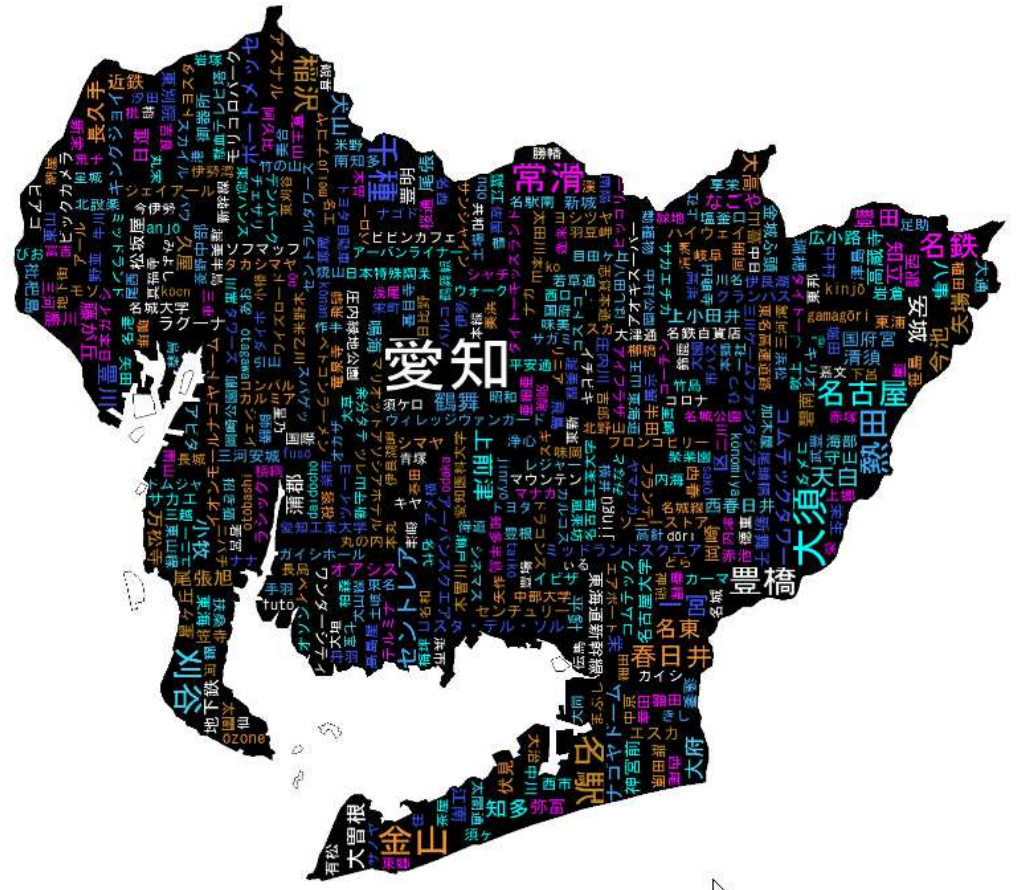
春アニメ



先行研究

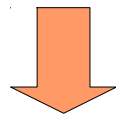
都道府県別タグクラウド



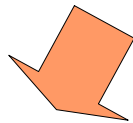


分析の手順

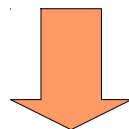
twitter 



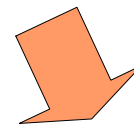
ジオタグ、ハッシュタグ付きツイート



県



県



県

データ収集期間：2013/10--2014/09 (約1年間)

Twitter Sample Stream から獲得

GPS 情報 (緯度, 経度) から都道府県を検出

分析の手順

ジオタグ、ハッシュタグ付きツイート

県

県

県

ハッシュタグ

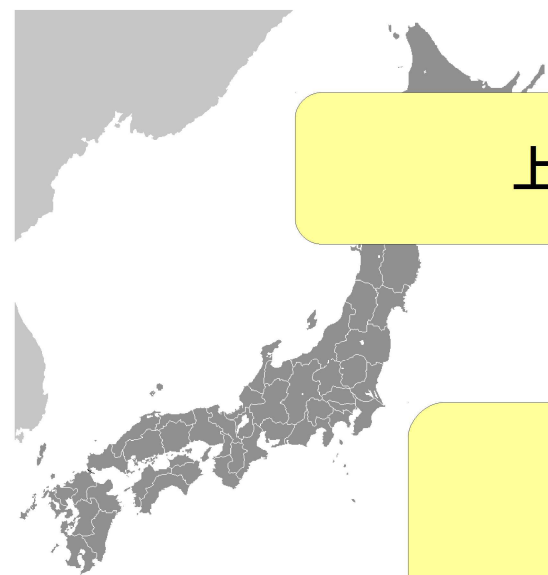
地域性の評価 (TF-IDF)

地域性でランキング

手作業です

上位 10 位以内での「土地」関係タグの集計

日本地図上に配置



地域性

その地域（だけ）で，どれだけ繰り返し替えし現れているか

$$\text{TF-IDF} = \text{TF} \times \text{IDF}$$

TF

繰り返し現れる

×

IDF

その地域で

文書内にいくつ出現？

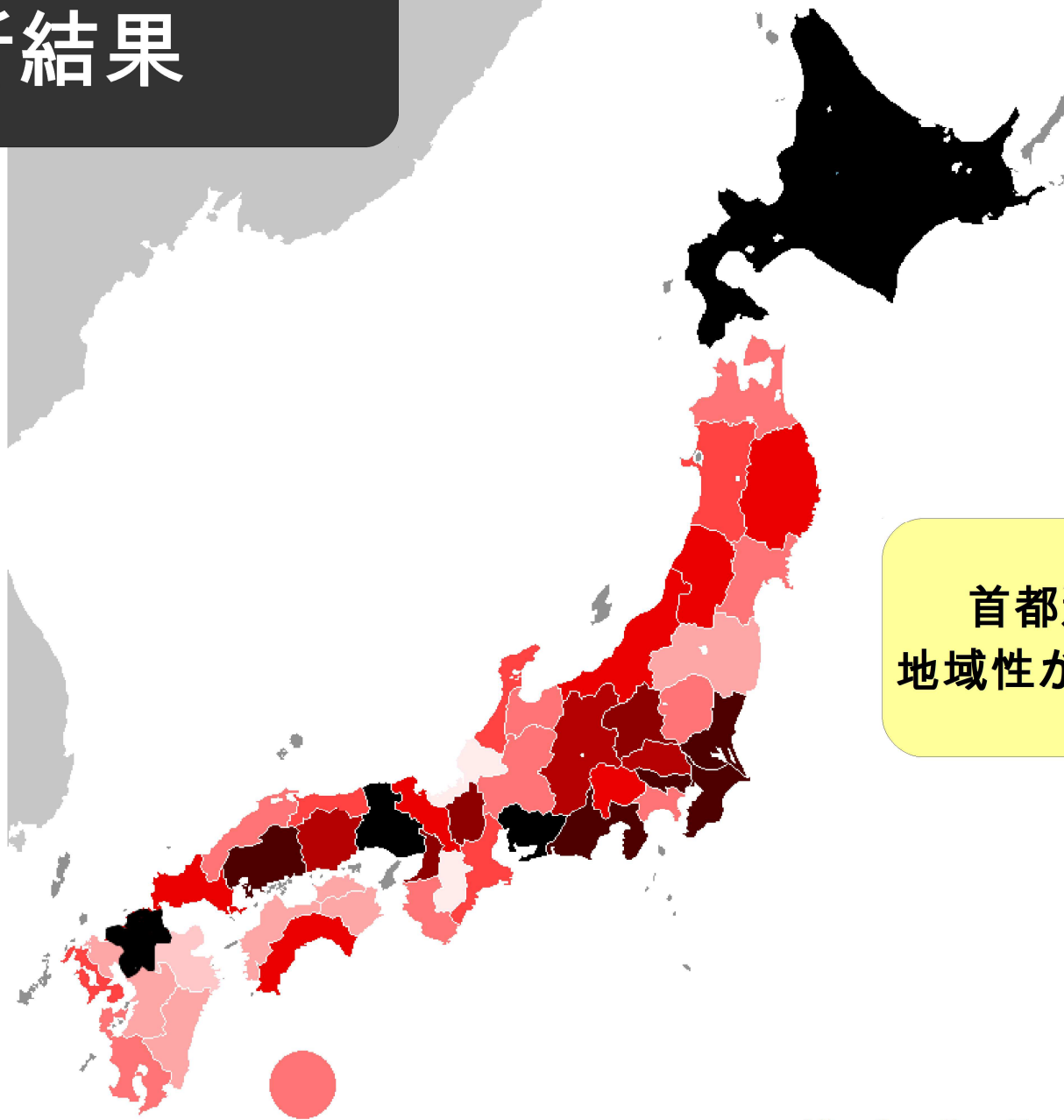
出現率

いくつの文書に出現？

逆数を取る

文書 = 都道府県

分析結果

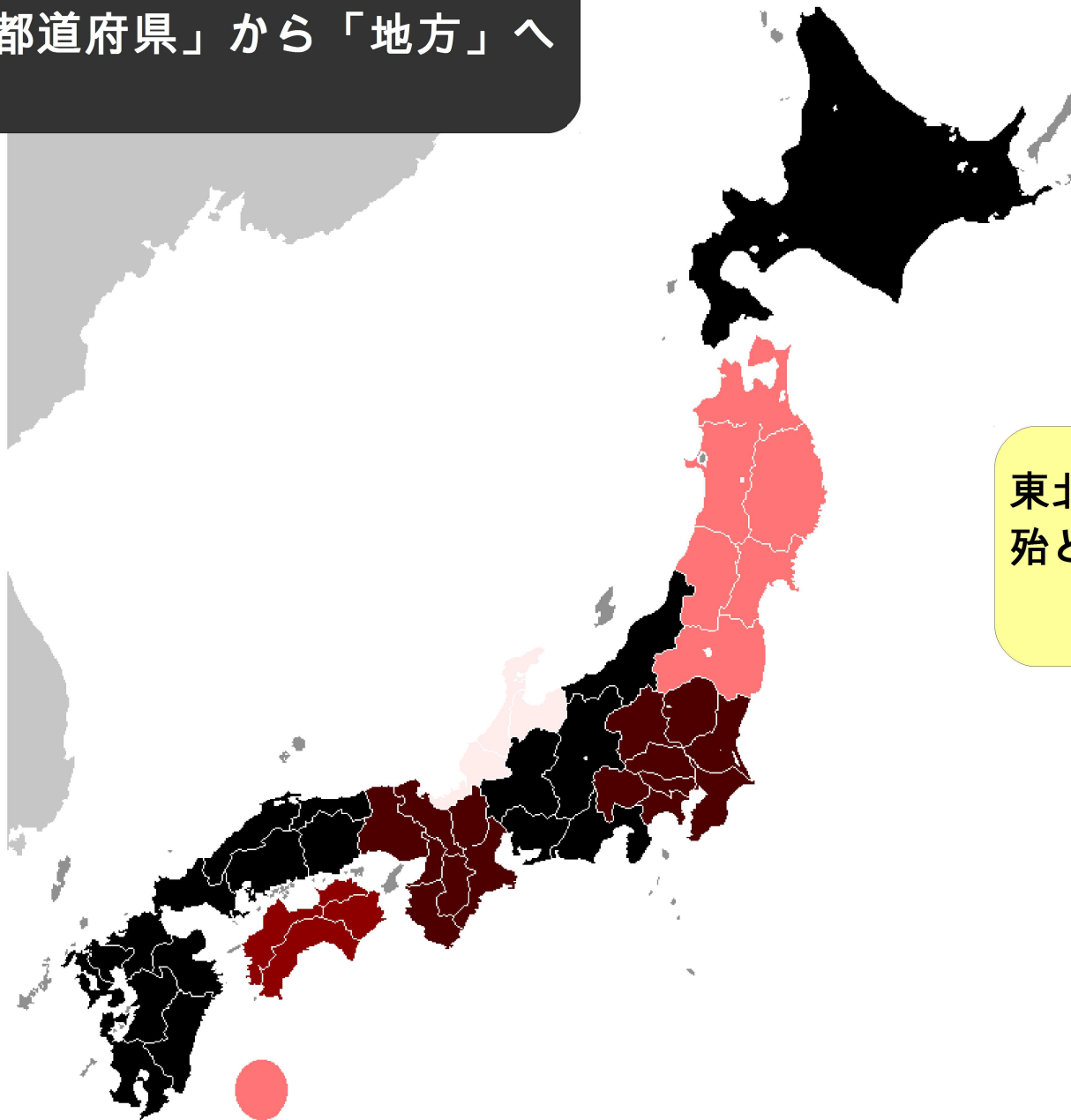


首都近辺において
地域性が高くなっている

10 9 8 7 6 5 4 3 2 1



Roll Up : 「都道府県」から「地方」へ

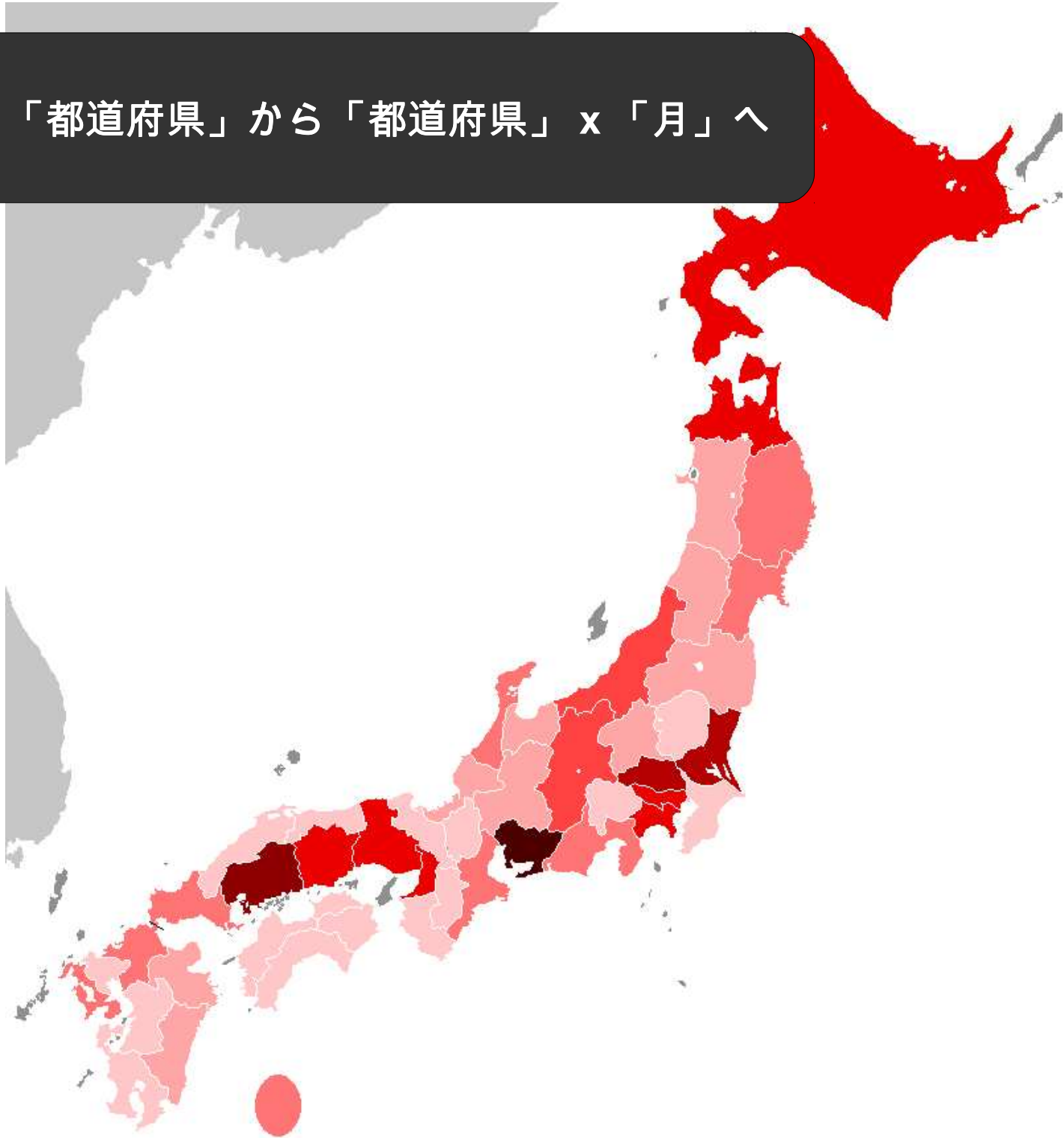


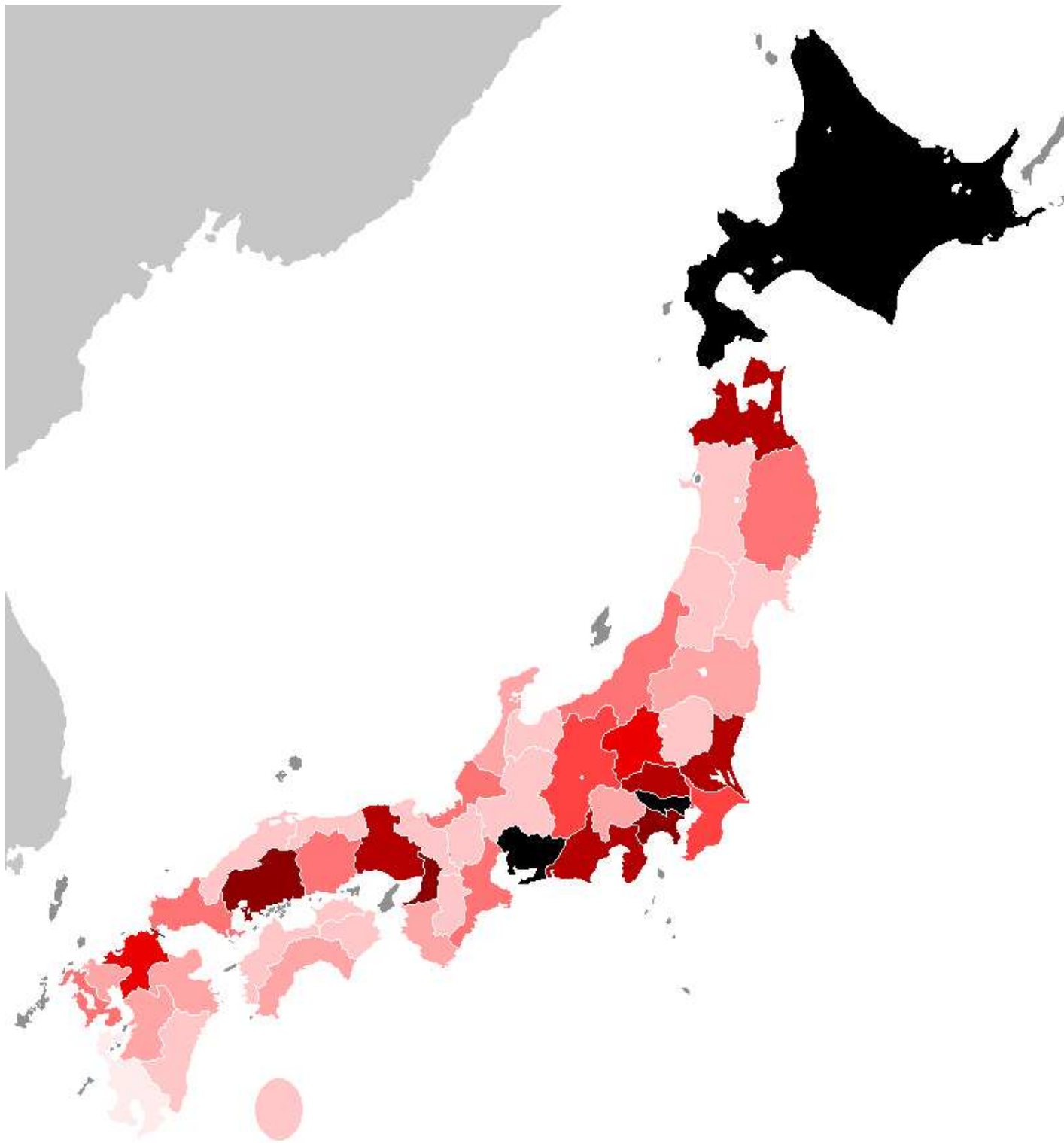
東北、沖縄地方を除く
殆どの地方が地域性が
高くなっている

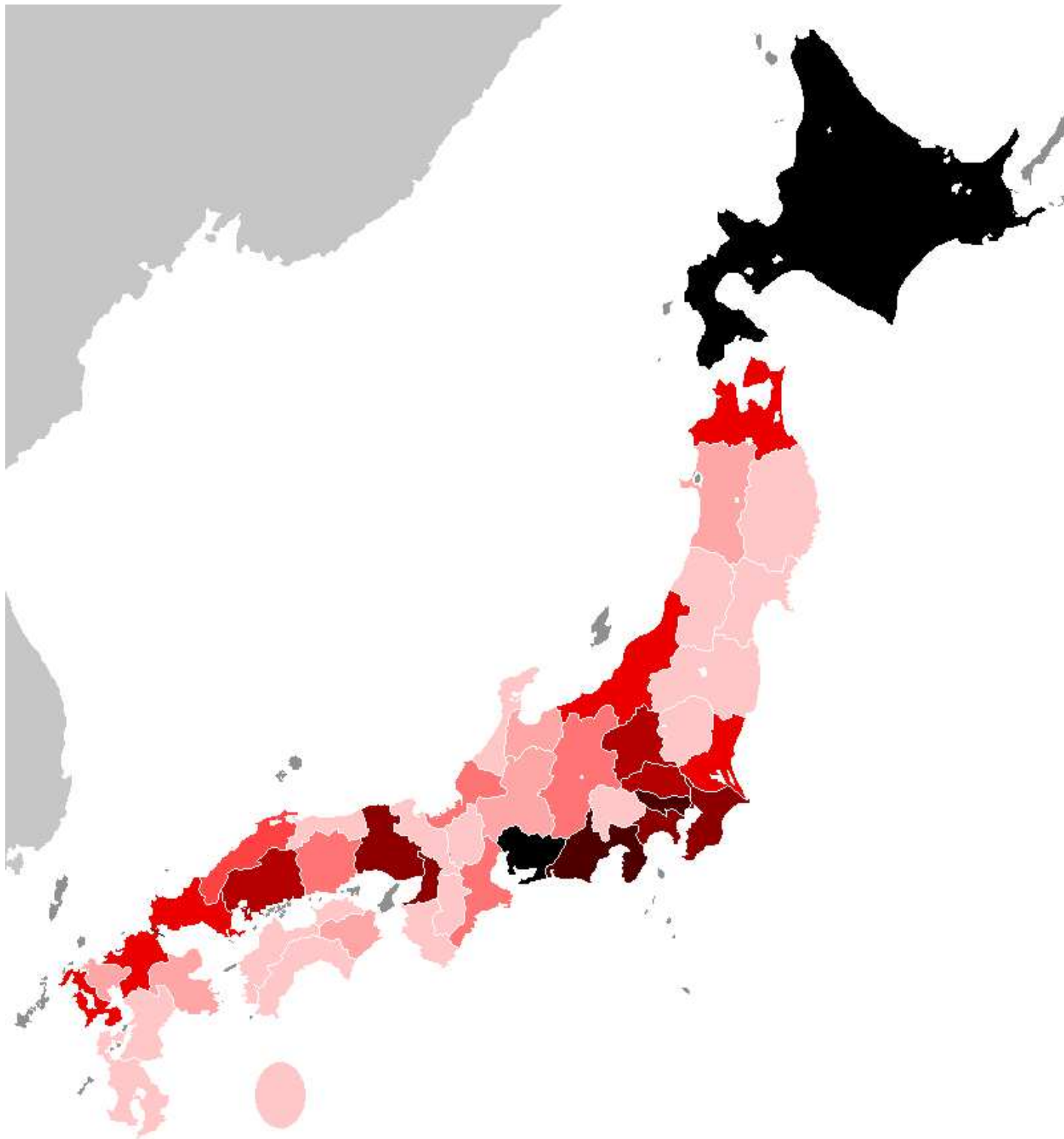
10 9 8 7 6 5 4 3 2 1

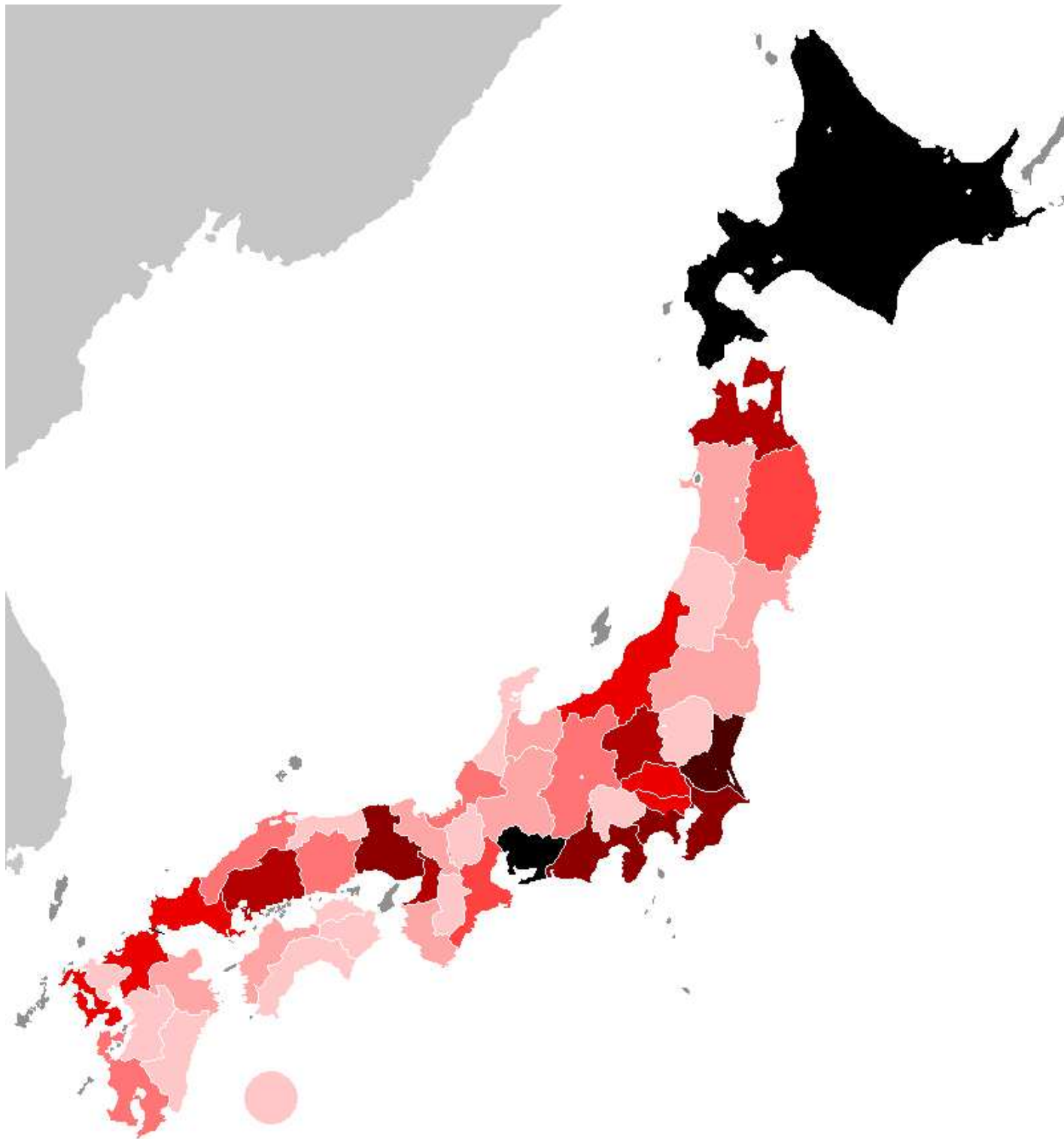


Drill Down : 「都道府県」から「都道府県」 x 「月」へ

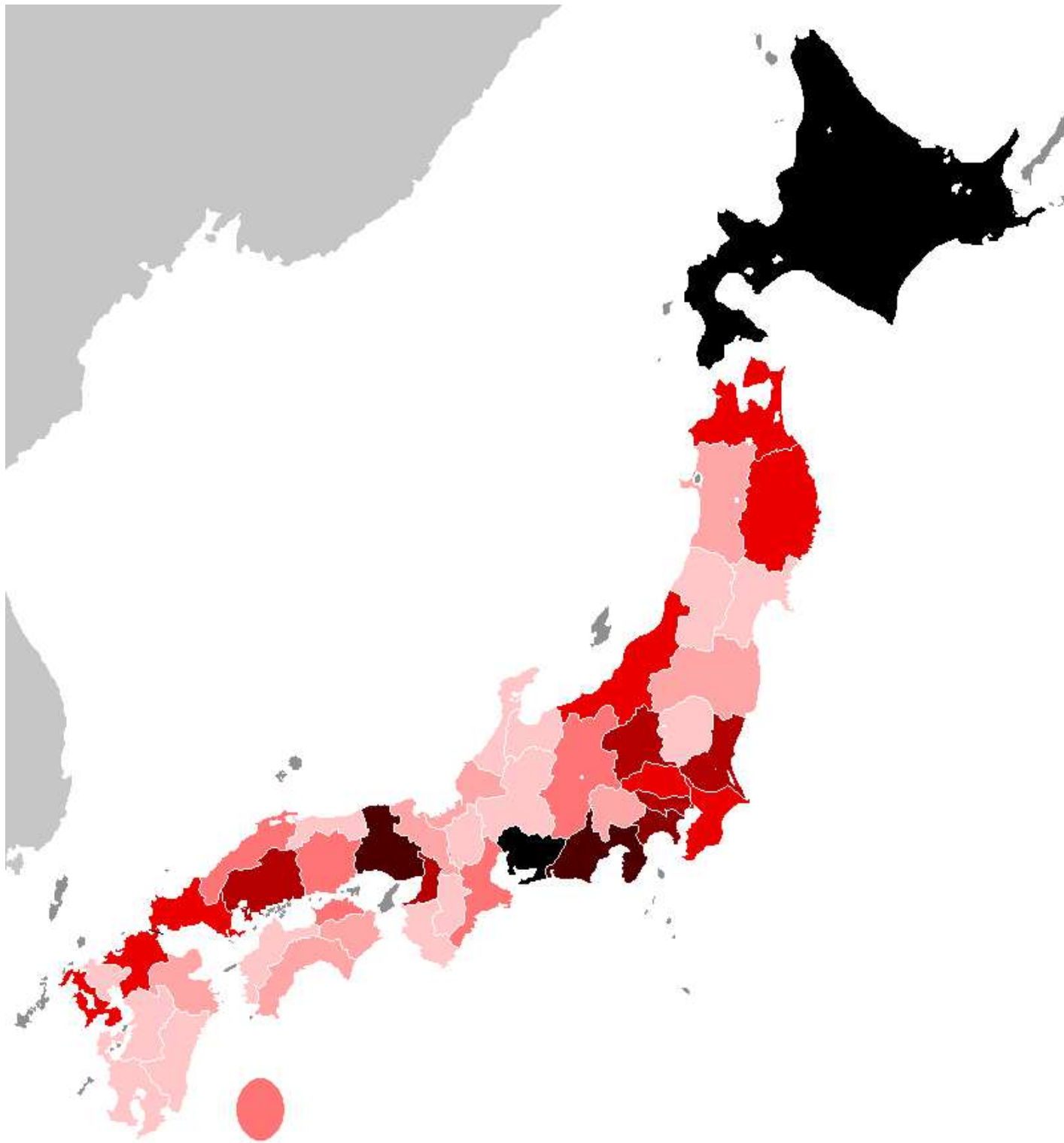


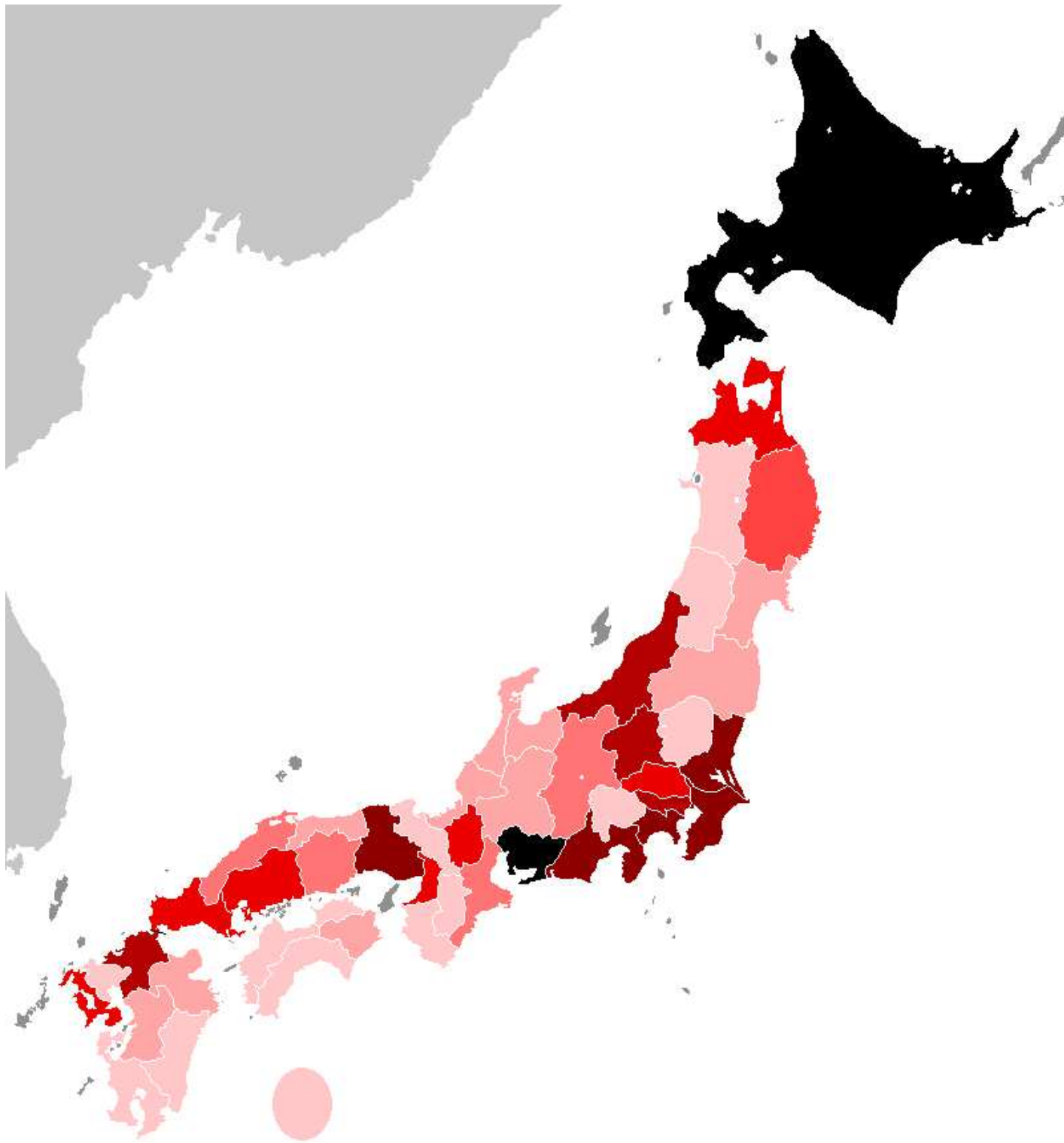


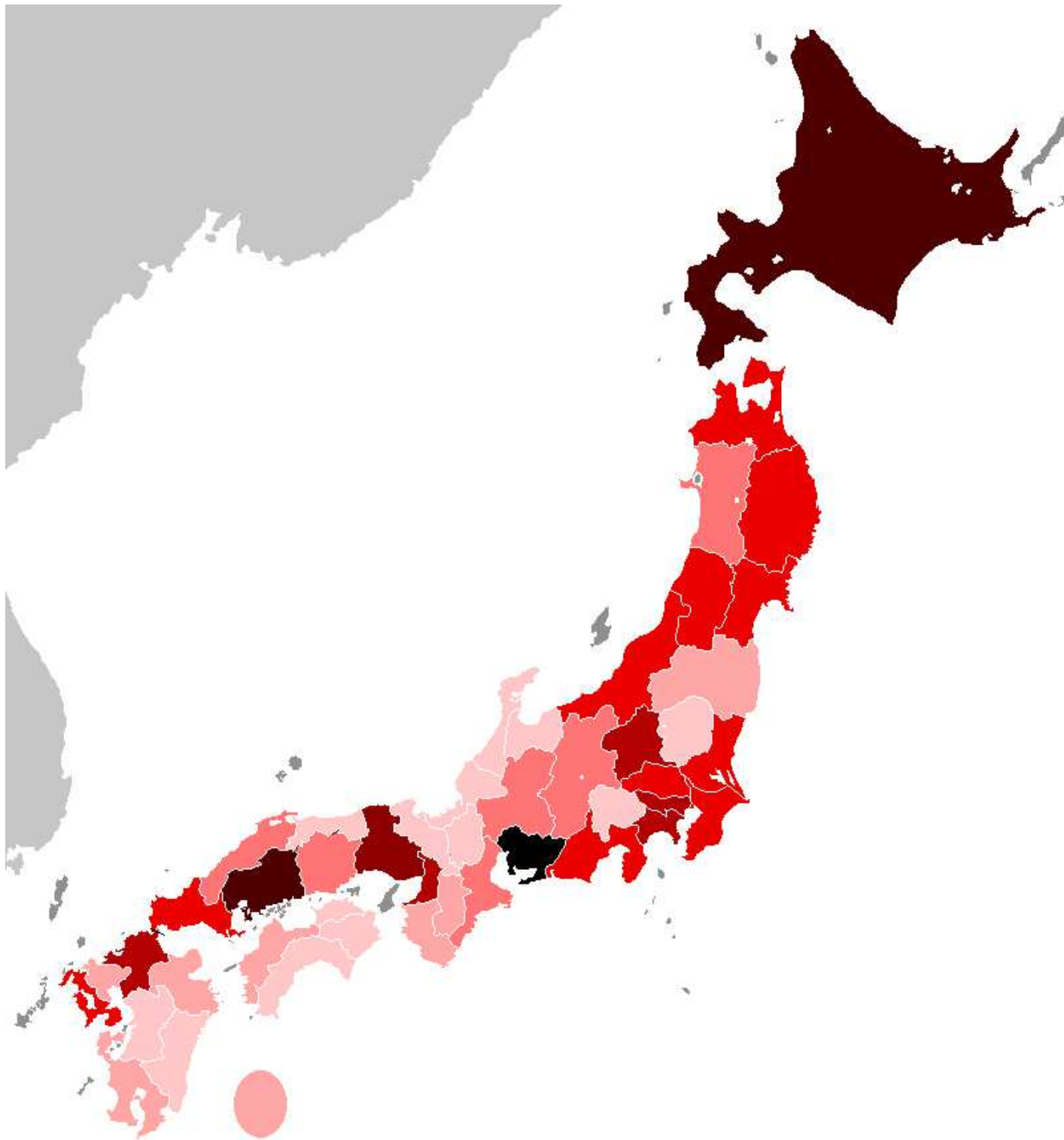


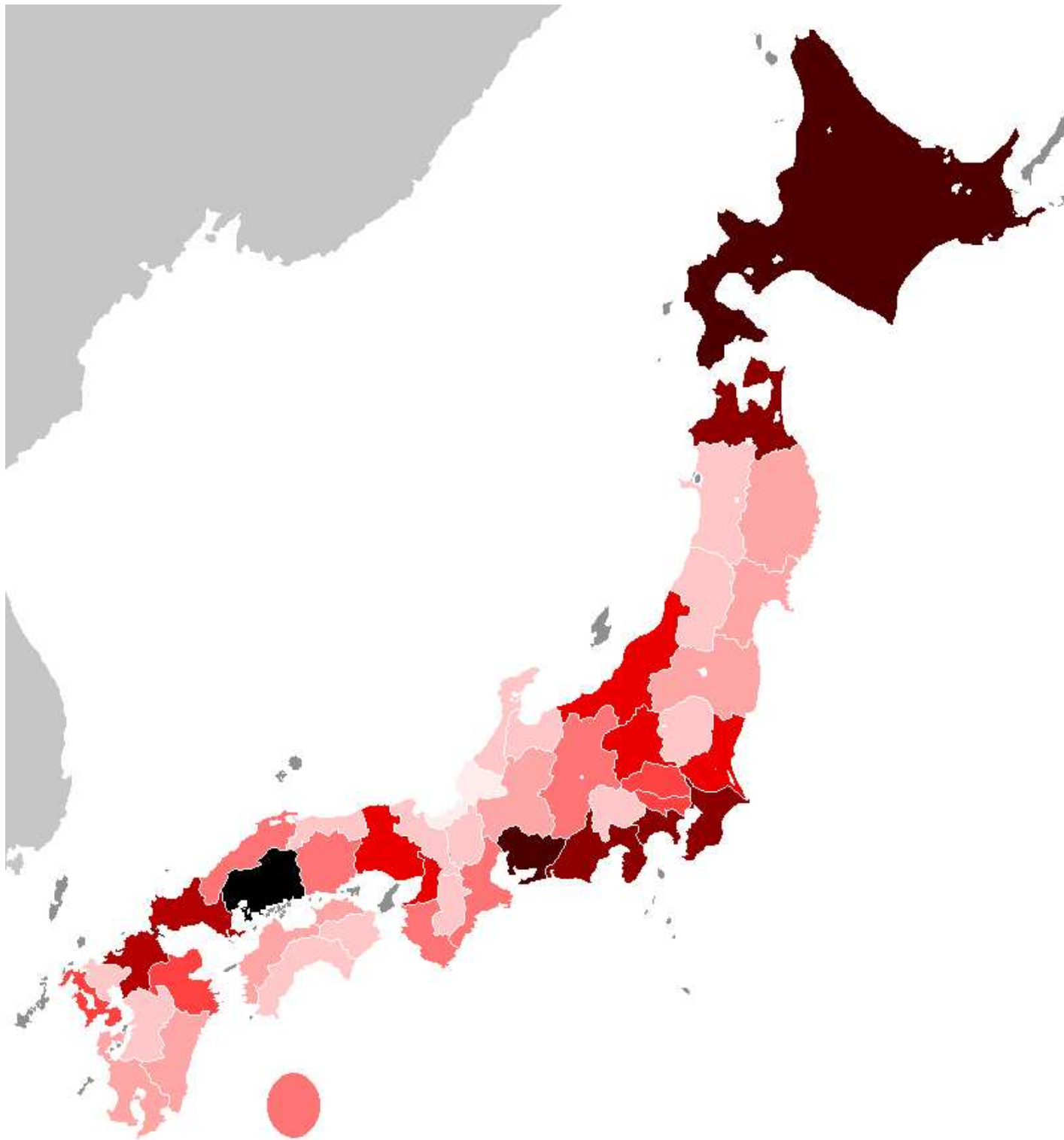


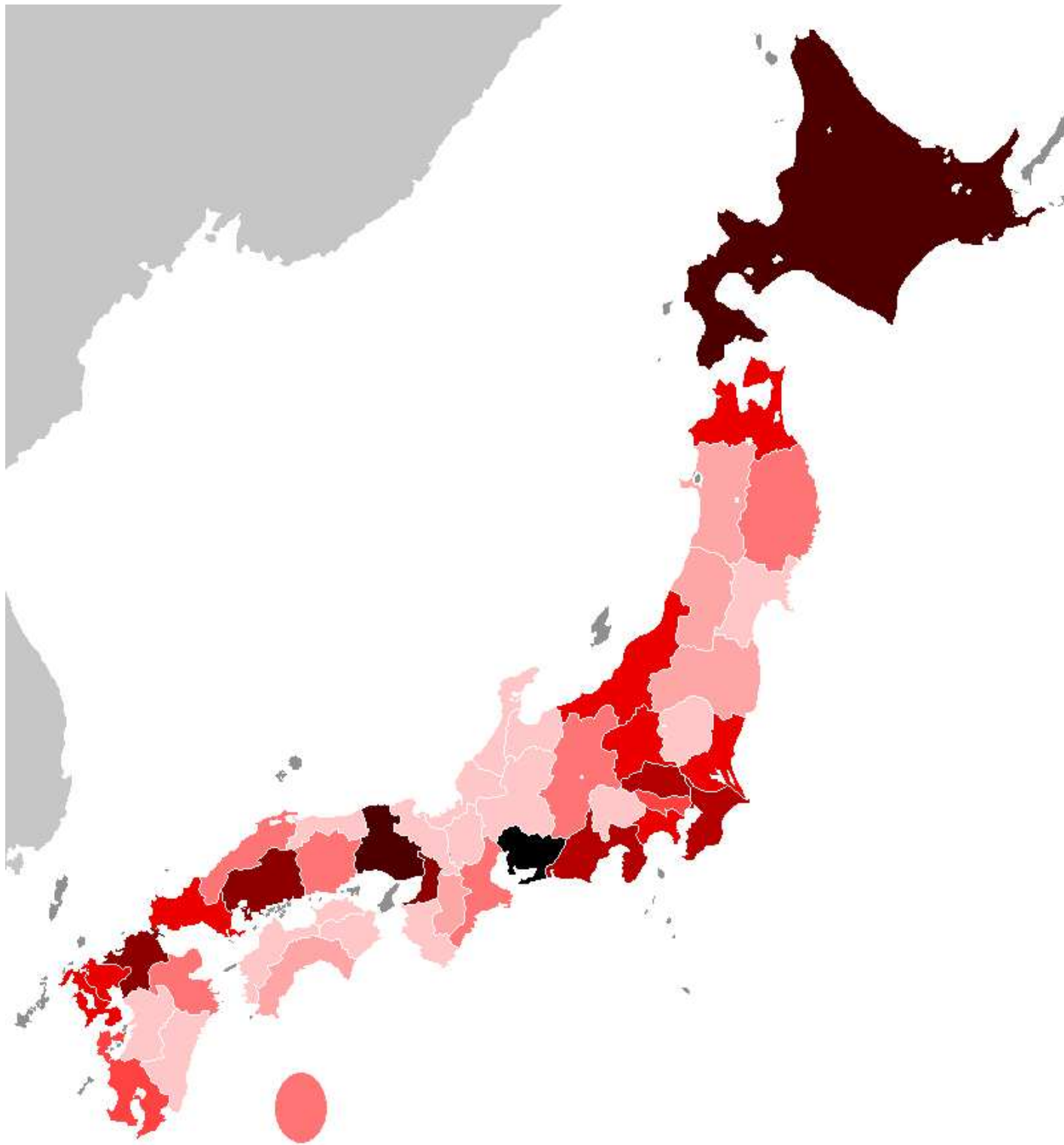
2014/1

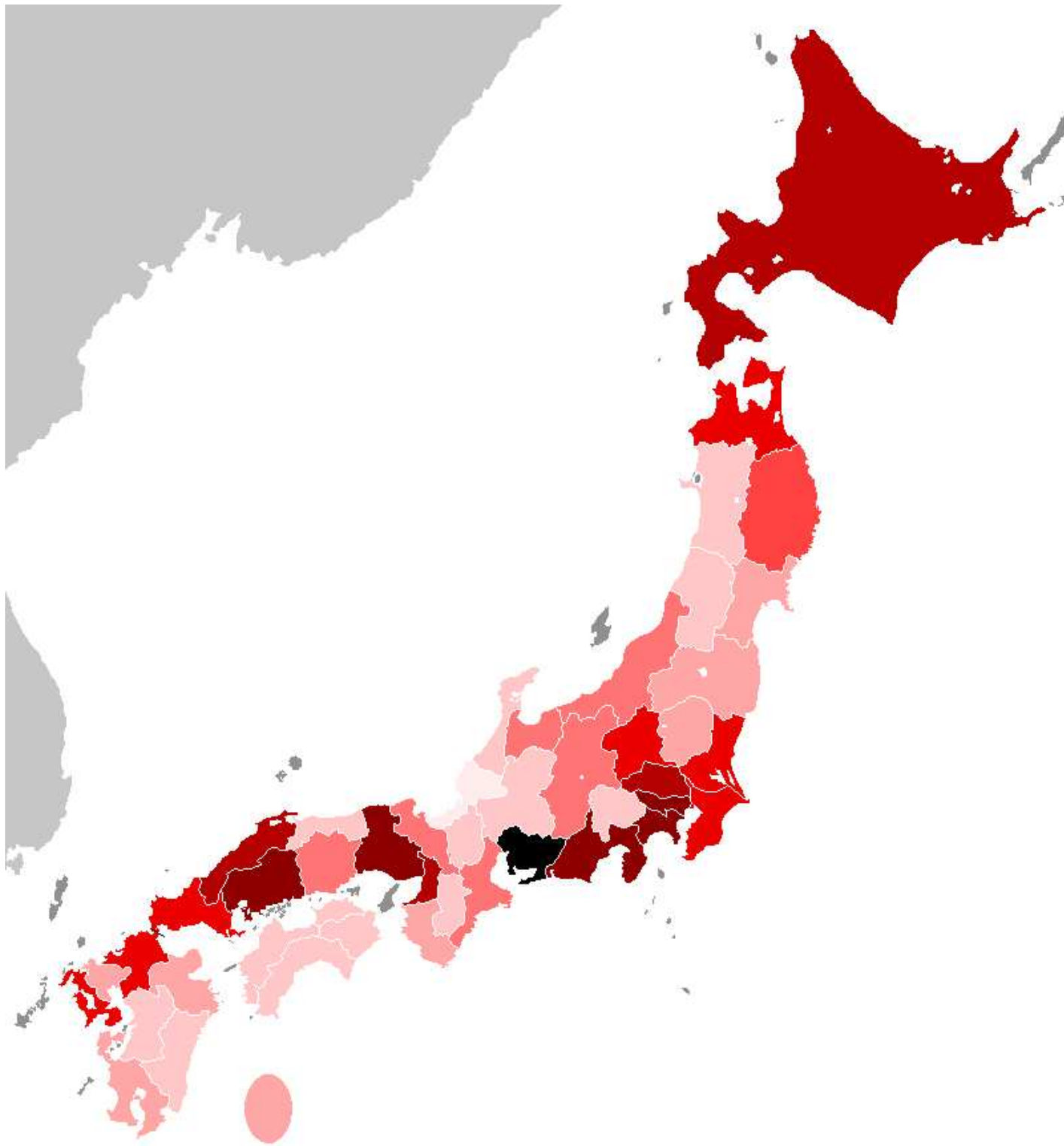


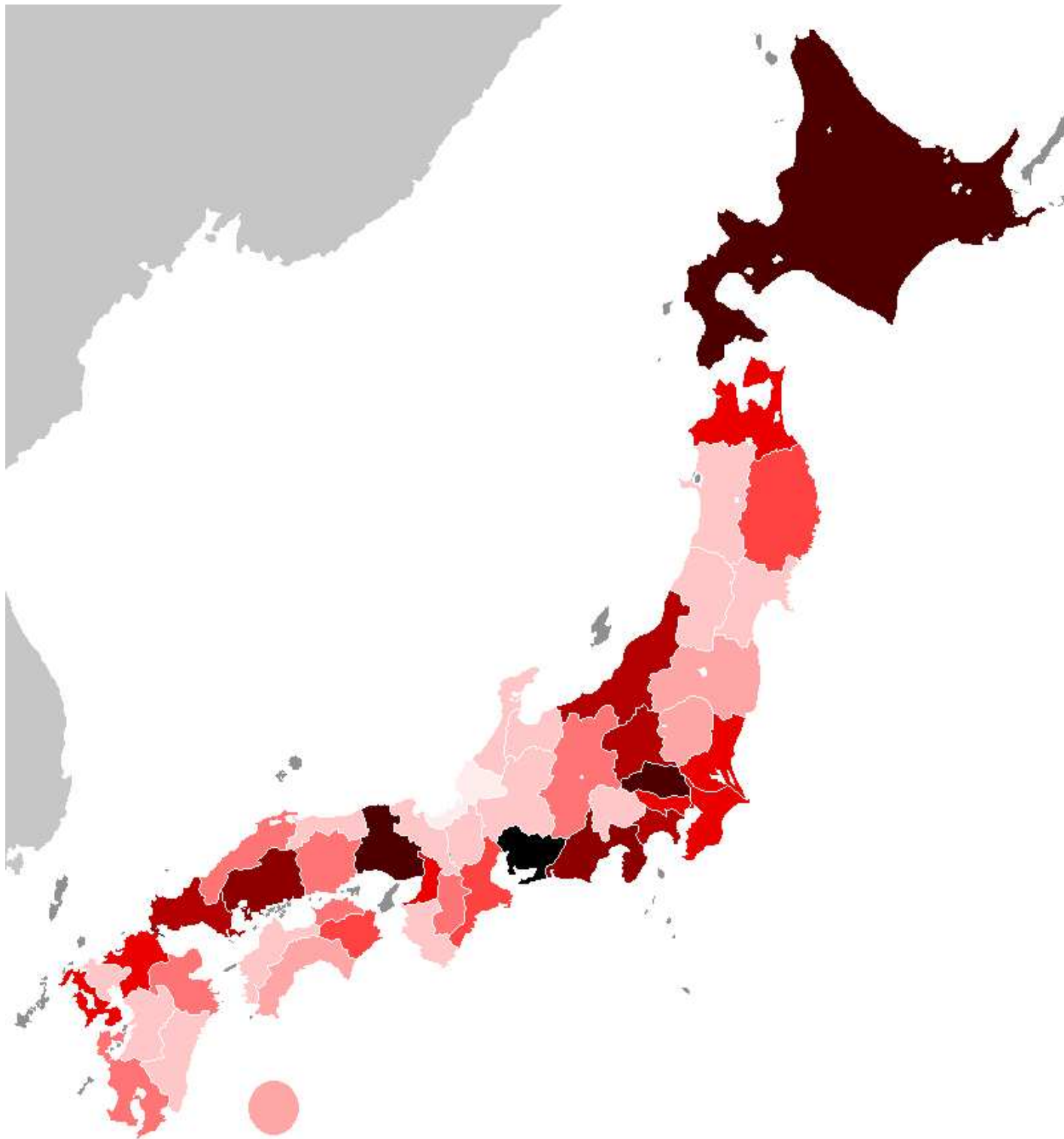


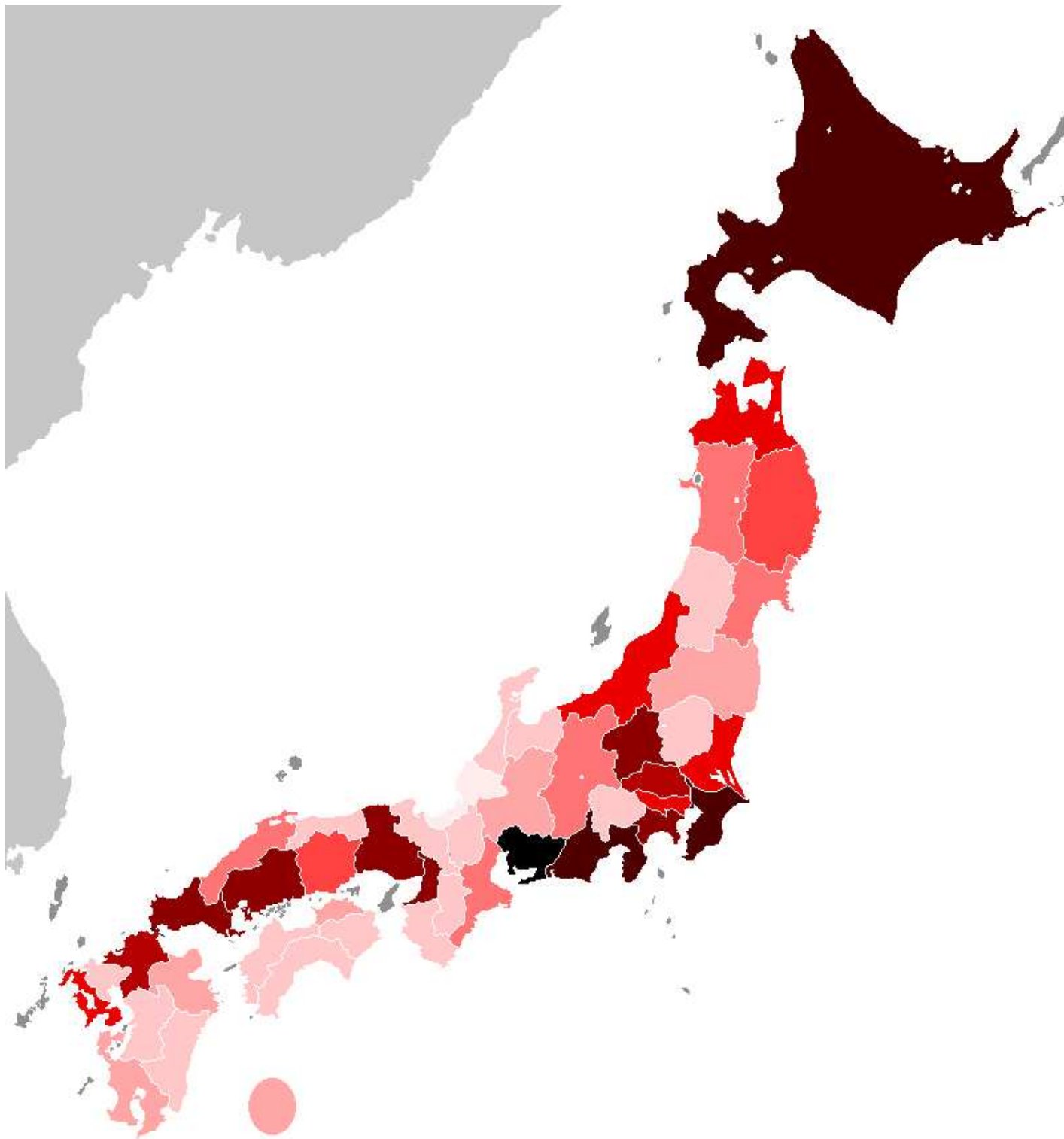


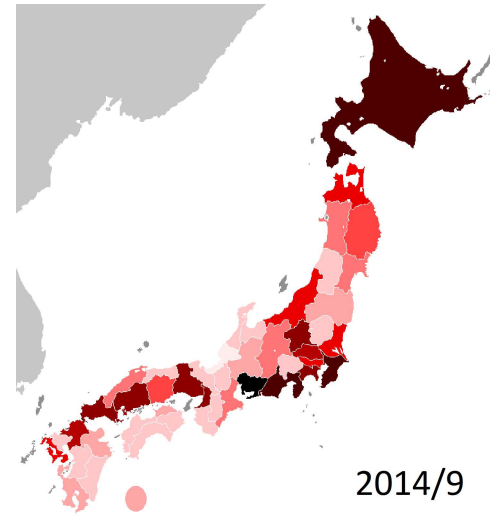
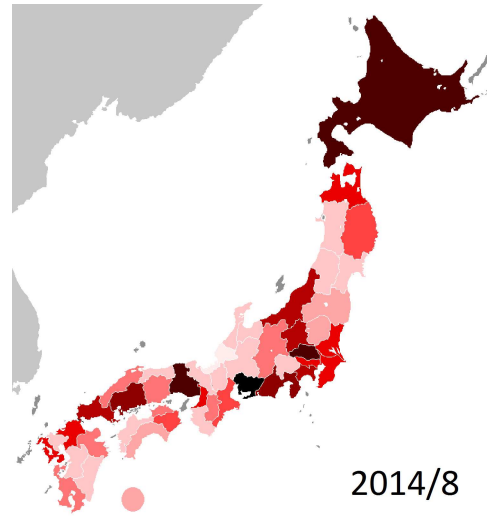
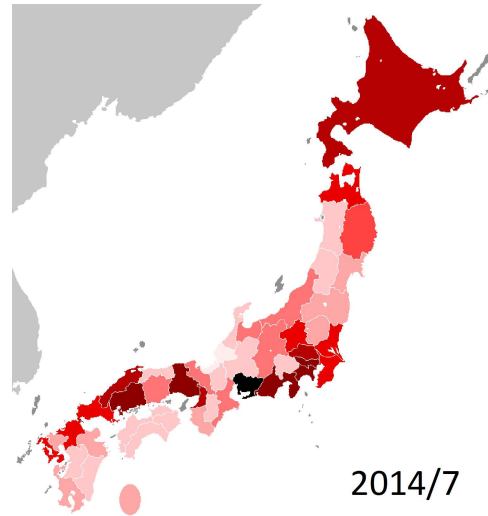
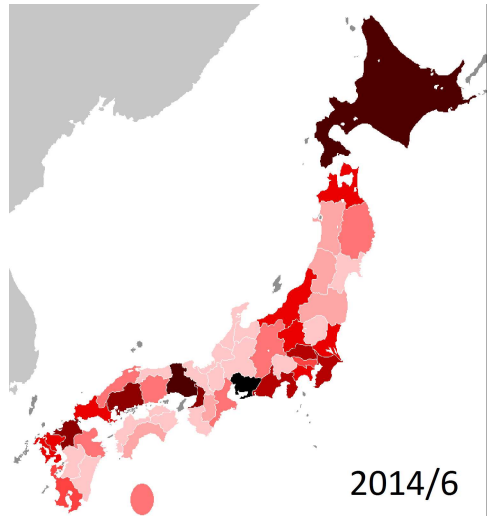
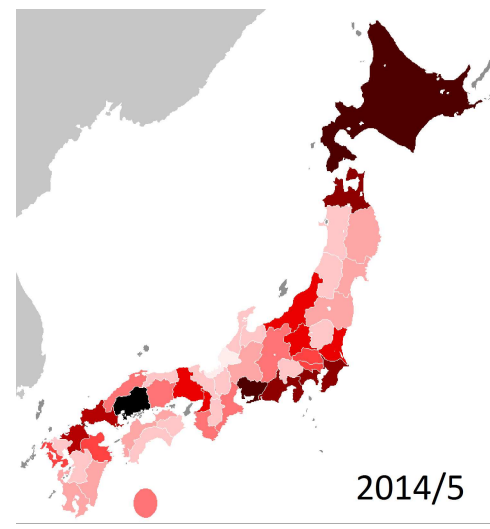
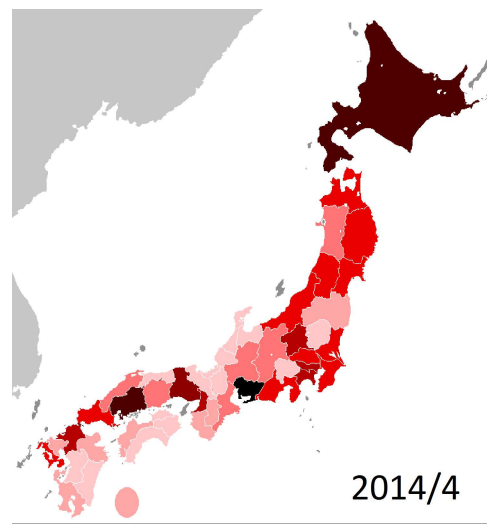
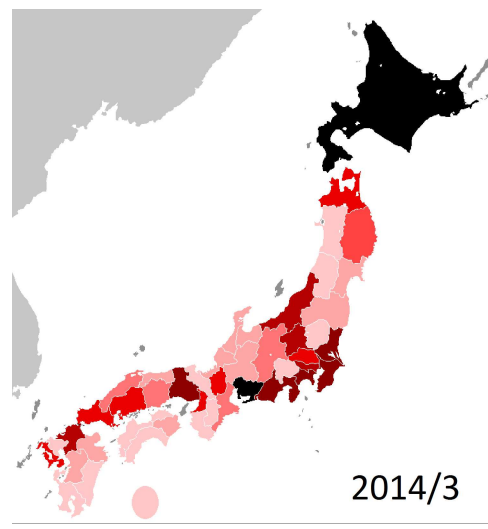
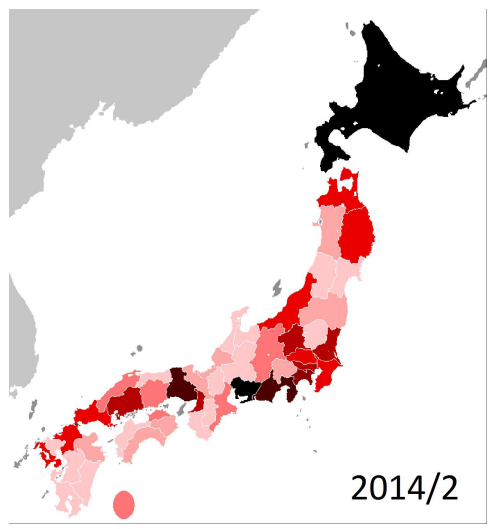
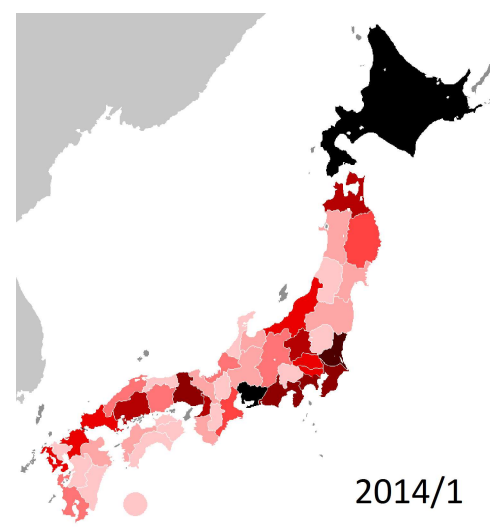
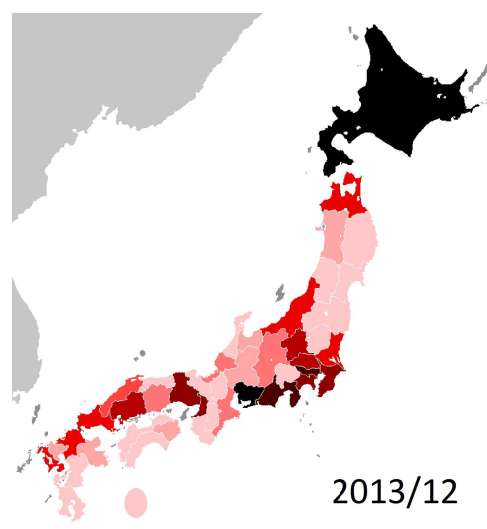
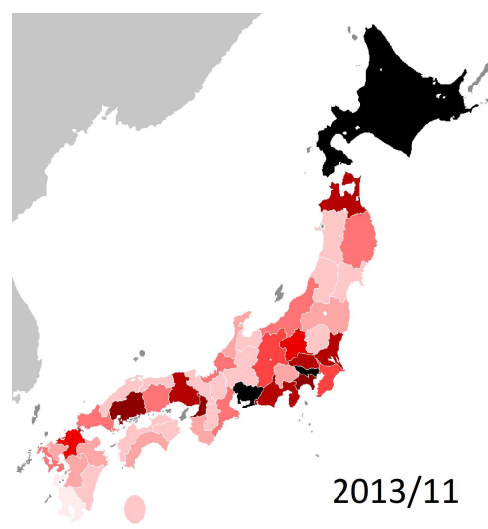
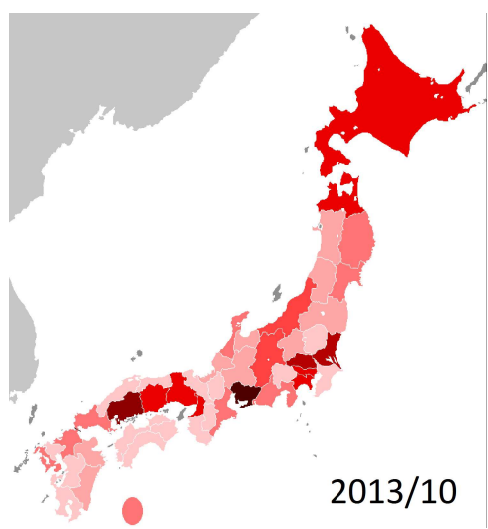




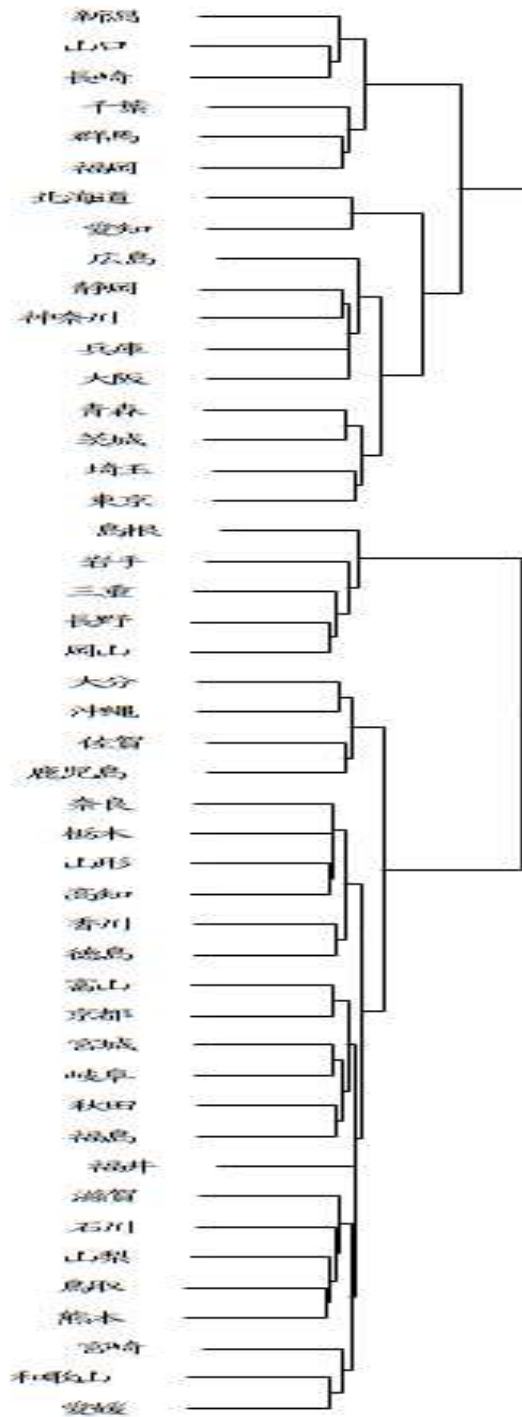






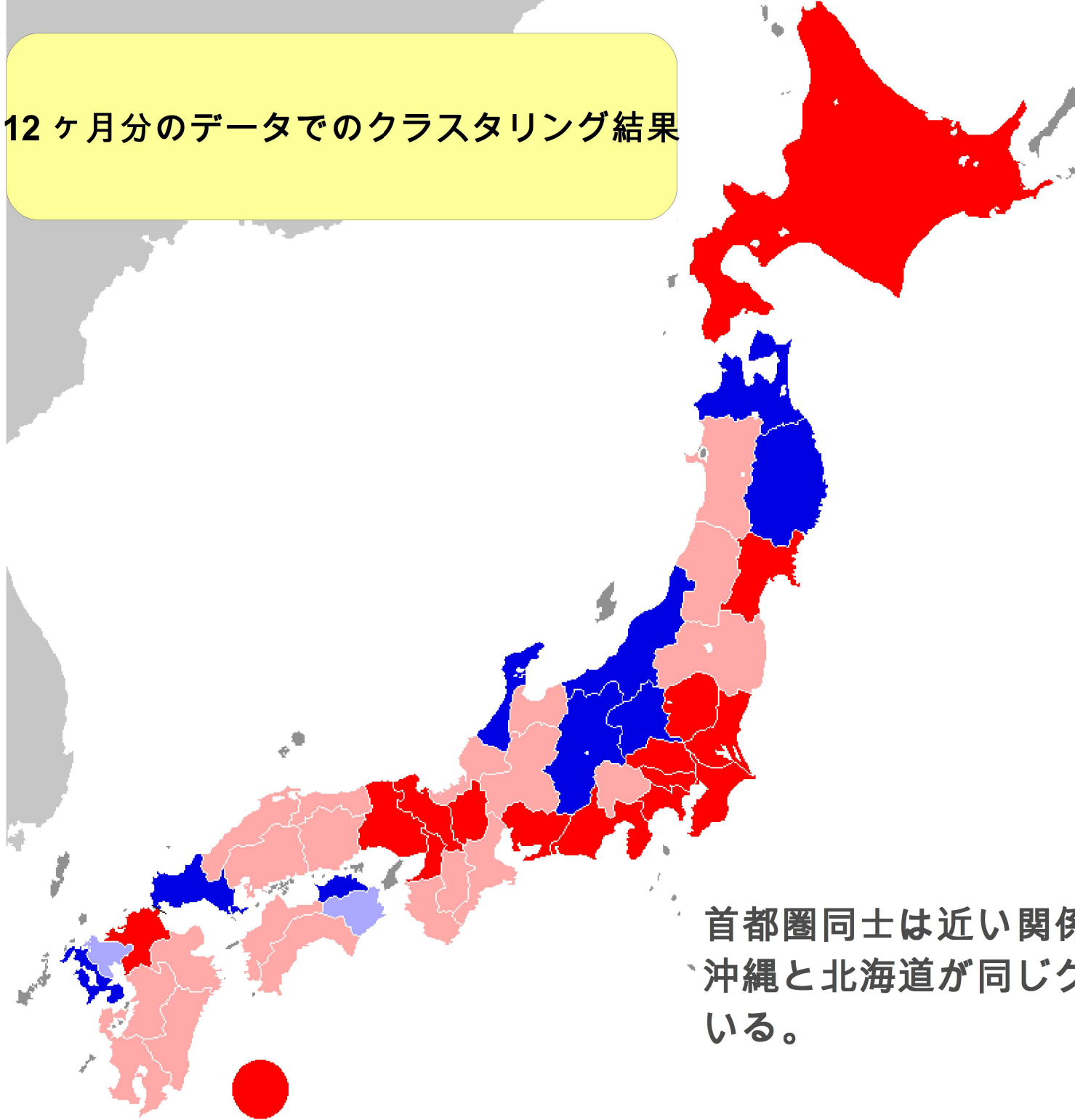


12ヶ月分のデータでのクラスタリング結果



12次元時系列データの ward 法による階層的クラスタリング

12ヶ月分のデータでのクラスタリング結果



首都圏同士は近い関係性となっている。
沖縄と北海道が同じグループとなっている。

まとめ

- ・ 地域毎のハッシュタグによる大域的な傾向を示した。
- ・ 首都圏では地域性が高くなる傾向にある。

今後の課題

- ・ 県 × 曜日 , 県 × 時間等のドリルダウン