VASA: mapping spatiotemporal structure of mobility patterns during the COVID-19 pandemic

Evgeny Noi, Alexander Rudolph, Somayeh Dodge

> GIScience 2021, Poznan, Poland (Virtual)

I. The problem

Introduction

- First and second order spatial effects violate assumptions of classic inferential statistic (independence and iid)
- Exploratory tools from spatial statistics and econometrics that help us assess phenomena while taking into consideration these spatial effects
    - Moran's I, Geary's C, Ripley's K, etc.
    - ~~Space-time data-mining and visual analytics~~

Problem

- Most common visual display in the geostack (PySAL, GeoDa, ArcGIS, QGIS, spdep, etc.) utilize choropleth maps, where statistically significant clusters (hotspots/coldspots) are denoted in blue/red.
- While this graphical display is ideally suited for purely spatial data, spatiotemporal and multivariate representations are problematic. How do we plot study areas with large number of spatial units?

II. The software 💾

VASA - Visual Analytics for Spatial Associations

- FOSS written on Python
- Ideally suited for exploratory spatial analysis on spatio-temporal areal data (e.g. Census units). Works with vector data.
- Offers several utility functions for spatial aggregation and visualizations

- Check out the Github repo
- Navigate example notebooks
- Install via pip (currently on test, alpha version)

```
pip install -i https://test.pypi.org/simple/VASA
```

III. The Case Study 🏞️

Assessing COVID-19 induced mobility in USA in 2020

- 2 data sources: *SafeGraph* (Open) & *Cuebiq* (Proprietary)
- 2 metrics: *%-sheltered* at home & *distance travelled*
- Approximately 3,200 counties over the period of 2020 (daily data)

General Workflow

0. Import data and clean
1. Create a VASA Object
2. Aggregate by spatial units
3. Downsample to Week
4. Calculate Local Moran's I for each week and county
5. Map *recency* | *consistency* | *week number* to visual variables
6. Plot on a graphic display

Create a VASA Object

```
v = VASA(
    df=data,
    gdf=counties_map,
    group_summary=lambda g: g[:2], # state level
    df_group_col="fips",
    gdf_group_col="GEOID",
    date_col="ds_m4",
    temp_res="week",
    date_format="%Y-%m-%d"
).group()
```

Impute missing values

Caution! VASA is very finicky about missing values!

```
v.pct_full_missing()
```

> array([0.03217503])

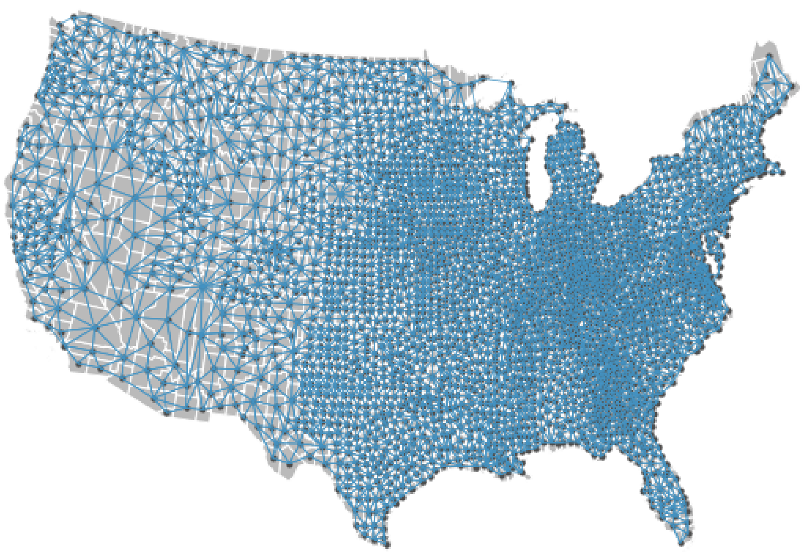
```
v.fill_missing()
```

> <VASA.vasa.VASA at 0x7f844422e250>

Build Spatial Weights Matrix

Since VASA utilizes Moran's I, we need to construct spatial weights matrix (SWM). There are different methods to create SWM (Rook's, Queen's, kNN, etc.). Please refer to PySAL docs to learn more.

```
v.show_weights_connection(k=3, type="union")
```



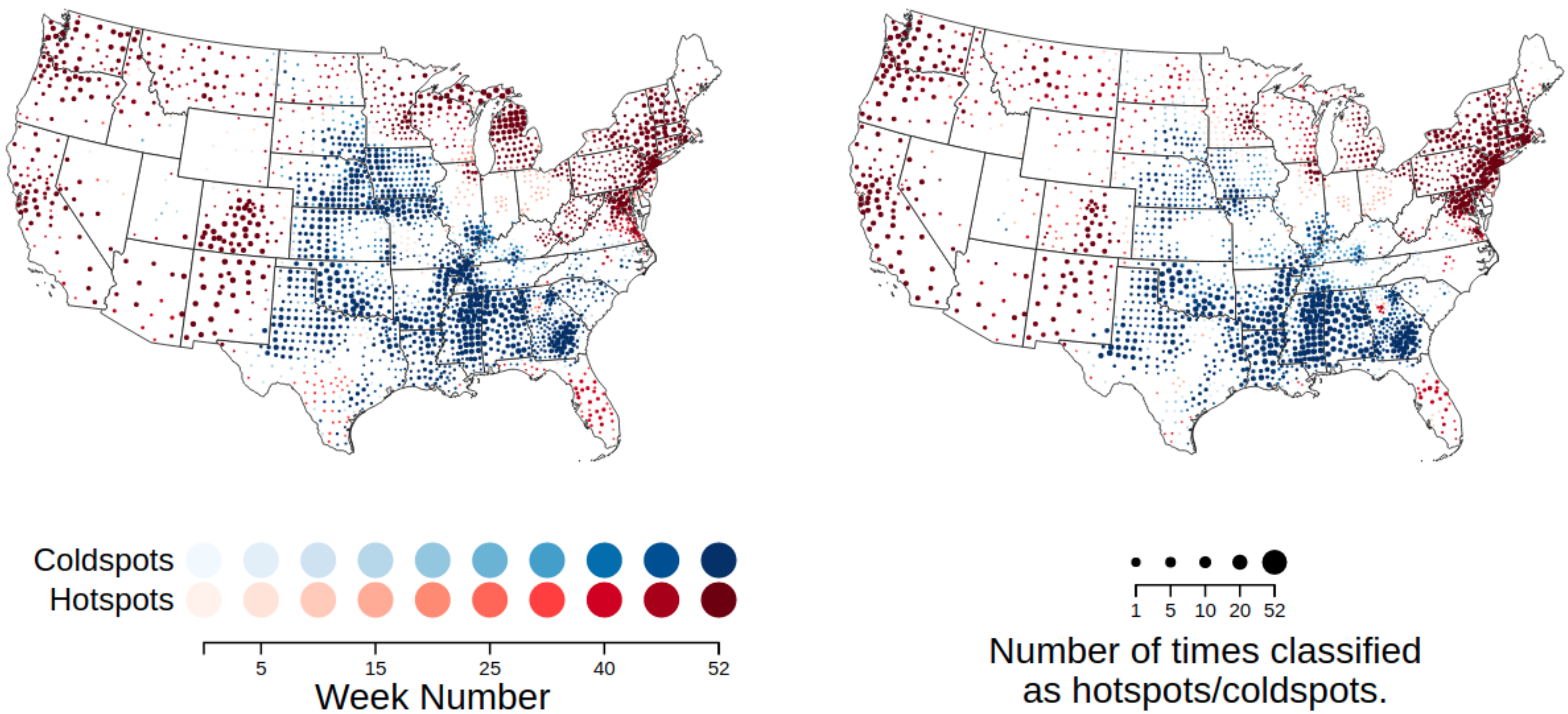
Recency and Consistency Map

> Better suited to study the aggregate (summary-level) details.

```
stacked.plot_both(a=2500, b=600)
```

hotspots:high values co-located with other high values.
coldspots: low values co-located with other low values.
Clustering consistency is illustrated by **size**.
Clustering recency is depicted by **color intensity**.

Percentage of Sheltered in Place (Cuebiq) | Percentage of Sheltered in Place (Safegraph)



Line-path scatter plot

> Better suited to study the finer (object-level) details.

```
from VASA.scatter import Scatter
sc = Scatter(v)
sc.plot(highlight="06") # this filters out CA
```

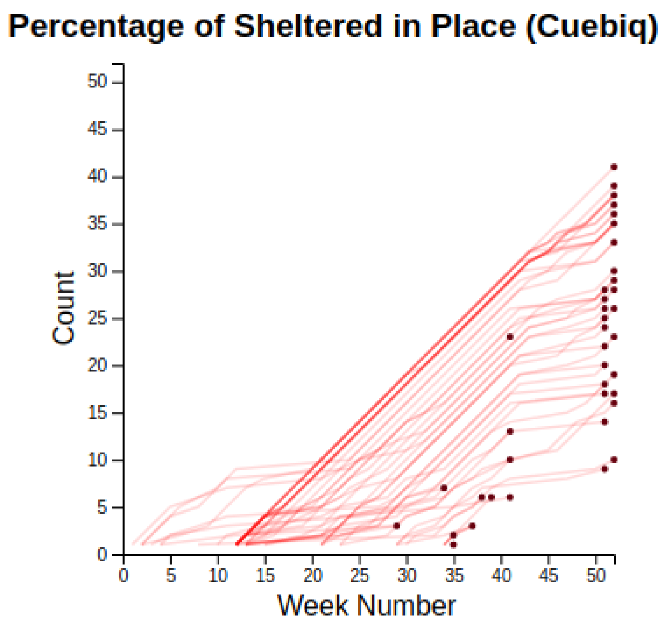
hotspots:high values co-located with other high values.

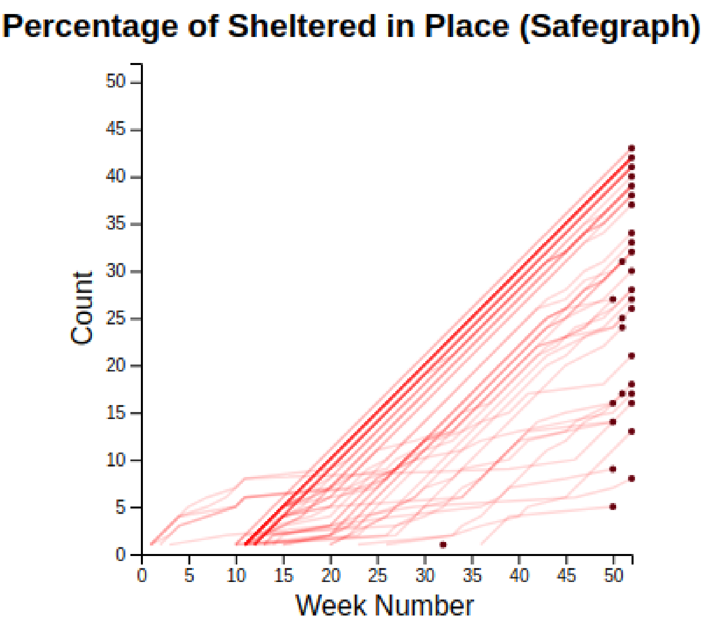
coldspots: low values co-located with other low values.

Week number is mapped onto X-axis.

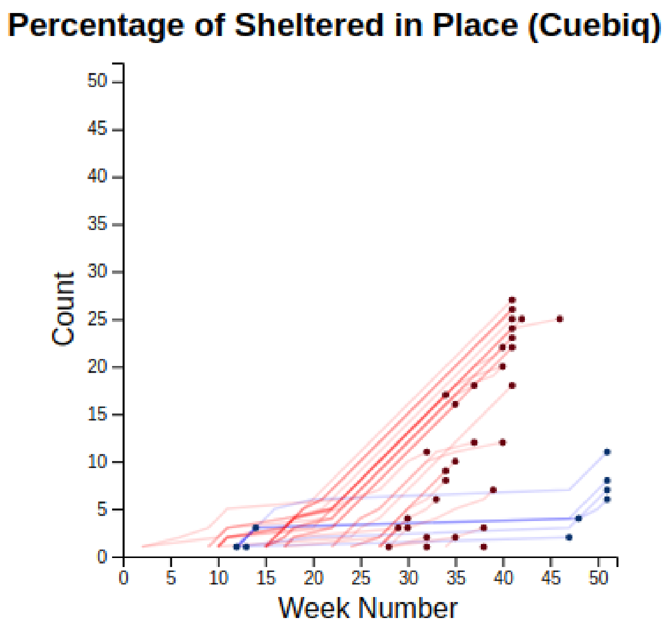
Number of times classified as hs/cs is mapped onto Y-axis.

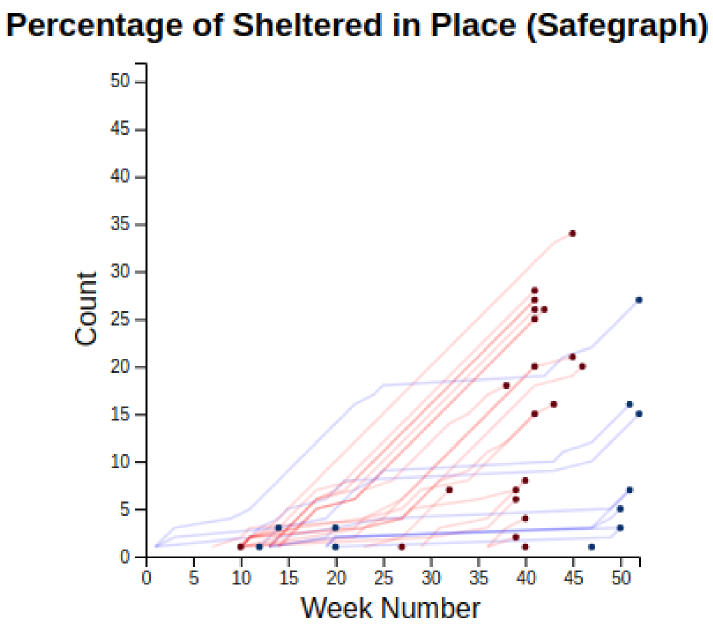
Social distancing in California





Social distancing in Florida





IV. Conclusion 🚀

- Proposed methodology provides a robust framework for assessing spatiotemporal structure of the data (NOT only mobility!).
- VASA allows exploratory spatial data analysis at the object-level and summary-level, which makes comparison of multiple spatiotemporal series easy
- VASA adds novel visual displays to traditional choropleth mapping prevalent in spatial econometrics
- *Planned: Adding more visual displays for micro-scale analysis on areal data (e.g. LA Census Block Groups ~ 25,000)*

Questions?

Thank you for Zoom-listening!

Evgeny Noi

Alex Rudolph

Somayeh Dodge

Connect with us

The authors gratefully acknowledge the support from the National Science Foundation Award BCS # 2043202 “CAREER: Modeling Movement and Behavior Responses to Environmental Disruptions”.