A **small-brief** Intro to Space-Time Data in R

Nabil Abdurehman
github.com/nabilabd
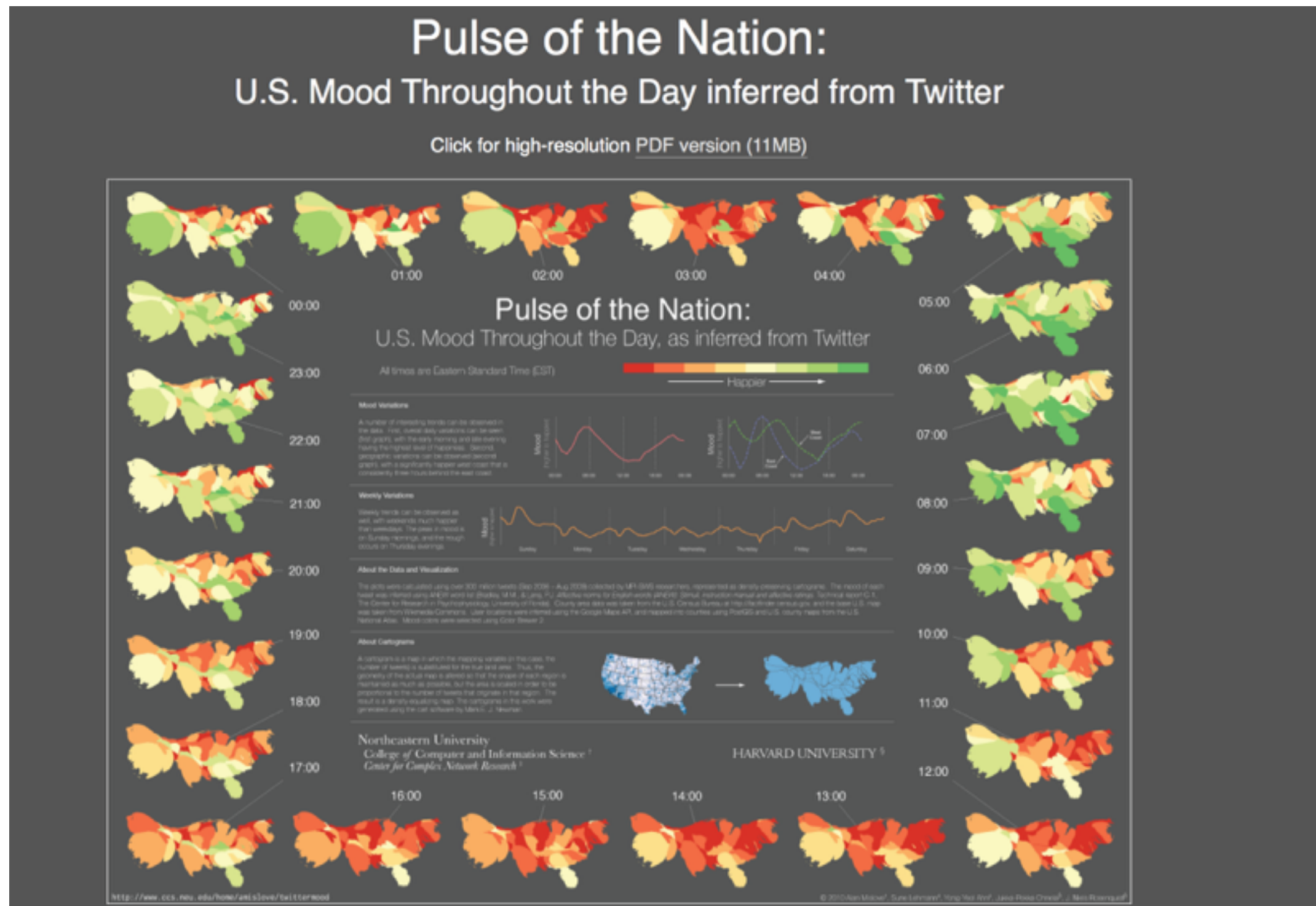
Who am I?

- B.A. in Math, Princeton '11

- Taught English in China

- Data Analyst at Tech Startup

- Finishing Masters in Statistics at Georgia Tech

Motivation

- Much data implicitly spatial (or, spatio-temporal)

  - Examples: Todd Schneider's "Analyzing 1.1 billion taxicab trips", air quality measurements
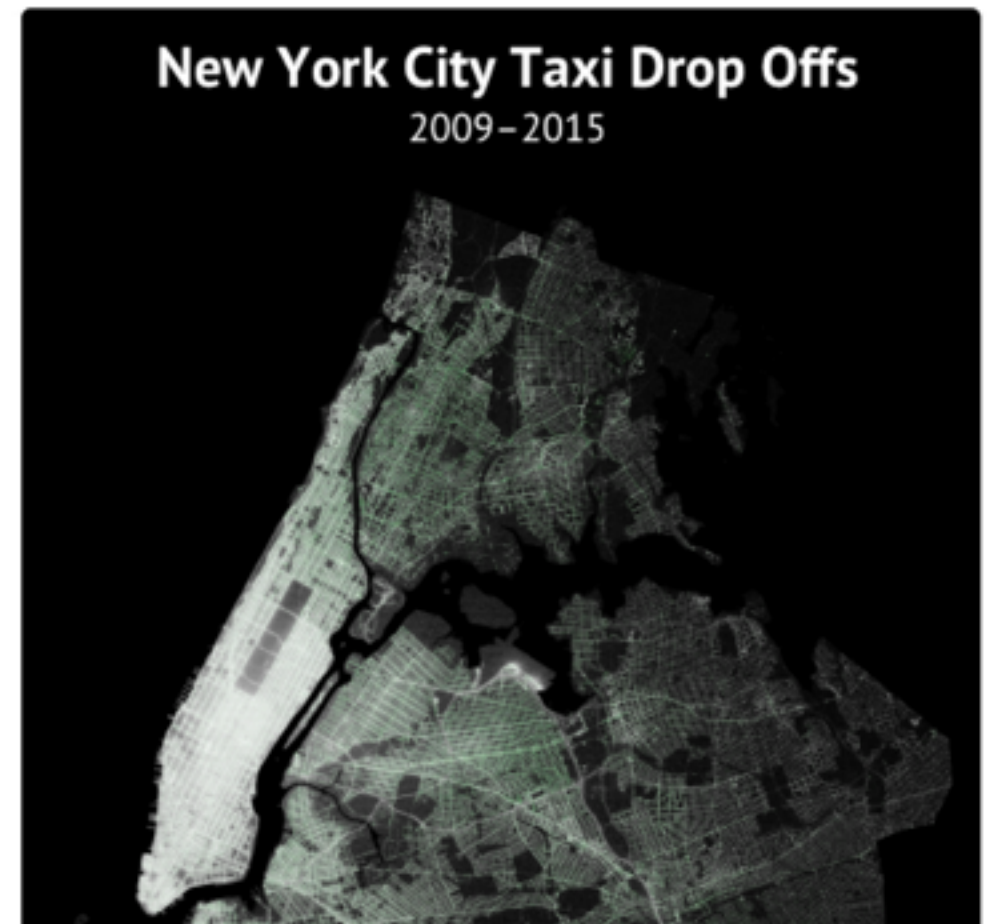
Inferring US Mood from Tweets

Taxi-Cab Trips

Todd Schneider Analyzed 1.1 Billion Taxi and Uber rides

Maps

I'm certainly not the first person to use the public taxi data to make maps, but I hadn't previously seen a map that includes the entire dataset of pickups and drop offs since 2009 for both yellow and green taxis. You can click the maps to view high resolution versions:

New York City Taxi Pickups
2009–2015

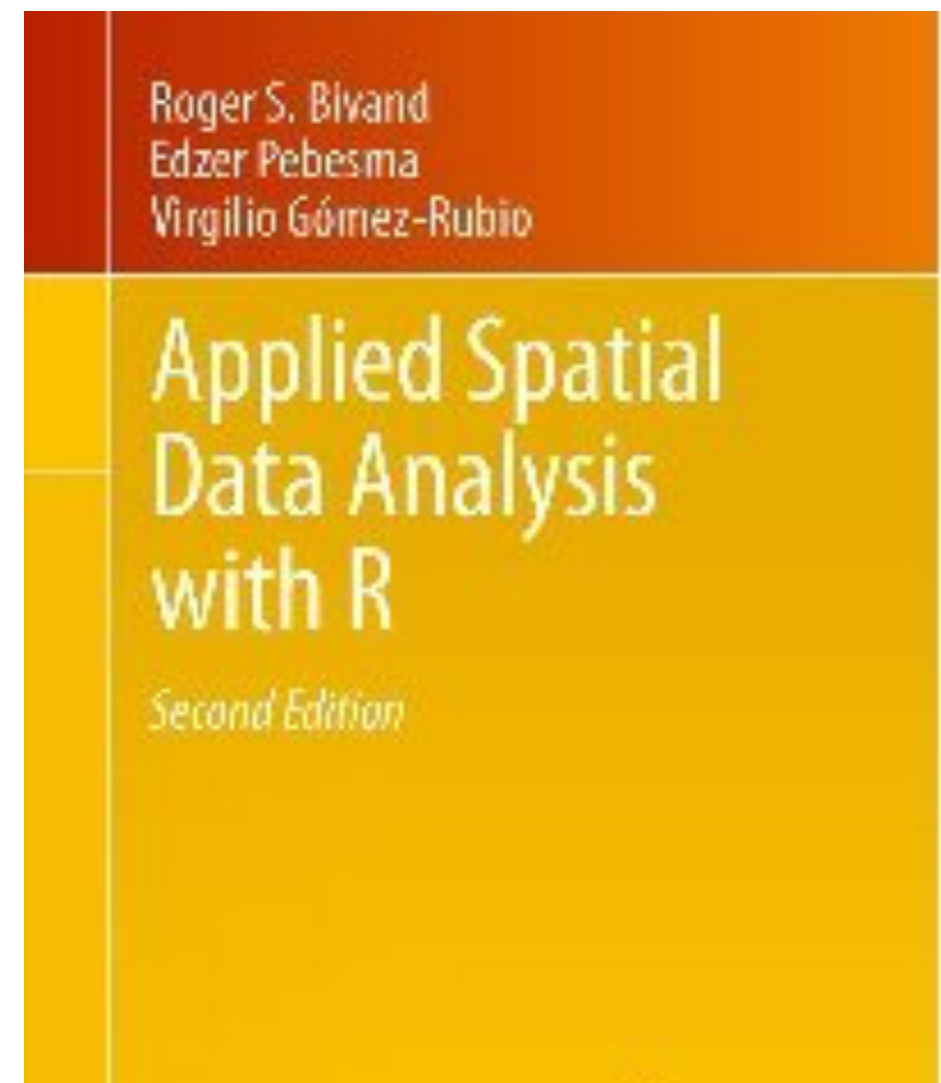New York City Taxi Drop Offs
2009–2015

Motivation

- But, how to identify if data inherently spatial?

  - Column names might be ambiguous

  - Which units meant?

- Want structures explicitly storing this information

R Spatial "Toolbox"

- Luckily, have R packages capable of handling spatial (or ST) data.

- Here, we focus on two:

    1. sp

    2. gstat

Reference ("ASDAR"):

Roger S. Bivand
Edzer Pebesma
Virgilio Gómez-Rubio

Applied Spatial Data Analysis with R

Second Edition

sp

Overview

- Published around 2003, by Edzer Pebesma and Roger Bivand

- Sought to standardize representation of spatial data

- Widely adopted by other packages

- (INSERT GRAPH HERE)

An Example

- We want to go over an example to get familiar with it, before going into more theoretical aspects of structure

- See: 1-data.R

Spatial Class Hierarchy

data type	class	attributes	contains
points	SpatialPoints	No	Spatial
points	SpatialPointsDataFrame	data.frame	SpatialPoints
multipoints	SpatialMultiPoints	No	Spatial
multipoints	SpatialMultiPointsDataFrame	data.frame	SpatialMultiPoints
pixels	SpatialPixels	No	SpatialPoints
pixels	SpatialPixelsDataFrame	data.frame	SpatialPixels SpatialPointsDataFrame
full grid	SpatialGrid	No	SpatialPixels
full grid	SpatialGridDataFrame	data.frame	SpatialGrid
line	Line	No	
lines	Lines	No	Line list
lines	SpatialLines	No	Spatial, Lines list
lines	SpatialLinesDataFrame	data.frame	SpatialLines
polygons	Polygon	No	Line
polygons	Polygons	No	Polygon list
polygons	SpatialPolygons	No	Spatial, Polygons list
polygons	SpatialPolygonsDataFrame	data.frame	SpatialPolygons

Source: Vignette in `sp` package

Spatial Class Hierarchy

- As you can see, there are many spatial classes, though. We've only dealt with one.

- How it's organized is in the abstract, with a top-down approach.

- "Spatial" is general, but doesn't contain any data. Depending on the kind of data the object can hold, is a different class.

- Some are used when, e.g., reading .shp files.

Accessing Properties

- The slots have helper functions (improving readability and safety)

- Note that in transforming to SPDF, the "data" portion loses out on

Building Objects Manually

- TODO

- See: 3-spatial_classes.R

Limitations

- Package designed to aid in organizing data

- Sometimes want to perform computations

- How to build off of that?

gstat

Overview

- Functions for modelling, plotting spatial and ST data

- Includes different interpolation routines (e.g., IDW)

- In many applications, one particular approach is common:

Kriging

- Named after Daniel G. Krige, South African professor

- Sought to identify estimate mineral deposits of gold

- Need to estimate spatial variability

(Semi-)Variogram

- WP: "function describing the degree of spatial dependence of a spatial random field or stochastic process"

- For this, need accurate measure of distance

- See: 4-variogram.R

Projections

- (Pictures with Mercator, others)

- Projections affect how distance is perceived or quantified. So, important to get right.

- Mercator not always sufficient, in real life or computation.

- See: 5-projections.R

Kriging, part 2

- Now, can measure distance, spatial variability, so can interpolate

- See: 6-kriging.R

Visualization

- With results, want to effectively communicate

- Different packages/tools for displaying results

- What are you trying to convey about your data?

- See: 7-visualization.R

Visualization

- Questions to ask:

  - Representation of discrete data: dense points or separate tiles?

  - Interactivity important?

  - Can the spatial data be meaningfully aggregated?

Application: Source Apportionment

- Particulate matter formed from different sources

- Want to quantify various contributions to PM mass

- One method incorporates simulated and observed concentrations of ambient chemical elements

- Once revised estimates are produced, can interpolate from point locations across the US

- For more: https://github.com/nabilabd/hybridSA

Alternatives

- When kriging can fail, inverse distance weighting can produce similar results

- For Bayesian methods, have different packages (e.g., spatstat). Can look into spTimer:

  - Pros: Good documentation (paper in JSS). Efficient code, can handle larger datasets. Implements recent Bayesian methods. ST interpolation in single step, not two.

  - Cons: Not deterministic (and so not reproducible), but rather slight randomness in different iterations.

- `spacetime` package for classes to deal with spatio-temporal data. Can use in conjunction with gstat (e.g., ST kriging methods, different ST variogram models)

  - NB: Recent (i.e., last couple of years) so documentation might be lacking, steep learning curve for the uninitiated

Acknowledgements

- Profs Ted Russell and Jim Mulholland. Cesunica Ivey. Derek Norton.