Notebook: Applying spatial and temporal filtering techniques.

Jul 10th, 2018

This is a notebook for exploring and implementing spatial/temporal filtering techniques introduced by such papers as Griffith (2013) and Hughes and Haran (2013). While `ngspatial` contains a proper full-fledged model framework introduced in Hughes and Haran (2013) and Hughes (2018), that code work will be separated as a raw code since it demands computational expense.

Some exploratory visualization work on the county-level areal dataset containing mortality rates is also done.

```
## -- Attaching packages --------------------------------------------------------------- tidyverse 1.2.1 --
## √ ggplot2 2.2.1     √ purrr   0.2.5
## √ tibble  1.4.2     √ dplyr   0.7.5
## √ tidyr   0.8.1     √ stringr 1.3.1
## √ readr   1.1.1     √ forcats 0.3.0
## -- Conflicts ------------------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
##
## Attaching package: 'maps'
## The following object is masked from 'package:purrr':
##
##     map
## Checking rgeos availability: FALSE
##    Note: when rgeos is not available, polygon geometry    computations in maptools depend on gpclib,
##      which has a restricted licence. It is disabled by default;
##      to enable gpclib, type gpclibPermit()
## Loading required package: Rcpp
## Loading required package: batchmeans
## batchmeans: Consistent Batch Means Estimation of Monte Carlo Standard Errors
## Version 1.0-3 created on 2016-07-03.
## copyright (c) 2012-2016, Murali Haran, Penn State University
##                          John Hughes, University of Colorado Denver
## For citation information, type citation("batchmeans").
## Type help(package = batchmeans) to get started.
## ngspatial: Fitting the Centered Autologistic and Sparse Spatial Generalized
## Linear Mixed Models for Areal Data
## Version 1.2-1 created on 2018-01-12.
## copyright (c) 2013-2018, John Hughes, University of Colorado Denver
## For citation information, type citation("ngspatial").
## Type help(package = ngspatial) to get started.
```

Data

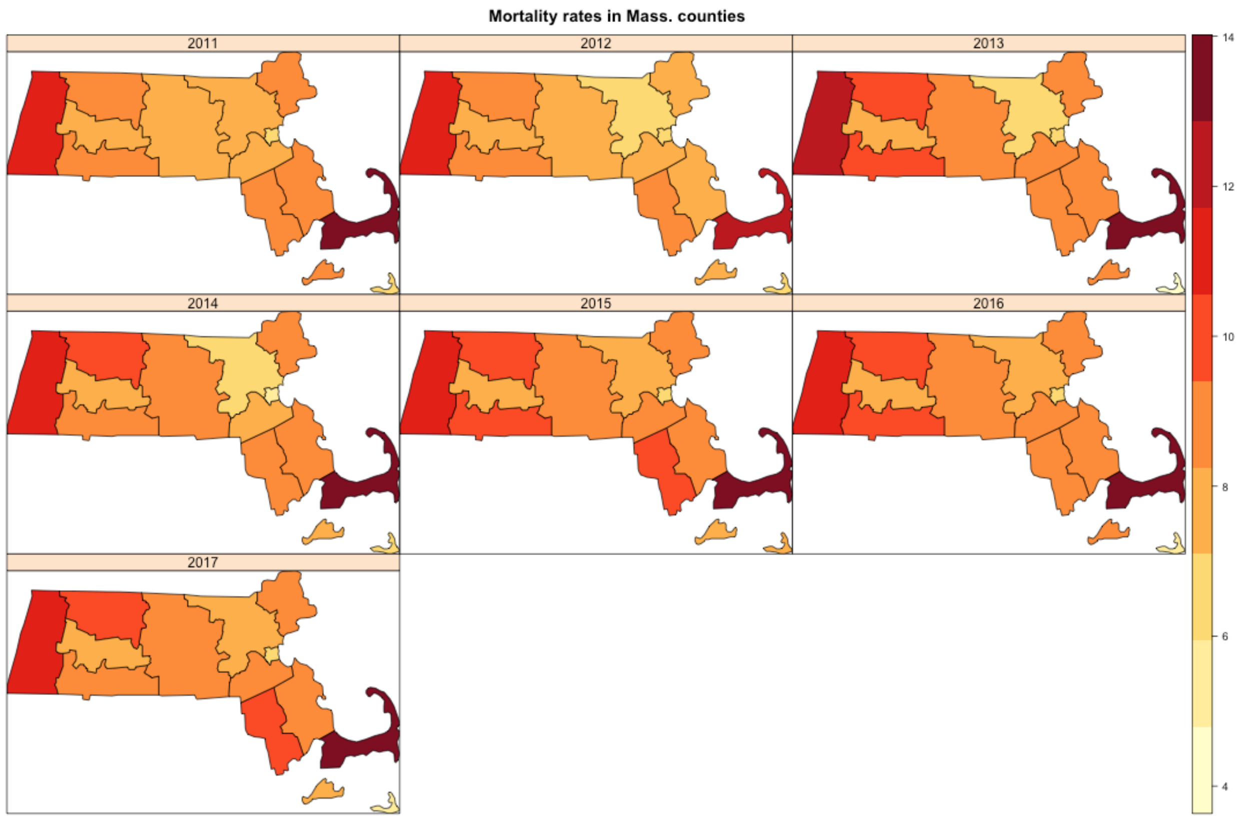
County-level datasets on Massachusetts state data, available from USDA ERS website (https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/). They consist of:

- Education: 10-year intervals from 1970 to 2000, and the 5-year average of 2012-2016.
- Population data: Available from 2010 to 2017.

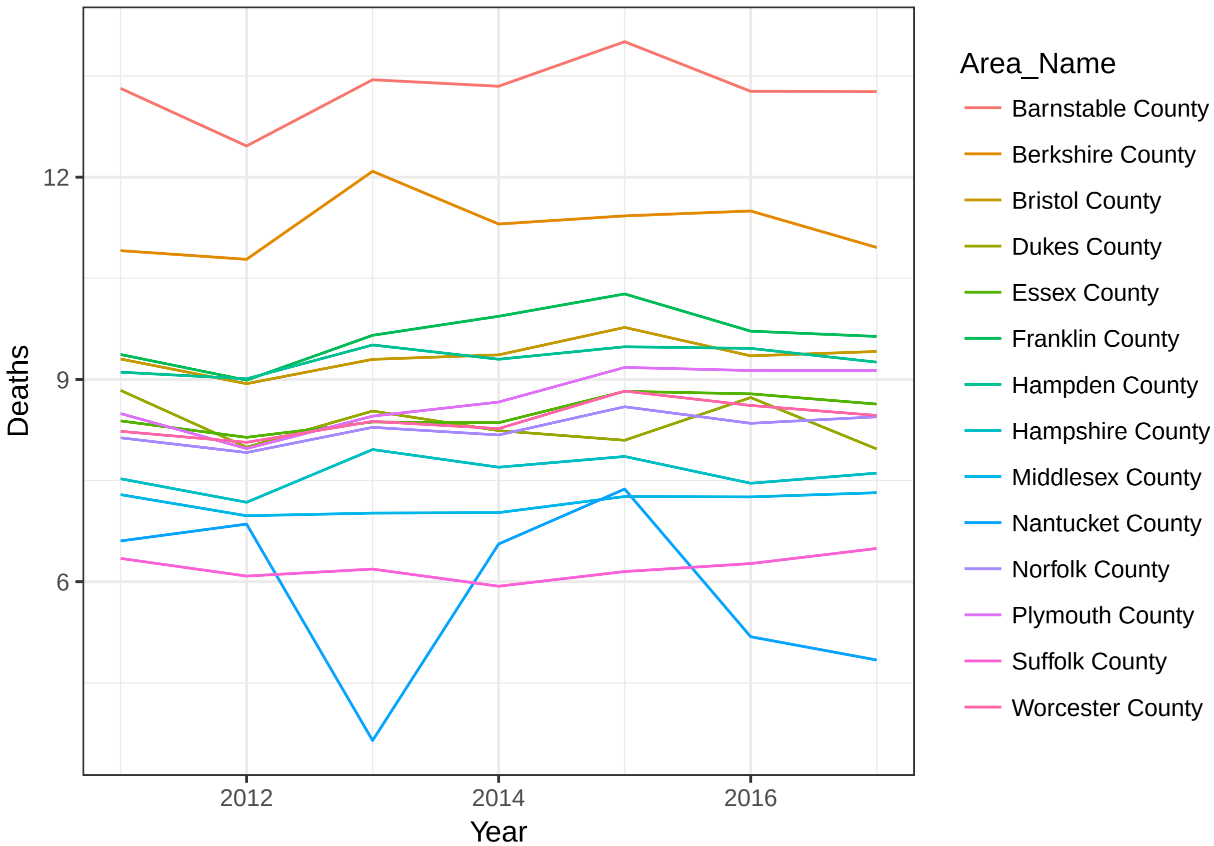
- Unemployment: Available from 2007 to 2017.

Here we focus on mortality rate as a response of interest. Time window is 2011-2017; area is here restricted to counties in Massachusetts for visual purposes.

```
## pdf
##   2
```

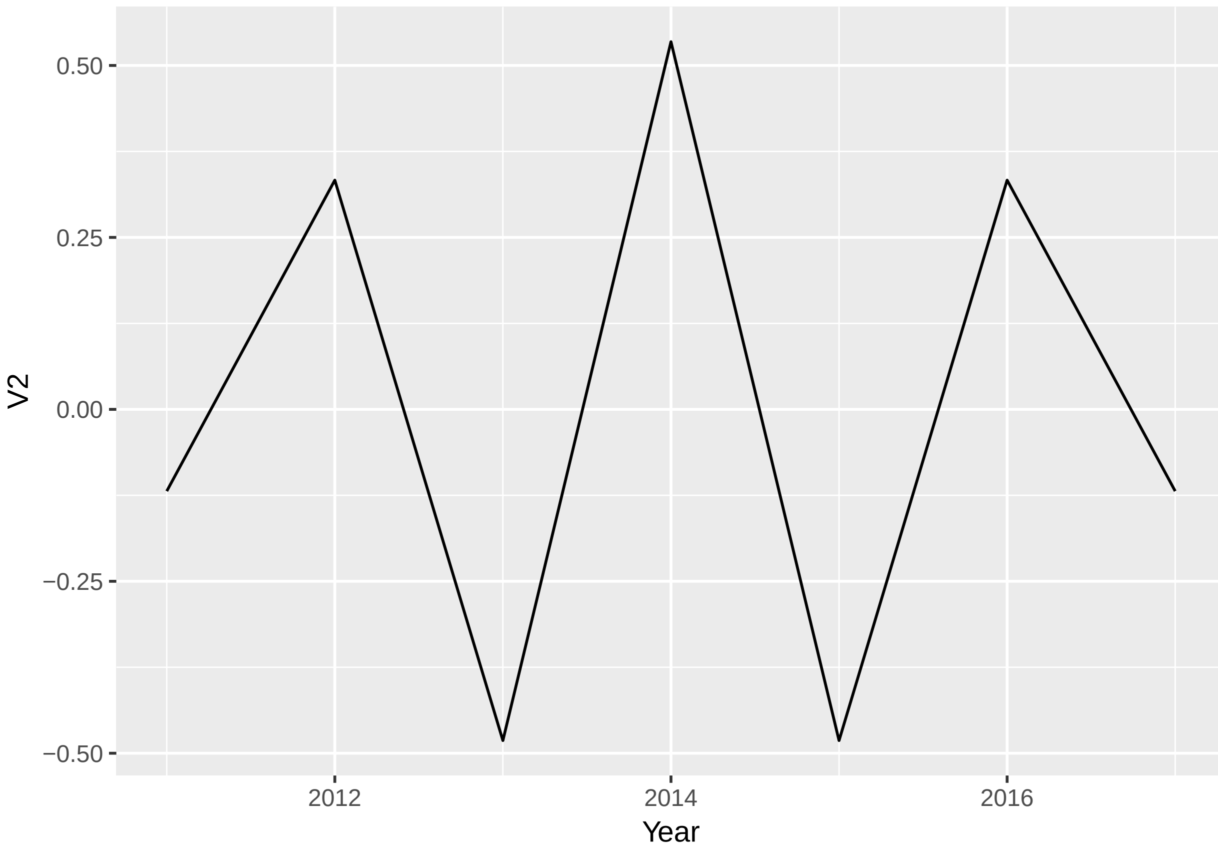


A time trend across years in different counties can be useful visualization to distinguish similar patterns and fit a simple model to separate a mean trend across locations.



Examining eigenvectors of time adjacency matrix (Griffith, 2010)

A neat visual explanation of the interpretability of filtering eigenvectors (with orthogonal restriction to the covariates) is given in Hughes and Haran, 2013. How, then, do we interpret and regress against the eigenvectors as obtained in a time series? Here we restrict our focus to mortality rates in Berkshire county (one spatial location across time). Since Griffith's model operates in a regular, fixed-effects regression setting, we need to separate possible trend element as a polynomial of time.

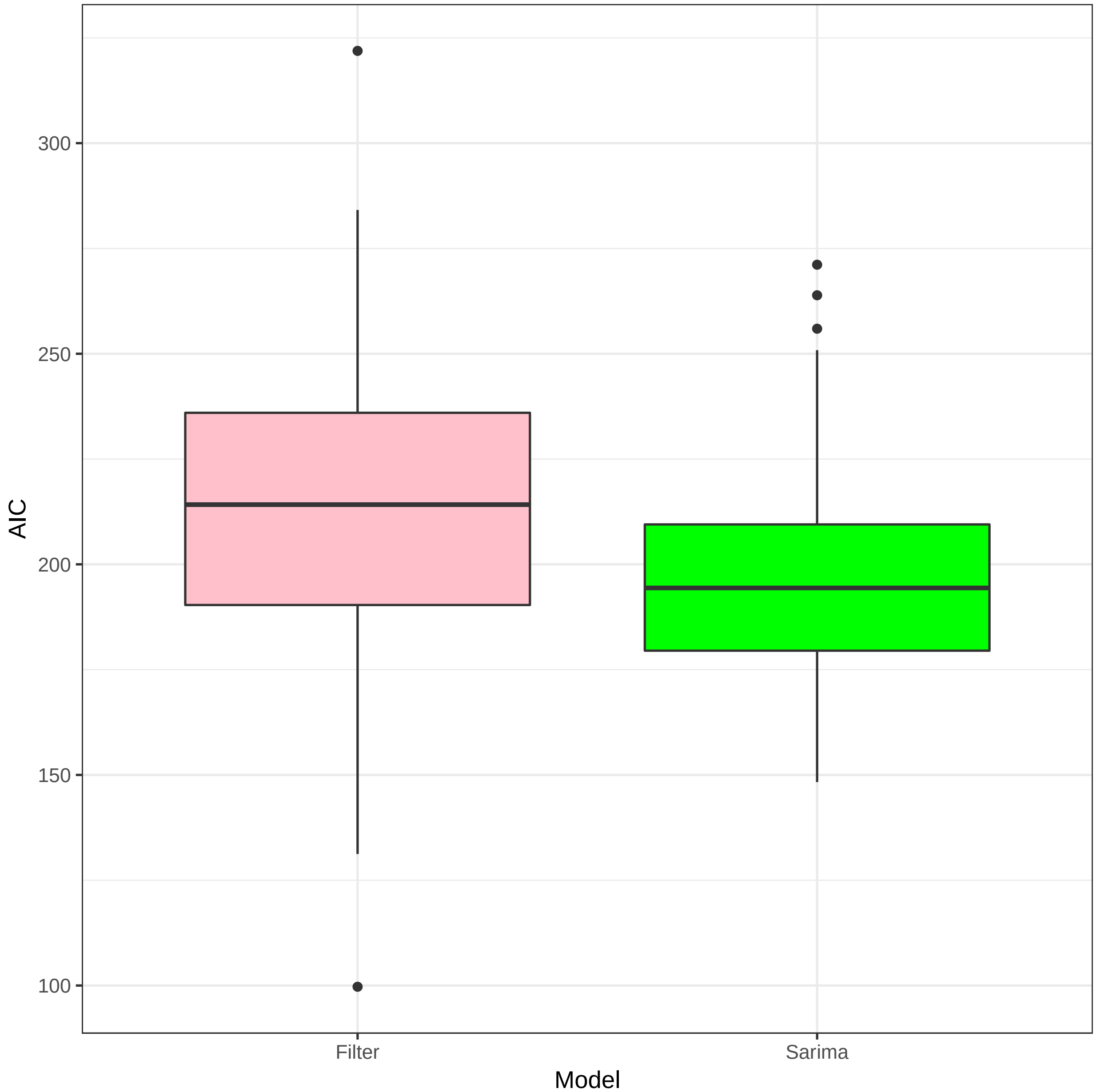


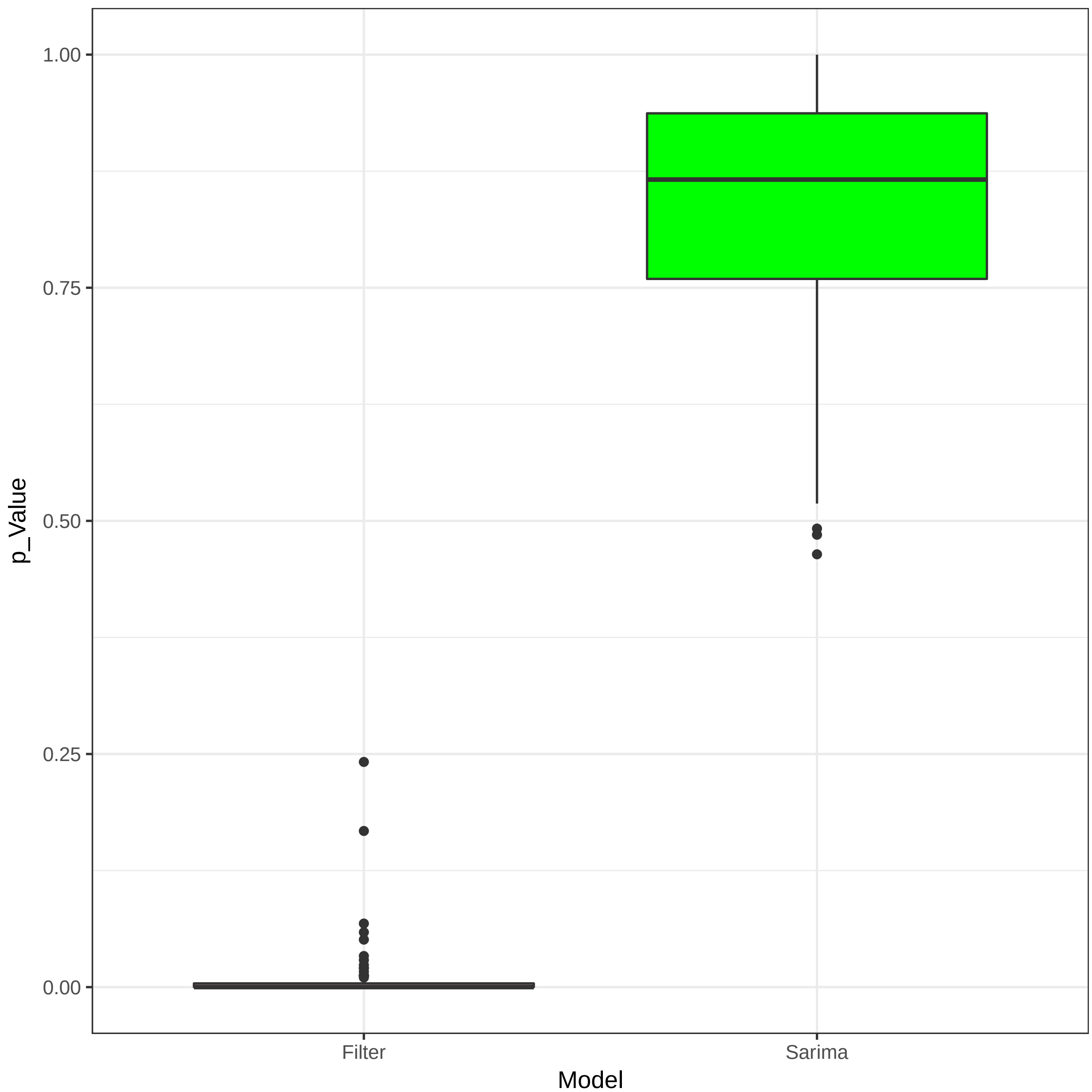
Can Griffith's version adequately capture long-lagged seasonal dependence structures? That is a question of acute interest.

Let us simualte 100 time series of varying shapes, which all have annual seasonlity (lag of 12) in its autoregressive component and random effects. For reference, an SARIMA model fitting that tries to account for period-12 seasonality is compared.

In the comparison, two important factors matter:

- How good is the fit? (We use AIC as the comparison criterion).
- How whitened is the residual process? (Perform Ljung-Box test).


300
250
200
150
100
AIC
Filter
Sarima
Model



Given that Griffith's proposed adjacency matrix does not allocate weights to each point based on the inherent properties of the time series, it is not surprising that the eigenfilter methodology performs poorly than a regular SARIMA fit, both in terms of model AIC and the residual goodness-of-fit test. In particular, almost all residual processes of eigenfilter model fit failed to pass the Ljung-Box test, which casts much doubt on the efficacy of the model methodology.

Thus, a concern regarding a more adequate structure of the adjacency matrix $\mathbf{A}_T$ that accounts for possibly complex seasonal dependence structure of a series arises.

Simulating spatiotemporal data

In the below simulation model, for example, the response is a linear combination of spatial coordinates $\mathbf{x}$ and $\mathbf{y}$. It has AR(1) dependence and a contemporaneous spatial dependence structure.

Some of the basic assumptions:

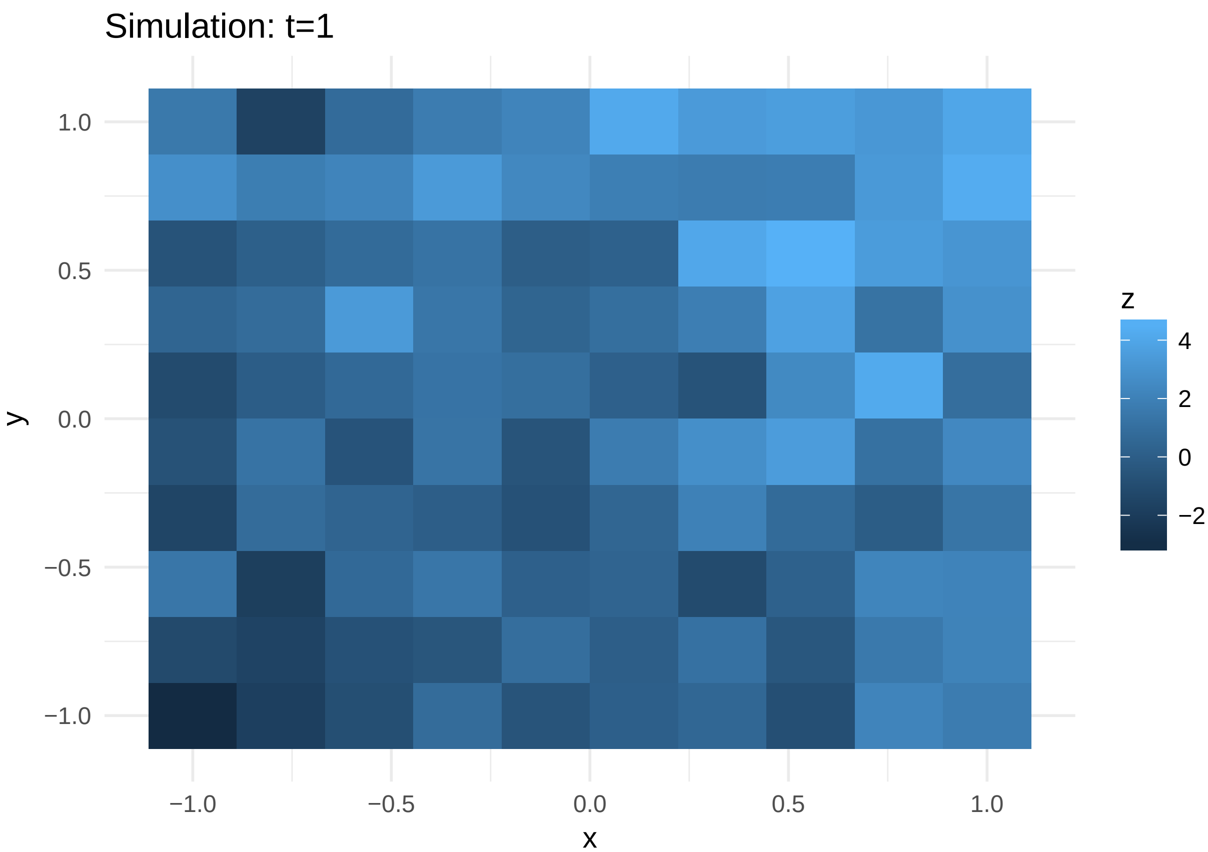
i) Spatial isotropy and stationarity.
ii) Temporal stationarity.
iii) Interaction between space and time.

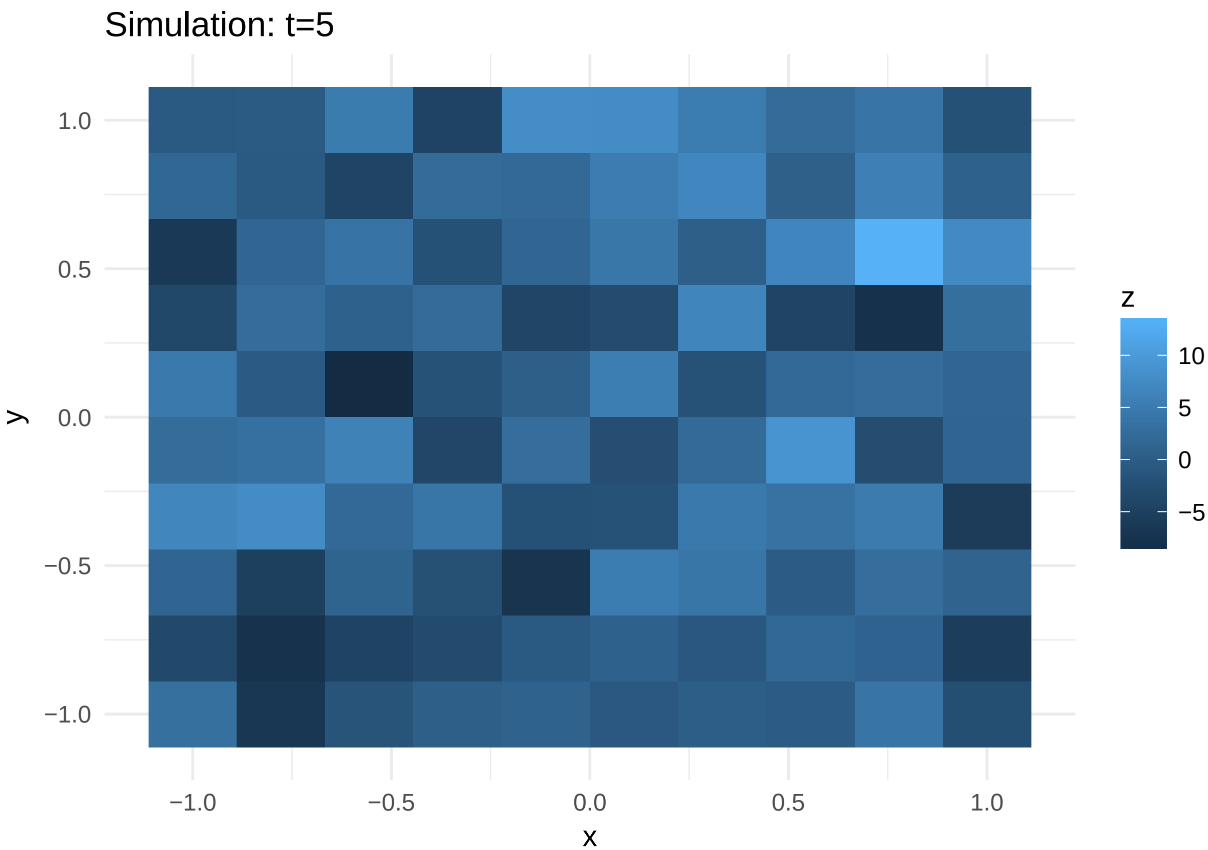
The time series is assumed to be causal; Z_1 is simply determined as the spatial mean plus gaussian error term. $\mathbf{s}$, which determines the interaction, is a complex, microscale spatial effect term that follows a Gaussian distribution with relatively small variance.

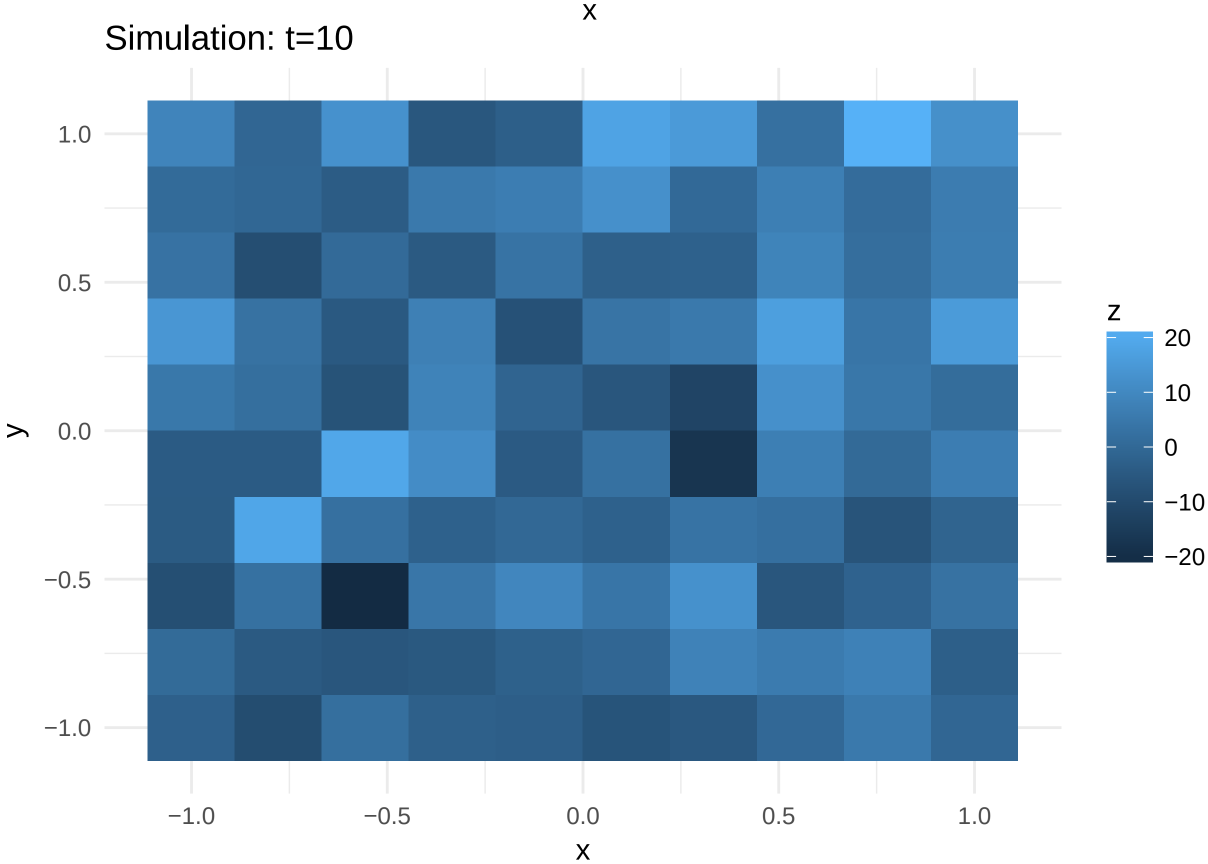
$$Z_t = (1 + 1.5\mathbf{x} + \mathbf{y} + t\mathbf{s}) + 0.1Z_{t-1} + \epsilon_t,\ \mathbf{s} \sim N(0, 0.5I),\ \epsilon_t \sim N(0, I),\ t > 1.$$

Model simulation for $n = 100$ areal units for $T = 10$ discrete time periods. Total number of simulated responses are 100×10 that follow a known gaussian distribution.

Different spatial heatmaps resulting from the simulation are plotted below. These are snapshots of responses at time periods $t = 1, 5, 10$.







Pattern-wise, we may be able to specify functional response curves from neighboring locations as being “closer” to each other. Interesting question may involve: **clustering**.