

lydemapr: an R package to map *Lycorma delicatula*

Sebastiano De Bona*

Matthew R. Helmus[†]

18 March 2024

Introduction

The Spotted lanternfly (*Lycorma delicatula*, White 1841) is an agricultural pest native of China and Southeast Asia, first discovered in the United States in 2014 in Berks County, PA. Since then, this planthopper has spread throughout the Mid-Atlantic and Midwest regions of the country, threatening the wine and fruit industry and damaging ornamental trees.

Since its first discovery, many sources have collected data on the presence/absence and population density of this species in order to monitor its spread and impact. The **lydemapr** package contains two anonymized datasets (at 1 km² and 10 km² resolution) resulting from an effort to combine, organize, and aggregate all available sources of data. In addition, this package contains useful functions to visualize the data within R.

The **lydemapr** package was built with the intent to increase accessibility to key data on this species of interest, and to improve reproducibility and consistency of modeling efforts.

We are constantly looking to expand the data sources to have a full representation of SLF's presence and abundance in the US. If you wish to contribute to this effort please contact the package authors.

Data Summary

To begin, let's take a look at the data structure:

source_agency	year	bio_year	latitude	longitude	state	lyde_present	lyde_established	lyde_density
iNaturalist	2015	2015	40.41	-75.66	PA	TRUE	FALSE	Unpopulated
iNaturalist	2016	2016	40.33	-75.64	PA	TRUE	FALSE	Unpopulated
iNaturalist	2016	2016	40.36	-75.48	PA	TRUE	FALSE	Unpopulated
iNaturalist	2016	2016	40.37	-75.62	PA	TRUE	FALSE	Unpopulated
iNaturalist	2016	2016	40.38	-75.72	PA	TRUE	FALSE	Unpopulated

Each data point contains information on its source and specific dataset of origin ("source_agency"). The data is organized by year (specified as both calendar "year" and "bio_year", running from May 1st to April 30th), coordinates, and state. Additional columns define whether SLF was found during the survey in that location (even as an anecdotal individual record, "lyde_present"), whether an established population was found there ("lyde_established"), and what the estimated population density of SLF was there ("lyde_density"). For additional information on the variables included, please consult the help file associated with the data by typing `?lyde` in the RStudio console. A Metadata file can also be found in the compressed folder **lyde_data.zip** contained in **download_data/**.

*Temple University, sebastiano.debona@gmail.com

[†]Temple University, mrhelmus@temple.edu

The package function `lyde_summary()` breaks the data down into a quick summary, with data organized by different axes. We can take a look at the data split across year and States. It's important to notice that the data is arranged yearly according to the **biological year** of SLF, and not calendar year. This allows for the appropriate inclusion of egg masses discovered during the winter months which were laid during the previous calendar year's summer/fall.

	2014	2015	2016	2017	2018	2019	2020	2021	2022
AZ	0	0	0	0	0	10	139	120	197
CA	0	0	0	0	0	0	0	0	1
CT	0	0	0	0	0	3	2081	1442	1436
DC	0	0	0	0	8	21	10	5	0
DE	0	0	0	0	1075	2208	4546	6962	5175
FL	0	0	0	0	0	0	0	1	0
IN	0	0	1	0	79	101	102	352	5
KS	0	0	0	0	0	0	0	21	0
KY	0	0	0	0	0	3	2	20	12
MA	0	0	0	0	0	0	893	2859	2056
MD	0	0	0	1	39	2398	17408	4734	1470
ME	0	0	0	0	0	0	0	20	85
MI	0	0	0	0	0	0	1	133	307
MO	0	0	0	0	0	15	18	0	0
NC	0	0	0	0	0	14067	5	110	86
NH	0	0	0	0	0	0	0	0	60
NJ	0	0	0	0	2443	9528	13075	83704	39397
NM	0	0	0	0	0	0	10	28	26
NY	0	0	0	0	18474	27047	18228	9063	19180
OH	0	0	0	0	0	0	681	575	304
OR	0	0	0	0	0	0	92	15	73
PA	372	7677	9269	9231	77055	150181	90476	69436	79699
RI	0	0	0	0	0	0	45	18	285
SC	0	0	0	0	0	2	7	49	70
TN	0	0	0	0	0	0	0	1	0
TX	0	0	0	0	0	0	0	0	75
UT	0	0	0	0	0	0	1	0	0
VA	0	0	0	2	1523	4353	4100	2574	2748
VT	0	0	0	0	0	0	0	2	25
WV	0	0	0	0	3	995	2368	2101	1450

Maps of the Spread of SLF

Two functions allow the user to plot the data: `map_spread()` and `map_yearly`.

The first function produces a snapshot of the SLF spread in the United States, with reference to the sampling effort associated with surveying the spread. Surveys finding an established population are plotted on the map as filled tiles, color coded by the year of first discovery. Surveys finding no established population are plotted as grey tiles.

The default function displays data aggregated at the 10km² (Figure 1). The function can be customized to show the data at higher spatial resolution (1k²), by setting the function option `resolution` to "1k". This will take considerably longer, so saving the result as a pdf is preferable in this instance as well.

The function displays data in a slightly different fashion at the 1km² resolution (Figure 2). At 10km² the data is plotted at filled tiles. This improves the visualization by representing the grid in which the data is

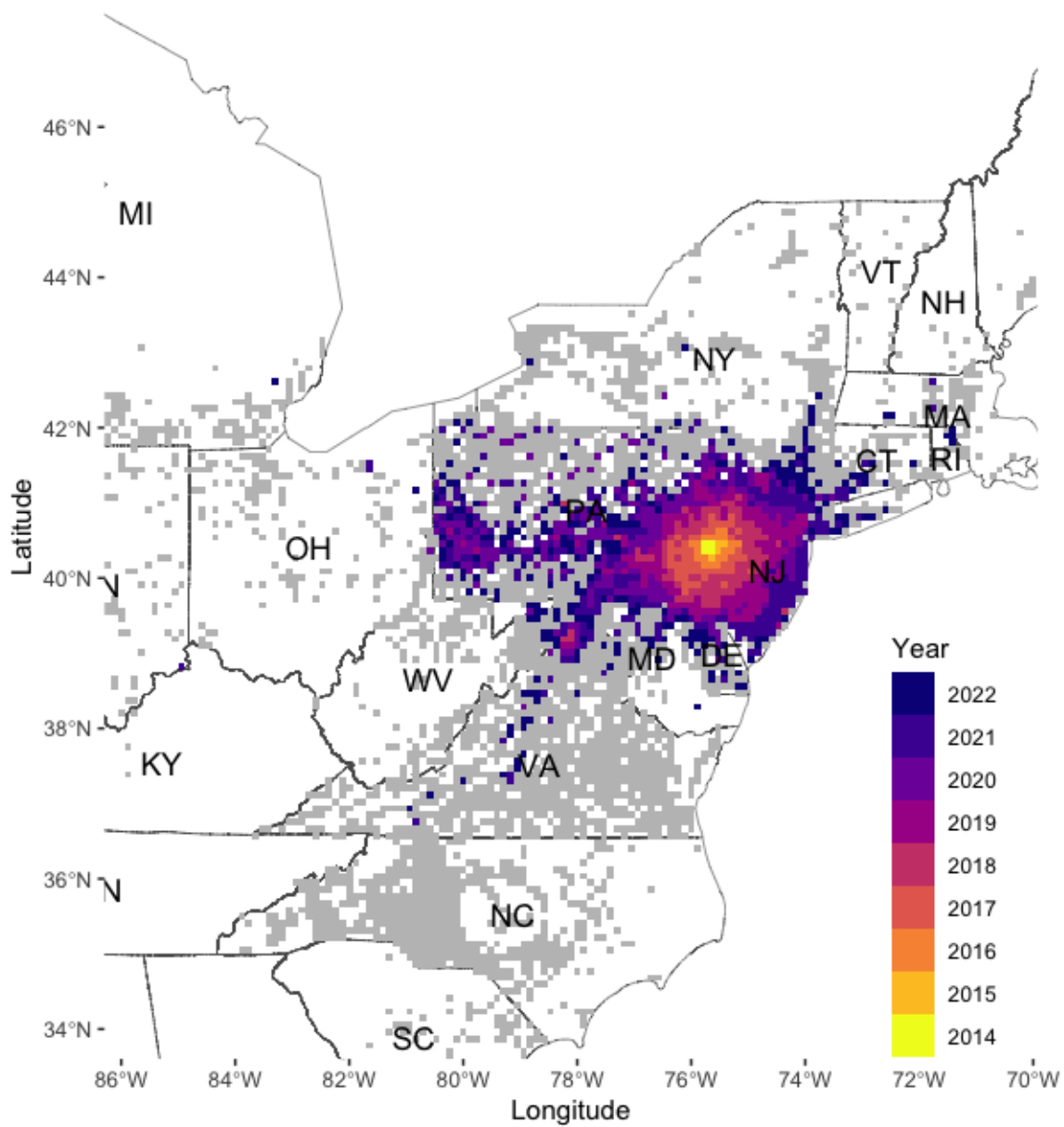


Figure 1: Output of the `map_spread()` function, plotted at the 10km resolution

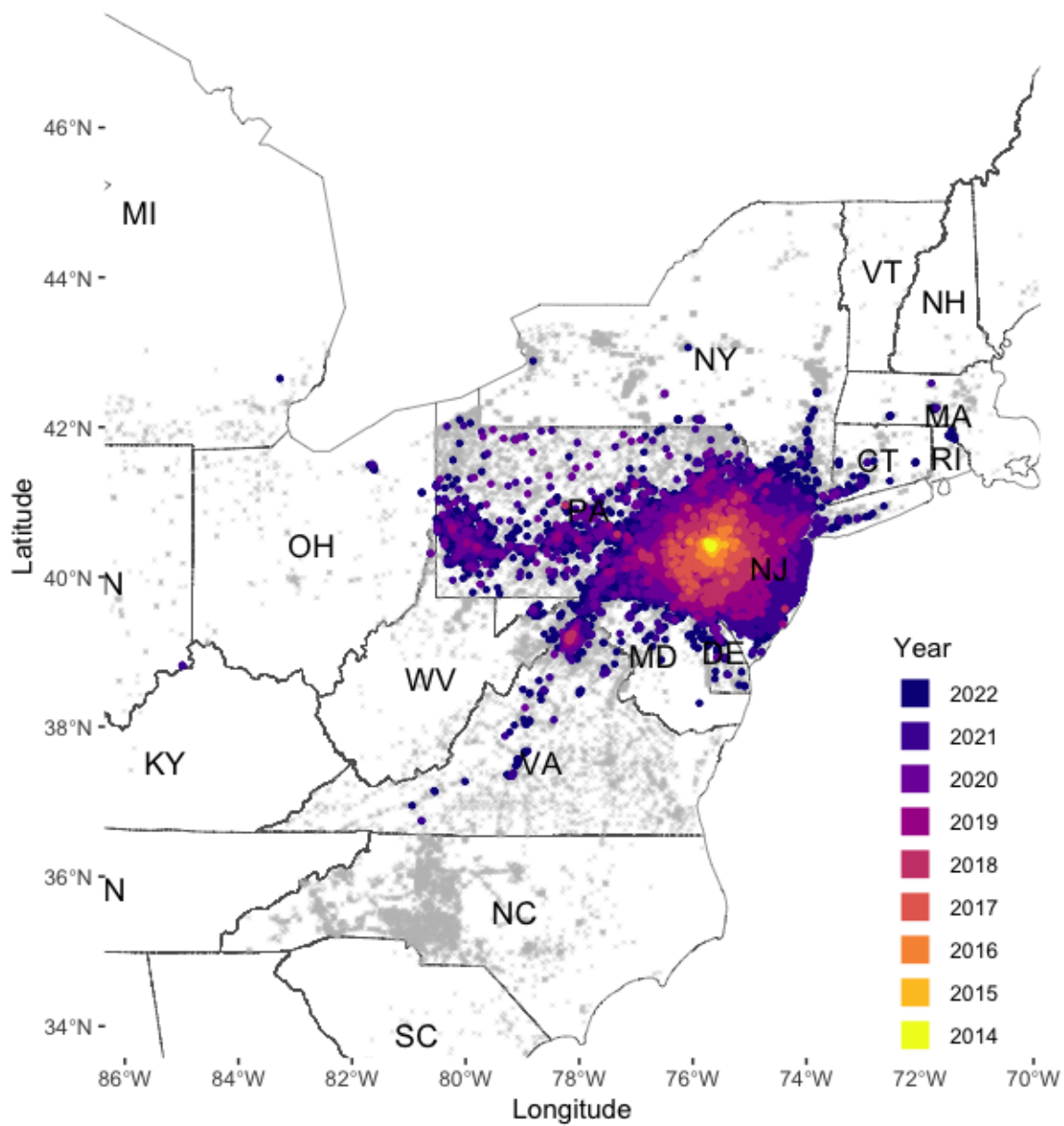


Figure 2: Output of the `map_spread()` function now plotted at a finer 1km resolution

organized more clearly. As tiles of size 1km are much smaller, we prefer to display survey points at this resolution as points on the map.

If the user wishes to visualize the data for a smaller area of the United States, the function allows them to specify which area should be mapped, by setting the `zoom` variable to “custom” and specifying the boundaries of the mapped area through `xlim_coord` (longitude) and `ylim_coord` (latitude), as Longitude and Latitude coordinates using the WG84 projection. Here’s an example of how this can be achieved.

The second function, `map_yearly()` allows the user to visualize the progression of SLF establishment, with a focus on the estimated population density through time. Note that the data here is not cumulative, meaning only data from a given year is shown in any given panel of the figure.

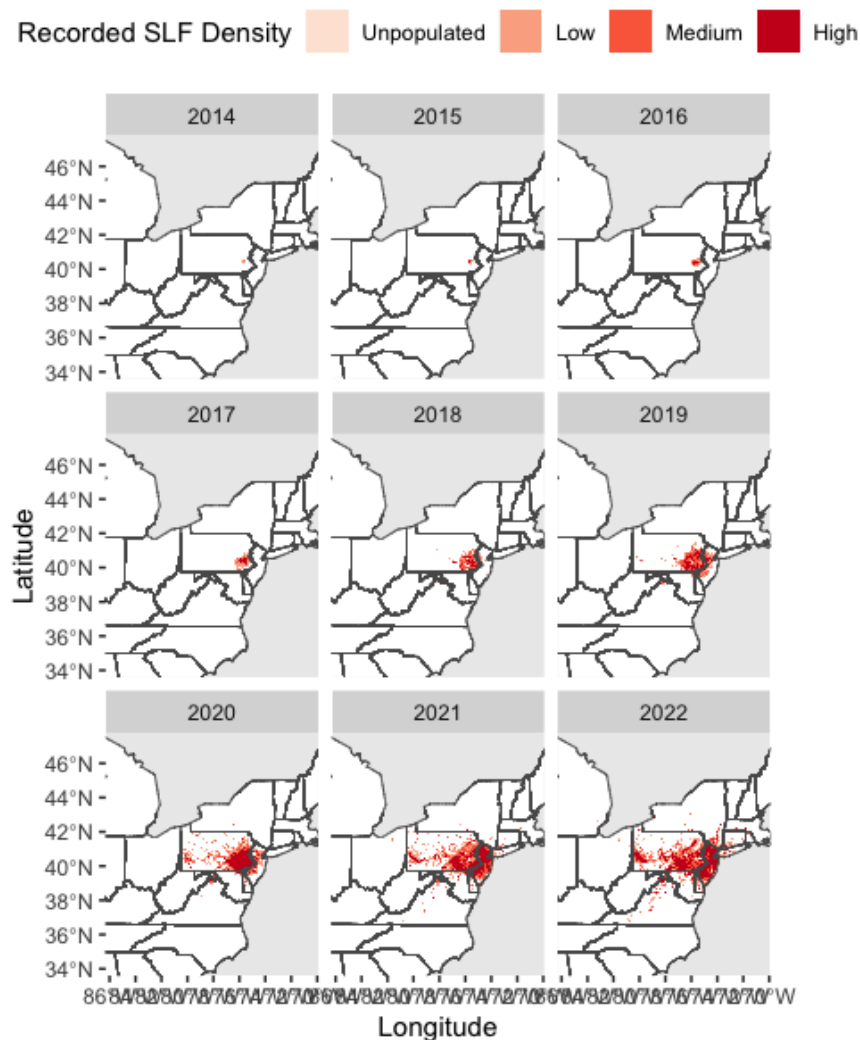


Figure 3: Yearly map of recorded population density