HMW2

Ivan Lopez

5/21/2022

Objectives.

- We would like to estimate the causal effect of using nitrogen on corn yield.
- We could get biased estimates if we worked with observational data only.
- Indeed, the farmers using nitrogen could be those who already know that their land is very productive.
- To rule out this possibility suppose we ran a randomized experiment on an 80-acre field. Moreover, such a procedure generates data on nitrogen doses (nitrogen rate) and yield measures.
- Finally, we have data on electric conductivity, a force that can interact with the nitrogen-based treatment. Thus, we must control electric conductivity (EC) to measure how it affects affects the marginal impact of nitrogen on corn.

Spatial Procedures

- read spatial data in various formats: R data set (rds), shape file, and GeoPackage file
  - use `sf::st_read()`
- create maps using the `ggplot2` package. Here, you will need to replace the use of tmap with `ggplot2`.
  - use `ggplot2::geom_sf()`
- create subplots within experimental plots
  - user-defined function that makes use of `st_geometry()`
- identify corn yield, as-applied nitrogen, and electric conductivity (EC) data points within each of the experimental plots and find their averages
  - use `sf::st_join()` and `sf::aggregate()`

Preparation for replication

- Run the following code to install or load (if already installed) the `pacman` package, and then install or load (if already installed) the listed package inside the `pacman::p_load()` function.

```r
# if (!require("pacman")) install.packages("pacman")
# pacman::p_load(
#   sf, # vector data operations
#   dplyr, # data wrangling
#   ggplot2, # for map creation
#   modelsummary, # regression table generation
#   patchwork # arrange multiple plots
# )
```

- Run the following code to define the theme for maps:

```r
# theme_for_map <-
#   theme(
#     axis.ticks = element_blank(),
#     axis.text = element_blank(),
#     axis.line = element_blank(),
#     panel.border = element_blank(),
#     panel.grid = element_line(color = "transparent"),
#     panel.background = element_blank(),
#     plot.background = element_rect(fill = "transparent", color = "transparent")
#   )
```

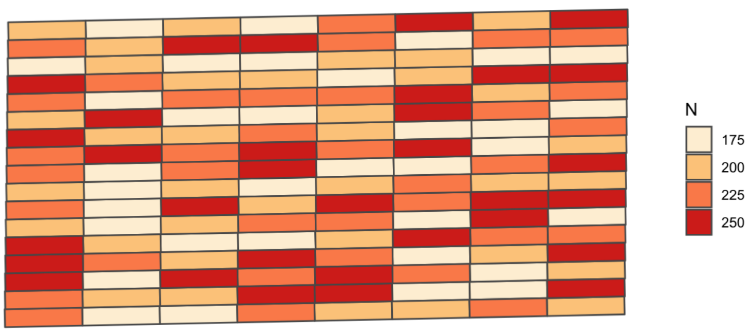
Instructions

- Read in `trial_design.rds` .
- The above dataset contains the nitrogen rates across aerial units. Please create a map showing the spatial distribution of the variable `NRATE` .
- Now, read in the information concerning the "as applied nitrogen" (`NH3.gpkg`), "electric conductivity" (`ec.shp`), and "collected yield." (`yield.rds`) . To do so, you will have to eomply `st_read()` and `readRDS` .
- Create a figure showing the spatial distribution of the three variables. First create a map of each variable, and then combine them into one figure using the `patchwork` package
  - To stack the figures vertically, label each individual figure as `g_yield`, `g_NH3`, and `g_ec`.
  - Then write: g_yield / g_NH3 / g_ec
- Instead of using `plot` as the observation unit (the term "plot" refers to a piece of land), we would like to create subplots inside each of the plots and make them the unit of analysis. This will avoid hiding the within-plot spatial heterogeneity of EC. Please divide each plot into six subplots.
- Create a figure displaying the subplots.
- Now identify the mean value of corn yield, nitrogen rate, and EC for each of the subplots using `sf::aggregate()` and `sf::st_join()`.
- Provide a visualization of the subplot-level data.
- Now you should be able to run a regression model.

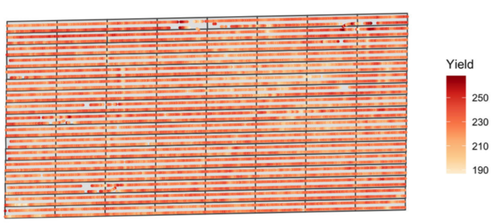
```r
# ols_res <- lm(yield ~ aa_NH3 + I(aa_NH3^2) + I(aa_NH3 * ec) + I(aa_NH3^2 * ec), data = reg_data)
#
# modelsummary(
#   ols_res,
#   stars = TRUE,
#   gof_omit = "IC|Log|Adj|Wamenableudo"
# )
```

- Please provide an interpretation of the above results and draw a DAG explaining why such results are amenable to causal interpretation.

Below you can find a guide for the outputs you must produce:

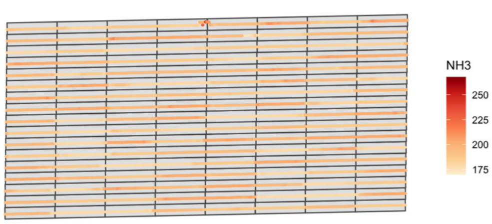


Randomized Experiment.

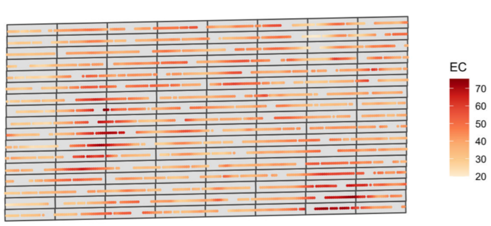


Spatial dbn of:

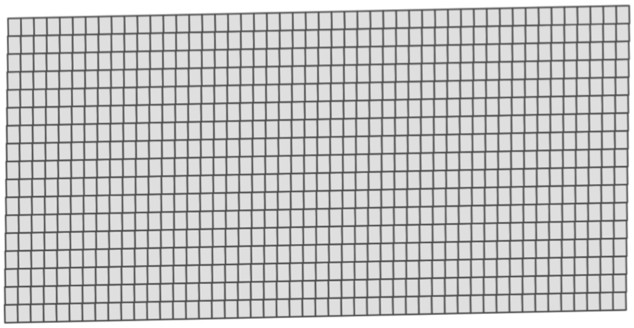
Yield



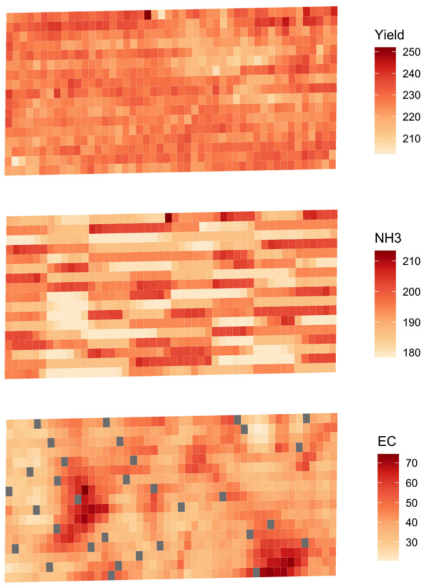
NH3



EC



Map of subplots



Spatial distribution of Yield, NH3 and EC at the subplot level.

	Model 1
(Intercept)	327.993**
	(125.638)
aa_NH3	−1.223
	(1.308)
I(aa_NH3^2)	0.004
	(0.003)
I(aa_NH3 * ec)	0.002
	(0.003)
I(aa_NH3^2 * ec)	0.000
	(0.000)
Num.Obs.	784
R2	0.010
F	2.023
RMSE	5.71

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Results from the regression model