Redesigning Yield Map Plots for Comprehension and Usability

2023 SDSS

Alison Kleffner

Department of Statistics, University of Nebraska - Lincoln

akleffner@huskers.unl.edu

https://alison-kleffner.netlify.app/

alisonkleffner

Slides: https://alisonkleffner.github.io/presentations/SDSS_2023/index.html#1

University of Nebraska Lincoln

Outline

- Data Intensive Farm Management
- Perceptual Issues with Current Maps
- Redesign Process
- Future Work

Data Intensive Farm Management

University of Nebraska Lincoln

Background

- With a projected increase in future crop demand, researchers are conducting studies on crop input application to increase yield, focusing on sustainability (Tilman, Balzer, Hill, and Befort, 2011)
- Crop Input Example: Nitrogen Fertilizer
  - Nitrogen is an essential component of food production as allows plants to photosynthesize efficiently (Maheswari, Murthy, and Shanker, 2017)
  - Nearly half of the nitrogen fertilizer supplied to the field is not used by the crops (Billen, Garnier, and Lassaletta, 2013)
  - This excess nitrogen can be harmful
- Hence, research needs to be conducted on determining input rates that increase crop yield, and are also more sustainable.

Data Intensive Farm Management (DIFM)

Problem: Address inefficient application of crop inputs to farm fields worldwide

Methods: On-Farm Precision Experimentation

- Conduct experiments using site-specific inputs
- GPS-reliant technology

Goals

- Develop infrastructure to develop and analyze these experiments
- Find economically optimal application rate to increase profit while reducing environmental impacts.

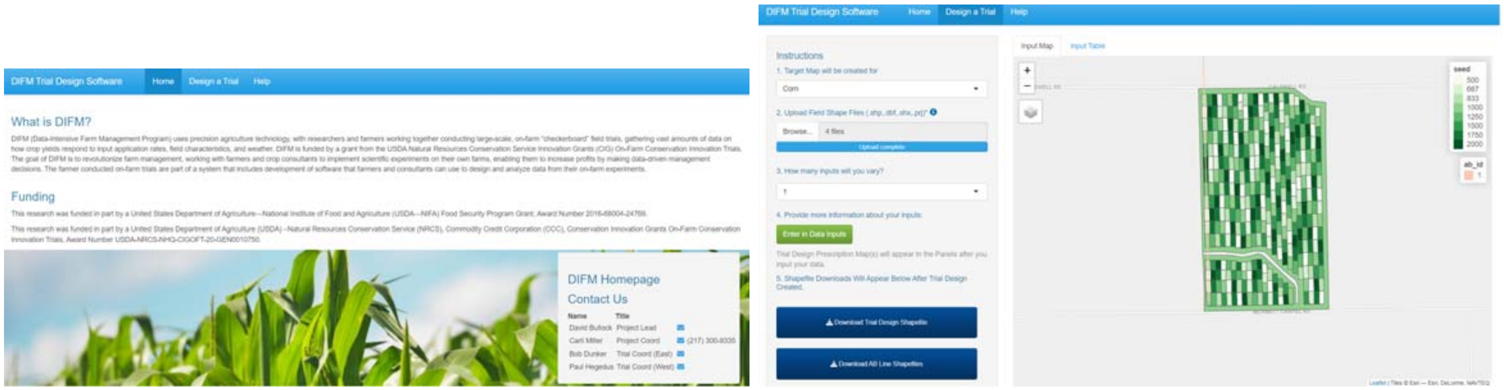
Project Website

DIFM-CIG
USDA NRCS
Natural Resources Conservation Service

Trial Design

Step 1: Develop infrastructure to develop experiments using site-specific inputs

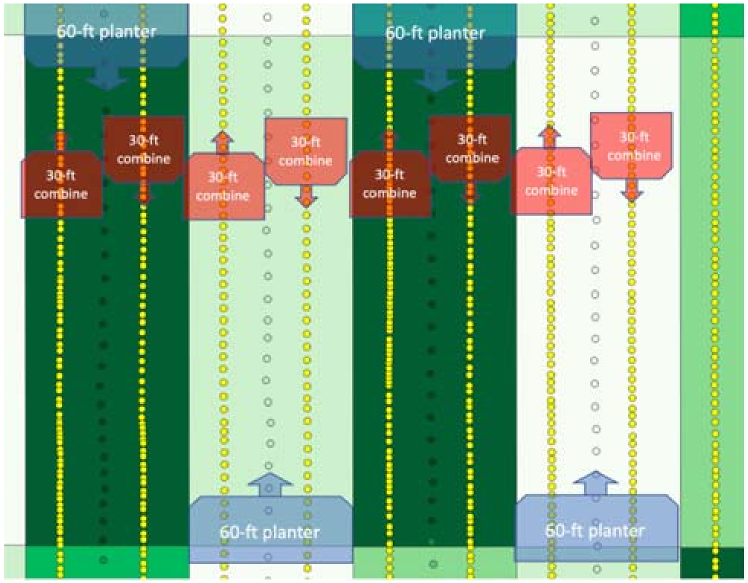
Trial Design Tool



Output: shape files that can be put into a farmer's tractor that allows them to carry out the experiments

Data Collection

Step 2: Conduct experiments and collect data



Examples of Data Collected:

- As-applied
- Yield
- Location of measurements

Explain the Results

Step 3: Explain the optimal management decisions, accounting for various factors

How: create a user interface, designed around explaining machine learning output to non-experts

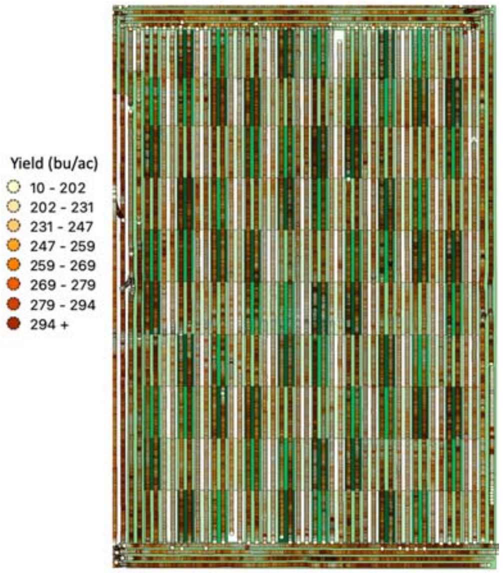
- Build trust in models
- Learn how crop yield responds to different input application rates, field characteristics, and weather to hopefully increase profits.

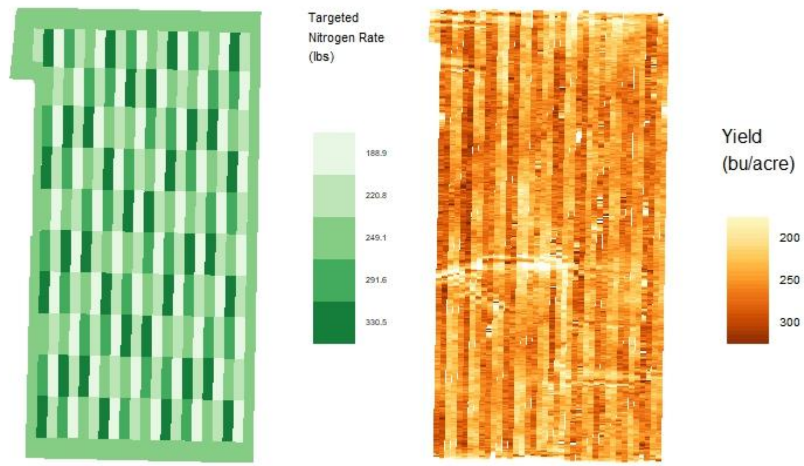
One way to do this: Visually explore the relationship between input application and yield through a graph

- Show the spatial correlations between the application/treatment and yield in a way that is understandable to farmers and consultants.
- Develop perceptually optimal plots that communication this relationship.

Current Maps

Yield Maps Currently Used by DIFM

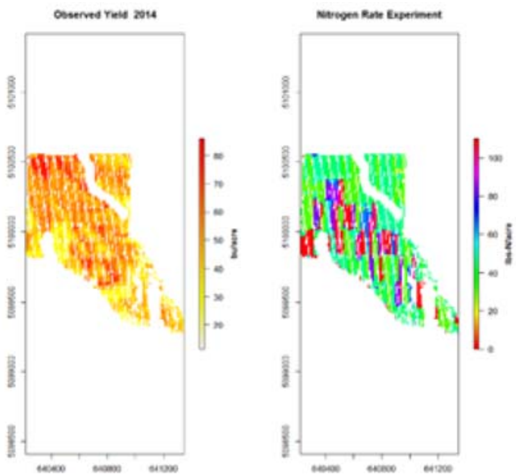


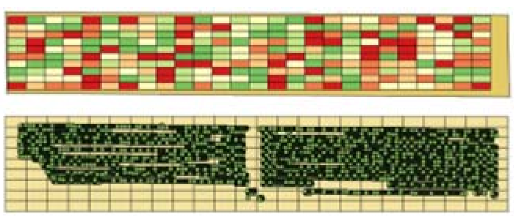


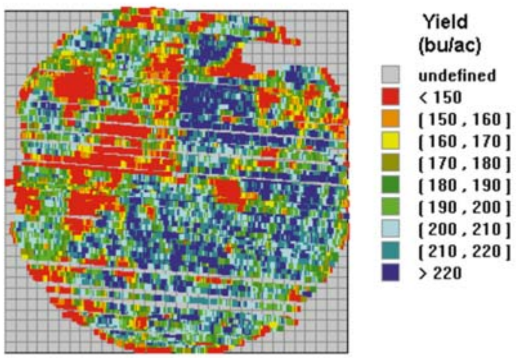
Examples of two versions of this plot currently given to farmers/crop consultants in DIFM reports

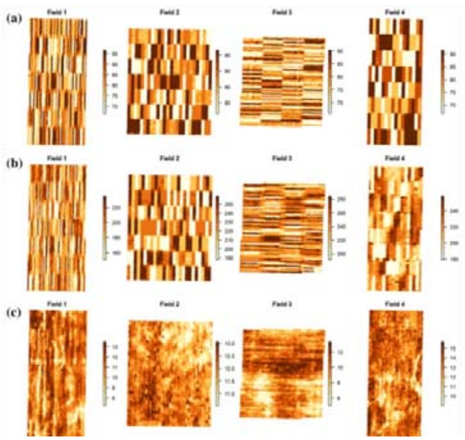
Other Maps in the Literature

Plenty of other iterations of trial design/yield maps used in the literature. Here's some examples

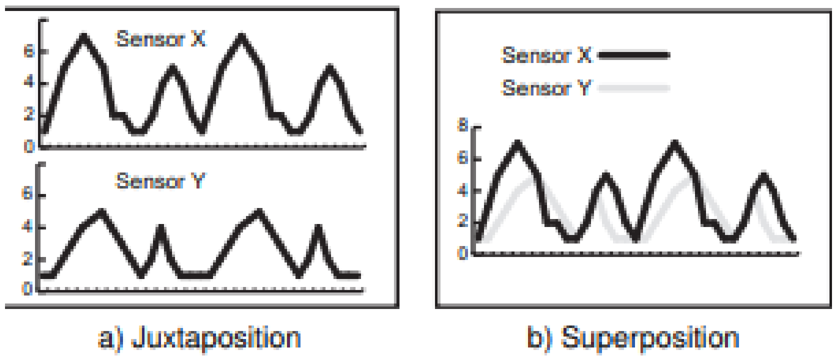








Juxtaposed vs Superimposed Graphs



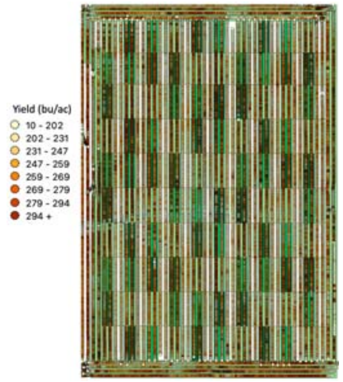
- **Juxtaposed graphs**: side by side (Gleicher, Albers, Walker, Jusufi, Hansen, and Roberts, 2011)
  - Benefits: less issues with visual clutter and easier to create
  - Drawbacks: comparative burden is placed on the user

- **Superimposed graphs**: multiple objects in same coordinate system (Gleicher, Albers, Walker, et al., 2011)
  - Benefits: Easier to compare as users can use perception rather than memory
  - Drawbacks: clutter
  - Useful when spatial location is a key component of the comparison (Wang, Haleem, Shi, Wu, Zhao, Fu, and Qu, 2018)

Data Clutter with Superimposed Graphs

Drawback: clutter

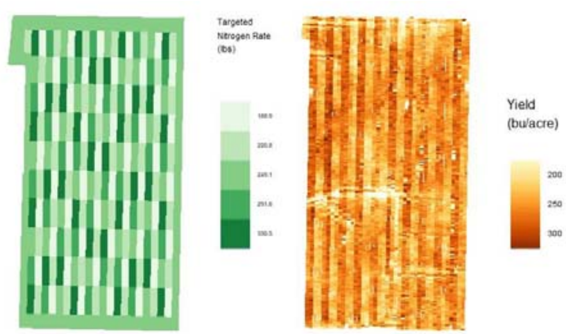
- General principle of graphical design: show data clearly (Cleveland, 1984; Gordon and Finch, 2015)
- Overlap - multiple dots on top of one another
  - Obscures true number of dots, harder to find patterns
  - Visual cues, like color, becomes partially obstructed, thereby reducing search efficiency (Bravo and Farid, 2004a; Bravo and Farid, 2004b)
  - Can overburden human perception, causing errors in performing tasks (Huang, Eades, and Hong, 2009)



Comparative Burden with Superimposed

Drawback: Most of the comparative burden placed on users' memory

- A mental image is relied on for comparison in these scenarios, as the user moves their eyes between images (shifting focus).
  - The plot contents may not be accurately formed in working memory, leading to potential errors when deriving patterns (Vanderplas, Cook, and Hofmann, 2020; LYi, Jo, and Seo, 2021)
- Lack of visual cues for locations



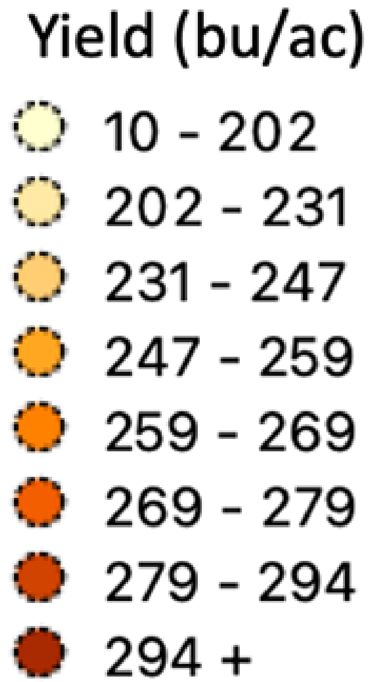
Color

- Red-green color blindness is experienced by approximately 8% of men and 0.5% of women of Northern European ancestry.
  - Difficult to discriminate between these colors (Wong, 2011)
- Stop-Light Color Scheme
  - Yellow has a highlighting effect
  - Univariate Scale more appropriate (magnitude only)
- Same Color scheme for multiple variables
  - May cause confusion
- Rainbow Color Scheme
  - No inherent ordering of magnitude (Light and Bartlein, 2004)
  - Extremes are visually close (red and violet) (Silva, Sousa Santos, and Madeira, 2011)

Number of Categories

Number of categorical scales should be limited to **5-7 categories** (Miller, 1956)

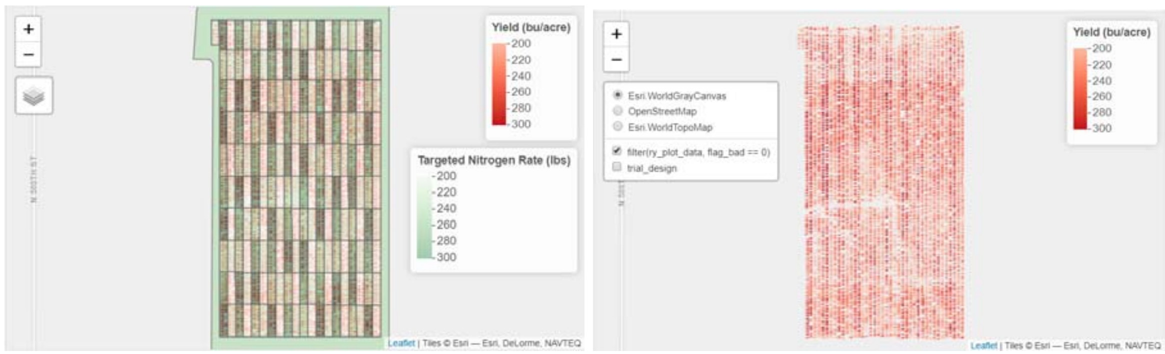
- Due to working memory limits (Macdonald, 1999)
    - Harder for the user to distinguish between colors and remember the meaning of colors.
- The load on the user's working memory leads to an increase in the time it takes for the user to comprehend the plot (Huang, Eades, and Hong, 2009)



Next: Redesign Process

Redesign

Redesign: Color Blending Part 1



Focus: Superimpose the treatment and yield plots, while reducing the clutter

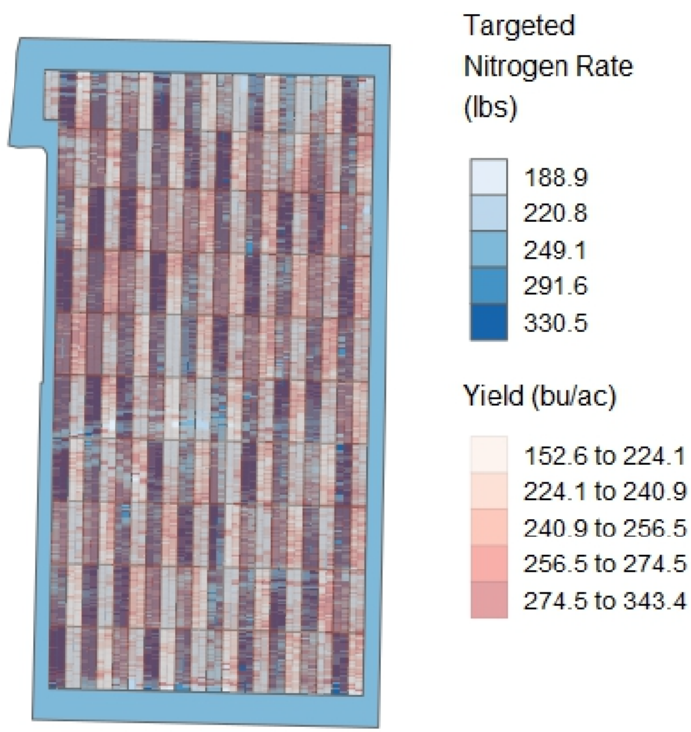
- Use of transparency to show both at same time
- Interactive plot with *leaflet* (Cheng, Karambelkar, and Xie, 2022)
  - Can add and remove trial plot layer

Work that still needs to be done:

- Color needs some work
  - different color schemes

Redesign: Color Blending Part 2

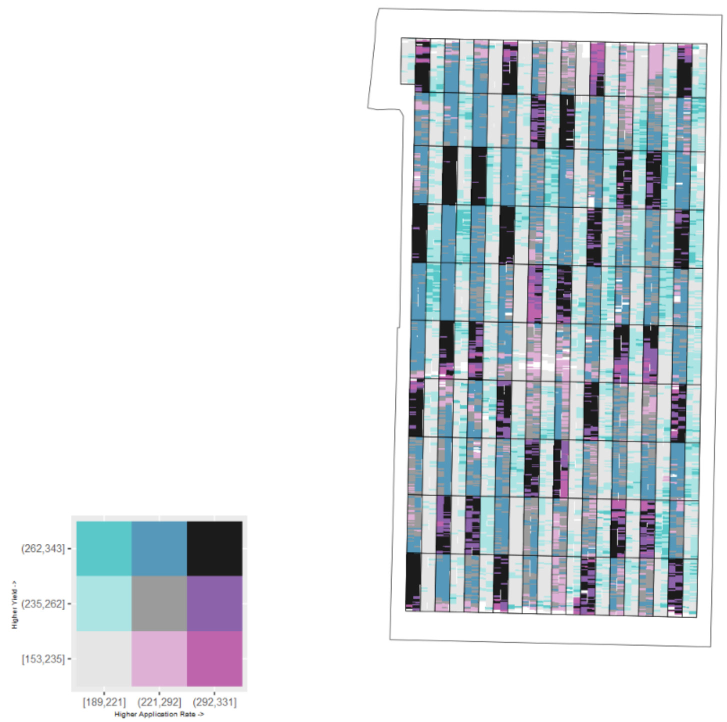
Focus: Color/Scales



- Working on blending using transparency
- Categorical Groups for both variables
  - Trial Design treatment is a factor
  - Reduce number of colors for yield
  - Use quantiles to determine categories for yield (Brewer and Pickle, 2002)

Redesign: Bivariate Color Plot

Alternative to color blending



- **Benefit**: relationship between the variables is most important (Elmer, 2013)
- Recommendation: 3x3 scale (Leonowicz, 2003)
    - ▪ Quantiles (Biesecker, Zahnd, Brandt, Adams, and J.M, 2020)
- **Focus**: diagonal
    - Diagonal: grayscale color scheme.
    - Upper left and lower right: complementary color scheme (Strode, Morgan, Thornton, Mesev, Rau, Shortes, and Johnson, 2020)
- **Drawback**: Lose more detailed information

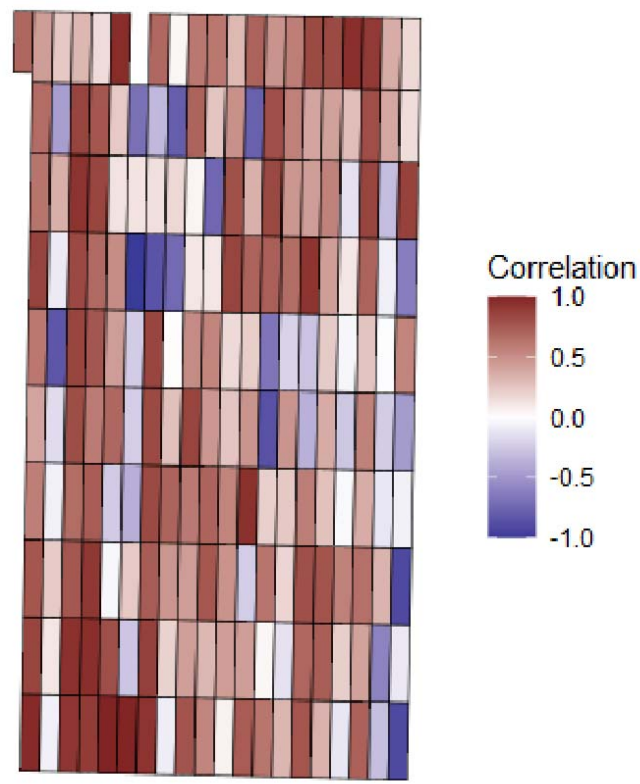
Redesign: Correlation

Directly encode correlations between As Applied treatments and Yield

Benefit: Direct statement of correlation while maintaining some spatial orientation

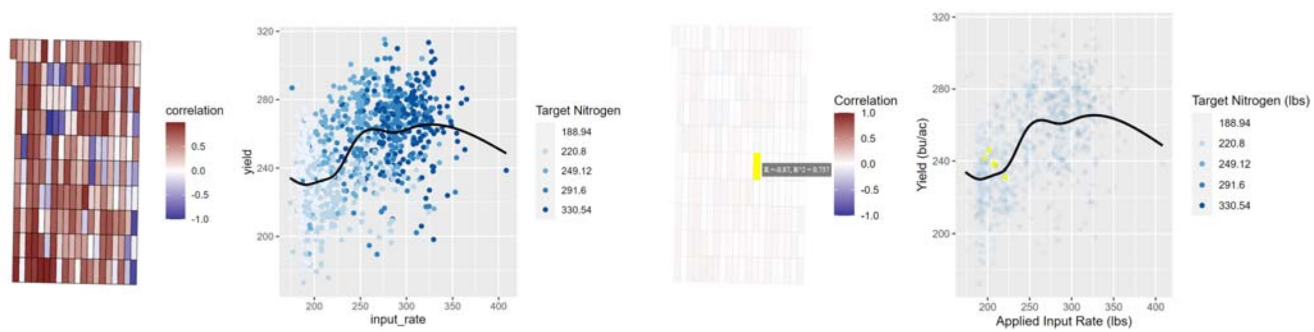
- Explicit Encoding Layout (Gleicher, Albers, Walker, et al., 2011)
  - Bivariate color scale
  - Maintain some spatial information.
- Correlations may be impacted by field location

Drawback: complicated to connect the displayed relationship back to the data (Gleicher, Albers, Walker, et al., 2011)



Redesign: Correlation with Scatterplot

Add some context back



- A standard practice to overcome weakness of decontextualization is utilizing a hybrid comparative layout (LYi et al., 2021).
  - Juxtaposed scatterplot to the correlation plot (combining the layouts of juxtaposition and explicit encoding).
- Interactivity connects the juxtaposed plots, where hovering over a trial plot in the correlation map highlights the corresponding points in the scatterplot used in the correlation calculation.

Link

Future Work

Conclusion/Future Work

Next Step: Obtain Feedback from those using the plots (farmers, crop consultants)

- Eventually do some testing between the layouts to see which farmers are reading more accurately.

Eventually: Develop a R Shiny app to explaining machine learning output to non-experts

- Build trust in the model predictions without requiring farmers to learn the details of statistical modeling.
- Will utilize these plots, among others
    - For example, accuracy of as applied treatment compared to trial design

References

Biesecker, C., W. Zahnd, H. M. Brandt, et al. (2020). "A Bivariate Mapping Tutorial for Cancer Control Resource Allocation Decisions and Interventions". In: Preventing chronic disease 17 (E01). DOI: 10.5888/pcd17.190254.

Billen, G., J. Garnier, and L. Lassaletta (2013). "The nitrogen cascade from agricultural soils to the sea: modelling nitrogen transfers at regional watershed and global scales". In: Philosophical Transactions of the Royal Society B: Biological Sciences 368.1621, p. 20130123. DOI: 10.1098/rstb.2013.0123.

Bravo, M. and H. Farid (2004a). "Recognizing and segmenting objects in clutter". In: Vision Research 44.4, pp. 385–396.

— (2004b). "Search for a category target in clutter". In: Perception 33.5, pp. 643–652.

Brewer, C. A. and L. Pickle (2002). "Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series". In: Annals of the Association of American Geographers 92.4, pp. 662-681.

Cheng, J., B. Karambelkar, and Y. Xie (2022). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.1.1. URL: https://CRAN.R-project.org/package=leaflet.

Cleveland, W. S. and R. McGill (1984). "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods". In: Journal of the American Statistical Association 79.387, pp. 531–554.

Elmer, M. E. (2013). "Symbol considerations for bivariate thematic maps". In: Proceedings of 26th International Cartographic Conference.

Gleicher, M., D. Albers, R. Walker, et al. (2011). "Visual Comparison for Information Visualization". In: Information Visualization 10.4, pp. 289-–309.

Gordon, I. and S. Finch (2015). "Statistician Heal Thyself: Have We Lost the Plot?" In: Journal of Computational and Graphical Statistics 24.4, pp. 1210–1229. DOI: 10.1080/10618600.2014.989324.

Huang, W., P. Eades, and S. Hong (2009). "Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective". In: Information Visualization 8.3, pp. 139–152.

Leonowicz, A. (2003). "RESEARCH ON TWO-VARIABLE CHOROPLETH MAPS AS A METHOD FOR PORTRAYING GEOGRAPHICAL RELATIONSHIPS". In: Proceedings of the 21st International Cartographic Conference. The International Cartographic Association.

Light, A. and P. Bartlein (2004). "The End of the Rainbow? Color Schemes for Improved Data Graphics". In: Eos, Transactions American Geophysical Union 85.40, pp. 385–391. DOI: 10.1029/2004EO400002.

LYi, S., J. Jo, and J. Seo (2021). "Comparative Layouts Revisited: Design Space, Guidelines, and Future Directions". In: IEEE Transactions on Visualization and Computer Graphics 27.2, pp. 1525–1535.

Macdonald, L. (1999). "Using color effectively in computer graphics". In: Computer Graphics and Applications, IEEE 19, pp. 20–35. DOI: 10.1109/38.773961.

Maheswari, M., A. Murthy, and A. Shanker (2017). "12 - Nitrogen Nutrition in Crops and Its Importance in Crop Quality". In: The Indian Nitrogen Assessment. Ed. by Y. P. Abrol, T. K. Adhya, V. P. Aneja, N. Raghuram, H. Pathak, U. Kulshrestha, C. Sharma and B. Singh. Elsevier, pp. 175–186. DOI: https://doi.org/10.1016/B978-0-12-811836-8.00012-4.

Maxwell, B., P. Hegedus, P. Davis, et al. (2018). "Can Optimization Associated with On-Farm Experimentation Using Site-Specific Technologies Improve Producer Management Decisions?" In: 14th International Conference on Precision Agriculture.

References

Miller, G. A. (1956). "The magical number seven plus or minus two: some limits on our capacity for processing information." In: Psychological review 63 (2), pp. 81–97.

Peerlinck, A., J. W. Sheppard, and B. D. Maxwell (2018). "Using Deep Learning in Yield and Protein Prediction of Winter Wheat Based on Fertilization Prescriptions in Precision Agriculture". In: 14th International Conference on Precision Agriculture.

Searcy, S. W. (1997). "Precision Farming: A New Approach to Crop Management, Texas Agricultural Extension Service". In: The Texas A&M University System, College Station, TX, USA.

Silva, S., B. Sousa Santos, and J. Madeira (2011). "Using color in visualization: A survey". In: Computers & Graphics 35.2, pp. 320–333. DOI: https://doi.org/10.1016/j.cag.2010.11.015.

Strode, G., J. Morgan, B. Thornton, et al. (2020). "Operationalizing Trumbo's Principles of Bivariate Choropleth Map Design". In: Cartographic Perspectives 94, pp. 5–24. DOI: 10.14714/CP94.1538.

Tilman, D., C. Balzer, J. Hill, et al. (2011). "Global food demand and the sustainable intensification of agriculture". In: Proceedings of the National Academy of Sciences 108.50, pp. 20260–20264. DOI: 10.1073/pnas.1116437108.

Trevisan, R., D. Bullock, and N. Martin (2021). "Spatial Variability of Crop Responses to Agronomic Inputs in On-Farm Precision Experimentation". In: Precision Agriculture 22.2, pp. 342–363. DOI: 10.1007/s11119-020-09720-8.

Vanderplas, S., D. Cook, and H. Hofmann (2020). "Testing Statistical Charts: What Makes a Good Graph?" In: Annual Review of Statistics and Its Application 7.1, pp. 61–88.

Wang, Y., H. Haleem, C. Shi, et al. (2018). "Towards Easy Comparison of Local Businesses Using Online Reviews". In: Computer Graphics Forum 37.

Wong, B. (2011). "Color blindness". In: nature methods 8.6, p. 441.

Questions?

akleffner@huskers.unl.edu

alisonkleffner