University of Vermont

UVM ScholarWorks

8-30-2019

US county-level agricultural crop production typology

Courtney R. Hammond Wagner
University of Vermont

Meredith T. Niles
University of Vermont

Eric D. Roy
University of Vermont

Follow this and additional works at: https://scholarworks.uvm.edu/calsfac

Part of the Agriculture Commons, and the Sustainability Commons

DATA NOTE

US county-level agricultural crop production typology

Courtney R. Hammond Wagner[1,2], Meredith T. Niles[2,3] and Eric D. Roy[1,2*]

Abstract

Objectives: Crop production is an important variable in social, economic and environmental analyses. There is an abundance of crop data available for the United States, but we lack a typology of county-level crop production that accounts for production similarities in counties across the country. We fill this gap with a county-level classification of crop production with ten mutually exclusive categories across the contiguous United States.

Data description: To create the typology we ran a cluster analysis on acreage data for 21 key crops from the United States Department of Agriculture's 2012 Agricultural Census. Prior to clustering, we estimated undisclosed county acreage values, controlled for acreage in other crop types, and removed counties with low agricultural production to produce proportional scores for each crop type in each county. We used proportional scores to control for the influence of county size in the cluster analysis and used internal and stability measures to validate the analysis. The final dataset features 2922 counties. Future research can leverage this typology as an input for county- or regional-level analysis.

Keywords: Agriculture, Farming systems, Typology, Crop production, United States, County-level

Objective

Agricultural production is an important variable for many different types of analyses, including, for example, regional economic models, environmental assessments of water quality or greenhouse gas emissions, and analyses of social trends. While ample data are available to examine various aspects of agricultural production across the US, we lack a typology of recent agricultural production across the country. We present a dataset with an agricultural production typology for counties across the contiguous US, derived from a cluster analysis of the 2012 county-level crop data [1].

The US Department of Agriculture (USDA) Economic Research Service's (ERS) Farm Resource Regions, defined in the early 2000s, are examples of agricultural regions partially based on county-level agricultural production [2]. The USDA uses the Farm Resource Regions to examine regional trends and determine program and funding priorities [3]. The ERS Farm Resource Regions were informed by a cluster analysis of county-level farming characteristics in the early 1990s [4]. Cropping patterns have shifted since 1991 due to economic trends, agricultural policy and climate change [5–8]. Using USDA crop data from the 2012 Agricultural Census, we provide a more up-to-date crop production typology.

The motivation for this dataset came from a need for a single categorical, county-level variable that incorporates the diversity of crop types grown throughout the US, where previous efforts were largely geographically focused. Additionally, we sought to capture similarities in agricultural production, including diversity of production and relative quantity of production, between counties. We anticipate that this dataset will be of use in analyses that seek to understand county-level patterns that may relate to crop production types, as we intend to do in a forthcoming publication focused on nitrogen dynamics.

*Correspondence: eroy4@uvm.edu
[1] Rubenstein School for Environment and Natural Resources, University of Vermont, Burlington, Vermont, USA
Full list of author information is available at the end of the article

Hammond Wagner *et al. BMC Res Notes*   (2019) 12:552

Page 2 of 3

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	1_Crop_cluster_data_wrangling_forRepository_071919	R script (.R)	Figshare: https://doi.org/10.6084/m9.figshare.8132867.v2
Data file 2	2_Cluster_Analysis_22crops_forRepository	R script (.R)	Figshare: https://doi.org/10.6084/m9.figshare.8132867.v2
Data file 3	Methods_and_Technical_Validation_US_county_agricultural_clusters_figshare	Word file (.docx)	Figshare: https://doi.org/10.6084/m9.figshare.8132867.v2
Data set 1	crop_production_typology_data_and_metadata_052819	Excel file (.xlsx)	Figshare: https://doi.org/10.6084/m9.figshare.8132867.v2

Data description

Our dataset identifies 10 mutually exclusive, agricultural crop production categories for the contiguous US based on crop production values in the 2012 USDA Agricultural Census Data [9]. Each crop production category is defined by the crops that emerged from the USDA crop data to drive membership in each county-level crop production cluster. These ten categories are: (1) corn silage and other crops, (2) tobacco, (3) hay, (4) barley, beans and sugar beets, (5) alfalfa and barley, (6) sorghum, sunflower and wheat, (7) oranges and sugarcane, (8) rice, (9) corn grain and soybeans, and (10) cotton and peanuts.

County-level crop acreage data was obtained from the USDA NASS Quick Stats Database from the US Agricultural Census of 2012 [9]. We downloaded county level crop acreage data for the 21 crops that, according to International Plant Nutrition Institute (IPNI), account for an average of 95% of harvested cropland acres reported in the agricultural census across the contiguous US [10]. The 21 crops are: apples, barley, canola, beans, corn grain, corn silage, cotton, alfalfa, oranges, peanuts, potatoes, rice, sorghum, soybeans, sugar beets, sugarcane, sunflower, sweet corn, tobacco, wheat and other hay (i.e. all hay acreage excluding alfalfa). Additionally, we downloaded the total county acreage and created a "22nd crop" which represents acreage of all other crops grown in the county that are unaccounted for in the 21 crops. The 22nd crop category captures acreage in the 55 crops included in the agricultural census that are less prevalent (i.e. combined represent only 5% of harvested cropland in the US), for example, cucumber, oats and cherries [9]. The initial data download from Quick Stats included 3060 counties, out of the total 3108 county equivalents in the contiguous US. The data cleaning process resulted in a final dataset of 2922 counties or 94% of the counties in the contiguous US.

USDA Agricultural Census data contains withheld data in the form of "(D)" and "(Z)" values in the dataset to avoid disclosing data for individual farms and to represent small figures, respectively [11]. To clean the dataset we changed all (Z) values to zero, as (Z) values represent a value of less than half an acre [11]. Then we followed the IPNI methodology [10] to estimate missing values for all (D)s in the dataset. We describe these methods in detail in the methods and technical validation document (see Table 1). We then created the 22nd crop variable by summing acreage for all 21 crops and subtracting this from the reported county total harvested cropland. To control for differences in county size, we transformed the absolute acreage values to proportional scores.

We then performed a k-means cluster analysis [12] on the 22 county-level crop acreage proportional scores. We determined the optimal number of clusters, or crop production categories, to be 10. We analyzed the cluster scree plot, as well as ran internal and stability measures using the clValid R package [13]. We describe in detail the technical validation of our cluster analysis in the methods document (see Table 1).

Limitations

We acknowledge that while we have justified and validated our selection of a 10-cluster solution, the selection of a different number of clusters would change the dataset.

Authors' contributions
CRHW collected (downloaded) and collated the data, ran the cluster analysis and validation, and led the writing of the manuscript. MTN co-designed the study, advised data collation, analysis and validation, and assisted in writing the manuscript. EDR co-designed the study, advised data collation, analysis and validation, and assisted in writing the manuscript. All authors read and approved the final manuscript.

Author details

[1] Rubenstein School for Environment and Natural Resources, University of Vermont, Burlington, Vermont, USA. [2] Gund Institute for Environment, University of Vermont, Burlington, Vermont, USA. [3] Department of Food and Nutrition Sciences & Food Systems Program, University of Vermont, Burlington, Vermont, USA.

References

1. Hammond Wagner CR, Niles MT, Roy ED. US county-level agricultural crop production typology. Figshare; 2019. https://doi.org/10.6084/m9.figshare.8132867.v2.
2. USDA ERS. Farm resource regions. United States Department of Agriculture; 2000 Sep. Report No.: Number 760.
3. Wade T, Claassen RL, Wallander S. Conservation-practice adoption rates vary widely by crop and region. United States Department of Agriculture, Economic Research Service; 2015. https://www.ers.usda.gov/webdocs/publications/44027/56332_eib147.pdf?v=42403.
4. Sommer JE, Hines FK. Diversity in U.S. agriculture: a new delineation by farming characteristics. Agricultural economic report (USA). 1991. http://agris.fao.org/agris-search/search.do?recordID=US9153983. Accessed date 5 Dec 2018.
5. Aguilar J, Gramig GG, Hendrickson JR, Archer DW, Forcella F, Liebig MA. Crop species diversity changes in the United States: 1978–2012. PLoS ONE. 2015;10(8):e0136580.
6. Iizumi T, Ramankutty N. How do weather and climate influence cropping area and intensity? Glob Food Sec. 2015;1(4):46–50.
7. Plourde JD, Pijanowski BC, Pekin BK. Evidence for increased mono-culture cropping in the Central United States. Agr Ecosyst Environ. 2013;15(165):50–9.
8. Westcott PC. Ethanol expansion in the United States: How will the agricultural sector adjust? Economic Research Service: United States Department of Agriculture; 2007. Report No.: FDS-07D-01.
9. USDA-NASS. Quick Stats. Washington, DC: National Agricultural Statistics Service, U.S. Department of Agriculture; 2019. https://quickstats.nass.usda.gov/.
10. IPNI. A nutrient use information system (NuGIS) for the U.S. Norcross; 2012. http://www.ipni.net/nugis.
11. USDA-NASS. 2012 Census of Agriculture: United States, Summary and State Data. United States Department of Agriculture; 2014. Report No: AC-12-A-51.
12. Everitt B, Hothorn T. An introduction to applied multivariate analysis with R. Berlin: Springer; 2011.
13. Brock G, Pihur V, Datta S, Datta S. clValid, an R package for cluster validation. J Stat Softw. 2011;25:1–22.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.