# Exploring a Big Data Approach to List Building

Linda J. Young
USDA National Agricultural Statistics Service, Research and Development Division
1400 Independence Avenue SW
Washington DC 20250, USA
Linda.Young@nass.usda.gov

Barbara R. Rater
USDA National Agricultural Statistics Service, Census and Survey Division
1400 Independence Avenue SW
Washington DC 20250, USA
Barbara.Rater@nass.usda.gov

Michael Hyman
USDA National Agricultural Statistics Service, Research and Development Division
1400 Independence Avenue SW
Washington DC 20250, USA
Michael.Hyman@nass.usda.gov

DOI: 10.1481/icasVII.2016.f29b

## ABSTRACT

The US Department of Agriculture's National Agricultural Statistics Service (NASS) has the responsibility of quantifying the nation's agricultural production. Historically it has focused on large, production agriculture. With increased interest and activity in the urban areas, NASS has begun exploring how to better quantify urban agriculture. This segment of agriculture is particularly challenging to enumerate because the agricultural holdings tend to be small, widely dispersed, and more transient than the predominantly large farms in rural areas. In collaboration with the Multi-Agency Collaboration Environment (MACE), a new approach to list building was explored in a pilot study conducted in the City of Baltimore, Maryland. Using a big data approach, areas of potential agricultural activity were identified by gathering information (state and local permits, facebook and twitter feeds, interest groups, *etc*.) via the web. A sample was drawn from the list, and an in-person survey was conducted to assess whether or not the identified areas were producing agriculture. Here the lessons learned from the study and next steps are discussed.

**Keywords:** urban agriculture, list building, web scraping

## 1. Introduction

Urban agriculture has been of increasing interest at the local, state, and national levels. Some cities, such as Detroit, Michigan, have goals of increasing their food resiliency by growing a large portion of the fruits and vegetables used by their citizens within the city limits (Colasanti, *et al*. 2010). Cities have also promoted urban agriculture to ensure that the vacant lots, abandoned buildings, and other under-utilized land are used productively (Goldsmith 2014, Santo, *et al.* 2016). Urban agriculture takes a variety of forms including backyard gardens, school or community gardens, urban farms, greenhouses, hoop houses, converted warehouses, and vertical gardens. At the local level, policies favourable to urban agriculture have been developed, access to land has been facilitated, and sometimes funding can be obtained to initiate agricultural activity. Federal funding programs for urban agriculture, such as The New Farmers and People's Garden Initiatives, have been established. The US Department of Agriculture (USDA) has developed a website with a wealth of information, including more funding opportunities, for urban farmers (USDA 2016). Organizations focused on urban agriculture have also been developed. These efforts have naturally led policy makers at the local and national levels to ask whether or not the programs are effective in increasing urban agriculture.

Although USDA's National Agricultural Statistics Service (NASS) has always included urban farms in its counts of farms and farm production, the estimates have not been as precise as those for other sectors of the agricultural economy. Because urban agriculture is responsible for a small portion of the total farm production, NASS has not devoted the resources—and has not had the resources—needed to provide better estimates. Policy makers at all levels are increasingly interested in the efficacy of programs designed to increase urban agriculture. Thus, the Under Secretary for USDA's Research, Education and Economics asked NASS to explore ways to better quantify the extent and food production of urban agriculture.

Urban farms tend to be smaller, more diverse, more transient, and more widely dispersed than the more traditional farms in rural areas of the U.S. This makes these farms challenging to identify and thus to quantify. The NASS list frame is thought to have substantial under-coverage of urban farms. During the 2012 Census of Agriculture, capture-recapture methods were employed to adjust for under-coverage, non-response, and misclassification, using the NASS Census mailing list and a sample from the NASS area frame, as the two independent samples (Young, *et al.* 2012). For a survey of urban agriculture, a sample from the NASS list frame, which is a list of all known confirmed or potential farms in the US, could serve as one sample in a capture-recapture analysis. However, it is cost prohibitive to obtain sufficient numbers of urban farms in a sample drawn from the NASS area frame because these farms tend to be dispersed (not concentrated) within the urban areas. Thus, NASS began to explore alternative ways to build a list of urban farms that would be independent of the NASS list frame, provide good coverage, and be relatively inexpensive, as a foundation for a second sample.

Taylor and Lovell (2014) identified potential areas of urban agricultural activity in Chicago, Illinois, from high-resolution aerial images in Google Earth in conjunction with ArcGIS. First, lists from several non-governmental organizations were combined to identify 1236 potential community gardens in Chicago. The sites were visited in 2010, and 12.9% were found to be food gardens with the others being ornamental garden/parks, a streetscaping projects, or no garden. The confirmed food and non-food gardens were analysed to develop a visual classification approach to classifying urban agriculture. Then the Google Earth images of Chicago were analysed visually, and 4493 potential areas of agricultural activity that were not included in the 1236 areas on the lists were identified. In a follow-up survey of 194 of the sites, 166 (89.6%) were found to have agriculture on them.

Although the methods used by Taylor and Lovell (2014) were effective for identifying urban agricultural sites, the approach was labor intensive, and it would be cost prohibitive when scaled to the national level. Thus, NASS began to consider ways in which to automate the approach. The results of that effort are presented in this paper. In Section 2, the target population for a national urban agricultural study is defined. A pilot study of a new approach to identifying urban agriculture in the City of Baltimore, with considerations for national implementation, is described in Section 3. The final section reviews the lessons learned and future directions.

## 2. Urban Agriculture

Historically, urban agriculture has not been reported separately; it has been combined with all other types of agriculture. To report on it separately, urban agriculture had to be defined so that an operation can be unambiguously identified as either being or not being urban agriculture. After consulting with the USDA's Economic Research Service, agriculture within the urbanized areas, as defined by the US Census Bureau, was defined to be within the target population (see Figure 1).

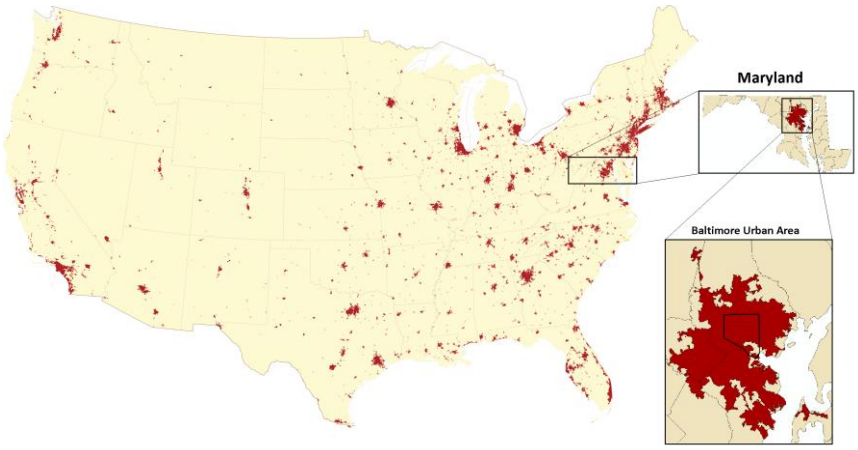


**Figure 1**: *Urbanized areas within the US (left), area of pilot study within US (upper right), and Baltimore City boundaries contrasted with Baltimore urbanized area (lower right)*

Then consideration turned to what should be included as agriculture within these urbanized areas. In the US, the definition of a farm is any operation that produces and sells or has the potential to sell $1000 or more of agricultural products in a year. Thus, an individual with a large backyard garden who sells the excess produce at a farmers market is a farmer if the produce sells for at least $1000 over the course of a year.

Community gardens are a common form of agriculture occurring in urbanized areas. For these types of operations, NASS has traditionally considered each plot as an operation, which either qualifies as a farm and is counted or found to not qualify as a farm and is not counted. Often none of the plots produces enough to qualify as a farm even though the total production of the community garden could well be enough to be classified as a farm. This could contribute to the perception that NASS is not fully reporting agricultural activity in urban areas. Thus, for the purposes of the pilot study, all agriculture was considered, but the entities meeting the definition of a farm were to be reported separately from those that did not qualify as a farm, such as ones that produced food only for home consumption.

What types of urban agriculture were within the target population? Community and school gardens were included as well as animal production, such as bees and chickens. In many urban

areas, abandoned warehouses have been converted for agricultural production, which were part of the target population. Greenhouses on the top of buildings, production in hoop houses, and aquaponics were also determined to be within the scope of urban agriculture.

What types of entities were not included in the target population of urban agriculture? Gardens on the top of buildings for enjoyment and not for production were excluded. Plants grown in window boxes or in pots on the porch were excluded. Also, mobile gardens, such as those on the top of a bus or in the bed of a pickup, were not in the target population.

## 3. City of Baltimore Pilot Study

A pilot study was conducted in the city of Baltimore. Recent city urban agricultural initiatives include the Baltimore Office of Sustainability's Land Leasing Initiative, the Urban Agriculture Tax Credit, and the Urban Agriculture Training Program (Baltimore Office of Sustainability 2016). With a plethora of vacant buildings and vacant lots, Baltimore City is focused on redevelopment strategies for the city's land use policies and on greening initiatives, such as urban agriculture. Based on available funds, the study area did not include the full urbanized area of Baltimore but was restricted to the City of Baltimore (see Figure 1).

To build a list of agricultural areas within Baltimore, NASS collaborated with the Multi-Agency Collaboration Environment (MACE). MACE is a consortium of Government agencies and contractors that solves complex data problems. In collaboration with the Air Force Research Laboratory, they explored the use of text and image analytics to create a list of urban farms within the City of Baltimore.

Traditional imagery analysis techniques were found to not be useful for automatically identifying the small areas of agriculture common in urban farms. The satellite imagery available had a 0.5-m resolution. An example of that imagery is provided on the left side of Figure 2. On the right is a 15-cm aerial imagery of the same area. (The green dots are at the same point in space on each.) The human eye is more effective than algorithms in detecting items within imagery. The agricultural activity evident in the 15-cm imagery is not visible in the 0.5 satellite imagery. Thus, satellite imagery was found to not be useful for this purpose. With the aerial imagery, the presence of agriculture is evident, and automatic identification is potentially possible. However, the aerial imagery was prohibitively expensive for NASS, so NASS was unable to take advantage of either satellite or aerial imagery. Therefore, the MACE list was based on the web-scraping techniques associated with the text analytics.

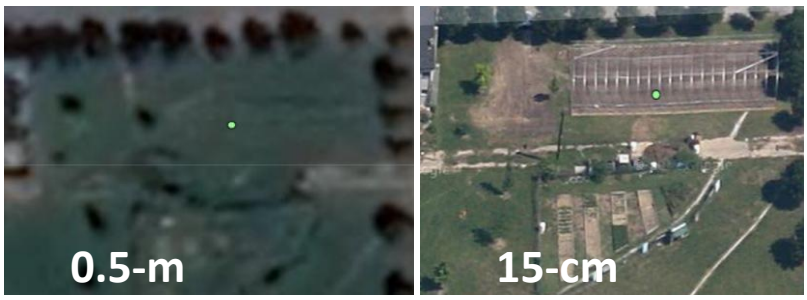


**Figure 2**: *Comparison of visibility of agricultural area using 0.5-m satellite imagery (left) compared to 15-cm aerial imagery (right). The green dots are at the same point in space.*

MACE identified 505 areas of potential agriculture. The City of Baltimore had a 2013 list of areas of agricultural activity that was not available for the development of either the NASS list frame or the MACE list. Of the 159 non-school garden areas on the list, one was on the NASS list frame, and 89 were on the MACE list. The NASS list frame has only farms and potential farms while the MACE list is broader, including agricultural activity of any size. This could explain some

of the difference. However, Baltimore's list included 13 urban farms; one of those was on the NASS list frame; the MACE list included all 13 farms.

To further evaluate the MACE list, an in-person survey was conducted to assess whether or not the identified areas had agricultural activity. A random sample of 266 of the potential areas were selected. For the community gardens, efforts were made to sample the plots within the community garden. Only 12 of the plot surveys were completed in six community gardens. It was extremely difficult to identify the plot operators for an interview; finding the manager or coordinator of the community garden was much easier. It became clear that interviewing the operators of the individual plots would be cost prohibitive. Of the 266 sites, 71% (188) of the interviews were completed. In 21% (73) of the cases, the operator could not be found, and there was a 2% (5) non-response rate.

If the operator could not be contacted, the interviewer was asked to make an effort to observe whether or not agriculture was present. In 5% of the cases, the presence or absence of agriculture could not be ascertained from either an interview or observation. For about half (52%) of the sites, agriculture was present. During the 2012 census, half of the operations on the list frame were identified as farms; the other half were non-farms. Of course, some of those on the MACE list identified as having agricultural activity were not farms.

From the 188 completed interviews, 108 unique operations were identified with agricultural activity. Backyard or home gardens (34), school gardens (29), and community gardens (20) were the most common types of operations. Urban farms, vacant lot gardens, roof top gardens, aquaponics, hydroponics, and a commercial enterprise were also present.

People grew a variety of produce on urban agricultural sites, with fruits and vegetables being most common, but a sizeable proportion raised or kept farm animals and produced animal products. As one would expect, the areas dedicated to agriculture tended to be small, with about 2/3 having an area of less than 1000 square feet.

## 3. Results and Discussion

Several lessons were learned from the pilot study. Only areas of at least 36 square feet were considered during the pilot study because it was anticipated that satellite imagery would be used to identify agricultural sites in urban areas. Since satellite imagery did not prove to be useful at the 0.5-m resolution available, the minimum area requirement has been dropped for future work.

Although people could generally report the total area available for agricultural activity, they had difficulty specifying the land devoted to each purpose, such as tomatoes, peppers, bees, *etc*. Many were unable to quantify the amount of agricultural products produced, either in terms of weight or in value. This was particularly true for those who sold no produce. Thus, if the full range of agricultural activity is to be reported, it would be best to separately quantify those agricultural operations that qualify as a farm and those that do not. In addition, the questionnaire should undergo extensive testing and perhaps revision so that the respondents can provide more accurate data.

Community gardens are challenging. Whereas few, if any, plots qualify as a farm, the total output of the community garden may qualify as a farm. Although each individual plot may be correctly classified as a non-farm for all community gardens within an area, it may be perceived that the contribution of community gardens is under-estimated. Reporting the activity for the community gardens instead of the plots within the gardens would result in another set of challenges. In particular, reporting the demographics of those engaged in urban agriculture would be difficult. The manager of the community garden may not have the demographic details of the plot operators, such as age, sex, race, and ethnicity of the individual plot operators.

Although the list building approach has good potential for providing the foundation for improved estimates of urban agriculture, funding will not be available for an enhanced urban agriculture effort during the 2017 Census of Agriculture.

The web scraping approach to list building is being used for the NASS 2015 Local Foods Marketing Study, and local foods producers tend to live in or near urban areas. The MACE list is used to draw a second sample in a capture-recapture framework with the first sample coming from the NASS list frame. The data from the local foods study will provide good insights into the viability of this approach to list building.

A pilot study in the state of Washington is being conducted to explore the use of web scraping to build a list of small farms. Three NASS staff are working with MACE personnel for this project in an effort to build capacity for web scraping within NASS. In general, identifying all types of small farms continues to be a challenge. For the 2012 US Census of Agriculture, capture-recapture was used to adjust for coverage, non-response, and misclassification. Two independent samples were the Census mail list and the June area frame. Could web scraping be the foundation for a third sample? Finally, some challenging methodological issues need to be addressed if this listing building approach is to be used, along with a capture-recapture approach. For example, it is likely that the probability of capture varies with list.

With constant pressure to provide statistics on emerging sectors of agriculture in short time frames, it is important to identify new, cost-effective approaches to addressing the questions of interest to policy makers and other stakeholders. Web scraping, technology, and secondary data sources for this purpose may be tools that are used increasingly.

## REFERENCES


Baltimore Office of Sustainability. (2016) Urban Agriculture. http://www.baltimoresustainability.org/projects/baltimore-food-policy-initiative/homegrown-baltimore/urban-agriculture-2/ Accessed August 19, 2016.

Colasanti, Kathryn, Litjens, Charlotte, Hamm Michael. (2010) *Growing Food in the City: The Production Potential of Detroit's Vacant Land.* East Lansing, MI: The C.S. Mott Group for Sustainable Food Systems at Michigan State University.

Goldsmith, Stephen. (2014) Milwaukee's push to turn vacant land into urban farms. *Governing the States and Localityies.* April 16, 2014. http://www.governing.com/blogs/bfc/gov-milwaukee-mayor-tom-barrett-home-grown-vacant-lots-urban-agriculture.html Accessed August 19, 2016.

Santo, Raychel, Palmer, Anne, Kim, Brent. (2016) *Vacant Lots to Vibrant Plots: A Review of the Benefits and Limitations of Urban Agriculture*. Johns Hopkins Center for a Livable Future. http://www.jhsph.edu/research/centers-and-institutes/johns-hopkins-center-for-a-livable-future/_pdf/research/clf_reports/urban-ag-literature-review.pdf Accessed August 19, 2016.

Taylor, John R., Lovell, Sarah Taylor. (2014) Mapping public and private spaces of urban agriculture in Chicago through the analysis of high-resolution aerial images in Google Earth. *Landscape and Urban Planning* 108(1): 57–70.

USDA. (2016) *Urban Agriculture*. http://www.usda.gov/wps/portal/usda/knowyourfarmer?navid=kyf-urban-agric Accessed August 16, 2016.

Young, Linda J., Lamas, Andrea C., Abreu, Denise A., Wang, Shu, Adrian, Daniel. (2012) Statistical Methodology for the 2012 US Census of Agriculture. http://www.statistics.gov.hk/wsc/IPS105-P3-S.pdf Accessed August 19, 2016.