



FIRE: THE FIRST-YEAR INNOVATION & RESEARCH EXPERIENCE

2020 APHIS-FIRE Summer Fellow Project:

Using Online Data to Identify Farms Selling Poultry and Eggs in Delaware, Maryland, and Virginia

Final Report

This project is a collaboration between the United States Department of Agriculture (USDA), Animal and Plant Health Inspection Service (APHIS), and the University of Maryland (UMD) [First-Year Innovation & Research Experience \(FIRE\)](#) to support endeavors that will promote the agricultural sciences and related fields within undergraduate research at UMD.

Each year, several FIRE students are selected to become Summer Fellows to work alongside the Research Educator for eight weeks. In 2020, four summer fellows, Anna Lazarus-Hall, Janna Chapman, Mishti Relan, and Shanil Kothari, worked under the guidance of the Research Educator, Dr. Thanicha Ruangmas, and Faculty Advisor, Dr. Lars Olson. The research group collaborated with the USDA and APHIS to identify farms selling poultry and eggs in Delaware, Maryland, and Virginia using publicly available online data. The research group has identified new methods to detect registered and unregistered farms selling poultry and egg products and regional trends. These methods could aid in identifying geographical areas prone to infection as well as the outreach of information.

As a result of this collaboration, the research group as produced the following deliverables:

1. An application with interactive maps showing farms selling poultry and egg products from online registries and Google Places: The online application can provide information about farms within a user-identified radius of a latitude and longitude centroid. A temporary version of this application can be found at <https://umdfire-aphis.shinyapps.io/farmsapp/>. In order to access the website, you need to be invited and create a ShinyApps.io account. Instructions on how to run the

application offline in a local computer and how to deploy the online application can be found in the README.txt file in the “1. APP” folder submitted together with this report.

2. Spreadsheets with address and contact information of farms selling poultry and egg products from federal, state, online directories and Google Places. These spreadsheets can be found in the “2. SPREADSHEETS” folder submitted together with this report.
3. A report that outlines methodology, best practices, and results from scraping data from online forums, public registries, Google Places, and Google Trends. This can be found in the “3. REPORT” folder.
4. A folder of codes in R that allows all results to be replicated. All the codes and accompanying flowcharts can be found in the “4. CODES” folder.



FIRE: THE FIRST-YEAR INNOVATION & RESEARCH EXPERIENCE

Written Report for 2020 APHIS-FIRE Summer Fellow Project:

Using Online Data to Identify Farms Selling Poultry and Eggs in Delaware, Maryland, and Virginia

Table of Contents	3
Introduction	5
1. Farms Selling Poultry and Eggs from Registries and Online Directories	5
1.1. Obtaining Data from State Registries	6
Maryland	6
Virginia	6
Delaware	7
1.2. Data Scraping Methodology and Best Practices	7
1.3. Findings	9
2. Farms Selling Poultry and Eggs from Google Places	11
2.1. Data Scraping Methodology	11
Preparation Before Scraping	11
Using Google Places API to Scrape	11
Obtaining Details of Each Place	12
2.2 Findings	12
3. Backyard Chicken Farmers from Forums	13
3.1. Data Scraping Methodology	13
Robots File	13
The Scraping Process	14
3.2. Findings	14
4. Regional and Seasonal Interest from Google Trends	15
4.1. Using Google Trends	15

4.2. Findings	16
Keywords Related to “Backyard Poultry”	16
Regional Interest Relative to the National Maximum	16
Seasonal Interest within Each Metro Area	17
Keywords Related to “Eggs”	18
Regional Interest Relative to the National Maximum	18
Seasonal Interest within Each Metro Area	19
Keywords Related to “Poultry For Sale”	20
Regional Interest Relative to the National Maximum	20
Seasonal Interest within Each Metro Area	21
Keywords Related to “Keeping Chickens”	23
Regional Interest Relative to the National Maximum	23
Seasonal Interest within Each Metro Area	24
Keywords Related to “Showing or Selling Poultry”	25
Regional Interest Relative to the National Maximum	26
Seasonal Interest within Each Metro Area	26
Keywords Related to “Types of Poultry”	28
Regional Interest Relative to the National Maximum	28
Seasonal Interest within Each Metro Area	29
Conclusion	29
Authors	31

Introduction

This collaboration's initial goal was to identify "gray markets" of live poultry and eggs in the DelMarVa Peninsula using data from online social media platforms. However, we learned early on that major social media websites do not allow data scraping under their terms of use. This prohibited us from identifying individuals raising or selling live poultry and eggs. Instead, we were able to use data from various federal and state registries, online directories, and Google Places to identify farms selling live poultry and eggs for the entire state of Delaware, Maryland, and Virginia. In addition to farm location, we were able to find their contact numbers and websites, which can aid in disseminating information during an outbreak. Furthermore, we find that many farms listed in online directories do not appear to be registered with APHIS' National Poultry Improvement Plan or the state registries that we were able to obtain; however, we did not obtain comprehensive poultry premise registry data for Delaware and Maryland.

To understand seasonal and regional trends in live poultry and eggs, we use data from Google Trends to monitor Google searches for various keywords. We found that the metro area of Richmond-Petersburg has one of the highest interest for fresh eggs in the nation, and the metro area of Bluefield-Beckley Oak Hill WV has one of the highest interest in chicken coop in the nation. Other areas in Delaware, Maryland, and Virginia have moderate interest for other terms. Within each area, we find that terms associated with chickens for sale tend to peak at the beginning at the end of each year. In addition, terms related to keeping chickens and eggs peaked around April to June.

Throughout this research project, the research group has identified new methods to detect registered and unregistered farms selling poultry and egg products and regional trends. We hope that these methods could help identify geographical areas prone to infection and the outreach of information for APHIS and other organizations.

1. Farms Selling Poultry and Eggs from Registries and Online Directories

We collected and consolidated data from federal, state, and online directories to obtain data of farms selling poultry and eggs. Below is a list of our data sources.

- [APHIS's National Poultry Improvement Plan](#) (NPIP)
- [Maryland Department of Agriculture](#) (MDA)

- [Virginia Department of Agriculture and Consumer Services](#) (VA Grown)
- [Local Hens](#)
- [Local Harvest](#)
- [Maryland's Best](#)

We used the programming language R to scrape farm data from each source and used Google Maps Geocode to obtain the formatted address and coordinates of each farm. Afterward, we consolidated data from each source. In the end, we were able to gather about 1,566 unique farms in Delaware, Maryland, and Virginia. Information about each farm's location and contact information (if available) is in the "allregistries.csv" spreadsheet submitted together with this report.

The rest of this section describes how we were able to obtain data from the three states and best practices for scraping data so that this methodology can be applied to find information about farms in other states. The end of this section also includes our findings from consolidating all farm information.

1.1. Obtaining Data from State Registries

In general, information from other states could be found by going to the state of interest's website for agriculture and looking for information about animals and poultry. If the data does not present itself clearly, there is contact information for a registry or animal department or a questions and comments department where the information may be requested.

Maryland

Data from Maryland was obtained from the Maryland Department of Agriculture's Egg Inspection Program. The PDF registry from January 2020 contained a list of packers and wholesalers registered with Maryland. Wholesalers in this dataset were located in many different states, so only locations in Maryland, Virginia, and Delaware were used.

Virginia

Data from Virginia was obtained from the Virginia Department of Agriculture and Consumer Services. Their website, known as VA Grown, contained the names, addresses, and phone numbers of farms selling different products, including poultry, meat, produce, and eggs. We found a list of farms that reported selling chicken and eggs. The list of these farms was exported into two separate PDF forms (one for each chicken, and one for eggs) and then to CSV files.

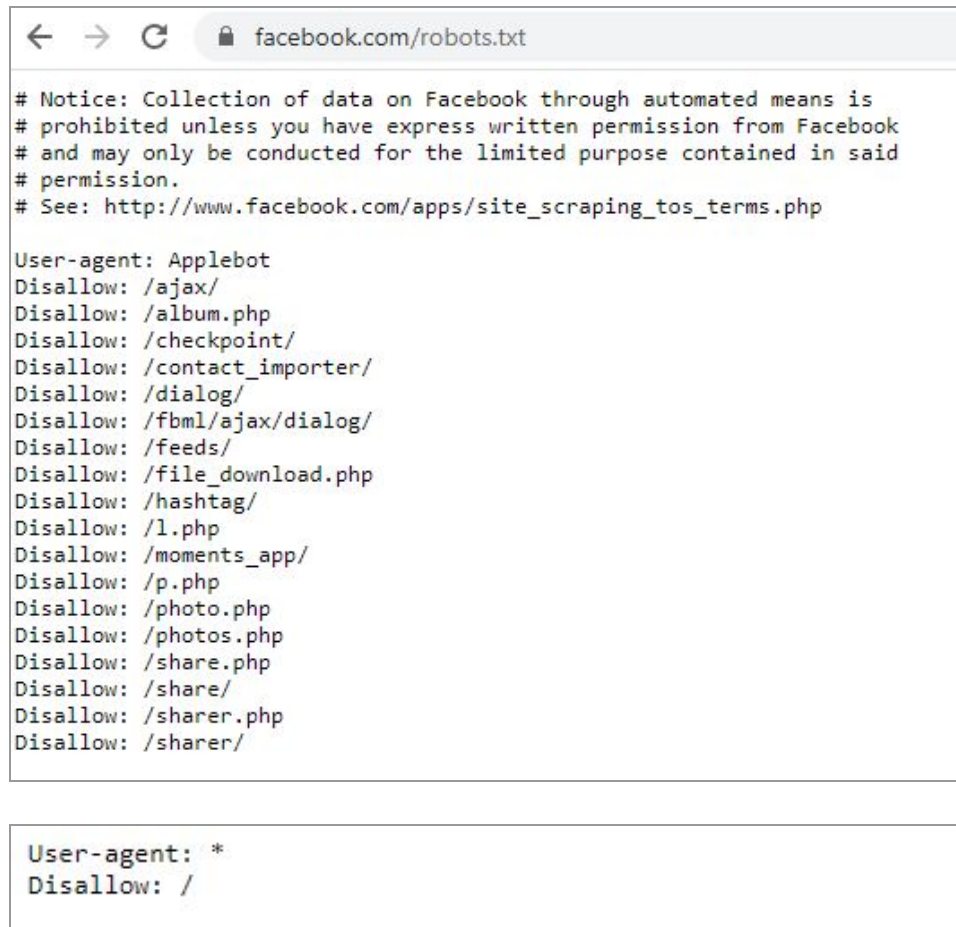
Delaware

We were unable to obtain data from Delaware. A Freedom of Information Act (FOIA) request was made on the state government website to access the list of registered packers and wholesalers. The request was denied due to state law that public records may only be requested by Delaware citizens. Another attempt was made to obtain the information using the “Questions and Complaints” email address to ask if it was possible to get the information, but were still unable to access the information. The only state registered data used for Delaware was from the Maryland Department of Agriculture.

1.2. Data Scraping Methodology and Best Practices

While the list of farms from NPIP and MDA are initially in a PDF format, data from other sources are listed online. However, before scraping is done, make sure that scraping is legal and ethical for each web page by going through robots.txt of the domain. This file lists the pages of a website that can or can't be scraped and any restrictions on the search engines that are allowed to scrape. The robots.txt file can be accessed by using the [get_robotstxt](#) function in R. Sometimes R could fail to load this file. If that is the case, open the robots.txt file in your browser by manually inspecting the robots' site, such as <https://facebook.com/robots.txt>.

Figure 1 shows Facebook's robots file. In the top panel, the general terms of use is shown. The robots file disallows scraping a number of specific features from the Applebot. The last rows of text in the robots file is shown in the bottom panel, which disallows the scraping for all other features by all other users.



```
< > ↻ facebook.com/robots.txt

# Notice: Collection of data on Facebook through automated means is
# prohibited unless you have express written permission from Facebook
# and may only be conducted for the limited purpose contained in said
# permission.
# See: http://www.facebook.com/apps/site_scraping_tos_terms.php

User-agent: Applebot
Disallow: /ajax/
Disallow: /album.php
Disallow: /checkpoint/
Disallow: /contact_importer/
Disallow: /dialog/
Disallow: /fbml/ajax/dialog/
Disallow: /feeds/
Disallow: /file_download.php
Disallow: /hashtag/
Disallow: /l.php
Disallow: /moments_app/
Disallow: /p.php
Disallow: /photo.php
Disallow: /photos.php
Disallow: /share.php
Disallow: /share/
Disallow: /sharer.php
Disallow: /sharer/

User-agent: *
Disallow: /
```

Figure 1: Robots File from Facebook

Instead of using major social media websites, we were able to use data from online directories such as Local Hens. We scraped data from VA Grown and Local Hens by following the steps below.

1. Identify the web page that we are scraping. After that, Load and parse the HTML file from that web page using the [read_html](#) from xml2 package.
2. Use [selectorgadget](#) to define a CSS element of interest.
3. Extract data from the web page by using [html_nodes](#) and [html_text](#) from the rvest package.

Alternatively, we used [ParseHub](#) to scrape data from Maryland's Best and Local Harvest. [ParseHub](#) is a powerful web scraping software that is available online. Their machine learning algorithm automatically transforms online data into spreadsheets.

After each farm information was transformed into a clean spreadsheet, we used [Google Maps Geocode API](#) tool to find the formatted address and coordinates of each farm. Because the

same farm address may be written differently in each directory, we use Google Maps Geocode API to autocorrect each farm's address so that it can be merged across different datasets.

It might be clearer to look at the codes to understand all the components involved in scraping and consolidating data from the online directories. Brief description of the detailed codes of each step are available in this report. The flowchart below shows the codes and CSV files associated with scraping and consolidating data from online directories. A larger version of the flowchart is submitted with the codes.

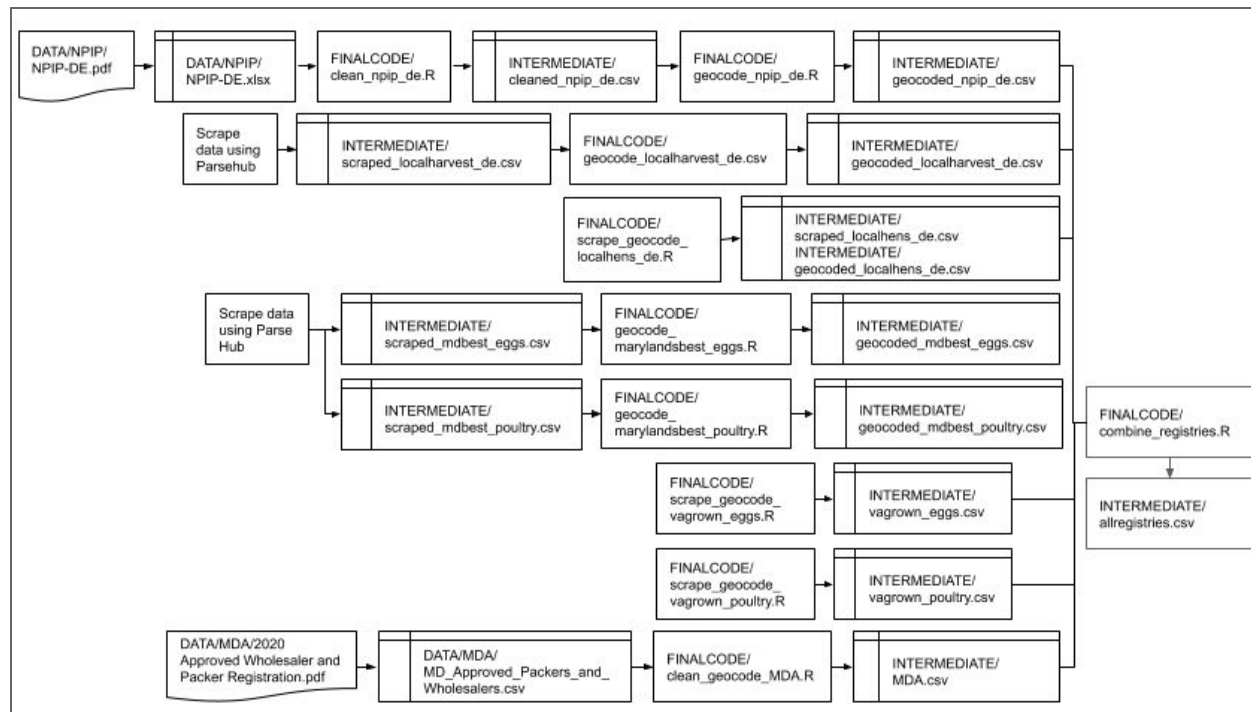


Figure 2: Consolidating Online Directories Flowchart

1.3. Findings

We were able to find the coordinates of 1,493 farms out of 1,566 farms in Delaware, Maryland, and Virginia. A map of the farms and the source of their information is Figure 3. An interactive map with a layout of major roads can be found in the application submitted together with this report.

Surprisingly, there is minimal overlap between different directories. Although all poultry owners are required to register with the state of Maryland and Virginia, many farms listed in Local Harvest, Local Hens, and Maryland's Best did not appear in the state registries that we were able to obtain. Specifically, we find 286 farms in Maryland (33.8% of all farms in Maryland

based on our data) did not register with the MDA. 119 farms in Virginia (20.9% of all farms in Virginia based on our data) did not register with the Virginia Department of Agriculture.

There is minimal overlap between farms that are registered with APHIS's NPIP and other directories. Of the 272 farms that are registered in APHIS's NPIP in Delaware, Maryland, and Virginia, only 4 are registered with the MDA, only 1 is registered with the Virginia Department of Agriculture, and only 1 is found in other directories. All 6 NPIP farms that are found in other directories are categorized as Subpart E (Special Provisions for Hobbyist and Exhibition Waterfowl, Exhibition Poultry, and Game Bird Breeding Flocks and Products). Figure 4 is a pie chart showing the number of farms that is registered in each directory.

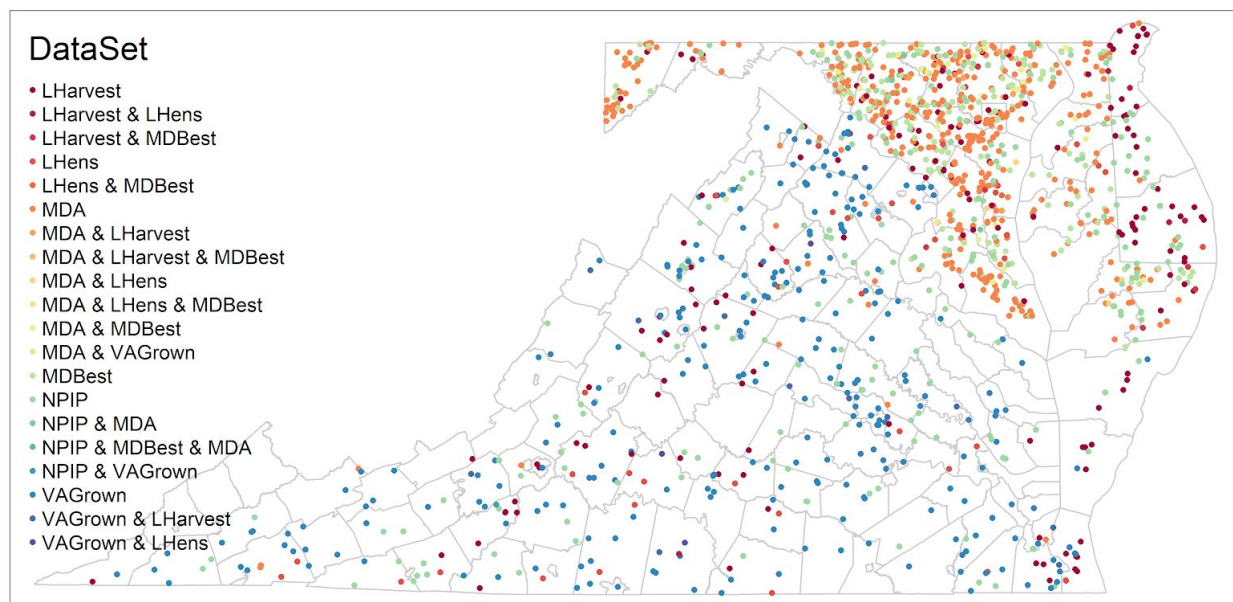


Figure 3: Map of Farms from Online Directories

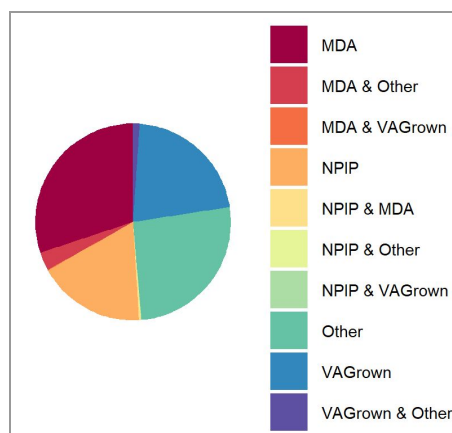


Figure 4: Fraction of Farms in Each Directory

2. Farms Selling Poultry and Eggs from Google Places

An alternative way to detect farms is to use [Google Places API](#). The tool can be used to provide a list of places found on Google Maps associated with a keyword. The user can provide a centroid, radius, and keyword to query all locations on Google Maps associated with the keyword surrounding a fixed radius around a centroid. In addition to the coordinates, we can obtain the location name, address, contact information, business status, and the top 5 reviews of each location.

For this project, we use Google Places API to find all locations in Delaware, Maryland, and Virginia associated with the following keywords: “baby chicks for sale,” “baby ducklings for sale,” “baby poultry for sale,” “baby poult for sale,” “chick hatchery,” “poultry hatchery,” “county fair,” “farm fair,” “state fair,” “farmers market,” “flea market,” “farms,” “poultry show.”

The rest of this section describes the methodology of obtaining data from Google Places API and a brief summary of our findings.

2.1. Data Scraping Methodology

Preparation Before Scraping

First, we need to identify the places that we would like to get. This includes getting the search term (ex. “farms”), the location (coordinates of the center), the radius (between 10 km and 50 km), and obtain an API key from [Google Cloud Platform](#). After getting the API key, we [enabled the API](#), which allows us to get the results from google.

After this, we used [this tool](#) available at mapdevelopers.com to identify centroids in which the 20 km radius would cover the entire Delaware, Maryland, and Virginia area. So, now we have the search term, coordinates, radius, and API key.

Using Google Places API to Scrape

The R package `googleway` allows us to obtain the places from google. The function [google_places](#) has to be used with the search term, coordinates, radius, and API key as arguments. It returns 20 places at once, and if there is another page (i.e., more results), we get a “next_page_token.” This has to be used to run the [google_places](#) function next time as the fifth argument. The best way to go about this is to put it inside a loop, so we do not need to run it constantly. The function with the “next_page_token” would return the next page of results.

This can also be done for multiple coordinates in a loop; this will make it completely automated and return all the places in the range from the coordinates for a search term. We could also add multiple search terms and loop it for all the search terms in all the coordinates.

Obtaining Details of Each Place

By now, we should have all the matching places in one table. We can use [google_place_details](#) and have the “place_id” we got from the [google_places](#) and API key as the arguments. It will return a list of all details of the place, including the reviews of the place. We can collect all this information and merge it with the original data we got from the [google_places](#). We can loop this for all the places we got from the [google_places](#).

The details can be used to obtain contact information, and reviews can be matched with specific words to check for relevance to backyard chicken in this case.

All the codes we used to obtain data from Google Places API are submitted together with this report. The flowchart below shows the codes and CSV files associated with scraping data from Google Places API. A spreadsheet of all locations can be found in the file, “places_reviews.csv” submitted together with this report.

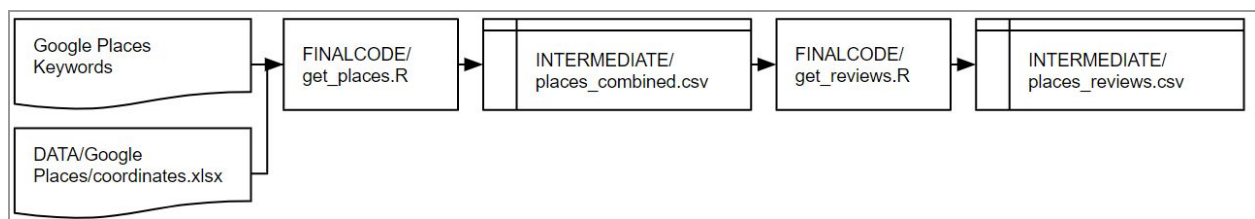


Figure 5: Google Places API Flowchart

2.2 Findings

Using Google Places API allowed us to find numerous locations, possibly selling poultry and eggs. For example, we are able to find 4,499 locations associated with the keyword “farms” in Delaware, Maryland, and Virginia. Because not all of these locations may be selling poultry and eggs, we count the number of times the term “chick,” “egg,” “poult,” and “bird” appears in the top 5 reviews. Figure 6 shows a map of all locations associated with the keyword “farms” and the colors of each dot represents the number of times the term “chick,” “egg,” “poult,” or “bird,” appears in the top 5 reviews. An interactive map with all other keywords that we have queried can be found in the application submitted together with this report.

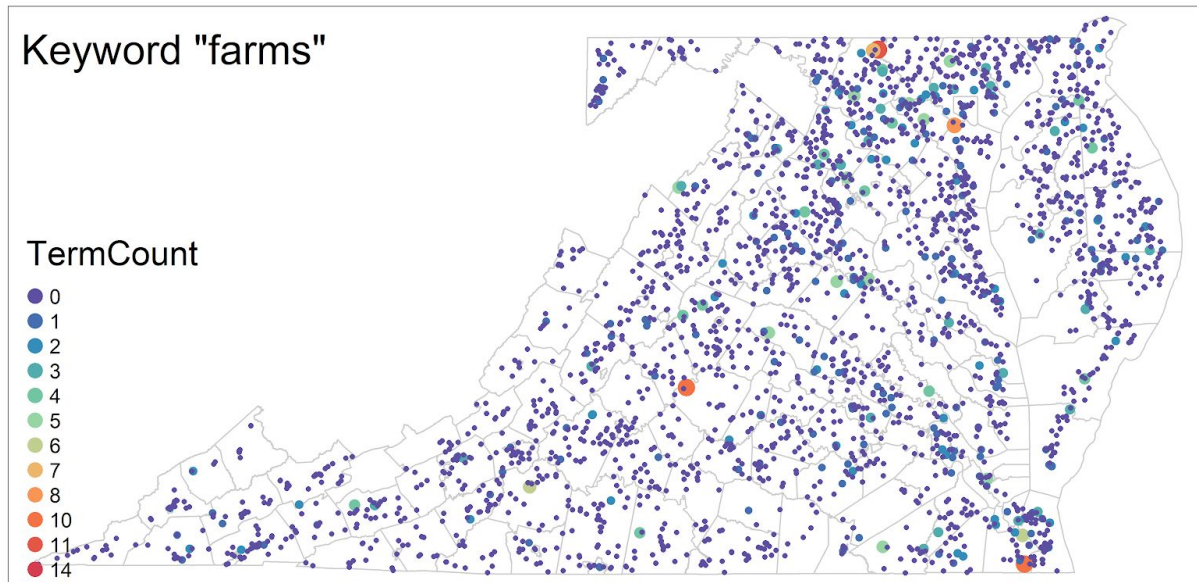


Figure 6: Farms from Google Places API

3. Backyard Chicken Farmers from Forums

Forums such as [backyardchickens.com](https://www.backyardchickens.com/) contain numerous information about where people are raising backyard chickens. However, their robots.txt file disallows scraping data from posts. However, we were able to scrape data from top members from <https://www.backyardchickens.com/> and <https://www.chickenforum.com/>. Many times, each member would indicate their location in their profile, and we can use that information to identify hotspots of backyard chickens in Delaware, Maryland, and Virginia.

The rest of this section outlines the methodology in scraping member data from forums and our brief findings.

3.1. Data Scraping Methodology

Robots File

As mentioned earlier in Section 1.2, before any scraping is done, make sure that scraping is legal and ethical for each web page by going through robots.txt of the domain.

While checking, make sure the part of the website you want to scrape is allowed. The robots.txt file will mention if the part of the website you want to scrape is allowed. Many times, the website gives out the things which are disallowed, and the rest would be allowed. For

forums, like any social media, the websites do not allow the posts, groups, or user data to be scraped. And anything which requires you to log in is mostly not allowed to scrape either.

Most forums allow articles and member leaderboards of most active and notable members to be scraped. The member pages can be used to scrape the location of top users to know the popularity and activity of backyard chickens in a region.

The Scraping Process

We scraped data from forums by following the same steps outlined in section 1.2. All the codes we used to obtain data from forums are submitted together with this report. The flowchart below shows the codes and CSV files associated with scraping data forums.

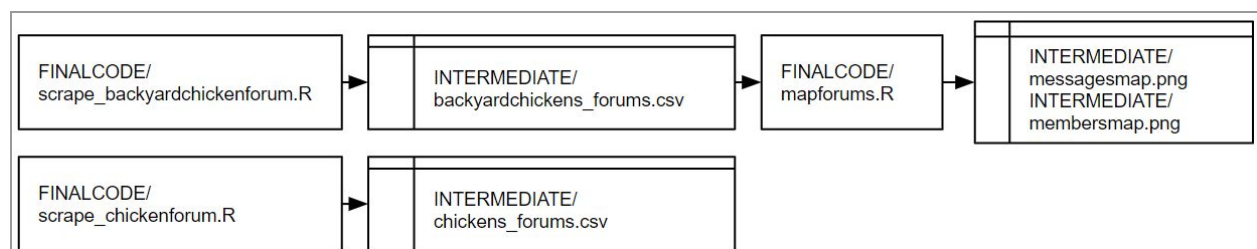


Figure 7: Scraping Forums Flowchart

3.2. Findings

In the end, we were not able to find beneficial information from scraping forums. First, we could not obtain data from all members, but just members with the most messages, most featured threads, most likes, most points, most showcase items, most media, most albums, most articles, and staff members.

Also, not many members indicate specific information about their location in their profile. Of 583 members that we were able to obtain information from <https://www.backyardchickens.com/> and <https://www.chickenforum.com/>, only 20 reported that they live in Delaware, Maryland, Virginia. Figure 8 shows the number of members from each state and the number of messages posted.

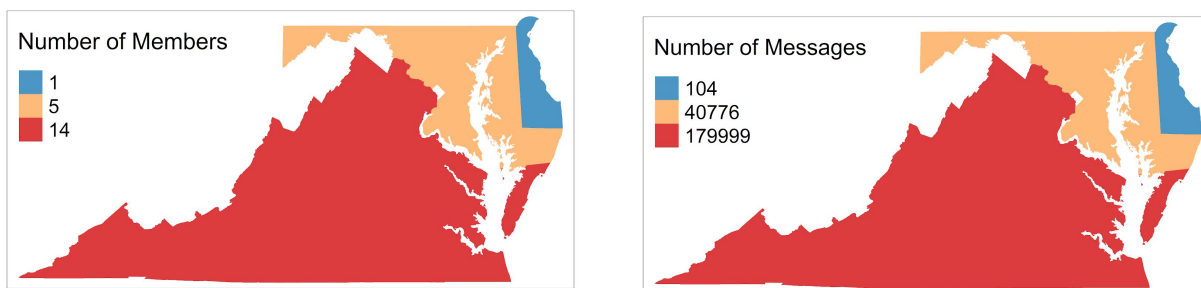


Figure 8: Number of Members and Messages from Forums

4. Regional and Seasonal Interest from Google Trends

Besides finding farms selling poultry and eggs, we can gauge regional and seasonal interest in backyard chickens using data from Google Trends. The tool provides the relative number of Google searches for each keyword in each area relative to the nation, or the number of Google searches for each keyword in each area.

The rest of this section describes best practices that we have learned from using Google Trends and analysis of using Google Trends to detect specific keywords related to backyard chicken in Delaware, Maryland, and Virginia.

4.1. Using Google Trends

The Google Trends Help Center (<https://support.google.com/trends/?hl=en#topic=4365530>) is very useful in getting started using Google Trends, particularly the “Explore Trends” tab. One major point to keep in mind while using Google Trends is that everything is relative. The results from each search are relative only to the keyword, location and time range chosen for that individual search. For example, the interest over time of a search for a keyword with the location set to a metro area will show interest levels reaching 100. However, the interest by DMA (designated market area) of another search for the same keyword with the location set to the entire United States may show that same metro area having an interest level of something like 26. This means that although the interest in that metro reaches 100 in the first search, it is only compared to other dates in that metro; search interest may still be low when compared to other locations. Given this, interest over time for separate searches should not be compared. However, up to five keywords can be compared by clicking the “+” button when choosing a keyword.

Another important tip for using Google Trends is to take a look at the related topics and related queries. These show you the kinds of searches included in the data you have collected. If the list

of keywords you are searching is small, it is definitely worthwhile to check and make sure the related topics and related queries match the topic that you are searching for. For example, I searched “incubator” and the top related topics included Pokemon. However, if you notice a particular unrelated topic comes up often for a certain keyword, you can type “keyword - unrelated keyword” in the search term in order to avoid unrelated searches in your results. Depending on the search term, it may also be beneficial to change the category in order to reflect the topic you are searching for. For example, the category “Pets & Animals”, and the category “Animal Products & Services” underneath that more specifically, may show better related results for search terms of interest to you. Throughout the process of collecting and analyzing data from Google Trends for the keywords related to this project, I found that the “Animal Products & Services” category was too specific in many cases. However, the “Pets & Animals” category seemed efficient in eliminating unrelated searches.

It should also be noted that time ranges that are three months or shorter will produce daily data and time periods that are five years or longer will produce monthly data when showing interest over time. Any length between these two constraints will produce weekly data.

4.2. Findings

We obtained Google Trends data for multiple keywords related to backyard chickens and grouped them into 6 categories: “backyard poultry,” “eggs,” “poultry for sale,” “keeping chickens,” “showing or selling poultry,” and “types of poultry.”

For each of the keywords, we found the average relative hits for each of the metro areas from 2010 through 2019. The results for regional trends for keywords with enough searches for data to be generated are mapped in Figure 9, 11, 13, 15, 17, and 19. Within each region, seasonal trends can be found in Figure 10, 12, 14, 16, 18, and 20.

Keywords Related to “Backyard Poultry”

Keywords searched: Raising chickens, raising ducklings, backyard chickens, backyard flock, small flock

Regional Interest Relative to the National Maximum

Throughout Maryland, Virginia, and Delaware, the metro areas of Washington DC and Philadelphia PA (metro areas of Philadelphia also includes some areas in the state of Delaware) are the two metro areas that showed interest, though mild, in the search term “backyard chickens”. Philadelphia PA is the only metro area that showed any interest in the search term “raising chickens” from 2010 through 2019. This is shown in Figure 9.

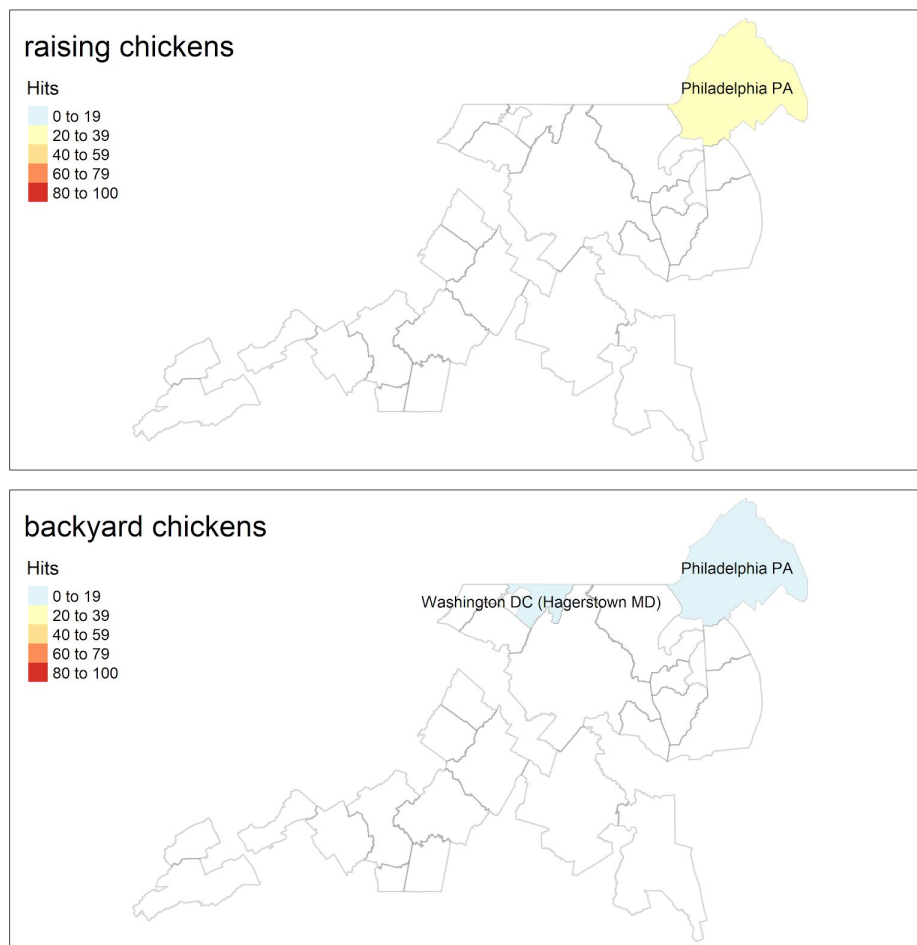


Figure 9: Regional Interest For Keywords “raising chickens” and “backyard chickens”

Note: The maps above illustrate interest in the chosen keyword relative to the national interest from 2010 through 2019. The number of searches, or “Hits”, in each metro area is scaled from 0 to 100. For example, the United States metro area with the highest number of searches would have 100 “Hits”. It is worth noting that metro areas without any color did not show enough interest in the keyword to collect any data. Also, this search interest data was collected with the category “Pets & Animals” specified.

Seasonal Interest within Each Metro Area

There is enough data for keywords related to backyard chickens to be analyzed only in the metro areas of Washington DC and Philadelphia PA. In Washington DC, interest in the search term “backyard chickens” tended to start the year on an incline, peak around the end of April, and then decline throughout the rest of the year. In Philadelphia PA, interest in this keyword tended to peak around the end of April, dip down to its lowest point around the end of October, then increase back to its starting point by the end of the year.

Search interest in the term “raising chickens” in Washington DC tended to peak around the end of March, dip down to a low point around the beginning of June, and then remain steady throughout the rest of the year. In Philadelphia PA, interest in this keyword tended to start the

year on a slight incline, peak around the end of May, reach a trough around the end of October, and then increase back to its starting point by the end of the year. This is shown in Figure 10.

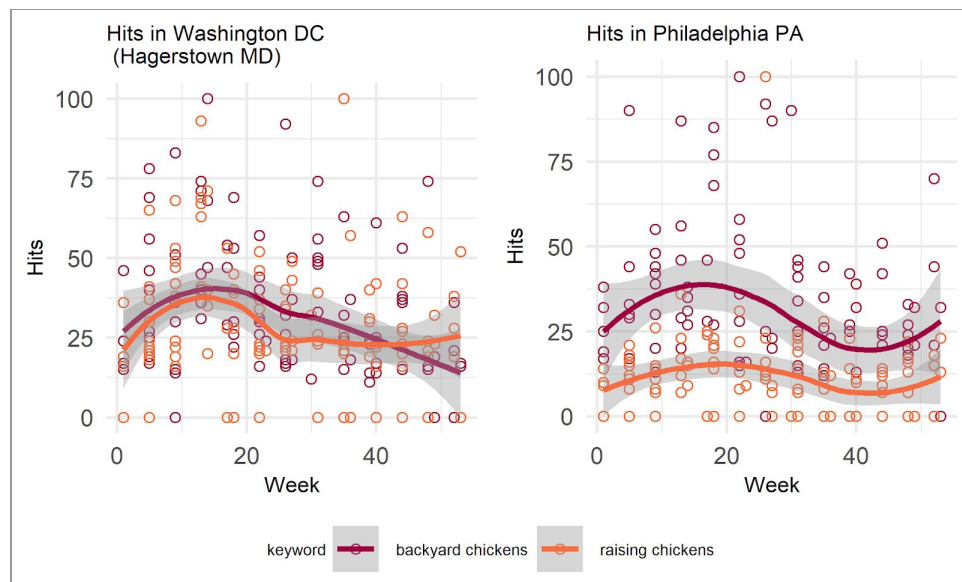


Figure 10: Seasonal Interest For Keywords “raising chickens” and “backyard chickens”

Note: The plots above show seasonal trends for searches across each year. Each dot represents the weekly search interest for each keyword in each area, relative to the maximum weekly search interest in that area. Each line represents the locally weighted smoothing of hits in that week across 10 years. This search interest data was collected with the category “Pets & Animals” specified. Because hits are relative to the maximum hits of the same keyword and area, hits between keywords should not be compared. For example, when the trend for one keyword is greater than the trend for another does not mean that there are more searchers for one relative to the other. They are plotted in the same graph in order to compare the differences in seasonal search interest.

Keywords Related to “Eggs”

Keywords searched: Poultry hatchery, chick hatchery, fresh eggs, hatching eggs, incubator

Regional Interest Relative to the National Maximum

More metro areas showed interest in the search term “incubator” compared to the search terms “fresh eggs” and “hatching eggs”. It is worth noting that Richmond-Petersburg VA had an extremely high interest in the search term “fresh eggs” compared to the rest of the nation. Washington DC and Philadelphia PA were the only metro areas that showed interest in all three of the search terms, as shown in Figure 11.

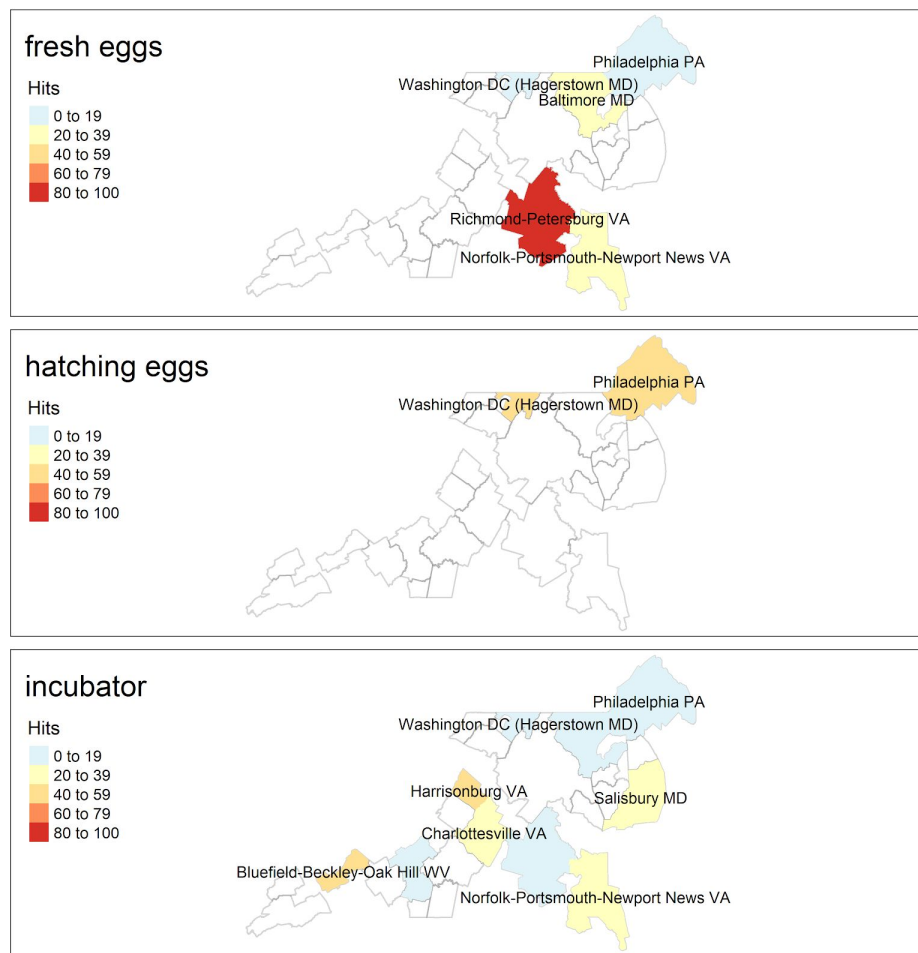


Figure 11: Regional Interest For Keywords “fresh eggs,” “hatching eggs,” and “incubator”

Note: The maps above illustrate interest in the chosen keyword relative to the national interest from 2010 through 2019. The number of searches, or “Hits”, in each metro area is scaled from 0 to 100. For example, the United States metro area with the highest number of searches would have 100 “Hits”. It is worth noting that metro areas without any color did not show enough interest in the keyword to collect any data. Also, this search interest data was collected with the category “Pets & Animals” specified.

Seasonal Interest within Each Metro Area

We were able to find seasonal trend data for the keyword “incubator” in Washington DC, Bluefield-Beckley-Oak Hill WV (this includes areas of Virginia as well), Charlottesville VA, Harrisonburg VA, Philadelphia PA, Richmond-Petersburg VA, and Salisbury MD. As all the trends are similar, we only show the seasonal trend for this keyword in Washington DC in Figure 12. In general, search interest in the term “incubator” tended to start the year on a slight incline, peak in the beginning of May, and then decline throughout the rest of the year.

We were able to find seasonal trend data for the keyword “fresh eggs” in Washington DC and Baltimore MD. In both metros, search interest in the term “fresh eggs” tended to remain

relatively steady throughout the year, peaking slightly around the end of June in Washington DC and the middle of April in Baltimore MD.

We were able to find seasonal trend data for the keyword “hatching eggs” in Washington DC. Search interest in the term “hatching eggs” tended to peak around the end of May and then reach a trough around the beginning of November, as shown in Figure 12.

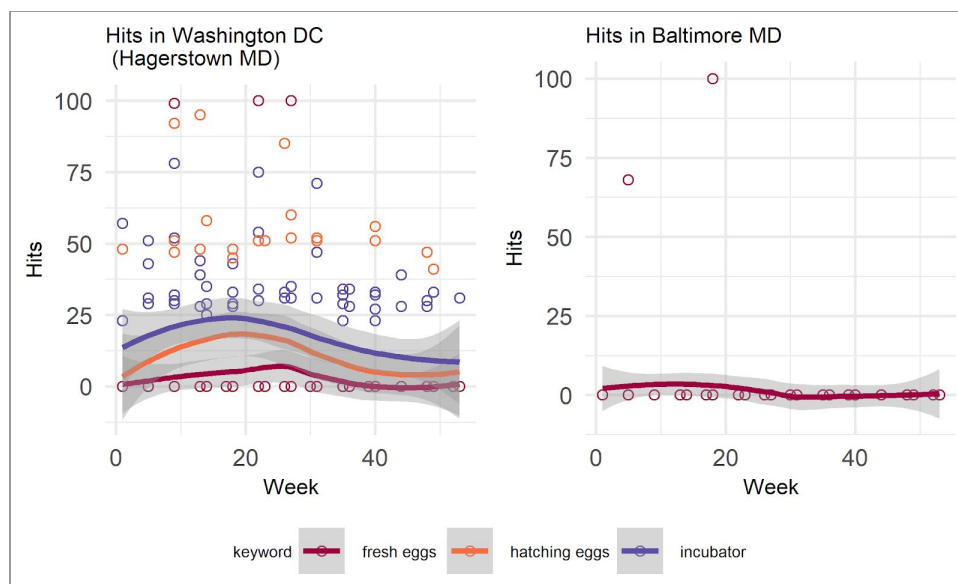


Figure 12: Seasonal Interest For Keywords “fresh eggs,” “hatching eggs,” and “incubator”

Note: The plots above show seasonal trends for searches across each year. Each dot represents the weekly search interest for each keyword in each area, relative to the maximum weekly search interest in that area. Each line represents the locally weighted smoothing of hits in that week across 10 years. This search interest data was collected with the category “Pets & Animals” specified. Because hits are relative to the maximum hits of the same keyword and area, hits between keywords should not be compared. For example, when the trend for one keyword is greater than the trend for another does not mean that there are more searchers for one relative to the other. They are plotted in the same graph in order to compare the differences in seasonal search interest.

Keywords Related to “Poultry For Sale”

Keywords searched: Baby poultry for sale, chicks for sale, chickens for sale, poults for sale, ducklings for sale

Regional Interest Relative to the National Maximum

Both Washington DC and Philadelphia PA showed interest in many of the search terms related to “for sale”. The most number of metro areas that showed interest in one of these search terms was for “chickens for sale”. However, it is worth noting that a relatively high level of interest compared to the rest of the nation was shown for the term “ducklings for sale” in Philadelphia PA. This is shown in Figure 13.

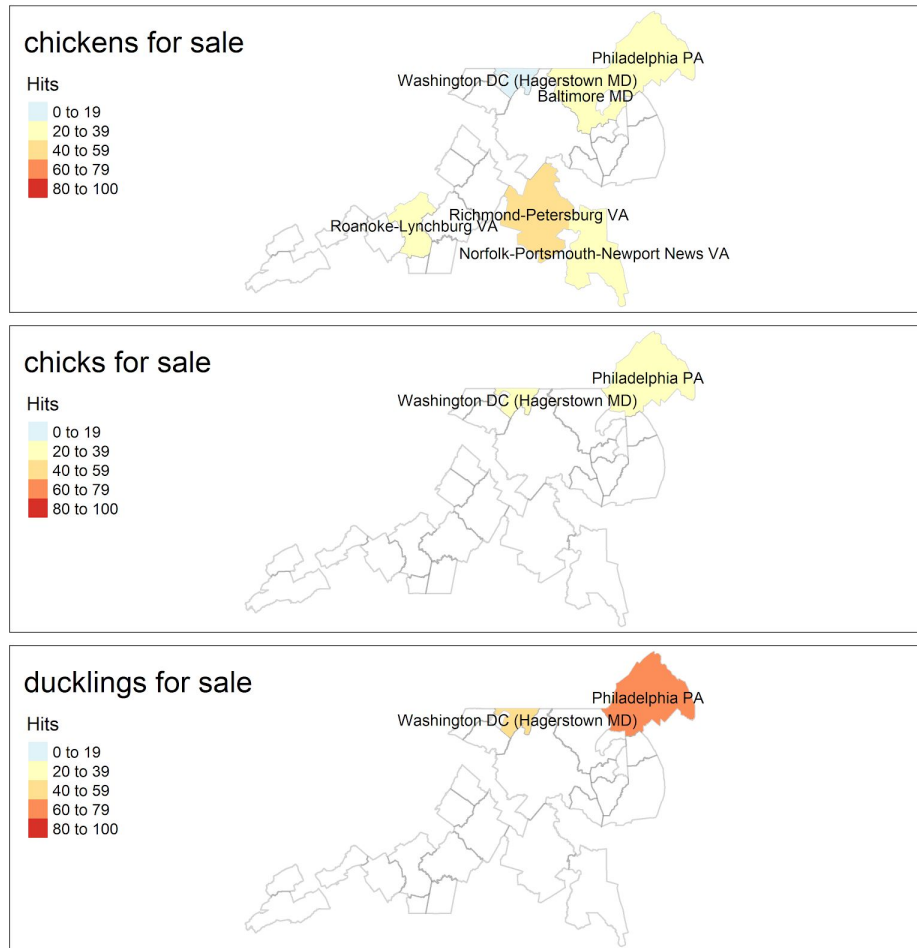


Figure 13: Regional Interest For Keywords “chickens for sale,” “chicks for sale,” and “ducklings for sale”

Note: The maps above illustrate interest in the chosen keyword relative to the national interest from 2010 through 2019. The number of searches, or “Hits”, in each metro area is scaled from 0 to 100. For example, the United States metro area with the highest number of searches would have 100 “Hits”. It is worth noting that metro areas without any color did not show enough interest in the keyword to collect any data. Also, this search interest data was collected with the category “Pets & Animals” specified.

Seasonal Interest within Each Metro Area

We were able to find seasonal trend data for the keyword “chickens for sale” in Washington DC, Baltimore MD, Norfolk-Portsmouth-Newport News VA, Philadelphia PA, and Richmond-Petersburg VA. In Norfolk-Portsmouth-Newport News VA, Baltimore MD, and Roanoke-Lynchburg VA, interest in the search term “chickens for sale” tended to peak around the end of March, reach a low point between the middle of September and the middle of October, and then incline again throughout the rest of the year. In Washington DC, interest in

this search term tended to peak around the middle of April and then decline throughout the rest of the year.

We were able to find seasonal trend data for the keyword “chicks for sale” only in Washington DC. Search interest in the term “chicks for sale” tended to peak around the end of March and then incline throughout the rest of the year, as shown in Figure 14.

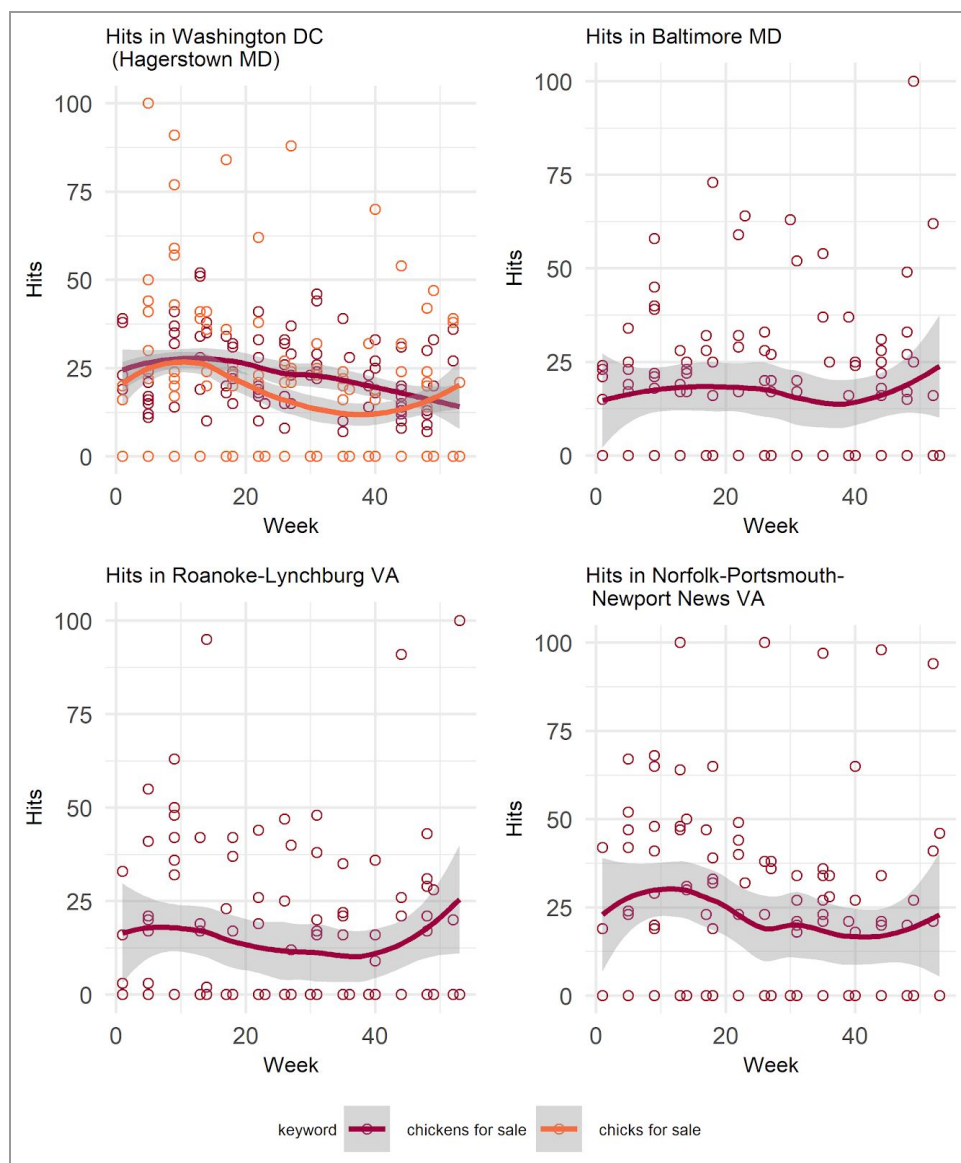


Figure 14: Seasonal Interest for Keywords “chickens for sale,” and “chicks for sale”

Note: The plots above show seasonal trends for searches across each year. Each dot represents the weekly search interest for each keyword in each area, relative to the maximum weekly search interest in that area. Each line represents the locally weighted smoothing of hits in that week across 10 years. This search interest data was collected with the category “Pets & Animals” specified. Because hits are relative to the maximum hits of the same keyword and area, hits between keywords should not be compared. For example, when the trend for one keyword is greater

than the trend for another does not mean that there are more searchers for one relative to the other. They are plotted in the same graph in order to compare the differences in seasonal search interest.

Keywords Related to “Keeping Chickens”

Keywords searched: Keeping chickens, chicken coop, chicken tractor

Regional Interest Relative to the National Maximum

Compared to the nation as a whole, Bluefield-Beckley Oak Hill WV showed extremely high interest in the search term “chicken coop”. This search term seems to be somewhat popular among metro areas in the three states, though interest levels varied throughout. Charlottesville VA and Harrisonburg VA showed relatively high interest in this search term as well. For the search term “chicken tractor”, which people in the three states showed much less interest in compared to the rest of the nation, Philadelphia PA and Washington DC showed mild interest. This is shown in Figure 15.

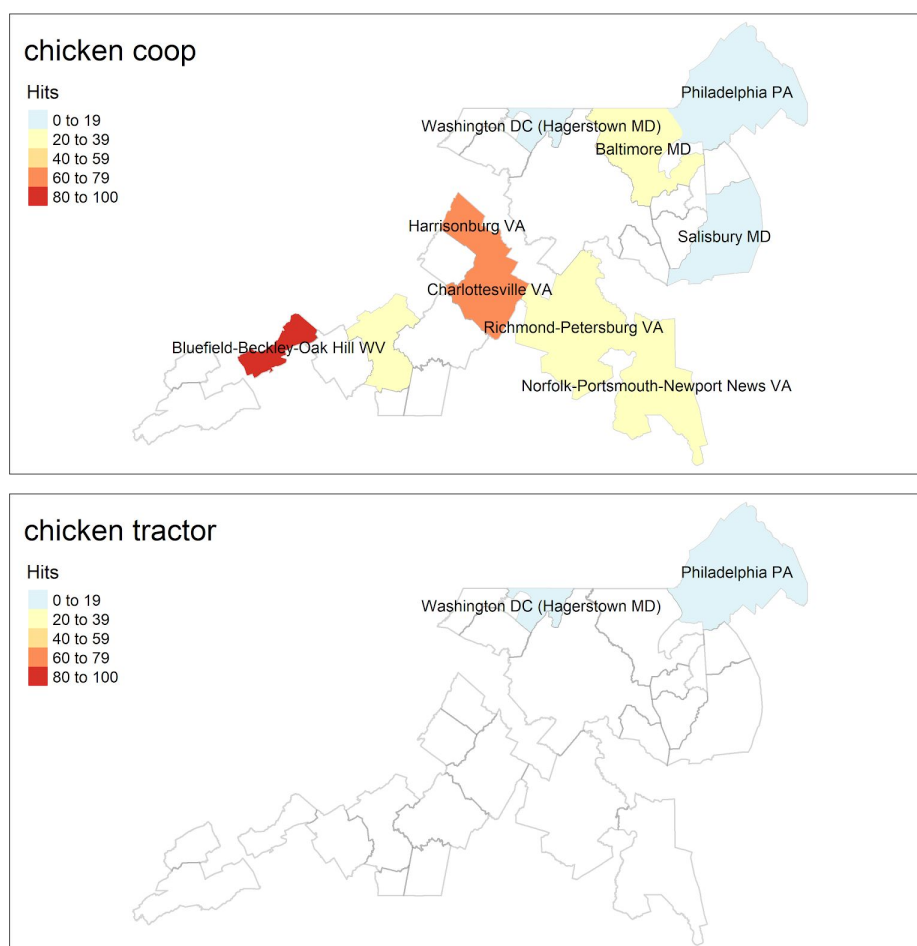


Figure 15: Regional Interest For Keywords “chicken coop” and “chicken tractor”

Note: The maps above illustrate interest in the chosen keyword relative to the national interest from 2010 through 2019. The number of searches, or “Hits”, in each metro area is scaled from 0 to 100. For example, the United States metro area with the highest number of searches would have 100 “Hits”. It is worth noting that metro areas without any color did not show enough interest in the keyword to collect any data. Also, this search interest data was collected with the category “Pets & Animals” specified.

Seasonal Interest within Each Metro Area

We were able to find seasonal trend data for the keyword “chicken coop” in Washington DC, Baltimore MD, Bluefield-Beckley-Oak Hill WV, Harrisonburg VA, Philadelphia PA, Richmond-Petersburg VA, Roanoke-Lynchburg VA, and Salisbury MD. For most of the metro areas in the three states, a trend can be found in which interest for the term “chicken coop” peaked in the end of April. After that, interest tended to decline somewhere between the beginning of August to the end of October and then increase slightly near the end of the year.

However, in Roanoke-Lynchburg VA, Norfolk-Portsmouth-Newport News VA and Harrisonburg VA, interest tended to peak around the end of June. In some of these metro areas, interest then began to rise slightly around the middle of October. Differing from trends in the other metro areas, Bluefield-Beckley Oak Hill tended to start the year at the highest point of interest, peak again around the end of June and then slightly decline throughout the rest of the year.

There is not much seasonal variation for interest in the search term “chicken tractor” as seasonal data can only be found for Washington DC and Baltimore MD. In Baltimore MD, interest for the term “chicken tractor” remained steady throughout the beginning of the year and then peaked around the end of October. In contrast, the highest point of interest for this term in Washington DC tended to be at the beginning of the year and then remained steady throughout the rest of the year, as shown in Figure 16.

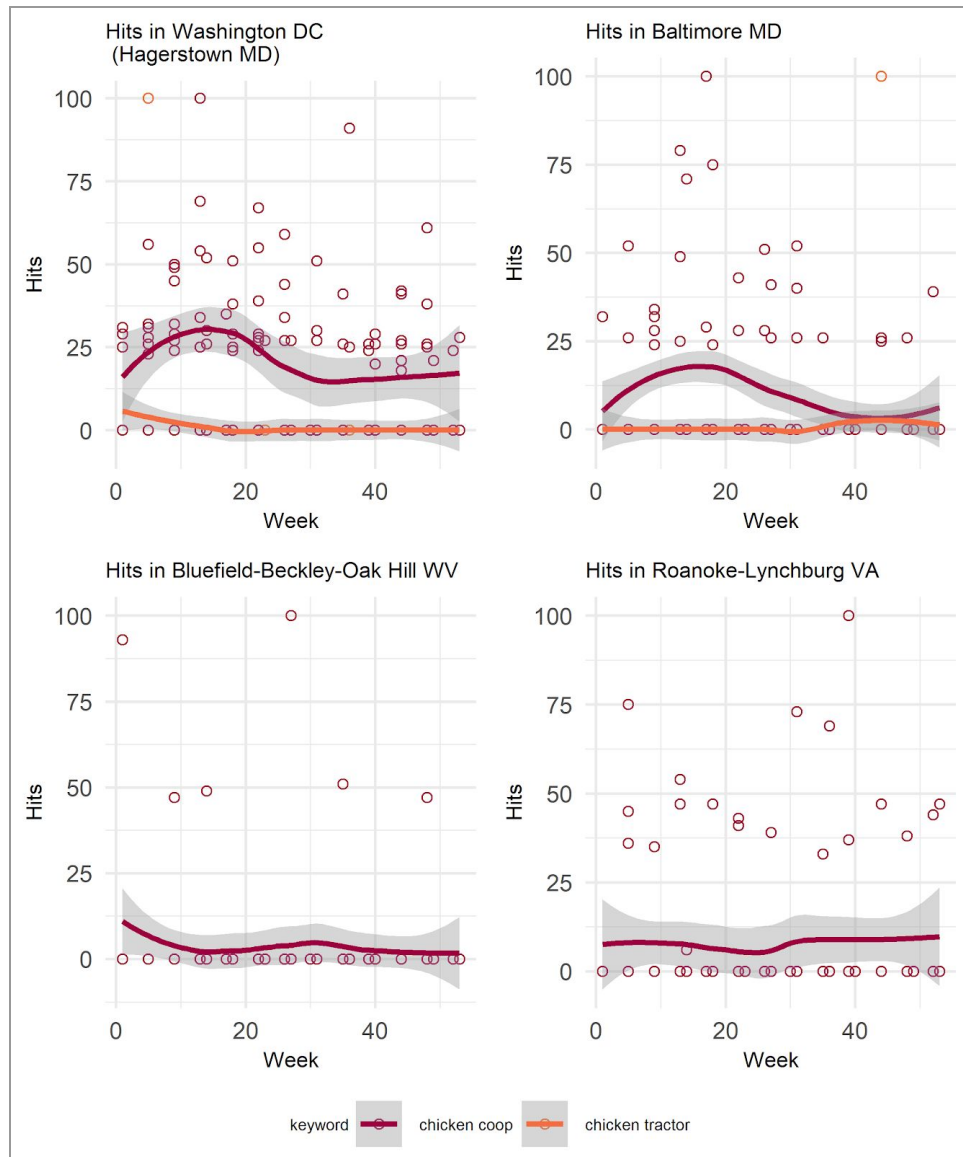


Figure 16: Seasonal Trends for Keywords “chicken coop” and “chicken tractor”

Note: The plots above show seasonal trends for searches across each year. Each dot represents the weekly search interest for each keyword in each area, relative to the maximum weekly search interest in that area. Each line represents the locally weighted smoothing of hits in that week across 10 years. This search interest data was collected with the category “Pets & Animals” specified. Because hits are relative to the maximum hits of the same keyword and area, hits between keywords should not be compared. For example, when the trend for one keyword is greater than the trend for another does not mean that there are more searchers for one relative to the other. They are plotted in the same graph in order to compare the differences in seasonal search interest.

Keywords Related to “Showing or Selling Poultry”

Keywords searched: Poultry show, poultry swap meet, show quality birds, flea market

Regional Interest Relative to the National Maximum

Although Salisbury MD showed the most interest, it is clear from Figure 17 that the metro areas in the three states showed little to no interest in the search term “flea market” compared to the rest of the nation.

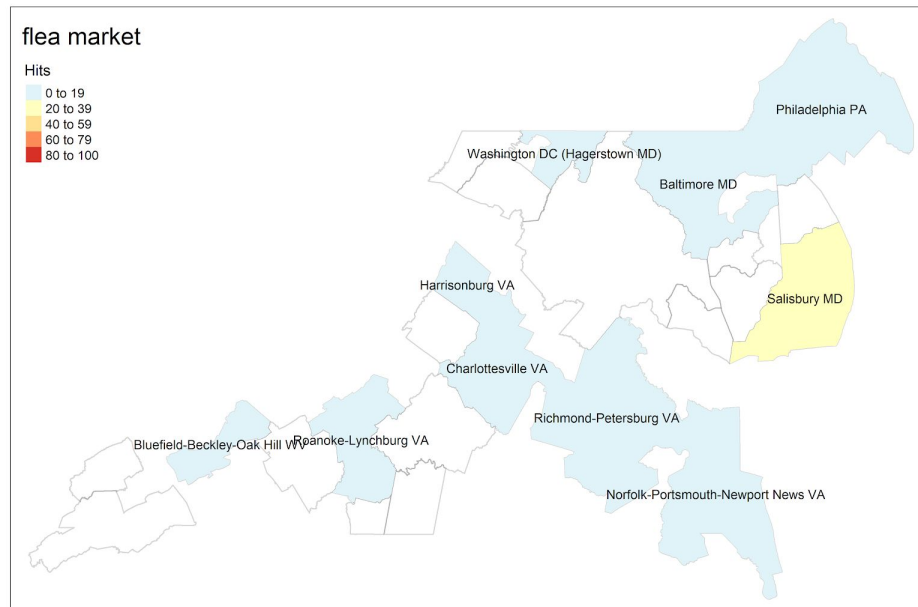


Figure 17: Regional Trends for Keyword “flea market”

Note: The maps above illustrate interest in the chosen keyword relative to the national interest from 2010 through 2019. The number of searches, or “Hits”, in each metro area is scaled from 0 to 100. For example, the United States metro area with the highest number of searches would have 100 “Hits”. It is worth noting that metro areas without any color did not show enough interest in the keyword to collect any data. Also, this search interest data was collected with the category “Pets & Animals” specified.

Seasonal Interest within Each Metro Area

We were able to find seasonal trend data for the term “flea market” in Washington DC, Baltimore MD, Bluefield-Beckley-Oak Hill WV, Charlottesville VA, Harrisonburg VA, Norfolk-Portsmouth-Newport News VA, Philadelphia PA, Richmond-Petersburg VA, and Roanoke-Lynchburg VA. In many of these metro areas, interest tended to peak around the middle of April and then remain steady throughout the rest of the year. Some of these trends were very slight, while others were more obvious. In some metro areas, interest tended to increase or even peak again near the end of the year.

In Baltimore MD, Bluefield-Beckley Oak Hill WV, Roanoke-Lynchburg VA, Norfolk-Portsmouth-Newport News VA, and Philadelphia PA, interest tended to peak somewhere between the end of June and the beginning of September. In Washington DC,

interest tended to start out at a high point, peak again around the end of June, and then decline throughout the remainder of the year. This is shown in Figure 18.

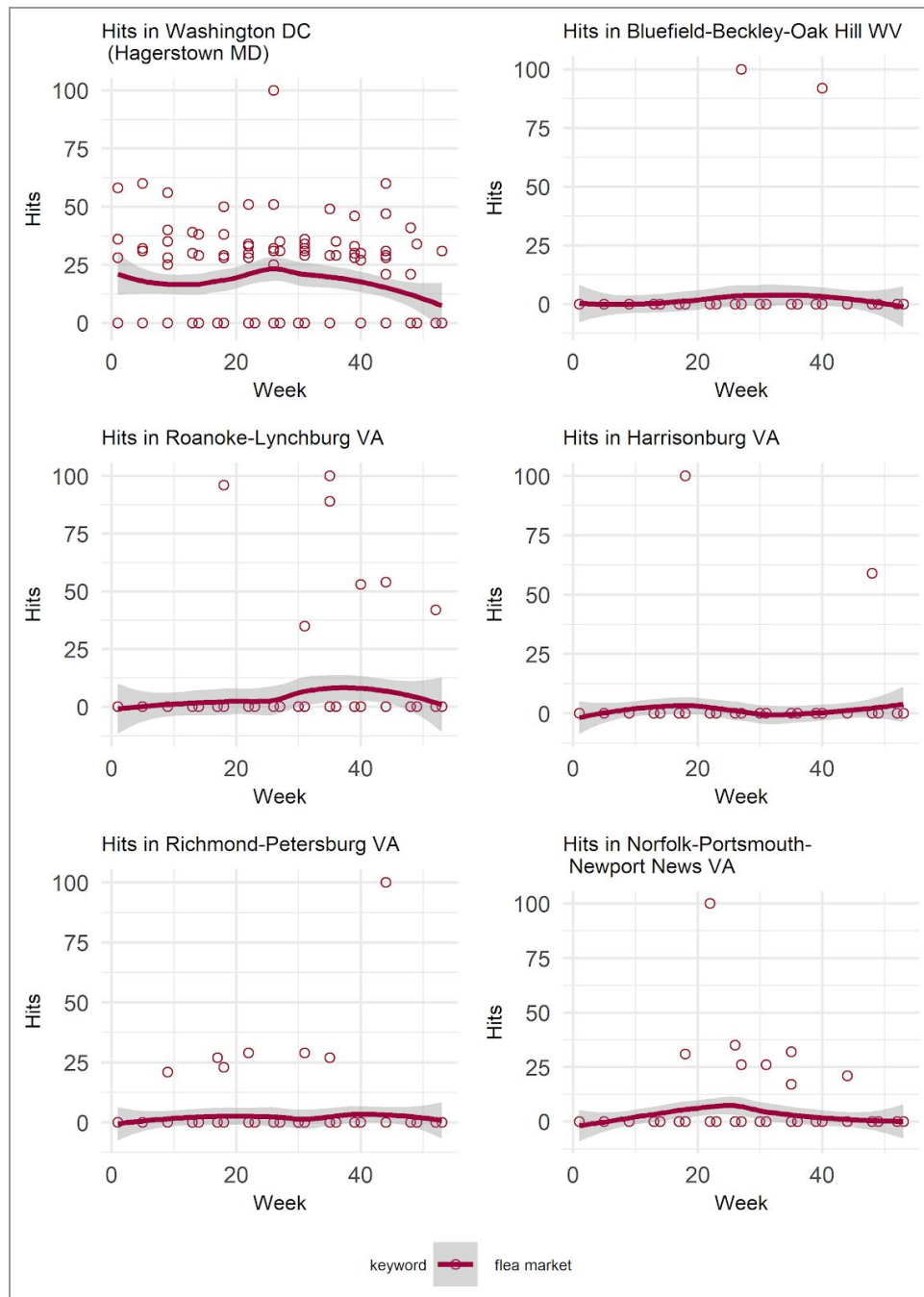


Figure 18: Seasonal Trends for Keyword "flea market"

Note: The plots above show seasonal trends for searches across each year. Each dot represents the weekly search interest for each keyword in each area, relative to the maximum weekly search interest in that area. Each line represents the locally weighted smoothing of hits in that week across 10 years. This search interest data was collected with the category "Pets & Animals" specified.

Keywords Related to “Types of Poultry”

Keywords searched: Pullet, started pullets, baby chicks, baby ducklings

Regional Interest Relative to the National Maximum

It is clear from Figure 19 that, compared to the rest of the nation, there was much higher interest in the search term “baby chicks” compared to the term “pullet” in the three states. Washington DC and Philadelphia PA both showed interest in each of these search terms.

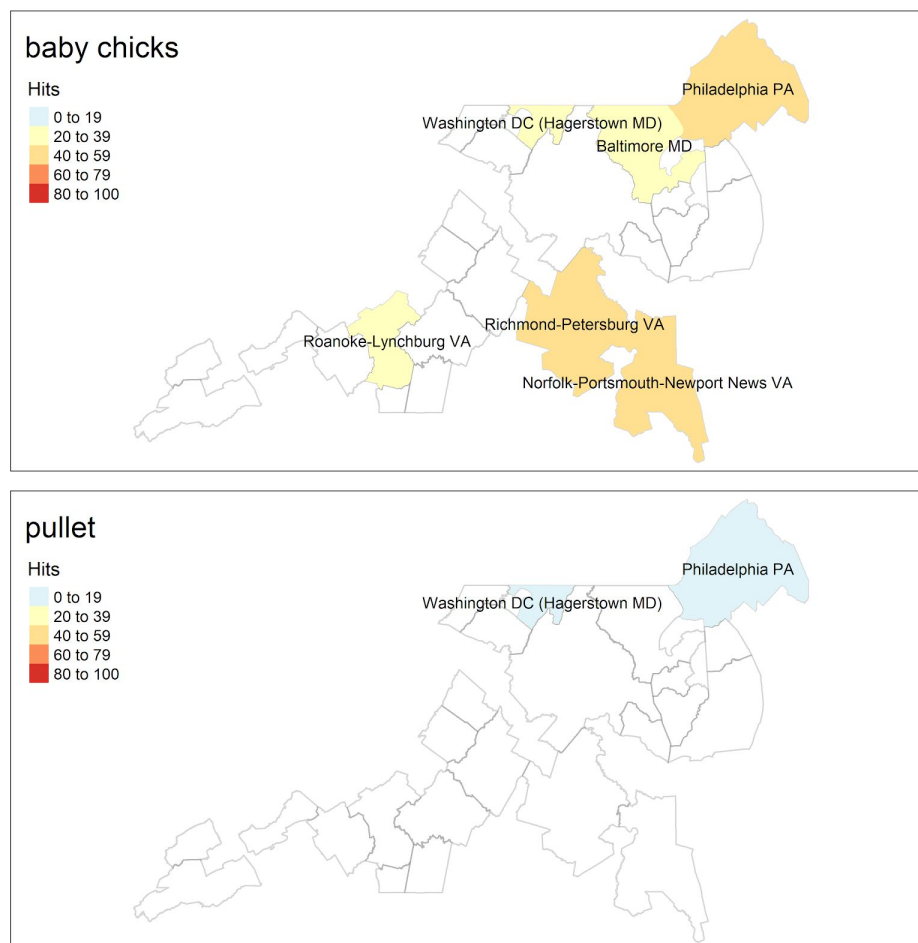


Figure 19: Regional Trends for Keywords “baby chicks” and “pullet”

Note: The maps above illustrate interest in the chosen keyword relative to the national interest from 2010 through 2019. The number of searches, or “Hits”, in each metro area is scaled from 0 to 100. For example, the United States metro area with the highest number of searches would have 100 “Hits”. It is worth noting that metro areas without any color did not show enough interest in the keyword to collect any data. Also, this search interest data was collected with the category “Pets & Animals” specified.

Seasonal Interest within Each Metro Area

We were able to obtain seasonal trend data for the keyword “baby chicks” in Washington DC and Philadelphia PA. Interest tended to peak around the end of March, reach a low point sometime in June, and then remain steady throughout the remainder of the year.

We were able to obtain data for the keyword “pullet” in Washington DC, Baltimore MD, and Philadelphia PA. There were not many notable seasonal changes throughout the year, as shown in Figure 20.

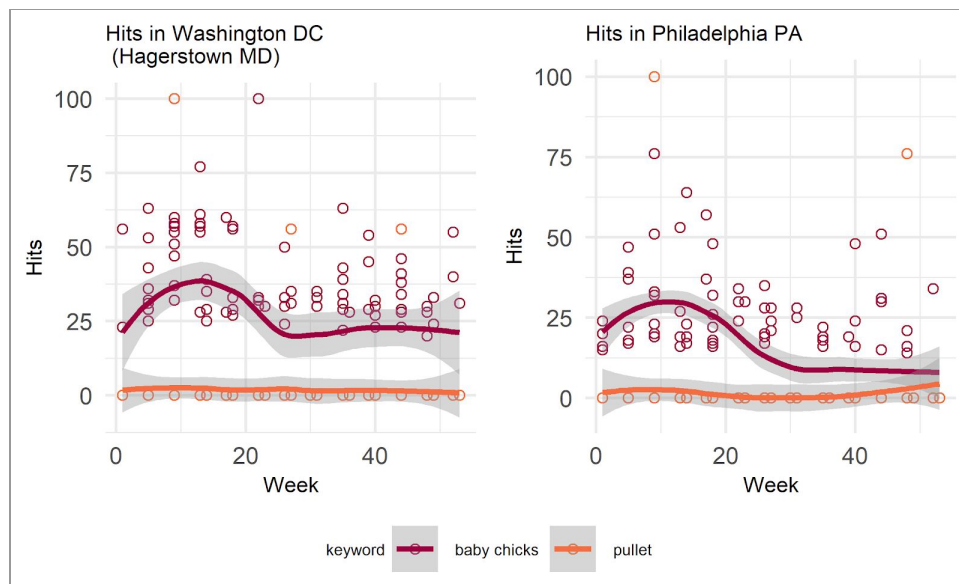


Figure 20: Season Trends for Keywords “baby chicks” and “pullet”

Note: The plots above show seasonal trends for searches across each year. Each dot represents the weekly search interest for each keyword in each area, relative to the maximum weekly search interest in that area. Each line represents the locally weighted smoothing of hits in that week across 10 years. This search interest data was collected with the category “Pets & Animals” specified. Because hits are relative to the maximum hits of the same keyword and area, hits between keywords should not be compared. For example, when the trend for one keyword is greater than the trend for another does not mean that there are more searchers for one relative to the other. They are plotted in the same graph in order to compare the differences in seasonal search interest.

Conclusion

From scraping online registries and data Google Places, we were able to collect location and contact information for more than 8,000 locations that can potentially be selling live poultry and eggs. As most of these locations do not register with APHIS or the state they are in, we encourage APHIS to keep monitoring these information sources so that information can be efficiently communicated when there is an outbreak.

Using Google Trends data, we were able to find that the metro area of Richmond-Petersburg has one of the highest interest for fresh eggs in the nation, and the metro area of

Bluefield-Beckley Oak Hill WV has one of the highest interest in chicken coop in the nation. There is moderate interest for all other keywords, with peak interest usually at the beginning of the year. Our analysis covered a very small scope of keywords that can be searched and we encourage APHIS to use Google Trends to monitor trends for symptoms of diseases prone to an outbreak in the area.

Lastly, we would like to thank APHIS for this opportunity. This is the first professional experience for us (the Summer Fellows) and doing research for a real group of users made the experience much more valuable. We learned to collaborate in a remote environment and apply our critical thinking skills and creativity to answer relevant questions. We also learned advanced functions and strategies in R to clean data and were able to experience trial and error. Lastly, we got to explore interesting data science applications to the agricultural sciences.

Authors



Thanicha Ruangmas (Principal Investigator)

Dr. Ruangmas is an Assistant Clinical Professor at the University of Maryland, College Park, where she leads the Sustainability Analytics FIRE research group since 2019. She received her Ph.D. in Agricultural and Applied Economics from the University of Wisconsin-Madison.

In this project, she coordinated, trained, and supervised the Summer Fellows in using data analysis tools to detect farms selling live poultry and eggs.



Lars Olson (Faculty Advisor)

Dr. Olson is Professor and Director of Undergraduate Studies in the Department of Agricultural Economics, University of Maryland, College Park. He is the founding faculty leader of Sustainability Analytics, the founding director of the cross-campus Global Poverty Minor, served as department chair from 2008-2013 and 2017-2018, and is co-lead faculty fellow on a new University Honors cluster on Information and Power to begin in Fall 2021. Professor Olson's research focuses on environmental and natural resource economics, with an emphasis on intertemporal problems under uncertainty. His research has been published in *J. Economic Theory*, *J. Public Economics*, *J. Environmental Economics and Management*, and *Environmental and Resource Economics*, among others. He has served as Associate Editor and on the Editorial Council of the *J. Environmental Economics and Management*.

Dr. Olson supported this project with his technical and non-technical guidance.



Anna Lazarus-Hall (Summer Fellow)

Anna Lazarus-Hall is a sophomore Materials Science and Engineering major. She contributed by experimenting in order to find the most practical use of Google Trends for this project. She was able to collect and analyze data from Google Trends regarding interest by metropolitan area as well as seasonal interest for relevant keywords.



Janna Chapman (Summer Fellow)

Janna Chapman is a sophomore Environmental Science and Technology major. She was responsible for the “Obtaining Data from State Registries” section of the report, as well as cleaning and geocoding data from MDA and VA Grown.



Mishti Relan (Summer Fellow)

Mishti Relan is a sophomore Computer Science major. Her contributions to this project include scraping farm details from web directories (Local Hens, Local Harvest, and MARYland’s Best) and creating an interactive web application that displays the location of the farms and their details on the map of the US based on the dataset they are included in. She also obtained the formatted Google addresses and geo coordinates of the farms.



Shanil Kothari (Summer Fellow)

Shanil Kothari is a sophomore Computer Science major. He contributed by using the Google Places API to get potential farms and their details. He also helped with cleaning the data of some state directories. He was also able to scrape a few Internet forums on Backyard Chickens to obtain some user details. He helped make the shiny app for the Google Places results and the “Farms in Range” for state registries as well as Google Places.