Title: **The Historical Phytogeography of *Cirsium arvense*, An Invasive Species in Pennsylvania**

Author: **Janet K. Mansaray,**
The Eli Kirk Price Endowed Flora of Pennsylvania Intern

Date Submitted: **March 6, 2018**

Abstract:

According to the Department of Conservation and Natural Resources, *Cirsium arvense* (Asteraceae family) is currently an invasive plant in the state of Pennsylvania. Invasive species pose a problem as they are detrimental to natural ecosystems and very costly to manage and eradicate. In this study, distribution of *C. arvense* in Pennsylvania was reconstructed using only herbarium records. Through detailed methodology, it was determined that there were no shifts in habit preference over time. With the data being specific to Pennsylvania, the objective was to determine if the distribution and habitat preference would align with the current literature on what is known about *C. arvense*. The data seemed to support the current literature in that *C. arvense* appeared to be widespread and prefers dry, disturbed areas like roadsides. However, with further analysis, the data was found to reflect trends in field collecting as opposed to the distribution of the species. One of the limitations of the study was collector bias in addition to procedural obstacles. From this study, valuable insight was gained about the future of botanical collecting techniques and the importance of phytogeographical studies.

Table of Contents

INTRODUCTION

Cirsium arvense (Canada Thistle) is known to directly compete with native vegetation, reduce species diversity, and alter habitat structure ("University of Georgia Center for Invasive Species and Ecosystem Health", 2017). A member of the Asteraceae family, its most considerable ecological impact is its allelopathic effect. Because of its ability to produce chemicals that negatively influence seed germination, its presence can decrease crop yields. There have been reports of the species vastly decreasing the yield of economically important crops such as wheat (Stachon, 1980).

Cirsium arvense was first accidently introduced to the United States in the 1600s from Europe. By 1795, it was considered a noxious weed in Vermont. In 1918, it became a noxious weed in 25 northern states (University of Georgia Center[…]", 2018). By 1954, it had been added to the noxious weeds lists of 43 states. Today, the species is distributed throughout Canada and the northern United States, from northern California to Maine and south to Virginia (Pennsylvania Department of Conservation and Natural Resources, 2018).

The literature stated that it is commonly found in dry, disturbed habitats such as old agriculture or abandoned fields, roadsides, and landfills. It is difficult to eradicate because the root system is extensive and must be removed completely. If not, the rhizomes allow the parent plant to propagate vegetatively (asexually), making mechanical methods of removal futile (Donald, 1994). According to the Department of Conservation and Natural Resources, *C. arvense*, is currently an invasive plant in Pennsylvania (Pennsylvania Department of Conservation and Natural Resources, 2018). The goal of this project was to examine if the historical distribution of *C. arvense* in Pennsylvania, aligned with the current literature using only herbarium records. It was predicted that habitat preference would be consistent throughout time and support the current literature. This meant that *C. arvense* would be more commonly found in disturbed areas, such as old fields or roadsides. In addition, it was predicted that the distribution would be widespread over time throughout the state due to it being commonly dispersed in Pennsylvania (Rhoads, 2007). It was important to study the spread of *C. arvense* throughout time in order to assess the feasibility of eradication.

METHODS

Using the Mid-Atlantic Herbaria Consortium database, herbarium records of *C. arvense* in Pennsylvania were found. Records lacking the collection date (year), locality, and habitat were excluded. Records indicating that the specimen was grown in greenhouses were excluded as well. From the database, an Excel sheet was exported that contained all records of *C. arvense* in Pennsylvania. The Excel sheet was modified to keep the information relevant for this project: the institution the record came from, catalog number, collector name, collector number, date (year, day, month), country, state, county, specific locality, and habitat. Each record was assigned coordinates using GeoLocate, a historical locality search engine. This software is designed with an algorithm that translates locality descriptions associated with biodiversity collections into geographic coordinates (Tulane University, 2014). Once all coordinates for the records were found, the next step was to see if there were any changes in habitat preferences. Thus, in order to distinctly compare the records to one another, they were divided into three time periods. Each time period contained an approximate 50-year interval to serve as a control when comparing the distribution throughout time. This was important because if the time interval was too small, such as ten years, changes in distribution or habitat preference might not be seen. If the intervals were disproportionate (one too small, another too large), then the data could have been skewed toward one time period than the other. CSV files of each time period were uploaded to Google Maps. Through Google maps, the coordinates were plotted thus generating the distribution maps for each time period. To differentiate between the time periods, the locality markers were given different colors (Fig. 2).

Once the maps were generated, habitat graphs were produced that showed the habitats occupied by *C. arvense* during each time period (Fig.3). The habitats were divided into four categories: transport habitats, man-made habitats, natural habitats and no data (Tab. 1). The "no data" bar in each habitat graph represents the specimen labels that did not contain habitat information. Transport habitats represented those that were used for transportation such as roads, railways, paths, roadside meadows and water courses. Water courses included water sources that move (as opposed to standing water like a swamp) or were used for travel such as streams, streamlets and canals (Pyšek, P., & Prach, K, 1993). Man-made habitats represented records that indicated man-made structures such as ore pits, open lots of abandoned buildings, and used fields. Used fields was a general term used to indicate records that described the fields as "old" or "for agriculture". Natural habitats included meadows, swamps, thickets, etc. Table 1 was used to create the graphs in Figure 3 and the habitat frequency graph in Figure 4. Finally, a habitat frequency graph was produced for all the habitats recorded over the time span. This showed with what frequency or intensity the habitats were colonized by the species (Fig. 4). The number of each habitat type occupied by *C. arvense* was summed and divided by the total number of records and then multiplied by 100 to get a percentage (Pysek, 1991).

RESULTS

Using the herbarium database, records were obtained from fifteen different herbarium: Carnegie Museum of Natural History Herbarium, The Academy of Natural Sciences, Field Museum of Natural History, Hillsdale College Herbarium, J. F. Bell Museum of Natural History Museum, Marshall University, Missouri Botanical Garden, Morris Arboretum of University of Pennsylvania, Muhlenberg College, Rutgers University-Chrysler Herbarium, New York Botanical Garden, University of Illinois Herbarium, University of Michigan Herbarium University of South Carolina Upstate Herbarium, and University of Wisconsin-Madison (Wisconsin State Herbarium). In total, there were 127 different collectors; 120 collectors were listed on the labels and 7 were not. Robert L. Schaeffer, Jr. contributed 98 records but all were from varying localities. The total records found on the Mid-Atlantic Herbaria Consortium database of *C. arvense* were 533. The earliest record was from 1864, representing the first recorded time the species was collected, and the most current record was from 2012. In total, the records span 149 years. This time frame was divided into three time periods. From 1864 to 1914, there were 71 records. Of the 71 records, 40 records had no habitat data. However, they did contain locality information. Of the records that had habitat information during this time period, most of the *C. arvense* records were found in transport habitats (Fig. 3A, Tab.1). From 1915 to 1965, there were 374 specimen records. The majority of records came from this time period. The habitats preferred during this time period were man-made with transport habitats being second (Fig. 3B, Tab. 1). There were clusters of records found in the southeastern portion of Pennsylvania. However, with an increase in records, it can be seen that that distribution starts to become more widespread toward the western portion of the state (Fig. 3B). The years1966-2012 consisted of a 46-year time span due to the lack of records after 2012. In total, there were 108 records from this time period. The habitat preference during this time period favored man-made and transport habitats (Fig. 3C, Tab. 1). The distribution of these records were similar to the previous time period. It was relatively widespread with a slight cluster of records overlapping in the southeastern region of Pennsylvania (Fig.3C). The habitat frequency graph showed the percentages of all habitats occupied by *C. arvense* throughout the time periods (Fig. 4). Man-made habitats comprised 49.53% of all the habitats, with used fields having the highest percentage of habitats recorded for *C. arvense* at 32.45%. The second highest habitats recorded for *C. arvense* were transport habitats at 26.26%, with roadsides/roads being the most frequently recorded at 16.88%. Natural habitats comprised 17.64% of overall habitats recorded, with meadows being the most frequently recorded in this category (Fig.4). When comparing the data from the three time periods, the distribution and habitat preferences are similar and seem to align with the current literature about *C. arvense*. *C. arvense* appeared to favor disturbed areas such as used fields, waste grounds, and roadsides (Fig 3, 4). The only difference between the time periods was the number of records. From 1915-1965 there were drastically more records than those in the other time periods. By looking at the maps and habitat graphs alone, the distribution and habitat preferences have not changed much over time (Fig 3, Fig 4).

DISCUSSION

In comparison to other geographical studies over larger regions that used a little over 700 herbarium records, 533 records was a sufficient sample size to reconstruct the spread of *C. arvense* (Lavoie, 2007). The spread of *C. arvense* seems to not have changed over the course of 149 years. The distribution seemed to be consistent throughout the state. Thus, it appeared that the data supported the hypotheses as well as the current literature. There were, however, overlapping localities that occurred in the southeastern region of Pennsylvania (Fig. 2). Despite the data seeming to support the hypotheses, the disproportionate number of records from each time period indicated a common bias found in studies using herbarium specimen. Collector bias is rooted in the practice of following the work of botanists who established extensive herbariums (Reuell, 2017). Collectors will often use these botanists and their past collection trips as a point of reference for where to search for certain plants. This often led to them collecting in similar areas (Reuell, 2017). It is possible the data of this study aligns with the current literature because collectors were following the literature of their time to find this species and not venturing into new or less accessible areas.

One obstacle encountered during this study was that the number of herbarium records increased from the first to the second time period but decreased from the second to the third time period. The literature stated that there is a tendency for collectors to collect non-native species (Reuell, 2017). Thus, one would expect that as a non-native species with a potential to become invasive increased its spread, the amount of collecting of this species would also increase. However, the fluctuation in record numbers does not reflect species population, but trends in botanical collecting. Therefore, the data seems to be more of an indication of common collecting spots rather than the distribution of *C. arvense*. Amongst collectors it is also common to collect plants close to roadsides rather than deeper in the landscape (Delisle, 2003). This could mean that the abundance of records found near roadsides could be skewed. There is also a strong preference for collecting plants in the summer versus the winter or fall (Reuell, 2017). Therefore, the habitat preferences seen in this study do not take into account collector bias or important habitat factors such as seasonality.

Another limitation to this study was the use of Google Maps to generate the distribution maps as opposed to using software like GIS. GIS is a computer program designed to capture, store, and display data related to positions on Earth's surface. It helps to better understand spatial patterns and relationships (National Geographic Society, 2012). Unlike Google Maps, GIS includes the option of adding climatic information to the maps. With GIS, one can generate distribution maps that take into account factors such as seasonality, elevation, temperature, and soil types. A detailed profile and distribution map could have helped to overcome some biases because there would have been more habitat data to compare other than the locality and general descriptions. Finally, another limitation to using herbarium specimens is human interpretation of labels. For example, if a collector writes on the label that the habitat is "field", it is hard to say what kind of field it is or what it was used for. Though herbarium labels sometimes provide detailed information, the details vary by collector. To combat this and produce the most accurate information, various papers were used as a standard for what is considered a field or meadow, etc.

Though the results do not account for bias, the results of this study are specific to Pennsylvania and can used to as a guide in how to re-design a similar project that takes into account biases. The methodological obstacles of using herbarium specimen data could be applied to studies done on a wide range of geographic and temporal scales.

CONCLUSIONS

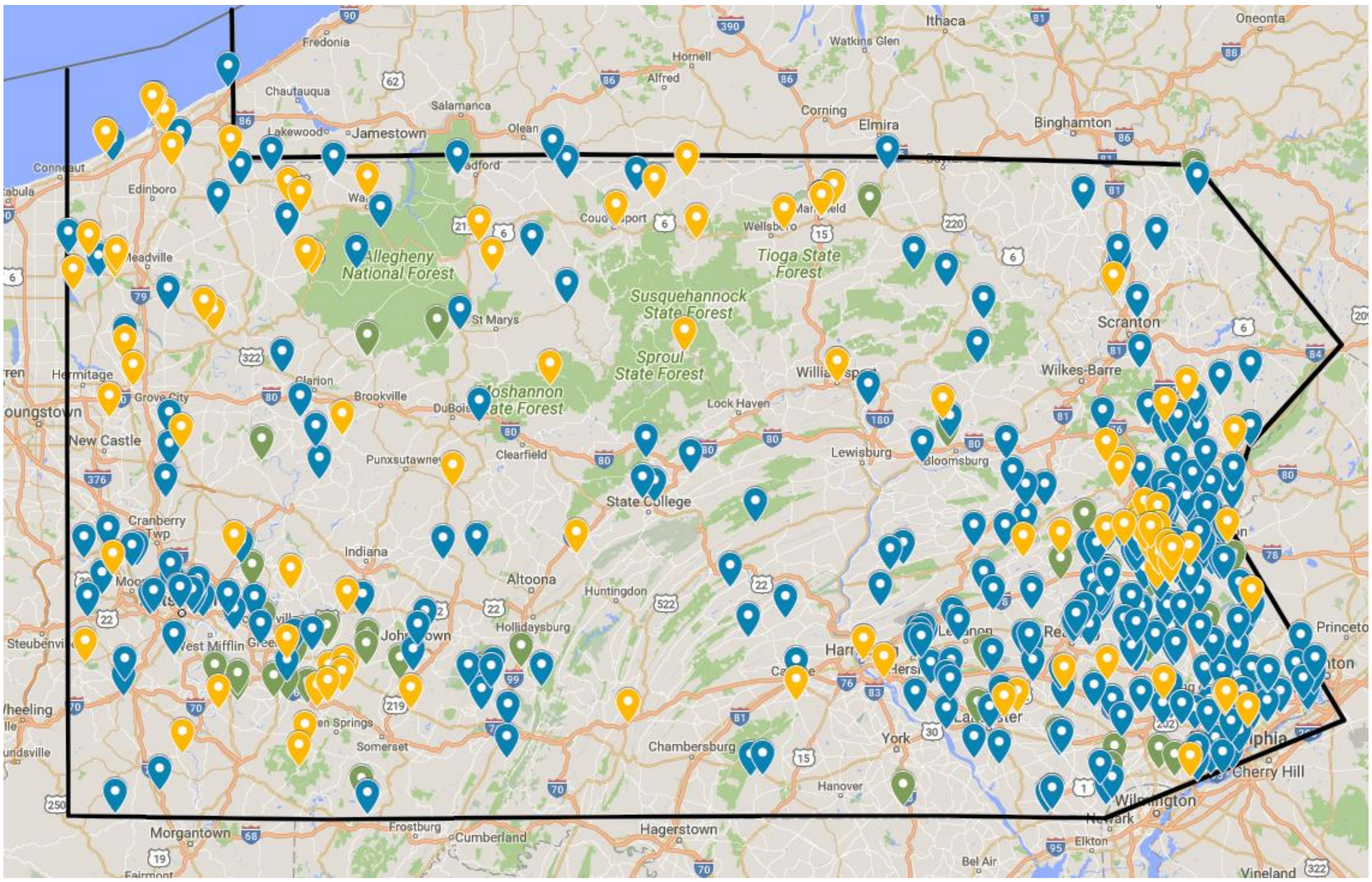
Phytogeographical studies allow researchers to classify organisms into fundamental geographic areas, establish conservation efforts, analyze species distribution range from the past and future, and understand the transformative effects of climate change (Lavoie, 2007). The main goals of this study were to provide an overview of the distribution of *C. arvense* in Pennsylvania and determine if there were any shifts in habitat preferences. Due to the lack of accountability of collector biases, the data is inconclusive. The question now is, how can this data be used?

Future botanists can use this information to strategize the best techniques for botanical collection. By understanding the obstacles faced in this study, a researcher could implement beneficial collecting practices. Beneficial for what the researcher may be working on in that immediate moment as well as beneficial for future researchers who also want to conduct phytogeography studies. For example, rather than prioritizing collecting at sites that are well known, botanists can challenge themselves to explore novel, diversity areas. Another example, would be to include more detailed labels that focus not only on the plant but the surrounding habitat. *C. arvense* has been prevalent in Pennsylvania for more than 150 years. If eradication is something the public wants to be possible, research projects like this are important. In learning the history of an invasive plant, especially in a specific area, we may discover the key to decreasing its spread in the present.

Figures and Tables

Figure 1. *C. arvense* in bloom. Photo by Jan Samanek (www.forestryimages.org)

Figure 2. Distribution of *C. arvense* over time.



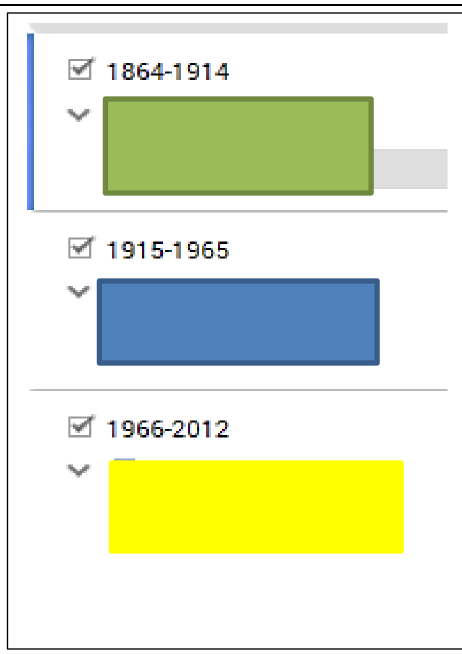
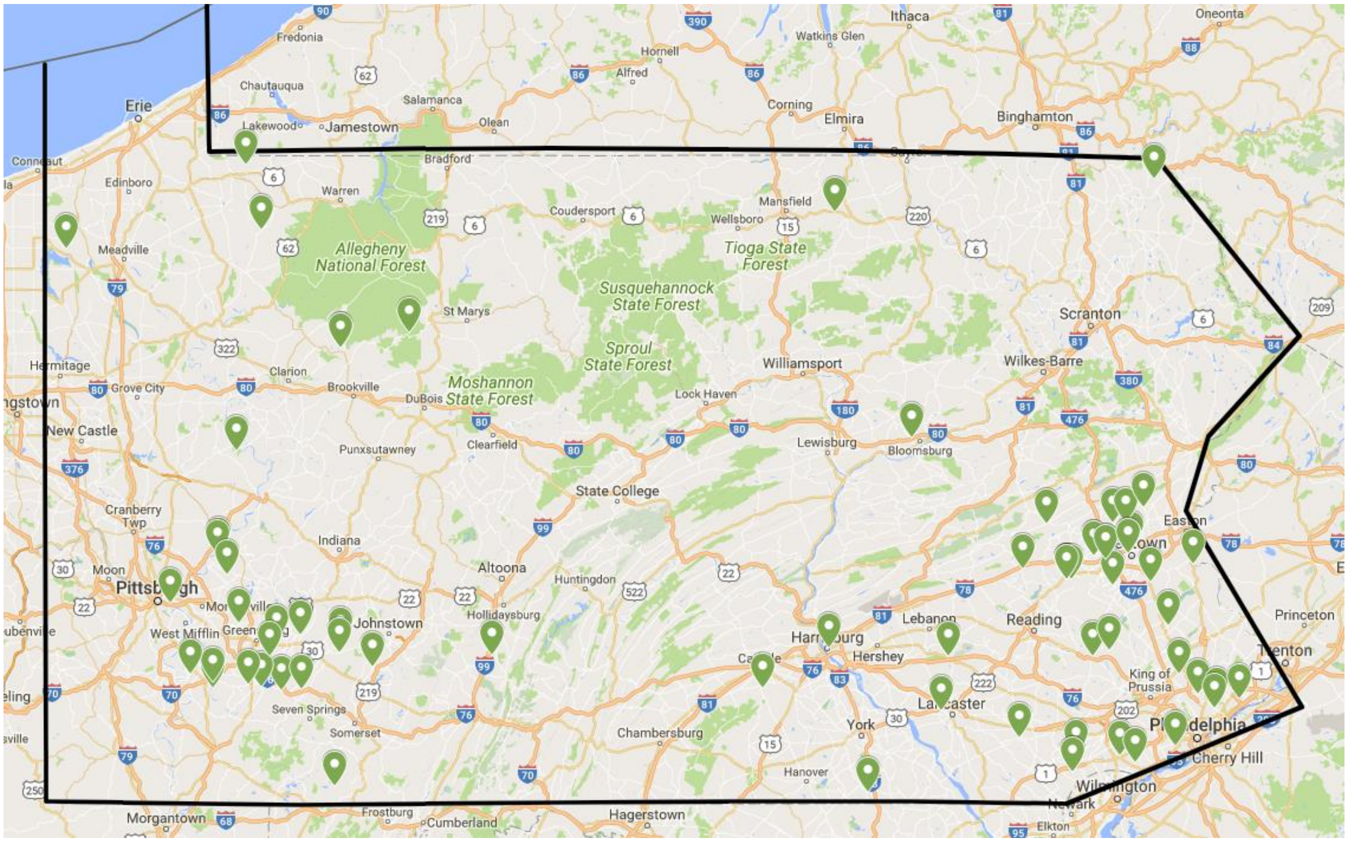


Figure 3A) Distribution Map of *C. arvense* from 1864-1914 and Habitat Graph



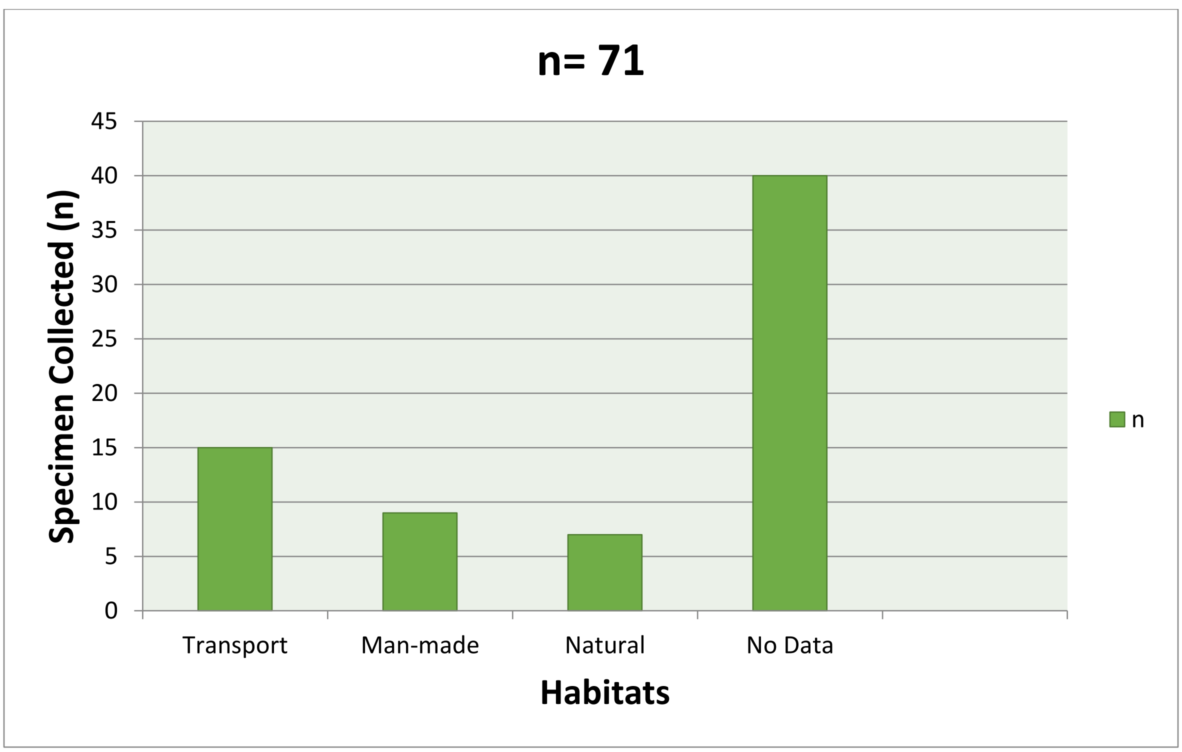
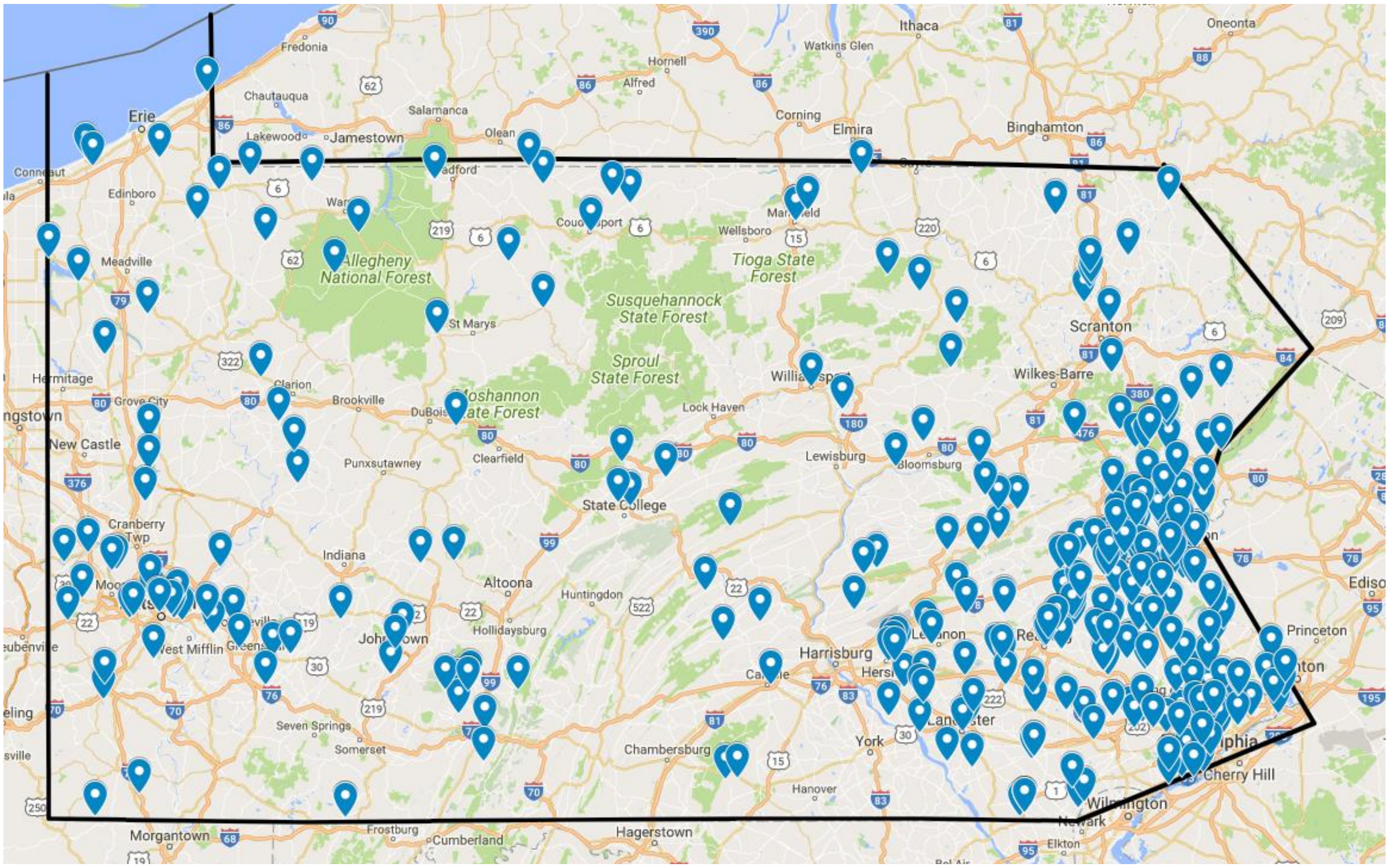


Figure 3B) Distribution Map of *C. arvense* from 1915-1965 and Habitat Graph



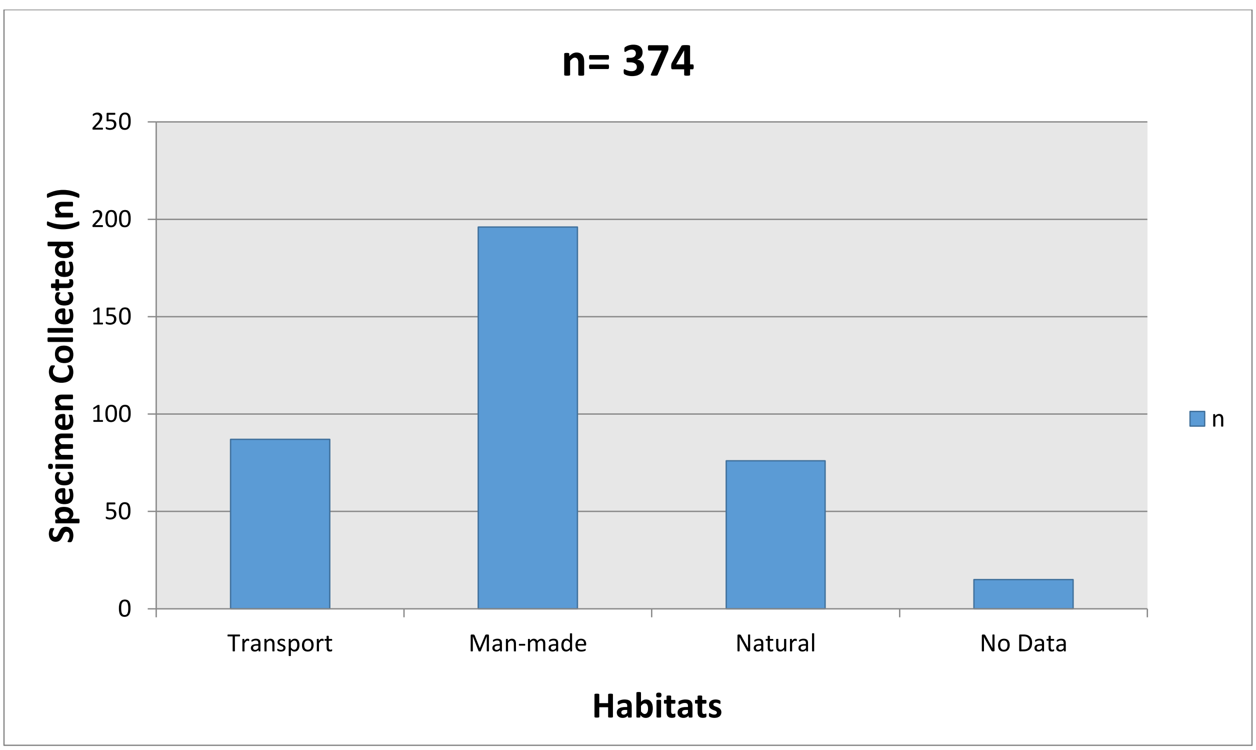
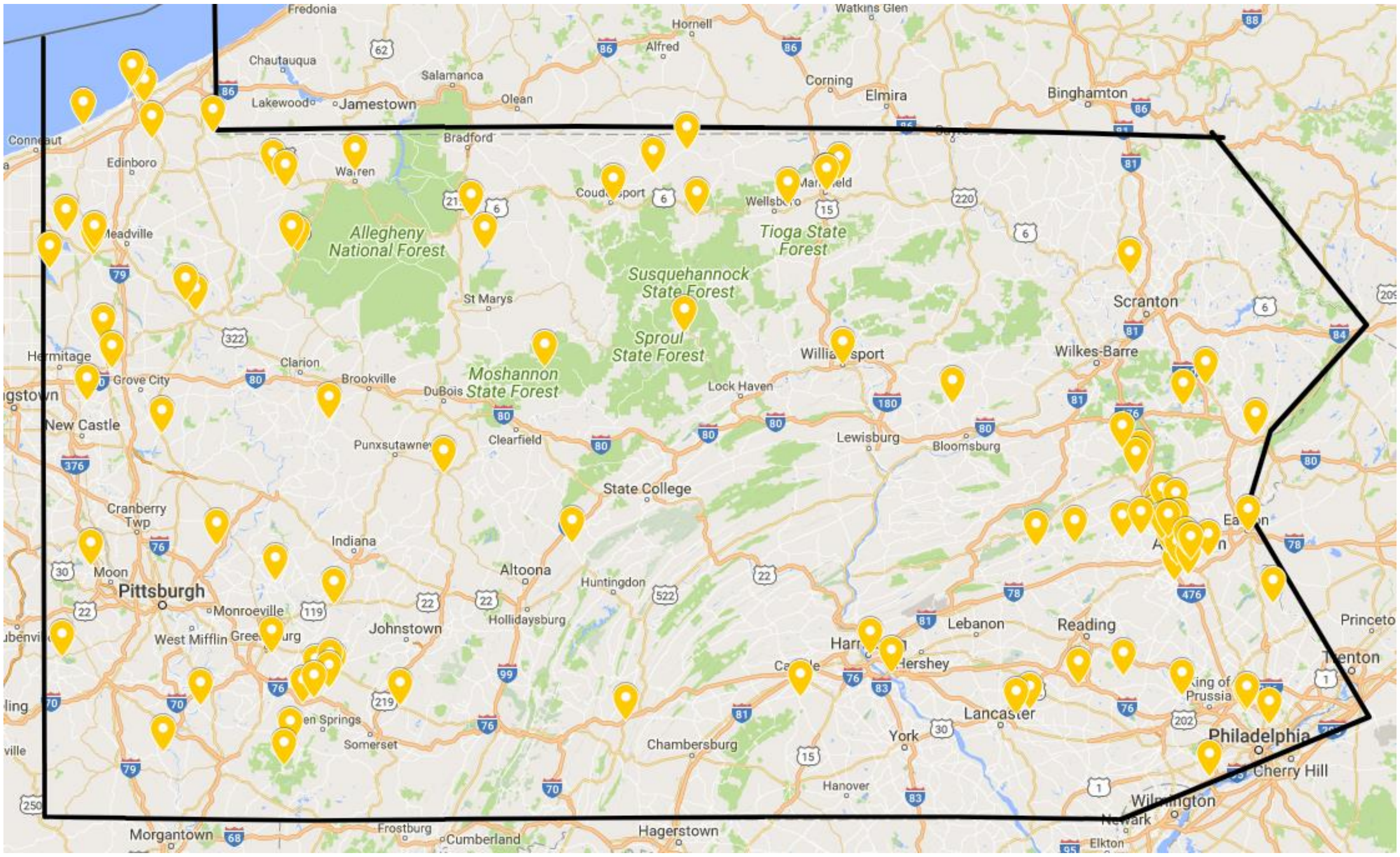


Figure 3C) Distribution Map of *C. arvense* from 1966-2012 and Habitat Graph



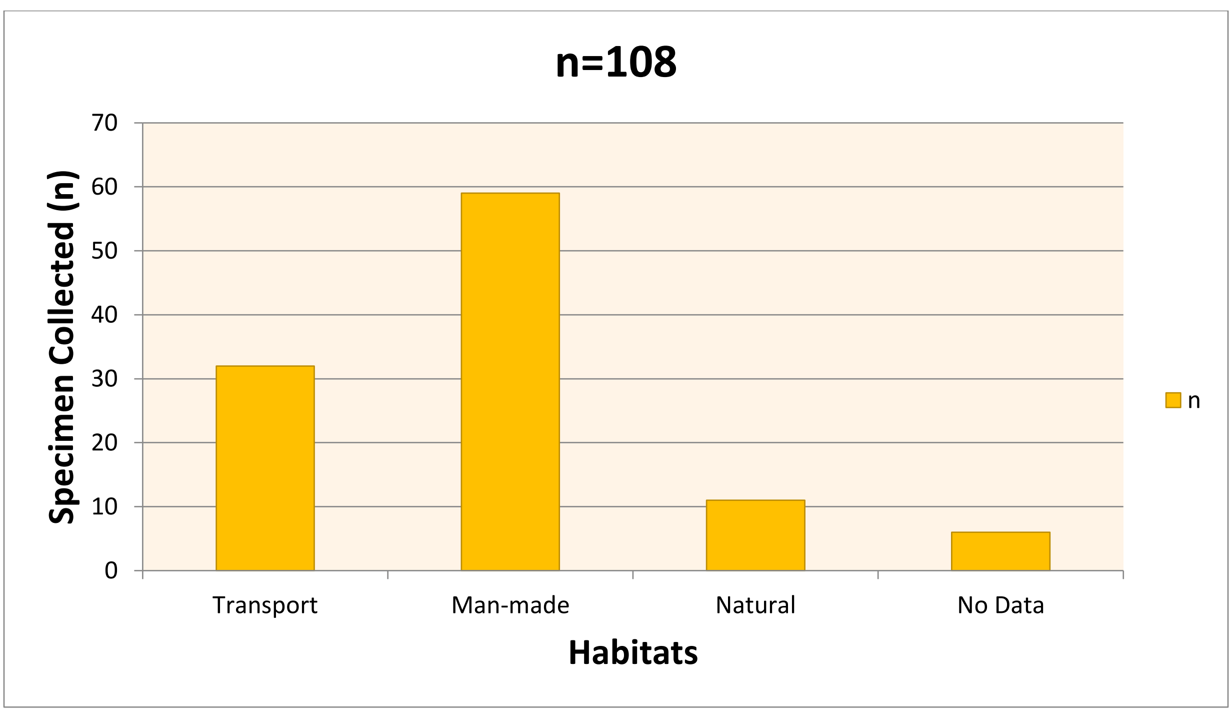


Figure 3. The distribution maps and habitat graphs of *C. arvense* from A) 1864-1914 B)1915-1965 C) 1966-2012.

Table 1. A table representing the various habitats colonized by *C. arvense* during each time period.

General Habitats		**1864-1914**	**1915-1965**	**1966-2012**
	Specific Habitat	# of Specimen Recorded	# of Specimen Recorded	# of Specimen Recorded
Transport	**railroad**	2	4	0
	roadside, roads	13	60	17
	water course	0	21	8
	paths	0	0	4
	roadside meadow	0	2	3
Man-made	**old field, field**	7	132	34
	ore pit	0	4	0
	open lots	0	4	4
	waste ground	2	44	19
	orchards, gardens	0	12	2
Natural	**forest, woods**	1	12	3
	meadow	6	39	7
	thicket	0	10	1
	hillside slope	0	11	0
	swamp	0	4	0
	No data	40	15	6

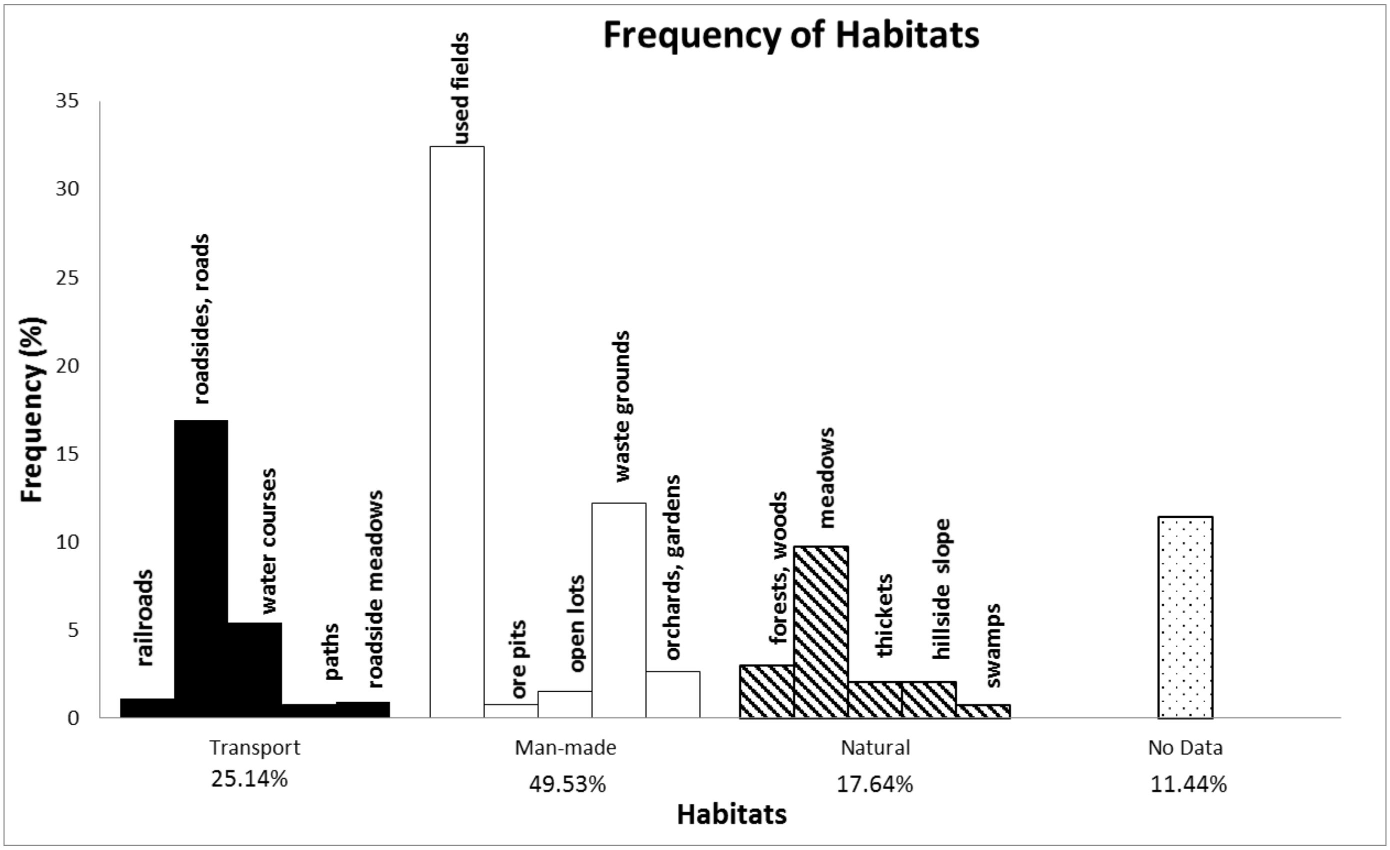


Figure 4. Frequency of habitats from all three of the time periods.

Works Cited

Delisle, F., Lavoie, C., Jean, M., & Lachance, D. (2003). Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. *Journal of Biogeography*, *30*(7), 1033-1042.

Donald, W. W. (1994). The biology of Canada thistle (*C. arvense*). Reviews of Weed Science, 6, 77-101.

Lavoie, C., Jodoin, Y., Merlis, D., & Goursaud, A. (2007). How did common ragweed (Ambrosia artemisiifolia L.) spread in Quebec? A historical analysis using herbarium records. *Journal of Biogeography*, *34*(10), 1751-1761.

Pennsylvania Department of Conservation and Natural Resources . (2018). Invasive Plants in Pennsylvania: Canada Thistle. Retrieved January 08, 2018, from http://www.docs.dcnr.pa.gov/cs/groups/public/documents/document/dcnr_010290.pdf

Pyšek, P. (1991). Heracleum mantegazzianum in the Czech Republic: dynamics of spreading from the historical perspective. *Folia geobotanica et phytotaxonomica*, *26*(4), 439-454.

Pyšek, P., & Prach, K. (1993). Plant invasions and the role of riparian habitats: a comparison of four species alien to central Europe. In *Ecosystem Management* (pp. 254-263). Springer, New York, NY.

Reuell, P. (2017, December 22). Harvard study illuminates botanical bias. Retrieved from https://news.harvard.edu/gazette/story/2017/12/harvard-study-illuminates-botanical-bias/

Rhoads, A. F., & Block, T. A. (2007). The plants of Pennsylvania: an illustrated manual. University of Pennsylvania Press. 884
Stachon, W. J., & Zimdahl, R. L. (1980). Allelopathic activity of Canada thistle (*C. arvense*) in Colorado. *Weed Science*, *28*(1), 83-86.

Tulane University . (2014). Overview. Retrieved January 08, 2018, from http://www.museum.tulane.edu/geolocate/about.html

University of Georgia Center for Invasive Species and Ecosystem Health. (2018). *C. arvense*. Retrieved August 18, 2017, from http://www.eddmaps.org/ipane/ipanespecies/herbs/cirsium_arvense.htm