

Off-Campus Recruiting: Methodology

1 Collecting Off-Campus Recruiting Data

Our project collects data on off-campus recruiting by colleges and universities. Many institutions advertise off-campus recruiting events on their admissions websites (e.g. “coming to your area” links). For example, Figure 1 shows part of a webpage that lists recruitment visits by the University of South Carolina during November 2016. Python, a general-purpose programming language (e.g., web-design, game development, data mining), was used to collect these recruiting data by web-scraping admissions websites. Python programs were automated to “crawl” relevant pages on admissions websites and “scrape” all information containing recruiting events once a week from 1/1/2017 to 12/31/2017.

To select the analysis sample for the broader project, we investigated the admissions websites of the following institutions: all public research-extensive universities as defined by the 2000 Carnegie Classification (N=102); all private universities in the top 100 of U.S. News and World Report National Universities rankings (N=58); and all private colleges in the top 50 of U.S. News and World Report Liberal Arts Colleges rankings (N=47). All institutions that posted off-campus recruiting visits on their admissions website were included in our analysis sample. The resulting data collection sample consists of 49 public research universities, 49 private research universities, and 42 private liberal arts colleges.

For each institution in the data collection sample, we investigated the entire university website searching for URLs containing data on off-campus recruiting events. This process was conducted independently by two members of the research team to avoid missing any relevant URLs. Our programs also scraped data about participation in national college fairs from the National Association for College Admission Counseling (NACAC) website. Additionally, we collected data about participation in “group travel tours” from websites advertising joint recruiting events by multiple universities (e.g. Peach State Tour by Georgia State University, Georgia Tech, and The University of Georgia). Since URLs containing data on off-campus recruiting events often change (e.g., a university creates a new URL or changes the formatting of an existing URL), we completed this investigation process for each university every 2 months and data collection scripts were updated accordingly.

2 Defining Off-Campus Recruiting

Off-campus recruiting events can be categorized along the dimensions of event type, event host, and event location. Event type includes college fairs (in which representatives from multiple colleges attend), day-time representative visits to a high school (in which a representative from one college attends), group travel visits, formal admissions interviews, admitted student events, and committed student events. Event hosts include paid staff, paid consultants (e.g. regional recruiters contracted by several institutions), alumni, and current students. Event locations include high schools, community colleges, hotels, conference/convention centers, and other public places (e.g., cafes).

For the purpose of our research, we define off-campus recruiting events as those focused on soliciting undergraduate admissions applications, hosted by paid personnel or consultants at any off-campus location. This definition excludes admitted and committed student events, but includes guidance counselor events.

Additionally, we excluded formal one-on-one formal interviews because these events are focused on determining the admission of one particular student rather than an open event soliciting applications from many prospective students. We excluded events hosted by alumni or student volunteers. Our rationale is that theories of organizational behavior suggest that duties performed by paid staff are a better indicator of organizational priorities than duties performed by volunteers (Meyer & Rowan, 1977).

3 Data Completeness

We do not claim to identify all off-campus recruiting events for institutions in our data collection sample. Based on prior research and conversations with admissions professionals, nearly all colleges and universities convene three broad types of off-campus recruiting events: (1) receptions/college fairs at hotels and convention centers; (2) evening college fairs at local high schools; and (3) day-time representative visits at local high schools. However, some institutions we collected data on did not post all three types of recruiting events on their website (e.g., posted evening college fairs at local high schools but not day-time visits to local high schools). When choosing sub-samples of institutions to analyze for a particular publication (e.g., AERA conference paper), we only considered institutions that posted all three types of off-campus recruiting events on their website.

However, even if an institution posted the three types of off-campus recruiting events on their website, specific events may have been omitted. Given this potential for missing data, our data should be interpreted as “where colleges and universities say they go” rather than a complete account of all off-campus recruiting events.

4 Data Processing & Secondary Data

We take a multi-step approach to processing information collected from admissions webpages and converting it into tabular recruiting event data. The first step begins with automated Python scripts, as described above, scraping all information on admission webpages and storing the information as HTML text into a Structured Query Language (SQL) database on a remote server. Separate scripts parse the HTML text into tabular data (e.g., columns for event date, event time, school name, address). Institutions advertise recruiting events weeks prior to the event date and typically remove events from admissions websites after they have occurred. Because our programs are automated to scrape websites weekly, our data collection includes multiple “scrapes” of the same event until it has occurred and is removed from the website. Duplicate event records caused by automated weekly scrapings are removed after HTML is converted into tabular data.

Due to variations in the location information provided for events across institutions, tabular data is geocoded. Geocoding is the process of converting limited location information (e.g., school name, city, state) into geographic coordinates. Geocoding scripts take location information, query the Google Maps Application Program Interface (API), and return more detailed geographic information for each event (e.g., latitude and longitude coordinates, county, city, state, full street address, zip code). For events with limited information provided from admissions websites (e.g. “Holy Spirit Preparatory School, Atlanta, GA”), this querying process returns multiple possibilities for geographic location (e.g, Possible Address #1: “4465 Northside Dr NW, Atlanta, GA 30327, USA”; Possible Address #2: “4449 Northside Dr NW, Atlanta, GA 30327, USA”). For events returning multiple options during geocoding, our research team “human” geocoded these events by selecting the correct option after investigating the given recruiting event location.

After parsing and geocoding, our events are categorized by their type and location, and merged accordingly to secondary data on community and school characteristics. For events not at a high school or community college, zip code is used to link recruiting events to detailed demographic and economic data from the U.S. Census Bureau’s American Community Survey (ACS). For events located at schools, latitude

and longitude coordinates were used to link recruiting events to data from the National Center for Education Statistics (NCES). Specifically, we obtain data on public high schools from the Common Core of Data (CCD), data on private high schools from the Private School Universe Survey (PSS), and data on community colleges (and the universities in the sample) from the Integrated Postsecondary Data System (IPEDS). Because there were large variations in street address and school name formatting in NCES datasets, we utilized longitude and latitude fields provided for all school records for merging recruiting data to NCES datasets.

Our geocoding process returned more precise latitude and longitude coordinates than those provided in NCES datasets (coordinates with 8 decimal points approximating a location to within 1 millimeter), making the merging based on exact match of coordinate points not possible. Due to this difference in precision, our merging programs linked “nearest” pairs of coordinate points between our recruiting data and NCES datasets based on shortest ellipsoidal distance (surface distance between two points on the surface of the earth). Merging was confirmed by exact matches in the following fields between recruiting events and NCES data: digits or “house numbering” within street address, city, state, and zip code.

Non-matches between recruiting events and NCES datasets were investigated by members of our research team. Non-matches were most often the result of universities visiting independent schools. Because NCES does not provide data for independent schools, these events were categorized as community events rather than school events. Other non-matches were due to schools moving locations, closed schools, and lack of precision in NCES latitude and coordinates for some school records.

5 High School Sample

5.1 Public high schools

Schools that satisfied the following criteria were included in the sample:

- Offers grades 9-12 and enrolls more than ten students in 12th grade
- Located in the 50 U.S. states, the District of Columbia, or land regulated by the Bureau of Indian Affairs
- Is not a special education school, an alternative school, a virtual school, or an independent school
- Is an open status school and reports enrollment to the Federal Department of Education

5.2 Private high schools

Schools that satisfied the following criteria were included in the sample:

- Offers grades 9-12 and enrolls more than ten students in 12th grade
- Located in the 50 U.S. states, the District of Columbia, or land regulated by the Bureau of Indian Affairs
- Is not a special education school, an alternative school, an early childhood center, or an independent school

References

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: formal structure as myth and ceremony. *The American Journal of Sociology*, 83(2), 340-363.

Figure 1: Partial listing of visits to high schools by University of South Carolina during November 2016



Office of Undergraduate Admissions

Admissions Events

Come meet us when we are in your area.

United States—California

- **Trabuco Hills High School**
Thursday, November 3 at 12:22 PM
Mission Viejo, CA
- **Tustin Unified College and Career Fair**
Monday, November 7 at 6:00 PM
Tustin, CA

United States—Georgia

- **Augusta Preparatory Day School**
Thursday, November 3 at 12:45 PM
Martinez, GA

United States—Kentucky

- **Cooper High School College Fair**
Thursday, November 17 at 8:00 AM
Union, KY
- **Randall Cooper High School College Fair**
Thursday, November 17 at 8:00 AM
Union, KY
- **Woodford County High School College Fair**
Thursday, November 17 at 6:30 PM
Versailles, KY
- **Covington Latin High School College Fair**
Monday, April 3 at 6:30 PM
Covington, KY

United States—Maryland

- **Mount Hebron High School**
Thursday, November 3 at 11:00 AM
Ellicott City, MD

United States—Massachusetts

- **Newton South High School**
Friday, November 4 at 7:45 AM
Newton Centre, MA
- **Dexter Southfield School**
Friday, November 4 at 9:00 AM
Brookline, MA
- **Brookline High School**
Friday, November 4 at 10:20 AM
Brookline, MA
- **Beaver Country Day School**
Friday, November 4 at 11:45 AM
Chestnut Hill, MA
- **Boston Latin School**
Tuesday, November 8 at 2:15 PM
Boston, MA
- **Walpole High School**
Tuesday, November 29 at 1:30 PM
Walpole, MA

United States—New Jersey

- **Summit High School College Fair**
Wednesday, November 2 at 7:00 PM
Summit, NJ