INSTITUTE FOR DEFENSE ANALYSES

Geographic Diversity in Military Recruiting

Matthew S. Goldberg, Project Leader
Karen Cheng
Nancy M. Huff
Dennis D. Kimko
Alexandra M. Saizan

November 2018
Approved for public release;
distribution is unlimited.
IDA Paper P-9079
Log: H 18-000184

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About this Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, project order number BE-7-4213, "Enlistment Supply by Geography and Familiarity with the Military," for the Director for Accession Policy, Office of the Assistant Secretary of Defense (Manpower and Reserve Affairs), Office of the Under Secretary of Defense (Personnel and Readiness). The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
Thank you to Dennis W. Kuo and Susan L. Rose for performing technical review of this document.

For More Information:
Matthew S. Goldberg, Project Leader
mgoldber@ida.org, (703) 578-2728

David J. Nicholls, Director, Cost Analysis and Research Division
dnicholl@ida.org, (703) 575-4991

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-9079

Geographic Diversity in Military Recruiting

Matthew S. Goldberg, Project Leader
Karen Cheng
Nancy M. Huff
Dennis D. Kimko
Alexandra M. Saizan

Executive Summary

Background

From the beginning of the all-volunteer force (AVF) in 1974—and, indeed, in the public dialogue leading up to the implementation of the AVF—concerns were raised about various aspects of diversity in the military. The Congress chartered the Defense Manpower Commission in 1973; the Commission reported its findings in 1976. The Commission pondered whether the diversity in race, gender, economic status, educational status, and tested aptitude that been achieved to varying degrees under the draft could be sustained in an AVF. The Commission's report highlighted geographical diversity as the one dimension along which the draft had largely been successful.[1]

Concerns about geographical diversity were publicly aired by the Department of Defense (DoD) as early as 1987, about a dozen years into the AVF. The 1987 report noted the stronger recruiting performance in the rural areas and in the south and southwest. The report attributed that phenomenon to a higher concentration of military installations in those areas, to larger numbers of military retirees, and to individuals with stronger military orientation.[2] Although the stellar recruiting performance from those areas of the United States was certainly welcome, the concern was whether the all-volunteer military might become isolated to largely just those areas.

Much more recently, in 2016, then-Secretary of Defense Ash Carter opined that

> too many of America's young men and women have no personal connection to our military. As a result, they give no real consideration to the possibility of joining us … It is my firm conviction that the Department of Defense must have access to 100 percent of America's population for our all-volunteer force to be able to recruit and retain the highly qualified men and women needed for the Force of the Future.[3]

Secretary Carter specifically identified geography as one of the dimensions along which the military needed to improve its diversity.

---

1 Defense Manpower Commission, Staff Studies and Supporting Papers, Volume III, *Military Recruitment and Accessions and Future of the All-Volunteer Force*, Appendix D, May 1976, D-5–D-6.

2 Department of Defense, *Population Representation in the Military Services, Fiscal Year 1987*, II-11.

3 Secretary of Defense Ash Carter, "Forging Two New Links to the Force of the Future," Memorandum, November 1, 2016, http://www.defense.gov/Portals/1/Documents/pubs/Forging-Two-New-Links-Force-of-the-Future-1-Nov-16.pdf.

This project was sponsored by the Office of the Under Secretary of Defense for Personnel and Readiness (OUSD(P&R)), Manpower and Reserve Affairs (M&RA), Military Personnel Policy (MPP), Director of Accession Policy (AP). The first objective of this research was to identify the demographic, economic, and cultural factors in a community that predict the percentage of its youth population that enlist in the active components of the US military. When viewed across the entire nation, those are the factors that determine geographical diversity. The second objective was to identify trends and events that could affect geographical diversity in the future.

Traditional Analysis of Enlisted Accessions

Youth unemployment is often portrayed as a major barometer of the recruiting climate. Although surely not the only determinant of enlisted accessions, it is interesting to examine the extent to which that one factor can explain the differences in recruiting success both across regions and over time. This paper begins with a traditional correlation analysis to measure the relationship between the youth unemployment rate and enlisted accessions over time within each of the four regions into which the Census divides the US. To varying degrees across the four regions, when a region is experiencing higher youth unemployment (i.e., a weaker economy) than the national average, the region's share of recruits relative to its youth population (its *representation ratio*) tends to go up. Conversely, a region with a higher-performing economy tends to supply fewer recruits to the military.

Next, we pursue a traditional regression analysis between accessions from a state and the deviation of the youth unemployment rate in the state from the average for that year within its Census region. Once again, the strength of the statistical relationship between a state's representation ratio and its youth unemployment rate varies across the four Census regions. Because the results of the correlation and regression analyses are mixed, we then proceed to a machine learning approach that introduces over 100 additional predictors of recruiting success, all measurable at finer levels of geographical detail than the entire state.

Machine Learning Analysis of Enlisted Accessions

Additional demographic, economic, and cultural factors may be measured at finer levels of geography such as counties. In addition, modern machine learning methods do not limit the researcher to a small number of explanatory variables in order to satisfy the traditional regression paradigm that the dataset be "long" (i.e., contain many fewer explanatory variables than the number of observations in the sample). Instead, machine learning methods are specifically designed for "wide" datasets, enabling the researcher to consider many more potential variables in a much more agnostic fashion.

We first apply the machine learning analysis to all accessions—both male and female, without regard to quality level. Then we apply the analysis to the high-interest subset of high-quality male accessions. "High-quality" is defined as high school graduates and seniors who score above the median (categories 1-3A) on the Armed Forces Qualification Test (AFQT), which is a subset of the Armed Services Vocational Aptitude Battery (ASVAB). Finally, we perform the corresponding analysis for high-quality female accessions. This section summarizes the most influential factors that affected all three populations studied.

Three measures of veteran presence in the community were among the most important predictors of recruiting success: the percentage of veterans among the population age 18 and older, the percentage of veterans among the population ages 35–54, and the percentage of veterans who served during the first Gulf War. The first of those measures is an indicator of contact with veterans, regardless of the ages of those veterans or the period in which they served. With increasing mortality among veterans from the World War II and Korean War eras, the veteran percentage has been consistently declining over the past decade, perhaps signaling a lower societal attachment to the military and tending to drive the number of accessions down. Although adverse trends can often be offset using tools such as recruiting effort or enlistment bonuses, there is no direct policy lever that the recruiting community can exercise to reverse the decline in the overall veteran population.

The next two measures pertain to slightly different groups of veterans but in essentially the same age range, 35–54. We find that for predicting recruiting success, the pure age effect seems more important than the particular era in which those veterans served. The age range 35–54 represents well family connections and mentors: parents, aunts and uncles, teachers, and coaches who may have served in the military. This variable, too, while having a positive association with accessions, has been declining over the past decade, tending to drive the number of accessions down. Nor is it susceptible to any direct policy lever.

The next important variable is the percentage of people currently attending college. College attendance works in two ways against military recruiting. First, youth currently enrolled in college are not immediately available for military service, although it must be remembered that about 40 percent of them do not complete college within six years and are an important pool of potential recruits. Second, high levels of college attendance in a community may signal a culture in which most high school graduates are expected to enroll in and complete college. College enrollment varies geographically and the national averages have shown some volatility with the economy and other factors. Many youth enrolled in college when civilian employment prospects were relatively poor during the Great Recession of 2007–2009, and remained in college to complete their four-year degrees (college attendance peaked in 2012 and remained elevated through 2013). Although trends in the national economy are not subject to policy intervention at the DoD level, they do form a predictable influence on the recruiting climate.

Two measures of Junior Reserve Officers' Training Corps (JROTC) density have positive associations with recruiting: the number of high schools that offer a JROTC program per youth population, and the number of cadets per youth population. The JROTC program is jointly funded by DoD (about $375 million per year) and the local school districts (about $225 million). However, DoD's ability to further invest in JROTC is constrained not only by its budget but also by the congressional cap of between 3,000 and 3,700 JROTC units at any point in time.

Finally, although a smaller effect, having more students take the ASVAB is associated with higher recruiting totals. An encouraging sign is that the number of students taking the ASVAB has increased over the past decade. Although DoD has the option to encourage more high schools to administer the ASVAB, that initiative is largely saturated to the point where the limiting factor is not DoD's policy or level of effort but rather the number of willing participants. Moreover, it is difficult to sort out the direction of causality—does more ASVAB testing increase the propensity to enlist, or does it just signal jurisdictions that fundamentally have a more positive sentiment toward military service?

Primacy of Cultural and Demographic Factors

Conspicuous by their absence from the sets of most important predictors are several economic variables that we included among the more than 100 total variables in the machine learning analysis. Although included in the preliminary correlation and regression analyses, we did not include youth unemployment rates in the machine learning analysis because age-specific unemployment rates by county were not available. Two of the important predictors that were included in the machine learning analysis may be construed as economic in nature: the percentage of people currently attending college, and the percentage of people having completed some college. Having controlled for college attendance and attainment, and notwithstanding the statistical association with youth unemployment rates, recruiting success in local areas appears to be driven largely by cultural and demographic factors.

Limitations

The Defense Manpower Data Center (DMDC) was able to supply the Institute for Defense Analyses (IDA) with recent information on numbers of recruiters for each active service branch at the recruiting station level. However, DMDC could not provide historical time-series data on recruiters to match the historical data on numbers of accessions and the other measured factors that are associated with accessions. The lack of historical data on recruiters may confound the estimated relationships that IDA established between recruiting success and the variables we were able to collect. IDA's findings should be tempered by that understanding.

Contents

1. Introduction

From the beginning of the all-volunteer force (AVF) in 1974—and, indeed, in the public dialogue leading up to the implementation of the AVF—concerns were raised about various aspects of diversity in the military. The Congress chartered the Defense Manpower Commission in 1973; the Commission reported its findings in 1976. The Commission pondered whether the diversity in race, gender, economic status, educational status, and tested aptitude that had been achieved to varying degrees under the draft could be sustained in an AVF. The Commission's report highlighted geographical diversity as the one dimension along which the draft had largely been successful.[1]

Concerns about geographical diversity were publicly aired by the Department of Defense (DoD) as early as 1987, about a dozen years into the AVF. The 1987 report noted the stronger recruiting performance in the rural areas and in the south and southwest. The report attributed that phenomenon to a higher concentration of military installations in those areas, to larger numbers of military retirees, and to individuals with stronger military orientation.[2] Although the stellar recruiting performance from those areas of the United States was certainly welcome, the concern was whether the all-volunteer military might become isolated to largely just those areas.

Much more recently, in 2016, then-Secretary of Defense Ash Carter opined that

> too many of America's young men and women have no personal connection to our military. As a result, they give no real consideration to the possibility of joining us … It is my firm conviction that the Department of Defense must have access to 100 percent of America's population for our all-volunteer force to be able to recruit and retain the highly qualified men and women needed for the Force of the Future.[3]

---

1 Defense Manpower Commission, Staff Studies and Supporting Papers, Volume III, *Military Recruitment and Accessions and Future of the All-Volunteer Force*, Appendix D, May 1976, D-5–D-6, https://www.dtic.mil/docs/citations/ADA029951.

2 Department of Defense (DoD), *Population Representation in the Military Services, Fiscal Year 1987*, II-11, http://www.cna.org/research/pop-rep.

3 Secretary of Defense Ashton Carter, "Forging Two New Links to the Force of the Future," Memorandum, November 1, 2016, http://www.defense.gov/Portals/1/Documents/pubs/Forging-Two-New-Links-Force-of-the-Future-1-Nov-16.pdf.

Secretary Carter specifically identified geography as one of the dimensions along which the military needed to improve its diversity.

Traditional, regression-based studies predict accessions often using data at the state or recruiting battalion level and include factors such as relative military pay, youth unemployment, various measures of recruiting effort (including marketing volume, enlistment bonuses, and the number of recruiters), population composition by race and ethnicity, and (more recently) wartime troop levels and casualties.[4] However, additional demographic, economic, and cultural factors may be measured at finer levels of geography such as counties. In addition, modern machine learning methods do not limit the researcher to a small number of explanatory variables in order to satisfy the traditional regression paradigm that the dataset be "long" (i.e., contain many fewer explanatory variables than the number of observations in the sample). Instead, machine learning methods are specifically designed for "wide" datasets, enabling the researcher to consider many more potential variables in a much more agnostic fashion.

This project was sponsored by the Office of the Under Secretary of Defense for Personnel and Readiness (OUSD(P&R)), Manpower and Reserve Affairs (M&RA), Military Personnel Policy (MPP), Director of Accession Policy (AP). The first objective of this research was to identify the demographic, economic, and cultural factors in a community that predict the percentage of its youth population that enlist in the active components of the US military. When viewed across the entire nation, those are the factors that determine geographical diversity. The second objective was to identify trends and events that could affect geographical diversity in the future.

Chapter 2 provides a history of the concerns regarding geographical diversity that have been raised since the advent of the AVF.

Chapter 3 first applies traditional correlation analysis to measure the relationship between the youth unemployment rate and non-prior-service (NPS) enlisted accessions over time within the four Census regions. Then follows a traditional regression analysis between accessions from a state and the deviation of the youth unemployment rate in the state from the average for that year within its Census region.

Chapter 4 exposits the machine learning approaches of regression trees and, in particular, a recent extension of regression trees known as random forests.

Chapter 5 discusses the collection of data at finer levels than states, including counties and a new concept we developed called pseudo-counties. The latter are clusters of

---

[4] For a recent example that accounts for many of those factors, see Lawrence Goldberg, Dennis D. Kimko, and Maggie X. Li, "Analysis and Forecasts of Army Enlistment Supply," IDA Document NS D-5466 (Alexandria, VA: Institute for Defense Analyses, April 2015).

contiguous counties that act as the least populous (smallest) geographical units to which native data from multiple sources can be consistently mapped.

Chapter 6 describes the sources of the data used in the machine learning analysis.

Chapter 7 presents the results of the machine learning analysis. The analysis is first applied to all accessions—both male and female, without regard to quality level. The second and third sections present the findings for high-quality male and female accessions, respectively. “High-quality” is defined as high school graduates and seniors who score above the median (categories 1-3A) on the Armed Forces Qualification Test (AFQT), which is a subset of the Armed Services Vocational Aptitude Battery (ASVAB).

Chapter 8 displays the regional trends over the past decade in the most important factors identified in the machine learning analysis, forming a partial prognosis for the future recruiting environment.

Finally, Chapter 9 presents the conclusions of the research.

2. History of Geographic Diversity in the All-Volunteer Force

This chapter begins by recounting some concerns about geographic diversity that were expressed by then-Secretary of Defense Ashton Carter in 2016. The analysis then looks back to the origins of the all-volunteer force in the 1970s, the concerns that were expressed at that time about the potential for geographic diversity in such a force, and the reality that later emerged. The analysis in this chapter is conducted at various levels of geographical detail, ranging from individual states to clusters of contiguous states as defined by the US Census Bureau. Although there have certainly been changes in the relative representation of the individual states over the past four decades, the relative representation of broad regions of the country (when properly normalized for the size of the youth population)—although unequal—have been remarkably constant over that period.

A. Recent Concerns and Historical Perspective

Secretary Carter noted that about 40 percent of all military recruits come from just six states. Although his statement is true, the situation is a bit more complicated than the statement might indicate. Figure 1 displays the six states that produced the most NPS enlisted accessions into the Active Components during Fiscal Year (FY) 2016—some 43 percent of the national total. The figure also displays, along the horizontal axis, the population aged 18–24 years in each of those states. In total, the six states that produced 43 percent of enlisted accessions also had nearly the same proportion of the national youth population—about 41 percent. However, the relationship between youth population and the number of accessions varies across the states. California, the nation's most populous state, yielded the most accessions, nearly 18,000. But relative to its youth population, California ranked only 34th among the states. Similarly, New York yielded the fifth-largest number of accessions, over 6,000, but ranked only 47th among the states relative to its youth population. The diagonal line in Figure 1 shows the number of accessions that a state with a given youth population would yield if the state performed at the national average level. Both California and New York lie below the line. In contrast, Florida lies above the line: Florida generated 81 percent more accessions than New York with only 88 percent of the youth population. Florida's performance ranked 2nd among the states in FY 2016; the other states shown in the figure are Georgia (ranked 1st), North Carolina (ranked 14th), and Texas (ranked 17th).

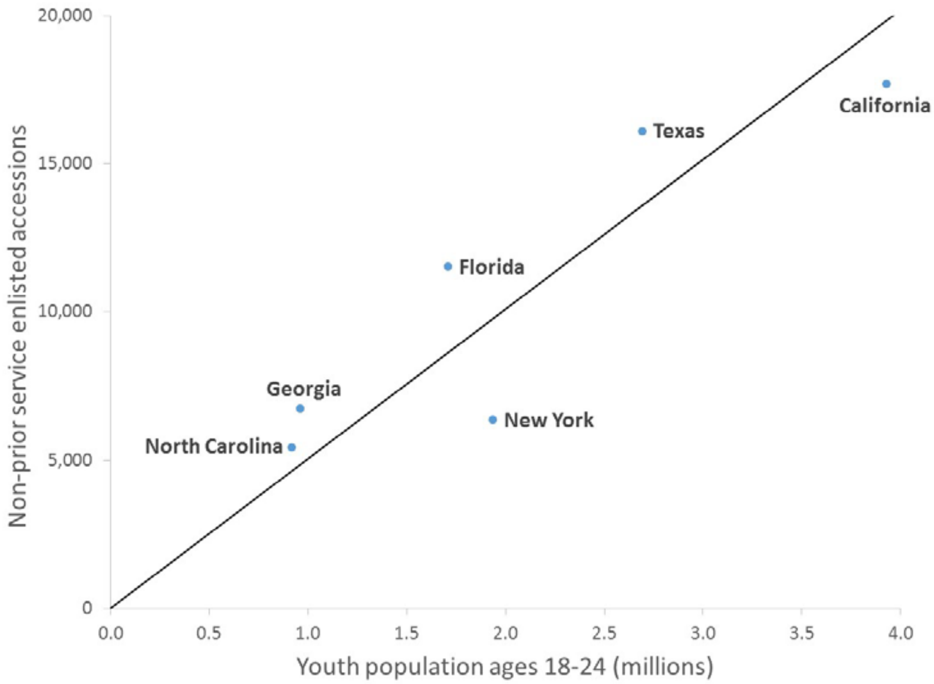


Figure 1. Six States that Yielded the Most Enlisted Accessions for the Active Components in FY 2016

Secretary Carter was concerned that "too many of America's young men and women have no personal connection to our military. As a result, they give no real consideration to the possibility of joining us." He went on to state that, "It is my firm conviction that the Department of Defense must have access to 100 percent of America's population for our all-volunteer force to be able to recruit and retain the highly qualified men and women needed for the Force of the Future."[5] Secretary Carter identified three types of diversity: geographic, demographic, and generational reach. The term "generational reach" reflects the observation that children of military Service members or of veterans are more likely than others to join the military. While Secretary Carter did not want to discourage those with family connections from joining, he also wanted to encourage those who do not have such connections.

Geographic diversity is primarily an issue for enlisted accessions into the Active Components. The three service academies recruit nationally, but they do not suffer from shortages of qualified potential officers. Military officers who did not attend one of the service academies would have been recruited either from specific colleges or universities having Reserve Officers' Training Corps (ROTC) programs, or from the general pool of college graduates who enter through Officer Candidate School (OCS). ROTC and OCS officers did not necessarily attend a college or university near their home town, or even

---

5 Carter, "Forging Two New Links to the Force of the Future."

one within their home state. Thus, the issue of filling the officer ranks is somewhat disconnected from where the potential officers grew up. Each state recruits its own National Guard units to local armories, so the issue of states' representation in a national recruiting pool for Guard members is not germane.

Concerns about various dimensions of diversity in the military go back to the beginning of the AVF in 1974.[6] The Congress chartered the Defense Manpower Commission in 1973; the Commission reported its findings in 1976. The Commission's report included an analysis of several dimensions of diversity: race, gender, economic status, educational status, mental group (i.e., percentile grouping on the AFQT), and geography. The Commission found that:

> From 1940 until the cessation of active inductions in 1973, the Selective Service insured [sic] a *generally representative geographical distribution of young men* [emphasis added] in the armed forces. Due to differences in regional educational systems, volunteerism rates, health conditions, etc., there were some differences; however, they were not significant.[7]

Continuing into the emerging AVF years:

> An analysis of accession and force characteristics during both the draft years and the current AVF years [at best through 1975 or partway into 1976] supports the conclusion that the armed forces never have been completely representative, *with the possible exception of geographical factors* [emphasis added]. In the other areas of prime interest — racial, educational, male-female and economic — there have been both wide ranges and changes in representation levels and limited past concern on the part of policy makers, the public or Congress as to the significance of these non-representative factors.[8]

DoD's Office of the Under Secretary of Defense for Personnel and Readiness (OUSD(P&R)) and its predecessor organizations have been reporting on population representation in the military annually since 1975.[9] Contrary to the second quotation from the Defense Manpower Commission's 1976 report, DoD found evidence of geographical

---

6 For a good summary of the early concerns regarding diversity in the AVF, see Bernard Rostker, *I Want You! The Evolution of the All-Volunteer Force*, MG-265-RC (Santa Monica, CA: RAND Corporation, 2006), especially pages 273–75 and 320–9, https://www.rand.org/pubs/monographs/MG265.html.

7 Defense Manpower Commission, *Military Recruitment and Accessions and Future of the All-Volunteer Force,* Appendix D, D-5.

8 Ibid., D-6.

9 The annual population representation reports were originally mandated by the Senate Committee on Armed Services (S. Rep. 93-884, May 1974). There have been several reorganizations over the years that changed the name of the responsible office within DoD. For example, the FY 1987 report came out under the imprimatur of the Office of the Assistant Secretary of Defense (Force Management and Personnel). In recent years the reports have been produced by CNA on behalf of the Under Secretary of Defense (P&R), and are catalogued at www.cna.org/research/pop-rep.

disparities about a dozen years into the AVF, consistent with the "generational reach" concept later expressed by Secretary Carter:

> In the early years of the all-voluntary military, there was considerable interest in the geographic distribution of new recruits and the population density (urban versus rural) of their hometowns. Interest in the geographic origin of enlistees stemmed largely from the view that an all-volunteer service would attract primarily young men from rural areas and from the south and southwest. Since these areas contain *a higher concentration of military installations, a larger number of military retirees, and individuals with stronger military orientation*, the Services have traditionally received their greatest acceptance and support there [emphasis added]. Thus, defenders of conscription were concerned that a regional bias among Service members might threaten to isolate the military ideologically from the rest of the Nation.[10]

B. Variation across States

As an alternative to the depiction of six states in Figure 1, the numbers of accessions per youth population may be displayed in a heat map for all 50 states, as in Figure 2, which pertains to FY 2010. The color spectrum from bright yellow to deep red corresponds to increasing values of the accession rate. Above-average rates (particular when compared to their immediately neighboring states) are prominent in Maine and New Hampshire; in a band of states in the southeast; and in the three Mountain states of Idaho, Montana, and Wyoming. Among the states with below-average accession rates are a band of states ranging from New Jersey and New York up into New England as far north as Vermont; North Dakota; and Utah.

---

[10] DoD, *Population Representation in the Military Services, Fiscal Year 1987*, II-11.

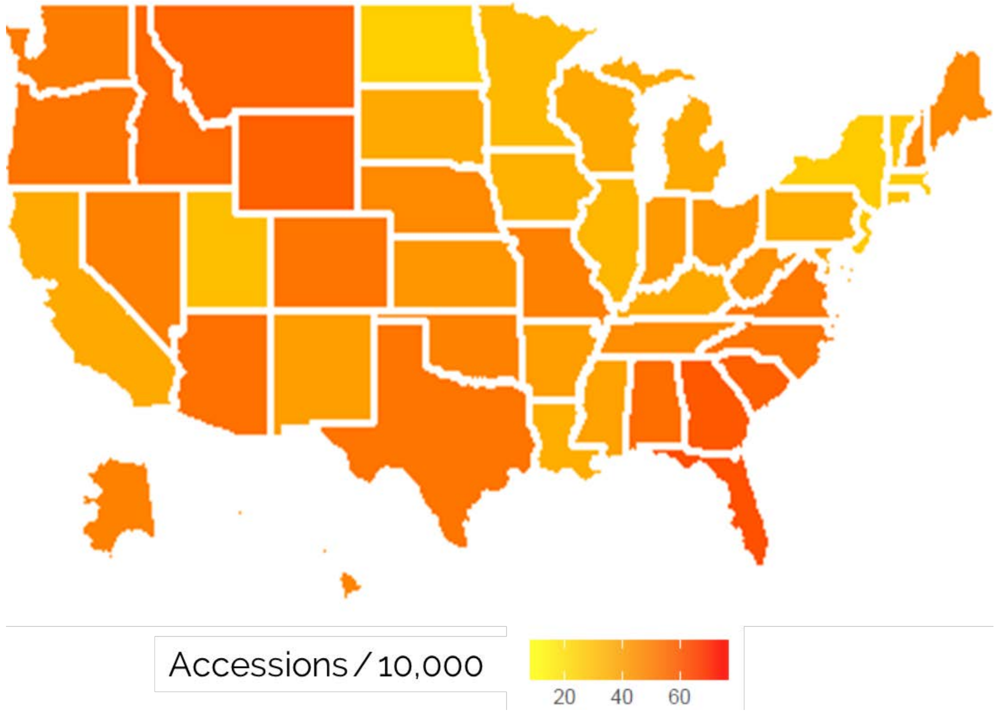


Figure 2. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2010

Accession rates vary over time as well as across states. Figure 3 depicts the situation three years later, in FY 2013. Idaho still had an above-average rate in FY 2013, but Montana regressed toward the national average (dropping from 71 to 62 per 10,000) and Wyoming fell to a rate slightly below the national average (dropping from 73 to 49 in FY 2013 when the national average rate was 54). The state-by-state differences over the three years essentially canceled out across the nation, as the total number of accessions (which is largely determined by the four Services' accession goals) increased only slightly from 159,000 to 165,000.

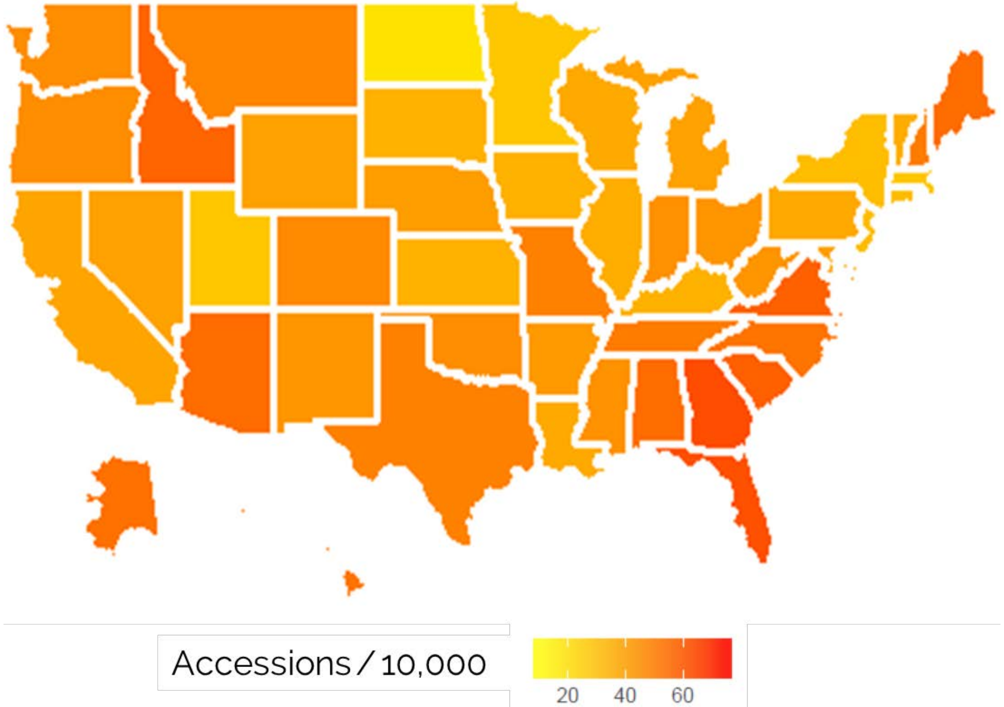


Figure 3. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2013

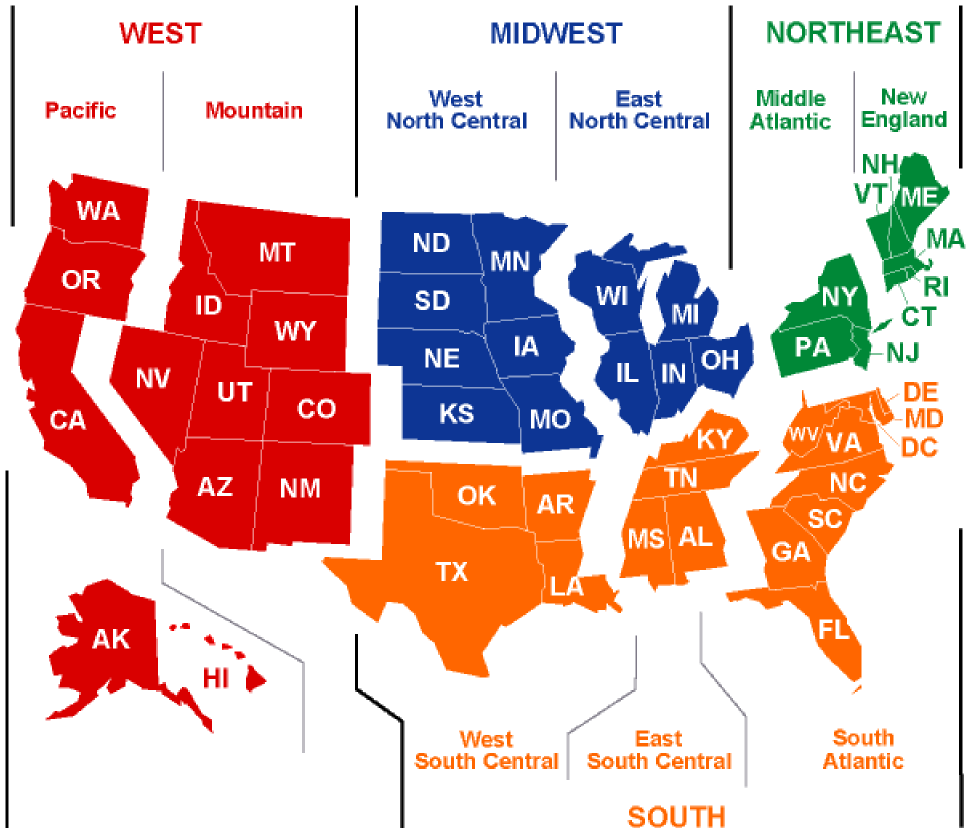
Appendix A contains a complete set of heat maps for each of the years 2007 through 2015.

C. Regional Trend Analysis

By pulling the trend analysis up from states to regions, we may shed additional light on the path from the Defense Manpower Commission's 1976 report—through four decades of DoD's population representation reports—to Secretary Carter's statements from 2016.

The Census Bureau divides the United States into the four regions shown in Figure 4: Midwest, Northeast, South, and West.[11] The four regions are further subdivided into nine divisions.

[11] The Midwest was known as the North Central Region until 1984. See https://www.census.gov/history /www/programs/geography/regions_and_divisions.html.



Source: Kristina Hamachi LaCommare and Joseph H. Eto, "Understanding the Cost of Power Interruptions to U.S. Electricity Consumers," LBNL-55718 (Berkeley, CA: Ernest Orlando Lawrence Berkeley National Laboratory, University of California Berkeley, September 2004), http://eta-publications.lbl.gov/sites/default /files/lbnl-55718.pdf.

Figure 4. US Census Regions and Divisions

Before embarking on the trend analysis, it is instructive to reconsider Figure 1, including all 51 "states" (including the District of Columbia), color-coded to indicate the Census regions to which they belong. (We adopted a different color-coding scheme from that in Figure 4, which we will observe from Figure 5 forward.) Both axes in Figure 5 are measured on logarithmic scales in order to distribute the states more evenly; on a linear scale, the states at the lower end of the distribution would be too tightly clustered so that their identifiers would overlap to the point of illegibility. The figure illustrates both large and small states, and those that over-produce relative to their youth population (they are vertically above the baseline) and those that under-produce (falling below the baseline).

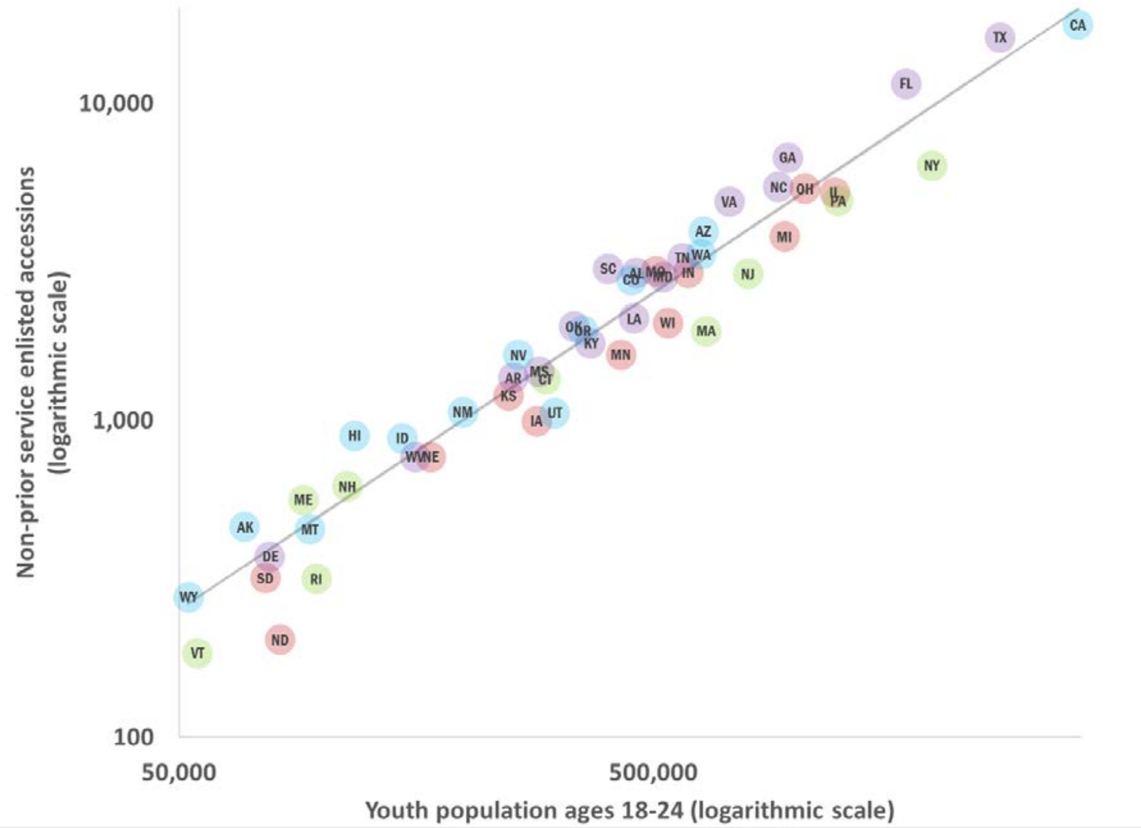
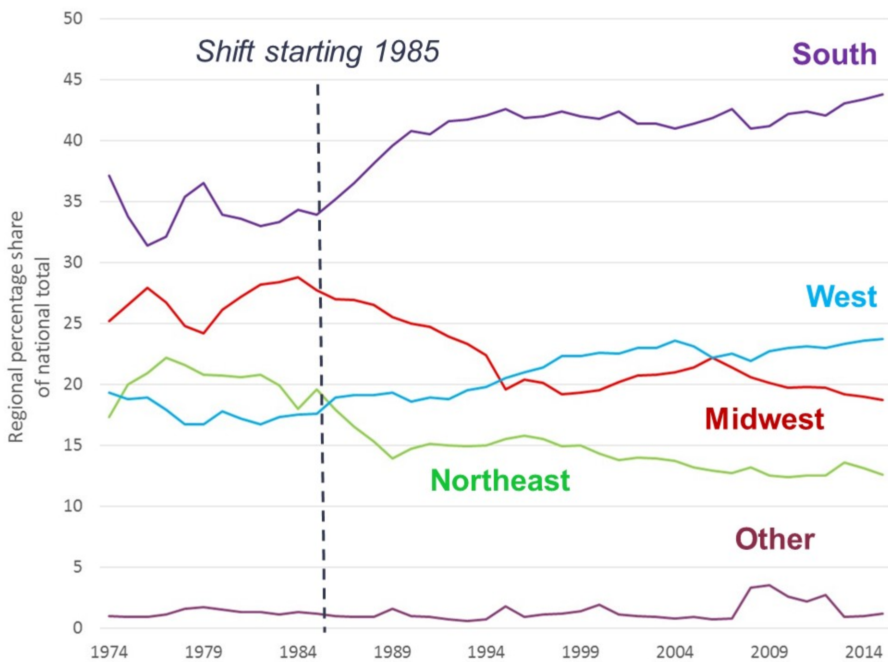


Figure 5. Success of the States and Regions in Recruiting for the Active Components, FY 2016

DoD, in its annual population representation reports at least since FY 2010, has noted that the proportion of NPS accessions began shifting in 1985 toward the South and away from the Midwest and Northeast (see Figure 6, which is reconstructed from the FY 2015 population representation report).[12] Because much of this shift is due to population migration and divergent birth rates across the states and regions, most researchers convert the number of NPS accessions to a *representation ratio*: a state's NPS accessions as a share of the national total, divided by the state's youth population (typically ages 18–24) as a share of the national total. States with representation ratios larger than 1.0 are over-performing relative to their youth population; they are states like Georgia, Florida, North Carolina, and Texas that lie above the "neutral" line in Figure 1. States with representation ratios smaller than 1.0 (like California and New York) are under-performing relative to youth population and lie below the neutral line. The population-weighted average of the representation ratios is exactly 1.0.

[12] DoD, *Population Representation in the Military Services: Fiscal Year 2015*, 23 and Table D-10.



Note: Reconstructed from DoD, *Population Representation in the Military Services: Fiscal Year 2015*. The "Other" region consists of other US territories such as Guam, Puerto Rico, and the Virgin Islands.

Figure 6. Regional Shares of Non-Prior-Service Enlisted Accessions into the Active Components

The trend in representation ratios is shown in Figure 7. The ratios from 1974 come from the Defense Manpower Commission, and the ratios since 1990 were computed from annual data provided to the Institute for Defense Analyses (IDA) by the Defense Manpower Data Center (DMDC). Note that the same four colors identify the four regions in both the squares for 1974 and the lines for 1990–2016. The seemingly paradoxical increase for the South in Figure 5 is removed in Figure 6 when the share of accessions is normalized by the share of youth population, which began increasing in that region in 1985 (the horizontal, dashed purple guideline indicating that the apparently upward-trending line is flattened out by this normalization). The southern share of the youth population ages 17–22 was 31.1 percent in 1974, and the share ages 17–21 was 32.6 percent in 1980. (Those are not the primary age ranges used in the current report, but they are the ones for which the earlier data could be easily obtained.[13]) The southern share of ages 18–24 grew to 34.0 percent by 1990, then continued a gradual increase to 35.1 percent in 2000, 35.8 percent in 2010, and finally 37.1 percent in 2016. Over the entire AVF period, the southern share has grown at an average rate of 1.3 percentage points per decade.

---

[13] Data for 1974 are from Defense Manpower Commission, *Military Recruitment and Accessions and Future of the All-Volunteer Force*, Appendix E, E-34. Data for 1980 are from DoD, *Population Representation in the Military Services, Fiscal Year 1980*, 2.

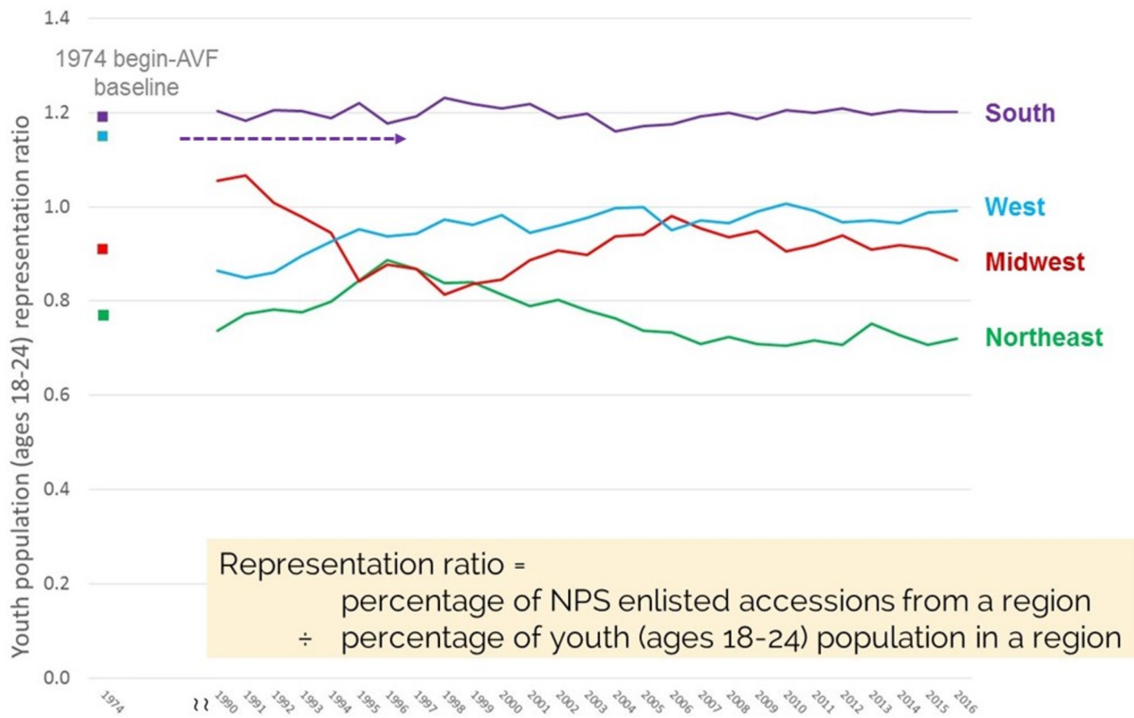


Figure 7. Regional Representation Ratios for the First Year of the AVF (1974) and the Trend since 1990

Disparities among the four regions' geographic representations are evident in Figure 7 as early as 1974, the very first year of the AVF. Tantalizingly, the ordering of the four regions was the same in 1974 as it has been more recently since 2007 (and, excepting one small bump, since 1999).

Both the chart and the story line are a bit more cluttered when selected additional pre-1990 years are added. Data for those years come from early versions of DoD's population representation reports. Not all of those earlier reports could be located, and a few that were located were judged unusable because the youth population counts appeared "stale" and may not have been extrapolated forward since the most recent decennial census. The addition of data from FY 1980, 1986, and 1987 is shown in Figure 8.

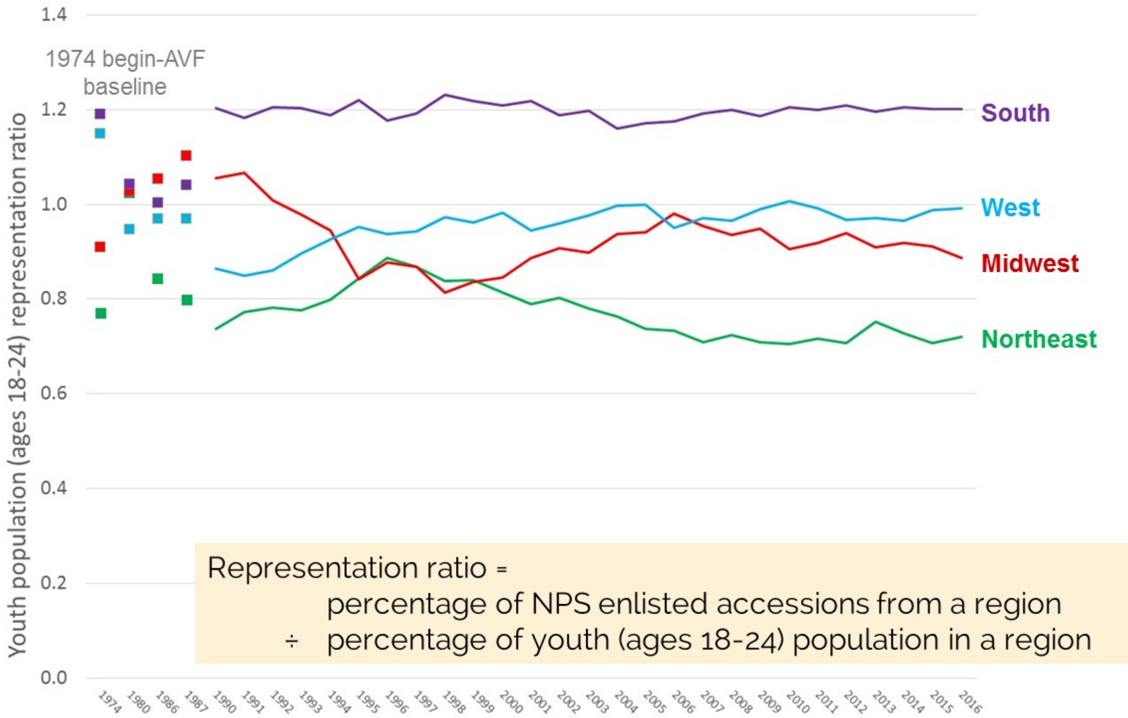


Figure 8. Regional Representation Ratios for Selected Early Years of the AVF and the Trend since 1990

Some interesting patterns are discernible in the squares for the earlier years in Figure 8, before the representation ratios settle into their longer-run trends. For example, the ratios for the South (colored purple) drop from 1.2 in 1974 to barely over 1.0 during the 1980s, before settling back to a reliable 1.2 (i.e., yielding 20 percent more recruits that would be expected from the region's youth population) from 1990 forward. Although the South consistently performs above the national average, the skyrocketing of the southern share of recruits shown earlier in Figure 6 is an artifact of that region's faster-growing youth population.

Another question is whether the four Services individually exhibit the same regional trends that we report for DoD as a whole. The trends for the Army most closely mimic those for all of DoD, perhaps not surprisingly in that Army accessions were, for example, 40 percent of the DoD total in 2015 (see Figure 9).

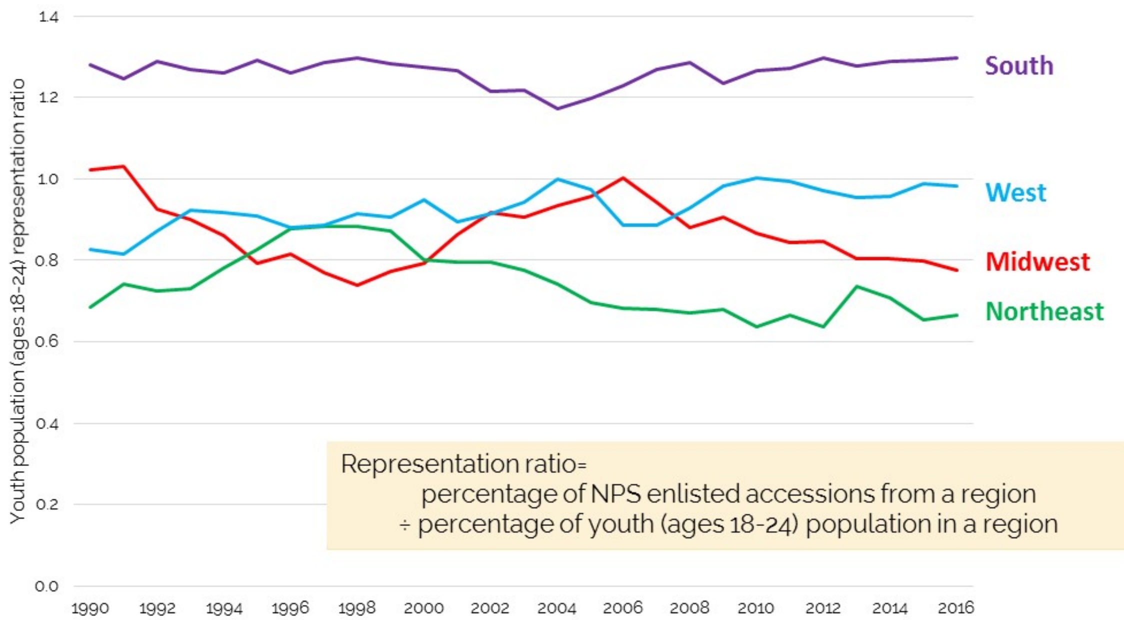


Figure 9. Regional Representation Ratios for Army Enlisted Accessions

The trends for the Navy also mimic those for all of DoD, although not as closely as for the Army (see Figure 10). Like the Army, the Navy had better recruiting success in the Midwest than in the West during the early years of the AVF, but that ordering reversed in more recent years.

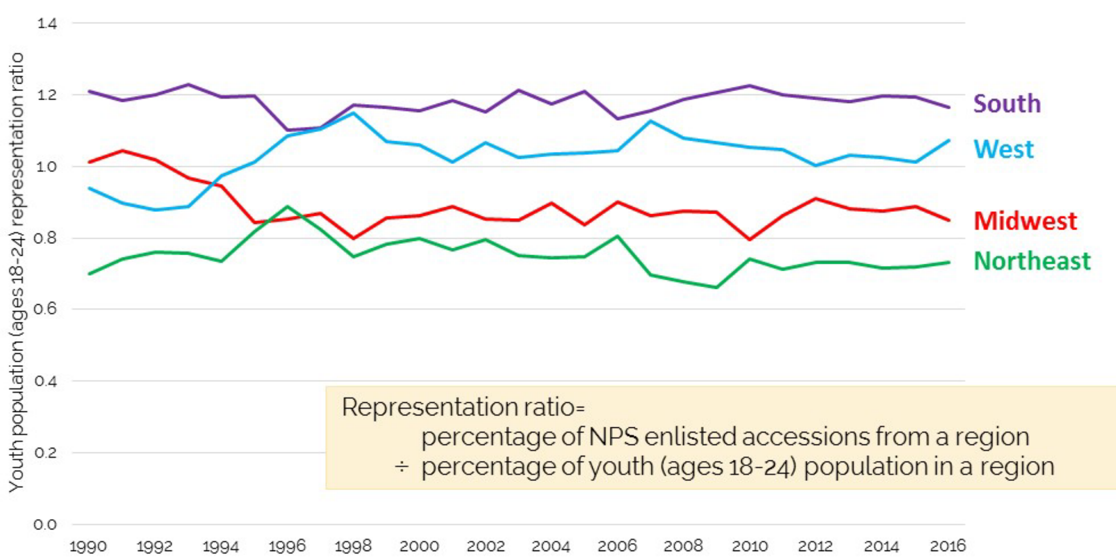


Figure 10. Regional Representation Ratios for Navy Enlisted Accessions

The pattern is roughly similar for the Air Force (see Figure 11). The South and the Northeast have been, respectively, the highest- and lowest-performing regions for the Air Force, with the Midwest and the West intermediate and about on par with each other.

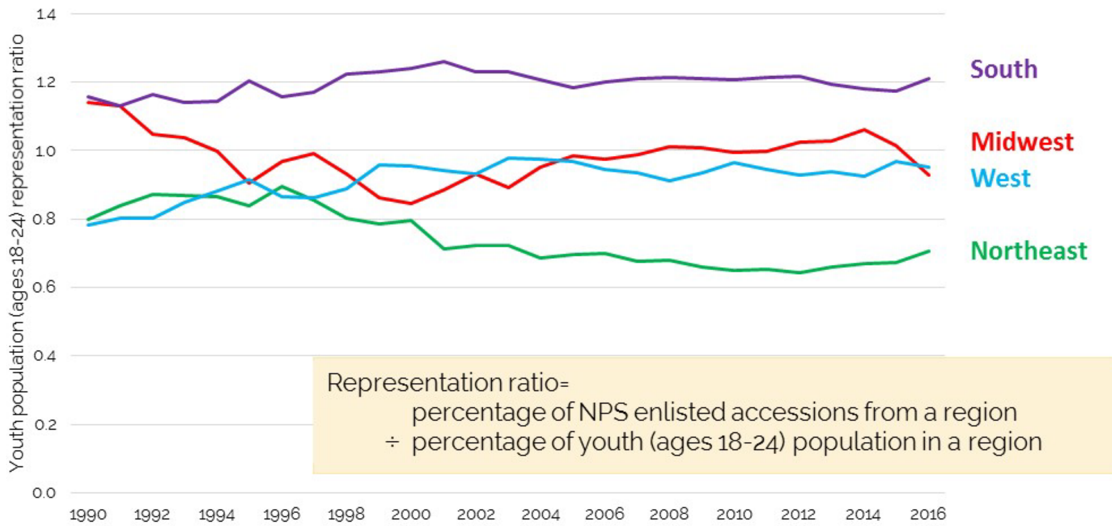


Figure 11. Regional Representation Ratios for Air Force Enlisted Accessions

Finally, among the four Services, the pattern for the Marine Corps is a bit different (see Figure 12). The Midwest and the South have been the highest-performing regions over roughly the past decade, with the Midwest overtaking the West in 2004 and overtaking the South in 2009.

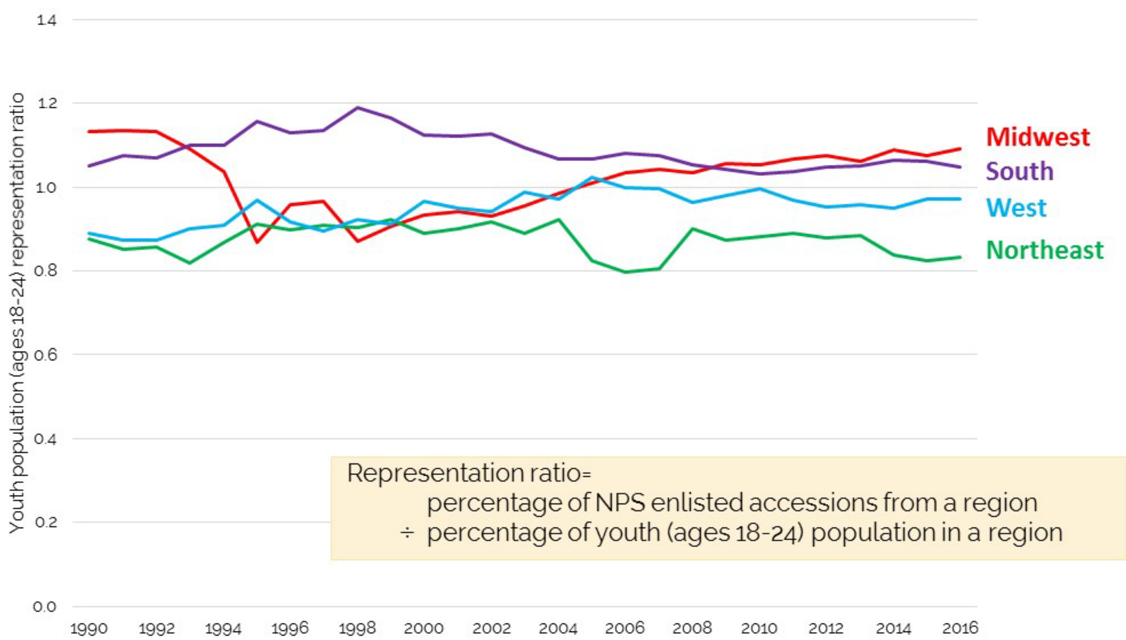


Figure 12. Regional Representation Ratios for Marine Corps Enlisted Accessions

The four Census regions can be subdivided into nine Census divisions, as shown in Figure 13. The representation ratios for the nine divisions move very little between 2015 and 2016, illustrating that the ratios (even at the finer geographic level) are stable from year to year and that most of redistribution takes place over the longer run. The representation ratio of the South Atlantic division increased over the 30 years (circled in green), while the ratio for the Pacific region decreased (circled in red). The redistribution between those two divisions is not one-for-one because some of the other divisions changed as well over those years. A drawback of this type of chart is that the weighted average of all the representation ratios is 1.0 every year: if one division goes up, another (or others) must go down to maintain the average. That is the case even if, for example, every one of the divisions improved its performance by some other metric such as accessions per youth population (measured in absolute terms, not relative to a national average standard).

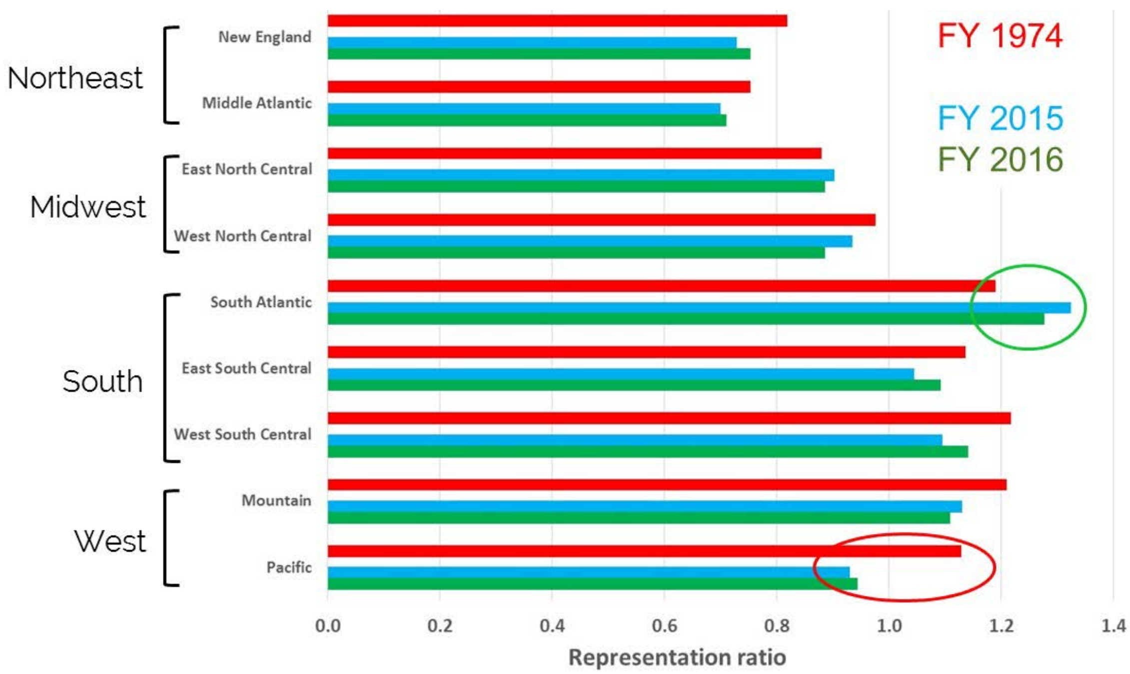


Figure 13. Representation Ratios for the Nine Census Divisions in Selected Years

3. Regional and State-Level Correlation and Regression Analysis

Our first line of attack is to attempt to explain the history of representation ratios at the same regional and state levels for which that history was presented in the preceding chapter. Youth unemployment is often portrayed as a major barometer of the recruiting climate.[14] However, the Bureau of Labor Statistics (BLS) does not have adequate sample sizes to estimate unemployment rates within age narrow brackets at finer geographical levels, such as by county. Hence, this chapter pursues the finest possible geographical resolution by examining the degree to which youth unemployment correlates with recruiting trends at the regional and state levels.

First we present a correlation analysis between representation ratios and youth unemployment rates for each of the four Census regions. To varying degrees across the four regions, when a region is experiencing higher youth unemployment (i.e., a weaker economy) than the national average, the region's representation ratio tends to go up. Conversely, a region with a higher-performing economy tends to supply fewer recruits to the military. In order to confirm those findings, we performed a regression analysis at the finer, state level. Once again, the strength of the statistical relationship between a state's representation ratio and its youth unemployment rate varies across the four Census regions.

Because the results of the correlation and regression analyses will be mixed, the following chapter commences the machine learning approach that will introduce over 100 additional predictors of recruiting success, all measurable at finer levels of geographical detail.

A. Correlation between Enlisted Accessions and Regional Unemployment Rates

Additional insight into the trends in representation ratios for the Census regions and divisions displayed in the preceding chapter may be gleaned with reference to youth unemployment rates. Although unemployment is surely not the only determinant of

[14] One of many recent examples is Nafeesa Syeed and Chloe Whiteaker, "Low U.S. Unemployment Is Making Army Recruiting Harder." *Bloomberg Politics*, March 13, 2018, https://www.bloomberg.com/news/articles/2018-03-13/trump-s-army-buildup-confronts-headwinds-of-tight-labor-market.

enlisted accessions, it is interesting to examine the extent to which that one factor explains the differences in representation ratios both across regions and over time.

The proportion of females among NPS Active Component enlisted accessions began at 7.9 percent in 1974 (the first year of the AVF), peaked at about 18 percent between 1998 and 2001, and then dipped to about 16 percent, before returning to 18 percent in 2015.[15] Because the military has disproportionately recruited males during the AVF years (and even more acutely during the draft years), the IDA team conducted a single analysis of all accessions (both genders combined) as a function of the unemployment rate for males ages 20–24. That is one of the age brackets in which the BLS reports employment statistics, but it is serendipitous because potential recruits might look ahead to their civilian employment prospects over the next few years of their lives. The male unemployment rate contains a bit more information about the state of the youth labor market than does the rate for females, but not much more. In 2014, for example, the labor force participation rate was 73.0 percent for males but 67.7 percent for females in the age range 20–24.[16]

One variable in the correlation analysis is the regional representation ratio. That variable moves in a relatively narrow band because it is theoretically bounded below by 0.0 (although, as a practical matter, it has been bounded between about 0.70 and 1.25 since 1990). Further, the representation ratio is centered at the neutral value of 1.0 that represents the weighted national average in any given year.

The IDA team constructed the other variable for the correlation analysis that is also, as a practical matter, bounded in a relatively narrow range; in this case, the variable is centered at a neutral value of 0.0. The raw, regional unemployment rate (averaged over the fiscal year) does not correlate well with the representation ratio because the former may swing wildly with movements in the national economy. However, the *deviation* of the regional unemployment rate from the national average rate, of course, averages to 0.0 across regions in any given year; it has varied in the range of about -2.00 to +2.75 (in percentage points) since 1990. The regional deviations are of either sign and correlate well with which regions over-perform (representation ratio > 1.0) or under-perform (representation ratio < 1.0) in a given year.

The correlation analysis for the Northeast region is shown in Figure 14. The correlation across the years equals 0.66. Although the national economy was strong during 1996, youth unemployment remained higher in the Northeast, 11.9 percent, versus the national average of 9.5 percent. Recruiting from the Northeast was strong in that year as well, and the unemployment deviation roughly tracks the local peak in the representation

---

15 DoD, *Population Representation in the Military Services: Fiscal Year 2015*, Table D-5.

16 "Civilian Labor Force Participation Rate by Age, Gender, Race, and Ethnicity," Bureau of Labor Statistics, October 2017, https://www.bls.gov/emp/tables/civilian-labor-force-participation-rate.htm.

ratio in that year. The two measures also track during the stronger regional economy (drop in the unemployment deviation) in 2009–2011 and the weaker regional economy in 2013–2014. In years when the regional labor market is stronger (e.g., the middle period of 2009–2011), this historically under-performing region sends even fewer recruits to the military.

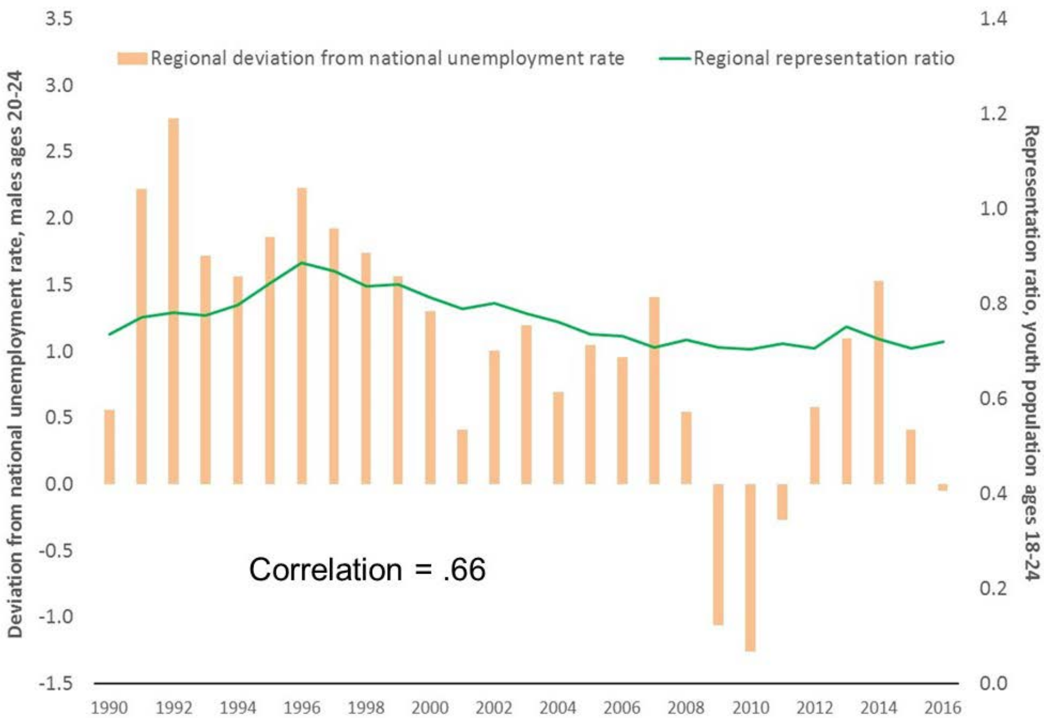


Figure 14. Regional-level Correlation Analysis: Northeast Region

The correlation analysis for the Midwest region is shown in Figure 15. The correlation across the years, at 0.37, is lower than for the Northeast region. The unemployment deviation in the Midwest is negative during the 1990s, indicating a stronger youth labor market than the national average, roughly correlated with a representation ratio that starts at 1.07 in 1991 but falls below 1.0 in 1993 and remains there for the remainder of the decade. The less prosperous regional economy between 2006 and 2009 is loosely related to a representation ratio that rises to between 0.95 and 0.98 during those years.

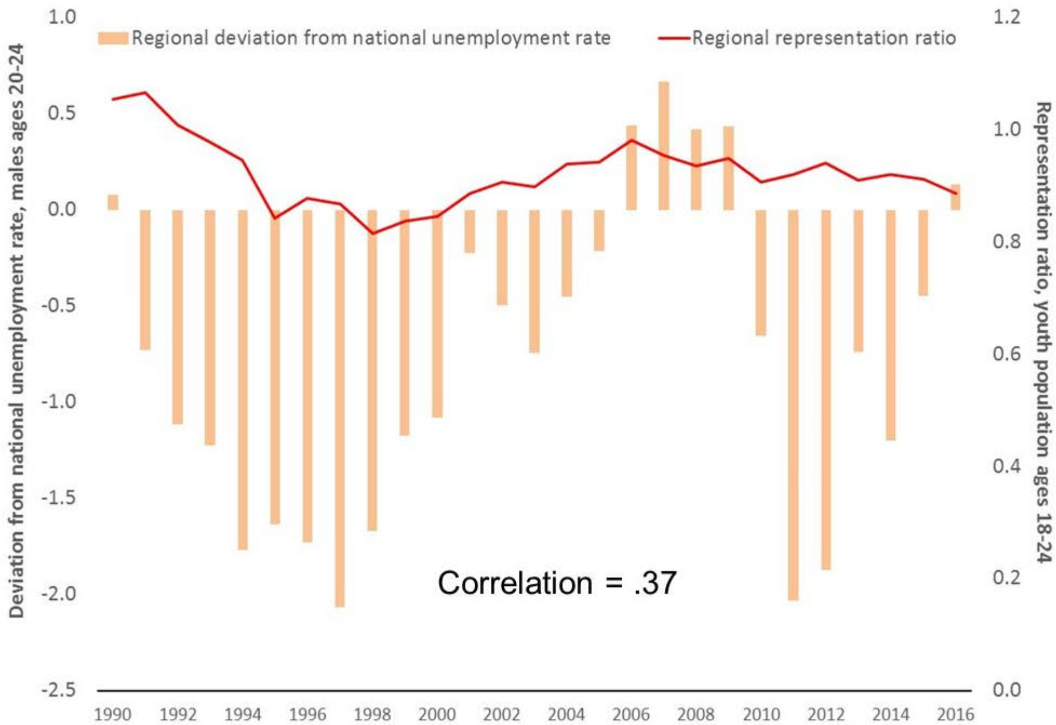


Figure 15. Regional-level Correlation Analysis: Midwest Region

The correlations for the West and the South are weaker still, at 0.29 and 0.10, respectively. The relatively flat representation ratio for the West (Figure 16) almost completely misses the strong regional economy (low unemployment) in 2001 and again in 2005–2007 (although the representation ratio drops slightly from 1.0 in 2005 to 0.95 in 2006). And the essentially flat representation ratio for the South (Figure 17) barely correlates at all with movements in the youth unemployment rate in that region.

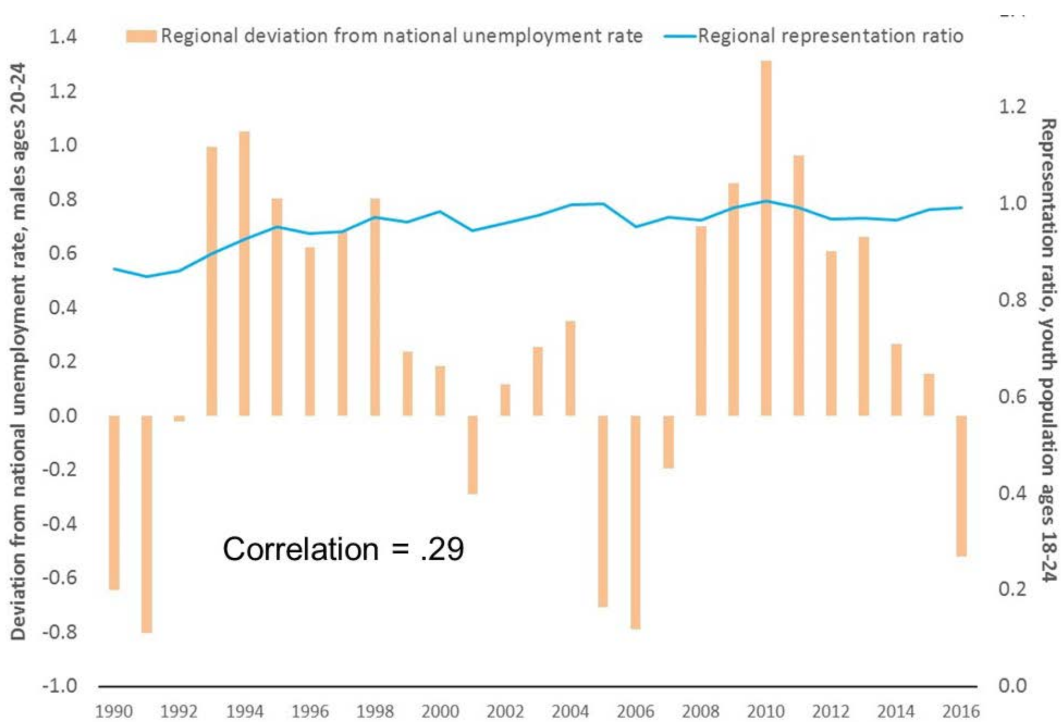


Figure 16. Regional-level Correlation Analysis: West Region

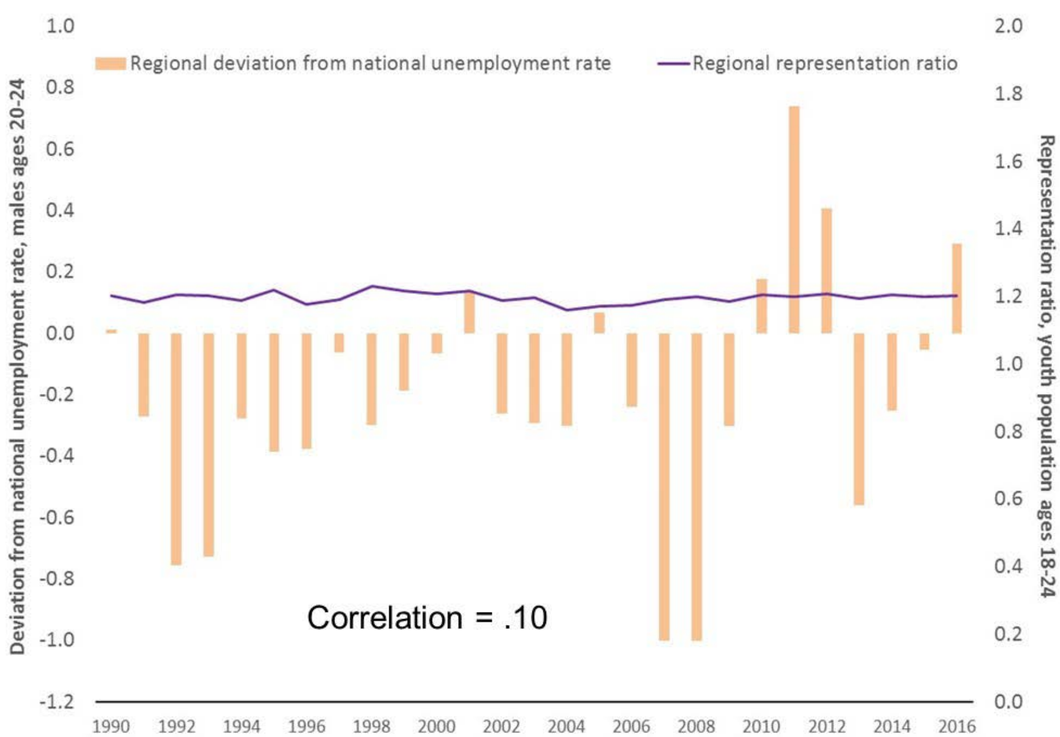


Figure 17. Regional-level Correlation Analysis: South Region

The outcome of the correlation analysis could have been predicted from the composite chart of regional representation ratios in Figure 7. When the representation ratio for a region like the South is essentially flat for more than 25 years, it will not correlate with movements in any other variable (be it unemployment or anything else) that is also measured at the regional level. However, by pushing the analysis down to finer levels such as states or counties, the indicators of recruiting success show more variation over time and begin to correlate with other factors that are also measured at the state or county level.

B. Regression Analysis of State-Level Enlisted Accessions and Unemployment Rates

This section applies regression analysis to further explore the relationship between accessions and unemployment at the state level. The analysis also incorporates information on the nine Census divisions that underlie the four Census regions, as introduced in Figure 4 above.

We begin with two regression models for the entire United States. The analysis covers the 51 "states" (including the District of Columbia) for the years 1990–2016, a total of 1,377 annual observations. The dependent (left-hand) variable is the representation ratio for one of the states in one of the years. The right-hand variable is the deviation of the state's unemployment rate among males aged 20–24 from the national average rate in the same year.

Table 1 shows the results of a simple regression with the unemployment rate deviation as the only explanatory variable. The coefficient of 0.0095 has the following interpretation: a 1.0-percentage point increase in the deviation of a state's unemployment rate from the national average is associated with a 0.95-percentage point increase in that state's representation ratio. That is, the estimated behavioral effect among potential recruits in the youth population is approximately one-for-one.

Table 1. Simple Regression Model for the Representation Ratio among 50 States and the District of Columbia

Variable	Coefficient	Standard Error	t-statistic	p-value (two-sided)	
Intercept	1.038 **	0.0076	136.52	0.000	
Unemployment deviation (national)	0.0095**	0.0024	3.93	0.000	
Sample size					1,377
R-squared					0.011

Note: ** indicates high statistical significance, p-value < 0.01 in a two-sided test.

Table 2 shows the results of a multiple regression when the single intercept is replaced by dummy variables for the four Census regions. This model explains the representation ratio much better, with the R-squared increasing from 0.011 to 0.944. The regional dummy variables have considerable explanatory power, and their ordering indicates that (having controlled for unemployment) the South and West regions are the most productive, followed by the Midwest and the Northeast. Conversely, having controlled for the regional effects, the coefficient on the unemployment rate deviation falls from 0.0095 to 0.0061. After controlling for region, the independent effect of the unemployment deviation is no longer one-for-one but about one-third smaller.

Table 2. Multiple Regression Model for the Representation Ratio among 50 States and the District of Columbia

Variable	Coefficient	Standard Error	t-statistic	p-value (two-sided)	
Unemployment deviation (national)	0.0061**	0.0023	2.65	0.008	
Midwest	0.936**	0.015	63.45	0.000	
Northeast	0.849**	0.016	52.23	0.000	
South	1.111**	0.012	93.36	0.000	
West	1.164**	0.014	85.64	0.000	
Sample size					1,377
R-squared					0.944

Note: ** indicates high statistical significance, p-value < 0.01 in a two-sided test.

Next we move on to four regional regression models, one for each Census region. As an example, nine states comprise the Northeast region, so the regression model for that region contains 9 states × 27 years = 243 annual observations (Table 3). The right-hand side includes a dummy variable to distinguish the two underlying Census divisions: New England and Middle Atlantic. Also, the right-hand side includes the deviation of the state's unemployment rate among males aged 20–24 from the *regional* average rate in the same year. For example, in 2010 the weighted average unemployment rate in the Northeast region was 16.6 percent; the most extreme rates in two of the constituent states were 19.5 percent in Pennsylvania and 11.1 percent in New Hampshire. The deviations for those two states would be coded as +2.9 percent (weaker state economy) and -5.5 percent (stronger state economy), respectively.

Table 3. Regression Model for the Representation Ratio among States in the Northeast Region

Variable	Coefficient	Standard Error	t-statistic	p-value (two-sided)	
Intercept	0.879**	0.022	39.22	0.000	
Unemployment deviation (regional)	-0.0081	0.0073	-1.10	0.273	
Middle Atlantic	-0.106**	0.037	-2.89	0.004	
Sample size					243
R-squared					0.051

Note: ** indicates high statistical significance, p-value < 0.01 in a two-sided test.

The regression models for the Midwest, West, and South regions are shown in Table 4 through Table 6.

Table 4. Regression Model for the Representation Ratio among States in the Midwest Region

Variable	Coefficient	Standard Error	t-statistic	p-value (two-sided)	
Intercept	0.960**	0.017	57.55	0.000	
Unemployment deviation (regional)	0.0150**	0.0045	3.36	0.001	
East North Central	-0.045	0.024	-1.88	0.061	
Sample size					324
R-squared					0.034

Note: ** indicates high statistical significance, p-value < 0.01 in a two-sided test.

Table 5. Regression Model for the Representation Ratio among States in the West Region

Variable	Coefficient	Standard Error	t-statistic	p-value (two-sided)	
Intercept	1.069**	0.024	43.79	0.000	
Unemployment deviation (regional)	0.0369**	0.0052	7.11	0.000	
Mountain	0.199**	0.033	5.98	0.000	
Sample size					351
R-squared					0.153

Note: ** indicates high statistical significance, p-value < 0.01 in a two-sided test.

Table 6. Regression Model for the Representation Ratio among States in the South Region

Variable	Coefficient	Standard error	t-statistic	p-value (two-sided)	
Intercept	1.070**	0.025	42.55	0.000	
Unemployment deviation (regional)	-0.0027	0.0038	-0.73	0.469	
South Atlantic	0.031	0.029	1.06	0.291	
West South Central	0.127**	0.035	3.61	0.000	
Sample size					459
R-squared					0.038

Note: ** indicates high statistical significance, p-value < 0.01 in a two-sided test.

Among the four regions, the unemployment coefficient is statistically significant for the Midwest (0.015) and the West (0.037). The interpretation of the magnitude for the Midwest is that a 1-percentage point increase in the deviation of a state's unemployment rate from the Midwest regional average is associated with a 1.5-percentage point increase in that state's representation ratio. In the West, a 1-percentage point deviation from that region's average unemployment rate is associated with a 3.7-percentage point increase in a state's representation ratio. Both effects measure the improved recruiting environment among states whose economies perform worse than the average in their respective region. For the Northeast and South regions, the annual and state-to-state variations in representation ratios are apparently driven by cultural or demographic factors, or possibly economic factors that are more subtle than simply the youth unemployment rate.

The magnitudes for the Midwest and the West may be compared to the results reported in Table 1 and Table 2 above, where each state's unemployment rate was compared to the *national* (not *regional*) average. The effects for the national-level regressions were smaller: 0.95-percentage points for the simple regression, and 0.61-percentage points for the (more properly specified) multiple regression that controls for regional effects. The larger effects at the regional level imply that in contemplating military service, youth in the Midwest and the West are more aware of their home state's economy relative to neighboring states in the same region than their home state's economy relative to the national situation. Among the choices facing youth in a state with a poor economy are joining the military or migrating to a neighboring state with a stronger economy. They seem less sensitive to the economies of more distant states, indicating a more regional focus and perhaps a disinclination to migrate to a different region of the country.

The regression analysis complements the correlation analysis from the previous section. Based on the previous analysis, the correlation between a *region's* representation ratio and its unemployment rate was strongest in the Northeast region, and successively weaker in the Midwest, West, and South (see Table 7, where the rows are displayed in that sequence). However, the ordering of the effects from the regression analysis is quite different. For *individual states,* the two statistically significant associations were for states in the West (strongest) and the Midwest (less strong but still statistically significant). For the Northeast, variation over time in the region's unemployment rate is correlated with the representation ratio that the entire region achieves (Figure 14, correlation = 0.66 and highly statistically significant), but cross-sectional variation in the unemployment rate among the constituent states is not statistically related to those states' representation ratios. For the South—where the representation ratio has been flat for over 25 years—neither the correlation nor the regression analysis shows a significant a statistical relationship with the unemployment rate.

Table 7. Summary of Correlation and Regression Analyses of Representation Ratios, by Region

Region	Correlation between regional unemployment deviation and representation ratio	Effect on a state's representation ratio of a 1-percentage point deviation above the regional average youth unemployment rate
Northeast	0.66 **	-0.8 percentage points †
Midwest	0.37 *	1.5 percentage points **
West	0.29	3.7 percentage points **
South	0.10	-0.3 percentage points †

Note: * indicates moderate statistical significance, p-value < 0.1; ** indicates high statistical significance, p-value < 0.01; † indicates unexpected sign but not statistically significant.

4. Machine Learning Approach

We applied a machine learning technique known as *random forests* to determine the most influential features of a community for predicting the number of non-prior-service enlisted accessions per youth population. This chapter provides a brief tutorial on the random forests technique. Chapter 6 describes the data sources for application of the technique, and Chapter 7 discusses the findings of that analysis.

The first step is to conceptualize a regression tree. Although we ultimately estimated regression trees at a finer level of geography, we began our exploration at the state level. Figure 18 is one tree for predicting the number of accessions per 10,000 youth population based on data by state (the 50 states plus the District of Columbia) over the years 2007–2015. The first branch in the tree is whether the percentage of college-enrolled among the population ages 17–29 is less than the value 0.389558 (39 percent). If "no" (i.e., the percentage exceeds that threshold), we move to the right and the model predicts an accession rate of 29.28 per 10,000 (that is, 0.29 percent of the youth population in a state with a high percentage of college-enrolled). If the percentage of college-enrolled is less that the threshold, we move instead to the left and the next branch is whether the percentage of veterans with more than a college degree is less than the value 0.606875 (61 percent). In a state where the percentage of "college+" is less than that threshold, the next branch involves the percentage of disabled veterans in the total population. But in a state where the percentage of college+ is greater than the threshold, the next two branches involve the ratio of veteran to non-veteran average income, and the percentage of people who commute to work more than 90 minutes. The numbers at the end of each node are the estimated accession rates down each path in the tree.

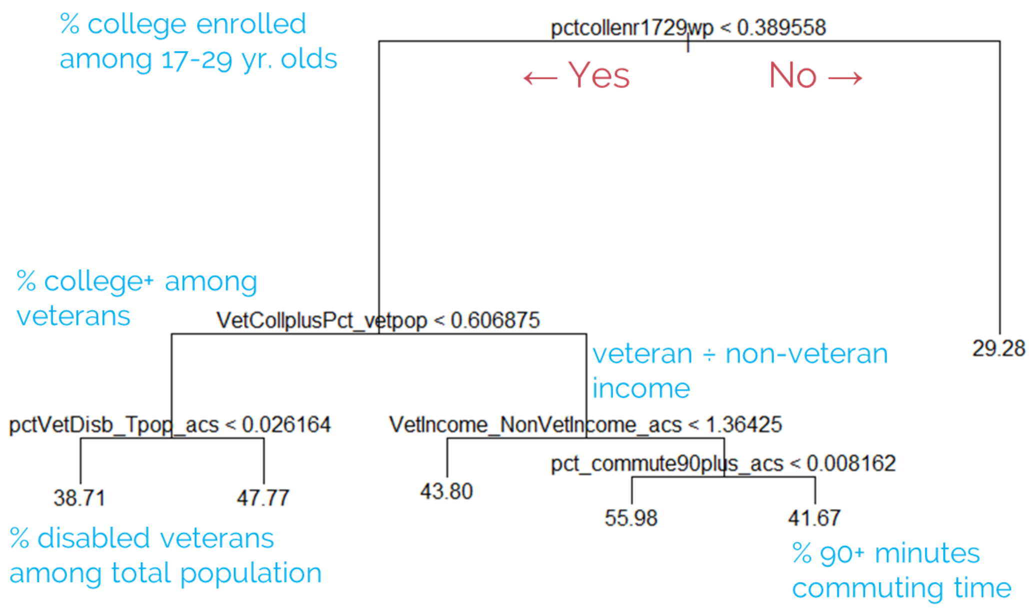


Figure 18. Example of a Regression Tree for Predicting Accessions per Youth Population at the State Level

The random forest method proceeds by sampling with replacement ("bootstrapping") different rows from the full dataset (in this example, a row is a combination of state and year) and estimating a regression tree; drawing a new bootstrap sample and repeating that process many (often hundreds of) times; and finally averaging the resulting trees.[17] While a single tree may yield imprecise estimates of the split variables and their respective thresholds, the process of averaging many trees greatly reduces the variance. An analogy may help illustrate the reduction in variance that results from averaging. If a fair coin is flipped 10 times, the expected number of heads is 5 but there is a 17 percent probability that 3 or fewer heads are observed. Now repeat that experiment 10 times and average the results. The expected number of heads is 50 (out of a total of 100 coin flips), but there is only a minuscule probability that 30 or fewer heads are observed and a 17 percent probability that 45 or fewer heads are observed. Put differently, whereas a "low" event (a small number of heads that occurs only 17 percent of the time) was 3 out of 10 heads, when averaged over 10 replications, the distribution of outcomes tightens such that the low event with the same 17 percent probability becomes 4.5 out of 10 heads.

The random forest method takes one additional step to reduce the correlations among the regression trees ("decorrelate" them) before taking the average. If two random variables are positively correlated, when one of them swings high, the second one also tends to do so, and when the first one swings low, the second one again tends to follow suit. As a result,

---

[17] Gareth James et al., *An Introduction to Statistical Learning, with Applications in R* (Heidelberg, Germany: Springer, 2015), Chapter 8.2.

both the highs and the lows are accentuated and the variance of the sum (or average) is increased. By decorrelating the two random variables, we allow the possibility that one will swing low while the other swings high, tending to offset each other and reduce the variance of the average. The random forest method reduces the correlation among trees by considering only a subset of predictor variables (the columns, rather than the rows of the dataset) at each split. That step eliminates the possibility that a single, dominant predictor would appear at the top level of virtually every tree that is being averaged, which, if allowed, would induce a positive correlation among the trees and increase the variance of the estimates.

Two of the parameters that must be set when using the random forest method are the maximum number of variables that are considered for each split (only one being selected), and the number of trees that are generated and averaged. We used the default specification that the maximum number of variables equals about one-third of the total number in the dataset. Specifically, with 106 potential predictors, we set the maximum at 33. Second, based on the standards that have developed in the literature, we generated and averaged 100 trees; a larger number of trees than that generally provides little additional accuracy.[18]

---

[18] Ibid.

5. Geographical Detail at Finer Than the State Level

Some of the data for this paper—such as the number of high school students taking the ASVAB, or the numbers of Junior Reserve Officers' Training Corps (JROTC) schools and cadets—are available at the ZIP code level. Other data—such as the number of degree-granting institutions—are available directly at the county level. The ASVAB and JROTC data can be "rolled up" or aggregated to the county level.

Still other data used in this paper—such as data from the American Community Survey (ACS) discussed in more detail in the following chapter—are available at a Census-defined geographical level called the Public Use Microdata Area (PUMA). Counties and PUMAs do not generally correspond one-to-one: a large county may contain several PUMAs, but a PUMA may also be constructed by combining several small (often rural) counties. In the first part of this chapter we develop a concept called a *pseudo-county*, which serves as the smallest geographic region that does not divide up a county or a PUMA. The second part of the chapter displays some trends in accession rates at the pseudo-county level.

A. Development of Pseudo-Counties

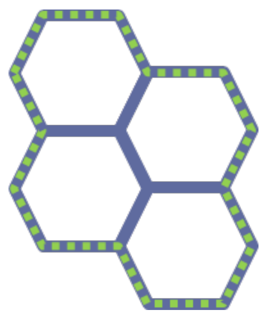
Because there is so much variation in demographic, economic, and cultural factors within any given state, we wished to analyze the factors that affect the geographic distribution of accessions at a smaller geographical unit than the state.

Some data sources are as fine as the county or ZIP code level—or even street address, in the case of the secondary school data from the Integrated Postsecondary Education Data System (IPEDS). However, other data sources such as the ACS aggregate data to the PUMA level to preserve the privacy of surveyed households and individuals. PUMAs are geographically contiguous regions that encompass the entire area of the United States, Puerto Rico, Guam, and the US Virgin Islands. The Census Bureau draws PUMAs so that they contain at least 100,000 people, and follow Census tracts and (where feasible) counties (or other jurisdictions such as independent cities).[19] Because the population changes over time, the Census Bureau redraws PUMA borders a few years after each decennial census.

[19] For example, Falls Church City is an independent city in the Commonwealth of Virginia. It is adjacent to both Arlington and Fairfax counties, but is not a part of either of them.

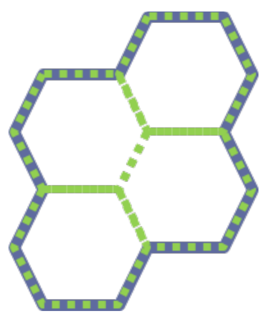
And because the borders of PUMAs are determined by population, they can be either larger or smaller than counties.

To merge these data sources while still preserving geographic variation below the state level, we defined a pseudo-county, which represents the smallest region that aligns counties and PUMAs. The Census Bureau may construct a large PUMA from several small (often rural) counties or independent cities. In those cases, the PUMA borders define the pseudo-county. Conversely, a large county with population equal to a multiple of 100,000 may contain several PUMAs. In that situation, the counties define the pseudo-county borders. Figure 19 and Figure 20 shows these respective cases, where the dashed green line represents the PUMA/pseudo-county border and the solid blue line represents the county borders.



Note: Dashed green lines represent pseudo-county border. Solid blue lines represent county borders.

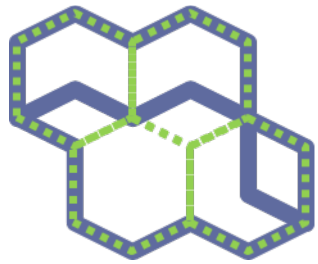
Figure 19. Pseudo-County Defined by PUMA Borders



Note: Dashed green lines represent pseudo-county border. Solid green lines represent PUMA borders.

Figure 20. Pseudo-County Defined by County Borders

In a few cases, PUMA borders and county borders do not line up neatly. In these cases, we define the pseudo-county as the smallest geographic region that does not divide up a county or a PUMA (the “least common denominator”). Figure 21 demonstrates an example of a pseudo-county that contains multiple PUMAs and multiple counties.



Note: Dashed green lines represent pseudo-county border. Heavy blue lines represent county borders, and solid green lines represent PUMA borders.

Figure 21. Pseudo-County Defined by Smallest Region Containing Entire PUMAs and Counties

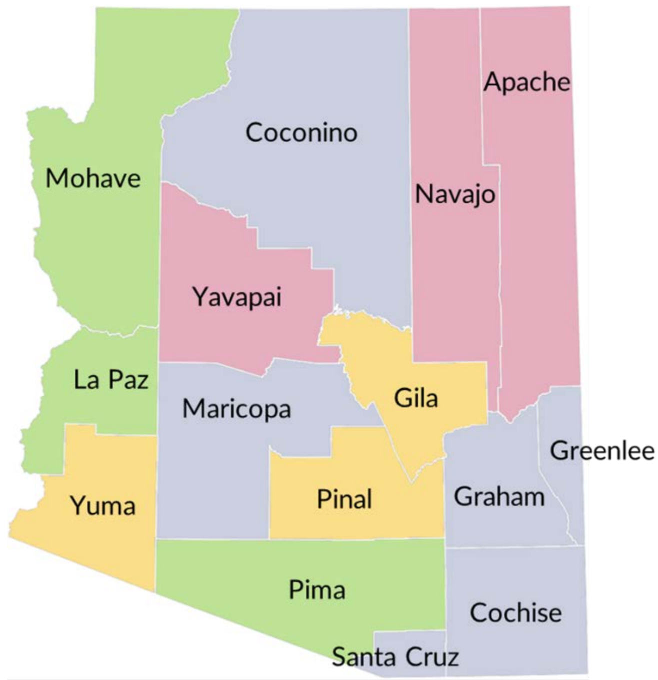
With these three examples, we can now map each county and each PUMA to a contiguous pseudo-county. For ZIP code-level data, we used ZIP-to-county crosswalks provided by the United States Department of Housing and Urban Development (HUD).[20] After the ZIP codes were matched to counties, we could then match them to pseudo-counties based on the county-to-pseudo-county crosswalks we created for the county-level data.

As a general rule, pseudo-counties will tend to aggregate counties in rural regions up to the PUMA population of around 100,000, whereas large-population counties will tend not to be aggregated with other counties. Figure 22 demonstrates how counties are aggregated in Arizona in 2007.[21]

---

[20] The HUD crosswalks are not available prior to 2010. We use the 2010 crosswalk to merge ZIP code data from earlier years.

[21] Because PUMAs are redrawn over time as the population changes, we defined separate pseudo-counties for each year of our data.



Note: Contiguous counties that are colored the same represent a single pseudo-county. Counties that are not contiguous but have the same color are not part of the same pseudo-county. For example, Navajo County and Apache County together form one pseudo-county. However, notwithstanding the reuse of the same color, Yavapai is its own separate pseudo-county.

Figure 22. Example of Pseudo-Counties in Arizona in 2007

Small-population counties like Santa Cruz, Cochise, Graham, and Greenlee are aggregated into a single PUMA in the southeast corner of the state, with a total 2006 population of 212,000. In contrast, Maricopa County (which includes Phoenix) had a 2006 population of 3,768,000 and stands alone as its own pseudo-county. The very largest counties in the United States also stand alone and form enormous but indivisible pseudo-counties, the most extreme example being Los Angeles County with a population of about 10 million.

In 2016, the 3,142 counties and 2,351 PUMAs combined into 982 pseudo-counties, an average of about 19 per state.

B. Variation across Pseudo-Counties

The pseudo-county-level heat map for male recruits in FY 2010 shown in Figure 23 provides finer geographic detail that the state-level map shown above as Figure 2. Although the grid in Figure 23 indicates individual counties, the data granularity is actually at the pseudo-county level. So, for example, if a particular pseudo-county is composed of three adjacent counties, the data value (which translates into the color on the map) will be exactly the same for all three constituent counties.

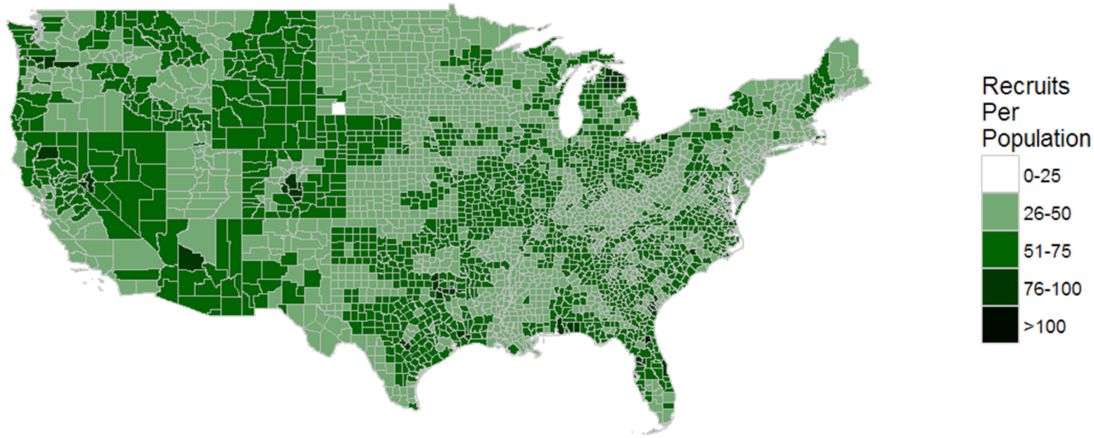


Figure 23. Non-Prior-Service Male Enlisted Accessions per 10,000 Youths Aged 18–24, by Pseudo-County, FY 2010

Figure 2 identifies Florida, Alabama, Georgia, and South Carolina as high-yield states in FY 2010. However, we learn from Figure 23 that southern Florida is less productive than other parts of the state, except for Broward County (which includes Ft. Lauderdale). Similarly, Texas is a high-yield state, but not extraordinarily so along the Mexican border except for the geographical extremes of El Paso County at the west and Cameron County (which includes Brownsville) at the southeast. In addition to these insights, a pseudo-county-level analysis provides more variation and thus more opportunities to identify meaningful statistical associations.

Figure 24 advances five years to FY 2015. The heat map is generally lighter than in FY 2010 because total male accessions fell from 133,000 to 119,000. Notwithstanding, some patterns are evident, such as the lightening (lower yield) in parts of Texas and the "filling-in" (higher yield) in Georgia and Mississippi.

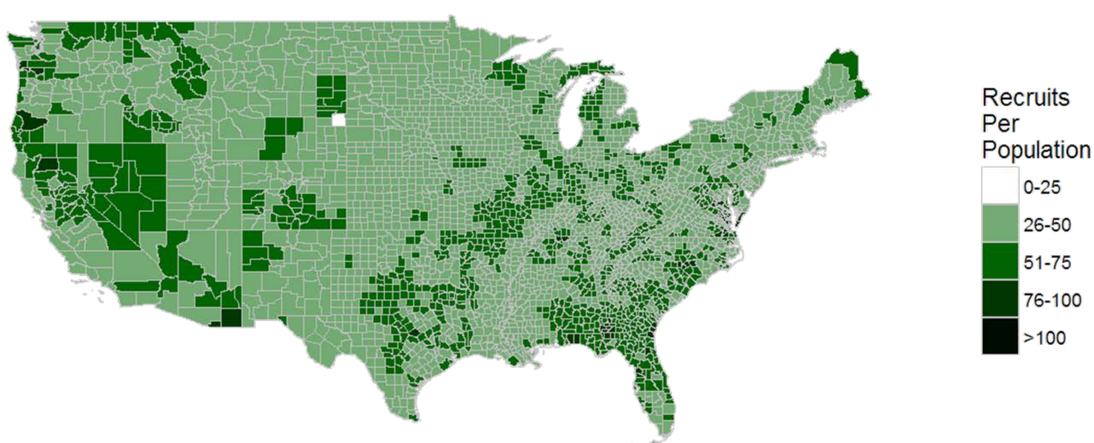


Figure 24. Non-Prior-Service Male Enlisted Accessions per 10,000 Youths Aged 18–24, by Pseudo-County, FY 2015

Figure 25 and Figure 26 are heat maps for female recruits in the same two years, 2010 and 2015. Total female accessions fell only slightly from 27,000 to 26,000, so the comparison between these two figures is almost purely one of redistribution. Again, some patterns are evident, such as the darkening of several counties (including the border counties of Santa Cruz and Cochise, among others) in Arizona.

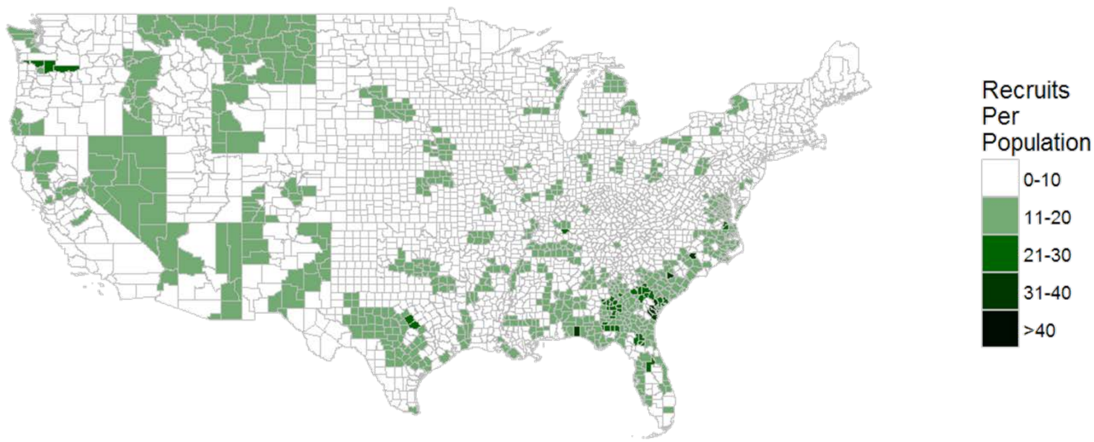


Figure 25. Non-Prior-Service Female Enlisted Accessions per 10,000 Youths Aged 18–24, by Pseudo-County, FY 2010

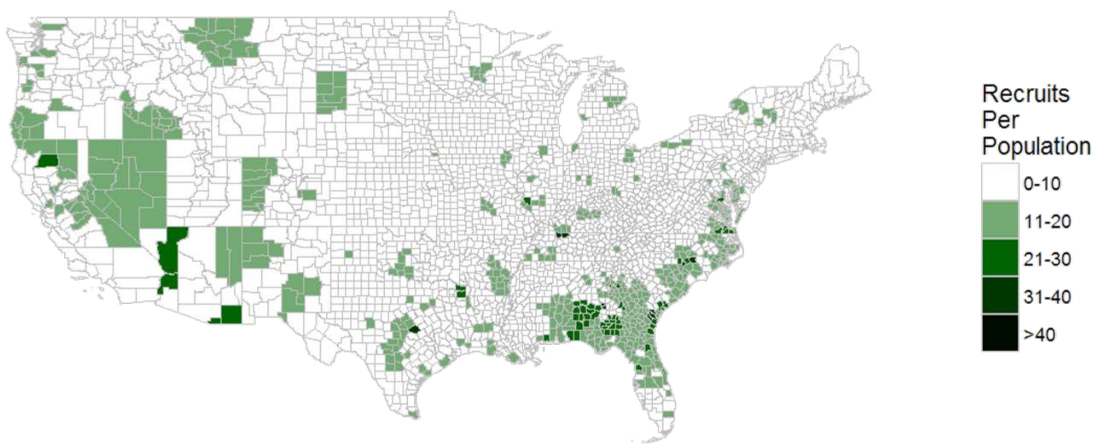


Figure 26. Non-Prior-Service Female Enlisted Accessions per 10,000 Youths Aged 18–24, by Pseudo-County, FY 2015

6. Data Sources

This chapter describes the sources of the data to which the machine learning analysis was applied. The first section details the data on accessions that are the subject of the analysis (i.e., the left-hand variable in the analysis). The second section provides summary descriptions of the data sources from which all of the predictors of accessions (the right-hand variables) were calculated. A complete list of variables, their method of calculation, and their sources can be found in Appendix B.

A. Accessions Data

DMDC supplied the IDA researchers with United States Military Entrance Processing Command (USMEPCOM) data on all Active and Reserve military and Coast Guard enlisted accessions from FY 2006–2015. We restricted the sample to all NPS candidates who shipped to boot camp for one of the four Active Duty military Services (Army, Navy, Marine Corps, or Air Force). To remove duplicates, we eliminated recruits who were over a Service's typical recruiting age,[22] or were above a standard entry-level rank.[23] We then aggregated the total number of accessions across all four military Services for each home-of-record ZIP code, and further aggregated accessions to pseudo-counties as described in the previous section. We performed this aggregation separately for six samples: all recruits, all male recruits, all female recruits, all high-quality[24] recruits, all high-quality male recruits, and all high-quality female recruits. Table 8 shows total accessions every year for each of the six samples.

---

[22] We dropped individuals who met the following criteria:

- Individual accesses into the Army and is above age 35 before 2005
- Individual accesses into the Army and is above age 40 after and including 2005
- Individual accesses into the Navy and is above age 34
- Individual accesses into the Marine Corps and is above age 35
- Individual accesses into the Air Force and is above age 27 before 2014
- Individual accesses into the Air Force and is above age 39 after and including 2014

[23] We removed individuals in paygrades above E-4.

[24] High-quality recruits are defined as recruits who are either high school seniors or graduates and who score as Category 1-3A on the AFQT.

Table 8. NPS Accessions per Fiscal Year

	All Accessions			High Quality Accessions		
Fiscal Year	Female	Male	Total	Female	Male	Total
2006	28,080	139,699	167,779	14,369	83,442	97,811
2007	25,513	133,600	159,113	13,137	78,799	91,936
2008	25,936	139,140	165,076	13,556	84,554	98,110
2009	26,225	133,466	159,691	14,384	84,278	98,662
2010	26,199	132,480	158,679	14,570	83,558	98,128
2011	25,394	126,865	152,259	14,655	83,305	97,960
2012	24,956	129,404	154,360	14,816	86,645	101,461
2013	27,747	137,031	164,778	15,565	88,163	103,728
2014	24,318	114,521	138,839	14,048	74,729	88,777
2015	26,183	119,129	145,312	14,883	77,650	92,533

Because population density varies significantly across the United States, we divided the accession counts in each pseudo-county by the total population aged 17–24 in the pseudo-county (from the ACS data described in the next section).

B. Other Data

Table 9 indicates some of variables that were used to predict accessions per youth population. More complete descriptions follow the table. Still other variables, for example measuring the recruiting effort in a pseudo-county, would have been interesting to include in the analysis. The IDA team was able to obtain contemporaneous information from DMDC on recruiting effort at the recruiting station level, which is two levels finer than the recruiting battalions at which much of the traditional analysis of enlisted accessions has been conducted (the intermediate level being, in the Army's terminology, the recruiting company). However, DMDC could provide only recent, not historical, time-series data on recruiting effort. The historical data, if available, might have enabled us to sort out prominent variations within states such as the high yields in Broward County, Florida and in El Paso County and Cameron County, Texas (as noted in Chapter 5).

Table 9. Partial List of Predictor Variables and their Sources

Data Source	Examples of Predictor Variables	Level of Geography
American Community Survey (ACS)	College enrollment and educational achievement	PUMA
	Commuting time	PUMA
	Number of vehicles in the household	PUMA
	Percentage of veterans in the population	PUMA
	79 other variables	PUMA
ASVAB Career Exploration Program (CEP)	Number of high school students available to take the ASVAB	ZIP code
Integrated Postsecondary Education Data System (IPEDS)	Number of degree-granting institutions	County
Junior Reserve Officers' Training Corps (JROTC)	Number of JROTC schools and cadets	ZIP code

1. American Community Survey (ACS)

The ACS is an annual questionnaire administered by the Census Bureau.[25] The ACS surveys 3.5 million people throughout the year and covers every state, the District of Columbia, and Puerto Rico. The ACS collects demographic, education, employment, and financial data for each household and for each resident in the household. ACS data come as one-, three-, and five-year estimates. We selected one-year estimates to match the frequency of the accessions data.

We selected ACS variables to capture community-level information that may factor into the decision to join the military, and obtained all data at the PUMA level.[26] Demographic variables include age, veteran status, and language preferences. Education variables include estimates of high school, college, and post-baccalaureate graduates. Employment variables include estimates of full- and part- time workers, both at the person-level and within a household. Financial data include average household-level income and mortgage payment in dollar values. These data cover calendar years 2006 through 2015.[27]

---

[25] "American Community Survey (ACS), Top Questions About the Survey," United States Census Bureau, https://www.census.gov/programs-surveys/acs/about/top-questions-about-the-survey.html.

[26] ACS data are also aggregated at the county level, but one-year estimates only cover counties with populations of 65,000 or greater. The PUMA is the smallest level of geographic aggregation that covers the entire US population in the one-year estimates.

[27] Before 2006, the Census did not survey individuals living in group quarters. Group quarters are group living arrangements, such as nursing homes, correctional facilities, and student housing.

To aggregate the PUMA-level ACS data to the pseudo-county level, we summed count variables (such as total population) and averaged non-count variables (such as average household income or veteran percentage). We weighted the averages of veteran-specific demographics by the total population of veterans older than 18. Similarly, we weighted the averages of non-veteran demographic variables by the over-18 non-veteran population. We weighted the averages of dollar-denominated household variables (such as average mortgage payments) by total household counts.

After merging these data from the ACS, we then normalized all count variables to represent rates per population. For individual-level count variables, we divided by the total population in each pseudo-county. For household-level count variables, we divided by the total number of households in each pseudo-county. We also created veteran-nonveteran indices for a select number of variables, including median income, percent college graduates, percent white, and percent living in poverty.

2. Centers for Disease Control and Prevention (CDC)

To join the military, prospective recruits must meet minimum standards for physical fitness, including body mass index (BMI), an indicator of obesity and underweight calculated by dividing one's weight in kilograms by the square of one's height in meters. As a result, higher obesity rates in some regions effectively reduce the eligible recruiting pool in these areas. We collected measures of obese population percentages from the CDC's county-level files. To get the obesity percentages for pseudo-counties, we averaged the county-level obesity rates, weighted by population. Obesity rates vary significantly across the country, from a minimum of 10.7 percent in 2012 Eagle County, Colorado to a maximum of 48.5 percent in 2011 Greene County, Alabama, with a weighted average obesity level of 27.2 percent across all counties and years.

3. Integrated Postsecondary Education Data System (IPEDS)

The IPEDS is an annual survey of all Title IV-eligible colleges, universities, and technical/vocational institutions in the US conducted by The National Center for Education Statistics (NCES).[28] The Higher Education Act of 1965 requires these postsecondary schools to submit information about institution characteristics such as enrollment, graduation, attendance costs, and student financial aid. While the ACS includes college attendance rates, the IPEDS data allow us to segregate attendance by type of postsecondary school (two-year institution, four-year institution, etc.). In addition, for some years, we can also observe the average net price (tuition, room and board, fees, less government financial

[28] Title IV-eligible institutions are those institutions that can participate in federal student financial aid programs.

aid) paid by enrolled students at each institution.[29] Unfortunately, unlike the ACS, all data are reported at the institution level, so while we can observe the geographic location of every postsecondary school in the database, we do not observe where their students are from. As a result, the presence of local technical schools and some smaller four-year schools may say more about the military recruiting propensity of a region than does the presence of larger schools that enroll a more geographically diverse set of students.

4. Quarterly Census of Employment and Wages (QCEW)

To supplement the economic data collected from the ACS, we also collected data on the percentage of workers employed by the government and the percentage employed in the private sector at the county level for the years 2006–2015, compiled by the BLS using the Quarterly Census of Employment and Wages (QCEW). We then calculated the weighted averages of these variables at the pseudo-county level. The two variables move independently and do not mechanically sum to 100 percent. QCEW only measures employment at the establishments that report to the federal Unemployment Insurance (UI) programs. Therefore, the self-employed are not counted, although some domestic and agricultural workers are counted when reported for UI purposes.

5. Armed Services Vocational Aptitude Battery Career Exploration Program (ASVAB CEP)

Military entrance processing centers administer the ASVAB to all prospective recruits to determine their eligibility for enlistment. The ASVAB's 10 subject areas test general knowledge such as arithmetic reasoning, word knowledge, and general science, as well as more job-specific skills such as electronics information, automotive and shop information, and mechanical comprehension. The AFQT score is the sum of the scores on three of the ASVAB subtests, and is used to determine eligibility for enlistment.[30] The Services use the scores on the remaining subtests to determine recruits' eligibility for specific career fields in the military.

In addition to the tests administered to military applicants, DoD offers the ASVAB as a free "career planning resource" to grade 10–12 students via the ASVAB Career Exploration Program (CEP). High school students who take the ASVAB through the CEP are encouraged to use their scores to consider a variety of career fields (not just the military) and are under no obligation to apply to the military. US high schools are not obligated to offer the ASVAB and, for the majority of high schools that do offer it, students are not obligated to take it. There are numerous potential reasons why particular schools may

---

[29] We convert all dollar values to real dollars using the Consumer Price Index.

[30] Specifically, $AFQT\ score = Arithmetic\ Reasoning\ score + Mathematical\ Knowledge\ score + (2 \times Verbal\ Expression\ score)$.

choose to administer the ASVAB and why students may choose to take it. Even so, overall geographic patterns in the frequency of ASVAB administration may reflect both cultural attitudes towards the military as well as young adults' exposure and access to the military as a potential career.

USMEPCOM provided us with a list of all schools (and their ZIP codes) that administered the ASVAB through the CEP from 2006 to 2015, as well as counts of all grade 10–12 students who were eligible to take the exam and counts of grade 10–12 students who actually took the exam for each school. From these data, we construct pseudo-county-level measures of the number of high school students with access to the ASVAB CEP, the number of high school students taking the ASVAB through the CEP, and the number of institutions administering the ASVAB CEP, normalized by the population aged 17–24 in each pseudo-county.

We show the current state of play by charting, for 2016, the number of high school students available to take the ASVAB per population ages 17–24 (Figure 27). High concentrations are observed in southern Alabama, South Dakota, Idaho, and portions of the other western states. Although the prevalence of ASVAB testing is potentially influenced by DoD policy, a high prevalence may simply signal counties and states that fundamentally have a more positive sentiment toward military service, and that are more willing to host a test that could boost the proclivity of some high school students to consider the military as a career. This analysis will demonstrate an association between ASVAB availability and recruiting success, but it does not sort out the direction of causality—which is the "chicken" and which is the "egg."

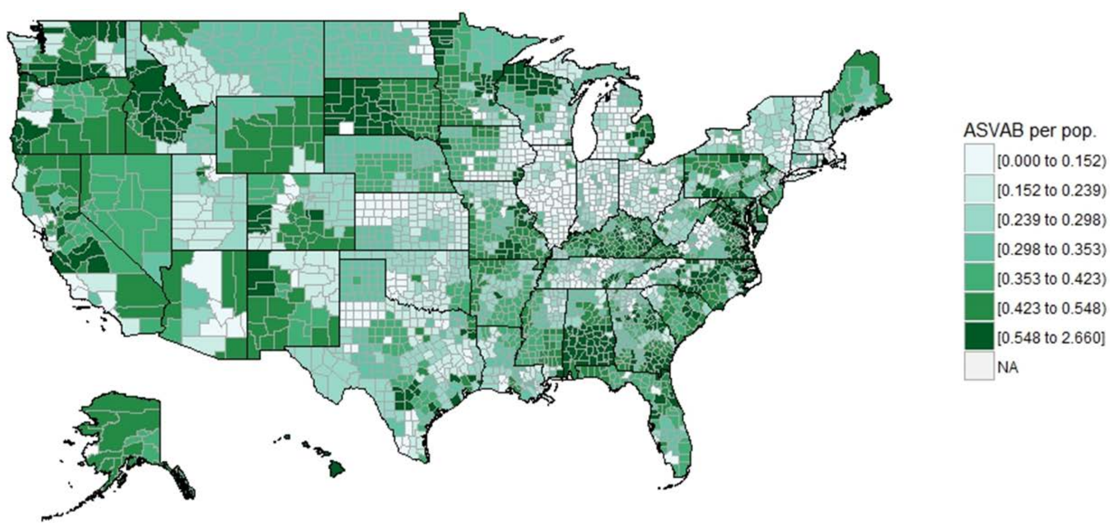


Figure 27. Number of High School Students Available to Take the ASVAB per Population Ages 17–24, by Pseudo-County, 2016

6. Junior Reserve Officers' Training Corps (JROTC) Programs

The JROTC is another high school program sponsored by DoD "to instill in students in [US] secondary educational institutions the values of citizenship, service to the United States, and personal responsibility and a sense of accomplishment."[31] The JROTC is not formally a recruiting program, but it does provide high school students with exposure to the possibility of a military career. High schools can choose whether to offer a JROTC program, and students can choose whether to participate. As with the ASVAB CEP, the distribution of schools offering JROTC programs and the participation rate of students may reveal something both about cultural attitudes and student exposure to the military.

We obtained a 2017 snapshot of all schools in the United States that currently offer a JROTC program. From these data, we calculated the total number of JROTC cadets and the total number of schools offering JROTC programs, normalized by the age 17–24 population in each pseudo-county. We conjectured that the JROTC variables would be important predictors of accessions, and we neither wanted to drop that pair of variables nor drop the early years of data from the larger dataset because the JROTC variables did not extend further back. Instead, we applied the 2017 values of the JROTC variables for each ZIP code to all of the years in our sample, 2006–2015. The definitions of the pseudo-counties changed somewhat over time, but for each year we were able to aggregate the JROTC data from ZIP codes to that year's collection of pseudo-counties.

Because, like the ASVAB, the prevalence of JROTC programs is potentially influenced by DoD policy, we display heat maps for JROTC cadets and JROTC schools in 2017 to show the current state of play (Figure 28 and Figure 29). Depending on the particular measure examined, high concentrations are observed throughout the southeast, in Arizona, Nevada, portions of Texas, and especially in New Mexico. The JROTC program is jointly funded by DoD (about $375 million per year) and the local school districts (about $225 million). However, under current congressional caps, DoD is limited to between 3,000 and 3,700 JROTC units.[32] There are some high schools on the waiting list that cannot be offered JROTC programs because of either funding limitations or bumping up against the congressional cap. While associated with recruiting success, DoD's ability to further propagate JROTC units hinges on relief from those two limitations.

---

[31] 10 U.S.C. § 2031 - Junior Reserve Officers' Training Corps.

[32] Public Law 112-239, National Defense Authorization Act for Fiscal Year 2013, Section 553, Modification of requirements on plan to increase the number of units of the Junior Reserve Officers' Training Corps.

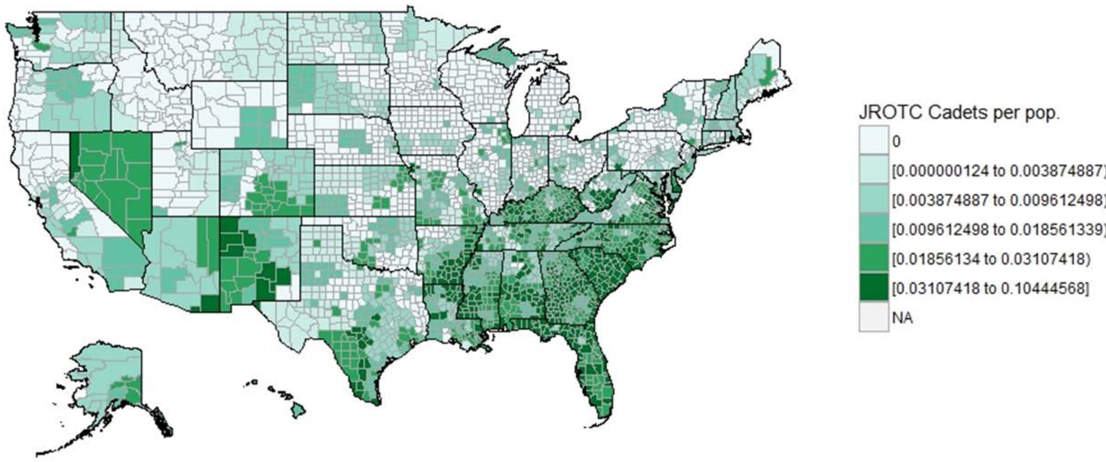


Figure 28. Number of JROTC Cadets per 10,000 Population Ages 17–24, by Pseudo-County, 2017

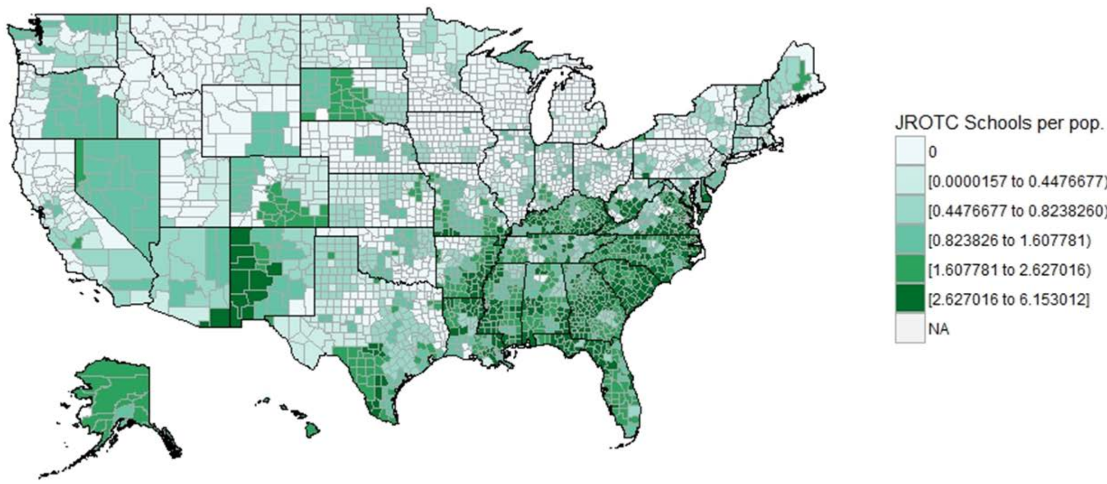


Figure 29. Number of JROTC Schools per 10,000 Population Ages 17–24, by Pseudo-County, 2017

7. Findings of the Machine Learning Analyses

This chapter contains the findings of the machine learning analyses. The analyses were conducted at the pseudo-county level for the years 2006 through 2015. Three response variables were modeled. The first section of the chapter presents the findings for all accessions—both male and female, without regard to quality level. The second section presents the findings for the high-interest subset of high-quality male accessions. "High-quality" is defined as high school graduates and seniors who score above the median (categories 1-3A) on the AFQT, a subset of the ASVAB. The third section presents the corresponding results for high-quality female accessions.

Appendix B enumerates the common set of 106 variables that were fed into the three machine learning models for potential inclusion. For each of the three analyses (all recruits, high-quality males, and high-quality females), the most important drivers of accessions were identified as those with the largest reduction in the mean-squared error when introduced one at a time as splits in the machine learning model.[33] Also, it should be remembered that machine learning only identifies variables that are statistically associated (correlated) with accession rates in the three samples. This approach does not reveal causation, so it would be an over-interpretation to assert that a policy change to one of the important variables would "cause" the accession rate to change in line with an estimated effect.

A. All Accessions

The first variable of interest is the percentage of veterans among the population age 18 and older in a pseudo-county. The left-hand panel of Figure 30 shows the strong effect of that variable on accessions, associated with an increase in the expected number of accessions in a range from about 5.0 to 7.0 per 1,000 youth population. However, the right-hand panel indicates that the yearly average values of the veteran percentage range between about 7.5 percent and 10.5 percent; the range across individual pseudo-counties is much wider than that of the annual averages. Within a reasonable range, and holding fixed other factors that also influence accessions, the percentage of veterans in the community can drive accessions in a range of about 5.0 to 6.0 per 1,000 youth population. However, with

---

[33] Gareth James et al., *An Introduction to Statistical Learning*, Chapter 8.2. We used the implementation in the R package called randomForest. See "Package 'randomforest'," dated March 25, 2018, at https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.

increasing mortality among veterans from the World War II and Korean War eras, the veteran percentage has been consistently declining over the past decade, perhaps signaling a lower societal attachment to the military and tending to drive the number of accessions down.

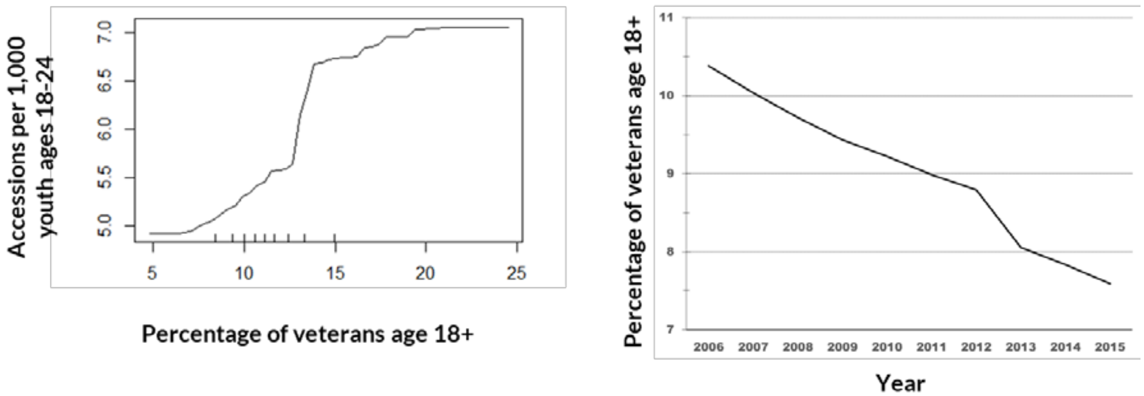


Figure 30. Machine Learning Estimates, All Accessions, for the Percentage of Veterans among the Population Age 18 and Older

The second variable of interest is the percentage of veterans among the population ages 35–54 in a pseudo-county (Figure 31). That age range represents well family connections and mentors: parents, aunts and uncles, teachers, and coaches who may have served in the military. This variable, too, has a positive association with accessions but has been declining over the past decade, tending to drive the number of accessions down.

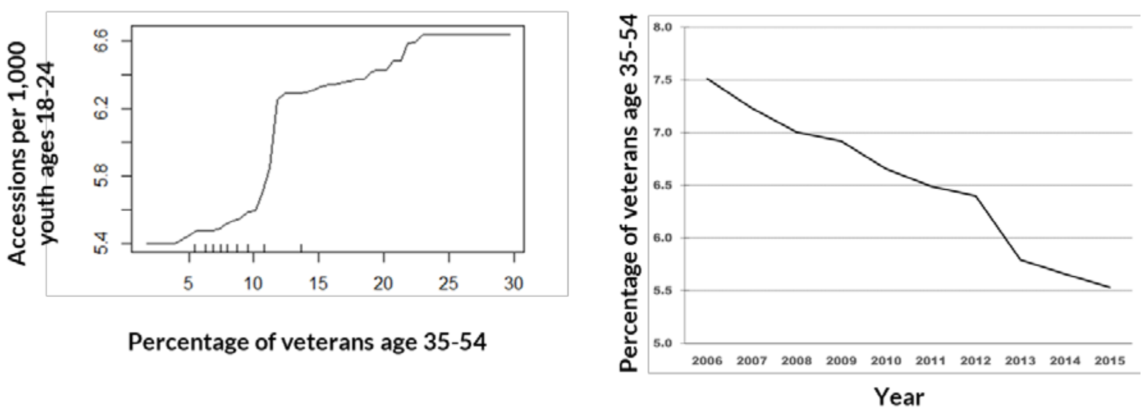


Figure 31. Machine Learning Estimates, All Accessions, for the Percentage of Veterans among the Population Ages 35–54

The next variable is the percentage of veterans, of all ages, who served during Gulf War I (Operations Desert Storm and Desert Shield). Most of those veterans are in the same age range as veterans depicted in the preceding figure. A relatively *young* veteran from

Gulf War I would have been age 20 during that conflict (in 1991), having been born in 1971. Over our sample period, that veteran's age would range from 35 in 2006 to 44 in 2015. A relatively *old* veteran might have been age 30 during the conflict, being born in 1961 (though many would have been still older). Over our sample period, that veteran's age would range from 45 in 2006 to 54 in 2015. The difference between Figure 32 and the preceding Figure 31 is that in Figure 31 we were measuring (essentially) Gulf War I veterans relative to the total population in the age range 35–54; in Figure 32, we are measuring Gulf War I veterans relative to all veterans regardless of age. The smaller effect in the latter case is indicated by the much more compressed scale on the Y-axis of the left-hand chart in Figure 32 as compared to the left-hand chart in Figure 31. That is, for predicting accessions, the percentage of veterans in the population is more important than the particular era in which those veterans served.

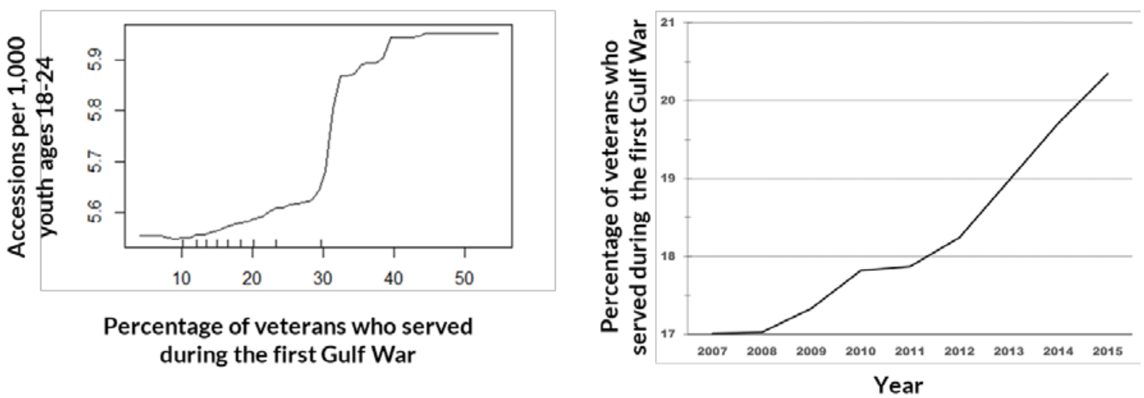


Figure 32. Machine Learning Estimates, All Accessions, for the Percentage of Veterans Who Served during the first Gulf War

College attendance works in two ways against military recruiting. First, youth currently enrolled in college are not immediately available for military service, although it must be remembered that 41 percent of them (by the most recent estimate) do not complete college and are an important pool of potential recruits.[34] Second, high levels of college attendance in a community may signal a culture in which most high school graduates are expected to enroll in and complete college. Indeed, high school juniors and seniors may already be predisposed against military service based on the prevailing local culture.

34 "Fast Facts," National Center for Education Statistics, https://nces.ed.gov/fastfacts/display.asp?id=40. According to that source, among first-time, full-time undergraduate students who began seeking a bachelor's degree at a four-year degree-granting institution in fall 2009, only 59 percent had completed a bachelor's degree by 2015 at the same institution where they started in 2009 (the "six-year graduation rate").

College enrollment certainly varies geographically, and the national averages have shown some volatility with the economy and other factors (see Figure 33). Many youth enrolled in college when civilian employment prospects were relatively poor during the Great Recession of 2007–2009, and remained in college to complete their four-year degrees (college attendance peaked in 2012 and remained elevated through 2013). It may be argued that other youth were pushed toward military service in response to higher unemployment rates during the Great Recession, but the unemployment effects estimated in Chapter 3 were quite variable across regions. In addition, the response by many young people to enroll in college was more enduring, in that college completion (for those who did complete) made those students basically unavailable to the military for at least four years.

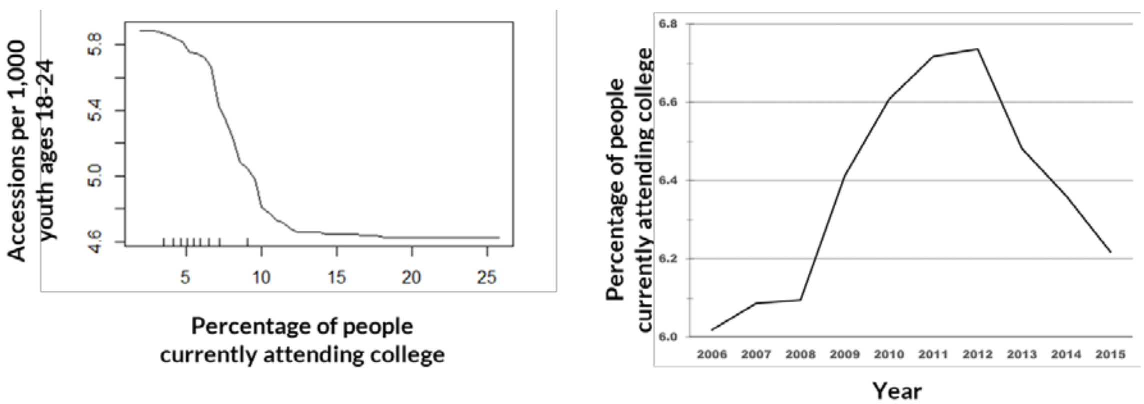


Figure 33. Machine Learning Estimates, All Accessions, for the Percentage of People Currently Attending College

The number of degree-granting institutions represents another aspect of colleges and universities competing against the military for the youth population (see Figure 34). This variable measures the proximity to (or density of) colleges and universities to youth living in a particular pseudo-county. The variable is inexact for our purposes in that some local youth attend distant colleges and, conversely, some student at local colleges come from distant places (e.g., out-of-state students). In any case, the effect is small as indicated by the much more compressed scale on the Y-axis of the left-hand chart in Figure 34 as compared to the left-hand chart in Figure 33.

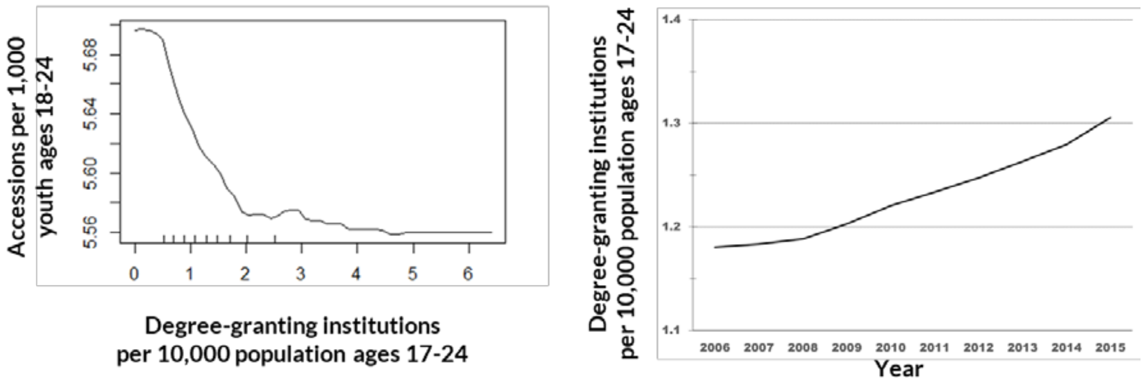


Figure 34. Machine Learning Estimates, All Accessions, for the Number of Degree-Granting Institutions per 10,000 Population Ages 17–24

The next variable is the number of high schools that offer a JROTC program per 1,000 youth population.[35] Although there is no mandatory service obligation for JROTC students, there is some direct (albeit dated) evidence that the presence of JROTC programs is correlated with enhanced military recruiting.[36] However, causation is more difficult to prove than correlation because the presence of JROTC programs could simply reflect a local culture that is more positive about military service. Our machine learning analysis, too, indicates a modest association between JROTC programs and accessions (see Figure 35). We cannot display the time trend in JROTC programs because the data were available for only the single year 2017.

---

[35] Some JROTC programs are housed at non-traditional schools such as Air Academy High, the Bedford Educational Center, the Diamond Oaks Career Development Center, the Marine Academy of Science and Technology, the Sarasota Military Academy, the Scarlet Oaks Career Center, and Utah Military Academy.

[36] Tyrone Walls, Major, United States Marine Corps, “Junior Reserve Officers’ Training Corps: A Comparison with Other Successful Youth Development Programs and an Analysis of Military Recruits who Participate in JROTC” (master’s thesis, Naval Postgraduate School, June 2003).

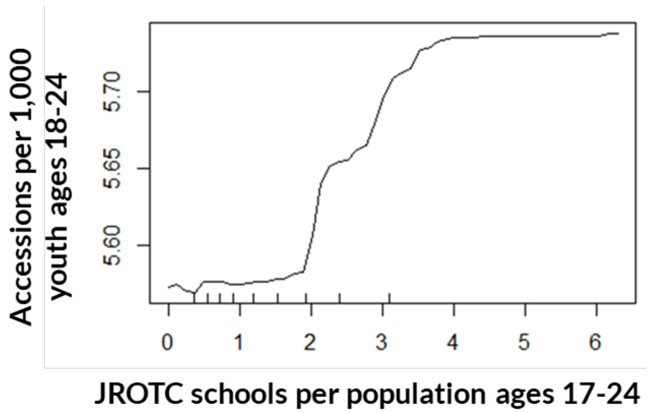


Figure 35. Machine Learning Estimates, All Accessions, for the Number of JROTC Schools per 10,000 Population Ages 17–24

Another indicator of JROTC presence in a community is the number of cadets per youth population (ages 17–24). The denominator here is not ideal because the majority of cadets would be at most age 18. However, population counts for single years of age (e.g., age 17) are not available for many of the smaller US counties (see Figure 36). Once again, we cannot display the time trend in JROTC programs because the data were available for only the single year 2017.

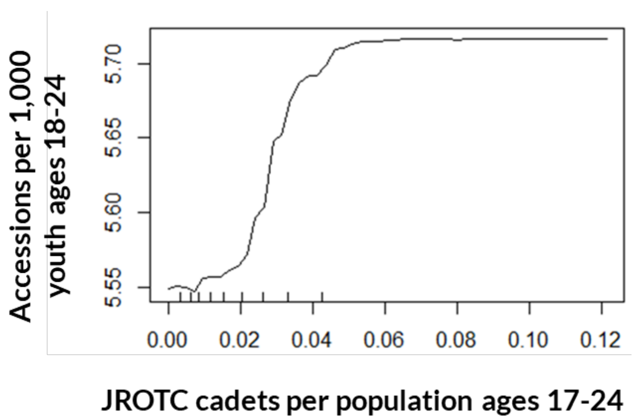


Figure 36. Machine Learning Estimates, All Accessions, for the Number of JROTC Cadets per Population Ages 17–24

The final important variable is the number of high school students available to take the ASVAB.[37] Once again, the denominator of youth population (ages 17–24) is not ideal. Within a relatively narrow range, having more students take the ASVAB is associated with higher recruiting totals. An encouraging sign is that the number of students taking the

[37] In addition to traditional high schools, the ASVAB is administered at career centers (e.g., A.W. Beattie Career Center), job corps (e.g., Pittsburgh Job Corps), and assorted other “centers” (e.g., Bullitt County, Kentucky, Detention Center; Job Academy; and Clay County, Kentucky, Adult Education Center).

ASVAB by this measure has increased by about 25 percent over the past decade (see Figure 37). Again, while demonstrating an association, this analysis does not sort out the direction of causality between ASVAB availability and recruiting success.

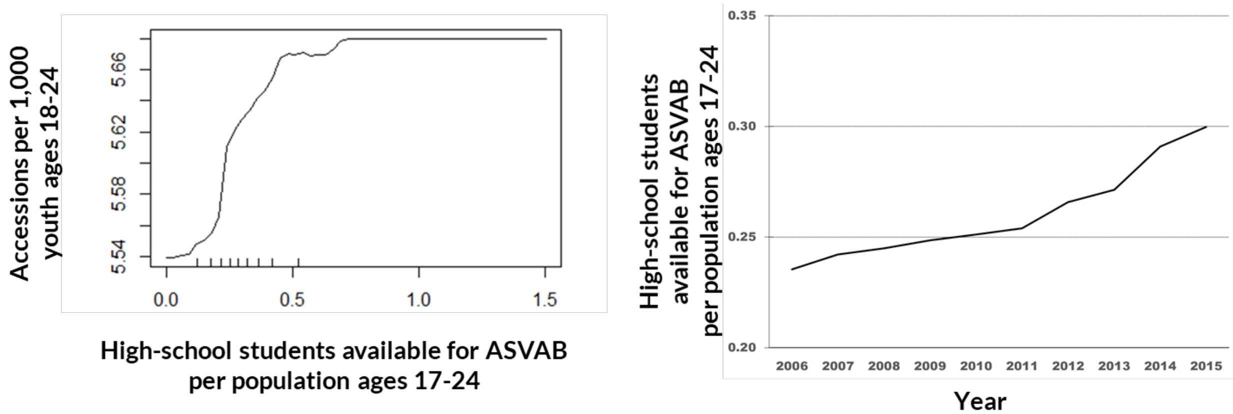


Figure 37. Machine Learning Estimates, All Accessions, for the Number of High School Students Available to Take the ASVAB per Population Ages 17–24

B. High-Quality Male Accessions

The machine learning estimates for high-quality males (male high school graduates and seniors who score above the median on the AFQT) are very similar to those reported in the preceding section for all recruits. The sets of potential predictor variables are identical in the two cases. Some of the variables, such as the number of degree-granting institutions, are only sensibly defined for all youth. Other variables, such as the three measures of veterans in the community, might have been measured for males only (i.e., male-only veterans in the two age ranges or who served during the Gulf War I era), although that approach would presume that male youth are more highly influenced by the presence of male veterans than female veterans. And it would have been extremely difficult to estimate the presence of *high-quality* male veterans in each pseudo-county (those who had scored above the median on the AFQT).

The findings for high-quality males are presented in Figure 38 through Figure 46.

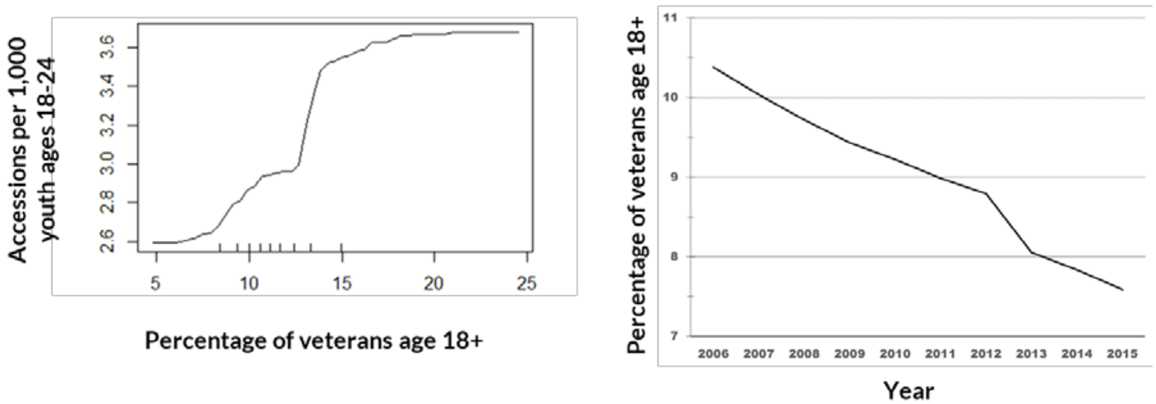


Figure 38. Machine Learning Estimates, High-Quality Male Accessions, for the Percentage of Veterans among the Population Age 18 and Older

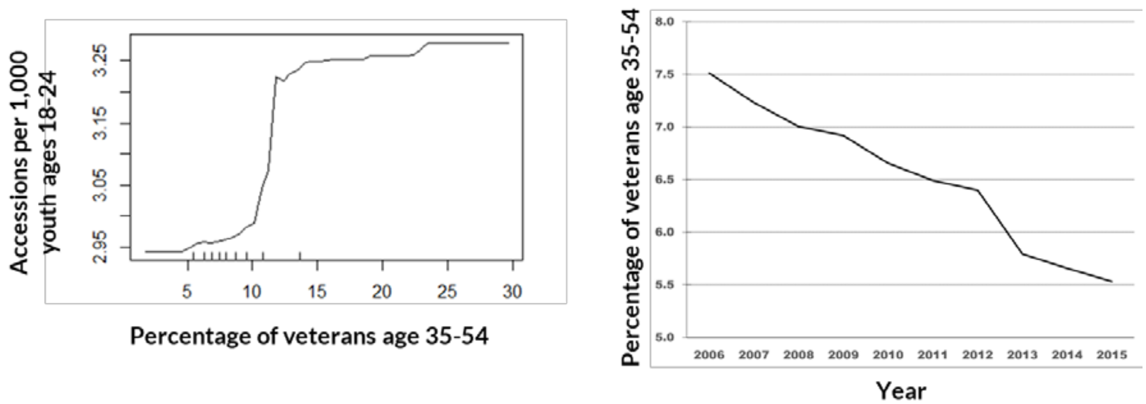


Figure 39. Machine Learning Estimates, High-Quality Male Accessions, for the Percentage of Veterans among the Population Ages 35–54

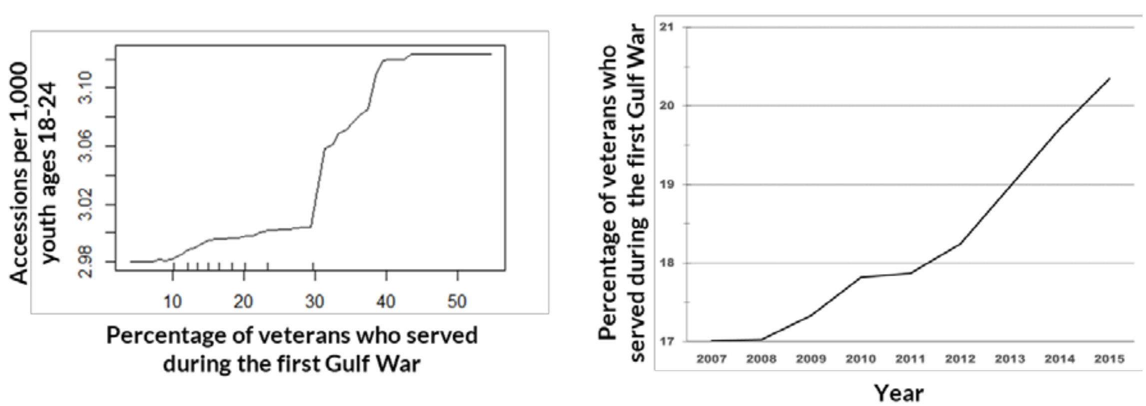


Figure 40. Machine Learning Estimates, High-Quality Male Accessions, for the Percentage of Veterans Who Served during the First Gulf War

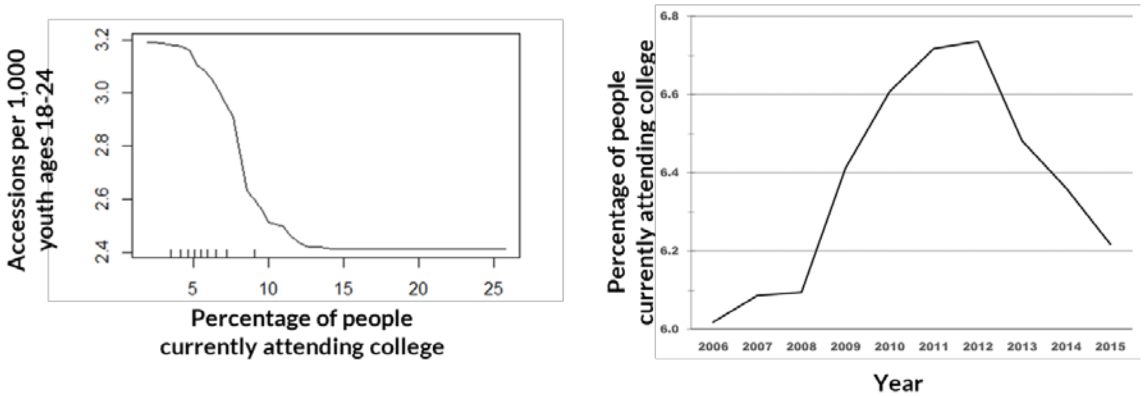


Figure 41. Machine Learning Estimates, High-Quality Male Accessions, for the Percentage of People Currently Attending College

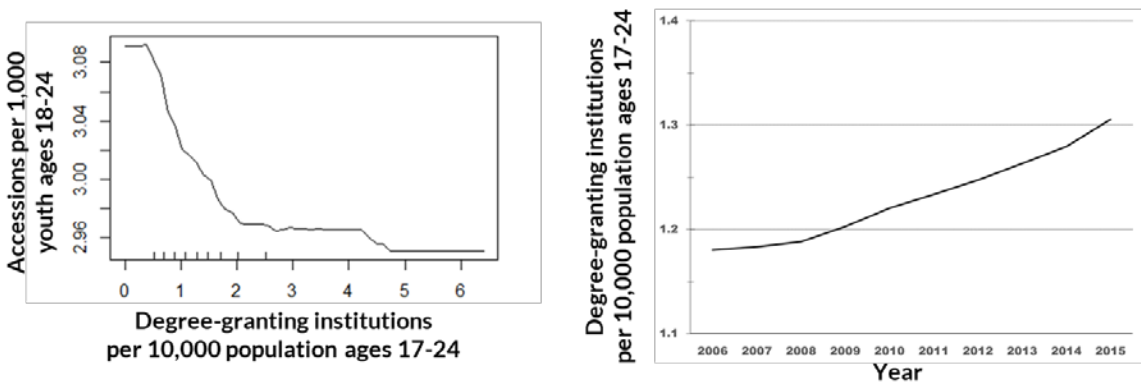


Figure 42. Machine Learning Estimates, High-Quality Male Accessions, for the Number of Degree-Granting Institutions per 10,000 Population Ages 17–24

The only additional important predictor relative to the preceding section is the percentage of people in the community with some college (less than 1 year, more than 1 year but less than a bachelor's degree, or an associate's degree) (Figure 43). The left-hand chart in Figure 43 shows the effect of that variable, associated with an increase in the expected number of high-quality male accessions in a range from about 3.0 to 3.1 per 1,000 youth population. However, the right-hand chart indicates that the yearly average values of the percentage with some college have increased from about 21.5 percent to 24.5 percent; the range across individual pseudo-counties is much wider than that of the annual averages. Within a reasonable range, and holding fixed other factors that also influence accessions, the percentage with some college is associated with increases in high-quality male accessions in a range of about 2.95 to 3.05 per 1,000 youth population. Although the magnitude here is modest, people who have not completed a four-year degree (many of whom will not—but recognizing that, for example, most college juniors will shortly complete their degrees) are an important pool of potential recruits. The recruiting community has the option to, for example, more aggressively recruit recent graduates with associate’s degrees.

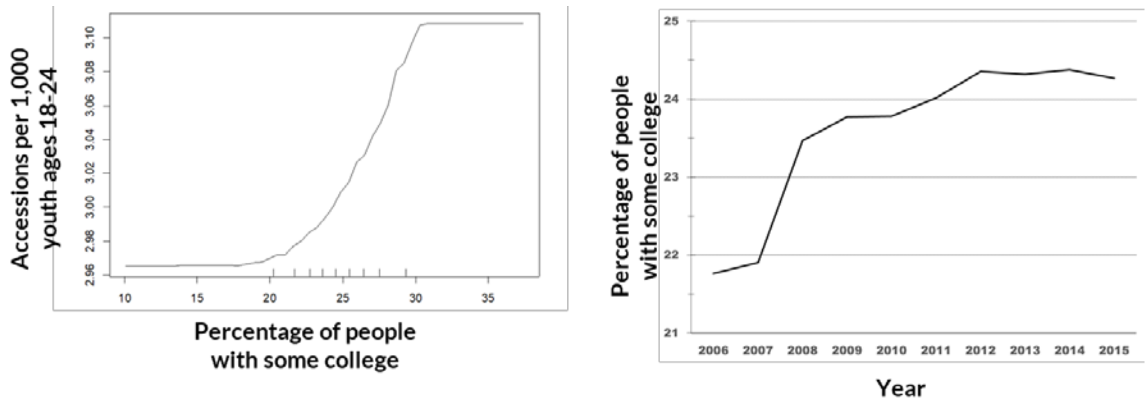


Figure 43. Machine Learning Estimates, High-Quality Male Accessions, for the Percentage of People with Some College (Less than a Bachelor’s Degree, or an Associate’s Degree)

The U-shaped patterns for JROTC programs that are observed here (Figure 44 and Figure 45) were not present in the findings for all recruits (Figure 35 and Figure 36). The U-shaped patterns are unique to high-quality *males*. The patterns for high-quality *females* (shown in the next section) are more conventional, associating near-steady increases in accessions with increases in either JROTC schools or JROTC cadets (the latter measured in total for both genders).

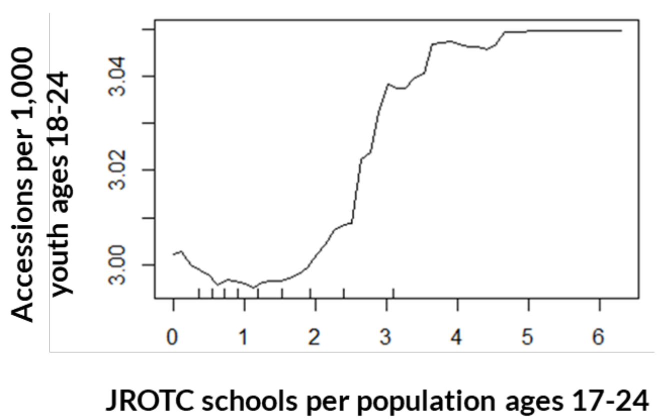


Figure 44. Machine Learning Estimates, High-Quality Male Accessions, for the Number of JROTC Schools per 10,000 Population Ages 17–24

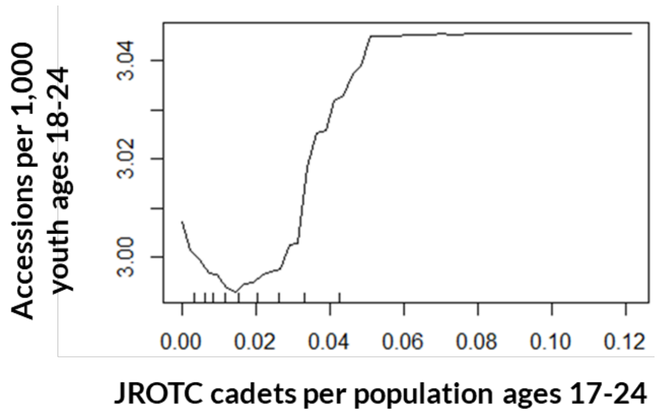


Figure 45. Machine Learning Estimates, High-Quality Male Accessions, for the Number of JROTC Cadets per Population Ages 17–24

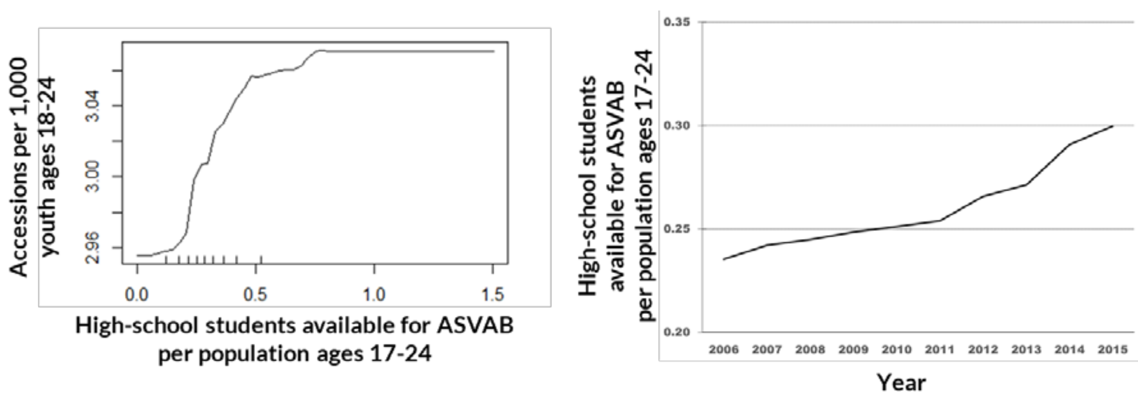


Figure 46. Machine Learning Estimates, High-Quality Male Accessions, for the Number of High School Students Available to Take the ASVAB per Population Ages 17–24

C. High-Quality Female Accessions

The machine learning estimates for high-quality females are similar to those reported in the preceding section for their male counterparts. The sets of potential predictor variables are the same for both genders. The findings for high-quality females are presented in Figure 47 through Figure 55. In comparison with the charts for high-quality males, the ranges along the Y-axes are more compressed here because the national number of high-quality female accessions is only about 20 percent the corresponding number for males (see Table 8).

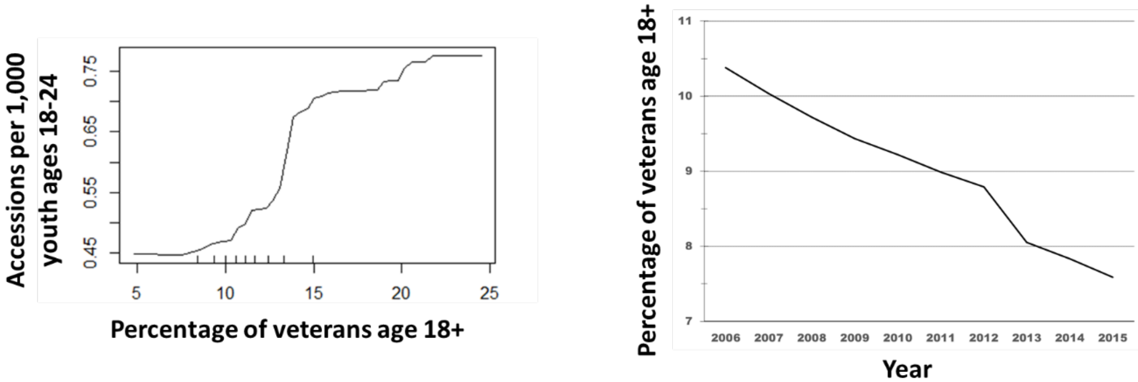


Figure 47. Machine Learning Estimates, High-Quality Female Accessions, for the Percentage of Veterans among the Population Age 18 and Older

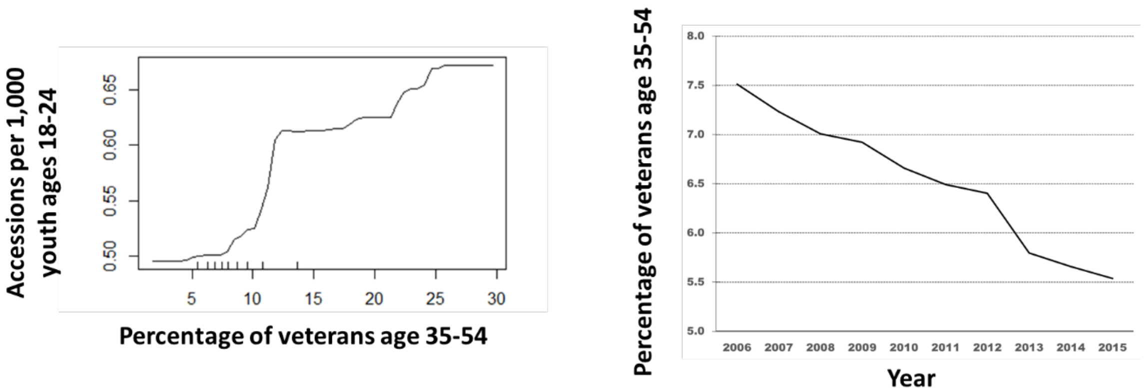


Figure 48. Machine Learning Estimates, High-Quality Female Accessions, for the Percentage of Veterans among the Population Ages 35–54

One difference from high-quality males is that, for females, the percentage of Gulf War veterans is not positively associated with the accession rate throughout its entire range (Figure 49). A small presence of Gulf War veterans in the community initially causes the accession rate to dip, but the accession rate starts to climb again when the percentage of veterans from that era reaches about 20 percent.

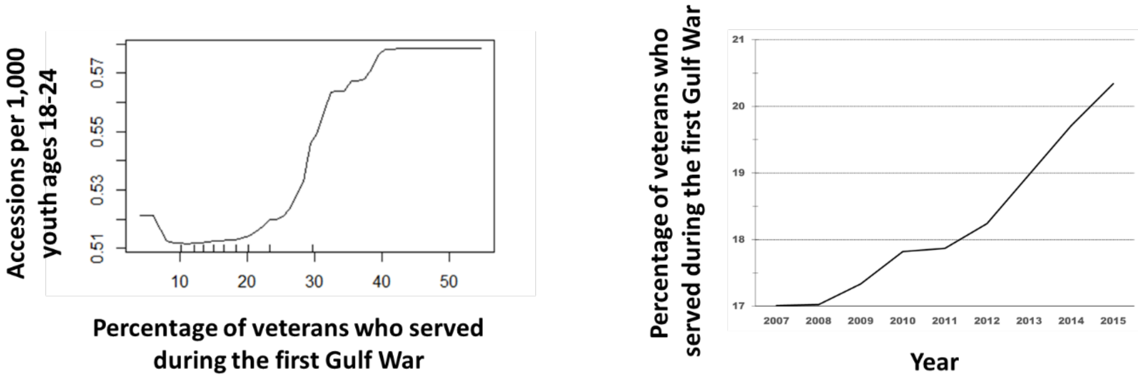


Figure 49. Machine Learning Estimates, High-Quality Female Accessions, for the Percentage of Veterans Who Served during the First Gulf War

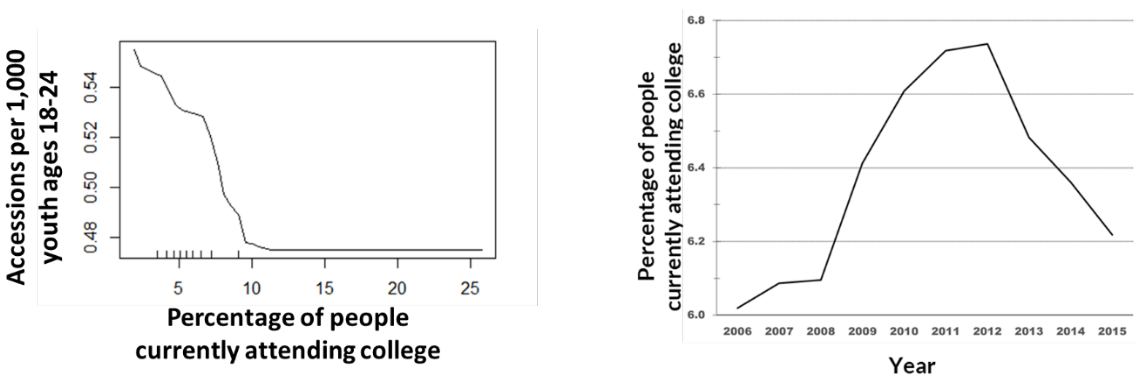


Figure 50. Machine Learning Estimates, High-Quality Female Accessions, for the Percentage of People Currently Attending College

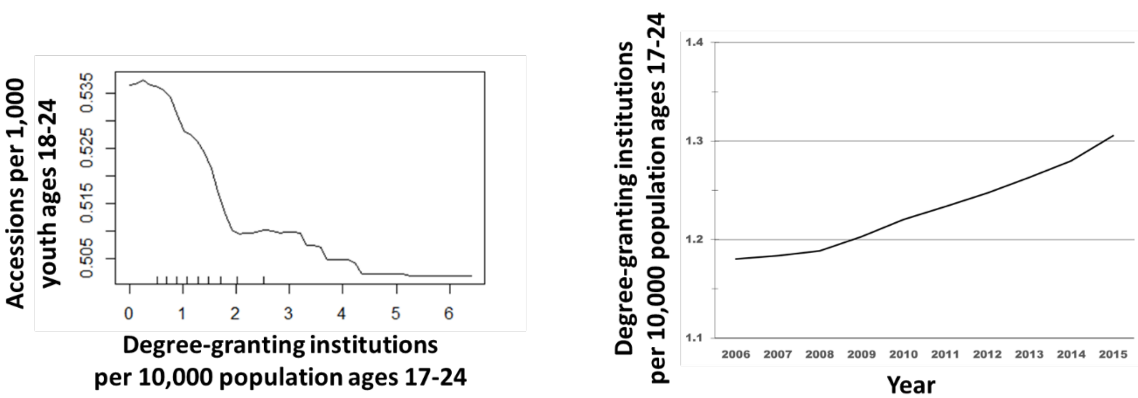


Figure 51. Machine Learning Estimates, High-Quality Female Accessions, for the Number of Degree-Granting Institutions per 10,000 Population Ages 17–24

Regarding JROTC programs, the patterns for high-quality females are more conventional than for high-quality males, associating near-steady increases in accessions with increases in either JROTC schools or JROTC cadets (the latter measured in total for both genders). These results are shown in Figure 52 and Figure 53.

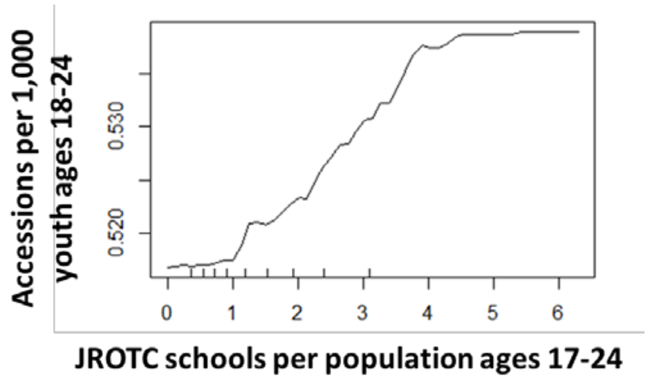


Figure 52. Machine Learning Estimates, High-Quality Female Accessions, for the Number of JROTC Schools per 10,000 Population Ages 17–24

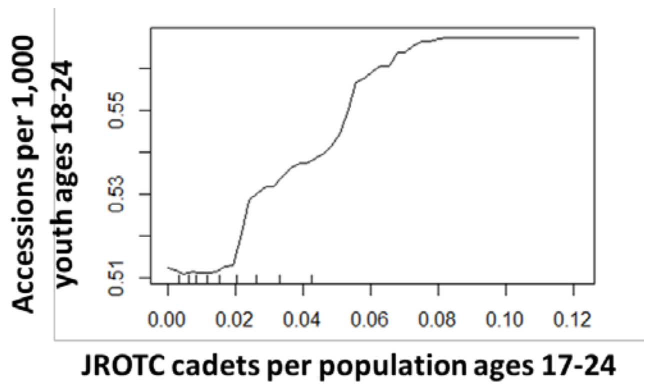


Figure 53. Machine Learning Estimates, High-Quality Female Accessions, for the Number of JROTC Cadets per Population Ages 17–24

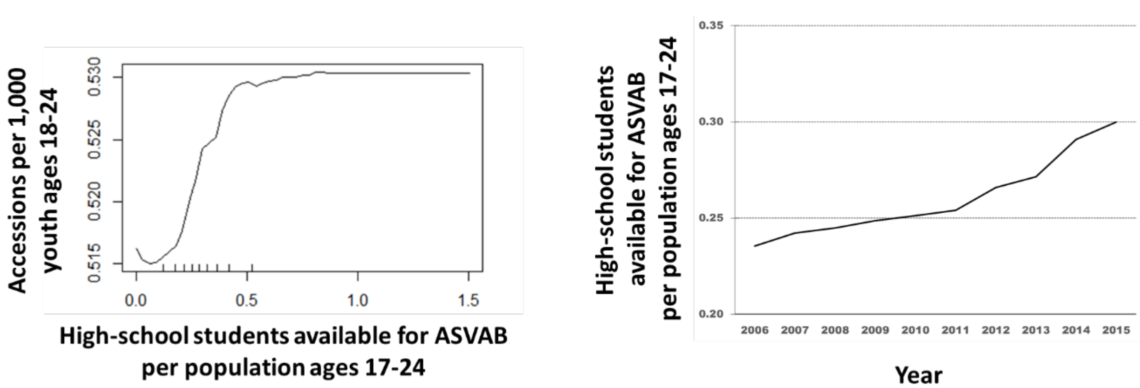


Figure 54. Machine Learning Estimates, High-Quality Female Accessions, for the Number of High-School Students Available to Take the ASVAB per Population Ages 17–24

The only additional important predictor for high-quality females is the ratio of the number of veterans in the community with a college degree or higher, divided by the number of non-veterans with a college degree or higher. The effect of this ratio is U-shaped with a trough at about 0.9 (the left-hand chart in Figure 55). For pseudo-counties in which veterans are generally less educated than non-veterans (i.e., low values of the ratio), a small increase in veteran education is associated with less successful recruiting of high-quality females. But when the ratio begins to exceed 0.9, further increases in veteran education are associated with improved recruiting. The yearly average values of the ratio exceed 1.0 (veterans are more likely than non-veterans to be college educated; see the right-hand chart in Figure 55), but the range across individual pseudo-counties is much wider than that of the annual averages. As an example, and holding fixed other factors that also influence accessions, an increase in the veteran education ratio from 1.0 to 1.2 might be associated with an increase in the accession rate for high-quality females from about 51.2 to 52.6 per 100,000 total youth ages 18–24 (or, equivalently, per 50,000 females in that age range).

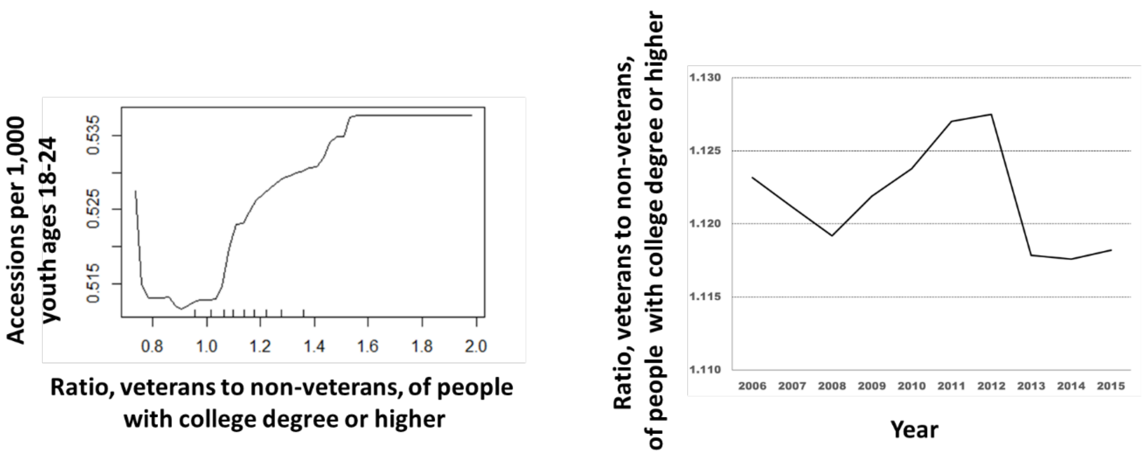


Figure 55. Machine Learning Estimates, High-Quality Female Accessions, for the Ratio of College Degree or Higher, Veterans to Non-Veterans

D. Summary of Machine Learning Results

Conspicuous by their absence from the sets of most important predictors are several economic variables. We explored the statistical relationship between the representation ratio and the youth unemployment rate in the correlation and regression analyses of Chapter 3, yielding mixed results. We did not include youth unemployment rates in the machine learning analysis because the BLS does not have adequate sample sizes to estimate unemployment rates within age narrow brackets by county. Two of the important predictors from the machine learning analysis may be construed as economic in nature: the percentage of people currently attending college, and the percentage of people having completed some college.

Several other economic factors were considered in the machine learning analysis, but none of them rose to the top of the list by the criterion of large reductions in the mean-squared error when predicting the accession rate. Among those economic factors and their sources are:

- ACS: average household income, additional measures of educational attainment, number of vehicles in a household, type of housing unit (e.g., number of bedrooms), rent versus purchase home, average mortgage payment.
- BLS QCEW: percentage of workers employed by the government (all levels), percentage of workers employed by private employers.

Having controlled for college attendance and attainment, and notwithstanding the statistical association with youth unemployment rates, recruiting success in local areas appears to be driven by largely cultural and demographic factors.

8. Recent Trends in Important Predictors of Recruiting Success

Having identified the most important factors that are associated with accessions, this chapter presents the trends in those factors over the past decade. Some factors, such as the prevalence of JROTC programs and of high school students taking the ASVAB, are to varying degrees within the control of the recruiting community, albeit at some cost. Other factors are beyond the recruiting community's policy levers, but are nonetheless worth noting because they describe the recruiting environment. The trends for seven key factors, by Census region, are shown in Figure 56 through Figure 62.

If a factor has a positive association with the accession rate and that factor is increasing (or can be further increased by policy actions), the prognosis for the future is optimistic. One example would be the increasing trend in the percentage of people with some college (less than a bachelor's degree, or an associate's degree) (Figure 61). Similarly, a factor having a negative association with the accession rate but that is declining over time also portends an optimistic future. When the two directions are opposite, however, the prognosis is pessimistic. For example, while two indicators of veteran representation in the population have positive associations with accession rates, those indicators are trending down over time (Figure 56 and Figure 57).

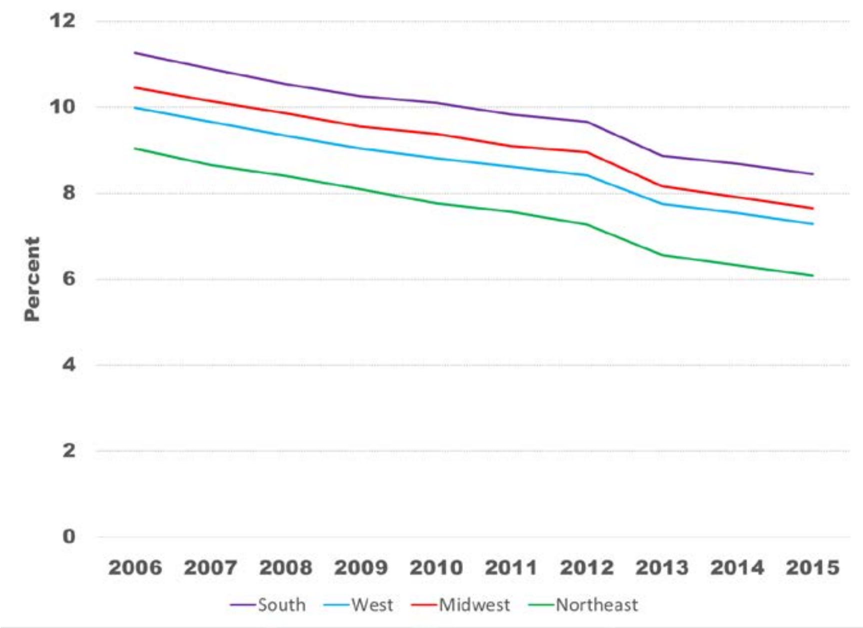


Figure 56. Regional Trends in Percentage of Veterans among the Population Age 18 and Older

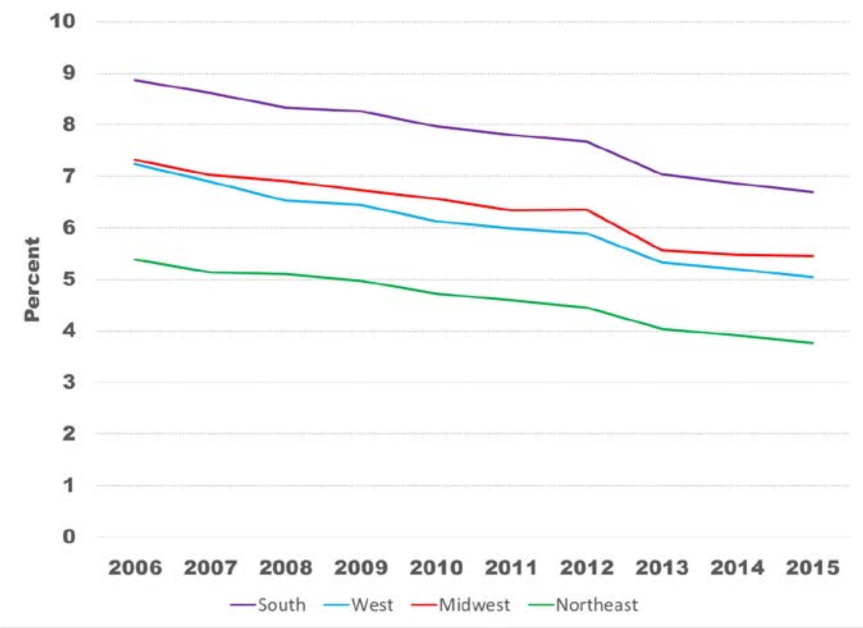


Figure 57. Regional Trends in Percentage of Veterans among the Population Ages 35–54

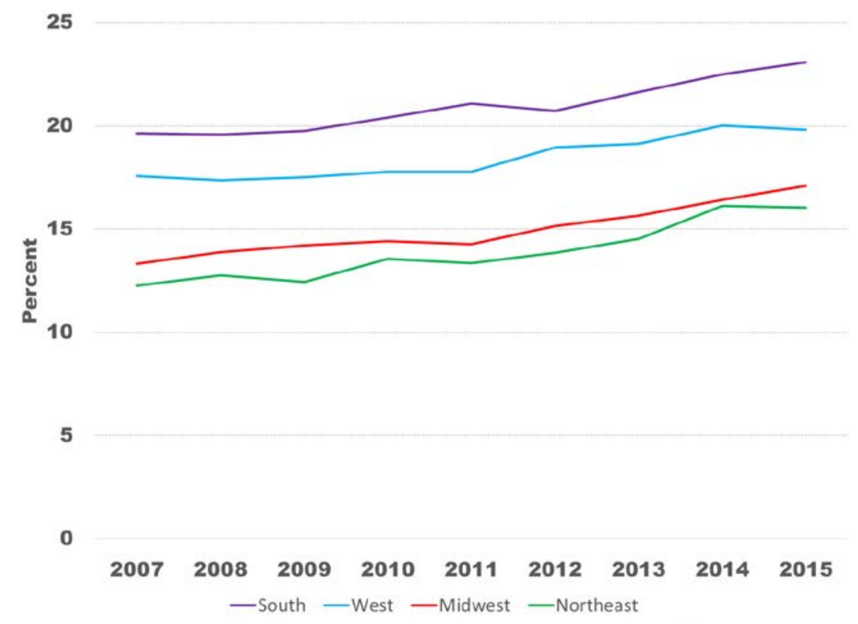


Figure 58. Regional Trends in Percentage of Veterans Who Served during the First Gulf War

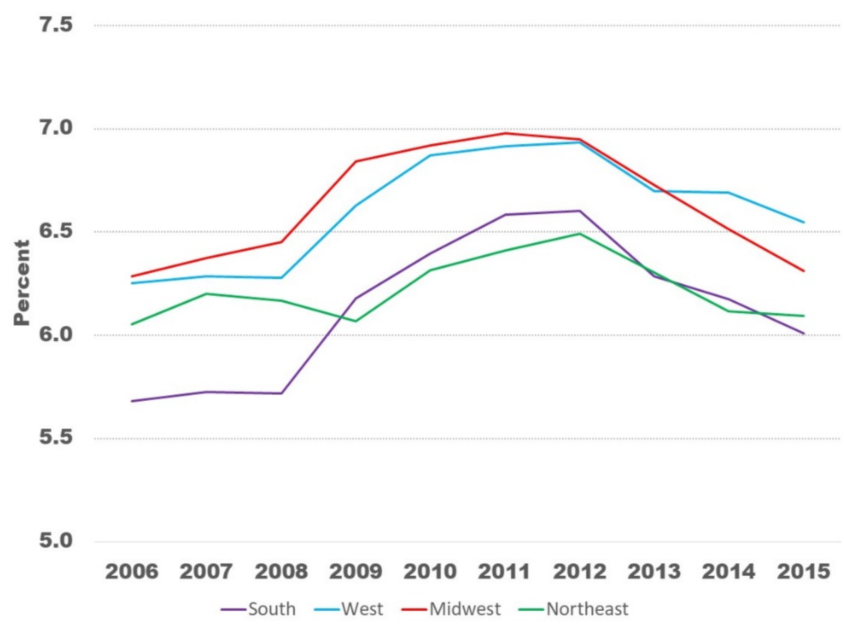


Figure 59. Regional Trends in Percentage of People Currently Attending College

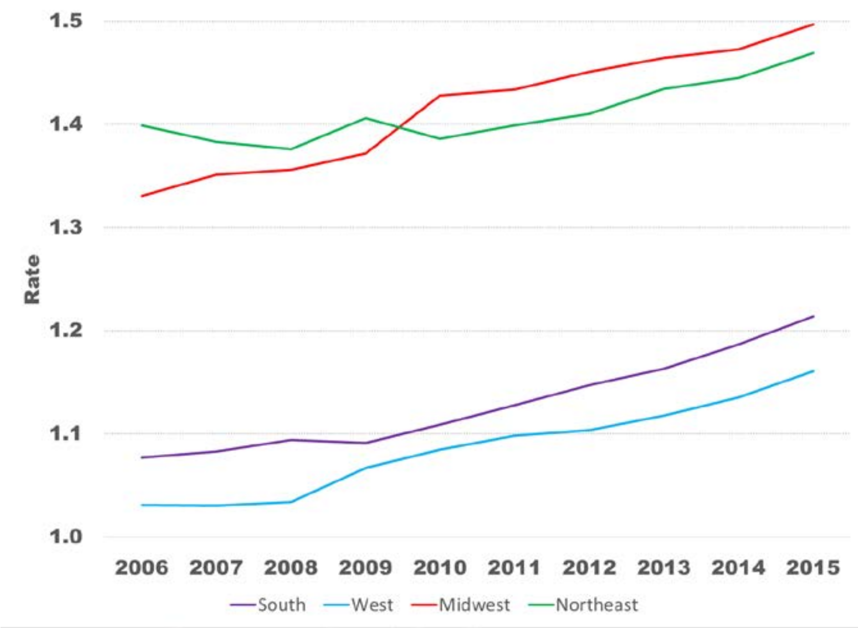


Figure 60. Regional Trends in Number of Degree-Granting Institutions per 10,000 Population Ages 17–24

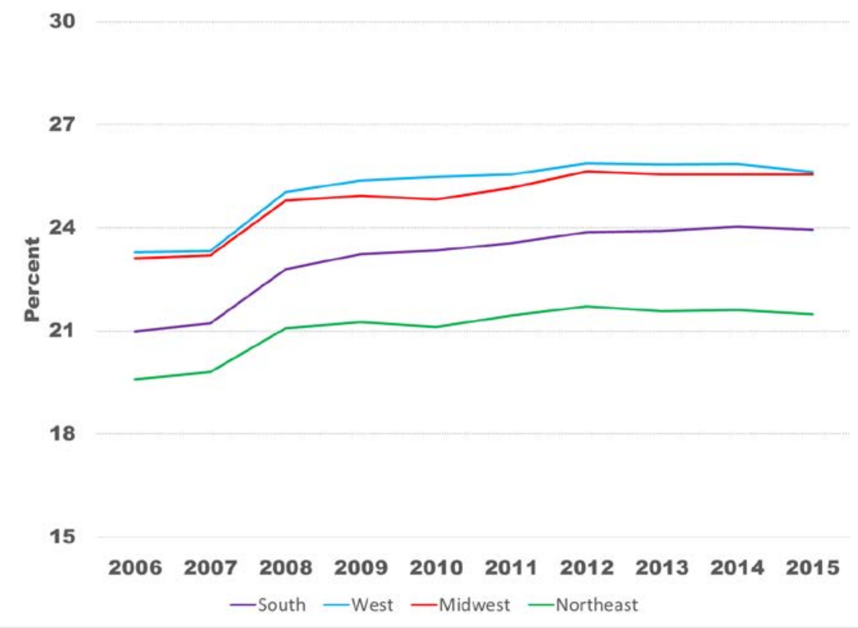


Figure 61. Regional Trends in Percentage of People with Some College (Less than a Bachelor's Degree, or an Associate's Degree)

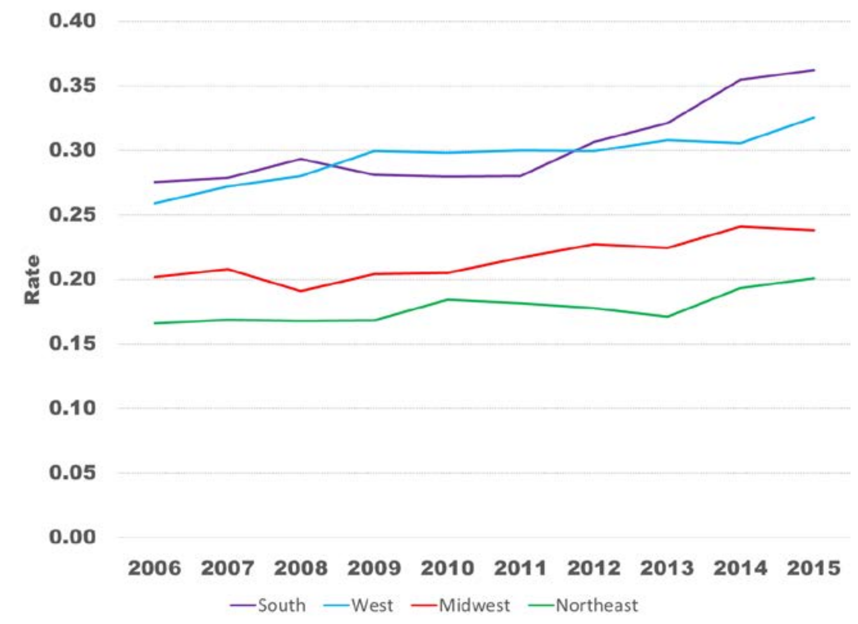


Figure 62. Regional Trends in High School Students Available for the ASVAB per Population Ages 17–24

9. Conclusions

The first objective of this project was to identify the demographic, economic, and cultural factors in a community that predict the percentage of its youth population that enlist in the military. When viewed across the entire nation, those are the factors that determine geographical diversity. The second objective was to identify trends and events that could affect geographical diversity in the future.

To achieve the first objective, the IDA team conducted a machine learning analysis of the determinants of non-prior-service enlisted accessions at the pseudo-county level for the years 2006 through 2015. We first considered all accessions, both male and female, without regard to quality level. Then we focused on the smaller, high-interest subset of high-quality male accessions: high-school graduates and seniors who score above the median (categories 1-3A) on the AFQT. Finally, we performed the corresponding analysis for high-quality female accessions.

Having identified the most important factors that are associated with accessions for those three groups, we presented the trends in those factors over the past decade. The prognosis is optimistic when either a factor having a positive association with accessions is trending upward ("more of a good thing"), or when a factor having a negative association is trending downward ("less of a bad thing"). The prognosis is pessimistic in the opposite two situations. For example, the percentage of people with some college (less than 1 year, more than 1 year but less than a bachelor's degree, or an associate's degree) is a positive factor that has been trending upward, perhaps signaling an improving recruiting environment. An offsetting phenomenon is that the presence of veterans between the ages of 35 and 54 in the community—a positive factor—has been trending downward.

A fundamental difference is between factors that are to some degree within the control of the recruiting community and those that describe the recruiting environment but for which there are no policy levers. Two examples of the former (i.e., where there are policy levers, albeit perhaps limited) are the prevalence of JROTC programs and of high school students taking the ASVAB. The JROTC program is jointly funded by DoD (about $375 million per year) and the local school districts (about $225 million). However, under current congressional caps, DoD is limited to between 3,000 and 3,700 JROTC units. Similarly, DoD has the option to encourage more high schools to administer the ASVAB. However, that initiative is largely saturated to the point where the limiting factor is not DoD's policy or level of effort, but rather the number of willing participants.

A. All Recruits

Looking first at all recruits, three measures of veteran presence in the community were among the most important predictors of recruiting success: the percentage of veterans among the population age 18 and older, the percentage of veterans among the population ages 35–54, and the percentage of veterans who served during the first Gulf War. The first of those measures is an indicator of contact with veterans, regardless of the ages of those veterans or the period in which they served. With increasing mortality among veterans from the World War II and Korean War eras, the veteran percentage has been consistently declining over the past decade, perhaps signaling a lower societal attachment to the military and tending to drive the number of accessions down. Although adverse trends can often be offset—using tools such as increases in marketing efforts, enlistment bonuses, or the number of recruiters—the recruiting community has no direct policy lever to reverse the decline in the overall veteran population.

The next two measures pertain to slightly different groups of veterans but in essentially the same age range, 35–54. We find that for predicting recruiting success, the pure age effect seems more important than the particular era in which those veterans served. The age range 35–54 represents well family connections and mentors: parents, aunts and uncles, teachers, and coaches who may have served in the military. This variable, too, while having a positive association with accessions, has been declining over the past decade, by itself contributing to a drop in the number of accessions. Nor is it susceptible to any direct policy lever.

The next important variable is the percentage of people currently attending college. College attendance works in two ways against military recruiting. First, youth currently enrolled in college are not immediately available for military service, although it must be remembered that about 40 percent of them (by the most recent estimate) do not complete college within six years and are an important pool of potential recruits. Second, high levels of college attendance in a community may signal a culture in which most high school graduates are expected to enroll in and complete college. College enrollment varies geographically and the national averages have shown some volatility with the economy and other factors. Many youth enrolled in college when civilian employment prospects were relatively poor during the Great Recession of 2007–2009, and remained in college to complete their four-year degrees (college attendance peaked in 2012 and remained elevated through 2013). Although trends in the national economy are not subject to policy intervention at the DoD level, they do form a predictable influence on the recruiting climate.

Another education variable, although with a smaller effect on recruiting, is the number of degree-granting institutions in the community. As constructed for this research, that variable roughly measures the proximity to (or density of) colleges and universities to youth living in a particular pseudo-county.

Two measures of JROTC density have positive effects on recruiting: the number of high schools that offer a JROTC program per youth population, and the number of cadets per youth population. Trends for those variables were not available to the IDA research team.

Finally for all recruits, although a smaller effect, having more students take the ASVAB is associated with higher recruiting totals. An encouraging sign is that the number of students taking the ASVAB has increased over the past decade. However, this analysis does not establish the direction of causation: Is it the case that greater penetration of ASVAB testing has lifted recruiting, or rather that a rising sentiment toward military service has encouraged more high schools to offer and more students to take the ASVAB? Further, even if the direction of causation could be determined (there is most likely some causality in each direction), it is no inherent contradiction that ASVAB testing has been a “good news” story at the same time that military recruiting has become, in many ways, more challenging. As this analysis demonstrates, many factors are at play—some favorable toward recruiting and others unfavorable.

All of these results should be tempered by the understanding that the IDA team was unable to obtain from DMDC historical time-series data on the numbers of recruiters at each recruiting station. The lack of that information may confound the estimated relationships between recruiting success and the variables for which historical data were available.

B. High-Quality Male Recruits

The machine learning estimates for high-quality males (male high school graduates and seniors who score above the median on the AFQT) are very similar to those just summarized for all recruits. We fed identical sets of potential predictor variables to the machine learning algorithm in both cases, and the most influential predictors were largely the same (albeit in a slightly different rank order).

The only additional important predictor for high-quality males is the percentage of people in the community with some college (less than 1 year, more than 1 year but less than a bachelor’s degree, or an associate’s degree). People who have not completed a four-year degree (many of whom will not—but recognizing that, for example, most college juniors will shortly complete their degrees) are an important pool of potential recruits. DoD has the option, for example, to more aggressively recruit recent graduates with associate’s degrees.

C. High-Quality Female Recruits

The machine learning estimates for high-quality females are largely similar to those for high-quality males but differ in a few interesting ways. The predictor for some college, which was important for high-quality males, did not turn out to be important for high-quality females. However, a new predictor that emerged for females is the ratio of the number of veterans in the community with a college degree or higher, divided by the number of non-veterans with a college degree or higher. For communities in which veterans are already more likely than non-veterans to be college educated (a ratio exceeding 1.0), further increases in veteran education are associated with improved recruiting of high-quality females. However, the effect is modest and this is not a factor over which the recruiting community holds policy levers.

D. Primacy of Cultural and Demographic Factors

Conspicuous by their absence from the sets of most important predictors are several economic variables that we included among the more than 100 total variables in the machine learning analysis. We explored the statistical relationship between accession rates and youth unemployment in preliminary correlation and regression analyses, yielding mixed results. We did not include youth unemployment rates in the machine learning analysis because the BLS does not have adequate sample sizes to estimate unemployment rates within age narrow brackets by county. Two of the important predictors that were included in the machine learning analysis may be construed as economic in nature: the percentage of people currently attending college, and the percentage of people having completed some college. Having controlled for college attendance and attainment, and notwithstanding the statistical association with youth unemployment rates, recruiting success in local areas appears to be driven by largely cultural and demographic factors.

Appendix A.
Heat Maps of Non-Prior-Service Accession Rates, Fiscal Years 2007–2015

This appendix provides heat maps for non-prior-service accession rates for the period 2007 through 2015. Total accessions were relatively stable, in the range between 152,000 and 168,000, from 2007 through 2013. For that subperiod, the year-to-year changes in the maps mostly consist of redistributions among the states of a relatively stable national total. However, total accessions fell from 165,000 in 2013 to 139,000 in 2014 (a drop of 16 percent), recovering slightly to 145,000 in 2015 (Table 8). The lightening of the maps for the latter two years may be interpreted, in part, as a more equal distribution across the states, but also as a general suppression in the accession rates required to meet a suddenly declining national total.

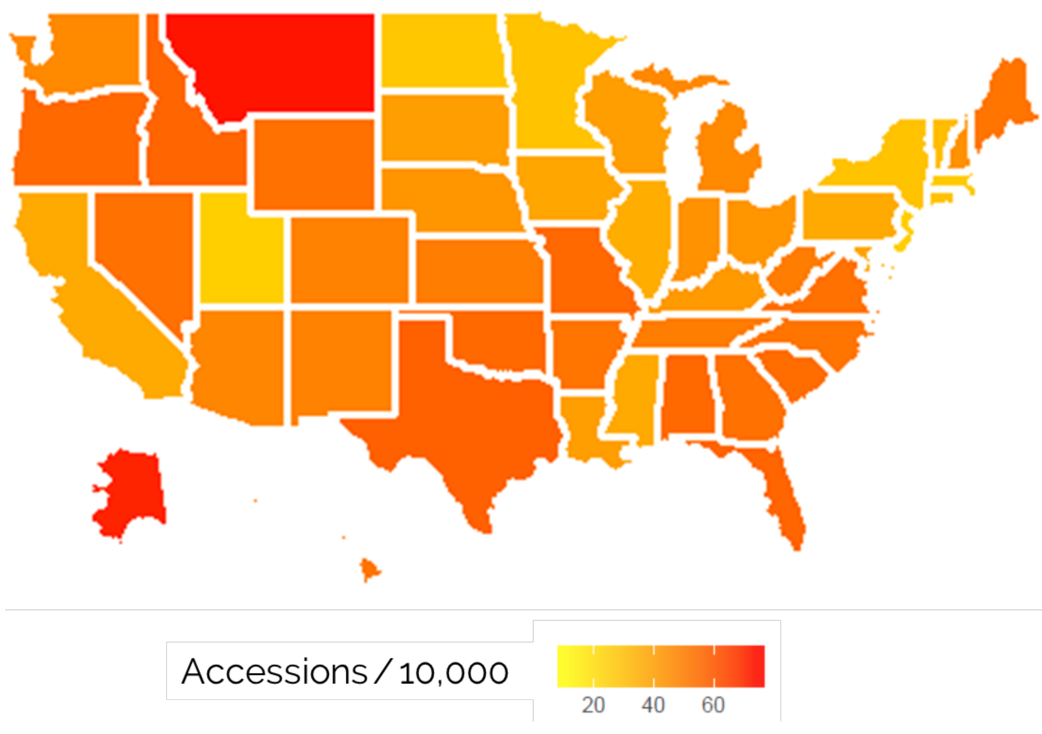


Figure A-1. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2007

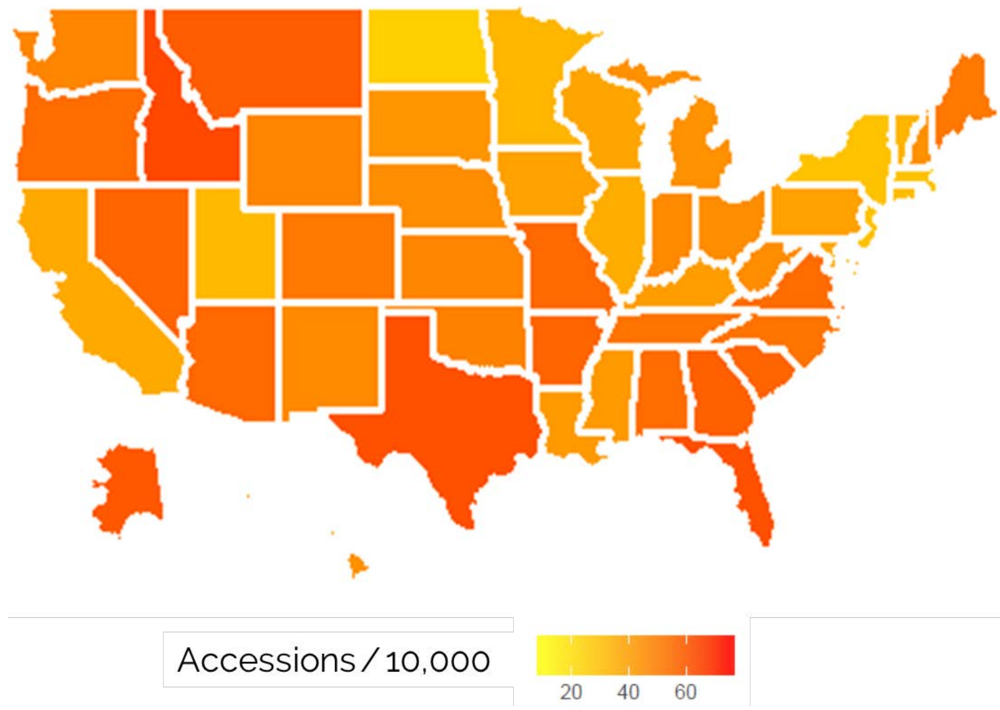


Figure A-2. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2008

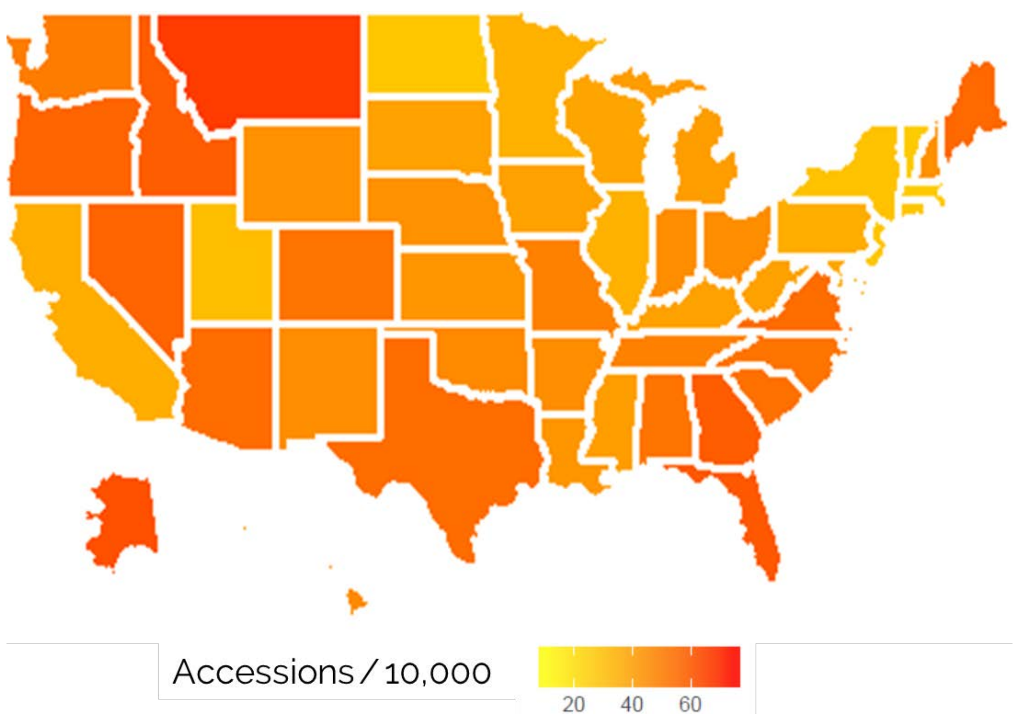


Figure A-3. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2009

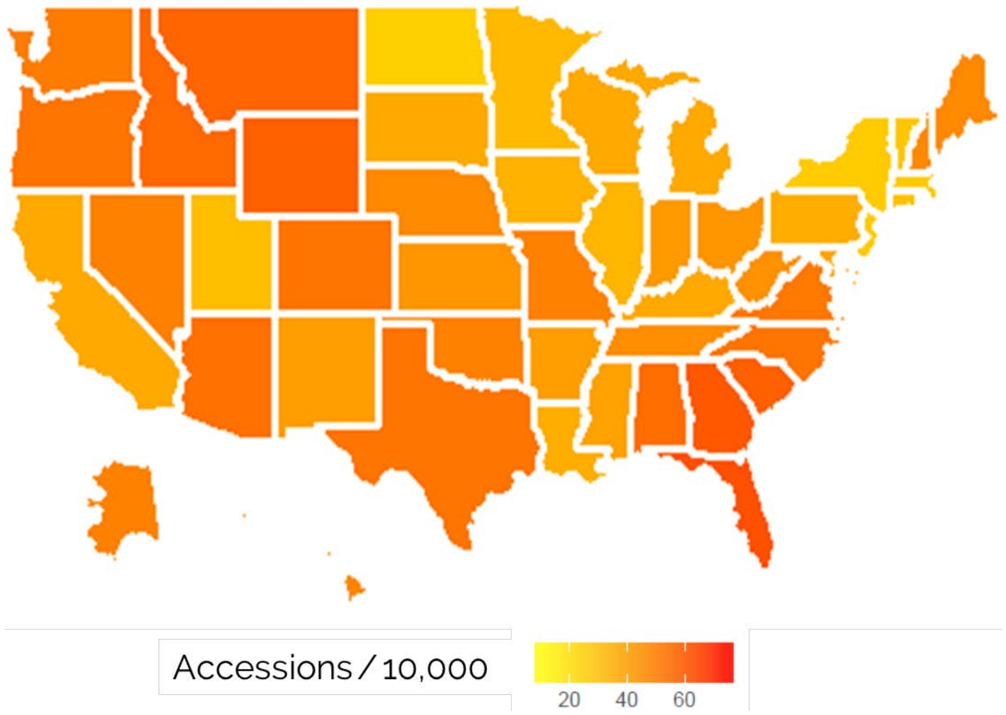


Figure A-4. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2010

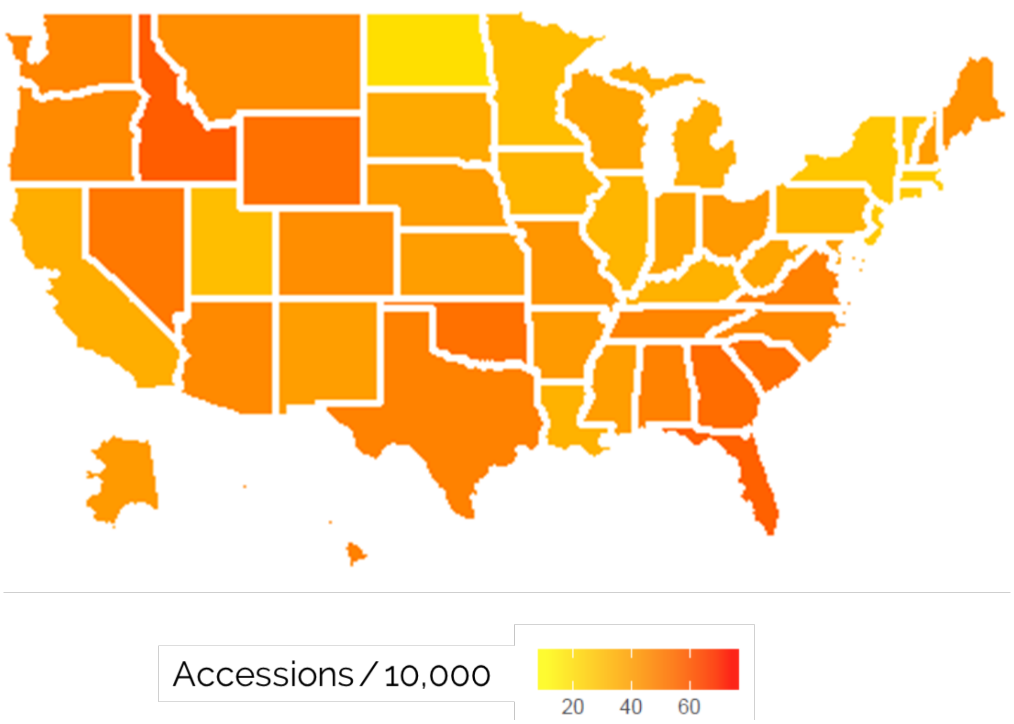


Figure A-5. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2011

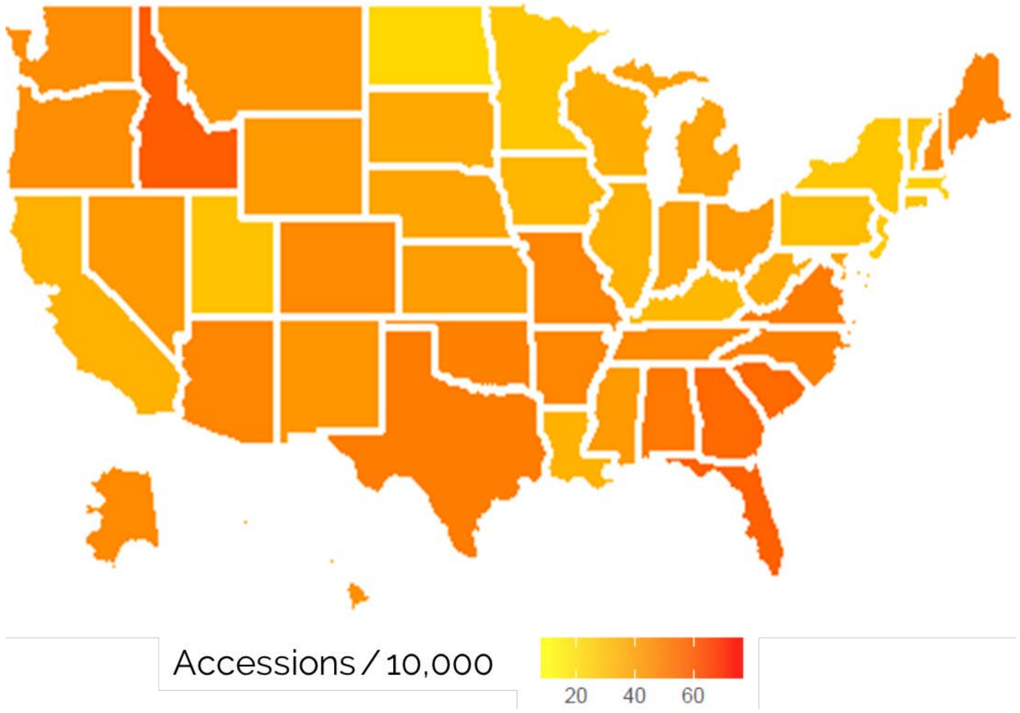


Figure A-6. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2012

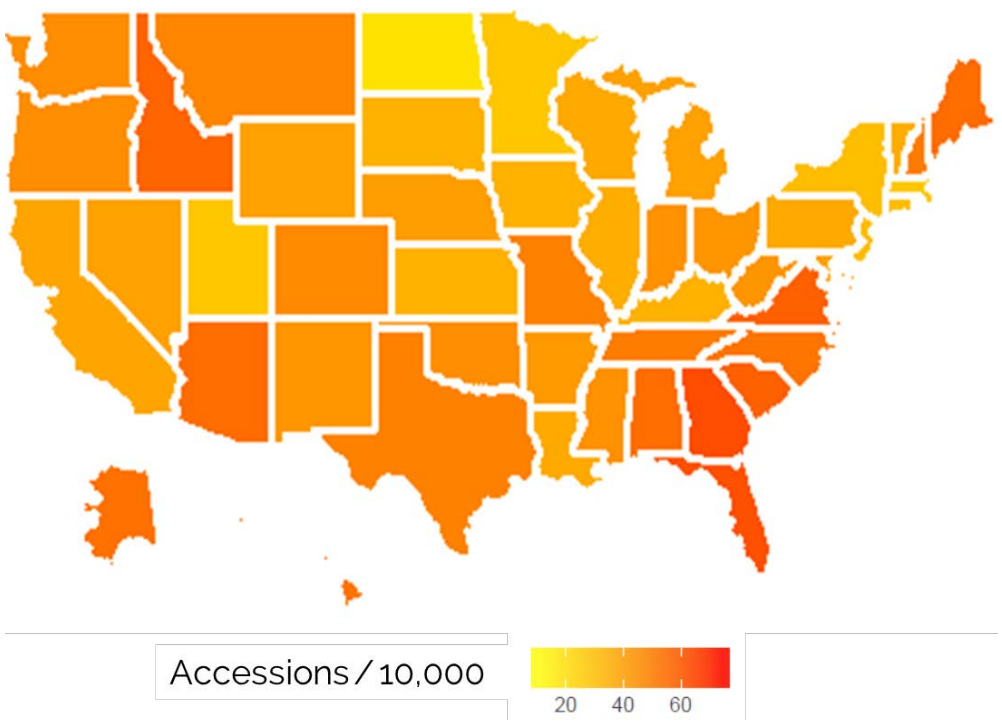


Figure A-7. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2013

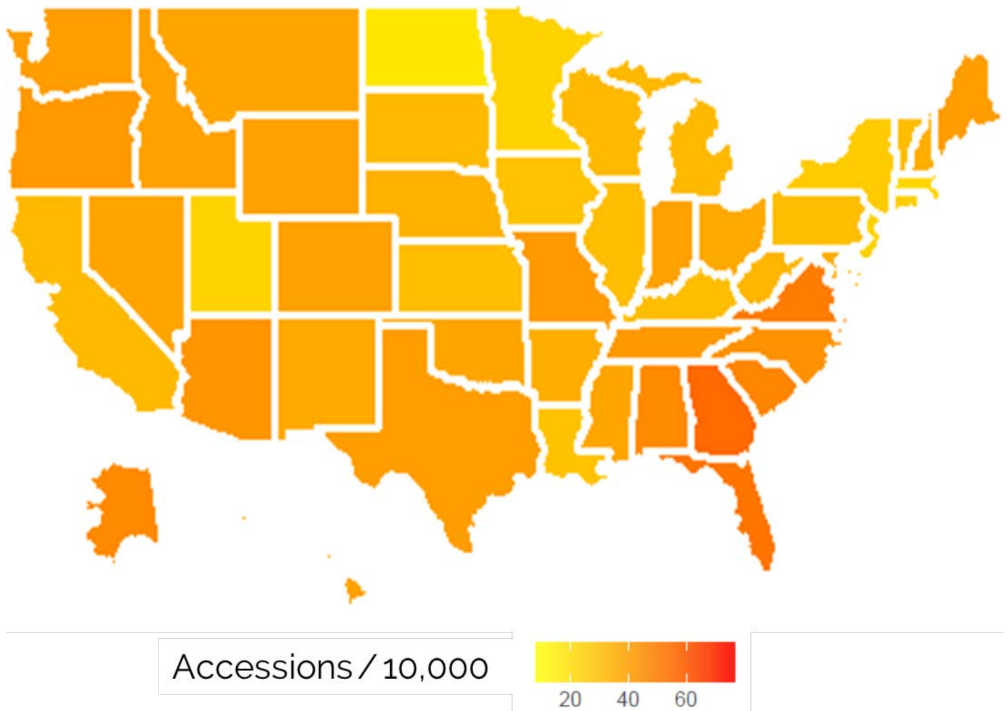


Figure A-8. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2014

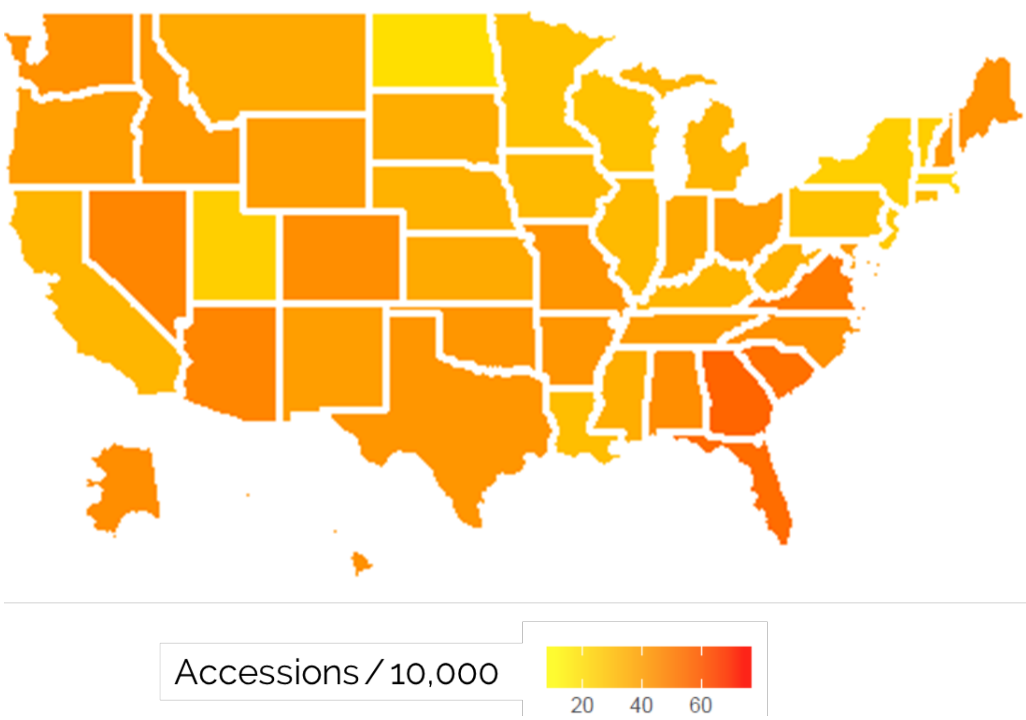


Figure A-9. Non-Prior-Service Enlisted Accessions per 10,000 Youths Aged 18–24, FY 2015

Appendix B.
Data Dictionary

Notes appear at the bottom of this table.

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
Pseudo-county	IDA-generated	N/A	Aggregation level	IDA-generated level of geographic aggregation, encompassing all counties within a PUMA when size of county < size of PUMA, or vice versa when size of PUMA < size of county	Y	2006-2015
year	N/A	N/A	Aggregation level	[Calendar] Year of record	N	2006-2015
RegTotal	DMDC	Individual	Response variable	(R): Number of total enlisted NPS recruits / Population age 17-24	Y	2001-2016
RegMale	DMDC	Individual	Response variable	(R): Number of male enlisted NPS recruits / Population age 17-24	Y	2001-2016
RegFemale	DMDC	Individual	Response variable	(R): Number of female enlisted NPS recruits / Population age 17-24	Y	2001-2016
HQTotal	DMDC	Individual	Response variable	(R): Number of High-Quality total enlisted NPS recruits / Population age 17-24	Y	2001-2016
HQMale	DMDC	Individual	Response variable	(R): Number of High-Quality male enlisted NPS recruits / Population age 17-24	Y	2001-2016
HQFemale	DMDC	Individual	Response variable	(R): Number of High-Quality female enlisted NPS recruits / Population age 17-24	Y	2001-2016

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
ASVBAavailableHS_perpop	ASVAB Career Exploration Program	Zip Code	Covariate	(R): No. HS students available for ASVAB / Population age 17-24	N	2003-2017
ASVABinstitutions_perpop	ASVAB Career Exploration Program	Zip Code	Covariate	(R): No. institutions administering ASVAB CEP / Population age 17-24	N	2003-2017
ASVABtestedALL_perpop	ASVAB Career Exploration Program	Zip Code	Covariate	(R) No. all taking ASVAB CEP/ Population age 17-24	N	2003-2017
ASVABtestedHS_perpop	ASVAB Career Exploration Program	Zip Code	Covariate	(R) No. HS students taking ASVAB CEP / Population age 17-24	N	2003-2017
avg_hhincome_adjusted	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(#): Average household income weighted by total households in a pseudo-county and adjusted by CPI	N	2006-2015
pct_mortgagetohhincome	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Annual mortgage payment, weighted by total households in a pseudo-county / Annual household income, weighted by total households in a pseudo-county	Y	2006-2015
avgnetprice	Integrated Postsecondary Education System Data	County	Covariate	(#): Weighted average net price of postsecondary schools	N	2006-2015
jrotccadets_perpop	Junior Reserve Officer's Training Corps	Zip Code	Covariate	(R): No. JROTC cadets / Population age 17-24	N	2006-2015
jrotcschools_perpop	Junior Reserve Officer's Training Corps	Zip Code	Covariate	(R): No. JROTC schools / Population age 17-24	N	2006-2015
p_est_vehicles0_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of households with no vehicles / Total households	N	2006-2015
p_est_vehicles1_2_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of households with 1-2 vehicles / Total households	N	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
p_est_vehicles_3plus_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of households with 3 or more vehicles / Total households	N	2006-2015
p_est_building_sfh_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households classified as single family homes / Total households	N	2006-2015
p_est_building_mobilehome_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households classified as mobile homes / Total households	N	2006-2015
p_est_building_aptlt10_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households in an apartment with less than 10 units / Total households	N	2006-2015
p_est_building_apt10_19_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households in an apartment with 10-19 units / Total households	N	2006-2015
p_est_building_apt20_49_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households in an apartment with 20-49 units / Total households	N	2006-2015
p_est_building_apt50plus_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households in an apartment with 50 or greater units / Total households	N	2006-2015
p_est_educ_LThsdg_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with less than a high school degree / Total population	N	2006-2015
p_est_educ_hsdg_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with a high school degree or GED ONLY / Total population	N	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
p_est_educ_collenrollees_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of individuals currently enrolled in undergraduate college / Total population	N	2006-2015
p_est_educ_collgrad_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with a bachelor's degree ONLY / Total population	N	2006-2015
p_est_educ_collplus_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with schooling above the level of a bachelor's degree (master's degree, professional degree, or doctorate) / Total population	N	2006-2015
p_est_educ_somecollplus_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with some college (less than 1 year, more than 1 year but less than a degree, or an associate's degree) / Total population	N	2006-2015
p_est_commute_lt30min_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with commute time less than 30 minutes / Total population	N	2006-2015
p_est_commute30_60min_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with commute time 30-60 minutes / Total population	N	2006-2015
p_est_commute60_90min_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with commute time 60-90 minutes / Total population	N	2006-2015
p_est_commute90_200min_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with commute time 90-200 minutes / Total population	N	2006-2015
p_est_workstatus_FTworkHH_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households where the household head worked full-time in the past 12 months / Total households	N	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
p_est_gphh_noparent_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households that are grandparent-headed with no parent present / Total households	N	2006-2015
p_est_gplivewith_acs	American Community Survey Public-Use Microdata Sample Population Record	PUMA	Covariate	(%): Number of people with grandparents living in the house / Total population	N	2006-2015
p_est_hh65plus1_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with one resident over age 65 / Total households	N	2006-2015
p_est_hh65plus2ormore_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with 2 or more residents over age 65 / Total households	N	2006-2015
p_est_hhbrd0_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with no bedrooms (studio or efficiency) / Total households	N	2006-2015
p_est_hhbrd1_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with one bedroom / Total households	N	2006-2015
p_est_hhbrd2_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with 2 bedrooms / Total households	N	2006-2015
p_est_hhbrd3_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with 3 bedrooms / Total households	N	2006-2015
p_est_hhbrd4_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with 4 bedrooms / Total households	N	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
p_est_hhbrd5plus_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with 5 or more bedrooms / Total households	N	2006-2015
p_est_hhvacstatus_hhforRent_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households for rent / Total Households	N	2006-2015
p_est_hhvacstatus_hhforSale_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households for sale / Total Households	N	2006-2015
p_est_hhvacstatus_seasonalHH_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with seasonal occupancy / Total Households	N	2006-2015
p_est_hhvacstatus_notOccdHH_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of unoccupied households / Total Households	N	2006-2015
p_est_hhkidsunder6andover6_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with children under 6 years of age ONLY / Total Households	N	2006-2015
p_est_hhkids6_17_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with children aged 6-17 ONLY / Total Households	N	2006-2015
p_est_hhkids_lt6_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households with children under 6 years old and 6-17 years old / Total households	N	2006-2015
p_hh_language_nonEngspeaker_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of non-English-speaking households / Total households	N	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
p_hh_language_Spanspeaker_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of Spanish-speaking households / Total households	N	2006-2015
p_est_payment_hhownednomort_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households owned with a mortgage / Total households	N	2006-2015
p_est_payment_hhownedmort_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households owned without a mortgage (mortgage paid off) / Total households	N	2006-2015
p_est_payment_hhrent_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households who rent / Total households	N	2006-2015
p_est_payment_hhfree_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households occupied without payment of rent / Total households	N	2006-2015
p_est_hhtype_marriedhh_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households led by a married couple / Total households	N	2006-2015
p_est_hhtype_nonfamHHfemale1_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of nonfamily households: Female householder: Living alone / Total households	N	2006-2015
p_est_hhtype_nonfamHHfemale2_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of nonfamily households: Female householder: not living alone / Total households	N	2006-2015
p_est_hhtype_nonfamHHmale1_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of nonfamily households: Male householder: Living alone / Total households	N	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
p_est_hhtype_nonfamHHmale2_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of nonfamily households: Male householder: not living alone / Total households	N	2006-2015
p_est_hhtype_othfamHHfemale_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of other family households: Female householder, no husband present / Total households	N	2006-2015
p_est_hhtype_othfamHHmale_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of other family households: Male householder, no wife present / Total households	N	2006-2015
p_est_workstatus_PTworkHH_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households where the householder worked less than full-time in the past 12 months / Total households	N	2006-2015
p_est_worksstatus_unemplHH_acs	American Community Survey Public-Use Microdata Sample Housing Record	PUMA	Covariate	(%): Number of households where the householder did not work in the past 12 months / Total households	N	2006-2015
pct_obese_age_adjusted_cdc_weigh	Centers for Disease Control County-level files	County	Covariate	(%): Weighted mean age-adjusted percent of individuals identified as obese, weight = numberObese_cdc	N	2005-2013
pct_govemp_totemp	Quarterly Census of Employment and Wages	County	Covariate	(%): Weighted mean percent of employees employed by the government, weight = total population	N	1994-2016
pct_threepriemp_totemp	Quarterly Census of Employment and Wages	County	Covariate	(%): Weighted mean percent of employees employed by the private employers, weight = total population	N	1994-2016
pct_gulfwar1vets_acs_weighted_me	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans who served in the first Gulf War / Veteran population ages 18+	N	2006-2015
pct_gulfwar2vets_acs_weighted_me	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans who served in the second Gulf War / Veteran population ages 18+	N	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
pct_vietnamwarvets_acs_weighted_	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans who served in the Vietnam War / Veteran population ages 18+	N	2006-2015
pct_koreanwarvets_acs_weighted_m	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans who served in the Korean War / Veteran population ages 18+	N	2006-2015
pct_ww2vets_acs_weighted_mean	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans who served in World War II / Veteran population ages 18+	N	2006-2015
perpop_enroll_grantdegree	Integrated Postsecondary Education System Data	County	Covariate	(R): Total enrollment in degree-granting institutions / 10K population age 17-24	N	2006-2015
perpop_grantdegree	Integrated Postsecondary Education System Data	County	Covariate	(R): No. degree-granting institutions / 10K population age 17-24	N	2006-2015
perpop_num_postsec	Integrated Postsecondary Education System Data	County	Covariate	(R): No. all postsec schools / 10K population age 17-24	N	2006-2015
perpop_undupug	Integrated Postsecondary Education System Data	County	Covariate	(R): 12-month unduplicated undergraduate enrollment / 10K population age 17-24	N	2006-2015
popdensity	IDA-generated	County	Covariate	(R): Total population / Land area	Y	2006-2015
popperhousehold	IDA-generated	County	Covariate	(R): Total population / Total households	Y	2006-2015
ratio_vetdisab_nonvetdisab	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who are disabled / Percent of nonveterans who are disabled	Y	2006-2015
ratio_vetincome_nonvetincome	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Veteran annual average income (adjusted to real values) / Nonveteran annual average income (adjusted to real values)	Y	2006-2015
ratio_vetpoverty_nonvetpoverty	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who are living in poverty / Percent of nonveterans who are living in poverty	Y	2006-2015
ratio_vetunemp_nonvetunemp	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who are unemployed / Percent of nonveterans who are unemployed	Y	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
aland9	IDA-generated	PUMA	Covariate	Land area	Y	2006-2015
ratio_collplus_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who have a college-level education or higher / Percent of nonveterans who have a college-level education or higher	Y	2006-2015
ratio_maleinc_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Veteran annual median income among males (adjusted to real values) / Nonveteran annual median income among males (adjusted to real values)	Y	2006-2015
ratio_medinc_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Veteran annual median income (adjusted to real values) / Nonveteran annual median income (adjusted to real values)	Y	2006-2015
ratio_femaleinc_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Veteran annual median income among females (adjusted to real values) / Nonveteran annual median income among females (adjusted to real values)	Y	2006-2015
ratio_malepop_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Male veteran population ages 18+ / Male nonveteran population ages 18+	Y	2006-2015
ratio_femalepop_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Female veteran population ages 18+ / Female nonveteran population ages 18+	Y	2006-2015
ratio_whitenothisp_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who are White and Not Hispanic / Percent of nonveterans who are White and Not Hispanic	Y	2006-2015
ratio_white_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who are White (including Hispanic) / Percent of nonveterans who are White (including Hispanic)	Y	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
ratio_somecoll_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who have completed some college or an associate's degree / Percent of nonveterans who have completed some college or an associate's degree	Y	2006-2015
ratio_hsgrad_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who are high school graduates (includes equivalency) / Percent of nonveterans who are high school graduates (includes equivalency)	Y	2006-2015
ratio_lths_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who completed less than a high school degree / Percent of nonveterans who completed less than a high school degree	Y	2006-2015
ratio_lfparticp_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who are in the labor force / Percent of nonveterans who are not in the labor force	Y	2006-2015
ratio_bachplus_vet_nonvet	American Community Survey FactFinder Table S2101	PUMA	Covariate	(R): Percent of veterans who completed a bachelor's degree or higher / Percent of nonveterans who completed a bachelor's degree or higher	Y	2006-2015
pct_vet_age18plus	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans aged 18 and over / Total population ages 18+	N	2006-2015
pct_vet_age18_34	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans ages 18-34 / Veteran population ages 18+	N	2006-2015
pct_vet_age35_54	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans ages 35-54 / Veteran population ages 18+	N	2006-2015
pct_vet_age55_64	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans ages 55-64 / Veteran population ages 18+	N	2006-2015

Field Name	Source Dataset	Original Level of Geographic Aggregation	Variable Type	Description	IDA-Generated?	Years Available
pct_vet_age65_74	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans ages 65-74 / Veteran population ages 18+	N	2006-2015
pct_vet_age75plus	American Community Survey FactFinder Table S2101	PUMA	Covariate	(%): Number of veterans ages 75 and over / Veteran population ages 18+	N	2006-2015

Notes: The symbol (%) indicates that the variable is a percent estimate.
The symbol (#) indicates that the variable is a numeric (count) estimate.
The symbol (R) indicates an IDA-generated ratio based on existing variables.

Note: High-quality recruits are those who score as Cat 1-3A on the AFQT and are high school graduates or seniors.

Illustrations

Figures

Tables

References

Bureau of Labor Statistics. "Civilian Labor Force Participation Rate by Age, Gender, Race, and Ethnicity." October 2017. https://www.bls.gov/emp/tables/civilian-labor-force-participation-rate.htm.

Defense Manpower Commission. Staff Studies and Supporting Papers, Volume III. *Military Recruitment and Accessions and Future of the All-Volunteer Force,* Appendix D. May 1976. https://www.dtic.mil/docs/citations/ADA029951.

Department of Defense. *Population Representation in the Military Services*, various years. http://www.cna.org/research/pop-rep.

Goldberg, Lawrence, Dennis D. Kimko, and Maggie X. Li. "Analysis and Forecasts of Army Enlistment Supply." IDA Document NS D-5466. Alexandria, VA: Institute for Defense Analyses, April 2015.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning, with Applications in R*. Heidelberg, Germany: Springer, 2015.

LaCommare, Kristina Hamachi, and Joseph H. Eto. "Understanding the Cost of Power Interruptions to U.S. Electricity Consumers." LBNL-55718. Berkeley, CA: Ernest Orlando Lawrence Berkeley National Laboratory, University of California Berkeley, September 2004. http://eta-publications.lbl.gov/sites/default/files/lbnl-55718.pdf.

National Center for Education Statistics. "Fast Facts." https://nces.ed.gov/fastfacts/display.asp?id=40.

"Package 'randomforest'." March 25, 2018. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.

Public Law 112-239, National Defense Authorization Act for Fiscal Year 2013, Section 553. Modification of requirements on plan to increase the number of units of the Junior Reserve Officers' Training Corps.

Rostker, Bernard. *I Want You! The Evolution of the All-Volunteer Force*. MG-265-RC. Santa Monica, CA: RAND Corporation, 2006. https://www.rand.org/pubs/monographs/MG265.html.

Secretary of Defense Ashton Carter, "Forging Two New Links to the Force of the Future," Memorandum. November 1, 2016. http://www.defense.gov/Portals/1/Documents/pubs/Forging-Two-New-Links-Force-of-the-Future-1-Nov-16.pdf.

Senate Committee on Armed Services. S. Rep. No. 93-884 (1974).

Syeed, Nafeesa, and Chloe Whiteaker. “Low U.S. Unemployment Is Making Army Recruiting Harder.” *Bloomberg Politics*, March 13, 2018. https://www.bloomberg.com/news/articles/2018-03-13/trump-s-army-buildup-confronts-headwinds-of-tight-labor-market.

United States Census Bureau. “American Community Survey (ACS): Top Questions about the Survey.” https://www.census.gov/programs-surveys/acs/about/top-questions-about-the-survey.html.

United States Census Bureau. “History: Regions and Divisions.” https://www.census.gov /history/www/programs/geography/regions_and_divisions.html.

Walls, Tyrone, Major, United States Marine Corps. “Junior Reserve Officers’ Training Corps: A Comparison with Other Successful Youth Development Programs and an Analysis of Military Recruits who Participate in JROTC.” Master’s thesis, Naval Postgraduate School, June 2003. www.dtic.mil/docs/citations/ADA417589.

Abbreviations

ACS	American Community Survey
AFQT	Armed Forces Qualification Test
AP	Accession Policy
ASVAB	Armed Services Vocational Aptitude Battery
AVF	All-Volunteer Force
BLS	Bureau of Labor Statistics
BMI	Body Mass Index
CDC	Centers for Disease Control and Prevention
CEP	Career Exploration Program
DMDC	Defense Manpower Data Center
DoD	Department of Defense
FY	Fiscal Year
HUD	US Department of Housing and Urban Development
IDA	Institute for Defense Analyses
IPEDS	Integrated Postsecondary Education Data System
JROTC	Junior Reserve Officers' Training Corps
M&RA	Manpower and Reserve Affairs
MPP	Military Personnel Policy
NCES	National Center for Education Statistics
NPS	Non-Prior-Service
OCS	Officer Candidate School
OUSD(P&R)	Office of the Under Secretary of Defense for Personnel and Readiness
PUMA	Public Use Microdata Area
QCEW	Quarterly Census of Employment and Wages
ROTC	Reserve Officers' Training Corps
UI	Unemployment Insurance
US	United States
USMEPCOM	United States Military Entrance Processing Command
ZIP	Zone Improvement Program

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
xx-11-2018	Final	Aug 2016 - Jul 2018

4. TITLE AND SUBTITLE
Geographic Diversity in Military Recruiting

5a. CONTRACT NUMBER
HQ0034-14-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER
BE-7-4213

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Goldberg, Matthew, S.
Cheng, Karen
Huff, Nancy, M.
Kimko, Dennis, D.
Saizan, Alexandra, M.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Paper P-9079
H 18-000184

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Director for Accession Policy, Office of the Assistant Secretary of Defense (Manpower and Reserve Affairs), Office of the Under Secretary of Defense (Personnel and Readiness)
The Pentagon, Room 3D1066
Washington, DC 20301

10. SPONSOR/MONITOR'S ACRONYM(S)
OUSD/M&RA/P&R

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The US military attempts to recruit personnel from all demographic, economic, and cultural segments of the nation's youth population. However, since the beginning of the all-volunteer military over 40 years ago, certain regions of the nation—especially the South—have supplied disproportionate numbers of recruits relative to the size of their youth population. This paper probes more deeply than just regional identification and summary economic indicators such as a state's youth unemployment rate. The supporting research uses machine learning methods (in particular, regression trees) to identify the demographic, economic, and cultural features of individual counties (or small collections of contiguous counties) that are most influential in predicting non-prior-service enlisted accessions. Among the most influential predictors are various measures of the proportion of veterans in the county's population. Other important predictors are the percentage of people attending college, the percentage who have completed some college (but not a bachelor's degree), the fraction of high school students available to take the Armed Services Vocational Aptitude Battery, and the prevalence of Junior Reserve Officers' Training Corps programs.

15. SUBJECT TERMS
military; recruiting; accession; geographic diversity; machine learning; regression tree

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Drogo, Dennis, J
Unclassified	Unclassified	Unclassified	Same as Report	103	19b. TELEPHONE NUMBER *(Include area code)* (703) 697-9268

Reset