Methods and data sources

This analysis follows our 2017 report, *Meet the out-of-work: Local profiles of jobless adults and strategies to connect them to employment, which sought to sort local out-of-work populations into groups useful to stakeholders in workforce development. For the most part, this analysis uses the same methods and data as in our earlier report. The major difference is that this analysis focuses on young adults, ages 18 to 24, whereas the prior analysis focused on the population age 25 to 64. Here we provide an overview of the methods and data; for detailed documentation and notes about how we developed these methods, see the 2017 report's technical appendix*.

Data and sample geographies

This analysis is based on 2013–2015 three-year American Community Survey (ACS) Public-Use Microdata Samples data. The U.S. Census Bureau ceased production of three-year ACS products in fall 2015, so we constructed our own three-year dataset by pooling single-year data for 2013, 2014, and 2015, and adjusting weights according to annual change in population using Population Estimates Program county population totals. All nominal dollars were converted to 2015 dollars.

In the prior report, we focused on 130 large jurisdictions (counties, cities, and county remainders net of large cities) that could be constructed neatly from 2010 Public-Use Microdata Areas (PUMAs) and provided sufficient sample size for our analysis. In this report we started from the same list of 130 jurisdictions and dropped those with samples of fewer than 150 unweighted observations of "out-of-work" individuals, as defined below. Consequently, we ended up with 119 jurisdictions with sample populations ranging from 155 out-of-work individuals (Seattle, Wash.) to 3,249 (Los Angeles County, Calif., net of Los Angeles city). (The numbers in the previous sentences refer to unweighted observations from the sample, not the estimated number of out-of-work individuals in a given jurisdiction.)

Sample population

The sample population in this analysis differs from our earlier report only in that it focuses on out-of-work young adults, ages 18 to 24. Otherwise, we operationalized "out-of-work" in the same way. We started with a sample of 155,400 unweighted observations of 18-to-24-year-olds, across the 119 sample geographies, who are not employed (they may be either "unemployed" or "not in the labor force"). We then subtracted the following groups:

- "Traditional" students, defined as all those not in the labor force (92,874), college students living in dormitories (28,297), and high school students living at home (18,392). We also subtract graduate and professional students (4,855), who may be "on the right track" to employment; 98,321 observations in all were dropped.
- Individuals receiving Supplemental Security Income, Social Security, or other disability or retirement income. Using these ACS variables, we are unable to distinguish disability benefits from retirement or survivor benefits; consequently, we

may have dropped out-of-work young adults receiving survivor benefits who are able to work; 7,329 observations were dropped.

- Stay-at-home parents, defined as married persons not in the labor force, with children, whose spouse is present and employed, and with family incomes at least twice the federal poverty level; 741 observations dropped.

The above categories are not mutually exclusive, and many individuals fall under more than one. The remaining 52,427 unweighted sample observations are the out-of-work young adult population used in this analysis.

Cluster analysis

Cluster analysis is a broad, flexible set of methods to create meaningful groups of similar objects (in this case, individual people) based upon user-defined characteristics.

Consistent with our earlier analysis, we first clustered out-of-work individuals within individual jurisdictions, and then sorted the resulting jurisdiction-level groupings across all sample geographies into "major groups." In both stages of the analysis, the methods used differ from those in our earlier analysis only in the list of clustering variables. The goal, as before, was to produce groupings useful to local workforce stakeholders.

In the first-stage, jurisdiction-level cluster analysis, we clustered observations within each jurisdiction based on the following variables:

- Years of education,
- Level of education completed (high school or less, some college or associate degree, bachelor's degree or higher),
- Whether the individual is a racial/ethnic minority,
- Whether the individual is an English language-learner,
- Age,
- Presence and severity of any disabilities,
- Whether the individual has children,
- Whether the individual worked in the past year,
- Ratio of the individual's family income to the poverty threshold, and
- Whether the individual is enrolled in school

Because the above variables are a mix of categorical and continuous measures, we used Gower's dissimilarity measure as our similarity measure. We used complete-linkage hierarchical clustering, and initially selected the clustering solution for each jurisdiction with the highest Calinski-Harabasz statistic, allowing for three to five groups. In 21 jurisdictions we used a solution with the second- or third-highest Calinski-Harabasz statistic because they were more practically meaningful, for example by splitting a large group of moderately educated individuals into two based on English language proficiency. In 11 of the 119 jurisdictions we relaxed the clustering criteria slightly by dropping a clustering variable (ratio of family income to the poverty threshold). Finally, we dropped two clusters with small samples. A total of 38 unweighted sample observations out of 52,427, across 12 of the 119 study jurisdictions, were not successfully assigned to clusters; these unassigned observations are reflected as "Cluster ." in the downloadable data tables. In all, we ended up with 557 jurisdiction-level clusters across all 119 sample geographies.

In the second stage, we aggregated the 557 jurisdiction-level clusters identified in the first step into *groups of groups. In this instance we used the following clustering variables:*

- Share of cluster population with a high school diploma or less,
- Share with some college completed or an associate degree,
- Share with a bachelor's degree or higher,
- Share that are racial/ethnic minorities,
- Share that are English language-learners,
- Share that are 18 to 21 years-old,
- Share that are 22 to 24 years-old,
- Share reporting any form of disability,
- Share with children,
- Average family income, and
- Share enrolled in school

We standardized all of these variables by their range, and used Ward's linkage hierarchical clustering, allowing for three to eight groups; we selected the solution with the highest Calinski-Harabasz statistic. We ended up with five of these "major groups;" for simplicity's sake, we refer to these five major groups as "clusters."