

# Executive Summary

May 2018

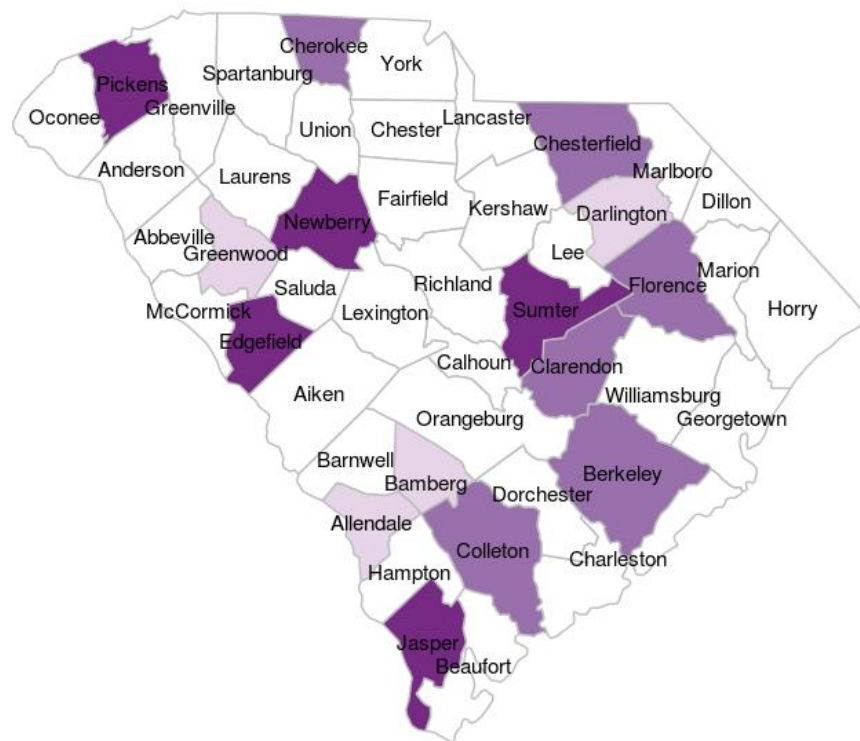
## Identifying County Outliers in the South Carolina Incident-Based Reporting System (SCIBRS): 2011–2015 Intimate-Partner Violent Victimization (IPVV)

### RESULTS AT A GLANCE

### SCIBRS 2011–2015 Intimate-Partner Violent Victimization

#### Outliers

- Allendale
- Bamberg
- Berkeley
- Cherokee
- Chesterfield
- Clarendon
- Colleton
- Darlington
- Edgefield
- Florence
- Greenwood
- Jasper
- Newberry
- Pickens
- Sumter



Outlier Identified by  
■ Both Analyses ■ Stonewall Analytics ■ SC SAC

The authors are grateful for funding received through the United States Department of Justice (DOJ), Office of Justice Programs (OJP), Bureau of Justice Statistics (BJS) Grant Number 2016-BJ-CX-K022. Any views expressed in this research—including those related to statistical, methodological, technical, or operational issues—are solely those of the authors and do not necessarily reflect the official position or policies of the DOJ, OJP, BJS, or the South Carolina Statistical Analysis Center located in the South Carolina Department of Public Safety's Office of Highway Safety and Justice Programs.

## PROJECT BACKGROUND

The South Carolina Law Enforcement Division (SLED) manages the South Carolina Incident-Based Reporting System (SCIBRS), which is National Incident-Based Reporting System (NIBRS)-certified by the Federal Bureau of Investigation (FBI). The SCIBRS stems from approximately 275 law enforcement agencies reporting information about victims, offenses, offenders, and arrestees (if applicable) for all criminal incidents known to police. SLED provides support to agencies through auditing, training, and guidance on coding individual incidents. SLED also stores every incident submitted by the law enforcement agencies on a state repository, submitting those same incidents to the FBI.

The SCIBRS can be used to study a variety of criminal justice subjects. The South Carolina Governor's Domestic Violence Task Force identified it as the primary source for domestic violence data. Accordingly, the integrity of the SCIBRS data must be ensured: quality data best guides policy and the distribution of resources for criminal justice agencies, government institutions, and nonprofit organizations in their mission to aid domestic violence victims. As a NIBRS-certified system, its crimes are categorized by general definitions; thus, the SCIBRS provides a unique opportunity to study domestic violence across jurisdictions—independent of statutory differences.

Because limited resources render it infeasible for SLED to visit all 275 reporting agencies, the South Carolina Statistical Analysis Center (SC SAC) designed a three-phase research project to aid SLED by statistically guiding its data integrity efforts. This project is the first of its kind in the nation. It develops methodologies that can be shared with uniform crime reporting programs (UCRP) throughout the United States, while also improving lives of domestic violence survivors in South Carolina. It is an attempt to establish a methodology by which UCRP might readily identify counties that are likely to have data quality issues. If successful, this methodology could be used by all such programs to allow for a more focused and efficient assessment of data quality.

In the first phase, the SC SAC's project focuses on a specific kind of domestic violence: intimate-partner violent victimization (IPVV) as recorded in the SCIBRS from 2011 through 2015. The SC SAC defines SCIBRS IPVV as any crime in which (1) the victim recorded in the SCIBRS is an intimate partner of at least one of the offenders (i.e., being in one of the following relationships: spouse, ex-spouse, common-law spouse, (ex-)boy/girlfriend, same-sex relationship), and (2) the offense recorded in the SCIBRS is a crime that the FBI includes in its Violent Crime Index (i.e., murder, sexual battery, robbery, aggravated assault).

During the first and current phase, two methods (those of the SC SAC and Stonewall Analytics <https://www.stonewallanalytics.com/>) are used to identify counties with SCIBRS IPVV that is outside expectation (i.e., either higher or lower than expected). These counties are deemed "outliers", although the cause is not determined or sought. A list of outlier counties will guide SLED during the second phase of the project in its refinement to an agency-level review. As SLED reviews agencies, the result will be a real-world comparison of methods to see which was more successful in pointing toward data issues. During the third phase, reporting agencies in South Carolina will benefit from SLED's resulting educational outreach.

## METHOD OF THE SOUTH CAROLINA STATISTICAL ANALYSIS CENTER

The first method for outlier identification is simple: the SC SAC applies descriptive statistics to SCIBRS IPVV rates. Using rates normalizes data. The SC SAC calculates IPVV rates per 10,000 people that pool victims and population over five years from 2011 through 2015. Counties with the highest and two lowest rates are included. The SC SAC groups the counties into five levels of SCIBRS IPVV rates and then maps them: any counties on the high or low end of IPVV rates that are also geographically “out-of-place”—meaning that they are surrounded by counties with different levels of rates—are included. Finally, the percent change in single-year IPVV rates (using 2011 and 2015) is examined: any counties that decrease more than 60% and drop by an IPVV rate of at least 15 victims per 10k are included. Counties with a percent change that increases are not included because they involve small rate increases; e.g., 4 victims per 10k. NB: the aforementioned “cutoffs” are not statistically-determined.

## METHOD OF STONEWALL ANALYTICS

The second method for outlier identification is more complex: Stonewall Analytics works with a supervised machine learning model—the random forest model. Random forest models tend to avoid model overfitting; are easy to interpret; and are completed quickly, even on large data sets.

The model is trained, tested, and validated with 2011 national aggravated assault arrest data from the Inter-University Consortium for Political and Social Research (ICPSR). Building upon prior research about intimate-partner violence, Stonewall Analytics incorporates county-level socioeconomic data from the American Community Survey (ACS) of the United States Census Bureau into the modeling. Ultimately, the random forest model predicts which counties are outliers; for this analysis, outliers are defined as counties that have IPVV counts of victims that are further than one standard deviation from what the model predicts would be the case for the county for at least three separate years ranging from 2011 through 2015.

Because the training/testing/validating ICPSR data does not have a high frequency of cases extending beyond 200 arrests per county, the SCIBRS IPVV data for each county also need to be restricted. Doing so improves the model’s ability to predict what the expected SCIBRS IPVV count of victims is for each county and year. Excluding counties with at least 141 victims (the 75<sup>th</sup> percentile) removes the following counties from the analysis: Anderson, Beaufort, Charleston, Greenville, Horry, Lexington, Richland, Spartanburg, and York.

The following ACS variables are used in the analysis. They are listed in order of decreasing importance, as they relate to the ICPSR data. Emphasis indicates the five most influential variables for ICPSR data.

***Receipt of supplemental security income (SSI); black race; median household income in past 12 months (inflation-adjusted dollars); marital status; white race;*** women who gave birth in the past 12 months; Hispanic; educational level of bachelor’s degree or higher; divorce; whether male aged 16 to 64 worked in the past 12 months; total working population; total population; sex; poverty status in past 12 months of families by household type by tenure; whether female aged 16 to 64 worked in the past 12 months; those who drive to work; average household size of occupied housing units by tenure; presence of own children under 18 for females aged 20 to 64; Gini Index (measure of wealth inequality). The model accounts for 45.6% of variability in ICPSR data.