Journal of Information Systems Applied Research

Volume 13, Issue 3
November 2020
ISSN: 1946-1836

In this issue:

iscap
Information Systems & Computing Academic Professionals

The **Journal of Information Systems Applied Research** (JISAR) is a double-blind peer reviewed academic journal published by ISCAP, Information Systems and Computing Academic Professionals. Publishing frequency is three issues a year. The first date of publication was December 1, 2008.

JISAR is published online (http://jisar.org) in connection with CONISAR, the Conference on Information Systems Applied Research, which is also double-blind peer reviewed. Our sister publication, the Proceedings of CONISAR, features all papers, panels, workshops, and presentations from the conference. (http://conisar.org)

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the JISAR journal. Currently the target acceptance rate for the journal is about 40%.

Questions should be addressed to the editor at editor@jisar.org or the publisher at publisher@jisar.org. Special thanks to members of EDSIG who perform the editorial and review processes for JISAR.

Journal of Information Systems Applied Research

Understanding Campus Crime with A Multi-University Analytics System

Uche Iheadindu
MS Computer Science and Information Systems

Douglas Kline
klined@uncw.edu
Information Systems

Ron Vetter
vetterr@uncw.edu
Computer Science

Ulku Clark
clarku@uncw.edu
Information Systems

University of North Carolina Wilmington
Wilmington NC 28403

Abstract

Due to budget challenges, the campus police department of University of North Carolina Wilmington engaged in a data-driven performance management effort. To support this effort, publicly available data from multiple sources was integrated into rigorous data model in a single MS SQL Server database with interactive reporting using MS SQL Server Reporting Services. The data consisted of publicly available crime statistics for 38 universities, as well as characteristics such as Carnegie classification, acreage, budget, number of students, etc. The purpose of the system was to benchmark the campus police department against peer, aspirant, and other similar universities. This paper describes the architecture of the system, the benefits to the police department, and sample analytics.

Keywords: Performance management, data warehouse, campus crime, analytics, frontier analysis, visualization, database management.

1. INTRODUCTION

Like many organizations, the campus police department of University of North Carolina Wilmington faced challenges in budget and resource constraints. In order to make better management decisions about scarce resource allocation, leadership desired usable information upon which to make these decisions. The initial motivation was to benchmark performance of campus police against other “similar” universities. In other words, how is this police department doing, compared with others? However, “similar” can have many meanings:

- Similar in size?, e.g. # students, acreage

- Similar in setting?, e.g. urban, rural
- Similar in type?, e.g. private, public
- Similar in police challenges?
- Similar in budget?

All of these, it turned out, were factors that would be important to the benchmarking effort. Depending on the stakeholder, each of these might be important. No single source of data had all of these data items.

Secondary motivations for the system included being able to make informed data-driven decisions with limited resources, being able to make compelling arguments for additional resources, and perhaps being inspired by and learning from other institutions. Having good data opened possibilities that transcended the initial requirements.

This paper describes the creation of a database integrating data from multiple sources and pertaining to many universities, and the architecture of the system for extracting meaningful information from the data. This should guide developers and decision makers on the challenges encountered in applying analytics to a particular domain.

2. BACKGROUND

Campus Police Department

As at many universities, at UNC Wilmington safety is identified clearly as a goal in the strategic plan. In this way, campus police play a vital role to the organization. The campus police department must not only enforce city, state, and federal laws, but also support the mission of the University. University police are state certified law enforcement officers, carry arms, have full powers of arrest and have the same authority as other state police officers. They work closely with local police agencies. In addition to enforcement, the department educates the university community, and works with and guides the university administration.

The police department's most visible division is the Patrol Division. This is the largest in terms of staff and the most visible to the community. It operates 24 x 7 on foot, bicycles, and in patrol vehicles. In addition to deterrence and intervention, they also provide support such as security checks, car unlocking, and escorts.

The Investigation Division performs follow up on reported crimes, and cooperates with local, state, and federal agencies as necessary. They also provide expert insight to University leadership in matters of crime prevention, substance abuse, sexual assault, and awareness.

The Support Services Division primarily gather, record, and report data for compliance purposes. This division consists of sworn officers and tele-communicators. This division must record and report crime data according to the Clery Act (Government, 2019). This requires all institutions of higher education that have federal support to gather and report specified crime statistics. The resulting data is aggregated and published in the Integrated Postsecondary Education Data System (IPEDS) from the National Center for Education Statistics (NCES) (National Center for Education Statistics, 2012) The annual IPEDS reports included crime statistics as well as enrollment, budget, tuition, accreditation, and many other items.

The IPEDS report was the main source of data for this analytics system. Additional data was gathered from the US Census Bureau (US Federal Government, n.d.) and the Federal Bureau of Investigation (US Federal Bureau of Investigation, 2013). The Census Bureau data provided context for the Metropolitan Statistical Area (MSA), aka, community, in which the universities resided. The FBI data provided the crime statistics for the MSA. Finally, some budget data was gleaned from individual university web sites.

3. SYSTEM ARCHITECTURE

Requirements

Several data challenges presented themselves through the requirements. All data needed to be "citable", with the ability to "drill-back" to the source of the data. Furthermore, data from these sources was aggregated across time, e.g., yearly data. Financial data was typically aggregated for the state fiscal year, enrollment data was aggregated for the academic year, and crime data was aggregated for a calendar year.

A second driving requirement was the need to be able to add new data sources, and new data items. For example, a local police department might also have data that could be integrated into the system in the future. The system had to be flexible enough to accommodate integrations with other data sets in the future.

Finally, the system had to be dynamic, with the ability to interact with the data. In the past, information was presented in periodic static reports. An annual report is quite stale even at the moment of publication. Also, the report is limited to what the authors decided to include. The police department wanted domain-expert police personnel to be able to explore the data interactively.

Platform Selection

The existing data processing occurred in long-lived and ad hoc spreadsheets, and in a home grown single-user Access database. This effectively created data silos and limited the utility of the data. Data quality was a concern due to manual, questionably repeatable data manipulation, and a lack of data integrity checks. Versioning of reports, spreadsheets, and the Access database had become problematic. Backups were ad hoc, manual, and had no consistent policy driven automation.

The decision was made to move to a more enterprise-level architecture. The desired architecture would be a fully transactional relational database server with a reporting server. MS SQL Server was chosen with SQL Server Reporting Services, both to be hosted and managed by the university technology services division. This ensured that servers were being monitored, that data was being backed up along with organizational policy, that the servers were in a data center with redundancy, and that the systems were robust to personnel changes.

Data Model

This system demanded a custom relational model to address the requirements. Central to the model was the ability to have many different types of metrics. Also required was the ability to add metrics as they became available. Figure 1 shows the data model for handling the different types of metrics and their values.

In this model each MetricValue represents a single data item, e.g., the number of sworn officers for a particular university obtained from a particular source, for a certain year with certain year type (yID). The current system has 208 Metric records, with 16,678 distinct MetricValues. Note that each Metric subtyped as a CrimeMetric, ControlMetric, or DiscMetric (discretionary). Discretionary Metrics included items such as budget, number of patrol officers, etc. Control Metrics are items like population of the MSA surrounding the university. Note that the use of subtypes is necessitated mainly by CrimeMetrics being related to other CrimeMetrics.

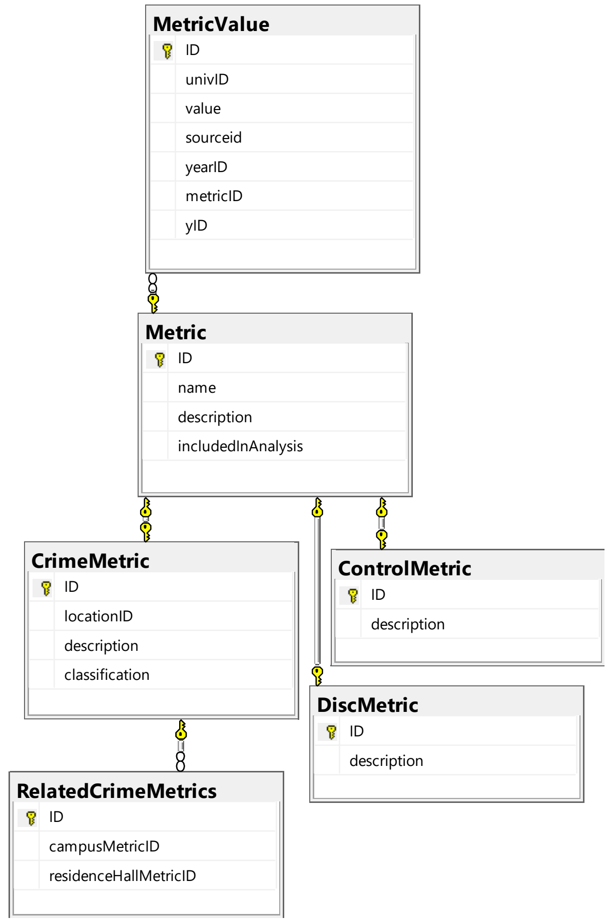


Figure 1: Metric Data Model

The data model in Figure 1 shows how data related to universities is stored. Figure 2 shows how Metropolitan Statistical Area (MSA) data is stored and related to universities. A single MSA can be the home of multiple universities. A refinement of this data model might have MetricValue as the supertype of a UniversityValue and MSAValue, since MetricValue and MSAValues are very similar except for foreign keys to University and MSA, respectively.

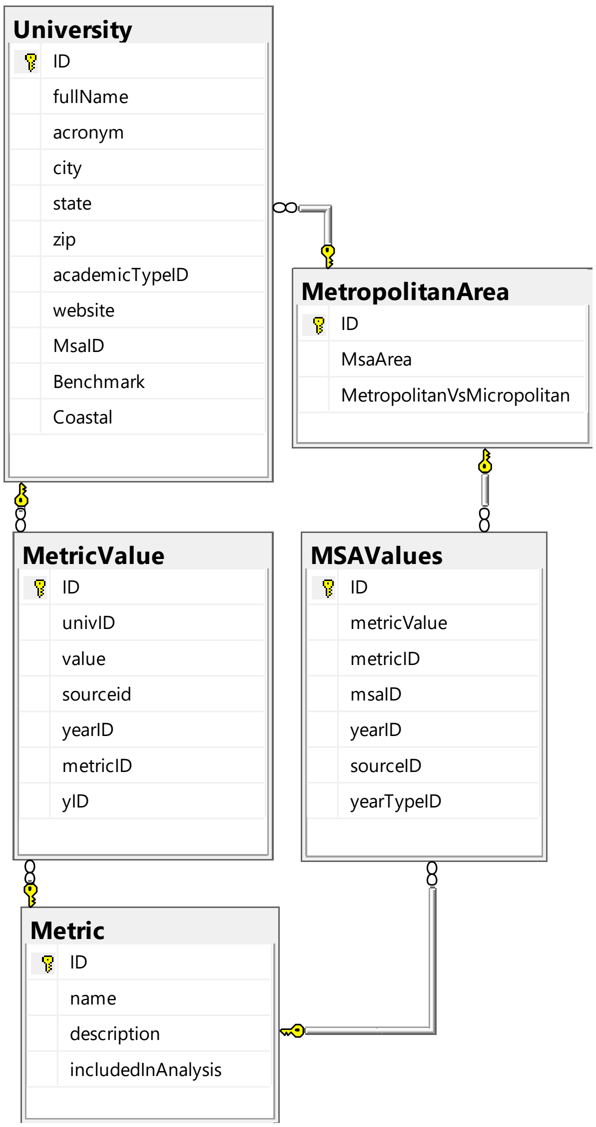


Figure 2: Metric Values Data Model

Crime metrics were quite complicated, with many “qualifications” on each metric. Figure 3 shows the data model for crime metrics. Note that CrimeMetric, also shown in Figure 1 is a subtype of metric, and that HateCrime is a subtype of CrimeMetric, i.e., some crimes are hate crimes and have additional data. Crimes can also be related to other crimes, so CrimeMetric has a many-to-many relationship with itself via RelatedCrimeMetrics. In Location, the location field is actually a location type with these values: Campus, Residence Hall, and OffCampus.

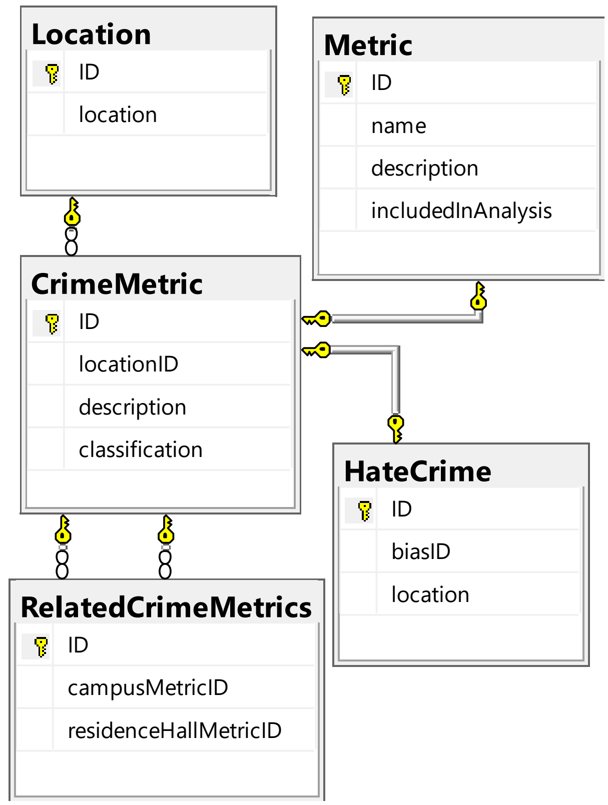


Figure 3: Crime Type and Location Data Model

Reporting Architecture

To better manage the complexity of necessary reporting capabilities, and to present a simpler mental model to users, we extensively used views. Figure 4 shows the flow of data from database tables to finished reports.

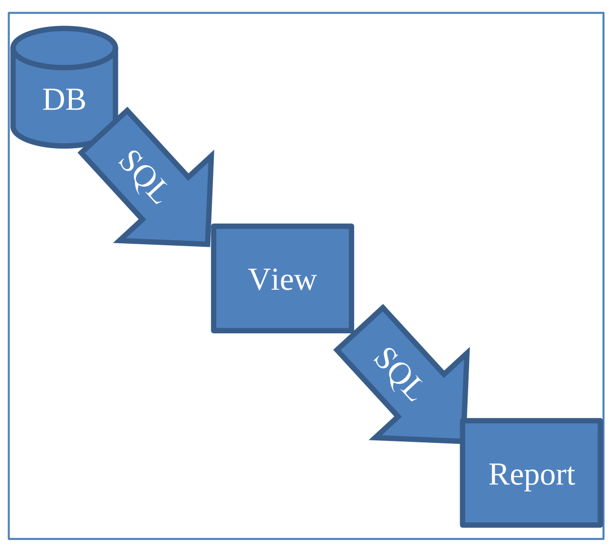


Figure 4

The first SQL statement in the process mainly joins tables appropriately. This eliminates the need for a report writing user to understand the relatively complex data model. For example, a view dealing with CrimeMetrics might need a self-referencing many-to-many outer join. Having a single View that performs this not only promotes code re-use and modularity but saves the report-writer time and potential mistakes.

The second set of SQL in the process might reside in the database server, or in the report itself. This SQL likely limits records with a WHERE clause or aggregates with GROUP BY. In this system, we used parameterized queries, where the parameter values were set via a drop-down or text field in the report and sent as parameters to the SELECT statement.

This design decision to use views was important but can be controversial. Views can have an impact on performance and with heavily nested (a view references another view) use can make the SQL code hard to read. In our situation, the size of the data is not likely to create a performance issue. The readability of the many views must be managed with naming conventions and very descriptive names. The current system has 123 views.

Reporting, Analytics, Visualization
In total 73 reports were designed for various purposes, and deployed using SQL Server Reporting Services (SSRS), where authenticated users could access them via a web page and click to download the data in .xls format. Simple tabular and graphical reports can be created by police department staff through the web-based report designer appropriate for technical novices. These can be saved and shared on the server for future use. An example of a repeatable tabular report might be the number of offenses per year for each university.

We also added interactive functionality where, for example, offenses per year could be drilled-down into counts of types of offense or clicking on a university drilled-through to a university-specific report. Many of the reports featured standard visualizations such as Pie-charts, bar charts, etc.

A particularly helpful interactive visualization is the scatterplot of universities on choose-able metrics. Here, the x-axis and y-axis are selectable from all metrics, and all universities are presented as dots in a scatterplot labeled with the university names. This allowed comparison of universities on any pairwise combination of metrics. This was very interesting to those knowledgeable in the campus crime domain, though perhaps unintelligible to the casual observer. For example, setting campus acreage on the y-axis and number of residence hall beds on the x-axis immediately showed how universities related to each other in a student-density sense. Then changing the x-axis to alcohol violations and back again gave a crude interactivity. This was highly engaging and had much more meaning to knowledgeable police professionals than to technical and analytics professionals.

The system was helpful for the benchmarking purpose. Although universities identify academic peer, sister, and aspirant schools, there is not an equivalent with respect to campus police challenges. Through interacting with the data, the police department was able to identify this subset of metrics to use in identifying universities with comparable crime challenges:

- Student headcount
- Total dormitory capacity
- Employee headcount
- MSA Population
- Total Acreage
- Total Operating Budget Expenses
- Operation and Maintenance Plant Expenses

Student headcount and employee headcount are measures of the population at the university. Total dormitory capacity indicates the proportion of residential population. The MSA population is a rough indicator of the urban/suburban/rural setting of the campus. Total acreage effective defines the area requiring monitoring. Total Operating Budget Expenses gives a rough measure of the total University budget, while Operation and Maintenance Plant Expenses (where Campus Police are generally placed) is an indicator of the police department budget. Using a normalized Euclidean distance measure, a set of six universities were determined to be very similar to UNC Wilmington.

Other analytic techniques were explored with limited success. An attempt was made to correlate discretionary metrics with crime metrics, e.g., number of sworn officers (discretionary) and number of alcohol violations. Unfortunately, the discretionary metrics are not available through the IPEDS data sets, and thus required surveying a subset of universities, resulting in only 6 universities participating. In the end, correlations were not significant at the alpha = 0.05 level.

Data Envelopment Analysis (Anderson, Sweeny, & Willimas, 1994) was then performed on this subset of similar universities. This technique gives a way to examine "efficiency" through comparison of discretionary metrics with outcomes (crime metrics). This analysis was foiled mainly by too few universities (6) with respect to number of metrics. This caused almost all universities to appear 100% efficient. Through this analysis, we also realized a domain-specific challenge: relationships between discretionary metrics and outcomes are complex and commonly non-linear. For example, adding more patrol officers increases tickets written linearly to a point, then adding more officers makes tickets decline. With few officers, the crime rate is constant, and additional officers merely *detect* more crime instances, resulting in more tickets. The presence of more officers begins to have a *deterrent* effect, actually decreasing the crime, and resulting in fewer tickets.

6. CONCLUSIONS

Design, construction, and use of this system was extraordinarily instructive, and useful to the campus police department. The main goal was met: to determine a data-driven approach to identifying benchmark universities with respect to campus police department challenges. The platform can be extended with user-authored reports as needed.

One glaring limitation is that the data loading was not scripted in such a way to be perfectly repeatable. Assuming the IPEDS, FBI, and US Census data report formats remain consistent, a scripted Extract, Transform, Load should be easily achievable with modern data manipulation tools.

In the future, we hope to add more data, mainly through annual IPEDS reports. With data sharing among the similar universities, it would also be possible to gain enough data to appropriately use statistical tests and techniques such as Data Envelopment Analysis.

7. REFERENCES

Anderson, R. D., Sweeny, D. J., & Willimas, T. A. (1994). *An Introduction to Management Science - Quantitative Approaches to Decision Making.* New York: West Publishing Company.

Government, U. (2019). Federal Code of Regulations. *EDUCATION*. Retrieved 7 15, 2019, from https://www.law.cornell.edu/uscode/text/20/1092#f

Iheadindu, U., Kline, D, Vetter, R., Clark, U., Donaldson, D. (2013) A Multi-University Data Analytics System for Understanding Campus Crime. Annals of the Master of Science in Computer Science and Information Systems at UNC Wilmington, 7(2) paper 6. http://csbapp.uncw.edu/data/mscsis/full.aspx

Johnston, T., & Weis, R. (2010). Managing Time in Relational Databases. San Francisco: Morgan Kauffman Publishers.

Larson, B. (2012). Microsoft SQL Server 2012: Reporting Services. New York: McGraw-Hill.

Microsoft. (2012, 10 3). Understanding Simple and Complex Expressions (Reporting Services). Retrieved from Microsoft: https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2008-r2/ms345242(v=sql.105)

Microsoft. (2013, 7 31). Authentication with Report Server. Retrieved from Microsoft: https://docs.microsoft.com/en-us/sql/reporting-services/security/authentication-with-the-report-server?view=sql-server-2017

Microsoft. (2013, 7 31). Working with Data in Reporting Services. Retrieved from Microsoft.com: https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2005/ms160324(v=sql.90)

National Center for Education Statistics. (2012). IPEDS Data Feedback Report. National Center for Education Statistics. Retrieved from https://nces.ed.gov/ipeds/DataCenter/DfrFiles/IPEDSDFR2012_199218.pdf

Russ, M., Ferrari, A., & Webb, C. (2012). SQL Server 2012 Analysis Services - The BISM Tabular Model. Sebastopol: Microsoft Corporation.

US Department of Education Office of Postsecondary Education. (2013). The

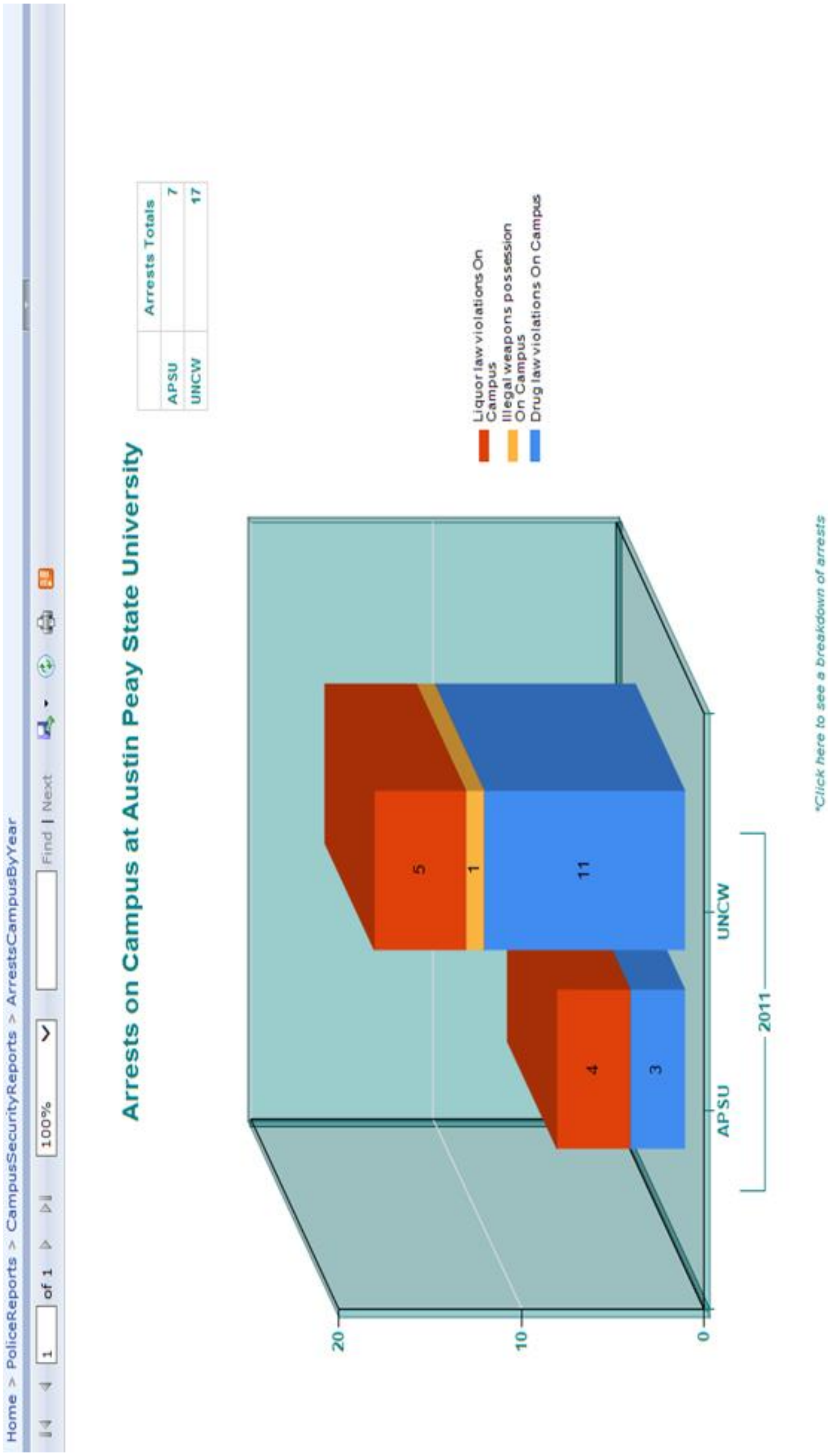
Handbook for Campus Safety and Security Reporting. US Department of Education. Retrieved from https://www2.ed.gov/admins/lead/safety/handbook.pdf

US Federal Bureau of Investigation. (2013). Uniform Crime Reporting Statistics. Retrieved from https://www.ucrdatatool.gov/

US Federal Government. (n.d.). US Census Bureah. Retrieved from https://www.census.gov/

Yeruva, S., Kumar, P., & Padmanabham, P. (2012). Distributed Warehouses: A Review on Design Methods and Recent Trends. International Journal of Computers and Distributed Systems, 60-61.

Appendix 1


Home > PoliceReports > CampusSecurityReports > ArrestsCampusByYear
1 of 1
100%
Find | Next
Arrests on Campus at Austin Peay State University
Arrests Totals
APSU 7
UNCW 17
Liquor law violations On Campus
Illegal weapons possession On Campus
Drug law violations On Campus
20
10
0
4
3
5
1
11
APSU
UNCW
2011
*Click here to see a breakdown of arrests