CALLS FOR 911 SERVICE: A COLLABORATION BETWEEN CLEVELAND STATE UNIVERSITY AND THE CLEVELAND CITY POLICE DEPARTMENT

John P. Holcomb, Jr.
Cleveland State University
Cleveland, OH 44115-2214 USA

Norean Radke Sharpe
Babson College
Babson Park, MA 02457 USA

For a period of time, professors from the Cleveland State University Department of Mathematics and Department of Sociology have been working closely with the City of Cleveland Police Department. This partnership has resulted in access to police records cataloging all emergency 911 calls for the city since 1995. Here, we share summary graphs and statistics as they relate to calls for service in the years 2001 – 2003 throughout the city, by district, and priority. We also discuss modeling approaches that can be used with students in second, or more advanced, courses in statistics to explore predicting the number of calls for service.

BACKGROUND

The city of Cleveland, Ohio is a metropolitan area located on the southern shores of Lake Erie. At the time of the 2000 census, the city had 478,403 residents. The year 2000 marked the first time the head count of Cleveland residents had fallen below 500,000. With dwindling tax revenues, the city government prioritized maximizing services to residents without increasing costs. In addition, the city was implementing a data-based approach to monitoring the dispensing of city services.

At this same time, Cleveland State University, a comprehensive and government supported institution of higher education located in downtown Cleveland, began to offer financial rewards to faculty who reached out to community agencies to form research partnerships. Faculty in the mathematics and sociology departments began a dialogue with the City Police Department to collect and analyze data related to calls of service. At that time, there was no comprehensive analysis of the volume, location, and nature of the calls.

As a result of this partnership, we received access to approximately 5 million records for calls for service (commonly known in the United States as 911 calls) from 1995-2003. This paper reports findings for 1,721,576 calls made from January 1, 2001 through December 31, 2003.

The goal of this paper is to describe pedagogical uses of these data for students in both introductory and advanced statistics courses. We describe some discussions and analyses that can be used for either level in this paper and provide links for teachers to obtain the data for use in their own classes. Instructors are encouraged to think of additional ways in which these data can be used and the authors welcome feedback from users of the data.

All the datasets described in this paper are available for download and use for pedagogical purposes from the website http://academic.csuohio.edu/holcombj/icots. Data are available in formats for MINITAB, Excel (when size allocations allow Excel use), SPSS, SAS, and comma – delimited (CSV). Specific datasets are referred to in the paper with their name and an extension denoted by .*** to denote the various versions of the data available.

USING POLICE DATA FOR INTRODUCTORY STATISTICS

Although data obtained from one city in the United States may not be universally interesting to students around the world, discussing data collection issues regarding calls for service from any local law enforcement agency should be relevant. One fruitful discussion might involve setting parameters for a study in one's local region and then discussing how the data would be collected. For the city of Cleveland, since the data are generated automatically from a Computer Automatic Dispatch system, the data are census data for the years in question because they are every call for service. Thought of in a grander scheme where the population of interest is all calls for service in the city of Cleveland since 1960, then the data can be considered as a sample.

With either approach, these kinds of data are wonderful to illuminate to students the power of simple summary statistics and graphs. It is our belief that students leave far too many introductory courses not realizing that much of statistics in the news are results and conclusions from simple summary analyses.

For example, a simple summary of the Cleveland data shows that there are 219 different identifying codes for the types of calls for police service. The table below gives the top ten categories for calls for service.

Table 1: Frequency table of top ten calls for service categories

Nature of Call	*Count*	*Percent*
Traffic Stop	186,123	10.8
District Assignment	79,032	4.6
Domestic Violence (Suspect present)	77,420	4.5
Burglar Alarm	71,686	4.2
Silent 911 Call	71,620	4.2
Residential Burglar Alarm	62,889	3.7
Drug Activity	54,195	3.1
Civil Disturbance	50,087	2.9
Suspicious Activity	34,750	2.0
Theft Report	34,312	2.0

What may be surprising to instructors and students is the third most common type of call for service – domestic violence involving assault or threats of assault with the suspect on the scene. Another area of surprise may be the large number of silent 911 calls. These calls are made to the call center, but the caller immediately hangs up or is silent. The City of Cleveland has a policy in place that officers respond to the geographic location of the call to determine if there is need of police service.

A data set as large as the calls for service may be too large for students to investigate on their own. The data set onesat.*** contains 1518 records for one single day – Saturday, January 6, 2001. The results here are similar to the above.

Of great concern to city council members is the number of calls per service broken down by district. The city of Cleveland has six districts with some known for higher crime than others. A simple frequency table for the six districts of Cleveland reveals a fairly even distribution of calls across the six districts (Table 2).

Table 2: Frequency table of calls for service by district

District	*Calls*	*Percent*
D1	290099	16.9
D2	296278	17.2
D3	270717	15.7
D4	309119	18.0
D5	271219	15.8
D6	284144	16.5

If, however, we take into account the population of each district and construct a bar graph, we see a different picture (Figure 1). This can prompt a useful discussion of the pros and cons of using per capita analysis. On the one hand, a regular frequency table shows useful information regarding where the calls originate. The per capita analysis illuminates how those calls compare in regard to the population of each district. We see district three has the fewest number of calls, but it also has the smallest population (see districts.*** for actual data counts). This zone

contains the downtown Cleveland business district which contains many businesses, restaurants, night clubs, and homeless people and each of these variables is associated with calls for police service.

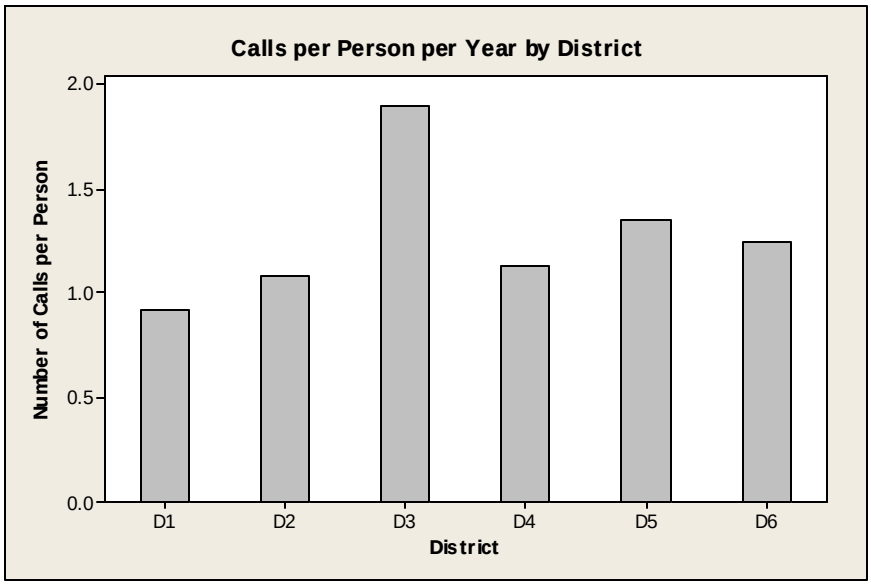

Figure 1: Bar chart of calls on a per capita basis

Another variable of interest is the priority of the service call. The City of Cleveland classifies calls into 5 priorities with priority 1 calls being the most important. Priority one calls consist of calls for service where crime is in progress – whether this be homicide, domestic violence, ethnic intimidation, shots being fired, or robbery, but the distinguishing feature is that the perpetrator is on the premises or close to the premises. Priority five calls are often police - initiated calls to the dispatch center that refer to patrolling one's district, filing reports, or walking through public areas, such as parks and neighborhoods.

Using district in conjunction with priority creates a two-way contingency table (see Table 3). This table provides an excellent exercise for students to practice calculations and concepts involving conditional probability. One can determine which district has the highest or lowest probability of receiving a priority one call.

Table 3: Frequency of Calls for Service by District and Priority

		Priority				
		1	2	3	4	5
	D1	28880	73859	86805	85551	15004
	D2	29996	79236	86972	81928	18146
District	D3	26386	73379	77880	74651	18421
	D4	34431	78275	84731	87393	24289
	D5	33905	72695	85449	62672	16498
	D6	33291	74006	84568	78880	13399

Another useful graph for students to construct is a time series plot to examine the mean number of calls per hour of the day (Figure 2). Here we see that the busiest hours of the day for calls for service are in the late afternoon/ early evening (hours 15 – 20). This is when traffic is at its worst and young people are home from school. Many students believe that police activity is busiest at night, but we see there is a steady decline in calls for police activity from 9:00 pm until 6:00 a.m.

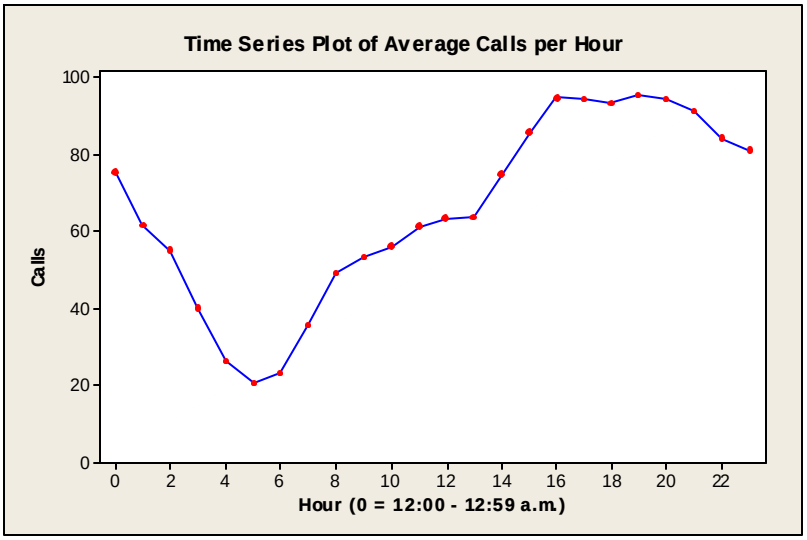

Figure 2: Average Calls per Hour

If we examine this six hour time frame (3:00 pm. to 8:59 p.m.) across the year on a monthly basis, we obtain Figure 3. Here we see that calls for service do increase dramatically during the hot summer months. For instance, in December (month 12), the median calls per hour is 74, while in June it is 102. To put this into perspective, 102 is a 38% increase in calls. The concept of *percent increase* is something not often discussed in the introductory statistics curriculum, but it is an essential part of quantitative literacy which is often lacking in university students in the United States.

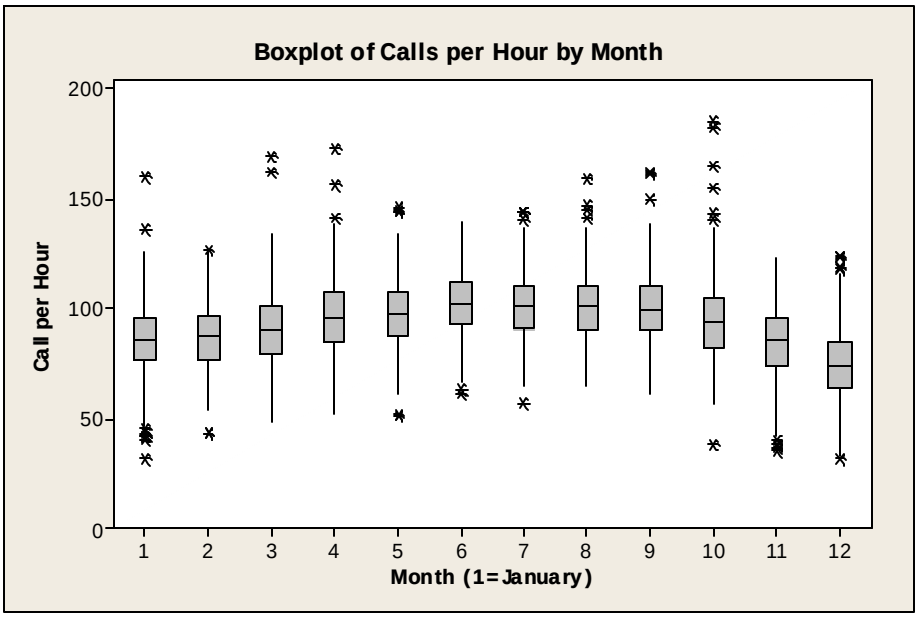

Figure 3: Calls Per Hour 3:00 p.m. – 8:59 p.m.

USING POLICE DATA FOR ADVANCED COURSES IN STATISTICS

Building on the fact that these 911 calls are collected over time, we can use regression and time series techniques to model the data using concepts presented in more advanced courses. Since the number of calls varies both by district and by priority level, it makes sense to model the time series for each district at each priority level. In this way, the forecasts provided by the

models may be more useful for the police for scheduling or staffing purposes. To enable a comparison of forecasts provided by each of the models with actual values, the last several months of our dataset may be used as a holdout sample. Figure 4 shows the number of monthly calls by district for each priority from January, 2001 through December, 2003.

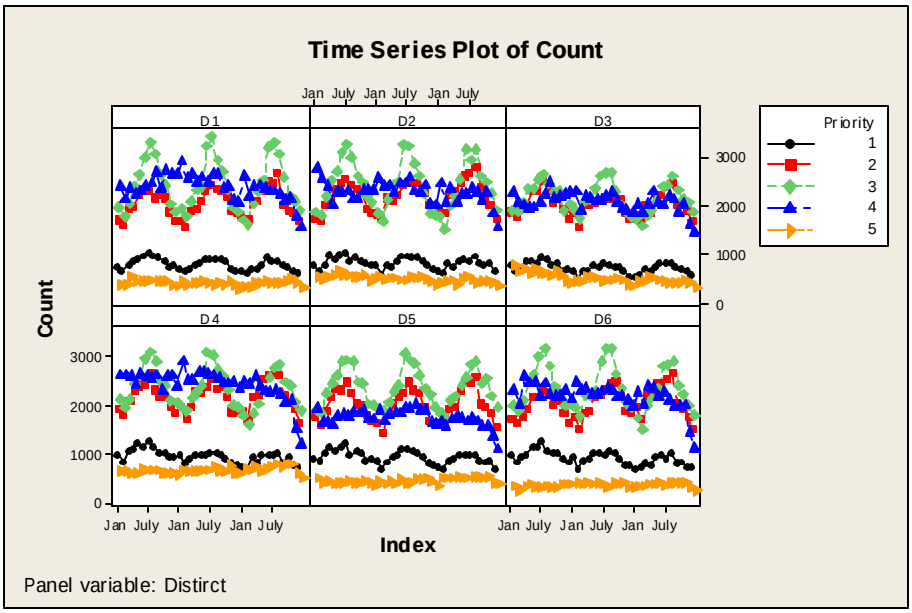

Figure 4. Time Series Plot of Calls per Month by Priority Level in each District

Note, that the level of calls for priority 1 (the most urgent) are relatively constant across the six city districts. The same is true of the priority 2 and priority 5 (administrative) calls. However, the priority 3 calls appear to be lower in district 3 and the priority 4 calls appear to be lower in district 5. Seasonal peaks in the summer are clearly visible in all districts for the most urgent calls: priorities 1,2, and 3.

Since these data reflect seasonality, as noted in Figures 3 and 4, they provide students an opportunity to use multiple time series approaches. Appropriate models include regressions with seasonal dummy variables and autoregression models with multiple lagged variables. The advantage of using the regression approach is that other variables (besides the dummy variables) can be added. However, the number of variables can quickly grow and additional variables may not be available at each time point. The advantage of the autoregression approach is that the only dataset needed is the time series itself. Decomposition methods may also be demonstrated on these data, where a seasonal index is obtained for each month. Students can use these data to explore the advantages and disadvantages of each modeling approach. The segregation of the dataset by priority and district also provides datasets for multiple teams of students in larger classes.

Another pedagogical use of the data is for use in a statistical programming course. The data allcalls.*** has at the observation level a record for each and every 911 call. It is excellent practice for students to use these data to make data sets ready for analysis at the hourly, monthly, or daily level described in the preceding paragraphs of this paper. Many students fail to realize the amount of preparation time that is needed to make data ready for analysis. If they obtain employment working in a statistical capacity, a large portion of their time will be devoted to the activity of data preparation.

CONCLUSION

This paper presents a case study using a dataset from the Police Department in the city of Cleveland consisting of 911 service calls. The descriptor variables in the dataset, including nature of call, district of call, and priority of call, provide different types of variables for students to explore both graphical and numerical descriptive approaches. In addition, the fact that the calls can be aggregated hourly, daily, or monthly enables students in more advanced statistics courses to apply regression and time series techniques to a real dataset. The seasonal nature of the data provides opportunities for students to apply the use of different variables (e.g., dummy and lagged variables), as well as residual diagnostics.