

# Challenges in Utilization of Statistical Analysis Software in Emergency Dispatch Data Analysis and Advances in Data and Technologies

Sashidhar R. Yerram, MS<sup>1</sup>

1. Department of Biomedical Informatics,  
University of Utah, Salt Lake City, Utah,  
USA

## Corresponding Author

Sashidhar R. Yerram  
Department of Biomedical Informatics  
University of Utah  
421 Wakara Way, Suite 140  
Salt Lake City, UT 84108  
sashidhar.sas@gmail.com

## Keywords

dispatch data, statistical analysis, big data

## Citation

Yerram SR. Challenges in utilization of statistical analysis software in emergency dispatch data analysis and advances in data and technologies. *Annals of Emergency Dispatch & Response*. 2016;4(2):5–6.

## INTRODUCTION

Emergency dispatch and emergency medical services (EMS) data keeps increasing every day—in quality, volume, and dimensionality. An Emergency Medical Dispatcher (EMD) at an emergency communication center is the primary link between the public caller requesting emergency medical assistance and EMS. Data collection starts when the EMD receives the call and ends when the patient is either treated on scene or admitted to the hospital. The data collected is mostly structured data containing both text and numerical data types.

Data generated from emergency dispatch is very valuable, as we can gain extremely beneficial insights from the data, such as identifying ways to reduce response times, determining the most commonly encountered types of emergencies (or Chief Complaints), and understanding the text data entered by the EMD to describe the caller's emergency problem. These insights in turn can help to evolve dispatch protocols and to design and implement better training for calltakers, dispatchers, and other emergency communication center staff. However, collecting and analyzing the data can present challenges. When attempting to apply one common statistical analysis software (called R) to one large dispatch dataset, we encountered challenges that are typical of those that researchers and data analysts will continue to encounter as dispatch data grows in size and dimensionality.

## CHALLENGES AND POTENTIAL SOLUTIONS

There are many tools that can be used for general data analysis and reporting results. Some of the tools are Tableau, MicroStrategy, and QlikView. These tools can be used for data analysis and for developing dashboards where emergency dispatchers can keep track of Key Performance Indicators in real time, depending on when they refresh the data from the database. For more sophisticated statistical data analysis, there are tools such as SAS, SPSS, STATA, and R program, which enable researchers to develop prediction models or perform other complicated analyses.

Even these programs, though, present their own difficulties. One of the problems with R program, for example, is that it was designed to have all of the data stored in memory—and it is memory-intensive. Emergency dispatch data can contain thousands of rows and can significantly exceed the memory of computers when we analyze large datasets. Hence, to alleviate memory problems, we need to split the datasets prior to analysis. However, splitting the data degrades the insights we can gather.

Another potential challenge for dispatch data analysis is so-called “Big Data.” Big Data can involve both structured and unstructured data, resulting in datasets greater than 1 terabyte (TB) in size. Big Data has three main characteristics: volume, velocity, and variety. Volume is the amount of data generated, velocity is how frequently data is generated, and variety refers to the number of types of data generated. Emergency dispatch data may already be categorized under Big Data based on volume, and sooner or later it will satisfy all three conditions. This is primarily due to the fact that, with time, as data volume increases, velocity also increases, because as population increases the total number of emergency cases may also exponentially increase. Also, as technology becomes cheaper, various other parameters may also be collected, which increases the variety of data.

Big Data cannot be analyzed using regular tools such as R program, SPSS, STATA, or Tableau. Using regular analysis tools, we pull data from data sources and directly load it into the tools for analysis; this is not a problem when the dataset is small. However, when the dataset size is big, it is difficult to pull it from the source and load it into the analysis tool. Hence, the analysis tools or programs need to reside where the data is—at the source. This way, analysis is performed without moving the data from its source. Hadoop framework software is the commonly used Big Data technology that has gained the most momentum recently. Hadoop framework consists of Hadoop Distributed File System (HDFS) and MapReduce. HDFS enables the storing of data. It divides the data in to smaller parts and distributes the data across various servers or nodes. MapReduce software framework helps in processing the data in parallel in the cluster—in small, easily-managed and -processed chunks of integrated datasets. Numerous vendors such as Amazon Web Services, Coludera, Hortonworks, and MapR technologies distribute open source Hadoop platforms. These may eventually offer solutions to some of the dispatch community's Big Data challenges.

## SUMMARY

Dispatch and EMS provide very rich data. Existing statistical analysis software tools can reveal trends and insights in that data that can help move dispatch forward. However, in the near future, as dispatch and EMS data get classified under Big Data, more sophisticated statistical tools such as Hadoop will evidently be needed in order to gain better insights in the data. Therefore, by keeping the growth in data (both size and dimensionality) in mind, it would be a rewarding investment to begin thinking about how to apply real-time analysis and reporting and big data technologies.

## ACKNOWLEDGEMENT

Thanks to Chris Olola, PhD; Isabel Gardett, PhD; and Greg Scott, MBA, EMD-Q® for their mentorship and supervision of the pilot application of the concept of using R analysis tool to mine problem description data in emergency dispatch research.