**R FONKSİYONUYLA KUTU GRAFİĞİNİN ÇİZİLMESİ ÜZERİNE YENİ BİR UYGULAMA**

**Özet**

*Verilerin özetlenmesinde tablolar ve hesaplamaların yanında grafikler de sıklıkla kullanılmaktadır. Grafiklerin kullanımı sayesinde, veri kümelerinde fark edilmeyen desenler rahatlıkla belirlenebilmekte ve yorumlanabilmektedir. Özellikle açık kaynak kodlu programlama dillerinin gelişmesiyle birlikte veri görselleştirme aracı olan grafiklerin özelliklerinin de geliştiği görülmektedir. Bu gelişim, verilerin özetlenmesinde sıklıkla kullanılan bazı grafik türlerindeki sakıncaların ortadan kaldırılmasına olanak sağlamaktadır. Görünüşte oldukça sade ve basit olan kutu grafiği, ilk etapta bir verinin en düşük, ilk çeyrek, medyan, üçüncü çeyrek ve en büyük değerlerinden oluşan özetine dayalı olarak dağılımını gösteren çok kullanışlı bir grafik türüdür. Bununla birlikte kutu grafiği, verinin simetrik veya çarpık olup olmadığı, verinin nerede yoğunlaştığı ve varsa aykırı değerler hakkında da bilgi vermektedir. Basit görüntüsüne rağmen kutu grafiğinin çok işlevsel bir gösterim olduğu söylenebilir. Kutu grafiği, değişkenlere ilişkin bazı betimsel istatistiklerin izlenmesini sağlamakla birlikte, bu grafikte aykırı değerler olarak belirlenen araştırma birimlerinin isimleri görünmemekte, bu birimler yıldız ve/veya nokta işaretleri ile temsil edilmektedir. Bu çalışmada, kutu grafiğindeki bu aykırı değerlerin adlarının görünmesini sağlamak ve bu sakıncayı gidermek amacıyla, kutu grafiği açık kaynak kodlu bir programlama dili olan R’de fonksiyon yazılarak yeniden oluşturulmuştur. Böylece, yazılan yeni R kodu ile veri kümesindeki aykırı değerler doğrudan görülebilmektedir. R kodlarıyla yazmış olduğumuz kutu grafiği fonksiyonu girdi olarak tüm veriyi kapsayan bir .csv dosyasını, verinin ait olduğu etiketi, sayısal ve faktör verilerinin bulunduğu sütun bilgisini ve ayrıca faktör seviyelerini kullanmaktadır. Bu bilgilere göre kod her bir faktör ve sayısal değer için çoklu kutu grafiği çizmekte, her bir kutu grafiği içinde varsa aykırı değerleri işaretleyerek isimlerini grafik üzerine yazmaktadır. Çalışmada, yazılan R kodu birkaç veri kümesi üzerinde denenmiş, elde edilen bulguların SPSS 21.0’den elde edilen bulgular ile örtüştüğü görülmüştür. Böylece çalışmada, araştırmacılar için daha kullanışlı olduğu düşünülen bir kutu grafiği kod sistemi tanıtılmıştır.*

*Anahtar Kelimeler: Kutu grafiği, R programlama, R fonksiyonu.*

**A NOVEL APPLICATION ON DRAWING BOXPLOT WITH THE R FUNCTION**

**Abstract**

*In addition to tables and calculations, graphs are often used to summarize data. Thanks to the use of graphics, unnoticeable patterns in data sets can be easily identified and interpreted. Especially with the development of open source programming languages, it is seen that the features of graphics, which are data visualization tools, have also improved. This development allows to eliminate the drawbacks of some types of graphs that are frequently used in summarizing data. The seemingly plain and simple boxplot is a very useful type of chart that initially shows the distribution of a data based on a summary of its lowest, first quartile, median, third quartile and highest values. In addition, the box plot also gives information about whether the data is symmetrical or skewed, where the data is concentrated, and outliers, if any. Despite its simple appearance, the boxplot can be said to be a very functional representation. Although the boxplot allows to monitor some descriptive statistics about the variables, the names of the research units determined as outliers are not visible in this graph, these units are represented by asterisks and/or dots. In this study, the boxplot was reconstructed by writing a function in R, an open source programming language, in order to make the names of these outliers appear in the boxplot and to eliminate this drawback. Thus, outliers in the dataset can be seen directly with the new R code written. The boxplot function that we wrote with R codes uses a .csv file that includes all the data, the label to which the data belongs, the column information containing the numerical and factor data, and also the factor levels. According to this information, the code draws a multi-box plot for each factor and numerical value, marks the outliers in each boxplot, if any, and writes their names on the chart. In the study, the written R code was tested on several datasets, and it was seen that the findings were consistent with the findings obtained from SPSS 21.0. Thus, a boxplot code system, which is thought to be more useful for researchers, is introduced in the study.*

*Keywords: Boxplot, R programming, R function.*