UNIVERSITI PUTRA MALAYSIA

MODIFIED BOXPLOT AND STAIRBOXPLOT FOR GENERALIZED EXTREME VALUE DISTRIBUTION

BABANGIDA IBRAHIM BABURA

IPM 2018 3

MODIFIED BOXPLOT AND STAIRBOXPLOT FOR GENERALIZED EXTREME VALUE DISTRIBUTION

By

BABANGIDA IBRAHIM BABURA

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

November 2017

DEDICATIONS

To my parents;
My mother Hajjiya Amina Ibrahim &
Father Late Malam Ibrahim Abubakar (May his soul rest in perfect peace, Ameen) .

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

MODIFIED BOXPLOT AND STAIRBOXPLOT FOR GENERALIZED EXTREME VALUE DISTRIBUTION

By

BABANGIDA IBRAHIM BABURA

November 2017

Chair: Associate Professor Mohd Bakri Adam, PhD
Institute: Institute for Mathematical Research

A boxplot is an exploratory data analysis tool for a compact distributional summary of a univariate dataset. It is designed to recognised all typical observations and displays the location, spread, skewness and the tail of the data. When the dataset is skewed such as extreme data, the precision of boxplot functionalities is less reliable and inaccurate. Many observations from extreme data were erroneously marked as outliers by the classical boxplot methods.

The Tukey's classical and Hubert's adjusted boxplots were utilize in the study based on outside rate per sample and a propose measure of fence sensitivity ratio to observe the suitability of the methods according to a simulation process from Generalized Extreme Value distribution. The adjusted method improves the classical method in extreme data capture but not sufficiently optimize to achieve the bench mark requirement in the literature.

The modified boxplot has been proposed with a fence adjustment of the existing boxplot method using the Bowley coefficient. The fence position was considered as a response to skewness in the simulated extreme data from GEV distribution and then fitted with resistance fit linear regression model. The propose fence adjustment enhance the boxplot to detect all atypical observations without any parametric assumption about an extreme data. The new boxplot displays some additional features other than the classical one such as a quantile region for the parameters of Generalized Extreme Value distribution in fitting an extreme data.

The modification of the entire boxplot display is also proposed as stairboxplot with combined features of boxplot, histogram and a dot plot. The stairboxplot divides the data points of a sample into four portions according to the range of the data set, such

that the individual points are inscribed in their respective range levels. However, stairboxplot displays each observation according to an introduce measure of outlyingness of a point called stairboxplot outlyingness.

The main findings and contributions in both modified boxplot and stairboxplot can generally be attributed to the enhancement of quality of a dataset by highlighting inconsistent observations from GEV distribution's modelling framework and diagnostic visualisation of extreme data to gain immediate information such as skewness, quantile estimate of GEV parameters region and data points display according to outlyingness.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PLOT-KOTAK TERUBAHSUAI DAN PLOT-TANGGA-KOTAK BAGI TABURAN NILAI EKSTRIM TERITLAK

Oleh

BABANGIDA IBRAHIM BABURA

November 2017

Pengerusi: Profesor Madya Mohd Bakri Adam, PhD
Institut: Institut Penyelidikan Matematik

Plot-kotak merupakan salah satu alat bagi analisis data terokaan untuk ringkasan taburan yang padat bagi set data univariat. Ianya direka untuk mengenalpasti kesemua jenis cerapan serta menunjukkan lokasi, serakan, kepencongan dan kepuncakan sesebuah data. Apabila data menunjukkan ciri pencongan seperti data ekstrim, ketepatan fungsi plot-kotak adalah kurang dipercayai dan tidak tepat. Kebanyakan cerapan daripada data ekstrim telah disalah tanda sebagai data terpencil oleh kaedah plot-kotak yang klasik.

Plotkotak yang klasik daripada Tukey dan yang diubahsuai oleh Hubert telah digunakan dalam kajian ini berdasarkan kadar luaran per sampel serta ukuran nisbah kepekaan pagar yang dicadangkan untuk melihat kesesuaian sesuatu kaedah berdasarkan proses simulasi dari taburan nilai ekstrim teritlak. Kaedah yang diubahsuai didapati adalah lebih baik dari kaedah klasik dalam mengenalpasti data ekstrim namun ianya tidak cukup di optimumkan untuk mencapai tahap keperluan dalam kajian tinjauan.

Plotkotak yang diubahsuai telah dicadangkan dengan pelarasan terhadap pagar plot-kotak yang sedia ada dengan menggunakan pekali Bowler. Kedudukan pagar telah diandaikan sebagai tindak balas kepada kepencongan di dalam data simulasi ekstrim daripada taburan nilai ekstrim teritlak dan kemudian dipadankan dengan model regresi rintangan linear. Pengubahsuaian pagar yang dicadangkan telah meningkatkan kebolehan plot-kotak dalam mengenalpasti semua jenis cerapan tanpa sebarang andaian parameter tentang data ektrim. Beberapa ciri tambahan telah ditunjukkan oleh plot-kotak baharu berbanding plot-kotak klasik seperti rantau kuantil bagi parameter-parameter taburan nilai ekstrem teritlak dalam mengesuaikan data ekstrim.

Pengubahsuaian terhadap keseluruhan paparan plot-kotak turut dicadangkan yang dikenali sebagai plot-tangga-kotak dengan menggabungkan ciri plot-kotak, histogram

dan plotdot. Plot-tangga-kotak membahagikan sampel data kepada empat bahagian berdasarkan julat set data supaya setiap cerapan ditempatkan di tahap julat masing-masing. Walau bagaimanapun, plot-tangga-kotak memaparkan setiap cerapan mengikut kepada ukuran keterpencilan titik yang diperkenalkan dan dikenali sebagai keterpencilan plot-tangga-kotak.

Penemuan dan sumbangan utama dalam kedua-dua plot-kotak dan plot-tangga-kotak yang diubahsuai boleh dikaitkan secara amnya untuk meningkatkan kualiti set data dengan menyoroti pemerhatian yang tidak konsisten daripada rangka model nilai ekstrim teritlak dan visualisasi diagnostik daripada data ekstrim untuk mendapatkan maklumat yang segera seperti kepencongan, anggaran quantil daripada rantau parameter nilai ekstrim teritlak, paparan titik data mengikut kepada keterpencilannya.

ACKNOWLEDGEMENTS

All gratitude goes to *Allah subhanahu wataala*, on whom ultimately we depend for sustenance, guidance and accomplishment. Special thanks to my Chairman Supervisory committee Associate Professor Dr Mohd Bakri Adam for the wonderful guide and inspiration to pursue a PhD research work to the level of accomplishment. I place in high esteem his vast wealth of knowledge, support and constructive criticism throughout the entire research journey. I would also like to express my appreciation to my co-supervisors; Dr Anwar Fitrianto and Associate Professor Dr Abdul Rahim Abdul Samad for their academic support and constructive criticism.

A special gratitude to my institution, Federal University Dutse, for my nomination in 2014 as a recipient of TETFUND PhD fellowship which become an instrument that facilitates a smooth pursuit of the research experience. Finally, I appreciate all those who gave me unconditional physical and emotional supports most especially my mother Hajia Amina, my wife Ramlat, my children, my relatives and friends. I thank you all and wish you Allah's blessing.

Babangida Ibrahim Babura

I certify that a Thesis Examination Committee has met on 28 November 2017 to conduct the final examination of Babangida Ibrahim Babura on his thesis entitled "Modified Boxplot and Stairboxplot for Generalized Extreme Value Distribution" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Shamarina binti Shohaimi, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Mohd Rizam bin Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Noor Akma binti Ibrahim, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Sutawanir Darwis, PhD
Professor
Bandung Islamic University
Indonesia
(External Examiner)

NOR AINI AB. SHUKOR, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 26 April 2018

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Mohd Bakri Adam, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

Anwar Fitrianto, PhD
Senior Lecturer
Faculty of Science
Universiti Putra Malaysia
(Member)

Abdul Rahim Abdul Samad, PhD
Associate Professor
Faculty of Economics and Management
Universiti Putra Malaysia
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature:____________________ Date:____________________

Name and Matric No: Babangida Ibrahim Babura, GS 40947

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________________
Name of Chairman of Supervisory Committee:
Associate Professor Dr. Mohd Bakri Adam.

Signature: ______________________________
Name of Member of Supervisory Committee:
Dr. Anwar Fitrianto.

Signature: ______________________________
Name of Member of Supervisory Committee:
Associate Professor Dr. Abdul Rahim Abdul Samad.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EDA	Exploratory Data Analysis
EVT	Extreme Value Theory
GEV	Generalized Extreme Value
GPD	Generalized Pareto Distribution

CHAPTER 1

INTRODUCTION

1.1 Background

The statistical field of exploratory data analysis (EDA) is endowed with simple but robust methods and techniques of understanding some immediate information about a dataset that may otherwise go unnoticed. The EDA techniques are typically applied before formal modelling commences and can help inform the development of more complex statistical models. The EDA practice was made popular by the work of John Tukey over 40 years ago among which the promotion of the concept of box and whiskers plot, see Tukey (1977). The popularity of boxplot which can be related with its philosophy of simplicity makes it as one of the most useful tool for EDA practice on univariate dataset (Hoaglin et al., 1983). The boxplots are particularly useful in detecting outliers and comparison between groups of dataset to visualized population difference or similarity of distributional properties.

The boxplot is a compact distributional summary which displays fewer details than other plots like histogram or kernel density but interestingly gives room to robust analysis and taking up less space. The robust summary statistics obtained from boxplot are usually located at actual data points with less computation and require no tuning parameters. The boxplot contains five conventional values of significance, namely; the two fences, the upper and lower hinges (quartiles), and the median (Tukey, 1977).

Extreme data are referred to a record of events that are more extreme than any that have already been observed within a particular block of time or over a determined threshold level. In the recorded history, work on extreme value may be traced to as early as 1709 when Nicolas Bernouilli discussed the mean largest distance from the origin given n points lying at random on a straight line of a fixed length l, a narration according to Coles (2001). Such observation can either be low extreme as minima or high extreme as maxima. The generalized extreme value (GEV) distribution is a limiting distribution describe by Generalized Extreme Value Theory. The GEV distribution is statistically useful in describing the likelihood of unusual behavior or rare events occurring. Its application is widely used in the areas of hydrology or environmental studies such as; flood frequency or associated rainfall intensity Rossi et al. (1984), Cooley et al. (2007) and Katz et al. (2002) , sea level analysis such as Méndez et al. (2007) and Tawn (1992) and other environmental trends Smith (1989), so also in finance/insurance as for; Jansen and De Vries (1991), Longin (1996) and Loretan and Phillips (1994).

The family of generalized extreme value (GEV) distribution comprising of Gumbel, Freichet and Weibull distributions possessed the distributional properties of asymmetry, heavy tail and skewness. These properties make it difficult for the standard boxplot to generate a good fence estimate that capture all typical observations within the fence markup area. We begin the research by studying the limitation of the existing boxplot methods and extend the conventional functions of boxplot not only on proper detection of outliers from an extreme dataset but to some additional functionalities. These modifications include; adjustment of the fence position to account for skewness of extereme dataset, additional display feature to the regular boxplot that display quantile region for location and scale parameters estimate along with skewness estimate of the shape parameter of the GEV distribution for fitting an extreme sample.

Furthermore, we proposed alternative boxplot called a stairboxplot. The plot has combine features of boxplot, histogram and a dot plot. The stairboxplot display individual points according to an introduce measure of outlyingness of a point. A simulated and real-life data were used to justify the advantages of this research work over those found in the literature.

1.2 Basic Configurations of Boxplot

1.2.1 Standard Boxplot

Turkey's boxplot consist of five components, strategically selected for a robust summary statistics of an ordered dataset $X_n = \{x_{(1)}, x_{(2)}, \cdots, x_{(n)}\}$. Figure 1.1 is the classical boxplot labeled with the five components and their descriptions as follows:

1. *The median*, denoted as Q_2 which is represented as the line that divided the box into two parts. It is located as the middle value when the dataset is arranged (sorted) in ascending order. So, for X_n, the $x_{\left(\frac{n+1}{2}\right)}$ observation is considered the median if n is odd, while the mid-point $\frac{x_{\left(\frac{n}{2}\right)}+x_{\left(\frac{n}{2}+1\right)}}{2}$ is the median for even n.
2. *The upper and lower hinges* corresponding to the upper and lower edges of the box with the edges passing through lower quartile (Q_1) and upper quartile (Q_3) of the dataset. The lower quartile is usually obtained as the middle item from the data points below the median, while the upper quartile is the middle item from data points above the median.
3. *The upper and lower fences* corresponding to two mark-up data points, a distance of h times the interquartile range ($IQR = Q_3 - Q_1$); below Q_1 for the lower fence and above Q_3 for the upper fence i.e $f_l = Q_1 - hIQR$ for the lower fence and $f_u = Q_3 + hIQR$ for the upper fence, where h is constant usually chosen to be 1.5

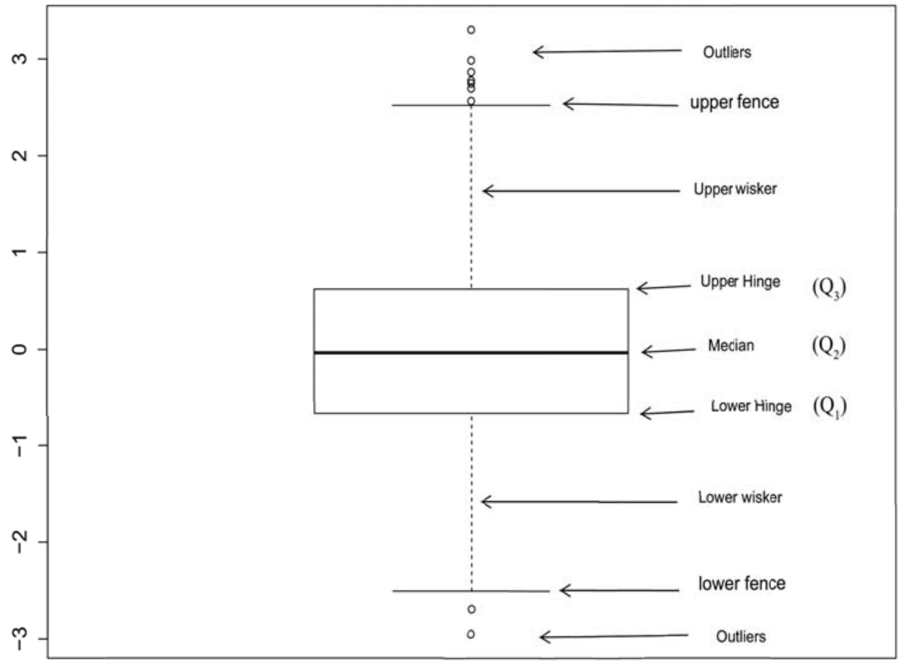

Figure 1.1: Classical boxplot.

for inner fence or 3.0 for outer fence. The description on how different choices of h are made is given in Chapter 3.

4. *The two whiskers* are straight lines which connect nearest data point above and below the lower and upper fences respectively to the two hinges.

5. *The outliers* are data points that deviates quantitatively from the majority of the data points, based on outlier-selection method above the upper fence or below the lower fence and are marked as points in the boxplot.

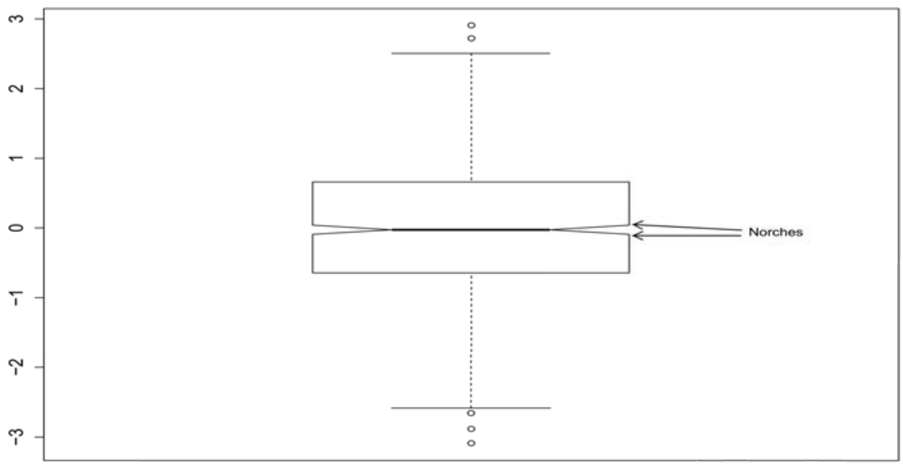

Figure 1.2: Notched boxplot.

1.2.2 Notched Boxplot

The notched boxplot is constructed in a similar way as the standard boxplot. The only difference is that it goes one step further by displaying confidence intervals around the medians, supporting the visual assessment of statistical significance. The length of the confidence interval is obtained so that non-overlapping intervals indicate (approximately) a difference at the 5% level, regardless of the underlying distribution. The notches are marked as indicated in Figure 1.2 and determined from the confidence interval around the median given by $Q_2 \pm 1.58IQR$ (McGill et al., 1978). Notch is the first visual enhancement made to boxplot classical display.

1.3 Other Variations in Boxplots

Boxplot has received a considerable interest from among variety of scholars. This makes it experience variations with significant enhancement in both plots, values of interest and applications. There are three early variants of boxplot as reported by Turkey (1978). The first incorporate a visual display of a measure of group size; the second features a highlight significance of differences between medians (or quantiles); and the third mixed both features of the first two. Additional boxplot variants that existed include; midgap plot by Tufte (2001) Tufte, colorful boxplot by Carr (1994), Boxplot for circular variables by Abuzaid et al. (2012). Wickham and Stryjewski (2012) categorise other variants of boxplot according to richer displays of density such as vaseplot (Benjamini, 1988), violinplot (Hintze and Nelson, 1998), beanplot Kampstra et al. (2008), raindrop plot Barrowman and Myers (2003) and more superior display of density according to Cohen and Cohen (2006) with sectioned-densityplot.

Boxplot was extended encompass bivariate data and is referred to as rangefinder plot (Becketti and Gould, 1987), the relplot (Goldberg and Iglewicz, 1992), quelplot (Goldberg and Iglewicz, 1992), bagplot (Rousseeuw et al., 1999), bivariate boxplot (Zani et al., 1998)(Zani et al., 1998), rotational boxplot (Muth et al., 2000) (Muth et al., 2000) and functional boxplot (Hyndman and Shang, 2010; Sun and Genton, 2011).

1.4 The Extreme Data

In this thesis we are interested in extreme (Maximum) observations in a dataset. These observations can be modelled using parametric models such as Gumbel, Freichet, Weibull distributions which all belong to the family of generalized extreme value distribution. A typical example of extreme value data include annual flood discharge level, insurance and financial data, teletraffic data in communication, minimum strength for the quality of materials, and a lot more extreme events in different scientific field of research.

Consider the following family of extreme value distributions for maxima $x \in \boldsymbol{X}$ where $\boldsymbol{X}$ is the set of block maximum. The GEV family which are described based on different shape parameters ξ is given by the theorem.

Theorem 1.1 *(Coles, 2001) If there exists sequences of constants $\{a_n > 0\}$ and $\{b_n\}$, as $n \to \infty$, such that $Pr\{(M_n - b_n)/a_n \leq x\} \to G(x)$ where M_n is a block maximum of n observations and G is a non-degenerate distribution function, then G is a member of the GEV family:*

$$G(x) = \exp\left\{ - \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi} \right\}$$

defined on $\{x : 1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0\}$, where $\sigma > 0$ and $\mu, \xi \in \Re$.

The varying shape parameter ξ define the GEV family upon the tail behavior. That is, if $\xi = 0$ the distribution is *Gumbel distribution* and decays exponentially. If $\xi < 0$, the distribution is negative *Weibull distribution* with finite short upper endpoint. While $\xi > 0$ the distribution is a *Fréchet distribution* with a heavy tail behavior to the right.

1.5 Problem Statement

Boxplot is one of the most routinely used EDA toolkit for outliers detection. Its popularity can also be associated with rich display of data summary in one simple plot. Tukey (1977) utilized the robust statistical techniques in constructing boxplot for univariate data. Boxplot is widely used in diverse scholarly works which include but not limited to Hoaglin et al. (1986), Kimber (1990), Davies and Gather (1993), Iglewicz and Banerjee (2001) and Banerjee and Iglewicz (2007) with approach according to the boxplot resistance rules. The existence of outliers are usually caused by the measurement error while recording observations. It is difficult and impracticable to avoid measurement errors. However existence or wrong measure in dataset can cause difficulty and inaccurate statistical inferences and hence the need of detecting and removing them. The present boxplot resistance rule doesn't have a particular general rule for all data type. We observe that dataset such as extreme data are particularly skewed in nature (Katz et al., 2002) and thus did not conform with Tukey's standard rule of thumb in defining boxplot fences. This problem made those interested in modelling extreme event avoid using boxplot for outlier detection (Spencer and McCuen, 1996).

However, the boxplot visualisation of extreme data reveals only descriptive statistical details obtainable from the three quartile. In a Gaussian set-up, the second quartile (median) can be utilized as a robust estimate of the location and interquartile range that spans the width of the box in boxplot as robust estimate of scale parameter. But in GEV distribution modelling framework, these three boxplot quartiles require such substance illustrated in the Gaussian set-up. This prompt enhancement of the existing boxplot set-up to account for the GEV distribution fitting parameters to enable a proper diagnostic of extreme data.

This research will extensively review the existing literature on boxplot and it's EDA diagnosing potentials in visualising extreme data by addressing the above listed problem as described fully in the aims and objectives section of this thesis.

1.6 Research Aim and Objectives

The main aim of this research is to improve the existing boxplot display to reflect specific characteristics of extreme data based on the following objectives:

1. To identify the characters of the existing boxplot methods, especially in the outlier labeling rules in visualizing extreme data.
2. To propose a new boxplot fence definition that reflects the skewness and capture all regular observations that align with the generalized extreme value distribution.

3. To enhance the display of boxplot by reflecting some additional diagnostic features of extreme data that account for the fitting parameters of GEV distribution.

4. To construct an alternative plot that maintains all the robust features of boxplot and overcome some limitations of the existing boxplot in visualizing individual observations and density of a dataset.

1.7 Limitation of the Study

We consider some limitations to the implementation of the objectives of our study. The simulation and real life dataset in the study are according to the block maximum extreme data modelling framework. Block maximum independent random variables where assumed to follow the popular extreme modelling tool of GEV distribution.

We adopt the existing theory and philosophy behind the boxplot in both boxplot methods, performance assessment, fence rule modification and additional visual features. However, we consider as necessary to stress that the newly incorporated features to the boxplot are to remain for diagnostic purposes which is the guiding philosophy of exploratory data analysis.

1.8 Structure of the Thesis

This thesis examine the application of boxplot as EDA diagnostic tool for extreme event modelling. There are eight chapters of the thesis that can be categorise into three phases. Chapter 1 to Chapter 3 are preliminaries on concept, literature review and methodology. Chapter 4 to Chapter 7 are presentation of results. Chapter 8 is summary of the entire research work.

In a more clear terms, Chapter 1 introduces the concept behind boxplot construction with its advantages over other visual EDA tools, introduction of the concept of extreme data and its statistical modelling tools. We also present in Chapter 1, a highlight on the research problem statement along with aim and objectives of the research all together with the limitation and structure of the thesis. An extensive review of literature in constructing different types of boxplot and outliers labelling rules and extremal events modelling are discussed in Chapter 2. In Chapter 3, we describe the methodology and philosophy involved in constructing the existing and proposed boxplots methods with other important statistical tools used in the entire research work.

The second phase begins with Chapter 4, it reviews the performance of the boxplots outlier rules and other boxplot characters in visualizing extreme samples based on simulation study. In Chapter 5, we propose a new fence definition for boxplot outlier rule using Bowley skewness estimate. Chapter 6 gives a modification of boxplot with additional diagnostic features of GEV modelling fit, with discussion of the new method on some real life extreme dataset. Finally, we propose an alternative plot called rangeplot and discuss its advantages over boxplot in Chapter 7.

The concluding part of the thesis is presented in Chapter 8, that gives summary, conclusion and recommendations for future research.

BIBLIOGRAPHY

Abuzaid, A. H., Mohamed, I. B., and Hussin, A. G. (2012). Boxplot for circular variables. *Computational Statistics*, pages 1–12.

Ashour, S. and El-Adl, Y. (1980). Bayesian estimation of the parameters of the extreme value distribution. *Egyptian Statistical Journal*, 24:140–152.

Aslam, M. and Khurshid, A. (1991). Shape-finder box plots. *ASQC Statistics Division Newsletter*, pages 9–11.

Balakrishnan, N. and Chan, P. (1992). Order statistics from extreme value distribution, ii: best linear unbiased estimates and some other uses. *Communications in Statistics-Simulation and Computation*, 21(4):1219–1246.

Banerjee, S. and Iglewicz, B. (2007). A simple univariate outlier identification procedure designed for large samples. *Communications in StatisticsSimulation and Computation®*, 36(2):249–263.

Barrowman, N. J. and Myers, R. A. (2003). Raindrop plots: A new way to display collections of likelihoods and distributions. *The American Statistician*, 57(4):268–274.

Becker, R. A. and Cleveland, W. S. (1993). Discussion of graphic comparisons of several linked aspects by john w. tukey. *Journal of Computational and Graphical Statistics*, 2(1):41–48.

Becketti, S. and Gould, W. (1987). Rangefinder box plots: A note. *The American Statistician*, 41(2):149–149.

Benjamini, Y. (1988). Opening the box of a boxplot. *The American Statistician*, 42(4):257–262.

Berred, M. (1995). K-record values and the extreme-value index. *Journal of Statistical Planning and Inference*, 45(1-2):49–63.

Bowley, A. L. (1926). *Elements of statistics*. PS King, London, 5th edition.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*. SIAM, Philadelphia.

Bru, B. and Hertz, S. (2001). *Maurice Fréchet*, pages 331–334. Springer New York, New York, NY.

Bruffaerts, C., Verardi, V., and Vermandele, C. (2014). A generalized boxplot for skewed and heavy-tailed distributions. *Statistics & Probability Letters*, 95:110–117.

Brys, G., Hubert, M., and Struyf, A. (2004). A robust measure of skewness. *Journal of Computational and Graphical Statistics*, 13(4):996–1017.

Brys, G., Hubert, M., and Struyf, A. (2006). Robust measures of tail weight. *Computational Statistics and Data Analysis*, 50(3):733–759.

Cai, Y. and Hames, D. (2010). Minimum sample size determination for generalized extreme value distribution. *Communications in Statistics - Simulation and Computation*, 40(1):87–98.

Carr, D. B. (1994). A colorful variation on boxplots. *Statistical Computing & Statistical Graphics Newsletter*, 5(3):19–23.

Castillo, E. (1988). *Extreme Value Theory in Engineering*. Academic Press, London.

Choonpradub, C. and McNeil, D. (2005). Can the box plot be improved. *Songklanakarin Journal of Science and Technology*, 27(3):649–657.

Christopeit, N. (1994). Estimating parameters of an extreme value distribution by the method of moments. *Journal of Statistical Planning and Inference*, 41(2):173–186.

Cohen, D. J. and Cohen, J. (2006). The sectioned density plot. *The American Statistician*, 60(2):167–174.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Lecture Notes in Control and Information Sciences. Springer, London.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

Cramér, H. (1953). Richard von mises' work in probability and statistics. *The Annals of Mathematical Statistics*, 24(4):657–662.

Davies, L. and Gather, U. (1993). The identification of multiple outliers. *Journal of the American Statistical Association*, 88(423):782–792.

Davis, R. and Resnick, S. (1984). Tail estimates motivated by extreme value theory. *The Annals of Statistics*, pages 1467–1487.

Dekkers, A. L., Einmahl, J. H., and De Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, pages 1833–1855.

Dietrich, D. and Hüsler, J. (1996). Minimum distance estimators in extreme value distributions. *Communications in Statistics-Theory and Methods*, 25(4):695–703.

Dodd, E. L. (1923). The greatest and the least variate under general laws of error. *Transactions of the American Mathematical Society*, 25(4):525–539.

Dupuis, D. and Field, C. (1998). A comparison of confidence intervals for generalized extreme-value distributions. *Journal of Statistical Computation Simulation*, 61(4):341–360.

Efron, B. and Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1(1):54–75.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge University Press.

Frigge, M., Hoaglin, D. C., and Iglewicz, B. (1989). Some implementations of the boxplot. *The American Statistician*, 43(1):50–54.

Goldberg, K. M. and Iglewicz, B. (1992). Bivariate extensions of the boxplot. *Technometrics*, 34(3):307–320.

Groeneveld, R. A. and Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician*, pages 391–399.

Hertz, S. (2001). *Ladislaus von Bortkiewicz: Statisticians of the Centuries*, pages 273–277. Springer New York, New York, NY.

Hill, B. M. et al. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5):1163–1174.

Hintze, J. L. and Nelson, R. D. (1998). Violin plots: a box plot-density trace synergism. *The American Statistician*, 52(2):181–184.

Hoaglin, D., Mosteller, F., and Tukey, J. (1983). *Understanding Robust and Exploratory Data Analysis*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, New York.

Hoaglin, D. C., Iglewicz, B., and Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396):991–999.

Hosking, J. R., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.

Hössjer, O. (1996). Incomplete generalized l-statistics. *The Annals of Statistics*, 24(6):2631–2654.

Hubert, M. and Van der Veeken, S. (2008). Outlier detection for skewed data. *Journal of Chemometrics*, 22(3-4):235–246.

Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis*, 52(12):5186–5201.

Hyndman, R. J. and Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, 50(4):361–365.

Hyndman, R. J. and Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45.

Iglewicz, B. and Banerjee, S. (2001). A simple univariate outlier identification procedure. In *Proceedings of the annual meeting of the american statistical association*, volume I, pages 1–4, Atlanta, USA. American Statistical Association, JSM Proceedings.

Ingelfinger, J. A., Mosteller, F., Thibodeau, L. A., and Ware, J. H. (1983). *The visual display of quantitative information*. Macmillan Publishing Co, New York.

Jansen, D. W. and De Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics*, 73(1):18–24.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions, vol. 2*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, New York.

Kampstra, P. et al. (2008). Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software*, 28(1):1–9.

Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, 25(8):1287–1304.

Kimber, A. (1990). Exploratory data analysis for possibly censored data from skewed distributions. *Applied Statistics*, pages 21–30.

Kovalenko, I. N. (2014). Boris vladimirovich gnedenko: a classical scholar in probability, statistics, queueing, and reliability. *Queueing Systems*, 76(2):113–124.

Lye, L., Hapuarachchi, K., and Ryan, S. (1993). Bayes estimation of the extreme-value reliability function. *IEEE Transactions on Reliability*, 42(4):641–644.

Mandel, M. and Betensky, R. A. (2008). Simultaneous confidence intervals based on the percentile bootstrap approach. *Computational Statistics and Data Analysis*, 52(4):2158–2165.

Maritz, J. S. and Munro, A. H. (1967). On the use of the generalised extreme-value distribution in estimating extreme percentiles. *Biometrics*, 23(1):79–103.

Marmolejo-Ramos, F. and Siva Tian, T. (2010). The shifting boxplot. a boxplot based on essential summary statistics around the mean. *International Journal of Psychological Research*, 3(1).

McGill, R., Tukey, J. W., and Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1):12–16.

McNeil, D. R. (1977). *Interactive data analysis: A practical primer*. Wiley, New York.

Méndez, F. J., Menéndez, M., Luceño, A., and Losada, I. J. (2007). Analyzing monthly extreme sea levels with a time-dependent gev model. *Journal of Atmospheric and Oceanic Technology*, 24(5):894–911.

Mukhopadhyay, N. and Ekwo, M. (1987). A note on minimum risk point estimation of the shape parameter of a pareto distribution. *Calcutta Statistical Association Bulletin*, 36(1-2):69–78.

Musah, A.-A. I. (2010). *Application of Extreme Value Theory for Estimating Daily Brent Crude Oil Prices*. Thesis, Kwame Nkrumah University of Science and Technology Kumasi.

Muth, S. Q., Potterat, J. J., and Rothenberg, R. B. (2000). Birds of a feather: using a rotational box plot to assess ascertainment bias. *International Journal of Epidemiology*, 29(5):899–904.

Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of statistics*, 8(3):154–168.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Potter, T. D. and Colman, B. R. (2003). *Handbook of weather, climate, and water: atmospheric chemistry, hydrology, and societal impacts*. Wiley-Interscience, Hoboken, N.J. OCLC: 123158897.

Prescott, P. and Walden, A. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724.

Qarmalah, N. M., Einbeck, J., and Coolen, F. P. (2016). k-boxplots for mixture data. *Statistical Papers*, pages 1–16.

Reiss, R. D. (1989). Approximations to distributions of extremes. In *Approximate Distributions of Order Statistics*, pages 151–205. Springer, New York.

Rossi, F., Fiorentino, M., and Versace, P. (1984). Two-component extreme value distribution for flood frequency analysis. *Water Resources Research*, 20(7):847–856.

Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999). The bagplot: a bivariate boxplot. *The American Statistician*, 53(4):382–387.

Seki, T. and Yokoyama, S. (1996). Robust parameter-estimation using the bootstrap method for the 2-parameter weibull distribution. *IEEE Transactions on Reliability*, 45(1):34–41.

Sim, C. H., Gan, F. F., and Chang, T. C. (2005). Outlier labeling with boxplot procedures. *Journal of the American Statistical Association*, 100(470):642–652.

Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, pages 367–377.

Spencer, C. S. and McCuen, R. H. (1996). Detection of outliers in pearson type iii data. *Journal of Hydrologic Engineering*, 1(1):2–10.

Stephenson, A. G. (2002). evd: Extreme value distributions. *R News*, 2(2):31–32.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334.

Tawn, J. A. (1992). Estimating probabilities of extreme sea-levels. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(1):77–93.

Thomas, B. M. and Nolan, J. D. (1997). Colorado extreme storm precipitation data study.

Tippett, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, 17(3/4):364–387.

Tufte, E. R. (2001). *The visual display of quantitative information*. Graphics Press, Cheshire, Conn, 2nd edition.

Tufte, E. R. and Graves-Morris, P. (1983). *The visual display of quantitative information*, volume 2. Graphics Press, Cheshire, Conn.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley Pub. Co, Reading, Mass.

Tukey, J. W. (1993). Graphic comparisons of several linked aspects: Alternatives and suggested principles. *Journal of Computational and Graphical Statistics*, 2(1):1–33.

Tukey, J. W. et al. (1990). Data-based graphics: visual display in the decades to come. *Statistical Science*, 5(3):327–339.

Velleman, P. F. and Hoaglin, D. C. (1981). *Applications, basics, and computing of exploratory data analysis*. Duxbury Press, Boston, Mass.

von Zwet, W. R. (2012). *Convex transformations: A new approach to skewness and kurtosis*, pages 3–11. Springer New York, New York, NY.

Wickham, H. and Stryjewski, L. (2012). 40 years of boxplots. Technical report, had.co.nz.

Woodbury, G. (2002). *An Introduction to Statistics*. Available Titles CengageNOW Series. Duxbury Canada.

Zani, S., Riani, M., and Corbellini, A. (1998). Robust bivariate boxplots and multiple outlier detection. *Computational Statistics & Data Analysis*, 28(3):257–270.