



# A GRAPHICAL APPROACH FOR ANALYSES OF DATA THIN NON-PARAMETRIC CONTINUOUS VARIABLE OF *BOTIA DARIO* WITH R PROGRAMMING LANGUAGE

***Bhakta Supratim Sarker***<sup>1</sup>, ***Shyamal Kumar Paul***<sup>2</sup>,  
***Kawser Kadir Maruf***<sup>3</sup>, ***Priyanka Rani Majumdar***<sup>4</sup>,  
***Golam Azom***<sup>5</sup>, ***Debasish Saha***<sup>6</sup>

<sup>1</sup> ORCID: 0000-0002-5108-4553

<sup>2</sup> ORCID: 0000-0003-0655-1623

<sup>4</sup> ORCID: 0000-0003-2220-9852

<sup>5</sup> ORCID: 0000-0003-3197-6658

<sup>6</sup> ORCID: 0000-0003-0223-4188

<sup>1–4,6</sup> Department of Fisheries and Marine Science, Faculty of Science

<sup>2</sup> Research Cell

Noakhali Science and Technology University, Noakhali, Bangladesh

<sup>5</sup> Department of Biology (Ecology and Evolutionary Biology Discipline)  
University of Oklahoma, USA

Key words: graphicacy, dot chart, gaussian kernel, sheather-jones bandwidth, distribution pattern.

## Abstract

*Botia dario* is categorized as endangered due to considerable drop in population over past two decades creates data thin condition where the nonparametric statistical methods are superior alternative approach for data drafting. Small data set necessitate graphical display reducing the chance of data compression by numerical analyses. Monthly mass and length density estimates, location, and spread were compared with R computing environment through charts and plots; proposing a graphical method for single discrete and continuous data analyses. Through graphicacy the novel method reveals the pattern for mass and length of *B. dario* by depicting the modes and skews of the kernel density estimates suggested wide fluctuations during pre-monsoon months; whereas the spreads and locations of boxplots draping dot-whiskers infer the Gaussian kernel by pairwise comparisons. The boxplot widths, notches of the boxplots and red dot-whiskers illustrate comprehensive variations. The novel method and suggestive narratives appeal for inclusive use excluding the limitations.

## Introduction

Literacy (communicate with written information), numeracy (communicate with numerical information) and graphicacy (communicate with graphical information) are the central ability of human cognitive inheritance (KEEN 2010). Perhaps the latter came first, as the ancient cave paintings indicated. Statistical data should be presented in such that it consumes little space and ink (TUFTE 2001), and at the same time accumulate “more relevant information within eye span” “to make a contrast, a comparison, a choice” (TUFTE 1990). Data reduction (calculating the central tendency and dispersion) based on numerical information is a common statistical process which can be balanced through graphical methods avoiding probable information loss and maintaining data expression (KEEN 2010). Data compression methods like mean could sometimes lead to misinterpretation of the population due to reduced number of sample and/or sampling error; commonly are being faced for some fish species whose numbers are dwindling and very difficult to capture, resulting poor representation of the population. On the other hand, graph can communicate most efficiently and represents the sample data as a whole to make an effective conclusion. The insights of graphicacy allow the observers to adhere coherent information from a chart and/or a plot precisely than a numerical table.

The Queen loach *Botia dario* (HAMILTON 1822) is a very popular SIS (Small Indigenous Species) fetching higher price in Bangladesh. Unfortunately, the natural population of the species is under threat as the population noticeably has dropped over the past two decades and now being considered as an ‘endangered’ species (IUCN Bangladesh 2015). Not only *B. dario* but a total of 64 fish species are being facing different levels of threats in Bangladesh (NEUMANN and ALLEN 2007) which ultimately demands more research on those to restore the population and at the same time, forces the researcher to work with small number of individuals (sample). Sometimes, it was quite difficult to collect representative samples of specific categorical variable while considering natural populations. A parametric statistical method based on poorly representable sample (without bootstrap or Monte Carlo simulations) might mislead succeeding models and inferences. An exploratory plot for the distribution of single continuous variable is empirical distribution function (EDF). The length frequency distributions are popularly presented in the fisheries literatures with the Kolmogorov-Smirnov (K-S) two-sample test (NEUMANN and ALLEN 2007) which determines whether the EDF are similar between groups or significantly different in the location (central tendency), spread

(data dispersion) and shape of the distributions (HOLLANDER et al. 2014). Again, if the discrete variable had more groups, then the pairwise comparison is strenuous to explain. For example, if the complied data require twelve groups' pairwise comparison (as in this study), then there are sixty-six  $p$ -values obtained by K-S two-sample test. Furthermore, when the comparison is extended to compare with the global location; at that point additional twelve pairwise comparisons would have added with mounting difficulties to conclude fairly. This numerical convention for data analysis might be replaced by graphical convenience smoothly. Therefore, Kernel density estimate – a nonparametric method (without assuming a model and estimating parameters from data), is probably the best alternative for single continuous variable; while variable width notched boxplots and dot whisker plots could be the best choice for comparing among categorical group variables. BURN (1993) offers an immaculate mnemonic, ACCENT for six principles (apprehension; clarity; consistency; efficiency; necessity; and truthfulness) of effective graphical communication. For producing effective graph following the ACCENT rules R programming languages (R Core Team 2021) with versatile packages can act as a prospective tool. Moreover, the inherent properties of R help to inspect the codes of complex statistical concepts step by step to get the results and the graphical output as well (AHO 2014 on the Preface page xix).

Although, there are few studies have been conducted on conservational view of *B. dario* (RAHMAN 2015, HOSSAIN et al. 2015, MAHFUJ et al. 2022), stock structure (MAHFUJ et al. 2022), and on sex ratio and fecundity (HUSSAIN 2007) but authors failed to found any on nonparametric method for any parameter of this species; ultimately attracted the authors to use the species as a model in this work. Conceiving a novel nonparametric approach with R for single continuous variables (mass and length of *B. dario*) separately compared by one discrete grouping variable (months) is imperious to transform conventional numerical analysis into effective graphical platform. The novel approach allows readers, researchers, and planners to get coherent information from small data set through the primitive inherent graphicacy and to conclude from heuristic to holistic one.

## Materials and Methods

### Study site, sample collection and recording of mass and length

Samples were collected 3 points of the ‘Sunamganj Beel’ (a lake-like wetlands with static waters might have channel to connect to river) namely ‘Dakshin Sunamganj’ (24°55’52.0”N 91°26’53.2”E), ‘Balaganj’ (24°39’13.1”N 91°49’35.8”E) and ‘Raznogor’ (24°39’08.9”N 91°49’55.1”E) of northeastern Bangladesh during August 2018 to July 2019. HAMILTON (1822) recorded the first specimen of *B. dario* from the northern rivers of Bengal and latter all other taxonomists (TALWAR and JHINGRAN 1991, RAHMAN 2005, RAHMAN and RUMA 2007) described this region as native of *B. dario*. The three points are connected by the Kushiya River with its tributaries and distributaries. Moreover, the ‘Sunamganj Beel’ is usually inundated in the monsoon, seemingly the entire area under water. Therefore, the site was considered to collect the species. A total of 226 individuals were captured by fine meshed seine net (Bengali name: Kapuri jal) and were preserved immediately with 10% formalin solution to avoid any spoilage. The samples were then transported to the laboratory of the Department of Fisheries and Marine Science, Noakhali Science and Technology University for further analysis. Mass and length were measured as whole-body mass to the nearest 0.01 g by an electrical balance (EK600 Dual) and total length (TL) to nearest 0.01 cm using a digital slide calipers having  $\pm 0.01$  mm accuracy (EAGems-B00Z5KETD4) following the instructions of JENNINGS et al. (2012).

### Months (discrete grouping variable)

Although the sampling was started in August 2018 and closed in July 2019, the order of months in the graphical displays (charts and plots) were presented as March to February following the rain dependent seasonal pattern viz., Mar.–May (pre-monsoon), June–Aug. (monsoon), Sept.–Nov. (post-monsoon) and Dec.–Feb. (winter).

### Dot chart

The dot chart (Fig. 1) was constructed sharing (with required modifications) the codes of KEEN 2010 which followed the inferred plotting standard of CLEVELAND and MCGILL 1984. The modifications were:

1. Ticks on the vertical axis for clear linking with data point to corresponding group labels.

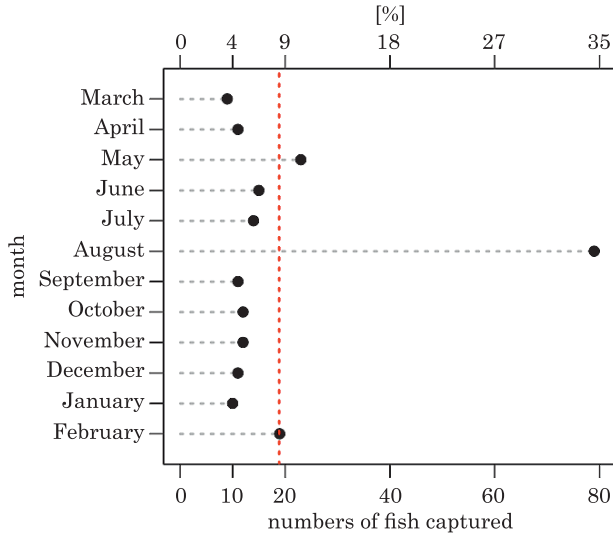


Fig. 1. Dot chart for numbers and percentages of *B. dario* captured from 'Sunamganj Beel' from August 2018 to July 2019. Month's order (top to bottom) started at the onset of pre-monsoon and end at the end of winter season. Vertical dashed red line indicating the mean of total captured. Ticks added on the horizontal axis at the data dense zone

2. Additional horizontal axis on top coordinate to see corresponding percentage of the values on the bottom coordinate.
3. Vertical dotted red line to mark the mean level of the count.
4. Additional ticks on both horizontal axes to compare the dot dense region.
5. Y-axis label to indicate the discrete grouping variable. The detailed code script is attached in Appendix 1.

### Trellis kernel density plots

Kernel density plots (Fig. 2 and Fig. 3) were produced using `densityplot()` function of **lattice** package with the syntaxes `kernel="gaussian"` for Gaussian kernel estimation and `plot.points="rug"` for adding rug on horizontal axis beneath the curve. WAND and JONES (1995) suggested objective method (automatic data dependent) for selecting bandwidth (`bw.SJ()`) SHEATHER and JONES (1991) were applied as it was explained that the subjective selection was time consuming and there was no idea about the outline of the distribution. The detailed code script is obtainable in Appendix 2 and Appendix 3.

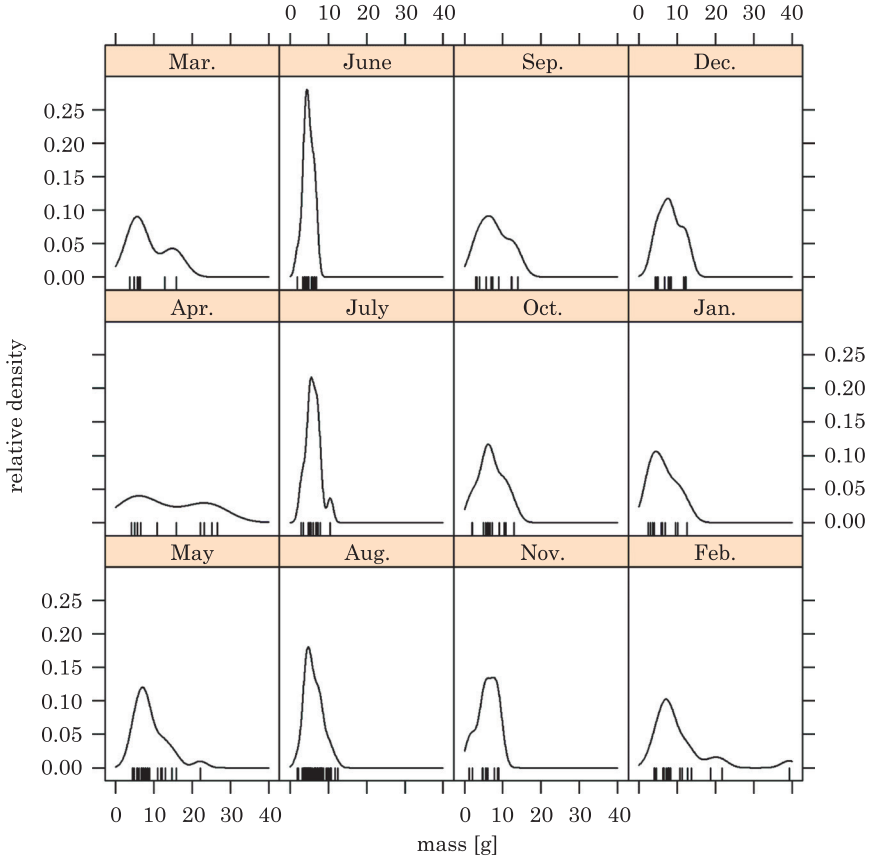


Fig. 2. Nonparametric kernel density plots for mass [g] of *B. dario* captured from ‘Sunamganj Beel’ from August 2018 to July 2019. The short names of the months in the strips ordered column-wise (top to bottom then to right). Each column represents a season viz. pre-monsoon, monsoon, post-monsoon and winter. Gaussian kernel with data dependent bandwidth is used to estimate the relative density

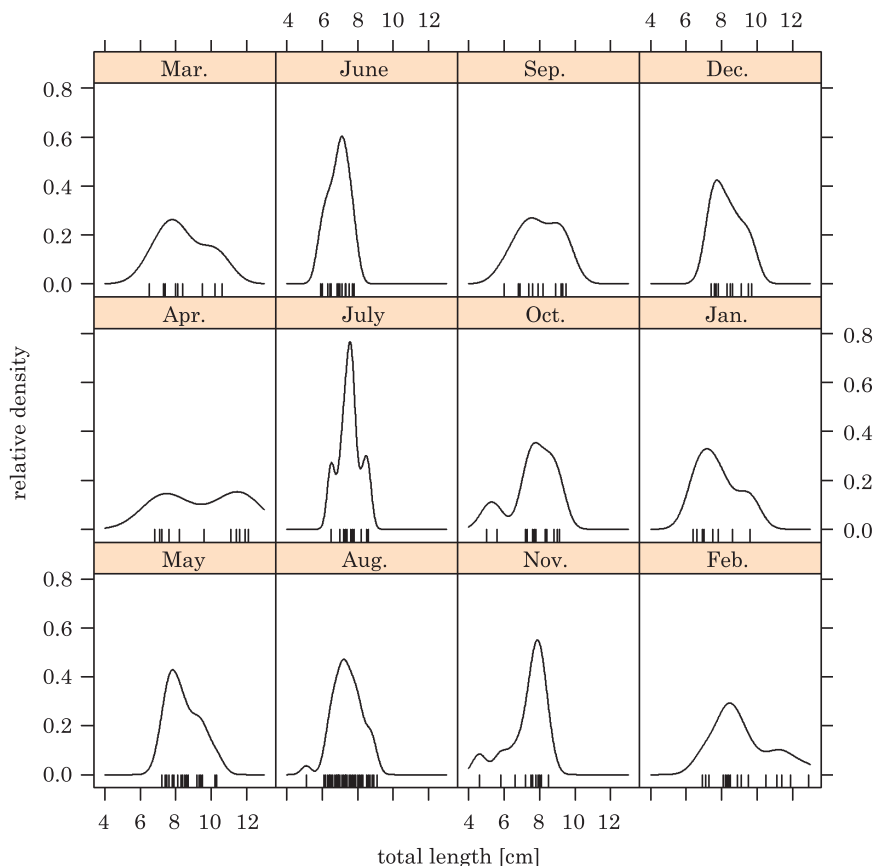


Fig. 3. Nonparametric kernel density plots for length [cm] of *B. dario* captured from 'Sunamganj Beel' from August 2018 to July 2019. The short names of the months in the strips ordered column-wise (top to bottom then to right). Each column represents a season viz. pre-monsoon, monsoon, post-monsoon and winter. Gaussian kernel with data dependent bandwidth is used to estimate the relative density

### Variable-width notched boxplot draping dot-whiskers

Variable-width notched outlier boxplots draping dot-whiskers (Fig. 4 and Fig. 5) were constructed using the codes of KEEN (2010) (function `stripchart()` was omitted, as `boxplot()` call active) after some modifications. The modifications were:

1. Variable-width notched boxplot and dot-whisker plot drafted within a single frame to compare between the locations (median and mean) and their spreads (standard error) simultaneously; and to assume their skew (without observing the density plot).

2. Total sample plotted (box plot, `add=TRUE` syntax added in the second plot call) with a light gray separation line as same visual like the box plot to compare between each group and total.

3. For both locations a reference line (median and mean of total marked by vertical dotted black line and red line, respectively) drawn to compare among each groups fairly definite with an ease.

4. Ticks added on the horizontal axis at a rough start (Fig. 5) and end (Fig. 4 and Fig. 5) of the values of observations to predict the ranges.

5. Logarithmic scale (argument `log="x"` inside plot call) added for the mass on the Figure 4 for superior comparison. The detailed code script is accessible in Appendix 4 and Appendix 5.

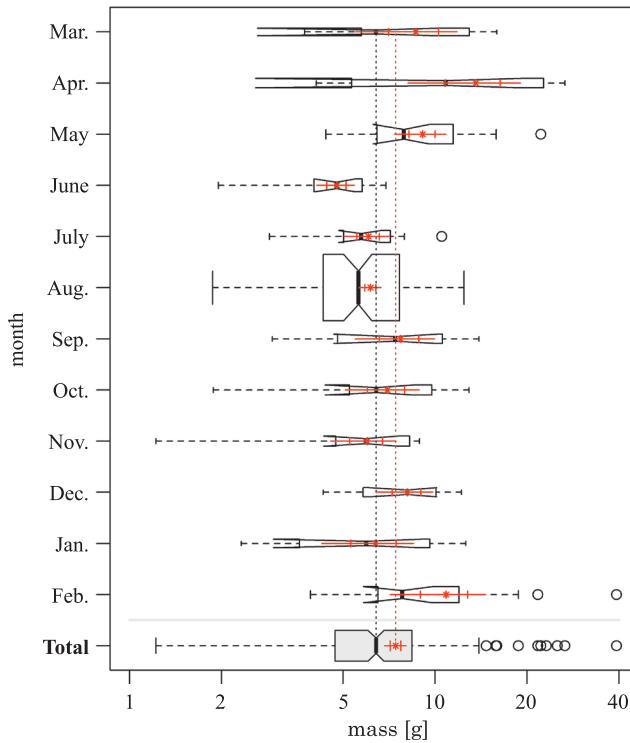


Fig. 4. Variable-width notched boxplot (box width proportional to sample size, except for the total, which was added with the syntax `add = TRUE`, where the proportion argument is inapplicable) draping dot-whiskers for monthly mass [g] with logarithmic scale of *B. dario* caught in ‘Sunamganj Beel’ from August 2018 to July 2019



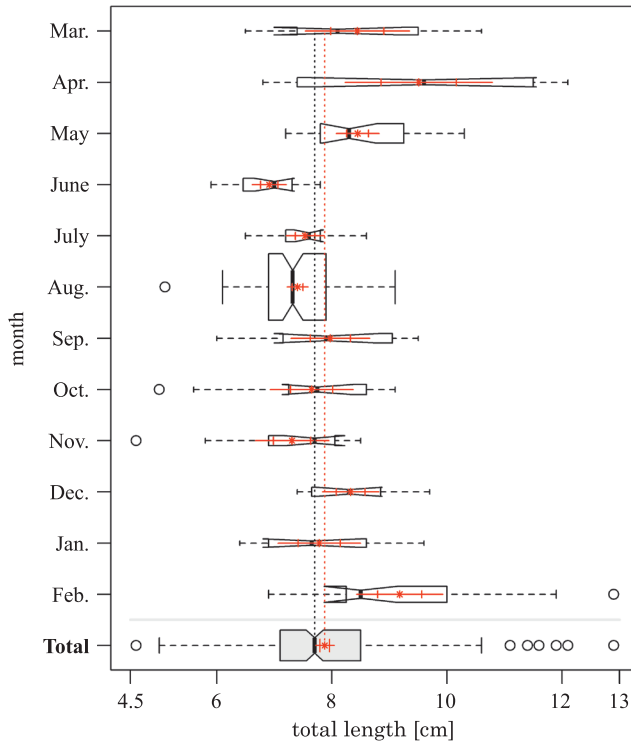


Fig. 5. Variable-width notched boxplot (box width proportional to sample size, except for the total, which is added with the syntax `add = TRUE`, where the proportion argument is inapplicable) draping dot-whiskers for monthly length [cm] of *B. dario* caught in 'Sunamganj Beel' from August 2018 to July 2019

## Data analyses

Raw data were entered into a spreadsheet (MS Excel 2016) and save as .csv file extension. The CSV file act as data frame (accessible in Appendix 5 – Table Appx. 5.1) for further data analyses using the R programming language (R Core Team 2021). Executions of basic stats were performed using core R, **dplyr** (WICKHAM et al. 2021) and **FSA** (OGLE et al. 2021). Package **graphics** (built in base R) and **lattice** (SARKAR 2008) were used to create and to modify charts and plots.

## Results

### Dot chart

Figure 1 is the dot chart corresponding the number of *B. dario* caught from 'Sunamganj Beel' by month. The highest occurrence (capture) was recorded on August (nearly 35%) and the lowest on March (nearly 4%). Only three points i.e., May, August and February were above to right) the mean of total captured.

### Trellis kernel density plots

Nonparametric kernel density plots for both mass (Fig. 2) and length (Fig. 3) exhibited pattern similarities with minor modal variations in column two to four. Those columns represented monsoon, post-monsoon and winter season, respectively. Most of the captured individuals during monsoon were found around 5 g (mass) and below 8 cm (length) constructing high narrow peaks for the both kernel density estimates indicating spawning came about at the end of pre-monsoon.

On winter season, although the majority of observed mass showed similarities with monsoon and post-monsoon but the length kernel density estimate made a bolder right skewed mode of above 8 cm. Considering the mass kernel density estimate, there was another spawning season possible at late post-monsoon whereas, the length kernel density estimate suggested the stock is struggling for food and space on winter months. In the first column (pre-monsoon), the positive skewed mode of over 20 g (Fig. 2) and 10 cm (Fig. 3) individuals on March indicated the stock was preparing for spawning and on April, was fully prepared. Therefore, the second mode of those bimodal kernel density estimates indicated the gravid dames. The relatively wider left peak modes for both kernel density estimates on May signposted spawning has been started.

### Variable-width notched boxplot draping dot-whiskers

Noticing both the plots (Fig. 4 and Fig. 5), it was observed that there were wide variations considering the standard deviation of the median (length of the notches) and standard error of the mean (red dot-whiskers) over the twelve months. Projecting the comparison of the notches for months it was found to be similar considering the median mass for all months except June. Whereas, all the months including the total were similar except June, August and February with respect to the median length ( $p < 0.05$ ). Considering median mass, approaching from left to right

a similar group of nine months excluding May, December and February were observed, while moving from right to left a group of ten months except June and August was revealed. Repeating the same considering median length, starting from left to right a similar group of eight months excluding April (additional see mass), May, December and February was found. While moving from right to the left a group of nine months except June, July (addition, see mass) and August was revealed ( $p < 0.05$ ).

On the other hand, all months including the total except April, June and August were found to be similar while projecting the comparison of the whiskers for months considering the mean mass. Whereas, all the months including the total were similar except April, May, June, August and February with respect to the mean length ( $p < 0.05$ ). For mean mass, approaching from left to right a similar group of five months (June, July, October, November and January) and moving from right to the left a group of eight months (except June, July, August and November) were found. Repeating the same for mean length, a group of four months (June, October, November and January) and a group of seven months (except June, July, August, October and November) were revealed ( $p < 0.05$ ).

Pairwise differences of locations (median by notches of the variable width boxplot and mean by red dot-whiskers) between months showed May and June were significantly ( $p < 0.05$ ) different for both mass and length.

Length of the notches is centered about the sample median and overlapping notches for any pair of months indicated no differences in sample medians ( $p < 0.05$ ). Dot-whiskers with red star dots give the locations for the sample means and the red bars sliced the red whiskers on both sides marked the length of one standard error, and overlapping red whiskers for any pair of months indicated no differences in sample means ( $p < 0.05$ ).

## Discussion

Although few other packages (ggplot2 (WICKHAM 2016), plotly (SIEVERT 2020), ggpubr (KASSAMBARA 2020) are available brilliantly to cope with R for plotting convention but in many cases the routine functions like `par()`, `axis()`, `points()` and `lines()` are not compatible with the high level plot functions of other packages. Those plot functions required their package specific syntaxes. That is why; the authors limit themselves within the base or core R and incorporation of the same function of other packages is beyond the scope of this article.

### Dot chart

The peak count on August suggested the spawning of *B. dario* commenced prior to August, presumably at the end of pre-monsoon which was fairly impossible to conclude without the mass and length density estimates. There was a third peak on late winter and whether was it indicated another spawning incidence or higher fishing pressure as a result of lower water level or was there any monthly (temporal) pattern evident? We try to resolve the queries as discussion progress.

Considering the data points on the chart the reference vertical red dotted mean line acted as the balance of the total vales of the dashed gray line portion. The large percentages (nearly 35%) for August forced the majority of the points left to the mean line, which might be a poor representation of the data set. However, this sampling majority did not show any recognizable influence further except slim width of the other boxplots; while the nonparametric protocol for graphical convenience was applied. Most of the dots in the dot chart showed close vicinity of the mean while excluding the count of August. Moreover, KEEN 2010 stated the principle of the central limit theorem which suggested “even for long-tailed distributions with sample sizes as small as ten” sufficient for “quite good” approximation. In the current study the sample sizes were “as small as ten” in few months which might not apprehend “quite good” deduction as the central limit theorem comprehend.

The modified dot chart is proposed for easy comparisons between the count and the percentages of the count specifically indicating the ratio of total by two horizontal axes; and additional ticks on the data dense portions in horizontal coordinates for comparing among the groups of the discrete variable on the vertical coordinate. Dot plot tick for vertical categories were absent while CLEVELAND and MCGILL 1984 designed it whereas, the proposed dot chart recommended it for better linking with the gray dashed line terminating the dots. Moreover, the dot plot is in the top most position leaving bar chart, pictograph, pie chart etc. as KEEN 2010 suggested hierarchy of preference for single discrete variable.

### Trellis kernel density plots

Nonparametric density estimates for monthly kernels pictured the seasonal influences over both mass and length giving a reasonably acceptable temporal pattern. Unimodal kernels of monsoon appeared to be smooth bimodal during post-monsoon and winter (except February) predicting a seasonal pattern similarity. Comparatively higher water levels during monsoon allowing the drifting stock more susceptible to be caught;

while progressing towards winter coupled with low water levels made possible to catch the bigger fish, as well. From February to May sharp decline of water level in Bangladesh context left the population into extreme fishing pressure depicted by the graduating right skewed bimodal curves.

The kernel density estimates were plotted four columns by three rows following the argument `layout=c(4, 3)` inside the `densityplot()` function. The month order was rearranged for showing the seasonal progress from top to bottom to right. Conventionally, the lattice package arranges the strip categories from left to right to up. The future researchers need to arrange their own layout and order according to the data set and the way they ought to present.

For selecting the kernel density estimator (the kernel and the bandwidth) the current study shared the view and precautions discussed by KEEN (2010) and after a critical examination the estimator was chosen. For comparison, the authors suggest the scholars to replace the argument `kernel="gaussian"` by their desired one ("`epanechnikov`", "`rectangular`", "`triangular`", "`biweight`", "`cosine`" and "`optcosine`"). And, for the bandwidth, one can try to check the changes in the plot by inserting the `bw=value` argument inside the `densityplot()` call after obtaining the value (number) by the functions `bw.SJ`, `bw.nrd0`, `bw.nrd`, `bw.bcv` or `bw.ucv`. Running different combinations of estimators, the current study relied on the "## sensible automatic choice" (R Core team 2021). Researcher should be strictly cautious while changing the default to avoid drafting a plot like hornbill beak.

### Variable-width notched boxplot draping dot-whiskers

Apart from finding the monotonous grouping similarities and pairwise comparisons in terms of spread with respect to the locations those plots (Fig. 4 and Fig. 5) might allow the spectator to pursue the overall scenario of the sampling distribution at a glance. The design of the plots discovered the layering and separation resemblances with the calligraphy of Uboku Nishitani when TUFTE (1990) wrote, "The saturated quality of the red partially offsets its lighter value and finer line (appropriate to meticulous annotation). Alone, each color makes a strong statement; together, a stronger one". In the graphical display the reader has the freedom of choice to compare locations, equate group ratio, form groups, match groups, identify skew, detect outliers, separate groups and might be more, "makes a strong plot; together, a stronger one".

In the boxplot draping the red dot-whiskers of November revealed an interesting puzzle. Putting the outlier inactive in the mass made a sym-

metric boxplot (in the logarithmic scale the left whisker appeared longer) by merging the locations seemingly within a point while, in the length the outlier act strongly by pulling mean apart from median. This marks the assumption about the second possible breeding during late post-monsoon under reconsideration.

The locations (median and mean) of mass was found the largest in April (although the outliers in February projected the larger individuals), while the smallest was in June. The scenarios were same for the locations of lengths except in November where the outlier represented the smallest individual.

## Conclusion

By the course of time through the results and the discussions the novel graphical displays showed their statistical analytical capabilities to figure out the arising questions inferentially. For plotting single discrete variable, the dot chart with proposed modified design could be a superior alternative, which is sufficient to compare among group aggregated values or counts or proportions. With respect to current study, the counts with percentages allow one to decipher the monthly capture pattern comparing with the mean value. The trellis density plot requests the managers for improving the threat status of *B. dario* to take initiatives intensively during late dry to early wet by minimizing recruitment overfishing. The scholars have liberty to replace the kernel and bandwidths and have choice to rearrange the layout to make their own. The ability of the boxplot draping dot-whiskers shown by its comparison with level of significance and separation by location reference lines. The versatility of the plot allow the users to control the significance level, reshape the box, change the scale etc. by adjusting the spread (standard error) algorithms, by muting and allowing specific syntaxes to run. Hope the modified designs and the analytical narratives of the proposed charts and plots will be accepted for graphical displays of single discrete and continuous variables.

## Applications and Limitations

The scripts (detailed codes) provided here with the supplementary .txt extension files are self-explanatory. The purposes of each succeeding codes were annotated with conventional #-marks. The readers have liberty to edit the code according to their requirements. The `par()` settings were

adjusted to accommodate the labels and leavings spaces for aesthetics; while, the `png()` function for the output image ratio. If the future user wish to control the plot ration rather than the image output then s/he would require to adjust the `pin`, `fin`, `mai` and `omi` arguments inside the `par()` function (MURRELL 2019).

Like other nonparametric methods this is robust for parametric applications also. Prediction of interactions among groups of a discrete variable is not possible through the novel graphical display.

## Acknowledgements

The authors acknowledge the NATP-2, Bangladesh Agriculture Research Council, Dhaka, Bangladesh for the Research grants (NATP-2/PIU-BARC/Research CRG/2017/553) awarded during the study period. Authors would like to appreciate the efforts of the reviewers and offering their heartfelt gratitude to make the article better.

Accepted for print 27.05.2022

## References

- AHO K.A. 2014. *Foundational and applied statistics for biologists using R*. Chapman & Hall/CRC, Boca Raton, FL, 1–148.
- BURN D.A. 1993. *Designing effective statistical graphics*. Computational statistics, Handbook of Statistics, 9, Chapter 22, pp. 745–773. Elsevier Science Publishers B.V., Amsterdam, The Netherlands.
- CLEVELAND W.S., MCGILL R. 1984. *Graphical perception: theory, experimentation, and application to the development of graphical methods*. J. Ame. Stat. Asso., 79: 531–554.
- HAMILTON F. 1822. *An account of the fishes found in the river Ganges and its branches*. Archibald Constable and Company, Edinburgh; and Hurst, Robinson, and Co. 90, Cheapside, London, <https://doi.org/10.5962/bhl.title.59540>.
- HOLLANDER M., WOLFE D.A., CHICKEN E. 2014. *Nonparametric statistical methods*. John Wiley & Sons, Inc., Hoboken, NJ, third edition, pp. 1–809.
- HOSSAIN M.Y., HOSSAIN M.A., AHMED Z.F., YAHYA K., RAHMAN M.M., AHAMED F., OHTOMI J. 2015. *Threatened fishes of the world: Botia dario (Hamilton, 1822) (Cypriniformes: Cobitidae)*. Croat. J. Fische., 73: 86–88, <https://doi.org/10.14798/73.2.812>.
- HUSSAIN M.A., KHATUN M.R., HOSSAIN M.A. 2007. *On the fecundity and sex-ratio of Botia dario (Hamilton) (Cypriniformes: Cobitidae)*. Univer. J. Zoo. Raj. Univer., 26: 27–29, <https://doi.org/10.3329/ujzru.v26i0.693>.
- IUCN Bangladesh, 2015. *Red list of Bangladesh: freshwater fishes*. IUCN, International Union for Conservation of Nature, Bangladesh Country Office, Dhaka, Bangladesh, 5: pp xvi+360.
- JENNINGS C.A., SLOSS B.L., LASEE B.A., BURTLE G.J., MOYER G.R. 2012. *Care, handling, and examination of sampled organisms*. Eds. A.V. Zale, D.L. Parrish, T.M. Sutton, A. Kassambara, 2020. ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0, <https://CRAN.R-project.org/package=ggpubr>.

- KASSAMBARA A. 2020. *ggpubr: 'ggplot2' based publication ready plots*. R package version 0.4.0, <https://CRAN.R-project.org/package=ggpubr>.
- KEEN K.J. 2010. *Graphics for statistics and data analysis with R*. Chapman and Hall/CRC, Taylor and Francis Group, Boca Raton, FL:1–443.
- MAHFUJ M.S., AHMED F.F., HOSSAIN M.F., ISLAM S.I., ISLAM M.J., ALAM M.A., HOSHAN I., NADIA Z.M. 2022. *Stock structure analysis of the endangered Queen Loach, Botia dario (Hamilton 1822) from five rivers of Northern Bangladesh by using morphometrics. Implications for conservation*. *Fishes*, 7: 41, <https://doi.org/10.3390/fishes7010041>.
- MURRELL P.R. 2019. *Graphics*. Chapman and Hall/CRC, Taylor and Francis Group, Boca Raton, FL.
- NEUMANN R.M., ALLEN M.S. 2007. *Size structure*. In: *Analysis and interpretation of freshwater fisheries data*. Eds. C.S. Guy, M.L. Brown, chapter 9. Ame. Fish. Soc., Bethesda, MD, pp. 375–421, <https://doi.org/10.47886/9781888569773.ch9>.
- OGLE D.H., WHEELER P., DINNO A. 2021. *FSA: fisheries stock analysis*. R package version 0.8.32, <https://github.com/droglenc/FSA>, access: 12.12.2021.
- CORE TEAM. R. 2021. *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>, access 10.12.2021.
- RAHMAN A.K.A. 2005. *Freshwater fishes of Bangladesh*. (2<sup>nd</sup> ed.). Zoological Society of Bangladesh, Department of Zoology, University of Dhaka, Dhaka. p. 394.
- RAHMAN A.K.A., RUMA F. 2007. *Botia dario*. In: K.U. Siddiqui, M.A. ISLAM, S.M.H. Kabir, M. Ahmad, A.T.A. Ahmed, A.K.A. Rahman, E.U. Haque, Z.U. Ahmed, Z.N.T. Begum, M.A. Hassan, M. Khondker, M.M. Rahman. *Encyclopedia of flora and fauna of Bangladesh. Freshwater fishes*. Asia. Soc. Bang., Dhaka, 23: 102.
- RAHMAN M.M. 2015. *Botia dario*. In: *IUCN Bangladesh. Red list of Bangladesh: Freshwater fishes*. IUCN, International Union for Conservation of Nature, Bangladesh Country Office, Dhaka, Bangladesh, 5: 67.
- SARKAR D. 2008. *Lattice: Multivariate data visualization with R*. Springer, New York, pp. 119–132.
- SHEATHER S.J., JONES M.C. 1991. *A reliable data-based bandwidth selection method for kernel density estimation*. *J. R Stat. Soc. – Series B*, 53: 683–690.
- SIEVERT C. 2020. *Interactive web-based data visualization with R, plotly, and shiny*. Chapman and Hall/CRC Florida.
- TALWAR P.K., JHINGRAN A.G. 1991. *Inland fishes of India and adjacent countries*. A.A. Balkema, Rotterdam, 1: 541.
- TUFTE E.R. 1990. *Envisioning information*. Graphics Press, Cheshire, CT, pp. 1–126.
- TUFTE E.R. 2001. *The visual display of quantitative information*. Graphics Press, Cheshire, CT, pp. 13–191.
- WAND M.P., JONES M.C. 1995. *Kernel smoothing*. Chapman and Hall, London, pp. 1–224.
- WICKHAM H. 2016. *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York, [https://doi.org/10.1007/978-3-319-24277-4\\_9](https://doi.org/10.1007/978-3-319-24277-4_9).
- WICKHAM H., FRANÇOIS R., HENRY L., MÜLLER K. 2021. *dplyr: A grammar of data manipulation*. R package version 1.0.4, <https://CRAN.R-project.org/package=dplyr>, access: 10.12.2021.



## Appendix 1

### Supplementary file – Fig\_1\_dotchart.txt

```
# call file rest in working directory (wd)
dr <- read.csv("dario.csv")

# Ordered and factor
dr$Month <-
ordered(dr$Month, levels=c("Feb", "Jan", "Dec", "Nov", "Oct", "Sep",
                           "Aug", "Jul", "Jun", "May", "Apr", "Mar"))

# package required for new data frame
library(dplyr)
library(FSA)

# new data frame
tem <- dr %>% group_by(Month) %>%
  summarize(n=n(), val.n=validn(tl),
            mean=round(mean(tl, na.rm=TRUE), 1)) %>%
  as.data.frame()

# For the dot plot new
NUM <- tem$n # x axis value
(mean1 <- mean(NUM)) # call the value of reference line

# for yaxis label
y <- c("February", "January", "December", "November", "October", "Sep-
tember",
       "August", "July", "June", "May", "April", "March")

# opening the printing device to make image as .jpeg
png("Fig_1_dotchart.jpeg", width = 5, height = 4.5, units = "in-
", res=588)

# change font
par(family="serif")
# par set
par(mar=c(bottom=3.9, left=6.9, top=3.9, right=1.3)+0.1)

# THE dot Chart of Cleveland and McGill
plot(NUM, y, type="n", xaxt="n", yaxt="n", xlim=c(0, 80),
      ylim=c(0, length(y)+1), xlab="numbers of fish captured",
      ylab='', xaxs="r", yaxs="i")
for (i in 1:12) lines(x=c(0, NUM[i]), y=c(i, i), lty=3, l-
wd=1.8, col="gray")
```

```

points(x=NUM,y=1:12,pch=19,cex=1.0)
axis(1,at=20*(0:4),labels=TRUE,tick=TRUE,outer=FALSE)
axis(2,at=1:12,labels=y,tick=T,outer=FALSE,las=2,
     hadj=1,padj=0.3)

# adding extra horizontal axis
axis(3,at=20*(0:4),labels = round((20*(0:4))*100/226))
# addind axis level
mtext("[%]",side = 3,line = 3)
mtext("month",side = 2,line = 6)

# create vertical mean ref line
lines(c(mean1,mean1),c(0,length(y)+1),lty=3,lwd=1.8,col="red")

# additional ticks on side 1 and 3
axis(1,10)
axis(3,at=10,labels = round(10*100/226))

# image in wd
dev.off()

```

## Appendix 2

### Supplementary file – Fig\_2\_densityplotM.txt

```

# changing default font settings
font.settings <- list(
  font = 1,
  cex = 1,
  fontfamily = "serif")

my.theme <- list(
  par.xlab.text = font.settings,
  par.ylab.text = font.settings,
  axis.text = font.settings,
  sub.text = font.settings,
  add.text = font.settings)

# call file rest in working directory (wd)
dr <- read.csv("dario.csv")

# ordering the strip names for layout=c(4,3)
dr$Month <- ordered(dr$Month,levels=c(
  "May","Aug","Nov","Feb",
  "Apr","Jul","Oct","Jan",
  "Mar","Jun","Sep","Dec"))

# package required
library("lattice")

# opening the printing device to make image as .jpeg
png("Fig_2_densityplotM.jpeg",width = 6, height = 6,units = "in",res=288)

```

```
# Mass plot call
densityplot( ~wt | Month,data=dr,xlab="mass [g]",
             type="density",kernel="gaussian",
             na.rm=TRUE,plot.points="rug",col="black",
             ylab="relative density",layout=c(4,3),from = 0,to=40,
             par.settings=my.theme)

# image in wd
dev.off()
```

## Appendix 3

### Supplementary file – Fig\_3\_densityplotL.txt

```
# changing default font settings
font.settings <- list(
  font = 1,
  cex = 1,
  fontfamily = "serif")

my.theme <- list(
  par.xlab.text = font.settings,
  par.ylab.text = font.settings,
  axis.text = font.settings,
  sub.text = font.settings,
  add.text = font.settings)

# call file rest in working directory (wd)
dr <- read.csv("dario.csv")

# ordering the strip names for layout=c(4,3)
dr$Month <- ordered(dr$Month,levels=c(
  "May","Aug","Nov","Feb",
  "Apr","Jul","Oct","Jan",
  "Mar","Jun","Sep","Dec"))

# package required
library("lattice")

# opening the printing device to make image as .jpeg
png("Fig_3_densityplotL.jpeg",width = 6, height = 6,units = "in",res=288)

# Length plot call
densityplot( ~tl | Month,data=dr,xlab="total length [cm]",
             type="density",kernel="gaussian",
             na.rm=TRUE,plot.points="rug",col="black",
             ylab="relative density",layout=c(4,3),from = 4,to=13,
             par.settings=my.theme)

# image in wd
dev.off()
```

## Appendix 4

### Supplementary file – Fig\_4\_boxplotM.txt

```
# call file rest in working directory (wd)
dr <- read.csv("dario.csv")

# ordering the grouping sequences
dr$Month <- ordered(dr$Month, levels=c("Feb", "Jan", "Dec",
                                       "Nov", "Oct", "Sep",
                                       "Aug", "Jul", "Jun",
                                       "May", "Apr", "Mar"))

# opening the printing device to make image as .jpeg
png("Fig_4_boxplotM.jpeg", width = 5, height = 6, units =
    "in", res=588)

# changing default par settings
par(mar=c(bottom=3.9, left=3.9, top=1.3, right=1.3)+0.1)

# changing default font
par(family="serif")

# calculating the width ratio for boxplot
dr$Count<-dr$wt
dr$Count[dr$Count >= 0]<-1
dr$Count[is.na(dr$Count)== TRUE]<-0
sums<-tapply(dr[, "Count"], INDEX=dr[, "Month"], FUN=sum, na.rm=TRUE)
sums<-sums/max(sums)

# VARIABLE-WIDTH notched BOXPLOT call
boxplot(wt ~ Month, data=dr, ylim=c(1, 40), xlim=c(0, 12),
        xlab="mass [g]", ylab="month",
        names=levels(dr$Month), horizontal=TRUE, las=1,
        width=sums, notch=T, boxwex=1.3, col="white", log="x")

# calculating the mean and se for dot whiskers
means<-tapply(dr[, "wt"], INDEX=dr[, "Month"], FUN=mean,
              na.rm=TRUE)
sdf<-tapply(dr[, "wt"], INDEX=dr[, "Month"], FUN=sd, na.rm=TRUE)
dr$Iswna<-!is.na(dr$wt)
counts<-tapply(dr[, "Iswna"], INDEX=dr[, "Month"], FUN=sum)
se<-sdf/sqrt(counts)

# taking 95% confidence interval
qf<-qnorm(0.975)
```

```

# draping with the dot whiskers
for (i in 1:12) points(means[i], i, pch=8, cex=.5, col="red")
for (i in 1:12) lines(c(means[i]-qf*se[i], means[i]+qf*se[i]),
                      c(i, i), lwd=1, col="red", lty="solid"),
                      c(i-0.07, i+0.07),
                      lwd=1, col="red", lty="solid")
for (i in 1:12) lines(rep(means[i]-se[i], 2), c(i-0.07, i+0.07),
                      lwd=1, col="red", lty="solid")
for (i in 1:12) lines(rep(means[i]+se[i], 2), c(i-0.07, i+0.07),
                      lwd=1, col="red", lty="solid")

# divide line separating total
lines(c(1, 40), c(0.5, 0.5), lwd=2, col="gray93", lty=1)

# calculating the mean and se for dot whiskers of total
dr$IswnaT<-!is.na(dr$wt)
meansT<-tapply(dr[, "wt"], INDEX=dr[, "IswnaT"], FUN=mean, na.rm=TRUE)
sdfT<-tapply(dr[, "wt"], INDEX=dr[, "IswnaT"], FUN=sd, na.rm=TRUE)
countsT<-tapply(dr[, "IswnaT"], INDEX=dr[, "IswnaT"], FUN=sum)
seT<-sdfT/sqrt(countsT)

# taking 95% confidence interval
qf<-qnorm(0.975)

# adding boxplot Total at 0 position on the active plot, add=TRUE
boxplot(dr$wt, data=dr, ylim=c(1, 41), at=0,
        xlab="mass [g]",
        horizontal=TRUE, las=1,
        notch=T, boxwex=1.2, col="gray93", log="x", add = T)

# draping with the dot whiskers
points(meansT, 0, pch=8, cex=.5, col="red")
lines(c(meansT-qf*seT, meansT+qf*seT),
      c(0, 0), lwd=1, col="red", lty="solid")
lines(rep(meansT-seT, 2), c(0-0.1, 0+0.1),
      lwd=1, col="red", lty="solid")
lines(rep(meansT+seT, 2), c(0-0.1, 0+0.1),
      lwd=1, col="red", lty="solid")

# add ref line for mean
lines(c(meansT, meansT), c(0, 12), lty=3, col="red")

# add ref line for median
med <- median(dr$wt) # calling the value
lines(c(med, med), c(0, 12), lty=3)

# add axis label for total
axis(2, at=0, labels = "Total", las=1, font = 2)

# additional approximate terminating tick value
axis(1, 40)

```

```
# Image at working directory, getwd()
dev.off()
getwd() # location where the image print as the name of the file
#####
par(mar=c(bottom=5,left=4,top=4,right=2)+0.1)#default lines
```

## Appendix 5

### Supplementary file – Fig\_5\_boxplotL.txt

```
# call file rest in working directory (wd)
dr <- read.csv("dario.csv")

# ordering the grouping sequences
dr$Month <- ordered(dr$Month,levels=c("Feb","Jan","Dec",
                                       "Nov","Oct","Sep",
                                       "Aug","Jul","Jun",
                                       "May","Apr","Mar"))

# opening the printing device to make image as .jpeg
png("Fig_5_boxplotL.jpeg",width = 5, height = 6,units = "in-
",res=588)

# changing default par settings
par(mar=c(bottom=3.9,left=3.9,top=1.3,right=1.3)+0.1)

# changing default font
par(family="serif")

# calculating the width ratio for boxplot
dr$Count<-dr$t1
dr$Count[dr$Count >= 0]<-1
dr$Count[is.na(dr$Count)== TRUE]<-0
sums<-tapply(dr[, "Count"],INDEX=dr[, "Month"],FUN=sum,na.rm=TRUE)
sums<-sums/max(sums)

# VARIABLE-WIDTH notched BOXPLOT call
boxplot(t1 ~ Month,data=dr,ylim=c(4.5,13),xlim=c(0,12),
        xlab="total length [cm]", ylab="month",
        names=levels(dr$Month),horizontal=TRUE,las=1,
        width=sums,notch=T,boxwex=1.3,col="white")

# calculating the mean and se for dot whiskers
means<-tapply(dr[, "t1"],INDEX=dr[, "Month"],FUN=mean,
              na.rm=TRUE)
sdf<-tapply(dr[, "t1"],INDEX=dr[, "Month"],FUN=sd,na.rm=TRUE)
dr$Iswna<-!is.na(dr$t1)
```

```

counts<-tapply(dr[, "Iswna"], INDEX=dr[, "Month"], FUN=sum)
se<-sdf/sqrt(counts)

# taking 95% confidence interval
qf<-qnorm(0.975)

# draping with the dot whiskers
for (i in 1:12) points(means[i], i, pch=8, cex=.5, col="red")
for (i in 1:12) lines(c(means[i]-qf*se[i], means[i]+qf*se[i]),
                      c(i, i), lwd=1, col="red", lty="solid"),
for (i in 1:12) lines(rep(means[i]-se[i], 2), c(i-0.07, i+0.07),
                      lwd=1, col="red", lty="solid")
for (i in 1:12) lines(rep(means[i]+se[i], 2), c(i-0.07, i+0.07),
                      lwd=1, col="red", lty="solid")

# divide line separating total
lines(c(4.5, 13), c(0.5, 0.5), lwd=2, col="gray93", lty=1)

# calculating the mean and se for dot whiskers of total
dr$IswnaT<-!is.na(dr$t1)
meansT<-tapply(dr[, "t1"], INDEX=dr[, "IswnaT"], FUN=mean, na.rm=TRUE)
sdfT<-tapply(dr[, "t1"], INDEX=dr[, "IswnaT"], FUN=sd, na.rm=TRUE)
countsT<-tapply(dr[, "IswnaT"], INDEX=dr[, "IswnaT"], FUN=sum)
seT<-sdfT/sqrt(countsT)

# taking 95% confidence interval
qf<-qnorm(0.975)

# adding boxplot Total at 0 position on the active plot, add=TRUE
boxplot(dr$t1, data=dr, ylim=c(4.5, 13), at=0,
        xlab="total length [cm]",
        horizontal=TRUE, las=1,
        notch=T, boxwex=1.2, col="gray93", add = T)

# draping with the dot whiskers
points(meansT, 0, pch=8, cex=.5, col="red")
lines(c(meansT-qf*seT, meansT+qf*seT),
      c(0, 0), lwd=1, col="red", lty="solid")
lines(rep(meansT-seT, 2), c(0-0.125, 0+0.125),
      lwd=1, col="red", lty="solid")
lines(rep(meansT+seT, 2), c(0-0.125, 0+0.125),
      lwd=1, col="red", lty="solid")

# add ref line for mean
lines(c(meansT, meansT), c(0, 12), lty=3, col="red")

# add ref line for median
med <- median(dr$t1) # calling the value
lines(c(med, med), c(0, 12), lty=3)

```

```
# add axis label for total
axis(2,at=0,labels = "Total",las=1,font = 2)

# additional approximate starting and terminating tick value
axis(1,c(4.5,13),c("4.5","13"))

# Image at working directory, getwd()
dev.off()
getwd() # location where the image print as the name of the file
#####
par(mar=c(bottom=5,left=4,top=4,right=2)+0.1)#default lines
```

Table Appx 1. CSV file-dario.csv

Month	t1	wt
August	8.3	7.6
August	8.8	10.52
August	7.4	4.22
August	7.9	7.89
August	8.6	10.1
August	7.2	4.95
August	8.9	11.74
August	7.5	5.86
August	6.5	4.1
August	7.9	8.43
August	8.7	10.78
August	8.2	8.15
August	8.1	7.76
August	9.1	12.43
August	6.9	4.16
August	6.2	3.18
August	7.6	6.33
August	7.7	7.17
August	7.2	5.51
August	8.1	8.12
August	7	5.25
August	8.5	9.45
August	6.6	3.85
August	8	8.1



cont. Table Appx. 1

August	8.8	10.64
August	6.1	3.47
August	6.6	3.73
August	8.7	9.72
August	8.3	8.39
August	8.1	8.75
August	7.2	5.36
August	6.4	7.32
August	6.4	3.64
August	7.9	7.61
August	7.6	7.1
August	7.5	6.45
August	7.1	5.1
August	8.1	6.7
August	7.9	5.1
August	6.9	8.8
August	5.1	1.87
August	6.5	3.7
August	6.92	4.63
August	6.8	4.22
August	6.71	4.18
August	7.9	7.69
August	8.2	7.1
August	7.2	4.5
August	6.5	3.74
August	7.1	4.38
August	6.7	5.83
August	7.3	5.66
August	7.6	6.42
August	7.9	7.2
August	6.7	4.1
August	8.8	9.78
August	7.2	5.15
August	7.3	5.24
August	7.4	5.27
August	7.7	6.65

cont. Table Appx. 1

August	6.6	3.92
August	7.2	4.1
August	7.5	6.71
August	8.7	7.1
August	7.3	5.2
August	7.6	4.62
August	5.1	1.93
August	7.8	8.1
August	7.1	4.88
August	7	4.2
August	6.3	3.72
August	7.7	5.93
August	7.5	6.34
August	7.1	5.6
August	6.42	2.1
August	6.9	5.2
August	6.84	4.75
August	6.93	4.42
August	7.32	5.61
September	9.3	12.16
September	7.6	6.91
September	7.4	5.64
September	7.9	7.4
September	8.9	8.95
September	6.8	3.22
September	6	2.93
September	9.2	12.3
September	9.5	13.9
September	6.9	3.94
September	8.2	7.4
October	7.8	6.26
October	7.2	5.94
October	7.3	5.56
October	8.4	9.1
October	5	2.1
October	8.3	7.2

cont. Table Appx. 1

October	8.8	12.9
October	9	10.83
October	7.6	6.58
October	5.6	1.88
October	9.1	10.42
October	7.7	4.92
November	6.6	4.77
November	7.2	5.75
November	7.9	8.7
November	5.8	2
November	8.5	7.72
November	4.6	1.22
November	7.8	6.1
November	8.1	8.75
November	8	4.69
November	8.1	8.89
November	7.5	7.81
November	7.6	5.5
December	9.7	12.2
December	9.1	11.69
December	7.7	4.9
December	7.4	6.72
December	7.6	8.1
December	9.5	12.11
December	7.4	4.69
December	8.3	8.42
December	8.6	7.7
December	8.5	8.37
December	7.8	4.3
January	9.6	12.6
January	7.5	5.8
January	7.8	3.9
January	6.9	3.6
January	9.6	10.1
January	7	6.1
January	8.6	6.8

cont. Table Appx. 1

January	6.6	2.95
January	6.4	2.32
January	7.8	9.6
February	10.5	18.7
February	11.9	21.68
February	8.4	6.3
February	8.3	8.1
February	8.1	6.6
February	8.4	8.26
February	8.4	7.8
February	8.2	7.7
February	11.1	12.73
February	12.9	39.24
February	11.4	13.73
February	9.1	10.7
February	9.5	11.2
February	6.9	3.91
February	9.5	7.1
February	8.5	7.5
February	8.9	6.4
February	7.3	4.6
February	7.1	4.1
March	10.2	15.91
March	9.5	12.92
March	10.6	15.8
March	8.4	5.74
March	7.3	6.4
March	6.5	3.74
March	7.4	4.8
March	7.99	6.42
March	8.1	6.1
April	11.6	23.12
April	11.1	22.16
April	12.1	26.6
April	7.2	4.08
April	11.9	15.9

cont. Table Appx. 1

April	11.4	25.1
April	9.6	10.83
April	8.2	5.7
April	7.6	6.62
April	6.8	4.95
April	7.1	4.1
May	9.3	10.98
May	7.6	5.55
May	8.6	8.83
May	8.4	7.9
May	7.8	6.78
May	8.5	8.24
May	7.5	4.5
May	10.3	13.02
May	7.8	5.72
May	8.7	6.8
May	8.1	7.33
May	10.2	22.15
May	7.5	7.11
May	9.4	15.82
May	7.8	4.39
May	9.2	14.69
May	7.9	6.12
May	9.5	11.91
May	9.5	12.11
May	7.4	7.69
May	8.3	8.42
May	7.2	4.92
May	7.8	8.25
June	7.3	4.95
June	5.9	1.95
June	7.8	6.9
June	7.1	4.88
June	7	4.2
June	6.3	3.72
June	7.7	5.93

cont. Table Appx. 1

June	7.5	6.34
June	7.1	6.5
June	6.42	3.82
June	6.9	4.2
June	6.84	4.75
June	5.99	4.42
June	7.32	5.61
June	6.5	3.32
July	8.2	6.7
July	7.7	7.13
July	7	5.34
July	8.6	10.5
July	7.6	6
July	7.8	7.34
July	7.2	5.45
July	8.5	7.93
July	7.3	6.89
July	7.4	5.39
July	6.5	3.49
July	7.6	5.01
July	7.6	4.7
July	6.5	2.87