JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

Box-and-Whisker Plots Applied to Food Chemistry

João E. V. Ferreira,*[†] Ricardo M. Miranda,[†] Antonio F. Figueiredo,[†] Jardel P. Barbosa,[‡] and Edykarlos M. Brasil[§]

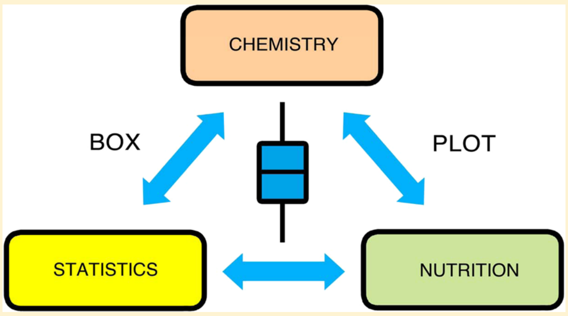
[†]Laboratory of Computational Chemistry, Federal Institute of Education, Science and Technology of Pará, CEP 66093-020, Belém, Pará, Amazon, Brazil

[‡]Laboratoy of Computational Chemistry, State University of Amapá, CEP 68900-070, Macapá, Amapá, Amazon, Brazil

[§]Laboratory of Chemistry Research, Federal University of Pará, CEP 66075-110, Belém, Pará, Amazon, Brazil

Supporting Information

ABSTRACT: Box-and-whisker plots or simply boxplots are powerful graphical representations that give an overview of a data set. In this work five different examples illustrate the applications of boxplots in food chemistry. The examples involve relative sweetness of sugars and sugar alcohols with respect to sucrose, the potassium content of fruits and vegetables, amino acid content of egg white and yolk, chemical composition of freshwater and saltwater fish, and change in fatty acid composition of soybean oil through traditional cultivation or genetic engineering techniques. Readers are guided to identify through boxplots key features present in some foods associated with inorganic elements and molecules. It is certainly an interdisciplinary way of studying concepts of statistics, nutrition, and chemistry.



KEYWORDS: *Upper-Division Undergraduate, Interdisciplinary/Multidisciplinary, Computer-Based Learning, Chemometrics, Consumer Chemistry*

INTRODUCTION

When we are dealing with a data set and looking for extracting relevant information from it, it is advisible to make use of graphics, because they are more practical and give a better visualization than spreadsheets. Box-and-whisker plots (Figure 1), which, as an abbreviation, are called boxplots,[1] are powerful graphical representations that give an overview and a numerical summary of a data set. They have been a standard statistical graph since the American statistician John W. Tukey (with his colleagues and students) publicized them energetically in the 1970s.[2] The graphic shows a rectangle (the box) with two lines (the whiskers) extending from opposite edges of the box. The ends of the whiskers represent the dispersion while the lower and upper edges of the box signal the first and third quartiles, respectively. Across the box there is a line parallel to the edges indicating the median of the distribution. Then, the boxplots are the graphical representation of the five-number-summary: two extremes, two quartiles, and the median.

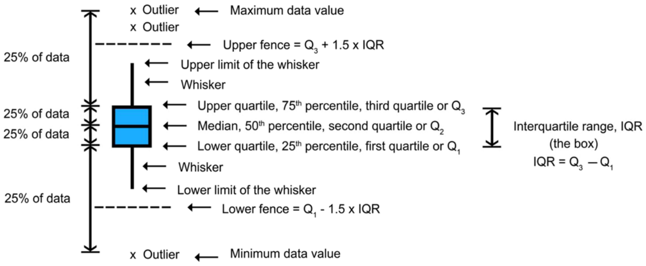


Figure 1. Main components of a boxplot: median, quartiles, whiskers, fences, and outliers.

A very good introduction of boxplots was presented in this *Journal* by Professor Larsen in 1985.[3] The author describes the construction and interpretation of these graphical formats, showing as examples applications involving bond energies, heats of formations, heats of solutions, and solubility products. He stated that presentation of data in this form provides considerable motivation for the "explanations" that we traditionally give for chemical and physical trends. According to him, tables of data that appear in introductory textbooks, especially alphabetically arranged data, seem rather uninformative and uninteresting, and should be accompanied by a boxplot showing the content of the table of data. He continuously emphasized that simply seeking highs and lows and particular groupings in a table does not produce the same results and give the clues that numerical detective work of exploratory data analysis achieves, and he states "box-and-whiskers plots are able to be rapidly constructed and thus provide a means for quickly assessing relative data values in a large (or small) data set consisting of chemical and physical properties." Another interesting use of these graphics is found in a paper published

Received: April 23, 2016
Revised: September 14, 2016
Published: October 14, 2016

DOI: 10.1021/acs.jchemed.6b00300

in 2008 by Canaes et al.,[4] who proposed a simple, didactic, and tasty (according to the authors) classroom experiment with colored candies to introduce the undergraduate students to important concepts of statistics and sampling techniques. During the activity, students classified the candies from 10 bags according to the color and constructed boxplot for each color, showing the distribution of the percentage of candies with a certain color that are present in each bag.

Now, in this work, boxplots are revisited and used in a different application: food chemistry. This area of chemistry is very important to our lives, and its study generally requires the analysis of so many data. The authors have selected five applications of these graphics in order to answer five questions related to food chemistry.

- Question 1: How can boxplots be used to classify sweeteners according to their sweetness power?
- Question 2: How can boxplots be used to compare the potassium content of fruits and vegetables?
- Question 3: How can boxplots be used to compare the amino acid compositions of egg white and egg yolk?
- Question 4: How can boxplots be used to compare the protein composition and fat composition of freshwater and saltwater fish?
- Question 5: How can boxplots be used to compare the fatty acid composition of different vegetable oils?

Boxplots are particularly useful in presenting data in graphical ways that facilitate making comparisons, finding tendencies, and providing additional insights. Readers are guided to identify through graphics (boxplots) key features present in some foods associated with inorganic elements and molecules and visualized through medians, quartiles, and outliers, which together describe the shape, central tendency, and variability of a distribution. This activity is appropriate for upper-division undergraduate students. The influence of the findings of this paper is to motivate students to apply these pictorial resources to investigate problems of chemical interest related to foods. It is certainly an interdisciplinary way of studying concepts of statistics, nutrition, and chemistry as can be seen in Table 1

Table 1. Key Concepts in This Interdisciplinary Activity

Chemistry (Inorganic Element, Organic Molecules)	Statistics (Boxplots)	Nutrition (Foods)
Potassium	Distribution	Sweeteners
Sugars	Median	Fruits
Sugar alcohols	Quartile	Vegetables
Lipids	Outlier	Egg
Proteins	Whisker	Fish
Amino acids	Interquartile range	Oils

■ METHODOLOGY

For those not familiar with boxplots, a brief description of the main components of the graphics is present (Figure 1). Interpretation of these elements helps understand how nutrients are distributed in foods.

The median is a measure that indicates the midpoint of the distribution. Computation of the median depends on the number of data. If the data set has an odd number of observations, then the median is the middle observation, considering all data ranked. But if the number of data is even, in this case the median is the mean of the two middle observations.

Quartiles are values that divide data into quarters, that is, groups containing 25% of the samples. The lower quartile (Q_1) is the value such that 25% (1/4) of the data lie at or below it; the second quartile (Q_2) is the median of the data, and the upper quartile (Q_3) is the value such that 75% (3/4) of the data lie at or below it.

The interquartile range (IQR) is the length of the box and envelops all of the data between Q_3 and Q_1, that is, the middle 50% of the ranked samples. IQR is little affected by the presence of outliers and is a measure that is very useful for comparing two data sets.

Fences are the limits above and below the box (generally not visualized) that are used to flag possible outliers. The upper fence is the upper limit computed as $Q_3 + 1.5 \times \text{IQR}$. Lower fence is the lower limit computed as $Q_1 - 1.5 \times \text{IQR}$.

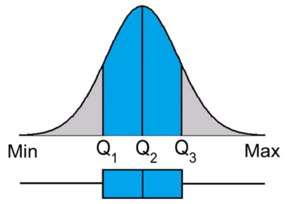
Whiskers are the lines that extend from Q_1 and Q_3, respectively, in the direction of the minimum and maximum values of the data set. In this work, the whiskers are extended to the data value just before the fences since this strategy is implemented in the computational package used by the authors. Moreover, sometimes the whiskers are represented ending in a small horizontal line. In this work, all boxplots are drawn without this horizontal bar.

The outlier is the value beyond the fence. There are many reasons to show an extreme value such as error while collecting data or while making measurements. However, we must be careful since extreme values may be correct, as in the examples presented, and in this case they are indeed very different from the rest of the data. Foods or food constituents flagged as outliers above the upper fence are those having large nutrient content. The contrary is valid for outliers below the lower fence.

Boxplots are mainly constructed to give an overview of the distribution of a data set. In Figure 2 we can see the correspondence between the boxplot shape and the distribution.

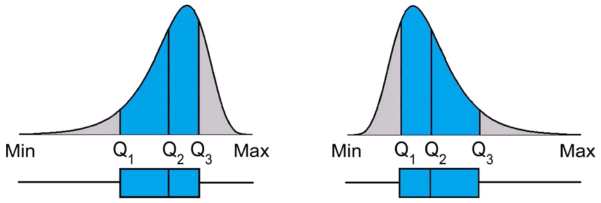
But when and how to construct a boxplot? These graphics are recommended when performing an exploratory data analysis. However, it is important to address that statistical tests should be used (with an appropriate confidence level) to compare data sets and not only to make a conclusion by visual inspection of the boxplot. Limitations of the method also occur for very small data sets. However, for large data sets the results can be very interesting. Then boxpots are mainly used in descriptive statistics when we have a large data set, and it is necessary to summarize information. For this purpose we can analyze some aspects related to the data set such as the central tendency, statistical dispersion, and shape of the distribution. Another point to highlight is that boxplots are nonparametric. Therefore, initially, data should be checked for normality.

Fortunately our ability to statistically analyze data has grown significantly with the maturing of computer hardware and software.[5] Available computer programs are capable to construct these graphics, but it is important to note that they may differ regarding the drawing. The way quartiles are computed differ too. The methodology to construct a boxplot is quite easy. In general, a column with data is selected in a spreadsheet. Then the option to create this graphic is chosen. Results presented here can be reproduced by employing Minitab 16.0 software,[6] widely used in educational and scientific research. The program has common statistical,



symmetric

(a)



asymmetric or skewed (negative or left)

(b)

asymmetric or skewed (positive or right)

(c)

Figure 2. Boxplots and the corresponding distributions of data.

plotting, and modeling functions available. Minitab shows a very common form of boxplot, showing the whiskers extending from Q_1 to the minimum value before the lower fence and from Q_3 to the maximum value before the upper fence. The program also shows outliers as separate points signaled by a star. Data used to generate all boxplots were taken from a popular textbook on food chemistry.[7] It is not the objective of this paper to get into details about the very complex aspects associated with the nutritional and physiological roles of the food ingredients investigated. In the same way, further detailed explanations involving mathematics and variations of boxplots are avoided, but can be found in a more specialized literature.[1−3,8,9] Data with nutritional information used to construct the boxplots presented in the five examples can be found in the Supporting Information.

This activity is better performed in a computer laboratory so that students can have practical experience with the construction and interpretation of boxplots by using software with statistical tools. The activity takes around 3 h and must be divided into two main parts. First, there is an explanation for the audience about the basic structure of the boxplots, as can be visualized in Figure 1. In the second part, instructional guidance is given on how to run the program to generate the graphics.

Initially, a small data set (1, 2, 3, 5, 7, 8, and 10, for example) is used in order to show how to construct and interpret boxplots. Students are guided to create and analyze the boxplot in terms of shape, center, spread, and presence of outliers (if any).

Steps students must follow while constructing boxplots are listed here:

- Computing the quartiles (Q_1, Q_2, and Q_3);
- Computing the IQR;
- Drawing the box limited by Q_1 and Q_3;
- Drawing the line inside the box indicating the median;
- Computing the fences;
- Flagging outliers (if present);
- Drawing the whiskers.

Steps students must follow while interpreting boxplots are listed here:

- Examining the length of the box;
- Examining the length of the whiskers;
- Examining the position of the median;
- Examining the reasons for the presence of outliers (if present).

We should reinforce the idea that outliers may result from error while collecting data or an observation whose parameter investigated is indeed very different from the other values. Additionally, students should be advised that there is not only one way to draw the graphics, including their formats. In this paper, special attention is dedicated to the way Minitab software draws the boxplots.

After the introduction of all these statistical concepts, students are given a spreadsheet with data set related to food chemistry so that the boxplot can be constructed and interpreted using Minitab statistical software. The examples authors traditionally have employed with students are described next and are easy and fast to use through computation. It is not imperative to use all the examples discussed in this text with students. They only represent some specific cases that could be employed.

Authors state that students can benefit from this activity, learning a variety of interrelated concepts involving statistics, chemistry, and nutrition. Additionally, teaching students how to create and interpret boxplots in the data analysis in real-world problems can decrease the rejection they usually give to statistics.

■ RESULTS

Example 1: Relative Sweetness of Sugars and Sugar Alcohols with Respect to Sucrose

Mono- and oligosaccharides and their corresponding sugar alcohols, with a few exceptions, are sweet. Sucrose is distinguished from the other sugars by its pleasant taste even at high concentrations and is the reference substance usually chosen to compare sweeteners.[7] Sugar alcohols occur in some fruits and are produced industrially as food ingredients. The relevance of some sugar alcohols as sweeteners to a diet lies in the fact that they are only slowly absorbed and can, therefore, be used in diabetic foods, possess reduced physiological calorie value, and are noncarcinogenic.[10]

A boxplot can show how sugars differ in intensity of sweetness. Figure 3 ranks sugars and sugar alcohols with respect to sucrose according to their relative sweetness, whose distribution of values follows a slightly symmetric shape with short whiskers. D-Galactose, whose relative sweetness is 63 (the median of the data set), indicates the midpoint of the sweetness scale. It is possible to categorize the sweeteners into three main classes in relation to the quartiles. A group of sweeteners with high relative sweetness power ($\geq Q_3$) comprises D-fructose, xylitol, sucrose, and invert sugar. All of them are sweeteners of importance in food processing,[7] and a boxplot points out this characteristic, that is, their high power to sweeten. Xylitol is as sweet as sucrose, and a boxplot ranks it just above sucrose. However, xylitol produces a cooling effect in the mouth when it dissolves. This effect is used in some candies. For these reasons, xylitol has been used as sugar substitute. A gap in the scale separates this group of important sweeteners from the other compounds with less sweetness power. The substances with intermediate sweetness power (Q_1 < relative sweetness < Q_3)

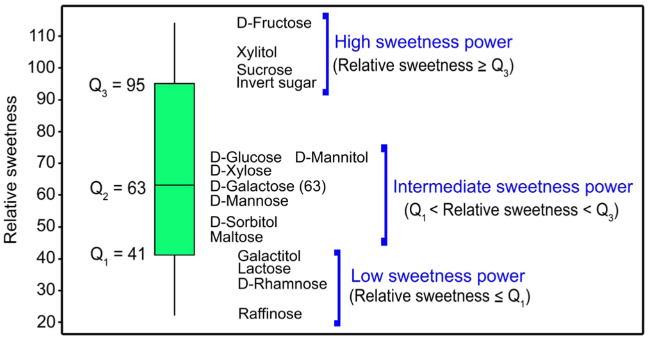


Figure 3. Boxplot for the relative sweetness of sugars and sugar alcohols with respect to sucrose. Quartiles are labeled beside the boxes. D-Galactose, whose relative sweetness is 63 (the median of the data set), indicates the midpoint of the relative sweetness. A classification is proposed according to the relative sweetness: low sweetness power (relative sweetness $\leq Q_1$), intermediate sweetness power ($Q_1 <$ relative sweetness $< Q_3$), and high sweetness power (relative sweetness $\geq Q_3$). Data from Table 1 (in the Supporting Information). Adapted with permission from ref 7. Copyright 2009 Springer-Verlag Berlin Heidelberg.

include D-glucose, D-mannitol, D-xylose, D-galactose, D-mannose, D-sorbitol, and maltose. The third class of sweeteners comprises the substances with the lowest relative sweetness ($\leq Q_1$), whose substances are galactitol, lactose, D-rhamnose, and raffinose. In all three classes of sweeteners, we can find mono- and oligosaccharides and sugar alcohols. No sample was classified as outlier, neither above the box nor below it.

In this example, the boxplot was employed with the ultimate objective of ranking substances based on how sweet they are. From D-fructose to raffinose, the sugars are ranked in a decreasing power of relative sweetness. The advantages of boxplots in comparison to simply seeking highs and lows and particular groupings in a table, as supported by Larsen, are more evident when dealing with many samples. A boxplot both ranks the samples and shows the distribution of values, which is one of the very useful sides of the graphic.

Example 2: Potassium Content of Fruits and Vegetables

Minerals classified as main elements, which include potassium, are essential for human beings in amounts >50 mg per day. The importance of minerals as food ingredients depends not only on their nutritional and physiological roles. They also contribute to food flavor and activate enzyme-catalyzed and other reactions, and they affect the texture of food.[7]

Results in Figure 4 show boxplots for the potassium content of 11 fruits and 17 vegetables listed in Table 2. Visualization of the boxplots reveals that vegetables, in general, present higher potassium content than fruits. In fact, fruits are generally less rich in minerals than vegetables.[10] This conclusion is drawn from the slightly upper position of the box and whiskers associated with vegetables and their larger median, which indicates the midpoint of the distribution of the mineral.

Authors propose, with this example, a classification of fruits and vegetables according to the potassium content: low potassium content ($\leq Q_2$ for fruits), intermediate potassium content (between Q_2 for fruits and Q_3 for vegetables), and high potassium content ($>Q_3$ for vegetables, except lentils, which are classified as having very high potassium content, 837 mg). In Table 3, the fruits and vegetables are listed according to the classification they receive on the basis of the potassium content they possess. By this criterion, we could say that vegetables with high potassium content include spinach, kale, and brussels sprouts. However, potatoes, carrots, peas, and tomatoes, for

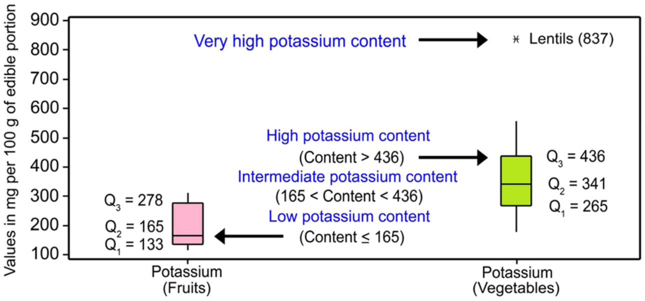


Figure 4. Boxplots for potassium content of some fruits and vegetables listed in Table 2. Quartiles are labeled beside the boxes. Foods that stand out as having very high content of this mineral are lentils. Fruits are generally less rich in potassium than vegetables. A classification is proposed for fruits and vegetables according to the potassium content: low potassium content ($\leq Q_2$ for fruits), intermediate potassium content (between Q_2 for fruits and Q_3 for vegetables), and high potassium content ($>Q_3$ for vegetables, except lentils, which are classified as having very high potassium content, 837 mg). Data from Table 2 (in the Supporting Information). Adapted with permission from ref 7. Copyright 2009 Springer-Verlag Berlin Heidelberg.

Table 2. Fruits and Vegetables Used To Generate the Boxplots in Figure 4

	Fruits	
Apple	Grapefruit	Cherries-sour
Orange	Rose hips	Plums
Apricots	Currants-red	Sea buckthorn
Strawberry	Currants-black	
	Vegetables	
Watercress	Kale	Brussels sprout
Mushroom, cultivated	Potatoes	Spinach
Chicory	Kohlrabi	Edible mushroom[a]
Endive	Head lettuce	Tomato
Peas, green	Lentils, dried	White cabbage
Lamb's lettuce	Carrots	

[a] Boletus edulis.

Table 3. Classification of Fruits and Vegetables According to the Potassium Content As Defined in Figure 4

	Fruits	
	Intermediate Potassium Content	
Currants-black	Rose hips	Apricots
Currants-red	Plums	
	Low Potassium Content	
Orange	Strawberry	Grapefruit
Sea buckthorn	Apple	Cherries-sour
	Vegetables	
	Very High Potassium Content	
Lentils, dried		
	High Potassium Content	
Spinach	Kale	Brussels sprout
	Intermediate Potassium Content	
Lamb̀s lettuce	Potato	Mushroom, cultivated
Endive	Edible mushroom	Kohlrabi
Carrots	Watercress	Peas, green
White cabbage	Tomato	Chicory
Head lettuce		

example, possess intermediate content in this mineral. Regarding fruits, currants-black, currants-red, rose hips, apricots, and plums can be classified as having intermediate potassium content, whereas oranges, strawberries, grapefruit, and apples, for example, present low potassium content. No fruit was classified as having high potassium content. In addition, no vegetable was classified as a poor source of this mineral.

With consideration of the shape of the distribution of potassium in the foods investigated, Figure 4 depicts a boxplot for vegetables with a symmetry somewhat like that in Figure 2a. However, boxplot for fruits lacks symmetry, being right skewed around the median, because the extension between Q_1 and Q_2 is visually shorter than that between Q_2 and Q_3. In this example, the comparison of the potassium content in fruits and vegetables reveals one of the advantages of boxplots. They allow a comparison of two or more data sets side by side (parallel). While comparing, we observe the distribution of values. The degree of similarity involving the data sets is measured qualitatively by the overlap of the boxes.

Example 3: Amino Acid Composition of Egg White and Egg Yolk

Chicken eggs are one of nature's perfect protein foods.[7] Moreover, proteins of animal origin, such as egg proteins, are widely used in fabricated foods.[11] Figure 5A,B show the boxplots for the amino acid content present in egg white and yolk. We clearly visualize that glutamine is by far the major constituent in both parts of the egg. It is interesting to analyze this amino acid specifically. The upper fence in the boxplot for egg white (dashed line above the box) lies at $Q_3 + 1.5 \times \text{IQR}$ ($0.77 + 1.5 \times 0.36 = 1.31$). However, the score for glutamine lies at 1.52. Then this amino acid is outside the fence and must be classified as an outlier. Now let's consider the boxplot for egg yolk. The upper fence lies at $Q_3 + 1.5 \times \text{IQR}$ ($1.18 + 1.5 \times 0.65 = 2.16$), but the score for glutamine lies at 1.95, within the fence. Then, this amino acid is not classified as outlier, even though the amino acid content in this case is larger than in the previous one. So being an outlier or not depends on the data set. The fact that glutamine is an outlier in the egg white means that glutamine is present in a much greater amount than the other amino acids.

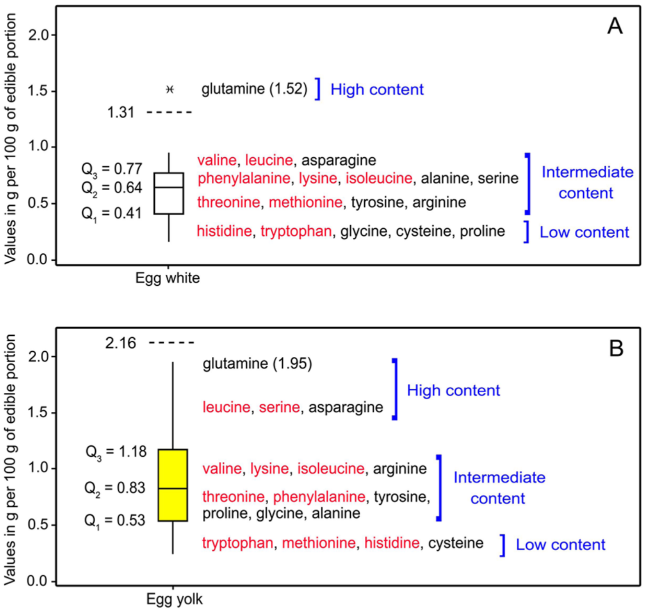


Figure 5. Boxplots for the amino acid content present in egg white (A) and egg yolk (B). Quartiles are labeled beside the boxes, and essential amino acids are written in red. A classification is proposed for amino acids according to the content they present in egg white and egg yolk. In egg white, there are amino acids in low content ($\leq Q_1$), intermediate content (between Q_1 and $Q_1 + 1.5 \times \text{IQR}$) and high content ($>Q_3 + 1.5 \times \text{IQR}$). In egg yolk, there are amino acids in low content ($\leq Q_1$), intermediate content (between Q_1 and Q_3), and high content ($>Q_3$). Data from Table 3 (in Supporting Information). Adapted with permission from ref 7. Copyright 2009 Springer-Verlag Berlin Heidelberg.

The boxplot for egg white is slightly negatively skewed whereas for egg yolk the shape looks a little bit positively skewed. Amino acids in egg yolk are present slightly in a larger amount than in egg white as we can note by comparing the relative position of the boxes, the whiskers, and medians (0.64 g, egg white, and 0.83 g, egg yolk). Moreover, the orders in which they are ranked do not differ greatly in position between egg white and yolk.

The authors propose, with this example, a classification for amino acids according to the content they present in egg white and egg yolk. In egg white (Figure 5A), amino acids present in low content ($\leq Q_1$) are histidine, tryptophan, glycine, cysteine, and proline; intermediate content (between Q_1 and $Q_3 + 1.5 \times \text{IQR}$) include valine, leucine, asparagine, phenylalanine, lysine, isoleucine, alanine, serine, threonine, methionine, tyrosine, and arginine; high content ($>Q_3 + 1.5 \times \text{IQR}$) include only glutamine. In egg yolk (Figure 5B), amino acids present in low content ($\leq Q_1$) are tryptophan, methionine, histidine, and cysteine; intermediate content (between Q_1 and Q_3) includes valine, lysine, isoleucine, arginine, threonine, phenylalanine, tyrosine, proline, glycine, and alanine. High content ($>Q_3$) includes glutamine, leucine, serine, and asparagine.

Example 4: Chemical Protein Composition and Fat Composition of Freshwater and Saltwater Fish

Fish and fish products play an important role in human nutrition as a source of biologically valuable proteins, fats, and fat-soluble vitamins.[7] They are particularly strongly consumed by those who live close to rivers or the coast. In this example the objective is to analyze if the environment in which the fish is found affects the content of proteins and fats present in fish meat. Table 4 lists the species from both classes investigated.

A fast look at Figure 6 reveals that fat composition of fishes listed in Table 4 varies greatly in comparison to the protein content. Other observation is that distribution of values for fat content is positively skewed, as we can see by the horizontal bar

Table 4. Freshwater and Saltwater Fish Used To Generate the Boxplots in Figure 6

Freshwater Fish		
Eel	Carp	Salmon
Perch	Tench	River trout
Zander	Pike	Smelt
Saltwater Fish		
Cod	Plaice	Herring (Baltic Sea)
Haddock	Flounder	Sardine
Ling	Commonsole	Mackerel
Hake	Halibut (but)	Tuna
Rockfish	Turbot (britt)	
Catfish	Herring (Atlantic)	

DOI: 10.1021/acs.jchemed.6b00300
J. Chem. Educ. 2016, 93, 2026−2032

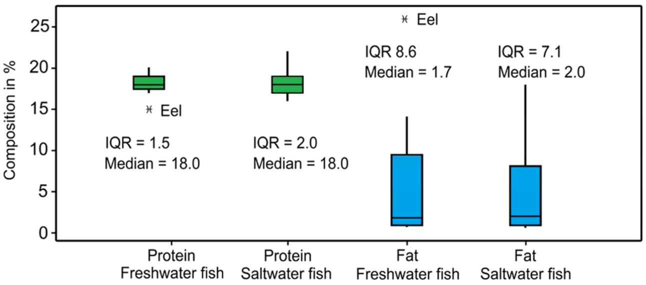


Figure 6. Boxplots for the chemical composition (protein and fat) of freshwater and saltwater fish. Overlap of the boxes for each food constituent occurs; consequently, there is no visual difference in nutritional composition between freshwater and saltwater fish regarding the meat of both classes of fish. Variations in composition are larger for fat than for protein. Data from Table 4 (in the Supporting Information). Adapted with permission from ref 7. Copyright 2009 Springer-Verlag Berlin Heidelberg.

for median very close to Q_1 and by the longer whisker above the box and a short whisker below it. This suggests that the lowest 50% of the ranked species of fish exhibit a narrow range for fat composition while the contrary is found for the other 50%.

Results also show overlap of the boxes for each food constituent. This argument is supported by the fact that there is a great similarity regarding shapes and positions in boxplots when we compare both classes of fish. In fact, the medians for protein composition are the same. For fat composition, medians are similar. Now comparing the IQRs, the similarities continue. Then, the first visual impression is that the nutritional composition between freshwater and saltwater fish regarding the meat considering the content in proteins and fats are similar. Obviously, differences may exist when only some species are compared. Curiously, eel is signaled as an outlier for having the lowest protein content but the highest fat content among freshwater fish.

Authors propose, with this example, a classification for fish according to the fat content they present in freshwater and saltwater fish (Figure 7 and Table 5). Freshwater fish with very low fat content ($\leq Q_2$ for saltwater fish) include perch, zander, tench, pike, and smelt. Low fat content (between Q_2 for saltwater fish and Q_3 for freshwater fish) include carp and river trout. Considering high fat content ($>Q_3$ for freshwater fish), we have salmon and eel. The classification for saltwater fish is quite similar. Saltwater fishes with very low fat content include cod, haddock, ling, plaice, flounder, commonsole, halibut (butt), and turbot (britt). Low fat content includes herring (Baltic Sea), hake, rockfish, and sardine. Considering high fat content, the examples are herring (Atlantic), mackerel, and tuna.

Table 5. Classification of Freshwater and Saltwater Fish According to the Fat Content As Defined in Figure 7

Freshwater Fish		
	High Fat Content	
Eel	Salmon	
	Low Fat Content	
Carp	River trout	
	Very Low Fat Content	
Perch	Zander	Tench
Pike	Smelt	
	Saltwater Fish	
	High Fat Content	
Herring (Atlantic)	Tuna	Mackerel
	Low Fat Content	
Herring (Baltic Sea)	Rockfish	Hake
Sardine		
	Very Low Fat Content	
Cod	Haddock	Ling
Plaice	Flounder	Commonsole
Halibut (butt)	Turbot (britt)	Catfish

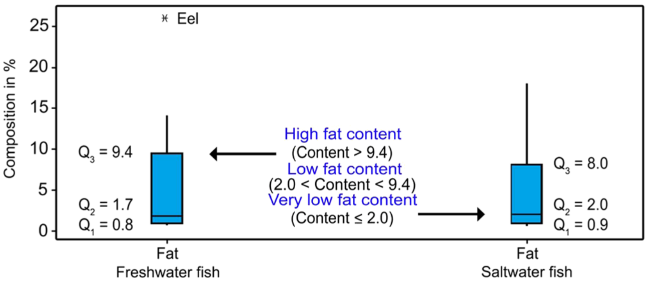


Figure 7. Boxplots for the fat composition of freshwater and saltwater fish. Quartiles are labeled beside the boxes. A classification is proposed for freshwater and saltwater fish according to the fat content: very low fat content ($\leq Q_2$ for saltwater fish), low fat content (between Q_2 for saltwater fish and Q_3 for freshwater fish) and high fat content ($>Q_3$ for freshwater fish). Data from Table 4 (in the Supporting Information). Adapted with permission from ref 7. Copyright 2009 Springer-Verlag Berlin Heidelberg.

Example 5: Difference in Fatty Acid Composition of Different Soybean Oils

Cultivation using traditional and genetic engineering techniques has made it possible to develop soybean genotypes which have a fatty acid composition which meets the different demands made on edible oils[7] because distinct oils display distinct physical chemical properties and nutritional aspects. This final application of boxplot is intended to show how this graphic can be employed to analyze the evolution of a process under different techniques. Therefore, a case involving the change in composition of fatty acids in soybean oil modified through different techniques to produce new types of oils is presented. The fatty acids analyzed are palmitic (16:0, saturated), stearic (18:0, saturated), oleic (18:1, unsaturated), linoleic (18:2, unsaturated), and linolenic (18:3, unsaturated).

Boxplots for low saturated oil (Figure 8B) display a low composition for both palmitic (saturated) and stearic (saturated) acids. In comparison to normal oil (Figure 8A), the decrease in composition of palmitic acid is from around 10% (normal oil) to around 5% (low saturated oil). Another observation is that the composition of oleic acid increased. Such a change in fatty acid composition may be of great importance considering oleic acid is involved in cholesterol in low density lipoproteins (LDL). Another point to consider is that normal and low saturated oils present boxplots with very similar shapes: they are positively skewed regarding the distribution of values for the composition of fatty acids in the oils investigated.

In high stearic oil (Figure 8C), stearic acid is ranked around the median. The composition of this fatty acid increased from a value less than 5% (normal oil) to 20% (high stearic oil). That is a significant increase: in normal oil, stearic acid ranks the lowest in composition but the third lowest in high stearic oil. For this oil, the distribution of values for the composition of fatty acids follows a somewhat symmetric shape.

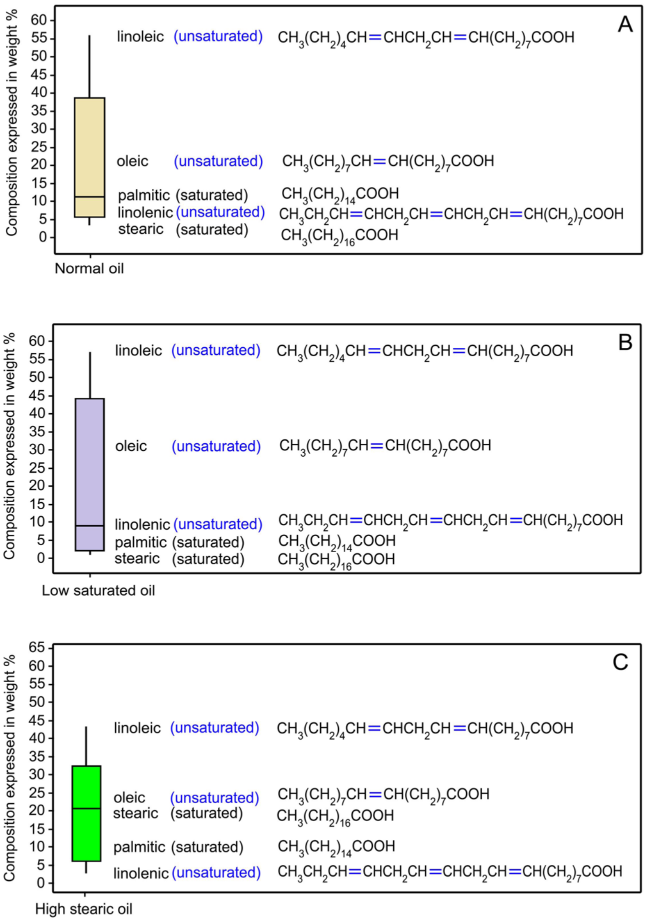


Figure 8. Boxplots for the fatty acid composition of normal (A), low saturated (B), and high stearic (C) soybean oil through traditional cultivation or using genetic engineering techniques. An evident change in the rank of composition occurs when normal oil is transformed into low saturated oil and high stearic oil. Data from Table 5 (in Supporting Information). Adapted with permission from ref 7. Copyright 2009 Springer-Verlag Berlin Heidelberg.

We propose a classification for fatty acids according to the content they present in the oils. So, in normal oil, palmitic, linolenic, and stearic acids are present in low quantities ($\leq Q_2$); only oleic acid is found in intermediate quantity (between Q_2 and Q_3), and only linoleic acid is found in high quantity ($>Q_3$). Considering low saturated oil, the same classification can be applied. In high stearic oil, palmitic and linolenic acids are present in low quantity (around Q_1); oleic and stearic acids are found in intermediate quantities (around Q_2), and only linoleic acid is found in high quantity ($>Q_3$).

CONCLUSION

Boxplots are graphics that are easy to be constructed and interpreted. They are very useful in food chemistry because they can give a good overview of our data, helping to better understand their characteristics, especially, but not only, when we are dealing with a large data set. In this work five different applications of these pictorial representations were discussed just to exemplify the great variety of situations in which they can be employed. Boxplots were useful in making comparisons and ranking samples of sweeteners according to their sweetness. The examples also showed how these graphics can identify the foods that stand out as being a rich source of potassium and how the contents of amino acids are distributed in egg white and yolk. Another utility was in the comparison of the chemical composition of freshwater and sea fish and the composition of distinct soybean oils. In fact, there are numerous possibilities for using these graphics whenever we have to perform statistical exploratory analysis in problems related to foods. Other examples of possibilities for the use of boxplots include many comparisons: fatty acid composition of different plant oils (sunflower, palm, olive, and others) or different meats (beef, chicken, pork, and fish); vitamin content of fruits and vegetables; caloric value of some food products; chemical compositions of different milks (cow, goat, buffalo, and sheep); phenols in white and red wine. However, applications of boxplots are not restricted to food chemistry but can be extended to other areas too. Then authors want with this paper to call attention to the advantages of applying these powerful graphic tools to study problems of chemical interest, making data analysis easier, visualizing data in a different way and, the most important, revealing hidden information.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.6b00300.

Data with nutritional information used to construct the boxplots presented in the five examples (PDF, DOCX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: joao.elias@yahoo.com.br.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Massart, D. L.; Smeyers-Verbeke, J.; Capron, X.; Schlesier, K. Visual Presentation of Data by Means of Box Plots. *LC.GC Europe* **2005**, *18*, 215−218.

(2) Cox, N. J. Speaking Stata: Creating and Varying Box Plots. *STATA J.* **2009**, *9* (3), 478−496.

(3) Larsen, R. D. Box-and-whisker Plots. *J. Chem. Educ.* **1985**, *62*, 302−305.

(4) Canaes, L. S.; Brancalion, M. L.; Rossi, A. V.; Rath, S. Using Candy to Learn About Sampling Techniques. *J. Chem. Educ.* **2008**, *85*, 1083−1088.

(5) Schlotter, N. E. A Statistics Curriculum for Undergraduate Chemistry Major. *J. Chem. Educ.* **2013**, *90*, 51−55.

(6) Minitab. *Minitab 16 Statistical Software*; Minitab Inc.: State College, PA, 2010.

(7) Belitz, H.-D.; Grash, W.; Schieberle, P. *Food Chemistry*, 4th ed.; Springer-Verlag: Berlin, Germany, 2009; pp 258−261, 421−423, 546−548, 617−622, 650−652.

(8) Dawson, R. How Significant is a Boxplot Outlier? *J. Statistics Educ.* **2011**, *19* (2), 1−13.

(9) McGill, R.; Tukey, J. W.; Larsen, W. A. Variations of Box Plots. *Am. Stat.* **1978**, *32* (1), 12−16.

(10) DeMan, J. M. *Principles of Food Chemistry*, 3rd ed.; Aspen Publishers, Inc.: Gaithersburg, MD, 1999; pp 167−182, 217.

(11) Damodaran, S. Amino Acids, Peptides, and Proteins. In *Food Chemistry*, 3rd ed.; Fennema, O. R., Ed.; Marcel Dekker, Inc.: New York, 1996; p 365.