

Proportional Visualization of Genotypes and Phenotypes with Rainbow Boxes: Methods and Application to Sickle Cell Disease

Al Hassim Diallo, Gaoussou Camara, Moussa Lo, Ibrahima Diagne, Jean-Baptiste Lamy

▶ **To cite this version:**

Al Hassim Diallo, Gaoussou Camara, Moussa Lo, Ibrahima Diagne, Jean-Baptiste Lamy. Proportional Visualization of Genotypes and Phenotypes with Rainbow Boxes: Methods and Application to Sickle Cell Disease. 2019 23rd International Conference Information Visualisation (IV), Jul 2019, Paris, France. pp.1-6, 10.1109/IV.2019.00010 . hal-03479276

HAL Id: hal-03479276
https://hal.science/hal-03479276

Submitted on 14 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proportional visualization of genotypes and phenotypes with rainbow boxes: methods and application to sickle cell disease

Diallo Al Hassim
LANI, Université Gaston Berger de Saint-louis
B.P. 234 Saint-Louis, Sénégal
Email: alhassimdiallobe@gmail.com

Gaoussou Camara
LIMA, Université Alioune Diop de Bambey
B.P. 30 Bambey, Sénégal
Email: gaoussou.camara@uadb.edu.sn

Moussa Lo
LANI, Université Gaston Berger de Saint-louis
B.P. 234 Saint-Louis, Sénégal
Email: moussa.lo@ugb.edu.sn

Ibrahima Diagne
CERPAD, Université Gaston Berger de Saint-louis
B.P. 234 Saint-Louis, Sénégal
Email: ibrahima.diagne@ugb.edu.sn

Jean-Baptiste Lamy
LIMICS, Université Paris 13, INSERM, Sorbonne Université
93017 Bobigny, France
Email: jean-baptiste.lamy@univ-paris13.fr

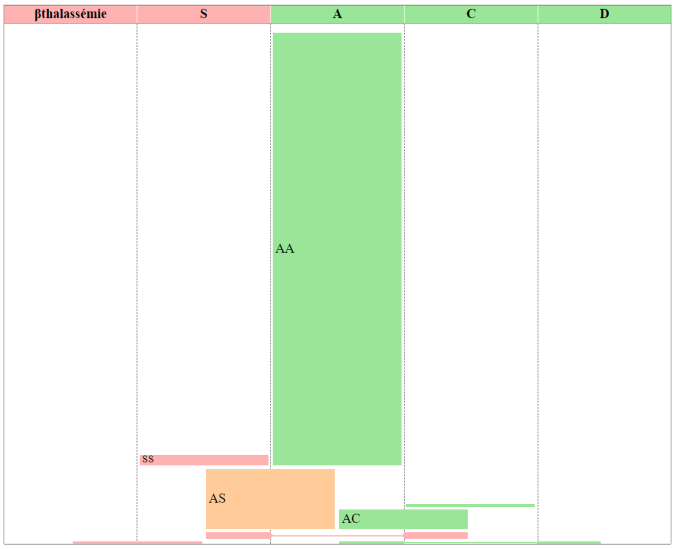


Figure 1. Rainbow boxes displaying the proportion of the various genotypes and the phenotypes of patient screened for sickle cell disease at CERPAD in Senegal, over a 23-month period from June 2017 to February 2019. Five alleles are present (represented in columns), and 8 genotypes are displayed as boxes. Each genotype Light red color represents diseased patients, light orange healthy carriers, and green healthy people. The first observation is that the proportion of healthy patients ("AA") is much higher than the proportions of the other seven genotypes. The proportion of patients with sickle cell disease ("SS", "SC", "Sβthalassemia") is relatively low. On the other hand, healthy carriers ("AS") have a significant proportion.

***Abstract*—Visualization tools allow visual and interactive data exploration for facilitating the interpretation of complex data sets. The Center for Research and Ambulatory Care of the Sickle Cell Disease (CERPAD), leading the neonatal screening of sickle cell program in Saint-Louis of Senegal, aims at setting up a long term database on which data analysis and reporting tools are built for epidemiological and socioanthropological studies.**

In this paper, we propose a tool for the proportional visualization of genotypes and phenotypes, using rainbow boxes, a recently introduced set visualization technique. We propose an improvement for rainbow boxes, featuring area-proportional boxes instead of height-proportional boxes. This new approach is integrated in the SIMENS module of the CERPAD for the visualization of sickle cell data proportion according to genotypes and phenotypes, and by ethnic groups. A qualitative evaluation is provided by two domain experts.

***Index Terms*—set visualization, genotype, phenotype, sickle cell disease.**

I. Introduction

Genetic diseases are among the most common health problems in the world and in Africa in particular. Almost 5% of all pregnancies worldwide result in the birth of a child with a genetic disease. In developing countries, hereditary diseases account for about 15-25% of perinatal and infant mortality [1]. There are more than 6000 different genetic diseases in the world [2]. They are all caused by genetic modifications. This further strengthens the relationships between genotypes and phenotypes [3], after the assumption issued since 1930 by Fisher on the existence of these, as part of his theory of the evolution of dominance [4]. Therefore, understanding, managing and preventing these diseases is essential to improve the quality of life of populations in general and provide them with better quality care. It then becomes imperative to predict the distribution of genotype and phenotype traits in populations for each generation.

Visualization tools allow visual and interactive data exploration. They are increasingly used in genomic epidemiology [5], [6]. Indeed, visualization is frequently used to facilitate the interpretation of complex data sets. Within genomics, visualizing the relationships between multiple genomes provides a framework onto which associated data (geographical, temporal, phenotypic and epidemiological) are added to generate hypotheses and to explore the dynamics of the system under investigation [7].

Many bioinformatics visualization tools exist to explore reference databases for genetic diseases (Example: OMIM, OrphaNet) [8]. These tools can provide insights into the distribution of genotypes and phenotypes, facilitating a prediction that can be directly useful, in helping to interpret the pathogenic character or not of sequence variation discovered within the scope of molecular diagnostics. For example: the PhenoStacks tool can be used to investigate the distribution of patient phenotypes within a single cohort or compare and contrast similarities and differences of phenotype distribution across patients between cohorts [9].

In this paper, we propose a tool for the proportional visualization of genotypes and phenotypes, using rainbow boxes [10], [11], a recently introduced set visualization technique. We propose an improvement for rainbow boxes, featuring area-proportional boxes instead of the previously proposed height-proportional boxes [12]. We also demonstrate the use of this tool on sickle cell disease in Senegal, on real data.

The rest of the paper is organized as follows. Section II presents related works in set visualization. Section III describes how we represented genotypes and phenotypes with rainbow boxes, and we improved rainbow boxes to achieve an area-proportional visualization. Section IV describes the application of the proposed visualization to sickle cell disease in Senegal, and gives the opinions of two medical experts. Section V discusses the method and the results, before section VI concludes.

II. Related works

In the literature, proportional Venn diagrams have been used for the visualization of the repartition and the percentages of patients having various subtypes of a given disorder, or various phenotypes. For example, several such diagrams have been proposed for Chronic Obstructive Pulmonary Disease (COPD) [13] and for Behçet's disease-related manifestations [14]. However, these diagrams are limited to the presentation of phenotypes, but they do not display the underlying genotypes.

Many other set visualization techniques have been proposed. They have been reviewed by Alsakallah *et al.* [15]. These authors distinguished 6 main approaches:

(1) Euler / Venn diagrams and their variants,
(2) overlays on a map,
(3) node-link diagrams,
(4) matrix-based techniques,
(5) aggregation-based techniques, and
(6) scatterplot-based techniques.

For our problem, *i.e.* the proportional visualization of genotypes and phenotypes, overlays on a map are not suitable because the problem has no geographic component. In addition, we need a proportional visualization, which is difficult to achieve with node-link diagrams. Finally, aggregation and scatterplot-based techniques display only aggregated values (*e.g.* means or principal components), but not the entire dataset. Consequently, Euler / Venn diagrams and matrix-based techniques seem the best approaches to this problem.

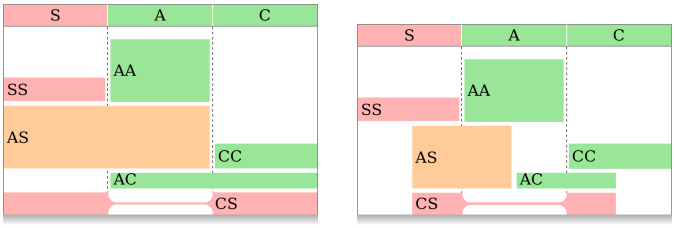


Figure 2. Standard height-proportional rainbow boxes displaying a toy genotype and phenotype dataset (left) and area-proportional rainbow boxes displaying the same dataset (right).

Recently, we introduced rainbow boxes, a matrix-based set visualization technique. Rainbow boxes [10], [11] display the elements to be compared in columns, and the sets in labeled rectangular boxes that cover all the columns corresponding to the elements in the set (see example Figure 2, left). Larger boxes are placed at the bottom and two boxes can be side-by-side as long as they do not cover the same columns. A box can have holes, if the elements in the set are not displayed in consecutive columns (*e.g.* the "CS" box in Figure 2, left). Finding the optimal column order that minimizes the number of holes is a combinatorial optimization problem with a factorial complexity. We proposed heuristic [10] and metaheuristic [16] algorithms to solve this problem in a satisfying time. Finally, we also proposed a proportional variant of rainbow boxes [12], in which box height is an additional visual variable.

III. General method

A. Visual encoding with rainbow boxes

The problem of the visualization of genotypes and phenotypes for a given genetic disorder can be formalized as follows. Let us consider a set of alleles $\mathcal{A} = \{a_1, ..., a_n\}$.

A genotype G consists of either two different alleles, or two copies of the same allele. It is owned by some patients, and it determines the phenotype. Consequently, a genotype can be formalized by a triplet $G = (a, n, p)$ where a is a subset of alleles $a \subseteq \mathcal{A}$ with $1 \leq |a| \leq 2$, $n \in \mathbb{R}$ is the proportion of patients having this genotype, and p is the phenotype. Phenotypes can be described by classifying each patient in a given class, which can be represented by colors. For most genetic disorders, three classes can be considered: diseased (noted Di and represented by a light red color), healthy carrier (Ca, light orange) and healthy non-carrier (He, light green). Healthy carrier patients are not diseased, but they have one recessive allele that can transmit the disease. Thus $p \in \{Di, Ca, He\}$.

The visualization of this type of dataset can be seen as a typed-set visualization problem: each genotype includes a set a with one or two alleles, and has two additional attributes, proportion n and phenotype p. Rainbow boxes can be used

to visualize the dataset as follows. Alleles are considered as elements and represented in columns. Genotypes' sets a are considered as sets and represented by boxes. The height of each box encodes n, the proportion of patients having the genotype, and the box color indicates p, the corresponding phenotype. In addition, column headers representing alleles can be colored in light red or green, for distinguishing pathological alleles from healthy ones.

Figure 2 (left) show an example corresponding to the following toy dataset with 3 alleles, 6 genotypes and 3 phenotypes:

$$\mathcal{A} = \{A, C, S\}$$
$$G_1 = (\{A\}, 30\%, He)$$
$$G_2 = (\{C\}, 12\%, He)$$
$$G_3 = (\{S\}, 11\%, Di)$$
$$G_4 = (\{A, C\}, 7\%, He)$$
$$G_5 = (\{A, S\}, 30\%, Ca)$$
$$G_6 = (\{C, S\}, 10\%, Di)$$

Rainbow boxes require to solve an optimization problem, for finding the optimal column order, *i.e.* the one that minimizes the number of holes in boxes. In this study, the number of columns is low, thus we used a brute-force algorithm that tests all possible orders. In addition, we systematically placed pathological alleles on the left, for facilitating the comparison of several visualizations.

B. Specific improvement of rainbow boxes

A problem is frequently encountered when using rainbow boxes for presenting proportional data: box height encodes the proportion but it is known that the human eye is more sensitive to area than to height [17]. For example, in Figure 2 (left), the "AS" box has the same height as the "AA" box, because both are associated with the same proportion (30%). However, since the "AS" box describes a genotype with two alleles, it covers two columns, while the "AA" box covers a single column. As a consequence, the area of the "AS" box is twice as much as the area of the "AA" box. Thus, one may abusively think that the "AS" box contains twice as many patients as the "AA" box.

The problem of genotype visualization is a particular type of set visualization, because each set cannot have more than 2 elements. This specific subtype of set visualization problem offers an opportunity for avoiding the "height *vs* area" problem described in the previous paragraph. In this specific problem, each box covers either 1 or 2 columns (but never more). Instead of having boxes covering 2 columns, we can put the box in the middle of the two columns, but with the same width as 1-column boxes (see example Figure 2, right). Similarly, boxes with a hole are also reduced to the same width (see "CS" box in Figure 2, right). In addition, this allows a tighter vertical stacking of the boxes, *e.g.* "AC" and "AS" boxes can now be next to each other.

	AA	AC	AS	SS	Total
Wolofs	3528	59	332	6	3925
Peulhs	726	22	62	1	811
Toucouleurs	279	4	24	2	309
Total	4533	85	418	9	5045

Table I
NUMBER OF PATIENTS FOUND IN EACH GENOTYPE, ACCORDING TO ETHNIC GROUPS.

IV. APPLICATIONS TO SICKLE CELL DISEASE IN SENEGAL

A. Context

Sickle cell disease is one of the most common genetic diseases in the world. It affects more than 300,000 babies worldwide each year. The majority of these are in sub-Saharan Africa [18], [19]. In Senegal, 1 in 10 people carry the sickle cell disease gene and each year about 1700 children are born with the disease. The SS homozygous form is manifested by anemia, susceptibility to infections and fit bone and/or abdominal attacks. These painful episodes are a major cause of morbidity [20]. Indeed, in the absence of appropriate care of this form SS, 50% of children die before the age of 5 years [21]. Thus, the Center for Research and Ambulatory Care of the Sickle Cell Disease (CERPAD) has been created at Saint-Louis in Senegal to systematically screen newborns in the maternity wards of the region [22]. The center also ensures follow up and care of positive-screened patients till 18 years old. It is equipped with a biological analysis laboratory to carry out screening tests and also a medical service with a care unit responsible for carrying out the care acts.

We have developed an Health Information System for sickle cell disease screening, as part of the follow-up and management of patients diagnosed with sickle cell disease [23]. It is a module integrated into the National Medical Information System for Senegal (SIMENS) [24], which is a modular HIS, initially designed for public health facilities in level 3 of the sanitary pyramid system of Senegal (PNDS 2009-2018, 2009). SIMENS provides generic modules for invoicing, outpatient clinics, hospitalization, biological analysis [25] and radiology. It also provides specialized modules for surgical units, emergency departments, maternity wards and ENT services. All these modules are implemented and tested in collaboration with physicians from the Regional Hospital (CHRSL) and the Gaston Berger University of Saint-Louis. However, the Rainbow boxes module is integrated into the SIMENS for visualization purposes.

B. Data

In this study, we used a dataset from the sickle cell neonatal screening program in Saint-Louis region of Senegal. We focused on the main maternity ward (CHRSL) which performs about half of deliveries, out of 15 maternities. For each newborn, an information sheet is filled regarding the baby's

medical data, the parent's marital status, contacts and socio-professional status, the ethnic group, etc. The dataset contains 5,045 records collected from 25/04/2017 to 26/02/2019 and involving three ethnic groups: Wolof, Peulh and Toucouleur (Table I).

C. Results

Figure 1 shows the repartition of the genotypes and phenotypes of the patients observed from 25/04/2017 to 26/02/2019, and present in SIMENS Electronic Medical Record for sickle cell disease. 5 alleles were observed: A, C, D, S and βthalassemia. The last two alleles (S and βthalassemia) are pathologic. A, C and S are the most frequent alleles. A is dominant over S, and therefore "AS" persons are healthy carriers. On the contrary, S is dominant over C, and thus "SC" patients suffer from sickle cell disease. "SS" patients, having two pathologic alleles, are of course diseased two.

Each allele is shown in a particular column. Each genotype is represented by a colored box, whose height is proportional to the number of persons having this genotype. In addition, popup bubbles display the exact number of patients in a given box, when the mouse is over the box (*i.e.* following the "details on demand" paradigm). Each box has a label indicating the corresponding genotype, provided that the box is tall enough to contain a label. Otherwise, the genotype can be easily deduced by looking at the column labels. The box horizontal position indicates the alleles of the genotype: if it contains a single allele (*e.g.* "AA" or "SS"), the box is in the corresponding column; if it contains two alleles (*e.g.* "AC" or "AS"), the box is in between the two corresponding columns. Whenever the two columns are not contiguous, the box has a hole: it is split in two parts, joined by a small line (*e.g.* box for genotype "SC" has a hole). Alleles were ordered in order to reduce the total hole's height and length, using a brute-force algorithm.

Box color indicated the phenotype (light red: diseased, light orange: healthy carrier, light green: healthy); the aforementioned dominance relations can be deduced by looking at the colors of the various boxes. On Figure 1, one can observe both the proportion of each genotype (height of boxes) and the proportion of each allele (by summing the heights of boxes in a given column). A is by far the most common allele, followed by S and C. "AA" is the most common genotype, followed by "AS", "AC", "SS" and "SC".

The proposed visualization can also be used for comparing the repartition of the genotypes and phenotypes between several subgroups of the population. For example, Figure 3 compares the genotypes and phenotypes of sickle cell disease, in three of the main ethnic groups in Senegal (Wolofs, Peulhs and Toucouleurs). While the overall "big picture" is similar for the three ethnic groups, some differences are visible. One can see that there are a higher proportion of healthy carriers among Wolofs (taller orange "AS" box), compared with the two other ethnic groups. In addition, the C allele seems less frequent among Touculeurs (much smaller "AC" box). Finally, it seems that there are fewer patient suffering from sickle cell disease among Peulhs (smaller "SS" red box). This is quite unexpected, because the proportion of healthy carriers is similar between these two ethnic groups. However, the difference is clearly not significant due to the very low number of patients involved (actually, 1 and 2 patients).

D. Expert opinions

The proposed visualization (Figure 1 and 3) were shown to two specialists of sickle cell disease screening at the CERPAD biological analysis laboratory. They found the approach very interesting and they appreciated the ability to see in a single image the repartition of alleles, genotypes and phenotypes, simultaneously. The visualization stratified by ethnic groups also interested experts, in order to analyze the importance of ethnicity with regard to sickle cell disease, in the Saint-Louis region, in which Peulhs are an important group.

Experts also suggested the combined visualization of additional phenotypes. Actually, alleles C, D and βthalassemia correspond to hemoglobinopathy. While these alleles may be healthy with regard to sickle cell disease, they can cause other diseases. Experts proposed the use of additional colors for representing the various diseases.

V. Discussion

In this paper, we proposed the use of rainbow boxes for the proportional visualization of genotypes and phenotypes for a given disease. We improved rainbow boxes with area-proportional boxes. We applied the proposed visualization to real data, from the newborns systematic screening program of sickle cell disease in Saint-Louis of Senegal. Two medical experts were introduced to this new visualization, and were greatly interested.

We described a particular set visualization problem, in which sets include only one or two elements. This specific problem would be very difficult to achieve using the traditional approach, *i.e.* proportional Venn diagram. Several tools exist or drawing proportional Venn diagrams [26], [27], but they are usually approximate [28]. In this specific case, the approximation would be very problematic: in facts, it is not possible to draw a four-set proportional Venn diagram using ellipses and displaying all intersection of degree 2, without showing any of a higher degree. Moreover, previous experimental studies have shown that rainbow boxes are easier to read than Euler / Venn diagrams [29].

For some diseases, the phenotype might not be determined only by the genotype, but it may also depend on the environment. In this case, the proposed method can be adapted by splitting genotype boxes in several parts. For example, if only 40% of the "SS" patients developed the disease, we may split the "SS" box in two parts: a red part representing 40% of the total "SS" box height, and a green part representing the remnant 60%.

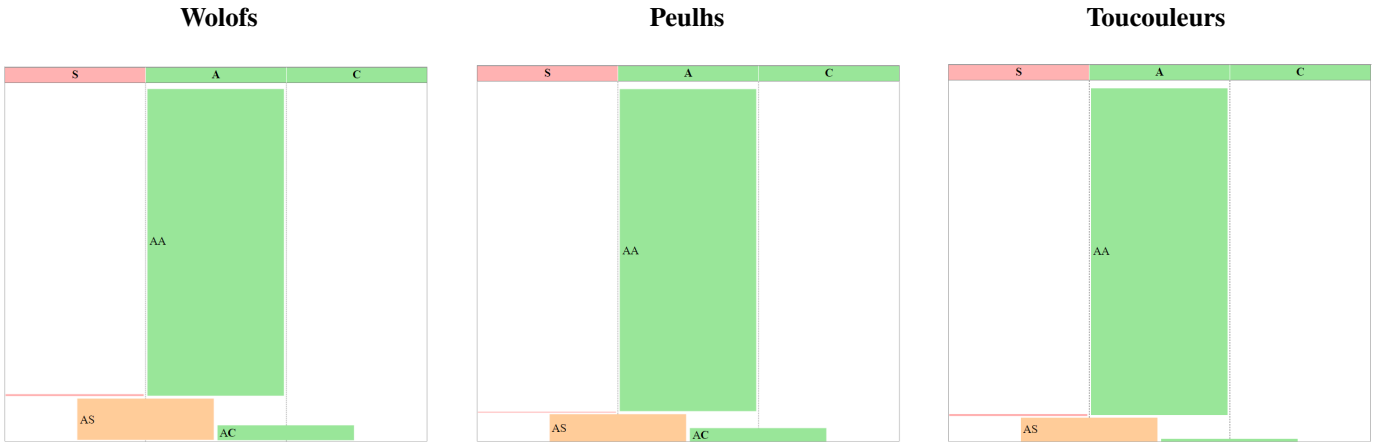


Figure 3. Three rainbow boxes visualizations of genotypes and phenotypes of sickle cell disease, corresponding to three of the main ethnic groups in Senegal (Wolofs, Peulhs and Toucouleurs). These visualizations were limited to the three main alleles (A, C, S) for the sake of simplicity.

For rare diseases, the proportion of diseased phenotypes may be very small compared to the proportion of healthy one. In this case, we suggest to apply a logarithm on the values, as it is a common practice in biology.

VI. Conclusion

We proposed a proportion visualization of the genotypes and phenotypes in a given disorder, using rainbow boxes, and we illustrated our approach on sickle cell disease. The proposed visualization is being integrated in SIMENS for the follow-up of this particular disease. The perspectives of this work include its application to other genetic disorders, the visualization of several phenotypes as suggested by experts, and the implementation of additional subgroup analyses, such as by sex, countries, geographic areas or maternity wards, socioeconomic groups, or time period (*e.g.* years) to visualize the evolution of the disease.

Acknowledgements

This work is part of the SIMENS project which is supported by the Senegalese High Study and Research Ministry through the Scientific and Technic Research Impetus Funding (FIRST), the African Center of Excellence in Mathematics Computer Science and ICT (CEA-MITIC), and the Pierre Fabre Foundation.

References

[1] D. C. Wertz, J. C. Fletcher, and K. Berg, "Les problèmes éthiques rencontrés en génétique médicale," 2001.

[2] T. Morgan, J. Schmidt, C. Haakonsen, J. Lewis, M. Della Rocca, S. Morrison, B. Biesecker, and K. A. Kaphingst, "Using the internet to seek information about genetic and rare diseases: a case study comparing data from 2006 and 2011," *JMIR research protocols*, vol. 3, no. 1, p. e10, 2014.

[3] P. Lopes and J. L. Oliveira, "An innovative portal for rare genetic diseases research: the semantic Diseasecard," *J Biomed Inform*, vol. 46, no. 6, pp. 1108–15, 2013.

[4] Feingold F, "Les gènes modificateurs dans les maladies héréditaires," *Société Française de Génétique*, vol. 16, no. 4, pp. 1–5, 2000.

[5] T. Neugebauer, E. Bordeleau, V. Burrus, and R. Brzezinski, "DNA Data Visualization (DDV): Software for Generating Web-Based Interfaces Supporting Navigation and Analysis of DNA Sequence Data of Entire Genomes," *PloS one*, vol. 10, no. 12, p. e0143615, 2015.

[6] C. B. Nielsen, M. Cantor, I. Dubchak, D. Gordon, and T. Wang, "Visualizing genomes: techniques and challenges," *Nature methods*, vol. 7, no. 3 Suppl, pp. S5–S15, 2010.

[7] S. Argimón, K. Abudahab, R. J. E. Goater, A. Fedosejev, J. Bhai, C. Glasner, E. J. Feil, M. T. G. Holden, C. A. Yeats, H. Grundmann, B. G. Spratt, and D. M. Aanensen, "Microreact: visualizing and sharing data for genomic epidemiology and phylogeography," *Microbial genomics*, vol. 2, no. 11, p. e000093, 2016.

[8] P. N. Robinson, C. J. Mungall, and M. Haendel, "Capturing phenotypes for precision medicine," *Cold Spring Harbor molecular case studies*, vol. 1, no. 1, p. a000372, 2015.

[9] M. Glueck, A. Gvozdik, F. Chevalier, A. Khan, M. Brudno, and D. Wigdor, "PhenoStacks: Cross-Sectional Cohort Phenotype Comparison Visualizations," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 191–200, 2017.

[10] J. B. Lamy, H. Berthelot, C. Capron, and M. Favre, "Rainbow boxes: a new technique for overlapping set visualization and two applications in the biomedical domain," *Journal of Visual Language and Computing*, vol. 43, pp. 71–82, 2017.

[11] J. B. Lamy, H. Berthelot, and M. Favre, "Rainbow boxes: a technique for visualizing overlapping sets and an application to the comparison of drugs properties," in *International Conference Information Visualisation (iV)*, Lisboa, Portugal, 2016, pp. 253–260.

[12] J. B. Lamy and R. Tsopra, "Translating visually the reasoning of a perceptron: the weighted rainbow boxes technique and an application in antibiotherapy," in *International Conference Information Visualisation (iV)*, London, United Kingdom, 2017, pp. 256–261.

[13] E. Marsh S, J. Travers, M. Weatherall, V. Williams M, S. Aldington, M. Shirtcliffe P, L. Hansell A, R. Nowitz M, A. McNaughton A, B. Soriano J, and W. Beasley R, "Proportional classifications of COPD phenotypes," *Thorax*, vol. 63, no. 9, pp. 761–7, 2008.

[14] A. Dinc, A. Bayir, I. Simsek, H. Erdem, S. Pay, and M. Turan, "The proportional Venn diagram of Behçet's disease-related manifestations among young adult men in Turkey," *Clin Exp Rheumatol*, vol. 23, no. 4 Suppl 38, pp. S86–90, 2005.

[15] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers, "Visualizing sets and set-typed data: State-of-the-art and future challenges," in *Eurographics conference on visualization (EuroVis)*, 2014, pp. 1–21.

[16] Lamy JB, *Advances in nature-inspired computing and applications*. Springer, 2019, ch. Artificial Feeding Birds (AFB): a new metaheuristic inspired by the behavior of pigeons, pp. 43–60.

[17] Ware C, *Visual thinking for design*. Burlington, USA: Morgan Kaufmann, 2008.

[18] D. Diallo and G. Tchernia, "Sickle cell disease in Africa," *Curr Opin Hematol*, vol. 9, no. 2, pp. 111–6, 2002.

[19] F. B. Piel and T. N. Williams, "Subphenotypes of sickle cell disease in Africa," *Blood*, vol. 130, no. 20, pp. 2157–2158, 2017.

[20] J. Stinson and B. Naser, "Pain management in children with sickle cell disease," *Paediatr Drugs*, vol. 5, no. 4, pp. 229–41, 2003.

[21] L. Thiam, A. Dramé, I. Z. Coly, F. N. Diouf, N. Seck, D. Boiro, A. A. Ndongo, I. Basse, B. Niang, I. Deme/Ly, A. Sylla, I. Diagne, and O. Ndiaye, "Profils épidemiologiques, cliniques et hématologiques de la drépanocytose homozygote SS en phase inter critique chez l'enfant à Ziguinchor, Sénégal," *The Pan African medical journal*, vol. 28, p. 208, 2017.

[22] C. S. Minkovitz, H. Grason, M. Ruderman, and J. F. Casella, "Newborn Screening Programs and Sickle Cell Disease: A Public Health Services and Systems Approach," *Am J Prev Med*, vol. 51, no. 1 Suppl 1, pp. S39–47, 2016.

[23] A. H. Diallo, G. Camara, J. B. Lamy, M. Lo, I. Diagne, D. Makalou, M. Diop, and D. Doupa, "Toward an information system for sickle cell neonatal screening in Senegal," in *Studies in health technology and informatics (STC)*, vol. 258, 2019, pp. 95–99.

[24] G. Camara, A. H. Diallo, M. Lo, J. N. Tendeng, and S. Lo, "A National Medical Information System for Senegal: Architecture and Services," *Stud Health Technol Inform*, vol. 228, pp. 43–7, 2016.

[25] A. H. Diallo, G. Camara, J. B. Lamy, M. Lo, I. Diagne, D. Makalou, M. Diop, and D. Doupa, "SIMENS-LIS4SC, a laboratory information system for biological tests of sickle cell screening and healthcare," in *Studies in health technology and informatics (MEDINFO)*, vol. accepted, 2019.

[26] Wilkinson L, "Exact and approximate area-proportional circular Venn and Euler diagrams," *IEEE Transactions on visualization and computer graphics*, vol. 18, no. 2, pp. 321–331, 2012.

[27] J. G. Pérez-Silva, M. Araujo-Voces, V. Quesada, and J. Wren, "nVenn: Generalized, quasi-proportional Venn and Euler diagrams," *Bioinformatics*, vol. 34, no. 13, pp. 2322–2324, 2018.

[28] Rodgers P, "A survey of Euler diagrams," *Journal of Visual Languages and Computing*, vol. 25, no. 3, pp. 134–155, 2014.

[29] Lamy JB, "A new diagram for amino acids: User study comparing rainbow boxes to Venn/Euler diagram," in *International Conference Information Visualisation (iV)*, Salerno, Italy, 2018, pp. 361–366.