Illustrating the Theory of Numbers

Martin H. Weissman

Dept of Mathematics, University of California, Santa Cruz; weissman@ucsc.edu

Abstract

We describe challenges and opportunities that arise if one attempts to visualize number theory. The goal is to create graphics which convey a theorem or argument, which tell a story, which are clear and honest, and which enable exploration. In the context of statistics, this is the mainstay of data visualization. In the number-theoretic context, problems of visualization are less familiar but tractable. From our pursuit of graphical excellence, we derive a few broad principles for the visualization of pure and experimental mathematics.

Introduction

Number theory is a strange old field of mathematics. It is not unified by methodology – there is algebraic number theory and analytic number theory and arithmetic geometry. Rather, number theory is unified by its devotion to a set of problems. At its core, these are problems about whole numbers, addition, and multiplication. As its theoretical methods are diverse, so too are the methods of visualization of number theory. One must visualize large data sets, abstract algebraic concepts, and many kinds of geometry. The author encountered these challenges while writing *An Illustrated Theory of Numbers* [10].

The study of multiplication of whole numbers quickly leads to the set of prime numbers, which I view as the most interesting *data set* of mathematics. As data sets go, it belongs to the simplest class: a one-dimensional distribution of whole numbers. But this in itself raises challenges for visualization. Typical problems of data visualization relate to extracting a visual story from high-dimensional data. For primes, the interesting story lies in subtle properties of a one-dimensional distribution. How can one visually tell the story of this data set? The first section addresses this challenge.

The second section addresses the visualization of a finite cyclic group: the multiplicative group $\mathbb{F}_p^\times$ of nonzero numbers modulo a prime p. Finite cyclic groups might be the least interesting, from the standpoint of abstract algebra. But here the generators (called *primitive elements*) are scattered among the representatives $\{1, \ldots, p-1\}$. Abstractly simple, this group is the foundation for cryptographic protocols which, like Diffie-Hellman key exchange [3], are based on the discrete logarithm problem. How does one *illustrate* the structure of $\mathbb{F}_p^\times$, in order to convey its abstract simplicity along with the erratic distribution of its primitive elements?

The third section addresses a major modern (1960s) achievement of number theory: the solution of Gauss's class-number-one problem for negative discriminants. In the language of binary quadratic forms, a theorem of Heegner, Baker, and Stark (see Stark's exposition at [7]) proved a conjecture of Gauss [4, Art. 303].

Theorem 1 (Baker-Heegner-Stark)**.** *The following is the complete list of negative integers Δ for which there is a unique proper equivalence class of binary quadratic forms of discriminant Δ.*

$$-3, -4, -7, -8, -11, -12, -16, -19, -27, -28, -43, -67, -163.$$

Conway's *topograph* provides a visual approach to the study of binary quadratic forms. In particular, his results allow one to reinterpret the above theorem as a statement about triangles with integer side-length. This, in turn, allows one to convey the meaning of the above theorem to an audience outside of number theory.

Art cannot be cleanly separated from function, craft and design. An architect or designer is given a "brief" to follow, with goals (or functions) together with constraints such as audience and space and gravity (a building must stand). Their products are described as elegant, an aesthetic judgment to be sure, if form and function and constraint harmonize rather than conflict. The same is true of mathematical proofs, constrained by logic and prior knowledge and directed towards a theorem.

While I began work in visualization for pedagogical reasons – teaching number theory – the exercise of giving visual form to mathematics led to unexpected results. Though I did not think in this way at the time, I had given myself a brief: to create images which would convey my intuition for number theory (as a specialist) to an audience of students. Like a mathematical proof, the images had to be honest and without contradiction. They had to fit on the page, and my tools were limited to programming in Python, and LaTeX with TikZ.

To present the mathematical form visually – this imperative can drive the creation of beautiful images which convey deep mathematical ideas. Satisfying the imperative is not enough. Elegance seems to come from a dialogue between visual aesthetics and mathematical structure. A first attempt may yield a muddled graphic, ugly colors, awkward proportion. Fixing the visual can raise new mathematical questions – and sometimes, ideally, the visual solution requires an unexpected perspective on mathematics. There can be a moment when the visual form seems to enjoy its constraints, and the viewer can explore mathematics just by sight. That sort of elegance is a rare and valuable artistic reward.

Visualizing the Primes

Euclid (*Elements*, IX.20) proved that the set of prime numbers is infinite. The list

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47 \ldots$$

of prime numbers does not end. Some of the central problems of number theory, e.g., the Riemann Hypothesis, address the distribution of primes. Although their small-scale structure is quite irregular, their large-scale structure is amazingly regular. Heuristically, a whole number n has about a $1/\log(n)$ "chance" of being prime (where log denotes the natural logarithm). For example, among the thousand numbers between 1 000 001 and 1 001 000, one would expect about $1000/\log(1\,000\,000) \approx 72$ primes. (The actual number is 75.)

A minimalist approach to visualizing one-dimensional distributions is given by Brian Hayes in [5, Figure 1]. His method is to simply place a horizontal line segment for each member of a data set. This allows one to perceive the difference between even and irregular spacing, data points which "repel" one another and those that cluster, and trends in the density of data.

From this effective minimalist starting point, one can consider the distribution of primes at different scales: between 1 and 100, 1 and 1000, 1 and 10000. Each column of Figure 1 displays the primes between 1 and 10^e, with $e = 2, \ldots, 9$.

The experiment is one of *direct* display. In the leftmost column, each prime is displayed as a solid bar, 10 points thick. In the second column, each prime is displayed as a line segment, 1 point thick. In the third column, each horizontal line segment stands for a range of 10 numbers. Each segment is shaded gray, according to the number of primes within the range of 10. For example, if a range of 10 numbers contains 4 primes, it is shaded with 40% black ink. By the fourth column, each horizontal line segment stands for a range of 100 numbers. By the last column, each line segment stands for a range of one million numbers. The amount of black ink is precisely (as far as the printer can handle) the proportion of primes within the given range. Note that on some computer screens, a striped appearance is possible due to some "quantization" in digital image rendering.

In the first five columns, the eye can see the transition from small-scale irregularity to large-scale smoothness. The second column is evidently much more "choppy" than the third or fourth; one reason is that the gaps between primes do not grow proportionally to the primes themselves. For example, the average gap between primes near 1000 is about 7, which translates to 7 points of line-thickness in the scale of the second bar. But the average gap between primes near 10000 is about 9, which translates to just 0.9 points of line-thickness in the scale of the third bar. In the fourth bar, the average gap between primes appears only 0.1 points thick, and so most gaps cannot be seen. Even the *largest* gaps become undetectable; the largest gap between primes up to 1 million is a gap of 114 (between 492 113 and 492 227). Such a gap is only 0.114 points thick in the fifth column (where it would first appear). The *absence* of visible gaps in the gray bars reflects the fact that the largest gaps between primes grow much more slowly than the primes themselves.

Among primes up to...	100	1000	10 000	100 000	1 000 000	10 000 000	10 000 000	1 000 000 000
The largest gap is...	8	20	36	72	114	154	220	282

In the right five columns, the human eye is able to detect the slight lightening of the gray bars from left to right, and within each bar, from bottom to top, reflecting the gradual "spreading out" of primes. One could emphasize this by using more ink on the left and less on the right. But the perfect correspondence between proportion of primes and proportion of black ink yields a gradualism that honestly portrays the $1/\log(n)$ heuristic. This also fits with Tufte's design strategy [8] of the "smallest effective difference."

The red horizontal and diagonal lines in Figure 1 are influenced by the cover image of Edward Tufte's *The Visual Display of Quantitative Information* [9]. They are meant to lead the reader's eye from left to right, to imagine (for example) the primes between 1 and 100 squeezed into the bottom tenth of the primes between 1 and 1000, then squeezed into the bottom hundredth of the primes between 1 and 10000.

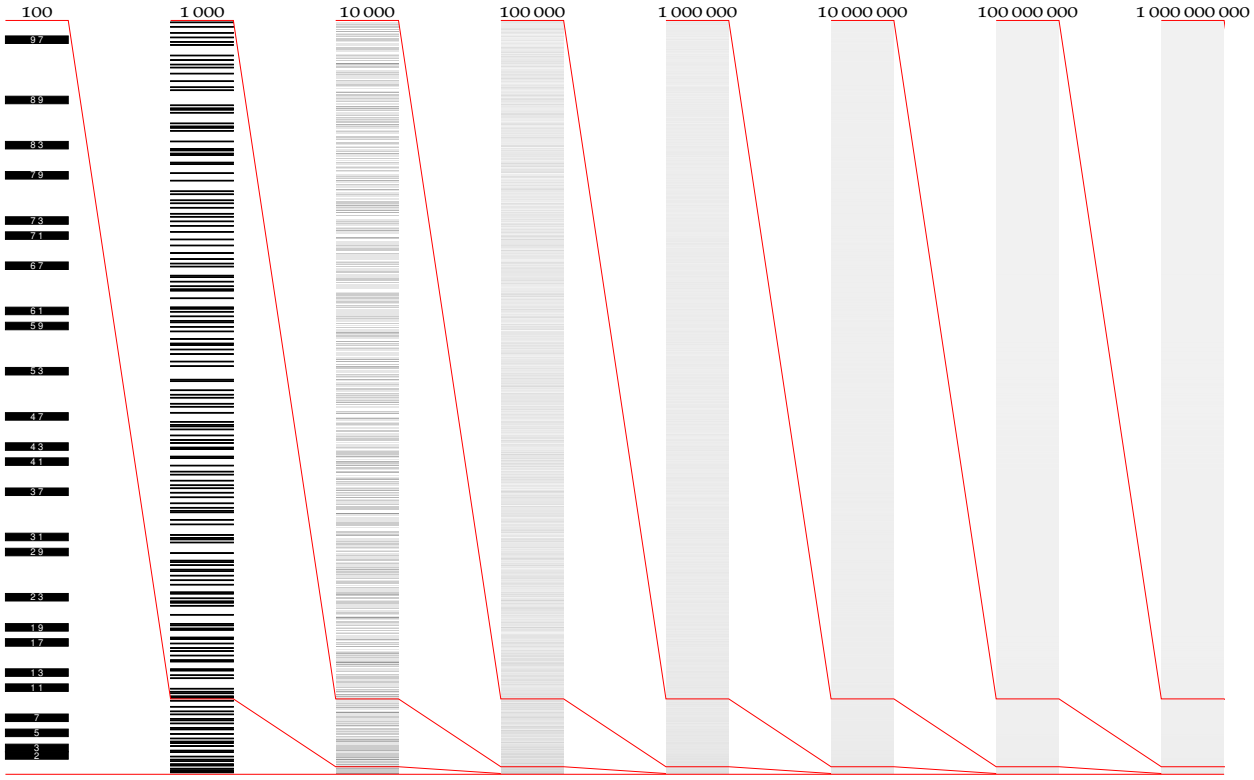

Figure 1: *Primes between one and one billion. From* The Distribution of Primes*, artwork by the author.*

Visualizing a Cyclic Group

Let p be a prime number. The abelian group $\mathbb{F}_p^\times$ may be described in elementary terms as follows. It consists of the set of numbers $\{1, 2, \ldots, p-1\}$ and the following rule for multiplication: the product of two elements a, b of the set is the remainder obtained after dividing ab by p. For example, in $\mathbb{F}_7^\times$, the product of 3 and 5 is 1. We write $3 \cdot 5 = 1 \bmod 7$. The statement that $\mathbb{F}_p^\times$ is an abelian group is the statement that this law of multiplication is commutative and associative, that it has an identity element (1), and that every element of $\mathbb{F}_p^\times$ has an inverse element. For example, the inverse of 3 is 5 in $\mathbb{F}_7^\times$, since $3 \cdot 5 = 1 \bmod 7$.

It is a theorem of Gauss that the group $\mathbb{F}_p^\times$ is *cyclic*. This means that there exists a *primitive element* – an element of $\mathbb{F}_p^\times$ which generates all of $\mathbb{F}_p^\times$ by repeated multiplication. For example, 3 is a primitive element in $\mathbb{F}_7^\times$, because repeated multiplication by 3 (mod 7) traverses *all* numbers in $\{1, \ldots, 6\}$:

$$1 \xrightarrow{\cdot 3} 3 \xrightarrow{\cdot 3} 2 \xrightarrow{\cdot 3} 6 \xrightarrow{\cdot 3} 4 \xrightarrow{\cdot 3} 5 \xrightarrow{\cdot 3} 1.$$

Here an arrow joins a to b if $a \cdot 3 = b \bmod 7$. Finite cyclic groups are the simplest objects of group theory. But here the structure is concealed because the *names* of the elements of the group do not reflect the *structure* of the group in a simple fashion. Contrast $\mathbb{F}_p^\times$ with the *additive* group $(\mathbb{F}_p, +)$ which is also cyclic; a primitive element of the additive group is 1 as seen by the repeated addition (modulo 7):

$$0 \xrightarrow{+1} 1 \xrightarrow{+1} 2 \xrightarrow{+1} 3 \xrightarrow{+1} 4 \xrightarrow{+1} 5 \xrightarrow{+1} 6 \xrightarrow{+1} 0.$$

In fact, every nonzero element of the cyclic group $(\mathbb{F}_p, +)$ is a primitive element (a generator of the cyclic group), while primitive elements of $\mathbb{F}_p^\times$ are distributed haphazardly.

We can interpret this as a statement about the *dynamics* of repeated multiplication in the group $\mathbb{F}_p^\times$. What happens when we start with a number a with $1 \le a \le p-1$, and repeatedly multiply by another number b (modulo p)? If b happens to be a generator, one obtains a cycle of length $p-1$. If not, one obtains a cycle whose length is a proper divisor of $p-1$. In fact, cyclic groups with $p-1$ elements are characterized by the fact that they have a *unique* subgroup of any order dividing $p-1$. The dynamics encode the structure.

To visualize the cyclic group $\mathbb{F}_p^\times$, Figure 3 displays *all* of the dynamics without making choices. A close-up is below in Figure 2.

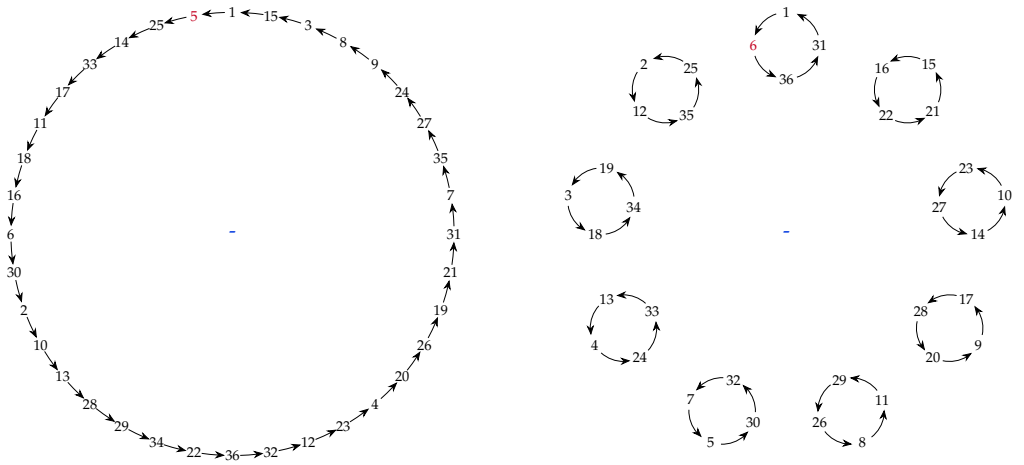

Figure 2: *A close-up from* Epicycles Modulo 37 *(artwork by the author). One the left, 5 is a primitive element in* $\mathbb{F}_{37}^\times$. *On the right, 6 generates a subgroup of order 4 (with 9 cosets).*

Every number b between 1 and 36 generates a subgroup of $\mathbb{F}_{37}^{\times}$. The full Figure 3 displays these subgroups and their cosets for all b. Thus the figure is made of 36 subfigures, as b ranges from 1 to 36 (reading across then down). The subfigures display a single large circle when b is a primitive element, i.e., when $b \in \{2, 5, 13, 15, 17, 18, 19, 20, 22, 24, 32, 35\}$. All in all, the figure displays 36×36 numbers, reflecting all possible multiplications in the group $\mathbb{F}_{37}^{\times}$. But a typical "multiplication table" with a 36x36 rectangular grid, is replaced by "dynamics table" which makes the cyclic structure evident.

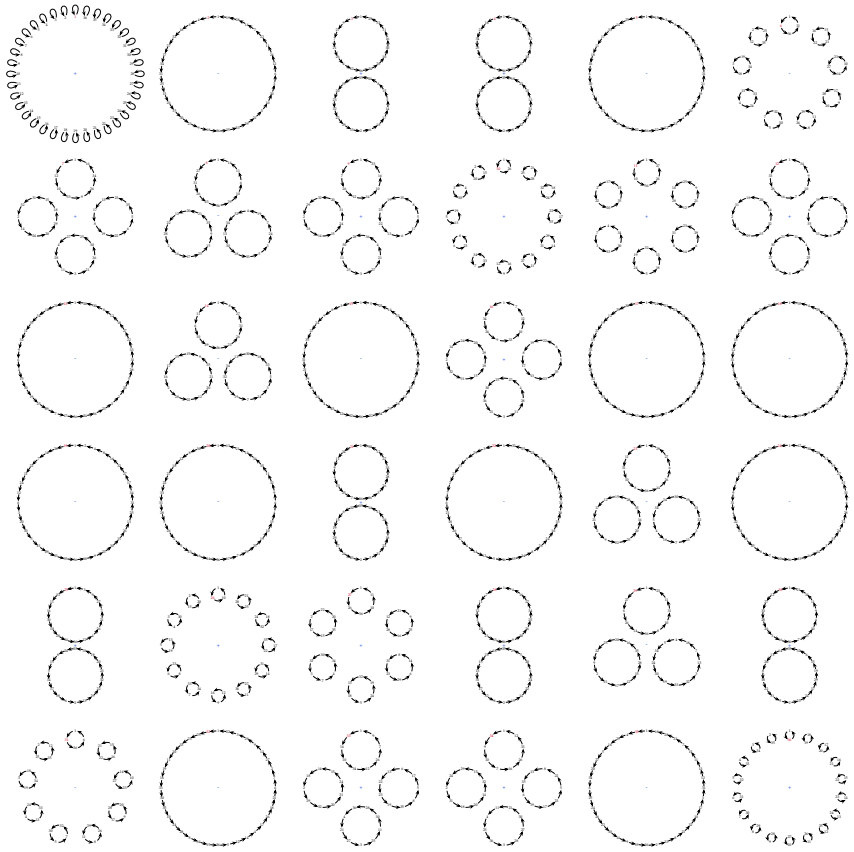

Figure 3: *The dynamics of multiplication, modulo* 37. *From* Epicycles Modulo 37*, artwork by the author.*

Quadratic reciprocity is a landmark theorem in number theory, first proven (in six different ways) by Gauss. Decades later, Zolotarev [11] proved quadratic reciprocity in yet another way. To start, Zolotarev observed that if we view multiplication-by-b as a permutation of $\mathbb{F}_p^{\times}$, then the Legendre symbol $\left(\frac{b}{p}\right)$ coincides with the sign of this permutation. The signs of these permutations are easily readable from a visualization like Figure 3. For example, $\left(\frac{6}{37}\right) = (-1)^9 = -1$, since multiplication by 6 yields 9 cycles of length four (see Figure 3), and every length-four cycle has sign -1. The Legendre symbol is evident in the figure.

An unexpected outcome of this visualization, and a source of its aesthetic appeal, is an *approximate* symmetry. If $x + y = 37$, then an exercise in group theory demonstrates that x and y generate subgroups of $\mathbb{F}_p^{\times}$ of the same order, or else one subgroup has twice the order of the other. For example, in the top-left subfigure and bottom-right subfigure, one finds subgroups of orders 1 and 2, respectively. When $x = 2$ and $y = 35$, we find subgroups of order 36 in both cases; both are primitive elements. When $x = 3$ and $y = 34$, we find subgroups of orders 18 and 9, respectively. This "almost-equality" (up to a possible factor of 2) gives the whole array a "near-symmetry" with respect to 180-degree rotation. Beyond aesthetics, the image allows the reader to explore and notice previously unseen patterns.

Visualizing Binary Quadratic Forms via Integer Triangles

Lastly, we turn to the visualization of a theorem: the complete solution to Gauss's class-number-one problem for binary quadratic forms of negative discriminant. This theorem is difficult to explain to a novice, but a brief survey follows. In his Disquisitiones [4], Gauss makes a detailed study of *binary quadratic forms* (BQFs, hereafter), functions $Q\colon \mathbb{Z}^2 \to \mathbb{Z}$ of the form

$$Q(x, y) = ax^2 + bxy + cy^2.$$

Gauss requires a, b, c to be integers; Gauss required b to be even, but today this assumption is relaxed. Each BQF has a *discriminant*: the number $\Delta(Q) = b^2 - 4ac$ familiar to those who know the quadratic equation. The number $\Delta(Q)$ is either a multiple of 4 (if b is even) or one more than a multiple of 4 (if b is odd).

Two binary quadratic forms are called *properly equivalent* if they are related by a change of variables matrix with determinant 1. To denote this, we write $Q \sim Q'$ if there exists a matrix $g \in SL_2(\mathbb{Z})$ such that $Q'(x, y) = Q((x, y) \cdot g)$ for all $x, y \in \mathbb{Z}$. Properly equivalent BQFs have the same discriminant.

Gauss proved that, for any nonzero Δ, there are finitely many proper equivalence classes of binary quadratic forms of discriminant Δ. The number of proper equivalence classes of *primitive* BQFs of discriminant Δ is called the class number and denoted $h^+(\Delta)$. Here, *primitive* means that $\mathrm{GCD}(a, b, c) = 1$. For negative discriminants, $h^+(\Delta)$ increases gradually (though not monotonically) as $|\Delta|$ increases. In particular, Gauss conjectured that there are only finitely many *negative* integers Δ for which $h^+(\Delta) = 1$. This was solved by Heegner, Baker, and Stark (Theorem 1): the complete list of negative Δ for which $h^+(\Delta) = 1$ is

$$-3, -4, -7, -8, -11, -12, -16, -19, -27, -28, -43, -67, -163.$$

In his *Sensual (quadratic) Form* [2], John H. Conway introduces the *topograph* – an arrangement of the values of a BQF on regions bounded by a ternary-regular tree in the hyperbolic plane. When Q has negative discriminant, Conway's topograph contains a unique *well* or *double-well*: a location at which the values are minimal, in a sense.

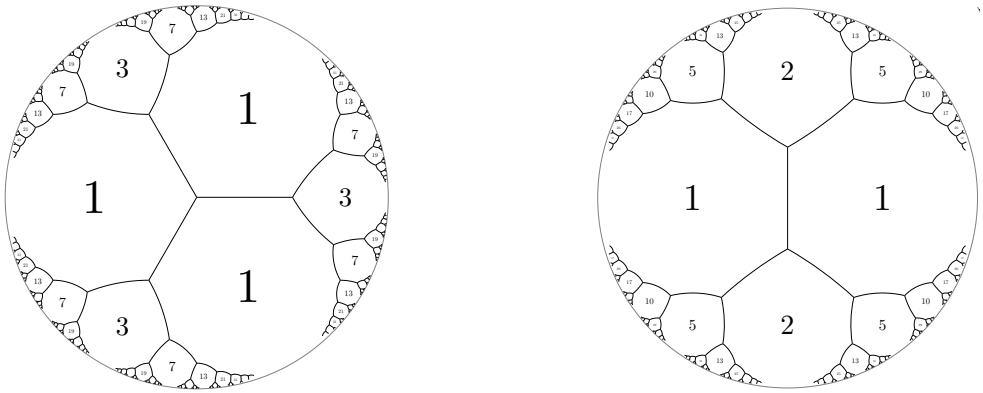

Figure 4: *Conway's topographs for $Q(x, y) = x^2 - xy + y^2$ and $Q(x, y) = x^2 + y^2$. The well on the left has the triple of values* $(1, 1, 1)$. *The double-well on the right has values* $(1, 1, 2)$.

More precisely, a well can be located on the topograph as a triple of numbers (u, v, w) surrounding a point, such that the (weak) triangle inequalities hold:

$$u + v \geq w, \quad v + w \geq u, \quad w + u \geq v.$$

It is a theorem, essentially of Gauss, but reformulated by Conway in terms of wells, that the equivalence class of a binary quadratic form is determined by the triple of numbers surrounding its well. Conversely, every

triple (u, v, w) of positive integers satisfying the triangle inequalities determines a unique equivalence class of binary quadratic forms. The discriminant of these BQFs is given by

$$\Delta = u^2 + v^2 + w^2 - 2uv - 2vw - 2wu.$$

Thus, to visualize the proper equivalence classes of BQFs of negative discriminants, it is equivalent to visualize all triangles with integer side-lengths (u, v, w). We must allow "degenerate triangles" – line segments with $u + v = w$, for example. In Figure 5, we organize these triangles horizontally by discriminant (the number Δ above) and vertically by their smallest side-length.

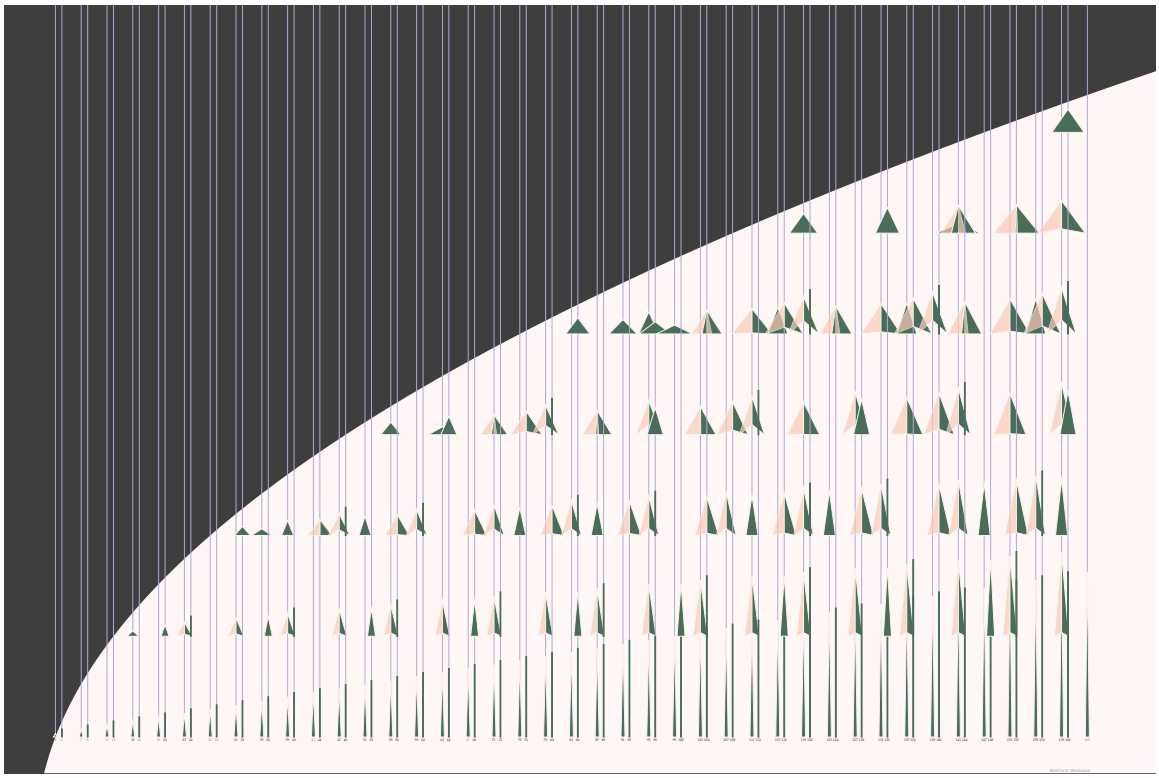

Figure 5: *Primitive triangles with integer sides u, v, w, arranged horizontally by discriminant $u^2 + v^2 + w^2 - 2uv - 2vw - 2wu$ and vertically by* $\min\{u, v, w\}$. *(Artwork by the author.)*

Negative discriminants, $-3, -4, -7, -8, -11, -12, \ldots, -163$ are marked by vertical "pinstripes," and numbered at the bottom. The thinnest triangles along the bottom correspond to the "principal classes" of binary quadratic forms (those with smallest side length 1). Some triangles are isosceles, and others are paired with their mirror-image. This enables the reader to compute $h^+(\Delta)$ by simply counting the number of triangles along a given vertical line. For Gauss's class number (based on *proper* equivalence), the non-isosceles triangles and their mirror images are counted separately. For example, $h^+(-31) = 3$. Isosceles triangles correspond to Gauss's *ambiguous forms*, capturing the "2-torsion of the class group" in modern terminology.

The rightmost line contains a single triangle, reflecting the endpoint of the Baker-Heegner-Stark theorem: -163 is the last negative discriminant with class number one. In fact, the whole theorem becomes a statement about triangles with integer side length – if one continued Figure 5 further to the right, one would never again find a vertical line with only one triangle on it.

The parabolic separation of light and dark reflects a classical bound: the smallest side-length of a triangle is bounded by $\sqrt{\Delta/3}$. Equivalently, the smallest nonzero value of a positive-definite BQF of discriminant Δ is bounded by $\sqrt{\Delta/3}$.

Principles for Visualizing Mathematics

From creating these images, I have distilled a few principles for visualizing mathematics.

First: Directly show the raw data. Try to represent your data points (e.g., all multiplications in $\mathbb{F}_{37}^{\times}$) directly, with simple dots, lines, numbers, etc. Do not bin your data or summarize until the direct portrayal has a chance to suggest patterns.

Second: Bring mathematics into the deep design. The image in Figure 5 began with a simple concept: draw all integer-side triangles, arranged by discriminant and smallest value. When standard data-graphics elements like axes with tick marks, grids, etc., failed, mathematics guided the way. The vertical lines of Figure 5 serve as gridlines at the same time as they exhibit the pattern of possible discriminants. To represent the smallest-value bound $\min\{u, v, w\} \leq \sqrt{\Delta/3}$, I drew a parabolic curve at first. But the curve seemed to clutter the image without conveying the message. Eventually, the curve became the border between dark and light, which underlies the composition of the piece. In terms of perception, the bound became the *ground* and the triangles the *figure*. Mathematics not only provided data, but deeply influenced the design.

Third: Visualize mathematics at multiple scales. When we look at an image, our eyes follow a jumpy path [6] guided by the image's composition. Our graphics can tell a nuanced story if they operate compositionally on a large scale and also offer fine detail (visually and mathematically). The broad structure of Figure 3 reflects the structure of a cyclic group. At a distance, one sees only circles arranged in circles arranged in a matrix. Close up (Figure 2), the fine detail gives the actual numbers. Words or numbers might play a crucial role, especially at the fine scale. Multiple scales are the foundation of Figure 1. The leftmost bar contains primes as individual numerically-labeled figures; the rightmost bar displays only a light gradient.

The field of data visualization is rapidly progressing, from Tufte's new foundations [9] to interaction and animation (e.g., the Javascript package `d3.js` [1]). The time is ripe for expositors of mathematics to understand this field and convey their ideas with a mix of text and graphics.

References

[1] M. Bostock, V. Ogievetsky, and J. Heer. "D3: Data-Driven Documents." *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.

[2] J. H. Conway. *The Sensual (quadratic) Form*, vol. 26 of *Carus Mathematical Monographs*. MAA, Washington, DC, 1997. With the assistance of Francis Y. C. Fung.

[3] W. Diffie and M. E. Hellman "New directions in cryptography." *IEEE Trans. on Information Theory* 22(6):644–654, November 1976.

[4] C. F. Gauss. *Disquisitiones Arithmeticae*. Springer-Verlag, New York, 1986. Translated and with a preface by Arthur A. Clarke.

[5] B. Hayes. "Computing science: The spectrum of Riemannium." *Amer. Scientist*, 91(4):296–300, 2003.

[6] R. Q. Quiroga and C. Pedreira. "How do we see art: an eye-tracker study." *Frontiers in Human Neuroscience*, 5(98):1–9, September 2011.

[7] H. M. Stark. "The Gauss class-number problems." In *Analytic Number Theory*, volume 7 of *Clay Math. Proc.*, pages 247–256. Amer. Math. Soc., Providence, RI, 2007.

[8] E. R. Tufte. *Visual Explanations*. Graphics Press, Cheshire, Connecticut, 1997.

[9] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut, 2001.

[10] M. H. Weissman. *An Illustrated Theory of Numbers*. Amer. Math. Soc., Providence, RI, 2017.

[11] G. Zolotareff. "Nouvelle démonstration de la loi de réciprocité de Legendre." *Nouvelles Annales de Mathématiques. 2e série*, 11:354–362, 1872.