The Book Review Column[1]
by Frederic Green

Department of Mathematics and Computer Science
Clark University
Worcester, MA 02465
email: fgreen@clarku.edu

In this column, we review the following 3 very different books:

1. **Art in the Life of Mathematicians,** edited by Anna Kepes Szemerédi. A compendium of articles by various mathematicians about their individual artistic pursuits, and their reflections on the relationships between math and art. Reviewed by Frederic Green.

2. **Practical Data Science with R,** by Nina Zumel and John Mount. A detailed textbook on data science explained using R as a medium. Reviewed by Allan M. Miller.

3. **Algebraic Coding Theory,** Revised Edition, by Elwyn Berlekamp. A new edition of a classic treatise about this subject by one of its most important contributors. Reviewed by S.V. Nagaraj.

BOOKS THAT NEED REVIEWERS FOR THE SIGACT NEWS COLUMN

Algorithms

1. *Distributed Systems: An algorithmic approach (second edition)*, by Ghosh
2. *Tractability: Practical approach to Hard Problems*, Edited by Bordeaux, Hamadi, Kohli
3. *Recent progress in the Boolean Domain*, Edited by Bernd Steinbach
4. *A Guide to Graph Colouring Algorithms and Applications*, by R.M.R. Lewis

Programming Languages

1. *Selected Papers on Computer Languages* by Donald Knuth

Miscellaneous Computer Science

1. *Algebraic Geometry Modeling in Information Theory* Edited by Edgar Moro
2. *Communication Networks: An Optimization, Control, and Stochastic Networks Perspective* by Srikant and Ying
3. *CoCo: The colorful history of Tandy's Underdog Computer* by Boisy Pitre and Bill Loguidice
4. *Introduction to Reversible Computing*, by Kalyan S. Perumalla
5. *A Short Course in Computational Geometry and Topology*, by Herbert Edelsbrunner
6. *Network Science*, by Albert-László Barabási
7. *Actual Causality*, by Joseph Y. Halpern
8. *The Power of Networks*, by Christopher G. Brinton and Mung Chiang

Computability, Complexity, Logic

1. *The Foundations of Computability Theory,* by Borut Robič
2. *Models of Computation*, by Roberto Bruni and Ugo Montanari
3. *Proof Analysis: A Contribution to Hilbert's Last Problem* by Negri and Von Plato.

Cryptography and Security

1. *Cryptography in Constant Parallel Time,* by Benny Appelbaum
2. *Secure Multiparty Computation and Secret Sharing*, Ronald Cramer, Ivan Bjerre Damgård, and Jesper Buus Nielsen
3. *A Cryptography Primer: Secrets and Promises*, by Philip N. Klein

Combinatorics and Graph Theory

1. *Finite Geometry and Combinatorial Applications*, by Simeon Ball
2. *Introduction to Random Graphs*, by Alan Frieze and Michał Karoński
3. *Erdös–Ko–Rado Theorems: Algebraic Approaches*, by Christopher Godsil and Karen Meagher
4. *Words and Graphs*, by Sergey Kitaev and Vadim Lozin

Miscellaneous Mathematics and History

1. *The Magic of Math*, by Arthur Benjamin
2. *Professor Stewart's Casebook of Mathematical Mysteries* by Ian Stewart

Review of[2]
Art in the Life of Mathematicians
by Anna Kepes Szemerédi
American Mathematical Society, 2015
292 pages, $45.00, Hardcover (ISBN: 978-1-4704-1956-1)

Review by
Frederic Green (fgreen@clarku.edu)
Department of Mathematics and Computer Science
Clark University, Worcester, MA

I am interested in mathematics only as a creative art.
– G.H. Hardy

1 Overview and (My) Background

This is a fascinating volume consisting of a number of essays by mathematicians about art and the role it plays in their lives. Most of them are about the artistic pursuits of the writers, rather than philosophical reflections on how the two fields relate, although the latter subject recurs in many of the articles. It is also not really about the idea that mathematics is itself a kind of art, with its own sense of aesthetics, although that notion is often discussed explicitly, and frequently plays an important role in many of the authors' reflections. It is, first and foremost, about the artistic pursuits of mathematicians and, as appropriate, how those pursuits interact with mathematics.

Needless to say, this is a not typical of AMS publications, and accordingly this will be a typical review neither in content nor in length. To begin with, very much in the spirit of the book, I thought it would be appropriate to say something about my own artistic background, which will make it clear why I was so eager to read (and review) this book.

While trained in theoretical physics and now active in theoretical computer science (long story in itself), I have a deep, life-long interest in art. It's not much of a stretch to say I'm the only non-professional artist in my immediate and extended family. For example, my father was an opera singer and scenic painter; my mother was a textile designer and also a scenic painter and designer (the two of them worked together for many years in that area). However, my primary artistic interest has always been music. Having been brought up with it, I have always loved opera, and I studied the piano (reluctantly at first) from the age of 8. But my active musical interests really took off when I built a harpsichord (from a kit) just before entering graduate school, and my graduate studies in physics were dovetailed with an intensive study of that instrument. For a time, I was also interested in building more instruments, and besides that first kit, built a renaissance lute from scratch, and later a much more elaborate French double-manual harpsichord (again from a kit). Ultimately, my main interest is in the playing and not the building, and over the years I have given a number of performances, both in formal and informal venues – some of you have heard me (on the piano of course) at Dagstuhl. I have the greatest affinity for baroque music, especially Bach, Scarlatti, Couperin, and Froberger, but I'm always happy to venture into repertoire as early as Cabezón and as late as jazz/pop harpsichordist Don Angle. Recently, I have been begun learning a new instrument, the baroque lute (a much different beast than its renaissance ancestor), especially the music of Sylvius Leopold Weiss, an almost exact contemporary (and acquaintance) of Bach.

So when I saw this book, I knew I absolutely *had* to read it! There are of course numerous other members of our community who will be keenly interested; lovers and practitioners of all the arts, and I know in particular several musicians and opera lovers in our midst.

There a few things about these articles that I found notable. One is that there is great commonality in artistic and mathematical creativity, and this is one of the central themes of the book. Another is that artistic interest informs mathematical inspiration much less than mathematical intuition and the problem-solving mentality informs artistic creativity and aesthetic judgement. Yet another notable aspect is diversity. Mathematicians have a great variety of artistic interests (music, art, theater, dance. . .), and in addition, relate their artistic and mathematical activities or interests in many different ways. Some keep them pretty much segregated (I put myself in this category); others manage to combine the two; and in other cases there is a middle ground.

Because of this diversity, it is only fair to discuss (most of) the individual articles, since, given their variety, it is a virtual impossibility to reduce the whole book to a small number of common themes. In the end, I'll return to an attempt to synthesize the overarching messages that this collection conveys, and questions that are inevitably raised.

2 The Essays

After a Preface by Anna Kepes Szemerédi, an Introduction by Aasaf Noar and AnneMarie Perl, and a brief introductory passage by Michael Atiyah on "The Art of Mathematics," the book really takes off with a lengthy essay by Béla Bollobás. In "Opera and Music – A Mathematician's View," Bollobás talks about this ultimate art form. Opera unites music, drama, dance, and all of the visual arts (through costumes, set design, stagecraft, architecture (in the form of the opera house itself), and even, in modern times, film). The number of variables that go into an operatic performance is staggering, even ignoring the work of the composer. The possibilities for error or even disaster are immense, and such possibilities have become actualities many times in operatic history. As Bollobás says, "It is a miracle that occasionally (and in a great opera house, quite often) most of these things click at the same time." This chapter is an intensely personal history of Bollobás's love of opera. He humbly avers that he has no authority to write about opera. However, over the span of many years, he has attended countless performances at many of the word's great opera houses, and his wife is an opera singer. Given that much familiarity and dedication, the result is a very informed exposition on the nature of opera. Even as an opera lover myself, I learned quite a bit from this, not just about his own fascination with the genre, but also about the plots of some operas with which I am not familiar, and many twists and turns in operatic history (including about various opera houses). There isn't much here about music *per se*; the emphasis is really on opera, and the way in which music interacts with it. This chapter goes into quite a bit of detail about specific operas, and the singers, conductors, directors etc. Some of these details might not be meaningful to all readers, but they are nevertheless essential, as they give the reader a sense of what this world is like. One of the passages that is certainly of universal interest was about personality, which, fittingly, is discussed in the context of Richard Wagner, by most accounts one of the great musical (and dramatic) geniuses of all time and, at the same time, a deplorable person. Inevitably, it seems, the case of Wagner raises the question of whether a person's creation should in any way be judged by the quality of the person? Bollobás answers a resounding "no." One can come up with many more examples in the arts (also in music, we have Carlo Gesualdo, who allegedly brutally murdered his wife and her lover; the Baroque painter Caravaggio apparently was also a killer; and Picasso destroyed many lives personally; and the list goes on), and the sciences and humanities have more than their fair share of unsavory characters (Bollobás cites Newton). But the odiousness of an artist's personality should not detract one iota from what that person has created. This seems to be something that many may find hard to accept in these times, but I nevertheless strongly agree with Bollobás's opinion.

"The Fragility of Beauty in Mathematics and Art," by Enrico Bomieri and Sarah Jones Nelson, is a brief philosophical essay on the nature of beauty, with special attention to the way it manifests itself in both areas, including ways in which artistic sensibility has an influence on mathematics, and vice versa. One usually thinks of the influence going more in one direction (from art to mathematics), but mathematics has had a profound influence on art, e.g., through the advent of perspective, as developed by Leon Battista Alberti and Piero della Francesca, the latter both a painter and a mathematician. The "fragility" of the title refers to the relative nature of artistic judgement (depending on "relative standards... that vary in time across cultures"). A similar fragility applies to mathematics, for example in the more subjective realms of peer review (which I think the authors construe more broadly than just what happens in the usual editorial process) and "any consensus of what is verifiably true and beautiful." Along the way the authors present a number of historical examples, in the arts mostly looking at painting and drawing (e.g., Dürer's 1514 engraving *Melencolia*) and, in mathematics, ideas revolving around the golden ratio and Cantor's diagonalization argument.

In "Mathematics, Love, and Tattoos," Edward Frenkel gives a brief history of the film "Rites of Love and Math," which he acted in and produced with Reine Graves. Some of this essay intersects with part of his recent book "Love and Math," also reviewed on these pages (SIGACT News, Volume 46 #2, June 2015). The film, an homage to Yukio Mishima's 1965 film "Rites of Love and Death," is about a mathematician who discovers the "formula of love" (in the event, for the benefit of those who may find the information illuminating, an equation relating two ways of computing correlation functions in a non-perturbative interacting theory of instantons). Because of the dangers the formula might present to the world, he decides to tattoo it onto the body of the woman he loves, paralleling some of the action of Mishima's film. There is an interesting connection between this article and the preceding one on the subject of "fragility." Frenkel is initially challenged by the fact that different viewers interpret the film in very different ways, in sharp contrast to the manner in which a mathematical result will be interpreted in essentially the same way by any mathematician.

In "A Life Not Chosen," Timothy Gowers writes a very personal and enlightening history of his deep relationship with music. His father, Patrick Gowers, was a composer for movies and television, including the notable British series "The Adventures of Sherlock Holmes," produced in the '80s and '90s; his mother was a piano teacher. Both of his parents, but especially his father, had a strong influence on his development as a musician, and he can not recall a time when he did not understand musical notation. He studied both piano and violin from a very young age, and at age 9, he boarded at King's College School, where he had passed the very competitive auditions for the world-famous choir. His interest in composition developed early as well. At Eton, he went on to sing in the College Chapel Choir, and continued studies in both violin (which he played on the main school orchestra) and piano. Later, at Cambridge, he cultivated a long-standing interest in jazz, which originated earlier with his father. There he played piano in a trio, which performed in clubs off-campus as well as at the university.

One subsection I found most intriguing was about the process of *rehearing* music, based on an article by Leonard Meyer. After the first time one hears a work of music, however great, its novelty will quickly fade (especially its surprising moments), and it can become difficult to listen to it again. This is particularly unfortunate for some of the greatest works (Beethoven's Fifth Symphony comes to mind), which are played so tediously often it's easy to forget how truly great they are. Visual works of art such as the Mona Lisa suffer a similar burden. There's not much one can do about this except be careful not to listen to the same piece (and, in the case of recordings, the identical performance) too often. On the other hand, there are some works of music that tolerate repeated hearings more than others, depending on the music and the listener. For example, it's hard for me to think of a work of Bach that I cannot listen to over and over again (within reason). This is especially true of works that I play myself. Gowers does remark on that phenomenon, citing pieces that he listens to as a practicing composer or jazz pianist. In this regard, I would have been very curious to hear more about Gowers' views on re*playing* music, performance being such a different experience than listening. Performers must repeat the same piece, often hundreds of times, and while this can "get old," there are ways to keep the music fresh. Of course, in the case of jazz, interpretations vary greatly from performance to performance, but still I imagine it's possible to get tired of the same tune, even with different improvisations.

While we all know what path Gowers ultimately took, it's quite clear from his background that he could easily have gone in another direction. As far as I can tell, his "Life Not Chosen" can only barely be characterized as "*not* chosen"; music is integral to his life as it exists. Gowers makes a good case (perhaps, in part, unintentionally) that, after all, he really is a musician in addition to being a mathematician, and at the very least blurs the boundary (which, in certain unusual cases, vanishes) between amateur and professional.

"Writing and Performing Mathematics as Metaphor," by Andrew and Jennifer Granville, is the highly unusual story of how a brother and sister team (Andrew is an eminent number theorist, and Jennifer plays many roles in film and theater, including as actor, producer, and writer) made a screenplay out of a mathematical paper. The mathematical ideas involve the intimate relationship between integers and partitions, and the resulting screenplay is entitled "Mathematical Sciences Investigation (MSI): The anatomy of integers and partitions." The name is inspired in part by the TV series "CSI: Crime Scene Investigation," and is written as a murder mystery with two victims, one representing the integers, the other representing permutations. The investigation includes characters based on Gauss and von Neumann, and the protagonist, a composite of Emmy Noether and Sophie Germaine by the name of Emmy Germaine, is a "wanna-be mathematical forensic pathologist." Ultimately, the team of detectives and forensics experts finds that the victims are in fact (in a sense) "twins," and the play's plot and dialog are designed as an exposition of the proof of that fact. Andrew's goal in this project was to avoid merely providing a dramatic scenario lightly decorated with mathematics (e.g., as in movies like *Good Will Hunting* or *Proof*), but rather to *substantively* explain the mathematical ideas. Jennifer's goal was to cast it in terms of a drama that would be suitable for a lay audience, without compromising the mathematical content. It was performed at the Institute for Advanced Study, MSRI and at the Canadian Mathematical Society annual meeting in 2012, and there are plans to make it into a graphic novel. Something to look forward to! The article concludes with an interview of the director, Michael Spencer.

In "Photography," Izabella Łaba talks about her passion for photography, and its relation to her mathematical thinking. It's clear that she regards the two pursuits as distinct activities, although similar problem-solving processes apply to both. Łaba begins by quoting Paul Halmos and G.H. Hardy, who both suggested that mathematicians should really stick to (and immerse themselves in) mathematics alone. Halmos says, "to the extent that ones loves can be ordered, the greatest love of a mathematician . . . is mathematics." Hardy says that a mathematician "would be silly if he surrendered any decent opportunity of exercising his one talent in order to do undistinguished work in other fields." These contentious opinions never sat well with me, and I found myself rooting for Łaba in expressing her contrary views. Even for some of the greatest (if not most) mathematicians, there is more to life than mathematics. In response she answers,

> I feel that "the extent that one's loves can be ordered" can be so limited as to be meaningless in

> practice, and furthermore, that the extent to which one's loves can even be arranged as separate and distinct entities suitable for ordering is not always sufficient to make this line of considerations worth my while.
>
> ...
>
> Contra Hardy, I am not nearly as attached to mathematics as a profession. My career could have forked in a different direction at any number of junctions... I have always had other interests, other passions, and never felt that I should give up on them on the grounds that my talents were lacking in versatility.

Her essay goes on to many interesting observations on the nature of science and mathematics and their relation to art, as well as her personal history on craft, photography, both in its artistic and technical aspects. Some of her images are included, and one of the most striking ones is used on the book's cover. (Incidentally, the volume itself is very attractively produced.)

"Painting and Mathematics," by Peter Lax is a very brief statement on the relation between these two fields. Lax says he does not see the similarity between music and mathematics; it should be obvious to the reader at this point that I think he's missing something. On the other hand, he does point out some key points that relate painting and mathematics, drawing especially on the nature of artistic realism/abstraction on the one hand versus applied mathematics/abstract (e.g. pure) mathematics on the other.

"Mathematics, Art, Civilization," by Yuri I. Manin, is one of the more abstract pieces, and looks more towards mathematics, as viewed through the lens of art and civilization. Manin's purpose is to address what makes mathematics "convincing" (in a similar way in which works of art can have that quality). He explains, through historical context, how language (broadly construed to encompass mathematical syntax and semantics in formal arguments) came to be the means of providing a convincing argument. However, he points to geometric intuition, represented in concrete pictures, as revealing a more fundamental truth. Consider, for example, the simple geometric proof of the identity $(a+b)^2 = a^2 + b^2 + 2ab$, or the more sophisticated geometric ideas underlying Pappus's theorem. Manin then asks what comes first, the discrete or the continuous? Contrary to our usual perceptions, the discrete can emerge from the continuous; for example, the integers can be derived from homotopy theory. By looking at developments in homotopic topology, foreshadowed by the advent of category theory, he argues that discrete ideas (manifested in language, algebra and logic) are becoming subordinate to intuition and imagination (as manifested in continuity and geometry).

This one essay is unfortunately marred by a glitch in production: It appears a page or more was omitted around pg. 178, as there is a jarring transition from the quote of E. H. Gombrich on pg. 177 to a fragment about commutative diagrams that begins in mid-sentence on pg. 179 (in addition to a full page (178) consisting of one figure that is never referenced in the text).

"De Novo Artistic Activity, Origins of Logograms, and Mathematical Intuition" is a second essay by Manin which deals with similar themes to the first, but from quite different perspectives. He begins with the left-brain/right-brain dichotomy, and how it is manifested in individuals, as well as in culture, and ultimately in mathematics, drawing on many observations from history, clinical studies, natural science and personal experience. For example, over time, images drawn by individuals who have been subject to electroconvulsive therapy (ECT), remarkably follow a similar evolutionary sequence (but in reverse) to the way in which logograms emerge out of primitive drawings in natural written language (e.g., how pictographs evolved into Chinese characters). Even going back deeply into prehistory (in primitive cave paintings), one finds evidence that right-brain-oriented thinking (which leads to the primacy of language over image) inhibits more intuitive (and realistic) imagery. The spatial/linguistic dichotomy that arises in the lateral asymmetry in the brain has its mathematical counterpart in the tension between algebra and geometry, or the discrete versus the continuous. Thus explicit ("linguistic") representations of numbers (e.g., the approximation to e used by Napier for his tables of logarithms) can be an obstruction to deeper understandings of numbers themselves (e.g., the relation $e^{i\pi} = -1$ which Napier could not even have guessed at from his calculations).

In "A Drifter of Dadaist Persuasion," Matilde Marcolli talks first about her background. Both of her parents being artists, trained as architects, her engagement in art started at an early age. The "drifting" in the title of this essay, based on a description of Allen Ginsberg's famous poem "Howl," alludes to Marcolli's peripatetic wanderings both in the geographic world, from institution to institution in different countries and continents, as well as from field to field in mathematics. Her art is visual, originally inspired by surrealism with dadaist tendencies, but in more recent times she has worked in a more abstract expressionist vein. The art expresses, in part, the darker sides of mathematics, including both the difficulties of doing mathematics due to its intrinsic challenges, as well as the contentiousness in the worldly social scene. In her 1995 "Tryptich," one panel expresses that hidden, as yet unformed and yet most crucial aspect of mathematical work, "the laboratory of ideas, the alchemical vessel where reactions finally occur." In the central panel of that work she also calls out the tension and conflict that arise in professional circles: Here we see a depiction of the "peacock-like self-aggrandizing narcissism and... reptilian aggressive territoriality, as well as ... spasms of human despair" that one encounters in the social environs of mathematics. These are contributing forces that, in several ways, drove her "professional itinerary through Mathematics and Physics that was far from linear and untroubled." This history hearkens back to that point made earlier in Bollobás's essay about personality versus achievement, and indeed Marcolli corroborates the fact that "beautiful mathematics can be, and often is, produced by horrible human beings ... This removes nothing from the beauty of the mathematics itself." Marcolli has also written several science fiction novels, and most recently poetry.

In "Inside Out, Outside In (Mathematics on a Single Snapshot)," Jaroslav Nešetřil (a mathematician) and Miroslav Petřiček (a philosopher of art) join forces to understand the dichotomy between depth and surface, and how it works in both mathematics and art. This notion of depth versus surface is central in painting, drawing, and photography. This dichotomy, which is closely related to the formal versus the intuitive, or the algebraic versus the geometric, is, in one form or another, a recurring theme in the book, especially among the more philosophical/historical articles. A central thesis of Nešetřil and Petřiček is that "there is nothing particular about mathematics here and that any deep thinking seems to share similar aspects. The similarity consists not in the content but in the form, in the way things are organised and revealed." The dichotomy is very subtle, especially because the two notions of "surface" and "depth" are really very interdependent (similar to the way in which language, formal or otherwise, can be parsed syntactically and superficially devoid of any meaning, but semantically conveys deep meaning). The authors focus on photography in part because a photograph really serves to give a more objective and abstract image of reality than our sense of sight does (as filtered and interpreted by the mind). Analogously, mathematical facts are abstract and objectively true. In this article, the authors focus on one particular photographic image (the "snapshot" of the title), the 1928 photograph "Radio Tower Berlin" by the famous photographer László Moholy-Nagy. By drawing on Moholy-Nagy's ideas, as well as philosophy and the history of photography, the authors analyze the depth versus surface aspects of the image, pointing out how this relates to both formal ("superficial") and intuitive ("deep") mathematical thinking.

"Conversation with Klaus Roth," by Anna Kepes Szemerédi is brief, entertaining and endearing conversation with Klaus Roth about dancing.

Balázs Szegedy writes two short and very eloquent essays. The first one, "Salsa and Mathematics," is about his passion for salsa (he calls himself a "salsaholic") and the relationship between dance and mathematics. Yes, there is one! And of course it's not about the conscious, deliberative side of mathematics, but rather the emergence of a sharp picture from hazy, provisional intuition. His second essay, "Thoughts on Art and Mathematics," expounds on this idea somewhat. This essay begins by asking what art is. This is not an easy question to answer, but one important point is addressed: Art is by no means necessarily beautiful. What makes great art is depth; art "reveals hidden layers of reality." So does mathematics. To underscore this point, Szegedy classifies two types of "pleasures" one derives from doing mathematics. One of them is the pleasure derived from the construction of working mathematical proofs, a relatively methodical activity that is gratifying when things fall into place. The complementary joy is a metaphysical one. If I understand him correctly, I would put it by saying that this reflects the deep truth that one is uncovering via the formal proof. The proof is a mere vehicle (as he puts it, a "machine") to convey that truth, which is in a sense independent of it, since one might find a different proof or a completely different (and more insightful) way of understanding it. It is the same in art.

In "Mathematics and Art," Cédric Villani talks about both the art in mathematics and the mathematics in art. The former is well-known to mathematicians, in their pursuit of beautiful results, structures, and proofs. For the latter, he imagines three types of "bridges." In the first, the artist uses mathematics directly (e.g., in the imagery of fractals). In the second, the artist uses math as inspiration. In the third, mathematical objects may be displayed outside of their mathematical context, which can serve as visual artifacts in and of themselves, divorced from their mathematical meaning. Among other examples, he cites Man Ray's photographs from the Institut Henri Poincaré in the '30's.

The book closes with "Into the Woods," a portfolio of beautiful nature photographs by Vladimir Voevodsky.

3 Summary, Conclusions, and a Question

There are several types of essays in this collection. One can broadly categorize them as follows. First, and most prevalent, are the personal recollections (e.g., Bollobás, Frenkel, Gowers, the Granvilles, Marcolli, and others). Second are the more abstract, philosophical/historical/art-critical syntheses (e.g., Bombieri, Manin, Nešetřil/Petřiček, etc.). Third, in most cases integrated into the other categories, are representative samples of the artworks themselves (Łaba, Marcolli, Voevodsky, etc.). It would have been great to see more of the latter – although of course, that would only be possible for visual art.

This is a rich, varied and fascinating book. I urge anyone who is interested in both art and mathematics to read it. Or, if you are just interested in mathematics, this book could kindle an interest in art. And vice versa.

I close with one issue that gradually came into focus as I was reading through this. Mathematicians are in a never-ending quest for the deepest and most beautiful ideas, relationships, and proofs. Those relationships, connections between seemingly disconnected areas of mathematics (e.g., between elliptic curves and modular forms) can be among the most profound conceptions in mathematics, and there is no denying their beauty. As G. H. Hardy said, "[beauty] is the first test: there is no permanent place in the world for ugly mathematics." I think there are few if any who would disagree with Hardy. However, here's the problem: As Szegedy points out in his essay, there's nothing that says art has to be beautiful. There is, unfortunately, a common misconception that in art, too, "beauty is the first test." But much of the time, and indeed for some of the very greatest works of art, it isn't. Is the Black Painting "Saturn Devouring His Son" of Goya beautiful? Picasso's "Guernica"? The paintings (often grotesque or dark) of Francis Bacon or Anselm Kiefer? Stravinsky's "Rites of Spring"? The opening bars of Strauss's "Elektra"? Even the first movement of Beethoven's Fifth Symphony, for that matter? Many works of art are meant to move, or express feelings, that are far from pleasant, to say nothing of beautiful. Nevertheless, great works of art are working towards a kind of truth that would not be apparent without that work. If you probe any of these works *deeply*, there nevertheless is a kind of beauty to them. The key here is abstraction – it is in the abstract that we may uncover the deep, frequently hidden, beauty in any work of art. Perhaps Andy Warhol was on to something when he said, "Everything has its beauty, but not everyone sees it."

Ultimately, art and mathematics have very similar goals: understanding the world around us more deeply, exploring new means of expressing ourselves. Therefore, it's not at all surprising that art plays such a deep role in the lives of so many mathematicians. Even the creative processes in both have much in common. It is only the outward form which each pursuit takes that makes them appear more different than they really are.

Review of
Practical Data Science with R
by Nina Zumel and John Mount
Published by Manning Publication, 2014
416 pages, Softcover, $49.99 (ISBN 9781617291562)

Review by
Allan M. Miller[3] **(**`allan.m.miller@berkeley.edu`**)**

1 Introduction

Data Science is a new and fast-growing field that arose from the exponential growth in electronic data being generated by mobile phones, tablets, computerized operational enterprise systems, and the World Wide Web. It combines traditional statistics with new statistical learning techniques such as data mining and machine learning to extract knowledge, usually from very large datasets (referred to as "big data"). It is being used in wide-ranging fields, such as bioinfomatics, supply chain optimization, health care, web optimization, marketing, and mobile phone application analytics. R is a programming language that is widely used for Data Science applications.

Zumel and Mount's book "Practical Data Science with R" provides exactly what the title says: practical coverage of commonly used data science methods using the R programming language. Topics include: how to conduct data science projects, manage and explore datasets, choose and evaluate modeling methods, and present results. The R programming language is used throughout the book for managing data, building models, and generating both graphical and tabular results. The book includes useful appendices on basic R language and tools, and relevant statistical concepts.

2 Who Would Benefit From Reading the Book

Because data science is a relatively new field, most practitioners have limited practical experience. The authors, both with Ph.D.'s (Robotics, Computer Science from Carnegie Mellon University, respectively), and substantial experience in analytics and data science, have been at the forefront of the field for a long time. As a result, they have accumulated a large amount of practical experience in the field. This book is perhaps the first of its kind, offering substantial technical as well as practical advice to readers on the data science process, modeling, and programming.

The book can serve as a useful advanced introduction to data science for readers with at least a basic understanding of statistics and computer programming. However, it is not designed to be a beginner's introduction. Readers seeking a purely introductory text would likely find it difficult to grasp many of the concepts covered, but at least can get a valuable overview of the subject. Most would likely return to the book many times as their level of understanding and experience doing data science grows.

[3] Allan Miller, Ph.D., is a San Francisco Bay Area based practicing freelance Data Scientist in the areas of economic, marketing, health care, and environmental analytics. He teaches highly popular courses in R and data science for the U.C. Berkeley Extension program, where he serves on their Data Science Advisory Committee. Dr. Miller is also the co-organizer of the Berkeley R users group.

3 Summary of Contents

Chapter 1, "The Data Science Process" covers the roles and stages in data science projects, emphasizing how to set appropriate expectations for a new project.

Major data science roles are:

The Project Sponsor, who represents the business interests that underly the project, the individual seeking the benefits of the results of the project. The success or failure of the project is defined by whether the sponsor accepts the results. Keys to successfully working with the sponsors are to (1) keep the sponsor informed, and (2) develop a clear understanding of the goals of the project with the sponsor.

The Client, who represents the model's end users' interests. One way to think of the client is as the direct consumer of the results. For example, a Bank President may be the Sponsor of a project, whereas a Loan Officer might be the Client; the Vice President of Marketing might be the Sponsor, the email Campaign Manager might be the Client. Working successfully with clients involves frequent communication, and taking their specific needs and comments to heart.

The Data Scientist is responsible for planning (and in some cases, managing) the data science project, choosing data sources, and the tools to be used to perform the analysis. The Data Scientist is also responsible for data exploration and performing statistical analysis, including the implementation of machine learning models and the evaluation of the results.

The Data Architect is responsible for storing and managing all of the data used in the project.

The Operations role is involved with both acquiring data and delivering the final results.

Chapter 1 also covers the data science process. The underlying principle of the data science process, according to the authors, is the interaction between the data scientist and project stakeholders. Project stages are defined as:

–Defining the goal of the project: sponsor needs, the shortcomings of current practices, resource requirements, and deployment plans. –Data collection: what is available, and is the quality of it good enough? –Modeling for classification, scoring, ranking, clustering, finding relations, and characterization (visualization and report generation). –Model evaluation and critique, assessing accuracy, generalizability, value, and domain theoretical validity. –Presentation and documentation of results, model deployment and maintenance.

Chapter 2, "Loading Data into R," covers the basics of loading data into R from the most common data scourses: small, structured datasets such as CSV, and connecting to relational databases in order to bring the data into R.

Chapter 3, "Exploring Data," covers the initial process of exploring in-hand data for a data science project: summary statistics, visualization for data exploration, and identifying and fixing problems with data. Data summaries often reveal problems with missing data, invalid values (such as negative numbers for age), outliers, and inconsistencies in measurement units.

The authors explain how to use R's summary statistics and graphical visualization capabilities to gain insights into datasets that are both informative in and of themselves, and useful for further analysis when modeling. The chapter includes techniques for building effective visualizations using "out of the box" base R and the more powerful and advanced ggplot package, including useful "rules of thumb" for when to use which type of chart (e.g., bar charts, density plot, or histogram), and when to use a logarithmic versus a normal numeric scale for visual representations. The chapter does not provide comprehensive coverage of R graphics, but does a good job of introducing R graphics programming and effective visualization techniques.

Chapter 4, “Managing Data.” Whereas Chapter 3 discusses how to identify problems with data, this chapter discusses how to fix them. Topics include: how to handle missing data, variable transformations, converting continuous variables to discrete, normalization and rescaling, and logarithmic transformations.

Several approaches are discussed for dealing with missing data. The first is to determine whether the number of observations with missing data is significant, and if not, whether to drop them. In the cases where dropping observations is not appropriate, an additional level may be added to categorical variables, e.g., “missing,” to signify that the observation has no value for this variable. For numeric variables, you may substitute an appropriate measure, such as the mean for non-missing values. The authors suggest that finding an appropriate substitute for missing numeric values warrants serious consideration, and suggest additional sources for guidance on this.

An example is presented showing how continuous variables may be transformed into categorical variables using the R cut function. For data where relative quantities are more meaningful, normalization and scaling are presented as potentially useful transformations for the data. For continuous variables that are skewed or “wide,” logarithmic transformations are suggested and an example is presented for how to implement this.

Sampling techniques are discussed as a means for working with datasets that are too large or otherwise impractical to work with. The authors describe several techniques for splitting datasets into separate test and training subsets for model building, including using sampling, and creating a group column consisting of randomly generated uniformly distributed values that may be used to segment the dataset. Finally, the authors point out the importance of data provenance, suggesting that one or more columns be added to record what version of the data transformation process was used to process the dataset.

Chapter 5, “Choosing and evaluating models,” is highly informative and one of the most valuable parts of the book. It focuses on three critical questions: (1) given a problem (such as credit scoring or fraud classification), how should one choose the appropriate model (problem-to-modeling method mapping), (2) evaluating models, and (3) validating models. The authors provide valuable, well-explained advice on how to choose the most effective type of model in order to solve it; how to evaluate/measure the effectiveness of the model in solving the problem, and how to diagnose problems, and fix them to insure model quality.

Chapter 6, “Memorization methods” (authors’ term), introduces models that perform classfication, or generate expected values, using a training subset of the original data. The chapter emphasizes the danger of using pseudo-independent variables (variables that are in fact a function of the outcome of the model) as predictors in this class of models. The chapter takes the reader on a tour that begins with constructing single variable models (using both categorical and numeric inputs), to more complex multivariate models that include multiple predictors. The code examples in the chapter, while very useful and highly annotated using comments, are a bit difficult to follow and could use more narrative explanation. But a reader who puts the time and effort to study and understand the code examples will be rewarded with a strong foundational understanding of basic data science concepts such as model training, overfitting, prediction, and evaluation.

Chapter 7, “Linear and logistic regression,” presents standard linear and logistic regression from a “data science” perspective. The authors emphasize that linear regression is “a good first technique for modeling quantities,” while logistic regression is “a good first technique for modeling probabilities.” The authors state: “Linear regression is the bread and butter prediction method for statisticians and data scientists . . . you should always try linear regression first.” This is a good example of the wisdom that comes from the experience of the authors, as beginning data scientists often overlook linear modeling and jump directly into using "sexier," more complex modeling methods, such as Random Forests or PCA.

The chapter begins with much of the standard explanation of ordinary least squares regression, starting at a very basic level of how to conceptualize a model, build it, and use it to make predictions in R. Early in the section, there's a short, useful section on how R stores training data in the model, and the problems associated with memory usage and how to overcome them. There's an excellent, detailed discussion of prediction quality, and a detailed and useful explanation of how to interpret the model summary generated by R, how to decompose the model summary component parts (coefficients, standard error, t-values, and p-values), and how to generate and evaluate overall model quality summaries. In general, this section presents a clear, succinct, and R-centric explanation of how to build, interpret, use, and evaluate simple linear models using R.

The section on logistic regression takes the same approach: conceptual overview, coding a logistic regression model in R, making predictions, characterizing model quality, interpreting the logistic regression model summary generated by R, and extracting information for model coefficients from the summary. An extremely helpful graphics-centered discussion of model quality is provided, and a short, excellent discussion of how to identify possibly collinear inputs. The authors also provide a useful, detailed explanation of the model summary generated by R, including overall model quality summaries such as using null and residual deviances to calculate the significance of the observed fit, a linear regression-like pseudo R-squared to measure the goodness of model fit, and the Akaike information criterion (AIC) used to determine the optimal number of input variables to use in a logistic regression model.

Chapter 8, "Unsupervised methods," focuses on cluster analysis and association rules for data exploration as a starting point for building effective experiments and models. The chapter discusses two clustering techniques: hierarchical and k-means clustering.

The chapter begins with in-depth explanations of several widely used distance measures (Euclidean, Hamming, Manhattan, and Cosine) as a means of determining similarity and dissimilarity between groups of observations, followed by some useful tips on data preparation specifically for building clustering models. Hierarchical clustering is then presented in detail, including how to perform the hierarchical clustering using R's hclust() function and plot the resulting dendogram output. In a somewhat technical, but clear and highly informative section, bootstrap evaluation of clusters is presented as a method for evaluating cluster results, followed by a discussion of the heuristics for choosing the number of clusters.

The widely used k-means clustering algorithm is then presented. The discussion is very practical, focusing on approaches for the most effective use of the algorithm, including how to effectively choose the number of clusters and generate a unique stopping point using the kmeansruns() and clusterboot() functions defined in the R fpc package.

Association rules are also discussed as a means for finding attributes of observations that frequently occur together. An overview of association rules is given, followed by a detailed example problem and some methods for discovering associations using the R apriori() function, inspecting the rules underlying the generated associations, and methods for evaluating associations.

Chapter 9, "Exploring advanced methods," presents advanced methods for improving the results generated by the more basic methods covered in earlier chapters. More specifically, the authors discuss bagging and random forests, generalized additive models, kernel methods for data separation, and support vector machines as a means of reducing training variance, relaxing the assumption of non-monotonic effects, and to move beyond the linear separability of data. This chapter is packed with explanations, useful examples, how to implement these methods and evaluate the results using R.

Chapter 10, "Documentation and Deployment," covers all aspects of project documentation: milestone documentation, commenting code, source code control and the deployment of demonstration versions of the final product.

The discussion on documentation focuses on how to use the R package knitr for "reproducible research," generating results that are well documented and easily reproducible. A step-by-step detailed example is presented for how to use RMarkdown and knitr to generate reproducible documents, including code chunks and options, "knitting" RMarkdown to output PDF's, LaTex, or HTML, best practices for commenting code, and a detailed explanation for how to carry out source code control using Git.

The section on deployment centers around a short discussion of how to deploy results as HTTP services using the R Rook HTTP server package, and how to export models built in R to other languages and systems. These techniques may be generally applied to other types of server deployments.

Chapter 11, "Producing effective presentations," covers the presentation of the final product to project sponsors and clients (see Chapter 2). This is a valuable chapter based on the experience the authors have had in presenting the results of data science projects. The key points are that the presentations should be focused on the audience specifics and the purpose of the project. They should include a recap of the goals of the project and how the project work has met those goals. Finally, the authors suggest that presentations should be calibrated to the technical level of the audience, and share the "interesting and convincing work" done by the data scientists.

4 Conclusion

"Practical Data Science with R" is a remarkable book, packed with both valuable technical material about data science, and practical advice for how to conduct a successful data science project. In a field that is so new, and growing so quickly, it is an essential guide for practitioners, especially for the large numbers of new data scientists moving into the field. It is not only a worthile read, it can serve as a useful ongoing technical reference and practical manual for the data science practitioner.

Review of[4]
Algebraic Coding Theory
Revised Edition
Elwyn Berlekamp
World Scientific, 2015
500 pages, Hardback, $118.00 (ISBN 978-981-4635-89-9)
eBook, $94 (ISBN 978-981-4635-91-2)

Review by
S.V. Nagaraj (`svnagaraj@acm.org`)
VIT University, Chennai Campus, India

1 Introduction

Algebraic coding theory is an interesting branch of mathematics which has a great charm for theoreticians and surprisingly has many applications in real life. The term *algebraic coding theory* denotes the sub-field of coding theory where the properties of codes are expressed in algebraic terms. It addresses the design of error-correcting codes for dependable transmission of information across noisy channels, and employs classical and modern algebraic techniques requiring concepts such as finite fields, group theory, and polynomial algebra. It has associations with areas of discrete mathematics as varied as number theory, cryptography, and design theory. The first edition of the book was published in the year 1968 (E. R. Berlekamp, Algebraic Coding Theory, McGraw-Hill Series in Systems Science, New York: McGraw-Hill, 1968) while its second edition in 1984 (E. R. Berlekamp, Algebraic Coding Theory, Laguna Hills, CA: Aegean Park Press, 1984).

2 Summary

The book has sixteen chapters, two appendices, bibliographies for the first and second editions, and an index.

The book begins with an introduction to basic binary codes and ushers in single error-correcting Hamming codes and double error-correcting Bose Chaudhuri Hocquenghem (BCH) codes. The decoding of binary BCH codes necessitates arithmetic operations in the field of binary polynomials modulo an irreducible binary polynomial. Hence, the next three chapters of the book focus on the algebra of polynomials and finite fields. The chapters study arithmetic operations modulo an irreducible binary polynomial, the number of irreducible q-ary polynomials of a given degree, and the structure of finite fields. The discussion then turns to cyclic binary codes. There is an outline of a general decoder for any cyclic binary code. Interestingly, in the index, the author points out that some authors have mistakenly used the term "BCH codes" to mean "cyclic codes," although only a small subset of cyclic codes are BCH codes. He also mentions that even though all BCH codes are cyclic, this fact was not highlighted in the research papers of Bose and Chaudhuri.

The design criteria of cyclic codes depends on the factorization of the polynomial $x^n - 1$. Hence there is a chapter on factorization of polynomials over finite fields. This chapter also talks about quadratic reciprocity, thereby bringing out the links of algebraic coding theory with number theory. Quadratic reciprocity leads to quadratic residue codes which are discussed elsewhere in this book. Here I should digress and mention that the author developed an algorithm for factoring polynomials over large finite fields which became well-known. Its variants have been used in computer algebra systems. This once again highlights the many original contributions of the author to algebraic coding theory. The next chapter is on binary BCH codes for correcting multiple errors, a natural generalization of concepts discussed in the first chapter. Then there is a short overview of non-binary coding which is useful for some situations. The next chapter is about negacyclic codes - a generalization of cyclic codes with respect to the Lee weight over an odd prime field. The Lee weight is an alternative to the weight used by Hamming. Generalized non-binary BCH codes originally attributed to Gorenstein and Zierler are then brought in along with an algorithm for decoding them.

There is a discussion on polynomials defined over finite fields whose roots over such fields may be found with much ease. Such polynomials were studied by the mathematician Ore who called them p polynomials. However, the author prefers to use the terms *linearized polynomials* and *affine polynomials*. The enumeration of information symbols in BCH codes, and its relation to the enumeration of certain types of sequences is discussed in the next chapter. Various interesting bounds with respect to the Hamming and Lee metrics are then mentioned. The author presents several upper and lower bounds relating the distance and the size of the codes.

Codes may be derived by modifying or combining other codes. The author looks at several ways of obtaining codes using known codes by performing operations such as extending, lengthening, puncturing, shortening, augmenting, expurgating, forming a direct product, and concatenating. A chapter provides an introduction to important codes and methods such as threshold decoding for decoding them. Such codes include Srivastava codes, residue codes, Reed-Muller codes, codes based on finite geometries, and convolutional codes. The last chapter of the book focuses on the relationship between weight distributions of codes and their duals. The two appendices contain information related to the polynomials $x^5 + x^2 + 1$ and $x^2 + x + 2$ over the Galois fields $GF(2^5)$ and $GF(5^2)$ respectively. The bibliographies of the first and second editions are also provided. The index is quite helpful.

3 Opinion

The first edition of the book appeared in the year 1968 while its second edition came out in the year 1984. In order to know what modifications were carried out in the editions, it is essential to have copies of them. Unfortunately, since the first edition was published about 48 years ago and the second edition about 32 years ago, copies of those books are hard to find in libraries. In fact, I had no access to them at the time of writing this review. Even obtaining their tables of contents has not been possible. However, on the basis of the prefaces to all the three editions contained in this book, my guess is that not many changes have been made over the years. From the preface to the second edition, it is clear that misprints and errors in the first edition were corrected, many circuit diagrams were revised, and the bibiography was also expanded. However, there appear to be very few differences between the 1984 edition and this revised 2015 edition. The author's updates to the second edition seems to be limited to the addition of the preface to the revised edition. The preface gives the author's historical overview of the evolution of algebraic coding theory from the time of Claude Shannon in the year 1948 (when he published "A Mathematical Theory of Communication", in two parts in the July and October issues of the Bell System Technical Journal) to the year 2002, when Madhu Sudan received the Nevanlinna prize for his numerous contributions, which also included an improved decoding algorithm for Reed-Solomon codes. Reed-Solomon codes and their variants were invented in the 1960s.

For forty years it was assumed that these codes could correct only a certain number of errors. By creating a new decoding algorithm, Sudan demonstrated that the Reed-Solomon codes could correct many more errors than previously thought possible. The author also reminisces on his activities and achievements during that period. The preface to the revised edition contains just six new references which were not mentioned in the second edition. Surprisingly, the bibliographies of the first and second editions are retained separately. So it appears that not much effort was made in producing this *revised* edition. Thus, those who had hoped for a revision may be disappointed.

Developments in the field of algebraic coding theory and its practical applications since the year 1984 have not been included in the revised edition. For example, in 2003, Ralf Koetter and Alexander Vardy presented a polynomial-time soft-decision algebraic list-decoding algorithm for Reed-Solomon codes, which was based upon the work by Sudan and Guruswami. References to the literature since 1984 are not available in the book except for four in the preface to the revised edition. An expanded bibliography providing references to recent publications would be a welcome addition.

The book covers a few error-correcting codes and their properties and some algorithms for decoding these codes. The primary focus of the book is on BCH codes. There is inadequate coverage of Reed-Solomon codes and no coverage of newer codes such as Low Density Parity Check (LDPC) codes (also known as Gallager codes) and turbo codes. Sometimes the codes may be impractical to implement when they are first designed. This was the case with Gallager codes. Now they are widely used in many applications. This highlights that it is important to study algebraic coding theory even from a theoretical perspective as applications may arise much later.

The book focuses more on the theory, mathematics, and design of algebraic codes and to a lesser extent on their implementation for solving engineering problems. The practical applications of algebraic codes have not been detailed in the book. Nonetheless, it would have been preferable to emphasize applications of algebraic coding theory in other areas, such as cryptography and mathematical games. Many digital communication systems, memory devices, semiconductor memories, optical and magnetic memories, and audio devices employ error-correcting codes. The average user may not be cognizant of lots of applications using error correction. Blu-ray discs, broadcast systems such as Digital Video Broadcasting, digital television standards such as ATSC (Advanced Television Systems Committee), two-dimensional bar codes, compact disc players, cell phones, data transmission technologies such as DSL and WiMAX, digital communication standards, digital storage devices, digital video discs, flash memory devices, mobile communications, modems, solid state drives, storage systems such as RAID 6, telephone transmissions, satellite and military communication, deep space communication, aerospace applications, and Quick Response (QR) codes all employ coding techniques. The book has virtually no focus on the numerous patents related to algebraic codes even though the author himself has obtained many patents. Some readers may tend to view the book as a mere aggregation of mathematical topics bereft of practical applications.

Nevertheless, this edition should be welcomed. As the promotional material for the second edition of the book states, "technology changes dramatically as newer integrated circuits and other electronic devices appear every few years, but the fundamental principles and algorithms which provide the best foundations for algebraic coding remain relatively unchanged". This is very true even today, which is why the book continues to be useful even after nearly half a century after its initial publication. Also, the previous two editions are out of print and hard to find. I was amazed to find a copy of the first edition retailing online for US$500. Berlekamp's textbook has been popular, widely cited and used by academics, engineers, and scientists. The author received an award from the IEEE Information Theory Society for this book. It was one of the foremost textbooks on algebraic coding theory. The book was even published in the Russian language in 1971. Much of the material is based on original research by the author. The book is quite suitable for pedagogy as it includes problems at the end of almost all the chapters. It lists starred sections that may be skipped on first reading by readers who do not wish to get bogged down by technical details. Because of its solid background in the enduring foundational principles, the book will continue to be very useful for students and teachers of mathematical sciences, electrical and electronics engineering, computer science, researchers, and practitioners.