# Python For Biologists A Complete Programming Course For Beginners

At last, here is a baseline book for anyone who is confused by cryptic computer programs, algorithms and formulae, but wants to learn about applied bioinformatics. Now, anyone who can operate a PC, standard software and the internet can also learn to understand the biological basis of bioinformatics, of the existence as well as the source and availability of bioinformatics software, and how to apply these tools and interpret results with confidence. This process is aided by chapters that introduce important aspects of bioinformatics, detailed bioinformatics exercises (including solutions), and to cap it all, a glossary of definitions and terminology relating to bioinformatics.

Wiley is proud to announce the publication of the first ever broad-based textbook introduction to Bioinformatics and Functional Genomics by a trained biologist, experienced researcher, and award-winning instructor. In this new text, author Jonathan Pevsner, winner of the 2001 Johns Hopkins University "Teacher of the Year" award, explains problem-solving using bioinformatic approaches using real examples such as breast cancer, HIV-1, and retinal-binding protein throughout. His book includes 375 figures and over 170 tables. Each chapter includes: Problems, discussion of Pitfalls, Boxes explaining key techniques and math/stats principles, Summary, Recommended Reading list, and URLs for freely available software. The text is suitable for professionals and students at every level, including those with little to no background in computer science.

Take Control of Your Data and Use Python with ConfidenceRequiring no prior programming experience, Managing Your Biological Data with Python empowers biologists and other life scientists to work with biological data on their own using the Python language. The book teaches them not only how to program but also how to manage their data. It shows how

A far-reaching course in practical advanced statistics for biologists using R/Bioconductor, data exploration, and simulation.

Concepts in Bioinformatics and Genomics takes a conceptual approach to its subject, balancing biology, mathematics, and programming while highlighting relevant real-world applications and providing students with the tools to compute and analyze biological data. It presents many thought-provoking exercises to stretch students' imaginations, giving them a deeper understanding of the molecular biology, basic probability, software programs, and program-codingmethodology underpinning this exciting field.

This book is suitable for use in a university-level first course in computing (CS1), as well as the increasingly popular course known as CS0. It is difficult for many students to master basic concepts in computer science and programming. A large portion of the confusion can be blamed on the complexity of the tools and materials that are traditionally used to teach CS1 and CS2. This textbook was written with a single overarching goal: to present the core concepts of computer science as simply as possible without being simplistic.

In today's data driven biology, programming knowledge is essential in turning ideas into testable hypothesis. Based on the author's extensive experience, Python for Bioinformatics, Second Edition helps biologists get to grips with the basics of software development. Requiring no prior knowledge of programming-related concepts, the book focuses on the easy-to-use, yet powerful, Python computer language. This new edition is updated throughout to Python 3 and is designed not just to help scientists master the basics, but to do more in less time and in a reproducible way. New developments added in this edition include NoSQL databases, the Anaconda Python distribution, graphical libraries like Bokeh, and the use of Github for collaborative development.

A Primer for Computational Biology aims to provide life scientists and students the skills necessary for research in a data-rich world. The text covers accessing and using remote servers via the command-line, writing programs and pipelines for data analysis, and provides useful vocabulary for interdisciplinary work. The book is broken into three parts: Introduction to Unix/Linux: The command-line is the "natural environment" of scientific computing, and this part covers a wide range of topics, including logging in, working with files and directories, installing programs and writing scripts, and the powerful "pipe" operator for file and data manipulation. Programming in Python: Python is both a premier language for learning and a common choice in scientific software development. This part covers the basic concepts in programming (data types, if-statements and loops, functions) via examples of DNA-sequence analysis. This part also covers more complex subjects in software development such as objects and classes, modules, and APIs. Programming in R: The R language specializes in statistical data analysis, and is also quite useful for visualizing large datasets. This third part covers the basics of R as a programming language (data types, if-statements, functions, loops and when to use them) as well as techniques for large-scale, multi-test analyses. Other topics include S3 classes and data visualization with ggplot2.

In biological research, the amount of data available to researchers has increased so much over recent years, it is becoming increasingly difficult to understand the current state of the art without some experience and understanding of data analytics and bioinformatics. An Introduction to Bioinformatics with R: A Practical Guide for Biologists leads the reader through the basics of computational analysis of data encountered in modern biological research. With no previous experience with statistics or programming required, readers will develop the ability to plan suitable analyses of biological datasets, and to use the R programming environment to perform these analyses. This is achieved through a series of case studies using R to answer research questions using molecular biology datasets. Broadly applicable statistical methods are explained, including linear and rank-based correlation, distance metrics and hierarchical clustering, hypothesis testing using linear regression, proportional hazards regression for survival data, and principal component analysis. These methods are then applied as appropriate throughout the case studies, illustrating how they can be used to answer research questions. Key Features: · Provides a practical course in computational data analysis suitable for students or researchers with no previous exposure to computer programming. · Describes in detail the theoretical basis for statistical analysis techniques used throughout the textbook, from basic principles · Presents walk-throughs · of data analysis tasks using R and example datasets. All R commands are presented and explained in order to enable the reader to carry out these tasks themselves. · Uses outputs from a large range of molecular biology platforms including DNA methylation and genotyping microarrays; RNA-seq, genome sequencing, ChIP-seq and bisulphite sequencing; and high-throughput phenotypic screens. · Gives worked-out examples geared towards problems encountered in cancer research, which can also be applied across many areas of molecular biology and medical research. This book has been developed over years of training biological scientists and clinicians to analyse the large datasets available in their cancer research projects. It is appropriate for use as a textbook or as a practical book for biological scientists looking to gain bioinformatics skills.

Python is rapidly becoming the standard language for many talks in scientific research, and is particularly popular in biology and bioinformatics. One of the great strengths of Python is the ecosystem of tools and libraries that have grown up around it. This book introduces the novice biologist programmer to tools and techniques that make developing Python code easier and faster and will help you to write more reliable, performant programs. Written by a biologist, it focusses on solving the problems that students and researchers encounter every day: How do I make my program run faster? How can I be sure that my results are correct? How do I share this program with my colleagues? How can I speed up the process of writing my code? Chapters include: Environments for development - learn how you can take advantage of different tools for actually writing code, including those designed specifically for scientific work. Organising and sharing code - learn how Python's module and packaging system works, how to effectively reuse code across multiple projects, and how to share your programs with colleagues and the wider world. Testing - learn how automated testing can make your code more reliable, how to catch bugs before they impact your work, and how to edit code with confidence. Performance - learn how to make your code run quickly even on large datasets, how to understand the scaling behaviour of your code, and explore the trade offs involved in designing code. User interfaces - learn how to make your code more user friendly, how to design effective interfaces, and how to automate record-keeping with Python's logging system. About the author Martin started his programming career by learning Perl during the course of his PhD in evolutionary biology, and started teaching other people to program soon after. Since then he has taught introductory programming to hundreds of biologists, from undergraduates to PIs, and has maintained a philosophy that programming courses must be friendly, approachable, and practical. In his academic career, Martin mixed research and teaching at the University of Edinburgh, culminating in a two year stint as Lecturer in Bioinformatics. He now runs programming courses for biological researchers as a full time freelancer. Praise for Martin's previous books "Great, great book. I think this is the perfect book for any biologist to who wants to start learning to code with Python... I didn't know a command-line from a hole in the ground when I first opened up this book, and mere days later I was impressing my colleagues with my own DNA analysis programs." "Zero to writing useful programs in a weekend... Python for Biologists arrived last Thursday, 6/16/16, I spent the whole weekend glued to my laptop in a 2 1/2 day frenzy of coding, and I just finished it -- and came on Amazon to order the next one!" "One of the BEST coding books I've used in a long time. Direct applications in bioinformatics. I bought the advanced python book too." "The most useful guide to Python I've found...I've tried a few Python books, and this is by far the best for me."

Python for BiologistsCreatespace Independent Pub

In biological research, we''re currently in a golden age of data. It''s never been easier to assemble large datasets to probe biological questions. But these large datasets come with their own problems. How to clean and validate data? How to combine datasets from multiple sources? And how to look for patterns in large, complex datasets and display your findings? The solution to these problems comes in the form of Python''s scientific software stack. The combination of a friendly, expressive language and high quality packages makes a fantastic set of tools for data exploration. But the packages themselves can be hard to get to grips with. It''s difficult to know where to get started, or which sets of tools will be most useful. Learning to use Python effectively for data exploration is a superpower that you can learn. With a basic knowledge of Python, pandas (for data manipulation) and seaborn (for data visualization) you''ll be able to understand complex datasets quickly and mine them for biological insight. You''ll be able to make beautiful, informative charts for posters, papers and presentations, and rapidly update them to reflect new data or test new hypotheses. You''ll be able to quickly make sense of datasets from other projects and publications - millions of rows of data will no longer be a scary prospect! In this book, Dr. Jones draws on years of teaching experience to give you the tools you need to answer your research questions. Starting with the basics, you''ll learn how to use Python, pandas, seaborn and matplotlib effectively using biological examples throughout. Rather than overwhelm you with information, the book concentrates on the tools most useful for biological data. Full color illustrations show hundreds of examples covering dozens of different chart types, with complete code samples that you can tweak and use for your own work. This book will help you get over the most common obstacles when getting started with data exploration in Python. You''ll learn about pandas'' data model; how to deal with errors in input files and how to fit large datasets in memory. The chapters on visualization will show you how to make sophisticated charts with minimal code; how to best use color to make clear charts, and how to deal with visualization problems involving large numbers of data points. Chapters include: Getting data into pandas: series and dataframes, CSV and Excel files, missing data, renaming columns Working with series: descriptive statistics, string methods, indexing and broadcasting Filtering and selecting: boolean masks, selecting in a list, complex conditions, aggregation Plotting distributions: histograms, scatterplots, custom columns, using size and color Special scatter plots: using alpha, hexbin plots, regressions, pairwise plots Conditioning on categories: using color, size and marker, small multiples Categorical axes:strip/swarm plots, box and violin plots, bar plots and line charts Styling figures: aspect, labels, styles and contexts, plotting keywords Working with color: choosing palettes, redundancy, highlighting categories Working with groups: groupby, types of categories, filtering and transforming Binning data: creating categories, quantiles, reindexing Long and wide form: tidying input datasets, making summaries, pivoting data Matrix charts: summary tables, heatmaps, scales and normalization, clustering Complex data files: cleaning data, merging and concatenating, reducing memory FacetGrids: laying out multiple charts, custom charts, multiple heat maps Unexpected behaviours: bugs and missing groups, fixing odd scales High performance pandas: vectorization, timing and sampling Further reading: dates and times, alternative syntax

Your research has generated gigabytes of data and now you need to analyse it. You hate using spreadsheets but it is all you know, so what else can you do? This book will transform how you work with large and complex data sets, teaching you powerful programming tools for slicing and dicing data to suit your needs. Written in a fun and accessible

style, this step-by-step guide will inspire and inform non-programmers about the essential aspects of Unix and Perl. It shows how, with just a little programming knowledge, you can write programs that could save you hours, or even days. No prior experience is required and new concepts are introduced using numerous code examples that you can try out for yourself. Going beyond the basics, the authors touch upon many broader topics that will help those new to programming, including debugging and how to write in a good programming style.

This book introduces Python as a powerful tool for the investigation of problems in computational biology, for novices and experienced programmers alike.

Do you have a biological question that could be readily answered by computational techniques, but little experience in programming? Do you want to learn more about the core techniques used in computational biology and bioinformatics? Written in an accessible style, this guide provides a foundation for both newcomers to computer programming and those interested in learning more about computational biology. The chapters guide the reader through: a complete beginners' course to programming in Python, with an introduction to computing jargon; descriptions of core bioinformatics methods with working Python examples; scientific computing techniques, including image analysis, statistics and machine learning. This book also functions as a language reference written in straightforward English, covering the most common Python language elements and a glossary of computing and biological terms. This title will teach undergraduates, postgraduates and professionals working in the life sciences how to program with Python, a powerful, flexible and easy-to-use language.

If you are either a computational biologist or a Python programmer, you will probably relate to the expression "explosive growth, exciting times". Python is arguably the main programming language for big data, and the deluge of data in biology, mostly from genomics and proteomics, makes bioinformatics one of the most exciting fields in data science. Using the hands-on recipes in this book, you'll be able to do practical research and analysis in computational biology with Python. We cover modern, next-generation sequencing libraries and explore real-world examples on how to handle real data. The main focus of the book is the practical application of bioinformatics, but we also cover modern programming techniques and frameworks to deal with the ever increasing deluge of bioinformatics data.

Discover modern, next-generation sequencing libraries from Python ecosystem to analyze large amounts of biological data Key Features Perform complex bioinformatics analysis using the most important Python libraries and applications Implement next-generation sequencing, metagenomics, automating analysis, population genetics, and more Explore various statistical and machine learning techniques for bioinformatics data analysis Book Description Bioinformatics is an active research field that uses a range of simple-to-advanced computations to extract valuable information from biological data. This book covers next-generation sequencing, genomics, metagenomics, population genetics, phylogenetics, and proteomics. You'll learn modern programming techniques to analyze large amounts of biological data. With the help of real-world examples, you'll convert, analyze, and visualize datasets using various Python tools and libraries. This book will help you get a better understanding of working with a Galaxy server, which is the most widely used bioinformatics web-based pipeline system. This updated edition also includes advanced next-generation sequencing filtering techniques. You'll also explore topics such as SNP discovery using statistical approaches under high-performance computing frameworks such as Dask and Spark. By the end of this book, you'll be able to use and implement modern programming techniques and frameworks to deal with the ever-increasing deluge of bioinformatics data. What you will learn Learn how to process large next-generation sequencing (NGS) datasets Work with genomic dataset using the FASTQ, BAM, and VCF formats Learn to perform sequence comparison and phylogenetic reconstruction Perform complex analysis with protemics data Use Python to interact with Galaxy servers Use High-performance computing techniques with Dask and Spark Visualize protein dataset interactions using Cytoscape Use PCA and Decision Trees, two machine learning techniques, with biological datasets Who this book is for This book is for Data data Scientistsscientists, Bioinformatics bioinformatics analysts, researchers, and Python developers who want to address intermediate-to-advanced biological and bioinformatics problems using a recipe-based approach. Working knowledge of the Python programming language is expected.

Life scientists today urgently need training in bioinformatics skills. Too many bioinformatics programs are poorly written and barely maintained--usually by students and researchers who've never learned basic programming skills. This practical guide shows postdoc bioinformatics professionals and students how to exploit the best parts of Python to solve problems in biology while creating documented, tested, reproducible software. Ken Youens-Clark, author of Tiny Python Projects (Manning), demonstrates not only how to write effective Python code but also how to use tests to write and refactor scientific programs. You'll learn the latest Python features and toolsâ??including linters, formatters, type checkers, and testsâ??to create documented and tested programs. You'll also tackle 14 challenges in Rosalind, a problem-solving platform for learning bioinformatics and programming. Create command-line Python programs to document and validate parameters Write tests to verify refactor programs and confirm they're correct Address bioinformatics ideas using Python data structures and modules such as Biopython Create reproducible shortcuts and workflows using makefiles Parse essential bioinformatics file formats such as FASTA and FASTQ Find patterns of text using regular expressions Use higher-order functions in Python like filter(), map(), and reduce()

Bioinformatics Algorithms: Design and Implementation in Python provides a comprehensive book on many of the most important bioinformatics problems, putting forward the best algorithms and showing how to implement them. The book focuses on the use of the Python programming language and its algorithms, which is quickly becoming the most popular language in the bioinformatics field. Readers will find the tools they need to improve their knowledge and skills with regard to algorithm development and implementation, and will also uncover prototypes of bioinformatics applications that demonstrate the main principles underlying real world applications. Presents an ideal text for bioinformatics students with little to no knowledge of computer programming Based on over 12 years of pedagogical materials used by

the authors in their own classrooms Features a companion website with downloadable codes and runnable examples (such as using Jupyter Notebooks) and exercises relating to the book

Offers a structured approach to biological data and the computer tools needed to analyze it, covering UNIX, databases, computation, Perl, data mining, data visualization, and tailoring software to suit specific research needs.

To help with the increasingly large data sets that many scientists deal with, this book illustrates how to use many freely available computing tools to work more powerfully and effectively. The book was born out of the authors' experiences developing tools for their research and to fix other biologist's computational problems.

An introduction to the fundamental concepts of the emerging field of Artificial Chemistries, covering both theory and practical applications. The field of Artificial Life (ALife) is now firmly established in the scientific world, but it has yet to achieve one of its original goals: an understanding of the emergence of life on Earth. The new field of Artificial Chemistries draws from chemistry, biology, computer science, mathematics, and other disciplines to work toward that goal. For if, as it has been argued, life emerged from primitive, prebiotic forms of self-organization, then studying models of chemical reaction systems could bring ALife closer to understanding the origins of life. In Artificial Chemistries (ACs), the emphasis is on creating new interactions rather than new materials. The results can be found both in the virtual world, in certain multiagent systems, and in the physical world, in new (artificial) reaction systems. This book offers an introduction to the fundamental concepts of ACs, covering both theory and practical applications. After a general overview of the field and its methodology, the book reviews important aspects of biology, including basic mechanisms of evolution; discusses examples of ACs drawn from the literature; considers fundamental questions of how order can emerge, emphasizing the concept of chemical organization (a closed and self-maintaining set of chemicals); and surveys a range of applications, which include computing, systems modeling in biology, and synthetic life. An appendix provides a Python toolkit for implementing ACs.

This hands-on tutorial text for non-experts demonstrates biological applications of a versatile modeling and optimization technique.

R is rapidly becoming the standard software for statistical analyses, graphical presentation of data, and programming in the natural, physical, social, and engineering sciences. Getting Started with R is now the go-to introductory guide for biologists wanting to learn how to use R in their research. It teaches readers how to import, explore, graph, and analyse data, while keeping them focused on their ultimate goals: clearly communicating their data in oral presentations, posters, papers, and reports. It provides a consistent workflow for using R that is simple, efficient, reliable, and reproducible. This second edition has been updated and expanded while retaining the concise and engaging nature of its predecessor, offering an accessible and fun introduction to the packages dplyr and ggplot2 for data manipulation and graphing. It expands the set of basic statistics considered in the first edition to include new examples of a simple regression, a one-way and a two-way ANOVA. Finally, it introduces a new chapter on the generalised linear model. Getting Started with R is suitable for undergraduates, graduate students, professional researchers, and practitioners in the biological sciences.

Python for biologists is a complete programming course for beginners that will give you the skills you need to tackle common biological and bioinformatics problems.

Computing is revolutionizing the practice of biology. This book, which assumes no prior computing experience, provides students with the tools to write their own Python programs and to understand fundamental concepts in computational biology and bioinformatics. Each major part of the book begins with a compelling biological question, followed by the algorithmic ideas and programming tools necessary to explore it: the origins of pathogenicity are examined using gene finding, the evolutionary history of sex determination systems is studied using sequence alignment, and the origin of modern humans is addressed using phylogenetic methods. In addition to providing general programming skills, this book explores the design of efficient algorithms, simulation, NP-hardness, and the maximum likelihood method, among other key concepts and methods. Easy-to-read and designed to equip students with the skills to write programs for solving a range of biological problems, the book is accompanied by numerous programming exercises, available at www.cs.hmc.edu/CFB.

Get a comprehensive, in-depth introduction to the core Python language with this hands-on book. Based on author Mark Lutz's popular training course, this updated fifth edition will help you quickly write efficient, high-quality code with Python. It's an ideal way to begin, whether you're new to programming or a professional developer versed in other languages. Complete with quizzes, exercises, and helpful illustrations, this easy-to-follow, self-paced tutorial gets you started with both Python 2.7 and 3.3— the latest releases in the 3.X and 2.X lines—plus all other releases in common use today. You'll also learn some advanced language features that recently have become more common in Python code. Explore Python's major built-in object types such as numbers, lists, and dictionaries Create and process objects with Python statements, and learn Python's general syntax model Use functions to avoid code redundancy and package code for reuse Organize statements, functions, and other tools into larger components with modules Dive into classes: Python's object-oriented programming tool for structuring code Write large programs with Python's exception-handling model and development tools Learn advanced Python tools, including decorators, descriptors, metaclasses, and Unicode processing

Scientific Python is a significant public domain alternative to expensive proprietary software packages. This book teaches from scratch everything the working scientist needs to know using copious, downloadable, useful and adaptable code snippets. Readers will discover how easy it is to implement and test non-trivial mathematical algorithms and will be guided through the many freely available add-on modules. A range of examples, relevant to many different fields, illustrate the language's capabilities. The author also shows how to use pre-existing legacy code (usually in Fortran77) within the Python environment, thus avoiding the need to master the original code. In this new edition, several chapters have been re-written to reflect the IPython notebook style. With an extended index, an entirely new chapter discussing SymPy and a substantial increase in the number of code snippets, researchers and research students will be able to quickly acquire all the skills needed for using Python effectively.

Bioinformatics Algorithms: an Active Learning Approach is one of the first textbooks to emerge from the recent Massive Online Open Course (MOOC) revolution. A light-hearted and analogy-filled companion to the authors' acclaimed online course (http://coursera.org/course/bioinformatics), this book presents students with a dynamic approach to learning bioinformatics. It strikes a unique balance between practical challenges in modern biology and fundamental algorithmic ideas, thus capturing the interest of students of biology and computer science students alike.Each chapter begins with a central biological question, such as "Are There Fragile Regions in the Human Genome?" or "Which DNA Patterns Play the Role of Molecular Clocks?" and then steadily develops the algorithmic sophistication required to answer this question. Hundreds of exercises are incorporated directly into the text as soon as they are needed; readers can test their knowledge through automated coding challenges on Rosalind (http://rosalind.info), an online platform for learning bioinformatics.The textbook website (http://bioinformaticsalgorithms.org) directs readers toward additional educational materials, including video lectures and PowerPoint slides.

Hands-on Data Science for Biologists using Python has been conceptualized to address the massive data handling needs of modern-day biologists. With the advent of high throughput technologies and consequent availability of omics data, biological science has become a data-intensive field. This hands-on textbook has been written with the inception of easing data

analysis by providing an interactive, problem-based instructional approach in Python programming language. The book starts with an introduction to Python and steadily delves into scrupulous techniques of data handling, preprocessing, and visualization. The book concludes with machine learning algorithms and their applications in biological data science. Each topic has an intuitive explanation of concepts and is accompanied with biological examples. Features of this book: The book contains standard templates for data analysis using Python, suitable for beginners as well as advanced learners. This book shows working implementations of data handling and machine learning algorithms using real-life biological datasets and problems, such as gene expression analysis; disease prediction; image recognition; SNP association with phenotypes and diseases. Considering the importance of visualization for data interpretation, especially in biological systems, there is a dedicated chapter for the ease of data visualization and plotting. Every chapter is designed to be interactive and is accompanied with Jupyter notebook to prompt readers to practice in their local systems. Other avant-garde component of the book is the inclusion of a machine learning project, wherein various machine learning algorithms are applied for the identification of genes associated with age-related disorders. A systematic understanding of data analysis steps has always been an important element for biological research. This book is a readily accessible resource that can be used as a handbook for data analysis, as well as a platter of standard code templates for building models.

An introduction to the quantitative modeling of biological processes, presenting modeling approaches, methodology, practical algorithms, software tools, and examples of current research. The quantitative modeling of biological processes promises to expand biological research from a science of observation and discovery to one of rigorous prediction and quantitative analysis. The rapidly growing field of quantitative biology seeks to use biology's emerging technological and computational capabilities to model biological processes. This textbook offers an introduction to the theory, methods, and tools of quantitative biology. The book first introduces the foundations of biological modeling, focusing on some of the most widely used formalisms. It then presents essential methodology for model-guided analyses of biological data, covering such methods as network reconstruction, uncertainty quantification, and experimental design; practical algorithms and software packages for modeling biological systems; and specific examples of current quantitative biology research and related specialized methods. Most chapters offer problems, progressing from simple to complex, that test the reader's mastery of such key techniques as deterministic and stochastic simulations and data analysis. Many chapters include snippets of code that can be used to recreate analyses and generate figures related to the text. Examples are presented in the three popular computing languages: Matlab, R, and Python. A variety of online resources supplement the the text. The editors are long-time organizers of the Annual q-bio Summer School, which was founded in 2007. Through the school, the editors have helped to train more than 400 visiting students in Los Alamos, NM, Santa Fe, NM, San Diego, CA, Albuquerque, NM, and Fort Collins, CO. This book is inspired by the school's curricula, and most of the contributors have participated in the school as students, lecturers, or both. Contributors John H. Abel, Roberto Bertolusso, Daniela Besozzi, Michael L. Blinov, Clive G. Bowsher, Fiona A. Chandra, Paolo Cazzaniga, Bryan C. Daniels, Bernie J. Daigle, Jr., Maciej Dobrzynski, Jonathan P. Doye, Brian Drawert, Sean Fancer, Gareth W. Fearnley, Dirk Fey, Zachary Fox, Ramon Grima, Andreas Hellander, Stefan Hellander, David Hofmann, Damian Hernandez, William S. Hlavacek, Jianjun Huang, Tomasz Jetka, Dongya Jia, Mohit Kumar Jolly, Boris N. Kholodenko, Markek Kimmel, Michal Komorowski, Ganhui Lan, Heeseob Lee, Herbert Levine, Leslie M Loew, Jason G. Lomnitz, Ard A. Louis, Grant Lythe, Carmen Molina-París, Ion I. Moraru, Andrew Mugler, Brian Munsky, Joe Natale, Ilya Nemenman, Karol Nienaltowski, Marco S. Nobile, Maria Nowicka, Sarah Olson, Alan S. Perelson, Linda R. Petzold, Sreenivasan Ponnambalam, Arya Pourzanjani, Ruy M. Ribeiro, William Raymond, William Raymond, Herbert M. Sauro, Michael A. Savageau, Abhyudai Singh, James C. Schaff, Boris M. Slepchenko, Thomas R. Sokolowski, Petr Šulc, Andrea Tangherloni, Pieter Rein ten Wolde, Philipp Thomas, Karen Tkach Tuzman, Lev S. Tsimring, Dan Vasilescu, Margaritis Voliotis, Lisa Weber

A concise introduction to key computing skills for biologists While biological data continues to grow exponentially in size and quality, many of today's biologists are not trained adequately in the computing skills necessary for leveraging this information deluge. In Computing Skills for Biologists, Stefano Allesina and Madlen Wilmes present a valuable toolbox for the effective analysis of biological data. Based on the authors' experiences teaching scientific computing at the University of Chicago, this textbook emphasizes the automation of repetitive tasks and the construction of pipelines for data organization, analysis, visualization, and publication. Stressing practice rather than theory, the book's examples and exercises are drawn from actual biological data and solve cogent problems spanning the entire breadth of biological disciplines, including ecology, genetics, microbiology, and molecular biology. Beginners will benefit from the many examples explained step-by-step, while more seasoned researchers will learn how to combine tools to make biological data analysis robust and reproducible. The book uses free software and code that can be run on any platform. Computing Skills for Biologists is ideal for scientists wanting to improve their technical skills and instructors looking to teach the main computing tools essential for biology research in the twenty-first century. Excellent resource for acquiring comprehensive computing skills Both novice and experienced scientists will increase efficiency by building automated and reproducible pipelines for biological data analysis Code examples based on published data spanning the breadth of biological disciplines Detailed solutions provided for exercises in each chapter Extensive companion website

Treat yourself to a lively, intuitive, and easy-to-follow introduction to computer programming in Python. The book was written specifically for biologists with little or no prior experience of writing code - with the goal of giving them not only a foundation in Python programming, but also the confidence and inspiration to start using Python in their own research. Virtually all of the examples in the book are drawn from across a wide spectrum of life science research, from simple biochemical calculations and sequence analysis, to modeling the dynamic interactions of genes and proteins in cells, or the drift of genes in an evolving population. Best of all, Python for the Life Sciences shows you how to implement all of these projects in Python, one of the most popular programming languages for scientific computing. If you are a life scientist interested in learning Python to jump-start your research, this is the book for you. What You'll Learn Write Python scripts to automate your lab calculations Search for important motifs in genome sequences Use object-oriented programming with Python Study mining interaction network data for patterns Review dynamic modeling of biochemical switches Who This Book Is For Life scientists with little or no programming experience, including undergraduate and graduate students, postdoctoral researchers in academia and industry, medical professionals, and teachers/lecturers. "A comprehensive introduction to using Python for computational biology... A lovely book with humor and perspective" -- John Novembre, Associate Professor of Human Genetics, University of Chicago and MacArthur Fellow "Fun, entertaining, witty and darn useful. A magical portal to the big data revolution" -- Sandro Santagata, Assistant Professor in Pathology, Harvard Medical School "Alex and Gordon's enthusiasm for Python is contagious" -- Glenys Thomson

Professor of Integrative Biology, University of California, Berkeley

Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tools provides a coherent and friendly treatment of bioinformatics for any student or scientist within biology who has not routinely performed bioinformatic analysis. The book discusses the relevant principles needed to understand the theoretical underpinnings of bioinformatic analysis and demonstrates, with examples, targeted analysis using freely available web-based software and publicly available databases. Eschewing non-essential information, the work focuses on principles and hands-on analysis, also pointing to further study options. Avoids non-essential coverage, yet fully describes the field for beginners Explains the molecular basis of evolution to place bioinformatic analysis in biological context Provides useful links to the vast resource of publicly available bioinformatic databases and analysis tools Contains over 100 figures that aid in concept discovery and illustration

The second edition of this best-selling Python book (over 500,000 copies sold!) uses Python 3 to teach even the technically uninclined how to write programs that do in minutes what would take hours to do by hand. There is no prior programming experience required and the book is loved by liberal arts majors and geeks alike. If you've ever spent hours renaming files or updating hundreds of spreadsheet cells, you know how tedious tasks like these can be. But what if you could have your computer do them for you? In this fully revised second edition of the best-selling classic Automate the Boring Stuff with Python, you'll learn how to use Python to write programs that do in minutes what would take you hours to do by hand--no prior programming experience required. You'll learn the basics of Python and explore Python's rich library of modules for performing specific tasks, like scraping data off websites, reading PDF and Word documents, and automating clicking and typing tasks. The second edition of this international fan favorite includes a brand-new chapter on input validation, as well as tutorials on automating Gmail and Google Sheets, plus tips on automatically updating CSV files. You'll learn how to create programs that effortlessly perform useful feats of automation to: • Search for text in a file or across multiple files • Create, update, move, and rename files and folders • Search the Web and download online content • Update and format data in Excel spreadsheets of any size • Split, merge, watermark, and encrypt PDFs • Send email responses and text notifications • Fill out online forms Step-by-step instructions walk you through each program, and updated practice projects at the end of each chapter challenge you to improve those programs and use your newfound skills to automate similar tasks. Don't spend your time doing work a well-trained monkey could do. Even if you've never written a line of code, you can make your computer do the grunt work. Learn how in Automate the Boring Stuff with Python, 2nd Edition.

Powerful, flexible, and easy to use, Python is an ideal language for building software tools and applications for life science research and development. This unique book shows you how to program with Python, using code examples taken directly from bioinformatics. In a short time, you'll be using sophisticated techniques and Python modules that are particularly effective for bioinformatics programming. Bioinformatics Programming Using Python is perfect for anyone involved with bioinformatics -- researchers, support staff, students, and software developers interested in writing bioinformatics applications. You'll find it useful whether you already use Python, write code in another language, or have no programming experience at all. It's an excellent self-instruction tool, as well as a handy reference when facing the challenges of real-life programming tasks. Become familiar with Python's fundamentals, including ways to develop simple applications Learn how to use Python modules for pattern matching, structured text processing, online data retrieval, and database access Discover generalized patterns that cover a large proportion of how Python code is used in bioinformatics Learn how to apply the principles and techniques of object-oriented programming Benefit from the "tips and traps" section in each chapter

This accessible and classroom-tested textbook/reference presents an introduction to the fundamentals of the emerging and interdisciplinary field of data science. The coverage spans key concepts adopted from statistics and machine learning, useful techniques for graph analysis and parallel programming, and the practical application of data science for such tasks as building recommender systems or performing sentiment analysis. Topics and features: provides numerous practical case studies using real-world data throughout the book; supports understanding through hands-on experience of solving data science problems using Python; describes techniques and tools for statistical analysis, machine learning, graph analysis, and parallel programming; reviews a range of applications of data science, including recommender systems and sentiment analysis of text data; provides supplementary code resources and data at an associated website.

For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—IPython, NumPy, Pandas, Matplotlib, Scikit-Learn, and other related tools. Working scientists and data crunchers familiar with reading and writing Python code will find this comprehensive desk reference ideal for tackling day-to-day issues: manipulating, transforming, and cleaning data; visualizing different types of data; and using data to build statistical or machine learning models. Quite simply, this is the must-have reference for scientific computing in Python. With this handbook, you'll learn how to use: IPython and Jupyter: provide computational environments for data scientists using Python NumPy: includes the ndarray for efficient storage and manipulation of dense data arrays in Python Pandas: features the DataFrame for efficient storage and manipulation of labeled/columnar data in Python Matplotlib: includes capabilities for a flexible range of data visualizations in Python Scikit-Learn: for efficient and clean Python implementations of the most important and established machine learning algorithms

Computational Genomics with R provides a starting point for beginners in genomic data analysis and also guides more advanced practitioners to sophisticated data analysis techniques in genomics. The book covers topics from R programming, to machine learning and statistics, to the latest genomic data analysis techniques. The text provides accessible information and explanations, always with the genomics context in the background. This also contains practical and well-documented examples in R so readers can analyze their data by simply reusing the code presented. As the field of computational genomics is interdisciplinary, it requires different starting points for people with different backgrounds. For example, a biologist might skip sections on basic genome biology and start with R programming, whereas a computer scientist might want to start with genome biology. After reading: You will have the basics of R and be able to dive right into specialized uses of R for computational genomics such as using Bioconductor packages. You will be familiar with statistics, supervised and unsupervised learning techniques that are important in data modeling, and exploratory analysis of high-dimensional data. You will understand genomic intervals and operations on them that are used for tasks such as aligned read counting and genomic feature annotation. You will know the basics of processing and quality checking high-throughput sequencing data. You will be able to do sequence analysis, such as calculating GC content for parts of a genome or finding transcription factor binding sites. You will know about visualization techniques used in genomics, such as heatmaps, meta-gene plots, and genomic track visualization. You will be familiar with analysis of different high-throughput sequencing data sets, such as RNA-seq, ChIP-seq, and BS-seq. You will know basic techniques for integrating and interpreting multi-omics datasets. Altuna Akalin is a group leader and head of the Bioinformatics and Omics Data Science Platform at

the Berlin Institute of Medical Systems Biology, Max Delbrück Center, Berlin. He has been developing computational methods for analyzing and integrating large-scale genomics data sets since 2002. He has published an extensive body of work in this area. The framework for this book grew out of the yearly computational genomics courses he has been organizing and teaching since 2015.

"This book is an introduction to the Python programming language for complete beginners – those who have never written a program before, or who are just getting started with programming"-back cover.

Advanced Python for Biologists is a programming course for workers in biology and bioinformatics who want to develop their programming skills. It starts with the basic Python knowledge outlined in Python for Biologists and introduces advanced Python tools and techniques with biological examples. You'll learn: - How to use object-oriented programming to model biological entities - How to write more robust code and programs by using Python's exception system - How to test your code using the unit testing framework - How to transform data using Python's comprehensions - How to write flexible functions and applications using functional programming - How to use Python's iteration framework to extend your own object and functions Advanced Python for Biologists is written with an emphasis on practical problem-solving and uses everyday biological examples throughout. Each section contains exercises along with solutions and detailed discussion.