The proposed book titled *Statistical Genetics Analyses with R* by Shirin Glander is potentially very exciting because it promises to include updated reproducible data analysis examples with the R programming language covering the following topics in statistical genetics: quantitative genetics, population genetics, evolutionary genetics, complex genetic diseases and genetic epidemiology. While the majority of these fields are well established, there are recent developments in the form of software that deserve to be demonstrated in the form of a book to be used the classroom.

In my opinion, the most useful contribution will be the demonstration of recent R packages that can be used to analyze statistical genetic data. While this has the potential to be impactful in the classroom, I have a few concerns that should be addressed:

- There are many published books that cover this material (theory) in great detail. I would recommend to the author of just briefly introducing these topics (referencing other books to learn about the theory) and focus the majority of the book on providing multiple case studies from these topics (e.g. in each chapter, provide a complete data analysis starting from the preprocessing the data, exploratory data analysis, applying models / inference, summarizing the results). The author can highlight various functions from R packages throughout the data analysis in a supporting way. This would be the most ideal way for an instructor to directly use the examples in this book in the classroom.
- Currently, the provided chapter in the proposed book feels like the author is just demonstrating the functions of various R packages in a set of disconnected datasets. The vignette of an R package can be used demonstrate the functions, without providing greater context of a data analysis. Instead, I think the book would be greatly enhanced and be much more widely used in the classroom if the author would frame each chapter as a complete data analysis of one or two datasets, demonstrating how the R functions can be used in various stages of the data analysis.

If the author simply demonstrates the functions of the R packages without framing it in the context of a complete data analysis, I would not suggest using the book. If the author provides multiple case-studies by focusing on one to two data sets per chapter, I would be much more likely to use it.

In addition, I have a few other concerns and recommendations to the author before writing this book:

- I am concerned about the target audience. The author states she will “primarily address users without mathematical background but with a strong foundation in biology. A basic knowledge of R is assumed”. In my experience, the students who have a strong foundation in biology without a mathematical background do not have a basic knowledge in R. If this book is to be used in the classroom, I would encourage the author to introduce the basics of R or at a minimum provide a set of resources that can be used to learning R prior to starting this book.

- I believe the author needs to incorporate the use of messy, high-throughput data examples. It is common for books such the one proposed here to use “small” or contrived data examples instead what is commonly found when analyzing real data. Including the length of time to run various functions on different sizes of data would also be valuable.
- I would recommend that the author create a corresponding GitHub repository to provide online exercises. A good example of this from a recently published Chapman & Hall book is here: http://mdsr-book.github.io. This provides a venue for the exercises to be updated if mistakes are found and provides tangible examples that can be reproduced in code provided online. Without this, instructors who will use this book will have to read the code in the book, and translate it by hand based to R.
- I would recommend the above examples be written in a reproducible document such as R Markdown or R Notebooks.
- I would recommend the author include a section on simulating genetic data. Methods developers and instructors who will use this in the classroom may want students to be able to generate their own genetic data to test out proposed methods or apply different genetic algorithms in R packages.

Review 2, proposal

1. Is this an important topic for research and what types of researchers would be interested? Would you consider using it as a text and if so, which text is used now?

In my opinion this book proposal is done in one of the hottest research topics. Today most human health researchers, human geneticists, plant and animal biologists are interested in statistical genetics and its applications.

"Would you consider using it as a text"
I certainly consider all books in the area of statistical genetics analysis as possible texts as support for my grad students as well as potential books for my class.
However, I stay a little away from all books that contain the words "with R" in the title. There has been recently many books that basically write code and publish their code in a book. However, from marketing proposes I think that more people will buy a book that specifically says that it shows how to do certain analysis with R than if that is not contemplated in the title. Also, when I consider books for my class I put 33% weight in the book to cover all important concepts and modern genomic analysis, 33% weight in the book to be clearly written, and 33% on books that would have hands on exercises with solutions, many many examples done, and if possible data downloadable to also have R exercises.
That saves time for homework preparations and assignation and gives the students material to explore more and practice more before the midterm/final exam.

"which text is used now?"
The text that I use the most for statistical genetics is 'quantitative genetics' by Falconer and MacKay. I complement my class with a lot of other materials, including the book by Andrea Foulkes. I also use "principles of population genetics" because it has many exercises and several with solutions, and for the initial classes of the course I use "essentials of Genetics" which is very handy because not only comes with a separated book with all the solutions to exercises, but it also comes with a DVD with a class per chapter including all the figures in the book. Planning classes and HW when using the 'essentials of genetics' is very very easy, however it's too basic and do not cover statistical genetics. Finally, I use the book by Lynch and Walsh entitled 'Genetics and Analysis of Quantitative Traits' for examples to develop in class.
The book that I say the students to buy is the one by Falconer and MacKay, and I have a couple of copies of all the other ones donated to the library of my department for the students to check out and read specific topics (complementing all the topics not covered in Falconer and MacKay book).
I would love to have comprehensive book covering concepts and modern statistical tools for estimation and analysis, with tons of examples, and homework. If that book become available, I would switch my students to buy that one.

2. What other books are available in this area? Do they have any particularly strong or weak features?

There are two type of books that I can imagine: One for our grad students and for us, which could be more technical, and explain how different R packages implement certain type of

analysis and can tackle new data from new technologies (I mean types of files that can become available for example for NGS, Affy files, etc. The second type of book that I ambition a lot to find a helpful and comprehensive one is for a initial level statistical genetics class. This second type of books would have a larger public but needs to be also more comprehensive.

'Handbook of Statistical Genetics'
although the topic coverage is great on this handbook it has drawbacks:
1-the series is too expensive for a course
2-each chapter is written by different authors, thus there is a lot of repetition, some chapters focus in historical notes, others are quite technical, others the level is for beginners, too heterogeneous both content and level.
3- do not provide exercises, not good for a course.

'the fundamentals of modern statistical genetics' by nan laird and Christoph Lange.
1-This book has a great coverage of topics and good organization, but I decided not to use it at all because the writing is mixed up and does not go deep into the fundamentals and conceptual topics.
2-Great price.
3-Has several exercises on each chapter, but without solutions

'statistical methods in genetic epidemiology' by D. C. Thomas.
-It has a focus in genetic epidemiology, which is a great trait.

'population genetics' by guillespie
it's an excellent book, but for students it's a little too concise, although fantastic to review topics and design a class.
-has many interesting and handy examples.

'handbook on analyzing human genetic data'
-written by several authors, I usually try to stay away from this types of books, level and focus varies chapter to chapter.
-no exercises available.

'primer to analysis of genomic data using R' by Gondro
I just acquired this book, I am not familiar with the content yet to comment.

'principles of population genetics' by Hartl and Clark
-Excellent book, well written clear, easy, with many examples and exercises with solutions.
-drawback is that does not have any R or other software example.
-does not cover the most modern topics either.

"Can you describe a need for a book in this area that differs in its topics, approach, or level?"
initial level book, very conceptual (not all organized around potential analysis and estimation), with extensions to 'how to' to analysis, R coding examples, available datasets, by-hand exercises (not all programming examples), with solutions.

3. Are there any changes that you would recommend in the topics included or the organization of the contents that would make this book more useful? Are the proposed prerequisites appropriate for the topics and audience?

I think that the organization of the chapters could be better, I find it a little mixed up, I would suggest to go to other tables of contents on other books for a better organization.
I would also suggest not to focus/organize the book on analysis.

4. What are your main recommendations to the author(s) before writing this book?
Read many other books, I don't agree with the document shared with me that the problem with the other books are always that the other books are old (7 yr or more). The fundamental principles of genetics do not change. The type of data available change. Thus, a great book would be one that keep a strong conceptual section in each chapter, and has a second section of 'how to do analysis and inference of… e.g. LD'

5. Please explain why you do or do not believe that this combination of author and topic is likely to produce a book that we will want to publish.
-I don't know the author, I cannot comment on her ability to write a book useful for the community.
-the topic is a hot topic, and everybody wants to know how to analyze it with R and the new tools. However, many people do not realize that their problem is not that they don't know how to do certain analysis with R, but they first need to know genetics/statistics, second learn how to do it and how to interpret the results and evaluate their quality.