Books

The goal of this analysis is to explore book genre popularity. Some questions I'll try to answer are: Have genre tastes changed over time? Does genre popularity have a pattern in the course of a year? Do demographics such as age or gender tend to report reading the same types of books?

The Data

This data was collected via a public API from a popular book social network. All information is self reported, including demographic information, books read, and genre categorization. Unfortunately it is not necessary when reporting having read a book to categorize it by genre but fortunately many users do.

Two types of data were collected: User data, and information about books users self-reported to be "currently reading"

A note about github: The collected CSV files are too large to share on github. However, you can see exactly how I collected the data at my repo called 'goodreads-analyses' I'm also happy to share the data I collected if you would like to reproduce my analyses or work with the data yourself. A word of caution: I collected this data too quickly for their tastes and ended up getting IP banned for about a month. Consume at your own risk.

User Data

```
user.cols <- c("userID", "name", "gender", "age", "location", "lastactive",
               "readcount", "currentcount", "wantscount")
users <- read.csv("data/users.csv", header=F, col.names=user.cols, stringsAsFactors=F)
head(users)
```

```
##   userID          name gender age            location lastactive readcount
## 1      1 Otis Chandler   male  36 San Francisco, CA    02/2014       361
## 2      2 odawg Diggity   male  36 San Francisco, CA    11/2013        75
## 3      3        Adrian   male  38 San Francisco, CA    12/2013        52
## 4      4       Isadora female  NA San Francisco, CA    06/2013         4
## 5      5     Elizabeth female  NA  Santa Monica, CA    01/2014       937
## 6      6         kelly female  40       Oakland, CA    01/2014        28
##   currentcount wantscount
## 1            3        360
## 2            7         19
## 3            0          0
## 4            0          0
## 5            7        568
## 6            1         12
```

```
dim(users)
```

```
## [1] 574919      9
```

Book Data

```
book.cols <- c("userID", "bookID", "shelf", "dateadded", "datestarted", "datefinished", "title",
               "isbn", "isbn13", "imageurl", "pages", "publisher", "publicationdate", "genres")
books <- read.csv("data/currently-reading.csv", header=F, col.names=book.cols,
                  stringsAsFactors=F, na.strings="")
head(books)
```

```
##   userID   bookID            shelf                    dateadded
## 1      1   123715 currently-reading Fri Nov 08 15:25:13 -0800 2013
## 2      1 12983631 currently-reading Wed Feb 22 11:08:15 -0800 2012
## 3      1  8677247 currently-reading Mon Jan 16 11:31:52 -0800 2012
## 4      2  6944712 currently-reading Fri Jul 30 18:27:39 -0700 2010
## 5      2    37553 currently-reading Thu Jul 29 22:38:19 -0700 2010
## 6      2    22463 currently-reading Wed Jul 14 22:56:36 -0700 2010
##                      datestarted datefinished
## 1 Fri Dec 27 16:10:42 -0800 2013         <NA>
## 2 Wed Feb 22 00:00:00 -0800 2012         <NA>
## 3 Mon May 28 17:32:49 -0700 2012         <NA>
## 4 Fri Jul 30 18:27:43 -0700 2010         <NA>
## 5 Thu Jul 29 22:38:19 -0700 2010         <NA>
## 6 Wed Jul 14 22:56:36 -0700 2010         <NA>
##                                                          title
## 1 Slack: Getting Past Burnout, Busywork, and the Myth of Total Efficiency
## 2                                    Book: A Futurist's Manifesto
## 3                                             Merchants of Culture
## 4                    Under the Amoral Bridge (The Bridge Chronicles, #1)
## 5                                            Sense and Sensibility
## 6                                            The Origin of Species
##         isbn        isbn13
## 1 0767907698 9780767907699
## 2       <NA>          <NA>
## 3 0745647863 9780745647869
## 4 1449509673 9781449509675
## 5 0192804782 9780192804785
## 6 0785819118 9780785819110
##                                                          imageurl
## 1   https://d202m5krfqbpi5.cloudfront.net/books/1320419657m/123715.jpg
## 2 https://d202m5krfqbpi5.cloudfront.net/books/1327887891m/12983631.jpg
## 3  https://d202m5krfqbpi5.cloudfront.net/books/1338542081m/8677247.jpg
## 4               https://www.goodreads.com/assets/nocover/111x148.png
## 5               https://www.goodreads.com/assets/nocover/111x148.png
## 6               https://www.goodreads.com/assets/nocover/111x148.png
##   pages        publisher publicationdate
## 1   256 Crown Business               2002
## 2   318      Pressbooks                NA
## 3   432            <NA>                NA
## 4   170     Createspace               2008
## 5   300            <NA>                NA
## 6    NA            <NA>                NA
##
## 1      ["business", "management", "non-fiction", "work", "agile", "nonfiction", "software-developmen
## 2                   ["nonfiction", "non-fiction", "publishing", "society", "technology", "kobo"
## 3 ["non-fiction", "nonfiction", "publishing", "books-about-books", "business", "writing", "university
## 4                   ["cyberpunk", "sci-fi", "kindle", "science-fiction", "my-ebooks", "scifi", "fict
```

```
## 5          ["classics", "fiction", "favorites", "romance", "classic", "jane-austen", "literature", "his
## 6              ["science", "non-fiction", "classics", "nonfiction", "biology", "evolution", "my-ebool
```

```r
dim(books)
```

```
## [1] 130396      14
```