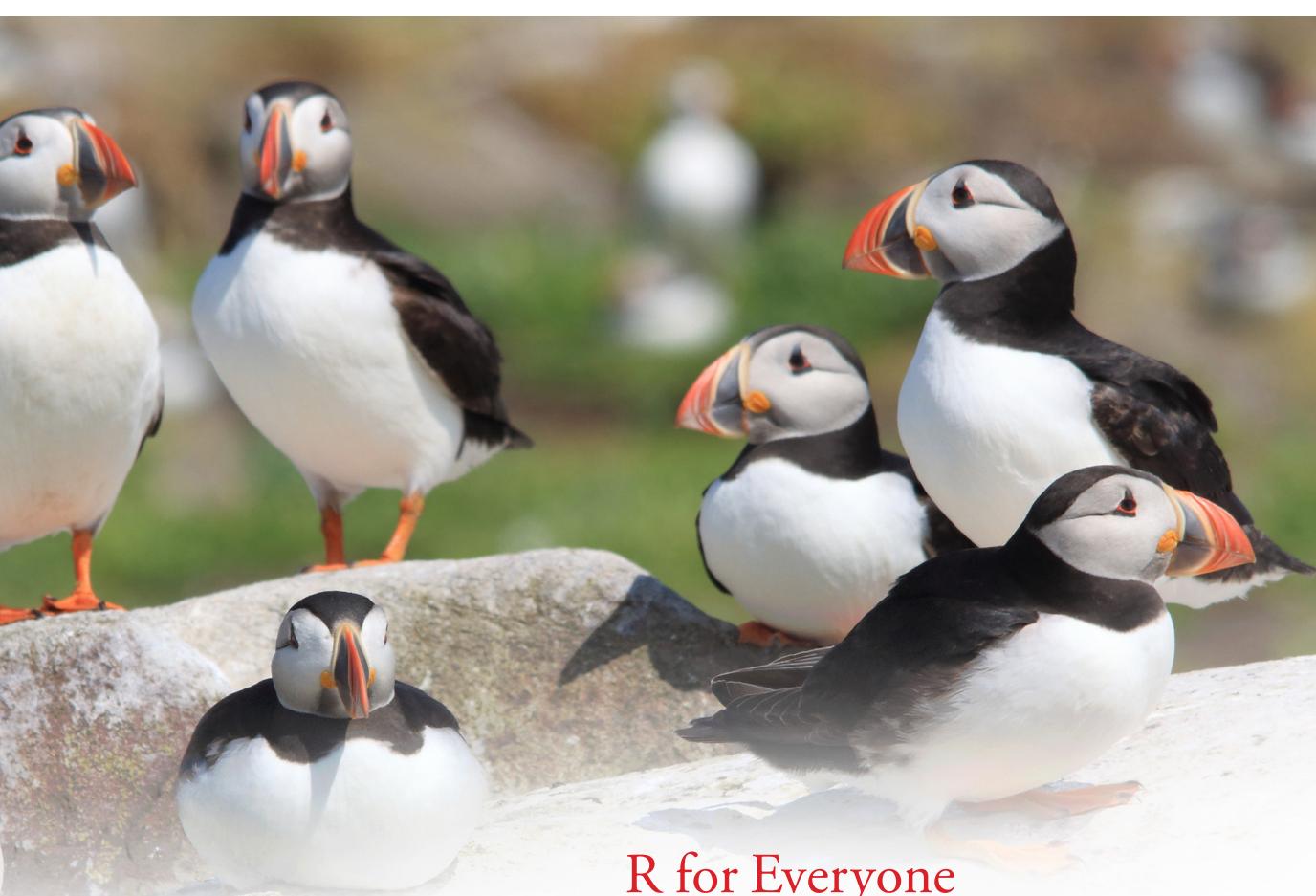


みんなのR

データ分析と統計解析の新しい教科書

Jared P. Lander [著] 高柳慎一、牧山幸史、簗田高志 [訳] Tokyo.R [協力]



R for Everyone
Advanced Analytics and Graphics

プロのデータサイエンティストから学ぼう!
"R"の基礎・モダンなデータ分析・統計解析と可視化。



みんなのR

データ分析と統計解析の新しい教科書

Jared P. Lander [著] 高柳慎一、牧山幸史、簗田高志 [訳] Tokyo.R [協力]

R for Everyone
Advanced Analytics and Graphics

マイナビ

Authorized translation from the English language edition, entitled R FOR EVERYONE: ADVANCED ANALYTICS AND GRAPHICS, 1st Edition, ISBN: 0321888030 by LANDER, JARED P., published by Pearson Education, Inc. publishing as Addison-Wesley Professional, Copyright © 2014.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

JAPANESE language edition published by MYNAVI CORPORATION, Copyright © 2015.

JAPANESE translation rights arranged with PEARSON EDUCATION, INC. through OWLS AGENCY INC., TOKYO JAPAN

- ・本書中に登場する会社名や商品名は一般に各社の商標または登録商標です。
- ・原書の公式サイトは以下になります（英語）。本書で取り上げたデータなどは以下のサイトから入手可能です（サイトの管理・運営はすべて米国で行っています）。

<http://www.jaredlander.com/r-for-everyone/>

- ・本書の正誤等に関するサポートサイトは以下になります。

<https://book.mynavi.jp/supportsite/detail/9784839955212.html>

翻訳者より

本書は2013年末に米国で発刊され、すでに4版まで増刷されたオープンソースの統計解析ソフトRの解説書です。ニューヨーク市でデータサイエンティストとして活躍しつつ、コロンビア大学でも教鞭を執っているJared P. Landerの執筆によるもので、Rの初学者向けに幅広いトピックを取り扱っています。

近年のビッグデータ・データサイエンティストブームの影響で、Rに関する書籍が数多く発刊されており、書店にも他の統計解析ソフトを凌駕する量の書籍が並んでいます。その中でも本書はとてもバランスのとれた内容で構成されています。著者が大学で教鞭を執っていることもあり、初めてデータサイエンスを学ぶ初学者がつまずきそうなポイントを丁寧に説明しています。他の書籍ではページ数の関係で省かれているようなコードや、可視化されたチャートなどを惜しげも無く多数記載しています。

データサイエンスに興味があるけれども、どのように学習をしたらよいか、何から始めたらよいか迷っている初学者に向けた適切な書籍となっているのではないでしょうか。

本書では、理論的な説明・数式だけでなくRのコードを多く引用しており、それぞれのトピックスについて実際にコードを動かしながら学ぶことができます。多くのコードを見て動かすことにより、各内容とRの挙動について深く学ぶことができると考えています。

トピックスとしては、Rのインストール、RStudioの使い方、Rの文法、データハンドリング、グラフィックス、確率分布、基礎統計、回帰、重回帰、一般化線形モデル、モデル評価、オーバーフィッティングを防ぐ方法、クラスタリング、時系列、レポートティングなどを網羅しており、この本を手元に1冊持っておけば、データサイエンスに関する課題・業務の課題解決の一助となるはずです。

網羅的なトピックスを扱っているので、データサイエンスに興味がある方、実際の業務でデータサイエンスを用いビジネスにインパクトを与えたい方、データサイエンティストを抱えている部署のマネジメントの方など、様々な人に自分の読み方で活用してもらえればと思います。

翻訳者を代表して 篠田高志

目次

翻訳者より	3
0.1 : はじめに	9
0.2 : まえがき	10
0.3 : 謝辞	13
1章 Rを手に入れる	17
1.1 : Rのダウンロード	17
1.2 : Rのバージョン	18
1.3 : 32bit と 64bit	18
1.4 : インストール	19
1.5 : Revolution R Community Edition	29
1.6 : まとめ	29
2章 Rの環境	31
2.1 : コマンドラインインターフェイス (CLI)	32
2.2 : RStudio	33
2.3 : Revolution Analytics RPE (R Productivity Environment)	45
2.4 : まとめ	46
3章 Rパッケージ	47
3.1 : インストールパッケージ	48
3.2 : パッケージのロード	51
3.3 : パッケージの作成	52
3.4 : まとめ	52
4章 Rの基本	55
4.1 : 基本的な数学	55
4.2 : 変数	56
4.3 : データ型	58
4.4 : ベクトル	63
4.5 : 関数 (Function) の呼び出し	69
4.6 : 関数ドキュメント	69
4.7 : 欠損値	70
4.8 : まとめ	72
5章 高度なデータ構造	73
5.1 : データフレーム (data.frame)	73
5.2 : リスト (List)	80
5.3 : マトリックス (Matrix)	87
5.4 : アレイ (Array)	90
5.5 : まとめ	91
6章 Rへのデータ取り込み	93
6.1 : CSVの読み込み	93
6.2 : Excel の読み込み	95
6.3 : データベースからの読み込み	95

6.4: 他社統計ツールからの読み込み	97
6.5: Rバイナリファイル	98
6.6: Rに入っているデータ	100
6.7: Web サイトからの抽出	101
6.8: まとめ	102
7章 統計グラフ	103
7.1: 基本グラフィクス	103
7.2: ggplot2	107
7.3: まとめ	121
8章 Rの関数を書く	123
8.1: ハロー、ワールド！	123
8.2: 関数の引数	124
8.3: 値の返却	127
8.4: do.call	128
8.5: まとめ	129
9章 コントロール文	131
9.1: if と else	131
9.2: switch	134
9.3: ifelse	135
9.4: 複合テスト	137
9.5: まとめ	138
10章 ループ・Rの方法ではない反復方法	139
10.1: for ループ	139
10.2: while ループ	141
10.3: ループの制御	142
10.4: まとめ	143
11章 グルーピング操作	145
11.1: Apply ファミリー	145
11.2: aggregate	149
11.3: plyr	152
11.4: data.table	158
11.5: まとめ	168
12章 データ整形	169
12.1: cbind と rbind	169
12.2: Join	170
12.3: reshape2	178
12.4: まとめ	182
13章 文字列操作	183
13.1: paste	183
13.2: sprintf	184
13.3: テキストの抽出	185
13.4: 正規表現	190
13.5: まとめ	198

14章 確率分布	199
14.1: 正規分布	199
14.2: 二項分布	204
14.3: ポアソン分布	209
14.4: その他の分布	212
14.5: まとめ	214
15章 基本統計	215
15.1: 要約統計	215
15.2: 相関と共分散	219
15.3: t検定	228
15.4: 分散分析	236
15.5: まとめ	238
16章 線形モデル	239
16.1: 単回帰	239
16.2: 重回帰	244
16.3: まとめ	259
17章 一般化線形モデル	261
17.1: ロジスティック回帰	261
17.2: ポアソン回帰	265
17.3: その他の一般化線形モデル	269
17.4: 生存時間分析	269
17.5: まとめ	274
18章 モデル評価	275
18.1: 残差	275
18.2: モデル比較	281
18.3: クロスバリデーション	285
18.4: ブートストラップ	290
18.5: ステップワイズ変数選択法	293
18.6: まとめ	297
19章 正則化と縮小(シュリンケージ)	299
19.1: Elastic Net	299
19.2: Bayesian 縮小	320
19.3: まとめ	324
20章 非線形モデル	325
20.1: 非線形最小二乗法	325
20.2: スプライン	328
20.3: 一般化加法モデル	332
20.4: 決定木	339
20.5: ランダムフォレスト	341
20.6: まとめ	343
21章 時系列と自己相関	345
21.1: 自己回帰移動平均	345
21.2: VAR	352

21.3 : GARCH	357
21.4 : まとめ	366
22章 クラスタリング	367
22.1 : K-means	367
22.2 : PAM	375
22.3 : 階層型クラスタリング	382
22.4 : まとめ	388
23章 knitrパッケージによる再現性・レポートとスライドショー	389
23.1 : LaTeX プログラムのインストール	389
23.2 : LaTeX 入門	390
23.3 : LaTeX を使った knitr	392
23.4 : マークダウンのヒント	398
23.5 : knitr とマークダウンの利用	399
23.6 : pandoc	399
23.7 : まとめ	402
24章 Rパッケージの構築	403
24.1 : フォルダ構成	404
24.2 : パッケージファイル	404
24.3 : パッケージドキュメンテーション	411
24.4 : チェック、構築とインストール	414
24.5 : CRANへの登録	416
24.6 : C++ コード	416
24.7 : まとめ	422
付録A 実生活での情報リソース	423
A.1 : Meetup	423
A.2 : Stack Overflow	424
A.3 : Twitter	425
A.4 : カンファレンス	425
A.5 : Web サイト	426
A.6 : ドキュメント	426
A.7 : 書籍	427
A.8 : まとめ	427
付録B 用語集	428
索引	441

母と父へ

0.1: はじめに

Rはこの3年間で大きく人気を伸ばしています。そう聞くと、読者はRを比較的新しく、新進気鋭の言語だと考えるでしょう。しかし驚いたことに、Rには1993年からの歴史があります。なぜこんなに急激に人気が出たのでしょうか？ その答えは、仕事と研究分野におけるデータサイエンスの高まりによるものようです。しかし、データサイエンスの土台は何十年も利用されてきました。統計、線形代数、オペレーションズ・リサーチ、人工知能、機械学習は最近のデータサイエンティストが利用するツールに貢献しています。Rは、他の言語よりも、1つの機能でこれらを呼び出すことができます。

これらの理由から、私はAddison-Wesley Data and Analytics シリーズの最初の本として本書を出版するのに興奮しています。Rはたくさんのデータサイエンスのタスクをこなすのに不可欠なツールです。予測や解析に利用できる多くのアルゴリズムを、少ないコードで利用でき、それらは最近のデータ解析の問題を解くのにとても合っています。データサイエンスは数学や統計学分野ではなく、プログラミングやインフラストラクチャの分野でもありません。この本は、Rの力や表現力の紹介を一般のみなさまにとてもバランスよく提供しています。

私には、Jared Landerさん以上にRの紹介をうまく出来る人は考えられません。Jaredと私は2009年後半のニューヨークの機械学習コミュニティで初めて会いました。それまで、ニューヨークのデータコミュニティはひとつの部屋で収まるほどとても小さく、他のデータ関連のカンファレンスはまだありませんでした。Jaredは4年以上、データサイエンスのプロフェッショナルとして最先端で活躍されています。

開かれた統計プログラミングのカンファレンスの運営、イベントでの講演、コロンビア大学でのRの授業を通して、Jaredはプログラマの教育、データサイエンティスト、ジャーナリスト、統計家などのコミュニティの成長を助けてきました。しかし、Jaredの専門知識は教育だけではありません。彼は実務家であり、これらの道具を使って、大企業から中小企業のコンサルティングを行っています。

この本は、常日頃Rを使われているユーザーに、Rのプログラミングと様々な統計手法を提供します。サンプルデータセットは、Jaredが準備してくれたWebサイトからアクセスでき、誰でも使えます。実際のデータや興味のある問題を解く際には、この本は最後まで役に立つでしょう。

—— Paul Dix, Addison-Wesley Data and Analytics シリーズの編集者

0.2: まえがき

我々の生活に関わるデータが増加傾向にあり、このデータの洪水に対して分析できる、新しく良い道具が求められています。これまで伝統的に2つの方法：ExcelやSPSSなどのツールを使って個人が分析する方法と、とても大変だがC++などを使って高性能に分析する方法がありました。

パソコンコンピュータの性能の向上が、対話型で使用でき、かつ、堅牢性もある分析領域を育みました。個人の（彼もしくは彼女）のコンピュータで行われる探索的データ解析の方法は、サーバ向けや先進的なビジネス・プロセスへと素早く転換されています。この分野はR、Python、その他のスクリプト言語の領域です。

Rは、ベル研究所のJohn Chambersが発明したS言語をオークランド大学のRobert GentlemanとRoss Ihakaが1993年に改良し、発明したものです。R言語は高級言語で、最初からユーザーがコマンドを入力、結果を得て、他のコマンドを実行できるように、インタラクティブ性を目的としていました。その後、システムに埋め込まれることもできて、複雑な問題にも取り組む言語に発展しました。

データの変換、分析に加えてRは驚くべきグラフィクスとレポーティングが生み出せるようになりました。いまでは、データの抽出や変換、モデルの適応、推論と予測、結果の作図とレポーティングなど、データ解析におけるフルスタックに使われるようになりました。

Rの人気は2000年代後半に急上昇し、学術界だけではなく、銀行、マーケティング、医薬、政治、遺伝子やその他の業界に広がっていきました。新しいユーザーはC++のようなコンパイルが必要な言語、SASやSPSSなどの他の統計ソフト、800ポンドのゴリラのようなやたらと重たいExcelから移ってきました。この期間、多くのアドオンパッケージ（Rの機能を拡張する予めコードが書かれたライブラリ）が公開されました。

Rは時々、特にプログラミングの経験が無い初心者に威圧感を与えることがあります。プログラミング分析は、ポイントアンドクリックの代わりに、より簡単に、より便利に、より信頼できるようになることを私は発見しました。学習プロセスを簡単に素早くすることが私のゴールです。

この本は、もし私が大学院でRを学習するとなったら、という前提で、私が希望する情報で配置されています。その出発点に戻るべく、本書の内容は、私がコロンビア大学で教えるデータサイエンスコースと同時に開発されました。Rの細かい部分はカバーできていませんが、むしろ20%の機能が仕事の80%を達成するのに必要です。この本の内容は以下の章で構成されています。

第1章 Rを手に入れる：どこでRをダウンロードするのか、インストールの方法について

て。様々なOSと32ビット、64ビットバージョンの違いを扱います。また合わせてどこにRをインストールしたらよいかアドバイスしています。

第2章 Rの環境: Rの概要、特にRStudioの概要を説明しています。RStudioプロジェクトとGitの統合はカスタマイズとRStudioのナビゲーションとしてカバーされています。

第3章 Rパッケージ: Rパッケージの見つけ方、インストールとロードについて説明しています。

第4章 Rの基本: Rを数学で使います。`numeric`, `character`, `date`を`vector`として詳細を説明します。関数の呼び出しやドキュメントの探し方の簡単な紹介をしています。

第5章 高度なデータ構造: 強力で、よく使われるデータ構造の`data.frame`を`matrices`や`list`と共に説明しています。

第6章 Rへのデータ取り込み: 分析をする前にRにデータを取り込む必要があります。CSVデータとデータベースから取り込む方法を含む様々な方法があります。

第7章 統計的なグラフィックス: グラフの作成は予備的分析と結果を説明するのに決定的な部分です。Rは強力な作図ユーティリティを使って美しいグラフを作成することができます。ベーシックなプロット関数とggplot2パッケージを使って詳細を説明しています。

第8章 関数を書く: 繰り返しの分析はユーザー定義関数でしばしば簡単になります。構造、引数、返り値について議論しています。

第9章 コントロール文: `if`, `ifelse`と複雑な構文チェックを使ってプログラミングコードをコントロールします。

第10章 ループ、Rの方法ではない反復方法: `for`と`while`を使った繰り返し。Rでは一般的には推奨されませんが、知っておくべき重要な点です。

第11章 グルーピング操作: 一度にすべての要素を扱うかなりの反復処理を行う、ループやベクトルではない代替方法。ここではより効果的な、`apply`関数と`plyr`パッケージを説明します。

第12章 データ合成: 積み上げや結合などの複数データの合成は一般的に必要です。`plyr`パッケージや`reshape2`パッケージは良い関数で、基本関数の`rbind`、`cbind`や`merge`と共に利用することで、データ加工を行うことができます。

第13章 文字列操作: ほとんどの人は統計で文字列データを関連付けて考えられませんが、データの中では重要です。Rは文字列を扱うため、多くの機能を提供しており、文字列の連結や抽出などを含んでいます。正規表現についても説明しています。

第14章 確率分布: 正規分布、二項分布、ポアソン分布の概要を説明しています。多くの分布の式や関数も記載しています。

第15章 基本統計: 平均、標準偏差やt検定など、ほとんどの人が最初に教えられる統計の説明をします。

第16章 線形モデル: 統計において、非常に強力かつ一般的なツールで、広範囲に説明

しています。

第17章 一般化線形モデル：ロジスティック回帰とポアソン回帰を含んだ拡張された線形モデル。生存時間分析もカバー。

第18章 モデル評価：モデルの質の決定と残差、AIC、クロスバリデーション、ブートストラップ、ステップワイズ変数選択法を使った変数選択の説明します

第19章 正則化と縮小：Elastic Netとベイジアン法を使ってモデルのオーバーフィットを防ぐ方法を説明します

第20章 非線形化モデル：線形モデルが適さない場合に、非線形モデルはよいソリューションです。非線形最小モデル、スプライン、一般化加法モデル、決定木、ランダムフォレストを説明しています。

第21章 時系列と自己相関：時系列データに対する一変量、多変量分析の手法。

第22章 クラスタリング：データをグルーピングするクラスタリングは、K-meansや階層化型クラスタリングなどの手法で実現できます。

第23章 knitrパッケージによる再現性、レポートとスライドショー：レポートの生成、RによるスライドショーとWebページはknitrパッケージとLatexとMarkdownで簡単に作成できます。

第24章 Rパッケージの作成：Rパッケージはポータブルで再利用性のあるコードです。devtoolsパッケージとRcppパッケージの出現で、パッケージの構築は驚くほど簡単にできるようになりました。

Appendix A 実生活でのRの情報源：Rをより学ぶための情報源と、コミュニティとの対話ができるリスト。

Appendix B 用語集：この本で利用されている言葉の用語集。

この本の中のテキストのより良いところは、Rコードだけでなく、コードの実行結果もついているところです。コードと結果はテキストの離れたブロックにしばしば配置され、以下の例にあるように、独特のフォントで表現しています。異なるコードのパートは異なる色で表示しています。コードは>で始まり、1行から次の行に移る際には、+で表現されています。

```
> # これはコメントです。
```

```
>
```

```
> # 基本計算
```

```
> 10 * 10
```

```
[1] 100
```

```
>
```

```
> # 関数の呼び出し
```

```
> sqrt(4)
```

```
[1] 2
```

インラインで表示される場合のコードは `sqrt(4)` のように表示されます。

いくつかの箇所で数式が必要で、方程式は余白とナンバーが振られています。

$$e^{i\pi} + 1 = 0 \quad (1)$$

方程式中で、通常の変数はイタリック (`x`) で表示され、ベクトルは太字の小文字 (`x`) で、マトリックスは太字の大文字 (`X`) で表現しています。 α や β のギリシャ文字は、同じ規則に従っています。

関数名は `join` のように書かれ、パッケージ名は `plyr` のように書かれています。コードで生成されたオブジェクトはテキストで参照される場合は、`object1` と書かれます。

Rを学習する経験は、たくさんのタスクをこなす生活を簡単にしてくれます。
私と一緒に学ぶことを楽しんでください。

0.3: 謝辞

始める前に、数学を専攻するように勇気づけてくれた、私の母の Gali Lander に感謝します。この後押しがなければ、統計とデータサイエンスの道を歩んではいなかったかもしれません。同じように、授業料の請求書のすべてを払ってくれた父、Howard Landar にも感謝します。彼は私の人生を通してたくさんのアドバイスと手引をしてくれました。同時に両親は私が何をしているのかを理解するのに固執しなかったし、私が行うこと好きになってくれ、いつも助けてくれました。家族と共にいて、私の妹と義理の弟の Aimee と Eric Schechtermen に、彼らの5歳の子供の Noah に数学を教えてくれるチャンスをくれたことに感謝します。

数年に渡って私を鍛えてくれた多くの先生がいらっしゃいました。最初は Rochelle Leck で、私の先生が私に、「君は数学の能力はもっていない」と言った時の、中学校の数学の家庭教師でした。

次に Beth Edmondson で Princeton Day School の precalc (数学) の先生です。高校の最初の半年を平凡な生徒として無駄に過ごしたあと、彼女は「来年の AP Calc を受験する気はあるのか?」と聞いていました。彼女のクラスで C から A+ に成績が上がれば AP Calc を受験してもよいと合意してくれましたが、私は決してチャンスとは考えませんでした。しかし、3ヶ月後、A+ をとるどころか、すべての成績が向上しました。彼女は私の人生を

変え、もし彼女がいなかったら、いまの私はいなかつたでしょう。彼女が私の先生だったことをずっと感謝しています。

Muhlenberg大学の最初の2年間は、私はビジネスとコミュニケーションを専攻することに決められていたのですが、私にとって自然に思えたため、数学のクラスを取りました。Penny Dunham教授, Bill Dunham教授,Linda McGuire教授は数学を専攻することに納得してくれましたし、この決定が私の人生を決定づけたのです。Greg Cicconetti教授は最初に厳格な統計を垣間見せてくれ、最初のリサーチの機会を与えてくれ、卒業後に大学院へ進むアイディアを考えさせてくれました。

M.A (Master of Arts) をコロンビア大学で取得する間、統計とプログラミングの素晴らしい考え方で囲まれていました。Dabid Madigan教授は最新の機械学習に対して開眼させてくれ、Bodhi Sen教授は統計プログラミングについて考えさせてくれました。私はAndrew Gelman教授と一緒に研究する特権を持ち、彼の考えは計り知れないほど私にとって影響がありました。Richard Garfield教授は、最初にアサインされたミャンマーで災害や戦闘地域で人々を助けるために統計をどう使うかを見せてくれました。長年に渡る彼のアドバイスと友情は私にとって大切なものです。

Jingchen Liu教授は、ニューヨークのpizzaについての論文を書くことを後押ししてくれ、それは大変な注目を集めました¹。

コロンビア大学にいる間、よい友だちに会えました。一度TAとして教えてくれた、Ivor Cribben教授は私の知識の足りない部分を埋めてくれました。彼を通じて、名誉なことにコロンビア大学で共に教えている、とてもよいアドバイスをくれるRachel Schutt教授に会えました。

大学院は、Shanna Leeの勇気づけとサポートがなかったら卒業できなかつたでしょう。彼女は私が2つのこと、授業とコロンビア大学のホッケーチームにとても入れ込んでしまっている間に正気を保つてられるように助けてくれました。彼女無しでは、この2つはできなかつたどうかわかりません。

Steve CzettyはSky IT Groupで分析の仕事を与えてくれ、データベースを教えてくれる一方で、突拍子も無いプログラミングの経験を与えてくれました。このことがデータと統計について興味を引き起しました。Bardess GroupのJoe DeSiena、Philip du PlessisとEd Bobrinは私がプレッシャーの中働いている中でも最高の人々で、今日においても最高の人と仕事ができたと誇りに思っています。Revolution Analytics社のMike Minelly、Rich Kittler、MarkBarry、David Smith、Joseph Rickert、Norman Nie教授、James Peruvankal、Neera TalbertとDave Richは私が考えうる最高の仕事をしてくれました。ビジネスを生業としている人に対してなぜRを使う必要があるのかを説明してくれまし

¹ <http://slice.serioouseats.com/archives/2010/03/the-moneyball-of-pizzastatistician-uses-statistics-to-find-nyc-best-pizza.html>

た。Big Computing の Kirk Mettler、Richard Shultz、Bryan Lewis 教授と Jim Winfield は、興味ある問題に対して、R で解決するのを楽しむように勇気づけてくれました。ゴールドマン・サックスの Vincent Saulys、John Weir と Saar Golde 教授は楽しみと教育の両方に時間を作ってくれました。

この本を書くことを通して、多くの人が助けてくれました。一番最初は Yin Cheung で、私が常にストレスを感じているのを見ていて、昼も夜も費やしてサポートしてくれました。

私の編集者、Debra Williams はどうやれば私を勇気づけられるかを知っており、彼女の導きは計り知れません。Paul Dix はこのシリーズのエディタかつ良き友達で、彼がこういった本を書いたらどうかを提案してくれ、彼がいなからたらこの本はできなかつたでしょう。偉大なコピーエディターの Caroline Senay と Andrea Fox には感謝を申し上げたい。彼ら無しではこのような仕上がりにはならなかつたでしょう。Robert Mauriello のテクニカルレビューはこの本のプレゼンテーションに磨きをかけてくれました。

RStudio の人々、特に JJ Allaire と Josh Paulsoon は他の製品よりもとても簡単に本を書き進められる製品を作ってくれました。knitr パッケージの Yihui Xie はこの本を書くために数多くの機能の変更を提供してくれました。彼のソフトウェアと私のリクエストを実装するスピードに大変感謝しています。

Chris Bethel、Dirk Eddelbuettel 教授、Ramnath Vaidyanathan 教授、Eran Bellin 教授、Avi Fisher、Brian Ezra、Paul Puglia、Nicholas Galasinao、Aaron Shumaker、Adam Hogan、Jeffrey Arnold と John Houston など、この本を作るのに多くの人が価値のあるフィードバックをしてくれました。

昨秋の私の講義の、コロンビア大学の 2012 年秋のデータサイエンス導入クラスに参加してくれた生徒が、この本の最終的な材料のモルモットになってくれたことに感謝します。

この道を助けてくれたすべての人々に感謝を申し上げます。

Rは統計解析、可視化、レポーティングのための素晴らしいツールです。その有用さを知るには、Rが広く様々な分野で使われていることを見るのが一番です。我々が関わっただけでも、銀行、政治運動、技術系スタートアップ、食品系スタートアップ、国際開発援助機関、病院、不動産開発者たちのプロジェクトにおいて、Rを使用しました。我々が見てきた他の分野として、インターネット広告、保険、生態学、遺伝学、医学においてもRが使用されていました。Rを使用するのは、高度な機械学習のトレーニングを受けた統計家、他のプログラミング言語に精通したプログラマー、そして、高度なデータ分析について訓練されているとは言えないが、Excelを使うことに疲弊した人たちです。

Rを使用するには、ダウンロードしてインストールする必要があります。その作業は、他のプログラムをインストールするのに比べて、難しいということはありません。

1.1: Rのダウンロード

Rを使用するための最初のステップは、コンピューター上に入手することです。C++のような言語とは異なり、Rを実行するためには、コンピューターにインストールする必要があります¹。プログラムはRの開発元であるComprehensive R Archive Network (CRAN) <http://cran.r-project.org/> から簡単に入手することができます。ページトップにWindows向け、Mac OS X/OS X向け、Linux向けのダウンロードリンクがあります。

Windowsと(Mac) OS Xにはあらかじめビルトされたインストーラーがあります。一方、Linuxではソースからコンパイルするのが一般的です。いずれのプラットフォームでも、Rのインストール方法は、他のプログラムのインストール方法と全く同じです。

Windowsユーザーは、Download R for Windows、base、Download R 3.x.x for Windowsというリンクを順次クリックしてください。xにはRのバージョン番号が入ります。この番号は、バージョンアップによって定期的に変わります。

同様に、MacユーザーはDownload R for (Mac) OS X、R-3.x.x.pkgを順次クリックします。再び、xにはRのバージョン番号が入ります。このインストーラーは、32ビット

¹ 技術的には、C++はコンパイラ無しで単独にセットアップすることはできません。いずれにしろ何かをインストールする必要があります。

バージョンと64ビットバージョンの両方をインストールします。

(訳注: R-3.1.3-snowleopard.pkgとR-3.2.0-mavericks.pkgがあるようです)

Linuxユーザーは、ディストリビューションが持つ標準の仕組みによってRをダウンロードしてください。例えば、UbuntuやDebianならapt-get、SUSEならzypperなどです。これにより、Rはビルドされ、インストールされます。

1.2: Rのバージョン

本書(原著)の執筆時点では、Rのバージョンは3.0.2です。これは、前のバージョンである2.15.3から大きく変わりました。CRANは一年ごとのリリースサイクルに従っています。これはメジャーバージョンアップと呼ばれ、バージョンの3つの数字のうち、真ん中の数字が増やされます。例えば、バージョン3.0.0は2013年にリリースされました。2014年には、バージョン3.1.0となるでしょうし、2015年には3.2.0がリリースされるでしょう(訳注: 2015年5月現在で3.2.0。本翻訳書はこのバージョンで確認しています)。3つの数字のうち一番右の数字は、現在のメジャーバージョンに対するマイナーアップデートのために使われます。

通常、Rの大部分の機能は、以前のバージョンとの後方互換性があります。

1.3: 32bitと64bit

コンピューターが64ビットをサポートしており(最近の多くのマシンではそうですが)、そのデータサイズでの動作を保証しているなら、32ビットを使うか64ビットを使うかの選択ができます。64ビットバージョンは任意の大容量メモリ(もしくはRAM)を扱うことができるので、そちらを使った方がいいでしょう。

これはバージョン3.0.0から始まった特に重要なことなのですが、64ビット整数のサポートが追加されました。つまり、より大きな量のデータをRオブジェクトとして保存できるようになったのです。

過去には、いくつかのパッケージが32ビットバージョンのRを要求していましたが、今ではそのようなパッケージは極めて少なくなりました。今日32ビットバージョンをインストールする理由としては、引き継いだ古い分析をサポートしたい場合か、Intelの低電力Atomチップのような32ビットプロセッサのマシンを使う場合くらいです。

1.4: インストール

Windows または Mac での R のインストールは、他のプログラムのインストールと変わりありません。

1.4.1: Windows でのインストール

ダウンロードしたインストーラーを見つけてください。Windows ユーザーには図 1.1 のように見えるでしょう。

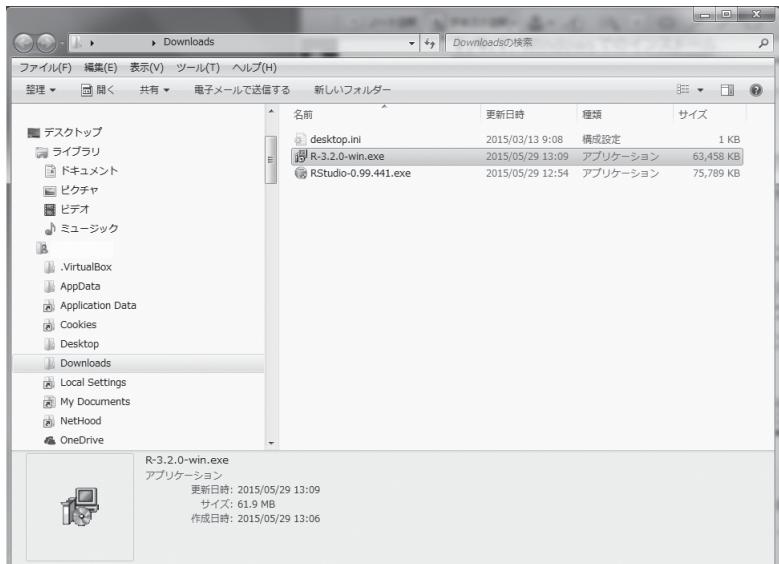


図 1.1 R インストーラーの場所

R は管理者権限でインストールします。つまり、インストーラーを右クリックして、管理者権限で実行を選択してください。すると、プロンプトが立ち上がるので、管理者のパスワードを入力します。

最初のダイアログ (図 1.2) は、言語の選択です。デフォルトは日本語です。適切な言語を選択して、OK をクリックしてください。

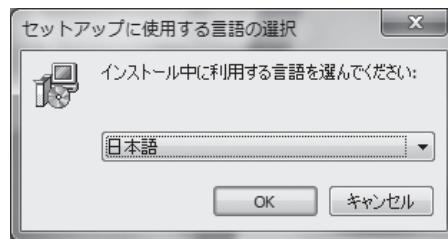


図1.2 言語の選択

次に、図1.3のような注意が表示されます。これは、他のプログラムを全て停止させることを推奨しています。このアドバイスは、めったに守られないか、もう必要ではありません。「次へ」をクリックするのが適切です。



図1.3 Windowsの現バージョンでは、この提案は問題なく無視することができる

続いてソフトウェアライセンスが表示されます（図1.4）。Rはこの（重要な）ライセンスに同意することなく使用することはできません。「次へ」をクリックしてください。



図1.4 Rを使うには、このライセンスに同意しなければならない

インストーラーは次に、インストール場所を聞いてきます。CRANは公式に、Rは空白を含まないディレクトリにインストールすべきだと言っています。ほとんどの場合、デフォルトのインストールディレクトリはProgram Files¥Rになりますが、これは、C++やFORTRANなどでコンパイルされたコードを要求するパッケージをビルドしようとしたときにトラブルの原因となります。図1.5はこのダイアログを示しています。

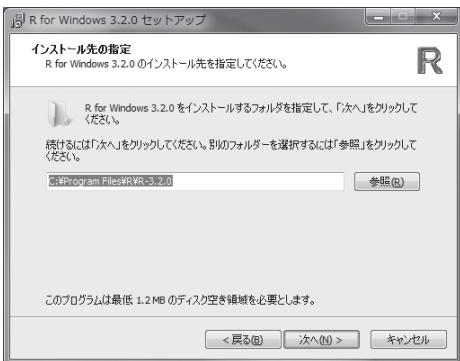


図1.5 空白を含まないフォルダ名をインストール先に指定することが重要

この場合、「参照」ボタンを押し、図1.6に示すようなフォルダオプションを表示させます。



図1.6 このダイアログは、インストール先のフォルダーを選択するために使用する

インストール先としてC:ドライブ(または、他のハードディスクドライブ)もしくは、マイドキュメントを選択するのが最良です。マイドキュメントはユーザーに分かりやすい名前ですが、実際の場所はC:¥Users¥<ユーザー名>¥Documentsなので空白を含んでいません。図1.7に適切なインストール先を示しています。

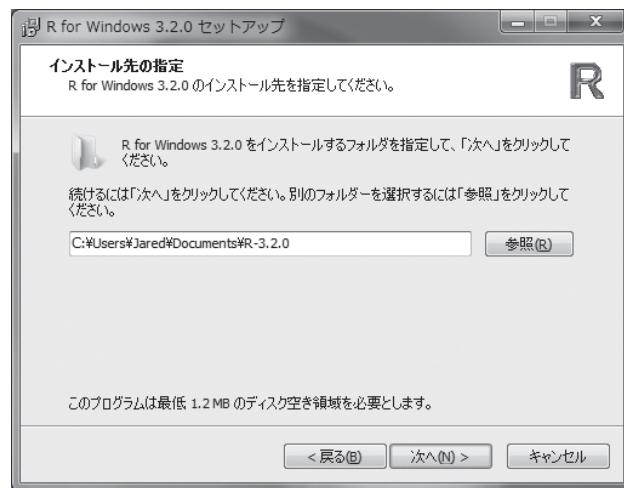


図1.7 名前に空白を含まない、適切なインストール先

次に、図1.8に、インストールするコンポーネントのリストを示します。（64ビットOSで使用する場合は）32ビットファイルは、特に必要がない限り、チェックを外してください。その他は全てチェックをつけてください。



図1.8 32ビットコンポーネント以外全て選択するのが最良

起動時オプションはいいえ（デフォルトのまま）にしてください（図1.9）。なぜなら、多くのオプションがあるわけでもないし、我々はフロントエンドとしてRStudioを使用することを推奨するからです。



図1.9 スタートアップオプションはデフォルトのままにする。我々はフロントエンドとしてRStudioを使うことを推奨するため、この選択はそれほど重要ではない

次に、スタートメニューショートカットをどこに置くかを選択します。我々は、単にRを使い、そこに全てのバージョンを置きます（図1.10）。

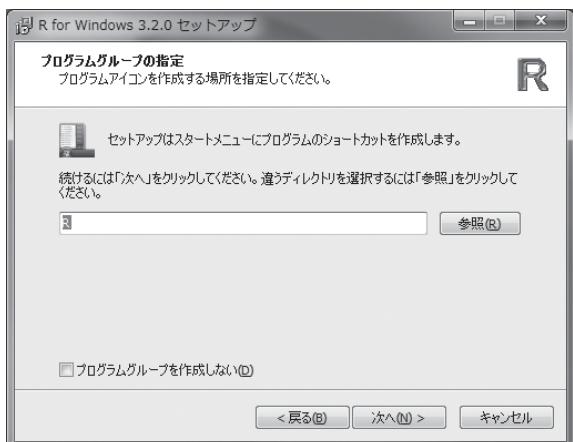


図1.10 ショートカットがインストールされるスタートメニューフォルダーを選択します。

我々は1つのコンピューターに複数のバージョンのRをインストールしており、その全てが同じスタートメニューフォルダーにあります。これにより、異なるバージョンでのテストができます。この状況を示したのが図1.11です。

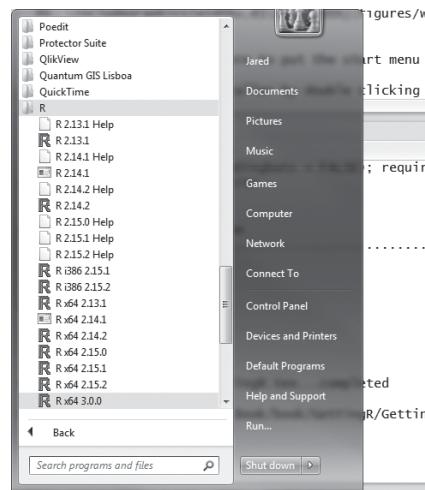


図1.11 我々は複数のバージョンのRをインストールしており、異なるバージョンを使った開発とテストを行うことができる

最後のオプションは、デスクトップにアイコンを追加するか（RStudioを使うなら必要ありません）というような、ちょっととした追加作業を行うかどうかの選択です。我々は、バージョン番号をレジストリに保存し、RDataファイルをRと紐づけることを強く勧めます。これらのオプションを図1.12に示します。



図1.12 我々は、バージョン番号をレジストリに保存し、RDataファイルをRに紐づけることを強く勧める

「次へ」をクリックすると、インストールが始まり、プログレスバーが表示されます（図1.13）。

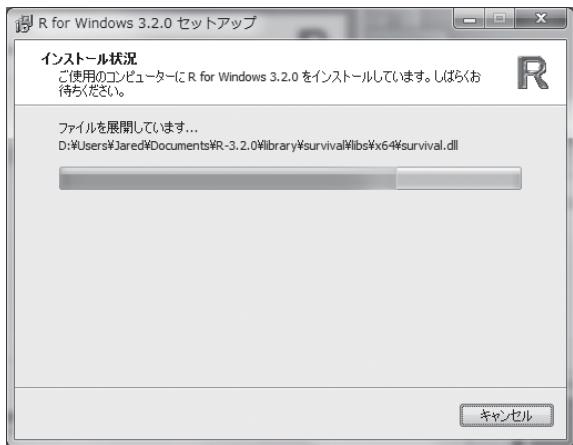


図1.13 インストール中はプログレスバーが表示される

最後のステップとして、完了ボタンを押しインストールを完了してください（図1.14）。



図1.14 インストールが完了したことの確認

1.4.2: Macでのインストール

.pkgで終わるインストーラーファイルを、ダブルクリックして起動します。図1.15に示すように、説明の画面が表示されます。続けるをクリックしてインストール作業を始めましょう。



図1.15 Macでのインストールに対する説明画面

インストールしようとしているRのバージョンについての情報が表示されます。ここでは、続けるをクリックする以外できません（図1.16）。

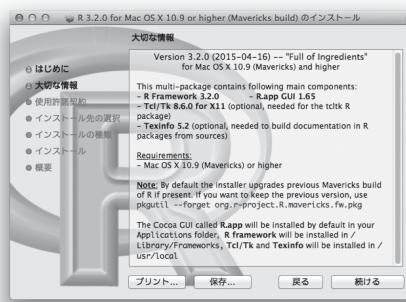


図1.16 バージョンの選択

続いてライセンス情報が表示されます（図1.17）。続けるをクリックして進めてください。それがRを使用するための唯一の現実的な選択です。

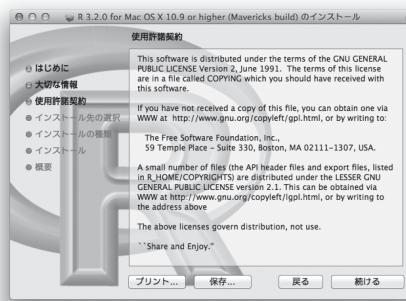


図1.17 Rを使うには、このライセンス契約を認めなければならない

ライセンスに同意するために、同意するをクリックしてください。図1.18で示されている通り、これはRを使用するためには必要な義務です。

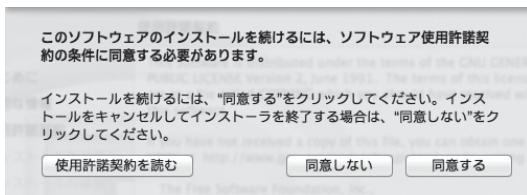


図1.18 ライセンス契約に同意しなければならない

すべてのユーザーに対してRをインストールするために、このコンピュータのすべてのユーザ用にインストールをクリックしてください。もしくは、特定のディスクにインストール...をクリックして、別の場所を指定してください。この状況を図1.19に示します。



図1.19 デフォルトではRはすべてのユーザーに対してインストールされますが、特定の場所を選択するオプションもある

図1.20に示すようなプロンプトが表示されたら、パスワードを入力する必要があります。



図1.20 インストールのために管理者パスワードが要求されるかもしれない

図1.21に示すように、インストールが開始され、プログレスバーが表示されます。



図1.21 インストール中はプログレスバーが表示される

インストールが完了すると、成功画面が表示されます（図1.22）。



図1.22 これは、インストールが成功したことを示している

1.4.3: Linuxでのインストール

ディストリビューションが持つ標準の仕組みからRを検索すれば、Rのダウンロード、ビルド、インストールをワンステップで実行できます。

1.5: Revolution R Community Edition

Revolution Analytics は、Visual Studio をベースとした統合開発環境を伴う、彼ら独自の R ビルドのコミュニティバージョンを提供しています。これは、Intel Math Kernel Library (MKL) を使ってビルドされており、より高速な行列演算を可能にします。これは、次の Web サイトで無償で提供されています。

<http://www.revolutionanalytics.com/products/revolution-r.php>

(訳注：現在この URL にアクセスすると

<http://www.revolutionanalytics.com/revolution-r-enterprise>

に飛ばされます。コミュニティエディションの URL は

<http://www.revolutionanalytics.com/community>

に変わったようです)

彼らは非常に大きなデータでも動く特別なアルゴリズムを使用できる有償版も提供しています。有償版についての更なる情報は次の Web サイトにあります。

<http://www.revolutionanalytics.com/products/revolution-enterprise.php>

(訳注：これもリンク切れですが、Revolution Analytics Enterprise のページにちゃんと飛ばされます)

1.6: まとめ

この章で、R の全ての機能が使用可能になりました。ただし、その GUI は作りが雑です。2.2 節で詳しく述べますが、RStudio をインストールして、そのインターフェイスを使うのが一番です。RStudio をインストールするためは、他のプログラムと同様に、インストーラーをダウンロードして起動します。

前章でRのダウンロードとインストールを行いました。次はRの使い方に慣れましょう。WindowsにおけるRの標準インターフェイスは図2.1に示すように、かなり質実剛健といった感じです。Macのインターフェイス（図2.2）は若干気が利いています。そしてLinuxでは完全にターミナルと同じです。

他の言語と違って、Rは非常にインタラクティブな言語です。すなわち、1つのコマンドを入力するごとに結果を見るすることができます。C++のような言語では、実行結果を見るためには、コードのまとまりを完全に書いてから、コンパイルして実行する必要があります。Rのオブジェクトの状態と結果は、どんな時でも確認できます。このインタラクティブ性は、Rを動かす上で最も驚くべき性質です。

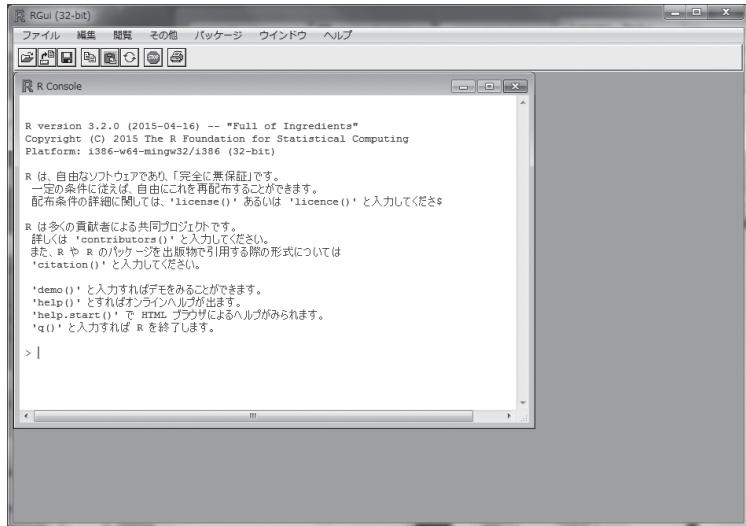


図2.1 WindowsにおけるRの標準インターフェイス

Rをビルドするための統合開発環境（IDE）には様々なものがあります。本書の目的のために、我々はRStudioが使われる場合を仮定します。RStudioについては、2.2節で論じます。

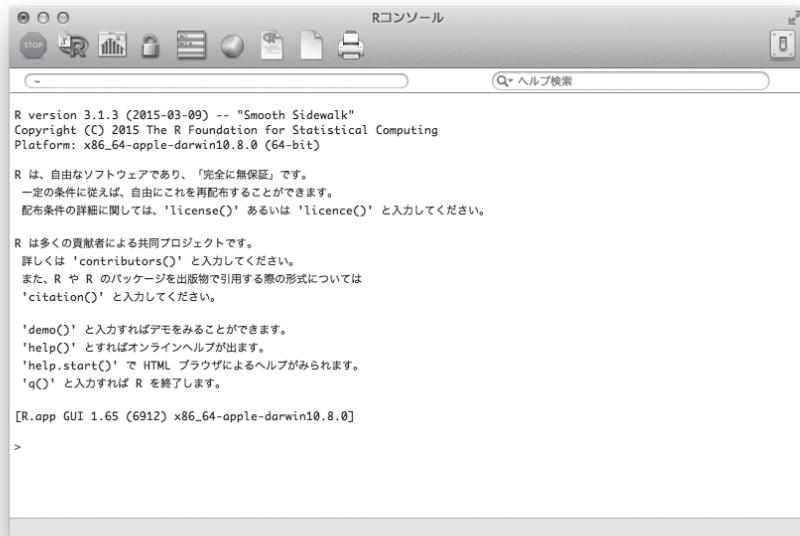


図2.2 (Mac) OS XにおけるRの標準インターフェイス

2.1: コマンドラインインターフェイス (CLI)

コマンドラインインターフェイスはRの持つ非常にパワフルな機能です。しかし、それでRを学んでいると不満が溜まります。Rcmdrのような、Rをマウスのクリック操作で動かすための試みもありましたが、本当にうまくいったとは言えません。これは、コマンドを打つ方がマウスを使うより優れているということの証拠です。特にExcelユーザーにとってはにわかには信じがたいかもしれません、時間とともにコマンドを打つという方法は簡単に思えてきます。そしてその方が間違いが少なくなる傾向にあります。

例えば、Excelで回帰分析を行うときは、少なくとも7回以上マウスをクリックしなければなりません:データ > データ分析 > 回帰分析 > OK > 入力Y範囲 > 入力X範囲 > OK。これにちょっとした変更を加えたいとか、新しいデータを追加したいというときでも、すべての手順をもう一度行わなければなりません。さらに、同僚に電子メールでこれらのステップを丹念に教えるのはもっと難しいでしょう。対照的に、同じことをコマンドで行うときは、1行で済みます。繰り返すのもコピーアンドペーストするのもとても簡単です。最初は信じがたいかもしれません、しばらくコマンドラインを使ってみれば、人生はもっと安らかなものになるでしょう。

Rでコマンドを実行するには、コンソール内の > という記号の横にコマンドを打ち込んでエンターキーを押します。入力するコマンドは数字の2のように単純だったり、8章で見るような複雑な関数だったりします。

直前に打った1行のコードを再び実行したいときは、単に上矢印キーを押してエンターを打つだけです。一度打ったコマンドは全て保存されており、上下の矢印キーによって再び利用可能です。

コマンドの実行を中断するには、Windowsではエスケープキー (Esc)、MacとLinuxではCtrl-Cを押します。

大規模な分析に取り組むときには、使われるコードをファイルとして持つのが良いでしょう。つい最近までは、これを取り扱う最も一般的な方法は、TextPadやUltraEditのようなテキストエディタ¹を用いてコードを書き、Rのコンソールにコピーして貼り付けるというものでした。このような方法で分析を行うと、プログラム間を頻繁に行き来するはめになり、ぐちゃぐちゃになります。

2.2: RStudio

RのIDEはいくつか存在しますが、現在最も良い選択はRStudioです。RStudioはColdFusionやWindows Live Writerを作ったJJ Alaireが率いるチームによって作られました。RStudioはWindows、Mac、Linuxの全てに対応していて、その見た目や操作感は、どのプラットフォームでも同じです。さらに印象的なのは、RStudioサーバーです。RStudioサーバーはLinuxサーバー上でRのインスタンスを走らせ、これによりユーザーはWebブラウザ上でRStudioの標準的なインターフェイスを使ってコマンドが実行できます。RStudioはR 2.11.1以上の全てのバージョンで動き、Revolution Analytics社のRevolution Rでも動きます。RStudioには、やや圧倒されてしまうほど多くのオプションがあります。ここでは、最も役に立つ機能と頻繁に使う機能について紹介します。

RStudioは高度にカスタマイズ可能ですが、基本的なインターフェイスはだいたい図2.3のようになります。この図では、左下のボックスはRコンソールです。これは、通常のRコンソールと全く同じように使えます。左上のボックスはテキストエディターの代わりですが、はるかに強力です。右上のボックスには、ワークスペース、コマンド履歴、現在のフォルダにあるファイル、そしてGitによるバージョンコントロールについての情報があります。右下のボックスは、プロット、パッケージ情報、ヘルプファイルを表

¹ これはMicrosoft Wordのようなワープロソフトではなく、プログラミングテキストエディターを意味します。テキストエディタがテキストの構造を保持するのに対して、ワープロソフトはテキストに整形を加えることがあるため、コンソールに挿入できないようになるかもしれません。

示します。

エディタからコンソールにコマンドを送って実行する方法は、いくつかあります。1行のコマンドを実行したい場合には、実行したい行にカーソルを置いて **Ctrl+Enter** を押します (Mac の場合は **Command+Enter**)。選択した行を実行したい場合には、単に選択したまま **Ctrl+Enter** を押します。ファイル全体のコードを実行したい場合は、**Ctrl+Shift+S** です。

オブジェクト名や関数名を入力しているときに **Tab** を打つとコードが自動補完されます。もし入力された文字に複数のオブジェクトや関数が対応するなら、図2.4に示すように候補から選択できるダイアログが現れます。

Ctrl+1 と打つと、カーソルはテキストエディタに移動し、**Ctrl+2** と打つとコンソールに移動します。テキストエディタで1つ前のタブに移動したいときは、Windowsなら **Ctrl+Alt+←**、Linuxなら **Ctrl+PageUp**、Macなら **Ctrl+Option+←** を押します。次のタブに移動したいときは、Windowsなら **Ctrl+Alt+→**、Linuxなら **Ctrl+PageDown**、Macなら **Ctrl+Option+→** を押します。RStudioで使える全てのショートカットのリストは **Help >> Keyboard Shortcuts** で見ることができます。

(訳注：Windowsで **Ctrl+Alt+←** を押してみたら画面の向きが変わりました。もとに戻すには **Ctrl+Alt+↑**)

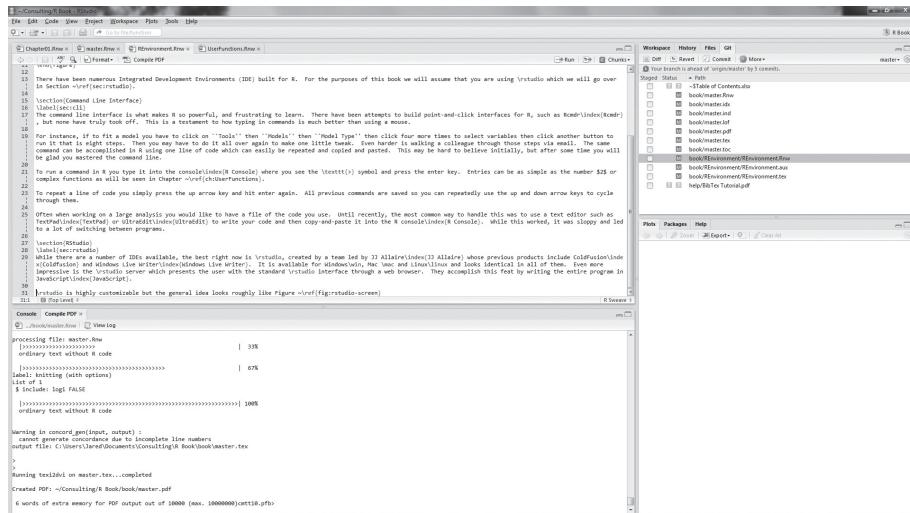
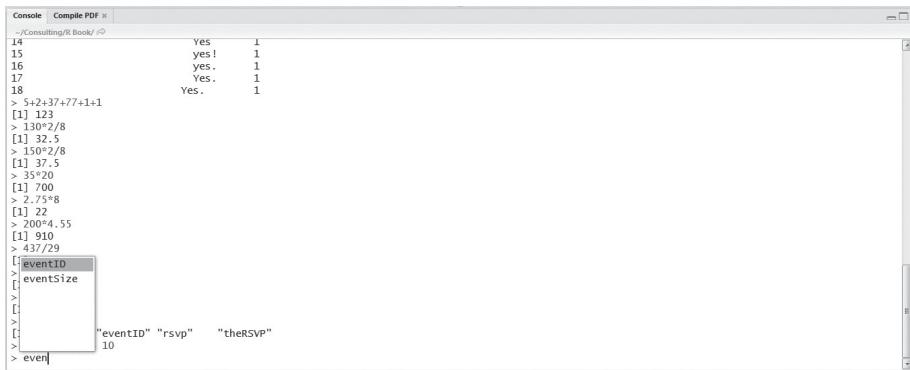


図2.3 RStudioの一般的なレイアウト



The screenshot shows the RStudio interface with the 'Console' tab selected. The code in the console is:

```

Console  Compile PDF x
~/Consulting/R Book/ / 
14      yes     1
15      yes!    1
16      yes.    1
17      yes.    1
18      Yes.    1
> 5+2+3+7+7+1+1
[1] 123
> 130^2/8
[1] 32.5
> 100^2/8
[1] 37.5
> 35^2/20
[1] 700
> 2.75^8
[1] 22
> 200^4.55
[1] 910
> 437/29
[1] 15
> even
[1] eventID
[2] eventSize
[3] 
[4] 
[5] 
[6] 
[7] 
[8] 
[9] 
[10] "eventID" "rsvp"      "theRSVP"
> even

```

The code completion dropdown for 'even' is open, showing suggestions: 'eventID', 'eventSize', and 'even'. The 'eventID' suggestion is currently selected.

図2.4 RStudioの自動補完

2.2.1 : RStudioプロジェクト

RStudioを使う上で最も重要な概念はプロジェクトです。プロジェクトとは、互いに関連しあったファイル(データ、結果、グラフも含む)の集まりです²。各プロジェクトは自分のワーキングディレクトリを持ちます。これはプロジェクトがしっかりとまとまりを保つための素晴らしい方法です。

一番シンプルに新しいプロジェクトを始めるには、File >> New Project をクリックします(図2.5)。

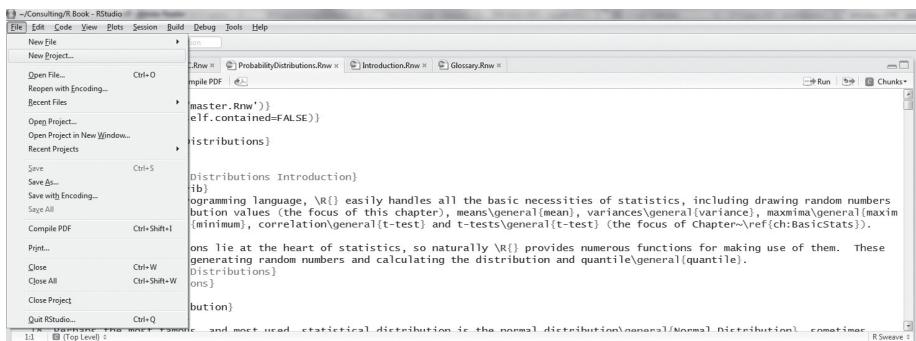


図2.5 File >> New Project をクリックしてプロジェクトの作成を始める

図2.6に示すように、次の3つのオプションがあります:新しいディレクトリで新しい

2 プロジェクトはRセッションとは異なる概念です。Rセッションとは、Rの利用中にメモリ上に保持される全てのオブジェクトと実行結果のことであり、それはRを再起動するとリセットされます。

プロジェクトを始める (New Project)、既存のディレクトリにプロジェクトを関連付ける (Existing Directory)、Git や SVN のようなバージョン管理リポジトリからプロジェクトをチェックアウトする (Version Control)。3つのケース全てにおいて、結果となるディレクトリに .Rproj ファイルが配置され、プロジェクトに関する情報が保存されます。

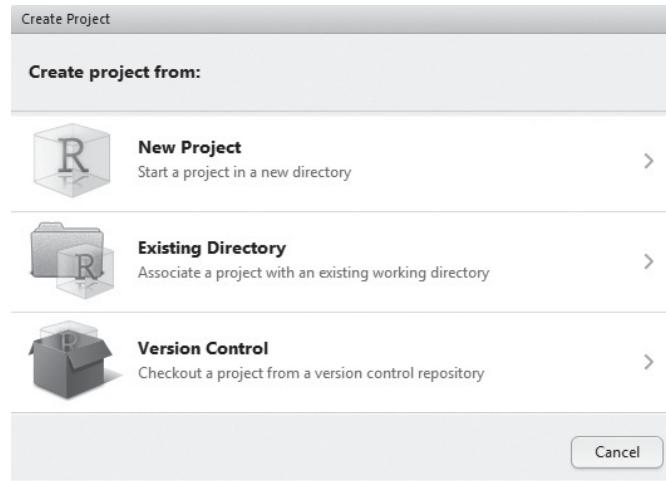


図2.6 新しいプロジェクトを始める際に、3つのオプションがある：新しいディレクトリ、既存のディレクトリにプロジェクトを紐づける、バージョン管理リポジトリからプロジェクトをチェックアウトする (New Project、Existing Directory、Version Control)

新しいディレクトリの作成を選ぶと、プロジェクト名とどこに新しいディレクトリを作成するかを入力するダイアログが表示されます (図2.7)。

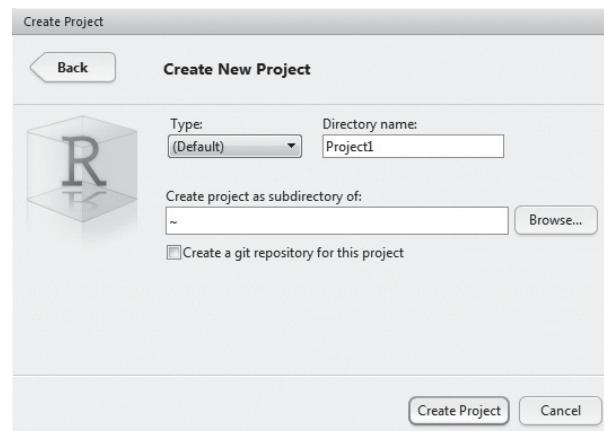


図2.7 新しいプロジェクトディレクトリをどこに作成するかを選ぶダイアログ

既存のディレクトリの使用を選ぶと、ディレクトリの名前を聞かれます（図2.8）。

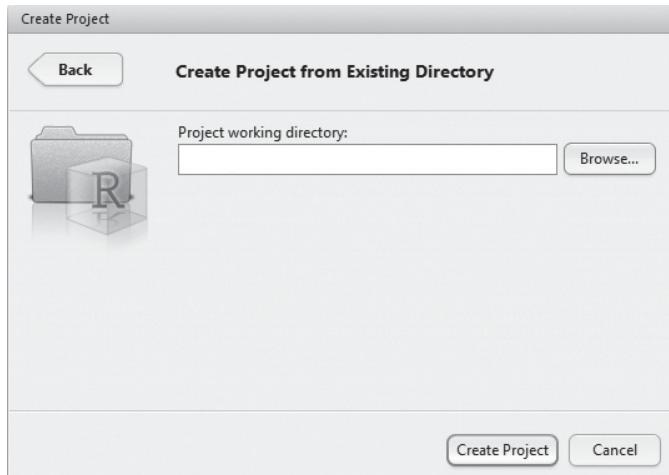


図2.8 どの既存ディレクトリでプロジェクトを始めるかを選択するダイアログ

バージョン管理の使用を選ぶと、まずはGitを使うかSVNを使うかを聞かれます（図2.9）。

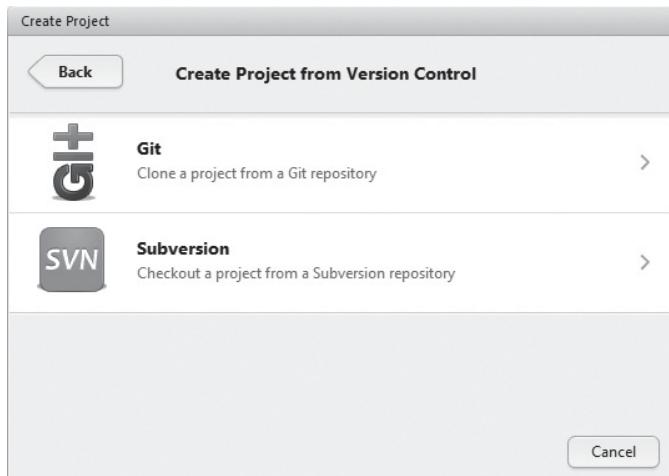


図2.9 どのタイプのリポジトリで新しいプロジェクトを始めるかを選択する

Gitを選択すると、リポジトリのURL（例えば`git@github.com:jaredlander/coefplot.git`）を聞かれます（図2.10）。URLを入力するとプロジェクトのディレクトリ名が自動的に入ります（訳注：変更可能です）。新しいディレクトリの作成のときと同様に、どこにディレクトリを作成するかについても聞かれます。

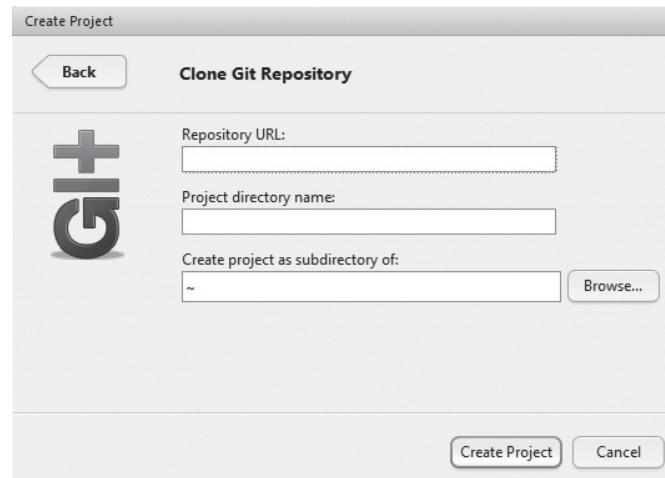


図2.10 GitリポジトリのURLと、どのフォルダーにクローンするのかを入力する

2.2.2 : RStudioツール

RStudioには多数のオプションがあり、高度にカスタマイズ可能です。オプションのほとんどはTools > Optionsをクリックすることで開くオプションダイアログに含まれます。

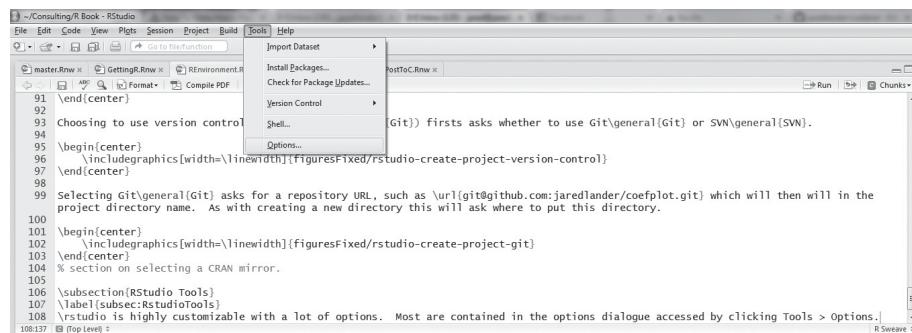


図2.11 Tools > Optionsをクリックすることで、RStudioのオプション画面が出てくる
(訳注: 現在はTools > Global Optionsで全体の、Tools > Project Optionsでプロジェクト単位のオプションが指定できます)

最初に表示されるのはGeneral（全般）オプションです（図2.12）。ここでは、どのバージョンのRを使うかを設定できます。これは、コンピューターにいくつかのバージ

ヨンのRが入っているときに非常に役立ちます。ただし、Rのバージョンを変えたあと、RStudioを再起動させなければなりません。将来的には、RStudioは各プロジェクトに対して異なるバージョンのRを設定できるようになる予定です。また、起動時に.RDataを復元しないようにすることと、終了時に.RDataを保存しないようにすることは良いアイデアです³。

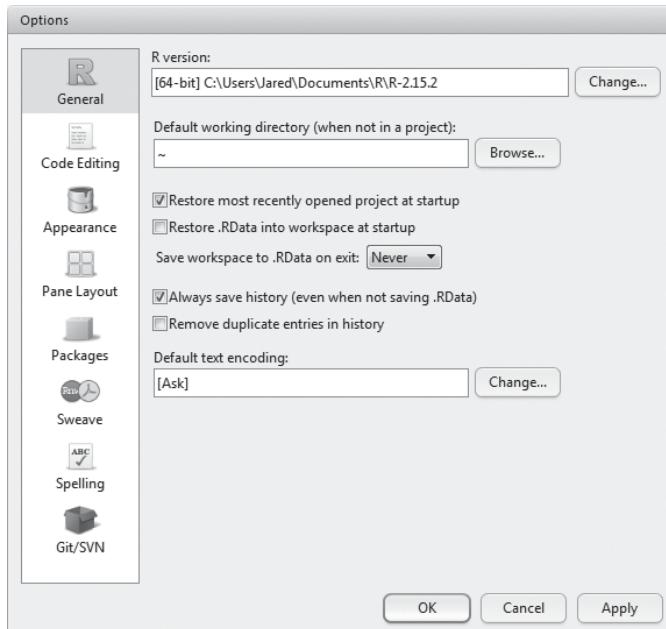


図2.12 RStudio全般に関するオプション

Code Editing（コード編集）オプション（図2.13）はテキストエディターでのコードの入力法と表示法を管理します。一般的に、タブをスペース（2個か4個）に置き換えることは良い習慣です。コアなプログラマー達はvimモードが使えることを高く評価するでしょう。今のところEmacsモードはありません。

3 .RDataファイルは、Rのオブジェクトを保存して共有する便利な方法です。詳しくは6.5節で論じます。

（訳注：起動時に.RDataを復元しないようにするには、Restore .RData into workspace at startupのチェックを外します。終了時に.RDataを保存しないようにするには、Save workspace to .RData on exitをNeverにします）

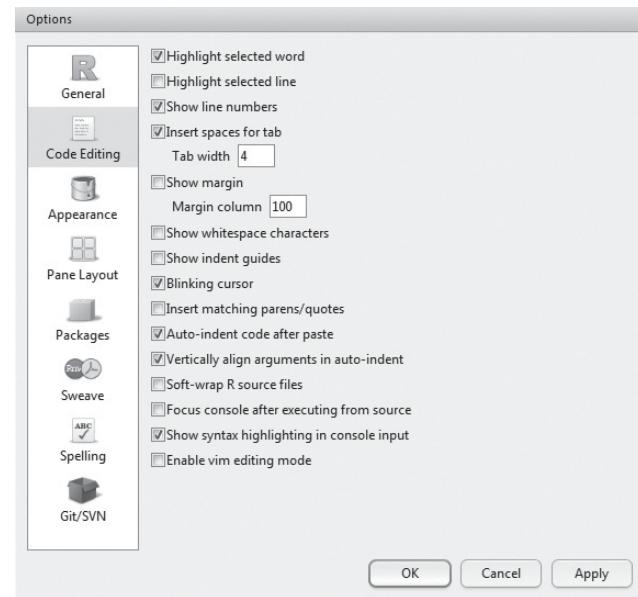


図2.13 コード編集ウインドウをカスタマイズするオプション

Appearance (外観) オプション (図2.14) は、コードの見た目を変えます。ここでは、背景やテキストの色、フォント、サイズをカスタマイズすることができます。

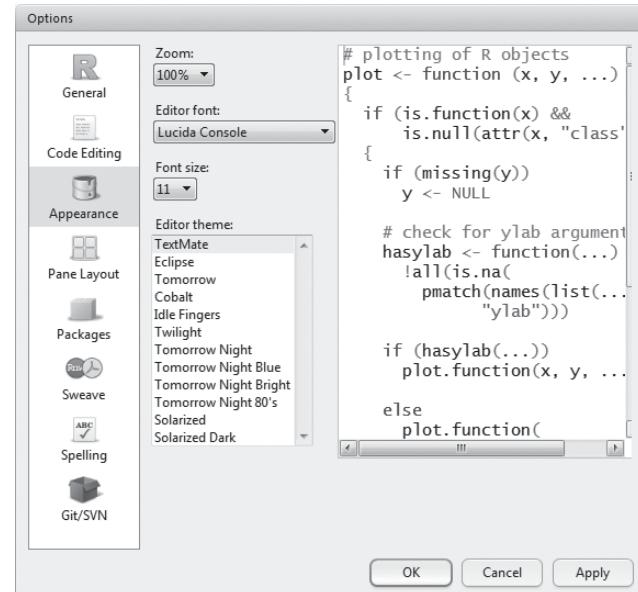


図2.14 コードの外観を変更するオプション

Pane Layout(ウインドウ配置)オプション(図2.15)は、RStudioを構成するウインドウを再配置します。

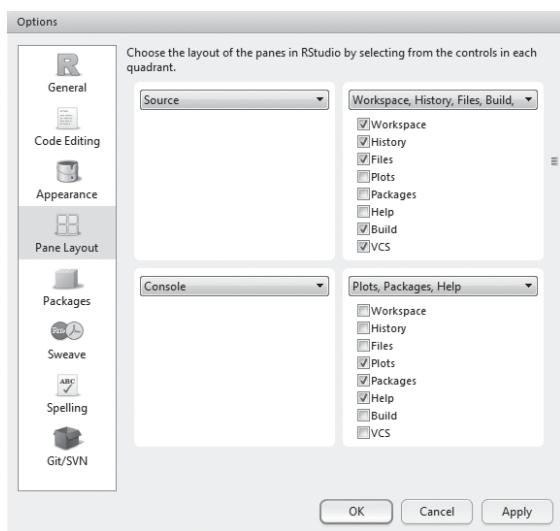


図2.15 RStudio内の様々なウインドウの配置に対するオプション

Packages(パッケージ)オプション(図2.16)は、パッケージに関するオプションを設定します。最も重要なのはCRANミラーの設定です。これはコンソールからも変更可能ですが、ここではデフォルトのCRANミラーを設定します。地理的に最も近いミラーを設定するのが良いでしょう。

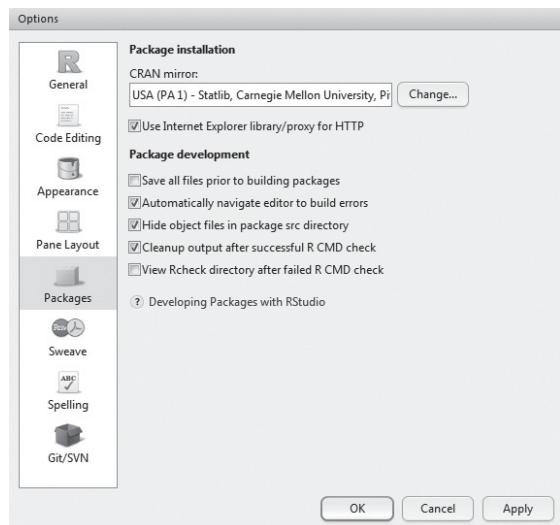


図2.16 パッケージに関するオプション。CRANミラーの選択が最も重要

Sweave (図2.17) は、ちょっと名前が紛らわしくて、Sweaveとknitrのどちらを使うかを選ぶことができます。どちらもPDF文書を生成するのに使われるツールであり、knitrはHTML文書も生成することができます。23章で詳しく述べますが、knitrを使う方が断然良いです。ただし、3.1節で説明するように、knitrは最初にインストールする必要があります。Sweaveオプションでは、PDFビューアーを選択することもできます。

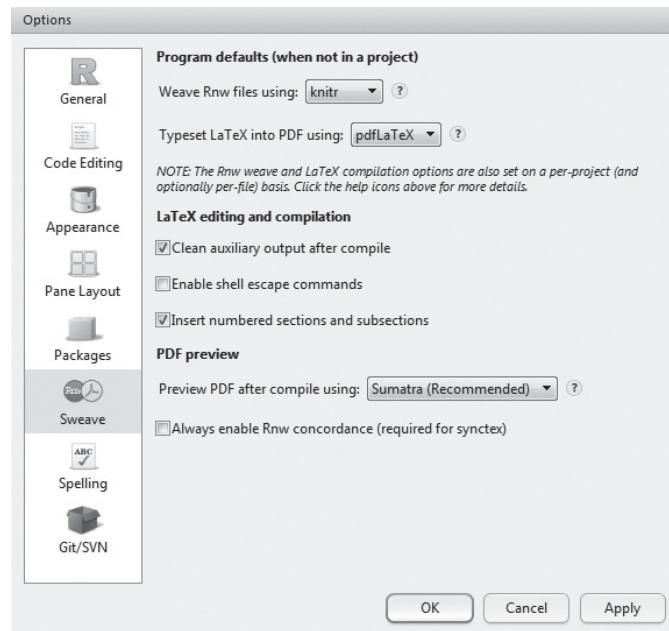


図2.17 Sweaveとknitrのどちらを使うかを選択する。また、PDFビューアを選択する

RStudioはLaTeXとMarkdown文書(knitrを使うことが望ましい)を書く際に使える、スペルチェッカーを持っています。この設定はSpellingオプション(図2.18)で行います。ここでは、特に何も設定しなくても大丈夫でしょう。

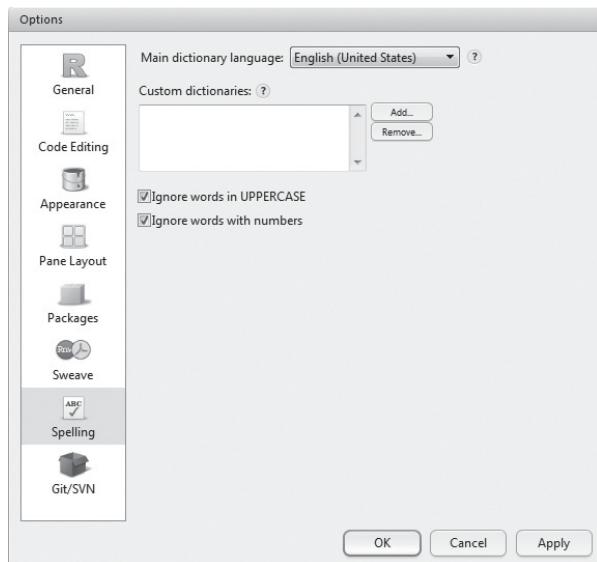


図2.18 スペルチェック用辞書のためのオプション。言語の選択とカスタム辞書の設定が行える

最後のオプションはGit/SVNです(図2.19)。これは、GitとSVNの実行ファイルがどこにあるかを指定します。このオプションはバージョン管理を使うために一度だけ設定する必要があります。

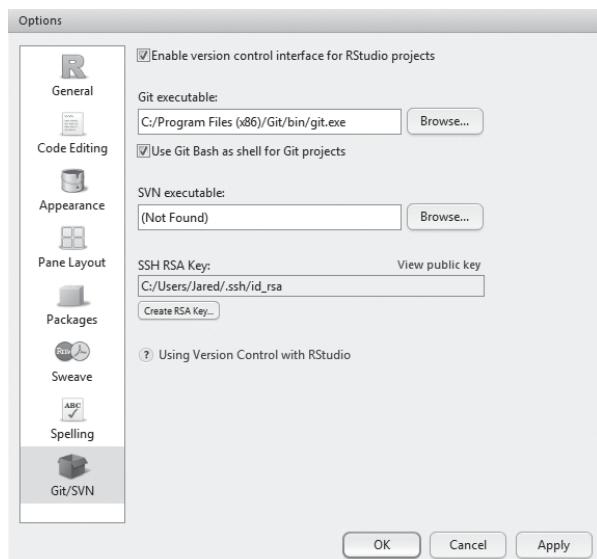


図2.19 RStudioで使用するGitとSVNの実行ファイルがどこにあるかを設定する



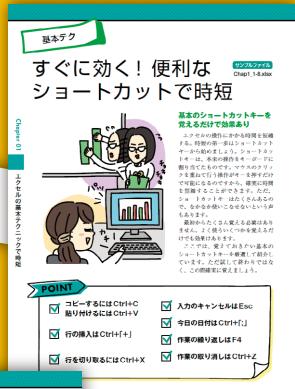
試し読みはお楽しみ
いただけましたか？

ここからはManatee
おすすめの商品を
ご紹介します。

Tech Book Zone
Manatee 

Excelの機能をフル活用して効率化&高速化!

請求書を発行したり、売上管理を行ったりと、仕事のさまざまなシーンで使うエクセルは、煩雑で時間がかかる作業が多いもの。でもエクセルには、そうした面倒な作業を効率化する機能がたくさん用意されているのです。本書は「複数シートを別々に作業するのが大変」「表の体裁をすばやく整えてスッキリ見せる」「ピボットテーブルでパパッと集計する」などのシーンに分けて、身につければ作業時間を減らせる時短テクを解説します。



どっつきやすいイラストで、読み手の理解をサポート!



すぐに試したい時短テクニックを数多く用意。面倒な作業を効率化しよう

テキパキこなす!
ゼッタイ定時に帰る
エクセルの時短テク 121

インプレス
尾崎裕子・タトラエディット（著者）
128 ページ
価格：972 円（PDF）



劇的にプレゼンが変わる! 目からウロコのテクニックが満載!

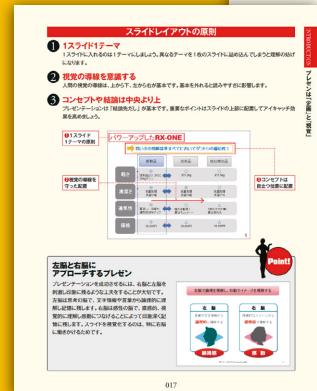
ちょっとしたPowerPointのテクニックで、いつものプレゼンが劇的に変わります。たとえばスライドがゴチャゴチャしている場合、「4分割は魔法のレイアウト! 4つに区切れば必ずスッキリする」のように、本書は“今日から使える”75個ものテクニックを紹介。スライドデザイン／文字／色使い／表／グラフ／図解／写真・イラスト／動き／プレゼン本番のカテゴリで、それぞれ使えるテクニック・押さえておくべきポイントを解説します。

「ビジネスアプリ」

企画立案やスライド作成など、
プレゼンを行う前に
検討すべきポイントを解説



016



017

スライド作りのカギは、
レイアウト、カラーリング、
図解作成

魅せる PowerPoint テクニック

マイナビ出版
野々山美紀（著者）
240 ページ
価格：2,138 円（PDF）



