今日の流れとしては、先ほど説明したとおり、簡単な自己紹介の後に講義概要をお話しさせていただいた後、R Markdownを学ぶということで、この中でR Markdownを実際にやったことがあるという人はどのくらいいますか。

00：10：04

　何名かいますね。ありがとうございます。

　なかなか国内でもR Markdownをチュートリアル形式で教えているところは少なくて、あと書籍とかウェブサイトもあまりないのが現状です。●先生、どちらで？自分で勉強されてという形になりますか。●のほうですか。ありがとうございます。なかなかそういう形式はないと思います。私も独学で勉強した部分もありますので、誤ったことはなるべくお伝えしないようにしたいと思いますが、皆さんのほうでも今回の講義を受けて自分の研究のほうに引用していってもらいたいと思っています。

　私はスポーツ科学や心理学が専門ですが、今まで私自身は特に調査研究をメインにやってきました。実験や脳画像の研究はしてこなかったのですが、調査研究の場合には統計解析ソフトを多く使います。皆さんの中でも結構使われているかもしれませんが、小池先生のほうからぼろくそに言われていました〓SPSS〓は私も特に大学院時代はよく使っていました。

　さらに、最近はエクセルシートでできるHADというソフトもあります。こちらをご存じの先生はおられますか。ありがとうございます。これはとても有用で、これをやってしまうとあまり●ではないかと思うのですけれども、SPSSよりもいろいろな結果が算出できますし、かつ無料でできます。これは関西学院大学の清水先生が作っているのですけれども、「関西学院大学、清水、HAD」で検索するとすぐにできます。Macは最近になって使えるようになったのではないかと思いますし、Windowsは完全に使えます。これも●構造分析からいろいろな分析までたくさん使えますので、Perlのプログラムが少し苦手な人がいたらHADを使うといいのではないかと思います。

　今日、紹介するのはRということですが、私は実はここの研究室に所属して約1年たつのですが、それまではRを全く触ったことがありませんでした。少し＜聞き取り不能＞に使ったぐらいで実際に分析してこなかったのですけれども、こちらのほうに所属するということで臨床情報のQCをやってほしいということからR Markdownを勉強して1年たつわけなのですが、今小池先生がいないのであれなのですが、それまではぼろくそに言われたSPSSの本を書いた先生のところにいました。黄色い本を読まれた人はいますか。SPSSの小塩真司という早稲田大学の先生ですが、どれくらいいますか。もしかしたら結構いるかなと思うのですが、私はそこに3年間ほど学振のPDで所属していまして、その前から7年ほど共同研究をしているのですが、実はこの小塩先生の裏話ではないですが、この本を書かれているのですがSPSSはそれほど使っていないのです。うちの研究室は●今回チューターで入っていますが、もっと自由なソフトを使って、SPSSだけではなくHADやRなどいろいろなソフトを使いながら研究を進めています。その1つとして今回はRを紹介しますが、他でもできる分析はたくさんあると思いますので、皆さんの用途に合わせて使っていただきたいと思っています。

　ということで、今日の講義概要ですけれども、先ほどお話ししたようにまずはR Markdownって何？から始めて、それを学んでもらいたいと思っています。

　そこから実践編に移っていきます。実践編は2つです。まずはR Markdownの使用方法を学んでもらいます。本当に簡単なことから始めてもらいますが、基本的には皆さんが先ほどやったR Studio上ですぐにできます。多分1分もかからずに、「こんなに簡単にできるんだ」ぐらいでできますので、皆さん自身もぜひ今回の講義を受けて今後使ってもらいたいと思っています。

　さらに、臨床情報Quality Checkということで、RでもできるのですがR Markdownのメリットとしては実はドキュメント形式でできます。皆さんの中にはRになじみがなかったり、あとは他の先生と情報を共有したいときに、Rを使用したことがない先生が来るときはコードをいちいち説明しないといけません。しかし、R MarkdownはHTMLの●形式で算出されますので、それを利用することによって他の先生との共有もできますし、さらにMarkdownという機能がついていますので文章の入力ができます。

00：15：13

　まず説明文を入力しておけば、こういうコードで入力して結果はこうなったのだなというのを報告書形式で見られますので、かなり有用なのではないかと思っています。

　ですから、今回の講義では特に新しいことをやるというよりは、このRを使ったものをいかにして可視化していろいろな形で共有していくか、そういうところに注目して紹介したいと思っています。皆さん自身、今回学んだことを全てうのみにしないことということで、1時間半程度の講義になりますが全て紹介し切れません。一応QCまではやりたいと思っていますが、皆さん自身の研究にうまくアレンジしていってほしいと思っています。

　さらに、入門編であることということで、R Markdownはとてもいろいろな使い方ができるのです。今回はHTML形式で出すということですが、PowerPointなど論文としてまとめるいろいろな形がありますので、ぜひ今回をきっかけにしていろいろな形で見てもらいたいと思っています。日本ではまだまだ普及されていないのですが、海外では多く使われていますので、Googleで調べればいろいろな形で出てくると思いますので、ぜひ利用してもらいたいと思っています。

　Rの方法については小池先生のほうから紹介がありましたので、細かいコードの話に関しては触れませんのでご容赦ください。先ほどお話ししたとおり、私は統計の専門家ではありませんので、この点に関してはご容赦いただければと思います。

　それでは、よろしくお願いします。最初にR Markdownを学ぶということで始めていきたいと思います。

　まずR Markdownに関する書籍のご紹介ということで、2018年まではR Markdownに関する書籍は全くなかったのですが……。

0：17：01

00：00：00

上野：私が知らないだけでもっとあるのかもしれないですが、今のところ私が知る限りでは3つあります。

　まずは、専修大学の小杉先生という心理統計では有名な先生ですが、『Rでらくらく心理統計』というこの本の一部として入ってきます。あと、『はじめてのR Studio』ということで書かれているのですが、この両外側の書籍に関しては本当に1ページ、2ページしか書かれていません。『はじめてのR Studio』だと少し書かれているかなというぐらいですが、それでもあまり多くないです。両方を読んでもらうのもいいと思いますが、この真ん中の『RユーザのためのR Studio実践入門』という本があるのですが、こちらのほうに関しては結構R Markdownの話が書かれています。私は今回この書籍からも一部抜粋して紹介していますので、ぜひ今日R Markdownの話をしてもう少し勉強したいという先生がいましたら、こちらの本を読まれるといいのではないかと思います。

　ちなみに、この本を書いた著者の1人である前田先生という先生がいるのですが、この前田先生はなんとウェブサイトで内容を紹介しています。ですから、前田先生には申し訳ないですが私は面識がないので紹介してしまいますが、わざわざ購入しなくてもある程度R Markdownの話が書いてある「kazutan」という、もしかしたら聞いたことがあるかもしれませんが、この「R Markdown入門 - kazutan on web」で調べると、R Markdown入門編ということでR Markdownの基礎的なものから実際にどう動かしていくのというのを細かく説明されていますので、ぜひ参考にしつつやってもらいたいと思っています。

　そもそもRでできて何がR Markdownでできるのかという話ですが、R Markdownでできることというのは、簡単にまとめますとRで分析して可視化を行ったものを、その報告書としてドキュメント形式でまとめることが可能になっています。

　例えば、こういう経験があるのではないかと思いますが、RやSPSSなど何でもいいのですが、実際に分析した後にエクセルでまとめたり、違うソフトでまとめるということで間違いが起きやすいケースが多いと思います。

　あとは、図を作るときに、いちいち平均値、標準偏差を用いて新しくプロットを作るとなると結構大変だと思うのですが、R Markdownの便利なところは、Rで使ったコードをそのまま報告書に反映できますので、Rのコードはたくさんありますが、そちらのほうで自分にフィットするようなコードを探して、それで可視化したものをそのままドキュメント形式でまとめることができます。

　あとは、共同研究者と共有するときということで、Rが全く分からない研究者と共有するときに、コードが何が何だか分からない人も多いと思うのですが、その場合に関してもMarkdownの機能も付いています。先ほどもお話ししましたが、基本的には説明文も全部入れることができますので、あとは結果●に関しても記載しておいたほうがいいかもしれません。そうすることによって、あの結果からこういうふうに解釈したというのを間違いなく共有することができるのではないかと思います。

　これが例えば全部手でやってしまうと、結果自体が間違ったものをそのまま共同研究者と共有してしまい、最終的に変な考察になっていって、後々もう1回分析をかけたらというこのルーティンがなくなり、きちんとした正しい結果に基づき、共同研究者と共有し、考察ができるのではないかと思っています。

　そして、Markdownの機能でとても便利なのは、グラフをものすごく作るときにRなどでやってしまうと1個ずつ図で出るので、1個ずつ確認すればいいのですけれども、それが結構面倒くさいときなども基本的にはMarkdownで並べられますので、100個以上のグラフでも報告書形式でまとめることが可能になります。

　実践例として、皆さんのほとんどの方がR Markdownってどういう感じになっているのか分からないと思いますので、私がRを実際に使ったのは1年前という話でしたが、そこから2週間でなんとなくRの使い方を学んで、R Markdownで1本論文を書いたものがありますので、そちらのR Markdownを少し紹介したいと思います。

　前側を見ていただきたいと思います。私は心理学の研究をしているということで、心理学の1つの概念でレジリエンスという精神的回復力という概念があるのですが、全画面ではないので見づらいかもしれませんが、これは基本的に皆さんが先ほどやりましたRで分析したものになります。それをHTML形式でR Markdownでアウトプットするとこのような感じの画面になります。例えばタイトルを入力したり、あとは中の構造を報告書形式で文書を作ることができます。

00：05：04

　さらに、横のサイドメニューバーも作れますので、どこを見たいかによって〓飛ばすことも〓できます。例えば、「本研究で分析するデータの確認」と押すと、このような感じで分析の形が出てきます。先ほどお話しした小池先生の中では、まずリードをCSV関数でデータを読み込みますと、データを読み込んだ後に●関数を使ってこのようにざっと出てきます。これも結構きれいですよね。先ほどのR上では何が何だか分からないというのも、報告書形式でまとめてくれるので、このような感じで丁寧にきれいにまとめてくれます。しかもこれは押すとこのような感じで横に行くことができます。それぞれのデータがどのような感じで入っているのかというのも見ることができます。

　さらに、Rの場合にはグラフ〓化〓にもたけていますので、例えば性別の度数に関してはこのような感じで入力してグラフにすることもできますし、例えばこれは、私は都道府県のデータも採っています。全国2万人のデータを利用したものになるのですが、このデータも例えば、少し見づらいですけれども、このような感じで度数分布表も作れます。

　さらに、先ほどお話しした相関関係のようなものもこのような感じで見ることができます。これだけでは少し見づらいので違うのを見せたいのですが、例えば先ほどお話ししていた相関マップの図表に関しては、これは〓ヒート〓マップを利用しているのですが、これも報告書形式のこの中のR Markdown上で記載することができます。

　あとは下のほうに、先ほど小池先生のところでも紹介があったものもそのままアウトプットとして出ますし、このような感じで円による相関係数の表記の仕方といったようなこともできます。これは完全に自分自身でどうまとめていくかによって変わるのですが、いろいろな形でページができます。ですから、共同研究者の好みに合わせてというか、自分自身が見たいものだけではなくて、いろいろな人に共有するデータであれば、いろいろな形で出しておいてあげたほうが今後分かりやすくなるのではないかと思います。

　今日は皆さんに実際に、このような形でHTML形式で算出ができるのだというのを最後にアウトプットとして学んでいただきたいと思っています。

　行ったり来たりになってしまうのですが奥のほうをご覧ください。

　今回ですけれども、R Markdownを学ぶということでRから始めていくわけですが、Rの便利なところは、まずデータさえあればそこから全部作業ができます。ですから、基本的に間違いが起きないのです。エクセル●などの関数を使っていろいろやっていると最終的に何が起きているか分からない部分があると思います。それがR1つで、さらにこういうふうにR Markdownで落とすことができれば、最終的に間違いに気付くことができます。分析の間違いに気付くのはすごく大切なことなのですが、なかなか気付かないと思います。

　例えば〓SKSS〓、ワンクリックでできるような分析ソフトの場合、自分1人ではできると思いますが他の人に共有ができないですよね。自分がどういうふうにクリックしていくか他の人は分からないですので。ですから、そういうときは作業として結構怖いです。ただ、Rの場合にはコードを入力しますので、他の人もコードを確認して、明らかに間違いがあった場合は指摘することができます。

　あとは、先ほどもお話ししました、それを他の研究者と共有ができますし、さらに何がいいかというと分析の再現可能性があるのです。一度R Markdownで使ったスクリプトは他のデータにも適用ができます。ですから、1回作ってしまえば、例えばデータ1で同じような結果が出ているかどうかデータ2で確認することができますので無駄にはなりません。プログラムを作るときって1回しか使えないのではないかと思う方がいるかもしれませんが、何度でも使えますし汎用性が高いです。ですから、1度プログラムを作ってもらうと、その間違いにも気付けますし、さらに再現可能性が高まるのではないかと思います。

　ということで、簡単にR Markdownをまとめますと、Rでやったことをドキュメント形式でまとめることができます、という話になります。

　ここまでがR Markdownを学ぶという形になりますが、この時点で質問がある方はいますか。大丈夫ですか。ありがとうございます。それでは、立て続けに進めていきたいと思います。

　では、R Markdownを実践するということで行っていきたいと思いますけれども、本日は既に小池先生のほうからRとR Studioの連結に関して〓指示があり〓、皆さまも午前中に特に問題なく動かせましたので問題はないかと思います。基本的にR Markdownを使う条件としては、RとR Studioをダウンロードしてもらって連結させる必要があります。この際の注意点としては、Rでも起きることなのですけれども、R StudioとRのバージョンによってR Markdownがうまく使えないというケースがあるのです。

00：10：04

　私が今動かしているR Studioは1個前のものになります。最新版では動かしていなかったのですが、小池先生の研究室の先生方にも確認してもらい、特に問題なく動かせましたので、最新版でも問題なくできるのではないかと思っています。

　ですから、念のためR〓統一〓のデータということで、Windowsしか私自身が確保できていなかったので、もし今日動かせなかったり、今後問題がありそうであれば、一応私自身が事前ダウンロードのところで、このバージョンだったら動かせますという感じで添付してありますので、そちらのほうをダウンロードしていただいて動かすといいと思います。

　これはRでもそうなのですが、文字化けすることが結構あります。特にR Markdownの場合には日本語を記入することが多いです。ですから、R StudioのGlobal Optionにて、UTF-8に文字コードを変更した上で進めていくことが大切になります。

　では、そのOptionに関してどうやってやるかということですが、前側をご覧ください。文字が小さくて恐縮ですけれども、この〓Tools〓というのがあります。これを押していただいて、Global Optionsというのがありますのでこれを押していただいてもよろしいですか。そうするとこのような感じで、すごく大きく映るのですけれども、この部分のCodeというのがあります。Codeの〓Saving〓というところに、このDefault Text Encodingのところにいろいろな文字があるのですが、UTF-8でもう登録されている先生でしたらそのままで結構ですが、もしそうでなければこちらの文字で登録しておかないと、日本語を読み取ったときに何かすごくよく分からない文字コードで反映されてしまって文字化けが起きてしまいます。ですから、もしこれが済んでいない先生がいましたら、こちらのほうを登録だけお願いできればと思います。よろしくお願いします。

　では、R Markdownに進めていく前に、簡単にというわけではないですけれども、行ったり来たりで申し訳ないですが奥側のスライドを見ていただきたいのですが、基本的にRの分析というのは、自分自身が〓持てるように〓なっている分析、誰かしら絶対にやっているはずだというのを私はとある先生から聞いたことがあります。ですから、皆さまがこれから分析したいものがあれば、Google上で何かしら調べれば出てくる可能性がありますので、今回はR Markdownの臨床情報QCという話だけしかできませんが、それ以外の分析に関しましても、ぜひGoogle上を調べてもらえればいろいろな形で●できるのではないかと思います。

　特にRがいい点としては、すごくすてきな図を作ってくれるのです。例えば、先ほどお話ししたような相関図の話や、あとは年齢の発達曲線とか、こんな感じの●曲線図とか、あとはこれは遺伝的アルゴリズムという分析ですけれども、そういう分析とか、あとこんな感じでいろいろな分析の形をグラフィカルにまとめることができます。最近の研究ですとそういうグラフィカルも結構重要視されていますので、それを学びつつ、それは実際にR Markdownでも表現できますので、ぜひともいろいろやってみてほしいと思っています。

　それでは、R Markdownの起動方法ということで、これからチュートリアルを実施していきたいと思います。

　それでは、最初にRを立ち上げていただいてもよろしいですか。R Studioを先ほどと同様に起動していただければと思います。これは私はもう出してしまっていますが、私が用意しているR Markdownのものはまだ出さなくて結構です。まずはR Markdownをどのように作ればいいかというのを先に説明したいと思っていますのでよろしくお願いします。

　午前中は開けたけれども、午後になって開けなくなったというのを聞きますが大丈夫ですか。ありがとうございます。

　それでは、順番を追ってやっていきたいと思います。少し私のほうが早ければスライドのほうにも手順が書いてありますので、右、左と目が回ってしまうかもしれませんが、もし早ければスライドを見ながら進めていってください。

　それでは、まずRと違うのは、R Markdownの場合はまた別ファイルとして開く必要があります。具体的にはR StudioのFileという画面があります。このFileからNew File、下のR Markdownというのがありますので、こちらのほうをクリックしてください。

　そうするとこんな感じの画面が出てきます。

00：15：05

　ここにTitleとAuthorというのがあります。私の場合は特に記入する必要がありませんのでこのままいきたいと思いますが、皆さまがこれからやる研究やQCなどであれば、事前にTitleを入力して、Authorも入力できますのでここで設定してもらいたいと思います。

　今回ですが、HTMLでアウトプットをします。PDFやWordもあるのですが、なぜHTMLでアウトプットするかというと、先ほど話したようにすごく量が多いとサイドメニューバーがないと〓最後〓分かりづらいのです。特に臨床情報などはとても変数が多くなるとすごく変数の数が多くなって、後々で見づらくなってしまいますので、今回に関してはHTML形式にしたいと思いますのでよろしくお願いします。

　できましたらOKボタンを押してください。

　そうしますと、R Studio上にこのような感じでスクリプトが表記されます。先ほどと違うのは、ここにR Markdownみたいな感じの赤丸で多分RMDというのが表記されています。こちらをまずは保存をしていただきたいと思いますので、名前は何でも結構ですので、まず分析前にこのフロッピーディスクマークがありますので、そちらで名前を付けて保存をしてください。

　こちらは特にデータを読み込みませんので、デスクトップでも結構ですしUSBでも結構ですので、自分が分かりやすいところに保存していただければと思います。

　ここまで分からない先生はおられますか。大丈夫そうですか。

00：17：02

00：00：00

上野：もし早かったら遠慮なくお申し付けください。

　そうしますと、実際にFileからNew File、R Markdownを開いたわけですけれども、こんな感じの画面が出てきます。先ほど入力したtitle、私の場合は名前を付けていませんのでuntitledで、Authorで私の名前、dateで今日の日付ですね。アウトプットにhtml\_documentということで設定がされています。

　では、このあとにどうやったらHTML形式でR Markdownが表記されるのかということですけれども、後ろ側を見ていただきたいのですが、基本的にはR Markdownというのはこのまま分析可能なのですけれども、R Markdown〓情報〓の簡単な説明をしますと、ここの一番上、titleなどでくくられている部分はフロントマター部分ということで、ドキュメント全体設定を記述します。私が先ほど提示した実践例もそうですが、ここは皆さん自身でどういうデザインにして、どういう形でドキュメントとしてまとめたいかというデザイン性を決められるところになります。

　そのあとに、Markdownによる文章の記述。簡単に言いますと、日本語や英語などで文章を記述するというものになります。

　さらに、Markdownと違うのはRというものが付いています。Rというのはいわゆるコードを入力して分析を実行するわけなのですが、そのコードを入力するところがあります。これをR Markdownの他にRチャンクと言われるものがあります。このRチャンク内に先ほど学んだRコードを入力してあげることによって、最終的にR Markdownを実施したときに分析が実行されて、分析実行した結果が〓算出〓される形になります。ですから、このRチャンク内にRコードが入力されていないと何も起きませんので注意してください。

　R Markdownの基本的な条件としては、Knitrパッケージというのがあります。これは特にダウンロードする必要はありませんが、このパッケージを利用してRコードを評価して、その後Pandocアプリケーションでドキュメントへ変換するということで、ここは結構複雑な構造になっているのですが、基本はワンクリックで全て終了しますので、この過程に関しては覚える必要はありません。

　それでは、R Markdownを試しにこの状態で実施していきたいと思います。これはもうデフォルトで決まっている中身になっていますが、ではどうしたらできるかということで、このKnitというのがあります。Knit to HTMLというのがありますので、こちらのほうの画面で今実践例の画面を出していますので、このKnitの三角部分を押していただくとHTMLというのがありますので、これを押してみてください。

　そうすると、下のほうに先ほどまでは出てこなかったコンソール上でRの場合には分析していたと思いますが、R Markdownの場合にはR Markdownという新しいタブが出てきます。こちらのところで、今回入力されたもともとデフォルトのものなのですが、それで分析がされます。基本的にこのパーセンテージが何を意味しているかというと、上から入力されていったMarkdownに機能、さらにフロントマター部分、そしてRチャンクが反映されているかどうかをパーセンテージで表記されます。ですから、100％にならないとHTML形式で出されません。途中で間違っているといった場合には、こちらのどこかでエラーが起きていますので、それを確認いただければと思います。

　大体、起きるエラーとしては、Rコードが間違っていることがほとんどです。ですから、Rコードで間違いが起きないように丁寧に入力する必要がありますが、こちらのところでも何が問題なのかも確認できますので見ていただければと思います。

　最後に、output fileということで、このcreatedの後に、私の場合はqcという名前を付けたのですがqc.htmlでHTML形式で算出されますので少し見てみましょうか。

　そうすると、このような感じで今皆さまに作っていただいたものが、そのままHTML形式のR Markdown上で出てきます。これもデフォルトで入っているものになりますがこのような感じで、先ほどのRとは違うのは文字がまず入っているということです。さらに、タイトルは付けていませんがタイトルが付いていたり、あとは実際にサマリーに関数を使いました。このデータの中身の構造も確認できますし、その後の図表も確認することができます。

00：05：04

　そのような形で、せっかくRを学ぶのであれば、そのままドキュメントにして、さらにグラフィカルにまとめたほうがすごくきれいで見やすくなりますよね。ですから、それを皆さまにやってもらいたいと思っています。

　R Markdownなのですが、使用手順をもう一度おさらいしたいと思っています。今、簡単にもともとあるR Markdownを立ち上げて、その後にKnit to HTMLでHTML形式で出しましたが、R Markdownの使用手順はとても簡単です。

　まずは皆さま、New fileからR Markdownファイルを準備してください。その後に文章、MarkdownやRチャンクを書いてください。そして、先ほどやったKnitです。Knit to HTMLをレンタリング、分析実行を押せば最後にドキュメントが出てきます。この繰り返しになります。

　一番重要なのは、どういう研究目的で、どういう解析をするかが大切になりますので、一番●なのはRチャンクをどう書くかというのが大切になります。

　では、文章やRチャンクに関して、細かい点ではありますが、今回紹介する部分を簡単に説明していきたいと思いますのでよろしくお願いします。

　まずドキュメントの設定に関して、こちらですが結構細かい設定ができるのです。フロントマター部分ということで、実際にどういうデザインにするかという設計なのですが、こちらに関してはデフォルトの状態だと見づらくなりますので、臨床情報QCで私が作ったもののときに丁寧に説明したいと思います。簡単に概要ではないですけれども、このような感じでできるのだということを覚えていただきたいと思います。

　このフロントマター部分のドキュメント全体設定を記述するということで、デフォルトの場合だとこれしか載っていません。これだけだったら先ほど皆さんが出した……ちょっと微妙ですよね。デフォルトを作った人に申し訳ないのですが、たくさん文字だけが載って何が何だかよく分からない形で、これだったらRでも変わらないよねという話になってしまうのですが、これをもう少し見やすくするためにはどうすればいいかというのがドキュメントの設定で行えます。

　このドキュメントの設定なのですが、とてもたくさんあります。これはQCで書いた情報をここに今書いてあるのですが、何を意味しているかというと、基本的にはデフォルトでは5番まで入力がされています。5番以降に関してはオプションです。ですから皆さまがどういうデザインでR Markdownを作りたいかによって変わってきますのでよろしくお願いします。

　私は丁寧に説明文を全部付けています。これは先ほどお話しした書籍の一部を抜粋しています。細かい設定に関してはこちらのほうをご参照いただきたいのですが、私がこの1年間、いろいろな形でやってきた中では、基本的にはこの11項を押さえておけば大体きれいなR Markdownができるのではないかと思います。もし、ここにいる●の方で、ぜひこれを付けたほうがいいよというのがあれば後で教えていただきたいと思っています。これに関しては後でもう一度説明したいと思いますのでよろしくお願いします。

　では、先ほどやったNewfileから出したR Markdownに関して、もう少し見やすくできないかということでやっていきたいと思います。ちなみにこの中でR Markdownは使ったことがないけれどもMarkdownというのを使ったことがあるという先生はいますか。ありがとうございます。基本的にはそれと使い方が似ていると思うのですが、Markdownの基本というのはいわゆる言語です。これをどう反映するかになるのですが、例えば見出しのレベルや箇条書きというのを独自に設定していくわけになります。

　先ほど出したR Markdownのファイルを見てもらい、デモのほうを見ていただきたいのですが、例えば14行目のこのThis isから始まる文章のところというのは、そのまま入力していてもいいのですが少し見づらいなというのが正直私の中であります。これをこのような感じで出てくるのです。これを例えば箇条書きのような感じにしたら見やすいのではないかと思いますので、箇条書きにしたいと思います。

　箇条書きの仕方なのですが、基本的には一番奥のスライドにも書いてあるとおり、半角横棒の半角スペースを押せば、これでもう箇条書きになります。同じように●から始まるところも同じように入力してみてください。半角スペースがないと箇条書きになりませんので注意してください。

　今、箇条書きを設定したのですけれども、では反映されているかどうかをもう1回分析したいと思いますので、同じようにKnitからKnit to HTMLを押していただけますか。

00：10：05

　そうすると下にまた同じように100％が出てきて、多分成功した先生であればこんな感じで丸が付くということで、箇条書きは段落が多くなりますが少し見やすくなるのではないかと思います。これは完全に好みによりますので、私の場合にはこちらのほうが見やすいかなということで箇条書きを付けましたが、先生方におかれましても自分の好みに合わせて箇条書きを使ってもらいたいと思います。

　次に1番目にある見出しというところで、例えばサマリー〓カールズ〓ということで、もともとあるデフォルトのデータを使って分析がかかっているのですが、ここに見出しを作りたいと思います。見出しのレベルですがシャープで作ります。シャープが1個かシャープが2個か、シャープが3個か、シャープが4個かによって見出しのレベルが変わります。いわゆる文字のフォントと大きさが変わるというように思ってください。今回の場合には見出しレベルを、本当はフロントマター部分で設定するのですが、このような感じで見出しが作れますよというのを簡単にやっていただきたいと思います。

　最初にやることは、シャープを入力して、あとは見出しは何でもいいです。「1番目見出し」と入力していただければ大きい見出しで表示されます。これに2個シャープが付けば、その大きい見出しよりもワンランク下がったフォントの大きさで出てきます。これを全部フロントマター部分で見出しレベルを何個まで下げますかというのを設定しないといけないのですが、ですから今回は動かしても1個でも2個でも大きさは変わりません。次のQCのところで細かくやっていきたいと思います。

　これも同じようにKnit to HTMLを押してください。そうするとR Markdownという表記があるのですが、これは2個のシャープが付いているのですが、後で見ていただきたいのですけれどもこれは1個の見出しの画面になります。1個のほうが少し大きいです。これをよりフォントをはっきり目立たせたい場合にはフロントマター部分で設定する必要が出てくるという形になります。

　あとは、段落や強制改行などの話があるのですが、強制改行などに関しては好みがあるとは思いますが、割とよく使われるのは強調表示ということで、太文字にしたり斜体にしたりというのがあります。

　では、斜体にしてもらいたいと思いますので、先ほどの見出しのところ、「1番見出し」というのがありますので、これをアスタリスクでできます。アスタリスク1個、半角でくくっていただけますか。アスタリスク、1ドット見出し、アスタリスクとやるとこのような感じで赤表記になります。何もなしだと白表記のままですが、アスタリスクで囲むことによってこのような感じできれいに●ができます。

　これを押していただくと、同じように分析をすると、このような感じで斜体になります。1ドット見出しで斜体になります。大丈夫そうですか。

　今のところはシャープで見出しレベルを1に設定していただいて、アスタリスク1個で斜体にしています。これが2個になると何かというと、奥のスライドにも書いてありますが太文字になります。見出しレベルが1なので、もともと大きいのであまり変わりありませんが、ちょっと2個にしてみましょうか。これで同じように、これはKnit to HTMLを押さなくてもKnitだけでも分析はできるのです。もともとデフォルトで設定していますので、これで分析していただくと先ほどより少し太文字になっているという感じになると思います。R Markdownの文字を見ると、太文字が強調されているのではないかと思います。

　その他のMarkdown機能もたくさんありますので、簡単にこの4つぐらいを覚えておけば、ある程度のドキュメントが作れるのではないかと思っています。ここまでがMarkdownの基本ということで4つ紹介しました。

　次にRチャンクの基本ということで、「Rチャンクって何ぞや？」ということなのですが、基本的にはR Markdownのドキュメント内にRのコードを埋め込んで実行できることということで、ただ単にR Studioで動かすだけでは動かないのです。先ほどまではコードだけ入力すれば反映するのですが、ドキュメント形式で見せますので、よりきれいに見せるためにRチャンクという独自のものを設定した上で、そのコードを反映させるかどうかというのを決定する権利があります。それを決定することによって、例えばですがこのような感じの図表をMarkdown上で表記できるという形になります。

00：15：03

　このRチャンクがすごく難しいのです。いろいろなものがあります。難しいというよりは、どれを選んだらいいかというのが分かりません。先ほど、皆さまにやっていただいたR Markdown上でも、例えばサマリーカールズという上にRカールズと書いたと思うのですが、ここの部分のRチャンクをどう設定していくかということになります。基本的にはRのコードは変わりませんが、このチャンクラベルというのを設定することによって、どのドキュメントへ、どのコードをどういうふうに反映させるかが変わってきます。

　私個人のお勧めというかいろいろやってきた経験の中では、このecho=TRUEとは何かというと、下に後で表記しますが、基本的にはチャンク内のコードをドキュメントに表示させるかどうかを指定するチャンクオプションになっています。その後にwarning=FALSE、 message=FALSEと書いてありますけれども、この3つは私が個人的に使っているものになります。皆さんに今回お渡しした臨床情報QCの中にも、同じようにチャンクオプションが書いてあるのですけれども、このwarningとmessageは何かというと、警告をするというのと、あとは分析をした後にメッセージが追加されるような場合があるのです。●のパッケージでも出てくるのですが、それが出てきてもいいのですが、それがすごく多くなると結構見づらいドキュメントになってしまいます。一度は警告画面やメッセージを確認する必要がありますが、特に問題がなければFALSEにしておいたほうがきれいなドキュメントになりますので、やってみるといいのではないかと思います。

　それ以外のデフォルトで決まっているのはこのような感じです。これは先ほど紹介した書籍のところを抜粋していますが、このような感じでechoの場合にはTRUEとか、includeがTRUEということでもともと決まっています。ですから別にechoもTRUEにしなくてもいいのですが、念のため私の中ではとりあえず自分自身が作ったコードをドキュメントに反映したいと思っていますので、それはきちんとTRUEとして出し……。

00：17：02

00：00：00

上野：あとはデフォルトで決まっていない部分に関して変更したい点というのをプラスで付けてあげるということになります。ですから、実際に皆さまが分析していく過程において、その分析結果をどういうふうに反映したいのか。例えば隠したい分析というのもあると思うのです。細かい分析なので、それを反映してしまうとすごくドキュメントをとってしまいますという場合には、この分析のここだけはechoはFALSEにしてくださいとすればきれいになります。という形で、これは本当に好みや●の仕方によって変わりますので、いろいろやっていく中でこのチャンクオプションというのを使い分けてもらいたいと思います。

　ということですが、ここまで少し駆け足になりましたが、簡単にR Markdown実践編ということで、Rからの流れでどういうふうにR Markdownを使うかというのを紹介しました。

　R Markdownの使用上の注意点ということで3つあります。まずはRを学んでいただければと思います。そもそもRが分からなければ、R Markdownできれいにまとめるということができません。まずはRを理解してRコードの内容を理解した上で、ではこれを他の先生と共有したい、何かで発表したいとなった場合にはR Markdownとしてまとめるといいのではないかと思います。

　ただ、私個人として、Rを今まで学んでこなかった中で急にR Markdownから学んだのです。メリットとしてはR Markdownでまとめていったほうが、このコードがどのような意味合いで、どういうふうになったらどうなるかというのが全部自分でMarkdown形式でまとめることができますので、急に学ぶ中でもR Markdownを先にやって、その中でRのコードなどを学んでいくのも1つなのではないかと思います。同時並行で学んでいっても結構だと思いますが、Rを知らないとそもそもR Markdownで分析がいきなりできるかというとできませんので、その点に関しては注意してください。

　Rと少し違うのは、結構警告やエラーがたくさん出ます。これはなぜかというと、R Markdownの場合には、いろいろなパッケージを設定した中で1つのファイルとしてまとめます。ファイルが競合してしまって分析が途中でエラーを起こしてしまうことが、3番目に関係しているところですけれども、例えばサイコパッケージとggプロットは結構相性が悪いようで、先にサイコをやってからggプロットを付けると競合してしまって、R Markdownが出てこないというケースがあるのです。いわゆるマスクと言われる現象なのですが、その場合にはどちらを先にするかなど、そういう競合の部分が出てきてしまいますので、それはやっていく中でいろいろ試すといいのではないかと思います。

　ですから、警告やエラーなどが起きるのですけれども、大体の間違いはRコードが多いので、Rコードを確認してHTML形式のKnitで押すと大体最後はしっかり出てきますので、それを頻回に確認していくといいと思います。

　ということで、R Markdownを学ぶ、さらにR Markdownの実践編ということで簡単に紹介しました。ここで大体3分の2は終了するわけですが、ここまでで質問はありますか。大丈夫そうですか。今ちょうど50分程度終わりましたので、大体残り30分間で、最後の臨床情報QCということで私自身が作成しましたファイルを使って説明に進みたいと思います。

　それでは、最後の章になりますが、私自身が事前にお送りしましたR Markdown、こちらのほうを立ち上げていただければと思います。こちらのR Markdownを事前にお送りしましたが問題なく動きましたか。問題があった先生はおられますか。具体的にどういう問題がありましたか。

Ａ：KnitでHTMLが出てくれなくて。

上野：ありがとうございます。あとでチューターの方が確認します。他の先生はおられますか。もしいましたらチューターの方が確認しに行きますので。あとは大丈夫そうですか。ありがとうございます。

　基本的にKnitで出てこない場合には、事前にパッケージをダウンロードしていなかった場合や、私はWindowsで作ったのですがMacとWindowsは若干違うみたいなのです。あとは初めてインストールする場合などでも変わってきますので、そういう点で変わってきますので確認……あとは警告やエラーで何が足りないというのが出てきますので、その警告やエラーに従って進めていただければ大丈夫かなと思います。

　それでは、最後の臨床情報Quality Checkということです。皆さんはデータの管理はどうしていますかということで、基本的に私の場合は調査研究で、さらに臨床情報のデータを扱って整理しているということですがこのような感じでデータが出てきます。

00：05：02

　例えばCSV形式やエクセル形式のデータがあります。このデータを急に使うということはないですよね。急に使ってハッピーな結果が出ればいいですけれども、大体ハッピーな結果は出ませんし、さらにデータに間違いがあるということも結構多いのではないかと思います。

　さらに、例えば1つの施設だけで採っていれば問題が少ないかもしれませんが、いろいろな施設からデータを採ってくる場合には回答の入力方法が変わってきます。ですから、誤った入力がされていて問題が起きるという可能性も出てきます。特に最後の論文に出す直前でデータの間違いに気付くというケースもあるのではないかと思います。

　ですから、事前にやることとしてはこのQuality Checkということで、データ前処理が必要になります。

　特にQuality Checkをやるものに関しましてはいろいろな種類があるのですが、まずはデータに誤りがあるかという確認をする作業とか、データに欠損があるとか、あとはデータに外れ値があるといった、大枠でこの3つを確認する作業があると思います。この3つを確認する作業をした後に実際にデータ解析をしていくということが大切になります。

　特にデータに誤りがあるというのは、いろいろな施設で採っている場合にはなかなか分かりづらい部分があると思います。それを実際にR Markdownを使って可視化することによって、1人ではもしかしたら気付かないかもしれないけれども、いろいろな先生方と共有していく中で、特にドキュメント形式で出てきますので見やすくもあり、さらに実際に間違いも発見しやすくなるという形になります。

　では、今回私が紹介する臨床情報QCということで、いろいろな方法があるのですが簡単なプログラムを紹介したいと思います。

　まずは、先ほどと同じようにCSV形式でデータを読み込んでいただいて、その後に欠損値の確認方法、外れ値の確認方法という形で、この3つのそれぞれの手続きを進めていきたいと思います。

　その辺りに使うパッケージはこのような感じになります。これは別に欠損値だからといってこのパッケージを全部使う必要はなくて、先生方の研究や目的に応じてこれもアレンジしていってもらいたいと思っています。

　それでは、実際に臨床情報QCのR Markdownを立ち上げていただいたと思いますので、中身を確認していきたいと思います。

　ちなみに今回使った臨床情報Quality Checkのプログラムのデータの関しましては、ありきたりなデータなのですが、とあるブログで書かれたデータが結構使いやすかったので、そのブログのデータをさらに私のほうで加工して使ったという形です。原本からもほぼ形がないような形なのですが、そちらのほうを引用しましたので、もし興味がある方がいましたらそちらのほうに飛んでいただけるといいのではないかと思います。

　実際に動かしていきたいと思いますが、先ほどお話ししたフロントマター部分に関して先に説明をしたいと思います。

　こちらがフロントマター部分になるのですが、先ほどのデフォルトではこのhtml\_documentまでしか書かれていないのですが、この下の部分に設定をしていくという形になります。現時点でこれを1回Knitしてもらいたいのですが、Knitすると100％になってHTML形式で出てくると思いますが、ちょっとその画面を覚えておいてください。このような感じです。横にサイドメニューバーがあって、下に行くとこのような感じでコードが入力されて、それでデータが実際にアウトプットされてという画面になっています。

　さらに、データの確認となったときに、この見出しレベルも2つ下げていますので、このような感じで少し段落が下がっています。こういう形で見やすくなっていると思いますけれども、これを例えばtocというのが最初にあります。これが今trueになっているのですがfalseにしてもらえますか。それでもう一度Knitしていただければと思います。

　そうすると、なんとサイドメニューバーがなくなるのです。これはどうですか。結構見づらくなってしまいますよね。特に臨床情報QCの場合には量が多いと思うので、サイドメニューバーでいろいろな見出しを作ってあげて、自分がチェックしたい項目を確認するということが大切になりますので、このtocというのがtrueになっただけでもだいぶ違うと思います。

00：10：04

　ですから、ぜひ使う場合には、このtocに関してはtrueのままにしてください。ということなので、このfalseはなくしてもらってtrueにしてください。

　では、この下の部分です。3と書いてありますが、これは見出しレベルです。見出しを何個まで仮定するかによります。私の場合は今回のR Markdownの場合にはシャープが1個ではすごく強調されてしまうのです。結構大きいフォントで表記されるので、私は個人的に嫌だなというのがありますので2個目から始めています。2個目から始めて3個目、その次に下げるときには3個にしているという形になりますので、それ以上の4個というのは私は設定していません。ですからここの部分は3と入力されています。見出しレベルをもっとどんどん下げたいという人がいたら、最大4ぐらいだと思いますけれども、5は多分できなかったと思うのですが下げていただけるといいのではないかと思います。

　こちらのfloatと書いてあるところはtrueになっているのですが、先ほど1点説明に誤りがありましたので訂正したいと思います。奥側のスライドを見ていただきたいのですが、先ほどtocに関しましては、ここに「文書の先頭にコンテンツ一覧表を表示」と書いてあります。これは下の部分のサイドメニューは今からやるところになるのですが、こちらが入力されてないとそもそも一覧表が表記されないということになります。ですから、今やるサイドメニューのところに関してtrueになっていると思いますが、このfloatの部分をfalseにしてもらうとサイドメニューが消えて、見出しだけがばっと出てきますので、それでもう一度Knitしていただけますか。

　そうするとこのような感じで出てきます。それでも見やすいかどうか分からないですけれども、一応メニューバーなので押せば飛べるのですが、でも横に貼ったほうが見やすいと思います。特に下まで行ったときにこれは戻ることができません。メニューバーが消えてしまいます。ですから、先ほどお話ししたようにtocをtrueして、さらにこのfloat部分にもtrueしてあげたほうが、横で表示されて、かつ見出しの一覧表も出ますのでやっていただけるといいと思います。ですから、ここに関してもfalseではなくてtrueにしておきます。

　この下の〓dfプリント〓に関しては〓paged〓となっているのですが、これはものによって設定するかしないか悩ましいところです。何かというと、既存で作られているRコードを読み●そのR Markdownで表記されたときにデザイン性が問われるのです。特に、例えば最初の●のような感じでサマリー関数を使ったり〓ヘッド〓を使ったときに、きれいに出したいときにはこのpagedというのを設定しておくことによって、R Markdown上できれいに表示することができます。

　細かいことに関しては全部書籍に投げてしまって申し訳ないのですけれども、自分自身の好みで私自身はこれを付けておいたほうがきれいに見えるので一応pagedとしているのですが、もし皆さまの中でこれは少し汚いなというがあれば違うのに変更していただけるといいかと思います。

　そして10番目はkeed meということで、これを押すことによってMarkdownが勝手に保存されます。基本的には押すと勝手に実際の自分のファイルの中にデータの分析結果が保存されるのですが、それをtrueにするのかfalseにするのかという形で設定ができます。基本的に私はtrueにしてあるのですが、上書きをしたくない場合はfalseにしておくといいと思います。

　最後のthemeというのは、これは●みたいな形で私は作っているのですが、R Markdownのデザインってとても多いのです。今のデザインはこういうデザインになっているのですが、これを例えばunitedというのにしてもらっていいですか。スペルに関しては奥にも書いてありますが、theme unitedと入力していただいてKnitを押してください。

00：15：00

　このような感じで赤文字になるのです。結構私はこの〓フォント体〓が好きなのですが、これも完全に好みです。R Markdownにはいろいろなデザインがありますので、それこそGoogleで調べればいろいろな形があります。私がいつも使っているのは、この〓フラットリー〓は結構お勧めですごくきれいなのですが、unitedも赤文字形式でみやすいです。あとフォントの形も変わります。自分のデザインをいろいろと試してみて、自分の好みに合わせてみるといいのではないかと思います。

　ちなみに、最後にjournalをやってみますか。themeのところにjournalと打っていただけますか。journalとフラットリーは近いのであまり変わりはないかもしれませんが、それでKnitを押してください。ちょっとフォント体が変わります。journalとunitedが近いですね。こういう形でデザイン性が少し変わります。これはRを学んだ後のデザインを決めるという形になりますので、これも好き好きで設定していただければと思います。私はフラットリーが好きなので●に戻します。

　ここがフロントマター部分の説明になります。それでは、QC情報に移っていきますが、ここまででご質問がある先生はいますか。大丈夫そうですか。ありがとうございます。

　それではQCのところに移っていきます。基本的には時間の関係もありまして、入力して何か分析するというわけではなくて、私が作ったQCを確認しながら説明……。

00：17：02

00：00：00

上野：……したいなと思います。お昼休みの後なので少し眠くなってしまうかもしれませんが、確認しながら自分の研究にアレンジできそうだなというところはどんどん使っていただければと思います。

　ダウンロードパッケージということで事前確認にも書いたのですが、基本的にはRコードで入力するか、それともPackagesを選択して入力するかで、2つインストールする方法がありますので皆さまが使いやすいほうを使ってインストールしていただければと思います。ただ、Macはバージョンによってはこの〓アームカール〓というパッケージが必要になるケースがありますので、もしご自身が使われていて他の人にも何か使ってほしいとなった場合には、そちらのパッケージが必要な場合がありますのでその点に関してもご理解ください。

　まず、CSVデータの取り込みということで、基本的にはもう皆さんは分析が実行できていますので下のコードを確認していきたいと思いますが、1番目はデータの確認ということで、Rチャンクに関してもecho=TRUE, warning=FALSE, message=FALSEになっています。

　Rがいいところはこの後に、例えば下にDTパッケージと書いてあるところがあると思いますがコメントが付けられるのです。ただ、Rチャンクに関してはコメントが付けられませんのでご容赦ください。

　ですから、このパッケージが何を意味しているかのようなことはコメントとシャープを付ければよいのですが、チャンク内の意味合いに関してはシャープマークで付けられませんので、ご自身でこの下に例えば箇条書き設定でこのechoは◯◯のように書いておくと後々忘れないのではないかと思います。

　そうしておくと、後でパッケージをKnitした後に箇条書きでCSVデータを取り込むと下に表記されますのでメモ書きでも使えるのです。例えばRコードでR Studioで動かすとなったときに、わざわざシャープマークを作ってそこで入力するか、後はテキストファイルを他でWordなどで入力するケースがあるかもしれませんが、もうR Markdownはこのような感じでメモ書きも下で作れますので、これは日本語でも入力できますので、こういうふうに下で箇条書き設定をするといいかもしれません。私は結構よく使っています。

　最初にファイルの中に入っているRデータを読み込んでもらうという形で一応関数が設定されています。細かい話に関しては、もう皆さまのほうに配布資料で提示していますが、一応奥側のスライドをご覧いただければと思います。

　CSVデータを読み取る場合にはread.csv関数を使うのですが、このような感じで設定をして実際に読み込むという形になります。

　先ほどのように細かい設定に関して、もともと私はこちらのほうで発表する前にはこのまま下に文章を書いていました。箇条書きで書いてある文章をこの下にそのままコピペして箇条書きで設定していました。ですから、これを使う場合には別にパワポで説明しながらこれでも良かったですけれども、こちらの下に箇条書きで書いておくことによって二重で確認する必要がありませんよね。ですから、メモ書きとして奥のCSVデータの取り込みというスライドの内容を書いておくことによって、後々、自分自身も忘れずにできるのではないかということで前までやっていたのですが、一応今回はチュートリアルということだったのでこちらのほうは消させていただきましたが、皆さまが使いやすいように●使っていただければと思います。

　CSVデータを取り込んでいただくと、その後にlibrary(DT)ということでDTパッケージを読み込みます。

　すみません、いったん戻るのですが、こちらのデータの読み込みということで先生方の中には1個ずつのコードを一応確認したいという先生がいるかもしれません。一応Knit to HTMLでさらに分析結果を確認できるのですけれども、その前にどこにコードに間違いがあるかとか、一応中身を確認しながら並行してR MarkdownとRをどちらもやりたいという先生もいるかもしれません。やり方はすごく簡単です。こちらのほうにコードが入力されていますので、これをcontrol、enterを押していただくと、下にこのような感じで入力で反映されます。ワーニングメッセージが出てくるのですが、特に問題は何もありませんので心配しないでください。こういうふうにワーニングメッセージがたくさんR Markdown上で表示されてしまうので、私はwarningやmessageはfalseにしているというのがあります。特にこの問題が起きているからといってデータに何か問題があるというワーニングメッセージではありませんのでご安心ください。ですから、それをcontrol、enterを押せば下に表記されますので、そうするとこれを読み込んだという話になります。

00：05：08

　次にDTパッケージを使うわけですけれども、DTパッケージというのはR Markdownでよく使われるパッケージです。DTパッケージのdatatable関数というのを使うことによってたくさんあるデータを検索することができます。先ほど作っていただいたものを1回見ていただきたいのですが、多分皆さんは1回Knitして消してしまっていると思いますので、もう一度Knitしていただいてもよろしいですか。多分Knitしていただいて出た画面のHTMLを見ていただきたいのですけれども、量が多くなると時間がかかってしまうのは弱点ではないのですけれども見づらいかなと思います。

　先ほどのCSVデータを取り込むというのを押したことによって、箇条書きのところにechoは◯◯、warningは◯◯のような形で反映がされます。

　先ほど話したように、CSVデータの確認ということでDTパッケージを使ったわけですが、その中のdatatable関数を使うことによって、こんな感じでデータが下に表記されます。

　このdatatable関数のいいところは、サーチということでデータを確認することができますし、これ以外の使い方もたくさんあります。ですから、DTパッケージを使用することでさまざまな表も作れますので、この表はちょっと嫌だなという先生がおられましたら、下を参考にしながら違う表を作ってみるといいのではないかと思います。

　一応、datatable上で出てくるのはこのような感じの画面が出てきます。行ったり来たりして申し訳ないのですが奥側のスライド画面をご覧ください。今回に関しては10名のデータになります。性別、年齢、身長、体重、最高血圧、血糖値というデータになっていますが、こちらのデータを使ってやるわけですが、もう既にブランクの部分が何個かあります。このブランクって何かというと欠損値に値します。この時点で欠損がどこにあるかというのが把握できると思いますが、それをより細かく次の章で見ていきたいと思います。

　ここまでがCSVデータの取り込みに関してになりますが、ご質問はありますか。大丈夫ですか。ありがとうございます。

　それでは次、欠損値の確認方法に移りたいと思います。

　欠損値の確認方法はたくさんあるのですけれども、まずはmiceのパッケージを使ってやっていきたいと思います。libraryでこれを〓実施〓することによって可視化できるのですけれども、このような感じでこのパッケージを利用した後のデータ解析の結果が出てきます。

　こちらのデータ解析の結果の解釈ですが、このような感じで使ったデータで、先ほどブランクが5カ所ぐらいありました。そのブランクの部分に関して表記されるのですけれども、このような感じで一番左上に関しましては、全ての項目で欠損なしが6名、下のBS欠損者が1名、Weight欠損者が2名、SBP欠損者が2名、BSとSBP欠損者が2名ということでそれぞれの欄が見られます。赤色が欠損が起きている欄になります。

　という形で、実際に先ほどのmiceパッケージを使うことによって、文字データだけでも分かりやすいのですけれども、取りあえずざっと確認したいとなった場合にはこちらのパッケージを使うと可視化できますので結構見やすいのではないかと思います。

　ただ、弱点としては件数が20や30あると結構見づらい表になってしまいますので、大体10～20内ぐらいがお勧めかなと思います。ですから、この組み合わせで今回分析して、論文をまとめるといった件数だけに絞ってやったほうがいいかもしれません。

　あとは、このmiceパッケージを使うとmd.pairsという関数も使えます。このmd.pairsでは何ができるかというと、2項目の組み合わせごとに欠損の部分が表記されます。2項目なので使う場面は少ないと思いますが一応ご紹介したいと思います。

　基本的には、こちらで皆さんに紹介しているデータのコードはとても簡単で、関数を表示して、その後に括弧でデータ名を出せば、あとはもうそれで分析が可能になりますので、すぐ実行することは可能かと思います。

　今回のデータに関して、例えばAgeとWeightを例にしますとこんな感じで出てくるわけですけれども、実際にHTMLではこの下のmd.pairsになったときにrrとかrmとかmrとかmmといった形でそれぞれいろいろな組み合わせがあります。

00：10：07

　これは何を意味しているかというと、rrに関してはともに観察されているものということで、例えばAgeとWeightだったら8名が観察されていますよ、Ageのみ欠損は1人もいませんよ、Weightのみの欠損者は2人いましたよと。あとはAgeとWeightの両方が欠損したという人はいませんよという形が見えます。

　基本的にはrは非欠損値を表し、mは欠損値。先頭が行で、後頭が列というような形で表記されていますので、もし2項目を挙げてその値を見たいというときにはこのmd.pairs関数を使うといいのではないかと思います。

　次にlibrary(VIM)というパッケージがあります。先ほどのパッケージと何が違うのかというと、あまり変わりはありません。これも好き好きです。皆さんにはいろいろな欠損値の可視化の方法を提示するということでこちらのパッケージも出しました。

　こちらのパッケージのメリットとしては、棒グラフとパターンの2種類で可視化することができます。md.pairsだったら文字データだけで分かりづらいという話になってしまうのですが、このような感じでVIMを使うとそれぞれの欠損者の値が見えます。

　例えば、ID、Sex、Age、Heightのところまでは欠損値がないので赤色の棒グラフが書かれていません。ただ、WeightとSBPとBSに関しては欠損値がありますので棒グラフが描かれています。さらに、md.pairsで可視化した2項目の組み合わせ間での人数比率といったようなものもそれぞれ出るのが分かります。

　ですから、VIMを使うのか、先ほどのmiceパッケージを使うかというのは完全に好みになりますので、使う用途に合わせてやっていただけるといいと思います。

　さらに、このVIMパッケージを使うとpbox関数というのが使えます。このpbox関数というのは、いわゆるボックスプロットを作ってくれるのですが、欠損値だからといって消していいかどうかを考える判断の1つになります。欠損値ってブランクなので消してしまっていいと判断する人が多いと思うのですけれども、私は心理出身なので、心理の中では●代入表で何かの値を代入した上で使ったほうがいいとか、あとは人数が限られていて結構もったいないよねという話もあると思うのです。例えば精神疾患のデータで、3名ぐらい消えてしまうけれども3名いたらもっといろいろな分析ができるのに、という話もあるかもしれません。

　そういう場合には、まずは欠損値の傾向を確認しましょうという形になります。実際にスライドのほうを確認していただきたいのですが、例えばこれは身長と体重のデータになっていますが、体重の中に欠損値が今回ありました。欠損値があるものが赤で表記されていて、ブルーが欠損値がないものになっています。それぞれどれだけ変わりがあるかという話なのですが、縦軸が身長なのであまり信ぴょう性はないかもしれませんが、実際Weightの中での欠損値のありなしでは身長はあまり変わらないけれども、SBPの中では少し差が出てきます。

　ですので、基本的に身長と体重だけの組み合わせだったら分析データとして使ってもいいかもしれないけれども、体重が欠損値の場合、SBPの値は割と高めにでるから、きちんと欠損値を消すような対策を考えないと、もしかしたら分析結果が変わってくる可能性が出てくるかもしれません。

　ですから、ただ単に欠損値だから消すという作業ではなく、このようにまずは可視化して、欠損値の傾向を確認した上で分析に移るといいのではないかと思います。

　ここまでは可視化ということですが、もう少し欠損値の具体的なものを確認したい、IDを誰なのかというのを確認したい場合、これはRでもできるのですが、このような感じでis.na関数というのを使うことによって、今回のデータの実際の欠損の傾向をデータ上から確認することができます。これはパッケージのダウンロードは必要ありません。

　確認しますとこのような感じで見られます。データが10名で少ないのですが、TRUEが欠損です。FALSEが欠損でないということで、今回の場合にはTRUEが欠損になりますので5名ほど欠損がいましたという話になります。

　では、そのデータに関して、欠損があった対象者は誰だったのかということで、下に全体データが出てきます。

00：15：00

　このデータを確認することによって、その●に当たることができます。ただ、これの場合では変数が多いとどんどん分かりづらくなるのです。特に欠損が1個、2個とか対象者が数十名だったら割と簡単に検索できますが、これが1,000名や2,000名などすごい数になってしまうと、なかなか欠損を確認するまで時間がかかってしまいます。

　さらにヒューマンエラーが起きやすくなりますので、完全にこの欠損値のIDは誰だということを確認したい場合には、次のdplyr、これは小池先生のほうでもパッケージ紹介がありました。こちらのフィルター関数を使うことによって特定のIDを〓算出〓することができます。

　具体的にやりますと、こちらのパッケージの分析を行いますと、このような感じで実際に今回は体重のところで欠損がありましたが、体重の欠損のIDは誰だということで、こちらのfilter関数を使ってRdataでis.na関数を使い、最後に＄マークでIDをくくると、6番と10番が今回の体重の欠損者として見つかりましたと。

　では次に、SBPは誰だというと2番と4番、そしてBSだと誰だということで4番が出てきます。という形で、人数が多くても特定のIDだけ知りたいとなった場合には、このdplyrを使って以下の関数を設定してあげることで特定のIDを使うことができます。

　ですから、欠損を確認する方法としては、先ほどお話ししたように、まずは可視化してあげて、そこからより文字列から確認し、最後に特定のIDを見つけてあげることによっていろいろな対応策ができるのではないかと思います。さらに、その中でpbox関数を使ってあげることによって、本当にその欠損値を消すのか、何か当てはめてあげるのか、それを考える機会になるのではないかと思います。結構、欠損値の……。

00：17：02

00：00：00

上野：取り扱いに関しては昨今、すごく問題になっていますので、ぜひ可視化して数値を確認しながら、皆さまの中で判断して使ってもらいたいと思っています。

　それでは最後、外れ値の確認方法に移りたいと思います。

　外れ値ですけれども、これはたくさんあります。よく使われるのは度数分布表ですけれども、データがたくさんある場合は結構見づらくなってしまうことが多くなります。どうするかというと、summarytoolsパッケージというのがあります。こちらも小池先生のほうでご紹介がありましたが、これをR Markdownではとても有能に使うことができます。

　実際に使ってもらうと、次のページに出ているような形できれいに出てくるのですが、1個だけ注意点があります。外れ値の確認方法の中でsummarytoolsパッケージを使った場合には、通常Rコード以外にもいろいろなものを設定しなければなりません。

　R Markdownのこちらのところで関数を確認したいのですが、Rチャンクのところで先ほどまではecho=TRUE, warning=FALSE, message=FALSEだけを設定していたのですが、その後にplain.ascii=FALSEという設定をしています。これを設定しておかないとR Markdown上で表が作成できないということに私自身が気付きまして、これを必ずsummarytoolsパッケージを使った場合だけプラスで付けてあげてください。

　こちらです。前側にコードのところのRチャンクに表記されていますが、ここの部分、必ずsummarytoolsを使う場合には反映をよろしくお願いします。

　さらに、dfSummaryだけでは結果が算出されませんので、先ほどのdfSummaryを使って小池先生のほうではご紹介がありましたが、プラスprint関数で間を挟んであげて、method='render'のような形でこのような関数を描かないと出てきません。この間にあるこれは0.50と指定されているのですが、これはグラフ幅です。これをとても大きくしてしまうとすごく見づらくなりますので、大体0.50ぐらいがいいのではないかなと、いろいろな経験をした中ではありました。

　このsummarytoolsパッケージは非常に有用ですので、こちらも以下のサイトにいろいろな使い方が書いてありますので、ぜひ見ていただければと思います。

　ということで、これを使うと何が見られるかという話ですが、先ほどは2チャンネルみたいな形で出てきたという話でしたが、これは可視化がすごくきれいに出てきます。ここは文字列しか赤でくくっていませんが、その隣に度数分布表が描かれていると思います。

　実際にいろいろなデータを確認する上で、度数分布表などは基礎統計ですごく大切になります。それを確認する上で、いつもsummarytoolsパッケージを使ってあげるといいのではないかと思います。

　この確認をする中で何が問題か、なんとなく見えてくると思います。例えばHeightの値を見てください。これは年齢が8歳～11歳のデータなのですが、身長のところを見てもらうと178というのが飛び抜けて1人だけいるのです。下は70というのがあります。

　果たして8歳～11歳の中で70と178というのが正しい値なのかというのを少し考えていただくきっかけになりますし、その後のWeightでも10というのは絶対におかしいですよね。どういう組み合わせで10になっているか分かりませんが、仮に一番低い組み合わせでも70センチ、10キロってないですよね。その10キロが本当に正しいのかというのを考えることになりますし、ざっと度数分布表を描いてあげることによって間違いに気付くことができます。

　先ほど話したSBPのような最高血圧に関しても150ってちょっと診療所に問題があるのではないかとか、そういう話になりますし、BSの値に関しても血糖値が170って少しやばいよねという話になってしまうので、それをざっと確認するのがsummarytoolsでできますので、まずこの値を確認して、しかもパーセンテージも書いてありますのでそれを確認していってあげるという形になります。

　これ以外にもやり方としては、皆さんもご存じかもしれませんがBoxplotを描いてあげてその中で確認する作業があります。

　例えば今回の度数分布表だけでは、実際に外れ値かどうかって分からないですよね。それを実際にBoxplotを使ってあげることによってより可視化した上で、どのIDが間違っているのかというのを確認することができます。

　以下、コードになります。

00：04：59

　これはggplot2というパッケージを使ってやるのですが、詳しい説明をすると長くなってしまうのですが、このような感じでggplotの関数でキャンパスを作って、その後でboxplotの設定やラベルの設定や、IDを表記するかといった設定ですけれども、これは最初は結構苦労しました。私は苦肉の策でこの関数を設定しているのですが大体ここまでで終わりなのです。本当ならgeom\_textまでで●関数で終わって、そのまま表記すればIDを表記できるのですけれども、ここまでだとIDを表記したときにBoxplot上でくっきり出てこないのです。それで、その後に私は何かいいのはないかということで、それこそ先ほども言いましたが、絶対に世界中の誰か1人ぐらいやっているだろうというので調べたらやっている人がいましたので、このggrepelというパッケージを利用することによって、よりIDをくっきり表示させることができるというパッケージを利用してやっていました。

　その理由としては、見てもらえれば分かるのですが、このような感じで一応図ではggplotで描いてIDを可視化しているわけですけれども、こちらに関しては、今、四角で囲って1番や7番と書いてあると思いますが、あちらのggrepelパッケージでないと全くその四角で囲ったものがない、ただ数字だけが書いてある状況になります。

　私が違うデータの臨床情報QCをしたときに、それが300人、400人もいたのです。300人、400人をこれで表記するとなると、もう被ってしまって何のIDか全然分からないという現象が起きてしまいました。

　そこで使う1つとしてこちらのパッケージを利用することによって、そのIDを引っ張ってもらい、さらにこのような感じできれいに出していただきました。ここら辺がこういうふうに被ってしまいますよね。被ってしまうので少し引っ張って6番、2番、4番という形で、もともとない場合にこれは4番だけの表記とか、6番とか9番になってしまうのですけれども、それを引っ張ってあげて●形になります。

　ただ、これを200人、300人やったときに、小池先生の指摘もあったのですが、最後はクジャクのようになります。すごく引っ張られてとても見づらくなってしまうのです。ここの部分が今は10人なので結構きれいに見えますが、300人がここの中にあるとごちゃごちゃして、しかもこちらの部分もごちゃごちゃしてしまうので、人数が多いときには一応IDは見られるのですけれども少し改良の余地はあるかと思います。人数が少ないデータをQCする場合にはこちらが利用できますが、人数が多い場合には一応確認できますが、皆さまのほうでも●しながら使っていただけるといいのではないかと思います。

　特に、このggplotはすごく使い勝手がよろしいですので、ggplotに関する書籍のご紹介ということでいろいろな本があります。真ん中の書籍に関しては先ほどお話ししたR Markdownの入門編も書いてありますし、R Studioの使い方や、あとは実際のggplotの使い方に関してもあります。別に私はこの中に知り合いの先生がいるわけでもなく、回し者でもないですけれども、すごく分かりやすい本ではないかと思います。

　それ以外で、もっとggplotを使って描いて、今回のような可視化をもっと先に進めていきたいという方がいたら、両隣の本も参考になりますのでぜひともやっていただきたいと思っています。

　1時間半という予定でした。少し間延びしてしまったわけですけれども、一応こちらで私のほうの紹介は終わりにしたいと思いますが、何かご質問等ありますか。大丈夫そうですか。

　特にこのR Markdownに関しては、やはりRを学ばないといけないというのは皆さまのほうでもなんとなく分かっていただけたのではないかという点と、かつ、R Markdownを使うことによってせっかく学べたものをもっときれいに見せて、さらにいろいろな研究者で共有するときに使い勝手がいいのではないかと、少しでも思っていただければうれしく思います。

　今回、参考にした文献やサイトは結構いろいろなものを参考にしながらやってきました。R Markdownに関してはまだまだそれほど普及されていませんので、なかなか情報が少ないですけれども、特に海外の方だったら多いですし、日本の中にも先ほどの紹介書籍や、あとはちょこちょこウェブサイト上でも紹介がありますので、私の中で学び切れていないとか、こういうことをやりたかったという人がいましたら、ぜひともググってもらえればいろいろな形で出てくると思いますので利用してみてください。

　あとは使用したパッケージはこちらになります。

　それでは、本日のまとめですけれども、一応皆さまには今回R Markdownというソフトの使用方法を学んでもらいました。

00：10：04

　さらに臨床情報QCのプログラムということで、外れ値の簡単な見方、そして欠損値の見方という2種類をやってきたわけですけれども、あくまでもこれは参考例に過ぎません。自身の研究で実践しまして、最後にプログラムをアレンジしてもらいたいと思っています。

　まずは皆さまに配ったコードに関していろいろな書き換えをしながら、自身の研究に使えるアレンジをして使えるのではないかと思いますので、いろいろ試して見てください。

　以上で私のチュートリアルは終了となります。ありがとうございました。