

RStudio の利用ガイド

2022 年 5 月

1. 本ガイドにつきまして

本ガイドは、4つの部分からできています。

- 2.では、R、RStudio の概要をお示ししております。
- 3.では、R というソフトウェアをインストールする方法を説明しております。
- 4.では、RStudio というソフトウェアをインストールする方法を説明しております。
- 5.では、RStudio を用いて匿名化を行う方法を説明しております。

作業手順を一つずつ説明しているため、やや長くなっておりますが、このガイド通りに作業していただければ、プログラミングの知識などは一切不要となっております。Excel を用いるよりも作業が楽かと思しますので、ぜひご検討ください。

作業時間の目安ですが、

- R のインストール：5 分
- RStudio のインストール：5 分
- 匿名化：自治体によって 10 分～数時間（※）

を見込んでおります（お使いの PC の性能に応じてこれより長くなる可能性もございます。）。

※ プログラムが動作している間、別の作業をしながらお待ちいただいて構いません。

※ 当方の PC（プロセッサ：Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz 2.30 GHz、実装 RAM：16.0 GB）で、人口約 2 万の自治体のデータ（データ項目数：約 600）10 年分を処理したところ、約 10 分で処理が完了しました。必ずしも処理時間が人口やデータ項目数、年数に比例するわけではありませんが、参考にしてください。

2. R、RStudio の概要

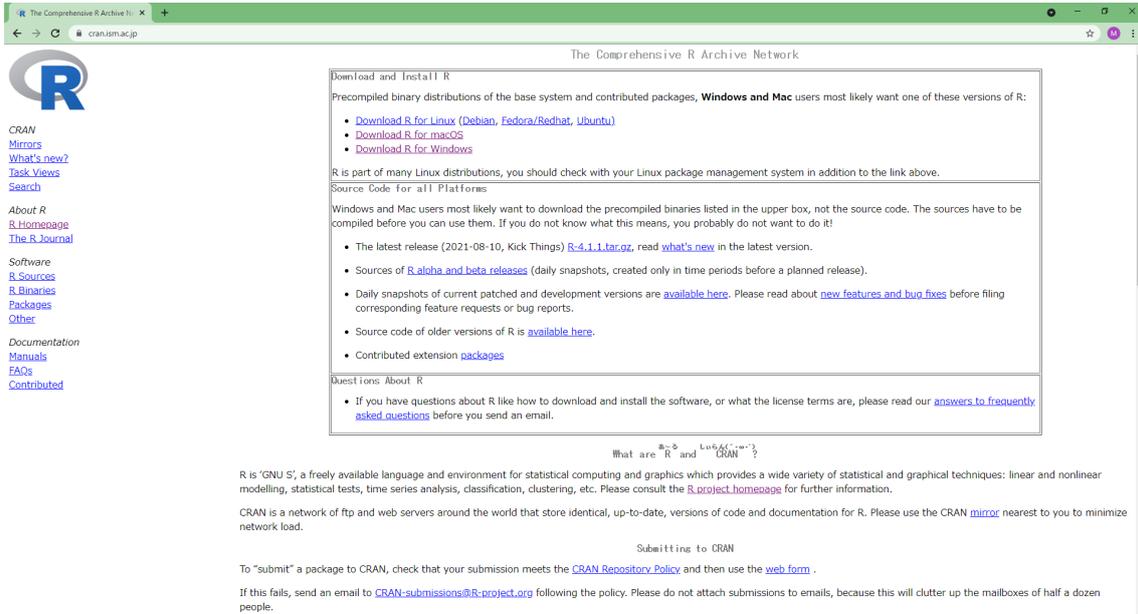
R は、データを扱う機能を持ったソフトウェアです。データを整理して利用しやすい形にしたり、データを分析したりする際に用います。様々な機能が無償で利用できるため、研究機関や教育機関、民間企業など、多くの組織で活用されております。

RStudio は、R をより使いやすくするためのソフトウェアです。コードや分析結果をより見やすく表示させることができ、大変便利です。そのため、今回もメインで使うのは RStudio の方となっております。ただし、R をインストールしないと RStudio は動作しないので、R と RStudio の両方をインストールしていただきます。

3. R のインストール

① 以下のサイト(統計数理研究所が提供しております)にアクセスします。

i. <https://cran.ism.ac.jp/>



The screenshot shows the CRAN website with the 'Download and Install R' section highlighted. The page title is 'The Comprehensive R Archive Network'. The main content area is titled 'Download and Install R' and contains the following text:

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2021-08-10, Kick Things) [R-4.1.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alphas and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and GNU?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

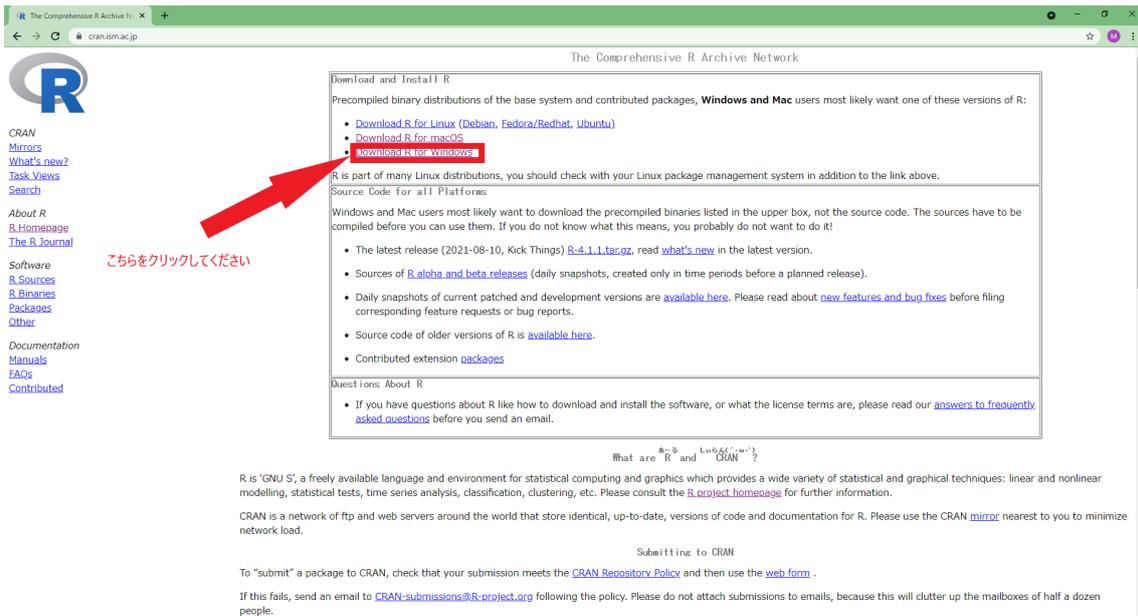
CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to minimize network load.

Submitting to CRAN

To "submit" a package to CRAN, check that your submission meets the [CRAN Repository Policy](#) and then use the [web form](#).

If this fails, send an email to CRAN-submissions@R-project.org following the policy. Please do not attach submissions to emails, because this will clutter up the mailboxes of half a dozen people.

② 「Download R for Windows」というリンクをクリックします。



The screenshot shows the same CRAN website as above, but with a red arrow pointing to the 'Download R for Windows' link in the list. A red text label 'こちらをクリックしてください' (Click here) is placed next to the arrow.

The main content area is titled 'Download and Install R' and contains the following text:

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2021-08-10, Kick Things) [R-4.1.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alphas and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and GNU?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

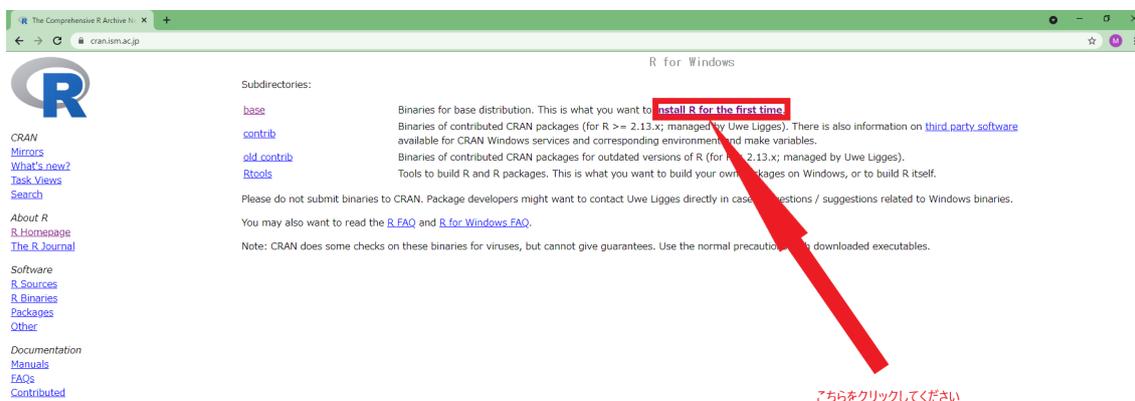
CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to minimize network load.

Submitting to CRAN

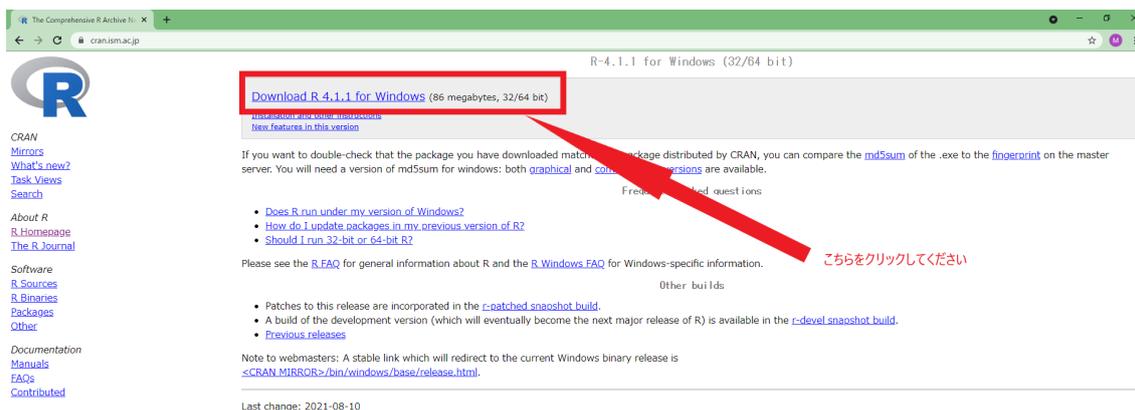
To "submit" a package to CRAN, check that your submission meets the [CRAN Repository Policy](#) and then use the [web form](#).

If this fails, send an email to CRAN-submissions@R-project.org following the policy. Please do not attach submissions to emails, because this will clutter up the mailboxes of half a dozen people.

③ 上方の「install R for the first time」をクリックします。



④ 上方の「Download R 4.1.1 for Windows」をクリックします。すると、ファイルのダウンロードが始まります。



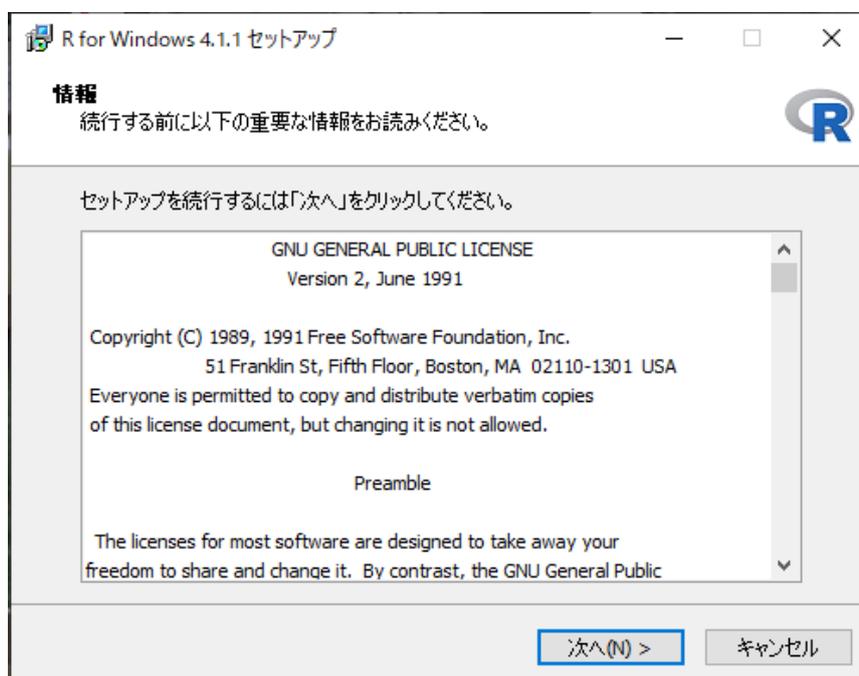
⑤ ダウンロードしたファイルを開きます。

The screenshot shows the CRAN website page for R 4.1.1 for Windows. The main heading is "Download R 4.1.1 for Windows (86 megabytes, 32/64 bit)". Below this, there are links for "Installation and other instructions" and "New features in this version". A paragraph explains that the downloaded package can be compared to the master server using md5sum. There are sections for "Frequently asked questions" and "Other builds". A red arrow points to the download link, with the text "こちらをクリックして、ファイルを開いてください" (Click here to open the file) written in red next to it.

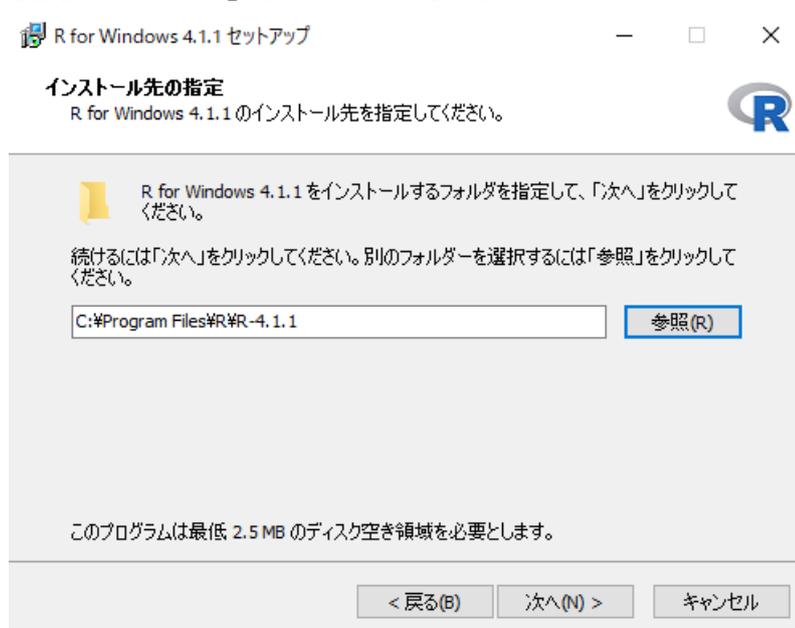
⑥ 使用する言語は「日本語」にして、「次へ」をクリックします。

The screenshot shows a language selection dialog box titled "セットアップに使用する言語の選択" (Select language for setup). The dialog contains a message: "インストール中に使用する言語を選んでください。" (Please select the language to use during installation.). Below the message is a dropdown menu with "日本語" (Japanese) selected. At the bottom of the dialog are two buttons: "OK" and "キャンセル" (Cancel).

- ⑦ 規約が表示されますので、一読して、「次へ」をクリックします。

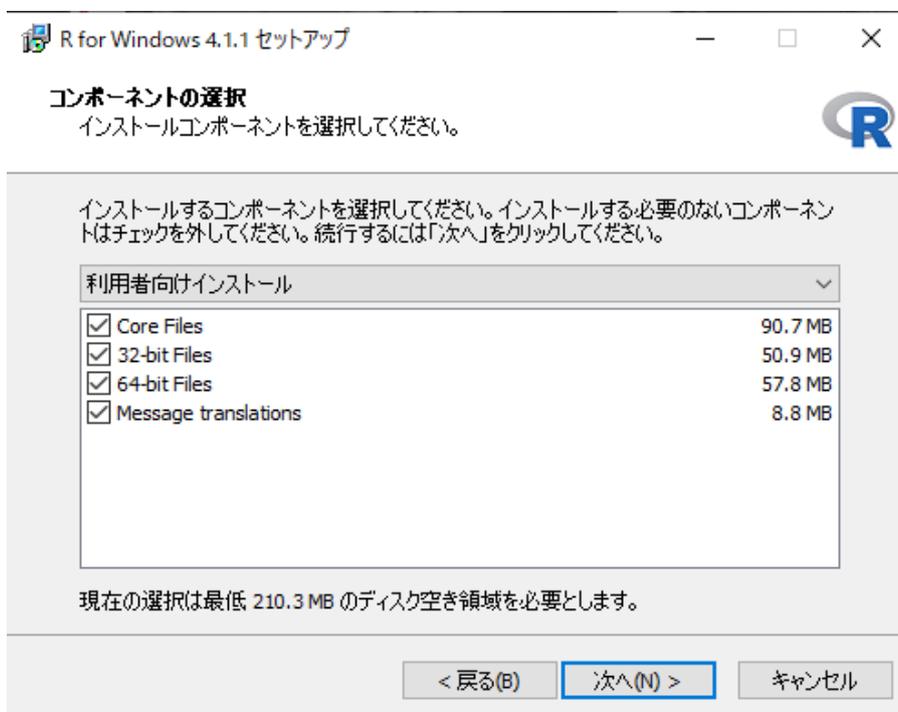


- ⑧ 変更せず、「次へ」をクリックします。

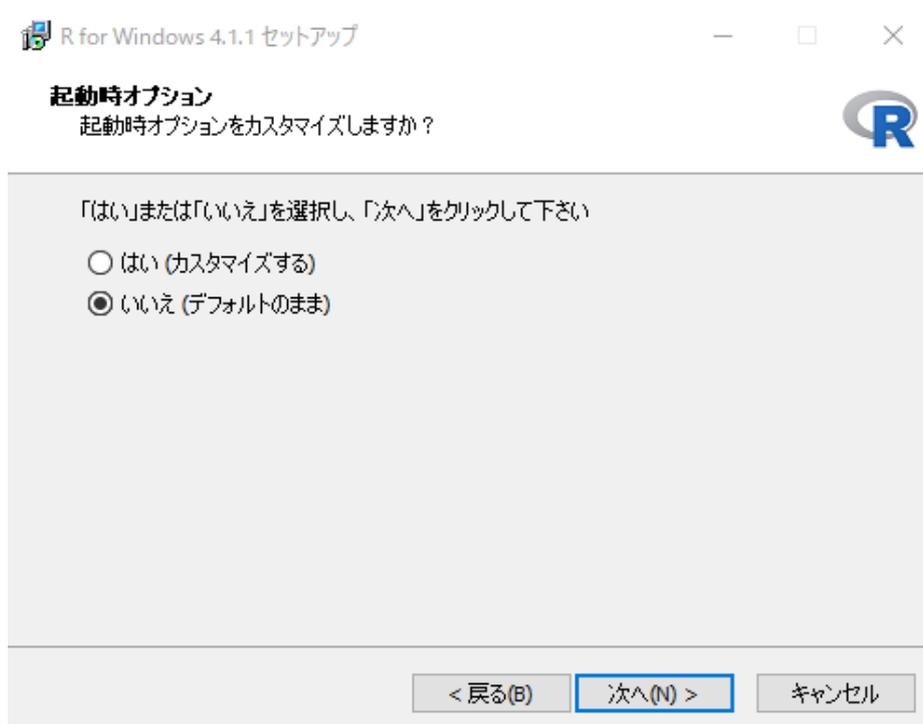


- ⑨ Windows が 32-bit か 64-bit か確認し（確認方法は下記のリンクを参照）、Windows が 64-bit の場合は、変更せず、「次へ」をクリックします。Windows が 32-bit の場合は、「64-bit Files」のチェックを外した上で「次へ」をクリックします。

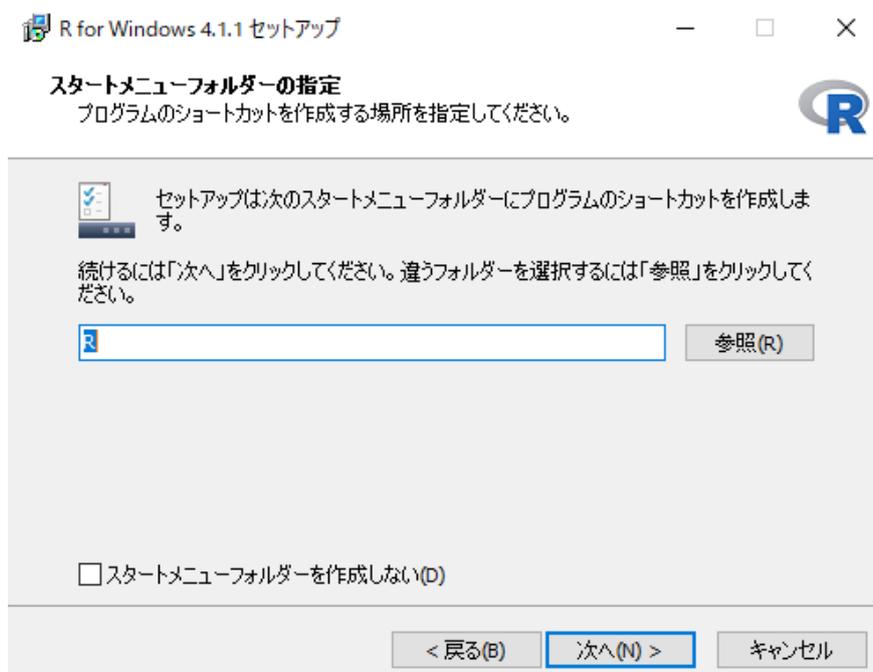
<https://support.microsoft.com/ja-jp/windows/32-%E3%83%93%E3%83%83%E3%83%88%E3%81%A8-64-%E3%83%93%E3%83%83%E3%83%88%E3%81%AE-windows-%E3%82%88%E3%81%8F%E5%AF%84%E3%81%9B%E3%82%89%E3%82%8C%E3%82%8B%E8%B3%AA%E5%95%8F-c6ca9541-8dce-4d48-0415-94a3faa2e13d>



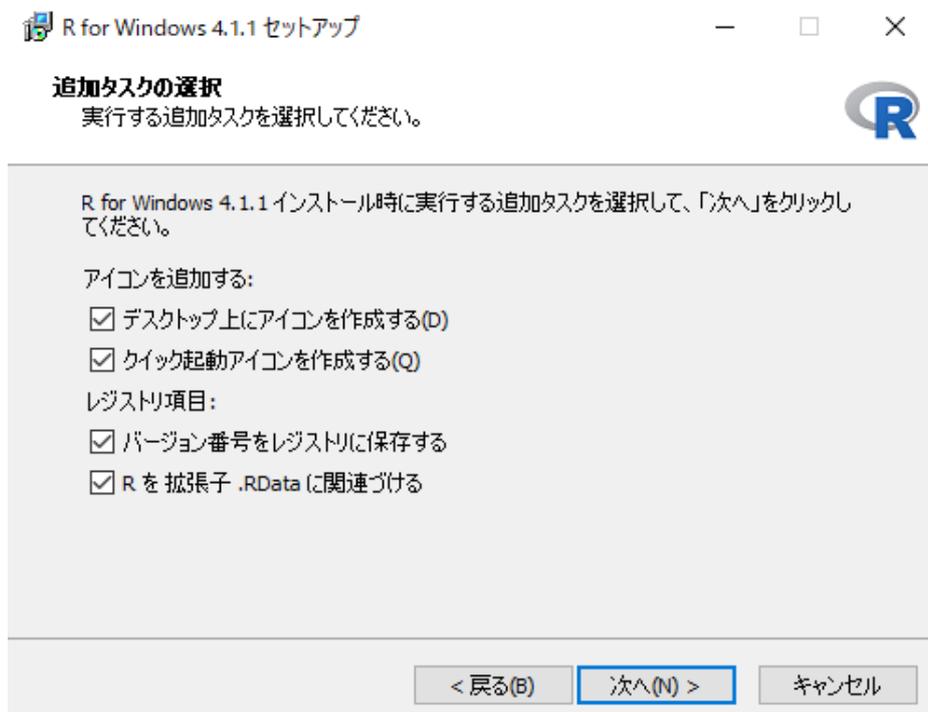
- ⑩ デフォルトのまま、「次へ」をクリックします。



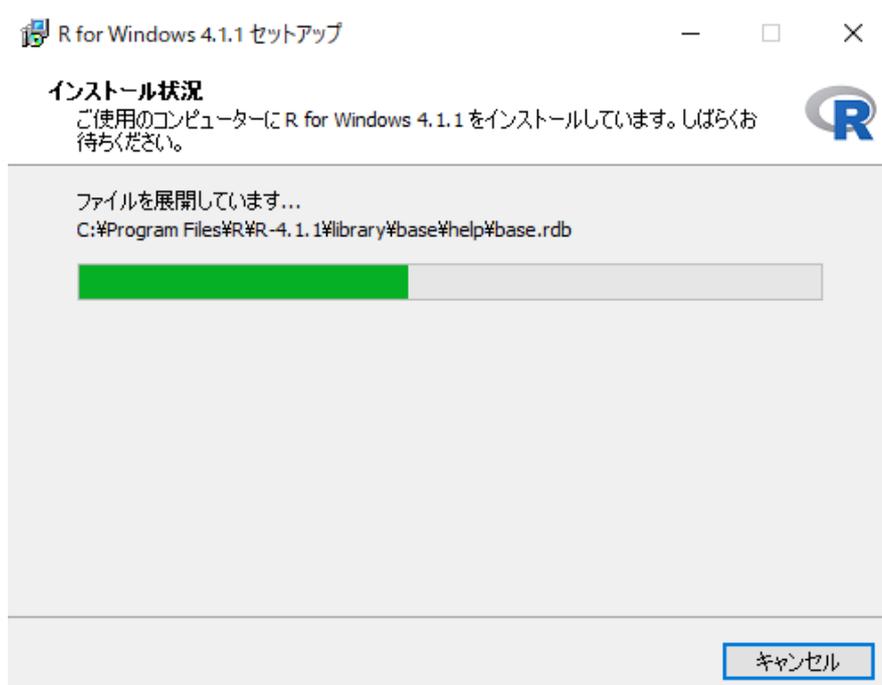
- ⑪ 変更せず、「次へ」をクリックします。



- ⑫ 上2つのチェックボックスにチェックをつけ、「次へ」をクリックします。



- ⑬ インストールが始まりますので、完了するまで待ちます。



- ⑭ この画面になりましたら、インストール完了です。「完了」をクリックします。



4. RStudio のインストール

- ⑮ RStudio のダウンロードサイトにアクセスします。
 - i. <https://www.RStudio.com/products/RStudio/download/>
- ⑯ 下記の画面の RStudio Desktop の下にある「Download」をクリックします。

Product	License	Price	Action	Additional Info
RStudio Desktop	Open Source License	Free	DOWNLOAD	Learn more
RStudio Desktop Pro	Commercial License	\$995 /year	BUY	Learn more
RStudio Server	Open Source License	Free	DOWNLOAD	Learn more
RStudio Workbench	Commercial License	\$4,975 /year (5 Named Users)	BUY	Evaluation Learn more

- ⑰ ご使用中の OS に合わせて、RStudio をダウンロードします。
 - i. Windows なら、「Windows 10」のリンクをクリックしてください。

RStudio Desktop 2021.09.0+351 - Release Notes

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:



Requires Windows 10 (64-bit)



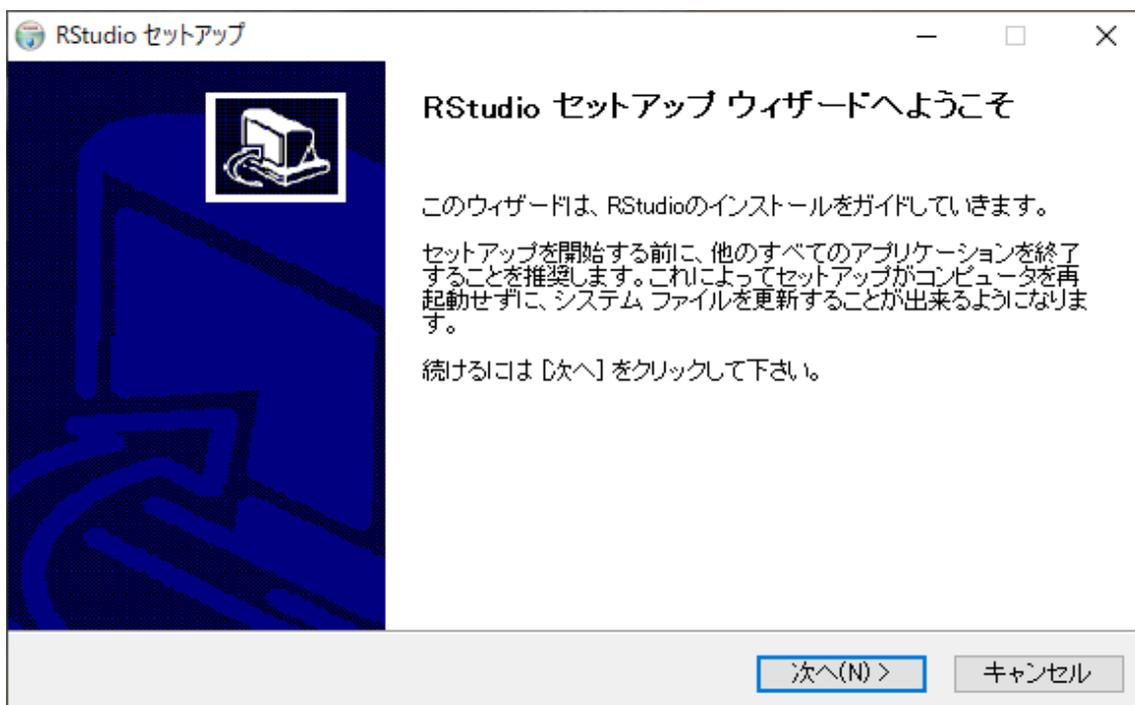
All Installers

Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy.
RStudio requires a 64-bit operating system. If you are on a 32-bit system, you can use an older version of RStudio.

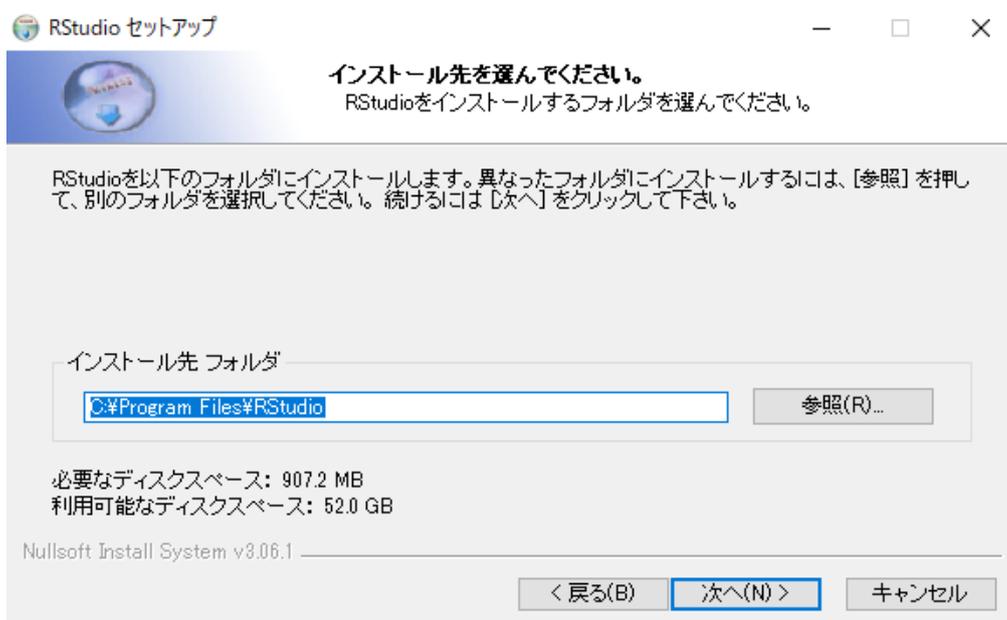
OS	Download	Size	SHA-256
Windows 10	RStudio-2021.09.0+351.exe	196.88 MB	f68644c2
macOS 10.14+	RStudio-2021.09.0+351.dmg	196.28 MB	f6e87cad
Ubuntu 18/Debian 10	rstudio-2021.09.0+351-amd64.deb	116.53 MB	0c7ef292
Fedora 19/Red Hat 7	rstudio-2021.09.0+351-x86_64.rpm	133.82 MB	3c953521
Fedora 28/Red Hat 8	rstudio-2021.09.0+351-x86_64.rpm	133.84 MB	1c4d342b
Debian 9	rstudio-2021.09.0+351-amd64.deb	116.79 MB	809c7c7c
OpenSUSE 15	rstudio-2021.09.0+351-x86_64.rpm	119.27 MB	1f8efae4

- ⑱ ダウンロードしたファイルをダブルクリックし、開きます。

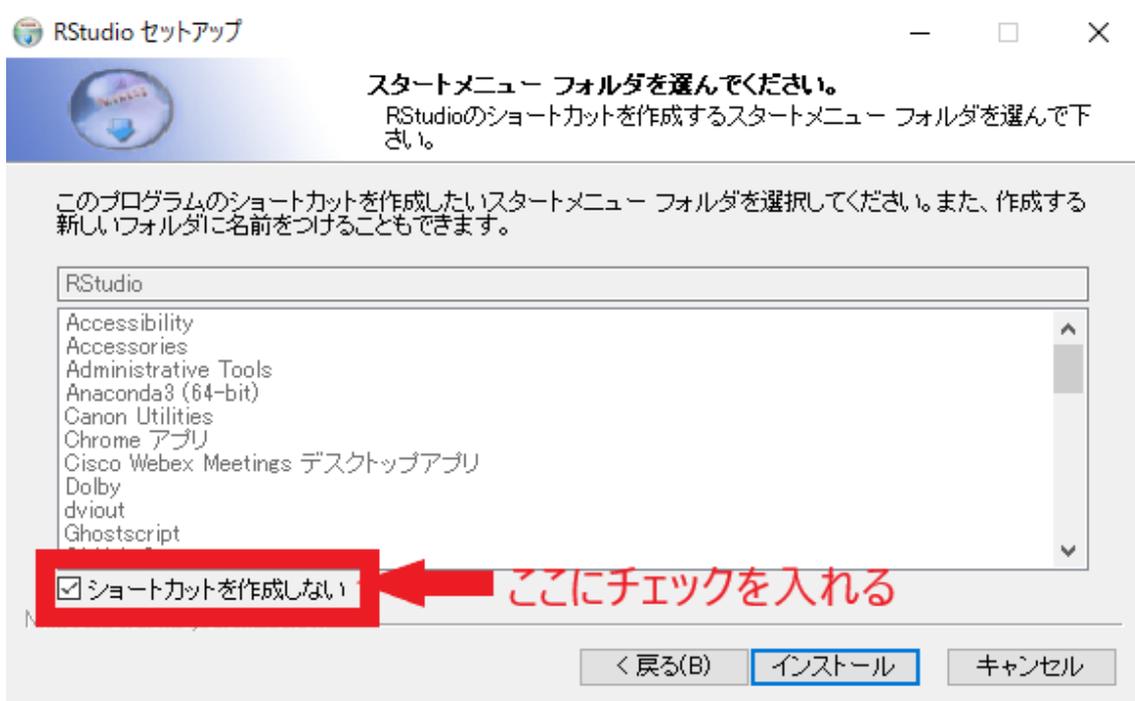
- ⑱ セットアップウィザードが開くので、「次へ」をクリックします。



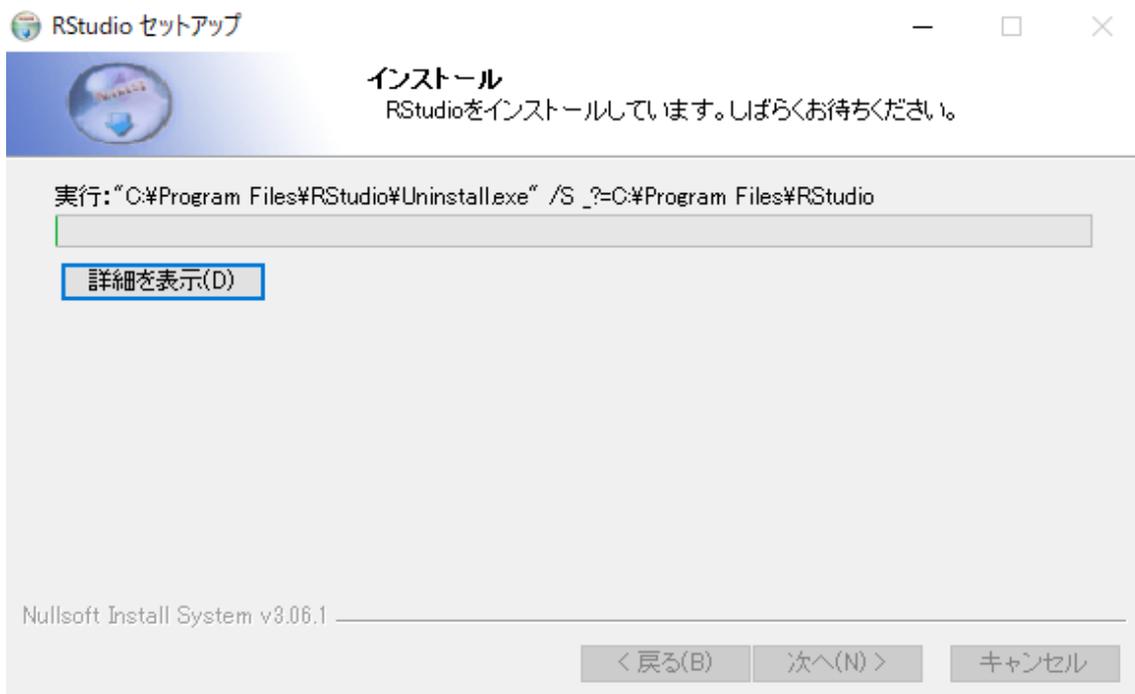
- ⑳ 変更せず、「次へ」をクリックします。



- 21 「ショートカットを作成しない」をチェックし、「インストール」をクリックします



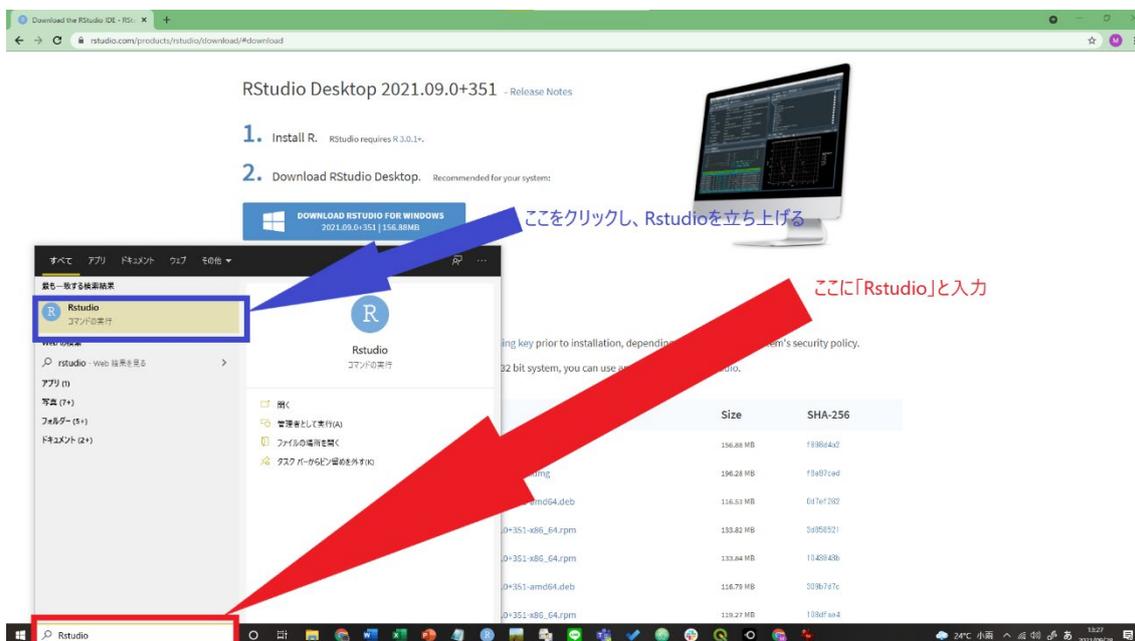
- 22 インストールが始まるので、完了まで待ちます。



23 インストールが完了したら、「完了」をクリックします。



24 画面下部の検索欄に「RStudio」と入力し、RStudio をクリックすると、RStudio が起動します。



5. RStudio による匿名化の方法

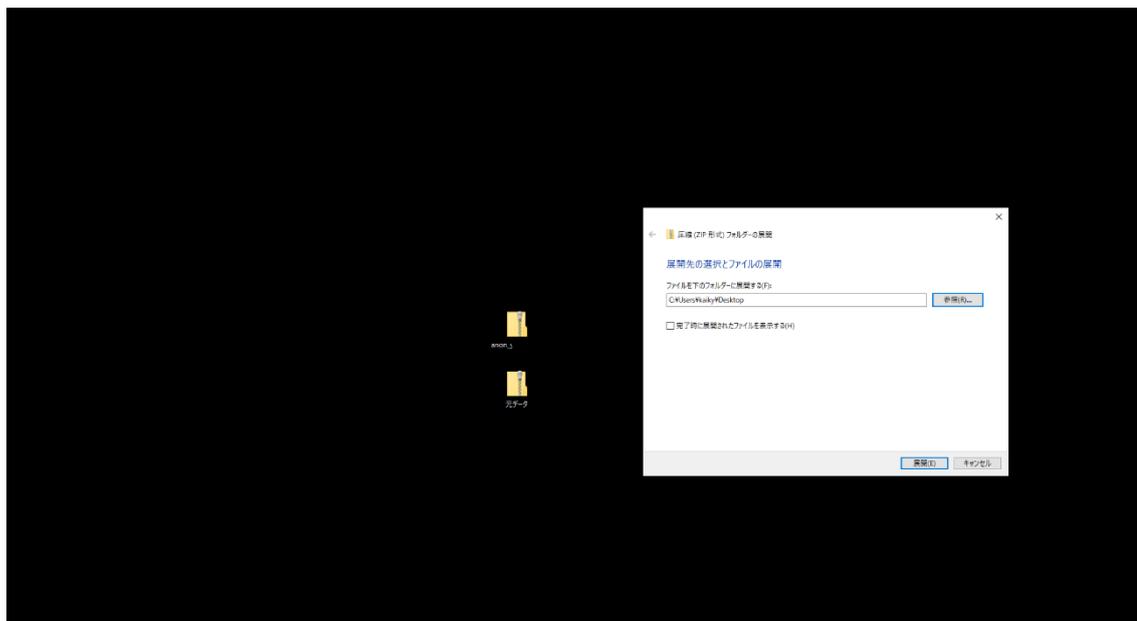
I. 全体的な注意点

- 次のファイルは、中身の変更をしないでください。
 - 「config」フォルダの中にあるファイル
 - 「anonymization.R」
- 匿名化には、ややお時間をいただく可能性があります。自治体の規模によっては、数時間いただく可能性もございます。プログラムが停止していない限りは、終わるまでお待ちください。
 - プログラムが動作している間、別の作業をしながらお待ちいただいて構いません。
 - 当方の PC（プロセッサ：Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz 2.30 GHz、実装 RAM：16.0 GB）で、人口約 2 万の自治体のデータ（データ項目数：約 600）10 年分を処理したところ、約 10 分で処理が完了しました。必ずしも処理時間が人口やデータ項目数、年数に比例するわけではありませんが、参考にしてください。
- もしプログラムが正常に動作しなかった場合は、CREPE 側で検証を行う必要があります。IIIでご案内しております手順に従ってください。

II. 具体的な手順

① ファイルを解凍する

こちらから別途お送りする「anon_〇〇.zip」(〇〇は自治体の名前)という zip ファイルを解凍します。展開先は、どこでも構いませんが、以下ではデスクトップを指定しています。



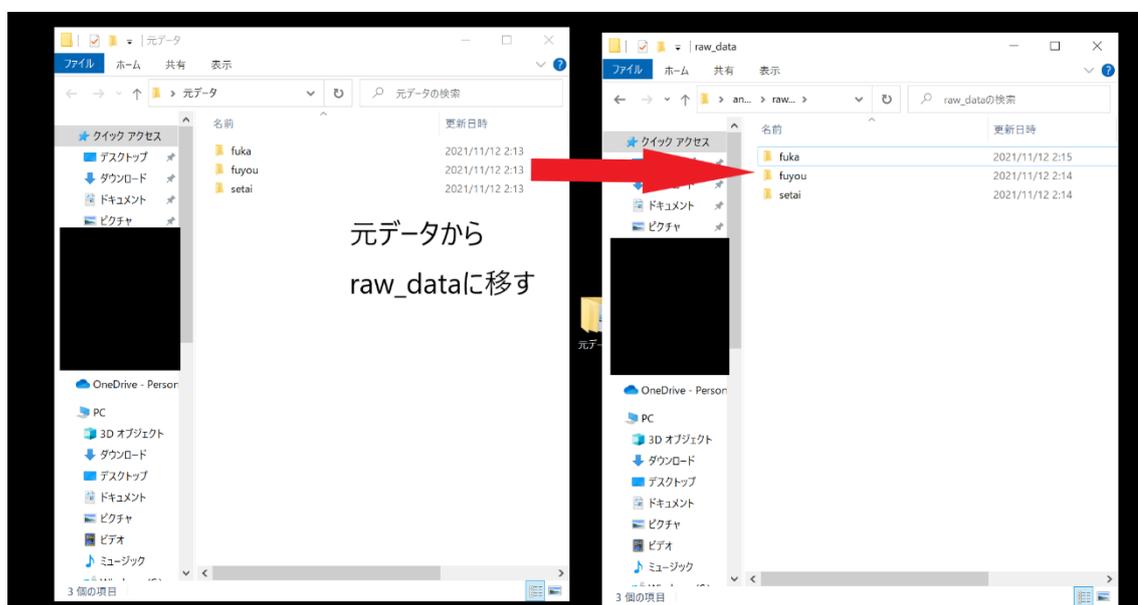
② 元データをチェックする

元データをチェックします。元データには複数のフォルダがあると思いますが、そのフォルダ内にあるファイルについては、列名がすべてそろっていることを確認してください。

③ 動作チェック用に、データを格納する

生成されたフォルダを開きます。「anon_〇〇」フォルダの中にある「raw_data」というフォルダの中に、元データを格納していただきます。

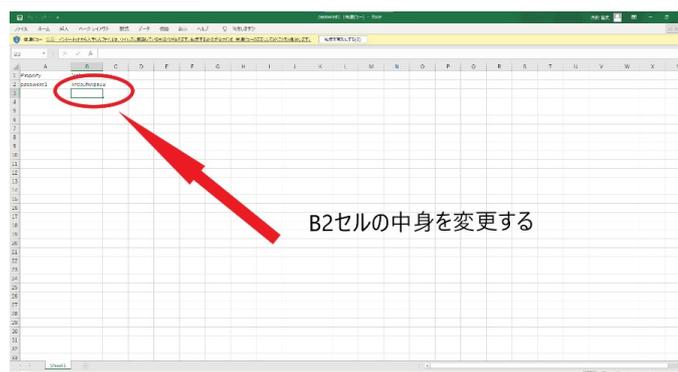
- 匿名化処理には、場合によって数時間を要する可能性があります。そのため、いきなり全てのデータを用いて匿名化処理を行うのではなく、まずは、1年度のみのデータで、動作チェックを行います。
- そこで、各フォルダにつき、データセット中最も古い年度の1ファイルのみになるようにして、データを格納してください
- 事前にお知らせいただいたフォルダ構造・フォルダ名のまま、データを格納してください。フォルダ構造やフォルダ名を変えてしまいますと、プログラムが動作しなくなってしまいます。



④ パスワードを変更する

「anon_〇〇」フォルダに、「passwords」というフォルダがあります。その中の「password1.xlsx」、「password2.xlsx」を開いてください。開きましたら、B2セルに入っている文字列を、別の文字列（何でも結構です。）に変更してください。

- 「password1」と「password2」に、それぞれ異なる文字列を入力してください。
- 変更後の文字列は、自治体の中で忘れないように保管してください。**CREPE側には知らせないでください。**



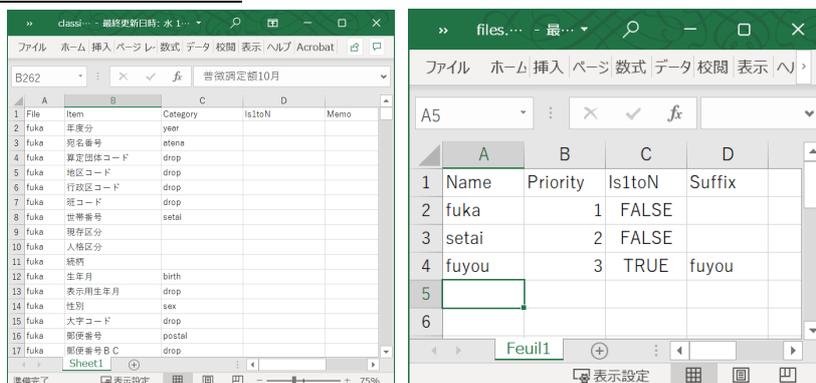
⑤ さらにチェックを行う

「config」フォルダの中に、「classification.xlsx」、「files.xlsx」というファイルがあります。それぞれ開いていただきます。

- 「classification.xlsx」には、いただくデータ項目の一覧と、匿名化方法が記されております。事前にお知らせした匿名化方針と相違ないか、ご確認ください。

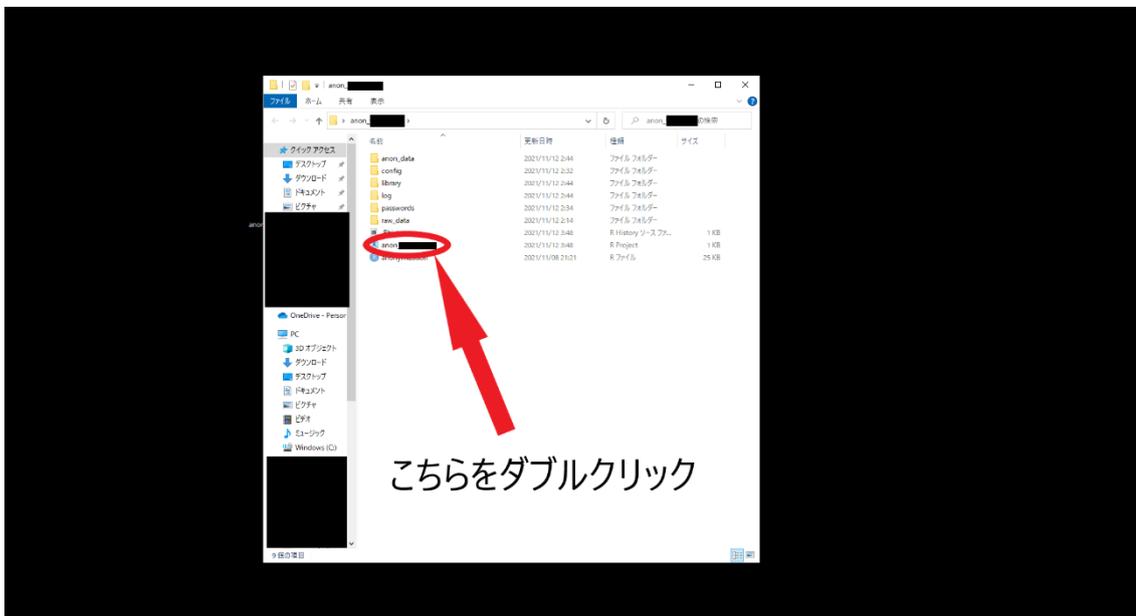
Category にある分類	意味	「匿名加工についての考え方(案)」3に基づく匿名化処理
year	年度	
atena	宛名番号	①②に沿って、ハッシュ化を行います。
setai	世帯番号	①に沿って、特異な世帯の世帯番号を秘匿します。 ①②に沿って、ハッシュ化を行います。
birth	生年月日	③⑤に沿って、月単位に丸めます。 ⑥に沿って、「3-匿名性」を判定し、「3-匿名性」を満たすまで秘匿化します。
sex	性別	⑥に沿って、「3-匿名性」を判定し、「3-匿名性」を満たすまで秘匿化します。
postal	郵便番号	⑥に沿って、「3-匿名性」を判定し、「3-匿名性」を満たすまで秘匿化します。
topcode	トップコーディング対象	④に沿って、トップコーディングを行います。
hash	ハッシュ化対象	ハッシュ化を行います。
drop	削除対象	データセットから削除します。

- 「files.xlsx」には、Nameの列に「raw_data」に格納していただくフォルダの一覧が示されております。手順③で格納したフォルダとあっているかどうか、ご確認ください。Name以外の列は無視してください。
- これら2つのファイルについては、ご確認いただきたいものの、**中身を変更しないようにしてください。**変更すると、プログラムが動作しなくなります。



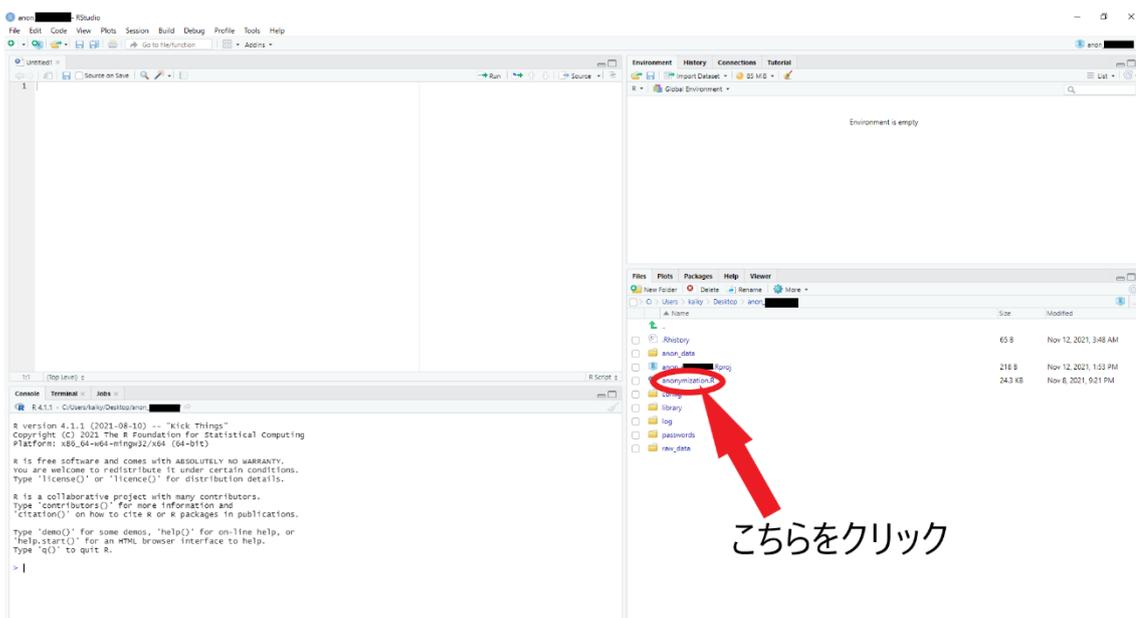
⑥ RStudio を開く

ここから、RStudio でプログラムを動作させます。「anon_〇〇」フォルダのなかにある、「anon_〇〇.Rproj」というファイルをダブルクリックして開きます。



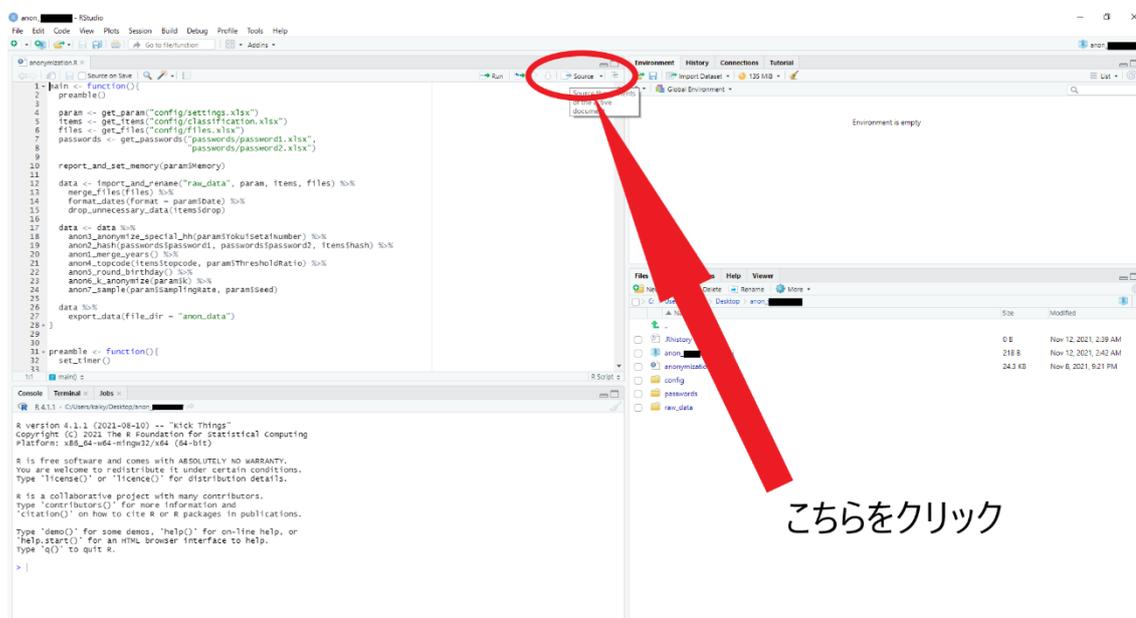
⑦ 匿名化プログラムを開く

画面右下に表示される「anonymization.R」というファイルをクリックして開きます。



⑧ プログラムを実行する

画面中央にある「Source」というボタンをクリックします。すると、プログラムが動作し始めます。

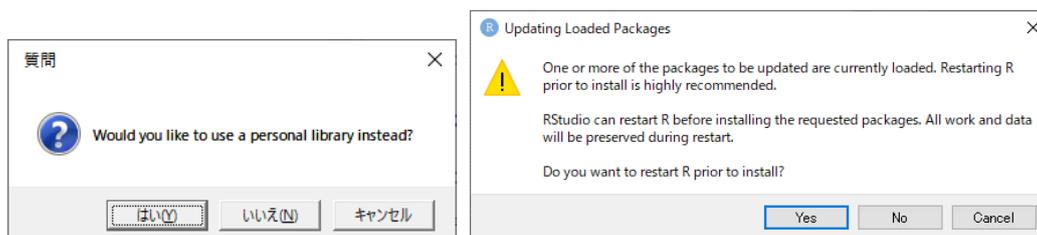


⑨ 確認メッセージに対応する

動作中に、以下のようなメッセージが表示されることがあります。

これは、プログラムを動作させるためのパッケージをインストールするためのものですので、「はい」又は「Yes」をクリックしてください。

もし特段、このようなメッセージが表示されなければ、手順⑪に進んでください。



⑩ プログラムを再実行する

⑨の操作を行った後、画面左下に“>”が表示されたものの、「ANONYMIZATION SUCCESSFUL」が表示されていない場合は、パッケージのインストール後、改めてプログラムを実行する必要がありますので、もう一度、画面中央にある「Source」というボタンをクリックします。

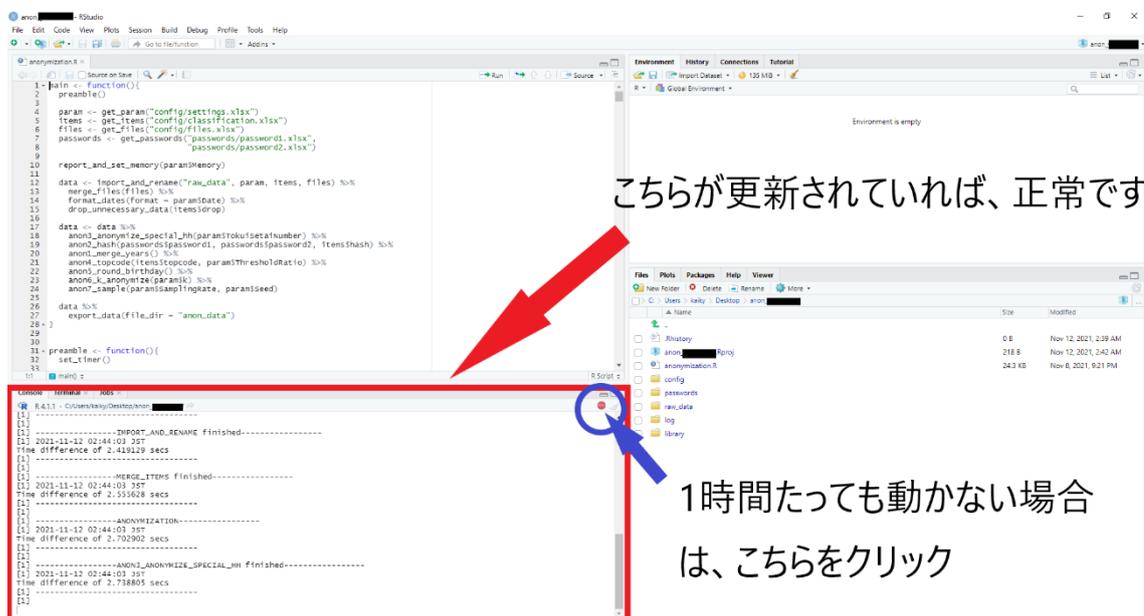
5分ほど待ってみても左下に“>”が表示されない場合は、手順⑪に進んでください。

⑪ プログラムが動作しているか確認する

画面左下に“>”が表示されれば、プログラムの動作が完了しています。手順⑫に進んでください。

画面左下に“>”が表示されていない場合は、プログラムが動作中です。

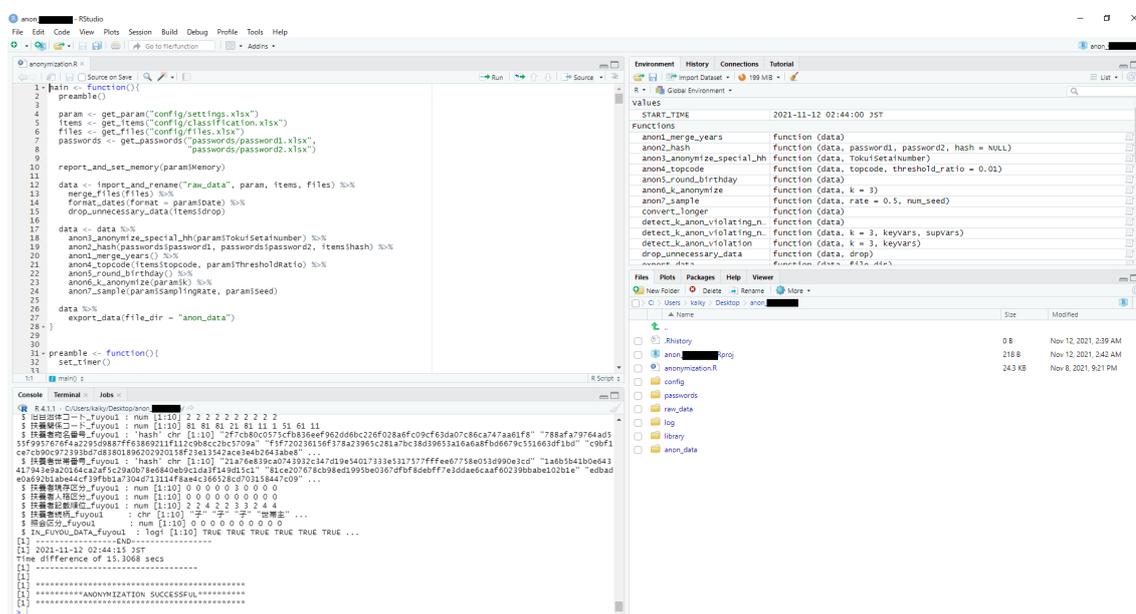
- 画面左下が更新され続けているようであれば、プログラムは正常に動作しています。そのまま待ってください。別の作業をしながら、プログラムの終了を待っていただいても構いません。
- 1時間経っても画面左下に変化がない場合は、プログラムに何らかの問題が生じている可能性があります。画面中央の赤色の「STOP」をクリックした上で、画面右上の「×」をクリックして手順Ⅲに進んでください。保存するかどうかを聞かれた場合、「いいえ」又は「No」をクリックしてください。



⑫ プログラムの完了を確認する

画面左下に“>”が表示されれば、プログラムの動作が完了しています。

- “>”の上に「ANONYMIZATION SUCCESSFUL」と出ていれば、プログラムの動作が無事に完了しています。これで、プログラムの動作チェックは以上ですので、次から本番の作業に移っていきます。手順⑬に進んでください。
- “>”の上に「ANONYMIZATION SUCCESSFUL」と出ていなければ、プログラムに何らかの問題が生じている可能性があります。画面右上の「×」をクリックして手順Ⅲに進んでください。保存するかどうかを聞かれた場合、「いいえ」又は「No」をクリックしてください。



⑬ データをすべて格納する

手順③と同様に、「raw_data」フォルダに、元データを入れていきます。今回は、全てのデータを格納してください。

⑭ 匿名化を実行する

手順⑧と同様、「Source」ボタンをクリックし、プログラムを実行してください。同様に、手順⑨～⑪に従ってください。

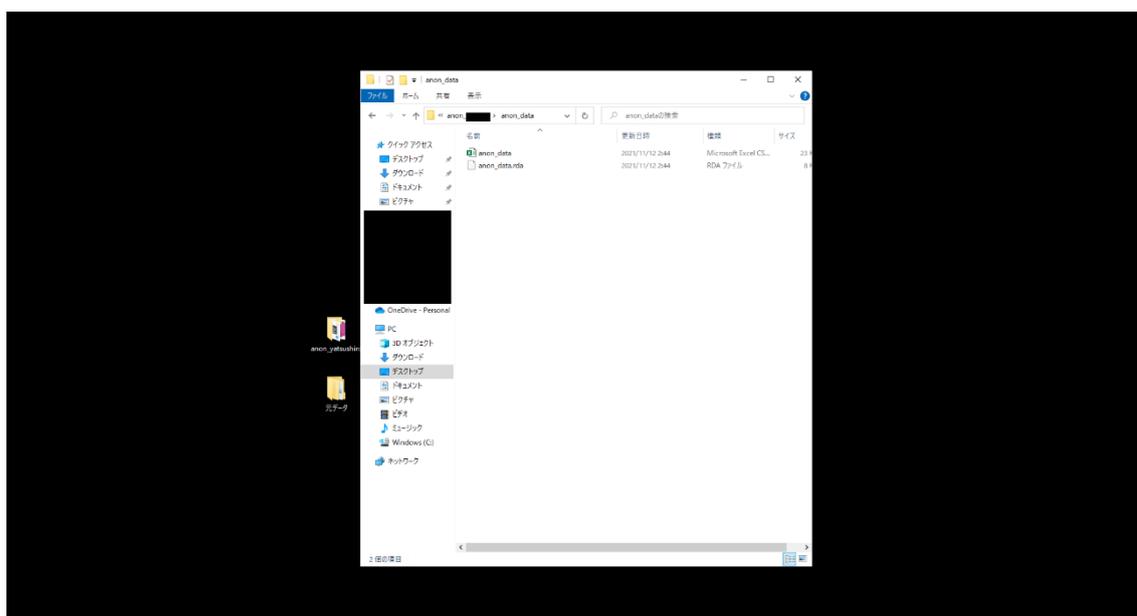
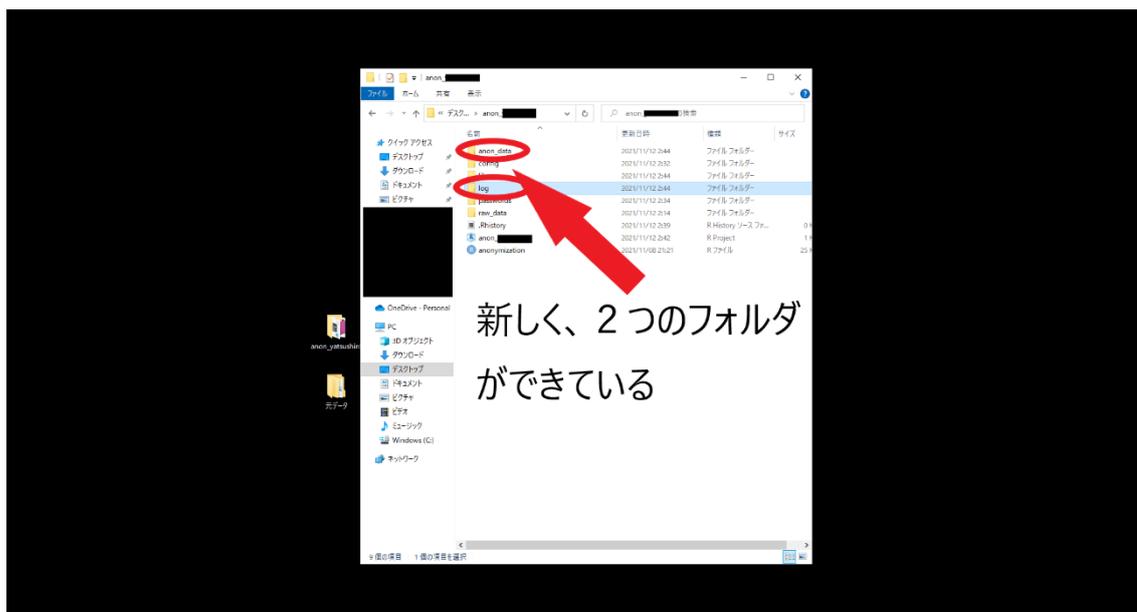
⑮ プログラムの終了を確認する

手順⑫と同様、画面左下に“>”が表示されれば、プログラムの動作が完了しています。

- “>”の上に「ANONYMIZATION SUCCESSFUL」と出ていれば、プログラムの動作が無事に完了しています。画面右上の「×」をクリックして手順⑮に進んでください。保存するかどうかを聞かれた場合、「いいえ」又は「No」をクリックしてください。
- “>”の上に「ANONYMIZATION SUCCESSFUL」と出ていなければ、プログラムに何らかの問題が生じている可能性があります。画面右上の「×」をクリックして手順Ⅲに進んでください。保存するかどうかを聞かれた場合、「いいえ」又は「No」をクリックしてください。

⑩ 加工済みデータを確認する

「anon_〇〇」フォルダに、「anon_data」フォルダが新しくできています。それを開き、「.csv」で終わるファイルと、「.rda」で終わるファイルの2種類のファイルができていることを確認してください。この2種類は、ファイル形式は異なりますが、内容は同一のものです。適切な匿名化が施されているか確認したい場合は、「.csv」で終わるファイルを開いて確認していただくことが可能です。



⑰ 加工済みデータを送信する

最後に、加工済みデータを CREPE まで送っていただきます。送っていただきたいファイルは次の 2 種類です。

- 「anon_data」のフォルダの中にある「.rda」で終わるファイル全て
 - 「.rda」で終わるファイルが複数ある場合は、全て送ってください。
 - 「.csv」で終わるファイルを送っていただく必要はありません。
- 「log」フォルダの中身全て
 - 複数の.txt ファイルが生成されています。プログラムが正常に動作しているか、ログを記録しているものですので、全てお送りください。

III. もしエラーが発生したら

手順⑪、⑫、⑮で、画面左下に“>”が表示されたものの、その上に「ANONYMIZATION SUCCESSFUL」と出ていなければ、プログラムに何らかの問題が生じている可能性があります。

その場合、CREPE 側で検証を行う必要がありますので、

- 「log」フォルダの中にある.txt ファイル全てを CREPE までお送りください。