CSVファイル

Andrew Ba Tran

目	次	
	csv ファイルはどのようなものか	1
	CSV ファイルの中身はどうなっているのか	2
cs	V ファイルのインポート	2
	データを取得する二つの方法	3
	URL の入手	3
	read.csv()	3
	データをインポートする別の方法:ダウンロード	4
	覚えていますか:RStudioでディレクトリを変更する方法	5
	ローカルの CSV のインポート	5
	stringsAsFactors=F	6
	readr パッケージから read_csv()を使う	6
cs	∨ファイルのエクスポート `NA`があるデータフレームのエクスポート	 6 7
	演習	7

This is from the second chapter of learn.r-journalism.com.

コンマで区切られたファイルは、Microsoftの有料プログラムを使わずにスプレッドシートを保存する最も 一般的な方法です。

csv ファイルはどのようなもの

か CSVのファイル名は.csv で終わりま す。

Admissions_to_DMHAS_Addiction_Treatment_by_Town_Yearand_Month.csv											
0	안 🛅 🗟 🗟 👗 🗳 📫 🗳 🛛 🔍 Search in Sheet 🔊 »										
Home Layout Tables Charts SmartArt Formulas											
E	dit	Font		Alignment	Numb	ber	Format	Cell	s		
	Calib	ri (Body)	▼ 12		General	•			- A		
Pa	ste B	ΙU		Align	S • 9	0) Co Fo	nditional rmatting Sty	les Actio	ns Tł		
	A1	\$ ⊗	🔘 (* f:	FiscalYe	ar				-		
	A	В	С	D	E	F	G	Н	=		
1	FiscalYear	MonthYear	Town	AdmMonth	FYMonthOrd	AdmYear	MonthTotal				
2	2014	13-Jul	Ansonia	7	1	2013	42				
3	2014	13-Aug	Ansonia	8	2	2013	43				
4	2014	13-Sep	Ansonia	9	3	2013	39				
5	2014	13-Oct	Ansonia	10	4	2013	33				
6	2014	13-Nov	Ansonia	11	5	2013	38				
7	2014	13-Dec	Ansonia	12	6	2013	45				
8	2014	14-Jan	Ansonia	1	7	2014	35				
0	2014	14 5-1-	A	2	0	2014	20				

CSV ファイルの中身はどうなっているのか

画像のファイル名は、コンマで区切られた値を説明しています。

Admissions_to_DMHAS_Addiction_Treatment_by_Town_Yearand_Month.csv	
FiscalYear,MonthYear,Town,AdmMonth,FYMonthOrder,AdmYear,MonthTotal	ñ
2014,July-13,Ansonia,7,1,2013,42	ļ
2014,August-13,Ansonia,8,2,2013,43	
2014, September-13, Ansonia, 9, 3, 2013, 39	
2014,October-13,Ansonia,10,4,2013,33	
2014,November-13,Ansonia,11,5,2013,38	
2014,December-13,Ansonia,12,6,2013,45	
2014, January-14, Ansonia, 1, 7, 2014, 35	
2014,February-14,Ansonia,2,8,2014,38	
2014,March-14,Ansonia,3,9,2014,45	
2014,April-14,Ansonia,4,10,2014,46	
2014,May-14,Ansonia,5,11,2014,40	
2014,June-14,Ansonia,6,12,2014,56	
2015,July-14,Ansonia,7,1,2014,54	
2015,August-14,Ansonia,8,2,2014,53	
2015, September-14, Ansonia, 9, 3, 2014, 50	
2015,October-14,Ansonia,10,4,2014,60	
2015,November-14,Ansonia,11,5,2014,41	

CSV ファイルのインポート

- CSV のインポートは base R でできるので、パッケージは不要です。

- しかし、ここではあえて readr パッケージを使ってみましょう。

データを取得する二つの方法

* URL アドレスをお持ちの場合

* csv ファイルがインターネット上にある場合は、ローカルにダウンロードしてからインポートする必要はありません。リンクを使用して Web から直接 R にインポートできます。

*コンピュータにファイルがある場合

URL の入手

CSV ファイルのリンクがあるならば、[データのリンク](https://data.ct.gov/Health-and-Human-Services/Admissions-to-DMHAS-Addiction-Treatment-by-Town-Ye/iyru-82zq/data)を右クリックし、さらに**リンク** のアドレスをコピーをクリックします。このデータセットは、[Connecticut Open Data Portal](https://data.ct.gov/Health-and-Human-Services/Admissions-to-DMHAS-Addiction-Treatment-by-Town-Ye/iyru-82zq/)にあります。

1	SODA API
8	OData (
9	Print
10	Download -
11	
12	Download a copy of this dataset in a static format
1	
2	Download As
3	Open Link in New Tab
4	CSV Open Link in New Window
5	JSO Open Link in Incognito Window
6	RDF Save Link As
7	RSS Copy Link Address
8	тѕу Сору
9	XML Search Google for "CSV"
10	Print
11	1 Password
12	Block element

read.csv()

CSV ファイルをインポートするための関数は`read.csv()`です。URL アドレスを引用符で囲み、(文字列を因子型として読み込まないよう)`stringsAsFactors=F`を追加します。(ここでは関数 `head()`を使用しています。デフォルトでは冒頭の6行を返しますが、10行見てみたいので head(data, 10)のように指定します。)

df_csv <- read.csv("https://data.ct.gov/api/views/iyru-82zq/rows.csv?accessType=DOWNLOAD", stringsAsFact head(df_csv, 10)

 ##
 FiscalYear
 MonthYear
 Town AdmMonth FYMonthOrder AdmYear

 ## 1 2014 July-13 Ansonia 7 1 2013 ## 2 2014 August-13 Ansonia 8 2 2013 ## 3 2014
 September-13 Ansonia 9 3 2013 ## 4 2014 October-13 Ansonia 10 4 2013 ## 5 2014

 November-13 Ansonia 11 5 2013 ## 6 2014 December-13 Ansonia 12 6 2013 ## 7 2014
 January-14 Ansonia 1 7 2014 ## 8 2014 February-14 Ansonia 2 8 2014 ## 9 2014 March-14 Ansonia 3 9 2014 ## 10 2014 April-14 Ansonia 4 10 2014

##	wonthiota	l
## 1	42 ## 2	43
## 3	39 ## 4	33
## 5	38	
## 6	4	5
## 7	3	5
## 8	3	8
## 9	4	5
## 10	4	6
データ	マをインオ	ペートする別の方法:ダウンロード

リンクのアドレスをコピーするのではなく、リンクを右クリックして**名前を付けてリンク先を保存**をクリックします

10	Download 🗸
11	
12	format
1	
2	Download As
3	Ce Open Link in New Tab
4	CS Open Link in New Window
5	JS Open Link in Incognito Window
6	RI Save Link As
7	Copy Link Address
8	тя Сору
9	XN Print
10	
11	(1) 1Password
12	md Copy as Markdown
1	► DataMiner-Get Similar
2	Increat

ワーキングディレクトリに保存してください。

右側の更新を表す矢印をクリックしてファイルを更新し、保存されていることを確認します。

New F	older 🧕 Delete 👍 Rename 🔮 More 🗸		(G
	Name	Size	Modified	Refresh file listing
t.				
i 🖳 🗆	mporting_data.Rpres	2.9 KB	Feb 21, 2017, 9:03 PM	
🗆 🗷 i	mporting_data.Rproj	205 B	Feb 21, 2017, 8:33 PM	
🗆 🧰 i	mporting_data-figure			
🗆 🧰 i	mages			
🗆 🔟 i	mporting_data.md	303.6 KB	Feb 21, 2017, 9:04 PM	

覚えていますか: RStudio でディレクトリを変更する方法

`setwd("/directory/where/you/want")` と入力するか、メニューから Session > Set Working Directory > Choose Directory...と進んでください。

Code	View	Plots	Session	Build	Debug	Profile	Tools	Window	Help		
			New Se	ssion						~/Docu	
থି_ + 😋	rting_data.	🗗 🖴 Rpres ×	Interrup Termina	ot R ate R			ldins 👻				
	26	ABC	Restart	R		企					
81 82 -	strings	AsFactor	Set Wo	rking Dii	rectory	•	To Project Directory				
82 ▼ ===================================			Load Workspace Save Workspace As				To Files Pane Location				
85 86	Blame st	tatistic	Clear W	/orkspac	ce		Choo	se Director	y ^	Select an (个企H)	
87 88 80	Back when R was Quit Session						it as we use it now				
90 91 - 92	The othe	er way t	o import	the dat	a: Downlo	oad it		=			

ローカルの CSV のインポート

先ほどの同じですが、URLの代わりにファイル名を使います。

Note: ワーキングディレクトリが csv ファイルがある場所に設定されている場合にのみ機能します。

df_csv <- read.csv("data/Admissions_to_DMHAS_Addiction_Treatment_by_Town__Year__and_Month.csv", stringsA

stringsAsFactors=F

なぜこれが必要なのでしょう。

統計学者のせいです。

R が作成された当時、現在のようにさまざまな文字列を扱うことはありませんでした。

`stringsAsFactors=F`を使わないと

df_csv <- read.csv("data/Admissions_to_DMHAS_Addiction_Treatment_by_Town__Year__and_Month.csv") str(df_csv)

readr は矩形データを素早く読み込むことができるパッケージで、デフォルトでは文字はストリングであって、因子型ではないと仮定しています。

)

readr をインストールしていなければ、以下の行のコメントを外して実行してください。 *#install.packages("readr")* **library(readr)**

df_csv <- read csv("data/Admissions_to_DMHAS_Addiction_Treatment_by_Town__Year__and_Month.csv"

Parsed with column specification:

cols(

- ## FiscalYear = col_integer(),
- ## MonthYear = col_character(),
- ## Town = col_character(),
- ## AdmMonth = col_integer(),
- ## FYMonthOrder = col_integer(),
- ## AdmYear = col_integer(),
- ## MonthTotal = col_integer()

##)

ご覧のとおり、`read_csv()` 関数は MonthYear と Town 列を`read.csv()`のように因子型としてではなくストリングとして解釈しました。

CSV ファイルのエクスポート

データの分析や変換が終了したら、readr パッケージの`write_csv()`を使用してデータフレームを CSV ファイ ルとして保存しましょう。 #write_csv() 関数にデータフレームの名前と、ファイル名を付けて渡します。

write_csv(df_csv, "transformed_data.csv")

ファイルはワーキングディレクトリに保存されますが、関数を使ってサブディレクトリを指定することができます。

write_csv() 関数にデータフレームの名前と、ファイル名を付けて渡します。

write_csv(df_csv, "data/transformed_data.csv")

`NA` があるデータフレームのエクスポート

警告:エクスポートされたファイルに含まれる`NA`を置き換えるには、変数 `na="whatever"`を渡します。 # これは NA を空白に置きかえます。

s write_csv(df_csv, "data/transformed_data.csv", na="")

演習

[これらの演習](http://code.r-journalism.com/chapter-2/#section-csvs) でこのセクションの知識を身につけましょう。

エクササイズアプリの実行に関する説明は、このセクションの [紹介ページ](http://learn.r-journalism.com/en/importing_exporting/) にあります。