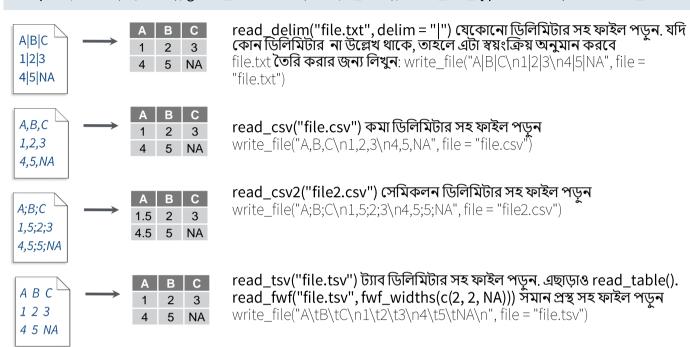
tidyverse এর মাধ্যমে ডেটা আমদানি :: চিট শিট

readr

readr এর মাধ্যমে ট্যাবূলার ডেটা পড়া

read_*(file, col_names = TRUE, col_types = NULL, col_select = NULL, id = NULL, locale, n_max = Inf,
 skip = 0, na = c("", "NA"), guess_max = min(1000, n_max), show_col_types = TRUE) See ?read_delim

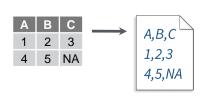


দরকারি ফাইল পড়ার আর্গুমেন্ট

A B C 1 2 3 4 5 NA	কলাম শিরোনাম ছাড়া read_csv("file.csv", col_names = FALSE)	1 2 3 4 5 NA	লাইন এড়িয়ে যান read_csv("file.csv", skip = 1)
x y z A B C	শিরোনাম সহ	A B C 1 2 3	কিছু লাইন পড়ন read_csv("file.csv", n_max = 1)
1 2 3 4 5 NA	read_csv("file.csv", col_names = c("x", "y", "z"))	A B C NA 2 3 4 5 NA	নিরুদ্দিস্ত হলে মান উল্লেখ করুন read_csv("file.csv", na = c("1"))
→	একাধিক ফাইল এক টেবিলে পড়ুন read_csv(c("f1.csv", "f2.csv", "f3.csv"), id = "origin_file")	A;B;C 1,5;2;3,0	দশমিক চিহ্ন উল্লেখ করুন read_delim("file2.csv", locale = locale(decimal_mark = ","))

readr দিয়ে ডাটা সংরক্ষণ করুন

write_*(x, file, na = "NA", append, col_names, quote, escape, eol, num_threads, progress)



write_delim(x, file, delim = " ") যেকোনো ডিলিমিটার দিয়ে ফাইল লিখুন
write_csv(x, file) কমা ডিলিমিটার দিয়ে ফাইল লিখুন
write_csv2(x, file) সেমিকোলন ডিলিমিটার দিয়ে ফাইল লিখুন
write_tsv(x, file) ট্যাব ডিলিমিটার দিয়ে ফাইল লিখুন

যেকোনো প্রকল্পের অন্যতম প্রথম ধাপ হলে বাইরের ডেটা R-এ আমদানি করা. ডেটা বেশিরভাগ সময় ট্যাবুলার বিন্যাসে থাকে। যেমনঃ csv ফাইল বা স্প্রেডশিট



এই শিট এর প্রথম পৃষ্ঠায় থাকছে কিভাবে readr এর মাধ্যমে text ফাইল পড়া এবং লেখা যায়।



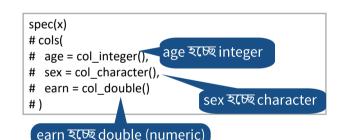
পরের পৃষ্ঠায়ে থাকছে কিভাবে readxl দিয়ে Excel ফাইল পড়া যায় এবং googlesheets4 দিয়ে Google Sheets পড়া যায় অন্যান্য প্রকারের ডেটা নিম্নোক্ত প্যাকেজগুলো ব্যবহার করুন:

- haven SPSS, Stata, এবং SAS ফাইল
- DBI ডেটাবেজ
- jsonlite json
- xml2 XML
- httr-Web APIs
- rvest HTML (ওয়েব স্ক্রাপিং)
- readr::read lines() text ডেটা

readr এর মাধ্যমে কলাম নির্দিষ্টকরণ

কলাম নির্দিষ্টকরণ এর মাধ্যমে ঠিক করা হয় আমদানিকৃত কলামগুলোর ডেটা এর প্রকার কি হবে। গতানুগতিক ভাবে readr একটি কলাম নির্দিষ্টকরণ সৃষ্টি করবে এবং সারসংক্ষেপ আউতপুট আকারে দেখাবে।

spec(x) আমদানিকৃত data frame এর সম্পূর্ণ কলাম নির্দিষ্টকরণ বের করুন।



কলামের প্রকার

প্রত্যেক কলাম প্রকারের একটা ফাংশন এবং string সংক্ষেপ আছে।

- col_logical() "l"
- col_integer() "i"
- col_double() "d"
- col_number() "n"
- col character() "c"
- col_factor(levels, ordered = FALSE) "f"
- col datetime(format = "") "T"
- col date(format = "") "D"
- col_time(format = "") "t"
- col_skip() "-", "_"
- col guess() "?"

দরকারি কলাম আর্গুমেন্ট কলাম নির্দিষ্টকরণ বার্তা লুকান read *(file, show col types = FALSE)

আমদানি করার জন্য কলাম নির্বাচন করুণ নাম, অবস্থান, অথবা selection helpers ব্যবহার করুন read_*(file, col_select = c(age, earn))

কলাম এর প্রকার অনুমান করুন কলান এর প্রকার অনুমান করতে read_*() প্রথম 1000 সারির ডেটা দেখে। guess_max ব্যবহার করে এটা বাড়াতে পারেন।

read_*(file, guess_max = Inf)

col type = "?ilc"

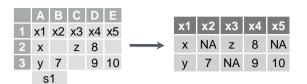
```
কলাম নির্দিষ্টকরণ নির্ধারণ
গতানুগতিক প্রকার নির্ধারণ করুন
read_csv(
file,
col_type = list(.default = col_double())
)
কলাম প্রকার অথবা string সংক্ষেপ ব্যবহার করুন
read_csv(
file,
col_type = list(x = col_double(), y = "l", z = "_")
)
একক string সংক্ষেপ ব্যবহার করুন
# col types: skip, guess, integer, logical, character
read_csv(
file,
```



স্প্রেডশিট আমদানি

readxl এর মাধ্যমে

EXCEL ফাইল পড়ন



read_excel(path, sheet = NULL, range = NULL) .xls অথবা .xlsx ফাইল পড়ন ফাইল extension এর উপর ভিত্তি করে। আরো reåd আর্গুমেন্ট এর জন্য প্রথম পৃষ্ঠা দেখুন। এছাড়া read_xls() এবং read_xlsx(). read excel("excel file.xlsx")

শিট পড়ন



read excel(path, sheet = NULL) কোন শিট পড়তে হবে তা অবস্থান অথবা নাম দিয়ে নির্ধারণ করুন।

read_excel(path, sheet = 1) read excel(path, sheet = "s1")



excel_sheets(path) শিটগুলোর নামের vector পান

excel sheets("excel file.xlsx")



একাধিক শিট পড়নঃ

- 1. File path ব্যবহার করে শিটগুলোর নামের vector পান
- 2. vector নামগুলো শিট এর নাম হিসেবে নির্ধারণ করুন
- 3. purrr::map dfr() ব্যবহার করে একাধিক ফাইল এক data frame এ আমদানি করুন

path <- "your_file_path.xlsx" path %>% excel sheets() %>% set names() %>% map dfr(read excel, path = path)

অন্যান্য দরকারি EXCEL প্যাকেজ

Excel ফাইল এ ডেটা লেখার জন্য দেখুনঃ

- openxlsx
- writexl

নন-টাবুলার Excel ডেটা এর জন্য দেখুনঃ

tidvxl



READXL এর কলাম নির্দিষ্টকরণ

কলাম নির্দিষ্টকরণ এর মাধ্যমে ঠিক করা হয় আমদানিকৃত কলামগুলোর ডেটা এর প্রকার কি হবে। read_excel() এর col_types আর্গুমেন্ট ব্যবহার করে কলাম নির্দিষ্টকরণ নির্ধারণ করুন।

কলাম এর প্রকার অনুমান করুন

কলান এর প্রকার অনুমান করতে read excel() প্রথম 1000 সারির ডেটা দেখে। guess max ব্যবহার করে এটা বাডাতে

read excel(path, guess max = Inf)

সকল কলামকে একই প্রকারের করুন। যেমনঃ character

read excel(path.col types = "text")

প্রত্যেক কলাম আলাদা ভাবে নির্ধারণ করুন

read excel(col types = c("text", "guess", "guess", "numeric")

কলাম এর প্রকার

logical	numeric	text	date	list
TRUE	2	hello	1947-01-08	hello
FALSE	3.45	world	1956-10-21	1

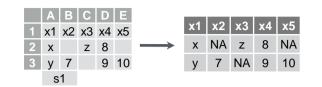
- skip
- logical
- date

- guess
- numeric
 list
- text

যেসব কলামে একাধিক প্রকারের ডেটা আছে, সেখানে list list-column ডেটা এর জন্য tidyr এবং purrr দেখুন।

googlesheets4 এর মাধ্যমে

শিট পড়ন



read_sheet(ss, sheet = NULL, range = NULL) URL, শিট ID অথবা googledrive প্যাকেজ এর dribble থেকে শিট পড়ন।প্রথম পূর্চা দেখুন আরও read আর্গুমেন্ট এর জন্য। range read() এর মতই।

শিট এর METADATA

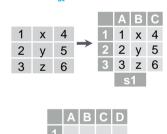
URI এর গঠনঃ

https://docs.google.com/spreadsheets/d/ SPREADSHEET_ID/edit#gid=SHEET_ID

gs4 get(ss) স্প্রেডশিট এর meta data পান gs4 find(...) সব স্প্রেডশিট ফাইল গুলোর ডেটা পান

sheet_properties(ss) প্রত্যেক worksheet এর বৈশিষ্ট্য সম্বলিত tibble পান। এছাড়া sheet names()

শিটে লিখন



2

2 y 5

3 z 6

write sheet(data, ss = NULL, sheet = NULL) নতন অথবা বিদ্যমান শিটে data frame লিখন

gs4_create(name, ..., sheets = NULL) নামের vector,data frame, অথবা a (নাম সম্বলিত) data frame এর list দিয়ে নতুন শিট বানান

sheet_append(ss, data, sheet = 1) worksheet এর শেষে একটি সারি যুক্ত করুন

GOOGLESHEETS4 এর কলাম নির্দিষ্টকরণ

কলাম নির্দিষ্টকরণ এর মাধ্যমে ঠিক করা হয় আমদানিকৃত কলামগুলোর ডেটা এর প্রকার কি হবে।

googlesheets

read_sheet()/range_read() এর col_types আর্গুমেন্ট ব্যবহার করে কলাম নির্দিষ্টকরণ নির্ধারণ করুন।

কলাম এর প্রকার অনুমান করুন কলান এর প্রকার অনুমান করতে read sheet()/range read() প্রথম 1000 সারির ডেটা দেখে। guess max ব্যবহার করে এটা বাডাতে পারেন। read sheet(path, guess max = Inf)

সকল কলামকে একই প্রকারের করুন। যেমনঃ character

read sheet(path, col types = "c")

প্রত্যেক কলাম আলাদা ভাবে নির্ধারণ করুন

col types: skip, guess, integer, logical, character read_sheets(ss, col_types = "?ilc")

কলাম এর প্রকার

I	n	С	D	L	
TRUE	2	hello	1947-01-08	hello	
FALSE	3.45	world	1956-10-21	1	
skip -guesslogicainteger	- "?" ıl - "l"	•	 datetime - "T" character - "c" list-column - "L" cell - "C" Returns 		

double - "d" • numeric - "n"

list of raw cell data.

• date - "D"

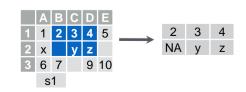
যেসব কলামে একাধিক প্রকারের ডেটা আছে, সেখানে list ব্যবহার করুন। list-column ডেটা এর জন্য tidyr এবং purrr দেখুন।

ফাইল স্তরীয় অপারেশান

এছাড়াও googlesheets4 শিটের বিভিন্ন দিক পরিবর্তন করার উপায় প্রদান করে। (যেমনঃ সারি freeze করা, কলামের প্রস্থ নির্ধারণ করা. (work)sheets বাবস্থাপনা করা) l আরও জানতে googlesheets4.tidyverse.org এ যানí

সম্পূর্ণ ফাইল অপারেশানের (যেমনঃ rename করা, share করা, folder এ রাখা) জন্য tidyverse এর অন্তর্ভক্ত প্যাকেজ googledrive দেখন googledrive.tidyverse.org.

READXL এবং GOOGLESHEETS4 এর জন্য কলাম নির্দিষ্টকরণ



readxl::read_excel() অথবা googlesheets4::read_sheet() এর range আর্গুমেন্ট ব্যবহার করে একটি শিট এর cell গুলোর উপসেট পড়ন read excel(path, range = "Sheet1!B1:D2") read sheet(ss, range = "B1:D2")

ABC

1 x1 x2 x3

3 2 y 5

→ 2 1 x 4

এছাড়া range আর্গুমেন্ট কলাম নির্দিষ্টকরণ ফাংশন এর সাথে ব্যবহার করুন cell limits(), cell rows(), cell cols(), and anchored().

