

# Package ‘excelstrippr’

January 7, 2021

**Type** Package

**Title** Extracts Tabular Data from Excel Reports

**Version** 0.1.2

**Description** Removes metadata and summary formatting from Excel reports, so that data from those reports can more easily be loaded into R. Currently works with .xlsx and .xls format files.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**Imports** tidyr, dplyr, readxl

**RoxygenNote** 7.1.1

**NeedsCompilation** no

**Author** Christopher Burch [aut, cre] (<<https://orcid.org/0000-0001-6934-3325>>)

**Maintainer** Christopher Burch <cmburch@cmburch.me>

**Repository** CRAN

**Date/Publication** 2021-01-07 18:10:18 UTC

## R topics documented:

extract_data . . . . .	2
percent_missing . . . . .	3
promote_colnames . . . . .	3
strip_metadata . . . . .	4
strip_na_percent . . . . .	4

<b>Index</b>	<b>6</b>
--------------	----------

---

extract_data	<i>Extract Data from a Report</i>
--------------	-----------------------------------

---

## Description

Extract Data from a Report

## Usage

```
extract_data(  
  .df,  
  checkcol = NA,  
  promote_colnames = TRUE,  
  strip_na_greater_than = 0.75  
)
```

## Arguments

`.df` a data frame object.

`checkcol` numeric. The column number to check for NA values. If not specified, `extract_data` will attempt to guess the appropriate column.

`promote_colnames` logical. Should the first row be promoted to column names?

`strip_na_greater_than` double between 0 - 1. Drops rows with greater than this percentage of missing values. Defaults to 0.75, or 75 percent or greater missing values

## Value

a data frame object.

## Examples

```
df <- data.frame(`...1` = c("Excel Report", NA, NA, NA),  
                 `...2` = c(NA, "X", "x_1", "x_2"),  
                 `...3` = c(NA, "Y", "y_1", "y_2"),  
                 `...4` = c(NA, "Z", "z_1", "z_2"))  
extract_data(df)
```

---

percent_missing	<i>Determine Row Percentage NA Values</i>
-----------------	---

---

**Description**

Determine Row Percentage NA Values

**Usage**

```
percent_missing(x)
```

**Arguments**

x                    a vector

**Value**

a numeric vector containing the percentage of NA values in x

---

promote_colnames	<i>Promote Row to Column Names</i>
------------------	------------------------------------

---

**Description**

Promote Row to Column Names

**Usage**

```
promote_colnames(.df, rownum = 1)
```

**Arguments**

.df                    a data frame  
rownum                numeric. The row in the data frame to be promoted to column names

**Value**

a data frame with the promoted row as column names

**Examples**

```
df <- data.frame(v1 = c("X", "123.45"), v2 = c("Y", "345.67"))  
promote_colnames(df)
```

---

strip_metadata	<i>Strip Report MetaData From Excel File</i>
----------------	--

---

**Description**

Strip Report MetaData From Excel File

**Usage**

```
strip_metadata(from, promote = NA, ...)
```

**Arguments**

from	the file path to the excel report.
promote	logical. Should first non-NA row be promoted to column names? Default tries to make a reasonable determination.
...	extra parameters to pass to the read_excel function

**Value**

a data frame containing the tabular data from the report

**Examples**

```
## Not run:
strip_metadata("../man/example/example-report")
# specify a particular sheet
strip_metadata("my-excel-file.xlsx", sheet = "report summary")

## End(Not run)
```

---

strip_na_percent	<i>Remove Rows of a Data Frame by Percent NA Values</i>
------------------	---

---

**Description**

Remove Rows of a Data Frame by Percent NA Values

**Usage**

```
strip_na_percent(.df, greater_than = 0.25)
```

**Arguments**

.df	a data.frame object
greater_than	double. The cutoff percent of NA values in a row to remove that row. Defaults to 0.25

**Value**

a data frame with specified rows removed.

**Examples**

```
df <- data.frame(A = runif(5), B = runif(5), C = runif(5))
df$B[3:4] <- NA
df$C[4] <- NA
strip_na_percent(df, greater_than = .5)
```

# Index

`extract_data`, [2](#)

`percent_missing`, [3](#)

`promote_colnames`, [3](#)

`strip_metadata`, [4](#)

`strip_na_percent`, [4](#)