



Data import with the tidyverse :: CHEAT SHEET

Read Tabular Data with readr

```
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
```

A B C	1 2 3	4 5 NA
A	B	C
1	2	3
4	5	NA

read_delim("file.txt", delim = "|") Read files with any delimiter. If no delimiter is specified, it will automatically guess.

To make file.txt, run: `write_file("A|B|C\n1|2|3\n4|5|NA", file = "file.txt")`

A,B,C	1,2,3	4,5,NA
A	B	C
1	2	3
4	5	NA

read_csv("file.csv") Read a comma delimited file with period decimal marks.

`write_file("A,B,C\n1,2,3\n4,5,NA", file = "file.csv")`

A;B;C	1;5;2;3	4;5;5;NA
A	B	C
1	5	2
4	5	NA

read_csv2("file2.csv") Read semicolon delimited files with comma decimal marks.

`write_file("A;B;C\n1,5;2;3\n4,5;5;NA", file = "file2.csv")`

A B C	1 2 3	4 5 NA
A	B	C
1	2	3
4	5	NA

read_tsv("file.tsv") Read a tab delimited file. Also **read_table()**.

read_fwf("file.tsv", fwf_widths(c(2, 2, NA))) Read a fixed width file.

`write_file("A\tB\tC\n1\t2\t3\n4\t5\tNA", file = "file.tsv")`

USEFUL READ ARGUMENTS

A	B	C
1	2	3
4	5	NA

No header
`read_csv("file.csv", col_names = FALSE)`

x	y	z
A	B	C
1	2	3
4	5	NA

Provide header

`read_csv("file.csv",
col_names = c("x", "y", "z"))`

A	B	C
NA	2	3
4	5	NA

Read multiple files into a single table

`read_csv(c("f1.csv", "f2.csv", "f3.csv"),
id = "origin_file")`

Save Data with readr

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

A	B	C
1	2	3
4	5	NA

write_delim(x, file, delim = " ") Write files with any delimiter.

write_csv(x, file) Write a comma delimited file.

write_csv2(x, file) Write a semicolon delimited file.

write_tsv(x, file) Write a tab delimited file.

One of the first steps of a project is to import outside data into R. Data is often stored in tabular formats, like csv files or spreadsheets.



The front page of this sheet shows how to import and save text files into R using **readr**.



The back page shows how to import spreadsheet data from Excel files using **readxl** or Google Sheets using **googlesheets4**.

OTHER TYPES OF DATA

Try one of the following packages to import other types of files:

- **haven** - SPSS, Stata, and SAS files
- **DBI** - databases
- **jsonlite** - json
- **xml2** - XML
- **httr** - Web APIs
- **rvest** - HTML (Web Scraping)
- **readr::read_lines()** - text data

Column Specification with readr

Column specifications define what data type each column of a file will be imported as. By default **readr** will generate a column spec when a file is read and output a summary.

spec(x) Extract the full column specification for the given imported data frame.

```
spec(x)
# cols(
#   age = col_integer(),
#   sex = col_character(),
#   earn = col_double()
# )
```

age is an integer

sex is a character

earn is a double (numeric)

COLUMN TYPES

Each column type has a function and corresponding string abbreviation.

- **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()** - "?"

DEFINE COLUMN SPECIFICATION

Set a default type

```
read_csv(
  file,
  col_type = list(.default = col_double())
)
```

Use column type or string abbreviation

```
read_csv(
  file,
  col_type = list(x = col_double(), y = "l", z = "_")
)
```

Use a single string of abbreviations

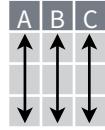
```
# col types: skip, guess, integer, logical, character
read_csv(
  file,
  col_type = "_?ilc"
)
```


Data tidying with `tidyr` :: CHEAT SHEET



Tidy data is a way to organize tabular data in a consistent data structure across packages.

A table is tidy if:



Each **variable** is in its own **column**

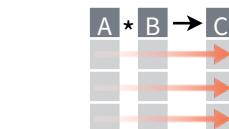
&



Each **observation**, or **case**, is in its own row



Access **variables** as **vectors**



Preserve **cases** in vectorized operations

Tibbles

AN ENHANCED DATA FRAME

Tibbles are a table format provided by the **tibble** package. They inherit the data frame class, but have improved behaviors:

- **Subset** a new tibble with `[`, a vector with `[[` and `$`.
- **No partial matching** when subsetting columns.
- **Display** concise views of the data on one screen.

`options(tibble.print_max = n, tibble.print_min = m, tibble.width = Inf)` Control default display settings.

`View()` or `glimpse()` View the entire data set.

CONSTRUCT A TIBBLE

tibble(...) Construct by columns.

`tibble(x = 1:3, y = c("a", "b", "c"))`

Both make this tibble

A tibble: 3 × 2
`x` <int> <chr>
 1 1 a
 2 2 b
 3 3 c

as_tibble(x, ...) Convert a data frame to a tibble.

enframe(x, name = "name", value = "value")

Convert a named vector to a tibble. Also `deframe()`.

is_tibble(x) Test whether x is a tibble.

Reshape Data

- Pivot data to reorganize values into a new layout.

table4a

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K



country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K

table2

country	year	type	count
A	1999	cases	0.7K
A	1999	pop	19M
A	2000	cases	2K
A	2000	pop	20M
B	1999	cases	37K
B	1999	pop	172M
B	2000	cases	80K
B	2000	pop	174M
C	1999	cases	212K
C	1999	pop	1T
C	2000	cases	213K
C	2000	pop	1T



country	year	cases	pop
A	1999	0.7K	19M
A	2000	2K	20M
B	1999	37K	172M
B	2000	80K	174M
C	1999	212K	1T
C	2000	213K	1T

Split Cells

- Use these functions to split or combine cells into individual, isolated values.

table5

country	century	year
A	19	99
A	20	00
B	19	99
B	20	00



country	year
A	1999
A	2000
B	1999
B	2000

table3

country	year	rate
A	1999	0.7K/19M
A	2000	2K/20M
B	1999	37K/172M
B	2000	80K/174M



country	year	cases	pop
A	1999	0.7K	19M
A	2000	2K	20M
B	1999	37K	172M
B	2000	80K	174M

table3

country	year	rate
A	1999	0.7K/19M
A	2000	2K/20M
B	1999	37K/172M
B	2000	80K/174M



country	year	rate
A	1999	0.7K
A	1999	19M
A	2000	2K
A	2000	20M
B	1999	37K
B	1999	172M
B	2000	80K
B	2000	174M

pivot_longer(data, cols, names_to = "name", values_to = "value", values_drop_na = FALSE)

"Lengthen" data by collapsing several columns into two. Column names move to a new names_to column and values to a new values_to column.

```
pivot_longer(table4a, cols = 2:3, names_to = "year", values_to = "cases")
```

pivot_wider(data, names_from = "name", values_from = "value")

The inverse of pivot_longer(). "Widen" data by expanding two columns into several. One column provides the new column names, the other the values.

```
pivot_wider(table2, names_from = type, values_from = count)
```

Expand Tables

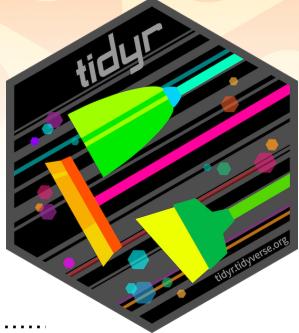
Create new combinations of variables or identify implicit missing values (combinations of variables not present in the data).

x	x1	x2	x3
A	1	3	
B	1	4	
B	2	3	

x	x1	x2	x3
A	1	3	
B	1	4	
B	2	3	
B	2	3	

expand(data, ...) Create a new tibble with all possible combinations of the values of the variables listed in ...

Drop other variables.
`expand(mtcars, cyl, gear, carb)`



Nested Data

A **nested data frame** stores individual tables as a list-column of data frames within a larger organizing data frame. List-columns can also be lists of vectors or lists of varying data types.

Use a nested data frame to:

- Preserve relationships between observations and subsets of data. Preserve the type of the variables being nested (factors and datetimes aren't coerced to character).
- Manipulate many sub-tables at once with **purrr** functions like `map()`, `map2()`, or `pmap()` or with **dplyr** `rowwise()` grouping.

CREATE NESTED DATA

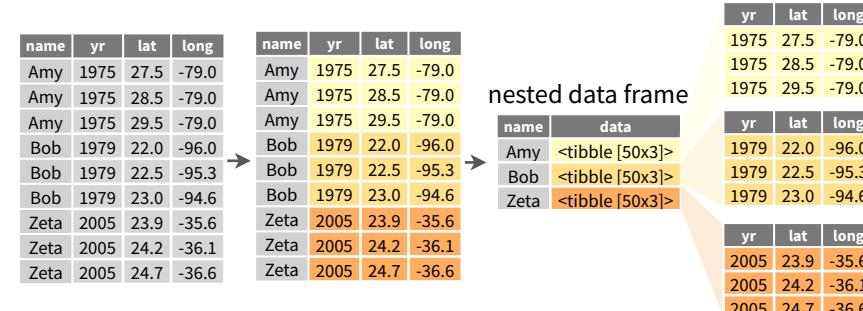
nest(data, ...) Moves groups of cells into a list-column of a data frame. Use alone or with `dplyr::group_by()`:

1. Group the data frame with `group_by()` and use `nest()` to move the groups into a list-column.

```
n_storms <- storms %>%
  group_by(name) %>%
  nest()
```

2. Use `nest(new_col = c(x, y))` to specify the columns to group using `dplyr::select()` syntax.

```
n_storms <- storms %>%
  nest(data = c(year:long))
```



Index list-columns with `[[[]]]`. `n_storms$data[[1]]`

CREATE TIBBLES WITH LIST-COLUMNS

tibble::tribble(...) Makes list-columns when needed.

```
tribble(~max, ~seq,
       3, 1:3,
       4, 1:4,
       5, 1:5)
```

tibble::tibble(...) Saves list input as list-columns.

```
tibble(max = c(3, 4, 5), seq = list(1:3, 1:4, 1:5))
```

tibble::enframe(x, name="name", value="value")

Converts multi-level list to a tibble with list-cols.
`enframe(list('3'=1:3, '4'=1:4, '5'=1:5), 'max', 'seq')`

OUTPUT LIST-COLUMNS FROM OTHER FUNCTIONS

dplyr::mutate(), transmute(), and summarise() will output list-columns if they return a list.

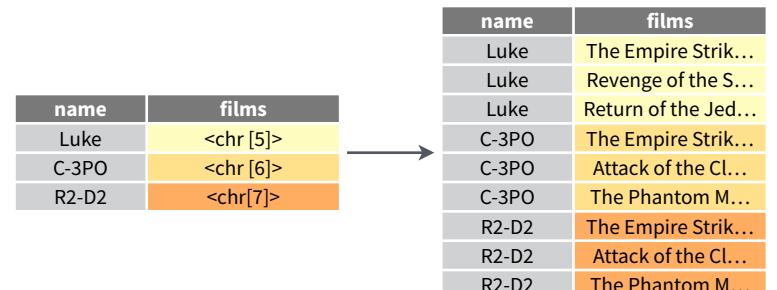
```
mtcars %>%
  group_by(cyl) %>%
  summarise(q = list(quantile(mpg)))
```

RESHAPE NESTED DATA

unnest(data, cols, ..., keep_empty = FALSE) Flatten nested columns back to regular columns. The inverse of `nest()`.
`n_storms %>% unnest(data)`

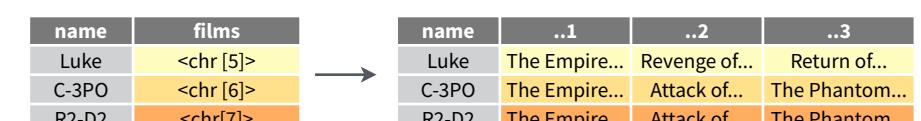
unnest_longer(data, col, values_to = NULL, indices_to = NULL)
Turn each element of a list-column into a row.

```
starwars %>%
  select(name, films) %>%
  unnest_longer(films)
```



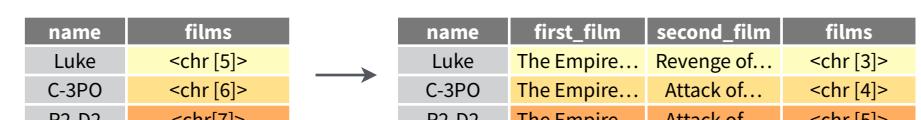
unnest_wider(data, col) Turn each element of a list-column into a regular column.

```
starwars %>%
  select(name, films) %>%
  unnest_wider(films)
```



hoist(.data, .col, ..., .remove = TRUE) Selectively pull list components out into their own top-level columns. Uses `purrr::pluck()` syntax for selecting from lists.

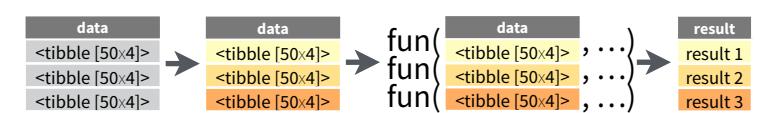
```
starwars %>%
  select(name, films) %>%
  hoist(films, first_film = 1, second_film = 2)
```



TRANSFORM NESTED DATA

A vectorized function takes a vector, transforms each element in parallel, and returns a vector of the same length. By themselves vectorized functions cannot work with lists, such as list-columns.

dplyr::rowwise(.data, ...) Group data so that each row is one group, and within the groups, elements of list-columns appear directly (accessed with `[]`, not as lists of length one. **When you use `rowwise()`, dplyr functions will seem to apply functions to list-columns in a vectorized fashion.**



Apply a function to a list-column and **create a new list-column**.

`n_storms %>%`
`rowwise() %>%`
`mutate(n = list(dim(data)))`

`dim()` returns two values per row

wrap with `list` to tell `mutate` to create a list-column

Apply a function to a list-column and **create a regular column**.

`n_storms %>%`
`rowwise() %>%`
`mutate(n = nrow(data))`

`nrow()` returns one integer per row

Collapse **multiple list-columns** into a single list-column.

`starwars %>%`
`rowwise() %>%`
`mutate(transport = list(append(vehicles, starships)))`

`append()` returns a list for each row, so col type must be list

Apply a function to **multiple list-columns**.

`starwars %>%`
`rowwise() %>%`
`mutate(n_transports = length(c(vehicles, starships)))`

`length()` returns one integer per row

See **purrr** package for more list functions.

Data transformation with dplyr :: CHEAT SHEET



dplyr functions work with pipes and expect **tidy data**. In tidy data:



Each **variable** is in its own **column**



Each **observation**, or **case**, is in its own **row**



`x %>% f(y)` becomes `f(x, y)`

Summarise Cases

Apply **summary functions** to columns to create a new table of summary statistics. Summary functions take vectors as input and return one value (see back).



`summarise(.data, ...)`
Compute table of summaries.
`summarise(mtcars, avg = mean(mpg))`

`count(.data, ..., wt = NULL, sort = FALSE, name = NULL)` Count number of rows in each group defined by the variables in ... Also **tally()**.
`count(mtcars, cyl)`

Group Cases

Use **group_by(.data, ..., .add = FALSE, .drop = TRUE)** to create a "grouped" copy of a table grouped by columns in ... dplyr functions will manipulate each "group" separately and combine the results.

`mtcars %>% group_by(cyl) %>% summarise(avg = mean(mpg))`

Use **rowwise(.data, ...)** to group data into individual rows. dplyr functions will compute results for each row. Also apply functions to list-columns. See tidyverse cheat sheet for list-column workflow.

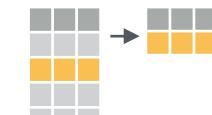
`starwars %>% rowwise() %>% mutate(film_count = length(films))`

ungroup(x, ...) Returns ungrouped copy of table.
`ungroup(g_mtcars)`

Manipulate Cases

EXTRACT CASES

Row functions return a subset of rows as a new table.



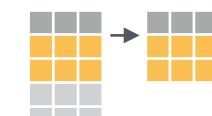
filter(.data, ..., .preserve = FALSE) Extract rows that meet logical criteria.
`filter(mtcars, mpg > 20)`



distinct(.data, ..., .keep_all = FALSE) Remove rows with duplicate values.
`distinct(mtcars, gear)`



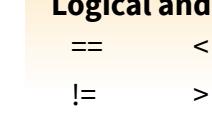
slice(.data, ..., .preserve = FALSE) Select rows by position.
`slice(mtcars, 10:15)`



slice_sample(.data, ..., n, prop, weight_by = NULL, replace = FALSE) Randomly select rows. Use n to select a number of rows and prop to select a fraction of rows.
`slice_sample(mtcars, n = 5, replace = TRUE)`



slice_min(.data, order_by, ..., n, prop, with_ties = TRUE) and **slice_max()** Select rows with the lowest and highest values.
`slice_min(mtcars, mpg, prop = 0.25)`



slice_head(.data, ..., n, prop) and **slice_tail()** Select the first or last rows.
`slice_head(mtcars, n = 5)`

Logical and boolean operators to use with filter()

<code>==</code>	<code><</code>	<code><=</code>	<code>is.na()</code>	<code>%in%</code>	<code> </code>	<code>xor()</code>
<code>!=</code>	<code>></code>	<code>>=</code>	<code>!is.na()</code>	<code>!</code>	<code>&</code>	

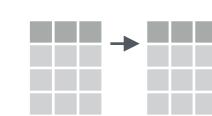
See [?base::Logic](#) and [?Comparison](#) for help.

ARRANGE CASES



arrange(.data, ..., .by_group = FALSE) Order rows by values of a column or columns (low to high), use with **desc()** to order from high to low.
`arrange(mtcars, mpg)`
`arrange(mtcars, desc(mpg))`

ADD CASES



add_row(.data, ..., .before = NULL, .after = NULL) Add one or more rows to a table.
`add_row(cars, speed = 1, dist = 1)`

Manipulate Variables

EXTRACT VARIABLES

Column functions return a set of columns as a new vector or table.



pull(.data, var = -1, name = NULL, ...) Extract column values as a vector, by name or index.
`pull(mtcars, wt)`



select(.data, ...) Extract columns as a table.
`select(mtcars, mpg, wt)`



relocate(.data, ..., .before = NULL, .after = NULL) Move columns to new position.
`relocate(mtcars, mpg, cyl, .after = last_col())`

Use these helpers with select() and across()

e.g. `select(mtcars, mpg:cyl)`

<code>contains(match)</code>	<code>num_range(prefix, range)</code>	: e.g. <code>mpg:cyl</code>
<code>ends_with(match)</code>	<code>all_of(x)/any_of(x, ..., vars)</code>	- e.g. <code>-gear</code>
<code>starts_with(match)</code>	<code>matches(match)</code>	everything()

MANIPULATE MULTIPLE VARIABLES AT ONCE



across(.cols, .funs, ..., .names = NULL) Summarise or mutate multiple columns in the same way.
`summarise(mtcars, across(everything(), mean))`



c_across(.cols) Compute across columns in row-wise data.
`transmute(rowwise(UKgas), total = sum(c_across(1:2)))`

MAKE NEW VARIABLES

Apply **vectorized functions** to columns. Vectorized functions take vectors as input and return vectors of the same length as output (see back).



mutate(.data, ..., .keep = "all", .before = NULL, .after = NULL) Compute new column(s). Also **add_column()**, **add_count()**, and **add_tally()**.
`mutate(mtcars, gpm = 1 / mpg)`



transmute(.data, ...) Compute new column(s), drop others.
`transmute(mtcars, gpm = 1 / mpg)`



rename(.data, ...) Rename columns. Use **rename_with()** to rename with a function.
`rename(cars, distance = dist)`



Vectorized Functions

TO USE WITH MUTATE ()

mutate() and **transmute()** apply vectorized functions to columns to create new columns. Vectorized functions take vectors as input and return vectors of the same length as output.

vectorized function →

OFFSET

dplyr::lag() - offset elements by 1
dplyr::lead() - offset elements by -1

CUMULATIVE AGGREGATE

dplyr::cumall() - cumulative all()
dplyr::cumany() - cumulative any()
 cummax() - cumulative max()
dplyr::cummean() - cumulative mean()
 cummin() - cumulative min()
 cumprod() - cumulative prod()
 cumsum() - cumulative sum()

RANKING

dplyr::cume_dist() - proportion of all values <=
dplyr::dense_rank() - rank w ties = min, no gaps
dplyr::min_rank() - rank with ties = min
dplyr::ntile() - bins into n bins
dplyr::percent_rank() - min_rank scaled to [0,1]
dplyr::row_number() - rank with ties = "first"

MATH

+, -, *, /, ^, %/%, %% - arithmetic ops
log(), log2(), log10() - logs
<, <=, >, >=, !=, == - logical comparisons
dplyr::between() - x >= left & x <= right
dplyr::near() - safe == for floating point numbers

MISCELLANEOUS

dplyr::case_when() - multi-case if_else()
starwars %>%
 mutate(type = case_when(
 height > 200 | mass > 200 ~ "large",
 species == "Droid" ~ "robot",
 TRUE ~ "other")
)
dplyr::coalesce() - first non-NA values by element across a set of vectors
dplyr::if_else() - element-wise if() + else()
dplyr::na_if() - replace specific values with NA
 pmax() - element-wise max()
 pmin() - element-wise min()

Summary Functions

TO USE WITH SUMMARISE ()

summarise() applies summary functions to columns to create a new table. Summary functions take vectors as input and return single values as output.

summary function →

COUNT

dplyr::n() - number of values/rows
dplyr::n_distinct() - # of uniques
 sum(!is.na()) - # of non-NAs

POSITION

mean() - mean, also **mean(!is.na())**
median() - median

LOGICAL

mean() - proportion of TRUE's
sum() - # of TRUE's

ORDER

dplyr::first() - first value
dplyr::last() - last value
dplyr::nth() - value in nth location of vector

RANK

quantile() - nth quantile
min() - minimum value
max() - maximum value

SPREAD

IQR() - Inter-Quartile Range
mad() - median absolute deviation
sd() - standard deviation
var() - variance

Row Names

Tidy data does not use rownames, which store a variable outside of the columns. To work with the rownames, first move them into a column.

A	B	C	A	B
1	a	t	1	t
2	b	u	2	u
3	c	v	3	NA

tibble::rownames_to_column()
Move row names into col.
a <- rownames_to_column(mtcars, var = "C")

A	B	C	A	B
1	a	t	1	a
2	b	u	2	b
3	c	v	3	c

tibble::column_to_rownames()
Move col into row names.
column_to_rownames(a, var = "C")

Also **tibble::has_rownames()** and **tibble::remove_rownames()**.

Combine Tables

COMBINE VARIABLES

X	y	=
A B C	E F G	
a t 1	a t 3	
b u 2	b u 2	
c v 3	d w 1	

bind_cols(..., .name_repair) Returns tables placed side by side as a single table. Column lengths must be equal. Columns will NOT be matched by id (to do that look at Relational Data below), so be sure to check that both tables are ordered the way you want before binding.

RELATIONAL DATA

Use a "Mutating Join" to join one table to columns from another, matching values with the rows that they correspond to. Each join retains a different combination of values from the tables.

A B C D	left_join(x, y, by = NULL, copy = FALSE, suffix = c(".x", ".y"), ..., keep = FALSE, na_matches = "na")
a t 1 3	Join matching values from y to x.

A B C D	right_join(x, y, by = NULL, copy = FALSE, suffix = c(".x", ".y"), ..., keep = FALSE, na_matches = "na")
a t 1 3	Join matching values from x to y.

A B C D	inner_join(x, y, by = NULL, copy = FALSE, suffix = c(".x", ".y"), ..., keep = FALSE, na_matches = "na")
a t 1 3	Join data. Retain only rows with matches.

A B C D	full_join(x, y, by = NULL, copy = FALSE, suffix = c(".x", ".y"), ..., keep = FALSE, na_matches = "na")
a t 1 3	Join data. Retain all values, all rows.

COLUMN MATCHING FOR JOINS

A B x C B y D
a t 1 t 3
b u 2 u 2
c v 3 NA NA

Use **by = c("col1", "col2", ...)** to specify one or more common columns to match on.
left_join(x, y, by = "A")

A x B x C A y B y
a t 1 d w
b u 2 b u
c v 3 a t

Use a named vector, **by = c("col1" = "col2")**, to match on columns that have different names in each table.
left_join(x, y, by = c("C" = "D"))

A1 B1 C A2 B2
a t 1 d w
b u 2 b u
c v 3 a t

Use **suffix** to specify the suffix to give to unmatched columns that have the same name in both tables.
left_join(x, y, by = c("C" = "D"), suffix = c("1", "2"))

COMBINE CASES

X	y	=
A B C	A B C	
a t 1	c v 3	

bind_rows(..., id = NULL)
Returns tables one on top of the other as a single table. Set **.id** to a column name to add a column of the original table names (as pictured).

X	y	=
A B C	A B C	
a t 1	x a t 1	

Use a "Filtering Join" to filter one table against the rows of another.

X	y	=
A B C	A B D	
a t 1	a t 3	

semi_join(x, y, by = NULL, copy = FALSE, ..., na_matches = "na") Return rows of x that have a match in y. Use to see what will be included in a join.

X	y	=
A B C	A B C	
c v 3	NA NA	

Use a "Nest Join" to inner join one table to another into a nested data frame.

X	y	=
A B C	<tibble [1x2]>	
a t 1	a t 1	

nest_join(x, y, by = NULL, copy = FALSE, keep = FALSE, name = NULL, ...) Join data, nesting matches from y in a single new data frame column.

SET OPERATIONS

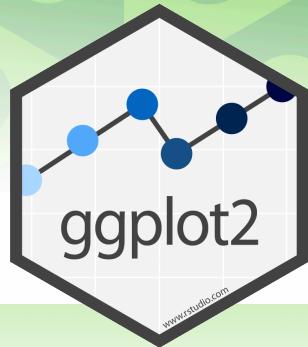
A B C	c v 3
a t 1	a t 1

intersect(x, y, ...)
Rows that appear in both x and y.

A B C	a t 1
a t 1	a t 1

setdiff(x, y, ...)
Rows that appear in x but not y.

Data visualization with ggplot2 :: CHEAT SHEET



Basics

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data** set, a **coordinate system**, and **geoms**—visual marks that represent data points.



To display values, map variables in the data to visual properties of the geom (**aesthetics**) like **size**, **color**, and **x** and **y** locations.



Complete the template below to build a graph.

```
ggplot (data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>),
  stat = <STAT>, position = <POSITION>) +
  <COORDINATE_FUNCTION> +
  <FACET_FUNCTION> +
  <SCALE_FUNCTION> +
  <THEME_FUNCTION>
```

required
Not required, sensible defaults supplied

`ggplot(data = mpg, aes(x = cty, y = hwy))` Begins a plot that you finish by adding layers to. Add one geom function per layer.

`last_plot()` Returns the last plot.

`ggsave("plot.png", width = 5, height = 5)` Saves last plot as 5' x 5' file named "plot.png" in working directory. Matches file type to file extension.

Aes Common aesthetic values.

color and **fill** - string ("red", "#RRGGBB")

linetype - integer or string (0 = "blank", 1 = "solid", 2 = "dashed", 3 = "dotted", 4 = "dotdash", 5 = "longdash", 6 = "twodash")

lineend - string ("round", "butt", or "square")

linejoin - string ("round", "mitre", or "bevel")

size - integer (line width in mm)

shape - integer/shape name or a single character ("a")


Geoms

Use a geom function to represent data points, use the geom's aesthetic properties to represent variables.
Each function returns a layer.

GRAPHICAL PRIMITIVES

```
a <- ggplot(economics, aes(date, unemploy))
b <- ggplot(seals, aes(x = long, y = lat))
```

- a + geom_blank()** and **a + expand_limits()**
Ensure limits include values across all plots.
- b + geom_curve(aes(yend = lat + 1, xend = long + 1, curvature = 1))** - x, yend, alpha, angle, curvature, linetype, size
- a + geom_path(lineend = "butt", linejoin = "round", linemitre = 1)** - x, y, alpha, color, group, linetype, size
- a + geom_polygon(aes(alpha = 50))** - x, y, alpha, color, fill, group, subgroup, linetype, size
- b + geom_rect(aes(xmin = long, ymin = lat, xmax = long + 1, ymax = lat + 1))** - xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size
- a + geom_ribbon(aes(ymin = unemploy - 900, ymax = unemploy + 900))** - x, ymax, ymin, alpha, color, fill, group, linetype, size

LINE SEGMENTS

common aesthetics: x, y, alpha, color, linetype, size

- b + geom_abline(aes(intercept = 0, slope = 1))**
- b + geom_hline(aes(yintercept = lat))**
- b + geom_vline(aes(xintercept = long))**
- b + geom_segment(aes(yend = lat + 1, xend = long + 1))**
- b + geom_spoke(aes(angle = 1:1155, radius = 1))**

ONE VARIABLE continuous

- ```
c <- ggplot(mpg, aes(hwy)); c2 <- ggplot(mpg)
```
- c + geom\_area(stat = "bin")** - x, y, alpha, color, fill, linetype, size
  - c + geom\_density(kernel = "gaussian")** - x, y, alpha, color, fill, group, linetype, size, weight
  - c + geom\_dotplot()** - x, y, alpha, color, fill
  - c + geom\_freqpoly()** - x, y, alpha, color, group, linetype, size
  - c + geom\_histogram(binwidth = 5)** - x, y, alpha, color, fill, linetype, size, weight
  - c2 + geom\_qq(aes(sample = hwy))** - x, y, alpha, color, fill, linetype, size, weight

### discrete

```
d <- ggplot(mpg, aes(f1))
```

- d + geom\_bar()** - x, alpha, color, fill, linetype, size, weight

### TWO VARIABLES both continuous

```
e <- ggplot(mpg, aes(cty, hwy))
```

- e + geom\_label(aes(label = cty), nudge\_x = 1, nudge\_y = 1)** - x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust
- e + geom\_point()** - x, y, alpha, color, fill, shape, size, stroke
- e + geom\_quantile()** - x, y, alpha, color, group, linetype, size, weight
- e + geom\_rug(sides = "bl")** - x, y, alpha, color, linetype, size
- e + geom\_smooth(method = lm)** - x, y, alpha, color, fill, group, linetype, size, weight
- e + geom\_text(aes(label = cty), nudge\_x = 1, nudge\_y = 1)** - x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

### one discrete, one continuous

```
f <- ggplot(mpg, aes(class, hwy))
```

- f + geom\_col()** - x, y, alpha, color, fill, group, linetype, size
- f + geom\_boxplot()** - x, y, lower, middle, upper, ymax, ymin, alpha, color, fill, group, linetype, shape, size, weight
- f + geom\_dotplot(binaxis = "y", stackdir = "center")** - x, y, alpha, color, fill, group
- f + geom\_violin(scale = "area")** - x, y, alpha, color, fill, group, linetype, size, weight

### both discrete

```
g <- ggplot(diamonds, aes(cut, color))
```

- g + geom\_count()** - x, y, alpha, color, fill, shape, size, stroke
- e + geom\_jitter(height = 2, width = 2)** - x, y, alpha, color, fill, shape, size

### THREE VARIABLES

```
seals$z <- with(seals, sqrt(delta_long^2 + delta_lat^2)); l <- ggplot(seals, aes(long, lat))
```

- l + geom\_contour(aes(z = z))** - x, y, z, alpha, color, group, linetype, size, weight
- l + geom\_contour\_filled(aes(fill = z))** - x, y, alpha, color, fill, group, linetype, size, subgroup
- l + geom\_raster(aes(fill = z), hjust = 0.5, vjust = 0.5, interpolate = FALSE)** - x, y, alpha, fill
- l + geom\_tile(aes(fill = z))** - x, y, alpha, color, fill, linetype, size, width

### continuous bivariate distribution

```
h <- ggplot(diamonds, aes(carat, price))
```

- h + geom\_bin2d(binwidth = c(0.25, 500))** - x, y, alpha, color, fill, linetype, size, weight
- h + geom\_density\_2d()** - x, y, alpha, color, group, linetype, size
- h + geom\_hex()** - x, y, alpha, color, fill, size

### continuous function

```
i <- ggplot(economics, aes(date, unemploy))
```

- i + geom\_area()** - x, y, alpha, color, fill, linetype, size
- i + geom\_line()** - x, y, alpha, color, group, linetype, size
- i + geom\_step(direction = "hv")** - x, y, alpha, color, group, linetype, size

### visualizing error

```
df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)
j <- ggplot(df, aes(grp, fit, ymin = fit - se, ymax = fit + se))
```

- j + geom\_crossbar(fatten = 2)** - x, y, ymax, ymin, alpha, color, fill, group, linetype, size
- j + geom\_errorbar()** - x, ymax, ymin, alpha, color, group, linetype, size, width  
Also **geom\_errorbarh()**.
- j + geom\_linerange()** - x, ymin, ymax, alpha, color, group, linetype, size
- j + geom\_pointrange()** - x, y, ymin, ymax, alpha, color, fill, group, linetype, shape, size

### maps

```
data <- data.frame(murder = USArrests$Murder, state = tolower(rownames(USArrests)))
```

```
map <- map_data("state")
```

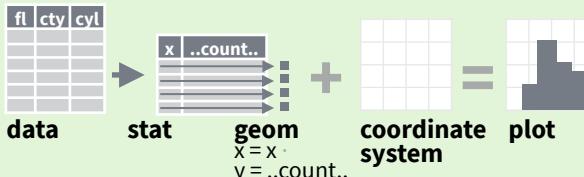
```
k <- ggplot(data, aes(fill = murder))
```

- k + geom\_map(aes(map\_id = state), map = map) + expand\_limits(x = map\$long, y = map\$lat)**  
map\_id, alpha, color, fill, linetype, size

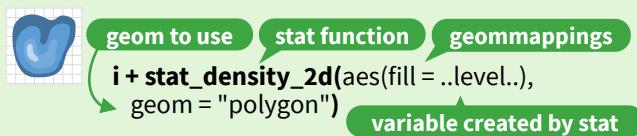
# Stats

An alternative way to build a layer.

A stat builds new variables to plot (e.g., count, prop).



Visualize a stat by changing the default stat of a geom function, `geom_bar(stat="count")` or by using a stat function, `stat_count(geom="bar")`, which calls a default geom to make a layer (equivalent to a geom function). Use `..name..` syntax to map stat variables to aesthetics.



```

c + stat_bin(binwidth = 1, boundary = 10)
x, y | ..count.., ..ncount.., ..density.., ..ndensity..
c + stat_count(width = 1) x, y | ..count.., ..prop..
c + stat_density(adjust = 1, kernel = "gaussian")
x, y | ..count.., ..density.., ..scaled..

```

```

e + stat_bin_2d(bins = 30, drop = T)
x, y, fill | ..count.., ..density..
e + stat_bin_hex(bins = 30) x, y, fill | ..count.., ..density..
e + stat_density_2d(contour = TRUE, n = 100)
x, y, color, size | ..level..
e + stat_ellipse(level = 0.95, segments = 51, type = "t")

```

```

l + stat_contour(aes(z = z)) x, y, z, order | ..level..
l + stat_summary_hex(aes(z = z), bins = 30, fun = max)
x, y, z, fill | ..value..
l + stat_summary_2d(aes(z = z), bins = 30, fun = mean)
x, y, z, fill | ..value..

```

```

f + stat_boxplot(coef = 1.5)
x, y | ..lower.., ..middle.., ..upper.., ..width.., ..ymin.., ..ymax..
f + stat_ydensity(kernel = "gaussian", scale = "area") x, y
| ..density.., ..scaled.., ..count.., ..n.., ..violinwidth.., ..width..

```

```

e + stat_ecdf(n = 40) x, y | ..x.., ..y..
e + stat_quantile(quantiles = c(0.1, 0.9),
formula = y ~ log(x), method = "rq") x, y | ..quantile..
e + stat_smooth(method = "lm", formula = y ~ x, se = T,
level = 0.95) x, y | ..se.., ..x.., ..y.., ..ymin.., ..ymax..

```

```

ggplot() + xlim(-5, 5) + stat_function(fun = dnorm,
n = 20, geom = "point") x | ..x.., ..y..
ggplot() + stat_qq(aes(sample = 1:100))
x, y, sample | ..sample.., ..theoretical..
e + stat_sum() x, y, size | ..n.., ..prop..

```

```

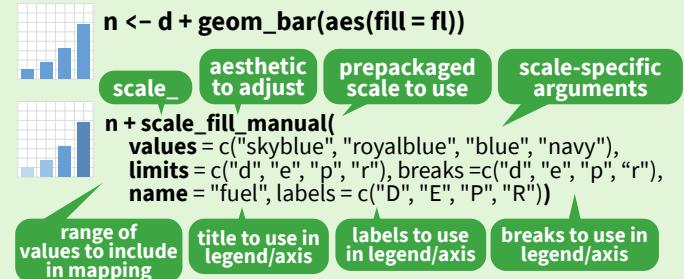
e + stat_summary(fun.data = "mean_cl_boot")
h + stat_summary_bin(fun = "mean", geom = "bar")
e + stat_identity()
e + stat_unique()

```

# Scales

Override defaults with `scales` package.

**Scales** map data values to the visual values of an aesthetic. To change a mapping, add a new scale.



## GENERAL PURPOSE SCALES

Use with most aesthetics

`scale_*_continuous()` - Map cont' values to visual ones.  
`scale_*_discrete()` - Map discrete values to visual ones.  
`scale_*_binned()` - Map continuous values to discrete bins.  
`scale_*_identity()` - Use data values as visual ones.  
`scale_*_manual(values = c())` - Map discrete values to manually chosen visual ones.  
`scale_*_date(date_labels = "%m/%d")`,  
`date_breaks = "2 weeks"` - Treat data values as dates.  
`scale_*_datetime()` - Treat data values as date times.  
Same as `scale_*_date()`. See `?strptime` for label formats.

## X & Y LOCATION SCALES

Use with x or y aesthetics (x shown here)

`scale_x_log10()` - Plot x on log10 scale.  
`scale_x_reverse()` - Reverse the direction of the x axis.  
`scale_x_sqrt()` - Plot x on square root scale.

## COLOR AND FILL SCALES (DISCRETE)

`n + scale_fill_brewer(palette = "Blues")`  
For palette choices:  
RColorBrewer::display.brewer.all()  
`n + scale_fill_grey(start = 0.2,`  
`end = 0.8, na.value = "red")`

## COLOR AND FILL SCALES (CONTINUOUS)

`o <- c + geom_dotplot(aes(fill = ..x..))`  
`o + scale_fill_distiller(palette = "Blues")`  
`o + scale_fill_gradient(low = "red", high = "yellow")`  
`o + scale_fill_gradient2(low = "red", high = "blue",`  
`mid = "white", midpoint = 25)`  
`o + scale_fill_gradientn(colors = topo.colors(6))`  
Also: rainbow(), heat.colors(), terrain.colors(),  
cm.colors(), RColorBrewer::brewer.pal()

## SHAPE AND SIZE SCALES

`p <- e + geom_point(aes(shape = fl, size = cyl))`  
`p + scale_shape() + scale_size()`  
`p + scale_shape_manual(values = c(3:7))`  
`0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25`  
`□○△×+×◊▽★◆⊕⊖⊕⊖□○△○○□○△○○`  
`p + scale_radius(range = c(1,6))`  
`p + scale_size_area(max_size = 6)`

# Coordinate Systems

`r <- d + geom_bar()`

`r + coord_cartesian(xlim = c(0, 5))` - xlim, ylim  
The default cartesian coordinate system.

`r + coord_fixed(ratio = 1/2)`  
ratio, xlim, ylim - Cartesian coordinates with fixed aspect ratio between x and y units.

`ggplot(mpg, aes(y = fl)) + geom_bar()`  
Flip cartesian coordinates by switching x and y aesthetic mappings.

`r + coord_polar(theta = "x", direction=1)`  
theta, start, direction - Polar coordinates.

`r + coord_trans(y = "sqrt")` - x, y, xlim, ylim  
Transformed cartesian coordinates. Set xtrans and ytrans to the name of a window function.

`π + coord_quickmap()`  
`π + coord_map(projection = "ortho", orientation = c(41, -74, 0))` - projection, xlim, ylim  
Map projections from the mapproj package (mercator (default), azequalarea, lagrange, etc.).

# Position Adjustments

Position adjustments determine how to arrange geoms that would otherwise occupy the same space.

`s <- ggplot(mpg, aes(fl, fill = drv))`

`s + geom_bar(position = "dodge")`  
Arrange elements side by side.

`s + geom_bar(position = "fill")`  
Stack elements on top of one another, normalize height.

`e + geom_point(position = "jitter")`  
Add random noise to X and Y position of each element to avoid overplotting.

`e + geom_label(position = "nudge")`  
Nudge labels away from points.

`s + geom_bar(position = "stack")`  
Stack elements on top of one another.

Each position adjustment can be recast as a function with manual `width` and `height` arguments:  
`s + geom_bar(position = position_dodge(width = 1))`

# Themes

`r + theme_bw()`  
White background with grid lines.

`r + theme_gray()`  
Grey background (default theme).

`r + theme_dark()`  
Dark for contrast.

`r + theme_classic()`  
`r + theme_light()`

`r + theme_linedraw()`  
`r + theme_minimal()`

`r + theme_void()`  
Empty theme.

`r + theme()` Customize aspects of the theme such as axis, legend, panel, and facet properties.  
`r + ggtitle("Title") + theme(plot.title.position = "plot")`  
`r + theme(panel.background = element_rect(fill = "blue"))`

# Faceting

Facets divide a plot into subplots based on the values of one or more discrete variables.

`t <- ggplot(mpg, aes(cty, hwy)) + geom_point()`

`t + facet_grid(cols = vars(fl))`  
Facet into columns based on fl.

`t + facet_grid(rows = vars(year))`  
Facet into rows based on year.

`t + facet_grid(rows = vars(year), cols = vars(fl))`  
Facet into both rows and columns.

`t + facet_wrap(vars(fl))`  
Wrap facets into a rectangular layout.

Set `scales` to let axis limits vary across facets.

`t + facet_grid(rows = vars(drv), cols = vars(fl), scales = "free")`

x and y axis limits adjust to individual facets:  
`"free_x"` - x axis limits adjust  
`"free_y"` - y axis limits adjust

Set `labeler` to adjust facet label:

`t + facet_grid(cols = vars(fl), labeler = label_both)`

`fl: c fl: d fl: e fl: p fl: r`

`t + facet_grid(rows = vars(fl), labeler = label_bquote(alpha ^ .(fl)))`

`αc αd αe αp αr`

# Labels and Legends

Use `labs()` to label the elements of your plot.

`t + labs(x = "New x axis label", y = "New y axis label", title = "Add a title above the plot", subtitle = "Add a subtitle below title", caption = "Add a caption below plot", alt = "Add alt text to the plot", <AES> = "New <AES> legend title")`

`t + annotate(geom = "text", x = 8, y = 9, label = "A")`  
Places a geom with manually selected aesthetics.

`p + guides(x = guide_axis(n.dodge = 2))` Avoid crowded or overlapping labels with `guide_axis(n.dodge` or `angle`).

`n + guides(fill = "none")` Set legend type for each aesthetic: colorbar, legend, or none (no legend).

`n + theme(legend.position = "bottom")`  
Place legend at "bottom", "top", "left", or "right".

`n + scale_fill_discrete(name = "Title", labels = c("A", "B", "C", "D", "E"))`  
Set legend title and labels with a scale function.

# Zooming

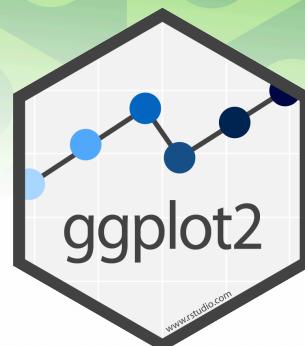
Without clipping (preferred):

`t + coord_cartesian(xlim = c(0, 100), ylim = c(10, 20))`

With clipping (removes unseen data points):

`t + xlim(0, 100) + ylim(10, 20)`

`t + scale_x_continuous(limits = c(0, 100)) + scale_y_continuous(limits = c(0, 100))`



# rmarkdown :: CHEAT SHEET

## What is rmarkdown?



**.Rmd files** • Develop your code and ideas side-by-side in a single document. Run code as individual chunks or as an entire document.

**Dynamic Documents** • Knit together plots, tables, and results with narrative text. Render to a variety of formats like HTML, PDF, MS Word, or MS Powerpoint.

**Reproducible Research** • Upload, link to, or attach your report to share. Anyone can read or run your code to reproduce your work.

## Workflow

- 1 Open a **new .Rmd file** in the RStudio IDE by going to *File > New File > R Markdown*.
- 2 **Embed code** in chunks. Run code by line, by chunk, or all at once.
- 3 **Write text** and add tables, figures, images, and citations. Format with Markdown syntax or the RStudio Visual Markdown Editor.
- 4 **Set output format(s) and options** in the YAML header. Customize themes or add parameters to execute or add interactivity with Shiny.
- 5 **Save and render** the whole document. Knit periodically to preview your work as you write.
- 6 **Share your work!**

## Embed Code with knitr

### CODE CHUNKS

Surround code chunks with `{{r}}` and `{{` or use the Insert Code Chunk button. Add a chunk label and/or chunk options inside the curly braces after {{r}}.

```
```{r chunk-label, include=FALSE}
summary(mtcars)
```
```

### SET GLOBAL OPTIONS

Set options for the entire document in the first chunk.

```
```{r include=FALSE}
knitr::opts_chunk$message = FALSE
```
```

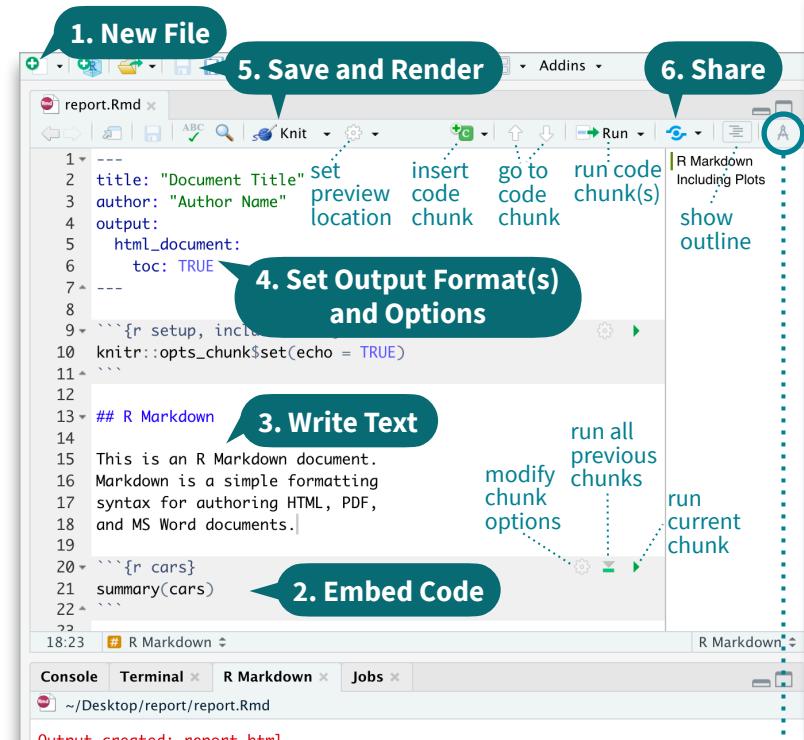
### INLINE CODE

Insert `{{r <code>}}` into text sections. Code is evaluated at render and results appear as text.

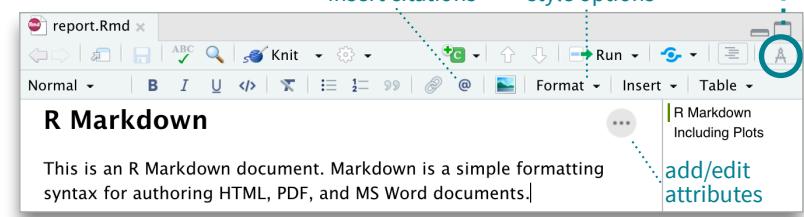
"Built with `{{r getRversion()}}`" --> "Built with 4.1.0"



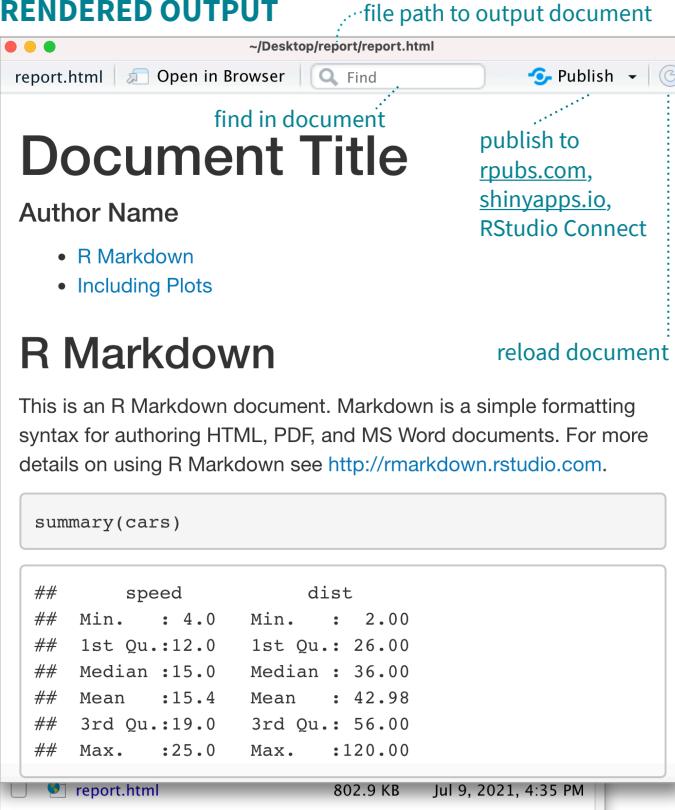
### SOURCE EDITOR



### VISUAL EDITOR



### RENDERED OUTPUT



## Write with Markdown

The syntax on the left renders as the output on the right.

Plain text.

Plain text.

End a line with two spaces to start a new paragraph.

End a line with two spaces to start a new paragraph.

Also end with a backslash\ to make a new line.

Also end with a backslash\ to make a new line.

**italics\*** and **\*\*bold\*\***

**italics** and **bold**

superscript<sup>2</sup>/subscript<sub>2</sub>

superscript<sup>2</sup>/subscript<sub>2</sub>

~~strikethrough~~

strikethrough

escaped: `\*` \`

escaped: \* \`

endash: --, emdash: ---

endash: -, emdash: --

### Header 1 Header 2

...

Header 6

- unordered list

• item 2

- item 2a (indent 1 tab)

• item 2b

1. ordered list

2. item 2

- item 2a (indent 1 tab)

• item 2b

<link url>

[This is a link.](link url)

[This is another link][id].

This is another link.

<http://www.rstudio.com/>

This is a link.

This is another link.



Caption.

verbatim code

multiple lines of verbatim code

> block quotes

block quotes

equation:  $e^{i\pi} + 1 = 0$

equation block:

$$E = mc^2$$

horizontal rule:

| Right | Left | Default | Center |
|-------|------|---------|--------|
| 12    | 12   | 12      | 12     |
| 123   | 123  | 123     | 123    |
| 1     | 1    | 1       | 1      |

### HTML Tabsets

```
Results {.tabset}
Plots text
text
```

## Tables more text

### Results

|       |        |
|-------|--------|
| Plots | Tables |
|-------|--------|

text





# Set Output Formats and their Options in YAML

Use the document's YAML header to set an **output format** and customize it with **output options**.

```

```

```
title: "My Document"
author: "Author Name"
output:
 html_document:
 toc: TRUE

```

**Indent format 2 characters,  
indent options 4 characters**

| OUTPUT FORMAT           | CREATES                      |
|-------------------------|------------------------------|
| html_document           | .html                        |
| pdf_document*           | .pdf                         |
| word_document           | Microsoft Word (.docx)       |
| powerpoint_presentation | Microsoft Powerpoint (.pptx) |
| odt_document            | OpenDocument Text            |
| rtf_document            | Rich Text Format             |
| md_document             | Markdown                     |
| github_document         | Markdown for Github          |
| ioslides_presentation   | ioslides HTML slides         |
| slidy_presentation      | Slidy HTML slides            |
| beamer_presentation*    | Beamer slides                |

\* Requires LaTeX, use `tinytex::install_tinytex()`  
Also see `flexdashboard`, `bookdown`, `distill`, and `blogdown`.

| IMPORTANT OPTIONS   | DESCRIPTION                                                                            | HTML    | PDF | MS Word | MS PPT |
|---------------------|----------------------------------------------------------------------------------------|---------|-----|---------|--------|
| anchor_sections     | Show section anchors on mouse hover (TRUE or FALSE)                                    | X       |     |         |        |
| citation_package    | The LaTeX package to process citations ("default", "natbib", "biblatex")               | X       |     |         |        |
| code_download       | Give readers an option to download the .Rmd source code (TRUE or FALSE)                | X       |     |         |        |
| code_folding        | Let readers to toggle the display of R code ("none", "hide", or "show")                | X       |     |         |        |
| css                 | CSS or SCSS file to use to style document (e.g. "style.css")                           | X       |     |         |        |
| dev                 | Graphics device to use for figure output (e.g. "png", "pdf")                           | X X     |     |         |        |
| df_print            | Method for printing data frames ("default", "kable", "tibble", "paged")                | X X X X |     |         |        |
| fig_caption         | Should figures be rendered with captions (TRUE or FALSE)                               | X X X X |     |         |        |
| highlight           | Syntax highlighting ("tango", "pygments", "kate", "zenburn", "textmate")               | X X X   |     |         |        |
| includes            | File of content to place in doc ("in_header", "before_body", "after_body")             | X X     |     |         |        |
| keep_md             | Keep the Markdown .md file generated by knitting (TRUE or FALSE)                       | X X X X |     |         |        |
| keep_tex            | Keep the intermediate TEX file used to convert to PDF (TRUE or FALSE)                  | X       |     |         |        |
| latex_engine        | LaTeX engine for producing PDF output ("pdflatex", "xelatex", or "lualatex")           | X       |     |         |        |
| reference_docx/_doc | docx/pptx file containing styles to copy in the output (e.g. "file.docx", "file.pptx") | X X     |     |         |        |
| theme               | Theme options (see Bootswatch and Custom Themes below)                                 | X       |     |         |        |
| toc                 | Add a table of contents at start of document (TRUE or FALSE)                           | X X X X |     |         |        |
| toc_depth           | The lowest level of headings to add to table of contents (e.g. 2, 3)                   | X X X X |     |         |        |
| toc_float           | Float the table of contents to the left of the main document content (TRUE or FALSE)   | X       |     |         |        |

Use `?<output format>` to see all of a format's options, e.g. `?html_document`

## More Header Options

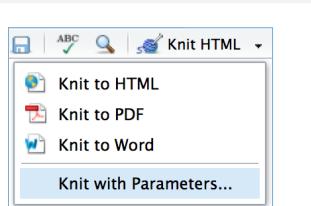
### PARAMETERS

Parameterize your documents to reuse with new inputs (e.g., data, values, etc.).

1. **Add parameters** in the header as sub-values of `params`.
2. **Call parameters** in code using `params$<name>`.
3. **Set parameters** with Knit with Parameters or the `params` argument of `render()`.

### REUSABLE TEMPLATES

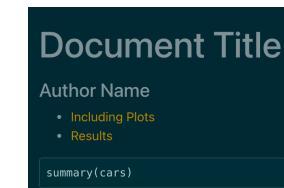
1. **Create a new package** with a `inst/rmarkdown/templates` directory.
2. **Add a folder** containing `template.yaml` (below) and `skeleton.Rmd` (template contents).
3. **Install** the package to access template by going to **File > New R Markdown > From Template**.



### BOOTSWATCH THEMES

Customize HTML documents with Bootswatch themes from the `bslib` package using the theme output option.

Use `bslib::bootswatch_themes()` to list available themes.



```

```

```
title: "Document Title"
author: "Author Name"
output:
 html_document:
 theme:
 bootswatch: solar

```

### CUSTOM THEMES

Customize individual HTML elements using `bslib` variables. Use `?bs_theme` to see more variables.

```

```

```
output:
 html_document:
 theme:
 bg: "#121212"
 fg: "#E4E4E4"
 base_font:
 google: "Prompt"

```

More on `bslib` at [pkgs.rstudio.com/bslib/](http://pkgs.rstudio.com/bslib/).

### STYLING WITH CSS AND SCSS

Add CSS and SCSS to your document by adding a path to a file with the `css` option in the YAML header.

```

```

```
title: "My Document"
author: "Author Name"
output:
 html_document:
 css: "style.css"

```

Apply CSS styling by writing HTML tags directly or:

- Use markdown to apply style attributes inline.

Bracketed Span  
A [green]{.my-color} word.

A green word.

Fenced Div  
:::{.my-color}  
All of these words  
are green.  
:::

All of these words  
are green.

- Use the Visual Editor. Go to **Format > Div/Span** and add CSS styling directly with Edit Attributes.

.my-css-tag ...  
This is a div with some text in it.

## Render

When you render a document, rmarkdown:

1. Runs the code and embeds results and text into an .md file with knitr.
2. Converts the .md file into the output format with Pandoc.



**Save**, then **Knit** to preview the document output. The resulting HTML/PDF/MS Word/etc. document will be created and saved in the same directory as the .Rmd file.

Use `rmarkdown::render()` to render/knit in the R console. See `?render` for available options.

## Share

### Publish on RStudio Connect

to share R Markdown documents securely, schedule automatic updates, and interact with parameters in real time.

[rstudio.com/products/connect/](https://rstudio.com/products/connect/)



### INTERACTIVITY

Turn your report into an interactive Shiny document in 4 steps:

1. Add `runtime: shiny` to the YAML header.
2. Call Shiny input functions to embed input objects.
3. Call Shiny render functions to embed reactive output.
4. Render with `rmarkdown::run()` or click **Run Document** in RStudio IDE.

```

```

```
output: html_document
runtime: shiny

```

```
```{r, echo = FALSE}
numericInput("n",
  "How many cars?", 5)
renderTable({
  head(cars, input$n)
})
```

speed	dist
1	4.00
2	4.00
3	7.00
4	7.00
5	8.00
	2.00
	10.00
	4.00
	22.00
	16.00

Also see Shiny Prerendered for better performance.
rmarkdown.rstudio.com/authoring_shiny_prerendered

Embed a complete app into your document with `shiny::shinyAppDir()`. More at bookdown.org/yihui/rmarkdown/shiny-embedded.html.

String manipulation with stringr :: CHEAT SHEET



The **stringr** package provides a set of internally consistent tools for working with character strings, i.e. sequences of characters surrounded by quotation marks.

Detect Matches

	str_detect(string, pattern, negate = FALSE) Detect the presence of a pattern match in a string. Also str_like() . str_detect(fruit, "a")
	str_starts(string, pattern, negate = FALSE) Detect the presence of a pattern match at the beginning of a string. Also str_ends() . str_starts(fruit, "a")
	str_which(string, pattern, negate = FALSE) Find the indexes of strings that contain a pattern match. str_which(fruit, "a")
	str_locate(string, pattern) Locate the positions of pattern matches in a string. Also str_locate_all() . str_locate(fruit, "a")
	str_count(string, pattern) Count the number of matches in a string. str_count(fruit, "a")

Subset Strings

	str_sub(string, start = 1L, end = -1L) Extract substrings from a character vector. str_sub(fruit, 1, 3); str_sub(fruit, -2)
	str_subset(string, pattern, negate = FALSE) Return only the strings that contain a pattern match. str_subset(fruit, "p")
	str_extract(string, pattern) Return the first pattern match found in each string, as a vector. Also str_extract_all() to return every pattern match. str_extract(fruit, "[aeiou]")
	str_match(string, pattern) Return the first pattern match found in each string, as a matrix with a column for each () group in pattern. Also str_match_all() . str_match(sentences, "(a the) ([^ +])")

Manage Lengths

	str_length(string) The width of strings (i.e. number of code points, which generally equals the number of characters). str_length(fruit)
	str_pad(string, width, side = c("left", "right", "both"), pad = " ") Pad strings to constant width. str_pad(fruit, 17)
	str_trunc(string, width, side = c("right", "left", "center"), ellipsis = "...") Truncate the width of strings, replacing content with ellipsis. str_trunc(sentences, 6)
	str_trim(string, side = c("both", "left", "right")) Trim whitespace from the start and/or end of a string. str_trim(str_pad(fruit, 17))
	str_squish(string) Trim whitespace from each end and collapse multiple spaces into single spaces. str_squish(str_pad(fruit, 17, "both"))

Mutate Strings

	str_sub() <- value. Replace substrings by identifying the substrings with str_sub() and assigning into the results. str_sub(fruit, 1, 3) <- "str"
	str_replace(string, pattern, replacement) Replace the first matched pattern in each string. Also str_remove() . str_replace(fruit, "p", "-")
	str_replace_all(string, pattern, replacement) Replace all matched patterns in each string. Also str_remove_all() . str_replace_all(fruit, "p", "-")
	str_to_lower(string, locale = "en")¹ Convert strings to lower case. str_to_lower(sentences)
	str_to_upper(string, locale = "en")¹ Convert strings to upper case. str_to_upper(sentences)
	str_to_title(string, locale = "en")¹ Convert strings to title case. Also str_to_sentence() . str_to_title(sentences)

Join and Split

	str_c(..., sep = "", collapse = NULL) Join multiple strings into a single string. str_c(letters, LETTERS)
	str_flatten(string, collapse = "") Combines into a single string, separated by collapse. str_flatten(fruit, ",")
	str_dup(string, times) Repeat strings times times. Also str_unique() to remove duplicates. str_dup(fruit, times = 2)
	str_split_fixed(string, pattern, n) Split a vector of strings into a matrix of substrings (splitting at occurrences of a pattern match). Also str_split() to return a list of substrings and str_split_n() to return the nth substring. str_split_fixed(sentences, " ", n=3)
	str_glue(..., .sep = "", .envir = parent.frame()) Create a string from strings and {expressions} to evaluate. str_glue("Pi is {pi}")
	str_glue_data(.x, ..., .sep = "", .envir = parent.frame(), .na = "NA") Use a data frame, list, or environment to create a string from strings and {expressions} to evaluate. str_glue_data(mtcars, "[rownames(mtcars)] has {hp} hp")

Order Strings

	str_order(x, decreasing = FALSE, na_last = TRUE, locale = "en", numeric = FALSE, ...)¹ Return the vector of indexes that sorts a character vector. fruit[str_order(fruit)]
	str_sort(x, decreasing = FALSE, na_last = TRUE, locale = "en", numeric = FALSE, ...)¹ Sort a character vector. str_sort(fruit)

Helpers

	str_conv(string, encoding) Override the encoding of a string. str_conv(fruit, "ISO-8859-1")
	str_view_all(string, pattern, match = NA) View HTML rendering of all regex matches. Also str_view() to see only the first match. str_view_all(sentences, "[aeiou]")
	str_equal(x, y, locale = "en", ignore_case = FALSE, ...)¹ Determine if two strings are equivalent. str_equal(c("a", "b"), c("a", "c"))
	str_wrap(string, width = 80, indent = 0, exdent = 0) Wrap strings into nicely formatted paragraphs. str_wrap(sentences, 20)

¹ See bit.ly/ISO639-1 for a complete list of locales.

Need to Know

Pattern arguments in string are interpreted as regular expressions *after any special characters have been parsed*.

In R, you write regular expressions as *strings*, sequences of characters surrounded by quotes ("") or single quotes('').

Some characters cannot be represented directly in an R string. These must be represented as **special characters**, sequences of characters that have a specific meaning., e.g.

Special Character	Represents
\\\	\
'"	"
\n	new line

Run `?"""` to see a complete list

Because of this, whenever a \ appears in a regular expression, you must write it as \\ in the string that represents the regular expression.

Use `writeLines()` to see how R views your string after all special characters have been parsed.

```
writeLines("|\.")  
# \.
```

```
writeLines("\| is a backslash")  
# \| is a backslash
```

INTERPRETATION

Patterns in stringr are interpreted as regexs. To change this default, wrap the pattern in one of:

`regex(pattern, ignore_case = FALSE, multiline = FALSE, comments = FALSE, dotall = FALSE, ...)`
Modifies a regex to ignore cases, match end of lines as well of end of strings, allow R comments within regex's , and/or to have . match everything including \n.
`str_detect("i", regex("i", TRUE))`

`fixed()` Matches raw bytes but will miss some characters that can be represented in multiple ways (fast). `str_detect("\u0130", fixed("i"))`

`coll()` Matches raw bytes and will use locale specific collation rules to recognize characters that can be represented in multiple ways (slow). `str_detect("\u0130", coll("i", TRUE, locale = "tr"))`

`boundary()` Matches boundaries between characters, line_breaks, sentences, or words. `str_split(sentences, boundary("word"))`

Regular Expressions -

Regular expressions, or *regexp*s, are a concise language for describing patterns in strings.

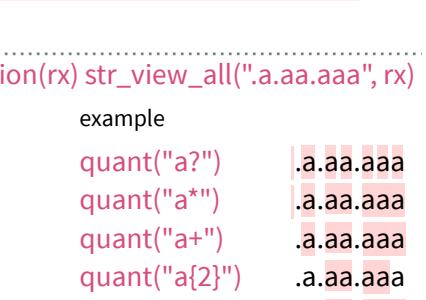
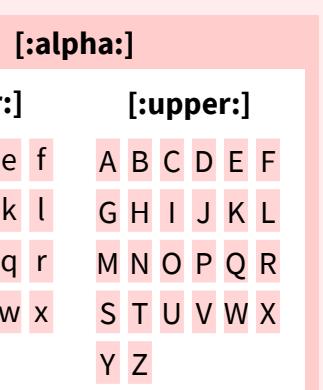
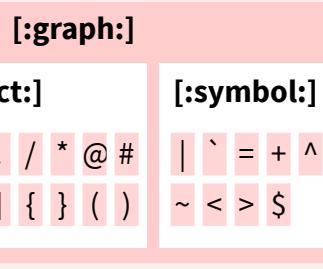
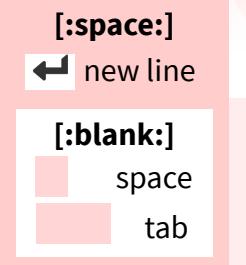
MATCH CHARACTERS

string (type this)	regexp (to mean this)	matches (which matches this)	example
	a (etc.)	a (etc.)	see("a")
\\.	\\.	.	see("\\.")
\\!	\\!	!	see("\\!")
\\?	\\?	?	see("\\?")
\\\\	\\\\	\	see("\\\\")
\\(\\((see("\\()")
\\)	\\))	see("\\)")
\\{	\\{	{	see("\\{")
\\}	\\}	}	see("\\}")
\\n	\\n	new line (return)	see("\\n")
\\t	\\t	tab	see("\\t")
\\s	\\s	any whitespace (S for non-whitespaces)	see("\\s")
\\d	\\d	any digit (D for non-digits)	see("\\d")
\\w	\\w	any word character (W for non-word chars)	see("\\w")
\\b	\\b	word boundaries	see("\\b")
	[:digit:] ¹	digits	see("[:digit:]")
	[:alpha:] ¹	letters	see("[:alpha:]")
	[:lower:] ¹	lowercase letters	see("[:lower:]")
	[:upper:] ¹	uppercase letters	see("[:upper:]")
	[:alnum:] ¹	letters and numbers	see("[:alnum:]")
	[:punct:] ¹	punctuation	see("[:punct:]")
	[:graph:] ¹	letters, numbers, and punctuation	see("[:graph:]")
	[:space:] ¹	space characters (i.e. \s)	see("[:space:]")
	[:blank:] ¹	space and tab (but not new line)	see("[:blank:]")
.	.	every character except a new line	see(".")

see <- function(rx) str_view_all("abc ABC 123\\t.!?\n\\n", rx)

abc ABC 123 .!?\n
abc ABC 123 .!?

see <- function(rx) str_view_all("abc ABC 123\\t.!?\n\\n", rx)



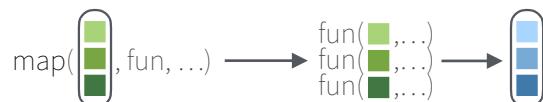
Apply functions with purrr :: CHEAT SHEET



Map Functions

ONE LIST

map(.x, .f, ...) Apply a function to each element of a list or vector, return a list.
`x <- list(1:10, 11:20, 21:30)
l1 <- list(x = c("a", "b"), y = c("c", "d"))
map(l1, sort, decreasing = TRUE)`



map_dbl(.x, .f, ...)
Return a double vector.
`map_dbl(x, mean)`

map_int(.x, .f, ...)
Return an integer vector.
`map_int(x, length)`

map_chr(.x, .f, ...)
Return a character vector.
`map_chr(l1, paste, collapse = "")`

map_lgl(.x, .f, ...)
Return a logical vector.
`map_lgl(x, is.integer)`

map_dfc(.x, .f, ...)
Return a data frame created by column-binding.
`map_dfc(l1, rep, 3)`

map_dfr(.x, .f, ..., .id = NULL)
Return a data frame created by row-binding.
`map_dfr(x, summary)`

walk(.x, .f, ...) Trigger side effects, return invisibly.
`walk(x, print)`

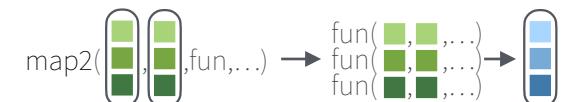
Function Shortcuts

Use `~.` with functions like **map()** that have single arguments.

`map(l, ~ . + 2)`
becomes
`map(l, function(x) x + 2)`

TWO LISTS

map2(.x, .y, .f, ...) Apply a function to pairs of elements from two lists or vectors, return a list.
`y <- list(1, 2, 3); z <- list(4, 5, 6); l2 <- list(x = "a", y = "z")
map2(x, y, ~ .x * .y)`



map2_dbl(.x, .y, .f, ...)
Return a double vector.
`map2_dbl(y, z, ~ .x / .y)`

map2_int(.x, .y, .f, ...)
Return an integer vector.
`map2_int(y, z, `+`)`

map2_chr(.x, .y, .f, ...)
Return a character vector.
`map2_chr(l1, l2, paste, collapse = "", sep = ":")`

map2_lgl(.x, .y, .f, ...)
Return a logical vector.
`map2_lgl(l2, l1, `%in%`)`

map2_dfc(.x, .y, .f, ...)
Return a data frame created by column-binding.
`map2_dfc(l1, l2, ~ as.data.frame(c(.x, .y)))`

map2_dfr(.x, .y, .f, ..., .id = NULL)
Return a data frame created by row-binding.
`map2_dfr(l1, l2, ~ as.data.frame(c(.x, .y)))`

walk2(.x, .y, .f, ...) Trigger side effects, return invisibly.
`walk2(objs, paths, save)`



MANY LISTS

pmap(.l, .f, ...) Apply a function to groups of elements from a list of lists or vectors, return a list.
`pmap(list(x, y, z), ~ ..1 * (.2 + ..3))`



pmap_dbl(.l, .f, ...)
Return a double vector.
`pmap_dbl(list(y, z), ~ .x / .y)`

pmap_int(.l, .f, ...)
Return an integer vector.
`pmap_int(list(y, z), `+`)`

pmap_chr(.l, .f, ...)
Return a character vector.
`pmap_chr(list(l1, l2), paste, collapse = "", sep = ":")`

pmap_lgl(.l, .f, ...)
Return a logical vector.
`pmap_lgl(list(l2, l1), `%in%`)`

pmap_dfc(.l, .f, ...) Return a data frame created by column-binding.
`pmap_dfc(list(l1, l2), ~ as.data.frame(c(.x, .y)))`

pmap_dfr(.l, .f, ..., .id = NULL) Return a data frame created by row-binding.
`pmap_dfr(list(l1, l2), ~ as.data.frame(c(.x, .y)))`

pwalk(.l, .f, ...) Trigger side effects, return invisibly.
`pwalk(list(objs, paths), save)`



LISTS AND INDEXES

imap(.x, .f, ...) Apply `.f` to each element and its index, return a list.
`imap(y, ~ paste0(y, ": ", .x))`



imap_dbl(.x, .f, ...)
Return a double vector.
`imap_dbl(y, ~ .y)`

imap_int(.x, .f, ...)
Return an integer vector.
`imap_int(y, ~ .y)`

imap_chr(.x, .f, ...)
Return a character vector.
`imap_chr(y, ~ paste0(y, ": ", .x))`

imap_lgl(.x, .f, ...)
Return a logical vector.
`imap_lgl(l1, ~ is.character(y))`

imap_dfc(.x, .f, ...)
Return a data frame created by column-binding.
`imap_dfc(l2, ~ as.data.frame(c(x, y)))`

imap_dfr(.x, .f, ..., .id = NULL)
Return a data frame created by row-binding.
`imap_dfr(l2, ~ as.data.frame(c(x, y)))`

iwalk(.x, .f, ...) Trigger side effects, return invisibly.
`iwalk(z, ~ print(paste0(y, ": ", .x)))`

Use `~ .x .y` with functions like **imap()**. `.x` will get the list value and `.y` will get the index, or name if available.

`imap(list(a, b, c), ~ paste0(.y, ": ", .x))`
outputs "index: value" for each item

Use `~ .x .y` with functions like **map2()** that have two arguments.

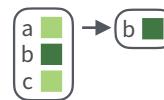
`map2(l, p, ~ .x + .y)`
becomes
`map2(l, p, function(l, p) l + p)`

Use a **string** or an **integer** with any map function to index list elements by name or position. `map(l, "name")` becomes `map(l, function(x) x[["name"]])`

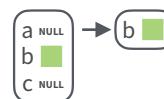


Work with Lists

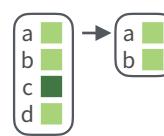
Filter



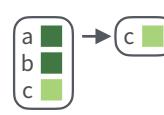
keep(.x, .p, ...)
Select elements that pass a logical test.
Conversely, **discard()**.
`keep(x, is.na)`



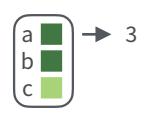
compact(.x, .p = identity)
Drop empty elements.
`compact(x)`



head_while(.x, .p, ...)
Return head elements until one does not pass.
Also **tail_while()**.
`head_while(x, is.character)`



detect(.x, .f, ..., dir = c("forward", "backward"), .right = NULL, .default = NULL)
Find first element to pass.
`detect(x, is.character)`



detect_index(.x, .f, ..., dir = c("forward", "backward"), .right = NULL) Find index of first element to pass.
`detect_index(x, is.character)`



every(.x, .p, ...)
Do all elements pass a test?
`every(x, is.character)`



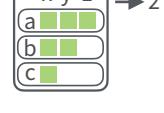
some(.x, .p, ...)
Do some elements pass a test?
`some(x, is.character)`



none(.x, .p, ...)
Do no elements pass a test?
`none(x, is.character)`

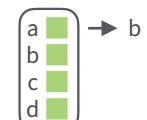


has_element(.x, .y)
Does a list contain an element?
`has_element(x, "foo")`

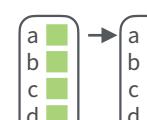


vec_depth(x)
Return depth (number of levels of indexes).
`vec_depth(x)`

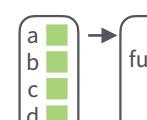
Index



pluck(.x, ..., .default=NULL)
Select an element by name or index. Also **attr_getter()** and **chuck()**.
`pluck(x, "b")`
`x %>% pluck("b")`

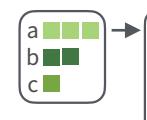


assign_in(x, where, value)
Assign a value to a location using pluck selection.
`assign_in(x, "b", 5)`
`x %>% assign_in("b", 5)`



modify_in(.x, .where, .f)
Apply a function to a value at a selected location.
`modify_in(x, "b", abs)`
`x %>% modify_in("b", abs)`

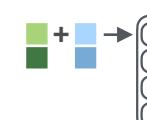
Reshape



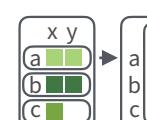
flatten(.x) Remove a level of indexes from a list.
Also **flatten_chr()** etc.
`flatten(x)`



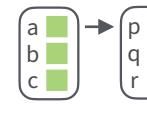
array_tree(array, margin = NULL) Turn array into list.
Also **array_branch()**.
`array_tree(x, margin = 3)`



cross2(.x, .y, .filter = NULL)
All combinations of .x and .y.
Also **cross()**, **cross3()**, and **cross_df()**.
`cross2(1:3, 4:6)`

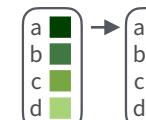


transpose(.l, .names = NULL)
Transposes the index order in a multi-level list.
`transpose(x)`

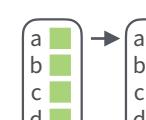


set_names(x, nm = x)
Set the names of a vector/list directly or with a function.
`set_names(x, c("p", "q", "r"))`
`set_names(x, tolower)`

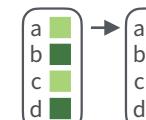
Modify



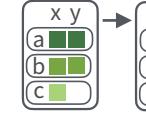
modify(.x, .f, ...) Apply a function to each element. Also **modify2()**, and **imodify()**.
`modify(x, ~.+ 2)`



modify_at(.x, .at, .f, ...) Apply a function to selected elements.
Also **map_at()**.
`modify_at(x, "b", ~.+ 2)`

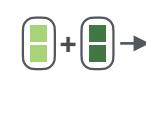


modify_if(.x, .p, .f, ...) Apply a function to elements that pass a test.
Also **map_if()**.
`modify_if(x, is.numeric, ~.+2)`

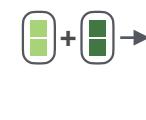


modify_depth(.x, .depth, .f, ...) Apply function to each element at a given level of a list. Also **map_depth()**.
`modify_depth(x, 2, ~.+ 2)`

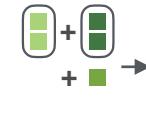
Combine



append(x, values, after = length(x)) Add values to end of list.
`append(x, list(d = 1))`



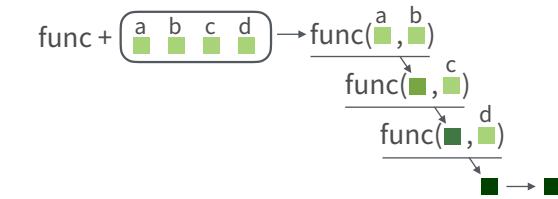
prepend(x, values, before = 1) Add values to start of list.
`prepend(x, list(d = 1))`



splice(...) Combine objects into a list, storing S3 objects as sub-lists.
`splice(x, y, "foo")`

Reduce

reduce(.x, .f, ..., .init, .dir = c("forward", "backward")) Apply function recursively to each element of a list or vector. Also **reduce2()**.
`reduce(x, sum)`



List-Columns

List-columns are columns of a data frame where each element is a list or vector instead of an atomic value. Columns can also be lists of data frames. See **tidyverse** for more about nested data and list columns.

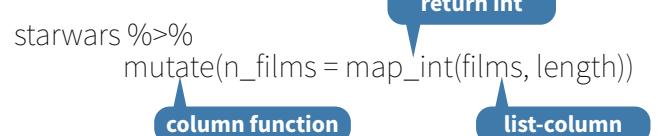
WORK WITH LIST-COLUMNS

Manipulate list-columns like any other kind of column, using **dplyr** functions like **mutate()** and **transmute()**. Because each element is a list, use **map functions** within a column function to manipulate each element.

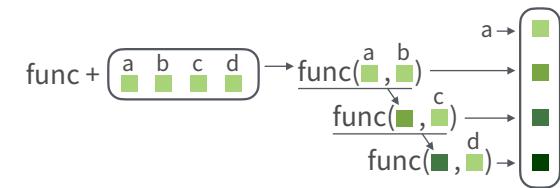
map(), **map2()**, or **pmap()** return lists and will **create new list-columns**.



Suffixed map functions like **map_int()** return an atomic data type and will **simplify list-columns into regular columns**.



accumulate(.x, .f, ..., .init) Reduce a list, but also return intermediate results. Also **accumulate2()**.
`accumulate(x, sum)`





Factors withforcats :: CHEAT SHEET

The **forcats** package provides tools for working with factors, which are R's data structure for categorical data.

Factors

R represents categorical data with factors. A **factor** is an integer vector with a **levels** attribute that stores a set of mappings between integers and categorical values. When you view a factor, R displays not the integers, but the levels associated with them.

<code>a c b a</code>	<code>a c b a</code>	<i>Create a factor with factor()</i>
	<code>1 = a 2 = b 3 = c</code>	<code>factor(x = character(), levels, labels = levels, exclude = NA, ordered = is.ordered(x), nmax = NA)</code> Convert a vector to a factor. Also <code>as_factor()</code> . <code>f <- factor(c("a", "c", "b", "a"), levels = c("a", "b", "c"))</code>
<code>a c b a</code>	<code>a b c</code>	<i>Return its levels with levels()</i> <code>levels(x)</code> Return/set the levels of a factor. <code>levels(f); levels(f) <- c("x", "y", "z")</code>

Use unclass() to see its structure

Inspect Factors

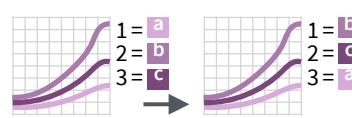
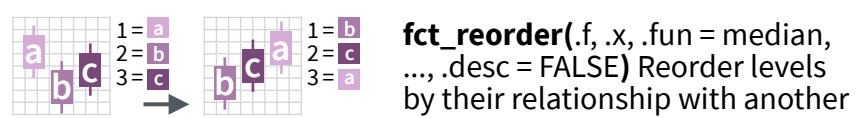
<code>a c b a</code>	<code>f n</code>	<code>fct_count(f, sort = FALSE, prop = FALSE)</code> Count the number of values with each level. <code>fct_count(f)</code>
<code>a b a</code>	<code>a 2 b 1 c 1</code>	<code>fct_match(f, lvls)</code> Check for lvls in f. <code>fct_match(f, "a")</code>
<code>a b a</code>	<code>a b 1 = a 2 = b</code>	<code>fct_unique(f)</code> Return the unique values, removing duplicates. <code>fct_unique(f)</code>

Combine Factors

<code>a c 1 = a 2 = c</code>	<code>b a 1 = a 2 = b</code>	<code>fct_c(...)</code> Combine factors with different levels. Also <code>fct_cross()</code> . <code>f1 <- factor(c("a", "c"))</code> <code>f2 <- factor(c("b", "a"))</code> <code>fct_c(f1, f2)</code>
<code>a b 1 = a 2 = b</code>	<code>a b 1 = a 2 = b 3 = c</code>	<code>fct_unify(fs, levels = lvl_union(fs))</code> Standardize levels across a list of factors. <code>fct_unify(list(f2, f1))</code>

Change the order of levels

<code>a c b a</code>	<code>a c b a</code>	<code>fct_relevel(.f, ..., after = 0L)</code> Manually reorder factor levels. <code>fct_relevel(f, c("b", "c", "a"))</code>
<code>c c a</code>	<code>c c a</code>	<code>fct_infreq(f, ordered = NA)</code> Reorder levels by the frequency in which they appear in the data (highest frequency first). Also <code>fct_inseq()</code> . <code>f3 <- factor(c("c", "c", "a"))</code> <code>fct_infreq(f3)</code>
<code>b a</code>	<code>b a</code>	<code>fct_inorder(f, ordered = NA)</code> Reorder levels by order in which they appear in the data. <code>fct_inorder(f2)</code>
<code>a b c</code>	<code>a b c</code>	<code>fct_rev(f)</code> Reverse level order. <code>f4 <- factor(c("a", "b", "c"))</code> <code>fct_rev(f4)</code>
<code>a b c</code>	<code>a b c</code>	<code>fct_shift(f)</code> Shift levels to left or right, wrapping around end. <code>fct_shift(f4)</code>
<code>a b c</code>	<code>a b c</code>	<code>fct_shuffle(f, n = 1L)</code> Randomly permute order of factor levels. <code>fct_shuffle(f4)</code>



Change the value of levels

<code>a c b a</code>	<code>v z x v</code>	<code>fct_recode(.f, ...)</code> Manually change levels. Also <code>fct_relabel()</code> which obeys purrr::map syntax to apply a function or expression to each level. <code>fct_recode(f, v = "a", x = "b", z = "c")</code> <code>fct_relabel(f, ~ paste0("x", .x))</code>
<code>a c b a</code>	<code>2 1 3 2</code>	<code>fct_anon(f, prefix = "")</code> Anonymize levels with random integers. <code>fct_anon(f)</code>
<code>a c b a</code>	<code>x c x x</code>	<code>fctCollapse(.f, ..., other_level = NULL)</code> Collapse levels into manually defined groups. <code>fct_collapse(f, x = c("a", "b"))</code>
<code>a c b a</code>	<code>a Other Other a</code>	<code>fct_lump_min(f, min, w = NULL, other_level = "Other")</code> Lumps together factors that appear fewer than min times. Also <code>fct_lump_n()</code> , <code>fct_lump_prop()</code> , and <code>fct_lump_lowfreq()</code> . <code>fct_lump_min(f, min = 2)</code>
<code>a c b a</code>	<code>a b Other Other b a</code>	<code>fct_other(f, keep, drop, other_level = "Other")</code> Replace levels with "other." <code>fct_other(f, keep = c("a", "b"))</code>

Add or drop levels

<code>a b 1 = a 2 = b 3 = x</code>	<code>a b 1 = a 2 = b</code>	<code>fct_drop(f, only)</code> Drop unused levels. <code>f5 <- factor(c("a", "b"), c("a", "b", "x"))</code> <code>f6 <- fct_drop(f5)</code>
<code>a b 1 = a 2 = b</code>	<code>a b 1 = a 2 = b 3 = x</code>	<code>fct_expand(f, ...)</code> Add levels to a factor. <code>fct_expand(f6, "x")</code>
<code>a b NA</code>	<code>a b x 3 = x</code>	<code>fct_explicit_na(f, na_level = "(Missing)")</code> Assigns a level to NAs to ensure they appear in plots, etc. <code>fct_explicit_na(factor(c("a", "b", NA)))</code>