

# Package ‘tidyseurat’

November 19, 2020

**Type** Package

**Title** Brings Seurat to the Tidyverse

**Version** 0.1.17

**Description** It creates an invisible layer that allow to see the 'Seurat' object as tibble and interact seamlessly with the tidyverse.

**License** GPL-3

**Depends** R (>= 4.0.0),  
Seurat

**Imports** tibble,  
dplyr,  
magrittr,  
tidyr,  
ggplot2,  
rlang,  
purrr,  
lifecycle,  
methods,  
plotly,  
tidyselect,  
utils,  
ellipsis,  
pillar,  
stringr,  
cli,  
fansi

**Suggests** testthat,  
knitr,  
GGally,  
markdown

**VignetteBuilder** knitr

**RdMacros** lifecycle

**Biarch** true

**biocViews** AssayDomain, Infrastructure, RNASeq, DifferentialExpression, GeneExpression, Normalization, Clustering, QualityControl, Sequencing, Transcription, Transcriptomics

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

## R topics documented:

as_tibble . . . . .	2
bind . . . . .	3
cell_type_df . . . . .	4
count . . . . .	4
ggplot . . . . .	5
join_transcripts . . . . .	6
pbmc_small . . . . .	7
pbmc_small_nested_interactions . . . . .	7
pivot_longer . . . . .	8
plot_ly . . . . .	9
print . . . . .	13
tidy . . . . .	14
unnest . . . . .	14
<b>Index</b>	<b>17</b>

---

as\_tibble

Coerce lists, matrices, and more to data frames

---

### Description

```
‘r lifecycle::badge("maturing")‘
```

‘as\_tibble()’ turns an existing object, such as a data frame or matrix, into a so-called tibble, a data frame with class [‘tbl\_df’]. This is in contrast with [tibble()], which builds a tibble from individual columns. ‘as\_tibble()’ is to [‘tibble()’] as [base::as.data.frame()] is to [base::data.frame()].

‘as\_tibble()’ is an S3 generic, with methods for: \* [‘data.frame’][base::data.frame()]: Thin wrapper around the ‘list’ method that implements tibble’s treatment of [rownames]. \* [‘matrix’][methods::matrix-class], [‘poly’][stats::poly()], [‘ts’][stats::ts()], [‘table’][base::table()] \* Default: Other inputs are first coerced with [base::as.data.frame()].

### Arguments

x	A data frame, list, matrix, or other object that could reasonably be coerced to a tibble.
...	Unused, for extensibility.
rownames	How to treat existing row names of a data frame or matrix: * ‘NULL’: remove row names. This is the default. * ‘NA’: keep row names. * A string: the name of a new column. Existing rownames are transferred into this column and the ‘row.names’ attribute is deleted. Read more in [rownames].
.name_repair	see tidyrr For compatibility only, do not use for new code.

### Value

A tibble

## Row names

The default behavior is to silently remove row names.

New code should explicitly convert row names to a new column using the ‘rownames’ argument.

For existing code that relies on the retention of row names, call ‘pkgconfig::set\_config("tibble::rownames" = NA)’ in your script or in your package’s [.onLoad()] function.

## Life cycle

Using ‘as\_tibble()’ for vectors is superseded as of version 3.0.0, prefer the more expressive maturing ‘as\_tibble\_row()’ and ‘as\_tibble\_col()’ variants for new code.

## See Also

[tibble()] constructs a tibble from individual columns. [enframe()] converts a named vector to a tibble with a column of names and column of values. Name repair is implemented using [vctrs::vec\_as\_names()].

## Examples

```
pbmc_small %>% tidy %>% as_tibble()
```

---

bind

*Efficiently bind multiple data frames by row and column*


---

## Description

This is an efficient implementation of the common pattern of ‘do.call(rbind, dfs)’ or ‘do.call(cbind, dfs)’ for binding many data frames into one.

## Arguments

- ... Data frames to combine.  
Each argument can either be a data frame, a list that could be a data frame, or a list of data frames.  
When row-binding, columns are matched by name, and any missing columns will be filled with NA.  
When column-binding, rows are matched by position, so all data frames must have the same number of rows. To match by value, not position, see [mutate-joins].
- .id Data frame identifier.  
When ‘.id’ is supplied, a new column of identifiers is created to link each row to its original data frame. The labels are taken from the named arguments to ‘bind\_rows()’. When a list of data frames is supplied, the labels are taken from the names of the list. If no names are found a numeric sequence is used instead.
- add.cell.ids from Seurat 3.0 A character vector of length(x = c(x, y)). Appends the corresponding values to the start of each objects’ cell names.

## Details

The output of ‘bind\_rows()’ will contain a column if that column appears in any of the inputs.

Value

‘bind\_rows()’ and ‘bind\_cols()’ return the same type as the first input, either a data frame, ‘tbl\_df’, or ‘grouped\_df’.

Examples

```
`%>%` = magrittr::`%>%`
tt = pbmc_small %>% tidy
bind_rows(  tt, tt  )

tt_bind = tt %>% select(nCount_RNA ,nFeature_RNA)
tt %>% bind_cols(tt_bind)
```

---

cell_type_df	<i>Example data set 2</i>
--------------	---------------------------

---

Description

Example data set 2

Usage

```
cell_type_df
```

Format

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 80 rows and 2 columns.

---

count	<i>Count observations by group</i>
-------	------------------------------------

---

Description

‘count()’ lets you quickly count the unique values of one or more variables: ‘df’ ‘df’ ‘count()’ is paired with ‘tally()’, a lower-level helper that is equivalent to ‘df’ switching the summary from ‘n = n()’ to ‘n = sum(wt)’. ‘add\_count()’ are ‘add\_tally()’ are equivalents to ‘count()’ and ‘tally()’ but use ‘mutate()’ instead of ‘summarise()’ so that they add a new column with group-wise counts.

Usage

```
count (
  x,
  ...,
  wt = NULL,
  sort = FALSE,
  name = NULL,
  .drop = group_by_drop_default(x)
)
```

**Arguments**

<code>x</code>	A data frame, data frame extension (e.g. a tibble), or a lazy data frame (e.g. from dbplyr or dtplyr).
<code>...</code>	<code>&lt;[‘data-masking’][dplyr_data_masking]&gt;</code> Variables to group by.
<code>wt</code>	<code>&lt;[‘data-masking’][dplyr_data_masking]&gt;</code> Frequency weights. Can be ‘NULL’ or a variable: * If ‘NULL’ (the default), counts the number of rows in each group. * If a variable, computes ‘sum(wt)’ for each group.
<code>sort</code>	If ‘TRUE’, will show the largest groups at the top.
<code>name</code>	The name of the new column in the output. If omitted, it will default to ‘n’. If there’s already a column called ‘n’, it will error, and require you to specify the name.
<code>.drop</code>	For ‘count()’: if ‘FALSE’ will include counts for empty groups (i.e. for levels of factors that don’t exist in the data). Deprecated in ‘add_count()’ since it didn’t actually affect the output.

**Value**

An object of the same type as ‘.data’. ‘count()’ and ‘add\_count()’ group transiently, so the output has the same groups as the input.

**Examples**

```
`%>%` = magrittr::`%>%`
pbmc_small %>% tidy %>% count(groups)
```

---

ggplot

---

*Create a new ggplot from a tidyseurat object*


---

**Description**

‘ggplot()’ initializes a ggplot object. It can be used to declare the input data frame for a graphic and to specify the set of plot aesthetics intended to be common throughout all subsequent layers unless specifically overridden.

**Arguments**

<code>.data</code>	Default dataset to use for plot. If not already a data.frame, will be converted to one by [fortify()]. If not specified, must be supplied in each layer added to the plot.
<code>mapping</code>	Default list of aesthetic mappings to use for plot. If not specified, must be supplied in each layer added to the plot.
<code>...</code>	Other arguments passed on to methods. Not currently used.
<code>environment</code>	DEPRECATED. Used prior to tidy evaluation.

## Details

'ggplot()' is used to construct the initial plot object, and is almost always followed by '+' to add component to the plot. There are three common ways to invoke 'ggplot()':

The first method is recommended if all layers use the same data and the same set of aesthetics, although this method can also be used to add a layer using data from another data frame. See the first example below. The second method specifies the default data frame to use for the plot, but no aesthetics are defined up front. This is useful when one data frame is used predominantly as layers are added, but the aesthetics may vary from one layer to another. The third method initializes a skeleton 'ggplot' object which is fleshed out as layers are added. This method is useful when multiple data frames are used to produce different layers, as is often the case in complex graphics.

## Value

A ggplot

## Examples

```
# Generate some sample data, then compute mean and standard deviation
# in each group
```

---

join_transcripts	<i>Extract and join information for transcripts.</i>
------------------	------------------------------------------------------

---

## Description

join\_transcripts() extracts and joins information for specified transcripts

## Usage

```
join_transcripts(
  .data,
  transcripts = NULL,
  all = FALSE,
  exclude_zeros = FALSE,
  shape = "long"
)
```

## Arguments

.data	A tidyseurat object
transcripts	A vector of transcript identifiers to join
all	If TRUE return all
exclude_zeros	If TRUE exclude zero values
shape	Format of the returned table "long" or "wide"

## Details

This function extracts information for specified transcripts and returns the information in either long or wide format.

**Value**

A 'tbl' containing the information for the specified transcripts

**Examples**

```
pbmc_small %>%  
  tidy %>%  
  join_transcripts(transcripts = c("HLA-DRA", "LYZ"))
```

---

pbmc\_small

*Example data set*

---

**Description**

Example data set

**Usage**

```
pbmc_small
```

**Format**

An object of class `Seurat` with 230 rows and 80 columns.

---

pbmc\_small\_nested\_interactions

*Example data set 2*

---

**Description**

Example data set 2

**Usage**

```
pbmc_small_nested_interactions
```

**Format**

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 100 rows and 9 columns.

---

pivot\_longer

---

*Pivot data from wide to long*


---

## Description

### Maturing

'pivot\_longer()' "lengthens" data, increasing the number of rows and decreasing the number of columns. The inverse transformation is [pivot\_wider()]

Learn more in 'vignette("pivot")'.

## Arguments

data	A data frame to pivot.
cols	<['tidy-select'] [tidyr_tidy_select]> Columns to pivot into longer format.
names_to	<p>A string specifying the name of the column to create from the data stored in the column names of 'data'.</p> <p>Can be a character vector, creating multiple columns, if 'names_sep' or 'names_pattern' is provided. In this case, there are two special values you can take advantage of:</p> <ul style="list-style-type: none"> <li>* 'NA' will discard that component of the name.</li> <li>* '.value' indicates that component of the name defines the name of the column containing the cell values, overriding 'values_to'.</li> </ul>
names_prefix	A regular expression used to remove matching text from the start of each variable name.
names_sep, names_pattern	<p>If 'names_to' contains multiple values, these arguments control how the column name is broken up.</p> <p>'names_sep' takes the same specification as [separate()], and can either be a numeric vector (specifying positions to break on), or a single string (specifying a regular expression to split on).</p> <p>'names_pattern' takes the same specification as [extract()], a regular expression containing matching groups ('()').</p> <p>If these arguments do not give you enough control, use 'pivot_longer_spec()' to create a spec object and process manually as needed.</p>
names_repair	What happens if the output has invalid column names? The default, "check_unique" is to error if the columns are duplicated. Use "minimal" to allow duplicates in the output, or "unique" to de-duplicated by adding numeric suffixes. See [vctrs::vec_as_names()] for more options.
values_to	A string specifying the name of the column to create from the data stored in cell values. If 'names_to' is a character containing the special '.value' sentinel, this value will be ignored, and the name of the value column will be derived from part of the existing column names.
values_drop_na	<p>If 'TRUE', will drop rows that contain only 'NA's in the 'value_to' column. This effectively converts explicit missing values to implicit missing values, and should generally be used only when missing values in 'data' were created by its structure.</p>



```
names_transform, values_transform
```

A list of column name-function pairs. Use these arguments if you need to change the type of specific columns. For example, `'names_transform = list(week = as.integer)'` would convert a character week variable to an integer.

```
names_ptypes, values_ptypes
```

A list of column name-prototype pairs. A prototype (or ptype for short) is a zero-length vector (like `'integer()'` or `'numeric()'`) that defines the type, class, and attributes of a vector. Use these arguments to confirm that the created columns are the types that you expect.

If not specified, the type of the columns generated from `'names_to'` will be character, and the type of the variables generated from `'values_to'` will be the common type of the input columns used to generate them.

```
...
```

Additional arguments passed on to methods.

### Details

`'pivot_longer()'` is an updated approach to `[gather()]`, designed to be both simpler to use and to handle more use cases. We recommend you use `'pivot_longer()'` for new code; `'gather()'` isn't going away but is no longer under active development.

### Value

A tidyseurat object or a tibble depending on input

### Examples

```
# See vignette("pivot") for examples and explanation

library(dplyr)
pbmc_small %>% tidy %>% pivot_longer(c(orig.ident, groups), names_to = "name", values_to
```

---

plot\_ly

*Initiate a plotly visualization*

---

### Description

This function maps R objects to `[plotly.js]`(<https://plot.ly/javascript/>), an (MIT licensed) web-based interactive charting library. It provides abstractions for doing common things (e.g. mapping data values to fill colors (via `'color'`) or creating [animation]s (via `'frame'`)) and sets some different defaults to make the interface feel more 'R-like' (i.e., closer to `[plot()]` and `[ggplot2::qplot()]`).

### Usage

```
plot_ly(
  data = data.frame(),
  ...,
  type = NULL,
  name = NULL,
  color = NULL,
  colors = NULL,
  alpha = NULL,
```

```

stroke = NULL,
strokes = NULL,
alpha_stroke = 1,
size = NULL,
sizes = c(10, 100),
span = NULL,
spans = c(1, 20),
symbol = NULL,
symbols = NULL,
linetype = NULL,
linetypes = NULL,
split = NULL,
frame = NULL,
width = NULL,
height = NULL,
source = "A"
)

```

## Arguments

data	A data frame (optional) or [crosstalk::SharedData] object.
...	Arguments (i.e., attributes) passed along to the trace 'type'. See [schema()] for a list of acceptable attributes for a given trace 'type' (by going to 'traces' -> 'type' -> 'attributes'). Note that attributes provided at this level may override other arguments (e.g. 'plot_ly(x=1:10, y=1:10, color=I("red"), marker=list(color="blue"))').
type	A character string specifying the trace type (e.g. "scatter", "bar", "box", etc). If specified, it <i>always</i> creates a trace, otherwise
name	Values mapped to the trace's name attribute. Since a trace can only have one name, this argument acts very much like 'split' in that it creates one trace for every unique value.
color	Values mapped to relevant 'fill-color' attribute(s) (e.g. [fillcolor](https://plot.ly/r/reference#scatter-fillcolor), [marker.color](https://plot.ly/r/reference#scatter-marker-color), [textfont.color](https://plot.ly/r/reference#scatter-textfont-color), etc.). The mapping from data values to color codes may be controlled using 'colors' and 'alpha', or avoided altogether via [I()] (e.g., 'color=I("red")'). Any color understood by [grDevices::col2rgb()] may be used in this way.
colors	Either a colorbrewer2.org palette name (e.g. "YlOrRd" or "Blues"), or a vector of colors to interpolate in hexadecimal "#RRGGBB" format, or a color interpolation function like 'colorRamp()'.
alpha	A number between 0 and 1 specifying the alpha channel applied to 'color'. Defaults to 0.5 when mapping to [fillcolor](https://plot.ly/r/reference#scatter-fillcolor) and 1 otherwise.
stroke	Similar to 'color', but values are mapped to relevant 'stroke-color' attribute(s) (e.g., [marker.line.color](https://plot.ly/r/reference#scatter-marker-line-color) and [line.color](https://plot.ly/r/reference#scatter-line-color) for filled polygons). If not specified, 'stroke' inherits from 'color'.
strokes	Similar to 'colors', but controls the 'stroke' mapping.
alpha_stroke	Similar to 'alpha', but applied to 'stroke'.
size	(Numeric) values mapped to relevant 'fill-size' attribute(s) (e.g., [marker.size](https://plot.ly/r/reference#scatter-marker-size), [textfont.size](https://plot.ly/r/reference#scatter-textfont-size), and [error_x.width](https://plot.ly/r/reference#scatter-error_x-width)). The mapping

	from data values to symbols may be controlled using ‘sizes’, or avoided altogether via [I()] (e.g., ‘size=I(30)’).
sizes	A numeric vector of length 2 used to scale ‘size’ to pixels.
span	(Numeric) values mapped to relevant ‘stroke-size’ attribute(s) (e.g., [marker.line.width](https://plot.ly/reference#scatter-marker-line-width), [line.width](https://plot.ly/r/reference#scatter-line-width) for filled polygons, and [error_x.thickness](https://plot.ly/r/reference#scatter-error_x-thickness)) The mapping from data values to symbols may be controlled using ‘spans’, or avoided altogether via [I()] (e.g., ‘span=I(30)’).
spans	A numeric vector of length 2 used to scale ‘span’ to pixels.
symbol	(Discrete) values mapped to [marker.symbol](https://plot.ly/r/reference#scatter-marker-symbol). The mapping from data values to symbols may be controlled using ‘symbols’, or avoided altogether via [I()] (e.g., ‘symbol=I("pentagon")’). Any [pch] value or [symbol name](https://plot.ly/r/reference#scatter-marker-symbol) may be used in this way.
symbols	A character vector of [pch] values or [symbol names](https://plot.ly/r/reference#scatter-marker-symbol).
linetype	(Discrete) values mapped to [line.dash](https://plot.ly/r/reference#scatter-line-dash). The mapping from data values to symbols may be controlled using ‘linetypes’, or avoided altogether via [I()] (e.g., ‘linetype=I("dash")’). Any ‘lty’ (see [par]) value or [dash name](https://plot.ly/r/reference#scatter-line-dash) may be used in this way.
linetypes	A character vector of ‘lty’ values or [dash names](https://plot.ly/r/reference#scatter-line-dash)
split	(Discrete) values used to create multiple traces (one trace per value).
frame	(Discrete) values used to create animation frames.
width	Width in pixels (optional, defaults to automatic sizing).
height	Height in pixels (optional, defaults to automatic sizing).
source	a character string of length 1. Match the value of this string with the source argument in [event_data()] to retrieve the event data corresponding to a specific plot (shiny apps can have multiple plots).

## Details

Unless ‘type’ is specified, this function just initiates a plotly object with ‘global’ attributes that are passed onto downstream uses of [add\_trace()] (or similar). A [formula] must always be used when referencing column name(s) in ‘data’ (e.g. ‘plot\_ly(mtcars, x=~wt)’). Formulas are optional when supplying values directly, but they do help inform default axis/scale titles (e.g., ‘plot\_ly(x=mtcars\$wt)’ vs ‘plot\_ly(x=~mtcars\$wt)’)

## Value

A plotly

## Author(s)

Carson Sievert

## References

<<https://plotly-r.com/overview.html>>

**See Also**

- For initializing a plotly-geo object: [plot\_geo()]
- For initializing a plotly-mapbox object: [plot\_mapbox()]
- For translating a ggplot2 object to a plotly object: [ggplotly()]
- For modifying any plotly object: [layout()], [add\_trace()], [style()]
- For linked brushing: [highlight()]
- For arranging multiple plots: [subplot()], [crosstalk::bscols()]
- For inspecting plotly objects: [plotly\_json()]
- For quick, accurate, and searchable plotly.js reference: [schema()]

**Examples**

```
## Not run:
# plot_ly() tries to create a sensible plot based on the information you
# give it. If you don't provide a trace type, plot_ly() will infer one.
plot_ly(economics, x=~pop)
plot_ly(economics, x=~date, y=~pop)
# plot_ly() doesn't require data frame(s), which allows one to take
# advantage of trace type(s) designed specifically for numeric matrices
plot_ly(z=~volcano)
plot_ly(z=~volcano, type="surface")

# plotly has a functional interface: every plotly function takes a plotly
# object as it's first input argument and returns a modified plotly object
add_lines(plot_ly(economics, x=~date, y=~ unemploy / pop))

# To make code more readable, plotly imports the pipe operator from magrittr
economics %>%
  plot_ly(x=~date, y=~ unemploy / pop) %>%
  add_lines()

# Attributes defined via plot_ly() set 'global' attributes that
# are carried onto subsequent traces, but those may be over-written
plot_ly(economics, x=~date, color=I("black")) %>%
  add_lines(y=~uempmed) %>%
  add_lines(y=~psavert, color=I("red"))

# Attributes are documented in the figure reference -> https://plot.ly/r/reference
# You might notice plot_ly() has named arguments that aren't in this figure
# reference. These arguments make it easier to map abstract data values to
# visual attributes.
p <- plot_ly(iris, x=~Sepal.Width, y=~Sepal.Length)
add_markers(p, color=~Petal.Length, size=~Petal.Length)
add_markers(p, color=~Species)
add_markers(p, color=~Species, colors="Set1")
add_markers(p, symbol=~Species)
add_paths(p, linetype=~Species)

## End(Not run)
```

print

*Printing tibbles***Description**

```
'r lifecycle::badge("maturing")'
```

One of the main features of the `'tbl_df'` class is the printing:

\* Tibbles only print as many rows and columns as fit on one screen, supplemented by a summary of the remaining rows and columns. \* Tibble reveals the type of each column, which keeps the user informed about whether a variable is, e.g., `'<chr>'` or `'<fct>'` (character versus factor).

Printing can be tweaked for a one-off call by calling `'print()'` explicitly and setting arguments like `'n'` and `'width'`. More persistent control is available by setting the options described below.

**Usage**

```
## S3 method for class 'tidyseurat'
print(x, ..., n = NULL, width = NULL, n_extra = NULL)
```

**Arguments**

<code>x</code>	Object to format or print.
<code>...</code>	Other arguments passed on to individual methods.
<code>n</code>	Number of rows to show. If <code>'NULL'</code> , the default, will print all rows if less than option <code>'tibble.print_max'</code> . Otherwise, will print <code>'tibble.print_min'</code> rows.
<code>width</code>	Width of text output to generate. This defaults to <code>'NULL'</code> , which means use <code>'getOption("tibble.width")'</code> or (if also <code>'NULL'</code> ) <code>'getOption("width")'</code> ; the latter displays only the columns that fit on one screen. You can also set <code>'options(tibble.width = Inf)'</code> to override this default and always print all columns.
<code>n_extra</code>	Number of extra columns to print abbreviated information for, if the width is too small for the entire tibble. If <code>'NULL'</code> , the default, will print information about at most <code>'tibble.max_extra_cols'</code> extra columns.

**Value**

Nothing

**Package options**

The following options are used by the tibble and pillar packages to format and print `'tbl_df'` objects. Used by the formatting workhorse `'trunc_mat()'` and, therefore, indirectly, by `'print.tbl()'`.

\* `'tibble.print_max'`: Row number threshold: Maximum number of rows printed. Set to `'Inf'` to always print all rows. Default: 20. \* `'tibble.print_min'`: Number of rows printed if row number threshold is exceeded. Default: 10. \* `'tibble.width'`: Output width. Default: `'NULL'` (use `'width'` option). \* `'tibble.max_extra_cols'`: Number of extra columns printed in reduced form. Default: 100.

- `pillar.bold`: Use bold font, e.g. for column headers? This currently defaults to `FALSE`, because many terminal fonts have poor support for bold fonts.
- `pillar.subtle`: Use subtle style, e.g. for row numbers and data types? Default: `TRUE`.

- `pillar.subtle_num`: Use subtle style for insignificant digits? Default: FALSE, is also affected by the `pillar.subtle` option.
- `pillar.neg`: Highlight negative numbers? Default: TRUE.
- `pillar.sigfig`: The number of significant digits that will be printed and highlighted, default: 3. Set the `pillar.subtle` option to FALSE to turn off highlighting of significant digits.
- `pillar.min_title_chars`: The minimum number of characters for the column title, default: 15. Column titles may be truncated up to that width to save horizontal space. Set to Inf to turn off truncation of column titles.
- `pillar.min_chars`: The minimum number of characters wide to display character columns, default: 0. Character columns may be truncated up to that width to save horizontal space. Set to Inf to turn off truncation of character columns.

## Examples

```
library(dplyr)
pbmc_small %>% tidy() %>% print()
```

---

<code>tidy</code>	<i>tidy for seurat</i>
-------------------	------------------------

---

## Description

tidy for seurat

## Usage

```
tidy(object)
```

## Arguments

`object`                      A Seurat object

## Value

A tidyseurat object

---

<code>unnest</code>	<i>unnest</i>
---------------------	---------------

---

## Description

Given a regular expression with capturing groups, ‘`extract()`’ turns each group into a new column. If the groups don’t match, or the input is NA, the output will be NA.

Convenience function to paste together multiple columns into one.

Given either a regular expression or a vector of character positions, ‘`separate()`’ turns a single character column into multiple columns.

**Arguments**

<code>cols</code>	<[‘tidy-select’][tidyr_tidy_select]> Columns to unnest. If you ‘unnest()’ multiple columns, parallel entries must be of compatible sizes, i.e. they’re either equal or length 1 (following the standard tidyverse recycling rules).
<code>names_sep</code>	<p>If ‘NULL’, the default, the names will be left as is. In ‘nest()’, inner names will come from the former outer names; in ‘unnest()’, the new outer names will come from the inner names.</p> <p>If a string, the inner and outer names will be used together. In ‘nest()’, the names of the new outer columns will be formed by pasting together the outer and the inner column names, separated by ‘names_sep’. In ‘unnest()’, the new inner names will have the outer names (+ ‘names_sep’) automatically stripped. This makes ‘names_sep’ roughly symmetric between nesting and unnesting.</p>
<code>keep_empty</code>	See tidyr::unnest
<code>names_repair</code>	See tidyr::unnest
<code>ptype</code>	See tidyr::unnest
<code>.drop</code>	See tidyr::unnest
<code>.id</code>	tidyr::unnest
<code>.preserve</code>	See tidyr::unnest
<code>.data</code>	A tbl. (See tidyr)
<code>.names_sep</code>	See tidyr::nest
<code>into</code>	Names of new variables to create as character vector. Use ‘NA’ to omit the variable in the output.
<code>regex</code>	a regular expression used to extract the desired values. There should be one group (defined by ‘()’) for each element of ‘into’.
<code>convert</code>	<p>If ‘TRUE’, will run [type.convert()] with ‘as.is = TRUE’ on new columns. This is useful if the component columns are integer, numeric or logical.</p> <p>NB: this will cause string “NA”’s to be converted to ‘NA’.</p>
<code>data</code>	A data frame.
<code>col</code>	<p>The name of the new column, as a string or symbol.</p> <p>This argument is passed by expression and supports [quasiquote][rlang::quasiquote] (you can unquote strings and symbols). The name is captured from the expression with [rlang::ensym()] (note that this kind of interface where symbols do not represent actual objects is now discouraged in the tidyverse; we support it here for backward compatibility).</p>
<code>...</code>	<[‘tidy-select’][tidyr_tidy_select]> Columns to unite
<code>na.rm</code>	If ‘TRUE’, missing values will be remove prior to uniting each value.
<code>remove</code>	If ‘TRUE’, remove input columns from output data frame.
<code>sep</code>	<p>Separator between columns.</p> <p>If character, ‘sep’ is interpreted as a regular expression. The default value is a regular expression that matches any sequence of non-alphanumeric values.</p> <p>If numeric, ‘sep’ is interpreted as character positions to split at. Positive values start at 1 at the far-left of the string; negative value start at -1 at the far-right of the string. The length of ‘sep’ should be one less than ‘into’.</p>

<code>extra</code>	<p>If ‘sep’ is a character vector, this controls what happens when there are too many pieces. There are three valid options:</p> <ul style="list-style-type: none"> <li>* "warn" (the default): emit a warning and drop extra values.</li> <li>* "drop": drop any extra values without a warning.</li> <li>* "merge": only splits at most ‘length(into)’ times</li> </ul>
<code>fill</code>	<p>If ‘sep’ is a character vector, this controls what happens when there are not enough pieces. There are three valid options:</p> <ul style="list-style-type: none"> <li>* "warn" (the default): emit a warning and fill from the right</li> <li>* "right": fill with missing values on the right</li> <li>* "left": fill with missing values on the left</li> </ul>

### Value

A tidyseurat objector a tibble depending on input  
 A tidyseurat objector a tibble depending on input  
 A tidyseurat objector a tibble depending on input  
 A tidyseurat objector a tibble depending on input  
 A tidyseurat objector a tibble depending on input

### See Also

[`separate()`] to split up by a separator.  
 [`separate()`], the complement.  
 [`unite()`], the complement, [`extract()`] which uses regular expression capturing groups.

### Examples

```
library(dplyr)
pbmc_small %>% tidy %>% nest(data = -groups) %>% unnest(data)
```

```
library(dplyr)
pbmc_small %>% tidy %>% nest(data = -groups) %>% unnest(data)
```

```
pbmc_small %>% tidy %>% extract(groups, into = "g", regex = "g([0-9])", convert = TRUE)
```

```
pbmc_small %>% tidy %>% unite("new_col", c(orig.ident, groups))
```

```
un = pbmc_small %>% tidy %>% unite("new_col", c(orig.ident, groups))
un %>% separate(col = new_col, into= c("orig.ident", "groups"))
```



# Index

- \* **datasets**
  - cell\_type\_df, [4](#)
  - pbmc\_small, [7](#)
  - pbmc\_small\_nested\_interactions,  
[7](#)
- as\_tibble, [2](#)
- bind, [3](#)
- cell\_type\_df, [4](#)
- count, [4](#)
- extract (unnest), [14](#)
- ggplot, [5](#)
- join\_transcripts, [6](#)
- nest (unnest), [14](#)
- pbmc\_small, [7](#)
- pbmc\_small\_nested\_interactions,  
[7](#)
- pivot\_longer, [8](#)
- plot\_ly, [9](#)
- print, [13](#)
- separate (unnest), [14](#)
- tidy, [14](#)
- unite (unnest), [14](#)
- unnest, [14](#)