

my_df %>% **filter()**



my_df %>% **select()**



my_df %>% **mutate()**



my_df %>% **summarize()**

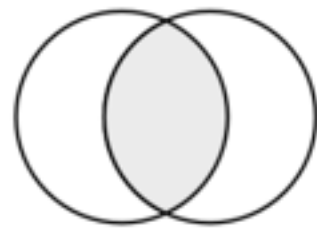


```
my_df %>% group_by() %>% summarize()
```

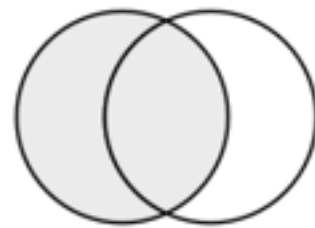


```
my_df %>% group_by() %>% mutate()
```

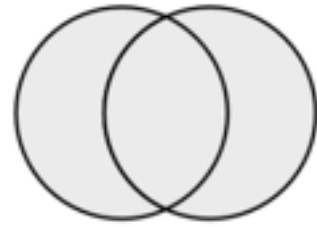




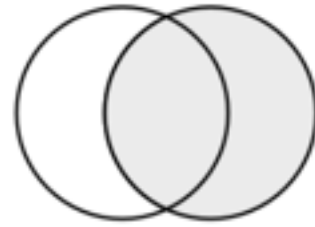
inner_join(x, y)



left_join(x, y)



full_join(x, y)



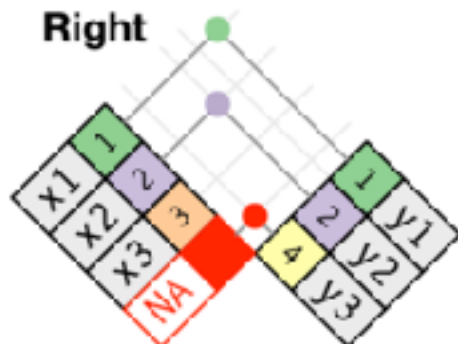
right_join(x, y)

Left



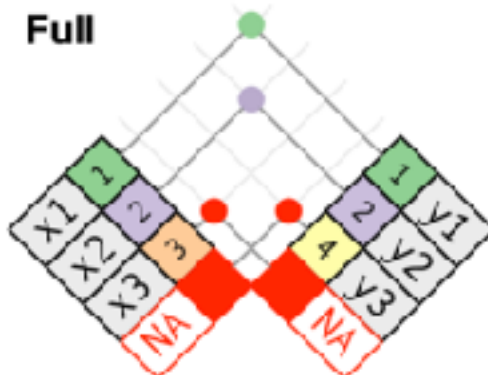
key	val_x	val_y
1	x1	y1
2	x2	y2
3	x3	NA

Right



key	val_x	val_y
1	x1	y1
2	x2	y2
4	NA	y3

Full



key	val_x	val_y
1	x1	y1
2	x2	y2
3	x3	NA
4	NA	y3

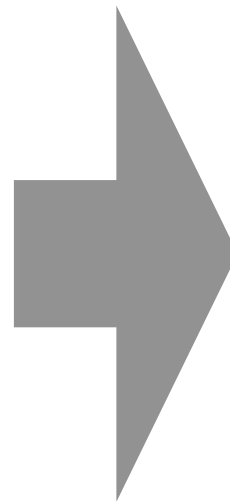
left_join()

Preserves original rows even if there isn't a match.

Make it your default join.

**certain data sets are
built-in to R**

iris
diamonds
mpg
table1...table5



not visible in Environment

- always available
- use for testing functions
- teaching examples

**your data must be
'read-in' to R**

visible in Environment

```
entrez_GO <- read_csv("~/Desktop/entrez_GO.csv")
```

any variable that you assign

visible in Environment

```
entrezGO_short <- entrezGO %>% head()
```

tidy data

country	year	cases	population
Afghanistan	1999	745	15987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	216766	1280425583

variables

each column is a
variable

country	year	cases	population
Afghanistan	1999	745	15987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	216766	1280425583

observations

each row is an
observation

country	year	cases	population
Afghanistan	1999	745	15987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	216766	1280425583

values

each value has its
own cell

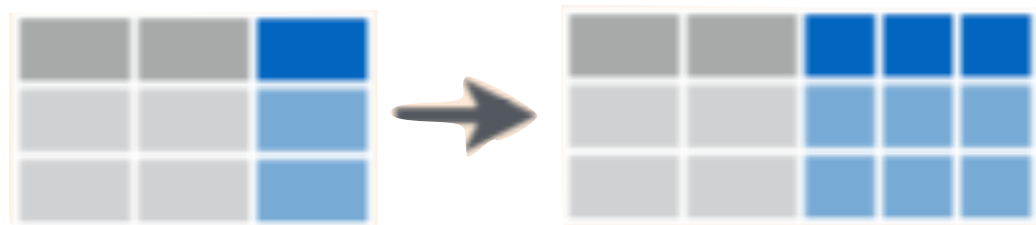
`tidyr::gather(key, value)`



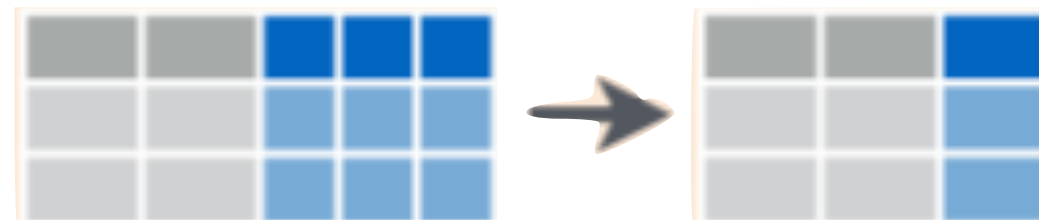
`tidyr::spread(key, value)`



`tidyr::separate()`



`tidyr::unite()`



AUTOMATION

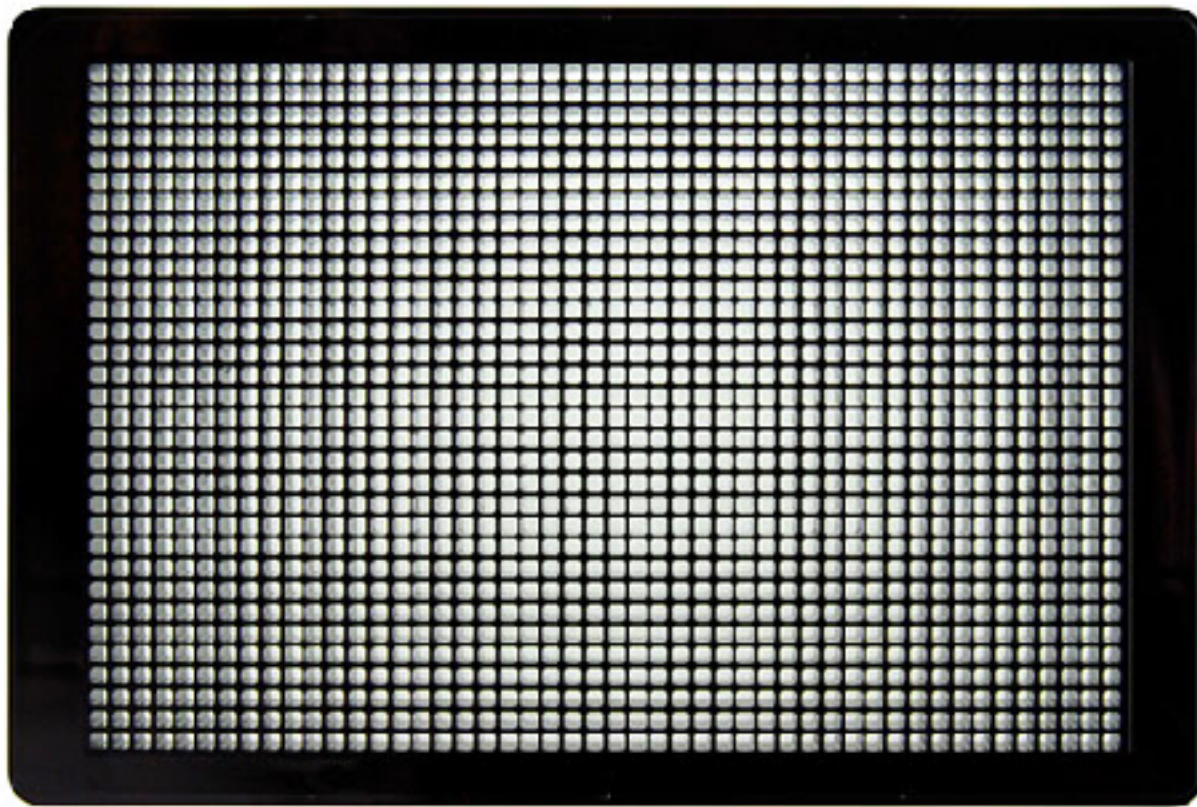
Iterate through many jobs that require the same analysis

```
job_ids <- c('job1','job2','job3')
```

```
for( i in job_ids ){ ...do some R ...}
```

- minimizes human error
- minimizes drudge work
- maximizes reproducibility

1532-well plate experimental results



x5

<http://tinyurl.com/screening-plates>

Which compounds significantly affect cell growth?

$$z = (\text{obs} - \text{mean}) / \text{sd}$$



R gGraduates - Over to you!



R for Data Science

text r4ds.had.co.nz

solutions jrnold.github.io/r4ds-exercise-solutions

Data Analysis for the Life Sciences

rafalab.github.io/pages/harvardx.html



DataCamp

datacamp.com/courses/dplyr-data-manipulation-r-tutorial

Statistical and Data Sciences via R

<https://moderndive.com/>

Mailing lists e.g. [sharpsightlabs](https://sharpsightlabs.com/)

R Users @ WEHI

Learn together

Help each other

YOUR COMMENTS PLEASE! surveyMonkey.com/r/RCourse_Mar19



R gRaduates - Over to you!



R HACKY HOUR

Thursday fortnights 3pm - 4pm

WEHI tea room

Bioinf. PhD student consultant

Help you solve your R issues

All welcome!!



R Users @ WEHI

Learn together

Help each other

YOUR COMMENTS PLEASE! surveymonkey.com/r/RCourse_Mar19