



tidyverse %>%

Manipulating data

Whereas filter() subsets a dataframe by row, select() returns a subset of the columns.

dplyr verbs

filter() ≡

str_detect() ≡

%in% ≡

is.na() ≡

complete.cases() ≡

select() ≡

contains() ≡

starts_with() ≡

everything() ≡

arrange() ≡

desc() ≡

ifelse() ≡

mutate() ≡

case_when() ≡

str_remove() ≡

rename() ≡

pull() ≡

summarize() ≡

n() ≡

group_by() ≡

group_by() %>% summarize() ≡

group_by() %>% mutate() ≡

groups() ≡

ungroup() ≡

count() ≡

sample_n() ≡

shortcut function

```
diamond_df %>%
  group_by(clarity, cut) %>%
  summarize(mean_price = mean(price),
            sd_price = sd(price),
            min_price = min(price),
            max_price = max(price),
            n_rows = n())
```

```
mpg_df %>%
  filter(class=='midsize') %>%
  select(class,manufacturer,displ,year) %>%
  arrange(displ) %>%
  ggplot(aes(x=class,y=displ)) + geom_boxplot()
```

```
diamond_df %>% select(-x,-y,-z) %>%
  mutate(price_label = ifelse(price > 5000,'expensive','cheap')) %>%
  ggplot(aes(x=price, fill = price_label)) +
  geom_histogram()
```

```
plate_tagged <- plate_long_num %>%
  mutate(well_tag = case_when(COL_num %in% c(1,23,25,47) ~ 'posCTRL',
                             COL_num %in% c(2,24,26,48) ~ 'negCTRL',
                             TRUE ~ 'test'))
```

Tidying & joining data

Reshaping data

pivot_wider() ≡

pivot_longer() ≡

Separating and uniting columns

separate() ≡

separate_rows() ≡

unite() ≡

complement function

```
GO_entrez_sep %>%
  separate_rows(entrez_multi, sep = ",") %>%
  rename(entrez_id = entrez_multi)
```

Joining dataframes

inner_join() ≡

left_join() ≡

right_join() ≡

full_join() ≡

bind_rows() / bind_cols ≡

Basic function

data_frame() ≡

tibble() ≡

as_tibble() ≡

read_excel() ≡

read_csv() ≡

read_delim() ≡

read_tsv() ≡

write_tsv() ≡

Pipeline

