

Tidy Data

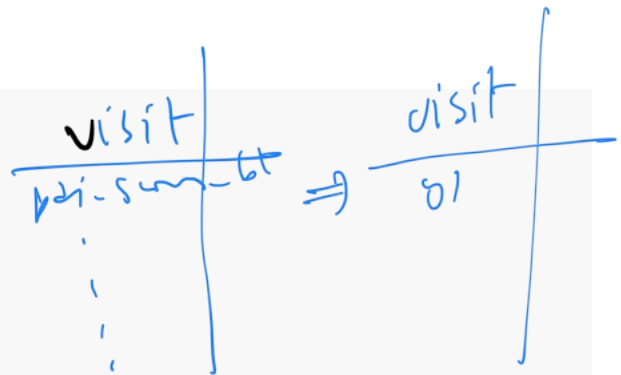
This document will show how to tidy data.

Pivot longer

```
pulse_df =  
  read_sas("data/public_pulse_data.sas7bdat") |>  
  janitor::clean_names()
```

This needs to go from wide to long format.

```
pulse_tidy_df =  
  pulse_df |>  
  pivot_longer(  
    cols = bdi_score_b1:bdi_score_12m,  
    names_to = "visit",  
    values_to = "bdi_scores",  
    names_prefix = "bdi_score_"  
  ) |>  
  mutate(  
    visit = replace(visit, visit == "b1", "00m")  
  ) |>  
  relocate(id, visit)
```



Do one more example

```
litters_df =  
  read_csv("data/FAS_litters.csv", na = c("NA", "", ".")) |>  
  janitor::clean_names() |>  
  pivot_longer(  
    cols = gd0_weight:gd18_weight,  
    names_to = "gd_time",  
    values_to = "weight"  
  ) |>  
  mutate(  
    gd_time = case_match(  
      gd_time,  
      "gd0_weight" ~ 0,  
      "gd18_weight" ~ 18  
    )  
  )
```

types Δ

```
## Rows: 49 Columns: 8
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr (2): Group, Litter Number
## dbl (6): GD0 weight, GD18 weight, GD of Birth, Pups born alive, Pups dead @ ...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
view(litters_df)
```

Pivot wider

Let's make up an analysis result table.

```
analysis_df =
  tibble(
    group = c("treatment", "treatment", "control", "control"),
    time = c("pre", "post", "pre", "post"),
    mean = c(4, 10, 4.2, 5)
  )
```

Pivot wider for human readability.

```
analysis_df |>
  pivot_wider(
    names_from = time,
    values_from = mean
  ) |>
  knitr::kable()
```

group	pre	post
treatment	4.0	10
control	4.2	5

Bind tables.

```
fellowship_ring =
  read_excel("data/LotR_Words.xlsx", range = "B3:D6") |>
  mutate(movie = "fellowship_ring")

two_towers =
  read_excel("data/LotR_Words.xlsx", range = "F3:H6") |>
  mutate(movie = "two_towers")

return_king =
  read_excel("data/LotR_Words.xlsx", range = "J3:L6") |>
  mutate(movie = "return_king")

lotr_df =
  bind_rows(fellowship_ring, two_towers, return_king) |>
  janitor::clean_names() |>
```

```

pivot_longer(
  cols = female:male,
  names_to = "sex",
  values_to = "words"
) |>
relocate(movie) |>
mutate(race = str_to_lower(race))

```

Join FAS datasets

Import litters dataset.

```

litters_df =
  read_csv("data/FAS_litters.csv", na = c("NA", "", ".")) |>
  janitor::clean_names() |>
  mutate(
    wt_gain = gd18_weight - gd0_weight
  ) |>
  separate(
    group, into = c("dose", "day_of_treatment"), sep = 3
  )

```

```

## Rows: 49 Columns: 8
## -- Column specification -----
## Delimiter: ","
## chr (2): Group, Litter Number
## dbl (6): GD0 weight, GD18 weight, GD of Birth, Pups born alive, Pups dead @ ...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

```

```
litters_df
```

```

## # A tibble: 49 x 10
##   dose day_of_treatment litter_number gd0_weight gd18_weight gd_of_birth
##   <chr> <chr>           <chr>         <dbl>      <dbl>      <dbl>
## 1 Con  7                #85           19.7       34.7       20
## 2 Con  7                #1/2/95/2     27        42        19
## 3 Con  7                #5/5/3/83/3-3 26        41.4      19
## 4 Con  7                #5/4/2/95/2   28.5       44.1      19
## 5 Con  7                #4/2/95/3-3   NA        NA        20
## 6 Con  7                #2/2/95/3-2   NA        NA        20
## 7 Con  7                #1/5/3/83/3-3/2 NA        NA        20
## 8 Con  8                #3/83/3-3     NA        NA        20
## 9 Con  8                #2/95/3       NA        NA        20
## 10 Con 8                #3/5/2/2/95   28.5       NA        20
## # i 39 more rows
## # i 4 more variables: pups_born_alive <dbl>, pups_dead_birth <dbl>,
## #   pups_survive <dbl>, wt_gain <dbl>

```

Import pups dataset next.

```
pups_df =
  read_csv("data/FAS_pups.csv", na = c("NA", "", ".")) |>
  janitor::clean_names() |>
  mutate(
    sex = case_match(
      sex,
      1 ~ "male",
      2 ~ "female"
    )
  )
```

```
## Rows: 313 Columns: 6
## -- Column specification -----
## Delimiter: ","
## chr (1): Litter Number
## dbl (5): Sex, PD ears, PD eyes, PD pivot, PD walk
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
pups_df
```

```
## # A tibble: 313 x 6
##   litter_number sex   pd_ears pd_eyes pd_pivot pd_walk
##   <chr>         <chr>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 #85          male     4      13      7      11
## 2 #85          male     4      13      7      12
## 3 #1/2/95/2    male     5      13      7      9
## 4 #1/2/95/2    male     5      13      8      10
## 5 #5/5/3/83/3-3 male     5      13      8      10
## 6 #5/5/3/83/3-3 male     5      14      6      9
## 7 #5/4/2/95/2  male    NA      14      5      9
## 8 #4/2/95/3-3  male     4      13      6      8
## 9 #4/2/95/3-3  male     4      13      7      9
## 10 #2/2/95/3-2 male     4      NA      8      10
## # i 303 more rows
```

Join the datasets!

```
fas_df =
  left_join(pups_df, litters_df, by = "litter_number") |>
  relocate(litter_number, dose, day_of_treatment)
```

```
litters_df = read_csv("./data/FAS_litters.csv")
```

```
## Rows: 49 Columns: 8
## -- Column specification -----
## Delimiter: ","
## chr (4): Group, Litter Number, GD0 weight, GD18 weight
## dbl (4): GD of Birth, Pups born alive, Pups dead @ birth, Pups survive
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```