II Data Reshaping

```
suppressWarnings(library(tidyverse))
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
suppressWarnings(library(data.table))
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

From Long to Wide

Convert all unique levels in a column (`names_from`) into new columns;

Fill rows under those columns with values from other columns (`values_from`).

```
d <- us_rent_income[c("GEOID", "NAME", "variable", "estimate")]
d %>% head
```

```
## # A tibble: 6 x 4
##   GEOID NAME    variable estimate
##   <chr> <chr>   <chr>       <dbl>
## 1 01    Alabama income      24476
## 2 01    Alabama rent          747
## 3 02    Alaska  income      32940
## 4 02    Alaska  rent         1200
## 5 04    Arizona income      27517
## 6 04    Arizona rent          972
```

```
d %>%
  pivot_wider(
    names_from = variable, #new columns
    values_from = estimate #values
  ) %>% head
```

```
## # A tibble: 6 x 4
```

```
##   GEOID NAME       income  rent
##   <chr> <chr>       <dbl> <dbl>
## 1 01    Alabama     24476   747
## 2 02    Alaska      32940  1200
## 3 04    Arizona     27517   972
## 4 05    Arkansas    23789   709
## 5 06    California  29454  1358
## 6 08    Colorado    32401  1125
```

```
d <- us_rent_income
d %>% head
```

```
## # A tibble: 6 x 5
##   GEOID NAME    variable estimate   moe
##   <chr> <chr>   <chr>       <dbl> <dbl>
## 1 01    Alabama income      24476   136
## 2 01    Alabama rent          747     3
## 3 02    Alaska  income      32940   508
## 4 02    Alaska  rent         1200    13
## 5 04    Arizona income      27517   148
## 6 04    Arizona rent          972     4
```

```
d %>%
  pivot_wider(
    names_from = variable, #new columns
    values_from = c(estimate, moe) #values
  ) %>% head
```

```
## # A tibble: 6 x 6
##   GEOID NAME       estimate_income estimate_rent moe_income moe_rent
##   <chr> <chr>                <dbl>         <dbl>      <dbl>    <dbl>
## 1 01    Alabama              24476           747        136        3
## 2 02    Alaska               32940          1200        508       13
## 3 04    Arizona              27517           972        148        4
## 4 05    Arkansas             23789           709        165        5
## 5 06    California           29454          1358        109        3
## 6 08    Colorado             32401          1125        109        5
```

```
d %>%
  pivot_wider(
    names_from = variable, #new columns
    values_from = estimate #values
  ) %>% head
```

```
## # A tibble: 6 x 5
##   GEOID NAME      moe income  rent
##   <chr> <chr>   <dbl>  <dbl> <dbl>
## 1 01    Alabama   136  24476    NA
## 2 01    Alabama     3     NA   747
## 3 02    Alaska    508  32940    NA
## 4 02    Alaska     13     NA  1200
## 5 04    Arizona   148  27517    NA
## 6 04    Arizona     4     NA   972
```

```
d %>%
  pivot_wider(
    names_from = variable,
```

```
    values_from = c(estimate, moe),
    names_vary = "slowest" #different column ordering
  ) %>% head
```

```
## # A tibble: 6 x 6
##   GEOID NAME       estimate_income moe_income estimate_rent moe_rent
##   <chr> <chr>                <dbl>      <dbl>         <dbl>    <dbl>
## 1 01    Alabama              24476        136           747        3
## 2 02    Alaska               32940        508          1200       13
## 3 04    Arizona              27517        148           972        4
## 4 05    Arkansas             23789        165           709        5
## 5 06    California           29454        109          1358        3
## 6 08    Colorado             32401        109          1125        5
```

From Wide to Long

Convert columns (`cols`) into levels of a new column (`names_to`);

Create a new column (`values_to`) to store values under those columns (`cols`).

```
d <- relig_income
d %>% head
```

```
## # A tibble: 6 x 11
##   religion  `<$10k` `$10-20k` `$20-30k` `$30-40k` `$40-50k` `$50-75k` `$75-100k`
##   <chr>       <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>      <dbl>
## 1 Agnostic       27        34        60        81        76       137        122
## 2 Atheist        12        27        37        52        35        70         73
## 3 Buddhist       27        21        30        34        33        58         62
## 4 Catholic      418       617       732       670       638      1116        949
## 5 Don't kn~      15        14        15        11        10        35         21
## 6 Evangeli~     575       869      1064       982       881      1486        949
## # ... with 3 more variables: `$100-150k` <dbl>, `>150k` <dbl>,
## #   `Don't know/refused` <dbl>
```

```
d %>%
  pivot_longer(!religion, #cols, columns to be converted into levels
               names_to = "income", #new column to store current column names
               values_to = "count" #new column to store current column values
               ) %>% head
```

```
## # A tibble: 6 x 3
##   religion income  count
##   <chr>    <chr>   <dbl>
## 1 Agnostic <$10k      27
## 2 Agnostic $10-20k    34
## 3 Agnostic $20-30k    60
## 4 Agnostic $30-40k    81
## 5 Agnostic $40-50k    76
## 6 Agnostic $50-75k   137
```