Counts and percentages

Data source

This program uses data from a study of sharing services (like sharing an automobile) and produces counts and percentages for a few demographic variables. There is a [data dictionary](#) that provides more details about the data.

Libraries

Here are the libraries you need for this program.

```r
library(readxl)
library(tidyverse)
```

Reading the data

Here is the code to read the data and show a glimpse. There are 31 columns total, but I am showing just a few of the columns here.

```r
fn <- "../data/sharing.xlsx"
sharing <- read_excel(fn)
glimpse(sharing[ , c(1, 5:7)])
```

```
Rows: 190
Columns: 4
$ id                <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1…
$ age               <chr> "36-45", "36-45", "over 60", "46-60", "36-45", "over…
$ gender            <chr> "M", "F", "F", "M", "F", "F", "F", "F", "F", "F", "F…
$ employment_status <chr> "employed", "employed", "employed", "employed", "emp…
```

Calculate counts and percentages for age group

```r
sharing |>
  group_by(age) |>
  summarize(n=n()) |>
  mutate(total=sum(n)) |>
  mutate(pct=100*n/total)
```

```
# A tibble: 6 × 4
  age        n total   pct
  <chr>  <int> <int> <dbl>
1 18-25     99   190  52.1
```

```
2 26-35        33    190 17.4
3 36-45        27    190 14.2
4 46-60        19    190 10
5 over 60       6    190  3.16
6 <NA>          6    190  3.16
```

The survey respondents were younger than the general population. About half of the survey respondents were 18 to 25 years old. Only 3% were over 60. Six ages were missing.

```
        sharing |>
          group_by(gender) |>
          summarize(n=n()) |>
          mutate(total=sum(n)) |>
          mutate(pct=100*n/total)
```

```
# A tibble: 3 × 4
  gender      n total   pct
  <chr>   <int> <int> <dbl>
1 F         114   190 60
2 M          70   190 36.8
3 <NA>        6   190  3.16
```

So, most of the people in this group are women, with men making up a little more than a third, and a small percentage didn't specify their gender.

```
        sharing |>
          group_by(employment_status) |>
          summarize(n=n()) |>
          mutate(total=sum(n)) |>
          mutate(pct=100*n/total)
```

```
# A tibble: 7 × 4
  employment_status         n total   pct
  <chr>                 <int> <int> <dbl>
1 employed                108   190 56.8
2 entrepreneur             13   190  6.84
3 full-time student        37   190 19.5
4 self-employed            16   190  8.42
5 temporarily unemployed    5   190  2.63
6 unemployed                5   190  2.63
7 <NA>                      6   190  3.16
```

Most people in this group are employed, with a significant portion being full-time students. A smaller number of people are entrepreneurs, self-employed, or unemployed, and a few did not specify their employment status.