`summarise` performs an aggregation, where it takes all rows in the data and reduces it to a single value.

Like `mutate`, you can specify the summaries you want to compute in the form `new_col = summary of (col)` and specify multiple columns in a single call.

Using the tidy long form TB data (https://github.com/datascienceprogram/ids_course_data/blob/master/tb_long.rds) you can compute a five number summary of the `count` variable for all observations in the data using summarise.

You can do this by copying and running the following code chunk in RStudio on your computer:

```
summarise(tb_long,
          min = min(count, na.rm = TRUE),
          first_quartile = quantile(count, 0.25, na.rm = TRUE),
          median = median(count, na.rm = TRUE),
          third_quartile = quantile(count, 0.75, na.rm = TRUE),
          max = max(count, na.rm = TRUE))
```

```
## # A tibble: 1 x 5
##     min first_quartile median third_quartile   max
##   <dbl>          <dbl>  <dbl>          <dbl> <dbl>
## 1     0              2     28            221 90830
```

It's worth noting that you can also perform summaries over different parts of the data using `group_by`.

Give it a go!

Continue to develop your skills with wrangling verbs by making your way through this exercise.

Within the **Comments**, share with other learners your understanding of the meaning of `na.rm = TRUE` in the `summarise` statement listed earlier in this step.

Then, return to RStudio on your computer and using `summarise` compute the:

- **number of observations in the data** (hint use `n()`).
- **number of observations that have missing counts** (hint use `is.na`).
- **proportion of observations that have missing counts.**

Tell us how you went

Share with other learners your results of using `summarise` to compute the number of observations in the data and the missing counts, and the proportions of observations with missing counts.
