

Midterm Exam: *Your Name Here*

All of these problems are from the *R for Data Science* book. The section numbers (e.g., “3.2.4 Exercises”) refer to sections in this book.

When solving these problems, you are allowed to use any method from the book or class, even if that method wasn’t yet covered when the exercise was presented in the book.

Because this is an exam, you need to do the work by yourself. Do not collaborate on this exam.

This exam is out of 60 points.

12.2.1 Exercises

(1) 12.2.1 Exercise 2 (10 pts)

Compute the `rate` for `table2`, and `table4a + table4b`. You will need to perform four operations:

1. Extract the number of TB cases per country per year.
2. Extract the matching population per country per year.
3. Divide cases by population, and multiply by 1000.
4. Store back in the appropriate place.

Which representation is easiest to work with? Which is hardest? Why?

12.3.3 Exercises

(2) 12.3.3 Exercise 1 (10 pts)

Why are `pivot_longer()` and `pivot_wider()` not perfectly symmetrical? Carefully consider the following example:

```
stocks <- tibble(  
  year   = c(2015, 2015, 2016, 2016),  
  half   = c( 1,    2,    1,    2),  
  return = c(1.88, 0.59, 0.92, 0.17)  
)  
stocks %>%  
  pivot_wider(names_from = year, values_from = return) %>%  
  pivot_longer(`2015`:`2016`, names_to = "year", values_to = "return")
```

(Hint: look at the variable types and think about column *names*.)

`pivot_longer()` has a `names_transform` argument. What does it do?

16.3.4 Exercises

(3) 16.3.4 Exercise 2 (10 pts)

Compare `dep_time`, `sched_dep_time` and `dep_delay`. I recommend looking at the distributions over an hour. Are they consistent? Explain your findings.

(4) 16.3.4 Exercise 4 (10 pts)

How does the average departure delay change over the course of a day? Should you use `dep_time` or `sched_dep_time`? Why?

(5) 16.3.4 Exercise 5 (10 pts)

On what day of the week should you leave if you want to minimize the chance of a departure delay?

(6) 16.3.4 Exercise 6 (10 pts)

What makes the distribution of `diamonds$carat` and `flights$sched_dep_time` similar?