013-Exercise-Question-3

Question 3 - R for Data Science (Page 366)

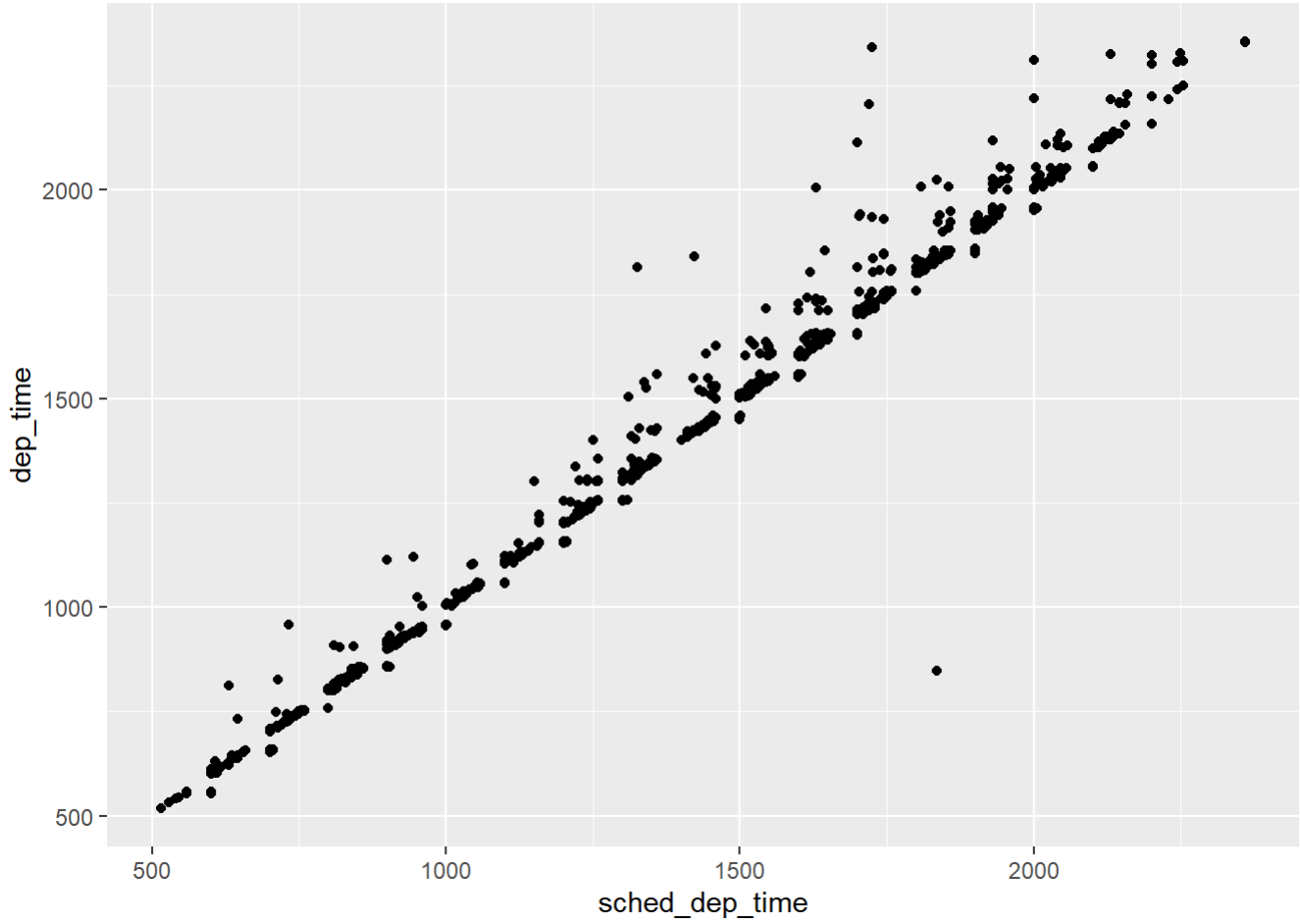
Converting `dep_time` and `sched_dep_time` to continuous time

```
# ------------------------------------------------------------
# Libraries
# ------------------------------------------------------------
library(tidyverse)
```

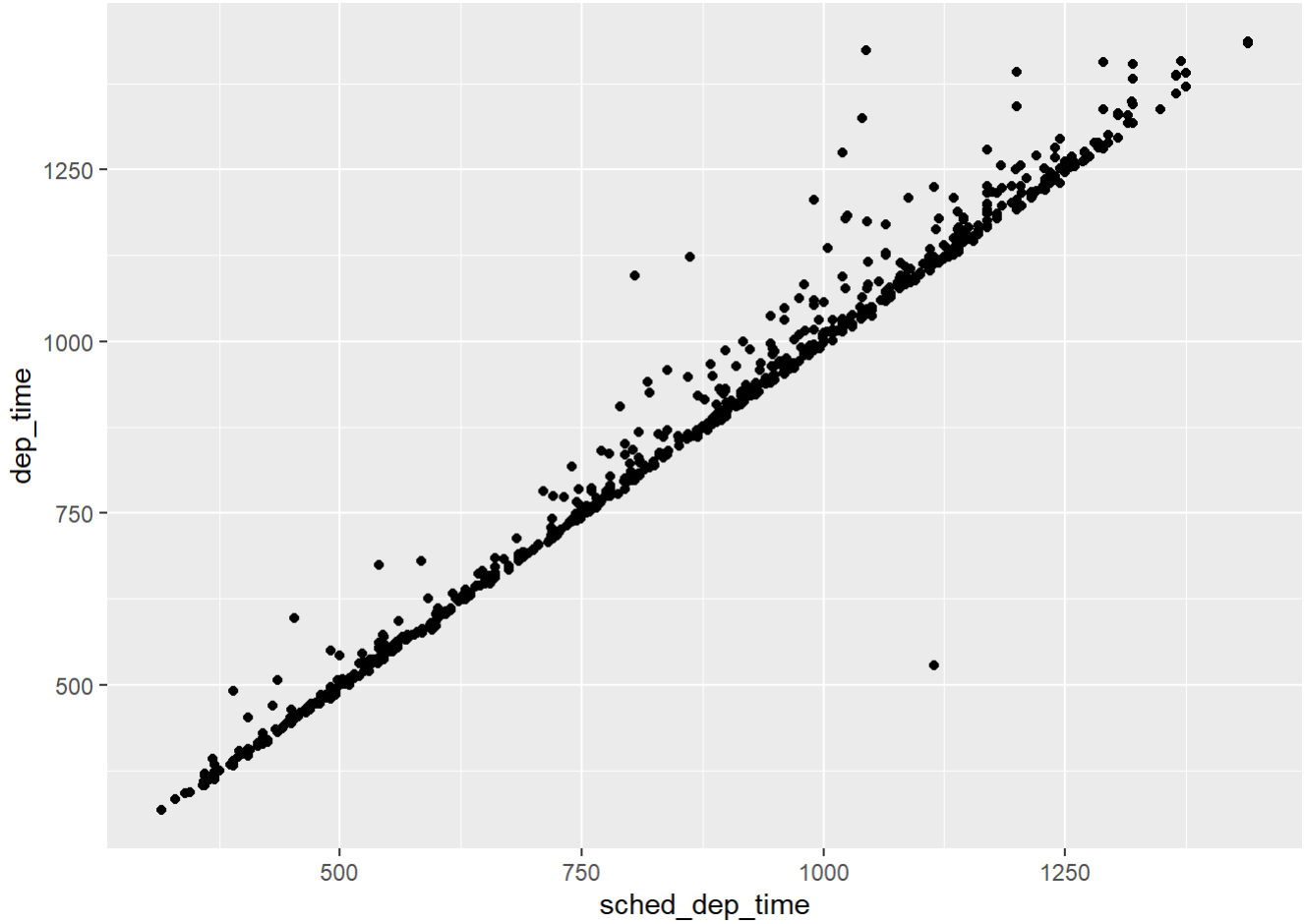
```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✓ dplyr     1.1.4     ✓ readr     2.1.5
✓ forcats   1.0.0     ✓ stringr   1.5.1
✓ ggplot2   4.0.1     ✓ tibble    3.2.1
✓ lubridate 1.9.4     ✓ tidyr     1.3.1
✓ purrr     1.0.4
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(nycflights13)

# ------------------------------------------------------------
# Problem description
# ------------------------------------------------------------
# The variables `dep_time` and `sched_dep_time` are stored in HHMM
# format (e.g. 517 = 5:17 AM). Although they look numeric, they are
# not continuous, which causes gaps between hours when plotting.

# ------------------------------------------------------------
# Demonstration of the problem
# ------------------------------------------------------------
flights |>
  filter(month == 1, day == 1) |>
  ggplot(aes(x = sched_dep_time, y = dep_time)) +
  geom_point()
```

```
Warning: Removed 4 rows containing missing values or values outside the scale range
(`geom_point()`).
```




```
# ------------------------------------------------------------
# Solution
# ------------------------------------------------------------
# To create a truthful representation of time, we convert HHMM
# values into minutes since midnight.
#
# Steps:
# 1. Extract hours using integer division (%/% 100)
# 2. Extract minutes using modulo (%% 100)
# 3. Convert hours to minutes and add minutes component

flights |>
  filter(month == 1, day == 1, !is.na(dep_time)) |>
  mutate(
    sched_dep_time = (sched_dep_time %/% 100) * 60 +
                     (sched_dep_time %% 100),
    dep_time       = (dep_time %/% 100) * 60 +
                     (dep_time %% 100)
  ) |>
  ggplot(aes(x = sched_dep_time, y = dep_time)) +
  geom_point()
```




```
# ------------------------------------------------------------
# Interpretation
# ------------------------------------------------------------
# After converting to minutes since midnight, the artificial gaps
# between hours disappear. Time is now represented as a continuous
# variable, making both plotting and computation meaningful.
```

Because `dep_time` and `sched_dep_time` are stored in HHMM format, they are not continuous; converting them to minutes since midnight removes artificial gaps and allows valid computation and visualization.