Solution to exercise 2

1. Manipulate `airquality` dataset

i) Load the `airquality` dataset

```
data(airquality)
```

```
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

ii) Rename the column headers to lower case

```
# convert column names to lower case in the classical way
airquality1 <- airquality
colnames(airquality1) <- tolower(colnames(airquality1))

# convert column names to lower case with dplyr
library(dplyr)
# option 1
airquality1 <- rename(airquality, ozone = Ozone, solar.rad = Solar.R,
                      wind = Wind, temp = Temp, month = Month, day = Day)
```

```
# option 2 (advanced)
airquality1 < airquality %>%
  # convert to lower case
  rename_all(tolower)
```

```
# Additional example:
# convert dot (.) to underscore (_)
# option 1
colnames(airquality1)[2] <- "solar_r"
# option 2 (advanced)
colnames(airquality1) <- gsub("\\.", "_", colnames(airquality1))

# Additional example with dplyr (advanced):
airquality1 <- airquality %>%
  # convert to lower case
  rename_all(tolower) %>%
  # replace dots (.) with underscores (_)
  rename_all(~gsub("\\.", "_", .))
```

iii) Add a column with the variable `year`

```
# Look at the help file of the dataset airquality
?airquality

airquality_year <- mutate(airquality1, year = 1973)
head(airquality_year)
```

```
##   ozone solar_r wind temp month day year
## 1    41     190  7.4   67     5   1 1973
## 2    36     118  8.0   72     5   2 1973
## 3    12     149 12.6   74     5   3 1973
## 4    18     313 11.5   62     5   4 1973
## 5    NA      NA 14.3   56     5   5 1973
## 6    28      NA 14.9   66     5   6 1973
```

iv) Create a new `date` column in the format (`YYYY-MM-DD`, e.g. `2019-09-25`)

```
# Create the new column as vector
date_vec <- paste(airquality_year$year, airquality_year$month,
                  airquality_year$day, sep = "-")
date_vec[1:6]
```

```
## [1] "1973-5-1" "1973-5-2" "1973-5-3" "1973-5-4" "1973-5-5" "1973-5-6"
```

```
# Add the new column to airquality
airquality_date <- mutate(airquality1, date = date_vec)
class(airquality_date$date)
```

```
## [1] "character"
```

```
# Convert the date from class character to class date
airquality_date <- mutate(airquality_date, date = as.Date(date, format = "%Y-%m-%d"))
class(airquality_date$date)
```

```
## [1] "Date"
```

```
str(airquality_date)
```

```
## 'data.frame':    153 obs. of  7 variables:
##  $ ozone  : int  41 36 12 18 NA 28 23 19 8 NA ...
##  $ solar_r: int  190 118 149 313 NA NA 299 99 19 194 ...
##  $ wind   : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
##  $ temp   : int  67 72 74 62 56 66 65 59 61 69 ...
##  $ month  : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ day    : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ date   : Date, format: "1973-05-01" "1973-05-02" ...
```

v) Try to do `i)-iv)` with `dplyr` and piping (`%>%`)

```
# load original airquality dataset
data(airquality)

airquality2 <- airquality %>%
  # rename columns
  rename(ozone = Ozone, solar_rad = Solar.R, wind = Wind, temp = Temp,
         month = Month, day = Day) %>%
  # add year column
  mutate(year = 1973) %>%
  # add date column
  mutate(date = paste(year, month, day, sep = "-")) %>%
  # change class of date column to date
  mutate(date = as.Date(date))

head(airquality2)
```

```
##   ozone solar_rad wind temp month day year       date
## 1    41       190  7.4   67     5   1 1973 1973-05-01
## 2    36       118  8.0   72     5   2 1973 1973-05-02
## 3    12       149 12.6   74     5   3 1973 1973-05-03
## 4    18       313 11.5   62     5   4 1973 1973-05-04
## 5    NA        NA 14.3   56     5   5 1973 1973-05-05
## 6    28        NA 14.9   66     5   6 1973 1973-05-06
```

```
str(airquality2)
```

```
## 'data.frame':    153 obs. of  8 variables:
##  $ ozone    : int  41 36 12 18 NA 28 23 19 8 NA ...
##  $ solar_rad: int  190 118 149 313 NA NA 299 99 19 194 ...
##  $ wind     : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
##  $ temp     : int  67 72 74 62 56 66 65 59 61 69 ...
##  $ month    : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ day      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ year     : num  1973 1973 1973 1973 1973 ...
##  $ date     : Date, format: "1973-05-01" "1973-05-02" ...
```

2. Convert table from `wide` to `long` format

i) Load the file `tree_growth_data_wide.rds` from the `01_Data` folder and give it a name (e.g. `wide_table`)

```
# maybe you need to change the working directory or the file path
# (remember to include the filename extension '.rds')
wide_table <- readRDS("01_Data/tree_growth_data_wide.rds")
head(wide_table)
```

```
## # A tibble: 6 x 4
##   ts                  dendrometer1_ch3 dendrometer2_ch1 temperature_site_1
##   <dttm>                         <dbl>            <dbl>              <dbl>
## 1 2019-05-31 23:00:00            8336.            2708.               14.2
## 2 2019-05-31 23:10:00            8336.            2706.               14.7
## 3 2019-05-31 23:20:00            8336.            2705.               14.6
## 4 2019-05-31 23:30:00            8336.            2704.               13.8
## 5 2019-05-31 23:40:00            8336.            2703.               13.9
## 6 2019-05-31 23:50:00            8336.            2702.               14.0
```

ii) Install the package `tidyr`

```
install.packages("tidyr")
```

```
library(tidyr) # library needs to be loaded after the installation to be available
```

iii) Convert the table to the format shown below using the function `pivot_longer` from the `tidyr` package

```
long_table <- pivot_longer(data = wide_table, cols = 2:4, names_to = "series",
                           values_to = "value") %>%
  # sort by series
  arrange(series)

head(long_table) # first six rows of a table
```

```
## # A tibble: 6 x 3
##   ts                  series           value
##   <dttm>              <chr>            <dbl>
## 1 2019-05-31 23:00:00 dendrometer1_ch3 8336.
## 2 2019-05-31 23:10:00 dendrometer1_ch3 8336.
## 3 2019-05-31 23:20:00 dendrometer1_ch3 8336.
## 4 2019-05-31 23:30:00 dendrometer1_ch3 8336.
## 5 2019-05-31 23:40:00 dendrometer1_ch3 8336.
## 6 2019-05-31 23:50:00 dendrometer1_ch3 8336.
```

```
tail(long_table) # last six rows of a table
```

```
## # A tibble: 6 x 3
##   ts                  series             value
##   <dttm>              <chr>              <dbl>
```

```
## 1 2019-06-02 15:40:00 temperature_site_1  6.37
## 2 2019-06-02 15:50:00 temperature_site_1  6.29
## 3 2019-06-02 16:00:00 temperature_site_1  6.00
## 4 2019-06-02 16:10:00 temperature_site_1  5.94
## 5 2019-06-02 16:20:00 temperature_site_1  5.81
## 6 2019-06-02 16:30:00 temperature_site_1  5.76
```

iv) Save the table to the `01_Data` folder with a new name (e.g. `long_table`)

```
saveRDS(object = long_table, file = "01_Data/long_table.rds")
```