Reshaping Data: Exercises

These exercises accompany the Reshaping Data tutorial: http://rpubs.com/NateByers/Reshaping. The exercises use data frames from the region5air library. Run the following code to clean out your global environment and load the data you need:

rm(list = ls())  
library(tidyr)  
library(dplyr)  
library(region5air)  
data(airdata)  
data(chicago\_air)

## Exercises

1. The chicago\_air data frame is in a wide format. Use gather() to make a long data frame named chicago\_air\_long.

[Solution 1](#ex1)

1. The airdata data frame is in a long format. Use the filter() function to create a data frame called site22. Filter down to site "840180890022" and a poc of 1 (remember to use ==). Use the select() function to select only the "datetime", "parameter", and "value" columns. Use spread() on site22 to make a wide data frame called site22\_wide with separate columns for each parameter. **Hint:** you want to spread the "parameter" column, so identify that column as the key in the spread() function. The "value" column should be identified as the value in the function.

[Solution 2](#ex2)

1. Use the filter() function on airdata to create a data frame called pm25. Filter down to parameter "88101". Use the select() function to select only the “datetime”, “site”, and “value” columns. Use spread() on pm25 to make a wide data frame called pm25\_wide with separate columns for each site. **Hint:** you want to spread the "site" column, so identify that column as the key in the spread() function.

[Solution 3](#ex3)

#### Advanced Exercises

1. Use ggplot2 to plot the chicago\_air\_long data frame that was created in exercise 1. First make sure to convert the "date" column to a Date class using as.Date(). Use facet\_grid() in the plot to make separate facets for each parameter, and be sure to set the scales to "free".

[Solution 4](#ex4)

## Solutions

#### Solution 1

chicago\_air\_long <- gather(chicago\_air, key = "parameter", value = "value",   
 ozone:solar)

[Back to exercises](#ex)

#### Solution 2

site22 <- filter(airdata, site == "840180890022", poc == 1)  
site22 <- select(site22, datetime, parameter, value)  
site22\_wide <- spread(site22, key = "parameter", value = "value")

[Back to exercises](#ex)

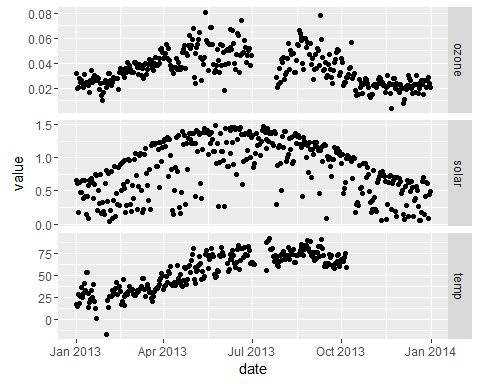
#### Solution 3

pm25 <- filter(airdata, parameter == "88101")  
pm25 <- select(pm25, datetime, site, value)  
pm25\_wide <- spread(pm25, key = "site", value = "value")

[Back to exercises](#ex)

#### Solution 4

library(ggplot2)  
chicago\_air\_long$date <- as.Date(chicago\_air\_long$date)  
ggplot(chicago\_air\_long, aes(date, value)) +   
 geom\_point() + facet\_grid(parameter ~ ., scales = "free")



[Back to exercises](#ex)