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

2. 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

3. 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

Advanced Exercises

4. 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

Solutions

```
chicago_air_long <- gather(chicago_air, key = "parameter", value = "value",
                           ozone:solar)
```

Solution 1 Back to exercises

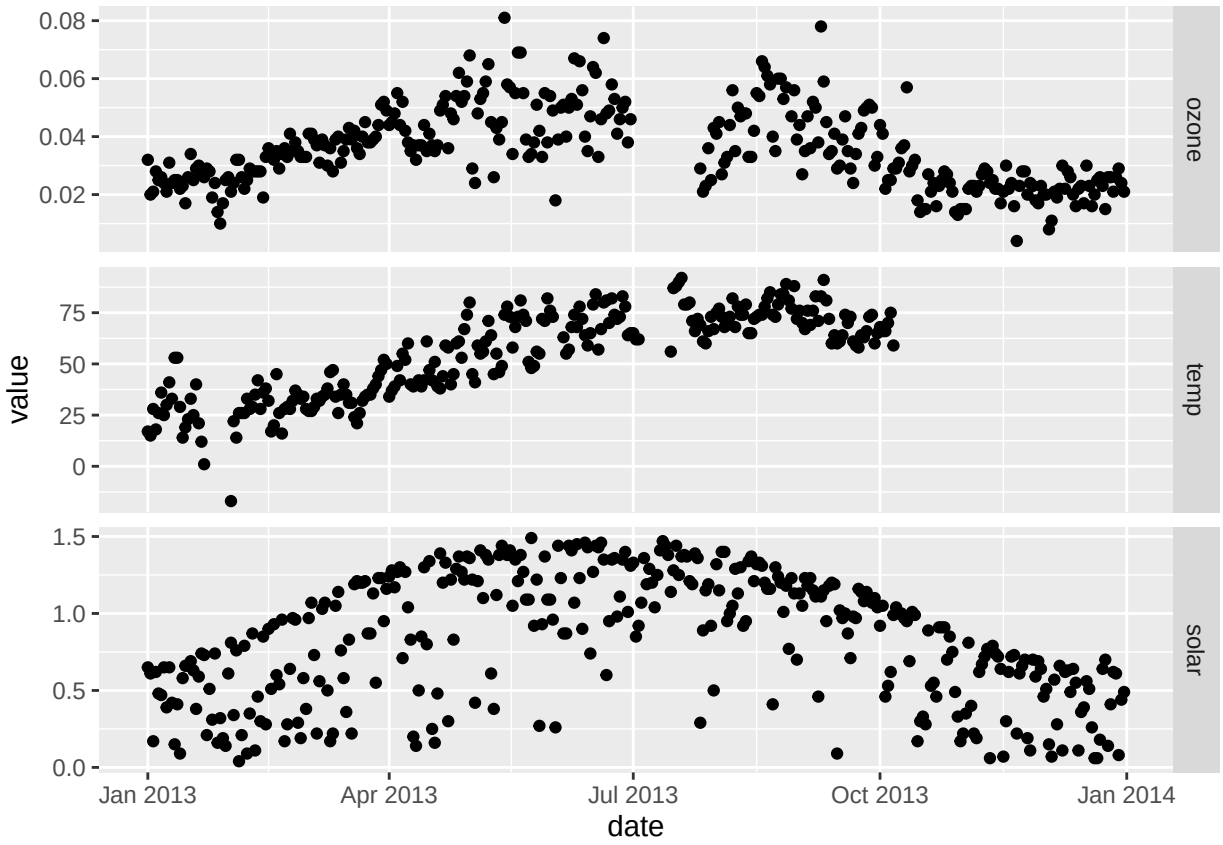
```
site22 <- filter(airdata, site == "840180890022", poc == 1)
site22 <- select(site22, datetime, parameter, value)
site22_wide <- spread(site22, key = "parameter", value = "value")
```

Solution 2 Back to exercises

```
pm25 <- filter(airdata, parameter == "88101")
pm25 <- select(pm25, datetime, site, value)
pm25_wide <- spread(pm25, key = "site", value = "value")
```

Solution 3 Back to exercises

```
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")
```



Solution 4

Back to exercises