Tutorial Exercises Week 4

Read in the dataset: tutorial-data-cleaning.csv. After reading in the "raw" data, the first six rows of the data should look like this:

```
  Sales_Data      Date      Sales Promotion.Sales
1         NA 03.16.18       9657                NA
2         NA 02.08.18       8886                NA
3         NA 04.13.18 Promotion             42312
4         NA 04.14.18 Promotion             35969
5         NA 02.04.18       6500                NA
6         NA 03.24.18       4854                NA
```

The goal of this exercise is to clean this dataset and provide some summary statistics about the cleaned data. When the data is cleaned, the first six rows should look like this:

```
        date sales promotion
1 2018-02-01 22455      TRUE
2 2018-02-02 43011      TRUE
3 2018-02-03  6471     FALSE
4 2018-02-04  6500     FALSE
5 2018-02-05 26509      TRUE
6 2018-02-06  2247     FALSE
```

Complete the following steps to clean the data to get it to look like the 2nd data extract:

- Drop the variable `Sales_Data`.
- Correctly format the "Date" variable as a date.
- Sort the dataset by date. Create a logical variable called `promotion` which is `TRUE` whenever there was a promotion (indicated by a non-NA value in the `Promotion.Sales` variable or the word `Promotion` in the `Sales` variable) and `FALSE` otherwise.
- Whenever the word `"Promotion"` appears in the `Sales` variable, replace it with the corresponding value in `Promotion Sales`.
- Drop the `Promotion.Sales` variable.
- Convert the `Sales` variable from character to numeric.
- Dropping any remaining rows with missing values.
- Convert all variable names to lower case.

Use the techniques discussed in Chapter 13 of the online book to create these data, and use the resulting data to answer the following questions.

Question 1

How many rows are in the final cleaned dataset?

Question 2

On how many days were there promotions?

Question 3

What is the average of the cleaned `sales` variable?

Question 4

What is the average daily sales on days where there were promotions?

Question 5

On which date in April is the median date of the cleaned dataset?