Data cleaning in-class assignment

Download the data MEASLES_Cases_1928-2003.csv from Moodle. This is a particularly dirty dataset that displays the reported measles cases for each US state for each week from 1928 through 2003. Some issues, that you may recognize from this week's homework, are that the columns are values (state names) instead of variables (like "state"), missing values (lots of them), and missing values not coded `NA`, amongst others. Let's get started:

1. Read the data into R. What class are most of the columns of the data? What issues can you find with the data?
2. Change the column names from upper case to lower case.
3. Transform the data into long format from wide format.
4. Re-name the missing measle values as `NA`.
5. Change the values for the measle cases from characters to numbers.
6. Find the largest number of measle cases for a single week in any state. Which state was it?
7. In the step (8) we will want to plot the time series of measles cases. In order to do that, we need the weeks to indicate the time between years. Thus they must be scaled to between 0 and 1, with the first week of the year equaling 0 and the last week of the year coming just before 1.
8. Choose one state and make a line plot with the state's measles cases on the y-axis versus year and time-in-year on the x-axis.

Bonus: 9. Make a table or data frame that shows the number of `NA` values for each state.

If you have extra time, explore the data further! See what other plots you can make! Now that you have replaced the `NA` values and made the values numbers, you could `spread()` the data to more easily compare across provinces.