

Data cleaning hints

1. Use `View()`, `str()`, `summary()`, or other functions to check out the data. If you used `read.csv` with the default settings, most of the columns are **factors**. Re-read them in with the option `stringsAsFactors=F`.
2. Use the function `tolower()` on `colnames(measles)`.
3. Use `gather()` from the `tidyr` package to change the data into long format from wide format. Use `?gather` if you need help using the function.
4. Use the `str_replace()` function from the `stringr` package.
5. Using `as.numeric()` on the measles cases column.
6. Use `max()` to find the maximum measles number. Use `which.max()` to find the row number of the maximum value. Can you use that to find the state name of the maximum value?
7. Find the maximum value of week with `max()`. Then create a new variable for time in year based on `(week-1)/max(week)`.
8. To choose one state from your data frame, use `filter(measles, state=="state_name")` from the `dplyr` package. Use `geom_line()` with `aes(x=year+time_in_year, y=cases)` from the `ggplot2` package to make the plot.
- 9a. Try using the `group_by()` and `summarize()` functions from `dplyr`.
- 9b. Use `sum(is.na())` in your `summarize()` call.