I studied how tidy data is created that makes it easy to manipulate and analyze data. Here, similar to pivot, melt was used among the pandas reshaping functions. The most memorable thing is that among the principles of making tidy data, several variables should be in one column. However, I heard that deciding which one to use as a column and which to use in a dataset is a difficult task because it varies every time depending on the characteristics and problems of the data.

When variables that should be in one column are scattered across several columns, we learned how to use the melt function to make one. id_vars means a reference column, and var_name and value_name give new names to the column to be restructured. The data frame to which the melt function is applied can also be converted into a pivot table. Afterwards, the result visualized by box plot was confirmed with this data frame.

We learned how to split the value by slicing and put it in another column when the previous letter is gender information and the latter number is age information, such as 'm014' in one column. The new data set, IMDB, was able to review and visualize this process once again in exercise.

After lunch, I studied data visualization. Before studying Matplotlib, I learned a new thing about IBCS Standards, a principle for good visualization. It is memorable that a bar chart is better to see and interpret than a pie chart, and charts like spaghetti are not good, which is just a guideline and it is said that it is not necessary to follow. I learned about various visualization functions that begin with plt., and I found that plt.gca() means get current axes, and ax.plot is a little different from plt.plot. In execise, I made a stacked bar chart applying pivot_table and data visualization. It was a good time to reconfirm the knowledge learned today.