First, I use pandas to import excel file into Jupyter Notebook and convert it to DataFrame format. The data consists of 9 tables. CONTENTS is an explanatory document for the remaining 8 tables. Table 1 - Table 6 are the data to be analyzed next. ANNEX and NOTES supplement the field data in these six tables. In the following analysis, I mainly use the data of Table1 - Table6. ANNEX data supplement the above data.

**Principle 1: column names need to be informative, variable names and not values**

One problem with most tables is that there are values in the table column names. They record gender or year in column names, making data processing and analysis difficult. I omit the original column names of the table, read the data in the correct format, and then use “melt” function in Pandas to transform the table in order to expressing gender or year in the data instead of the column names.

Table 1 Raw data

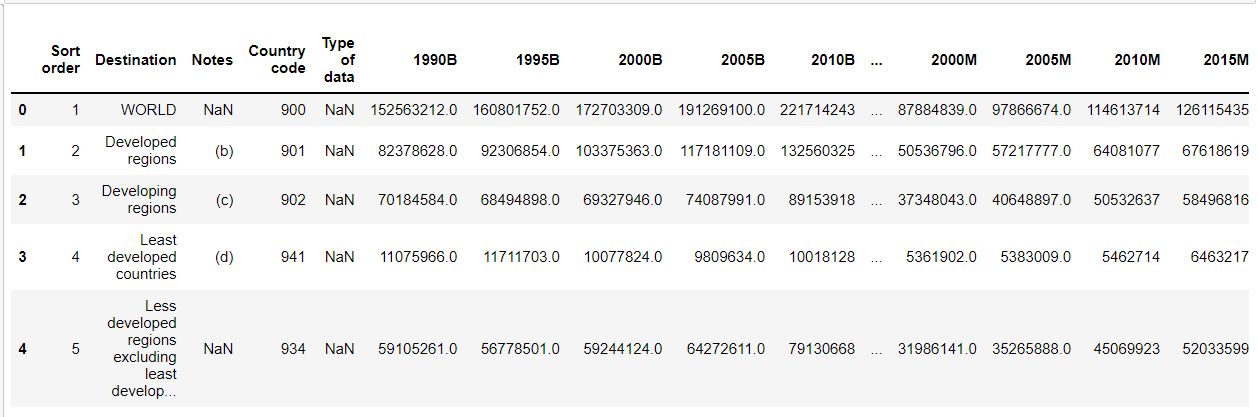
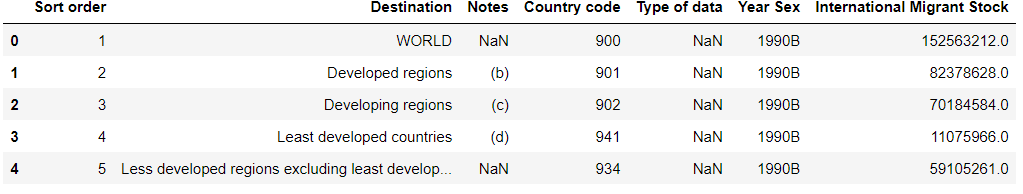


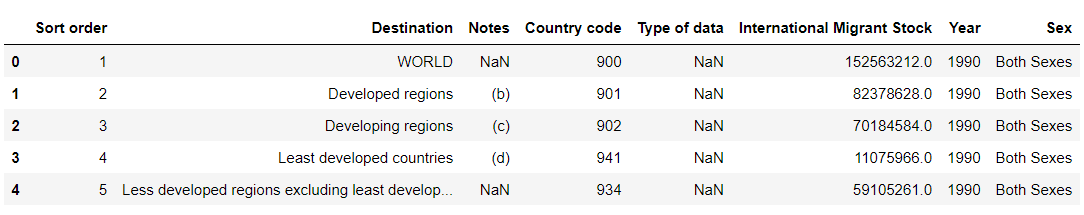
Table1 After “melt”



**Principle 2: each column needs to consist of one and only one variable**

After melting the year values, the column contains a combination of two variables, which is the population of international migrant stock and male and female. To distinguish them, I use the “apply” function in Pandas to split this mixed column into two separate columns.

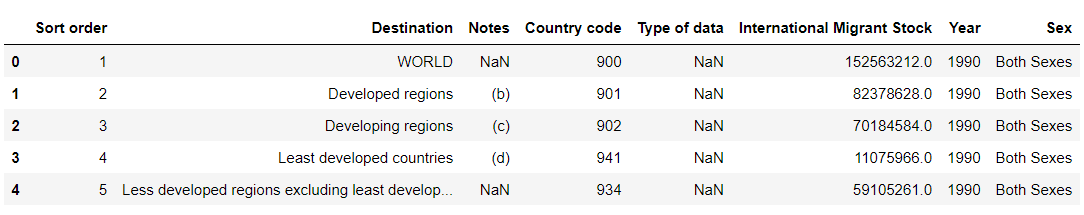
Table1 After “apply”



**Principle 3: variables need to be in cells, not rows and columns**

According to the completed processing, the column names of data no longer contain value information. Through observation, there is no data information in the data row. Therefore, the Principle 3 has been reached. Variables are only in cells.

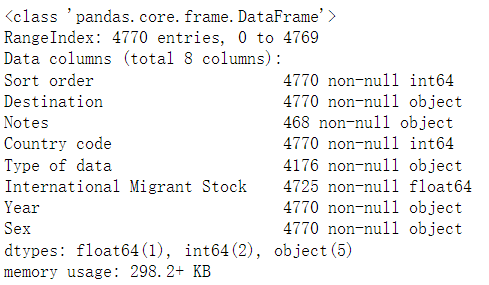
Table1 Values in cells



**Principle 4: each table column needs to have a singular data type**

The original data has missing values and abnormal values, such as the “International Migration Stock” in Tabel1. Normally, this column should be data of float type. But the value '..' in the data is a string. So the data type of this column cannot be unified. I replace the value '..' with np.NAN and unify the format to float type.

Table1 After “replace”



**Principle 5: a single observational units must be in 1 table**

The data in Table 1-Table 6 are in the same column when dividing regions and sub regions. They cannot be distinguished only by the data in the table. So I borrow the data in the ANNEX table. I use the merge function in Pandas to connect data in the “inner” way with the “Country code” as the primary key to obtain more information about the same data.

Table1 After “merge”

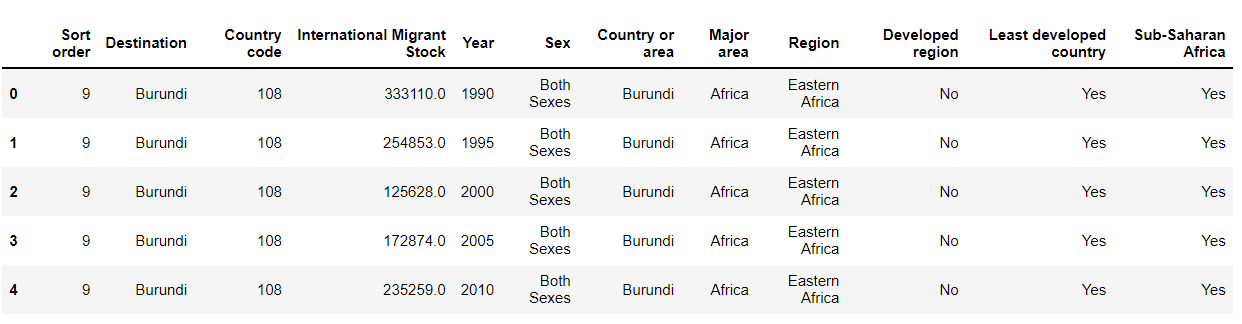


Table 6 is the data related to refugees in the migrant population. I analyze the estimated number of refugees from 1990 to 2015 and the proportion of refugees in the migrant population. According to the histogram, refugees decreased year by year from 1990 to 2005 and increased year by year from 2005 to 2015. There were more refugees in 2015 than in 1990. According to the line chart, the proportion of refugees decreased year by year from 1990 to 2005, remained stable in 2010, and increased in 2015. According to the information in the two charts, the number of refugees has gradually increased in recent years, with only a large number of refugees migrating around 2010. In the rest of the time, the proportion of refugees in the migrant population was relatively stable.

At the same time, I observe the changes in the stock of refugees on different continents. Changes in most continents are consistent with the world. Only Latin America has a large number of refugee flows, which is the main reason for the increase of refugees in the world from 2005 to 2010.

Table 5 is the data related to the annual change rate of population migration. According to the broken line chart, the migration trend of men and women remain the same, with an increasing migration rate from 1990 to 2010 and a slight decline from 2010 to 2015. However, the migration rate of men has been increasing, which was still lower than that of women in 1990, but higher than that of men after 2000. This also shows that men have a higher desire to migrate.

Through the distribution of different continents, I can also see that there are differences between men and women in different continents. In Africa and Europe, the distribution of migration rates for men and women is almost the same. In Asia and Latin America, the distribution of migration rates between men and women is only slightly different. In South America, the mobility variance of men is higher than that of women, which shows that men's life in South America is not stable. In Oceania, the migration change rate of women is higher than that of men. It can be seen that women have more desire to like Oceania more than men.