Hands-on Project

Submission instructions:

- Upload a .Rmd file, and the corresponding HTML file to Moodle
- One person of your team should upload the corresponding files on Moodle
- You should list your team members at the top of your file
- There is no need to show warning and messages for each of the chunk in your file, so you might want to include {r message=FALSE, warning=FALSE} as part of your chunk definition.

Project description:

You have received the "Get_to_know_survey_2022.csv" file, and you need to do some data manipulation in order to show a summary of the results collected in the file.

After loading the data into R, you notice a few inconsistencies in the data before you can summarize it. The following objectives will help you to solve some of those inconsistencies:

1. Create a new column called Birth_Month which has only the month portion of Birth_Month_Year.
2. Create a new column called Birth_Year which has only the year portion of Birth_Month_Year.
3. Since the birth month and birth year were open ended fields, there are many inconsistencies in the data. Standardize the values for Birth_Month and Birth_Year.
4. Show the five most frequent birth months, and the five least frequent birth months.
5. Show the five most frequent birth years, and the five least frequent birth year.
6. The State column was also open ended, so there are many inconsistencies. Clean this column to eliminate any non-US state, and to standardize the values for State.
7. Show the five most frequent states, and the five least frequent state.
8. Similar to State, the column Country needs to be standardized. Once you standardize it, show the five most frequent countries, and the five least frequent countries.
9. We want to clean the How_Hear_About_Program column by eliminating the number and any extra character portion in those responses (so we keep only the text portion). Show the five most frequent responses to this column.
10. We want to transform the data from wide to long on column Languages by duplicating the values of the other columns. You can use the comma as a separator for Languages. Show the five most frequent languages.