Final project

Escola Tècnica Superior d'Enginyeria

UNIVERSITAT ROVIRA I VIRGILI

Student: **Eduard Bel**

Table of Contents

Information about the dataset

For the final project I have chosen the Astronauts[1] dataset from the #tidytuesday challenge. This dataset is a compilation of all the astronauts that have gone to space (in one or more missions) before January 15th of 2020.

The variables of the dataset are:

- Id: represents the row number.
- Number: identificatory of the astronaut.
- nationwide_number: Number within country.
- name: name of the astronaut.
- original_name: name of the astronaut in their original language.
- sex: gender of the astronaut (represented as binary).
- year_of_birth: year when the astronaut was born.
- Nationality: nation where the astronaut was born.
- military_civilian: military status.
- selection: name of the selection program.
- year_of_selection: year the astronaut was selected.
- mission_number: mission number for the astronaut.
- total_number_of_missions: number of missions the astronaut has been in.
- occupation: task of the astronaut on the mission.
- year_of_mission: when did the mission take place.
- mission_title: title of the mission.
- ascend_shuttle: name of the ascend shuttle.
- in_orbit: name of the shuttle orbiting.
- descend_shuttle: name of the descend shuttle.
- hours_mission: hours that the mission lasted.
- total_hrs_sum: Total hours the astronaut has been in mission in their life.
- field21: instances of EVA[2] by mission.
- eva_hrs_mission: hours the astronaut was in EVA on the mission.
- total_eva_hrs: total EVA hours of the astronaut in their life.

[1] https://github.com/rfordatascience/tidytuesday/tree/master/data/2020/2020-07-14

[2] Extra Vehicular Activities

Approach followed and results

The goal when analysing the data and making the plots was answering five different questions I had when I first discovered the dataset. Those are:

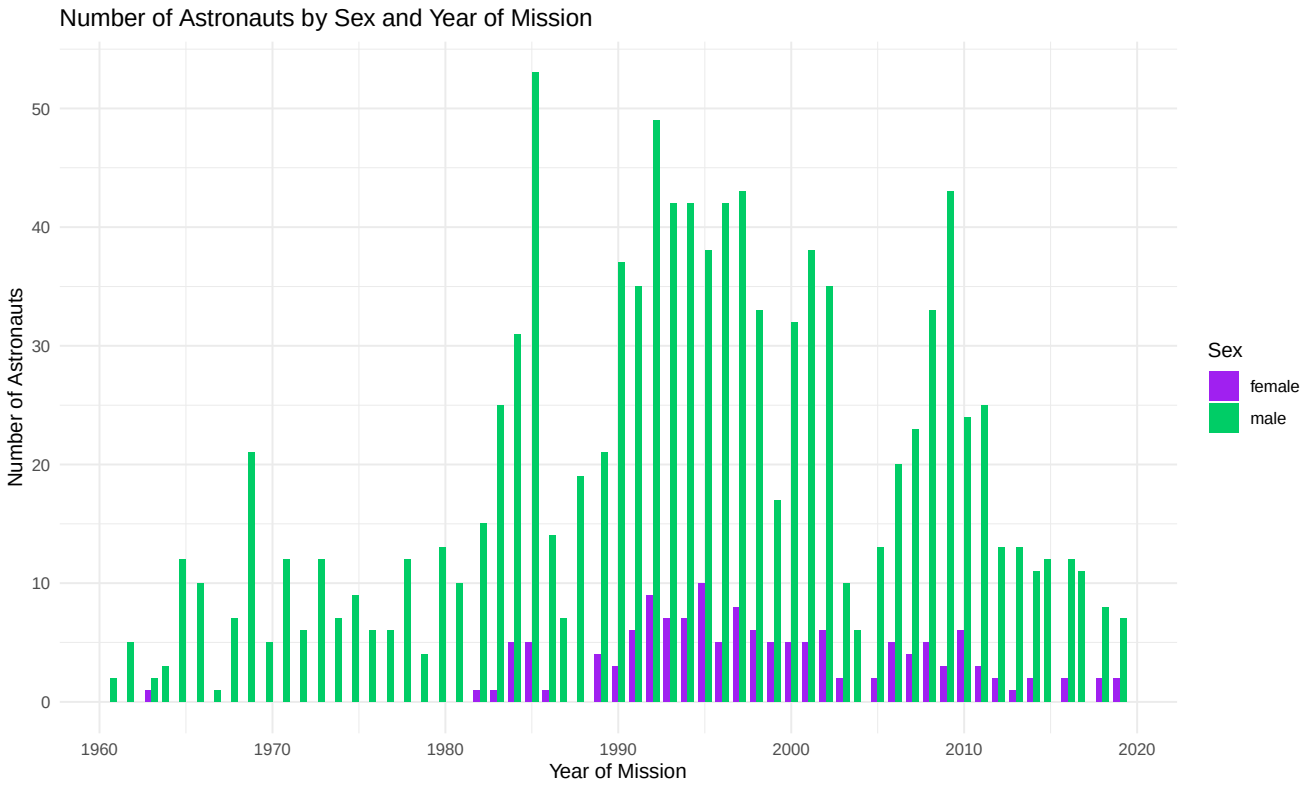
How many astronauts of each nation have been sent to space?

This question can be answered with a bar plot representing the nationalities vs the count of astronauts that have been on a mission (counting them only once).

We can easily see that the countries that have sent more people to space are USA and Russia. This is because they were the two involved countries on the Space Race. Because of that they are also the countries with a more developed space program.

How many men/women have gone to space by years? Has the proportion gotten more similar throughout the years?

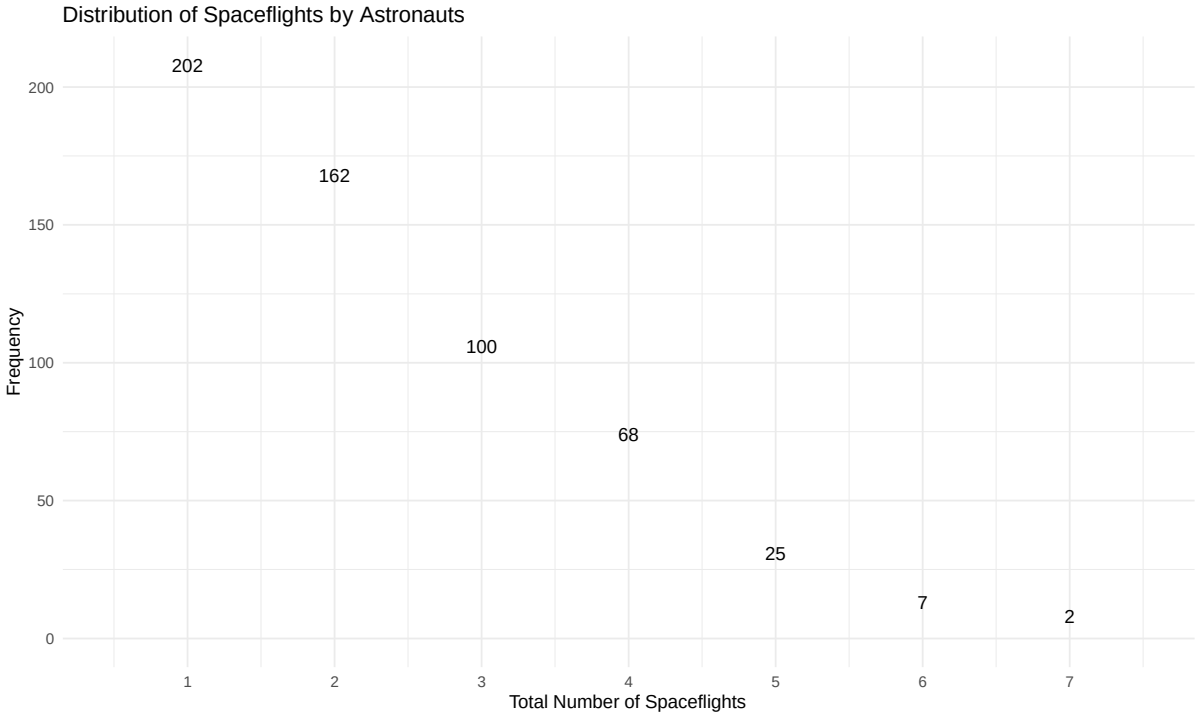
This question can be answered again with a bar plot representing the year of mission vs the number of astronauts sent to space that year. We also add a third dimension with color to distinguish gender.



From the plot we can extract that the first woman (Tereshkova, Valentina) was sent to space in 1963, but it wasn't until the early 80's that women began to go on space missions more frequently. The proportion of women compared to men sent to space has increased since then.
Finally, we can see that the proportion is still not equal at all.

Which is the distribution of spaceflights by astronauts?

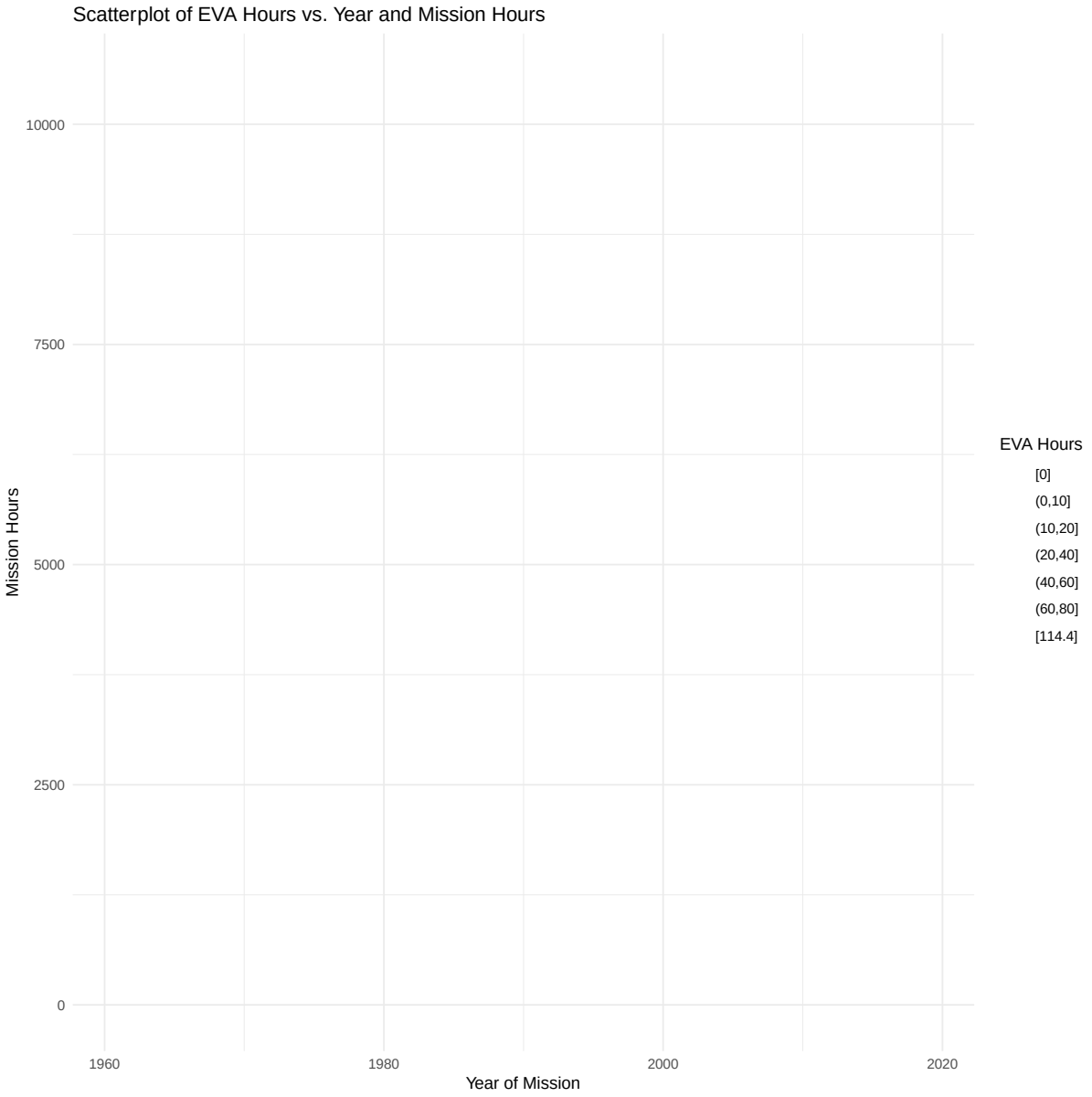
The third question can be answered again by a histogram representing the number of spaceflights of each astronaut vs the frequency.



We can easily see that the most common thing is to go to space only once. As the number of spaceflights increase, the frequency of astronauts doing them decreases. This is because it's a dangerous vocation, besides it's not that well paid. Astronauts get between $70k-120k per year.

Does the EVA (Extra-Vehicular Activity) time have some relation with the duration of the mission? Has the length of mission varied throughout the years?

Both questions can be answered using the same plot, a scatterplot. It represents the year of mission vs the mission hours and we get a third dimension with color representing a binning of the sum of EVA hours spent on that mission.



We can see that until 2010 the most common thing was to do short missions, but since the 70's the mission time has increased by a lot. We can also extract that having a long mission doesn't imply having to do EVA to repair outside machinery or for any other reason.

Implementation

The tools used are those seen on the course. The three libraries used are: ggplot2, dplyr and forcats.

I have found some problems to get the unique astronauts at the beginning.
Then my main problem was with the last plot:

- Aggregating the number of EVA hours for each mission.
- Binning the hours of EVA (I had to use a trick).

In the end I'm happy with how this project turned out, I got to do it about something I like and generated me some curiosity and I have been able to answer my own questions working with the data.