

Use this template to begin typing the first page (summary page) of your electronic report. This template uses a 12-point Times New Roman font. Submit your paper as an Adobe PDF electronic file (e.g. 1111111.pdf), typed in English, with a readable font of at least 12-point type.

Do not include the name of your school, advisor, or team members on this or any page.

Be sure to change the control number and problem choice above.

You may delete these instructions as you begin to type your report here.

Follow us @COMAPMath on X or COMAPCHINAOFFICIAL on Weibo for the most up to date contest information.

1 Data Cleanup

Before beginning any analysis of the data, we had to clean and organize the data provided for this problem.

1.1 Considered Countries

Due to a myriad of historic and geopolitical reasons, countries once prevalent at the Olympics (e.g., the Soviet Union) no longer exist, others have split (e.g., Czechoslovakia), split and then re-merged (e.g., East and West Germany), and yet others have changed their names (e.g., Rhodesia to Zimbabwe). Moreover, there are several multinational teams, such as the mixed team (MIX) in the 1896, 1900, and 1904 Olympics, the West Indies Federation (WIF) in the 1960 Olympics, and the Unified Team (EUN), consisting of multiple former Soviet states, at the 1992 Olympics. To create a working list of states to consider for our analysis, without prejudice to current or past geopolitical events, for our analysis we only consider countries for each unique three-letter country code (NOC) in the `summerOly_athletes.csv` file. For the purposes of this report, we use "country" to refer to those states.

The primary drawbacks of this approach is that it ignores the continuity in the athletic ability of several countries (e.g., Ukraine, Czechia), but such changes to the geographical base from which to draw athletes intrinsically kneecaps any attempt at continuity in these situations. Perhaps, with substantially more time and data, this could be competently addressed, but for the analysis we present we consider them as separate nations.

We will be considering a total of 232 countries, each with a distinct NOC code which we will use for our analysis.

Question, should we include countries that have only competed at one olympics in our analysis?
Question, should we include pre WW1 olympics? pre ww2?