# IPC 144 Project - Milestone 1

**Name**: Ved Viralbhai Makwana

**Student number**: 113366249

**Seneca Username**:vvmakwana

## Project Problems:

1. Get your problems by logging onto matrix
2. Then run the program: ~catherine.leung/getproject

Note that this is completely individualized to you. Your classmates will have a different set. Your profs will check to make sure you are doing the set of problems assigned to you

**Copy and paste the output of getproject here**:

Based on your student number: 113-366-249

Your Problem set is as follows:

A4

B7

C3

Look at the files provided to you in the olympic.zip file (highly recommend opening the files in Excel and answer the following questions. See milestone 1 specs for clarification.

### Problem 1 (Easy level):

a) **What is the problem you are doing for this part (copy the text of the problem, not just the problem number from the specs)**?

Given the edition of the Olympics (the year and either Summer or Winter), and a country, how many athletes did that country send to that edition of the Olympics?

b) **What data file(s) (from the set of provided files consisting of Olympic statistical information) will you need to answer the problem**?

Olympic\_Athlete\_Event\_Results (to determine participation by edition and country)

c) **What column(s) (from within each data file) will you need to solve the problem**?

From Olympic\_Athlete\_Event\_Results:

edition (to filter by the specified Olympic edition, e.g., "2008 Summer Olympics")

country (to filter by the specified country, e.g., "USA")

athlete\_id (to count unique athletes)

d) **Is there any missing/incomplete data for the problem**?

No significant missing data is apparent for this problem:

The Olympic\_Athlete\_Event\_Results file includes edition, country, and athlete\_id, which are sufficient to count unique athletes for a given country in a specific edition.

Some entries have "DNS" (Did Not Start) or "DNF" (Did Not Finish), but these athletes still count as participants if entered in the event.

Potential gaps exist if some athletes’ participation is unrecorded (e.g., historical omissions), but I’d assume the dataset is complete for the editions considered.

e) **Other than the data in the files, what other information will you need to solve the problem**?

The specific edition of the Olympics (e.g., "2008 Summer Olympics") and the country (e.g., "USA") must be provided as input by the user or problem context.

A method to handle duplicates (e.g., if an athlete participates in multiple events, they should be counted once per edition).

f) **For cases where the problems involve multiple files, how do you relate the data between the files** ?

This problem can be solved using only Olympic\_Athlete\_Event\_Results, as it contains all necessary data. No other files are required, so no data relation between files is needed.

### Problem 2 (Intermediate level):

a) **What is the problem you are doing for this part (copy the text of the problem, not just the problem number from the specs)**?

Given the edition of the Olympics (the year and either Summer or Winter), find the number of unique events at the Olympics.

b) **What data file(s) (from the set of provided files consisting of Olympic statistical information) will you need to answer the problem**?

Olympic\_Athlete\_Event\_Results (to list and count unique events)

c) **What column(s) (from within each data file) will you need to solve the problem**?

From Olympic\_Athlete\_Event\_Results:

edition (to filter by the specified Olympic edition, e.g., "2008 Summer Olympics")

event (to identify and count unique events)

d) **Is there any missing/incomplete data for the problem**?

No significant missing data is apparent:

The Olympic\_Athlete\_Event\_Results file includes edition and event, sufficient to count unique events.

If some events were held but not recorded (e.g., due to missing participation), the count might be underreported, but I’d assume the dataset includes all relevant events.

Events like "Long Jump" and "3,000-metre" are distinct in the event column, so I’ll assume it accurately represents unique events.

e) **Other than the data in the files, what other information will you need to solve the problem**?

The specific edition of the Olympics (e.g., "2008 Summer Olympics") must be provided as input by the user or problem context.

A definition of "unique event" (e.g., are "Men’s 100m" and "Women’s 100m" one or two events?). I’ll assume the event column distinguishes them, as seen in the sample data.

f) **For cases where the problems involve multiple files, how do you relate the data between the files** ?

This problem can be solved using only Olympic\_Athlete\_Event\_Results, so no data relation between files is needed.

### Problem 3 (Hard level):

a) **What is the problem you are doing for this part (copy the text of the problem, not just the problem number from the specs)**?

Given the edition of the Olympics (the year and either Summer or Winter), produce a histogram using ascii values of the 10 top ranked countries based on the total number medals won. Break ties with number of golds, then number of silvers, then number of bronze.

b) **What data file(s) (from the set of provided files consisting of Olympic statistical information) will you need to answer the problem**?

Olympic\_Athlete\_Event\_Results (to count medals by country)

noc\_country.csv (to map country codes to full names for the histogram)

c) **What column(s) (from within each data file) will you need to solve the problem**?

From Olympic\_Athlete\_Event\_Results:

edition (to filter by the specified Olympic edition, e.g., "2008 Summer Olympics")

country (to group medals by country)

medal (to count gold, silver, and bronze medals)

isTeamSport (to adjust for team event medal allocation)

From noc\_country.csv:

noc (to match with country from Olympic\_Athlete\_Event\_Results)

country (to display full country names in the histogram)

d) **Is there any missing/incomplete data for the problem**?

Yes, there could be incomplete data:

The medal column includes "-" (no medal), "Silver", etc. Missing medals (e.g., historical gaps) are treated as zero.

For isTeamSport=TRUE (e.g., "Water Polo"), multiple athletes are listed per event, but only one medal is awarded per team. I’d deduplicate medals by event and country to count one medal per team event.

If fewer than 10 countries win medals in an edition, the histogram will include fewer entries, which is acceptable.

e) **Other than the data in the files, what other information will you need to solve the problem**?

The specific edition of the Olympics (e.g., "2008 Summer Olympics") must be provided as input by the user or problem context.

A histogram scaling method (e.g., 1 ASCII character = 5 medals) to ensure readability, which can be defined for implementation in a later milestone.

f) **For cases where the problems involve multiple files, how do you relate the data between the files** ?

Link Olympic\_Athlete\_Event\_Results to noc\_country.csv by matching country (NOC code) with noc to retrieve full country names for the histogram.

Process: Aggregate medals from Olympic\_Athlete\_Event\_Results by country for the specified edition, adjust for team events using isTeamSport, rank the top 10 using total medals (breaking ties with gold, silver, bronze counts), and map to noc\_country.csv for names. (Note: The actual histogram creation will be implemented in a later milestone based on this plan.)