



**MGSC 661 Finals Project**

Date: 5 December 2024

## Introduction

The Olympics stands as the most prestigious global sporting event, providing athletes around the world with an unparalleled platform to demonstrate their skills and compete at the highest level. Since its inception in 1896, the modern Olympic Games have evolved into a grand spectacle. Among the sports that grace the Olympic stage, two stand out for their unique blend of physicality, precision, and artistry: Gymnastics and Alpine Skiing.

Gymnastics, a stunning fusion of artistry and athleticism demands not only extraordinary physical strength and technical precision, but also exceptional grace. Today, its pinnacle is embodied by athletes like Simone Biles, whose performances redefine the limits of human capability. Watching Biles perform has been remarkably captivating, as her routines epitomise the perfect balance of intricate precision and graceful artistry. In contrast, Alpine Skiing offers a thrilling mix of speed, technical skill, and the beauty of mountain landscapes—elements that resonate with my love for the outdoors and adrenaline. As athletes carve their way down steep slopes at astonishing speeds, they demonstrate a mastery of balance and control that mirrors the precision and discipline seen in gymnastics. Both sports, though vastly different in setting, require unparalleled focus, physical conditioning, and a fearless pursuit of perfection.

Thus, this report explores the characteristics of these two groups of athletes, aiming to understand how their physical attributes and performance metrics differ. To do so, I employed Principal Component Analysis (PCA) to uncover underlying patterns within each group. Then, tree-based models were used to classify athletes based on these characteristics and explore how these elements vary between the two disciplines.

## Data Description

### *Data Pre-processing & Feature Engineering*

From the Olympics dataset, Gymnastics (26707 entries) and Alpine Skiing (8829 entries) data were filtered. Given the limited attributes for each athlete – Age, Height, Weight – it was essential to remove all missing values in these columns (31.6% Gymnastics, 28.5% Alpine Ski). Using mean values would mask the diversity in athletes' characteristics. Besides, physical standards of athletes, particularly gymnasts evolved significantly over time. A brief look into the unique team values revealed they were not representative of the countries athletes were from. Thus, the additional NOC dataset from the same Kaggle source along with online resources were used to map NOC codes to the respective countries. Missing values in the "Medal" column were also filled with "None".

Gymnastics events with fewer than 10 participations were recategorised to ensure more meaningful groupings. The Men's Rope Climbing and Women's Portable Apparatus events (0.3% data) were removed because they did not align with the artistic dimension of gymnastics; they were originally included when gymnastics had not yet fully developed its current identity.

To better capture the athletes' experience levels and performance, five new features were created. First, Years of Experience was calculated by subtracting the year of an athlete's first Olympic participation from their most recent entry, providing an estimate of their experience levels. The second feature – Number of Participations, counts the total number of Olympic events an athlete has competed in, reflecting their level of involvement. Next, Participation Gap was calculated as the time difference between an athlete's last two entries, offering insight into consistency and stability of their performances. Previous Wins was also created to quantify the number of medals an athlete had won prior to a given year. This variable serves as a proxy for prior success, allowing

for analysis of its influence on future performance. Lastly, Body Mass Index (BMI) was also computed as a composite of height and weight to account for physical attributes.

### ***Trends in Athlete Participation: Shifts in Gymnastics and Alpine Skiing Events***

Despite having the second-highest number of overall participations (Fig A1), Gymnastics has experienced a surprising decline in participation over the decades (Fig A2a). This decline likely stems from the sport's increasing emphasis on specialisation and elite performance, limiting the number of athletes entering the sport. In contrast, Alpine Skiing has seen a consistent rise in participation (Fig A2b). Despite the decline in gymnastics participation, gymnasts still compete in more events on average compared to alpine skiers. Gymnasts often compete in multiple disciplines such as beam, bars, and floor, which require a diverse skill set, while alpine skiers specialise in fewer events like slalom or downhill, each emphasising specific aspects of speed and technical precision. This underscores gymnastics' multifaceted nature, where athletes demonstrate versatility, while specialisation in distinct skill areas is key in alpine skiing.

Historically, both sports have been male dominated, especially in the earlier years (Fig A3a, b). However, a noticeable shift has occurred in recent years in gymnastics; the gender gap has narrowed, particularly in events like the floor routine, which now dominates the sport (Fig A4a). This shift reflects the growing inclusion of women in gymnastics, in contrast to alpine skiing, where the male participation remains significantly higher (Fig A4b).

### ***Rise of China and Russia in the Gymnastics Scene***

In gymnastics, Russia, the USA, Romania, China, and Japan have firmly established themselves as dominant forces (Fig A5a). Notably, both Russia and China entered the global gymnastics scene later than other leading nations. Russia, following the Soviet Union's dissolution, quickly cemented its legacy, while China's rapid ascent within a relatively short period is particularly commendable (Fig A6a). Further analysis of medal distributions by gender revealed unique trends. As expected, both male and female gymnasts from Russia excelled in total and average medal counts (Fig 1). In the USA, female gymnasts significantly outshone their male counterparts, reflecting a strong emphasis on women's gymnastics. Conversely, Japan displayed the opposite trend, with male gymnasts achieving far greater success than their female peers.

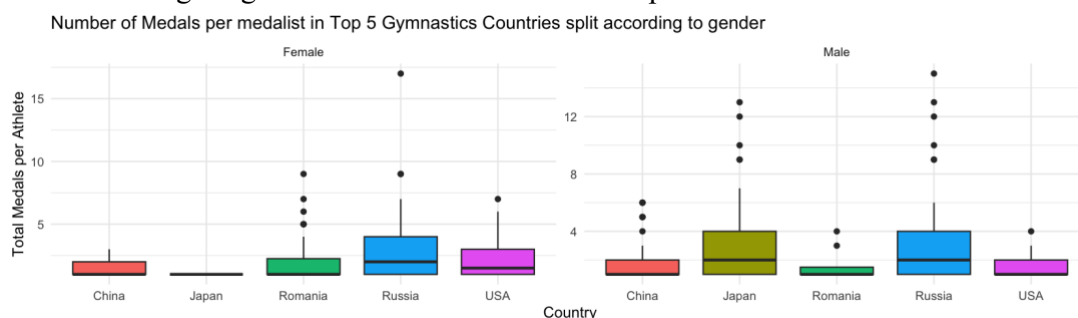


Fig 1: Medals obtained per medalist in top 5 performing countries in gymnastics

### ***The Allure of Alpine Skiing in Mountainous Regions***

Austria, Switzerland, France, Germany, and the USA emerged as the top-performing countries in Alpine Skiing (Fig A5b), a trend closely tied to their mountainous terrains, which provide ideal conditions for training and competition. Austria has maintained its position as the dominant force since 1990s, consistently outshining other nations in medal counts. Meanwhile, the USA has shown remarkable progress in recent years, with increasing success on the global stage (Fig A6b). Though male participation in alpine ski was significantly higher than in gymnastics (Fig A3b), female skiers



outperformed their male counterparts across the top five performing countries (Fig 2). This contrasts with gymnastics, where the performance gap between genders is more pronounced, and top-performing female gymnasts were typically concentrated within a few elite athletes (Fig 1).

Number of medals per medalist in Top 5 Alpine Ski Countries according to gender

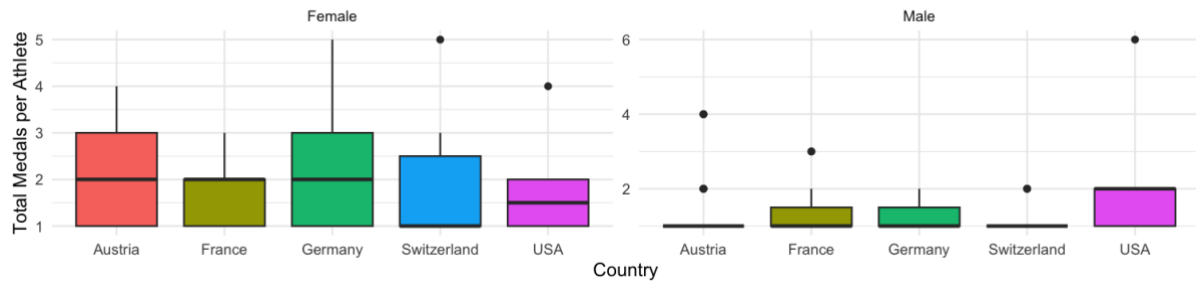


Fig 2: Medals obtained per medalist in top 5 performing countries in alpine ski

### ***Evolution of Medalists' attributes***

Historically, female gymnasts were much younger than their male counterparts, with many top performers in their teens during the 80s and 90s (Fig A7a). This was due to the sport's emphasis on flexibility and a smaller frame. However, the trend is slowly shifting with recent athletes like Simone Biles, who continues to perform at a high level into her late 20s, highlighting a more mature approach to training. In contrast, skiers often peak later, as alpine ski requires years of experience to master the technical aspects of skiing such as speed control and risk management (Fig A7b).

Physically, gymnasts are generally smaller and lighter than skiers, due to the sport's prioritisation on agility and body control (Fig A7a). Skiers, in contrast, have a more robust build (Fig A7b) to handle the physical demands of high-speed descents, extreme weather and slalom courses which necessitate higher muscle mass and greater stature for stability and power. However, in recent years, more female gymnasts are taller and heavier, likely due to evolving beauty standards and an increasing appreciation of muscularity. This reflects a broader trend in sports where strength and power are increasingly valued.

### **Model Selection & Methodology**

To prepare the data for modelling, univariate variables and identifiers were removed. Thus, the eventual gymnastic and ski datasets each comprised of 8 numerical variables, and 3 categorical variables. A correlation matrix was executed to identify any highly correlated numerical variables. As expected, weight and height, as well as BMI and weight showed significant correlations ( $>0.8$ ). Additionally, years of experience was strongly correlated with both participation gap and total participations ( $\sim 0.8$ ). Despite these correlations, the variables were retained for PCA to assess their combined influence and determine the most informative features.

For the PCA model, a separate variable “Win” was also engineered, which was coded as 1 for athletes who won a medal, and 0 for those who did not. Initially, separate variables for “Gold”, “Silver” and “Bronze” medals were also created to capture distinctions among medal types. However, their relationships with the remaining numerical variables were found to align closely to the consolidated “Win” variable. As a result, “Win” was retained to simply the PCA analysis while effectively representing medal-winning performance. This consolidated feature provides a focused lens for analysing how other numerical variables correlate with medal success. Events and countries were not dummified and were excluded from PCA to avoid the drastic increment in number of variables, which would lead to overcomplication and dilute meaningful insights from the model.

With the selected features from PCA, regression trees and random forest models were constructed for both groups of athletes. Initially, overfitted trees were grown to explore classification differences across four distinct subsets of athletes in each discipline: (i) Athletes who won versus those who did not in medal-winning countries (ii) All medalists (iii) Athletes who won versus those who did not in the top five performing countries, and (iv) All medalists from the top five performing countries. For each subset, the medal, event and country categories were dummified to allow for classification across the different outcomes. To refine these trees, cross-validation was employed to determine the optimal complexity parameter (cp), ensuring that the models avoided overfitting while retaining interpretability. Feature importance was then derived from the pruned trees, providing insights into the predictors most relevant to performance outcomes within each subset.

Interestingly, the variable 'Events' did not receive a variance importance score in any regression tree, thus excluded from further analysis. This likely reflects minimal variability in this feature across athletes, making it less relevant for predicting performance outcomes. To complement and validate the regression tree results, random forests were applied to the same datasets. The use of random forests, with their robustness against overfitting, enhanced the reliability of the feature importance rankings derived from the regression trees, ensuring that the most relevant predictors were consistently identified across both modelling techniques.

## Result & Discussion

### *PCA Comparisons between Gymnasts and Alpine Skiers*

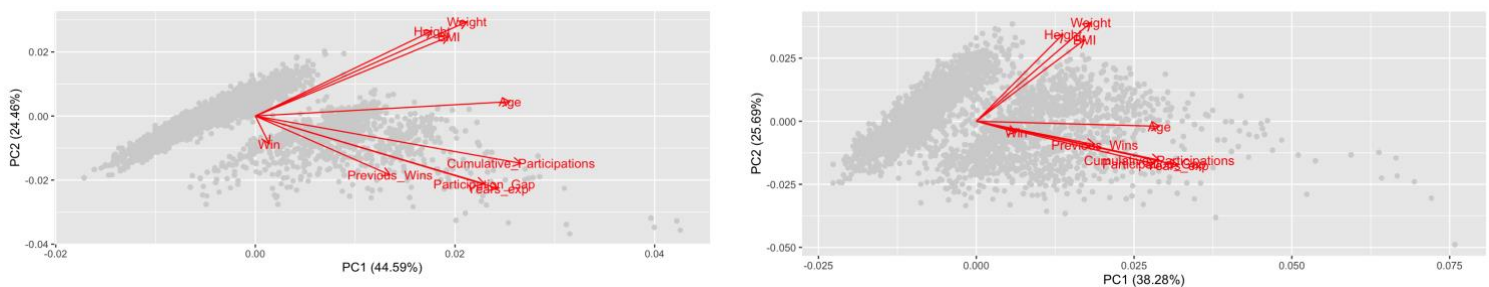


Fig 2: PCA of numerical variables for gymnasts (left, 2a) and alpine skiers (right, 2b)

Based on PCA biplots in Fig 2, BMI, Height, and Weight are correlated, as reflected in the correlation matrix. However, BMI does not fully capture body composition, particularly for gymnasts, where lean builds are more advantageous. Given the higher loadings of height and weight, height and weight were retained in favour of BMI. Interestingly, these physical traits are almost orthogonal to the "Win" vector for gymnastics data. This indicates that while physical traits such as being smaller and lighter are beneficial in gymnastics for ease of movement and control, performance outcomes may be influenced more by skill, technique, and mental factors than simple physical characteristics. Though these 2 components are not entirely orthogonal in the alpine ski data, physical traits are also likely to have minimal influence on performance outcomes.

Among the engineered features, there is an interesting relationship with the "Win" vector. For alpine skiers, these features are nearly aligned along the same line, indicating that experience and participation closely correlate with performance. The alignment suggests that experienced athletes with more participations and previous wins are more likely to succeed, highlighting that continued involvement and proven track records contribute significantly to success in alpine skiing. For gymnasts, however, the relationship differs slightly. Previous wins and years of experience are most

closely aligned with the "Win" vector. This implies that while participation is important, past victories and accumulated experience are stronger predictors of success.

It can also be observed that age is more closely aligned with success in alpine skiing than in gymnastics. This suggests that experience and maturity are key factors in a skier's performance. Skiing, as a sport that often values endurance, technical refinement, and mental strategy, benefits from the accumulated skills that come with age. Conversely, age might have a less direct impact on gymnasts' performances and success compared to other factors, which is particularly interesting given gymnastics has favoured younger athletes, with peak performance typically thought to occur in their teens. However, this finding, along with recent trends such as the continued success of athletes like Simone Biles, and most notably the podium finishes in the recent Women's All-Round event at Paris Olympics 2024 (2 of 3 athletes were 25 years old and above), demonstrate that older gymnasts can still achieve peak performances. This shift in trends reflects how evolving training, better athlete-coach partnerships, and a greater understanding of physical and mental resilience allow athletes to extend their careers well beyond the previously expected age limits.

### ***Regression Tree Results***

Among the regression trees grown for the four subsets of each discipline, only subsets 2 and 4 successfully produced meaningful splits and had their cp tuned. Subsets 1 and 3, which include athletes who did not win any medals, had a significantly larger proportion of non-medalists compared to medalists. As a result, the distribution across the three medal types showed even less variability. With little to no variation in the target variable, the regression tree algorithm was unable to identify meaningful splits, resulting in failed tree growth (Table A1). Comparing the regression tree models for subsets 2 and 4 of gymnasts and alpine skiers, there were rather significant and interesting difference between the two disciplines (Table 1, Figs A8).

	Height	Country	Weight	Age	Years of experience	Sex	Previous Wins
Gymnastics	17	40	19	8	4	6	5
Alpine Ski	22	22	19	14	10	8	5

Table 1: Variance importance for subset 2 (all medalists) of both sports

Variance importance for medalists subsets revealed that height plays a markedly different role in the two sports. For gymnastics, it has a relatively low importance, whereas it is more critical in alpine skiing, reflecting the distinct physical demands of the sport. Country stands out as the most important feature for gymnastics, underscoring the influence of nation-specific training programmes, facilities and historical dominance in the sport. In alpine skiing, while the country of origin remains significant, its relatively lower importance is surprising, especially given that countries with mountainous regions typically have better access to specialised training grounds and facilities. These natural advantages, such as consistent snow conditions and well-developed ski runs, often provide athletes with more opportunities for early skill development and year-round training. Experience is also substantially more critical for alpine skiing than for gymnastics. This aligns with skiing's technical complexity and reliance on accumulated knowledge of courses and conditions. Gymnastics, by contrast, prioritises physical aptitude and training intensity over sheer duration of experience.

	Height	Country	Weight	Age	Years of experience	Sex	Previous Wins
--	--------	---------	--------	-----	---------------------	-----	---------------

Gymnastics	25	14	26	17	5	5	7
Alpine Ski	31	11	18	17	11	5	6

Table 2: Variance importance for subset 4 (medalists from top 5 countries) of both sports

The regression tree models for medalists from the top five countries reveal intriguing shifts in the factors that influence success in gymnastics and alpine skiing (Table 2, Figs A9). For gymnastics, the importance of country diminishes significantly compared to the overall medalist dataset, suggesting that athletes from top-performing nations share a level playing field regarding access to resources and training environments. Instead, physical attributes like height and weight dominate, with both factors becoming key predictors of success. This shift highlights the role of body composition in achieving peak performances in gymnastics, especially among elite athletes from established gymnastics countries. These countries, known for their strong traditions and success in the sport, often have well-developed infrastructures, including advanced training facilities and access to expert coaching. As a result, when resources and opportunities are relatively uniform at this level, the focus shifts toward refining individual physical attributes, such as body composition, to gain a competitive edge in performance. In alpine skiing, years of experience continues to remain a dominant factor, underscoring its significance in mastering the technical nuances of the sport. Interestingly, the importance of height becomes even more pronounced, further emphasising its advantage in providing stability and speed on the slopes, which are critical for performance at the elite level. The role of country in skiing is less significant, mirroring the observation that skiing excellence is less centralised compared to gymnastics.

### ***Random Forest Results***

The performance of random forest models varied significantly across subsets, highlighting key limitations. For subsets 1 and 3, which largely comprised of non-medalists, the models excelled at predicting the majority class, with error rates below 20%. However, they struggled at classifying medalists, with class error rates above 90%. In contrast, for subsets 2 and 4, which focus exclusively on medalists, the models had higher error rates exceeding 50% for both disciplines. This indicates that existing predictors are insufficient at capturing medal outcomes, underscoring the complexity of performance in these sports, which are likely influenced by factors beyond dataset.

For gymnasts, medal outcomes are often driven by factors such as technical precision, execution scores, and artistry, which are assessed subjectively by judges and are not directly reflected in the physical or demographic variables. Training regimens, access to elite coaching, and the quality of facilities play a significant role in an athlete's performance. Furthermore, psychological factors like mental resilience, which is crucial for executing high-stakes routines under pressure, were not accounted for either. For alpine skiers, factors like environmental adaptability, quick decision-making, and technical proficiency are critical for medal success. While experience is a critical factor, it is insufficient on its own. Training intensity, access to advanced ski equipment, and support from national teams or sponsors are pivotal in shaping competitive outcomes, while external factors like weather conditions and course familiarity can significantly affect medal results but are not part of the current feature set. Though the data may fall short in capturing the intricate, sport-specific factors that ultimately determine medal success, the predictors still offer a foundational view of athletic performance.

First, height and weight emerged as significant predictors for both gold and silver medals in both sports for athletes that successfully won medals at their respective sports, as indicated by the high

MeanDecreaseGini scores (Tables A2, A3). However, the variance importance from the regression tree was much lower (Table 1), suggesting that regression tree does not effectively capture the influence of these physical attributes on medal outcomes. Second, age was found to have strong predictive power, particularly for gold medals in gymnastics. However, the variance importance for age in the regression tree was significantly lower than expected, indicating that the regression tree underestimates the impact of age. For medalists from top five performing countries, random forest model reinforced the idea that the main factors influencing medal outcomes were physical attributes. However, one key aspect that the regression trees failed to effectively capture was the importance of years of experience (Table A4, A5). Interestingly, the random forest model revealed that weight had a more significant influence than what was captured in the regression tree for alpine skier from the top performing countries.

This comparison shows that random forest model better captures the complexities of feature importance, particularly for physical attributes and years of experience in athletes from top-performing countries, which the regression tree model overlooks or underestimates.

### **Final thoughts**

Overall, the dataset provides valuable insights into athletes' physical attributes and performance history, but falls short of capturing the full range of factors contributing to success. Crucial elements such as psychological resilience, training environments, and injury histories remain underrepresented, limiting the depth and applicability of the analysis. A key finding is the significant influence of country of origin on athletic success. Nations that prioritise and invest in sports, like Russia with its emphasis on gymnastics excellence, demonstrate how robust infrastructures, advanced training facilities, and experienced coaching can elevate athletes to global prominence. Russia's decades-long gymnastics dominance has been exemplified by legends such as Svetlana Khorkina and Larisa Latynina, who benefited from these investments. Similarly, the emergence of nations like Brazil in gymnastics highlights how strategic investment can transform countries into competitive contenders. Rebeca Andrade's historic achievements, including her Olympic successes (who won 2<sup>nd</sup> to Simone Biles in the All-Around and Vault finals in Paris 2024 Olympics), underscore how targeted efforts in training facilities and talent development can break traditional dominance, paving the way for broader global representation.

It is also encouraging to see that the gender gap in sports is narrowing, reflecting progress toward equity in access to resources and opportunities for female athletes. However, there is still room for improvement. Promoting female participation further and ensuring equal support can enhance diversity in competition and unlock the full potential of athletes worldwide. Beyond individual success, nations stand to gain broader cultural and economic benefits by investing in sports, from fostering national pride to inspiring future generations and enhancing their global reputation. By addressing these disparities and incorporating more holistic data, we can not only refine performance predictions but also foster the continued evolution of global athletic development strategies.



## Appendix

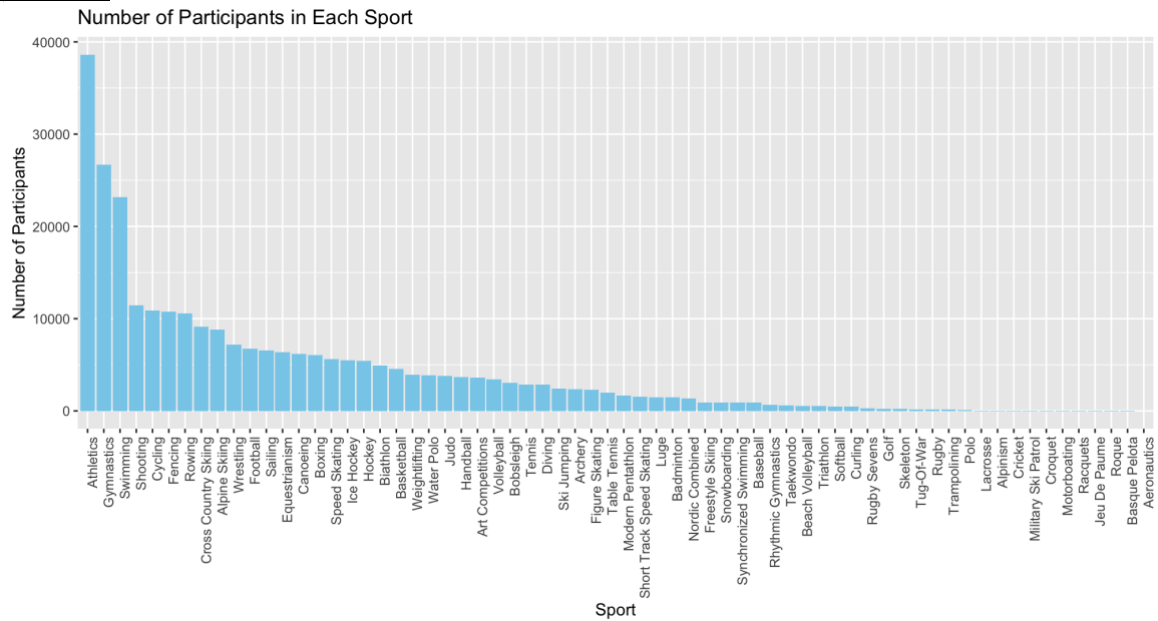


Fig A1: Total number of participations for each Olympic sport from 1896 to 2016 (descending order), with Gymnastics at 2<sup>nd</sup> overall and Alpine Skiing as 2<sup>nd</sup> highest among winter sports

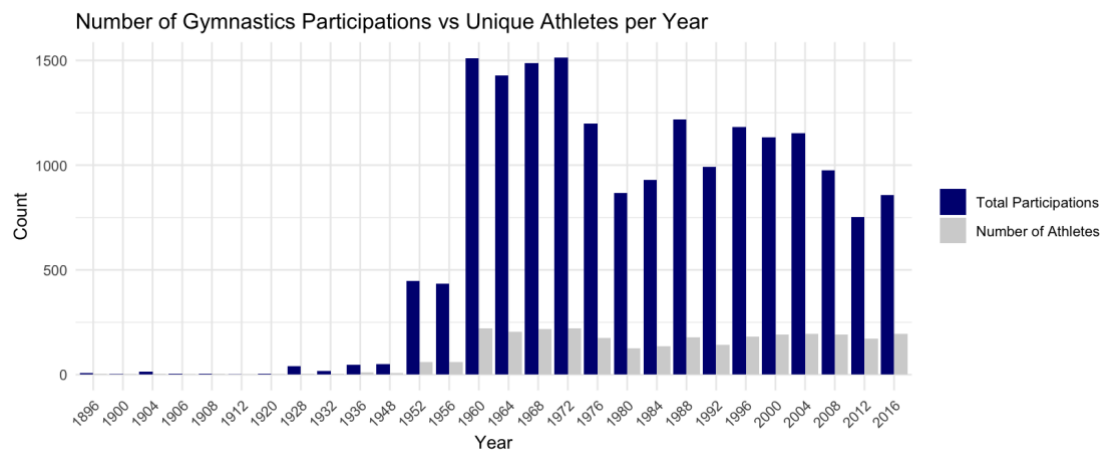


Fig A2a: Number of participations vs unique number of gymnasts over the years

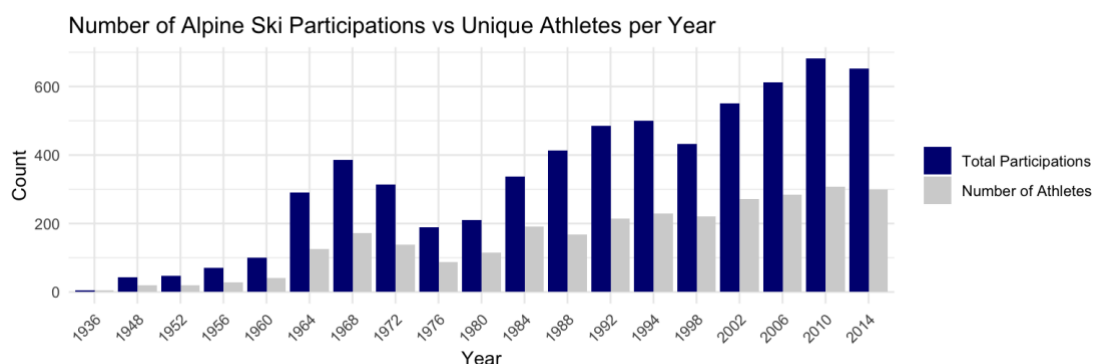


Fig A2b: Number of participations vs unique number of Alpine Skiers over the years

Comparing Figs A2a and A2b reveals intriguing trends in athlete participation across events in gymnastics and alpine skiing over the last century. The total number of participations in events far exceeds the number of athletes registered, reflecting the multi-event nature of these disciplines. Notably, the ratio of participations to athletes is significantly higher in gymnastics than in alpine skiing, underscoring the broader range of events that gymnasts typically engage in. However, an

interesting shift emerges over time. In gymnastics, the average number of events per athlete appears to have declined compared to the 1960s and 1970s. This could suggest an increased emphasis on specialisation or changes in event formats. Conversely, the average for alpine skiers has risen, possibly due to greater emphasis on versatility in skiing disciplines or evolving competition structures.

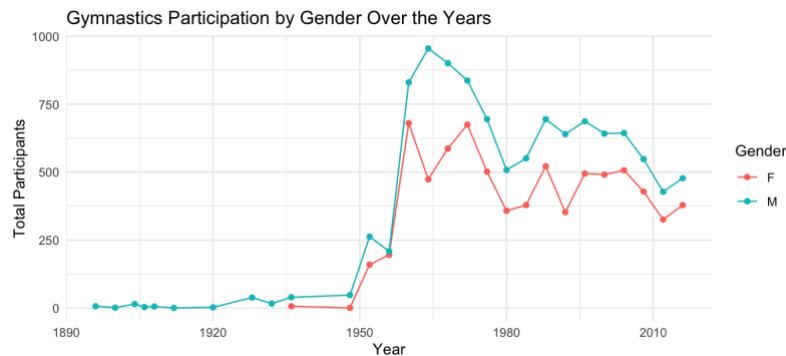


Fig A3a: Number of participations in Gymnastics over the years split according to Gender



Fig A3b: Number of participations in Alpine Ski over the years split according to Gender

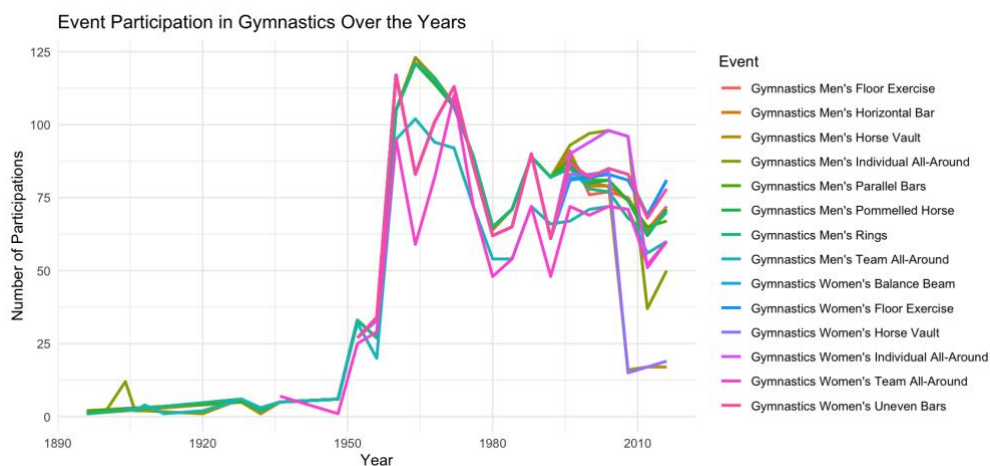


Fig A4a: Number of participations in each gymnastics event over the century

Notably, vault events saw significant decline in recent years, possibly accounting for the drop in number of participations across gymnastics events. Floor routines, on the other hand, are becoming the dominant events for gymnastics, particularly among female athletes, who participate in these events at higher rates than their male counterparts. This disparity may reflect the unique nature of floor routines in women's gymnastics, which often emphasise artistic expression, choreography, and music in addition to technical skills. The greater female participation could also suggest that floor routines serve as a platform for gymnasts to showcase a broader range of talents, making them a more appealing event for female athletes.

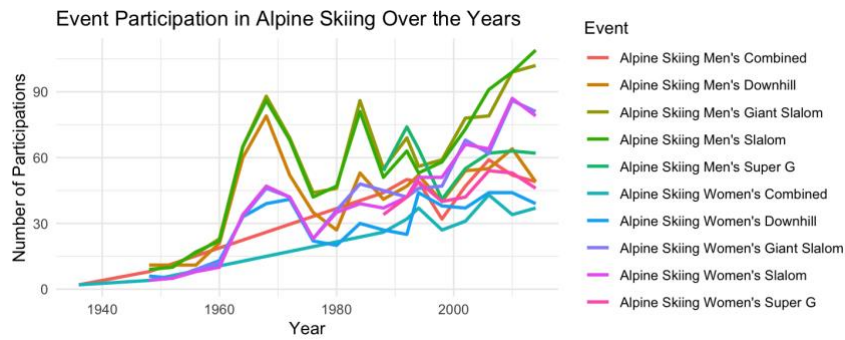


Fig A4b: Number of participations in each alpine ski event over the century

Though alpine ski events for both women and men saw an overall increment over the decades, the number of male participants saw a larger exponential increase in recent years.

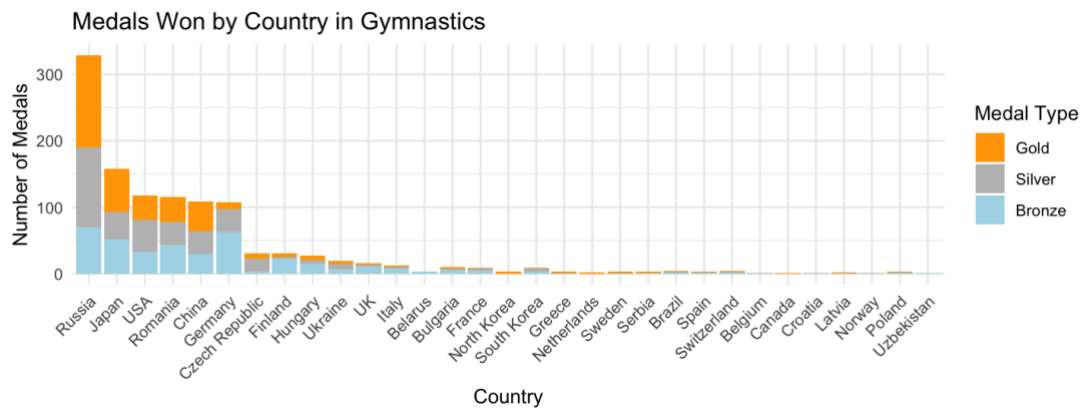


Fig A5a: Number of gymnastics medals won by each country over time

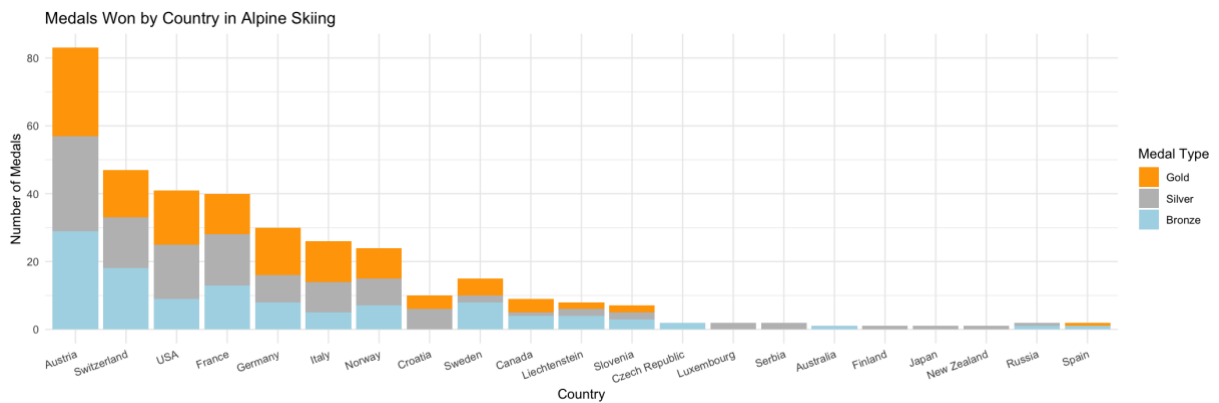


Fig A5b: Number of alpine ski medals won by each country over time

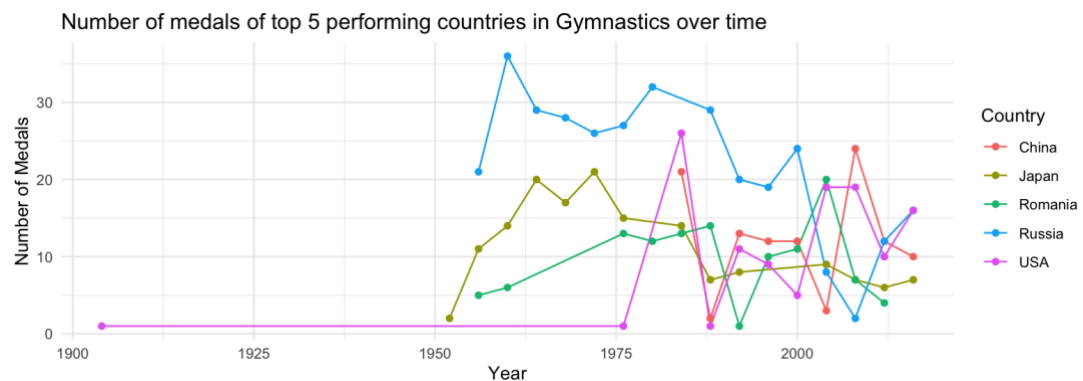


Fig A6a: Number of medals won by the 5 powerhouses in gymnastics over time

China's relatively late entry into international gymnastics competitions compared to other top-performing nations is noteworthy, yet its rapid ascent to elite status is commendable. Over the last few decades, China has established itself as a formidable force, leveraging its structured athlete development programs, disciplined training regimens, and a cultural emphasis on technical excellence. This remarkable progress highlights China's commitment to investing in sports infrastructure and cultivating a strong talent pipeline to compete at the highest levels. In contrast, Russia, which inherited the gymnastics legacy of the Soviet Union, experienced a period of dominance throughout much of the 20th century. The Soviet Union's centralised training system and state-sponsored athlete development programs positioned it as the global leader in gymnastics for decades. However, following the dissolution of the Soviet Union in 1991, Russia faced significant challenges. The loss of a unified infrastructure, economic instability, and the migration of top coaches to other countries undermined its ability to sustain this dominance. Despite these challenges, Russia has had moments of resurgence, such as the success of athletes like Aliya Mustafina, who earned numerous medals in the 2010s. However, the country has not returned to the level of dominance it once enjoyed during the Soviet era.

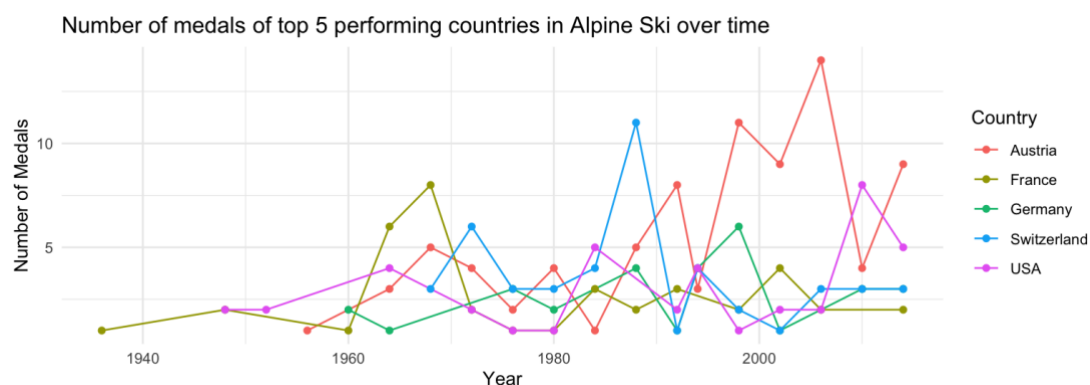


Fig A6b: Number of medals won by top performing countries for alpine ski over time

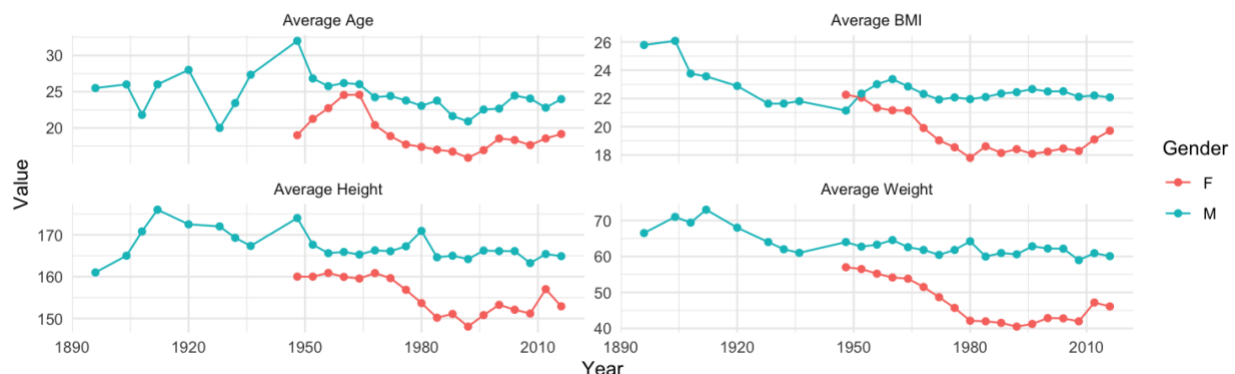


Fig A7a: Physical attributes of medalists in gymnastics over time

Between the 1960s and 1990s, the average age of gymnasts competing at the Olympic level experienced a noticeable decline. This period was marked by the dominance of younger gymnasts, particularly teenage athletes. The shift was driven by changes in the sport's scoring system and a greater emphasis on routines requiring extreme flexibility, agility, and a lower centre of gravity—

attributes often more pronounced in younger athletes. This trend was epitomised by the successes of gymnasts like Nadia Comaneci, who achieved historic milestones as a teenager. In recent years, however, the average age of gymnasts has been on the rise again. Several factors contribute to this reversal, including advancements in sports science, injury prevention, and training techniques, which have extended the careers of elite gymnasts. Moreover, the evolving Code of Points has shifted to reward execution, artistry, and technical mastery alongside difficulty, allowing more experienced athletes to remain competitive. The same trends are evident in gymnasts' physiques. During the mid-20th century, a shift toward leaner, smaller frames became prevalent, aligning with the demands of the sport's dynamic routines. Today, there is a slight move toward more robust and muscular physiques, reflecting the increased emphasis on strength and power needed for the high-difficulty tumbling and vaulting elements in modern gymnastics.

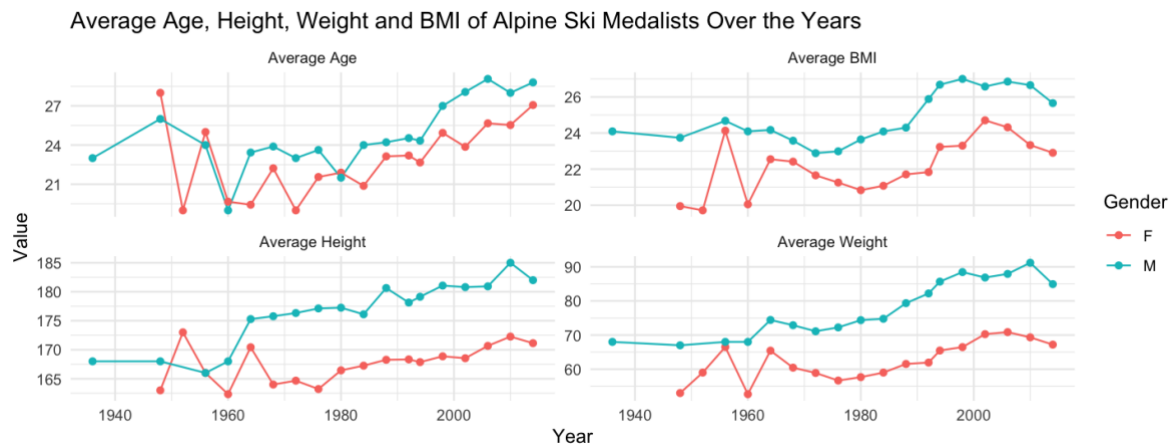


Fig A7b: Physical attributes of medalists in alpine ski over time

In contrast to gymnastics, alpine skiing exhibits a consistent trend of rising ages and physical attributes among medalists. The sport requires significant lower-body strength and endurance to handle high speeds, sharp turns, and varying terrain, favouring athletes who have had years to develop and refine these physical traits. These trends underscore the long career trajectories of alpine skiers compared to gymnasts, emphasising the endurance and sustained excellence needed to compete at the highest levels of the sport.

	Gymnastics	Alpine Ski
Subset 1 (all athletes in medal-winning countries)	<div>None</div> <div>.02 .03 .92 .03</div> <div>100%</div>	<div>None</div> <div>.04 .04 .88 .04</div> <div>100%</div>
Subset 3 (all athletes in top 5 performing countries)	<div>None</div> <div>.02 .02 .93 .02</div> <div>100%</div>	<div>None</div> <div>.05 .07 .83 .06</div> <div>100%</div>

Table A1: Regression trees that failed to grow for subsets 1 & 3 of gymnastics and ski datasets

As seen from Table A1, no regression tree was successfully grown for both subsets 1 and 3 of the gymnastics and alpine ski datasets. This is due to the significant imbalance in the medal distribution, where the proportion of non-medalists far outweighs the other three categories (gold, silver, and bronze). Such an imbalance in the target variable can impede the regression tree's ability to meaningfully split the data and capture distinctions among the smaller, underrepresented categories.



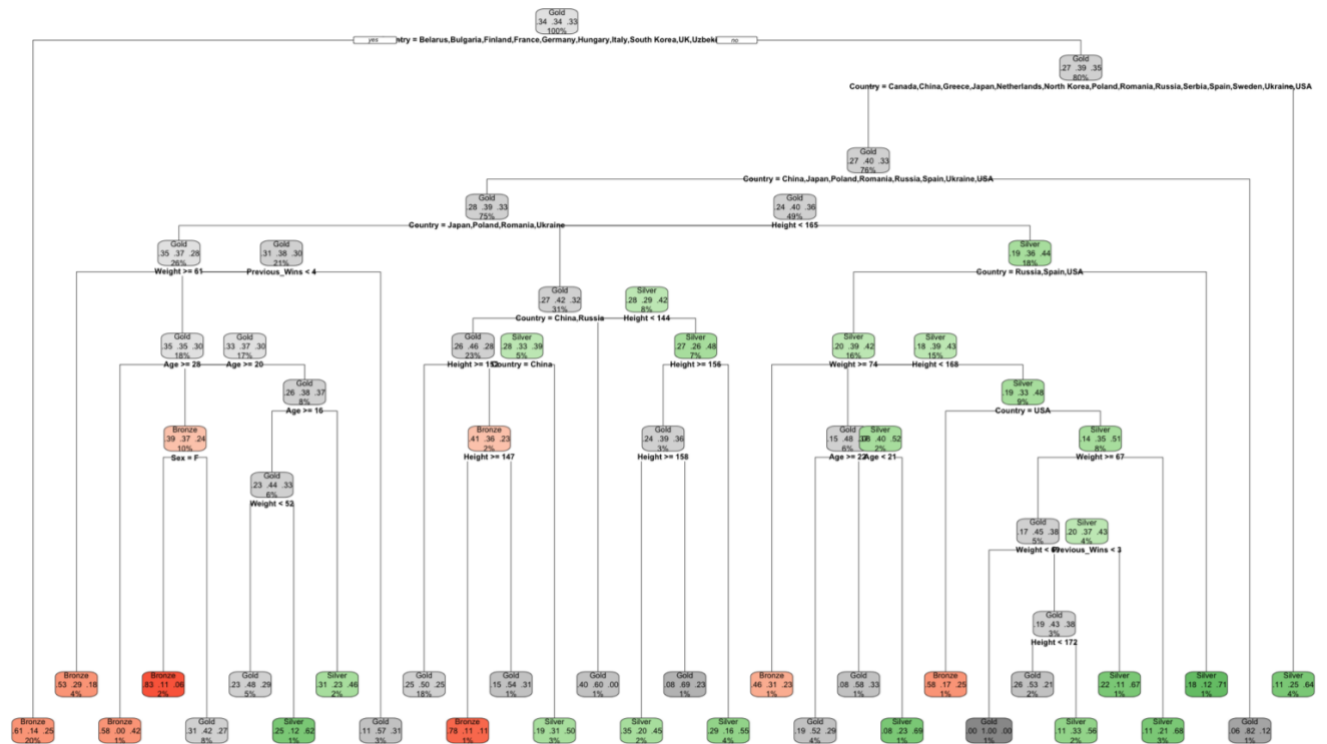


Fig A8a: Regression tree for all medalists in gymnastics

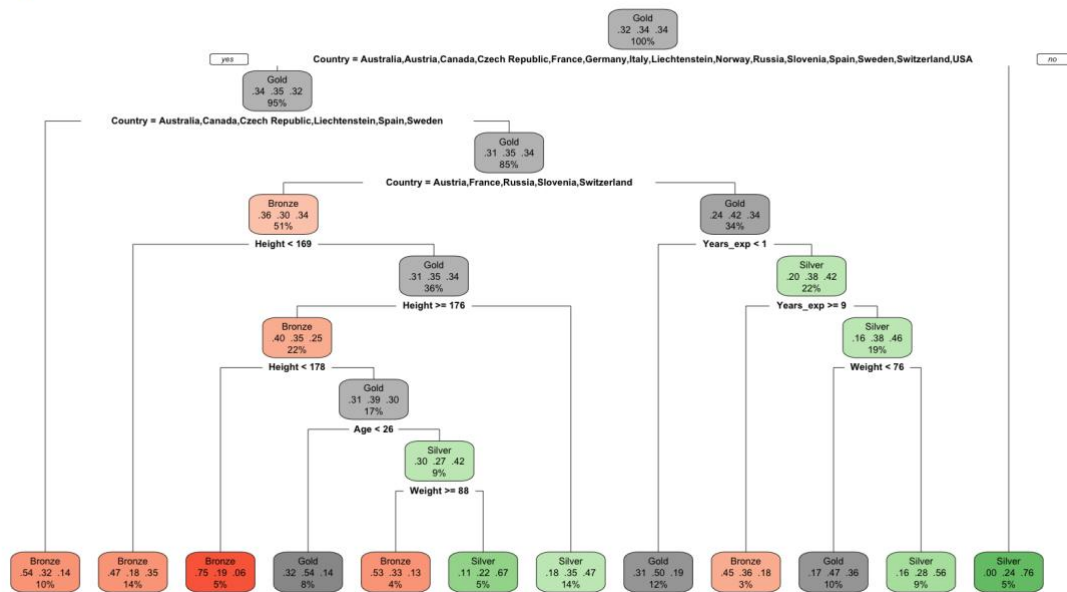


Fig A8b: Regression tree for all medalists in alpine ski

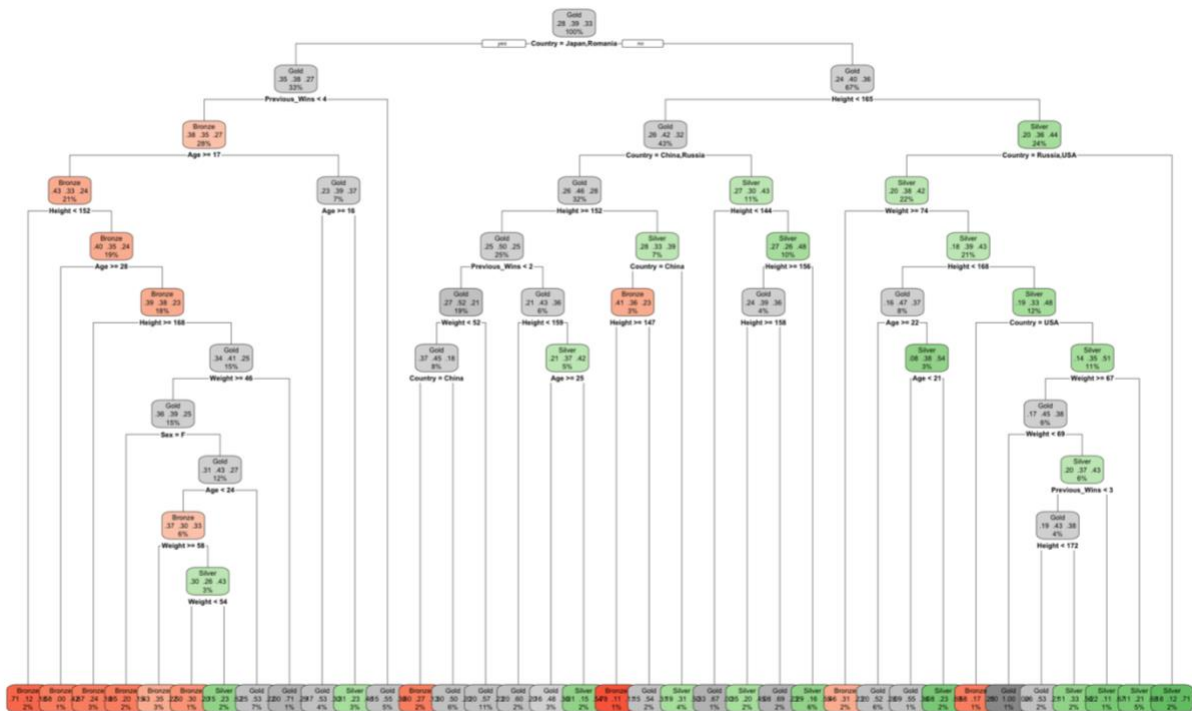


Fig A9a: Regression tree for medalists from top 5 countries in gymnastics

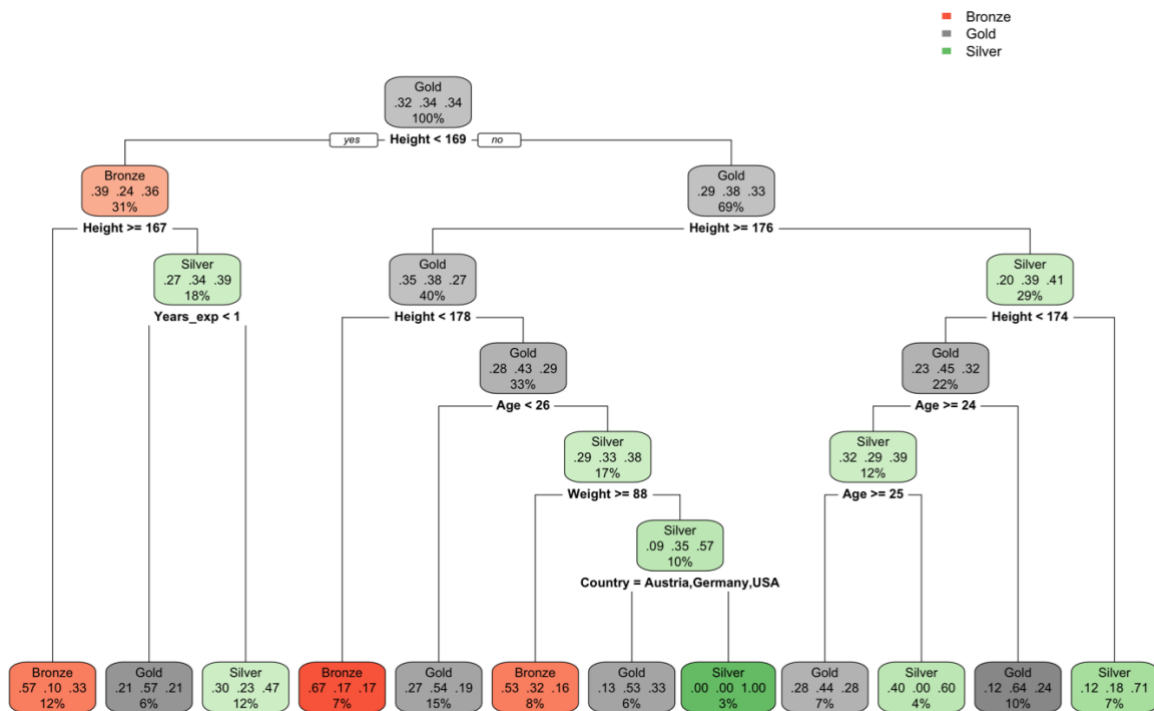


Fig A9b: Regression tree for medalists from top 5 countries in alpine ski

	Bronze	Gold	Silver	MeanDecreaseAccuracy	MeanDecreaseGini
Age	-11.84	19.34	-0.50	6.60	77.70
Height	1.64	14.61	3.98	15.9	88.04
Weight	-0.59	15.04	-2.70	9.86	88.00
Years of experience	-8.51	8.37	1.41	1.52	20.06

Previous wins	-1.22	3.52	3.36	4.42	28.17
Sex	0.99	6.31	2.11	9.11	8.34
Country	15.77	36.58	13.91	41.04	109.34

Table A2: Feature importance table from random forest for subset 2 of gymnastics data  
The influence of country is by far the most significant factor in determining medal outcomes for gymnasts. With very high scores for both MeanDecreaseAccuracy and MeanDecreaseGini, this highlights the systemic advantage that athletes from countries with established and well-funded sporting programs (such as Russia and the United States) possess.

	Bronze	Gold	Silver	MeanDecreaseAccuracy	MeanDecreaseGini
Age	4.22	9.61	2.86	10.19	34.26
Height	-4.03	8.20	-0.98	2.12	36.64
Weight	-4.19	10.53	1.25	5.49	40.26
Years of experience	4.45	4.33	-3.45	3.50	18.23
Previous wins	-3.08	5.44	0.77	2.23	10.93
Sex	0.04	1.59	-1.77	0.03	4.88
Country	1.01	9.66	-2.45	4.91	40.25

Table A3: Feature importance table from random forest for subset 2 of alpine ski data  
Similar to gymnastics, country of origin also has a huge impact on medal outcomes among alpine skiers, however, weight seems to have a more pronounced effect, surprisingly. Physical attributes in this discipline also seem to have a somewhat huge influence on success compared to gymnastics. This signifies that these traits contribute strongly to a skier's overall performance, shaping their ability to navigate courses, manage speed, and execute technical turns.

	Bronze	Gold	Silver	MeanDecreaseAccuracy	MeanDecreaseGini
Age	-12.82	18.58	2.66	10.37	54.08
Height	-1.57	13.82	6.70	16.82	63.23
Weight	-5.95	15.18	0.50	9.65	61.76
Years of experience	-2.93	3.87	3.87	4.38	12.85
Previous wins	-1.23	-0.10	2.06	0.49	19.93
Sex	-4.82	8.08	5.78	11.01	6.44
Country	-1.01	26.51	10.50	24.98	33.24

Table A4: Feature importance table from random forest for subset 4 of gymnastics data  
Comparing to feature importance in Table A2, though country remains a significant factor in both groups, the impact of physical attributes (height, weight) and experience appears to be more pronounced among gymnasts from the top-performing nations. This suggests that while national systems and resources play a critical role, individual characteristics and athlete development processes are key differentiators at the elite level.

	Bronze	Gold	Silver	MeanDecreaseAccuracy	MeanDecreaseGini
Age	2.53	5.67	-0.36	4.76	26.19
Height	-3.16	6.85	-1.80	0.94	28.39
Weight	1.23	8.76	-4.14	3.87	30.90
Years of experience	5.28	0.86	-5.11	0.55	12.69

Previous wins	-3.03	2.26	-2.78	-2.16	7.19
Sex	0.74	-0.07	-0.41	0.29	3.84
Country	5.39	1.45	-4.24	1.49	16.66

Table A5: Feature importance table from random forest for subset 4 of alpine ski data

Again, among elite skiers from top-performing countries, country of origin does not have as significant an influence on medal success compared to physical attributes. This is also a trend observed among gymnasts from the top-performing countries, which aligns with the broader understanding of elite sports. Most notably, the relative importance of years of experience among these skiers has a much significant influence on medal outcome compared to gymnast medalists in top performing countries. This indicates that sports with higher complexity, where athletes need to master varying terrains and weather conditions would benefit from refined skills and tactical awareness honed throughout the years.

## References

1. “120 Years of Olympic History: Athletes and Results.” *Kaggle*, 15 June 2018, [www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results/data?select=noc\\_regions.csv](https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results/data?select=noc_regions.csv).
2. Carayol, Tomaini. “Biles Is Epitome of Shift in Gymnastics That Prioritises Autonomy Over Age.” *The Guardian*, 2 Aug. 2024, [www.theguardian.com/sport/article/2024/aug/02/gymnastics-simone-biles-autonomy-age-paris-2024-olympic-games](https://www.theguardian.com/sport/article/2024/aug/02/gymnastics-simone-biles-autonomy-age-paris-2024-olympic-games).
3. Vladimírova, Alexandra, and Russkiy Reporter. “What Happened to the Soviet Union’s Gymnastics Coaches?” *Russia Beyond*, 3 Dec. 2024, [www.rbth.com/sport/2013/12/22/what\\_happen\\_to\\_the\\_soviet\\_unions\\_gymnastic\\_coaches\\_32791.html](https://www.rbth.com/sport/2013/12/22/what_happen_to_the_soviet_unions_gymnastic_coaches_32791.html).