



## **Rapport Jeux Olympiques**

**Rapport écrit par:**

**Enzo LERICHE, Samuel DARMALINGON, Franck TANKHAPANYA  
BUT Science des données FI EMS  
2023-11-13**

## Table des matières

Introduction: .....	3
Données.....	4
Importation des données.....	4
Description de la base de données .....	4
Analyses descriptives .....	6
Histogrammes pour les variables numériques.....	6
Graphiques en barre pour les variables catégorielles.....	7
Comparaison entre été et hiver pour le nombre de pays et de sportifs.....	10
Nombre de pays .....	10
Nombre de sportifs distincts par pays.....	11
Analyse comparative .....	15
Analyse de la distribution bivariées chez les sportifs.....	15
Poids/Genre.....	15
Taille/Genre.....	16
Poids/Taille/Genre .....	17
Poids/Genre selon la saison.....	18
Taille/Genre selon la saison.....	20
Poids/Taille/Genre selon la saison .....	22
Médailles remportées au total.....	24
Global.....	24
Comparaison entre Eté et Hiver .....	26
Top 10 des pays les plus médaillés .....	28
Global.....	28
Comparaison entre Eté et Hiver .....	30
TOP 5 des sports les plus médaillés .....	33
Global.....	33
Comparaison entre Eté et Hiver .....	34
Pays les plus médaillés dans le top 5 sports.....	36
Eté.....	36
Hiver.....	39
Pays des sportifs les plus médaillés .....	41
Global.....	42
Comparaison entre Eté et Hiver .....	43
Conclusion.....	45

## Introduction:

Les Jeux olympiques, en tant que rendez-vous sportif de renommée mondiale, symbolisent l'apogée des rêves et des ambitions pour les athlètes et les espoirs du monde du sport. Leur célébration transcende les barrières nationales, faisant des Jeux Olympiques un événement d'envergure universelle. Notre démarche de recherche s'est focalisée sur une analyse approfondie des performances des athlètes qui ont pris part aux Jeux Olympiques, englobant une période significative de 1896 à 2016. Les données nécessaires à cette étude ont été extraites du site "Kaggle.com". Afin de poursuivre notre exploration, nous avons opté pour le langage Python.

La base de données que nous avons utilisée offre une richesse d'informations, permettant une comparaison approfondie de multiples variables. Cette diversité de données nous offre l'opportunité d'explorer et de déceler d'éventuels liens entre différentes caractéristiques des athlètes et leurs performances aux Jeux Olympiques. Par exemple, nous pourrions analyser les variations de poids entre les athlètes participant aux Jeux olympiques d'été et d'hiver, cherchant à identifier des tendances significatives ou des différences marquées.

L'objectif principal de notre analyse consiste à déterminer quels pays se sont distingués dans les différentes disciplines des Jeux Olympiques et à examiner les éventuelles disparités entre les Jeux Olympiques d'Hiver et d'Été. De plus, nous avons cherché à comprendre les caractéristiques propres aux athlètes qui influent sur leur capacité à remporter des médailles aux Jeux olympiques. En se penchant sur la période de 1896 à 2016, nous nous attardons particulièrement sur l'évolution des performances aux Jeux olympiques d'été et d'hiver au cours de ces trois décennies.

### ***Problématique :***

*Comment les performances sportives des pays participants aux Jeux Olympiques se distinguent-elles, et quelles sont les tendances qui émergent en termes de succès et de régularité?*

*Et comment les caractéristiques individuelles des athlètes, influent-elles sur leurs performances aux Jeux Olympiques d'été et d'hiver?*

## Données

### Importation des données

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("C:/Users/User/OneDrive/Documents/BUT/2eme annee/python/proj
et/2eme/athlete_events.csv")
```

Dans notre script Python, nous avons initié le processus en important plusieurs bibliothèques qui seront essentielles pour mener à bien nos analyses. Celles-ci comprennent :

- **import pandas as pd**: Cette instruction nous permet d'utiliser la bibliothèque Pandas pour créer et manipuler des structures de données de type data.frame, offrant une puissante gestion des données tabulaires en Python.
- **import matplotlib.pyplot as plt**: Cette ligne importe la bibliothèque Matplotlib, principalement utilisée pour générer des visualisations et des graphiques à partir de nos données. L'utilisation de plt comme alias facilite son utilisation dans le code.
- **import seaborn as sns**: L'instruction importe la bibliothèque Seaborn qui, basée sur Matplotlib, fournit une interface simplifiée pour créer des graphiques statistiques. Cette bibliothèque est appréciée pour ses capacités à produire des graphiques informatifs et esthétiquement agréables.

Ces importations constituent la base de notre code, permettant ainsi d'exploiter les fonctionnalités avancées offertes par ces bibliothèques tout au long de notre analyse de données.

Ensuite, nous avons importé le fichier *athlete\_events.csv* pour initier notre étude statistique.

### Description de la base de données

*#Informations générales sur Le jeu de données*

```
df.info()

## <class 'pandas.core.frame.DataFrame'>
## RangeIndex: 271116 entries, 0 to 271115
## Data columns (total 15 columns):
## #   Column   Non-Null Count  Dtype
## ---  -
## 0    ID      271116 non-null  int64
## 1    Name     271116 non-null  object
## 2    Sex      271116 non-null  object
## 3    Age      261642 non-null  float64
## 4    Height   210945 non-null  float64
## 5    Weight   208241 non-null  float64
## 6    Team     271116 non-null  object
## 7    NOC      271116 non-null  object
## 8    Games    271116 non-null  object
```

```

## 9   Year      271116 non-null  int64
## 10  Season    271116 non-null  object
## 11  City      271116 non-null  object
## 12  Sport     271116 non-null  object
## 13  Event     271116 non-null  object
## 14  Medal     39783 non-null   object
## dtypes: float64(3), int64(2), object(10)
## memory usage: 31.0+ MB

#Nom des colonnes
df.columns.tolist()

## ['ID', 'Name', 'Sex', 'Age', 'Height', 'Weight', 'Team', 'NOC', 'Games', '
Year', 'Season', 'City', 'Sport', 'Event', 'Medal']

#Nombre de valeurs manquantes par colonne
df.isnull().sum()

## ID                0
## Name              0
## Sex               0
## Age              9474
## Height           60171
## Weight           62875
## Team              0
## NOC               0
## Games             0
## Year              0
## Season            0
## City              0
## Sport             0
## Event             0
## Medal            231333
## dtype: int64

#Statistiques descriptives
df.describe()

##              ID              Age  ...              Weight              Year
## count  271116.000000  261642.000000  ...  208241.000000  271116.000000
## mean    68248.954396    25.556898  ...    70.702393    1978.378480
## std     39022.286345     6.393561  ...    14.348020     29.877632
## min       1.000000    10.000000  ...    25.000000    1896.000000
## 25%     34643.000000    21.000000  ...    60.000000    1960.000000
## 50%     68205.000000    24.000000  ...    70.000000    1988.000000
## 75%    102097.250000    28.000000  ...    79.000000    2002.000000
## max    135571.000000    97.000000  ...   214.000000    2016.000000
##
## [8 rows x 5 columns]

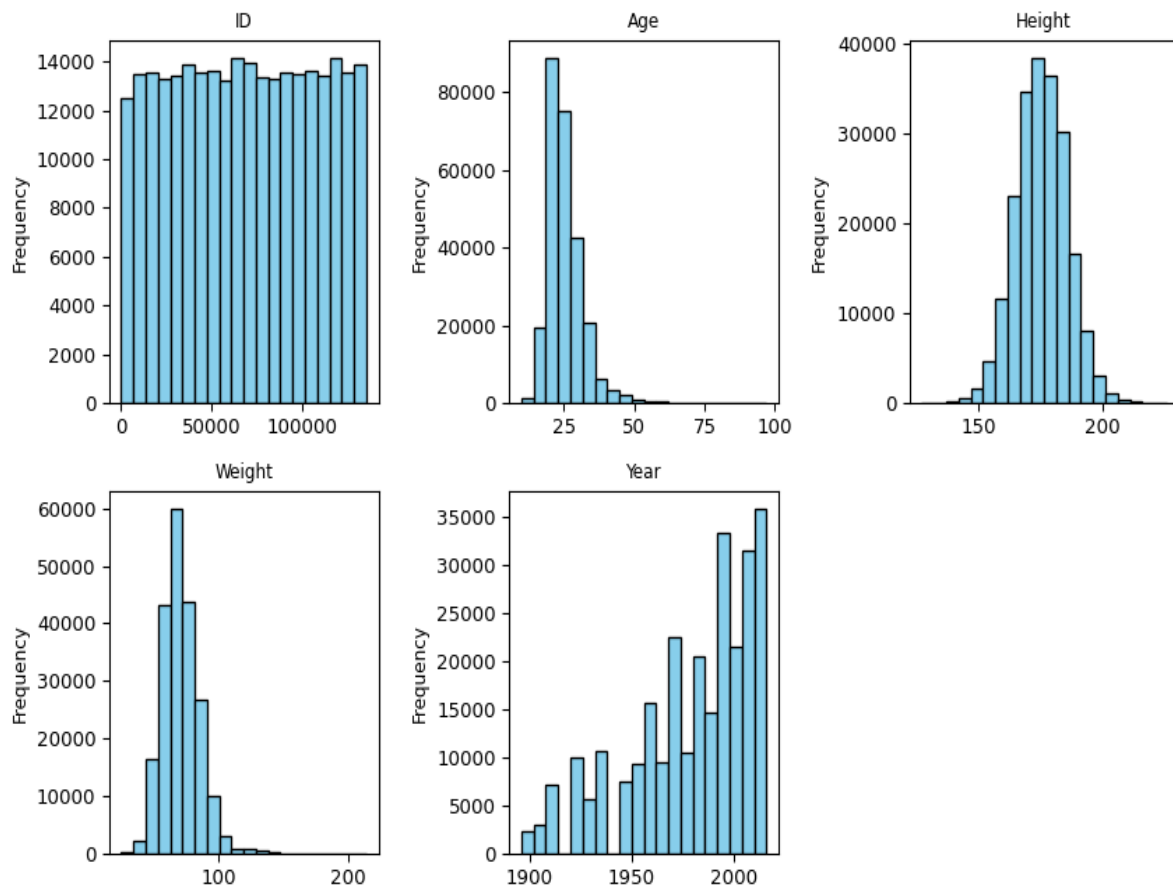
```

Nous disposons d'une base de données comprenant 271116 lignes pour 15 colonnes. Cette base de données se concentre sur des informations relatives aux athlètes. Il est important de noter que certains athlètes peuvent ne pas avoir remporté de médailles, ce qui explique pourquoi la colonne "Médailles" présente le plus grand nombre de données manquantes dans notre ensemble de données.

## Analyses descriptives

### Histogrammes pour les variables numériques

```
numeric_columns = df.select_dtypes(include='number').columns
plt.figure(figsize=(9, 6))
for col in numeric_columns:
    plt.subplot(2, 3, numeric_columns.get_loc(col) + 1)
    plt.hist(df[col].dropna(), bins=20, color='skyblue', edgecolor='black')
    plt.title(col, fontsize=9)
    #plt.xlabel(col, fontsize=9)
    plt.ylabel('Frequency', fontsize=9)
plt.tight_layout()
plt.show()
```



Parmi les cinq histogrammes que nous avons générés, quatre d'entre eux fournissent des insights pertinents. En effet, la représentation visuelle de l'ID ne présente pas d'intérêt particulier, car cette variable fonctionne simplement comme un code unique pour chaque sportif.

Le graphique d'âge offre une vue sur la répartition des sportifs en fonction de leur âge, mettant en évidence une concentration significative entre 20 et 30 ans. La moyenne d'âge est de 25,55 ans, avec une médiane de 24 ans, renforçant cette observation.

L'histogramme de la variable "height" illustre une tendance autour de 170 à 180 cm, avec une moyenne de 175,33 cm.

De même, l'histogramme de la variable "Weight" indique une concentration entre 60 et 75 kg, avec une moyenne de 70 kg.

L'histogramme représentant les années est essentiel pour observer l'évolution du nombre de sportifs au fil des Jeux olympiques. On constate une croissance significative, passant d'environ 5000 sportifs au début à plus de 30000 sportifs lors des derniers JO en 2016. Cette évolution suggère des hypothèses intéressantes, telles que l'augmentation du nombre d'épreuves au fil des années.

### Graphiques en barre pour les variables catégorielles

```
categorical_columns = df.select_dtypes(include='object').columns
for col in categorical_columns:
    if df[col].nunique() <= 10: #Afficher uniquement les variables avec 10 catégories ou moins (préciser dans le rapport pourquoi)
        plt.figure(figsize=(9, 5))
        sns.countplot(data=df, x=col, palette='viridis')
        plt.title(f'Graphique en barre pour la variable {col}')
        plt.xlabel(col)
        plt.ylabel('Count')
        plt.xticks(rotation=45)
        plt.tight_layout()
        plt.show()

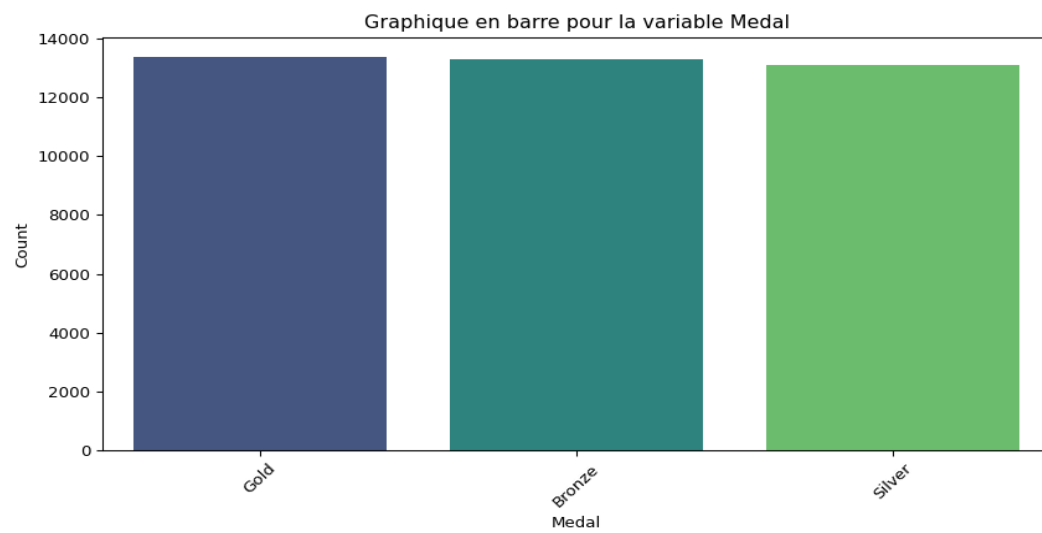
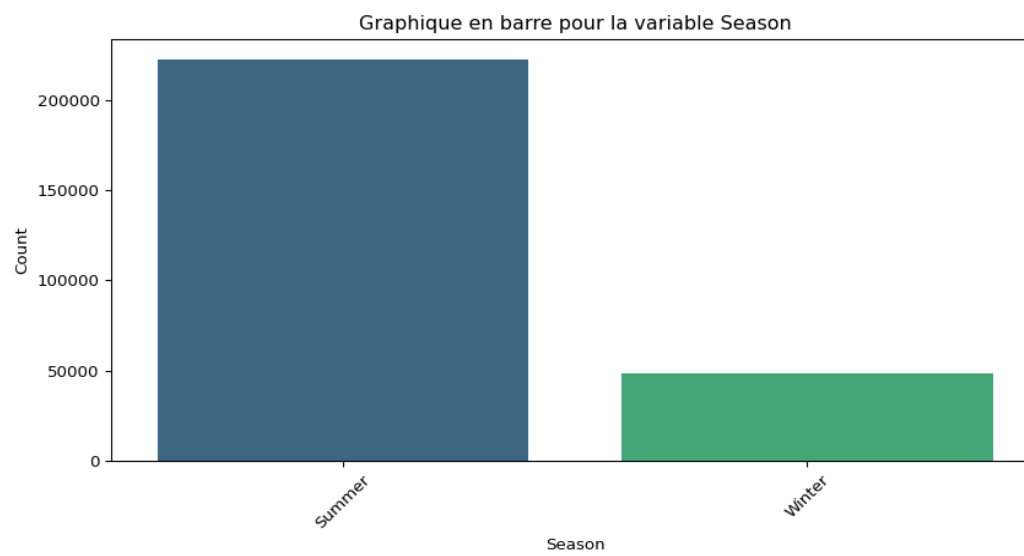
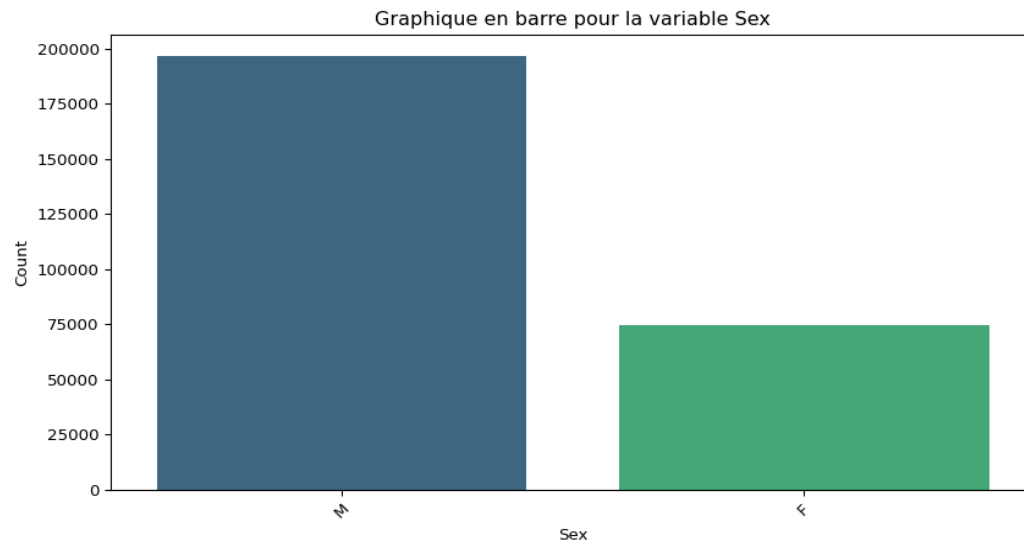
## <Figure size 900x500 with 0 Axes>
## <Axes: xlabel='Sex', ylabel='count'>
## Text(0.5, 1.0, 'Graphique en barre pour la variable Sex')
## Text(0.5, 0, 'Sex')
## Text(0, 0.5, 'Count')
## ([0, 1], [Text(0, 0, 'M'), Text(1, 0, 'F')])
## <Figure size 900x500 with 0 Axes>
## <Axes: xlabel='Season', ylabel='count'>
## Text(0.5, 1.0, 'Graphique en barre pour la variable Season')
## Text(0.5, 0, 'Season')
## Text(0, 0.5, 'Count')
## ([0, 1], [Text(0, 0, 'Summer'), Text(1, 0, 'Winter')])
## <Figure size 900x500 with 0 Axes>
## <Axes: xlabel='Medal', ylabel='count'>
## Text(0.5, 1.0, 'Graphique en barre pour la variable Medal')
```

```

## Text(0.5, 0, 'Medal')
## Text(0, 0.5, 'Count')
## ([0, 1, 2], [Text(0, 0, 'Gold'), Text(1, 0, 'Bronze'), Text(2, 0, 'Silver'
)])
##
## <string>:4: FutureWarning:
##
## Passing `palette` without assigning `hue` is deprecated and will be remove
d in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the
same effect.
##
## <string>:4: FutureWarning:
##
## Passing `palette` without assigning `hue` is deprecated and will be remove
d in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the
same effect.
##
## <string>:4: FutureWarning:
##
## Passing `palette` without assigning `hue` is deprecated and will be remove
d in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the
same effect.

```





Nous avons créé des graphiques pour des variables catégorielles, en choisissant des variables avec un maximum de 10 catégories. Ce choix s'explique par le souci de rendre les graphiques lisibles et faciles à interpréter.

Les deux premiers graphiques en barres nous interrogent sur l'évolution des Jeux olympiques. Le premier graphique représente la répartition des sexes des sportifs aux Jeux olympiques d'été et d'hiver. On observe une nette prédominance des sportifs masculins, ce qui peut être expliqué par l'histoire des JO, les femmes ayant été autorisées à participer seulement à partir des années 2000, avec un nombre initial d'épreuves limité.

Le deuxième graphique présente la répartition du nombre de sportifs participant aux JO d'été et d'hiver. La disparité entre les deux est évidente, avec un nombre beaucoup plus élevé de sportifs aux JO d'été. Cette différence peut être attribuée au début ultérieur des JO d'hiver, mais aussi à la disparité du nombre d'épreuves entre les deux saisons. En 2016, par exemple, les JO d'été comportaient environ 306 épreuves dans 28 sports, tandis que les JO d'hiver de 2014 comptaient environ 98 épreuves dans 15 disciplines.

Quant au dernier graphique, il représente le nombre de médailles d'or, d'argent et de bronze. On n'observe pas de grandes différences, ce qui est logique étant donné que chaque épreuve distribue une médaille d'or, une d'argent et une de bronze.

## Comparaison entre été et hiver pour le nombre de pays et de sportifs

*On va séparer le jeu de données en 2 pour pouvoir avoir un jeu pour les jeux d'été et un jeu pour les jeux d'hiver.*

```
df_ete = df[df['Season'] == 'Summer']
df_hiver = df[df['Season'] == 'Winter']
```

### Nombre de pays

*Global*

```
num_pays = df['NOC'].nunique()
print(f"Nombre de pays : {num_pays}")
## Nombre de pays : 230
```

*Nombre de pays distinct en été*

```
num_pays_ete = df_ete['NOC'].nunique()
print(f"Nombre de pays qui participent aux JO d'été : {num_pays_ete}")
## Nombre de pays qui participent aux JO d'été : 230
```

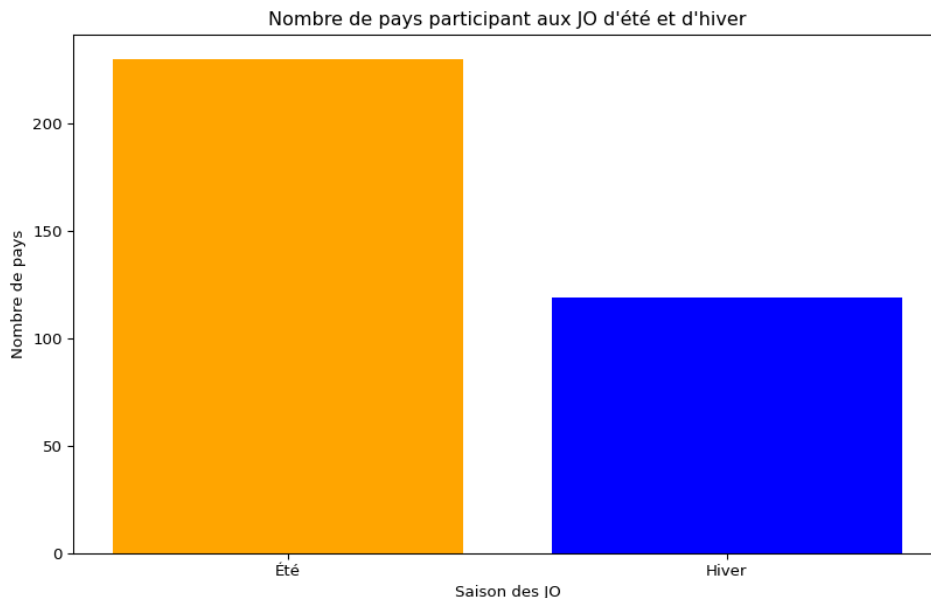
*Nombre de pays distinct distinct en hiver*

```
num_pays_hiver = df_hiver['NOC'].nunique()
print(f"Nombre de pays participant aux JO d'hiver : {num_pays_hiver}")
```

```
## Nombre de pays participant aux JO d'hiver : 119
```

*Graphique pour le nombre de pays participant aux JO d'été et d'hiver*

```
plt.figure(figsize=(10, 6))
plt.bar(['Été', 'Hiver'], [num_pays_ete, num_pays_hiver], color=['orange', 'blue'])
plt.title("Nombre de pays participant aux JO d'été et d'hiver")
plt.xlabel("Saison des JO")
plt.ylabel("Nombre de pays")
plt.show()
```



Le graphique illustre la participation, passée et actuelle, des pays aux Jeux olympiques d'été et d'hiver. Il met en évidence de manière frappante la disparité du nombre de pays entre les deux saisons. Aux Jeux d'été, nous recensons la participation de 230 pays, tandis que pour les Jeux olympiques d'hiver, ce chiffre s'élève à 119.

Cette disparité peut être attribuée à plusieurs facteurs, notamment la diversité des sports estivaux et hivernaux, le climat géographique des pays et le développement historique des deux événements. Les JO d'été offrent un éventail plus large de disciplines, ce qui attire un nombre supérieur de nations. Par ailleurs, les pays situés dans des zones à climat plus tempéré peuvent être davantage représentés aux JO d'été, tandis que les JO d'hiver suscitent la participation de nations bénéficiant de conditions hivernales propices aux sports d'hiver.

### Nombre de sportifs distincts par pays

*Nombre de sportif total*

```
sportifs_total = df['Name'].nunique()
print("\nNombre total de sportifs distincts :")
```

```
##  
## Nombre total de sportifs distincts :  
print(sportifs_total)  
## 134732
```

*Nombre de sportif en été*

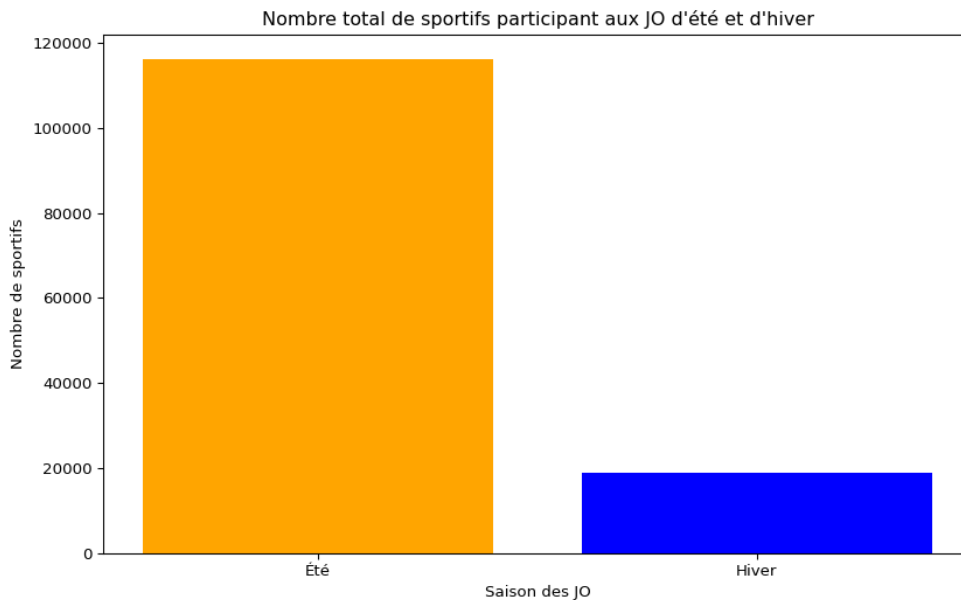
```
sportifs_total_ete = df_ete['Name'].nunique()  
print("\nNombre total de sportifs distincts participant aux JO d'été :")  
  
##  
## Nombre total de sportifs distincts participant aux JO d'été :  
print(sportifs_total_ete)  
## 116122
```

*Nombre de sportif en hiver*

```
sportifs_total_hiver = df_hiver['Name'].nunique()  
print("\nNombre total de sportifs distincts participant aux JO d'hiver :")  
  
##  
## Nombre total de sportifs distincts participant aux JO d'hiver :  
print(sportifs_total_hiver)  
## 18923
```

*Graphique pour le nombre de sportifs participants aux JO d'été et d'hiver*

```
plt.figure(figsize=(10, 6))  
plt.bar(['Été', 'Hiver'], [sportifs_total_ete, sportifs_total_hiver], color=['orange', 'blue'])  
plt.title("Nombre total de sportifs participant aux JO d'été et d'hiver")  
plt.xlabel("Saison des JO")  
plt.ylabel("Nombre de sportifs")  
plt.show()
```



Le graphique illustre le nombre de sportifs distincts ayant participé ou participant aux Jeux olympiques d'été et d'hiver. Une disparité marquée entre les deux saisons est perceptible.

Auparavant, nous avons déterminé le nombre spécifique de sportifs pour chaque édition des Jeux Olympiques, soit 116 122 pour les JO d'été et 18 923 pour les JO d'hiver. Notons que le nombre total de sportifs distincts s'élève à 134 732. Ces chiffres soulignent clairement que les athlètes des JO d'été représentent plus de 85% du total des sportifs participant aux Jeux Olympiques.

Cette disparité peut être expliquée par divers facteurs, tels que la diversité des sports estivaux par rapport à ceux d'hiver, ainsi que la plus longue histoire et l'ancrage culturel plus étendu des JO d'été.

#### *Nombre de sportifs par pays*

```
sportifs_par_pays = df.groupby('NOC')['Name'].nunique() #on peut utiliser la
variable "name" ou "ID", vu que "ID" correspond à un code unique donnée à un
sportif
print("Nombre de sportifs distincts par pays :")

## Nombre de sportifs distincts par pays :

print(sportifs_par_pays)

## NOC
## AFG      103
## AHO       55
## ALB       45
## ALG      373
## AND       61
## ...
```

```
## YEM      28
## YMD       5
## YUG    1269
## ZAM     145
## ZIM     196
## Name: Name, Length: 230, dtype: int64
```

*Nombre de sportifs distincts par pays en été*

```
sportifs_par_pays_ete = df_ete.groupby('NOC')['Name'].nunique() #on peut utiliser la variable "name" ou "ID", vu que "ID" correspond a un code unique donnée a un sportif
```

```
print("Nombre de sportifs distincts par pays participant aux JO d'été :")
```

```
## Nombre de sportifs distincts par pays participant aux JO d'été :
```

```
print(sportifs_par_pays_ete)
```

```
## NOC
## AFG      103
## AHO       52
## ALB       43
## ALG      366
## AND       33
## ...
## YEM       28
## YMD        5
## YUG     1035
## ZAM      145
## ZIM      195
## Name: Name, Length: 230, dtype: int64
```

*Nombre de sportifs distincts par pays en hiver*

```
sportifs_par_pays_hiver = df_hiver.groupby('NOC')['Name'].nunique() #on peut utiliser la variable "name" ou "ID", vu que "ID" correspond a un code unique donnée a un sportif
```

```
print("Nombre de sportifs distincts par pays participant aux JO d'hiver :")
```

```
## Nombre de sportifs distincts par pays participant aux JO d'hiver :
```

```
print(sportifs_par_pays_hiver)
```

```
## NOC
## AHO        3
## ALB        2
## ALG        7
## AND       28
## ARG      136
## ...
## USA     1697
## UZB       21
```

```
## VEN      5
## YUG     234
## ZIM      1
## Name: Name, Length: 119, dtype: int64
```

## Analyse comparative

Le but maintenant va être de faire des analyses de comparaison pour pouvoir répondre à notre problématique. Notre but étant de détecter de potentielle caractéristique qui font que des pays ou des sportifs remporte plus de médailles que d'autres.

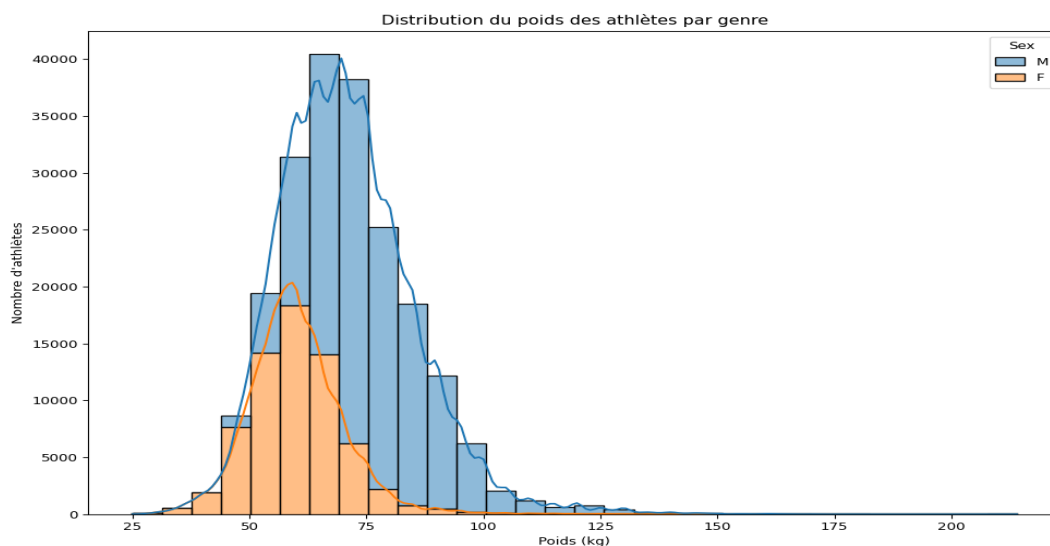
## Analyse de la distribution bivariées chez les sportifs

Pour entamer notre analyse, nous allons examiner les tendances potentielles relatives aux variables telles que le poids et la taille. Nous observerons les tendances générales en comparant l'ensemble des données, que ce soit pour les Jeux d'été ou d'hiver.

Par ailleurs, nous effectuerons des comparaisons saisonnières afin d'identifier d'éventuelles différences entre l'été et l'hiver. Cette démarche nous permettra d'explorer plus en profondeur les caractéristiques physiques des athlètes et de dégager des observations significatives.

### Poids/Genre

```
plt.figure(figsize=(12, 8))
sns.histplot(data=df, x='Weight', hue='Sex', multiple='stack', bins=30, kde=True)
plt.title("Distribution du poids des athlètes par genre")
plt.xlabel("Poids (kg)")
plt.ylabel("Nombre d'athlètes")
plt.show()
```

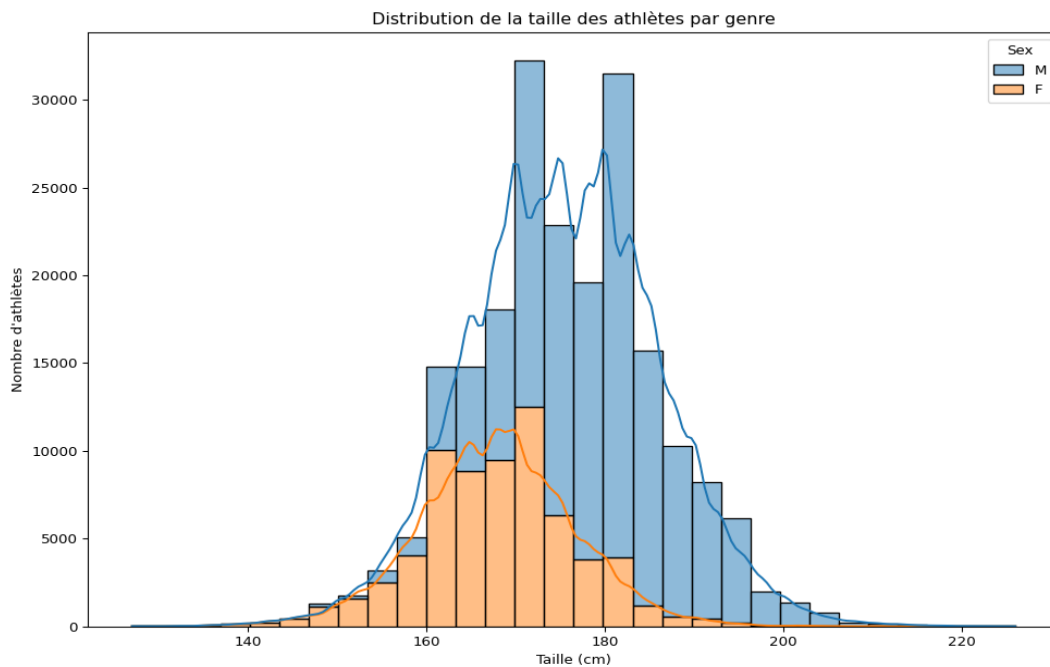


Le diagramme illustre deux histogrammes distincts, l'un décrivant la distribution des poids chez les femmes et l'autre chez les hommes. Une disparité évidente émerge, mais elle s'explique naturellement par la tendance générale : en moyenne, les femmes ont tendance à avoir un poids inférieur à celui des hommes.

Le graphique intègre également une courbe de densité, générée avec l'option "kde=True", offrant ainsi une visualisation de la distribution estimée des données. Dans le cas des femmes, la moyenne se situe autour de 60 kg, tandis que pour les hommes, elle se positionne davantage aux alentours de 75 kg. Cette représentation graphique permet de saisir rapidement les tendances de poids au sein des deux groupes de genre.

### Taille/Genre

```
plt.figure(figsize=(12, 8))
sns.histplot(data=df, x='Height', hue='Sex', multiple='stack', bins=30, kde=True)
plt.title("Distribution de la taille des athlètes par genre")
plt.xlabel("Taille (cm)")
plt.ylabel("Nombre d'athlètes")
plt.show()
```



Le graphique présente deux histogrammes distincts, l'un décrivant la distribution des tailles chez les femmes et l'autre chez les hommes. Une disparité évidente émerge, mais elle s'explique naturellement par la tendance générale : en moyenne, les femmes ont tendance à être de taille inférieure à celle des hommes.



Le graphique intègre également une courbe de densité, générée avec l'option "kde=True", offrant ainsi une visualisation de la distribution estimée des données. Dans le cas des femmes, la moyenne de taille se situe autour de 165 cm, tandis que pour les hommes, elle se positionne davantage aux alentours de 180 cm. Cette représentation graphique permet de saisir rapidement les tendances de taille au sein des deux groupes de genre.

### Poids/Taille/Genre

```
correlation_by_sex = df.groupby('Sex')[['Height', 'Weight']].corr()

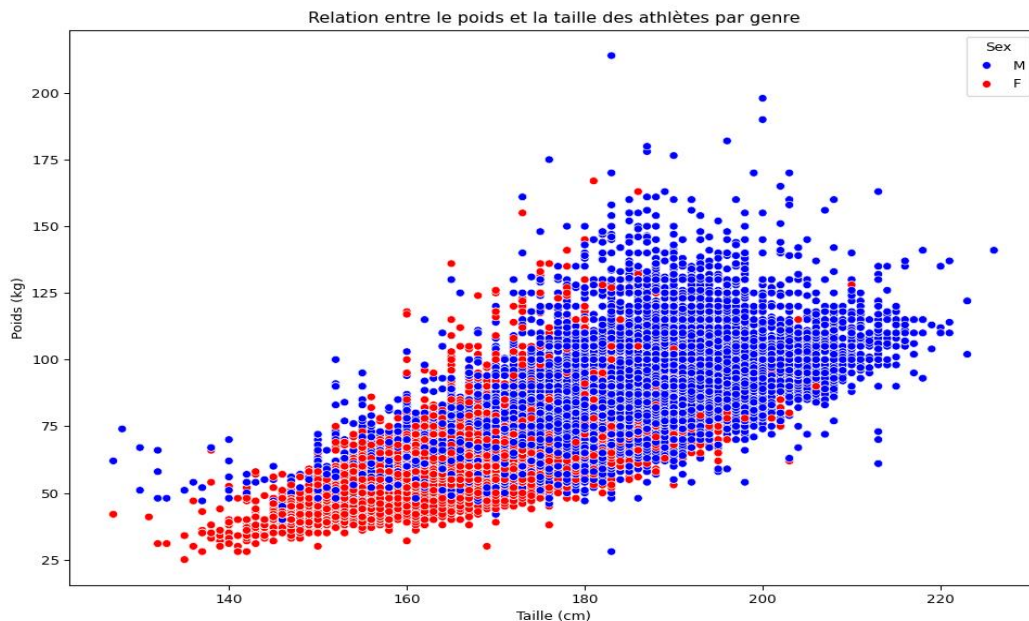
print("Corrélation entre la taille et le poids par sexe :")

## Corrélation entre la taille et le poids par sexe :

print(correlation_by_sex)

##           Height    Weight
## Sex
## F   Height  1.000000  0.740115
##      Weight  0.740115  1.000000
## M   Height  1.000000  0.726983
##      Weight  0.726983  1.000000

plt.figure(figsize=(12, 8))
sns.scatterplot(data=df, x='Height', y='Weight', hue='Sex', palette={'M': 'blue', 'F': 'red'})
plt.title("Relation entre le poids et la taille des athlètes par genre")
plt.xlabel("Taille (cm)")
plt.ylabel("Poids (kg)")
plt.show()
```



Le nuage de points ci-dessus représente la relation entre le poids et la taille des athlètes, avec une distinction entre les sexes. On observe une tendance intéressante : les points situés dans la partie inférieure gauche du graphique, où la taille et le poids sont plus petits, sont majoritairement associés aux femmes. En revanche, les points dans la partie supérieure droite, où la taille et le poids sont plus importants, sont principalement liés aux hommes.

Cette observation suggère une corrélation entre la taille, le poids et le sexe des athlètes. En général, les femmes tendent à avoir des tailles et des poids plus modestes, tandis que les hommes présentent des valeurs plus élevées dans ces deux variables. La corrélation entre la taille et le poids, calculée à partir des données, confirme cette observation avec une valeur de 0.74 pour les femmes et 0.73 pour les hommes. Ces coefficients positifs indiquent une relation positive modérée entre la taille et le poids pour les deux sexes.

L'analyse de ce nuage de points offre ainsi une perspective visuelle sur la répartition des athlètes en fonction de la taille, du poids et du sexe, mettant en évidence des tendances significatives et étayées par des mesures de corrélation.

### Poids/Genre selon la saison

```
#moyenne du poids des femmes pour Les JO d'été et d'hiver
poids_moyen_femme_ete = df_ete[df_ete['Sex'] == 'F']['Weight'].mean()
poids_moyen_femme_hiver = df_hiver[df_hiver['Sex'] == 'F']['Weight'].mean()

print(f"Moyenne du poids des femmes aux JO d'été : {poids_moyen_femme_ete:.2f
} kg")

## Moyenne du poids des femmes aux JO d'été : 60.09 kg

print(f"Moyenne du poids des femmes aux JO d'hiver : {poids_moyen_femme_hiver
:.2f} kg")

## Moyenne du poids des femmes aux JO d'hiver : 59.76 kg

#moyenne du poids des hommes pour Les JO d'été et d'hiver
poids_moyen_homme_ete = df_ete[df_ete['Sex'] == 'M']['Weight'].mean()
poids_moyen_homme_hiver = df_hiver[df_hiver['Sex'] == 'M']['Weight'].mean()

print(f"Moyenne du poids des hommes aux JO d'été : {poids_moyen_homme_ete:.2f
} kg")

## Moyenne du poids des hommes aux JO d'été : 75.60 kg

print(f"Moyenne du poids des hommes aux JO d'hiver : {poids_moyen_homme_hiver
:.2f} kg")

## Moyenne du poids des hommes aux JO d'hiver : 76.36 kg

#poids maximum des hommes pour Les JO d'été et d'hiver
poids_max_homme_ete = df_ete[df_ete['Sex'] == 'M']['Weight'].max()
poids_max_homme_hiver = df_hiver[df_hiver['Sex'] == 'M']['Weight'].max()

print(f"Poids maximum des hommes aux JO d'été : {poids_max_homme_ete} kg")
```

```

## Poids maximum des hommes aux JO d'été : 214.0 kg

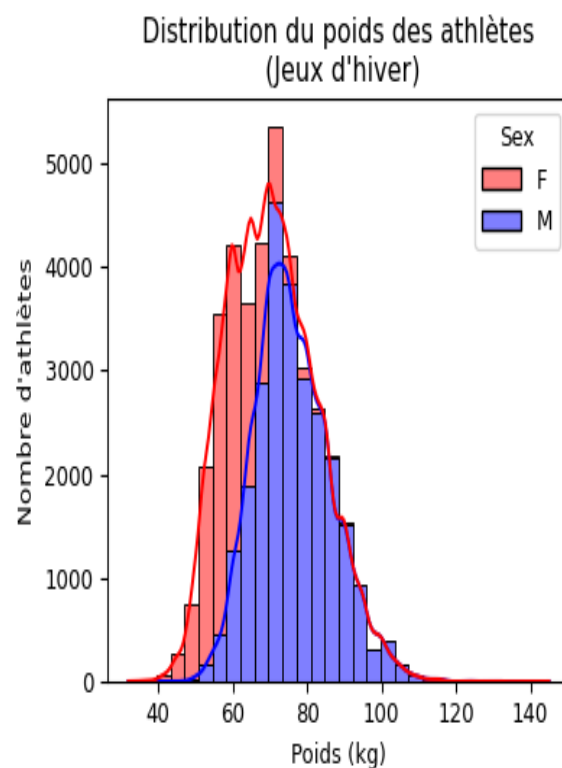
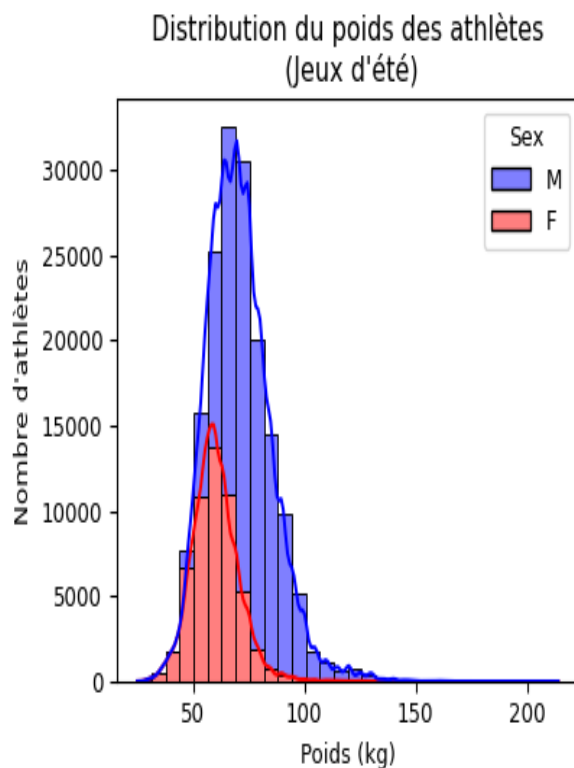
print(f"Poids maximum des hommes aux JO d'hiver : {poids_max_homme_hiver} kg"
)

## Poids maximum des hommes aux JO d'hiver : 145.0 kg

plt.figure(figsize=(9, 4))
# Distribution du poids pour Les Jeux d'été
plt.subplot(1, 2, 1)
sns.histplot(data=df_ete, x='Weight', hue='Sex', multiple='stack', bins=30, kde=True, palette={'M': 'blue', 'F': 'red'})
plt.title("Distribution du poids des athlètes\n (Jeux d'été)")
plt.xlabel("Poids (kg)")
plt.ylabel("Nombre d'athlètes")

# Distribution du poids pour Les Jeux d'hiver
plt.subplot(1, 2, 2)
sns.histplot(data=df_hiver, x='Weight', hue='Sex', multiple='stack', bins=30, kde=True, palette={'M': 'blue', 'F': 'red'})
plt.title("Distribution du poids des athlètes\n (Jeux d'hiver)")
plt.xlabel("Poids (kg)")
plt.ylabel("Nombre d'athlètes")
plt.subplots_adjust(wspace=20)
plt.tight_layout()
plt.show()

```



À gauche, le graphique représente le poids des athlètes aux Jeux olympiques d'été, tandis qu'à droite, nous observons la distribution pour les Jeux olympiques d'hiver. La juxtaposition de ces deux graphiques permet une comparaison visuelle entre les deux saisons olympiques. Une observation intéressante réside dans le fait que, pour les Jeux d'hiver, les femmes semblent avoir légèrement moins de poids par rapport aux Jeux d'été.

En analysant de plus près, nous constatons que la moyenne du poids des femmes pour les Jeux d'été est de 60 kg, tandis qu'elle diminue légèrement à 59 kg pour les Jeux d'hiver. Pour les hommes, le poids moyen semble rester relativement constant entre les deux saisons. Notons cependant que le poids maximum atteint 214 kg aux Jeux d'été, tandis qu'il est de 145 kg aux Jeux d'hiver.

Cette disparité entre les poids maximums peut potentiellement s'expliquer par la nature des épreuves olympiques. Les Jeux d'été comprennent un large éventail de disciplines, y compris des sports de force tels que l'haltérophilie et la musculation, où des poids plus élevés peuvent être atteints. En revanche, les Jeux d'hiver mettent en avant des sports nécessitant des compétences spécifiques, peut-être moins axés sur la force brute. Ainsi, la différence dans les épreuves olympiques entre été et hiver pourrait contribuer à cette variation dans les poids maximums observés.

La moyenne du poids des hommes reste relativement constante, avec 75,6 kg pour les Jeux d'été et 76,3 kg pour les Jeux d'hiver. Cette analyse offre un éclairage intéressant sur les variations de poids en fonction des saisons olympiques et soulève des questions intrigantes sur l'influence des disciplines sportives spécifiques sur les caractéristiques physiques des athlètes.

#### Taille/Genre selon la saison

```
# Moyenne de la taille des femmes pour Les JO d'été et d'hiver
taille_moyenne_femme_ete = df_ete[df_ete['Sex'] == 'F']['Height'].mean()
taille_moyenne_femme_hiver = df_hiver[df_hiver['Sex'] == 'F']['Height'].mean(
)

print(f"Moyenne de la taille des femmes aux JO d'été : {taille_moyenne_femme_ete:.2f} cm")

## Moyenne de la taille des femmes aux JO d'été : 168.17 cm

print(f"Moyenne de la taille des femmes aux JO d'hiver : {taille_moyenne_femme_hiver:.2f} cm")

## Moyenne de la taille des femmes aux JO d'hiver : 166.53 cm

# Moyenne de la taille des hommes pour Les JO d'été et d'hiver
taille_moyenne_homme_ete = df_ete[df_ete['Sex'] == 'M']['Height'].mean()
taille_moyenne_homme_hiver = df_hiver[df_hiver['Sex'] == 'M']['Height'].mean(
)

print(f"Moyenne de la taille des hommes aux JO d'été : {taille_moyenne_homme_ete:.2f} cm")
```

```

## Moyenne de la taille des hommes aux JO d'été : 178.90 cm

print(f"Moyenne de la taille des hommes aux JO d'hiver : {taille_moyenne_homme_hiver:.2f} cm")

## Moyenne de la taille des hommes aux JO d'hiver : 178.67 cm

# Taille maximum des hommes pour les JO d'été et d'hiver
taille_max_homme_ete = df_ete[df_ete['Sex'] == 'M']['Height'].max()
taille_max_homme_hiver = df_hiver[df_hiver['Sex'] == 'M']['Height'].max()

print(f"Taille maximum des hommes aux JO d'été : {taille_max_homme_ete} cm")

## Taille maximum des hommes aux JO d'été : 226.0 cm

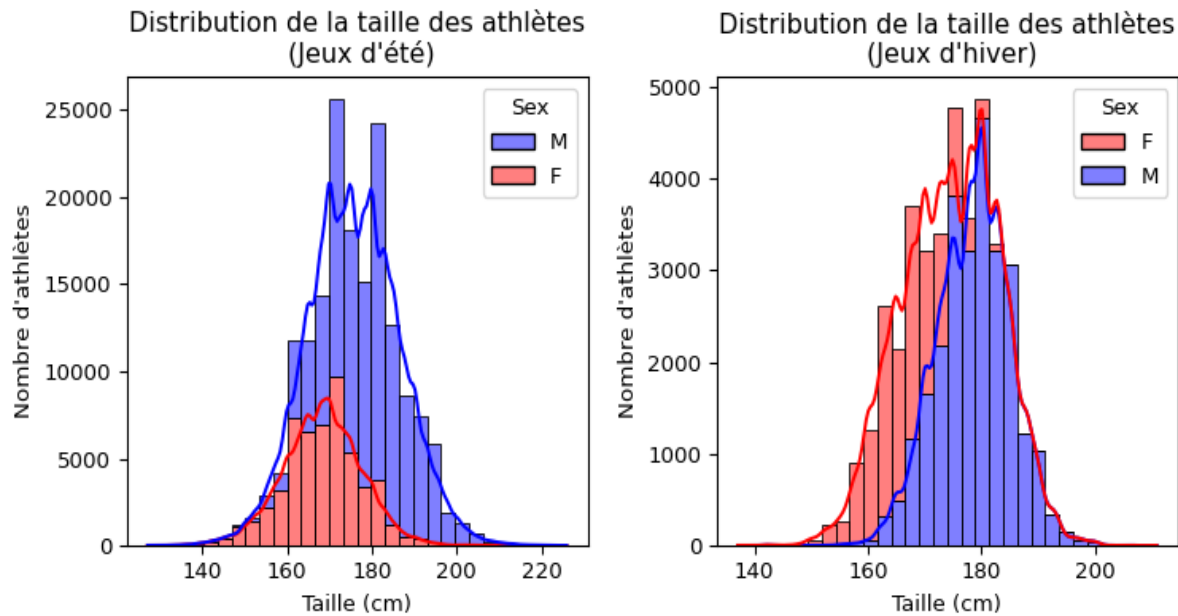
print(f"Taille maximum des hommes aux JO d'hiver : {taille_max_homme_hiver} cm")

## Taille maximum des hommes aux JO d'hiver : 211.0 cm

plt.figure(figsize=(8, 4))
# Distribution de la taille pour les Jeux d'été
plt.subplot(1, 2, 1)
sns.histplot(data=df_ete, x='Height', hue='Sex', multiple='stack', bins=30, kde=True, palette={'M': 'blue', 'F': 'red'})
plt.title("Distribution de la taille des athlètes\n (Jeux d'été)")
plt.xlabel("Taille (cm)")
plt.ylabel("Nombre d'athlètes")

# Distribution de la taille pour les Jeux d'hiver
plt.subplot(1, 2, 2)
sns.histplot(data=df_hiver, x='Height', hue='Sex', multiple='stack', bins=30, kde=True, palette={'M': 'blue', 'F': 'red'})
plt.title("Distribution de la taille des athlètes\n (Jeux d'hiver)")
plt.xlabel("Taille (cm)")
plt.ylabel("Nombre d'athlètes")
plt.subplots_adjust(wspace=20)
plt.tight_layout()
plt.show()

```



À gauche, le graphique représente la taille des athlètes aux Jeux olympiques d'été, tandis qu'à droite, nous observons la distribution pour les Jeux olympiques d'hiver. La juxtaposition de ces deux graphiques permet une comparaison visuelle entre les deux saisons olympiques. Une observation intéressante réside dans la légère diminution de la taille moyenne des femmes aux Jeux d'hiver par rapport aux Jeux d'été.

En analysant de plus près, nous constatons que la moyenne de la taille des femmes pour les Jeux d'été est de 168,17 cm, tandis qu'elle diminue légèrement à 166,53 cm pour les Jeux d'hiver. Pour les hommes, la moyenne de la taille semble rester relativement constante entre les deux saisons. La moyenne de la taille des hommes reste relativement constante, avec 178,90 cm pour les Jeux d'été et 178,67 cm pour les Jeux d'hiver. Cette analyse offre un éclairage intéressant sur les variations de taille en fonction des saisons olympiques et soulève des questions intrigantes sur l'influence des disciplines sportives spécifiques sur les caractéristiques physiques des athlètes.

Notons cependant que la taille maximale atteint 226,0 cm aux Jeux d'été, tandis qu'elle est de 211,0 cm aux Jeux d'hiver. Cette variation dans les tailles maximales peut s'expliquer par les types d'épreuves olympiques. Les sports d'été peuvent inclure des disciplines nécessitant une plus grande envergure physique, tandis que les sports d'hiver peuvent favoriser des compétences spécifiques, éventuellement avec moins d'accent sur la taille.

#### Poids/Taille/Genre selon la saison

*#corrélation entre la taille et le poids par saison et par sexe*

```
correlation_by_season_sex = df.groupby(['Season', 'Sex'])[['Height', 'Weight']]
    .corr()
```

```
print("Corrélation entre la taille et le poids par saison et par sexe :")
```

```
## Corrélation entre la taille et le poids par saison et par sexe :
```

```

print(correlation_by_season_sex)

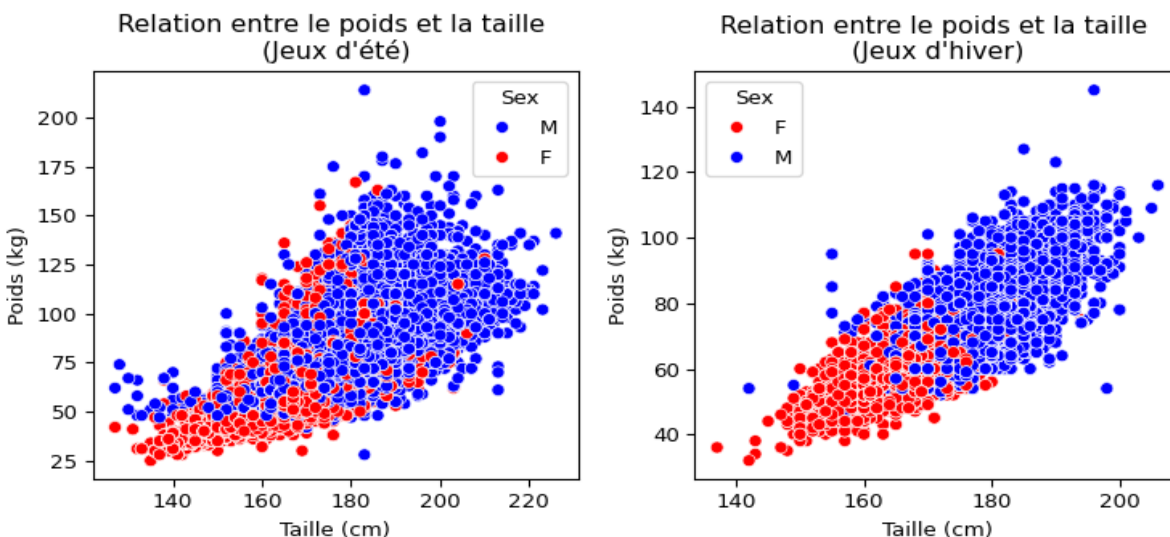
##                Height    Weight
## Season Sex
## Summer F    Height  1.000000  0.748157
##           Weight  0.748157  1.000000
##           M    Height  1.000000  0.734233
##           Weight  0.734233  1.000000
## Winter F    Height  1.000000  0.675347
##           Weight  0.675347  1.000000
##           M    Height  1.000000  0.672621
##           Weight  0.672621  1.000000

plt.figure(figsize=(8, 4))

# Relation entre le poids et la taille pour Les Jeux d'été
plt.subplot(1, 2, 1)
sns.scatterplot(data=df_ete, x='Height', y='Weight', hue='Sex', palette={'M':
'blue', 'F': 'red'})
plt.title("Relation entre le poids et la taille \n(Jeux d'été)")
plt.xlabel("Taille (cm)")
plt.ylabel("Poids (kg)")

# Relation entre le poids et la taille pour Les Jeux d'hiver
plt.subplot(1, 2, 2)
sns.scatterplot(data=df_hiver, x='Height', y='Weight', hue='Sex', palette={'M':
': 'blue', 'F': 'red'})
plt.title("Relation entre le poids et la taille \n(Jeux d'hiver)")
plt.xlabel("Taille (cm)")
plt.ylabel("Poids (kg)")
plt.subplots_adjust(wspace=20)
plt.tight_layout()
plt.show()

```





Les deux graphiques représentent la relation entre le poids et la taille par saison. Comme nous l'avons observé précédemment, il n'y a pas de légères différences.

Une tendance intéressante se dégage : les points situés dans la partie inférieure gauche du graphique, où la taille et le poids sont plus petits, sont majoritairement associés aux femmes. En revanche, les points dans la partie supérieure droite, où la taille et le poids sont plus importants, sont principalement liés aux hommes.

Cette observation suggère une corrélation entre la taille, le poids et le sexe des athlètes. Pour les femmes aux Jeux olympiques d'été, la corrélation entre la taille et le poids est de 0,748, indiquant une relation positive modérée. En revanche, aux Jeux olympiques d'hiver, cette corrélation est légèrement inférieure, avec une valeur de 0,675. Cela suggère que, pour les femmes, la corrélation entre la taille et le poids est un peu moins marquée aux Jeux d'hiver par rapport aux Jeux d'été.

Pour les hommes, une tendance similaire est observée. Aux Jeux olympiques d'été, la corrélation entre la taille et le poids est de 0,734, tandis qu'aux Jeux olympiques d'hiver, elle est légèrement inférieure, avec une valeur de 0,673. De manière similaire aux femmes, cela indique que la corrélation entre la taille et le poids est légèrement moins prononcée aux Jeux d'hiver pour les hommes par rapport aux Jeux d'été.

L'analyse de ce nuage de points offre ainsi une perspective visuelle sur la répartition des athlètes en fonction de la taille, du poids et du sexe, mettant en évidence des tendances significatives et étayées par des mesures de corrélation.

## Médailles remportées au total

### Global

```
medailles_totales = df['Medal'].count()
print(f"Nombre total de médailles remportées : {medailles_totales}")

## Nombre total de médailles remportées : 39783

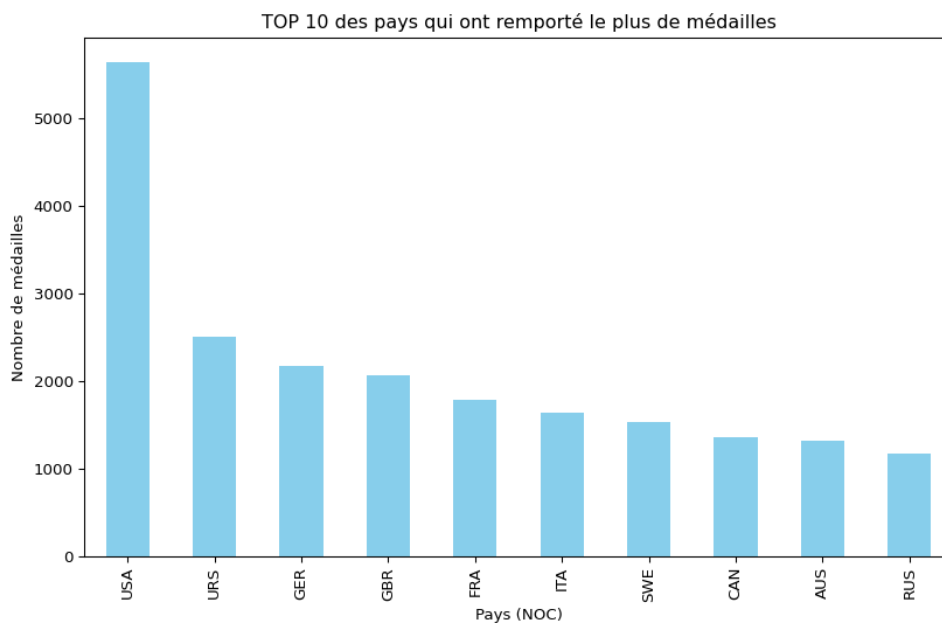
#TOP 10 des pays qui ont remportés le plus de médailles
df_medailles = df[df['Medal'].notnull()]
top_10_pays = df_medailles['NOC'].value_counts().head(10)
print(f"TOP 10 des pays qui ont remporté le plus de médailles : {top_10_pays}")

## TOP 10 des pays qui ont remporté le plus de médailles : NOC
## USA      5637
## URS      2503
## GER      2165
## GBR      2068
## FRA      1777
## ITA      1637
## SWE      1536
## CAN      1352
```



```
## AUS    1320
## RUS    1165
## Name: count, dtype: int64

#graphique
plt.figure(figsize=(10, 6))
top_10_pays.plot(kind='bar', color='skyblue')
plt.title("TOP 10 des pays qui ont remporté le plus de médailles")
plt.xlabel("Pays (NOC)")
plt.ylabel("Nombre de médailles")
plt.show()
```



Le diagramme en barres présenté illustre le palmarès des dix pays ayant remporté le plus grand nombre de médailles, en prenant en compte le total des médailles accumulées par tous les athlètes les composant. Il est crucial de noter que ces données ne reflètent pas simplement le nombre distinct de médailles, car un sport d'équipe sur une même épreuve la même année peut générer plusieurs médailles.

Sans conteste, les États-Unis se distinguent comme le pays le plus médaillé, avec une impressionnante collection dépassant les 5000 médailles. Le trio de tête est constitué des États-Unis, de la Russie (URS), en dépit de son évolution historique, et l'Allemagne (GER), se classent également parmi les nations les plus décorées sur la scène olympique.

La France, quant à elle, se positionne à la cinquième place du classement, totalisant environ 1800 médailles (précisément 1777 médailles). Il est important de noter que ces chiffres englobent l'ensemble des disciplines et des éditions des Jeux olympiques, mettant en évidence la performance globale des athlètes français au fil des années.

## Comparaison entre Été et Hiver

Il serait particulièrement intéressant d'explorer le top 10 des pays les plus performants aux Jeux Olympiques d'été et d'hiver, afin de mettre en lumière d'éventuelles différences entre les deux événements.

```
#Été

nb_medailles_or_ete = df_ete[df_ete['Medal'] == 'Gold'].shape[0]
#Nombre total de médailles remportées
medailles_totales_ete = df_ete['Medal'].count()
print(f"Nombre total de médailles remportées : {medailles_totales}")

## Nombre total de médailles remportées : 39783

#TOP 10 des pays qui ont remportés le plus de médailles
df_medailles_ete = df_ete[df_ete['Medal'].notnull()]
top_10_pays_ete = df_medailles_ete['NOC'].value_counts().head(10)
print(f"TOP 10 des pays qui ont remporté le plus de médailles : {top_10_pays}")

## TOP 10 des pays qui ont remporté le plus de médailles : NOC
## USA      5637
## URS      2503
## GER      2165
## GBR      2068
## FRA      1777
## ITA      1637
## SWE      1536
## CAN      1352
## AUS      1320
## RUS      1165
## Name: count, dtype: int64

##Hiver

#Nombre total de médailles remportées
medailles_totales_hiver = df_hiver['Medal'].count()
print(f"Nombre total de médailles remportées : {medailles_totales}")

## Nombre total de médailles remportées : 39783

#TOP 10 des pays qui ont remportés le plus de médailles
df_medailles_hiver = df_hiver[df_hiver['Medal'].notnull()]
top_10_pays_hiver = df_medailles_hiver['NOC'].value_counts().head(10)
print(f"TOP 10 des pays qui ont remporté le plus de médailles : {top_10_pays}")

## TOP 10 des pays qui ont remporté le plus de médailles : NOC
## USA      5637
## URS      2503
## GER      2165
```

```

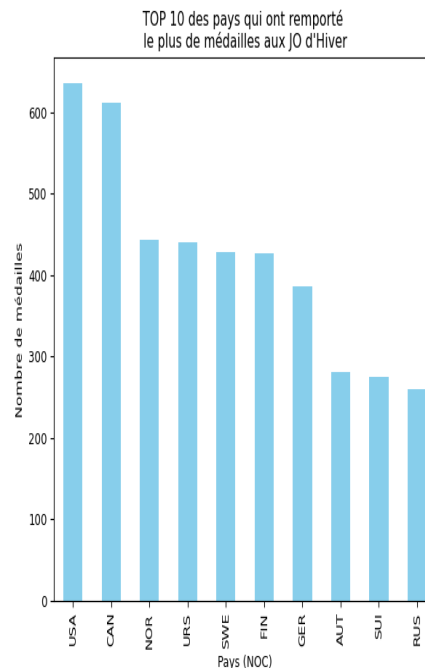
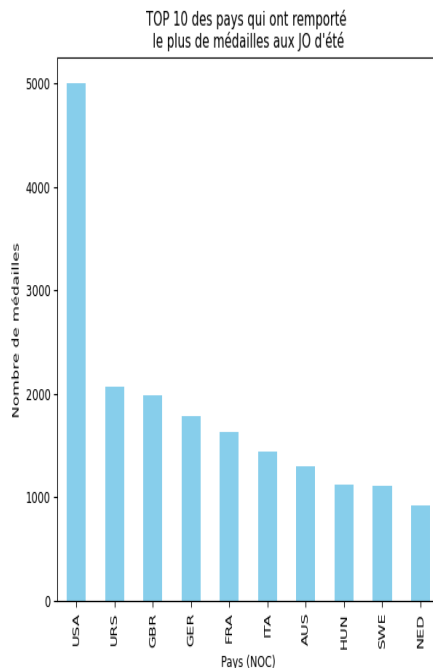
## GBR    2068
## FRA    1777
## ITA    1637
## SWE    1536
## CAN    1352
## AUS    1320
## RUS    1165
## Name: count, dtype: int64

plt.figure(figsize=(18, 6))

# Graphique pour L'été
plt.subplot(1, 2, 1)
top_10_pays_ete.plot(kind='bar', color='skyblue')
plt.title("TOP 10 des pays qui ont remporté \nle plus de médailles aux JO d'été")
plt.xlabel("Pays (NOC)")
plt.ylabel("Nombre de médailles")

# Graphique pour L'hiver
plt.subplot(1, 2, 2)
top_10_pays_hiver.plot(kind='bar', color='skyblue')
plt.title("TOP 10 des pays qui ont remporté \nle plus de médailles aux JO d'Hiver")
plt.xlabel("Pays (NOC)")
plt.ylabel("Nombre de médailles")
plt.subplots_adjust(wspace=0.55)
# Afficher la figure
plt.show()

```



Les deux graphiques présentent des diagrammes en barres illustrant le top 10 des pays qui ont remporté le plus de médailles aux Jeux Olympiques d'été et d'hiver. Une observation initiale révèle une disparité significative dans le nombre total de médailles entre les deux événements, avec le meilleur pays aux JO d'hiver ayant un total inférieur au 10e meilleur aux JO d'été.

Un point d'intérêt majeur réside dans les différences marquées au sein du top 3 de chaque classement. Aux Jeux d'été, le trio de tête est occupé par les États-Unis, l'URSS et le Royaume-Uni, tandis qu'aux Jeux d'hiver, ce sont les États-Unis, le Canada et la Norvège qui dominent. Cette variation souligne des performances distinctes entre les nations dans les deux contextes olympiques.

Il est également pertinent de noter une tendance géographique intéressante : les pays nordiques et froids semblent prédominer aux Jeux d'hiver, ce qui pourrait être lié aux conditions climatiques favorables à certaines disciplines hivernales. Cette observation suggère une corrélation entre la localisation géographique des pays et leurs succès respectifs aux Jeux Olympiques d'hiver.

En explorant ces différences, il est possible de dégager des tendances plus profondes et de mieux comprendre les dynamiques qui influent sur les performances sportives aux Jeux Olympiques, qu'ils soient estivaux ou hivernaux.

## Top 10 des pays les plus médaillés

Après avoir examiné le classement des 10 meilleurs pays aux Jeux Olympiques d'été et d'hiver, explorons à présent les médailles individuelles remportées par ces nations. En se penchant sur la répartition des médailles, on peut identifier les pays qui gagne le plus de médailles d'or, d'argent ou de bronze.

### Global

```
df_medailles = df[df['Medal'].notnull()]
top_10_pays = df_medailles['NOC'].value_counts().head(10).index
df_top_10_pays = df_medailles[df_medailles['NOC'].isin(top_10_pays)]

# Convertir les catégories 'Medal' en valeurs numériques
df_top_10_pays['Medal'] = pd.Categorical(df_top_10_pays['Medal'], categories=
['Gold', 'Silver', 'Bronze'], ordered=True)

## <string>:3: SettingWithCopyWarning:
## A value is trying to be set on a copy of a slice from a DataFrame.
## Try using .loc[row_indexer,col_indexer] = value instead
##
## See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

df_top_10_pays['MedalNumeric'] = df_top_10_pays['Medal'].cat.codes
```

```

## <string>:1: SettingWithCopyWarning:
## A value is trying to be set on a copy of a slice from a DataFrame.
## Try using .loc[row_indexer,col_indexer] = value instead
##
## See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

# Créer un tableau croisé des médailles
df_pivot = df_top_10_pays.groupby(['NOC', 'MedalNumeric']).size().unstack(fill_value=0)

# Trier par le total de médailles
df_pivot['Total'] = df_pivot.sum(axis=1)
df_pivot = df_pivot.sort_values(by='Total', ascending=False).drop(columns='Total')

# Définir la palette de couleurs
palette_couleurs = {'Gold': '#FFD700', 'Silver': '#C0C0C0', 'Bronze': '#CD7F32'}

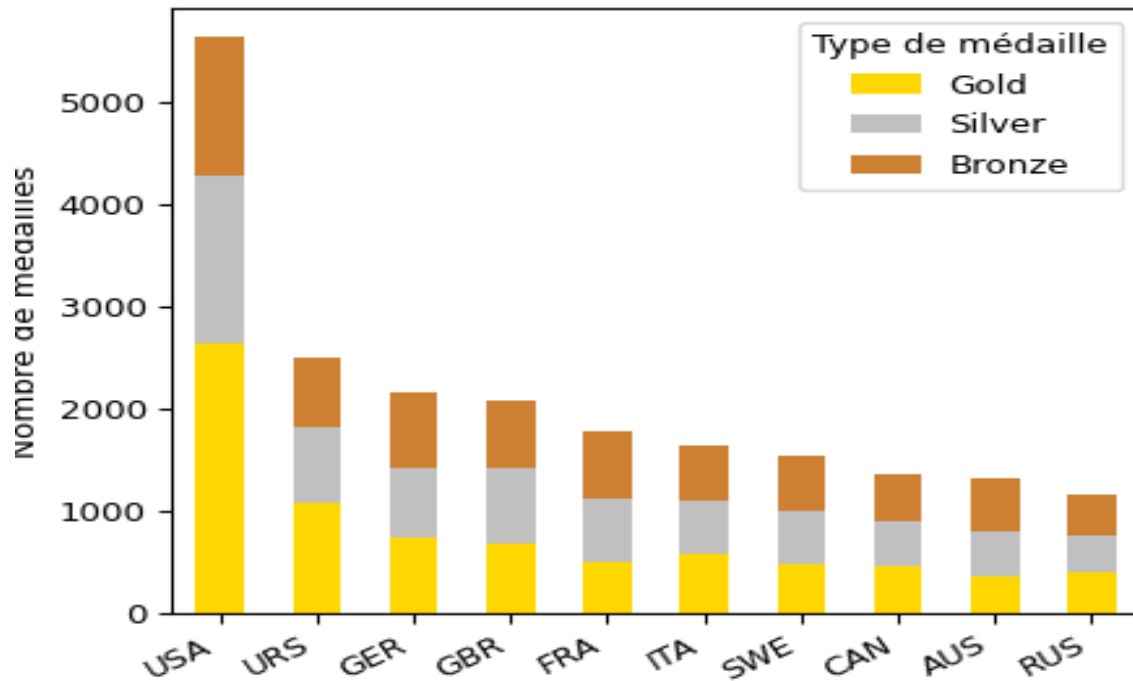
# Créer le graphique en barre empilée
plt.figure(figsize=(12, 8))
df_pivot.plot(kind='bar', stacked=True, color=palette_couleurs.values())
plt.title("Nombre de médailles d'or, d'argent et de bronze pour \nles 10 pays les plus médaillés")
plt.ylabel("Nombre de médailles")
plt.xlabel('')
plt.legend(title='Type de médaille', labels=palette_couleurs.keys())
plt.xticks(rotation=30, ha='right')

## (array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]), [Text(0, 0, 'USA'), Text(1, 0, 'URS'), Text(2, 0, 'GER'), Text(3, 0, 'GBR'), Text(4, 0, 'FRA'), Text(5, 0, 'ITA'), Text(6, 0, 'SWE'), Text(7, 0, 'CAN'), Text(8, 0, 'AUS'), Text(9, 0, 'RUS')])

# Afficher le graphique
plt.show()

```

## Nombre de médailles d'or, d'argent et de bronze pour les 10 pays les plus médaillés



Ce schéma graphique offre une visualisation claire du nombre de médailles d'or, d'argent et de bronze remportées par les 10 pays les plus médaillés dans l'histoire des Jeux Olympiques. Une observation frappante émerge dès le premier coup d'œil : les États-Unis dominent le tableau en termes de médailles d'or, d'argent et de bronze, démontrant leur excellence dans une diversité de disciplines.

Passons maintenant à une analyse comparative entre les Jeux Olympiques d'été et d'hiver, car c'est là que l'étude devient particulièrement intrigante. En explorant les différences entre ces deux éditions, on peut identifier des tendances distinctes et évaluer les performances des nations sous des contextes climatiques et sportifs variés.

### Comparaison entre Été et Hiver

# Été

```
top_10_pays_ete = df_medailles_ete['NOC'].value_counts().head(10).index
df_top_10_pays_ete = df_medailles_ete[df_medailles_ete['NOC'].isin(top_10_pays_ete)]
```

```
df_top_10_pays_ete['Medal'] = pd.Categorical(df_top_10_pays_ete['Medal'], categories=['Gold', 'Silver', 'Bronze'], ordered=True)
```

```
## <string>:2: SettingWithCopyWarning:
```

```
## A value is trying to be set on a copy of a slice from a DataFrame.
```

```
## Try using .loc[row_indexer,col_indexer] = value instead
```

```
##
```

```
## See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```

df_top_10_pays_ete['MedalNumeric'] = df_top_10_pays_ete['Medal'].cat.codes

## <string>:1: SettingWithCopyWarning:
## A value is trying to be set on a copy of a slice from a DataFrame.
## Try using .loc[row_indexer,col_indexer] = value instead
##
## See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

df_pivot_ete = df_top_10_pays_ete.groupby(['NOC', 'MedalNumeric']).size().unstack(fill_value=0)
df_pivot_ete['Total'] = df_pivot_ete.sum(axis=1)
df_pivot_ete = df_pivot_ete.sort_values(by='Total', ascending=False).drop(columns='Total')

# Hiver
top_10_pays_hiver = df_medailles_hiver['NOC'].value_counts().head(10).index
df_top_10_pays_hiver = df_medailles_hiver[df_medailles_hiver['NOC'].isin(top_10_pays_hiver)]

df_top_10_pays_hiver['Medal'] = pd.Categorical(df_top_10_pays_hiver['Medal'], categories=['Gold', 'Silver', 'Bronze'], ordered=True)

## <string>:2: SettingWithCopyWarning:
## A value is trying to be set on a copy of a slice from a DataFrame.
## Try using .loc[row_indexer,col_indexer] = value instead
##
## See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

df_top_10_pays_hiver['MedalNumeric'] = df_top_10_pays_hiver['Medal'].cat.codes

## <string>:1: SettingWithCopyWarning:
## A value is trying to be set on a copy of a slice from a DataFrame.
## Try using .loc[row_indexer,col_indexer] = value instead
##
## See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

df_pivot_hiver = df_top_10_pays_hiver.groupby(['NOC', 'MedalNumeric']).size().unstack(fill_value=0)
df_pivot_hiver['Total'] = df_pivot_hiver.sum(axis=1)
df_pivot_hiver = df_pivot_hiver.sort_values(by='Total', ascending=False).drop(columns='Total')

# Palette de couleurs
palette_couleurs = {'Gold': '#FFD700', 'Silver': '#C0C0C0', 'Bronze': '#CD7F32'}

# Créer une figure avec deux sous-graphiques côte à côte

```

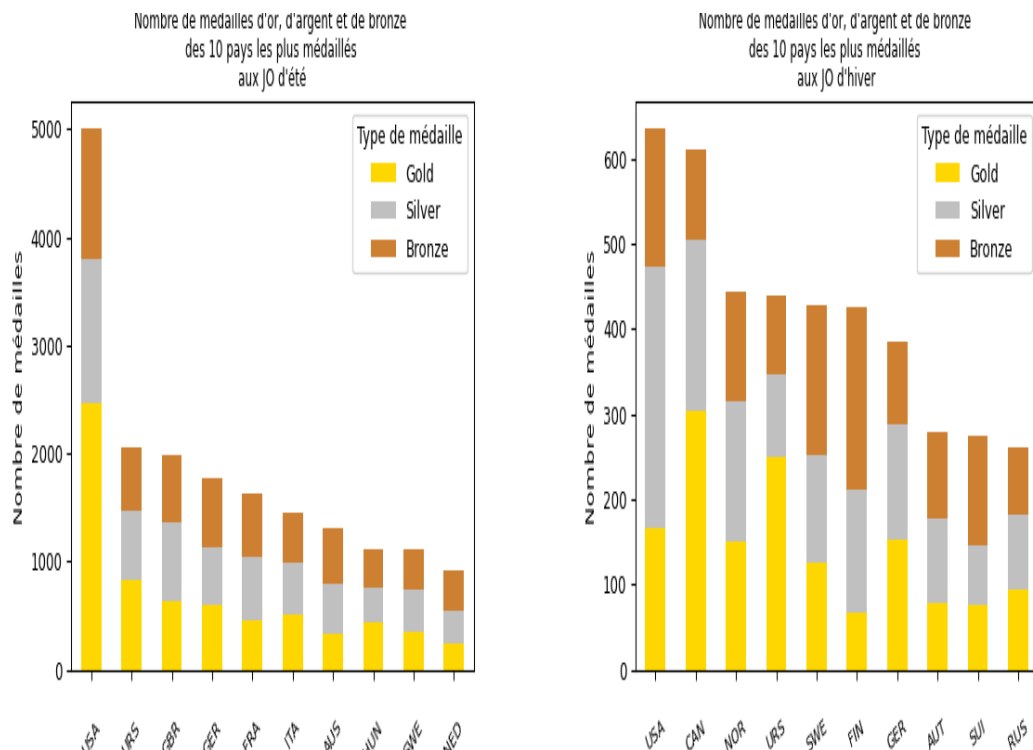
```
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(14, 4))

# Graphique en barre empilée pour L'été
df_pivot_ete.plot(kind='bar', stacked=True, color=palette_couleurs.values(),
ax=ax1)
ax1.set_title("Nombre de médailles d'or, d'argent et de bronze \ndes 10 pays
les plus médaillés \naux JO d'été", fontsize=9)
ax1.set_ylabel("Nombre de médailles")
ax1.legend(title='Type de médaille', labels=palette_couleurs.keys())
ax1.tick_params(axis='x', rotation=45)
ax1.tick_params(axis='x', which='major', pad=15, labels=8)

# Graphique en barre empilée pour L'hiver
df_pivot_hiver.plot(kind='bar', stacked=True, color=palette_couleurs.values()
, ax=ax2)
ax2.set_title("Nombre de médailles d'or, d'argent et de bronze \ndes 10 pays
les plus médaillés \naux JO d'hiver", fontsize=9)
ax2.set_ylabel("Nombre de médailles")
ax2.legend(title='Type de médaille', labels=palette_couleurs.keys())
ax2.tick_params(axis='x', rotation=30)
ax2.tick_params(axis='x', which='major', pad=15, labels=8)

# Ajuster L'espace entre Les deux sous-graphiques
plt.subplots_adjust(wspace=0.4)

# Afficher la figure
plt.show()
```





Le graphique à gauche représente le nombre de médailles d'or, d'argent et de bronze pour les 10 pays les plus médaillés aux Jeux Olympiques d'été, tandis que celui à droite concerne les JO d'hiver. Une première observation saisissante est la similitude du classement pour les JO d'été, confirmant la constance des performances des nations dans ces disciplines diversifiées.

Cependant, le véritable questionnement se trouve dans le graphique des JO d'hiver, où une dynamique différente se dessine. Contrairement aux attentes, le Canada se positionne en tête pour le nombre de médailles d'or, détrônant les États-Unis. Cette variation peut être attribuée à la remarquable performance canadienne dans des sports hivernaux spécifiques.

Ce qui mérite une attention particulière, c'est que la première place des États-Unis dans le classement global des JO d'hiver est largement influencée par leur impressionnante récolte de médailles d'argent.

## TOP 5 des sports les plus médaillés

Après avoir examiné les nations les plus médaillées, concentrons-nous désormais sur les sports qui ont accumulé le plus grand nombre de médailles.

### Global

```
df_medailles = df[df['Medal'].notnull()]
palette_couleurs = {'Gold': '#FFD700', 'Silver': '#C0C0C0', 'Bronze': '#CD7F32'}

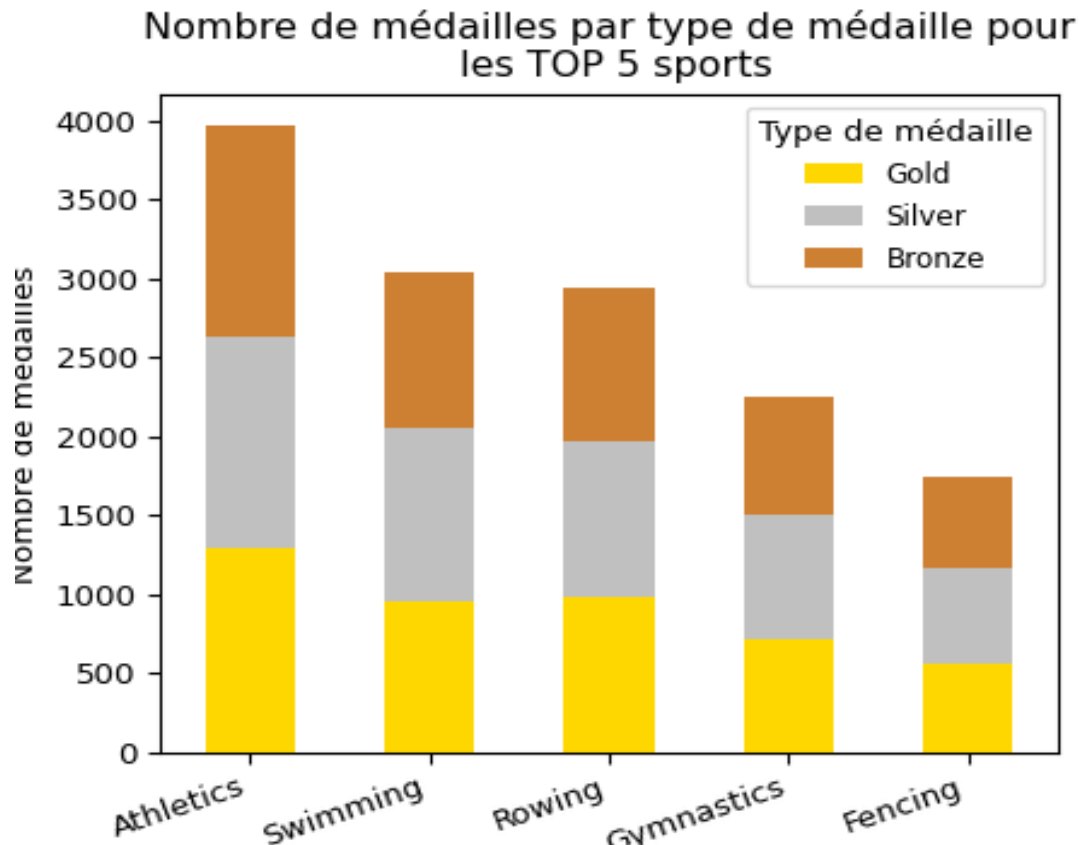
# TOP 5 des sports les plus médaillés (graphique en barre)
top_5_sports = df_medailles['Sport'].value_counts().head(5).index
df_top_5_sports = df_medailles[df_medailles['Sport'].isin(top_5_sports)]
sport_order = df_top_5_sports['Sport'].value_counts().index

plt.figure(figsize=(3, 2))
df_pivot = df_top_5_sports.groupby(['Sport', 'Medal']).size().unstack(fill_value=0)
df_pivot = df_pivot.loc[df_pivot.sum(axis=1).sort_values(ascending=False).index]

df_pivot.plot(kind='bar', stacked=True, color=palette_couleurs.values())
plt.title("Nombre de médailles par type de médaille pour\n les TOP 5 sports")
plt.xlabel("Sport")
plt.ylabel("Nombre de médailles")
plt.legend(title='Type de médaille', fontsize=9, labels=palette_couleurs.keys())
plt.xticks(rotation=20, ha='right') # Retirez le paramètre 'labels'

## (array([0, 1, 2, 3, 4]), [Text(0, 0, 'Athletics'), Text(1, 0, 'Swimming'),
Text(2, 0, 'Rowing'), Text(3, 0, 'Gymnastics'), Text(4, 0, 'Fencing')])

plt.show()
```



D'un point de vue global, les cinq disciplines ayant remporté le plus grand nombre de médailles sont l'athlétisme, la natation, l'aviron, la gymnastique et l'escrime, comme le révèle le diagramme en barres empilées.

L'examen de ce graphique révèle que l'athlétisme se distingue en tant que discipline avec le plus grand nombre de médailles, approchant la barre des 4000 médailles, réparties presque équitablement entre médailles d'or et d'argent. Ce total est suivi de près par la natation et l'aviron, qui affichent des chiffres médians similaires. Il en découle que l'athlétisme est la discipline où les athlètes obtiennent les performances les plus remarquables.

Par ailleurs, nous allons maintenant nous pencher sur une possible disparité dans la distribution du nombre de médailles entre les différentes saisons des Jeux Olympiques.

#### Comparaison entre Été et Hiver

*# Été*

```
top_5_sports_ete = df_medailles_ete['Sport'].value_counts().head(5).index
df_top_5_sports_ete = df_medailles_ete[df_medailles_ete['Sport'].isin(top_5_sports_ete)]
sport_order_ete = df_top_5_sports_ete['Sport'].value_counts().index
```

*# Hiver*

```
top_5_sports_hiver = df_medailles_hiver['Sport'].value_counts().head(5).index
```

```

df_top_5_sports_hiver = df_medailles_hiver[df_medailles_hiver['Sport'].isin(top_5_sports_hiver)]
sport_order_hiver = df_top_5_sports_hiver['Sport'].value_counts().index

palette_couleurs = {'Gold': '#FFD700', 'Silver': '#C0C0C0', 'Bronze': '#CD7F32'}

# Créer une figure avec deux sous-graphiques côte à côte
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(17, 5))

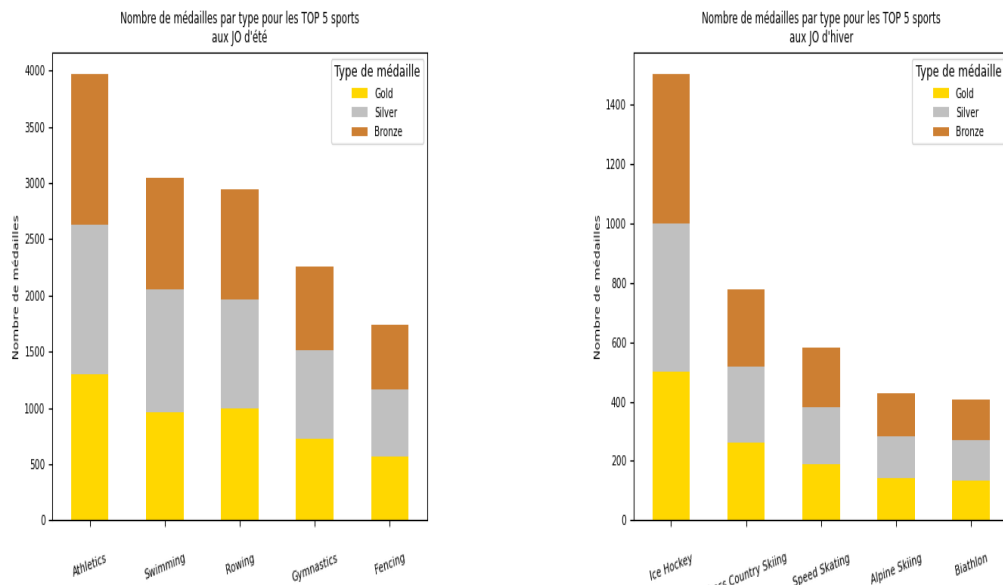
# Graphique été
df_pivot_ete = df_top_5_sports_ete.groupby(['Sport', 'Medal']).size().unstack(fill_value=0)
df_pivot_ete = df_pivot_ete.reindex(sport_order_ete)
df_pivot_ete.plot(kind='bar', stacked=True, color=palette_couleurs.values(), ax=ax1)
ax1.set_title("Nombre de médailles par type pour les TOP 5 sports\nnaux JO d'été", fontsize=9)
ax1.set_xlabel("Sport", fontsize=10) # Augmenter la taille du label de l'axe des x
ax1.set_ylabel("Nombre de médailles", fontsize=8)
ax1.legend(title='Type de médaille', labels=palette_couleurs.keys(), fontsize=8)
ax1.tick_params(axis='x', rotation=15)
ax1.tick_params(axis='x', which='major', pad=15, labelsiz=6)
ax1.tick_params(axis='both', labelsiz=8) # Ajuster la taille de la graduation

# Graphique hiver
df_pivot_hiver = df_top_5_sports_hiver.groupby(['Sport', 'Medal']).size().unstack(fill_value=0)
df_pivot_hiver = df_pivot_hiver.reindex(sport_order_hiver)
df_pivot_hiver.plot(kind='bar', stacked=True, color=palette_couleurs.values(), ax=ax2)
ax2.set_title("Nombre de médailles par type pour les TOP 5 sports\nnaux JO d'hiver", fontsize=9)
ax2.set_xlabel("Sport", fontsize=10) # Augmenter la taille du label de l'axe des x
ax2.set_ylabel("Nombre de médailles", fontsize=8)
ax2.legend(title='Type de médaille', labels=palette_couleurs.keys(), fontsize=8)
ax2.tick_params(axis='x', rotation=15)
ax2.tick_params(axis='x', which='major', pad=15, labelsiz=6)
ax2.tick_params(axis='both', labelsiz=8) # Ajuster la taille de la graduation

# Ajuster l'espace entre les deux sous-graphiques
fig.subplots_adjust(wspace=0.55)

```

```
# Afficher la figure
plt.show()
```



Dans le diagramme de gauche illustrant le Top 5 du nombre de médailles aux Jeux Olympiques d'Été sous forme de barres empilées, nous notons que les mêmes cinq disciplines occupent les premières places en termes de médailles sur l'ensemble de la base de données, avec près de 4000 médailles, ce qui semble cohérent étant donné que plus de 85% des sportifs de la base de données ont participé aux Jeux Olympiques d'Été.

Cependant, sur le graphique de droite représentant le Top 5 du nombre de médailles aux Jeux Olympiques d'Hiver sous forme de barres empilées, la dynamique change. Contrairement aux Jeux d'Été, le hockey sur glace se hisse en tête du classement des Jeux Olympiques d'Hiver avec environ 1500 médailles, suivi du ski de fond avec près de 800 médailles. Ainsi, nous constatons une prédominance de la participation au hockey sur glace aux Jeux Olympiques d'Hiver.

## Pays les plus médaillés dans le top 5 sports

Ensuite, nous avons examiné le nombre de médailles remportées par les pays dans cinq disciplines majeures lors de chaque saison des Jeux Olympiques, là où le nombre de médailles est le plus élevé. L'objectif était d'analyser s'il existe un lien éventuel entre le pays et ces disciplines.

### Été

```
sports_medailles_ete = ['Athletics', 'Swimming', 'Rowing', 'Gymnastics', 'Fencing']
```

```
plt.figure(figsize=(15, 10))
plt.suptitle("TOP 5 des sports aux JO d'été", fontsize=16)
```

```

for i, sport in enumerate(sports_médailles_ete, 1):
    plt.subplot(2, 3, i)
    df_sport = df_médailles[df_médailles['Sport'] == sport]
    df_top_10_pays = df_sport[df_sport['NOC'].isin(top_10_pays.tolist())] #
    Convert to list
    palette_couleurs = {'Gold': '#FFD700', 'Silver': '#C0C0C0', 'Bronze': '#C
D7F32'}
    df_pivot = df_top_10_pays.groupby(['NOC', 'Medal']).size().unstack(fill_v
alue=0)
    df_pivot = df_pivot.loc[df_pivot.sum(axis=1).sort_values(ascending=False)
.index]
    df_pivot.plot(kind='bar', stacked=True, color=palette_couleurs.values(),
ax=plt.gca())
    plt.title(f"{sport}")
    plt.xlabel("Pays (NOC)")
    plt.ylabel("Nombre de médailles")
    plt.legend(title='Type de médaille', labels=palette_couleurs.keys())
    plt.xticks(rotation=45, ha='right')

## <Axes: >
## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Athletics')
## Text(0.5, 0, 'Pays (NOC)')
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251B97B5600>
## (array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]), [Text(0, 0, 'USA'), Text(1, 0, 'GB
R'), Text(2, 0, 'URS'), Text(3, 0, 'GER'), Text(4, 0, 'SWE'), Text(5, 0, 'FRA
'), Text(6, 0, 'RUS'), Text(7, 0, 'CAN'), Text(8, 0, 'AUS'), Text(9, 0, 'ITA
')])
## <Axes: >
## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Swimming')
## Text(0.5, 0, 'Pays (NOC)')
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251B98E7190>
## (array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]), [Text(0, 0, 'USA'), Text(1, 0, 'AU
S'), Text(2, 0, 'GER'), Text(3, 0, 'GBR'), Text(4, 0, 'URS'), Text(5, 0, 'CAN
'), Text(6, 0, 'FRA'), Text(7, 0, 'SWE'), Text(8, 0, 'RUS'), Text(9, 0, 'ITA
')])
## <Axes: >
## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Rowing')
## Text(0.5, 0, 'Pays (NOC)')
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251B995FBB0>
## (array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]), [Text(0, 0, 'USA'), Text(1, 0, 'GB
R'), Text(2, 0, 'GER'), Text(3, 0, 'CAN'), Text(4, 0, 'ITA'), Text(5, 0, 'URS
'), Text(6, 0, 'AUS'), Text(7, 0, 'FRA'), Text(8, 0, 'RUS'), Text(9, 0, 'SWE
')])
## <Axes: >

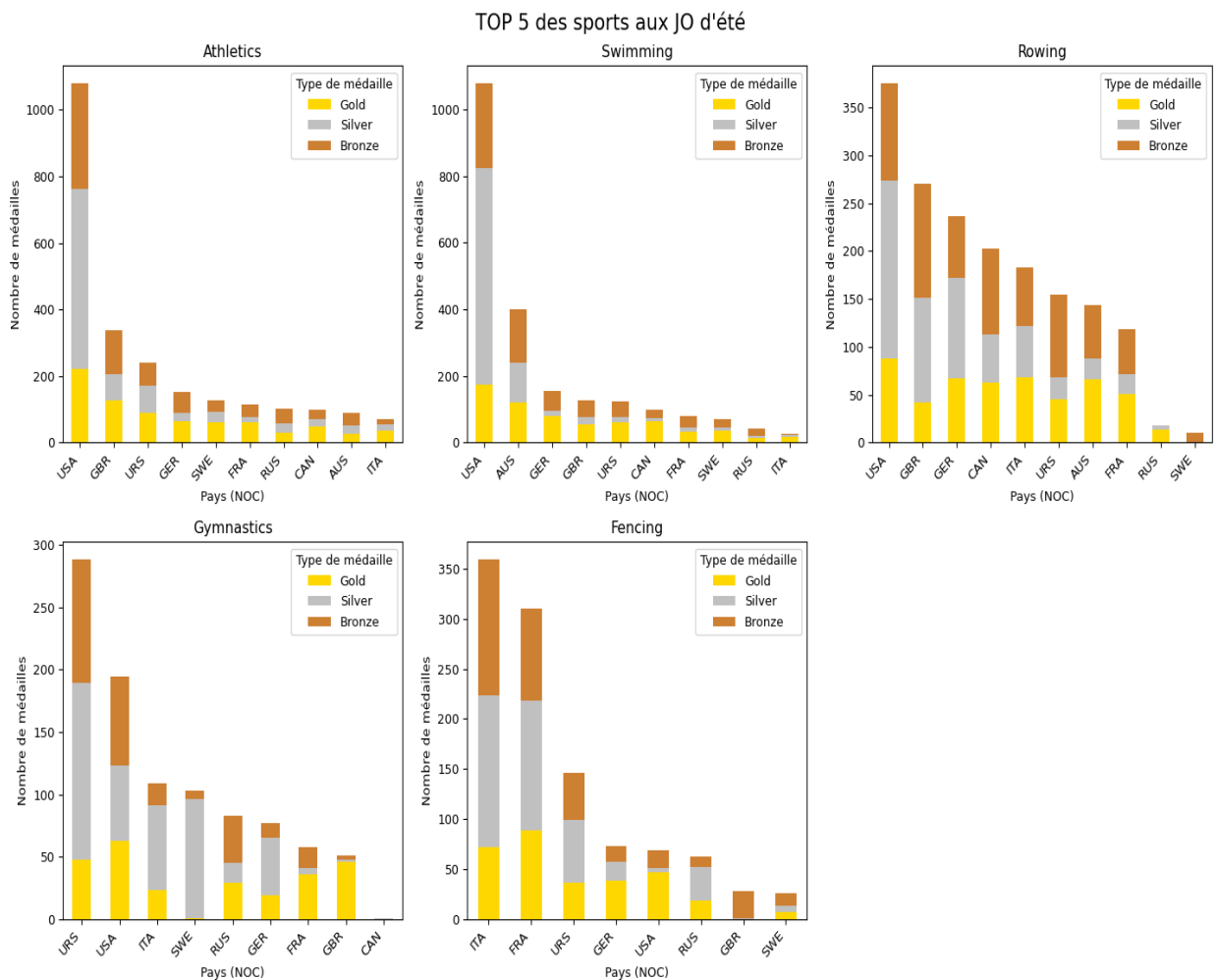
```

```

## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Gymnastics')
## Text(0.5, 0, 'Pays (NOC)')
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251B98E7E50>
## (array([0, 1, 2, 3, 4, 5, 6, 7, 8]), [Text(0, 0, 'URS'), Text(1, 0, 'USA'),
Text(2, 0, 'ITA'), Text(3, 0, 'SWE'), Text(4, 0, 'RUS'), Text(5, 0, 'GER'),
Text(6, 0, 'FRA'), Text(7, 0, 'GBR'), Text(8, 0, 'CAN')])
## <Axes: >
## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Fencing')
## Text(0.5, 0, 'Pays (NOC)')
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251BBC451B0>
## (array([0, 1, 2, 3, 4, 5, 6, 7]), [Text(0, 0, 'ITA'), Text(1, 0, 'FRA'),
Text(2, 0, 'URS'), Text(3, 0, 'GER'), Text(4, 0, 'USA'), Text(5, 0, 'RUS'),
Text(6, 0, 'GBR'), Text(7, 0, 'SWE')])

plt.tight_layout()
plt.show()

```



Sur ce graphique représentant le classement des pays en été pour leur nombre de médailles dans les cinq disciplines les plus médaillées sous la forme d'un diagramme empilé, les États-Unis dominent l'athlétisme, la natation et l'aviron en remportant le plus grand nombre de médailles. Cependant, en gymnastique, c'est la Russie qui prend la tête avec près de 300 médailles dans cette discipline, suivie des États-Unis. Dans l'escrime, c'est l'Italie qui se distingue avec environ 400 médailles, devançant la France. À noter que les États-Unis n'apparaissent même pas dans le classement de cette discipline, contrairement aux autres.

Ainsi, il est possible d'affirmer que lors des Jeux Olympiques d'Été, les États-Unis demeurent les leaders dans les disciplines les plus médaillées, démontrant leur excellence dans une diversité de disciplines.

### Hiver

```
sports_medaillés_hiver = ['Ice Hockey', 'Cross Country Skiing', 'Speed Skating', 'Alpine Skiing', 'Biathlon']

plt.figure(figsize=(15, 10))
plt.suptitle("TOP 5 des sports aux JO d'hiver", fontsize=16)
for i, sport in enumerate(sports_medaillés_hiver, 1):
    plt.subplot(2, 3, i)
    df_sport = df_medailles[df_medailles['Sport'] == sport]
    df_top_10_pays = df_sport[df_sport['NOC'].isin(top_10_pays.tolist())] #
    Convert to list
    palette_couleurs = {'Gold': '#FFD700', 'Silver': '#C0C0C0', 'Bronze': '#CD7F32'}
    df_pivot = df_top_10_pays.groupby(['NOC', 'Medal']).size().unstack(fill_value=0)
    df_pivot = df_pivot.loc[df_pivot.sum(axis=1).sort_values(ascending=False).index]
    df_pivot.plot(kind='bar', stacked=True, color=palette_couleurs.values(),
ax=plt.gca())
    plt.title(f"{sport}")
    plt.xlabel("Pays (NOC)")
    plt.ylabel("Nombre de médailles")
    plt.legend(title='Type de médaille', labels=palette_couleurs.keys())
    plt.xticks(rotation=45, ha='right')

## <Axes: >
## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Ice Hockey')
## Text(0.5, 0, 'Pays (NOC)')
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251BAA4CE50>
## (array([0, 1, 2, 3, 4, 5, 6]), [Text(0, 0, 'CAN'), Text(1, 0, 'USA'), Text(2, 0, 'SWE'), Text(3, 0, 'URS'), Text(4, 0, 'RUS'), Text(5, 0, 'GBR'), Text(6, 0, 'GER')])
## <Axes: >
## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Cross Country Skiing')
## Text(0.5, 0, 'Pays (NOC)')
```

```

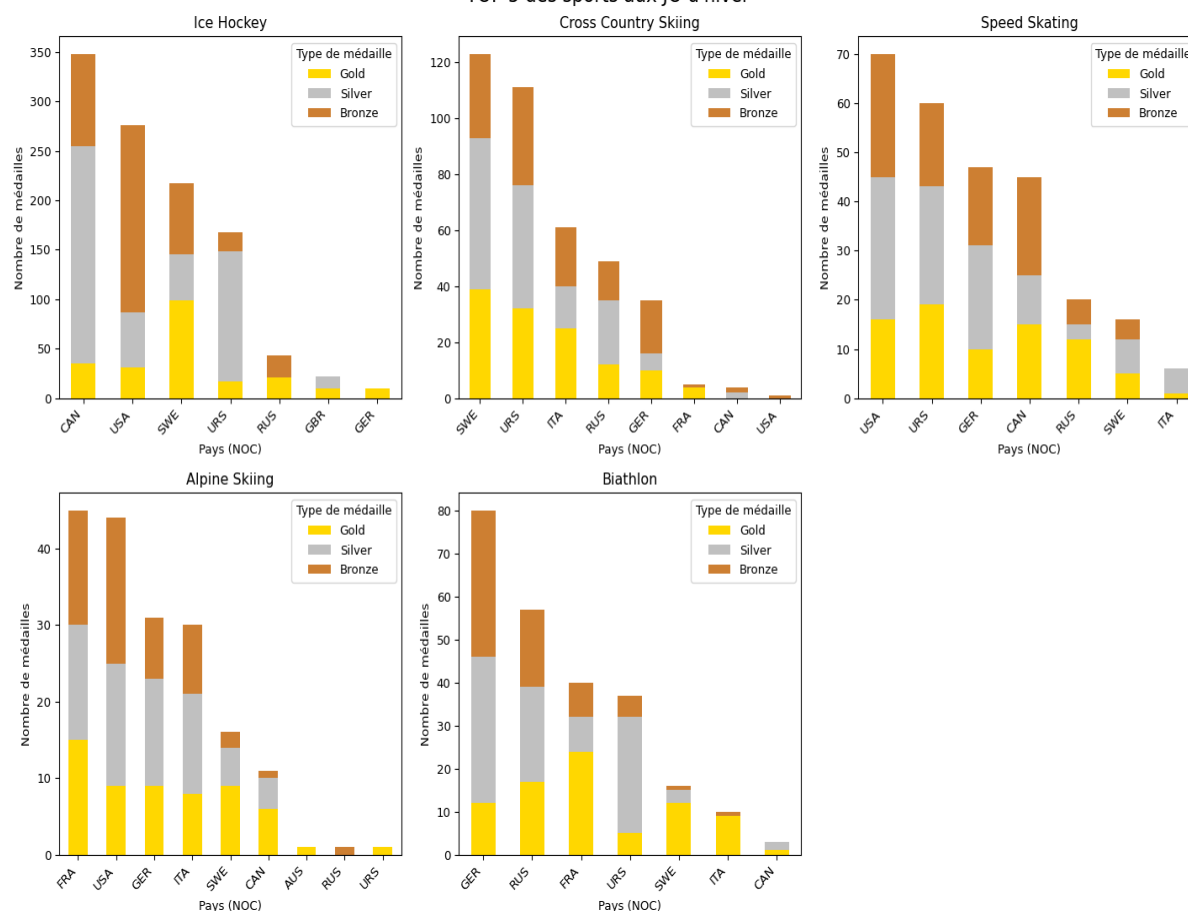
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251BB83D750>
## (array([0, 1, 2, 3, 4, 5, 6, 7]), [Text(0, 0, 'SWE'), Text(1, 0, 'URS'), Text(2, 0, 'ITA'), Text(3, 0, 'RUS'), Text(4, 0, 'GER'), Text(5, 0, 'FRA'), Text(6, 0, 'CAN'), Text(7, 0, 'USA')])
## <Axes: >
## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Speed Skating')
## Text(0.5, 0, 'Pays (NOC)')
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251CB929780>
## (array([0, 1, 2, 3, 4, 5, 6]), [Text(0, 0, 'USA'), Text(1, 0, 'URS'), Text(2, 0, 'GER'), Text(3, 0, 'CAN'), Text(4, 0, 'RUS'), Text(5, 0, 'SWE'), Text(6, 0, 'ITA')])
## <Axes: >
## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Alpine Skiing')
## Text(0.5, 0, 'Pays (NOC)')
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251BAA07FA0>
## (array([0, 1, 2, 3, 4, 5, 6, 7, 8]), [Text(0, 0, 'FRA'), Text(1, 0, 'USA'), Text(2, 0, 'GER'), Text(3, 0, 'ITA'), Text(4, 0, 'SWE'), Text(5, 0, 'CAN'), Text(6, 0, 'AUS'), Text(7, 0, 'RUS'), Text(8, 0, 'URS')])
## <Axes: >
## <Axes: xlabel='NOC'>
## Text(0.5, 1.0, 'Biathlon')
## Text(0.5, 0, 'Pays (NOC)')
## Text(0, 0.5, 'Nombre de médailles')
## <matplotlib.legend.Legend object at 0x00000251B7E8B400>
## (array([0, 1, 2, 3, 4, 5, 6]), [Text(0, 0, 'GER'), Text(1, 0, 'RUS'), Text(2, 0, 'FRA'), Text(3, 0, 'URS'), Text(4, 0, 'SWE'), Text(5, 0, 'ITA'), Text(6, 0, 'CAN')])

plt.tight_layout()
plt.show()

```



### TOP 5 des sports aux JO d'hiver



Dans les cinq sports d'hiver les plus médaillés, le classement des pays diffère de celui observé pendant les Jeux Olympiques d'Été, comme illustré par ce diagramme en barres empilées.

En hockey sur glace, le Canada occupe la première place avec environ 350 médailles, suivi des États-Unis. Pour le ski de fond, c'est la Suisse qui détient le record avec près de 125 médailles, suivie de la Russie. Cependant, dans le patinage de vitesse, ce sont les États-Unis qui dominent en termes de médailles. En ski alpin, c'est la France qui se positionne en tête avec environ 45 médailles. Quant au biathlon, l'Allemagne est le pays le plus médaillé avec 80 médailles.

Les résultats montrent une diversité de pays leaders dans les différentes disciplines, contrairement aux Jeux Olympiques d'Été. Notamment, ces pays leaders sont souvent des pays froids et montagneux tels que le Canada, la Suisse, et la Russie.

### Pays des sportifs les plus médaillés

Après avoir examiné les données concernant les athlètes, notre attention s'est tournée vers les sportifs d'élite, en mettant l'accent sur les 100 meilleurs d'entre eux. Il est important de noter que parmi les plus de 130 000 sportifs distincts, cette focalisation sur les 100 meilleurs vise à simplifier notre analyse.

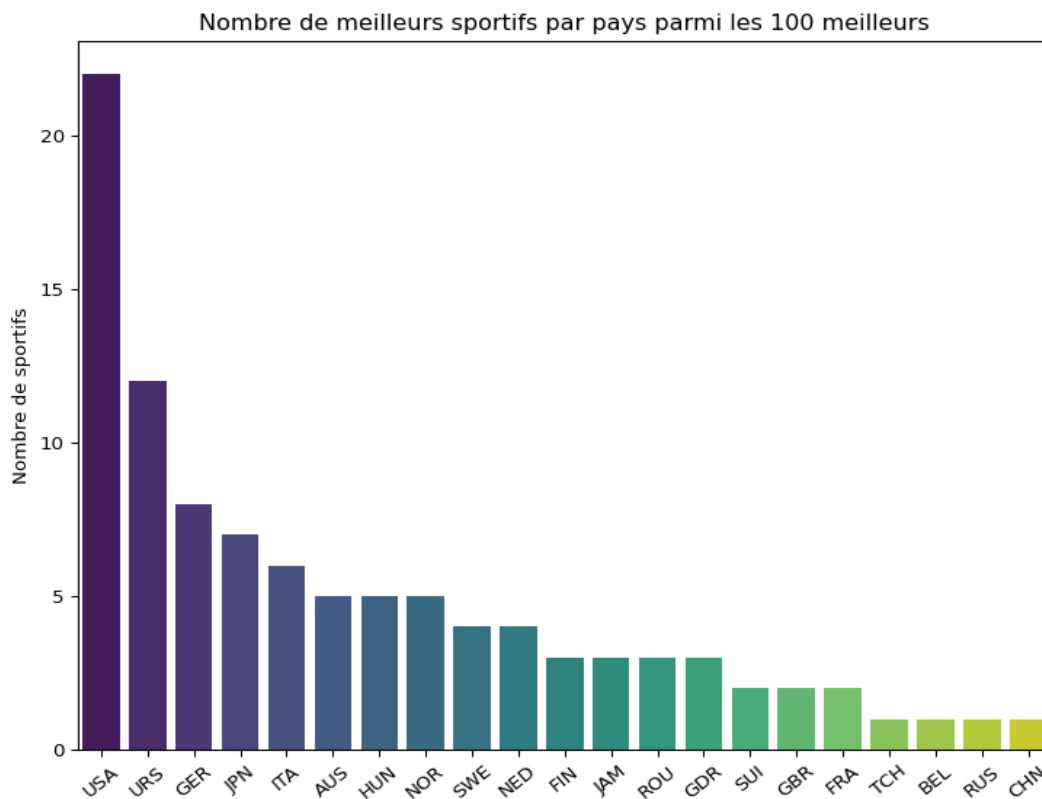
## Global

```
df_sportifs_pays = df_medailles.groupby(['Name', 'NOC'])['Medal'].count().reset_index(name='TotalMedals')
meilleurs_sportifs = df_sportifs_pays.sort_values(by='TotalMedals', ascending=False).head(100)
top_pays = meilleurs_sportifs['NOC'].value_counts()

plt.figure(figsize=(9, 7))
sns.barplot(x=top_pays.index, y=top_pays.values, palette='viridis')
plt.title("Nombre de meilleurs sportifs par pays parmi les 100 meilleurs")
plt.xlabel('')
plt.ylabel("Nombre de sportifs")
plt.xticks(rotation=40)

## ([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20], [Text(0, 0, 'USA'), Text(1, 0, 'URS'), Text(2, 0, 'GER'), Text(3, 0, 'JPN'), Text(4, 0, 'ITA'), Text(5, 0, 'AUS'), Text(6, 0, 'HUN'), Text(7, 0, 'NOR'), Text(8, 0, 'SWE'), Text(9, 0, 'NED'), Text(10, 0, 'FIN'), Text(11, 0, 'JAM'), Text(12, 0, 'ROU'), Text(13, 0, 'GDR'), Text(14, 0, 'SUI'), Text(15, 0, 'GBR'), Text(16, 0, 'FRA'), Text(17, 0, 'TCH'), Text(18, 0, 'BEL'), Text(19, 0, 'RUS'), Text(20, 0, 'CHN')])

plt.show()
```



Ce diagramme à barres illustre la répartition des 100 meilleurs sportifs par pays, mettant en lumière les performances exceptionnelles des États-Unis qui occupent la première position avec plus de 20 athlètes d'élite. La Russie se classe en deuxième position, bien qu'avec un nombre de meilleurs sportifs moindre, leur nombre reste quand même remarquable.

Nous nous tournons maintenant vers la comparaison entre les Jeux Olympiques d'été et d'hiver, cherchant à identifier des différences significatives qui pourraient susciter un intérêt particulier.

### Comparaison entre Été et Hiver

```
#Été
df_sportifs_pays_ete = df_medailles_ete.groupby(['Name', 'NOC'])['Medal'].count().reset_index(name='TotalMedals')

meilleurs_sportifs_ete = df_sportifs_pays_ete.sort_values(by='TotalMedals', ascending=False).head(100)

top_pays_ete = meilleurs_sportifs_ete['NOC'].value_counts()

#Hiver

df_sportifs_pays_hiver = df_medailles_hiver.groupby(['Name', 'NOC'])['Medal'].count().reset_index(name='TotalMedals')

meilleurs_sportifs_hiver = df_sportifs_pays_hiver.sort_values(by='TotalMedals', ascending=False).head(100)

top_pays_hiver = meilleurs_sportifs_hiver['NOC'].value_counts()

# Créer une figure avec deux sous-graphiques côte à côte
plt.figure(figsize=(30, 8))

# Graphique pour Les JO d'été
plt.subplot(1, 2, 1)
sns.barplot(x=top_pays_ete.index, y=top_pays_ete.values, palette='viridis')
plt.title("Nombre de meilleurs sportifs par pays \nparmi les 100 meilleurs au x JO d'été")
plt.xlabel("Pays")
plt.ylabel("Nombre de sportifs")
plt.xticks(rotation=40)

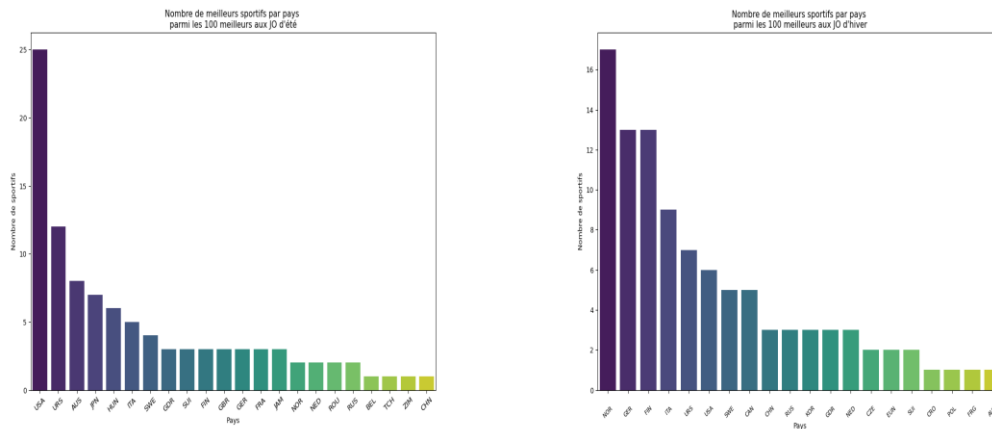
## ([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21], [Text(0, 0, 'USA'), Text(1, 0, 'URS'), Text(2, 0, 'AUS'), Text(3, 0, 'JPN'), Text(4, 0, 'HUN'), Text(5, 0, 'ITA'), Text(6, 0, 'SWE'), Text(7, 0, 'GDR'), Text(8, 0, 'SUI'), Text(9, 0, 'FIN'), Text(10, 0, 'GBR'), Text(11, 0, 'GER'), Text(12, 0, 'FRA'), Text(13, 0, 'JAM'), Text(14, 0, 'NOR'), Text(15, 0, 'NED'), Text(16, 0, 'ROU'), Text(17, 0, 'RUS'), Text(18, 0, 'BEL'), Text(19, 0, 'TCH'), Text(20, 0, 'ZIM'), Text(21, 0, 'CHN')])
```

```
# Graphique pour Les JO d'hiver
plt.subplot(1, 2, 2)
sns.barplot(x=top_pays_hiver.index, y=top_pays_hiver.values, palette='viridis')
plt.title("Nombre de meilleurs sportifs par pays \nparmi les 100 meilleurs au x JO d'hiver")
plt.xlabel("Pays")
plt.ylabel("Nombre de sportifs")
plt.xticks(rotation=40)

## ([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19], [
Text(0, 0, 'NOR'), Text(1, 0, 'GER'), Text(2, 0, 'FIN'), Text(3, 0, 'ITA'), Text(4, 0, 'URS'), Text(5, 0, 'USA'), Text(6, 0, 'SWE'), Text(7, 0, 'CAN'), Text(8, 0, 'CHN'), Text(9, 0, 'RUS'), Text(10, 0, 'KOR'), Text(11, 0, 'GDR'), Text(12, 0, 'NED'), Text(13, 0, 'CZE'), Text(14, 0, 'EUN'), Text(15, 0, 'SUI'), Text(16, 0, 'CRO'), Text(17, 0, 'POL'), Text(18, 0, 'FRG'), Text(19, 0, 'AUT')])

plt.tick_params(axis='x', which='major', pad=15, labelsz=8)
# Ajuster l'espace entre les deux sous-graphiques
plt.subplots_adjust(wspace=0.4)

# Afficher la figure
plt.show()
```



Le graphique à gauche illustre le nombre de meilleurs sportifs par pays parmi les 100 meilleurs aux JO d'été, tandis que celui à droite concerne les JO d'hiver. Pour les JO d'été, les États-Unis maintiennent leur position dominante avec plus de 20 athlètes d'élite. Cependant, aux JO d'hiver, la Norvège prend la première place avec 17 sportifs d'élite, marquant une différence notable. Il est important de noter que, contrairement à leur performance aux JO d'été, les États-Unis se classent 6ème aux JO d'hiver. Cette distinction entre les deux événements souligne l'impact des conditions climatiques, avec une prédominance des nations nordiques dans le classement des JO d'hiver.

## Conclusion

Dans l'exploration minutieuse des données olympiques, notre analyse a dévoilé des facettes intrigantes des performances athlétiques au fil des Jeux d'été et d'hiver. Les nuances significatives entre ces deux événements emblématiques révèlent la complexité des succès sportifs à travers les décennies.

Les États-Unis émergent en tant que titan indiscutable du podium, accumulant une impressionnante collection de plus de 5000 médailles. Toutefois, la dynamique de ce triomphe varie substantiellement entre les Jeux d'été et d'hiver. Là où les rayons du soleil illuminent les exploits estivaux, les Jeux d'hiver voient les nations nordiques s'imposer.

En scrutant la répartition des médailles parmi les nations les plus honorées, les États-Unis se tiennent au sommet dans toutes les catégories aux Jeux d'été. Un pivotement vers les Jeux d'hiver, cependant, révèle une prouesse exceptionnelle du Canada dans la quête des médailles d'or.

Les performances des sportifs d'élite, au sein du cercle restreint des 100 meilleurs, confirment la prédominance continue des États-Unis, suivis de près par la Russie.

Une comparaison saisonnière souligne une corrélation modérée mais significative entre la taille et le poids des athlètes, amplifiée chez les femmes lors des Jeux d'été. Les différences saisonnières ajoutent une nuance captivante, suggérant des influences spécifiques liées aux disciplines pratiquées.

Enfin, il est impératif de souligner que les variations dans les épreuves olympiques entraînent des fluctuations substantielles dans les poids et tailles des athlètes, impactant les relations observées. Cette diversité des épreuves offre une perspective éclairante sur la nature hétérogène des compétitions olympiques.

Dans cette synthèse, émergent des éléments clés : la constance des États-Unis, les différences saisonnières significatives, et la nécessité de considérer la diversité des sports pour une compréhension global des données olympiques.