

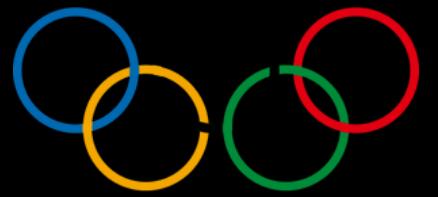


ANALYSE STATISTIQUE DES JO

THP Data analyse - Projet final
Printemps 2023

30 mai 2023





SOMMAIRE

Problématiques

Équipe

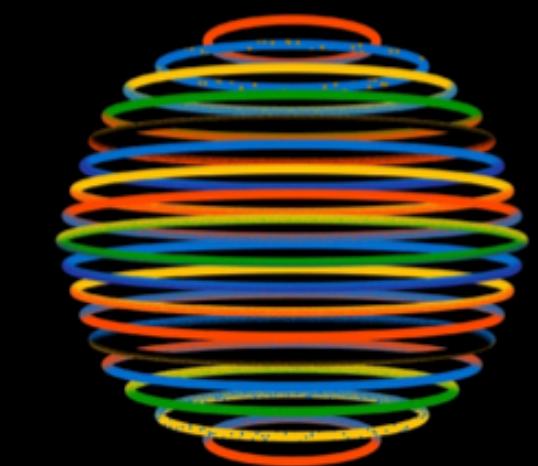
Données

Méthodologie

Cadrage

Timeline

Analyses





PROBLÉMATIQUE

Quels sont les facteurs explicatifs des performances aux JO ?

4 dimensions explorées :

- Tendances temporelles de performance
- Facteurs physique de réussite des athlètes médaillés
- Impact du genre sur la sélection et les performances
- Facteurs macro-sociaux d'explication des performances



ÉQUIPE



David BERTRAND
Phd Data Scientist



Suzanne CHRISTÉ
Data analyste



Jocelyn EPAILLARD
Data analyste



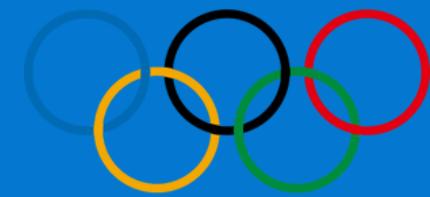
Ruth ETIENNE
Data analyste

" Les performances individuelles, ce n'est pas le plus important. On gagne et on perd en équipe "
-Z. Zidane



DONNÉES

- Athlètes et médailles d'Athènes 1896 à Rio 2016 (Kaggle)
- Gender Inequality Index (Human Development Reports)
- WORLD BANK, "GDP (constant, US \$ 2015), World Bank national accounts data, and OECD National Accounts data files.
- WORLD BANK, "GNI (constant 2015 US\$), World Bank national accounts data, and OECD National Accounts data files.
- WORLD BANK, "Population, total", United Nations Population Division. World Population Prospects: 2022 Revision
- WORLD BANK, "Population ages 65 and above, total", United Nations Population Division's World Population Prospects: 2022 Revision.
- Tokyo 2020 Olympic Summer Games (Kaggle)



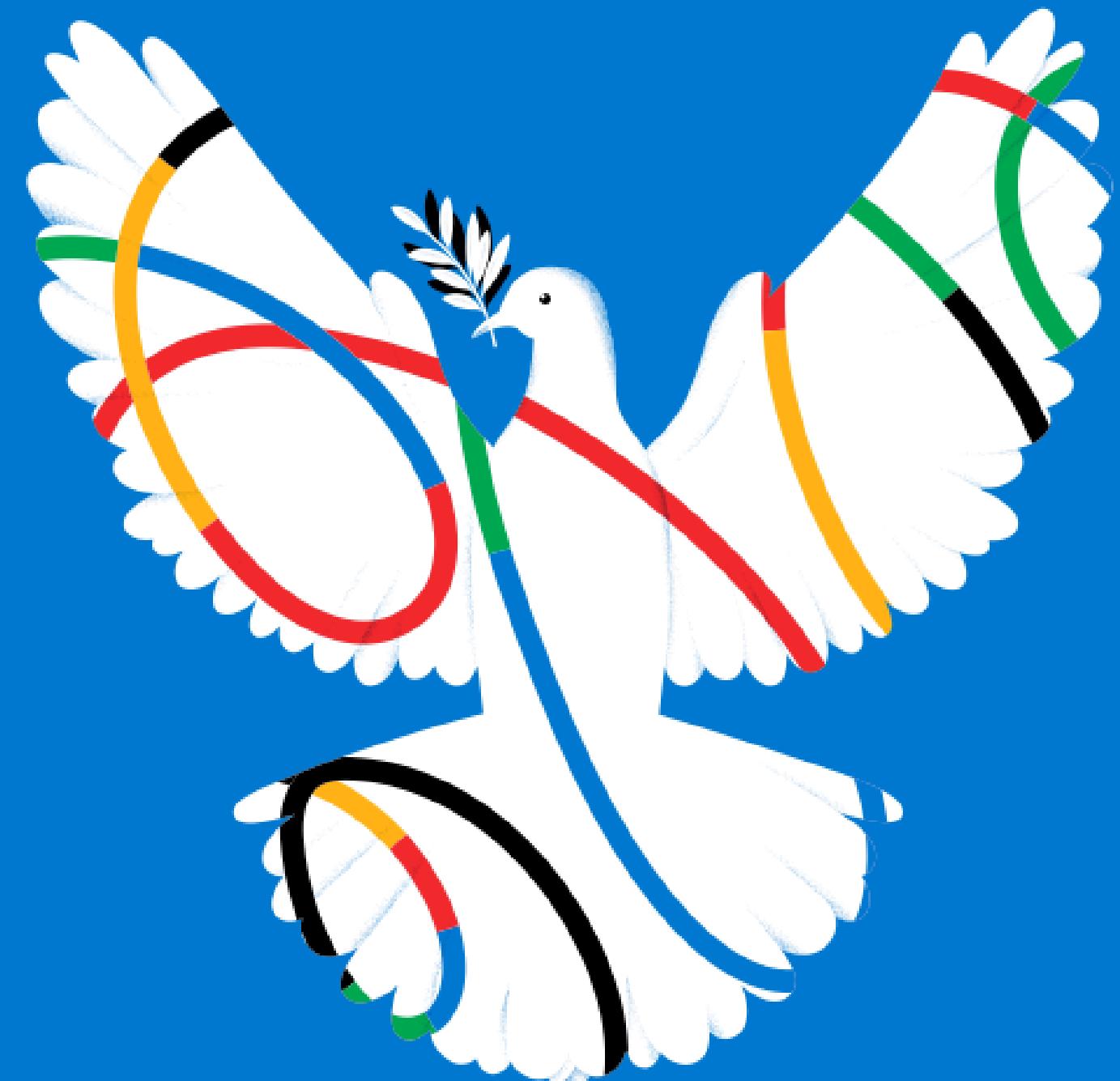
MÉTHODOLOGIE

Gestion de projet

- Agile en scrum

Outils utilisés

- Notions
- Github
- Jupyter Notebook
- Tableau Public
- Canva





CADRAGE

- Période : du 22 au 30 mai 2023
- Limites du projet :
 - Analyse sur les jeux d'été uniquement
 - De 1896 à 2016 (29 éditions)
 - Pas de données complètes pour les jeux de Tokyo 2021
- Livrables : pdf présentation de l'analyse, repository Github, page Notion, graphiques interactifs Tableau Public



TIMELINE

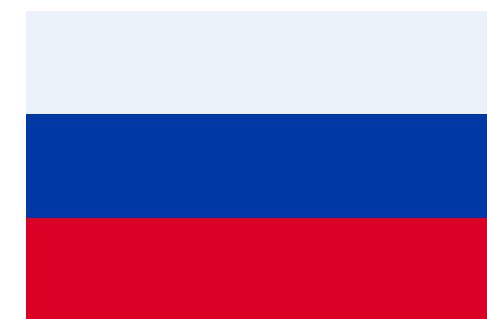
» mai 2023		Semaine ▾ < Aujourd'hui >											
L	22	M	23	J	25	V	26	S	27	D	28	L	29
<input type="checkbox"/> Orienter les analyses pour le storytelling													
			<input type="checkbox"/> Analyse des données										
<input type="checkbox"/> Prévisualiser la timeline													
<input type="checkbox"/> Cadrer le projet													
<input type="checkbox"/> Attribution des analyses MVP													
<input type="checkbox"/> Analyse datasets													
<input type="checkbox"/> Préparation des données													
			<input type="checkbox"/> Rédiger la présentation finale										
<input type="checkbox"/> Organiser le Github (arborescence)													
+ Nouvelle page													





PERFORMANCE GLOBALES

Tous sports confondus



2

RUSSIE

582 médailles d'or
1 514 médailles au total



1

USA

1 067 médailles d'or
2 615 médailles au total



3

ALLEMAGNE

448 médailles d'or
1 380 médailles au total



PERFORMANCE GLOBALES

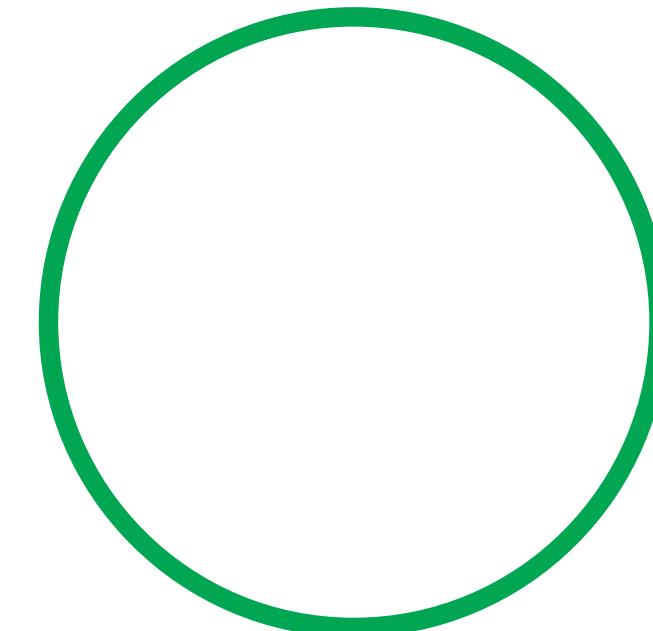
Tous sports confondus



43 épreuves en 1896



339 épreuves en 2021



Création d'un KPI : part de médailles remportée par édition

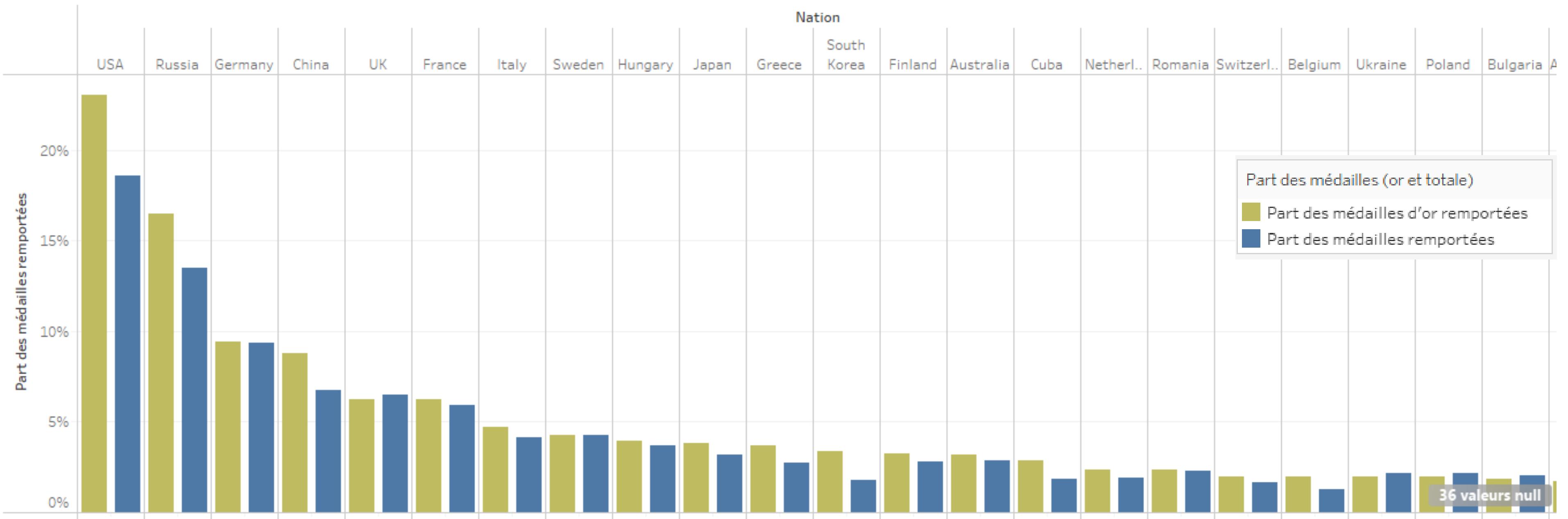


EVOLUTION DES PAYS DOMINANTS

Tous sports confondus



Part des médailles (or et totale) remportées par pays



Data visualisation sur Tableau Public



QUELLES CARACTÉRISTIQUES PHYSIQUES POUR LES MÉDAILLÉS ?

4 sports les plus cités dans les médias français en 2016 :

- Athlétisme
- Natation
- Handball
- Judo



Âge moyen

Taille moyenne

Poids moyen

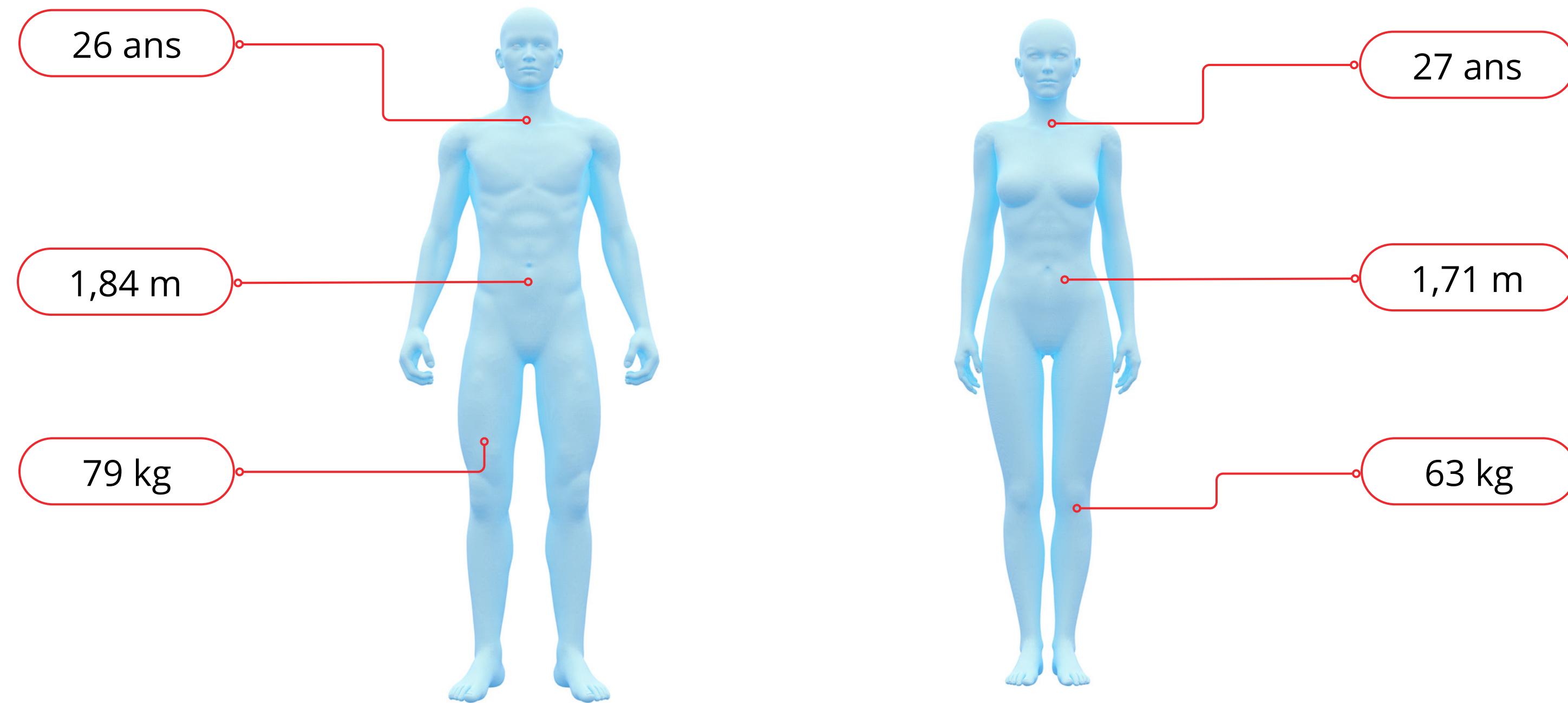
Année 2016





PORTRAIT ROBOT DES MÉDAILLÉS

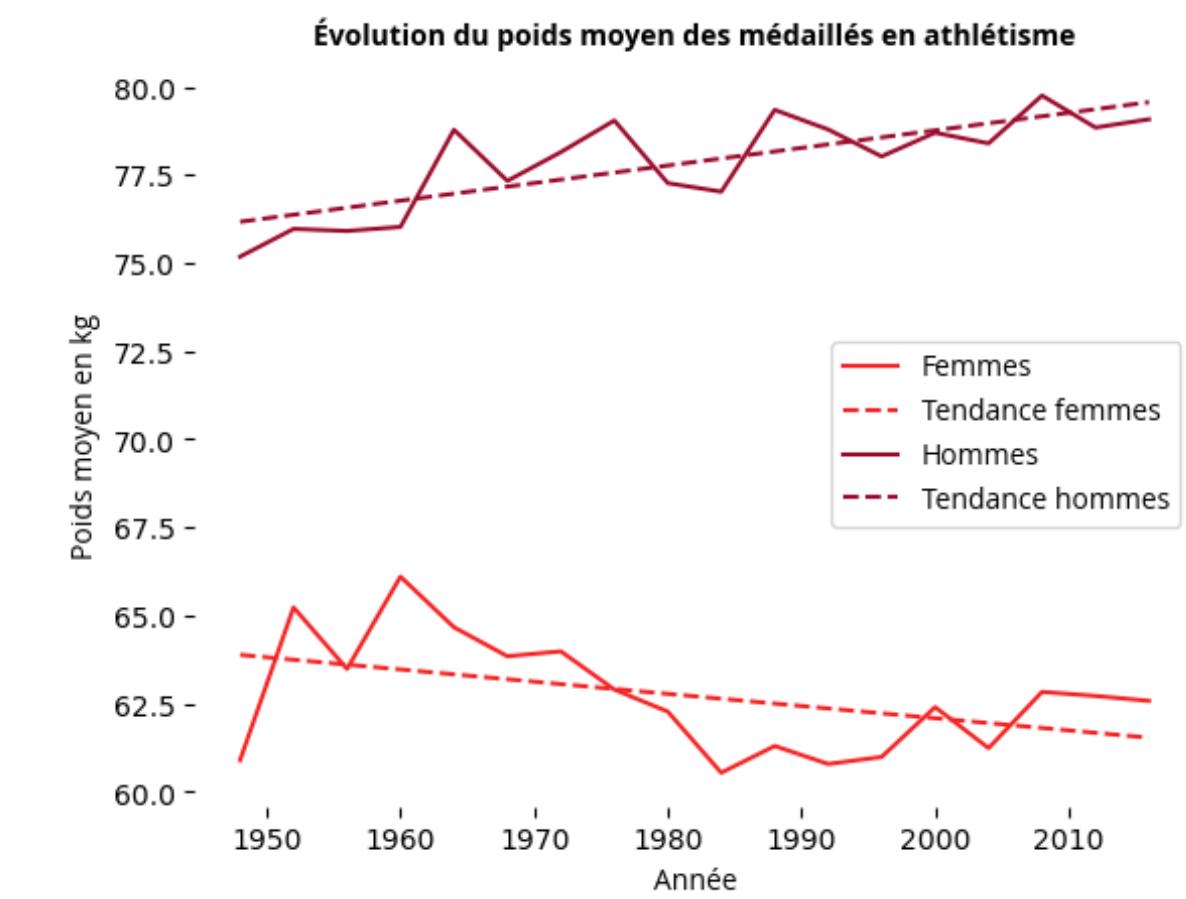
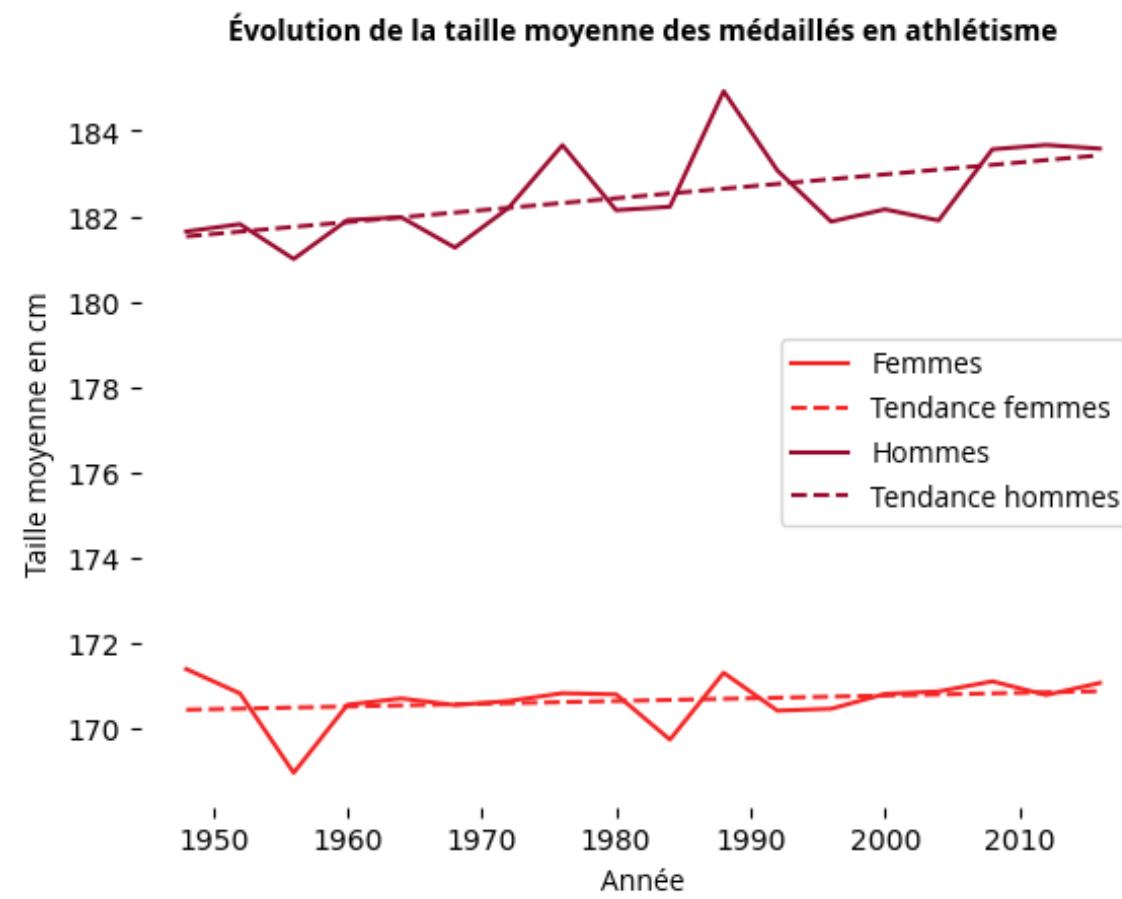
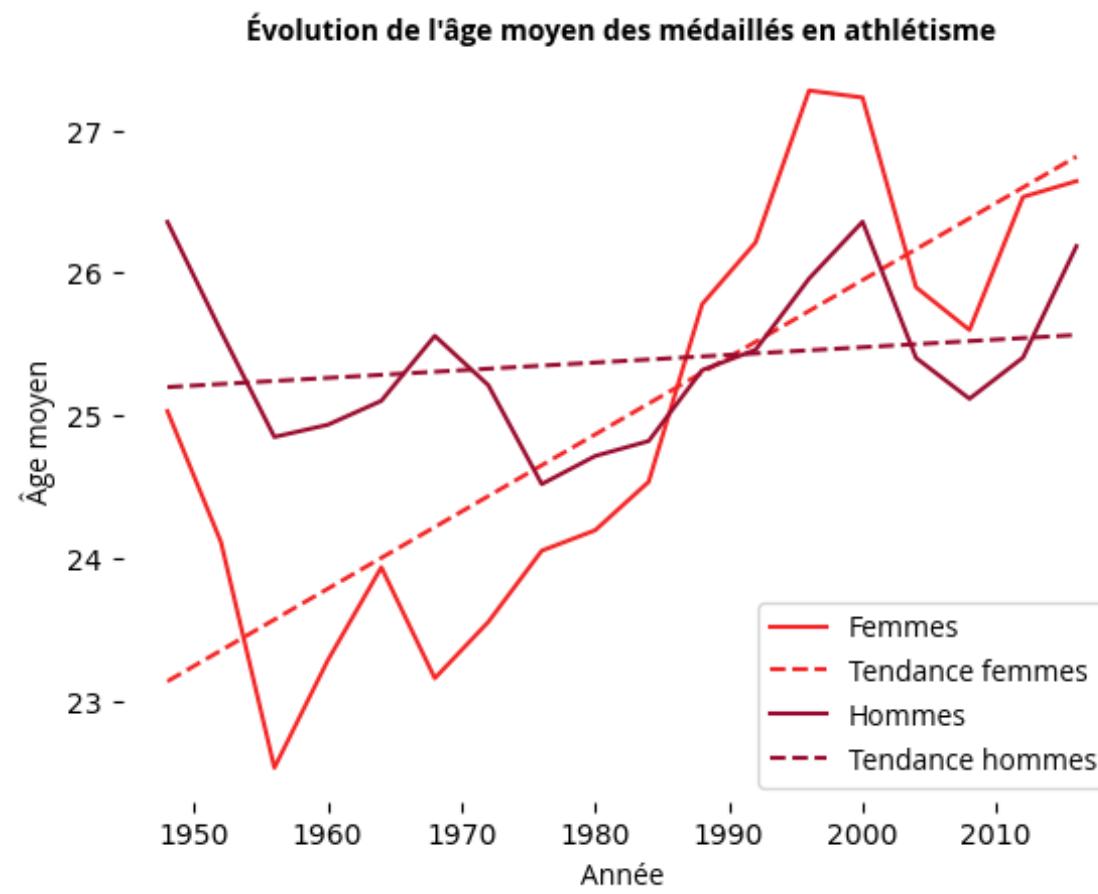
Athlétisme





PORTRAIT ROBOT DES MÉDAILLÉS

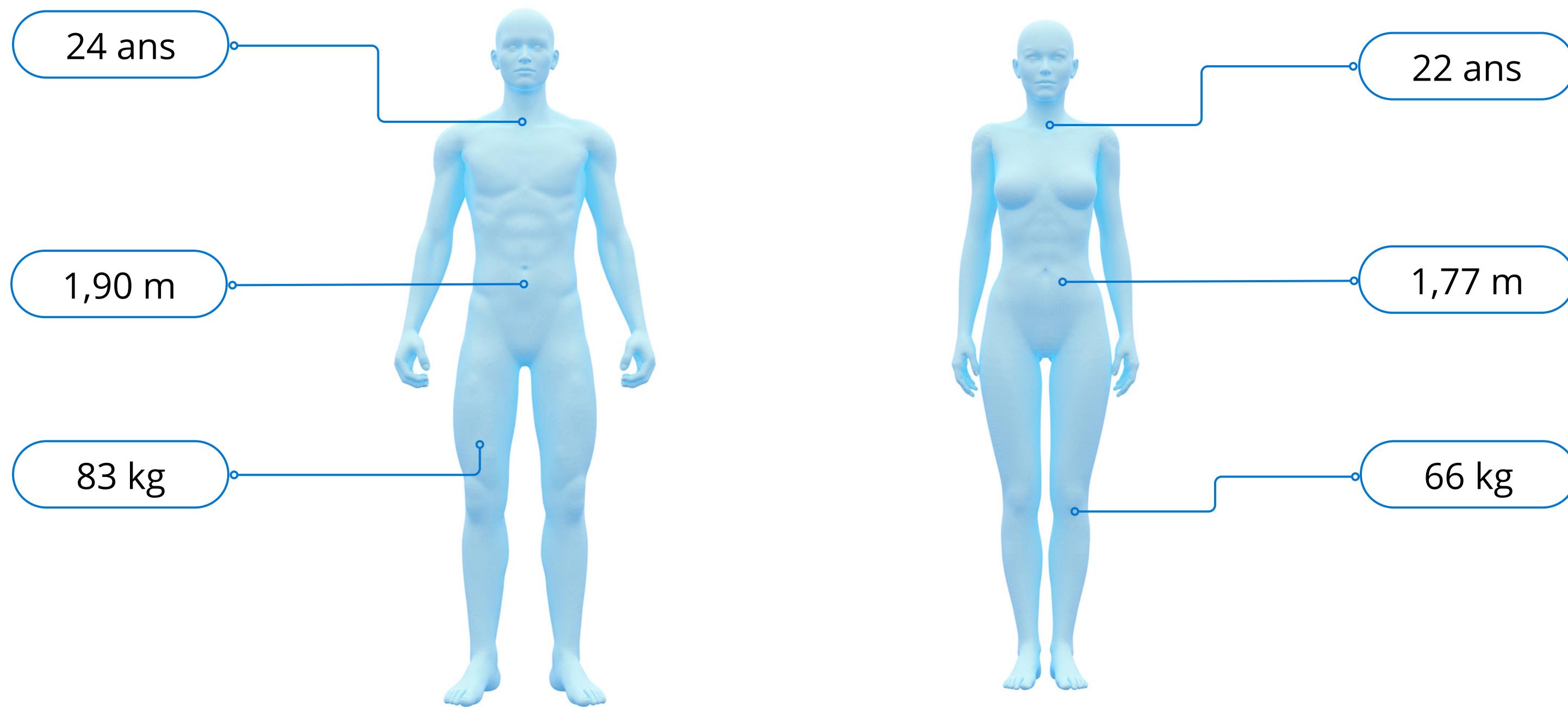
Évolution des caractéristiques physiques en athlétisme





PORTRAIT ROBOT DES MÉDAILLÉS

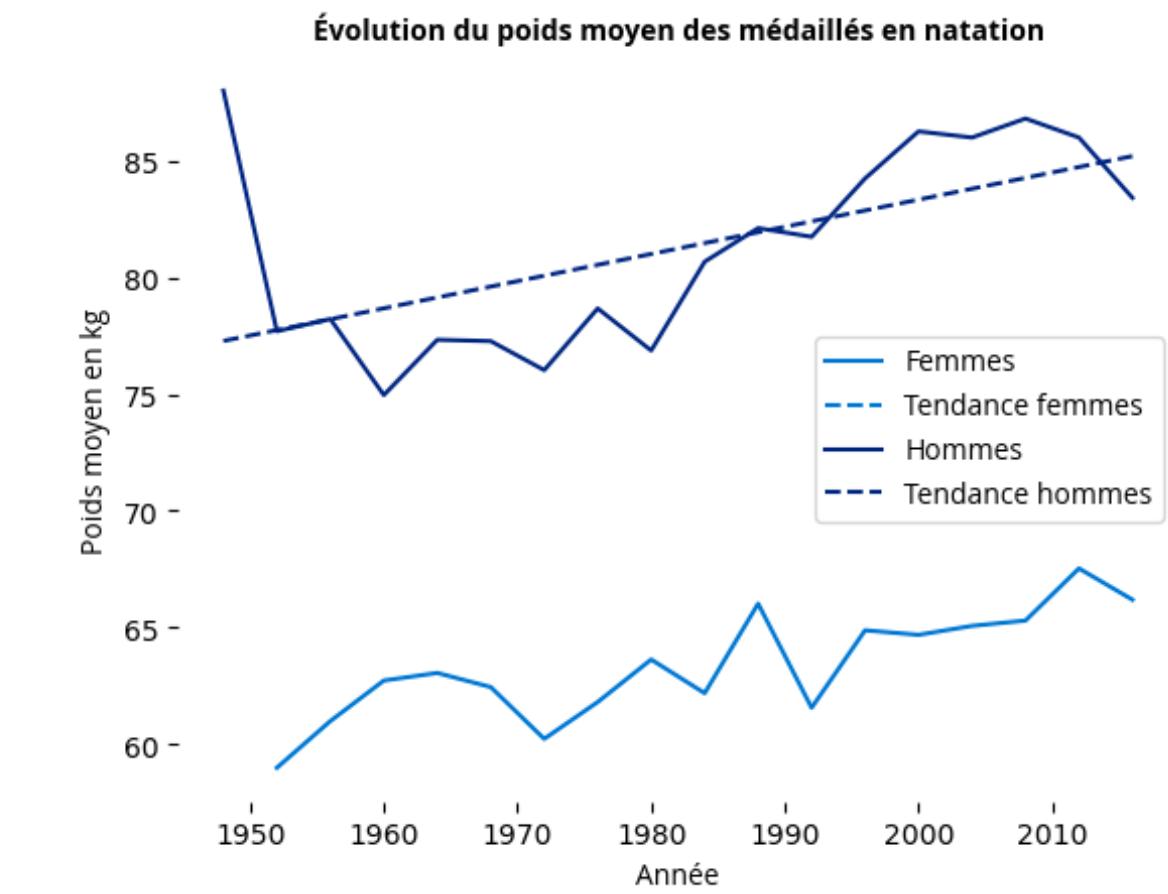
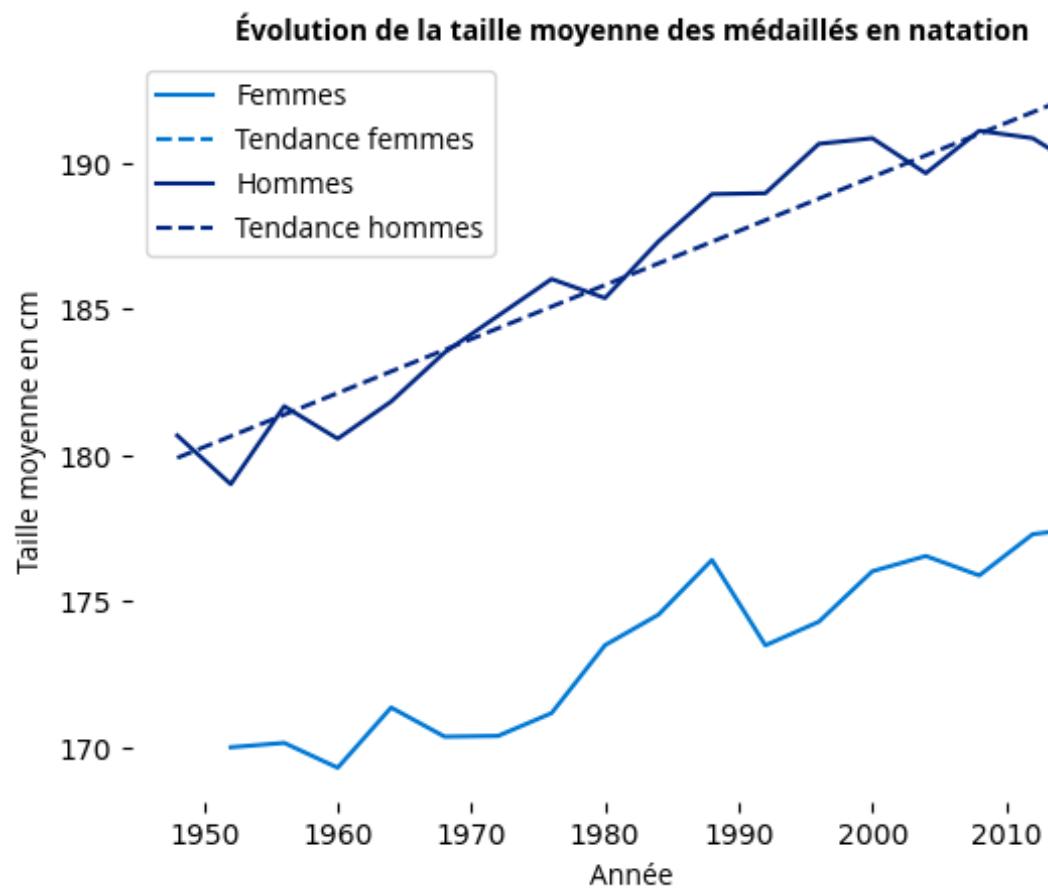
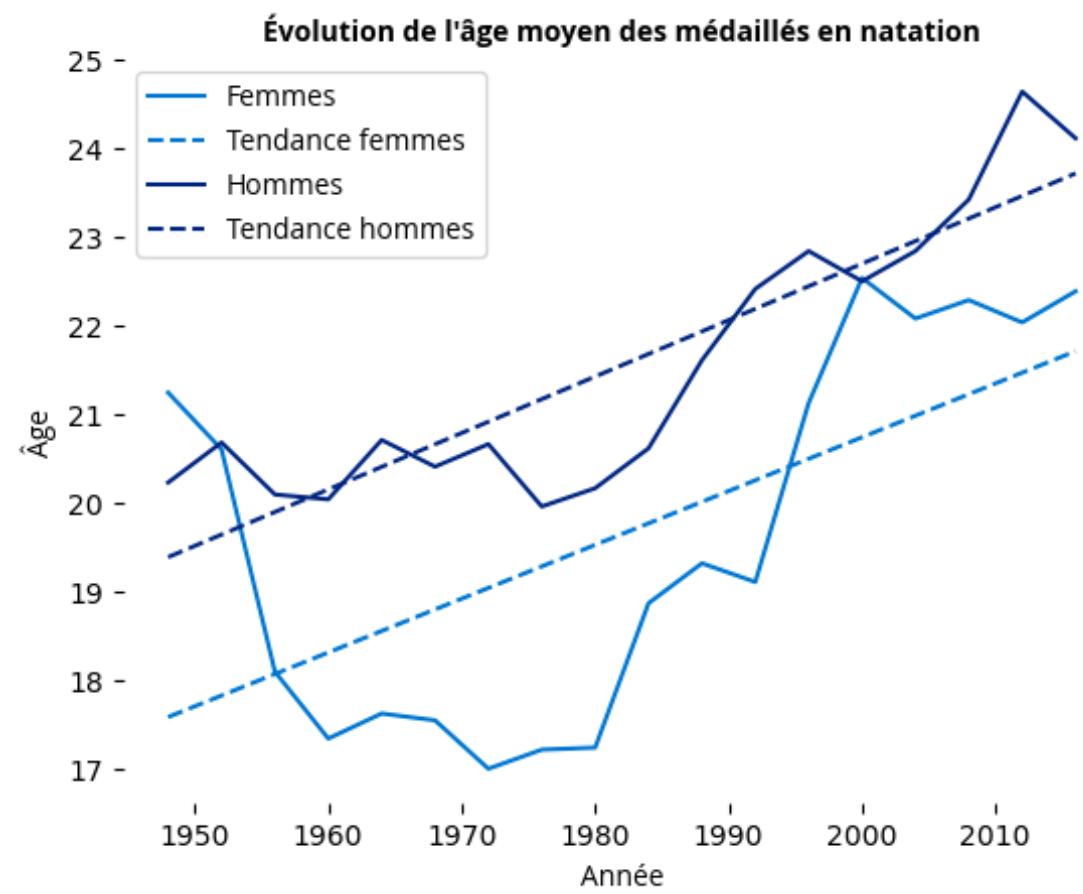
Natation





PORTRAIT ROBOT DES MÉDAILLÉS

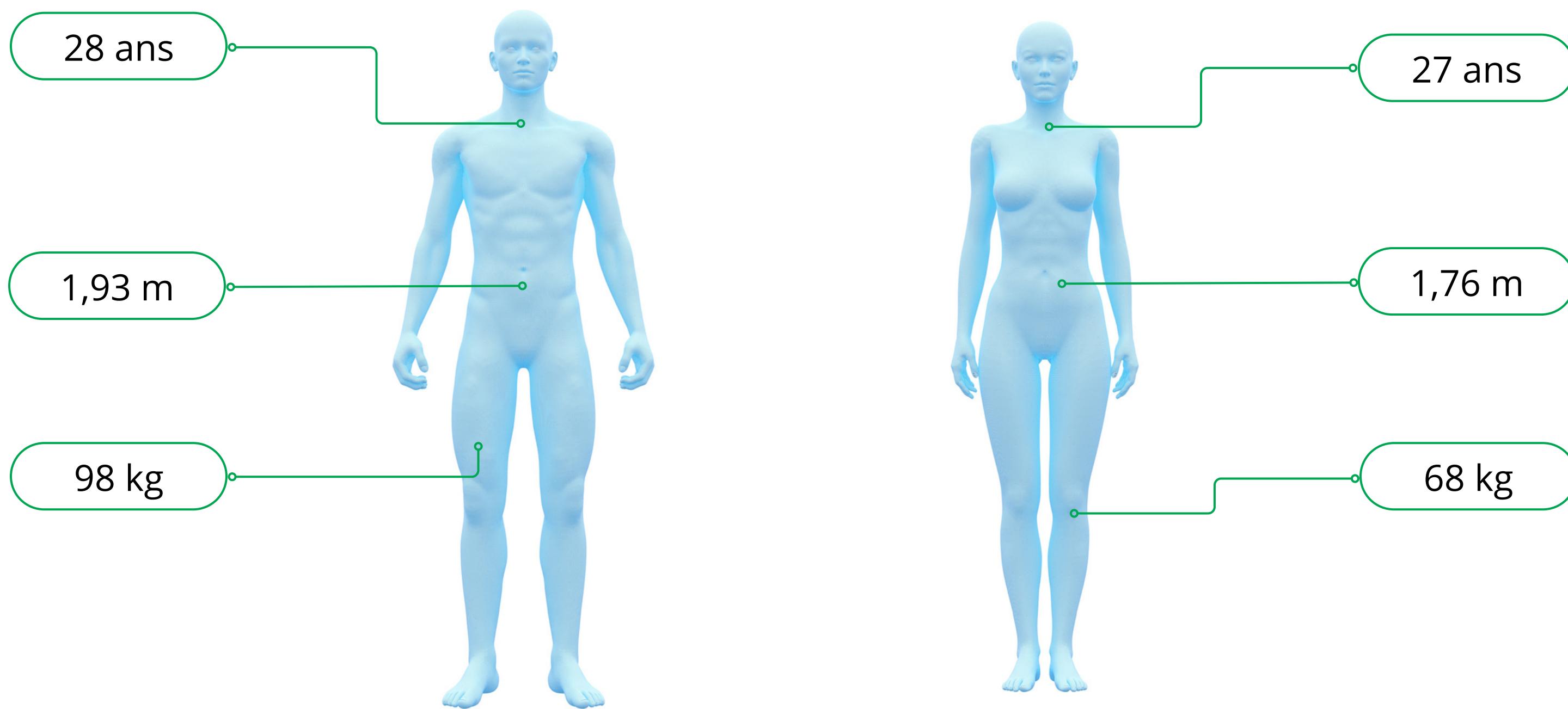
Évolution des caractéristiques physiques en natation





PORTRAIT ROBOT DES MÉDAILLÉS

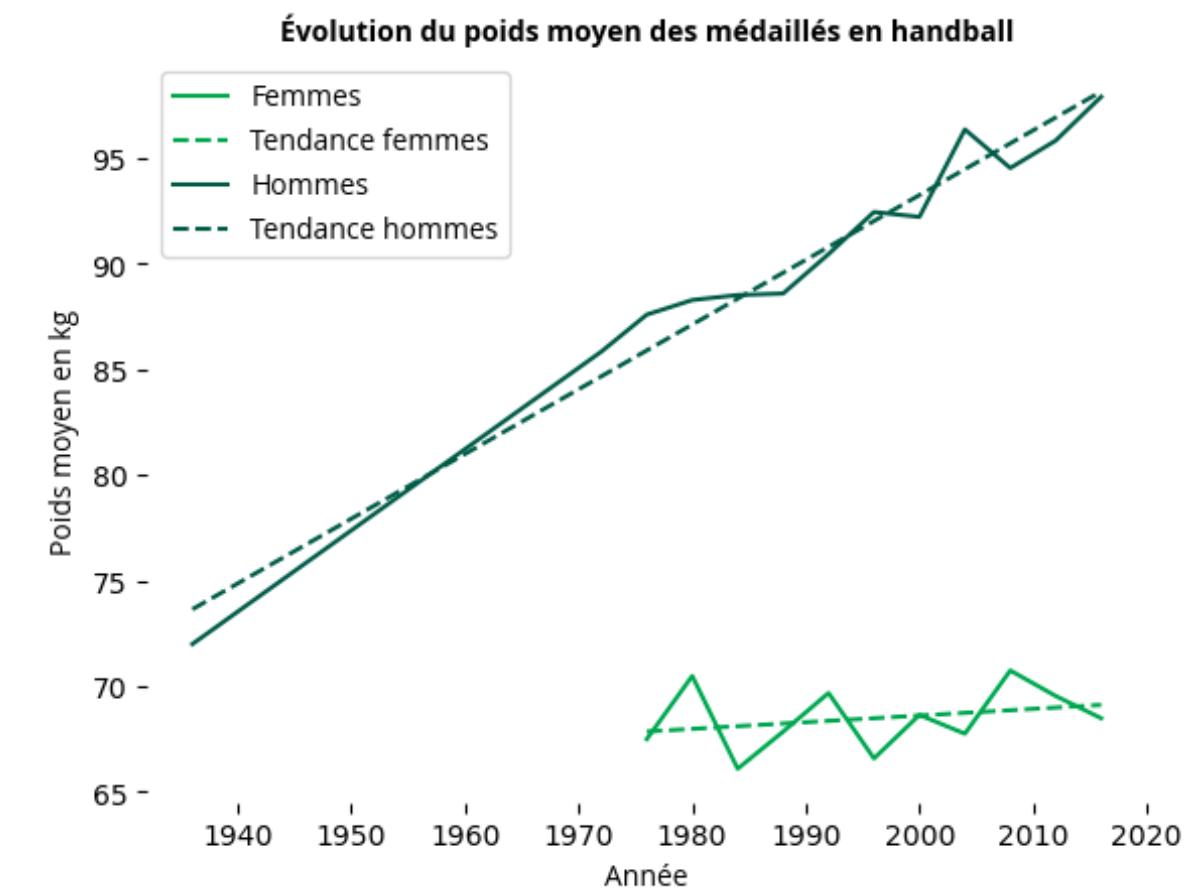
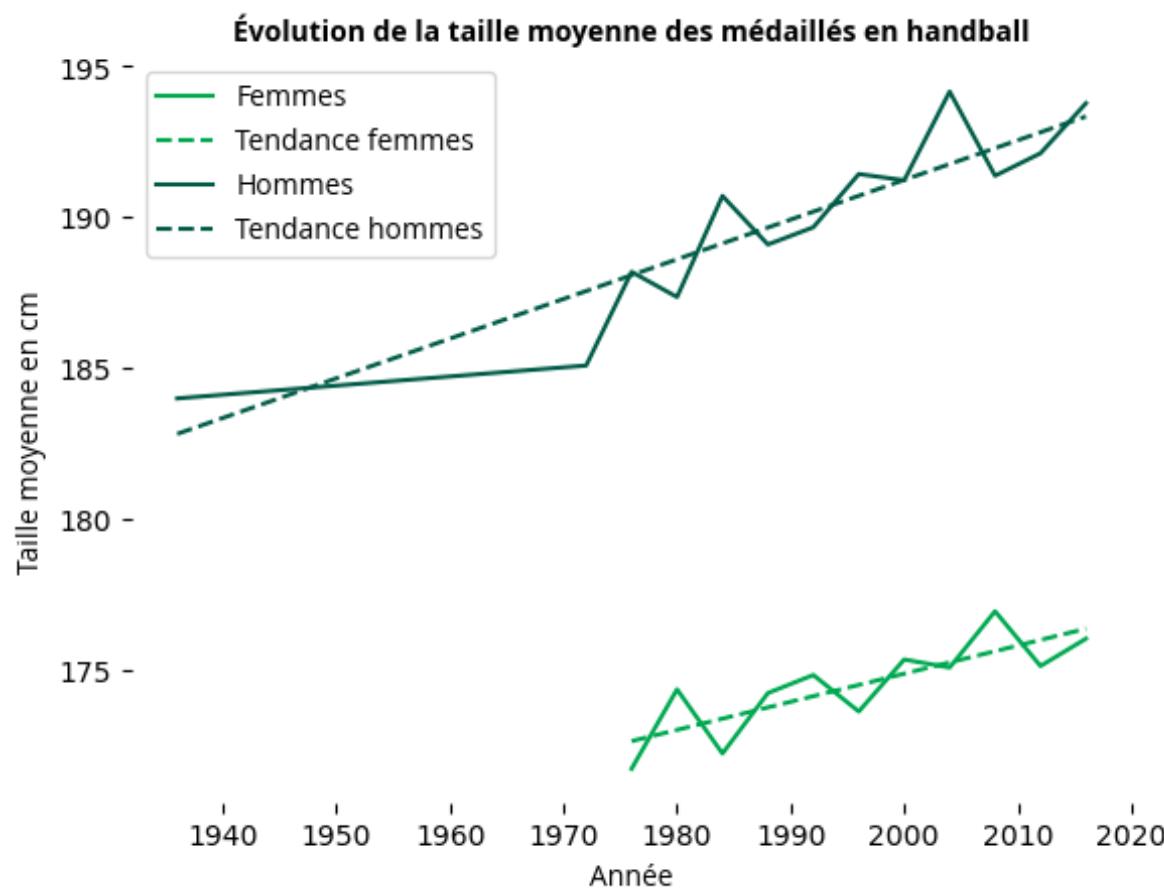
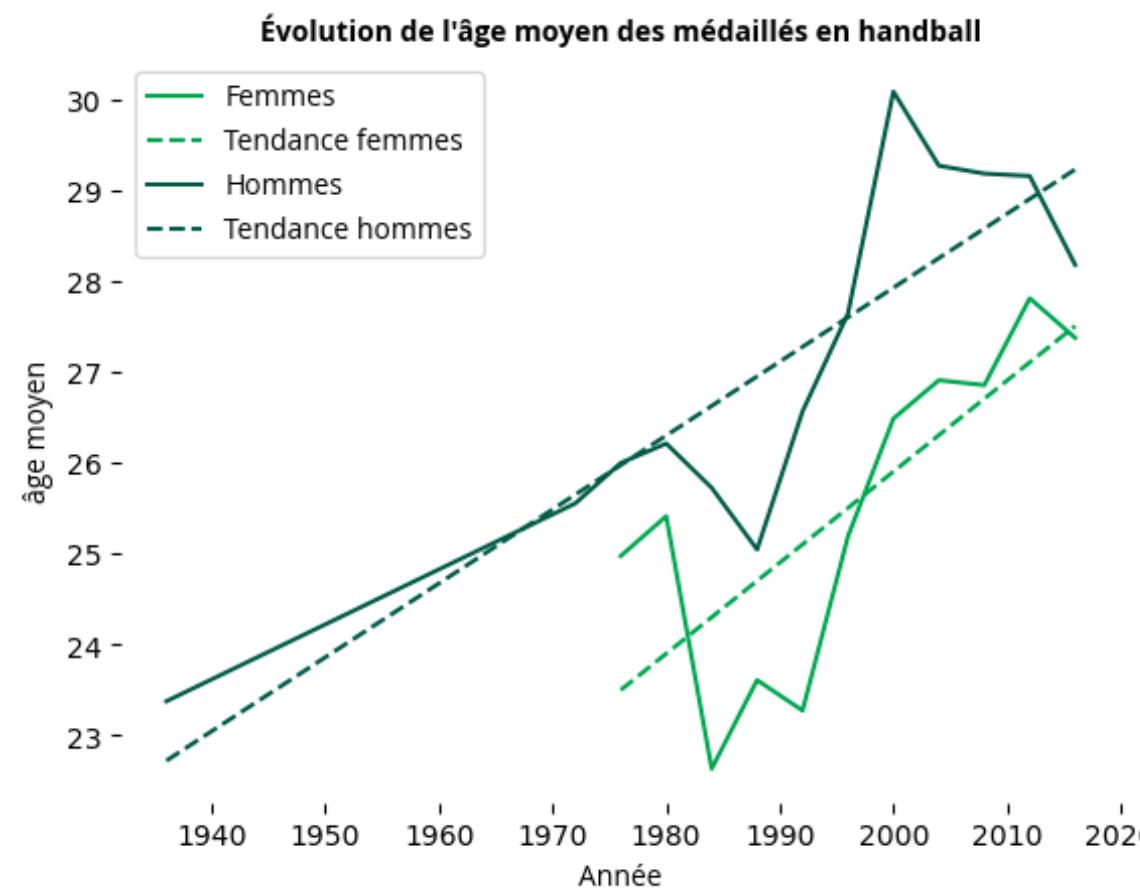
Handball





PORTRAIT ROBOT DES MÉDAILLÉS

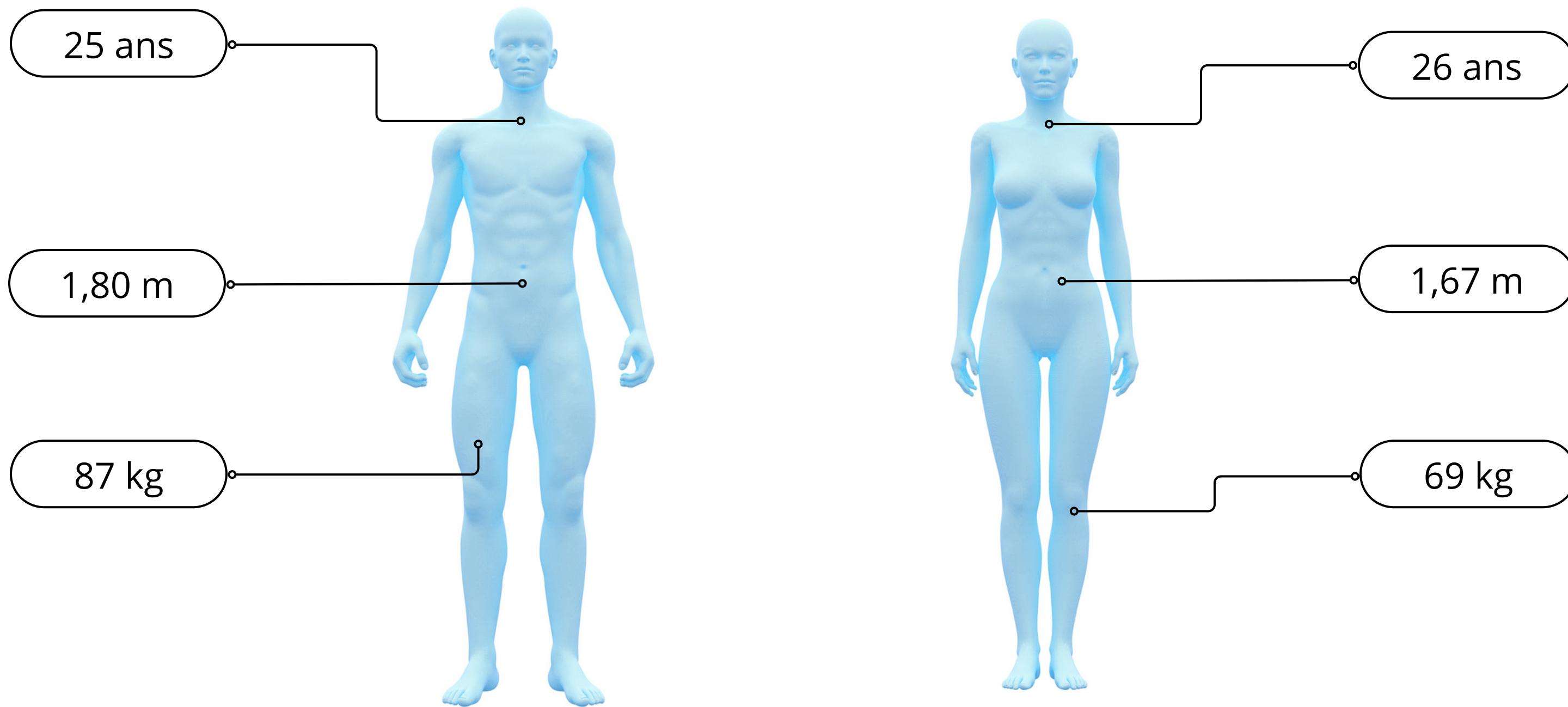
Évolution des caractéristiques physiques en handball





PORTRAIT ROBOT DES MÉDAILLÉS

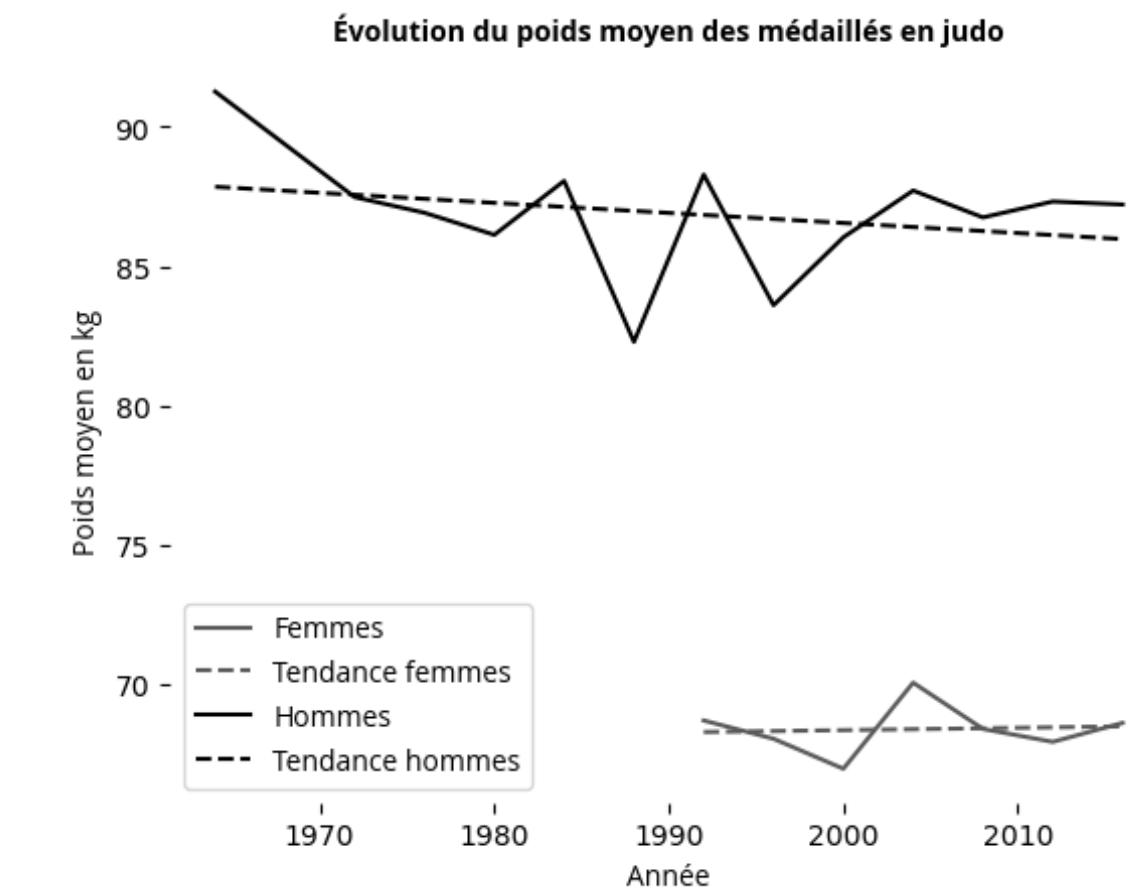
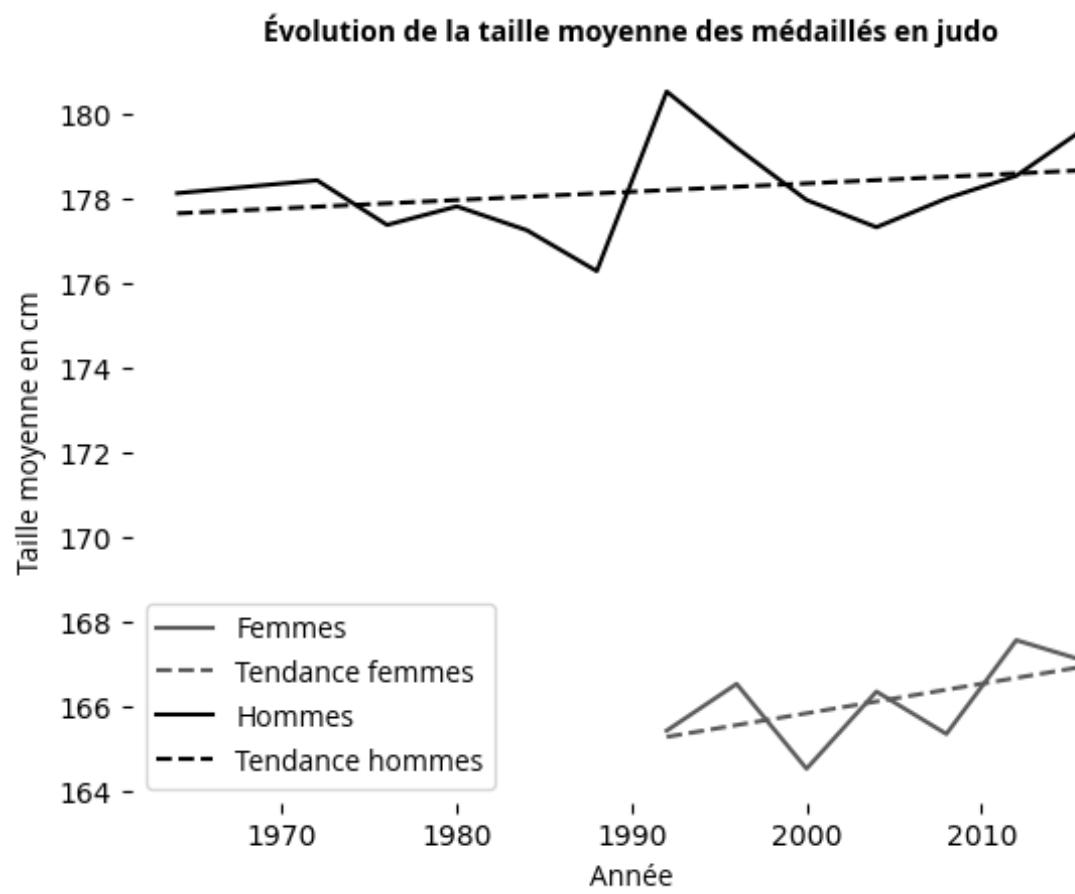
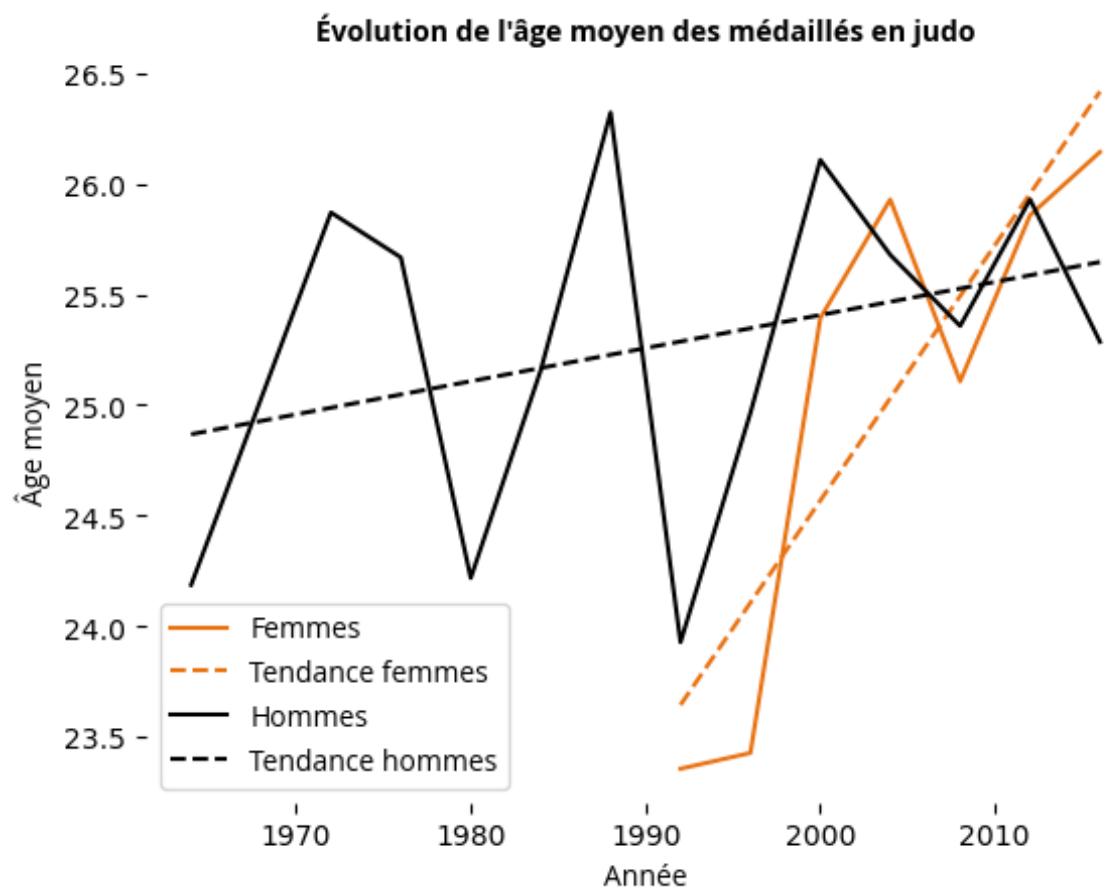
Judo





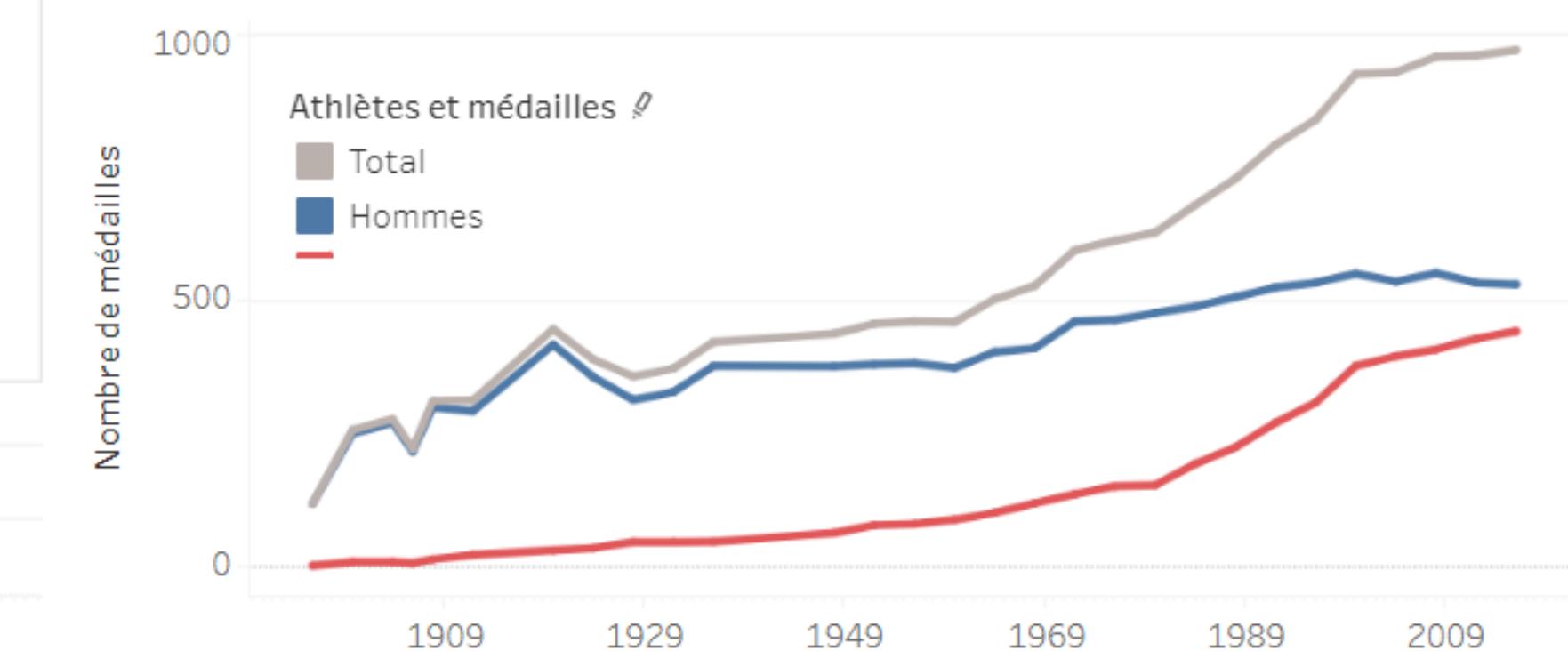
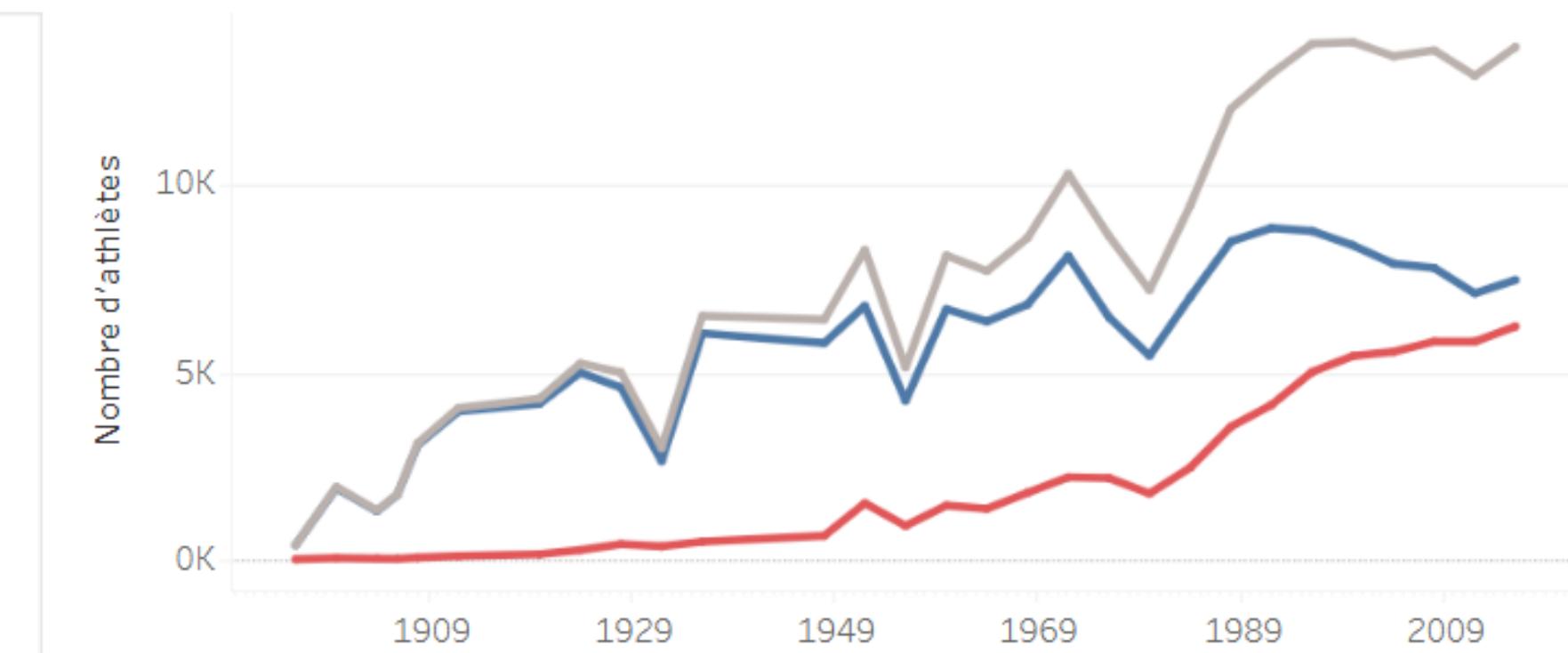
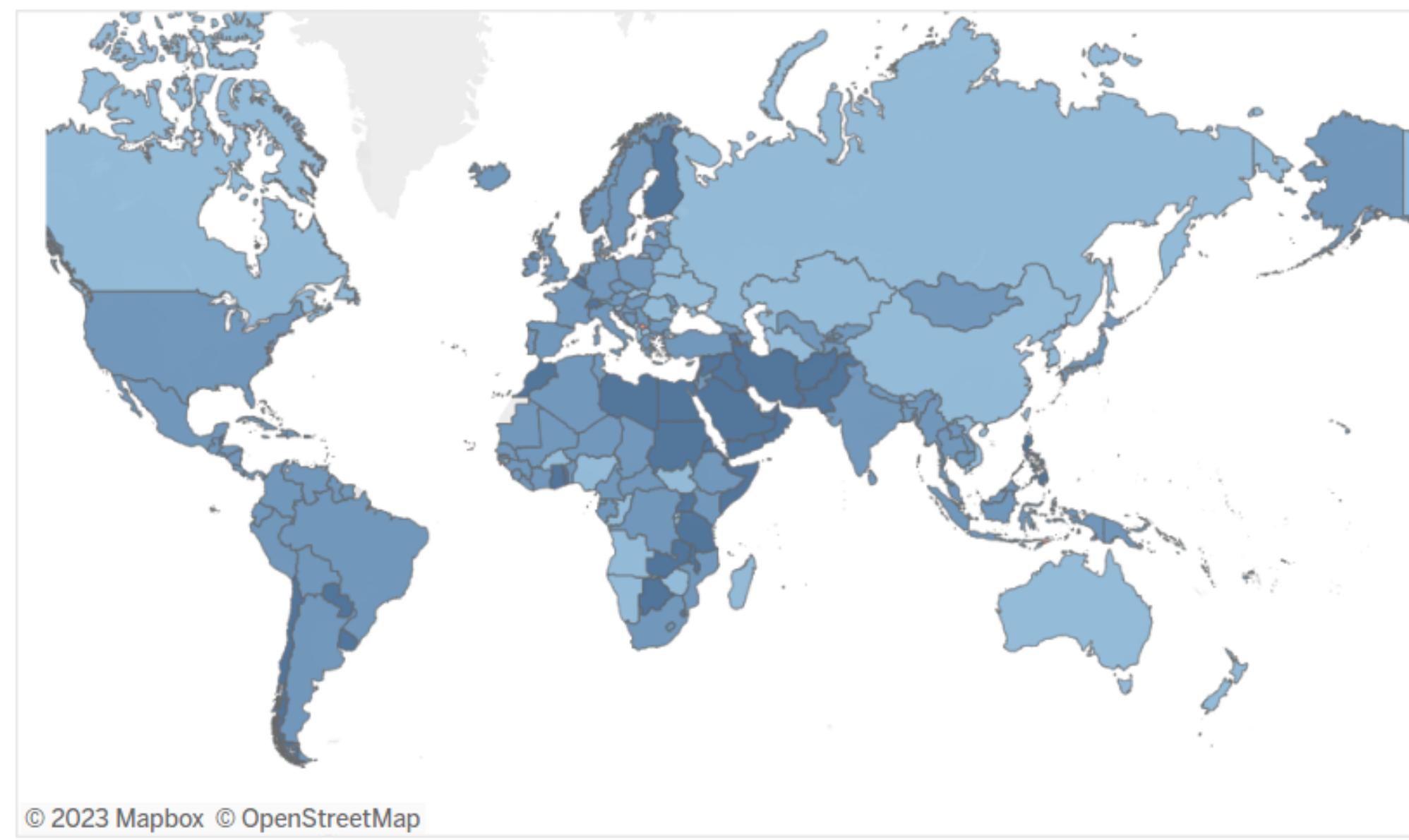
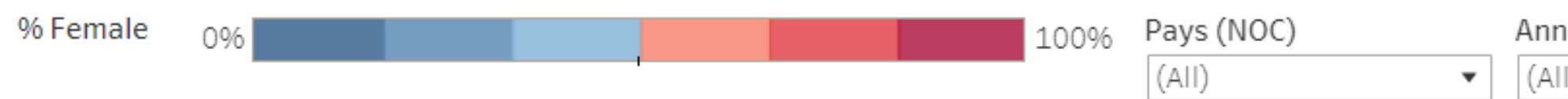
PORTRAIT ROBOT DES MÉDAILLÉS

Évolution des caractéristiques physiques en judo





INÉGALITÉS DE GENRE AUX JO





INÉGALITÉS DE GENRE AUX JO

Médailles femmes

11 61

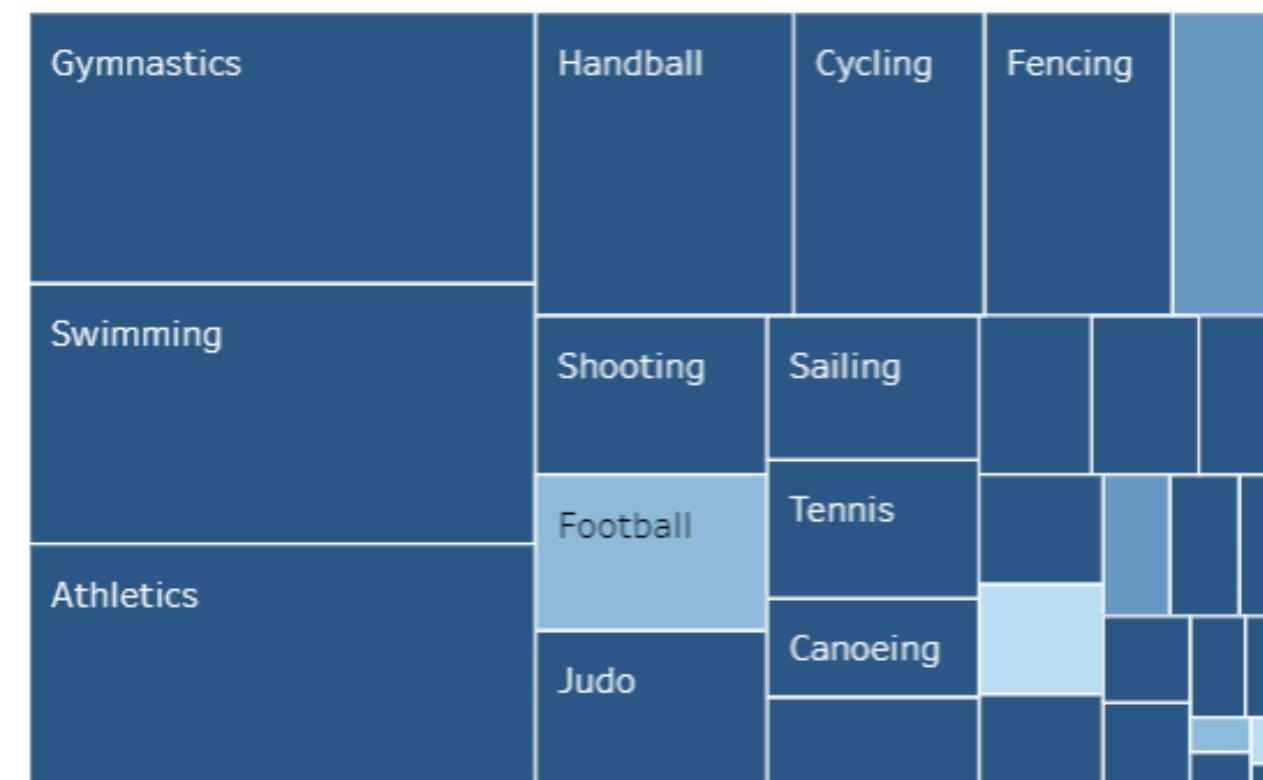
% Female

0%

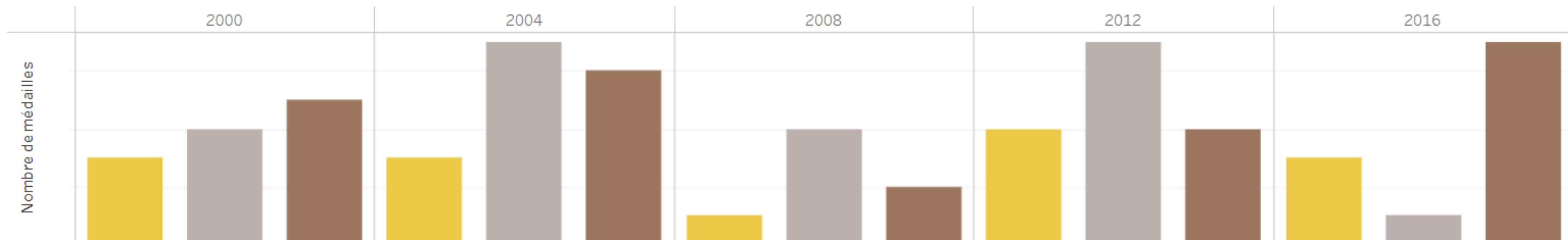
Pays (NOC)

France

Top Sports Femmes



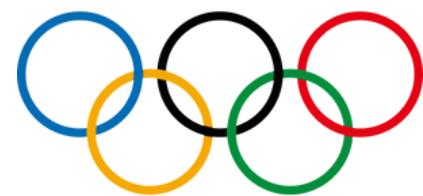
Sport	2000	2004	2008	2012	2016
Football				17	17
Rhythmic Gymnastics	7			1	1
Synchronized Swimming	10	2	2	2	2
Taekwondo	1	2	1	2	3
Badminton	2	2	1	1	1
Diving	6	1	2	3	1
Rugby Sevens					13
Basketball	12			12	12
Golf					2
Judo	7	6	7	7	7
Handball	15	15	14	14	16
Swimming	11	23	40	25	26
Modern Pentathlon	1	2	1	2	1
Gymnastics	32	32	27	19	20
Triathlon	3	1	2	3	2



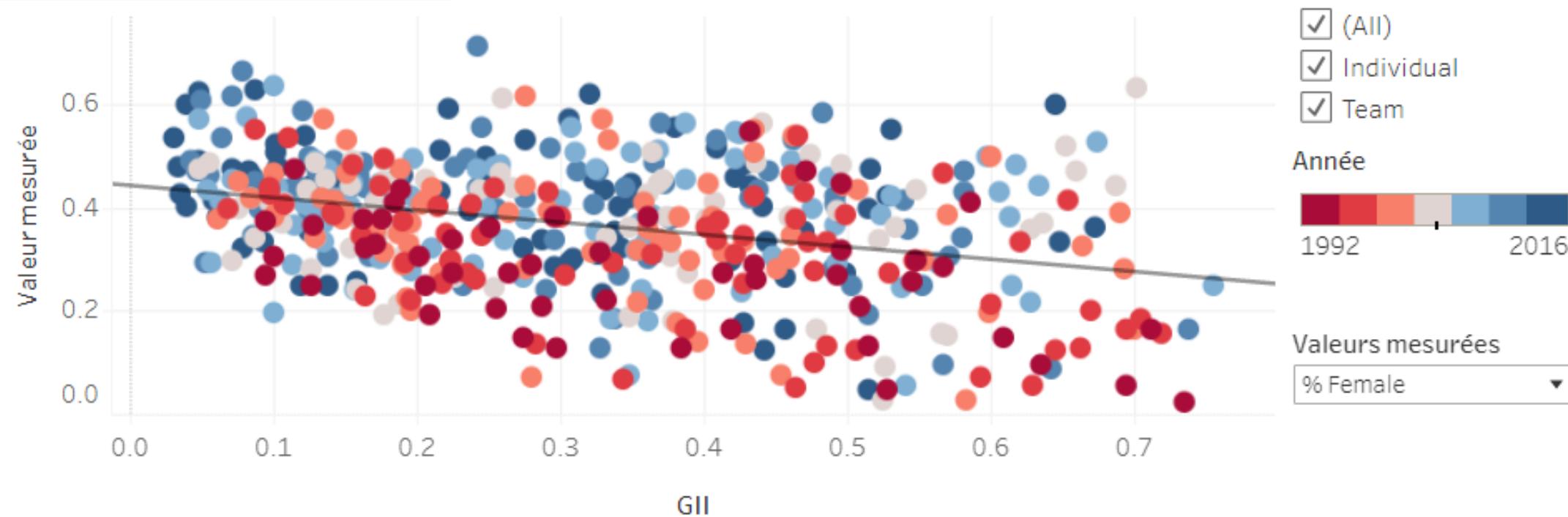
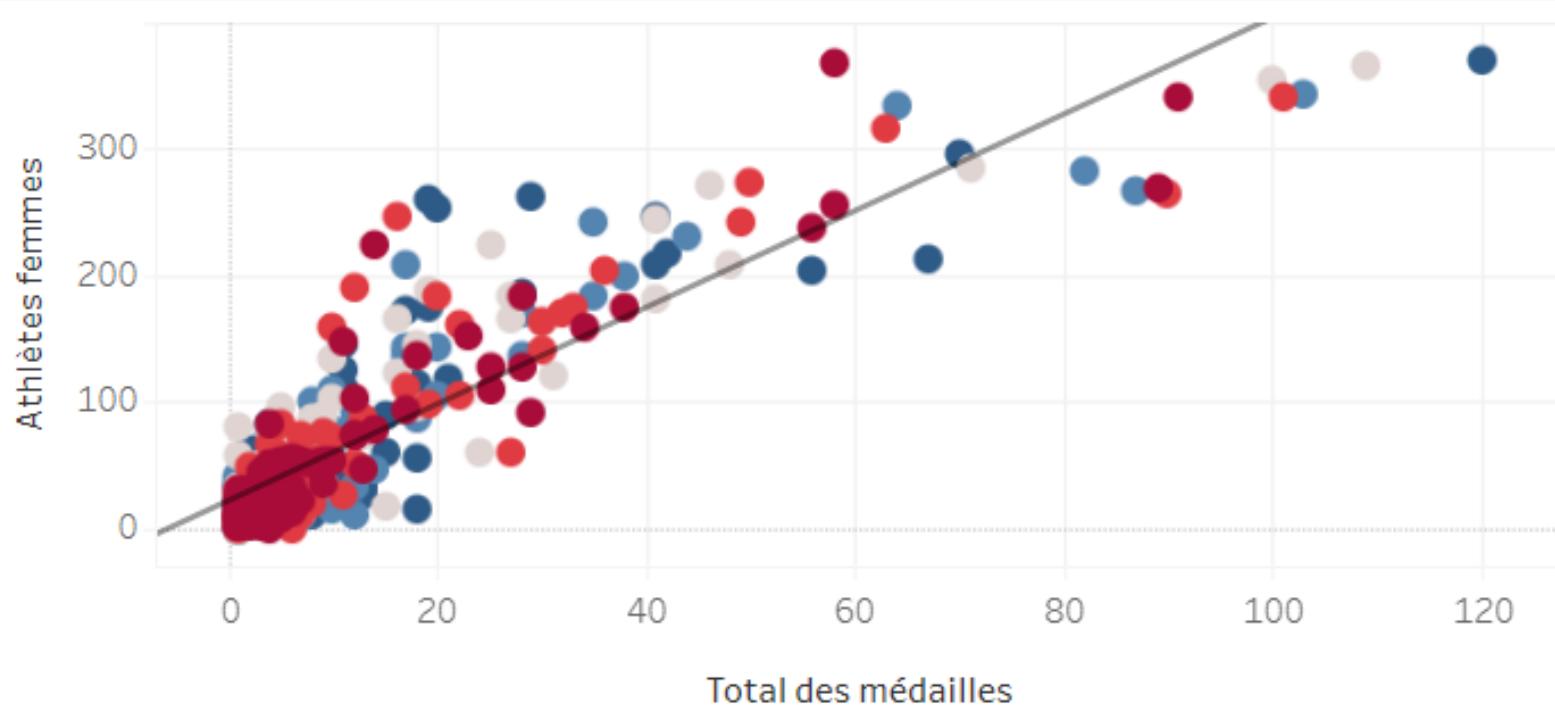
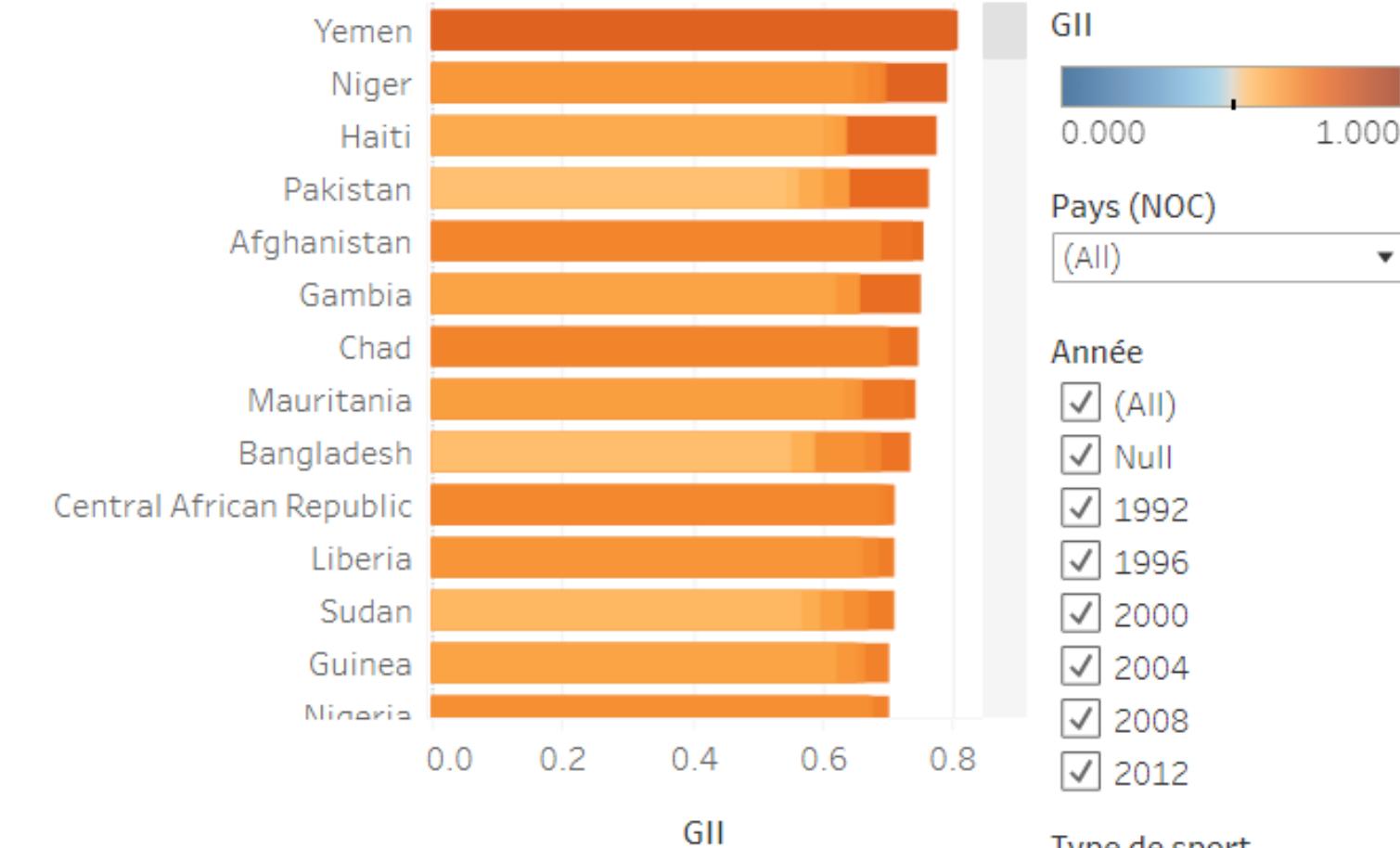
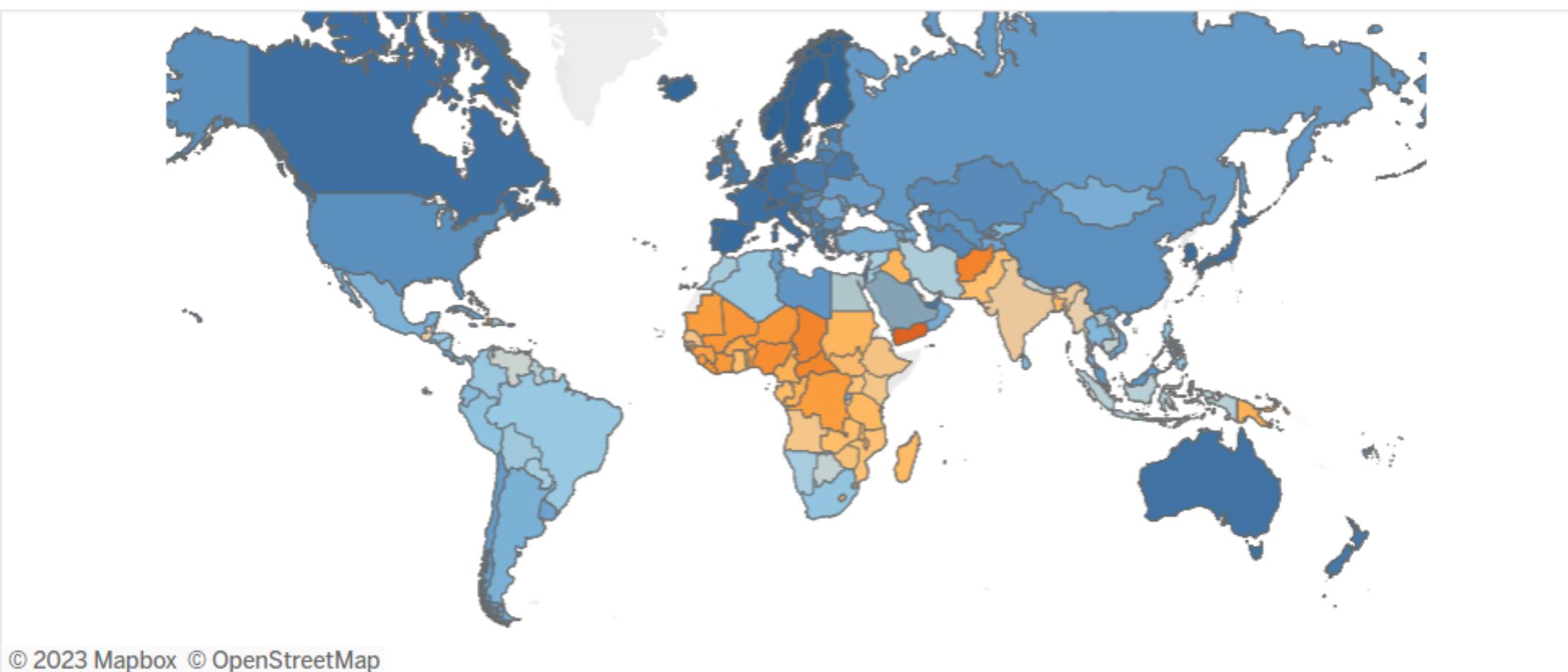
Année
(Multiple values)

Type de sport
 (All)
 Individual
 Team

Médailles
 Or Femmes
 Argent Femmes
 Bronze Femmes



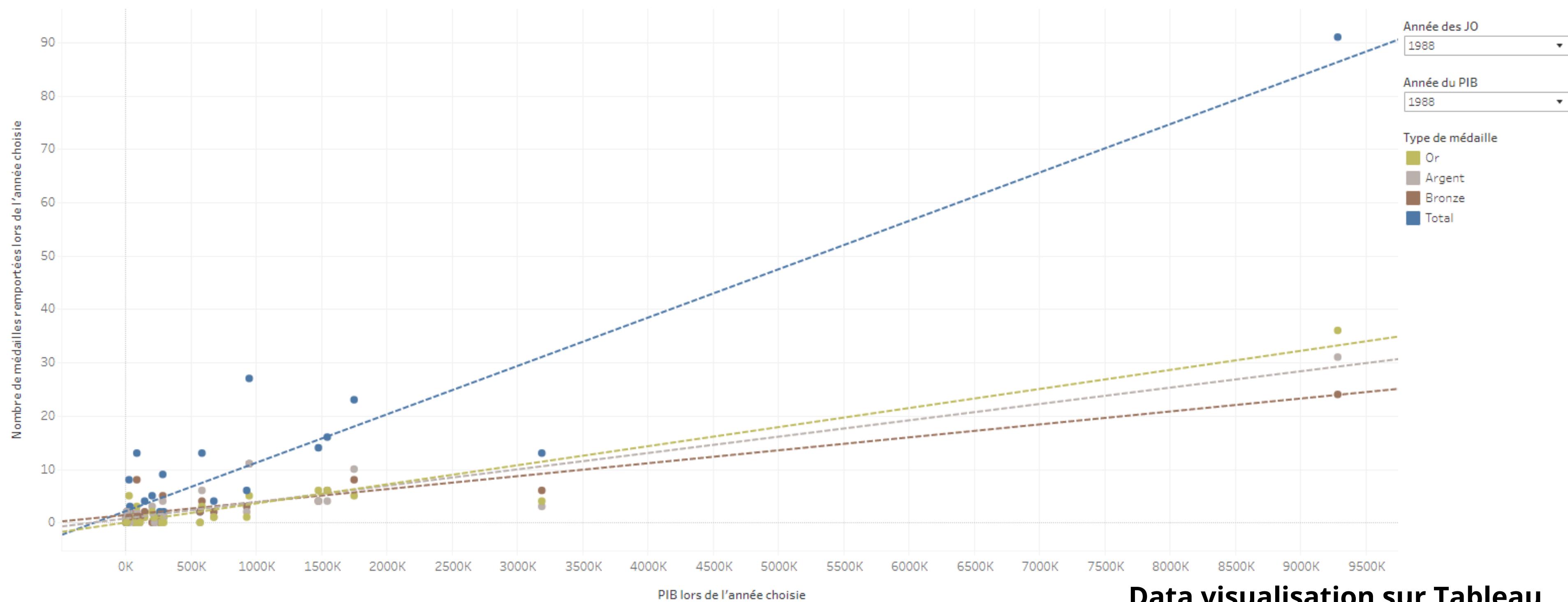
INÉGALITÉS DE GENRE AUX JO





FACTEURS DE RÉUSSITE AUX JO

PIB et médailles



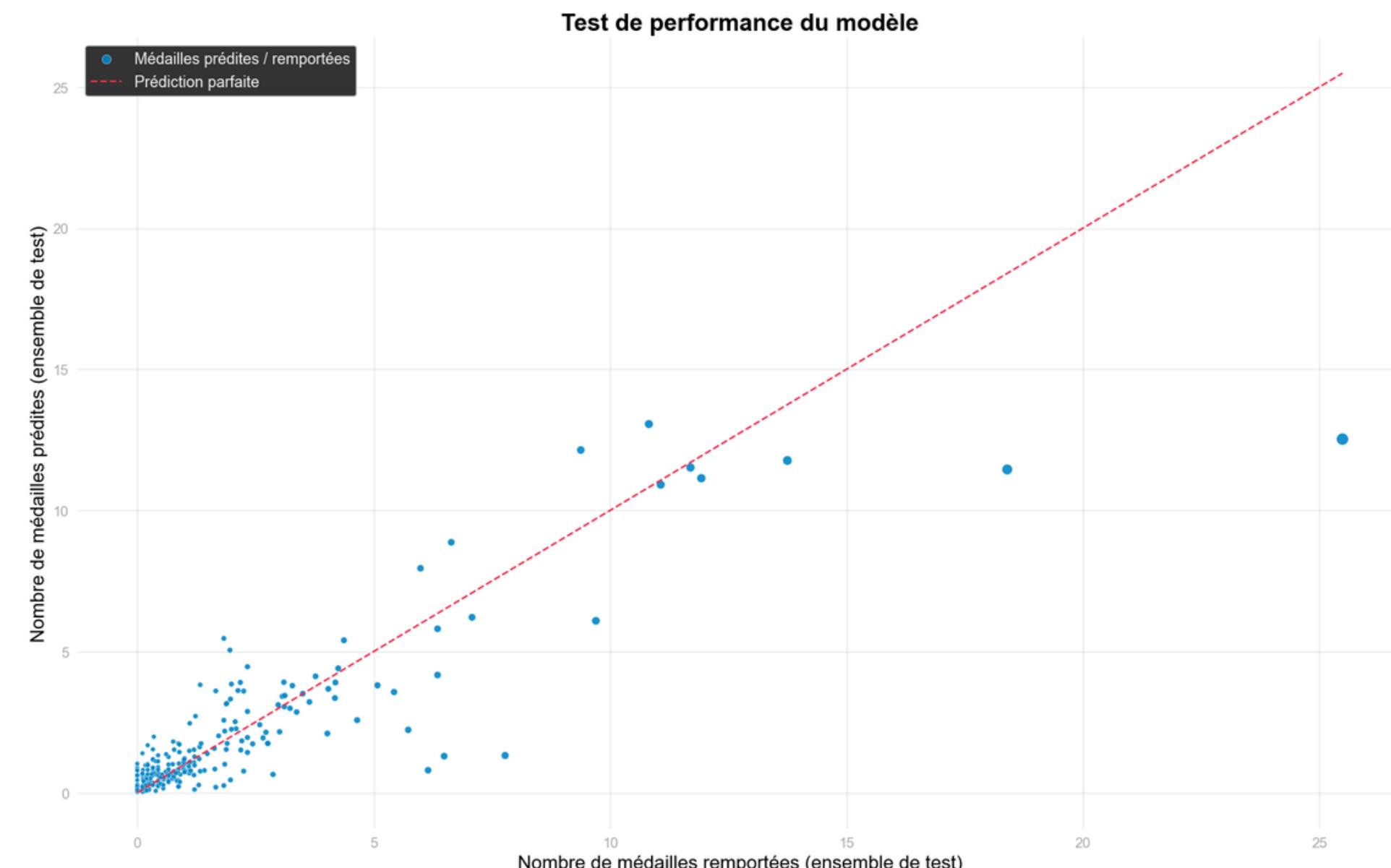


MODÈLE PRÉdictif DES PERFORMANCES AUX JO

Variables retenues

- PIB
- PIB / habitant
- RNB
- Population
- Pourcentage de la population âgée de 65 ans et +

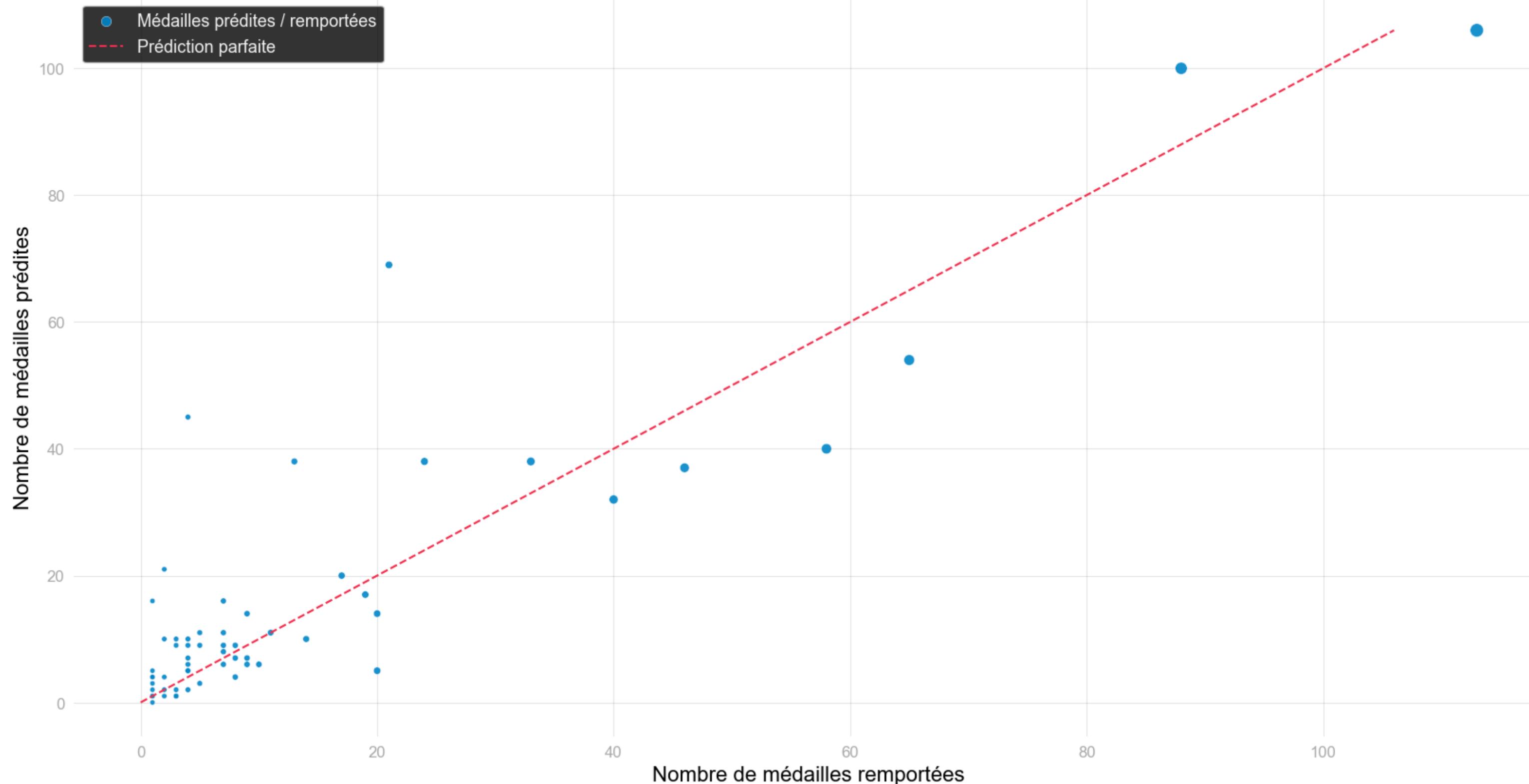
Modèle prédictif retenu : random forest regressor





PRÉDICTIONS POUR TOKYO 2020

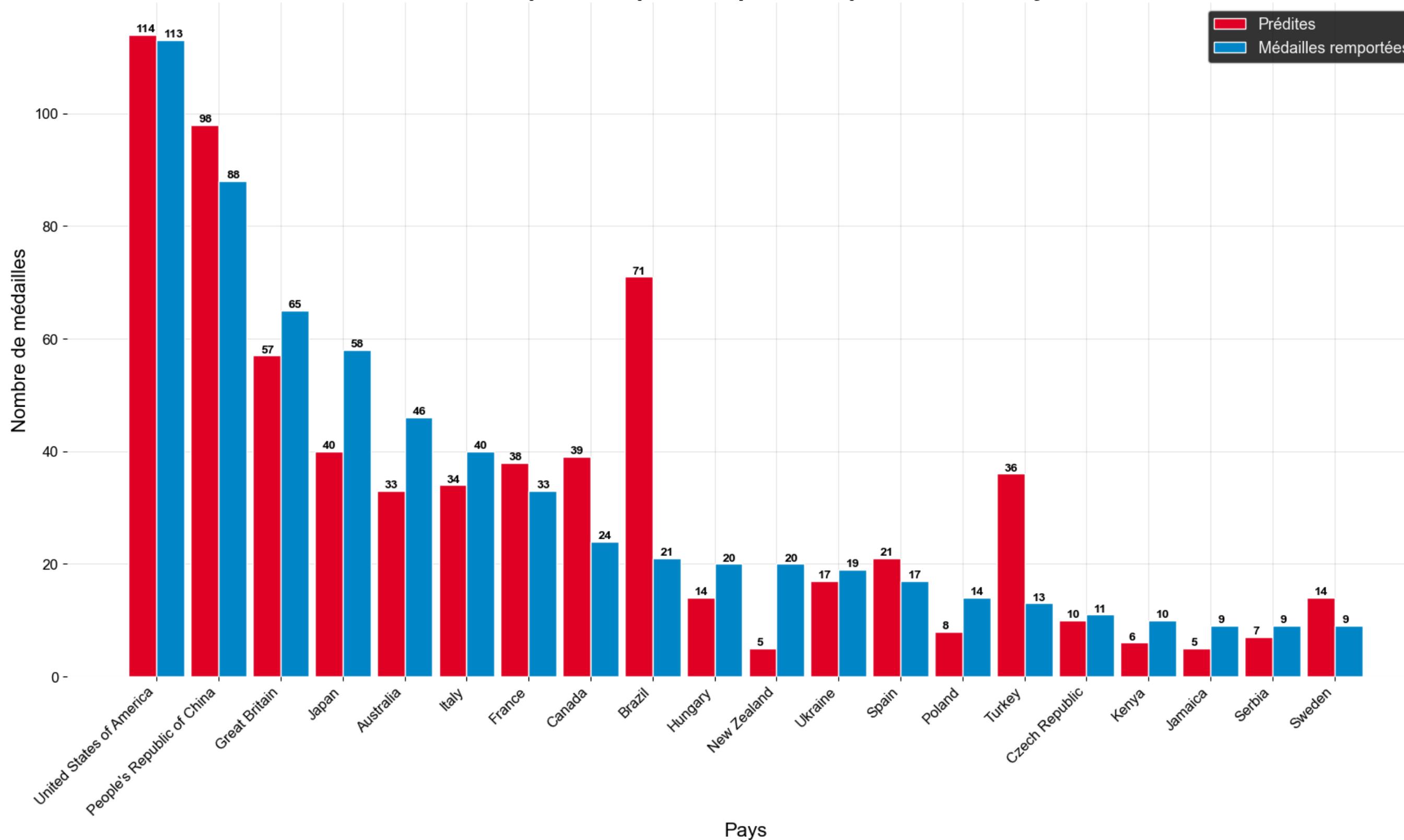
Comparaison entre nombre de médailles obtenues et prédites aux JO de Tokyo 2020

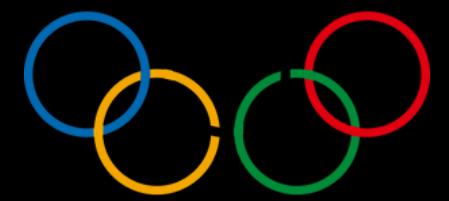




PRÉDICTIONS POUR LE TOP 20

Médailles remportées / prédites pour le top 20 aux JO Tokyo 2020



A blurry, high-speed photograph of a runner in motion, wearing a white and blue athletic outfit, serving as the background for the text.

Merci
À vous pour les questions !