

# SAé 2.04 – Partie analyse de données

Baptiste Bertout et Killian Lawson

## Table des matières

Question 1 :

Question 2 :

Question A :

Question B :

Question C :

Question 3 :

Question A :

Question B :

Question I :

Question II :

Question III :

Question IV :

Question V :

Question VI :

## Question 1 :

Déterminer les 20 athlètes affichant le plus de participations aux J.O :

Pour savoir les 20 athlètes affichant le plus de participations aux J.O, nous avons effectuer la requête suivantes :

```
SELECT name
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
GROUP BY name
ORDER BY COUNT(DISTINCT (name,year)) DESC
LIMIT 20;
```

SQL

Nous avons trouver le resultat suivant :

```

name
-----
Oksana Aleksandrovna Chusovitina
Joo Filipe Gaspar Rodrigues
Noriaki Kasai
Yekaterina Anatolyevna Khodotovich-Karsten
Albert Mikhaylovich Demchenko
Giovanni Pellielo
Jess ngel Garca Bragado
Nino Salukvadze (-Machavariani)
Segun Moses Toriola
Tinne Eva Caroline Wilhelmsson-Silfvn
Leander Adrian Paes
Alessandra Sensini
Fehaid Al-Deehani
Evgeniya Nikolova Radanova
Dragutin Topi
Afanasijs Kuzmins
Emil Ivanov Milev
Franck Dumoulin
Claudia Pechstein
Armin Zggeler
(20 lignes)

```

Figure 1: Les 20 athlètes affichant le plus de participations

## Question 2 :

### Question A :

#### Remplir le tableau statistique

Pour cette question nous avons l'année 1992 et nous avons choisi de prendre 'Summer' comme type de jeu.

Pour voir le tableau rempli, il faut visualiser le pdf qui s'appelle **tableau\_question2\_a.pdf**

Pour remplir ce tableau, nous avons utilisé les requêtes suivantes :

#### Requête pour la colonne 'pays'

```

SELECT DISTINCT region
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) JOIN
regions r USING(noc)
WHERE year=1992 AND season='Summer';

```

SQL

#### Requête pour la colonne 'Moyenne d'âge'

```

SELECT ROUND(AVG(CAST(age AS integer)),2)
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) JOIN
regions r USING(noc)
WHERE year=1992 AND season='Summer' AND age <> 'NA'
GROUP BY region;

```

SQL

#### Requête pour la colonne 'Nb sportifs'

SQL

```
SELECT count(DISTINCT name)
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) JOIN
regions r USING(noc)
WHERE year=1992 AND season='Summer'
GROUP BY region;
```

### Requête pour la colonne 'Min âge'

SQL

```
SELECT min(CAST(age AS integer))
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) JOIN
regions r USING(noc)
WHERE year=1992 AND season='Summer' AND age <> 'NA'
GROUP BY region;
```

### Requête pour la colonne 'Max âge'

SQL

```
SELECT max(CAST(age AS integer))
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) JOIN
regions r USING(noc)
WHERE year=1992 AND season='Summer' AND age <> 'NA'
GROUP BY region;
```

En ce qui concerne les graphiques sur cette question, ils se trouvent dans le dossier pdf > pdf\_question2\_a

## Question B :

### Comparer l'âge moyen des médaillés et l'âge moyen des participants

Pour l'âge moyen nous pouvons faire de deux manière :

- La première est de passé par une requête SQL pour avoir l'âge moyen des participants et l'âge moyen des médaillés directement. Pour cela nous utilisons les requêtes suivantes :

#### Requêtes pour l'âge moyen des participants

SQL

```
SELECT ROUND(AVG(CAST(age AS integer)),2)
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND age <> 'NA';
```

#### Requête pour l'âge moyen des médaillés

SQL

```
SELECT ROUND(AVG(CAST(age AS integer)),2)
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND age <> 'NA' AND medal<>'NA';
```

- Pour la seconde nous pouvons passer par les calculs. Nous passons par une requête pour obtenir la liste des âges des différents participants et médaillés, et ensuite nous passons par le calcul d'une moyenne mathématique avec la formule :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$$

Figure 2: Formules mathématiques pour le calcul de la moyenne

Pour ce faire on utilise les requêtes suivantes :

#### Requêtes pour la liste de l'âge des participants

```
SELECT age
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND age <> 'NA';
```

SQL

#### Requête pour la liste de l'âge des médaillés

```
SELECT age
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND age <> 'NA' AND medal<>'NA';
```

SQL

Ensuite on utilise la formule ce qui nous donne :

→ Age moyen des participants :  $(24+26+20+\dots+25+20+20)/12934 = 24.4$

→ Age moyen des médaillés :  $(26+29+45+\dots+24+27+27)/1712 = 24.69$

Ainsi, nous avons la comparaison suivante :

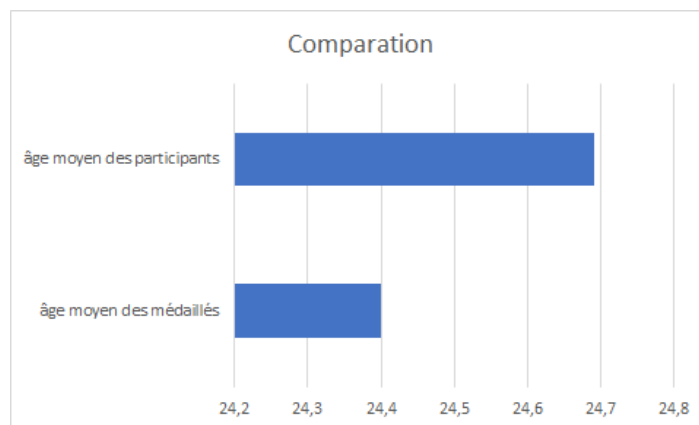


Figure 3: Comparaison des âges entre les participants et les médaillés

## Question C :

Comparer le poids moyen des médaillé(e)s et le poids moyen des participant(e)s

Pour le poids moyen nous pouvons faire de deux manière :

- La première est de passé par une requête SQL pour avoir le poids moyen des participant(e)s et le poids moyen des médaillé(e)s directement. Pour cela nous utilisons les requêtes suivantes :

#### Requêtes pour le poids moyen des participants

```
SELECT AVG(CAST(weight AS numeric))
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 and season='Summer' AND sex='M' AND weight<>'NA';
```

SQL

#### Requêtes pour le poids moyen des participantes

```
SELECT AVG(CAST(weight AS numeric))
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND sex='F' AND weight<>'NA';
```

SQL

#### Requête pour le poids moyen des médaillés

```
SELECT AVG(CAST(weight AS numeric))
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND sex='M' AND medal<>'NA' AND weight<>'NA';
```

SQL

#### Requête pour le poids moyen des médaillées

```
SELECT AVG(CAST(weight AS numeric))
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND sex='F' AND medal<>'NA' AND weight<>'NA';
```

SQL

- Pour la seconde nous pouvons passer par les calculs. Nous passons par une requête pour obtenir la liste des poids des différents participant(e)s et médaillé(e)s, et ensuite nous passons par le calcul d'une moyenne mathématique avec la formule :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$$

Figure 4: Formules mathématiques pour le calcul de la moyenne

Pour ce faire on utilise les requêtes suivantes :

#### Requêtes pour la liste de poids des participants

```
SELECT weight
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND sex='M' AND weight<>'NA' AND medal<>'NA';
```

SQL

#### Requêtes pour la liste de poids des participantes

```
SELECT weight
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND sex='F' AND weight<>'NA';
```

SQL

*Requête pour la liste de poids des médaillés*

SQL

```
SELECT weight
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND sex='M' AND weight<>'NA' AND medal<>'NA';
```

*Requête pour la liste de poids des médaillées*

SQL

```
SELECT weight
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno)
WHERE year=1992 AND season='Summer' AND sex='F' AND weight<>'NA' AND medal<>'NA';
```

Ensuite on utilise la formule ce qui nous donne :

→ Poids moyen des participants :  $(80+107+64+\dots+78+80+80)/7258 = 76.24$

→ Poids moyen des participantes :  $(64+45+52+\dots+64+58+58)/3215 = 59.64$

→ Poids moyen des médaillés :  $(92+50+89+\dots+83+90+90)/1020 = 80.31$

→ Poids moyen des médaillées :  $(81+68+68+\dots+68+68+63)/509 = 63.43$

Ainsi, nous avons la comparaison suivante :

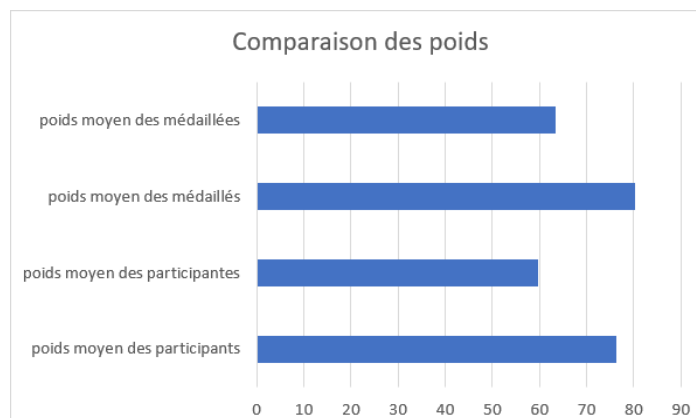


Figure 5: Comparaison des âges entre les participants et les médaillés

### Question 3 :

### Question A :

Faire la liste des 15 pays qui ont gagné le plus de médailles (1992 à 2016)

Pour avoir ce résultat nous utilisons la requête suivante. Nous ne spécifions pas les années dans la requête puisque tous les athlètes présents dans la base donnée sont enregistrés entre les années 1992 et 2016.

*Requête des 15 pays qui ont gagné le plus de médailles*

```
SELECT region,count(medal)
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) join
regions r USING(noc)
WHERE medal<>'NA'
GROUP BY region
ORDER BY count(medal) DESC
LIMIT 15;
```

## Question B :

## Question I :

Choisir 5 pays parmi ces 15 et représenter l'évolution du nombre de participants

Pour étudier le nombre de participants par année et par pays, nous avons effectué la requête suivante :

*Requêtes pour le nombre de participants par pays et par année*

```
SELECT count(*)
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) JOIN
regions r USING(noc)
WHERE region='nom_du_pays'
GROUP BY year
ORDER BY year;
```



Nous devons remplacer 'nom\_du\_pays' par le nom du pays que l'on souhaite. Dans notre cas les 5 pays sont : USA, Russia, Germany, Italy, France.

Suite à cela nous avons le graphique suivant, avec en abscisse les années et en ordonnée le nombre de participants.

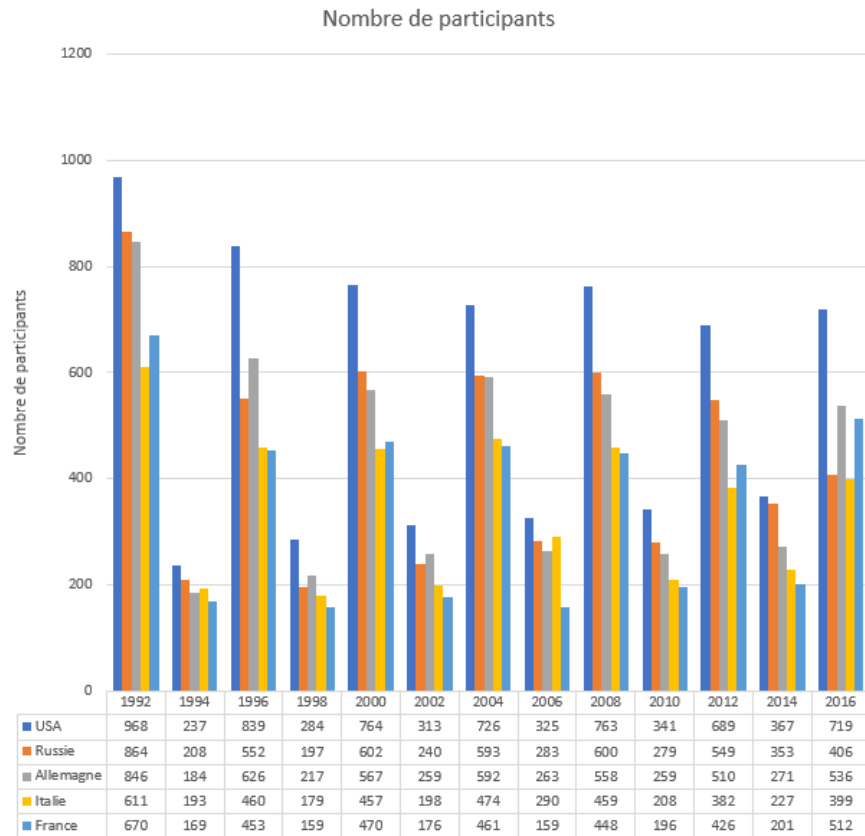


Figure 6: Nombre de participants par pays et par année

## Question II :

Choisir 5 pays parmi ces 15 et représenter l'évolution du nombre de médaillés

Pour étudier le nombre de médaillés par année et par pays, nous avons effectué la requête suivante :

*Requêtes pour le nombre de médaillés par pays et par année*

SQL

```
SELECT count(*)
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) JOIN
regions r USING(noc)
WHERE region='nom_du_pays' AND medal<>'NA'
GROUP BY year
ORDER BY year;
```



Comme précédemment, il faut remplacer nom\_du\_pays.

Suite à cela nous avons le graphique suivant, avec en abscisse les années et en ordonnée le nombre de médaillés.



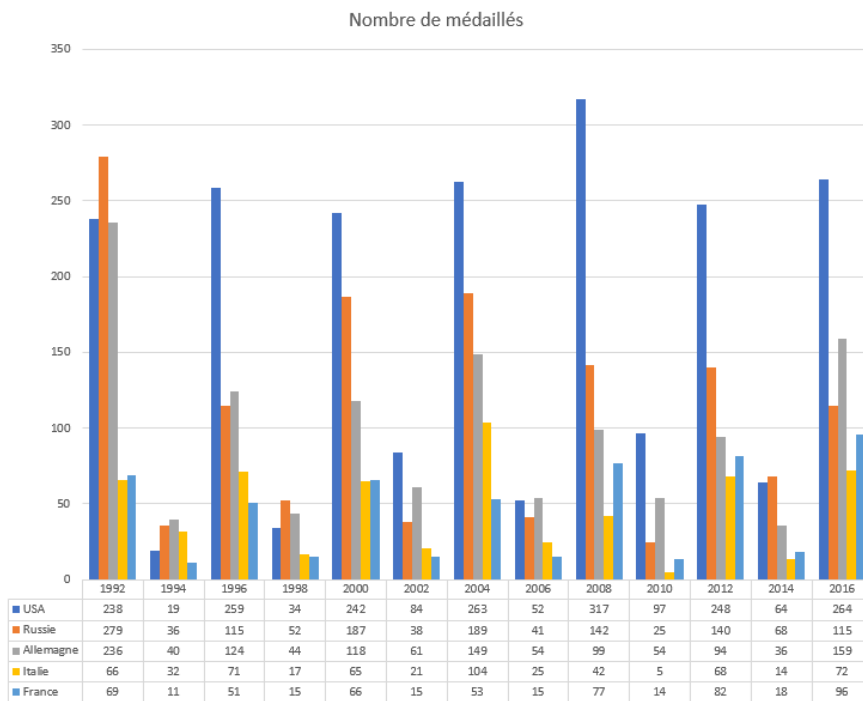


Figure 7: Nombre de participants par pays et par année

### Question III :

Choisir 5 pays parmi ces 15 et représenter l'évolution du nombre de femmes participantes

Pour étudier le nombre de femmes participantes par année et par pays, nous avons effectué la requête suivante :

*Requêtes pour le nombre de femmes participantes par pays et par année*

```
select count(*)
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) join
regions r USING(noc)
where region='nom_du_pays' and sex='F'
group by year
order by year;
```

SQL



Comme précédemment, il faut remplacer nom\_du\_pays.

Suite à cela nous avons le graphique suivant, avec en abscisse les années et en ordonnée le nombre de femmes participantes.

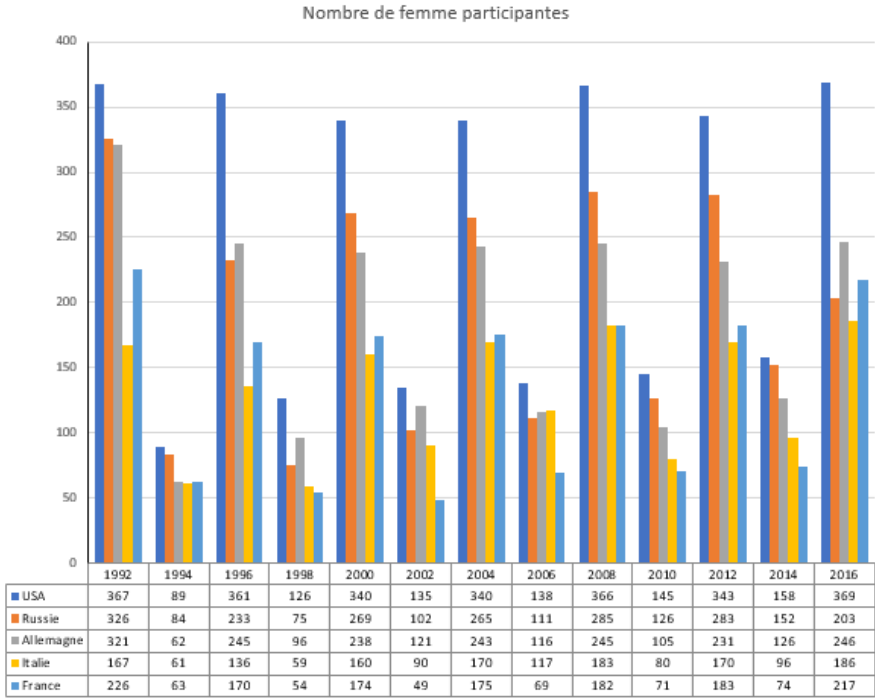


Figure 8: Nombre de participants par pays et par année

Question IV :

Choisir 5 pays parmi ces 15 et représenter l'évolution de la proportion de femmes participantes

Pour étudier la proportion de femmes participantes par année et par pays, nous devons effectué un calcul de proportion avec le nombre de participants et le nombre de femme participantes.

Suite à cela nous obtenons le graphique suivant, avec en abscisse les années et en ordonnée la proportion de femmes participantes.

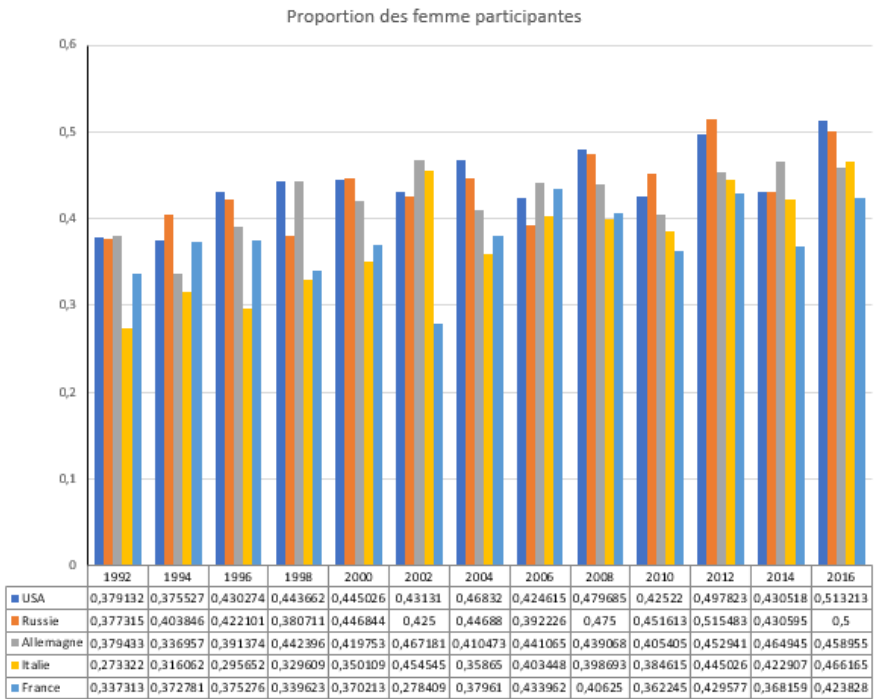


Figure 9: Nombre de participants par pays et par année

Nous pouvons également calculer le pourcentage d'évolution de cette proportion entre les années.

Nous pouvons calculer cela grâce à la formule ci-dessus :

$$p = e / N$$

Avec  $p$  = proportion,  $e$  = effectif du caractère recherché et  $N$  = effectif total

Suite aux calculs de l'évolution des proportions des femmes participantes, nous obtenons le graphique suivant :

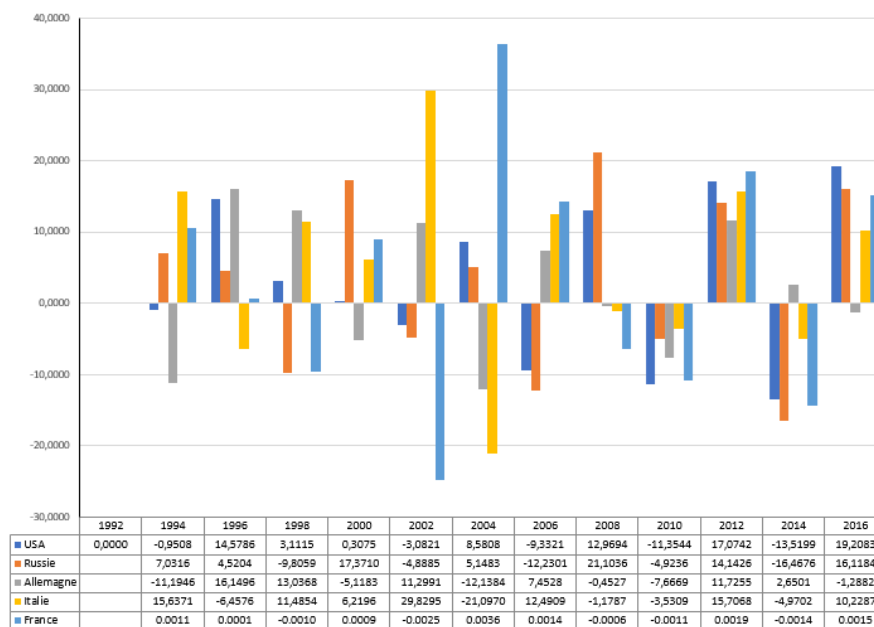


Figure 10: Graphique représentant l'évolution de la proportion des femmes participantes (en %)

Pour ce faire nous avons utilisé la formule suivante :

$$\% = ((Vf - Vi) / Vi) * 100$$

Avec % = pourcentage cherché,  $Vf$  = valeur finale et  $Vi$  = valeur initiale

Par valeur d'arrivée, nous entendons valeur de l'année que l'on cherche. Et par valeur de départ, nous entendons valeur de l'année dernière.

Par exemple, pour les USA (ou tout autre pays), pour le premier calcul :

- $Va$  = 1993
- $Vb$  = 1992

## Question V :

## Choisir 5 pays parmi ces 15 et représenter la proportion de médaillées parmi les femmes

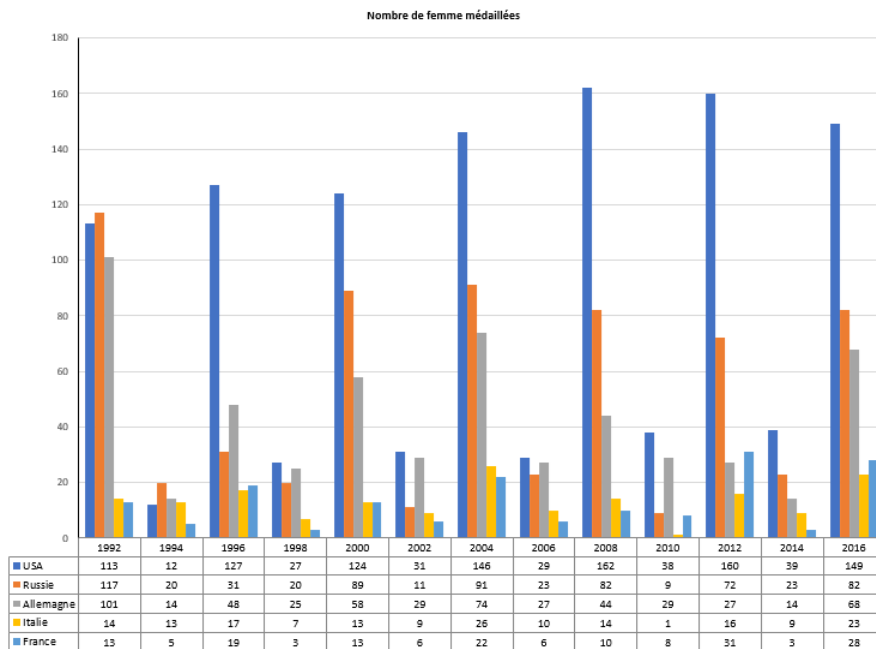
Pour étudier la proportion de médaillées parmi les femmes par année et par pays, nous devons effectuer un calcul de proportion avec le nombre de femme participantes et le nombre de femme médaillées. Pour savoir le nombre de femme médaillées, il faut utiliser la requête suivante :

### Requêtes pour le nombre de femmes médaillées par pays et par année

```
select count(*)
FROM equipe e JOIN athlete a USING(eno) JOIN participe p USING(ano) JOIN competition c USING(cno) join
regions r USING(noc)
where region='USA' and sex='F' and medal <> 'NA'
group by year
order by year;
```

SQL

Suite à cela nous obtenons le graphique suivant, avec en abscisse les années et en ordonnée le nombre de femmes médaillées.



*Figure 12: Nombre de femme médaillées par pays et par année*

Maintenant nous devons calculer la proportion de femme médaillées par rapport au nombre de femmes participantes. Nous utilisons donc la formule vue au dessus et nous obtenons le graphique ci-dessous :

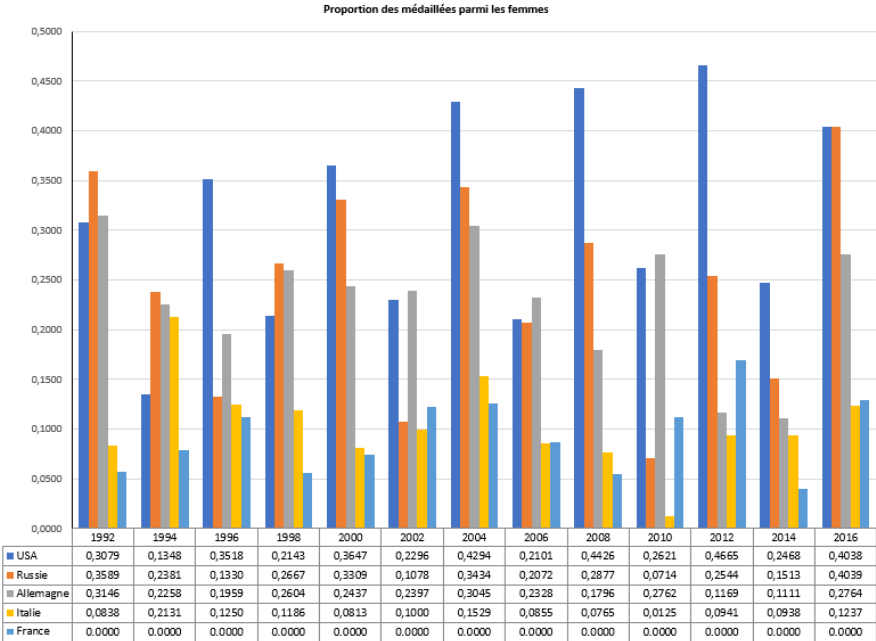


Figure 13: Proportion des médaillées parmi les femmes.

Suite aux calculs de l'évolution des proportions des femmes médaillées, que nous avons effectué grâce à la formule utilisée à la question IV, nous obtenons le graphique suivant :

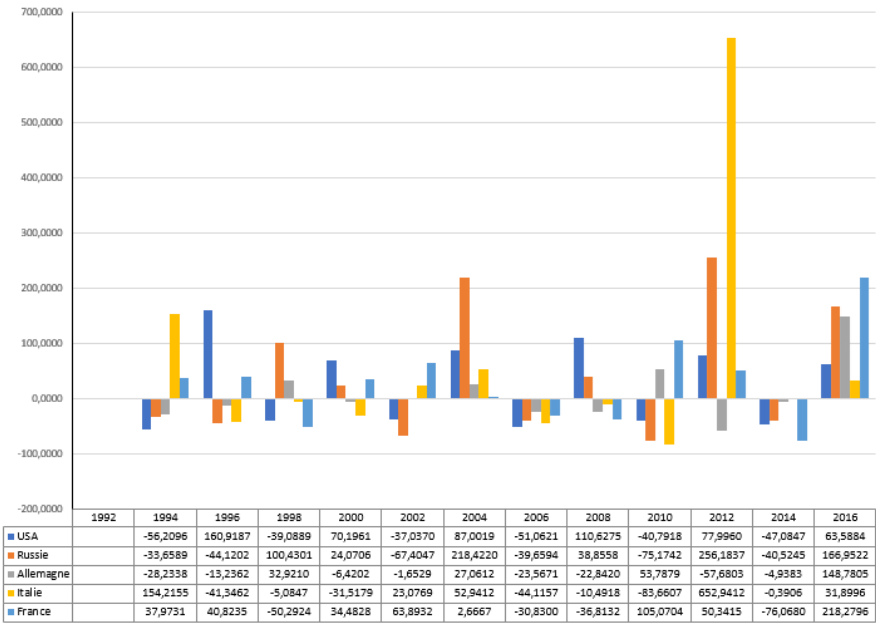


Figure 14: Graphique représentant l'évolution de la proportion des médaillées parmi les femmes (en %)

Question VI :

Choisir 5 pays parmi ces 15 et représenter la proportion de femme médaillées parmi les médaillés

Pour étudier la proportion de femmes médaillées par année et par pays, nous devons calculer la proportion de femme médaillées par rapport au nombre de médaillés. Nous utilisons donc la formule vue au dessus et nous obtenons le graphique ci-dessous :

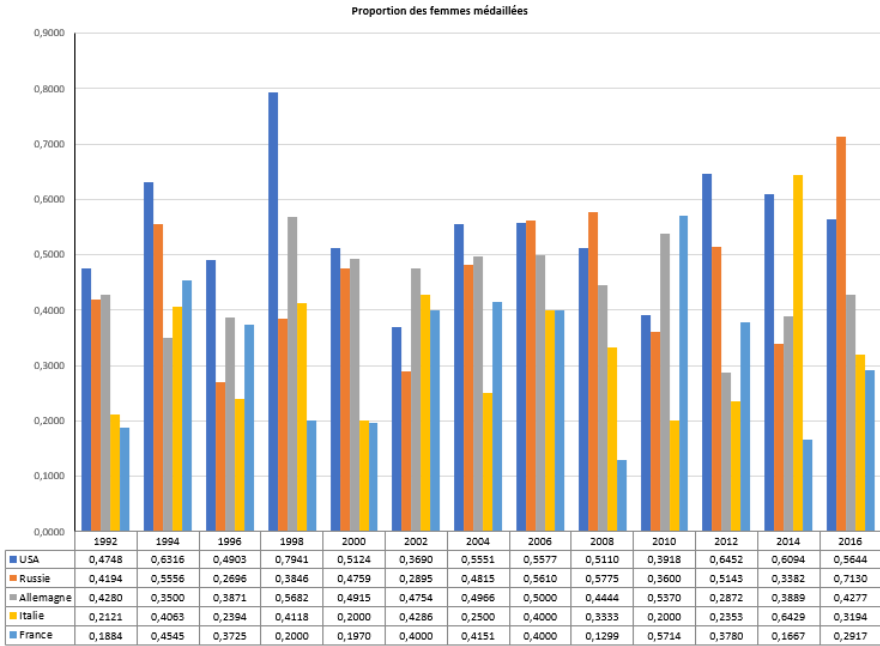


Figure 15: Proportion des femmes médaillées parmi les médaillés

Suite aux calculs de l'évolution des proportions des femmes médaillées, que nous avons effectué grâce à la formule utilisée à la question IV, nous obtenons le graphique suivant :

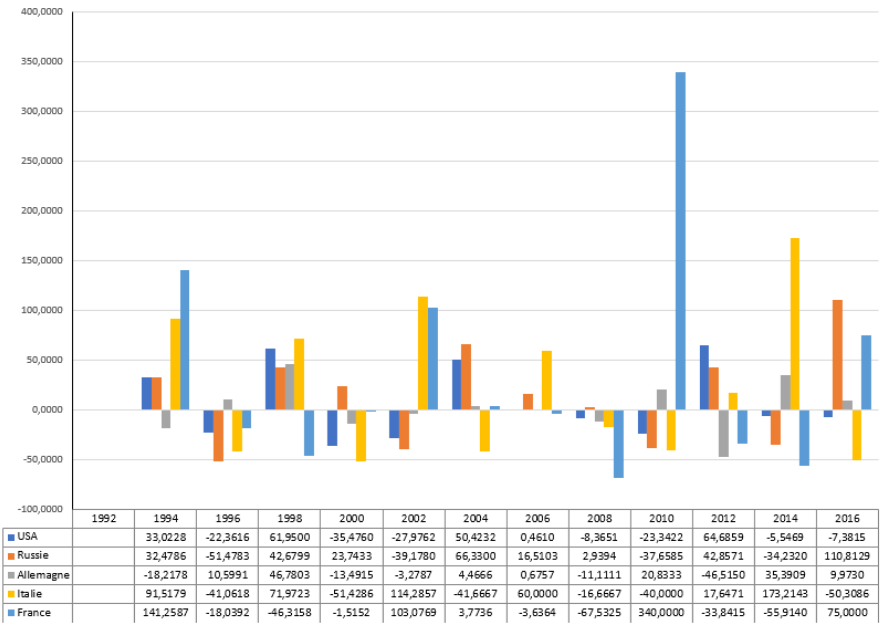


Figure 16: Graphique représentant l'évolution de la proportion des femmes médaillées (en %)