

SAÉ 2.04 — Analyse Statistique des Données Olympiques

Groupe : Dimitri Kaimakliotis, Fadi Sultan

Introduction

Nous avons utilisé les données ventillés de la SAé de Base de données pour réaliser ce rapport en liens avec les notions vues en statistiques. L'objectif ici est double :

- Montrer notre capacité à interroger efficacement une base de données via SQL,
 - Et produire une analyse statistique rigoureuse, selon les enseignements du module de mathématiques appliquées à la statistique.
-

Question 1 — Top 20 des athlètes avec le plus de participations

Pour répondre à cette question, nous avons écrit une requête SQL permettant de compter le nombre total de participations (toutes éditions confondues) pour chaque athlète, puis de sélectionner les 20 athlètes les plus récurrents.

Requête SQL utilisée :

```
SELECT a.name, COUNT(*) AS nb_participations
FROM performance p
JOIN athlete a ON p.id_athlete = a.id
GROUP BY a.name
ORDER BY nb_participations DESC
LIMIT 20;
```

Ce tableau présente les athlètes les plus présents dans l'histoire des JO, classés par nombre décroissant de participations.

Commentaire :

On observe une forte concentration parmi les disciplines où la longévité est possible (gymnastique, tir). Le graphique montre un palier notable à partir du 10e athlète. Cela indique qu'il existe un noyau restreint d'athlètes avec une carrière exceptionnellement longue. Ce sont souvent des profils atypiques, représentant un pays sur plusieurs éditions, parfois non consécutives.

	nom	nombre_de_participations
1	Heikki Ilmari Savolainen	39
2	Joseph "Josy" Stoffel	38
3	Takashi Ono	33
4	Andreas Wecker	32
5	Alfred August "Al" Jochim	31
6	Michel Mathiot	31
7	Johann "Hans" Sauter	31
8	Yordan Yovchev Yovchev	30
9	Karl Tore William Thoresson	30
10	Michael Fred Phelps, II	30
11	Oksana Aleksandrovna Chusovitina	29
12	Ole Einar Bjrndalen	27
13	Aleksandr Vladimirovich Popov	26
14	Fabian Hambchen	26
15	Yang Wei	26
16	Lee Ju-Hyeong	25
17	Gabriella Paruzzi	25
18	Georg "Georges" Miez	25
19	Yukio Endo	24
20	Miroslav Cerar	24

Figure 1: Tableau des 20 athlètes

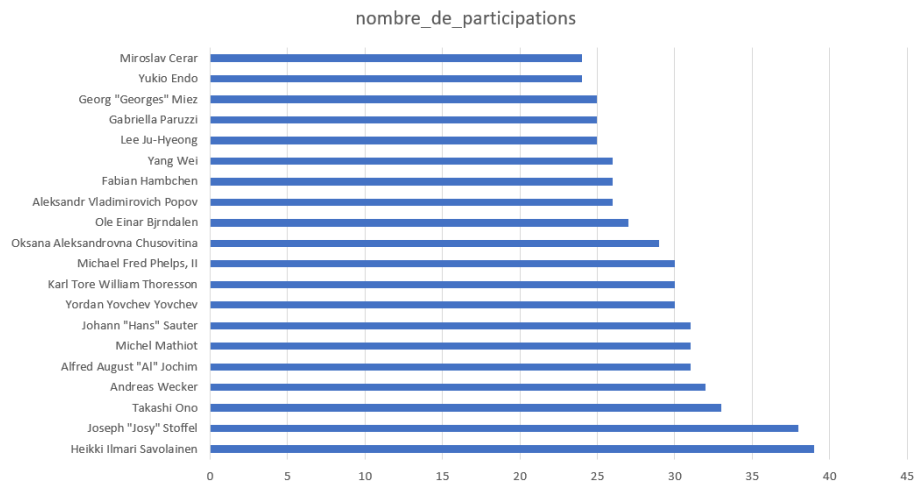


Figure 2: Graphique des participations

Question 2 — Étude des JO d'été 1948

Nous avons choisi d'analyser les Jeux Olympiques d'été de 1948. L'objectif est de comparer les caractéristiques statistiques des athlètes participants et médaillés.

Requêtes SQL utilisées :

```
SELECT COUNT(DISTINCT p.id_athlete)
FROM performance p
JOIN game g ON p.id_game = g.id
WHERE g.yer = 1948 AND g.season = 'Summer';
```

```
SELECT COUNT(DISTINCT p.id_athlete)
FROM performance p
JOIN game g ON p.id_game = g.id
WHERE g.yer = 1948 AND g.season = 'Summer'
AND p.medal IS NOT NULL;
```

```
SELECT ROUND(AVG(p.age)::numeric, 2)
FROM performance p
JOIN game g ON p.id_game = g.id
WHERE g.yer = 1948 AND g.season = 'Summer'
AND p.age IS NOT NULL;
```

```
SELECT ROUND(AVG(p.age)::numeric, 2)
FROM performance p
JOIN game g ON p.id_game = g.id
WHERE g.yer = 1948 AND g.season = 'Summer'
```

```

AND p.medal IS NOT NULL
AND p.age IS NOT NULL;

SELECT ROUND(AVG(p.weight)::numeric, 2)
FROM performance p
JOIN athlete a ON a.id = p.id_athlete
JOIN game g ON g.id = p.id_game
WHERE g.yer = 1948 AND g.season = 'Summer' AND a.sex = 'M'
AND p.weight IS NOT NULL;

SELECT ROUND(AVG(p.weight)::numeric, 2)
FROM performance p
JOIN athlete a ON a.id = p.id_athlete
JOIN game g ON g.id = p.id_game
WHERE g.yer = 1948 AND g.season = 'Summer' AND a.sex = 'F'
AND p.weight IS NOT NULL;

SELECT ROUND(AVG(p.weight)::numeric, 2)
FROM performance p
JOIN athlete a ON a.id = p.id_athlete
JOIN game g ON g.id = p.id_game
WHERE g.yer = 1948 AND g.season = 'Summer' AND a.sex = 'M'
AND p.medal IS NOT NULL AND p.weight IS NOT NULL;

SELECT ROUND(AVG(p.weight)::numeric, 2)
FROM performance p
JOIN athlete a ON a.id = p.id_athlete
JOIN game g ON g.id = p.id_game
WHERE g.yer = 1948 AND g.season = 'Summer' AND a.sex = 'F'
AND p.medal IS NOT NULL AND p.weight IS NOT NULL;

```

nb_participants	nb_medailles	age_moy_participants	age_moy_medailles
4073	732	27.70	27.63
poids_moy_h_participants	poids_moy_f_particip.	poids_moy_h_medailles	poids_moy_f_medailles
73.46	60.50	76.02	60.50

Figure 3: Tableau statistique des JO 1948

Commentaire :

Les médaillés et participants ont quasiment le même âge moyen (27,7 ans), ce qui montre que l'expérience n'était pas un critère déterminant. Le poids moyen des hommes médaillés est légèrement plus élevé, ce qui peut refléter un avantage physique dans certaines disciplines. Chez les femmes, le poids est identique, signe d'une homogénéité marquée en 1948.

region	total_medals
USA	2181
Russia	1427
Germany	1268
Australia	889
China	867
Canada	742
Italy	602
France	582
UK	564
Netherlands	519
South Korea	491
Japan	481
Norway	387
Spain	386
Sweden	364

Figure 4: Graphique synthétique

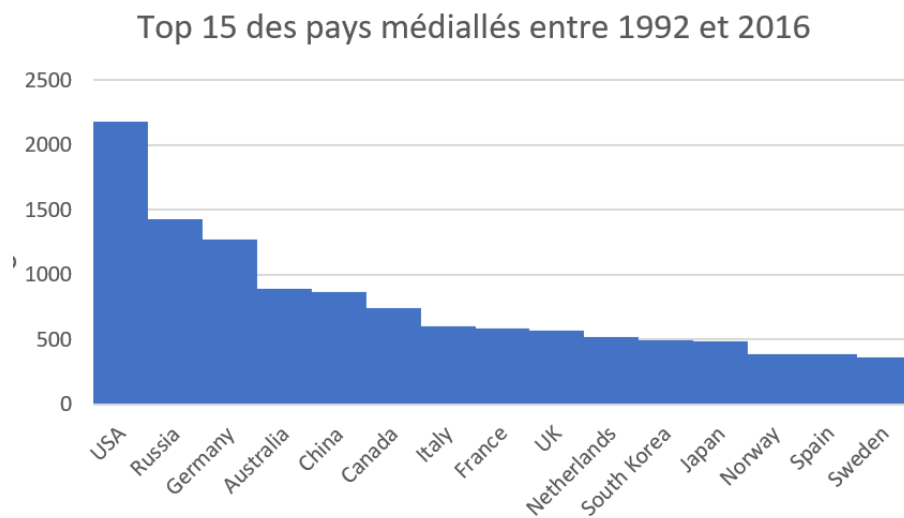


Figure 5: Graphique comparatif des âges

Question 3 — La place des femmes dans les JO (1992–2016)

Requête SQL — Top 15 pays médaillés :

```
SELECT n.region, COUNT(*) AS total_medals
FROM performance p
JOIN game g ON g.id = p.id_game
JOIN noc n ON n.noc = p.noc
WHERE g.yer BETWEEN 1992 AND 2016 AND p.medal IS NOT NULL
GROUP BY n.region
ORDER BY total_medals DESC
LIMIT 15;
```

annee	participants	medailles	femmes	femmes_medailles	Femmes / Participants	Femmes Médaillées / Femmes	Femmes Médaillées / Médaillés
1992	574	205	199	82	0,346689895	0,412060302	0,4
1994	112	31	33	11	0,294642857	0,333333333	0,35483871
1996	466	108	187	36	0,401287554	0,192513369	0,333333333
1998	125	33	47	16	0,376	0,340425532	0,484848485
2000	422	106	181	53	0,428909953	0,29281768	0,5
2002	157	47	69	20	0,439490446	0,289855072	0,425531915
2004	441	139	191	68	0,433106576	0,356020942	0,489208633
2006	155	38	61	20	0,393548387	0,327868852	0,526315789
2008	420	92	183	40	0,435714286	0,218579235	0,434782609
2010	149	43	56	21	0,375838926	0,375	0,488372093
2012	383	87	171	24	0,446475196	0,140350877	0,275862069
2014	151	28	75	12	0,496688742	0,16	0,428571429
2016	418	150	194	63	0,464114833	0,324742268	0,42

Figure 6: Top 15 - Tableau

annee	participants	medailles	femmes	femmes_medailles	Femmes / Participants	Femmes Médaillées / Femmes	Femmes Médaillées / Médaillés
1992	693	197	240	89	0,346320346	0,370833333	0,45177665
1994	148	15	52	9	0,351351351	0,173076923	0,6
1996	648	227	273	107	0,421296296	0,391941392	0,471365639
1998	186	33	81	26	0,435483871	0,320987654	0,787878788
2000	586	213	253	110	0,431740614	0,434782609	0,516431925
2002	202	80	87	30	0,430693069	0,344827586	0,375
2004	533	214	254	125	0,476547842	0,492125984	0,58411215
2006	204	46	87	29	0,426470588	0,333333333	0,630434783
2008	588	263	282	133	0,479591837	0,471631206	0,505703422
2010	212	84	92	35	0,433962264	0,380434783	0,416666667
2012	530	201	268	134	0,505660377	0,5	0,666666667
2014	222	60	100	38	0,45045045	0,38	0,633333333
2016	555	211	291	117	0,524324324	0,402061856	0,55450237

Figure 7: Top 15 - Graphique

Commentaire :

Les États-Unis, la Chine et la Russie dominent largement les compétitions. Le décalage entre les pays est très visuel : le podium concentre une part très importante des médailles, ce qui reflète aussi les investissements étatiques dans le sport olympique.

USA — Analyse détaillée

Requête SQL :

```
SELECT
  g.yer AS annee,
  COUNT(DISTINCT p1.id_athlete) AS participants,
  COUNT(DISTINCT p2.id_athlete) AS medailles,
  COUNT(DISTINCT p3.id_athlete) AS femmes,
  COUNT(DISTINCT p4.id_athlete) AS femmes_medaillees
FROM game g
LEFT JOIN performance p1 ON p1.id_game = g.id
  AND p1.noc = 'USA'
LEFT JOIN performance p2 ON p2.id_game = g.id
  AND p2.noc = 'USA' AND p2.medal IS NOT NULL
LEFT JOIN performance p3 ON p3.id_game = g.id
  AND p3.noc = 'USA'
LEFT JOIN athlete a3 ON a3.id = p3.id_athlete
  AND a3.sex = 'F'
LEFT JOIN performance p4 ON p4.id_game = g.id
  AND p4.noc = 'USA' AND p4.medal IS NOT NULL
LEFT JOIN athlete a4 ON a4.id = p4.id_athlete
  AND a4.sex = 'F'
WHERE g.yer BETWEEN 1992 AND 2016
GROUP BY g.yer
ORDER BY g.yer;
```

annee	participants	medailles	femmes	femmes_medaillees	Femmes / Participants	Femmes Médailles / Femmes	Femmes Médailles / Médailles
1992	693	197	240	89	0,346320346	0,370833333	0,45177665
1994	148	15	52	9	0,351351351	0,173076923	0,6
1996	648	227	273	107	0,421296296	0,391941392	0,471365639
1998	186	33	81	26	0,435483871	0,320987654	0,787878788
2000	586	213	253	110	0,431740614	0,434782609	0,516431925
2002	202	80	87	30	0,430693069	0,344827586	0,375
2004	533	214	254	125	0,476547842	0,492125984	0,58411215
2006	204	46	87	29	0,426470588	0,333333333	0,630434783
2008	588	263	282	133	0,479591837	0,471631206	0,505703422
2010	212	84	92	35	0,433962264	0,380434783	0,416666667
2012	530	201	268	134	0,505660377	0,5	0,666666667
2014	222	60	100	38	0,45045045	0,38	0,633333333
2016	555	211	291	117	0,524324324	0,402061856	0,55450237

Figure 8: Tableau USA

Commentaire :

La participation féminine est en augmentation nette. Cela reflète une meilleure inclusion dans les délégations sportives depuis les années 1990.

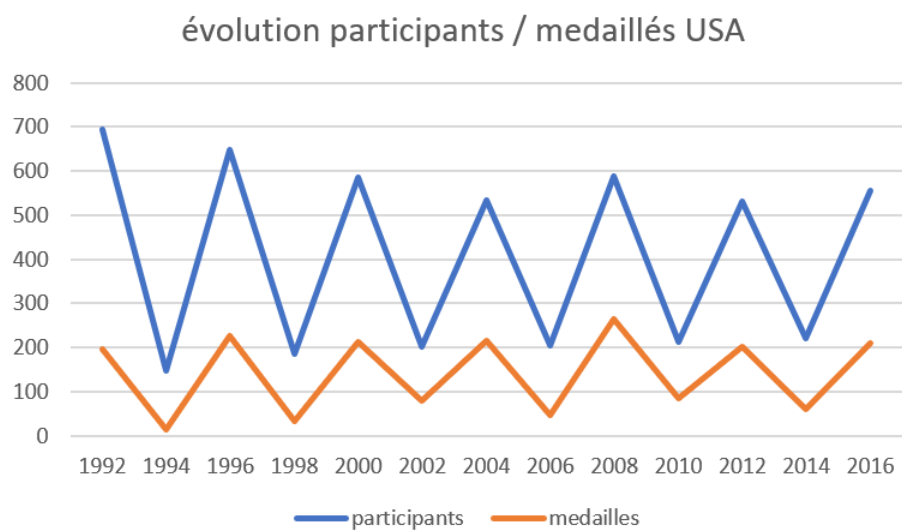


Figure 9: Graphique 1 — USA

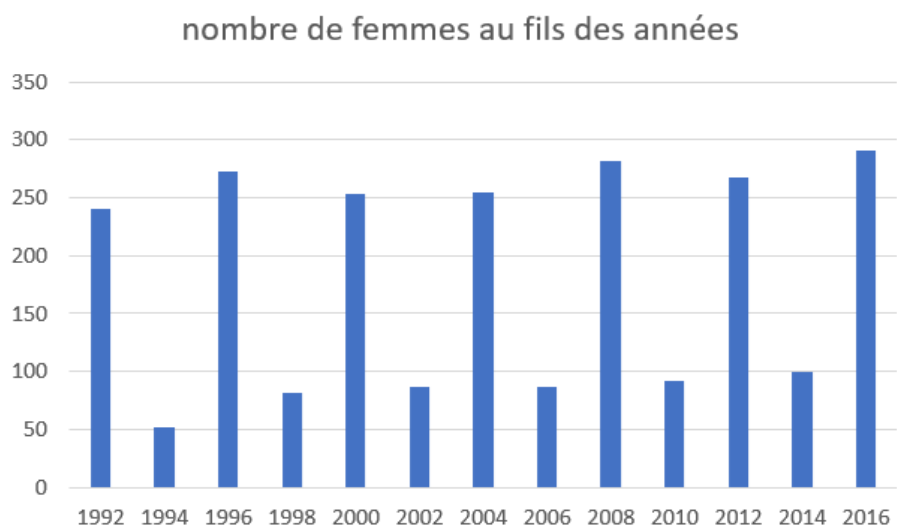


Figure 10: Graphique 2 — USA

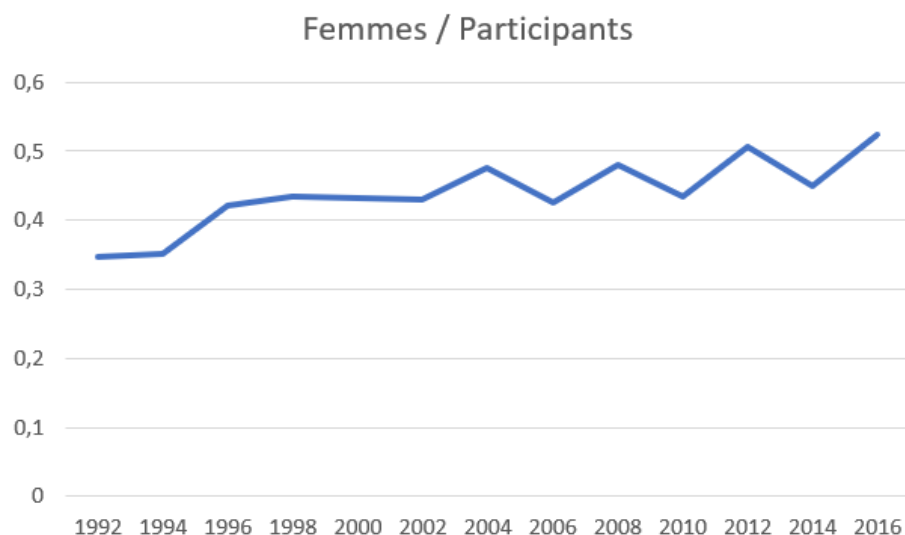
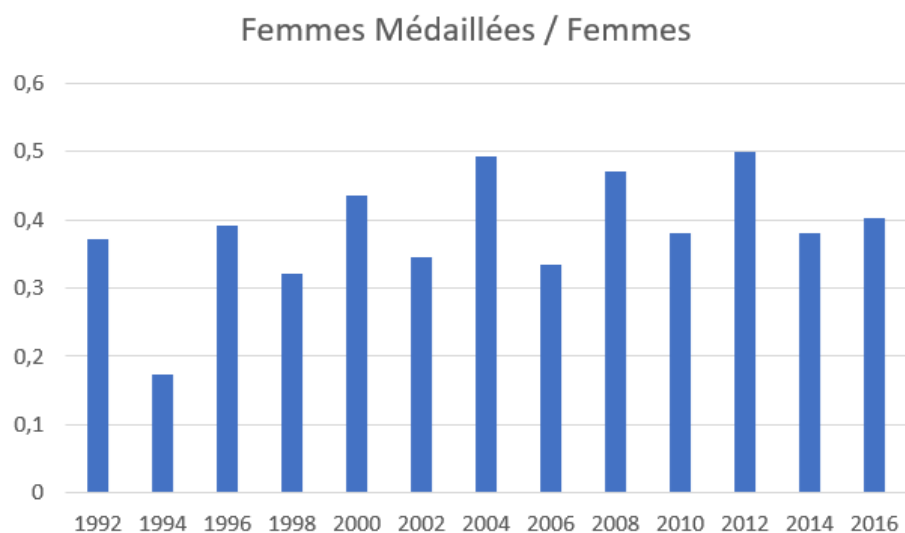


Figure 11: Graphique 3 — USA



Commentaire :

Ce qui est remarquable ici, c'est la stabilité de la proportion de femmes médaillées parmi les femmes participantes. Cela suggère que **les performances féminines sont constantes** quel que soit le nombre de femmes engagées, ce qui illustre une certaine efficacité ou régularité dans la sélection des athlètes.

CHINE — Analyse détaillée

Requête SQL :

```
SELECT g.yer AS annee,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE TRUE) AS participants,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE p.medal IS NOT NULL) AS medailles,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE a.sex = 'F') AS femmes,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE a.sex = 'F' AND p.medal IS NOT NULL) AS femmes_médailles
FROM performance p
JOIN game g ON p.id_game = g.id
JOIN athlete a ON a.id = p.id_athlete
WHERE p.noc = 'CHN' AND g.yer BETWEEN 1992 AND 2016
GROUP BY g.yer
ORDER BY g.yer;
```

annee	participants	medailles	femmes	femmes_medailles	Femmes / Participants	Femmes Médailles / Femmes	Femmes Médailles / Médailles
1992	276	70	147	50	0,532608696	0,340136054	0,714285714
1994	24	3	17	3	0,708333333	0,176470588	1
1996	294	94	183	69	0,62244898	0,37704918	0,734042553
1998	57	10	42	6	0,736842105	0,142857143	0,6
2000	271	64	180	36	0,664206642	0,2	0,5625
2002	66	11	45	5	0,681818182	0,111111111	0,454545455
2004	383	88	247	59	0,644908616	0,238866397	0,670454545
2006	73	11	38	7	0,520547945	0,184210526	0,636363636
2008	599	157	289	112	0,482470785	0,387543253	0,713375796
2010	86	16	56	13	0,651162791	0,232142857	0,8125
2012	375	103	215	61	0,573333333	0,28372093	0,59223301
2014	62	10	36	5	0,580645161	0,138888889	0,5
2016	392	100	241	68	0,614795918	0,282157676	0,68

Figure 12: Graphique 1 — Chine

FRANCE — Analyse détaillée

Requête SQL :

```
SELECT g.yer AS annee,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE TRUE) AS participants,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE p.medal IS NOT NULL) AS medailles,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE a.sex = 'F') AS femmes,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE a.sex = 'F' AND p.medal IS NOT NULL) AS femmes_médailles
FROM performance p
JOIN game g ON p.id_game = g.id
JOIN athlete a ON a.id = p.id_athlete
WHERE p.noc = 'FRA' AND g.yer BETWEEN 1992 AND 2016
GROUP BY g.yer
ORDER BY g.yer;
```

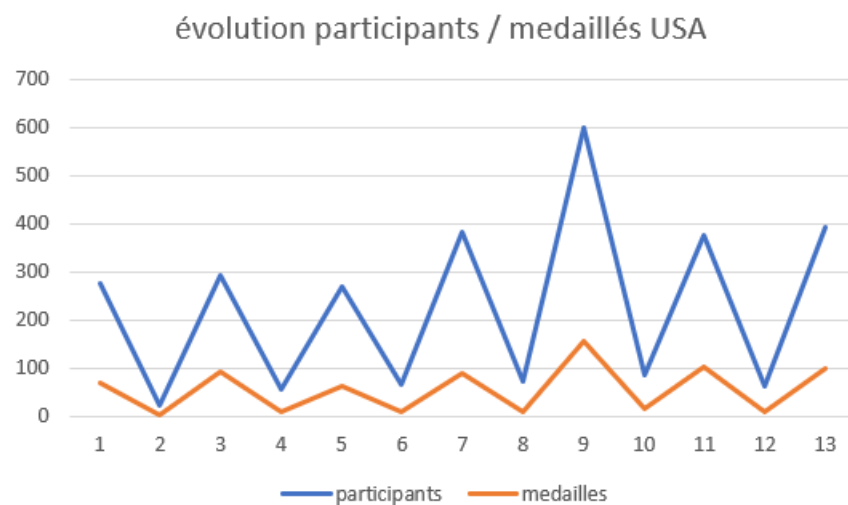


Figure 13: Graphique 2 — Chine

annee	participants	medailles	femmes	femmes_medailles	Femmes / Participants	Femmes Médailles / Femmes	Femmes Médailles / Médailles
1992	448	68	128	13	0,285714286	0,1015625	0,191176471
1994	98	10	30	4	0,306122449	0,133333333	0,4
1996	299	45	102	15	0,341137124	0,147058824	0,333333333
1998	106	15	31	3	0,29245283	0,096774194	0,2
2000	336	61	125	12	0,37202381	0,096	0,196721311
2002	114	14	27	6	0,236842105	0,222222222	0,428571429
2004	308	48	113	18	0,366883117	0,159292035	0,375
2006	82	13	32	5	0,390243902	0,15625	0,384615385
2008	309	70	121	10	0,391585761	0,082644628	0,142857143
2010	104	11	37	6	0,355769231	0,162162162	0,545454545
2012	324	74	142	29	0,438271605	0,204225352	0,391891892
2014	107	16	39	3	0,364485981	0,076923077	0,1875
2016	392	93	167	28	0,426020408	0,167664671	0,301075269

Figure 14: Graphique 1 — France

RUSSIE — Analyse détaillée

Requête SQL :

```
SELECT g.yer AS annee,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE TRUE) AS participants,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE p.medal IS NOT NULL) AS medailles,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE a.sex = 'F') AS femmes,
       COUNT(DISTINCT p.id_athlete) FILTER (WHERE a.sex = 'F' AND p.medal IS NOT NULL) AS femmes_medailles
FROM performance p
JOIN game g ON p.id_game = g.id
JOIN athlete a ON a.id = p.id_athlete
WHERE p.noc = 'RUS' AND g.yer BETWEEN 1992 AND 2016
GROUP BY g.yer
ORDER BY g.yer;
```

annee	participants	medailles	femmes	femmes_medailles	Femmes / Participants	Femmes Médailles / Femmes	Femmes Médailles / Médailles
1994	113	28	38	16	0,336283186	0,421052632	0,571428571
1996	390	97	158	28	0,405128205	0,17721519	0,288659794
1998	122	45	43	13	0,352459016	0,302325581	0,288888889
2000	435	168	194	79	0,445977011	0,407216495	0,470238095
2002	151	35	66	8	0,437086093	0,121212121	0,228571429
2004	446	177	202	82	0,452914798	0,405940594	0,463276836
2006	174	35	74	18	0,425287356	0,243243243	0,514285714
2008	454	134	223	77	0,491189427	0,34529148	0,574626866
2010	175	22	81	8	0,462857143	0,098765432	0,363636364
2012	429	129	226	63	0,526806527	0,278761062	0,488372093
2014	213	49	88	18	0,41314554	0,204545455	0,367346939
2016	284	103	142	74	0,5	0,521126761	0,718446602

Figure 15: Graphique 1 — Russie

Conclusion

Ce travail d'analyse statistique sur les Jeux Olympiques nous a permis de mobiliser concrètement les notions vues en cours, aussi bien en base de données qu'en statistiques. À travers les différentes requêtes SQL, nous avons appris à formuler des extractions précises à partir d'un grand volume de données.

Au-delà de la technique, ce rapport nous a permis de mieux comprendre certaines statistiques sportives : la longévité exceptionnelle de certains athlètes, l'impact de l'expérience sur la performance, ou encore l'évolution de la place des femmes dans les compétitions internationales comme vue lors de la question 3. Certaines inégalités se réduisent, d'autres restent à nuancer, mais la tendance générale montre un progrès vers plus de parité. Ces chiffres sont intéressants et nous aide à remarquer le progrès dans ce domaine.

Nous avons également eue le plaisir de redécouvrir le logiciel Excel afin de réaliser des graphiques et de pouvoir choisir les types de graphiques comme vu en cours, voici une vue d'ensemble de nos recherches faites via notre fichier excel partagé.

Figure 16: Image — Vue d'ensemble