

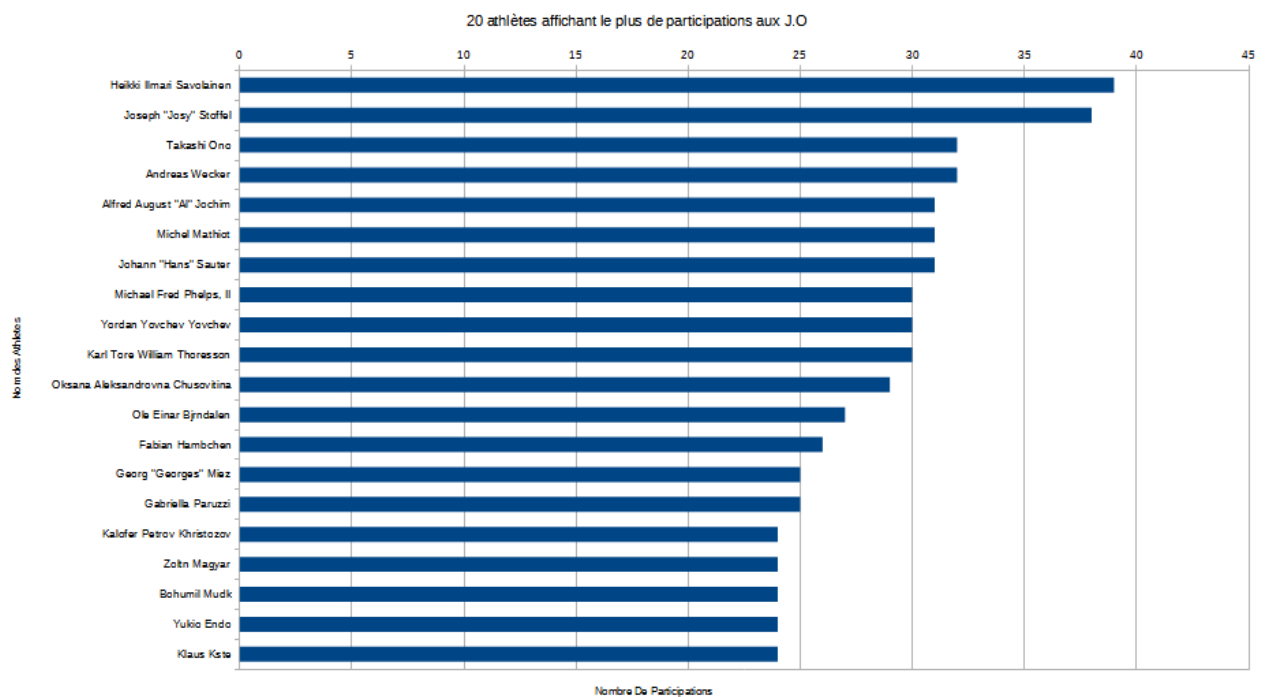
SAE2.04 - Statistiques

Question 1 : Déterminer les 20 athlètes affichant le plus de participations aux J.O. (tout sexe confondu)

Requete :

```
\copy (SELECT 'Question 1' as Question,a.id,a.name,count(*)
FROM participe as p , athlete as a
WHERE p.id=a.id
GROUP BY a.id,a.name
ORDER BY count(*) desc limit 20)
TO './Question1.csv' with (FORMAT CSV , Delimiter ',' , HEADER);
```

Graphique :



SAE2.04 - Statistiques

Question 2 : Pour cette question nous avons choisi les Jeux Olympiques d'Hiver de 1988 qui ont eu lieu au Canada.

Question 2a :

```
Requete: \copy (SELECT 'Question 2 a' as Question ,
pa.Region,AVG(p.age) as Moyenne ,count(Distinct p.id) as Nb
,min(p.age) as Min ,max(p.age) as Max
FROM participe as p , jo as j,TEAM as t , Pays as pa
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND j.annee=1988
AND j.season='Winter'GROUP BY pa.NOC,pa.Region )
TO 'Question2a.csv' WITH (format CSV , DELIMITER ',' , HEADER);
```

Tableau Statistique :

<i>Pays</i>	<i>Moyenne d'âge</i>	<i>Nombre De Sportifs</i>	<i>Min âge</i>	<i>Max âge</i>
Andorra	17.11	4	16	18
Argentina	19.33	15	16	28
Australia	23.69	19	16	41
Austria	23.90	81	17	38
Belgium	22	1	22	22
Boliva	26.75	6	23	32

SAE2.04 - Statistiques

Bulgaria	24.08	27	18	34
Canada	23.63	112	17	32
Chile	18.63	5	16	22
China	19.60	13	11	24
Costa Rica	35	2	31	41
Curacao	33.67	2	33	35
Cyprus	18.13	3	17	20
Czech Republic	23.08	59	17	33
Denmark	20	1	20	20
Fiji	26	1	26	26
Finland	25.66	53	18	32
France	24.48	68	17	39
Germany	23.66	53	16	39
Germany	24.92	90	18	33
Greece	20.6	6	16	25
Guam	38	1	38	38
Guatemala	24.2	6	19	41
Hungary	24.14	5	19	29
Iceland	23.67	3	20	25
India	22.67	3	21	24

SAE2.04 - Statistiques

Italy	24.82	58	17	37
Jamaica	23.83	4	23	25
Japan	23.18	48	14	32
Lebanon	24.11	4	21	27
Liechtenstein	23.13	13	18	40
Luxembourg	24	1	24	24
Mexico	26.19	11	17	31
Monaco	29	3	29	29
Mongolia	21	3	20	23
Morocco	19	3	15	23
Netherlands	23.12	11	20	27
New Zealand	24.74	9	21	41
North Korea	19.33	6	12	22
Norway	24.5	63	20	36
Philippines	35	1	35	35
Poland	25.9	32	19	35
Portugal	22.75	5	18	28
Puerto Rico	29.41	9	17	41
Romania	21.28	11	17	29
Russia	24.03	101	16	35

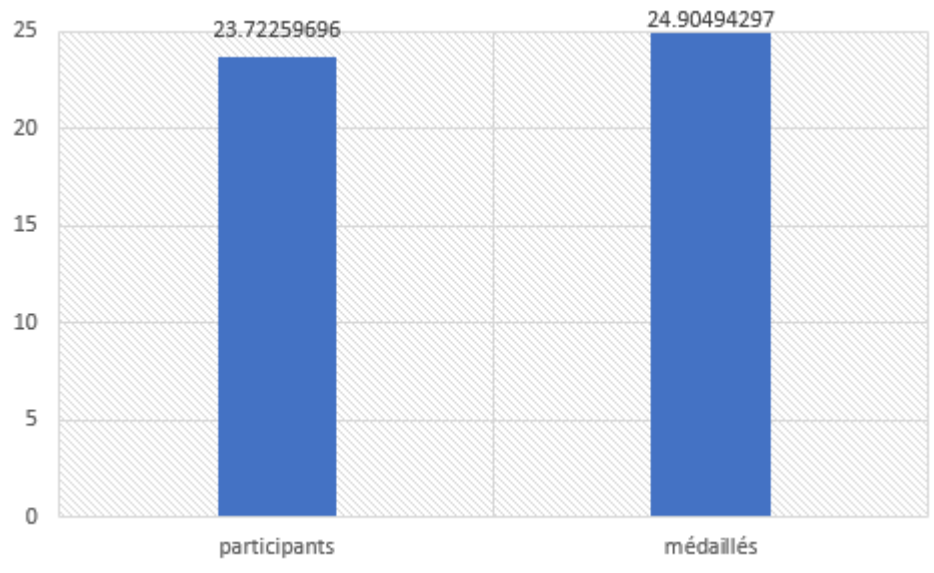
SAE2.04 - Statistiques

San Marino	21.75	5	18	30
Serbia	22.98	22	17	32
South Korea	20.71	22	15	25
Spain	22.62	12	19	32
Sweden	24.28	67	17	34
Switzerland	24.945	70	19	43
Taiwan	22.59	13	18	28
Turkey	20.47	8	17	26
UK	23.73	48	16	31
USA	23.92	117	17	39
Virgin Islands, US	36.48	6	14	52

Question 2b : Comparer l'âge moyen des médaillés et l'âge moyen des participants (tout sexe confondu).

Requete : \copy (SELECT 'Question 2 b' as Question , m.NomMedal<>'NA' as
Medaille,AVG(p.age) as Moyenne
FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND j.annee=1988
AND j.season='Winter'AND p.idM=m.idM
GROUP BY m.NomMedal<>'NA')
TO 'Question2b.csv' WITH (format CSV , DELIMITER ';' , HEADER);

SAE2.04 - Statistiques



On peut voir que l'âge moyen des participants est de 23.72 et que l'âge moyen des médaillés est de 24.90. Les deux valeurs obtenues sont très proches mais l'âge moyen des médaillés reste cependant au dessus de celui des participants.

Question 2c : Comparer le poids moyen des médaillés et le poids moyen des participants (distinguer les hommes et les femmes)

```
Requete : \copy (SELECT 'Question 2 c' as Question , m.NomMedal<>'NA' as
Medaillé,a.SEX,AVG(p.poid) as Moyenne ,count(p.id) as Nb
FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m, athlete as a
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND j.annee=1988
AND j.season='Winter'
AND p.idM=m.idM
AND p.ID=a.ID GROUP BY m.NomMedal<>'NA',a.SEX) TO 'Question2c.csv' WITH (format
CSV , DELIMITER ',' , HEADER);
```

Question 3a : a. Sur la période 1992 à 2016, faire la liste des 15 pays qui ont gagné le plus de médailles cumulées sur cette période (jeux d'été et hiver confondus).

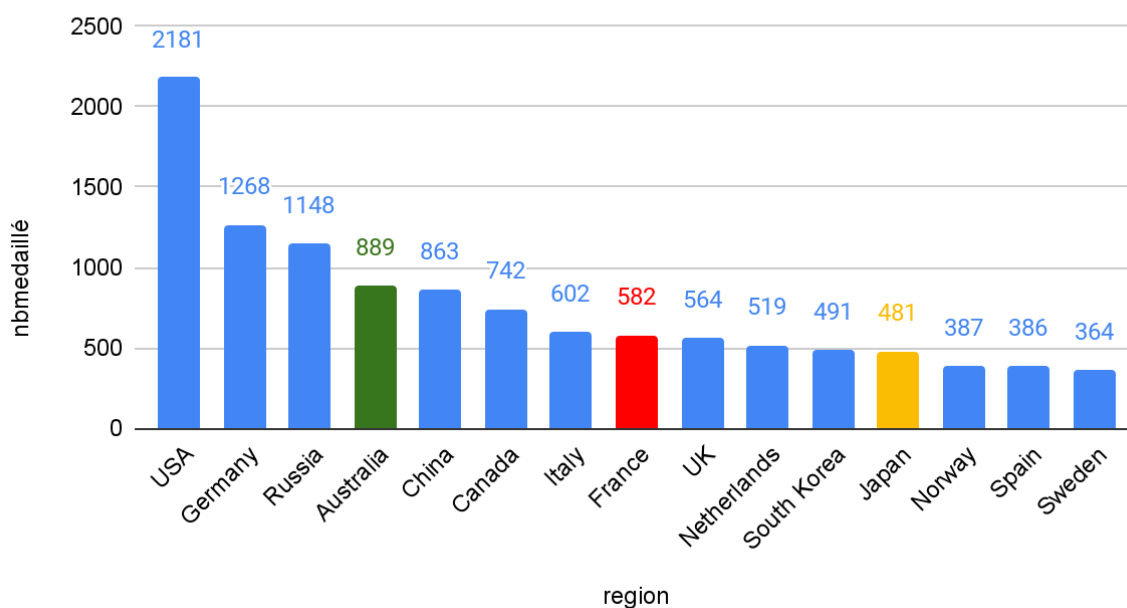
SAE2.04 - Statistiques

Requete :

```
\copy (SELECT 'Question 3 a' as Question ,pa.region, count(*) as NBMedaillé
FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m, athlete as a
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND j.annee>=1992 And j.annee<=2016
AND p.idM=m.idM
AND p.ID=a.ID
AND m.NomMedal<>'NA'
GROUP BY t.NOC,pa.region
ORDER BY count(*) desc
limit 15 )
TO 'Question3a.csv' WITH (format CSV , DELIMITER ',' , HEADER);
```

Graphique :

Nombre de médaillé par rapport à région



SAE2.04 - Statistiques

- Le graphique ci-dessus indique le nombre de médailles remporté par pays durant la période 1992 à 2016 en ne gardant que les 15 premiers.
- Calcul :
 - Moyenne :
 - $(2181 + 1268 + 1148 + 889 + 863 + 742 + 602 + 582 + 564 + 519 + 491 + 481 + 387 + 386 + 364) / 15 = 764,47$
 - Mediane et Quartiles :
 - Position de la Mediane = $(15 / 2) + 1$ (+1 car 15 impair) = 8 ;
 - Position du premier Quartile = $(15+1)/4$ (+1 car 15 impair) = 4 ;
 - Position du troisième Quartile = $((15+1)/4)*3$ (+1 car 15 impair) = 12 ;

SAE2.04 - Statistiques

- 2181 ; 1268 ; 1148 ; 889 (**12 donc troisième Quartile**) ; 863 ; 742 ; 602 ; **582 (8 donc mediane)** ; 564 ; 519 ; 491 ; **481(4 donc premier Quartile)** ; 387 ; 386 ; 364

- Analyse :

La moyenne de médaille remportée parmi ces 15 pays est d'environ 746 cependant on constate que les Etats-Unis ont remporté bien plus de médailles que les autres et va donc faire augmenter la moyenne c'est pour cela que nous avons décidé de vous représenter la médiane et les quartiles sur le graphes. La médiane est représenté en rouge et est égale á 582, cela signifie que 50% des pays ont au moins 582 médailles et que 50% ont moins de 582 médailles. Le premier quartile est représenté en jaune et est égale á 481, cela signifie que 75% des pays ont au moins 481 médailles et que 25% ont moins de 481 médailles. Et pour finir le troisième quartile est représenté en vert et est égale á 889, cela signifie que 25% des pays ont au moins 889 médailles et 75% ont moins de 889 médailles. Donc 50% des 15 premiers pays ont remporté entre 481 et 889 médailles.

Question 3b :

Pour cette question les 5 Pays que nous avons choisi sont la France, L'Allemagne, le Japon, La Corée du Sud , et le Royaume Uni.

Question 3bi : représenter l'évolution Du nombre de participants

Requete :

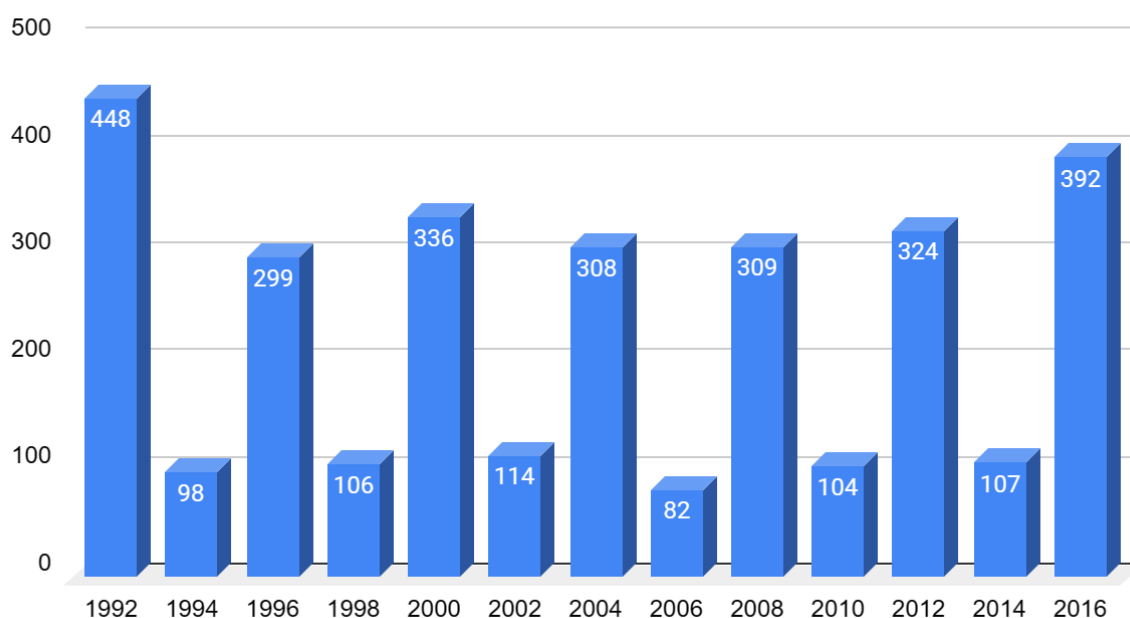
```
\copy (SELECT 'Question 3 b i' as Question ,pa.region, j.annee,count(Distinct p.id) as NbParticipants
```


SAE2.04 - Statistiques

Lorsque nous généront le graphique de nos données nous obtenons ceci, cependant nous avons décidé de découper notre graphique en 5 sous graphiques afin de mieux pouvoir étudier l'évolution du nombre de participants dans chaque pays.

France :

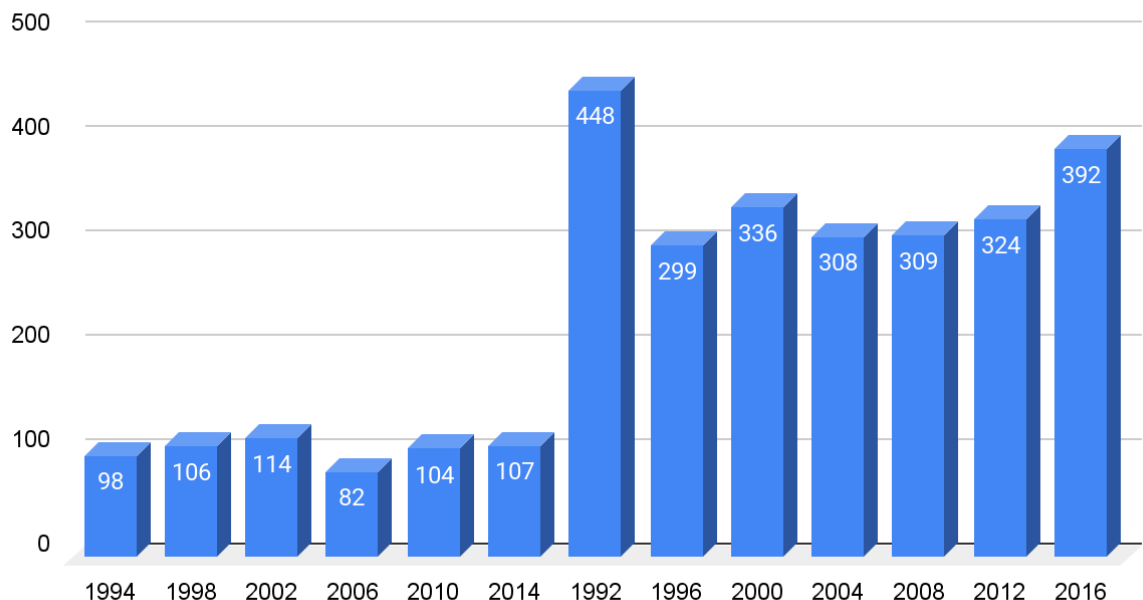
Nombre de Participants Français par An



- Tout d'abord on observe que notre graphique à l'air découpé en deux, en effet les années durant lesquelles il y a peu de participants correspondent aux JO d'hiver (Nous observons le même phénomène dans chaque pays c'est pour cela que les graphiques seront déjà découpés par la suite). Cependant si nous séparons les JO d'été et d'hiver nous obtenons des données qui sont de manière générale assez stables (sans valeurs extrêmes), nous pouvons donc calculer deux moyennes, la moyenne du nombre de participants aux JO d'été et la moyenne du nombre de participants aux JO d'hiver.

SAE2.04 - Statistiques

Nombre de Participants Français par An



- Moyenne Participation JO Hiver :
 - $(98 + 106 + 114 + 82 + 104 + 107) / 6 = 101,8$ participants en moyenne
 - Calcul de l'écart-type :
 - Racine Carrée($((98^2 + 106^2 + 114^2 + 82^2 + 104^2 + 107^2) / 6) - 101,8^2$) = 10,4

L'écart type est faible ce qui montre que nombre de participant ne varie pas énormément est qu'il reste autour de 101 participants

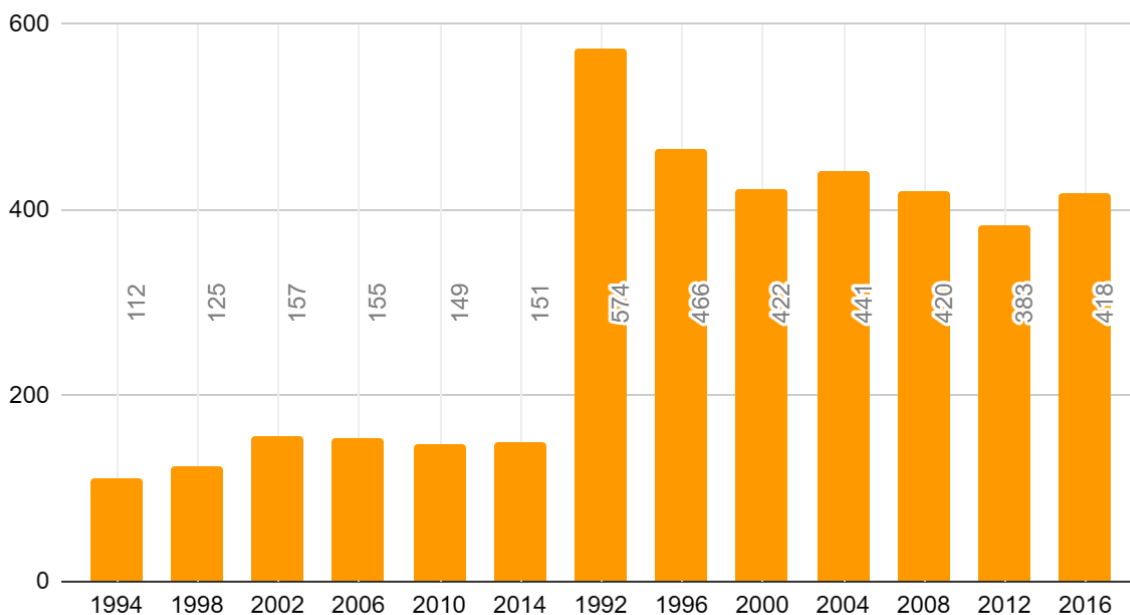
- Moyenne Participation JO Été :
 - $(448 + 299 + 336 + 308 + 309 + 324 + 392) / 7 = 345$
 - Calcul de l'écart type :
 - Racine Carrée($((448^2 + 299^2 + 336^2 + 308^2 + 309^2 + 324^2 + 392^2) / 7) - 345^2$) = 51,8

L'écart type dans ce cas est moins faible que pour les JO d'hiver car l'année 1992 est bien plus élevée que les autres, cependant l'écart type reste assez faible car il ne représente que 15% de la moyenne.

Allemagne :

SAE2.04 - Statistiques

Nombre de Participants Allemands par An



- Pour ce graphique nous avons déjà séparé les JO d'été et d'hiver, et on peut constater visuellement que le nombre de participations aux JO d'hiver a l'air plus stable qu'en été. Pour vérifier cela nous allons comparer les écarts-type
- Moyenne Participation JO Hiver :
 - $(112 + 125 + 157 + 155 + 149 + 151) / 6 = 141,5$ participants en moyenne
 - Calcul de l'écart-type :
 - Racine Carrée($((112^2 + 125^2 + 157^2 + 155^2 + 149^2 + 151^2) / 6) - 141,5^2$) = 16,8

L'écart type est faible (11% de la moyenne) ce qui montre que le nombre de participants ne varie pas énormément et qu'il reste autour de 141 participants
- Moyenne Participation JO Été :
 - $(574 + 466 + 422 + 441 + 420 + 383 + 418) / 7 = 446$ participants en moyenne
 - Calcul de l'écart type :
 - Racine Carrée($((574^2 + 466^2 + 422^2 + 441^2 + 420^2 + 383^2 + 418^2) / 7) - 446^2$) = 59,2

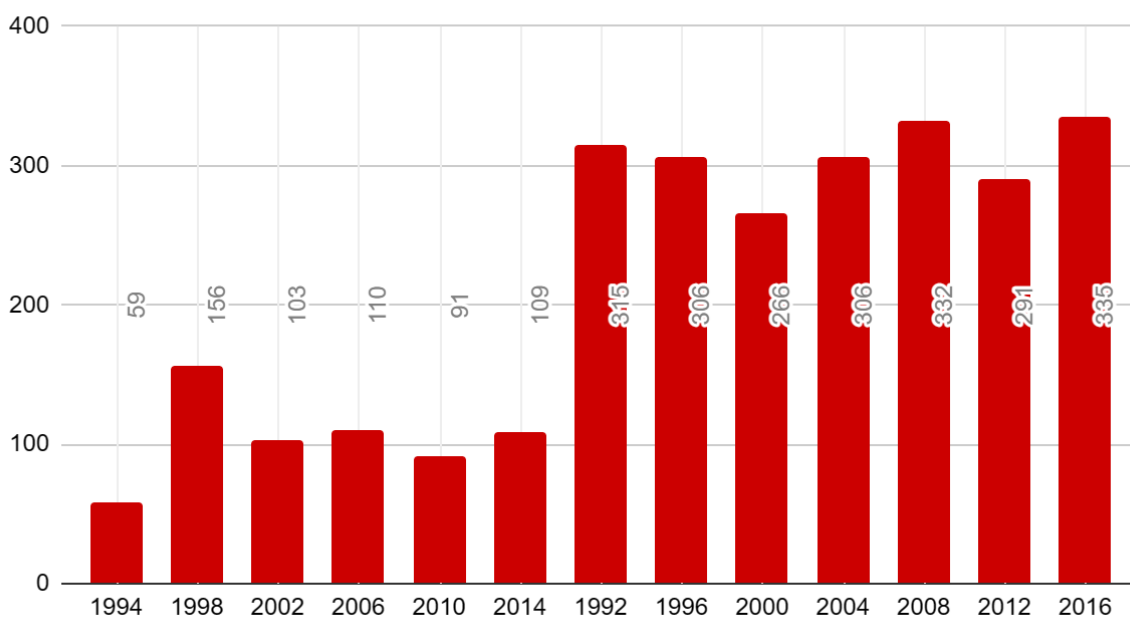
SAE2.04 - Statistiques

L'écart type est faible (11% de la moyenne) ce qui montre que le nombre de participants ne varie pas énormément, il reste autour de 446 participants.

Nous pouvons donc constater que ce que nous pensions au départ était erroné car après nos calculs les écarts types étaient plus ou moins équivalents par rapport à la moyenne de participants.

Japon :

Nombre de Participants Japonais par An



- Pour ce graphique nous avons déjà séparé les JO d'été et d'hiver, et on peut constater visuellement que contrairement aux deux premiers pays le nombre de participations aux JO d'hiver a l'air moins stable qu'en été. Pour vérifier cela nous allons comparer les écarts-type.
- Moyenne Participation JO Hiver :
 - $(59 + 156 + 103 + 110 + 91 + 109) / 6 = 104,6$ participants en moyenne
 - Calcul de l'écart-type :

SAE2.04 - Statistiques

- Racine Carré(((59² + 156² + 103² + 110² + 91² + 109²) / 6) - 141,5²) = 28

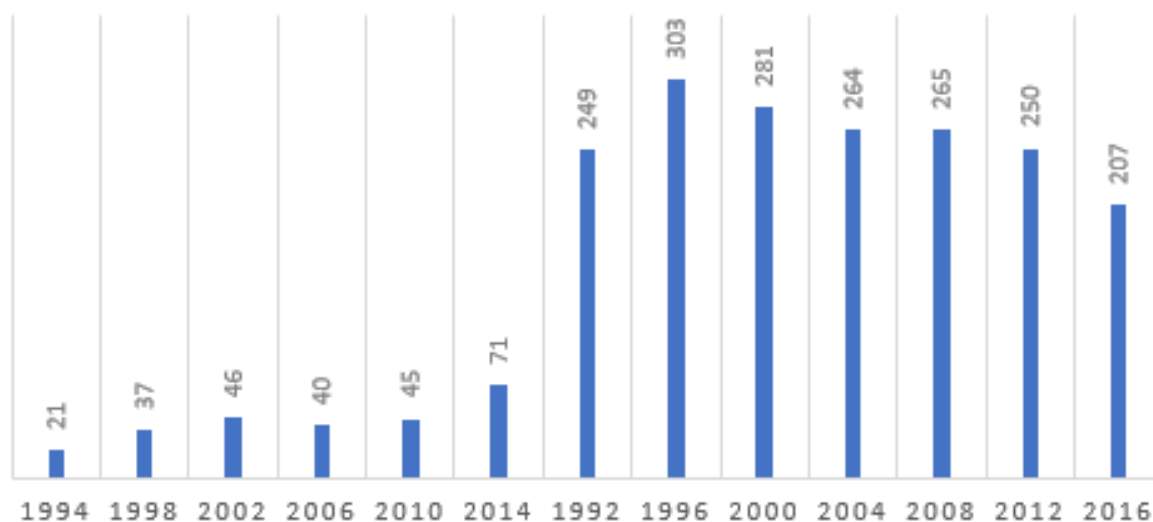
L'écart type est élevé (26,7% de la moyenne) ce qui montre que nombre de participants varie beaucoup c'est pour cela que dans ce cas nous allons calculer les quartiles .

- Moyenne Participation JO Été :
 - (315 + 306 + 266 + 306 + 332 + 291 + 335) / 7 = 307.3 participants en moyenne
 - Calcul de l'écart type :
 - Racine Carré(((315² + 306² + 266² + 306² + 332² + 291² + 335²) / 7) - 307.3²) = 21.8

La variation dans le graphique pour la partie JO été se confirme grâce à l'écart type même si il n'est pas très élevé il reste néanmoins bien présent.

Corée du sud :

NOMBRE DE PARTICIPANTS SUD-CORÉEN PAR AN



- Pour ce graphique nous avons déjà séparé les JO d'été et d'hiver, nous pouvons constater que le nombre de participant au JO d'hiver augmente au fil des des années, contrairement au JO d'été qui depuis 1996 ne font quasiment que descendre.

SAE2.04 - Statistiques

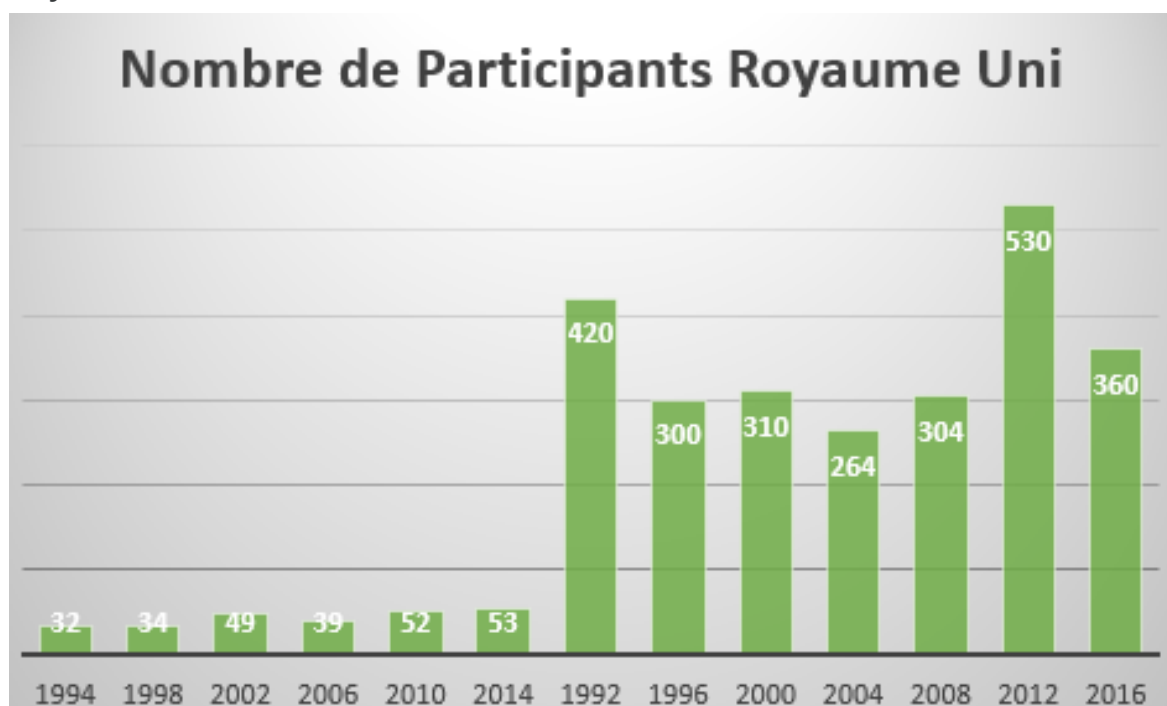
- Moyenne Participation JO Hiver :
 - $(21+37+46+40+45+71) / 6 = 43.3$ participants en moyenne
 - Calcul de l'écart-type :
 - Racine Carré($((21^2 + 37^2 + 46^2 + 40^2 + 45^2 + 71^2) / 6) - 43.3^2$) = 14.9

L'écart type est faible (13% de la moyenne) ce qui montre que le nombre de participants ne varie pas énormément et qu'il reste autour de 43.3 participants

- Moyenne Participation JO Été :
 - $(249 + 303 + 281 + 264 + 265 + 250 + 207) / 7 = 259.9$ participants en moyenne
 - Calcul de l'écart type :
 - Racine Carré($((249^2 + 303^2 + 281^2 + 264^2 + 265^2 + 250^2 + 207^2) / 7) - 259.9^2$) = 27.2

L'écart type est faible (10% de la moyenne) ce qui montre que le nombre de participants ne varie pas énormément et qu'il reste autour de 259 participants

Royaume Uni:



SAE2.04 - Statistiques

- Pour ce graphique nous avons déjà séparé les JO d'été et d'hiver, nous pouvons observer que pour les JO d'hiver il y a une très faible augmentation. Alors que les JO d'été varient beaucoup d'année en année.

- Moyenne Participation JO Hiver :

- $(32+34+49+39+52+53) / 6 = 43.1$ participants en moyenne

- Calcul de l'écart-type :

- Racine Carrée($((32^2 + 34^2 + 49^2 + 39^2 + 52^2 + 53^2) / 6) - 43.1^2$) = 8.84

L'écart type est plutôt faible (20% de la moyenne) ce qui montre que le nombre de participants ne varie pas énormément, il reste autour de 43.1 participants

- Moyenne Participation JO Été :

- $(420 + 300 + 310 + 264 + 304 + 530 + 360) / 7 = 355.4$ participants en moyenne

- Calcul de l'écart type :

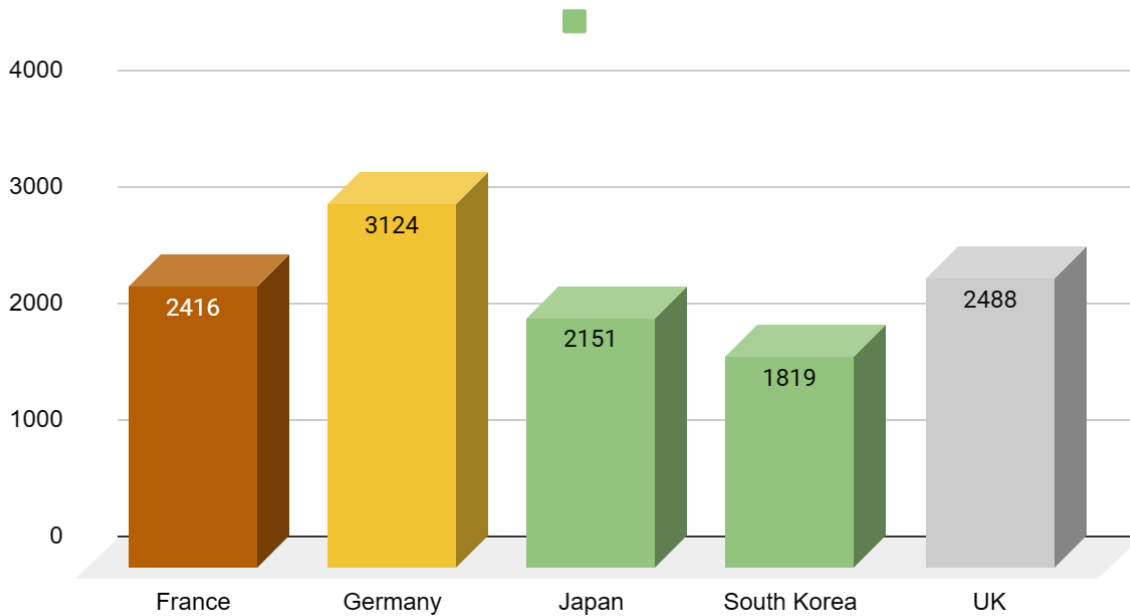
- Racine Carrée($((420^2 + 300^2 + 310^2 + 264^2 + 304^2 + 530^2 + 360^2) / 7) - 355.4^2$) = 54.8

L'écart type est plutôt élevé ce qui se confirme quand on regarde directement le graphique

Pour terminer cette analyse nous souhaitons faire un classement des pays ayant le plus de participants au JO d'hiver et ensuite au JO d'été puis au JO d'Hiver

SAE2.04 - Statistiques

Nombre de Participation pour JO Été et Hiver



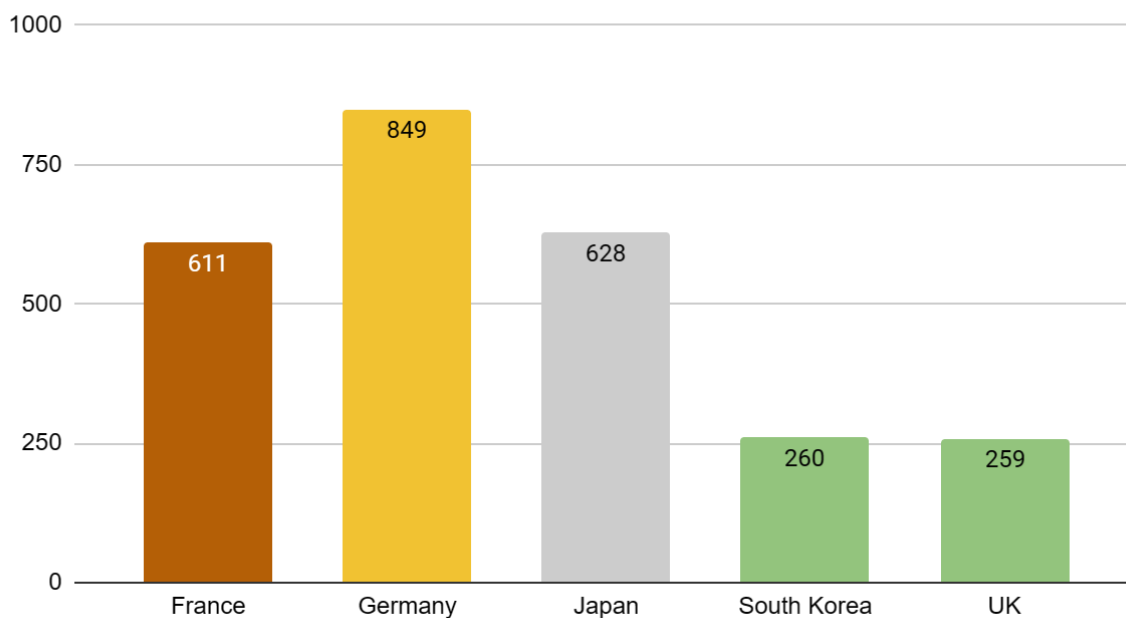
Voici Le Classement que l'on obtient après avoir comptabiliser le nombre de participants pour les JO d'été par Pays :

- 1er : L'Allemagne
- 2eme : Le Royaume Uni
- 3eme : La France
- 4eme : Le Japon
- 5eme : La Corée Du Sud

Mais le classement reste-il le même pour les JO d'hiver ?

SAE2.04 - Statistiques

Nombre de Participation pour JO Hiver



La réponse est non on peut d'ailleurs le voir sur le graphique ci-dessus.
Ce graphique donne le classement suivant:

- 1er : L'Allemagne
- 2eme : Le Japon
- 3eme : La France
- 4eme : La Corée Du Sud
- 5eme : Le Royaume Uni

On observe notamment que le Royaume Uni passe de 3eme à dernier.
Le classement n'est donc clairement pas le même.

SAE2.04 - Statistiques

Question 3bii : Du nombre de médaillés

Requete SQL :

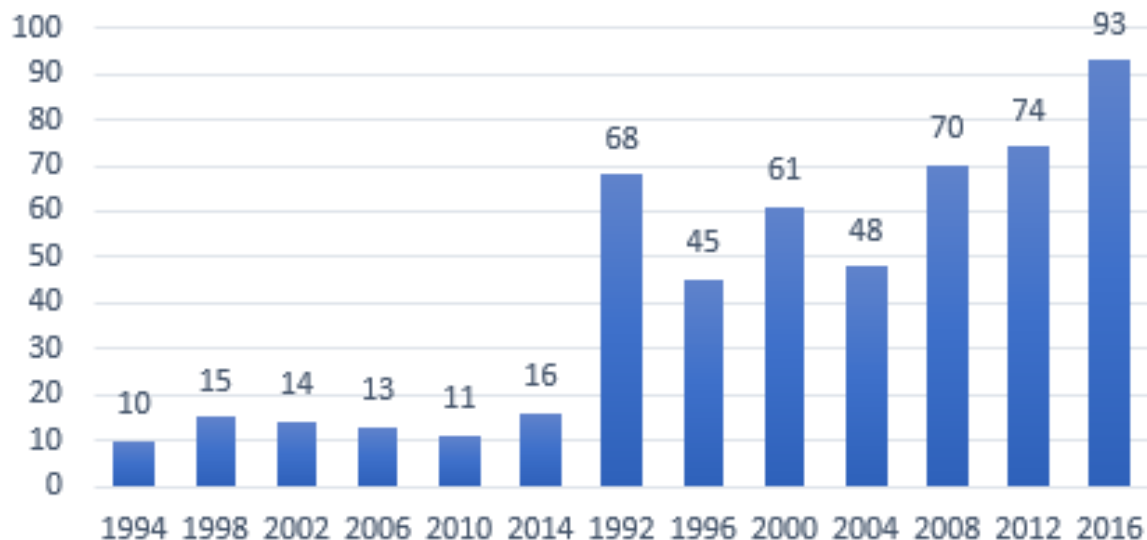
```
\copy (SELECT 'Question 3 b ii' as Question ,pa.region, j.annee,count(Distinct p.id)
as NbParticipants
FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m, athlete as a
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND j.annee>=1992
And j.annee<=2016
AND p.idM=m.idM
AND p.ID=a.ID
AND( pa.region='France'
      OR pa.region='UK'
      OR pa.region='Japan'
      OR pa.region='Germany'
      OR pa.region='South Korea')
AND m.NomMedal<>'NA'
GROUP BY t.NOC,pa.region,j.annee
ORDER BY pa.region , j.annee )
TO 'Question3bii.csv' WITH (format CSV , DELIMITER ',' , HEADER);
```

Graphiques :

France :

SAE2.04 - Statistiques

Nombre de Médaille de La France



- Pour ce graphique nous avons déjà séparé les JO d'été et d'hiver, nous pouvons observer que pour les JO d'hiver il y a une très faible augmentation. Alors que les JO d'été varient beaucoup d'année en année.

- Moyenne Médailles gagnées aux JO Hiver :
 - $(10 + 15 + 14 + 13 + 11 + 16) / 6 = 13,1$ de médaille gagnée en moyenne
 - Calcul de l'écart-type :
 - Racine Carrée($((10^2 + 15^2 + 14^2 + 13^2 + 11^2 + 16^2) / 6) - 13,1^2$) = 2.49

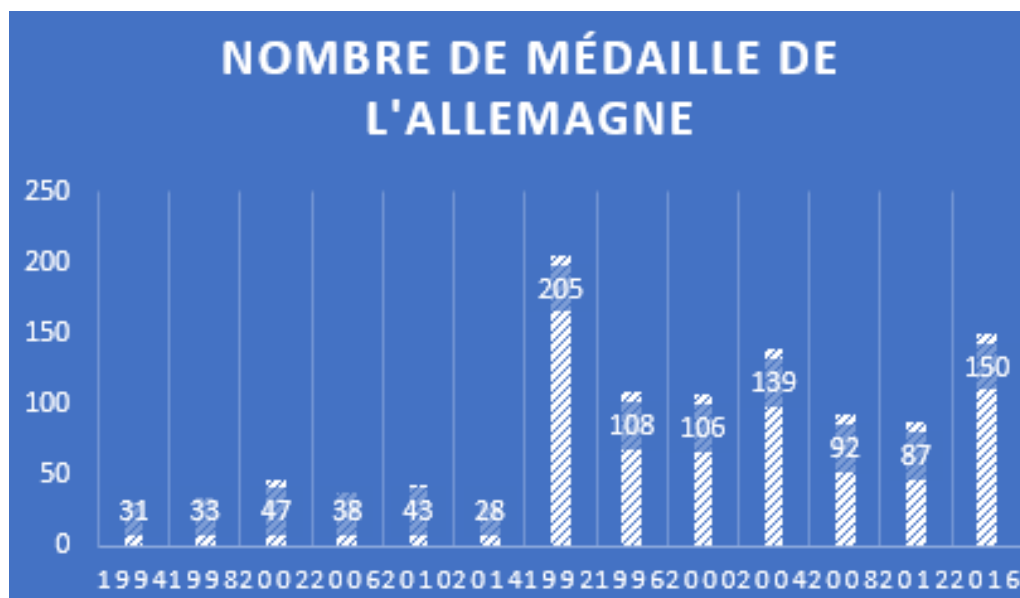
L'écart type est faible ce qui montre que le nombre de médailles ne varie pas énormément, il reste autour de 13 médailles

- Moyenne Participation JO Été :
 - $(68 + 45 + 61 + 48 + 70 + 74 + 93) / 7 = 65.5$
 - Calcul de l'écart type :
 - Racine Carrée($((68^2 + 45^2 + 61^2 + 48^2 + 70^2 + 74^2 + 93^2) / 7) - 65.5^2$) = 13.1

L'écart type des médailles pour les JO d'été est plutôt faible mais on peut quand même voir une petite variation du nombre de médailles gagnées selon les années

SAE2.04 - Statistiques

L'Allemagne :



- Pour ce graphique nous avons déjà séparé les JO d'été et d'hiver, nous pouvons observer que pour les JO d'hiver il y a une très faible augmentation. Alors que les JO d'été varient beaucoup d'année en année.

- Moyenne Médailles gagnées aux JO Hiver :

- $(31 + 33 + 47 + 38 + 43 + 28) / 6 = 36.8$ de médaille gagnée en moyenne

- Calcul de l'écart-type :

- Racine Carrée $\left(\frac{(31^2 + 33^2 + 47^2 + 38^2 + 43^2 + 28^2)}{6} - 36.8^2 \right) = 6.8$

L'écart type comme on peut le voir sur graphique est très faible le nombre de médaille est d'environ 36.8

- Moyenne Participation JO Été :

- $(205 + 108 + 106 + 139 + 92 + 87 + 150) / 7 = 126.7$

- Calcul de l'écart type :

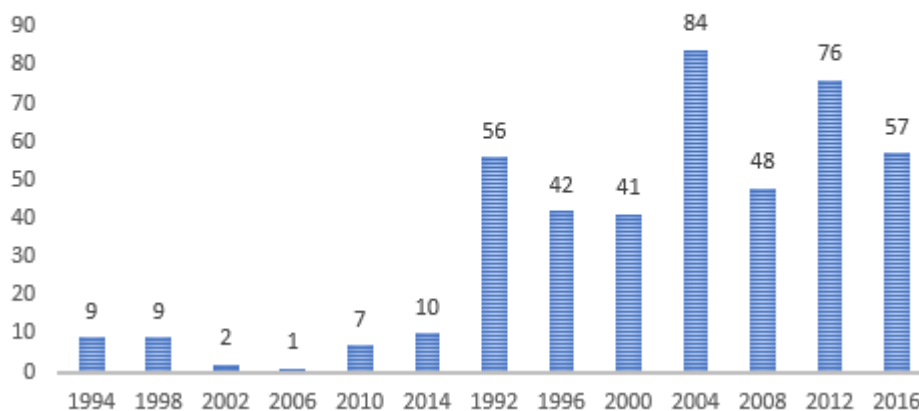
SAE2.04 - Statistiques

- Racine Carré(((205² + 108² + 106² + 139² + 92² + 87² + 150²) / 7) - 126.7²) = 38.5

L'écart type est de 38 ce qui reste non négligeable et qui peut être remarqué simplement regardant le graphique, le nombre de médaille est d' environ 126

Le Japon :

NOMBRE DE MÉDAILLE DU JAPON



- Pour ce graphique nous avons déjà séparé les JO d'été et d'hiver, nous pouvons observer que pour les JO d'hiver il a une très faible augmentation. Alors que les JO d'été varie beaucoup d'année en année.

- Moyenne Médaillés gagner aux JO Hiver :
 - $(9 + 9 + 2 + 1 + 7 + 10) / 6 = 6.33$ de médaille gagner en moyenne

- Calcul de l'écart-type :
 - Racine Carré(((9² + 9² + 2² + 1² + 7² + 10²) / 6) - 6.33²) = 3.5

L'écart type comme on peut le voir sur graphique est très faible le nombre de médaille est d' environ 6.33

- Moyenne Participation JO Été :
 - $(56 + 42 + 41 + 84 + 48 + 76 + 57) / 7 = 57.7$

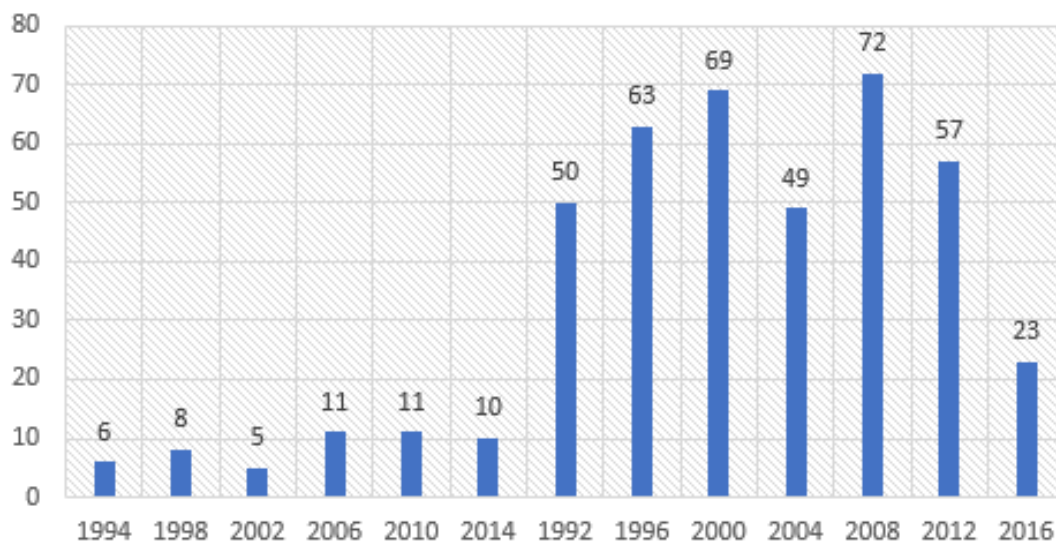
SAE2.04 - Statistiques

- Calcul de l'écart type :
 - Racine Carrée($((56^2 + 42^2 + 41^2 + 84^2 + 48^2 + 76^2 + 57^2) / 7) - 57.7^2$) = 15.4

L'écart type remarquable lorsque l'on regarde le graphique avec des écarts qui creuse entre certaine années , le nombre de médaille est d' environ 57.7

Corée du Sud :

Nombre de Médaille de la Corée du Sud



- Pour ce graphique nous avons déjà séparé les JO d'été et d'hiver, nous pouvons observer que pour les JO d'hiver il a une très faible augmentation. Alors que les JO d'été varie beaucoup d'année en année.
- Moyenne Médailles gagner aux JO Hiver :
 - $(6 + 8 + 5 + 11 + 11 + 10) / 6 = 8.5$ de médaille gagner en moyenne

- Calcul de l'écart-type :

SAE2.04 - Statistiques

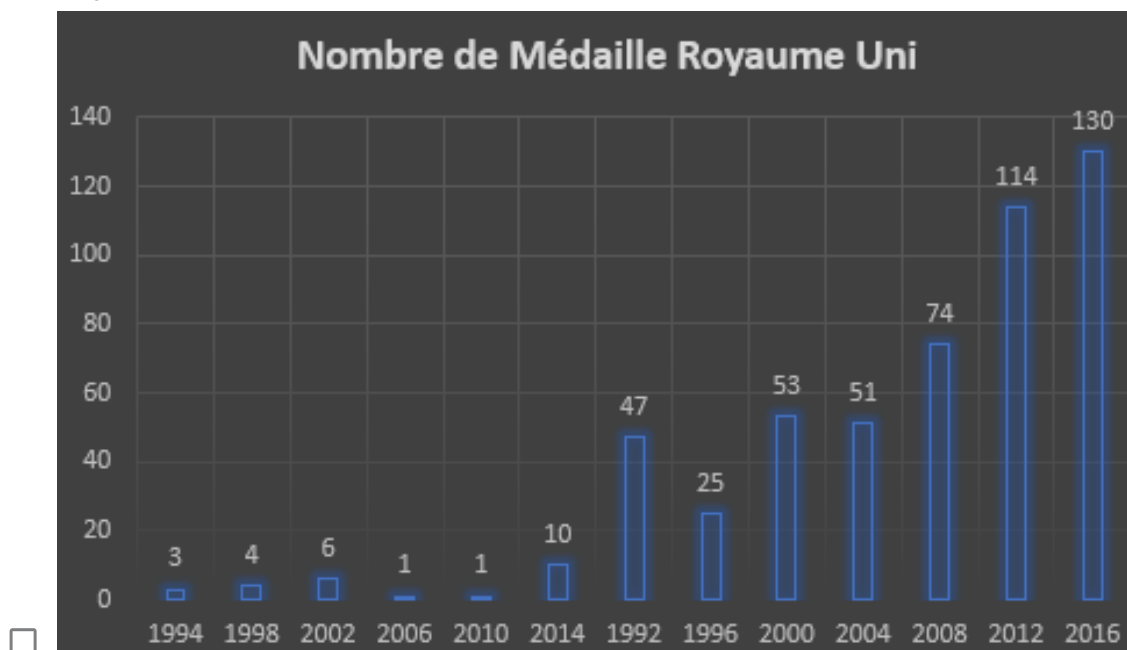
- Racine Carré($((6^2 + 8^2 + 5^2 + 11^2 + 11^2 + 10^2) / 6) - 8.5^2$) = 2.3

L'écart type est extrêmement faible ce qui montre le nombre de médaille gagner est environ le même soit de 8.5

- Moyenne Participation JO Été :
 - $(50 + 63 + 69 + 49 + 72 + 57 + 23) / 7 = 54.7$
- Calcul de l'écart type :
 - Racine Carré($((50^2 + 63^2 + 69^2 + 49^2 + 72^2 + 57^2 + 23^2) / 7) - 54.7^2$) = 15.3

On peut remarquer que l'écart est quasiment égal à celui du japon pour les épreuves d'été. On peut aussi remarquer que le nombre de médaille gagner varie beaucoup ce qui impacte l'écart type, mais aussi que les les épreuves d'été depuis 2008 le nombre de médaille ne fait que baisser

Royaume Uni:



- Pour ce graphique nous avons déjà séparé les JO d'été et d'hiver, nous pouvons observer que pour les JO d'hiver il a une très faible augmentation. Alors que les JO d'été varie beaucoup d'année en année.

SAE2.04 - Statistiques

- Moyenne Médailleurs gagner aux JO Hiver :
 - $(3 + 4 + 6 + 1 + 1 + 10) / 6 = 4.1$ de médaille gagner en moyenne

- Calcul de l'écart-type :
 - Racine Carré($((3^2 + 4^2 + 6^2 + 1^2 + 1^2 + 10^2) / 6) - 4.1^2$) = 10.3

L'écart type est altéré à cause de l'année 2010 qui la plus élevé qui va aussi faire remonter la moyenne à 4.1

- Moyenne Participation JO Été :
 - $(47 + 25 + 53 + 51 + 74 + 114 + 130) / 7 = 70.5$
- Calcul de l'écart type :
 - Racine Carré($((47^2 + 25^2 + 53^2 + 51^2 + 74^2 + 114^2 + 130^2) / 7) - 70.5^2$) = 35.5

On peut remarquer que le nombre de médailles augmente drastiquement au fil des années pour passer de 47 en 1992 à 130 en 2016 ce démontré que l'écart type est aussi élevé.

Question 3biii : Du nombre de femmes participantes

Requete :

```
\copy (SELECT 'Question 3 b iii' as Question ,pa.region, j.annee,count(Distinct
p.id) as NbDeFemme
FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m, athlete as a
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND j.annee>=1992
And j.annee<=2016
AND p.idM=m.idM
AND p.ID=a.ID
AND( pa.region='France'
OR pa.region='UK'
OR pa.region='Japan'
OR pa.region='Germany'
OR pa.region='South Korea')
and a.SEX='F'
```

SAE2.04 - Statistiques

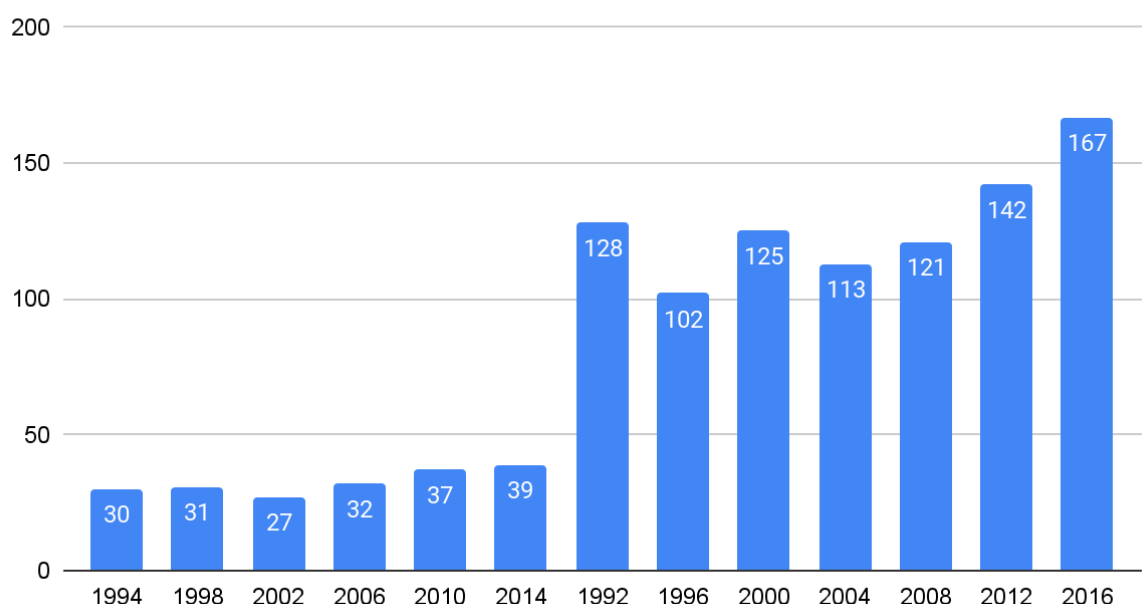
```
GROUP BY t.NOC,pa.region,j.annee  
ORDER BY pa.region , j.annee)  
TO 'Question3biii.csv' WITH (format CSV , DELIMITER ',' , HEADER);
```

Graphique :

Nous avons séparé le graphique en 5 sous graphiques par pays

France :

Nombre de Participantes Françaises



Pour ce graphique comme pour les prochains, nous voulons voir s'il y a une augmentation du nombre de femmes lorsque les années passent parmi les athlètes des JO, pour cela nous allons analyser la covariance et le coefficient de corrélation linéaire afin de voir si cette dernière est positive et forte. Cependant pour ne pas biaiser les calculs nous allons une nouvelle fois séparer tous nos sous-graphiques en 2 ; Une partie JO été et une autre hiver.

- Moyenne Participation JO Hiver :
 - $(30+31+27+32+37+39) / 6 = 32,7$ participants en moyenne
 - Calcul de l'écart-type :

SAE2.04 - Statistiques

- Racine Carré($((30^2 + 31^2 + 27^2 + 32^2 + 37^2 + 39^2) / 6) - 32,7^2$) = 3,8

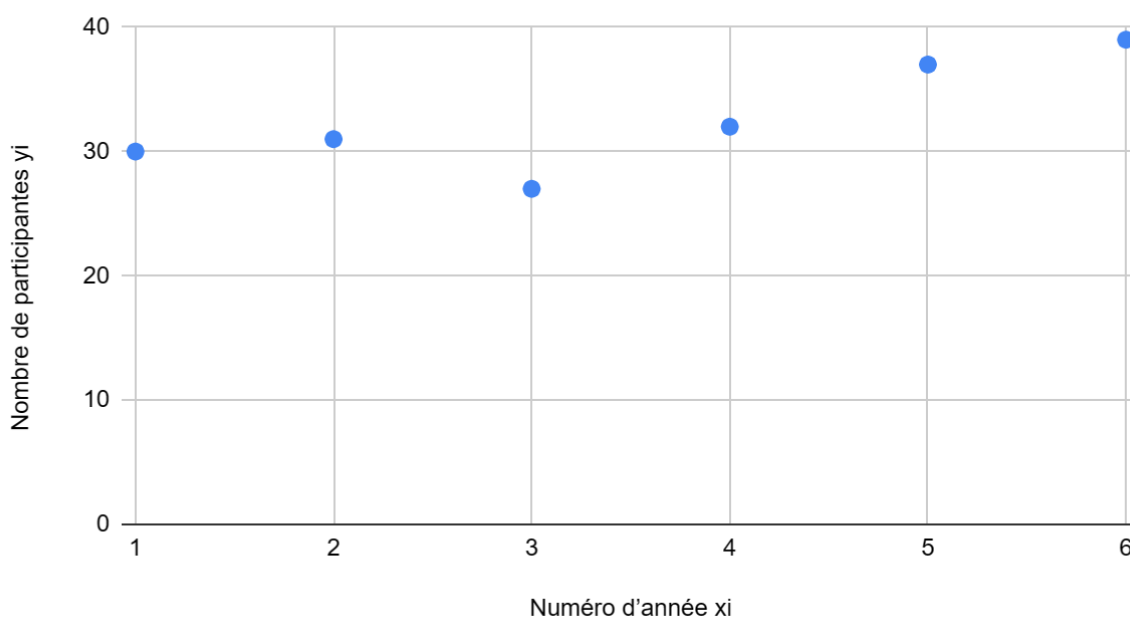
L'écart type est faible (11,6% de la moyenne) ce qui montre que nombre de participantes ne varie pas énormément et qu'il reste autour de 32.7 participantes

Nous obtenons ensuite ce tableau de nuage de points :

Numéro d'année xi	1	2	3	4	5	6
Nombre de participantes yi	30	31	27	32	37	39

Nous obtenons le nuage de points suivant :

Nombre de participantes yi par rapport à Numéro d'année xi



- Moyenne Année JO Hiver :
 - $(1+2+3+4+5+6)/6 = 3,5$
 - Calcul de l'écart-type :
 - Racine Carré($((1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) / 6) - 3,5^2$) = 2,91
- Calcul de Covariance :
 - $((30*1+31*2+27*3+32*4+37*5+39*6)/6)-(3,5*32,7)=5,55$

SAE2.04 - Statistiques

- Calcul de Corrélation linéaire :
 - $5,55/(2,91*3,8)=0,5$

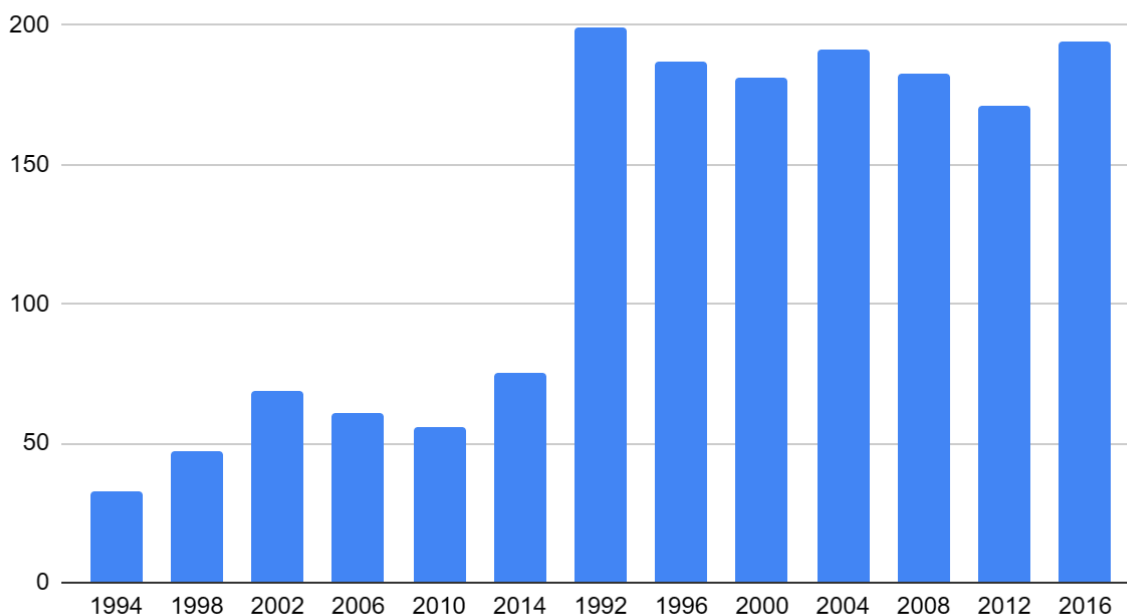
On constate que le coefficient de corrélation est inférieur à 0,7 , il y a donc une faible corrélation entre les deux variables cela démontre qu'il n'y a pas forcément de lien entre les années qui passent et l'augmentation du nombre de participantes .

C'est grâce à ce calcul que nous nous rendons compte qu'il manque une partie des données pour vérifier qu'il y a eu une augmentation du nombre de femmes lors des JO. Ce qui nous manque ce sont des données sur la proportion de femmes dans les équipes car en effet le nombre d'athlètes change chaque année tantôt il augmente, tantôt il diminue ce qui biaise ce calcul.

Pour les prochains pays, nous ne montrerons que les graphiques étant donné que les calculs ne sont pas représentatifs.

Allemagne :

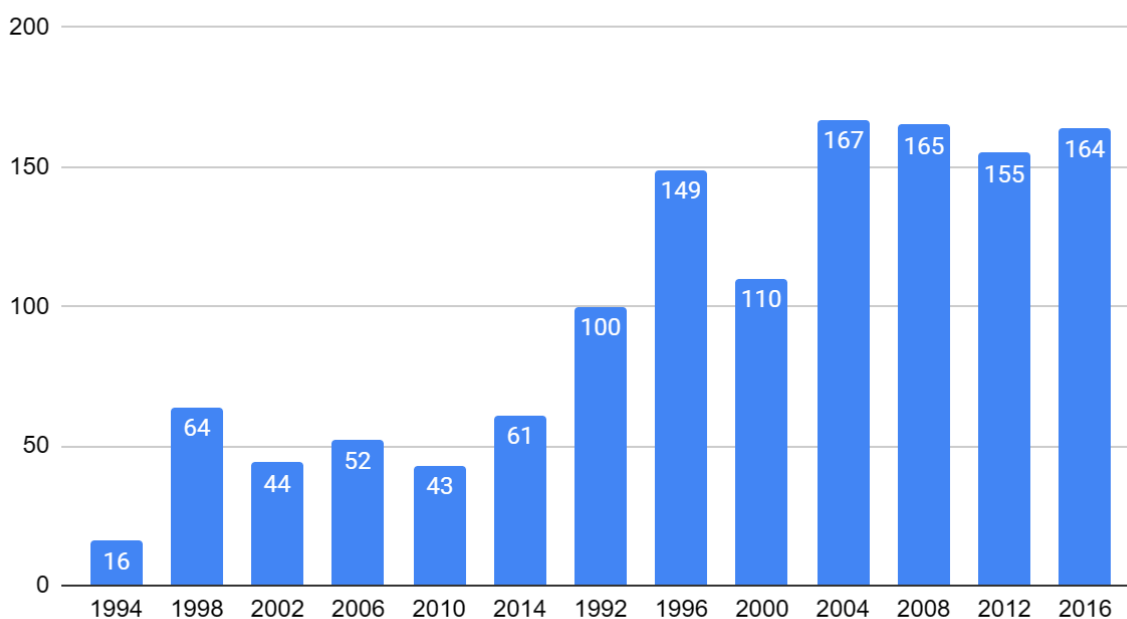
Nombre de Participantes Allemandes



Japon :

SAE2.04 - Statistiques

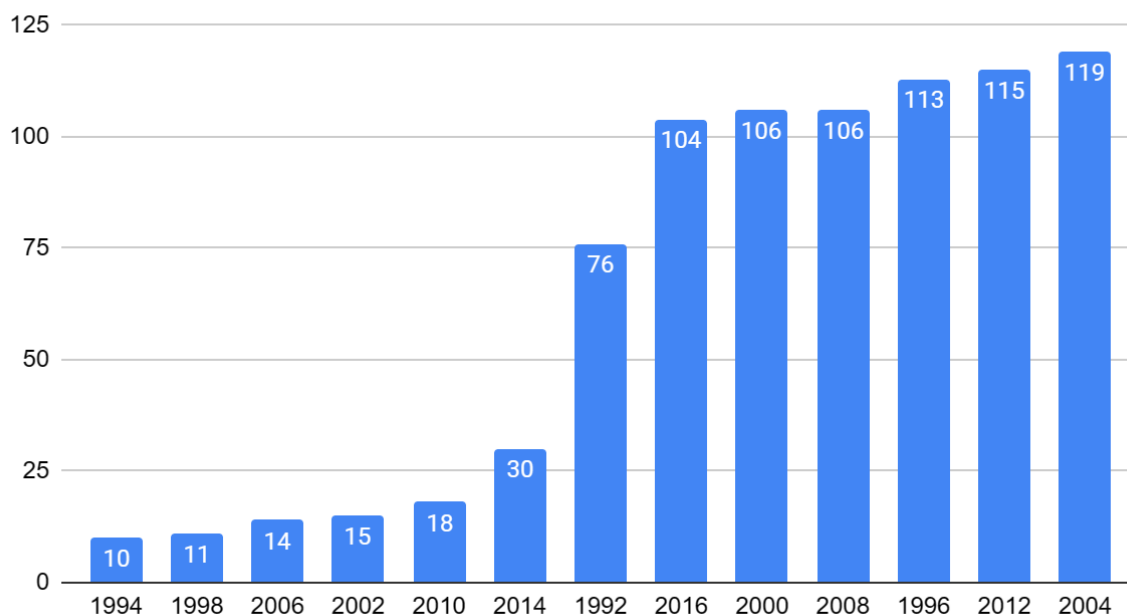
Nombres de participants Japonaise



Corée du sud :

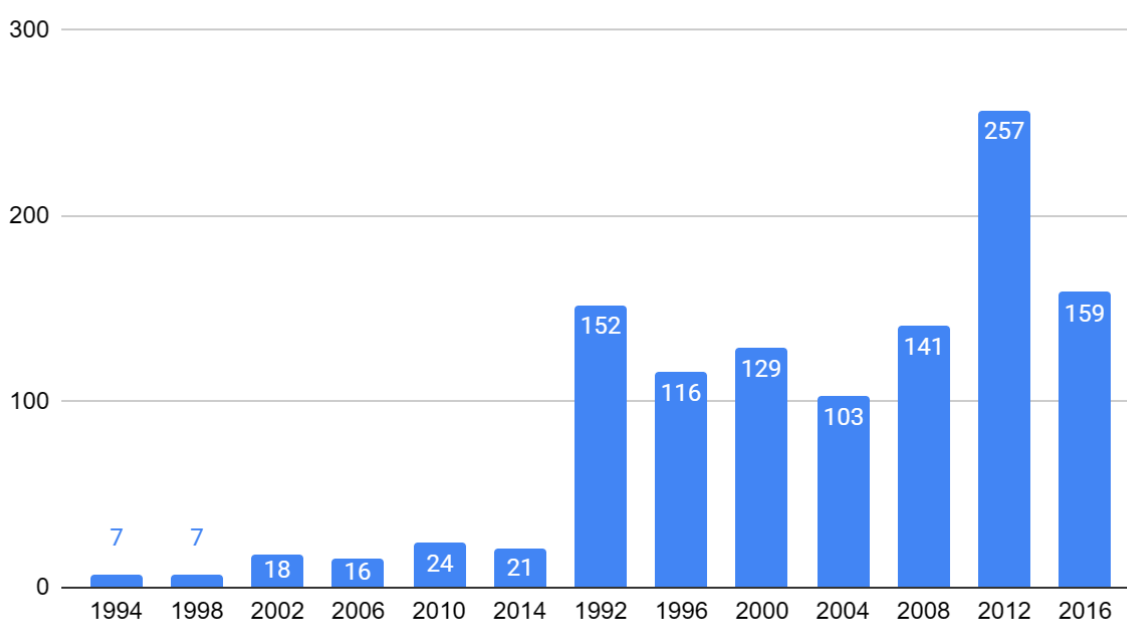
SAE2.04 - Statistiques

Nombre de participantes Sud Corréenne



Royaume Uni :

Nombre de Participantes Du Royaume Uni



SAE2.04 - Statistiques

Question 3biv : De la proportion de femmes participantes (en comparaison de la proportion d'hommes)

Requetes :

Pour cette requete nous avons été confronté a un probleme, les proportions en effet lorsque nous avons décidé de compter le nombre de femmes, nous avons aussi perdu le nombre total de participants par pays, c'est pour cela que nous avons donc dû créer une nouvelle table contenant les données de proportion :

Drop Table Effect;

– Ici nous allons créer la table permettant d'habriter les proportions par années et par pays

```
CREATE Table Effect(  
    NOC char(3),  
    Year int,  
    effectif int,  
    Constraint pk_eff PRIMARY KEY(NOC,Year)  
);
```

– Ici nous allons insérer toutes les données dans la table

```
Insert Into Effect  
    SELECT Distinct t.NOC,j.annee,count(DISTINCT p.id)  
    FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m, athlete as a  
    WHERE p.idjo=j.idjo  
    AND p.idteam=t.idteam  
    AND t.NOC=pa.NOC  
    AND j.annee>=1992  
    And j.annee<=2016  
    AND p.idM=m.idM  
    AND p.ID=a.ID  
    AND( pa.region='France'  
    OR pa.region='UK'  
    OR pa.region='Japan')
```


SAE2.04 - Statistiques

```
OR pa.region='Germany'  
OR pa.region='South Korea')  
Group By t.NOC,j.annee;
```

– Ici il s'agit de la commande SQL permettant de répondre à la question 3biv

```
\copy (SELECT 'Question 3 b iv' as Question ,pa.region, a.SEX,j.annee,CAST  
(CAST(count(Distinct p.id)AS Numeric(6,3))/(e.effecfif) AS Numeric(6,3))As proportion  
FROM participe as p , jo as j,TEAM as t ,Effect as e, Pays as pa,Medal as m, athlete as a  
WHERE p.idjo=j.idjo  
AND p.idteam=t.idteam  
AND t.NOC=pa.NOC  
AND e.NOC=t.NOC  
And e.year=j.annee  
AND j.annee>=1992  
And j.annee<=2016  
AND p.idM=m.idM  
AND p.ID=a.ID  
AND( pa.region='France'  
      OR pa.region='UK'  
      OR pa.region='Japan'  
      OR pa.region='Germany'  
      OR pa.region='South Korea')  
GROUP BY t.NOC,pa.region,j.annee,a.SEX,e.effecfif  
ORDER BY pa.region , j.annee)  
TO 'Question3biv.csv' WITH (format CSV , DELIMITER ',' , HEADER);
```

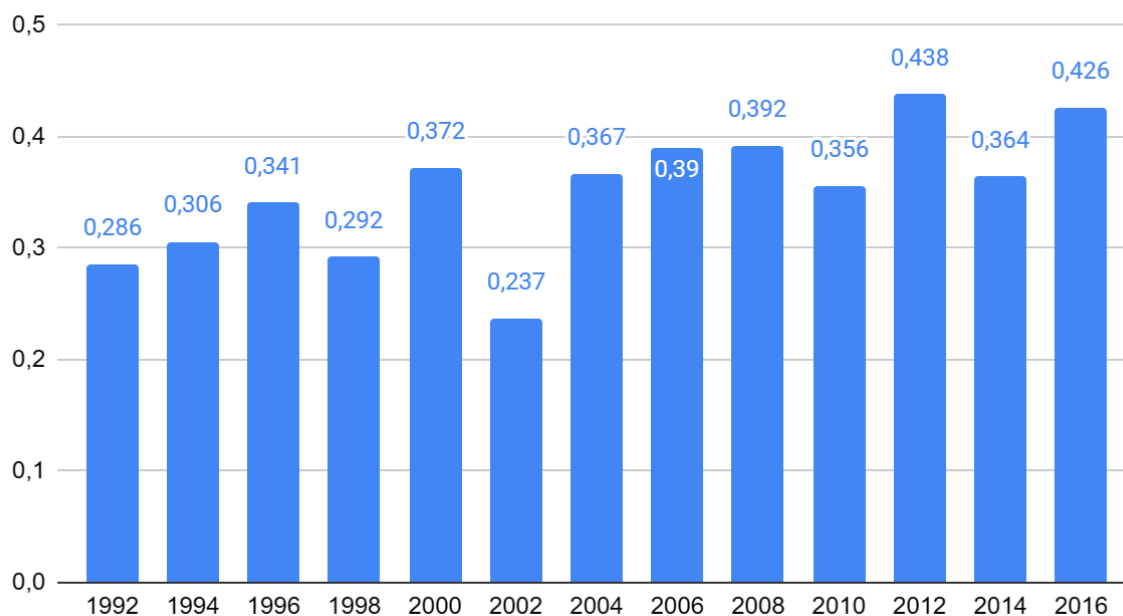
Graphique :

Ici nous avons découpé notre graphiques en 5 sous graphiques organisés par pays.

France :

SAE2.04 - Statistiques

Proportion de Femmes Athlètes Dans l'équipe de France



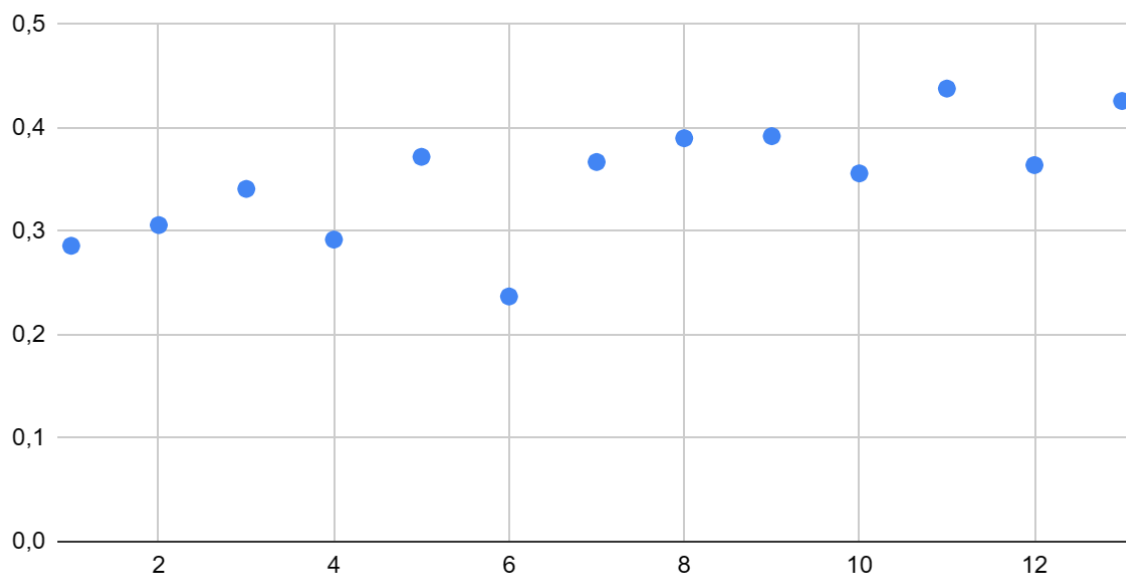
Pour cette variable il semble plus intéressant de calculer le coefficient de covariance;
Nous obtenons ce tableau de nuage de points

Année xi	Proportion Participants yi
0	0,286
1	0,306
2	0,341
3	0,292
4	0,372
5	0,237
6	0,367
7	0,39
8	0,392
9	0,356
10	0,438
11	0,364
12	0,426

SAE2.04 - Statistiques

Nous Obtenons ensuite ce nuage de points :

Evolution de la proportion des Femmes dans l'équipe de France Au fil des Années



Calcul de l'écart type :

écart type des proportions = 0,05554815598 ;

écart type des années (Nous pourrions réutiliser cette écart types pour les prochains pays) = 3,894440482;

Calcul de Covariance :

Covariance = 0,1463846154 ;

Calcul du Coefficient de corrélation :

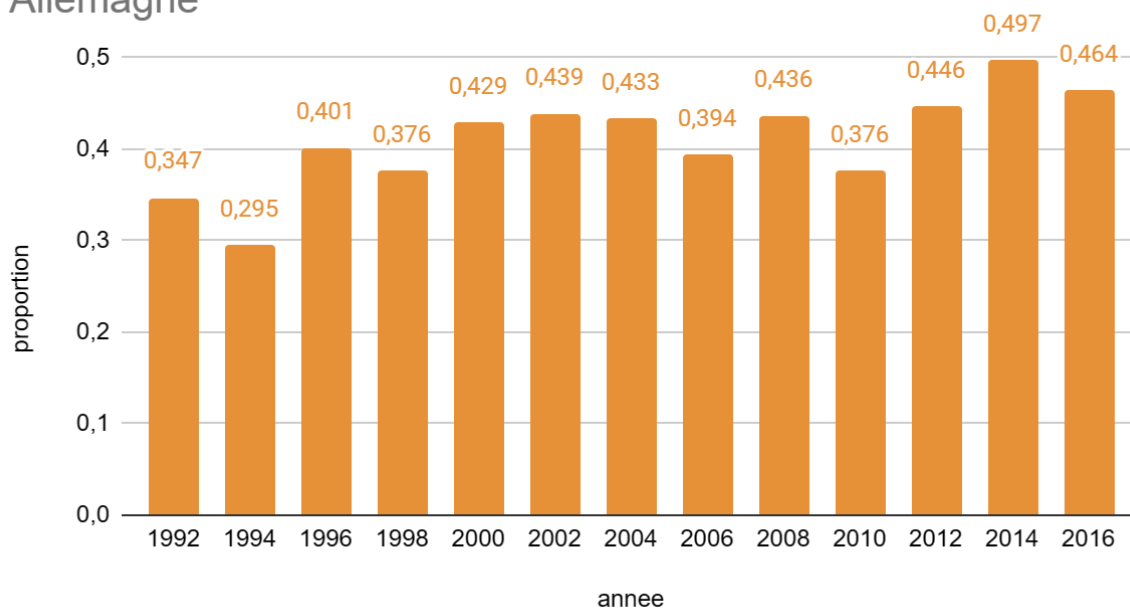
Coefficient de corrélation : 0,7043066221

Le coefficient de corrélation étant supérieur à 0,7 alors on dit qu'il y a une forte corrélation cela pourrait vouloir dire que plus les JO passent plus les femmes sont présentes lors des JO en France.

SAE2.04 - Statistiques

Allemagne :

proportion du nombre de femme par rapport à l'année en Allemagne



Pour cette variable il semble plus intéressant de calculer le coefficient de covariance;
Nous obtenons ce tableau de nuage de points

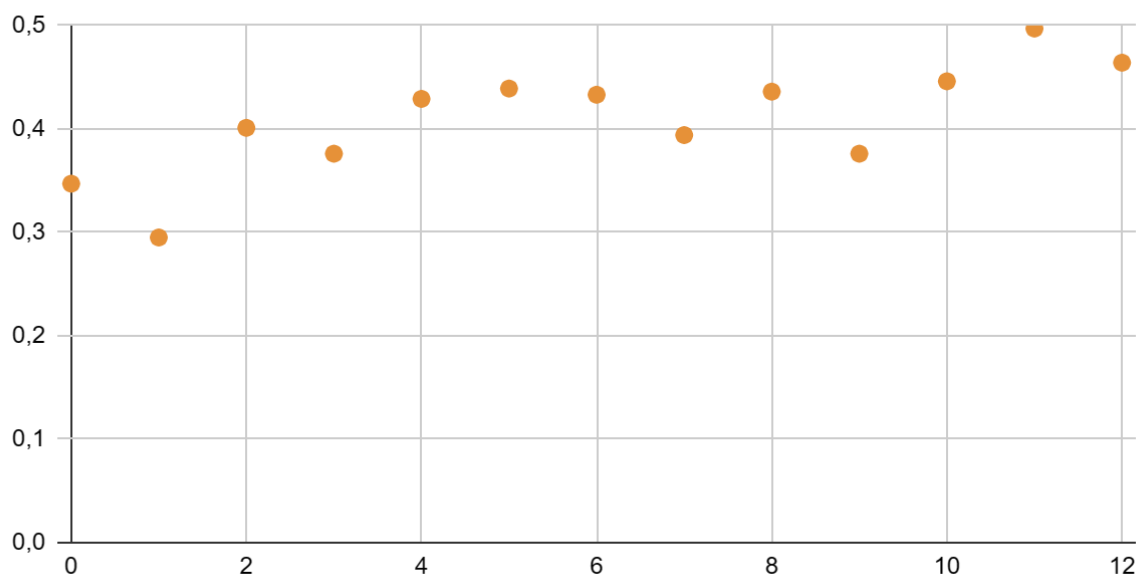
Année xi	Proportion Participants yi
0	0,347
1	0,295
2	0,401
3	0,376
4	0,429
5	0,439
6	0,433
7	0,394
8	0,436

SAE2.04 - Statistiques

9	0,376
10	0,446
11	0,497
12	0,464

Nous Obtenons ensuite ce nuage de points :

Evolution de la proportion des Femmes dans l'équipe
Allemande au fil des Années



Calcul de l'écart type :

écart type des proportions = 0,05325278372 ;
écart type des années = 3,894440482;

Calcul de Covariance :

Covariance = 0,1431538462 ;

Calcul du Coefficient de corrélation :

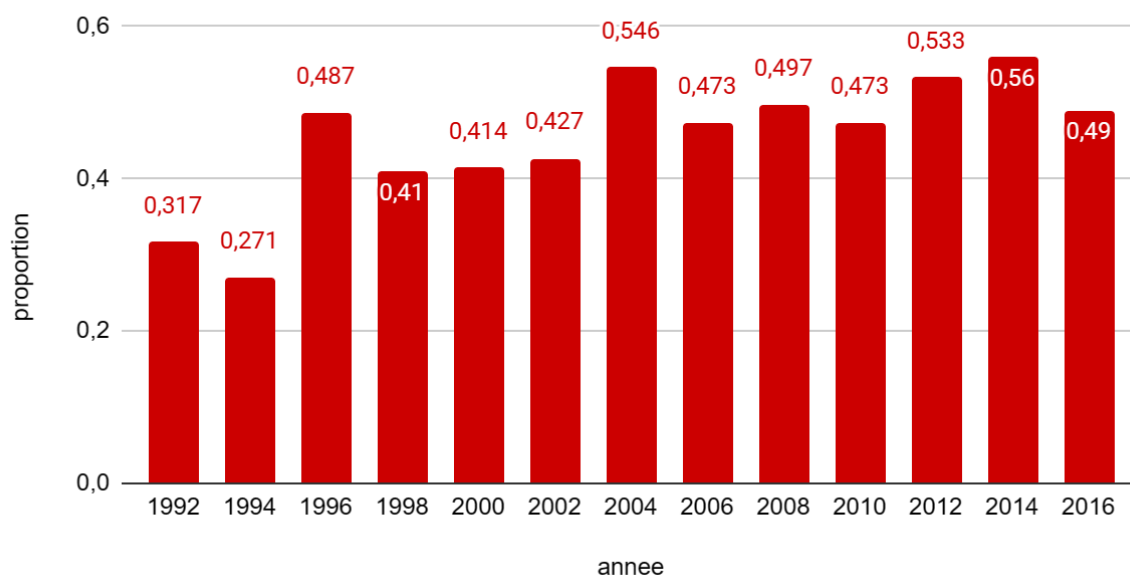
Coefficient de corrélation : 0,7477867079

SAE2.04 - Statistiques

Le coefficient de corrélation étant supérieur à 0,7 alors on dit qu'il y a une forte corrélation cela pourrait vouloir dire que plus les passent plus les femmes sont présentes lors des JO en Allemagne.

Japon:

proportion de Femme dans l'équipe Japonaise par rapport à l'année



Pour cette variable il semble plus intéressant de calculer le coefficient de covariance;
Nous obtenons ce tableau de nuage de points

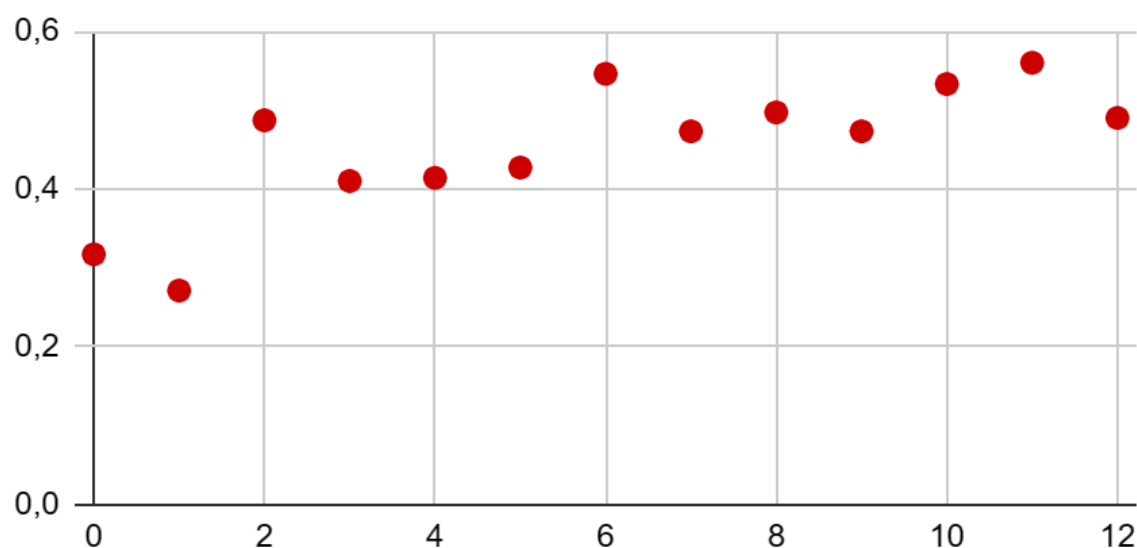
Année xi	Proportion Participants yi
0	0,317
1	0,271
2	0,487

SAE2.04 - Statistiques

3	0,41
4	0,414
5	0,427
6	0,546
7	0,473
8	0,497
9	0,473
10	0,533
11	0,56
12	0,49

Nous Obtenons ensuite ce nuage de points :

Evolution de la proportion des Femmes dans l'équipe Japonaise au fil des Années



Calcul de l'écart type :

écart type des proportions = 0,08535161062 ;

écart type des années = 3,894440482;

SAE2.04 - Statistiques

Calcul de Covariance :

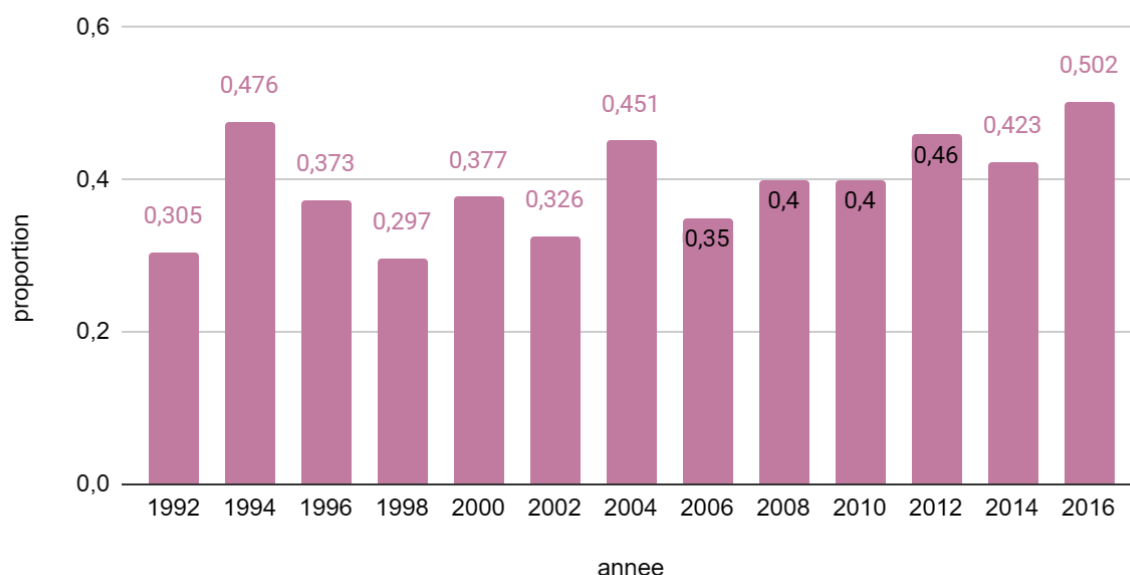
Covariance = 0,236 ;

Calcul du Coefficient de corrélation : 0,7691611098;

Le coefficient de corrélation étant supérieur à 0,7 alors on dit qu'il y a une forte corrélation cela pourrait vouloir dire que plus les passent plus les femmes sont présentes lors des JO en Japonaise.

Corée du sud :

proportion de Femme dans l'équipe Sud Coréenne par rapport à l'année



Pour cette variable il semble plus intéressant de calculer le coefficient de covariance;

SAE2.04 - Statistiques

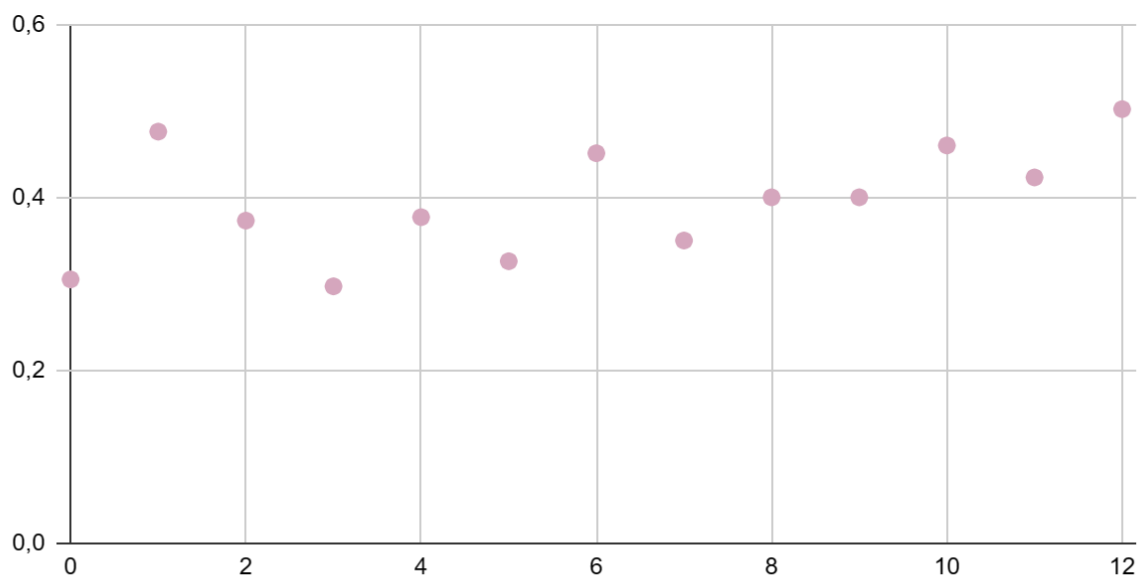
Nous obtenons ce tableau de nuage de points

Année xi	Proportion Participants yi
0	0,305
1	0,476
2	0,373
3	0,297
4	0,377
5	0,326
6	0,451
7	0,35
8	0,4
9	0,4
10	0,46
11	0,423
12	0,502

Nous Obtenons ensuite ce nuage de points :

SAE2.04 - Statistiques

Evolution de la proportion des Femmes dans l'équipe Sud Coréenne au fil des Années



Calcul de l'écart type :

écart type des proportions = 0,06551149831 ;

écart type des années = 3,894440482;

Calcul de Covariance :

Covariance = 0,1264615385;

Calcul du Coefficient de corrélation : 0,5369798187;

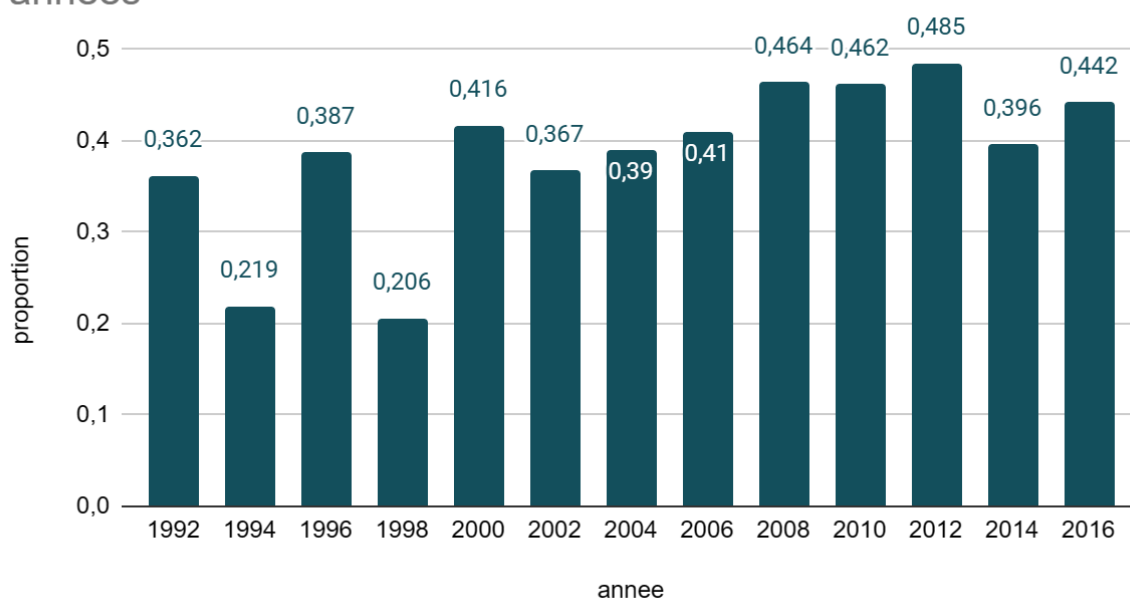
Le coefficient de corrélation étant inférieur à 0,7 alors on dit qu'il y a une faible corrélation cela pourrait vouloir dire que la Corée du Sud n'a pas fait réellement augmenter la proportion de femmes dans son équipe.

Cependant on observe que la proportion de femme atteint déjà voire dépasse les 50%, cela explique donc que la proportion de femme stagne d'où ce coefficient de corrélation faible, mais cela ne veut pas dire que la Corée du Sud n'inclut pas les femmes, pour vérifier cela nous devrions comparer les moyennes de chaque pays ce que nous ferons par la suite.

SAE2.04 - Statistiques

Royaume Uni:

proportion Femmes dans l'équipe du Royaume Uni au fil des années



Pour cette variable il semble plus intéressant de calculer le coefficient de covariance;
Nous obtenons ce tableau de nuage de points

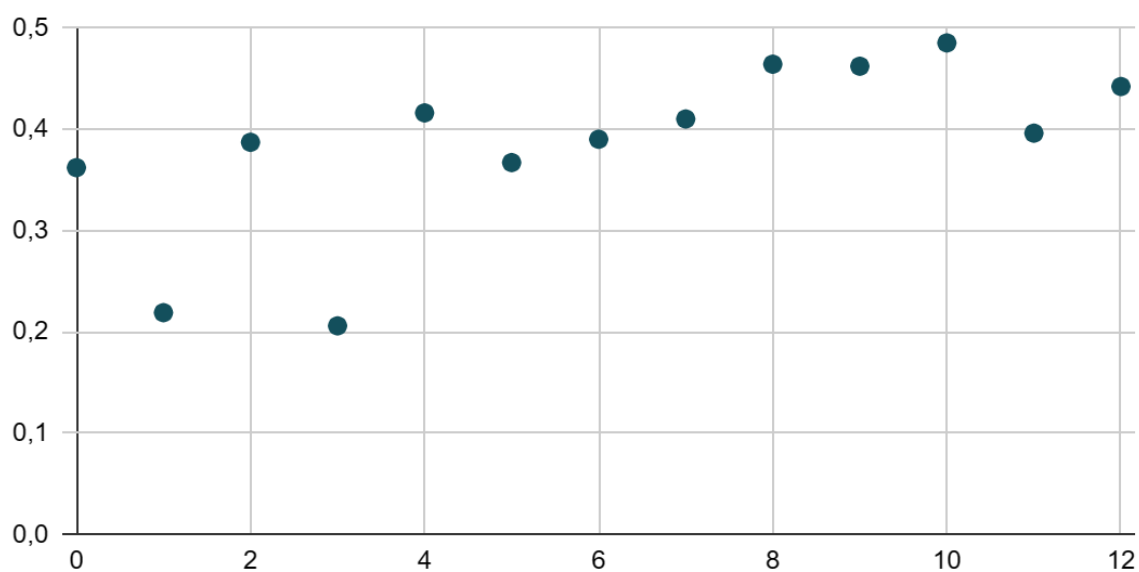
Année xi	Proportion Participants yi
0	0,362
1	0,219
2	0,387
3	0,206
4	0,416
5	0,367
6	0,39

SAE2.04 - Statistiques

7	0,41
8	0,464
9	0,462
10	0,485
11	0,396
12	0,442

Nous Obtenons ensuite ce nuage de points :

Evolution de la proportion des Femmes dans l'équipe du Royaume Uni au fil des Années



Calcul de l'écart type :

écart type des proportions = 0,08544243826 ;

écart type des années = 3,894440482;

Calcul de Covariance :

Covariance = 0,2049230769;

SAE2.04 - Statistiques

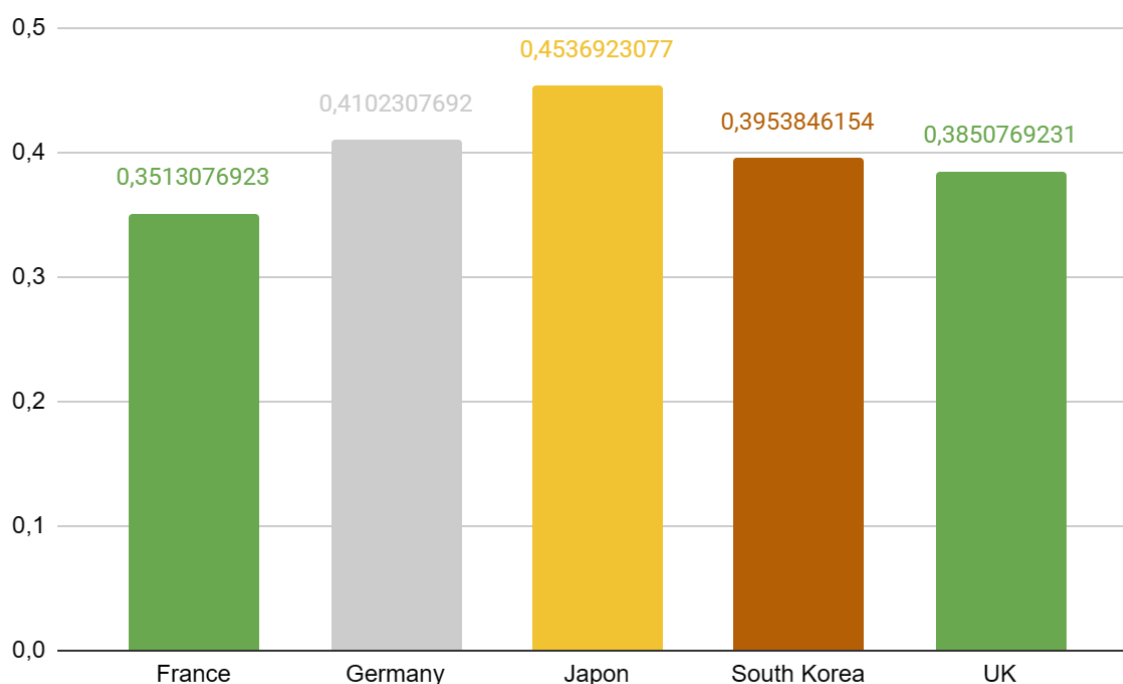
Calcul du Coefficient de corrélation : 0,6671665599;

Le coefficient de corrélation étant inférieur à 0,7 alors on dit qu'il y a une faible corrélation cela pourrait vouloir dire que le Royaume Uni n'a pas fait réellement augmenter la proportion de femmes dans son équipe.

Cependant on observe que la proportion de femme a tendance à se rapprocher des 50%, cela explique donc que la proportion de femme stagne d'où ce coefficient de corrélation faible, mais comme pour la Corée du sud, cela ne veut pas dire que le Royaume Uni n'inclut pas les femmes, pour vérifier cela nous allons comparer les moyennes de chaque pays.

CONCLUSION :

Après avoir calculé les moyennes des proportions nous obtenons le classement suivant :



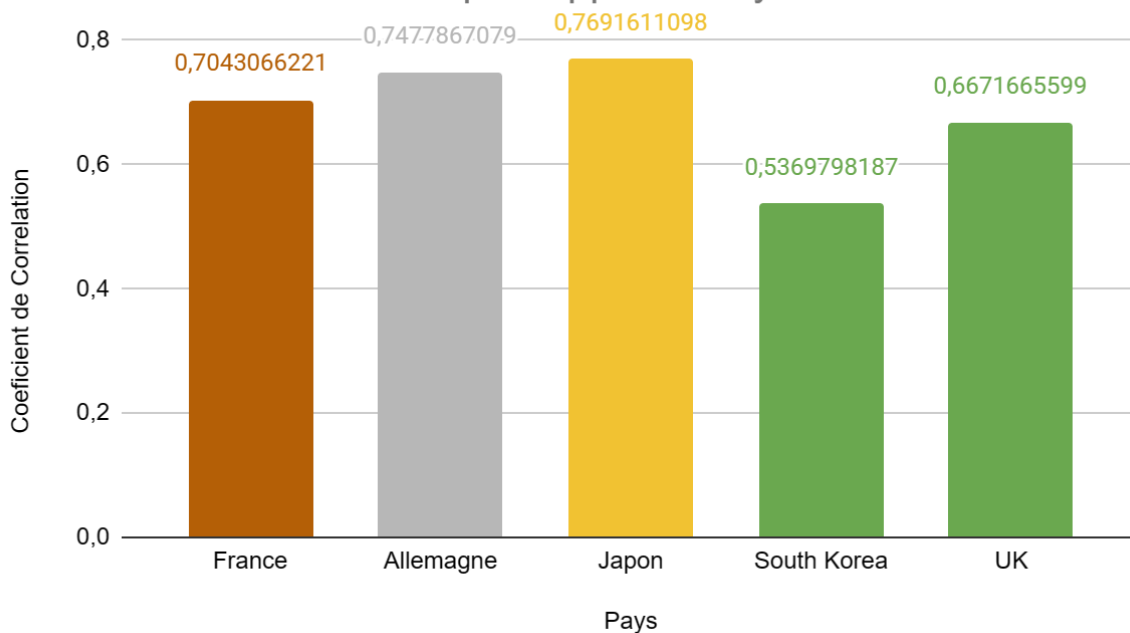
(Tableau de la moyenne des proportions de femmes parmi les athlètes par pays)

Le Japon est le pays qui inclut le plus de femmes aux JO suivi par l'Allemagne et la Corée du Sud, la France se trouvant en dernière position.

SAE2.04 - Statistiques

Mais maintenant intéressons nous au classement des pays ayant montré le plus d'évolution pour cela nous allons utiliser les coefficient de corrélation de chacun :

Coefficient de Correlation par rapport à Pays



Le Japon est le pays qui a le plus évolué de 1992 à 2016 en incluant de plus en plus les femmes aux JO suivi par l'Allemagne et la France, la Corée du Sud se trouvant en dernière position .

On observe donc que le Japon est le pays qui a le plus évolué et qui inclut le plus de femmes.

Question 3bv. De la proportion de médaillés parmi les femmes :

Requete : Pour cette requete nous avons eu le même problème que pour le précédent (proportion), nous l'avons donc résolu de la même manière

`DELETE FROM effect;`

SAE2.04 - Statistiques

Insert Into Effect

```
SELECT Distinct t.NOC,j.annee,count(DISTINCT p.id)
FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m, athlete as a
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND j.annee>=1992
And j.annee<=2016
AND p.idM=m.idM
AND p.ID=a.ID
AND( pa.region='France'
OR pa.region='UK'
OR pa.region='Japan'
OR pa.region='Germany'
OR pa.region='South Korea')
AND a.SEX='F'
Group By t.NOC,j.annee;
```

```
\copy (SELECT 'Question 3 b v' as Question ,pa.region, a.SEX,m.nomMedal<>'NA' as
Medaille, j.annee,Cast(count(DISTINCT p.id )AS numeric(6,3))/e.effecfif as
NbDeFemme
FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m, athlete as a , effect as
e
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND e.NOC=t.NOC
AND e.year=j.annee
AND j.annee>=1992
And j.annee<=2016
AND p.idM=m.idM
AND p.ID=a.ID
AND( pa.region='France'
      OR pa.region='UK'
      OR pa.region='Japan'
      OR pa.region='Germany'
      OR pa.region='South Korea')
AND a.SEX='F'
AND m.nomMedal<>'NA'
GROUP BY t.NOC,pa.region,j.annee,a.SEX,m.nomMedal<>'NA',e.effecfif
```

SAE2.04 - Statistiques

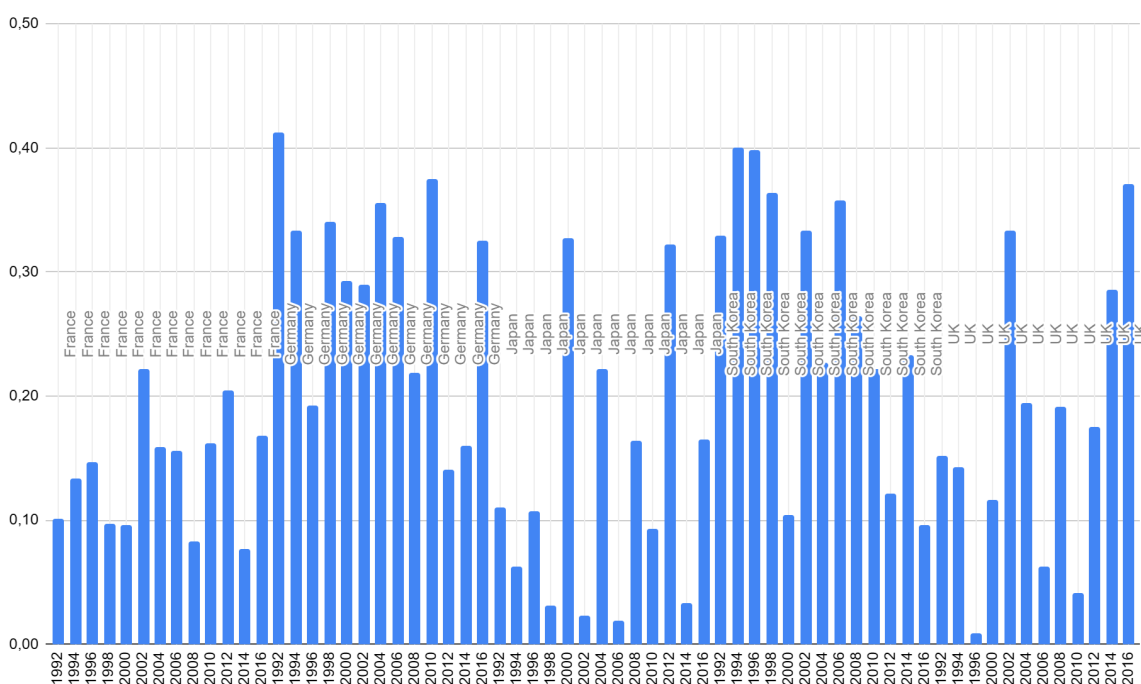
ORDER BY pa.region , j.annee) TO 'Question3bv.csv' WITH (format CSV , DELIMITER
' , , HEADER);

Nous avons décidé de traiter cette question de deux facons différentes , nous
souhaitions savoir quel pays avait en moyenne le plus haut taux de médaillés parmi les
femmes.

Cependant nous souhaitions auusi savoir quelle année avait vu le plus de médaillés
parmi les femmes.

Pour ce faire commencons par voir quel pays avait en moyenne le plus haut taux de
médaillés parmi les femmes :

Nous avons pour base ce graphique :



Ce graphique est beaucoup trop dense pour être utilisé, c'est pour cela que nous allons
utiliser un graphique différent se basant sur les moyenne des proportion des médaillés
parmi les femmes par pays .

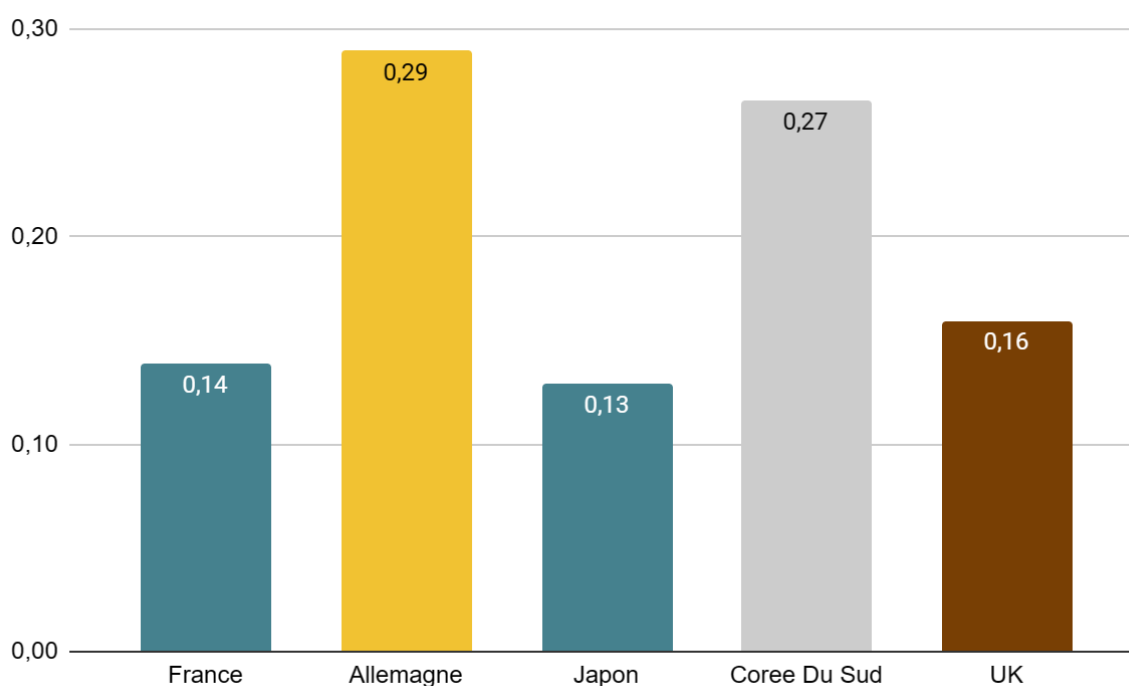
Calcul des Moyennes :

SAE2.04 - Statistiques

- Pour la France :
 $(0,10+0,13+0,15+0,10+0,10+0,22+0,16+0,16+0,08+0,16+0,20+0,08+0,17)/13 =$
0,14
- Pour l'Allemagne:
 $(0,41+0,33+0,19+0,34+0,29+0,29+0,36+0,33+0,22+0,38+0,14+0,16+0,32)/13 =$
0,29
- Pour le Japon :
 $(0,11+0,06+0,11+0,03+0,33+0,02+0,22+0,02+0,16+0,09+0,32+0,03+0,16)/13 =$
0,13
- Pour la Corée Du Sud:
 $(0,33+0,40+0,40+0,36+0,10+0,33+0,23+0,36+0,26+0,22+0,12+0,23+0,10)/13 =$
0,27
- Pour le Royaume Uni :
 $(0,15+0,14+0,01+0,12+0,33+0,19+0,06+0,19+0,04+0,18+0,29+0,37)/13 = 0,16$

(**Remarque:** en effet nous divisons par 13 le résultat alors qu'il n'y a que 12 proportion car lors des JO de 1998 le Royaume Uni n'a remporté qu'une médaille et elle a été remporté par un homme, c'est pour cela qu'il n'y a pas de proportion pour cette année cependant cela veut seulement dire qu'il y a eu 0% de médaillé parmi les femmes).

Nous obtenons après le calcul des moyennes le graphiques suivant :



SAE2.04 - Statistiques

On obtient alors le classement suivant :

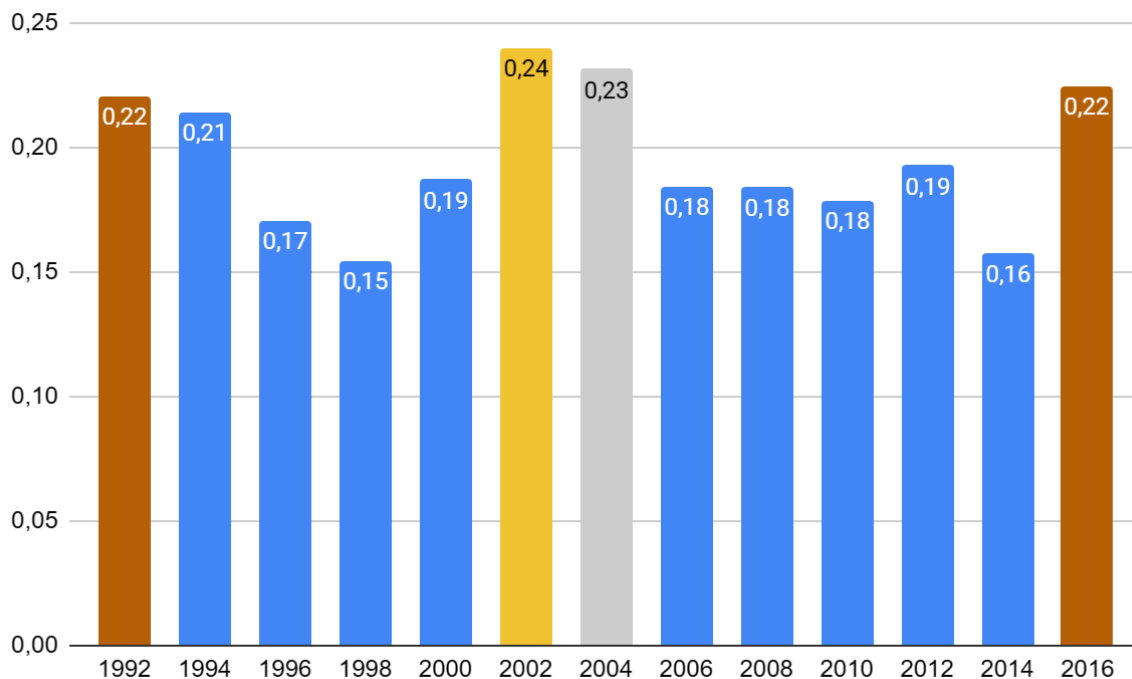
L'Allemagne est en première position ce qui veut dire que les femmes y sont le plus médaillées suivi par la Corée du sud et le Japon.

Maintenant nous souhaitons savoir quelle année a vu le plus de médaillés parmi les femmes, après avoir fait les calculs nous avons obtenu le tableau suivant :

Annee	Moyenne Proportion
1992	0,22
1994	0,21
1996	0,17
1998	0,15
2000	0,19
2002	0,24
2004	0,23
2006	0,18
2008	0,18
2010	0,18
2012	0,19
2014	0,16
2016	0,22

Ce tableaux nous donne ensuite le graphique suivant :

SAE2.04 - Statistiques



Nous concluons donc que l'année 2002 est l'année durant laquelle il y a eu en moyenne parmi nos 5 équipes le plus haut taux de médaillés parmi les femmes.

Question 3bvi. De la proportion de femmes parmi les médaillés :

Requetes :

```
DELETE FROM effect;
```

Insert Into Effect

```
SELECT Distinct t.NOC,j.annee,count(p.id)
FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m, athlete as a
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND j.annee>=1992
And j.annee<=2016
```

SAE2.04 - Statistiques

```
AND p.idM=m.idM
AND p.ID=a.ID
AND( pa.region='France'
OR pa.region='UK'
OR pa.region='Japan'
OR pa.region='Germany'
OR pa.region='South Korea')
AND m.nomMedal<>'NA'
Group By t.NOC,j.annee;
```

```
\copy (SELECT 'Question 3 b vi' as Question ,pa.region, a.SEX,m.nomMedal<>'NA' as
Medaille, j.annee,Cast(count(p.id )AS numeric(6,3))/e.effectif as NbDeFemme
FROM participe as p , jo as j,TEAM as t , Pays as pa,Medal as m, athlete as a , effect as
e
WHERE p.idjo=j.idjo
AND p.idteam=t.idteam
AND t.NOC=pa.NOC
AND e.NOC=t.NOC
AND e.year=j.annee
AND j.annee>=1992 And j.annee<=2016
AND p.idM=m.idM
AND p.ID=a.ID
AND( pa.region='France'
OR pa.region='UK'
OR pa.region='Japan'
OR pa.region='Germany'
OR pa.region='South Korea')
AND m.nomMedal<>'NA'
GROUP BY t.NOC,pa.region,j.annee,a.SEX,m.nomMedal<>'NA',e.effectif
ORDER BY pa.region , j.annee)
TO 'Question3bvi.csv' WITH (format CSV , DELIMITER ',' , HEADER);
```

Graphique :

Pour cette dernière question nous avons décidé de regarder nos données sous le même angle que a la question précédente c'est à dire effectuer un classement par Pays

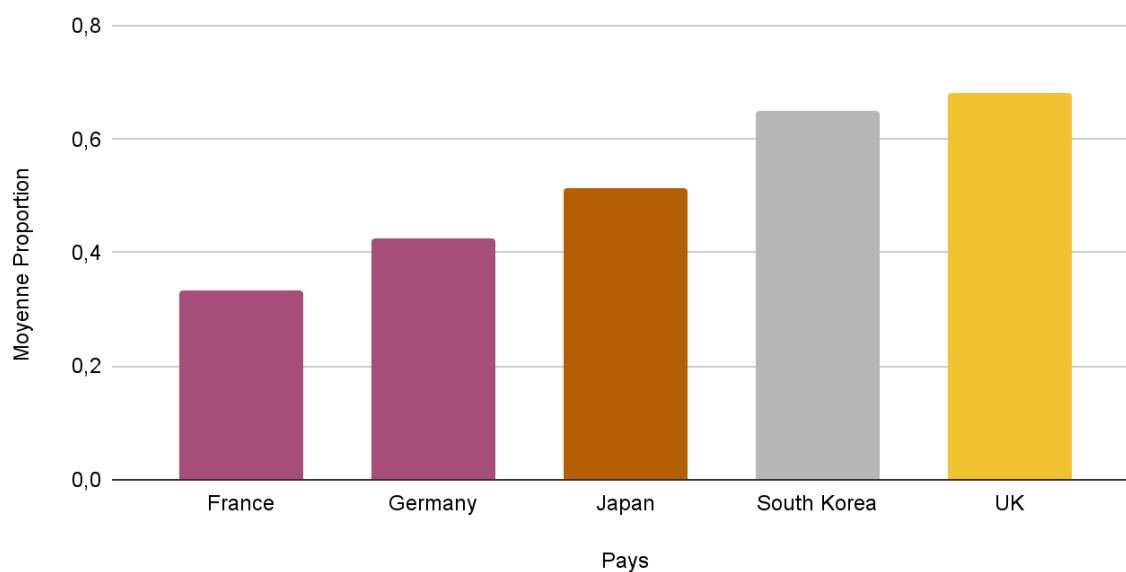
Après avoir calculé les moyennes nous obtenons ce tableaux :

SAE2.04 - Statistiques

Pays	Moyenne Proportion
France	0,3333333333
Germany	0,425
Japan	0,51328125
South Korea	0,6510416667
UK	0,68

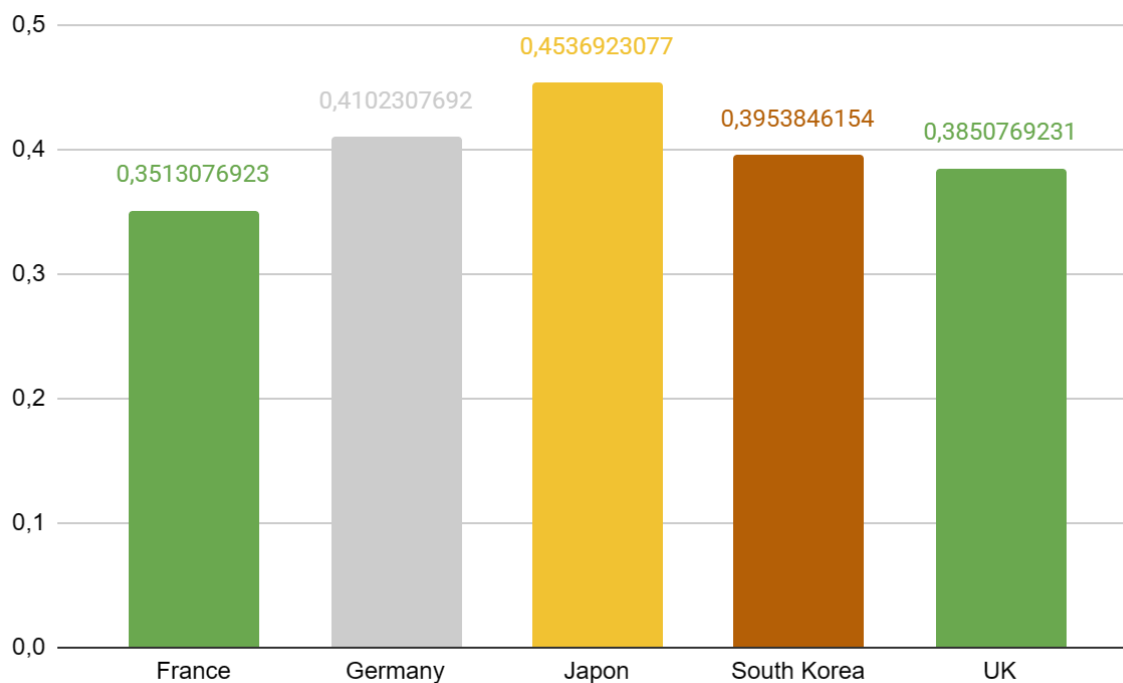
Et obtenons le graphique suivant :

Moyenne Proportion de Femmes Parmi Les Médaille par rapport au Pays



Et ce qui est intéressant c'est qu' il semblerait logique que le pays ayant le plus inclut les femmes ait la plus haute proportion de femmes parmi les médaillés et inversement, cependant si on reprend le graphique sur la moyenne des proportion des femmes parmi les athletes par pays on se rend compte que cela n'est pas forcément vrai.

SAE2.04 - Statistiques



(Tableau de la moyenne des proportion de femmes parmi les athletes par pays)

En effet on peut voir que même si le Royaume Uni était l'avant dernier pays à le plus inclure les femmes, il se trouve être le premier lorsqu'il s'agit des médailles attribuées aux femmes parmi les médaillés.