


Nom de la SAE	SAE projet		Semestre / Période	Semestre 2
volume horaire consacré par l'étudiant	Avec enseignant	16h	en autonomie	25 h
Coéquipiers :	Rachid Sahli		Hanna sysoieva	
	Ndoumbe Seck			

Sujet spécifique	Analyse de la parité Hommes Femmes aux jeux olympiques
Objectifs	<ul style="list-style-type: none"> <li>Faire comprendre à l'étudiant l'interdisciplinarité de la statistique et de l'informatique décisionnelle</li> <li>L'initier à la conduite de projet et à la planification</li> <li>Le confronter à toutes les étapes d'un projet en science des données</li> <li>Lui faire prendre conscience des problématiques de confidentialité des données.</li> </ul>
Livrables	 <p>The deliverables include:</p> <ul style="list-style-type: none"> <li><b>PROJET EN BUT 1 SD</b>: La parité entre femmes et hommes existe-elle dans les jeux olympiques d'hiver ? Réalisé par Ayoub Errahmani, Ndoumbé Seck, Hanna Sysoieva, Rachid Sahli.</li> <li><b>Répartition des participants par pays pour les jeux d'hiver</b>: A bar chart showing the number of participants by country.</li> <li><b>Evolution d'âge moyen aux jeux olympiques d'hiver</b>: A line chart showing the average age of participants from 1920 to 2016 for men and women.</li> <li><b>Répartition des participants par sport et année pour les jeux d'hiver</b>: A complex chart showing the distribution of participants by sport and year.</li> <li><b>CONCLUSION SCIENTIFIQUE</b>: Progression significative: 26% de femmes en 1988 -&gt; 45% de femmes en 2016. Nombre de disciplines exclusivement féminines: 6 en 1988 -&gt; 13 en 2016.</li> <li><b>CONCLUSION DE RÉCUL</b>: Recherche des données, Travail en équipe, Remplacement de membre de groupe, Matériau de R.</li> </ul>

## Bilan de la SAE

(reproduire le tableau autant de fois que de compétences mobilisées dans la SAÉ)

Compétence	Analyser statistiquement des données
Apprentissages critiques sollicités	<ul style="list-style-type: none"> <li>Comprendre qu'une analyse correcte ne peut émaner que de données propres et préparées</li> </ul>
	<ul style="list-style-type: none"> <li>Comprendre l'intérêt des synthèses numériques et graphiques pour mettre en évidence des liaisons entre variables.</li> </ul>
	<ul style="list-style-type: none"> <li>Réaliser que les sources de données ont des caractéristiques propres à considérer (variation, précision, mise à jour...)</li> </ul>
Composantes essentielles à respecter	En tenant compte du contexte de l'étude (économique, socio-démographique, commerciale, clinique...)
	En identifiant et en mettant en œuvre les techniques adaptées aux données complexes (données massives, données mal structurées, flux de données...)
	En mettant en évidence les grandes tendances et les informations principales

Compétence	Valoriser une production dans un contexte professionnel
Apprentissages critiques sollicités	<ul style="list-style-type: none"> <li>Identifier l'importance de contextualiser ses données</li> </ul>
	<ul style="list-style-type: none"> <li>Lors de la restitution des résultats, mesurer l'importance d'explicitement également la démarche suivie.</li> </ul>
	<ul style="list-style-type: none"> <li>Mesurer l'importance d'une expression précise et nuancée dans la communication en français et dans une langue étrangère des résultats</li> </ul>
Composantes essentielles à respecter	En interprétant et contextualisant les résultats (citations, vérification des sources, esprit critique)
	En utilisant la forme de restitution adaptée
	En s'exprimant correctement, aussi bien en français que dans une langue étrangère, à l'oral comme à l'écrit

### Ma démarche

Savoirs / connaissances	Savoir-faire	Savoir-être
Analyse de la parité des hommes et des femmes aux jeux olympiques .	<ul style="list-style-type: none"> <li>Analyser</li> <li>Programmer</li> <li>Comparer</li> <li>Traiter</li> <li>Valoriser</li> </ul>	<ul style="list-style-type: none"> <li>Collaboration en équipe</li> </ul>

### Evaluation du résultat

- Ce que je trouve bien réalisé, pourquoi ?

## Fiche bilan SAE

Les graphiques des Jeux Olympiques, illustrant la participation des hommes et femmes dans les sports d'été et d'hiver, montrent l'équité de genre de manière efficace. Cette approche, en valorisant autant les athlètes masculins que féminins, encourage l'égalité et l'appréciation des talents sans distinction de genre. C'est une démarche significative pour promouvoir l'inclusion et inspirer les futures générations, illustrant bien l'évolution vers une plus grande égalité dans le sport.

- Ce que je n'ai pas bien compris ; ce qui serait à améliorer pour une prochaine fois : pourquoi ? comment ?

Bien que la comparaison entre disciplines et l'analyse sur plusieurs années soient essentielles, l'emploi de visualisations interactives, comme des cartes de chaleur (heatmaps), des graphiques en bulles dynamiques ou des visualisations en temps réel, pourrait rendre les données plus accessibles et captivantes pour un public plus large.

De plus, l'intégration d'analyses prédictives pour anticiper les tendances futures en matière de participation et de performances des athlètes féminines et masculines pourrait offrir une valeur ajoutée significative

## Eléments de preuve, ce que je peux montrer

(Choisir des éléments précis à mettre annexe)

### 1) Extrait de code

```
install.packages("dplyr")
install.packages("reshape2")
install.packages("spread")
install.packages("tidyverse")
library(tidyverse)
library(readr)
library(dplyr)
library(reshape2)

athlete_events <- read_csv("C:/Users/ayoub/OneDrive/Documents/sae projet/athlete_events.csv")

# Filtrer les données pour les sports d'hiver
athlete_events_winter <- athlete_events %>% filter(Season == "Winter")

# Regrouper les données par discipline sportive et sexe
sex_discipline <- athlete_events_winter %>%
  group_by(Sport, Sex) %>%
  summarise(count = n()) %>%
  pivot_wider(names_from = Sex, values_from = count, names_prefix = "") %>%
  mutate(Total = F + M,
         Female_pct = F / Total,
         Male_pct = M / Total,
         Diff_pct = abs(Female_pct - Male_pct)) %>%
  arrange(desc(Diff_pct))

# Afficher les 10 disciplines sportives les plus paritaires et les moins paritaires
cat("Disciplines sportives les plus paritaires :\n")
print(head(sex_discipline, 10))
cat("\nDisciplines sportives les moins paritaires :\n")
print(tail(sex_discipline, 10))

# Créer un graphique en barres montrant la participation relative des femmes et des hommes dans chaque discipline sportive
barplot_df <- melt(sex_discipline, id.vars = "Sport", measure.vars = c("Female_pct", "Male_pct"), variable.name = "Sex", value.name = "Participation_pct")

ggplot(barplot_df, aes(x = Sport, y = Participation_pct, fill = Sex)) +
  geom_bar(stat = "identity", position = "stack", color = "black") +
  geom_text(aes(label = round(Participation_pct, 2)), position = position_stack(vjust = 0.5), color = "white", size = 3) +
  scale_fill_manual(values = c("blue", "orange"), labels = c("Femme", "Homme")) +
  theme_classic() +
  coord_flip() +
  labs(title = "Participation relative des femmes et des hommes aux sports d'hiver",
       x = "Discipline sportive",
       y = "Pourcentage de participation") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.y = element_text(size = 9))

ggplot(barplot_df, aes(x = Sport, y = Participation_pct, fill = Sex)) +
  geom_bar(stat = "identity", position = "stack", color = "black") +
  geom_text(aes(label = round(Participation_pct, 2)), position = position_stack(vjust = 0.5), color = "white", size = 5) +
  scale_fill_manual(values = c("blue", "orange"), labels = c("Femme", "Homme")) +
  theme_classic() +
  coord_flip() +
  labs(title = "Participation relative des femmes et des hommes aux sports d'hiver",
       x = "Disciplines sportives",
       y = "Pourcentage de participation") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.y = element_text(size = 9),
        axis.text.x = element_text(size = 9))
```

2) Extrait de graphique réalisé :

