

Notebook_chapitre_7_Visualisation_avec_Seaborn_light

February 1, 2023

1 Importation des données et nettoyage

```
[1]: import pandas as pd
donnees=pd.read_csv("../datasets/
↳120-years-of-olympic-history-athletes-and-results.csv", index_col=[1])
donnees['Medal']=donnees['Medal'].fillna("None")
donnees_NaN_supprimes = donnees[pd.notnull(donnees['Age'])]
donnees_NaN_supprimes = donnees_NaN_supprimes[pd.
↳notnull(donnees_NaN_supprimes['Height'])]
donnees_NaN_supprimes = donnees_NaN_supprimes[pd.
↳notnull(donnees_NaN_supprimes['Weight'])]
donnees[['Age', 'Height', 'Weight']] = donnees[['Age', 'Height', 'Weight']].
↳fillna(donnees[['Age', 'Height', 'Weight']].mean())
donnees_NaN_replaces=donnees.drop_duplicates()
donnees_NaN_supprimes=donnees_NaN_supprimes.drop_duplicates()
print("Tableau avec NaN remplacés")
print(donnees_NaN_replaces.shape)
print("Tableau avec suppression des lignes contenant des NaN")
print(donnees_NaN_supprimes.shape)
```

Tableau avec NaN remplacés

(269731, 14)

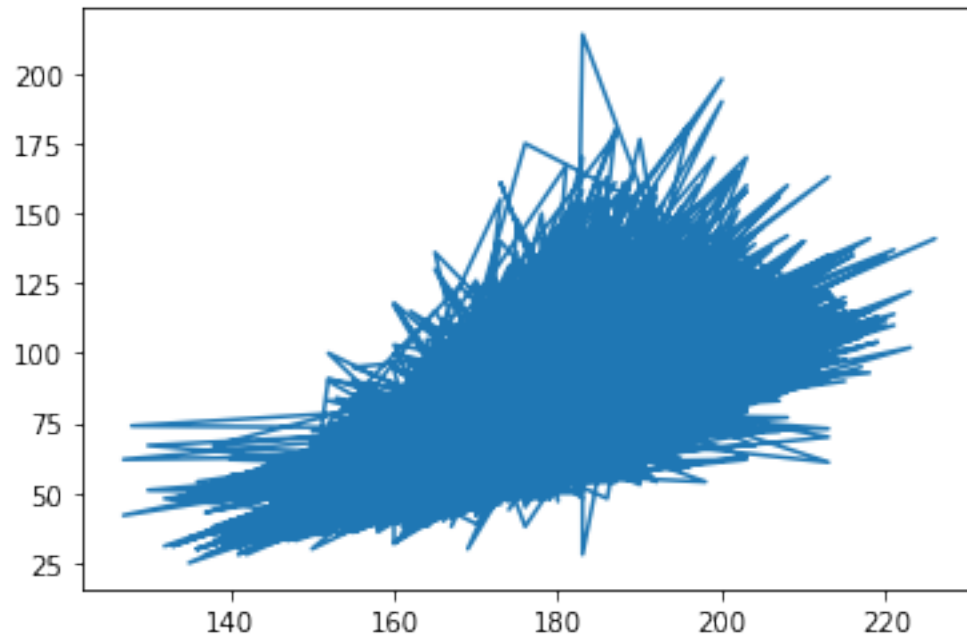
Tableau avec suppression des lignes contenant des NaN

(206152, 14)

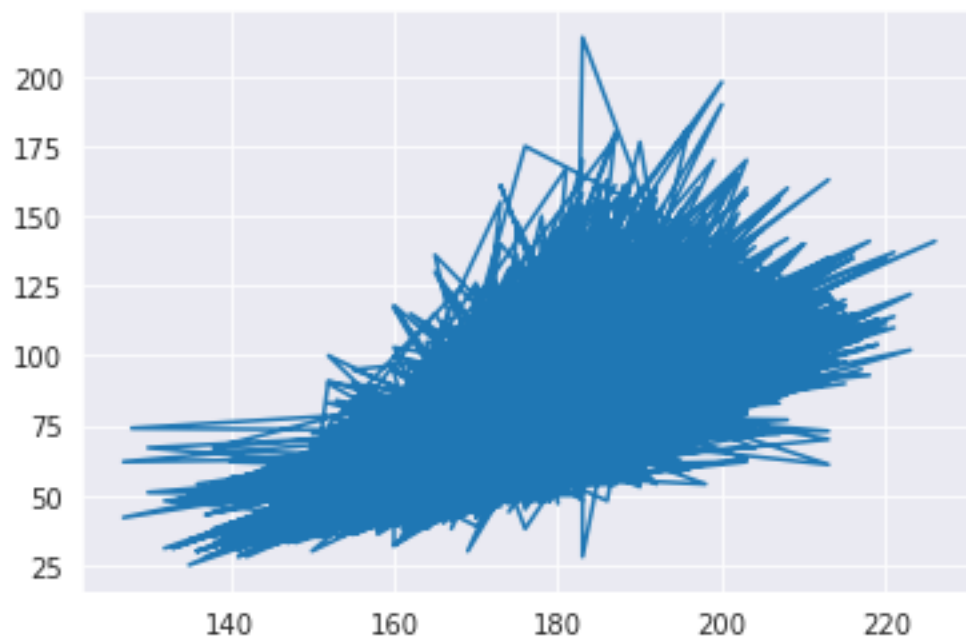
2 2. L'esthétique des figures avec Seaborn (Aesthetic)

2.1 2.1. Paramétrer les styles Seaborn (thèmes)

```
[2]: import matplotlib.pyplot as plt
plt.figure()
plt.subplot(111)
plt.plot(donnees_NaN_supprimes["Height"], donnees_NaN_supprimes["Weight"])
plt.show()
```

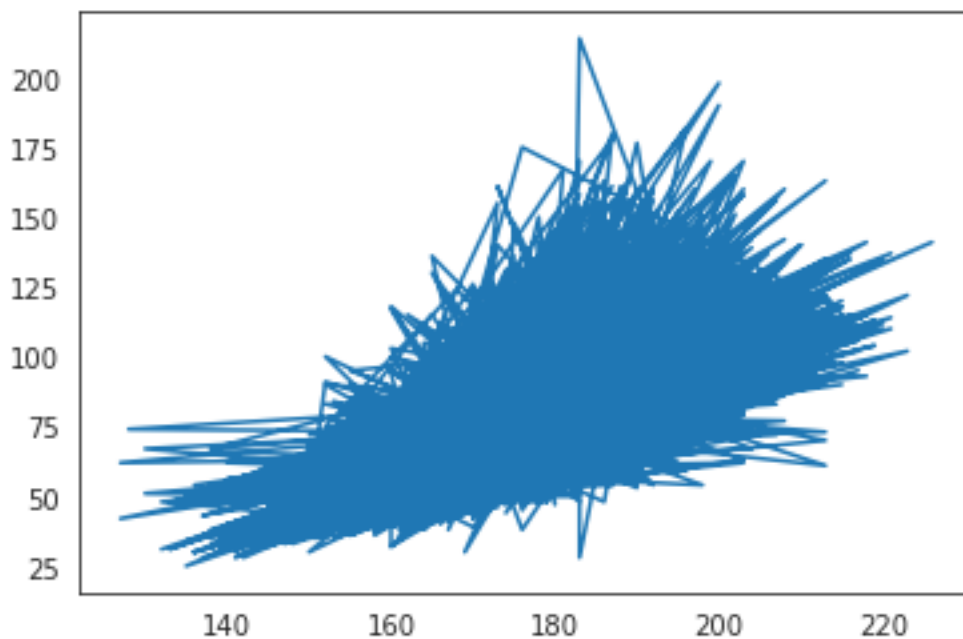


```
[3]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("darkgrid")
plt.figure()
plt.subplot(111)
plt.plot(donnees_NaN_supprimes["Height"], donnees_NaN_supprimes["Weight"])
plt.show()
```

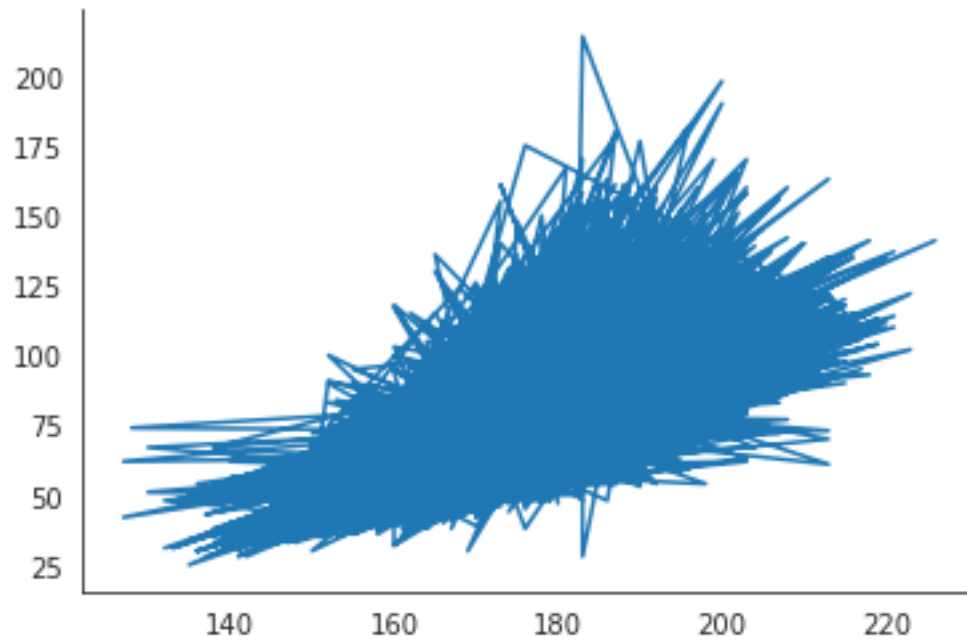


2.2 2.2. Supprimer les axes

```
[4]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("white")
plt.figure()
plt.subplot(111)
plt.plot(donnees_NaN_supprimes["Height"], donnees_NaN_supprimes["Weight"])
plt.show()
```

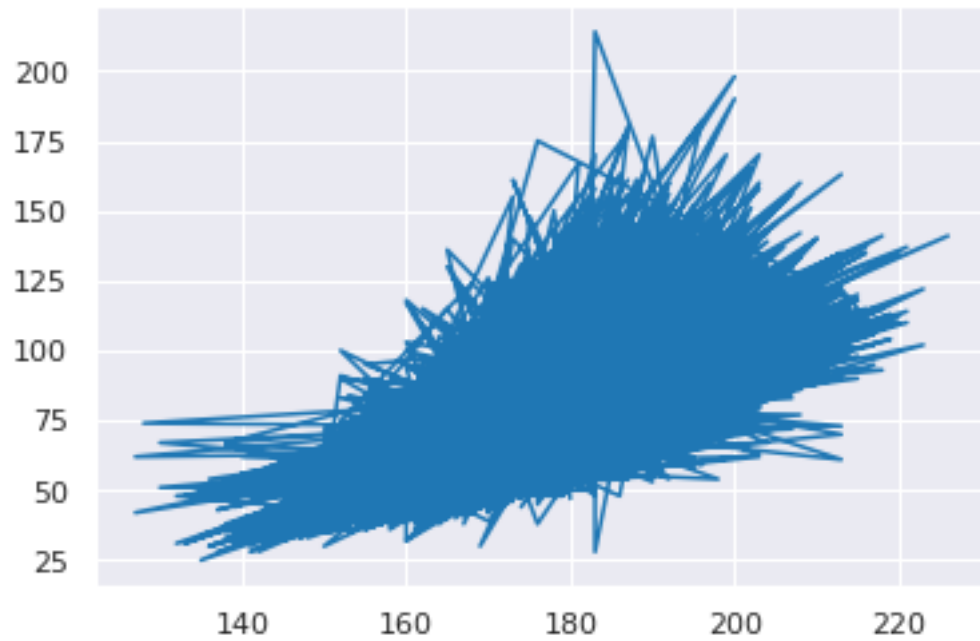


```
[5]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("white")
plt.figure()
plt.subplot(111)
plt.plot(donnees_NaN_supprimes["Height"], donnees_NaN_supprimes["Weight"])
sns.despine()
plt.show()
```

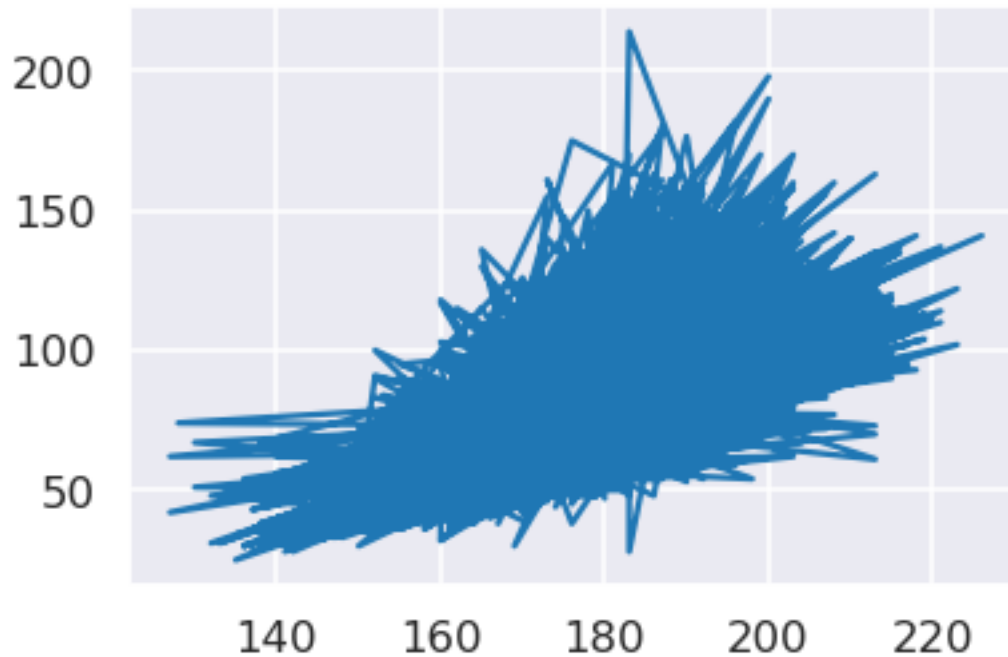


2.3 2.3. Paramétrer les contextes avec Seaborn

```
[6]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("darkgrid")
sns.set_context("notebook")
plt.figure()
plt.subplot(111)
plt.plot(donnees_NaN_supprimes["Height"], donnees_NaN_supprimes["Weight"])
plt.show()
```



```
[7]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("darkgrid")
sns.set_context("talk")
plt.figure()
plt.subplot(111)
plt.plot(donnees_NaN_supprimes["Height"], donnees_NaN_supprimes["Weight"])
plt.show()
```



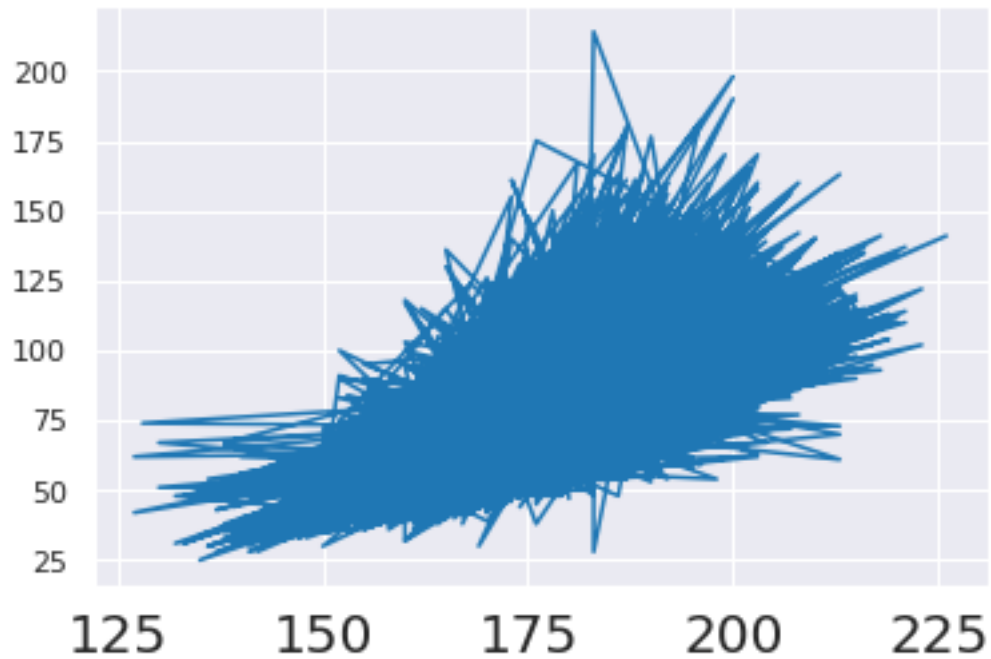
```
[8]: sns.set_context("notebook")
sns.plotting_context()
```

```
[8]: {'font.size': 12.0,
      'axes.labelsize': 12.0,
      'axes.titlesize': 12.0,
      'xtick.labelsize': 11.0,
      'ytick.labelsize': 11.0,
      'legend.fontsize': 11.0,
      'axes.linewidth': 1.25,
      'grid.linewidth': 1.0,
      'lines.linewidth': 1.5,
      'lines.markersize': 6.0,
      'patch.linewidth': 1.0,
      'xtick.major.width': 1.25,
      'ytick.major.width': 1.25,
      'xtick.minor.width': 1.0,
      'ytick.minor.width': 1.0,
      'xtick.major.size': 6.0,
      'ytick.major.size': 6.0,
      'xtick.minor.size': 4.0,
      'ytick.minor.size': 4.0,
      'legend.title_fontsize': 12.0}
```

```
[9]: sns.set_context("talk")
sns.plotting_context()
```

```
[9]: {'font.size': 18.0,
      'axes.labelsize': 18.0,
      'axes.titlesize': 18.0,
      'xtick.labelsize': 16.5,
      'ytick.labelsize': 16.5,
      'legend.fontsize': 16.5,
      'axes.linewidth': 1.875,
      'grid.linewidth': 1.5,
      'lines.linewidth': 2.25,
      'lines.markersize': 9.0,
      'patch.linewidth': 1.5,
      'xtick.major.width': 1.875,
      'ytick.major.width': 1.875,
      'xtick.minor.width': 1.5,
      'ytick.minor.width': 1.5,
      'xtick.major.size': 9.0,
      'ytick.major.size': 9.0,
      'xtick.minor.size': 6.0,
      'ytick.minor.size': 6.0,
      'legend.title_fontsize': 18.0}
```

```
[10]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("darkgrid")
sns.set_context("notebook",rc={"xtick.labelsize": 20})
plt.figure()
plt.subplot(111)
plt.plot(donnees_NaN_supprimes["Height"],donnees_NaN_supprimes["Weight"])
plt.show()
sns.set_context("notebook",rc={"xtick.labelsize": 16.5})
```



2.4 2.4. Les palettes de couleur avec Seaborn

2.4.1 2.4.1. Choisir une palette de couleurs existante

```
[11]: sns.palplot(sns.color_palette())
```



```
[12]: sns.palplot(sns.color_palette("dark",15))
```



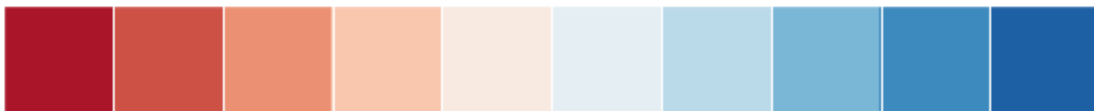
```
[13]: sns.palplot(sns.color_palette("Paired"))
```




```
[14]: sns.palplot(sns.color_palette("OrRd",10))  
sns.palplot(sns.color_palette("Blues",10))
```



```
[15]: sns.palplot(sns.color_palette("RdBu", 10))
```



2.4.2 2.4.2. Créer sa propre palette de couleurs

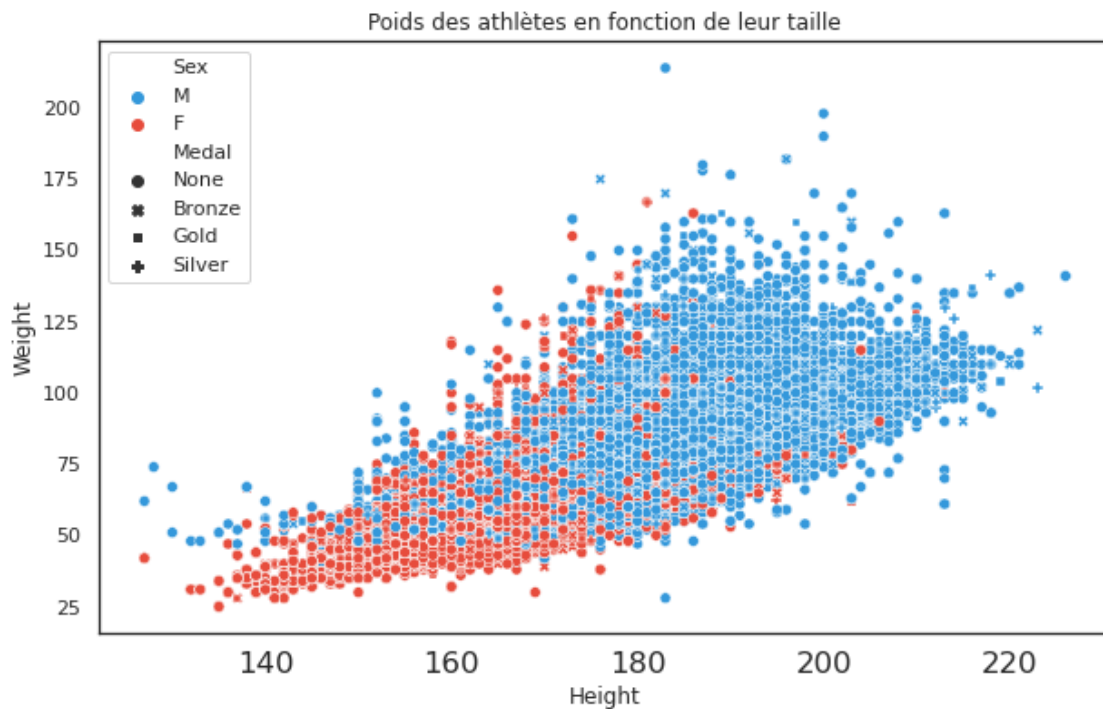
```
[16]: palette= sns.color_palette(["#9b59b6", "#3498db", "#e74c3c", "#34495e", "#2ecc71"])  
sns.set_palette(palette)  
sns.palplot(palette)
```



3 3. Les différents types de graphiques

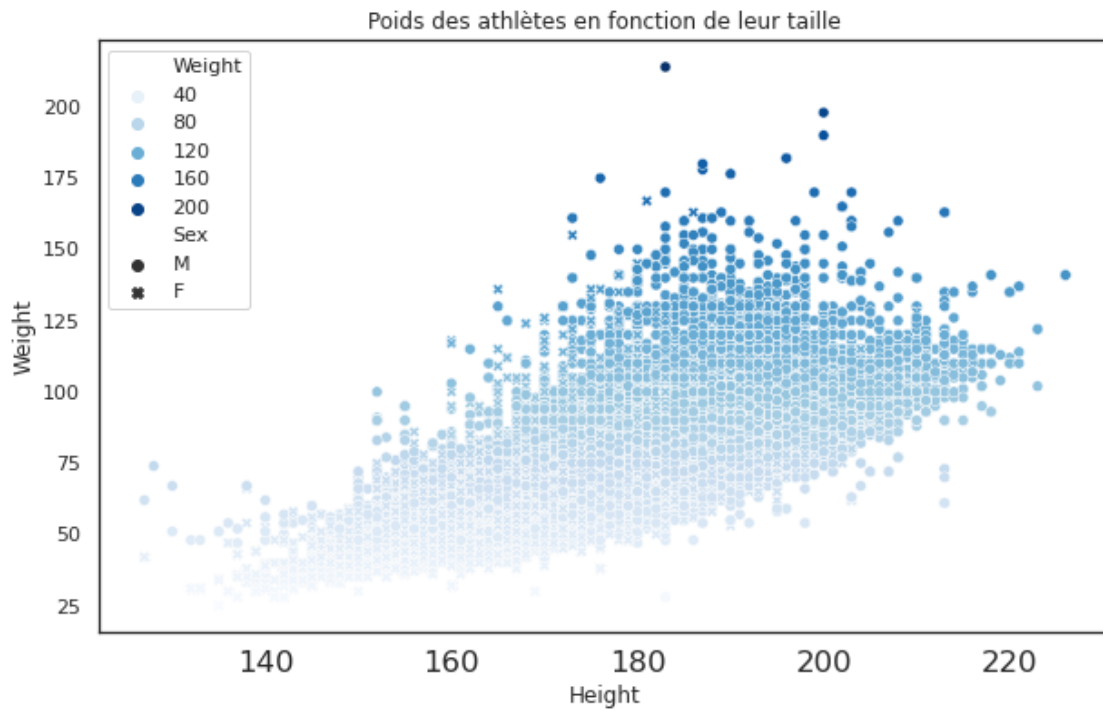
3.1 3.2. Nuage de points (scatterplot)

```
[17]: import matplotlib.pyplot as plt
import seaborn as sns
palette= sns.color_palette(["#3498db","#e74c3c"])
sns.set_palette(palette)
sns.set_style("white")
plt.figure(figsize=[10,6])
plt.subplot(111)
sns.scatterplot(data=donnees_NaN_supprimes, x='Height', y='Weight', hue='Sex', style='Medal')
plt.title("Poids des athlètes en fonction de leur taille")
plt.legend(loc='upper left')
plt.show()
```



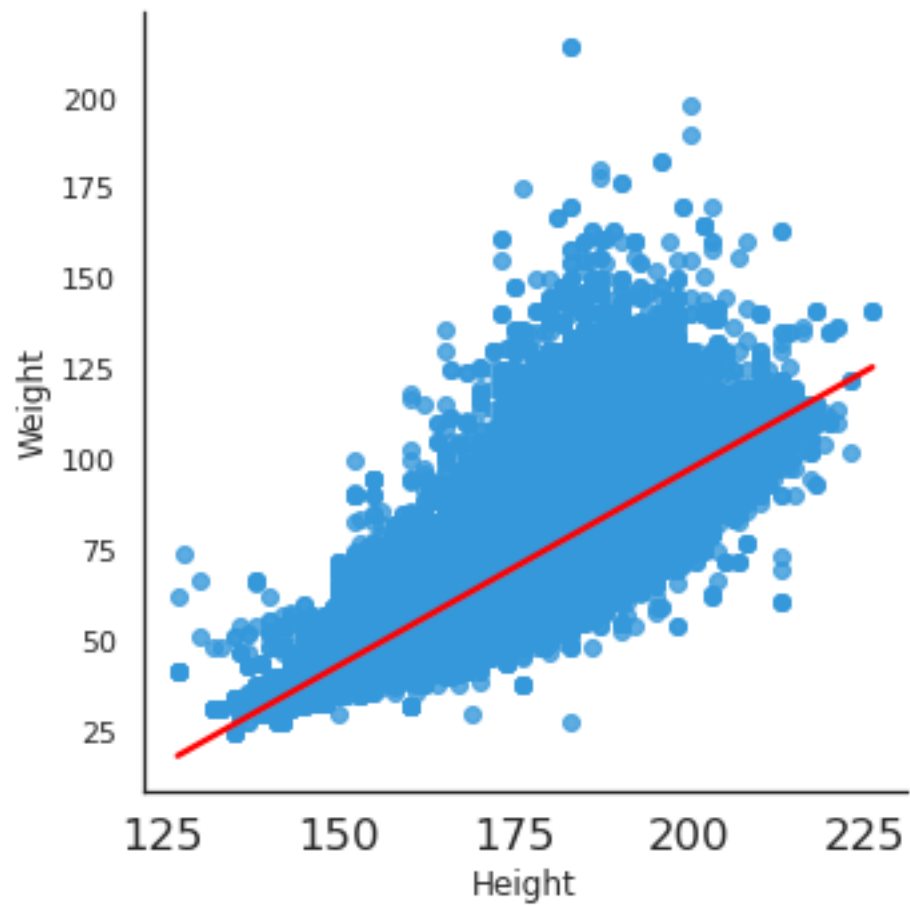
```
[18]: import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=[10,6])
plt.subplot(111)
sns.scatterplot(data=donnees_NaN_supprimes, x='Height', y='Weight', hue='Weight', style="Sex", palette="Blues")
plt.title("Poids des athlètes en fonction de leur taille")
```

```
plt.legend(loc='upper left')
plt.show()
```

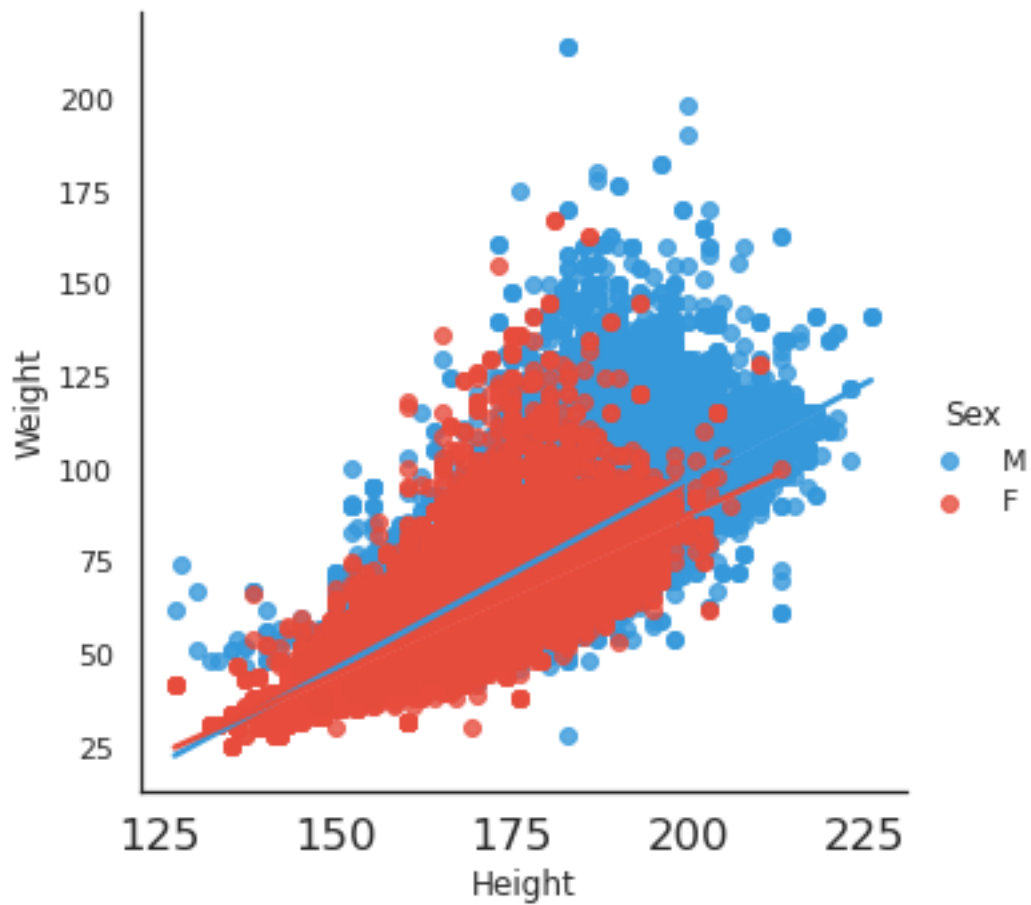


3.2 3.3. Graphiques de régression

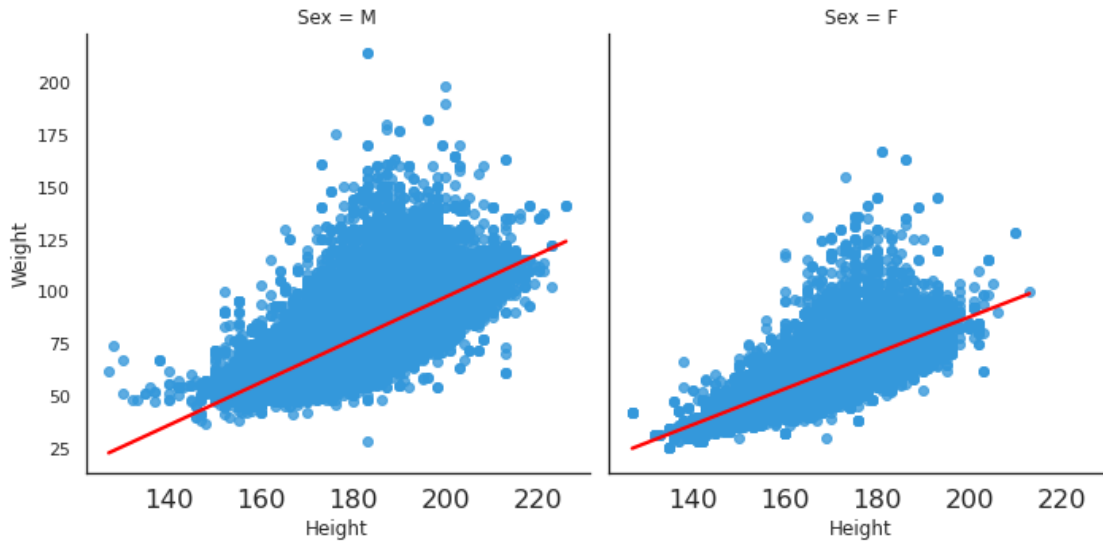
```
[19]: import matplotlib.pyplot as plt
import seaborn as sns
#sns.set(rc={'figure.dpi':200})
sns.lmplot(x='Height', y='Weight', data = donnees_NaN_supprimes, line_kws={'color': 'red'})
plt.show()
```



```
[20]: import matplotlib.pyplot as plt
import seaborn as sns
sns.lmplot(x='Height', y='Weight', data = donnees_NaN_supprimes, hue="Sex")
plt.show()
```

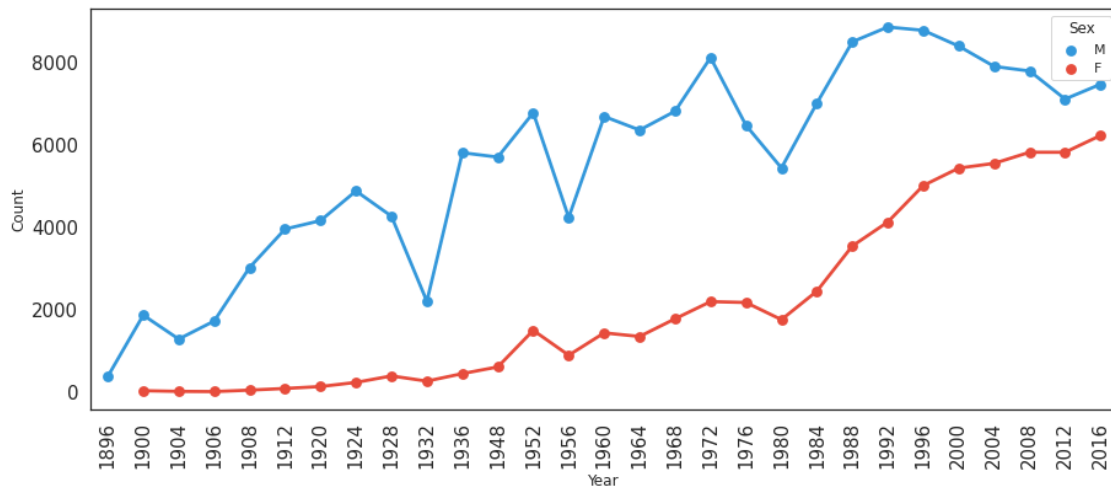


```
[21]: import matplotlib.pyplot as plt
import seaborn as sns
sns.lmplot(x='Height', y='Weight', data = donnees_NaN_supprimes,
           col="Sex", line_kws={'color': 'red'})
plt.show()
```



3.3 3.4. Point plot

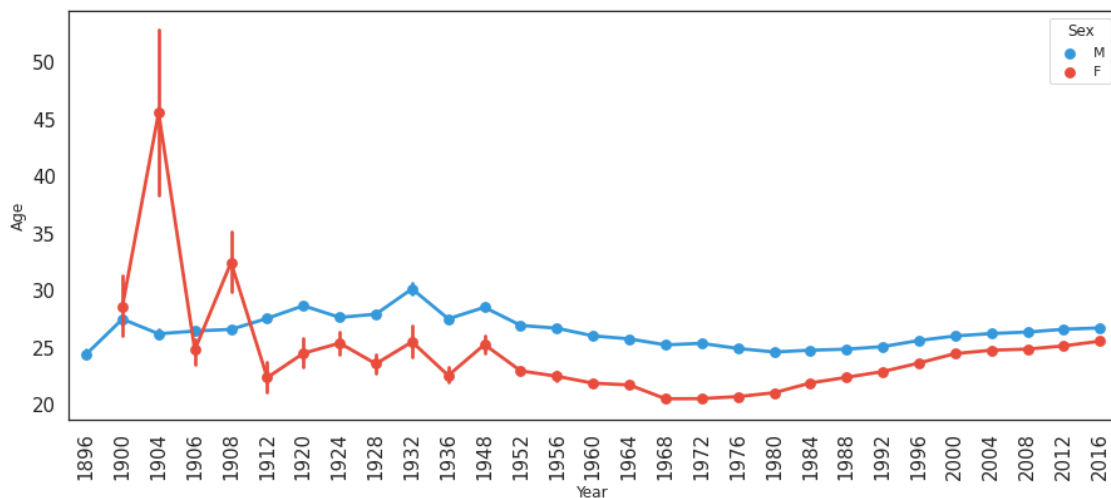
```
[22]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_context("notebook",rc={"xtick.labelsize": 15,"ytick.labelsize": 15})
sns.set_style("white")
plt.figure(figsize=[15,6])
plt.subplot(111)
donnees_summer=donnees_NaN_replaces[donnees_NaN_replaces["Season"]=="Summer"]
donnees_pointplot=donnees_summer.groupby(['Year','Sex']).size().
    ↪reset_index(name='Count')
sns.pointplot(x='Year', y='Count', data = donnees_pointplot, hue="Sex")
plt.xticks(rotation=90)
plt.show()
```



```
[23]: donnees_pointplot.head()
```

```
[23]:   Year Sex  Count
0  1896   M    380
1  1900   F     33
2  1900   M   1865
3  1904   F     16
4  1904   M   1285
```

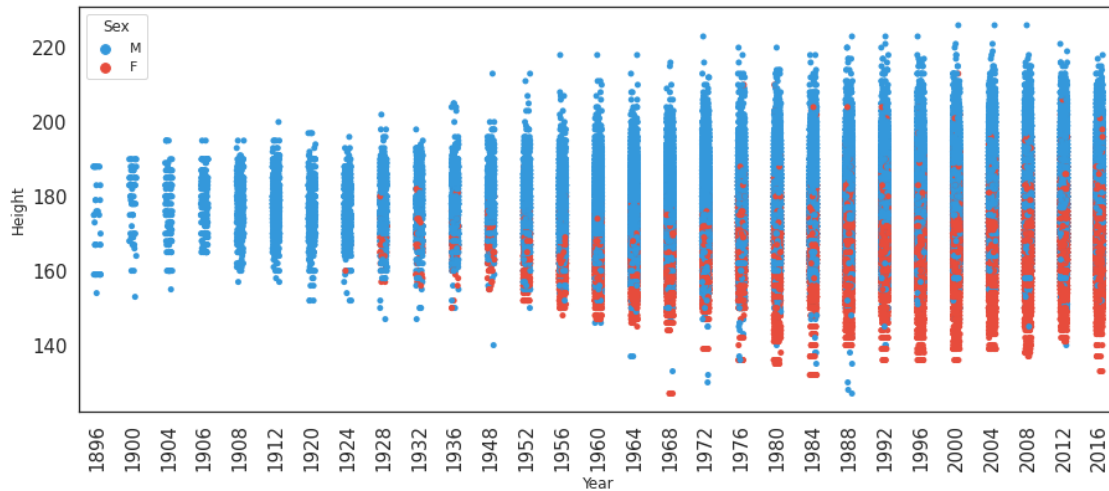
```
[24]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_context("notebook",rc={"xtick.labelsize": 15,"ytick.labelsize": 15})
plt.figure(figsize=[15,6])
plt.subplot(111)
donnees_summer=donnees_NaN_replaces[donnees_NaN_replaces["Season"]=="Summer"]
sns.pointplot(x='Year', y='Age', data = donnees_summer, hue="Sex")
plt.xticks(rotation=90)
plt.show()
```



3.4 3.5 Nuage de points avec une variable qualitative : stripplot

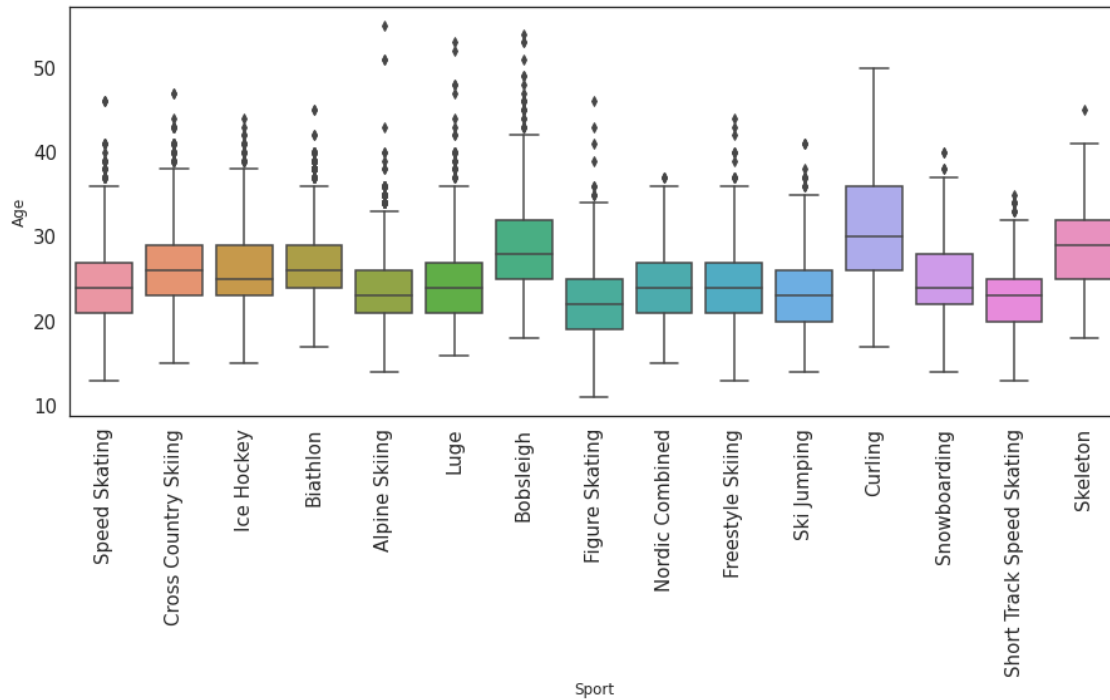
```
[25]: import matplotlib.pyplot as plt
import seaborn as sns
donnees_summer=donnees_NaN_supprimes[donnees_NaN_supprimes["Season"]=="Summer"]
palette= sns.color_palette(["#3498db","#e74c3c"])
sns.set_palette(palette)
plt.figure(figsize=[15,6])
```

```
plt.subplot(111)
sns.stripplot(x='Year', y='Height', hue='Sex', data=donnees_summer)
plt.xticks(rotation=90)
plt.show()
```



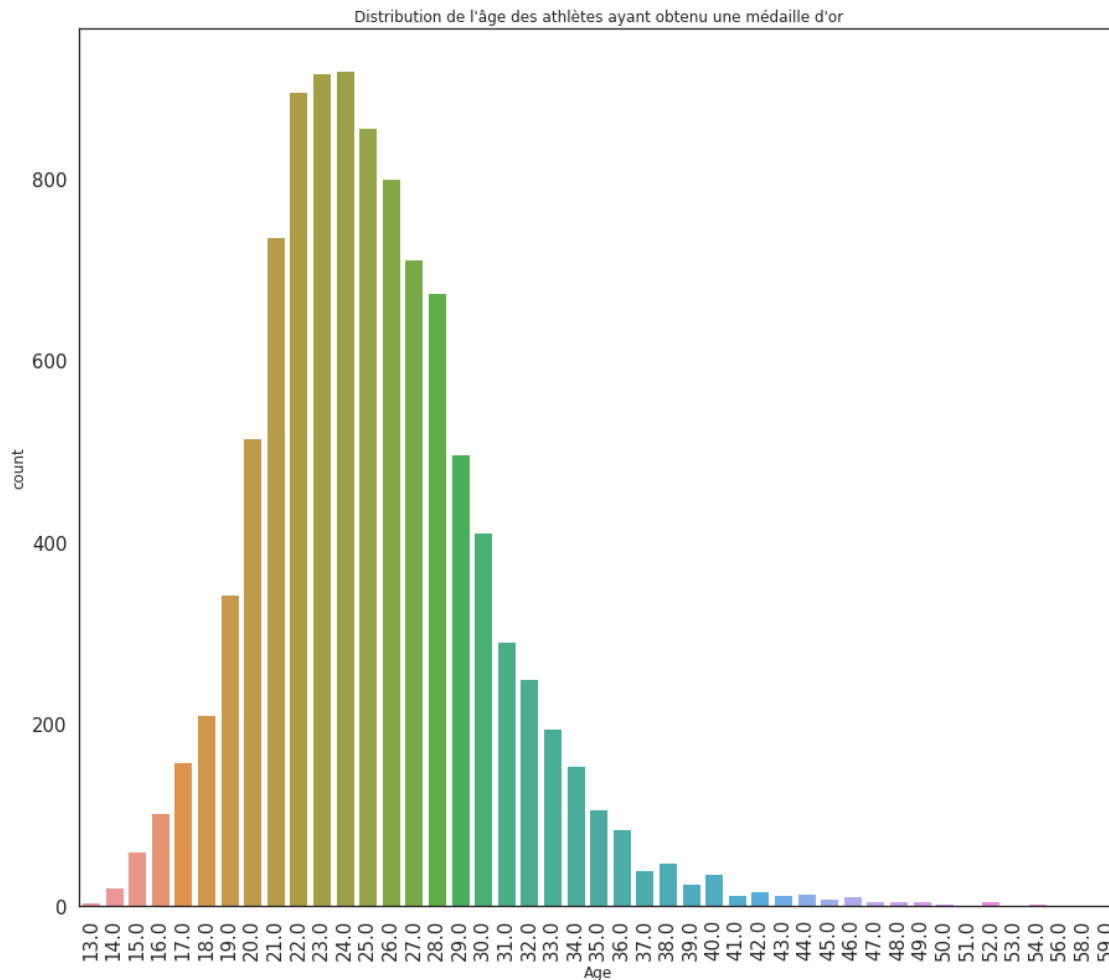
3.5 3.6. Boxplots

```
[26]: import matplotlib.pyplot as plt
import seaborn as sns
donnees_Winter=donnees_NaN_supprimes[donnees_NaN_supprimes["Season"]=="Winter"]
plt.figure(figsize=[15,6])
plt.subplot(111)
sns.boxplot(x='Sport', y='Age', data=donnees_Winter)
plt.xticks(rotation=90)
plt.show()
```

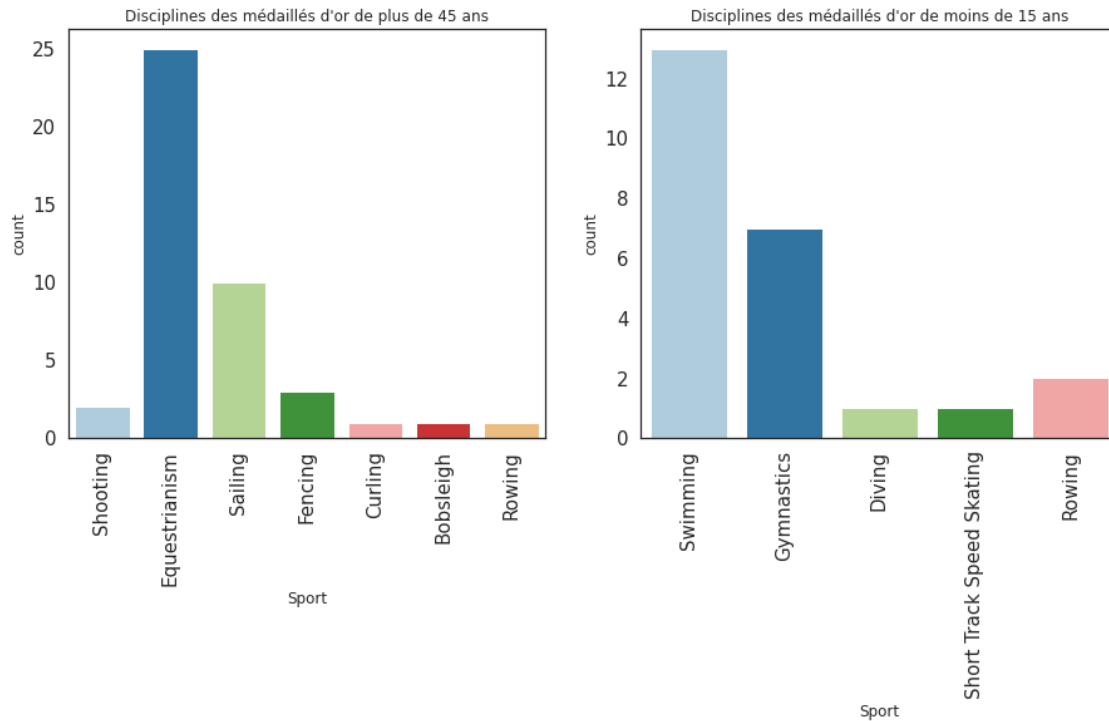
3.6 3.7. Graphique à barres : countplot

```
[27]: import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=[15,13])
plt.subplot(111)
donnees_medaille_or=donnees_NaN_supprimes.
↳loc[donnees_NaN_supprimes["Medal"]=="Gold"]
sns.countplot(data=donnees_medaille_or, x="Age")
plt.title("Distribution de l'âge des athlètes ayant obtenu une médaille d'or")
plt.xticks(rotation=90)
plt.show()
```



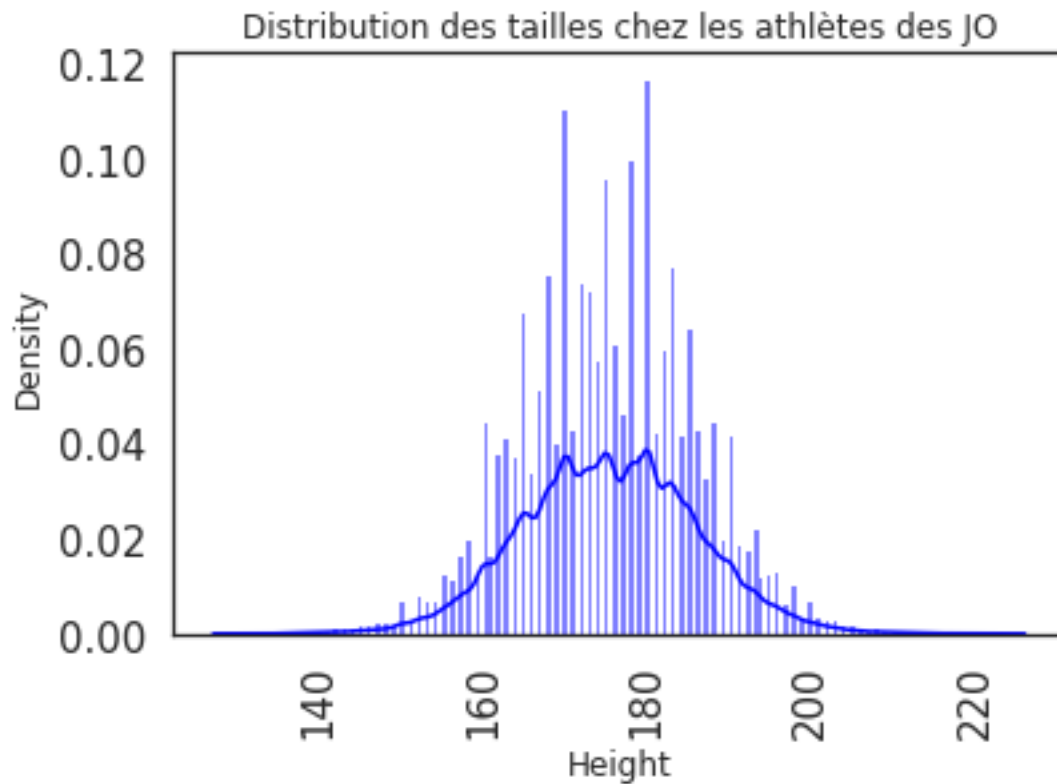
```
[28]: import matplotlib.pyplot as plt
import seaborn as sns
palette=sns.color_palette("Paired")
sns.set_palette(palette)
disciplines_ages = donnees_NaN_supprimes.
    ↳loc[(donnees_NaN_supprimes["Medal"]=="Gold") &
    ↳(donnees_NaN_supprimes["Age"]>45)]
disciplines_jeunes = donnees_NaN_supprimes.
    ↳loc[(donnees_NaN_supprimes["Medal"]=="Gold") &
    ↳(donnees_NaN_supprimes["Age"]<15)]
plt.figure(figsize=[15,6])
plt.subplot(121)
sns.countplot(data= disciplines_ages, x="Sport")
plt.title("Disciplines des médaillés d'or de plus de 45 ans")
plt.xticks(rotation=90)
plt.subplot(122)
```

```
sns.countplot(data= disciplines_jeunes, x="Sport")
plt.title("Disciplines des médaillés d'or de moins de 15 ans")
plt.xticks(rotation=90)
plt.show()
```

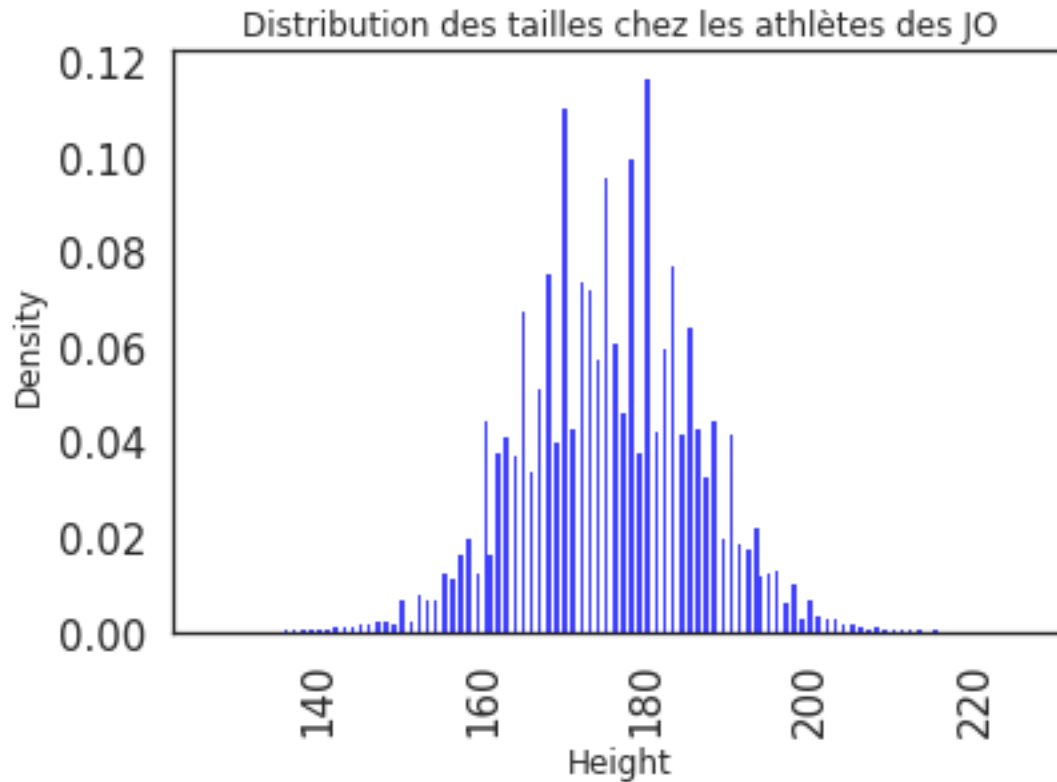


3.7 3.8. Histogrammes

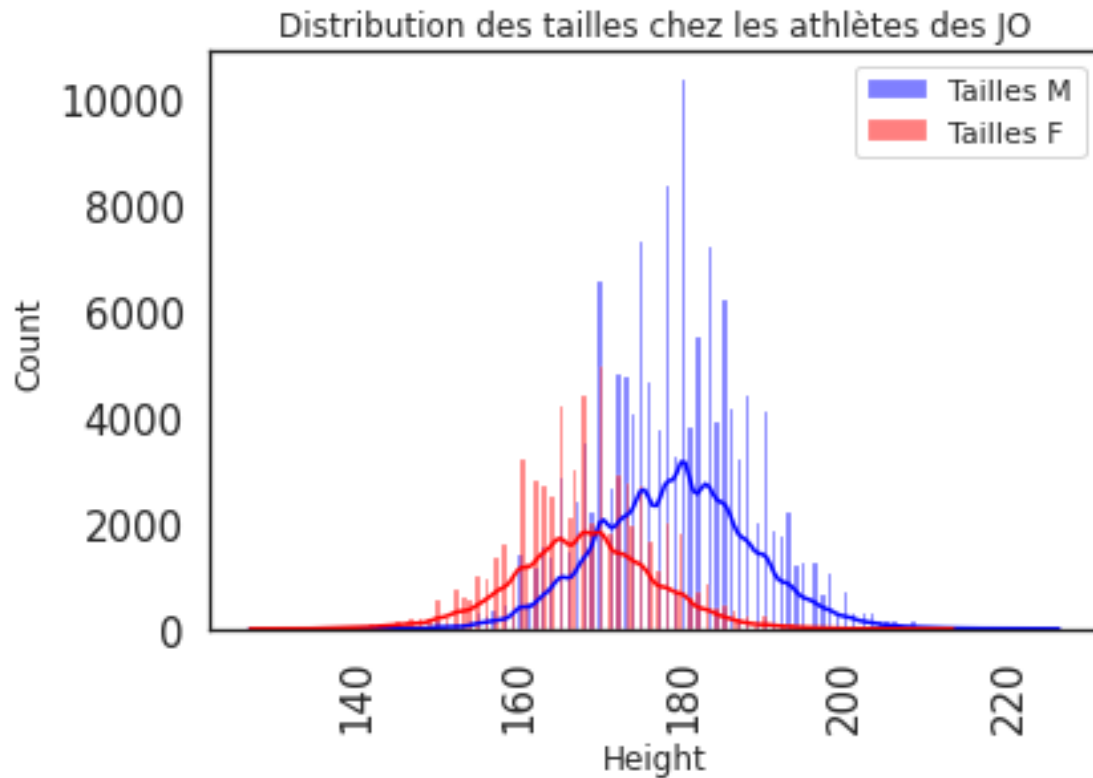
```
[29]: import matplotlib.pyplot as plt
import seaborn as sns
palette= sns.color_palette(["#3498db","#e74c3c"])
sns.set_palette(palette)
plt.figure()
#plt.figure(figsize=[8,6], dpi=300)
plt.subplot(111)
#sns.distplot(donnees_NaN_supprimes["Height"])
#sns.histplot(a, color="red", label="100% Equities", kde=True, stat="density",
↳linewidth=0)
sns.histplot(donnees_NaN_supprimes["Height"],color="blue", label="100%_↳
↳Equities", kde=True, stat="density", linewidth=0)
plt.title("Distribution des tailles chez les athlètes des JO")
plt.xticks(rotation=90)
plt.show()
```



```
[30]: import matplotlib.pyplot as plt
import seaborn as sns
palette= sns.color_palette(["#3498db","#e74c3c"])
sns.set_palette(palette)
plt.figure()
#plt.figure(figsize=[8,6], dpi=300)
plt.subplot(111)
#sns.distplot(donnees_NaN_supprimes["Height"], kde=False)
sns.histplot(donnees_NaN_supprimes["Height"],color="blue", label="100%_
↳Equities", kde=False, stat="density", linewidth=0)
plt.title("Distribution des tailles chez les athlètes des JO")
plt.xticks(rotation=90)
plt.show()
```

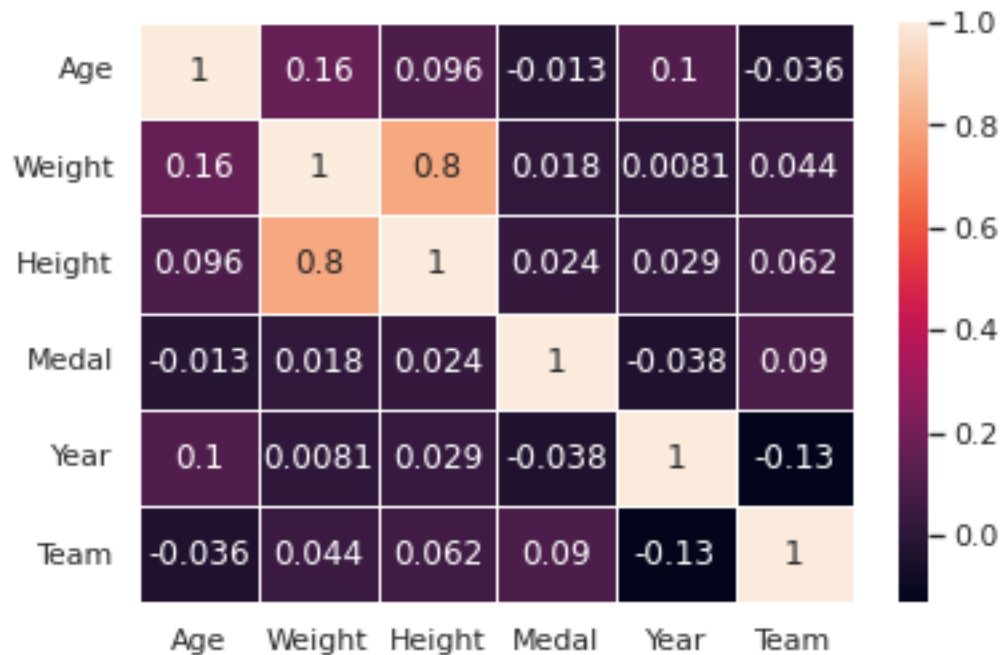


```
[31]: import matplotlib.pyplot as plt
import seaborn as sns
palette= sns.color_palette(["#3498db","#e74c3c"])
sns.set_palette(palette)
valeurs_athletes_feminines=donnees_NaN_supprimes.
    ↪loc[(donnees_NaN_supprimes["Sex"]=="F")]
valeurs_athletes_masculins=donnees_NaN_supprimes.
    ↪loc[(donnees_NaN_supprimes["Sex"]=="M")]
plt.figure()
#plt.figure(figsize=[8,6], dpi=300)
plt.subplot(111)
#sns.distplot(valeurs_athletes_masculins["Height"], label="Tailles M")
#sns.distplot(valeurs_athletes_feminines["Height"], label="Tailles F")
sns.histplot(valeurs_athletes_masculins["Height"], color="blue", label="Tailles_
    ↪M", kde=True)
sns.histplot(valeurs_athletes_feminines["Height"], color="red", label="Tailles_
    ↪F", kde=True)
plt.title("Distribution des tailles chez les athlètes des JO")
plt.xticks(rotation=90)
plt.legend()
plt.show()
```



3.8 3.9. Heatmap

```
[36]: import matplotlib.pyplot as plt
import seaborn as sns
donnees_heatmap=donnees_NaN_supprimes.
↳loc[(donnees_NaN_supprimes["Medal"]=="Gold") |
↳(donnees_NaN_supprimes["Medal"]=="Silver") |
↳(donnees_NaN_supprimes["Medal"]=="Bronze")]
donnees_heatmap=donnees_heatmap[["Sex","Age","Weight","Height","Medal","Year","Team"]]
dictionnaire = {"Bronze" : 1, "Silver" : 2, "Gold" : 3}
donnees_heatmap["Medal"] = donnees_heatmap["Medal"].map(dictionnaire)
donnees_heatmap["Team"]=donnees_heatmap["Team"].astype("category").cat.codes
correlation = donnees_heatmap.corr()
plt.figure()
plt.subplot(111)
sns.heatmap(correlation, annot=True, linewidths=1.0)
plt.show()
```



```
[ ]: donnees_NaN_supprimes.head()
```

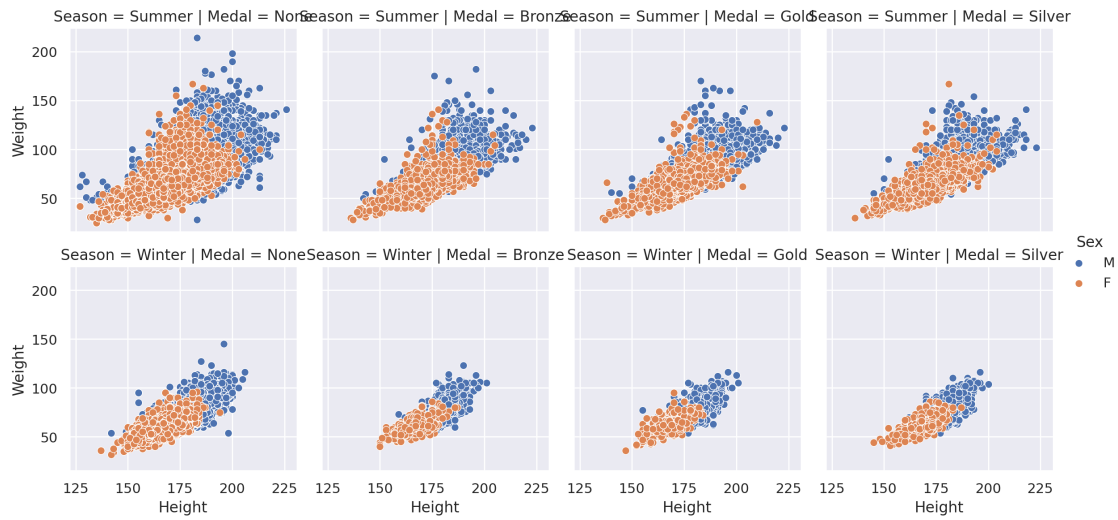
4 4. Graphiques multi-grilles

4.1 4.1. FacetGrid

4.2 Etude de la relation entre le poids et la taille des athlètes aux Jeux Olympiques d'été et d'hiver

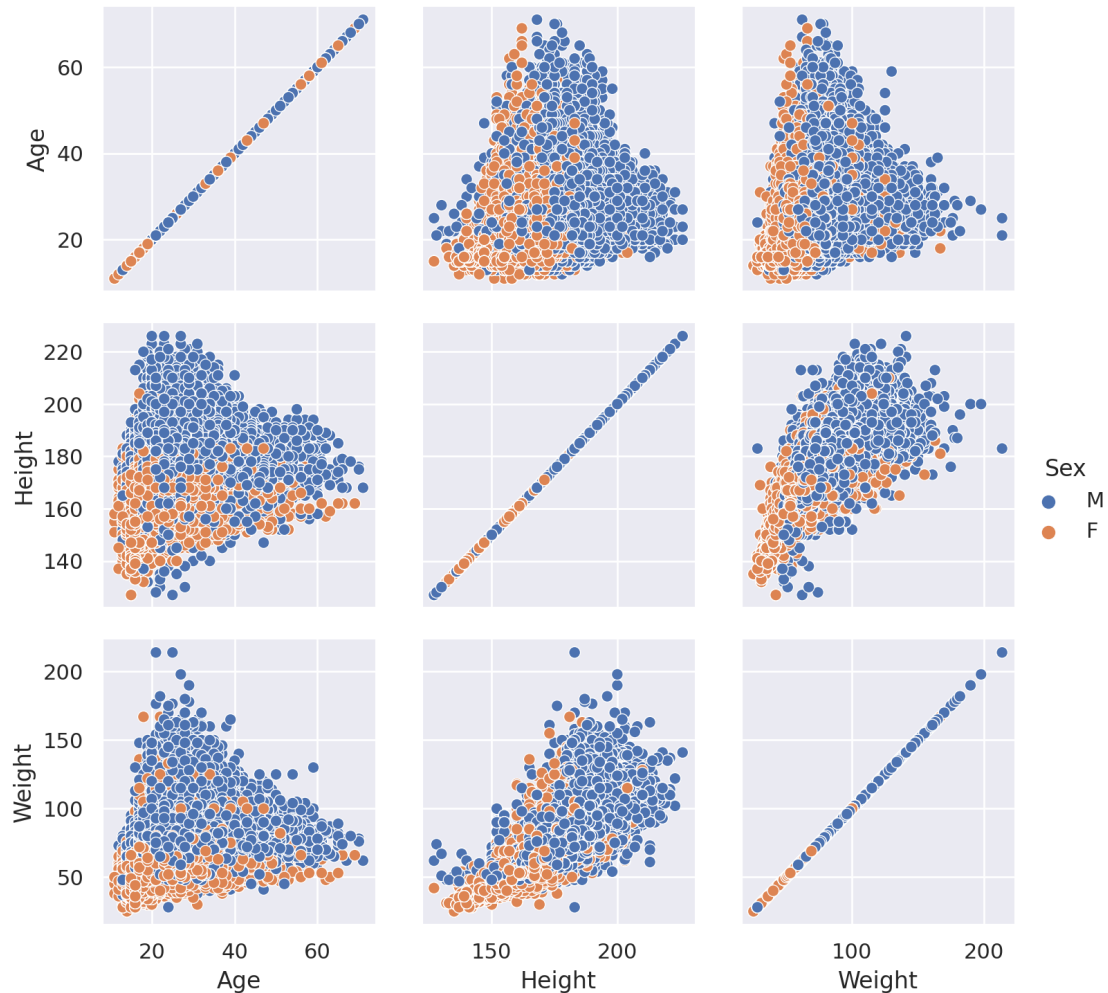
Nuages de points séparés selon le type de médaille obtenue par les athlètes et colorés selon le genre des athlètes

```
[37]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set(rc={'figure.dpi':200})
mon_facetgrid = sns.FacetGrid(donnees_NaN_supprimes, row = 'Season', col = 'Medal', hue = 'Sex')
mon_facetgrid.map(sns.scatterplot, 'Height', 'Weight')
mon_facetgrid.add_legend()
plt.show()
```



4.3 4.2. PairGrid

```
[38]: import matplotlib.pyplot as plt
import seaborn as sns
mon_pairgrid = sns.
    ↳ PairGrid(donnees_NaN_supprimes, vars=["Age", "Height", "Weight"], hue="Sex")
mon_pairgrid.map(sns.scatterplot)
mon_pairgrid.add_legend()
plt.show()
```

```
[57]: import matplotlib.pyplot as plt
import seaborn as sns
mon_pairgrid = sns.
    ↪PairGrid(donnees_NaN_supprimes, vars=["Age", "Height", "Weight"], hue="Sex")
mon_pairgrid.map_diag(sns.distplot)
mon_pairgrid.map_offdiag(sns.scatterplot)
mon_pairgrid.add_legend()
plt.show()
```

/home/hocquard/anaconda3/envs/sage/lib/python3.9/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

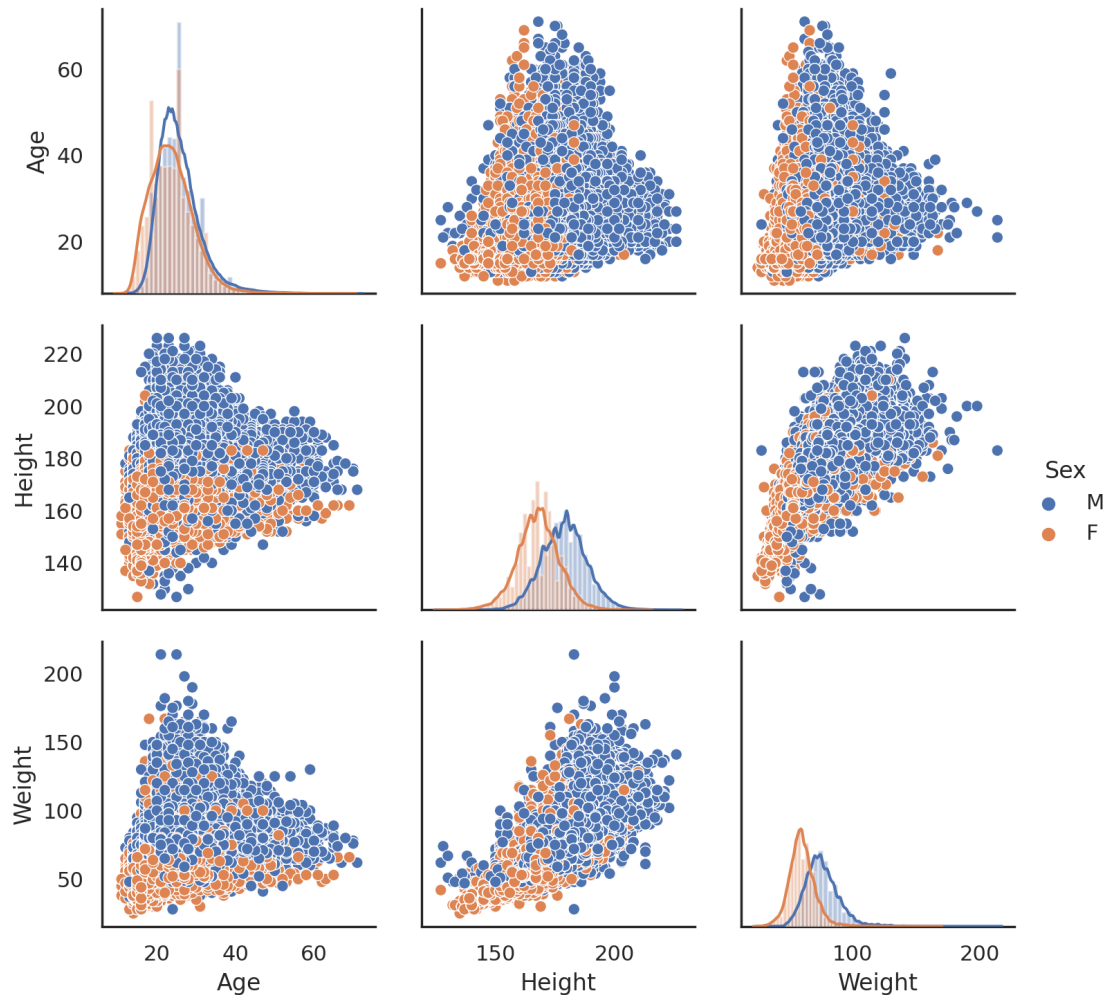
warnings.warn(msg, FutureWarning)

/home/hocquard/anaconda3/envs/sage/lib/python3.9/site-

```

packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar flexibility)
or `histplot` (an axes-level function for histograms).
    warnings.warn(msg, FutureWarning)
/home/hocquard/anaconda3/envs/sage/lib/python3.9/site-
packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar flexibility)
or `histplot` (an axes-level function for histograms).
    warnings.warn(msg, FutureWarning)
/home/hocquard/anaconda3/envs/sage/lib/python3.9/site-
packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar flexibility)
or `histplot` (an axes-level function for histograms).
    warnings.warn(msg, FutureWarning)
/home/hocquard/anaconda3/envs/sage/lib/python3.9/site-
packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar flexibility)
or `histplot` (an axes-level function for histograms).
    warnings.warn(msg, FutureWarning)
/home/hocquard/anaconda3/envs/sage/lib/python3.9/site-
packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar flexibility)
or `histplot` (an axes-level function for histograms).
    warnings.warn(msg, FutureWarning)

```



4.4 4.3. JointGrid

```
[54]: import matplotlib.pyplot as plt
import seaborn as sns
mon_JointGrid=sns.JointGrid(data=donnees_NaN_supprimes,x="Height",y="Weight")
mon_JointGrid.plot(sns.scatterplot, sns.distplot)
plt.show()
```

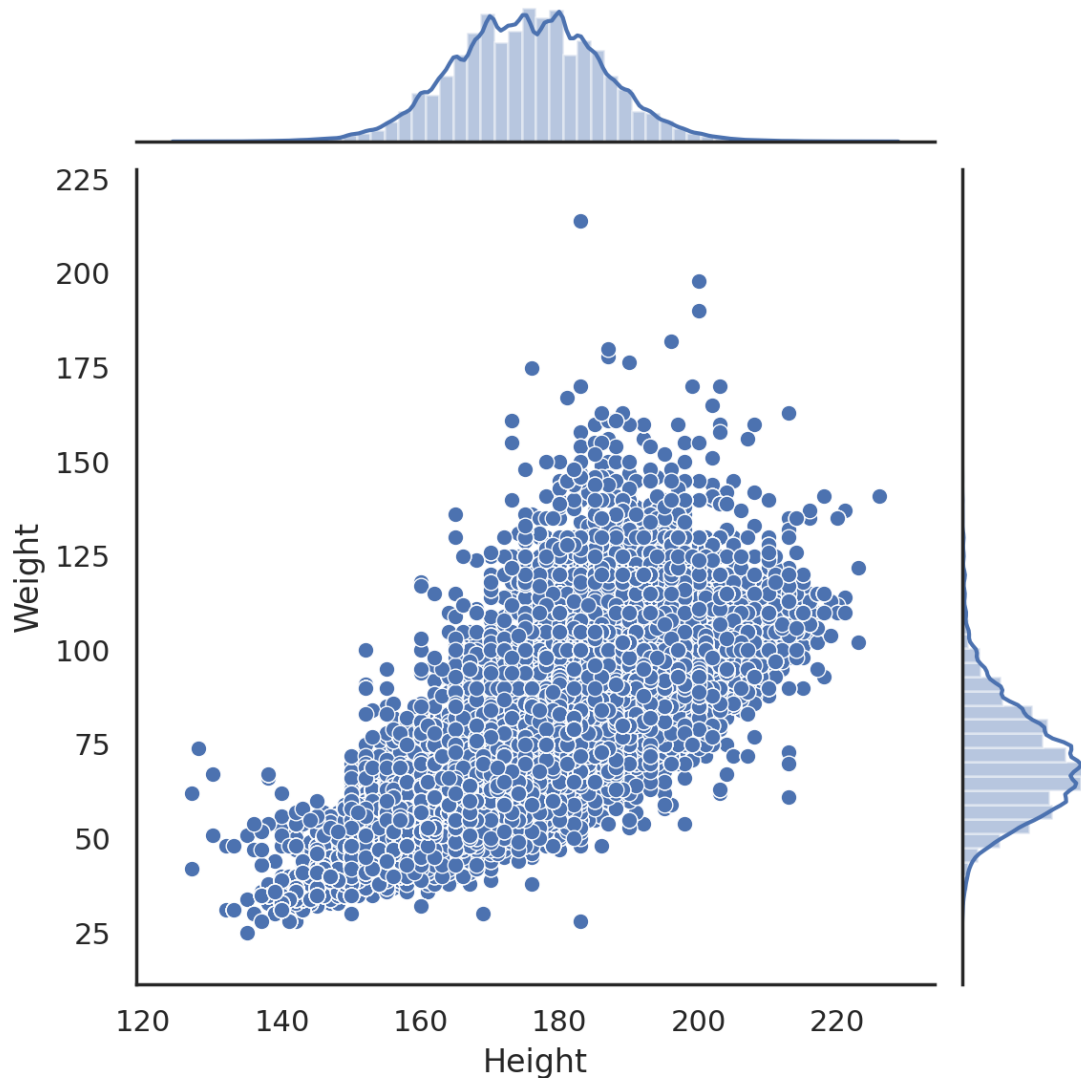
/home/hocquard/anaconda3/envs/sage/lib/python3.9/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)

/home/hocquard/anaconda3/envs/sage/lib/python3.9/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a

deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

```
warnings.warn(msg, FutureWarning)
/home/hocquard/anaconda3/envs/sage/lib/python3.9/site-
packages/seaborn/distributions.py:1689: FutureWarning: The `vertical` parameter
is deprecated and will be removed in a future version. Assign the data to the
`y` variable instead.
warnings.warn(msg, FutureWarning)
```



[]: