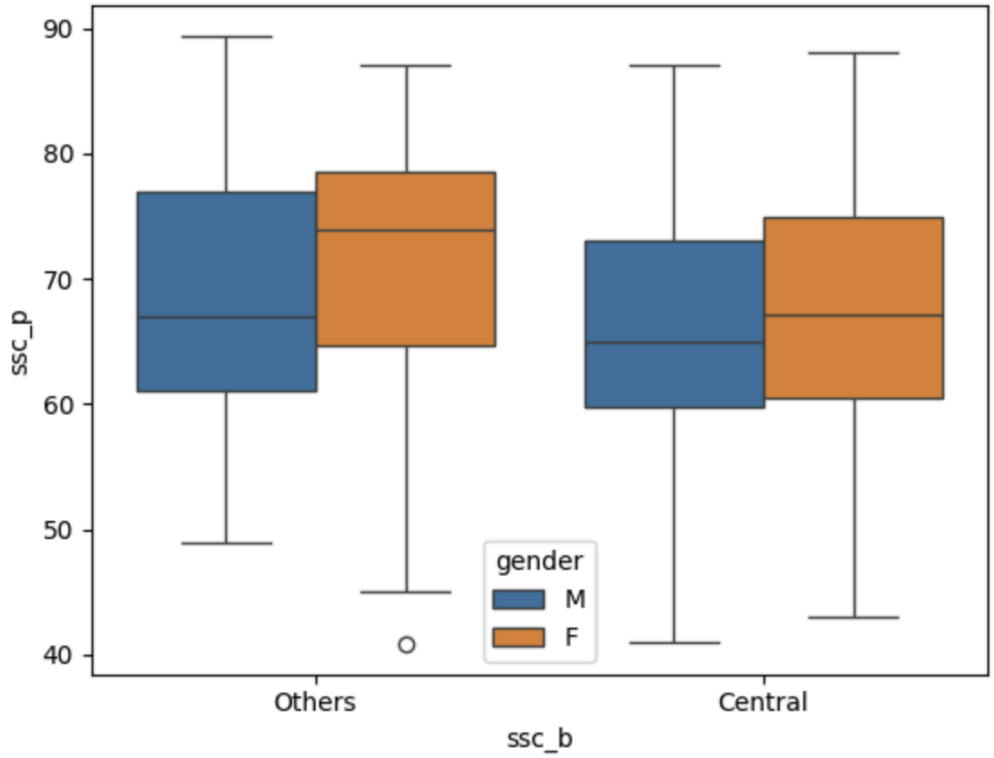
# Seaborn Assignment

## Box Plot:

```
sns.boxplot(x= 'ssc_b', y='ssc_p', data=dataset, hue='gender')
```

```
<Axes: xlabel='ssc_b', ylabel='ssc_p'>
```




**Summary:**

**Others Specialisation:**

1. Male:
    a. Males are getting 50 as the initial mark, 25% of Males are getting up to 60 marks, 50% of Males are getting up to 68 marks, 75% of Males are getting up to 77 marks and the maximum mark of the males is 89.
2. Female:
    a. Females are getting 46 as the initial mark, 25% of females are getting up to 65 marks, 50% of Females are getting up to 75 marks, 75% of females are getting up to 78 marks and the maximum mark of the Female in others is 87. We can see the outliers in Female marks i.e 40

**Central Specialisation:**

1. Male:

a. Males are getting 42 as the initial mark, 25% of Males are getting up to 60 marks, 50% of Males are getting up to 66 marks, 75% of Males are getting up to 73 marks and the maximum marks of the male is 87

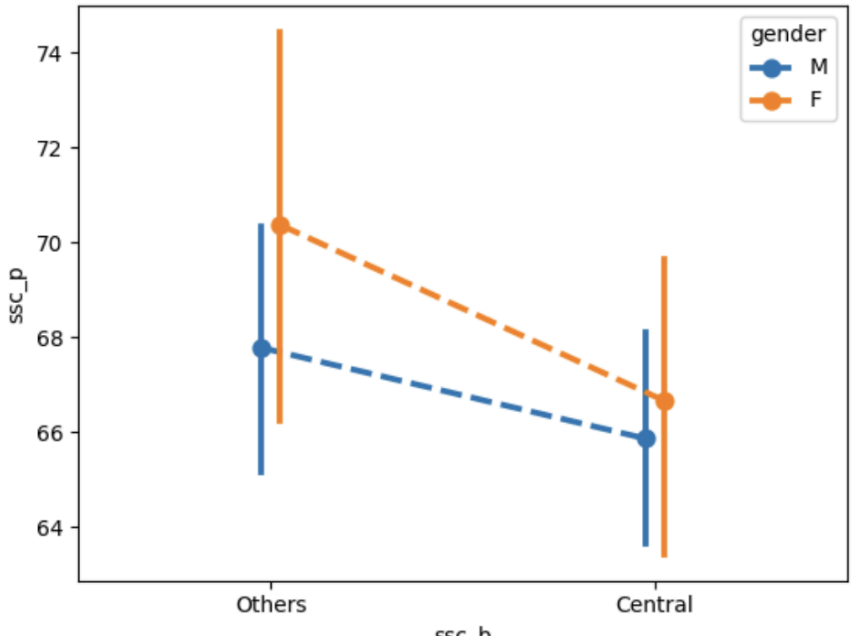
2. Female:
    a. Females are getting 44 as the initial mark, 25% of females are getting up to 62 marks, 50% of Females are getting up to 67 marks, 75% of females are getting up to 74 marks and the maximum mark of the Female in others is 88

Overall, Females in both specialisations got better average marks than the males in both specialisations. The maximum marks got by Males in Others specialisation and the minimum mark got by Female in Central specialisation.

As per gender point of view, Males under Other specialisation got better marks than Males under Central specialisation. Females under Other specialisation got better marks than Females under Central specialisation

**Point Plot:**

```
sns.pointplot(x= 'ssc_b', y='ssc_p', data=dataset, hue='gender', linestyles= 'dashed', dodge=True)
# Dodge - separate the points for each level of the `hue` variable along
# the categorical axis. Setting to `True` will apply a small default
plt.show()
```




A point plot uses scatter plot points to represent the central tendency of numeric data.These plots make use of error bars to indicate any uncertainty around the numeric variables.

If you want to visualize how these scores change over time in different categorical variables then use Point plots. It  can be more useful than bar plots for focusing comparisons between different levels of one or more categorical variables.

Imagine each time point on the x-axis as a dot, and the height of that dot represents the average score at that time. Then, you might see some 'error bars' above and below each dot. These bars show you how much the scores varied around the average at that time point. The taller the bars, the more uncertain the scores were at that time

**Summary**:
**Others specialisation:**

1. Male:
    a. The average ssc_p marks got by Males is approx 68 and as per the error bars, we could be some (low  is 65 and high is 70.5)  deviation from the average marks
2. Female:
    a. The average ssc_p marks got by Females is approx  71 and as per the error bars, we could be some (low  is 66 and high is 74.5)  deviation from the average marks

Overall, Females have got  high average marks than Males  and lot of uncertainty around the average marks

**Central Specialisation:**

1. Male:
    a. The average ssc_p marks got by Males is approx 66 and as per the error bars, we could be some (low  is 63.5 and high is 69)  deviation from the average marks
2. Female:
    a. The average ssc_p marks got by Females is approx  67 and as per the error bars, we could be some (low  is 63 and high is 70)  deviation from the average marks

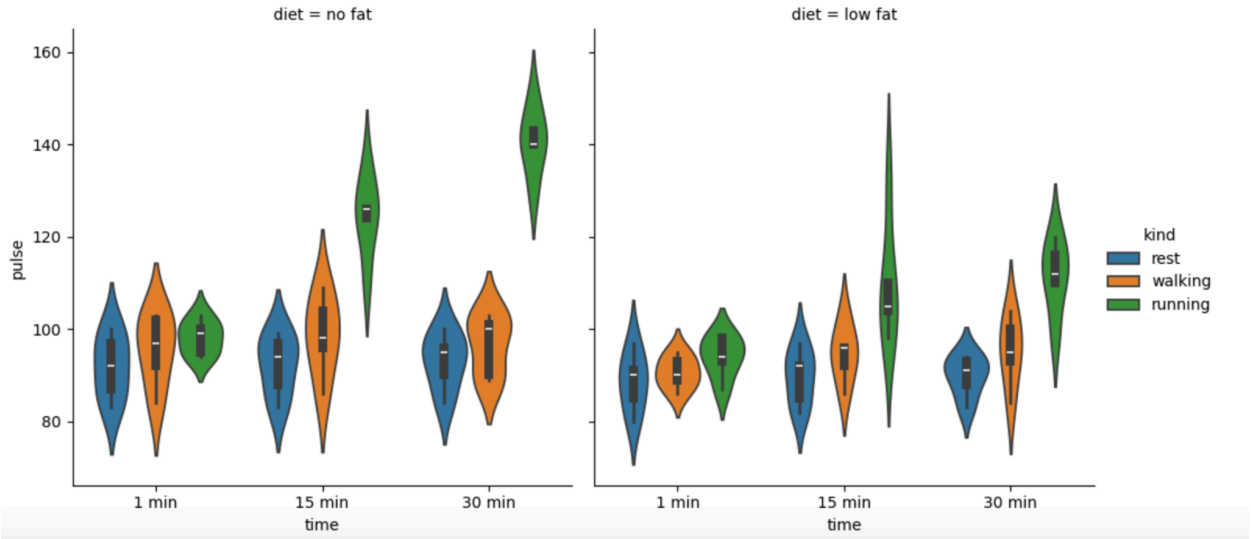
Overall, Females have got  high average marks than Males  and lot of uncertainty around the average marks

To conclude, Females from other specializations got higher average marks than others and also there is a  lot of uncertainty around the average marks. As per the specialisation, students from others specialisation got high average mark than central specialisation

## Factor or Cat plot:

A factor plot helps us understand how a numerical variable varies across different categories. It's useful for comparing groups or categories and identifying patterns or trends

```
#seaborn.factorplot was renamed to seaborn.catplot in seaborn 0.9 and has been marked as deprecated since then.
# It was definitely removed in seaborn 0.12 (see release notes).
# I am having seaborn version 0.13.2 so the factorplot has been removed so I used catplot
sns.catplot(x= 'time', y= 'pulse', hue= 'kind', kind= 'violin', col='diet', data=df)
plt.show()
```




**Summary:**

**No fat:**

1. **1 min:**
    a. Rest: pulse rate distribution  starts from 70 to 110 . As per the box plot, Initial value is 82, 25% of people are getting up to 85 pulse, 50% of people are getting up to 90 pulse, 75% of people are getting up to 97 pulse and the maximum pulse of the people is 110.
    b. Walking: pulse distribution starts from 70 to 117. As per the box plot, Initial value is 82, 25% of people are getting up to 90 pulse, 50% of people are getting up to 97 pulse, 75% of people are getting up to 102 pulse
    c. Running: pulse distribution starts from 90 to 105. As per the box plot, 25% of people are getting up to 90.5 pulse, 50% of people are getting up to 100 pulse, 75% of people are getting up to 102 pulse and the maximum pulse of the people is 103
2. **15 mins:**
    a. Rest: pulse rate distribution  starts from 70 to 108 . As per the box plot, Initial value is 82, 25% of people are getting up to 85 pulse, 50% of people are getting up to 94 pulse, 75% of people are getting up to 97 pulse and the maximum pulse of the people is 98.
    b. Walking: pulse distribution starts from 70 to 120. As per the box plot, Initial value is 83, 25% of people are getting up to 96 pulse, 50% of people are getting up to 97 pulse, 75% of people are getting up to 103 pulse and the maximum pulse of the people is 107
    c. Running: pulse distribution starts from 100 to 150. As per the box plot, 25% of people are getting up to 123 pulse, 50% of people are getting up to 125 pulse,

75% of people are getting up to 126 pulse and the maximum pulse of the people is 126

3. **30 mins:**
    a. Rest: pulse rate distribution  starts from 73 to 110 . As per the box plot, Initial value is 83 , 25% of people are getting up to 90 pulse, 50% of people are getting up to 97 pulse, 75% of people are getting up to 98 pulse and the maximum pulse of the people is 100.
    b. Walking: pulse distribution starts from 80 to 115. As per the box plot, 25% of people are getting up to 90 pulse, 50% of people are getting up to 100 pulse, 75% of people are getting up to 103 pulse and the maximum pulse of the people is 105
    c. Running: pulse distribution starts from 120 to 160. As per the box plot, 25% of people are getting up to 140 pulse, 50% of people are getting up to 142 pulse, 75% of people are getting up to 145 pulse

**Low fat:**

1. **1 min:**
    a. Rest: pulse rate distribution  starts from 70 to 110 . As per the box plot, Initial value is 80 , 25% of people are getting up to 83 pulse, 50% of people are getting up to 90 pulse, 75% of people are getting up to 93 pulse and the maximum pulse of the people is 97.
    b. Walking: pulse distribution starts from 83 to 100. As per the box plot, Initial value is 85, 25% of people are getting up to 87 pulse, 50% of people are getting up to 88 pulse, 75% of people are getting up to 90 pulse and the maximum pulse of the people is 91
    c. Running: pulse distribution starts from 80 to 105. As per the box plot, Initial value is 85, 25% of people are getting up to 90 pulse, 50% of people are getting up to 92 pulse, 75% of people are getting up to 98 pulse

2. **15 mins:**
    a. Rest: pulse rate distribution  starts from 75 to 105 . As per the box plot, Initial value is 80 , 25% of people are getting up to 82 pulse, 50% of people are getting up to 90 pulse, 75% of people are getting up to 92 pulse and the maximum pulse of the people is 97.
    b. Walking: pulse distribution starts from 77 to 117. As per the box plot, Initial value is 85, 25% of people are getting up to 90 pulse, 50% of people are getting up to 97 pulse, 75% of people are getting up to 98 pulse
    c. Running: pulse distribution starts from 80 to 150. As per the box plot, Initial value is 97, 25% of people are getting up to 98 pulse, 50% of people are getting up to 102 pulse, 75% of people are getting up to 115 pulse

3. **30 mins:**
    a. Rest: pulse rate distribution  starts from 77 to 98 . As per the box plot, Initial value is 83 , 25% of people are getting up to 85 pulse, 50% of people are getting up to 90 pulse, 75% of people are getting up to 93 pulse
    b. Walking: pulse distribution starts from 75 to 118. As per the box plot, Initial value is 84, 25% of people are getting up to 90 pulse, 50% of people are getting up to 93 pulse, 75% of people are getting up to 103 pulse  and the maximum pulse of the people is 105.
    c. Running: pulse distribution starts from 90 to 130. As per the box plot,25% of people are getting up to 115 pulse, 50% of people are getting up to 117 pulse, 75% of people are getting up to 119 pulse  and the maximum pulse of the people is 121.

To conclude, As per the diet, people who are following no fat diet, have high pulse rate while at running kind based on the increase in the time limits whereas the people who are following low fat diet have high pulse rate at running kind only if they reach 15 mins.

As per the different kind, people who are following no fat diet, have minimal changes in rest kind , walking kind and high changes in running as the time increases whereas people who are following low fat diet, have minimal pulse rate while at rest, max pulse rate while at walking and uncertainty pulse rate while at running as the time increases

As per the box plot, There is a minimal difference in pulse rate for most of the people while at rest irrespective of diets. Most of the people who are following a no fat diet, have higher pulse rate while at walking and running than a low fat diet group

Overall, people who follow low fat diets, have better pulse rate in different kinds as compared to people who are following no fat diet. This shows we need to include fat in our diets to reduce the spike in pulse rate when we increase the time of different physical activities