

1. Swarm Plot :

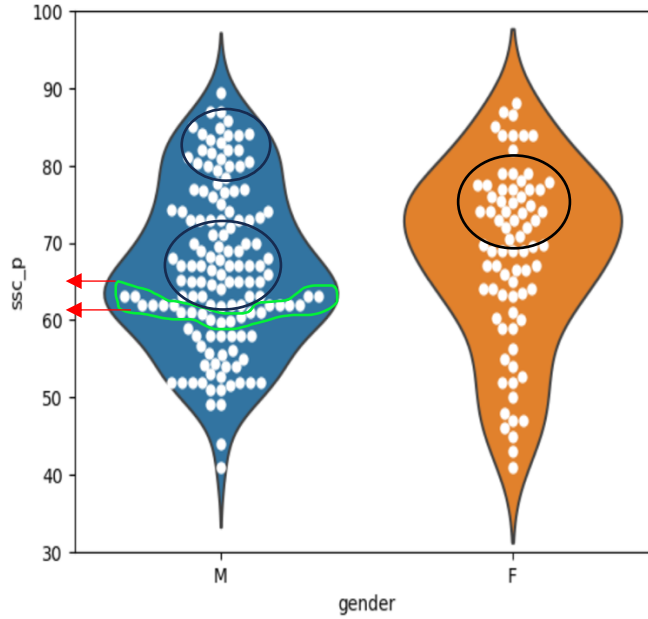
A swarm plot, also known as a bee swarm plot, is a type of data visualization that is used to display the distribution of categorical data points along a continuous axis. It is particularly useful for visualizing the distribution of data when we have a relatively small number of data points and want to avoid overlapping points that can occur in a traditional scatter plot.

Here's how to interpret a swarm plot:

1. **Vertical Axis:** The vertical axis typically represents the categorical variable we are interested in analyzing or comparing. Each unique category in our data is represented as a tick mark or label on this axis.
2. **Horizontal Axis:** The horizontal axis represents the continuous variable we want to examine within each category. It could be any numerical value, such as a measurement or a score.
3. **Data Points:** In a swarm plot, individual data points are plotted along the continuous axis for each category. Each data point is represented as a point or dot. The position along the continuous axis corresponds to the value of the data point.
4. **Distribution:** A swarm plot provides a visual representation of the distribution of data points within each category. We can observe how data points are distributed along the continuous axis. Clusters of dots indicate where there may be a concentration of data points, and the spread of dots shows the overall distribution.
5. **Density:** Swarm plots are helpful for visualizing the density of data points within each category. Denser areas have more data points clustered together, while sparser areas have fewer points.
6. **Outliers:** Outliers, if present, can be identified in a swarm plot. These are data points that are far away from the main cluster of points within a category.
7. **Comparison:** If we have multiple categories, we can easily compare their distributions by looking at how the points are spread along the continuous axis for each category. This allows us to make visual comparisons between groups.
8. **Insights:** Swarm plots are useful for spotting patterns, trends, or differences in our data. For example, we might notice that one category has a wider spread of data points compared to another, or that certain categories have more outliers.
9. **Limitations:** Keep in mind that swarm plots are most effective with relatively small datasets. With larger datasets, the density of points can become overwhelming, and it may be more appropriate to use other types of plots, such as box plots or violin plots, to visualize the distribution.

In summary, swarm plots are a valuable tool for visualizing the distribution of categorical data points along a continuous axis, helping us to gain insights into the characteristics of our data and make comparisons between categories.

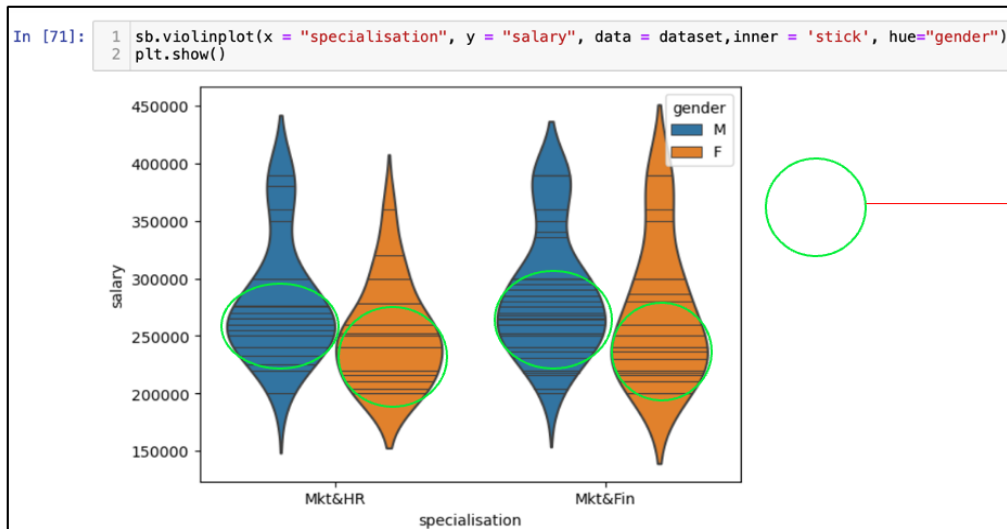
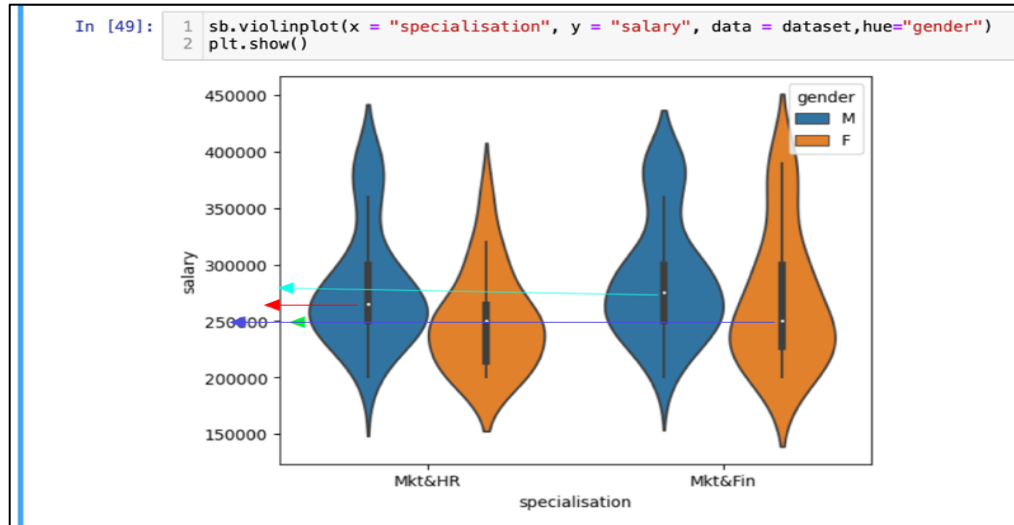
```
In [67]: 1 ax=sb.violinplot(x = "gender", y = "ssc_p",data = dataset,width=0.8,inner= None)
2 ax.set_ylim(30, 100)
3 sb.swarmplot(x = "gender", y = "ssc_p", data = dataset,color = 'white',size = 6);
4 #plt.show()
```



Plot Inference :

From the plot it is obvious that male count is $>$ than the female count. All the circled regions shows the clusters (aggregation of data) implying cluster of genders securing the same marks within that specified range. The extended oblate shape which is in green color indicated that more males have secured within \pm average score whereas the count of females securing the same mark is very less when compared to males.

2. Violin Plot Inference(Specialization vs Salary):

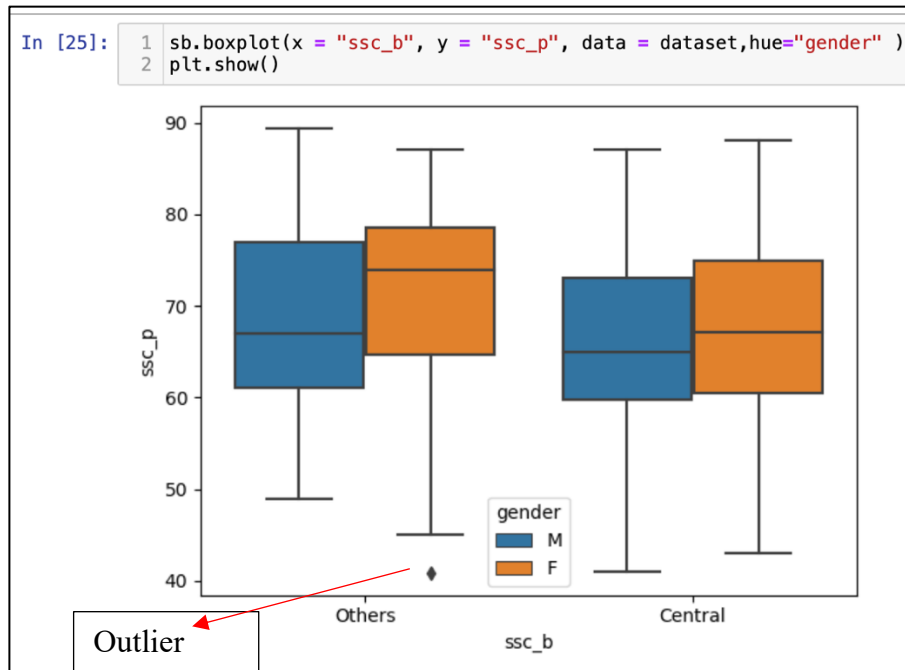


Regions of
data
condensation

Plot Inference :

This plot deals with the relationship of salary w.r.t specialisation for male & female cadre. All the bulged region represents aggregation of data in that region which can be visualised in the second picture, where the lines become more crowded. Median (represented by white dot inside the box plot) salary of males under specialisation Mkt & Fin is higher when compared to rest of them. Females falling under both the specialisations has the same median salary. On the other hand, females under Mkt & Fin gets both higher & lower salary when compared to rest of them. Besides female distribution under both specialisation exhibits well pronounced + skewness indicating the presence of very few number of female students receive higher salary. On average males under both specialisation & female under specialisation Mkt & Fin receive decent salary.

3. Box Plot Inference:



Plot Inference :

This plot deals with the relationship of ssc_p w.r.t ssc_b for male & female cadre. Median (represented by white dot inside the box plot) pass mark of females under board others is higher when compared to rest of them. Highest pass mark is secured by males under board others when compared to rest of them. Lowest pass mark is secured by males under board central when compared to rest of them. On average both male & female under the board others secured decent ssc pass mark with an exceptional outlier (shown by black coloured diamond shape), when compared to the genders under the board central. Besides, Females under both the boards exhibit negative skewness (by seeing the left whisker size & lower quartile Q1 being large), with female under others board being well pronounced which indicates the presence of few number of low pass mark scorers. Whereas, males under both the boards exhibit slight positive skewness in the distribution.

Skewness Proof:

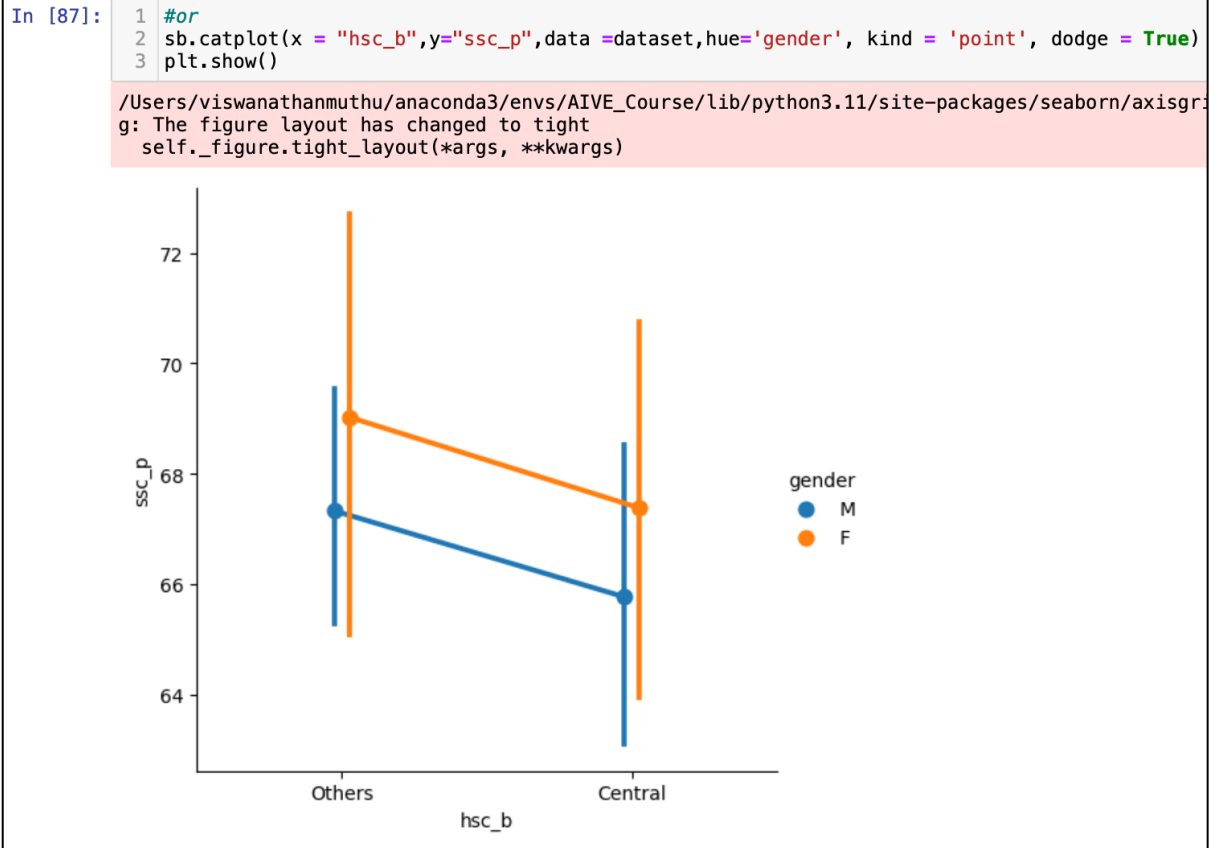
```
In [70]: 1 import scipy.stats as stats
         2 groups = dataset.groupby(['ssc_b', 'gender'])
         3 skewness_results = groups['ssc_p'].apply(stats.skew)
         4 skewness_results
```

Out[70]:

ssc_b	gender	skewness
Central	F	-0.380987
	M	0.105774
Others	F	-0.847248
	M	0.132999

Name: ssc_p, dtype: float64

4. Point Plot Inference:



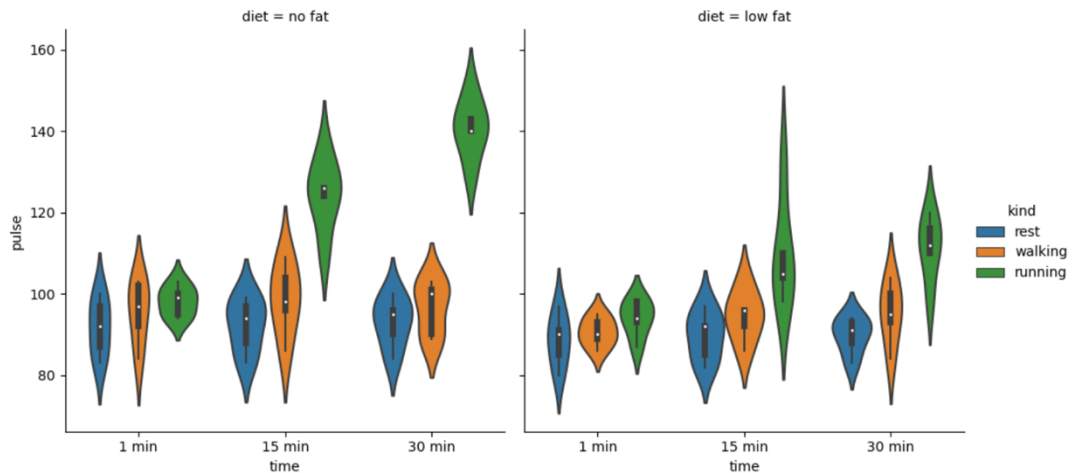
Plot Inference:

Point plots allow us to compare the central tendency of the ssc_p across different categories like hsc_b for genders. The central tendency by default ssc_p mean of others board for females is the highest of all. In both the boards females are better performers, where the females under others board secured the highest mark among all. Males under central board secured the least score as starting score when compared to all. The direction of the points (upward or downward) indicates whether the central tendency of the ssc_p tends to increase or across the others board both for female & male category.

5. Factor Plot Inference:

```
In [33]: 1 import pandas as pd
2 import seaborn as sb
3 from matplotlib import pyplot as plt
4 df = sb.load_dataset('exercise')
5 sb.catplot(x = "time", y = "pulse", hue = "kind", kind = 'violin', col = "diet", data = df);
6 plt.show()
```

/Users/viswanathanmuthu/anaconda3/envs/AIVE_Course/lib/python3.11/site-packages/seaborn/axisgrid.py:117: UserWarning: The figure layout has changed to tight
self._figure.tight_layout(*args, **kwargs)



Plot Inference:

This grouped violin plot provides a visual summary of how pulse rates are distributed across different exercise durations, types, and diets. It helps us to identify the patterns and variations in pulse rates and their relationships with these categorical variables.

For detailed inference refer table.

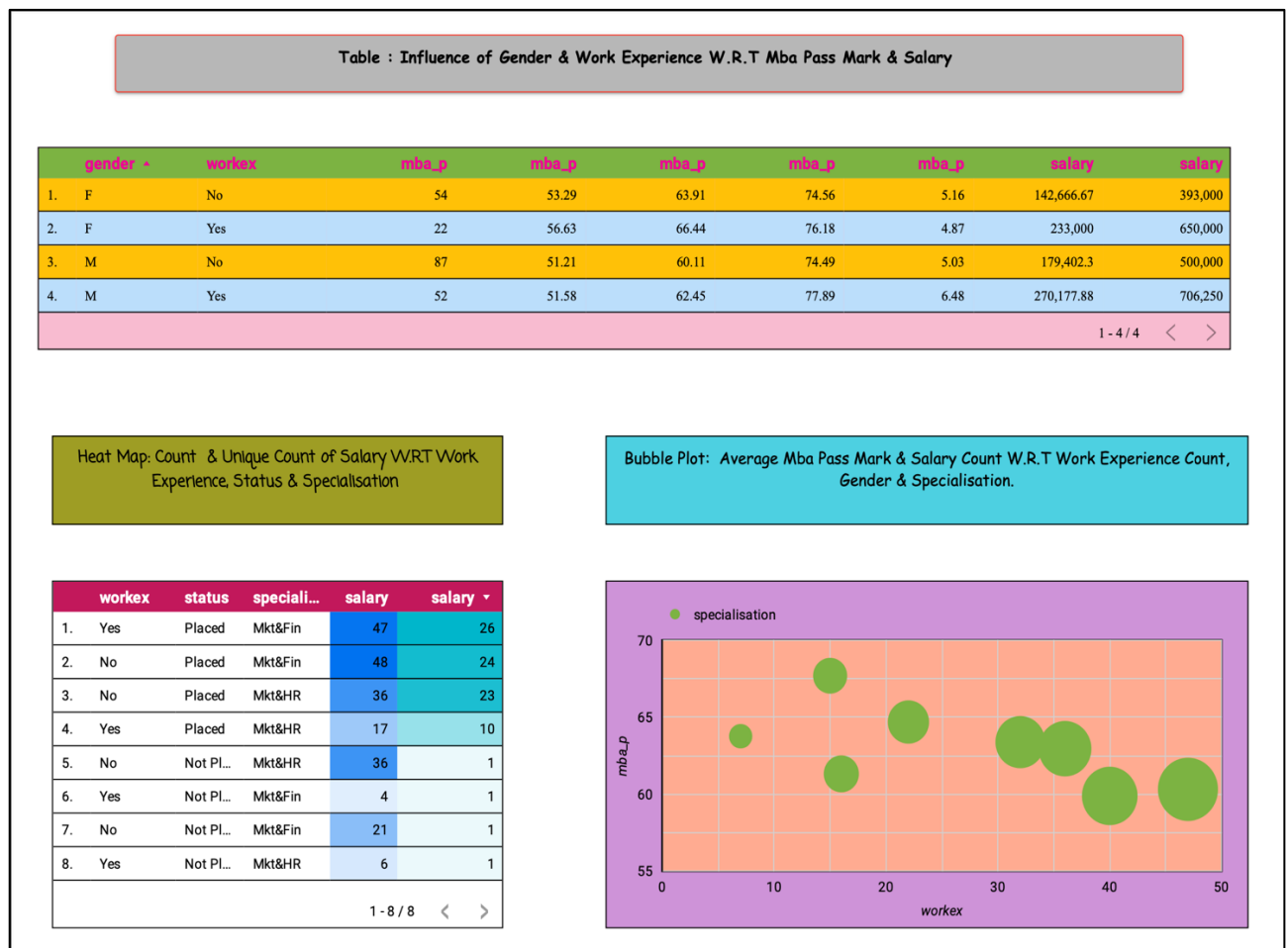
Diet = No fat				Diet = Low fat	
Duration	Kind	Pulse Concentration (Approximately)	Inference	Pulse Concentration (Approximately)	Inference
1 min	Rest	80-95	Starting pulse rate < 80 Ending pulse rate = 110	80-95	Starting pulse rate < 80 Ending pulse rate = 110
	Walking	90-100	Starting pulse rate < 80 Ending pulse rate = 112	90-95	Starting pulse rate > 80 Ending pulse rate = 98
	Running	95-100	Starting pulse rate = 90 Ending pulse rate = 106	95-100	Starting pulse rate > 90 Ending pulse rate = 108
15 min	Rest	90-100	Starting pulse rate < 80 Ending pulse rate = 110	90-95	Starting pulse rate < 80 Ending pulse rate = 110
	Walking	100-110	Starting pulse rate < 80 Ending pulse rate = 120	90-98	Starting pulse rate < 80 Ending pulse rate = 110
	Running	120-130	Starting pulse rate = 100 Ending pulse rate = 142	95-110	Starting pulse rate = 80 Ending pulse rate = 150
30 min	Rest	90-100	Starting pulse rate < 80 Ending pulse rate = 110	90-95	Starting pulse rate < 80 Ending pulse rate = 100
	Walking	90-95 & 100-105	Starting pulse rate > 80 Ending pulse rate = 112	90-100	Starting pulse rate < 80 Ending pulse rate = 110
	Running	135 - 145	Starting pulse rate = 120 Ending pulse rate = 160	103-120	Starting pulse rate = 90 Ending pulse rate = 123
Take Away:	Better pulse rate for all the kinds of exercise with persons running exhibiting exceptionally high range of pulse rate w.r.t duration 15 mints & 30 mints. Diet has greater influence on pulse rate w.r.t time.			Pulse rate range varies slightly for all the kinds of exercise w.r.t all the durations. Diet has only smaller influence on pulse rate w.r.t time.	

6. Creating Dashboard For Placement Dataset (Using Data Studio Tool):

Questions:

1. Find the influence of Gender & Work Experience W.R.T Mba Pass Mark & Salary using a Table.
2. Get the Count & Unique Count of Salary W.R.T Work Experience, Status & Specialisation using Heat Map.
3. Do the analysis for Average Mba Pass Mark & Salary Count W.R.T Work Experience Count, Gender & Specialisation using Bubble Plot.

- Google Studio Dashboard Link : <https://lookerstudio.google.com/s/jEV5goLQxjo>



-----End Of Document-----

முத்து வசுமதி