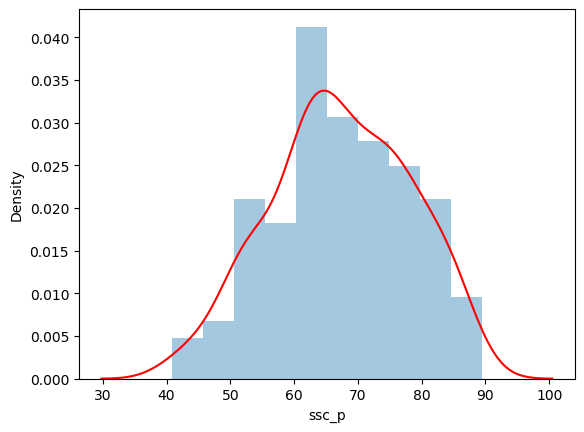
**Univarite Analysis**

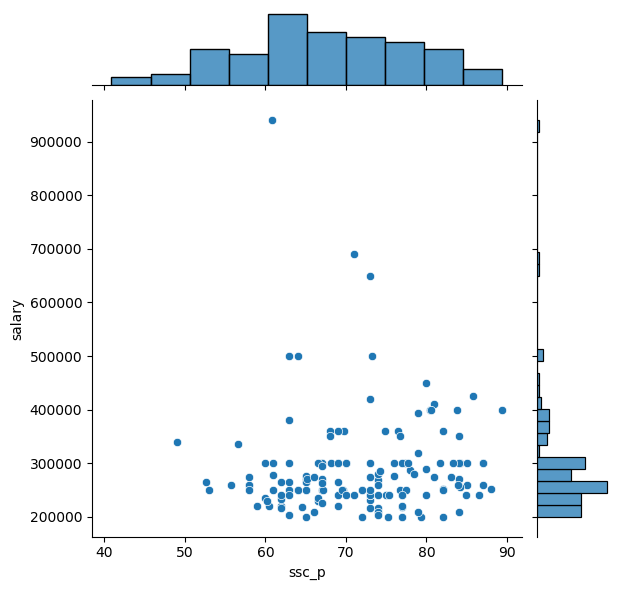


The diagram shows the distribution of ssc\_p (likely student scores):

* **Blue Histogram**: Represents the frequency of scores. Most students scored between 60-70%, shown by the tallest bar.
* **Red KDE Curve**: Smoothly estimates the probability of scores. The peak at 60-65% confirms that this range is the most common. The curve tapers off, indicating fewer students with very low or very high scores.

Together, the histogram and KDE show a normal-like distribution, with most scores around the middle (60-70%).

**Bivariate Analysis : Distribution Plot**

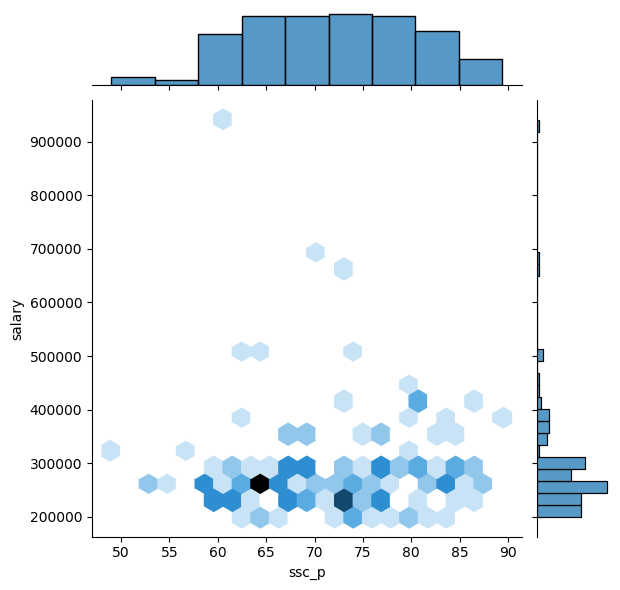


This diagram is a **scatter plot** of ssc\_p (secondary school percentage) versus salary, with **marginal histograms** for both variables:

* **Scatter Plot**: It shows the relationship between ssc\_p and salary. Most points are clustered between 200,000 and 400,000 in salary, regardless of ssc\_p. There are some outliers with very high salaries, especially around 60-80% ssc\_p, but overall, there is no strong visible correlation between higher scores and higher salaries.
* **Top Histogram (ssc\_p)**: Most students have scores between 60-70%.
* **Right Histogram (Salary)**: Most salaries fall between 200,000 and 300,000.

In summary, the plot shows a scattered relationship between ssc\_p and salary, with most data points concentrated in mid-ranges for both.

**Bivariate Analysis : Distribution Plot; Kind=hue**

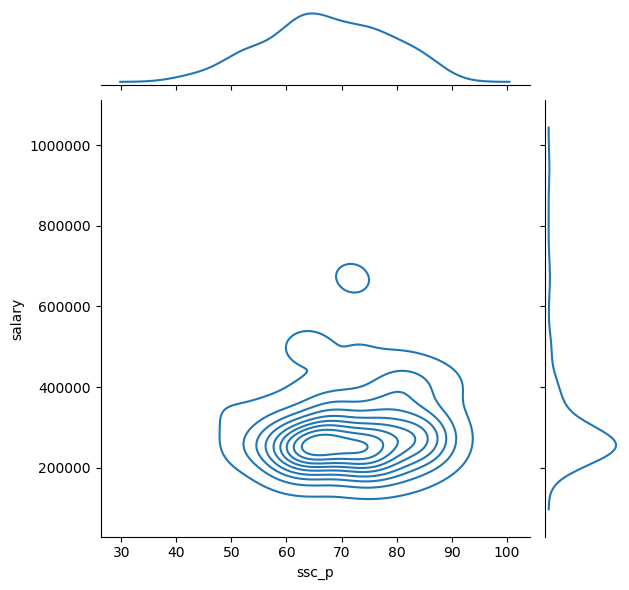


This diagram is a hexbin plot of ssc\_p (secondary school percentage) versus salary, with marginal histograms. It shows the density of data points by grouping them into hexagonal bins:

* Hexbin Plot: The color intensity of each hexagon represents the concentration of data points in that area. Darker hexagons indicate more students within that particular ssc\_p and salary range.
  + The darkest hexagons are found around ssc\_p scores between 60-70% and salaries around 200,000-300,000. This indicates that the majority of students in this dataset have scores in this range and receive salaries within this bracket.
  + A few lighter hexagons are scattered at higher salary ranges (500,000-900,000), suggesting some outliers, but these are much less frequent.
* Top Histogram (ssc\_p): Most students have ssc\_p scores in the range of 60-70%.
* Right Histogram (Salary): The majority of salaries fall between 200,000 and 300,000.

The hexbin plot highlights that most students have ssc\_p scores around 60-70% and receive salaries between 200,000-300,000. There are some outliers with higher salaries, but they are much less frequent.

**Bivariate Analysis : Distribution Plot; Kind=Kde**

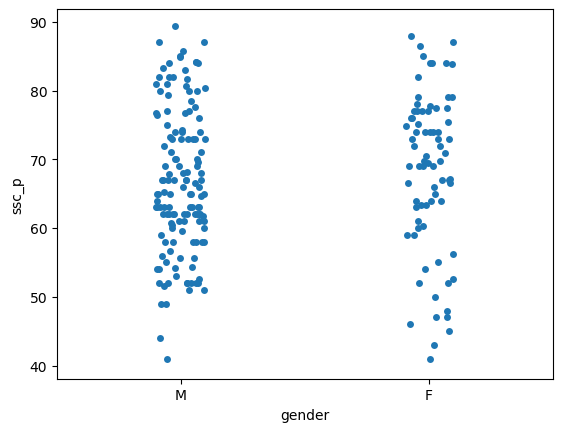


The plot shows the relationship between **secondary school percentage (ssc\_p)** and **salary**:

* **X-axis (ssc\_p)**: Most values range between **60-70%**.
* **Y-axis (salary)**: Most salaries are between **200,000 and 400,000**.
* **Contours**: These show data density, indicating a concentration of data points in the **60-70% ssc\_p range** with salaries in the **200,000 to 400,000** range.
* **Marginal KDE plots**: The top plot shows the distribution of ssc\_p, with a peak around **65-70%**, and the right plot shows the distribution of salaries, peaking around **200,000 to 400,000**.
* There is also a small, isolated cluster representing higher salaries, up to **1,000,000**, likely indicating a few outliers.

This visual clarifies that **ssc\_p between 60-70%** is associated with most salaries in the dataset, primarily around **200,000 to 400,000**.

**For Categorical Plot**



The plot is a strip plot (scatter plot) showing the distribution of secondary school percentage (ssc\_p) based on gender (M/F).

Explanation:

* X-axis (gender): Two categories, M for male and F for female.
* Y-axis (ssc\_p): The percentage scores from secondary school.

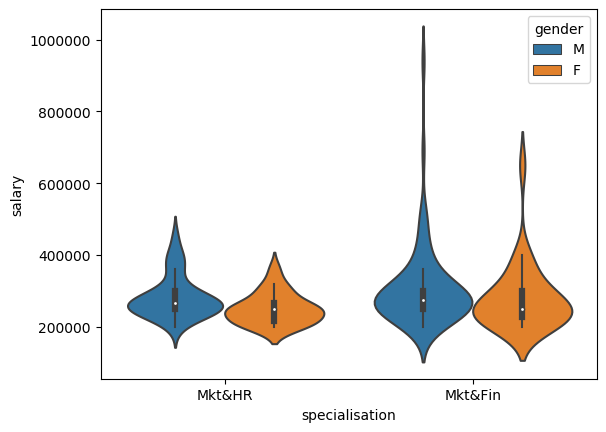
Observations:

1. Distribution of ssc\_p for males (M):
   * The scores range from 40% to 90%.
   * There is a concentration around 60-80%, meaning most males have ssc\_p in this range.
2. Distribution of ssc\_p for females (F):
   * Similarly, the scores for females also range from 40% to 90%.
   * There is a cluster around 60-80%, similar to the male distribution, but slightly more spread out across this range.

Overall Insight:

* Both genders have similar secondary school percentages, primarily between 60-80%, with a few outliers on the lower (40-50%) and higher (80-90%) ends.
* There doesn't seem to be a significant difference in the overall distribution of ssc\_p between males and females.

**Distribution Observation-Violin Plot**

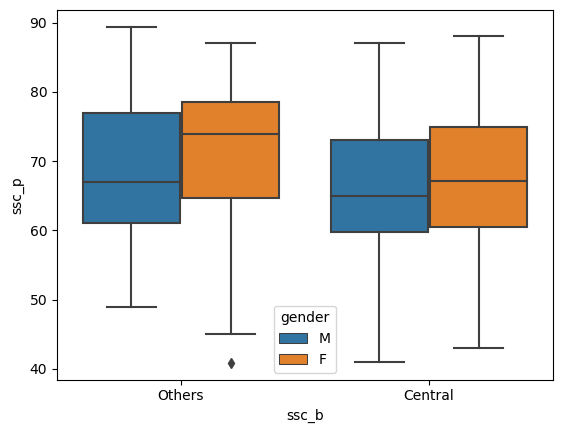


The plot shows salary distribution by specialization and gender:

* **Mkt&HR**:
  + Both males and females have similar salaries, mostly ranging between **200,000 and 300,000**.
* **Mkt&Fin**:
  + Males have a broader salary range, extending up to **1,000,000**, with most between **300,000 and 500,000**.
  + Females have a narrower range, mostly between **300,000 and 500,000**, with fewer outliers.

In summary, **Mkt&Fin** has higher and more varied salaries, particularly for males, while **Mkt&HR** has more clustered and lower salaries for both genders.

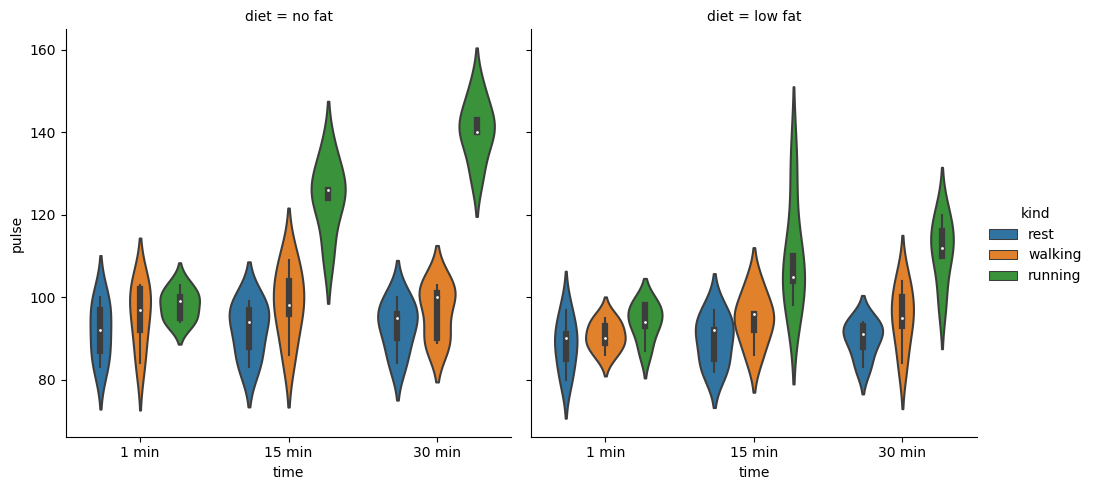
**Box Plot**



This boxplot shows the distribution of "ssc\_p" (possibly a score or percentage) based on two categories of "ssc\_b" (Others and Central) and separated by gender (M for Male, F for Female). Here's the breakdown:

* **Others**:
  + Males: The median (middle line in the box) is slightly lower than females. The spread (interquartile range, or IQR) is wide, indicating variability in the data.
  + Females: Median is slightly higher, with a more compact IQR, indicating more consistency in scores.
* **Central**:
  + Males: Median is similar to females, with a similar IQR.
  + Females: Median is close to males, but the distribution seems slightly more spread out.

There is one outlier in the "Others" group for males. Overall, the scores are fairly similar across gender and "ssc\_b" categories.

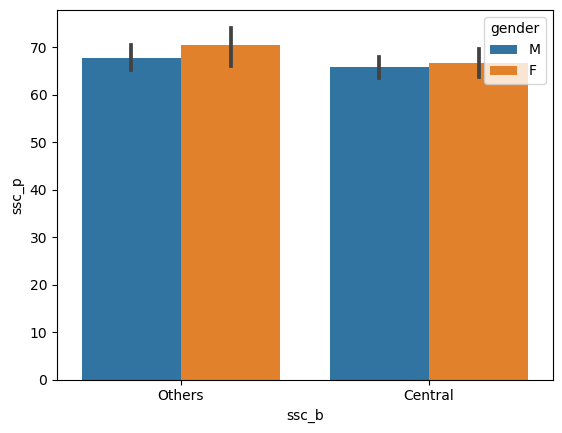


This is a violin plot showing the distribution of pulse rates over time (1 min, 15 min, 30 min) for two different diets (no fat and low fat). The activities are categorized into three types: rest, walking, and running.

* **No Fat Diet (left plot)**:
  + **Rest** (blue): Pulse remains fairly stable with low variation.
  + **Walking** (orange): Slight increase in pulse over time.
  + **Running** (green): Significant rise in pulse, especially at 15 and 30 minutes, with wider distributions, indicating more variation in heart rates.
* **Low Fat Diet (right plot)**:
  + **Rest** (blue): Pulse remains stable and similar to the no-fat diet.
  + **Walking** (orange): Similar to the no-fat group, with a moderate pulse.
  + **Running** (green): A noticeable increase in pulse, with peaks at 15 and 30 minutes, though the rise at 30 minutes is slightly more intense than in the no-fat diet.

The width of each shape represents the density of data points, showing how pulse values are distributed across time and activities.

**Bar Plot**



This bar plot shows the average "ssc\_p" (possibly scores) based on "ssc\_b" (Others and Central) for males (blue) and females (orange).

* **Others**: Females have slightly higher scores than males.
* **Central**: Males and females have nearly identical scores.

The bars include small error lines showing variation in the data.