Sports Data Analysis – Decathlon event (Vizualisation on R)

Introduction

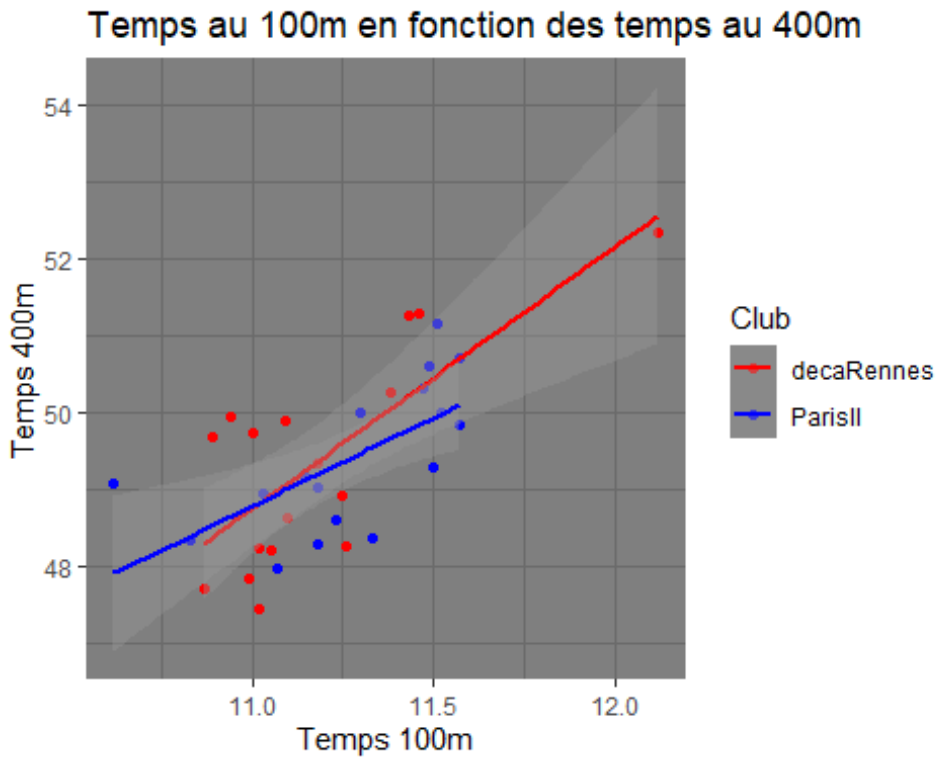
This report aims to analyze the performances of athletes from Rennes and Paris across several decathlon events. Various statistical visualizations have been conducted to identify trends and differences between the clubs.

1. Analysis of 100m and 400m Times

Visualization : Scatter Plot with Linear Regression

```
ggplot(
  DECATHLON2, aes(x=C100, y=C400, color=Club, group=Club)) +
  geom_point() +
  labs(title="Temps au 100m en fonction des temps au 400m",
       x="Temps 100m",
       y="Temps 400m") +
  theme_dark() +
  scale_color_manual(values=c("red","blue")) +
  geom_smooth(method="lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Observations

- There is a positive correlation between 100m and 400m performances: athletes who are fast in the 100m tend to be fast in the 400m as well.
- An extreme value disrupts the overall trend, suggesting an abnormal performance in the 400m for one athlete.
- In general, **Rennes athletes perform better than Parisian athletes** in these distances.

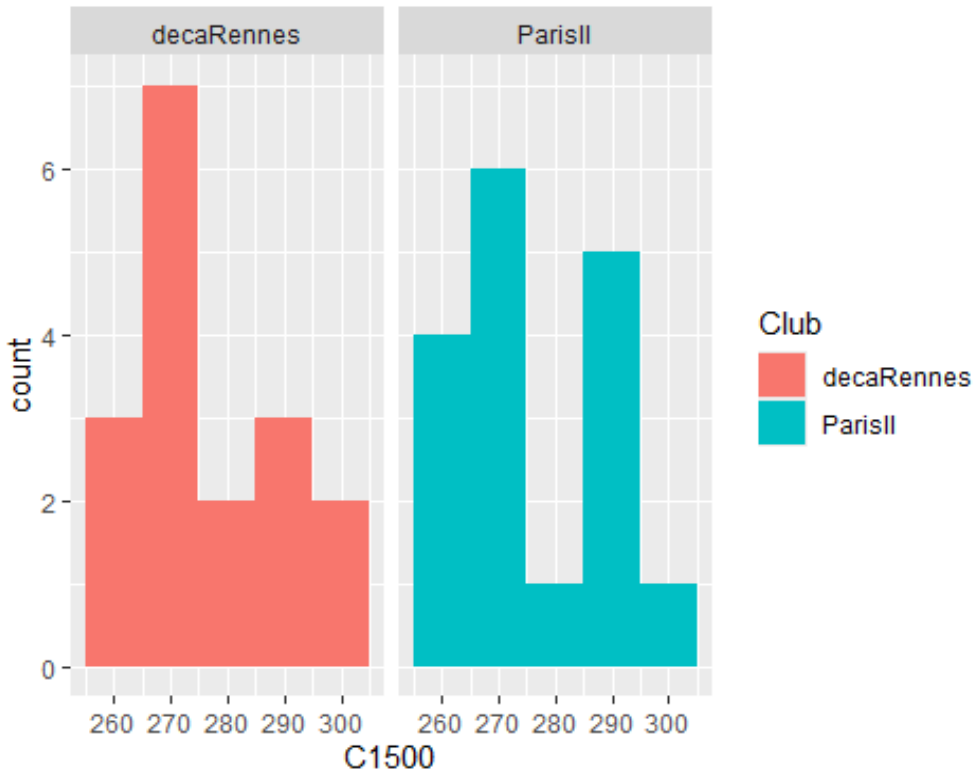
Interpretation

- Rennes' superiority in sprinting could be attributed to a more effective training program for speed events.
- The anomaly in 400m performance could be due to an injury, fatigue, or an athlete's poor condition during the competition.

2. Analysis of 1500m Times

Visualization: Histogram with Club Facets

```
ggplot(DECATHLON2, aes(x=C1500, fill=Club)) +
  geom_histogram(binwidth = 10) +
  facet_wrap(~Club)
```



Observations

- The top 10 Rennes and Paris athletes have similar performances, all finishing under 275 seconds.
- There is no significant gap between the two clubs in this event.

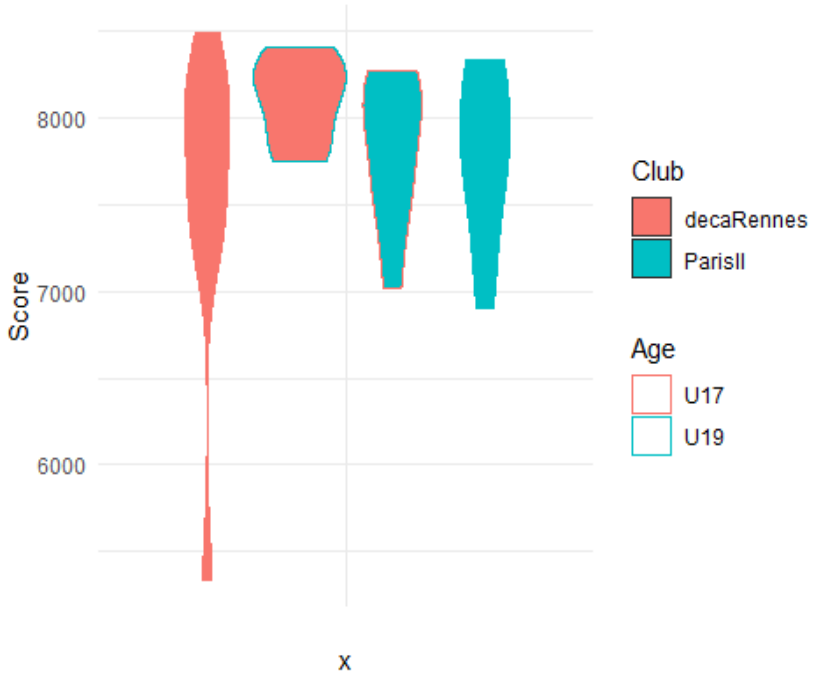
Interpretation

- Unlike sprint events, **endurance seems to be an area where both clubs are equivalent**.
- The management of endurance events may be similar in both clubs in terms of training and race strategy.

3. Analysis of Scores by Age and Club

Visualization: Violin Plot

```
ggplot(DECATHLON2) +
  aes(x = "", y = Score, fill = Club, colour = Age) +
  geom_violin(adjust = 1L, scale = "area") +
  scale_fill_hue(direction = 1) +
  scale_color_hue(direction = 1) +
  theme_minimal()
```



Observations

- The scores of **U17 Rennes athletes** are more dispersed than those of U17 Paris athletes.
- Conversely, among **U19 athletes**, the Paris group shows greater variability.
- **The performance gap between U17 and U19 athletes in Rennes is very significant**, whereas it is much smaller in Paris.

Interpretation

- **In Rennes, U17 and U19 athletes have distinctly different levels**, suggesting significant progression with age and training.
- **In Paris, U17 and U19 athletes have more homogeneous levels**, which could indicate a more rigorous selection strategy at younger ages.
- The greater dispersion of scores among U19 Paris athletes might indicate a **more marked difference in skill levels between the best and weaker athletes** in this category.

Conclusion

The analysis of Rennes and Parisian athletes' performances reveals several clear trends:

- Rennes athletes have an advantage in sprint events (100m, 400m), likely due to more effective training.
- The 1500m competition is balanced, with no notable differences between clubs.
- The performance progression is more pronounced in Rennes between the U17 and U19 categories, whereas Paris maintains a more consistent level across these age groups.

These observations provide insights for future performance analysis and improvements in both clubs.