HSR Season Analysis

Angelo Capasso

2024-03-27

Introduction

This report contains a detailed analysis about training sessions and matches of a football club in order to establish an evaluation on the entire season. The related dataset contains information about training sessions and matches and it is composed by several features:

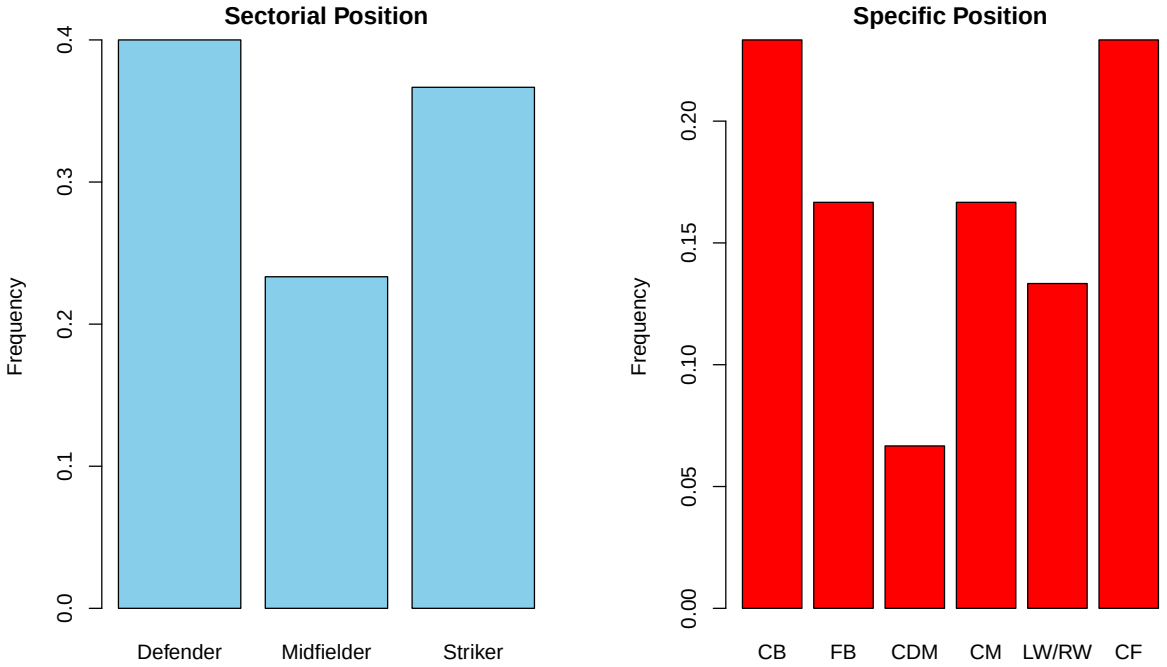
1) **IDNR** : Identification Number of the Player;
2) **Sectorial_Position** : categorical variable indicating the role covered by the player (1 = Defender; 2 = Midfielder; 3 = Striker);
3) **Specific Position** : categorical variable which indicates the specific role of the player (1 = Center Back, 2 = Full Back, 3 = Defending Midfielder, 4 = Center Midfielder, 5 = Winger, 6 = Center Forward);
4) **Time Session** : continuous variable expressing daily training time session;
5) **RPE** : continuous feature expressing daily perception of player effort during training sessions;
6) **S_RPE** : RPE for daily session, generated by the product between **Time Session** and **RPE**;
7) **S_RPE_Week** : Weekly Session RPE;
8) **HSR** : Daily training High Speed Running;
9) **HSR_Week** : Weekly training measurements of High Speed Running;
10) **HSR_Match_Championship** : HSR measured during Championship matches;
11) **HSR_Match_Cup** : HSR measured during Cup matches;
12) **Part_Match_Championship** : binary feature which is equal to 1 when a certain player participates on the Championship match and 0 otherwise;
13) **Part_Cup_Championship** : binary feature which is equal to 1 when a certain player participates on the Cup match and 0 otherwise;
14) **Time_Match_Championship** : game time on Championship match;
15) **Time_Match_Cup** : game time on Cup match.

Methodology

The analysis has been conducted using programming language R. More specifically, the backend calculations have been made using dplyr and tidyverse while the graphs have been creating through the use of ggplot2, pheatmap and ggofortify. These implemented packages have given me the possibility to build up different complex plots without renouncing to information's interpretability; in addition, they gave me the possibility to group the variables for doing the correct calculations.

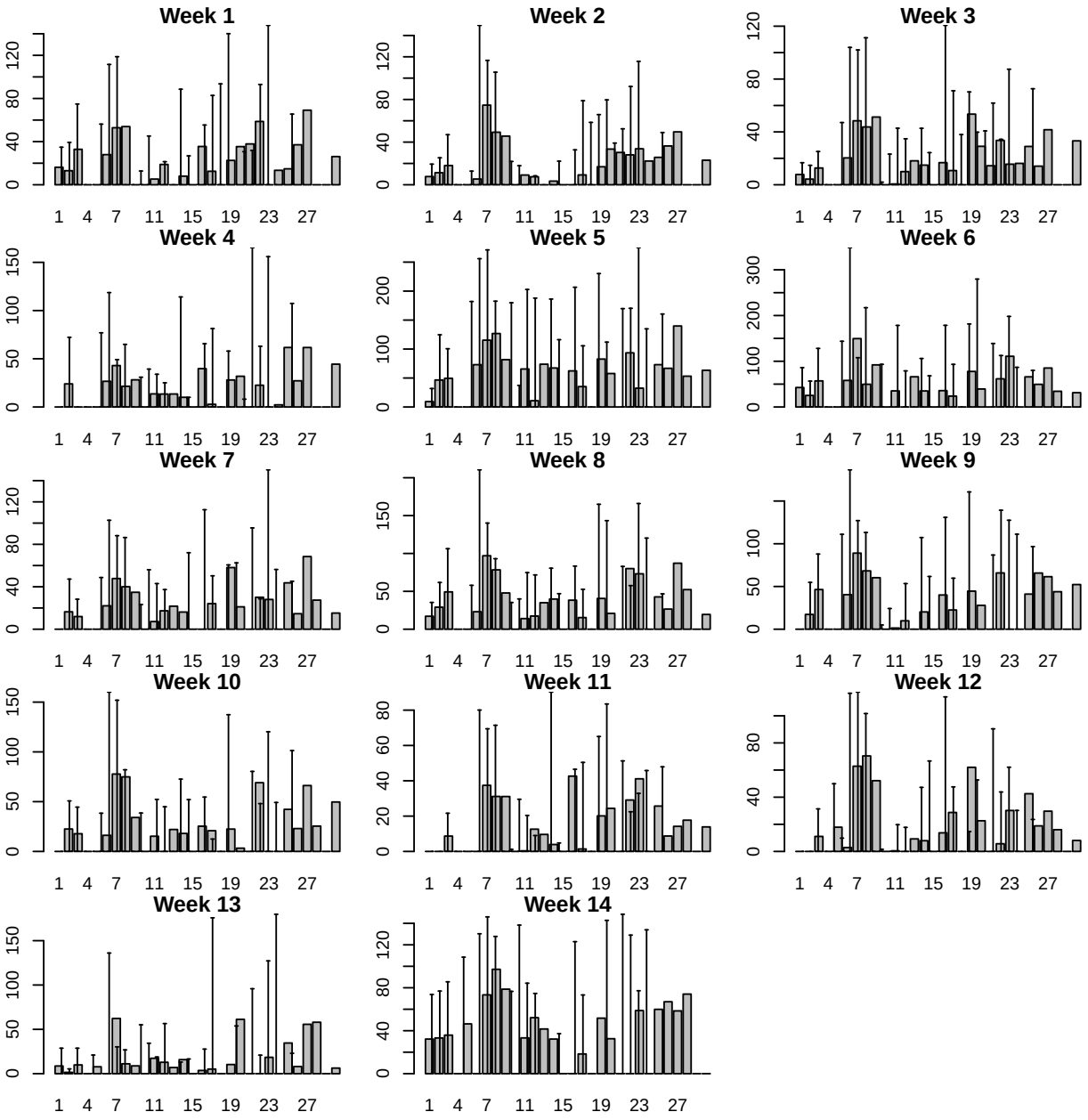
Descriptive Analysis of Player Position

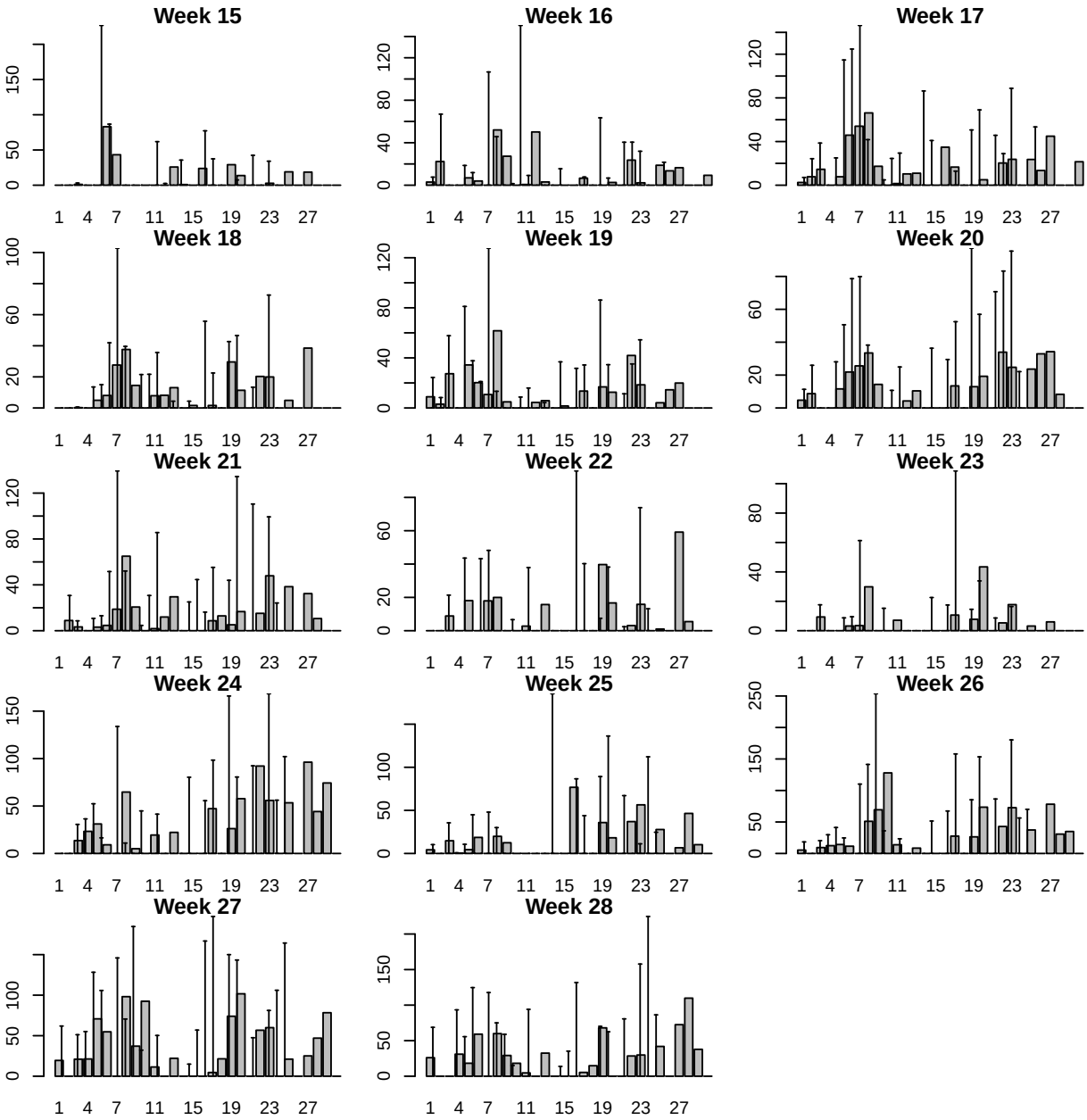
We can start our analysis taking a look to the team composition in terms for role and specific position. The plot in the bottom shows us this information and we can see that the team is composed by more defenders and strikers than midfielders. This can be involved with the preference of the coach into using formations with less center midfielders and wingers placed in left and right midfield, according to different scenarios.

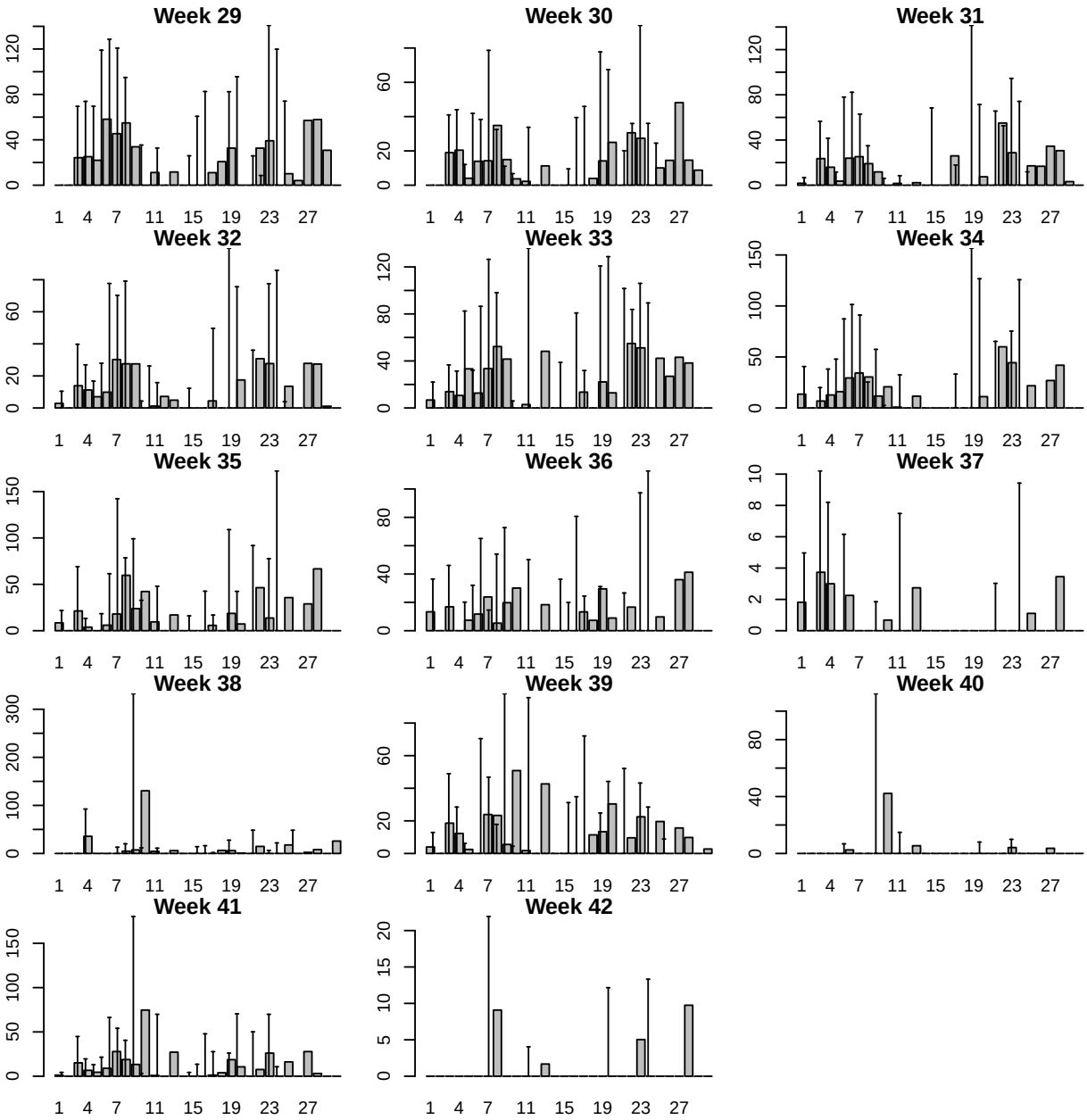


HSR Weekly Load Analysis

Analysing High Speed Running and evaluating player's performances in training is very important for improving the team's match performances because it give us to find out valuable insights in order to organize better the sessions Here there an overview of the mean weekly HSR measured for all 30 players:

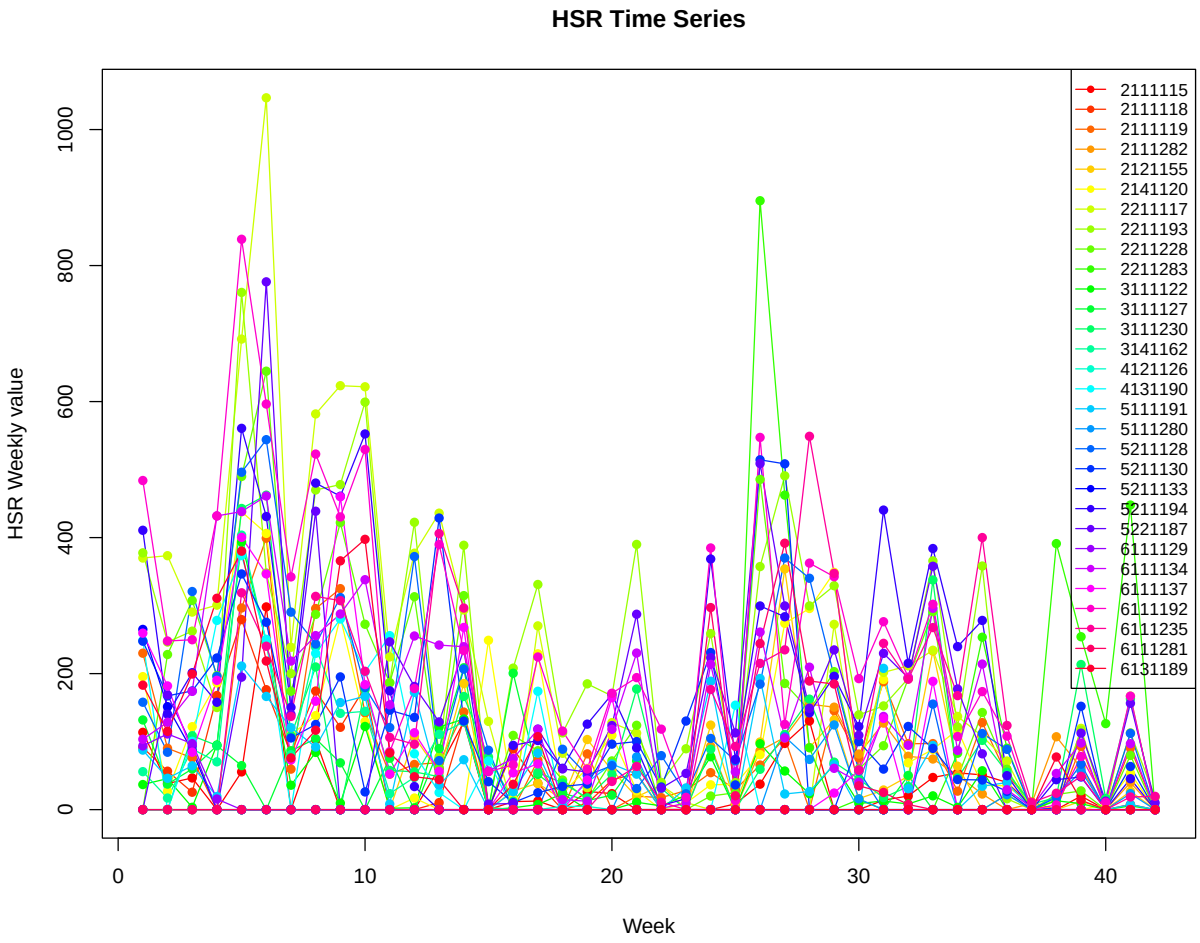

Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Week 9
Week 10
Week 11
Week 12
Week 13
Week 14
1 4 7 11 15 19 23 27


Week 15
Week 16
Week 17
Week 18
Week 19
Week 20
Week 21
Week 22
Week 23
Week 24
Week 25
Week 26
Week 27
Week 28



The height of the bar represents the Mean Weekly HSR for each player while the height of the error bar measures the standard deviation of weekly HSR training measurements. As we can see, the major intensity of training load have been registered in the first 14 weeks because the team had to face up both Championship and Cup matches. We can also see some bars with height near the 0 or equal to 0. When this scenario is repeated, it probably means that the player is injured and so the structure of training have to be modified. High variations in training loads during daily sessions can raise injury risks, so it can be useful to keep in mind the consistency during the season. Monitoring variations and weekly loads can help coach to organize sessions in a better way, adapting loads to the characteristics of the player. At the middle/end of the season, training loads decrese beacuse of the cup matches' absence.

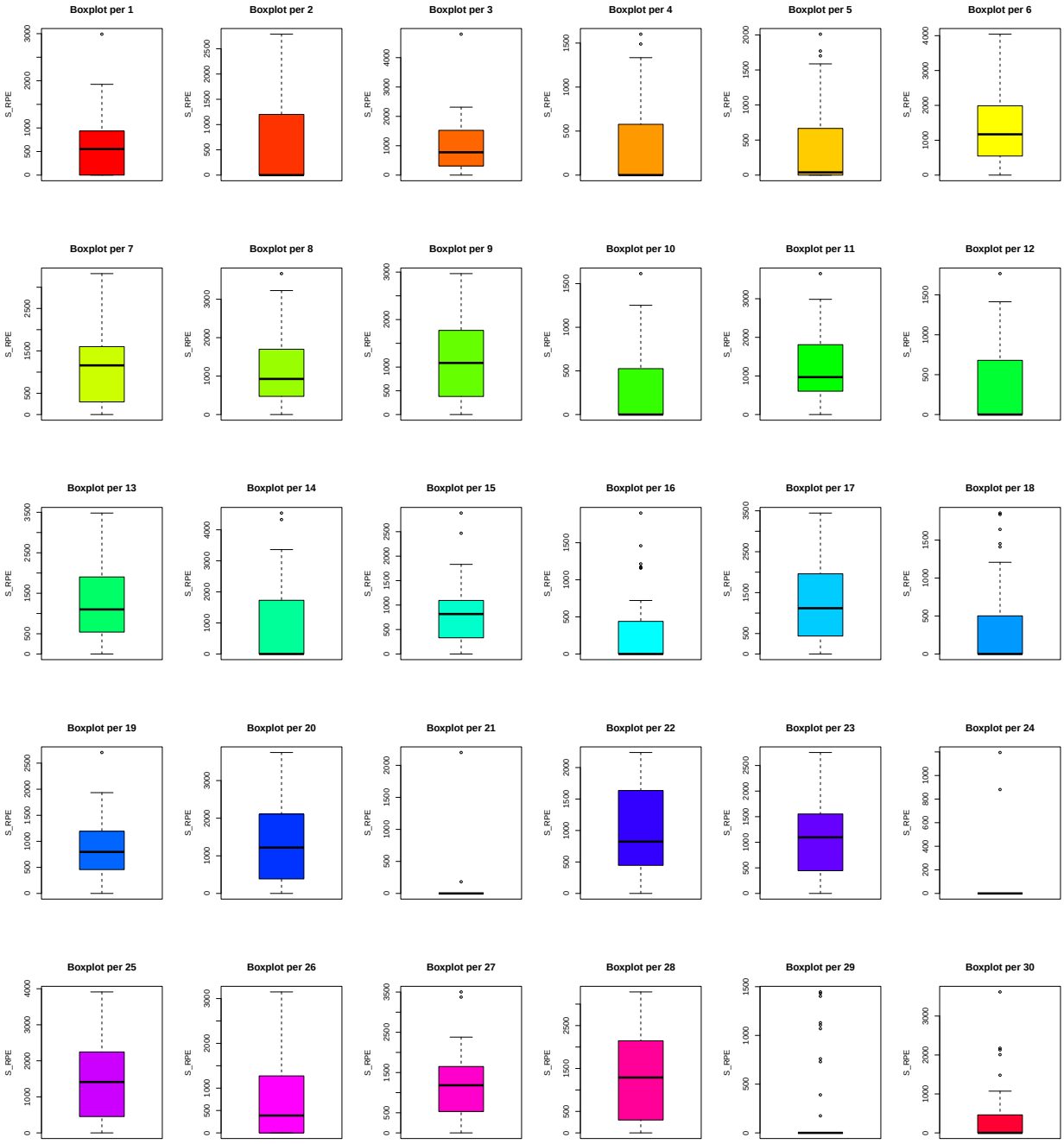
In addition, HSR can be studied as a time series, showing this result:



Perception of Fatigue

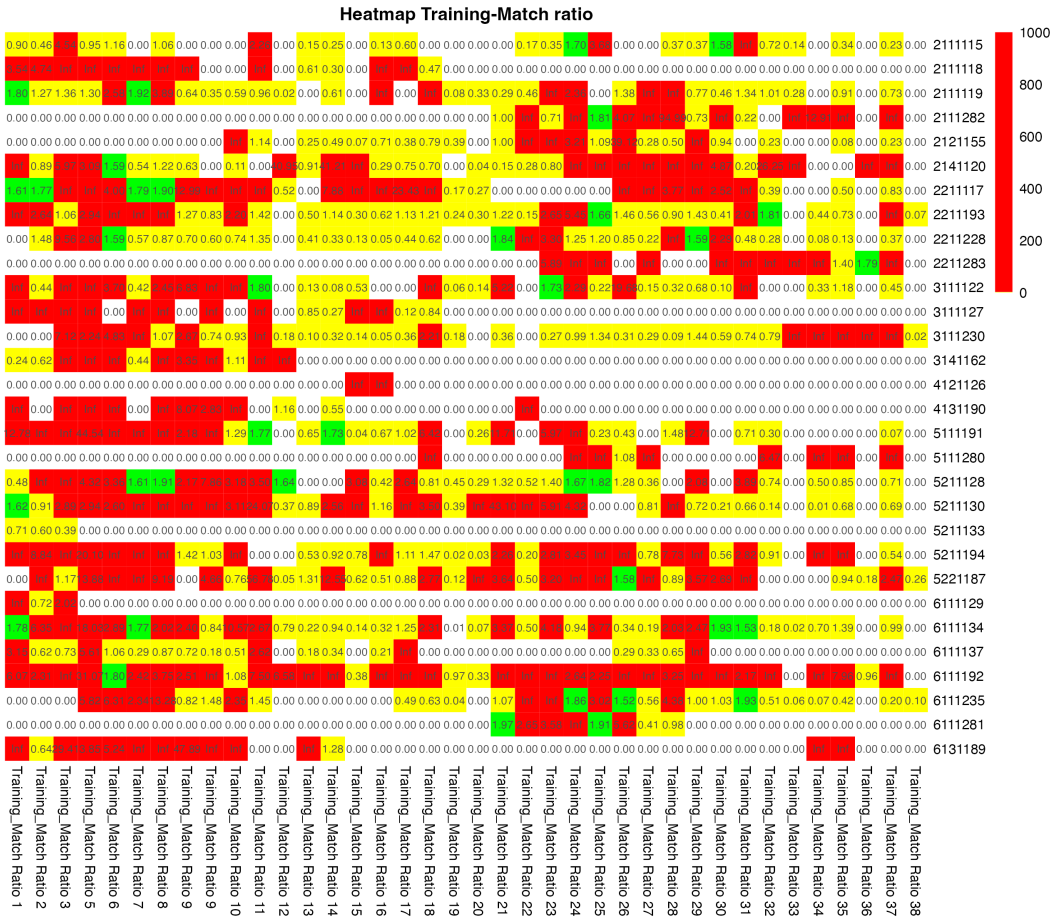
It refers to the subjective experience or awareness of fatigue by athletes during sports activities. It involves how athletes perceive their own level of tiredness or exhaustion during training or competition and how this perception influences their performance, decision-making, and overall well-being. When studying RPE to improve performances, it is recommended to consider both physical and psychological condition of the player and to relate to the other measured features (in this case HSR). In the next page a visualization of S_RPE distribution for every player is provided, crossing the entire season.

As you can see, players like 2 or 5 or 10 (which have the mean S_RPE equal or near to 0) can have suffered an injury during the season because of too intensive training sessions or training loads. For these kind of players, reorganizing training loads can result effective in order to maximize their performances in later games.


Boxplot per 1
Boxplot per 2
Boxplot per 3
Boxplot per 4
Boxplot per 5
Boxplot per 6
Boxplot per 7
Boxplot per 8
Boxplot per 9
Boxplot per 10
Boxplot per 11
Boxplot per 12
Boxplot per 13
Boxplot per 14
Boxplot per 15
Boxplot per 16
Boxplot per 17
Boxplot per 18
Boxplot per 19
Boxplot per 20
Boxplot per 21
Boxplot per 22
Boxplot per 23
Boxplot per 24
Boxplot per 25
Boxplot per 26
Boxplot per 27
Boxplot per 28
Boxplot per 29
Boxplot per 30
S_RPE

The importance of balancing training loads for improving performances

An important metric for evaluating athlete's fitness and effectiveness of training sessions in relation to matches is the Training-Match ratio. This ratio is an important consideration in sports performance and injury prevention, as it impacts the overall workload and stress placed on athletes. Below you can see a heatmap realized with values of training-match ratio for each week and for each player. Through the scale of colors, we can visualize better the information and to highlights important insights:



The values are grouped in 4 "clusters":

1) **White cells** : they always contains **0** and they are linked to players that have not trained in the specific week and they are injured;

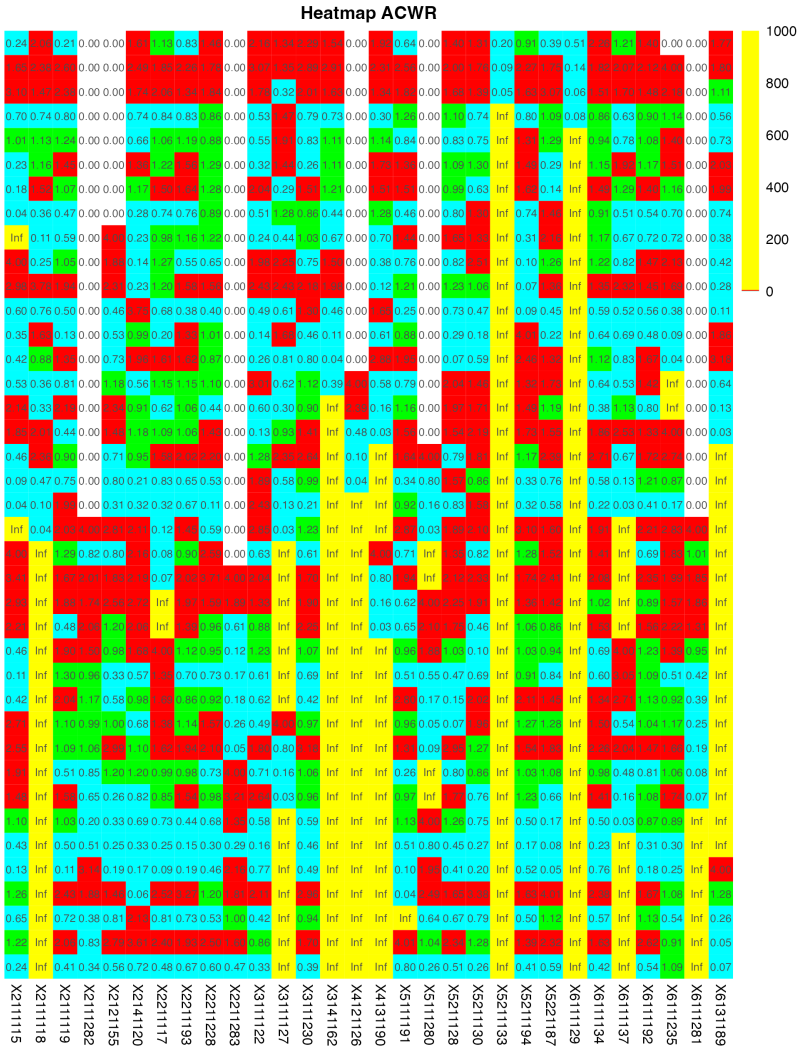
2) **Yellow cells** : they highlight a scenario when the player is "undertrained", so they identify a situation when the weekly training workload is insufficient for facing up the upcoming match and so the risk of increase increases a lot;

3) **Green cells** : they are associated to the best scenario, where the training workloads are optimal in order to face up the competitions. This give players the possibility to maximize their performances in games;

4) **Red cells** : they contains values expressing overtraining situations. When a player is in overtraining, the match demand is way less than training preparation and this can raise injury risks. Some values (infinites) indicate a scenario where the player trained during the week but does not participate on the game.

ACWR (Acute Chronic Workload Ratio) Analysis

ACWR a metric used in sports science and athletic training to assess the balance between the short-term (acute) and long-term (chronic) training loads experienced by athletes. This ratio is calculated by dividing the acute workload (typically representing the workload over the past week) by the chronic workload (typically representing the average workload over the past month or longer). The ACWR is used to monitor training loads and help prevent overtraining or injury by identifying periods of high acute workload relative to chronic workload. In this specific case, ACWR has been calculated for assessing that players have trained with optimal loads throughout the season:



The heatmap shows us some different scenarios:

1) **White cells** : Like the Training-Match ratio, an ACWR equal to 0 represents players that have not trained during those sessions (injured);

2) **Cyan cells** : they displays a situation where players' training load is underbalanced, i.e. it is too mild and these can generate cases that increase risk of injury. Since they represents the major composition of this matrix, we can assume that training load have not been adapted and this might have generated some of the injuries during the season;

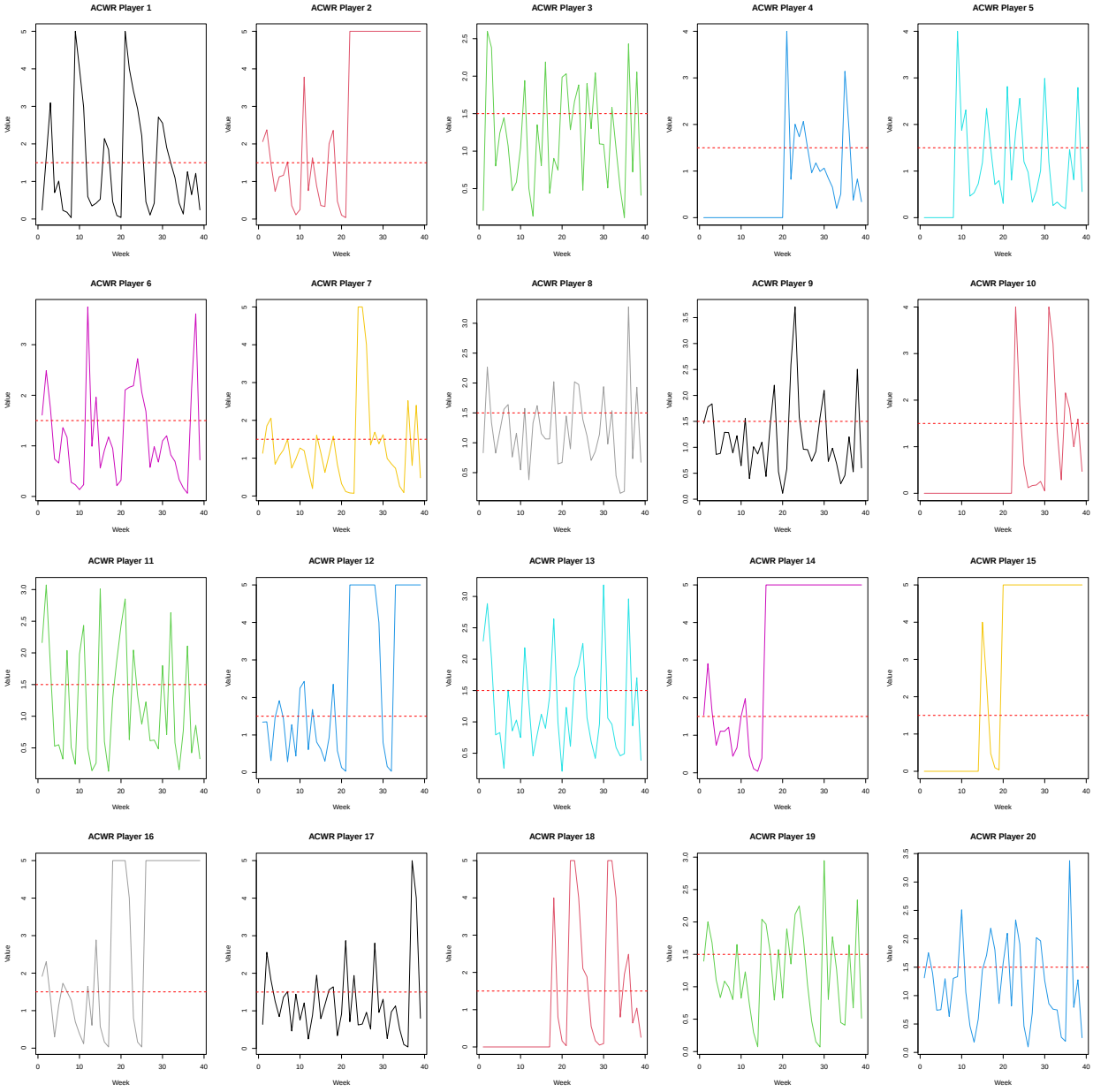
3) **Green cells** : they highlight the optimal situation where the ratio is contained in its optimal range. That means training loads have been well-balanced and this can help team to maximize the outcome during the game and to reach high level of fitness;

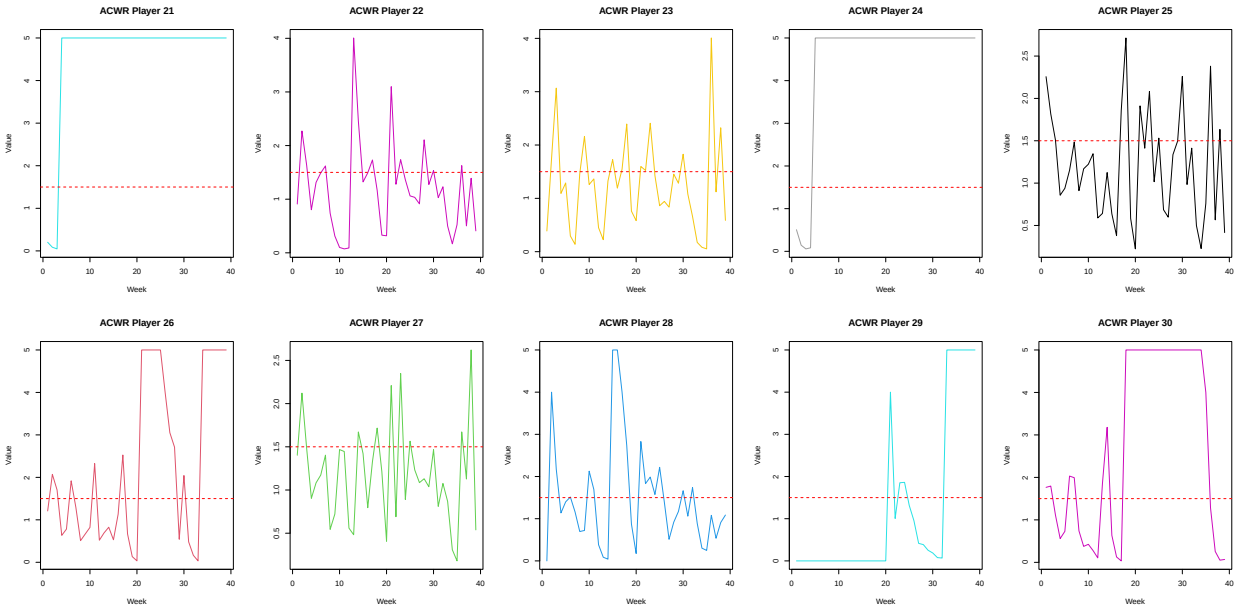
4) **Red cells** : they are linked to the worst scenario: training acute load has been overcharged than during previous sessions. This can produce as consequence more fatigue perceived by the player during sessions

and a higher risk of injury;

5) **Yellow cells** : they represents the case where a player was injured and then he returned to train or the scenario where a player showed a lower level of consistency in terms of training.

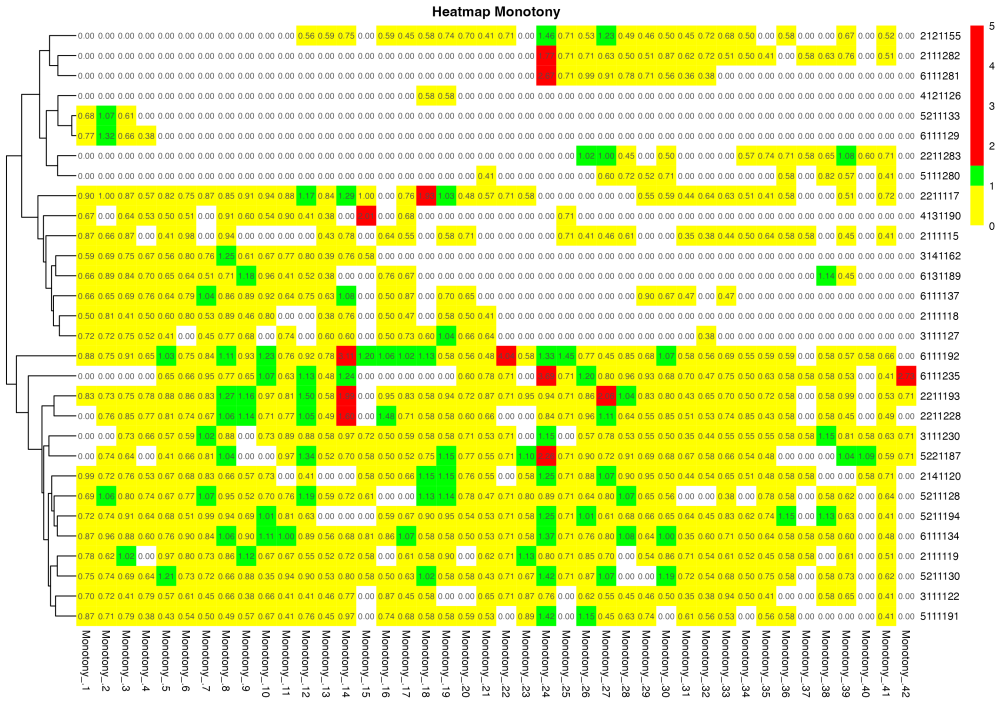
In addition, an overall trend analysis has been provided in order to understand the variations for each player. In particular, we can spot better the differences between the players in order to adapt changes according to the context.





Evaluating players' motivation during training sessions

Studying psychological aspect is also very important if we want to reach the goal of maximum physical condition for the team. In this case, training monotony is a useful metric in order to give some valuable insights on players' motivation. Training monotony refers to a condition in which an athlete's training program becomes repetitive and lacks variation in terms of intensity, volume, or exercises. When athletes engage in the same training routine repeatedly without significant changes, they may experience reduced motivation, performance plateaus, and increased risk of overuse injuries. Monitoring training monotony is important for coaches and athletes to prevent staleness and optimize performance by incorporating variety and periodization into the training regimen. Below you can diplay a plot which shows this information for each player and week:



The heatmap is implemented together with a hierarchical clustering algorithm (displayed as a dendrogram) and it highlights the presence of 3 cluster of players grouped in this colors:

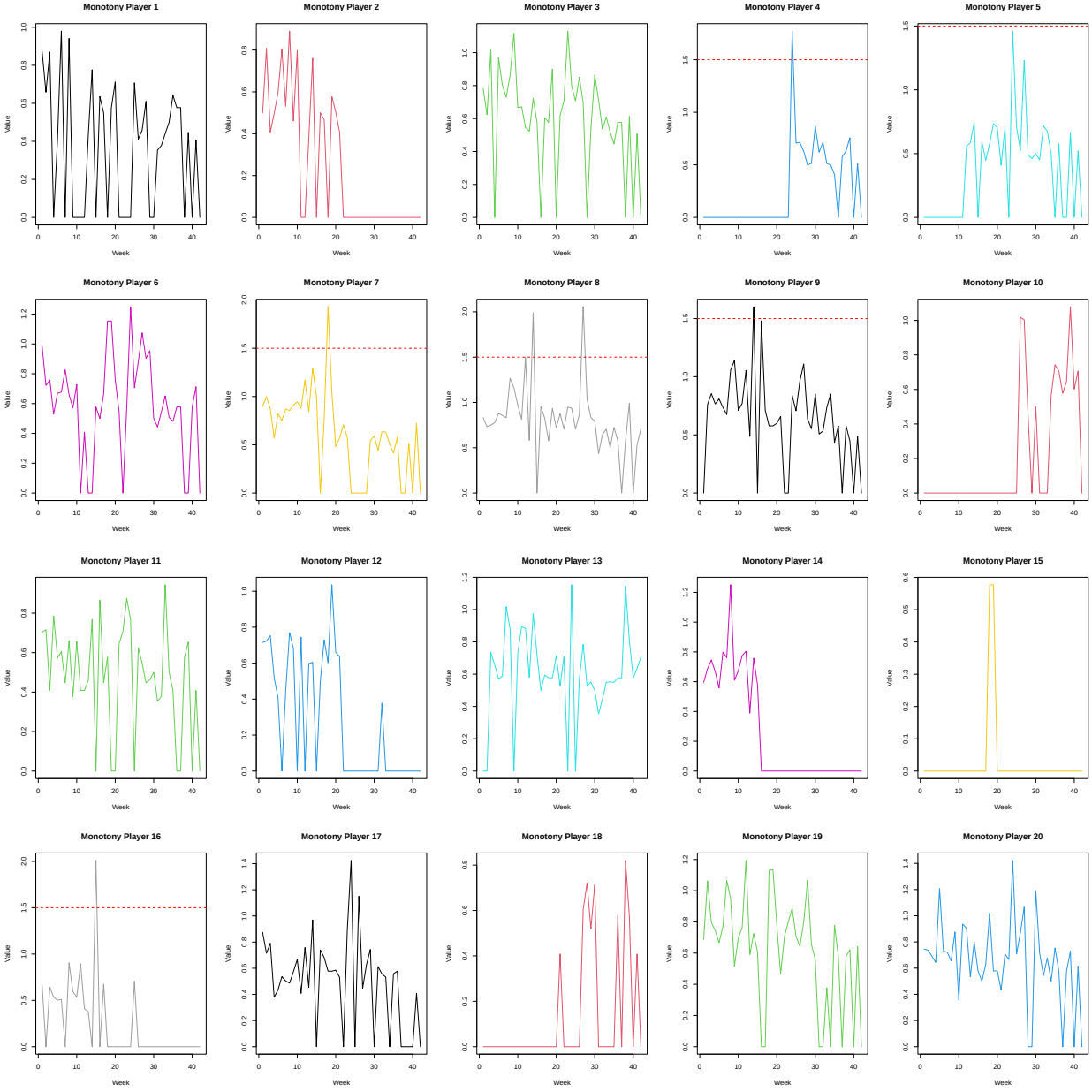
1) **White cells** : they identify the first cluster composed by players that have not trained because they are injured;

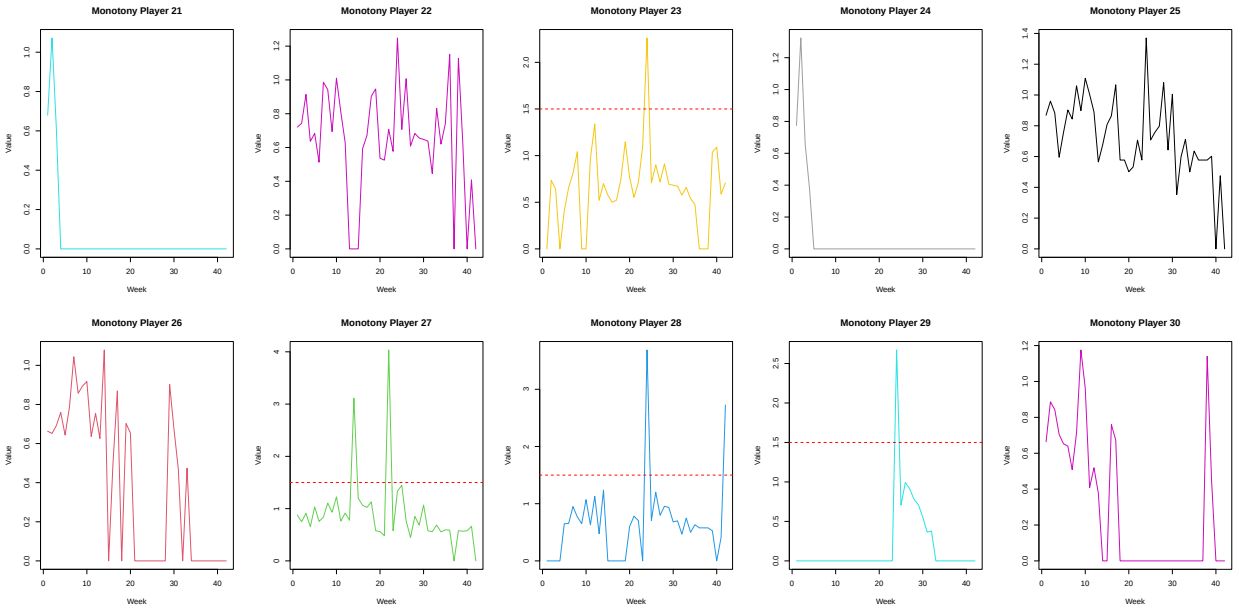
2) **Yellow cells** : they represents a scenario where players does not experience the right balancing of load but they are very different from a day to another and this can relate to lack of maximum performance in matches;

3) **Green cells** : they express the right trade-off between variations of training loads and consistency in training., which means that players can express their maximum potential;

4) **Red cells** : they express the situation when a player couldn't experience the changes in training loads but a sort of repetitiveness and this can produce a lack of players' motivation, influentiating the performance too.

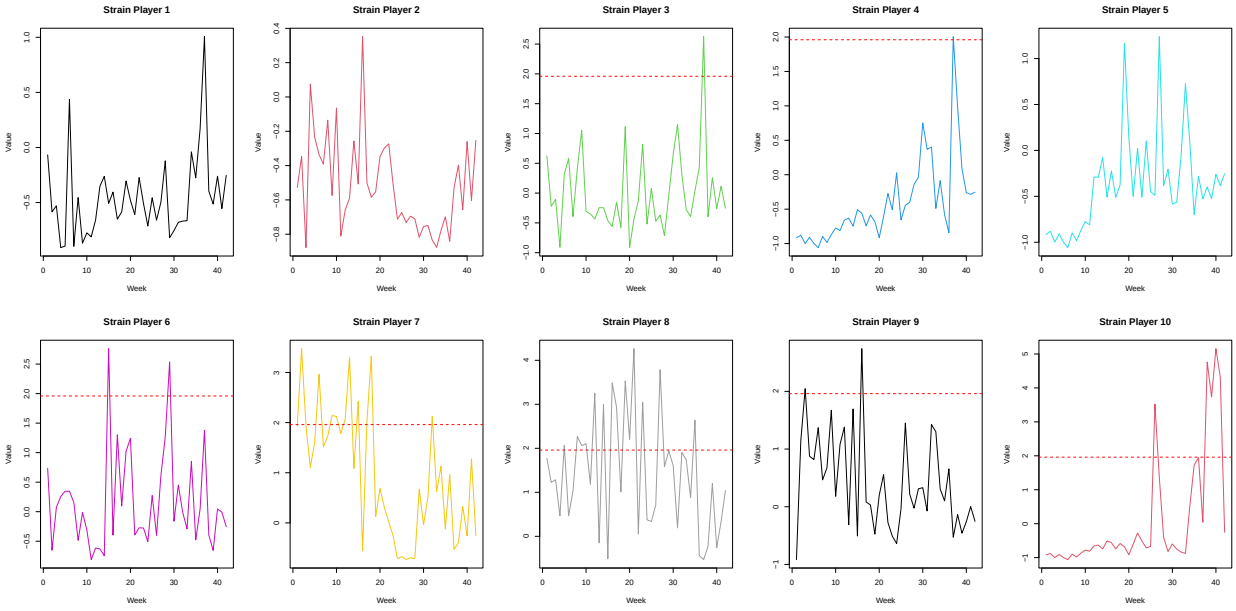
A better visualization of monotony's variations over time for each player can be spotted in the following subplot:

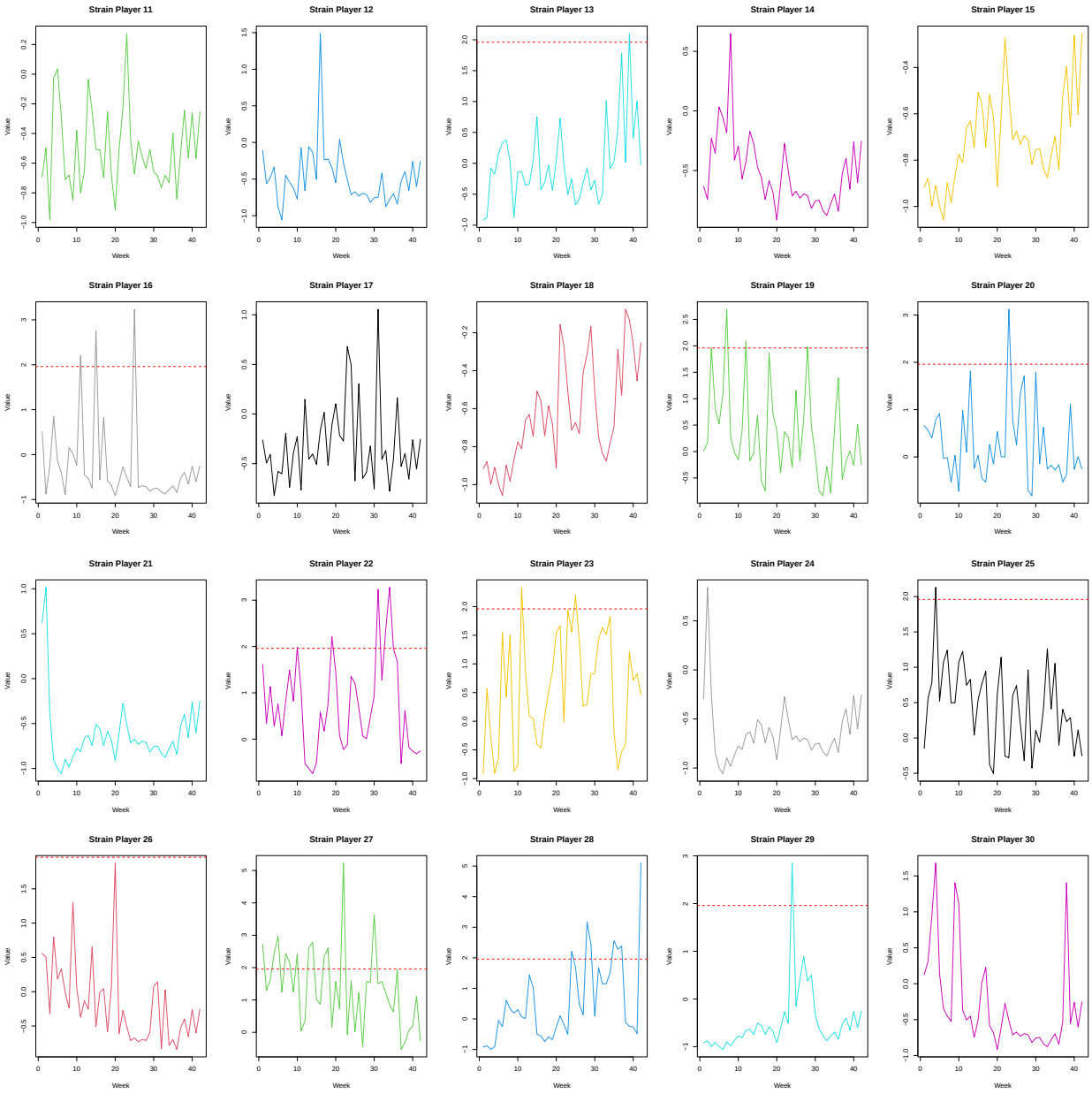




Training Strain Analysis

Training strain refers to the cumulative physiological and psychological stress placed on an athlete's body as a result of training. It encompasses various factors such as intensity, volume, frequency, duration, and the overall demands of training sessions. Monitoring training strain is essential for optimizing performance while minimizing the risk of overtraining and injury. Following there is an overall report of scaled strain for each player and each week:





In order to give valuable insights and support decisions making, a method for evaluating critical scenarios in terms of combination fo physical and psychological aspects has to be provided. One of the most used is related to substancial variations' analysis over time to identify situation of overtraining or high risk of injury. In fact, since I scaled every player distribution using standardized normal distribution Z, I considered a 10% of error threshold for both right and left tails in order to visualize an optimal range of "feasible" variations. Considering all these factors, when the line rimains in the range, it means that players have experienced a good balance of training sessions' structure, while when the line get out of the range, it highlights a too intensive variations in training and it can be linked with a higher risk of injury.