League of Legends Champions Dataset

Multivariate Analysis

Luísa Fontanete, Haonan Jin and Hang Yao

2024-12-23

Contents

1 Introduction

This data set contains detailed information about **167** champions from **League of Legends**. This is a PvP (Player versus Player) game that has outstanded from the other games since its released date. The game has basically 5 roles, which consists of **Top**, **Jungle**, **Mid**, **Bottom** and **Support**, and each role has its way to play the game. In addition, this data set also includes some initial champion statistics and his growth capacity.

The data set can be used for various purposes such as statistical analysis. The main objective consists to determine the best champion for each role, meaning the one who has the highest statistics in it determined role. Let's explain the variables of the data:

- Name: the name of the League of Legends champion.
- Tags: tags or legacy classes of the champion. List of tags: **Assassin**, **Fighter**, **Mage**, **Marksman**, **Support**, **Tank**.
- Role: the primary role or a position played by the champion.
  - Top: focus on defensive stats (HP, armor, magic resistance) and sustainable (HP regeneration). These champions typically are durable and have a few of attack damage.
  - Jungle: emphasize base armor, and attack speed. These champions often need good early-game stats for jungle clear efficiency.
  - Middle: look at mana-related stats, burst damage potential (high mana stats, high ability damage). Versatility and damage output are key factors.
  - Bottom: prioritize attack damage, attack speed, and attack range. These champions are expected to have high offensive stats but lower defensive stats.
  - Support: look at utility stats (mana regeneration, armor) and sustain stats (HP regeneration). Supports are less focused on damage and more on enabling and protecting their teammates.
- Range type: whether the champion is melee or ranged.
- Resourse type: the resource that a champion can generate and consume when using abilities or a basic attack.
- Base HP: the champions base health points (HP) at level 1.
- HP per lvl: the amount of hp the champion gains per level.
- Base mana: the champion's base mana points at level 1.
- Mana per lvl: the amount of mana points (MP) the champion gains per level.
- Movement speed: the champion's base movement speed.
- Base armor: the champion's base armor at level 1.
- Armor per lvl: the amount of armor the champion gains per level.
- Base magic resistance: the champion's base magic resistance at level 1.
- Magic resistance per lvl: the amount of magic resistance the champion gains per level.
- Attack range: the range of the champion's basic attacks.
- HP regeneration: determines the rate that a unit's current health passively regenerates, measured per 5 seconds.
- HP regeneration per lvl: the amount of health regeneration the champion gains per level.

- Mana regeneration: determines the rate that a unit's current mana passively regenerates, measured per 5 seconds.
- Mana regeneration per lvl: the amount of mana regeneration the champion gains per level.
- Attack damage: the champion's base attack damage at level 1.
- Attack damage per lvl: the amount of attack damage the champion gains per level.
- Attack speed per lvl: the amount of attack speed the champion gains per level (%).
- Attack speed: the champion's base attack speed (AS).
- AS ratio: adjusts the effectiveness of bonus AS from all sources.

In this project, we will apply various methods acquired in this master course, including **Principal Component Analysis**, **Multidimensional Scaling**, **Cluster Analysis**, and others, as our data set is well-suited for these analytical techniques. However, we have chosen to focus **Multiple Correspondence Analysis** instead of **Correspondence Analysis**, due to the results of the MCA is more suitable for the dataset, providing significant results to analyze them.

In the following sections, we will analyze several interesting variables. For each method, we will explain its objective and describe its specific role in the analysis. Finally, we will integrate the results from all methods to present a comprehensive conclusion.

2 Exploratory Data Analysis

2.1 Data Transformation

In this section, we have converted all categorical variables to factor variables for the posterior analysis, specifically the variables `Tags`, `Role` and `Range.type`. Also, we have observed that some rows in the `Resourse.type` have **NULL** values (" "), so we have converted these rows to the category **Nothing**, since that doing a categorical variables imputation is not all correct because each observation has itself type of resource.

We have fixed and transformed the variables `Role` and `Tags`, since these two variables contain more than one category for each observations, so we have decided to keep the first category that appears in each row because the main role and tag of an individual is the first ones. Additionally, we have converted the name of each observation to it row name.

Finally, as we can see in the figure 9, the numerical variable `Attack.range` is mostly like a categorical variable, since the distribution of this variable is non normally distributed. We have checked this behavior using **Shapiro Wilk Normality test** (table 1), and the p-value (4e-15) is less than 0.05, so the null hypothesis (the variable is normally distributed) is rejected. Moreover, it only has 19 different values, so converting to a categorical variable will be better to analyze.

We have used quantiles to assign the range type to each observation according to it attack range value, we have divided to three categories (**Close Range**, **Medium Range** and **Long Range**), this variable now is converted to a categorical variable with three levels.

2.2 Data Visualization

Firstly, let's analyze some categorical variables. The `Attack.range` variable, which we have converted to categorical variable, shows that the **Long Range** category is the highest one, suggesting most of observations are included to this type. Another interesting categorical variable is `Tags` variable, which indicates the tag for each observation. The most frequent one is the **Fighter** category and the least one is the **Support** category. That means most of the champions are fighter and only a small proportion are support. As we can see in the figure 10.

We could see boxplots of several numerical variables, which we have analyzed its behavior. As we could see that the `Base.HP` has two outlier champions (lower outliers), indicating these champions have low health at the level 1. Another boxplot, where has presence of outliers is the `Attack.speed` variable, one lower outlier and six upper outliers, meaning its `Attack.speed` is far from the mean value. The existence of outliers will show some interesting behavior between the points in the posterior analysis because the outlier variables might be the variables that distinct a set of observations from others sets. The other two variables do not have presence of outliers since the boxplots do not show any outliers. As we can see the figure 11 in the appendix.

2.3 Data Summary

In this subsection, we wanna analyze the proportion of the champions that matches their `Tags` with `Role` (table 2). We have observed a distribution and they seemed to be mostly correct distributed by the concept of each `Role` (we know that each `Role` has different tasks to do. For example, the tag **assassin** will be the perfect profession to go for the roles **Middle** and **Jungle**).

For another distribution (table 3), `Role` and `Attack.range`, we could observe that **Close Range** and **Medium range** categories do not have any champions with **Bottom** role, the most frequent role in these two categories is the **Top** role, and followed by **Jungle** role. On the **Long Range** category, we could see that **Bottom** and **Middle** roles are the dominant ones in it. In the case of **Top** role the proportion of **Long range** is very poor.

We have chosen various variables that we considered interesting to analyze their correlations. So we got some of the base statistics, such as attack damage and attack speed. Looking through the correlation plot, we can see most of the variables are not even correlated but only one of them seems to be positive correlated such as `Base.armor` and `Attack.damage`. Also, some negative correlation between these numerical variables, such as `Base.armor` and `Base.mana`. As we can see in the figure 12 and table 4.

3 Principal Component Analysis

In our project, the PCA is used to analyze the correlations among the different statistical variables of the champions, and to understand the relationships between the individuals and their variables. This analysis helps us identify the strongest characteristics of each champion and explore how champions relate to each other based on their statistics. We included all numerical variables, except `Movement.speed` and `AS.ratio`, since these variables are not correlated with any other variables. Also, the **"best"** champion (of each role) is not depended on these two variables.

According to the Kaiser Rule, the dimensions 1-5 are chosen. But according to thumb rule, the dimensions 1-6 are chosen. Since the dimension 6 has a eigenvalue close to 1, so we have decided to choose the dimensions 1-6. As we can observe the figure 13 and table 5 in the appendix section.

Looking at the figure 14, the first dimension is dominated by the variables `Base.HP`, `Base.armor`, `Magic.resistance.per.lvl`, `HP.regeneration`, `Attack.damage`, so we can denominate this dimension like **Durability and damage statistics**, since these aspects match to those champions (durable with high damage). In the second dimension, we could name it as **Mana growth**, due to the dimension is related with the mana growth variables. In the third dimension could be named like **Attack speed growth and regeneration capacity**, given that the correlated variable is `Attack.speed.per.lvl`, `HP.regeneration.per.lvl` and `Mana.regeneration`.

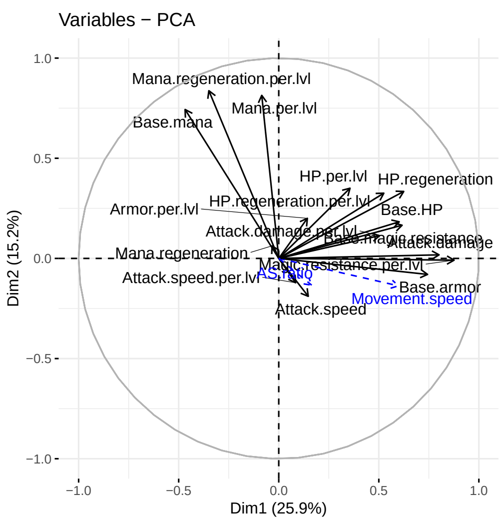
In the dimension 4, the most related variable is `Armor.per.lvl`, so is named as **Armor growth**. The `Attack.speed` variable is the one that most correlated with the dimension 5, so it could be named as

Basic attack speed. Then, the dimension 6 can be denominated as **Health growth**, since the variable `HP.per.lvl` is dominant one.

As we said previously, the below figure (figure 1) shows the variables most correlated to each dimensions. The first plot, which indicating the dimensions 1 and 2, the individuals that situated at the right side have a good durability and damage statistics, and the ones, who positioned at the second quadrant have a good mana growth capacity. Moreover, the champions that positioned to the negative side of the dimensions are not "good" champions in term of durability, damage and mana growth.

Looking at the second plot of the figure, which representing the dimensions 2 and 3. We could observe that the dominant variables of these dimensions are related to regeneration and growth, such as `Mana.regeneration`, `HP.regeneration.per.lvl`, `Mana.generation.per.lvl`, `Attack.speed.per.lvl` and `Mana.per.lvl`, `Base.mana`. Indicating the champions that positioned to the positive side are those that have good regeneration and statistics as the level increases. The negative side of the dimensions indicating good durability, damage and attack speed at level 1 and great attack speed as the level increases, those individuals are better champions at the level 1 compare to anothers ones, as they have good statistics only at the first level.

On the other plots (figure 15), reflect the same interpretation as the plots that we have interpreted previously, only changing the combination between dimensions. In conclusion, the individuals situated at the positive side of the dimensions are the ones most correlated with the dominant variables in those dimensions.



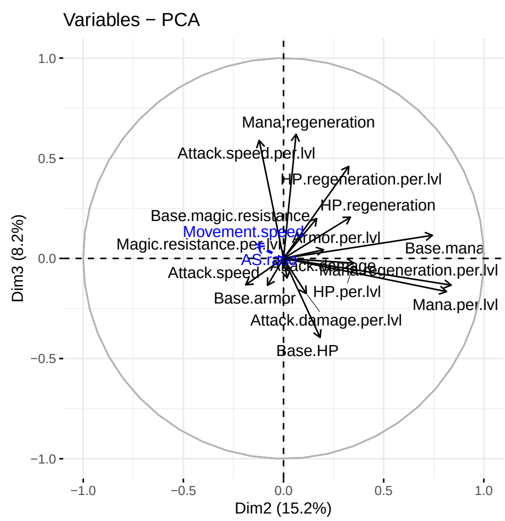


Figure 1: PCA Variables Dimensions 1, 2 and 3

Another behavior to demonstrate the most correlated or dominant variables in each dimension is use the contribution of variables to dimensions. The variables which have high contribution in dimensions are the dominant ones. In the figure 16, we could see the contribution of variables to dimensions 1-2 and 2-3. The results shown that the most contribution of each pair dimensions are the ones that appear in the anterior plots.

Let's analyzing the individuals relation with the dimensions, the following figure 2 also show the relevance of champions in the dimensions (the individuals related with the dimension are the relevant ones). As we said previously, the dimension 1 is dominant per high durability and damage champions. The dimension 2, the champions with high mana growth. So, in the first plot, which projects dimensions 1 and 2, the champions that situated at right side are those that have great durability and damage statistics, such as "Camille", "Darius", "Tryndamere" or "Shyvana", but the case of "Tryndamere" and "Shyvana" are situated at the quart quadrant, indicating a high durability and damage statistics with a low mana growth. In the second quadrant, we could observe various correlated champions with the dimension 2, such as "Syndra", "Ryze", "Zilean", "Kassadin", with a high mana growth but a low durability and basic attack statistics. The extreme champion "Briar", which situated at the bottom of the third quadrant, has a low mana growth, durability and damage statistics, as it is totally at opposite side of the correlated variables in these two dimensions. Additionally, we want to highlight the champions positioned around of the center of the plot, which have a low variance in this plot, meaning they are not or a few correlated with the dominant variables in these two dimensions. Then these champions variables, specifically the dominant ones in the dimensions, are not significance to these dimensions.

Observing the second plot, which projects the individuals in the principal components 2 and 3. We could see interesting aspect, where most of the champions are around of the center. Only the far ones showing a high variance with the dimensions, since they are more spread around the plot. Looking at the most right side of the plot, the champions "K'Sante", "Ryze", "Naafiri" have a high relation with the dimension 2, indicating a great mana growth compare to another ones. On the top of the second quadrant, we could see that the individuals are more spread because the relevance or variance of their variables are different. For example, "Akali" are the one that best matched with the dimension 3 (high attack speed growth and regeneration). On the other hand, "Kennen", "Shen", "Lee Sin" are correlated with the dimension 3, but not as much as "Akali", this is caused that they have some correlated variables with low values.

We could observe that the bottom of the third quadrant, have some relevant champions, such as "Briar", "Zac", "Rengar". They are negative correlated with the dominant variables in these two dimensions, meaning they do not have a nice mana growth neither a good attack speed nor regeneration capacity. This information is useful because we can know which individuals are "better" or "worst" in the dimensions. Moreover, we can conclude that closely positioned observations share similar dominant variables within the dimensions. For example, in the first plot, the bottom of the fourth quadrant, the champions "Zac" and "Riven" are very close, suggesting they have similar values in `Mana.regeneration.per.lvl`, `Base.mana`, `Magic.resistance.per.lvl`, `Attack.damage`, `Base.armor`, `HP.regeneration`.

The other plots (figure 17) have no sense to analyze since the most of champions are around of the center, only a few have significant variance in those dimensions. So, the dimensions which have more variance are the dimensions 1 and 2, since they capture most of variance of the variables.

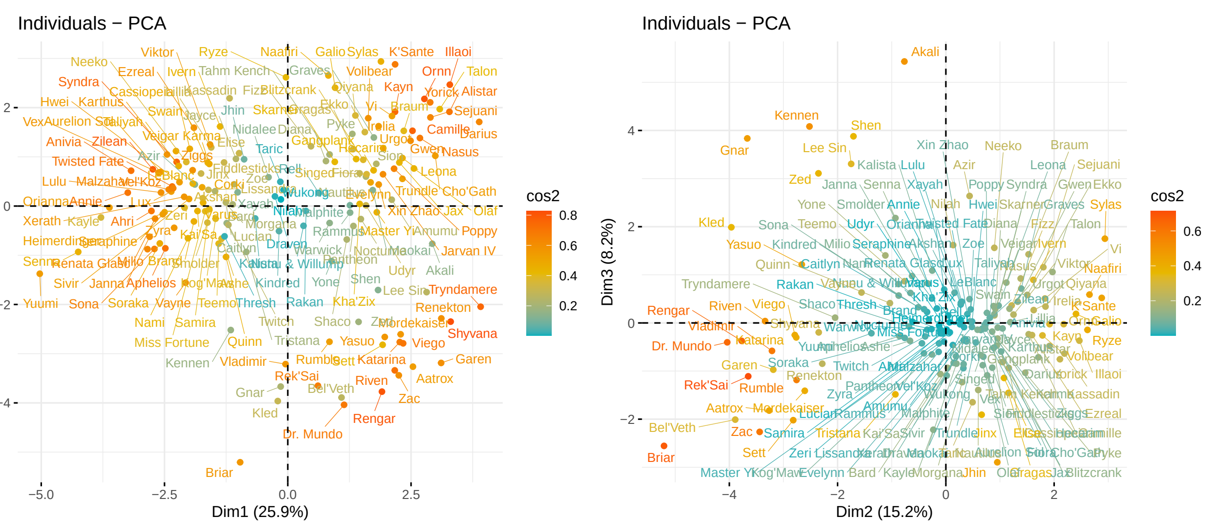


Figure 2: PCA Individuals Dimensions 1, 2 and 3

The PCA successfully demonstrated the correlations among the statistical variables of the champions, such as the champions with a high `Attack.damage`, also have a high `HP.regeneration` or `Base.armor`, since these attributes are related in the reduce dimensions. Also, provided ideas into the relationships between champions and their variables to determine the strongest ones.

After analyzing the dominant variables of each dimension, and according to the role descriptions mentioned in the **Introduction** section (each role has it strongest variables). We could determine the **"best"** champion for the roles using this relationships between role and variables (if it is possible).

We only could use the first two principal components to determine the **"best"** champions according to the roles, due to that the dominant variables in other dimensions are not totally matched with the "role variables" (the strongest statistics of each role). The correlated variables to the dimension 1 are those related to defensive stats and damage, these variables match perfect to the **Top** role, so the strongest champions in this role are those that have high correlation within the dimension 1, such as "Tryndamere", "Darius", "Illaoi". On the second principal component, the dominant variables are related to the mana stats. It match perfectly with the **Middle** role, so the champions positively correlated with this dimension are the most powerful ones, such as "Sylas", "Ryze", "Naafiri", "Galio". As we can observe in the figure 18.

4 Multidimensional Scaling

We use MDS because we want to visualize how similar or different characters are to each other based on their attributes (HP, mana, armor, etc.) in a way that's easy to interpret visually and also see if characters with similar roles cluster together, suggesting they share common characteristics. In this case, we do not use all the variables, rather we chose for the numerical only theirs base statistics (without including the variables that contains "per.lvl") and include all the categorical ones. As we can see the coordinate distributions in the following figure 3.

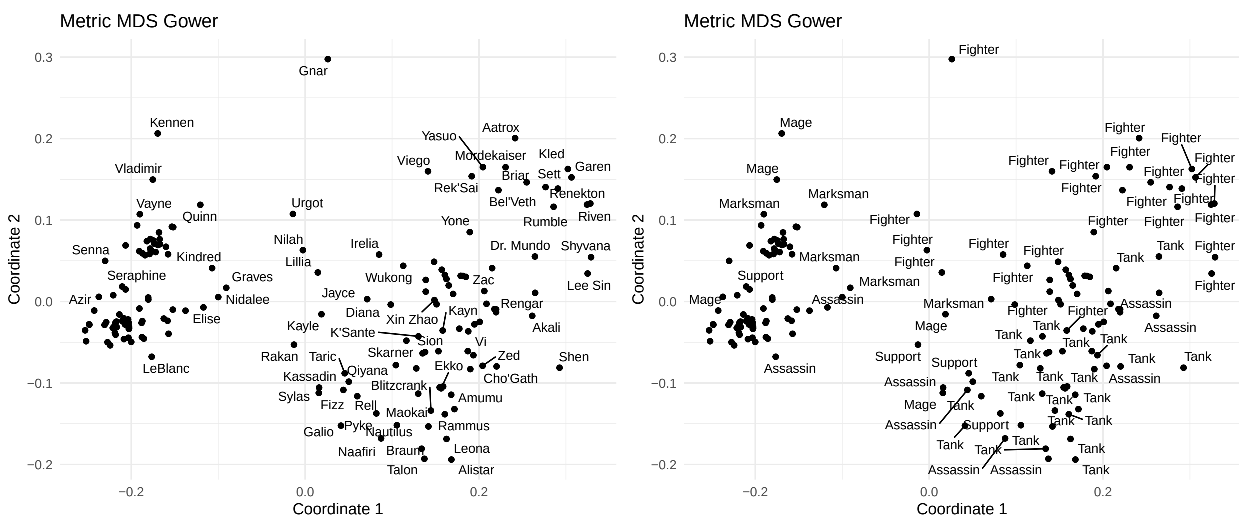


Figure 3: MDS Coordinates

Let's first take a look into the individuals that are separated as long as possible referenced in coordinate x . Those will be "Shyvana", "Lee sin" compared with "Senna". Comparing these champions we can see that "Lee sin" and "Shyvana" has similar attributes in numerical which are `Base.armor`, `HP.regeneration`, `Attack.Damage`, and for the categorical we got the Tag of **Fighter** and role of **Jungle**. As we can see "Senna" is quite "opposite" from them as she has not high `Base.HP` as the other two and also it differs from the `Tags`, `Role`, `Range.type"`. As we can observe in the table 6.

Now let's continue with the extreme of the y axis in the middle. The possible candidates will the individuals represented in the below table. According to the following table we can see that "Alistar" seems to be more durable and tank as it has a high number of `Base.Armor`, `Base.HP`, `HP.regeneration`, while "Gnar" seems to be more fragile (low level of `Base.armor` and `HP.regeneration`) compared to "Alistar", so he is separated from the others cluster (isolated). But we got some champions like "Aatrox", that is also positioned in the top but it also has high **Fighter** attributes. As we can see in the table 7.

Next, we see that we got a huge cluster in the left side in the plot, let's take a look the reason why they are in that place. We see that all their `Base.magic.resistance` are the same and they are all long attack range. Moreover all these champions, are from the role **Bottom** or **Middle**. All these similar attributes is the reason why they are positioned along the plot. We can observe that this huge cluster is separated to two sub-clusters, this is caused that the individuals have different roles and tags, **Bottom** and **Middle** role mentioned previously, and **Mage** and **Marksman** tags. In the table 8, we can see the common variables.

To conclude this about the coordinates of the 2-dimensional visualization. We see that the coordinate 1 seems to differs the champion like separating the ranged champions (left side) to the melee champions (right side). Thus, as we know most of the ranged type champions are **Mage** and **Marksman**, so it kind of separates also from the **Mage/Marksman** to the **Fighter/Assassin/Tank**. The second coordinate, we can not conclude anything as we do not have evidences by looking the positions that these champions are distributed. Even thought, this visualization allows us to see which characters have the most similar overall attributes and which ones are more distinct. For example, **Mage** and **Support** roles cluster closely together, suggesting they have quite similar characteristics. Meanwhile, **Tank** and **Marksman** occupy more distinct regions of the plot, given that the variables are very different between them, such as `Range.type` or `Base.armor`.

5 Multiple Correspondence Analysis

The purpose of using MCA is because we want to understand the relationships between categorical variables (`Tags`, `Role`, `Range.type`, etc) that describe the champions and also see how different categories of variables associate with each other. For example, do certain roles tend to have specific range types or resource types? Or maybe identify clusters in how character attributes are distributed across different roles and tags. The variables analyzed are all the categorical variables, these variables describe different attributes making MCA an ideal choice.

For all those dimension that have a eigenvalue higher than 1/12 (we got 12 dimensions), this is the condition to take the number of dimensions. So cause of that, we will take 8 dimension. In the operation of MCA, we got a variable `Resourse.type` that is not influence too much in our analysis, so we put this variable as `quali.sup`. As we can observe the figure 19 and table 9 in the appendix section.

The right plot of below figure 4 represents the contributions of the original variables to the two main dimensions identified by MCA. Dimension 1 is strongly influenced by variables related to attack characteristics, such as `Range.type` and `Attack.range`. These variables has a significant contribution along this dimension, it means that this dimension captures a really important information for the association between the ranged and melee characteristics for all the champions. Dimension 2 seems to be linked to champions roles and their styles of combat. Thus the second dimension is influenced by the variables `Role` and `Tags`. These variables have a high contribution to Dimension 2, indicating that this coordinate captures information about the different categories based on roles and play style.

The left plot shows the positions of individual categories for each variable, demonstrating the associations between them. We can observe that categories like **Role_Bottom** and **Tags_Marksman** are positioned far along Dimension 1, indicating strong associations with long-range characteristics. This is certainly true, as we know all the champions that play the tag of bottom marksman, their attack range is long. **Close Range** and **Melee** categories are positioned opposite to **Ranged** and **Long Range**, reinforcing the interpretation of Dimension 1 as capturing range-related variability. In addition, **Role_Support**, **Role_Middle**, **Tags_Mage**, and **Tags_Support** seems to cluster together, suggesting similarities in their attributes, such as resource type of mana and long attack range.

Another interesting point is that **Tags_Mage** are very close to **Role_Middle** and also **Tags_Fighter** are very close to **Role_Top**, by seeing this we can extract the information that most of the champions that goes to **Middle** they are **Mage** and by the same reason for the **Top** and for the others roles. Other combination of dimensions are showed in the appendix with the same interpretation (figure 20).

Looking at the below two plots (figure 5). The left one, where individual data points (champions) are colored based on their roles. While the second plot shows a MCA factor map with the quality of representation

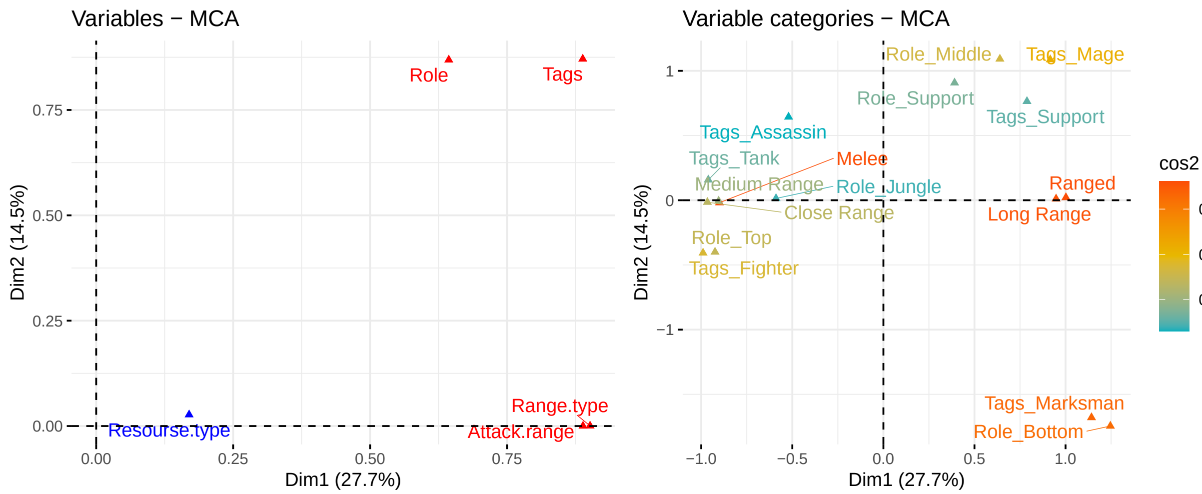


Figure 4: MCA Variables and Categories in Dimensions 1 and 2

indicated by a color gradient. High quality values (in red) indicate champions that are well-represented by the first two dimensions. For instance, champions like "Jhin", "Kai'Sa", "Miss Fortune", and "Caitlyn" have high cos2 (bottom-right side), suggesting that their characteristics align strongly with the captured dimensions (**Long Range**, **Tags_Marksman**). Low cos2 values (in yellow) suggest champions that are less well-explained by these dimensions, possibly due to having different roles or unique attribute combinations. Champions like "Viego" and "Urgot" appear in the middle of the plot (center), indicating their attributes do not fit neatly into the primary dimensions of attack range or utility.

Moreover the right plot can be classify by clusters, the bottom-right cluster is dominated by marksman ("Jhin", "Kai'Sa"), highlighting their shared attributes such as high attack range. The top-right cluster contains many mages ("Orianna", "Lux") and supports ("Lulu", "Janna"), indicating a grouping based on **Middle** and **Support** roles due to they share some similar variables. The bottom-left section is contributed with melee fighters and tanks ("Yasuo", "Garen", "Riven"), aligning with **Close Range**, and finally the top-left section features assassin champions ("Zed", "Katarina").

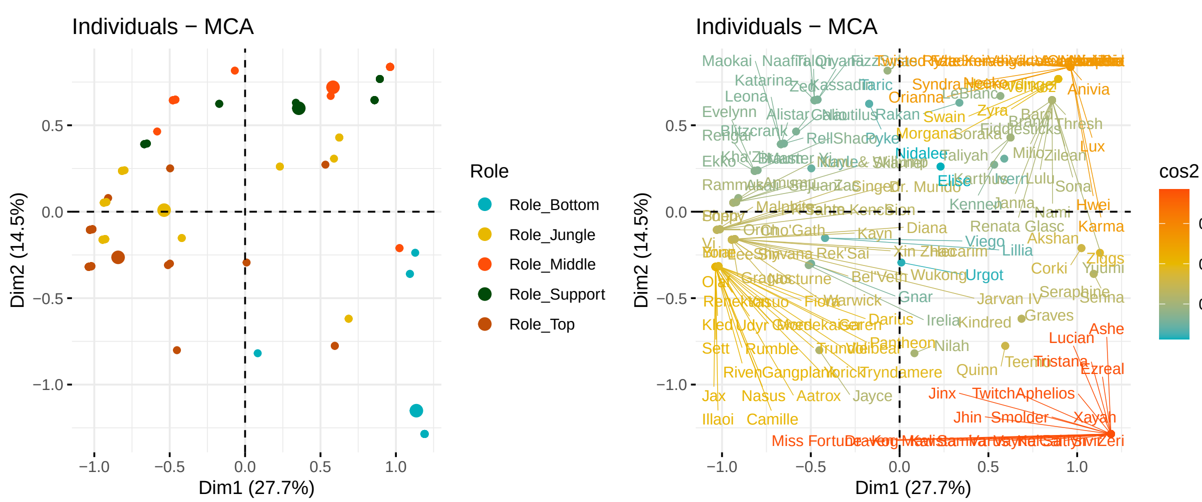


Figure 5: MCA Individuals and Categories in Dimensions 1 and 2

Now let's analyze the MCA grouping with Dimensions 3 and 4 (figure 21). We can observe that champions like "Rengar", "Evelynn", "Ekko", and "Khazix" have high cos2 values (top-left side), meaning they are well-represented by these dimensions. These champions are known for as playing the role in **Middle** and **Jungle**, aligning with the interpretation of Dimension 4. While the other champions such as “Rakan”, “Milio”, and “Taric” are well represented in Dimension 3 (Middle-right side), likely due to their roles and tags as being a **Support**. Finally **Top** and **Bottom** champions like “Tahm Kench”, “Nasus”, “Volibear”, “Caitlyn”, “Kai’sa” have lower cos2 values (around of the center of plot), suggesting they possess attributes not well-explained by Dimensions 3 and 4. Additionally, in the appendix has more plots of other dimensions (figure 22).

All these MCA plots successfully illustrate the relationships between champions based on their roles and other attributes. Champions are distinctly separated into clusters based on their roles (bottom marksmen, middle mages, top tanks), like the `Role` and `Tags` variables are related.

In Dimension 1, captures differences based on attack range type (melee or ranged), while Dimension 2 captures variations in the tag according to the range type. For example, the right side separates the **Mage** and **Support** from **Marksman**, and the left side separates **Tank** and **Assassin** from **Fighter**.

Dimension 3 seems to capture differences in champion aggression, burst potential, separating utility supports from offensive champions roles like **Middle**, **Jungle**. Finally, Dimension 4 has high contribution champions like “Evelynn” and “Rengar” playing the Tags of **Assassin** and role of **Jungle**.

6 Cluster Analysis

The objective is to determine the optimal number of clusters k, which grouping champions based on their gameplay characteristics, allowing us to compare and analyze how similar or distinct they are. For that matter, we will calculate the resulting optimal k from applying either hierarchical or unhierarchical clustering methods. To achieve this, we only use the numerical variables in the dataset. By combining base stats with scaling (variables that include “per lvl”) we get a more comprehensive view of the attributes of each champion.

6.1 Cluster Selection

After doing a study comparison between all the different types of dendrograms available for hierarchical clustering (figure 23), the **Ward’s** method dendrogram gave the best results in our case. Additionally, we see that the champion “Thresh” is an outlier individual, since it is isolated from the other observations, indicating that none has common characteristics with this champion. So, we decided to remove it from the cluster selection. As we can see this plot in the appendix (figure 24).

The structure of this dendrogram (figure 6) shows very distinctively the clusters, making it an ideal choice for understanding the grouping of the data. The clusters are noticeable at a higher height threshold, demonstrating the effectiveness of **Ward’s** method in minimizing intra-cluster variance.

In hierarchical clustering, the goal is to find clusters that are both interpretable and representative of the data’s natural structure. Based on the visual structure of the dendrogram and the distribution of elements in the figure 25, we can conclude that $k = 3$ is the better choice, capturing the main groupings in the data. While $k = 4$ does not add significant value as it results in an unbalanced partition with one cluster being too small to provide meaningful insights.

In our analysis, we determined that the optimal number of clusters differs depending on the clustering approach user. The **Ward’s** method suggests that the optimal number of clusters is 3, while **K-Means** method identifies 4 clusters as the optimal solution. As we can see the result of unhierarchical clustering methods in the appendix (figure 26).

For this specific analysis of **League of Legends** champions, where the goal is to balance interpretability with capturing distinct champion patterns, we have chosen the **3-cluster** solution from **Ward’s Method**. This allows for clearer insights into the unique roles and characteristics of champions while maintaining a balance between simplicity and explanatory power.

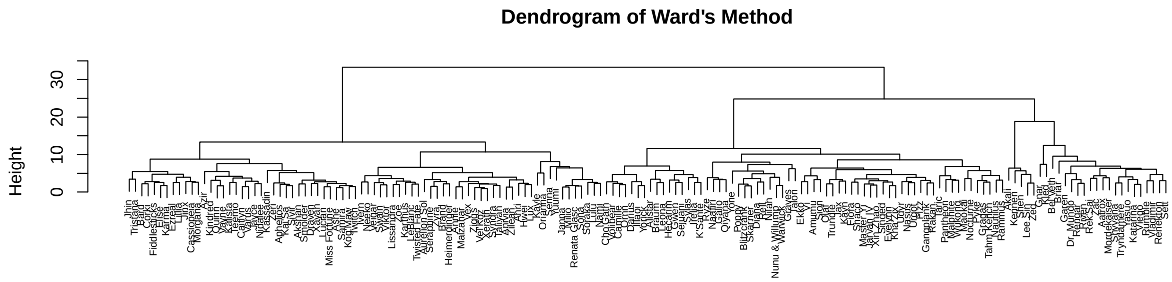


Figure 6: Ward's Dendrogram

6.2 Profiling Analysis

In the previous section, we selected **3 clusters** as the optimal number and chose PCA over MCA for cluster profiling. PCA provides tighter clusters (figure 7), making it easier to interpret, whereas MCA results in more dispersed clusters (figure 27), reducing clarity. Thus, PCA offers a more focused and centered clusters to analyze the behavior of the champions, even thought Clusters 1 and 3 has a few overlap, meaning some characteristics are similar.

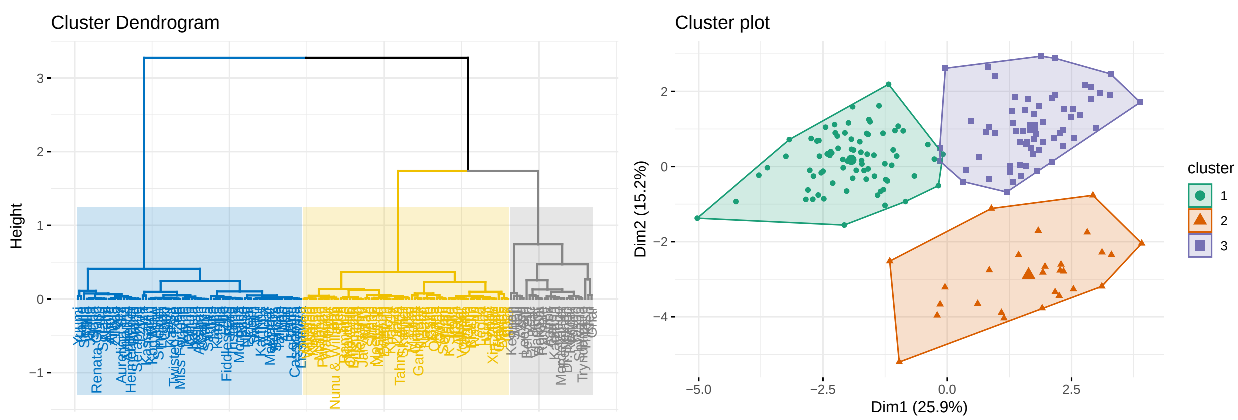


Figure 7: Profiling Analysis PCA

The most significant categorical variable for clusters are `Range.type` (p-value=4.194589e-28) and `Attack.range` (p-value=4.807783e-25). Variables `Tags` and `Role` are the least significant ones, indicating these variables are not relevant to distinct the clusters, as the first two variables. As we can observe the table 10 in the appendix section.

We see that Cluster 1 is characterized by **Long Range** and **Ranged** champions, since more than 90% of individuals in this cluster are strong associated with these two categories. Also, the dominant `Role` are **Marksman** and **Mage**, which related `Tags` are **Bottom** and **Middle** respectively, indicating a huge proportion of individuals with these categories are in this cluster, such as **Marksman** with Cla/Mod=96.428571 (96% of individuals that have this category are in Cluster 1) or **Mage** with Cla/Mod=88.571429. As we can see in the table 11.

Cluster 2 is related with all resource types that are not mana (mana is the basic type), such as **Nothing** (not provide any resource), **Energy** or **Fury**. This behavior is indicated by the high proportion of distinct resource types, where all observations without mana resource are in this cluster. Additionally, champions with **Fighter** tag, **Top** role and **Melee** range type are located in this cluster, suggesting 88% of members in this cluster are **Melee** and 66% of individuals are **Fighter**. The table 12 shows the information analyzed.

The most significant categories in Cluster 3 are **Melee** range type, **Tank** and **Fighter** tags, **Top** and **Jungle** roles, indicating this cluster is dominated by the champions with these characteristics, since that the 95% of individuals in this cluster are **Melee** and the proportion of **Tank** is high, with 87% of all champions. Additionally, more than 50% of total champions, who are categorized as **Fighter**, **Top** or **Jungle**, are located in this cluster. As we can observe in the table 13.

On the other hand, the most correlated numerical variables are `Magic.resistance.per.lvl`, `Mana.per.lvl`, and `Mana.regeneration.per.lvl`, indicating these variables are the most significant among the clusters with the highest Eta2 (0.84469719, 0.72972407, 0.71309699 respectively). While, the variables like `AS.ratio`, `HP.per.lvl` and `Attack.damage.per.lvl` are the less significant ones. As we can see in the table 14.

In Cluster 1, the most significant variables `Base.Mana`, `Mana.regeneration.per.lvl` and `HP.per.lvl`, with v.test values as 6.137196, 5.102791 and -2.368723 respectively. Moreover, we can affirm that the champions in this cluster have a high `Base.mana` and `Mana.regeneration.per.lvl`, since the mean value of these variables in Cluster 1 are higher than the overall mean (mean value of all cluster). In contrast, the mean of `Base.armor` or `Magic.resitance.per.lvl` of this cluster are lower than the overall mean, indicating the weak capacity of durable. We need to highlight that these characteristics correspond to **Mage** or **Marksman** champions, as we affirmed previously that Cluster 1 is dominated by **Mage** and **Marksman**. As we can observe in the table 15.

Oppositely, Cluster 2 is characterized by `Magic.resistance.per.lvl`, `Base.armor` and `Attack.damage`, affirmed by v.test and mean value of these variables in the cluster. Since the mean of variables `Magic.resistance.per.lvl`, `Base.armor` and `Attack.damage` are higher than the overall mean, such as the mean value of `Magic.resistance.per.lvl` is 341.666667 and it overall mean is 313.5868263. Additionally, the mean values of `Mana.per.lvl` and `Mana.regeneration.per.lvl` are 0, affirming the analysis of previous part, where Cluster 2 is dominated by champions without mana resource. Combining the both analysis, we can approve that the champions in Cluster 2 have high durability and damage without mana resource, which are **Fighter** and **Top** champions. As shown in the table 16.

Lastly, the most meaningful variables of Cluster 3 are `Mana.regeneration.per.lvl`, `Base.armor`, `HP.regeneration`, `Attack.damage` and `Base.magic.resistance`, and the mean value of these variables also are higher than the corresponding overall mean value, indicating this cluster groups champions with high durability and damage. These patterns are similar to Cluster 2, but the champions in this cluster have mana resource type, different to Cluster 2.

Combining the both informations, we can mentioned that Cluster 3 is dominated by champions with mana resource and high durability and attack damage, which are **Tank** and **Fighter** champions with **Top** and **Jungle** tags. Moreover, we observe that all the associated variables with this cluster have mean higher than the overall mean, suggesting the champions have a good statistics compare to the other clusters. As we can see the table 17 in the appendix.

In conclusion, the profiling analysis explains well the significant patterns of each cluster. Cluster 1 exhibits champions with high mana capacity, with v.test values showing over-representation, it is strongly associated with **Mage**, **Marksman** and **Long Range** categories, reflecting high mana capacity patterns. Cluster 2 demonstrates high durability and damage, which are strongly linked with no mana resource, **Fighter** and **Top** champions. Cluster 3, while associated with only mana resource and melee champions, is characterized by extreme durability and damage of champions with **Tank** or **Fighter** roles in **Top** or **Jungle** tags.

7 Discriminant Analysis

7.1 Hotelling T^2 Test

The purpose of the Hotelling T^2 test is to analyze the variance between the groups within the categorical variable `Range.type`, as it is the only variable suitable for this method (two distinct groups).

We selected numerical variables based on their normality using `Shapiro-Wilk` and `Kolmogorov-Smirnov` tests, and plots(`histogram` and `Q-Q plot`). The `Shapiro-Wilk test` (table 18), which is sensitive to small deviations, indicates that none of the variables followed a normal distribution. However, the `Kolmogorov-Smirnov test` (table 19) and plots (figure 28) shows that some variables (`Base.armor`, `HP.regeneration`, `Attack.damage`, `Attack.speed.per.lvl`) have a significant distribution indicated by p-values (less than 0.05). Moreover, observing at plots, we can affirm that some variables could still be considered normally distributed. Both tables present a selection of numerical variables that are the most significant for evaluating this assumption.

Additionally, the p-value obtained by `Kolmogorov-Smirnov test` for the variable `Base.HP` was close to 0.05, so we decided to apply a transformation to this variable, which resulted in significant change in its distribution (table 20). As we can see in the figure 29, exists some residual outliers in the extreme quantiles in the `Q-Q` plot, which deviate from the diagonal line.

Using `boxM()` function, we can observe that the null hypothesis of the homogeneity on variance is rejected (p-value=2.26e-05), indicating that the covariance matrices are not equal, suggesting unequal variance across the groups. As we can observe in the table 21.

After checking the variables, we use Hotelling's T^2 test to compare the multivariate means of the selected variables `Base.HP`, `Base.armor`, `HP.regeneration`, `Attack.damage`, and `Attack.speed.per.lvl` between the **Melee** and **Ranged** groups in `Range.type` variable (table 22). The test resulted in a T^2 statistic of approximately 261.9775 with a p-value of $<$ 2.2e-16. This indicates a statistically significant difference between the **Melee** and **Ranged** groups at the multivariate level for these variables.

Specifically, this suggests that **Melee** and **Ranged** champions differ significantly in their average values for `Base.HP`, `Base.armor`, `HP.regeneration`, `Attack.damage`, and `Attack.speed.per.lvl`. This means that the **Melee** group likely prioritizes higher durability and regeneration attributes (e.g., `Base.armor` and `HP.regeneration`), while the **Ranged** group may emphasize offensive and scaling attributes (e.g., `Attack.damage` and `Attack.speed.per.lvl`). These differences reflect distinct gameplay roles and design philosophies for melee versus ranged champions in **League of Legends**.

7.2 MANOVA

We aim to determine whether the variation observed in the numerical variables can be attributed to the different levels of the categorical variables, or if the differences between the groups are negligible. We take the levels of categorical variables `Role` and `Tags` to test the mean of the numerical variables, the ones that we are checked in the previous subsection, which have normal distribution. Therefore, we want to determine the similar variables between the different roles or tags, observing the common characteristics that might have between the champions.

Using the `boxM()` function, we observed that not all numerical variables selected have the same variance across the groups defined by `Role` and `Tags`, suggesting that at least one group exhibits significantly different variance (p-value $<$ 0.05). Further analysis revealed that the variable `Base.armor` has a distinctly different variance when grouped by `Tags`, indicating heterogeneity among these groups. In contrast, within the `Role` group, only two numerical variables, `Attack.damage` and `Attack.speed.per.lvl`, demonstrated consistent variance across all groups, reflecting homogeneity in this case. We could see the test results in the table 23.

The `Role` group contains only two numerical variables, making it unnecessary to perform a `MANOVA` test, as testing mean equality with just two variables is less informative. Instead, these variables can be used directly in `discriminant analysis`. On the other hand, the `Tags` group, it is necessary to perform some `MANOVA` tests to determine if the selected numerical variables have significantly different mean across the groups, ensuring that the final selected variables contribute meaningfully to the analysis and help to identify the distinct groups.

We only apply `MANOVA` tests to `Tags` group (table 24), we can observe that all numerical variables have significant effect on `Tags` groups, showing a p-value less than 0.05. Indicating that none group has the same mean in these numerical variables. However, the variable `Attack.speed.per.lvl` exhibits a p-value of 0.009879, which is less significant that the other variables, indicating that this variable might have less variability across the `Tags` groups compared to the other variables.

Using `Tukey's HSD test`, we compared the means of each group for numerical variables (those that complete the assumptions). The results show that the variables `Base.HP`, `HP.regeneration` and `Attack.damage` have several groups with significantly different means, as indicate p-values less than 0.05 (table 25, 26 and 27). For example, in the variable `Base.HP`, the `Mage` and `Tank` groups show a meaningful difference in their means, suggesting that champions with the `Tank` tag tend to have significantly higher base health compare to those with the `Mage` tag.

In contrasts, for the variable `Attack.speed.per.lvl` (table 28), the mean differences between most groups are not significantly. Only the mean difference between the `Mage` and `Marksman` groups shows a slightly significant difference, with p-value of 0.0234610. This indicates that `Attack.speed.per.lvl` has limited utility to distinct the groups, as the differences in means for most groups in `Tags` are not meaningful.

After analyzing the `MANOVA` test for the different categorical variables, we retained `Base.HP`, `HP.regeneration` and `Attack.damage` as the most significant variables for distinguishing observations based on `Tags`, given their strong differences across groups. These will be used in `discriminant analysis` to differentiate champions by `Tags`. For the variable `Role`, `Attack.damage` and `Attack.speed.per.lvl` are identified as significant variables and will be used in **discriminant analysis**.

7.3 Discriminant Analysis

In the previous section, we discussed two possible categorical variables, `Tags` and `Role`, that can be analyzed using discriminant analysis. Both variables have more than five levels, making them suitable for this technique. However, for the purpose of this section, we will focus on the `Tags` variable. While both variables are interesting to explore, choosing one does not significantly affect the insights gained.

In this section, we aim to apply discriminant analysis to identify the features that are most predictive of a champion's tags. This will help us better understand how numerical attributes influence champion roles and classes. Such analysis not only aids in categorizing existing champions but also provides valuable insights for predicting the classification of newly introduced champions. For example, if a new champion is introduced with specific attribute values, the discriminant model can assign the champion to its most likely Tag with a quantifiable level of accuracy.

We first performed **Linear Discriminant Analysis** (LDA) for the target variable `Tags`, using the predictors identified in the **MANOVA** section. This process involved carefully selecting the best candidate predictors for the LDA model. However, the results were not particularly strong, with a correct classification rate of approximately **56%**. While this accuracy is not very high, it still provided some insights into the relationships between attributes and the `Tags` variable.

Given the limitations of LDA, we applied **Quadratic Discriminant Analysis** (QDA) to the same target variable. Unlike LDA, QDA does not assume homogeneity of variance across groups, making it suitable when this assumption is violated. The predictors used for QDA included attributes such as `Base.HP`, `HP.regeneration`, `Attack.damage`, `Attack.speed.per.lvl`, and `Base.armor`.

We verified the lack of homogeneity of variance for these attributes using the `boxM()` function (table 29). The results showed a p-value significantly smaller than 0.05, leading us to reject the null hypothesis and conclude that these attributes do not exhibit homogeneity of variance. This made them appropriate for QDA, which is designed to handle such cases.

In the figure 30, we got the prior probabilities for each class, representing the proportion of observations in each class before considering any predictor variables. For example, approximately 21% of the dataset belongs to the **Mage** class, while around 28% corresponds to the **Fighter** class, which has the largest proportion in the dataset.

Next, we examined the mean values of the predictors for each class (table 30). For instance, the **Marksman** class has the highest mean value for `Attack.speed.per.lvl`, suggesting that champions with higher values for this attribute are more likely to belong to the **Marksman** class. Similarly, when looking at the mean of `Base.armor`, we observe that the **Fighter** class has a mean value of 34.28, while the **Tank** class has a mean of 36.12. This indicates that champions with higher `Base.armor` values are more likely to be classified as **Tank** or **Fighter**.

The table 31 presents the confusion matrix from our QDA model. In this matrix, the rows represent the predicted classes from the QDA model, while the columns correspond to the actual classes in the dataset. From the table, we can observe that most observations are classified correctly looking at diagonal line. Let's calculate the percentage of correct classifications to evaluate the model's performance.

Finally, we achieved a correct classification rate of approximately **70%** with the QDA model (figure 31), which is a significant improvement compared to the **56%** obtained with the LDA model (figure 32). Additionally, the class with the highest percentage of correct classifications is **Support**, at around **81%**. This could be attributed to the fact that the predictors used in the analysis show distinct mean values for the **Support** class, making it easier to differentiate from other classes.

On the other hand, the **Tank** class has the lowest percentage of correct classifications. Examining the confusion matrix, we observe that five observations from the **Tank** class were misclassified as **Fighter**. This is not surprising, as we previously noted that several predictors have similar mean values for **Tank** and **Fighter** classes, leading to challenges in distinguishing between them. A similar issue arises when classifying **Fighter** and **Tank** classes due to their shared characteristics.

The graph below (figure 8) illustrates the predicted versus actual classes. It provides a visual representation of how observations are distributed across the different classes in `Tags`. The comparison between actual and predicted classes provides insight into how well the QDA model performs and areas where it faces challenges.

From the graph, it is evident that the QDA model performs differently across classes. The **Fighter** class is the one that got the highest frequency, both in actual and predicted values because as we know from the previous explanations the 28% of the dataset are from the class **Fighter**. For the **Mage** and **Marksman** classes, the predictions align relatively well with the actual classes, indicating that the model captures their distinct features reasonably well. However, there are still some instances of misclassification, such as predictions of **Fighter** or other classes for these actual categories.

The classes **Assassin**, and **Tank** display notable misclassifications. Predictions for these classes are dispersed across multiple categories, particularly **Fighter** for **Tank** class. This dispersion suggests that the QDA model struggles to clearly distinguish these classes, likely due to overlapping feature distributions or insufficient representation in the dataset.

Finally, the QDA model demonstrates varying success in classifying observations based on different predictor pairs. Some combinations (figure 33), such as `HP.regeneration` vs. `Base.armor`, result in lower error rates (0.377). Thus, it plays a crucial role in distinguishing classes in this case, as they contribute to clearer decision boundaries and lower error rates. While others, like `Base.HP` and `HP.regeneration`, show higher error rates (0.521), reflecting less effective classification.

Significant overlap is observed between certain classes, such as **Fighter** and **Tank**, which share similar predictor values. This overlap makes it challenging for the model to distinguish these classes, resulting in higher misclassification rates. Conversely, classes like **Support** exhibit better separation due to distinct predictor means, leading to higher classification accuracy.

In conclusion, the application of QDA demonstrates its potential for predicting the `Tags` variable with a reasonable degree of accuracy. The analysis underscores the importance of understanding predictor distributions and ensuring balanced representation across classes.

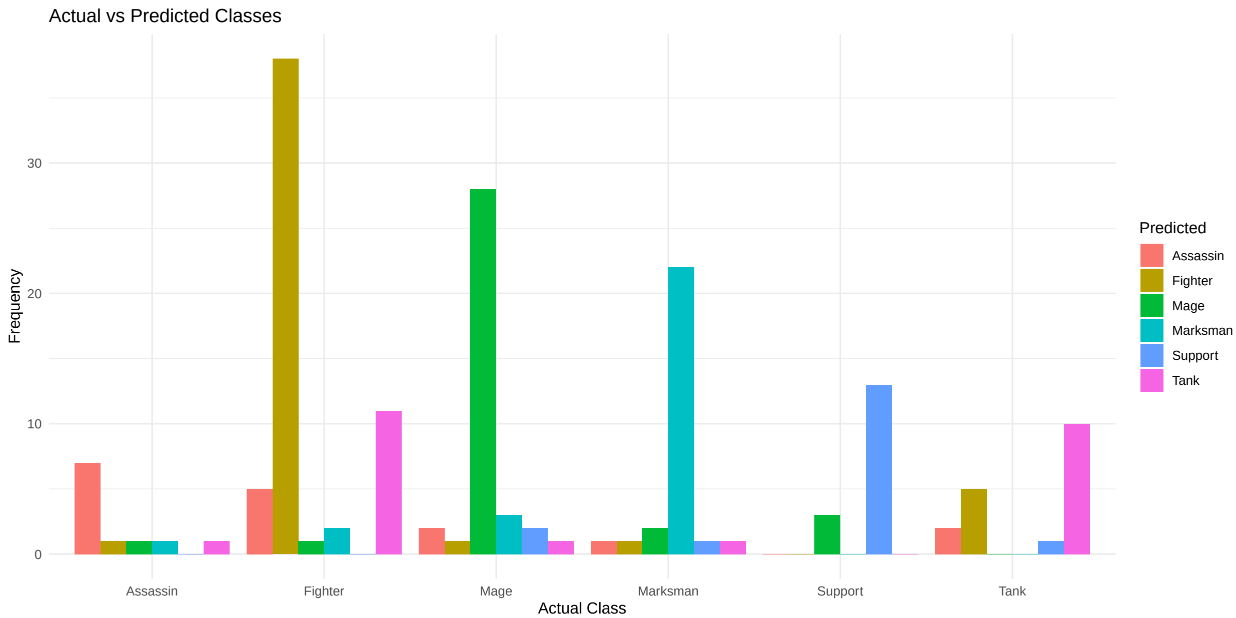


Figure 8: QDA Actual vs Predicted Classes

8 Conclusion

Once analyzed all the analytical methods proposed, we will present the relevant results in a comprehensive summary, which reflects the objective mentioned in each section.

- **PCA:** The significant dimensions are the first two, as they capture most of the variance. The first dimension presents the champions with powerful durability and damage capacities, such as "Tryndamere", "Darius", and "Illaoi". On the other hand, the dimension 2 is dominated by champions as "Sylas", "Ryze", and "Naafiri", suggesting a potentially mana growth statistics. Using professional knowledge about the game, we could affirm that the champions correlated in the first dimension corresponding to the role **Top**, and the role **Middle** for the second dimension.

- **MDS:** The two dimensional coordinates shows different clusters group by several common characteristics, indicating the champions within the same cluster have similar game statistics. The champions are grouped principally by their tags, suggesting the champions with same tag have common statistic behaviors, such as "Shyvana" and "Lee Sin" have similar statistics on `Base.armor`, `HP.regeneration`, and `Attack.Damage`.

- **MCA:** The significant dimensions reveal distinct patterns, where Dimension 1 captures variations in attack range (ranged vs melee), and Dimension 2 highlights differences in tags related to range type. Combining these behaviors, the dimensions distinguish champions with tags like **Marksman**, **Mage**, which have **Long Range**, from **Fighter** and **Tank**, which are melee. For example, it differentiates "Jinx" (**Marksman** and **Long Range**) from "Illaoi" (**Fighter** and **Close Range**), suggesting that their statistics cannot be directly compared as their roles and tags are fundamentally different.

- **Cluster Analysis:** The optimal number of cluster is 3 using Ward's method. Cluster 1 is associated with high mana capacity, strongly linked to **Mage**, **Marksman**, and **Long Range** categories. Cluster 2 highlights champions with high durability and damage, tied to no mana resource, **Fighter**, and **Top** roles. Cluster 3 represents extreme durability and damage, regardless of resource type, associated with melee champions in **Tank** or **Fighter** roles, principally in **Top** or **Jungle** tags.

- **Discriminant Analysis:** Particularly using Quadratic Discriminant Analysis (QDA), demonstrated its utility in classifying champions based on their statistical attributes. While the model showed varying levels of accuracy depending on the predictor pairs, it was particularly effective in distinguishing classes with distinct means, such as **Support**, while facing challenges in separating overlapping classes like **Fighter** and **Tank**. This underscores the importance of understanding the distribution of predictors and the limitations of classification models in contexts of high overlap.

In conclusion, the combination of these analytical techniques provided a holistic view of the dataset, uncovering meaningful patterns and relationships. The findings not only enhance our understanding of champion characteristics but also demonstrate the potential of statistical and computational methods in deriving actionable insights from the datasets.

9 Bibliography

[1] Kassambara. (2017-09-23). PCA - Principal Component Analysis Essentials. Retrieved November 25, 2024, from https://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/.

[2] Kassambara. (2017-09-24). MCA - Multiple Correspondence Analysis in R: Essentials. Retrieved November 27, 2024, from https://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/114-mca-multiple-correspondence-analysis-in-r-essentials/.

[3] Kassambara. (2017-09-25). HCPC - Hierarchical Clustering on Principal Components: Essentials. Retrieved December 2, 2024, from https://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/117-hcpc-hierarchical-clustering-on-principal-components-essentials/.

10 Appendix

This appendix shows all referenced figures and tables in the previous sections.

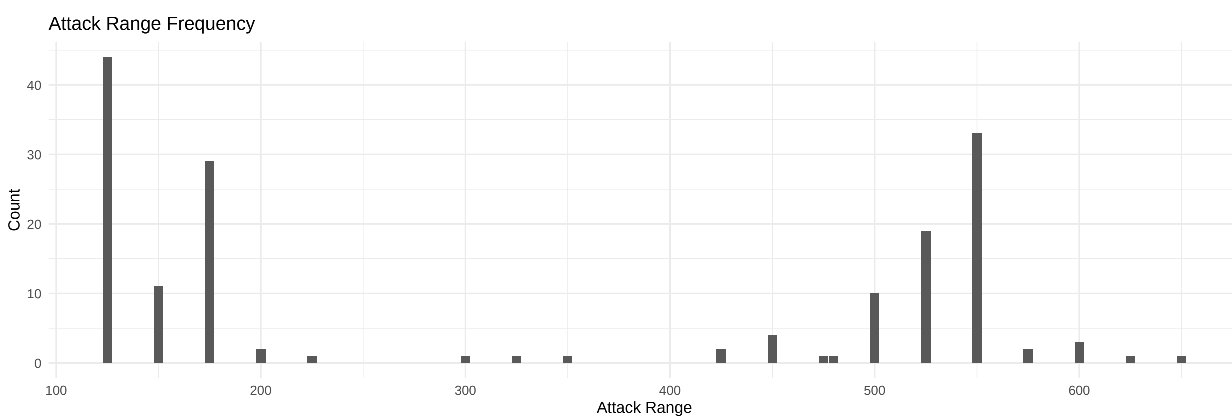


Figure 9: Attack Range Histogram

Table 1: Shapiro Wilk Normality Test for Attack Range

Variable	Statistic	P_value
Attack.range	0.7627058	4e-15

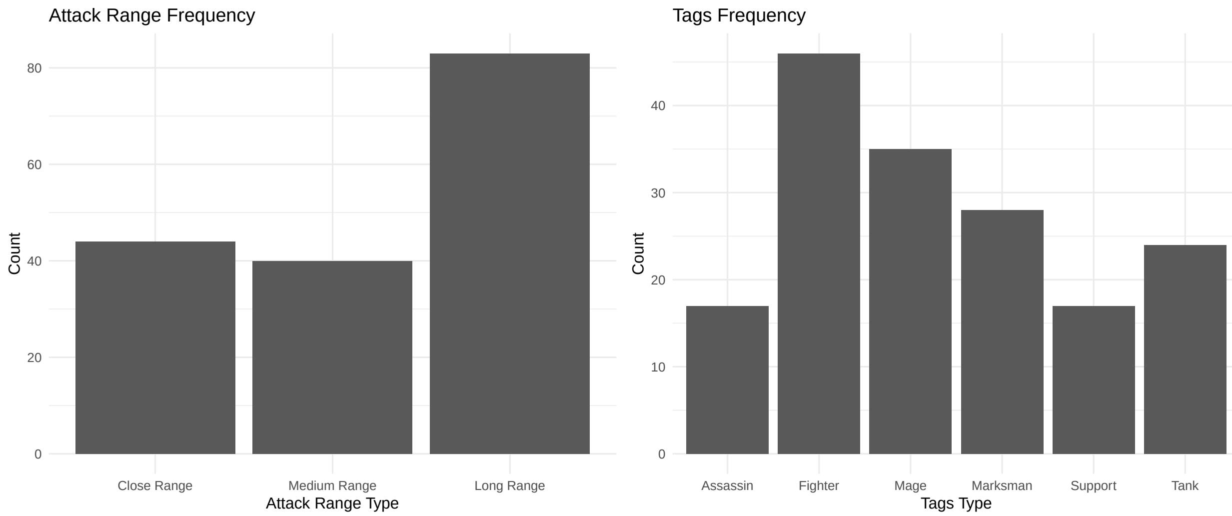


Figure 10: Frequency Plots

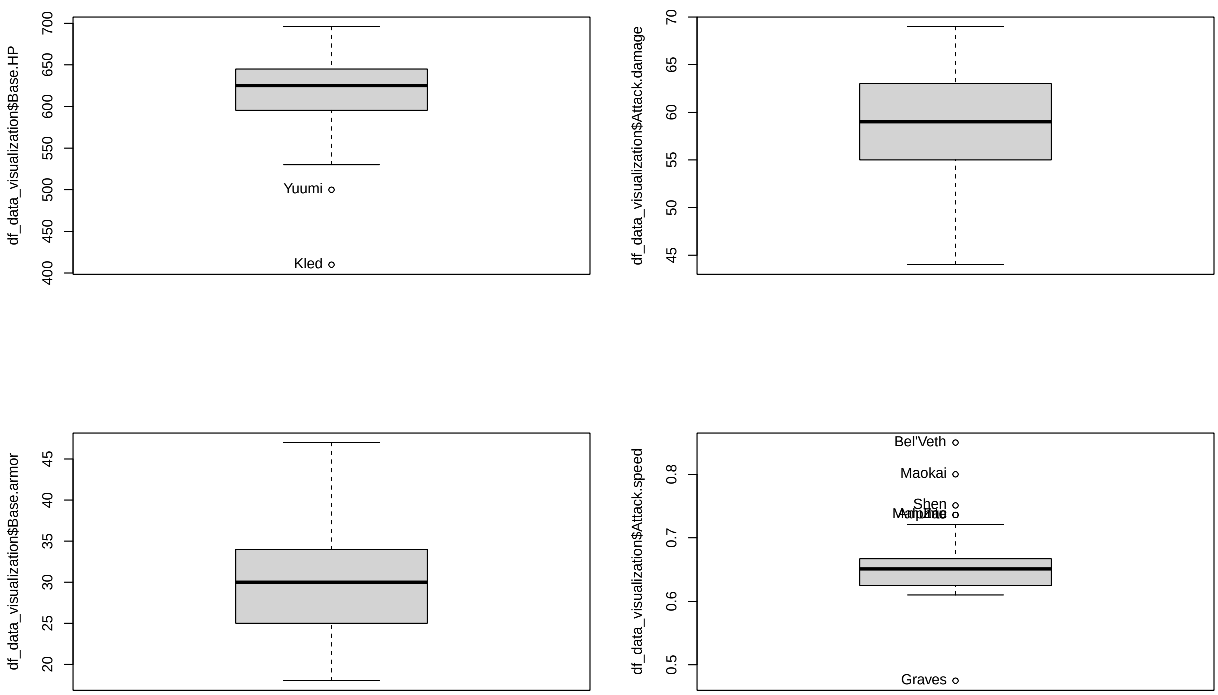


Figure 11: Relevant Boxplots

Table 2: Proportional Table: Tags vs Role

	Bottom	**Jungle**	**Middle**	**Support**	**Top**
Assassin	0.0000000	0.4705882	0.4705882	0.0000000	0.0588235
Fighter	0.0217391	0.3695652	0.0000000	0.0000000	0.6086957
Mage	0.0285714	0.1142857	0.6571429	0.1428571	0.0571429
Marksman	0.7500000	0.0714286	0.0714286	0.0000000	0.1071429
Support	0.1176471	0.0588235	0.0000000	0.8235294	0.0000000
Tank	0.0000000	0.2500000	0.0416667	0.2916667	0.4166667

Table 3: Proportional Table: Role vs Attack.range

	Close Range	**Medium Range**	**Long Range**
Bottom	0.0000000	0.000	0.3012048
Jungle	0.3409091	0.300	0.1325301
Middle	0.0909091	0.125	0.3012048
Support	0.1136364	0.100	0.2048193
Top	0.4545455	0.475	0.0602410

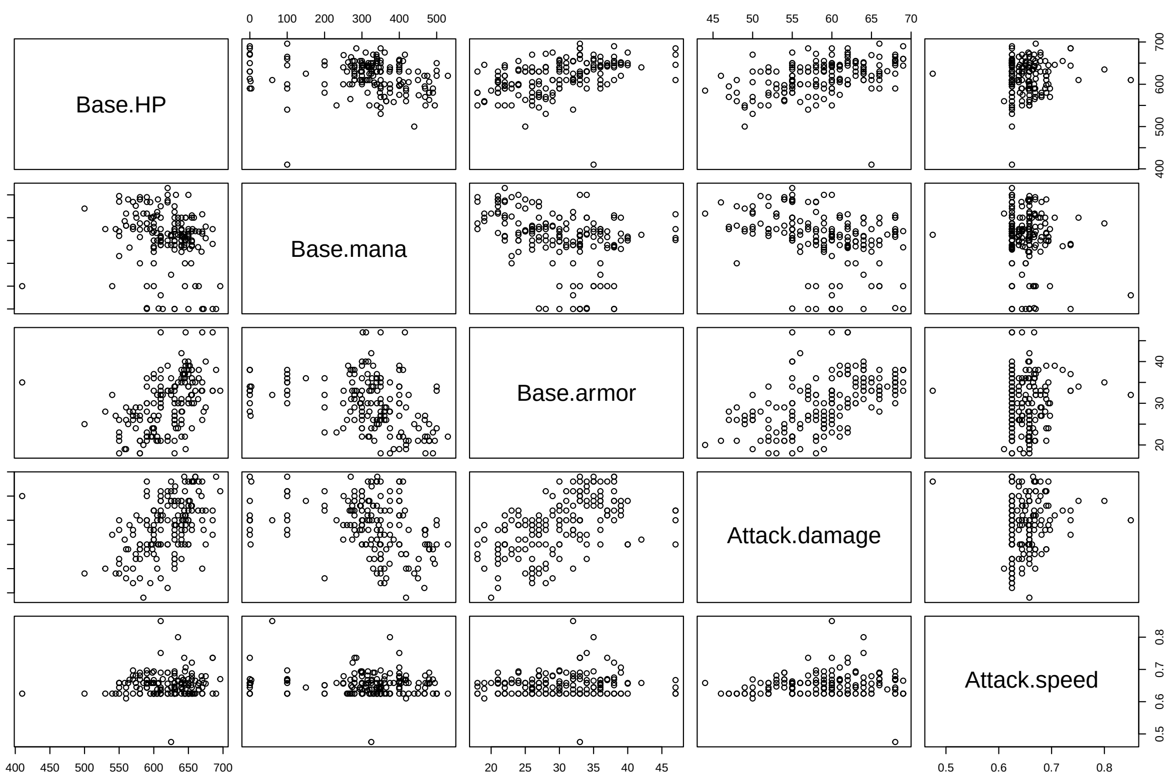


Figure 12: Correlation Plots

Table 4: Correlation Table

	Base.HP	**Base.mana**	**Base.armor**	**Attack.damage**	**Attack.speed**
Base.HP	1.0000000	-0.1916291	0.4471465	0.4368721	0.0939491
Base.mana	-0.1916291	1.0000000	-0.3694285	-0.3427977	-0.1052514
Base.armor	0.4471465	-0.3694285	1.0000000	0.5334466	0.0809114
Attack.damage	0.4368721	-0.3427977	0.5334466	1.0000000	0.1284963
Attack.speed	0.0939491	-0.1052514	0.0809114	0.1284963	1.0000000

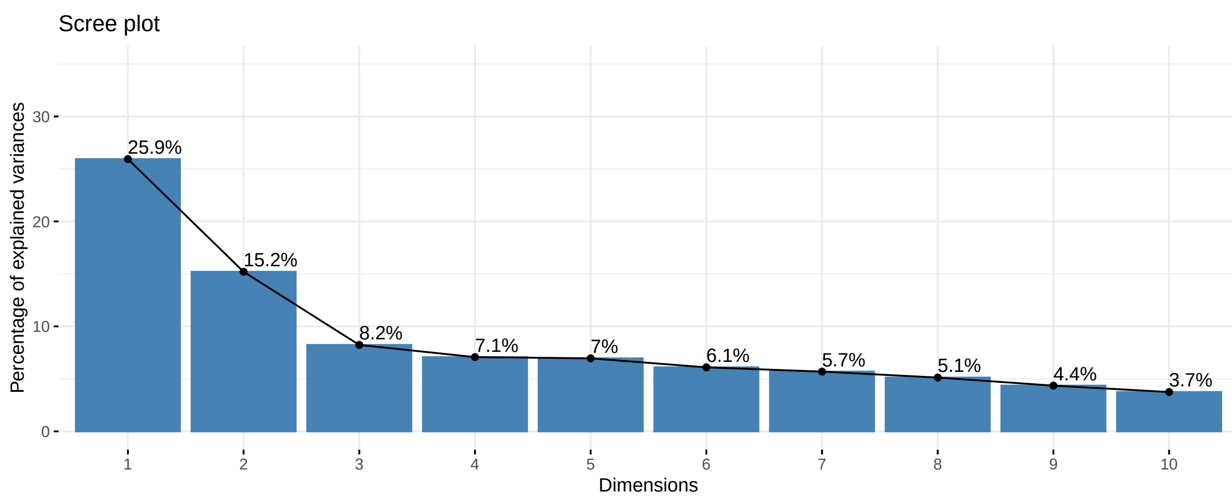


Figure 13: PCA Scree Plot

Table 5: Eigenvalue of PCA Dimensions

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	4.1496368	25.935230	25.93523
comp 2	2.4319778	15.199861	41.13509
comp 3	1.3169784	8.231115	49.36621
comp 4	1.1317108	7.073193	56.43940
comp 5	1.1127268	6.954543	63.39394
comp 6	0.9755733	6.097333	69.49128
comp 7	0.9104642	5.690402	75.18168
comp 8	0.8198917	5.124323	80.30600
comp 9	0.6971407	4.357129	84.66313
comp 10	0.5993390	3.745869	88.40900

Table 6: MDS Table 1

	Tags	Role	Range.type	Resourse.type	Base.HP	Base.mana	Base.armor	Base.magic.resistance	Attack.range	HP.regeneration	Mana.regeneration	Attack.damage	Attack.speed
Shyvana	Fighter	Jungle	Melee	Fury	665	100	38	32	Close Range	8.5	0.0	66	0.658
Lee Sin	Fighter	Jungle	Melee	Energy	645	200	36	32	Close Range	7.5	50.0	66	0.651
Senna	Support	Bottom	Ranged	Mana	530	350	28	30	Long Range	3.5	11.5	50	0.625

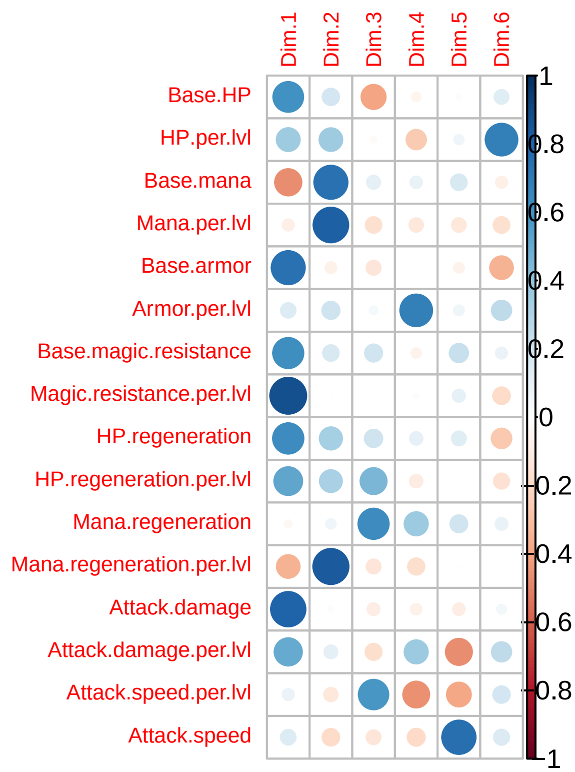


Figure 14: PCA Correlation Plot

Table 7: MDS Table 2

	Tags	Role	Range.type	Resourse.type	Base.HP	Base.mana	Base.armor	Base.magic.resistance	Attack.range	HP.regeneration	Mana.regeneration	Attack.damage	Attack.speed
Gnar	Fighter	Top	Ranged	Rage	540	100	32	30	Medium Range	4.5	0.0	57	0.625
Alistar	Tank	Support	Melee	Mana	685	350	47	32	Close Range	8.5	8.5	62	0.625
Aatrox	Fighter	Top	Melee	Blood Well	650	0	38	32	Medium Range	3.0	0.0	60	0.651

Table 8: MDS Table 3

	Tags	Role	Range.type	Resourse.type	Base.HP	Base.mana	Base.armor	Base.magic.resistance	Attack.range	HP.regeneration	Mana.regeneration	Attack.damage	Attack.speed
Aphelios	Marksman	Bottom	Ranged	Mana	580	348	26	30	Long Range	3.25	6.50	55	0.640
Draven	Marksman	Bottom	Ranged	Mana	675	361	29	30	Long Range	3.75	8.05	62	0.679
Azir	Mage	Middle	Ranged	Mana	550	320	22	30	Long Range	3.50	8.00	52	0.625
Corki	Marksman	Middle	Ranged	Mana	640	350	30	30	Long Range	5.50	7.40	52	0.644
Ahri	Mage	Middle	Ranged	Mana	590	418	21	30	Long Range	2.50	8.00	53	0.668
Zeri	Marksman	Bottom	Ranged	Mana	600	250	24	30	Long Range	3.25	6.00	56	0.658
Ziggs	Mage	Bottom	Ranged	Mana	606	480	21	30	Long Range	6.50	8.00	55	0.656

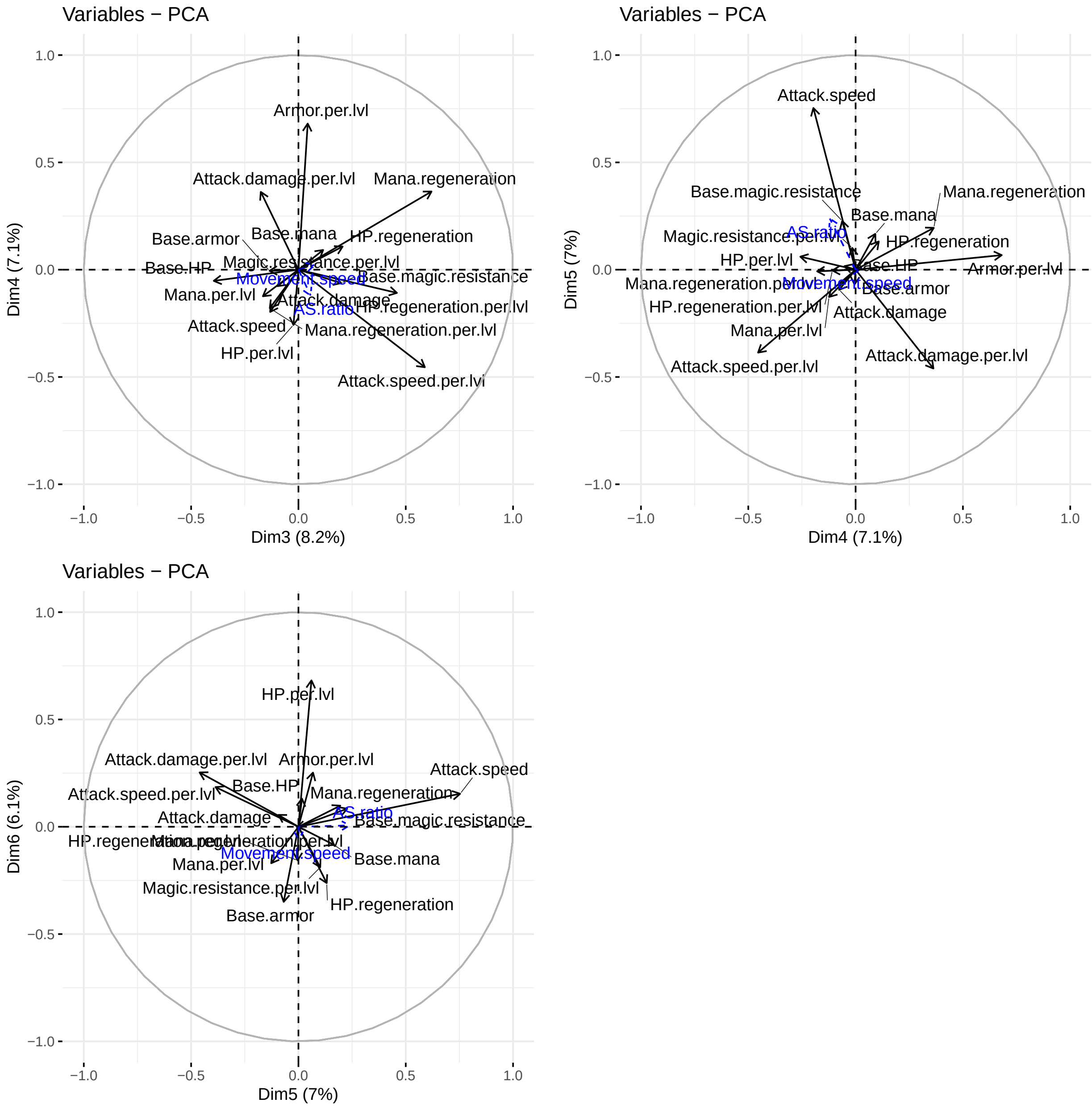


Figure 15: PCA Variables Dimensions 3, 4 and 5

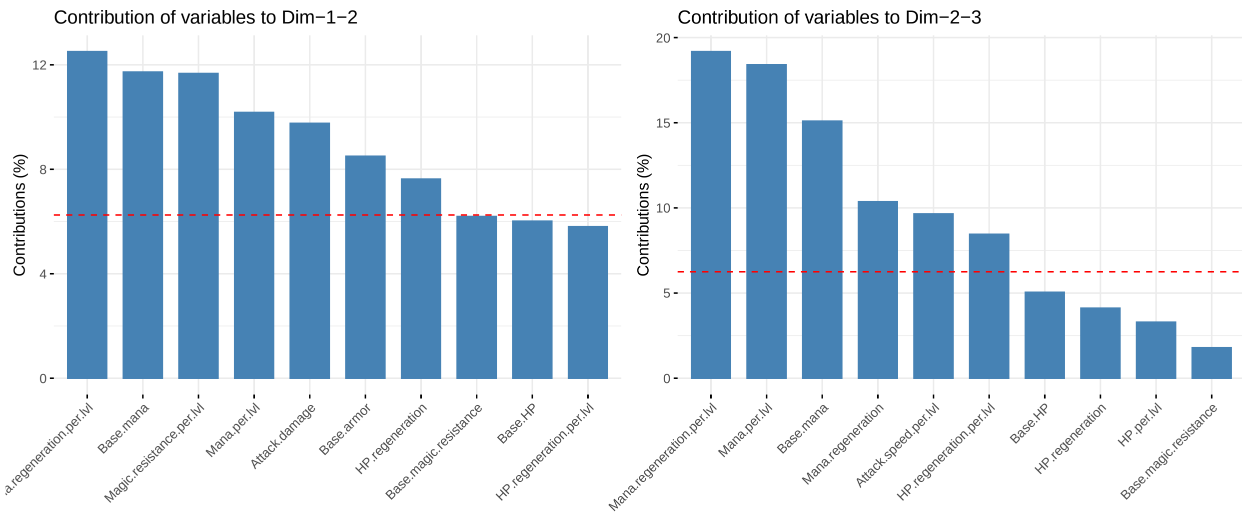


Figure 16: Contribution Histograms in Dimensions 1, 2 and 3

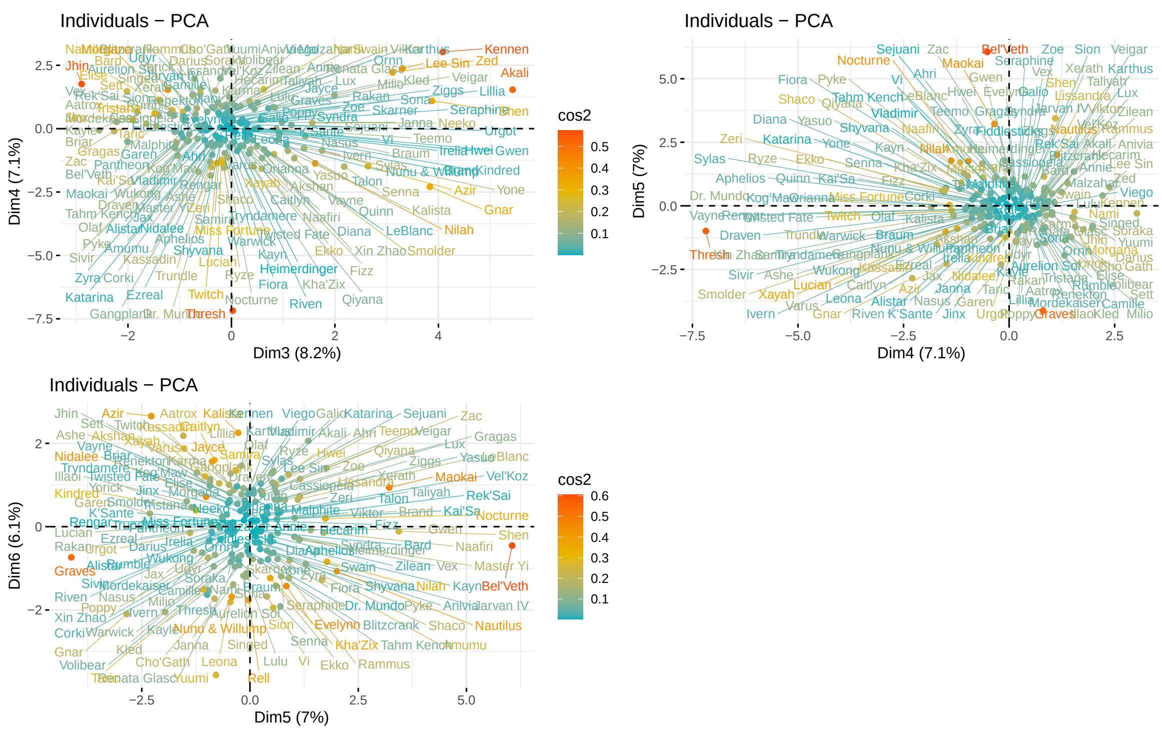


Figure 17: PCA Individuals Dimensions 3, 4 and 5

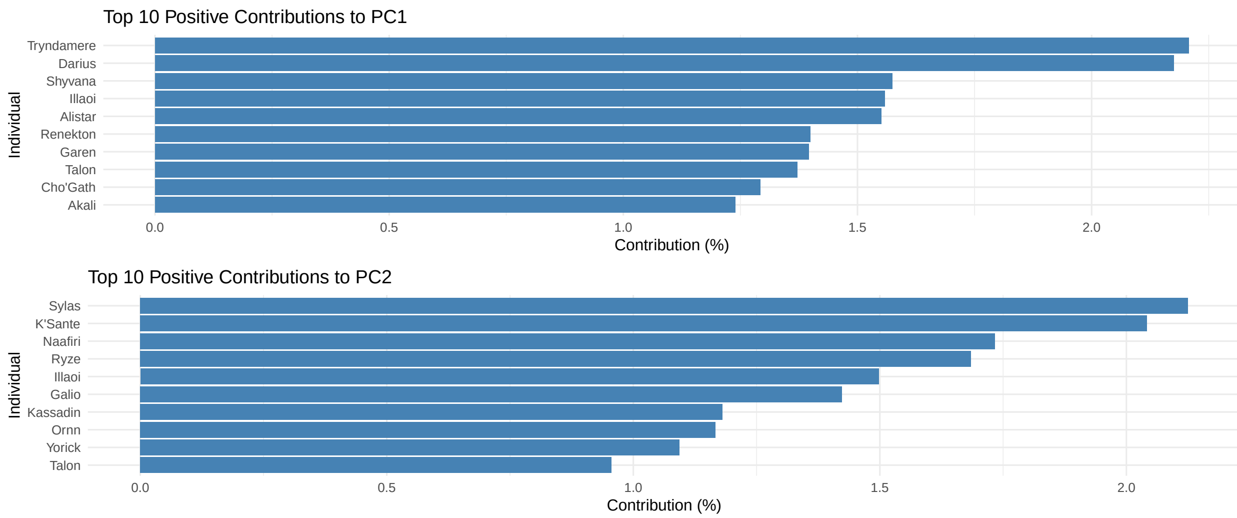


Figure 18: PCA Individuals Contributions

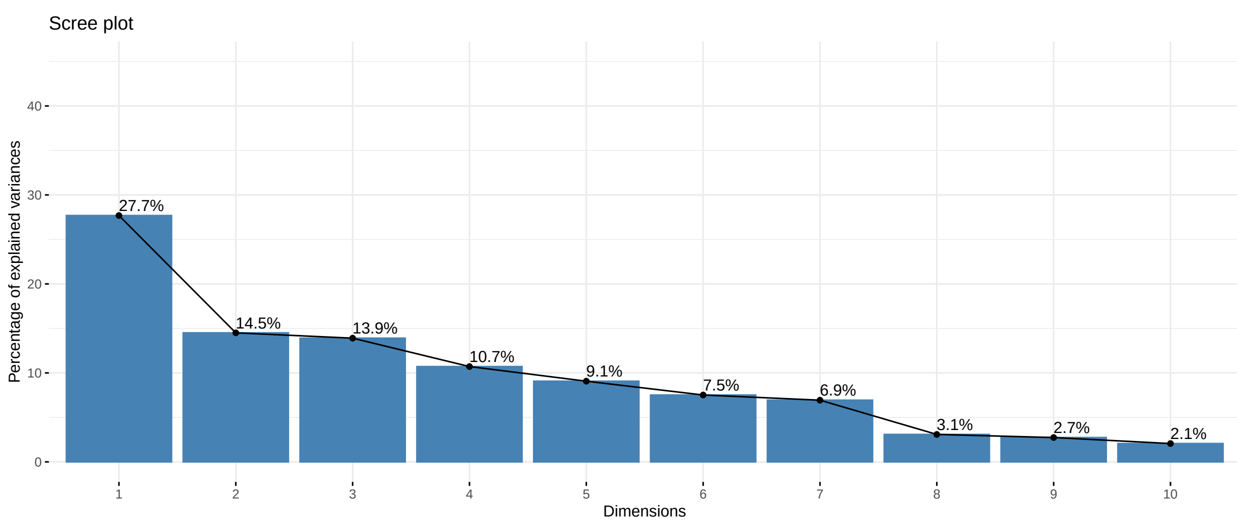


Figure 19: MCA Scree Plot

Table 9: Eigenvalue of MCA Dimensions

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.8305550	27.6851679	27.68517
dim 2	0.4350850	14.5028321	42.18800
dim 3	0.4171234	13.9041122	56.09211
dim 4	0.3214341	10.7144688	66.80658
dim 5	0.2720741	9.0691359	75.87572
dim 6	0.2255396	7.5179860	83.39370
dim 7	0.2080705	6.9356836	90.32939
dim 8	0.0925549	3.0851629	93.41455
dim 9	0.0821151	2.7371693	96.15172
dim 10	0.0618797	2.0626561	98.21437
dim 11	0.0375412	1.2513730	99.46575
dim 12	0.0160276	0.5342522	100.00000

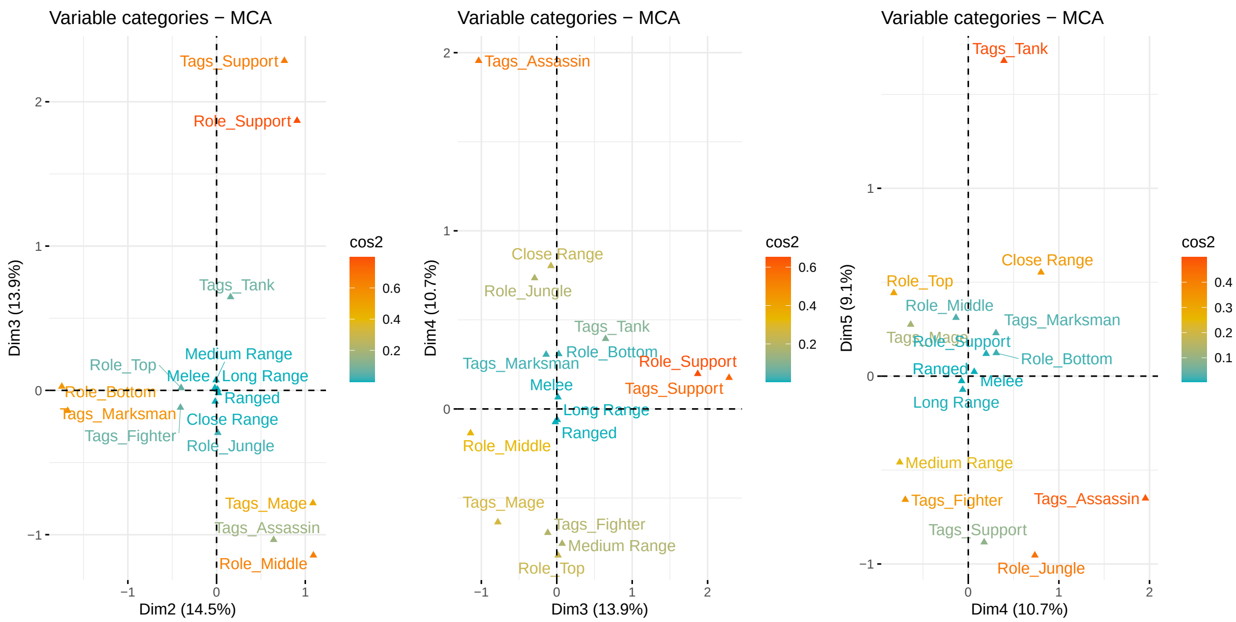


Figure 20: MCA Variables Dimensions 2, 3, 4 and 5

Table 10: Categorical variables profiling

	p.value	df
Range.type	4.000000e-28	2
Attack.range	4.808000e-25	4
Resourse.type	2.159110e-23	24
Tags	3.574060e-23	10
Role	1.875327e-12	8

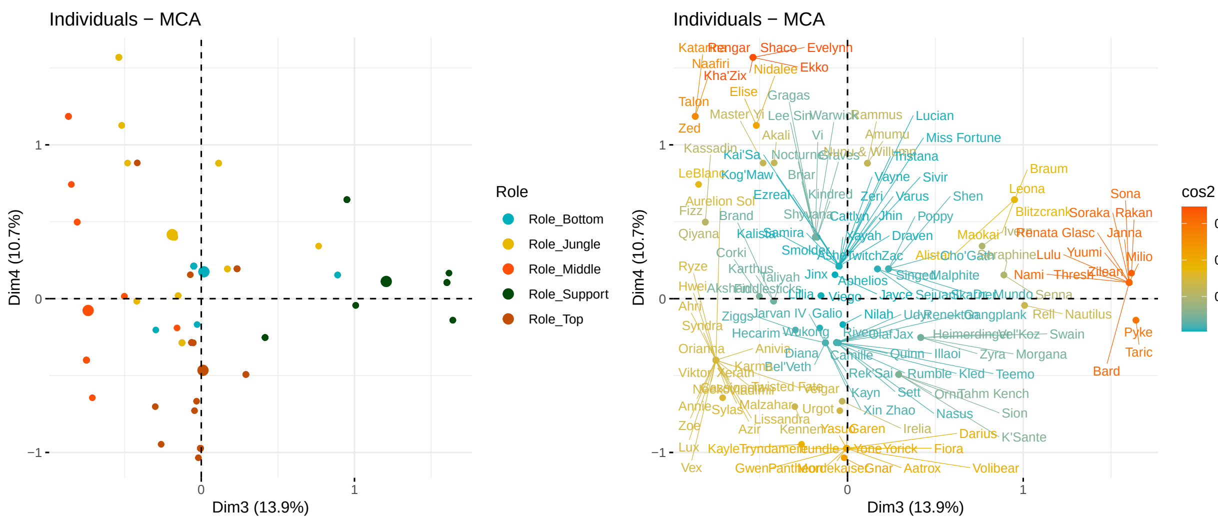


Figure 21: MCA Individuals Dimensions 3 and 4

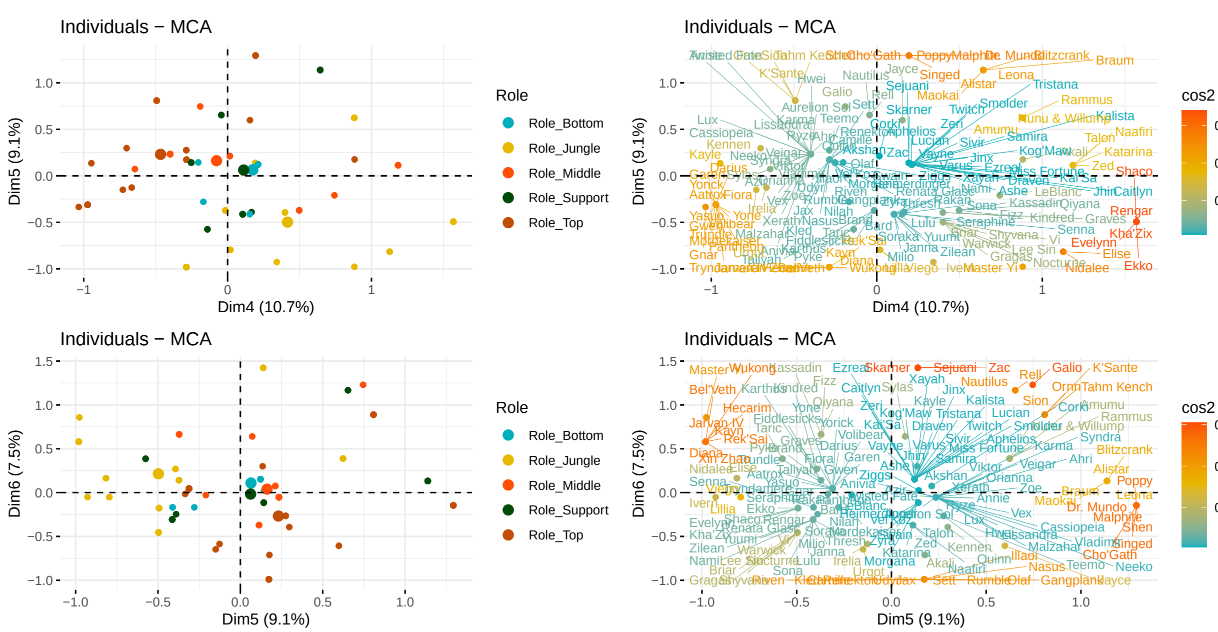


Figure 22: MCA Individuals Dimensions 4, 5 and 6

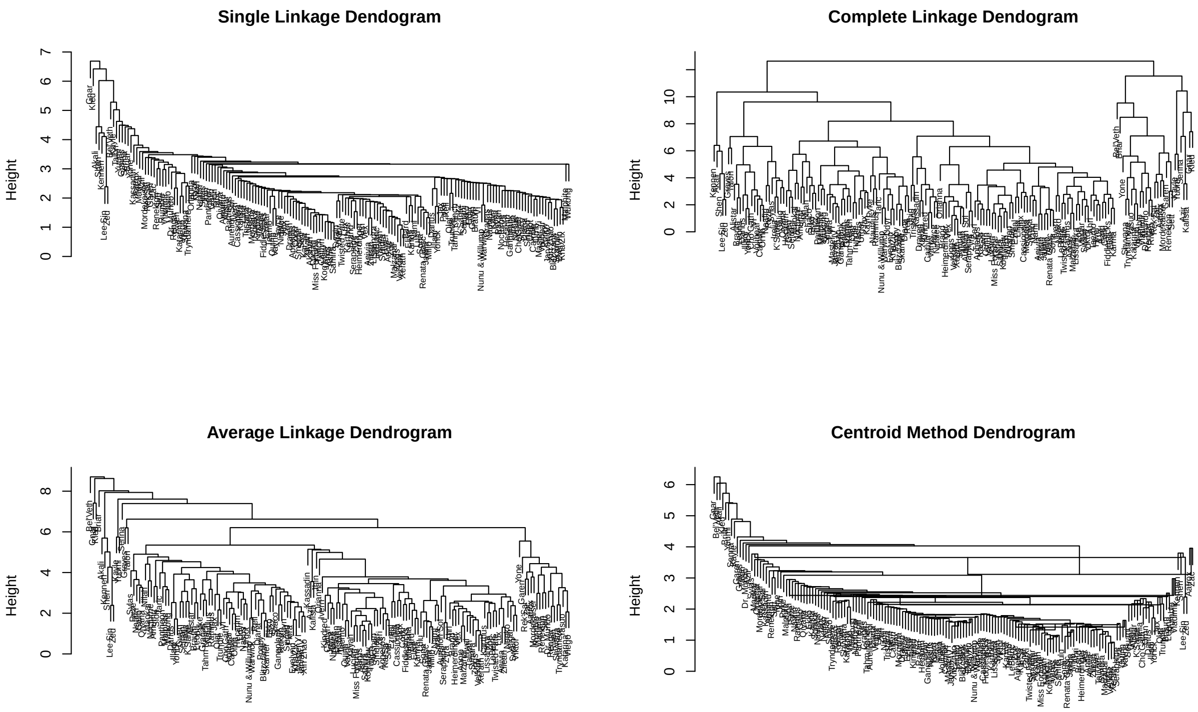


Figure 23: Dendrograms with different methods

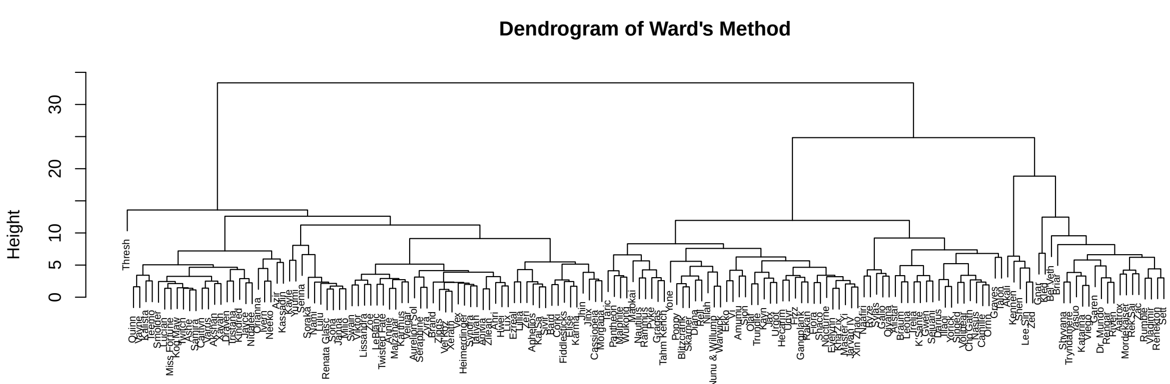


Figure 24: Dendrogram with Outlier

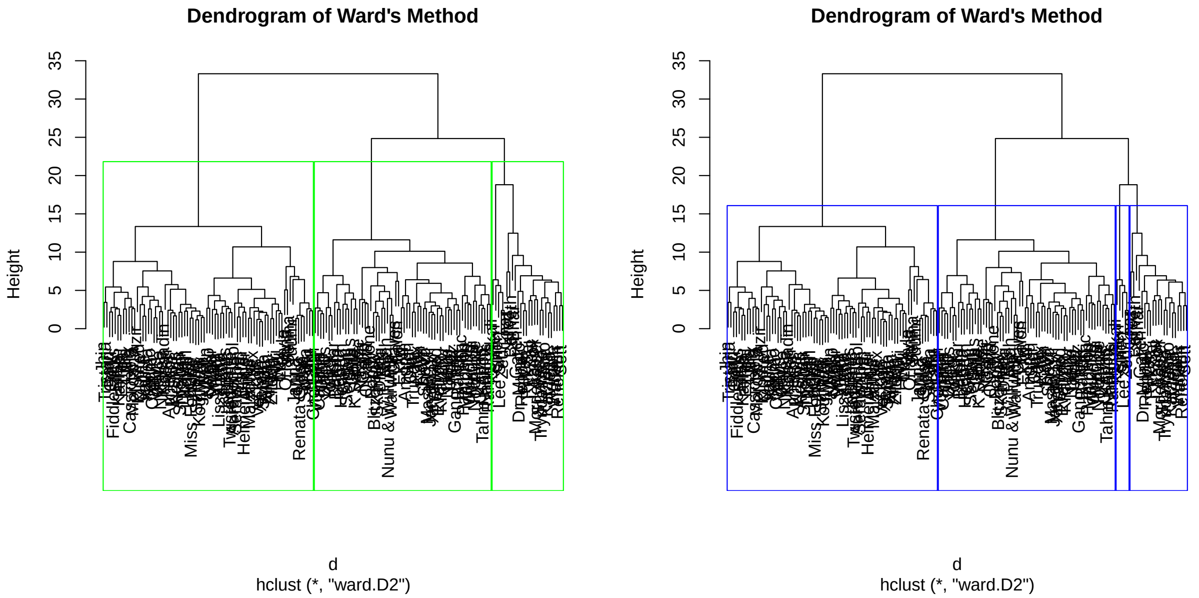


Figure 25: Dendrogram with different clusters

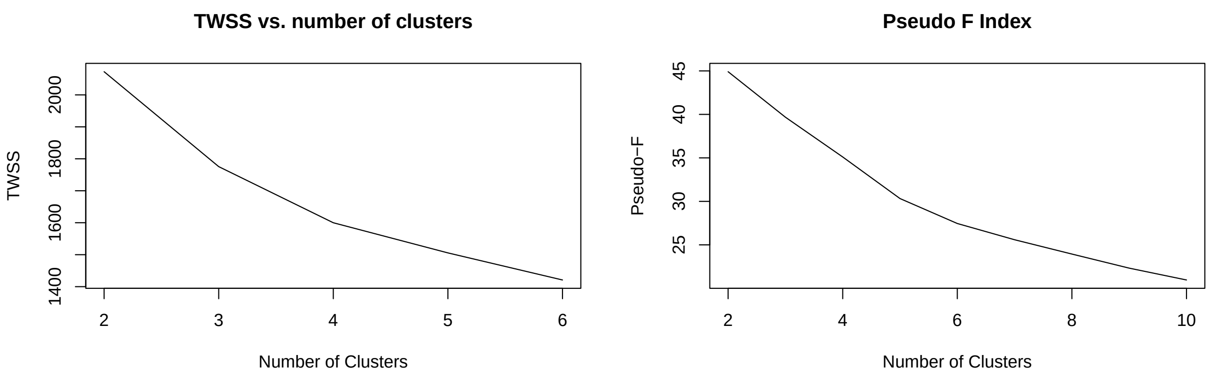


Figure 26: Unhierarchical Clustering

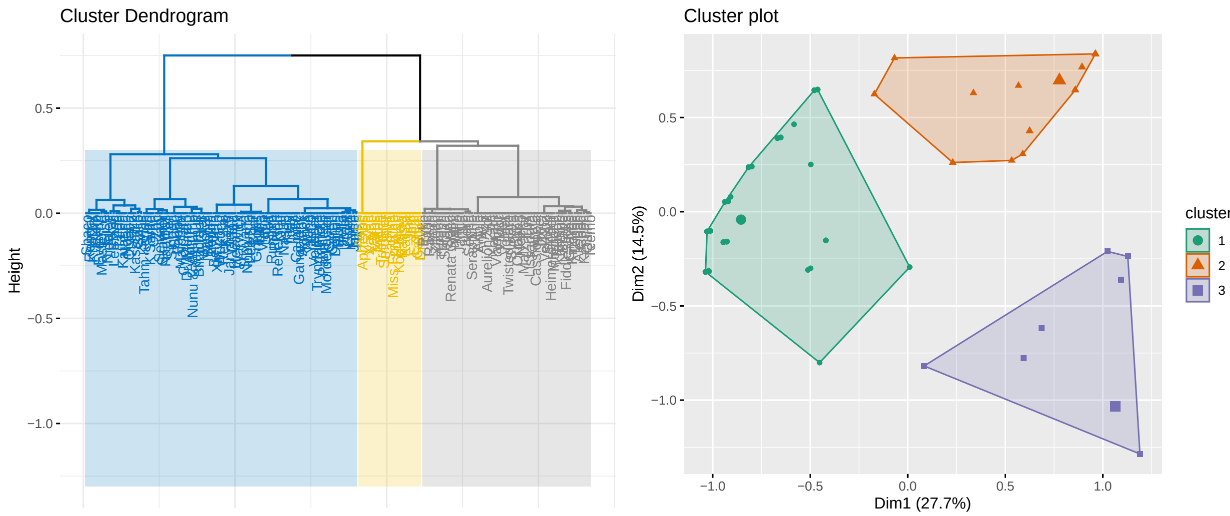


Figure 27: Profiling Analysis MCA

Table 11: Categorical variables profiling of Cluster 1

	X1.Cla.Mod	X1.Mod.Cla	X1.Global	X1.p.value	X1.v.test
Range.type=Ranged	92.405063	93.589744	47.305389	1.000000e-33	12.078803
Attack.range=Long Range	89.156627	94.871795	49.700599	8.000000e-32	11.739679
Tags=Tags_Marksman	96.428571	34.615385	16.766467	1.211101e-09	6.078767
Resourse.type=Mana	55.714286	100.000000	83.832335	4.696451e-09	5.857585
Tags=Tags_Mage	88.571429	39.743590	20.958084	1.142433e-08	5.708101
Role=Role_Bottom	96.000000	30.769231	14.970060	2.051577e-08	5.607595
Role=Role_Middle	70.588235	30.769231	20.359281	1.983369e-03	3.092711
Tags=Tags_Support	82.352941	17.948718	10.179641	2.076692e-03	3.079039
Tags=Tags_Assassin	23.529412	5.128205	10.179641	4.629170e-02	-1.992724
Resourse.type=Energy	0.000000	0.000000	2.994012	4.073515e-02	-2.046216
Resourse.type=Nothing	0.000000	0.000000	4.191617	1.088895e-02	-2.546243
Role=Role_Jungle	26.315789	12.820513	22.754491	4.196814e-03	-2.862977
Tags=Tags_Tank	0.000000	0.000000	14.371257	5.130733e-08	-5.446719
Attack.range=Medium Range	7.500000	3.846154	23.952096	2.056011e-09	-5.993319
Role=Role_Top	9.090909	5.128205	26.347305	1.154730e-09	-6.086407
Tags=Tags_Fighter	4.347826	2.564103	27.544910	4.860785e-13	-7.229135
Attack.range=Close Range	2.272727	1.282051	26.347305	8.540171e-14	-7.461718
Range.type=Melee	5.681818	6.410256	52.694611	1.000000e-33	-12.078803

Table 12: Categorical variables profiling of Cluster 2

	X2.Cla.Mod	X2.Mod.Cla	X2.Global	X2.p.value	X2.v.test
Resourse.type=Nothing	100.000000	25.925926	4.191617	1.403387e-06	4.824523
Tags=Tags_Fighter	39.130435	66.666667	27.544910	3.907919e-06	4.616220
Range.type=Melee	27.272727	88.888889	52.694611	2.201832e-05	4.243375
Resourse.type=Energy	100.000000	18.518519	2.994012	7.922755e-05	3.946725
Role=Role_Top	36.363636	59.259259	26.347305	8.563387e-05	3.928060
Resourse.type=Fury	100.000000	14.814815	2.395210	5.614822e-04	3.449569
Attack.range=Close Range	31.818182	51.851852	26.347305	2.169386e-03	3.066007
Resourse.type=Rage	100.000000	7.407407	1.197605	2.532285e-02	2.236442
Resourse.type=Flow	100.000000	7.407407	1.197605	2.532285e-02	2.236442
Tags=Tags_Support	0.000000	0.000000	10.179641	4.211358e-02	-2.032396
Role=Role_Bottom	0.000000	0.000000	14.970060	8.249936e-03	-2.641666
Role=Role_Support	0.000000	0.000000	15.568862	6.681286e-03	-2.712326
Tags=Tags_Marksman	0.000000	0.000000	16.766467	4.360098e-03	-2.850861
Range.type=Ranged	3.797468	11.111111	47.305389	2.201832e-05	-4.243375
Attack.range=Long Range	3.614458	11.111111	49.700599	6.830531e-06	-4.498902
Resourse.type=Mana	0.000000	0.000000	83.832335	9.700000e-32	-11.722721

Table 13: Categorical variables profiling of Cluster 3

	X3.Cla.Mod	X3.Mod.Cla	X3.Global	X3.p.value	X3.v.test
Range.type=Melee	67.045455	95.161290	52.694611	2.810535e-19	8.975897
Tags=Tags_Tank	87.500000	33.870968	14.371257	6.105989e-08	5.415669
Resourse.type=Mana	44.285714	100.000000	83.832335	8.549630e-07	4.922384
Attack.range=Close Range	65.909091	46.774194	26.347305	6.997606e-06	4.493761
Attack.range=Medium Range	67.500000	43.548387	23.952096	8.996130e-06	4.439994
Tags=Tags_Fighter	56.521739	41.935484	27.544910	1.804396e-03	3.120671
Role=Role_Top	54.545455	38.709677	26.347305	6.586437e-03	2.717061
Role=Role_Jungle	52.631579	32.258065	22.754491	2.847442e-02	2.190687
Resourse.type=Nothing	0.000000	0.000000	4.191617	3.596961e-02	-2.097271
Role=Role_Middle	20.588235	11.290323	20.359281	2.445200e-02	-2.249952
Role=Role_Bottom	4.000000	1.612903	14.970060	5.724749e-05	-4.023878
Tags=Tags_Marksman	3.571429	1.612903	16.766467	1.155421e-05	-4.385834
Tags=Tags_Mage	5.714286	3.225806	20.958084	3.379319e-06	-4.646306
Attack.range=Long Range	7.228916	9.677419	49.700599	1.407775e-16	-8.264079
Range.type=Ranged	3.797468	4.838710	47.305389	2.810535e-19	-8.975897

Table 14: Numerical variables profiling

	Eta2	P-value
Magic.resistance.per.lvl	0.84469719	5.000000e-67
Mana.regeneration.per.lvl	0.72972407	2.559787e-47
Mana.per.lvl	0.71309699	3.421989e-45
Base.mana	0.62162963	2.450214e-35
Base.armor	0.52159598	5.535734e-27
Attack.damage	0.42362467	2.386940e-20
Movement.speed	0.39095750	2.193325e-18
HP.regeneration	0.38215579	7.113390e-18
Base.magic.resistance	0.30945580	6.511808e-14
HP.regeneration.per.lvl	0.25719344	2.581131e-11
Base.HP	0.19276376	2.366210e-08
Attack.damage.per.lvl	0.11151807	6.154392e-05
HP.per.lvl	0.04657549	2.002063e-02
AS.ratio	0.03908751	3.802547e-02

Table 15: Numerical variables profiling of Cluster 1

	X1.v.test	X1.Mean.in.category	X1.Overall.mean	X1.sd.in.category	X1.Overall.sd	X1.p.value
Base.mana	6.137196	375.5128205	313.5868263	70.73999768	121.70526120	8.399086e-10
Mana.regeneration.per.lvl	5.102791	0.6993590	0.5742515	0.16515214	0.29572079	3.346812e-07
HP.per.lvl	-2.368723	102.3589744	104.0598802	7.96301444	8.66108414	1.784959e-02
AS.ratio	-2.466692	0.6229615	0.6365329	0.08702290	0.06636145	1.363674e-02
Attack.damage.per.lvl	-4.250803	2.9633538	3.1939617	0.56398651	0.65434898	2.130056e-05
Base.HP	-5.343703	600.4230769	617.8562874	33.95302910	39.34969515	9.106654e-08
HP.regeneration.per.lvl	-5.916780	0.5839744	0.6580838	0.07048076	0.15107572	3.283060e-09
Base.magic.resistance	-6.957980	29.7820513	30.7604790	1.26741647	1.69610184	3.451859e-12
HP.regeneration	-7.409035	5.1314103	6.2979042	1.38196045	1.89900977	1.272220e-13
Movement.speed	-7.955032	331.5128205	336.1556886	4.70336374	7.03964483	1.790850e-15
Attack.damage	-8.343855	54.9487179	58.8562874	4.44308000	5.64867156	7.190740e-17
Base.armor	-9.235428	25.0000000	29.8862275	3.88620176	6.38150318	2.572548e-20
Magic.resistance.per.lvl	-11.786253	1.3153846	1.6847305	0.09684368	0.37797572	5.000000e-32

Table 16: Numerical variables profiling of Cluster 2

	X2.v.test	X2.Mean.in.category	X2.Overall.mean	X2.sd.in.category	X2.Overall.sd	X2.p.value
Movement.speed	4.429450	341.666667	336.1556886	4.5133547	7.0396448	9.447355e-06
Magic.resistance.per.lvl	3.804630	1.938889	1.6847305	0.2664351	0.3779757	1.420165e-04
Base.armor	2.760805	33.000000	29.8862275	3.4854194	6.3815032	5.765903e-03
Attack.damage	2.666683	61.518519	58.8562874	4.5327644	5.6486716	7.660397e-03
Base.mana	-9.902171	100.592593	313.5868263	122.1581426	121.7052612	4.073369e-23
Mana.per.lvl	-10.533105	0.000000	36.7125749	0.0000000	19.7210644	6.079556e-26
Mana.regeneration.per.lvl	-10.987330	0.000000	0.5742515	0.0000000	0.2957208	4.397400e-28

Table 17: Numerical variables profiling of Cluster 3

	X3.v.test	X3.Mean.in.category	X3.Overall.mean	X3.sd.in.category	X3.Overall.sd	X3.p.value
Magic.resistance.per.lvl	9.271906	2.0387097	1.6847305	0.1305754	0.3779757	1.828494e-20
Base.armor	7.433200	34.6774194	29.8862275	5.3090738	6.3815032	1.060016e-13
HP.regeneration	7.360257	7.7096774	6.2979042	1.3396810	1.8990098	1.835564e-13
Attack.damage	6.584241	62.6129032	58.8562874	3.9690530	5.6486716	4.572160e-11
Base.magic.resistance	6.199681	31.8225806	30.7604790	1.3139965	1.6961018	5.657754e-10
HP.regeneration.per.lvl	6.182097	0.7524194	0.6580838	0.1395348	0.1510757	6.325558e-10
Mana.per.lvl	6.120005	48.9032258	36.7125749	10.6084411	19.7210644	9.357236e-10
Base.HP	5.133125	638.2580645	617.8562874	23.9589269	39.3496951	2.849701e-07
Movement.speed	4.839494	339.5967742	336.1556886	6.6779842	7.0396448	1.301701e-06
Attack.damage.per.lvl	3.520069	3.4266129	3.1939617	0.6404835	0.6543490	4.314348e-04
Mana.regeneration.per.lvl	3.102974	0.6669355	0.5742515	0.1713193	0.2957208	1.915867e-03
HP.per.lvl	2.715545	106.4354839	104.0598802	9.0368796	8.6610841	6.616684e-03

Table 18: Shapiro-Wilk normality test

Variable	Statistic	P_value
Base.HP	0.9403490	0.0000019
Base.armor	0.9771872	0.0074269
HP.regeneration	0.9697078	0.0010267
Attack.damage	0.9806891	0.0198913
Attack.speed.per.lvl	0.9631268	0.0002073

Table 19: Asymptotic one-sample Kolmogorov-Smirnov test

Variable	Statistic	P_value	Alternative
Base.HP	0.1118568	0.0306278	two-sided
Base.armor	0.0639300	0.5022297	two-sided
HP.regeneration	0.0921325	0.1173990	two-sided
Attack.damage	0.0650198	0.4802642	two-sided
Attack.speed.per.lvl	0.0817870	0.2138997	two-sided

Table 20: New Base.HP Test Results

Type	Statistic	P_value	Alternative
Shapiro-Wilk normality test	0.9665080	0.0004643	
Asymptotic one-sample Kolmogorov-Smirnov test	0.1046254	0.0516642	two-sided

Table 21: Box's M Test Result on Range Type

	Group	M Statistic	df1	P-Value
Chi-Sq (approx.)	Range Type	48.31404	15	2.26e-05

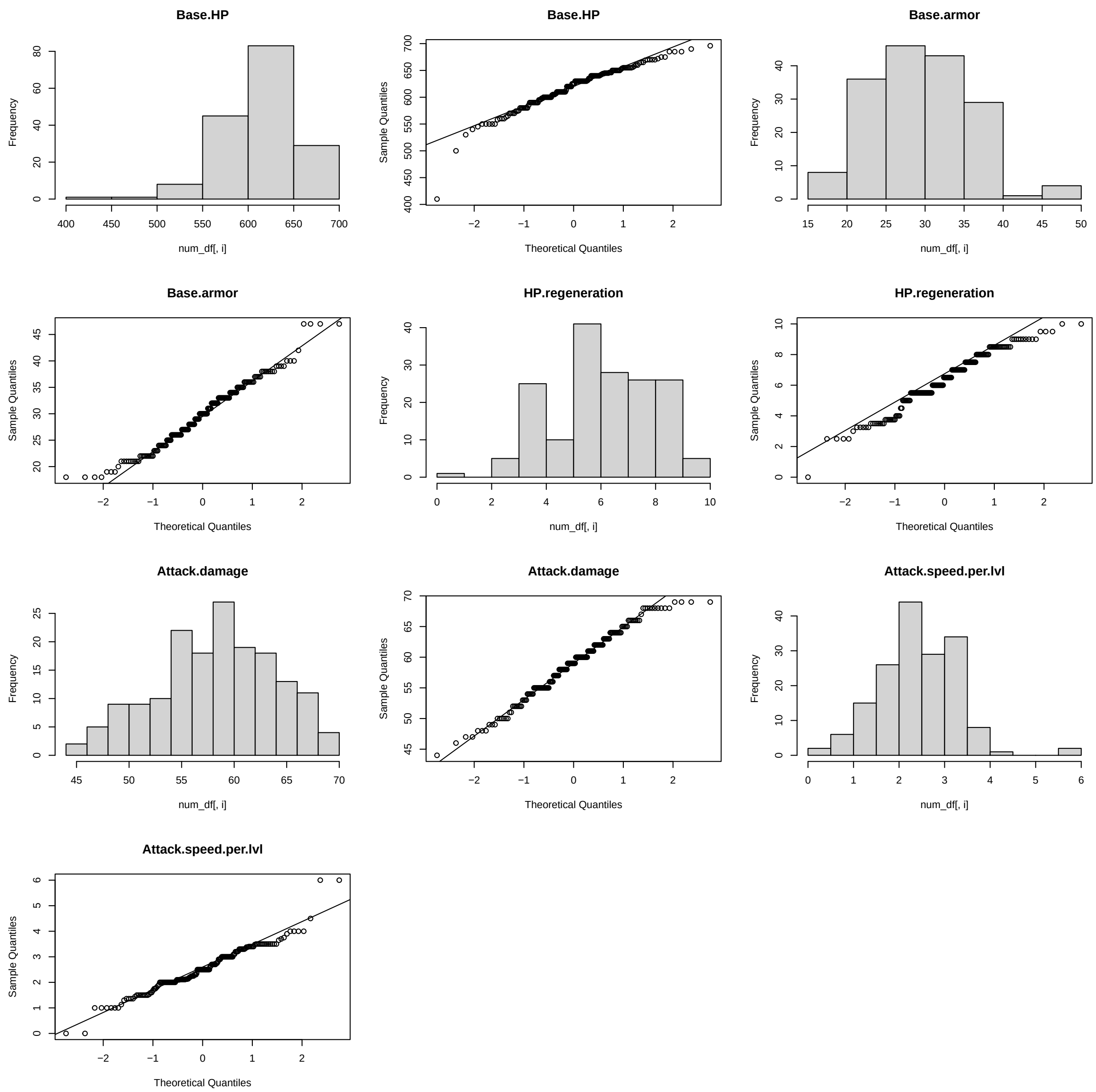


Figure 28: Variable Distributions

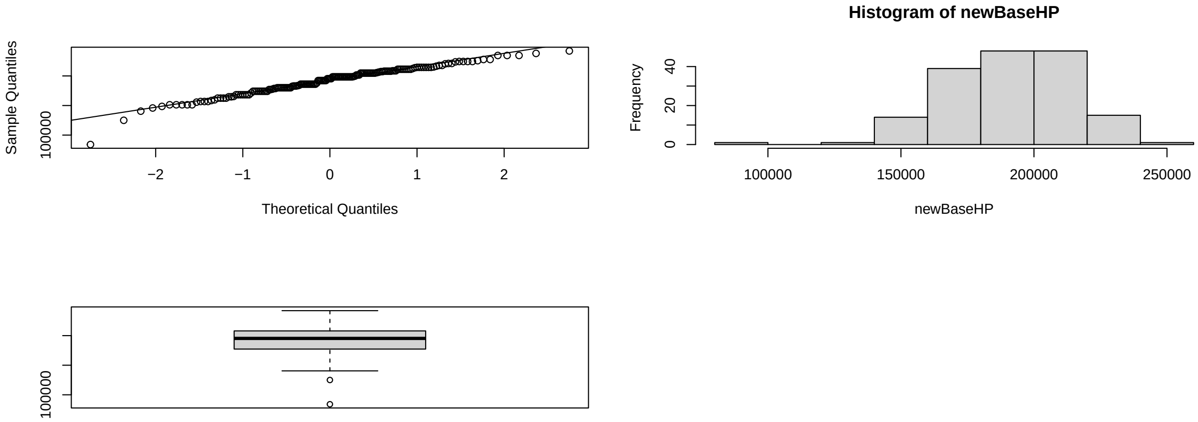


Figure 29: Base.HP transformation

Table 22: Hotelling's two sample T2-test

	Group	T.2 Statistic	df	P-Value	Alternative	Null Value
T.2	Range Type	261.9775	5	0	two.sided	c(0,0,0,0,0)

Table 23: Box's M Test Results

	Group	M Statistic	df1	P-Value
Chi-Sq (approx.)	Role	20.08615	12	0.0654738
Chi-Sq (approx.)1	Tags	62.47965	50	0.1107610

Table 24: Summary for MANOVA Results

Response	Effect	Df	Sum of Squares	Mean Square	F Value	Pr(>F)
Response Base.HP	Tags	5	2.132358e+10	4.264716e+09	9.695314	4.130712e-08
Response Base.HP	Residuals	161	7.081971e+10	4.398740e+08	NA	NA
Response HP.regeneration	Tags	5	2.225558e+02	4.451117e+01	18.874279	9.600000e-15
Response HP.regeneration	Residuals	161	3.796859e+02	2.358298e+00	NA	NA
Response Attack.damage	Tags	5	3.019950e+03	6.039899e+02	42.121775	0.000000e+00
Response Attack.damage	Residuals	161	2.308601e+03	1.433914e+01	NA	NA
Response Attack.speed.per.lvl	Tags	5	1.169842e+01	2.339684e+00	3.139347	9.878505e-03
Response Attack.speed.per.lvl	Residuals	161	1.199896e+02	7.452771e-01	NA	NA

Table 25: Tukey HSD Test Results for the Effect of Tags on Base HP

	diff	lwr	upr	p adj
Fighter-Assassin	-1504.698	-18675.034	15665.6385	0.99986
Mage-Assassin	-23281.797	-41165.402	-5398.1909	0.00325
Marksman-Assassin	-11106.543	-29706.617	7493.5312	0.51914
Support-Assassin	-29474.471	-50223.722	-8725.2190	0.00092
Tank-Assassin	2337.784	-16838.910	21514.4786	0.99928
Mage-Fighter	-21777.099	-35345.900	-8208.2982	0.00011
Marksman-Fighter	-9601.845	-24101.906	4898.2154	0.39985
Support-Fighter	-27969.773	-45140.109	-10799.4369	0.00008
Tank-Fighter	3842.482	-11390.200	19075.1641	0.97829
Marksman-Mage	12175.254	-3162.761	27513.2678	0.20438
Support-Mage	-6192.674	-24076.280	11690.9318	0.91775
Tank-Mage	25619.581	9587.191	41651.9707	0.00012
Support-Marksman	-18367.928	-36968.002	232.1468	0.05507
Tank-Marksman	13444.327	-3383.534	30272.1891	0.19835
Tank-Support	31812.255	12635.561	50988.9492	0.00006

Table 26: Tukey HSD Test Results for the Effect of Tags on HP Regeneration

	diff	lwr	upr	p adj
Fighter-Assassin	-0.65537	-1.91260	0.60186	0.66250
Mage-Assassin	-1.37059	-2.68004	-0.06113	0.03440
Marksman-Assassin	-3.26523	-4.62715	-1.90332	0.00000
Support-Assassin	-1.97059	-3.48987	-0.45131	0.00342
Tank-Assassin	0.30025	-1.10389	1.70438	0.98968
Mage-Fighter	-0.71522	-1.70874	0.27830	0.30492
Marksman-Fighter	-2.60986	-3.67157	-1.54815	0.00000
Support-Fighter	-1.31522	-2.57245	-0.05799	0.03456
Tank-Fighter	0.95562	-0.15974	2.07097	0.13893
Marksman-Mage	-1.89464	-3.01771	-0.77158	0.00004
Support-Mage	-0.60000	-1.90945	0.70945	0.77268
Tank-Mage	1.67083	0.49693	2.84474	0.00089
Support-Marksman	1.29464	-0.06727	2.65656	0.07270
Tank-Marksman	3.56548	2.33332	4.79763	0.00000
Tank-Support	2.27083	0.86670	3.67497	0.00009

Table 27: Tukey HSD Test Results for the Effect of Tags on Attack Damage

	diff	lwr	upr	p adj
Fighter-Assassin	2.75064	-0.34946	5.85074	0.11362
Mage-Assassin	-7.16303	-10.39191	-3.93414	0.00000
Marksman-Assassin	-1.95588	-5.31412	1.40236	0.54707
Support-Assassin	-8.82353	-12.56981	-5.07725	0.00000
Tank-Assassin	0.83578	-2.62657	4.29814	0.98214
Mage-Fighter	-9.91366	-12.36351	-7.46382	0.00000
Marksman-Fighter	-4.70652	-7.32451	-2.08854	0.00001
Support-Fighter	-11.57417	-14.67427	-8.47407	0.00000
Tank-Fighter	-1.91486	-4.66512	0.83541	0.34225
Marksman-Mage	5.20714	2.43787	7.97642	0.00000
Support-Mage	-1.66050	-4.88939	1.56838	0.67539
Tank-Mage	7.99881	5.10416	10.89346	0.00000
Support-Marksman	-6.86765	-10.22589	-3.50940	0.00000
Tank-Marksman	2.79167	-0.24660	5.82994	0.09139
Tank-Support	9.65931	6.19696	13.12167	0.00000

Table 28: Tukey HSD Test Results for the Effect of Tags on Attack Speed per lvl

	diff	lwr	upr	p adj
Fighter-Assassin	-0.09041	-0.79717	0.61635	0.99909
Mage-Assassin	-0.58138	-1.31750	0.15474	0.20917
Marksman-Assassin	0.10819	-0.65742	0.87381	0.99853
Support-Assassin	-0.24235	-1.09643	0.61172	0.96378
Tank-Assassin	-0.54469	-1.33404	0.24465	0.35240
Mage-Fighter	-0.49097	-1.04949	0.06755	0.12006
Marksman-Fighter	0.19860	-0.39825	0.79545	0.92982
Support-Fighter	-0.15194	-0.85871	0.55482	0.98942
Tank-Fighter	-0.45428	-1.08129	0.17272	0.29786
Marksman-Mage	0.68957	0.05823	1.32091	0.02346
Support-Mage	0.33903	-0.39710	1.07515	0.76886
Tank-Mage	0.03668	-0.62324	0.69661	0.99999
Support-Marksman	-0.35055	-1.11616	0.41507	0.77324
Tank-Marksman	-0.65289	-1.34555	0.03978	0.07710
Tank-Support	-0.30234	-1.09169	0.48701	0.87882

Table 29: Box's M Test Result on Tags

	Group	M Statistic	df1	P-Value
Chi-Sq (approx.)	Tags	110.1551	75	0.0051209

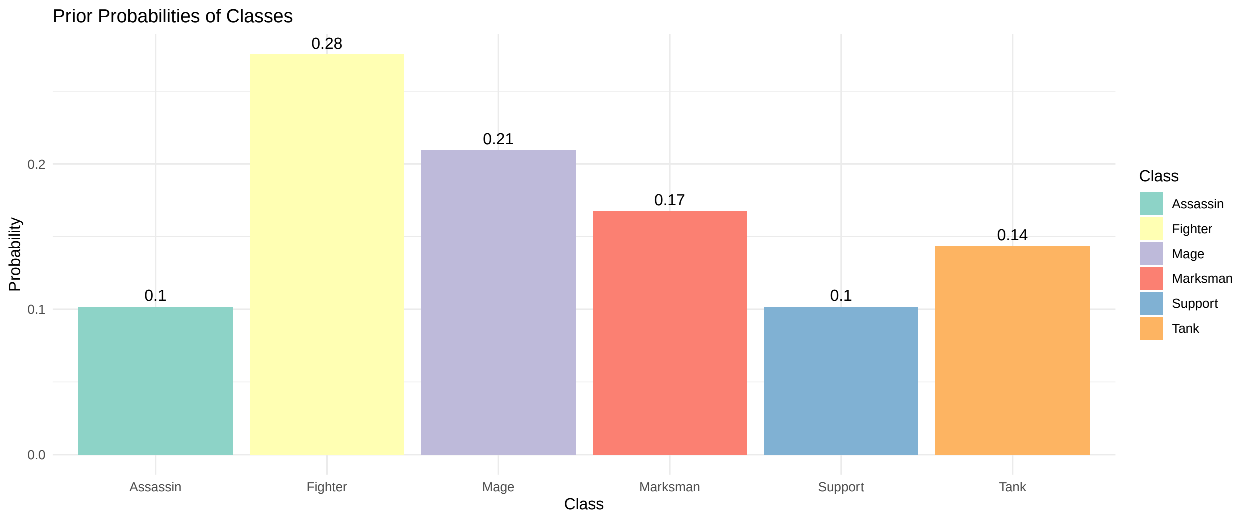


Figure 30: QDA - Probabilities of Tags

Table 30: Group Means for Each Class

Class	Base.HP	HP.regeneration	Attack.damage	Attack.speed.per.lvl	Base.armor
Assassin	201467.4	7.470588	60.70588	2.768235	29.29412
Fighter	199962.7	6.815217	63.45652	2.677826	34.28261
Mage	178185.6	6.100000	53.54286	2.186857	22.85714
Marksman	190360.8	4.205357	58.75000	2.876429	26.46429
Support	171992.9	5.500000	51.88235	2.525882	29.88235
Tank	203805.2	7.770833	61.54167	2.223542	36.12500

Table 31: Confusion Matrix: Predicted vs Actual Classes

	Assassin	Fighter	Mage	Marksman	Support	Tank
Assassin	7	1	1	1	0	1
Fighter	5	38	1	2	0	11
Mage	2	1	28	3	2	1
Marksman	1	1	2	22	1	1
Support	0	0	3	0	13	0
Tank	2	5	0	0	1	10

```
## [1] "QDA Correct Classification Rate: 0.706586826347305"
```

```
## [1] "LDA Correct Classification Rate: 0.562874251497006"
```

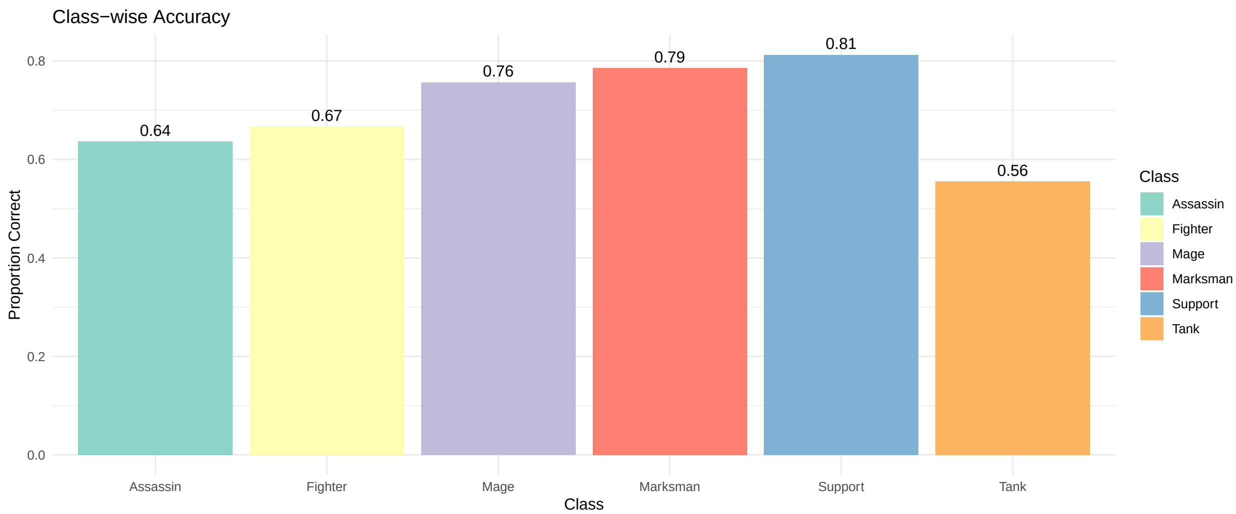


Figure 31: QDA Class-wise Accuracy

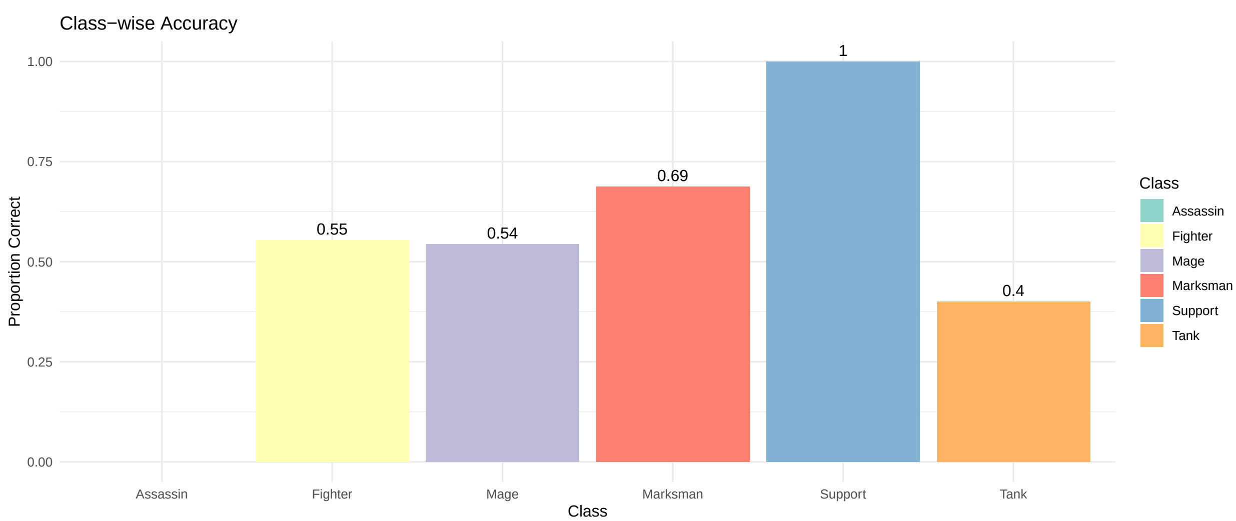


Figure 32: LDA Class-wise Accuracy

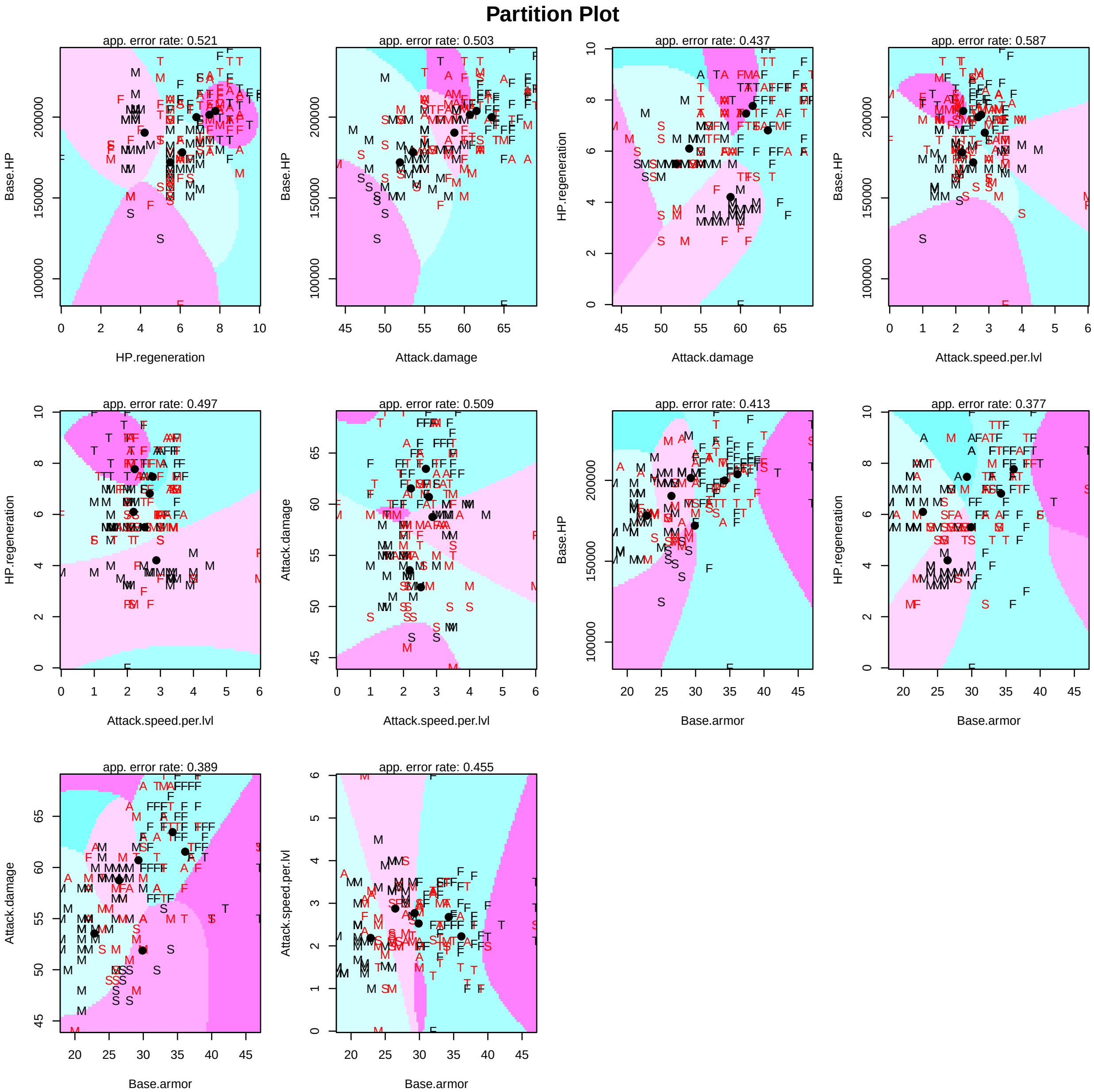


Figure 33: QDA Partition Plot