Clustering Analysis of FIFA Player Positions

Why So Series?

David Buksa (buksa2)
Roberto Sanchez-Vegas (snchzvg2)
Favian Tirado (ftirad2)
Conrad Manaugh (manaugh2)
Meghan Clodfelter (clodfel2)

May 5, 2019

Background

We obtained the "FIFA 19 Complete Player Dataset" from Kaggle, and the raw data for this was obtained from SoFIFA. SoFIFA gathers this data in purposes of observing each player and comparing their individual attributes, as well as evaluating teams and positions as a whole. At SoFIFA.com, you are able to compare and sort all of these factors, as well as view photos of players, jerseys, and team flags in efforts to learn as much as possible about all of the FIFA players. This 2019 FIFA data was brought to Kaggle 5 months ago (at the beginning of 2019) by user Karan Gadiya. Taken from real-life players and adapted for the video game of the same name, the relevant data to our project explored the various skills and abilities players in each position possess. We observed these skill levels and performed clustering analysis on each of the positions, categorizing them into various archetypes that provide a better insight into their unique role on a soccer team.

Before we can explore our in-depth look into the player's archetypes and clusters, we will review the basics. Soccer, (or football in any country other than the United States), is a game where each team consists of eleven players, and the object of the game is to score more points than the opposing team. A team scores a point when the soccer ball makes it into the other team's goal. Even though the game seems to be simple, there are many positions that players specialize in and play at the professional level.

The first and most important of these positions is the goalkeeper (GK in the data). Their main job is to block the ball from entering the goal. Unlike all other positions, they are allowed to use their hands and arms to touch the ball. However, they can only do this in a small area near their team's goal. Outside of this area, they are like every other player; they are allowed to use everything but their arms and hands to manipulate the ball.

Next, we have the defensive positions: center back (CB), left and right full back (LCB, RCB), and left and right wing back (LWB, RWB). These positions all have common jobs of protecting the goalkeeper, blocking shots, and stopping the opposing team's offensive players from passing, receiving, and shooting at the goal. However, each of these positions have their own unique roles. To be specific, the center back will hang back even when the rest of the team is on the offensive to help protect the goal. In contrast, the full back and the wing back positions play at the wider parts of the field and help support the more offensive positions when attacking (or trying to score a goal). The main difference between the two is that the wing backs are a more aggressive position, since they stay wide in order to assist their offensive teammates, and full backs usually protect the wide sides of the field while also assisting action in the center of the field when needed.

The midfield positions for players include the following: left, center, and right defensive midfielder (LDM, CDM, RDM), left, center, and right central midfielder (LCM, CM, RCM), and left, center, and right attacking midfielder (LAM, CAM, RAM). They serve as the connection between the defensive and offensive players because they can play across the field depending on where the ball is. Because of this, midfielders have to be able to run back and forth across the field. The defensive midfielders' more specialized job is to keep the ball from advancing past them, either by intercepting passes or stealing the ball from the other team's players. Their main job is to pass the ball and to support their teammates on both the offensive and defensive. Given this information, it makes sense that they would be skilled in ball handling and passing. The attacking

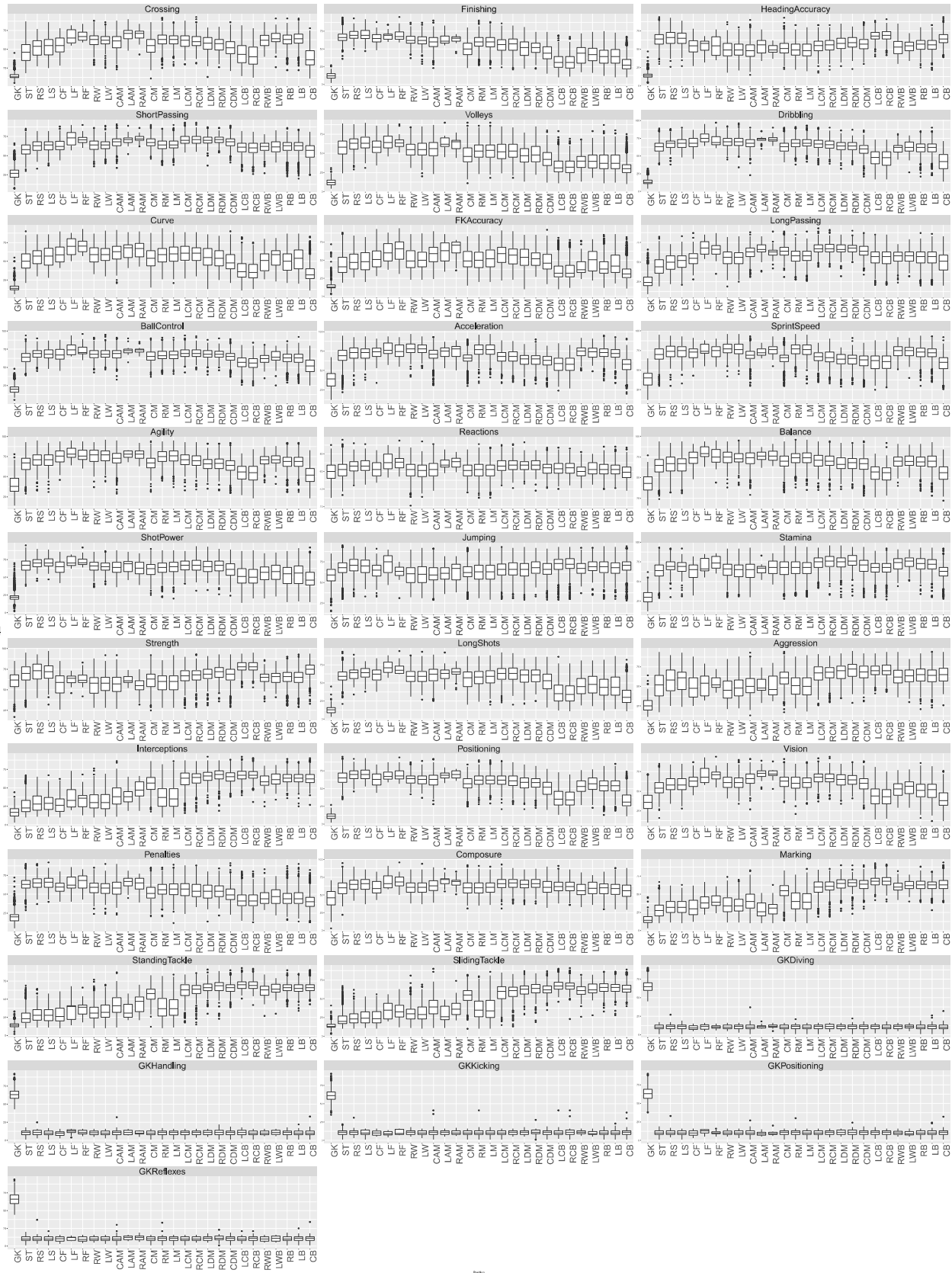
midfielders are positioned in between the midfield and offensive lines and will usually be skilled in dribbling and are able to create plays for their offensive teammates to score a goal. The left/right versions of all specializations of midfielders will usually have good one-on-one skills and are expected to be able to pass to their offensive players or even take a shot at scoring a goal themselves from further away since they operate on the outside.

The last of the player positions are the offensive player positions which include the strikers (LS, RS, ST), forwards (CF, LF, RF), and wings (LW, RW). The strikers and the forwards have the same main job of scoring goals while also having skills in dribbling and speed to get past the defensive players. They should also be skilled in heading. When they are not in possession of the ball, their job is to apply pressure to the opposing team's defensive players. The difference between these two general positions is that strikers are positioned closer to the enemy team's goal. The wings will stay at both edges of the field and try to pass the ball to the striker or forward with the hope of eventually scoring a goal. They can also attempt to score a goal themselves if they see the opportunity. They are usually skilled in one-on-one skills such as ball control and dribbling as well as being fast.

We proceeded to use unsupervised learning to perform an analysis on the data to see whether the players fell into clusters that separated them from other positions based on a multitude of variables. The methods used in our analysis include k-means clustering, hierarchical clustering, and gaussian mixture models. The main goal of this analysis is too take a closer look at the players and their abilities and create new sets of archetypes that can possibly describe the players more accurately than the current positions. With this analysis, we hope to be able to make custom offense and defense lineups that maximize the individual talents of the players.

Summary of Data

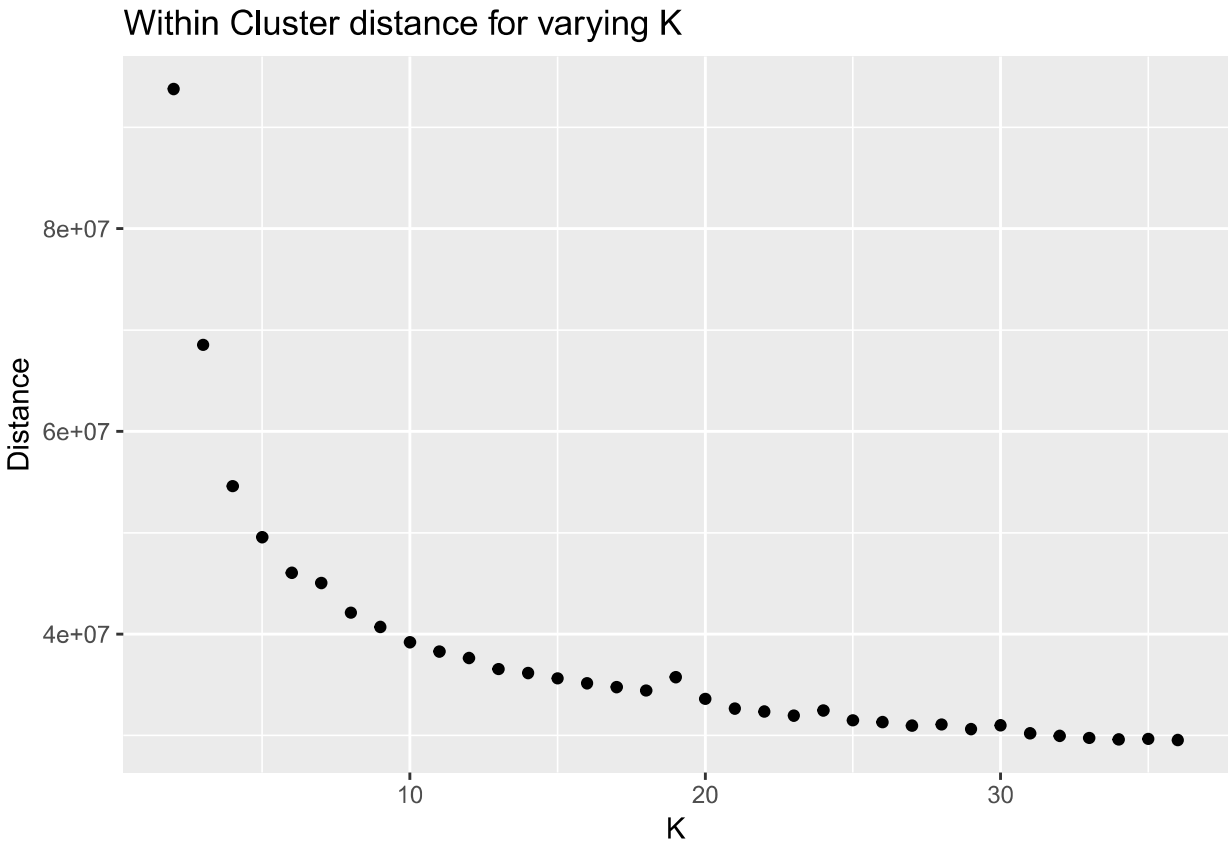
	N	Mean	SD	Min	Q1	Median	Q3	Max
Crossing	18147	49.74	18.36	5	38	54	64	93
Finishing	18147	45.55	19.53	2	30	49	62	95
HeadingAccuracy	18147	52.30	17.38	4	44	56	64	94
ShortPassing	18147	58.70	14.70	7	54	62	68	93
Volleys	18147	42.91	17.70	4	30	44	57	90
Dribbling	18147	55.38	18.91	4	49	61	68	97
Curve	18147	47.18	18.40	6	34	48	62	94
FKAccuracy	18147	42.87	17.48	3	31	41	57	94
LongPassing	18147	52.72	15.33	9	43	56	64	93
BallControl	18147	58.37	16.69	5	54	63	69	96
Acceleration	18147	64.61	14.93	12	57	67	75	97
SprintSpeed	18147	64.73	14.65	12	57	67	75	96
Agility	18147	63.50	14.77	14	55	66	74	96
Reactions	18147	61.84	9.01	21	56	62	68	96
Balance	18147	63.96	14.14	16	56	66	74	96
ShotPower	18147	55.47	17.24	2	45	59	68	95
Jumping	18147	65.09	11.82	15	58	66	73	95
Stamina	18147	63.22	15.90	12	56	66	74	96
Strength	18147	65.32	12.55	17	58	67	74	97
LongShots	18147	47.11	19.26	3	33	51	62	94
Aggression	18147	55.88	17.37	11	44	59	69	95
Interceptions	18147	46.70	20.70	3	26	52	64	92
Positioning	18147	49.96	19.53	2	38	55	64	95
Vision	18147	53.41	14.15	10	44	55	64	94
Penalties	18147	48.55	15.70	5	39	49	60	92
Composure	18147	58.65	11.44	3	51	60	67	96
Marking	18147	47.29	19.90	3	30	53	64	94
StandingTackle	18147	47.70	21.66	2	27	55	66	93
SlidingTackle	18147	45.67	21.29	3	24	52	64	91
GKDiving	18147	16.62	17.70	1	8	11	14	90
GKHandling	18147	16.39	16.91	1	8	11	14	92
GKKicking	18147	16.23	16.50	1	8	11	14	91
GKPositioning	18147	16.39	17.04	1	8	11	14	90
GKReflexes	18147	16.71	17.96	1	8	11	14	94


Crossing
Finishing
HeadingAccuracy
ShortPassing
Volleys
Dribbling
Curve
FKAccuracy
LongPassing
BallControl
Acceleration
SprintSpeed
Agility
Reactions
Balance
ShotPower
Jumping
Stamina
Strength
LongShots
Aggression
Interceptions
Positioning
Vision
Penalties
Composure
Marking
StandingTackle
SlidingTackle
GKDiving
GKHandling
GKKicking
GKPositioning
GKReflexes
GK ST RS LS CF LF RF RW LW CAM LAM RAM CM RM LM LCM RCM LDM RDM CDM LCB RCB RWB LWB RB LB CB

Based on these plots we can identify key skills for each position, skills such as `StandTackle`, `Interceptions`, `Marking` and `SlideTackle` are important for defenders while not so much for offensive players. For offence skills such as `Finishing`, `Volleys` and `ShotPower` are the most important. As goalies have their own skills, that specificially pertain to goalies, So we can see that they excel in `GK` skills and have poor skills otherwise.

Analysis and Classification of Cluster Archetypes

Cluster Initializations



Based on the elbow method, 18 clusters seems to be appropriate. We will analyze k = 4, 18, and 28.

K = 4 (Generalized Positions)

```
analysis_generalized = kmeans(fifa_19_filtered[,-ncol(fifa_19_filtered)],4)
table(analysis_generalized$cluster, generalized_positions)
```

```
##    generalized_positions
##     Defender Forward   GK Midfielder
##   1        0       0 2025          0
##   2     4008      13    0        985
##   3       14    3218    0       2612
##   4     1844     187    0       3241
```

Using 4 clusters had reasonable results. The goalies are isolated, and the defenders are isolated from the forwards. The midfielders are split, but this is a promising start.

K = 18 (Ideal K for Bias vs Variance)

```
analysis_18 = kmeans(fifa_19_filtered[, -35], 18, iter.max = 20, nstart = 3)
```

We found 18 clusters had a good balance between distinct and accurate clusters, so we did further analysis on this below.

Without Goalies

```
XnoGoalie = fifa_19_filtered[fifa_19_filtered$Position!="GK", -(30:34)]
analysis_no_goalie = kmeans(XnoGoalie[,-ncol(XnoGoalie)], 16, iter.max = 20, nstart = 3)
```

We tried removing the goalies (and goalie statistics), but they were already well isolated. This did not improve our clusters in any way.

K = 28 (Natural Clustering based on Positions)

```
analysis_natural = kmeans(fifa_19_filtered[,-ncol(fifa_19_filtered)], 28, iter.max = 20, nstart = 3)
```

We wanted to see the proportions of each position in each cluster if we used as many clusters as there are positions. This seemed to overfit the data and create too many clusters. In particular, distinguishing between left and right positions was not useful for us.

Cluster Analysis

The variable `analysis_18$centers` is a cluster matrix with the columns being different attributes a player can be skilled in and the rows are the clusters that will hopefully become our categories. We also have a matrix of cluster positions, with each row being the cluster and each column being the original positions. With these two matrices we can learn about each cluster composition as well as the attributes that those clusters have.

We will gather some statistics for each of the clusters. The mean of each cluster is a good metric for general skill of each archetype. The Size will help us know if they are an outlier cluster with only a few members, The Types will let us know what positions compose the cluster. The Bests will tell us the skills that are likely used for the archetype. To avoid clutter we only show those positions that compose most of the cluster. We also only show skills that are above a 60 on average for each cluster. We are also changing the cluster names retroactively for easy viewing. Lastly we have desc_order, which will order the clusters by overall skill. This will help us categorize them.

Now we will call the summary statistics for each cluster in order of best average.

AllStars

```
##              Allstar Playmakers
## Cluster Size          919.00000
## Cluster Mean           61.79268
## [1] "Cluster_Bests"
##         Stamina    ShortPassing     BallControl       ShotPower
##        79.17193        76.28727        76.01741        74.52992
##       Dribbling         Agility       Reactions         Balance
##        74.13602        74.08487        73.97933        73.39064
```

```
##       Composure     LongPassing    Acceleration      Aggression
##        73.17519        72.80087        72.79869        72.51578
##          Vision     SprintSpeed        Crossing  StandingTackle
##        72.11099        71.92165        71.80849        71.03264
##           Curve   Interceptions       LongShots         Jumping
##        70.83134        70.76170        70.34059        69.40696
##     Positioning        Strength   SlidingTackle         Marking
##        69.07726        68.62894        68.24592        68.07182
##      FKAccuracy       Penalties         Volleys       Finishing
##        65.53863        62.70511        62.10120        61.60174
## HeadingAccuracy
##        60.94124
## [1] "Cluster_Types"
##     CM     LB     RB    RCM    LCM    CDM
## 0.1774 0.1523 0.1110 0.0925 0.0903 0.0729
```

We start by looking at the best players, who we call AllStars. One of the types of Allstar is the Allstar Playmaker, the most versatile player archetype of those we found. They have many midfielders and quite a few backs, but judging by the low percentage of shown composition this archetype has a wide variety of player positions. This makes sense as this archetype is like an Allstar Midfielder. Midfielders can make use of defensive or offensive skills. The Allstar Playmaker has an excess of passing, dribbling, shooting, passing, and defending skills. Although not the best at everything, they are decent at most things and quite good at playmaking, allowing them to run an offense. They can position themselves and other players using passing and advanced dribbling plays to score.

```
##              Allstar Finishers
## Cluster Size         745.00000
## Cluster Mean          58.66881
## [1] "Cluster_Bests"
##      Agility Acceleration    Dribbling  BallControl      Balance
##     80.62282     78.91812     78.62416     78.08725     77.35302
##  SprintSpeed    ShotPower        Curve ShortPassing       Vision
##     76.94765     75.14228     74.88859     74.49799     73.95034
##  Positioning    Composure    Reactions    LongShots     Crossing
##     73.86309     73.44027     73.29128     72.75973     72.03490
##    Finishing      Stamina   FKAccuracy      Volleys    Penalties
##     71.35302     71.06846     69.70201     68.90604     68.46980
##  LongPassing      Jumping     Strength
##     67.77315     64.66980     60.83490
## [1] "Cluster_Types"
##    CAM     LM     RM     ST     RW     LW
## 0.2121 0.1893 0.1785 0.1060 0.0819 0.0698
```

The next Allstar archetype is known for their skills at offense, and especially weaving through the defense to score. We see a good deal of aggressive midfielders, as well as Left and Right midfielders. The left and right midfielder are other names for wings, which along with aggressive midfielders are the most like strikers. So it makes sense that all of these positions are paired together, as they all excel at sprinting, ball control, dribbling, and shooting. Because these players are gifted Allstars they are also decent at passing abilities like crossing and volleys, but these are not their main weapon shooting dribbling and sprinting are. These players are the best finishers, as they run and dribble around defenders to shoot, so we call them the Allstar Finishers.

```
##              Allstar Guardians
## Cluster Size         720.00000
## Cluster Mean          56.93456
## [1] "Cluster_Bests"
```

```
##        Strength      Aggression    ShortPassing         Stamina
##        74.04306        73.38333        71.56111        71.18333
##       ShotPower   Interceptions  StandingTackle     LongPassing
##        70.87917        70.17917        70.10972        69.52778
##     BallControl       Composure       Reactions         Marking
##        69.48472        69.14861        68.92083        68.34444
##         Jumping   SlidingTackle          Vision HeadingAccuracy
##        67.73194        66.69167        65.85972        64.66389
##       LongShots       Dribbling         Agility        Crossing
##        64.38750        64.24444        61.21389        61.17222
##         Balance           Curve
##        61.11944        60.92361
## [1] "Cluster_Types"
##    CDM     CM    RDM     CB    RCM    LCM
## 0.2528 0.1542 0.0889 0.0736 0.0736 0.0722
```

This next archetype has midfielders as the previous one does, but they are closer to the defensive side of things. They also have a multitude of different kinds of defender. This means the archetype has a variety of position it can play, defensive or midfielders who focus defense. This player archetype caries some classic midfielder traits like passing skills, and even some shooting skills. What makes them different than an average midfielder is the series of defensive skills like tackling, marking, heading, and strength. These traits are often seen in defenders. This archetype excels at responding to opposing players entering their side of the field. In these situations these are the best of the first responders, and so we call them the Allstar Guardians, as they guard the edge of the defensive zone. They possess stamina to chase down invading players, but are not very fast players.

```
##              Allstar Defenders
## Cluster Size        1466.00000
## Cluster Mean          55.69092
## [1] "Cluster_Bests"
##        Stamina    SprintSpeed   Acceleration         Agility        Balance
##       77.25989       73.09891       73.05321       71.59004       71.18690
##        Jumping     Aggression StandingTackle       Strength   ShortPassing
##       71.10778       69.67667       68.06071       67.80491       67.68213
##    BallControl  SlidingTackle  Interceptions      Reactions         Marking
##       67.33356       66.62415       66.58458       66.34789       65.36698
##      Dribbling      Composure       Crossing    LongPassing      ShotPower
##       65.26808       64.28308       63.94816       63.29809       62.98772
## [1] "Cluster_Types"
##     RB     LB    CDM     CM    RCM    LCM
## 0.2538 0.2524 0.0955 0.0914 0.0525 0.0416
```

The next, and last Allstar player we have is the Allstar Defenders. These defenders are composed of the defensive position or the defensive midfielder, many left and right backs as we see, but other varieties exist as well. They excel mostly at defensive skills like Marking, Reactions, and Tackling. As Allstars they also have other skills in passing, shooting, and dribbling, but typically the archetype is defined by its defense. With great stamina, and better speed the Guardians, these defenders are meant to chase down any players from the opposite team near their goal.

Our four allstars have two roles on offense and two on defense. They also have two that are focused exclusively on those roles, Defenders and Finishers. The other two are midfielders who focus offense or defense, so they specialize less. Overall the Playmaker is an offensive midfielder with higher than average stats then the Finisher mainly because it can do more things well, while the Finisher specializes in only doing a few things very well. We have a similar situation with Guardians vs Defenders, the Guardian has higher stats as it is a more versatile Midfield position.

Elites

Beyond the Allstar we have the Elite players. Elite players usually have an Allstar version that can do the job better, but some archetypes may have some unique traits. Most of all though Elites are solid soccer players, not everyone can be an Allstar but Elites are still some of the best players on the field.

```
##              Elite Playmakers
## Cluster Size       1020.00000
## Cluster Mean         55.18573
## [1] "Cluster_Bests"
##      Agility      Balance Acceleration      Stamina  SprintSpeed
##     73.09902     72.31765     70.30588     69.78137     69.36961
##  BallControl ShortPassing    Dribbling       Vision    ShotPower
##     68.79216     68.59510     67.87745     66.59706     66.51667
##  LongPassing        Curve    Composure    Reactions     Crossing
##     64.70980     64.40490     64.38529     64.20196     63.72549
##    LongShots  Positioning      Jumping     Strength   FKAccuracy
##     63.56961     63.40294     62.28039     61.98333     61.73137
## [1] "Cluster_Types"
##     CM    CAM     RM     LM    LCM    RCM
## 0.2363 0.1618 0.1363 0.1196 0.0696 0.0637
```

The next archetype is called the Elite Playmaker, names after the Allstar Playmaker as they have many similarities. These players also excel at running, passing, and ball control. They can even do some shooting.

```
##              Elite Strikers
## Cluster Size      741.00000
## Cluster Mean       54.73057
## [1] "Cluster_Bests"
##        Strength       ShotPower     Positioning         Jumping
##        75.66532        74.09447        73.43050        72.91903
##       Finishing HeadingAccuracy     BallControl       Reactions
##        72.63428        71.10256        70.95007        69.92308
##     SprintSpeed       Composure       Dribbling       Penalties
##        69.92173        69.58704        69.17004        69.07287
##         Stamina    Acceleration         Agility         Volleys
##        68.64777        68.37247        68.34143        68.18489
##       LongShots    ShortPassing      Aggression         Balance
##        66.72335        65.93792        64.92848        63.55870
##          Vision           Curve
##        63.26991        61.50472
## [1] "Cluster_Types"
##     ST     RS     LS     LM     RM    CAM
## 0.6599 0.0945 0.0918 0.0445 0.0364 0.0283
```

Another elite that looks like an Allstar archetype. Although they have major difference in strength, generally this archetype has the similar abilities to the Allstar Finisher. However this archetype is entirely strikers, while they had many midfielders and wing players along with some strikers. To note that this archetype is mostly strikers we call it Elite Striker, they are similar to Allstar Finishers in shooting ability, but worse dribblers and sprinters. Where the Finisher could shoot and set up a shot, these Elite Strikers are best used just for shooting.

```
##              Elite Finishers
## Cluster Size      1451.00000
## Cluster Mean        51.71993
## [1] "Cluster_Bests"
```

```
## Acceleration      Agility  SprintSpeed      Balance    Dribbling
##     79.56168     78.90489     78.38594     76.24328     70.21985
##  BallControl      Stamina    ShotPower ShortPassing  Positioning
##     68.40455     66.05513     65.25637     64.36389     63.89524
##      Jumping    Finishing    Reactions       Vision     Crossing
##     62.97519     62.67333     62.31013     62.21985     61.85320
##    Composure    LongShots        Curve
##     61.62164     60.77464     60.66850
## [1] "Cluster_Types"
##     LM     RM    CAM     ST     LW     RW
## 0.2205 0.2040 0.1537 0.1241 0.0999 0.0951
```

This player is similar to the above in that they are both offensively oriented with many strikers. They are different in that the Elite Striker gains value from its shooting, strength and scoring, it excels at the finishing strike. This archetype is okay at finishing, and can do some passing, but they excel incredibly at dribbling and sprinting. They are also composed of many wings and left and right midfielders. They are the Elite version of Finisher, those that can dribble around defenders and shoot to score.

```
##              Elite Guardians
## Cluster Size      1029.00000
## Cluster Mean        51.49194
## [1] "Cluster_Bests"
##        Strength      Aggression  StandingTackle         Jumping
##        77.73178        72.99028        69.83576        69.28766
##         Stamina   Interceptions         Marking HeadingAccuracy
##        68.74150        68.27697        67.62002        67.48980
##   SlidingTackle       Reactions    ShortPassing       Composure
##        67.13022        65.33236        65.23810        63.89018
##     LongPassing     BallControl       ShotPower
##        61.90087        61.49077        61.06025
## [1] "Cluster_Types"
##     CB    CDM    RCB    LCB     RB    RDM
## 0.2721 0.1749 0.1642 0.1574 0.0389 0.0379
```

Looking at this archetype it seems comparable to the Elite version of the Guardian. Mostly Backs or defensive midfielders who are not the best sprinters. They guard the entrance to the defensive zone, intercepting those with the ball in order to clear it.

```
##              Elite Defenders
## Cluster Size      1221.00000
## Cluster Mean        50.02575
## [1] "Cluster_Bests"
##        Stamina    SprintSpeed   Acceleration        Jumping        Balance
##       72.45946       71.35954       71.00983       69.81736       67.96478
##        Agility       Strength StandingTackle     Aggression  SlidingTackle
##       67.67649       66.89599       65.37510       64.27355       63.98935
##  Interceptions        Marking   ShortPassing    BallControl      Reactions
##       62.76740       62.29320       61.68878       61.22359       61.02211
## [1] "Cluster_Types"
##     RB     LB     CB    CDM    RCB    LCB
## 0.3153 0.2842 0.0942 0.0885 0.0508 0.0450
```

As we just saw the Elite Guardians it is no surprise that the next defensive archetype is an elite version of the Defender. They have mostly defensive skills, and enough stamina and sprinting to chase down invading players. Lastly the players all come from the defensive position on the field.

```
##              Elite Goalie
```

```
## Cluster Size   1134.00000
## Cluster Mean     35.01095
## [1] "Cluster_Bests"
##    GKReflexes      GKDiving GKPositioning    GKHandling     GKKicking
##      70.61640      69.54409      67.67901      66.64638      65.05291
##     Reactions      Strength       Jumping
##      64.77690      64.43739      63.05203
## [1] "Cluster_Types"
##  GK  CB  LB  RB LWB RWB
##   1   0   0   0   0   0
```

We go a bit out of order to bring us the last elite, the elite goalie. Worth noting is because goalies are only good at goalie skills they have the lowest average for clusters. Despite this these Goalies are adept at all the attributes a goalie needs like diving reactions and jumping.

This ends our group of elite archetypes. Compared with Allstars we see that most of the time the archetypes stayed similar. We saw Allstar Finishers split into Finishers and Strikers, those who are good at out dribbling defenders to score and those good at shooting well to score. We also added the Goalie archetype.

Roleplayers

From here on we go to Roleplayers. Roleplayers are the more average soccer pro soccer player, but they still make sizable contributions to any team. As the archetypes get worse less and less notable things can be said about them, so they become more difficult to distinguish.

```
##              RolePlaying Midfielder
## Cluster Size             1219.00000
## Cluster Mean               49.67304
## [1] "Cluster_Bests"
##      Balance Acceleration      Agility  SprintSpeed      Stamina
##     69.41509     68.04266     67.44381     67.35439     65.12387
## ShortPassing  BallControl      Jumping    Dribbling
##     63.53322     61.90812     60.89664     60.50287
## [1] "Cluster_Types"
##     CM     RM    CDM    CAM     LB     LM
## 0.3692 0.0853 0.0829 0.0829 0.0714 0.0697
```

We see an archetype composed of an assortment of midfielders, and a few backs. They are mostly good at running, as midfielders tend to be, and dribbling. As they are not very notable for anything other then generic midfielder skills we call them the Roleplaying Midfielder.

```
##              RolePlaying Forwards
## Cluster Size            928.00000
## Cluster Mean             48.52298
## [1] "Cluster_Bests"
##        Strength     SprintSpeed         Jumping    Acceleration
##        74.80172        69.27586        68.00431        67.45043
##       ShotPower       Finishing     Positioning         Stamina
##        67.34267        66.46875        65.69073        65.54526
## HeadingAccuracy     BallControl         Agility       Dribbling
##        65.52802        63.77909        63.64655        62.21875
##       Penalties       Reactions
##        61.70582        61.33297
## [1] "Cluster_Types"
##     ST     LS     RS     LM     RM     LW
## 0.7446 0.0700 0.0647 0.0366 0.0280 0.0194
```

The next cluster contains a lot of strikers and wings, who are fast and good at finishing. This looks similar to the Finisher and Striker categories, but unlike the elite categories that favor sprinting or running this player looks like they have the Finishers shooting, and the Strikers running and ball control. They are still offensive, but have lost specialization. So we call them RolePlaying Forwards.

```
##              RolePlaying Backs
## Cluster Size         1184.0000
## Cluster Mean           45.8483
## [1] "Cluster_Bests"
##        Strength      Aggression  StandingTackle         Jumping
##        78.77703        69.61233        69.10726        68.83530
## HeadingAccuracy         Marking   Interceptions   SlidingTackle
##        68.30405        66.92399        66.49662        66.31757
##         Stamina       Reactions       Composure
##        63.69595        62.31334        60.09459
## [1] "Cluster_Types"
##     CB    RCB    LCB    CDM     LB     RB
## 0.4823 0.2255 0.2221 0.0253 0.0169 0.0169
```

As before we see another Roleplaying archetype with the notable attributes for any Back player. The cluster is also composed of Back players. So we call these players Roleplaying Backs.

```
##              RolePlaying Goalie
## Cluster Size          891.00000
## Cluster Mean           28.39318
## [1] "Cluster_Bests"
## GKReflexes
##   60.35578
## [1] "Cluster_Types"
##  GK  CB  LB  RB LWB RWB
##   1   0   0   0   0   0
```

Our last Roleplayers are the goalies. It should be noted that only two goalie clusters exist, so it is likely that many of these goalies are also bench level goalies.

That concludes the roleplayer archetypes. As the players become less skilled and less can be said about them we have to call them broader and broader terms. We switched from things like striker and defender to just the position, indicating that these positions have the skills you would expect from their position at nothing else surprising.

Bench

Lastly we will look at the worst players, who we call bench players. As very few attributes will be large not much can be said on these archetypes.

```
##              Bench Wings
## Cluster Size   986.00000
## Cluster Mean    45.12135
## [1] "Cluster_Bests"
##      Balance Acceleration  SprintSpeed      Agility    Dribbling
##     73.18864     72.10142     71.04665     70.02840     61.80527
##  BallControl
##     60.43408
## [1] "Cluster_Types"
##     RM     LM    CAM     LW     RW     CM
## 0.2465 0.2221 0.1917 0.0852 0.0842 0.0771
```

Midfielders and wings with some skills in dribbling, we call the Bench Wings.

```
##              Bench Chaser
## Cluster Size    912.00000
## Cluster Mean     43.96898
## [1] "Cluster_Bests"
##      Balance Acceleration  SprintSpeed      Agility      Jumping
##     67.17873     66.84211     65.87390     62.29825     62.26316
##      Stamina
##     61.36513
## [1] "Cluster_Types"
##     LB     RB     CM    CDM     CB     RM
## 0.2752 0.2555 0.1491 0.1184 0.0768 0.0230
```

Backs with some sprinting ability, we call them Bench Chaser, as all they are good at is running after offenders

```
##              Bench Forward
## Cluster Size     726.00000
## Cluster Mean      42.84958
## [1] "Cluster_Bests"
##  SprintSpeed Acceleration      Balance      Jumping      Agility
##     68.25895     67.94490     65.39532     64.27410     63.53581
##    Finishing     Strength
##     60.31680     60.01653
## [1] "Cluster_Types"
##     ST     RM     LW     RW     LS     RS
## 0.8320 0.0427 0.0220 0.0207 0.0207 0.0179
```

Strikers with some basic stiker skills, slightly worse at dribbling then wings, but better shooters. We will call them Bench Forwards.

```
##              Bench Backs
## Cluster Size   855.00000
## Cluster Mean    40.45862
## [1] "Cluster_Bests"
##       Strength        Jumping StandingTackle        Stamina
##       71.39883       68.91579       62.36491       60.09825
## [1] "Cluster_Types"
##     CB    RCB    LCB     RB     LB    CDM
## 0.7439 0.1111 0.0947 0.0234 0.0199 0.0035
```

The last cluster is full of backs who are not good at running, so we will call them Bench Backs.

Our bench archetypes are vague, we split the defensive players into chasers and backs and split the offensive players into forwards and wings.

In summary it seems like players are split along multiple metrics. First they are split on whether they are offense or defense. Midfielders may specialize either way. They also split on whether they specialize or become more versatile. Midfielders are more versatile on both offense and defense, while backs and strikers are specialized. In some cases on offense we see further specialization in either dribbling, passing, or shooting to score around defenders. On defense we see some players being slow but persistent and strong, and others being faster and quick to respond in order to steal and clear the ball on defense..

Discussion

We tried some additional methods, but did have time to fully analyze them.

Standardized Data

```
standardized_fifa_19 = fifa_19_filtered[, -35] - rowMeans(fifa_19_filtered[, -35])
analysis_standardized_18 = kmeans(standardized_fifa_19, 18, iter.max = 20, nstart = 3)
```

Standardizing the data was not as simple as we expected. The player's average skill was still captured in their goal keeping statistics, since those have very low variance. It would take more thorough standardization for this to be useful.

Hierarchical Clustering

```
hclust = hclust(dist(fifa_19_filtered), method = "complete")
hclust.cut = cutree(hclust, 18)
```

We tried running hierarchical clustering, but it did not seem useful. On some machines, this was too intense to even finish running. The results did not seem very helpful, so we did not explore further.

Gaussian Mixture Models

```
predGMM = function(data, centers, cov, weights){ # Gaussian with highest contribution
  rel_data = sweep(centers, 1, data)
  dist = colSums(rel_data^2/cov)
  which.max(weights * exp(-0.5 * dist)/sqrt(apply(cov, 2, prod)))
}
```

```
analysis_GMM = GMM(fifa_19_filtered[, -35], 18)
pred_GMM = apply(fifa_19_filtered[, -35], 1, function(x){
  predGMM(x, t(analysis_GMM$centroids),
          t(analysis_GMM$covariance_matrices), analysis_GMM$weights)})
```

We wrote a Gaussian mixture model clustering, but did not have time to do a detailed analysis of it. The output looks initially promising. For one thing, there is only a single goalie cluster.

Conclusion

Our first challenge was proven when we obtained the dataset. With such a large number of variables, we had to do extensive cleaning in order to determine which variables are most useful and which are merely distractions in our efforts. We had a lot of data that we observed as NA, and many of the variable columns were unusable in our analysis due to the fact that FIFA includes erroneous information in this dataset. After careful observation and consideration, we took careful choice into which clusters were going to prove most meaningful in benefiting the results of our analysis.

Additionally, we realized that keeping a large number of variables in our analysis lead to problems when attempting to build a clustering model. We explored a number of variable selections before building a model.

After extensive analysis, we obtained valuable information from our clustering analysis regarding the categorization of player skills and archetypes. Firstly, having an "All-Star" at any position is a valuable asset to the team and the game. This is a logical conclusion given that All-Stars are archetypes based on their ability to be stellar at their individual positions. It would make sense that allowing the best possible player to be able to perform on the field would be beneficial for your respective team to win the game. When observing offense, we learned that offensive action can be executed in 1 of 2 ways: either the "Playmakers" execute an action plan and collaborate with the "Finishers" who continually pass across the field, or one very skilled dribbler is elected for his ability to outrun all other opponents. Similarly to the opportunities on offense, defenders can also be categorized into 2 different archetypes. There are defenders who are able to intercept and outrun the opponents, and guardians who use force rather than speed to create an intended defensive action.

In terms of midfielders, they are able to specialize as either offensive or defensive due to their diverse set of skills that will benefit them in either position. Regardless to their specialty, however, endurance is a key factor in performing well as a midfielder. Finally, the most important lesson that we learned during this clustering analysis is that k-means clustering tends to group players by skill level, not necessarily how those skills are utilized. It is due to this reason that observing the different clusters allows us to see that some clusters are just simply good or bad based on the overall skill of their included players.

After the completion of this project, we have reflected upon various methods that would improve this clustering analysis, if we were to complete this analysis again with more time and resources. First, we could utilize a Gaussian Model Mixture. This method would allow us to model oblong clusters within the data, and we could then use these more specific clusters to observe the variance of each trait in our analysis. Secondly, our current observations of the group indicate that due to the data being numeric, groups tend to be split based on skill level instead of variety of skill. If we could perform a good level of normalization on the data and each of the players, perhaps this issue could be resolved.

Overall, we obtained much information from our clustering analysis of FIFA player positions, and each group member is satisfied with our findings.