

빅데이터 분석 (c) 미니 프로젝트

FIFA 21 데이터셋을 활용한 데이터 분석

TEAM **ballon d'or**

김선홍 장소영

INDEX

01

주제 선정 배경 및 팀 구성

- 주제 소개 및 선정 배경
- 팀구성 및 역할

02

데이터 분석

- 데이터 수집 및 전처리
- EDA
- UEFA 리그 포인트
- 발롱도르
- Overall 예측
- 원발잡이 축구선수

03

느낀점

01

주제 선정 배경 및 팀 구성

- 주제 소개 및 선정 배경
- 팀 구성 및 역할

1. 주제 선정 배경 및 팀 구성

주제 소개 및 선정 배경

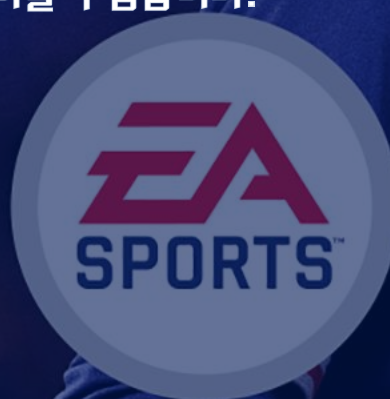
FIFA는 실제 축구를 반영하는 대표적인 축구 게임

EA sports는 총 6000명 이상의 인력을 투입하여,
선수의 능력치에 대한 데이터를 수집합니다.

주제 선정에 대해 이야기를 나누다 팀원 모두가 축구에 많은 관심을
가지고 있는 공통점을 발견했습니다.

미니프로젝트인 만큼 좋아하는 분야에 대해 프로젝트를
진행하는 것이 좋을 것 같아

비교적 정형화되어 있는 FIFA 21 게임의 선수 데이터 셋을 통해
평소에 궁금했던 내용을 토대로 분석을 진행하였습니다.



FIFA 21



1. 주제 선정 배경 및 팀 구성

팀 구성



김선홍

데이터 수집 및 전처리
EDA 분석
원발잡이 축구선수



장소영

데이터 수집 및 전처리
EDA 분석
UEFA
Ballondor
PCA Overall

02

데이터 분석

- 데이터 수집 및 전처리
- EDA
- UEFA 리그 포인트
- 발롱도르
- Overall 예측
- 왼발잡이 축구선수

2. 데이터 분석

데이터 수집

The Kaggle logo, featuring the word "kaggle" in a light blue, lowercase, sans-serif font.

<https://www.kaggle.com/bryanb/fifa-player-stats-database>

FIFA 21의 선수 데이터

The Selenium logo, consisting of a green square with a white checkmark and the letters "Se" in white, followed by the word "Selenium" in a green, sans-serif font.

https://en.wikipedia.org/wiki/2019_Ballon_d%27Or
<https://www.uefa.com/memberassociations/uefarankings/country/#/yr/2021>

발롱도르 수상 순위, 포인트 데이터
UEFA 클럽 랭킹 데이터

2. 데이터 분석

EDA 탐색적 데이터 분석

FIFA 21 게임 데이터셋의 EDA 분석

본격적인 데이터 분석에 앞서 탐색적 데이터 분석을 통해 해당 데이터의 탐색과 이해를 높였습니다.

Cofn	Terminal	Jobs							
C:\Users\jefaz1> cd C:\Program Files\Foxit Software\Foxit Reader									
[list output truncated]									
summarize(faz2)									
soffid_id	player_url	short_name	long_name	age	dob	height_cm	weight_lb		
Min.: 41	Length:18944	Length:18944	Length:18944	Min.: 16.00	Length:18944	Min.: 115.0	Min.: 50.00		
Max.: 12187	Class :character	Class :character	Class :character	1st Qu.: 21.00	Class :character	1st Qu.: 175.0	1st Qu.: 70.00		
Median :23215	Mode :character	Mode :character	Mode :character	Median :25.00	Mode :character	Median :181.0	Median : 75.00		
Mean :22642				Mean :25.23		Mean :181.2	Mean : 75.00		
3rd Qu.:24076				3rd Qu.: 29.00		3rd Qu.:188.0	3rd Qu.: 80.00		
NA's :259570				NA's :53.00		NA's :206.0	NA's :110.00		
nationality	club_name	league_name	league_rank	overall	potential	value_eur	wage_eur		
Length:18944	Length:18944	Length:18944	Min.: 1.000	Min.: 47.00	Min.: 47.00	Min.: 0	Min.: 0		
Class :character	Class :character	Class :character	1st Qu.:1.000	1st Qu.:61.00	1st Qu.:67.00	1st Qu.: 300000	1st Qu.: 1000		
Mode :character	Mode :character	Mode :character	Median :1.000	Median :65.00	Median :71.00	Median : 350000	Median :1500		
			Mean :1.357	Mean :65.08	Mean :71.00	Mean : 222483.3	Mean : 8676		
			3rd Qu.:1.000	3rd Qu.:70.00	3rd Qu.:75.00	3rd Qu.:1000000	3rd Qu.: 7000		
			NA's :225	NA's :93.00	NA's :95.00	NA's :105500000	NA's :500000		
player_positions	preferred_foot	international_reputation	weak_foot	skill_moves	work_rate	body_type	real_face		
Length:18944	Length:18944	Min.: 1.000	Min.: 1.000	Min.: 1.000	Length:18944	Length:18944	Length:18944		
Class :character	Class :character	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:1.000	Class :character	Class :character	Class :character		
Mode :character	Mode :character	Median :1.000	Median :1.000	Median :1.000	Mode :character	Mode :character	Mode :character		
		Mean :1.052	Mean :1.937	Mean :1.363					
		3rd Qu.:1.000	3rd Qu.:1.000	3rd Qu.:1.000					
		NA's :5.000	NA's :5.000	NA's :5.000					
release_clause	play_tags	team_position	team_jersey_number	loaned from	joined	contract_until	until_national_position		
Min.: 9900	Length:18944	Length:18944	Min.: 1.00	Length:18944	Length:18944	Min.: 2020	Length:18944		
1st Qu.: 525000	Class :character	Class :character	1st Qu.: 9.00	Class :character	Class :character	1st Qu.:2021	Class :character		
Median :100000	Mode :character	Mode :character	Median :15.00	Mode :character	Mode :character	Median :2022	Mode :character		
Mean :429633			Mean :10.59			Mean :2022			
3rd Qu.:320000			3rd Qu.:27.00			3rd Qu.:2023			
NA's :995			NA's :225			NA's :2020			
nation_jersey_number	pace	shooting	passing	dribbling	defending	physic	gk_diving	gk_handling	
Min.: 25.00	Min.: 16.00	Min.: 25.00	Min.: 25.00	Min.: 15.00	Min.: 25.00	Min.: 58.00	Min.: 45.00	Min.: 58.00	
1st Qu.: 6.00	1st Qu.:62.00	1st Qu.:42.00	1st Qu.:50.00	1st Qu.:57.00	1st Qu.:36.00	1st Qu.:58.00	1st Qu.:60.00	1st Qu.:58.00	
Median :67.00	Median :58.00	Median :54.00	Median :58.00	Median :54.00	Median :56.00	Median :65.00	Median :65.00	Median :65.00	
Mean :67.67	Mean :52.77	Mean :57.14	Mean :62.47	Mean :51.32	Mean :58.83	Mean :65.16	Mean :65.16	Mean :65.16	
3rd Qu.:12.00	3rd Qu.:75.00	3rd Qu.:63.00	3rd Qu.:64.00	3rd Qu.:69.00	3rd Qu.:64.00	3rd Qu.:72.00	3rd Qu.:70.00	3rd Qu.:63.00	
Max.: 27.00	Max.: 96.00	Max.: 93.00	Max.: 93.00	Max.: 91.00	Max.: 91.00	Max.: 90.00	Max.: 90.00	Max.: 92.00	
NA's :2083	NA's :2083	NA's :2083	NA's :2083	NA's :2083	NA's :2083	NA's :18661	NA's :18661	NA's :18661	
nk_kicking	gk_reflexes	gk_speed	gk_positioning	player_traits	attacking_crossing	attacking_finishing	attacking_heading_accuracy		
Min.: 135.00	Min.: 44.00	Min.: 12.0	Min.: 138.00	Length:18944	Min.: 6.00	Min.: 3.0	Min.: 5.00		
1st Qu.:157.00	1st Qu.:60.00	1st Qu.:57.00	1st Qu.:57.00	Class :character	1st Qu.:30.00	1st Qu.:30.00	1st Qu.:30.00		
Median :61.00	Median :65.00	Median :37.0	Median :63.00	Mode :character	Median :54.00	Median :49.0	Median :55.00		
Mean :61.72	Mean :66.11	Mean :37.2	Mean :63.17		Mean :49.61	Mean :45.0	Mean :51.07		
3rd Qu.:156.00	3rd Qu.:72.00	3rd Qu.:45.0	3rd Qu.:69.00		3rd Qu.:63.00	3rd Qu.:62.0	3rd Qu.:64.00		
Max.: 93.00	Max.: 96.00	Max.: 65.0	Max.: 91.00		Max.: 94.00	Max.: 95.0	Max.: 93.00		
NA's :18661	NA's :18661	NA's :18661	NA's :18661						
attacking_sprint_passing	attacking_volley	attacking_dribbling	skill_curve	skill_fk_accuracy	skill_long_passing	skill_ball_control	movement_acceleration		
Min.: 7.00	Min.: 3.00	Min.: 5.00	Min.: 4.00	Min.: 5.00	Min.: 5.00	Min.: 5.00	Min.: 13.00		
1st Qu.:154.00	1st Qu.:130.00	1st Qu.:49.00	1st Qu.:135.00	1st Qu.:121.00	1st Qu.:43.00	1st Qu.:154.00	1st Qu.:157.00		
Median :62.00	Median :44.00	Median :41.00	Median :61.00	Median :56.00	Median :56.00	Median :63.00	Median :57.00		
Mean :58.75	Mean :42.67	Mean :35.55	Mean :47.19	Mean :43.36	Mean :52.65	Mean :50.48	Mean :46.29		
3rd Qu.:160.00	3rd Qu.:156.00	3rd Qu.:66.00	3rd Qu.:61.00	3rd Qu.:55.00	3rd Qu.:64.00	3rd Qu.:69.00	3rd Qu.:74.00		
NA's :94.00	NA's :96.00	NA's :96.00	NA's :94.00	NA's :95.00	NA's :93.00	NA's :96.00	NA's :97.00		
movement_sprint_speed	movement_agility	movement_reactions	movement_balance	power_shot	power_punching	power_stamina	power_strength		
Min.: 157.00	Min.: 157.00	Min.: 157.00	Min.: 157.00	Min.: 157.00	Min.: 157.00	Min.: 157.00	Min.: 157.00		
1st Qu.:157.00	1st Qu.:155.00	1st Qu.:156.00	1st Qu.:156.00	1st Qu.:158.00	1st Qu.:158.00	1st Qu.:155.0	1st Qu.:157.00		
Median :67.00	Median :66.00	Median :62.00	Median :66.00	Median :59.00	Median :65.00	Median :66.00	Median :66.00		
Mean :62.00	Mean :65.33	Mean :61.61	Mean :57.73	Mean :64.56	Mean :64.56	Mean :64.56	Mean :64.56		
3rd Qu.:174.00	3rd Qu.:174.00	3rd Qu.:174.00	3rd Qu.:174.00	3rd Qu.:168.00	3rd Qu.:173.00	3rd Qu.:173.0	3rd Qu.:174.00		
Max.: 96.00	Max.: 96.00	Max.: 95.00	Max.: 97.00	Max.: 95.00	Max.: 95.00	Max.: 97.0	Max.: 97.00		
power_long_shots	mentality aggression	mentality interceptions	mentality positioning	mentality vision	mentality mental penalties	mentality composure			
Min.: 4.00	Min.: 3.00	Min.: 3.00	Min.: 2.00	Min.: 9.00	Min.: 6.00	Min.: 12.00			
1st Qu.:157.00	1st Qu.:125.00	1st Qu.:125.00	1st Qu.:125.00	1st Qu.:125.00	1st Qu.:125.00	1st Qu.:125.00			
Median :51.00	Median :58.00	Median :52.00	Median :55.00	Median :55.00	Median :49.00	Median :59.00			
Mean :46.76	Mean :55.49	Mean :46.25	Mean :50.26	Mean :53.83	Mean :40.05	Mean :57.08			
3rd Qu.:62.00	3rd Qu.:64.00	3rd Qu.:64.00	3rd Qu.:64.00	3rd Qu.:64.00	3rd Qu.:60.00	3rd Qu.:66.00			
NA's :96.00	NA's :96.00	NA's :91.00	NA's :96.00	NA's :95.00	NA's :96.00	NA's :96.00			

defending	standing	tackle	defending	sliding	tackle	goalkeeping	diving	goalkeeping	handling	goalkeeping	kicking	goalkeeping	positioning
Mode : logical	Min. : 5.00	Min. : 7.00	Min. : 5.00	Min. : 6.00	Min. : 8.00	Min. : 1.00	Min. : 1.00	Min. : 1.00	Min. : 1.00	Min. : 1.00	Min. : 1.00	Min. : 1.00	Min. : 1.00
NA's : 19944	1st Qu.: 27.00	1st Qu.: 24.00	1st Qu.: 27.00	1st Qu.: 24.00	1st Qu.: 27.00	1st Qu.: 8.00	1st Qu.: 8.00	1st Qu.: 8.00	1st Qu.: 8.00	1st Qu.: 8.00	1st Qu.: 8.00	1st Qu.: 8.00	1st Qu.: 8.00
	Median : 53.00	Median : 52.00	Median : 53.00	Median : 52.00	Median : 53.00	Median : 11.00	Median : 11.00	Median : 11.00	Median : 11.00	Median : 11.00	Median : 11.00	Median : 11.00	Median : 11.00
	Mean : 47.50	Mean : 45.55	Mean : 47.50	Mean : 45.55	Mean : 47.50	Mean : 16.45	Mean : 16.24	Mean : 16.24	Mean : 16.1	Mean : 16.1	Mean : 16.23	Mean : 16.23	Mean : 16.23
	3rd Qu.: 65.00	3rd Qu.: 60.00	3rd Qu.: 65.00	3rd Qu.: 60.00	3rd Qu.: 65.00	3rd Qu.: 14.00	3rd Qu.: 14.00	3rd Qu.: 14.00	3rd Qu.: 14.00	3rd Qu.: 14.00	3rd Qu.: 14.00	3rd Qu.: 14.00	3rd Qu.: 14.00
	Max. : 99.00	Max. : 99.00	Max. : 99.00	Max. : 99.00	Max. : 99.00	Max. : 22.00	Max. : 22.00	Max. : 22.00	Max. : 22.00	Max. : 22.00	Max. : 22.00	Max. : 22.00	Max. : 22.00

goalkeeping_reflexes	ls	st	rs	lw	lf	cf	rf
Min. : 1.00	Length:18944	Length:18944	Length:18944	Length:18944	Length:18944	Length:18944	Length:18944
1st Qu.: 8.00	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character
Median :11.00	Node :character	Node :character	Node :character	Node :character	Node :character	Node :character	Node :character
Mean :16.55							
3rd Qu.:14.00							
Max. :99.00							

rw	lan	cen	ram	ln	lcn	cn	rqn
Length:18944	Length:18944	Length:18944	Length:18944	Length:18944	Length:18944	Length:18944	Length:18944
Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character
Node :character	Node :character	Node :character	Node :character	Node :character	Node :character	Node :character	Node :character

rm	lub	ldn	cdm	rdm	nub	lb	lcb
Length:18944	Length:18944	Length:18944	Length:18944	Length:18944	Length:18944	Length:18944	Length:18944
Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character
Node :character	Node :character	Node :character	Node :character	Node :character	Node :character	Node :character	Node :character

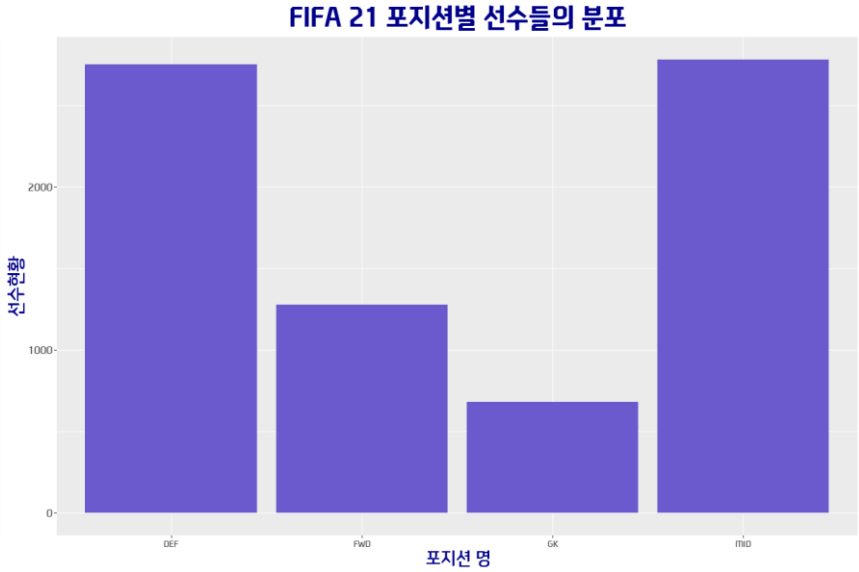
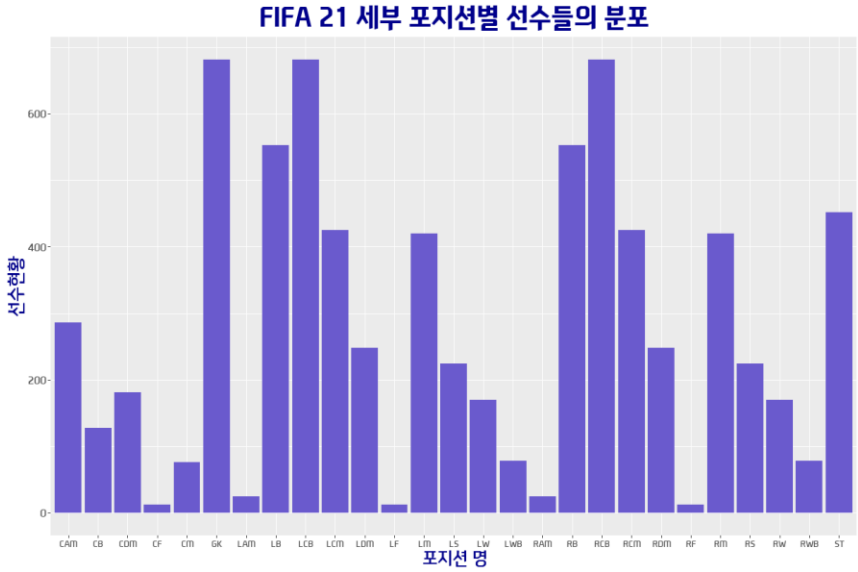
cb	rcb	rb
Length:18944	Length:18944	Length:18944
Class :character	Class :character	Class :character
Node :character	Node :character	Node :character

선수의 정보(국적, 소속 클럽, 소속 리그, 나이, 몸무게, 키)와

FIFA 21에서의 오버롤, 세부 능력치로 구성되어 있는 데이터셋임을 확인했습니다.

2. 데이터 분석

EDA 탐색적 데이터 분석

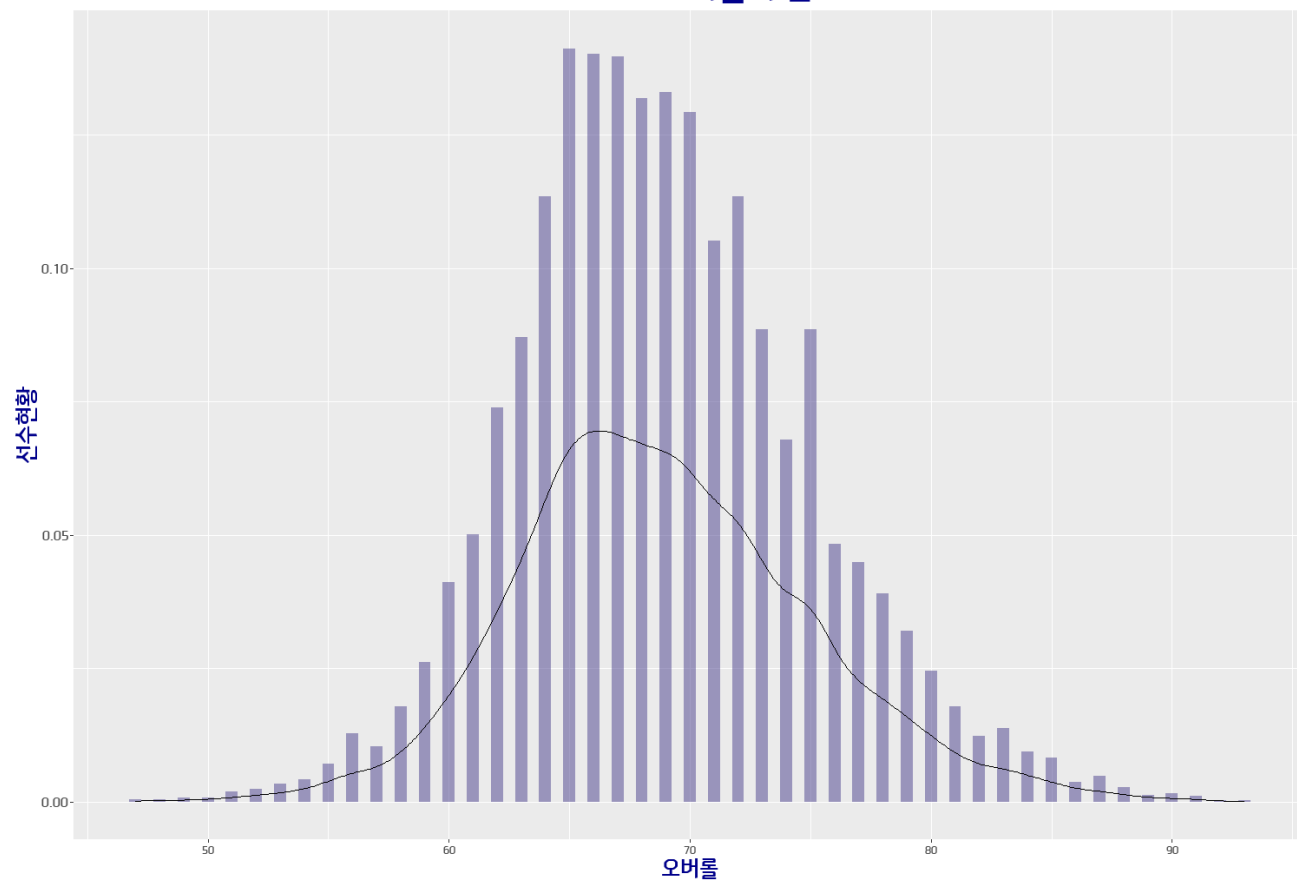


FIFA 21은 공격수, 수비수, 미드필더, 골키퍼의 4종류의 포지션이 아니라 구체적인 선수의 포지션이 정해져 있는 게임인 것을 확인할 수 있었습니다. 또한 은퇴 선수, 후보 선수, 2군 선수의 데이터까지도 포함되어 있는 것을 확인하였습니다. 분석의 목적에 따라, 은퇴 선수, 후보 선수, 2군 선수는 데이터에서 제외하였으며, 분석을 용이하게 작업하기 위해 다음과 같이 포지션을 재그룹화해 데이터 전처리를 수행했습니다.

2. 데이터 분석

EDA 탐색적 데이터 분석

FIFA 21 오버롤의 분포



FIFA 21에서 '오버롤'이란 선수의 평균적인 능력치를 뜻합니다.

FIFA21 데이터셋의 모든 선수의 오버롤을

히스토그램과 커널밀도함수(KDE)로 나타내보았습니다.

그 결과, 다음과 같이 평균에 가까울수록 많은 분포값을 보이는

종 모양의 형태를 보이는 것을 확인할 수 있었습니다.

즉, 능력치의 수준이 높아지거나 낮아질수록

선수의 분포는 더 작아지는 것을 확인할 수 있었습니다.

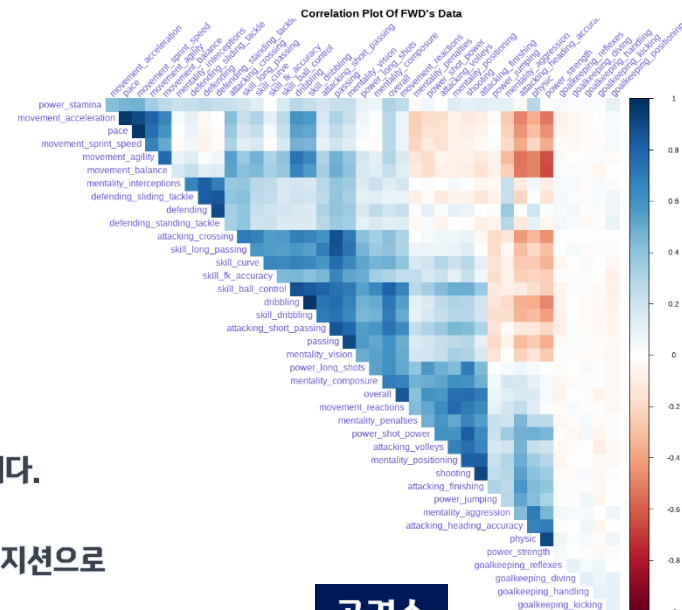
2. 데이터 분석

EDA 탐색적 데이터 분석

FIFA 21에서 선수의 포지션은 다양하게 구성되어 있습니다.

따라서 공격수, 미드필더, 수비수, 골키퍼의 4가지 대분류의 포지션으로

재 분류하는 작업을 시행했습니다.



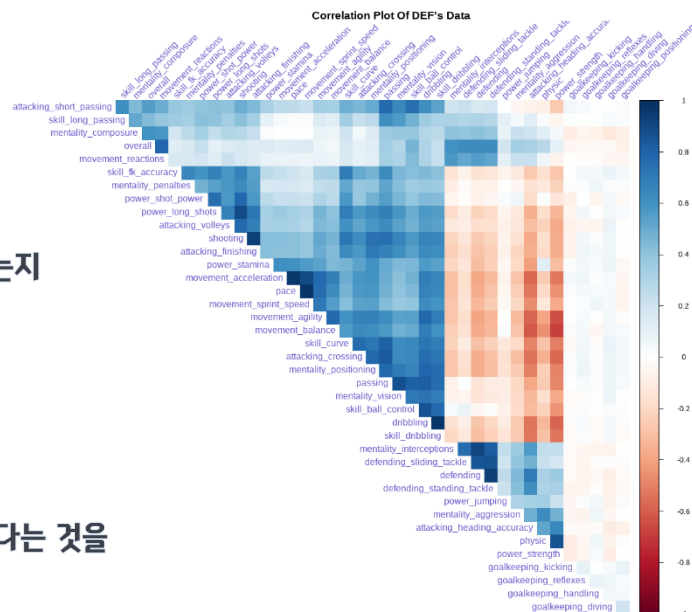
그리고 각 포지션별로 재분류하여

각각의 세부 능력치들이 어떠한 상관관계를 가지고 있는지

파악했습니다.

그 결과, 다음과 같이 포지션별로 상관관계의 양상이 다양하다는 것을

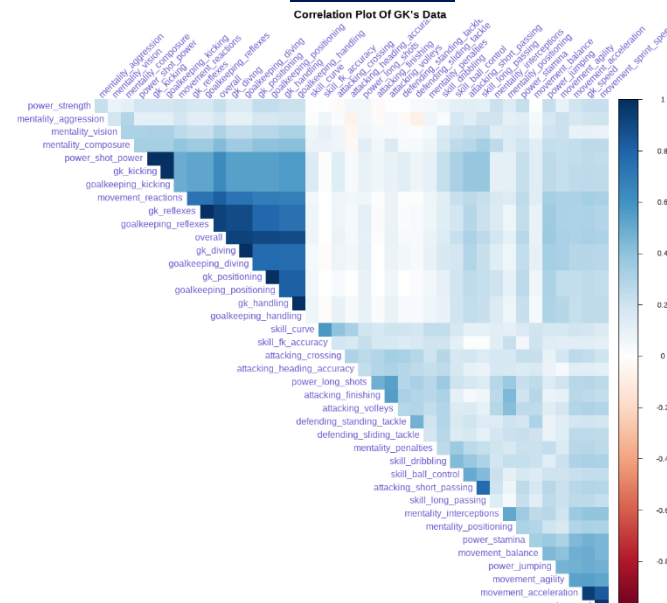
파악할 수 있었습니다.



수비수



미드필더



골키퍼

2. 데이터 분석

분석 주제 1

UEFA 리그 포인트와 선수 Overall은 상관관계가 있을까?

UEFA는 '리그포인트'제도를 통해 챔피언스 리그, 유로파 리그,

유로파 콘퍼런스 리그의 팀 수를 정합니다.

이러한 리그 포인트는 그 리그의 경쟁력을 보여줄 수 있는 지표입니다.

따라서 UEFA 리그 포인트와 FIFA21 게임에서 선수들의 Overall의

상관관계를 분석했습니다.

Pearson's product-moment correlation

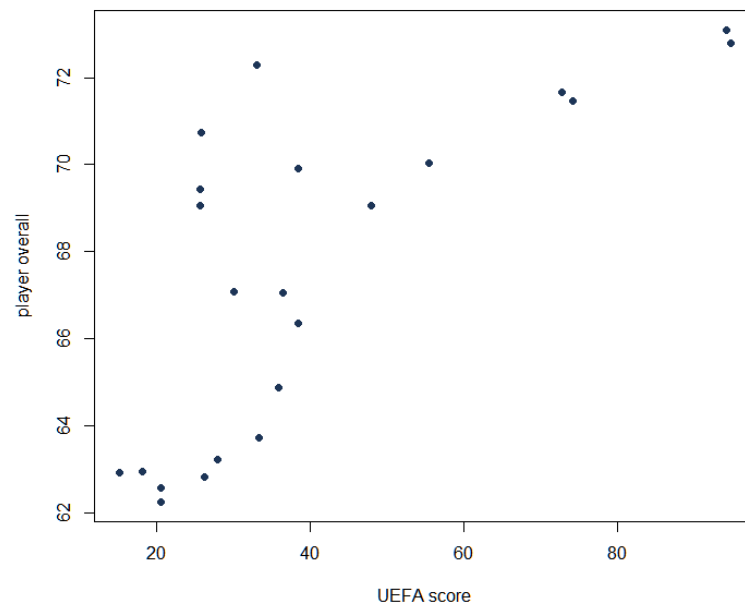
$t = 4.7078$, $df = 20$, $p\text{-value} = 0.000135$

$cor: 0.7250179$

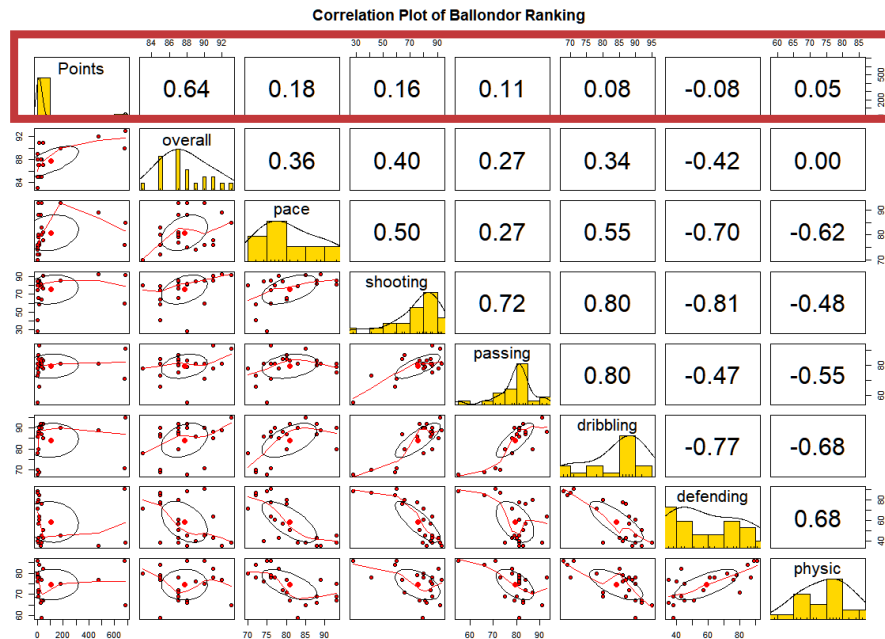
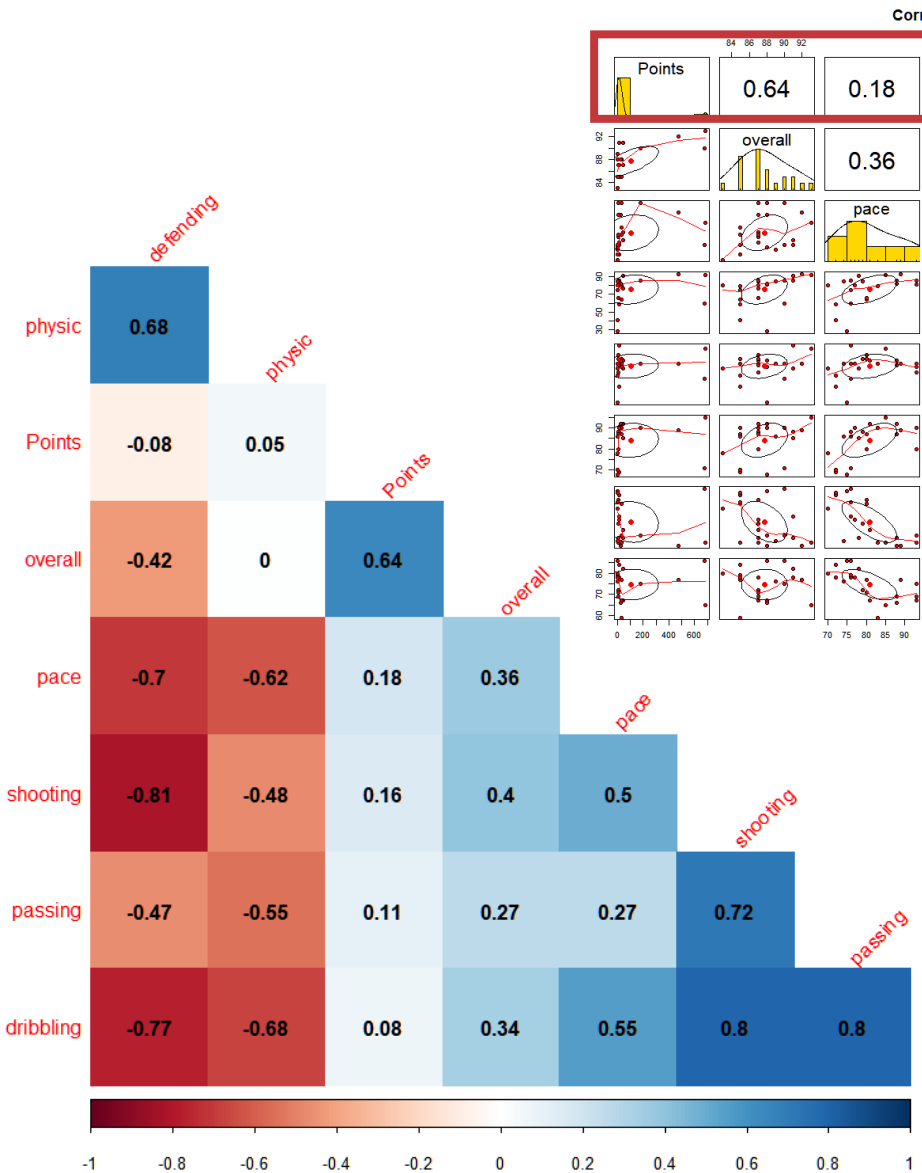
Country coefficients

		2017/18	2018/19	2019/20	2020/21		
Pos	Country	16/17	17/18	18/19	19/20	Pts	Clubs
1	England	14.928	20.071	22.642	18.571	19.500	95.712
2	Spain	20.142	19.714	19.571	18.928	17.357	95.712
3	Italy	14.250	17.333	12.642	14.928	15.428	74.581
4	Germany	14.571	9.857	15.214	18.714	14.928	73.284
5	France	14.416	11.500	10.583	11.666	7.416	55.581
6	Portugal	8.083	9.666	10.900	10.300	9.200	48.149
7	Netherlands	9.100	2.900	8.600	9.400	9.000	39.000
8	Russia	9.200	12.600	7.583	4.666	4.333	38.382
9	Belgium	12.500	2.600	7.800	7.600	6.000	36.500
10	Austria	7.375	9.750	6.200	5.800	6.700	35.825
11	Scotland	4.375	4.000	6.750	9.750	8.500	33.375
12	Ukraine	5.500	8.000	5.600	7.200	6.800	33.100

UEFA League score and Overall of Players



상관계수가 +0.73, P-value가 0.0001로 리그 포인트와 Overall은
강한 상관관계가 있음을 확인할 수 있습니다.



발롱도르는 '황금빛 공'이라는 뜻을 가진 프랑스 축구 전문지에서

주관하는 올해의 유럽 남자 축구 선수에게 수여하는 상입니다.

발롱도르의 순위와 선수의 능력치 상관관계를 분석했습니다.

발롱도르 순위는 발롱도르 포인트를 통해 정해지기 때문에

발롱도르 점수와 선수 능력치의 수치형 데이터간의 상관분석을

진행했습니다.

(단, 2020년의 발롱도르의 경우 코로나의 영향으로 진행되지 않아 2019년의 데이터를 사용했습니다.)

발롱도르 점수는 선수의 전반적인 능력치 평가인 Overall과 0.64로 비교적 강한 양의 상관관계를 보입니다.

반면, 스탯(페이스,슈팅,패스,드리블,수비,체력)과는 약한 상관관계 혹은 거의 관계가 없음을 확인할 수 있습니다.

발롱도르 순위와 선수 능력치는 상관관계가 있을까?

2. 데이터 분석

분석 주제 2

2. 데이터 분석

분석 주제 3

FIFA21 세부 능력치의 주성분(PCA) 분석과 Overall 예측

attacking_crossing	attacking_finishing	attacking_heading_accuracy	attacking_short_passing	attacking_volleys	skill_dribbling	skill_curve	
84	95	90	82	86	88	81	
71	94	85	84	89	85	79	
85	87	62	87	87	95	88	
78	91	73	83	83	92	79	
	movement_balance	power_shot_power	power_jumping	power_stamina	power_strength	power_long_shots	mentality_aggression
	71	94	95	84	78	93	63
	82	89	84	76	86	85	81
	83	80	62	81	50	84	51
	82	86	77	86	76	79	62
skill_fk_accuracy	skill_long_passing	skill_ball_control	movement_acceleration	movement_sprint_speed	movement_agility	movement_reactions	
76	77	92	87	91	87	95	
85	70	88	77	78	77	93	
89	81	95	94	89	96	91	
63	70	90	96	96	92	92	
	mentality_interceptions	mentality_positioning	mentality_vision	mentality_penalties	mentality_composure	defending_standing_tackle	
	29	95	82	84	95	32	
	49	94	79	88	88	42	
	36	87	90	92	93	30	
	38	91	80	70	84	34	
defending_sliding_tackle	goalkeeping_diving	goalkeeping_handling	goalkeeping_kicking	goalkeeping_positioning	goalkeeping_reflexes		
24	7	11	15	14	11		
19	15	6	12	8	10		
29	9	9	15	15	11		
32	13	5	7	11	6		
38	10	10	15	7	14		

FIFA21에서 선수들의 세부 능력치는 다음과 같은 컬럼으로 다양하게 구성되어 있다.

예를 들어 크로싱, 피니싱, 정확성 등의 공격에 관한 요소와

스탠딩 태클, 슬라이딩 태클 등의 수비에 관한 요소 등

여러 요소를 통하여 선수들의 능력치가 산출됨을 확인할 수 있다.

따라서 여러 데이터들이 모여 하나의 분포를 이루는 주성분을 분석하는 방법, PCA로

고차원의 데이터를 저차원의 데이터로 환원시켜 분석을 진행하였다.

2. 데이터 분석

분석 주제 4

FIFA21 선수 데이터 특성 상

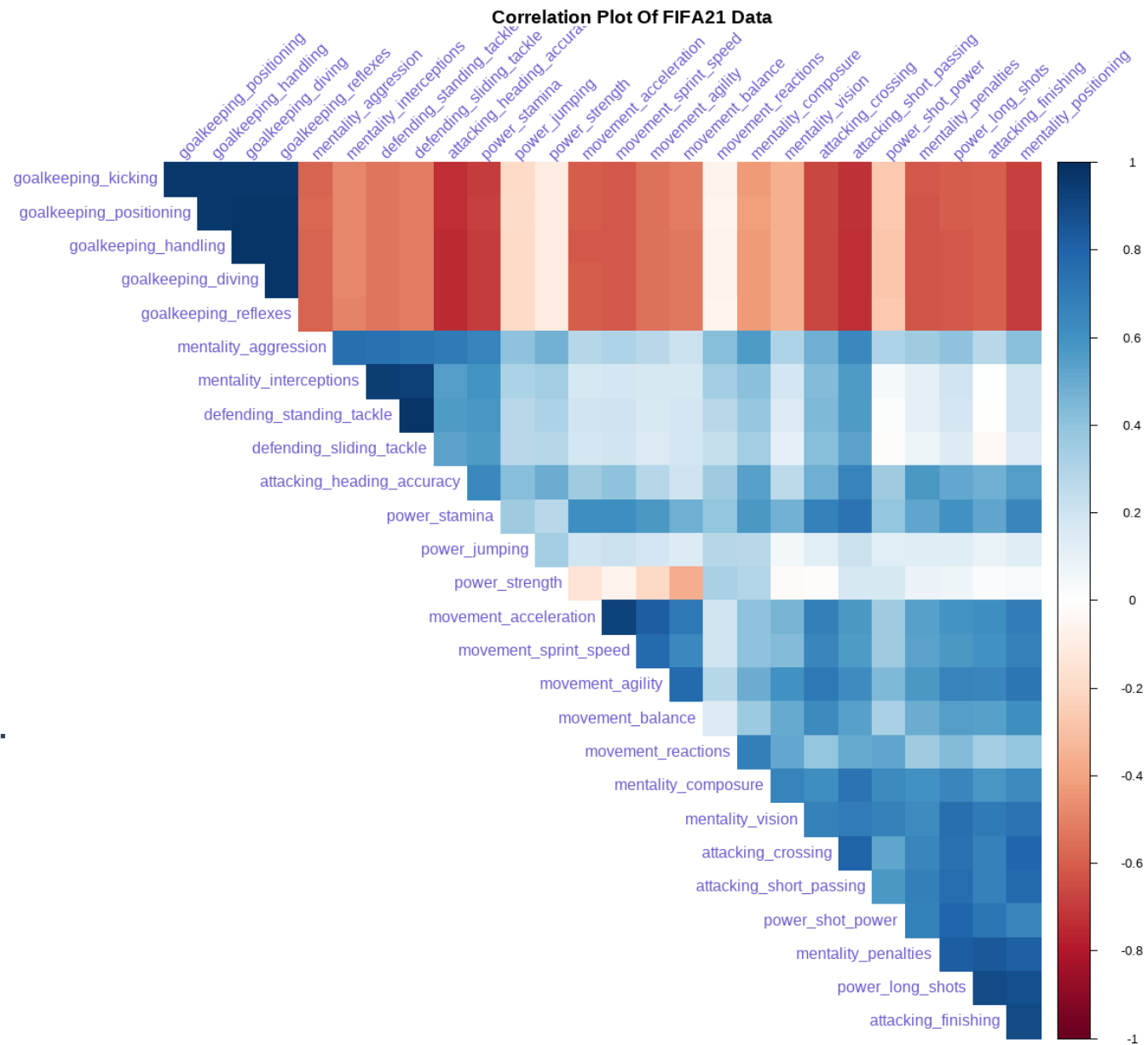
골키퍼(GK)에게만 할당되는 세부 능력은 다른 포지션(DEF,MID,FWD)에는

NA로 할당되어 있기 때문에

본 분석은 4가지 포지션으로 우선 분류를 진행하였고,

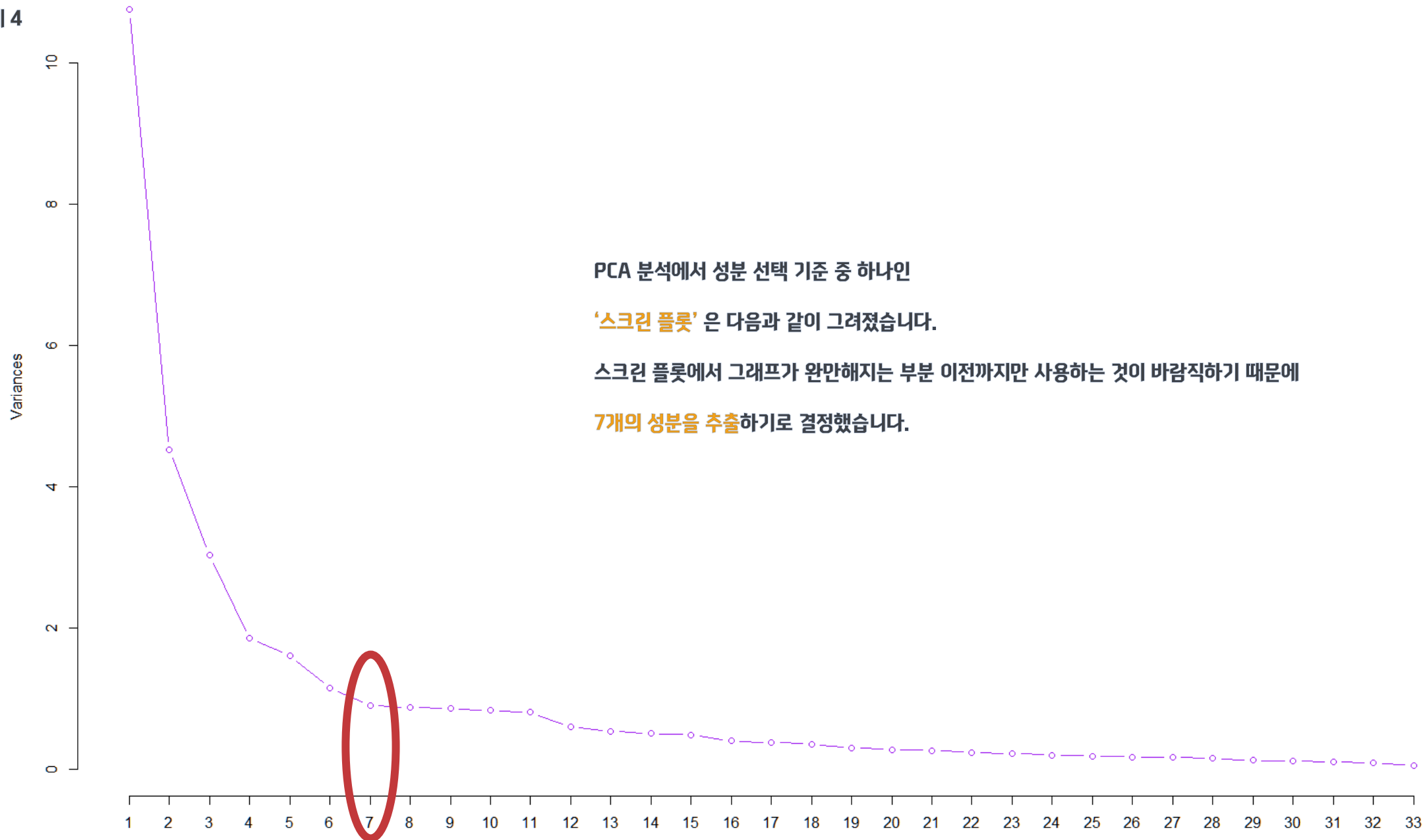
이후 골키퍼는 제외했습니다.

또한, 전체 데이터에서 포지션별로 각각 500명을 산출하여 주성분을 분석했습니다.



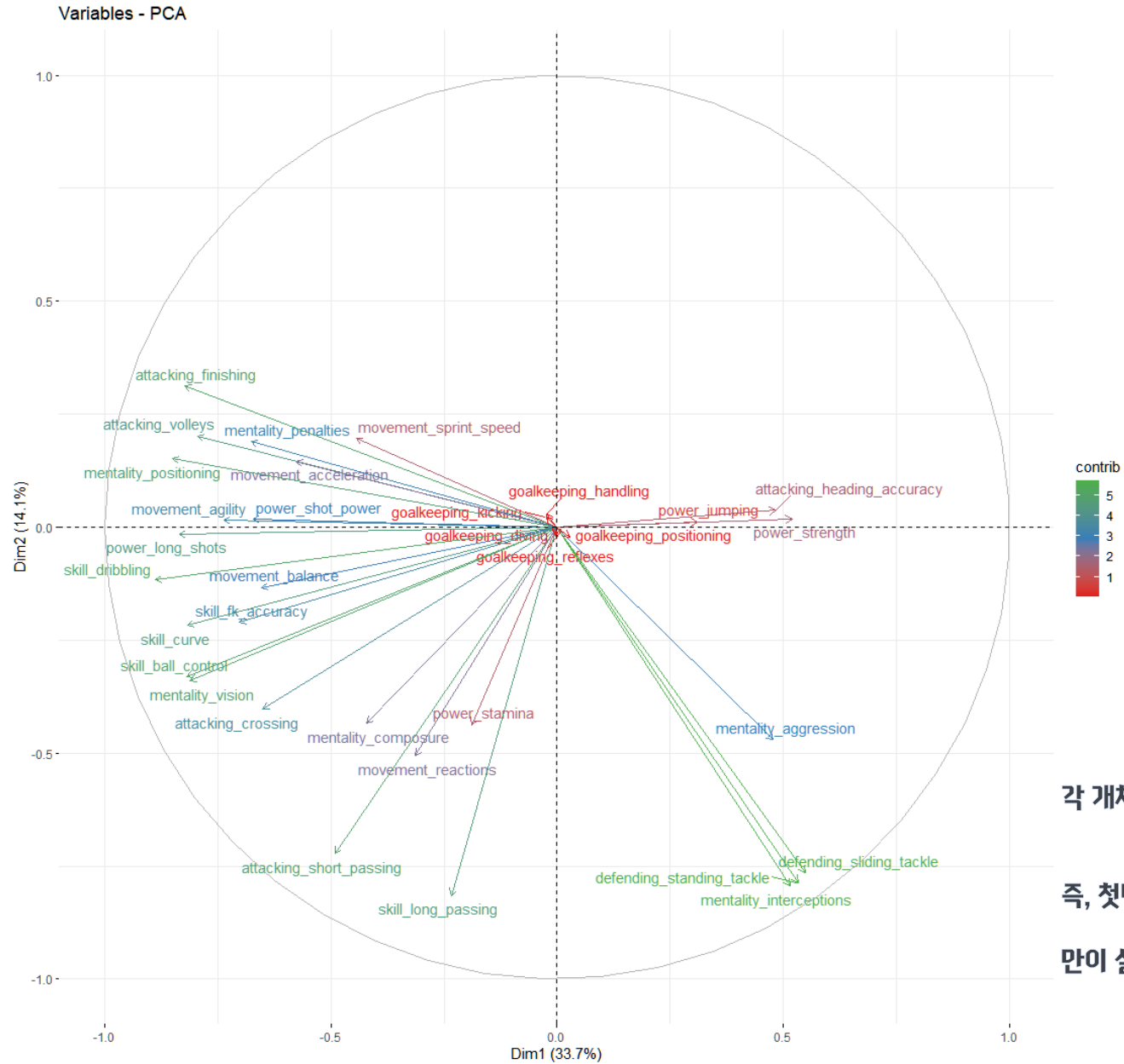
2. 데이터 분석

분석 주제 4



2. 데이터 분석

분석 주제 4



각 개체에 대한 첫번째, 두번째 주성분 행렬도(biplot)는 다음과 같습니다.

즉, 첫번째 주성분과 두번째 주성분으로 데이터의 47.8 % (14.1%+33.7%)

만이 설명 가능하다고 해석할 수 있습니다.

2. 데이터 분석

분석 주제 4

```
> p5_os <- principal(best500[,8:40], nfactors=7, rotate="varimax")
> print(loadings(p5_os), digits=2, cutoff=0.4, sort=TRUE)
```

Loadings:

	RC1	RC2	RC3	RC4	RC6	RC7	RC5
attacking_finishing	0.78	-0.47					
attacking_volleys	0.79						
skill_dribbling	0.61		0.51				
skill_curve	0.72						
skill_fk_accuracy	0.73						
skill_ball_control	0.61				0.55		
power_shot_power	0.85						
power_long_shots	0.88						
mentality_positioning	0.76						
mentality_vision	0.69						
mentality_penalties	0.74						
skill_long_passing		0.67					
mentality_aggression		0.58		-0.52			
mentality_interceptions		0.86					
defending_standing_tackle		0.87					
defending_sliding_tackle	-0.40	0.86					
movement_acceleration			0.91				
movement_sprint_speed			0.87				
movement_agility			0.78				
movement_balance			0.67				
power_stamina		0.49	0.52				
attacking_heading_accuracy				-0.77			
power_jumping				-0.71			
power_strength			-0.46	-0.72			
attacking_short_passing	0.41	0.45			0.61		
movement_reactions					0.79		
mentality_composure					0.75		
goalkeeping_diving						0.61	
goalkeeping_handling						0.79	
goalkeeping_kicking							0.72
goalkeeping_positioning							0.71
attacking_crossing	0.47		0.46				
goalkeeping_reflexes							

	RC1	RC2	RC3	RC4	RC6	RC7	RC5
SS loadings	7.50	4.37	4.29	3.13	2.63	1.16	1.16
Proportion Var	0.23	0.13	0.13	0.09	0.08	0.04	0.04
Cumulative Var	0.23	0.36	0.49	0.58	0.66	0.70	0.73

스크린 플롯을 통해 7개의 성분을 추출하기로 결정하였으므로

다음과 같이 **principal()** 함수를 이용해 분석을 진행하였습니다.

결과적으로 7개의 요인을 선택하여 분석을 진행하게 되면

하단의 **7개의 요인**으로 **그룹핑**을 할 수 있습니다.

그리고 이에 따른 **누적확률**은 약 73%임을 확인할 수 있습니다.

1요인 : 기술, 멘탈
2요인 : 수비
3요인 : 가속도
4요인 : 협응성 (힘)
5요인 : 패스
6요인 : 골키퍼 능력
7요인 : 크로싱

2. 데이터 분석

분석 주제 4

	fit	lwr	upr	V4
1	87.69496	83.66137	91.72855	92
2	85.09908	81.08218	89.11597	91
3	85.19464	81.17044	89.21884	91
4	84.37325	80.35594	88.39057	90
5	86.80058	82.77701	90.82415	90
6	85.12881	81.11188	89.14574	90
7	85.78897	81.76378	89.81416	89
8	81.91749	77.90636	85.92863	88
9	84.40376	80.38556	88.42196	88
10	83.22835	79.20483	87.25187	88
11	83.73243	79.71728	87.74759	88
12	82.55476	78.53926	86.57025	87
13	81.85090	77.84281	85.85899	87
14	83.04160	79.02659	87.05661	87
15	83.96249	79.93728	87.98769	87
16	85.71752	81.69940	89.73564	87
17	86.86597	82.84922	90.88271	87
18	82.48653	78.47188	86.50119	86
19	82.02441	78.00896	86.03986	86
20	80.44018	76.43248	84.44788	85

21	81.24861	77.23561	85.26161	85
22	82.43573	78.42236	86.44910	85
23	81.20282	77.18445	85.22118	85
24	81.05365	77.03906	85.06823	85
25	80.05950	76.03781	84.08120	85
26	80.09651	76.08815	84.10487	85
27	80.32262	76.31203	84.33320	84
28	83.28133	79.26804	87.29461	84
29	78.60658	74.60040	82.61277	84
30	80.66784	76.65977	84.67590	84
31	81.27778	77.26586	85.28969	84
32	83.57130	79.55594	87.58666	84
33	79.39088	75.37892	83.40283	84
34	82.52567	78.51275	86.53859	84
35	78.50766	74.49581	82.51950	83
36	82.13564	78.12427	86.14701	83
37	80.88103	76.85704	84.90501	83
38	80.78907	76.77065	84.80749	83
39	80.58158	76.57513	84.58803	83
40	79.62499	75.60870	83.64127	83

568	79.72020	75.70857	83.73183	82
569	80.11331	76.10368	84.12293	82
570	79.61668	75.61110	83.62225	82
571	81.39399	77.38446	85.40353	82
572	82.87154	78.85784	86.88524	82
573	81.27251	77.26768	85.27734	81
574	80.59500	76.58878	84.60123	81
575	77.73473	73.72155	81.74790	81
576	79.83033	75.82664	83.83403	81
577	78.27169	74.25604	82.28733	81
578	80.99040	76.97832	85.00248	81
579	79.99006	75.97930	84.00083	81
580	78.28602	74.27470	82.29733	81
581	83.12307	79.11592	87.13022	81
582	77.91963	73.90886	81.93040	81
583	80.86215	76.84261	84.88169	81
584	81.15162	77.14308	85.16016	81
585	80.61865	76.61326	84.62404	81
586	78.50747	74.49863	82.51632	81
587	81.53410	77.52708	85.54111	81

주성분 분석을 통한 능력치와 실제 Overall을 비교하였을 때

어느정도 유사함을 확인할 수 있습니다.
(하한과 상한 사이에 fit이 존재함)

반면, 상위의 Overall의 경우에는 실제의 값이 더 높은 경향을 보이는데

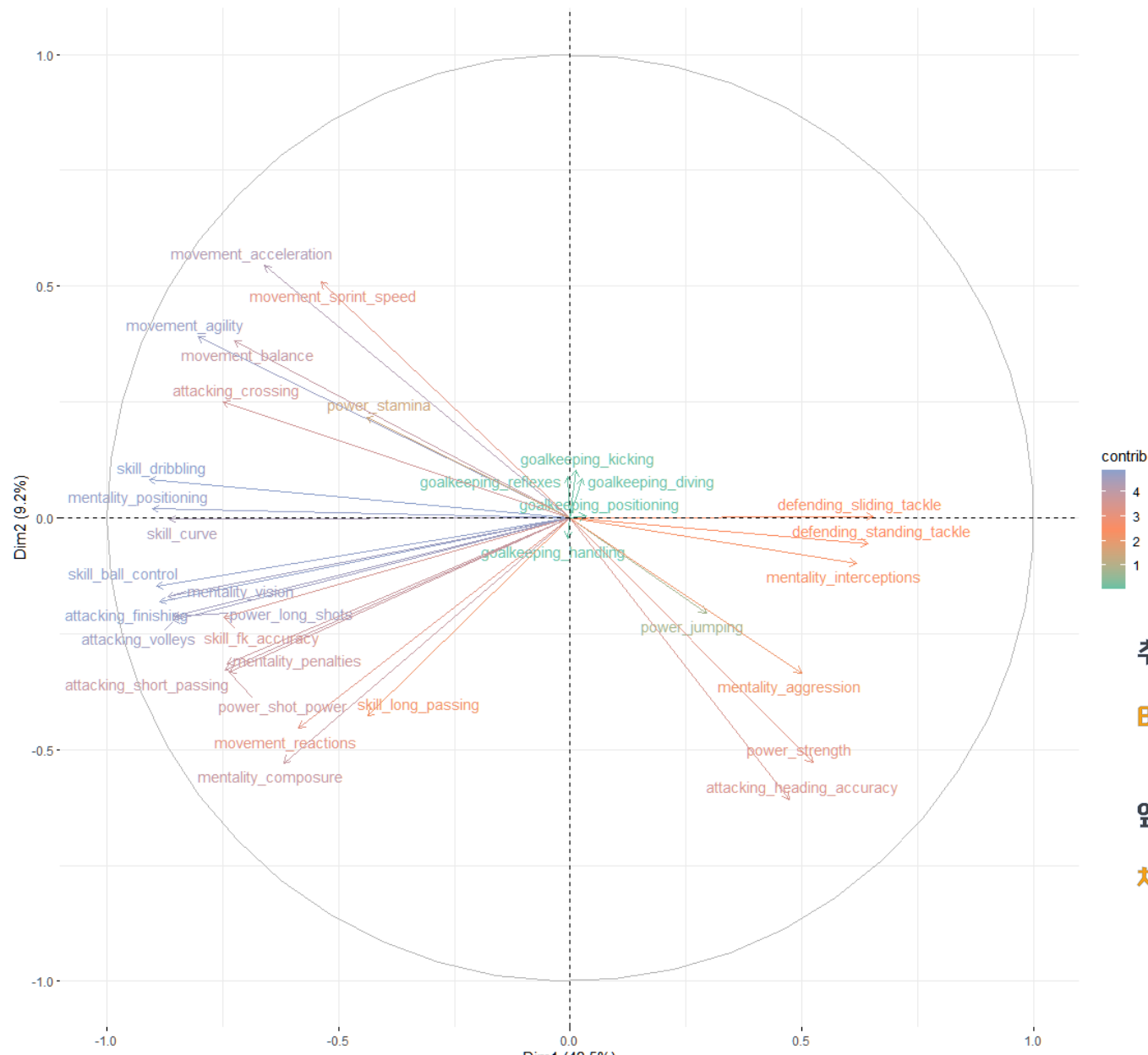
이는 스타플레이어의 영향과

본 분석에서의 주성분 분석의 한계 때문일 것으로 추측해보았습니다.

2. 데이터 분석

분석 주제 4

Variables - PCA



추가적으로 포지션별로 100명씩 추출한 표본에서 진행한 PCA분석의 **Biplot**은 다음과 같습니다.

앞서 살펴본 포지션별 500명을 추출한 표본과 차이가 있는 것을 확인할 수 있습니다.

2. 데이터 분석

분석 주제 5

원발잡이의 능력치가 더 높을까?

TOP 500 선수 중 공격수, 미드필더, 수비수로 분류하여 분석을 진행했습니다.

골키퍼의 경우에는 발보다 손을 쓰는 포지션이므로 이상치로 제거한 후 회귀 분석을 진행하였습니다.

1) NA값을 먼저 제거한 후 상관성을 살펴보았습니다.

> # NA값 제거

> sum(is.na(Attacker))

[1] 175

> colSums(is.na(Attacker))

Attacker

Goalkeeper

Defender

Midfielder

Forward

Attacker

Goalkeeper

Defender

Midfielder

Forward

> Attacker_D <- na.omit(Attacker)

> |

Attacker

Goalkeeper

Defender

Midfielder

Forward

Attacker

Goalkeeper

Defender

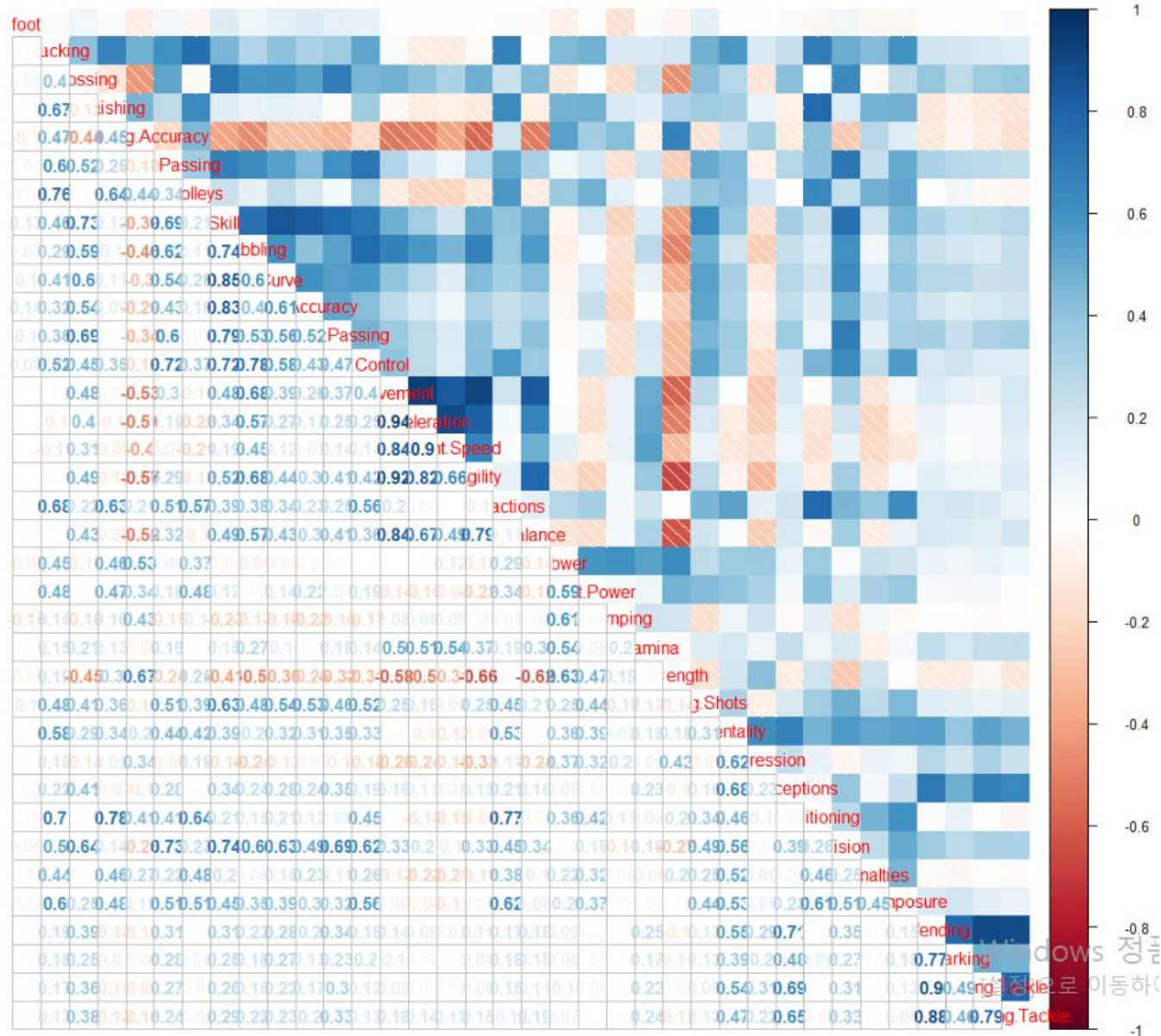
Midfielder

Forward

	Attacker	Goalkeeper	Defender	Midfielder	Forward
Foot	1.0000000000	-0.0193956211	0.0526825165	0.01964370	0.04777323
Attacking	0.450184262	-0.478612820	0.401479261	0.30565352	0.486477516
Crossing	0.052682516	-0.4033334190	0.665100400	0.12224825	0.43881104
Finishing	0.039643697	0.4033334190	0.665100398	0.122248245	0.4388110386
Heading.Accuracy	0.097138679	0.4727533345	-0.4388110386	0.45174446	0.122248245
Short.Passing	0.047773226	0.6031945650	0.5217263455	0.25125210	0.637408186
Volleys	0.039643697	0.6031945650	0.5217263455	0.25125210	0.637408186
Skill	0.134993368	0.4573435156	0.7288342115	0.12297405	0.09995191
Dribbling	0.066470727	0.2917519440	0.5873699893	0.09995191	0.09995191
Curve	0.104189986	0.4084204104	0.5960156061	0.10772503	0.29730344
FK.Accuracy	0.140249352	0.3189045991	0.5383303465	0.09075057	0.29162693
Long.Passing	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Ball.Control	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Movement	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Acceleration	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Sprint.Speed	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Agility	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Reactions	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Balance	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Stamina	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Long.Shots	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Strength	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Mentality	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Defending	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Marking	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Sliding.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Standing.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Composure	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Positioning	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Penalties	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Shot.Power	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Jumping	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Stamina	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Long.Shots	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Strength	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Mentality	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Defending	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Marking	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Sliding.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Standing.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Composure	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Positioning	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Penalties	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Shot.Power	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Jumping	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Stamina	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Long.Shots	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Strength	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Mentality	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Defending	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Marking	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Sliding.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Standing.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Composure	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Positioning	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Penalties	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Shot.Power	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Jumping	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Stamina	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Long.Shots	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Strength	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Mentality	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Defending	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Marking	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Sliding.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Standing.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Composure	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Positioning	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Penalties	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Shot.Power	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Jumping	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Stamina	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Long.Shots	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Strength	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Mentality	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Defending	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Marking	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Sliding.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Standing.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Composure	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Positioning	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Penalties	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Shot.Power	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Jumping	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Stamina	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Long.Shots	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Strength	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Mentality	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Defending	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Marking	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Sliding.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Standing.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Composure	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Positioning	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Penalties	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Shot.Power	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Jumping	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Stamina	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Long.Shots	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Strength	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Mentality	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Defending	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Marking	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Sliding.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Standing.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Composure	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Positioning	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Penalties	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Shot.Power	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Jumping	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Stamina	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Long.Shots	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Strength	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Mentality	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Defending	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Marking	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Sliding.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Standing.Tackle	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Composure	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Positioning	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Penalties	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Shot.Power	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
Jumping	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000

2. 데이터 분석

분석 주제 5



2) 상관분석에서 주력으로 쓰는 발과 상관성이 있는 부분을 찾아보았습니다.

2. 데이터 분석

분석 주제 5

```
Call:
lm(formula = foot ~ ., data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-0.5347 -0.2188 -0.1241 -0.0161  0.9808

Coefficients: (6 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.916699    0.520557   1.761  0.0791 .
Attacking     -0.006102    0.004698  -1.299  0.1948
Crossing       0.003313    0.005534   0.599  0.5497
Finishing      0.017421    0.008897   1.958  0.0510 .
Heading.Accuracy 0.004802    0.006506   0.738  0.4610
Short.Passing  0.005519    0.008730   0.632  0.5277
Volleys        NA          NA        NA      NA
Skill          0.016432    0.008941   1.838  0.0669 .
Dribbling     -0.023856    0.014605  -1.633  0.1033
Curve         -0.013664    0.009555  -1.430  0.1536
FK.Accuracy   -0.013792    0.009383  -1.470  0.1424
Long.Passing  -0.011902    0.009242  -1.288  0.1986
Ball.Control  NA          NA        NA      NA
Movement     -0.002185    0.002633  -0.830  0.4071
Acceleration  0.001230    0.006386   0.193  0.8474
Sprint.Speed  0.004529    0.005137   0.882  0.3786
Agility       -0.001324    0.005831  -0.227  0.8205
Reactions     0.006418    0.007312   0.878  0.3807
Balance        NA          NA        NA      NA
Power          0.003193    0.004761   0.671  0.5029
Shot.Power    -0.008283    0.009232  -0.897  0.3702
Jumping       -0.006699    0.005247  -1.277  0.2025
Stamina       -0.003478    0.005664  -0.614  0.5396
Strength      -0.003855    0.005522  -0.698  0.4856
Long.Shots     NA          NA        NA      NA
Mentality     -0.002856    0.003055  -0.935  0.3505
Aggression     0.005327    0.003548   1.501  0.1342
Interceptions  0.003463    0.004051   0.855  0.3932
Positioning    -0.009221    0.009520  -0.969  0.3334
Vision         0.002830    0.005941   0.476  0.6341
Penalties      NA          NA        NA      NA
Composure     -0.008884    0.004996  -1.778  0.0762 .
Defending      0.003062    0.003095   0.989  0.3232
Marking       -0.002268    0.003951  -0.574  0.5663
Standing.Tackle -0.008120    0.005486  -1.480  0.1397
Sliding.Tackle NA          NA        NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3802 on 359 degrees of freedom
Multiple R-squared:  0.09589,    Adjusted R-squared:  0.02286
F-statistic: 1.313 on 29 and 359 DF,  p-value: 0.1332
```

3) 하지만 **상관관계가 양/음으로 큰 것이 많아,**

PCA를 진행한 후 나온 주성분으로 진행하는 것이

더 효과적인 것 같아 **회귀분석**을 먼저 진행하였습니다.

2. 데이터 분석

분석 주제 5

```
Call:
lm(formula = foot ~ ., data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-0.5347 -0.2188 -0.1241 -0.0161  0.9808

Coefficients: (6 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.916699   0.520557   1.761  0.0791 .
Attacking      -0.006102   0.004698  -1.299  0.1948
Crossing        0.003313   0.005534   0.599  0.5497
Finishing       0.017421   0.008897   1.958  0.0510 .
Heading.Accuracy 0.004802   0.006506   0.738  0.4610
Short.Passing   0.005519   0.008730   0.632  0.5277
Volleys         NA         NA         NA      NA
Skill           0.016432   0.008941   1.838  0.0669 .
Dribbling      -0.023856   0.014605  -1.633  0.1033
Curve          -0.013664   0.009555  -1.430  0.1536
FK.Accuracy    -0.013792   0.009383  -1.470  0.1424
Long.Passing   -0.011902   0.009242  -1.288  0.1986
Ball.Control    NA         NA         NA      NA
Movement       -0.002185   0.002633  -0.830  0.4071
Acceleration    0.001230   0.006386   0.193  0.8474
Sprint.Speed    0.004529   0.005137   0.882  0.3786
Agility         -0.001324   0.005831  -0.227  0.8205
Reactions       0.006418   0.007312   0.878  0.3807
Balance         NA         NA         NA      NA
Power           0.003193   0.004761   0.671  0.5029
Shot.Power     -0.008283   0.009232  -0.897  0.3702
Jumping        -0.006699   0.005247  -1.277  0.2025
Stamina        -0.003478   0.005664  -0.614  0.5396
Strength       -0.003855   0.005522  -0.698  0.4856
Long.Shots      NA         NA         NA      NA
Mentality      -0.002856   0.003055  -0.935  0.3505
Aggression      0.005327   0.003548   1.501  0.1342
Interceptions   0.003463   0.004051   0.855  0.3932
Positioning     -0.009221   0.009520  -0.969  0.3334
Vision          0.002830   0.005941   0.476  0.6341
Penalties       NA         NA         NA      NA
Composure      -0.008884   0.004996  -1.778  0.0762 .
Defending       0.003062   0.003095   0.989  0.3232
Marking        -0.002268   0.003951  -0.574  0.5663
Standing.Tackle -0.008120   0.005486  -1.480  0.1397
Sliding.Tackle  NA         NA         NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3802 on 359 degrees of freedom
Multiple R-squared:  0.09589,    Adjusted R-squared:  0.02286
F-statistic: 1.313 on 29 and 359 DF,  p-value: 0.1332
```

```
> # 결과 값 비교
> Attacker_D_pred <- predict(Attacker_D_lm, newdata = test)
경고메시지(들):
In predict.lm(Attacker_D_lm, newdata = test) :
  prediction from a rank-deficient fit may be misleading
> Attacker_D_pred <- round(Attacker_D_pred)
> result <- data.frame(Attacker_D_pred, test$foot)
> prop.table(table(ifelse(result$Attacker_D_pred == result$test.foot, '0', 'X')))
```

	0	X
	0.8372093	0.1627907

```
> |
```

4) 처음 회귀분석을 돌렸을 시,

P-value 값과 R-squared 값이 좋지 않아

주성분 분석을 시행했습니다.

결과값 비교도 83%가 나와 높지 않은 지표를 보여주었습니다.

2. 데이터 분석

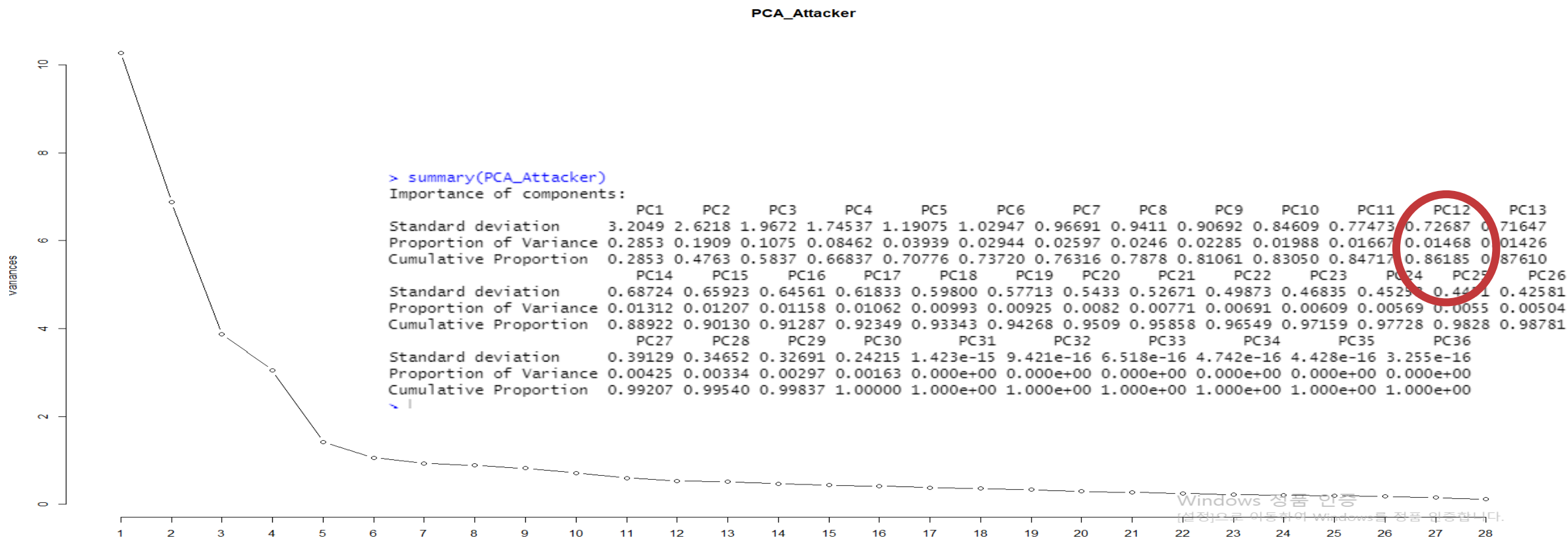
분석 주제 5

5) 주성분 분석을 통해 변수가 28개 인 것을 파악했습니다. 시각화를 통해 기울기가 급격하게 줄어드는 부분을 파악했고,

그후 summary를 통해 85%가 넘는 부분을 확인하고, 85% 이상인 주성분을 사용하기로 결정하였습니다.

Standard deviations (1, .., p=36):

[1]	3.204877e+00	2.621826e+00	1.967179e+00	1.745370e+00	1.190754e+00	1.029474e+00	9.669072e-01	9.410636e-01	9.069177e-01
[10]	8.460851e-01	7.747253e-01	7.268699e-01	7.164659e-01	6.872366e-01	6.592291e-01	6.456099e-01	6.183300e-01	5.980006e-01
[19]	5.771264e-01	5.432919e-01	5.267107e-01	4.987261e-01	4.683473e-01	4.525256e-01	4.450675e-01	4.258146e-01	3.912880e-01
[28]	3.465238e-01	3.269051e-01	2.421481e-01	1.423251e-15	9.421250e-16	6.517634e-16	4.741505e-16	4.428203e-16	3.255419e-16



2. 데이터 분석

분석 주제 5

```
Call:
lm(formula = foot ~ ., data = train_pca)

Residuals:
    Min       1Q   Median       3Q      Max
-6.7055 -1.3205  0.0832  1.6060  8.7017

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.340245    2.722324   2.329  0.02038 *
PC1          -0.001421    0.004349  -0.327  0.74406
PC2          -0.025745    0.005183  -4.968 1.03e-06 ***
PC3           0.017349    0.005493   3.159  0.00171 **
PC4          -0.089523    0.004757 -18.818 < 2e-16 ***
PC5           0.305895    0.011093  27.575 < 2e-16 ***
PC6          -0.980840    0.018009 -54.463 < 2e-16 ***
PC7          -0.016141    0.009074  -1.779  0.07605 .
Attacker_D.foot -2.120348    0.332430  -6.378 5.20e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.435 on 380 degrees of freedom
Multiple R-squared:  0.9316,    Adjusted R-squared:  0.9302
F-statistic: 647.4 on 8 and 380 DF,  p-value: < 2.2e-16
```

6) 주성분 분석을 통해 회귀분석을 돌려본 결과 **R-squared도 93%로 유의미한 결과**를 보였고,

P-value 값도 유의미한 결과를 보였습니다.

회귀분석의 한계로 인하여

왼발잡이와 오른발잡이가 누가 더 우위에 있다고는 알 수 없었지만,

왼발과 오른발이 주발인 선수들이 차이가 있음을 확인할 수 있었습니다.

03

느낌점

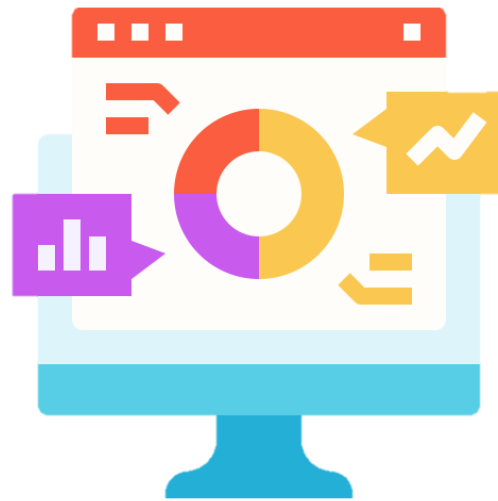
3. 느낀점



데이터 전처리

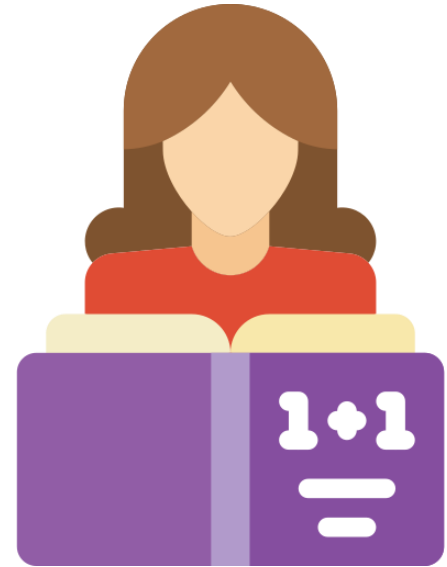
생각보다 데이터 전처리가 오래 걸리는 것을
깨닫게 되었습니다.

또한 생각보다 무의미한 값이 많이 떠서
시행착오를 많이 겪기도 하였지만, 계속된 도전 후
유의미한 값을 보게 되어 보람찬 프로젝트였습니다.



필드에 대한 이해

게임의 선수 이름이 별명인 경우가 있어서
데이터 전처리. 작업에 있어서
그 필드에 대한 사전지식도 중요하다 생각하게 되었습니다.



수학, 통계 지식

어떤 통계 기법을 적용할지, 혹은 기법에 대한 이해를 위해
수학과 통계에 대한 지식이 중요함을
깨닫게 되었습니다.

감사합니다