Northeastern University
College *of* Engineering

IE6600 Project Presentation

Analysis of the Future Evaluation System
Based on the "FIFA" World

Reported by :

Yuxi Chen, Haotian Chen, Junfei Ren, Rundong Xu

Overall framework

Three Part

Reasons and ‘’**Data**’’

’’R**Shiny**’’

Introduction | Dataset | Methods | Results | Conclusion | Q&A

Approaches to ‘’**Visualization**’’

1. Introduction

Introduction

What is the Topic?

*"Analysis of the future **Evaluation system** based on the **"FIFA" world**"*

"FIFA" World

Goal	Question
Helping the club Find top wages player to balance their team cost.	1.What is wage distribution of top 50 player?
Find the	What is the wage level and interval of top 50 player in their overall level?
Helping the game player, and club to compare the Play in different condition.	2. Compare the different Player's Capability Map
Help fans or pre-fans to find the development of Soccer industry in the world and promoting it.	3.What is the typical distribution of Top-athletes' nationality and clubs in the world?

VS

Future Evaluation System

Question	Goal
1.Top50 wage occupied by members on your dream organization (Company, Soccer Club, Field, Industry, Market etc)	Help the people find who are the Top Ranking person's wages or other Filter dimensions (Value, Overall, Potential)
2.Candidate Competition, Peer Comparation	Help the organization to make optimal choice.
3. Outstanding Person's distribution from their nationality/degree level/age, and where organization they employed	Profile of ''Outstanding'', which help the organization to target person background to modelling, also help the person to find the outstanding organization

Introduction

Why we choose this Topic ?

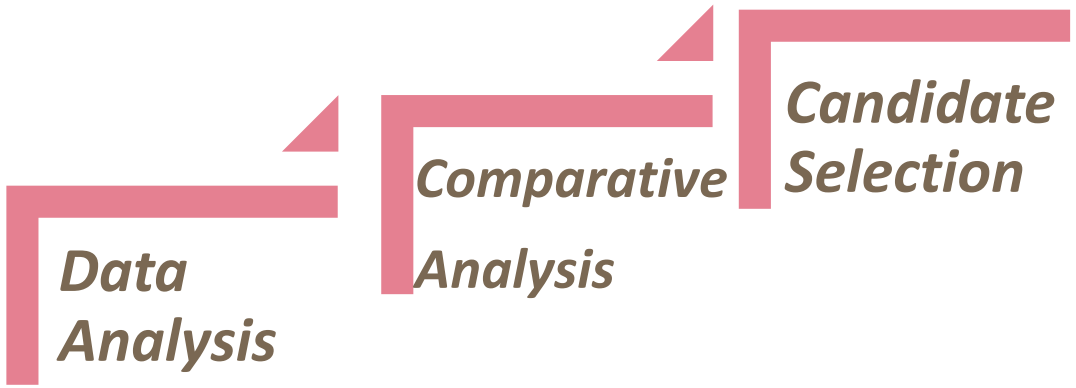
*“Analysis of the **”Future”** **Evaluation system** based on the **"FIFA" world**”*

Performance

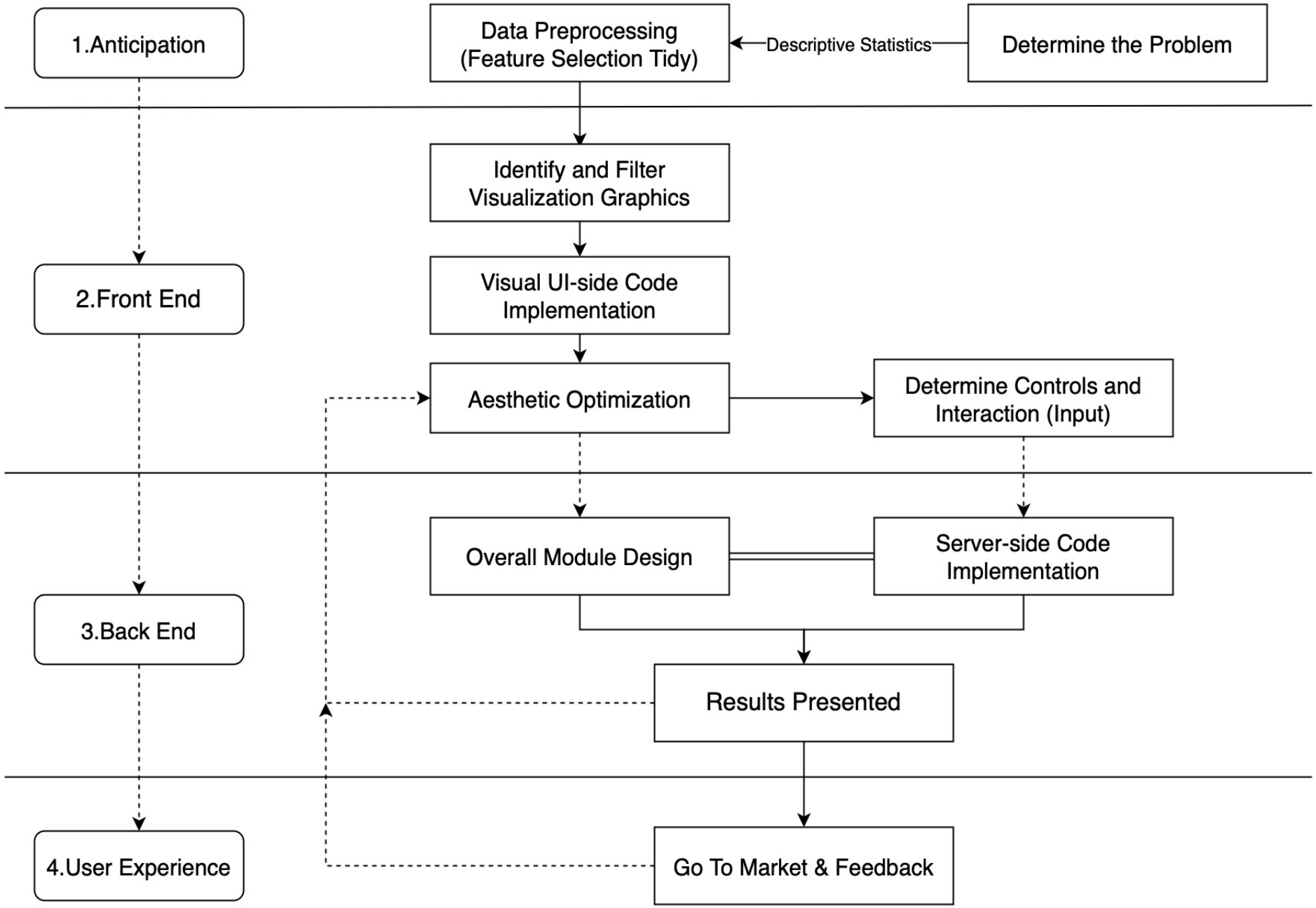
- Playback analysis from recording
- Optimize the Performance of Player, Coaching and Club
- Select and Arrange Team member

Business Value

- New Evaluation system for Recruitment
- Managing a Valuable Club
- Balance the cost of buy new Player



Workflow



That's How we **Co**llaborating

Teamwork

2. Data

Dataset

Introduce the background, and structure of your data set

Data: FIFA19(Video Game)

Link: https://sofifa.com / https://www.kaggle.com,

Features: 18000players , 88 attributes.

Data Cleaning

Feature selection:

Problem1:What is the correlation between player value and salary?

Feature: 'Name', 'Photo', 'Overall', 'Potential' ,'Value', 'Wage'

Problem2: How to evaluate a player's ability attributes, and the influence of ability attributes on player value.

Feature: 'Name', 'Photo', 'Age' , Ability(Comprehensive ability)

Problem3: What is the typical distribution of high-value athletes nationality and clubs in the world?

Feature: 'Nationality', 'Name', 'Photo', 'Club' , 'Wage', 'Overall', 'League'

Data Cleaning

Problem:

(A). There are special characters in the variable, and the parameter cannot be called in the later drawing.

Value
€110.5M
€77M
€118.5M
€72M
€102M
€93M
€67M
€80M
€51M
€68M
€77M
€76.5M
€44M
€60M
€63M
€89M

Special characters: € , M

Solution: Use function gsub().

```
data1$Value <- as.numeric(gsub("\\€|\\M","",data1$Value))
data1$Wage <- as.numeric(gsub("\\€|\\K","",data1$Wage))
```

Value	Wage
3.2	24
3	6
1.9	15
NA	2
4.6	12
NA	2
3.9	16
NA	2
NA	2
NA	1
NA	1

Data Cleaning

Problem:

(B). There is a duplicate name, after drawing or Rshiny production will appear some error.

2372	222352	A. Ajeti	21	https://cdn.	Switzerland	https://cdn.	74
8492	203458	A. Ajeti	24	https://cdn.	Albania	https://cdn.	67

3469	228179	A. Castro	23	https://cdn.	Argentina	https://cdn.	72
7153	220805	A. Castro	24	https://cdn.	Colombia	https://cdn.	68
10464	158595	A. Castro	31	https://cdn.	Mexico	https://cdn.	65

Solution: Rename, number people who have the same name

Step: 1. Choosing duplicated name.

```
data1 <- datac[duplicated(datac$Name),]
```

2. Edit the number after the name.

```
data1$Name <- paste(data1$Name,"1",sep = "")
```

3. Checking if there are still have duplicate name
4. Repeating step2&3
5. Combining new data frame with origin data.

```
data3 <- merge(data1,data2,all=T)
```

R. Thomas	https://cdn.	76	76	5.5	21	30	6'4	Angers SCO	France	Ligue 1	France
R. Thomas1	https://cdn.	75	80	9	13	23	5'9	PSV	New Zealan	Eredivisie	Netherlands

Data Cleaning

Curve	BallControl	Aggression	Interceptior	Positioning	Vision	Composure	Crossing	ShortPassin	LongPassing	Acceleratio	SprintSpeec	Agility	Reactions	Balance	Jumping	Stamina	Strength	Finishing	Volleys	FKAccuracy	ShotPower	LongShots	Penalties
93	96	48	22	94	94	96	84	90	87	91	86	91	95	95	68	72	59	95	86	94	85	94	75
81	94	63	29	95	82	95	84	81	77	89	91	87	96	70	95	88	79	94	87	76	95	93	85
88	95	56	36	89	87	94	79	84	78	94	90	96	94	84	61	81	49	87	84	87	80	82	81
21	42	38	30	12	68	68	17	50	51	57	58	60	90	43	67	43	64	13	13	19	31	12	40
85	91	76	61	87	94	88	93	92	91	78	76	79	91	77	63	90	75	82	82	83	91	91	79
83	94	54	41	87	89	91	81	89	83	94	88	95	90	94	56	83	66	84	80	79	82	80	86
85	93	62	83	79	92	84	86	93	88	80	72	93	90	94	68	89	58	72	76	78	79	82	82
86	90	87	41	92	84	85	77	82	64	86	75	82	92	83	69	90	83	93	88	84	86	85	85
74	84	88	90	60	63	82	66	78	77	76	75	78	85	66	93	84	83	60	66	72	79	59	75
13	16	34	19	11	70	70	13	29	26	43	60	67	86	49	76	41	78	11	13	14	22	12	11
77	89	80	39	91	77	86	62	83	65	77	78	78	90	78	84	78	84	91	89	86	88	84	88
86	90	60	82	79	86	85	88	92	93	64	62	70	89	71	30	75	73	76	82	84	87	92	73
49	76	89	88	48	52	82	55	79	70	68	68	58	85	54	91	66	88	42	47	51	67	43	50
82	94	57	50	89	92	93	84	93	87	70	64	92	90	90	64	78	52	76	82	77	72	75	75
49	80	90	92	71	79	85	68	86	81	82	78	82	93	92	77	96	76	65	56	49	71	69	54
88	92	48	32	84	87	84	82	87	75	87	83	91	86	85	75	80	65	84	88	88	82	88	86
78	84	76	35	93	80	89	75	80	82	68	72	71	91	71	78	89	84	94	84	68	88	85	90
84	90	69	35	91	83	87	82	83	76	88	85	90	90	80	90	83	62	90	87	78	80	82	79
18	18	43	22	11	69	69	15	36	42	38	50	37	85	43	79	35	79	14	14	12	22	10	25
19	23	23	15	13	44	66	14	33	35	46	52	61	84	45	68	38	70	14	12	20	36	17	27
66	88	85	87	77	87	90	62	89	82	50	52	66	87	52	66	86	77	67	44	68	61	54	60
77	82	84	48	93	77	82	70	78	52	75	76	77	91	59	88	92	78	89	90	76	87	79	85
14	48	29	30	12	70	70	15	55	59	54	60	51	84	35	77	43	80	13	11	11	25	16	47
82	89	65	24	92	83	90	70	81	64	88	80	86	90	91	81	76	73	93	85	73	88	83	83
60	57	92	88	28	50	84	58	59	59	63	75	54	82	55	89	65	89	33	45	31	78	49	50
77	91	62	38	88	82	86	77	82	73	96	96	92	87	83	75	83	71	88	78	63	79	78	70
83	88	63	55	90	82	91	78	82	72	94	91	91	91	88	68	84	70	90	73	60	77	83	61
59	78	87	87	69	77	84	52	85	82	59	65	62	84	66	88	87	89	59	53	74	86	79	66
89	90	64	55	80	89	87	90	89	83	73	67	83	85	76	54	70	68	83	90	86	86	92	81
87	93	34	26	83	87	83	86	85	78	94	86	94	83	93	53	75	44	77	74	77	75	84	61
88	95	58	64	78	89	86	75	89	83	75	69	87	77	90	64	70	59	79	65	76	69	87	76
86	91	46	56	83	91	88	88	91	88	75	73	79	88	81	50	92	58	80	77	87	84	89	67
91	92	59	49	84	90	85	79	88	83	89	75	92	83	93	59	79	61	79	75	86	83	93	70
80	82	43	48	90	77	86	77	77	64	93	95	76	87	70	79	76	76	88	86	74	82	79	76
65	81	69	92	56	79	91	64	81	85	53	64	63	87	60	68	66	84	55	60	53	71	51	68
85	92	84	85	85	82	86	90	84	76	83	82	86	88	86	76	91	78	70	54	67	83	70	59
90	85	65	59	85	79	86	87	85	80	94	95	82	85	65	87	75	80	86	85	87	92	91	76
11	34	31	27	10	30	65	13	50	50	65	62	55	85	54	74	41	43	10	11	10	23	14	40
74	85	50	20	92	74	86	68	75	59	73	73	75	86	69	79	70	85	92	90	62	86	80	70
61	80	76	89	59	72	81	60	80	80	70	72	68	82	68	90	74	82	38	63	64	71	68	60
12	16	25	22	12	41	69	12	36	34	51	55	47	83	36	78	41	71	10	12	14	22	19	23
20	28	38	28	12	50	70	13	37	35	49	43	55	79	49	75	39	69	15	17	13	39	13	22

Data Cleaning

Solution:

Fusing th variable to evaluate players

Method:

Using fur

Defending	General	Mental	Passing	Mobility	Power	Rating	Shooting
29	89	70.8	87	90.75	73.5	94	88.1666667
27.3333333	88	72.8	80.6666667	90.75	83	94	88.3333333
28	85.25	72.4	80.3333333	93.5	68.75	92.5	83.5
16.3333333	25.5	43.2	39.3333333	66.25	54.25	92	21.3333333
59	79.25	81.2	92	81	76.25	91.5	84.6666667
27.6666667	83.25	72.4	84.3333333	91.75	74.75	91	81.8333333
69.6666667	80.75	80	89	83.75	77.25	91	78.1666667
48.3333333	85	77.8	74.3333333	83.75	81.25	91	86.8333333
90	78	76.6	73.6666667	78.5	81.5	91	68.5
19	14	40.8	22.6666667	64	61	91.5	13.8333333
31.6666667	84	74.6	70	80.75	81	90	87.6666667
73.3333333	77.75	78.4	91	71.25	62.25	90	82.3333333
89.3333333	67.5	71.8	68	69.75	74.75	90	50
47	79.75	76.2	88	79	71	90	76.1666667
88.6666667	65.5	83.4	78.3333333	83.75	85.25	89.5	60.6666667
21	85	67	81.3333333	86.75	76.25	91.5	86
43.3333333	81.75	74.6	79	75.5	80.5	90	84.8333333
51.3333333	86.5	73	80.3333333	88.25	78.75	89.5	82.6666667
16	16	42.8	31	52.5	59	90.5	16.1666667
18	17	32.2	27.3333333	60.75	55.25	89.5	21
85.3333333	75.5	85.2	77.6666667	63.75	70.25	89	59
45.3333333	82	76.8	66.6666667	79.75	79.25	89	84.3333333
12.6666667	29.25	42.2	43	62.25	58.75	89	20.5

```
data2$Defending <- a
data2$General <- app
data2$Mental <- appl
data2$Passing <- app
data2$Mobility <- ap
data2$Power <- apply
data2$Rating <- appl
data2$Shooting <- ap
```

```
)
e')],1,mean)
)
'Penalties")],1,mean)
```

3. Method

Statistical Methods

Statistics Methods Support

Descriptive statistics Related analysis

Personal Characteristics VS Social Value Measurement

Comparative Analysis

Ability Index VS External Evaluation System Score

Distribution Analysis

Distribution of Organization Where Outstanding Humans Gather

Feature Engineering

Why?

- Removing noise
- Dimension Reduction
- Feature Selection

----------By Using **Data Mining**--PCA, K-means etc.

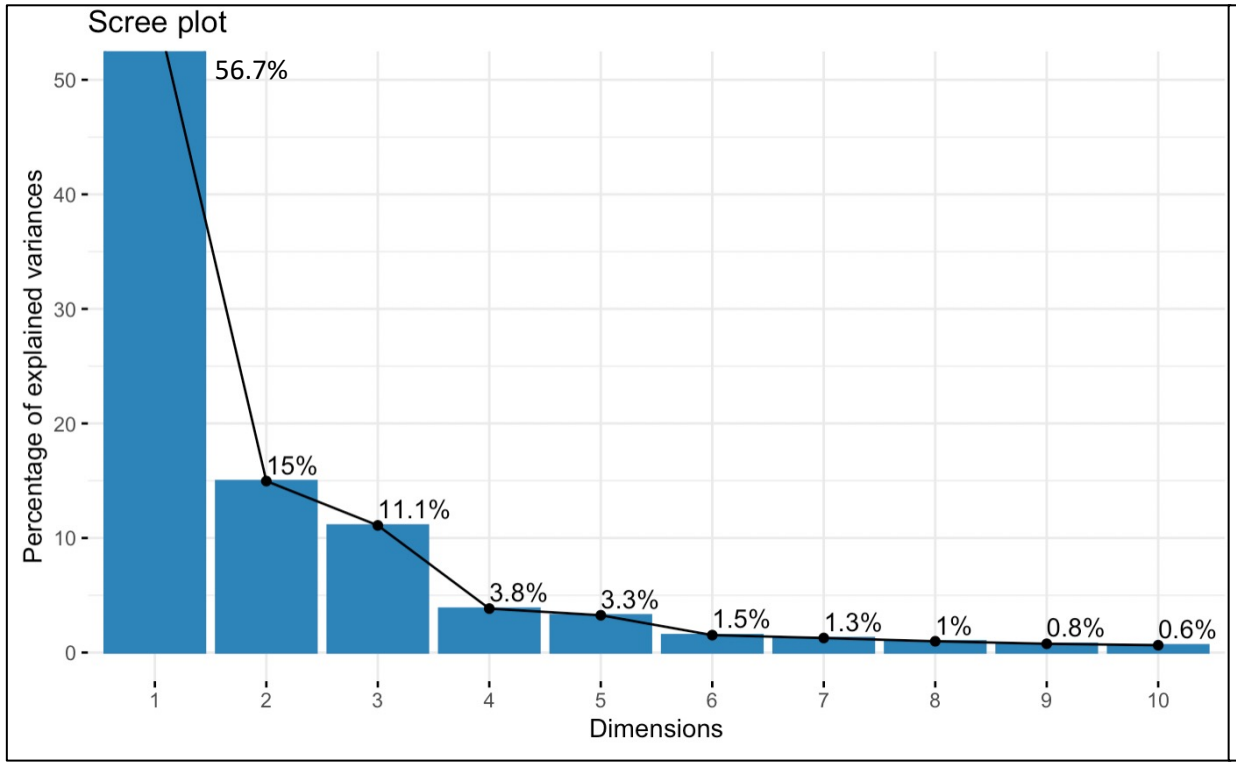
----------By incoporating **Domain Knowledge**

Anticipation:

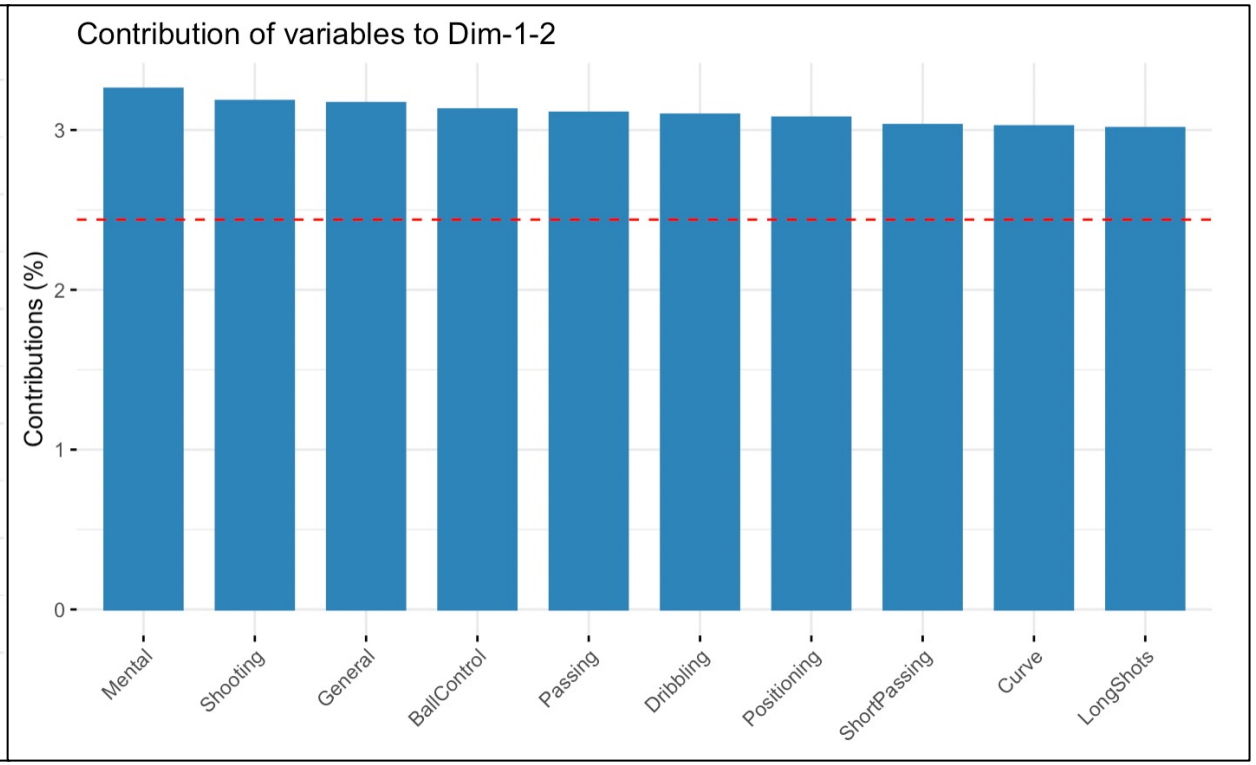
- Interpreting data and variable clearly in low-dimension and clustering
- Obtain the Correlation between Variables
- Specify and select the features for our usage

PCA

Dimension



Contribution

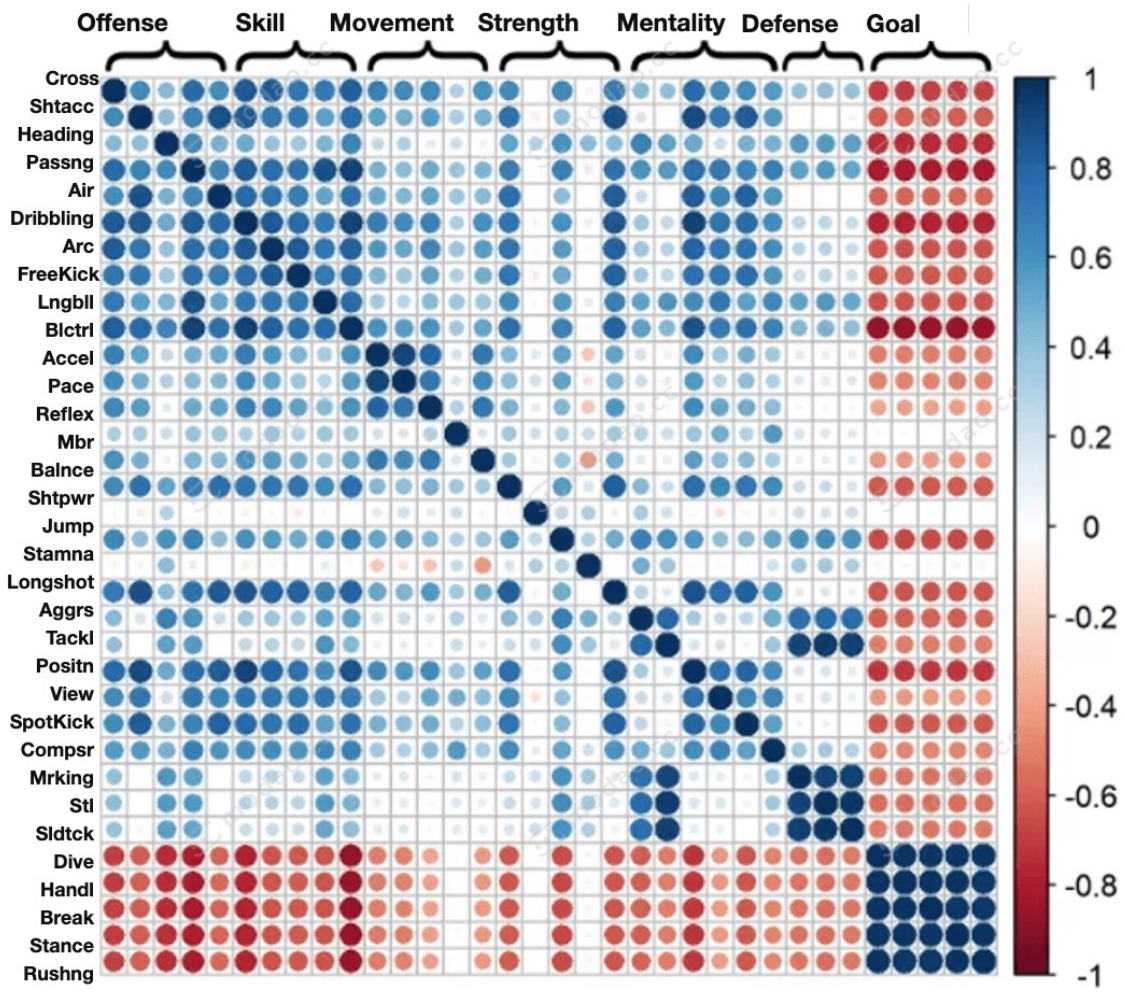


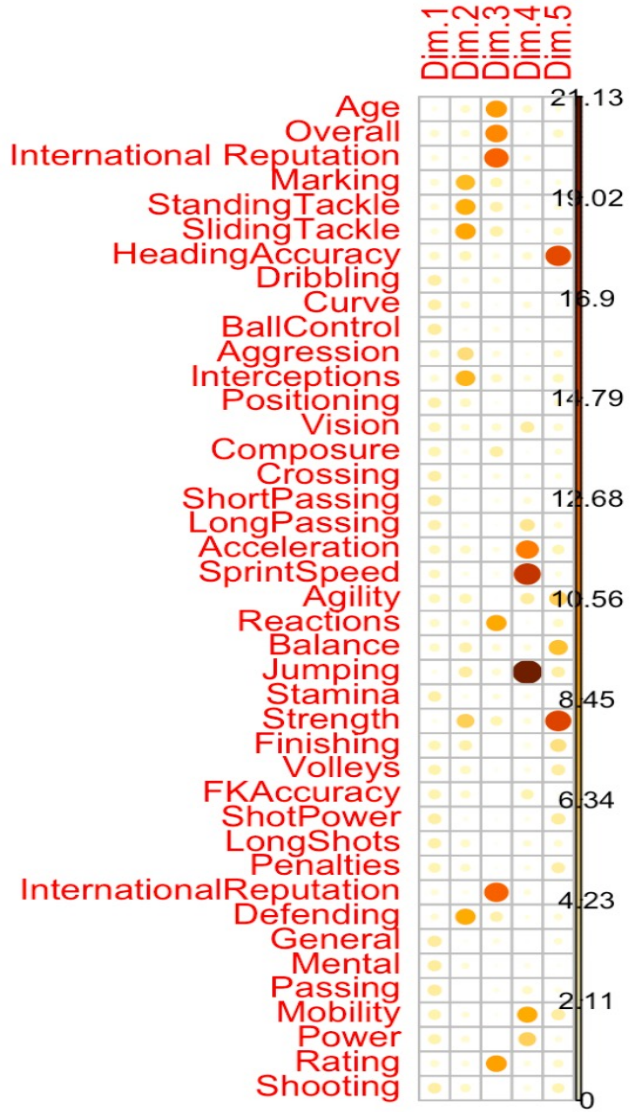
Stop at 5th principal component
Which could represent the 89.9% information (Variance)

The 'Mental' , 'Shooting' and 'General' occupy Top 3 high contribution to PC1 and PC2

PCA ----Simple But Powerful Data Mining

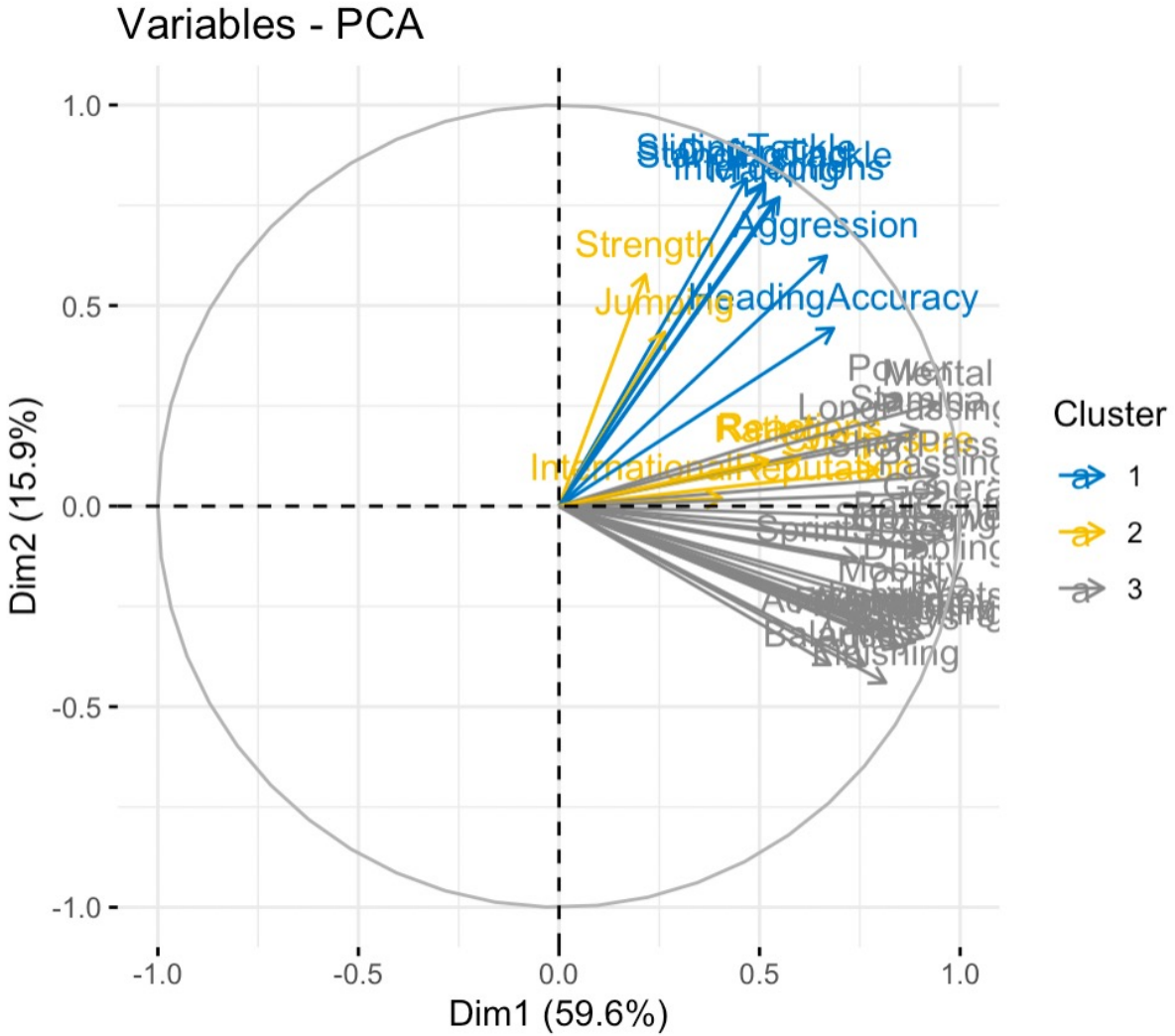
Ability Correlation Analysis





K-means Clustering

Correlation circle



Clustering

We still find the high related variables in clustering by using K-means

3 Clustering to help us knowns the Correlation between Variables

Cos2 Value

- Close to the circle : Represent better
- Close to the center : Less important

4. Result

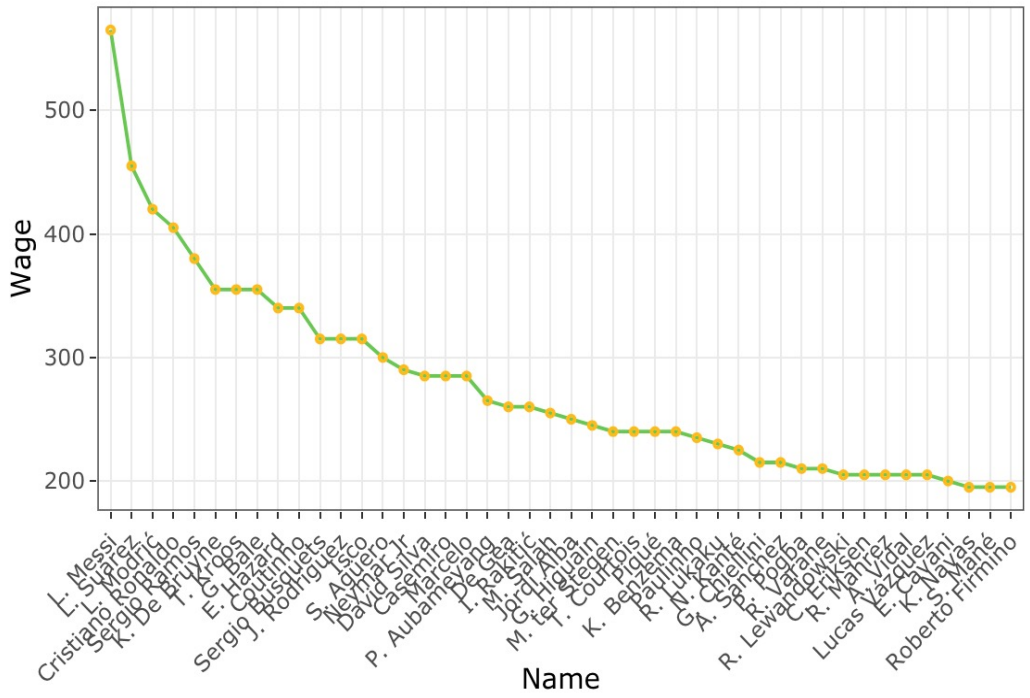
Present the results which approach your goal/s

Result

Personal Characteristics VS Social Value Measurement

Single Variable Analysis

NOW, BASED ON YOUR SETTING, THE OPTIMAL CHOICE FOR YOU IS
L. Messi.

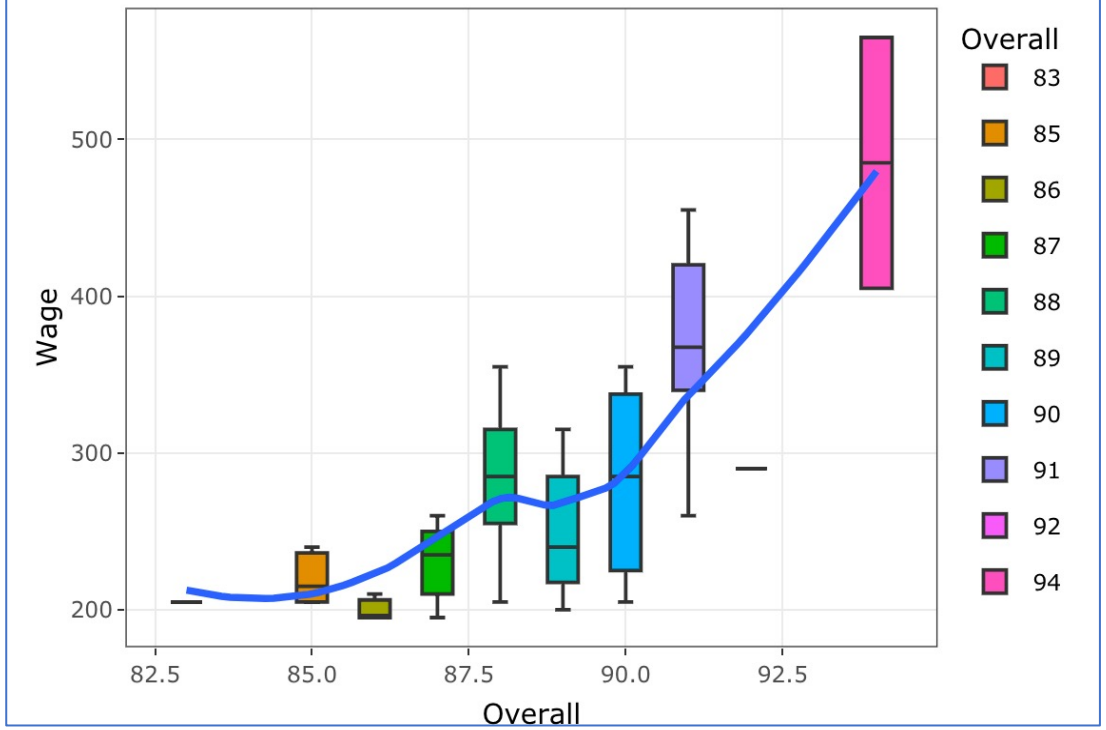


Intelligent Recommendation

Correlation

Binary Variable

NOW, BASED ON YOUR SETTING, THE OPTIMAL CHOICE FOR YOU IS
L. Messi and Cristiano Ronaldo.



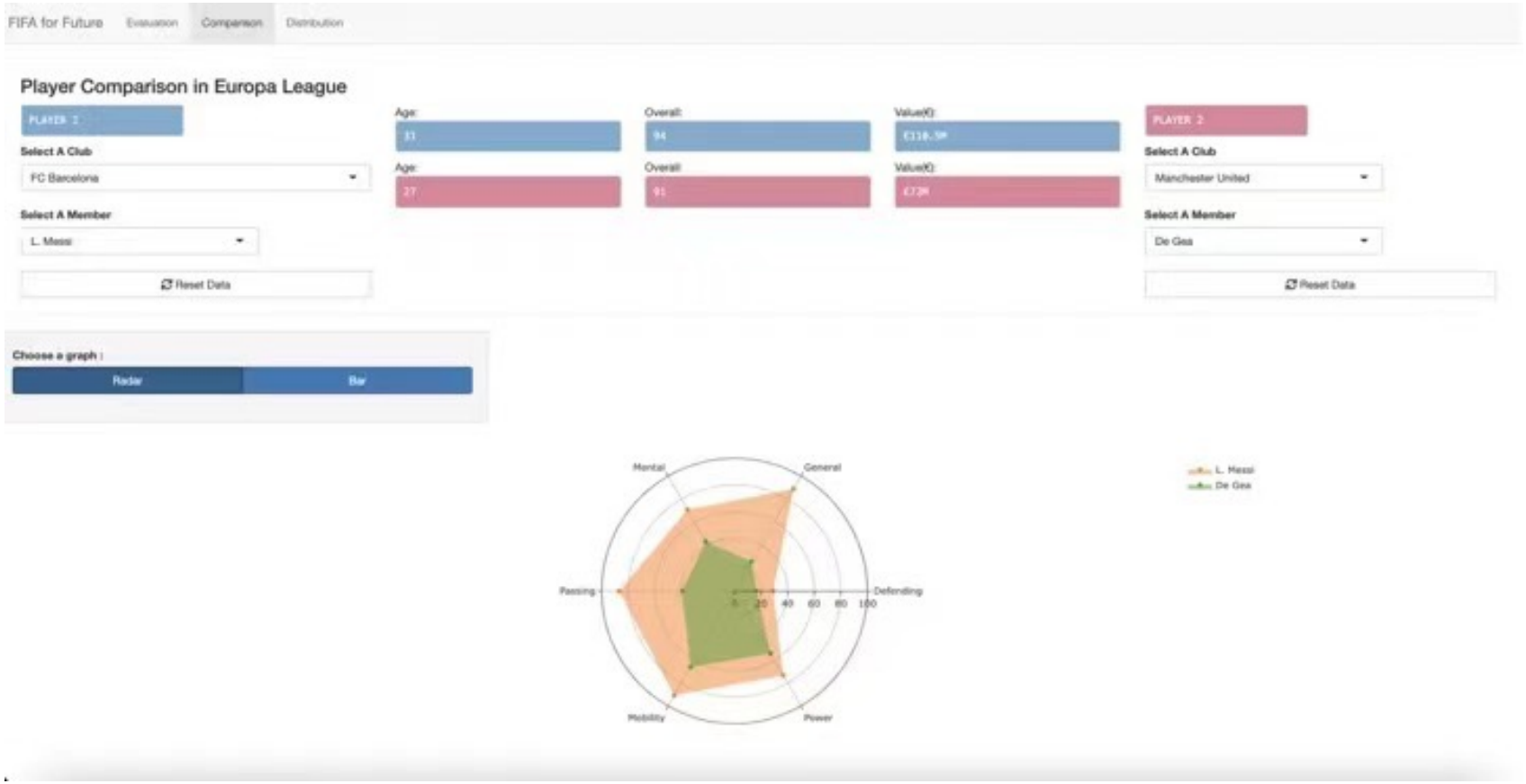
The plot shows that who are the
Top 50 wages of players; L.Messi is best one In age(25-35).
And 70%+ of them occupy wage interval [200-300]

Control Variates

Overall Score Ranking in Top 50 wages of players, L.Messi and C.Ronaldo are the best two In age (25-35).
And 70%+ of them occupy wage interval [200-300]

Result

Ability Index VS External Evaluation System Score

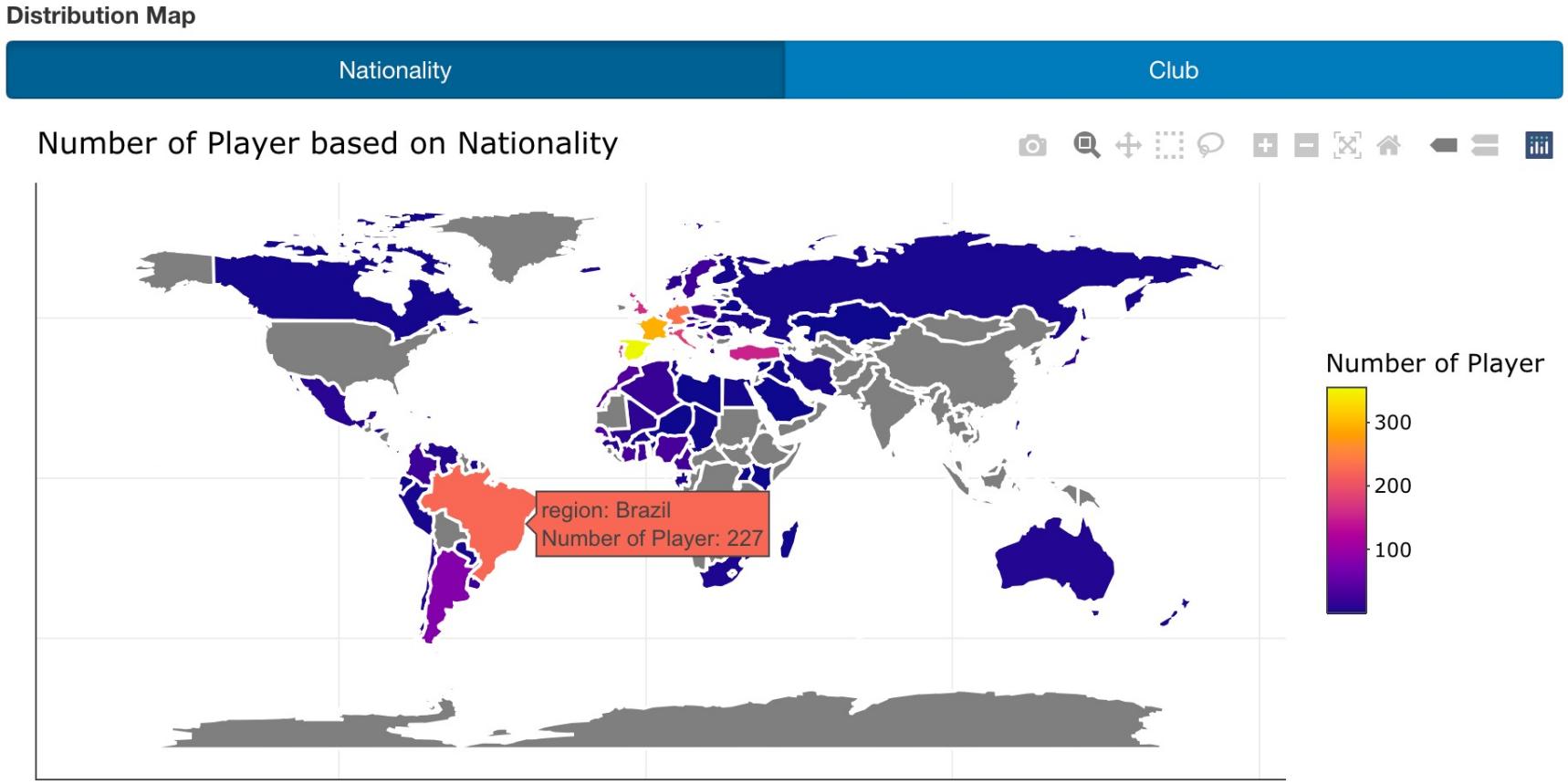


The Messi and De gega have similar overall score,

but obviously the area of Messi's ability map big than Dega, which help us certain that the overall score evaluating standard and weight are different from all case, and the ability are not the only factor it considered

Result

Distribution of Organization Where Outstanding Humans Gather



Most High-level player come from hispanic nation, and they usually gathering in Europe to soccering.
The Europe occupy high rate of gathering, and culture of soccer are the most famous.

5. Let's Shiny

R shiny

Evaluation Part 1

Design Framework and Procedure :

Widget	Server	Output
selectInput(variable x/y) (single&binary)	select a x/y - axis variable, can be null	-> plot from the filter setting
radioGroupButtons(principal)	-> choose x or y variable to be the ordering standard	Change the different axis for plotting
TabPanel	-> if Evaluation ->	Intelligent Decision System

Difficult Problem: **Needing using binary event to plotting two analysis result**

Solution:

```
observeEvent(c(input$variablex, input$variabley), {
  if(input$variablex != 'Choose' & input$variabley != 'Choose'){
    output$obs <- renderUI({
      radioGroupButtons(
        'Principal',
        label = 'Choose Principal Variable',
        choices = c('X', 'Y'),
        status = 'primary',
        checkIcon = list(
          yes = icon('ok',
                     lib = 'glyphicon'),
          no = icon('remove',
                    lib = 'glyphicon'))
      )
    })
```

Learning from: https://www.w3schools.com/colors/colors_picker.asp

R shiny

Comparation Part 2

Design Framework and Procedure ：

Widget	Server	Output
selectInput	-> indicator input ->	filter the data by indicator
radioGroupButtons	-> number of top	filter the data by number of top
radioGroupButtons	-> if radar->run function radar plot	output radar chart
radioGroupButtons	-> if bar -> run function bar plot	-> output bar chart

Difficult Problem: Custom css style
Needing Using ''Html'' to optimize aesthetic

Solution:

```
                                    tags$style(HTML("
pre {
  color: white;
  background-color: #e1849a;
}
.myclass pre {
  color: white;
  background-color: #79acd2;
  }
}"))
                                  ).
```

Learning from: https://www.w3schools.com/colors/colors_picker.asp

R shiny

Cosmopolitan Distribution Part 3

Design Framework and Procedure :

Widget	Server	Output
selectInput	> indicator input ->	filter the data by indicator
radioGroupButtons	-> number of top	filter the data by number of top
radioGroupButtons	-> type of graph	plotting the data using nationality function or plotting the data using club function

Difficult Problem: **Needing Use Plot_ly to draw the Raday plot in Rshiny**

Solution:

```
df_x <- df_part2_5 %>% filter(Name == Namex) %>% select(Defending,General,Mental,Passing,Mobility,Power)
plot_ly(
  type = 'scatterpolar',
  r = as.numeric(df_x[1,]),
  theta = c('Defending','General','Mental','Passing','Mobility','Power'),
  fill = 'toself'
) %>%
  layout(
    polar = list(
      radialaxis = list(
        visible = T,
        range = c(0,100)
      )
    ),
    showlegend = F
  )
}
}
```

Learning from: https://plotly.com/r/radar-chart/

5.Conclusion & Summary

Conclusion & Summary

OKR

Evaluation of Player and Correlation

Distribution of"Outstanding"

Ability Index VS External Evaluation System Score

Objectives	Key Conclusions
Helping the club Find top wages player to balance their team cost.	Overall Score Ranking in Top 50 wages of players, L.Messi and C.Ronaldo are the best two In age (25-35).And 70%+ of them occupy wage interval [200-300]. So if the club manager want to introduce overall interval[>82.5], they need budget not lessan than 200.
Helping the game player, and club to compare the Play in different condition.	The Messi and De gega have similar overall score, but obviously the area of Messi ‘s ability map big than Dega, so the club could choose their candidate based different evaluating standard
Help fans or pre-fans to find the development of Soccer industry in the world and promoting it.	Most High-level player come from hispanic nation, and they usually gathering in Europe to soccering. The Europe occupy high rate of gathering, and culture of soccer are the most famous.

Thanks for Everyone to Listening !!

Thanks for Instructor: Professor Lu

Questions

&

Answers

Problem Defintion : Rundong Xu(Pricinpal)

Feature Selection : Rundong Xu(Pricinpal)

Data Cleaning : Junfei Ren(Pricinpal)

Data Visualization : Yuxi Chen (Pricinpal),
Haotian Chen, Rundong Xu, Junfei Ren

Shiny : Haotian Chen (Pricinpal), Yuxi Chen,
Rundong Xu, Junfei Ren