

Sprawozdanie

Analiza i wizualizacja danych

Prowadzący:

Dr. Jarosław Olejniczak

PROJEKT

Demidziuk Jakub I5B2S1

1) Zbiór danych

Zbiór danych przedstawia listę piłkarzy dostępnych w grze FIFA 18 wraz z ich statystykami, które dobrze odzwierciedlają ich umiejętności w prawdziwym życiu. Zbiór zawiera kolumny:

ID-numer id zawodnika

Name- imie i nazwisko zawodnika

Club- klub w którym gra zawodnik

Age- wiek zawodnika

League- liga w której gra zawodnik

height_cm- wysokość zawodnika

weight_kg- waga zawodnika

nationality- narodowość zawodnika

eur_value- wartość zawodnika w euro

eur_wage- zarobki zawodnika w euro

overall- ocena ogólna zawodnika

potential-potencjał zawodnika

pac- tempo zawodnika

sho- umiejętności strzałów zawodnika

pas- umiejętności podań zawodnika

dri- umiejętności dryblingu zawodnika

def- umiejętności podań zawodnika

phy- fizyczność zawodnika

international_reputation- reputacja międzynarodowa zawodnika

preferred_foot- preferowana stopa(prawo/lewnonożny)

ID	name	club	age	league	height_cm	weight_kg	nationality	eur_value	eur_wage	overall	potential	pac	sho	pas	dri	def	phy	international_reputation	preferred_foot
20801	Cristiano Ronaldo	Real Madrid CF	32	Spanish Primera División	185.0	80.0	Portugal	95500000.0	565000.0	94	94	90	93	82	90	33	80	5	Right
158023	L. Messi	FC Barcelona	30	Spanish Primera División	170.0	72.0	Argentina	105000000.0	565000.0	93	93	89	90	86	96	26	61	5	Left
190871	Neymar	Paris Saint-Germain	25	French Ligue 1	175.0	68.0	Brazil	123000000.0	280000.0	92	94	92	84	79	95	30	60	5	Right
176580	L. Suárez	FC Barcelona	30	Spanish Primera División	182.0	86.0	Uruguay	97000000.0	510000.0	92	92	82	90	79	87	42	81	5	Right
167495	M. Neuer	FC Bayern Munich	31	German Bundesliga	193.0	92.0	Germany	61000000.0	230000.0	92	92	91	90	95	89	60	91	5	Right
188545	R. Lewandowski	FC Bayern Munich	28	German Bundesliga	185.0	79.0	Poland	92000000.0	355000.0	91	91	81	88	75	86	38	82	4	Right
193080	De Gea	Manchester United	26	English Premier League	193.0	76.0	Spain	64500000.0	215000.0	90	92	90	85	87	90	58	86	4	Right
183277	E. Hazard	Chelsea	26	English Premier League	173.0	76.0	Belgium	90500000.0	295000.0	90	91	90	82	84	92	32	66	4	Right
182521	T. Kroos	Real Madrid CF	27	Spanish Primera División	182.0	78.0	Germany	79000000.0	340000.0	90	90	56	81	89	81	73	70	4	Right
167664	G. Higuaín	Juventus	29	Italian Serie A	184.0	87.0	Argentina	77000000.0	275000.0	90	90	79	87	70	83	25	74	4	Right
155862	Sergio Ramos	Real Madrid CF	31	Spanish Primera División	183.0	75.0	Spain	52000000.0	310000.0	90	90	76	63	71	71	88	83	4	Right
192985	K. De Bruyne	Manchester City	26	English Premier League	181.0	68.0	Belgium	83000000.0	285000.0	89	92	75	84	88	85	45	75	4	Right
192119	T. Courtois	Chelsea	25	English Premier League	199.0	94.0	Belgium	59000000.0	190000.0	89	92	85	91	69	88	49	86	4	Left

2) Kod w R

```
fifadata = read.csv("C:/Users/Jakub/Desktop/Zeszyt6.csv")

library(ggplot2)

library(data.table)

library(plyr)

##leagues

#wykres1

m_data<-aggregate(fifadata,list(league = fifadata$league), mean)

setorder(m_data,-overall)

barplot(m_data$overall[2:7], names=m_data$league[2:7],cex.names=0.55,col="RED", main="Sredni overall w
ligach",ylim=c(0,80),xlab="league", ylab="mean overall")

#wykres2

barplot(m_data$age[2:7], names=m_data$league[2:7],cex.names=0.55,col="PURPLE", main="Sredni wiek w
ligach",ylim=c(0,40),xlab="league", ylab="mean age")

#wykres3

x_data<-fifadata

x_data<-x_data[x_data$overall > 80,]

x_data<-count(x_data,vars="league")

setorder(x_data,-freq)

barplot(x_data$freq[1:7], names=x_data$league[1:7],cex.names=0.55,col=c(1,2,3,4,5,6,7), main="Zawodnicy z overallem
80+",ylim=c(0,160),xlab="league", ylab="liczba zawodnikow")

#wykres4

m_data<-fifadata

setorder(m_data,-league)

x_data<-m_data[m_data$age>25,]

x2_data<-m_data[m_data$age<=25,]

x_data<-count(x_data,vars="league")

x2_data<-count(x2_data,vars="league")

barplot(rbind(x_data$freq,x2_data$freq),names.arg=x_data$league,las=2

,cex.names=0.6,beside=TRUE,xlim=c(0, 120), ylim=c(0, 500),ylab="liczba zawodnikow",main="zawodnicy w wieku 25+ oraz ci ponizej 25
lat")

##kluby

#wykres5

m_data<-aggregate(fifadata,list(club = fifadata$club), mean)

setorder(m_data,-overall)

barplot(m_data$overall[1:15], names=m_data$club[1:15],cex.names=0.55,col=rainbow(16),hor=TRUE,las=2, main="Sredni overall w
klubach",xlim=c(0,100),xlab="league", ylab="mean overall")

#wykres6

barplot(m_data$age[1:15], names=m_data$club[1:15],cex.names=0.55,col=rainbow(16),hor=TRUE,las=2, main="Sredni wiek w
klubach",xlim=c(0,40),xlab="league", ylab="mean age")
```

```

#wykres7

m_data<-fifadata[fifadata$overall>=85,]

m_data<-count(m_data,vars=("club"))

pie(c(m_data$freq),label=m_data$club,col=rainbow(23),main="piłkarze z ocena 85+",cex=0.6)

##zawodnicy

#wykres8

boxplot(pac~age,data=fifadata,main="zaleznosc szybkości zawodników od wieku",xlab="wiek",ylab="szybkość",col="LIGHTBLUE")

#wykres9

boxplot(eur_value~overall,data=fifadata,main="zaleznosc wartości od oceny",xlab="ocena",ylab="wartość",col="GREEN")

#wykres10

boxplot(eur_wage~overall,data=fifadata,main="zaleznosc zarobków od oceny",xlab="ocena",ylab="zarobki",col="GREEN")

#wykres11

boxplot(overall~preferred_foot,data=fifadata,main="zaleznosc oceny od preferowanej
stopy",xlab="stopa",ylab="ocena",col="DARKGREEN")

#wykres12

qplot(weight_kg,height_cm, data=fifadata, colour=phy,

main="zaleznosc fizyczności od wysokości i ciężaru",xlab="waga w kg",ylab="wysokość w cm")

#wykres13

x_data<-aggregate(fifadata,list(age = fifadata$age), mean)

qplot( age,overall, data=x_data,

main="zaleznosc oceny od wieku",

geom="path",xlab="wiek",ylab="ocena")

#wykres14

qplot( age,international_reputation, data=x_data,

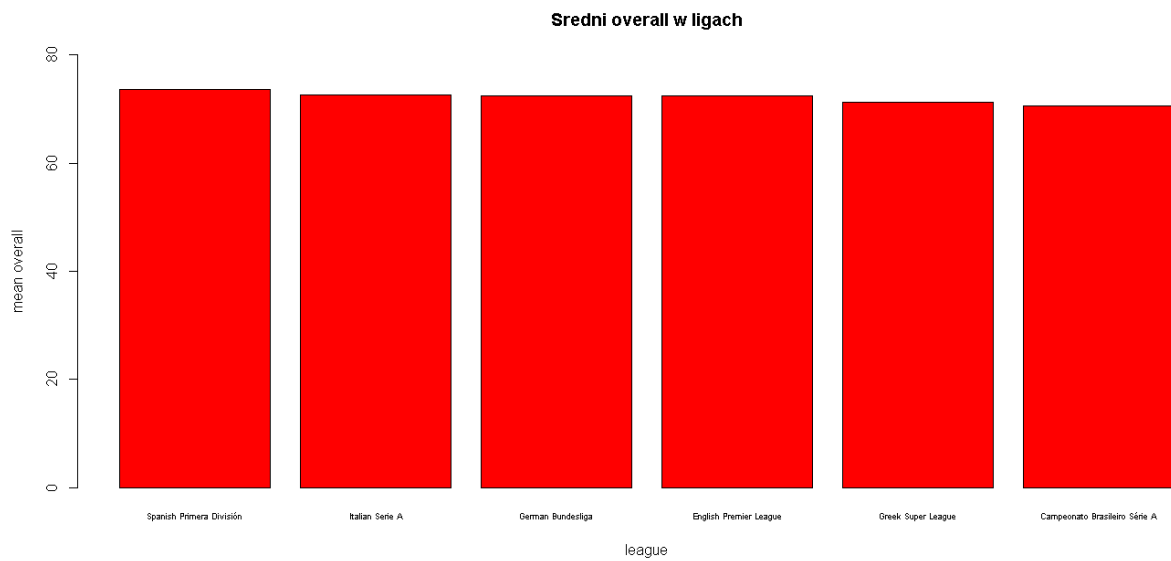
main="zaleznosc reputacji międzynarodowej od wieku",

geom="path",xlab="wiek",ylab="reputacja")+xlim(18,40)

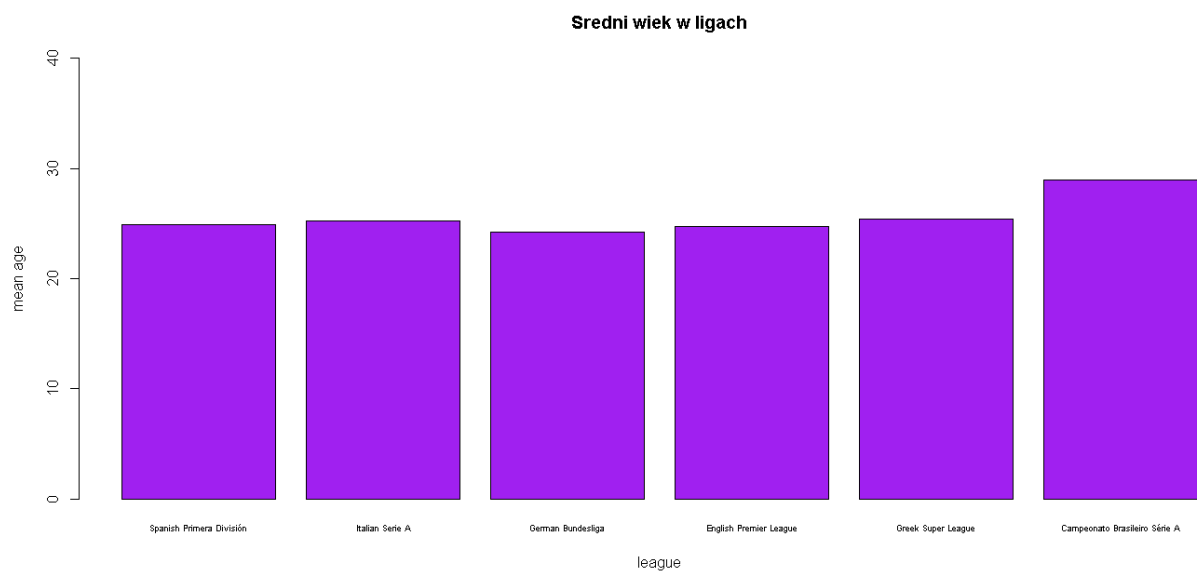
```

3) Wykresy i omówienie

a) Wykres1

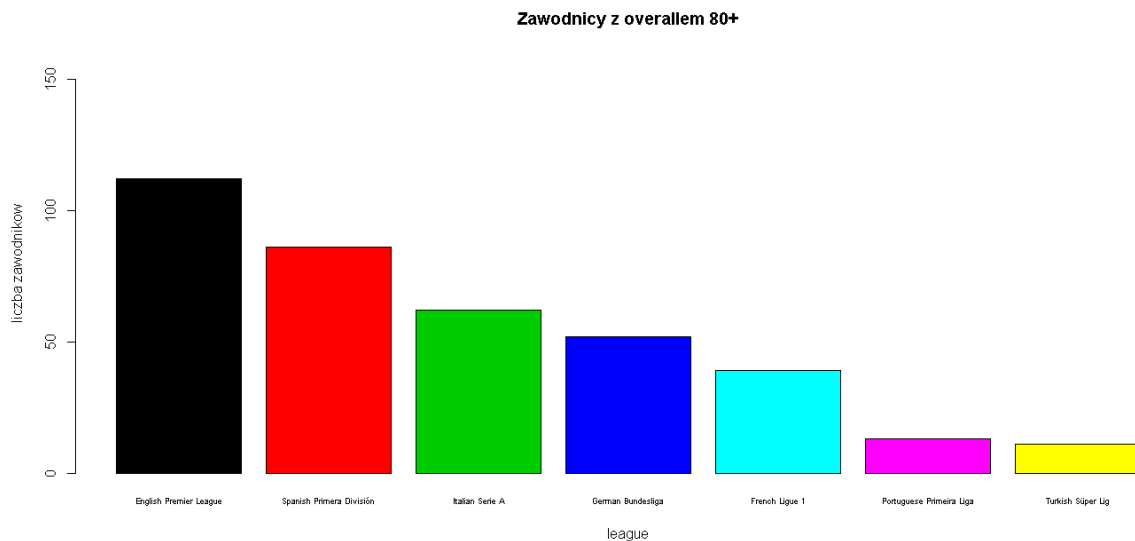


b) Wykres2



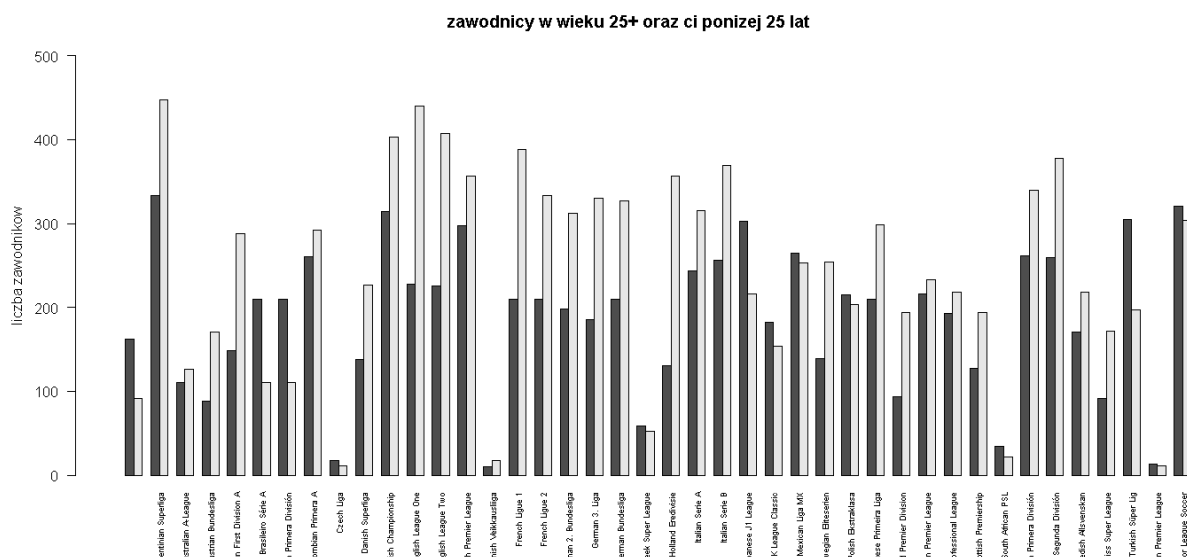
Wykres1 oraz wykres2 przedstawiają odpowiednio średnią ocenę ogólną dla 6lig z największą oceną oraz średnią wieku dla tych lig. Pominięta została liga ukraińska, ponieważ w bazie znajdują się tylko zawodnicy jednego klubu z tej ligi. To samo wpłynęło na wysoki wynik ligi greckiej(4kluby). Poza tym można zauważyć, że średnia ocena w najlepszych ligach jest bardzo podobna i wynosi około 77.

c) Wykres3



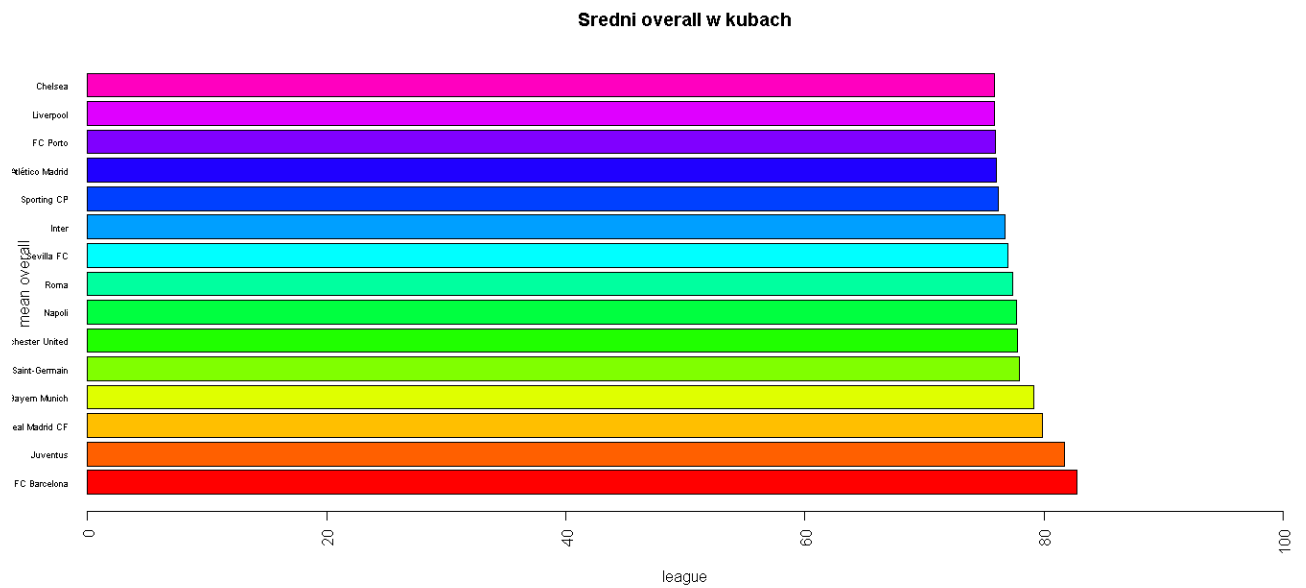
Wykres3 pokazuje rzeczywistą siłę lig poprzez policzenie zawodników z oceną 80+ którzy tam występują. Widać, że w angielskiej lidze występuje zdecydowanie więcej tak dobrych zawodników niż w innych ligach. Na dalszych miejscach znajdują się odpowiednio: hiszpańska, włoska, niemiecka tak jak na wykresie1.

d) Wykres4

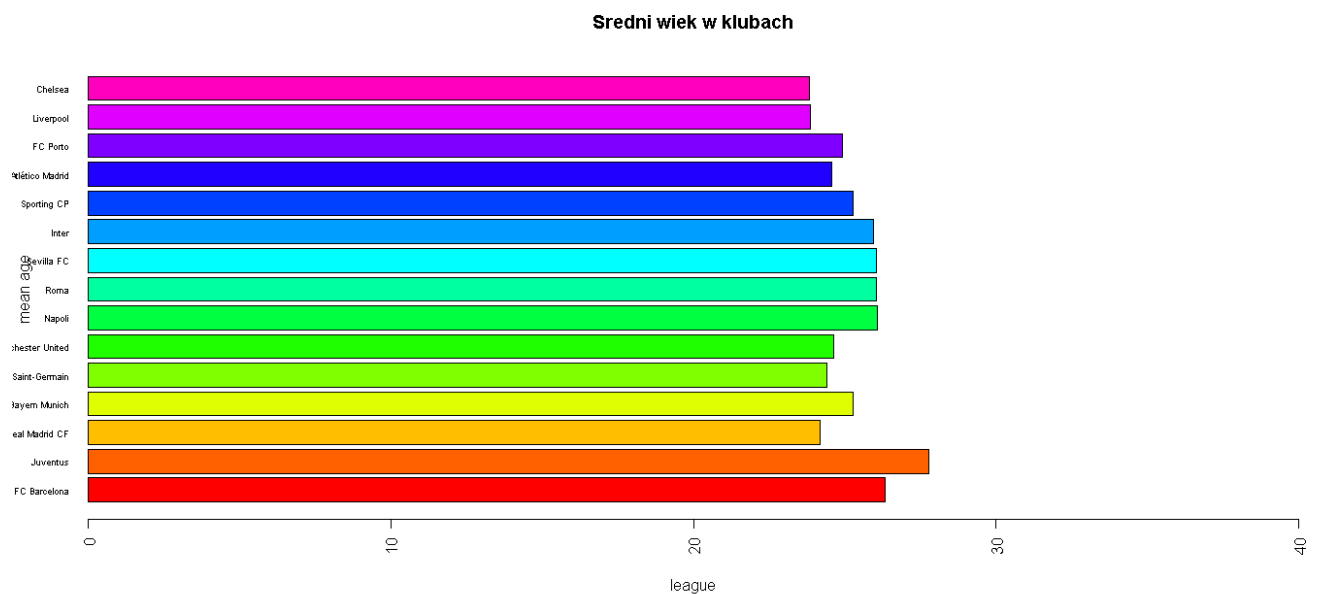


Wykres4 przedstawia ilość zawodników 25+ oraz poniżej 25 lat w ligach. Z tego wykresu można zaobserwować w których ligach baza zawiera informacji o kilku klubach a w których baza jest pełna. Ponadto wykres pokazuje, że w większości lig przeważają młodzi zawodnicy, natomiast w amerykańskiej przez wielu traktowanej jako „piłkarska emerytura” nieznacznie przeważają zawodnicy starsi. Tak samo dzieje się wśród zawodników bez klubu oraz w lidze tureckiej.

e) Wykres5

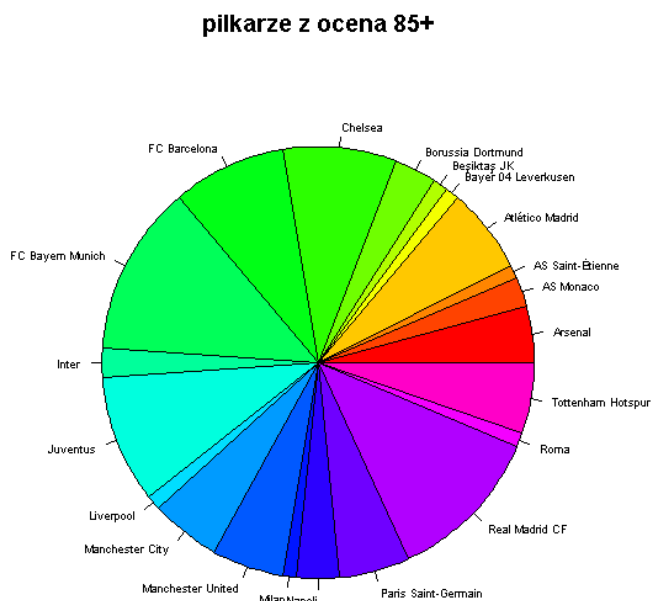


f) Wykres6



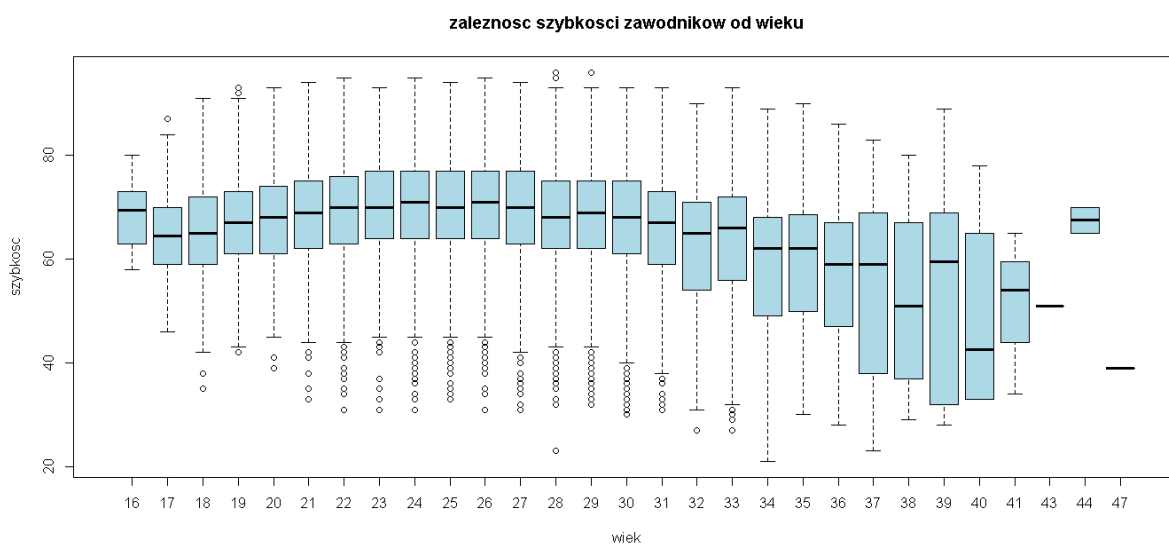
Wykresy 5 i 6 są analogiczne do wykresów 1 oraz 2, z różnicą że przedstawiają te wartości dla klubów. Jak widać największą średnią ocenę mają takie kluby jak Barcelona, Juventus, Real Madryt czy Bayern Monachium. „Najstarszym” klubem wśród tych najsilniejszych jest Juventus a następnie Barcelona.

g) Wykres7



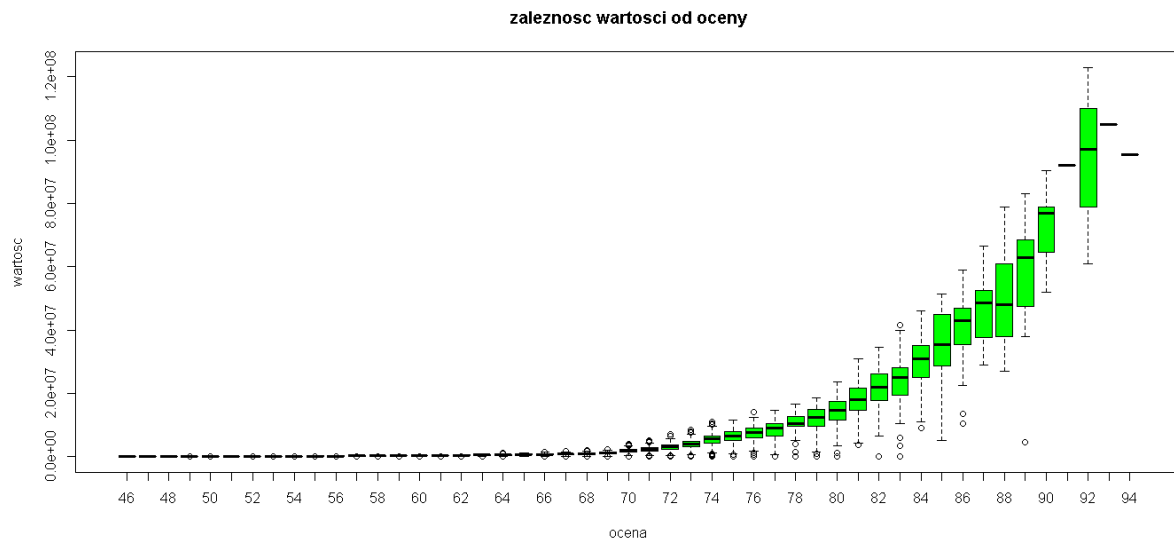
Wykres ten przedstawia ilość gwiazd (piłkarzy z bardzo wysoką oceną) mają poszczególne kluby. Najwięcej takich zawodników ma Real oraz Bayern. Zaraz za nimi jest Chelsea, Juventus i Barcelona. Można z tego wywnioskować że Barcelona i Juventus ma bardziej wyrównaną kadrę od Realu czy Bayernu, ponieważ mimo nieco mniejszej ilości gwiazd mają nieco większą ocenę średnią zespołu.

h) Wykres8

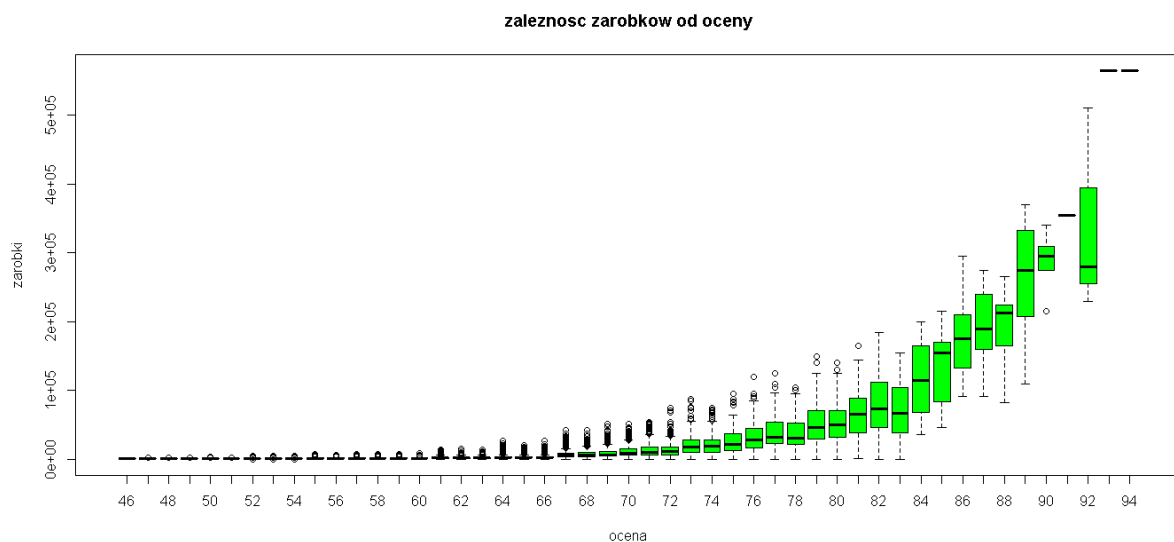


Na podstawie tego wykresu można zauważyć że piłkarze w wieku powyżej 35 lat nie biegają już tak szybko jak piłkarze młodszy.

i) Wykres9

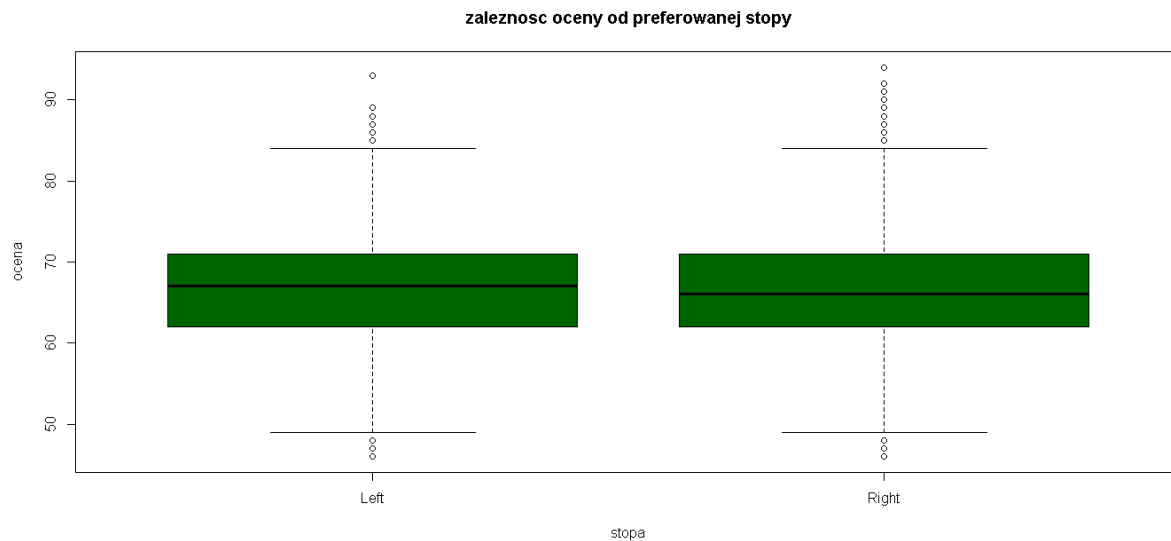


j) Wykres10



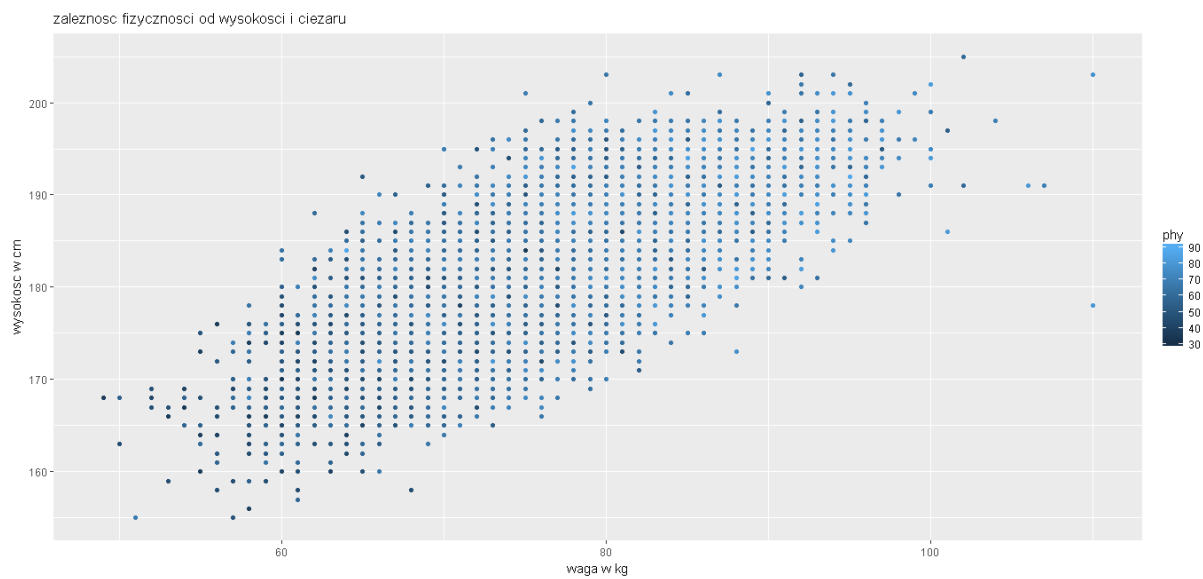
Bez zaskoczenia wykres 9 oraz 10 pokazuje, że wartość oraz zarobki piłkarzy rosną wraz z ich oceną ogólną.

k) Wykres11



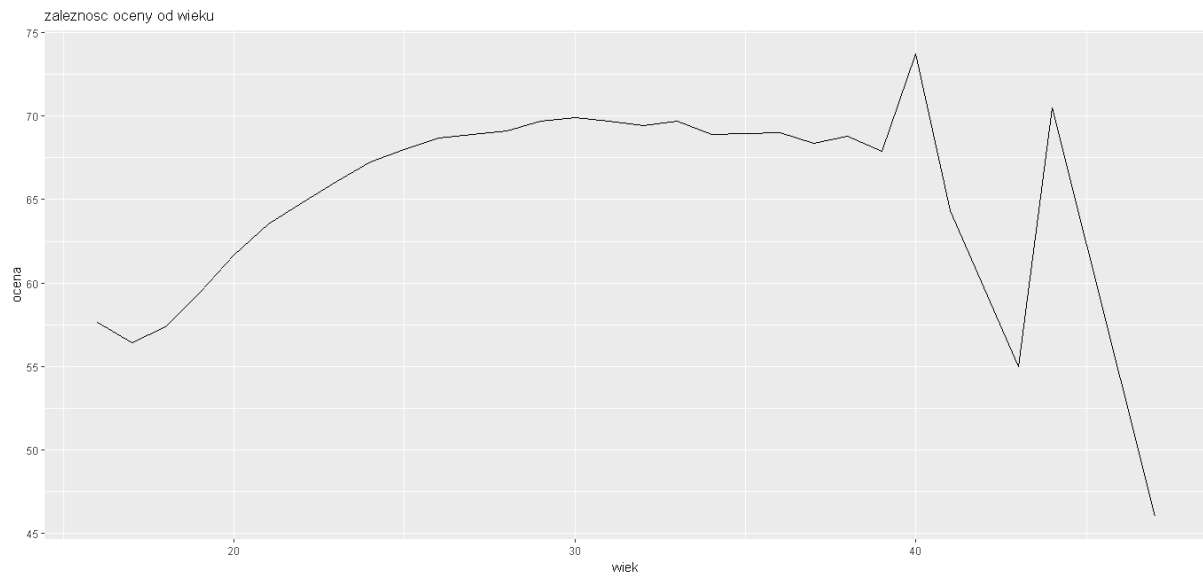
Dzięki wykresowi 11 możemy zobaczyć, że to czy zawodnik jest prawonożny czy lewnożny nie wpływa znacząco na jego ocenę.

I) Wykres12



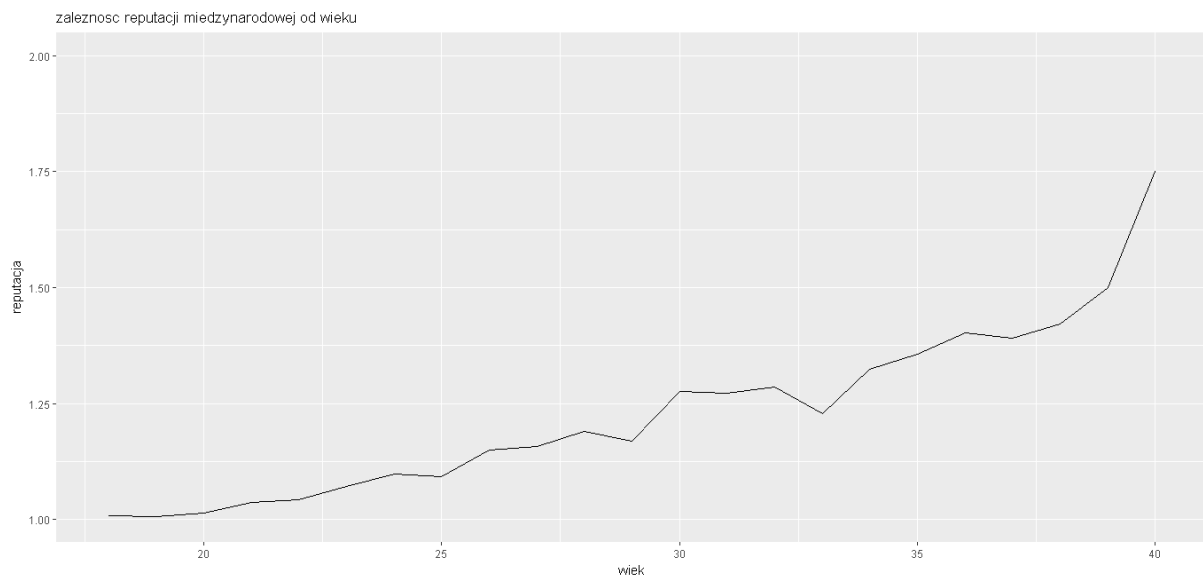
Wykres ten przedstawia wartość fizyczności w zależności od wzrostu oraz wagi zawodnika.

m) Wykres13



Jak widać na wykresie13 najlepszym wiekiem dla zawodnika są okolice 30lat.(baza zawiera małą liczbę zawodników w wieku 40+ dlatego nie bierzemy pod uwagę skoku oceny przy wieku 40lat)

n) Wykres14



Z tego wykresu możemy odczytać, że najmłodsi zawodnicy mają jeden punkt reputacji międzynarodowej i wraz z wiekiem reputacja wzrasta.