

# EST0133 - INTRODUÇÃO À MODELAGEM DE BIG DATA

## Projeto I

Jaylhane Nunes

22/11/2021

```
library(factoextra)
library(ggfortify)
library(gt)
library(hrbrthemes)
library(kableExtra)
library(NbClust)
library(rvest)
library(stringr)
library(tidyverse)
```

## Parte 1 - Clusterização

### 10/10 Questão 1

```
fifa <- read.csv("fifa.csv")

head(fifa)

summary(fifa)
```

Como desejamos identificar padrões podemos remover as variáveis `id`, `club_number` e `national_number`, pois, elas não representam características que contribuem para análise, uma vez que os valores atribuídos nessas variáveis, para cada observação, são escolhidas ao acaso (de acordo com o site Torcedores.com, atualmente, a numeração da camisa é escolhida pelo gosto do jogador).

```
fifa <- fifa %>%
  select(-c(id, club_number, national_number))
```

## 10/10 Questão 2

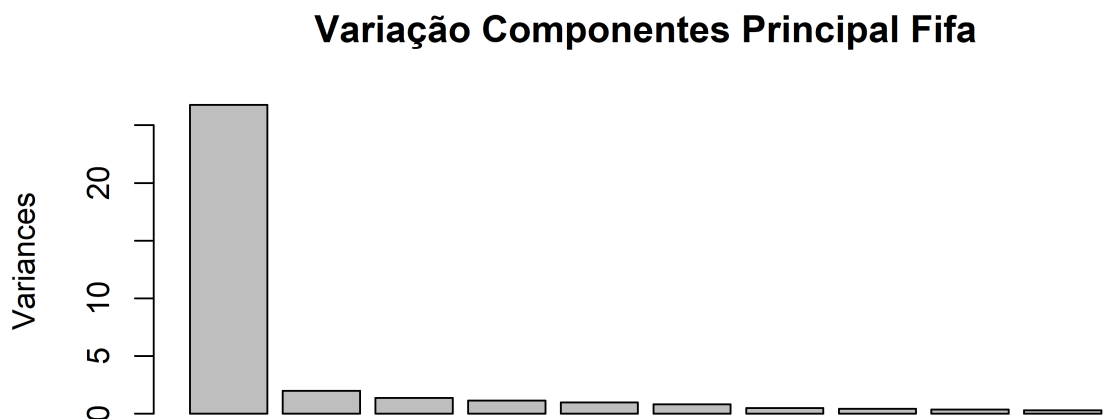
```
fifa.pca <- prcomp(fifa, center = TRUE, scale. = TRUE)

summary(fifa.pca)$importance[,1:10]
```

Tabela 1: Sumário das 10 primeiras componentes principais

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	5.174	1.414	1.176	1.065	0.986	0.898	0.710	0.665	0.617	0.550
Proportion of Variance	0.724	0.054	0.037	0.031	0.026	0.022	0.014	0.012	0.010	0.008
Cumulative Proportion	0.724	0.778	0.815	0.846	0.872	0.894	0.907	0.919	0.930	0.938

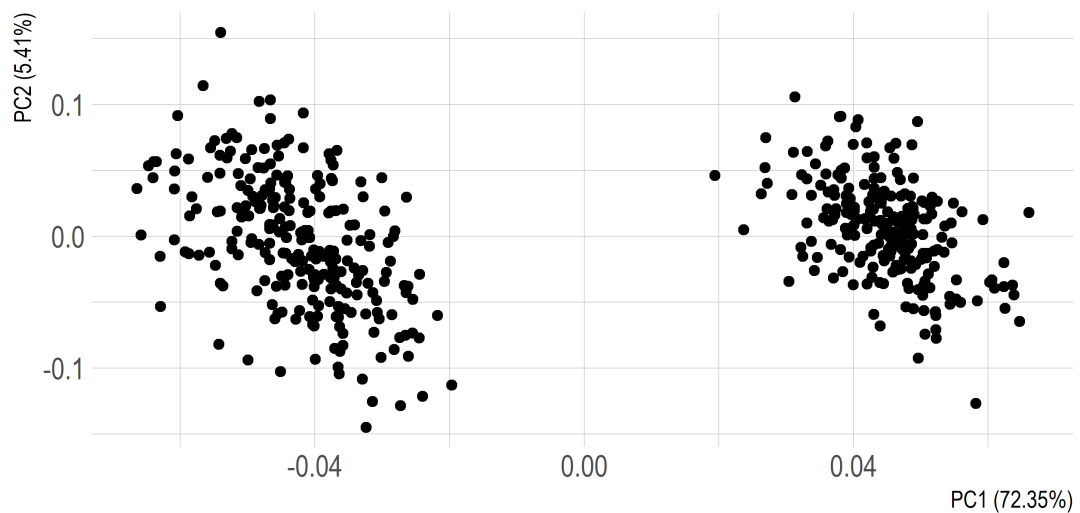
```
plot(fifa.pca,
     main = "Variação Componentes Principal Fifa")
```



De acordo com o **summary** da análise de componentes principais, e o gráfico, a primeira componente principal (PC1) explica a maior parte dos dados, sendo responsável por aproximadamente 72% da variação. Dessa forma, a escolha da primeira e da segunda componente (PC2) são suficientes para prosseguir com a análise, dados que a inclusão da PC2 acrescentou apenas 5% na proporção de variação dos dados.

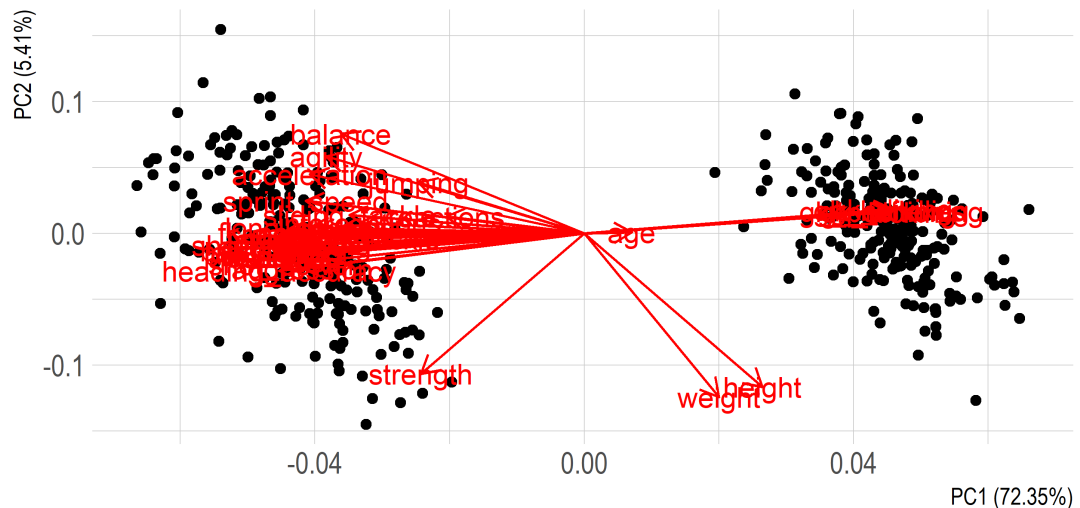
Para continuar a análise iremos verificar como fica a dispersão dos dados para essas duas componentes:

```
autoplot(fifa.pca)
```



Verificando o gráfico de dispersão das componentes, é possível perceber claramente dois grupos distintos. A seguir iremos incluir os vetores das variáveis responsáveis pela formação desses grupos para verificar se é possível identificar de forma visual quais variável exercem maior influência na formação dessas componentes:

```
autoplot(fifa.pca, loadings.label = TRUE)
```

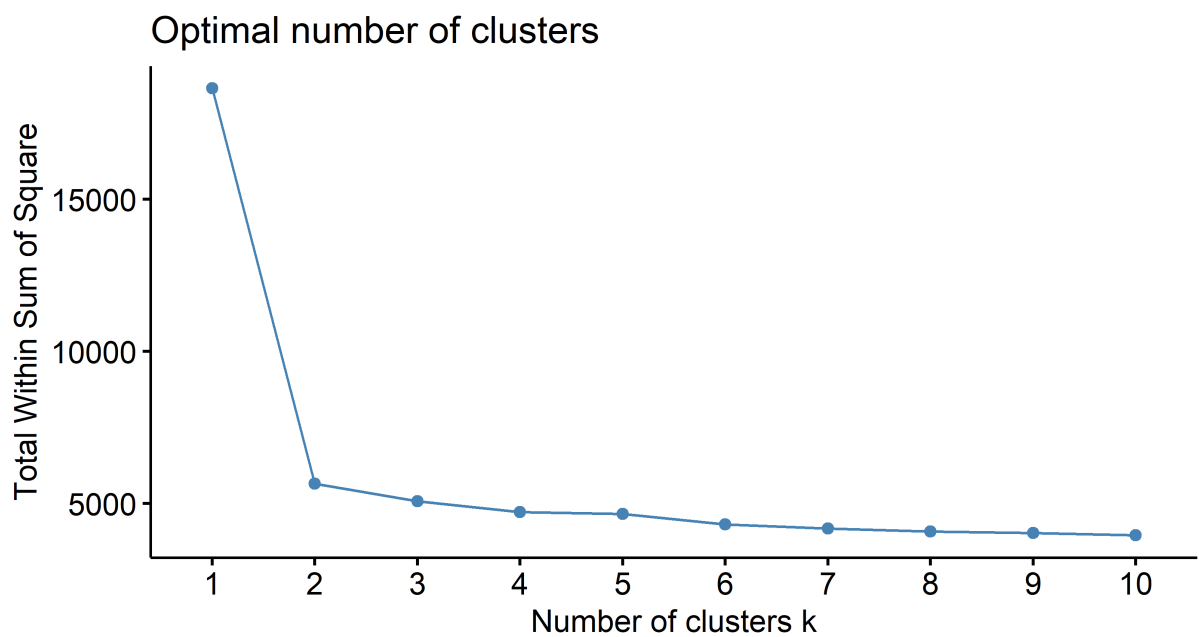


Apesar de não ser possível confirmar a grafia dos nomes devido a sobreposição (*o que já era esperado dado a quantidade de variáveis em análise*), é possível supor que o grupo a direita, e formado pelas variáveis influentes na PC1, é de goleiros, pois, entre as características físicas adequadas a um goleiro apontadas por Csanádi (apud Scopel, Andrade e Levandowski, 2006), estão a estatura (entre 1,75 e 1,90m) e peso

proporcional à altura, que no nosso conjunto de dado são apresentadas em duas variáveis (**height** e **weight**) que estão presente positivamente na PC1.

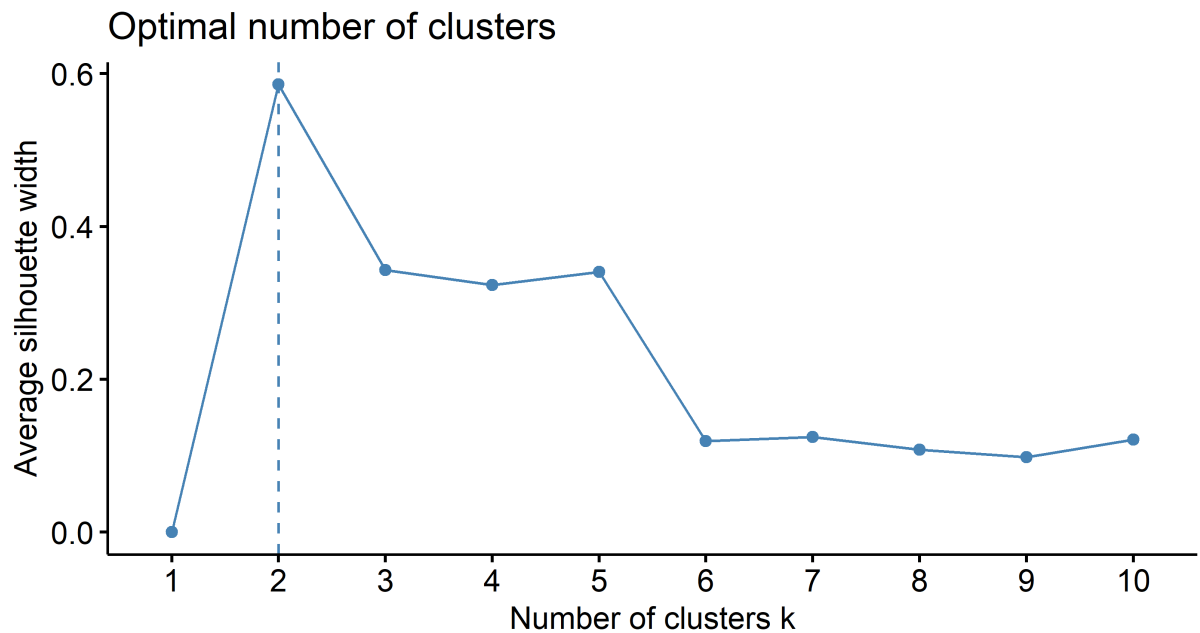
### 10/10 Questão 3

```
fifa.scale <- scale(fifa)
fviz_nbclust(fifa.scale, kmeans, method="wss")
```



De acordo com o método do cotovelo seria necessário dois clusters para dividir os grupos, pois é quando percebemos uma mudança na curvatura, indicando que a partir dali há perda de acurácia na formação dos clusters.

```
fviz_nbclust(fifa.scale, kmeans, method = "silhouette")
```



Já de acordo com o método da silhueta, dois clusteres também seriam suficientes para dividir os grupos, mas seria possível a divisão em 5 clusters de forma parcialmente distinta.

Mas, levando em consideração a quantidade ótima de clusteres do método da silhueta, teríamos dois clusteres, assim como no cotovelo e, concomitantemente, duas posições.

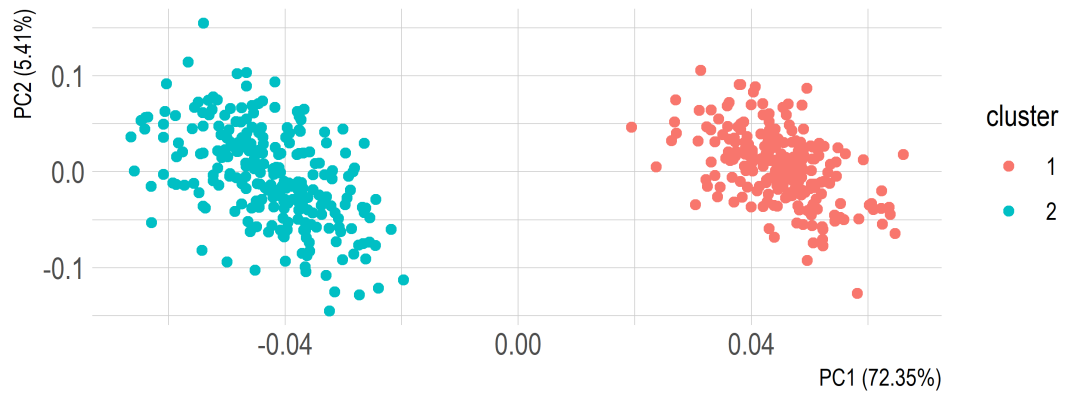
#### 8/10 Questão 4

Verificando graficamente as opções de clusters para as quantidades iguais e maiores do que 2, temos:

```
set.seed(102)
autoplot(kmeans(fifa.scale,2), data = fifa.scale)+
  labs(title = "Gráfico de Análise de Componentes Principais",
        subtitle = "Com 2 clusters")
```

## Gráfico de Análise de Componentes Principais

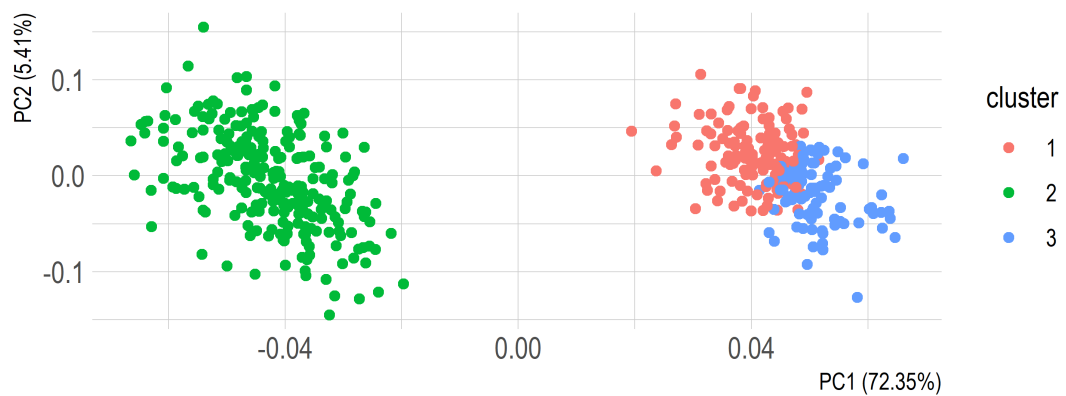
Com 2 clusters



```
set.seed(102)
autoplot(kmeans(fifa.scale,3), data = fifa.scale)+
  labs(title = "Gráfico de Análise de Componentes Principais",
        subtitle = "Com 3 clusters")
```

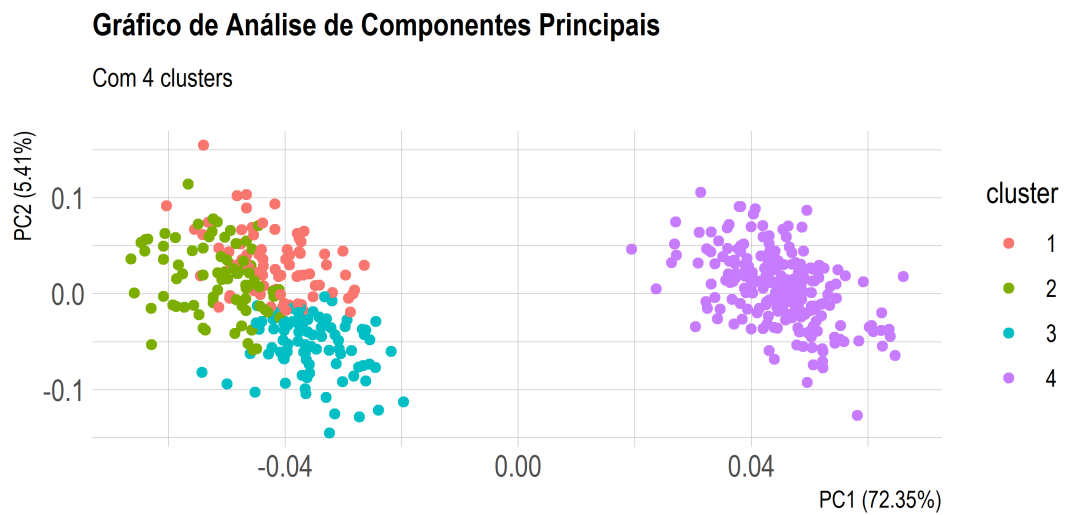
## Gráfico de Análise de Componentes Principais

Com 3 clusters

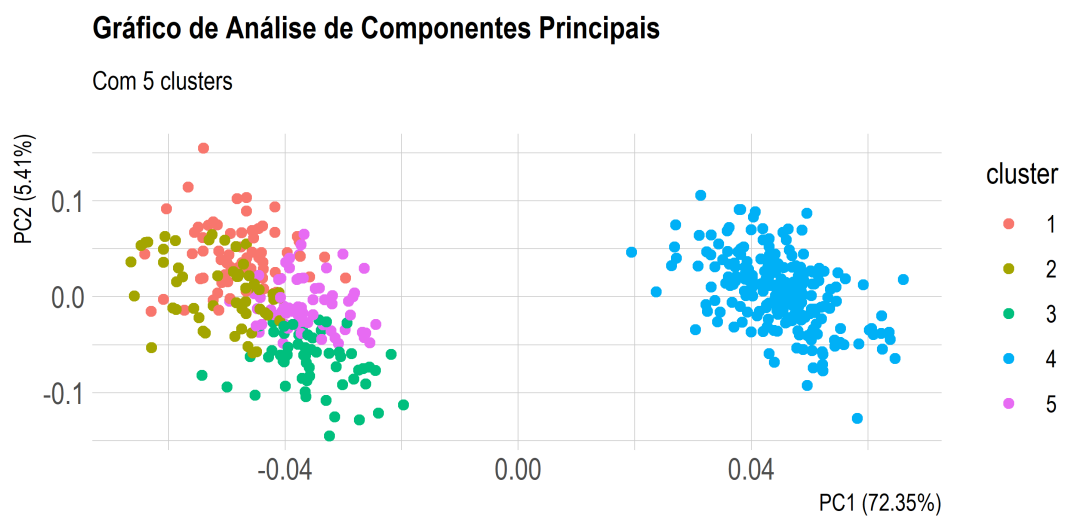


Com 3 clusteres a conclusão referente a posição de goleiro é questionável, pois, o cluster da direita definido como sendo de goleiros, a partir dos **loadings** vistos na questão 2, agora apresenta dois clusteres, o que poderia significar que esse cluster poderia ser de goleiros e defesas, para o caso dessas posições terem características semelhantes.

```
set.seed(102)
autoplot(kmeans(fifa.scale,4), data = fifa.scale)+
  labs(title = "Gráfico de Análise de Componentes Principais",
        subtitle = "Com 4 clusters")
```



```
set.seed(102)
autoplot(kmeans(fifa.scale,5), data = fifa.scale)+
  labs(title = "Gráfico de Análise de Componentes Principais",
        subtitle = "Com 5 clusters")
```



A partir de 4 clusteres já temos um resultado parecido com o de dois clusteres, em que o grupo da direita pode ser constituído exclusivamente pelos goleiros e a diferença estaria no grupo da esquerda, no qual as demais posições se encaixariam.

Algumas suposições que podemos levantar:

- esses clusteres podem estar sendo influenciados pelas habilidades da posição de goleiro, já que, assim como em 4 clusteres, com 5 clusteres o grupo da direita continua sendo homogêneo, mas o grupo da esquerda é heterogêneo.
- já que com mais clusteres perde-se acurácia na separação, incluir mais clusteres misturam as habilidades de acordo com as características das variáveis, de forma que os goleiros passam a formar grupos específicos com relação as características dos não-goleiros e os não-goleiros passam a ser vistos de forma homogênea com relação as características dos goleiros (e por isso eles ficam em grupos distintos).

Em outras palavras, a medida que aumenta-se a quantidade de clusters, o algoritmo de ~~classificação~~ <sup>clusterização</sup> busca dentro dos grupos as similaridades entre as variáveis que formam os grupos, até que chega em um ponto que a similaridade entre o grupo da esquerda (*o grupo de não goleiros*) é a ausência das habilidades de goleiro.

Mas esta análise altera, de alguma maneira, o que foi concluído na Questão 3? Senti falta de objetividade nesta resposta.

## 10/10 Questão 5

De acordo com os números de clusteres sugeridos pelos métodos do cotovelo e da silhueta (*dois clusteres*) iremos estabelecer suposições a respeito de duas posições dos jogadores.

Dessa forma, é possível perceber no conjunto de dados que as variáveis relacionadas às habilidades do goleiro tem como mediana o valor 16, conforme pode ser observado abaixo:

```
sumario_gk <- summary(fifa[,33:37])
```

Tabela 2: Sumário das habilidades de goleiros

gk_diving	gk_handling	gk_kicking	gk_positioning	gk_reflexes
Min. : 1.00	Min. : 1.00	Min. : 1.00	Min. : 1.00	Min. : 1.00
1st Qu.: 9.00	1st Qu.:11.00	1st Qu.:10.00	1st Qu.:10.00	1st Qu.:10.00
Median :16.00	Median :16.00	Median :16.00	Median :16.00	Median :16.00
Mean :42.86	Mean :41.91	Mean :40.28	Mean :42.57	Mean :43.66
3rd Qu.:77.00	3rd Qu.:74.00	3rd Qu.:72.00	3rd Qu.:76.00	3rd Qu.:79.00
Max. :91.00	Max. :92.00	Max. :93.00	Max. :92.00	Max. :90.00

Assim, espera-se que os goleiros sejam aqueles que apresentem valor maior que 16 nas variáveis acima. Para averiguar a suposição, escolhemos ao acaso a variável de posição (**gk\_position**) para definir o grupo, por supor que seria a variável com menor possibilidade de erro.

Com relação ao segundo grupo, o grupo localizado a esquerda, é interessante perceber a partir dos **loadings** na questão 2 que as variáveis que influenciam positivamente a formação da componente 2 (PC2) são as variáveis de: balanço (**balance**), agilidade (**agility**) e um pouco por aceleração (**acceleration**), ar-rancada (**sprint**) e velocidade(**speed**), além dessas, negativamente teríamos a força (**strength**) constituindo essa componente (assim como na PC1).

Baseado nessas características da PC2, podemos definir que o grupo da esquerda é composto pela posição de meio-campo, pois, assim como define FC Futebol, a força é uma característica marcante para a



posição de atacante e defesa e como nesses grupos a presença da força não está correlacionada, podemos considerar que essa habilidade não está presente e portanto essas posições também não. No entanto, já que para a posição de meio-campo exige-se agilidade, podemos dizer que o grupo no cluster a esquerda é do meio-campo.

```
set.seed(102)
fifa.kmeans <- kmeans(fifa.scale, centers = 2)

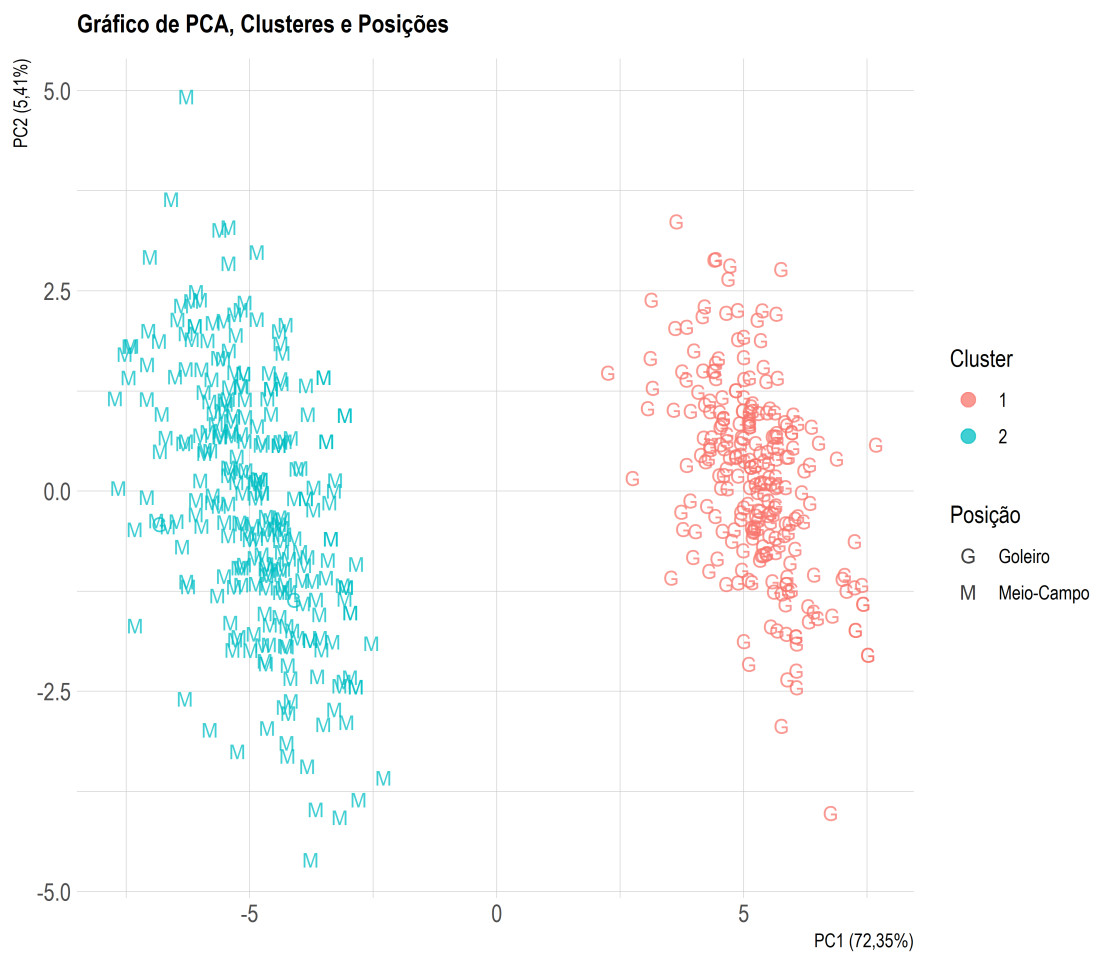
fifa.plot <- data.frame(fifa,
                        fifa.pca$x[,1:2],
                        Posicao=as.character(ifelse(fifa$gk_positioning>16,
                                                    "Goleiro",
                                                    "Meio-Campo")),
                        Cluster=as.character(fifa.kmeans$cluster))

head(fifa.plot)
```

Tabela 3: Exemplo tabela com os clusters e as posições

	gk_kicking	gk_positioning	gk_reflexes	PC1	PC2	Posicao	Cluster
161	9	7	9	-5.622219	-0.1873889	Meio-Campo	2
84	10	6	12	-3.934272	-1.1124188	Meio-Campo	2
136	11	7	8	-5.415313	1.2962229	Meio-Campo	2
431	76	73	77	5.118089	1.4175449	Goleiro	1
497	59	72	76	6.519233	0.6022589	Goleiro	1
144	7	8	14	-5.281930	1.9474138	Meio-Campo	2

```
ggplot(fifa.plot, aes(x=PC1, y=PC2))+
  geom_point(aes(shape=as.factor(Posicao),colour=Cluster), size=3, alpha=.75)+
  scale_shape_manual("Posição",values = c(71,77))+
  labs(x="PC1 (72,35%)", y="PC2 (5,41%)",
       title="Gráfico de PCA, Clusters e Posições")
```



Calculando a média das variáveis por posição temos:

```
medias.posicoes <- fifa.plot %>%
  group_by(Posicao) %>%
  summarise_all(list(mean))
```

Tabela 4: Valores médios das habilidades por posição

Posicao	Goleiro	Meio-Campo
age	29.76829	28.18533
height	189.7520	183.5637
weight	84.07724	78.81853
crossing	15.36992	61.19691
finishing	13.61789	80.29344
heading_accuracy	15.20732	75.85714
short_passing	33.26423	72.84556
volleys	14.14634	74.69498
dribbling	16.50813	76.28571
curve	16.58130	66.94981
fk_accuracy	15.52439	57.23938
long_passing	31.91057	59.25869
ball_control	23.89837	77.72587
acceleration	45.34146	74.13514
sprint_speed	45.44309	75.97683
agility	47.35366	72.28185
reactions	73.51626	77.56371
balance	45.23984	67.49421
shot_power	54.58537	80.35521
jumping	65.06504	75.62162
stamina	34.92683	72.74903
strength	66.94309	78.10039
long_shots	14.47561	72.83012
aggression	29.69106	67.52124
interceptions	19.71545	34.02317
positioning	13.08943	80.64865
vision	49.13415	69.53282
penalties	22.75610	73.82625
composure	54.93089	75.71815
marking	17.60976	34.58687
standing_tackle	15.11789	33.91506
sliding_tackle	14.49187	28.45946
gk_diving	77.471545	9.984556
gk_handling	74.65854	10.81467
gk_kicking	71.85366	10.28185
gk_positioning	76.58943	10.25097

## Parte 2 - Web scraping

### 10/10 Questão 6

```
url <- "https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population"

pagina <- url %>%
  read_html()

pagina <- pagina %>%
  html_table(fill = TRUE)

tabela_atual <- pagina[[1]] %>%
  select(`Country or dependent territory`, Region, Population) %>%
  rename("pais" = "Country or dependent territory",
         "regiao" = "Region",
         "populacao" = "Population")

tabela_atual <- tabela_atual[-1,]

tabela_atual$populacao <- as.numeric(str_replace_all(tabela_atual$populacao, "\\,", ""))

head(tabela_atual)
```

Tabela 5: Observações iniciais dados mais recentes

pais	regiao	populacao
China	Asia	1411778724
India	Asia	1384652149
United States	Americas	332751452
Indonesia	Asia	271350000
Pakistan	Asia	225200000
Brazil	Americas	213980100

```
summary(tabela_atual)
```

Tabela 6: Sumário dos dados mais recentes

pais	regiao	populacao
Length:241	Length:241	Min. :4.000e+01
Class :character	Class :character	1st Qu.:3.013e+05
Mode :character	Mode :character	Median :5.227e+06
		Mean :3.230e+07
		3rd Qu.:1.968e+07
		Max. :1.412e+09

## 10/10 Questão 7

```
url <- "https://en.wikipedia.org/wiki/List_of_countries_by_population_in_2015"

pagina <- url %>%
  read_html()

pagina <- pagina %>%
  html_table(fill = TRUE)

tabela_2015<- pagina[[2]] %>%
  select(`Country / territory`, `Area (km²)[1]`, `Population2015(UN estimate)`) %>%
  rename("pais"="Country / territory",
         "area"="Area (km²)[1]",
         "populacao"="Population2015(UN estimate)")%>%
  mutate(pais = as.character(str_replace_all(pais, "\\.[*\\]", ""))) %>%
  mutate(area = as.numeric(str_replace_all(area, "\\.", "")) ) %>%
  mutate(populacao = as.numeric(str_replace_all(populacao, "\\.", "")) )

tabela_2015 <- tabela_2015[-c(1,200),]

head(tabela_2015)
```

Tabela 7: Observações iniciais dados de 2015

pais	area	populacao
China	9596961	1376048943
India	3287263	1311050527
United States	9833520	321418820
Indonesia	1904569	257563815
Brazil	8515767	207847528
Pakistan	881913	188924874

```
summary(tabela_2015)
```

Tabela 8: Sumário dados de 2015

pais	area	populacao
Length:198	Min. : 722	Min. :1.008e+05
Class :character	1st Qu.: 65122	1st Qu.:1.712e+06
Mode :character	Median : 240302	Median :7.158e+06
	Mean : 850827	Mean :3.624e+07
	3rd Qu.: 685146	3rd Qu.:2.297e+07
	Max. :17075200	Max. :1.376e+09
	NA's :42	

10/10 Questão 8

```
tabela_final <- inner_join(tabela_2015,tabela_atual,by="pais", suffix=c(".2015",".atual"))
head(tabela_final)
```

Tabela 9: Observações iniciais tabela final

pais	area	populacao.2015	regiao	populacao.atual
China	9596961	1376048943	Asia	1411778724
India	3287263	1311050527	Asia	1384652149
United States	9833520	321418820	Americas	332751452
Indonesia	1904569	257563815	Asia	271350000
Brazil	8515767	207847528	Americas	213980100
Pakistan	881913	188924874	Asia	225200000

```
summary(tabela_final)
```

Tabela 10: Sumário dados tabela final

pais	area	populacao.2015	regiao	populacao.atual
Length:178	Min. : 722	Min. :1.008e+05	Length:178	Min. :9.953e+04
Class :character	1st Qu.: 66632	1st Qu.:2.775e+06	Class :character	1st Qu.:2.785e+06
Mode :character	Median : 240302	Median :8.447e+06	Mode :character	Median :9.617e+06
	Mean : 864544	Mean :3.964e+07		Mean :4.285e+07
	3rd Qu.: 702282	3rd Qu.:2.445e+07		3rd Qu.:3.082e+07
	Max. :17075200	Max. :1.376e+09		Max. :1.412e+09
	NA's :28			

10/10 Questão 9

Esta resposta não é o que eu estava pensando na hora em que criei o problema, mas gostei da tua interpretação e dos gráficos criados. Ela me lembrou infográficos de jornais e revistas para o público em geral.

```
tabela_final_plot <- as.data.frame(tabela_final %>%
  group_by(regiao) %>%
  summarise(`2015`=sum(populacao.2015),
    `Atual`=sum(populacao.atual),
    area=sum(area,na.rm = TRUE)))

tabela_final_plot %>%
  GGally::ggparcoord(columns = 2:3,
    groupColumn = "regiao",
    scale = "globalminmax",
    showPoints = TRUE,
    title = "Evolução da População nas Regiões")+
  geom_line(size=1.5)+
  ggrepel::geom_text_repel(data= . %>%
    filter(variable=="Atual"),
```

```

aes(x=variable, y=value, label=regiao),
xlim = c(2, NA))+
scale_y_continuous(limits=c(-1e8,5e9), breaks=seq(0,5e9,1.5e9),
labels = scales::dollar_format(prefix = "",
suffix=" Bilhões",
scale = 1/1e9,
decimal.mark = ",",
big.mark = ".",
accuracy=1.5))+

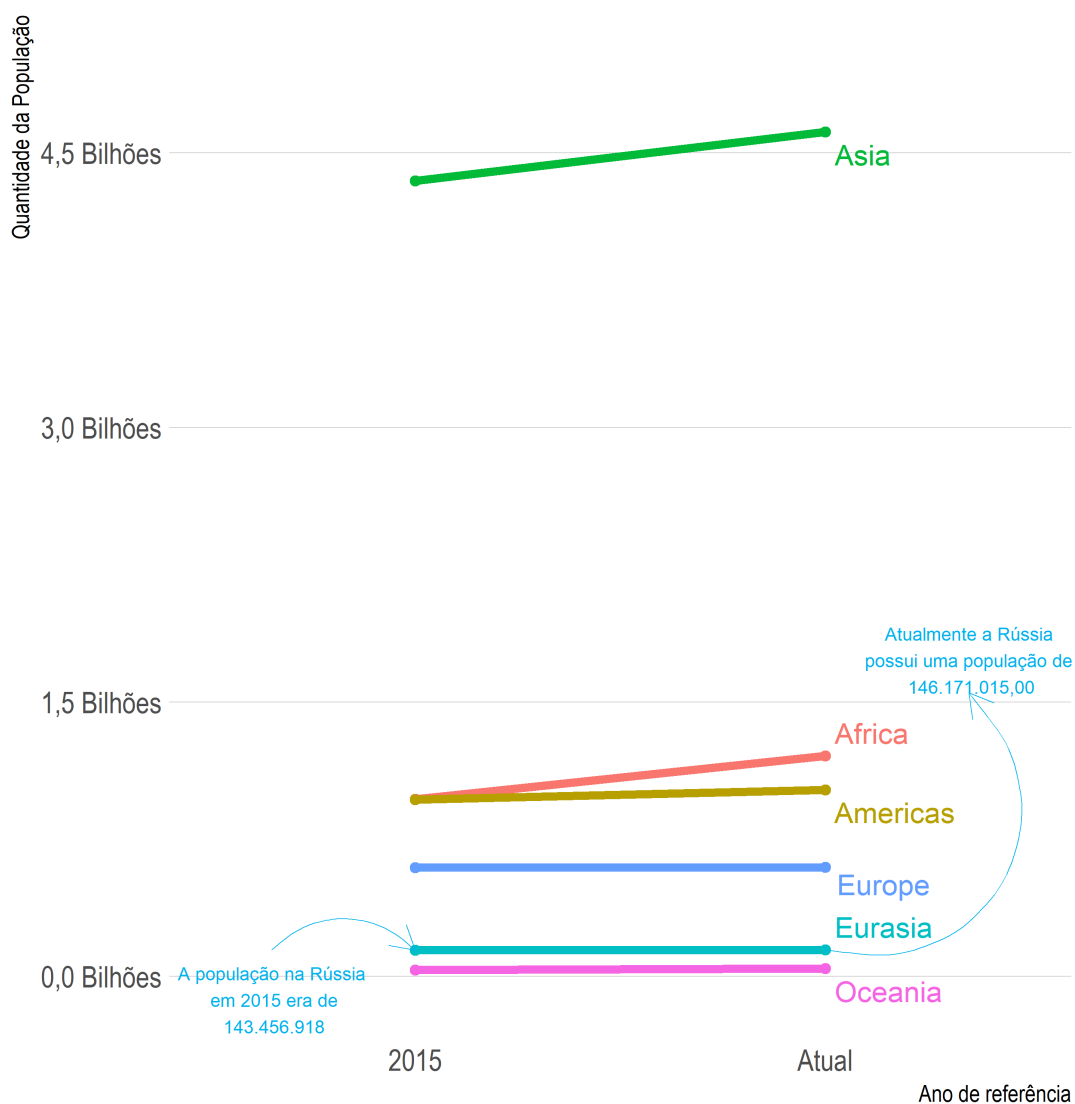
labs(y="Quantidade da População",
x="Ano de referência",
subtitle = "Eurásia é uma região encontrada entre a Europa e a Ásia, formado pelos países:Armêni.
\nAzerbaijão, Cazaquistão,Chipre, Geórgia, Rússia e Turquia. Neste conjunto de dados
\n temos a presença apenas da Rússia,o maior país transcontinental do mundo!
\n A Rússia sozinha possui uma população maior que a Oceania que é formada por 14 países
\n (dos quais 10 estão presentes nesse conjunto de dados).")+
annotate("text",
x=.65,
y=143456918,
vjust=1.25,
label="A população na Rússia\n em 2015 era de\n 143.456.918",
color="deepskyblue2",
size=2.5)+
geom_curve(x=.65,xend=1,
y=143456918+999999, yend=143456918,
curvature = -.45,
color="deepskyblue2",
arrow = arrow(length = unit(0.03, 'npc')),
size=.2)+
annotate("text",
x=2.35,
y=2e9,
vjust=1.25,
label="Atualmente a Rússia\n possui uma população de\n 146.171.015,00",
color="deepskyblue2",
size=2.5)+
geom_curve(x=2,xend=2.35,
y=143456918+999999, yend=1.55e9,
curvature = .85,
color="deepskyblue2",
arrow = arrow(length = unit(0.03, 'npc')),
size=.2)+
theme(legend.position = "none",
panel.grid.minor.x = element_blank(),
panel.grid.major.x = element_blank(),
panel.grid.minor.y = element_blank())

```

## Evolução da População nas Regiões

Eurásia é uma região encontrada entre a Europa e a Ásia, formado pelos países: Armênia, Azerbaijão, Cazaquistão, Chipre, Geórgia, Rússia e Turquia. Neste conjunto de dados temos a presença apenas da Rússia, o maior país transcontinental do mundo!

A Rússia sozinha possui uma população maior que a Oceania que é formada por 14 países (dos quais 10 estão presentes nesse conjunto de dados).





```

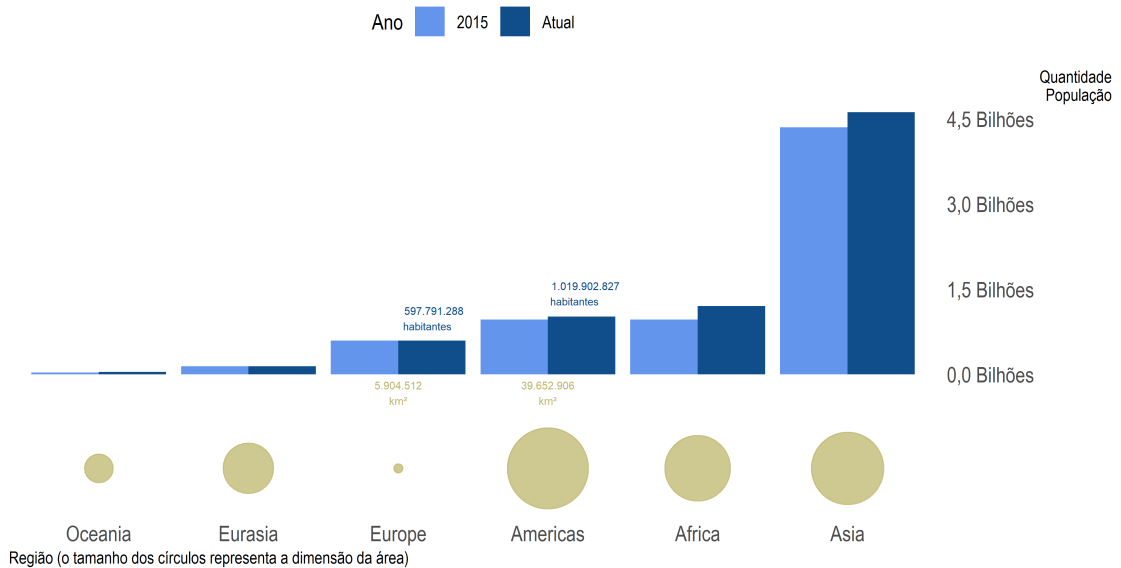
tabela_final_plot%>%
  select(regiao, `2015`,Atual) %>%
  pivot_longer(!regiao,names_to = "ano",values_to = "populacao") %>%
  arrange(populacao) %>%
  ggplot()+
  geom_col(aes(x=fct_inorder(regiao), y=populacao, fill=factor(ano)),position = "dodge")+
  theme(legend.position = "top",
        legend.justification = "center",
        axis.title.y.right = element_text(angle = 0,vjust = 1),
        axis.title.x.bottom = element_text(hjust = 0))+
  labs(title = "Evolução da população por área",
        subtitle = "A região da Europa é quase 7 vezes menor que a área das Américas,
        \nmas atualmente seu número de habitantes representa quase metade da população das Américas",
        color="dodgerblue4",
        x="Região (o tamanho dos círculos representa a dimensão da área)",
        y="Quantidade\n População",
        fill="Ano")+
  scale_fill_manual(values=c("cornflowerblue","dodgerblue4"))+
  scale_y_continuous(limits=c(-2.2e9,5e9), breaks=seq(0,5e9,1.5e9),
                     labels = scales::dollar_format(prefix = "",
                                                       suffix=" Bilhões",
                                                       scale = 1/1e9,
                                                       decimal.mark = ",",
                                                       big.mark = ".",
                                                       accuracy=1.5),
                     position = "right")+
  geom_point(data=tabela_final_plot,aes(x=regiao, y=-1.65e9,size=area),
            alpha=0.75, color="darkkhaki", show.legend = FALSE) +
  scale_size(range = c(2, 20), name="Area (km²)")+
  annotate("text",x=3, y=-2e8,vjust=.8,
          label="5.904.512\nkm²", color="darkkhaki", size=2)+
  annotate("text",x=4, y=-2e8,vjust=.8,
          label="39.652.906\nkm²", color="darkkhaki", size=2)+
  annotate("text",x=3, y=7e8,vjust=-.2, hjust=-.1,
          label="597.791.288\nhabitantes", color="dodgerblue4", size=2)+
  annotate("text",x=4, y=1.2e9,vjust=-.05, hjust=-.05,
          label="1.019.902.827\nhabitantes", color="dodgerblue4", size=2)+
  ggExtra::removeGrid()

```

### Evolução da população por área

A região da Europa é quase 7 vezes menor que a área das Américas,

mas atualmente seu número de habitantes representa quase metade da população das Américas



De forma geral, é nítido que a Ásia se destaca com relação a quantidade de população (sendo a região mais populosa), dado que sua área é similar a da África e menor que a das Americas.

Além disso, África e Ásia foram as regiões que apresentaram maior crescimento populacional em número absoluto comparando com as outras regiões, sendo que elas aparentam ter a mesma área, mas a Ásia apresententa quase o triplo de habitantes da África.

### 10/10 Questão 10

```

variacao_pop <- tabela_final %>%
  select(-area) %>%
  group_by(regiao) %>%
  mutate(diferenca = round(100*(populacao.atual - populacao.2015)/populacao.2015,2)) %>%
  mutate(media_regiao=mean(diferenca)) %>%
  arrange(-diferenca) %>%
  arrange(-media_regiao)

names(variacao_pop) <- c("País", "2015", "Região", "Atual", "Diferença", "Var. Média na Região")

variacao_pop <- variacao_pop %>%
  gt() %>%
  fmt_percent(columns = 5:6, scale_values = FALSE, decimals = 2)

```

A tabela abaixo apresenta a África como a região com maior média de variação percentual entre seus países (média de 26,04%), seguida da: Ásia (17,14%), Oceania (16,30%), Americas (10,08%), Eurasia (1,89%)

e Europa (-0,36%). Observa-se ainda que as duas últimas praticamente não tiveram variação populacional de 2015 à atualidade.

País	2015	Atual	Diferença	Var. Média na Região
Africa				
Somalia	9358602	16360000	74.81%	26.04%
Angola	18992708	32097671	69.00%	26.04%
Niger	15203822	24112753	58.60%	26.04%
Gabon	1501266	2233272	48.76%	26.04%
Burundi	8518862	12574571	47.61%	26.04%
Chad	11714904	16818391	43.56%	26.04%
Sierra Leone	5835664	8297882	42.19%	26.04%
Gambia	1750732	2487000	42.05%	26.04%
Zambia	13257269	18400556	38.80%	26.04%
Senegal	12509434	17223497	37.68%	26.04%
Tanzania	43187823	59441988	37.64%	26.04%
Mozambique	22416881	30832244	37.54%	26.04%
Burkina Faso	15730977	21510181	36.74%	26.04%
Benin	9211741	12506347	35.77%	26.04%
Mali	15370000	20856000	35.69%	26.04%
Malawi	13947592	18898441	35.50%	26.04%
Uganda	31784600	42885900	34.93%	26.04%
Madagascar	20146442	26923353	33.64%	26.04%
Mauritania	3217383	4271197	32.75%	26.04%
São Tomé and Príncipe	165397	214610	29.75%	26.04%
Togo	6191155	7886000	27.38%	26.04%
Ghana	24223431	30792608	27.12%	26.04%
Guinea	10323755	12907395	25.03%	26.04%
Central African Republic	4505945	5633412	25.02%	26.04%
Zimbabwe	12644041	15790716	24.89%	26.04%
Rwanda	10412820	12955768	24.42%	26.04%
Algeria	35978000	44700000	24.24%	26.04%
Egypt	83483000	102594320	22.89%	26.04%
Cameroon	19958352	24348251	22.00%	26.04%
South Africa	49991300	60142978	20.31%	26.04%
Botswana	2029307	2410338	18.78%	26.04%
Ethiopia	99390750	117876000	18.60%	26.04%
Kenya	40862900	47564296	16.40%	26.04%
Nigeria	182201962	211401000	16.03%	26.04%
Western Sahara	530000	612000	15.47%	26.04%
Namibia	2212037	2550226	15.29%	26.04%
Equatorial Guinea	1313000	1505588	14.67%	26.04%
Morocco	31894000	36434318	14.24%	26.04%
Lesotho	1891830	2159000	14.12%	26.04%
Liberia	4101767	4661010	13.63%	26.04%
Comoros	675000	758316	12.34%	26.04%
Tunisia	10547100	11746695	11.37%	26.04%
Djibouti	879053	976107	11.04%	26.04%
Cape Verde	512582	563198	9.87%	26.04%
Libya	6545619	6959000	6.32%	26.04%
Sudan	42338426	43971505	3.86%	26.04%
Guinea-Bissau	1647380	1646077	-0.08%	26.04%
Mauritius	1283415	1266334	-1.33%	26.04%

Eritrea	5223994	3601000	−31.07%	26.04%
Asia				
Jordan	6113000	11083072	81.30%	17.14%
Lebanon	3785655	6769000	78.81%	17.14%
Qatar	1699435	2799202	64.71%	17.14%
Oman	2773479	4513666	62.74%	17.14%
Afghanistan	24485600	32890171	34.32%	17.14%
Yemen	23153982	30491000	31.69%	17.14%
Kuwait	3566437	4670713	30.96%	17.14%
Palestine	4048403	5227193	29.12%	17.14%
Saudi Arabia	27136977	35013414	29.02%	17.14%
Iraq	32481000	41190700	26.81%	17.14%
Uzbekistan	28001400	35217340	25.77%	17.14%
Tajikistan	7616400	9504000	24.78%	17.14%
Israel	7623600	9433340	23.74%	17.14%
Kyrgyzstan	5418300	6663000	22.97%	17.14%
Mongolia	2780800	3408767	22.58%	17.14%
Bahrain	1234571	1501635	21.63%	17.14%
Maldives	319738	383135	19.83%	17.14%
Pakistan	188924874	225200000	19.20%	17.14%
Laos	6230200	7337783	17.78%	17.14%
Kazakhstan	16442000	19043456	15.82%	17.14%
Malaysia	28334135	32729000	15.51%	17.14%
United Arab Emirates	8264070	9503738	15.00%	17.14%
Azerbaijan	8997586	10139196	12.69%	17.14%
Turkmenistan	5479800	6118000	11.65%	17.14%
Philippines	100981437	111125245	10.05%	17.14%
Brunei	414400	453600	9.46%	17.14%
Cambodia	14302779	15552211	8.74%	17.14%
Bhutan	695822	756129	8.67%	17.14%
Nepal	28043744	30378055	8.32%	17.14%
Turkey	77266814	83614362	8.22%	17.14%
Singapore	5076700	5453600	7.42%	17.14%
Iran	79109272	84965255	7.40%	17.14%
Sri Lanka	20653000	22156000	7.28%	17.14%
Bangladesh	160995642	171741894	6.67%	17.14%
Vietnam	91508084	97580000	6.64%	17.14%
India	1311050527	1384652149	5.61%	17.14%
North Korea	24325701	25660000	5.49%	17.14%
Indonesia	257563815	271350000	5.35%	17.14%
South Korea	49410366	51671569	4.58%	17.14%
China	1376048943	1411778724	2.60%	17.14%
Taiwan	23162120	23430948	1.16%	17.14%
Japan	126573481	125070000	−1.19%	17.14%
Thailand	67959359	66720717	−1.82%	17.14%
Myanmar	59780000	55294979	−7.50%	17.14%
Armenia	3249482	2963900	−8.79%	17.14%
Syria	20619000	18276000	−11.36%	17.14%
Georgia	4436391	3728573	−15.95%	17.14%
Oceania				
Solomon Islands	530669	728041	37.19%	16.30%
Papua New Guinea	6744955	9122994	35.26%	16.30%

Vanuatu	245036	301295	22.96%	16.30%
Kiribati	100835	120740	19.74%	16.30%
New Zealand	4367800	5135664	17.58%	16.30%
Australia	22299800	25902487	16.16%	16.30%
Samoa	183123	199853	9.14%	16.30%
Fiji	850700	898402	5.61%	16.30%
Micronesia	102624	105754	3.05%	16.30%
Tonga	103365	99532	−3.71%	16.30%
Americas				
Belize	312971	430191	37.45%	10.08%
Ecuador	14483499	17861648	23.32%	10.08%
Panama	3504483	4278500	22.09%	10.08%
Haiti	9855000	11743017	19.16%	10.08%
Guatemala	14361666	17109746	19.13%	10.08%
Honduras	8045990	9450711	17.46%	10.08%
Chile	17094270	19678363	15.12%	10.08%
Paraguay	6459727	7353038	13.83%	10.08%
Nicaragua	5822265	6595674	13.28%	10.08%
Bolivia	10426154	11797257	13.15%	10.08%
Costa Rica	4563539	5163038	13.14%	10.08%
Argentina	40518951	45808747	13.06%	10.08%
Canada	34108752	38484238	12.83%	10.08%
Suriname	531170	598000	12.58%	10.08%
Colombia	45508205	51049498	12.18%	10.08%
Peru	29461933	33035304	12.13%	10.08%
Bahamas	353658	393450	11.25%	10.08%
El Salvador	6194126	6825935	10.20%	10.08%
Dominican Republic	9884371	10535535	6.59%	10.08%
Uruguay	3356584	3554915	5.91%	10.08%
Barbados	276300	288000	4.23%	10.08%
Trinidad and Tobago	1317714	1367558	3.78%	10.08%
United States	321418820	332751452	3.53%	10.08%
Grenada	109553	113000	3.15%	10.08%
Brazil	207847528	213980100	2.95%	10.08%
Saint Lucia	174000	178696	2.70%	10.08%
Saint Vincent and the Grenadines	109284	110696	1.29%	10.08%
Jamaica	2701200	2734093	1.22%	10.08%
Venezuela	28833845	28705000	−0.45%	10.08%
Cuba	11241161	11181595	−0.53%	10.08%
Mexico	127017224	126014024	−0.79%	10.08%
Guyana	761442	743699	−2.33%	10.08%
Eurasia				
Russia	143456918	146171015	1.89%	1.89%
Europe				
Luxembourg	502066	634730	26.42%	−0.36%
Malta	414372	514564	24.18%	−0.36%
Iceland	317630	374830	18.01%	−0.36%
Ireland	4467854	5011500	12.17%	−0.36%
Sweden	9340682	10435447	11.72%	−0.36%
Switzerland	7785806	8696088	11.69%	−0.36%
Norway	4858199	5415166	11.46%	−0.36%

Austria	8375290	8976467	7.18%	−0.36%
Belgium	10839905	11580185	6.83%	−0.36%
Netherlands	16574989	17660681	6.55%	−0.36%
Denmark	5534738	5867412	6.01%	−0.36%
Germany	78728000	83129285	5.59%	−0.36%
France	64395345	67505000	4.83%	−0.36%
United Kingdom	64715810	67081234	3.66%	−0.36%
Finland	5351427	5516184	3.08%	−0.36%
Spain	45989016	47394223	3.06%	−0.36%
Slovenia	2046976	2108977	3.03%	−0.36%
Czech Republic	10506813	10702942	1.87%	−0.36%
Montenegro	616411	621306	0.79%	−0.36%
Slovakia	5424925	5459781	0.64%	−0.36%
Poland	38167329	38153000	−0.04%	−0.36%
Estonia	1340127	1330068	−0.75%	−0.36%
Belarus	9480178	9349645	−1.38%	−0.36%
Italy	60340328	59108671	−2.04%	−0.36%
Portugal	10637713	10347892	−2.72%	−0.36%
Hungary	10014324	9730772	−2.83%	−0.36%
Greece	11305118	10718565	−5.19%	−0.36%
Serbia	7291436	6871547	−5.76%	−0.36%
Croatia	4425747	4036355	−8.80%	−0.36%
Bulgaria	7563710	6875040	−9.10%	−0.36%
Ukraine	45782592	41319838	−9.75%	−0.36%
Romania	21462186	19186201	−10.60%	−0.36%
Albania	3195000	2829741	−11.43%	−0.36%
Bosnia and Herzegovina	3844046	3320954	−13.61%	−0.36%
Latvia	2248374	1879700	−16.40%	−0.36%
Lithuania	3329039	2780427	−16.48%	−0.36%
Kosovo	2208107	1782115	−19.29%	−0.36%
Cyprus	1102677	888005	−19.47%	−0.36%
Moldova	3563695	2597100	−27.12%	−0.36%

---