

Analizando o Fluxo Imigratório usando dados do Banco Mundial

A imigração vem sendo um tema central nos debates políticos para presidência, gerando muitas dúvidas nos eleitores, por conta de *fake news* que acabam surgindo, e também muita esperança para as pessoas que buscam mudar de vida com a mudança de país. Esta onda imigratória está cada vez chamando mais a atenção da mídia, por conta da grave crise na Syria, Yemen e Venezuela, que fez com que milhões de pessoas decidissem sair de seus lares.

Esta atenção da mídia fez com que estudos mais aprofundados sobre o tema surgissem, visando desmistificar alguns fatos sobre o tema. Um estudo que me chamou bastante a atenção foi o realizado pela [The Economist](#), onde se conduziram diversas reportagens sobre aspectos do fluxo imigratório, como o impacto nos países que recebem as pessoas, empregabilidade, bem-estar, entre outros. A reportagem foi muito bem feita e o tema me atrai bastante, por isso resolvi fazer este post, para tentar analisar os dados do fluxo imigratório.

Na tentativa de complementar os *insights* gerados pela reportagem, vou usar os dados do Banco Mundial, através de uma API que pode ser acessada pelo R com o pacote **wbstats**. Esta API me permitirá acessar o banco de dados desta instituição, que me parece ser a melhor fonte para dados de fluxo imigratório.

A imagem abaixo é de uma das reportagens da revista. Fantástica!



Antes de iniciar a análise dos dados, vou listar abaixo as principais conclusões das reportagens desenvolvidas nesta edição da The Economist.

- 1) Pessoas que conseguem migrar de seus países, geralmente saindo de um país pobre e imigrando para um país rico, auferem um aumento de 3 a 6 vezes em sua renda, por conta de que países ricos possuem melhores instituições, melhor alocação de capital e empresas mais modernas.
- 2) Se todas as pessoas que desejassem se mudar de país conseguissem, o PIB mundial poderia dobrar, por conta de que a mão de obra e o capital estaria sendo melhor utilizado.
- 3) Imigrantes têm mais chance de abrir seu próprio negócio no país em que entram, pois conseguem notar com mais facilidade demandas não atendidas, e também por conta de que descobrem soluções alternativas para problemas existentes na sociedade em que entram. Em outras palavras, os imigrantes olham para as necessidades da sociedade em que entram com outra perspectiva, pensando fora da caixa.

- 4) O maior “problema” com relação à imigração é a mudança cultural (se é que se pode chamar isso de um problema), que ocorre rapidamente em locais que recebem um fluxo migratório, por conta de que os imigrantes se agrupam em locais específicos. Além de ser o principal problema, é algo difícil de se solucionar, pois é difícil mensurá-lo e contestá-lo.
- 5) As pessoas são mais tolerantes ao fluxo migratório em seus países quando notam que seu governo está sob controle da situação e também das fronteiras de seu país
- 6) Os imigrantes na maioria das vezes compartilham seus rendimentos com seus familiares que ficaram no país de origem. Esta forma de investimento é chamada de [remittances](#). Estas remessas já constituem a maior fonte de investimento direto externo de alguns países e seu potencial de impacto é alto, visto que é um investimento que vai direto para as pessoas que precisam e o risco de serem desviada por corrupção é menor. Atualmente, em 28 países, estas remessas representam mais de 10% de seus respectivos PIBs.

Analizando os Dados do Fluxo Migratório

Agora que entendemos alguns aspectos do fluxo migratório, nós podemos analisar alguns datasets da base de dados do Banco Mundial para nos aprofundar um pouco mais no tema. Para isso, vamos usar os seguintes pacotes.

```
#install.packages("wbstats")
#install.packages("patchwork")
#install.packages("pander")

library(tidyverse) # data science and data manipulation packages
library(wbstats)   # World Bank API
library(patchwork)
```

A primeira informação que vale a pena ser vista sobre o tema é a quantidade de pessoas consideradas imigrantes pelo Banco Mundial. O gráfico abaixo irá mostrar a série temporal desta variável.

O código abaixo salva na variável ‘migrants_pop’ o resultado da chamada da API do Banco Mundial. Veja que é necessário especificar qual indicador se deseja trazer, o período e quais países para ser retornado.

```
# Creating a variable with the data.frame of the requested data
migrants_pop <- wbstats::wb(indicator = "SM.POP.TOTL",
                             country = "countries_only",
                             startdate = 1980,
                             enddate = 2020)
```

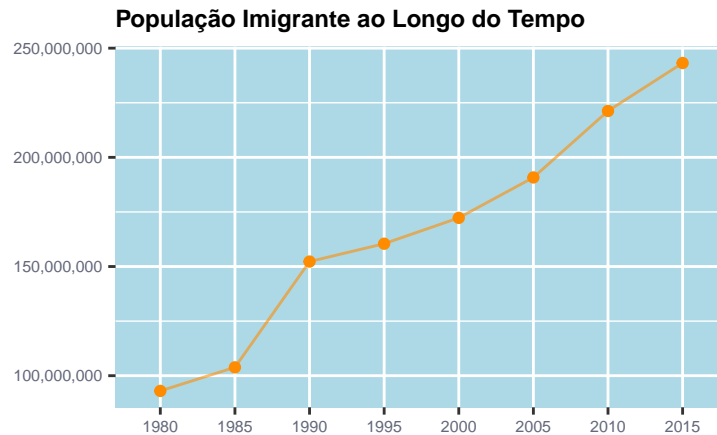
Após fazer a requisição, é possível plotar a série temporal.

```
migrants_pop %>%
  group_by(date) %>%
  summarise(immigrants = sum(value)) %>%
  ggplot()+
  geom_line(mapping = aes(x = date, y = immigrants, group = 1),
            size = .5,
            color = "dark orange",
            alpha = .6)+
  geom_point(mapping = aes(x = date, y = immigrants),
```

```

    color = "dark orange")+
  scale_y_continuous(labels = scales::comma)+
  theme_graph()+
  labs(title = "População Imigrante ao Longo do Tempo",
       x = "",
       y = "")

```



Veja que o dataset trouxe dados até 2015, mesmo eu tendo solicitado até 2020. Em 2015 a população imigrante era de 250 milhões (maior do que a população do Brasil). Agora vamos verificar os países que possuem maior proporção de imigrantes em seus territórios. O próximo gráfico irá trazer os 20 países com maior proporção de população imigrante. Provavelmente Canadá e Austrália estarão presentes neste rank. Irei remover países com população abaixo de 1 milhão de habitantes.

```

# Trazendo a população dos países

```

```

population <- wbstats::wb(indicator = "SP.POP.TOTL",
                          country = "countries_only")

```

```

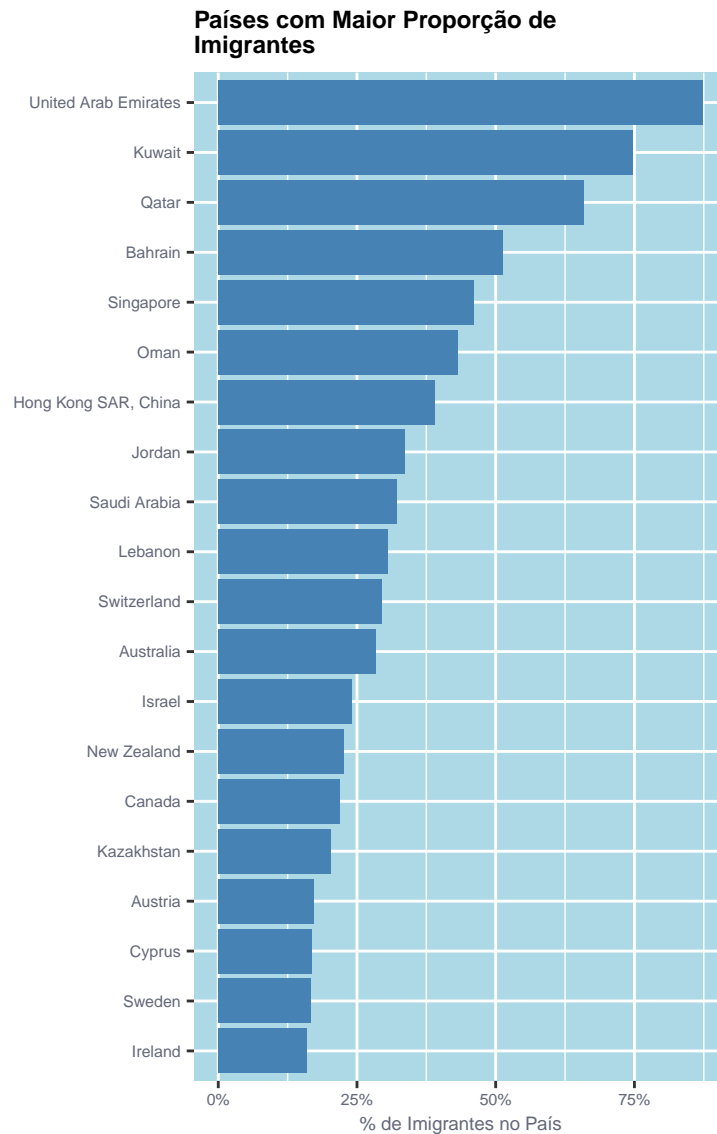
# Juntando os dois datasets de população e imigrantes

```

```

migrants_pop %>%
  filter(date == "2015") %>%
  select(country, date, value) %>%
  inner_join(population, by = c("country" = "country", "date" = "date")) %>%
  mutate(share_immigrants = round(value.x/value.y, digits = 4)) %>%
  filter(value.y >= 1000000) %>% # Retirando países com baixa população
  arrange(desc(share_immigrants)) %>%
  head(20) %>%
  mutate(country = fct_reorder(country, share_immigrants)) %>%
  ggplot()+
  geom_col(mapping = aes(x = country, y = share_immigrants),
          fill = "steel blue")+
  scale_y_continuous(labels = scales::percent)+
  theme_graph()+
  labs(title = "Países com Maior Proporção de \nImigrantes", # \n indica quebra de linha para
       x = "",
       y = "% de Imigrantes no País")+
  coord_flip()

```



Alguns países me surpreenderam estar na lista, como por exemplo Arábia Saudita. Este [artigo](#) reforça esse número.

Se por um lado, alguns países receberam muitos imigrantes, por outro, alguns países perderam muitos residentes. Vamos ver os países que mais perderam residentes nos últimos anos. O Banco Mundial possui um indicador que mostra a saída líquida de pessoas, e com essa métrica, podemos classificar os países do menor valor para o maior e selecionar os 20 primeiros. Com isso, o resultado será os 20 países que mais perderam pessoas.

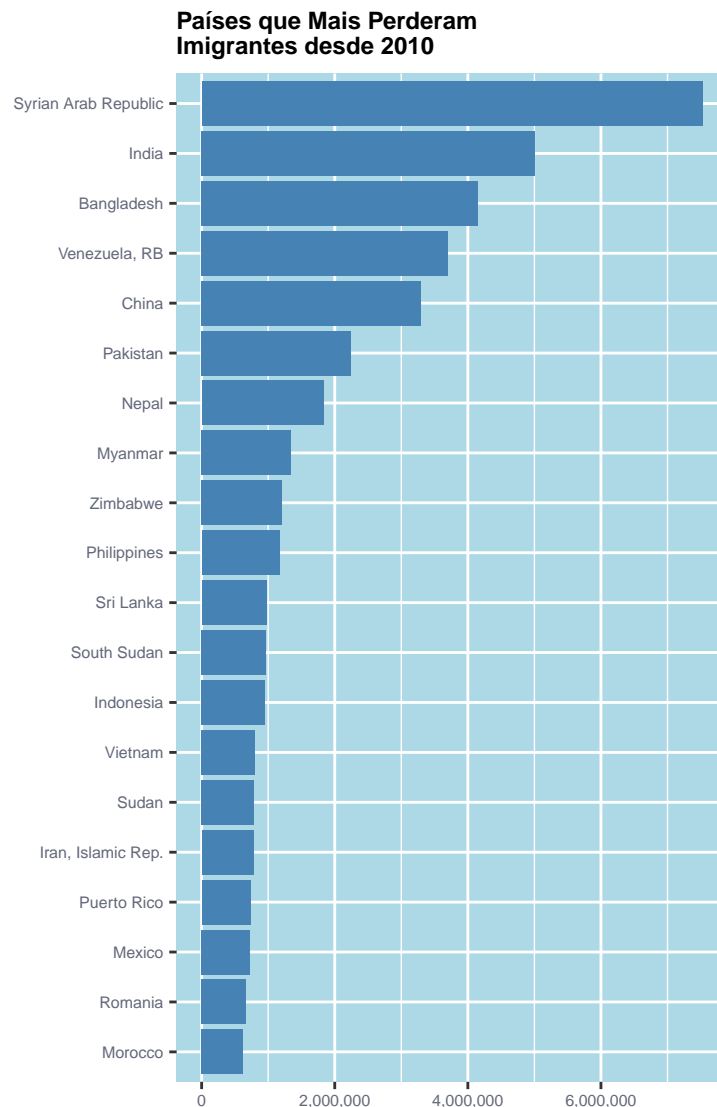
```
net_immigration <- wbstats::wb(indicator = "SM.POP.NETM",
                                country = "countries_only",
                                startdate = 2010,
                                enddate = 2019) %>%
  select(country, value, date)
```

```
net_immigration %>%
  group_by(country) %>%
  summarise(net_immigration = sum(value)) %>%
  arrange(net_immigration) %>%
```

```

head(20) %>%
mutate(net_immigration = net_immigration * -1,
       country = fct_reorder(country, net_immigration)) %>%
ggplot()+
geom_col(mapping = aes(x = country, y = net_immigration),
         fill = "steel blue")+
scale_y_continuous(labels = scales::comma)+
theme_graph()+
coord_flip()+
labs(title = "Países que Mais Perderam \nImigrantes desde 2010",
     x = "",
     y = "")

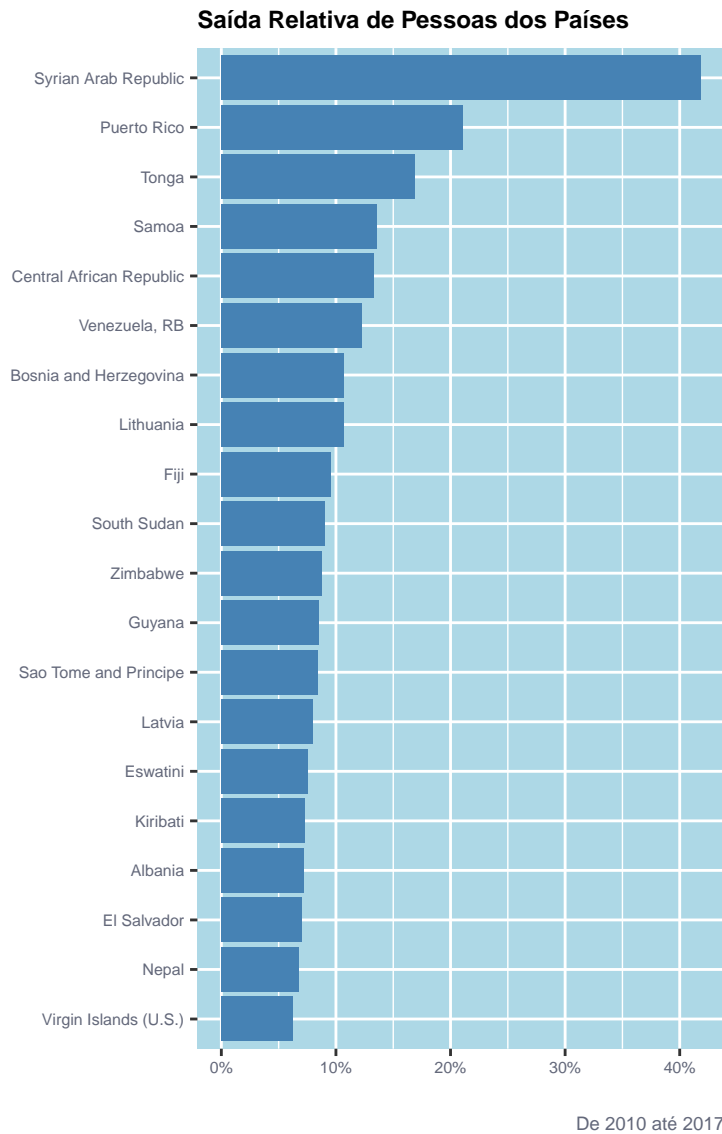
```



Síria e Venezuela são dois países que eu esperava ver neste gráfico. Porém Índia e China aparecem por conta de sua grande população, pois um pequeno movimento de seus habitantes é algo grande comparado aos outros. O ideal é normalizar este dataset usando a população de cada país, de forma que a saída de pessoas seja relativa à sua população. Para isso, basta cruzarmos este dataset com o dataset de população do país para criar a métrica de ‘saída líquida relativa de pessoas’.

```
# Net immigration summed during the period
net_immigration %>%
  group_by(country) %>%
  summarise(net_immigration = sum(value)) -> net_immigration

# Joinning the population and net immigration data
population %>%
  filter(date == 2015) %>%
  mutate(population = value) %>%
  select(country, population) %>%
  inner_join(net_immigration, by = c("country" = "country")) %>%
  mutate(net_immigration_share = round(net_immigration/population, digits = 4)) %>%
  arrange(net_immigration_share) %>%
  head(20) %>%
  mutate(net_immigration_share = net_immigration_share * -1,
         country = fct_reorder(country, net_immigration_share)) %>%
  ggplot()+
  geom_col(mapping = aes(x = country, y = net_immigration_share),
          fill = "steel blue")+
  scale_y_continuous(labels = scales::percent)+
  theme_graph()+
  coord_flip()+
  labs(title = "Saída Relativa de Pessoas dos Países",
       x = "",
       y = "",
       caption = "De 2010 até 2017")
```



Com essa visão ajustada, podemos ver outros países compondo o ranking. Seguindo adiante, vamos verificar as 'remittances' que são enviadas por imigrantes para seus familiares que estão em dificuldades. Um dado importante indicado nas reportagens da revista é que este fluxo de investimento direto representa mais de 10% do PIB de 29 países. Vamos ver se este indicador é corroborado pelo dataset do Banco Mundial.

Por sorte o Banco Mundial já mensura o percentual de 'remittances' sobre o PIB, desta forma desenvolver esta tabela ficou bem simples.

```
remittances_over_gdp <- wbstats::wb(indicator = "BX.TRF.PWKR.DT.GD.ZS",
                                     country = "countries_only")

remittances_over_gdp %>%
  filter(date == "2018",
         country != "Lesotho",
         value > 10) %>%
  mutate("Remittance over gdp" = paste(round(value,digits = 2),"%",sep = "")) %>%
  arrange(desc(value)) %>%
  mutate(rows = row_number()) %>%
  select(rows, date, country, "Remittance over gdp") %>%
  knitr::kable(format.args = list(decimal.mark = ",", big.mark = "."))
```


rows	date	country	Remittance over gdp
1	2018	Tonga	37.49%
2	2018	Kyrgyz Republic	32.51%
3	2018	Tajikistan	28.12%
4	2018	Somalia	25.33%
5	2018	Nepal	25.03%
6	2018	El Salvador	20.72%
7	2018	Bermuda	20.02%
8	2018	Honduras	19.85%
9	2018	Haiti	18.03%
10	2018	Samoa	17.97%
11	2018	West Bank and Gaza	17.41%
12	2018	Moldova	16.04%
13	2018	Jamaica	15.9%
14	2018	Kosovo	15.68%
15	2018	Comoros	14.52%
16	2018	Uzbekistan	14.46%
17	2018	Marshall Islands	14.19%
18	2018	Liberia	13.44%
19	2018	Guatemala	12.89%
20	2018	Lebanon	12.71%
21	2018	Gambia, The	12.69%
22	2018	Armenia	11.94%
23	2018	Cabo Verde	11.93%
24	2018	Georgia	11.56%
25	2018	Nicaragua	11.55%
26	2018	Ukraine	11.23%
27	2018	Bosnia and Herzegovina	11.21%
28	2018	Montenegro	10.69%
29	2018	Senegal	10.5%
30	2018	Jordan	10.42%
31	2018	Egypt, Arab Rep.	10.22%
32	2018	Kiribati	10.18%

Aparentemente são 31 países nesta situação, o que corrobora o que foi indicado na reportagem, visto que alguns países desta tabela são bem pequenos, e podem ter sido desconsiderados pela revista.

Outra informação indicada nas reportagens é de que as remessas enviadas pelos familiares aos seus parentes é um fluxo acíclico. Isso quer dizer que o fluxo não varia em conformidade com a variação do PIB do país que recebe este investimento. Essa hipótese tem um fundamento sólido, visto que o imigrante, ao ver seus parentes precisando de ajuda (por conta de uma crise econômica), envia mais dinheiro para que eles consigam passar por esta situação adversa. Apesar deste embasamento, é necessário ver empiricamente esta afirmação.

Vou selecionar alguns países com baixo PIB per capita para ver se o fluxo de 'remittances' de fato assume um comportamento acíclico. Os países serão: Suriname, Somalia, Serra Leoa, Senegal, Sudão, Ruanda, Nepal, Nigéria, Malawi, Mali, Líbia, Quênia, Etiópia, Eritreia e Congo.

```
# Creating a variable with the selected countries
country_view <- c("Yemen, Rep.", "Tanzania", "Uganda",
                  "Tunisia", "Chad", "Syrian Arab Republic",
```


Table 1: Amostra do dataset

	date	country	remittances_adj
55	2014	Malawi	2.3627
1	2020	Congo, Rep.	2.2893
51	2018	Malawi	9.9799
109	2020	Sierra Leone	22.1928
133	2016	Chad	0.0000

```
" Suriname","Somalia","Sierra Leone",
"Senegal","Sudan","Rwanda","Nepal",
"Nigeria","Malawi","Mali","Libya",
"Kenya","Ethiopia","Eritrea",
"Congo, Rep.")
```

Agora vamos ver o fluxo de 'remittances' e o PIB per Capita.

```
remittances_received <- wbstats::wb(indicator = "BX.TRF.PWKR.CD.DT",
                                   country = "countries_only")

gdp_pc <- wbstats::wb(indicator = "NY.GDP.PCAP.CD",
                     country = "countries_only")
```

Tendo as duas variáveis, antes de compará-las, será necessário normalizar o fluxo de 'remittances' pela população. O seguinte código fará esta normalização.

```
remittances_received %>%
  mutate(remittances = value) %>%
  select(date, country, remittances) %>%
  filter(country %in% country_view,
         date > 2010) %>%
  inner_join(population, by = c("date" = "date", "country" = "country")) %>%
  select(date, country, remittances, value) %>%
  mutate(remittances_adj = round(remittances/value, digits = 4)) %>%
  select(date, country, remittances_adj) -> remittances

# Gerando uma amostra do dataset
remittances[sample(nrow(remittances),5), ] %>%
  knitr::kable(caption = "Amostra do dataset")
```

Agora eu posso cruzar os dois datasets para ver a correlação das duas variáveis para os países selecionados.

```
gdp <- gdp_pc %>%
  mutate(gdp_pc = value) %>%
  select(date, country, gdp_pc) %>%
  inner_join(remittances, by = c("date" = "date", "country" = "country")) %>%
  group_by(date) %>%
  summarise(gdp = sum(gdp_pc),
            remittance = sum(remittances_adj)) %>%
```

```

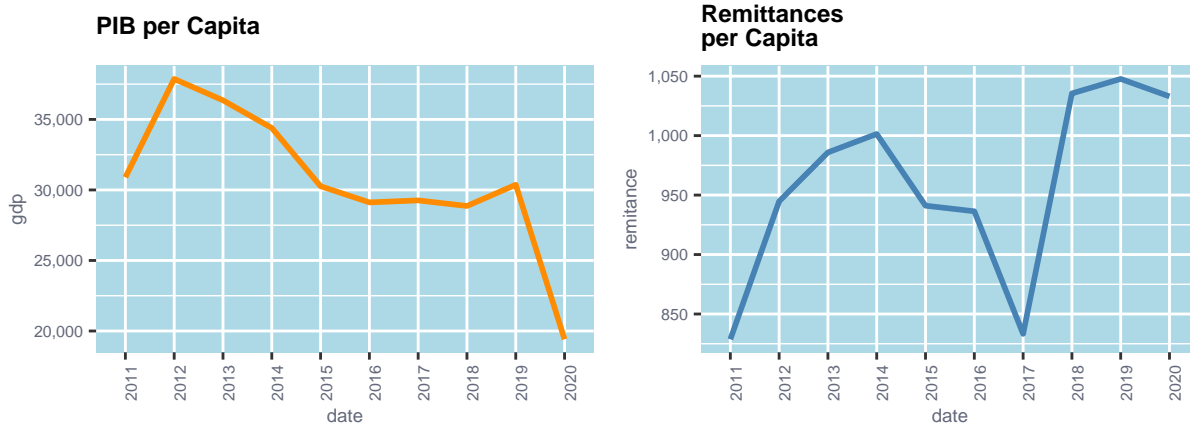
ggplot()+
  geom_line(mapping = aes(x = date, y = gdp, group = 1),
            color = "dark orange", size = 1)+
  scale_y_continuous(labels = scales::comma)+
  theme_graph()+
  theme(axis.text.x = element_text(angle = 90))+
  labs(title = "PIB per Capita")

remittances <- gdp_pc %>%
  mutate(gdp_pc = value) %>%
  select(date, country, gdp_pc) %>%
  inner_join(remittances, by = c("date" = "date", "country" = "country")) %>%
  group_by(date) %>%
  summarise(gdp = sum(gdp_pc),
            remittance = sum(remittances_adj)) %>%

ggplot()+
  geom_line(mapping = aes(x = date, y = remittance, group = 1),
            color = "steel blue", size = 1)+
  scale_y_continuous(labels = scales::comma)+
  theme_graph()+
  theme(axis.text.x = element_text(angle = 90))+
  labs(title = "Remittances \nper Capita")

gdp+remittances

```



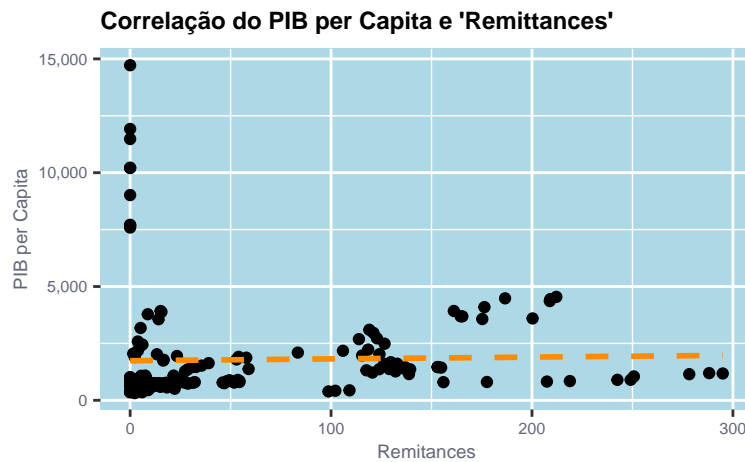
Nota-se que a relação não é direta, porém também não é inversa. Desta forma, esta relação acíclica não fica evidente ao vermos estas duas séries temporais. O gráfico abaixo mostra a correlação destas duas séries temporais.

```

gdp_pc %>%
  mutate(gdp_pc = value) %>%
  select(date, country, gdp_pc) %>%
  inner_join(remittances, by = c("date" = "date", "country" = "country")) %>%
  ggplot()+
  geom_point(mapping = aes(x = remittances_adj, y = gdp_pc))+

```

```
geom_smooth(mapping = aes(x = remittances_adj, y = gdp_pc),
            method = "lm",
            formula = y ~ x,
            se = F,
            lty = 2,
            color = "dark orange")+
scale_y_continuous(labels = scales::comma)+
scale_x_continuous(labels = scales::comma)+
theme_graph()+
labs(title = "Correlação do PIB per Capita e 'Remittances'",
     y = "PIB per Capita",
     x = "Remittances")
```



Eu esperava ver uma correlação mais forte, onde o PIB per Capito reduzido iria trazer um valor de 'remittances' per Capita alto. Porém não ficou tão clara esta correlação. O modelo de regressão linear para este contexto pode ser do seguinte formato:

$$R = \beta_0 + \beta_1 pc.gdp + \mu$$

Este modelo tentará explicar variações do fluxo de 'remittances' per Capita em função do PIB per Capita, de forma que se possa prever o valor da variável dependente com um dado grau de confiança.

```
gdp_pc %>%
  mutate(gdp_pc = value) %>%
  select(date, country, gdp_pc) %>%
  inner_join(remittances, by = c("date" = "date", "country" = "country")) %>%
  lm(formula = remittances_adj ~ gdp_pc) %>%
  summary() %>%
  pander::pander()
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	54.28	7.071	7.676	1.224e-12
gdp_pc	0.0008202	0.00247	0.3321	0.7402

Table 3: Fitting linear model: remittances_adj ~ gdp_pc

Observations	Residual Std. Error	R^2	Adjusted R^2
172	72.54	0.0006483	-0.00523

Com o output do modelo, podemos reescrever a equação inserindo os valores do intercepto e inclinação.

$$R = 19.29 + 0.03542pc.gdp + \mu$$

Embora o resultado tenha saído diferente do que eu esperava, tanto o intercepto quanto a inclinação se mostram estatisticamente significativos, com um valor-p abaixo de 5%.