

# Trabalho Final

## MC536

### Cavaleiros da Query Estruturada

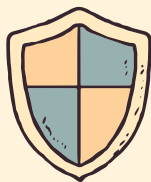
Willian Takayuki Ozako - 207267

Frederico Meletti Rappa - 216047

João Alberto Moreira Seródio - 218548



# Sumário



## Introdução

- Motivação
- Objetivos



## Base de Dados

- Bases de Dados



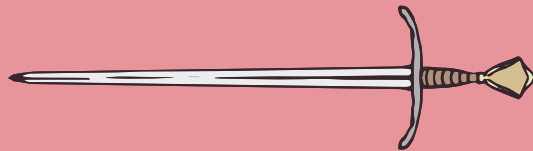
## Metodologia e Resultados

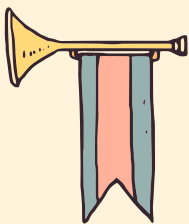
- Modelo Conceitual
- Modelos Lógicos
- Tratamento de Dados
- Queries (SQL e Cypher)
- Visualização



01

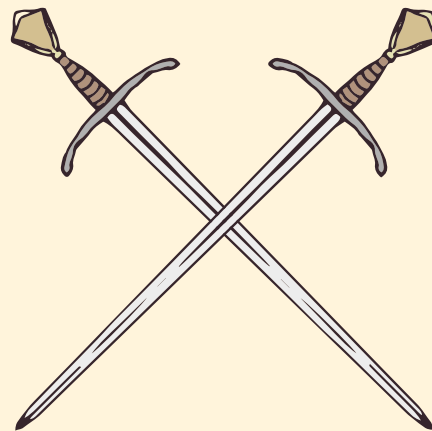
# Introdução





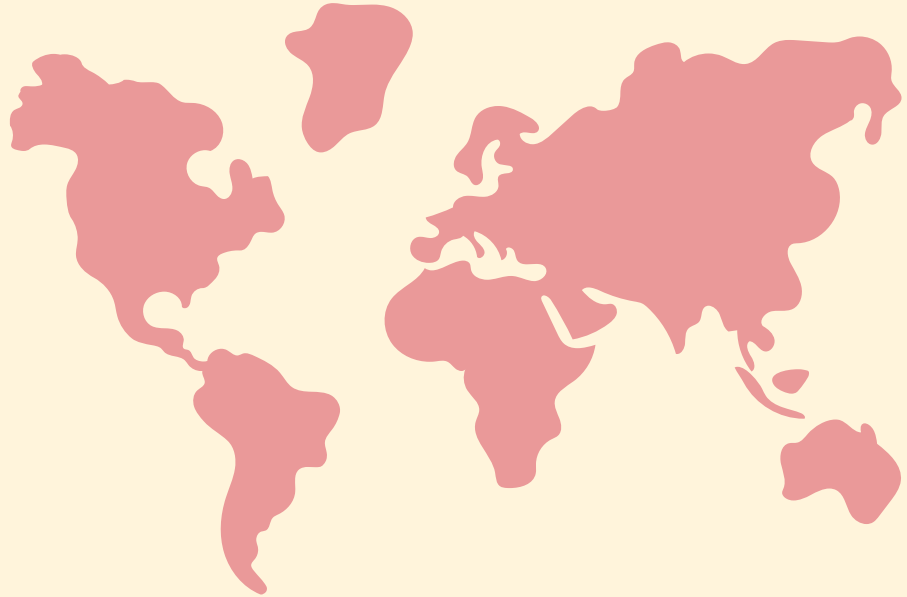
# Motivação

Visualizar possíveis correlações entre o número de casos de DSTs, aspectos socioeconômicos e políticas públicas, a fim de determinar fatores chave no controle de novos casos.



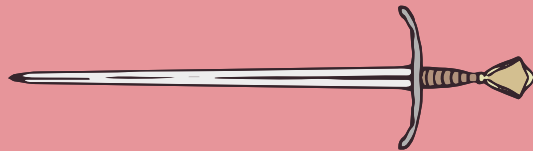
# Análise de DSTs e aspectos socioeconômicos

- Foco em casos de infecção por HIV
- Busca por correlação entre quantidade de novas infecções, aspectos socioeconômicos no país (e.g. grau de escolaridade, renda) e políticas públicas



02

## Bases de Dados



# WHO GHO



## Modelo Lógico

Hierárquico



## Formato

JSON



## Obtenção

API

(Athena API)



## Local

Mundo



## Dados

- Casos
- Aspectos socioeconômicos



## Relevância

Alta



# UNAIDS Laws and Policies



## **Modelo Lógico**

Tabular



## **Formato**

CSV



## **Obtenção**

Exportação pelo site



## **Local**

Mundo



## **Dados**

Políticas e Legislações  
relacionadas ao HIV



## **Relevância**

Média





# DBPedia



## **Modelo Lógico**

Grafo



## **Formato**

JSON



## **Obtenção**

API



## **Local**

Mundo



## **Dados**

Aspectos socioeconômicos



## **Relevância**

Alta



# Países Vizinhos



**Modelo Lógico**

Tabular



**Formato**

CSV



**Obtenção**

Exportação pelo site



**Local**

Mundo



**Dados**

Países vizinhos



**Relevância**

Alta

# UK Gov Statistics (descartada)



## Modelo Lógico

Tabular



## Formato

ODS



## Obtenção

Exportação pelo site



## Local

Reino Unido



## Dados

DSTs no Reino Unido



## Relevância

Baixa

# AtlasPlus CDC (descartada)



## Modelo Lógico

Tabular



## Formato

CSV



## Obtenção

Exportação pelo site



## Local

EUA



## Dados

Doenças nos EUA



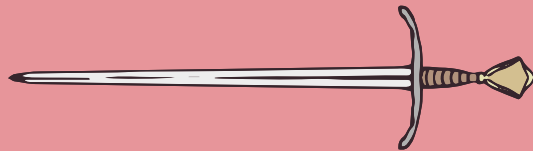
## Relevância

Baixa

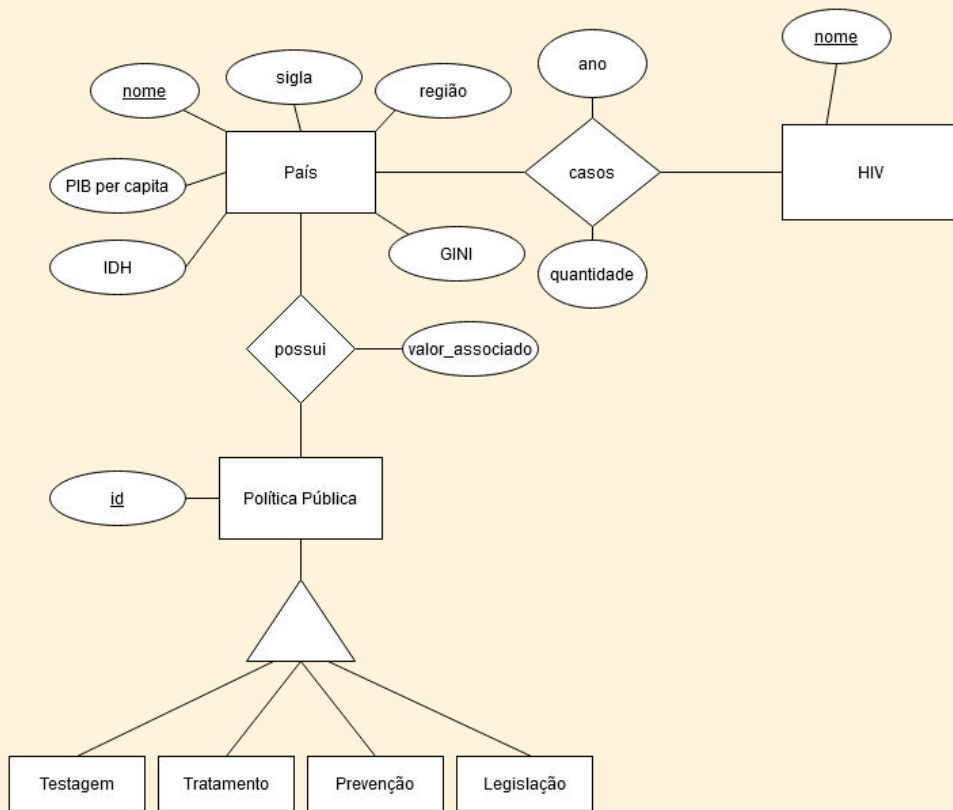


03

## Metodologia e Resultados



# Modelo Conceitual



# Modelo Lógico Relacional

País							
nome	sigla	região	wb_class	GINI (x 100)	IDH (x 1000)	PIB per Capita	População
Algeria	DZA	Africa	Upper middle income	353	736	14610.0	40400000

Política		
id	política	tipo
P1	Country adopted 2015 WHO guidelines ...	law

Infecções por HIV				
país (ou GLOBAL)	ano	mínimo	nominal	máximo
ARG	2000	5200.00	6200.00	7400.00

PolíticaPaís				
país	política	éNumérico	valor	valorNumérico
Afghanistan	P1	False	Yes, partially	

Testagem				
país	Política AT	Política BT	...	Política ATZ
Afghanistan	Yes	Yes		No

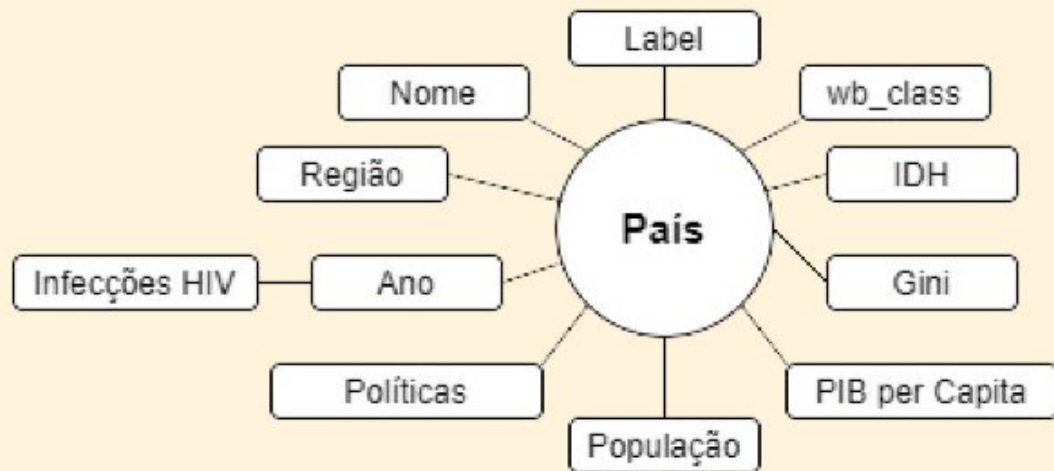
Tratamento				
país	Política ATr	Política BTr	...	Política ATrZ
Afghanistan	Yes	Yes		Countrywide

Prevenção				
país	Política AP	Política BP	...	Política APZ
Afghanistan	Yes	Yes		Yes

Legislação				
país	Política AL	Política BL	...	Política ALZ
Afghanistan	Yes, patially	No		Countrywide

# Modelo Lógico de Grafos

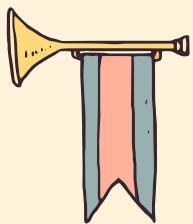
## Nó com atributos



## Políticas utilizadas para criar arestas conectando Países (nós)

- Índices socioeconômicos com valores próximos;
- Semelhança entre políticas adotadas;
- Proximidade geográfica (países vizinhos).





# Tratamento dos Dados - SQL

Utilizou-se Jupyter Notebooks em Python para confeccionar as tabelas utilizadas no modelo lógico tabular.

Os dados das fontes de dados utilizadas são tratados e então inseridos em um DataFrame, que após completado o processo é exportado como CSV.

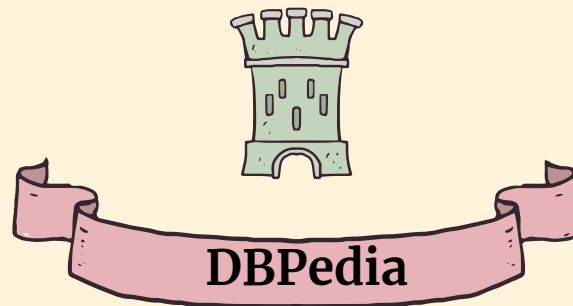


# Tabela de Países



Lista de Países (170)

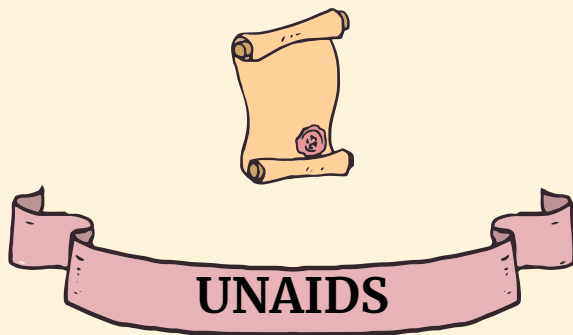
- Nome
- Label
- Região
- Classificação World Bank



Utilizando a lista de países,  
adicionamos os seguintes campos:

- Gini (124)
- PIB per Capita PPP (138)
- IDH (137)
- População (116)

# Tabelas de Políticas



Divisão das Políticas em 4 tipos:

- Testagem
- Tratamento
- Legislação
- Prevenção



Criação de tabelas que facilitam queries:

- Políticas (id, política, tipo)
- País-Políticas (país, políticaID, éNumérico, valor, valorNumérico)

# Tabela de Infecções



HIV\_0000000026 GHO

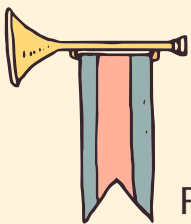
- Região/País
- Ano
- Nominal
- Mínimo
- Máximo

# Tabelas de Casos de DSTs no Reino Unido (descartada)



- DST
- Ano
- Sexo
- Casos reportados por não especialistas
- Casos reportados por especialistas
- Total

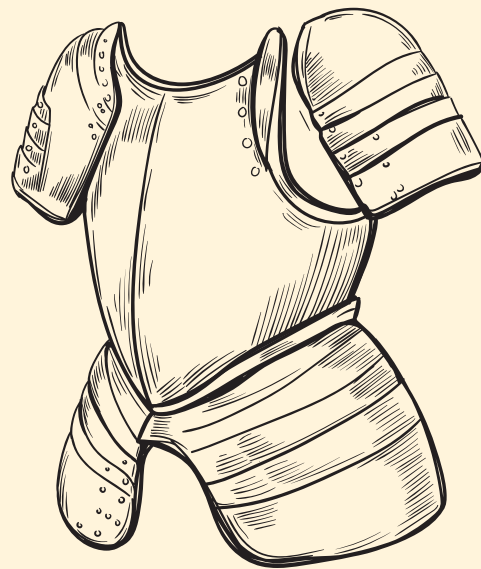




# Queries - SQL

Foram realizadas 7 queries buscando as seguintes análises:

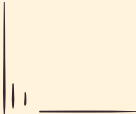
- Qtd de Políticas x Renda QR
- Região x Novas Infecções QR
- Socioeconômico x Novas Infecções QRR2
- Tratamento x Renda QR
- Distribuição Preservativos x Região QR
- Preservativos por Habitante QR
- Estratégia x Região QR



# Quantidade de Políticas x Renda

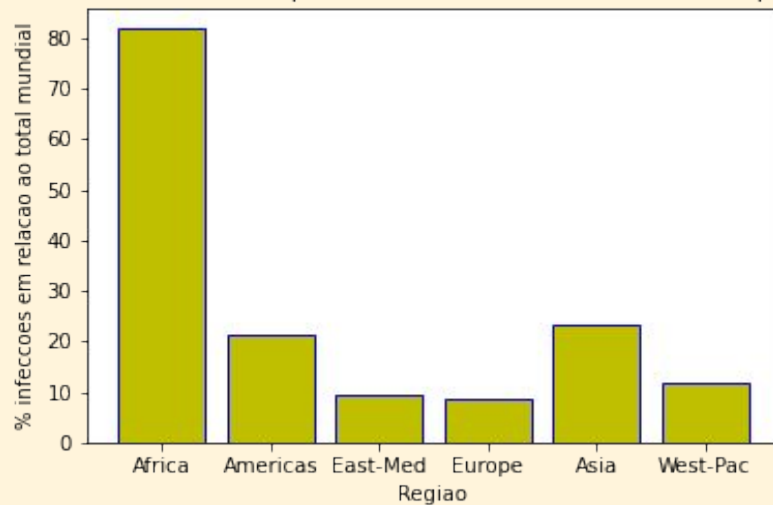


RENDA	Media_Políticas
Upper middle income	116.622222222222222222222222222222
Low income	135.0294117647058823529411765
Lower middle income	133.325
High income	64.2



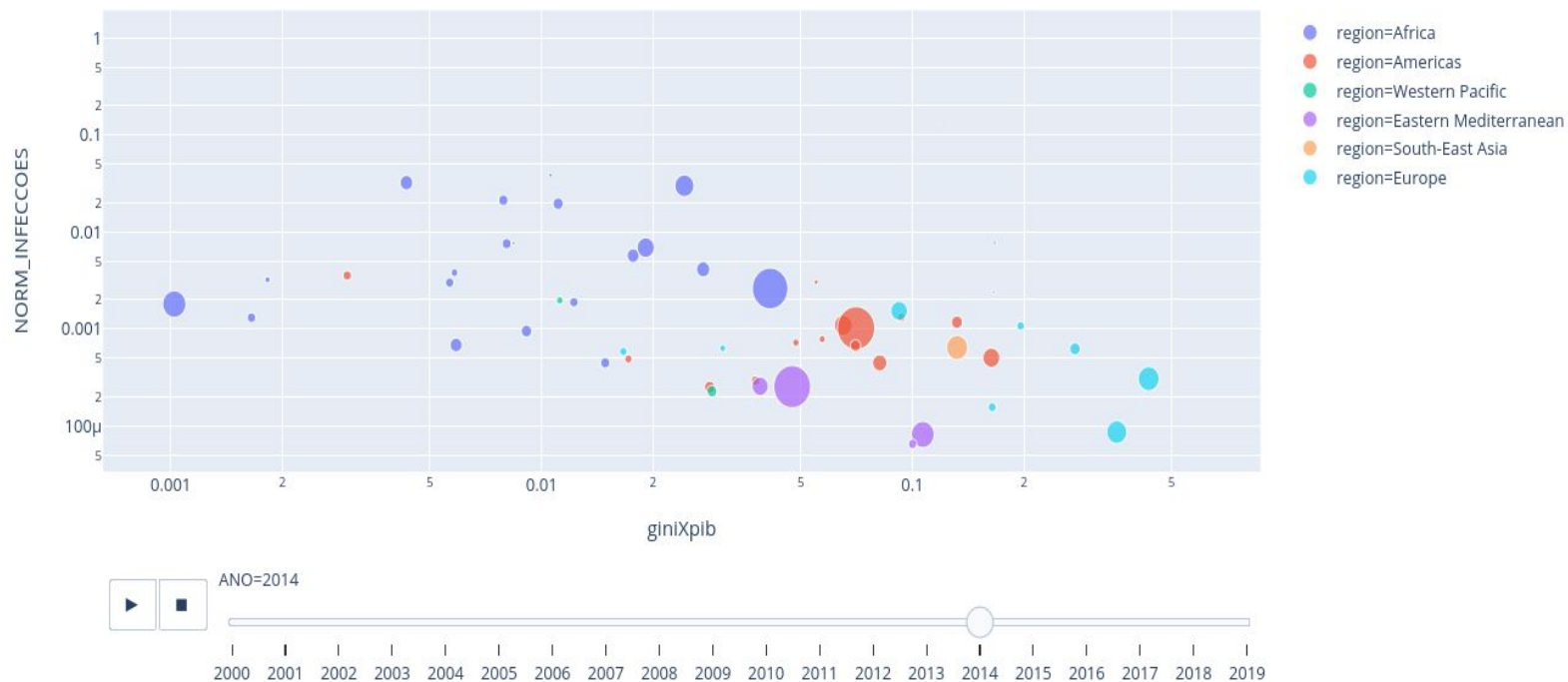
# Região x Novas Infecções

Quantidade de infeccoes por HIV em relacao ao valor mundial por regia

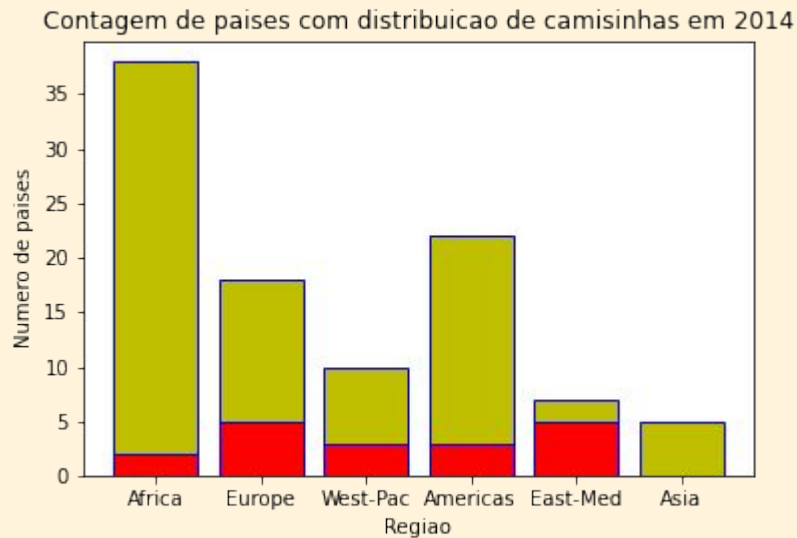




# Dado socioeconômico x Novas Infecções



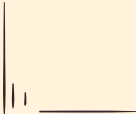
# Distribuição de Preservativos x Região



# Preservativos por Habitante

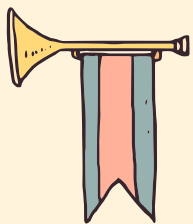


COUNTRY	Número de camisinhas masculinas por habitante	Número de camisinhas femininas por habitante	Número de camisinhas por habitante	National condom strategy/plan	Prosecution/pu
Algeria	0.07944237623762376	6.93069306930693E-4	0.08013544554455446	Yes	No
Argentina	0.4766722712301633			Yes	No
Armenia	0.06179150270126059			Yes	No
Azerbaijan	0.7199128021708221	0.0	0.7199128021708221	Yes	No
Bahamas				Yes	No
Bangladesh	0.13613514432262339			Yes	No
Bolivia	1.31456128138526			Yes	No
Botswana				Yes	No
Brazil	1.6164436447534487	0.007750404050361157	1.6241940488038098	Yes	No
Burkina Faso	1.1979095060635707	0.03710486459576156	1.2350143706593324	Yes	No
Burundi	1.1768060620519636	0.0233620042578349	1.2001680663097987	Yes	No



# Política nacional x Região

REGIAO	National strategy/policy guiding AIDS response	QUANTIDADE	Porcentagem de países
Africa	Yes, a stand-alone AIDS strategy or policy	34	75.55555555555556
Africa	No	1	2.222222222222223
Africa	Yes, a health strategy or policy that integrates the AIDS response	6	13.333333333333334
Europe	Yes, a stand-alone AIDS strategy or policy	18	36.0
Europe	No	1	2.0
Europe	Yes, a health strategy or policy that integrates the AIDS response	6	12.0
Western Pacific	Yes, a stand-alone AIDS strategy or policy	8	53.333333333333336
Western Pacific	No	2	13.333333333333334
Western Pacific	Yes, a health strategy or policy that integrates the AIDS response	3	20.0
Americas	Yes, a health strategy or policy that integrates the AIDS response	11	37.93103448275862
Americas	Yes, a stand-alone AIDS strategy or policy	15	51.724137931034484
Eastern Mediterranean	Yes, a stand-alone AIDS strategy or policy	12	60.0
Eastern Mediterranean	No	2	10.0
Eastern Mediterranean	Yes, a health strategy or policy that integrates the AIDS response	2	10.0
South-East Asia	Yes, a health strategy or policy that integrates the AIDS response	2	18.181818181818183
South-East Asia	Yes, a stand-alone AIDS strategy or policy	5	45.45454545454545

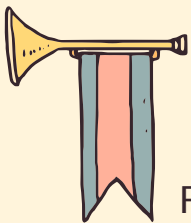


# Tratamento dos Dados – Grafos

Foram utilizadas as tabelas resultantes das queries em SQL, além do csv que relacionava países vizinhos.

Os arquivos csv foram utilizados na plataforma Neo4j, em que foram elaboradas queries para a criação de grafos, que foram visualizados no Cytoscape.

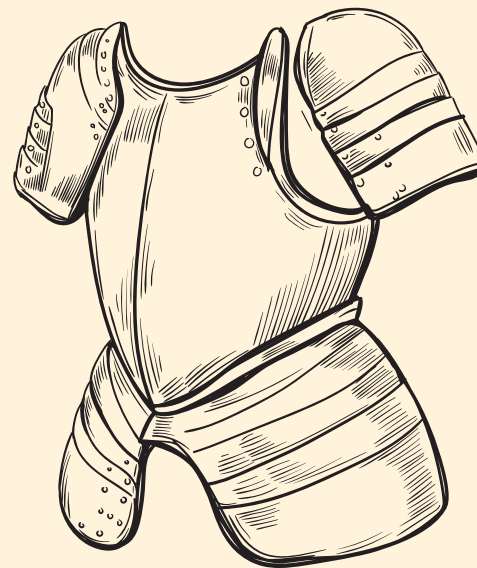




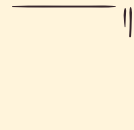
# Queries – Grafos

Foram realizadas 3 queries usando Cypher:

- Clustering socioeconomics
  - Query
  - Resultados: idh, gini
- Políticas e Países
  - Query
  - Resultados: high, low, total
- Regional Infections
  - Query
  - Resultados: África, América do Sul, Ásia



# Agrupando países com dados socioeconômicos

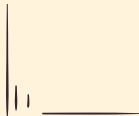


## Que combinação utilizar?

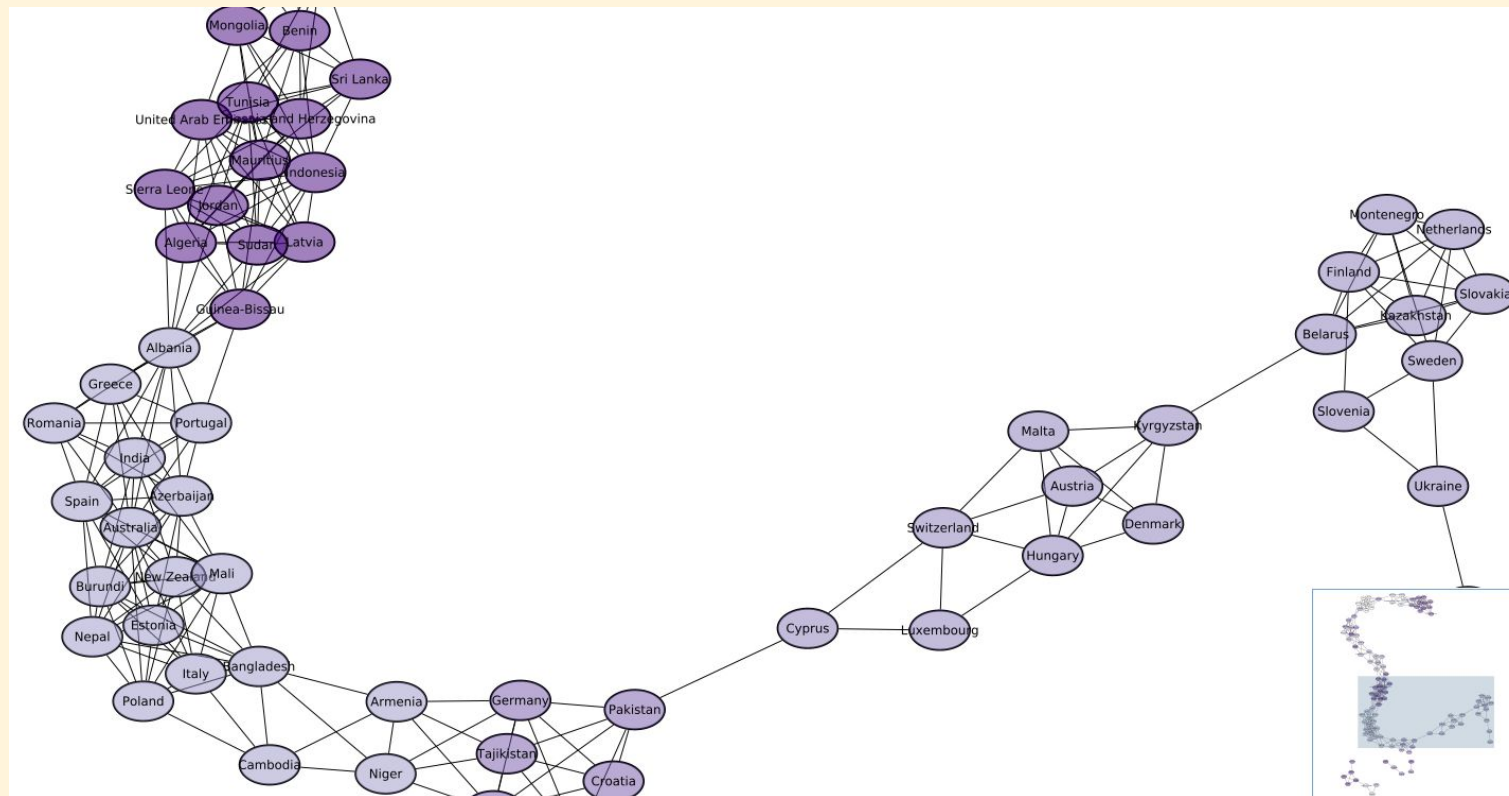
- PIB per Capita PPP
- IDH
- Gini
- IDH + Gini
- IDH + Gini + PIB per Capita

## Algoritmo:

- Se dois países tem um dado socioeconômico semelhante, cria uma aresta entre eles e quanto maior a proximidade do valor, maior o peso
- Aplicar Louvain com e sem peso
- Verificar manualmente se o agrupamento faz sentido

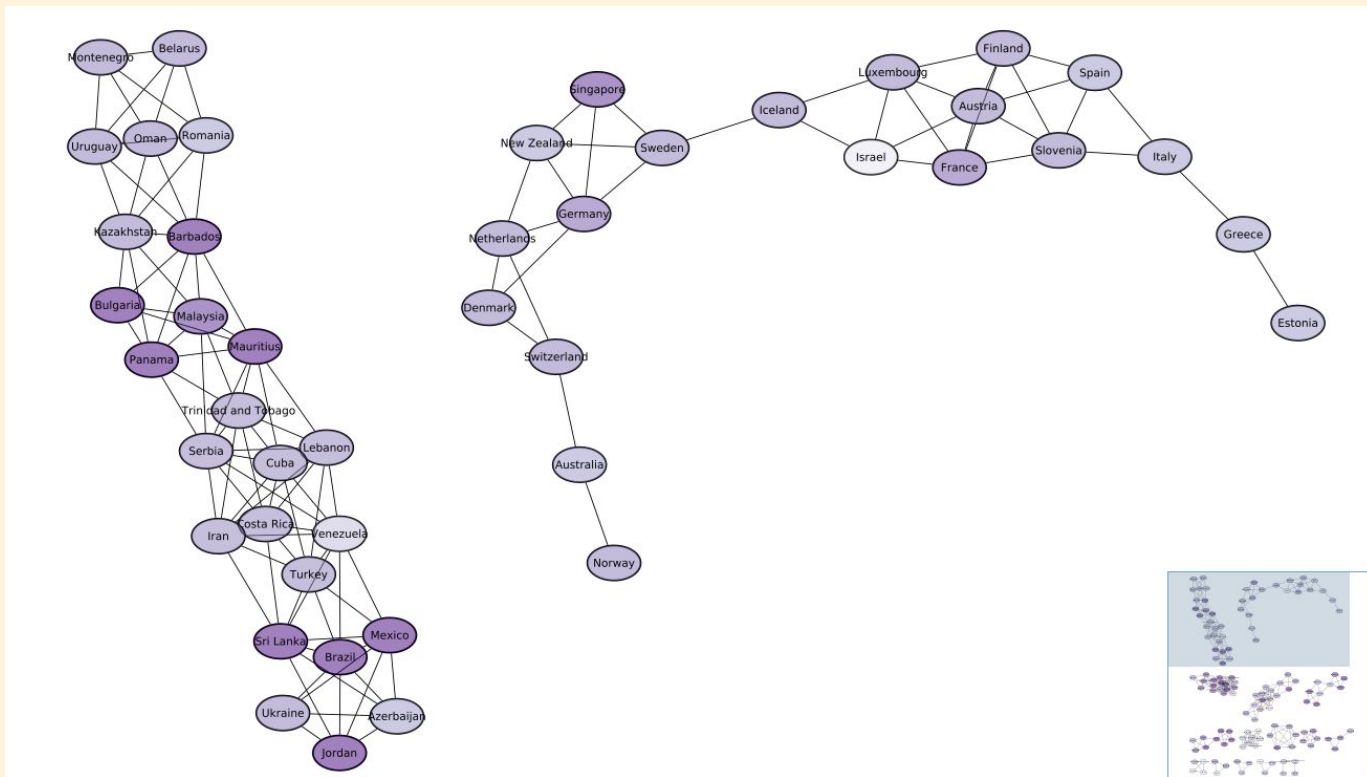


# Gini + IDH com Peso - grafo de Gini





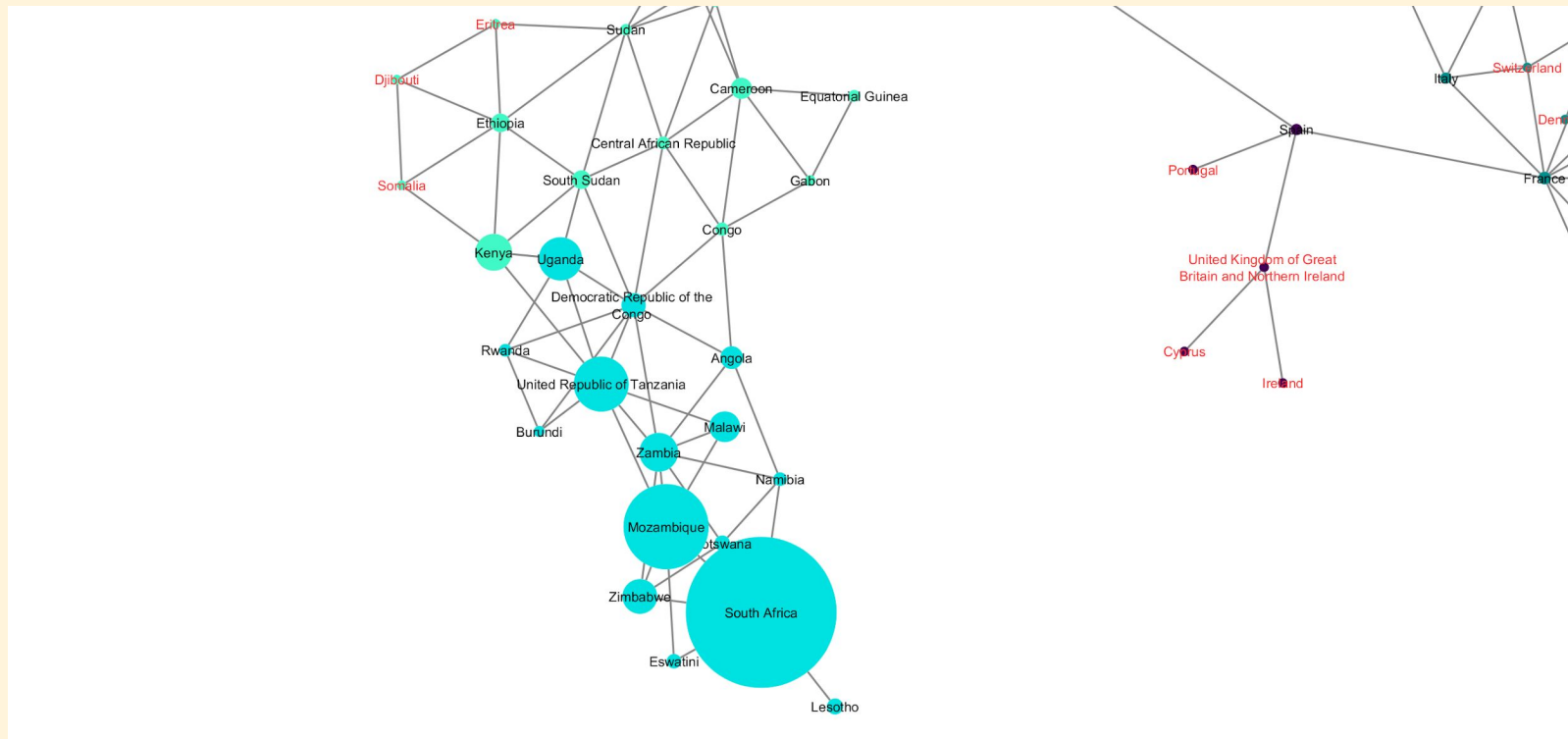
# Gini + IDH com Peso - grafo do IDH



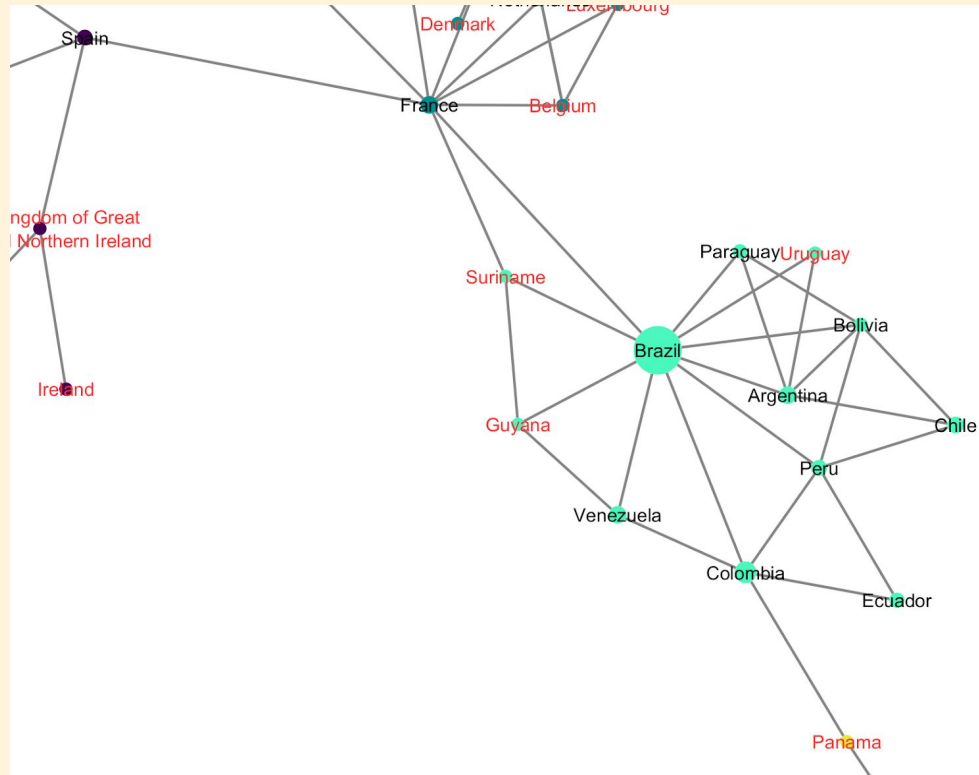
# Grupo x Infecções

ID	NUMBER_OF_COUNTRIES	IDH	GINI	QTDE_INFECCOES
155	6	470	422	25400.0
78	8	558	462	125860.0
52	7	597	451	278200.0
165	11	621	458	108180.0
113	5	626	387	14340.0
154	10	629	478	506400.0
103	10	642	353	23900.0
133	12	651	423	58600.0
117	6	790	302	25480.0

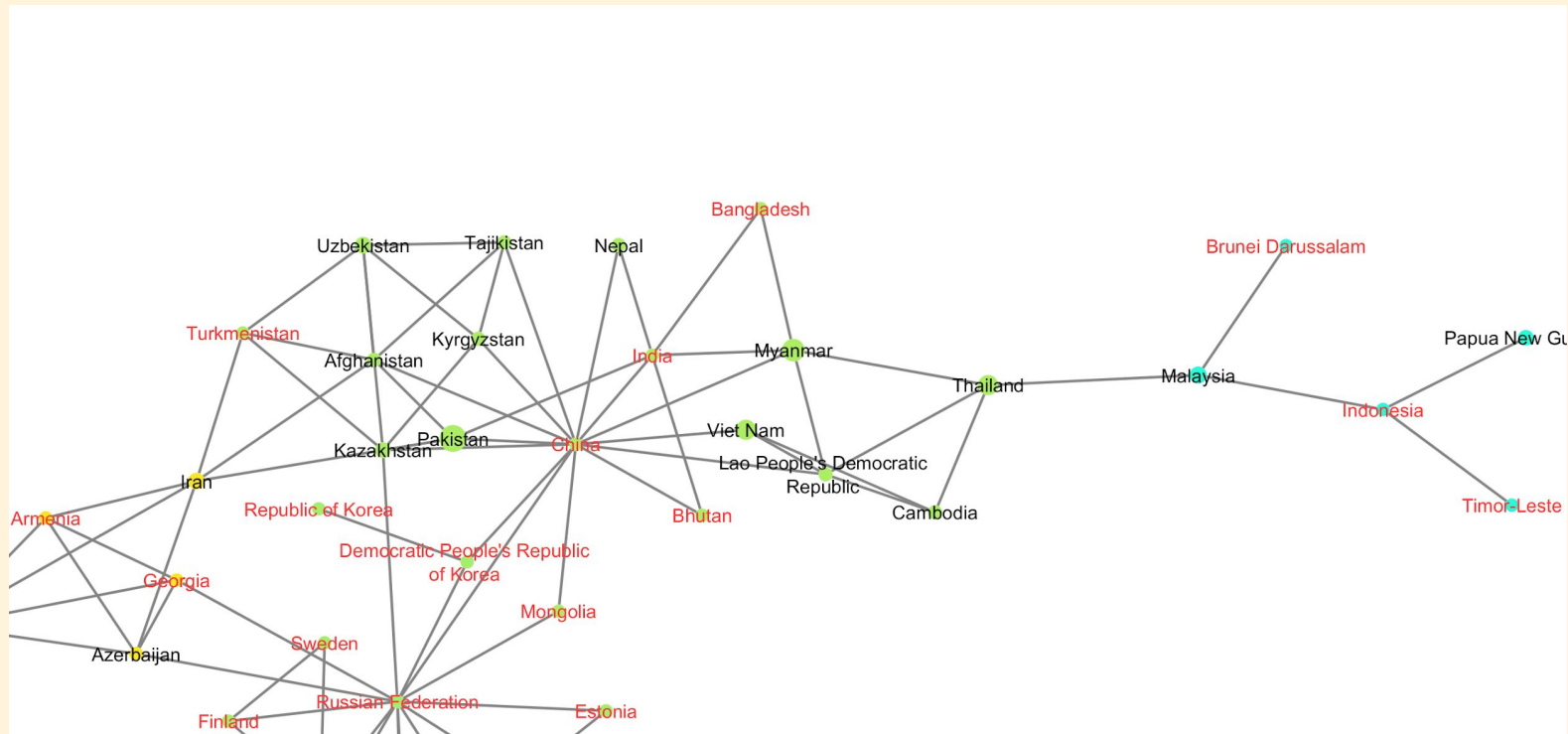
# Infecções regionais - África



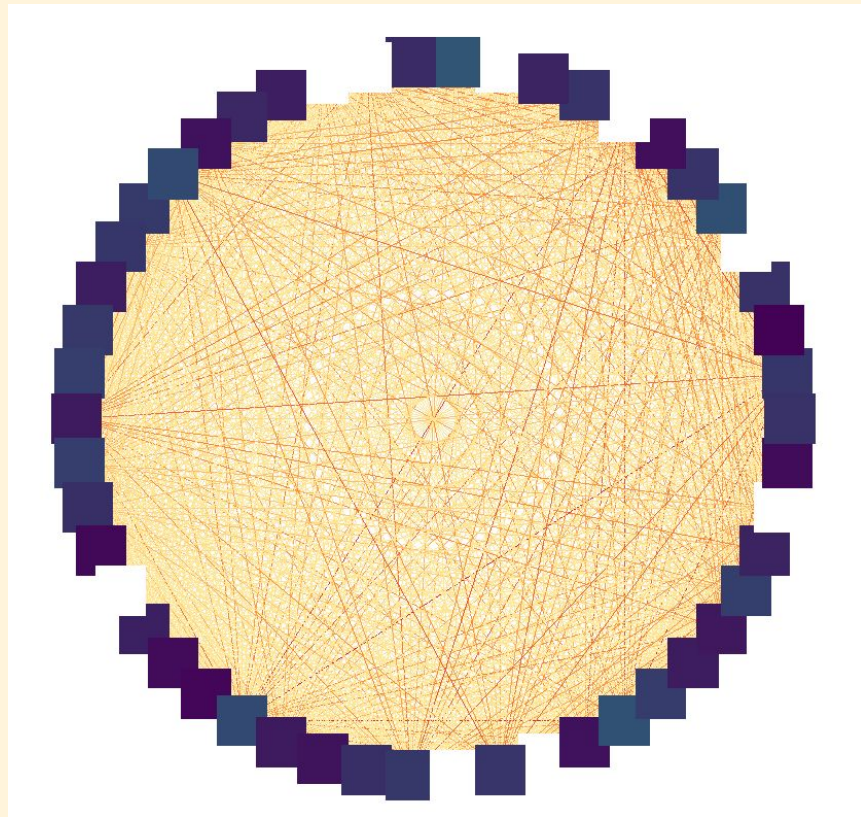
# Infecções regionais - América do Sul

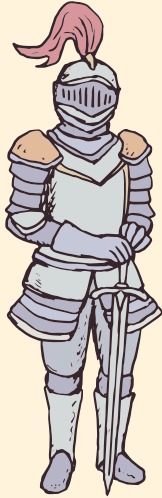


# Infecções regionais - Ásia



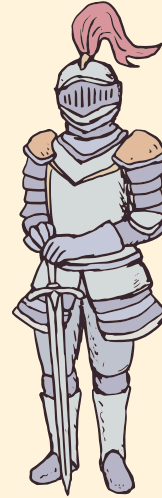
# Políticas compartilhadas - Alta renda





# Obrigado!

Willian Takayuki Ozako - 207267  
Frederico Meletti Rappa - 216047  
João Alberto Moreira Seródio - 218548



**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.

