

Trabalho Final

MC536

Cavaleiros da Query Estruturada

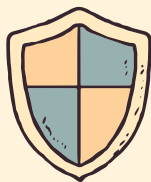
Willian Takayuki Ozako - 207267

Frederico Meletti Rappa - 216047

João Alberto Moreira Seródio - 218548



Sumário



Introdução

- Motivação
- Objetivos



Base de Dados

- Bases de Dados



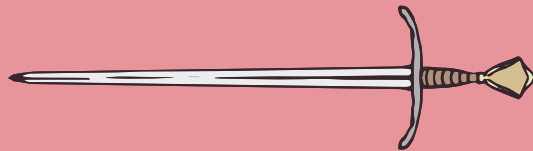
Metodologia e Resultados

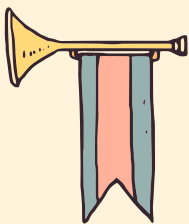
- Modelo Conceitual
- Modelos Lógicos
- Tratamento de Dados
- Queries (SQL e Cypher)
- Visualização



01

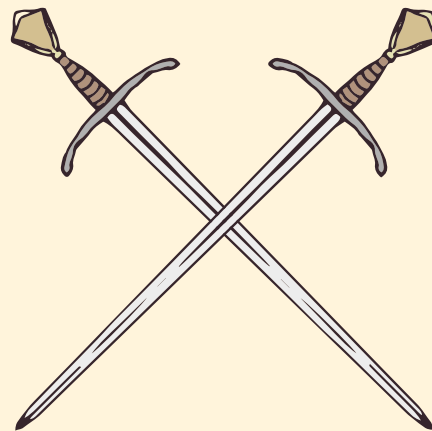
Introdução





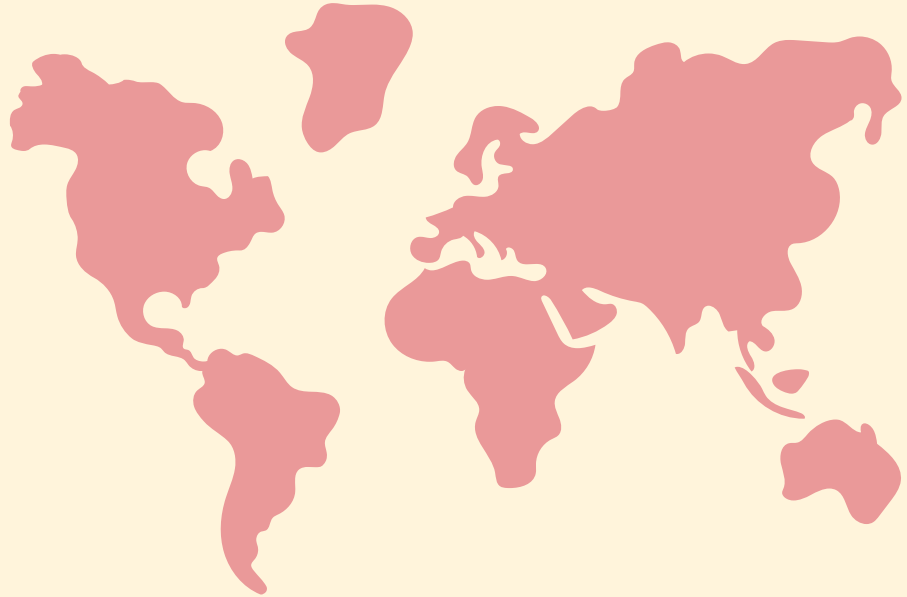
Motivação

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.



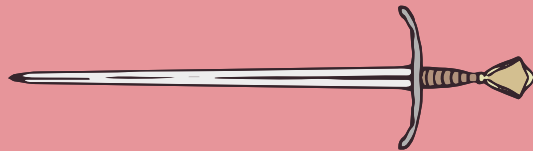
Análise de DSTs e aspectos socioeconômicos

- 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



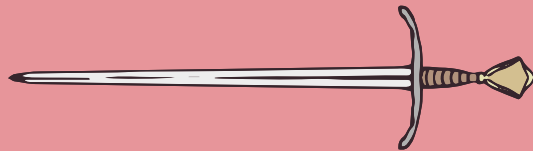
Relevância

Baixa

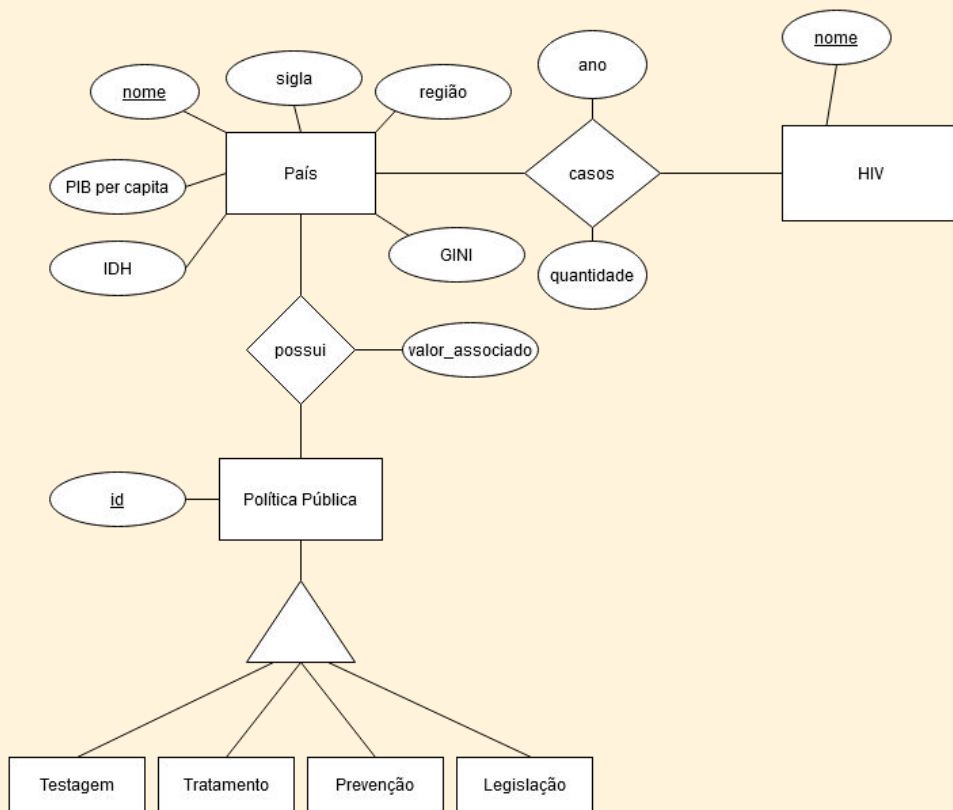


03

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

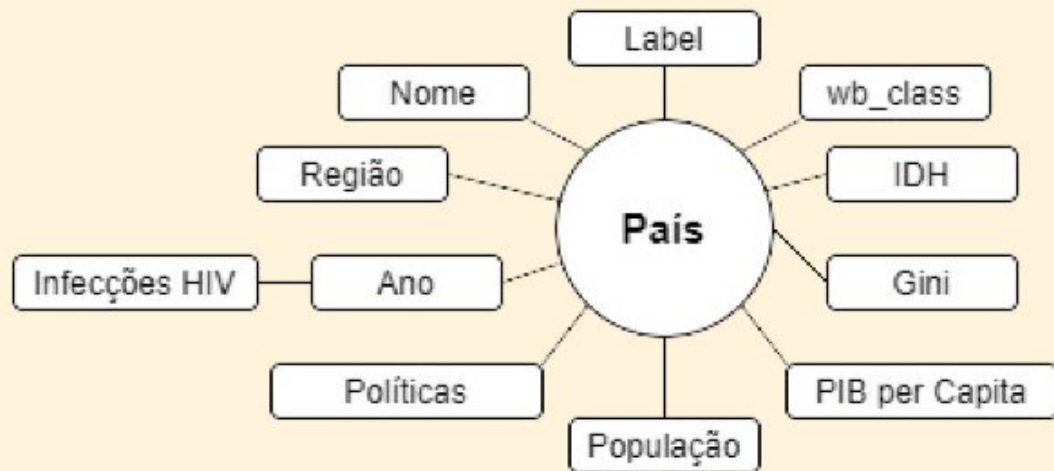
Tratamento				
país	Política ATr	Política BTr	...	Política ATrZ
Afghanistan	Yes	Yes		Countrywide

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

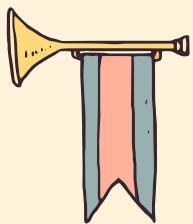
Nó com atributos



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.

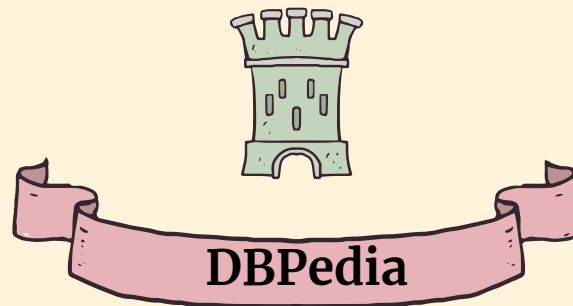


Tabela de Países



Lista de Países (170)

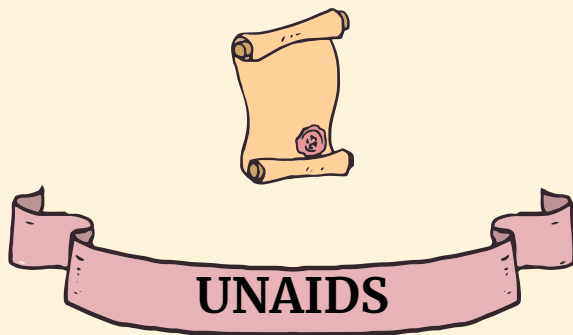
- Nome
- Label
- Região
- Classificação World Bank



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

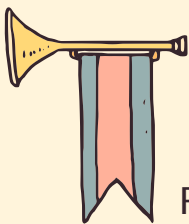
- Região/País
- Ano
- Nominal
- Mínimo
- Máximo

Tabelas de Casos de DSTs no Reino Unido (descartada)



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

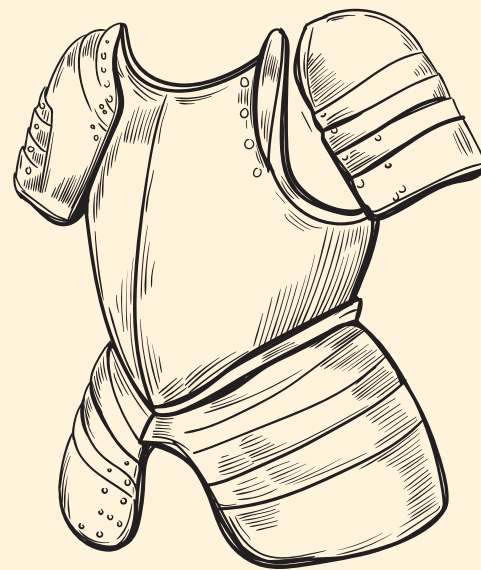




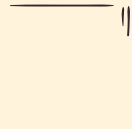
Queries - SQL

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

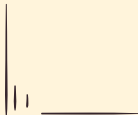


Quantidade de Políticas x Renda



```
DROP VIEW IF EXISTS PoliticasRenda;
DROP VIEW IF EXISTS PoliticasPaises;
CREATE VIEW PoliticasPaises AS
SELECT Pais.nome nome, Pais.classificacaoRenda renda, COUNT (*) politicas
  FROM Pais JOIN CountryPolicies CP ON Pais.nome = CP.country
 WHERE CP.isNumeric = False AND
        CP.value <> 'No'
 GROUP BY Pais.nome;

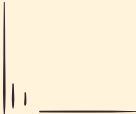
CREATE VIEW PoliticasRenda AS
SELECT P1.renda, SUM(P1.politicas)/(SELECT COUNT(*) FROM PoliticasPaises P2 WHERE P2.politicas <> 0 AND P2.renda = p1.renda) Media_Politicas
  FROM PoliticasPaises P1
 WHERE P1.politicas <> 0
 GROUP BY P1.renda;
CALL CSVWRITE('../..//saida/politicas-x-renda.csv', 'SELECT * FROM PoliticasRenda');
SELECT * FROM PoliticasRenda;
```



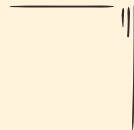
Quantidade de Políticas x Renda



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



```
CREATE VIEW RegiaoxInfeccoes AS
SELECT DP.ano, P.regiao,
       CAST(SUM(DP.quantidade) AS DOUBLE) / CAST(SUM(DG.quantidade) AS DOUBLE)*100.0 Infeccoes
FROM Pais P, DST DP, DST DG
WHERE P.sigla = DP.regiao
      AND DG.ano = DP.ano
      AND DP.quantidade IS NOT NULL
GROUP BY P.regiao, DP.ano
ORDER BY DP.ano,
         P.regiao;

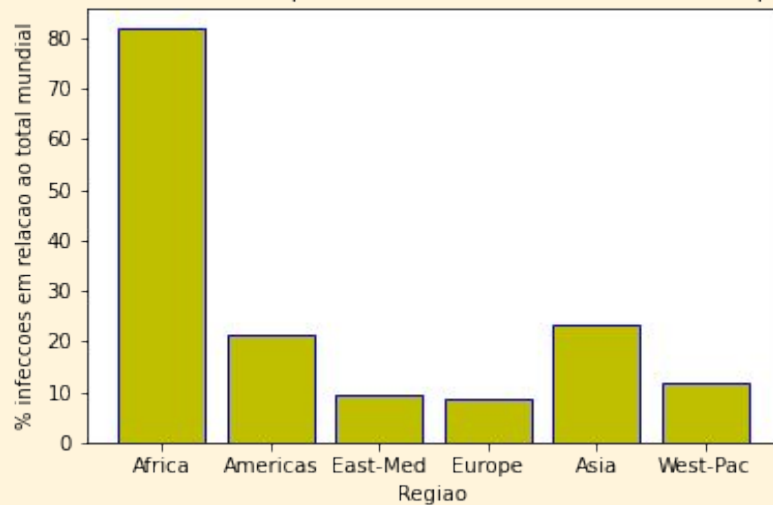
SELECT * FROM RegiaoxInfeccoes;

CALL CSVWRITE('../saida/regiao-x-infeccoes.csv', 'SELECT * FROM RegiaoxInfeccoes');
```



Região x Novas Infecções

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



Dado socioeconômico x Novas Infecções

```
CREATE VIEW tabelaInicial AS
SELECT D.ano, P.sigla,
       CAST(P.gini AS DOUBLE) / 1000.0 AS gini,
       P.PIBperCapita,
       CAST(D.quantidade AS DOUBLE) / CAST(P.populacao AS DOUBLE) taxaInfeccoes,
       CAST(P.IDH AS DOUBLE) / 1000.0 AS IDH
FROM Pais P, DST D
WHERE P.sigla = D.regiao AND P.gini IS NOT NULL
      AND P.PIBperCapita IS NOT NULL
      AND P.populacao IS NOT NULL
ORDER BY D.ano, gini, P.PIBperCapita
```

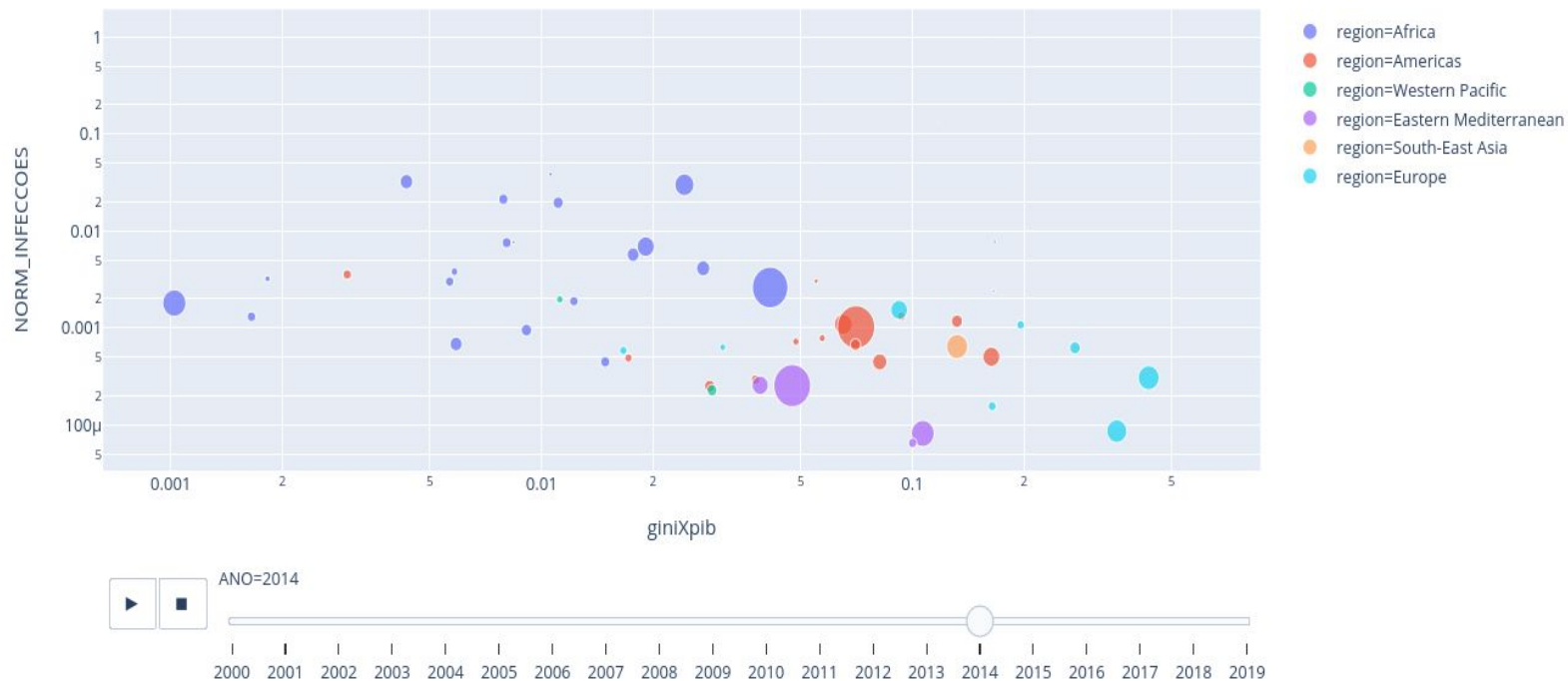
```
CREATE VIEW parNormInfecoes AS
SELECT ano,
       MIN(taxaInfecoes) minInf,
       MAX(taxaInfecoes) maxInf
FROM tabelaInicial
GROUP BY ano
ORDER BY ano
```

```
CREATE VIEW parNormPibGini AS
SELECT MIN(pibpercapita) minPib,
       MAX(pibpercapita) maxPib,
       MIN(1.0 - gini) minConvGini,
       MAX(1.0 - gini) maxConvGini,
       MIN(IDH) minIDH,
       MAX(IDH) maxIDH
FROM tabelaInicial
```

```
CREATE VIEW socioeconomicoXinfecoes AS
SELECT T.ano, T.sigla,
       (1.0 - T.gini - PNPG.minConvGini) / (PNPG.maxConvGini - PNPG.minConvGini) Corr_Gini,
       CAST(T.pibpercapita - PNPG.minpib AS DOUBLE) / CAST(PNPG.maxpib - PNPG.minpib AS DOUBLE) Norm_PIB,
       (T.taxaInfecoes - PNI.minInf) / (PNI.maxInf - PNI.minInf) Norm_infecoes,
       (T.IDH - PNPG.minIDH) / (PNPG.maxIDH - PNPG.minIDH) Norm_IDH
FROM tabelaInicial T, parNormInfecoes PNI, parNormPibGini PNPG
WHERE T.ano = PNI.ano
ORDER BY ano, corr_gini, norm_pib
```

```
SELECT * FROM socioeconomicoXinfecoes;
```

Dado socioeconômico x Novas Infecções



Queries de Políticas

```
DROP VIEW AgrupadoPorRegiao IF EXISTS;
DROP VIEW PoliticaNacionalAIDS IF EXISTS;
DROP VIEW Paisesss IF EXISTS;
DROP VIEW PaísesPorRegiao IF EXISTS;

CREATE VIEW Paisesss AS
SELECT Pr."National condom strategy/plan" AS "Tem uma política ou estratégia nacional de distribuição de camisinhas?", Pa.regiao, Pa.classificacaoRenda FROM Prevencao Pr, Pais Pa WHERE Pr.country = Pa.nome AND Pr."National condom strategy/plan" <> '';

CREATE VIEW PaísesPorRegiao AS
SELECT regiao, COUNT(*) "Quantidade de países na região"
FROM Pais
GROUP BY Regiao;

CREATE VIEW AgrupadoPorRegiao AS
SELECT regiao, "Tem uma política ou estratégia nacional de distribuição de camisinhas?", COUNT(*) Quantidade
FROM Paisesss
GROUP BY "Tem uma política ou estratégia nacional de distribuição de camisinhas?", regiao
ORDER BY regiao;

CREATE VIEW PoliticaNacionalAIDS AS
SELECT regiao, "National strategy/policy guiding AIDS response", COUNT(*) Quantidade
FROM (
    SELECT Pr."National strategy/policy guiding AIDS response" AS "National strategy/policy guiding AIDS response", Pa.regiao, Pa.classificacaoRenda
    FROM Legislacao Pr, Pais Pa
    WHERE Pr.country = Pa.nome AND Pr."National strategy/policy guiding AIDS response" <> ''
)
GROUP BY "National strategy/policy guiding AIDS response", regiao
ORDER BY regiao;

CREATE VIEW PossibleToStartART AS
SELECT classificacaoRenda, "Possible to start ART on the same day as HIV diagnosis", COUNT(*) Quantidade
FROM (
    SELECT Pr."Possible to start ART on the same day as HIV diagnosis"
    AS "Possible to start ART on the same day as HIV diagnosis", Pa.classificacaoRenda
    FROM Legislacao Pr, Pais Pa
    WHERE Pr.country = Pa.nome AND Pr."Possible to start ART on the same day as HIV diagnosis" <> ''
)
GROUP BY "Possible to start ART on the same day as HIV diagnosis", classificacaoRenda
ORDER BY classificacaoRenda;
```

Queries de Políticas

```
SELECT Pr.country,
       CAST(Pr."Number of male condoms distributed in the previous calendar year: total" AS FLOAT) / Pa.populacao AS "Número de
camisinhas masculinas por habitante",
       CAST(Pr."Number of female condoms distributed in the previous calendar year: total" AS FLOAT) / Pa.populacao AS "Número d
e camisinhas femininas por habitante",
       CAST(Pr."Number of male condoms distributed in the previous calendar year: total" + Pr."Number of female condoms distribut
ed in the previous calendar year: total" AS FLOAT) / Pa.populacao AS "Número de camisinhas por habitante",
       Pr."National condom strategy/plan",
       Pr."Prosecution/punishment of individuals for carrying condoms"
FROM Prevencao Pr, Pais Pa WHERE Pr.country = Pa.nome AND (Pr."Number of male condoms distributed in the previous calendar yea
r: total" IS NOT NULL OR Pr."Number of female condoms distributed in the previous calendar year: total" IS NOT NULL);
--
--

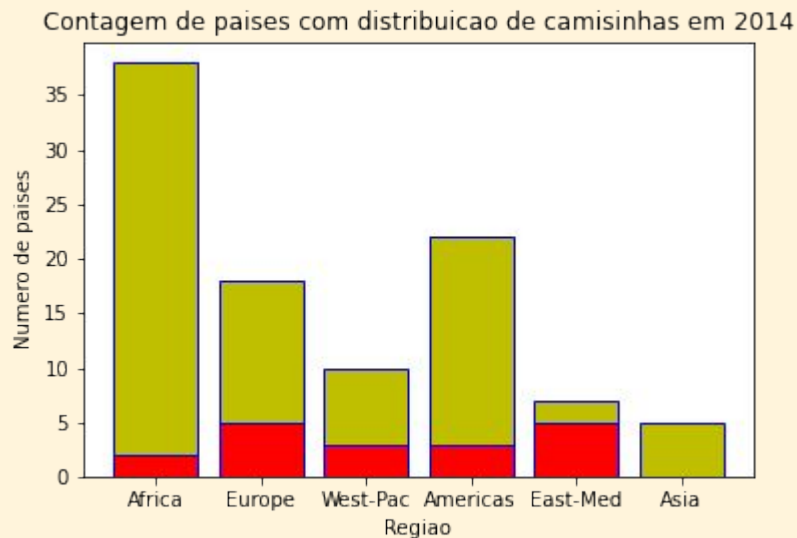
SELECT APR.regiao, APR."Tem uma política ou estratégia nacional de distribuição de camisinhas?", APR.Quantidade, CAST(APR.Quan
tidade*100 AS FLOAT) / PPR."Quantidade de países na região" "Porcentagem de países"
FROM AgrupadoPorRegiao APR, PaísesPorRegiao PPR
WHERE APR.regiao = PPR.regiao;
--
--

SELECT PNA.regiao, PNA."National strategy/policy guiding AIDS response", PNA.Quantidade, CAST(PNA.QUANTIDADE*100 AS FLOAT) / P
PR."Quantidade de países na região" "Porcentagem de países"
FROM PoliticaNacionalAIDS PNA, PaísesPorRegiao PPR
WHERE PPR.regiao = PNA.regiao;
--
--

SELECT ART.classificacaoRenda, ART."Possible to start ART on the same day as HIV diagnosis", ART.QUANTIDADE "Número de paíse
s", CAST(ART.QUANTIDADE*100 AS FLOAT) / CR.QUANTIDADE "Porcentagem de países"
FROM PossibleToStartART ART, (SELECT classificacaoRenda, COUNT(*) quantidade FROM Pais GROUP BY classificacaoRenda) CR
WHERE ART.classificacaoRenda = CR.classificacaoRenda;
```



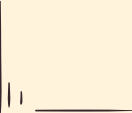
Distribuição de Preservativos x Região



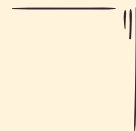
Possibilidade de tratamento x Classificação



CLASSIFICACAORENDA	Possible to start ART on the same day as HIV diagnosis	Número de países	Porcentagem de países
Lower middle income	No	11	27.5
Lower middle income	Yes	23	57.5
Upper middle income	Yes	23	51.111111111111114
Upper middle income	No	8	17.777777777777778
High income	Yes	10	19.607843137254903
High income	No	5	9.803921568627452
Low income	No	2	5.882352941176471
Low income	Yes	27	79.41176470588235



Preservativos por Habitante

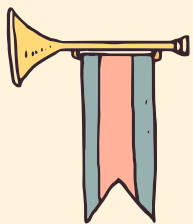


COUNTRY	Número de camisinhas masculinas por habitante	Número de camisinhas femininas por habitante	Número de camisinhas por habitante	National condom strategy/plan	Prosecution/pu
Algeria	0.07944237623762376	6.93069306930693E-4	0.08013544554455446	Yes	No
Argentina	0.4766722712301633			Yes	No
Armenia	0.06179150270126059			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	75.5555555555556
Africa	No	1	2.22222222222223
Africa	Yes, a health strategy or policy that integrates the AIDS response	6	13.3333333333334
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	53.3333333333333
Western Pacific	No	2	13.3333333333334
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	18.1818181818183
South-East Asia	Yes, a stand-alone AIDS strategy or policy	5	45.4545454545455

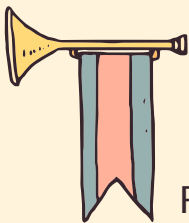


Tratamento dos Dados – Grafos

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.

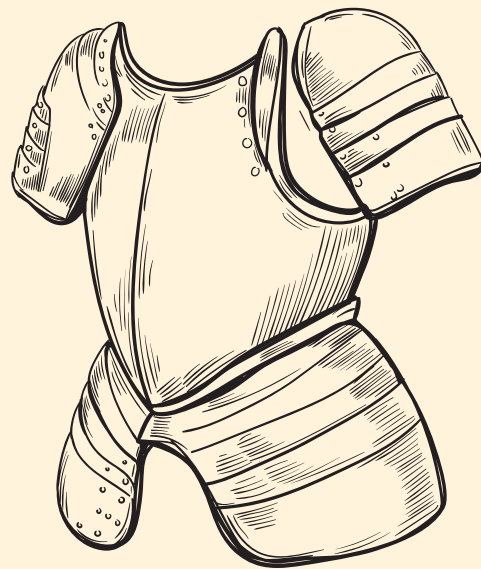




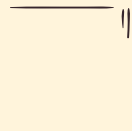
Queries – Grafos

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

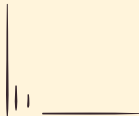


Que combinação utilizar?

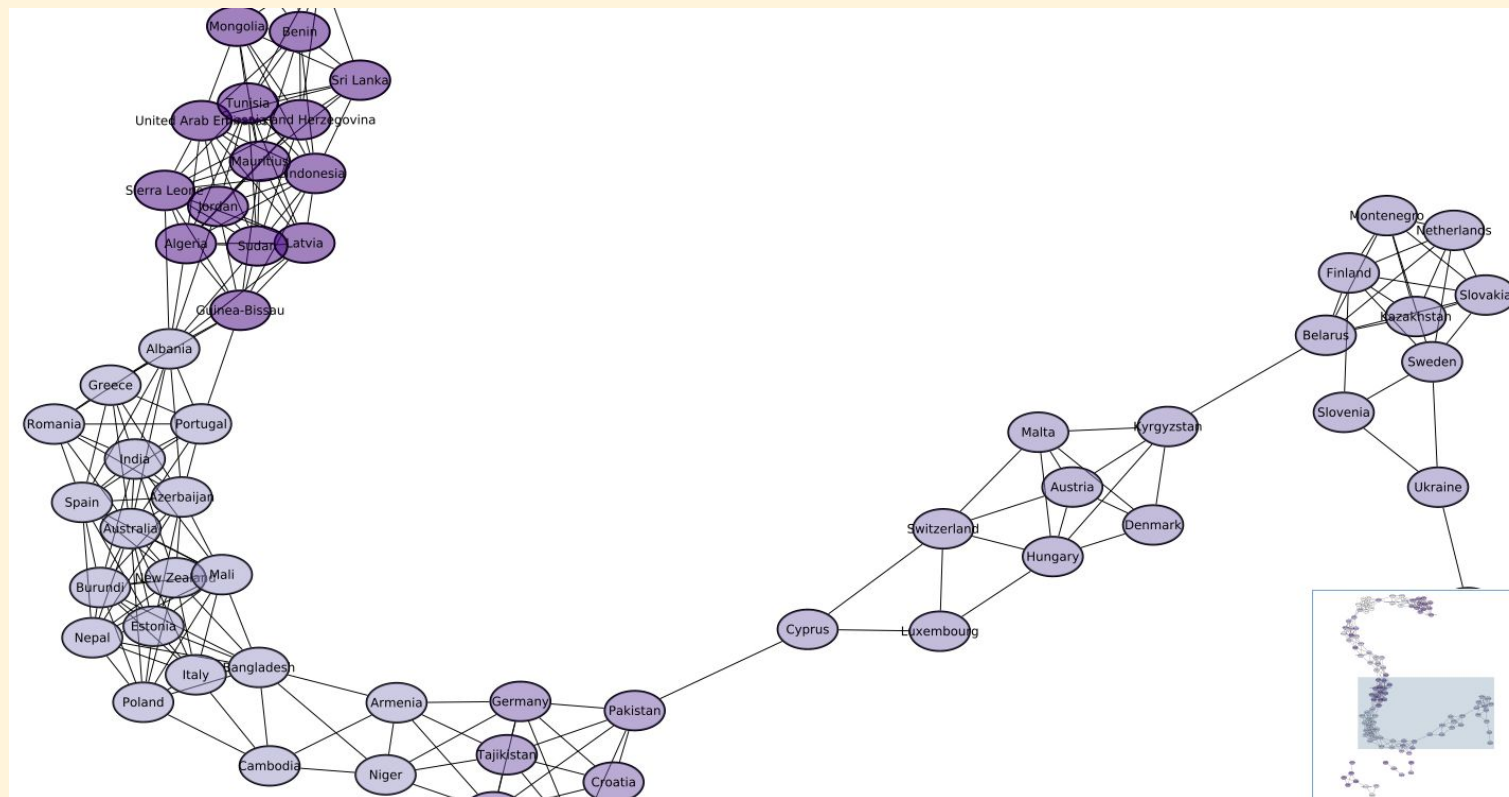
- PIB per Capita PPP
- IDH
- Gini
- IDH + Gini
- IDH + Gini + PIB per Capita

Algoritmo:

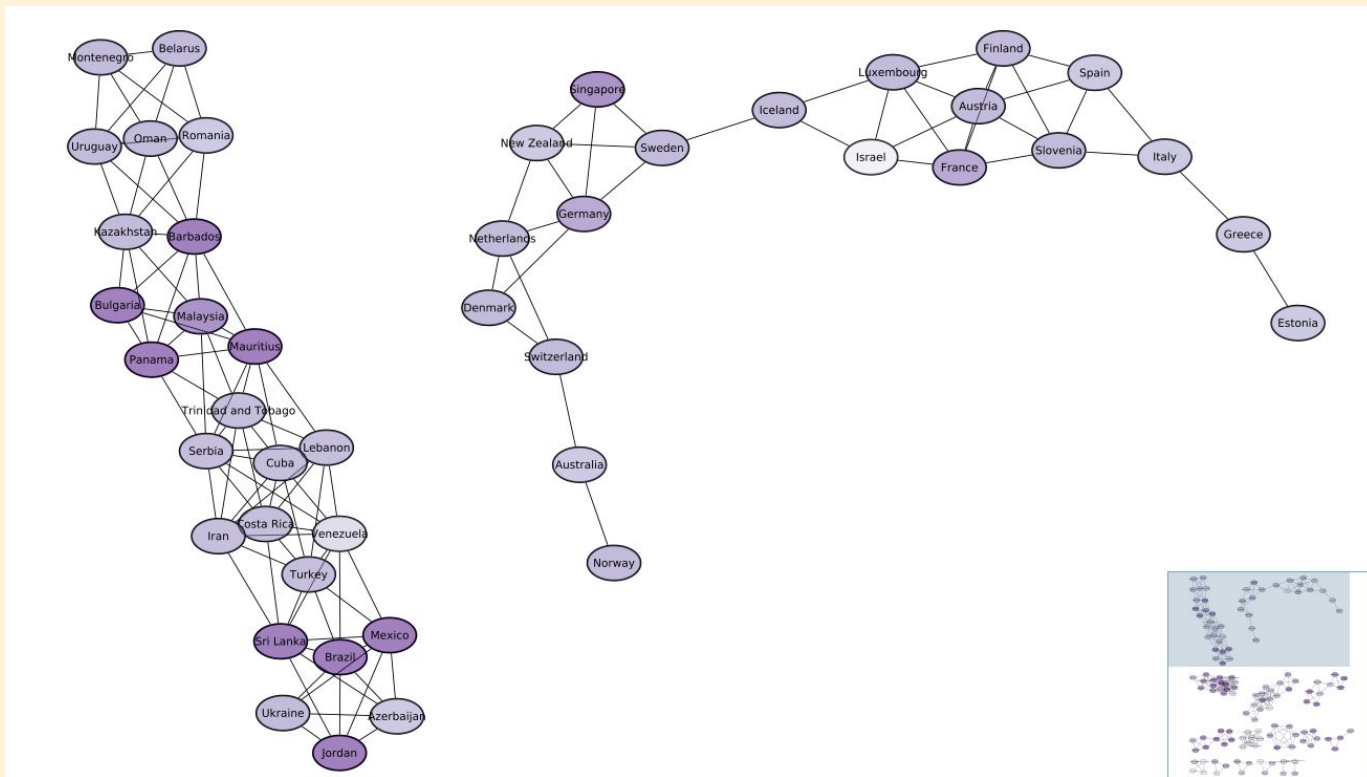
- Se dois países tem um dado socioeconômico semelhante, cria uma aresta entre eles e quanto maior a proximidade do valor, maior o peso
- Aplicar Louvain com e sem peso
- Verificar manualmente se o agrupamento faz sentido



Gini + IDH com Peso - grafo de Gini



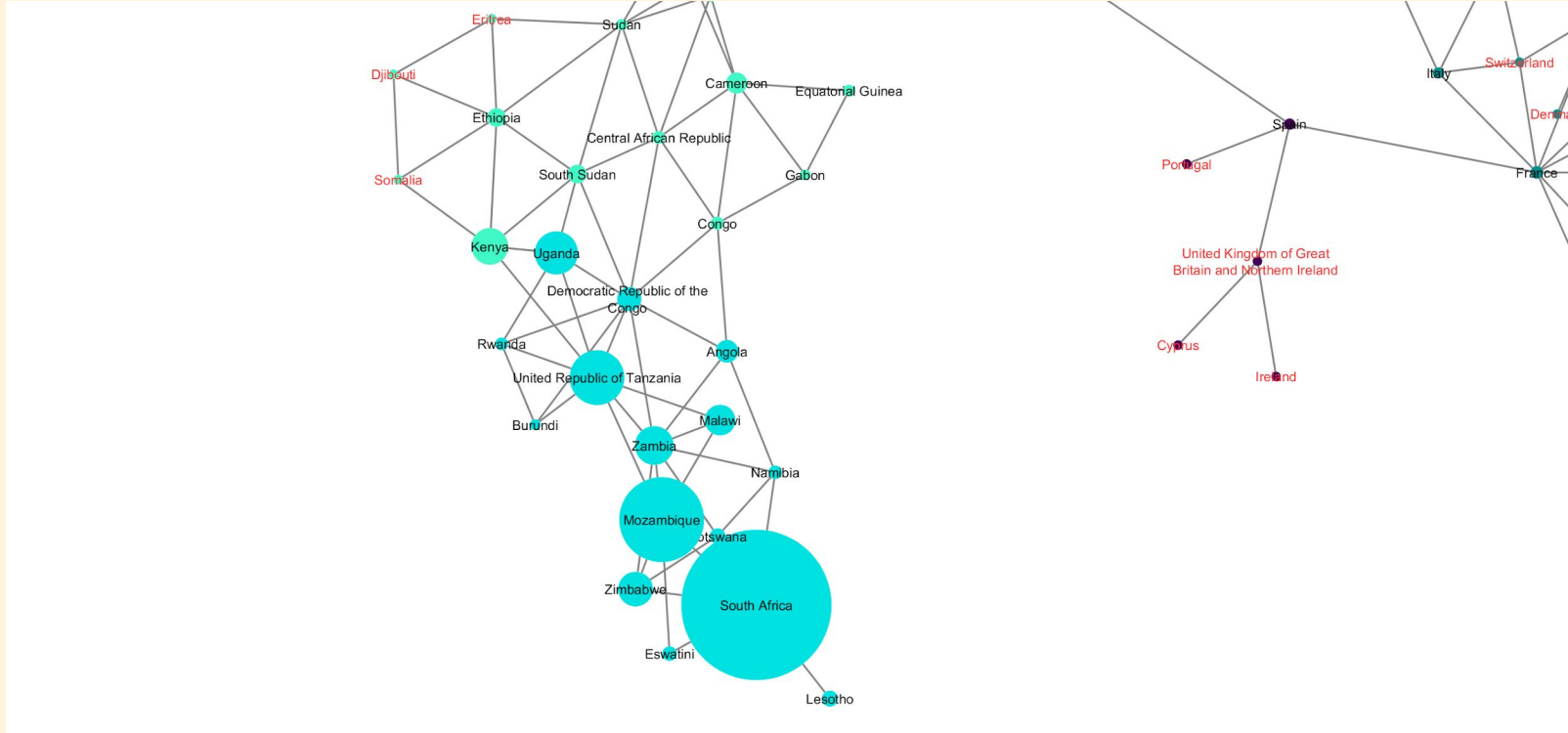
Gini + IDH com Peso - grafo do IDH



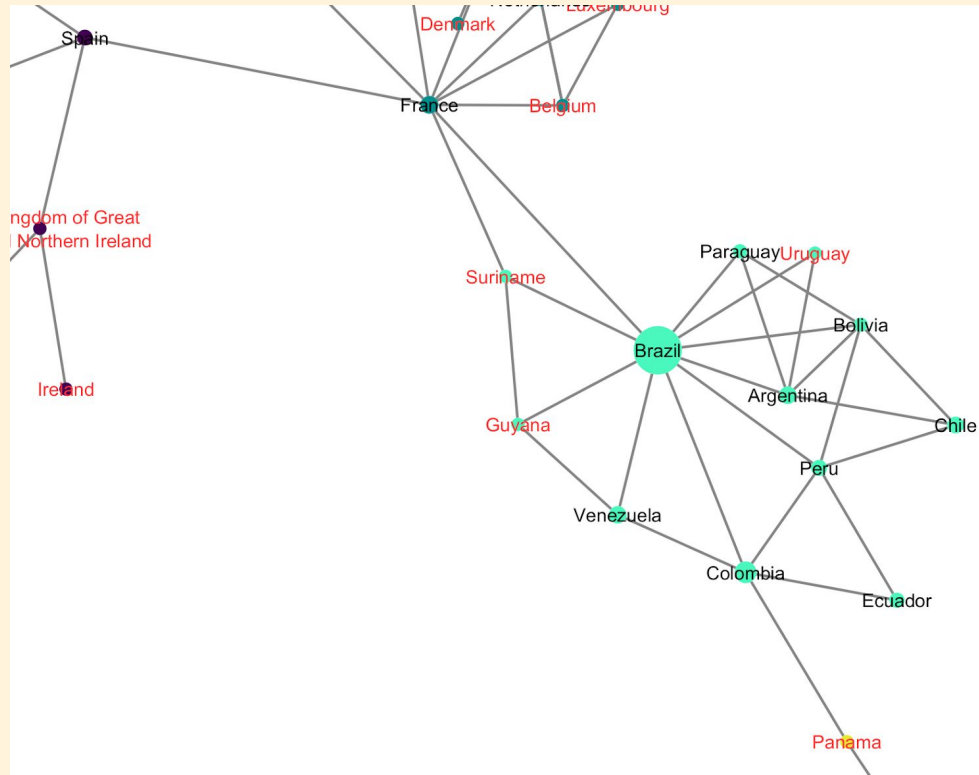
Grupo x Infecções

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

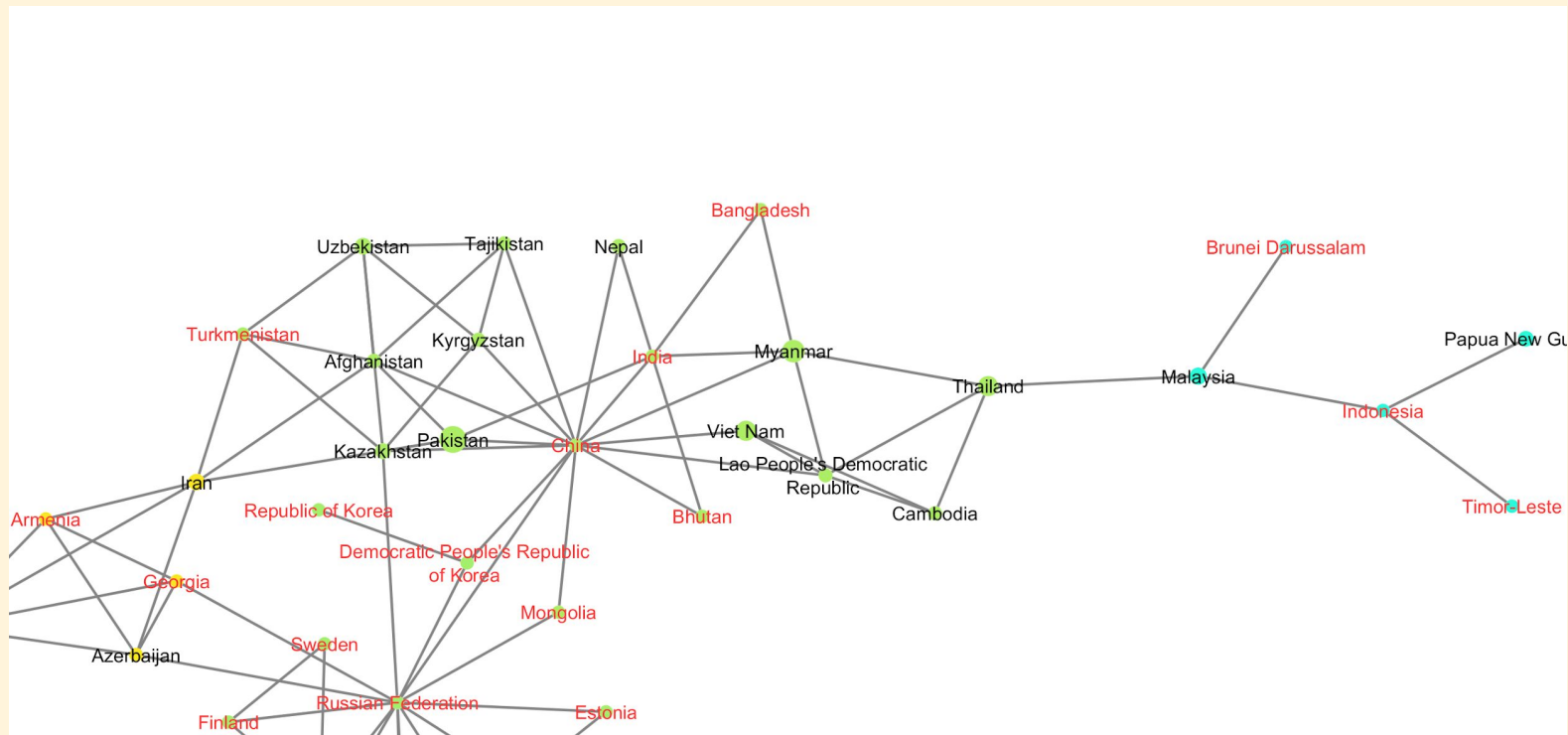
Infecções regionais - África



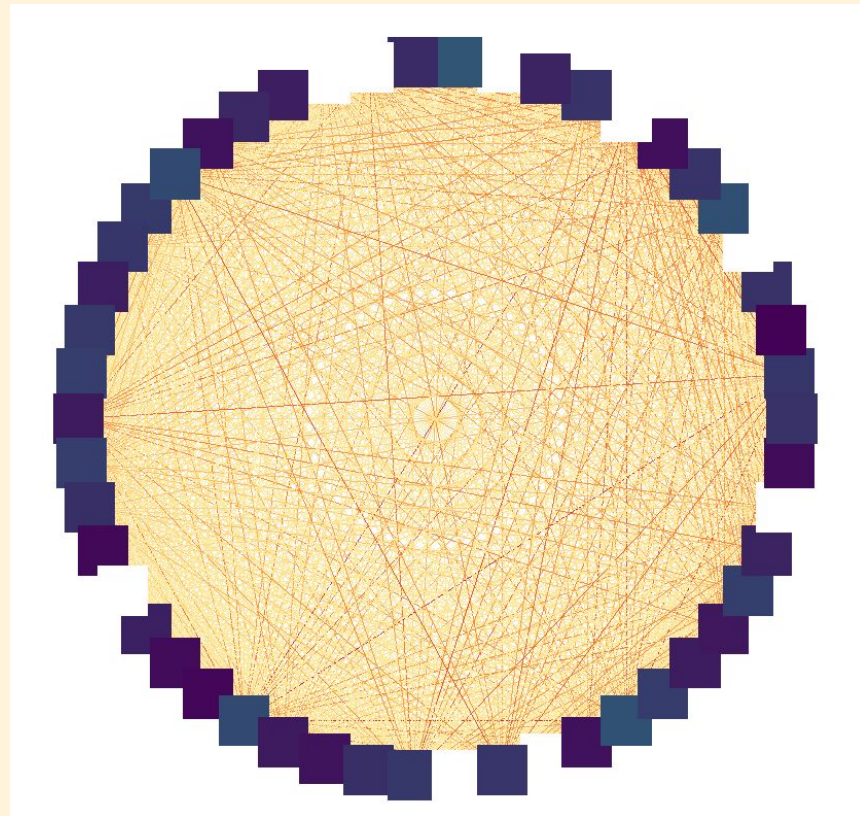
Infecções regionais - América do Sul

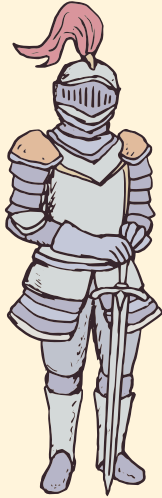


Infecções regionais - Ásia



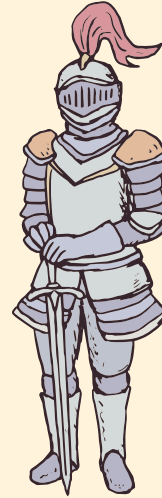
Políticas compartilhadas - Alta renda





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.

