

# TRABALHO PRÁTICO

## FUNDAMENTOS DE BANCO DE DADOS 2020/1

---

Sergio Mergen

# Introdução

- Com o surgimento da pandemia de Covid-19, muitas perguntas surgiram, como as seguintes
  - Quais as características (geográficas, culturais, ...) de uma região (país, cidade) que favorecem a contaminação?
  - Quais estratégias de combate apresentam melhor resultado?
  - Que pessoas estão mais suscetíveis?
  - Como os números atuais se comparam com demais causas de morte?
  - Como o isolamento vem afetando a economia?

# Introdução

- Para entender essas questões, pode-se colher informações e extrair estatísticas a partir delas
- Exemplos de informações
  - Local (Cidade, estado, país, ...)
  - Setor econômico
  - Número de curados, infectados, mortos
  - Densidade populacional
  - Causa da morte
  - Data (dia, mês, ano)
  - Fronteiras geográficas
  - Clima
  - Ações na bolsa
  - Características pessoais
  - ...

# Exemplo 1

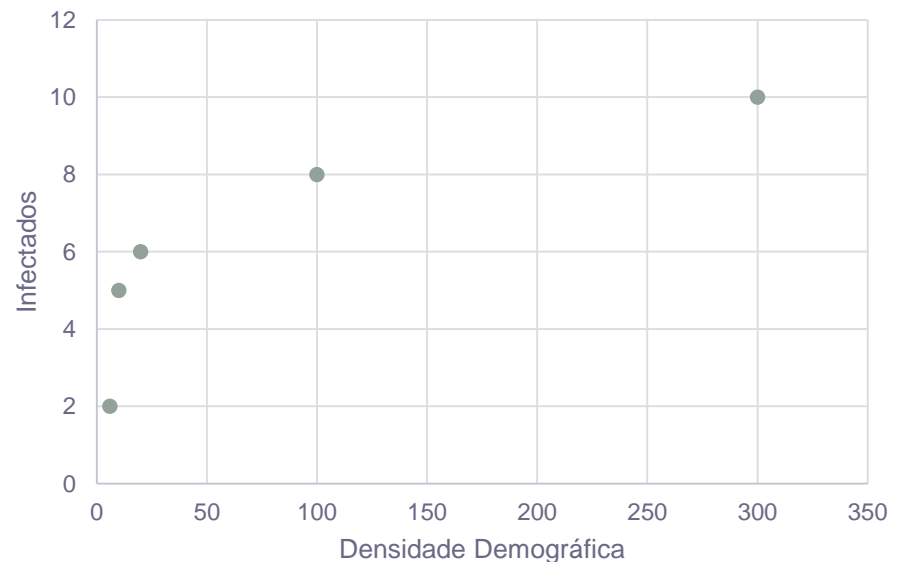
- A fonte de dados ao lado (fictícia) traz características das pessoas que vieram a óbito
- Agrupamentos podem mostrar estatísticas interessantes
- Ex.
  - média de idade: 61 anos
  - Percentual de homens: 63%
  - Percentual de mulheres: 37%
  - Média de idade dos homens: 73 anos
  - Média de idade dos homens: 39 anos
- Consultas SQL podem fazer esses agrupamentos

sexo	idade
M	65
M	89
F	55
M	74
F	20
M	77
F	45
M	66
M	89
F	38
M	55

# Exemplo 2

- As informações (fictícias) de densidade demográfica e número de infectados por 1 milhão de habitantes mostram que existe uma relação entre essas duas variáveis.
- Como obter os dados para popular essa tabela?

Estado	Densidade demográfica	Infectados
Estado A	6	2
Estado B	100	8
Estado C	300	10
Estado D	10	5
Estado E	20	6
...	...	...



# Exemplo 2

- Talvez os dados estejam dispostos de forma diferente
  - E precisem ser ajustados durante a integração

Estado	Densidade demográfica
Estado A	6
Estado B	100
...	...

Estado	Infectados
A	2
B	8
...	...



Estado	Densidade demográfica	Infectados
Estado A	6	2
Estado B	100	8
...	...	...

# Exemplo 2

- Talvez as fontes estejam em formatos diferentes
  - O que exige a criação de códigos que façam a conversão para o formato relacional

Estado	Densidade demográfica
Estado A	6
Estado B	100
...	...

Dados da wikipedia

Estado	Infectados
Estado A	2
Estado B	8
...	...

Dados de uma planilha excel

# Objetivo Geral

- O trabalho é dividido em 3 etapas
  - **(2 pontos)** Criação de um banco de dados
  - **(4 pontos)** Alimentação de um banco de dados
  - **(4 pontos)** Consultas sobre o banco de dados
- Pode ser usado qualquer fornecedor de banco de dados



# Etapa 1

- Criação de um banco de dados
- Passos
  - Escolher uma ou mais fontes de dados referentes ao tema definido
    - Ex. arquivos csv, planilhas excel, páginas web, ...
  - Gerar um esquema de banco de dados relacional **normalizado** referente a essa(s) fonte(s)

# Etapa 1

- Os exemplos ficam interessantes quando
  - A quantidade de dados é grande
  - Existe a possibilidade de realizar agrupamentos
  - Existe a possibilidade de integrar informações que venham de fontes distintas

# Etapa 1

- Os dados não precisam tratar especificamente sobre covid-19, mas questões que estejam relacionadas
- Ex.
  - infra-estrutura em saúde
  - Causas de óbito
  - Situação econômica
  - ...
- Também não se limita ao Brasil.
- Pode-se usar dados referentes a outras países/regiões

# Etapa 2

- **Carga dos dados**
- Nesta etapa o objetivo é alimentar o banco de dados com informações vindas das fontes
  - A alimentação pode ser manual ou automatizada (via ferramenta ou código-fonte)

# Etapa 2

- A nota será relativa ao grau de complexidade envolvido para fazer a importação
  - Número de registros
  - Número de fontes integradas
  - Processo de extração utilizado
  - Necessidade de adequação dos dados
- Ex.
  - Uma fonte com 1000 registros que permita agrupamentos é bastante interessante
  - Já uma fonte com 10 registros é de baixíssima complexidade para importação
    - Mas o trabalho fica interessante caso essa fonte seja integrada uma outra que traz dados complementares.

# Etapa 3

- **Consultas**

- Nesta etapa o objetivo é realizar consultas sobre o banco de dados criado
- As consultas devem trazer dados
  - Resumidos e estatísticos
    - ex. através do group by
  - Informativos
    - Mostrando dados que ajudem a identificar tendências e relações

# Escolha da(s) Fonte(s)

- O trabalho é individual.
- Cada aluno poderá escolher uma ou mais fontes para realizar o trabalho
  - Desde que essas fontes já não estejam sendo usadas por outro aluno ou que o uso dos dados seja distinto
- O aluno deve informar ao professor a respeito de quais fontes deseja trabalhar para fazer a reserva

# Apresentação dos Resultados

- Deve-se preparar uma apresentação contendo os seguintes tópicos
  - Descrição da(s) fonte(s) escolhida(s)
  - Script de criação do banco
  - Esquema relacional gráfico
  - Dificuldades encontradas para realizar a importação
  - Scripts/códigos de importação
  - Consultas elaboradas
  - Análise dos resultados



# Apresentação dos Resultados

- Durante a suspensão de aulas presenciais, as apresentações serão feitas por videoconferência
  - Alunos podem aguardar o retorno das aulas para fazer a apresentação em sala de aula

# Apresentação dos Resultados

- Aos alunos que optarem por apresentar de forma remota
  - Não serão fixadas datas para apresentação
  - Assim que os alunos forem terminando, serão organizados dias de seminário
    - Os apresentadores deverão informar o tempo médio de apresentação para organização do seminário
  - Todos alunos serão convidados a assistir as apresentações
    - Mas os seminários não contarão como presença
- Obs. Avisem o professor quanto à conclusão do trabalho, para a organização dos seminários