

Ciclano de tal

Rodrigo Anes Sena de Araújo, 9763064, rodrigo.anes.araujo@usp.br

William Luis Alves Ferreira, 9847599, william.luis.ferreira@usp.br

Estudo de grande bases de dados: COVID

Brasil

Dez, 2021

Ciclano de tal

Rodrigo Anes Sena de Araújo, 9763064, rodrigo.anes.araujo@usp.br

William Luis Alves Ferreira, 9847599, william.luis.ferreira@usp.br

Estudo de grande bases de dados: COVID

Entrega de projeto para avaliação na disciplina **Mineração a partir de Grandes Bases de Dados - SCC0244** junto ao docente **Caetano Traina Junior** e monitor **Erikson Júlio de Aguiar**.

Universidade de São Paulo – USP

Escola de Engenharia de São Carlos – EESC

Instituto de Ciências Matemáticas e Computação – ICMC

Programa de Pós-Graduação

Brasil

Dez, 2021

Sumário

	Introdução	4
1	QUESTÕES	5
1.1	Preparação e exploração da Base de Dados	5
1.1.1	Execício 1	5
1.1.2	Execício 2	7
1.1.3	Execício 3	9
1.2	Exercícios sobre Funções de Janelamento em SQL	10
1.2.1	Execício 4	10
1.2.2	Execício 5	10
1.2.3	Execício 6	12
1.2.4	Execício 7	13
1.2.5	Execício 8	14
1.2.6	Execício 9	14
1.3	Exercícios sobre Histogramas	15
1.3.1	Execício 10	15
1.3.2	Execício 11	15
1.4	Exercícios sobre Fractais	17
	Considerações finais	18
	REFERÊNCIAS	19
	APÊNDICES	20
	APÊNDICE A – QUISQUE LIBERO JUSTO	21
	APÊNDICE B – NULLAM ELEMENTUM URNA VEL IMPER- DIET SODALES ELIT IPSUM PHARETRA LI- GULA AC PRETIUM ANTE JUSTO A NULLA CURABITUR TRISTIQUE ARCU EU METUS	22
	ANEXOS	23
	ANEXO A – MORBI ULTRICES RUTRUM LOREM.	24

ANEXO B – CRAS NON URNA SED FEUGIAT CUM SOCIIS NATOQUE PENATIBUS ET MAGNIS DIS PARTU- RIENT MONTES NASCETUR RIDICULUS MUS . .	25
ANEXO C – FUSCE FACILISIS LACINIA DUI	26

Introdução

Este documento contempla a resolução das questões propostas como entrega única para disciplina **Mineração a partir de Grandes Bases de Dados - SCC0244** junto ao docente **Caetano Traina Junior** e monitor **Erikson Júlio de Aguiar**, no qual utiliza-se comandos **SQL** no Sistema de Gerenciamento de Banco de Dados (SGBD) [PostgreSQL](#) - versão 14 a fim de explorar o conjunto de dados do repositório [FAPESP COVID-19 DataSharing/BR](#) que armazena conjuntos de dados estruturados sobre Covid-19 em 5 instituições de saúde: Beneficência Portuguesa de São Paulo ,Hospital das Clinicas da Faculdade de Medicina da Universidade de Sao Paulo, GrupoFleury, Hospital Israelita Albert Einstein e Hospital Sírio-Libanês.

Ressalva-se que todo o documento esta associado ao repositório no github disponível em [illiamw/BigDataCovid](#) para armazenamento e disponibilidade dos *scripts* completos para cada questão.

1 Questões

Para as resoluções com *scripts* completos e perenes para execução do projeto será feito o compartilhamento via *hyperlink*, logo, apenas *scripts* curtos ou que possuem caráter intermediário serão destacados entrelinhas neste documento.

1.1 Preparação e exploração da Base de Dados

1.1.1 Execício 1

Deve ser criada uma base de dados que contenha as 3 relações básicas para cada hospital. Para isso, as tabelas devem ser:

- criadas
- carregadas
- executados procedimentos simples de limpeza de dados, incluindo:
 - formatação e definição correta dos tipos de dados e definição das restrições de integridade das tabelas
 - correção de valores fora do padrão indicados na criação das tabelas para valores nulos

Resolução

Carregamento e limpeza prévia:

Utilizando o meta-arquivo de descrição ***HSL_Diccionario_4*** foram criadas as tabelas, e, posteriormente carregados os dados dos arquivos na extensão CSV, porém duas inconsistências foram corrigidas, como:

- “HSL_Exames_4.csv” - Coluna “DE_ORIGEM” consta formato de 4 caracteres alfanuméricos, mas os dados apresentam comprimento variado acima de 4 caracteres. Para esta inconsistência foi modificado o formato para ***text*** (texto livre)
- “HSL_Desfechos_4.csv” - Coluna “DT_DESFECHO” apresentam datas inválidas com valor “DDMMAA”, provavelmente dados faltantes. Para esse foi realizado a substituição via por “ ” (vazio) já que **PostgreSQL** não aceita valor ***null***.

Com tudo, os tipos de dados (formatos) e restrições segue o meta-arquivo de descrição, para o carregamento foi utilizado a função **COPY** atentando-se ao delimitador utilizado no arquivo, por fim, após o carregamento foi feita verificação da unicidade das colunas nas quais foram identificadas as restrições (i.e, *constraints* em inglês).

Logo, é possível obter o seguinte script [1_inicializacao.sql](#). Por fim, após o início da questão 5 teve a necessidade de carregar tabelas exames dos demais hospitais na base de dados Fapesp para covid, os hospitais são: Beneficência Portuguesa de São Paulo, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Grupo Fleury, Hospital Israelita Albert Einstein e Hospital Sírio-Libanês.

Avaliando base e limpando:

Com os dados carregados foi executado uma avaliação coluna a coluna dos valores distintos através do *script*:

```
SELECT DISTINCT "COLUNA"  
FROM TABELA  
ORDER BY "COLUNA";
```

Tabela pacientes:

Coluna **AA_NASCIMENTO**: verificou-se a presença de valores como “AAAA” e “YYYY”, como valores indefinidos para este caso substituímos pela média de aproximada 1978.

Demais colunas seguem sem inconsistências, porém com valores nulos não resolúveis.

Tabela desfechos:

Coluna **DT_DESFECHO**: verificou-se a presença de valores como *null* que associa-se a ocorrência de algum tipo de óbito, ao verificar os valores de DE_DESFECHO verifica-se a informação temporal que ocorreu o óbito, logo a princípio foram mantidos os valores *null* para tratamento posterior em análises.

Tabela exames:

As colunas **DE_RESULTADO**, **CD_UNIDADE** e **DE VALOR REFERENCIAL** apresentam variação na representações de valores, primeiro, trata-se de exames com domínio de valores distintos, segundo, ocorre formatação distinta em exames de mesmo valor tratado posteriormente conforme desenvolvimento das questões.

O *script* completo para executar da análise e eventual limpeza está disponível em: [1_2AvaliacaoLimpeza](#).

1.1.2 Execício 2

Faça uma análise exploratória das tabelas, avaliando:

1. Os principais indicadores estatísticos sobre cada atributo de cada relação (valores distintos, variância, nulos, etc.)
2. Tabela Pacientes:
 - Qual a quantidade de pacientes presente na base de dados? Quantos são homens e quantos são mulheres?
 - Qual é faixa etária dos pacientes homens e mulheres?
 - Qual a distribuição dos quartis dentre de cada faixa?
 - Qual a distribuição em cada gênero por década de vida?
3. Tabela exames:
 - Qual a maior quantidade de exames solicitados para um único paciente?
 - Qual é a média de exames pedidos para homens e para mulheres?
 - Quantos exames de Coronavírus (2019-nCoV) foram solicitados? E quantos foram positivos?
 - Para cada idade, mostre os resultados dos exames de Coronavírus (2019-nCoV)
4. Tabela Desfechos:
 - Qual é o desfecho para a maioria dos casos registrados?
 - E para cada distribuição por gênero e por década de vida?

Resolução

Item 1: Como iniciado no item 1.1.1 (Exercício 1) foi verificado os valores distintos para todas as colunas, além de verificar valores inconsistentes ou nulos, já para execução dos demais indicadores estatísticos realizou-se uma análise descritiva através da sumarização dos dados não categórico (por exemplo, no atributo AA_NASCIMENTO), logo temos a sumarização:

```
SELECT COUNT(*),  
MIN("AA_NASCIMENTO"),  
MAX("AA_NASCIMENTO"),  
AVG("AA_NASCIMENTO") AS mean,  
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY "AA_NASCIMENTO") AS median,  
ROUND(STDDEV("AA_NASCIMENTO"), 2) AS std
```


FROM pacientes;

Figura 1 – Sumarização coluna AA_NASCIMENTO

	count bigint	min integer	max integer	mean numeric	median double precision	std numeric
1	14673	1931	2020	1977.6024671164724324	1979	16.92

Fonte: Pelos próprios autores

disponibilizado na integra como [2_1Estatistica](#) e [1_2AvaliacaoLimpeza](#) para todas as tabelas.

Item 2:

O total de pacientes registrados na tabela **pacientes** é 14673.

	HOMENS	MULHERES
Qtd. Pacientes	7381	7292
Amplitude	89 anos	89 anos
Faixa	0 - 89 anos	0 - 89 anos

Tabela 1 – Sumarização pela idade e faixa etária

Mínimo	Q1	Q2 (Mediana)	Q3	Máximo
0	31	41	53	89

Tabela 2 – Distribuição dos quantils pela faixa etária

FAIXA (DÉCADAS)	1-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Recorrência Total	343	728	2059	3672	3431	2115	1230	726	369
Recorrência Mulheres	147	368	1122	1964	1708	976	529	306	172
Recorrência Homens	196	360	937	1708	1723	1139	701	420	197

Tabela 3 – Distribuição por década

Item 3:

Paciente com id=376CA07DC7EC73C108D1F44579C7134F realizou 18428 exames, esse valor exorbitante se deve a internação do paciente em UTI.

Média de exames solicitadas é aproximadamente 2568 para Homens e 1653 para Mulheres.

Para o analito (atributo DE_ANALITO) com valor “Coronavírus (2019-nCoV)” temos um total de 8207 solicitações, no qual destes 5370 foram **positivos**.

Para cada idade, mostre os resultados dos exames de Coronavírus (2019-nCoV), devido ao comprimento do resultado desta questão esta disponível em [2_2PositivosPorIdade.csv](#).

Item 4:

O desfecho mais comum é o de '*Alta Administrativa*', com 66908 casos, seguido de '*Alta médica melhorado*' com 21702 casos e '*Desistência do atendimento*' com 401 casos.

O desfecho mais comum para todas os grupos de gênero e década de vida continua sendo '*Alta Administrativa*'.

Para a resolução dos **Itens 2, 3 e 4** os *scripts* estão disponíveis em [2_2EstatisticasEspecificas.sql](#).

1.1.3 Execício 3

Considerando as tabelas e as consultas solicitadas anteriormente, escreva/projete uma consulta para extrair algum conhecimento da base de dados que não foi descoberto pelas consultas anteriores. Apresente uma breve justificativa do objetivo da consulta e, por que esse objetivo é relevante:

Resolução

A consulta adicional que foi desenvolvida agrupa as datas da coluna ***DT _ATENDIMENTO*** da tabela *desfechos* em grupos de mesmo mês e ano. A consulta contabiliza a quantidade de atendimentos para cada par mês-ano.

Mês-ano	02/20	03/20	04/20	05/20	06/20	07/20	08/20	09/20	10/20
Atendimentos	514	4468	3257	3951	4923	5813	5915	5750	5586
	11/20	12/20	01/21	02/21	03/21	04/21	05/21	06/21	
	7091	7076	6630	5938	7844	5644	7080	2457	

Tabela 4 – Quantidade de atendimentos por mês-ano

O objetivo desta consulta é observar como a frequência de atendimentos varia com o tempo. É possível extrair informações como o período que teve maior pico de atendimentos (Março/2021) e o que teve menor quantidade (Fevereiro/2020). Além de observar as oscilações de quantidade de atendimentos de mês a mês. O *script* completo disponível em [3_ConsultaEspecial.sql](#).

1.2 Exercícios sobre Funções de Janelamento em SQL

1.2.1 Execício 4

Considere que se pretende obter os pacientes ‘mais novos’ e ‘mais velhos’ em cada cidade, na base Fapesp-Covid. Escreva um comando que responda a essa consulta:

- Com uma sub-consulta usando apenas a cláusula ‘GROUP BY’
- Com sub-consultas usando a construção CTE (Common Table Expression ‘WITH queries’);
- Usando ‘Window functions’:

Resolução

Foi escrito um comando que encontra todos os pacientes que se encontram nos extremos de maior ou menor ano de nascimento (***AA_NASCIMENTO***) de seu respectivo município. Foram encontrados 104 pacientes que satisfazem tais condições.

As 3 formas de consulta se encontram em [4_Janelamento](#).

1.2.2 Execício 5

A tabela de Exames reporta uma medida sobre um analito em cada tupla. Portanto, os exames que medem diversos analitos são representados em diversas tuplas. No entanto, pode-se assumir que, se foram registrados dois exames iguais no mesmo dia para o mesmo paciente, pode-se assumir como valor a ser considerado a média dos valores medidos em cada analito.

- Escreva uma consulta que mostre quais analitos podem ser medidos em exames de ‘hemograma’, em cada hospital.
- Compare os nomes dos analitos entre os diferentes hospitais, e execute um processo de atualiza,

Resolução

Após realizar uma busca pelos valores distintos da coluna ***DE_EXAME*** para os casos em que ***DE_ANALITO*** = ‘Hemograma’. Temos os seguintes valores distintos de analitos para os *Hospital Sírio Libanês*:

- | | | |
|-----------------|------------------|---------------|
| • Basófilos | • Bastonetes | • Blastos |
| • Basófilos (%) | • Bastonetes (%) | • Blastos (%) |

- CHCM
- Eosinófilos
- Eosinófilos (%)
- Eritrócitos
- Fração Imatura de Plaquetas
- HCM
- Hematócrito
- Hemoglobina
- Índice de Green & King
- Leucócitos
- Linfócitos
- Linfócitos (%)
- Metamielócitos
- Metamielócitos (%)
- Mielócitos
- Mielócitos (%)
- Monócitos
- Monócitos (%)
- Morfologia, SB
- Morfologia, SVE
- Neutrófilos
- Neutrófilos(%)
- Plaquetas
- Plasmócitos (%)
- Plasmóticos
- Promielócitos
- Promielócitos (%)
- RDW
- Segmentados
- Segmentados (%)
- VCM
- Volume plaquetário médio

Figura 2 – Trecho da consulta para verificar os valores distintos entre todas as tabelas exames

1	DE_ANALITO	?column?
2		HC
3	.	HC
4	Basófilos	BPSP
5	Basófilos	HSL
6	Basófilos	HC
7	Basófilos	EINSTEIN
8	Basófilos	FL
9	Basófilos #	EINSTEIN
10	Basófilos (%)	BPSP
11	Basófilos (%)	HSL
12	Basófilos (%)	FL
13	Bastonetes	EINSTEIN

Fonte: Pelos próprios autores

Esses valores foram utilizados como referência para as tabelas exames para os demais hospitais: Beneficência Portuguesa de São Paulo ,Hospital das Clinicas da Faculdade de Medicina da Universidade de Sao Paulo, GrupoFleury, Hospital Israelita Albert Einstein e Hospital Sírio-Libanês, foram unidos por **UNION ALL** onde foi conferido os valores distintos como no *output* [5_ComparacaoHemogramas.csv](#) (trecho na figura ??) , por fim

Figura 3 – Trecho da consulta para verificar os valores distintos ao fim do script



1	DE_ANALITO
2	
3	
4	Basófilos
5	Basófilos (%)
6	Bastonetes
7	Bastonetes (%)
8	Blastos
9	Blastos (%)
10	Células Anômalas
11	CHCM
12	Concentração de Hemoglobina Corpuscular
13	Eosinófilos

Fonte: Pelos próprios autores

realizado o **UPDATE** com a clausula **CASE** e conferido os valores distintos novamente como no *output* [5_ComparacaoHemogramasVerificacao.csv](#) (trecho na figura ??).

O código deste *script* completo usado nessa questão encontra-se em [5_Hemograma](#).

1.2.3 Execício 6

Escreva uma consulta que gere a relação de todos os exames de colesterol que foram efetuados, de maneira que cada tupla dessa relação inclua as medidas de todos analitos correspondentes desse exame (executar o pivotamento da relação de exames). Para isso, considere que cada exame de cada paciente é realizado em um único dia, e que se houver repetição de medidas do mesmo analito, deve ser considerada a média de todas as medidas desse analito. Analitos não medidos num exame devem ficar nulos. Inclua nessa tabela o desfecho que o paciente teve para o atendimento onde esse exame foi feito.

Resolução

Foram encontrados 7 tipos de valores no campo **DE_ANALITO**: *'Colesterol não-HDL, soro'*, *'VLDL-Colesterol'*, *'Hdl-Colesterol'*, *'HDL-Colesterol'*, *'V-Colesterol'*, *'LDL Colesterol'*, *'Colesterol total'*.

Esses 7 valores dão origem a 5 novas colunas: *'HDL'*, *'VLDL'*, *'LDL'*, *'Não HDL'*, *'TOTAL'*

Alguns valores no campo **DE_RESULTADO** não estavam em forma numérica e foram filtrados para fora da busca utilizando um regex,

Por fim, foi realizado um JOIN com a tabela **desfechos** para acrescentar a informação da descrição dos desfechos nas tuplas. As 10 primeiras tuplas desta query

foram:

	ID_PACIENTE character (32)	DT_COLETA date	HDL numeric	VLDL numeric	LDL numeric	Não HDL numeric	TOTAL numeric	DE_DESFECHO text
1	000F0BC139D2846DB86AA32B8F05B215	2021-02-11	38.00	26.00	171.00	197.00	235.00	Alta Administrativa
2	000F0BC139D2846DB86AA32B8F05B215	2020-09-24	39.00	28.00	180.00	208.00	247.00	Alta Administrativa
3	002B919CC409B11DE52FB212379BE2CB	2020-09-23	79.00	15.00	152.00	167.00	246.00	Alta Administrativa
4	004688799FD293C3ABE0A07209FD8B75	2020-11-23	67.00	33.00	198.00	[null]	298.00	Alta Administrativa
5	004688799FD293C3ABE0A07209FD8B75	2021-02-10	36.00	28.00	106.00	[null]	170.00	Alta Administrativa
6	004688799FD293C3ABE0A07209FD8B75	2020-09-08	36.00	28.00	90.00	[null]	154.00	Alta Administrativa
7	004688799FD293C3ABE0A07209FD8B75	2020-10-01	41.00	43.00	174.00	[null]	258.00	Alta médica melhorado
8	004688799FD293C3ABE0A07209FD8B75	2020-05-09	36.00	27.00	95.00	[null]	158.00	Alta Administrativa
9	005B118C512EE0B624AB7293F42B1D54	2020-06-26	53.00	17.00	95.00	[null]	165.00	Alta Administrativa
10	005B118C512EE0B624AB7293F42B1D54	2021-05-06	42.00	17.00	124.00	141.00	183.00	Alta Administrativa

O código usado nesse exercício encontra-se em [6_Colesterol](#).

1.2.4 Execício 7

Escreva uma consulta equivalente à anterior, agora para os exames de hemograma que foram efetuados. Nessas tabelas, cada tipo de exame seguiu uma estrutura diferente. Neste caso a principal diferença para gerar as duas tabelas é que, enquanto para obter os exames de colesterol cada medida é independente, e a escolha das tuplas teve que ser feita diretamente pelo atributo ‘De Analito’, os exames de hemograma são identificados por um único valor no tipo de exame (embora hospitais diferentes possam usar nomes diferentes para o mesmo exame) e portanto o atributo ‘De Exame’ pode ser usado como filtro de seleção.

Resolução

Para realizar uma consulta equivalente à consulta proposta na questão 6, temos que criar uma nova coluna para cada valor distinto de **DE_ANALITO**. Contudo temos 38 valores distintos. Isso geraria tuplas muito longas e com muitos valores nulos, já que cada exame em geral analisa somente uma pequena parte dos analitos.

As 10 primeiras tuplas desta query foram:

	ID_PACIENTE character (32)	DT_COLETA date	BASOFILOS numeric	BASOFILOS % numeric	BASTONETES numeric	BASTONETES % numeric	BLASTOS numeric	BLASTOS % numeric	CHCM numeric	EOSINOFILOS numeric	EOSINOFIL numeric
1	00017961865C4F766FDBB3CD8FE0BF80	2020-08-25	40.00	[null]	[null]	[null]	[null]	[null]	[null]	60.00	
2	0003B368F65004E14332CD44BEE6E600	2020-12-28	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	
3	00293E8E4B5FE4DABA62131274B3685A	2021-02-15	30.00	[null]	[null]	[null]	[null]	[null]	[null]	[null]	
4	00293E8E4B5FE4DABA62131274B3685A	2021-02-06	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	
5	00293E8E4B5FE4DABA62131274B3685A	2021-02-10	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	
6	002B919CC409B11DE52FB212379BE2CB	2020-05-29	60.00	[null]	[null]	[null]	[null]	[null]	[null]	710.00	
7	003051C9B19101D1C10C5DC654384017	2020-06-12	[null]	[null]	[null]	[null]	[null]	[null]	[null]	60.00	
8	003F1F4C194763E4A00FD809AF5FA6AA	2020-04-21	[null]	[null]	[null]	[null]	[null]	[null]	[null]	110.00	
9	003F1F4C194763E4A00FD809AF5FA6AA	2020-04-01	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	
10	003F1F4C194763E4A00FD809AF5FA6AA	2020-04-11	30.00	[null]	[null]	[null]	[null]	[null]	[null]	70.00	

Por causa do comprimento das tuplas, não foi possível apresentar todas as colunas na imagem.

O código usado nesse exercício encontra-se em [7_Hemograma_parte2](#).

1.2.5 Execício 8

Considerando exames de Covid, substitua os valores do atributo ‘De Resultado’ que tenham valores numéricos para ‘Positivo’ e ‘negativo’ considerando o atributo ‘CD ValorReferencia’.

Resolução

Para esta questão temos a análise de valores distintos para compreender as variações de representação de covid, com:

```
SELECT DISTINCT "DE_ANALITO", "DE_RESULTADO", "DE_VALOR_REFERENCIA"
FROM exames
WHERE(LOWER("DE_ANALITO") LIKE '%ovid'
OR
LOWER("DE_ANALITO") LIKE '%oron%')
order by "DE_ANALITO"
```

após está análise foi verificado a coluna **DE_VALOR_REFERENCIA** para realizar a substituição dos valores numéricos em “Positivo” e “Negativo” conforme valor de referência. Em seguida, foi extraído o valor numérico do formato **string** criado para coluna **DE_RESULTADO** e convertido no formato numérico **::numeric**, com isso testado em clausula **SELECT** concluímos executando o **UPDATE** utilizando o **SELECT** para o **FROM** deste **UPDATE**.

O código deste *script* completo usado nesse exercício encontra-se em [8_replaceCovid.sql](#).

1.2.6 Execício 9

Faça uma consulta equivalente ‘a de exames de hemograma, agora para exames vinculados a testes de covid, usando o resultado da consulta anterior. Inclua na relação resultante o número de dias entre dois exames que tenham resultado mudado a medida entre ‘positivo’ e ‘negativo’ para Covid.

Resolução

(QUESTÃO 9)

O código deste *script* completo usado nesse exercício encontra-se em [9_Covid_parte2.sql](#).

1.3 Exercícios sobre Histogramas

1.3.1 Execício 10

Escreva um comando em SQL que crie:

- O histograma equi-largura de distribuição das idades dos pacientes, de maneira que a largura de cada bin do histograma corresponda a “duas idades”. Atente para que todas as “idades possíveis”, desde 0 até a maior idade registrada esteja representada no histograma.
- Modifique esse comando para gerar um histograma equi-largura com 10 bins.

Resolução

Temos os histogramas de equi-largura com as distribuições de bins 2 (figuras 4 e 5) e 10 (figuras 6), tanto com as faixas definidas quanto o bin que a quantidade participa, além de outras variações contidas no *script* completo.

Figura 4 – Histograma Idade equi-largura: 2 bins, idade inicial

	idade numeric	conta bigint
1	0	8786
2	45	5887

Fonte: Pelos próprios autores

Figura 5 – Histograma Idade equi-largura: 2 bins, com faixas

	ini numeric	fim numeric	qtd bigint
1	0	44	8786
2	44	89	5887

Fonte: Pelos próprios autores

O código deste *script* completo usado nesse exercício encontra-se em [10_HistogramaIdade.sql](#).

1.3.2 Execício 11

Escreva um comando em SQL que crie o histograma tri-dimensional equi-largura de distribuição de exames (da tabela ExamLabs), tendo por dimensões:

- DE_Hospital

Figura 6 – Histograma Idade equi-largura: 10 bins, com faixas

	ini numeric	fim numeric	qtd bigint
1	0	8	307
2	8	17	539
3	17	26	1497
4	26	35	2840
5	35	44	3603
6	44	53	2298
7	53	62	1697
8	62	71	972
9	71	80	607
10	80	89	313

Fonte: Pelos próprios autores

- DE_Origem, contabilizando as origens em (Hosp)ital, (Lab)oratório, (inter)nação ou (pronto) socorro para caracterizar: (Hosp)ital, (Lab)oratório, (Atend)imento e os demais como (Outros)
- DE_Exame, contabilizando os exames como sendo de (Hemogr)ama, (colest)erol, e (covid) ou (pcr) ou (igm) ou (igg) para caracterizar (Hemograma), (Colesterol), (Covid) e (outros).

Resolução

Parte do *script* desenvolvido na questão 5 foi utilizado para gerar a coluna “DE_Hospital”, além de complementar as colunas “DE_Origem” e “DE_Exame”.

Com o **SELECT** da questão 5:

```
(select "DE_EXAME", "DE_ORIGEM", 'HSL' as "DE_HOSPITAL"
from exames
UNION ALL
select "DE_EXAME", "DE_ORIGEM", 'BPSP' as "DE_HOSPITAL"
from examesbpsp)
UNION ALL
(select "DE_EXAME", "DE_ORIGEM", 'EINSTEIN' as "DE_HOSPITAL"
from exameseinstein
UNION ALL
select "DE_EXAME", "DE_ORIGEM" , 'HC' as "DE_HOSPITAL"
from exameshc)
UNION ALL
select "DE_EXAME", "DE_ORIGEM", 'FL' as "DE_HOSPITAL"
from examesfl)
```

Em seguida foi realizado a simplificação das colunas conforme especificado na questão, logo temos como resultado final a tabela [11_Histograma.csv](#) disposta trecho na figura 7.

Figura 7 – Histograma tri-dimensional equi-largura

1	hospitaiscol	origemcol	examescol	contagem
2	BPSP	Hospital	Colesterol	11119
3	BPSP	Hospital	Covid	39517
4	BPSP	Hospital	Hemograma	2223416
5	BPSP	Hospital	Outros	2434859
6	BPSP	Laboratório	Colesterol	29088
7	BPSP	Laboratório	Covid	28455
8	BPSP	Laboratório	Hemograma	249176
9	BPSP	Laboratório	Outros	323663
10	EINSTEIN	Hospital	Covid	159035
11	EINSTEIN	Hospital	Hemograma	1480520
12	EINSTEIN	Hospital	Outros	1775600
13	FL	Laboratório	Colesterol	2087905
14	FL	Laboratório	Covid	5496589
15	FL	Laboratório	Hemograma	13157011
16	FL	Laboratório	Outros	18826263

Fonte: Pelos próprios autores

O código deste *script* completo usado nesse exercício encontra-se em [11_HistogramaMultidimensional.sql](#).

1.4 Exercícios sobre Fractais

Considerações finais

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Phasellus id magna. Duis malesuada interdum arcu. Integer metus. Morbi pulvinar pellentesque mi. Suspendisse sed est eu magna molestie egestas. Quisque mi lorem, pulvinar eget, egestas quis, luctus at, ante. Proin auctor vehicula purus. Fusce ac nisl aliquam ante hendrerit pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi wisi. Etiam arcu mauris, facilisis sed, eleifend non, nonummy ut, pede. Cras ut lacus tempor metus mollis placerat. Vivamus eu tortor vel metus interdum malesuada.

Sed eleifend, eros sit amet faucibus elementum, urna sapien consectetur mauris, quis egestas leo justo non risus. Morbi non felis ac libero vulputate fringilla. Mauris libero eros, lacinia non, sodales quis, dapibus porttitor, pede. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi dapibus mauris condimentum nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Etiam sit amet erat. Nulla varius. Etiam tincidunt dui vitae turpis. Donec leo. Morbi vulputate convallis est. Integer aliquet. Pellentesque aliquet sodales urna.

Referências

Apêndices

APÊNDICE A – Quisque libero justo

Quisque facilisis auctor sapien. Pellentesque gravida hendrerit lectus. Mauris rutrum sodales sapien. Fusce hendrerit sem vel lorem. Integer pellentesque massa vel augue. Integer elit tortor, feugiat quis, sagittis et, ornare non, lacus. Vestibulum posuere pellentesque eros. Quisque venenatis ipsum dictum nulla. Aliquam quis quam non metus eleifend interdum. Nam eget sapien ac mauris malesuada adipiscing. Etiam eleifend neque sed quam. Nulla facilisi. Proin a ligula. Sed id dui eu nibh egestas tincidunt. Suspendisse arcu.

APÊNDICE B – Nullam elementum urna vel imperdiet sodales elit ipsum pharetra ligula ac pretium ante justo a nulla curabitur tristique arcu eu metus

Nunc velit. Nullam elit sapien, eleifend eu, commodo nec, semper sit amet, elit. Nulla lectus risus, condimentum ut, laoreet eget, viverra nec, odio. Proin lobortis. Curabitur dictum arcu vel wisi. Cras id nulla venenatis tortor congue ultrices. Pellentesque eget pede. Sed eleifend sagittis elit. Nam sed tellus sit amet lectus ullamcorper tristique. Mauris enim sem, tristique eu, accumsan at, scelerisque vulputate, neque. Quisque lacus. Donec et ipsum sit amet elit nonummy aliquet. Sed viverra nisl at sem. Nam diam. Mauris ut dolor. Curabitur ornare tortor cursus velit.

Morbi tincidunt posuere arcu. Cras venenatis est vitae dolor. Vivamus scelerisque semper mi. Donec ipsum arcu, consequat scelerisque, viverra id, dictum at, metus. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut pede sem, tempus ut, porttitor bibendum, molestie eu, elit. Suspendisse potenti. Sed id lectus sit amet purus faucibus vehicula. Praesent sed sem non dui pharetra interdum. Nam viverra ultrices magna.

Aenean laoreet aliquam orci. Nunc interdum elementum urna. Quisque erat. Nullam tempor neque. Maecenas velit nibh, scelerisque a, consequat ut, viverra in, enim. Duis magna. Donec odio neque, tristique et, tincidunt eu, rhoncus ac, nunc. Mauris malesuada malesuada elit. Etiam lacus mauris, pretium vel, blandit in, ultricies id, libero. Phasellus bibendum erat ut diam. In congue imperdiet lectus.

Anexos

ANEXO A – Morbi ultrices rutrum lorem.

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

ANEXO B – Cras non urna sed feugiat cum sociis natoque penatibus et magnis dis parturient montes nascetur ridiculus mus

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

ANEXO C – Fusce facilisis lacinia dui

Phasellus id magna. Duis malesuada interdum arcu. Integer metus. Morbi pulvinar pellentesque mi. Suspendisse sed est eu magna molestie egestas. Quisque mi lorem, pulvinar eget, egestas quis, luctus at, ante. Proin auctor vehicula purus. Fusce ac nisl aliquam ante hendrerit pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi wisi. Etiam arcu mauris, facilisis sed, eleifend non, nonummy ut, pede. Cras ut lacus tempor metus mollis placerat. Vivamus eu tortor vel metus interdum malesuada.