

**CENTRO FEDERAL DE EDUCACAO TECNOLÓGICA
CELSO SUCKOW DA FONSECA**

Visualização de Dados Institucionais

Gabriel Guimarães Garcia
Luiz Antonio da Ponte Junior

Professor Orientador:
Renato Campos Mauro, M.Sc.

**Rio de Janeiro
Setembro de 2017**

**CENTRO FEDERAL DE EDUCACÃO TECNOLÓGICA
CELSO SUCKOW DA FONSECA**

Visualização de Dados Institucionais

Gabriel Guimarães Garcia

Luiz Antonio da Ponte Junior

Projeto final apresentado em cumprimento às
normas do Departamento de Educação
Superior do Centro Federal de Educação
Tecnológica Celso Suckow da Fonseca,
CEFET/RJ, como parte dos requisitos para
obtenção do título de Tecnólogo em Sistemas
para Internet.

Professor Orientador:

Renato Campos Mauro, M.Sc.

**Rio de Janeiro
Setembro de 2017**

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

G216 Garcia, Gabriel Guimarães
Visualização de dados institucionais / Gabriel Guimarães
Garcia, Luiz Antonio da Ponte Junior.—2017.
xii, 45f. + apêndices : il.color. , grafs. ; enc.

Projeto Final (Tecnólogo) Centro Federal de Educação
Tecnológica Celso Suckow da Fonseca , 2017.
Bibliografia : f. 43-45
Orientador : Renato Campos Mauro

1. Internet – Programas de computador. 2. Visualização da
informação. I. Ponte Junior, Luiz Antonio da. II. Mauro, Renato
Campos (Orient.). III. Título.

CDD 004.678

AGRADECIMENTOS

Agradece-se ao Professor e Orientador Renato Campos Mauro por ter nos indicado esse projeto e por toda a ajuda e paciência durante o entendimento da biblioteca D3.js.

RESUMO

O avanço da tecnologia e o aumento do número de dispositivos eletrônicos colaboram para o volume de dados gerados diariamente. Este volume possui dimensões tão grandes que extrair informações deixa de ser uma tarefa simples. Analisar dados brutos de um conjunto pequeno de dados não é algo complicado e pode ser uma tarefa que não demande de uma máquina para ser executada, porém dependendo da complexidade e do tamanho deste conjunto de dados, a extração de informações pode deixar de ser viável caso não se utilize computadores para tal. Ao se extrair as informações de um conjunto de dados, as mesmas devem ser exibidas de alguma forma. Decidir qual forma utilizar é crucial no processo da transmissão do conteúdo das informações, pois ela irá impactar na compreensão que os usuários têm dos resultados obtidos. A visualização de dados é então uma ferramenta utilizada para exibir os dados já selecionados de um conjunto de dados. Ela é usada muitas vezes para analisar os dados, obter e gerar estatísticas e auxiliar em processos de tomada de decisão. Nesse contexto a visualização de dados mostrou-se útil para realizar uma análise dos dados referentes ao processo seletivo de cursos de ensino de nível Médio e Técnico da unidade Maracanã da instituição de ensino Centro Federal de Educação Tecnológica Celso Suckow da Fonseca. Para tal foram realizadas consultas à base de dados dos candidatos a fim de se obter resultados para critérios específicos e em seguida foram selecionadas diferentes representações gráficas para exibir esses resultados e assim se obter uma melhor compreensão dos mesmos.

Palavras-chave: visualização de dados; gráficos; representações gráficas; análise de dados; instituições de ensino; processos seletivos

ABSTRACT

The advancement of technology and the increase of the number of electronic devices collaborates to the bulk of data that are daily generated. This bulk has massive dimensions that extract them stops being an easy task. Analyze raw data from a small dataset is not so complicated and may be a task which does not need a help from a machine to be executed, however depending on the complexity and the size of this dataset, the extraction of data can stop being practicable if a machine is not used for it. When the information is extracted from a dataset, it must be displayed somehow. Decide which visualizations' form to use is crucial in the process of transmitting the content of information because this will impact in the comprehension that users have about the obtained results. Data visualization is a tool that is used to display previously selected data from a dataset. It is often used to analyze data, obtain and generate statistics and to help during decision-making processes. In this context data visualization showed itself useful to perform a analysis of data about a selective process of courses of High School and Technical School of the campus Maracanã from the educational institution Centro Federal de Educação Tecnológica Celso Suckow da Fonseca. To perform that were made queries to the database of candidates willing to obtain results to comply with specific criteria and then were selected distinct graphical representations to display these results and have a better comprehension of them.

Keywords: data visualization; graphics; graphical representation; data analysis; educational institutions; selective processes

SUMÁRIO

1	Introdução	1
2	Fundamentação Teórica	4
2.1	Ferramentas Utilizadas	4
2.1.1	PHP	4
2.1.2	MySQL	5
2.1.3	SQL	5
2.1.4	Biblioteca D3	5
2.1.5	JSON	6
2.1.6	Conversão para JSON	6
2.1.7	TSV	6
2.1.8	CSV	6
2.1.9	SVG	6
2.1.10	Armazenamento dos Dados	7
2.2	Conceitos	7
2.2.1	Mapeamento dos Dados	7
2.3	Trabalhos Relacionados	10
2.3.1	The Ignorance Project	10
2.3.2	Fatores da Percepção Visual Humana na Visualização de Dados	11
2.3.3	VISDADOSENM	12
3	Desenvolvimento	13
3.1	Definição das Representações	13
3.2	Critérios de Escolha de Atributos	14
3.3	Precauções	14
3.4	Estrutura e Organização do Dataset	15
3.5	Visualizações	15
3.5.1	Bubble Menu	15
3.5.2	Escolha do Bubble Menu	18
3.5.3	Bar Chart	18
3.5.4	Bubble Chart	18

3.5.5	Donut Chart	21
3.5.6	Dual Scale D3 Bar Chart	23
3.5.7	Grouped Bar Chart	25
3.6	Processamento das Consultas	26
3.7	Comunicação com a Ferramenta D3	28
4	Avaliação Experimental	29
4.1	Usabilidade	29
4.2	Método Usability Scale (SUS)	30
4.2.1	Como Aplicar o Método SUS	30
4.2.2	Resultados Obtidos ao Aplicar o SUS no Projeto	33
4.3	Formulário de Avaliação da Transmissão de Informações	35
4.4	Roteiro de Avaliação da Aplicação	36
5	Conclusão	38
5.1	Problemas e Dificuldades Encontrados	39
5.2	Trabalhos Futuros	42
	Referências Bibliográficas	42
A	Roteiro para Avaliação da Aplicação	46
B	Formulário System Usability Scale (SUS)	48
C	Formulário de Avaliação da Transmissão de Informações	51
D	Resultados Obtidos do Formulário de Avaliação da Transmissão de Informações	55
D.1	Resolução do Formulário	63
E	Questionário Socioeconômico	64

LISTA DE FIGURAS

FIGURA 1:	Opções básicas de exibição. Figura obtida de [1]	8
FIGURA 2:	Tipos de representação agrupados pelas 4 categorias: comparação, composição, distribuição e relacionamento. Figura obtida de [2]	9
FIGURA 3:	Bubble Menu sem bolha selecionada	17
FIGURA 4:	Bubble Menu com bolha Bubble Chart - Bairro selecionada	17
FIGURA 5:	Gráfico Bar Chart para representar número de ocorrências para o atributo etnia na unidade Maracanã	20
FIGURA 6:	Gráfico Bubble Chart para representar a quantidade de candidatos por bairros existentes no dataset da unidade Maracanã	22
FIGURA 7:	Gráfico Donut Chart para representar a quantidade de candidatos por curso desejado na unidade Maracanã	24
FIGURA 8:	Gráfico Dual Scale D3 Bar Chart usado para representar a quantidade de candidatos do sexo masculino e do sexo feminino por curso da unidade Maracanã	26
FIGURA 9:	Gráfico Grouped Bar Chart para representar o número de candidatos que se enquadram e que não se enquadram em categorias de cotas agrupando por cursos da unidade Maracanã	27
FIGURA 10:	Conversão de conteúdo armazenado na variável em PHP para o formato JSON	28
FIGURA 11:	Exemplo de formulário utilizando o método SUS. Figura obtida de [3]	31
FIGURA 12:	Cálculo da pontuação de um formulário utilizando o método SUS. Figura obtida de [4]	31
FIGURA 13:	Pontuações finais obtidas pelo método SUS. Figura obtida de [5]	32
FIGURA 14:	Resultado das pontuações finais de 50 usuários usando o método SUS. Figura obtida de [6]	33
FIGURA 15:	Resultado das pontuações finais de usuários usando o método SUS para a aplicação desenvolvida	34
FIGURA 16:	Número de ocorrências de cada nota por questão do formulário SUS para os 32 usuários	35

FIGURA 17:	Gráfico Donut Chart para representar o número de ocorrências de candidatos por bairro da unidade Maracanã	41
FIGURA 18:	Formulário System Usability Scale (SUS) página 1 de 3	48
FIGURA 19:	Formulário System Usability Scale (SUS) página 2 de 3	49
FIGURA 20:	Formulário System Usability Scale (SUS) página 3 de 3	50
FIGURA 21:	Formulário de Avaliação da Transmissão de Informações página 1 de 4	51
FIGURA 22:	Formulário de Avaliação da Transmissão de Informações página 2 de 4	52
FIGURA 23:	Formulário de Avaliação da Transmissão de Informações página 3 de 4	53
FIGURA 24:	Formulário de Avaliação da Transmissão de Informações página 4 de 4	54
FIGURA 25:	Resultados do formulário de Avaliação da Transmissão de Informações página 1 de 8	55
FIGURA 26:	Resultados do formulário de Avaliação da Transmissão de Informações página 2 de 8	56
FIGURA 27:	Resultados do formulário de Avaliação da Transmissão de Informações página 3 de 8	57
FIGURA 28:	Resultados do formulário de Avaliação da Transmissão de Informações página 4 de 8	58
FIGURA 29:	Resultados do formulário de Avaliação da Transmissão de Informações página 5 de 8	59
FIGURA 30:	Resultados do formulário de Avaliação da Transmissão de Informações página 6 de 8	60
FIGURA 31:	Resultados do formulário de Avaliação da Transmissão de Informações página 7 de 8	61
FIGURA 32:	Resultados do formulário de Avaliação da Transmissão de Informações página 8 de 8	62
FIGURA 33:	Questionário socioeconômico página 1 de 3	64

FIGURA 34:	Questionário socioeconômico página 2 de 3	65
FIGURA 35:	Questionário socioeconômico página 3 de 3	66

LISTA DE TABELAS

TABELA 1:	Representações gráficas e breve descrição de cada uma	14
TABELA 2:	Estrutura do dataset contendo nome dos atributos e breve descrição de cada um	16
TABELA 3:	Atributos utilizados na representação do gráfico Bar Chart	19
TABELA 4:	Atributos utilizados na representação do gráfico Bubble Chart	21
TABELA 5:	Atributos utilizados na representação do gráfico Donut Chart	23
TABELA 6:	Atributos utilizados na representação do gráfico Dual Scale D3 Bar Chart	25
TABELA 7:	Atributos utilizados na representação do gráfico Grouped Bar Chart	25
TABELA 8:	Respostas corretas para cada questão do formulário	63

LISTA DE ABREVIACÕES

ANSI	<i>American National Standards Institute</i>
CEFET-RJ	Centro Federal de Educação Tecnológica Celso Suckow da Fonseca
CEP	Código de Endereçamento Postal
CPF	Cadastro de Pessoa Física
CSS	<i>Cascading Style Sheets</i>
CSV	<i>Comma-separated Values</i>
D3	<i>Data-Driven Documents</i>
ENEM	Exame Nacional do Ensino Médio
HTML	<i>HyperText Markup Language</i>
JSON	<i>JavaScript Object Notation</i>
PHP	<i>PHP: Hypertext Preprocessor</i>
QUIS	<i>Questionnaire for User Interaction Satisfaction</i>
SGBD	Sistema Gerenciador de Banco de Dados
SQL	<i>Structured Query Language</i>
SUMI	<i>Software Usability Measurement Inventory</i>
SUS	<i>System Usability Scale</i>
SVG	<i>Scalable Vector Graphics</i>
TCC	Trabalho de Conclusão de Curso
TSV	<i>Tab-separated Values</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>EXtensible Markup Language</i>

Capítulo 1

Introdução

O avanço tecnológico acelerado, a ubiquidade da computação e a quantidade maciça de dados sendo gerados constantemente acarreta em pontos positivos e negativos. A computação vem permitindo uma mais-valia na vida das pessoas. Contudo, a ineficiência humana de acompanhar e processar tamanho volume de informação diminui consideravelmente o valor destas informações.[7] [8]

Com o avanço da tecnologia, o volume de dados gerados diariamente pelas pessoas tem aumentado. Hoje em dia já é comum uma pessoa ter mais de um aparelho celular, um tablet ou qualquer outro dispositivo inteligente que se conecte à internet.

A partir do início do século XXI a quantidade de dados gerados e armazenados cresce exponencialmente, sendo esses dos mais variados tipos de atributos. Tentar transmitir todo esse volume tão complexo de dados através de métodos convencionais pode ser ineficiente. Por isso, há uma necessidade de um estudo mais aplicado à amostragem das informações obtidas a partir desses dados.

Segundo um relatório da empresa Cisco, em 2010 o número de dispositivos conectados à internet era de 10 bilhões, o que já ultrapassava o número de pessoas do planeta. Esse mesmo relatório estima que em 2020, esse número seja de 50 bilhões [9].

Ao utilizar redes sociais, seus usuários estão diariamente gerando dados. No Facebook, a cada 60 segundos 510 mil comentários são feitos, 293 mil status são atualizados e 136 mil fotos são enviadas [10].

Contudo esses exemplos correspondem a dados gerados por pessoas e representam apenas uma parcela dos dados existentes. Máquinas, sensores, dispositivos inteligentes, também geram dados constantemente e muitos desses dados são de interesse para pesquisa [10].

Apenas os dados em si e a grande quantidade de dados existentes, podem por sua vez acabar dificultando o processo de extração das informações referentes a estes dados. Números e palavras soltas, por exemplo, podem num primeiro momento não fazer muito sentido. Porém ao serem agrupados de acordo com algum critério ou com alguma característica em comum, podem começar a transmitir informações relevantes.

A visualização de dados procura facilitar esse processo, pois permite exibir de forma mais clara e concisa as informações contidas nos dados e pode ser usada para diferentes áreas, não se limitando apenas à área da computação. Ela permite agrupar os dados de acordo com critérios escolhidos e em seguida extrair as informações que se deseja exibir. A exibição é feita geralmente na forma de gráficos, podendo ser feita também na forma de tabelas ou de outras representações.

A visualização de dados, como o próprio nome diz, é responsável apenas pela parte da exibição dos dados, isto é, ela não extrai os dados. Porém ao ser utilizada em conjunto com alguma ferramenta de extração de dados, torna-se uma ferramenta poderosa.

Devido ao grande volume de dados existentes e ao poder que a exibição dos mesmos fornece para auxiliar na extração de informações, optou-se por utilizar uma das ferramentas de visualização de dados existentes. A ferramenta utilizada chama-se *Data-Driven Documents* (D3) [11] [12] e é uma biblioteca desenvolvida em linguagem Javascript.

A motivação do trabalho consiste em analisar um conjunto de dados, também chamado de dataset, de candidatos aos cursos de ensino médio e técnico da unidade Maracanã do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET-RJ) para o ano de 2017 e procurar selecionar atributos que possam ser úteis para extrair informações existentes ou gerar estatísticas, e através da ferramenta D3, exibir esses dados na forma de gráficos, facilitando assim a compreensão das informações.

O dataset utilizado continha informações referentes aos dados dos candidatos como por exemplo nome, data de nascimento, sexo, endereço, Cadastro de Pessoa Física (CPF), curso pretendido, respostas do questionário sócio-econômico. A estrutura do dataset e seus campos serão melhor explicados e detalhados no Capítulo 3.

Armazenou-se o dataset em um sistema de banco de dados MySQL [13] de forma que fosse possível realizar consultas (queries) para obter os resultados de forma mais rápida e eficaz. Esses resultados oriundos das consultas eram convertidos para um formato de arquivo *JavaScript Object Notation* (JSON) [14] e enviados para a ferramenta D3, de modo que a mesma pudesse lê-los e exibi-los na forma de gráficos.

O Capítulo 2 contém explicações acerca das tecnologias e ferramentas utilizadas para o desenvolvimento da aplicação deste projeto. Nesse mesmo capítulo são abordados alguns trabalhos relacionados e que serviram de consulta durante o desenvolvimento deste projeto.

No Capítulo 3 são definidas e explicadas quais representações gráficas foram utilizadas

para exibir os dados e também explica-se quais critérios foram utilizados para selecionar os dados que seriam exibidos e ainda como determinar quais formas de representação gráfica são adequadas para os tipos de dados que se deseja exibir [15] [16] [17].

Também é abordado quais precauções se deve ter ao utilizar um dataset e como ocorre a comunicação entre o dataset, o mecanismo de consulta ao dataset para extrair os resultados e a ferramenta de visualização de dados responsável por selecionar a representação gráfica para aquele resultado retornado pela consulta.

No Capítulo 4 é descrito como foi realizado o processo de avaliação da aplicação e quais métodos e ferramentas foram utilizadas para tal. Neste capítulo explica-se os questionários criados para medir e avaliar o quanto os gráficos haviam sido eficazes para transmitir o conteúdo da informação [18]. Nesses questionários era avaliado o quanto a representação gráfica se mostrava mais eficaz do que a simples exibição dos dados do dataset.

No Capítulo 5 são abordados os trabalhos futuros, além das próprias conclusões e resultados obtidos durante o desenvolvimento do projeto.

Capítulo 2

Fundamentação Teórica

Neste Capítulo 2 serão abordadas as ferramentas, tecnologias e linguagens utilizadas no desenvolvimento do trabalho e seus conceitos e também alguns trabalhos relacionados à área de visualização de dados.

2.1 Ferramentas Utilizadas

Nesta seção serão explicadas as ferramentas, tecnologias e linguagens utilizadas no desenvolvimento do trabalho.

2.1.1 PHP

Segundo o site oficial do PHP, o nome PHP é um acrônimo recursivo para *PHP: Hypertext Preprocessor* (PHP) e é uma linguagem de script de código aberto, cujo código é executada no lado do servidor [19].

O PHP não se limita apenas ao desenvolvimento WEB, embora seja muito utilizado neste meio. O código PHP pode ser facilmente embutido nas páginas *HyperText Markup Language* (HTML), fornecendo assim mais recursos e ferramentas e evitando que as páginas sejam estáticas. Com isso é possível também gerar páginas HTML dinamicamente no servidor e enviá-las ao cliente [19] [20].

Para executar scripts em PHP é preciso ter o interpretador do PHP instalado na máquina, um servidor WEB compatível - para a realização deste trabalho foi utilizado o servidor Apache Web Server, versão 2.4 - e um navegador. Os scripts serão interpretados e depois executados no servidor WEB e em seguida seus resultados poderão ser visualizados através do navegador [20].

A versão do PHP utilizada no trabalho foi a versão 7 e os scripts foram utilizados para fazer o acesso ao banco de dados MySQL, processar os resultados e convertê-los no formato JSON para enviá-los à ferramenta D3.

2.1.2 MySQL

O MySQL [13] é um Sistema Gerenciador de Banco de Dados (SGBD) que faz uso da linguagem *Structured Query Language* (SQL) ou Linguagem de Consulta Estruturada. Ele foi comprado pela Sun Microsystems e pertence atualmente à Oracle, pois a mesma comprou a Sun Microsystems.

Grandes empresas como Facebook, Google e Adobe utilizam o MySQL para armazenar seus dados [21] e o mesmo possui versões pagas e também possui a versão gratuita chamada Community Edition. Para o projeto foi utilizada a versão Community do MySQL Server.

2.1.3 SQL

O nome SQL é um acrônimo para *Structured Query Language* ou Linguagem de Consulta Estruturada e é uma linguagem de consulta a banco de dados relacionais. É uma linguagem declarativa, onde sua sintaxe determina a forma do resultado a ser obtido e não o método de obtê-lo.

A autoridade *American National Standards Institute* (ANSI) [22] é responsável dentre outras coisas por estabelecer padrões para linguagens de programação. Contudo por existir diversos bancos de dados relacionais de diferentes fabricantes, alguns bancos de dados possuem sintaxes específicas durante a execução de algumas consultas (queries) SQL.

2.1.4 Biblioteca D3

A biblioteca D3 [11] [12] utilizada para representar os dados através de gráficos é uma biblioteca JavaScript para manipular documentos baseados em dados e que faz uso de elementos HTML, *Scalable Vector Graphics* (SVG) e *Cascading Style Sheets* (CSS). Por seguir os padrões WEB, ela roda nos browsers modernos oferecendo sua funcionalidade completa.

Os gráficos e elementos visuais são gerados dinamicamente através de suas funções que leem e manipulam os dados de um documento. Esse documento é geralmente um arquivo no formato JSON (JavaScript Object Notation), *Tab-separated Values* (TSV) ou *Comma-separated Values* (CSV).

2.1.5 JSON

O JSON é uma formatação para transmissão de dados no formato texto e está baseado em um subconjunto da linguagem JavaScript. Os dados representados no formato JSON podem ser de três tipos primitivos: numérico, booleano e caractere (String). Com isso é possível ainda representar vetores (arrays) e objetos [14].

2.1.6 Conversão para JSON

A conversão do resultado obtido pelo acesso à base por meio da consulta executada pelo script PHP é feita através da função PHP chamada `json_encode($var)`. Nessa função é passado um parâmetro que contém o conteúdo que deseja-se converter para o formato JSON.

2.1.7 TSV

O formato TSV, assim como o formato JSON, também é um formato para transmissão de dados. Neste formato, os dados são delimitados através do caractere de tabulação. Os dados de um arquivo TSV são representados na forma de uma tabela e cada linha do arquivo representa uma linha da tabela [23].

2.1.8 CSV

O formato CSV se assemelha ao formato TSV e também se destina à transmissão de dados. Neste formato o diferencial é que os dados são delimitados através do caractere vírgula (em inglês comma). Os dados de um arquivo CSV são representados também na forma de uma tabela e cada linha desse arquivo equivale à uma linha da tabela [24].

2.1.9 SVG

O padrão SVG criado pela *World Wide Web Consortium* (W3C) é baseado no formato *Extensible Markup Language* (XML) e serve para definir imagens e gráficos bidimensionais, estáticos ou dinâmicos, com a vantagem de não perder a qualidade da resolução ao ampliá-la. O navegador interpreta as instruções contidas no SVG e assim consegue converter as informações que o SVG contém em imagens [25] [26].

2.1.10 Armazenamento dos Dados

Os dados da instituição de ensino, contendo informações dos alunos, professores e cursos, estão armazenados em datasets. Os datasets são arquivos que armazenam dados e seguem uma estrutura, de modo que seja possível identificar cada dado específico no momento em que a aplicação realiza a leitura do mesmo. Estes arquivos estão armazenados no MySQL.

2.2 Conceitos

Nesta seção são explicados conceitos utilizados para a visualização de dados. Esses conceitos envolvem seleção e mapeamento de atributos e representações gráficas para as diversas formas de visualizações existentes.

2.2.1 Mapeamento dos Dados

Patrick Lundblad explica que toda seleção de um gráfico para uma representação começa com um fator comum: o dado. Ele define então o que chama de três pilares para o mapeamento de dados.

O primeiro pilar engloba a questão dos tipos dos atributos dos dados [15]. O segundo pilar engloba a questão das formas de exibição desses dados, como a utilização das cores e das formas [16]. O terceiro pilar é responsável por agrupar e categorizar as formas de representação existentes e ajudar no processo de seleção da representação adequada à visualização dos dados [17].

PRIMEIRO PILAR

Quando se trata de atributos de dados, existem dois tipos: dados quantitativos e dados qualitativos. Dados quantitativos são valores numéricos em uma escala ascendente. Dados qualitativos se referem a valores que não podem ser medidos numericamente, mas podem ser descritos através da linguagem [15].

Dentre os fatores existentes na escolha de um gráfico para a visualização dos dados estão: os tipos de dados - quantitativos ou qualitativos - e o modelo daqueles dados - se são dados de temperatura, dados temporais ou dados financeiros, por exemplo.

SEGUNDO PILAR

No segundo pilar é abordado a forma mais adequada de exibir os dados usando cores, for-

mas, tamanhos e posição. Segundo uma pesquisa realizada por William S. Cleveland e Robert McGill foi concluído que pessoas têm diferentes percepções de visualizações. Os pesquisadores estabeleceram através de testes, uma série de orientações baseadas em qual marcador visual é mais preciso [16].

Dentre as opções básicas de exibição, destacam-se: posição, comprimento, ângulo, área, volume, diferença, cor, saturação de cor, contraste e textura. Cada uma dessas opções possui uma ordem de precisão, fazendo com que a escolha de uma dessas opções seja mais adequada perante outra [16].

A figura 1 [1] demonstra as principais opções de exibição utilizadas em representações gráficas: posição, comprimento, ângulo/declive, área, volume, diferença, tonalidade, saturação, contraste e textura.

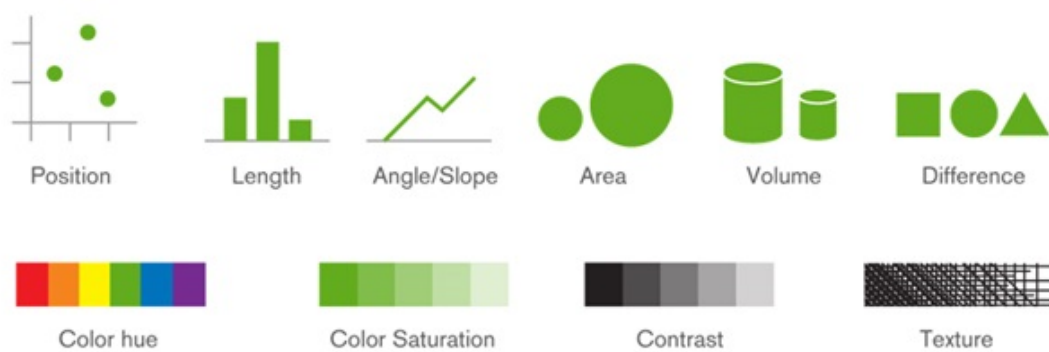


Figura 1: Opções básicas de exibição. Figura obtida de [1]

Alguns pontos importantes a se considerar: posição é a opção mais adequada para dados nominais e formas também são úteis para esse tipo de dados; para dados ordinais, o contraste das cores ajuda a destacar a ordem dos valores em um ranking; para dados de intervalos, a forma mais adequada de exibição é por meio de posição ou comprimento [16].

TERCEIRO PILAR

Pode-se agrupar as formas de representações gráficas em diferentes categorias. Para tornar mais simples, foram definidas as 4 seguintes categorias: comparação, composição, distribuição e relacionamento.

COMPARAÇÃO

As representações gráficas dessa categoria estão relacionadas ao tempo e ao tamanho dos dados. São usadas para comparar a magnitude dos valores entre os elementos e encontrar o valor mínimo e máximo. Podem ser usadas para comprar valores atuais com valores antigos de

modo a observar se houve um crescimento ou decrescimento dos valores com a passagem de tempo [17].

COMPOSIÇÃO

As representações gráficas dessa categoria referem-se a dados que mudam ao longo do tempo ou incluem dados estatísticos. São usados para exibir como a parte se compara com o todo e como o valor total pode ser dividido em partes [17].

DISTRIBUIÇÃO

As representações gráficas dessa categoria são usadas para exibir como os valores quantitativos se distribuem ao longo do eixo do menor ao maior valor. Através da forma do dado, é possível extrair características como faixa de valores e tendência central [17].

RELACIONAMENTO

As representações gráficas dessa categoria são usadas para identificar relacionamentos existentes entre os dados para encontrar correlações entre eles [17].

A figura 2 [2] mostra um resumo dos tipos de representação existentes agrupados pelas 4 categorias: comparação, composição, distribuição e relacionamento.

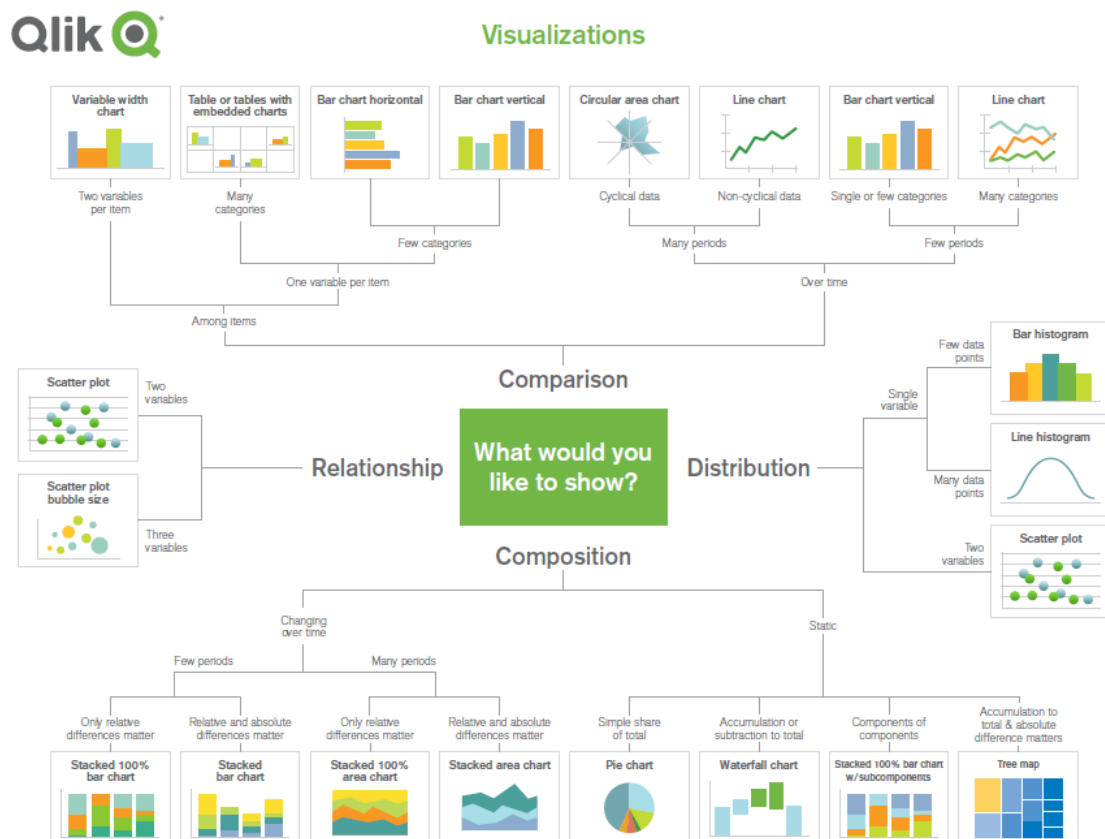


Figura 2: Tipos de representação agrupados pelas 4 categorias: comparação, composição, distribuição e relacionamento. Figura obtida de [2]

2.3 Trabalhos Relacionados

Esta seção aborda trabalhos relacionados ao tema de visualização de dados.

2.3.1 The Ignorance Project

O professor de Saúde Internacional da universidade Karolinska Institutet da Suécia, Hans Rosling, obteve resultados interessantes ao realizar um pequeno questionário com seus alunos. O questionário pedia para que os alunos selecionassem um país de cada um dos cinco pares de países nos quais eles acreditavam que a taxa de mortalidade infantil fosse o dobro da taxa do outro país [27].

Através deste questionário, o professor Rosling observou que o resultado do teste foi pior do que o esperado. O professor brinca ainda que seus estudantes se saíram piores do que os chimpanzés, pois se os nomes dos países estivessem escritos em 5 pares de bananas, os chimpanzés acertariam em média 50% devido ao fato de eles escolherem as bananas aleatoriamente [27].

A conclusão obtida pelo professor foi que a maioria das pessoas não tem ideia do que acontece ao redor do mundo. As pessoas possuem ideias preconcebidas e por isso se saíram piores neste teste do que se tivessem respondido as questões de forma arbitrária. O resultado dessas ideias preconcebidas, segundo Hans Rosling, é oriundo de três fatores: viés pessoal, materiais e fontes de aprendizado desatualizados e a mídia - jornais e noticiários, por exemplo [28].

O viés pessoal consiste no ambiente onde cada pessoa vive e foi criada. Este ambiente não oferece uma visão geral do mundo. Nele é possível ter uma visão local.

Os materiais e fontes de aprendizado desatualizados consistem nos materiais usados em escolas e universidades e a metodologia de ensino das escolas. Não existe uma preocupação em atualizar o conteúdo dos materiais impressos e nem de atualizar os métodos de ensino. Isto é, os professores aprenderam algo quando estavam estudando e hoje eles ensinam o mesmo e do mesmo jeito [28].

A mídia também é responsável, pois ela seleciona as notícias e acontecimentos que se tornarão manchetes. Normalmente acontecimentos incomuns são os mais selecionados, pois são mais interessantes.

Visando combater os problemas gerados por esses três vieses, o professor Hans Rosling e seu filho Ola Rosling iniciaram o projeto The Ignorance Project [28] que pertence à empresa

deles Gapminder [29]. O projeto consiste em elaborar questionários com perguntas relacionadas a desenvolvimento global e medir o índice de acertos e erros das pessoas.

Além do projeto, a empresa desenvolveu o software Trendalyzer que é responsável por converter dados estatísticos em gráficos interativos. O software, de código aberto, possui diversas representações e modelos de gráficos. A empresa Gapminder [29] procura também incentivar o uso e a difusão do Trendalyzer por professores em escolas.

O professor Hans Rosling, apesar de possuir formação relacionada à área de saúde, é conhecido por seus trabalhos envolvendo a área de visualização de dados e por meio de seus trabalhos demonstra que a visualização de dados nos auxilia na análise e compreensão de informações.

2.3.2 Fatores da Percepção Visual Humana na Visualização de Dados

Em seu artigo, os autores Dulclerci Alexandre e João Tavares explicam que os métodos tradicionais para a visualização e a análise dos dados existentes não se mostram suficientes e eficientes.

A visualização dos dados explora a percepção humana e a rapidez da capacidade do processamento visual humano, fornecendo apoio cognitivo para uma melhor compreensão das informações existentes e que estão sendo transmitidas pelos dados [7].

Segundo eles a forma como os humanos reagem às visualizações influencia no entendimento dos dados e de sua informação. Com isso esses fatores humanos contribuem no processo de visualização e devem ser levados em consideração na construção de ferramentas adequadas à visualização de dados.

As características das imagens e dos objetos influenciam na percepção humana. Essas características podem ser utilizadas de forma a se obter um maior proveito das visualizações geradas, destacando informações ou induzindo o expectador a perceber propositadamente determinadas informações [7].

Os autores explicam como as etapas de análise das informações provenientes das visualizações de dados ocorrem na mente humana e como o ser humano utiliza sua capacidade de percepção visual para processar essas informações e gerar conclusões. Para ser efetiva, uma visualização deve transmitir as informações de forma rápida, clara e objetiva, sem induzir a erros, atendendo às capacidades da percepção da visão humana [7].

2.3.3 VISDADOSENEM

Em seu Trabalho de Conclusão de Curso (TCC), os alunos César Henrique Cardoso dos Santos e Maykon Carlos de Freitas desenvolveram uma aplicação web nomeada VISDADOSE-NEM. Tal aplicação utiliza a biblioteca D3 com os dados do Exame Nacional do Ensino Médio (ENEM), para gerar gráficos dinâmicos e facilitar o processo de análise dos dados e tomada de decisões [8].

Durante o trabalho deles foi estudado sobre as técnicas de representações visuais, caracterização de dados e visualização de dados. Através dessas técnicas os estudantes foram capazes de gerar gráficos com propósito específico para cada tipo de informação que desejavam obter [8]. Algumas das ferramentas usadas foram similares as utilizadas neste trabalho: linguagem PHP, banco de dados MySQL, biblioteca D3 e o formato JSON. Porém no trabalho deles foi utilizado também as bibliotecas NVD3 que é uma biblioteca baseada na D3 que utiliza gráficos prontos e reutilizáveis.

Neste Capítulo 2 foram explicadas as ferramentas, tecnologias e linguagens utilizadas na realização da aplicação e também foram explicados conceitos de mapeamento e escolha de atributos para a visualização de dados além de trabalhos relacionados com a área de visualização de dados. O Capítulo 3 aborda como o desenvolvimento da aplicação ocorreu e quais artefatos foram produzidos.

Capítulo 3

Desenvolvimento

Neste Capítulo 3 serão abordados os temas relativos à definição das formas de representação, onde será explicado quais representações foram escolhidas, dentre os exemplos existentes e quais critérios foram usados para determinar quais representações seriam utilizadas.

Explica-se também a estrutura do dataset, quais atributos foram escolhidos e quais critérios foram utilizados para sua escolha. É abordado também precauções que se devem ter quando se trabalha com dados armazenados em datasets.

Em seguida são descritas as representações utilizadas, quais tipos de atributos melhor se adequam à mesma e quais atributos foram selecionados para que a mesma os exibisse. Para cada representação é exibida uma imagem para ilustrar como os dados foram representados graficamente pela mesma.

Por fim procurou-se explicar como as consultas à base são processadas e como a comunicação entre o resultado obtido pelas consultas e as páginas que fazem uso da ferramenta D3 para exibi-los ocorre.

3.1 Definição das Representações

Para exibir os dados retornados pelas consultas realizadas aos dataset, foram escolhidas 6 representações, onde apenas campos relevantes do dataset foram utilizados. Além das 6 representações, um menu interativo [30] - chamado Bubble Menu - que leva para cada uma das representações foi escolhido.

O D3, conforme já explicado no Capítulo 2, pode ler arquivos em formato como JSON, CSV e TSV e através de funções próprias, gerar gráficos, elementos visuais e animações.

No site oficial do D3 [11], além da documentação contendo explicações acerca das funções próprias da ferramenta, existem diversos exemplos de código aberto [31]. Dentre os exemplos existentes, procurou-se aqueles que melhor se adequavam ao contexto do projeto.

A tabela 1 informa os exemplos escolhidos e uma breve descrição de cada um.

Nome da visualização	Descrição
Bar Chart	Gráfico de barras para representar número de ocorrências de 1 único atributo.
Bubble Chart	Gráfico cujos elementos são representados na forma de bolhas. Também é usado para 1 único atributo.
Bubble Menu	Menu da aplicação para navegar entre as páginas.
Donut Chart	Gráfico similar ao gráfico de pizza para representar número de ocorrências de 1 atributo.
Dual Scale D3 Bar Chart	Gráfico de barras para número de ocorrências de 2 atributos.
Grouped Bar Chart	Gráfico de barras para número de ocorrências de grupos de atributos.

Tabela 1: Representações gráficas e breve descrição de cada uma

3.2 Critérios de Escolha de Atributos

Em seguida foram selecionados os campos do dataset que seriam mais relevantes para a visualização dos dados. Por exemplo: o campo cpf não é interessante, pois é único e com isso em um gráfico de eixo (x,y), o eixo x teria n colunas, onde n seria igual a quantidade de candidatos do dataset e o eixo y possuiria o mesmo valor, que seria o valor 1, pois o cpf é único.

Com isso, campos cuja distribuição da visualização se assemelhe ao exemplo do CPF, não são desejados. Para o projeto foram buscados campos cujos valores do eixo x não fossem tão grandes e os valores do eixo y não fossem tão pequenos ou tão próximos.

Por fim buscou-se para cada gráfico o campo ou o conjunto de campos que poderiam gerar uma visualização relevante, pois conforme explicado, nem todos os atributos existentes no dataset são interessantes para se gerar alguma visualização. Atributos como CPF, telefone e email, por exemplo, não são interessantes.

Conforme já mencionado no Capítulo 2, o processo para a escolha dos atributos e também para a escolha das representações gráficas segue o que Patrick Lundblad chama de três pilares para o mapeamento de dados para visualizações [15] [16] [17].

3.3 Precauções

Foi necessário normalizar o dataset, pois alguns campos possuíam diferentes formas de grafia para a mesma palavra e inclusive possuíam palavras escritas com erros de digitação. Isso deve ter ocorrido, pois o formulário de inscrição dos candidatos na internet devia possuir campos onde o próprio candidato digitava as informações ao invés de selecioná-las. Isso resultou em

situações onde a cidade Rio de Janeiro aparecia como “RIO DE JANEIRO” ou “rio de janeiro” e também com erros de digitação, como por exemplo “Rio de Janiero”.

Caso não fosse feita normalização no dataset, o resultado da consulta não seria preciso, pois consideraria as entradas “Rio de Janeiro” e “Rio de Janiero” como duas cidades diferentes.

Para minimizar esse tipo de problema, o método de coleta dos dados deveria ser melhor elaborado. Para esse exemplo o candidato poderia informar seu Código de Endereçamento Postal (CEP) e a aplicação realizaria uma consulta à uma base de dados para obter outras informações como os nomes da cidade, da rua e do bairro. Com isso impediria que o candidato informasse valores inválidos ou mesmo valores válidos, porém com erros de digitação.

3.4 Estrutura e Organização do Dataset

O dataset apresenta ao todo 78 atributos para cada candidato do concurso, onde 33 dos 78 atributos correspondem às respostas das questões do questionário socioeconômico. A tabela 2 contém os 78 atributos e uma descrição do que cada atributo significa.

3.5 Visualizações

Nesta sessão serão abordadas as formas de visualizações de dados utilizadas na aplicação. Serão explicados o menu e os gráficos utilizados e os atributos selecionados para cada tipo de visualização.

3.5.1 Bubble Menu

O Bubble Menu [30] é um modelo de visualização de links que os representa na forma de bolhas. Cada bolha representa um link para uma página e ao clicar numa bolha, se é redirecionado para a página apontada por aquele link. Esse modelo foi escolhido para representar o menu da aplicação e através dele é possível navegar por todos os gráficos da mesma.

A figura 3 ilustra o Bubble Menu sem que alguma bolha seja selecionada para exibir seus gráficos.

Ao posicionar o cursor do mouse sobre uma das bolhas que representam os gráficos, a mesma é destacada das demais e exibe suas formas de representações existentes por meio de bolhas de tamanho inferior. Cada uma dessas bolhas menores é um link para uma página web

Tabela 2: Estrutura do dataset contendo nome dos atributos e breve descrição de cada um

Campo	Descrição
inscricao	número de inscrição do candidato.
candidato	nome do candidato.
cargo_ID	id do curso.
cargo	nome do curso.
cpf	CPF do candidato.
rg	RG do candidato.
sexo	sexo do candidato.
etnia	etnia do candidato.
nascimento	data de nascimento do candidato.
tipo_logradouro	tipo do logradouro do endereço.
endereço	endereço do candidato.
numero	número da casa/prédio.
complemento	complemento da casa/prédio.
bairro	bairro do candidato.
cidade	cidade do candidato.
cep	CEP do candidato.
estado	Estado do candidato.
ddd_telefone	DDD do telefone.
telefone	telefone do candidato.
ddd_celular	DDD do celular.
celular	celular do candidato.
senha	senha de inscrição do candidato.
lembrete	campo vazio.
email	e-mail do candidato.
email2	segundo e-mail do candidato.
mae	nome da mãe do candidato.
pne	se é portador de necessidades especiais.
cid	classificação internacional de doença.
nis	número de identificação social do candidato.
descricao	descrição do cid.
org_emissor	nome do órgão emissor.
UFemissor	nome da UF emissora.
uf_emissor	código da UF emissora.
pai	nome do pai do candidato.
nacionalidade	nacionalidade do candidato.
naturalidade	naturalidade do candidato.
estado_civil	estado civil do candidato.
sangue	tipo sanguíneo do candidato.
datahora	data da inscrição do candidato para o concurso.
n_eleitor	número de eleitor do candidato.
s_eleitor	seção eleitoral do candidato.
z_eleitor	zona eleitoral do candidato.
ensino	se o candidato está na cota de ensino.
renda	se o candidato está na cota de renda.
autodeclaracao	se o candidato está na cota de autodeclaração.
q01 a q33	respostas das questões 01 a 33 do questionário socioeconômico.

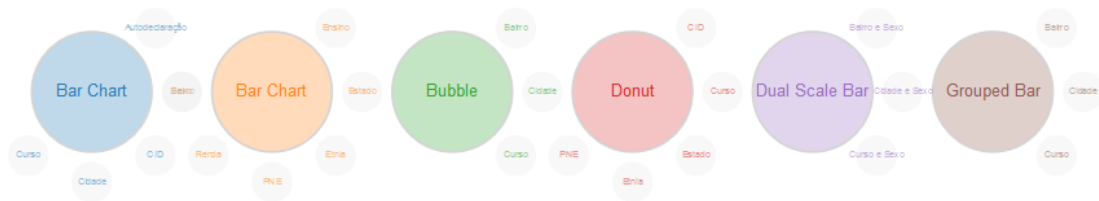


Figura 3: Bubble Menu sem bolha selecionada

que contém a representação gráfica selecionada.

A figura 4 ilustra o Bubble Menu quando a bolha Bubble Chart é destacada das demais e quando o cursor se posiciona sobre sua bolha inferior “Bairro”. Note que na parte inferior da página é também exibida uma descrição da representação gráfica selecionada, contendo os atributos utilizados na mesma.



Figura 4: Bubble Menu com bolha Bubble Chart - Bairro selecionada

O Bubble Menu recebe os dados de um arquivo JSON contendo as informações a respeito das bolhas maiores que representam os tipos de representações gráficas - Bar Chart, Bubble Chart, Donut Chart, por exemplo - e das bolhas menores - Bairro, Cidade, Estado, por exemplo - que representam cada uma das representações gráficas geradas. As bolhas maiores possuem nome, descrição e suas próprias bolhas menores que pertencem ao tipo de representação gráfica da bolha maior. Cada bolha menor possui nome e um link para a página da representação gráfica.

3.5.2 Escolha do Bubble Menu

Assim como todo tipo de representação gráfica passa por um processo seletivo para determinar qual representação é a mais adequada para o objetivo e para os atributos que se deseja exibir, a escolha do Bubble Menu [30] para representar a página inicial e o menu da aplicação ocorreu devido ao fato de o mesmo já ser uma representação gráfica, isto é, uma forma de visualização.

Por o trabalho se tratar de visualização de dados e representá-los visualmente e graficamente, achou-se oportuna a adoção desse exemplo - Bubble Menu - como página inicial da aplicação. O mesmo é também uma forma de chamar a atenção do usuário e de tornar a navegação mais convidativa e interativa.

3.5.3 Bar Chart

O gráfico Bar Chart [32] corresponde ao gráfico em barras e para o projeto foi representado apenas um campo. No eixo x localizam-se os valores daquele campo e no eixo y a quantidade de ocorrência para cada um dos valores. Para esse gráfico não é interessante selecionar atributos que possuam um número muito grande de valores, pois a sua exibição no eixo x seria difícil de representar, devido à extensão dos valores. Isso poderia comprometer a legibilidade e simplicidade da visualização [15] [16] [17].

Para esse tipo de gráfico foram selecionados 11 atributos, implicando na geração de 11 gráficos do tipo Bar Chart. Os 11 atributos são demonstrados na tabela 3.

Para cada um dos atributos foi representada a ocorrência de candidatos por atributo. Por exemplo: para o atributo cidade, foi representado no gráfico a ocorrência de candidatos para aquela cidade.

A figura 5 ilustra uma das 11 representações para o gráfico Bar Chart. Nessa representação foi selecionado o atributo etnia para exibir o número de ocorrências de etnias dentro do conjunto de candidatos aos cursos da unidade Maracanã.

3.5.4 Bubble Chart

O gráfico Bubble Chart [33] representa cada elemento na forma de uma bolha ou círculo. O tamanho da bolha é proporcional ao tamanho do elemento. Por exemplo: se dois elementos A e B forem representados por meio desse gráfico, e A possui valor ou tamanho igual a 4x e B

Tabela 3: Atributos utilizados na representação do gráfico Bar Chart

Nome do atributo	Descrição
autodeclaracao	O gráfico exibe a quantidade de candidatos que se enquadram e que não se enquadram na categoria de cota de auto-declaração.
bairro	O gráfico exibe a quantidade de candidatos por cada bairro existente no dataset.
cargo	O gráfico exibe a quantidade de candidatos por curso da unidade Maracanã.
cidade	O gráfico exibe a quantidade de candidatos por cada cidade existente no dataset.
cid	O gráfico exibe a quantidade de candidatos que possuem e que não possuem alguma doença do quadro CID existente no dataset.
ensino	O gráfico exibe a quantidade de candidatos que se enquadram e que não se enquadram na categoria de cota por ensino em rede pública.
estado	O gráfico exibe a quantidade de candidatos por cada Estado existente no dataset.
etnia	O gráfico exibe a quantidade de cada etnia de acordo com as etnias dos candidatos do dataset.
pne	O gráfico exibe a quantidade de candidatos que são portadores de necessidades especiais e de candidatos que não são portadores de necessidades especiais.
renda	O gráfico exibe a quantidade de candidatos que se enquadram e que não se enquadram na categoria de cota por renda.
sexo	O gráfico exibe a quantidade de candidatos do sexo masculino e do sexo feminino do concurso para a unidade Maracanã.

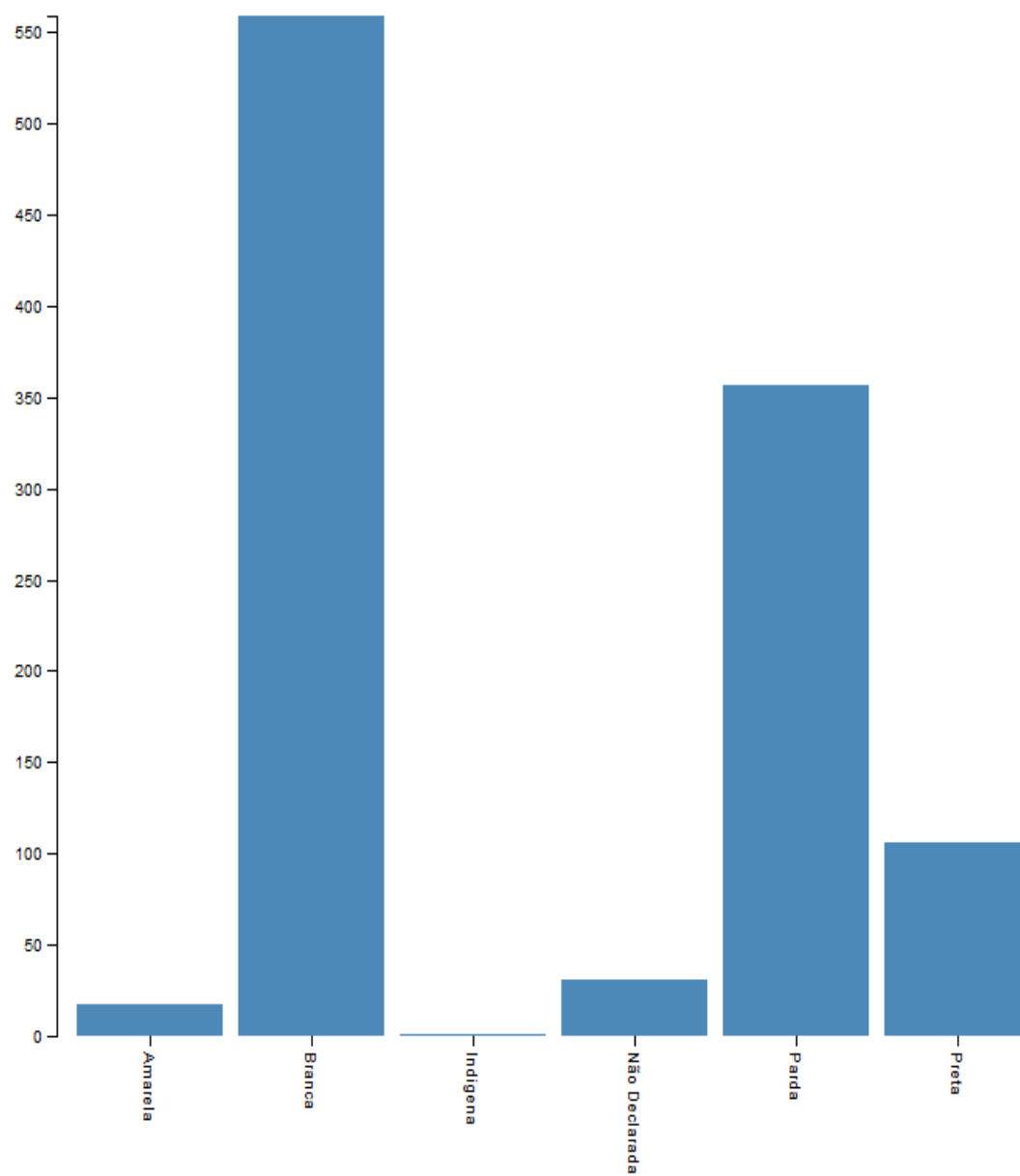


Figura 5: Gráfico Bar Chart para representar número de ocorrências para o atributo etnia na unidade Maracanã

possui valor ou tamanho igual a x , o elemento A será uma bolha de tamanho 4 vezes maior que o tamanho da bolha correspondente ao elemento B.

Esse gráfico, assim como o Bar Chart, utiliza apenas um atributo e para este atributo é exibido o número de ocorrências. Um dos pontos que diferem esse gráfico do Bar Chart é o fato de os elementos não estarem representados seguindo o eixo (x,y) . Esse gráfico se mostra interessante para atributos que possuam um número grande de valores, pois como os elementos não são representados seguindo o eixo (x,y) , a representação gráfica não fica comprometida.

Dos atributos existentes no dataset, foram selecionados 3 atributos, pois estes possuíam diversos valores e devido a este motivo, sua representação pelo Bubble Chart foi mais interessante que pelo Bar Chart. Para cada um dos atributos selecionados foi gerado um gráfico contendo o número de ocorrência de candidatos para cada valor existente. A tabela 4 demonstra os atributos utilizados.

Tabela 4: Atributos utilizados na representação do gráfico Bubble Chart

Nome do atributo	Descrição
bairro	O gráfico exibe a quantidade de candidatos por cada bairro existente no dataset.
cargo	O gráfico exibe a quantidade de candidatos para cada curso da unidade Maracanã.
cidade	O gráfico exibe a quantidade de candidatos por cada cidade existente no dataset.

A figura 6 ilustra uma das três representações geradas através do gráfico Bubble Chart. Nessa representação, o atributo selecionado foi o atributo bairro e por meio dela é exibido o número de candidatos aos cursos da unidade Maracanã por cada bairro existente no dataset.

3.5.5 Donut Chart

O Donut Chart [34] é similar ao gráfico de pizza, onde os elementos são representados na forma de fatias e cuja soma de todas as fatias forma um círculo. O que o diferencia do gráfico de pizza é que o Donut Chart possui um espaço em branco em seu centro.

Para cada representação foi selecionado apenas 1 atributo. Cada ocorrência daquele atributo é exibida na forma de fatia. A tabela 5 demonstra os atributos selecionados para a representação por meio do Donut Chart.

A figura 7 ilustra uma das 5 representações geradas por meio do gráfico Donut Chart. Nessa representação foi selecionado o atributo cargo e com isso foi exibida a quantidade de candidatos

Tabela 5: Atributos utilizados na representação do gráfico Donut Chart

Nome do atributo	Descrição
cargo	O gráfico exibe todos os cursos ofertados na unidade Maracanã, onde cada curso ocupa uma fatia do círculo e o tamanho dessa fatia é referente à quantidade de candidatos que selecionaram esse curso como curso desejado no momento da inscrição.
cid	O gráfico exibe os códigos internacional de doença (CID) existentes no dataset junto com a ausência do código - representando os candidatos que não possuem doença. Cada CID e sua ausência ocuparão uma fatia do círculo e o tamanho dessa fatia é referente à quantidade de candidatos que possuem ou não alguma doença.
estado	O gráfico exibe todos os Estados existentes no dataset, onde cada Estado ocupará uma fatia do círculo e o tamanho dessa fatia é referente à quantidade de candidatos daquele Estado.
etnia	O gráfico exibe todas as etnias existentes no dataset, onde cada etnia ocupará uma fatia do círculo e o tamanho dessa fatia é referente à quantidade de candidatos daquela etnia.
pne	O gráfico exibirá a quantidade de candidatos portadores de necessidades especiais (PNE), onde cada fatia possui tamanho referente à condição de possuir e de não possuir alguma necessidade especial.

por curso ofertado na unidade Maracanã. Cada curso ocupa uma fatia do círculo e o tamanho de cada fatia é proporcional à quantidade de candidatos que selecionaram o curso dessa fatia como curso desejado no momento da inscrição. Observe que os valores das fatias são superiores às quantidades de vagas dos cursos ofertados, pois esses valores não correspondem aos candidatos aprovados.

3.5.6 Dual Scale D3 Bar Chart

O gráfico Dual Scale D3 Bar Chart [35] é muito similar ao Bar Chart, contendo apenas um diferencial. Nele é possível representar 2 atributos ao invés de apenas um. Ele também faz uso do eixo (x,y) e procura-se seguir os critérios de seleção de atributos usados no Bar Chart.

Para um desses gráficos decidiu-se representar a distribuição de candidatos por curso, agrupados por sexo. Isto é, o número de ocorrências de candidatos do sexo masculino e de candidatos do sexo feminino para cada curso. A tabela 6 demonstra os atributos utilizados em cada gráfico e uma descrição a respeito da representação da visualização por meio dos atributos selecionados.



Figura 7: Gráfico Donut Chart para representar a quantidade de candidatos por curso desejado na unidade Maracanã

Tabela 6: Atributos utilizados na representação do gráfico Dual Scale D3 Bar Chart

Atributo principal	Atributo 1	Descrição
sexo	bairro	O gráfico exibe a quantidade de candidatos por cada bairro existente no dataset, onde essa quantidade está agrupada por sexo; isto é, exibe a quantidade de candidatos do sexo masculino e do sexo feminino por cada bairro do dataset.
sexo	cargo	O gráfico exibe a quantidade de candidatos por cada curso da unidade Maracanã, onde essa quantidade está agrupada por sexo; isto é, exibe a quantidade de candidatos do sexo masculino e do sexo feminino por cada curso da unidade Maracanã.
sexo	cidade	O gráfico exibe a quantidade de candidatos por cada cidade existente no dataset, onde essa quantidade está agrupada por sexo; isto é, exibe a quantidade de candidatos do sexo masculino e do sexo feminino por cada cidade do dataset.

A figura 8 ilustra uma das três representações geradas por meio do gráfico Dual Scale D3 Bar Chart. Nessa representação foram selecionados os atributos sexo e cargo e com isso foi possível exibir a quantidade de candidatos do sexo masculino e do sexo feminino para cada curso da unidade Maracanã.

3.5.7 Grouped Bar Chart

O Grouped Bar Chart [36] assim como o Bar Chart e o Dual Scale D3 Bar Chart é um gráfico de barras. Ele permite haver um número variado de elementos a serem comparados. Seleciona-se primeiro um atributo e para este atributo, seleciona-se outros atributos dos quais deseja-se verificar o número de ocorrências.

Para esse gráfico foram escolhidas 3 combinações para gerar as visualizações. A tabela 7 ilustra as combinações que contém os 4 atributos.

Tabela 7: Atributos utilizados na representação do gráfico Grouped Bar Chart

Atributo principal	Atributo 1	Atributo 2	Atributo 3
cargo	autodeclaracao	ensino	renda
bairro	autodeclaracao	ensino	renda
cidade	autodeclaracao	ensino	renda

O atributo principal significa o atributo que representará cada agrupamento do eixo x. Para a combinação cargo, autodeclaracao, ensino e renda, o gráfico irá exibir os demais valores de ocorrências dos 3 atributos autodeclaracao, ensino e renda para cada valor existente no atributo

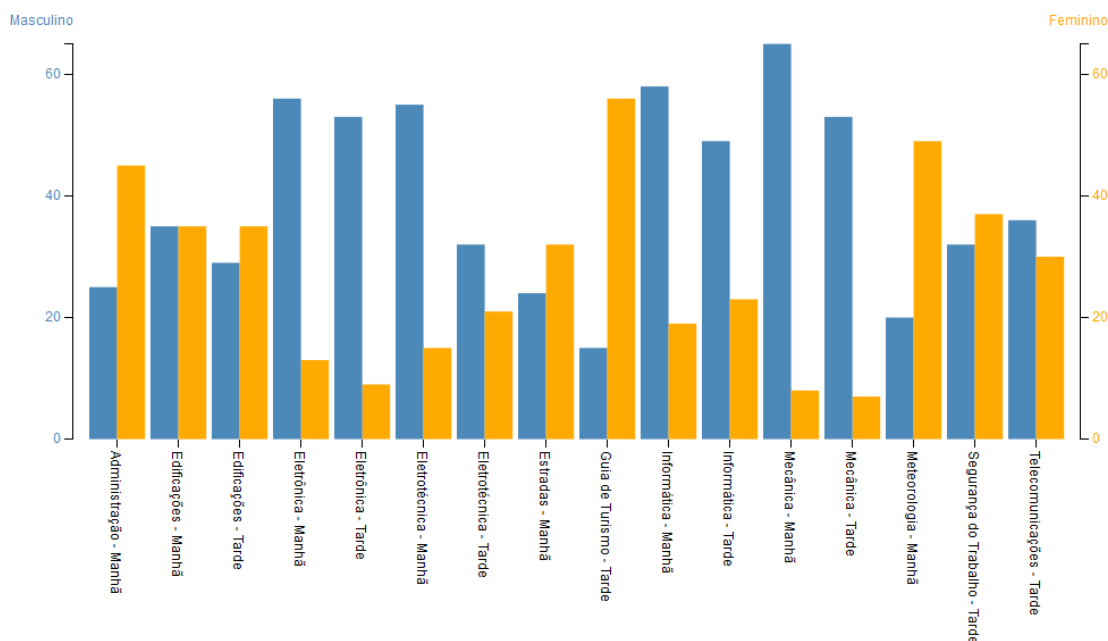


Figura 8: Gráfico Dual Scale D3 Bar Chart usado para representar a quantidade de candidatos do sexo masculino e do sexo feminino por curso da unidade Maracanã

cargo. Entre outras palavras, exibe o número de candidatos que se enquadram na categoria autodeclaração, na categoria ensino e na categoria renda - note que são situações independentes. Ou seja, um candidato pode se enquadrar numa categoria e não se enquadrar em outra - de cada curso ofertado pelo concurso.

O mesmo raciocínio é válido para as 2 demais combinações. Isto é, exibe o número de candidatos que se enquadram nas categorias autodeclaração, ensino e renda - novamente são situações independentes - para cada bairro e cidade existente no dataset.

A figura 9 ilustra uma das 3 representações geradas por meio do gráfico Grouped Bar Chart. Nessa representação os atributos autodeclaracao, ensino e renda foram agrupados por cargo. Com isso foram exibidos para cada curso da unidade Maracanã o número de candidatos que se enquadram e que não se enquadram nas categorias de cota de autodeclaração, ensino e renda.

3.6 Processamento das Consultas

Para cada gráfico existe um arquivo PHP correspondente. No arquivo PHP existem funções responsáveis por fazer a conexão com o banco de dados MySQL e realizar a consulta que retorna o resultado específico para aquela visualização.

Com isso, ao clicar em um dos links do Bubble Menu [30], o arquivo PHP correspondente é chamado e o mesmo irá acessar o banco de dados, realizando a query SQL salva em sua função.

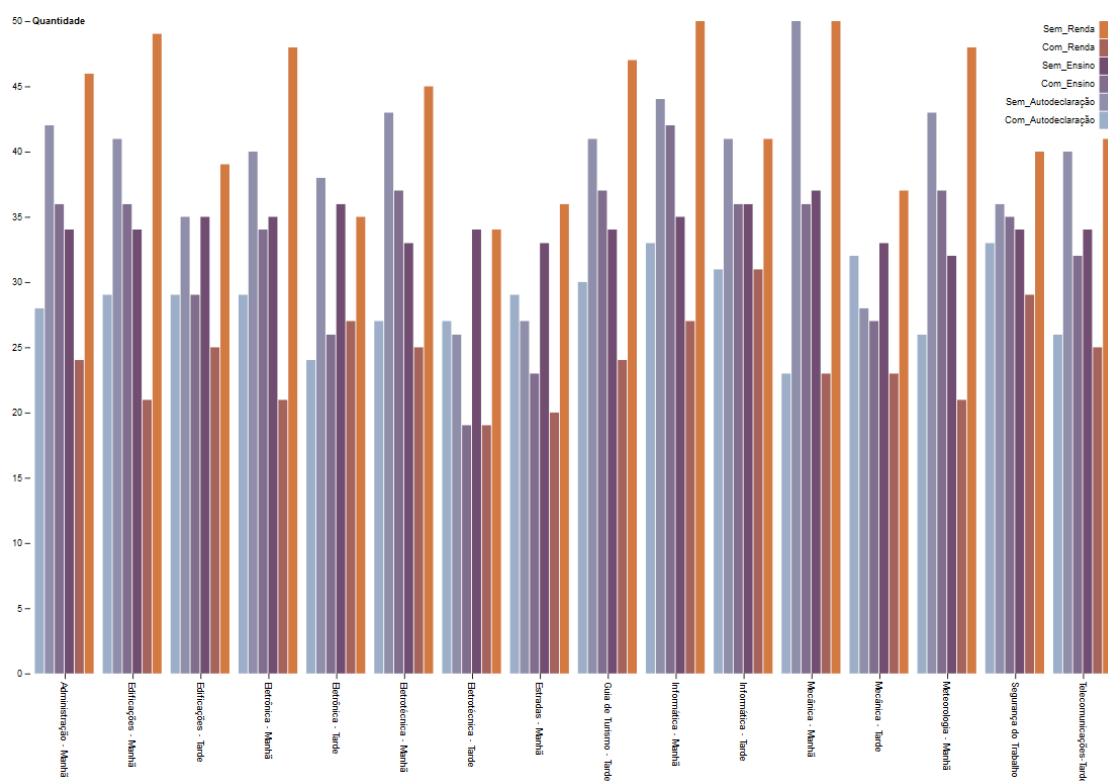


Figura 9: Gráfico Grouped Bar Chart para representar o número de candidatos que se enquadram e que não se enquadram em categorias de cotas agrupando por cursos da unidade Maracanã

3.7 Comunicação com a Ferramenta D3

Como dito anteriormente, cada gráfico utiliza a ferramenta D3 para ser exibido e cada gráfico possui também um arquivo PHP correspondente. É por meio do arquivo PHP que a ferramenta D3 consegue obter os dados selecionados do banco pela query.

A ferramenta D3 recebe dados de fontes externas e alguns dos formatos suportados dessas fontes são o JSON, o CSV e o TSV. Nesse projeto foi utilizado o formato JSON e por isso em alguns dos exemplos, foram necessárias algumas modificações, pois os formatos que eles liam eram o CSV e o TSV.

O arquivo PHP faz uso de uma função nativa que é responsável por fazer a conversão para JSON. Com isso ele lê a base de dados e faz a chamada a essa função nativa. Em seguida envia os dados já convertidos em JSON para o arquivo onde a ferramenta D3 é utilizada para a representação dos dados.

A figura 10 ilustra a conversão para JSON dos valores armazenados na variável \$resp pela função nativa do PHP.

```
// Indica que o formato é Json
header("Content-type:application/json");

// Imprime o Json na tela
echo json_encode($resp);
```

Figura 10: Conversão de conteúdo armazenado na variável em PHP para o formato JSON

Neste Capítulo 3 foram abordadas as formas de visualizações de dados utilizadas na aplicação e foi explicado para cada uma delas quais atributos foram selecionados e quais critérios foram utilizados para a seleção. No final do Capítulo 3 foi explicado como é feita a comunicação entre os dados selecionados da base de dados e as páginas em PHP responsáveis por conter o código JavaScript das formas de visualizações de dados.

No Capítulo 4 é explicado como foi feito o processo de avaliação da aplicação e quais metodologias e ferramentas foram utilizadas para realizar tal procedimento.

Capítulo 4

Avaliação Experimental

Neste Capítulo 4 será descrito o processo realizado com o objetivo de avaliar a aplicação desenvolvida. Para o processo de avaliação foi utilizado um tipo de formulário - o formulário que utiliza o método *System Usability Scale* (SUS) [37] [18] - e foi elaborado juntamente um roteiro para que os usuários seguissem ao utilizar a aplicação e ao final poderem responder as perguntas do formulário.

O Capítulo 4 ainda explica a importância de medir ou quantificar a usabilidade de uma aplicação, quais técnicas podem ser utilizadas para tal medição, o que é o método SUS, qual a sua utilidade e porque foi decidido utilizá-lo durante o processo de avaliação da aplicação.

4.1 Usabilidade

Medir a usabilidade de uma aplicação, seja ela um website, um software, ou um aplicativo para dispositivos móveis, por exemplo, nem sempre é algo trivial. Geralmente é necessário realizar testes com os usuários que utilizam tal aplicação e para isso é necessário desenvolver roteiros que determinem passos a serem seguidos pelos usuários durante os testes e ao final realizar uma pesquisa com os usuários para obter um retorno dos mesmos com relação ao grau de satisfação que os mesmos obtiveram ao usar a aplicação [18].

Saber o nível de usabilidade de uma aplicação é muito importante e se possível é recomendado procurar medi-la ainda durante o processo de desenvolvimento, pois uma vez finalizada, mudanças na aplicação se tornam mais trabalhosas do que se realizadas durante o processo de desenvolvimento.

Pode se definir a usabilidade como o nível de facilidade em que o usuário tem ao se utilizar a aplicação. Uma aplicação complexa e que requer muito esforço do usuário para utilizá-la ou ainda para aprender a utilizá-la, pode acabar desmotivando o mesmo a usá-la e com isso levar a aplicação ao fracasso [18].

Para se medir o grau de usabilidade foram desenvolvidas técnicas e métodos, dentre eles: *System Usability Scale* (SUS), *Software Usability Measurement Inventory* (SUMI), *Question-*

naire for User Interaction Satisfaction (QUIS).

O método adotado para o processo de avaliação da aplicação desenvolvida no projeto foi o System Usability Scale (SUS) devido à sua simplicidade. O SUS será explicado na seção seguinte.

4.2 Método Usability Scale (SUS)

O método SUS é um dos métodos existentes para se avaliar o grau de usabilidade de aplicações fazendo uso de escalas numéricas. O SUS foi criado por John Brooke em 1986 e consiste em perguntas ou afirmações a serem respondidas ou julgadas pelos usuários de modo a avaliar a utilização da aplicação [37].

As perguntas são simples e não devem ser extensas a fim de não tomar muito tempo dos usuários ao responderem o formulário. Isso também é um ponto importante a ser considerado, pois formulários extensos e complexos tendem a desmotivar o usuário a respondê-lo por completo e isso pode afetar o resultado da medição da usabilidade.

O SUS avalia a efetividade da aplicação, isto é, se os usuários conseguem realizar suas tarefas de modo a alcançar o objetivo, a eficiência, isto é, quanto esforço foi necessário investir para realizar as tarefas, a satisfação do usuário, isto é, se o mesmo conseguiu realizar as tarefas, alcançando os objetivos, obtendo uma experiência satisfatória ao utilizar a aplicação [18].

O formulário do método SUS contém 10 questões ou afirmações, onde cada uma é respondida em uma escala de 1 a 5, onde a escala 1 significa que o usuário discorda completamente com a pergunta ou afirmação e a escala 5 significa que o usuário concorda completamente com a pergunta ou afirmação.

A figura 11 [3] ilustra um modelo de formulário que utiliza o método SUS, contendo 10 perguntas ou afirmativas e para cada uma delas existe uma resposta que varia em uma escala de 1 a 5 pontos.

A figura 12 [4] demonstra como é feito o cálculo para o mesmo formulário da figura 11 após o mesmo ter sido preenchido.

4.2.1 Como Aplicar o Método SUS

O formulário deve ser aplicado após os usuários terem executado um grupo de tarefas. Essas tarefas podem ser definidas por meio de um roteiro que descreve os testes que serão realizados

		Strongly Disagree				Strongly Agree
1.	I think that I would like to use this mobile app frequently.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	I found this mobile app unnecessarily complex.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	I thought this mobile app was easy to use.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.	I think that I would need assistance to be able to use this mobile app.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.	I found the various functions in this mobile app were well integrated.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6.	I thought there was too much inconsistency in this mobile app.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7.	I would imagine that most people would learn to use this mobile app very quickly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8.	I found this mobile app very cumbersome/awkward to use.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9.	I felt very confident using this mobile app.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.	I needed to learn a lot of things before I could get going with this mobile app.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figura 11: Exemplo de formulário utilizando o método SUS. Figura obtida de [3]

		Strongly Disagree			Strongly Agree	Scale Position	Calculation	Score Contribution
1.	I think that I would like to use this mobile app frequently.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	4	4-1	3
2.	I found this mobile app unnecessarily complex.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2	5-2	3
3.	I thought this mobile app was easy to use.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	5	5-1	4
4.	I think that I would need assistance to be able to use this mobile app.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1	5-1	4
5.	I found the various functions in this mobile app were well integrated.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	4	4-1	3
6.	I thought there was too much inconsistency in this mobile app.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2	5-2	3
7.	I would imagine that most people would learn to use this mobile app very quickly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	5	5-1	4
8.	I found this mobile app very cumbersome/awkward to use.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1	5-1	4
9.	I felt very confident using this mobile app.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	4	4-1	3
10.	I needed to learn a lot of things before I could get going with this mobile app.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1	5-1	4
TOTAL								35
x 2.5								87.5

Figura 12: Cálculo da pontuação de um formulário utilizando o método SUS. Figura obtida de [4]

para uma determinada funcionalidade, por exemplo.

A pontuação final, isto é, a soma obtida pela pontuação de cada uma das 10 perguntas, varia de 0 a 100 pontos e a média do SUS é de 68 pontos. Resultados abaixo de 68 pontos indicam que a usabilidade da aplicação pode estar comprometida e que seria interessante realizar mudanças no design da mesma.

O cálculo da pontuação final não consiste apenas em somar a pontuação individual das perguntas. Antes disso é realizado um processo para cada pergunta, onde nesse processo será efetuada uma conta para cada pontuação individual.

Para cada questão ímpar (1, 3, 5, 7 e 9) subtraia 1 do valor e para cada questão par (2, 4, 6, 8 e 10) subtraia 5 do valor, ignorando o sinal. Por exemplo: para a questão 3, se o usuário responder 3 soma-se 2, pois subtraindo 1 de 3 é igual a 2 e se o usuário responder 4 soma-se 3, pois subtraindo 1 de 4 é igual a 3. Para a questão 4, se o usuário responder 4 soma-se 1, pois subtraindo 5 de 4 e ignorando o sinal é igual a 1 e se o usuário responder 3 soma-se 2, pois subtraindo 5 de 3 e ignorando o sinal é igual a 2. Ao final é realizado um somatório das pontuações individuais e por fim multiplica-se o somatório por 2,5 [38].

A figura 13 [5] demonstra os resultados das pontuações finais obtidos ao se utilizar o método SUS.

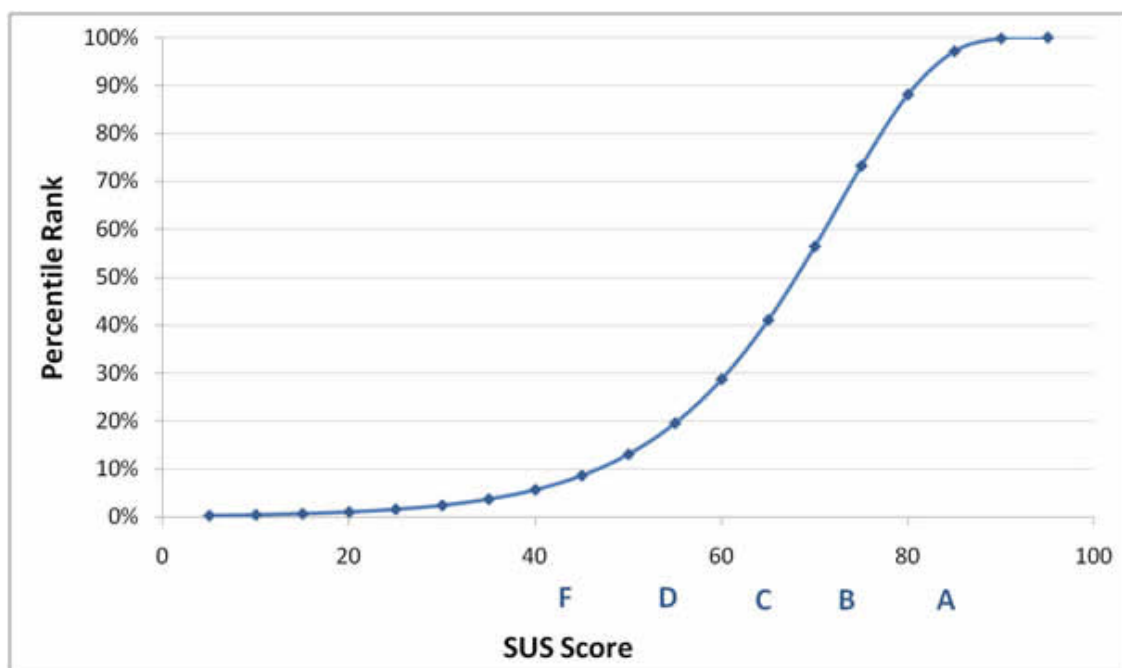


Figura 13: Pontuações finais obtidas pelo método SUS. Figura obtida de [5]

O eixo x indica a pontuação final obtida e o eixo y indica a porcentagem de ocorrência daquela pontuação. Note que a média (50%) de ocorrências consiste numa pontuação próxima

de 70 pontos - aproximadamente 68 pontos. E na parte inferior do gráfico é ainda exibido um grau para a pontuação obtida, variando de A a F. O grau A equivale a um resultado muito bom, onde a pontuação final é igual ou superior a 80 pontos e o grau F equivale a um resultado muito ruim, onde a pontuação final é igual ou inferior a 50 pontos.

A figura 14 [6] demonstra um exemplo para os resultados obtidos após aplicar o formulário para 50 usuários.

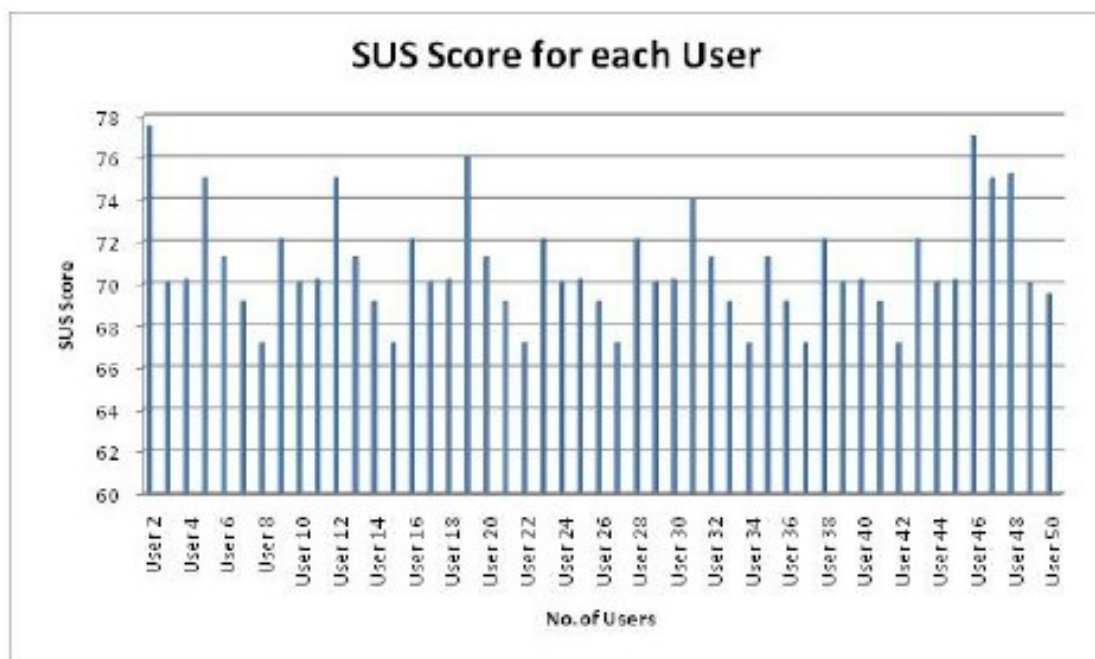


Figura 14: Resultado das pontuações finais de 50 usuários usando o método SUS. Figura obtida de [6]

4.2.2 Resultados Obtidos ao Aplicar o SUS no Projeto

Para aplicar o método SUS durante o processo de avaliação da aplicação foi criado um formulário simples no Google Forms contendo as seguintes afirmações a serem avaliadas:

1. Eu gostaria de usar essa aplicação frequentemente.
2. Eu achei essa aplicação desnecessariamente complexa.
3. Eu achei essa aplicação fácil de utilizar.
4. Eu acho que eu precisaria de ajuda para utilizar essa aplicação.
5. Eu achei as funcionalidades dessa aplicação bem integradas.

6. Eu achei que a aplicação contém muitas inconsistências.
7. Eu imagino que a maioria das pessoas aprenderia a utilizar essa aplicação rapidamente.
8. Eu achei essa aplicação muito estranha de se utilizar.
9. Eu me senti muito confiante ao utilizar a aplicação.
10. Eu preciso aprender muitas coisas antes de começar a utilizar essa aplicação.

Foi solicitado aos usuários responder o formulário somente após realizar as tarefas definidas no roteiro de avaliação. O formulário criado que utiliza a metodologia SUS se encontra no Apêndice B deste documento. O roteiro criado será explicado na próxima seção.

Ao final do período de avaliação realizou-se a coleta dos resultados e com isso se obteve o seguinte resultado final, demonstrado pela figura 15.

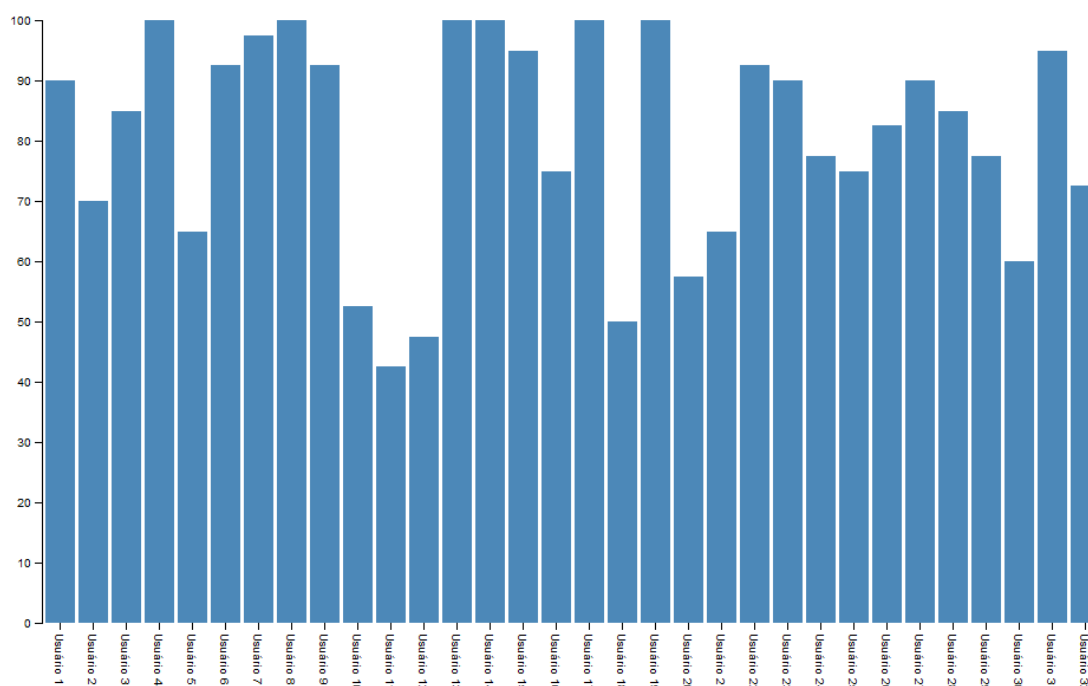


Figura 15: Resultado das pontuações finais de usuários usando o método SUS para a aplicação desenvolvida

Ao todo 32 usuários responderam o formulário e como se pode observar na figura 15 as duas pontuações finais máximas foram de 100 e 97,5 pontos e as duas pontuações finais mínimas foram de 42,5 e 47,5 pontos.

No total 8 candidatos atribuíram pontuação final abaixo da média - isto é, abaixo de 68 pontos. Ao calcular a média de todas as pontuações finais dos 32 candidatos o valor aproximado

obtido foi de 80,47 pontos. Esse valor atribui o grau A à aplicação, indicando que a mesma está acima da média de 68 pontos.

A figura 16 demonstra o número de ocorrências de cada uma das notas - de 1 a 5 - em cada uma das 10 questões respondidas pelos 32 usuários. Para essa representação utilizou-se o gráfico Grouped Bar Chart.

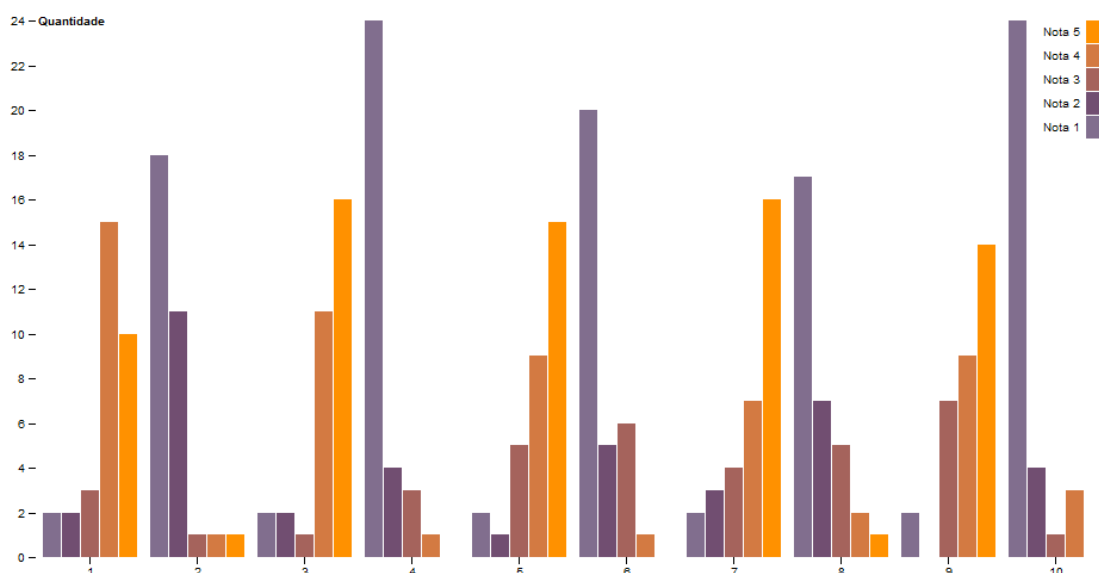


Figura 16: Número de ocorrências de cada nota por questão do formulário SUS para os 32 usuários

Após a realização do experimento e da coleta dos dados, os resultados obtidos possibilitaram uma melhor compreensão acerca do grau de usabilidade da aplicação e com isso pode-se concluir que a mesma não apresenta dificuldades aos usuários quanto à sua utilização.

4.3 Formulário de Avaliação da Transmissão de Informações

Juntamente com o formulário que utiliza a metodologia SUS foi elaborado um segundo formulário a fim de avaliar a qualidade das informações transmitidas pelas representações gráficas.

Esse formulário contém perguntas acerca de interpretação gráfica, onde o usuário deveria selecionar a alternativa correta referente àquela pergunta.

Foram elaboradas 13 perguntas no total, procurando abranger cada um dos tipos de representações gráficas utilizados: Bar Chart, Bubble Chart, Donut Chart, Dual Scale D3 Bar Chart e Grouped Bar Chart.

Para responder as perguntas os usuários precisavam utilizar a aplicação e selecionar a representação gráfica indicada a fim de visualizar os dados e obter as informações para aquela

pergunta.

Procurou-se elaborar questões simples e rápidas de serem respondidas e que ao mesmo tempo pudessem se mostrar úteis na avaliação da transmissão das informações por meio da visualização de dados.

Por meio dos resultados obtidos pode-se constatar que não houve dificuldade em responder as perguntas e também não houve problemas em analisar e extrair as informações que as representações gráficas exibem.

O formulário e as respostas dos usuários que realizaram o experimento encontram-se no Apêndice C e no Apêndice D deste documento.

4.4 Roteiro de Avaliação da Aplicação

Para avaliar a aplicação hospedou-se a mesma em um servidor web destinado à hospedagem gratuita para que os usuários pudessem acessá-la remotamente e seguir os passos descritos pelo roteiro. Ao final foi solicitado aos usuários que respondessem um breve formulário contendo 10 afirmações, seguindo a metodologia SUS [18], conforme explicado na seção anterior. Foi solicitado também aos usuários que respondessem um segundo formulário com perguntas relativas à interpretação dos gráficos a fim de avaliar a qualidade das informações transmitidas pelas representações gráficas.

O roteiro completo encontra-se no Apêndice A deste documento e o mesmo foi escrito de modo que o usuário realizasse uma sequência de passos a fim de alcançar os objetivos descritos no próprio roteiro.

Os objetivos visavam familiarizar o usuário com a aplicação, fazendo com que o mesmo navegasse pela mesma e buscasse compreender seu funcionamento. Contudo esse não era o único objetivo dos passos do roteiro.

Posteriormente era solicitado ao usuário para que o mesmo encontrasse a exibição adequada para um ou mais atributos especificados. Inicialmente isso era feito informando explicitamente o nome do tipo da visualização e solicitando ao usuário que a exibisse. Em outros passos o nome da visualização e do seu tipo já não eram mais informados e cabia ao usuário decidir qual ou quais visualizações existentes seriam adequadas para a tarefa do roteiro, pois nesse momento o usuário já estaria familiarizado com os tipos de visualizações que a aplicação dispõem.

Ao final, caso o usuário ainda não tivesse selecionado alguma visualização, ele era encorajado a selecioná-la para ver e entender seu funcionamento. Ao terminar de executar todos os

passos do roteiro, o usuário deveria preencher o formulário, cujo endereço estava indicado no próprio roteiro da aplicação.

Neste Capítulo 4 foi abordado o processo de avaliação da aplicação e foram explicadas quais metodologias e formas de avaliações foram utilizadas. Em seguida foram apresentados os resultados obtidos por meio da avaliação onde os usuários deveriam seguir os passos definidos no roteiro e avaliar as afirmativas do formulário que utiliza a metodologia SUS e responder as perguntas do formulário acerca das informações transmitidas pelas representações gráficas.

Capítulo 5

Conclusão

Conforme explicado a visualização de dados é uma área em que pode ser utilizada em conjunto com diversas outras, pois seu objetivo principal é a exibição das informações contidas nos dados. A exibição dessas informações dá-se muitas vezes na forma de gráficos e estes ainda se subdividem em diversos tipos, como por exemplo os tipos usados na aplicação desenvolvida neste trabalho: gráfico de barras [32], gráfico de donut [34] - uma variação do gráfico de pizza, gráfico de bolhas [33], gráfico de barras agrupadas [36] e outros tipos existentes, porém não utilizados no trabalho.

Ao exibir os dados é preciso antes realizar um processo de seleção dos dados, isto é definir quais atributos serão exibidos, pois nem sempre é possível exibir todos os dados de uma só vez. Isso geraria uma poluição visual e comprometeria o entendimento das informações transmitidas pela representação gráfica.

Esse processo de seleção de dados foi explicado no Capítulo 2 e no Capítulo 3, no qual foram explicados os chamados 3 pilares para o mapeamento de dados para visualizações, definidos por Patrick Lundblad [15] [16] [17].

Após a escolha dos atributos ainda é preciso também definir qual forma de representação será utilizada para gerar a visualização. Esse processo, também explicado no Capítulo 3, deve levar em consideração a quantidade de atributos envolvidos e também a complexidade e os tipos de dados que esses atributos possuem [15] [16] [17].

Ao escolher a forma de representação adequada para o grupo de atributos selecionados dos quais deseja-se exibir as informações, o processo de compreensão dessas informações é facilitado, pois o ser humano utiliza suas habilidades perceptivas para captar as informações e o uso da visualização de dados maximiza esse processo [7].

O ser humano é capaz de pegar os dados brutos existentes em um dataset ou em uma planilha, por exemplo, e sozinho extrair as informações de que precisa, selecionando os atributos necessários. Porém dependendo do volume de dados e de sua complexidade, essa tarefa pode deixar de ser viável ou pode levar muito tempo para ser executada [7].

Utilizar uma aplicação que será responsável por realizar a busca dos dados necessários e

agrupá-los para gerar os resultados e estatísticas desejadas, torna essa tarefa em algo simples e relativamente rápido de ser executado - dependendo do volume dos dados envolvidos. A visualização de dados funciona então como uma ferramenta adicional que irá utilizar os resultados e estatísticas obtidos pelas consultas realizadas pela aplicação para representá-los visualmente e facilitar o processo de compreensão das informações existentes naqueles resultados, possibilitando também processos como tomada de decisões e análise e acompanhamento temporal dos resultados.

Após realizar a avaliação experimental da aplicação desenvolvida no trabalho com os usuários por meio de um roteiro de passos e um formulário que utiliza o método SUS [18], explicado no Capítulo 4, pôde-se confirmar que a visualização de dados mostra-se uma ferramenta eficiente para facilitar o processo de percepção e compreensão dos dados.

Aplicar a visualização de dados no dataset de candidatos aos cursos da unidade Maracanã do CEFET-RJ, possibilitou gerar estatísticas com relação aos cursos mais procurados, ao sexo, à localidade e à renda dos candidatos, por exemplo.

Isso pode-ser aplicado para possibilitar um maior entendimento com relação à procura dos cursos do CEFET-RJ e ainda pode ser usado para realizar um acompanhamento temporal, envolvendo dados coletados de concursos anteriores, por exemplo.

Portanto utilizar a visualização de dados auxilia o processo de extração de informações e estatísticas, sendo fundamental escolher a representação gráfica adequada dos dados que se deseja exibir para que isso não afete o processo da percepção visual humana [7].

5.1 Problemas e Dificuldades Encontrados

Conforme explicado no Capítulo 3 foi necessário normalizar a base de dados dos candidatos, pois dados como bairro e cidade continham valores inválidos provenientes de erros de digitação. Exemplos como esse comprometem o resultado da visualização dos dados, pois nem sempre é simples de implementar uma lógica no software que trate erros de digitação, de modo que considere algumas entradas como a mesma.

Um exemplo prático foi o bairro Rio Comprido, em que houve uma ocorrência em que um candidato escreveu Rio Cumprido. Caso não fossem realizadas consultas à base em busca dos bairros e das cidades para normalizá-los a fim de corrigir erros de digitação e de ortografia, o software consideraria para esse exemplo a existência de 2 bairros e para o bairro Rio Cumprido existiria apenas 1 candidato.

Para contornar esse problema o método de coleta dos dados dos candidatos deve ser reformulado a fim de eliminar inconsistências na base e evitar o trabalho de normalizá-la. Para o exemplo de coleta de dados de candidatos de um concurso, onde os mesmos informam dados como nome, endereço, idade, sexo, curso desejado é preferível usar alguma lógica para obter os dados de campos que sejam derivados de outros ao invés de permitir que o usuário informe os dados para esses campos também. Isto é, os campos de endereço, bairro e cidade, por exemplo, podem ser obtidos através de uma lógica implementada que busque suas informações no momento em que o usuário informar o seu CEP.

Esse simples procedimento já reduz o trabalho de normalizar a base para 3 atributos que são derivados do atributo cep. Esse tipo de procedimento pode ser aplicado para outros tipos de atributos. Outro problema encontrado diz respeito à escolha da representação gráfica e dos atributos que serão utilizados por ela. Conforme explicado no Capítulo 2, Patrick Lundblad definiu 3 pilares para mapear dados para visualizações [15] [16] [17]. Ele explica que o tipo do dado influencia na escolha da visualização, pois seu tipo pode comprometer a visualização e também a transmissão de suas informações como explica o trabalho acerca da compreensão das informações por meio das capacidades cognitivas e perceptivas humanas [7].

Esse problema foi observado na escolha do gráfico Donut Chart para representar a ocorrência de candidatos por bairro. Devido à grande quantidade de bairros, a escolha do atributo bairro para esse tipo de representação não foi bem sucedida. A figura 17 exibe o gráfico Donut Chart para a ocorrência de candidatos por bairro e como é possível observar, o label, isto é, o rótulo onde fica a descrição com o nome do bairro, sobrepõe a área de bairros adjacentes, comprometendo a legibilidade da visualização.

A biblioteca D3 conforme explicado no Capítulo 2 e no Capítulo 3 utiliza a linguagem JavaScript e possui funções próprias, além das funções nativas da linguagem JavaScript. A dificuldade encontrada ao se utilizar a biblioteca D3 foi compreender o comportamento próprio do D3, pois o mesmo também utiliza elementos como o SVG para representar figuras e vetores matematicamente e aprender um pouco mais da parte avançada do JavaScript. O que se observou com relação ao D3 foi que grande parte do material disponível para aprendizado acerca de seu funcionamento encontra-se em seu site oficial. O site oficial possui muitos exemplos de visualizações de dados e também a documentação oficial dividida pelas versões existentes do D3. Contudo foi difícil encontrar muitas informações fora do site oficial.

Conhecer bem a linguagem JavaScript ajuda bastante a entender o funcionamento do D3,

porém não é suficiente. Ainda é preciso recorrer à documentação oficial e aos exemplos do site para se ter uma melhor compreensão de sua utilização.

5.2 Trabalhos Futuros

Este trabalho utilizou um dataset que continha apenas os dados do concurso realizado no ano de 2016 para os candidatos que desejavam ingressar nos cursos das unidades do CEFET-RJ no ano de 2017.

Deste dataset foi utilizado e analisado apenas os dados referentes à unidade Maracanã. É interessante realizar uma análise para todas as unidades do CEFET-RJ existentes no dataset, podendo assim gerar outras possibilidades de comparações.

Outro ponto interessante e que não pôde ser abordado é a utilização de dados de concursos de anos anteriores, de modo a gerar uma análise temporal dos resultados de todas as unidades do CEFET-RJ e realizar assim um acompanhamento.

O dataset contém ainda 33 questões de um questionário socioeconômico que se encontra anexado ao Apêndice E deste trabalho. É relevante e também interessante realizar uma análise dos dados dessas 33 questões respondidas pelos candidatos, pois isso geraria mais estatísticas e informações sobre os candidatos. Apesar de ter-se utilizado o atributo de renda - para cota - esse atributo apenas definia se o candidato se enquadrava na categoria de renda. Isso não informava a situação econômica do candidato e por isso não era possível extrair tantas informações quanto se fizesse uso do questionário socioeconômico.

Referências Bibliográficas

- [1] Figura com principais opções de exibições utilizadas em representações gráficas. http://global.qlik.com/~media/Images/Blog/entries/posts-large/VisualEncoding-Large_625x349.ashx?h=349&w=625, 2017.
- [2] Figura com resumo dos principais tipos de representações gráficas existentes agrupados por categorias. http://global.qlik.com/~media/Images/Blog/entries/posts-large/GoodChart-Large_625x349.ashx?h=349&w=625, 2017.
- [3] Figura de formulário SUS em branco. <http://satoriinteractive.com/wp-content/uploads/2015/09/blank-mobile-app-sus.jpg>, 2017.
- [4] Figura de formulário SUS preenchido. <http://satoriinteractive.com/wp-content/uploads/2015/09/sample-sus-score.jpg>, 2017.
- [5] Figura para ranking em porcentagem da pontuação final para o formulário SUS. <http://www.uxpamagazine.org/wp-content/uploads/2011/08/10-3-Sauro-Fig2.jpg>, 2017.
- [6] Figura de formulário SUS para pesquisa de 50 usuários. https://www.researchgate.net/profile/Shoba_Chigarapalle/publication/272370711/figure/fig10/AS:281049637638150@1444018863735/Fig-12-SUS-Score-of-50-users.png, 2017.
- [7] João Manuel R. S. Tavares Dulclerci Sternadt Alexandre. Factores da percepção visual humana na visualização de dados. June 2007.
- [8] Maykon Carlos de Freitas César Henrique Cardoso dos Santos. Visdadosenem: Visualização de dados do enem com a biblioteca d3. December 2016.
- [9] Estimativa de número de dispositivos no futuro. <https://www.cnet.com/news/how-much-is-the-internet-of-everything-worth-cisco-says-19-trillion/>, 2017.
- [10] Estatísticas do Facebook. <https://zephoria.com/top-15-valuable-facebook-statistics/>, 2017.
- [11] Site Data-Driven Documents. <https://d3js.org/>, 2017.

- [12] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.
- [13] MySQL. <https://www.mysql.com/>, 2017.
- [14] Definição do formato JSON. <http://www.json.org/json-pt.html>, 2017.
- [15] Primeiro pilar de mapeamento de dados para visualização. <http://global.qlik.com/us/blog/posts/patrik-lundblad/mapping-data-to-visualizations-data-attributes>, 2017.
- [16] Segundo pilar de mapeamento de dados para visualização. <http://global.qlik.com/us/blog/posts/patrik-lundblad/visual-encoding>, 2017.
- [17] Terceiro pilar de mapeamento de dados para visualização. <http://global.qlik.com/fr/blog/posts/patrik-lundblad/third-pillar-of-mapping-data-to-visualizations-usage>, 2017.
- [18] Formulário SUS. <https://brasil.uxdesign.cc/o-que-%C3%A9-o-sus-system-usability-scale-e-como-us%C3%A1-lo-em-seu-site-6d63224481c8>, 2017.
- [19] O que é PHP. https://secure.php.net/manual/pt_BR/intro-what-is.php, 2017.
- [20] O que o PHP pode fazer. https://secure.php.net/manual/pt_BR/intro-whatcando.php, 2017.
- [21] Empresas que usam MySQL. <https://www.mysql.com/why-mysql/>, 2017.
- [22] Site American National Standards Institute. <https://www.ansi.org/>, 2017.
- [23] Definição do formato TSV. <http://www.cs.tut.fi/~jkorpela/TSV.html>, 2017.
- [24] Definição do formato CSV. <https://tools.ietf.org/html/rfc4180>, 2017.
- [25] Explicação do padrão SVG. <https://helpx.adobe.com/br/illustrator/using/svg.html>, 2017.
- [26] Origem do padrão SVG. https://www.w3.org/Graphics/SVG/WG/wiki/Secret-Origin_of_SVG, 2017.

- [27] Palestra sobre pesquisa que pedia aos alunos para categorizar os países pela taxa de natalidade. <http://www.gapminder.org/videos/google-zeitgeist/google-zeitgeist-europe-2008/>, 2017.
- [28] The Ignorance Project. <http://www.gapminder.org/ignorance/>, 2017.
- [29] Empresa Gapminder. <http://www.gapminder.org/>, 2017.
- [30] Bubble Menu. <http://sunsp.net/demo/BubbleMenu/>, 2017.
- [31] Exemplos do D3. <https://github.com/d3/d3/wiki/Gallery>, 2017.
- [32] Bar Chart. <https://bl.ocks.org/mbostock/3885304>, 2017.
- [33] Bubble Chart. <https://bl.ocks.org/mbostock/4063269>, 2017.
- [34] Donut Chart. <https://bl.ocks.org/mbostock/3887193>, 2017.
- [35] Dual Scale D3 Bar Chart. <https://github.com/liuflly/Dual-scale-D3-Bar-Chart>, 2017.
- [36] Grouped Bar Chart. <https://bl.ocks.org/mbostock/3887051>, 2017.
- [37] System Usability Scale (SUS). <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>, 2017.
- [38] Como calcular os resultados do formulário SUS. <http://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-we> 2017.

Apêndice A

Roteiro para Avaliação da Aplicação

Por favor siga os passos desse roteiro e ao final preencha os formulários para avaliar a aplicação.

1. Nessa aplicação existem 5 tipos de representações gráficas: Bar Chart (gráfico em barras), Bubble Chart (gráfico de bolhas), Donut Chart (gráfico de donut, semelhante ao gráfico de pizza), Dual Scale Bar Chart (gráfico de barras com duas escalas) e Grouped Bar Chart (gráfico de barras agrupadas). Escolha uma visualização de dados de cada uma das representações gráficas para exibi-la. Por exemplo: Duas possíveis escolhas de visualização de dados são a de candidatos por curso da representação gráfica Bar Chart e a de candidatos por bairro da representação gráfica Bubble Chart.
2. Agora selecione para o Bar Chart a visualização responsável por exibir os candidatos por cidade e procure identificar qual cidade possui o maior número de candidatos e qual possui o menor.
3. Para o Bubble Chart selecione a visualização responsável por exibir os candidatos por bairro e procure identificar qual bairro possui o maior número de candidatos e qual possui o menor número.
4. Para o Donut Chart selecione a visualização responsável por exibir os candidatos por etnia e procure identificar qual etnia possui o maior número de candidatos e qual possui o menor.
5. Para o Dual Scale Bar Chart selecione a visualização responsável por exibir os candidatos por curso e sexo e procure identificar qual curso possui a maior quantidade de candidatos dos sexos masculino e feminino e qual curso possui a menor quantidade.
6. Para o Grouped Bar Chart selecione a visualização responsável por exibir os candidatos por autodeclaração, ensino, renda e curso e procure identificar para os cursos as taxas de autodeclaração de ensino e de renda e identificar qual curso possui a maior dessas taxas e qual possui a menor.

7. Procure agora selecionar uma ou mais visualizações (se existirem) que você ache adequadas para exibirem a quantidade de candidatos por bairro. Observação: a visualização pode conter ainda outros atributos.
8. Procure agora selecionar uma ou mais visualizações (se existirem) que você ache adequadas para exibirem a quantidade de candidatos por curso. Observação: a visualização pode conter ainda outros atributos.
9. Procure agora selecionar uma ou mais visualizações (se existirem) que você ache adequadas para exibirem a quantidade de candidatos por pne (portador de necessidades especiais). Observação: a visualização pode conter ainda outros atributos.
10. Caso ainda existe alguma visualização que você não tenha exibido, selecione-a para ver como é feita a sua exibição.

Após ter executado esses passos, por favor preencha os dois formulários criados por meio do Google Forms a seguir para avaliar a aplicação. É necessário fazer login com uma conta do Google e isso serve apenas para controlar o número de respostas ao formulário. Nenhuma informação sobre sua conta do Google é coletada. Nós agradecemos por sua colaboração.

Apêndice B

Formulário System Usability Scale (SUS)

Formulário Método System Usability Scale

Formulário para medir o grau de usabilidade do sistema. Por favor avalie 10 as afirmações abaixo utilizando a escala de 1 a 5 para 1 como Discordo Completamente e 5 como Concordo Completamente.

***Obrigatório**

1. 1) Eu gostaria de usar essa aplicação frequentemente. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

2. 2) Eu achei essa aplicação desnecessariamente complexa. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

3. 3) Eu achei essa aplicação fácil de utilizar. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

Figura 18: Formulário System Usability Scale (SUS) página 1 de 3

4. 4) Eu acho que eu precisaria de ajuda para utilizar essa aplicação. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

5. 5) Eu achei as funcionalidades dessa aplicação bem integradas. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

6. 6) Eu achei que a aplicação contém muitas inconsistências. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

7. 7) Eu imagino que a maioria das pessoas aprenderia a utilizar essa aplicação rapidamente. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

8. 8) Eu achei essa aplicação muito estranha de se utilizar. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

Figura 19: Formulário System Usability Scale (SUS) página 2 de 3

9. 9) Eu me senti muito confiante ao utilizar a aplicação. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

10. 10) Eu preciso aprender muitas coisas antes de começar a utilizar essa aplicação. *

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Completamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo Completamente

Com tecnologia



Figura 20: Formulário System Usability Scale (SUS) página 3 de 3

Apêndice C

Formulário de Avaliação da Transmissão de Informações

Gráfico de Barras (Bar Chart)

Dado os Gráficos de Barras do site, responda as seguintes questões. Observe que ao colocar o mouse sobre a barra é exibido a quantidade de candidatos daquela característica.

***Obrigatório**

1. Qual é o Curso mais concorrido e o menos concorrido, respectivamente? *

Marcar apenas uma oval.

- ☐ Informática - Turno Manhã e Meteorologia - Turno Manhã
- ☐ Eletrônica - Turno Tarde e Guia de Turismo - Turno Tarde
- ☐ Administração - Turno Manhã e Edificações - Turno Tarde
- ☐ Informática - Turno Manhã e Eletrotécnica - Turno Tarde

2. O Gráfico de Ensino informa quantos candidatos cursaram em escolas públicas. Após analisar o gráfico, selecione a opção correta sobre o número de candidatos que estudaram em escolas públicas em relação aos que não estudaram. Sendo "sim" para aqueles que estudaram em escola pública e "não" para os que estudaram em escola particular. *

Marcar apenas uma oval.

- ☐ O número de candidatos é o mesmo.
- ☐ Aqueles que estudaram em escola PÚBLICA é ligeiramente maior.
- ☐ Aqueles que estudaram em escola PARTICULAR é ligeiramente maior.
- ☐ Aqueles que estudaram em escola PÚBLICA é muito maior.

3. Qual é a porcentagem mais provável dos candidatos do sexo masculino e feminino, respectivamente? Considerando que o total de candidatos foi 1071. *

Marcar apenas uma oval.

- ☐ 26% e 74%
- ☐ 50% e 50%
- ☐ 59% e 41%
- ☐ 66% e 34%

Gráfico de Bolha (Bubble Chart)

Dados os Gráficos de Bolha do site, responda as seguintes questões. Observe que ao colocar o mouse sobre as bolhas é exibido a quantidade de candidatos daquela característica.

4. De qual Bairro é a maioria dos candidatos? *

Marcar apenas uma oval.

- ☐ Campo Grande
- ☐ Centro
- ☐ Bangu
- ☐ Tijuca

5. De quantas Cidades diferentes os candidatos vieram? *

Marcar apenas uma oval.

- ☐ 16
- ☐ 31
- ☐ 25
- ☐ 22

6. Qual é o turno do Curso mais concorrido? *

Marcar apenas uma oval.

- ☐ Manhã
- ☐ Tarde

Gráfico de Barras de Duas Escalas (Dual Scale)

Figura 22: Formulário de Avaliação da Transmissão de Informações página 2 de 4

Dado os Gráficos de Barras de Duas Escalas do site, responda as seguintes questões. Observe que ao colocar o mouse sobre a barra é exibido a quantidade de candidatos daquela característica.

7. Em qual Curso é possível notar a maior disparidade de candidatos do sexo masculino e feminino? *

Marcar apenas uma oval.

- ☐ Mecânica - Turno da Tarde
- ☐ Mecânica - Turno da Manhã
- ☐ Eletrônica - Turno da Tarde
- ☐ Guia de Turismo - Turno da Tarde

8. De qual Bairro veio a maior quantidade de candidatos do sexo masculino? E do sexo feminino? *

Marcar apenas uma oval.

- ☐ Irajá e Taquara
- ☐ Taquara e Campo Grande
- ☐ Ambos de Campo Grande
- ☐ Campo Grande e Taquara

Gráfico de Barras Agrupadas (Grouped Bar Chart)

Dado os Gráficos de Barras Agrupadas do site, responda as seguintes questões. Observe que ao colocar o mouse sobre a barra é exibido a quantidade de candidatos daquela característica.

9. De qual Cidade veio o maior número de candidatos que estudaram em escolas públicas? *

Marcar apenas uma oval.

- ☐ Duque de Caxias
- ☐ Magé
- ☐ Rio de Janeiro
- ☐ Niterói

10. **Dentre os cursos, qual foi o que teve a maior diferença em relação a quem se declarou de baixa renda e quem não? ***

Marcar apenas uma oval.

- ☐ Administração - Turno da Manhã
- ☐ Edificações - Turno da Manhã
- ☐ Mecânica - Turno da Manhã
- ☐ Meteorologia - Turno da Manhã

Gráfico de Donut (Donut Chart)

Dado os Gráficos de Donuts do site, responda as seguintes questões. Observe que ao colocar o mouse sobre o atributo é exibido a quantidade de candidatos daquela característica.

11. **Qual a Etnia predominante entre os candidatos? ***

Marcar apenas uma oval.

- ☐ Parda
- ☐ Branca
- ☐ Preta
- ☐ Indígena

12. **Sobre a distribuição dos candidatos por Curso, é possível notar que: ***

Marcar apenas uma oval.

- ☐ A maioria deseja estudar a tarde
- ☐ É muito equilibrada.
- ☐ A maioria deseja estudar de manhã
- ☐ A minoria deseja fazer Eletrônica

13. **Qual das alternativas corresponde respectivamente ao número de candidatos das etnias Indígena, Amarela, Preta, Parda e Branca? ***

Marcar apenas uma oval.

- ☐ 1, 17, 106, 357, 559
- ☐ 5, 15, 100, 360, 559
- ☐ 10, 10, 106, 357, 650
- ☐ 11, 24, 50, 380, 730

Figura 24: Formulário de Avaliação da Transmissão de Informações página 4 de 4

Apêndice D

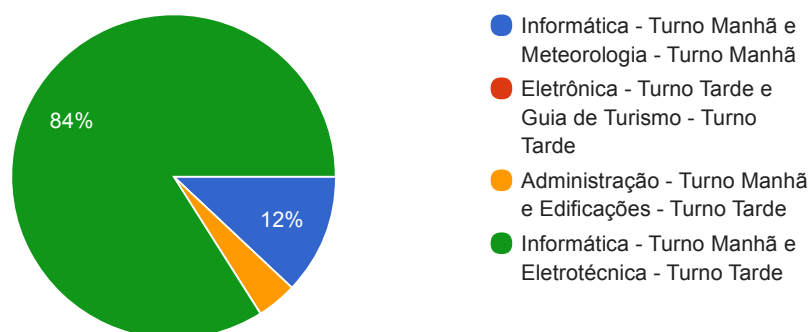
Resultados Obtidos do Formulário de Avaliação da Transmissão de Informações

Gráfico de Barras (Bar Chart)

25 respostas

Qual é o Curso mais concorrido e o menos concorrido, respectivamente?

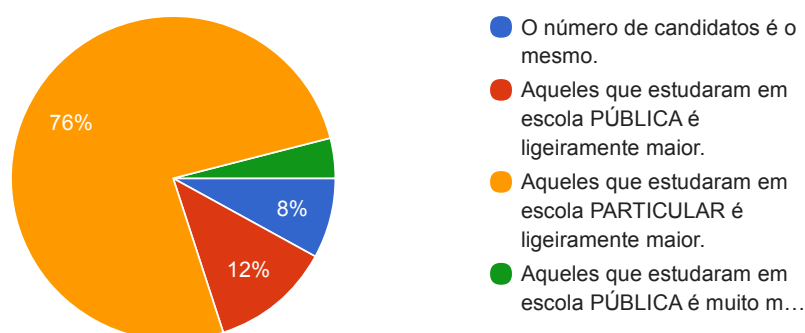
25 respostas



O Gráfico de Ensino informa quantos candidatos cursaram em escolas públicas. Após analisar o gráfico, selecione a opção correta sobre o número de candidatos que estudaram em escolas públicas em relação aos que não estudaram. Sendo "sim" para aqueles que estudaram em escola pública e "não" para os que estudaram em escola particular.

25 respostas

Figura 25: Resultados do formulário de Avaliação da Transmissão de Informações página 1 de 8



Qual é a percentagem mais provável dos candidatos do sexo masculino e feminino, respectivamente? Considerando que o total de candidatos foi 1071.

25 respostas

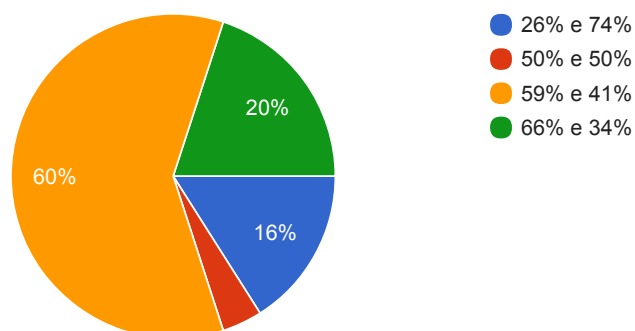
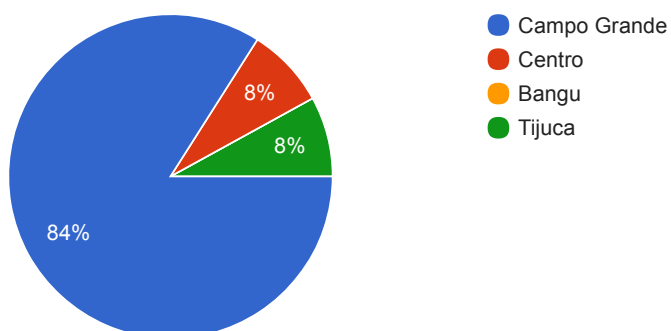


Gráfico de Bolha (Bubble Chart)

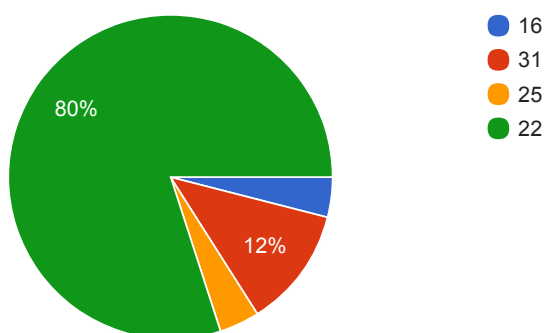
Figura 26: Resultados do formulário de Avaliação da Transmissão de Informações página 2 de 8

25 respostas



De quantas Cidades diferentes os candidatos vieram?

25 respostas



Qual é o turno do Curso mais concorrido?

25 respostas

Figura 27: Resultados do formulário de Avaliação da Transmissão de Informações página 3 de 8

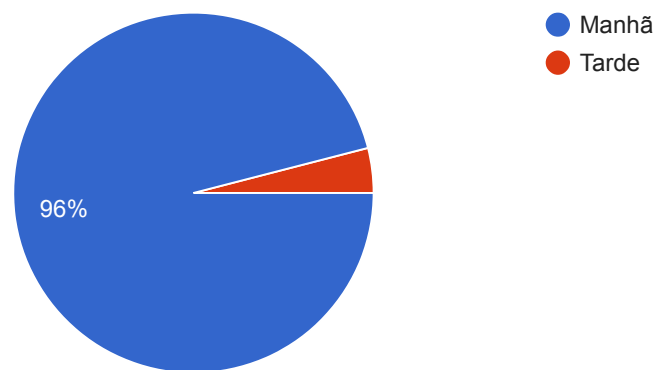
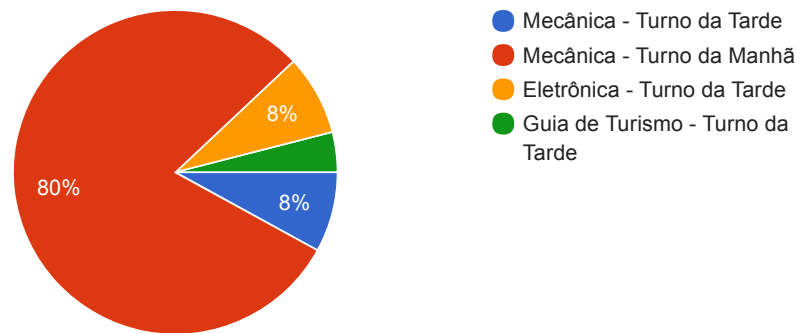


Gráfico de Barras de Duas Escalas (Dual Scale)

Em qual Curso é possível notar a maior disparidade de candidatos do sexo masculino e feminino?

25 respostas

Figura 28: Resultados do formulário de Avaliação da Transmissão de Informações página 4 de 8



De qual Bairro veio a maior quantidade de candidatos do sexo masculino? E do sexo feminino?

25 respostas

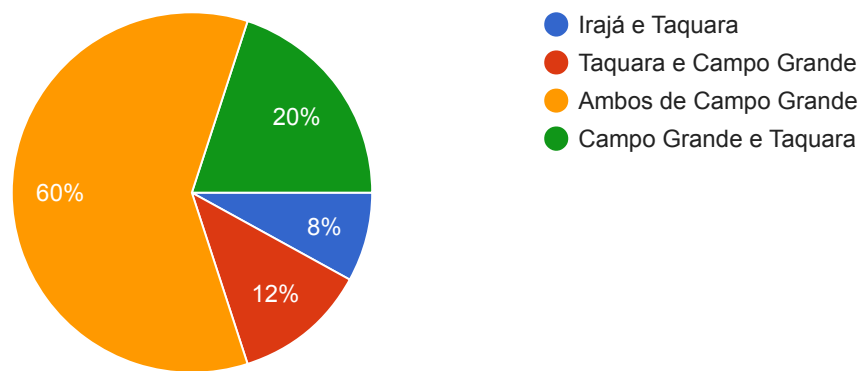
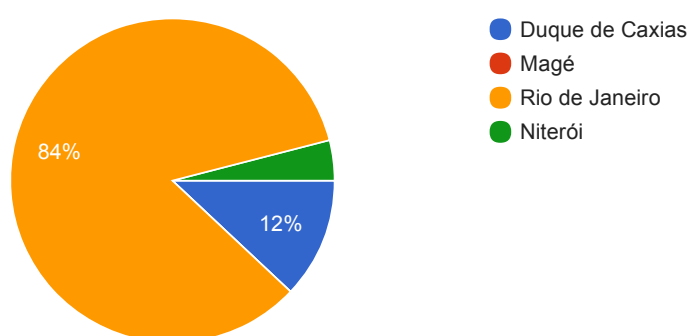


Gráfico de Barras Agrupadas (Grouped Bar Chart)

Figura 29: Resultados do formulário de Avaliação da Transmissão de Informações página 5 de 8

De qual Cidade veio o maior número de candidatos que estudaram em escolas públicas?

25 respostas



Dentre os cursos, qual foi o que teve a maior diferença em relação a quem se declarou de baixa renda e quem não?

25 respostas

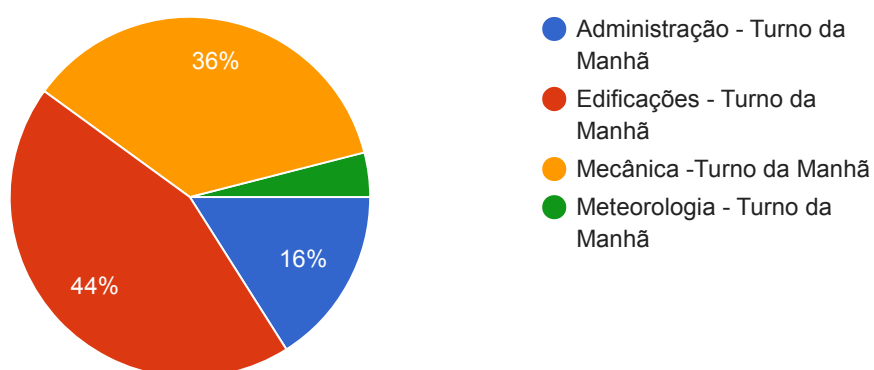
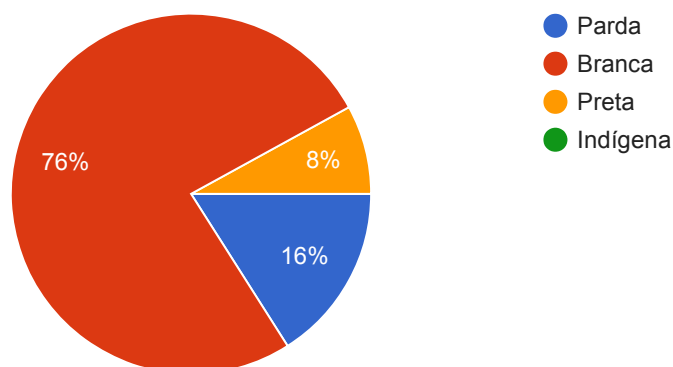


Figura 30: Resultados do formulário de Avaliação da Transmissão de Informações página 6 de 8

Gráfico de Donut (Donut Chart)

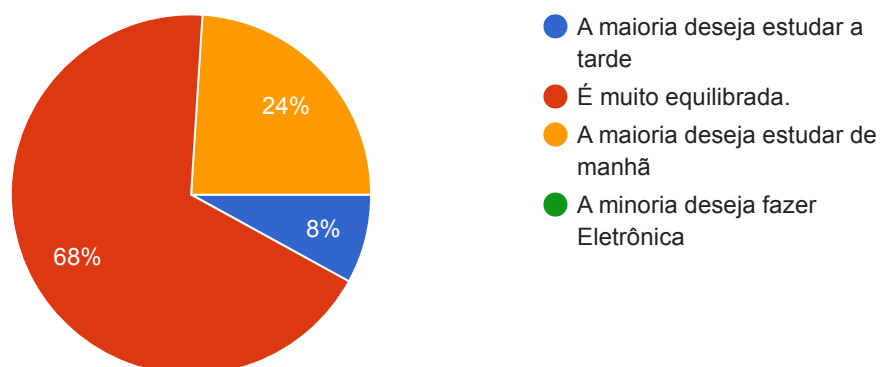
Qual a Etnia predominante entre os candidatos?

25 respostas



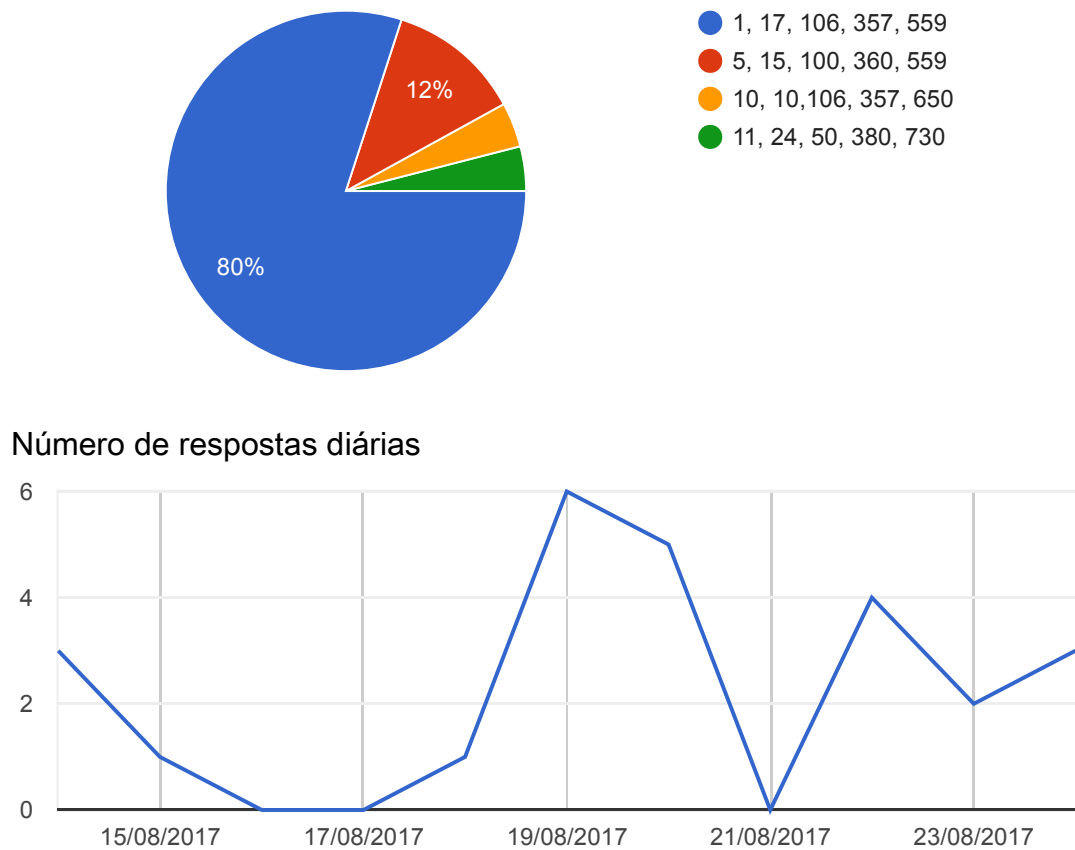
Sobre a distribuição dos candidatos por Curso, é possível notar que:

25 respostas



Qual das alternativas corresponde respectivamente ao número de candidatos das etnias Indígena, Amarela, Preta, Parda e Branca?

25 respostas



Este conteúdo não foi criado nem aprovado pela Google. Denunciar abuso - Termos de Utilização
- Termos adicionais

Google Formulários

Figura 32: Resultados do formulário de Avaliação da Transmissão de Informações página 8 de 8

D.1 Resolução do Formulário

A tabela 8 contém as respostas corretas para cada uma das 13 questões do formulário de Avaliação da Transmissão de Informações da aplicação.

Tabela 8: Respostas corretas para cada questão do formulário

Pergunta	Resposta Correta
Qual é o Curso mais concorrido e o menos concorrido, respectivamente?	Informática - Turno Manhã e Eletrotécnica - Turno Tarde
O Gráfico de Ensino informa quantos candidatos cursaram em escolas públicas. Após analisar o gráfico, selecione a opção correta sobre o número de candidatos que estudaram em escolas públicas em relação aos que não estudaram. Sendo "sim" para aqueles que estudaram em escola pública e "não" para os que estudaram em escola particular.	Aqueles que estudaram em escola PARTICULAR é ligeiramente maior
Qual é a porcentagem mais provável dos candidatos do sexo masculino e feminino, respectivamente? Considerando que o total de candidatos foi 1071.	59% e 41%
De qual Bairro é a maioria dos candidatos?	Campo Grande
De quantas Cidades diferentes os candidatos vieram?	22
Qual é o turno do Curso mais concorrido?	Manhã
Em qual Curso é possível notar a maior disparidade de candidatos do sexo masculino e feminino?	Mecânica - Turno da Manhã
De qual Bairro veio a maior quantidade de candidatos do sexo masculino? E do sexo feminino?	Ambos de Campo Grande
De qual Cidade veio o maior número de candidatos que estudaram em escolas públicas?	Rio de Janeiro
Dentre os cursos, qual foi o que teve a maior diferença em relação a quem se declarou de baixa renda e quem não?	Edificações - Turno da Manhã
Qual a Etnia predominante entre os candidatos?	Branca
Sobre a distribuição dos candidatos por Curso, é possível notar que:	É muito equilibrada
Qual das alternativas corresponde respectivamente ao número de candidatos das etnias Indígena, Amarela, Preta, Parda e Branca?	1, 17, 106, 357, 559

Apêndice E

Questionário Socioeconômico

PERFIL DOS CANDIDATOS

INFORMAÇÕES SOCIOCULTURAIS

Estas informações objetivam conhecer o perfil dos candidatos inscritos no concurso. Desse modo, sua resposta ao questionário é muito importante para a Instituição.

As informações não interferirão no processo seletivo. Elas são de caráter confidencial e servirão exclusivamente para o diagnóstico e planejamento das atividades de ensino.

Ao realizar sua inscrição, preencha com cuidado, na FICHA DE INSCRIÇÃO, os campos correspondentes aos itens aqui apresentados, **ASSINALANDO APENAS UMA OPÇÃO DE RESPOSTA PARA CADA PERGUNTA.**

1. É a primeira vez que você está prestando concurso para o CEFET/RJ?
 - a) Sim
 - b) Não
2. Você frequentou curso preparatório para prestar este concurso para o CEFET/RJ?
 - a) Não
 - b) Sim, por um semestre
 - c) Sim, o ano inteiro
 - d) Sim, por mais de um ano
3. Neste ano, você pretende prestar concurso para outras escolas?
 - a) Não
 - b) Sim, para escolas militares
 - c) Sim, para escolas de ensino médio
 - d) Sim, para outras escolas de ensino técnico
 - e) Sim, para outra escola de ensino técnico junto com ensino médio
4. Qual o tipo de escola de nível médio de seu maior interesse?
 - a) Técnicas Federais
 - b) Técnicas Estaduais
 - c) Militares
 - d) Médio Convencional, sem o técnico
5. Se você pudesse escolher o curso técnico de sua preferência, qual seria ele? **(apenas para os candidatos da Unidade Maracanã)**
 - a) Edificações
 - b) Estradas
 - c) Meteorologia
 - d) Eletrônica
 - e) Eletrotécnica
 - f) Mecânica
 - g) Informática
 - h) Segurança do Trabalho
 - i) Administração
 - j) Turismo e Entretenimento
 - k) Telecomunicações
6. Você pretende cursar uma faculdade?
 - a) Sim
 - b) Não
 - c) Talvez
7. Por que você escolheu o CEFET/RJ para estudar?
 - a) Porque é uma escola reconhecida pela formação que dá aos alunos
 - b) Porque oferece o melhor curso técnico
 - c) Por recomendação de professores, parentes, amigos, etc
 - d) Para me preparar para o ensino superior na área tecnológica.
 - e) Por outras razões
8. Você conhece as atividades da profissão correspondente ao curso de sua preferência?
 - a) Sim
 - b) Não
 - c) Não tenho certeza
9. Qual o nível de instrução de seu pai?
 - a) Ele não frequentou escola
 - b) Fundamental incompleto
 - c) Fundamental completo
 - d) Médio incompleto
 - e) Médio completo
 - f) Técnico de nível médio
 - g) Superior incompleto
 - h) Superior completo
 - i) Pós-graduação
 - j) Não sei
10. Qual o nível de instrução de sua mãe?
 - a) Ela não frequentou escola
 - b) Fundamental incompleto
 - c) Fundamental completo
 - d) Médio incompleto
 - e) Médio completo
 - f) Técnico de nível médio
 - g) Superior incompleto
 - h) Superior completo
 - i) Pós-graduação
 - j) Não sei

Figura 33: Questionário socioeconômico página 1 de 3

11. Com quem você mora?
- Com meus pais
 - Com meu pai
 - Com minha mãe
 - Com outros familiares que não meu pai ou minha mãe
 - Com outra(s) pessoa(s)
 - Sozinho
12. Qual o principal responsável pelo sustento de sua família?
- Pai
 - Mãe
 - Pai e mãe
 - Parente
 - Eu mesmo
 - Outro
13. Qual a situação de trabalho do(a) principal responsável pelo sustento de sua família?
- Trabalha regularmente com vínculo empregatício
 - Tem seu próprio negócio
 - Presta serviços por conta própria
 - Trabalha em atividade remunerada, dentro de casa
 - Está desempregado(a)
 - É aposentado(a)
 - É aposentado(a) e continua trabalhando regularmente
 - Vive de renda
 - Nenhuma das opções acima
 - Não tenho informações
14. Qual a renda mensal de sua família?
- Até 1 salário mínimo
 - Mais de 1 a 2 salários mínimos
 - Mais de 2 a 3 salários mínimos
 - Mais de 3 a 5 salários mínimos
 - Mais de 5 a 10 salários mínimos
 - Mais de 10 salários mínimos
15. Qual a sua participação na vida econômica da família?
- Não trabalho
 - Trabalho, mas recebo ajuda financeira da minha família
 - Trabalho e sou responsável pelo meu sustento
 - Trabalho e contribuo para o sustento da família
 - Trabalho e sou o responsável pelo sustento da família
16. Qual a situação do imóvel em que você reside?
- Próprio, já quitado
 - Próprio, ainda não quitado
 - Alugado
 - Emprestado
 - Outra
17. Em relação ao local de sua residência, como se situa o CEFET/RJ?
- Perto
 - Distante
 - Muito distante, mas no mesmo município
 - Muito distante, em município diferente
18. Se você vier a ser aluno do CEFET/RJ, que meio de transporte você utilizará para chegar à escola?
- Somente ônibus
 - Somente trem/metrô
 - Ônibus e trem/metrô
 - Carro
 - Bicicleta
 - Nenhum
19. Sua família possui automóvel?
- Não
 - Sim
20. Qual o meio de informação que você mais utiliza para se manter atualizado?
- TV
 - Rádio
 - Jornal
 - Revista
 - Internet
 - Conversa com outras pessoas
 - Não tenho buscado informações
21. Que tipo de leitura você prefere?
- Romance
 - Ficção científica
 - Poesia
 - Livros técnicos/científicos
 - Outros
 - Nenhum
22. Você lê jornal?
- Não
 - Sim, ocasionalmente
 - Sim, todos os domingos
 - Sim, diariamente

Figura 34: Questionário socioeconômico página 2 de 3

23. Que seção do jornal você prefere?
- Esportes
 - Informática
 - Notícias locais
 - Notícias internacionais
 - Política
 - Economia
 - Cultura
 - Quadrinhos
 - Outras
 - Nenhuma
24. Excetuando-se os livros didáticos, quantos livros você lê por ano?
- de um a três
 - de quatro a cinco
 - de seis a dez
 - não costumo ler livros
25. Você assiste à TV?
- Não
 - Sim, ocasionalmente
 - Sim, diariamente
26. Qual o tipo de programação preferida na TV?
- Telejornal
 - Filmes
 - Desenhos animados
 - Novelas
 - Esportes
 - Programas humorísticos e/ou de auditório
 - Programas de entrevista e/ou documentários
 - Programas culturais
 - Outros
 - Nenhum
27. De que tipo de atividade você **mais gosta** para se divertir?
- Ir à praia
 - Ir ao cinema
 - Praticar esporte e/ou assistir a eventos esportivos
 - Ouvir música
 - Frequentar shopping
 - Ir à LAN House
 - Tirar fotografia
 - Ler
 - Outros
 - Nenhum
28. Durante a semana, você tem algum tipo de atividade extraescolar?
- Não
 - Sim, curso de língua estrangeira
 - Sim, curso de informática
 - Sim, prática esportiva ou ginástica
 - Sim, outro
29. Você usa computador?
- Não
 - Sim, em casa
 - Sim, em outros locais
30. Qual a principal razão para você usar computador?
- Para fazer trabalhos escolares
 - Por diversão
 - Para cumprir as tarefas no meu local de trabalho
 - Não sei usar computador
31. Como você soube do concurso para o qual está se inscrevendo?
- Pelo jornal, televisão e/ou rádio
 - Por cartazes e/ou folhetos
 - Por amigos ou parentes
 - Por palestra na escola e/ou no curso preparatório
 - Por outros meios
32. Qual é a sua maior expectativa em relação a estudar no CEFET/RJ?
- Fazer amigos
 - Adquirir conhecimentos em geral
 - Obter formação profissional técnica
 - Conseguir trabalhar ao final do curso
 - Preparar-me para prestar vestibular
 - Ter uma profissão para continuar estudando
 - Outra
33. Qual a sua cor/etnia?
- Branca
 - Parda
 - Negra
 - Amarela
 - Outros

Figura 35: Questionário socioeconômico página 3 de 3