



Universidade do Estado do Rio de Janeiro
Centro de Tecnologia e Ciências
Escola Superior de Desenho Industrial

Luiz Frederico Sarkis Arbex

Visualização dos dados estatísticos da UERJ: proposta de dashboards baseados no trabalho de Jacques Bertin

Visualização dos dados estatísticos da UERJ: proposta de dashboards baseados no trabalho de Jacques Bertin

Luiz Frederico Sarkis Arbex

**Visualização dos dados estatísticos da UERJ: proposta de dashboards baseados
no trabalho de Jacques Bertin**



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-graduação em Design, da Universidade do Estado do Rio de Janeiro.

Orientador: Prof. Dr. Washington Dias Lessa

Coorientador: Prof. Dr. André Soares Monat

Rio de Janeiro

2013

Luiz Frederico Sarkis Arbex

**Visualização dos dados estatísticos da UERJ: proposta de dashboards baseados no
trabalho de Jacques Bertin**

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-graduação em Design, da Universidade do Estado do Rio de Janeiro.

Aprovada em 25 de setembro de 2013

Banca Examinadora:

Prof. Dr. Washington Dias Lessa (Orientador)

Escola Superior de Desenho Industrial – UERJ

Prof. Dr. Marcos André Franco Martins

Escola Superior de Desenho Industrial – UERJ

Prof.^a Dra. Carla Maria Dal Sasso Freitas

Universidade Federal do Rio Grande do Sul

Prof. Dr. André Soares Monat (Coorientador)

Escola Superior de Desenho Industrial – UERJ

Rio de Janeiro

2013

DEDICATÓRIA

Dedico este trabalho ao grande designer Goebel Weyne,
pioneiro, mestre, amigo, crítico

(e muito possivelmente, entre nós, o introdutor dos
estudos bertinianos no âmbito acadêmico do design).

AGRADECIMENTOS

A Washington Dias Lessa, professor estimulante e orientador dedicado, cujo interesse por Bertin e demais afinidades temáticas resultaram em um frutífero período de aprendizagem e intensa troca de ideias.

Aos professores Carla Freitas e Marcos Martins, que gentilmente aceitaram fazer parte da banca e cujas contribuições no exame de qualificação foram fundamentais para o desenvolvimento posterior do trabalho.

Ao professor André Monat pelo interesse e incentivo iniciais, pela sugestão de tema e pelos livros emprestados (a serem todos devidamente devolvidos).

Ao professor Rodolfo Capeto por minha primeira (e única) aula sobre Semiologia gráfica, e pelas frequentes opiniões, sugestões e esclarecimentos posteriores e, ainda, pelo interesse e incentivo constantes.

À professora Mailsa Carla Passos pelas relevantes considerações sobre os aspectos sociais e políticos presentes no trabalho.

Ao professor Marcelo Rubens pelas orientações estatísticas.

À professora Doris Kosminsky pela generosidade aceitando participar da suplência da banca examinadora.

À equipe do *Projeto de BI UERJ*, em especial, Erika Silva e Carlos Eduardo Rodrigues, pela receptividade inicial e pelos esclarecimentos fundamentais para o desenrolar da pesquisa.

À equipe do DataUERJ e do Niesc, Núcleo de Informação e Estudos de Conjuntura da UERJ, pelos dados disponibilizados.

Ao amigo Gustavo Ferreira, pela inestimável ajuda em programar as visualizações iniciais, trabalho sem o qual o resultado apresentado seria impensável.

Ao amigo de sempre Pablo Rossi, pelo interesse no tema e pela também inestimável ajuda na produção do vídeo de apresentação da defesa.

À companheira Elaine Moreira, pelo amor, incentivo, compreensão e paciência ao longo de todo o processo de desenvolvimento deste trabalho.

À minha mãe, Vera Lúcia Arbex, pelo suporte e incentivo.

À Sonia Moreira e Piedade Carvalho pela providencial ajuda em ocasiões diversas.

Aos amigos do mestrado, que propiciaram um ambiente de importantes reflexões e contribuíram com sugestões, referências, debates. Em especial, Clara Meliande, Sérgio Boiteux, Almir Mirabeau, André Malheiro e Eduardo Vieira.

Aos demais amigos, que dando informações, indicações e opiniões, emprestando equipamentos ou livros, conversando especificamente sobre a pesquisa ou sobre as diversas questões acadêmico-existenciais contribuíram em algum momento pontual, fazendo com que a engrenagem continuasse rodando até a conclusão do trabalho.

À equipe da biblioteca da Esdi, pela dedicação, interesse e paciência nas minhas visitas, consultas e solicitações constantes.

À equipe do laboratório de informática da Esdi, pelo suporte na montagem da estrutura para a defesa.

À secretaria da Pós-Graduação, pela eficiência e prontidão ao longo do curso.

RESUMO

ARBEX, L. F. S. *Visualização dos dados estatísticos da UERJ: proposta de dashboards baseados no trabalho de Jacques Bertin*. 2013. 190 f. Dissertação (Mestrado em Design) – Escola Superior de Desenho Industrial, Universidade do Estado do Rio de Janeiro. Rio de Janeiro, 2013.

Os painéis de gráficos estatísticos conhecidos como *dashboards* são utilizados comumente na área de *Business Intelligence (BI)* para a visualização de grandes sistemas organizados de dados. A presente dissertação propõe embasar o projeto de dashboards pelas teorias de Jacques Bertin, formuladas nas obras *Sémiologie Graphique* e *La Graphique et le Traitement Graphique de l'information*. Considerando este referencial, e ainda parâmetros do design de informação e da visualização de dados, foram desenvolvidos dashboards que apresentam dados sobre a política de reserva de vagas da Universidade do Estado do Rio de Janeiro, sistematizados pelo projeto de BI dessa instituição. O objetivo foi não apenas o de atender aos requisitos convencionais de um dashboard, mas sobretudo o de apresentar outras perspectivas informativas. Nesse sentido, investigam-se as especificidades dos métodos de Bertin e sua expansão para o domínio dos sistemas interativos.

Palavras-chave: Semiologia gráfica. Dashboard. Jacques Bertin. Visualização de dados. Design de informação.

ABSTRACT

ARBEX, L. F. S. *Visualization of State University of Rio de Janeiro statistical data: proposed dashboards based on the work of Jacques Bertin*. 2013. 190 f. Dissertação (Mestrado em Design) – Escola Superior de Desenho Industrial, Universidade do Estado do Rio de Janeiro. Rio de Janeiro, 2013.

The panels of statistical charts known as dashboards are employed in the area of Business Intelligence (BI) for the visualization of large-scale organized systems of data. This Masters Thesis proposes to base the design of dashboards by theories of Jacques Bertin, formulated in the books *Sémiologie graphique* and *La graphique et le traitement de l'information*. Considering this framework, and further parameters of information design and data visualization, it were developed dashboards that present data about the policy of quotas from the State University of Rio de Janeiro, systematized by the BI project of that institution. The goal was not only to meet the requirements of a conventional dashboard, but rather to present other perspectives of information. Accordingly, we investigate the specific methods of Bertin and its expansion into the area of interactive systems.

Keywords: Semiology of Graphics. Jacques Bertin. Dashboard. Data Visualization. Information Design.

LISTA DE GRÁFICOS

Figura 1 – Desenhos na caverna de Lascaux, França.....	26
Figura 2 – Visualidade característica de dashboards para aplicações diversas.....	35
Figura 3 – Dashboard sobre frequência de visitantes nos parques nacionais americanos.....	37
Figura 4 – Componentes de um arquitetura de alto nível de BI.....	43
Figura 5 – Proposta de dashboard apresentada por Stephen Few.....	47
Figura 6 – Semiologia gráfica nos sistemas de significação fundamentais.....	55
Figura 7 – Representação da relação entre os componentes tempo e quantidade de francos.....	57
Figura 8 – Os três níveis de informação.....	63
Figura 9 – As três dimensões da imagem.....	64
Figura 10 – As três formas de intervenção gráfica.....	69
Figura 11 – Esquematização dos tratamentos gráficos heterogêneos e homogêneos.....	70
Figura 12 – Tabelas de alocação.....	73
Figura 13 – Tabelas de alocação com mais de dois componentes.....	74
Figura 14 – Tabelas de alocação: uso da notação [\].....	75
Figura 15 – Relação entre tabelas de alocação e esquemas de homogeneidade.....	76
Figura 16 – Tabela de pertinência: ocupação de um hotel.....	78
Figura 17 – Tabela de pertinência: planejamento de território.....	79
Figura 18 – Gráficos no nível elementar de informação.....	81
Figura 19 – Uso de gráficos no nível elementar de informação em dashboards.....	81
Figura 20 – Construção normal: produção de carne no Mercado Comum Europeu.....	82
Figura 21 – Etapas do processo de permutação.....	83
Figura 22 – Sinopse das construções gráficas.....	84
Figura 23 – Escala absoluta e escala relativa.....	85
Figura 24 – Aplicações da lei da visibilidade.....	86
Figura 25 – Gráfica de comunicação.....	88
Figura 26 – Arquitetura básica do sistema de BI da UERJ.....	94
Figura 27 – Interface do recurso de tabela dinâmica.....	98
Figura 28 – Dashboard básico.....	102

Figura 29 – Tabela de alocação (cubo <i>graduação</i>).....	107
Figura 30 – Esquema de homogeneidade (cubo <i>graduação</i>).....	109
Figura 31 – Esquema de homogeneidade simplificado.....	110
Figura 32 – Tabela de pertinência.....	112
Figura 33 – Consultas individuais à base de dados do BI.....	116
Figura 34 – Transcrição visual da tabela numérica.....	118
Figura 35 – Exemplificação do recurso de mouse over para revelar informações secundárias	120
Figura 36 – Questões da representação em escala absoluta	122
Figura 37 – Escala de expansão simples.....	123
Figura 38 – Escala de expansão pelo mínimo.....	124
Figura 39 – Escala de expansão pela métrica (média).....	125
Figura 40 – Escalas de expansão pela métrica: questões de interpretação.....	126
Figura 41 – Comparação entre escalas.....	127
Figura 42 – Comparação entre escalas – diagrama 1: panorama.....	129
Figura 43 – Visibilidade dos itens acima da média e permutação.....	131
Figura 44 – Aspectos da gráfica de comunicação.....	133
Figura 45 – Gráfica de comunicação: destaques visuais.....	135
Figura 46 – Combinação entre indicadores: sexo e conclusão.....	138
Figura 47 – Combinação entre indicadores: rede de ensino e conclusão.....	139
Figura 48 – Combinação entre indicadores: rede de ensino e conclusão.....	141
Figura 49 – Dashboard derivado 1.....	142
Figura 50 – Síntese do diagrama 2 a partir de componentes do diagrama 1.....	143
Figura 51 – Diagrama 2: comparação entre cotistas e não cotistas.....	145
Figura 52 – Processo de síntese dos indicadores de CR por curso.....	146
Figura 53 – CR por curso: agrupamento por predominância entre cotistas e não cotistas..	148
Figura 54 – CR por curso: agrupamento por tendência evolutiva.....	149
Figura 55 – CR por cursos: agrupamento por classes de CR.....	149
Figura 56 – Dashboard derivado 2: três perspectivas sobre o coeficiente de rendimento.....	151
Figura 57 – Detalhamento – CRs mais baixos.....	152
Figura 58 – Detalhamento – CRs mais baixos por tipo de cota.....	152
Figura 59 – Diagrama 3: ocupação dos cursos por cotistas.....	154
Figura 60 – Dashboard básico: estudo de correlações 1.....	155

Figura 61 – Dashboard básico: estudo de correlações 2.....	156
Figura 62 – Dashboard básico: estudo de correlações 3.....	156

Anexo

Figura 63 – Características da percepção visual do plano.....	171
Figura 64 – Propriedades do plano.....	172
Figura 65 – Três categorias de imposição.....	173
Figura 66 – Tipos de imposição.....	174
Figura 67 – Variáveis retinianas.....	175
Figura 68 – Nível de organização das variáveis retinianas.....	177
Figura 69 – Nível de organização das variáveis visuais.....	181
Figura 70 – Regras gerais de legibilidade.....	182

LISTA DE TABELAS

Tabela 1 – As três etapas sucessivas da intervenção gráfica.....	68
Tabela 2 – Símbologia usada na tabela de alocação.....	72
Tabela 3 – Dados do cubos de desempenho acadêmico.....	96
Tabela 4 – Registro de dados por aluno.....	104
Tabela 5 – Formulário estatístico: aluna 10027, ano 2005.....	105
Tabela 6 – Tabela homogênea.....	117

Anexo

Tabela 7 – Planilha diagrama 1: características gerais.....	184
Tabela 8 – Comparação entre cotistas e não cotistas: características gerais.....	185
Tabela 9 – Planilha diagrama 2, drill down - CR por cursos entre cotistas e não cotistas.....	187
Tabela 10 – Planilha diagrama 3: ocupação de cotistas e não cotistas por curso (%).....	190

SUMÁRIO

	INTRODUÇÃO	16
1	DASHBOARDS	33
1.1	Características gerais e definições	33
1.2	Dashboard no contexto da <i>Business Intelligence</i>	40
1.2.1	<u>Dashboards na arquitetura de um sistema de <i>Business Intelligence</i></u>	42
1.3	Dashboard como artefato de comunicação	44
1.3.1	<u>Dashboard no contexto do design de informação</u>	47
1.4	Considerações gerais	49
1.4.1	<u>Proposta de síntese para definição de dashboards</u>	50
2	SEMIOLOGIA GRÁFICA E TRATAMENTO GRÁFICO DA INFORMAÇÃO	52
2.1	Jacques Bertin: <i>Sémiologie graphique</i>	52
2.1.1	<u>O aspecto monossêmico da percepção visual</u>	53
2.1.2	<u>Análise da informação</u>	55
2.1.2.1	Componentes e invariantes.....	56
2.1.2.2	Número de componentes.....	57
2.1.2.3	Comprimento dos componentes.....	58
2.1.2.4	Nível de organização dos componentes.....	59
2.1.3	<u>Regras do sistema gráfico</u>	61
2.1.3.1	O problema gráfico.....	61
2.1.3.2	A teoria da imagem.....	61
2.1.3.3	As três funções da representação gráfica.....	65
2.1.3.4	As regras de construção.....	66
2.2.1	<u>Análise matricial</u>	68
2.2.1.1	Processo de análise matricial.....	71
2.2.2	<u>Construção normal e permutação</u>	80
2.2.3	<u>Questões de visibilidade</u>	85
2.2.4	<u>Gráfica de comunicação</u>	87
2.2.5	<u>Síntese</u>	89

3	PROPOSIÇÃO	90
4	APLICAÇÃO E DISCUSSÃO	92
4.1	Portal de Indicadores UERJ	92
4.1.1	<u>Estrutura técnica: noções gerais</u>	92
4.1.1.1	Arquitetura técnica	93
4.2	Síntese sobre a discussão em torno das cotas	99
4.3	Teoria matricial	102
4.3.1	<u>Registro dos dados discentes</u>	103
4.3.2	<u>Formulário estatístico</u>	103
4.3.4	<u>Esquema de homogeneidade</u> (<i>le schéma d'homogénéité</i>)	108
4.3.5	<u>Tabela de pertinência</u> (<i>le tableau de pertinence</i>)	111
4.3.6	<u>Tabela de dados</u>	115
4.4	Dashboards	117
4.4.1	<u>Tratamento de informação básico: diagrama</u>	117
4.4.1.1	Tabela homogênea e transcrição	117
4.4.1.2	As questões de escala	121
4.4.1.2.1	Escala absoluta	121
4.4.1.2.2	Escala relativa	122
4.4.1.2.3	Comparação entre escalas	127
4.4.1.3	Análise e interpretação	129
4.4.1.4	Permutação	130
4.4.1.5	Gráfica de comunicação	133
4.4.2 A	A perspectiva digital: diagramas 2 e 3	136
4.4.2.1	<i>Drill down</i>	137
4.4.2.2	Diagrama 2: comparação entre cotistas e não cotistas	142
4.4.2.1.1	Análise	146
4.4.2.1.2	Estudo dos CRs mais baixos	151
4.4.2.2	Diagrama 3: ocupação dos cursos por cotistas	153
4.4.2.3	Síntese	155

CONCLUSÃO	158
REFERÊNCIAS	163
ANEXO A – Demais tópicos da Semiologia gráfica.....	168
ANEXO B – Planilhas de base.....	184

INTRODUÇÃO

Concentrando-se no estudo de painéis de gráficos estatísticos, *dashboards*, esta dissertação escolhe como objeto de análise um artefato específico, empregado em contextos igualmente bem delimitados. Como referencial teórico, os livros do cartógrafo francês Jacques Bertin apresentam uma teoria consistente para a representação visual de dados e, nesse sentido, parecem adequados ao objeto escolhido. Apesar desse recorte, o presente trabalho está contido na área mais ampla do *design de informação* que, por se tratar de uma ramificação das atividades de design, envolve questões específicas de projeto e de comunicação que não podem ser abordadas aqui de modo extenso. Entretanto, alguns desses aspectos serão brevemente delineados nessa introdução contextualizando este estudo no universo mais amplo do qual faz parte.

Dashboards são tipicamente empregados como instrumentos de monitoração e análise em situações nas quais o acompanhamento de sistemas e seus processos contribuem para a eficiência de uma operação. É o caso das áreas de *business*, usuária habitual de dashboards, em que a velocidade de compreensão e de reação a uma situação global de negócios ou ao controle de um tópico particular são demandas normais. Tais tarefas são exemplificadas pelo monitoramento do volume de vendas em determinado departamento e pela avaliação do desempenho de certo funcionário. Assim, dashboards são, em sua essência, mecanismos de controle, condição que poderia sugerir seu estudo sob uma perspectiva crítico-filosófica, como a ótica *foucaultiana* do dispositivo e seu papel na sociedade disciplinar, ou ainda a contextualização dos dashboards na sociedade de controle descrita por Deleuze.

Entretanto, a presente dissertação volta-se aos aspectos que procuram contribuir para a eficiência dos dashboards, direcionando-se, portanto, a um viés mais técnico/tecnológico do que crítico. Neste sentido, pode parecer que tal objetivo reforça os aspectos de dominação e controle levantados pelos filósofos franceses por meio de técnicas mais eficientes. Não há dúvida de que se trata de uma interpretação a ser levada em conta, considerando que melhorias em mecanismos de controle potencialmente intensificariam seus aspectos negativos. Por outro lado, as possibilidades do cenário tecnológico atual permitem entrever um viés que talvez possa contrabalançar essas tendências opressivas. Trata-se de recursos que permitem acrescentar a uma perspectiva unicamente instrumental uma outra, potencialmente emancipatória, na qual o dashboard atuaria como artefato que potencializa a comunicação objetiva, em um contexto

em que a abundância de dados tende a obstar a trocas de informações e a produção de novos significados. É nessa perspectiva que o presente trabalho se desenvolve.

Nesse âmbito, ganha relevo o significado do próprio conceito de *informação*. A espantosa capacidade de processamento dos computadores é historicamente bastante recente, e mesmo que hoje ela seja tida como evidente, as decorrências dessa capacidade são sentidas de modo enfático em nossa sociedade. Considerando as proporções atuais, a codificação binária, base das tecnologias digitais, poderia ser interpretada como um processo de replicação e expansão digital do mundo concreto. Segundo estudo de Hilbert e López (2011), desde os anos 2000 nossos registros tecnológicos em formato digital ultrapassaram os analógicos e, em 2007, 94% dos dados mundiais estavam registrados digitalmente. Nesse mesmo ano, 99,9% das telecomunicações operavam com tecnologia digitais. O estudo concluiu ainda que, entre 1986 e 2007 a capacidade computacional de uso geral vem crescendo 58% ao ano.

A copiosidade do mundo computacional para o armazenamento da cultura humana parece ir em sentido contrário à escassez de recursos e espaço do mundo real, sugerindo que as possibilidades de expansão daqui para frente apontam para o universo dos *bits*. Diante dessa enorme capacidade de processar, armazenar e transmitir dados ou, por outras palavras, face à disponibilidade tecnológica para registrar nossas memórias e vestígios, e sintetizar novos significados a partir deles, a questão que se impõe é: como se orientar em meio à incomensurabilidade do mundo digital. Mais especificamente, trata-se de buscar por sentido informativo em meio ao acúmulo de dados dispersos.

Em relação à crescente neurose do excesso de informação, Wurman há algum tempo comentou que “a grande era da informação é, na verdade, uma explosão da *não-informação* (...)” (1991, p.43, grifo do autor). Remediar essa questão, que se coloca como um desafio à toda a sociedade atual, é prerrogativa de diversas disciplinas. Dentre as mais próximas ao universo desta dissertação, estão aquelas que procuram abordá-la pelo viés da visualidade, como o design de informação e a visualização de dados (*data visualization*). Contudo, mesmo as disciplinas que buscam ser facilitadoras na produção de mensagens informativas, lidam com a possibilidade de que seus produtos tornem-se, eles mesmos, parte da homegeneidade amorfa da não informação. Esse risco não diz respeito apenas à qualidade intrínseca ao produto, mas é inerente à própria relatividade do conceito *informação*, tendo em vista que a mesma mensagem pode ser informativa para uns e lugar-comum para outros. Evidentemente, isso só acontece se

entendemos como *informação* aquilo que resulta no *novo*, isto é, algo que se soma ao que já se sabe; mas esse é apenas um dos muitos significados que o conceito pode ter.

Na linguagem escolástica, o ato de informar qualquer coisa significava passar da “virtualidade ao ato”¹. A etimologia latina aponta para o vocábulo *informatio, onis* como a “ação de formar, de fazer, fabricação etc.”². Já o entendimento moderno de informação ecoa metáforas de conceitos da engenharia e da física do século 19³. Nesse sentido, Wurman (ibid., p.42) aponta que até meados do século 20, havia relativo consenso sobre a palavra *informação*, interpretada como a “a ação de informar; formação ou moldagem da mente ou do caráter, treinamento, instrução, ensinamento, comunicação de conhecimento instrutivo”. Entretanto, o termo adquiriu novo significado com a *Teoria Matemática da Comunicação*, de Claude Shannon e, sobretudo, a partir de sua edição em livro prefaciado por Warren Weaver. Nesse texto, Weaver (1998) procura expandir a teoria original de Shannon (que referia-se aos parâmetros de eficiência na transmissão de sinais entre aparelhos) para abarcar a comunicação humana. Ele aponta as três instâncias dos problemas de comunicação: a) o problema técnico (quanto à transmissão precisa); b) o problema semântico (quanto à comunicação do significado pretendido); c) o problema da efetividade (quanto à indução de um comportamento esperado).

No *paper* de Shannon, nenhuma das 61 ocorrências do termo *informação* define estritamente esse conceito. Nesse sentido, é esclarecedor notar que *informação* completa expressões iniciadas por palavras como “medida”, “armazenamento”, “unidades”, “um bit” etc., evidenciando seu caráter quantitativo, condizente à afirmação de Shannon (1948a. p.379): “Os aspectos semânticos são irrelevantes para o problema da engenharia”. Fica claro então, que a teoria de Shannon se situa no primeiro nível dos problemas de comunicação. Desse modo, Weaver chama a atenção para o necessário do entendimento do referido problema técnico, argumentando que ele descreve as condições que viabilizam os outros níveis. Mas em última instância, Weaver considera que a separação em níveis é artificial, dado o estreito inter-

¹ Verbete *informer* em LEGRAND, Gérard. Dictionnaire de philosophie. Bordas: Paris, 1983.

² Verbete *informação* em Dicionário eletrônico Houaiss da língua portuguesa. 2007. Versão monousuário.

³ Em especial nos estudos ligados à termodinâmicas dos gases e à questão da perda de calor para o entorno a partir do trabalho de Sadi Carnot e, posteriormente, em Rudolf Clausius, James Maxwell e Lord Kelvin, dentre outros, quando o fenômeno passa a ser designado pelo conceito de “entropia”.

relacionamento entre eles. Nesse sentido, o autor sugere formas de expandir o esquema original de comunicação de Shannon⁴ de modo a incluir as comunicações semânticas.

A partir daí, o conceito de *informação* foi desenvolvido em duas vertentes principais: as ciências e tecnologias computacionais e os estudos de comunicação. Na área tecnológica, os conceitos de Shannon passaram a ser designados como “teoria da informação”. A palavra *informação* passou a substituir o termo tecnicamente mais estrito *sinais*, no sentido daquilo que, independentemente do seu significado, é transferido através de um *canal de transmissão*. Essa acepção foi amplamente adotada pelas áreas da computação (eventualmente nomeadas como “tecnologias da informação”) e por força da crescente proeminência e ubiquidade desses campos, tal significado torna-se de uso corrente. Pelas palavras de Wurman, o vocábulo *informação*: “De repente, podia ser aplicado a algo que, a rigor, não era necessariamente informativo.” (ibid.), notando que o termo se descolou dos aspectos que dão sentido a uma comunicação. Ele argumenta ainda que “a palavra *informe* foi destacada do substantivo *informação*, e a *forma*, ou estrutura, desapareceu do verbo *informar*.” (ibid., grifo do autor). Nesse sentido, Wurman argumenta em favor de uma distinção (que hoje é normalmente aceita) para resguardar a peculiaridade comunicativa do termo: “Grande parte daquilo que supomos ser informação é, na verdade, apenas *dado*, às vezes nem isso.” (ibid., grifo nosso). O uso do termo *dado* para designar elementos que podem ser armazenados, transmitidos, trocados e processados, independentemente dos aspectos semânticos é uma acepção mais contemporânea que vem se consolidando. Textos atuais sobre informação confirmam essa tendência. Assim, para Logan “o que Shannon analisa em sua teoria da informação é a transmissão de sinais ou dados.”—, considerando que “dados são os fatos puros e simples (...)” (2012, p.53). Em contraponto, ele reafirma o caráter semântico da informação, que não se trata de “uma invariante, [pois] depende do quadro de referência ou do contexto no qual é usada.” (ibid. p.16). Sendo assim, essa distinção entre dado e informação é empregada na presente dissertação, reservando-se o termo *dados* para os valores numéricos e elementos estatísticos registrados nos bancos de dados, e *informação* para os padrões visuais que podem responder às perguntas formuladas em um contexto específico.

Já pela via dos estudos de comunicação, além das propostas de Warren Weaver, novos

⁴ Trata-se do conhecidíssimo esquema de Shannon, composto por sinais emitidos por uma *fonte de informação* que são codificados através de um *transmissor* para serem transportados por um *canal* até um *receptor* que os decodifica e entrega à sua *destinação*.

conceitos sobre *informação* baseados tanto na teoria de Shannon quanto no conceito físico de *entropia*, são propostos por Norbert Wiener. O proponente dos estudos de *cibernética* argumentou um paralelo entre as pressões entrópicas, que tendem a reduzir a diversidade estrutural do Universo à homogeneidade amorfa e os conceitos comunicação. Para Wiener:

É possível, realmente, encarar conjuntos de mensagens como se fossem dotados de entropia, à semelhança de conjuntos de estados do mundo exterior. Assim como a entropia é uma medida de desorganização, a informação conduzida por um grupo de mensagens como sendo, essencialmente, o negativo de sua entropia e o logaritmo negativo de sua probabilidade. Vale dizer, quanto mais provável seja a mensagem, menor será a informação que propicia. Os chavões, por exemplo, são menos alusivos que os grandes poemas. (WIENER, 2007, p. 21)

Essa interpretação de *informação* como aquilo que é inversamente proporcional à sua possibilidade de ocorrência repercutiu de modo bastante amplo nos estudiosos das questões da linguagem. Por exemplo, Décio Pignatari associa *informação* à *entropia negativa* argumentando que, mesmo intuitivamente, a noção de *informação* liga-se à ideia de surpresa e originalidade, de modo que “Quanto menos previsível, ou mais rara, uma mensagem, maior sua informação (...).” (2003, p.57). Isso decorre do conceito mais amplo no qual pela “desdiferenciação de formas e funções, teríamos a tendência caótica ou entrópica, cujo ponto extremo seria a uniformidade geral.”. Nessa situação, seria impossível a produção ou transmissão de informação, que “só começa a existir onde houver um mínimo de diferenciação, um mínimo de alternativa (...)” (ibid. p.58).

O filósofo Vilém Flusser expande esse conceito entendendo a comunicação humana como fenômeno negativamente entrópico e, portanto, contranatural, pois se opõe à tendência natural do mundo a se desinformar “e é isso sobretudo que caracteriza o homem: ele é um animal que encontrou truques para acumular informações adquiridas.”. (2010, p.93). Segundo Flusser, isso acontece porque a humanidade engaja-se na ação de *informar* o mundo, de lhe dar forma por meio de coisas. Para o filósofo, informar o mundo é *arte*, termo que é a raiz de *artificial*, de *artifício*, sublinhando a condição inatural da atividade humana. Criando, transmitindo e conservando *informação*, o homem se opõe à tendência entrópica produzindo *arte*, termo entendido em sentido amplo, englobando arte, ciência, técnica. Trata-se, mesmo, da efetiva produção de *cultura*, que emerge das sucessivas lutas da humanidade contra a obstinação dos objetos em se deixar informar. Nesse âmbito, o ato de informar assume um caráter fundamental, confundindo-se à própria condição do homem como ente criador.

Vê-se, assim, a abrangência das múltiplas interpretações para a palavra *informação*,

que vão desde o conceito mais específico e técnico de Shannon até sua acepção como processo fundador, como a entende Flusser. Esse largo espectro sem dúvida traz dificuldades se se busca uma definição consolidada. Entretanto, para uma área de comunicação como o design de informação, que trabalha sob uma perspectiva generalista e se propõe a lidar com fenômenos complexos, a diversidade de acepções (e, portanto, de pontos de vista) pode enriquecer a reflexão sobre sua prática. Do universo de interpretações sobre *informação*, o que talvez se possa deduzir é que este, mais do que outros termos, é um conceito *em formação*. Assim, acompanhar seus constantes desenvolvimentos talvez possa contribuir inclusive para formulações mais adequados às nuances práticas do design de informação.

Da perspectiva específica dessa atividade, o conceito de *informação* assume um viés mais objetivo, ainda que necessariamente amplo. Por conta disso, qualquer formulação descritiva deve prever que, nesse âmbito, *informação* abrange desde respostas triviais a questões como “para onde fica a saída?” até necessidades mais complexas, como o projeto de certas interfaces, livros ou exposições em que os aspectos e atitudes analíticas mais estritas predominam sobre as tendencialmente mais polissêmicas (condição que permite situar tais projetos no escopo do *design de informação*). No entanto, definições com tal abrangência são raras. Normalmente, mesmo no círculo do design de informação, o conceito de *informação* é mais presumido do que explicitado, de modo que é preciso que seja inferido a partir de determinadas propriedades de produtos como gráficos ou visualizações científicas.

Em muitos trabalhos da área de visualização de dados, tradicionalmente ligada aos campos das ciências da computação, não é incomum o entendimento genérico de informação, como tudo aquilo que pode ser transmitido ou armazenado no código binário (o que, pelo comentado, seria designado mais estritamente como *dado*) ou, ainda, definições restritas a contextos específicos. Nesse viés, Card et al. (1999, p.6) propõem que a “visualização de informação é o uso de representações visuais interativas de dados abstratos para amplificar a cognição.”, sugerindo vagamente que a *informação* acontece pelo processo de otimização dos aspectos relativos à aquisição e ao processamento de conhecimento por meio de visualizações.

Também Tufte não se ocupa em definir o termo, algumas vezes usando-o como equivalente a *dados*, mas propõe que visualizar informação (*envision information*) “é trabalhar na interseção da imagem, palavra, número, arte. Os instrumentos são aqueles da escrita e da tipografia, da gestão de grandes conjuntos de dados e análises estatísticas, de linha, layout e cor.” (1990, introdução). Sugere ainda que a informação pode ser encontrada na fuga do plano

(*escape from flatland*), isto é, na propriedade de uma visualização transmitir dados densos e altamente complexos. Sugerem-se assim os meios de transmissão de informação visual e um padrão de exigência no qual a qualidade informativa relaciona-se à quantidade e à complexidade dos dados resultando em suficiente massa crítica capaz para possibilitar frutíferos *insights*.

Wurman, por outro lado, propõe uma definição mais estrita. Para ele, “informação é aquilo que reduz a incerteza, a causa profunda da ansiedade.” acrescentando ainda que “deve ser aquilo que leva à compreensão.”, e ainda acentuando seu caráter relativo e cultural, ele observa que: “Cada um precisa dispor de uma medida pessoal para definir a palavra. O que constitui informação para uma pessoa pode não passar de dados para outra.” (1991, p. 43). Esse conceito é amplo o suficiente para abarcar de uma sinalização a um atlas, dando abertura conceitual para que, mesmo tais exemplos clássicos do design de informação possam ser ignorados, por exemplo, por quem não precisa se orientar ou por quem não sabe ler mapas. Por esta formulação, o conceito absoluto de informação é relativizado em favor de uma situação mais realista, baseada em um encontro específico entre mensagem e destinatário.

Jacques Bertin define *informação* na representação gráfica como “o conteúdo traduzível de um pensamento [que é constituído] essencialmente por uma ou várias correspondências originais entre um conjunto finito de conceitos de variação e um invariante.” (2005, p.9). Trata-se de uma definição acoplada ao seu sistema, inclusive empregando seu jargão (como será visto adiante). Como a definição sugere apenas correspondências entre conceitos, termina por descrever um processo de codificação genérico. Entretanto, *informação* adquire um significado mais abrangente por meio de outro conceito do autor: os *níveis de informação* da imagem (ou *níveis de leitura*). Trata-se das três escalas de observação que podem ser depreendidas de um gráfico ou de um mapa: a *leitura elementar* (que associa dados pontuais), a *leitura média* (que designa os diversos subconjuntos em que os dados podem se dividir) e a *leitura de conjunto* (que simplifica todo o conjunto de dados aos níveis mais significativos). A leitura de conjunto é o estágio de informação mais alto que, quando atingido dá resposta a todos os anteriores. A relevância dessa parametrização está em se referir propriamente às interpretações possíveis diante de um conjunto de dados, estabelecendo de antemão quais escalas podem conter informação. Isso direciona a busca por uma resposta em particular, desmembrando o conceito monolítico de informação em diversas nuances, e também sistematiza, no âmbito dos gráficos e mapas, a relatividade da informação notada por Wurman.

Essas diversas perspectivas para o conceito não devem ser consideradas de modo

excludente e sim complementar. No que concerne a esta dissertação, ela reverbera um entendimento que considera *informação* como um conceito aberto. Em um nível prático, tal perspectiva pode ser exemplificada por projetos que possibilitam múltiplos caminhos de análise e compreensão. Considera-se assim a relatividade contextual do termo, ecoando a noção de Wurman e abrindo mão de um conceito absoluto em favor de uma noção de *percepções informativas*, no sentido de juízos de valor sobre a utilidade de uma mensagem. Todavia, mesmo nesse âmbito de entendimento e, sem perder de vista o entendimento de Logan sobre a “informação não ser um invariante”, há certos tópicos relacionados à comunicação visual de mensagens que são bastante estáveis em contextos específicos, além de outros aspectos que parecem poder, de fato, ser tomados como invariáveis.

Os aspectos mais constantes da comunicação visual são aqueles que possibilitam ao design de informação projetar âmbitos de uso baseados em forte aderência semântica entre a comunicação projetada e sua leitura, minimizando assim interpretações mais polissêmicas. No caso de comunicações imagéticas, essa aderência baseia-se no uso de propriedades da visão que, potencialmente, tenderiam a uma interpretação universal: o que Bertin nomeia como o *nível monossêmico* da percepção visual. A existência de parâmetros visuais de abrangência universal (tendo em conta o universo daqueles que compartilham uma percepção visual considerada normal) é a pedra de toque dentre os autores e praticantes que se ocupam do projeto de comunicações informativas, envolvendo questões como objetividade, neutralidade, retórica e subjetividade. Por exemplo, Tufte considera que “os princípios do design de informação são universais –como a matemática– e não estão ligados a questões particulares de uma língua ou cultura (1990, p.10). Já Bertin, ao propor o *nível de organização dos componentens* declara que eles “formam o campo das significações universais, das analogias fundamentais (...)” (2005, p.10). Tomadas isoladamente, essas declarações, podem sugerir a existência de um único resultado visual necessário, uma “conclusão” decorrente da representação dos dados por tais premissas universais, negligenciando assim os diversos aspectos semânticos, retóricos e persuasivos envolvidos no ato de comunicar. Diversos autores argumentam por esse viés, discussão onde se entrevê certo antagonismo entre universalidade e subjetividade.

Kinross (1985) chama a atenção para a existência de uma “retórica da neutralidade”, mesmo na mais insuspeita informação visual, propondo que os pressupostos universalizantes do design de informação podem ser rastreados até o período modernista do entre-guerras. Por esse entendimento, a ausência de retórica e a universalidade ecoaria a crença modernista em formas

geométricas simples, não ornamentadas, preferência que não se daria por estilo ou preferência, mas por necessidade projetual. Kinross insurge-se contra essa interpretação advertindo que “nada é livre de retórica, que as manifestações visuais emergem de circunstâncias históricas particulares, que não existe o vácuo ideológico.” (ibid. p.29).

Mais recentemente, ao estudar os diversos tipos de retórica presentes nas visualizações de dados, Kostelnick distingue uma “retórica da clareza” e uma “retórica da ciência” presentes nos trabalhos de Tufte, Bertin, Card et al. e Cleveland & McGill que “deriva seus princípios da pesquisa empírica sobre visualizações eficazes.” (2007, p. 281) e teria sua origem em Otto Neurath e seu envolvimento com o positivismo e princípios da *gestalt* visual. O autor chama a atenção que seria possível pensar que essa abordagem racional “seria essencialmente não retórica porque assume que os leitores são universais.”. Por outro lado, o fato de ser orientada ao usuário (*user oriented*) garantiria “benefícios retóricos – sobretudo, a eficiência e a previsibilidade das respostas (...)” (ibid. p.282) Nesse sentido, Kostelnick expande o entendimento de Kinross (baseado em critérios retóricos em relação à forma), entendendo que mesmo a conclusão de um resultado esperado seria uma formulação retórica. Além disso, ele nota que a própria racionalidade dessas formulações incorpora um componente ético, implicando “que o leitor merece uma apresentação íntegra dos dados (...) e que o designer tem um imperativo moral de fornecê-la” (ibid.), e, ainda, um componente estético, no projeto de visualizações belas, elegantes e bem projetadas.

Pesados esses argumentos, convém investigar se as corretas constatações sobre a existência de elaborações retóricas no design de informação são de fato incompatíveis à prerrogativa de níveis de entendimento universal. Nessa discussão, o que está em jogo é a liberdade formal e semântica do designer e do leitor/observador/usuário para projetar e interpretar visualizações, ou, por outras palavras, o grau de determinação desses processos. Nesse ponto, cabe retomar a declaração de Tufte em relação aos princípios do design de informação serem universais e reter o termo *princípios*, pois ele indica que se pode separar o entendimento sobre essa atividade em partes, e assim analisá-las por vieses distintos. É nessa perspectiva que Bertin discute o sistema monossêmico na percepção visual, que se dá quando “o conhecimento da significação de cada signo precede a observação do conjuntos de signos.” (2005, p.6). Ele se refere às significações baseadas em *aspectos fisiológicos* de nossa percepção visual, de modo que se poderia esperar um consenso tácito entre os observadores. Por exemplo, em uma folha de papel, dois círculos iguais em diferentes posições são percebidos como elementos distintos

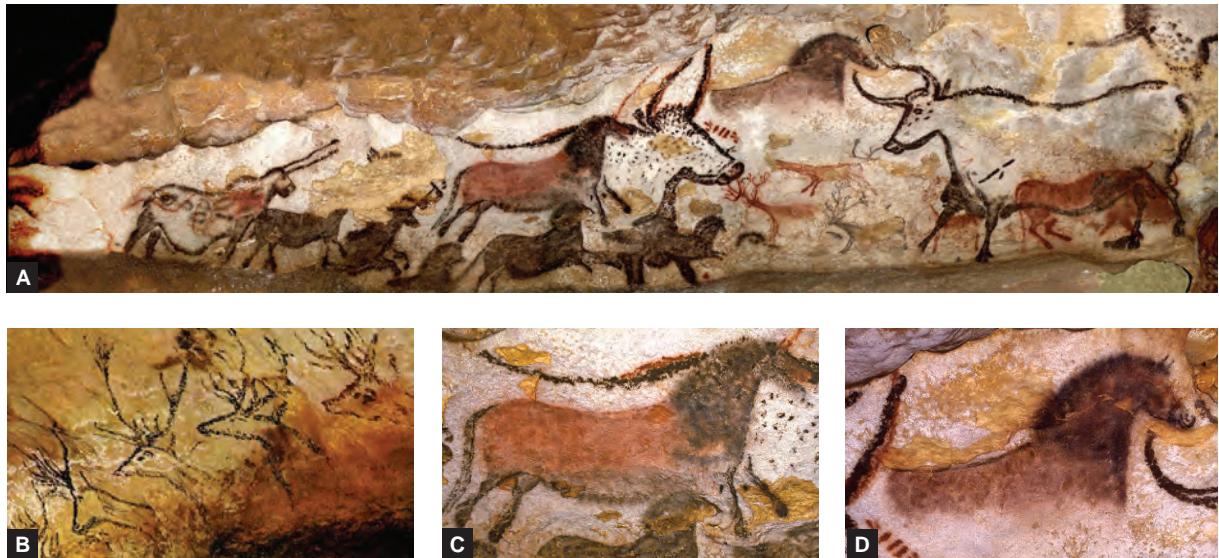
(significação *seletiva*), mas se acrescentarmos um triângulo, os dois círculos serão visualmente agrupados por semelhança (significação *associativa*) diferenciando-se do triângulo (novamente, significação *seletiva*). Dado três quadrados A, B, C de tamanhos progressivamente maiores (significação *quantitativa*) e a tarefa de identificá-los em ordem de grandeza, pode-se fazê-lo de modo crescente A, B, C ou decrescente C, B, A, mas não B, C, A (significação *ordenada*). Existem ainda gradações nessas significações. Em meio a diversos elementos multiformes e multicoloridos, é mais fácil identificar e agrupar visualmente todos os vermelhos do que todos os octógonos. Nesse caso, vê-se que a significação seletiva, presente tanto na variação de forma quanto na de cor, é mais intensa nessa última⁵. Fenômenos como esses constituem um nível básico de significação, composto por relações de maior/menor, sequenciamento, diferenciação e associação que independem de codificações convencionadas. Entretanto, importa ressaltar que se trata de relações em um âmbito restrito, como coloca Bertin: “A monosssemia é a condição da lógica. Mas ela permite também medir seus limites. Com efeito, a monosssemia só pode existir no interior de um domínio finito de objetos e relações.”. (2005, p.7).

Aceitar essa hipótese leva a concluir que a monosssemia, como entendida por Bertin, independe de instâncias culturais ou históricas específicas, uma vez que as interpretações de imagens sob tais aspectos são essencialmente polissêmicas, variando conforme o código particular de uma época e de um lugar. Isso é exemplificado no entendimento de Panofsky sobre a perspectiva como construção simbólica, no qual o autor argumenta pela pluralidade das estruturas perspécticas em diferentes culturas como indícios de visões de mundo igualmente plurais. Ou ainda, na análise de Nelson Goodman sobre a o aspecto essencialmente de *convenção*, presente nas representações visuais realistas. Segundo o autor: “considerando-se imagem e objeto, há geralmente um sistema de representação –um plano de correlação– sob o qual a imagem representa o objeto.”. ou, mais explicitamente: “a imagem, particularmente aquela impressa em papel é um símbolo altamente convencionado, que a criança criada na cultura Ocidental aprendeu a interpretar.” (1968).

Por outro lado, a natureza da comunicação monossêmica parece não ser a da convenção. Não se trata de um “plano de correlação” que funciona por meio de um código, embora a

⁵ Para uma explicação mais aprofundada sobre os níveis de significação, ver os tópicos *Nível de organização dos componentes* (capítulo 2); *Nível de organização do plano*; *Nível de organização das variáveis retinianas* (ambos no Anexo).

Figura 1 – Desenhos na caverna de Lascaux, França



Fonte: LASCAUX: visite de la grotte. République Française, Ministère de la Culture et de la Communication.
Disponível em: <www.lascaux.culture.fr>. Acesso em: 22 jul. 2013.

monossemia embasa a elaboração de tais planos e a formulação de seus códigos, como acontece na correlação entre uma tabela numérica e um gráfico que a visualiza. Sendo assim, como possibilidade de afastamento dos aspectos culturais, talvez seja útil um rápido olhar aventureiro para imagens de um tempo em tudo diferente do nosso, como aquele que produziu os desenhos na gruta de Lascaux, França, procurando entrever indícios de uma percepção monossêmica. A Figura 1A mostra o “painel do unicórnio”, na “sala dos touros” e sugere a compreensão de relações de *proporcionalidade* entre as imagens, revelando que, independentemente da distância temporal entre o desenho das imagens desse painel, a diferença de *tamanho* entre elas foi notada, já que há significativa variação na proporção dos animais, com os touros se sobressaindo aos demais. A Figura 1B mostra o friso dos cervos, na seção conhecida com “a nave”, onde o desenho sugere que o artista previu o significado *associativo* entre as partes individuais, ou seja, que se tratavam todas de cervos, desenhando-as similares, inclusive com todas as cabeças voltadas ao mesmo lado. Parece que compreendia também a propriedade *seletiva* do plano, já que é a separação espacial entre os cervos que permite individualizá-los e percebê-los distintamente. A figura 1C, o “grande cavalo vermelho e negro”, revela um indício do conhecimento da significação *seletiva* da cor, pois o contorno do desenho é delineado em preto e a parte interna, preenchida com tons avermelhados, separando-se assim linha e preenchimento, como elementos distintos. Já o “cavalo marrom” (1D), também no painel do

unicórnio, sugere a compreensão da seletividade por graus de luminosidade diferentes, havendo uma gradação do corpo mais claro, o pescoço e a cabeça mais escuros e a crina ainda mais escura. Nesse sentido, poderia ser aventado que o artista intuía também um sentido de *ordenação* no uso das tonalidades. Por fim, o conjunto de todos os desenhos pode indicar as significações *seletivas* e *associativas* das variações de forma, já que suas diferentes ocorrências pelas salas da caverna podem ser diferenciadas (por exemplo, um cavalo de um touro), mas também reunidas por semelhança (o grupo dos cavalos, dos cervos, dos auroques). A tais observações poderia ser argumentado também que o fato de conseguirmos entender e identificar o aspecto formal desses desenhos, separados de nós por milênios, sugere que há algo em relação à compreensão de imagens que independe de fatores culturais cronologicamente localizados.

Nesse âmbito, o psicólogo Colin Ware (2004, p.5) propõe uma “ciência da visualização”, baseada em uma “semiótica empírica baseada na percepção”. Isso depende da identificação de propriedades da percepção visual que permanecem estáveis independentemente do contexto, e cujos comportamentos e características possam ser validados cientificamente. Ware entende que esse grupo constitui a categoria dos *símbolos sensoriais*, referindo-se aos “aspectos da visualização que derivam seu poder expressivo da sua habilidade de empregar a capacidade de processamento do cérebro dispensando o aprendizado [dos símbolos]”. Por oposição há os *símbolos arbitrários* que “definem aspectos de representação que devem ser aprendidos, pois as representações não têm embasamento perceptual.” (ibid., p.10). Ware explica que “as representações sensoriais são efetivas (ou enganosas) porque elas são bem adaptadas aos estágios iniciais do processamento neural. Elas tendem a ser estáveis em todos os indivíduos, culturas e tempos.” (ibid.). Por esse entendimento, a ocorrência de padrões de percepção visual “universais” daria-se por conta de termos “todos [o] mesmo sistema visual. [Sendo assim,] é provável que nós todos vejamos do mesmo modo, pelo menos como primeira aproximação. Consequentemente, as mesmas visualizações serão efetivas para todos nós.” (ibid., p.12).

A distinção entre *símbolos sensoriais* e *arbitrários* parece correlacionar-se ainda aos dois tipos de comportamentos dos organismos vivos, como colocados por Maturana e Varela: os *instintivos* e os *aprendidos*. Os biólogos apontam que “o âmbito das condutas possíveis de um organismo é determinado por sua estrutura, já que é ela que especifica seus domínios de interação.” (2005, p.191). Assim, considerada uma mesma espécie, caso desenvolvam-se “certas estruturas independentes das peculiaridades de suas histórias de interação, diz-se que

tais estruturas estão geneticamente determinadas, e que o comportamento que elas possibilitam (caso ocorram) são *instintivos*” (ibid., grifo nosso). Por outro lado, se os desenvolvimentos dependem de uma “história particular de interações”, define-se uma *conduta aprendida*.

Assim, aspectos *monossêmicos*, *símbolos sensoriais* e *condutas instintivas*, parecem embasar a hipótese de um domínio que é independente das particularidades culturais dos indivíduos e, desse modo, poderia ser considerado comum a todos os membros de uma mesma espécie. Apesar disso, cabe ressaltar que tais fatores, de modo nenhum, anulam as questões culturais envolvidas no design de informação. Considerar a complementariedade entre as instâncias constantes e as variáveis, e também as especificidades de cada uma, pode desfazer a impressão de haver contradições entre as formulações retóricas e as propriedades universais da visão. Em última análise, equivale a compreender o design de informação sob um ponto de vista mais complexo, considerando como sua especificidade uma natureza comunicativa híbrida, parte determinada monossêmicamente, parte projetada e interpretada polissêmicamente.

Certamente, o design de informação atua em um âmbito em que prevalecem mensagens potencialmente informativas. Aplicando o conceito de informação dos teóricos da comunicação, poderíamos entender que o contraponto a tais mensagens informativas não é apenas que elas possam ser “desinformativas”, significando que havia uma demanda por informação, mas que, apesar disso, foi transmitida uma mensagem falsa, uma não informação. Mas, para além do âmbito da integridade da comunicação, o outro contrário do vocábulo *informação* é o termo *indiferença*, no sentido de não alterar certo estado presente, caso a mensagem não fosse comunicada, ou seja, se fosse *indiferente* saber de tal coisa ou não. Essa observação sublinha a necessidade de certos artifícios retóricos para assegurar que a comunicação aconteça e que o destinatário a entenda como planejado pelo emissor, no caso em questão, como informação. A conjunção entre tais fatores é resumida por McCoy ao considerar que entre os processos comunicativos e retóricos há “uma interação complexa entre as intenções do emissor, o conteúdo da mensagem, as motivações do público/usuário, o contexto das comunicações e as estratégias do designer.” (2000, p.92).

Assim, reafirma-se que a especificidade do design de informação em relação ao design de comunicação deriva da condição de que parte de suas mensagens baseia-se em propriedades mais resistentes a interpretações abertas, e essa constância semântica efetiva-se, sobretudo, por meio de estruturas monossêmicas e de correspondências estritas entre dado e representação. Por outro lado, a contraparte desse processo emprega artifícios retóricos como estratégia de

comunicação para viabilizar, explicar, ressaltar, selecionar, investigar, comentar e comparar os dados. Por essa formulação, atingir o desejado estágio da transformação de dados em informação depende da hábil integração entre os níveis monossêmicos e retóricos.

A presente dissertação situa-se nessa perspectiva e, embora seu escopo não seja um estudo direcionado à retórica dos dashboards ou dos métodos de Bertin, ela perpassa diversas questões semânticas envolvidas nas visualizações propostas. Tais questões aparecem sobretudo no tratamento gráfico proposto por Bertin –em especial, nos aspectos ligados às escalas da representação gráfica– e , ainda, na etapa da *gráfica de comunicação* em que as descobertas da análise são expostas, revelando a interpretação do designer sobre o fenômeno.

Portanto, diante do exposto, reitera-se que a proposta ampla desta dissertação é discutir a possibilidade de que dashboards possam ser projetados embasados pela metodologia de Jacques Bertin. Para isso, são delineadas as características principais de tais interfaces para nortear visualizações que sejam, ao mesmo tempo, compatíveis às demandas dos dashboards e referenciadas pelas propostas de Bertin . Cabe a explicação de que não se trata aqui da elaboração de um dashboard funcional, em razão das diversas questões operacionais que isso envolve e que excedem o escopo de uma pesquisa de mestrado. Entre elas, a programação computacional de tal protótipo e o desenvolvimento de uma interface em todos os aspectos necessários à sua efetividade. Nesse sentido, as visualizações apresentadas não devem ser interpretadas como um produto acabado e sim como uma perspectiva de abordagem inicial. Para tanto, elas devem ser contextualizadas ainda em relação a dois pontos.

O primeiro deles refere-se ao conteúdo típico dos dashboards (dados privados sobre *business*), que aqui foram substituídos por dados públicos sobre os alunos de uma instituição pública. Consequentemente, os usuários costumeiros de dashboards (analistas e executivos) foram substituídos pela figura de um estudioso interessado em políticas afirmativas e que precisa comunicar seus dados. Esse deslocamento é possível se entendermos, como alguns autores, o dashboard como um instrumento de comunicação com suficiente autonomia para abordar aspectos distintos, guardadas as peculiaridades de cada área. Mais ou menos como o *formato* livro pode abarcar diversos assuntos e dirigir-se a diferentes leitores, mantendo suas características definidoras. A despeito de abordar temas de *business*, dados sobre cotas –ou qualquer outro tema da universidade, como titulação dos professores e distribuição dos funcionários– as características de concisão que definem um dashboard são preservadas. Ressalta-se assim, a compreensão ampla dos dashboards para além de um viés puramente

instrumental para entendê-los possivelmente como um tipo de “linguagem”. Trata-se de uma tipologia específica de artefato visual cuja principal característica é o aspecto de síntese (resumindo informações) e de índice (direcionando para dados adicionais).

O segundo ponto diz respeito à aplicação dos métodos de Bertin ao ambiente digital. A *Semilogia gráfica* descreve o tipo de situação imagética com a qual a teoria opera, cuja principal característica é tratar-se de uma imagem estática e bidimensional. Sendo assim, como essa imagem situa-se no âmbito das possibilidades cinéticas do computador? Os recursos da imagem em movimento trariam ganhos para a compreensão da informação? No âmbito das mídias interativas que vinculam conteúdos potencialmente informativos, a variação de estados no decurso do tempo acontece basicamente em duas situações: as animações e as atualizações dinâmicas. No caso das animações, além de serem usados como transição entre estados estáticos contribuindo no entendimento do fluxo de navegação, podem ser aplicadas aos gráficos quantitativos ou mapas para a visualização do desenvolvimento de processos ao longo do tempo. Apesar de tais possibilidades, esses recursos não são abordados neste estudo, que se concentra nas propriedades da imagem estática que conservam sua validade mesmo no ambiente digital.

Entretanto, é na possibilidade de análises baseadas na atualização dinâmica dos dados que se dirige a maior parte dos desenvolvimentos propostos aqui em relação à teoria original de Bertin. Importa notar que um dos aspectos mais característicos dessa teoria é a mobilidade interna da imagem, constituindo uma *gráfica móvel*, no dizer de Bertin. Disso, resulta possibilidades analíticas que se baseiam fortemente na alternância entre múltiplas visualizações de um mesmo conjunto de dados, tarefa que é exponencialmente facilitada pelo uso do processamento computacional. Assim, é proposto aqui que as próprias visualizações funcionem como interface, potencializando o acesso aos diferentes modos de se estudar os dados. Para tanto, indicam-se algumas dessas possibilidades, de modo que, por meio delas, também pode-se sugerir o comportamento da metodologia de Bertin no domínio digital.

Para abordar de modo amplo a metodologia de Bertin, é preciso que as visualizações refiram-se a um conjunto de dados em particular, estabelecendo-se um contexto de referência para as análises e interpretações específicas. Com isso, é possível que as muitas questões que se colocam desde a escolha dos dados até sua comunicação sejam analisadas e discutidas em um âmbito onde essas operações adquirem sentido.

Como mencionado, esse âmbito é fornecido pelos dados sobre o corpo discente da

UERJ, sobretudo aqueles disponibilizados pelo projeto de *Business Intelligence* da instituição, complementados, em alguns casos, pelos anuários estatísticos da universidade. Os dados efetivamente utilizados são uma parte de todo o conjunto disponível, em razão da inviabilidade de lidar com todos eles no contexto desta dissertação. Assim, ela concentra-se nos dados que parecem ter relação com a política de cotas. Como as teorias de Bertin preveem a formulação de hipóteses que delimitam aquilo que um gráfico pode informar, fez-se um breve levantamento da discussão atual em torno das cotas para identificar argumentações contrárias e favoráveis a serem investigados pelos tratamentos gráficos. Isso parece resultar em uma visão dualista (e, portanto, simplista) de uma questão que envolve diversas nuances, não se resumindo a posições a favor ou contra, ou ainda, a dados meramente quantitativos. Mas a demarcação desses polos permite abranger a gama de variantes entre eles consubstanciando análises que ultrapassam as questões quantitativas apresentadas pelos dashboards propostos e também a dicotomia entrevista nas hipóteses iniciais.

Sendo assim, diante do panorama apresentado e de acordo com os objetivos apontados, a presente dissertação estrutura-se em cinco partes principais.

O primeiro capítulo (*Dashboards*) refere-se a tais interfaces apresentando suas características gerais e definições propostas por empresas e autores da área. Discute ainda a inserção dos dashboards no contexto das tecnologias de *Business Intelligence* e também sua condição como artefato de comunicação, investigando seus atributos em relação às prerrogativas da área de design de informação. Por fim, levanta as características consideradas como definidoras dos dashboards, relacionando os parâmetros que norteiam o desenvolvimento desses artefatos no presente trabalho.

O segundo capítulo (*Semiologia gráfica e tratamento gráfico da informação*) trata dos tópicos dos livros de Bertin que mais diretamente relacionam-se ao universo desta dissertação. Nesse sentido, é apresentada a primeira parte da teoria geral da Semiologia gráfica (*Análise da informação*), além de alguns itens da terceira parte (*As regras do sistema gráfico*), como o conceito de *imagem* e dos *níveis de informação*. São descritos também aspectos do segundo livro de Bertin (*La graphique et le traitement graphique de l'information*), especialmente com relação à *análise matricial* e tópicos como *diagramas inúteis*, *construção normal*, *escalas comuns e específicas*, o desenvolvimento de *estudos homogêneos* baseados em uma única tabela de dados e aspectos da *gráfica de comunicação*. No anexo da dissertação, encontra-se

uma síntese dos pontos restantes da teoria geral: a segunda parte do livro (*Os meios do sistema gráfico*), e ainda as *regras gerais de visibilidade*.

O terceiro capítulo (*Proposição*) formula a hipótese geral que será investigada nesta dissertação com base no exposto nos capítulos anteriores.

O quarto capítulo (*Aplicação e discussão*) emprega as metodologias de Bertin apresentando propostas de dashboards com informações relativas a três características principais sobre o sistema de cotas. São elas: um panorama da universidade antes e depois da implantação do sistema; um comparativo entre características de alunos cotistas e não cotistas em questões relacionadas ao desempenho e um quadro da ocupação dos cursos da UERJ por cotistas. Com esse contexto, propõe-se algumas operações de interação e uso desses dados, indicando outras possibilidades para os recursos de detalhamento (*drill-down*) tradicionalmente presentes nos dashboards. Empregando exemplos visuais, são discutidas as principais características das visualizações propostas apresentando um encaminhamento de análise possível baseado em tais visualizações.

1 DASHBOARDS

1.1 Características gerais e definições

O termo *dashboard* como usado neste trabalho refere-se aos painéis que apresentam dados de natureza estatística, predominantemente por meios visuais como acontece em gráficos e mapas. A palavra, de origem inglesa, é uma justaposição dos vocábulos “dash” (salpico de lama) e “board” (placa). Referia-se à tábua colocada na frente da carruagem para impedir que os respingos de lama levantados pelos cavalos resvassem no condutor⁶. Com a introdução do automóvel, o termo passou a designar o painel de instrumentos do veículo, componente que agrupa os indicadores necessários à sua condução. Daí, o termo *dashboard* passou às áreas da computação e da informática com o significado amplo de “central de controle”. Mantendo essa acepção genérica, dashboards têm características específicas conforme o segmento onde são empregados, apresentando funcionalidades, objetivos e repertório visual próprios a determinado domínio. Desse modo, aquele que se detém nos artefatos referidos como dashboards, irá se deparar com relativa diversidade funcional e formal. Talvez o denominador que permita ao senso comum associar todas as ocorrências de dashboards seja a visualização conjunta de uma multiplicidade de elementos que podem ser acessados a partir de uma tela. Nesse âmbito, são denominados dashboards tanto a interface do sistema operacional *OS X* que acessa pequenos aplicativos (*widgets*) como calculadora, relógio etc., quanto a tela para as configurações do console de videogame *XBOX* ou ainda um painel com gráficos estatísticos sobre o desempenho de uma empresa. O presente trabalho concentra-se no âmbito desses últimos. São dashboards empregados prioritariamente com a finalidade de analisar, monitorar e visualizar dados estatísticos. Tais artefatos disseminaram-se com a implantação dos sistemas computadorizados de gestão, nos quais o dashboard é uma das interfaces com o usuário.

As diversas tecnologias de gestão desenvolvidas entre os anos 1980 e 1990 consolidaram-se em tipologias de sistemas que hoje são designados sob o termo *Business Intelligence*

⁶ *Online Etymology Dictionary*. Disponível em: <<http://www.etymonline.com>>. Acesso em: 30 ago. 2013.

(frequentemente referido pelo acrônimo BI⁷). Em termos gerais, designa a atividade que, por meio de suporte computacional, habilita o uso estratégico dos dados coletados pelas empresas visando melhorias na performance da instituição. Cabe notar que mesmo nesse domínio, o termo dashboard não é inequívoco, designando, por exemplo, tanto dashboards interativos que permitem navegação em seus dados, quanto um folha impressa com a imagem do dashboard digital. Além disso, não raro, há confusões entre dashboards e outros instrumentos de comunicação de dados empresariais como *balanced scorecards* e relatórios com gráficos estatísticos. Mesmo no domínio estrito dos dashboards interativos, há distinções devido aos protocolos, às necessidades e à cultura de uso de cada área que os utiliza. Assim, há dashboards na área médica cujo objetivo é monitorar processos em um hospital; há os que atendem à demanda por acompanhamento do fluxo de tráfico na internet, empregados nas ferramentas conhecidas como *analytics*. Existem ainda os que se dedicam a apresentar informações sobre as atividades empresariais como despesa, lucratividade e desempenho de setores e funcionários. A Figura 2 apresenta, apenas para referência visual, algumas ocorrências de dashboards de diferentes áreas.

No âmbito do BI, os vários segmentos de aplicação e de ferramentas para construção de dashboards resultam em uma diversidade de situações de uso, visualizações e layouts. Assim, sintetizar uma definição suficientemente abrangente depende de localizar os elementos comuns a essa grande variedade de dashboards, ressaltando aspectos gerais, como acontece no verbete do Gartner Group⁸ em seu glossário de termos relacionadas à BI:

Dashboards são um mecanismo de comunicação que agregam e exibem métricas e indicadores-chave de desempenho (KPIs – *key performance indicators*), permitindo seu exame em um relance por diversos tipos de usuários preliminarmente a outras investigações que empreguem outros recursos de *business intelligence*, gerenciamento de desempenho (PM – *performance management*) e ferramentas analíticas. Dashboards ajudam a otimizar a tomada de decisão revelando e comunicando uma visão contextual sobre o desempenho dos negócios. Eles exibem KPIs ou métricas de negócios utilizando visualização intuitiva, incluindo gráficos, mostradores, medidores e “semáforos” que indicam o progresso dos KPIs em relação às metas definidas. Permitem que os usuários se aprofundem (*drill-down*) através dos sucessivos níveis de detalhe definidos pelo desenvolvedor do painel, para analisar um KPI fora da meta. Dashboards podem ser implantados em uma miríade de tipos de dispositivos (desktops, laptops, tablets e / ou smartphones) em modo conectado ou desconectado.
(DASHBOARD In: GARTNER. Disponível em: <<http://www.gartner.com/it-glossary/dashboard/>>. Acesso em 07 set. 2012)

⁷ Apesar de algumas ocorrências do termo *Business Intelligence* traduzido por *inteligência de negócios*, esta não é uma prática frequente na literatura em português, razão pela qual esta dissertação emprega a expressão original

⁸ O *Gartner Group* é uma renomada consultoria empresarial americana sediada em Stanford, Connecticut, à qual é atribuída a introdução do termo *Business Intelligence*, em meados dos anos 1990.

Figura 2 – Visualidade característica de dashboards para aplicações diversas



A Métricas financeiras e vendas; **B** Gestão de recursos humanos; **C** Ciclo de pagamento de entidade governamental; **D** Monitoramento de recursos energéticos; **E** Gestão de impressoras em rede; **F** Alocação de ativos e outros dados mercadológicos; **G** Gerenciamento de projetos; **H** Monitoramento do sistema de trânsito.

Esse verbete indica que o principal objetivo de um dashboard é a melhoria dos processos de tomada de decisões sugerindo que essa meta é atingida quando se transmite uma visão contextual dos dados por meio de visualizações específicas e recursos interativos. Essas características delimitam a especificidade de um dashboard distinguindo-o dentre outros instrumentos empregados na área de BI. Mas a variabilidade formal e funcional dos dashboards sugerem que se leve em conta definições mais específicas.

Cabe ressaltar que o universo do BI, onde se concentra a literatura sobre dashboards e, portanto, algumas tentativas de definições mais consequentes, envolve acirrada competição entre fabricantes e prestadores de serviço. Nessa arena, eles disputam a venda de serviços/infra-estrutura/*know-how* para grandes empresas, atividades que em muitos casos envolvem valores da ordem dos milhões de dólares. Como será detalhado à frente, dashboards posicionam-se no “bojo” de uma estrutura complexa de ferramentas para análise de negócio constituída por diversos componentes. O objetivo dos grandes fabricantes (dentre os quais *Oracle*, *SAP*, *Microsoft*) é a venda de *suítes* completas de ferramentas argumentando pela ótima integração entre os componentes do seu sistema. Entretanto, módulos de diferentes fabricantes podem ser adquiridos em separado para integrar uma nova estrutura de BI de uma empresa ou se acoplar a um sistema já existente. Neste cenário, dashboards podem ser projetados e operados, prioritariamente, por três tipos de *softwares*: 1) por aplicativo específico, parte de uma grande plataforma, como o *Crystal Dashboard Design* do fabricante *SAP*; 2) por *software* independente, como as ferramentas do fabricante *Visual Mining, Inc.*; 3) por um aplicativo capaz de comunicação com banco de dados e de operações com planilhas e gráficos, como o *Microsoft Excel*. Uma quarta possibilidade, ainda que não seja prática corrente, é a construção do dashboard em qualquer software gráfico que permita a construção de gráficos a partir de dados tabulares. Aliás, essa possibilidade ressalta, pelo menos em um plano conceitual, a independência do artefato dashboard em relação a tecnologias específicas.

Assim, em torno de critérios gerais, exemplificados no verbete do *Gartner Group*, cada fabricante enfatiza determinadas características e funcionalidades de um dashboard com relação aos recursos dos seus *softwares*. Por sua vez, os autores que se ocupam da elaboração de dashboards, geralmente escrevem manuais técnicos, nos quais o dashboard é tratado no contexto de um programa específico. Percebe-se que as definições de dashboard orbitam um núcleo comum que, se por um lado é difuso, por outro, parece ser suficientemente flexível para novas variantes, como a proposta do presente trabalho. De todo modo, circunscrever minimamente este núcleo depende da análise de uma amostragem das definições de diferentes autores. Definições que, longe de serem antagônicas, são semelhantes e complementares, variando na ênfase sobre este ou aquele aspecto. Daí, a necessidade de sintetizá-las como ponto de partida para novas propostas.

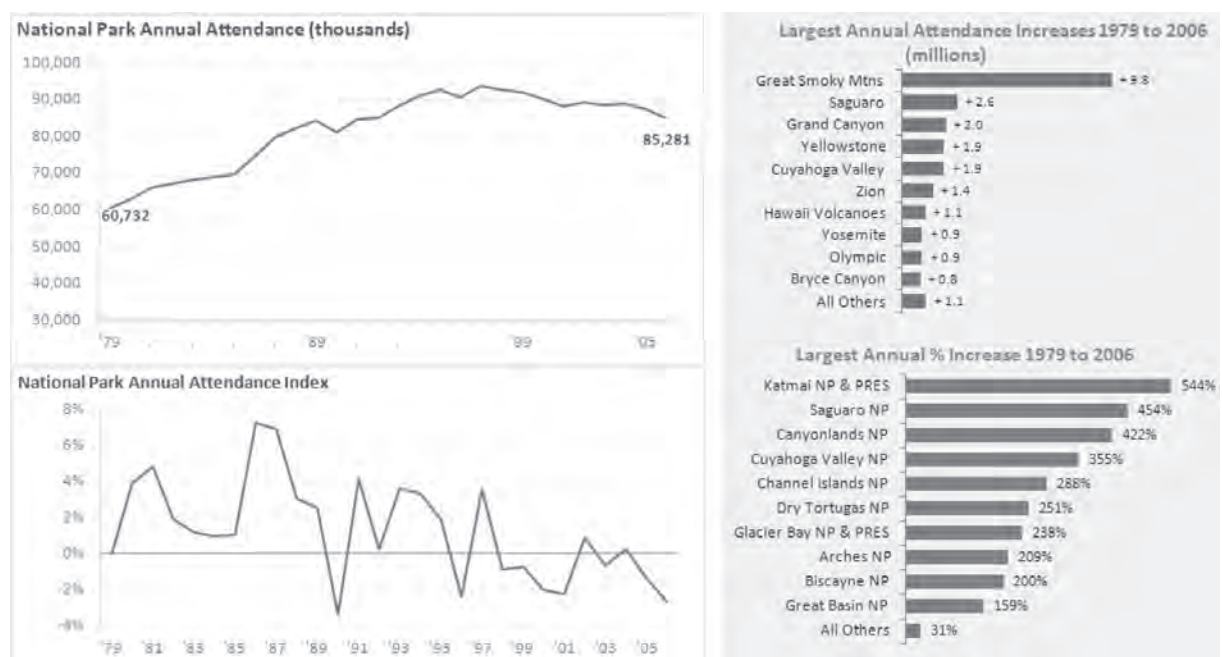
Por exemplo, Alexander e Walkenar (2010) entendem dashboards como instrumentos que comunicam conclusões de forma que o usuário seja dispensado de análises detalhadas.

Salientam que este aspecto diferencia os dashboards dos relatórios (*reports*) frequentemente usados em BI. Segundo os autores, o atributo principal de um relatório é “não conduzir o leitor a uma conclusão pré-definida.”. Isso demanda que o usuário aplique “seu próprio julgamento e análise aos dados” (ibid., p.12), pois os relatórios limitam-se a apresentar uma determinada configuração dos dados. Eles não direcionam a atenção para fatores que possam merecer atenção. Por outro lado, o dashboard é apontado como “uma interface visual que permite vistas de relance sobre métricas relevantes para um objetivo particular ou processo de negócio.” (ibid. p.12). Nesse sentido três propriedades básicas definem um dashboard, segundo os autores:

- 1 Apresentam dados visualmente (como em gráficos). Permitem visualizações que auxiliam o foco nas principais tendências, comparações e exceções.
 - 2 Apresentam apenas os dados relevantes para o objetivo do dashboard
 - 3 Contém conclusões pré-definidas que são relevantes para o objetivo do dashboard e desobrigam o leitor de fazer suas próprias análises.
- ALEXANDER; WALKENBACH, 2010 p.12.

Os autores ilustram tais atributos no dashboard da Figura 3, com estatísticas sobre a frequência de visitantes dos parques nacionais americanos. Uma visualização que permite “reconhecer rapidamente uma tendência global.” (ibid., p.13). Além disso, é seletiva na escolha

Figura 3 – Dashboard sobre frequência de visitantes nos parques nacionais americanos



Fonte: ALEXANDER; WALKENAR, 2010, p.13.

dos dados apresentando: “somente as principais partes de informação que suportam a meta deste dashboard.” (ibid. p.13). Por fim, “apresenta análise e conclusões sobre a tendência de frequência” [dos visitantes].

Por seu turno, Ron Person entende os dashboards como elementos que influenciam ativamente aspectos da cultura de uma corporação. Por uma perspectiva ambiciosa, na qual dashboards atuam como vetores de uma mudança comportamental na empresa, o autor acredita que “[...] dashboards deveriam ser a alavanca que muda a cultura de uma organização.” (2009, p.107). Deste modo, quando “dashboards são usados como ferramentas para construir organizações de aprendizagem⁹ de alta performance, eles monitoram performance, cultura e o modelo de negócios.” (ibid., p.109). Para tanto, o autor aponta as principais diretrizes para dashboards que propiciem a mudança pretendida resultando em melhor desempenho nos negócios:

- 1 Ser baseado nas relações causais que efetivam os objetivos organizacionais
- 2 Deduzir suas métricas usando processos científicos
- 3 Otimizar a velocidade, facilidade e precisão da tomada de decisões
- 4 Utilizar dados oportunos habilitando decisões que assegurem o cumprimento dos objetivos definidos
- 5 Comunicar o comportamento adequado à cultura da empresa
(PERSON, 2009, p. 107-108).

O conceito de “dashboard de desempenho” (*performance dashboards*) é sugerido por Wayne Eckerson como adequado a todas as situações de uso de um dashboard na área de BI. Para o autor “dashboards traduzem a estratégia da organização em objetivos, métricas, iniciativas e tarefas customizadas a cada grupo e indivíduo da organização.” (ECKERSON, 2011, p. 4). Tal como Person, este autor também sublinha o papel do dashboard na cultura da organização destacando o uso desse instrumento como um sistema de gerenciamento de desempenho que “comunica objetivos estratégicos e habilita o pessoal de negócios a medir, monitorar e gerenciar atividades chave e processos necessários para o cumprimento de suas metas.”. (ibid.). O autor ainda aponta três funcionalidades principais de dashboards:

- 1 **Monitorar** processos de negócios críticos e atividades usando métricas que disparam alertas quando a performance cai abaixo dos alvos pré-definidos.
- 2 **Analisar** a causa de problemas pela investigação de informação relevante e oportuna a partir de múltiplas perspectivas em vários níveis de detalhe.
- 3 **Gerenciar** pessoas e processos pra otimizar decisões, desempenho e conduzir a organização na direção correta. (ECKERSON, 2011, p. 5).

⁹ No original, *learning organization*. Termo cunhado pelo professor de Harvard, Chris Argyris, para designar as empresas que aprendem, desenvolvendo novas perspectivas e valores, à medida que os seus funcionários adquirem novos conhecimentos.

Pelo lado dos fabricantes, a definição proposta pela empresa *Visual Mining*, especializada em softwares de visualização, ressalta o aspecto agregador de dados dos dashboards. Por conta dessa característica, ele atua como o elemento de síntese de dados provenientes de diferentes origens.

Um dashboard é uma coleção de gráficos ou métricas tabulares de desempenho de negócio. Geralmente, eles retiram seus dados de múltiplas fontes para fornecer uma abrangente visão do ambiente de um usuário de negócios. (VISUAL MINING, 2006, p. 4).

Aponta ainda que um dashboard deve se basear em perguntas específicas e o foco do seu desenvolvimento está em responder perguntas como “Quais produtos não estão vendendo?”, “Para quais mercados acontecem as maiores vendas?”. O fabricante aponta a importância da formulação de questões específicas ao invés de colocações com sentido geral. “Tentativas de construir um dashboard baseadas em solicitações como ‘Apenas preciso de melhores dados de receita’ causarão infundáveis problemas.” (ibid.).

A grande empresa alemã de softwares de gestão de empresas, *SAP*, enfatiza nas suas diretrizes para projetos de dashboard, a propriedade de apresentar uma visão panorâmica a partir da qual pode-se aprofundar em investigações mais específicas:

Dashboards têm o propósito de cumprir um papel específico por meio de uma bem estruturada visão global sobre uma área específica [...] com orientação visual para importantes fatos de negócios possibilitando e conduzindo a navegação para detalhes adicionais [...]. (SAP AG, p. 5, 2011).

A empresa descreve ainda o aspecto visual característico dos dashboards. Trata-se de um “Layout geral composto de um ou mais painéis [no qual cada] painel apresenta informação gráfica, tal como um diagrama, gráfico, tabelas como micro-diagramas (*micro-charts*) ou uma coleção de indicadores chave de performance (KPI) [...]” (ibid. p.5). Lista ainda o que considera como as três “Características de bons dashboards”. *Simplicidade*, não pela redução de informação, “mas pela orientação clara aos fatos mais importantes”. *Acionabilidade*, por meio de “recursos básicos de interatividade incluindo a navegação no dashboard; apresentação progressiva da informação; ações de clique único e navegação adicional para análise detalhada e tarefas transacionais.”. E ainda *customização*, recursos disponibilizados para “assegurar que a informação apresentada é relevante para o trabalho em questão.”.

Assim, a partir das definições apresentadas, pode-se confirmar o dashboard como um artefato que apresenta dados por meios visuais. Esses dados devem permitir uma visão

abrangente dos fatos em questão, sendo representados por indicadores projetados criteriosamente para informar de modo preciso, em consonância aos objetivos da instituição. Importa ainda a prontidão na comunicação, critério atendido quando o dashboard sugere ao usuário conclusões pré-definidas, mas também recursos que permitam investigar e aprofundar a visão geral apresentada. Cabe ressaltar por fim que, apesar de todas as definições mencionarem visualizações como o elemento fundamental dos dashboards, em geral, não mencionam quais tipos de gráficos, mapas e diagramas devem ser empregados, deixando tais aspectos em aberto.

1.2 Dashboard no contexto de *Business Intelligence*

As tecnologias computacionais que se propõem a auxiliar decisões empresariais procuram responder às pressões relativamente recentes que a expansão do mercado global e de novos comportamentos de consumo vem impondo às corporações. Acirrada competição em âmbito transnacional, exigentes demandas do consumidor e da sociedade, pressões das regulamentações governamentais são vetores que, em conjunto, constituem um complexo ambiente de negócios envolvendo restrições e oportunidades. Nesse panorama, a premissa fundamental das tecnologias de *Business Intelligence* é a possibilidade de obter vantagens competitivas por meio da adequada gestão das informações empresariais.

Numa acepção ampla, *Business Intelligence* é o segmento da Tecnologia da Informação que se ocupa da coleta, armazenamento, análise e disponibilização dos dados empresariais visando embasar decisões estratégicas. A introdução do termo, em meados dos anos 1990, é atribuída à consultoria *Gartner Group* que, no glossário de seu *website*, define Business Intelligence como:

Um termo guarda-chuva que inclui as aplicações, arquiteturas, infraestrutura e ferramentas que permitem o acesso e a análise às informações para melhorar e otimizar decisões e desempenho. (BUSINESS INTELLIGENCE. In GARTNER. Disponível em: <<http://www.gartner.com/it-glossary/business-intelligence-bi/>> Acesso em 07 set. 2012).

Esse conceito engloba arquitetura de dados, ferramentas de manipulação, bancos de dados, aplicações e metodologias. Turban et al. (2009) apontam que ferramentas de gestão empresarial e suporte computadorizado às rotinas corporativas que eram inicialmente estáticas, bidimensionais e sem recursos de análise, porém as tecnologias sofisticaram-se possibilitando

relatórios dinâmicos e multidimensionais, previsões, detalhamento, análises de tendências etc. A difusão desses recursos entre um número de produtos comerciais cada vez maior em conjunção ao aparecimento de novos fabricantes e a constante evolução tecnológica conforma um panorama complexo, que dá ensejo à designação dessas ferramentas como *Business Intelligence*.

Ainda que tais ferramentas estejam disponíveis a empresas de diferentes escalas é em especial nas grandes corporações (onde tecnologias capazes de processar quantidade expressiva de dados estabelecem o contraponto aos diversos mecanismos de obtenção de dados) que são encontrados os sistemas mais completos, robustos e integrados. Face ao crescente acúmulo de dados sem tratamento e comunicação apropriados, as práticas de BI emergem como resposta à tal demanda prometendo suporte computadorizado para os processos de tomada de decisão. Sua prerrogativa é justamente lidar com o acúmulo de dados informatizados, mas sem destinação adequada. Nessa situação, tais dados não se apresentam convenientemente para o uso efetivo, permanecendo como imenso arquivo, cuja falta de praticidade em acessá-lo conforme as necessidades reais pode resultar na sua ineficiência como recurso estratégico. Turban et al. exemplificam essa problemática relatando que, na década de 1980, usuários de sistemas de gerenciamento de negócios “referiam-se a seus *mainframes* como ‘o buraco negro’, pois as informações entravam ali, mas nenhuma jamais saía.” (ibid., p.37). Contam ainda que obter um relatório dependia de sua programação prévia pela equipe de TI. “Consultas *ad hoc* em tempo real eram praticamente impossíveis.” (ibid.).

Em tal contexto, os autores descrevem como objetivos principais do BI: “Permitir o acesso interativos aos dados (às vezes, em tempo real), proporcionar a manipulação desses dados e fornecer aos gerentes e analistas de negócios a capacidade de realizar a análise adequada.” (ibid. p.27). Assim, as ferramentas de BI devem possibilitar que os dados armazenados estejam disponíveis a quem lhes couber e em tal forma que permita a manipulação e a análise necessárias. Nesse sentido, tais ferramentas viabilizam estudos sob múltiplas perspectivas, efetivando correlações entre aqueles que se supõe manter relações de causa e consequência revelando padrões significantes ou, ainda, anomalias e exceções. Outro aspecto importante são as características de uso das interface onde os dados são acessados e analisados. Essa interface deve ser conveniente, disponibilizando rapidamente uma pesquisa requisitada e cômoda, apresentando-se facilmente inteligível em relação às possibilidades de uso e acesso às funcionalidades. Aqui, despontam algumas características que se entremeiam ao âmbito do design de interação e da usabilidade desses sistemas. Em última instância, tais recursos

objetivam fazer com que os dados armazenados possam ser percebidos por seus usuários como *informação* e, sendo assim, deem suporte à tomada de decisão que resulta em ação efetiva, o que, no contexto empresarial, significa a melhoria de performance em algum aspecto.

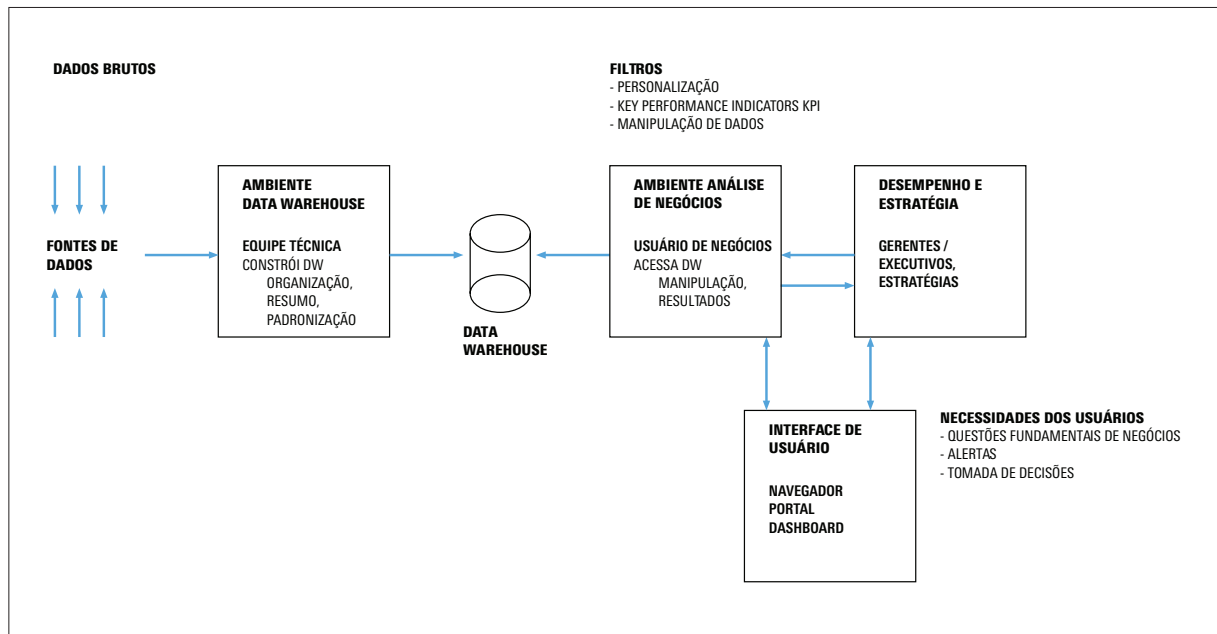
1.2.1 Dashboards na arquitetura de um sistema de *Business Intelligence*

A multiplicidade e abrangência da prática de BI envolve métodos, técnicas e ferramentas. Esses itens aparecem em diferentes equacionamentos e a ênfase sobre um ou mais deles varia conforme o praticante e a especificidade da situação operacional e do contexto de implantação. Trata-se de condições que propiciam que diversas abordagens sejam designadas genericamente como *Business Intelligence*, não havendo um padrão único das partes que constituem um sistema de BI. Entretanto, Turban et al. (2009, p.28) apontam os componentes presentes em uma arquitetura de alto nível de BI. São eles: um grande banco de dados conhecido como *data warehouse*, ferramentas para extração de dados de tal dispositivo e uma interface com o usuário, itens esquematizados na Figura 4.

Data warehouse: um repositório computacional de dados não voláteis com alta capacidade de armazenamento. Trata-se de um dispositivo de memória que centraliza dados provenientes de diversas fontes, como os diversos departamentos de uma empresa e mesmo dados de origem externa, como informações demográficas e geográficas disponibilizadas por terceiros. Tais dados são consolidados para fornecer o suporte adequado para as aplicações de tomada de decisões e, em geral, para impedir alterações após sua inserção no *data warehouse*, condição que permite uma única versão dos dados. Segundo Turban et al. “as aplicações variam de simples gerações de relatórios ou consultas [até] complexas otimizações” (2009, p.29).

Análise de negócios e ferramentas de manipulação de dados: A análise dos dados que objetiva responder às questões práticas de negócios se insere no contexto da *Análise de Negócios* (ou *BA – Business Analysis*). Segundo Turban et al. (ibid., p.101) trata-se da atividade de análise “da decisão por meio de acesso a todos os dados e informações relevantes” e, ainda, de “uma ampla categoria de aplicações e técnicas para reunir, armazenar, analisar e fornecer acesso aos dados [auxiliando] os usuários [a tomar] melhores decisões comerciais e estratégicas.” (ibid., p.104). Os autores listam dentre tais aplicações e técnicas o processamento analítico de dados

Figura 4 – Componentes de um arquitetura de alto nível de BI



Fonte: autor (adaptado de TURBAN et al., 2009)

online (OLAP), data mining, sistemas de informação geográfica (GIS), multidimensionalidade, visualização de dados e ainda técnicas de análise avançada. Resumidamente, as funcionalidades dessas ferramentas incluem:

- 1 a possibilidade de consultas sob demanda e elaboração de relatórios (OLAP),
- 2 a identificação de padrões significativos e eventualmente análise preditiva (data mining e técnicas avançadas),
- 3 a integração entre objetos e sua localização geográfica (GIS) precisa,
- 4 a análise e comparação de dados em múltiplas perspectivas (multidimensionalidade)
- 5 comunicação rápida por meio de gráficos estatísticos (visualização de dados)

Nesse panorama, esta dissertação está inserida no âmbito da visualização de dados organizados de forma multidimensional por meio de ferramentas OLAP, conforme será detalhado no capítulo que descreve a tecnologia empregada pelo projeto de BI da UERJ.

Interface com o usuário: O terceiro elemento de uma arquitetura de alto nível de BI são as interfaces que apresentam os dados e permitem sua manipulação. Incluem navegadores, portais e dashboards. Nesse nível operam todas as condicionantes envolvidas no design de

interação dessas interfaces, incluindo aspectos de arquitetura da informação e usabilidade. Dentre as interfaces empregadas, os dashboards são aquelas voltadas aos gestores e demais tomadores de decisão apresentando de modo sintético uma visão significativa sobre o negócio ou algum segmento dele.

1.3 Dashboard como artefato de comunicação

O posicionamento do dashboard na cadeia de BI como interface com o usuário assinala seu caráter como artefato de comunicação. O verbete do *Gartner Group* descreve esse atributo (“mecanismo de comunicação”) como característica definidora, constatação presente de modo mais ou menos explícito nas demais definições. Alexander e Walkenar (2010, p.14) são claros quanto a isso, apontando como importantes requisitos no desenvolvimento de dashboards, a identificação e coleta das necessidades dos usuários, caracterizando como tarefas prioritárias “definir a mensagem” e “estabelecer o público”. Nesse sentido, fica claro que adequar os dados às necessidades informativas dos usuários, concentrando-se em aspectos fundamentais da mensagem é uma importante prerrogativa dos dashboards. Trata-se de um objetivo amplo que demanda investigações adicionais por meio de recursos interativos e o uso de visualizações idealmente compreensíveis em um golpe-de-vista. Essa meta aponta para campos de estudo autônomos em relação à BI justificando o estudo dos dashboards sob diferentes pontos de vista. A condição do dashboard como interface interativa sugere abordá-lo pelos viés do design de interação e de disciplinas correlatas como a arquitetura da informação, a usabilidade e a análise ergonômica de interfaces. Por outra vertente, apresentar visualizações eficientes, amigáveis e informativas envolve um estudo das tipologias de visualizações que melhor atendem a uma situação de uso pré-determinada, estudo que tende ao segmento do design de informação e da visualização de dados.

Seja considerando aspectos interativos ou as questões de visualização, sublinha-se a natureza de comunicação do dashboard como veículo de comunicação. Nesse âmbito, verificam-se a ausência de estudos e propostas mais detalhadas sob esse viés analítico. Fabricantes enfatizam as funcionalidades do seu produto concentrando-se, como tende a acontecer na área de informática, na oferta de recursos sofisticados e abrangentes, mas que em muitas situações

não correspondem aos melhores critérios de uso. Autores da área de BI concentram-se no ensino de metodologias e na determinação de objetivos, mas poucos discutem em profundidade e conseqüentemente quais as melhores visualizações para que tais metas sejam atingidas ou de que modo uma visualização pode ser mais ou menos informativa e eficiente do que outra e, nesse sentido, contribuir para a compreensão efetiva de uma situação.

Um dos raros trabalhos sobre os aspectos visuais em dashboards foi escrito por Stephen Few, autor especializado na visualização de dados financeiros. Em um primeiro momento, o autor procurou definir de modo amplo um dashboard enfatizando a eficiência da sua natureza comunicativa como uma propriedade significativa:

Um dashboard é uma apresentação visual das informações mais importantes para a conclusão de um ou mais objetivos; [tais informações estão] consolidadas e arranjadas em uma única tela de modo que [...] possam ser monitoradas em um relance. (FEW, 2006, p.34).

Essa definição parece especialmente adequada às considerações projetuais sobre dashboards, abrindo um viés de entendimento que se descola da perspectiva restrita à área de BI. A definição de Few é flexível porque se baseia em critérios gerais de forma e *layout*, ao invés de citar visualizações específicas, como faz o verbete do *Gartner Group* e, além disso, não determina a natureza dos objetivos e sim a finalidade de cumpri-los. Por fim, introduz um critério de avaliação que permite balizar projetos de dashboard, tópicos analisados a seguir.

O trecho em questão ressalta ainda a característica de um dashboard como instrumento de visualização, ou seja, um artefato que apresenta dados de modo predominantemente visual. Isso não exclui a presença de elementos textuais sintéticos como tabelas ou textos topificados, mas enfatiza o aspecto visual de um dashboard, inscrevendo-o também na área mais ampla da *visualização de dados*, inclusive por se tratar, no mais das vezes, de uma interface digital.

A definição indica de modo geral o propósito dos dados apresentados, estabelecendo um critério para sua escolha: a potencialidade para, em um contexto particular, serem interpretados como informação. Few aponta como finalidade dos dashboards o cumprimento de “um ou mais objetivos”, condição que sugere apenas a inclusão de dados que contribuam para a compreensão desses objetivos. Além disso, ainda que de modo vago, essa noção sugere critérios para uma tipologia de dashboard baseada tanto em objetivos gerais, como monitorar, analisar, comunicar como em objetivos específicos de cada aplicação de dashboard. Como detalha Few, mais adiante: “Não se trata apenas da informação que é solicitada por executivos ou [...] gerentes; pode ser

informação que é necessária por qualquer um que tem objetivos a cumprir.” (Few, 2009, p.35).

Já a restrição a uma única tela é uma significativa peculiaridade formal dos dashboards ressaltada pelo autor. As implicações projetuais desta restrição sugere um planejamento específico para garantir que todos os elementos apresentados caibam em uma única superfície, abarcada em um golpe de vista. Isso exclui movimentações horizontais ou verticais (*scrolling*) para revelar partes das interface ocultas em um primeiro momento. “O objetivo é obter as informações mais importantes prontamente e sem esforço para que se possa absorver rapidamente o que e preciso saber.” (ibid. p.35). Tal restrição torna-se particularmente importante se consideramos a variedade formal dos suportes digitais atualmente disponíveis. Além do *desktop* padrão, o uso de interfaces portáteis como *tablets* e *smartphones* para atividades profissionais, sugere a dificuldade para adequar dashboards a telas de pequeno formato sem fazer uso do recurso da “rolagem”. Como veremos ao longo deste trabalho, esse é um problema projetual que demanda além da análise dos aspectos da visualização dos gráficos, um planejamento que priorize os dados estritamente necessários em razão não apenas da potencialidade informativa, mas também do espaço disponível para sua visualização.

Ainda segundo Few “um dashboard deve ser capaz de apontar rapidamente algo que mereça atenção e pode necessitar uma intervenção” (ibid. p.36). Geralmente, essa condição é satisfeita quando são incluídos elementos visuais destacando os tópicos que merecem detalhamento. Baseado em tais critérios, o autor propõe dashboards como os da Figura 5, dispensando gráficos na forma de velocímetros ou termômetros, usuais em dashboards convencionais, mas considerados por Few como vícios formais pouco informativos que consomem o precioso espaço restrito de uma única tela.

Sobre a questão da rapidez de leitura, ou seja a faculdade de se perceber prontamente algo que se quis comunicar, ela pode ser entendida não apenas em relação aos indicadores que apontam para um tópico potencialmente problemático, como os pontos vermelhos em Few, mas também às próprias visualizações, sendo um critério para avaliá-las segundo a velocidade da sua compreensão. Desse modo, é possível usá-lo como elemento de comparação entre diferentes visualizações para que seja possível qualificá-las, condição aventada por Jacques Bertin. Assim, o dashboard pode ser julgado também pelo modo como seus gráficos são mais ou menos rápidos na forma que evidenciam um padrão informativo.

1.3.1 Dashboard no contexto do design de informação

Figura 5 – Proposta de dashboard apresentada por Stephen Few



Dashboard de vendas: métricas diversas.

Fonte: FEW, 2009, p.177

Importa ainda relacionar dashboards ao campo de estudo do design de informação, área com a qual a comunidade do BI não está particularmente familiarizada de modo que tende a não considerá-la nas suas análises. Entretanto, uma vez que o objetivo deste trabalho é justamente aplicar conceitos de um autor clássico da área do design de informação aos dashboards, será feita uma breve contextualização do campo de modo a evidenciar o dashboard como um produto facilmente inserido no seu escopo de estudo e prática.

Ainda que com contornos indefinidos, o design de informação é o segmento do design que lida com os aspectos que visam facilitar a compreensão de dados, notadamente dados de relativa complexidade, de tal modo a potencializar seu entendimento como informação. Com essa finalidade, a atividade consiste da elaboração de estratégias que favoreçam o entendimento, em situações em que explicações puramente verbais ou textuais são insuficientes. Algumas imagens produzidas com esse objetivo tornaram-se emblemáticas da visualidade característica

do design de informação, atividade que não se resume à produção de imagens, mas tem nelas um elemento fundamental. Mapas diagramáticos dos transportes públicos, gráficos e diagramas nos relatórios anuais de grandes empresas, interfaces de sistemas de dados, infográficos em periódicos e sistemas de sinalização talvez sejam as primeiras a serem lembradas.

Apesar desses produtos característicos, a especificidade do design de informação encontra-se não no projeto de visualizações padronizadas, mas no projeto de âmbitos informativos relativamente complexos. Neles, os componentes da informação não acontecem isoladamente e sim articulando-se em um conjunto explicativo coeso. Imagens informativas avulsas (por exemplo as empregadas pela visualização técnico/científica, como uma planta-baixa, uma carta náutica ou uma tomografia) tendem a encerrar um caráter autossuficiente, e de resposta a uma situação bem definida segundo protocolos profissionais igualmente bem especificados. Mas os bons exemplos de design de informação costumam transmitir certo sentido de totalidade informacional em dado contexto, tratando-se, assim, do projeto de âmbitos informativos, em contraponto ao caráter unicamente instrumental dos exemplos acima.

Os produtos do design de informação compartilham com o design de comunicação o *modus operandi* de conferir certo sentido de ordem a elementos heterogêneos reunindo-os segundo alguma intenção comunicativa. É essa condição que coloca o projeto de âmbitos informativos como atividade a ser desenvolvida sob o ponto de vista projetual. O objetivo amplo de projetos dessa natureza é possibilitar algum tipo de compreensão, ambição que envolve, ainda, uma perspectiva ética, no que concerne à intenção comunicativa, como apontado em uma proposta para os projetos de graduação da Escola Superior de Desenho Industrial, que caracterizou o design de informação como:

[...] uma atitude de integridade na apresentação dos dados e de ênfase nos processos que levem o sujeito à compreensão desses dados [...] o intuito deverá ser, através da compreensão, dar ao sujeito a possibilidade de tomar decisões e adquirir controle sobre os processos que o afetam, podendo-se dizer então que o design de informação contribuirá para o *empowerment* do sujeito, possibilitando a sua ação efetiva. (DESIGN de informação: temas brasileiros – proposta para projetos de graduação. 2003. Escola Superior de Desenho Industrial. Universidade do Estado do Rio de Janeiro (arquivo do autor).

No contexto da propalada “explosão da informação”, a atividade do *designer de informação* ganha relativa proeminência destacando-o como um profissional apto a organizar dados complexos em formas mais significativas, capazes de suscitar interesse e facilitar a aquisição de conhecimento. Tendo em conta as definições do design como o “domínio da interface [entre

homens e coisas]” (Bonsiepe, 1999) ou ainda, como a “proposição que design é fazer sentido das coisas” (Krippendorff, 1989, p.9), vê-se que o design de informação pode ser referenciada por tais enunciados, bastando um ajuste de termos: “coisas” por “dados”. Desse modo, chega-se à noção de que o design de informação é uma atividade que se ocupa dos processos de tratamento e manipulação de dados objetivando dar sentido a eles, condição em que podem ser percebidos como informação. Sob esse ponto de vista, dashboards emergem como autênticos artefatos do design de informação. Podem ser considerados como a face visível dos enormes repositórios de dados que, ainda que constituam sistemas organizados de dados, o são apenas sob um ponto de vista computacional ou arquivístico. O processo de torná-los compreensíveis a um público mais amplo do que os projetistas dessas estruturas, possivelmente para aqueles que irão de fato ter demandas informativas, é uma atividade que requer uma mediação entre a instância maquínica e o usuário final. Esse processo abrange demandas informativas de diferentes intensidades. Entender tais contextos particulares e otimizar experiências de uso para efetivar a conversão de grande volume de dados abstratos em significado informativo é um propósito específico dos dashboards, mas circunscrito ao amplo escopo dos objetivos do design de informação.

1.4 Considerações gerais

As seções até aqui discutiram algumas descrições que procuram elucidar a natureza do artefato *dashboard*, além de apresentar um breve panorama das áreas que se intersectam quando os dashboards são analisados em maior detalhe. Importa notar que, além do dashboard ser um instrumento relativamente recente, com carência de estudos mais aprofundados, tanto as áreas de *Business Intelligence*, quanto do design de informação são incipientes e difusas e não chegam a ter alicerces bem consolidados. Mas, se por um lado isso dificulta definições rigorosas e bem estabelecidas, por outro, revela um grande campo a ser explorado, em especial pelos aspectos comunicativos que envolve. Em tal contexto, vale ressaltar que o campo de BI, parece reproduzir os problemas da área mais geral do projeto de *software*, quanto à carência de intervenções sistemáticas dos campos de design. Few observa que fornecedores da área de BI concentraram-se em desenvolver tecnologias para reunir, transformar, armazenar, acessar dados e apresentá-

los na forma de relatórios. “Tremendos progressos foram feitos nessas áreas, resultando em tecnologias robustas que podem lidar com enormes repositórios de dados”. O autor considera, entretanto, que “enquanto conseguimos armazenar grande volume de informação, fizemos pouco progresso em usar efetivamente essa informação” (FEW, 2006, p.6).

Essa deficiência talvez resulte do afastamento entre desenvolvedores e designers no projeto de *softwares* que objetivam contribuir em processos de aquisição de conhecimento. Como argumenta Bonsiepe “talvez não resulte óbvio à primeira vista que a apresentação do conhecimento requeira a intervenção de ações projetuais” (BONSIEPE, 2000-2001). Isso implica que os aspectos de uso, no caso do design de informação, aqueles que contribuem para que dados sejam entendidos como informação, podem não se efetivar da melhor maneira: “temos o paradoxo que a informática é possivelmente a disciplina que menos clareza possui sobre o conceito “informação””) (ibid.).

No caso de dashboards, o estudo das suas potencialidades informativas pode sugerir a possibilidade de destacá-los do campo de uso específico do BI, entendendo-o não apenas como um instrumento especializado, mas possivelmente como um método geral de organização de informação. Tal entendimento implica na observação de algumas características gerais como as apontadas por Few conjuntamente aos critérios relacionados nas definições anteriores. A síntese dessas formulações constitui o patamar de características definidoras de um dashboard sobre as quais a proposta deste trabalho desenvolve-se.

1.4.1 Proposta de síntese para definição de dashboards

- 1 Trata-se de painéis compostos por elementos visuais, o que significa que sua leitura acontece de modo global e simultâneo e não pela observação em diversos instantes, como acontece ao texto tabular. Tais elementos são organizados em uma única tela, sendo assim, devem “caber” no campo de visão humano, não se estendendo além dele.
- 2 Encerram um sentido de completude, sendo a consolidação de uma decisão por determinado conjunto de dados. Isso implica que essa escolha tende potencialmente a responder às perguntas formuladas que, explícita ou implicitamente, conformam a

seleção dos dados a serem visualizados. De modo que o dashboard é condicionado, além dos dados disponíveis, pelas perguntas que devem ser respondidas.

- 3 Sintetizam os atributos mais importantes para a compreensão de uma situação sugerindo determinado viés analítico e direcionando a atenção para os itens que parecem merecer maiores considerações.
- 4 Idealmente sua informação deve ser apreendida em pouco tempo. Não apenas se tal atributo está ou não dentro de um parâmetro pré-definido, mas ainda se a visualização que evidencia essa circunstância também é rapidamente perceptível.
- 5 Deve possibilitar o acesso à detalhamento adicional sobre um aspecto que se queira aprofundar a compreensão. Nesse sentido, o dashboard assume um caráter de índice, a partir do qual elementos específicos podem ser localizados e acessados.

Em suma, os itens 1–3 evidenciam o caráter de “edição” na formulação de dashboards apontando a relação entre escolha dos dados, resposta às demandas informativas e adequação a um espaço pré-determinado. O item 4 refere-se às condições de eficiência de uma visualização, ressaltando a velocidade de comunicação como aspecto importante. O item 5 foca mais diretamente nos recursos computacionais que possibilitam desdobramentos analíticos no estudo dos dados.

Esses tópicos compilam as principais características dos dashboards, consideradas pela literatura e pelos fabricantes, mas procuram contextualizá-las para propiciar uma abordagem mais geral, e potencialmente mais produtiva. Portanto, o que essas características propõem é uma compreensão dos dashboards que ultrapassa a área de BI, na qual eles se consolidaram, para adquirir autonomia como instrumento de visualização independentemente de usos específicos. Como mencionado, esses cinco tópicos compõem a estrutura de referência na qual os dashboards propostos aqui serão baseados e contra a qual serão discutidos no capítulo 4.

2 SEMIOLOGIA GRÁFICA E TRATAMENTO GRÁFICO DA INFORMAÇÃO

2.1 Jacques Bertin: *Sémiologie graphique*

O número 21 da revista da escola de ulm, iniciava sua sessão de resenhas com esta citação: “Informar-se é um dos meios para garantir a sobrevivência.”.¹⁰ Logo abaixo, lia-se uma curta nota sobre um livro lançado um ano antes, reputado como “altamente recomendável”, pois apresentava “um tratamento sistemático das representações gráficas” não contendo “apenas técnicas gráficas, mas acima de tudo uma heurística plenamente desenvolvida para traduzir informações codificadas linguisticamente em códigos visuais não verbais”. Tanto a citação quanto o livro são de autoria do cartógrafo e geógrafo francês, Jacques Bertin, fundador e diretor do Laboratório Cartográfico da *École Pratiques des Hautes Études en Sciences Sociales*.

Esse pequeno fragmento sobre o recém-publicado livro (que pode bem ser a primeira referência à teoria de Bertin em uma comunidade de designers) traz alguns entendimentos que viriam a se repetir em muitos textos sobre os aspectos visuais da informação e também nas referências ao trabalho do cartógrafo. Um desses juízos, presente na própria frase de Bertin, é a valorização da capacidade de lidar criticamente com informação como uma virtude determinante no cenário da nascente *Aldeia Global*. Nesse contexto, a *Semiologia gráfica* foi a primeira formulação teórica consistente sobre o uso de visualizações como instrumento para lidar de modo sistemático com o volume de dados em crescente expansão. Nas palavras do linguista Georges Mounin “é a primeira teoria orgânica de um sistema semiológico (ou semiótico) diverso das línguas naturais.”¹¹, sugerindo uma analogia que interpreta o trabalho de Bertin como uma espécie de gramática visual.

Diversos comentadores posteriores reiteraram aspectos que a resenha na revista de ulm descreveu como “tratamento sistemático” e sua condição como “heurística plenamente desenvolvida”. Tufte a considera a *Semiologia gráfica* um “importante trabalho” (2001, p.112).

¹⁰ Book Notes. *ulm 21, Journal of the Ulm School for Design*. Ulm: Hochschule für Gestaltung Ulm, p.32 abr. 1968.

¹¹ Fragmento de artigo publicado no jornal *Le Monde* em 16 de março de 1968 e impresso na 4ª capa da edição de 2003 da *Semiologia gráfica*.

Ware a designa como “obra mestra” (2004, p.6) e “trabalho seminal” (ibid., p.297). Card, MacKimlay e Shneiderman ressaltam a condição de “trabalho clássico”(1999, p.57). Spence refere-se a Bertin como um “pioneiro da visualização de informação.” (2007, p.52). Entretanto esses reconhecimentos foram relativamente tardios de modo que não obstaram que o trabalho de Bertin permanecesse em relativa obscuridade fora da Europa, talvez devido à dificuldade de acesso às traduções em línguas que não o francês.

Sémiologie graphique: les diagrammes, les réseaux, les cartes, de 1967 é a fundamental contribuição do cartógrafo para as disciplinas que lidam com as singulares faculdades da percepção visual para comunicar informações. O livro descreve um sistema orgânico para conversão de dados em construções gráficas, como as enumeradas no subtítulo da obra, evidenciando as informações que os dados possam conter. Dez anos mais tarde, o autor publicou um segundo livro *La graphique et le traitement graphique de l'information* contendo, além da síntese da Semiologia gráfica, uma expansão dos aspectos da *análise matricial*, no qual Bertin enfatiza e detalha as propriedades de certos gráficos, em especial da denominada *construção normal*, que possibilita a permutação de suas partes como recurso para o tratamento e comunicação de grandes volumes de dados.

Jacques Bertin converteu o laboratório de cartografia da Escola de Ciências Sociais em uma instituição de pesquisa de técnicas de visualização, mantendo constante diálogo com outras disciplinas da universidade. O cartógrafo e sua equipe projetaram visualizações para os trabalhos de estudiosos como Fernand Braudel, Emmanuel Le Roy Laudrie, Claude Levi Strauss, dentre outros. Nas várias décadas de atividade do laboratório, foram desenvolvidas diversas técnicas e extensa produção que corroboram os princípios da teoria. Nas próximas páginas serão sintetizados pontos relevantes de tal metodologia que estão diretamente relacionados ao presente estudo. Aspectos complementares encontram-se no anexo, onde são apresentados outros tópicos da teoria geral da Semiologia gráfica, referida a partir daqui pela sigla SG.

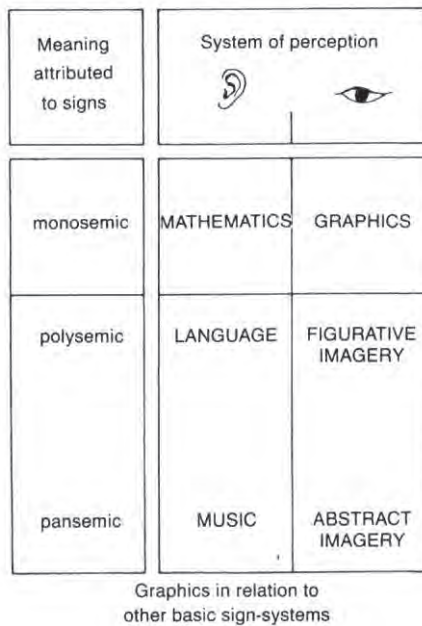
2.1.1 O aspecto monossêmico da percepção visual

As teorias de Bertin delimitam um âmbito específico de aplicação. Do ponto de vista formal, elas lidam com imagens estáticas, bidimensionais, abrangidas pelo campo de visão humano,

excluindo-se do seu escopo animações, anaglifos e imagens estereoscópicas. Pela perspectiva semiótica, ela se refere a significações de características tendencialmente universais baseadas em aspectos fisiológicos da percepção visual, de modo que se pode esperar um consenso entre os observadores no que concerne a tais ocorrências. Neste sentido, dois círculos de mesmo tamanho [● ●], mas em diferentes posições de uma folha de papel são percebidos como unidades distintas, mas se acrescentamos um losango [● ● ◆], os dois círculos serão visualmente agrupados por semelhança e diferenciados do quadrilátero. Dados três quadrados [■ ■ ■] de diferentes tamanhos e a tarefa de ordená-los, pode-se fazê-lo de modo crescente [■ ■ ■] ou, decrescente [■ ■ ■], mas não [■ ■ ■]. Em meio a formas diversas [◆ ▲ ■ ♣ ◆ ● ◆ ■ ●▲ ♥ ♠ ◆ ■], é mais fácil identificar e contabilizar prontamente aquelas que compartilham a cor vermelha do que todos os losângos. Esses são fenômenos visuais com as quais podemos esperar a concordância daqueles que compartilham uma percepção visual nos parâmetros considerados normais. Eles constituem, então, um nível básico de significação, constituído por relações de maior/menor, sequenciamento e diferenciação que precedem o uso de qualquer codificação. Essa condição fundamenta o nível *monossêmico* de significação da imagem. “Quando o conhecimento da significação de cada signo **precede** à observação do conjunto de signos.” (BERTIN, 2005, p.6, grifo do autor). Nesse âmbito, a imagem descrita na SG seria o equivalente visual da matemática e da lógica, atividades nas quais a significação de um símbolo antecede seu uso, assegurando o consenso sobre o seu significado.

A monossêmia é um dos três níveis de significação definidos na SG (Figura 6). Além dela, há os níveis *polissêmico* e *panssêmico*. A polissemia ocorre quando o significado atribuído a um signo sucede o contato com o interpretante, tendendo-se à multiplicidade de significações, como acontece na palavra escrita e na imagem figurativa. No nível panssêmico, o signo possibilita graus ainda mais abertos de significação podendo ser interpretado de diversos modos (como em uma pintura abstrata ou na música atonal). Em Bertin, a imagem simbólica ou *figuração*, que se movimenta entre esses dois níveis, é a quarta modalidade de ocorrência de uma imagem no plano, além dos *diagramas*, *redes* e *cartas*. Interpretá-la depende do conhecimento da codificação que lhe deu origem, tarefa que fundamenta disciplinas próprias, voltadas para relações que se estabelecem entre a imagem e o interpretante e não entre os elementos “concretos” de uma representação como aqueles que interessam a Bertin. A SG parte das propriedades monossêmicas da imagem para fundamentar visualizações que não sejam, em si, alvo de dúvidas semânticas em razão das convenções da representação. Qualquer

Figura 6 – Posicionamento da Semiologia gráfica (*Graphics*) nos sistemas de significação fundamentais



Fonte: Bertin, 2010, p.2

difficuldade quanto à visualização é inerente à forma como os dados são transpostos em imagem e, portanto, descolada do conteúdo informativo que os dados podem revelar. Assim o que Bertin propõe mantendo-se no nível monossêmico é estabelecer um parâmetro de concordância e eficiência básico que minimize interpretações equivocadas da visualização. Condição que permite direcionar a atenção ao significado externo da representação, isto é, para as relações entre os fatos que deram origem à visualização, âmbito onde ela reintegra-se aos significados convencionados, permitindo análise.

2.1.2 Análise da informação

A SG opera a partir de uma separação estrita entre os dados e sua visualização, entendidos como instâncias formalmente independentes, condição que faz com que um mesmo conjunto de dados originais possa ser representado por diferentes formas. Daí a necessidade da análise prévia dos dados para identificar sua condição e estabelecer correspondências entre eles e as respectivas

representações visuais, justificando a escolha por determinada forma de representação. Esse é o objetivo do primeiro elemento de estudo da SG: a análise da informação a ser visualizada.

2.1.2.1 Componentes e invariantes

No âmbito da SG, *informação* é “uma série de correspondências observadas entre um conjunto finito de conceitos de variação ou *componentes*.” (ibid. p.16, grifo nosso). Assim, a identificação dos *componentes* de determinado conjunto de dados é a etapa basilar da análise. Para exemplificar esse elemento na sua forma mais simples, Bertin apresenta esta situação: “Em 8 de julho de 1964, a ação X da bolsa de Paris foi cotada a 128 francos; em 9 de julho, foi cotada a 135 francos.” (BERTIN, 2005, p.8) Esta frase pode ser entendida sob a perspectiva da variação de cada um dos dois conceitos presente na declaração. São eles: o *tempo*, categorizado em dias e a *moeda*, referida em quantidade de francos. Trata-se dos componentes que podem ser transcritos graficamente, “os conceitos de variação utilizados”. (ibid., p.9). Identificar claramente a quantidade de componentes distintos, bem como o modo particular como eles variam constitui o ponto fundamental da análise da informação, pois estes aspectos estão biunivocamente relacionados à sua representação visual. No exemplo comentado, são utilizadas as duas dimensões do plano em um diagrama que coloca no eixo *x*, o tempo (os dias do mês) e no eixo *y*, a quantidade de francos (Figura 7).

O segundo elemento de análise são os invariantes ou a “definição completa e invariável comum a todos os dados.” (ibid., p.16). Trata-se das características que permanecem imutáveis em todo o conjunto de dados, possibilitando a comparação entre tais dados, sob os mesmos critérios. No exemplo anterior, as características invariantes são os valores cotados em francos da ação X; no mês de julho; na bolsa de Paris. Bertin ressalta que a análise dos conceitos de variação e dos elementos invariantes permite “compreender uma informação complexa, definir a melhor transcrição gráfica, redigir títulos e legendas”. (ibid., p.16). Em especial sobre a redação das identificações verbais de uma visualização, o autor adverte que há casos em que um mesmo conjunto de dados pode ser descritos de diferentes modos, situação possível quando é formulada uma hipótese de associação inédita, caso em que: “Somente a pesquisa da definição precisa dos invariantes e dos componentes permitem, em tal caso, compreender a informação.”

Figura 7 – Representação da relação entre os componentes tempo e quantidade de francos



Fonte: Bertin, 2010, p.16

(ibid., p.17). Este fato sublinha que a análise criteriosa dos elementos que antecedem à uma visualização “torna-se obrigatória se o redator¹² quer compreender o que ele deve comunicar e se o leitor quer compreender aquilo que é apresentado, caso contrário todos estão expostos à graves erros.” (ibid., p.17).

2.1.2.2 Número de componentes

Bertin ressalta a limitação inerente à percepção visual para discernir clara e inequivocamente variações visuais de forma que apenas um número reduzido delas pode ser visualizado simultaneamente de modo integro. Daí a relevância de analisar a quantidade de componentes de um conjunto de dados, pois esse número determina a quantidade de diferenciações visuais correspondentes na visualização de tal conjunto. “O número de diferenciações visuais necessárias à representação é pelo menos igual ao número de componentes da informação” (ibid., p.28).

¹² *rédacteur*, no original: no trabalho de Bertin, este termo costuma designar o agente que projeta uma visualização, ocorrendo ainda o termo *rédacteur graphique*.

2.1.2.3 Comprimento dos componentes

Cada componente pode apresentar certa divisibilidade interna de modo que um mesmo conceito de variação pode-se subdividir em categorias¹³ (propriedade que fundamenta o conceito de *classe* nos estudos estatísticos). Nesse sentido as regiões *Norte*, *Sul* etc. são subdivisões de um componente geográfico, as categorias *bovinos*, *ovinos*, *caprinos*, do componente *diferentes animais domésticos* ou ainda as divisões anuais são classes do componente *tempo*.

O *comprimento* do componente é um parâmetro que quantifica “o número de elementos ou categorias que ele [o componente] permite identificar” (BERTIN, 2005, p.9). Segundo o autor “a complexidade de uma figura está ligada ao número de categorias em cada componente” (ibid., p.9). Esta noção distingue, grosso modo, componentes *curtos* dos *longos*, conforme o comprimento de seus elementos. Deste modo, o componente *sexo* tem comprimento 2, o componente *Regiões do Brasil*, comprimento 5. Já estados brasileiros apresenta comprimento 26 e municípios brasileiros, comprimento 5.565. A identificação de componentes mais longos ou curtos definem visualizações apropriadas a cada caso, além da escolha por *variáveis visuais* que possibilitem, no mínimo, a mesma quantidade de distinções visuais quanto o comprimento do(s) componente(s) representado(s). Bertin ressalta ainda que “em um problema com mais de três componentes, o número mínimo de imagens necessárias é função do comprimento dos componentes.” (ibid, p.33). Cabe notar também que o número de patamares sensíveis das variáveis visuais costuma ser menor do que o comprimento dos componentes mais longos. Por exemplo, o número de níveis de cinza discerníveis da variável *valor* permite em torno de dez diferenciações, ao passo que as variações dos componentes são, em princípio, ilimitadas. Assim, os procedimentos de agregação são um recurso para a redução dos comprimentos longos a poucas classes, adequando os dados aos limites fisiológicos da percepção visual. Desse modo, o componente *idade humana*, cujo comprimento potencialmente chega à ordem da centena, poderia ser reduzido às classes significativas *criança*, *jovem*, *adulto*, *idoso* ou, ainda, os diversos segmentos produtivos de um país podem ser reduzidos aos setores econômicos *primário*, *secundário* e *terciário*. Nesse âmbito reside o delicado equacionamento entre o nível

¹³ Bertin trata como sinônimos os termos *catégories* (categorias) *éléments* (elementos), *classes* (classes) e *paliers* (níveis). Ver Bertin 2005, p.9.

de simplificação dos dados a classes significativas e a demanda por dados mais completos, além dos aspectos de como a formulação de categorias que são consolidadas a partir de tais classes podem influir na comunicação, na compreensão e nas ações em relação a determinado tema.

2.1.2.4 Nível de organização dos componentes

Para Bertin, o que seu método representa são as relações entre os componentes de um conjunto de dados. A SG sistematiza tais relações em três *níveis de organização* (*qualitativo*, *ordenado* e *quantitativo*) sob os quais todos os componentes podem ser classificados, assim como as variáveis visuais que venham a representá-los. Identificando-se o nível de organização do componente e da representação visual podem-se estabelecer as melhores correspondências e as visualizações mais eficientes.

O *nível qualitativo* (ou combinatório), “agrupa todos os conceitos de simples diferenciação (profissões, produtos, religiões, cores...)” (BERTIN, 2005, p.10), designando assim as noções de semelhança e diferença. Um componente pode ser classificado deste modo se suas “categorias não se ordenam de um maneira universal. Consequentemente, são ordenáveis de diferentes maneiras.” (ibid., p.36). Nesse nível, duas atitudes perceptivas são possíveis: associação, quando “isto é semelhante aquilo e eu posso reuni-los em um só grupo” e seleção, se “isto é diferente daquilo e está separado em outro grupo.” (ibid., p.10).

O *nível ordenado* “agrupa todos os conceitos propensos a ordenar elementos de modo universalmente aceito.” (ibid., p.10), em contraponto aos conceitos de ordem e desordem. Isto pode acontecer em séries cronológicas como acontece em *janeiro-fevereiro-março*; em percepções de luminosidade como *branco-cinza-preto*; em juízos de valor como *ruim-mediano-bom*. Os componentes ordenados permitem declarar que “isso é mais do que isso, mas menos que aquilo.” (ibid., p.10).

O *nível quantitativo* (ou métrico), ocorre “quando se dispõe de uma unidade contável” que permite declarar que “isto é o quarto, o triplo, quatro vezes aquilo.” (ibid., p.10). Assim, “uma série numérica é quantitativa quando tem por objetivo precisar a variação de distância entre as categorias”. (ibid., p.38). Bertin identifica duas situações para dados quantitativos. 1) Quando eles são dependentes das classes, como quando determinada população é contabilizada em

períodos desiguais de tempo, ou classes desiguais de idade. 2) Se as quantidades são autônomas em relação às classes, como quando certa população é analisada por sua distribuição espacial absoluta, independentemente das áreas específicas de cada região.

Para Bertin, “os níveis de organização formam o campo das significações universais, das analogias fundamentais que podem pretender à transcrição gráfica” (ibid., p.10). O autor sublinha o papel dessas relações no raciocínio lógico apontando que seu método limita-se à representação dos níveis de organização, “mas são as relações de similitude e de ordem, apoiadas em métricas, que constituem a base de toda reflexão.” (ibid., p.10). Ele considera também que as significações além desses três níveis são externas à representação gráfica, sendo, portanto, convenções que dependem de decodificação para que sejam interpretadas, como acontece às imagens figurativas. Da perspectiva operacional, quando se identifica a condição *qualitativa*, *quantitativa* ou *ordenada* dos componentes do problema, buscam-se no repertório de estímulos visuais aqueles que compartilham a mesma condição. “Cada variável visual tem suas propriedades particulares de nível [de organização] e comprimento. É importante que cada componente seja transcrito por uma variável com pelo menos o nível e o comprimento correspondentes.” (ibid. p.10).

Assim, para representar um fenômeno quantitativo como o número de habitantes, emprega-se uma variável visual que expresse um sentido também quantitativo, condição só presente na variação de *tamanho*. Logo, para transcrever essa variação populacional em um gráfico emprega-se um aumento na altura das barras, e não uma mudança na *forma* ou na *orientação* das barras. Para representar um fenômeno ordenado como a variação da altimetria de um terreno é adequado usar uma variável universalmente ordenável como a variação de valor do claro ao escuro ao invés de, por exemplo, a variação de cor, na qual não se pode estabelecer uma ordem inequívoca. Estes aspectos são detalhados na segunda parte da Semiologia gráfica que se dedica a estudar os recursos gráficos disponíveis para representação de dados. Para a descrição detalhada desses conceitos, e a correlação entre o *nível de organização dos componentes* e os estímulos visuais, ver a seção “Os meios do sistema gráfico” no anexo desta dissertação.

2.1.3 Regras do sistema gráfico

2.1.3.1 O problema gráfico

Tendo estabelecido a relação entre os componentes e suas representações visuais, resta definir a morfologia característica da visualização, isto é, o tipo de *imposição no plano* no qual ela se enquadra. Diante das múltiplas combinações entre variáveis visuais e construções gráficas surge um vasto universo das construções gráficas. Para exemplificar a extensão dessas possibilidades, Bertin (2005, p.100-137) demonstra como um mesmo conjunto de dados pode ser representado por 100 visualizações diferentes (como histogramas, gráficos de dispersão, mapas, cartogramas, anamorfoses etc.). Diante dessa variabilidade, o problema fundamental é justificar a escolha por determinada forma de visualização. “A utilidade do desenho não pode ser aceita, seu rendimento informativo não pode ser julgado sem que se responda com rigor à [...] questão: Que desenho é preciso fazer?” (ibid., p.100). Respondê-la depende da definição precisa de critérios de avaliação e as condições de uso, aspectos delineados na seção da *teoria da imagem*.

2.1.3.2 A teoria da imagem

Os fatores que facilitam a interpretação de uma construção gráfica compõem a *teoria da imagem* da SG. Tem como elemento fundamental o conceito de *eficácia* (*efficacité*) como critério que possibilita qualificar uma construção gráfica de modo preciso e mensurável. Desse modo, permite uma classificação baseada na melhor adequação a determinado fim, justificando a preferência por uma construção em particular. Com tal objetivo, a *eficácia* é definida na proposição:

Se para obter uma resposta correta e completa a uma questão dada, com todos os demais elementos sendo iguais, uma construção requer um tempo de observação mais curto que outra construção, se dirá que ela é mais eficaz para aquela questão.
(BERTIN, 2005, p. 139)

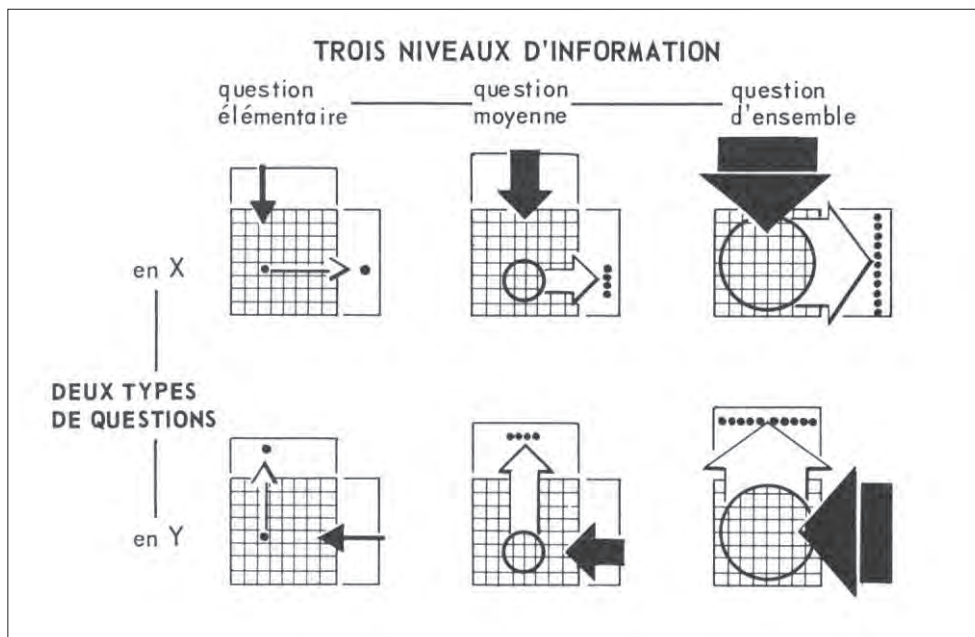
Bertin baseia esta definição como a aplicação à fisiologia visual do conceito de “custo mental” proposto por GK Zipf ¹⁴. A eficácia torna-se relevante quando o observador coloca uma questão precisa ao desenho, caso em que o tempo de resposta entre uma visualização eficaz e outra “pode exceder uma hora.” (ibid., p.139). Para estabelecer as regras para as construções eficazes, a teoria da imagem se desdobra em cinco tópicos de análise. Eles circunscrevem as etapas do processo de leitura de uma visualização, generalizam as questões que cabem ser colocadas, apresentam o conceito de imagem na SG, assim como sua construção e seus limites.

As etapas do processo de construção: Referem-se às operações necessárias à correta interpretação de uma visualização. Por meio da *identificação externa* em títulos e legendas, relacionam-se os componentes envolvidos no problema. Já a *identificação interna*, refere-se à clareza de comunicação sobre a relação entre as variáveis visuais e os componentes respectivos. Tais identificações estabelecem “a ligação entre o sistema gráfico e os outros sistemas. Ela é fornecida pela palavra, a escrita, títulos e legendas ou por analogias figurativas de forma e de cores” (ibid., p.12). Esses recursos precedem e consubstanciam a etapa da percepção das correspondências originais, e à compreensão do modo como os componentes se correlacionam, efetivando a elaboração de perguntas como “em tal período, qual o valor?”. Segundo Bertin “as questões podem ser muito diferentes. Mas para cada informação elas são em número finito e podem ser rigorosamente definidas” (ibid., p.140). Este critério aponta para o segundo tópico da teoria da imagem, a seguir.

As questões possíveis e os níveis de informação: Trata-se das questões que podem ser colocadas a uma visualização estabelecendo três níveis de leitura conforme a abrangência das suas colocações (Figura 8). O *nível elementar* classifica as perguntas introduzidas a um único elemento de um componente (“em tal país”) e o elemento correspondente de outro componente (“qual o PIB?”). Bertin salienta que tais questões tendem a ser externas à representação gráfica, de modo que este nível é usado como um reservatório de dados de onde se extrai uma informação elementar para transcrever em uma outra linguagem ou em outra imagem”. (ibid., p.141). O nível *médio* ocorre na relação entre um grupo de elementos ou categorias e um grupo de correspondências. Permite questionamentos sobre subconjuntos dos componentes como “no

¹⁴ ZIPF, Georges Kingsley. *The Psycho-biology of Language*. Harvard University: 1935. Referência citada em Bertin, 2005, p.139.

Figura 8 – Os três níveis de informação



Fonte: Bertin, 1977, p.13

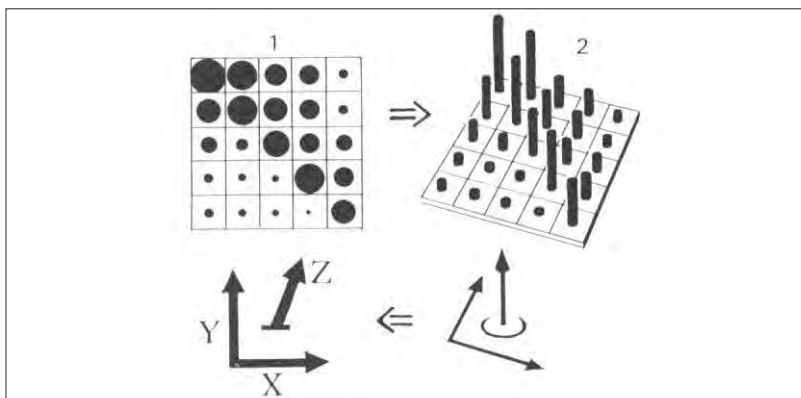
primeiro trimestre, qual a tendência?” São questões que “tendem a ‘reduzir’ o comprimento da componente para descobrir, em função da informação, os grupos de elementos ou de categorias homogêneas [...]” (p.141). Tal redução fragmenta os dados iniciais nos diversos grupos que sintetizam as respostas para inúmeras perguntas que dizem respeito a partes específicas do conjunto total. O *nível de conjunto* permite um juízo sobre a “feição” geral da organização dos dados, habilitando a identificação de padrões e tendências abrangentes. Bertin considera que este é o nível superior de leitura, apto a responder questões do tipo: “durante o período, qual foi a evolução?”, cuja possível resposta seria: “tendência de alta”. Este nível tende a reduzir a informação a uma relação de ordem entre os componentes hierarquizando-os segundo certos critérios de agrupamento. Esta característica facilita a retenção da mensagem e favorece a comparação com outras. Segundo a SG, quando todos os componentes de um conjunto de dados e seus respectivos comprimentos são definidos, pode-se listar a totalidade das questões passíveis de resposta, antes de qualquer tipo de visualização. Deste modo, “É possível definir em termos de componentes, variáveis e comprimento todos os objetivos que uma informação pode abranger.” (ibid., p.141).

Definição da imagem: A SG entende o conceito de *imagem* como uma unidade de

percepção visual. Trata-se da forma (“padrão”, “silhueta”) que permite abranger com um golpe de vista todas as correspondências que dão resposta a um questão. Pelas palavras de Bertin (ibid., p.13), imagem é a “forma visual significativa perceptível em um instante mínimo de visão.”. Este conceito introduz um critério de avaliação das construções visuais basilar na SG. Ele permite qualificar visualizações em razão da propriedade de sintetizar correspondências em uma imagem. Assim, as mais eficazes seriam aquelas em que “toda questão, independentemente do tipo ou do nível obtém uma resposta em um único instante de percepção, uma resposta perceptível em uma única imagem.” (ibid., p.13). Desse modo, pode-se estabelecer uma tipologia das construções (ver Figura 18, p.81) que não atendem a este requisito (nomeados por Bertin de “construções inúteis”) e aquelas que o satisfazem, dando “resposta espontânea a todas as questões que ela possa suscitar, independentemente do tipo ou do nível” (ibid., p.160).

Construção da imagem: Bertin considera que uma imagem comporta no máximo três variáveis homogêneas e ordenadas (Figura 9). Elaborar visualizações que definem uma imagem, depende do emprego das duas dimensões do plano de modo homogêneo, retilíneo e ortogonal e uma variável de 3ª dimensão ordenada (tamanho, valor ou grão). Portanto, “toda informação com três componentes ou menos pode ser construída em uma imagem.” (ibid., p.14). Respeitadas essas condições, independentemente do tipo ou do nível da questão, a resposta se apresentará visível em um único relance. Caso contrário, é provável que a resposta a algumas questões dependa da observação de diversas imagens, dificultando a compreensão de conjunto e a retenção da informação. Bertin nomeia tais construções como *figurações*, avaliando que serão menos eficazes que as *imagens*.

Figura 9 – As três dimensões da imagem



Os limites: Além de três variáveis, torna-se improvável a constituição de uma imagem de modo que conjuntos de dados com mais de três componentes demandam vários instantes de percepção. Dada tal restrição, torna-se necessário “escolher questões preferenciais introduzidas por um só instante de percepção e reservar às [...] menos úteis ou [...] prováveis as identificações de entrada que necessitam vários instantes de percepção.” (ibid, p.14). A retenção destas várias imagens é dificultada em proporção à quantidade de imagens necessárias para a compreensão de um problema. Neste sentido, coloca-se o tópico de como priorizar as questões, problema que circunscreve as funções da representação gráfica.

2.1.3.3 As três funções da representação gráfica

A limitação ao máximo de três componentes é apontada na SG como restrição fundamental diante da qual restaria somente a criteriosa escolha das formas de comunicá-las, pois “as informações com mais de três componentes constituem o essencial dos problemas de comunicação.” (Bertin, 2005, p.160). Para tanto, é necessário estudar a demanda por informação em um determinado âmbito de uso. Tais necessidades são generalizadas na SG sob as categorias funcionais de *registro, comunicação e tratamento* da informação.

Registrar a informação: Trata-se dos desenhos de inventário cuja finalidade é o registro exaustivo dos dados, constituindo uma memória artificial que evita a necessidade de memorização. Neste caso, a rapidez de leitura é menos importante do que completude dos dados. Logo, a constituição de *imagens* dos dados não é requisito determinante, de modo que, tendencialmente, os inventários respondem a questões do nível elementar de leitura.

Comunicar a informação: Tem como objetivo a elaboração de uma imagem memorizável, (ou *mensagem*, nos termos da SG) passível de ser lembrada por ocasião de um debate ou de uma tomada de decisão. Para isto, a informação exaustiva é simplificada por meio da priorização de certos elementos que tendem ao nível de conjunto. Devido às muitas possibilidades de simplificação de dados completos, o problema que se coloca é o de como justificar uma imagem simplificada em particular. Bertin desenvolve esse tópico discutindo a *gráfica de comunicação*, detalhada mais à frente.

Tratar a informação: Processo que fundamenta a simplificação dos dados originais

viabilizando sua comunicação. Isso acontece por meio de procedimentos de ordenação e categorização que possibilitam descobrir grupos significativos, resultado que tende a reduzir os dados extensos a poucos padrões relevantes. Assim, o tratamento da informação parte dos dados completos, objetivando simplificar a representação ao menor número possível de imagens memorizáveis que satisfaçam as questões ao nível de conjunto. Neste caso, a construção será memorizável para as comparações e exaustiva para as escolhas dos diversos agrupamentos possíveis, que variam conforme as perguntas feitas. Bertin sublinha a característica “lúdico-instrumental” desta função da representação gráfica, que permite ao pesquisador “jogar” com os dados de modo que ele os “classifica, ordena de diferentes maneiras, agrupa as imagens semelhantes, reconstrói as imagens ordenadas para descobrir o esquema sintético mais simples e mais significativo.” (ibid., p.164).

Segundo Bertin “uma informação com três componentes ou menos construirá uma imagem que responde às três funções da representação gráfica.”. Por outro lado, se exceder este limite “será construída diferentemente dependendo da função pretendida, isto é, segundo a natureza das questões úteis.” (ibid., p.14).

2.1.3.4 As regras de construção

Como discutido, as construções gráficas de alto nível informativo são aquelas que dão resposta ao *nível de conjunto* da informação, pois nesse caso, perguntas referentes aos níveis inferiores, *médio* e *elementar*, também podem ser respondidas. Mas a recíproca não é verdadeira, construções elementares não fornecem a informação de conjunto. Assim, para a elaboração de gráficos baseados nesse grau de exigência, Bertin delinea três parâmetros básicos.

Reduzir ao mínimo o número de imagens. A melhor situação acontece com um única imagem, mas quando isso não for possível deve-se buscar manter o menor número possível de instantes de percepção.

Simplificar sem perda informativa. Assegurar a integridade do número de correspondências no processo de simplificação é o parâmetro geral que concerne aos conjuntos de dados com um ou mais componentes ordenáveis.

Simplificar com redução informativa. Este caso se aplica aos conjuntos de dados com

vários componentes ordenados conseguindo, a despeito da inevitável perda informativa, clareza e eficácia de comunicação.

Tais regras são sintetizadas por Bertin em uma série de *esquemas de base* organizados segundo o tipo de gráfico (diagrama, rede, ou cartografia) e o número de componentes. A interrelação entre os componentes é visualizada em esquemas gerais que indicam, para os principais casos, as representações gráficas mais eficientes, conforme consta na seção sobre *imposição*, em anexo.

2.2 *La graphique et le traitement graphique de l'information*

Em 1977, Jacques Bertin publica seu segundo livro *La graphique et le traitement graphique de l'information*¹⁵. Essa obra apresenta uma síntese da Semiologia gráfica (denominada então como *Graphique*) e uma expansão das técnicas de tratamento gráfico da informação. Para os propósitos da presente dissertação, interessam quatro aspectos em especial.

1 O novo segmento sobre a *Análise matricial* de um problema e a elaboração da tabela de dados, instrumento que baseia toda visualização. São descritas as etapas para organização do conjunto de dados para a elaboração de uma única tabela que consolida os dados e permite dar resposta às hipóteses iniciais. Este é um aspecto fundamental no que concerne ao âmbito dos dashboards, pois eles visualizam sistemas que armazenam dados em inúmeras tabelas, de modo que o problema da definição pela tabela mais adequada coloca-se enfaticamente.

2 Análise das construções que tendem ao nível superior de informação em contraposição aos “diagramas inúteis” como gráficos de setores (“pizzas” e “borboletas”), histogramas justapostos e empilhados (*stacked bar graph*), visualizações frequentemente empregadas em dashboards, mas que se mantêm no nível elementar de informação. Bertin argumenta em favor da *construção normal* e suas variantes, que possibilitam manipulações e reordamentos, simplificando o conjuntos complexos de dados em poucos grupos significativos.

¹⁵ Coincidentemente, esse é o ano de publicação do livro *Exploratory Data Analysis* do estatístico norte-americano John Wilder Tukey, onde ele apresenta diversas técnicas fundamentais na análise estatística como o gráfico de caixa (*box-and-whisker plot*), o diagram de caule e folha (*stem-and-leaf plots*) e as comparações pareadas (*Tukey's paired comparisons*).

3 Os aspectos de visibilidade de um *diagrama*, como aqueles descritos na *lei da visibilidade*, que identificam procedimentos para otimizar a clareza nas representações diagramáticas. Tais procedimentos relacionam-se às diferentes possibilidades de escalas gráficas, que delineiam o conceito de escalas absolutas e escalas relativas.

4 Expansão dos argumentos sobre a *gráfica de comunicação*, em especial sobre os elementos gráficos que sinalizam e identificam os resultados obtidos no tratamento dos dados, destacando e justificando as informações, hipóteses, advertências etc.. Trata-se, então, de recursos que coincidem com o uso de alertas, como ocorre nos dashboards.

2.2.1 Análise matricial

Bertin aponta que, em um estudo gráfico, para cada etapa de decisão corresponde uma modalidade de intervenção gráfica, conforme resumido na Tabela 1 e exemplificadas na Figura 10. Nesse processo, o autor sustenta o caráter imprescindível do pensamento humano também na definição do problema, orientando a coleta dos dados necessários a um estudo, e a formulação de hipóteses, fases em que os processos automatizados não se mostram adequados. Bertin ressalta que, sobretudo em grandes pesquisas com muitas fontes de dados, construir uma tabela de dados depende das “hipóteses e dos meios disponíveis para reduzir os dados. Estes meios são os tratamentos matemáticos e gráficos da informação.” (BERTIN, 1977, p.17). Nesse

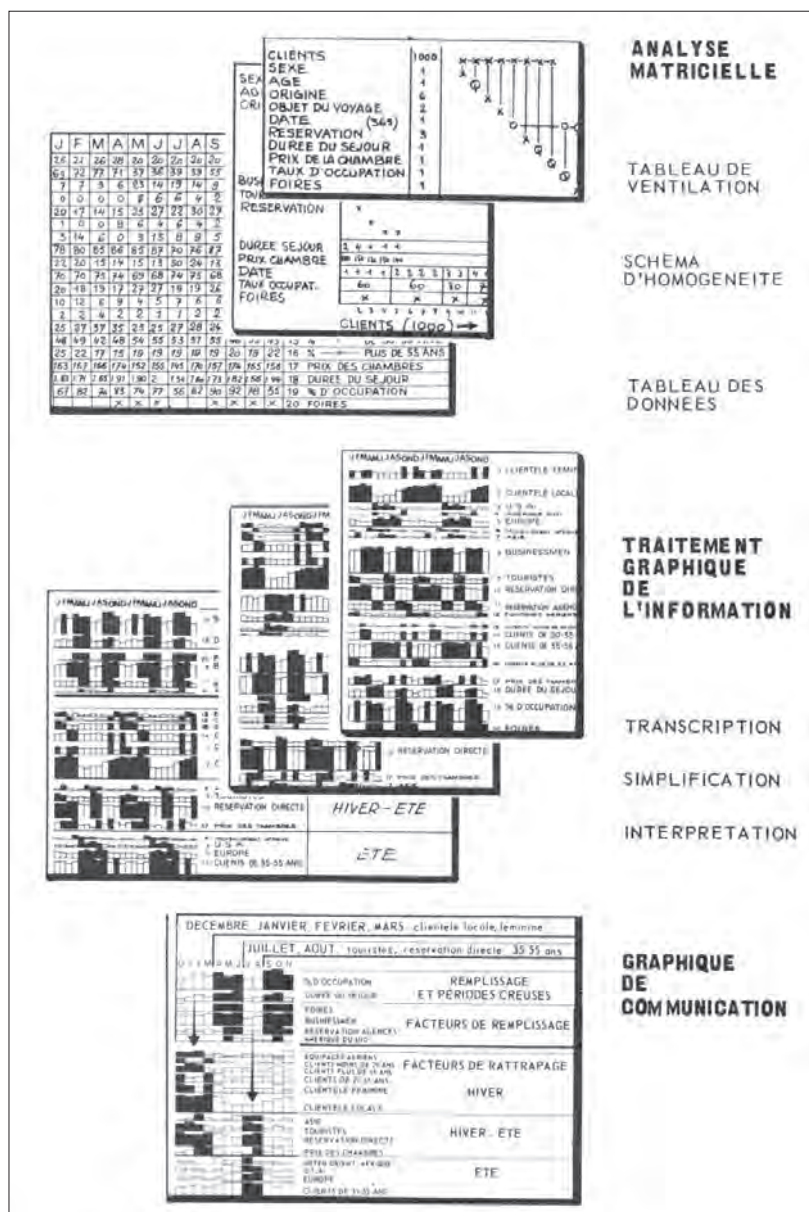
Tabela 1 – As três etapas sucessivas da intervenção gráfica

Etapas da decisão	Intervenção gráfica
definir o problema construir a tabela de dados	Análise matricial de um problema (definir as questões)
adotar uma linguagem de tratamento: tratar os dados (categorizar os dados exaustivos)	Tratamento gráfico da informação (descobrir as respostas)
interpretar para: decidir ou comunicar os dados simplificados	Gráfica de comunicação (comunicar as respostas, se apropriado)

Fonte: autor (adaptado de BERTIN, 1977, p.16).

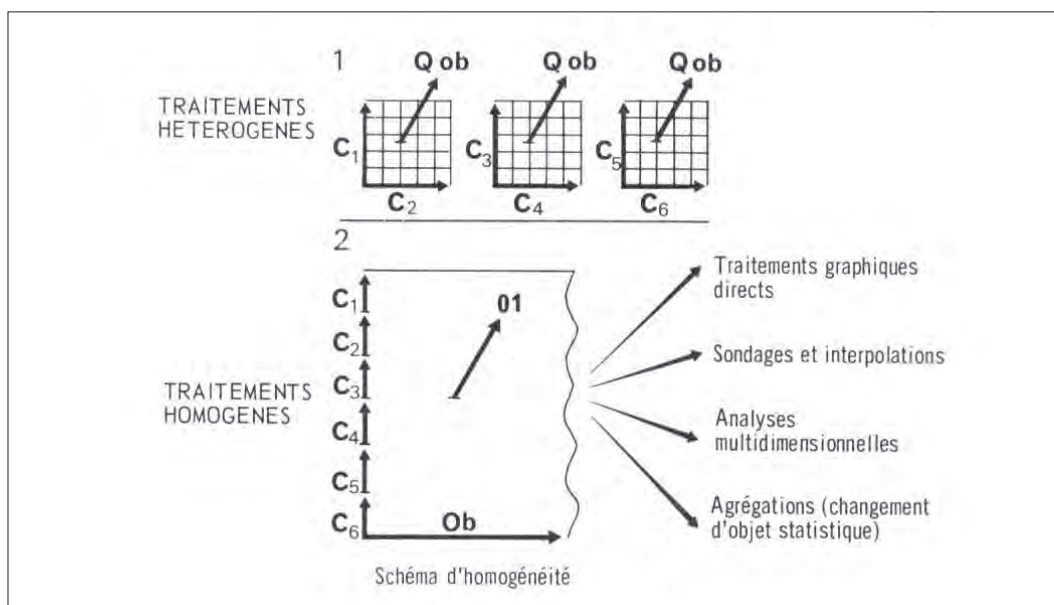
sentido, para elaborar hipóteses criteriosas é preciso estabelecer uma visão global dos dados disponíveis que possibilite identificar suas propriedades, esclarecer suas interdependências, notar elementos indisponíveis e a demanda por novos tipos de dados. Neste sentido a análise matricial “é um processo que permite ‘ver’, isto é, construir graficamente este conjunto e ver as escolhas possíveis e sua incidência.” (ibid). A Figura 10 localiza o processo da análise matricial como a primeira etapa de intervenção gráfica.

Figura 10 – As três formas de intervenção gráfica



Assegurar que todos os elementos de um estudo tenham um ponto de comparação é o requisito para relacionar de modo consequente os dados de um conjunto. Trata-se da questão da *homogeneidade* de um problema. Estudos homogêneos baseiam-se na identificação de um componente que é comum a todos os outros, caso em que ele pode ser situado no eixo x de um diagrama “tornando-se o objeto estatístico para o qual todos os outros componentes são caracteres, sendo suscetíveis de se posicionarem em y.”, condição ilustrada na Figura 11. Isso resulta em uma *tabela única*, que, segundo Bertin, é a condição da homogeneidade de um estudo, caracterizando-o como uma pesquisa embasada por uma demonstração lógica. Ao passo que diversas tabelas permitem apenas comparações dois a dois, recaindo em uma análise heterogênea, na qual não há um ponto de comparação. “É preciso construir o esquema (2) [Figura 11] no qual a ‘população de objetos’ (Ob) fornece o ponto de comparação. Trata-se do esquema de homogeneidade [...] que delinea a matriz dos dados.”. (Bertin 1977, p. 18). O *esquema de homogeneidade* é o segundo documento da análise matricial, antecedido pela *tabela de ventilação* e sucedido pela *tabela de pertinência*. A partir de tal esquema é possível optar pela melhor forma de análise como os tratamentos gráficos diretos, ou os tratamentos matemáticos, como as sondagens e interpolações, as análises multidimensionais e as agregações. Diante dessas possibilidades, a peculiaridade do tratamento gráfico é ser, ao mesmo tempo, instrumento de análise e evidência visual do processo.

Figura 11 – Esquematisação dos tratamentos gráficos heterogêneos e homogêneos



2.2.1.1 Processo de análise matricial

Em relação à problemática da decisão sobre a tabela que irá basear uma visualização, Bertin propõe o processo de análise matricial considerando que “Um estudo é um todo que se constitui apenas de uma única matriz. Neste sentido, a análise matricial “permite, quais sejam as dimensões do problema, concebê-lo como uma única tabela, a partir da qual todas as escolhas necessárias podem ser logicamente discutidas e encadeadas.” (Bertin, 1977, p.233). O objetivo desta análise é:

- definir o ponto de comparação assegurando a homogeneidade do estudo
 - ‘dimensionar’ os problemas com relação aos meios [disponíveis].
 - notar as questões de maneira homogênea aos dados e assegurar que o tratamento responderá a tais questões
 - superar a ambiguidade do discurso estatístico e apresentar a informação de modo preciso e conciso aos envolvidos [...]
 - expor a informação tendo em conta desde o início sua publicação e, portanto, justificar as escolhas e tratamentos que foram selecionados no conjunto do estudo.
- (BERTIN, 1977, p.234)

Este processo é composto pela elaboração de três documentos sucessivos. 1: a *tabela de alocação* (*tableau de ventilation*), que inventaria os componentes e assinala suas correlações. 2: o *esquema de homogeneidade* (*schemá d’homogénéité*), que permite planejar, para determinado conjunto de componentes, as tabelas possíveis, identificando seus pontos de comparação. 3: a *tabela de pertinência* (*tableau de pertinence*) anota na tabela estabelecida as hipóteses, os dados faltantes e os cálculos preparatórios para a tabela definitiva.

1 *Tabela de alocação.* Caracteriza-se por três funções: “anotar todos os componentes do discurso, anotar o comprimento de cada componente e seus cruzamentos com outros componentes; ser independente da ordem do discurso” (BERTIN, 1977, p. 235). Nessa etapa, Bertin ressalta que identificação precisa dos componentes não é uma tarefa simples, sendo potencialmente uma fonte de ambiguidades. Uma declaração aparentemente simples como “Eu conheço a população por profissão e por idade.”, pode-se referir a diferentes modos de organização dos dados. Talvez se conheça a profissão e a idade de cada indivíduo. Mas também é possível que só se conheça a profissão para cada classe de idade, ou ainda a idade média por grupos profissionais específicos. Devido a tais nuances, torna-se importante identificar estritamente a natureza dos componentes

do estudo em questão, sobretudo em situações complexas que envolvem grande número deles. Para tanto, Bertin propõe as *notações de cruzamentos*, utilizadas nas tabelas de alocação.

Trata-se de um quadro no qual a coluna da esquerda enumera os componentes. A próxima assinala seus comprimentos e as adjacentes demarcam suas relações empregando os símbolos relacionados na Tabela 2. A seguir, são apresentados alguns exemplos de diferentes casos de uso da tabela de alocação (Figura 2.7).

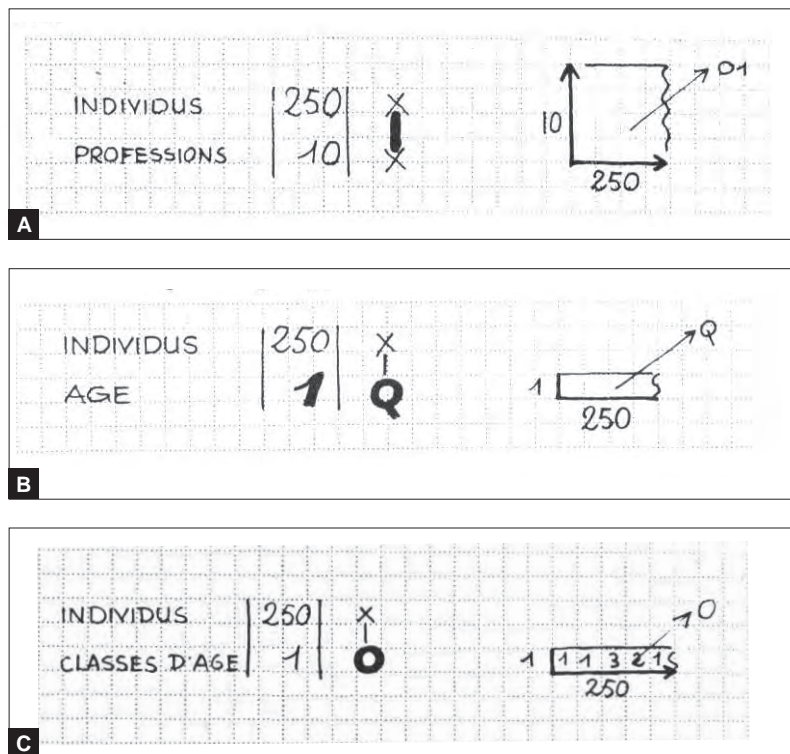
(Figura 12A) População de 250 indivíduos, cada qual desempenhando uma de dez profissões possíveis. Para o componente seletivo *indivíduos*, anota-se o comprimento 250 e demarca-se um [×]. Para o componente seletivo *profissões*, anotam-se o comprimento *dez* e um [×]. A ligadura vertical indica a construção de uma tabela de dupla entrada na qual é indiferente em qual eixo (x ou y) os componentes serão representados, pois as notações são independentes da ordem dos signos (“independentes da ordem do discurso.” (ibid., p.239)), sendo possível colocá-los em linha ou coluna. A tabela dá origem ao esquema de homogeneidade à sua direita, no qual as células da tabela são preenchidas ou deixadas em branco conforme um indivíduo específico pratique ou não determinada profissão, processo indicado pelo símbolo [01] no eixo z. Sendo assim, o esquema de homogeneidade possibilita contabilizar as dimensões da tabela, estaabelecendo as dimensões do problema. No caso, 250 colunas × 10 linhas, resultando em uma tabela de 2.500 células.

Tabela 2 – Símbologia usada na tabela de alocação

notação	relação representada
×	Componente seletivo, no qual não existe uma relação inerente de ordenação
Q	Componente quantitativo (como <i>coeficiente de rendimento</i>) posicionado no eixo z
↗ 01	Componente seletivo (respostas binárias do tipo “sim” ou “não”) posicionado no eixo z
↗ –	Presença parcial de certa característica
	Ligadura vertical: cruzamento entre os componentes que constituem uma tabela de dupla entrada
—	Ligadura horizontal: sinaliza elemento comum a duas tabelas correspondentes. O(s) componente(s) que unido(s) pela ligadura são identificados como potenciais pontos de comparação
\	Dados comuns; dados sem objetos para certas classes; dados desconhecidos

Fonte: autor (baseado em BERTIN, 1977).

Figura 12 – Tabelas de alocação



Fonte: BERTIN, 1977, p. 236

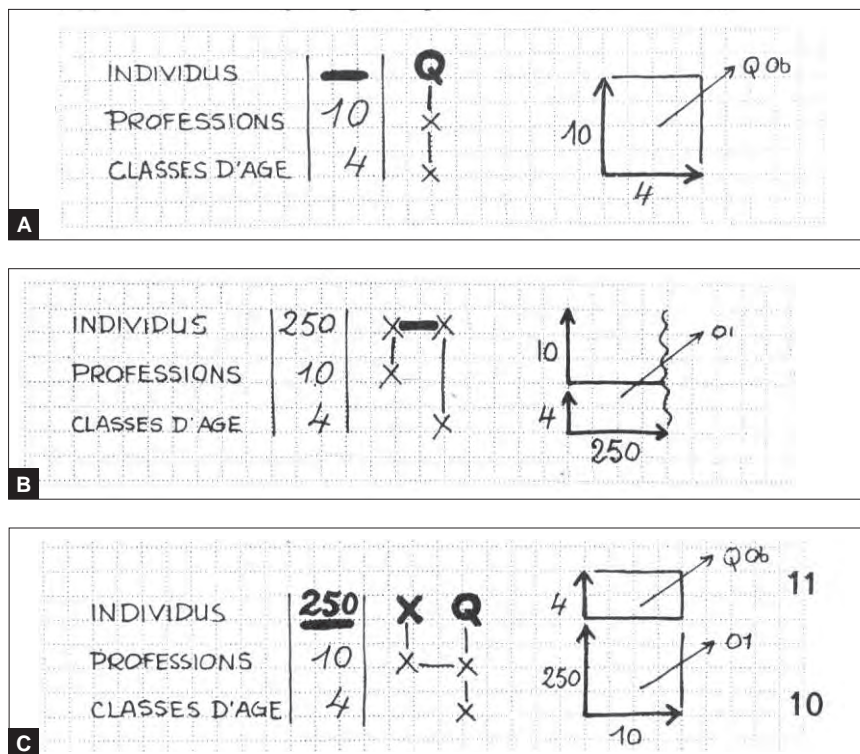
(Figura 12B) Caso em que a idade de cada indivíduo é conhecida, tratando-se assim de um componente quantitativo (*quantidade de anos*), indicado pelo símbolo [Q]. Nessa situação, os valores podem ser transcritos em uma única linha da tabela.

(Figura 12C) A diversidade de valores do componente *idade* poderia ser reduzida a três classes de idade, como (1) 0-17; (2) 18-35 e (3) 36-60 anos, tornando-a, assim, um componente ordenado e sendo identificada apenas pela numeração da respectiva classe (1, 2 ou 3). Nesse caso, ao invés do símbolo [Q], emprega-se a notação [O] indicando que em z será assinalada a classe de idade respectiva a cada indivíduo.

A figura 13 apresenta situações envolvendo mais de dois componentes, como:

(Figura 13A) Não se sabe a profissão e a idade de cada indivíduo, mas sim a quantidade de indivíduos segundo dez profissões para cada uma das quatro classe de idade. Isso poderia ser exemplificado considerando-se que, por exemplo, na classe de [31-40 anos] há dez professores, dois empresários, cinco médicos etc., na classe [41-50 anos], encontram-se dois atores, três produtores, cinco empresários etc. Neste caso, demarca-se um traço no componente *indivíduos*

Figura 13 – Tabelas de alocação com mais de dois componentes



Fonte: BERTIN, 1977, p. 236, 237.

indicando que eles não serão analisados unitariamente e anota-se um [Q] assinalando que os valores deste componente (a quantidade de objetos estatísticos, Qob) serão demarcados em z. As dimensões da tabela são 10×4 , resultando em 40 células.

(Figura 13B) Caso em que existe uma diferença nos registros sobre os indivíduos. Sabe-se a profissão de cada um, mas a idade é conhecida apenas segundo quatro faixas etárias. A ligação horizontal indica que a componente *indivíduo* é comum tanto às *profissões* quanto à *idade*, o que resulta em um esquema de homogeneidade com duas tabelas justapostas verticalmente, cuja dimensão total é dada pelo cálculo de suas áreas: $250 \times (10+4) = 3500$ células. No eixo z, a notação [01] indica que os dados são apresentados na forma do pertencimento ou não de cada um dos 250 indivíduos a determinada profissão e a uma classe de idade específica.

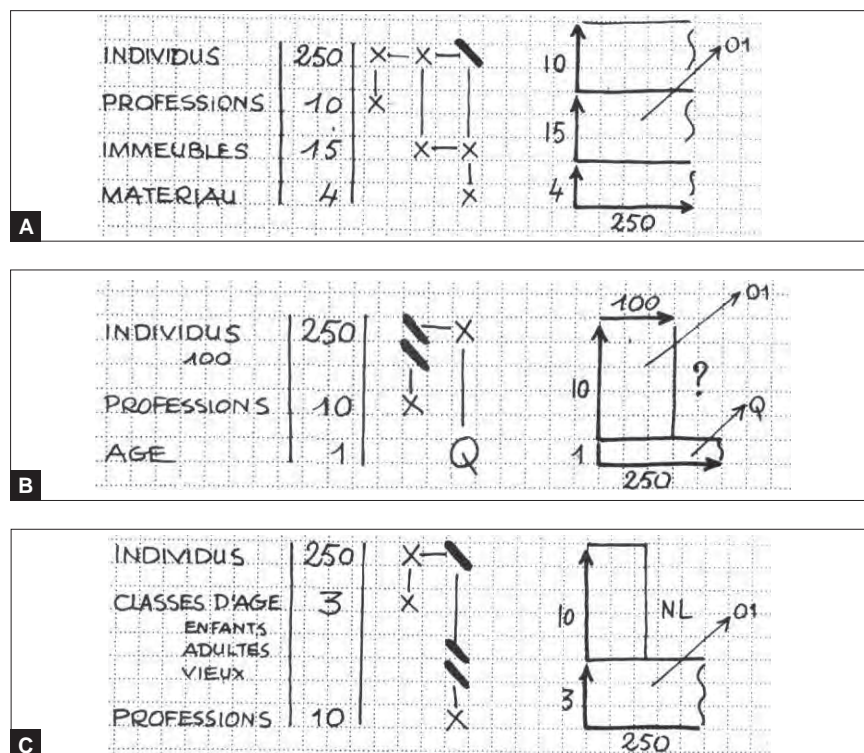
(Figura 13C) Quando se sabe a profissão para cada indivíduo, mas a idade é conhecida pela quantidade de indivíduos em cada classe de idade. Neste caso, o componente *indivíduos* apresenta um comportamento duplo: completo em relação à profissão, mas segmentado em relação à idade, situação representada pelos símbolos [250] e [—]. Essa notação correspondente a

um [X] e um [Q] nas colunas adjacentes significando a necessidade de duas tabelas para representar o componente em questão. Na primeira, ele acontece na forma “sim-não” [01] e na segunda, na forma de quantidades de objetos [Qob]. A ligadura horizontal no componente *profissões* indica que as duas tabelas podem ser comparadas tendo esse componente como elemento comum. Caso em que se tem uma tabela cujas dimensões são calculadas multiplicando-se 10 (profissões) por 250 (indivíduos) + 4 (classes de idade), perfazendo 2540 células.

Bertin utiliza ainda a notação [\] para indicar a ocorrência de dados comuns a diversos componentes, dados sem objetos para certas categorias de um componente e também a ausência de dados.

(Figura 14A) Dados comuns: Sabe-se a profissão de cada indivíduo, assim como o material que caracteriza o imóvel de cada indivíduo. Desse modo a componente *material* “caracteriza o imóvel e os indivíduos que nele habitam.” (Bertin, 1977, p.239). Esta condição é indicada pelo símbolo [\] em *indivíduos*, que permite agrupar na mesma coluna *imóveis* e *material*, indicando assim o uso de três e não de quatro tabelas.

Figura 14 – Tabelas de alocação: uso da notação [\]

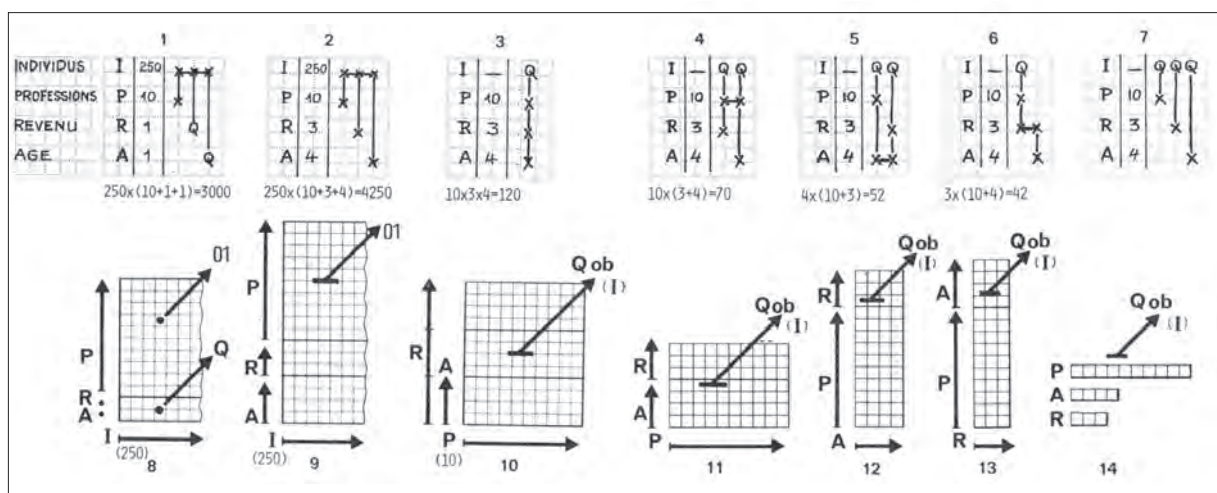


(Figura 14B) Ausência de dados. Indicação de que dos 250 indivíduos, sabe-se a profissão de 100 deles de modo que se reserva uma linha para os cruzamentos desse sub-grupo. No esquema de homogeneidade, isso é representado com uma área reservada aos dados desconhecidos, sugerindo seu preenchimento com base em tratamentos de interpolação (ibid., p.256-259).

(Figura 14C) Ausência de dados para certas categorias de um componente. Se a população inclui crianças, os dados sobre profissão não se aplicam a elas. Nessa situação, demarcam-se as categorias relevantes com a notação [\]. O esquema de homogeneidade correspondente apresenta uma área vazia para as categorias sem participação no conjunto de dados.

2 *Esquema de homogeneidade*: Trata-se do “esquema xyz da tabela de dupla entrada que os dados constroem” (Bertin, 1977, p. 240). Este instrumento apresenta uma visão sintética do conjunto de dados a ser visualizado sumarizando em um diagrama o tipo, o comprimento e as relações entre os seus componentes (Figura 2.10). Seu objetivo é “definir o ponto de comparação do conjunto de um estudo, isto é, o componente colocado em x, de modo que todos os componentes do estudo podem ser colocados em y.” (ibid.). Como já mencionado, tal ponto de comparação é indicado pela ligadura horizontal entre as cruces, podendo inclusive haver mais de um. Quando isso acontece, é preciso estudar as várias possibilidades. Assim, o esquema de homogeneidade expõe as escolhas possíveis e, nesse sentido, possibilita definir as dimensões da tabela necessária, avaliar e estabelecer os pontos de comparação e identificar o método de tratamento mais adequado. A Figura 15 indica diversas relações entre tabelas de

Figura 15 – Relação entre tabelas de alocação e esquemas de homogeneidade



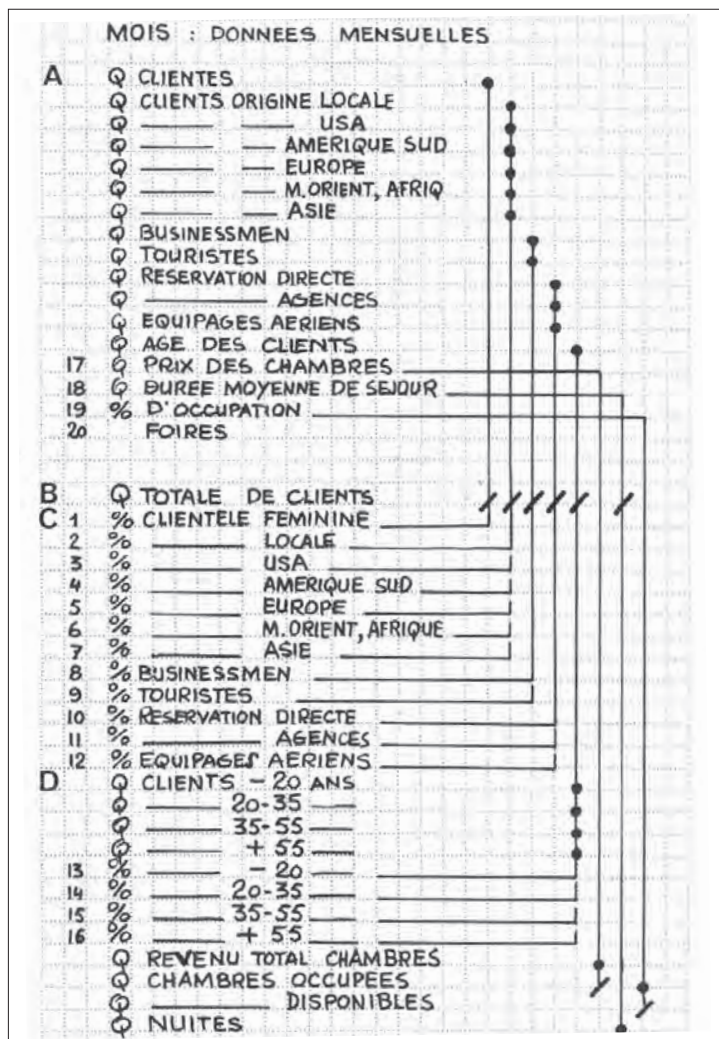
alocação e esquemas de homogeneidade. Dentre as várias situações apresentadas, nota-se, por exemplo, que quando só há uma linha vertical (item 3) qualquer componente demarcado com uma cruz pode ser colocado em x. Já a existência de várias linhas verticais, mas nenhuma horizontal (itens 7 e 14) é um caso que indica um estudo heterogêneo, no qual não há ponto de comparação entre os componentes.

3 *Tabela de pertinência:* Trata-se de um instrumento para notação e elaboração de hipóteses que permite arrolar as operações necessárias definindo os cálculos mais convenientes e a eventual necessidade de novos dados. Bertin ilustra o funcionamento da tabela de pertinência com uma análise do comportamento da clientela de um hotel ao longo do ano, a partir do registro de diversas características referentes a um grupo de 10.000 pessoas (Figura 16). O esquema de homogeneidade correspondente revela a necessidade de uma tabela com 18.000 casas, muito extensa para o tratamento. É necessário, então, reduzi-la agrupando os clientes por alguma das características. Neste caso, a informação procurada pode ser sintetizada na pergunta “Quais as características dos períodos de maior ocupação?”. A natureza cronológica do questionamento aponta para agregar os dados sobre os 1000 clientes estudados pela componente *data*. “Resta definir os cálculos utilizados para construir a tabela final e verificar se todos os indicadores utilizados são comparáveis e significativos. É o objetivo da tabela de pertinência.” (Bertin 1977, p.243).

Na tabela estão enumerados os componentes de modo que se possam estabelecer as relações entre sua forma inicial e final. Em relação à quantidade de clientes, opta-se pelos valores percentuais ao invés das quantidades absolutas, pois as relações expressas em porcentagens são mais significativas. Inicia-se com a demarcação do sinal (•) ao lado dos componente originais em (A), na parte superior da tabela. Tais componentes são divididos pelo valor total de clientes em (B), operação assinalada por (/) ao lado do valor absoluto e uma linha em direção à forma final (o valor em porcentagens). Nesse processo, as idades são sintetizadas em quatro classes (D) e o preço médio mensal dos quartos (17) é obtido pela divisão da receita dos quartos pelo número de quartos ocupados. A duração média da estadia (18) calcula o número total de noites dividido pelo total de clientes. Para a porcentagem de ocupação, divide-se o total de quartos disponíveis pelo total de quartos ocupados.

A Figura 17 apresenta uma tabela de pertinência mais complexa, que exemplifica o registro de hipóteses para um estudo de planejamento do território, pesquisa cujos dados

Figura 16 – Tabela de pertinência: ocupação de um hotel



Fonte: BERTIN, 1977, p. 244.

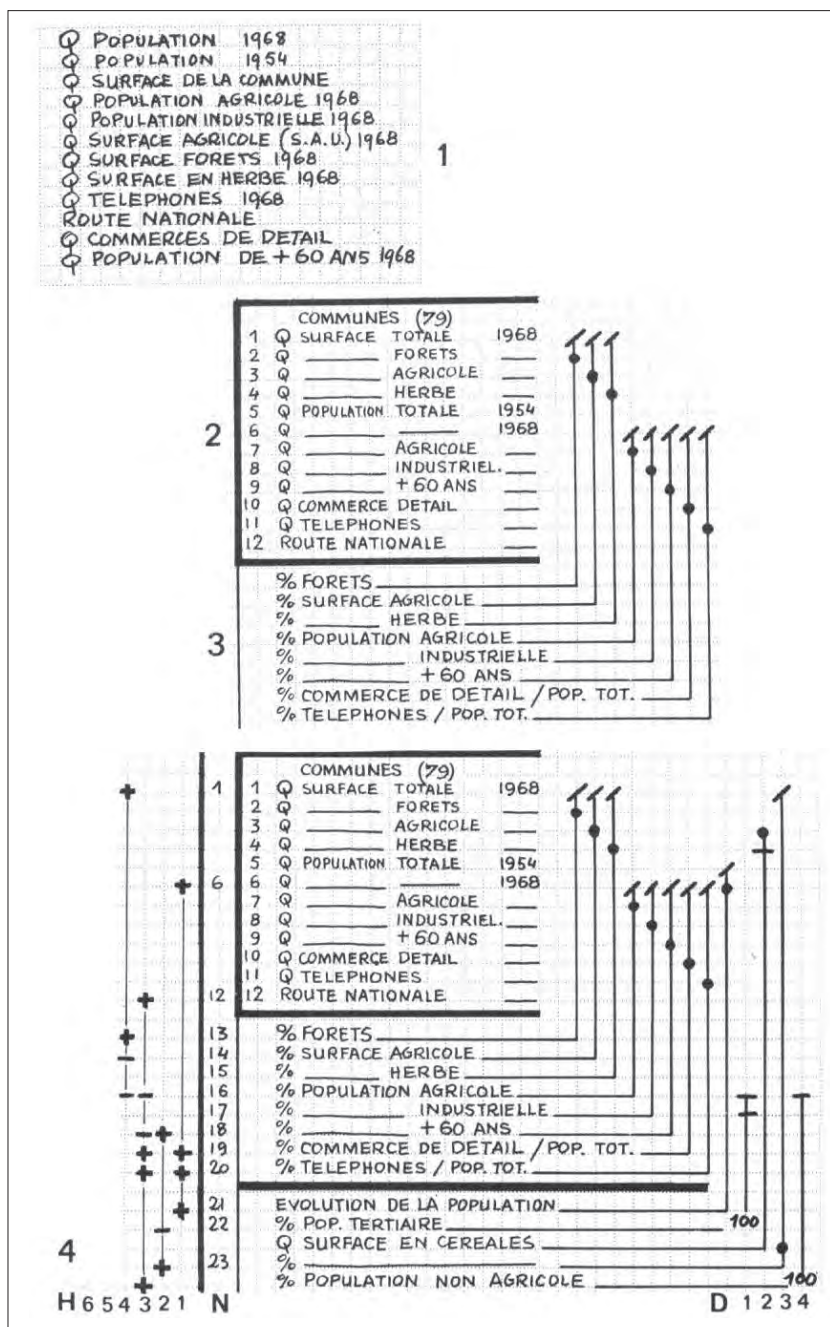
absolutos precisam ser substituídos por relações para evitar a discrepância na representação entre cidades grandes e pequenas. Esses cálculos são anotados em (3). A partir daí, é possível estabelecer hipóteses e anotá-las à esquerda da tabela.

1ª hipótese [H1]: Existe relação entre o aumento da população e o aumento de telefones e comércio? Na tabela (4), listam-se os elementos desta questão: + (quanto maior) a população, + (maior a %) de comércio, + (maior a %) de telefones. Descobre-se que resta saber a taxa de aumento da população (21).

2ª hipótese [H2]: O aumento da população idosa seria inversamente proporcional à população terciária e diretamente proporcional à cultura de cereais? Anota-se: [+] (quanto

maior) a população idosa, [-] (menor) a população terciária. Descobre-se que é preciso calcular a população terciária. Para tanto, subtrai-se de 100, a soma da percentagens das populações industrial e agrícola, como indica a coluna D1. Anota-se ainda [+] (quanto maior) superfície cultivada com cereais. É necessário calculá-la em duas etapas, pela quantidade de superfície (D2) e pela porcentagem (D3).

Figura 17 – Tabela de pertinência: planejamento de território



3ª hipótese [H3]: A estrada nacional seria favorável [+] ao comércio varejista, ao telefone e à população não agrícola? Mas desfavorável [-] aos idosos? É necessário a inversão da linha “população agrícola” para revelar a população “não agrícola”.

4ª hipótese [H4]: É verdade que as maiores [+] comunidades são as florestais e nelas a superfície agrícola e a população agrícola tendem a ser menores [-]? Nesse caso, todos os dados estão disponíveis.

5ª hipótese [H5]: As comunidades florestais situam-se no norte e favorecem [+] a população industrial? Para responder, é necessário um novo indicador “comunidades do norte” em um mapa (24).

Por fim, a coluna [N] assinala as questões pertinentes e os dados que serão efetivamente empregados para dar resposta a esse conjunto de hipóteses.

Dessas operações resultam os dados na forma em que serão transcritos para uma tabela que é a base da visualização. “A tabela final mantém apenas os dados correspondentes às questões pertinentes”. (Bertin 1977, p. 249). Sendo assim, a tabela de pertinência dá forma aos dados, sendo um instrumento para a transformação consciente de um conjunto extenso de dados em formas que se mostram mais úteis e inteligíveis para dar resposta às perguntas propostas no âmbito de determinado estudo.

2.2.2 Construção normal e permutação

A etapa seguinte diz respeito à definição das melhores construções gráficas para a visualização dos dados consolidados na análise matricial. Bertin entende como construção eficaz aquela em que se leva menos tempo para entrever os grupos formados pelos dados de modo que as comparações acontecem rapidamente, ao passo que nos diagramas inúteis, para se obter a mesma informação pode-se levar muito mais tempo. Assim, a eficácia é atingida quando a informação na construção gráfica é perceptível de modo instantâneo. “Fora da construção normal, a maior parte das construções não fornece resposta visual além do nível elementar de informação.” (BERTIN, 1977, p.15). A construção normal é um gráfico matricial $x \times y$ que modela a tabela de dados desde que: “1º [...] os números sejam transcritos em z por uma variação visual ordenada do branco ao preto ou do pequeno ao grande; 2º [...] as linhas e/ou colunas sejam reclassificadas para mostrar os agrupamentos.” (idem).

Mas de modo geral, e inclusive no âmbito dos dashboards, predominam gráficos que se restringem ao nível elementar de informação, como aqueles da Figura 18, que visualizam os dados da tabela à esquerda sobre a produção de carne nos países do Mercado Comum em 1966 (BERTIN, 1977, p.195). Nesse nível, o gráfico só oferece resposta a questões simples, como “em tal país, qual o maior rebanho?” ou “considerado o rebanho bovino, qual o país com maior produção?”. Bertin questiona a vantagem desses gráficos em relação à simples leitura da tabela, já que eles não são capazes de informar de modo abrangente o interrelacionamento específico deste grupo de dados. “Como se reagrupam ou se opõem as cinco nações no âmbito dessa informação sobre a produção?” Como não dão resposta a essa questão, o autor declara que “Estes diagramas são inúteis.” (ibid., p.195). A Figura 19, relaciona algumas telas de dashboards. Verifica-se que muitos dos gráficos utilizados estariam classificados nessa categoria, o que é indício também de um problema inerente à própria informação comunicada, situada no âmbito dos níveis mais básicos de informação.

Figura 18 – Gráficos no nível elementar de informação

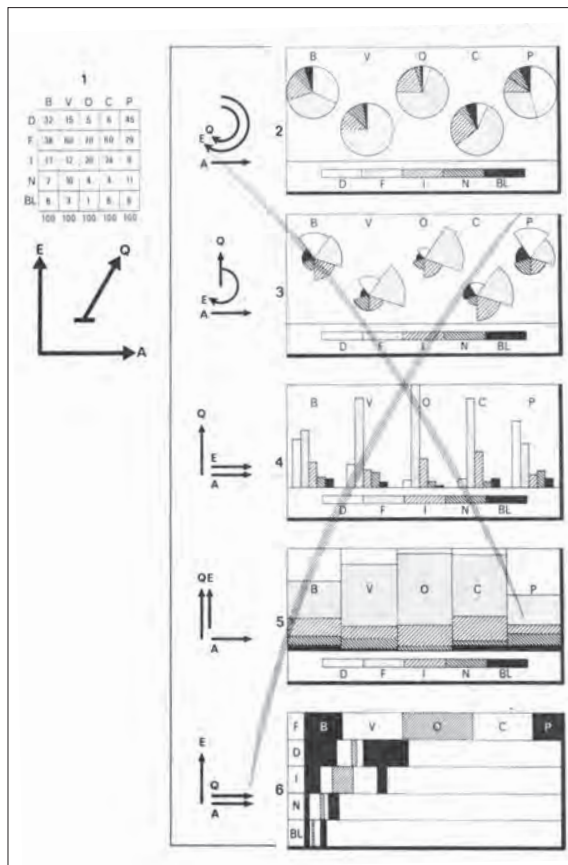


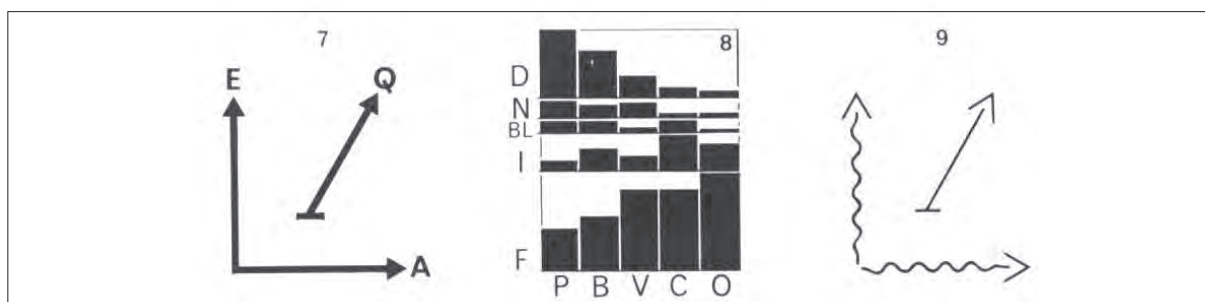
Figura 19 – Uso de gráficos no nível elementar de informação em dashboards



Em contraponto, a *construção normal* (Figura 20) é uma *imposição em diagrama* que revela, a partir do mesmo conjunto dos dados, “um problema estritamente político” (ibid., p.195). Na imagem construída, evidenciam-se dois grupos contrários: “a Alemanha e os Países-Baixos de uma parte; a Itália e a França de outra, que só podem ter políticas opostas... ou complementares. E assim a União Belgo-Luxemburguesa assume o papel de árbitro.” (ibid., p. 195). Para possibilitar esta conclusão é necessário reduzir ao mínimo as variações visuais para evidenciar os padrões de diferença e as anomalias. Por conta disso, o gráfico utilizado varia apenas no posicionamento nas duas dimensões no plano e no *tamanho* das barras que representam a quantidade de produção pecuária. Mas além disso, é o procedimento de rearranjo da ordem original dos componentes na tabela, que possibilita chegar aos padrões significantes.

Tais rearranjos são chamados por Bertin de *manipulações*, sendo “baseadas nas propriedades da percepção abrangente do olho. O olho vê os conjuntos, pode comparar e aproximar duas linhas semelhantes, isto é, suprimir as distâncias não significativas.” (ibid., p.20). Nesse sentido, o objetivo das manipulações, que podem acontecer nas chamadas *matrizes de permutação*, é simplificar a totalidade da visualização a uma imagem significativa, um padrão cujas partes são associadas em razão de seu significado visual. [A] “reclassificação visual permite definir os grupos característicos e as situações particulares (...)”. A figura 21 ilustra etapas possíveis em um processo de permutação em uma pesquisa de nove características em 16 cidades. Os diagramas partem da transcrição básica de um conjunto de dados específico (com suas possibilidades de reordenamento também específicas) até a formulação de novas categorias compiladas após a permutação.

Figura 20 – Construção normal: produção de carne no Mercado Comum Europeu (1966)

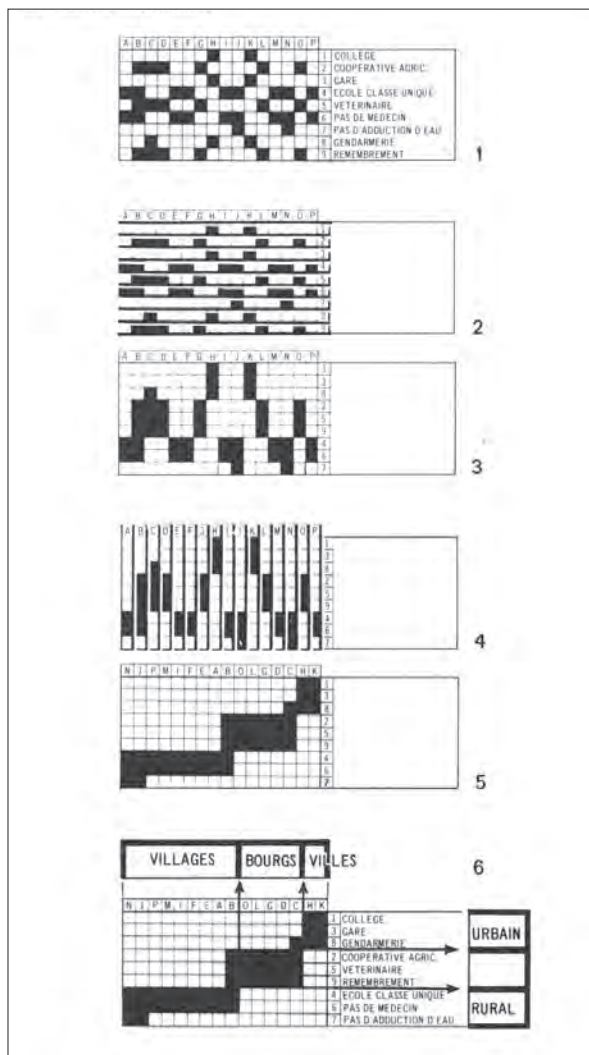


7: Esquematisação da construção normal / 9: as linhas onduladas indicam que o gráfico poder ser permutado tanto no eixo x, quanto no eixo y (por serem componentes sem ordem estabelecida)

8: eixo y: D. Alemanha, N. Países Baixos, BL. Bélgica-Luxemburgo, I. Itália, F. França
eixo x: P. porcos, B. bovinos, V. veados, C. cavalos, O. ovinos

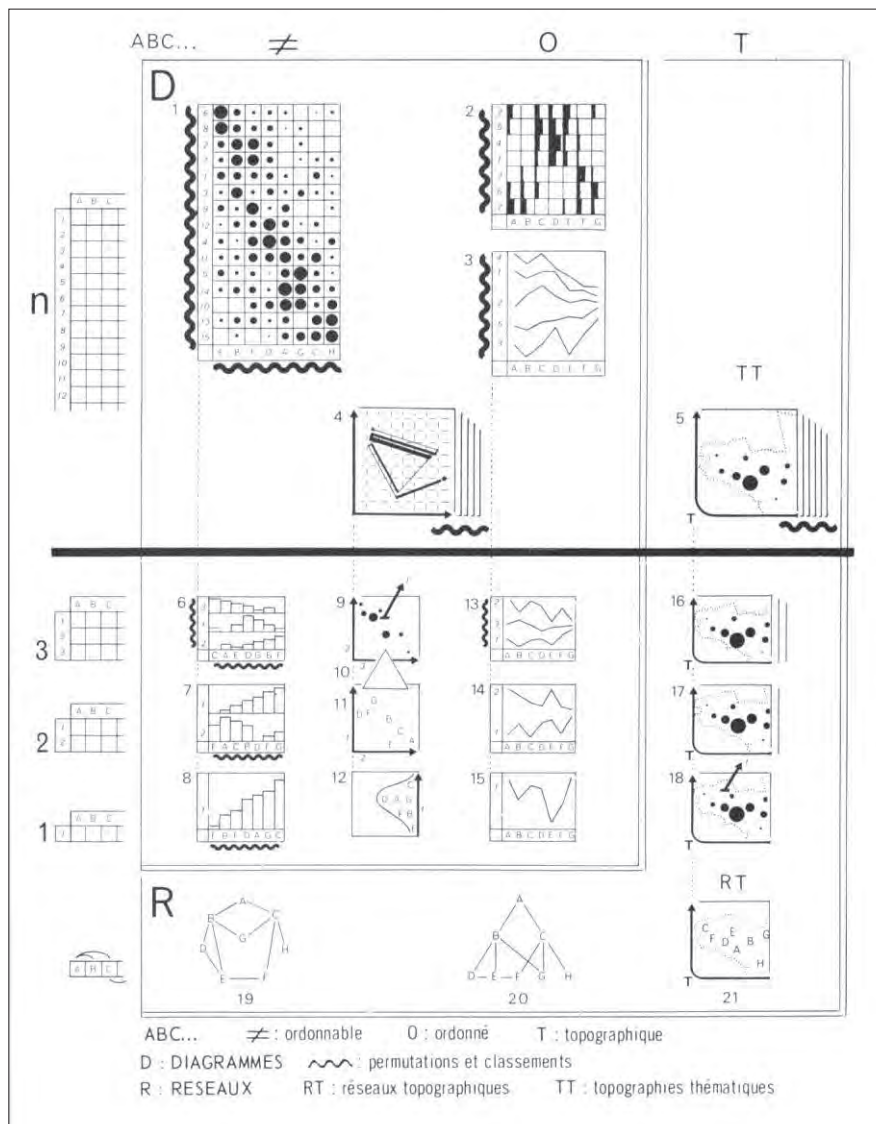
Para identificar os gráficos que permitem rearranjo, Bertin relaciona-os no sinóptico da Figura 22, que inclui diagramas, redes e cartografias, analisando-os em função do número de linhas presentes na tabela base (1, 2, 3 ou n). Para o presente trabalho importa apenas aquilo que se refere aos *diagramas* de n linhas, uma vez que os dados a serem visualizados têm mais de três linhas e não apresentam relação topológica entre eles. Nesse sentido, o diagrama básico é a *matriz ordenável*, “utilizável nos casos em que a tabela de dados tem a forma ($\neq \neq$) e não ultrapassa $x \times y = 10.000$, aproximadamente.” (BERTIN, 1977, p.32). Trata-se assim de um diagrama em que os componentes em x e em y, têm propriedades *ordenáveis* (propriedade identificada pelo sinal (\neq), não existindo uma ordem pré-estabelecida entre os dados, permitindo sua livre

Figura 21– Etapas do processo de permutação



Fonte: BERTIN, 1977, p. 33

Figura 22 – Sinopse das construções gráficas



Fonte: BERTIN, 1977, p. 29

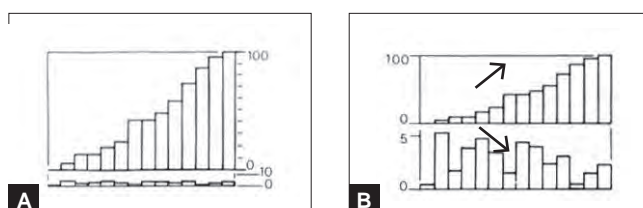
reorganização, indicada pela linha ondulada na imagem D1 da Figura 22. Por conta disso, a matriz ordenável é a construção fundamental dentre os diagramas que permitem permutação de modo que “todas as outras construções normais são casos particulares.” O diagrama que será empregado no presente trabalho é o *fichário-imagem*, (item 2, Figura 22). Trata-se “da construção normal de tabelas de dados (\neq O), ou seja, quando há um componente ordenado.”, (sendo os outros componentes *ordenáveis*). Nesse caso, só é possível fazer permutações em uma das dimensões do gráfico, pois na outra há um componente de organização fixa como dados cronológicos. Um exemplo de fichário-imagem pode ser visto no gráfico da Figura 25, em relação às características de ocupação de um hotel ao longo do ano.

2.2.3 Questões de visibilidade

Trata-se dos tópicos relacionados ao discernimento visual das variações em uma imagem e que dependendo do uso podem revelar ou ocultar padrões visuais de interesse analítico. Nesse sentido, Bertin discute o conceito de escalas absolutas (*échelles communes*) e escalas relativas (*échelles propres*), questão que se aplica em especial aos diagramas, como as *matrizes ordenáveis* e os *fichários-imagem*, situações em que há comparação entre diversos *perfis* (histogramas). Nesses casos, o uso de uma escala que é comum a todos os perfis referencia toda a série por esse mesmo padrão. Entretanto, esse procedimento que, *a priori* seria o mais preciso e fiel aos dados pode ocultar informações significantes (como na Figura 23A). Assim, Bertin afirma que nos diagramas simples “a escala absoluta pode ser empregada, desde que os perfis das linhas com menor valor sejam perfeitamente visíveis e comparáveis aos outros perfis.” (ibid. p.211). Por outro lado, quando isso não acontece é preciso que cada linha tenha sua escala específica expandida o suficiente para que se torne visível: “nas matrizes e no fichário-imagem, a escala relativa é a regra geral quando a variação visual disponível é muito reduzida.” (ibid.). Assim, a Figura 23B expande a proporção vertical do histograma inferior, revelando uma tendência oposta ao gráfico de cima, padrão que só se torna visível devido à expansão.

Outros aspectos análogos são tratados na seção sobre a *lei da visibilidade* que é formulada no sentido de garantir a separação visual entre os elementos de um gráfico. Bertin declara que “Todo elemento que não é discernível é inútil e reduz a visibilidade da imagem.”. Devido a isso, o autor recomenda “suprimir aquilo que é comum a todos os elementos da informação [pois] aquilo que é comum não estabelece separação visual.” Orienta ainda a “utilizar todo o

Figura 23 – Escala absoluta e escala relativa

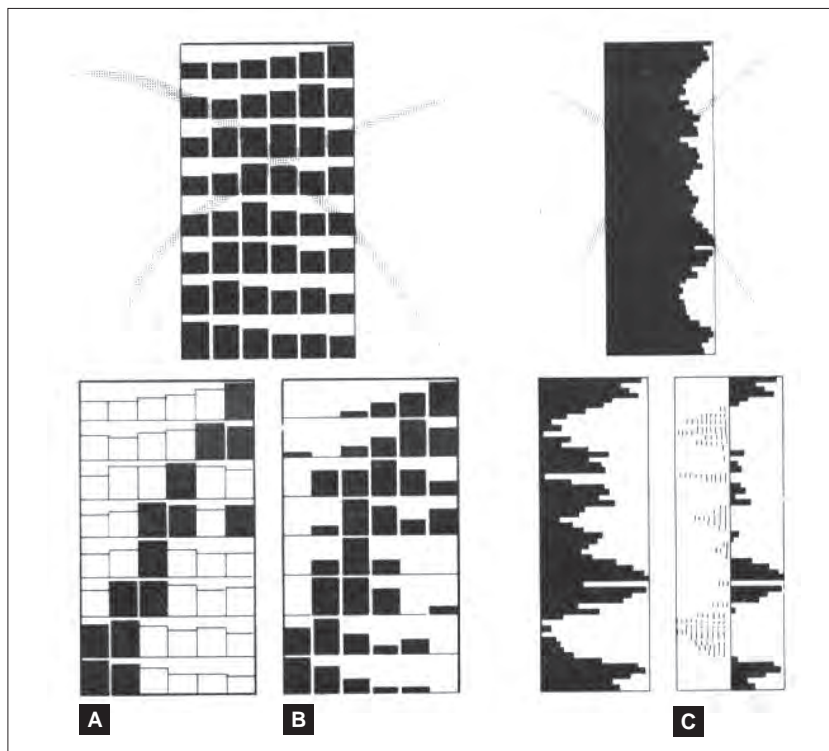


Fonte: Adaptado de BERTIN, 1977, p. 210.

comprimento da variável visual para transcrever a informação discernível” (ibid. p.228). Isso se traduz pelo uso de valores de luminosidade abrangendo toda a gama de variações do branco ao negro, não começando a série por um cinza intermediário, por exemplo, o que desperdiçaria as variações úteis dos tons de cinza mais claros.

A figura 24 aponta algumas das aplicações da lei da visibilidade. Em 24A, Bertin sugere manter em preto apenas os valores acima da média. Isso fornece um critério para a permutação baseado na semelhança entre perfis, caracterizando as correlações entre eles e identificando padrões relevantes. Em 24B os mesmos dados são tratados de modo que a parte comum a cada histograma seja suprimida. A diferença restante é então expandida, resultando em uma vista muito mais nuançada das diferenças entre os elementos. Bertin sugere ainda que “para um fichário imagem muito grande, a melhor solução é a que escurece aquilo que ultrapassa uma linha média.” (ibid. p. 229), como visto na figura 24C.

Figura 24 – Aplicações da lei da visibilidade



Fonte: BERTIN, 1977, p. 210

2.2.4 Gráfica de comunicação

Refere-se à etapa do tratamento gráfico que sucede o tratamento dos dados e cuja finalidade é precisamente comunicar os estudos realizados. “A gráfica de comunicação é um meio de fixar e de dizer aos outros o que foi descoberto. Sua meta: a percepção rápida e eventualmente a memorização da informação de conjunto. *Seu imperativo: a simplicidade*.”. (BERTIN 1977, p.22, grifo do autor). O autor sublinha que o procedimento de simplificação deve acontecer oportunamente, isto é, em função das necessidades de um âmbito informativo particular “é preciso evitar [...] redigir documentos simplificados quando o exaustivo é necessário, ou sobrepor documentos exaustivos acreditando ampliar a informação. A falta de visibilidade rapidamente torna o gráfico inútil.” (ibid., p.22). Nesse sentido, a gráfica de comunicação sintetiza o conjunto de procedimentos que viabilizam a compreensão de um gráfico, fazendo a intermediação entre os dados e o usuário, de modo que ele possa apropriar-se do conteúdo dos dados, condição em que eles podem ser considerados, de fato, informação.

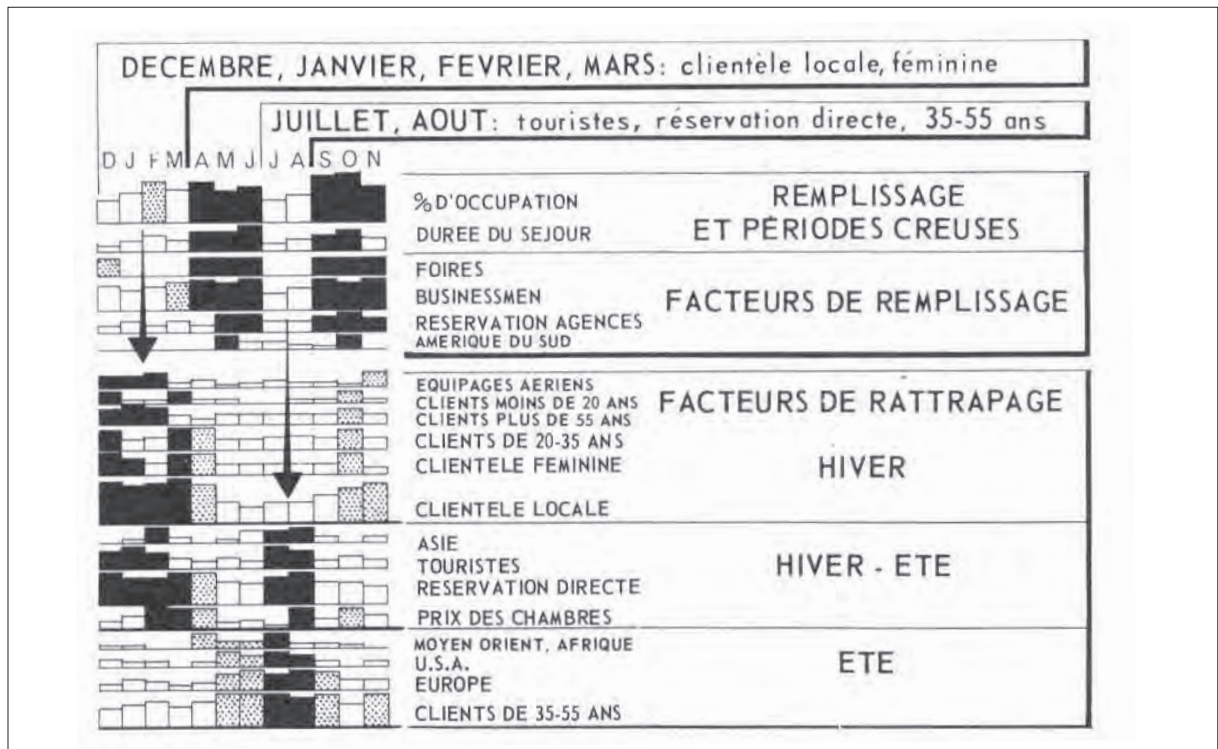
Isso é facilitado por meio de recursos visuais específicos que buscam explicar ou destacar uma circunstância, um padrão recorrente ou uma anomalia significativa. Vão desde a organização espacial dos elementos do gráfico, para que os grupos significativos encontrados após a permutação sejam claramente individualizados, até a inserção de elementos visuais chamando atenção para aspectos específicos. Ou ainda, a hierarquia tipográfica que possibilita compreender as ordens de prioridade na informação apresentada. Neste sentido, vê-se que tais estratégias de comunicação visando facilitar a compreensão relacionam-se estreitamente ao que hoje nomeamos comecemos *design de informação*, cuja finalidade apontada por Bonsiepe seria a de “facilitar o metabolismo do conhecimento, ou seja, a transmissão e a assimilação do saber.” (2001, p.7).

O exemplo do hotel (Figura 25) indica os tipos de recursos que Bertin se refere ao mencionar a gráfica de comunicação e o modo como ela procura ressaltar aquilo que foi descoberto na fase de tratamento. Trata-se de uma indicação clara da resposta à pergunta inicial “Quais as características dos períodos de maior ocupação?”. Respondê-la demanda o entendimento de diversos fatores correlacionados, e, desejavelmente, indicações para reverter a ocupação nos períodos de pouca frequência. Nesse sentido, o gráfico apresenta-se segmentado em duas partes principais, visíveis à primeira leitura: *Períodos de pico* e *Fatores de recuperação*. O primeiro

grupo é subdividido em *ocupação* (com os indicadores relativos à ocupação e duração da estadia e *fatores de ocupação*, que relacionam as causas que justificam essa demanda. O segundo grupo divide-se em três segmentos que nomeiam os três períodos de oportunidades de recuperação da clientela: *inverno*, *inverno-verão* e *verão*. Esses grupos concentram os indicadores com tendência oposta aos períodos de pico e, portanto, podem sugerir possibilidades de incrementar a ocupação do hotel por meio de ações especificamente orientadas ao grupo de clientes predominante nesses meses.

Tal padrão oposto é indicado pelas setas, que mostram que nos meses em que a ocupação é menor, existem componentes que indicam a tendência oposta, ocorrendo com maior intensidade. Este critério de comparação é dado pelos elementos que estão acima da média, indicados nas cores preto e cinza. (Note-se que o cinza é utilizado para rebaixar os elementos que não contribuem para delinear com clareza o padrão que se quer ressaltar, que foi enfatizado em preto).

Figura 25 – Gráfica de comunicação



2.2.5 Síntese

O diagrama do hotel sintetiza a abordagem de Bertin em relação às questões de comunicação evidenciando características análogas entre os aspectos da gráfica de comunicação e as propriedades e requisitos comumente presentes no âmbito dos dashboards, como por exemplo:

- 1 gráficos que evidenciam um padrão significativo de determinado conjunto de dados.
- 2 elementos que organizam, hierarquizam e explicam informações relevantes.
- 3 a natureza do relacionamento entre os dados, passível de ser percebida em um relance, constituindo-se como informação principal e foco de atenção.
- 4 elementos de ênfase que definem hierarquias de importância em relação àquilo que se quer saber, como a diferença de valor entre preto (destaque) e cinza.
- 5 elementos de indicação, como as setas que apontam para a evidência do tópico a ser comunicado.
- 6 elementos textuais sintéticos que descrevem clara e rapidamente os pontos principais em que a informação se divide, enfatizando os aspectos que podem embasar uma tomada de decisão, como os meses com maior incidência de ocupação.

Nesse sentido, a metodologia de Bertin sistematiza os recursos de ênfase, explicação, evidência, descrição, síntese e rapidez de comunicação, aspectos que tangenciam o âmbito habitual dos dashboards e que, portanto, mostram-se como critérios e técnicas coerentes para embasar os dashboards propostos nesta dissertação.

3 PROPOSIÇÃO

Neste ponto, retoma-se a proposta deste trabalho relacionando-a ao que foi exposto até aqui, para então formulá-la com maior definição. Dashboards são artefatos que objetivam efetivar a comunicação informativa de dados armazenados em bases computacionais. Tal comunicação se dá, prioritariamente, por meio de imagens (como gráficos, mapas e diagramas) que visualizam o tema a ser analisado em uma área delimitada. Nesse sentido, para orientar os estudos desenvolvidos, enumeram-se as principais características definidoras de um dashboard, como consideradas no presente trabalho:

- 1 Painel composto predominantemente por elementos visuais organizados de tal modo que estejam inscritos em uma única tela.
- 2 Sintetiza um sentido de completude em relação ao tema tratado para resultar em um conjunto explicativo coeso.
- 3 Indica de um viés de análise, orientando a atenção do leitor aos pontos considerados mais relevantes.
- 4 Favorecem o rápido reconhecimento dos indicadores potencialmente negativos, assim como a rápida compreensão dos fatores que evidenciam tal situação
- 5 Possibilitam investigações mais detalhadas a partir da condição inicial do dashboard.

As propostas de Jacques Bertin tratam da transcrição de dados em construções gráficas sistematizando os estímulos visuais que melhor comunicam determinada condição dos dados. Além disso, descreve um método para organizar o conjunto de dados de modo coeso; os meios de análise gráfica que podem revelar padrões significantes em um “golpe de vista” e também as formas de comunicá-los, atendendo a demandas informativas específicas. Neste sentido, parece haver uma convergência entre meios e objetivos que justifica a análise dos dashboards pelas teorias do cartógrafo francês.

Considerado isso, a hipótese principal da presente dissertação é a viabilidade do projeto de dashboards que empreguem a metodologia de Bertin, e com isso, comunicar informações que os dashboards convencionais (de modo geral, aqueles compostos predominantemente por gráficos situados no nível de informação *elementar*) não são capazes de fornecer. Será utilizada como fonte de dados o conteúdo disponibilizado pelo *Projeto de Business Intelligence*

da UERJ por meio de planilhas dinâmicas. Os dados escolhidos são aqueles que possibilitam uma análise da questão das políticas afirmativas implantadas na Universidade a partir de 2003. Além disso, pode-se entrever uma segunda investigação que será apenas principiada nesta dissertação: a transposição de alguns aspectos das teorias de Bertin para o domínio digital. Isso acontece por meio da sugestão de algumas possibilidades interativas (ainda que apenas por meio de esquemas visuais), indicando situações em que as técnicas de visualização propostas pelo cartógrafo francês poderiam ser otimizadas por meio dos recursos computacionais. Nesse sentido, como uma hipótese secundária, são esboçadas algumas sugestões de uso interativo, procurando entrever se metodologia de Bertin conserva sua validade (ou parte dela) quando técnicas de interação são utilizadas, e também investigar situações em que o aporte de recursos computacionais pode contribuir para o desenvolver as propostas de Bertin.

4 APLICAÇÃO E DISCUSSÃO

4.1 Portal de Indicadores UERJ

4.1.1 Estrutura técnica: noções gerais

Em 2003, a Universidade do Estado do Rio de Janeiro foi a primeira universidade brasileira a adotar a política de reserva de vagas na admissão de alunos. Esse programa enquadra-se no contexto das políticas de ações afirmativas objetivando incluir segmentos sociais historicamente ausentes do ensino superior. Para atender à demanda por informações que possam orientar o desenvolvimento do programa de cotas e responder ao interesse público sobre o tema, a instituição resolve integrar e sistematizar seus bancos de dados sobre a população universitária.

Isso acontece por meio do *Projeto de BI* da instituição, iniciado em 2009, conduzido pela Diretoria de Informática da universidade (Dinfo) e pela empresa privada W5, consultoria de tecnologia especializada em *Busines Intelligence*. Trata-se da aplicação dos recursos e conceitos de BI aos dados acadêmicos sobre a área de graduação, possibilitando consultas específicas, integração dos dados em sistemas pré-existentes na universidade e acesso *online* a eles. Um dos seus objetivos é subsidiar os gestores da universidade na tomada de decisões facultando recursos como o cruzamento de informações, possibilidade pouco viável nos sistemas anteriores devido à dificuldade de pesquisas comparativas.

Com esses objetivos, o Projeto de BI lida com os dados com base no que denomina *Modelo de Negócio da Graduação*, que corresponde ao *Ciclo de Vida do Aluno*, modelo que divide-se em três segmentos: *ingresso*, *curso* e *egresso*.

O primeiro segmento (*ingresso*) agrupa os dados relativos à modalidade de ingresso do aluno, como exame vestibular, transferência, convênio, aproveitamento de estudos, reingressos etc. Reúne dados de diversos anos como vagas por carreira, campus, tipo de cota, nota mínima etc. Em relação ao vestibular armazena informações cadastrais dos candidatos como idade, sexo, moradia, nota do ENEM, classificação etc. Dentre as pesquisas possíveis nesse segmento estão: a nota média no vestibular por curso, modalidade de cota, região etc.; porcentagem de vagas

preenchidas por cotistas; taxas de ociosidade e quantidade de classificados não ingressantes. Ou ainda, o número de ingressantes por escolas de ensino médio; a quantidade de ingressantes em determinado ano; a porcentagem de ocupação por curso etc.

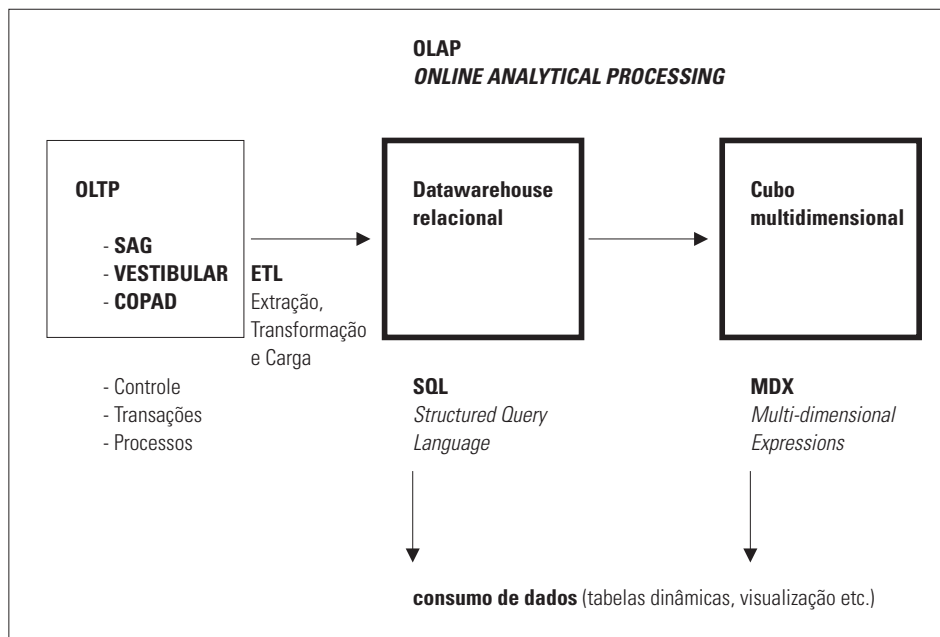
O segmento *curso* reúne os dados sobre a situação do aluno em determinado período, informação que envolve os processos de avaliação, inscrição em disciplinas, planejamento de turmas e planejamento das atividades docentes. Os dados podem ser pesquisados por categorias como período letivo, campus, carreira, turno, disciplina, faixa de nota, situação do aluno em uma disciplina específica etc. Nesse âmbito, podem-se investigar a nota média, desvio-padrão de notas por disciplina, porcentagem de notas não lançadas e informadas fora do prazo, porcentagem de abandono, desempenho entre cotistas e não cotistas.

O segmento sobre os *alunos egressos* agrupam registros como a data de egressão e da colação de grau e o período mínimo da grade. Vincula também dados sobre as diversas formas de egressão, como conclusão, transferências (internas e externas), abandono, solicitação do aluno, integralização esgotada, ausência de inscrição em disciplinas, afastamento por motivo de intercâmbio etc. Esta coleção permite questões com relação à quantidade de alunos formados e abandonos; o estágio do curso em que há maior incidência de abandonos; o tempo médio de conclusão entre cotistas e não cotistas, dentre outras.

4.1.1.1 Arquitetura técnica

A Figura 26 esquematiza a organização dos elementos na arquitetura básica dos segmentos do Projeto de BI - UERJ. Os dados da instituição são provenientes de três fontes: 1) Sistema dos vestibulares; 2) *Comissão permanente de carga horária e avaliação docente* (Copad); 3) *Sistema Acadêmico de Graduação* (SAG). Trata-se de sistemas baseados em aplicações OLTP (*Online Transaction Processing* – Processamento de transações *online*) voltadas para “o processamento de transações repetitivas em grandes quantidades e manipulações simples.” (Turban et al. p. 109). Por meio de uma ferramenta de Extração, transformação e carga (ETL), os dados são consolidados, categorizados, centralizados e carregados para o *data warehouse* da instituição, integrando-os em uma base de dados histórica e relacional. O projeto de BI - UERJ é baseado na plataforma de gestão *SQL Server 2008* que, além das ferramentas de gerenciamento de banco

Figura 26 – Arquitetura básica do sistema de BI da UERJ



Fonte: autor (a partir de informações fornecidas pela equipe de BI da UERJ).

de dados, permite integração aos aplicativos *Microsoft Office*. Em especial com as aplicações de gerenciamento de desempenho, destinadas ao usuário final, como os *software* de monitoramento *Performance Point Server* e o aplicativo de planilhas *Excel*.

O acesso ao *data warehouse* acontece por meio de um sistema *OLAP* (*Online Analytical Processing* – Processamento analítico online), empregado quando as necessidades de análise evolutiva envolvem “o exame de muitos itens de dados [...] em relacionamentos complexos.”. Nesses casos, a tecnologia *OLAP* possibilita um “método direto de suporte à decisão.” (Turban et al. p.109). Tais aplicações possibilitam também a organização dos dados em *cubos multidimensionais* de acesso *online*, modelados pela linguagem de programação *MDX* (*Multi-Dimensional Expressions*) que gerencia os cubos e possibilita pesquisa nos registros armazenados. Ainda segundo Turban et. al (ibid. p.118), cubos multidimensionais constituem um método de organização em que “conjuntos de dados altamente correlacionados [...] são organizados para permitir que os usuários combinem qualquer atributo de um cubo [...] com qualquer métrica [...]. Disto, resultam planilhas que possibilitam muitos tipos de análises, como detalhamentos, filtros, correlações etc. e ainda a intervenção de ferramentas de descoberta de padrões como as de *data mining*. O acesso aos usuários finais é disponibilizado por meio

dos softwares *Excel* e *SharePoint* que possibilitam a extração de dados na forma de relatórios, dashboards, planilhas, visões analíticas, planos e *scorecards*.

4.1.1.2 Cubos de desempenho acadêmico

Como visto, o sistema de BI da UERJ baseia-se no conceito de *multidimensionalidade* na organização dos seus registros. Mais especificamente, trata-se de uma técnica que disponibiliza os dados facultando ao usuário final organizá-los e pesquisá-los de modo prático e oportuno, conforme suas demandas. Nesse âmbito, Turban et al. (ibid. p.117) descrevem os três fatores fundamentais em relação à multidimensionalidade: *dimensões*, *medidas* e *tempo*. No BI UERJ, algumas das *dimensões* (determinado agrupamentos de dados que constitui uma categoria independente) são: aluno, tipo de cota, endereço, estado civil, carreira etc. A noção de *medidas* refere-se aos valores quantitativos tais como classificação no vestibular, quantidade de alunos, nota no Enem etc. O fator *tempo* qualifica os dados organizados em ordem cronológica de modo mensal, semestral, anual. Em tal contexto, uma pesquisa possível seria a de um diretor de unidade que queira saber sobre o desempenho de alunos provenientes de área geográfica específica, de uma turma específica em um semestre letivo em particular. Turban et al. relatam que respostas a esse tipo de pergunta podem ser obtidas de modo mais rápido “independentemente da estrutura do banco de dados, se os dados estiverem organizados em bancos de dados multidimensionais ou se a consulta [estiver projetada] para multidimensionalidade.” (ibid., p.118). Ainda segundo os autores, caso isto aconteça “os usuários podem navegar através de muitas dimensões e níveis de dados por meio de tabelas ou gráficos e fazer interpretações rápidas (ibid.).

O BI UERJ disponibiliza acesso ao banco de dados *Desempenho Acadêmico* mediante cadastro na Diretoria de informática (Dinfo). Essa coleção de dados subdivide-se em seis cubos: *Ingresso*, *Egresso*, *Graduação*, *Planejamento de atividades*, *Turmas* e *Vestibular*. Segundo a equipe de BI, os dois primeiros (*Ingresso* e *Egresso*) devem ser desconsiderados, pois os dados de ambos foram recompilados no cubo *Graduação*. A Tabela 3 detalha os indicadores presentes nos quatro cubos de desempenho acadêmico que estão operacionais. Mais especificamente: o cubo *Graduação* (que se refere à população dos ingressantes) vincula dados sobre desempenho do aluno, faixa de períodos cursados e tempo de integralização do curso etc. O cubo *Turmas*

Tabela 3 – Dados do cubos de desempenho acadêmico

cubo	dados
graduação	população de ingressantes quantidades de: abandonos, afastamentos, conclusões, egressos, eliminações, trancamentos, transferências, ingressantes, ingressantes por vestibular, rematrículas, alunos, alunos aptos, alunos em turma. percentuais de: conclusão, abandono. média de CR acumulado por: aluno, aluno apto, aluno em turma, egresso tipo de cota, período do ingresso, dados pessoais, dados sobre ensino médio, opções de vestibular, turno, faixa de períodos (cursados e estada), forma de saída, estruturas acadêmicas, cursos, duração da integralização, ênfase, habilitação, modalidade, forma de entrada, tempo de integralização
turmas	população de alunos em turma quantidades de: alunos, turmas, vagas acréscimo UERJ, vagas ADI, vagas totais, vagas UERJ, vagas vestibular, alunos avaliação, alunos efetivos. tipo de cota, período do ingresso, dados pessoais, dados sobre o ensino médio, opções de vestibular, turno, período letivo, disciplinas, etapa inscrição, situação do aluno em disciplina, situação da inscrição, situação da turma, tipo de requisição de inscrição
planejamento	população de docentes quantidades de docentes, docentes graduados, horas contratadas, horas planejadas. percentuais de: docentes por categoria, titulação, hora contratada por categoria e por titulação, hora planejada por categoria e por titulação. área de conhecimento, atividade, categoria docente, docente, estrutura acadêmica, titulação docente, período letivo
vestibular	população de candidatos quantidades de: candidatos total, desistentes, matriculados, pontos por disciplina, vagas distribuídas, vagas ociosas. percentuais de: desistência, ociosidade, candidato. média da nota final, média de pontos por disciplina, média dos pontos discursivos, média dos pontos qualificação, nota final máxima, nota final mínima, relação candidato/vaga, coeficiente, desvio-padrão, média, KPI candidato/vaga, ano, dados pessoais, dados sobre ensino médio, conceitos, disciplinas vestibular, etapas classificação, exames, faixa nota final, isenção, opção (campus, carreira, semestre, turno), ponto total disciplina, região da prova, situação do candidato

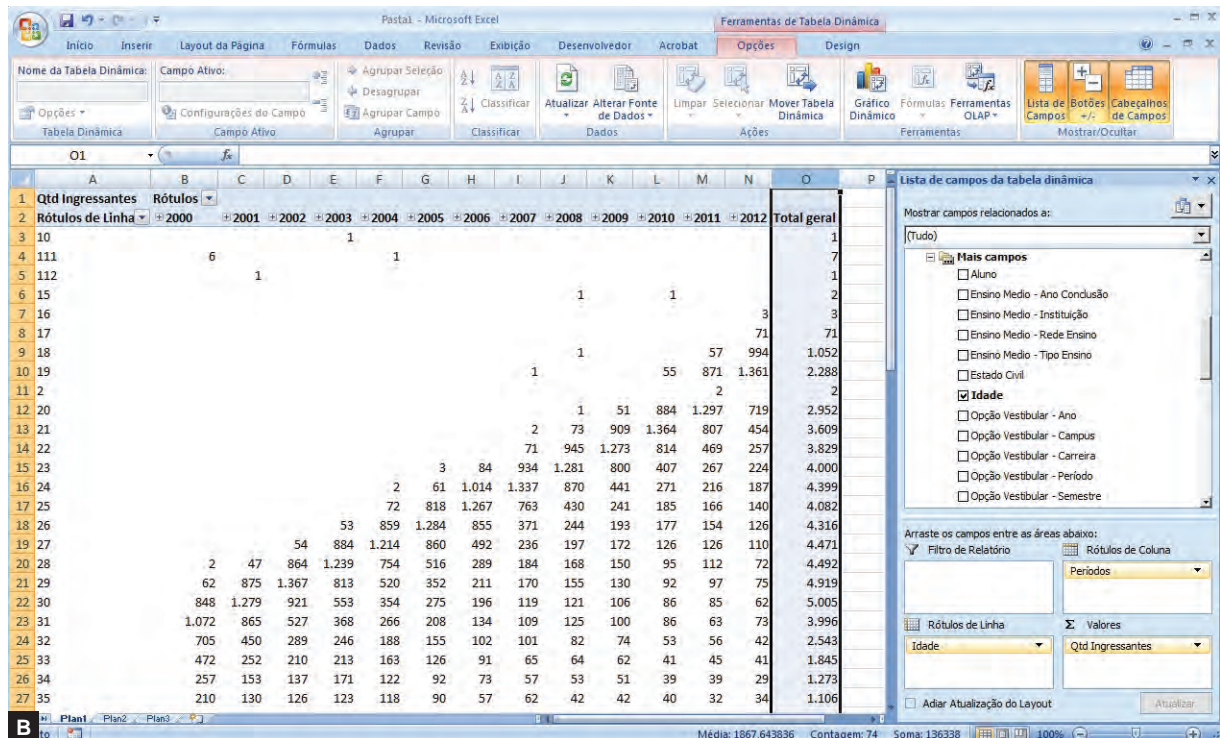
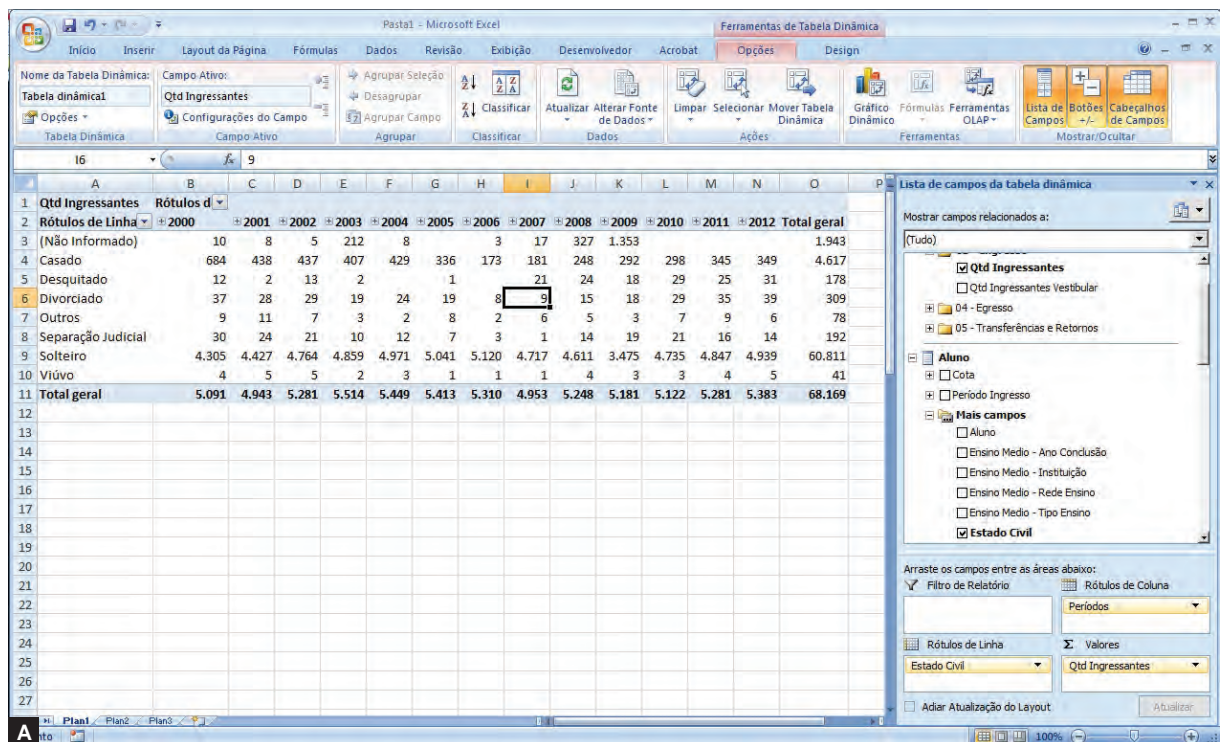
Fonte: autor (compilado a partir das tabelas dinâmicas disponibilizadas pelo BI UERJ).

(sobre a população de alunos em turma) apresenta a quantidade de turmas e de vagas, disciplinas oferecidas por áreas, situação do aluno em determinada disciplina (aprovado/cancelado/isento etc.), situação da turma (cancelada/ativa) etc. O cubo *Planejamento de atividades* (referente ao quadro docente) vincula dados como porcentagem de professores por tipo de titulação, quantidade de docentes, quantidade de horas trabalhadas, situação funcional, vinculação a determinada área do conhecimento etc. O cubo *Vestibular* (que reúne dados sobre os candidatos do vestibular) disponibiliza estatísticas como porcentagem de desistência, média da nota final, porcentagem de ociosidade das vagas oferecidas, isenção de matrícula, dados pessoais sobre os candidatos etc.

A análise dos dados armazenados em cubos é viabilizada por meio das mencionadas *tabelas dinâmicas* (*pivot tables*), ferramenta do software *Microsoft Excel*, adequadas à manipulação de grandes volumes de dados. Como quando “temos centenas de milhares de dados pontuais para rastrear.” [Nesses casos] “as tabelas dinâmicas possibilitam sumariá-los rapidamente em quase qualquer modo imaginável”, tratando-se assim de uma “poderosa ferramenta que pode ser usada para analisar os dados.” (WINSTON, 2007). Dentre seus recursos incluem-se alterações na estrutura de linhas e colunas das tabelas como a inversão, agrupamentos e sínteses, separação etc, que viabilizam pesquisas mais específicas.

A Figura 27 exibe um exemplo da interface da tabela dinâmica composta pelos dados do cubo *Graduação*. O menu à direita da tela divide-se em cinco áreas que organizam o acesso aos dados. A parte superior apresenta todas as categorias presentes no cubo, que acionadas, exibirão os dados correspondentes na forma de uma planilha na área principal do programa. Os itens *Rótulo de coluna* e *Rótulo de linha* recebem as categorias descritivas (as *dimensões* a serem analisadas, equivalentes aos *componentes* descritos por Bertin) que podem ser apresentados tanto no eixo horizontal, quanto no vertical (x ou y). O campo *Valores* refere-se às categorias de dados de natureza quantitativa (*componentes quantitativos*), tratando-se dos dados que preenchem as células da tabela após a escolha de uma ou mais categorias. O campo *Filtro de relatório* possibilita uma redução em um conjunto de dados selecionados de acordo com um critério de análise escolhido pelo usuário, como a seleção de determinado período letivo ou a filtragem de certo segmento de alunos. Na Figura 27A, a tela superior exemplifica uma consulta cruzando-se a *quantidade de ingressantes* por tipo de *estado civil*, nos *anos letivos* 2000 a 2012. A imagem 27B compara, para o mesmo período, a *quantidade de ingressantes* por *idade*. A efetividade da tabela dinâmica está na possibilidade de acesso a todos os dados contidos em um conjunto pré-estabelecido de modo a favorecer análises específicas. Entretanto, para consultas com mais de três dimensões, o sistema adota uma abordagem diferente da proposta por Bertin. A cada categoria inserida em *Rótulo de coluna* ou em *Rótulo de linha*, o sistema a internaliza como um subconjunto da anterior. Isso torna inviável consultas baseadas na organização tabular prevista na análise matricial, na qual os diversos *componentes* dos dados (dimensões) podem ser vistos independentemente por meio da justaposição vertical e agrupados por um ponto de comparação comum a todos eles. Portanto, para realizar tais pesquisas, é preciso consultar individualmente cada combinação de três componentes e, posteriormente, compilar todas essas pesquisas em uma planilha que agrupa todas elas.

Figura 27 – Interface do recurso de tabela dinâmica



Fonte: autor (captura de tela do software *Microsoft Excel*).

4.2 Síntese sobre a discussão em torno das cotas

Para contextualizar a análise dos dados em relação ao programa de cotas da UERJ é apresentada a seguir uma síntese dos tópicos principais que estão em jogo nessa questão. Como se sabe, as políticas de ações afirmativas relacionadas ao ingresso na universidade pública concentram intenso debate que, muitas das vezes, assume um viés marcadamente ideológico. A prerrogativa básica em que se baseia o programa de cotas é a inclusão de grupos sociais que historicamente estiveram à margem do ensino público de nível superior, tais como afro-descendentes, pobres, indígenas e deficientes físicos. UERJ, Unicamp e UNB estão dentre as universidades pioneiras na implementação de programas de cotas nos seus vestibulares, nos primeiros anos da última década. Mas foi a partir da sanção da Lei nº 12.711, em 2012, que instituiu a reserva de 50% das matrículas nas instituições federais de ensino “a alunos oriundos integralmente do ensino médio público, em cursos regulares ou da educação de jovens e adultos.”, que o debate se intensificou. Nesse momento, a contraposição entre os favoráveis e os contrários às cotas tende a correlacionar-se a posições conservadoras e progressistas, evidenciando os aspectos políticos envolvidos e que balizam os dois pólos ideológicos que delineiam esse debate.

O grupo contrário argumenta que as cotas ferem o princípio da meritocracia, “o alicerce que sustenta as modernas burocracias estatais”, segundo o sociólogo Demétrio Magnoli (2012). Segundo ele, ao sancionar a lei de cotas “O Estado brasileiro desembaraça-se do princípio do mérito alegando que se trata de critério “elitista”” (ibid.). Outro questionamento recorrente diz respeito à discriminação entendida como sendo inerente ao estabelecimento de cotas sociais. Esse argumento fundamentou a queixa do partido *Democratas* contra a constitucionalidade da lei de cotas, somando-se ainda à alegada impossibilidade de critérios objetivos para identificação de grupos sociais. Segundo o site noticioso *Deutsche Welle*, a advogada do DEM embasou-se em tais princípios para sustentar no Supremo Tribunal Federal que uma consequência provável da lei seria o risco do surgimento de culturas segregadas, em oposição à cultura nacional que seria intrinsecamente miscigenada. Outro questionamento do grupo contrário às cotas argumenta que o que se pode esperar com a composição de 50% das universidades federais por cotistas é a degradação no ensino dessas instituições. Por esse raciocínio, o biólogo Fernando Reinach acredita que “O mais provável é que esse aumento na heterogeneidade [devido ao ingresso de cotistas advindos de diversos tipos de escola] diminua a qualidade do ensino nas universidades

federais.”. Risco semelhante é levantado pelo sociólogo Simon Schwarzman, para quem “ou as universidades vão admitir essas pessoas hoje e daqui a um ano vão expulsá-las, por não conseguirem acompanhar os cursos, ou então vamos ter que baixar o nível para atender a essas pessoas”. Sendo assim, ele acredita que “as mais qualificadas [irão] embora, procurar outras instituições”. Na mesma linha, Magnoli profetiza um “impacto devastador nas universidades federais. Por motivos óbvios, elas estão condenadas a espelhar o nível médio das escolas públicas que fornecerão 50% de seus graduandos.”. Sendo assim, a solução aventada para a maior participação da sociedade nas universidades públicas depende de aprimoramentos dos ensinos fundamental e médio, tal como propõe Reinach, para quem “Precisamos não de cotas, mas de um ensino público melhor.”. E também Magnoli, que entende ser necessária uma “revolução no ensino público destinada a equalizar por cima a qualidade da educação oferecida aos jovens.”.

Entretanto, os favoráveis às cotas entendem que o tempo necessário para tal revolução pode ser desproporcional à urgência por incluir os grupos sociais em questão. Nesse sentido, o cientista político Luiz Felipe de Alencastro, em texto intitulado “Por que o Brasil precisa das cotas”, relembra os diversos fatores históricos que levaram à condição marginal do negro na sociedade. O autor aponta ainda estudos que provam que a desigualdade racial no Brasil tem características estruturais que se mantêm a despeito do progresso econômico do país. Em tal contexto, Alencastro conclui que os programas de cotas incluem-se em uma perspectiva ampla de integração social que depende “de inscrever a discussão sobre a política afirmativa no aperfeiçoamento da democracia, no vir a ser da nação.”. Esse entendimento parece ser corroborado pelo relator do referido processo movido pelo DEM, ministro Ricardo Lewandowski. Ao decidir pela constitucionalidade das cotas, ele afirma que elas são “uma forma de compensar essa discriminação culturalmente arraigada, não raro praticada de forma inconsciente, e à sombra de um Estado complacente.” No que se refere à queda na qualidade do ensino, os favoráveis às cotas sustentam que não se trata de um resultado necessário e permanente que decorre da política de cotas. Por esse entendimento, o coordenador do Núcleo de Estudos Afro-brasileiros da UNB, Nelson Inocêncio considera que “Existe um pensamento conservador muito forte nas universidades, de que essas políticas vão comprometer a qualidade do ensino. É um discurso perverso, conservador, mas que permanece nas universidades e é compartilhada por alunos e professores”. Ele acredita que mesmo havendo uma defasagem em um primeiro momento, a questão do acesso à universidade pública não pode ser balizada somente pelo discurso do mérito. Reforçando essa interpretação, o presidente da Associação Nacional dos Dirigentes de

Instituições Federais de Ensino Superior, Carlos Maneschy, acredita na possibilidade de cotas sem redução da qualidade no ensino. Segundo ele, “levantamentos mostram que, apesar das dificuldades iniciais apresentadas pelos cotistas, os índices de desempenho e de evasão entre alunos vindos de escolas públicas e de particulares são os mesmos.”

Fora da arena ideológica, os estudos abrangentes que lidam de modo objetivo com a questão ainda estão principiando. Os pesquisadores da UFF, Fábio Waltenberg e Márcia de Carvalho analisaram dados sobre o desempenho de cotistas no Enade de 2008. Nesse estudo, eles encontraram “diferenças razoáveis” que “não são catastróficas como previam alguns críticos das ações afirmativas, mas é importante registrar que existe uma diferença para não tapar o sol com a peneira”. Ainda assim, consideram que “o desnível atual é um preço baixo a se pagar pela maior inclusão.”. Além disso, a pesquisa encaminhada pelo economista Alvaro Mendes Junior sobre os dados da UERJ descobriu que o nível de evasão entre os cotistas tende a ser menor, mas “os alunos cotistas possuem um pior desempenho em termos de coeficientes [de rendimento].” (MENDES JUNIOR, 2013, p. 88). Por outro lado, os dados mostram que apesar disso, “os cotistas estão se graduando a taxas mais elevadas.”. Nesse sentido, o autor concluiu que “Mesmo diante de um alto diferencial de performance na entrada do vestibular, os cotistas conseguiram se graduar a taxas mais elevadas. Este se revelou um ponto importante para a justificativa de manutenção do programa.” (ibid. p.89).

Assim, por esse breve panorama, que busca apenas situar os tópicos principais da discussão, pode-se entender que a pedra de toque desse debate é a questão da qualidade do ensino. Apesar da discussão espinhosa que seria definir “qualidade” em relação ao processo de educação e aprendizado, os estudos mais objetivos sobre o programa de cotas costumam traduzir o aspecto da qualidade pelos indicadores de desempenho tradicionais como notas e índices de evasão. Sendo assim, esta dissertação também se vale desses critérios, mas ao considerá-los e evidenciá-los visualmente de modo mais nuançado e abrangente procura indicar um outro caminho de análise, talvez potencialmente mais apto a lidar com a complexidade de tais questões.

4.3 Teoria matricial

(condição exemplificada na Figura 11, p.70). Nesse estágio, pode-se proceder aos tratamentos de dados, entre o quais, o tratamento gráfico, pela permutação nos componentes de uma matriz ordenável. Isso resulta em uma singular visualização dos dados no que concerne às informações que apresenta. A partir daqui, são detalhadas as técnicas da *Teoria matricial* que embasaram a formulação do primeiro diagrama. A seguir, será detalhado de que modo os outros dois foram derivados do primeiro, sugerindo como, por meio dos recursos interativos, o tipo de análise proposta por Bertin pode ser transposta para o domínio digital.

4.3.1 Registro dos dados discentes

Para um entendimento dos dados disponibilizados pelo BI da UERJ a partir da coleta dos dados (considerando-se assim uma perspectiva *bottom-up*), eles podem ser sintetizados em uma lista, como a da Tabela 4. Esse registro apresenta os dados relativos a uma aluna da UERJ (identificada com o número 100271) desde a inscrição no vestibular até seu egresso após concluída a graduação em Pedagogia. Trata-se de um histórico das atividades e desempenho da aluna por meio do qual se sabe, por exemplo, que ela optou no vestibular pela carreira de Administração no ano letivo de 2005. Porém, ela cursou de fato a graduação em Pedagogia, no turno da noite, concluindo após um trancamento e um abandono. Históricos como esse são registrados para todos os 84.903 alunos incluídos no sistema, embora muitos tragam informações incompletas.

4.3.2 Formulário estatístico

A partir dos registros sobre os alunos, sua transposição para uma tabela requer uma padronização que depende de algumas operações realizadas de modo implícito ou explícito e que resultam em um *formulário estatístico* (BERTIN, 1977, p.242). Trata-se de um registro para cada indivíduo que aceita apenas valores numéricos ou marcações binárias (característica presente ou ausente), conforme a organização do dado (*quantitativa* ou *ordenada*, no primeiro caso; e *seletiva*

Tabela 4 – Registro de dados por aluno

identificação aluno	100271
situação	egresso / concluinte
média CR acumulado aluno	8,96
média CR acum. aluno apto	9,04
média CR acum. aluno em turma	9,11
média CR acumulado egresso	34,13
abandono	sim (1)
conclusão	sim (1)
egressão	sim (2) abandono + conclusão
eliminação	não
afastamento	não
rematrícula	sim (1)
trancamento	sim (1)
transferência de curso	não
cota	não declarado/não cota
período ingresso	2005.1
ensino médio: ano conclusão	2002
ensino médio: instituição	Colégio Ateneu do Rio de Janeiro
ensino médio: rede ensino	particular
ensino médio: tipo ensino	regular
estado civil	solteiro
idade	39
opção vestibular: ano	2005
opção vestibular: campus	UERJ / Rio
opção vestibular: carreira	Administração / Rio
opção vestibular: período	2005.1
opção vestibular: semestre	1º
opção vestibular: turno	manhã
sexo	feminino
estrutura acadêmica	Centro de Ciências Sociais (CCS)
curso	Pedagogia - Séries iniciais
curso: duração máxima integralização	14
curso: duração mínima integralização	8
ênfase	NI
habilitação	licenciatura plena nas séries iniciais do ensino fundamental e educ infantil
modalidade	NI
forma de entrada	vestibular
forma de saída	abandono + conclusão final
período letivo	2005-2010
períodos cursados	11 (abandono) 13 (conclusão)
períodos de estada	12 (abandono) 13 (conclusão)
tempo de integralização	3 após min até máx
turno	noite

Fonte: autor (a partir de dados do BI - UERJ).

no segundo). A tabela 5 exemplifica um formulário segundo esses critérios. O componente *coeficiente de rendimento* (CR) tem característica quantitativa, pois pode-se declarar que tal aluno obteve o dobro, o triplo, a metade do CR de outro estudante, tal como acontece ao componente *idade*. Já itens que registram datas, como *período de ingresso* ou *período letivo* têm significado ordenado, uma vez que se pode sequenciá-los do maior para o menor (mas teria pouco sentido uma identificação proporcional, declarando que tal data é a metade ou a terça parte de outro período). Por fim, componentes *seletivos* são aqueles de característica qualitativa, em que o componente divide-se em subcategorias ou classes mutuamente excludentes. Daí a marcação (×) na subcategoria correspondente a determinado dado individual.

Observa-se que tais conceitos podem ser modificados, e aqueles muito extensos

Tabela 5 – Formulário estatístico: aluna 10027, ano 2005

número aluno	10027	sexo	masculino	
			feminino	x
média CR acumulado aluno	8,96	estrutura acadêmica	Centro de Educação e Humanidades	
média CR acum. aluno apto	9,04		Centro de Tecnologia e Ciências	
média CR acum. aluno em turma	9,11		Centro Biomédico	
média CR acumulado egresso	34,13		Centro de Ciências Sociais	x
rematricula	1	curso	Administração	
trancamento	1		...	
transferência de curso (entrada)	0		Pedagogia - séries iniciais	x
transferência de curso (saída)	x		...	
cota	não declarado / não cota negro negro (SADE - 2003) port deficiência / indígena rede pública rede pública (SADE - 2003)	habilitação	Turismo	
			Administração de empresas	
período ingresso	2005.1		Artes plásticas	
ensino médio: ano conclusão	2002		...	
ensino médio: instituição	Colégio Ateneu do Rio de Janeiro	ênfase	Licenciatura plena nas séries iniciais...	x
ensino médio: rede de ensino	estadual federal municipal particular		Mag Mat Ped do 2ºGrau...	
			Construção civil	NI
ensino médio: tipo de ensino	curso regular curso supletivo exame supletivo		Elettricidade industrial	-
			...	-
estado civil	casado desquitado divorciado outros separação judicial solteiro viúvo	modalidade	Transportes	-
			Têxtil	-
idade	39		bacharelado	NI
opção vestibular : ano	2005		informática	-
opção vestibular : campus	UERJ / BFI UERJ / IPR UERJ / Res UERJ / Rio UERJ / SGo UERJ / Teresópolis		licenciatura	-
			médica	-
opção vestibular : carreira	Administração ... Direito ... Turismo	forma de entrada	aproveitamento de estudos convênio transferência ex-officio transferência externa vestibular	x
opção vestibular : período	2007.1	forma de saída	abandono afastamento conclusão outras eliminações trancamento transferência	x x
opção vestibular : semestre	1º			
opção vestibular : turno	integral manhã tarde noite manhã e tarde tarde e noite	período letivo		2005
		períodos cursados	abandono conclusão	11 13
		períodos estada	abandono conclusão	11 13
		tempo de integralização	antes do mín no mín 1 a 2 após mín até max 3 após mín até max no máx após máx	x
		turno	integral manhã tarde noite manhã e tarde tarde e noite	x

Fonte: autor (a partir de dados do BI - UERJ).

podem ser agrupados em classes, tornando-se mais significativos ao condensar uma série inconvenientemente longa. Por exemplo, componente *tempo de integralização* (cuja condição original é quantitativa “10 períodos, 15 períodos etc.) foi agregado em em seis categorias ordenadas.

4.3.3 Tabela de alocação (*le tableau de ventilation*)

Trata-se do primeiro de três documentos propostos por Bertin para a organização dos dados em uma tabela única, que será a base da visualização. Seu objetivo é estabelecer qual será o ponto de comparação entre os dados. Na primeira coluna registam-se os *componentes*; na segunda seu *comprimento*; e no corpo da tabela, a característica e as relações entre os componentes são representadas por símbolos e notações de cruzamento:

- × componentes seletivos;
- Q componentes quantitativos;
- O componentes ordenados;
- \ dados não informados;
- | relação entre duas componentes, permitindo justaposição vertical;
- componente comum.

A ligadura vertical [|] representa a possibilidade de uma tabela xy entre dois componentes. A tabela de alocação do cubo de graduação na Figura 29, indica a possibilidade de se desenhar uma tabela em que o componente *alunos* fique em x, e o componente *CR* esteja posicionado em y. Desse modo, nas células dessa tabela ocorrem os valores [Q] do CR. Outra possibilidade seria, ainda com os *alunos* em x, colocar-se o componente *ingressantes do ensino particular* em y, e as células conteriam a quantidade de ingressantes desse tipo de ensino.

A ligadura horizontal [—] representa um componente comum a duas tabelas correspondentes. Quando tal ligadura percorre diversas colunas identificam-se aqueles componentes que são comuns aos demais, podendo ser colocados no eixo x de uma tabela, com os demais no eixo y, justapostos verticalmente resultando na configuração característica de um estudo homogêneo (ver Figura 11, p.70). Na Figura 29, isso acontece ao componente *alunos* e seus desdobramentos (*alunos em turma*, *ingressantes* etc.), visto que *quase* todos os dados da tabela concernem a eles. Por outro lado, percebe-se uma segunda formação horizontal na parte inferior direita da tabela. Ela evidencia que os componentes *duração máxima da integralização* e *duração mínima da integralização* não se relacionam aos *alunos* e sim ao componente *cursos*, mostrando-se contrários à lógica que constitui o cubo *graduação*, o que sugere que esses indicadores sejam desconsiderados na análise.

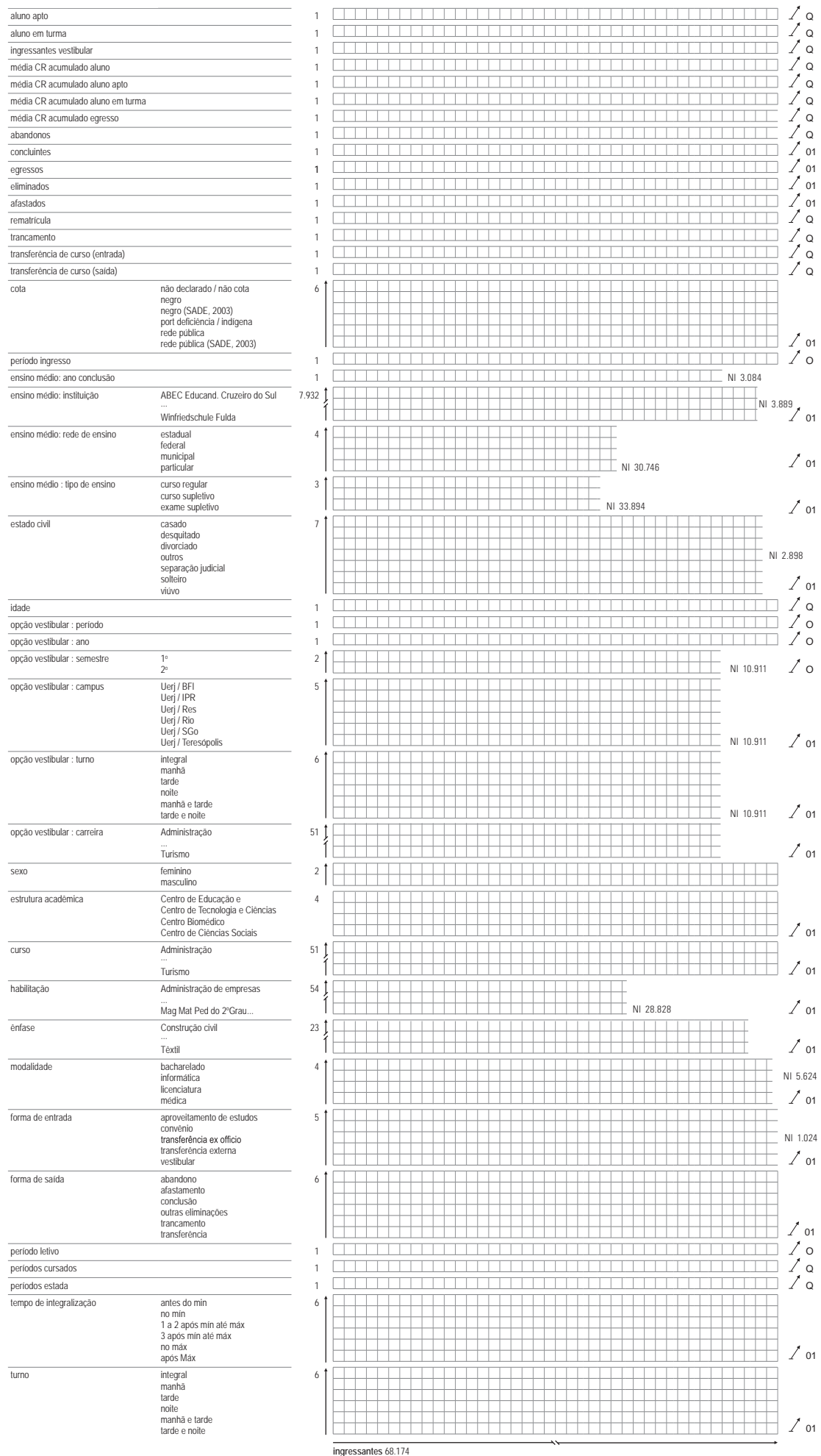
4.3.4 Esquema de homogeneidade (*le schéma d'homogénéité*)

Diagrama utilizado para dimensionar a tabela e definir a estratégia para seu tratamento e visualização. A partir da tabela de alocação, observam-se os eixos horizontais que identificam os componentes que são comuns aos demais, como *alunos*, *alunos aptos*, *alunos em turma*, *ingressantes* e *ingressantes vestibular*. Todos referem-se aos alunos, mas agrupando-os de formas particulares de modo que, a princípio, qualquer um deles poderia ser usado como a base do esquema de homogeneidade. Empregou-se então o componente *ingressantes* que se refere aos alunos que entraram na UERJ por meio de vestibular, transferência *ex-officio*, transferência externa, aproveitamento de estudos, convênio, cortesia e didática especial, perfazendo 68.174 indivíduos cadastrados. No esquema de homogeneidade da Figura 30, o algarismo à esquerda das setas verticais representam o comprimento de cada *componente* (correspondendo ao número de células no sentido vertical da tabela, no caso daqueles componentes com pequeno *comprimento*. O eixo z, definido pelos dados nas células da tabela é indicado, conforme sua natureza, pela seta oblíqua na lateral direita do diagrama. [Q] define os componentes quantitativos, [O] os ordenados e [01] os seletivos. Os vazios no corpo da tabela representam proporcionalmente os alunos para os quais não há informação de determinado tipo, o que é indicado pelo símbolo [NI] (não informado) antecedendo a quantidade de ausências.

Esta organização possibilita o cálculo da tabela hipotética que conteria os dados do cubo graduação. Tal dimensionamento acontece por meio da multiplicação do *número de alunos* pelo *comprimento de cada componente* para os quais há dados disponíveis. Assim, a soma das “áreas” das 44 tabelas resulta nas dimensões da tabela final. Neste caso, teríamos uma tabela com 560.414.204 células, dimensões impraticáveis para análise humana e para a visualização. Torna-se necessário, então, estabelecer critérios para a redução deste conjunto de dados baseados nos objetivos específicos de cada pesquisa. Para isso, a redução pode ocorrer tanto pela supressão de características não prioritárias, como pela agregação da população de objetos (os ingressantes) segundo um outro ponto de comparação. As ligaduras verticais da tabela de alocação indicam os componentes que podem ser usados para agregar os alunos. Como ressalta Bertin “Todo componente colocado em y no esquema [de homogeneidade] pode ser usado para reagrupar os objetos em x.” (1977, p. 243).

Nesse sentido, o ponto determinante para a redução dos dados são as perguntas para

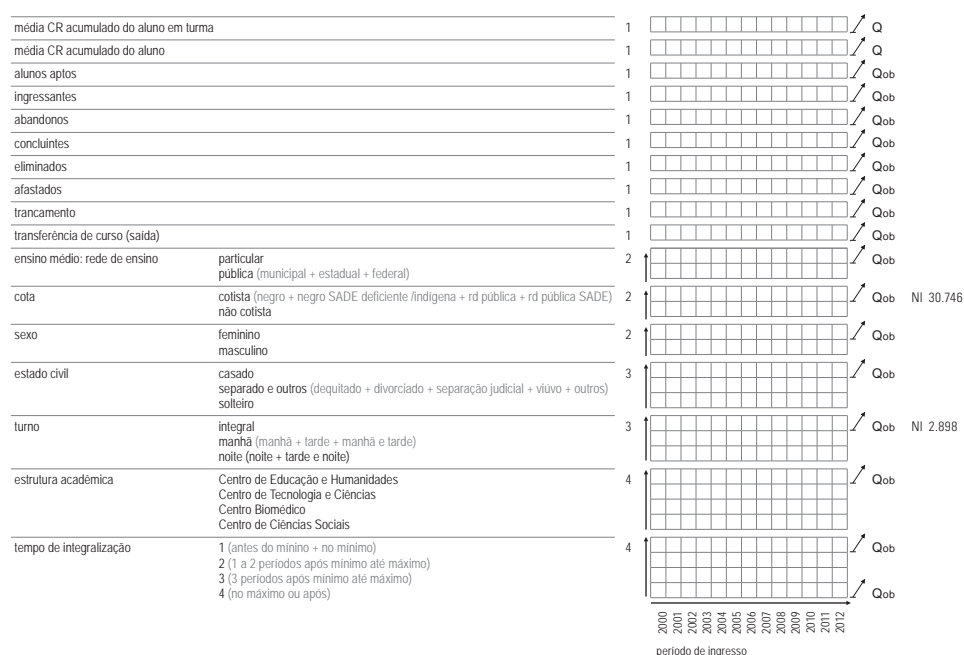
Figura 30 – Esquema de homogeneidade (cubo *graduação*)



as quais é preciso dar resposta. Portanto, considerando-se as posições em torno da questão das políticas de ação afirmativa, parece interessante como primeira abordagem, uma análise do desenvolvimento das características diretamente relacionadas ao universo do programa de cotas em dois momentos: antes e após sua implantação. Essa abordagem pode ser sintetizada pela abrangente pergunta: “Como a universidade vem se comportando após a implantação das cotas?”, questionamento que sugere o uso do componente *períodos de ingresso* para agregar os dados sobre os alunos, possibilitando análises evolutivas. A figura 31 apresenta um novo esquema de homogeneidade em que os anos registrados pelo sistema (2000–2012) atuam como ponto de comparação entre as categorias. Neste caso, o eixo z (células) é preenchido pela quantidade de objetos estatísticos [Qob] (a população de ingressantes) em cada ano letivo (eixo x), por categoria (componentes posicionados em y).

Além disso, a redução vertical do esquema de homogeneidade acontece em razão de questões pragmáticas e de critérios específicos ao conjunto de dados em questão. A questão prática prioritária é que os dados possam ser visualizados empregando-se os recursos disponíveis. Em razão da inexistência de *softwares* que processem dados pela metodologia de Bertin, as análises desta dissertação são limitadas pela possibilidade de realizá-las combinando-se, basicamente, procedimentos de programação e de desenho vetorial.

Figura 31 – Esquema de homogeneidade simplificado



Fonte: autor

foram mantidas apenas as características estritamente pertinentes ao estudo. Nesse sentido, desconsiderou-se todos os itens relativos às *opções do vestibular*, pois no que se refere às cotas, tendem a repetir dados sobre a graduação que já constam em outras categorias. Os tópicos *períodos cursados* e *períodos de estada* foram suprimidos em favor do critério mais abrangente *tempo de integralização*. Já o componente *idade* não foi usado em razão de uma suspeita de erro no sistema, que registra poucos alunos na faixa de idade que, a princípio, seria a mais comum (entre 17 e 25 anos), além de idades improváveis como ingressantes com 111 e 10 anos, como pode ser constatado na imagem da tabela dinâmica na Figura 27B, p. 98.

Outro recurso para a redução do esquema de homogeneidade foi a agregação de alguns componentes resultando em classes mais significativas. Assim, as cinco modalidades de cota foram agregadas sob o item *cotista*; os sete *estados civis* foram agrupados em três; as três redes de ensino público do Ensino Médio foram agregadas como *rede pública* etc. Esse procedimento deve ser considerado tendo em vista a possibilidade de que tais classes sejam novamente expandidas por meio dos recursos interativos, revelando assim todos os seus elementos constituintes, caso necessário. De modo análogo *habilitação*, *ênfase* e *modalidade* (referentes aos cursos de graduação) foram suprimidos, ao passo que o componente *curso* foi agregado, considerando-se que eles são parte da categoria mais abrangente *estrutura acadêmica* (que refere-se aos quatro centros universitários) e que, em uma interface digital, poderiam ser acessados através dela. Desse modo, chega-se a um esquema aceitável com 390 células, bastante abaixo do limite de 10.000 células apontado por Bertin para o tratamento gráfico direto (ibid., p.251), mas que viabiliza o desenvolvimento das visualizações deste trabalho. Por fim, esse esquema de homogeneidade é a base para a etapa da definição dos cálculos que antecede a tabela final.

4.3.5 Tabela de pertinência (*le tableau de pertinence*)

Terceiro documento da análise matricial, registra os cálculos preparatórios para a tabela base da visualização (Figura 32). Tais operações são necessárias quando, como acontece no cubo de graduação, têm-se apenas quantidades absolutas. Da perspectiva do projeto do banco de dados, isso é adequado pois trata-se do estado mais autônomo, mais “puro” dos dados, facultando

Figura 32 – Tabela de pertinência

1 hipóteses	2 operações principais	3 cálculos adicionais	
	1 Q média CR acumulado do aluno em turma		A dados originais
	2 Q média CR acumulado do aluno		
	3 Q alunos aptos		
	4 Q abandonos		
	5 Q concluintes		
	6 Q eliminações		
	7 Q afastamentos		
	8 Q trancamentos		
	9 Q transferências		
	10.1 Q ensino médio: rede particular		
	10.2 Q ensino médio: rede pública		
	11.1 Q cola: cotista		
	11.2 Q cola: não cotista		
	12.1 Q sexo: feminino		
	12.2 Q sexo: masculino		
	13.1 Q estado civil: casado		
	13.2 Q estado civil: separado e outros		
	13.3 Q estado civil: solteiro		
	14.1 Q turno: integral		
	14.2 Q turno: manhã		
	14.3 Q turno: noite		
	15.1 Q estrutura acadêmica CEH		B dados derivados
	15.2 Q estrutura acadêmica CTC		
	15.3 Q estrutura acadêmica CBIO		
	15.4 Q estrutura acadêmica CCS		
	16.1 Q integralização (1) no mínimo ou antes		
	16.2 Q integralização (2) 1-2 períodos após mínimo até máximo		
	16.3 Q integralização (3) 3 períodos após mínimo até máximo		
	16.4 Q integralização (4) após máximo		
	Q total de ingressantes		
	Q total de alunos		
	Q concluintes		
	1 Q INGRESSANTES COEFICIENTE DE RENDIMENTO DO ALUNO EM TURMA		
	2 Q INGRESSANTES COEFICIENTE DE RENDIMENTO DO ALUNO		
	3 % INGRESSANTES APTOS		
	4 % INGRESSANTES ABANDONOS		
	5 % INGRESSANTES CONCLUINTES		
	6 % INGRESSANTES ELIMINAÇÃO		
	7 % INGRESSANTES AFASTAMENTO		
	8 % INGRESSANTES TRANCAMENTO		
	9 % INGRESSANTES TRANSFERÊNCIA		
	10.1 % INGRESSANTES ORIGEM: REDE PARTICULAR		
	10.2 % INGRESSANTES ORIGEM: REDE PÚBLICA		
	11.1 % INGRESSANTES COTA: COTISTA		
	11.2 % INGRESSANTES COTA: NÃO COTISTA		
	12.1 % INGRESSANTES SEXO: FEMININO		
	12.2 % INGRESSANTES SEXO: MASCULINO		
	13.1 % INGRESSANTES ESTADO CIVIL: CASADO		
	13.2 % INGRESSANTES ESTADO CIVIL: SEPARADO E OUTROS		
	13.3 % INGRESSANTES ESTADO CIVIL: SOLTEIRO		
	14.1 % INGRESSANTES TURNO: INTEGRAL		
	14.2 % INGRESSANTES TURNO: MANHÃ		
	14.3 % INGRESSANTES TURNO: NOITE		
	15.1 % INGRESSANTES ESTRUTURA ACADÊMICA CEH		
	15.2 % INGRESSANTES ESTRUTURA ACADÊMICA CTC		
	15.3 % INGRESSANTES ESTRUTURA ACADÊMICA CBIO		
	15.4 % INGRESSANTES ESTRUTURA ACADÊMICA CCS		
	16.1 % INGRESSANTES INTEGRALIZAÇÃO: (1) NO MÍNIMO OU ANTES		
	16.2 % INGRESSANTES INTEGRALIZAÇÃO: (2) 1-2 PERÍODOS APÓS MÍNIMO ATÉ MÁXIMO		
	16.3 % INGRESSANTES INTEGRALIZAÇÃO: (3) 3 PERÍODOS APÓS MÍNIMO ATÉ MÁXIMO		
	16.4 % INGRESSANTES INTEGRALIZAÇÃO: (4) APÓS MÁXIMO		
	Q INGRESSANTES REPROVAÇÃO		C dados adicionais
	Q INGRESSANTES INSCRIÇÃO EM DISCIPLINAS		
	% INGRESSANTES REPROVAÇÃO		
	Q COEFICIENTE DE RENDIMENTO AL. EM TURMA COTISTA ESTRUTURA ACADÊMICA CEH		
	Q COEFICIENTE DE RENDIMENTO AL. EM TURMA COTISTA ESTRUTURA ACADÊMICA CTC		
	Q COEFICIENTE DE RENDIMENTO AL. EM TURMA COTISTA ESTRUTURA ACADÊMICA CBIO		
	Q COEFICIENTE DE RENDIMENTO AL. EM TURMA COTISTA ESTRUTURA ACADÊMICA CCS		
	Q VESTIBULANDOS ORIGEM: REDE PARTICULAR		
	Q VESTIBULANDOS ORIGEM: REDE PÚBLICA		
	Q VESTIBULANDOS		
	% VESTIBULANDOS ORIGEM: REDE PARTICULAR		
	% VESTIBULANDOS ORIGEM: REDE PÚBLICA		
	Q VAGAS ESTRUTURA ACADÊMICA CEH		
	Q VAGAS ESTRUTURA ACADÊMICA CTC		
	Q VAGAS ESTRUTURA ACADÊMICA CBIO		
	Q VAGAS ESTRUTURA ACADÊMICA CCS		
	Q INGRESSANTES COTISTAS ESTRUTURA ACADÊMICA CEH COTISTAS		
	Q INGRESSANTES COTISTAS ESTRUTURA ACADÊMICA CTC COTISTAS		
	Q INGRESSANTES ESTRUTURA ACADÊMICA CBIO COTISTAS		
	Q INGRESSANTES ESTRUTURA ACADÊMICA CCS COTISTAS		
	% OCUPAÇÃO ESTRUTURA ACADÊMICA CEH COTISTAS		
	% OCUPAÇÃO ESTRUTURA ACADÊMICA CTC COTISTAS		
	% OCUPAÇÃO ESTRUTURA ACADÊMICA CBIO COTISTAS		
	% OCUPAÇÃO ESTRUTURA ACADÊMICA CCS COTISTAS		
	Q INGRESSANTES CONCLUINTES COTISTAS		
	% INGRESSANTES CONCLUINTES COTISTAS		
1 2 3 4 5 6 H		C 1 3 4 5 6	

ao usuário analisá-los e transformá-los conforme suas necessidades. Mas para a comunicação informativa, quantidades absolutas têm pouca significação, pois tendem a não ser facilmente comparáveis. Daí, a opção por estabelecer relações pertinentes entre os componentes. Para tanto, a tabela de pertinência registra os cálculos necessários. Ela refere-se a um ano específico do conjunto determinado no esquema de homogeneidade considerando que os mesmos cálculos são válidos para todos os anos. Além disso, outro objetivo dessa tabela é registrar as hipóteses sobre o tema pesquisado e verificar se os dados necessários para avaliá-las estão disponíveis ou precisam ser levantados.

Consideradas essas funções, vê-se no segmento superior [A] a lista dos componentes definidos no esquema de homogeneidade assinalados com [Q], indicando sua condição como quantidade autônoma. O segmento intermediário [B] registra os dados derivados das operações aplicadas em [A], indicando a nova unidade de medida dos componentes (em sua maioria [%], representando sua condição percentual). O setor [C] refere-se aos dados adicionais, que não constavam do primeiro segmento, sendo preciso levá-los para verificar algumas das hipóteses formuladas. Na coluna à esquerda da identificação dos componentes [1], assinalam-se as hipóteses e na da extrema direita [3] são registrados as operações necessárias para avaliá-las.

Assim, para indicar as transformações, os dados em [A] (brutos) são referenciados por uma linha que se dirige até sua forma final em [B] (derivados). A operação matemática que originou as porcentagens é indicada pelo sinal de divisão [/] alinhado ao componente pelo qual os dados originais foram divididos; no presente caso, a maior parte das características foi dividida pelo total de geral de ingressantes. Como exemplo, a quantidade de alunos aptos [Q alunos aptos] é dividida pelo total de ingressantes [Q total de ingressantes] para encontrarmos o percentual de alunos aptos em relação ao total de alunos [% INGRESSANTES] [APTOS]. De modo análogo opera-se com os demais componentes, excetuando-se os *coeficientes de rendimento*, *vagas* e *quantidade de ingressantes* que permanecem em sua condição absoluta, pois não há relações significantes a estabelecer para eles no diagrama 1.

Tendo-se estabelecido a forma final dos dados, podem-se demarcar as hipóteses em relação ao conjunto de dados, delineando assim a gama de respostas que será procurada na visualização. Exemplificando esse processo, são aventadas seis hipóteses (identificadas pelo número na parte inferior esquerda da tabela de pertinência).

1 *É verdade que quanto menor o coeficiente de rendimento, maior a incidência de integralizações longas e maior a taxa de reprovações?* Para anotar essa hipótese assinalam-se o

sinal [–] (quanto menor) ao lado do *CR do aluno em turma* e o sinal [+] (maior incidência) ao lado das *integralizações 3 e 4*. Assim, falta localizar os índices de *reprovação*, que não estavam no conjunto inicial de dados. Na coluna da direita, registram-se os cálculos necessários. Os componentes que já se encontram disponíveis para responder à pergunta colocada são assinalados pelo círculo branco; aqueles que precisam ser levantados são marcados com o círculo preto. Para estabelecer a taxa de reprovações, é necessário identificar o total de *ingressantes reprovados* e dividi-lo pelo total de *inscrições em disciplinas*, operações assinaladas em [3], que resultam no primeiro grupo de dados adicionais em [C].

2 *Os coeficientes de rendimento dos cotistas é menor nas quatro estruturas acadêmicas?*

Os itens são assinalados com o sinal respectivo na coluna 1. É preciso estabelecer os CRs dos cotistas por cada estrutura acadêmica, mas para isso não há cálculos adicionais a fazer. Basta consultar o sistema para obter esses indicadores.

3 *A flutuação no número de ingressantes é equivalente nas quatro estruturas acadêmicas, (de modo que, por exemplo, um aumento no total geral de ingressantes significa um aumento correspondente nos ingressantes do CTC)?* Para identificar essa pergunta, emprega-se o sinal [=] ao lado das estruturas acadêmicas e da quantidade de ingressantes. Nesse caso, não há novos indicadores a estabelecer, nem são necessários cálculos adicionais, pois todos os dados já estão disponíveis.

4 *Há relação direta entre os vestibulandos e os ingressantes vindos de uma mesma rede de ensino (de modo que, por exemplo, quanto menor o número de vestibulandos do ensino público, menor a quantidade de ingressantes dessa rede)?* É preciso levantar os dados sobre os *vestibulandos de cada rede*, e ainda sobre o total de *vestibulandos*, de maneira que se possa dividir os primeiros pelos segundos, encontrando o percentual de vestibulandos por rede. Novamente, trata-se de dados que não estão presentes no cubo *graduação*, sendo necessário buscá-los no cubo *vestibular*.

5 *É verdade que no Centro de Tecnologia e Ciências o percentual de ocupação dos cursos pelos ingressantes cotistas é maior do que nas outras estruturas acadêmicas?* Para definir a taxa de ocupação é preciso descobrir o número de vagas disponibilizadas por estrutura acadêmica e então dividir o total de cotistas de cada estrutura pela quantidade de vagas oferecidas.

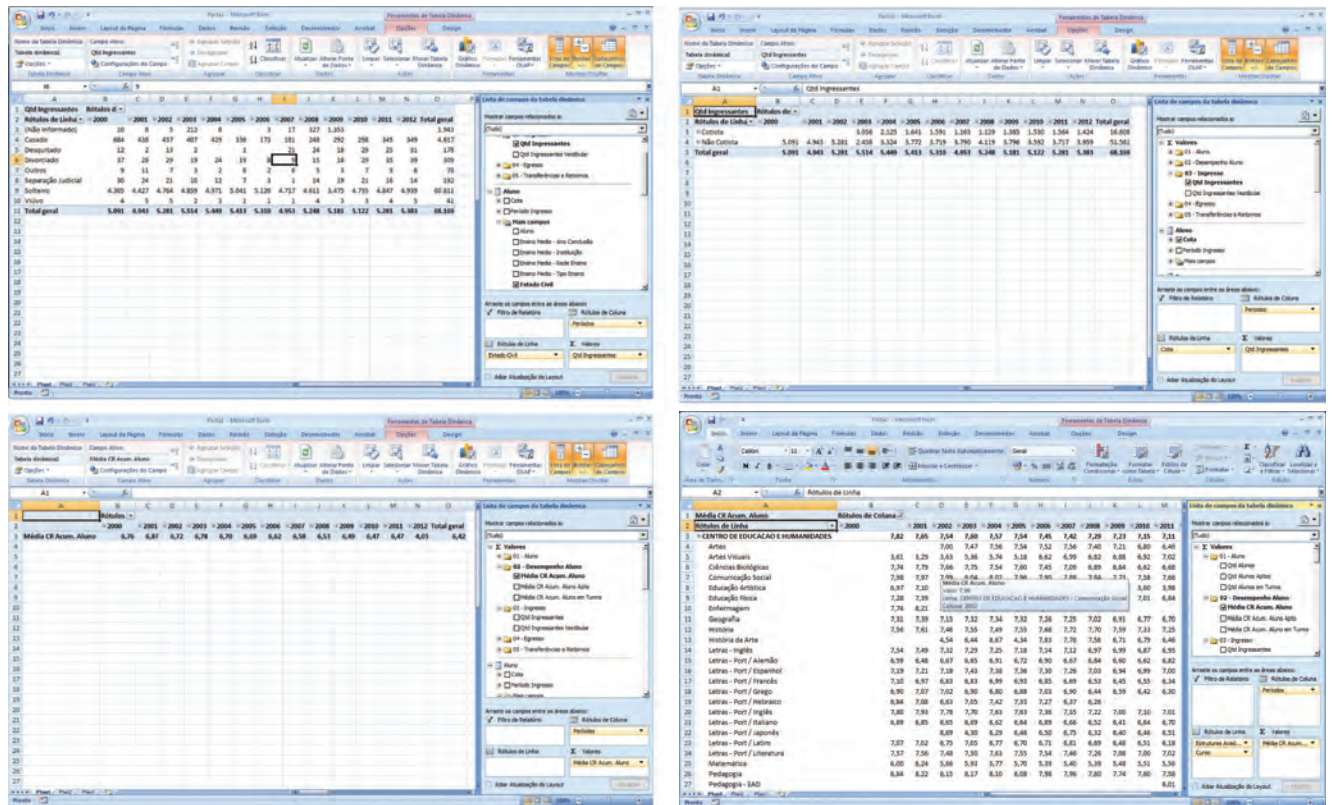
6 *É verdade que o percentual de concluintes cotistas é maior do que de não cotistas?* É preciso encontrar a quantidade de concluintes cotistas e não cotistas para que esses indicadores possam ser divididos pelo total de concluintes, revelando o percentual de concluintes para cada grupo.

Além dessas seis questões, outras certamente poderiam ser colocadas, mas levantar hipóteses é a seara específica dos estudiosos das políticas afirmativas. No contexto da presente dissertação, as questões levantadas procuram esboçar caminhos de pesquisa possíveis que serão investigados por meio dos gráficos. Assim, tais hipóteses conformam um enquadramento básico em relação a tópicos como ocupação das vagas, relação entre vestibulandos e ingressantes, diferenças entre cotistas e não cotistas etc. Além de exemplificar situações de naturezas distintas, as hipóteses em questão permitem desenvolvimentos no momento da análise dos dashboards. Isso acontece, por exemplo, nas perguntas sobre as estruturas acadêmicas que serão desdobradas para os cursos que integram tais estruturas. Além disso, comparações como a elaborada na hipótese seis, em relação aos concluintes cotistas, serão expandidas para diversas outras características, como abandonos, eliminação, trancamentos etc. resultando em um panorama mais detalhado dos alunos da instituição.

4.3.6 Tabela de dados

A partir deste ponto, podem-se consultar os indicadores que foram definidos na tabela de pertinência. Desse modo, com base nas tabelas dinâmicas do sistema de BI, são pesquisadas individualmente cada combinação entre duas dimensões qualitativas (eixos horizontal e vertical) e uma dimensão de valores (células da tabela), procedimento exemplificado na Figura 33. Posteriormente, as várias tabelas individuais são compiladas em uma única planilha do programa *Microsoft Excel*. Nesse documento são feitas as agregações que possibilitam reduzir a extensão de componentes muito longos resultando em poucas classes com maior significância, conforme definido no esquema de homogeneidade. Os cálculos estabelecidos na tabela de pertinência também são aplicados nessa planilha convertendo-se assim quantidades absolutas em unidades percentuais.

Figura 33 – Consultas individuais à base de dados do BI



Fonte: autor (captura de tela do software *Microsoft Excel*).

4.4 dashboards

4.4.1 Tratamento de informação básico: diagrama 1

4.4.1.1 Tabela homogênea e transcrição

Das operações descritas anteriormente resulta a Tabela 6, uma tabela homogênea que registra a forma final dos dados numéricos que serão a base das visualizações que se seguem. Trata-se de uma *construção normal* do tipo *fichário-imagem*, uma vez que apenas a dimensão (y)

Tabela 6 – Tabela homogênea

COMPONENTE	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
INGRESSANTES ABANDONO (%)	17,01	11,87	10,87	9,87	10,56	10,12	12,63	11,23	11,63	11,10	13,17	10,86	8,90
INGRESSANTES APTOS (%)	90,91	91,11	86,91	87,22	86,76	87,89	85,84	86,74	86,91	87,83	87,88	89,06	89,43
INGRESSANTES CONCLUINTE (%)	53,94	58,79	56,66	57,69	56,63	60,54	56,03	58,53	52,69	56,67	55,51	54,55	41,06
INGRESSANTES COEFICIENTE DE RENDIMENTO DO ALUNO	6,76	6,87	6,72	6,78	6,70	6,69	6,62	6,58	6,53	6,49	6,47	6,47	6,48
INGRESSANTES COEFICIENTE DE RENDIMENTO ALUNO EM TURMA	7,03	7,03	6,93	6,93	6,86	6,81	6,75	6,71	6,60	6,58	6,56	6,56	6,57
INGRESSANTES COTA COTISTA (%)	-	-	-	55,42	39,0	30,32	29,96	23,48	21,51	26,73	29,87	29,62	26,45
INGRESSANTES COTA NÃO COTISTA (%)	100	100	100	44,58	61,00	69,68	70,04	76,52	78,49	73,27	70,13	70,38	73,57
INGRESSANTES ELIMINAÇÃO (%)	1,11	1,07	1,30	2,89	1,36	1,77	1,73	1,11	1,33	1,06	0,88	1,02	1,78
INGRESSANTES ESTADO CIVIL CASADO (%)	13,46	8,88	8,28	7,79	8,03	6,45	3,52	3,85	5,29	7,67	5,84	6,55	6,53
INGRESSANTES ESTADO CIVIL SEPARADO E OUTROS (%)	1,81	1,42	1,42	0,68	0,77	0,67	0,28	0,77	1,18	1,56	1,74	1,69	1,74
INGRESSANTES ESTADO CIVIL SOLTEIRO (%)	84,73	89,71	90,30	91,53	91,20	92,89	96,19	95,38	93,53	90,77	92,42	91,76	91,72
INGRESSANTES ESTRUTURA ACADÊMICA CENTRO BIOMÉDICO (%)	8,39	8,36	7,84	8,25	8,45	8,32	8,30	8,48	8,65	8,46	8,09	7,78	8,04
INGRESSANTES ESTRUTURA ACADÊMICA CENTRO DE CIÊNCIAS SOCIAIS (%)	24,08	22,36	21,29	21,61	21,39	21,32	21,28	21,91	21,26	21,05	21,95	21,02	21,27
INGRESSANTES ESTRUTURA ACADÊMICA CENTRO DE EDUCAÇÃO E HUMANIDADES (%)	36,06	34,63	37,72	37,59	37,65	37,18	37,31	36,23	37,14	37,34	37,24	38,36	37,68
INGRESSANTES ESTRUTURA ACADÊMICA CENTRO DE TECNOLOGIA E CIÊNCIAS (%)	31,47	34,65	33,15	32,54	32,51	33,18	33,11	33,39	32,94	33,15	32,72	32,83	33,01
INGRESSANTES QUANTIDADE DE INGRESSANTES	5091	4943	5281	5514	5449	5413	5310	4953	5248	5181	5122	5281	5387
INGRESSANTES INTEGRALIZAÇÃO NO MÍNIMO OU ANTES (%)	5,26	4,39	4,06	3,82	3,88	3,73	3,32	3,50	3,82	3,46	3,75	3,25	4,56
INGRESSANTES INTEGRALIZAÇÃO 1-2 PERÍODOS APÓS MÍNIMO ATÉ MÁXIMO (%)	48,77	48,11	49,95	52,35	53,83	52,73	57,79	53,49	49,35	48,15	41,96	39,63	44,04
INGRESSANTES INTEGRALIZAÇÃO 3 PERÍODOS APÓS MÍNIMO ATÉ MÁXIMO (%)	27,07	28,00	25,84	25,66	23,91	24,73	21,61	24,32	27,69	29,18	30,68	33,71	30,96
INGRESSANTES INTEGRALIZAÇÃO MÁXIMO OU APÓS (%)	18,90	19,49	20,15	18,17	18,38	18,81	17,29	18,69	19,14	19,21	23,61	23,40	20,44
INGRESSANTES ORIGEM: REDE PÚBLICA (%)	37,06	32,5	37,66	60,03	51,64	51,76	49,68	44,11	43,59	45,53	43,97	47,32	46,71
INGRESSANTES ORIGEM: REDE PARTICULAR (%)	62,94	67,50	62,34	39,97	48,36	48,24	50,32	55,89	56,41	54,47	56,03	52,68	53,29
INGRESSANTES REPROVAÇÃO POR NOTA (%)	6,11	5,78	5,21	5,90	-	-	-	6,80	6,60	7,80	7,60	7,80	-
INGRESSANTES REPROVAÇÃO POR FREQUÊNCIA (%)	11,37	10,70	11,94	11,10	-	-	-	13,60	15,60	15,50	14,90	14,00	-
INGRESSANTES SEXO FEMININO (%)	46,95	45,36	44,08	46,06	47,72	45,22	47,50	48,29	46,82	47,13	46,86	45,77	47,11
INGRESSANTES SEXO MASCULINO (%)	53,05	54,64	55,92	53,94	52,28	54,78	52,50	51,71	53,18	52,87	53,14	54,23	52,89
INGRESSANTES TRANCAMENTO (%)	6,59	6,65	17,07	15,36	16,16	14,87	17,80	16,33	17,17	16,51	16,48	14,97	16,38
INGRESSANTES TRANSFERÊNCIA (%)	0,71	1,56	1,16	3,61	1,87	1,24	1,09	1,39	1,60	0,95	0,90	1,10	1,47
VESTIBULANDO ORIGEM: REDE PÚBLICA (%)	-	-	-	-	45,58	44,44	44,56	40,01	39,76	41,06	39,35	39,76	41,12
VESTIBULANDO ORIGEM: REDE PARTICULAR (%)	-	-	-	-	54,42	55,56	55,44	59,99	60,24	58,94	60,65	60,24	53,88
INGRESSANTES TURNO INTEGRAL (%)	10,17	10,19	9,35	9,75	10,68	10,88	10,56	10,21	10,60	10,37	9,84	9,22	9,80
INGRESSANTES TURNO MANHÃ (%)	38,49	39,42	41,48	42,22	40,91	40,35	41,07	41,17	42,25	41,94	42,52	40,79	40,53
INGRESSANTES TURNO NOITE (%)	51,34	50,39	49,16	48,03	48,42	48,77	48,37	48,62	47,15	47,69	47,64	49,99	49,67
VAGAS DISPONIBILIZADAS	-	-	-	4989	4989	4988	4992	4948	5035	5043	5107	5112	5136

Fonte: Projeto de BI, UERJ – cubo: desempenho acadêmico, graduação. 2013.

é permutável, pois os dados cronológicos (x) têm natureza ordenada, sendo, portanto, fixos. Os valores tabulados são então transcritos em séries de histogramas (*profils*, no jargão de Bertin) (Figura 34), cuja altura dos retângulos é proporcional às grandezas originais e as larguras são invariáveis, não tendo relação com os valores visualizados a não ser otimizar a

Figura 34 – Transcrição visual da tabela numérica

	00	01	02	03	04	05	06	07	08	09	10	11	12
INGRESSANTES ABANDONOS													
	13,01	11,87	10,87	9,87	10,54	10,12	12,43	11,23	11,43	11,10	13,17	10,84	8,90
INGRESSANTES APTOS													
	90,91	91,11	88,91	87,22	88,76	87,89	85,84	86,76	86,91	87,83	87,88	89,36	87,43
INGRESSANTES CONCLUINTES													
	53,96	58,79	54,66	57,69	56,43	60,54	54,03	58,53	52,69	54,07	55,51	54,55	61,06
COEFICIENTE DE RENDIMENTO ALUNO INSCRITO													
	6,78	6,87	6,72	6,78	6,70	6,69	6,62	6,58	6,53	6,49	6,47	6,47	6,48
COEFICIENTE DE RENDIMENTO ALUNO EM TURMA													
	7,03	7,03	6,93	6,93	6,86	6,81	6,75	6,71	6,60	6,58	6,54	6,54	6,57
INGRESSANTES COTISTA +	-	-	-										
				55,42	59,0	30,32	29,96	23,48	21,51	26,73	29,87	29,62	26,45
INGRESSANTES NÃO COTISTA	100	100	100	44,58	41,00	49,68	70,04	76,52	78,49	73,27	70,13	70,38	73,57
INGRESSANTES ELIMINADOS													
	1,11	1,07	1,20	2,89	1,34	1,77	1,73	1,19	1,33	1,06	0,88	1,03	1,78
INGRESSANTES ESTADO CIVIL CASADO													
	13,44	8,88	8,28	7,79	8,03	6,45	3,52	3,85	5,29	7,67	5,84	4,55	4,53
INGRESSANTES ESTADO CIVIL SEPARADO E OUTROS													
	1,81	1,42	1,42	0,48	0,77	0,47	0,28	0,77	1,18	1,54	1,76	1,49	1,74
INGRESSANTES ESTADO CIVIL SOLTEIRO													
	84,73	89,71	90,30	91,53	91,20	92,81	96,19	95,38	93,53	90,77	92,42	91,76	91,72
INGRESSANTES ESTRUTURA ACADÊMICA CENTRO BIOMÉDICO													
	6,39	6,34	1,84	6,25	6,45	6,32	6,30	6,48	6,45	6,46	6,39	1,78	6,04
INGRESSANTES ESTRUTURA ACADÊMICA CENTRO DE CIÊNCIAS SOCIAIS													
	24,08	22,36	21,29	21,61	21,39	21,32	21,28	21,91	21,26	21,85	21,95	21,02	21,27
INGRESSANTES ESTRUTURA ACADÊMICA CENTRO DE EDUCAÇÃO E HUMANIDADES													
	34,06	34,43	37,73	37,59	37,45	37,18	37,31	36,23	37,16	37,24	37,26	38,26	37,48
INGRESSANTES ESTRUTURA ACADÊMICA CENTRO DE TECNOLOGIA E CIÊNCIAS													
	31,47	34,45	33,15	32,54	32,51	33,18	33,11	33,39	32,76	33,15	32,72	32,83	33,81
INGRESSANTES QUANTIDADE DE INGRESSANTES													
	5891	4963	5281	5514	5649	5813	5310	4953	5048	5181	5222	5281	5387
INGRESSANTES INTEGRALIZAÇÃO NO MÍNIMO OU ANTES													
	5,24	6,39	4,06	3,82	3,88	3,73	3,32	3,50	3,82	3,64	3,75	3,25	4,54
INGRESSANTES INTEGRALIZAÇÃO 1-2 PERÍODOS APÓS MIN. ATÉ MÁX.													
	48,77	48,11	49,95	52,35	53,83	52,73	51,79	53,49	49,35	48,15	45,76	49,63	44,84
INGRESSANTES INTEGRALIZAÇÃO 3 PERÍODOS APÓS MIN. ATÉ MÁX.													
	27,07	28,00	25,84	25,44	25,91	26,73	21,61	26,22	27,49	29,18	30,48	33,71	28,94
INGRESSANTES INTEGRALIZAÇÃO MÁXIMO OU APÓS													
	18,90	19,49	20,15	18,17	18,38	18,81	17,29	18,49	19,14	19,21	22,61	22,48	20,44
INGRESSANTES ORIGEM: REDE PÚBLICA													
	37,06	32,5	37,66	40,03	51,66	51,76	49,68	46,11	43,59	45,53	42,97	47,32	46,71
INGRESSANTES ORIGEM: REDE PARTICULAR													
	62,96	67,50	62,34	59,97	48,34	48,24	50,32	55,89	56,41	54,47	56,03	52,68	53,29
INGRESSANTES REPROVAÇÃO POR NOTA													
	6,11	5,78	5,21	5,90				6,80	6,60	7,80	7,80		
INGRESSANTES REPROVAÇÃO POR FREQUÊNCIA													
	11,37	10,70	11,94	11,10				13,60	10,40	15,50	16,90	14,00	
INGRESSANTES SEXO FEMININO													
	66,75	63,36	64,08	66,96	67,72	65,22	67,50	68,29	68,82	67,12	68,86	65,77	67,11
INGRESSANTES SEXO MASCULINO													
	53,05	54,64	55,92	53,74	52,28	54,78	52,50	51,71	53,18	52,87	53,14	54,23	52,89
INGRESSANTES TRANCAMENTOS													
	6,09	6,45	77,07	75,34	74,14	76,87	77,80	76,33	75,17	76,51	76,48	74,77	76,28
INGRESSANTES TRANSFERÊNCIAS													
	0,71	1,54	1,36	3,61	1,87	1,24	1,09	1,39	1,60	0,95	0,90	1,10	1,47
VESTIBULANDO ORIGEM: REDE PÚBLICA	-	-	-	-									
					45,58	44,44	44,58	40,01	39,76	47,06	39,35	39,76	47,12
VESTIBULANDO ORIGEM: REDE PARTICULAR	-	-	-	-									
					56,42	55,54	55,44	59,99	60,24	58,76	60,45	60,24	53,88
ING T.INT TURNO INTEGRAL													
	16,77	10,19	9,35	9,75	10,64	10,88	10,56	10,21	10,40	10,37	9,86	9,22	8,80
ING T.MAN TURNO MANHÃ													
	38,49	39,42	41,68	42,22	40,91	40,35	41,07	41,77	42,25	41,94	42,52	40,79	40,53
ING T.NOITE TURNO NOITE													
	51,36	50,39	51,14	48,03	48,42	48,77	48,37	48,42	47,75	47,49	47,64	45,99	47,47
VAGAS DISPONIBILIZADAS	-	-	-	-									
					4789	4789	4789	4789	4789	5043	5107	5112	5136

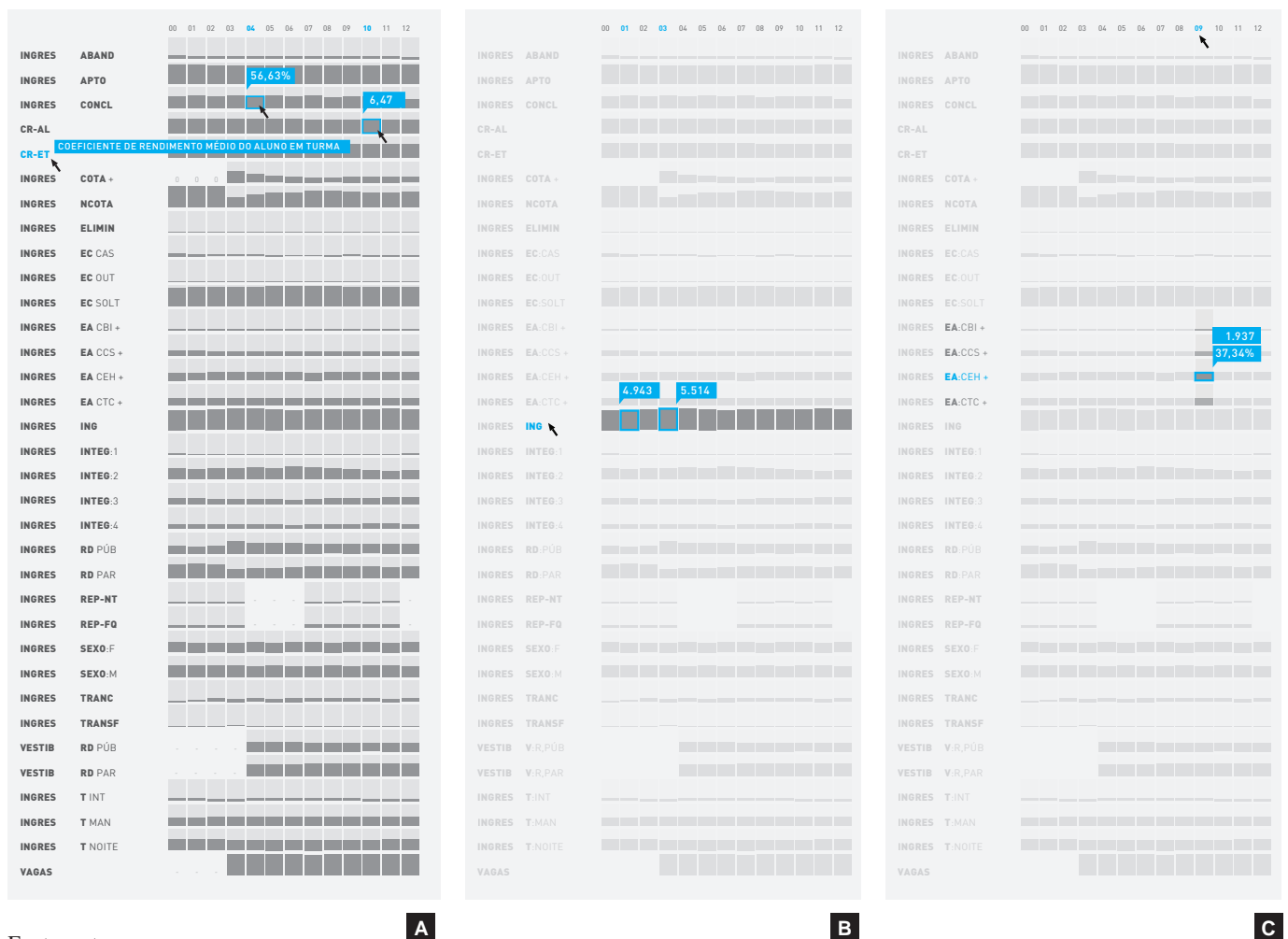
visibilidade. As barras em cinza claro utilizadas nos componentes com unidades percentuais (a maioria, excetuando-se os *CRs*, as *vagas* e os *ingressantes*, apresentados em valores absolutos) representam o percentual máximo de 100%, viabilizando em cada item a comparação da parte (em cinza médio) com o todo.

Considerando o domínio digital no qual estão inseridos os dashboards, algumas intervenções podem contribuir para uma melhor fruição do gráfico, como a racionalização do espaço disponível. Como mencionado, dashboards envolvem a questão muito concreta do aproveitamento da área em que são visualizados. Sendo assim, tendo em conta a superfície limitada das telas digitais, e ainda, a variabilidade de tamanhos e formatos em razão da vasta gama de aparelhos digitais, fica claro a importância de interfaces que se adaptem a tais espaços restritos. Um dos recursos que contribuem para isso é a alternância entre a visibilidade e a ocultação de itens, de modo a reduzir a quantidade de elementos na tela. Na verdade, a ausência da informação exposta em sua completude parece ser uma característica necessária do domínio digital, tornando-se mesmo parte de sua linguagem. Neste sentido, o presente trabalho considera esse contexto, assumindo que se podem suprimir os valores dos gráficos, assim como reduzir os elementos de identificação de cada *componente* para siglas. Para visualizá-los bastaria sobrepor o cursor (*mouse over*) ¹⁶ ao item para o qual se quer a informação completa, possibilidade ilustrada na Figura 35A.

Ainda neste estágio inicial, outro ponto a ser considerado com relação à perspectiva digital são os níveis de informação básicos. Após a transcrição, o gráfico situa-se no *nível elementar de informação*, isto é, nas questões de natureza mais simples. Elas envolvem uma pergunta feita ao componente no eixo x cuja resposta encontra-se nos componentes em y ou vice-versa. Perguntas do tipo “Em 2009, qual a estrutura acadêmica com maior (ou menor) percentual de alunos?”; ou ainda “Considerada a quantidade de ingressantes, em qual ano houve o maior (ou menor) número de ingressos?”. Essas questões, facilmente, respondidas pela análise do gráfico estático encontram resposta ainda mais rápida em uma interface digital que preveja esse tipo de consulta dentre suas possibilidades, como ilustra a Figura 35B e C. Entretanto, como comentado, o nível de informação elementar é atingido por todos os gráficos, de modo que interessa tratar daquilo que é peculiar ao método de Bertin: a análise no nível médio e no

¹⁶ Apesar desta seção voltar-se aos procedimentos que independem do uso do computador, esse recurso foi tratado aqui considerando-se sua trivialidade (dispensando maiores explicações) e porque com isso, as imagens daqui para frente podem de valer de uma ocupação mais racionalizada da página impressa devido à redução de elementos visuais.

Figura 35 – Exemplificação do recurso de *mouse over* para revelar informações secundárias



Fonte: autor

nível de conjunto. Isso envolve certas alterações na escala do histograma convencional, a fim de evidenciar tais níveis de leitura.

Concluídas as etapas da formulação da tabela homogênea e da transcrição, deve-se ter em conta que a peculiar organização de *componentes* sobrepostos descrita por Bertin levanta questões igualmente particulares devido à visualização conjunta dos diversos dados. Uma das mais interessantes está sintetizada na proposição da “lei da visibilidade”, descrita anteriormente e aplicada aqui. Ela diz respeito à adequação das diferentes escalas às necessidades de uso e comunicação específicos, envolvendo ainda questões que extrapolam a simples transcrição gráfica. A seguir são discutidas as propriedades dessas várias escalas.

4.4.1.2 As questões de escala

4.4.1.2.1 Escala absoluta

Trata-se da representação em que todos os valores percentuais são inscritos na mesma escala, de forma que o valor máximo de 100% é representado por uma coluna de mesma altura para todos os componentes e suas classes. Assim, as pequenas relações (p. ex. 1 – 5%) são representados por barras curtas, as intermediárias (40 – 50%), por barras médias e os valores mais altos são transcritos por barras elevadas. Por outro lado, esta representação, que pelo senso comum seria a mais “natural”, precisa e correta, pode deixar passar despercebida informações mais abrangentes. A Figura 36A destaca os componentes que, devido à pequena ordem de grandeza, têm a visibilidade prejudicada, como acontece às *integralizações* de nível 1, às *eliminações*, às *transferências* etc. Essa questão é exemplificada também ao se desdobrar o item cota revelando as modalidades de cotas. Vê-se que as reservas para negros e escolas públicas, e o percentual de ingressantes não cotistas são bastante discrepantes em relação aos cotistas deficientes e indígenas. Por causa disso, representar esses últimos em uma escala absoluta revela-se uma visualização inútil, pois ilegível. Corrigir esse problema, preservando a relação entre todas as modalidades de cota, demanda a ampliação de todas elas até que os valores das cotas para deficientes e indígenas tornem-se legíveis. A Figura 36B, amplia a escala original dez vezes, mas, mesmo assim, não se chega a uma leitura confortável. Entretanto, o espaço tomado somente pelo componente *cota* é maior do que o ocupado pelo quadro anterior com seus 20 componentes, contrariando o sentido de ocupação eficiente do espaço que é prerrogativa de bons dashboards. Certamente, uma interface digital poderia prever o recurso de alternar entre essa vista de *zoom* e a original ou mesmo justapô-las. Porém a visão de conjunto assim como a praticidade de uso ficariam comprometidas, sobretudo porque o problema em *cota* repete-se em outros componentes, como comentado. Sendo assim, apesar da precisão representacional da escala absoluta, ela é deficiente em situações como as descritas, casos em que as escalas relativas podem responder com mais flexibilidade.

Figura 36 – Questões da representação em escala absoluta



Fonte: autor

A

B

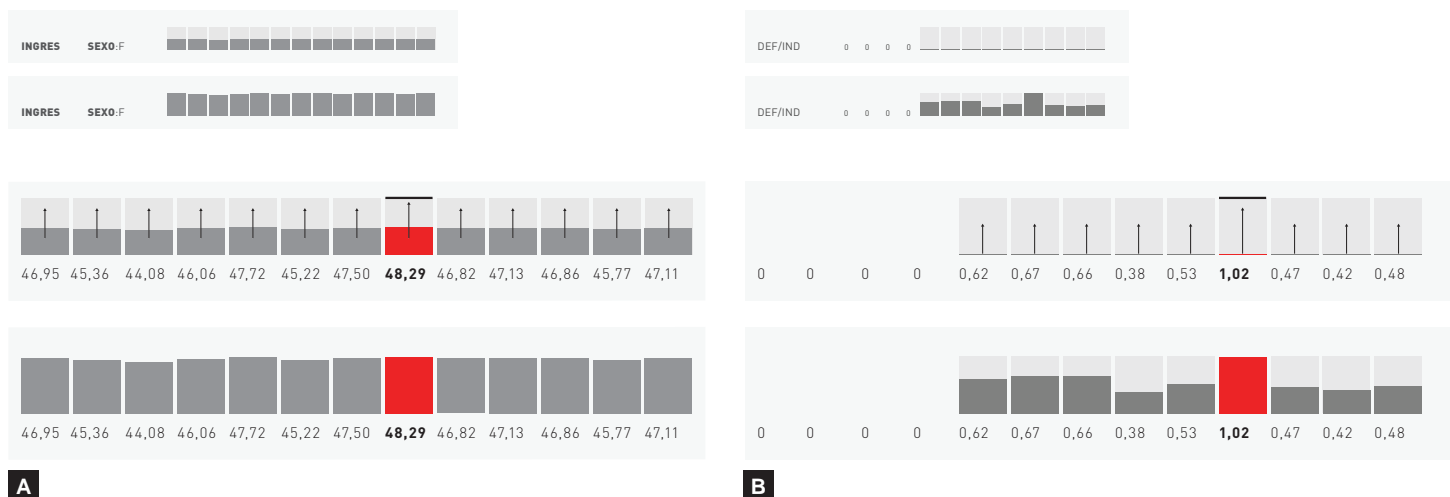
4.4.1.2.2 Escalas relativas

São escalas específicas a cada histograma, que não se baseiam em única gradação que referencia todos eles, como no caso anterior. Elas priorizam a visibilidade do gráfico em detrimento da

fidelidade quantitativa do conjunto. Nessa técnica, os maiores valores de cada histograma são balizados de acordo com uma altura máxima pré-estabelecida. Esse procedimento assegura a visibilidade dos gráficos, evidenciando tendências que possam existir como aumento, regressão etc. Há mais de uma forma de estabelecer uma escala visível mantendo-se a proporcionalidade entre os membros de uma mesma série, mas abrindo mão da proporcionalidade do conjunto. Todos essas formas envolvem uma expansão das alturas em comparação à escala absoluta e cada qual se adequa melhor a determinadas análises e situações de uso.

Expansão simples: Demonstrada pela Figura 37, trata-se da ampliação do maior elemento de um histograma até um limite pré-estabelecido, aumentando-se os demais elementos proporcionalmente. No exemplo, o limite determinado foi a altura da barra de 100% usada na escala absoluta das figuras anteriores. A expansão simples caracteriza-se por preservar as relações originais entre os elementos, de forma que as diferenças entre valores muito próximos tendem a permanecer sutis, mesmo após expandidas, como acontece em 37A. Por outro lado, oscilações mais heterogêneas, mesmo conservando relações entre os itens após a expansão, podem ser percebidas com maior clareza, como em 37B. Trata-se, assim, de uma espécie de *zoom* individual sobre cada histograma que compõem um diagrama, assegurando a visibilidade de todos eles, mas sem preocupações em evidenciar as distinções entre os elementos de uma série e, conseqüentemente, das séries entre si. Essa demanda é atendida pelas escalas que, a partir de um valor determinado, expandem os itens que a compõem.

Figura 37 – Escala de expansão simples



Acima (proporção original): comparação entre o mesmo histograma em escala absoluta e expansão simples.
Abaixo (histogramas ampliados 2,5×): processo de conversão, mostrando a expansão a partir do maior valor.
Fonte: autor

Expansão pelo valor mínimo: A supressão da parte que é comum a todos os elementos de um mesmo histograma é uma operação recomendada por Bertin para ressaltar as diferenças entre os itens da série. Esse procedimento acontece pela eliminação da porção correspondente ao menor valor de cada série e a expansão da parte restante até que a maior delas iguale a altura máxima delimitada, da mesma forma como na escala de expansão simples. Disso decorre que, em cada série, as ocorrências do item com valor menor são suprimidas. Evidentemente, havendo mais de uma ocorrência do valor mínimo, todas elas serão suprimidas. A Figura 38 demonstra o uso dessa escala. Nota-se a maior visibilidade das diferenças entre as grandezas do que na escala de expansão simples resultando em um gráfico mais nuançado. Por outro lado, esta é uma escala sensível a valores discrepantes, de forma que quando a série apresenta um valor sensivelmente mais alto que os demais, isto faz com que não haja tanta diferenciação entre os demais valores como aconteceria caso não houvesse um valor atípico e a série toda tendesse à homogeneidade, como é o caso da figura em questão.

Figura 38 – Escala de expansão pelo mínimo

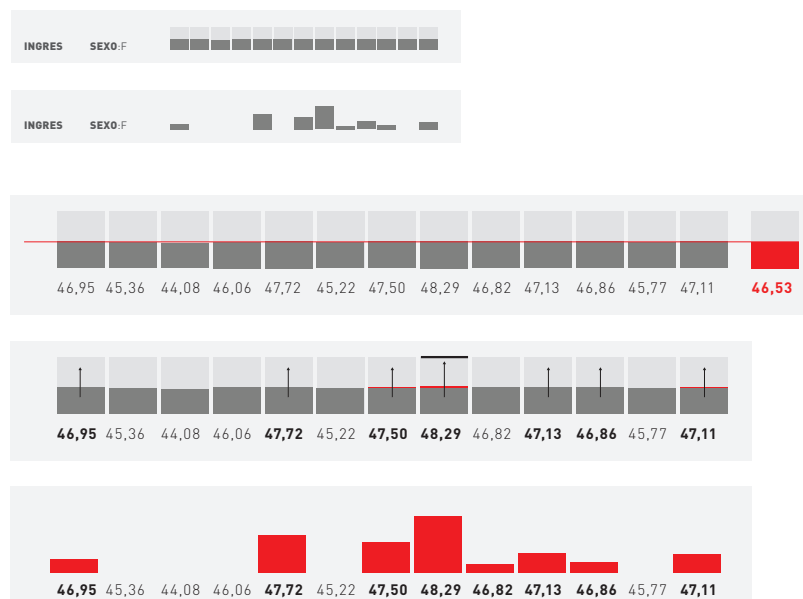


Acima (proporção original): comparação entre o mesmo histograma em escala absoluta e expansão a partir do valor mínimo. Abaixo (histogramas ampliados 2,5×): processo de conversão, mostrando a expansão a partir da maior diferença em relação ao valor mínimo.

Fonte: autor

Expansão pela métrica: Técnica similar à anterior, mas ao invés da expansão se dar a partir da parte comum, ela acontece a partir de uma métrica pré-determinada. A figura 39 ilustra uma expansão pela média dos valores da série, na qual somente os itens que excedem essa métrica são expandidos, como nas escalas anteriores. Os valores inferiores à média são suprimidos. A utilidade dessa escala está na visualização isolada dos valores de interesse em comparação a qualquer medida estatística, como a média, mediana, quartil etc. ou ainda um valor arbitrário, como “acima de 5.000 alunos”, ou “mais alto do que 50%”. Considerando-se os recursos digitais, também seria possível uma expansão dos itens abaixo da métrica, também com a finalidade de isolá-los e visualizar somente essas ocorrências. Apesar de possivelmente ser um recurso analítico de grande utilidade, as expansões por métricas não foram aplicadas no presente estudo. Isso acontece por dois motivos. O primeiro deve-se à tendência a histogramas com muitas descontinuidades (por conta da supressão de todos os valores abaixo da métrica), o que compromete a análise conjunta de diversos histogramas. O segundo ponto é a impossibilidade técnica de demonstrar as aplicações em que as expansões pela métrica mostram-se como um poderoso recurso analítico. Trata-se da alternância e comparação entre

Figura 39 – Escala de expansão pela métrica (média)



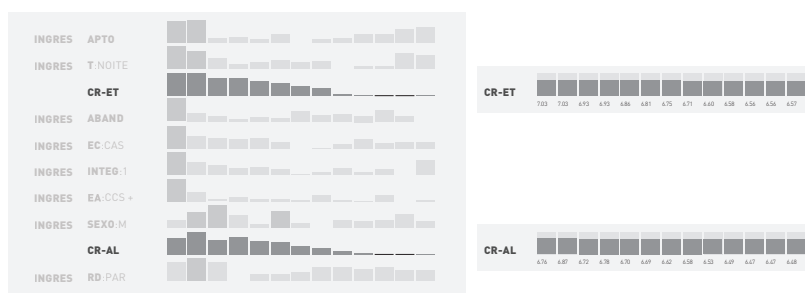
Acima (proporção original): comparação entre o mesmo histograma em escala absoluta e expansão a partir do valor mínimo. Abaixo (histogramas ampliados 2,5×): processo de conversão, mostrando a expansão a partir da maior diferença em relação ao valor mínimo.

Fonte: autor

as diversas possibilidades de métricas estatísticas ou aquelas definidas como metas a atingir, recursos que, em si, mereceriam um estudo à parte.

Apresentadas tais escalas, um ponto importante a ser considerado no uso das *escalas de expansão pelo mínimo e expansão pela métrica* refere-se à característica de ênfase, que é próprias dessas escalas. Nesse sentido, ressaltar as diferenças entre os itens de um histograma pode contribuir como evidência para corroborar análises às vezes dramatizadas pelo gráfico. Muitos comentadores já apontaram os modos como determinados vieses em gráficos podem direcionar o julgamento dos leitores para uma interpretação questionável do fenômeno em questão. Como exemplo, a Figura 40 destaca os componentes que apresentam os coeficientes de rendimento médio dos alunos em geral e dos alunos em turma [CR-AL]; [CR-ET] comparados na escala de expansão pelo mínimo e, à direita, na escala absoluta. No primeiro caso, entende-se o gráfico como um declínio consistente na nota dos alunos da instituição. Por conta disso, essa imagem poderia embasar argumentos contrário à política de cotas, ressaltando a derrocada no desempenho geral da universidade. Por outro lado, o gráfico em escala absoluta revela, na verdade, uma discreta involução. Evidentemente, a queda existe, como inclusive os valores podem atestar, mas a questão que se coloca é se é preciso ressaltá-la e, se for, em que condições. Essa situação ilustra como os recursos visuais podem influenciar determinadas interpretações dos gráficos direcionando a atenção em função de conclusões estabelecidas de antemão e colocando enfaticamente a questão da integridade na apresentação de dados. Por outro lado, o eventual mal uso desses recursos não deveria impedir seu emprego nas muitas situações em que eles se revelam mais adequados do que a mera transcrição dos dados em escala absoluta. No que concerne à esta dissertação, dois pontos permitem contextualizar esse risco. 1) Deve-se

Figura 40 – Escalas de expansão pela métrica: questões de interpretação



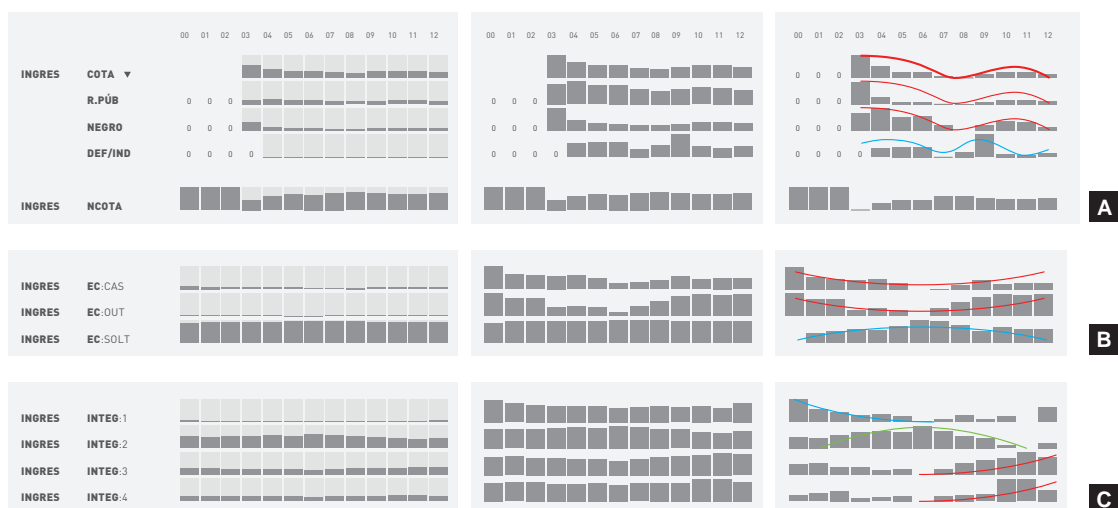
Fonte: autor

considerar que os recursos interativos permitem alternar facilmente entre uma escala e outra, conforme a finalidade do gráfico. 2) De modo geral, o problema em questão torna-se mais grave quando um gráfico é apresentado isoladamente o que não costuma ser o objetivo dos métodos de Bertin que, ao processar dados em grandes conjuntos, tem nas escalas relativas um poderoso recursos de diferenciação visual.

4.4.1.2.3 Comparação entre escalas

A questão dos ingressantes cotistas deficientes e indígenas é retomada na Figura 41 que compara as escalas de expansão simples e pelo mínimo com a escala original (absoluta). Observa-se que em ambas as escalas relativas, além de se ter clara a flutuação entre os ingressantes deficientes e indígenas [INGRES DEF/IND] através dos anos, revela-se que eles têm um padrão diferente dos outros cotistas. Ao passo que negros e estudantes de escola pública parecem constituir a tendência geral do conjunto dos cotistas (pois revelam padrão similar ao apresentado por tal conjunto), os deficientes e indígenas revelam um padrão anômalo, condição que só se torna visível pelo emprego das escalas relativas. Observações análogas podem ser feitas em relação aos componentes estado civil [INGRES EC] e integralizações [INGRES INTEG], no quais a

Figura 41 – Comparação entre escalas



Fonte: autor

relação entre suas classes constituintes pôde ser melhor compreendida devido ao uso de escalas relativas. No primeiro caso (Figura 41B), assinala-se a oposição entre os padrões de ingresso dos solteiros em relação tanto ao grupo dos casados quanto dos separados, ressaltando-se assim o comportamento semelhante na evolução desses últimos grupos. Já no item *integralizações* (Figura 41C) percebem-se três estruturas características. As integralizações de nível 1 sugerem uma tendência oposta às integralizações 3 e 4: enquanto a primeira mostra um comportamento tendencialmente involutivo, as demais revelam um aumento em anos recentes. Nesse conjunto, as *integralizações* de nível 2 são uma exceção apresentando um comportamento de ascensão inicial e queda nos anos recentes. Sendo assim, a análise dos níveis de integralização parece sugerir um movimento recente em direção às integralizações mais longas.

A Figura 42 compara as escalas absoluta (42A), expansão simples (42B) e expansão pelo mínimo (42C) aplicadas ao *diagrama 1* (panorama do programa de cotas) do dashboard básico (Figura 28, p.102) demonstrando o comportamento característico de cada escala na vista conjunta dos dados. Importa ressaltar que a natureza específica de um grupo de dados influencia na forma assumida pelos perfis e que o aspecto que se quer evidenciar (por exemplo, a semelhança ou a diferença entre os elementos de um histograma) direciona fortemente a escolha da escala. Pela observação da escala absoluta percebe-se que há relativa homogeneidade na maior parte dos componentes, pois, de modo geral, não há variações drásticas ao longo dos anos, tendendo-se mais a uma flutuação discreta. O segundo diagrama emprega a escala de expansão simples, que preservou a homogeneidade observada na escala absoluta dando visibilidade às baixas porcentagens. Em relação à imagem geral formada, a maior ocupação da altura das linhas resulta em um diagrama muito mais escuro e condensado do que o anterior. No terceiro caso, as diferenças originalmente sutis são ressaltadas de modo que se chega a histogramas mais “recortados” que permitem discernir melhor tendências de progressão, retrocesso etc.

As especificidades de cada uma das escalas sugere interessantes questões ligadas à interpretação do gráfico, que serão discutidas no contexto da perspectiva digital, em relação à possibilidade de alternância entre tais formas de visualização, permitindo vistas mais adequadas a usos particulares.

Figura 42 – Comparação entre escalas – diagrama 1: panorama



4.4.1.3 Análise e interpretação

Após as considerações prévias sobre as diversas escalas, a morfologia do diagrama 1 (*panorama*) utilizado a partir daqui será baseada na escala de expansão pelo mínimo, considerando que esse recurso pode revelar diferenças que de outro modo passariam despercebidas em razão da relativa homogeneidade na flutuação dos dados em questão. Assim, na Figura 42C, torna-se claro, por exemplo, a redução nos níveis de coeficiente de rendimento; a involução dos ingressantes casados até 2007, quando então passam a oscilar e também a tendência de redução dos ingressantes cotistas até 2008 etc. Entretanto, essas são observações pontuais e restritas a casos particulares que não contribuem para uma compreensão da totalidade do diagrama. O

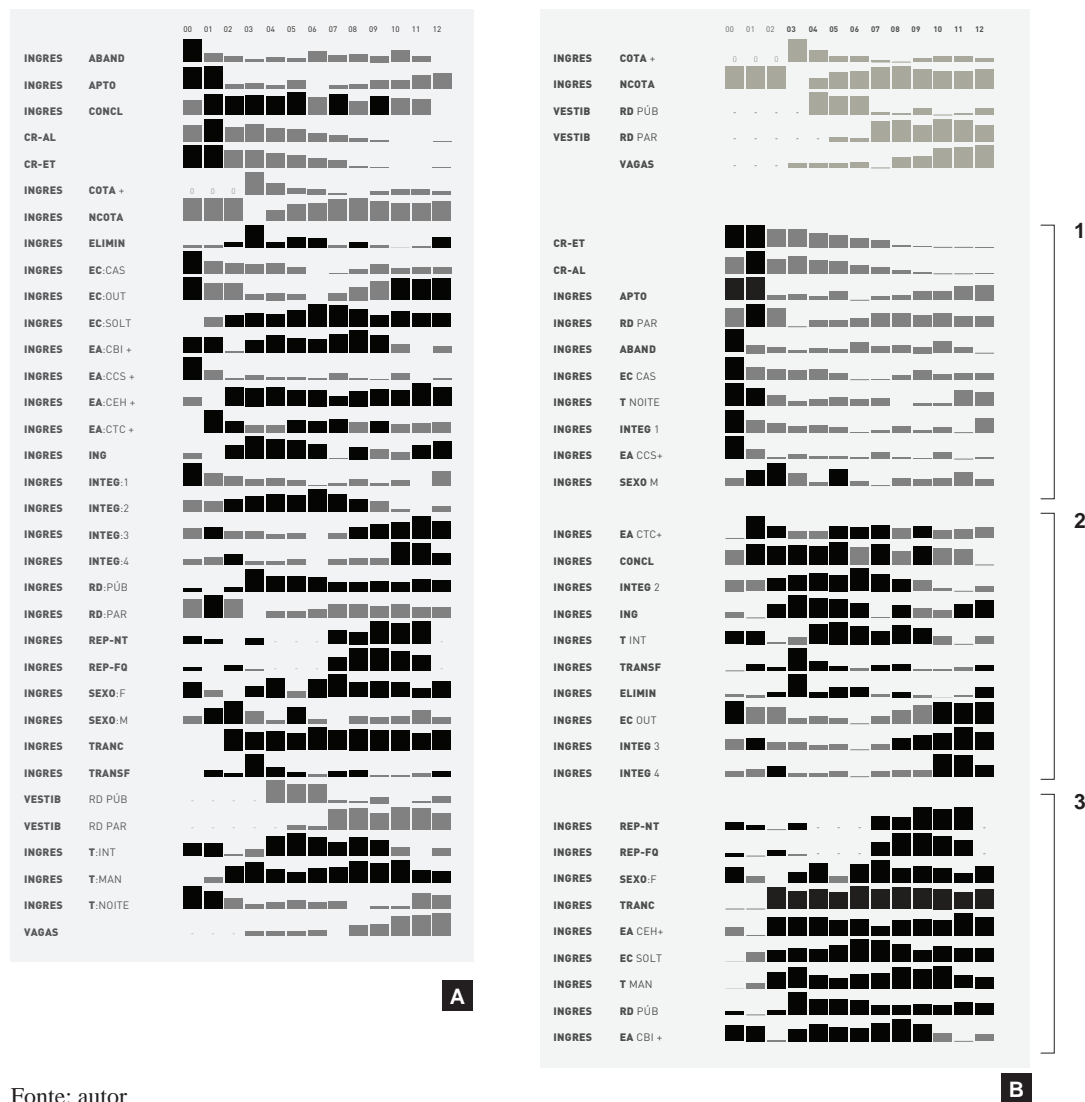
objetivo do dashboard apresentado é visualizar, em um mínimo de *imagens*, características gerais da universidade expressas por sua população de alunos antes e após a implantação da reserva de vagas. Nesse sentido, uma característica desejável da sua observação seria encontrar resposta rápida para questões de ordem geral, tais como: “O que mudou?” O que se manteve como era?” “O que aumentou, e o que diminuiu?”. Para tanto, é importante introduzir um fator de comparação que permita avaliar essas situações, e ainda estabelecer uma diferenciação visual para representá-lo, isolando assim os padrões que podem dar resposta às perguntas colocadas. Considerados os dados disponíveis, um fator de comparação lógico são os valores dos anos anteriores a 2003, quando as cotas começaram. Como o sistema só contém dados dos três anos imediatamente anteriores às cotas, a comparação limita-se a essa pequena amostragem. Mesmo assim, pode-se estabelecer um elemento comparativo pelo uso da *média* entre tais anos, assinalando quais componentes, em anos posteriores, ficaram acima ou abaixo da média entre 2000 a 2002. Assim, a Figura 43A considera, para cada linha do diagrama, os valores que estão acima dessa métrica¹⁷, assinalando-os em preto, e os que estão abaixo são preenchidos em cinza. Com essa marcação, percebe-se que, a despeito das oscilações, muitos componentes sofreram aumento em relação aos anos anteriores às cotas. Entretanto, ainda é difícil estabelecer uma visão de síntese do painel. Para tanto, é preciso rearranjar os componentes.

4.4.1.4 Permutação

Trata-se das operações de reordenamento das linhas individuais para que elas possam dar resposta às questões propostas inicialmente. Como discutido, elas são agrupadas por semelhança constituindo uma forma geral mais simples, reduzindo o conjunto de vários componentes a poucos grupos significantes. A Figura 43B apresenta uma versão reordenada do gráfico anterior. Essa permutação acontece em função do número de ocorrências de itens acima da média. Destacam-se, assim, três grupos: aquele dos componentes que apresentam redução consistente em relação à média entre 2000 e 2002 e cuja imagem é mais clara [1]. Há o grupo que reúne os componentes que oscilaram

¹⁷ Excepcionalmente para os componentes [REP-NT] (reprovação por nota) e [REP-FQ] (reprovação por frequência) a média foi calculada em relação aos anos 1998, 1999 e 2002, devido a inexistência de dados para os anos de 2000 e 2003. Também por isso, e somente em relação reprovações, os gráficos que ocupam o espaço reservado aos anos 2000 e 2001, na verdade, referem-se aos valores de 1998 e 1999, respectivamente.

Figura 43 – Visibilidade dos itens acima da média e permutação



Fonte: autor

[2], isto é, ora aumentaram e ora reduziram, podendo ainda ser percebido o subgrupo superior cujo aumento aconteceu entre os anos de 2003 e 2008 e outro subgrupo inferior, composto pelas ampliações em anos recentes. O último grupo [3] contém os itens que aumentaram consistentemente, condição percebida pela imagem mais escura que resulta desse agrupamento. (Há ainda um quarto grupo, no topo do gráfico, em cinza quente, que reúne os componentes que não tem significância de comparação pois só há informações para eles após 2002; entretanto, eles foram mantidos por trazerem informações contextuais, por exemplo a redução de ingressantes cotistas ao longo dos anos, e porque são importantes para o desdobramento do dashboard no âmbito digital). Desse modo, considerando os três grupos identificados, dois tópicos merecem ser melhor explicados.

Um deles é que, ao estabelecer um fator de comparação e assinalar o desvio em relação a ele em tons de claro e escuro, criam-se duas chaves de leitura do gráfico. A primeira é a convencional e mais nuançada, que acontece pela observação das alturas das barras, avaliando-se o desenvolvimento ao longo do tempo. A segunda é binária, isto é, basta saber o que está acima ou abaixo da média; o que tem interesse e o que não tem. No gráfico em questão, as duas leituras convivem, embora presuma-se a proeminência da leitura binária (que simplifica o conjunto) e, num momento posterior (depois que se sabe se o item está ou não no grupo de interesse), analisa-se seu desdobramento no tempo.

O segundo ponto diz respeito aos critérios para agrupamento que, tal como a opção por uma escala, também dependem de escolhas que não são automáticas, envolvendo critérios semânticos e/ou arbitrários. Boa parte dos componentes podem ser identificados claramente como membros do grupo em que foram organizados. Desse modo, o item [INGRES] [ABAND] (abandonos) é decididamente um integrante do grupo que reduziu em relação à média, assim como [INGRES] [RD PÚBL] (egressos do ensino público) pertence certamente ao grupo das ampliações. Por outro lado, há casos limítrofes em que a decisão pelo pertencimento a um grupo ou outro não é tão evidente. Isso acontece com os componentes [INGRES] [SEXO M] e [INGRES] [INTEG 4] (integralizações longas). A princípio, ambos poderiam ser colocados tanto no grupo das reduções quanto no das oscilações. Desse modo, é preciso arbitrar critérios para seu posicionamento que, no presente caso, são os seguintes: 1) a quantidade de ocorrências acima da média; e 2) a semelhança a outros membros do grupo. Por isso, [INGRES] [INTEG 4] foi classificado como comportamento variável. Arbitrou-se que mais de três anos marcados em preto indicariam que o item pertence ao grupo da alternância em relação à média. Além disso, a concentração dos anos acima da média em período recente assemelha-se ao padrão das *integralizações* de nível 4, que já é um membro do grupo das alternâncias, em função do primeiro critério.

No âmbito da interpretação dos grupos delineados, uma análise superficial desses agrupamentos já poderia inferir um processo de mudança no perfil do alunado da UERJ. Observa-se que no grupo das reduções encontram-se os ingressantes do turno da noite [INGRES] [T NOITE] e dos casados [INGRES] [EC CAS], fato que sugere alunos mais jovens (que não trabalham e ainda não se casaram), o que, evidentemente, implicaria em diminuição do número de ingressantes mais velhos. Nota-se também uma redução dos ingressantes masculinos [INGRES] [SEXO M], cuja contraparte é o maior ingresso de mulheres e ainda a maior participação de alunos provenientes do

ensino público [INGRES] [RD PUB], o que confirmaria o êxito pela inclusão de tais estudantes por meio das ações afirmativas. Essas são algumas das observações iniciais que poderiam interessar ao pesquisador. Além delas, outros tópicos podem ser esclarecidos e ressaltados empregando recursos de comunicação visual e textual.

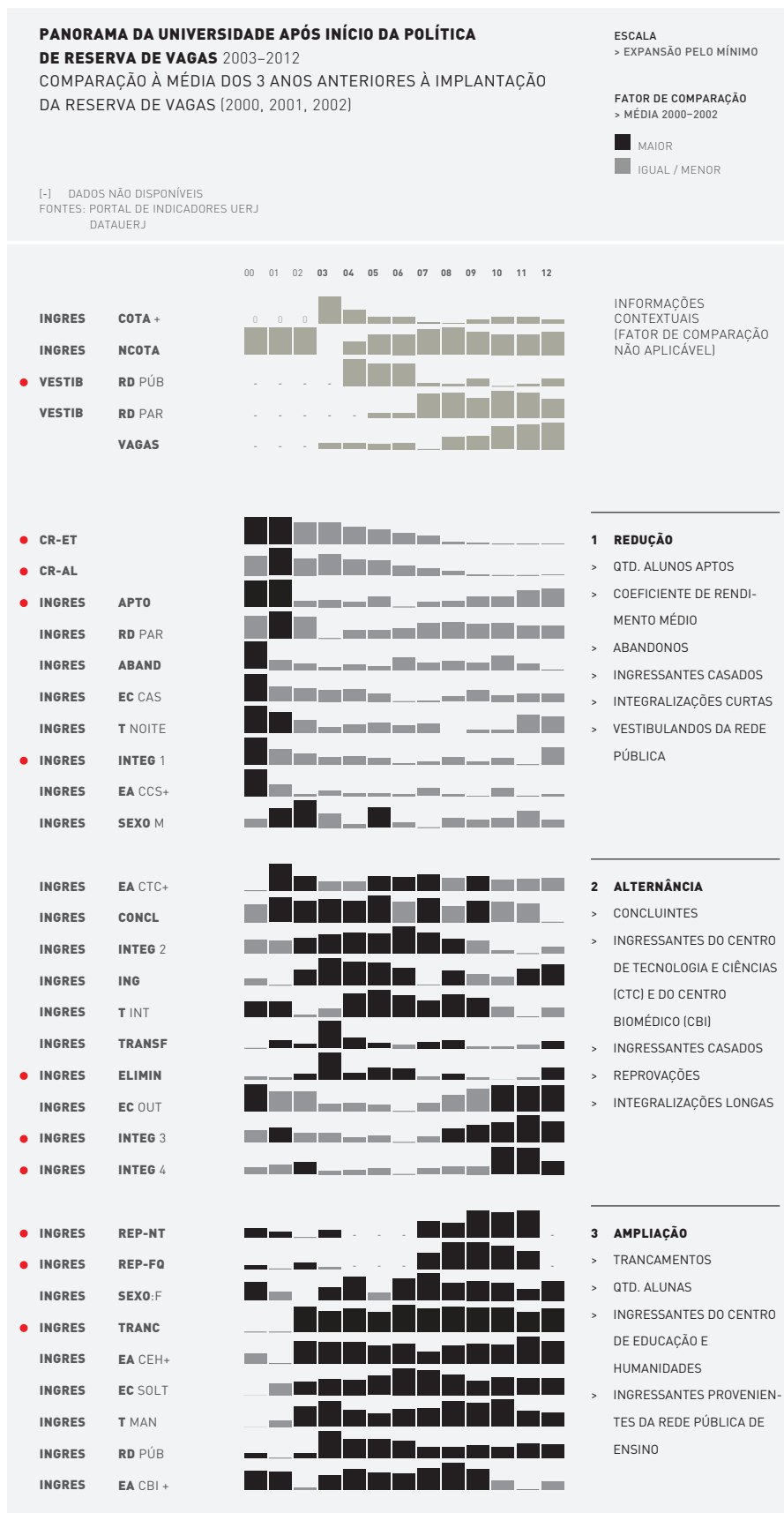
4.4.1.5 Gráfica de comunicação

Nesta etapa, são adicionados os elementos explicativos que viabilizam a compreensão dos padrões encontrados, efetivando sua comunicação. Na Figura 44, o cabeçalho agrupa as informações de identificação em um título e subtítulo. Bertin denomina essa identificação como a *informação externa*, no sentido em que estabelece a conexão entre o gráfico e a realidade não gráfica (não monossêmica), em contraposição à *informação interna*, aquela resultante dos aspectos concretos do gráfico que, por serem baseados nas propriedades monossêmicas, seriam independentes de contexto. Além disso, é descrito no cabeçalho o tipo de escala utilizada, assim como o fator de comparação e sua legenda. (Cabe notar que, em uma perspectiva digital, estes dois últimos itens poderiam atuar como operadores interativos, permitindo a troca de escala e do fator de comparação). À direita do gráfico, os grupos são identificados por suas características principais, listando ainda os principais membros que os compõem constituindo uma espécie de sinopse da informação. Marcadores em vermelho atuam como alertas para os tópicos que sugerem atenção especial, como a queda no CR e na quantidade de alunos aptos ou a ampliação dos trancamentos. Dessa forma, atendem-se às demandas tradicionais dos dashboards por meio da rápida comunicação dos pontos problemáticos.

O duplo registro constituído pelo gráfico e seus elementos explicativos constitui uma ponte entre os dados e o destinatário da comunicação, potencializando que tais dados seja percebidos pelo leitor/usuário como informação. Nesse processo, a complementariedade entre a informação visual e a textual desempenha importante papel. Assim, partindo-se da imagem chega-se à síntese verbal que a descreve, ao mesmo tempo que tal sinopse é corroborada por essa mesma imagem, resultando em uma aderência forte entre o signo visual e seu significado expresso tipograficamente, condição que tende a reforçar sua compreensão e contribuir para a retenção da mensagem.

Elementos de destaque: A Figura 45 (p.136) exemplifica o uso de elementos visuais

Figura 44 – Aspectos da gráfica de comunicação



Fonte: autor

explicativos que podem ser, tanto lembretes ao pesquisador, como destaques sobre aspectos importantes que deve ser comunicado aos destinatários do dashboard. Neste caso, esses recursos são empregados para evidenciar a resposta a algumas das hipóteses formuladas na tabela de pertinência: [H1], [H3], [H4]. A primeira delas questiona sobre a correlação entre os CRs, as integralizações longas e as taxas de reprovação. Os destaques em 43A envolvem esses componentes. Por meio das setas sugere-se uma provável conexão entre a involução nos coeficientes de rendimento (tanto dos alunos em geral, quanto do aluno em turma) e o aumento das integralizações mais longas (níveis 3 e 4) e das reprovações (tanto por frequência, quanto por nota). Nesse sentido, parece de fato existir uma correlação, pois, se não fosse assim, a redução no CR ocorreria sem alteração das outras taxas.

Figura 45 – Gráfica de comunicação: destaques visuais



Fonte: autor

A Figura 43B procura responder à terceira hipótese que questionava sobre a relação entre o total de ingressantes da universidade e o padrão de ingresso de cada Centro (estrutura acadêmica). Aventava ainda a possibilidade dos Centros reproduzirem o comportamento da universidade. As marcações na figura 43B sugerem que a relação é mais complexa. Apenas o CTC encontra-se no grupo das alternâncias, a que pertence o total de ingressantes (assinalado em azul). Os Centros Biomédico e de Educação e Humanidades mostram um aumento consistente no número de ingressantes, contrário, portanto, ao Centro de Ciências Sociais, que apresenta queda de inscritos. Nesse sentido, as marcações no gráfico questionam a possibilidade de uma “migração” dos ingressantes do CCS para os outros centros.

Já a hipótese [H4] diz respeito à relação entre os candidatos inscritos no vestibular cujo Ensino Médio foi cursado em rede pública e a quantidade de ingressantes da rede pública. Questionava se um aumento ou redução no número de vestibulandos originários de tal rede implicaria em uma variação análoga no número de ingressantes com a mesma origem. A alternativa a essa hipótese seria que a quantidade desses ingressantes fosse independente do número de vestibulandos, situação que sugeriria que o exame não estaria atuando como filtro para o ingresso na universidade. Mas os destaques em 43C assinalam esses dois componentes, nos quais se percebe uma tendência involutiva em ambos quando se analisa a flutuação do histograma, o que parece confirmar a correlação. Pode-se assinalar ainda que, apesar de tal tendência, a quantidade de ingressantes de escola pública vem se mantendo mais alta do que era, considerando-se a média dos três anos anteriores às cotas, observação denotada pela constância de retângulos marcados em preto após 2003. Por fim, importa ressaltar que os gráficos que dão resposta a tais hipóteses poderiam ser isoladas, caso fosse desejável uma comunicação mais direta de sua informação. Técnicas de análise seletiva são amplamente favorecidas pelos recursos interativos, domínio que será discutido a seguir.

4.4.2 A perspectiva digital: diagramas 2 e 3

A característica essencial da metodologia de Bertin como uma *gráfica móvel* faz com que seja natural potencializar tal propriedade por meio dos recursos interativos. Nesse sentido, há diversos tópicos a serem estudados. Um primeiro ponto (já abordado pelo próprio Bertin) diz

respeito à possibilidade de automatizar as permutações por meio de recursos computacionais que agrupem os gráficos de acordo com os critérios determinados. É desejável que esses algoritmos não resultem em um processo de automação completo com um único resultado, sobretudo considerando-se os aspectos subjetivos presentes na reorganização dos dados, comentados anteriormente. Sendo assim, a interface ideal seria aquela que: facultasse ao usuário o critério do agrupamento (semelhança, tendência, ordem etc.); executasse a operação de permutação segundo tal critério; e permitisse manipulações adicionais viabilizando ajustes finos. Nesse âmbito, será tratado adiante a questão das diferentes possibilidades de permutação aplicáveis a um mesmo conjunto de gráficos, mas esta dissertação não lidará com os aspectos de programação envolvidos em uma manipulação calculada pelo computador. Ao invés disso, alguns recursos digitais mais diretamente relacionados à análise e à visualização serão discutidos, como a combinação entre indicadores para a síntese de novas métricas e a alternância entre visualizações complementares.

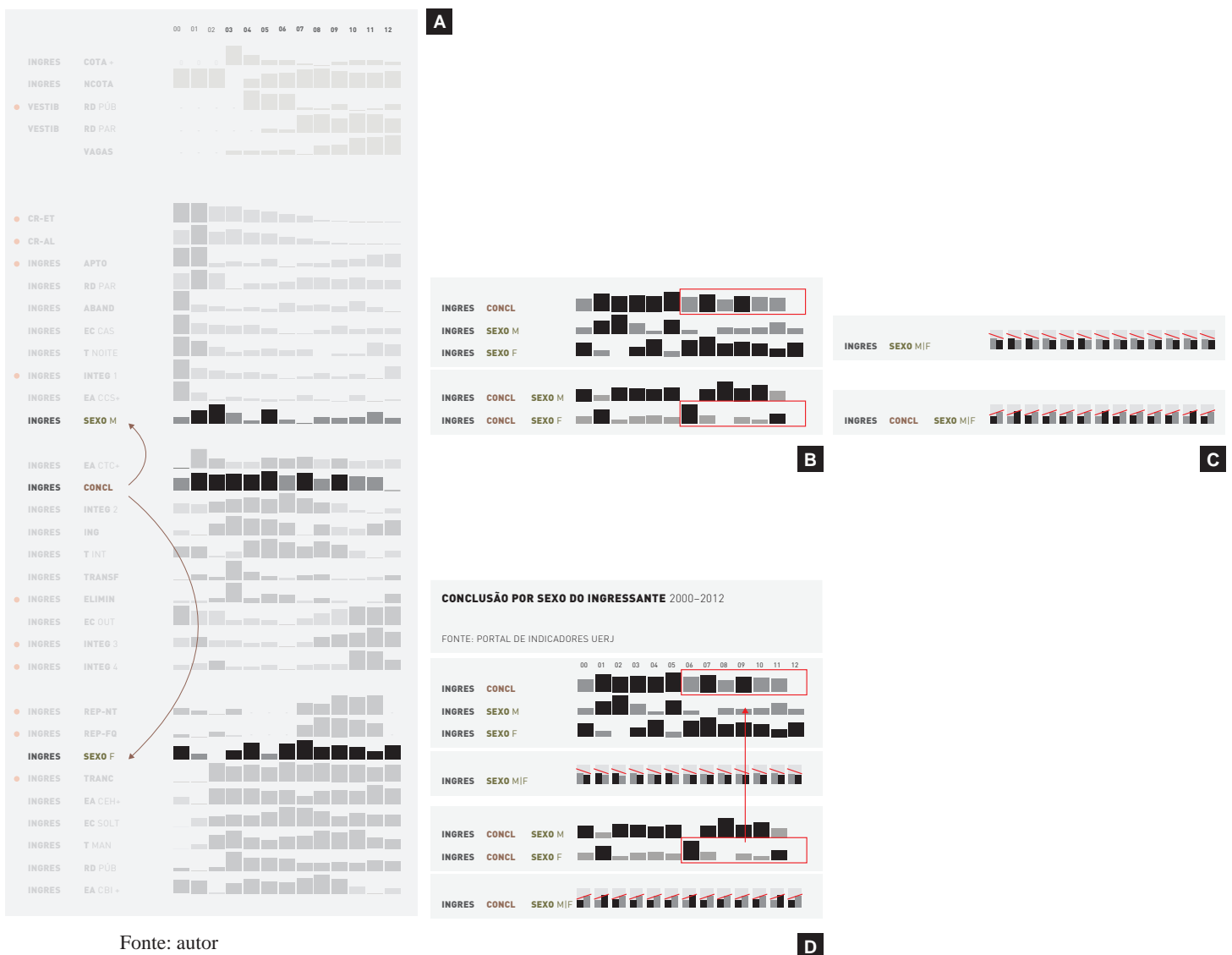
Baseando-se nesses dois tópicos, são esboçadas algumas indicações que resultariam em uma ferramenta de análise baseada na dinâmica de formulação e reformulação de gráficos. Disso, resulta um contínuo processo de pesquisa no qual, partindo-se de um dashboard básico que circunscreve o tema tratado, pode-se desenvolvê-lo combinando seus componentes e detalhando pontos de interesse. Essas operações resultam em gráficos adicionais, atendendo à consultas específicas que, por sua vez, podem ser consolidados em novos dashboards, sintetizados a partir dos elementos de interesse descobertos ao longo dessas pesquisas.

4.4.2.1 *Drill down*

Recurso presente na maior parte dos dashboards digitais, refere-se à possibilidade de “mergulhar” nos dados, aumentando o nível de detalhamento em aspectos específicos que demandam maior atenção. Trata-se, então, de um segundo momento de análise, possibilitado sobretudo pelas potencialidades interativas, que se destinam a investigações mais profundas a partir das informações gerais apresentadas no dashboard. Para exemplificar esse procedimento é inevitável aventurar-se em esboços de uma interface que viabilize tais recursos. Além disso, deve-se ter em conta que as possibilidades de *drill-down* são

inúmeras e estão associadas não só às especificidades técnicas do *software* do dashboard (e da linguagem de programação em que é baseado), mas também ao tipo de detalhamento necessário em determinada pesquisa. Nesse sentido, as figuras seguintes buscam dar uma ideia de alguns aprofundamentos prováveis, tanto em relação à interatividade quanto à lógica de pesquisa encaminhada até aqui, embora não se pretenda esgotar as possibilidades de análise. A principal característica de *drill-down* proposta aqui é a combinação entre *componentes* de modo a sintetizar novas métricas a partir das existentes. Essa é uma operação já possibilitada pelas tabelas dinâmicas disponibilizadas pelo sistema de BI da UERJ, entretanto o que se propõe aqui é uma abordagem mais intuitiva e visual, baseada na manipulação direta dos gráficos e não pelo acesso a menus, como acontece no *software Excel*. Para isso, a Figura 46A sugere como os *labels* dos componentes podem atuar tanto como controles, quanto

Figura 46 – Combinação entre indicadores: sexo e conclusão



como indicadores das operações combinatórias; nessa última situação funcionam um pouco como *breadcrumbs*, indicando a lógica percorrida para se chegar a determinada métrica.

A Figura 46A indica uma síntese das métricas originais [INGRESS] [SEXO] e [INGRES] [CONCL] para se chegar à métrica composta [INGRESS] [CONCL] [SEXO] de modo a pesquisar o percentual de concluintes por sexo. A imagem resultante dessa busca (46B) é encimada pelas métricas originais e, a seguir, apresentam-se os índices derivados. Nesse caso, pode-se inferir que a queda recente no número de mulheres concluintes em relação à média 2000-2002 correlaciona-se à queda no total geral de concluintes. A Figura 46C mostra, em escala absoluta, uma outra vista dos mesmos dados. Sendo assim, tem-se um comparativo em termos absolutos, tanto do número de ingressantes, quanto de concluintes por sexo. Visualizados lado a lado, fica clara a relação de oposição existente (destacada por fios indicando as tendências ao longo dos anos): o número de ingressantes homens é consistentemente maior do que mulheres, ao passo que a predominância nas conclusões é do sexo feminino. Essa informação contextualiza aquela sintetizada em 46B de forma que ambas, podem ser agrupadas em um mesmo conjunto diagramático, como demonstra a Figura 46D.

As Figuras 47A e B apresentam uma pesquisa semelhante à anterior, mas analisando o percentual de concluintes por rede de ensino de proveniência do ingressante (pública ou privada) [INGRES] [CONCL] [RD]. Analogamente ao que acontece na pesquisa sobre o gênero, a redução recente nos concluintes parece ter relação com a redução também recente nas conclusões dos alunos egressos da rede pública. Além disso, essa mesma redução relaciona-se à ampliação na quantidade de alunos egressos da rede pública a partir das cotas (2003) e que estariam concluindo seus cursos nos anos recentes. Sendo assim, o esperado seria um aumento equivalente nessas conclusões. Mas isso não acontece, condição demarcada no gráfico pela seta. Esta situação parece ser corroborada pelo aumento das repetências e integralizações mais longas, já apontado no painel principal. Ainda de forma semelhante à pesquisa sobre gênero, a Figura 47C apresenta comparativo em termos absolutos entre ingressantes e concluintes das duas redes de ensino. Observa-se assim, que a tendência histórica à maioria de estudantes vindos de escolas particulares foi modificada em 2003, condição que permaneceu até 2005, quando se retorna à predominância da rede particular, embora em menor proporção. Em relação aos concluintes, vê-se que a maior parte ainda pertence ao grupo que egressou da rede particular, embora a diferença venha diminuindo ao longo do tempo. Em 2012, as conclusões da rede pública predominaram.

Figura 47 – Combinação entre indicadores: rede de ensino e conclusão



Fonte: autor

Na figura 48A, combinam-se a rede de ensino de origem do ingressante [INGRES] [RD] à estrutura acadêmica [INGRES] [EA] para chegar ao percentual de alunos egressos da rede particular e pública de ensino em cada Centro da UERJ. Como nas análises anteriores, as métricas originais encimam as derivadas, organização vista em 48B, que apresenta os Centros organizados alfabeticamente. A permutação desses fatores, mostrada em 48C, sugere um padrão interessante. Parece haver uma correlação entre o aumento de ingressantes da rede pública (conforme visível na métrica original) e a ampliação do mesmo tipo de ingressante nos Centros Biomédico, de Ciências Sociais, e de Tecnologia e Ciências, de modo que tais instituições confirmam o padrão geral. Por outro lado, o Centro de Educação

Figura 48 Combinação entre indicadores: rede de ensino e conclusão



Fonte: autor

e Humanidades tem uma característica anômala em relação aos demais, apresentando aumento dos alunos das escolas públicas nos primeiros anos das cotas e, desde 2007, uma redução desses ingressantes.

Nesses exemplos, foram sintetizados novos gráficos a partir do diagrama original (diagrama 1). Esse processo pode servir a dois propósitos principais: tanto à visualização isolada de uma especificidade, como é comum em situações de *drill-down*; quanto à elaboração de novos dashboards a partir dos resultados obtidos. Além da hipótese de uso dos *labels* para a síntese de novos indicadores, essa é a segunda proposta para se avançar

em aspectos da interface. Nesse caso, os detalhamentos pesquisados atuam como módulos que, justapostos, podem dar origem a um dashboard autônomo. A Figura 49 ilustra essa possibilidade reunindo, além dos detalhamentos acima, o gráfico sobre o ingressantes por tipo de cota, constituindo assim um novo dashboard sobre os ingressantes e conclusões por gênero e rede de ensino.

4.4.2.2 Diagrama 2: comparação entre cotistas e não cotistas

A figura 50 apresenta o processo de síntese de um novo diagrama a partir da combinação dos componentes presentes no diagrama 1 do dashboard básico. Disso, resulta o segundo diagrama daquele mesmo dashboard, que visualiza uma comparação entre ingressantes cotistas e não cotistas, sobretudo em relação a características de desempenho acadêmico, visando identificar situações de equilíbrio e/ou a predominância de um ou outro grupo. Com isso, procura-se dar resposta às hipóteses [H2] e [H6] definidas na tabela de pertinências, além de outros

Figura 49 – Dashboard derivado 1

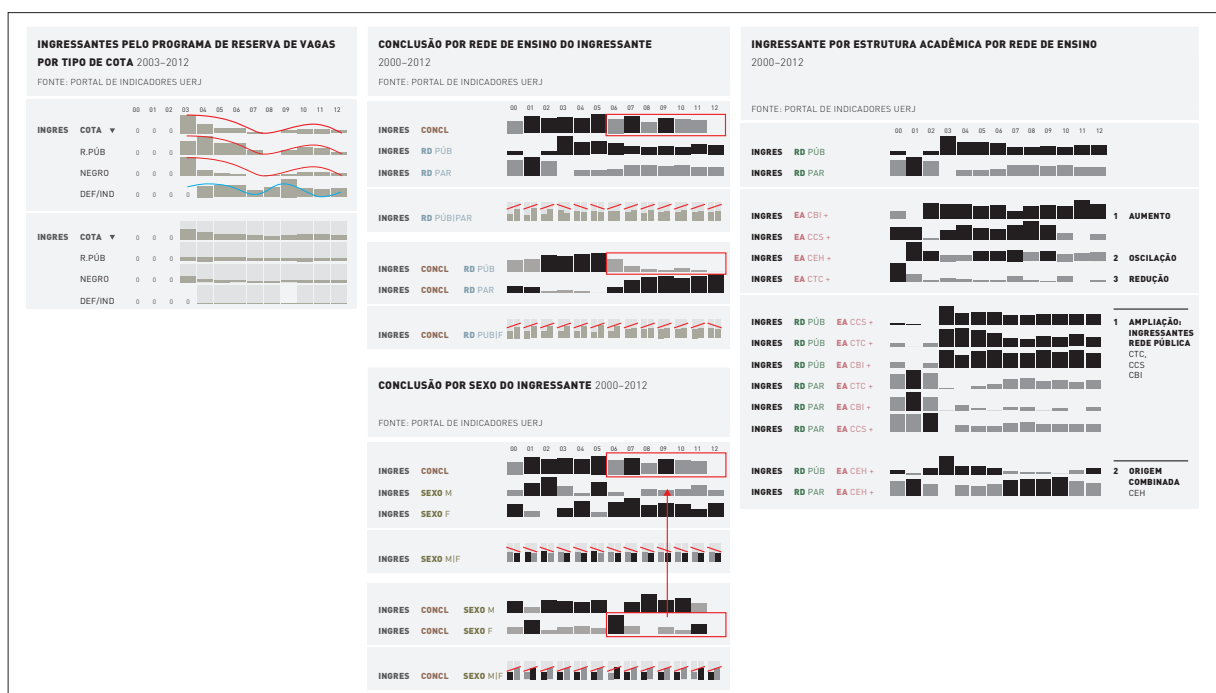


Figura 50 – Síntese do diagrama 2 a partir de componentes do diagrama 1



Fonte: autor

questionamentos análogos. Para tanto, associam-se [INGRESS] a [COTA] para determinar o universo de ingressantes cotistas. A seguir, acrescentam-se os componentes ligados ao desempenho para estabelecer os índices específicos:

[INGRES] [COTA] [APTO] (alunos não eliminados no sistema e aptos para a inscrição em disciplinas)

[INGRES] [COTA] [CONCL] (para analisar a taxa de conclusão);

[CONCL] [COTA] [INTEG] (considerando-se o tempo para integralizar o curso);

[INGRES] [COTA] [CR-AL] e [INGRESS] [COTA] [CR-ET] (para coeficientes de rendimento);

[INGRES] [COTA] [ELIM] (eliminações);
 [INGRES] [COTA] [ABAND] (abandonos);
 [INGRES] [COTA] [TRANC] (trancamentos);
 [INGRES] [COTA] [TRANSF] (transferências internas e externas).

O fator de comparação empregado é o universo dos ingressantes não cotistas [INGRES] [NCOTA], de tal forma que os casos em que os cotistas têm um percentual maior, a ocorrência é marcada em preto. No caso de percentual menor, o item é marcado em cinza. A escala empregada foi a de expansão simples, para manter a visibilidade dos itens com menor percentual e assegurar que ocorra a comparação em todos os anos, o que não aconteceria na expansão pelo mínimo, na qual o menor elemento é suprimido.

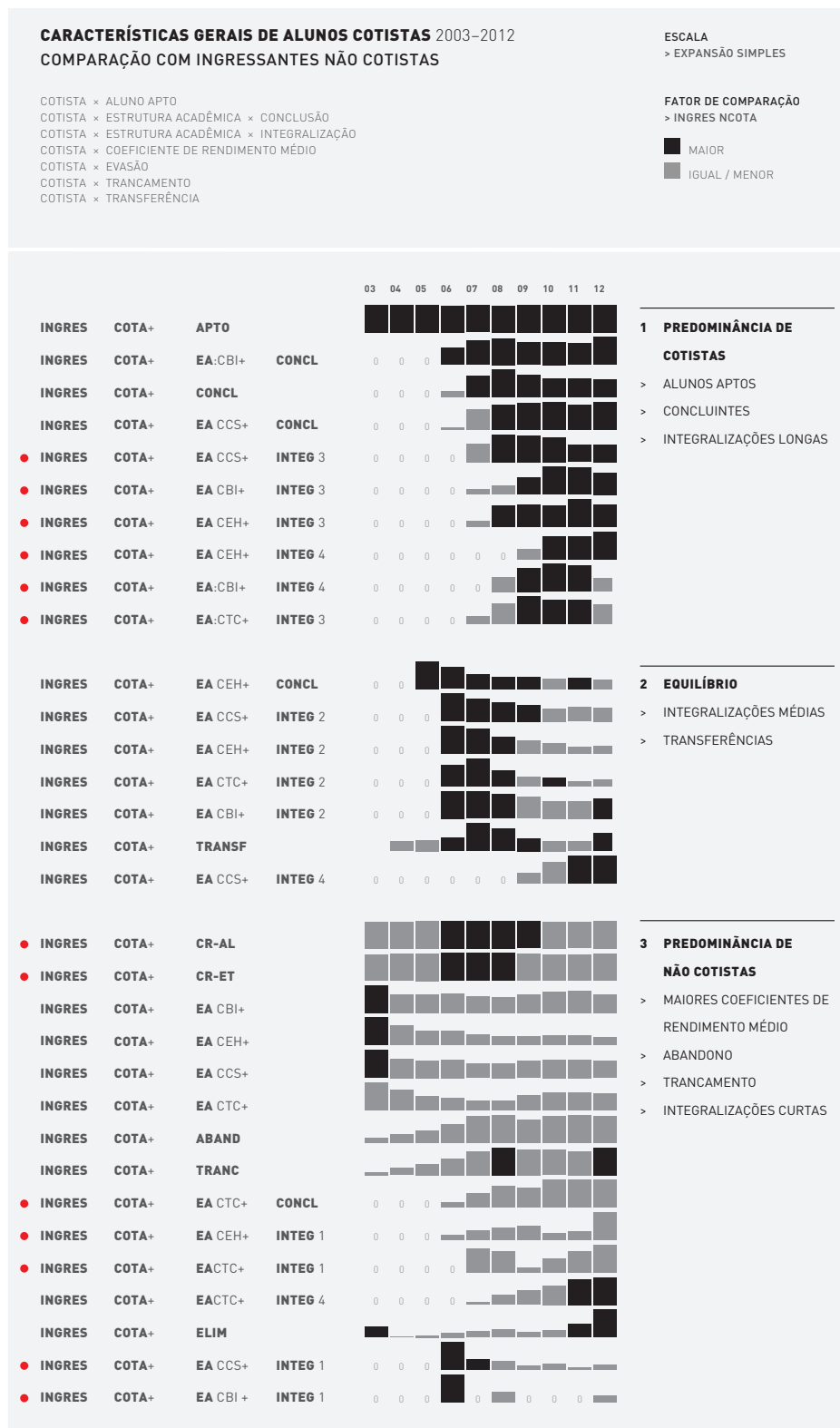
Baseado nesse critério de comparação, acontece a permutação e a comunicação, resultando no gráfico da Figura 51, que foi permutado para evidenciar três grupos característicos:

- 1) componentes com predominância de cotistas (sinalizados pela maior concentração de barras negras).
- 2) alternância entre predominância de cotistas e não cotistas, isto é, o padrão que seria o esperado em um situação de equilíbrio.
- 3) predominância consistente de não cotistas (maior ocorrência de barras cinza).

Esse procedimento evidencia mais algumas respostas para as hipóteses da tabela de pertinência. A hipótese [H6] questionava sobre uma taxa de conclusão mais alta em favor dos cotistas. A observação do gráfico confirma essa hipótese, evidenciando que, a partir de 2007, os cotistas da universidade vêm concluindo mais do que os não cotistas. Pode-se notar ainda que esse padrão espelha as taxas de conclusão nos Centros Biomédico e de Ciências Sociais, incluídos no grupo em que há consistência na predominância de cotistas. O Centro de Educação e Humanidades também tende a confirmar a predominância de cotistas nessa categoria, ainda que havendo ocasiões em que isso não aconteceu. Por outro lado, o Centro de Tecnologia e Ciências revela-se uma exceção [INGRES] [COTA] [EA CTC], observando-se que, a despeito do seu histograma mostrar um padrão de ascensão recente nas conclusões de cotistas, o percentual delas vem sendo consistentemente inferior aos não cotistas.

Nesse sentido, comparações semelhantes podem ser feitas analisando-se outras categorias, como o índice de abandono, de alunos aptos, trancamentos, constituindo um panorama mais abrangente em relação ao desempenho dos cotistas frente aos não cotistas. A dimensão mais evidente para isso é a análise do coeficiente de rendimento. A hipótese [H2]

Figura 51 – Diagrama 2: comparação entre cotistas e não cotistas



Fonte: autor

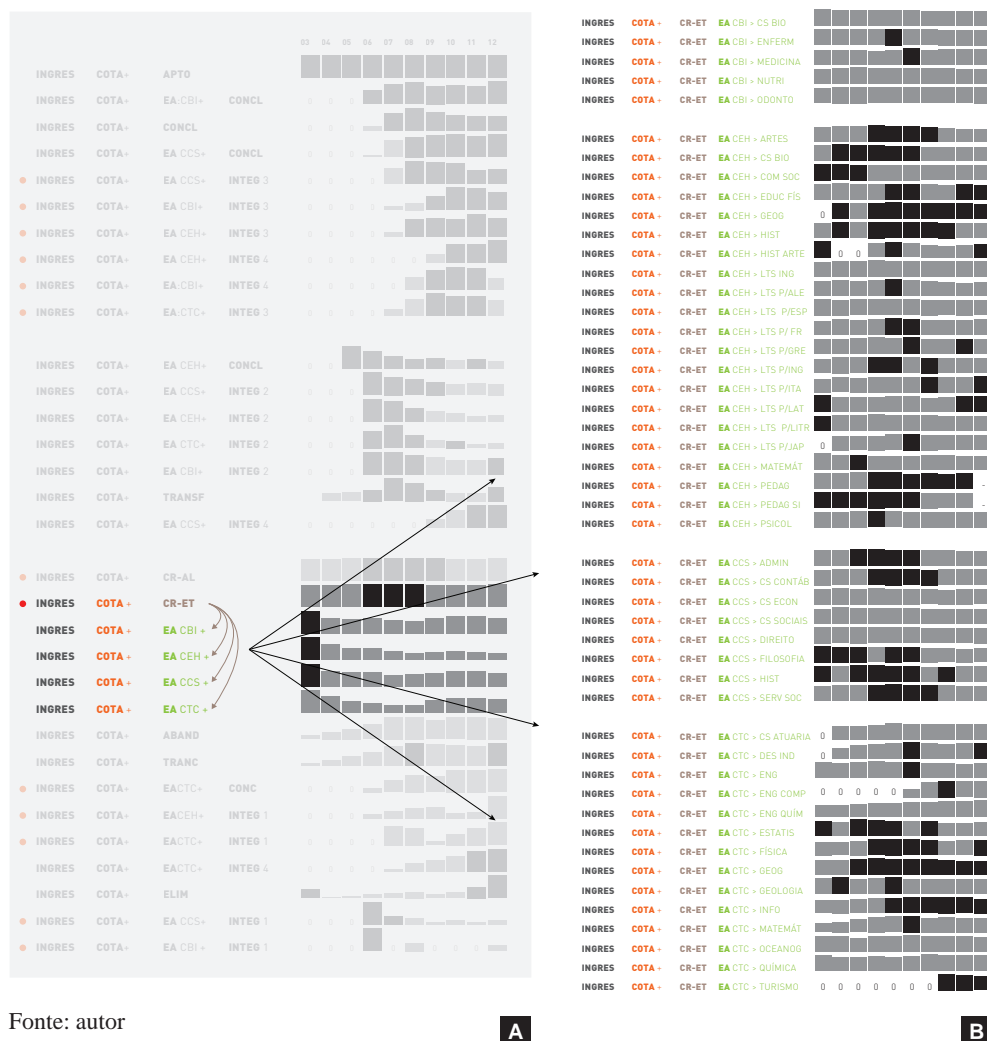
questionava sobre uma suposta defasagem dos cotistas, em todas as estrutura acadêmicas. Observando o estado atual do diagrama 2, percebe-se que as duas modalidades de CR utilizadas ([CR-ET] e [CR-AL]) parecem confirmar tal suspeita no que se refere aos coeficientes de todos os ingressantes. Entretanto, nos anos intermediários da séries concentram-se taxas favoráveis aos cotistas, revelando que não se trata de uma defasagem consistente ao longo de todo o tempo considerado. Nesse caso, torna-se interessante uma investigação mais detalhada visando identificar o comportamento tanto das estruturas acadêmicas, quanto dos cursos que as constituem.

4.4.2.1.1 Análise

Na Figura 52A são combinados os componentes [INGRES] [COTA] a [EA] e [CR-ET], obtendo-se o coeficiente médio do cotista em turma para cada estrutura acadêmica cujo *label*, quando expandido revelaria (em uma interface digital) os respectivos cursos de cada centro. Isso resulta na Figura 52B, em que os cursos são listados por Centro, em ordem alfabética, mantendo-se a escala de expansão simples. Com isso, pode-se dar uma resposta mais completa à hipótese [H2] . Vê-se que em relação às estruturas a resposta é positiva, mas salvo no caso do Centro Biomédico, em todos os demais há ocorrências de cursos com predominância de CR para os alunos cotistas.

A regularidade nos CRs ao longo dos anos sugere que seria útil ressaltar a diferenciação entre os itens para melhor discerni-los visualmente. Além disso, a ordenação alfabética é um agrupamento arbitrário, que não se relaciona às especificidades de um conjunto de dados em particular. Ao passo que a reorganização baseada em critérios “internos” (permutação) pode acontecer segundo diversas intenções comunicativas, de modo que o agrupamento em razão de um fator de comparação pré-estabelecido (como acontece nos diagramas 1 e 2) é somente um deles. As figuras seguintes exemplificam a complementariedade entre três formas de reorganização baseadas cada qual em um critério.

Figura 52 – Processo de síntese dos indicadores de CR por curso



A Figura 53A agrupa os itens por predominância entre cotistas e não cotistas, chegando-se a três grupos, cada um contendo quantidades diferentes de cursos, organização que já revela uma relevante informação na leitura do conjunto. Há um grupo pequeno no qual os cotistas têm CRs maiores, composto basicamente pelas ciências humanas, informática e pelo curso de turismo. O grupo intermediário, pouco maior que o anterior, reúne os cursos em que há alternância entre cotistas e não cotistas, caso de muitos cursos do CCS. Por fim, o grupo que concentra maior quantidade de cursos é aquele com CRs mais altos para não cotistas, composto de cursos diversos, sobretudo aqueles de alta procura e os de ciências biomédicas. Assim, o detalhamento nos cursos contextualiza o padrão visto no CR da totalidade dos ingressantes ([INGRES] [COTA] [CR-ET]) revelando ainda que a predominância de cotistas nos anos intermediários da série, notada naquela

Figura 53 – CR por cursos: agrupamento por predominância entre cotistas e não cotistas

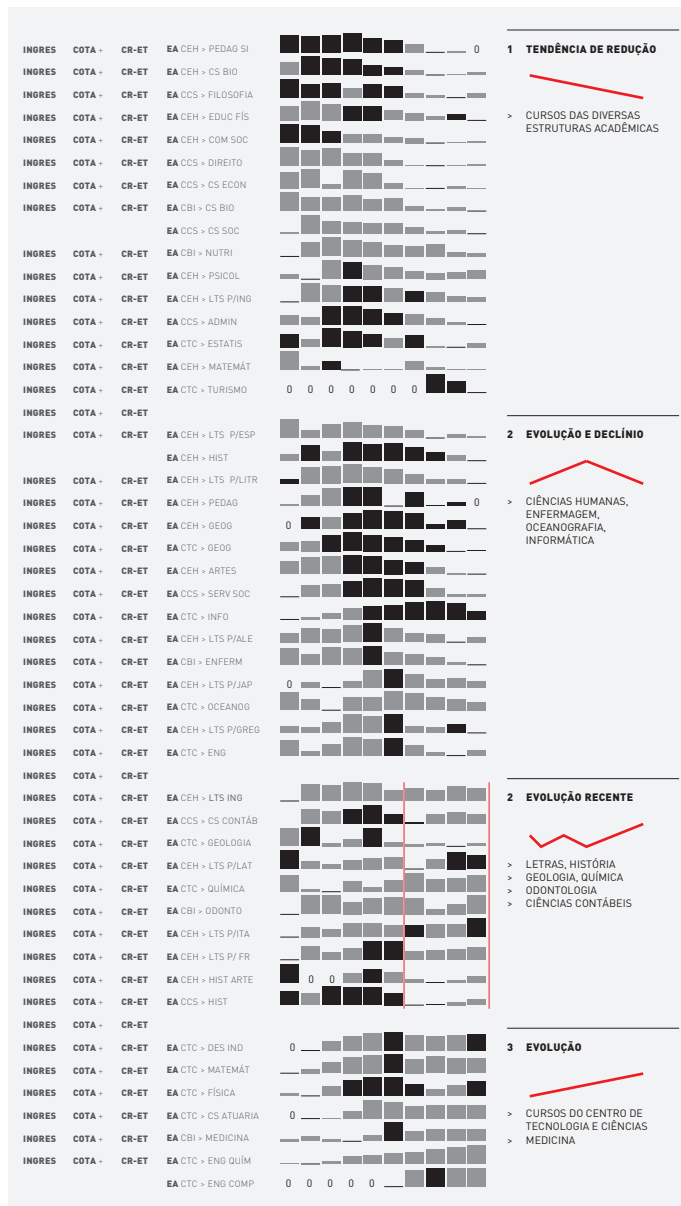


Fonte: autor

métrica, tende a repetir-se nos cursos. Isso é especialmente claro no grupo 2, e sobretudo no grupo 3 da Figura 53A, no qual os anos 2007 e 2008 concentram as ocorrências de CR mais alto para os cotistas, evidenciadas pela alinhamento central de barras pretas.

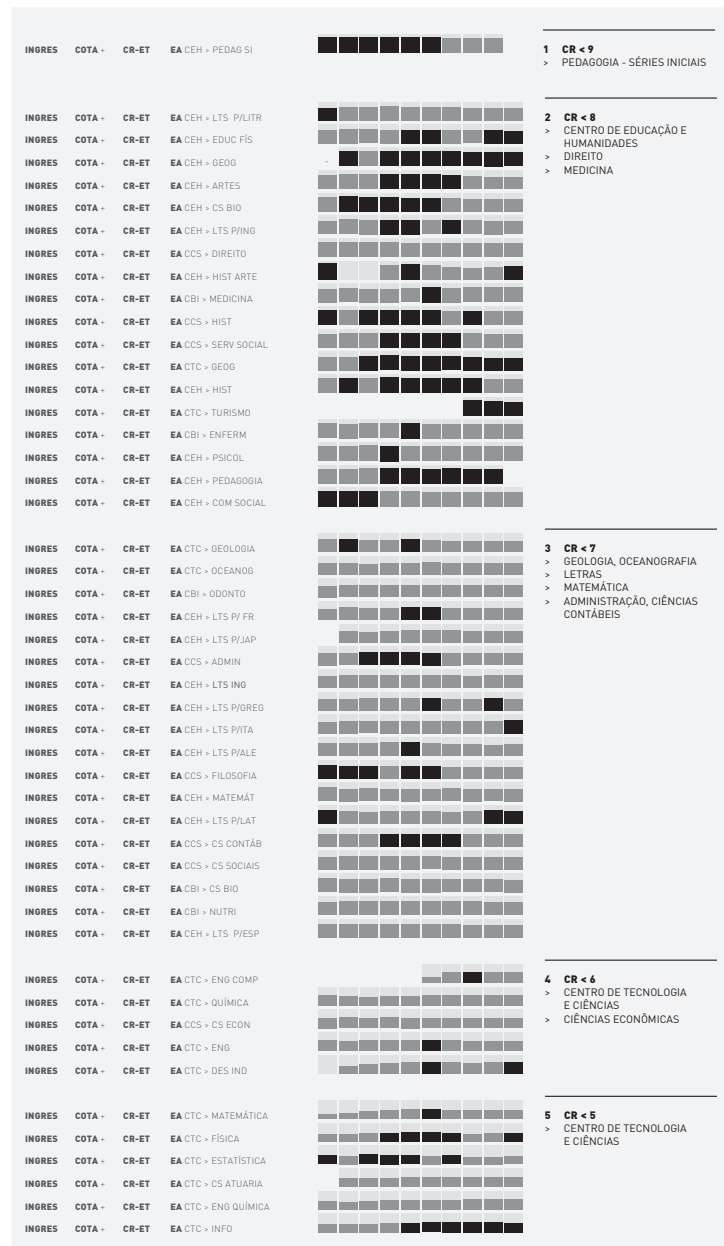
Para complementar a resposta à hipótese [H2], a Figura 53B diferencia os Centros por cor: CTC: vermelho; CBI: verde; CCS: laranja; CEH: ciano. Desse modo, pela observação dos grupos 1 (predominância de cotistas) e 2 (alternância) sabe-se que eles são compostos por cursos

Figura 54 – CR por cursos:
agrupamento por tendência evolutiva



Fonte: autor

Figura 55 – CR por cursos:
agrupamento por classes de CR



do CEH, CTC e CCS. Esse fato já permite responder negativamente à hipótese formulada, com a ressalva de que o único Centro em que as notas dos cotistas vem sendo consistentemente mais baixas do que as do não cotistas é o Centro Biomédico (CBI).

Por uma outra perspectiva, a Figura 54 rearranja as linhas segundo a tendência apresentada pelo CR do cotistas em cada curso. Assim, divisam-se quatro grupos: aquele com tendência à redução; um segundo grupo em que há uma evolução nos primeiros anos e um declínio recente; o terceiro grupo, com comportamento irregular nos primeiros anos e tendência

recente à evolução, e o último grupo que apresenta uma tendência de aumento consistente. Nesta forma de permutação (tal como na seguinte), o fator de comparação assume uma propriedade contextual, e não mais protagonista. A concentração dos itens em preto indica que em muitos dos cursos em que os CRs mais altos são dos cotistas, há uma tendência de diminuição do CR desse mesmo grupo.

Na figura 55 (p.149), a permutação acontece em relação às classes de CR baseadas na média de cada curso entre os anos 2003 e 2012, dado que aponta naturalmente para o uso da escala absoluta habilitando uma comparação direta entre as notas. Sendo assim, são divisadas cinco classes de nota, vistas da maior para a menor. Assim, o curso com média mais alta [PEDAG SI] (pedagogia – séries iniciais, média 8,06) é o único membro da classe que varia de 8,99 a 8,0. O curso [ENFERM] (enfermagem), cuja média é 7,76, pertence ao grupo de CR entre 7,99 e 7,0, e da mesma forma para os demais cursos. Neste diagrama, observa-se que os dois grupos que concentram a maior parte dos cursos variam seus coeficientes de rendimento entre 7,99 a 6,0. Eles são também aqueles em há a maior parte das ocorrências do CRs mais altos dos cotistas (em preto), em especial no grupo [CR<8], vantagem que também acontece no curso [PEDAG-SI], evidenciando assim que a predominância dos cotistas está distribuída pelos três grupos de maior CR. Isso sugere que os cursos em que há predominância de cotistas tendem a ser os mesmos cursos em que os cotistas têm melhor desempenho.

A figura 56 reúne os três diagramas anteriores constituindo um dashboard autônomo, com dados sobre o coeficiente de rendimento dos diversos cursos. A complementariedade entre as informações de cada diagrama pode ser exemplificada pela observação do comportamento de um mesmo curso nas três variações, como ilustra a imagem. Assim, sabe-se que no curso de geografia (em laranja) do Centro de Educação e Humanidades, os cotistas vem obtendo notas mais altas do que os não cotistas, mas que há uma tendência recente de queda do CR, cuja média entre 2003 e 2012 oscilou entre 7,0 e 7,9 pontos. Já no curso de estatística, do Centro de Tecnologia e Ciências, há uma flutuação na predominância de CRs mais altos entre cotistas e não cotistas, mas observa-se uma tendência de redução nas notas dos primeiros, cuja média histórica é das mais baixas dentre os cursos da universidade, como pode ser observado no terceiro diagrama do dashboard em questão.

Figura 56 – Dashboard derivado 2: três perspectivas sobre o coeficiente de rendimento

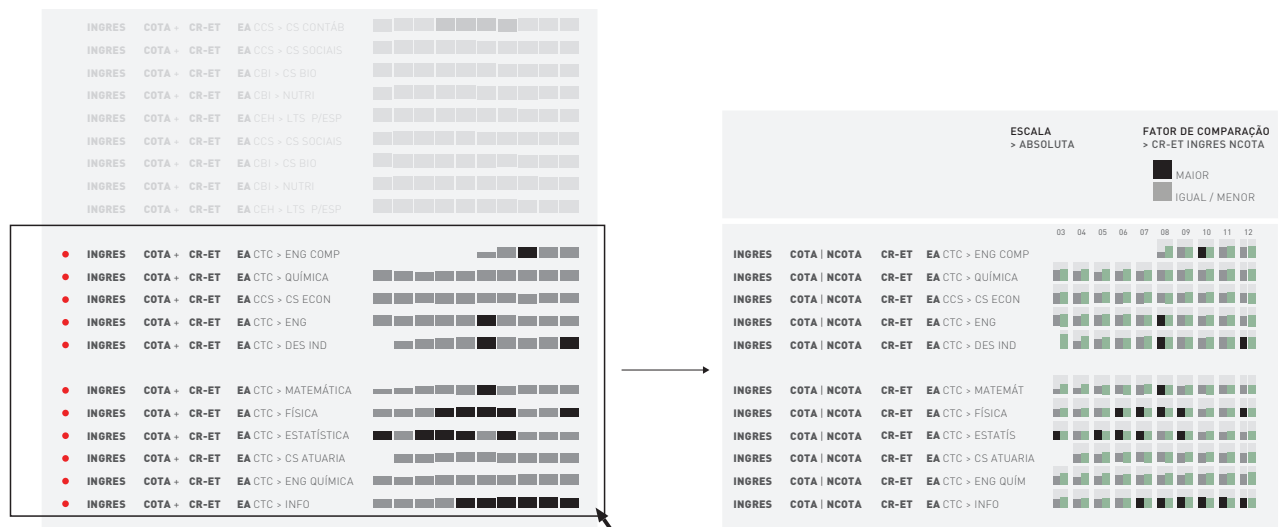


Fonte: autor

4.4.2.1.2 Estudo dos CRs mais baixos

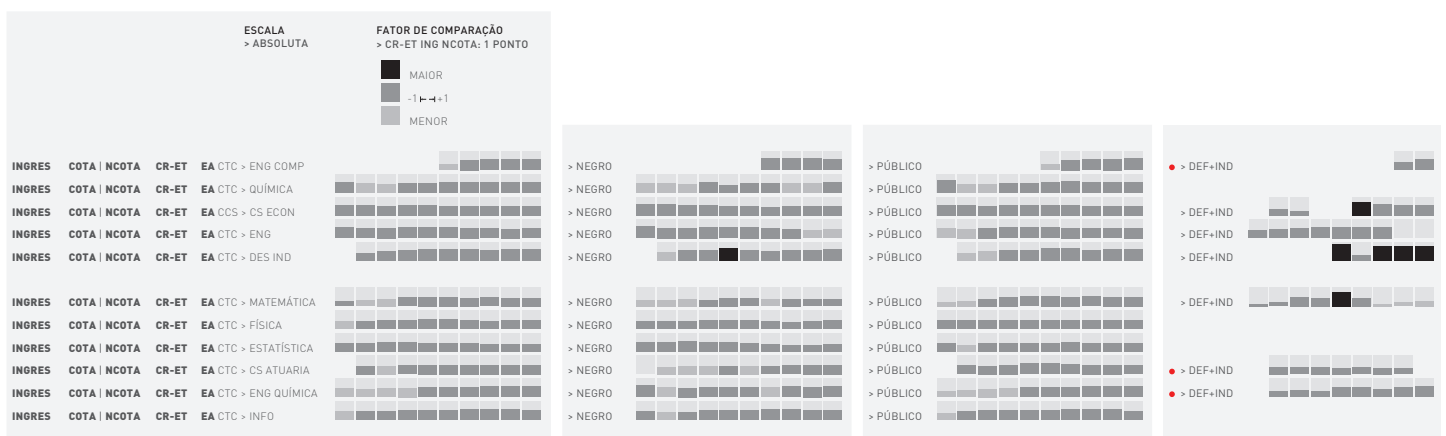
A partir do diagrama que agrupa os CRs por classe, pode-se querer investigar os grupos potencialmente mais problemáticos, aqueles cujo CR encontra-se entre 5,99 a 4,0 e que concentram os cursos do Centro de Tecnologia e Ciências. Essa situação sugere uma análise mais avançada em relação ao grau de diferença que separa os CRs dos cotistas e dos não cotistas. Para isso, pode-se exibir o fator de comparação [INGRES COTA CR-ET] visualizando-se assim os dois tipos de ingressantes em barras justapostas de cores diferentes, como no diagrama à direita da Figura 57. Comparando a altura das barras, vê-se que as diferenças não parecem muito grandes, sobretudo em anos recentes. Sendo assim, refinar a visualização de modo a identificar somente as grandes diferenças seria um desdobramento lógico da análise. Isso acontece na Figura 58,

Figura 57 – Detalhamento – CRs mais baixos



Fonte: autor

Figura 58 – Detalhamento – CRs mais baixos por tipo de cota



Fonte: autor

na qual retorna-se à visualização habitual e define-se como fator de comparação o intervalo de 1 ponto, para mais ou para menos. Desse modo, se o CR dos cotistas supera os do não cotistas por mais de 1 ponto, as ocorrências são vistas em preto. Os casos que se situam no intervalo delimitado de menos 1 até mais 1 ponto de diferença, aparecem em cinza médio. Quando a diferença é desfavorável aos cotistas excedendo esse limite, as ocorrências são marcadas em cinza claro. Isolam-se assim as diferenças mais significativas, e observa-se que nos cursos em que houve diferença desfavorável aos cotistas superior a 1 ponto, tal diferença aconteceu nos

primeiros anos das cotas. Percebe-se também a ausência de diferenças favoráveis aos cotistas superiores a um ponto. Por outro lado, a leitura do conjunto do gráfico delineia com clareza que na grande maioria dos casos a diferença de notas entre cotistas e não cotistas não ultrapassa 1 ponto. Conclui-se, então, que no grupo dos CRs mais baixos, as diferenças de desempenho entre cotistas e não cotistas não parece ser alarmante. Por fim, resta investigar se esse padrão repete-se em todas as modalidades de cota, o acontece pela expansão do componente [COTA]. Chega-se à Figura 58 (p.152), que confirma o padrão geral em relação aos negros e oriundos do ensino público, mas, novamente, mostra uma situação anômala no caso de deficientes e indígenas. Além de não haver representantes em todos os cursos, trata-se do grupo que concentra a maior incidência de notas abaixo e acima do intervalo delimitado.

4.4.2.2 Diagrama 3: ocupação dos cursos por cotistas

O último diagrama que compõe o dashboard básico diz respeito à ocupação dos cursos em relação ao parâmetro regulamentado pela lei ordinária nº 5.346 de 45% de vagas reservadas. Embora a lei date de 2008, no presente estudo essa proporção é tomada como fator de comparação para todos os anos estabelecendo um critério de análise consistente. Para formar o diagrama 3 são combinados os componentes [INGRES] [COTA] a [VAGAS] e [EA] resultando no percentual de cotistas em relação às vagas disponibilizadas por estrutura acadêmica, item que expandido revela seus respectivos cursos. A figura 59 apresenta o diagrama já permutado, em estágio de comunicação. Os histogramas foram agrupados em três grupos: o primeiro reúne as ocorrências de ocupação máxima, que acontece principalmente nos cursos de alta procura; há um grupo intermediário no qual a taxa de ocupação mostra alternância, embora as maiores ocupações tendam a se concentrar nos primeiros e nos últimos anos da série; por fim, há o maior grupo, que reúne os cursos nos quais sobram vagas, pois a ocupação fica abaixo da estabelecida.

Figura 59 – Diagrama 3: ocupação dos cursos por cotistas



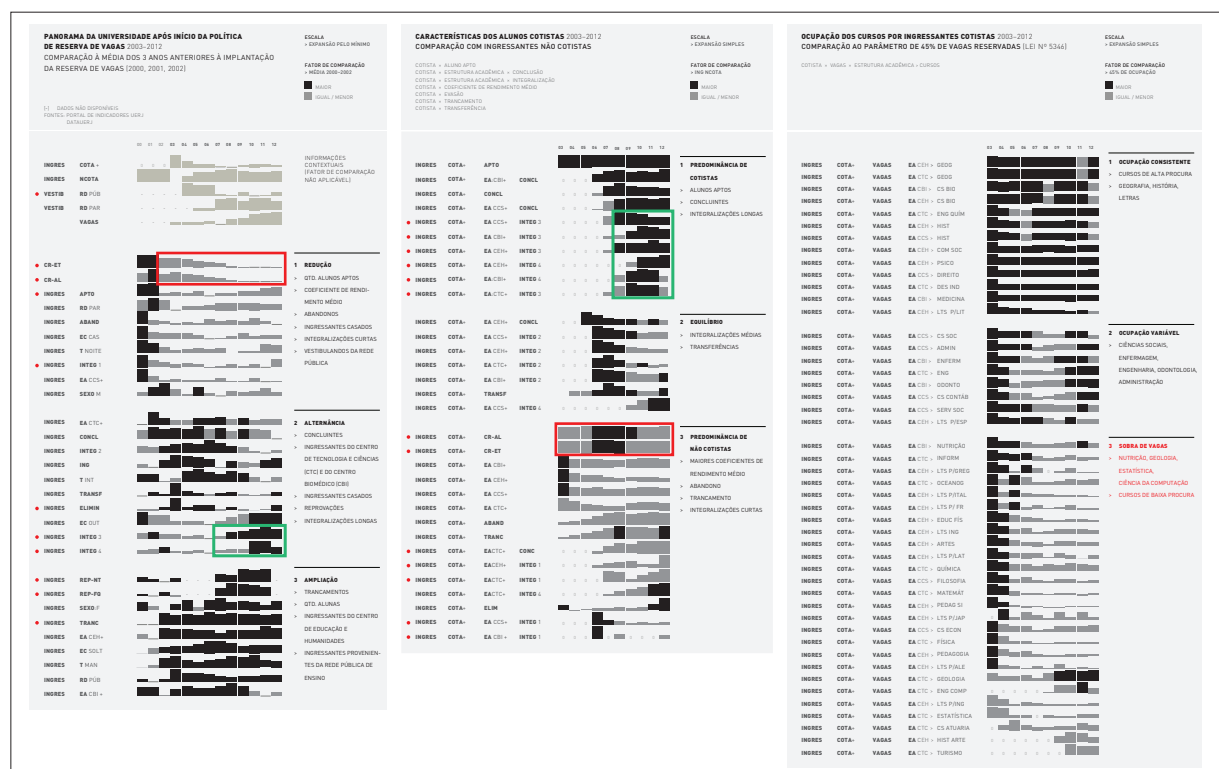
4.4.2.3 Síntese

Os três diagramas reunidos apresentam uma visão geral de questões ligadas ao sistema de cotas. Nesse sentido, eles constituem um dashboard que atua como sinopse do tema em estudo e índice, apontando para os desdobramentos possíveis. Além disso, a vista dos três diagramas permite comparações entre eles, como na Figura 60.

1 Os destaques na figura referem-se a tópicos de desempenho relacionando comportamentos de ordem geral (no diagrama 1) às ocorrências específicas dos cotistas. Assim, a redução no CR total parece repercutir os CRs mais baixos dos cotistas (diagrama 2), como sugerido pelos contornos em vermelho. Analogamente, o aumento das integralizações longas anotadas em verde no diagrama 1 relaciona-se à predominância de cotistas em integralizações nível 3 e 4 em todas as estruturas acadêmicas (diagrama 2).

2 Outras correlações podem ser entrevistas na Figura 61. Por exemplo, na comparação entre alunos aptos (diagrama 1) e cotistas, mostrando que a maior parte dos alunos em

Figura 60 – Dashboard básico: estudo de correlações 1



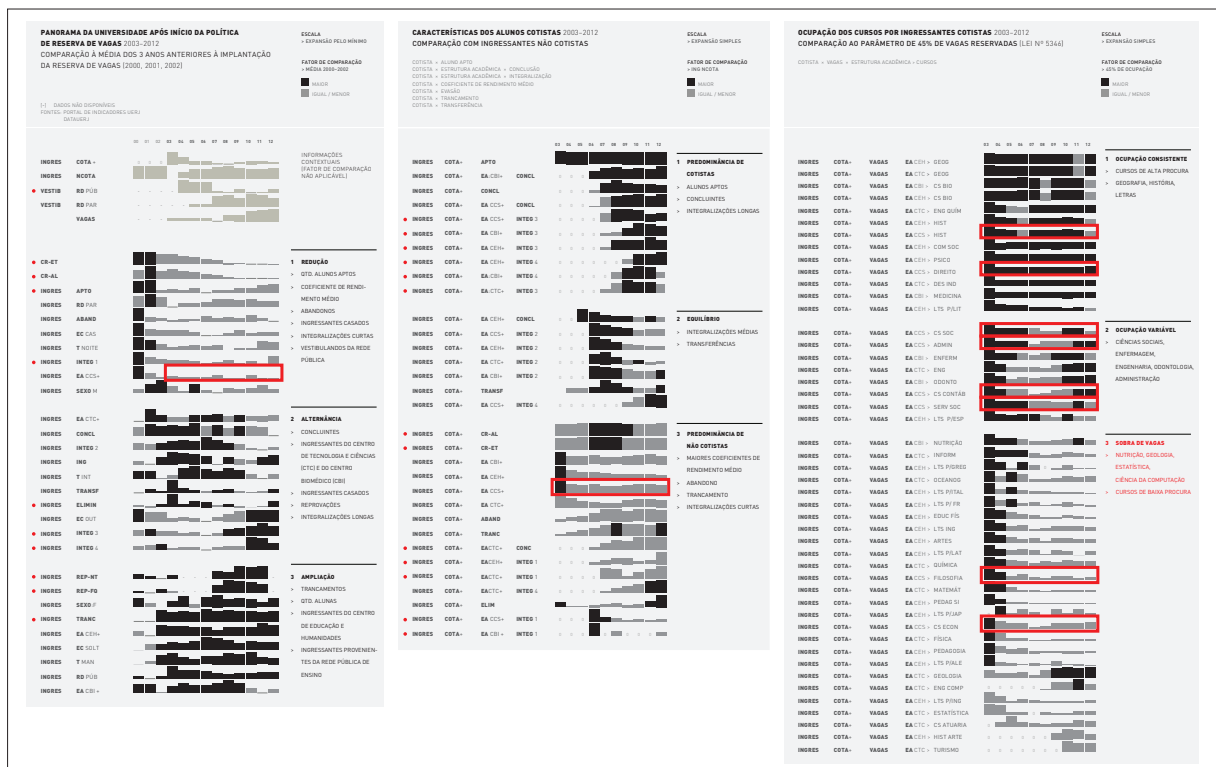
Fonte: autor

Figura 61 – Dashboard básico: estudo de correlações 2



Fonte: autor

Figura 62 – Dashboard básico: estudo de correlações 3



Fonte: autor

condições de fazer inscrição em disciplinas é cotista (diagrama 2). Ou ainda, que apesar da queda nas conclusões (diagrama 1), os cotistas concluem mais (diagrama 2); que a despeito da oscilação geral nas eliminações (diagrama 1), esse percentual é consistentemente menor entre cotistas (diagrama 2). E que, apesar dos índices de trancamento mais altos após a implantação das cotas (diagrama 1), entre os cotistas acontecem menos trancamentos (diagrama 2).

3 A Figura 62 (p.156) sinaliza que a redução de ingressantes no Centro de Ciências Sociais (diagrama 1) pode ser contextualizada em razão da menor participação de cotistas nesta instituição (diagrama 2), como indicam as marcações em vermelho. Mas, além disso, a observação da ocupação por cotistas dos cursos pertencentes ao CCS revela que nem todos têm baixa ocupação e sim que eles estão distribuídos pelas três categorias de ocupação delimitadas no diagrama 3.

Analogamente, inúmeras outras observações poderiam ser feitas e hipóteses levantadas, sobretudo pelos estudiosos do tema. No escopo desta dissertação o que mais importa é indicar as possibilidades analíticas que a associação entre a metodologia de Bertin, os recursos computacionais e as peculiaridades do dashboard poderiam alcançar. Em relação à aparência característica aos dashboards convencionais, a aplicação da teoria de Bertin parece apontar para outra visualidade, mais complexa; ao mesmo tempo que resguarda outros requisitos característicos aos dashboards. Assim, pelo que foi tratado neste trabalho, parece plausível sugerir que, em comparação ao dashboard convencional, o incremento na densidade de dados, como apresentado aqui (e já entrevisto nas propostas de Few e considerado uma virtude segundo Tufte) poderia contribuir para ampliar a potencialidade informativa dos dashboards.

CONCLUSÃO

A hipótese principal desta dissertação concerne à possibilidade de aplicação da metodologia de visualização proposta por Jacques Bertin ao projeto de dashboards. Nesse sentido, foram apresentados os principais aspectos das teorias de Bertin em relação a dados quantitativos e foram descritas as propriedades que usualmente caracterizam dashboards. Vimos que essas interfaces são artefatos de comunicação projetados para o domínio digital, inseridos no contexto das tecnologias de *Business Intelligence*. Caracterizam-se ainda como suporte de comunicações visuais que ocorrem em determinados âmbitos nos quais elas tendem a ser interpretadas como informação. Daí, a possibilidade de, até certo ponto, desvencilhar os aspectos interativos dos visuais, concentrando-se nesses últimos. Foi ressaltado também que tal recorte é sugerido pelo estado atual dos trabalhos sobre dashboards, que concentram-se em questões técnicas e nos recursos interativos, mas, salvo raras exceções, não se estendem aos aspectos de visualização, a despeito de sua importância para o processo comunicativo.

Para um estudo consequente de tais aspectos dos dashboards, assim como da metodologia de Bertin, é necessário a análise de dados reais, que estejam em consonância com as tecnologias de BI e também que sejam suficientemente abrangentes para que se possa discutir os diversos aspectos ligados à interpretação dos dados. Esses dois critérios são atendidos pelo banco de dados da UERJ, projetado com tecnologias de *Business Intelligence*, vincula indicadores estatísticos sobre a instituição. Os estudos encaminhados nesta dissertação a partir daí desenvolveram-se em duas perspectivas complementares: a aplicação da metodologia de Bertin diretamente aos dados coletados evidenciando uma utilização tradicional das técnicas, e a indicação de técnicas interativas buscando investigar algumas possibilidades de expansão da metodologia para o âmbito digital.

Em relação à organização preliminar dos dados a serem visualizados, os conceitos de *tabela única* e *estudo homogêneo* são especialmente relevantes no que concerne à organização e escolha dos dados que irão compor o dashboard. O uso da tabela de alocação, esquema de homogeneidade e tabela de pertinência (os documentos que antecedem à tabela única, base da visualização) permitem uma racionalização do estudo, reduzindo a infinitude de dados constante no *data warehouse* a uma proporção mais facilmente manipulável. Isso acontece não apenas pela identificação e supressão dos dados que não têm relação entre si ou com a população

em questão, como também pela sua seleção em razão das perguntas a que se quer responder. Nesse sentido, tal estudo preliminar sistematiza a composição do dashboard, estabelecendo um direcionamento para a construção dos gráficos. Além disso, nesse processo notou-se que a tabela dinâmica, interface dos dados disponibilizados pela UERJ, não prevê a organização em tabelas sobrepostas e orientadas por um elemento de comparação, técnica que caracteriza a homogeneidade do estudo e a construção da tabela única.

A partir da visualização da tabela, evidenciaram-se algumas questões características aos métodos de Bertin, especialmente no que se refere às escalas utilizadas para transcrever os dados: absolutas ou relativas. A escolha entre elas aponta para questões analíticas e de comunicação, envolvendo o contraponto entre a fidelidade na transcrição dos dados e a evidenciação das diferenças entre eles. Nesse âmbito, as escalas de expansão pela métrica possibilitam exagerar as distinções entre os dados, recurso que pode ser usado tanto do ponto de vista analítico, como para sugestionar o leitor acriticamente. Esses tópicos evocam a tensão entre questões retóricas e a atitude do designer ou do pesquisador na apresentação da informação.

Na verdade, questões ligadas ao propósito da comunicação perpassam todos os aspectos da metodologia de Bertin e evidenciam-se enfaticamente já nas etapas do tratamento gráfico. Assim, vimos que por meio da permutação entre os componentes, uma grande quantidade de categorias é reduzida a poucos grupos característicos que só se revelam após tal manipulação. A definição desses grupos caracteriza o nível mais alto de informação, segundo Bertin. Nesse processo há dois aspectos complementares: um de descoberta (uma vez que não se sabe de antemão quais padrões surgirão) e outro de intencionalidade, já que o critério do agrupamento é, este sim, definido previamente. Importa notar que, da injunção entre esses fatores, pode resultar uma mudança epistemológica no trato com o problema em questão. Isso poderia acontecer quando categorias, definidas previamente (em razão de necessidades na coleta dos dados, por exemplo) transformam-se em novas categorias em razão da reorganização dos dados originais. Tal possibilidade é um dos aspectos característicos da teoria de Bertin que sugere estudos posteriores.

Na etapa seguinte, a *gráfica de comunicação*, as formulações retóricas mostram-se mais presentes, inclusive na evidenciação das novas categorias mencionadas. Nesse caso, a própria escolha do nome de um grupo e o destaque sobre certos componentes considerados mais representativos (na síntese descritiva de cada grupo, que aparece na coluna da direita) envolve a questão retórica do destaque e da seleção. Além disso, a escolha dos tópicos que serão

realçados por elementos gráficos (como os alertas comuns em dashboards convencionais ou outras técnicas de seleção e indicação) obviamente evidenciam um juízo de valor por parte do autor do gráfico ao optar por ressaltar tal informação e não outra. Assim, evidencia-se uma forte característica de edição, mesmo nas comunicações informativas, propriedade que, sobretudo em relação à visualidade, é constituída por uma dialética na qual para algo ganhar relevo, algo deve permanecer plano.

Pesadas tais questões, pode-se responder positivamente à hipótese formulada de início sobre a relação entre dashboards e os trabalhos de Jacques Bertin, no que concerne aos aspectos de visualização e comunicação. Isso acontece sobretudo porque os atributos que caracterizam dashboards, tais como comunicação visual sintética, possibilidade de alertas e detalhamento adicional foram considerados nas propostas apresentadas.

Em relação à perspectiva computacional, a investigação conduzida trabalhou com princípios dos métodos de Jacques Bertin que, mesmo tendo sido desenvolvidos em um cenário pré-digital, sugeriam a possibilidade de transposição para o domínio computacional atendendo aos requisitos básicos dos dashboards, expandindo ainda suas propriedades informativas. Assim, em virtude dos dashboards estáticos apresentados nesta dissertação, evidentemente essa hipótese não pode ter uma resposta definitiva, o que demandaria a elaboração de um protótipo. Entretanto, o que foi apresentado parece sugerir, que alguns recursos interativos potencializam o tipo de análise proposta por Bertin em questões relevantes, tais como a combinação entre os dados, sua análise seletiva e ainda a alternância entre as várias formas que eles podem assumir. Assim as indicações de interatividade apresentadas concentraram-se nessas possibilidades.

Sobre a relação entre os dados, as visualizações procuraram indicar como os componentes podem ser combinados entre si, sintetizando novos indicadores. Por meio desse recurso, a análise dos dados acontece pela manipulação direta dos histogramas (perfis), viabilizando uma pesquisa visual, em que a morfologia específica do histograma de certo componente (ou de um grupo deles) pode sugerir e incentivar pesquisas adicionais, o que é dificultado com dados apenas numéricos, como acontece nas tabelas dinâmicas. Quanto à análise seletiva, trata-se de um recurso que possibilita a seleção de certos grupos de dados, visualizando-os de forma isolada, constituindo um útil instrumento de pesquisa em meio aos três níveis de informação. Entretanto, a característica interativa que norteou o estudo desenvolvido foram as possibilidades de alternar entre as diversas formas que uma construção normal pode assumir. Essa flexibilidade propiciada pelo computador é uma característica importante em razão da complementariedade

informativa entre tais formas. E também porque evidencia, como o próprio Bertin colocou, a característica provisória da gráfica móvel, que não têm uma forma fixa e sim várias possíveis, cuja escolha depende essencialmente da especificidade da pesquisa e de sua comunicação. Assim, o computador permite a alternância entre escalas, tanto de gráficos inteiros, quanto de partes tomadas isoladamente. Pode-se trabalhar com uma escala absoluta, alternar a visualização de algum dos componentes para uma escala relativa e, após a análise, retornar à situação original. Pode-se ainda justapor escalas visualizando versões complementares. Considerados em conjunto (e como uma pequena parte das perspectivas possíveis) esses recursos ressaltam o aspecto altamente dinâmico do tratamento e da comunicação de dados, sobretudo colocando em destaque suas nuances semânticas em contraponto à precisão matemática dos dados. Tais nuances alcançam outra proporção informativa em vista da praticamente infinita capacidade do computador em processar diferentes vistas dos dados. Nesse contexto, o desdobramento natural deste trabalho é a elaboração de protótipos de interfaces digitais que permitam inserir a metodologia de Bertin no domínio computacional, empregando técnicas de programação e design de interação.

No que concerne ao uso prático das visualizações desenvolvidas em um âmbito mais próximo à situação de uso esperada, elas foram apresentadas à Mailsa Passos, professora do Programa de Pós-graduação em Educação da Universidade do Estado do Rio de Janeiro e estudiosa da questão das políticas afirmativas. A pesquisadora ponderou o papel dos gráficos estatísticos na disseminação de informações ressaltando os aspectos semânticos que influenciam no seu uso como instrumento de comunicação. Comentou ainda que todo gráfico é desenvolvido com um objetivo em mente, visando direcionar o leitor para a interpretação desejada. Nesse sentido, a professora ressalta que essa comunicação estatística não é neutra, nem isolada de intenção. Ao contrário, qualquer sistema organizado de dados, desde o modo como registra os dados até a forma de comunicá-los encerra uma ação política, podendo ser interpretado e utilizado com diferentes objetivos. De modo análogo, todo tratamento de dados, sobretudo quando a finalidade é comunicá-lo a outrem, envolve a interpretação pessoal do analista que insere suas vivências, crenças e ideologias no resultado de seu trabalho. Além disso, a pesquisadora observou que dados ausentes desta proposta poderiam ser utilizados como fator de contextualização adicional, somando-se aos indicadores já empregados. Em especial, foi sugerido a inclusão dos dados sobre as bolsas que visam a permanência dos alunos na universidade. Tais indicadores interferem, por exemplo, na questão da redução de trancamentos e abandonos dos cotistas, assim como no seu

desempenho geral ¹⁸. Feitas essas observações, a professora Mailsa considerou as visualizações apresentadas como uma interessante proposta, potencialmente útil aos estudiosos das políticas afirmativas, fornecendo um viés analítico complementar.

Considerado isso, como conclusão final, importa ressaltar que as propostas e considerações apresentadas nortearam-se por uma metodologia altamente desenvolvida e facetada, o que seria suficiente para justificar a continuidade dos seus estudos, tanto em relação aos aspectos abordados, quanto àqueles que não fizeram parte do escopo desta dissertação. Mas além disso, tais estudos são necessários porque o legado bertiniano circunscreve um campo de estudo abrangente e rico em possibilidades de desenvolvimento. Em especial no que concerne à visualização dos dados quantitativos, as teorias de Bertin podem parecer, em um primeiro momento, limitantes, uma vez que reduzem a exuberante diversidade de gráficos possíveis a uma tipologia restrita. Por outro lado, importa observar que inúmeras questões técnicas e semânticas que dizem respeito ao universo da informação visual apresentam-se com clareza somente após estabelecidos parâmetros de uso e comunicação básicos, fundamentados logicamente. Nesse processo, parte-se daquilo que a Semiologia gráfica considera como o marco zero da percepção (i.e., seus aspectos monossêmicos), características tidas como invariáveis que constituem uma estrutura conceitual definida, cuja contraparte visual são alguns poucos modelos de visualizações específicos. É a partir da aceitação de tais visualizações padrão (exemplificados pela construção normal e pelas matrizes de permutação) que se pode alcançar outro patamar de investigações gráficas, no qual problemas de diversas ordens começam a aparecer, mais complexos e interessantes do que tópicos de ordem puramente estética. Nesse caso, encaminha-se a discussão em outra direção, deslocando-a de uma vertente formal/estética para uma formal/estética/epistemológica, uma vez que os recursos apresentados por Bertin (e os desenvolvimentos que podem se seguir) tendem a influenciar o próprio modo de conhecer um problema.

¹⁸ Apesar de constarem no menu do cubo de graduação, os dados sobre as bolsas não estão disponibilizados para pesquisa pela interface do projeto de BI. Por esse motivo não puderam ser incluídos na análise.

REFERÊNCIAS

AIGNER, W. et al. Bertin was Right: An Empirical Evaluation of Indexing to Compare Multivariate Time-Series Using Line Plots. *Computer Graphics Forum*, v.30, n. 1, p. 215-228, 2011.

ALENCASTRO, Luiz Felipe de. Por que o Brasil precisa das cotas? *Outras palavras*, 27 abr 2012. Disponível em: <<http://goo.gl/I5uRB5>>. Acesso em: 02 ago. 2013.

ALEXANDER, Michael; WALKENBACH, John. *Excel Dashboards & Reports*. Indianapolis: Wiley, 2010.

ARNHEIM, Rudolf. Percepção em mapas. In: _____. *Intuição e intelecto na arte*. São Paulo: Martins Fontes, 2004

BERTIN, Jacques. *A neográfica e o tratamento gráfico da informação*. Universidade Federal do Paraná: Paraná, 1986.

_____. *Graphics and Graphic Information Processing*. Walter de Gruyter, 1981.

_____. *La graphique et le traitement graphique de l'information*. Paris: Flammarion, 1977.

_____. *Semiology of Graphics: Diagrams, Networks, Maps*. Esri Press, 2010.

_____. *Sémiologie graphique: Les diagrammes, les réseaux, les cartes*. Paris: Éditions de l'EHESS, 2005.

BRASIL. Lei nº 12.711, de 29 de agosto de 2012. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. Disponível em: <<http://goo.gl/4N5HvH>>. Acesso em: 02, ago. 2013.

BONSIEPE, Gui. Ciência | Comunicação | Design. *Arcos*, v.3, número único, 2000-2001.

_____. *Del objeto a la interfase: mutaciones del Diseño*. Buenos Aires: Infinito, 1999.

_____. A Step Towards the Reinvention of Graphic Design. *Design Issues*, Cambridge, MA, v.10 n.1, The MIT Press, 1994

CARD, Stuart, MACKINLAY Jock, SHNEIDERMAN, Ben. *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann, Inc, 1999.

CLEVELAND, William. *Visualizing Data*. New Jersey: AT&T Bell Laboratories, 1993.

ECKERSON, Wayne W. *Performance Dashboards: Measuring, monitoring, and managing your business*. Hoboken, NJ: John Wiley & Sons, 2011.

FEW, Stephen. *Information Dashboard Design: The Effective Visual Communication of Data*. Sebastopol, CA: O'Reilly, 2006.

FLUSSER, Vilém. *Filosofia da caixa preta: ensaios para uma futura filosofia da fotografia*. 1. ed. Rio de Janeiro: Relume Dumará, 2002.

_____. *O mundo codificado: por uma filosofia do design e da comunicação*. 1. ed. São Paulo: Cosac Naify, 2007.

FRAGA, Érica. Cota não garante aluno de escola pública em vestibular. *Folha de S. Paulo*. São Paulo, 25 mar. 2013. Cotidiano C1.

GOODMAN, Nelson. *Languages of Art: An Approach to a Theory of Symbols*. Cambridge: Hackett Publishing, 1968.

HILBERT, Martin; LÓPEZ, Priscila. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*. v.332 n.6025, p 60-65, Feb. 2011.

JUNIOR, ALVARO A. F. MENDES. *Três ensaios sobre ações afirmativas no ensino superior brasileiro: Acesso, progressão e simulações de diferentes políticas de cotas para a Universidade do Estado do Rio de Janeiro*. 2013. Dissertação de mestrado. Programa de Pós-Graduação em Ciências Econômicas, Universidade Federal Fluminense, Niterói.

KINROSS, Robin. The rhetoric of neutrality. *Design Issues*, Cambridge, MA, v.2, n.2, p.18-30, Autumn, The MIT Press, 1985.

KOSTELNICK, Charles. The Visual Rhetoric of Data Displays: The Conundrum of Clarity. *IEEE Transactions On Professional Communication*, v.50, n.4, 2007

KRIPPENDORFF, Klaus. On the Essential Contexts of Artifacts or on the Proposition That "Design Is Making Sense (Of Things)". *Design Issues*, v. 5, n.2, p. 9-39, The MIT Press, Spring 1989.

_____. Mathematical Theory of Communication. In: LITTLEJOHN S. W.; FOSS K. A. (ed.). *Encyclopedia of Communication Theory*. Los Angeles: Sage, 2009. p.614-618..

LOGAN, Robert K. *Que é informação? A propagação da organização na biosfera, na simbolosfera, na tecnosfera e na ecosfera*. Rio de Janeiro: Contraponto, 2012.

MACCOY, Katherine. Information and persuasion: rival or partners? *Design Issues*, Cambridge, MA, v.16, n.3, p.80-83, The MIT Press, Autumn, 2000.

MAGNOLI, Demétrio. Os “amigos-do-povo” contra o mérito. *O Estado de S. Paulo*, São Paulo, 16 ago. 2012. Disponível em: <<http://goo.gl/JBrgKv>>. Acesso em: 02 ago. 2013.

MATURANA, Humberto R.; VARELA, Francisco J. *A árvore do conhecimento: as bases biológicas da compreensão humana*. São Paulo: Palas Athena, 2005.

REINACH, Fernando. Cota desrespeita inteligência. *O Estado de S. Paulo*, São Paulo, 16 ago. 2012. Disponível em: <<http://goo.gl/6tGD2X>>. Acesso em: 02 ago. 2013.

SÁ, Ericka de. A UnB e o sistema de cotas nas universidades brasileiras. *Deutsche Welle*, Brasília, 18 mai. 2012. Disponível em: <<http://dw.de/p/14xLF>>. Acesso em: 02 ago. 2013.

SÁ, Ericka de. Movimento negro festeja decisão do STF a favor das cotas nas universidades. *Deutsche Welle*, Brasília, 24 abr. 2012. Disponível em: <<http://dw.de/p/14m74>>. Acesso em: 02 ago. 2013.

SANTOS, Mariana. Cotas para alunos de escolas públicas dividem comunidade universitária brasileira. *Deutsche Welle*, 17 ago. 2012. Disponível em: <<http://dw.de/p/15rto>>. Acesso em: 02 ago. 2013.

PANOFSKY, Erwin. *A Perspectiva como forma simbólica*. Lisboa: Edições 70, 1999.

PERSON, Ron. *Balanced Scorecards & Operational Dashboards with Microsoft Excel*. Indianapolis, IN: Wiley, 2009.

PIGNATARI, Décio *Informação. Linguagem. Comunicação*. São Paulo: Atelie Editorial, 2003

RIO DE JANEIRO. Lei nº 5346, de 11 de dezembro de 2008. Dispõe sobre o novo sistema de cotas para ingresso nas universidades estaduais e dá outras providências. Disponível em <<http://goo.gl/ImAEgf>>. Acesso em: 02 ago. 2013.

ROD, Jan Ketil. The third choice. Cybergeog: European Journal of Geography, Dossiers, *Colloque 30 ans de sémiologie graphique*”, 2000.

SAP AG. SAP User Interface Guidelines for Dashboard Design. Disponível em: < <http://goo.gl/AHx70m>>. Acesso em 02 ago. 2013.

SAP AG. Collaboration Workspace from SAP. Disponível em: <<http://goo.gl/yD4uH3>>. Acesso em: 02 ago. 2012.

SHANNON, Claude E. A Mathematical Theory of Communication. *The Bell System Technical Journal*. v.27. Jul. 1948. Bell Labs, New Jersey , 1948. Disponível em: <[http:// www.alcatel-lucent.com/bstj/vol27-1948/articles/bstj27-3-379.pdf](http://www.alcatel-lucent.com/bstj/vol27-1948/articles/bstj27-3-379.pdf)>. Acesso em: 20 jar 2012.

_____. A Mathematical Theory of Communication. *The Bell System Technical Journal*. v. 27. Oct. 1948. Bell Labs, New Jersey, 1948. Disponível em: <<http://www.alcatel-lucent.com/bstj/vol27-1948/bstj-vol27-issue04.html>>. Acesso em: 20 jar 2012.

SHANNON, Claude E.; WEAVER, Warren. *The Mathematical Theory of Communication*. Illinois: University of Illinois Press, 1998.

SPENCE, Robert. *Information Visualization*. Essex: Pearson, 2001.

NETO, Antonio Teixeira. Haverá, também, uma semiologia gráfica? *Boletim Goiano de Geografia*, Goiás, v. 4, 5, 6. n. 1, 2. Universidade Federal de Goiás, Departamento de Geografia, Jan-Dez. 1984, 1985, 1986. Disponível em: <<http://goo.gl/VkhDZ5>>. Acesso em: 26 ago. 2013.

TUFTE, Edward. *Visual Explanations*. Connecticut: Graphics Press, 2003.

_____. *Beautiful Evidence*. Connecticut: Graphics Press, 2006.

_____. *Envisioning Information*. Connecticut: Graphics Press, 1995.

_____. *The Visual Display of Quantitative Information*. Connecticut: Graphics Press, 2001.

TURBAN, Efraim et al. *Business Intelligence: Um enfoque gerencial para a inteligência do negócio*. Porto Alegre: Bookman, 2009.

VIEGAS, José Carlos. *Atlas das áreas verdes e densidades demográficas do Rio de Janeiro*. 1979. Projeto de graduação (Graduação em Desenho Industrial) - Escola Superior de Desenho Industrial, Universidade do Estado do Rio de Janeiro, Rio de Janeiro.

VISUAL MINING, INC. Dashboards: Another Look at Data Visualization. Disponível em <<http://goo.gl/m2BWc2>>. Acesso em: 19 set. 2012.

VISUAL MINING, INC. A Dashboard Data Methodology. Visual Mining, Inc, 2006. Disponível em: <<http://goo.gl/iQ0SwK>>. Acesso em: 02 ago. 2013.

WARE, Colin. *Information Visualization: Perception for Design*. San Francisco: Morgan Kaufmann, 2010.

WIENER, Norbert. *Cibernética e sociedade: o uso humano de seres humanos*. 7.ed. São Paulo: Cultrix, 2007.

WILSON, Robert A.; KEIL, Frank C. *The MIT Encyclopedia of Cognitive Sciences*. Cambridge, MA : MIT Press, 1999.

WURMAN, Richard Saul. Ansiedade de informação. São Paulo: Cultura, 1981.
_____. Hats. *Design Quarterly*, n.145. MIT Press for the Walker Art Center, Minneapolis, 1989.

ANEXO A – Demais tópicos da Semiologia gráfica

Este anexo aborda conceitos da teoria geral da Semiologia gráfica que não foram apresentados na parte principal deste trabalho por não serem indispensáveis para a compreensão do que se desenvolveu, de modo que puderam ser deslocados como anexo contribuindo para a concisão da parte principal da dissertação. Entretanto, apesar de não serem centrais no trabalho, trata-se de tópicos cujo entendimento perpassa e consubstancia todo o estudo elaborado.

1 Os meios do sistema gráfico

A segunda parte da Semiologia gráfica sistematiza características dos fenômenos gráficos para estabelecer uma correspondência estrita entre eles e a natureza dos dados originais. Nesse sentido, Bertin analisa as propriedades do plano e as formas de ocorrência de uma imagem sobre ele assim como o repertório de estímulos visuais aptos a transcrever determinada relação entre componentes.

1.1 Delimitação do sistema

Destacando a necessidade de delimitar rigorosamente um sistema de signos, Bertin estabelece as condições sob as quais as regras da SG são aplicáveis e passíveis de análise. Sendo assim, exclui do estudo a imagem cinética, que acontece no tempo, instância difícil de ser compreendida em um nível operacional, em contraponto à percepção da imagem estática, que é atemporal. Além disso, na relação entre ambas, as características cinéticas tendem a se sobrepor às outras variáveis, hierarquizando a ordem de percepção. Assim, tremores de imagem, animações e o cinema são excluídos da análise. Tal como os diversos aspectos da tridimensionalidade física, como o relevo real e as variações de distância e espessuras; além da variação nas condições de iluminação, as imagens estereoscópicas e anaglifos. Com isso, o âmbito da SG limita-se àquilo “que é representado ou impresso:

- sobre uma folha de papel branco
 - de formato médio, visível em um “golpe de vista”
 - a uma distância de visão correspondente à leitura de um livro ou atlas
 - sob iluminação normal e constante (mas eventualmente tendo em conta a diferença entre a luz do dia e a iluminação artificial)
 - por todos os meios gráficos disponíveis.”
- (BERTIN, 2005, p. 42)

Nesses limites, a teoria analisa as manchas (*taches*) sobre um plano: a modulação da energia luminosa que difere da superfície visualmente homogênea do papel. Tais manchas podem ocupar qualquer posição nas duas dimensões do plano e podem variar em seu tamanho, valor, retícula, cor, orientação e forma conforme ilustrado na Figura A.5. Estes estímulos constituem as 8 variáveis visuais que, junto à noção de implantação, “se ajustam de modo homogêneo aos problemas práticos com os quais se defronta o redator gráfico.” (BERTIN, 2005, p.42). O autor sistematiza as propriedades dessas variáveis com relação ao seu *nível de organização e comprimento*, assim como seu modo de ocorrência no plano.

1.1.1 O plano

1.1.1.1 Modo de implantação

Trata-se das três significações que uma mancha visível pode assumir conforme a extensão que ocupa no plano: os *modos de implantação pontual, linear e zonal* ou “os três momentos do continuum sensível aplicado ao plano. São as três figuras elementares da geometria.” (BERTIN, 2005, p.11). Assim, um indivíduo pode ser representado por um ponto (“um momento do plano, sem comprimento ou superfície”) em um gráfico de dispersão; determinada grandeza pode ser indicada por uma linha (“momento do plano com comprimento mensurável, mas sem superfície.”) em um histograma; uma região geográfica representada por uma zona (“parte do plano com superfície mensurável.” *ibid.*, p.44) em um mapa. Importa notar a característica dessas figuras como mera significação no plano, condição que independente da extensão real da figura. Assim, a significação pontual pode ser representada por um círculo de determinada superfície tal como a significação linear pode ser representada por

uma linha com uma certa espessura e, ainda assim, transcrevem, respectivamente, dados zerodimensionais ou unidimensionais.

Bertin salienta (*ibid.*, p.44) que o conceito de implantação implica que tanto o comprimento dos componentes de terceira dimensão como a representação de quantidades variam conforme o modo de implantação, e ainda que a diferença entre as implantações são seletivas. Sendo assim, um mesmo conceito não pode ter, em uma mesma imagem, dois modos de implantação diferentes.

1.1.1.2 Continuidade e homogeneidade do plano

A continuidade do plano implica a possibilidade de uma subdivisão infinita restrita apenas pelos limites da percepção visual. Nesse sentido, “depois da variação de forma, o plano oferece a variável visual mais longa.” (*ibid.*, p.46), de modo que os componentes mais longos tendem a ser representadas diretamente no plano. Como exemplifica um mapa das inúmeras municipalidades de um estado, no qual elas são demarcadas por diversas segmentações na superfície do plano). Já no que concerne à homogeneidade do plano, Bertin declara que a certeza dessa propriedade “leva ao pressuposto da homogeneidade das convenções em um espaço significativo.”. Isso implica que:

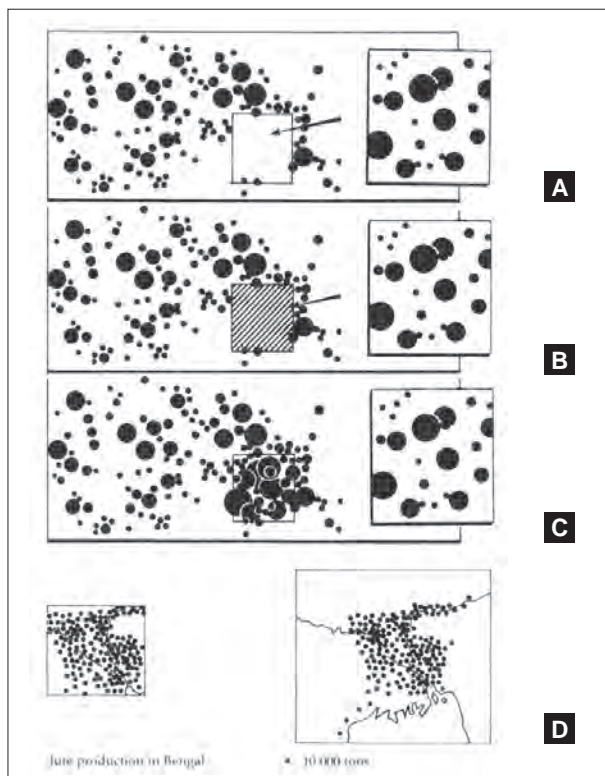
A ausência de sinais em um área significa ausência de fenômeno e não, por exemplo, falta de dados sobre o fenômeno (Figura 63A).

Toda variação visual é significativa, de modo que a introdução de elementos que não representam uma correspondência pode prejudicar a compreensão (63B).

Uma convenção deve permanecer invariável em um mesmo espaço pois a mudança é interpretada como transformação do fenômeno (63C).

Um quadro que circunscreva uma parte do plano delimita um recorte de significação importante, mas não o fenômeno em si, de modo que o leitor supõe que o fenômeno se prolonga para além do limite demarcado (63D).

Figura 63 – Características da percepção visual do plano



Fonte: BERTIN, 2010, p.47

1.1.1.3 Nível de organização do plano

O “nível de organização do plano é máximo. Suas duas dimensões fornecem as únicas variáveis com todas as propriedades perceptivas.” (BERTIN, 2005, p. 11). Isso significa que uma figura variando em posição no plano comporta as quatro atitudes perceptivas definidas na SG:

Seleção: quando é possível “isolar espontaneamente todas as correspondências pertencentes a uma mesma categoria [...] formando entre elas uma “família.” (ibid., p.48) : [●●●●●●●●] (i.e., grupo dos círculos vermelhos e grupo dos círculos pretos).

Associação: “quando se pode reagrupar espontaneamente todas as correspondências diferenciadas por certa variável. Essas correspondências são vistas como ‘todas as categorias confundidas’.” (ibid.). Essa característica tende à indistinção visual, como na variação de

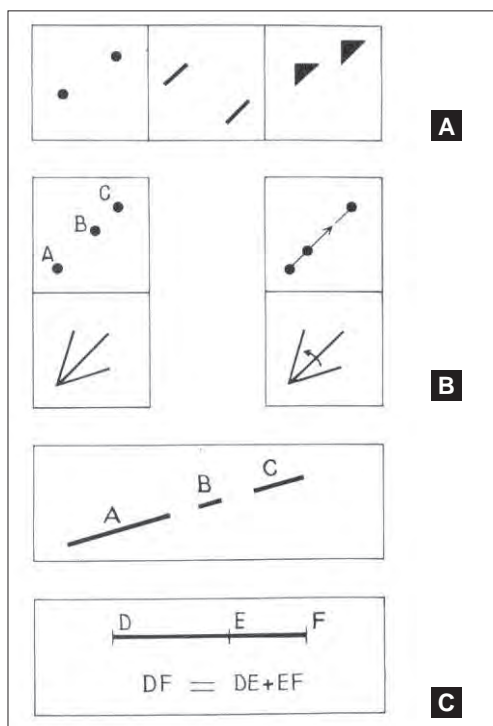
formas, em que nenhuma forma destaca-se visualmente das demais e, assim, não hierarquiza a ordem de percepção: [AEOFGCEDHKLOPVWAPV] (i.e., variações de forma).

Ordenação: “quando a classificação visual das categorias, de seus níveis é espontânea e universal.” (ibid.): [● ● ● ● ● ●] (i.e., variações de luminosidade).

Quantificação: “quando a distância visual entre as categorias de um componente ordenado podem se exprimir espontaneamente por uma relação numérica.” (ibid.): [■ ■ ■] (pode-se estimar que o último quadrado tem quatro vezes a área do quadrado central).

Bertin analisa as propriedades do plano observando que a variação de posição é *seletiva*, de modo que duas formas iguais, mas ocupando diferentes porções do plano são vistas como diferentes (Figura 64A); é *associativa*, pois as mesmas formas, apesar de estarem em posições diferentes, podem ser interpretados como semelhantes (idem); é *ordenada*, pois o alinhamento de figuras A, B, C pode ser percebido de modo universal como A, B, C ou C, B, A, mas não BCA (Figura 64B); é *quantitativa*, pois permite avaliar com certa precisão a diferença de grandeza entre os elementos A, B, C e a proporção desta relação (Figura 64C).

Figura 64 – Propriedades do plano



1.1.1.4 Imposição

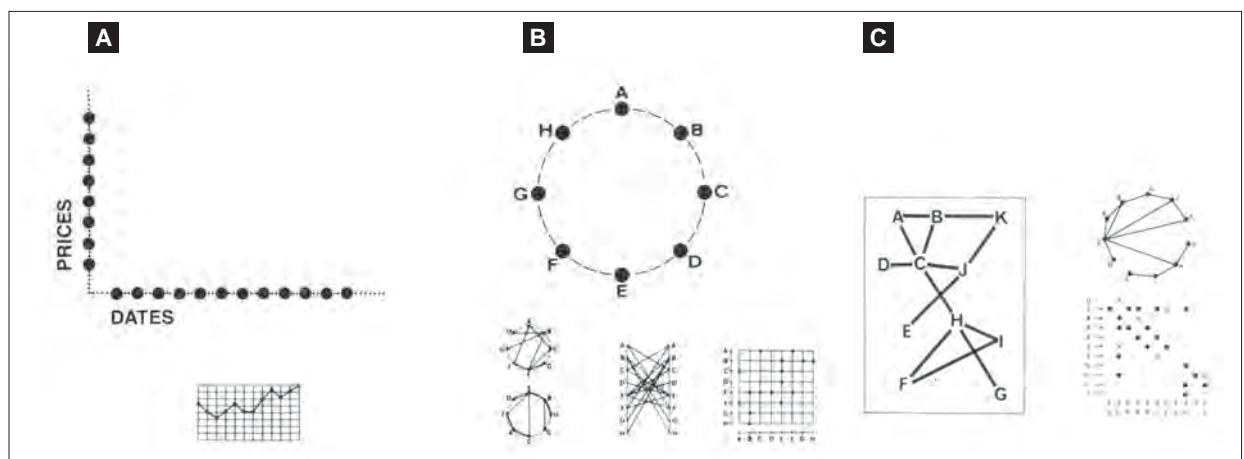
Este conceito define como as dimensões do plano serão usados para representar os dados. É o critério formal que distingue a especificidade das visualizações em razão da natureza dos dados e da maneira como as correspondências acontecem. Bertin descreve três modalidades de construções: *diagramas*, *redes*, e *cartas geográficas*. Classifica todas as outras ocorrências de imagens no plano como imposições *simbólicas*, nas quais as correspondências são externas à representação gráfica, sendo constituídas por analogias figurativas não passíveis de análise por critérios lógicos inequívocos e, portanto, não analisadas na teoria.

Assim, os *diagramas* ocorrem “quando as correspondências no plano podem se estabelecer [...] entre todas as divisões de um componente [...] e todas as divisões de um *outro* componente [...]” (BERTIN, 2005, p. 50 - grifo do autor). Exemplo dessa construção é a correlação entre os meses em determinado intervalo de tempo e os valores de certo produto visualizados no gráfico da Figura 65A.

Redes são definidas “quando as correspondências no plano podem se estabelecer [...] entre todas as divisões de um *mesmo* componente [...]” (p.50, grifo do autor). Bertin dá como exemplo as relações entre os diferentes indivíduos A, B, C, D... de um grupo, conforme visualizadas na Figura 65B.

As cartas geográficas ocorrem “quando as correspondências no plano podem se




















Figura 65 – Três categorias de imposição



estabelecer [...] entre todas as divisões de um mesmo componente, dispostos segundo a ordem geográfica [...]” (ibid., p.50). São exemplificadas por um inventário de rotas que visualiza as ligações entre um série de cidades demarcadas por sua localização geográfica (Figura 65C).

Essas três modalidades constituem os grandes grupos de imposição, primeiro estágio da construção de uma visualização. Mas, suas especificidades dependem ainda de outro parâmetro: os *tipos de imposição*, como indicados na Figura 66. Trata-se da tipologia que descreve as conformações espaciais particulares: *retilínea*, *circular*, *ortogonal* ou *polar*. Do equacionamento entre essas configurações e uma *variável de elevação* resulta a grande diversidade formal de mapas e gráficos. Bertin salienta que “a variedade de construções possíveis para um diagrama ou para uma rede coloca o problema da escolha da construção e, em vez disso, a ausência desse problema por meio do mapeamento.” (ibid. p.52). Por essa tabela podemos intuir a fonte da diversidade gráfica dos dashboards, embora essa variabilidade não esteja necessariamente associada à eficiência dessas visualizações.

Figura 66 – Tipos de imposição

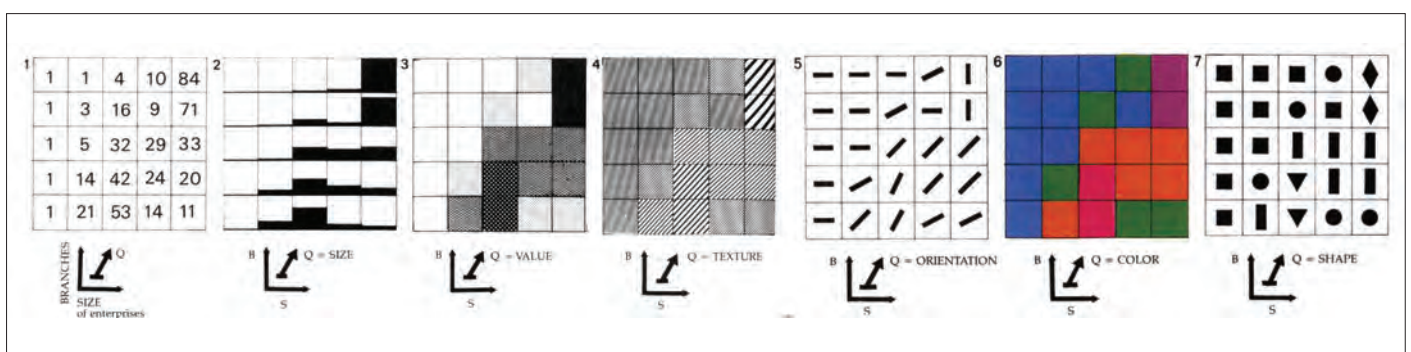
IMPOSITION		TYPES OF IMPOSITION						
		ARRANGEMENT	RECTILINEAR	CIRCULAR	ORTHOGONAL	POLAR		
GROUPS OF IMPOSITION	DIAGRAMS		 	 	 	 		
	NETWORKS	 	 	 	 			
	MAPS	 						
	SYMBOLS							

1.1.2 Variáveis retinianas

Provavelmente um dos tópicos mais debatidos da SG, este conceito elenca o repertório de estímulos visuais passíveis de representar a natureza *quantitativa*, *seletiva*, *associativa* ou *ordenada* dos dados. Bertin nomeia como “elevação ou variação de 3ª dimensão o uso das seis variáveis para além do plano ou variáveis retinianas (variáveis de 3ª dimensão).” (Bertin 2005, p. 11). São designadas deste modo pois representam o terceiro componente de um conjunto dados, fazendo-o “elevar-se” perceptivamente acima do plano. Trata-se das variáveis *tamanho*, *valor*, *retícula*, *cor*, *orientação* e *forma*, que, tal como os componentes, podem ser empregadas conforme o *nível de organização* e *comprimento* que são próprios a cada variável. Elas são necessárias “em todos os problemas gráficos com três componentes ou mais, quando as duas dimensões do plano são mobilizadas por um sistema de dupla identificação.” (ibid., p.62). Bertin exemplifica com a visualização de uma tabela de dupla entrada representada por cada uma das variáveis (Figura 67). Os dois componentes (dimensão da empresa e tipo de empresa) mobilizam as respectivas duas dimensões do plano e as quantidades são assinaladas nas células da tabela, correspondendo ao terceiro componente, aquele a ser representado graficamente empregando-se as variáveis retinianas.

Bertin observa que a correspondência entre as dimensões do plano e dois componentes “fixam de uma vez por todas a construção de base. [...] definindo também o campo de visão.” (ibid. 2005, p.62). Assim, delimitam um espaço semântico próprio, fora do qual não

Figura 67 – Variáveis retinianas



Tamanho, valor, retícula, orientação, cor, e forma.

Fonte: BERTIN, 2010, p.62-63.

há significação para os propósitos de representação gráfica dos dados. O autor observa que esse espaço pode ser separado do que ocorre “sobre” ele (por meio das variáveis retinianas) devido à propriedade da *preensão* do plano. Por tal hipótese, os movimentos oculares que circunscrevem o plano ao observá-lo definem sensibilidades “musculares” que distinguem-se das sensibilidades “retinianas” por meio das quais as variáveis “retinianas” são percebidas. Assim “na escala das percepções comuns que nos interessa aqui, as variáveis retinianas são psicologicamente diferentes das dimensões do plano.” (ibid., p.63).

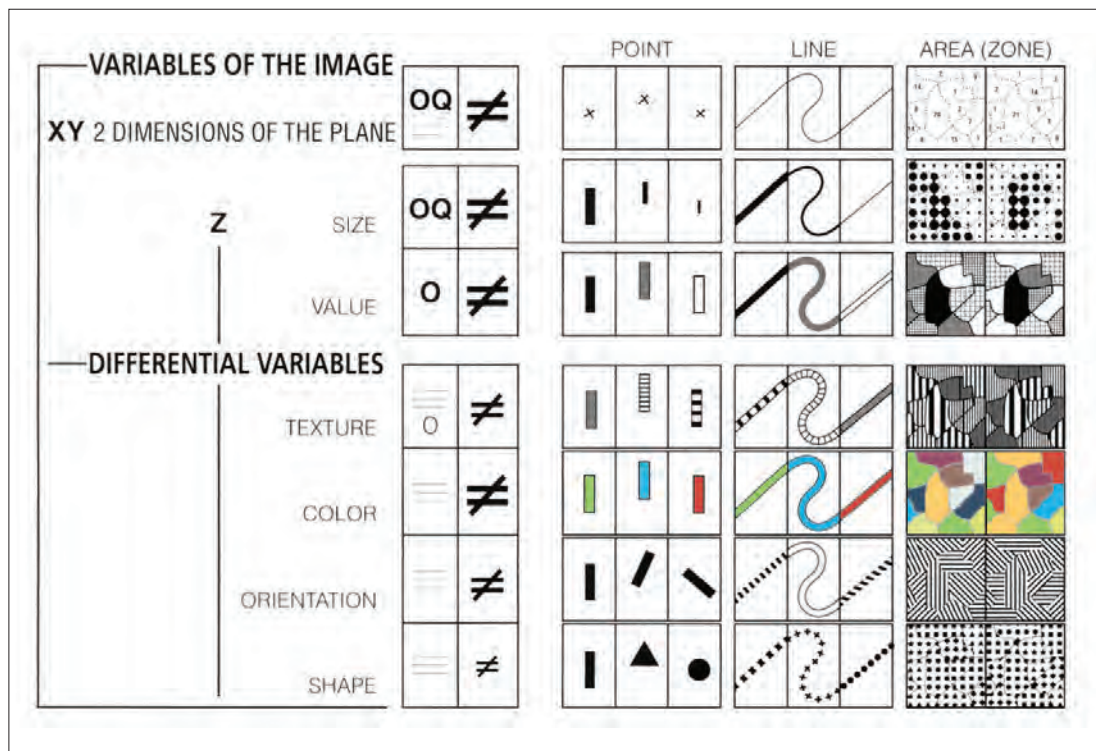
O conhecimento das propriedades do plano e das variáveis permite a melhor representação do dado original, contribuindo na comunicação da informação que possa conter. Objetivo que o psicólogo Rudolph Arheim sintetiza ao sugerir que os bons cartógrafos conseguem “traduzir os aspectos salientes da mensagem nas qualidades expressivas do meio de expressão, de tal modo que a informação seja obtida como um impacto direto de forças perceptivas.”. Em termos operacionais, esse alto grau de exigência significa reduzir ao máximo o recurso à legendas para que a informação seja apreendida diretamente pelo observador.

1.1.2.1 Nível de organização das variáveis

Somente o plano apresenta as propriedades de todos os níveis de organização, que ocorrem de diferentes modos quando aplicados às variáveis de 3ª dimensão devido tanto à especificidade da variável quanto de sua implantação. Tais propriedades são resumidas no quadro da Figura 68.

Percepção associativa: “utilizada quando se procura igualar uma variação, para reagrupar as correspondências [de modo que] ‘todas as categorias desta variação se confundam’”. (BERTIN, 2005, p. 65). Com esse conceito, o autor distingue as variáveis *associativas* (*forma, orientação, cor e retícula*), das *dissociativas* (*tamanho e valor*). Nas primeiras, a visibilidade dos signos não se altera, e então eles tendem a se confundir, tornando-se indistinguíveis. Predomina assim uma percepção homogênea de todo o conjunto, independentemente de como os signos variam. No segundo caso, os signos definem padrões heterogêneos em razão da potência dos signos maiores e mais escuros, que emergem em contraste aos menores e mais claros, hierarquizando a percepção e definindo grupos visuais distintos. Esse fenômeno é percebido nos três modos de implantação.

Figura 68 – Nível de organização das variáveis retinianas



Fonte: BERTIN, 2010, p.421

Percepção seletiva: “utilizada quando se procura responder [...] ‘onde está tal categoria?’ O olho deve isolar todos os elementos dessa categoria, abstrair todas as outras e perceber a imagem formada pela categoria buscada.” (ibid., p.67). Bertin aponta que tal percepção pode ser espontânea, como acontece com a variação de cor, variável seletiva por excelência, em que signos da mesma cor são rapidamente reunidos visualmente. No caso de variáveis pouco seletivas, exige-se uma comparação meticulosa (e normalmente mais lenta) de todos os signos. Apesar de todas as variáveis serem seletivas, a seletividade acontece em maior ou menor grau. Assim, no âmbito operacional, a SG não considera seletiva a variação de forma em nenhuma implantação e a variação de orientação em implantação zonal.

Percepção ordenada: “utilizada quando se procura comparar duas ou mais ordens [...]”. (ibid., p.67). A variação ordenada permite organizar os signos em uma ordem universalmente aceita, dispensando a leitura de legendas para identificar como uma categoria é ordenada. As variações de forma, cor e orientação não apresentam essa característica, que ocorre nas

variações de retícula, valor e tamanho, “que impõem à todas [as outras variáveis] uma ordem imediatamente sensível e universal” (ibid., p.67) no três modos de implantação.

Percepção quantitativa: “utilizada quando se procura definir a relação numérica entre dois signos [e] agrupar signos homogêneos, isto é, apresentar “distâncias” quantitativas aproximadas e definir assim os patamares naturais resultantes de uma contagem.” (ibid., p.69). Nesse nível de percepção, a relação proporcional entre dois signos manifesta-se sem recurso à legenda: “isto é o dobro [ou] oito vezes aquilo.” (ibid., p.69). Bertin considera que somente a variação de *tamanho*, em todas as implantações, apresenta tal propriedade, uma vez que um signo menor serve como unidade proporcional em relação a um outro maior. Essa possibilidade não ocorre, por exemplo, na variação de *valor*, em que “o branco não pode servir como unidade para medir um cinza, nem [...] o preto.” (ibid., p.69).

1.1.2.1 Propriedades das variáveis retinianas

Variação de tamanho: Trata-se do estímulo sensível propiciado pela variação na superfície do signo de modo que ele ocupe uma área progressivamente maior ou menor do plano. Quanto ao comprimento da variável, na percepção *quantitativa* e na *ordenada* as distinções possíveis seriam, a princípio, ilimitados, embora Bertin considere que “o olho não diferencia, em média, mais de 20 patamares entre dois pontos cuja relação seja de 1 para 10.” (BERTIN, 2005, p.71). Além disso, sua característica seletiva é curta, não admitindo mais que quatro ou cinco distinções nas quais pode-se isolar visualmente o grupo delineado por pontos de um mesmo tamanho. A variação é ainda *dissociativa*, pois os signos de maior tamanho hierarquizam a ordem de percepção e assim não se podem abstraí-los visualmente e enxergar um conjunto homogêneo.

Variação de valor: Refere-se à variação contínua na luminosidade de um signo em gradações do preto ao branco, intermediados pelos diversos níveis de cinza. Em percepção ordenada, o valor permite incontáveis gradações, mas na percepção seletiva, Bertin ressalta que “é prudente não ultrapassar 6 ou 7 patamares de valores, incluindo preto e branco.” (ibid., p.73). O comprimento dessa variável é sensível ao tamanho dos signos, sendo assim, naqueles muito pequenos o número de patamares seletivo é reduzido. Tal como o *tamanho*, o *valor* é uma

variável dissociativa, e por isso não é possível enxergar suas variações de modo homogêneo, pois os signos mais escuros predominam visualmente sobre os mais claros.

Variação de retícula: É “a sensação que resulta da sucessão de reduções fotográficas de uma retícula de pontos.” (ibid., p.79). Quando se trata de uma padrão regular, esse processo refere-se ao movimento de aproximação ou afastamento a partir de uma retícula dada. No primeiro caso, o processo enquadra um número menor de elementos progressivamente maiores; no segundo, ocorre o inverso. Independentemente do sentido do movimento, a variação de *grão*, diferentemente da variação de *tamanho*, não modifica a proporção entre figura e fundo presente no estado inicial da retícula. O comprimento dessa variável está condicionado à implantação. A implantação zonal permite padrões formados por pontos grandes, habilitando maior número de níveis identificáveis do que a linear, na qual em uma linha de 1 cm “conserva um certo comprimento (3 ou 4 patamares seletivos)” (ibid., p.79). Em implantação pontual, Bertin aponta a necessidade de pontos grandes que, ainda assim, “não fornecem mais que dois ou três patamares seletivos.” (ibid., p.79).

Variação de cor: Trata-se da “diferenciação sensível geralmente provocada pelas excitações coloridas diferentes que podemos perceber entre dois intervalos de mesmo valor.” (ibid., p.85). Nesse sentido, as propriedades dessa variável são consideradas de modo estrito, excluindo-se as variações de luminosidade (valor) que permeiam a noção comum do conceito “cor”. Assim, é possível estabelecer que a cor, isoladamente, não se apresenta como variável ordenada, pois não possibilita ordenar signos de modo inequívoco, inviabilizando a visualização de um componente ordenado. Por outro lado, apresenta boa seletividade, contribui para a memorização da mensagem e possibilita facilmente combinações com outras variáveis. Apesar disso, Bertin, mesmo reconhecendo o apelo psicológico da cor, sugere reservá-la para mensagens de natureza pedagógica considerando que “a cor não é estritamente indispensável aos problemas de pesquisa e tratamento.” (ibid., p.90).

Variação de orientação: “É a diferença de ângulo entre os campos criados por vários signos paralelos [...]” (p.93). Trata-se de uma variável com comprimento curto, permitindo poucas variações úteis. Em implantação pontual, a SG considera o uso máximo de quatro a cinco angulações em para assegurar a seletividade observando que os signos oblíquos tendem a constituir um grupo destacados dos signos ortogonais. Em implantação linear, possibilita duas variações: o eixo da linha e sua perpendicular. Na implantação zonal, apresenta-se como a variável menos seletiva.

Variação de forma: Refere-se à diversidade formal de um signo, cujo comprimento é praticamente ilimitado. É uma variável associativa, pois em um golpe de vista confunde todas as formas vistas em uma percepção homogênea. Não tem a característica da seletividade pois inviabiliza o discernimento de categorias visuais, por isso Bertin declara que “é inutilizável nos problemas de regionalização.” (ibid., p.95). Para otimizar sua característica seletiva é possível combiná-las com outras variáveis permitindo gerar séries seletivas suficientemente longas. O autor aponta que o forte atributo simbólico da forma, base dos sistemas de codificação, cujos significados não são de natureza universal, como acontece aos estímulos monossêmicos. Nesse âmbito, Bertin considera que a elaboração de codificações eficientes é prioritariamente um problema de ordem sintática, de diferenciação formal entre os signos do mesmo código. Assim, estaria menos relacionado “à descoberta de formas espetaculares [do que à] definição do campo de utilização no qual seu significado será constante e suficientemente repetido para tornar-se um hábito adquirido.” (p.95).

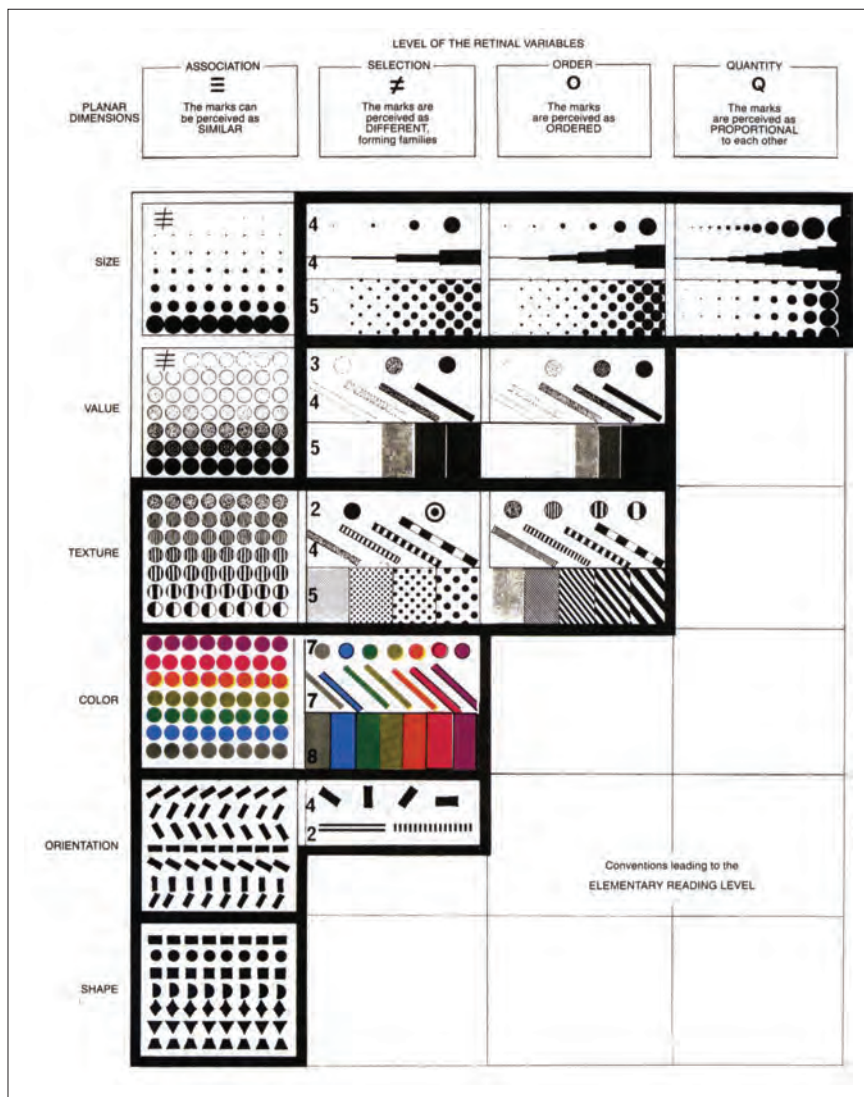
Bertin resume as considerações anteriores em uma tabela que compara as variáveis ao nível de organização indicando ainda os tipos de implantação e o comprimento das variáveis seletivas (Figura 69). As células vazias da tabela representam combinações que não ultrapassam a leitura elementar (aquela capaz de responder apenas a perguntas pontuais e não de avaliar tendências). O comprimento das variáveis seletivas é indicado pelo numeral à sua esquerda. Nos outros casos, o comprimento é potencialmente ilimitado.

2 As regras do sistema gráfico

2.1 As regras gerais de legibilidade

Após a escolha das variáveis visuais mais adequadas para um conjunto de dados em particular, importa otimizar tais estímulos por meio das distinções visuais mais eficazes para assegurar a legibilidade de um gráfico. Com esse objetivo, Bertin afirma que os princípios de legibilidade de uma construção gráfica são “as observações que permitem implementar as maiores diferenças sensíveis à visão.” (Bertin 2005, p.14). O autor compara tais regras aos princípios de clareza

Figura 69 – Nível de organização das variáveis visuais



Fonte: BERTIN, 2010, p.96

da oratória, que zelam por uma comunicação clara por meio da dicção e do volume adequados, lembrando a possibilidade de um bom discurso ser mau pronunciado e de um texto ruim ser declamado com excelência. Analogamente, mesmo que um gráfico atenda aos critérios de eficácia da SG, sua apreensão depende de se estabelecer suficientes inflexões visuais que o façam legível. Elas são específicas a cada variável ou combinação de variáveis sendo analisadas por seu comprimento e de acordo com o nível de significação que representa.

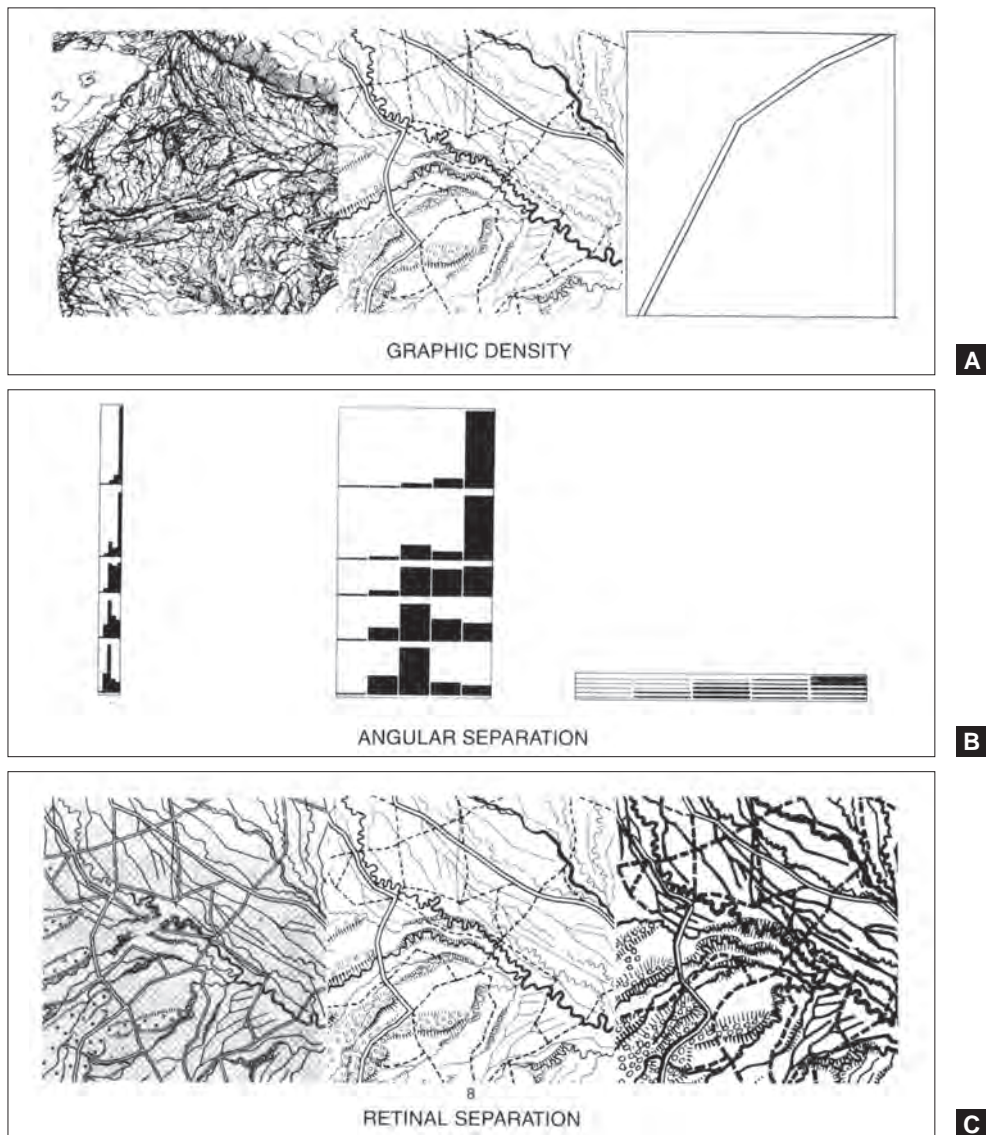
As regras de legibilidade são resumidas em três categorias (Figura 70):

Densidade gráfica: refere-se à quantidade adequada de pontos por cm^2 de modo a se estabelecer um equilíbrio visual entre a densidade que beira a ilegibilidade e a simplificação excessiva (70A).

Separação angular: trata da separação entre as duas dimensões do plano que previne o achatamento horizontal ou vertical do campo que contém o gráfico (70B).

Separação retiniana: ressalta o adequado destaque aos elementos importantes para

Figura 70 – Regras gerais de legibilidade



que eles se sobressaíam em relação aos não significativos. No caso das variáveis retinianas, aponta para a separação entre os patamares sensíveis evitando imagens muito tênues ou desnecessariamente carregadas, casos em que a escala de distinções visuais tende para um dos extremos e não para a utilização equilibrada das distinções ao longo da escala (70C).

Assim, somados ao segmento da *Análise da informação* e das *Regras do sistema gráfico*, constantes no capítulo 2 deste trabalho, completa-se a exposição da teoria geral da semiologia gráfica. A partir desse estágio, o livro de Bertin apresenta uma série de exemplos e casos particulares nos quais a teoria é aplicada, podendo interessar aos pesquisadores voltados às áreas de informação visual curiosos sobre os desdobramentos práticos da metodologia.

Anexo B – Planilhas de base

Tabela 7 – Planilha diagrama 1: características gerais

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Sexo masculino %	53,05	54,64	55,92	53,94	52,28	54,78	52,50	51,71	53,18	52,87	53,14	54,23	52,89
Sexo feminino %	46,95	45,36	44,08	46,06	47,72	45,22	47,50	48,29	46,82	47,13	46,86	45,77	47,11
Est-civil solteiro %	84,73	89,71	90,30	91,53	91,20	92,89	96,19	95,38	93,53	90,77	92,42	91,76	91,72
Est-civil casado %	13,46	8,88	8,28	7,79	8,03	6,45	3,52	3,85	5,29	7,67	5,84	6,55	6,53
Est-civil outros %	1,81	1,42	1,42	0,68	0,77	0,67	0,28	0,77	1,18	1,56	1,74	1,69	1,74
Rede pública %	37,06	32,50	37,66	60,03	51,64	51,76	49,68	44,11	43,59	45,53	43,97	47,32	46,71
Rede particular %	62,94	67,50	62,34	39,97	48,36	48,24	50,32	55,89	56,41	54,47	56,03	52,68	53,29
Turno integral %	10,17	10,19	9,35	9,75	10,68	10,88	10,56	10,21	10,60	10,37	9,84	9,22	9,80
Turno manhã %	38,49	39,42	41,48	42,22	40,91	40,35	41,07	41,17	42,25	41,94	42,52	40,79	40,53
Turno noite %	51,34	50,39	49,16	48,03	48,42	48,77	48,37	48,62	47,15	47,69	47,64	49,99	49,67
Estrut-acad CEH %	36,06	34,63	37,72	37,59	37,65	37,18	37,31	36,23	37,14	37,34	37,24	38,36	37,68
Estrut-acad CTC %	31,47	34,65	33,15	32,54	32,51	33,18	33,11	33,39	32,94	33,15	32,72	32,83	33,01
Estrut-acad CBI %	8,39	8,36	7,84	8,25	8,45	8,32	8,30	8,48	8,65	8,46	8,09	7,78	8,04
Estrut-acad CCS %	24,08	22,36	21,29	21,61	21,39	21,32	21,28	21,91	21,26	21,05	21,95	21,02	21,27
Integralização 1 %	5,26	4,39	4,06	3,82	3,88	3,73	3,32	3,50	3,82	3,46	3,75	3,25	4,56
Integralização 2 %	48,77	48,11	49,95	52,35	53,83	52,73	57,79	53,49	49,35	48,15	41,96	39,63	44,04
Integralização 3 %	27,07	28,00	25,84	25,66	23,91	24,73	21,61	24,32	27,69	29,18	30,68	33,71	30,96
Integralização 4 %	18,90	19,49	20,15	18,17	18,38	18,81	17,29	18,69	19,14	19,21	23,61	23,40	20,44
Conclusão %	53,94	58,79	56,66	57,69	56,63	60,54	56,03	58,53	52,69	56,67	55,51	54,55	41,06
Abandono %	17,01	11,87	10,87	9,87	10,56	10,12	12,63	11,23	11,63	11,10	13,17	10,86	8,90
Eliminação %	1,11	1,07	1,30	2,89	1,36	1,77	1,73	1,11	1,33	1,06	0,88	1,02	1,78
Trancamento %	6,59	6,65	17,07	15,36	16,16	14,87	17,80	16,33	17,17	16,51	16,48	14,97	16,38
Transferência %	0,71	1,56	1,16	3,61	1,87	1,24	1,09	1,39	1,60	0,95	0,90	1,10	1,47
CR em_turma	7,03	7,03	6,93	6,93	6,86	6,81	6,75	6,71	6,60	6,58	6,56	6,56	6,57
CR aluno	6,76	6,87	6,72	6,78	6,70	6,69	6,62	6,58	6,53	6,49	6,47	6,47	6,48
Reprovação-nota %	6,11	5,78	5,21	5,90	x	x	x	6,80	6,60	7,80	7,60	7,80	x
Reprovação-freq %	11,37	10,70	11,94	11,10	x	x	x	13,60	15,60	15,50	14,90	14,00	x
Vestibulando-púb %	x	x	x	x	45,58	44,44	44,56	40,01	39,76	41,06	39,35	39,76	41,12
Vestibulando-part %	x	x	x	x	54,42	55,56	55,44	59,99	60,24	58,94	60,65	60,24	58,88
Aptos %	90,91	91,11	86,91	87,22	86,76	87,89	85,84	86,74	86,91	87,83	87,88	89,06	89,43
Cota total %	x	x	x	55,42	39,00	30,32	29,96	23,48	21,51	26,73	29,87	29,62	26,45
Cota negro %	x	x	x	35,42	16,11	11,01	10,30	7,93	8,06	10,44	12,75	12,93	11,69
Cota def/ind %	x	x	x	0,00	0,62	0,67	0,66	0,38	0,53	1,02	0,47	0,42	0,48
Cota rede pública %	x	x	x	20,00	22,26	18,64	19,00	15,16	12,92	15,27	16,65	16,27	14,28
Cota não cota %	x	x	x	44,58	61,00	69,68	70,04	76,52	78,49	73,27	70,13	70,38	73,57
Concl x sexo f %	32,92	37,06	34,03	35,15	34,67	36,76	36,25	35,82	30,98	34,43	33,29	33,69	25,10
Concl x sexo m %	21,02	21,73	22,63	22,54	21,97	23,78	19,77	22,71	21,70	22,24	22,22	20,87	15,96
Concl x rd pública %	39,41	38,44	34,58	35,79	34,11	32,86	38,28	45,26	48,63	49,46	47,58	49,53	51,14
Concl x rd particular %	60,59	61,56	65,42	64,21	65,89	67,14	61,72	54,74	51,37	50,54	52,42	50,47	48,86
EA ceh x rd particular %	49,22	52,41	46,15	30,85	43,85	43,48	46,09	50,55	51,88	52,11	54,18	49,31	46,84
EA CEH x rd pública %	50,78	47,59	53,85	69,15	56,15	56,52	53,91	49,45	48,12	47,89	45,82	50,69	53,16
EA CTC x rd particular %	66,35	70,70	66,67	47,58	46,66	49,45	52,67	59,44	60,64	56,83	58,47	55,08	58,31
EA CTC x rd pública %	33,65	29,30	33,33	52,42	53,34	50,55	47,33	40,56	39,36	43,17	41,53	44,92	41,69
EA CBI x rd particular %	82,32	92,89	82,72	56,40	66,41	59,38	56,75	62,13	62,42	56,95	60,45	56,15	62,57
EA CBI x rd pública %	17,68	7,11	17,28	43,60	33,59	40,63	43,25	37,87	37,58	43,05	39,55	43,85	37,43
EA CCS x rd particular %	72,95	73,72	75,37	38,06	50,88	49,59	48,91	56,03	56,50	54,31	53,77	53,92	53,31
EA CCS x rd pública %	27,05	26,28	24,63	61,94	49,12	50,41	51,09	43,97	43,50	45,69	46,23	46,08	46,69

Tabela 8 – Comparação entre cotistas (COTA) e não cotistas (NCOTA): características gerais

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
COTA x est-civil solteiro	91,41	91,39	92,44	95,16	91,66	90,88	91,48	93,40	94,63	93,40
COTA x est-civil casado	7,89	7,81	7,13	4,65	6,71	6,47	6,28	4,12	3,96	4,78
COTA x est-civil separado+outros	0,70	0,80	0,43	0,19	1,63	2,66	2,24	2,48	1,41	1,83
COTA x rede pública	81,93	82,44	86,79	85,58	85,45	82,78	83,69	81,62	81,16	82,98
COTA x rede particular	18,07	17,56	13,21	14,42	14,55	17,22	16,31	18,38	18,84	17,02
COTA x turno integral	9,09	10,22	11,57	12,34	13,59	14,88	14,91	14,33	14,83	14,96
COTA x turno manhã	42,77	41,36	40,88	39,95	39,40	40,92	41,07	41,82	40,79	38,83
COTA x turno noite	48,14	48,41	47,55	47,71	47,01	44,46	44,02	43,84	44,37	46,21
COTA x abandono	5,99	11,11	15,42	23,07	33,96	34,90	28,95	33,79	34,21	33,15
COTA x conclusão	0,03	0,00	0,12	18,79	69,48	91,67	73,29	62,29	60,55	57,87
COTA x eliminação	4,02	0,28	0,73	1,76	2,15	2,83	1,81	2,55	5,05	10,32
COTA x trancamentos	9,42	19,15	30,41	44,56	64,49	74,14	69,53	67,65	63,81	73,10
COTA x transferência	0,00	0,80	0,91	1,13	2,32	1,86	1,08	0,85	0,83	1,47
COTA x sexo feminino	56,84	55,01	56,98	58,83	59,07	59,52	57,76	56,08	57,93	57,87
COTA x sexo masculino	43,16	44,99	43,02	41,17	40,93	40,48	42,24	43,92	42,07	42,13
COTA x cr-aluno	6,44	6,49	6,64	6,74	6,75	6,70	6,55	6,42	6,38	6,68
COTA x cr-em_turma	6,49	6,64	6,76	6,88	6,87	6,77	6,57	6,46	6,43	6,72
COTA x ea CEH	59,89	41,09	30,31	30,89	23,40	17,92	19,46	21,51	20,93	16,80
COTA x ea CTC	45,62	33,86	22,86	19,56	14,96	15,13	24,65	28,41	28,84	27,00
COTA x ea CBI	60,83	41,34	39,82	42,53	37,05	34,73	41,46	46,51	47,69	41,34
COTA x ea CCS	60,32	42,31	38,08	39,51	31,34	32,20	37,09	40,14	40,00	37,00
COTA x ea CEH x conclusão	0,00	0,00	100,00	78,26	52,97	42,51	44,14	37,46	40,97	33,98
COTA x ea CTC x conclusão	0,00	0,00	0,00	3,34	9,53	14,11	12,91	18,26	18,37	18,38
COTA x ea CBI x conclusão	0,00	0,00	0,00	10,70	15,22	16,91	14,19	14,59	13,83	17,69
COTA x ea CCS x conclusão	100,00	0,00	0,00	7,69	22,28	26,47	28,77	29,70	26,82	29,94
COTA x ea CEH x integ 1	0,00	0,00	100,00	1,28	2,31	3,16	3,55	1,67	2,05	4,90
COTA x ea CEH x integ 2	0,00	0,00	0,00	98,72	87,27	58,47	45,23	36,67	25,06	27,35
COTA x ea CEH x integ 3	0,00	0,00	0,00	0,00	10,42	38,37	40,58	38,33	49,87	39,59
COTA x ea CEH x integ 4	0,00	0,00	0,00	0,00	0,00	0,00	10,64	23,33	23,02	28,16
COTA x ea CTC x integ 1	0,00	0,00	0,00	40,00	3,90	3,42	0,76	2,30	3,45	4,55
COTA x ea CTC x integ 2	0,00	0,00	0,00	60,00	77,92	43,84	26,72	25,29	13,22	18,18
COTA x ea CTC x integ 3	0,00	0,00	0,00	0,00	14,29	37,67	50,38	43,68	44,25	34,85
COTA x ea CTC x integ 4	0,00	0,00	0,00	0,00	3,90	15,07	22,14	29,31	39,08	42,42
COTA x ea CBI x integ 1	0,00	0,00	0,00	3,13	0,00	1,14	0,00	0,00	0,00	0,79
COTA x ea CBI x integ 2	0,00	0,00	0,00	96,88	94,31	85,14	75,69	60,43	61,83	70,08
COTA x ea CBI x integ 3	0,00	0,00	0,00	0,00	5,69	10,29	20,14	33,09	32,06	25,98
COTA x ea CBI x integ 4	0,00	0,00	0,00	0,00	0,00	3,43	5,56	6,47	6,11	3,15
COTA x ea CCS x integ 1	100,00	0,00	0,00	13,04	4,92	4,00	1,71	2,80	1,18	2,33
COTA x ea CCS x integ 2	0,00	0,00	0,00	86,96	71,04	60,36	51,88	40,91	44,71	43,26
COTA x ea CCS x integ 3	0,00	0,00	0,00	0,00	24,04	35,64	34,47	31,47	21,96	22,79
COTA x ea CCS x integ 4	0,00	0,00	0,00	0,00	0,00	0,00	11,95	24,83	32,16	31,63
COTA x em_turma	99,04	95,84	95,35	93,94	93,32	92,00	94,54	93,87	93,30	94,27
COTA x aptos	93,74	93,26	93,17	91,15	91,50	91,34	92,35	92,01	92,21	92,95
NCOTA x est-civil solteiro	91,68	91,07	93,05	96,58	96,53	94,32	90,37	92,01	90,56	91,12
NCOTA x est-civil casado	7,67	8,17	6,18	3,09	2,97	4,94	8,45	6,57	7,64	7,17
NCOTA x est-civil separado+outros	0,65	0,75	0,77	0,32	0,50	0,74	1,18	1,42	1,80	1,72
NCOTA x rede pública	32,62	31,95	36,81	34,60	31,66	30,99	30,93	27,36	32,16	32,76
NCOTA x rede particular	67,38	68,05	63,19	65,40	68,34	69,01	69,07	72,64	67,84	67,24
NCOTA x turno integral	10,57	10,97	10,58	9,80	9,18	9,44	8,71	7,92	6,86	7,95
NCOTA x turno manhã	41,54	40,61	40,12	41,55	41,71	42,65	42,26	42,82	40,79	41,15
NCOTA x turno noite	47,89	48,42	49,30	48,65	49,11	47,91	49,03	49,26	52,35	50,91
NCOTA x abandono	79,58	62,61	53,76	65,04	55,09	53,61	56,11	68,35	53,48	40,52
NCOTA x conclusão	129,37	92,84	86,82	71,95	55,17	42,00	50,61	52,62	52,03	44,81
NCOTA x eliminação	20,50	8,78	10,26	9,49	5,83	6,43	5,72	4,43	4,22	4,57
NCOTA x trancamentos	123,64	94,46	75,61	86,56	75,44	72,95	73,81	74,78	66,77	70,86
NCOTA x transferência	8,10	2,56	1,38	1,08	1,11	1,53	0,90	0,92	1,21	1,46
NCOTA x sexo feminino	50,33	50,54	53,82	49,80	49,45	51,44	51,08	51,89	52,68	51,10
NCOTA x sexo masculino	49,67	49,46	46,18	50,20	50,55	48,56	48,92	48,11	47,32	48,90
NCOTA x cr-aluno	6,82	6,74	6,71	6,57	6,50	6,45	6,46	6,49	6,51	7,3

NCOTA x CR-em_turma	6,99	6,92	6,83	6,69	6,62	6,52	6,58	6,62	6,63	7,44
NCOTA x EA CEH	40,11	58,91	69,69	69,11	76,60	82,08	80,54	78,49	79,07	83,20
NCOTA x EA CTC	54,38	66,14	77,14	80,44	85,04	84,87	75,35	71,59	71,16	73,00
NCOTA x EA CBI	39,17	58,66	60,18	57,47	62,95	65,27	58,54	53,49	52,31	58,66
NCOTA x EA CCS	39,68	57,69	61,92	60,49	68,66	67,80	62,91	59,86	60,00	63,00
NCOTA x EA CEH x conclusão	42,64	41,32	41,13	39,16	37,88	39,19	40,24	38,73	40,80	38,63
NCOTA x EA CTC x conclusão	22,11	22,55	23,48	24,70	30,27	30,00	26,70	28,84	30,82	28,84
NCOTA x EA CBI x conclusão	11,01	11,05	10,32	10,69	8,75	6,53	10,36	10,48	8,12	10,84
NCOTA x EA CCS x conclusão	24,25	25,08	25,07	25,45	23,10	24,28	22,70	21,96	20,27	21,68
NCOTA x EA CEH x integ 1	3,23	3,76	3,37	2,66	4,42	4,26	3,73	2,59	2,40	4,75
NCOTA x EA CEH x integ 2	51,40	53,29	54,62	55,75	42,93	49,85	53,80	46,26	42,48	45,76
NCOTA x EA CEH x integ 3	29,07	27,39	26,69	26,50	29,67	22,50	24,97	29,80	36,41	34,07
NCOTA x EA CEH x integ 4	16,30	15,57	15,32	15,10	22,98	23,38	17,50	21,36	18,71	15,42
NCOTA x EA CTC x integ 1	4,54	4,88	5,20	4,07	4,10	3,08	3,70	6,04	6,21	6,83
NCOTA x EA CTC x integ 2	36,31	35,29	32,90	37,56	26,03	23,85	29,24	25,27	29,70	29,84
NCOTA x EA CTC x integ 3	33,62	30,70	35,11	31,37	37,54	35,00	31,19	33,52	35,74	38,72
NCOTA x EA CTC x integ 4	25,53	29,12	26,79	27,00	32,33	38,08	35,87	35,16	28,36	24,60
NCOTA x EA CBI x integ 1	2,00	1,47	3,25	2,80	4,37	1,77	2,50	4,55	0,64	1,82
NCOTA x EA CBI x integ 2	83,43	86,22	83,43	82,17	81,42	81,42	79,00	68,18	80,25	69,70
NCOTA x EA CBI x integ 3	9,14	5,28	7,99	9,44	6,01	13,27	16,00	24,75	14,65	24,85
NCOTA x EA CBI x integ 4	5,43	7,04	5,33	5,59	8,20	3,54	2,50	2,53	4,46	3,64
NCOTA x EA CCS x integ 1	4,91	4,24	2,92	3,65	2,26	6,40	6,19	6,73	5,08	5,44
NCOTA x EA CCS x integ 2	54,65	57,12	55,65	53,51	45,17	41,71	44,27	50,00	48,22	57,70
NCOTA x EA CCS x integ 3	19,90	20,28	18,71	19,15	24,02	18,25	21,79	18,03	20,30	19,94
NCOTA x EA CCS x integ 4	20,54	18,36	22,72	23,68	28,54	33,65	27,75	25,24	26,40	16,92
NCOTA x em turma	90,00	89,09	90,34	89,17	88,38	87,97	89,23	88,72	90,02	90,51
NCOTA x aptos	86,36	85,21	86,19	83,63	84,54	84,88	85,78	85,90	87,52	88,09

Tabela 9 – Planilha diagrama 2, drill down - CR por cursos entre cotistas e não cotistas

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
COTA	6,49	6,64	6,76	6,88	6,87	6,77	6,57	6,46	6,43	6,72
COTA x CR x CEH	7,44	7,65	7,67	7,72	7,61	7,45	7,30	7,08	7,12	8,16
COTA x CR x CEH Pedagogia	8,31	8,27	8,32	8,39	8,17	8,07	7,90	7,51	7,57	12,78
COTA x CR x CEH Inglês	5,89	6,60	6,58	6,72	6,65	6,36	6,42	6,35	6,53	6,43
COTA x CR x CEH Port / Alemão	6,37	6,79	6,69	6,99	7,21	6,74	6,32	6,25	5,54	6,01
COTA x CR x CEH Port / Espanhol	7,28	6,93	7,11	7,18	7,08	7,05	6,95	6,72	6,81	6,76
COTA x CR x CEH Port / Francês	5,63	6,42	6,05	6,29	6,73	6,68	6,12	6,19	6,23	6,34
COTA x CR x CEH Port / Grego	6,31	6,28	6,48	6,85	6,75	6,86	6,33	6,28	6,41	6,05
COTA x CR x CEH Port / Italiano	5,84	6,41	6,27	6,63	6,60	6,48	6,49	6,58	6,56	6,91
COTA x CR x CEH Port / Japonês	x	6,07	5,76	6,11	6,75	6,80	6,52	6,24	6,27	6,13
COTA x CR x CEH Port / Latim	7,08	6,51	6,34	6,58	6,68	6,72	6,12	6,61	6,99	6,81
COTA x CR x CEH Port / Literatura	6,81	7,38	7,41	7,50	7,35	7,28	7,04	6,79	6,79	6,58
COTA x CR x CEH Educação Física	7,27	7,45	7,35	7,29	7,29	7,10	7,02	6,87	6,95	6,72
COTA x CR x CEH Psicologia	7,64	7,53	7,99	7,94	7,88	7,81	7,72	7,68	7,70	7,74
COTA x CR x CEH Comunicação Social	8,37	8,31	8,07	7,96	7,95	7,83	7,80	7,59	7,62	7,66
COTA x CR x CEH Ciências Biológicas	7,44	7,94	7,80	7,67	7,29	7,13	6,91	6,56	6,63	6,71
COTA x CR x CEH Geografia	7,60	7,64	7,51	7,57	7,46	7,21	6,91	6,69	6,59	6,64
COTA x CR x CEH História	7,42	7,80	7,57	7,95	7,86	7,90	7,74	7,52	7,41	7,16
COTA x CR x CEH Port / Inglês	6,63	7,71	7,56	7,54	7,54	7,39	7,24	7,19	6,94	6,87
COTA x CR x CEH Port / Literatura	7,46	7,87	7,82	7,75	7,63	7,44	7,13	6,89	7,04	7,07
COTA x CR x CEH Matemática	5,23	5,13	5,40	5,16	5,35	5,30	5,11	4,65	5,02	4,93
COTA x CR x CEH Pedagogia	7,97	8,34	8,38	8,35	8,14	7,88	7,44	7,34	7,14	15,30
COTA x CR x CEH Geografia	x	7,60	7,61	7,74	7,88	7,84	7,76	7,37	7,50	7,20
COTA x CR x CEH Matemática	7,33	6,49	6,77	6,32	6,35	6,34	6,78	6,43	6,34	6,37
COTA x CR x CEH Pedagogia	7,62	7,90	8,12	8,14	8,15	7,57	8,00	7,60	7,69	15,64
COTA x CR x CEH Artes	7,29	7,64	7,76	7,89	7,74	7,43	7,16	6,68	6,18	6,06
COTA x CR x CEH Artes Visuais	6,46	6,46	5,26	7,16	7,32	6,98	6,70	6,67	7,02	7,29
COTA x CR x CEH História da Arte	8,67	x	x	7,45	8,02	7,54	6,70	6,25	6,58	7,13
COTA x CR x CTC	4,45	4,51	4,88	5,18	5,33	5,34	5,23	5,26	5,26	5,38
COTA x CR x CTC Engenharia	4,83	4,73	4,99	5,10	5,24	5,20	5,08	5,13	5,08	5,21
COTA x CR x CTC Cs Computação	4,36	3,28	3,50	3,70	4,22	4,28	4,21	4,20	4,51	4,61
COTA x CR x CTC Cs Atuariais		4,33	4,38	4,61	5,07	4,97	4,80	4,87	4,84	4,86
COTA x CR x CTC Estatística	4,75	4,41	5,14	4,98	4,74	4,52	4,71	3,99	3,89	4,17
COTA x CR x CTC Informática	4,26	4,44	4,65	4,98	5,11	5,21	5,41	5,49	5,34	4,81
COTA x CR x CTC Matemática	2,64	3,12	3,88	4,56	4,81	4,87	4,31	4,63	4,43	4,38
COTA x CR x CTC EngQuímica	4,28	4,20	4,42	4,79	4,89	5,06	5,16	5,42	5,53	5,76
COTA x CR x CTC Química	5,60	4,84	4,68	5,24	4,92	5,30	5,73	5,38	5,45	5,57
COTA x CR x CTC Física	3,93	3,81	4,31	4,65	4,83	4,80	4,45	4,16	4,37	4,60
COTA x CR x CTC Geografia	7,10	7,18	7,58	7,75	7,61	7,44	7,26	6,94	6,52	6,58
COTA x CR x CTC Oceanografia	5,31	6,70	6,05	6,64	7,31	x	x	x	x	x
COTA x CR x CTC Desenho Industrial	x	4,14	5,15	5,71	6,08	6,23	5,90	5,76	5,87	5,97
COTA x CR x CTC Geologia	6,36	6,40	5,84	5,98	6,39	5,86	5,80	5,81	5,71	5,81
COTA x CR x CTC Engenharia	4,99	5,03	5,08	5,07	5,49	5,65	5,77	5,82	5,70	5,86
COTA x CR x CTC Eng de Comput	x	x	x	x	x	3,04	5,37	5,73	5,68	5,73
COTA x CR x CTC Engenharia	5,62	5,11	5,44	5,77	5,67	5,70	5,37	5,04	4,88	5,15
COTA x CR x CTC Geografia	x	5,73	6,54	8,17	7,74	7,30	7,03	7,12	7,05	6,84
COTA x CR x CTC Turismo	x	x	x	x	x	x	x	8,31	7,84	7,08
COTA x CR x CTC Oceanografia	6,29	5,98	5,50	6,09	6,09	6,35	6,26	6,08	6,01	5,87
COTA x CR x CBI	6,94	7,14	7,18	7,12	7,21	7,42	7,15	7,11	7,04	7,04
COTA x CR x CBI Medicina	7,09	7,24	7,09	6,95	7,30	8,11	7,54	7,66	7,66	7,68
COTA x CR x CBI Odontologia	5,88	6,32	6,32	6,17	6,25	6,29	6,26	6,00	6,12	6,32
COTA x CR x CBI Enfermagem	7,89	7,73	7,90	7,88	7,94	7,78	7,74	7,66	7,55	7,49
COTA x CR x CBI Ciências Biológicas	7,19	6,97	6,97	7,05	6,91	6,94	6,63	6,50	6,58	6,44
COTA x CR x CBI Nutrição	6,45	7,12	7,34	7,23	7,12	6,99	6,94	7,04	6,66	6,62

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
COTA x CR x CCS	6,91	6,98	7,03	7,02	7,05	6,91	6,74	6,70	6,70	6,70
COTA x CR x CCS Direito	7,58	7,48	7,54	7,41	7,39	7,20	7,05	7,04	7,06	7,09
COTA x CR x CCS Administração	6,36	6,31	6,70	6,70	6,57	6,45	6,42	6,28	6,13	6,04
COTA x CR x CCS Cs Contábeis	6,41	6,77	6,72	6,76	6,87	6,65	6,45	6,65	6,68	6,63
COTA x CR x CCS Cs Econômicas	5,49	5,55	5,25	5,54	5,48	5,26	5,16	5,15	5,21	5,16
COTA x CR x CCS Serviço Social	7,19	7,59	7,65	7,74	7,86	7,81	7,79	7,48	7,29	7,27
COTA x CR x CCS Ciências Sociais	6,49	7,19	6,94	6,91	6,88	6,86	6,70	6,55	6,60	6,45
COTA x CR x CCS Filosofia	7,15	6,83	6,85	6,57	6,79	6,66	6,26	6,09	5,96	6,22
COTA x CR x CCS História	7,57	7,51	7,74	7,71	7,69	7,52	7,13	7,14	7,23	7,32
NCOTA	6,99	6,92	6,83	6,69	6,62	6,52	6,58	6,62	6,63	7,44
NCOTA x CR x CEH	7,76	7,74	7,64	7,47	7,47	7,30	7,33	7,29	7,24	9,20
NCOTA x CR x CEH Pedagogia	8,27	8,22	8,16	8,07	7,99	7,88	7,91	7,87	7,86	12,67
NCOTA x CR x CEH Inglês	7,49	7,60	7,51	7,49	7,48	7,23	7,31	7,24	7,27	7,22
NCOTA x CR x CEH Port / Alemão	6,74	7,14	7,01	7,04	6,77	6,79	6,59	6,60	7,27	6,78
NCOTA x CR x CEH Port / Espanhol	7,55	7,69	7,57	7,48	7,49	7,10	7,02	7,33	7,22	7,40
NCOTA x CR x CEH Port / Francês	7,13	7,24	7,19	7,16	6,72	6,64	6,56	6,73	6,51	6,67
NCOTA x CR x CEH Port / Grego	6,94	7,28	6,95	7,13	7,12	6,64	6,90	6,61	6,34	6,69
NCOTA x CR x CEH Port / Italiano	6,85	6,82	6,89	7,15	6,78	6,72	6,39	6,77	6,71	6,61
NCOTA x CR x CEH Port / Japonês	x	6,34	6,66	6,47	6,87	6,20	6,67	6,52	6,72	6,72
NCOTA x CR x CEH Port / Latim	6,99	6,96	6,89	6,72	7,02	6,92	6,81	6,64	6,09	6,40
NCOTA x CR x CEH Port / Literatura	7,51	7,75	7,54	7,55	7,63	7,44	7,39	7,42	7,27	7,16
NCOTA x CR x CEH Educação Física	7,51	7,55	7,59	7,44	7,24	7,03	7,18	7,20	6,90	6,70
NCOTA x CR x CEH Psicologia	8,17	8,19	8,07	7,92	8,00	7,91	8,01	8,11	7,98	8,16
	7,85	7,86								
NCOTA x CR x CEH Com. Social	8,12	8,07	7,93	7,96	8,07	7,90	7,82	7,67	7,85	7,86
NCOTA x CR x CEH	7,64	7,50	7,39	7,22	7,19	7,09	6,97	6,92	6,95	8,31
NCOTA x CR x CEH Cs Biológicas	7,87	7,65	7,60	7,44	7,04	7,01	7,01	6,77	6,78	6,88
NCOTA x CR x CEH Geografia	7,44	7,42	7,34	7,24	7,20	6,92	6,84	6,73	6,74	6,61
NCOTA x CR x CEH História	7,72	7,63	7,66	7,70	7,72	7,70	7,65	7,36	7,46	7,34
NCOTA x CR x CEH Port / Inglês	7,97	7,80	7,75	7,48	7,52	7,40	7,09	7,24	7,12	6,96
NCOTA x CR x CEH Port / Literatura	7,80	7,86	7,73	7,68	7,54	7,28	7,17	7,27	7,24	7,13
NCOTA x CR x CEH Matemática	6,07	5,82	5,65	5,30	5,24	5,08	5,00	5,12	5,12	4,89
NCOTA x CR x CEH Pedagogia	8,74	8,50	8,23	7,93	7,99	7,80	7,59	7,51	7,61	15,45
NCOTA x CR x CEH Geografia	x	7,57	7,93	7,53	7,53	7,49	7,38	7,32	7,09	6,86
NCOTA x CR x CEH // Matemática	7,88	7,56	6,72	6,34	6,65	6,58	7,08	7,04	6,87	6,49
NCOTA x CR x CEHPedagogia	8,26	8,26	8,23	7,73	8,10	7,17	7,96	7,59	7,45	14,68
NCOTA x CR x CEH Artes	7,61	7,71	7,80	7,39	7,69	7,32	7,11	6,71	6,42	7,00
NCOTA x CR x CEH Artes Visuais	4,68	5,48	4,95	6,70	7,06	6,81	7,03	7,29	7,24	7,03
NCOTA x CR x CEH História da Arte	3,55	x	x	7,88	7,69	7,78	7,15	7,22	6,67	6,83
NCOTA x CR x CTC	5,65	5,57	5,52	5,44	5,40	5,36	5,40	5,46	5,49	5,57
NCOTA x CR x CTC Engenharia	5,85	5,76	5,59	5,52	5,46	5,41	5,52	5,57	5,61	5,83
NCOTA x CR x CTC Cs Computação	3,88	3,80	4,73	4,48	4,67	4,70	4,77	5,07	5,34	5,34
NCOTA x CR x CTC Cs Atuariais	x	4,86	5,41	5,38	5,43	5,35	5,12	5,09	5,28	5,38
NCOTA x CR x CTC Estatística	4,46	4,43	4,63	4,54	4,54	4,58	4,44	4,48	4,43	4,50
NCOTA x CR x CTC Informática	5,36	5,21	5,15	5,10	5,08	5,16	5,32	5,25	5,00	4,79
NCOTA x CR x CTC Matemática	4,94	4,94	4,85	4,86	5,00	4,85	4,74	4,81	4,55	4,42
NCOTA x CR x CTC Eng Química	6,00	6,01	5,82	5,79	5,77	5,87	6,12	6,22	6,31	6,49
NCOTA x CR x CTC Química	6,14	6,28	6,21	5,82	5,31	5,80	6,08	6,13	6,20	6,11
NCOTA x CR x CTC Física	4,50	4,49	4,70	4,58	4,50	4,46	4,44	4,42	4,40	4,33
NCOTA x CR x CTC Geografia	7,49	7,51	7,39	7,18	6,89	6,78	7,03	6,75	6,50	5,64
NCOTA x CR x CTC Oceanografia	6,83	6,94	6,63	6,59	5,89	5,18	4,43	4,40	3,47	x
NCOTA x CR x CTC Des Industrial	7,60	6,63	5,99	5,77	6,40	5,90	5,97	5,81	6,08	5,93
NCOTA x CR x CTC Geologia	6,37	6,32	6,21	6,17	6,28	6,16	6,20	6,26	6,28	6,61
NCOTA x CR x CTC Engenharia	6,14	6,00	5,76	5,69	5,71	5,83	5,89	6,12	6,22	6,20
NCOTA x CR x CTC Eng Computação	x	x	x	x	x	6,02	5,37	5,45	6,02	6,24
NCOTA x CR x CTC Geografia	5,97	6,44	6,72	7,71	7,85	7,15	6,76	6,97	7,06	6,71
NCOTA x CR x CTC Turismo	x	x	x	x	x	x	x	7,37	7,39	7,07
NCOTA x CR x CTC Oceanografia	6,40	6,50	6,91	6,62	6,52	6,45	6,46	6,47	6,58	6,46

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
NCOTA x CR x CBI	7,99	7,85	8,03	7,99	7,81	7,51	7,59	7,66	7,69	7,58
NCOTA x CR x CBI Medicina	8,78	8,55	8,83	8,92	8,86	7,76	8,22	8,32	8,43	8,32
NCOTA x CR x CBI Odontologia	7,42	7,23	7,38	7,38	7,09	7,22	7,07	7,29	7,31	7,27
NCOTA x CR x CBI Enfermagem	7,94	7,83	8,08	7,91	7,43	7,83	7,82	7,88	7,86	7,86
NCOTA x CR x CBI Cs Biológicas	7,55	7,28	7,27	7,17	7,16	7,08	6,84	6,76	6,97	6,95
NCOTA x CR x CBI Nutrição	7,59	7,67	7,76	7,63	7,45	7,51	7,38	7,28	7,07	6,78
NCOTA x CR x CCS	7,38	7,37	7,26	7,17	7,04	7,04	7,08	7,17	7,18	7,19
NCOTA x CR x CCS Direito	8,19	8,18	8,17	8,17	8,09	8,25	8,28	8,37	8,41	8,48
NCOTA x CR x CCS Administração	6,94	6,69	6,64	6,39	6,25	6,37	6,43	6,53	6,59	6,51
NCOTA x CR x CCS Cs Contábeis	6,92	6,98	6,83	6,48	6,24	6,20	6,40	6,74	6,72	6,76
NCOTA x CR x CCS Cs Econômicas	5,95	6,03	5,77	5,61	5,62	5,52	5,65	5,68	5,71	5,59
NCOTA x CR x CCS Serviço Social	7,64	7,68	7,65	7,67	7,50	7,56	7,66	7,48	7,43	7,43
NCOTA x CR x CCS Cs Sociais	7,10	7,19	7,03	7,03	7,07	6,96	6,84	6,85	6,91	6,68
NCOTA x CR x CCS Filosofia	7,06	6,82	6,71	6,59	6,68	6,60	6,39	6,60	6,42	6,37
NCOTA x CR x CCS História	7,53	7,57	7,47	7,43	7,32	7,24	7,18	7,10	7,31	7,46

Tabela 10 – Planilha diagrama 3: ocupação de cotistas e não cotistas por curso (%)

cotistas por curso	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
cotista total	60,47	43,24	33,08	32,09	23,65	22,62	28,00	30,39	30,81	28,04
CEH	64,80	43,83	31,44	31,32	21,27	17,07	18,64	20,25	20,64	16,56
Artes	68,00	48,00	35,00	21,00	27,14	12,00	31,43	28,57	24,29	21,43
Ciências Biológicas	58,75	45,00	25,00	28,75	18,75	16,25	17,50	22,50	32,50	20,00
Comunicação Social	71,25	45,00	46,25	46,25	49,00	48,00	49,00	47,00	48,00	48,00
Educação Física	61,67	50,00	38,33	28,33	21,67	9,17	10,00	17,50	16,67	7,50
Geografia	50,56	45,56	45,56	38,33	17,78	20,56	19,44	20,56	24,44	16,11
História	67,50	45,00	43,75	46,25	35,00	33,75	33,75	20,00	30,00	13,75
História da Arte	0,00	0,00	0,00	0,00	0,00	x	16,67	26,67	23,33	13,33
Letras - Inglês	67,14	47,14	32,86	38,57	12,86	14,29	32,86	31,43	30,00	24,29
Letras - Port / Alemão	55,00	20,00	15,00	5,00	15,00	5,00	25,00	5,00	10,00	4,00
Letras - Port / Espanhol	79,55	50,00	45,45	47,73	47,73	43,18	29,55	50,00	38,64	38,64
Letras - Port / Francês	65,00	37,50	45,00	15,00	25,00	5,00	12,50	27,50	10,00	7,50
Letras - Port / Grego	70,00	50,00	10,00	50,00	30,00	0,00	10,00	30,00	6,67	6,67
Letras - Port / Inglês	35,00	21,25	5,00	13,13	8,13	5,63	3,75	5,00	6,25	6,88
Letras - Port / Italiano	75,00	34,09	47,73	27,27	20,45	15,91	11,36	11,36	13,64	4,55
Letras - Port / Japonês	0,00	46,67	13,33	33,33	20,00	6,67	33,33	13,33	33,33	13,33
Letras - Port / Latim	75,00	45,00	35,00	40,00	35,00	5,00	35,00	25,00	0,00	25,00
Letras - Port / Literatura	171,43	114,29	87,14	88,57	62,86	64,29	52,86	45,71	54,29	38,57
Matemática	36,67	35,00	12,50	10,83	4,17	3,33	5,83	9,17	6,67	4,17
Pedagogia	61,61	40,89	25,71	30,50	13,04	9,46	9,11	12,68	12,68	11,25
Pedagogia ,Séries Iniciais	100,00	46,67	16,67	14,17	9,17	4,17	5,00	7,50	7,50	7,50
Psicologia	64,58	45,83	45,83	45,83	47,92	45,36	48,96	47,92	47,92	47,92
CTC	48,53	39,20	26,29	22,18	16,56	17,76	28,71	31,52	32,69	31,69
info + ciencia comput	90,83	60,00	58,33	59,17	35,83	20,83	29,17	41,67	30,00	24,17
Ciências Atuariais	0,00	8,33	17,50	8,33	5,56	4,44	6,67	6,67	11,11	8,89
Desenho Industrial	88,57	45,71	45,71	45,71	45,71	45,71	45,71	45,71	45,71	45,00
Engenharia	65,40	46,60	29,80	26,60	20,40	19,60	43,33	48,24	56,08	51,37
Engenharia da Comput	x	x	x	x	x	4,00	36,00	36,00	48,00	24,00
Engenharia Química	65,00	47,50	42,50	36,25	45,00	45,00	47,50	47,50	45,00	46,67
Estatística	25,83	24,17	5,83	6,67	0,00	6,67	2,22	4,44	7,78	10,00
Física	50,56	38,20	14,04	11,24	3,37	2,81	9,55	7,87	7,87	8,43
Geografia	50,00	45,71	47,14	45,71	47,14	47,14	45,71	45,71	44,29	45,71
Geologia	36,67	23,33	23,33	26,67	10,00	36,67	46,67	50,00	46,67	50,00
Matemática	56,50	45,50	22,50	10,50	5,50	4,50	9,00	9,00	5,00	8,00
Oceanografia	47,50	27,50	45,00	35,00	12,50	12,50	17,50	25,00	22,50	25,00
Química	55,00	45,00	37,50	32,50	22,50	32,50	27,50	40,00	35,00	32,00
Turismo	x	x	x	x	x	x	x	15,00	12,50	12,50
CBI	65,55	47,13	44,74	46,65	38,76	39,00	45,93	47,13	48,09	42,08
Ciências Biológicas	46,88	45,83	46,88	51,04	50,00	35,42	48,96	48,96	47,92	38,33
Enfermagem	67,50	45,00	45,00	43,75	28,75	41,25	45,00	45,00	45,00	40,00
Medicina	79,79	45,74	45,74	45,74	45,74	45,74	45,74	50,00	46,81	46,81
Nutrição	75,00	45,45	45,45	43,18	25,00	25,00	36,36	39,77	47,73	30,68
Odontologia	63,33	46,67	26,67	38,33	35,00	43,33	40,00	46,67	46,67	50,00
CCS	67,63	46,40	41,29	42,04	31,93	33,68	37,92	41,39	40,66	38,83
Administração	76,67	45,00	45,00	45,00	25,00	42,50	44,17	44,17	45,00	45,00
Ciências Contábeis	64,17	45,00	41,67	42,50	12,50	18,33	35,00	44,17	45,00	45,83
Ciências Econômicas	59,29	42,86	29,29	30,71	9,29	16,43	30,00	32,14	30,71	35,00
Ciências Sociais	69,00	47,00	45,00	45,00	44,00	26,00	27,00	46,00	46,00	21,00
Direito	69,23	46,15	46,15	46,15	46,15	46,96	46,47	46,47	46,47	48,40
Filosofia	65,00	46,00	25,00	32,00	15,00	12,00	21,00	21,00	10,00	13,00
História	72,28	53,47	46,00	44,00	46,00	46,00	46,00	50,00	46,00	42,00
Serviço Social	63,16	48,68	47,37	47,37	44,74	43,42	38,16	39,00	46,00	39,00

não cotista por curso	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
não cotista	49,93	67,69	77,02	76,14	78,52	82,68	75,83	70,16	72,12	76,50
CEH	42,76	62,06	71,97	70,35	69,78	78,78	76,62	73,74	77,99	82,03
Artes	38,00	56,00	66,00	63,00	82,86	70,00	75,71	75,71	90,00	88,57
Ciências Biológicas	58,75	55,00	87,50	73,75	83,75	82,50	76,25	72,50	76,25	85,00
Comunicação Social	45,00	78,75	66,25	68,75	49,00	68,00	58,00	53,00	58,00	67,00
Educação Física	75,00	98,33	91,67	75,00	75,00	92,50	95,00	83,33	105,00	112,50
Geografia	28,89	54,44	57,78	64,44	83,89	80,00	81,11	81,67	78,33	85,00
História	38,75	60,00	63,75	56,25	67,50	67,50	85,00	71,25	75,00	82,50
História da Arte	x	x	x	x	46,67	x	90,00	73,33	86,67	106,67
Letras - Inglês	48,57	57,14	71,43	67,14	85,71	88,57	71,43	67,14	82,86	85,71
Letras - Port / Alemão	65,00	90,00	80,00	90,00	85,00	80,00	75,00	125,00	120,00	120,00
Letras - Port / Espanhol	54,55	63,64	63,64	56,82	52,27	63,64	90,91	61,36	77,27	70,45
Letras - Port / Francês	55,00	67,50	60,00	87,50	80,00	95,00	95,00	75,00	92,50	102,50
Letras - Port / Grego	70,00	60,00	90,00	50,00	70,00	100,00	80,00	80,00	60,00	106,67
Letras - Port / Inglês	20,00	29,38	46,25	41,25	45,63	44,38	45,00	41,25	43,13	41,88
Letras - Port / Italiano	43,18	70,45	54,55	72,73	79,55	88,64	84,09	84,09	93,18	90,91
Letras - Port / Japonês	6,67	53,33	86,67	66,67	80,00	100,00	80,00	93,33	66,67	93,33
Letras - Port / Latim	40,00	60,00	65,00	65,00	70,00	85,00	70,00	85,00	105,00	75,00
Letras - Port / Literatura	82,86	130,00	142,86	134,29	158,57	152,86	171,43	160,00	167,14	191,43
Matemática	54,17	65,83	85,00	71,67	58,33	61,67	78,33	72,50	58,33	77,50
Pedagogia	34,11	52,14	67,86	65,96	64,82	74,46	73,04	76,07	77,50	77,14
Pedagogia - Séries Iniciais	61,67	92,50	90,00	105,00	71,67	95,83	73,33	71,67	75,83	66,67
Psicologia	46,88	66,67	79,17	64,58	48,96	64,95	47,92	40,63	54,17	55,21
CTC	64,71	78,67	93,48	96,09	100,75	99,67	86,87	77,84	79,72	84,19
info + ciencia comput	40,83	82,50	95,00	102,50	135,83	85,83	67,50	50,83	67,50	78,33
Ciências Atuariais	0,00	70,83	90,83	96,67	102,22	106,67	93,33	86,67	90,00	91,11
Desenho Industrial	17,14	54,29	57,14	60,00	54,29	54,29	54,29	57,14	54,29	55,00
Engenharia	87,80	87,60	111,80	115,00	117,00	117,80	94,31	84,90	83,92	82,75
Engenharia de Computação	x	x	x	x	x	100,00	64,00	56,00	36,00	80,00
Engenharia Química	57,50	67,50	66,25	83,75	58,75	76,25	73,75	55,00	55,00	74,44
Estatística	75,83	82,50	80,83	95,00	88,89	94,44	95,56	92,22	93,33	90,00
Física	66,85	76,97	82,02	81,46	89,33	97,19	89,33	76,97	86,52	95,51
Geografia	48,57	62,86	81,43	67,14	65,71	67,14	75,71	70,00	95,71	82,86
Geologia	83,33	83,33	93,33	80,00	140,00	83,33	80,00	70,00	50,00	53,33
Matemática	57,50	70,00	89,00	89,50	89,50	94,00	86,50	90,50	81,00	95,00
Oceanografia	87,50	80,00	102,50	87,50	107,50	125,00	120,00	100,00	90,00	87,50
Química	82,50	85,00	77,50	67,50	72,50	80,00	67,50	37,50	60,00	68,00
Turismo	x	x	x	x	x	x	x	72,50	75,00	87,50
CBI	43,06	64,83	65,07	60,77	63,40	71,05	61,48	53,11	51,44	57,47
Ciências Biológicas	50,00	87,50	83,33	66,67	58,33	112,50	77,08	42,71	48,96	50,83
Enfermagem	40,00	56,25	58,75	56,25	77,50	60,00	50,00	51,25	52,50	56,25
Medicina	28,72	63,83	55,32	57,45	55,32	53,19	55,32	70,21	54,26	58,51
Nutrição	39,77	53,41	54,55	60,23	56,82	64,77	61,36	52,27	48,86	64,77
Odontologia	63,33	58,33	75,00	63,33	75,00	56,67	61,67	46,67	53,33	60,00
CCS	44,62	63,42	67,42	64,23	70,04	70,91	64,33	61,72	60,99	66,12
Administração	38,33	59,17	65,00	58,33	68,33	57,50	56,67	57,50	56,67	56,67
Ciências Contábeis	45,83	60,00	60,00	60,00	90,00	79,17	72,50	48,33	59,17	58,33
Ciências Econômicas	53,57	58,57	76,43	77,86	87,86	91,43	71,43	67,14	71,43	67,14
Ciências Sociais	40,00	55,00	60,00	63,00	53,00	70,00	71,00	55,00	53,00	78,00
Direito	33,33	60,26	59,62	55,77	57,37	55,91	57,69	55,45	54,81	57,69
Filosofia	44,00	76,00	89,00	69,00	85,00	99,00	80,00	101,00	91,00	107,00
História	69,31	62,38	79,00	76,00	70,00	72,00	52,00	62,00	52,00	54,00
Serviço Social	56,58	93,42	64,47	69,74	63,16	65,79	64,47	62,00	60,00	71,00