



UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PÓS-GRADUAÇÃO EM MATEMÁTICA
MESTRADO EM PROFISSIONAL EM MATEMÁTICA - PROFMAT

**Uso do Software R como complemento para o ensino
de Estatística no Ensino Médio**

Pedro Soares de Brito Neto

Teresina/PI - 2016

Pedro Soares de Brito Neto

Dissertação de Mestrado:

**Uso do Software R como complemento para o ensino de
Estatística no Ensino Médio**

Dissertação submetida à Coordenação Acadêmica Institucional do Programa de Mestrado Profissional em Matemática em Rede Nacional na Universidade Federal do Piauí, oferecido em associação com a Sociedade Brasileira de Matemática, como requisito parcial para obtenção do grau de Mestre em Matemática.

Orientadora:

Prof. Dra. Valmária Rocha da Silva Ferraz

Teresina - 2016

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca Setorial do CCN

B862u Brito Neto, Pedro Soares de.

 Usu do software R como complemento para o ensino de estatística no ensino médio / Pedro Soares de Brito Neto. – Teresina, 2016.
 69f.

 Dissertação (Mestrado Profissional) – Universidade Federal do Piauí, Centro de Ciências da Natureza, Pós-Graduação em Matemática, 2016.

 Orientadora: Profa. Dra. Valmária Rocha da Silva Ferraz.

 1. Matemática Estatística. 2. Estatística Descritiva. 3. Estatística – Ensino – Aprendizagem. I. Título

CDD 519.53



PROFMAT



UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
CENTRO DE EDUCAÇÃO ABERTA E À DISTÂNCIA
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL



Dissertação de Mestrado submetida à coordenação Acadêmica Institucional, na Universidade Federal do Piauí, do Programa de Mestrado Profissional em Matemática em Rede Nacional para obtenção do grau de **mestre em matemática** intitulada: **Uso do Software R como complemento para o ensino de Estatística no Ensino Médio**, defendida por **Pedro Soares de Brito Neto** em **20/07/2016** e aprovada pela banca constituída pelos professores:

Valmária R. S. Ferraz

Prof. Dra. Valmária Rocha da Silva Ferraz (UFPI) - Presidente

Fernando F. Nascimento

Prof. Dr. Fernando Ferraz do Nascimento (UFPI) - Examinador

Valtercio de Almeida Carvalho

Prof. Ms. Valtercio de Almeida Carvalho (IFPI) - Examinador Externo

Dedico esse trabalho a toda minha família e em especial ao meu pai Pedro Filho, a minha mãe Maria Briolanja e a minha amada esposa Érica Verusca.

Agradecimentos

Agradeço a Deus por me abençoar, por sua preseça e por permitir que concluisse dessa dissertação.

Agradeço aos meus pais Pedro Filho e Maria Briolanja pelo esforço e dedicação em minha criação e por todo amor que recebo de vocês em todos os momentos de minha vida.

Agradeço à minha querida e amada esposa Érica Verusca, pelo companheirismos, carinho, amor e dedicação nesses anos que estamos juntos.

Agradeço aos meus irmãos Paulo Rainan e Sara Simone pelo companheirismos e torcida pelo meu sucesso.

Agradeço a toda minha família que sempre torceram por mim e entenderam o meu afastamento temporário para me dedicar a este trabalho.

Agradeço à professora Dr^a Valmária Rocha da Silva Ferraz pela paciência e dedicação em sua orientação e incentivo que tornaram possível a conclusão dessa dissertação

Agradeço a todos os professores do PROFMAT que se dedicam ao programa e tornaram possível a obtenção desse título.

Agradeço à SBM e à CAPES pelo apoio ao longo da pós-graduação.

Agradeço às diretoras Socorro Mendes e Irene Rocha das escolas Lucídio Portella e Odolfo Medeiros, respectivamente, pela autorização e apoio na realização desse projeto, bem como aos alunos dessas escolas que se dedicaram ao curso e foram essencias para o seu sucesso.

Agradeço à minha afilhada Ellen Fernanda que amo muito e, mesmo sem compreender ainda, é para mim uma grande fonte de energia e inspiração.

Agradeço aos meus sogros Irandi e Maria das Dôres, aos amigos e a todos os que torcem por meu sucesso e felicidade.

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes”.

Marthin Luther King

Resumo

Este trabalho apresenta como proposta o uso de uma ferramenta tecnológica, o Software R, como auxílio ao ensino de Estatística para os estudantes da 3ª série do Ensino Médio. O uso desse programa em laboratório tem por objetivo complementar os conteúdos ministrados em sala de aula e assim ajudar no processo ensino-aprendizagem da Matemática, mais especificamente de Estatística. Neste trabalho encontra-se uma proposta de apostila para a utilização do R relacionando-o aos conteúdos propostos para Estatística no Ensino Médio, assim como um relatório com os resultados da aplicação do uso do software elaborado a partir de um questionário respondido pelos próprios alunos.

Palavras-chave: Software R, Estatística, Ensino-aprendizagem.

Abstract

This paper presents proposal as the use of a technological tool, R Software, as an aid to Statistics teaching for students of the 3rd year of High school . The use of the laboratory program intends to complement the contents taught in the classroom and thus help in the teaching and learning of mathematics, specifically Statistics. This work is a handout proposal for the use of R relating to the proposed content for Statistics in high school, as well as a report with the results of software application usage drawn from a questionnaire answered by the students.

Keywords: Software R, Statistics, Teaching-learning.

Sumário

Resumo	iv
Abstract	v
Introdução	1
1 Informações Preliminares	3
2 Fundamentação Teórica	7
2.1 Conceitos Básicos	7
2.2 Tabelas e Gráficos	9
2.2.1 Tabelas de Frequências	9
2.2.2 Gráfico de Linha (ou Segmentos)	11
2.2.3 Gráfico de Barras	12
2.2.4 Histograma	13
2.2.5 Gráfico de setores (Pizza)	14
2.3 Medidas de Posição	15
2.4 Medidas de Dispersão	19
3 Aplicando o Software R	21
3.1 Preparando o Laboratório	21
3.2 Operações Básicas	22
3.3 Armazenando dados	24
3.3.1 Criando um objeto	24
3.3.2 Criando Vetores	25
3.3.3 Criando Matrizes	26
3.3.4 Construindo Data frame	27

Sumário	vii
3.4 Estatística Descritiva	30
3.4.1 Produzindo uma tabela para ser analisada	30
3.4.2 Medidas de Posição	31
3.4.3 Medidas de Dispersão	34
3.4.4 Representação Gráfica	35
4 Relatório e Resultados	48
Considerações Finais	53
Referências Bibliográficas	55
Anexo I	57

Introdução

O ensino nas escolas públicas é alvo de preconceito por boa parte da comunidade, sendo tratado como de baixa qualidade e pouca eficiência. E avaliações nacionais como a Prova Brasil mostram que realmente o nível de conhecimento dos alunos de escola pública está abaixo do satisfatório.

Sabemos que o resultado das avaliações de Matemática geralmente são piores em relação a outras disciplinas e isso geralmente se dá pela complexidade da matéria conciliada com uma aula extremamente tradicionalista e que não tem mais efeito sobre o jovem atual.

Diante de tantas tecnologias que os rodeiam e a velocidade que elas trabalham, os alunos sentem-se desestimulados com o estilo tradicionalista de exclusividade das aulas expositivas. Eles querem dinamismo, trabalhar com ferramentas e situações próximas de sua realidade.

Esta dissertação tem como objetivo propor uma ferramenta tecnológica, que é o **Software R**, para o auxílio no ensino de Estatística na disciplina Matemática para o Ensino Médio. Irá ser proposto o uso desse software estatístico em que os alunos poderão utilizar os computadores do laboratório da escola e utilizar dados próximos a eles para assim complementar e fixar os conteúdos estudados em sala de aula.

A dissertação é dividida em quatro capítulos. No primeiro será exposto algumas informações preliminares onde analisa-se alguns trechos dos PCN e PCNEM, como a proposta curricular para o ensino de Estatística no Ensino Médio, serão citados alguns motivos de se trabalhar corretamente o ensino da Estatística e será apresentado o software R.

No segundo capítulo será apresentado o conteúdo a ser ministrado em sala de aula, que são os conceitos, definições, cálculos de medidas e onde será apresentado e analisado as principais formas de organizar dados, que são as tabelas de frequência e gráficos.

O terceiro capítulo consiste na aplicação do R em laboratório com propostas de uso desse software pelos alunos.

O quarto capítulo finaliza esta dissertação com um relatório do curso que foi ministrado em duas escolas públicas estaduais bem como análise dos resultados obtidos através de um questionário respondido pelos alunos relacionado ao seu aprendizado.

A partir dos resultados desse questionário pode-se conferir a eficácia da estratégia proposta.

Capítulo 1

Informações Preliminares

Uma das maiores dificuldades dos professores em sala de aula, principalmente os de matemática, está relacionada a certa falta de dedicação e disciplina no estudo por parte dos alunos. Isso porque os jovens, atualmente, diferentemente daqueles que estudavam há alguns anos, estão cada vez mais dinâmicos, sempre querendo saber o porquê de tudo e demonstram maior interesse em conteúdos que podem ver de imediato sua utilização e aplicabilidade. Quem nunca ouviu a pergunta: “Onde vou utilizar isso professor?”. Em outras palavras, o que temos é a ineficiência do ensino de matemática através unicamente da tradicional aula expositiva em que o professor fala e o aluno escuta. Deve-se “adaptar-se” a essa nova modalidade de aluno, mudar as técnicas de ensino. É preciso estimulá-los através das tecnologias e aproximar as situações-problema estudadas ao seu ambiente sociocultural.

Os Parâmetros Curriculares para o Ensino Médio (PCNEM) dizem:

“Não somente em Matemática, mas até particularmente nessa disciplina, a resolução de problemas é uma importante estratégia de ensino. Os alunos, confrontados com situações-problema, novas mas compatíveis com os instrumentos que já possuem ou que possam adquirir no processo, aprendem a desenvolver estratégia de enfrentamento, planejando etapas, estabelecendo relações, verificando regularidades, fazendo uso dos próprios erros cometidos para buscar novas alternativas; adquirem espírito de pesquisa, aprendendo a consultar, a experimentar, a organizar dados, a sistematizar resultados, a validar soluções; desenvolvem sua capacidade de raciocínio, adquirem auto-confiança e sentido de responsabilidade; e, finalmente, ampliam sua autonomia e capacidade de comunicação e de argumentação.”
(PCNEM 2000, p. 52)

Os PCNEM reforçam assim a necessidade trabalhar o ensino de matemática de forma

mais dinâmica, com experimentos, análise de resultados, fazer com que o aluno coloque a “mão na massa”, para que construa seu conhecimento através de experiências vividas por ele.

A Estatística está presente em muitas das áreas do conhecimento e também no dia-a-dia de todos, isso torna o conteúdo de Estatística um dos melhores a se ministrar para os alunos do Ensino Médio, justamente por ela estar à vista deles seja em qual área eles identifiquem-se mais. Os alunos devem ser estimulados a identificar o uso de Estatística seja em uma simples tabela com a distribuição do total da população brasileira por regiões, em uma notícia no jornal a respeito de uma complexa análise financeira de riscos de investimentos por uma determinada empresa ou ainda nos gráficos de uma partida online que eles costumam jogar em computadores ou celulares. E além da visualização, precisa-se trabalhar também a interpretação, análise, previsão e tomada de decisão através dos dados estatísticos.

Segundo os PCN+ (Orientações Educacionais Complementares aos Parâmetros Curriculares Nacionais):

“Uma das grandes competências propostas pelos PCNEM diz respeito à contextualização sociocultural como forma de aproximar o aluno da realidade e fazê-lo vivenciar situações próximas que lhe permitam reconhecer a diversidade que o cerca e reconhecer-se como indivíduo capaz de ler e atuar nesta realidade. A Matemática do ensino médio pode ser determinante para a leitura das informações que circulam na mídia e em outras áreas do conhecimento na forma de tabelas, gráficos e informações de caráter estatístico. Contudo, espera-se do aluno nessa fase da escolaridade que ultrapasse a leitura de informações e reflita mais criticamente sobre seus significados. Assim, o tema proposto deve ir além da simples descrição e representação de dados, atingindo a investigação sobre esses dados e a tomada de decisões.” (PCN+ 2007, p. 126)

Os professores devem trabalhar essa contextualização sociocultural proposta pelos PCNEM e com certeza irão superar a desmotivação e desatenção dos alunos, pois estes estarão vendo algo que os cerca tanto no âmbito escolar como social.

Os conteúdos e habilidades propostos pelos PCN+, para o ensino de Estatística no Ensino Médio, são:

“Estatística: descrição de dados; representações gráficas; análise de dados: médias, moda e mediana, variância e desvio padrão.

- *Identificar formas adequadas para descrever e representar dados numéricos e informações de natureza social, econômica, política, científico-tecnológica ou abstrata.*
- *Ler e interpretar dados e informações de caráter estatístico apresentados em diferentes linguagens e representações, na mídia ou em outros textos e meios de comunicação.*
- *Obter médias e avaliar desvios de conjuntos de dados ou informações de diferentes naturezas.*
- *Compreender e emitir juízos sobre informações estatísticas de natureza social, econômica, política ou científica apresentadas em textos, notícias, propagandas, censos, pesquisas e outros meios.”* PCN+(2007, p. 127)

Através destes conteúdos ministrados e essas habilidades assimiladas pelos alunos espera-se um imenso desenvolvimento cognitivo dos mesmos, tornando-os mais críticos em relação ao ambiente que vivem e capazes de analisarem as possibilidades para tomar a melhor decisão em uma situação proposta.

Outro ponto que torna a Estatística um dos conteúdos a ser tratado com muito cuidado e dedicação no Ensino Médio pelos professores de Matemática é o grande índice de questões, envolvendo esse conteúdo, no Exame Nacional do Ensino Médio (ENEM). O ENEM vem dando grande importância aos estudos estatísticos, sendo um dos conteúdos mais abordados em questões do exame. E como hoje a maior porta de entrada para o Ensino Superior no Brasil é o ENEM, a Estatística deve ser trabalhada de forma especial para que os alunos compreendam todos os tópicos ministrados e assim alcancem o acesso a uma Instituição de Ensino Superior e assim continuar sua caminhada acadêmica.

Como a Estatística é um conteúdo que trabalha muito o abstrato do aluno torna-se um conteúdo muitas vezes de difícil assimilação por eles. Isso porque na maioria dos casos não basta fazer o cálculo, também é preciso fazer uma análise minuciosa do contexto do problema e essa parte é a de maior dificuldade de compreensão por parte do alunado, acarretando um maior cuidado e muita dedicação no trabalho não só dos professores de matemática, mas de todas as disciplinas. Trabalha-se hoje com um jovem imediatista, que não para pra pensar e interpretar as informações do problema, que não tem o hábito da leitura e quer, no caso da Matemática, ir direto ao cálculo sem antes analisar o importante contexto da questão proposta, na maioria das vezes.

Nesse momento é onde entra a ferramenta tecnológica, proposta aos professores de matemática, o **Software R** que proporciona agilidade na parte de organização de dados

e cálculos estatísticos assim como diversidade em opções de demonstração dos dados (gráficos), o que acarreta um ganho de tempo para que o professor possa trabalhar a parte contextual do problema tendo assim um resultado final do processo ensino-aprendizagem mais satisfatório. Sem falar, que o uso de uma ferramenta nova desencadeia nos alunos um sentimento de curiosidade e por ser tecnológica, a vontade de manusear é ainda mais intensa e com isso pode-se criar uma expectativa positiva em relação ao plano de uso do software .

O Software R é de domínio público, livre, ou seja, você pode baixar gratuitamente e está disponível em versões para os variados sistemas operacionais como Windows e Linux, que são geralmente os encontrados nas escolas. O seu download é feito em <http://cran.r-project.org/> onde escolhe-se a versão de acordo com o sistema operacional.

A apostila *Mini Curso de Estatística Básica: Introdução ao software R* do Programa de Educação Tutorial - Engenharia Elétrica da Universidade Federal de Santa Maria [10] define o R assim:

“R é uma série integrada de instalações de softwares para manipulação de dados, cálculo e exibição gráfica. Dentre outras coisas, possui:

- *uma manipulação de dados eficaz e facilidade de armazenamento;*
- *uma série de operadores para cálculos com arranjos, especialmente matrizes;*
- *uma extensa, coerente e integrada coleção de ferramentas intermediárias para análise de dados;*
- *instalações gráficas para análises de dados e exibição tanto direta no computador quanto para cópia permanente (impressões);*
- *uma bem desenvolvida, simples e eficaz linguagem de programação (chamada ‘S’) a qual inclui condições, loops, funções recursivas definidas pelo usuário e instalações de entradas e saídas.”*

No R utiliza-se uma linguagem de programação em inglês, mas isso não deve ser encarado como ponto negativo, ao contrário, deve-se levar em consideração o mundo globalizado de hoje e o uso constante de smartphones, tablets, jogos online em que quase sempre os jovens usam a língua inglesa, portanto com o uso do R os alunos podem expandir o seu vocabulário, o que acaba sendo algo positivo para eles.

Deve ser ressaltado que a finalidade do uso do R neste trabalho não é ensinar Estatística com ele e sim utilizá-lo como uma ferramenta para complementar e auxiliar o ensino da disciplina acarretando uma aprendizagem mais consistente e sólida pelos alunos.

Capítulo 2

Fundamentação Teórica

Fazer uma pesquisa, seja para qual finalidade for, é uma atividade rotineira no dia-a-dia das pessoas. Seja uma pesquisa de preços de um mesmo produto em lojas diferentes ou uma procura por um ponto comercial para abrir um estabelecimento qualquer, etc. Geralmente uma pesquisa é realizada para ajudar a tomar uma decisão ou compreender uma situação. Isso é possível através das informações que são recolhidas. Pois aí está o princípio da Estatística.

Manoel Paiva [5] define Estatística assim:

“Detalhando, a Estatística é uma ciência fundamentada na teoria das probabilidades e em um conjunto de técnicas e métodos de pesquisa que abrange, entre outros temas: planejamento de experimentos, coleta e organização de dados, representação de dados numéricos por meio de tabelas e gráficos, análise de dados, previsões e tomadas de decisões em situações de incerteza com base na análise de dados. Ela pode ser aplicada nas mais diversas áreas do conhecimento, desde pesquisas sobre temas do cotidiano até pesquisas científicas.”(2013, p. 8)

2.1 Conceitos Básicos

Em uma pesquisa estatística nem sempre é possível consultar a todos os envolvidos, que chamamos **população estatística** ou **universo estatístico**, por exemplo, caso queira saber a média de filhos dos casais brasileiros, é quase impossível que consiga entrevistar a todos os casais brasileiros, portanto deve-se fazer uma pesquisa por **amostragem**, ou seja, selecionar uma certa quantidade de casais(chamamos **amostra**) e a partir desses dados específicos generalizar os resultados. Cada elemento que compõe a amostra é chamado **indivíduo** ou **objeto**. No exemplo descrito, os indivíduos são os casais.

A **amplitude** de uma amostra corresponde à diferença entre o maior valor e o menor valor obtidos na pesquisa. Por exemplo, se foi feita uma pesquisa perguntando a altura, em metros, de algumas pessoas e obtendo a amostra: 1,65 - 1,68 - 1,65 - 1,70 - 1,66 - 1,72 - 1,57 - 1,69 - 1,68 - 1,67. Observa-se a maior e a menor altura encontrada que são 1,72m e 1,57m, daí a amplitude da amostra é $1,72\text{m} - 1,57\text{m} = 0,15\text{m}$. Isso significa que entre as pessoas pesquisadas a maior diferença entre suas alturas é de 15cm.

Nos dois exemplos anteriores, onde pesquisa-se a média de filhos dos casais brasileiros e a altura das pessoas chama-se a *quantidade de filhos* e a *altura* de **variáveis aleatórias**, isto é, são os atributos pesquisados de cada indivíduo. As variáveis podem ser **quantitativas**, que são aquelas expressas por valores numéricos e subdividem-se em *discretas* (números inteiros) e *contínuas* (números reais), ou **qualitativas**, que são aquelas expressas por características que não podem ser contadas ou medidas (qualidades). Nos dois exemplos citados, as variáveis eram quantitativas, sendo a quantidade de filhos uma variável quantitativa discreta e a altura uma variável quantitativa contínua. Em uma pesquisa sobre a nacionalidade das pessoas (brasileiro, argentino, ...) ou a cor de seus olhos (castanho, verde, ...) tem-se variáveis qualitativas.

Cada resultado diferente de uma amostra é denominado **classe** e a quantidade de vezes que esta classe se repete é chamada **frequência absoluta**. A razão entre a frequência de uma classe e a soma da frequência de todas as classes é chamada **frequência relativa** e geralmente é representada na forma de porcentagem. Tem-se também a **frequência acumulada** que é a soma das frequências das classes anteriores e a frequência da classe atual. Para melhor compreensão observe a amostra a seguir com o resultado de uma pesquisa da idade, em anos, de algumas pessoas:

19 20 18 19 21 18 25 20 20 23 20 21 20 18 21 20 18 19 25 20

Nessa amostra cada valor de idade representa uma classe e a quantidade de vezes que cada idade repete é a frequência dessa idade, por exemplo, a frequência da idade 18 anos é 4. Já a frequência relativa da idade 18 anos, observando que 20 é o total de resultados, será a razão $\frac{4}{20} = 20\%$.

Os dados mostrados podem ser organizados em **rol** que é a simples organização desses dados em ordem crescente ou decrescente:

18 18 18 18 19 19 19 20 20 20 20 20 20 20 21 21 21 23 25 25

As principais formas de organizar e expor os dados coletados de uma pesquisa são as tabelas de frequências e os gráficos.

2.2 Tabelas e Gráficos

2.2.1 Tabelas de Frequências

Para montar uma tabela de frequência pode-se tomar como fonte uma pesquisa a um grupo de alunos fictícios onde foi perguntado a idade(em anos), a altura(em metros), e a cor dos olhos de cada um. Sendo obtido os seguintes resultados:

Ana(16/1,54/castanho); André(18/1,65/castanho) ; Paulo(17/1,71/azul);
 Carla(16/1,64/castanho) ; Felipe(16/1,67/preto) ; Pedro(18/1,76/verde);
 Letícia(17/1,59/castanho) ; Antônia(15/1,49/preto) ; João(19/1,88/verde);
 Maria(16/1,62/castanho); Fátima(18/1,68/verde) ; Israel(17/1,73/azul);
 Nara(15/1,55/castanho) ; Ellen(16/1,61/preto) ; Érica(18/1,77/Castanho);
 Flávia(16/1,65/castanho) ; Lara(18/1,47/preto) ; José(17/1,78/azul);
 Sara(16/1,53/castanho) ; Fabrício(17/1,67/castanho).

Observe a tabela de frequência da variável idade, que é uma variável quantitativa discreta, onde cada número inteiro correspondente a uma idade será uma classe.

IDADE (anos)	FREQUÊNCIA ABSOLUTA (nº de alunos)	FREQUÊNCIA ABSOLUTA ACUMULADA	FREQUÊNCIA RELATIVA	FREQUÊNCIA RELATIVA ACUMULADA
15	2	2	$\frac{2}{20} = 0,1 = 10\%$	10%
16	7	9	$\frac{7}{20} = 0,35 = 35\%$	45%
17	5	14	$\frac{5}{20} = 0,25 = 25\%$	70%
18	5	19	$\frac{5}{20} = 0,25 = 25\%$	95%
19	1	20	$\frac{1}{20} = 0,05 = 5\%$	100%
TOTAL	20	-----	100%	-----

Figura 2.1: Tabela de frequência da variável idade

A tabela de frequência da variável altura, que é uma variável quantitativa contínua, está exposta a seguir. Para esse tipo de variável as classes são formadas por intervalos. No caso da pesquisa, a amostra foi dividida em intervalos de 0,1m.

ALTURA (metros)	FREQUÊNCIA ABSOLUTA (nº de alunos)	FREQUÊNCIA ABSOLUTA ACUMULADA	FREQUÊNCIA RELATIVA	FREQUÊNCIA RELATIVA ACUMULADA
1,40m ┆ - - - 1,50m	2	2	$\frac{2}{20} = 0,1 = 10\%$	10%
1,50m ┆ - - - 1,60m	4	6	$\frac{4}{20} = 0,2 = 20\%$	30%
1,60m ┆ - - - 1,70m	8	14	$\frac{8}{20} = 0,4 = 40\%$	70%
1,70m ┆ - - - 1,80m	5	19	$\frac{5}{20} = 0,25 = 25\%$	95%
1,80m ┆ - - - 1,90m	1	20	$\frac{1}{20} = 0,05 = 5\%$	100%
TOTAL	20	-----	100%	-----

Figura 2.2: Tabela de frequência da variável altura

Na tabela de frequência da variável cor dos olhos, que é uma variável qualitativa, cada cor será uma classe.

COR DOS OLHOS	FREQUÊNCIA ABSOLUTA (nº de alunos)	FREQUÊNCIA ABSOLUTA ACUMULADA	FREQUÊNCIA RELATIVA	FREQUÊNCIA RELATIVA ACUMULADA
Azul	3	3	$\frac{3}{20} = 0,15 = 15\%$	15%
Castanho	10	13	$\frac{10}{20} = 0,5 = 50\%$	65%
Preto	4	17	$\frac{4}{20} = 0,2 = 20\%$	85%
Verde	3	20	$\frac{3}{20} = 0,15 = 15\%$	100%
TOTAL	20	-----	100%	-----

Figura 2.3: Tabela de frequência da variável cor dos olhos

2.2.2 Gráfico de Linha (ou Segmentos)

A tabela a seguir mostra a venda de carros em uma certa concessionária durante o primeiro semestre do ano.

MÊS	Nº DE CARROS VENDIDOS
JANEIRO	55
FEVEREIRO	87
MARÇO	45
ABRIL	39
MAIO	82
JUNHO	124

Figura 2.4: Tabela com número de carros vendidos no semestre

Em um plano cartesiano marca-se os pontos de acordo com os pares ordenados (classe, frequência), ou seja, para esse caso, os pares(mês, nº de carros). Depois liga-se os pontos com segmentos de reta.

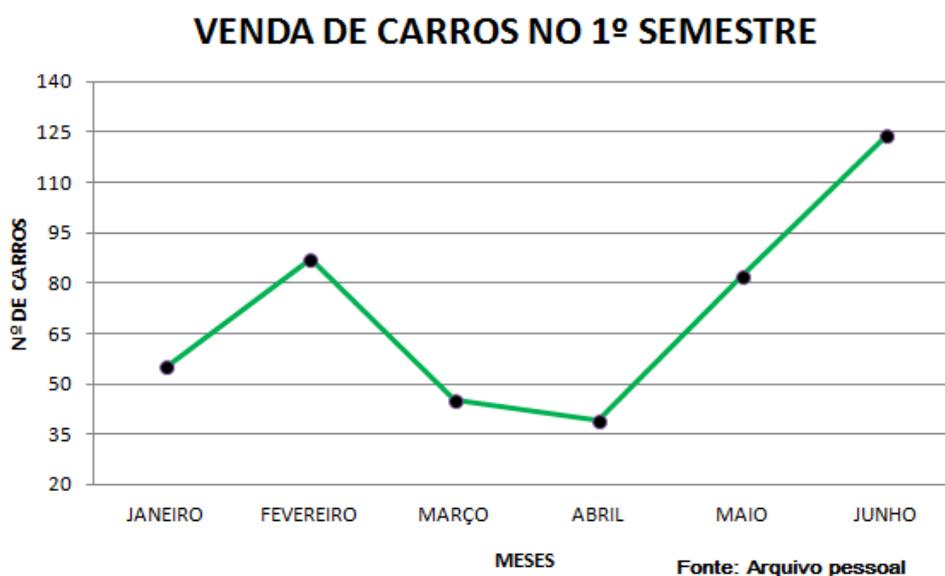


Figura 2.5: Gráfico de Linha ou Segmentos

Nesse tipo de gráfico, através da inclinação dos seguimentos, é possível notar um crescimento, decréscimo ou estabilidade dos valores da variável. Nele são fixados dados coletados no decorrer de um certo período afim de se observar o seu comportamento.

No exemplo citado, observa-se que o mês de junho foi o melhor em vendas para a loja e que abril foi o pior. Também nota-se uma queda nas vendas entre os meses de fevereiro e abril, bem como um crescimento entre os meses de abril e junho.

O gráfico de linha é utilizado geralmente quando pretende-se observar a evolução dos valores de uma variável, como os lucros de uma empresa ou o desempenho de um candidato à eleição nas últimas pesquisas de opinião.

2.2.3 Gráfico de Barras

A tabela a seguir traz as notas de um aluno nas provas de algumas disciplinas em um determinado mês.

DISCIPLINA	NOTA
PORTUGUÊS	7,0
HISTÓRIA	9,0
GEOGRAFIA	6,0
MATEMÁTICA	10,0
BIOLOGIA	5,0
QUÍMICA	7,0

Figura 2.6: Tabela de notas

Pode-se representar os dados da tabela em retângulos paralelos, todos de mesma largura e comprimentos proporcionais às suas frequências. Esse é o gráfico de barras. Os retângulos podem ser expostos na vertical (colunas), onde o eixo das frequências será o vertical, ou na horizontal, onde o eixo das frequências será o horizontal.

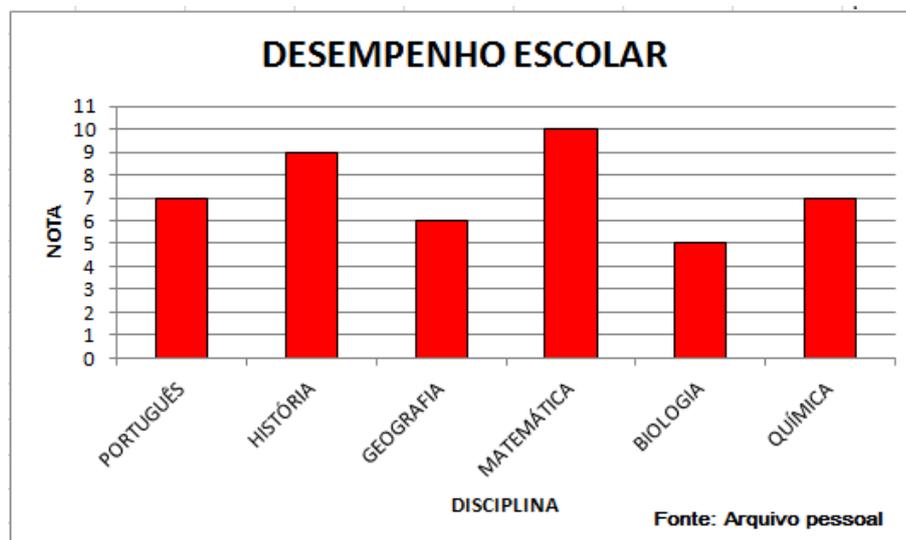


Figura 2.7: Gráfico de Barras Verticais

O gráfico de barras ajuda a analisar valores absolutos das variáveis. No exemplo, é possível notar claramente em quais disciplinas o aluno obteve melhores e piores desempenhos.

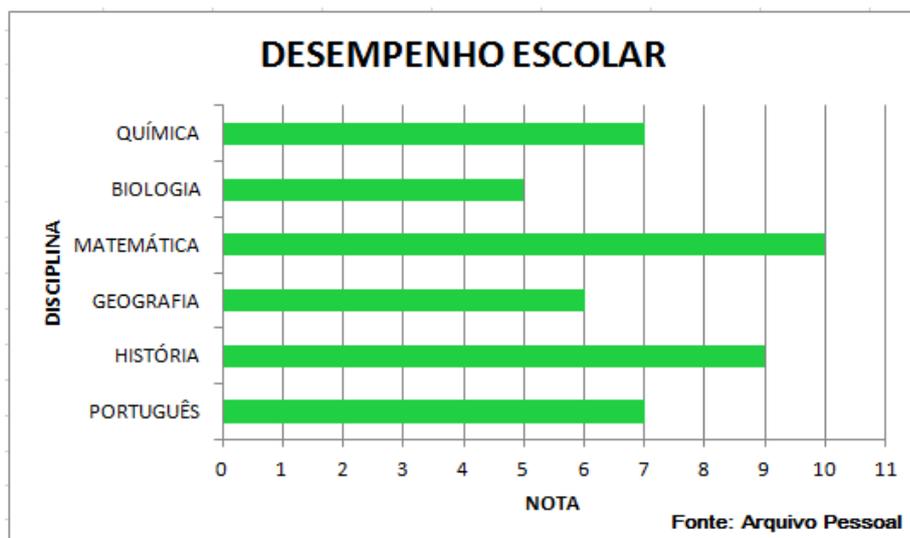


Figura 2.8: Gráfico de Barras Horizontais

2.2.4 Histograma

Quando os valores das variáveis são trabalhados na forma de intervalos, isto é, uma distribuição de frequência com dados agrupados, geralmente é utilizado o gráfico que chama-se Histograma. Observe o gráfico abaixo.

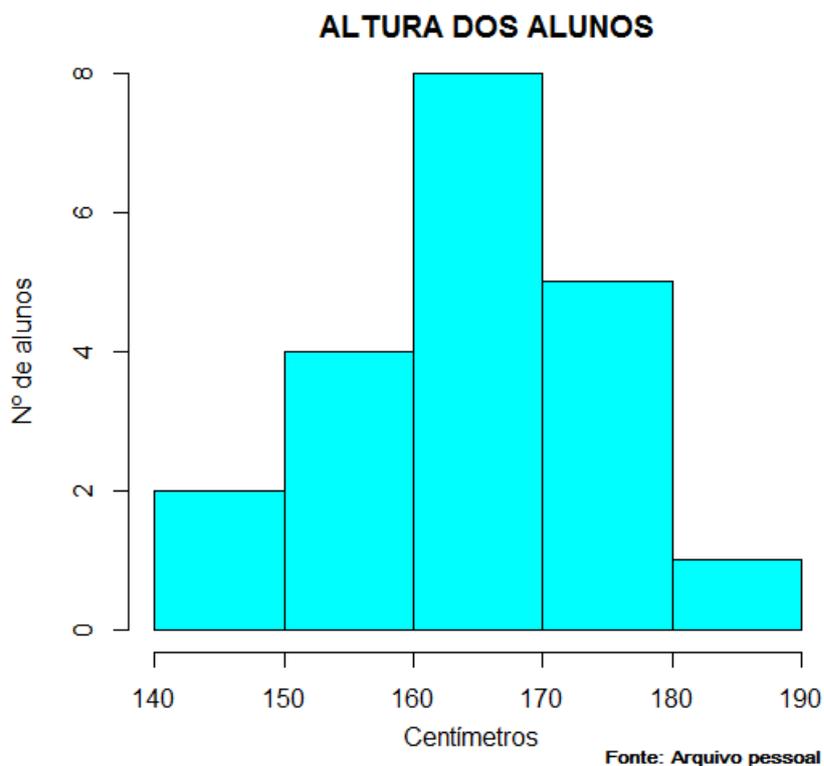


Figura 2.9: Histograma

Giovanni e Bonjorno [3] definem Histograma como um gráfico formado por um conjunto de colunas retangulares. No eixo das abscissas são marcadas as classes, cujas amplitudes correspondem às bases dos retângulos. No eixo das ordenadas marca-se as frequências absolutas, que correspondem às alturas dos retângulos. Os pontos médios das bases dos retângulos coincidem com os pontos médios dos intervalos das classes.

Para a construção do gráfico anterior, foram utilizados como referência os dados mostrados anteriormente da pesquisa com alunos em que consta suas alturas e com ele é possível notar que a maior quantidade de alunos tem altura medindo entre 1,60m e 1,70m.

2.2.5 Gráfico de setores (Pizza)

O gráfico de setores consiste em dividir um círculo em setores (fatias), daí também ser chamado de gráfico de pizza, com ângulos centrais proporcionais à frequência de cada classe. Geralmente utiliza-se a frequência relativa quando demonstra-se os dados nesse tipo de gráfico.

Para demonstração do gráfico de setores será utilizada a tabela com a cor dos olhos dos alunos (Figura 2.3) mostrada anteriormente. Primeiramente calcula-se o ângulo central de cada setor:

- Azul : $\frac{360^\circ \cdot 15}{100} = 54^\circ$
- Castanho : $\frac{360^\circ \cdot 50}{100} = 180^\circ$
- Preto : $\frac{360^\circ \cdot 20}{100} = 72^\circ$
- Verde : $\frac{360^\circ \cdot 15}{100} = 54^\circ$

Observe o gráfico:

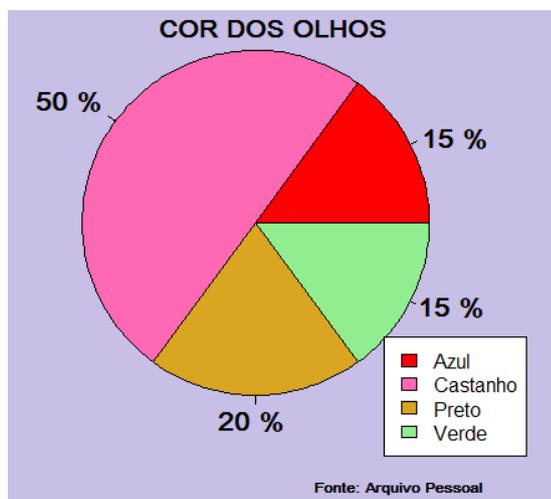


Figura 2.10: Gráfico de Setores

2.3 Medidas de Posição

As principais medidas de posição são as *medidas de tendência central*: **média aritmética**, **mediana** e **moda**. Mas também serão descritas as *separatrizes*: **quartis** e **percentis**.

A **média aritmética** é a mais comum no cotidiano das pessoas, portanto é a mais conhecida entre as medidas de tendência central. Veja a seguinte situação e entenda melhor. Ronaldo adora futebol e participa de jogos regularmente com os amigos. Nos últimos 5 jogos ele marcou essa quantidade de gols: 5, 2, 1, 4 e 2. Para encontrar a média aritmética calcula-se:

$$\bar{x} = \frac{5 + 2 + 1 + 4 + 2}{5} = 2,8$$

Portanto Ronaldo atualmente possui uma média de 2,8 gols por partida.

Quando se quer calcular a média aritmética de alguma amostra pega-se o quociente do somatório dos valores da amostra com a quantidades de parcelas dessa soma. Seja \bar{x} a média aritmética e $x_1, x_2, x_3, \dots, x_n$ os valores da amostra (ou variável) tem-se:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Para situações em que os dados estão agrupados em uma tabela de frequência, Crespo [1] diz que nesse caso, como as frequências são números indicadores de intensidade de cada valor da variável, elas funcionam como fatores de ponderação, o que nos leva a calcular a **média aritmética ponderada** que é dada pela fórmula:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + x_3 \cdot f_3 + \dots + x_n \cdot f_n}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i}$$

Analisando uma situação para melhor compreensão, suponha que em uma pesquisa sobre a quantidade de TVs que cada família possui em casa, tenha-se a tabela a seguir como resultado.

NÚMERO DE TVs	FREQUÊNCIA
0	5
1	36
2	44
3	24
4	8
5	3

Figura 2.11: Tabela do número de TVs por casa

Nesse caso, a média aritmética será dada por:

$$\bar{x} = \frac{0.5 + 1.36 + 2.44 + 3.24 + 4.8 + 5.3}{5 + 36 + 44 + 24 + 8 + 3} = 2,025$$

Portanto as famílias entrevistadas possuem uma média de 2,025 TVs em suas casas.

A **moda** corresponde ao valor de maior frequência em um determinado grupo de dados cuja frequência dos valores não sejam todas iguais.

Observe a série numérica 2, 5, 8, 3, 3, 5, 4, 2, 3, 8, 3, 5, 3. É fácil perceber que o valor com maior frequência é 3, portanto ele é a moda dessa série.

Em série de dados que não apresentam valor modal dizemos que a série é **amodal**, em séries que possuem dois valores modais dizemos que essa série é **bimodal** e caso apresentem mais de dois valores com maior frequência dizemos que essa série é **multimodal**.

Quando os dados estão organizados em rol, ou seja, dispostos em ordem crescente ou decrescente, a **mediana** corresponde ao valor central, isto é, posicionado ao meio da distribuição.

Dada uma amostra com os números 6, 4, 7, 4, 2, 9 e 6, para determinar a mediana basta organizar os dados em ordem e observar quem é o termo central.

$$\begin{array}{ccc} \underbrace{2 \quad 4 \quad 4} & 6 & \underbrace{6 \quad 7 \quad 9} \\ 3 \text{ termos} & \uparrow & 3 \text{ termos} \\ & \text{Mediana} & \end{array}$$

Daí pode-se concluir que a mediana desses dados é 6. Caso a quantidade de valores da amostra seja par, onde não há um único termo central, deve-se calcular a média aritmética dos dois números que posicionam-se ao centro da distribuição. Tomando a distribuição descrita anteriormente acrescentando o 5 temos:

$$\begin{array}{ccc} \underbrace{2 \quad 4 \quad 4} & \underbrace{5 \quad 6} & \underbrace{6 \quad 7 \quad 9} \\ 3 \text{ termos} & \uparrow & 3 \text{ termos} \\ & \frac{5 + 6}{2} = 5,5 & \end{array}$$

Portanto a mediana agora é 5,5.

Com os dados organizados em ordem crescente determina-se os **quartis**(Q1, Q2, Q3) que são valores que dividem a distribuição em quatro partes iguais.

- **Primeiro Quartil (Q1)** : número que deixa 25% dos valores da pesquisa abaixo dele.

- **Segundo Quartil (Q2)** : número que deixa 50% dos valores da pesquisa abaixo dele (mediana).

- **Terceiro Quartil (Q3)** : número que deixa 75% dos valores da pesquisa abaixo dele.

Uma maneira simples de calcular cada quartil é pela fórmula $Q_k = \frac{k \cdot (n + 1)}{4}$ onde $k = 1, 2, 3$ e n o total de elementos da amostra. Com essa fórmula é possível encontrar a posição do termo correspondente ao quartil. Caso o resultado não seja inteiro basta calcular a média aritmética entre os termos das posições imediatamente inferior e superior.

Observe os dados de uma pesquisa organizados em ordem crescente: 1,5 ; 2,4 ; 2,8 ; 3,4 ; 3,9 ; 4,5 ; 5,1 ; 5,3 ; 5,6 ; 6,2.

Nesse exemplo:

- $Q1 = \frac{1 \cdot (10 + 1)}{4} = 2,75$, ou seja, deve-se fazer a média aritmética entre o 2º e o 3º termo: $\frac{2,4 + 2,8}{2} = 2,6$. Portanto $Q1 = 2,6$, isto é, 25% das amostras são inferiores a 2,6 e 75% são superiores.

- $Q2 = \frac{2 \cdot (10 + 1)}{4} = 5,5$, ou seja, deve-se fazer a média aritmética entre o 5º e o 6º termo: $\frac{3,9 + 4,5}{2} = 4,2$. Portanto $Q2 = 4,2$, isto é, 50% das amostras são inferiores a 4,2 e 50% são superiores.

- $Q3 = \frac{3 \cdot (10 + 1)}{4} = 8,25$, ou seja, deve-se fazer a média aritmética entre o 8º e o 9º termo: $\frac{5,3 + 5,6}{2} = 5,45$. Portanto $Q3 = 5,45$, isto é, 75% das amostras são inferiores a 5,45 e 25% são superiores.

Representando em uma reta tem-se:

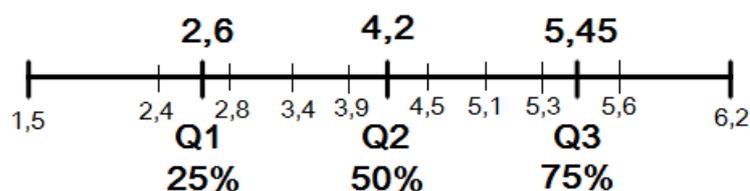


Figura 2.12: Reta Numérica destacando os quartis

Os **percentis** correspondem aos 99 valores que dividem a série numérica dos dados em 100 partes iguais. Assim é feita a mesma análise dos quartis, isto é, o 20º percentil (P_{20}) é o valor que divide a sequência, ficando 20% dos valores abaixo dele e 80% acima dele. Observe que $P_{25} = Q1$, $P_{50} = Q2$ (Mediana) e $P_{75} = Q3$.

Existe um tipo de gráfico que utiliza os valores dos quartis, é o **boxplot**. Este gráfico consiste em uma caixa compreendida entre o $Q1$ (1º quartil) e o $Q3$ (3º quartil) com uma linha seccionando a caixa na altura do $Q2$ (2º quartil) que é o termo central da amostra (mediana), ou seja, no espaço delimitado por esta caixa encontram-se 50% dos valores da amostra, o que nos dá uma tendência central dos valores. Duas semirretas partindo das laterais da caixa estendem-se até as extremidades da amostra, geralmente limitando de 10% a 90%. Os valores muito distantes da mediana são representados por pontos fora desses limites e são conhecidos por *outliers* e essa é uma das principais utilidades do boxplot, mostrar quem são valores dispersos na amostra, isto é, ficam distantes da concentração.

Suponha que uma pesquisa sobre o peso, em quilogramas, de um grupo de pessoas teve como resultado os valores 56, 65, 59, 70, 57, 56, 74, 72, 68, 70, 66, 83, 72, 55, 49, 79, 120, 85, 54, 64, 72. Com esses dados pode-se construir o gráfico:

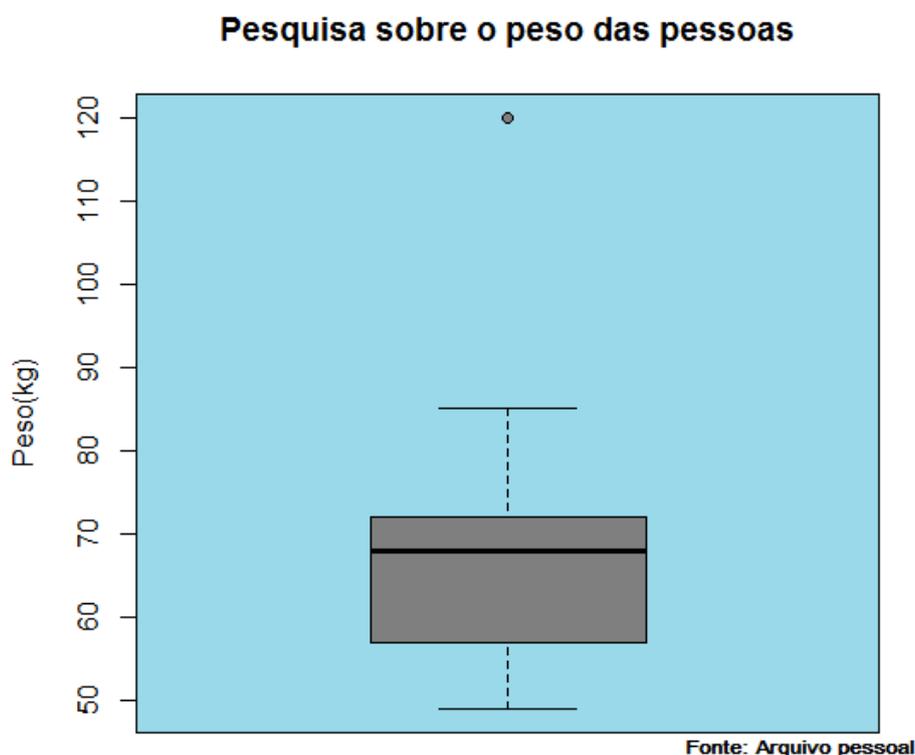


Figura 2.13: Gráfico Boxplot

Note que o ponto referente ao 120 fica fora dos limites demarcados, ele é um outliers. Conhecer os outliers ajuda pois em certos estudos deve-se desconsiderar esse resultado por ele estar tão disperso em relação aos demais.

2.4 Medidas de Dispersão

As medidas de posição são muito úteis na hora de analisar uma variável mas nem sempre elas são suficientes para se caracterizar corretamente essa amostra.

Em uma certa empresa, para o funcionário ser promovido ele precisa participar de um pequeno concurso interno que consiste em cinco provas com os conteúdos: Conhecimentos de Informática, Português, Matemática, Inglês e Conhecimentos Gerais. O funcionário que obtiver a maior média nas provas será promovido.

Na tabela abaixo estão as notas dos dois funcionários com melhor média: Paulo e Sara.

CONTEÚDO	PAULO	SARA
INFORMÁTICA	9,0	8,5
PORTUGUÊS	6,0	9,0
MATEMÁTICA	10,0	7,5
INGLÊS	6,0	7,0
CONHEC. GERAIS	9,0	8,0
MÉDIA	8,0	8,0

Figura 2.14: Tabela com as notas de Paulo e Sara

Nesse caso, como os dois obtiveram a mesma média, teremos que abordar um novo critério para selecionar o candidato a ser promovido. Esse critério será a **dispersão** ou **variabilidade** das notas em relação à média, ou seja, avaliar o quanto as notas de cada um estão distantes da média. Aquele que obter notas mais homogêneas, ou seja, menos distantes (dispersas) em relação à média será o funcionário mais capacitado para a promoção.

As principais medidas de dispersão são: **amplitude total**, **variância** e **desvio padrão**.

Calculando a amplitude total, que já foi mencionada anteriormente e é dada pela diferença entre o maior e o menor valor observado, tem-se:

$$\text{PAULO} : 10 - 6 = 4$$

$$\text{SARA} : 9 - 7 = 2$$

Através dessa medida vemos que Sara possui uma menor dispersão em suas notas e portanto seria a promovida. Mas pelo fato de a amplitude total utilizar apenas os dois valores extremos ela acaba tornando-se limitada.

O ideal é uma medida de abrangência todos os valores da amostra. E uma opção é a variância.

Para calcular a variância utiliza-se os desvios dos valores obtidos em relação à média aritmética, isto é, a diferença entre esses valores e a média. A variância é dada pela média aritmética dos quadrados dos desvios e é representada por σ^2 . Sendo $x_1, x_2, x_3, \dots, x_n$ os valores da amostra e \bar{x} a média aritmética calculada-se a variância pela fórmula:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

Assim, no exemplo anterior tem-se:

$$\sigma_{\text{Paulo}}^2 = \frac{(9 - 8)^2 + (6 - 8)^2 + (10 - 8)^2 + (6 - 8)^2 + (9 - 8)^2}{5} = 2,8$$

$$\sigma_{\text{Sara}}^2 = \frac{(8,5 - 8)^2 + (9 - 8)^2 + (7,5 - 8)^2 + (7 - 8)^2 + (8 - 8)^2}{5} = 0,5$$

Novamente se vê que o resultado de Sara é mais homogêneo.

Como a variância é determinada tomando o quadrado dos desvios, sua unidade de medida não será a mesma dos valores da variável e isso atrapalha na interpretação do seu resultado. Com isso, uma outra opção de medida de dispersão é o desvio padrão, que é a raiz quadrada da variância e é representada por σ . Assim a sua unidade volta a ser a mesma dos valores analisados.

$$\sigma_{\text{Paulo}} = \sqrt{2,8} \cong 1,7$$

$$\sigma_{\text{Sara}} = \sqrt{0,5} \cong 0,70$$

Daí observa-se que as notas de Paulo variam em média 1,7 pontos em relação à média, enquanto que as notas de Sara variam apenas 0,7 ponto em relação à média. Portanto conclui-se definitivamente que Sara é a pessoa mais capacitada para assumir o novo cargo.

No capítulo a seguir será apresentada uma apostila para o curso com o software R que foi elaborada tomando como principais referências os materiais: *Estatística com R* da FGV/IBRE [7] e *Utilização do R no Ensino da Estatística Básica I* de Carolina Valani Cavalcante e Luciane Ferreira Alcoforado [12].

Capítulo 3

Aplicando o Software R

3.1 Preparando o Laboratório

Inicialmente deve-se fazer o download e instalar o software R em todos os computadores do laboratório visando deixar os equipamentos aptos a receberem os alunos e para que não haja atrasos no início do curso. Assim, chegando ao laboratório deve-se pedir a todos que liguem os computadores e abram o software clicando sobre o ícone do R. A tela abaixo abrirá.

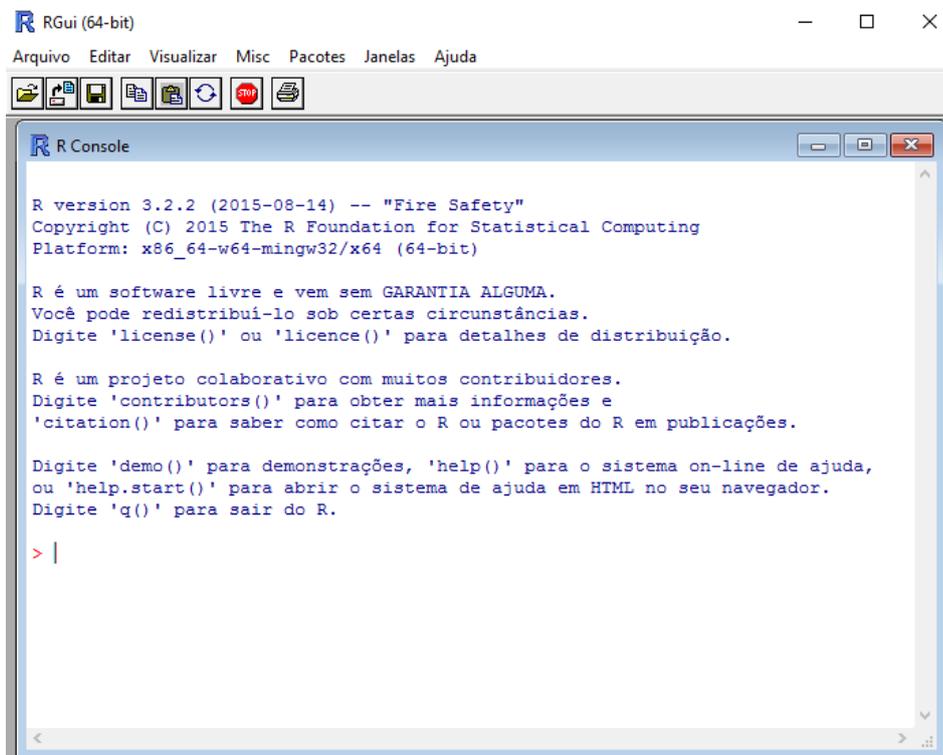


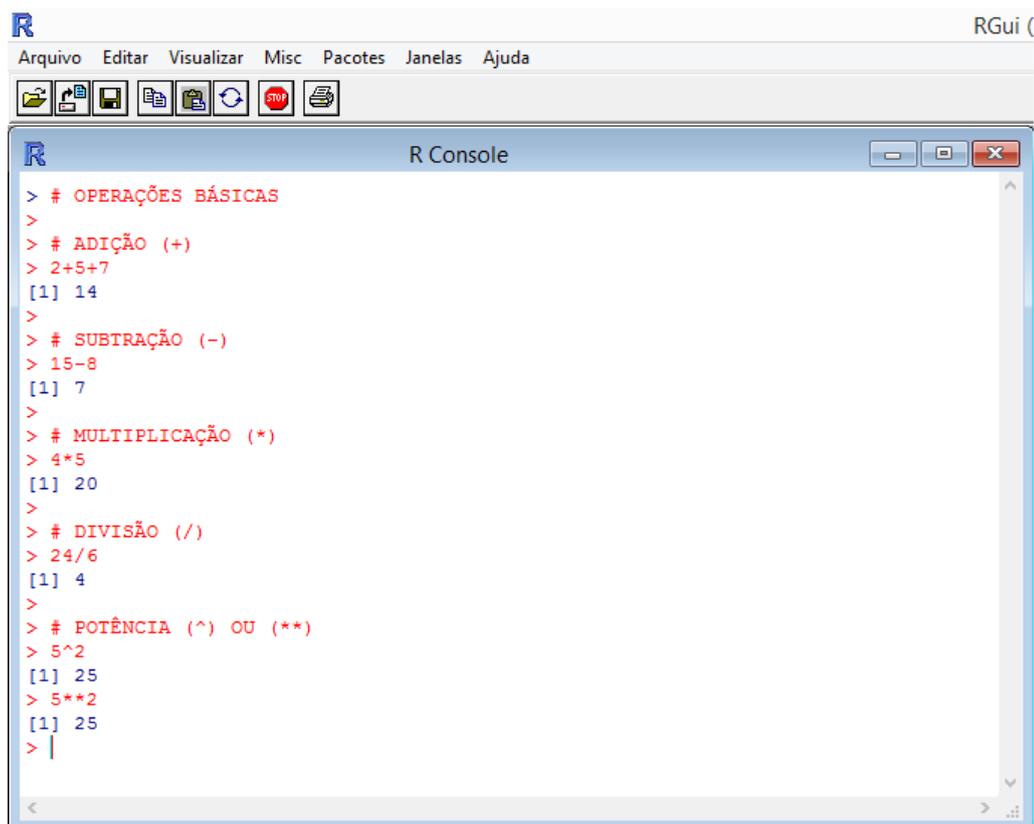
Figura 3.1: Tela inicial do Software R

Para limpar a tela basta pressionar **CTRL + L**. E agora podemos iniciar as atividades práticas do curso.

3.2 Operações Básicas

Antes de ir direto ao conteúdo de estatística pode-se mostrar algumas funções básicas e essenciais, afim de que os alunos se familiarizarem com o software R. O programa efetua todas as operações básicas podendo assim ser utilizado como uma calculadora. Os comandos são os seguintes:

- Adição : +
- Subtração : -
- Multiplicação : *
- Divisão : /
- Potência : ^ ou **

The image shows a screenshot of the R GUI (RGui) window. The title bar reads 'RGui (R)'. The menu bar includes 'Arquivo', 'Editar', 'Visualizar', 'Misc', 'Pacotes', 'Janelas', and 'Ajuda'. Below the menu bar is a toolbar with icons for file operations and execution. The main window is titled 'R Console' and contains the following text:

```
> # OPERAÇÕES BÁSICAS
>
> # ADIÇÃO (+)
> 2+5+7
[1] 14
>
> # SUBTRAÇÃO (-)
> 15-8
[1] 7
>
> # MULTIPLICAÇÃO (*)
> 4*5
[1] 20
>
> # DIVISÃO (/)
> 24/6
[1] 4
>
> # POTÊNCIA (^) OU (**)
> 5^2
[1] 25
> 5**2
[1] 25
> |
```

Figura 3.2: Operações básicas no R

OBS: O símbolo # trata-se de um comentário.

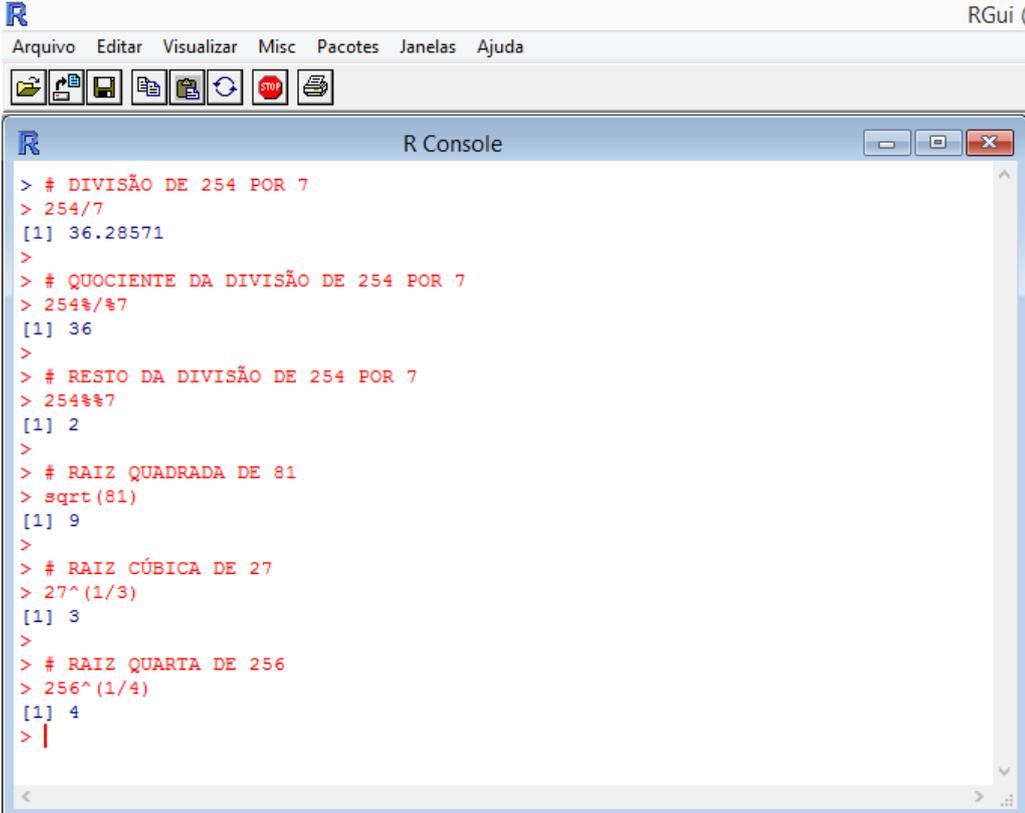
Após mostrar esses comandos os alunos podem ficar a vontade para criarem e calcularem expressões numéricas envolvendo as operações fundamentais.

Através de comandos simples pode-se encontrar informações diretas de uma divisão como o quociente e o resto. Encontrar o resto, por exemplo, ajuda a saber se um número é divisível por outro.

- Quociente da divisão: `% / %`
- Resto da divisão: `% %%`

Para calcular a raiz quadrada utiliza-se o comando `sqrt()` que é a abreviação de square root (raiz quadrada em inglês).

Caso queira calcular uma raiz de outra ordem pode utilizar a potenciação com expoente fracionário. A raiz cúbica de 27 ($\sqrt[3]{27}$), por exemplo, calcula-se com `27^(1/3)`. Coloca-se parênteses no expoente, pois caso contrário, primeiro será calculado a potência `27^1` e posteriormente a divisão por 3.

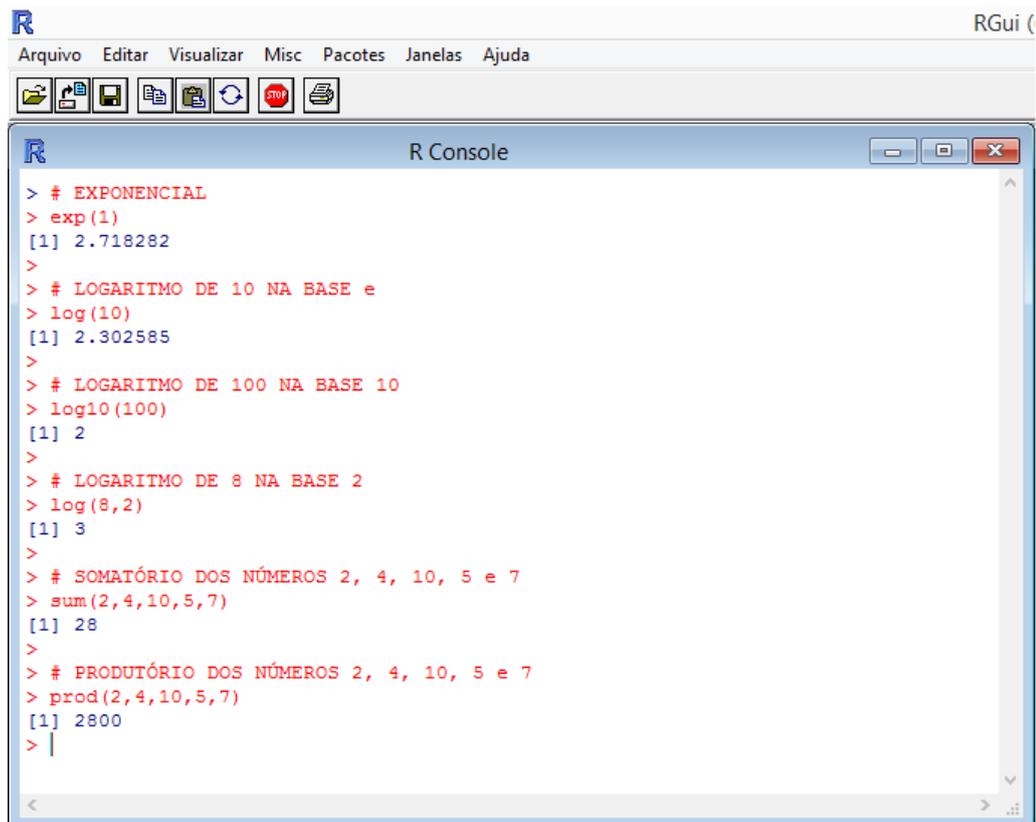


```
R
Arquivo Editar Visualizar Misc Pacotes Janelas Ajuda
R Console
> # DIVISÃO DE 254 POR 7
> 254/7
[1] 36.28571
>
> # QUOCIENTE DA DIVISÃO DE 254 POR 7
> 254%/%7
[1] 36
>
> # RESTO DA DIVISÃO DE 254 POR 7
> 254%%7
[1] 2
>
> # RAIZ QUADRADA DE 81
> sqrt(81)
[1] 9
>
> # RAIZ CÚBICA DE 27
> 27^(1/3)
[1] 3
>
> # RAIZ QUARTA DE 256
> 256^(1/4)
[1] 4
> |
```

Figura 3.3: Calculando quociente e resto de uma divisão e radiciação

Algumas outras funções conhecidas dos alunos também podem ser calculadas usando esse software, como exponencial, logarítmica, somatório e produtório.

- Exponencial com base e : **exp()**
- Logaritmo na base e : **log()**
- Logaritmo na base 10 : **log10()**
- Logaritmo de a na base b : **log(a,b)**
- Somatório : **sum()**
- Produtório : **prod()**



```
> # EXPONENCIAL
> exp(1)
[1] 2.718282
>
> # LOGARITMO DE 10 NA BASE e
> log(10)
[1] 2.302585
>
> # LOGARITMO DE 100 NA BASE 10
> log10(100)
[1] 2
>
> # LOGARITMO DE 8 NA BASE 2
> log(8,2)
[1] 3
>
> # SOMATÓRIO DOS NÚMEROS 2, 4, 10, 5 e 7
> sum(2,4,10,5,7)
[1] 28
>
> # PRODUTÓRIO DOS NÚMEROS 2, 4, 10, 5 e 7
> prod(2,4,10,5,7)
[1] 2800
> |
```

Figura 3.4: Algumas funções e operações matemáticas

3.3 Armazenando dados

3.3.1 Criando um objeto

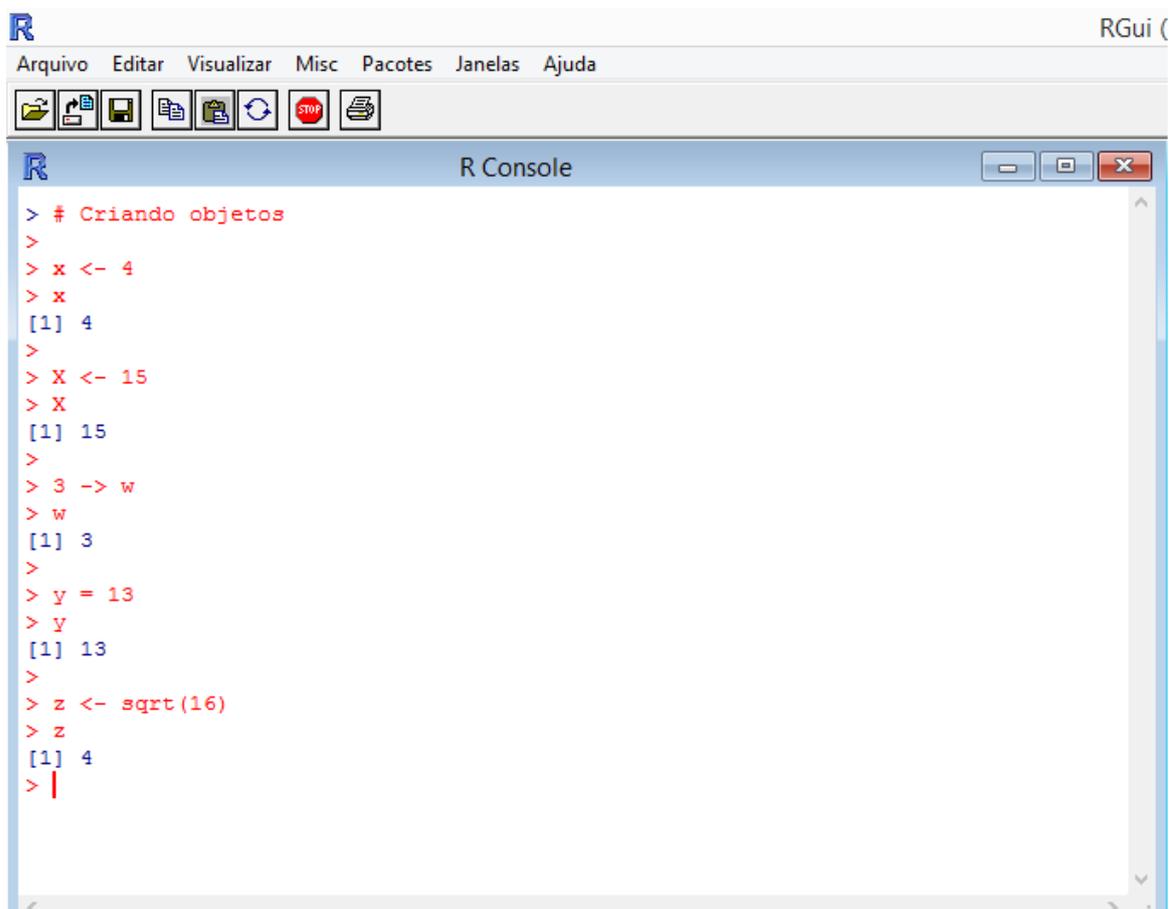
Para criar um objeto deve-se dar um nome a ele utilizando “uma seta” formada pelos símbolos “<” (menor que) ou “>” (maior que) com “-” (menos) dependendo em que direção está o nome dado ao objeto ou ainda utilizar o = (igual).

- `x <- 4` : o objeto x receberá o valor 4
- `X <- 15` : o objeto X receberá o valor 15

- `3 -> w` : o objeto `w` receberá o valor 3
- `y = 13` : o objeto `y` receberá o valor 13
- `z <- sqrt(16)` : o objeto `z` receberá o valor da raiz quadrada de 16

OBS: Deve-se observar que o software diferencia letra maiúscula de minúscula e também que o nome do objeto deve começar com uma letra e não poderá contar com símbolos referentes a funções, caso contrário ocorrerá erro na operação. Por exemplo:

- `x+ <- 4` : o nome do objeto possui o sinal da adição
- `2y <- 10` : o nome do objeto começa com número
- `w -> 5` : a seta esta para o lado contrário

The image shows a screenshot of the R GUI window. The title bar reads 'RGui ('. The menu bar includes 'Arquivo', 'Editar', 'Visualizar', 'Misc', 'Pacotes', 'Janelas', and 'Ajuda'. Below the menu bar is a toolbar with icons for file operations and execution. The main window is titled 'R Console' and contains the following code and output:

```
> # Criando objetos
>
> x <- 4
> x
[1] 4
>
> X <- 15
> X
[1] 15
>
> 3 -> w
> w
[1] 3
>
> y = 13
> y
[1] 13
>
> z <- sqrt(16)
> z
[1] 4
> |
```

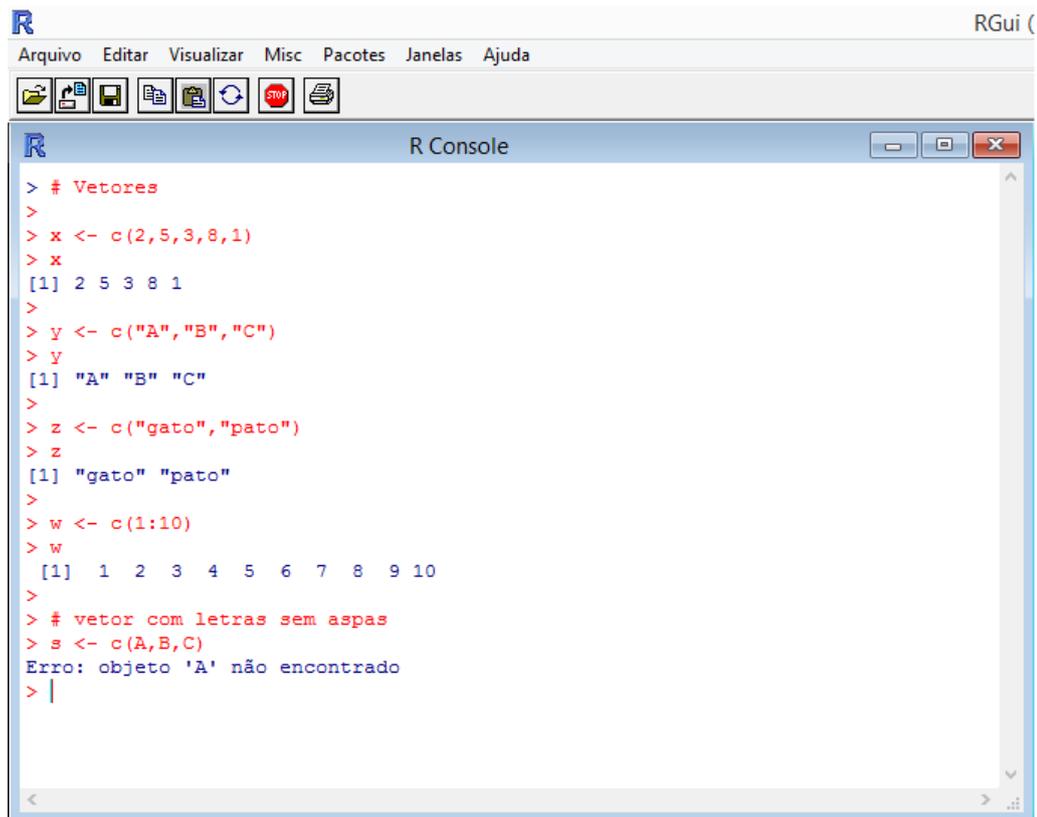
Figura 3.5: Criando objetos

3.3.2 Criando Vetores

Os vetores são objetos que armazenam mais de um valor. Para isso utiliza-se a função `c()`. Se o vetor for composto por letras ou nomes é preciso coloca-los entre aspas (“ ”),

caso contrário ocorrerá um erro. Se trabalhar com um intervalo numérico não é preciso escrever um a um, basta utilizar dois pontos (:) entre as extremidades do intervalo.

- `x <- c(2,5,3,8,1)` : vetor de nome x com os valores numéricos 2,5,3,8 e 1.
- `y <- c("A","B","C")` : vetor de nome y com os caracteres A, B e C.
- `z <- c("gato","pato")` : vetor de nome z com os nomes gato e pato.
- `w <- c(1:10)` : vetor de nome w com os números de 1 a 10.



```
R
Arquivo Editar Visualizar Misc Pacotes Janelas Ajuda
R Console
> # Vetores
>
> x <- c(2,5,3,8,1)
> x
[1] 2 5 3 8 1
>
> y <- c("A","B","C")
> y
[1] "A" "B" "C"
>
> z <- c("gato","pato")
> z
[1] "gato" "pato"
>
> w <- c(1:10)
> w
[1] 1 2 3 4 5 6 7 8 9 10
>
> # vetor com letras sem aspas
> s <- c(A,B,C)
Erro: objeto 'A' não encontrado
> |
```

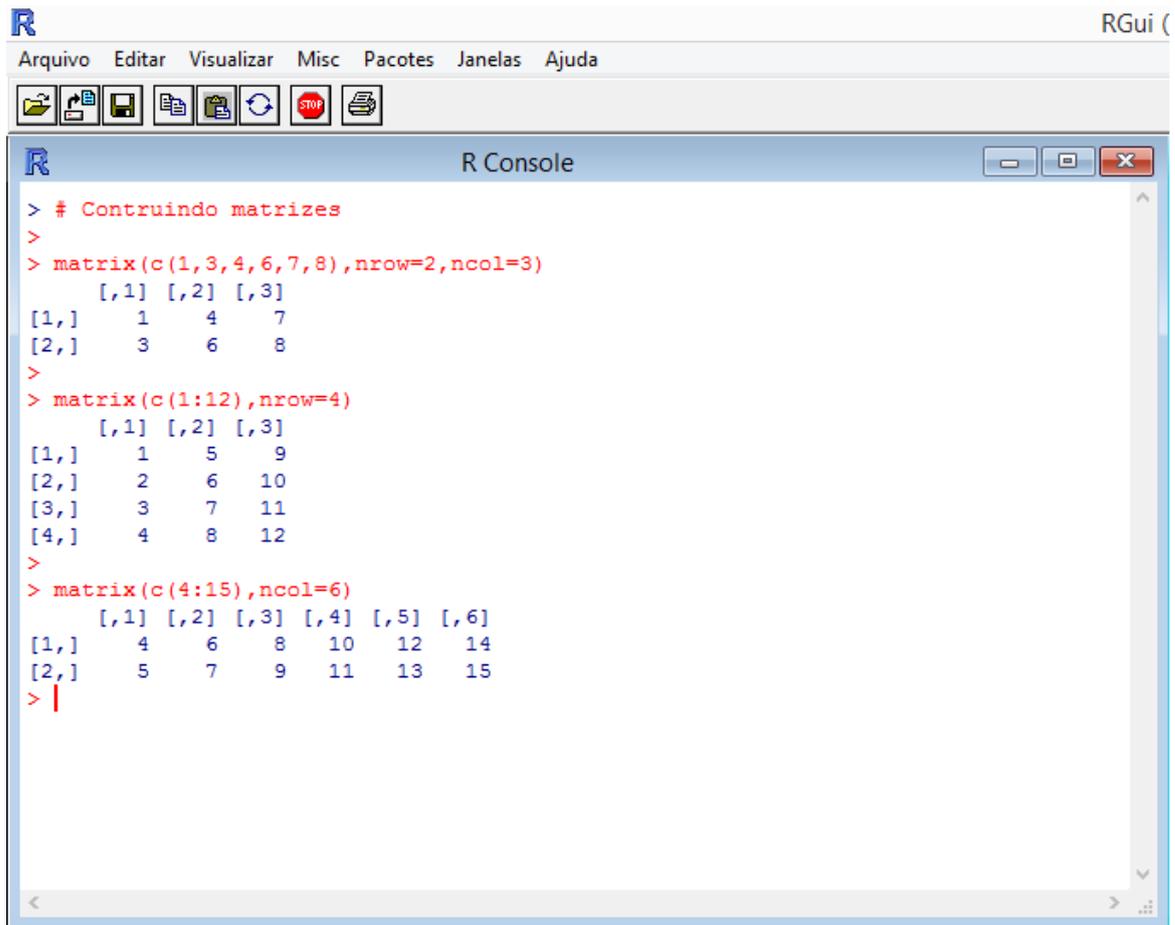
Figura 3.6: Criando vetores

3.3.3 Criando Matrizes

Uma matriz é criada pela função `matrix()`. O preenchimento da matriz com os valores dados será por coluna da primeira para a última.

- `matrix(c(1,3,4,6,7,8), nrow=2, ncol=3)` : matriz formada pelos elementos 1, 3, 4, 6, 7, 8 com 2 linhas e 3 colunas.
- `matrix(c(1:12), nrow=4)` : matriz formada pelos números da sequência de 1 a 12 com 4 linhas.

- `matrix(c(4:15), ncol=6)` : matriz formada pelos números da sequência de 4 a 15 com 6 colunas.



```
> # Construindo matrizes
>
> matrix(c(1,3,4,6,7,8),nrow=2,ncol=3)
  [,1] [,2] [,3]
[1,]  1   4   7
[2,]  3   6   8
>
> matrix(c(1:12),nrow=4)
  [,1] [,2] [,3]
[1,]  1   5   9
[2,]  2   6  10
[3,]  3   7  11
[4,]  4   8  12
>
> matrix(c(4:15),ncol=6)
  [,1] [,2] [,3] [,4] [,5] [,6]
[1,]  4   6   8  10  12  14
[2,]  5   7   9  11  13  15
> |
```

Figura 3.7: Criando matrizes

3.3.4 Construindo Data frame

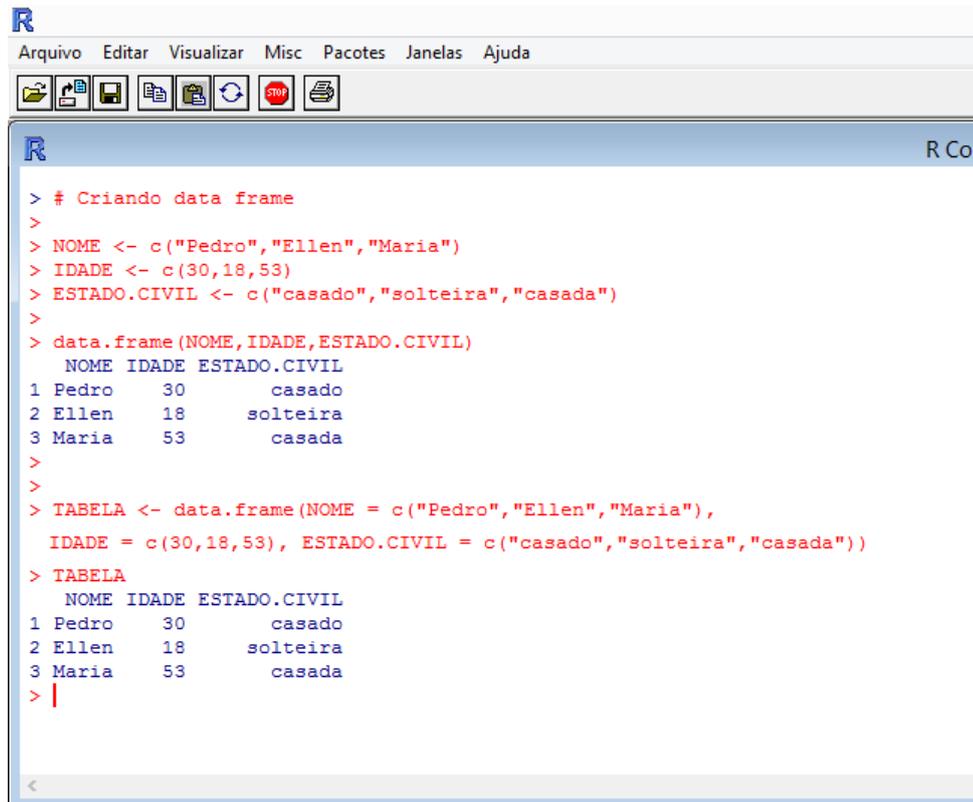
Este tipo de objeto será o mais utilizado no conteúdo de estatística, pois através dele pode-se armazenar tabelas onde cada coluna envolve dados numéricos ou outros caracteres. Ele também é importante por ser o formato que o R utiliza para ler dados importados de outros softwares como Excel e Bloco de Notas.

O comando é dado por `data.frame()`. Para montar uma tabela com os dados de uma pesquisa que envolve NOME, IDADE, ESTADO CIVIL, existem duas maneiras:

- Pode-se criar cada vetor individualmente, por exemplo `NOME<-c("Pedro","Ellen", "Maria")`, `IDADE <- c(30,18,53)` e `ESTADO.CIVIL <- c ("casado", "solteira", "casada")`, depois aplicar a função `data.frame()` para os vetores escrevendo `data.frame(`

```
NOME, IDADE, ESTADO.CIVIL);
```

- Criar a tabela em um único comando englobando as duas funções gravando-a com um nome qualquer, por exemplo `TABELA <- data.frame(NOME = c("Pedro", "Ellen", "Maria"), IDADE = c(30, 18, 53), ESTADO.CIVIL = c("casado", "solteira", "casada"))`.



```
> # Criando data frame
>
> NOME <- c("Pedro","Ellen","Maria")
> IDADE <- c(30,18,53)
> ESTADO.CIVIL <- c("casado","solteira","casada")
>
> data.frame(NOME,IDADE,ESTADO.CIVIL)
  NOME IDADE ESTADO.CIVIL
1 Pedro   30      casado
2 Ellen   18     solteira
3 Maria   53      casada
>
>
> TABELA <- data.frame(NOME = c("Pedro","Ellen","Maria"),
  IDADE = c(30,18,53), ESTADO.CIVIL = c("casado","solteira","casada"))
> TABELA
  NOME IDADE ESTADO.CIVIL
1 Pedro   30      casado
2 Ellen   18     solteira
3 Maria   53      casada
> |
```

Figura 3.8: Criando data frame (tabelas)

Para adicionar uma nova coluna ao data frame basta repetir o seu nome, colocar \$ e o nome da nova coluna com seus valores. No exemplo exposto na figura anterior (Figura 3.8), para criar uma nova coluna com o peso (em kg) de cada pessoa basta escrever `TABELA$PESO <- (90,57,65)`.

As funções `cbind()` ou `rbind()` são utilizadas para incluir uma coluna ou uma linha, respectivamente, ao data frame.

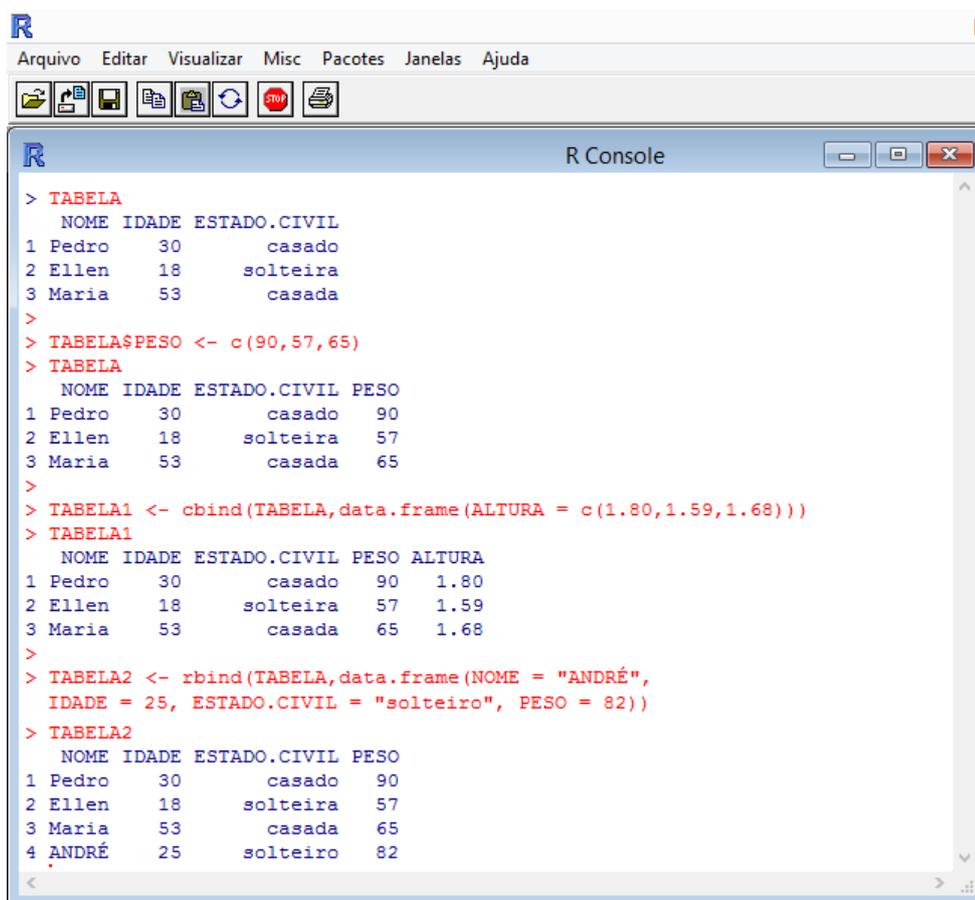


Figura 3.9: Incluindo linha e coluna ao data frame

Há ainda a função `edit (data.frame())`.

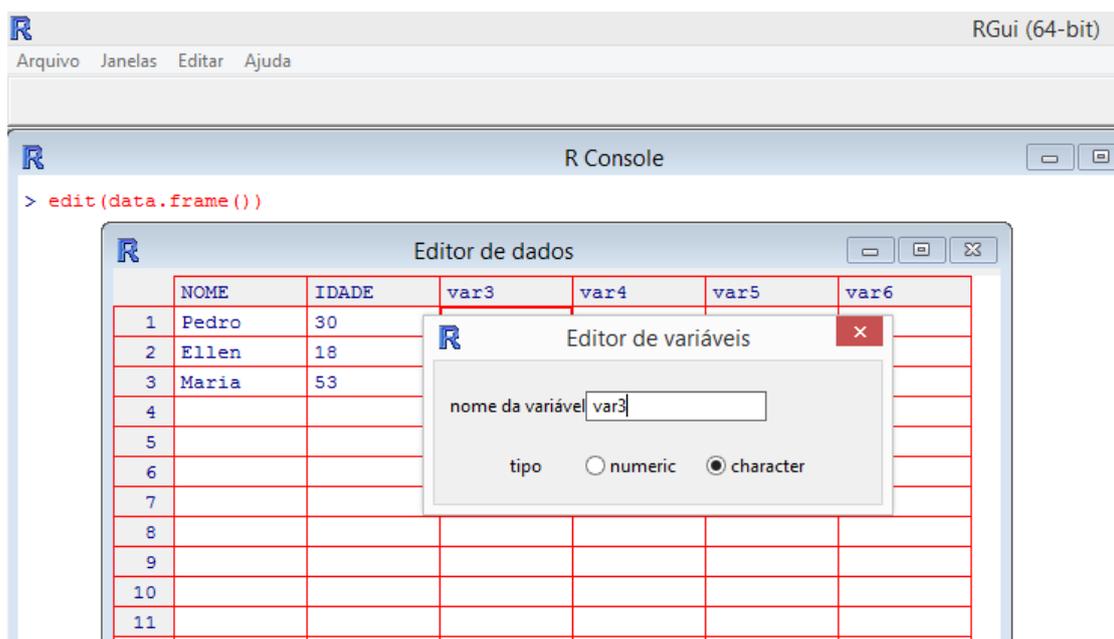


Figura 3.10: Criando data frame com a função `edit(data.frame())`

Utilizando essa função aparece uma janela idêntica ao Excel em que preenchemos o nome da variável e marcamos se essa variável trata-se de valores numéricos ou caracteres. Essa é a melhor opção para quem quer digitar uma planilha grande.

Após essa pequena introdução do software R pode-se por em prática o que foi estudado na teoria em sala.

3.4 Estatística Descritiva

No estudo de estatística, vários valores e medidas utilizados para as análises dos dados são encontrados através de cálculos e observações que, em muitas vezes, são complexos e assim demandam um certo tempo para decifrá-los e/ou calculá-los, mesmo dispondo da ajuda de uma calculadora.

Pois bem, com o software R é possível encontrar esses resultados apenas utilizando comandos que, em sua maioria, são bem simples.

3.4.1 Produzindo uma tabela para ser analisada

Inicialmente precisa-se de dados para serem organizados e analisados posteriormente. Para isso, uma opção bem interessante, é fazer uma pesquisa com os alunos da própria turma, montar uma tabela com esses dados e esta será a base para as análises estatísticas. Caso prefira, o professor também pode levar uma tabela previamente montada e repassar aos computadores do laboratório para que os alunos possam importar para o R.

Optando-se por levar a tabela pronta lembre-se de salvar no formato *.txt*, por exemplo, se a tabela foi preparada no Excel, deve ser salva com o formato *Texto (Separado por tabulações)* para que o R seja capaz de reproduzi-la.

Caso opte por realizar a pesquisa entre os alunos, o ideal seria pedir que montem a tabela no Excel e sigam essa instrução na hora de salvar, ou ainda, como foram demonstradas duas outras maneiras de se inserir um data frame, pode-se deixar a cargo de cada um escolher a forma que achar mais conveniente das três. Como uma estrutura de tabela facilitaria no preenchimento dos dados, entre as duas demonstradas no tópico 3.3.4, recomenda-se utilizar o comando `edit(data.frame())`. O nome de “tabela” pode ser dado ao data frame. Assim basta digitar `tabela <- edit(data.frame())` e preencher a tabela.

Supondo que os alunos escolham fazer a pesquisa entre eles e utilizar o excel, deve-se montar o banco de dados, ou seja, a tabela. Ordenar os dados de cada aluno de acordo com o número correspondente a ele na lista de frequência. Podem ser pesquisados, por exemplo, a idade (em anos), o sexo (masculino ou feminino), o peso (em kg) e a cor do cabelo (preto, castanho, loiro ou ruivo).

OBS: Nesta apostila os dados são fictícios já que sua finalidade é apenas demonstrativa.

Para importar o arquivo utiliza-se o comando **read.table()**. Mas antes de usar o comando deve-se clicar (na janela do R) em **Arquivo**, escolher a opção **Mudar dir...** e localizar a pasta onde se encontra salvo o arquivo que será importado.

Feito isso, usa-se o comando para importar o banco de dados, mas não esqueça de dar um nome ao vetor. Supondo que o arquivo foi salvo no computador com o nome `tabela1`, escreve-se `tabela <- read.table("tabela1.txt", header=TRUE)`. O `.txt` corresponde ao formato do arquivo e o complemento `header=TRUE` é para informar ao programa que a primeira linha da tabela corresponde ao nome das variáveis.

Após importar a tabela deve-se digitar o comando **attach(nome do objeto)** para que o programa entenda qual vetor estará sendo utilizado como referência para as análises que serão feitas a seguir.

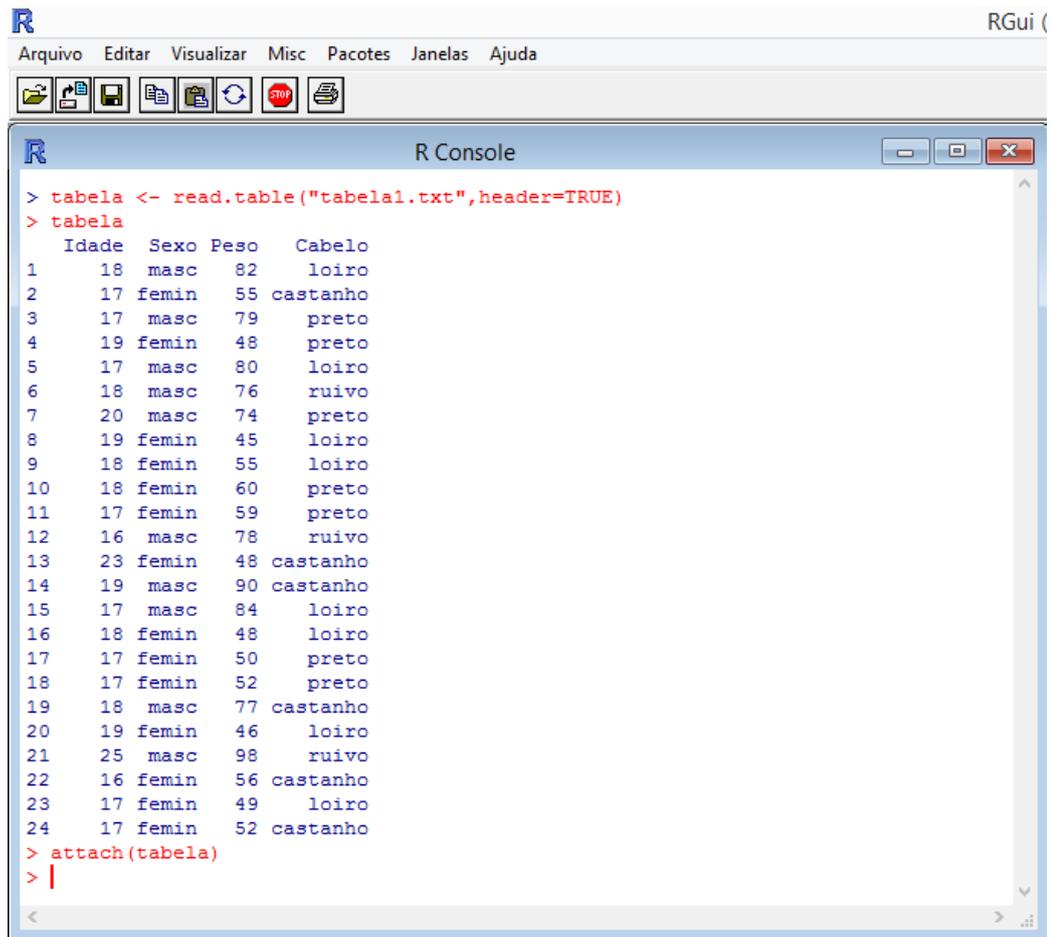
3.4.2 Medidas de Posição

Para começar o estudo do banco de dados observe alguns comandos: **length()**, **min()**, **max()** e **range()**.

- `length(tabela)` : mostra o número de variáveis analisadas.
- `length(Idade)` : mostra a quantidade de valores da variável Idade.
- `min(Peso)` : mostra o menor valor de peso apresentado.
- `max(Peso)` : mostra o maior valor de peso apresentado.
- `range(Idade)` : mostra o menor e o maior valor descrito para a variável Idade.

Assim pode-se, por exemplo, calcular a amplitude dos valores da variável `Peso` escrevendo a operação `max(Peso) - min(Peso)`.

Para calcular o total das variáveis quantitativas, basta utilizar o comando **sum()** que já foi citado anteriormente. Portanto, para calcular a soma de todos os valores da variável `Idade`, por exemplo, basta usar `sum(Idade)`.



```
> tabela <- read.table("tabela1.txt",header=TRUE)
> tabela
  Idade  Sexo  Peso  Cabelo
1    18  masc   82   loiro
2    17  femin  55 castanho
3    17  masc   79   preto
4    19  femin  48   preto
5    17  masc   80   loiro
6    18  masc   76   ruivo
7    20  masc   74   preto
8    19  femin  45   loiro
9    18  femin  55   loiro
10   18  femin  60   preto
11   17  femin  59   preto
12   16  masc   78   ruivo
13   23  femin  48 castanho
14   19  masc   90 castanho
15   17  masc   84   loiro
16   18  femin  48   loiro
17   17  femin  50   preto
18   17  femin  52   preto
19   18  masc   77 castanho
20   19  femin  46   loiro
21   25  masc   98   ruivo
22   16  femin  56 castanho
23   17  femin  49   loiro
24   17  femin  52 castanho
> attach(tabela)
> |
```

Figura 3.11: Importando uma tabela gravada previamente no computador

Continuando a tratar das medidas de posição tem-se a **média**, a **moda** e a **mediana**.

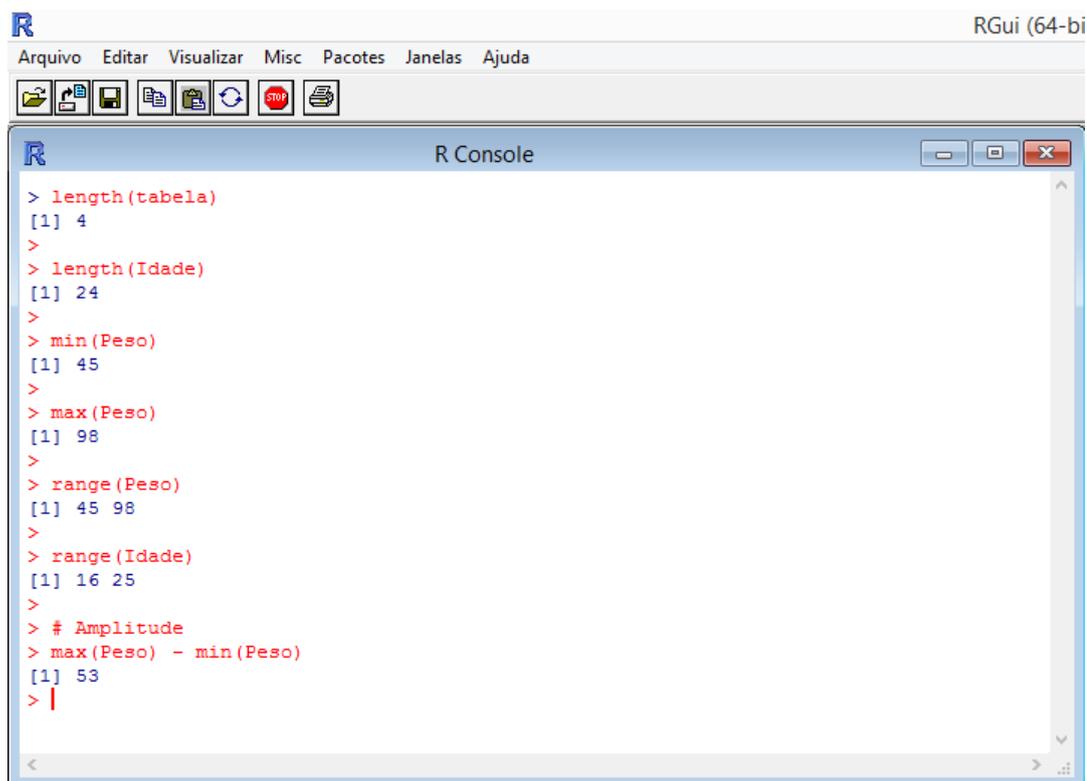
Já foi explicado que a média de uma variável quantitativa é encontrada somando todos os valores dessa variável dividindo pela quantidade desses valores, então pode-se escrever uma expressão para obter esse resultado utilizando as funções estudadas até o momento, que seria `sum(nome da variável) / length(nome da variável)`. Acontece que através do simples comando `mean()` obtem-se esse mesmo resultado.

O mesmo acontece para encontrar a mediana, onde é preciso apenas utilizar o comando `median()` para encontrá-la.

Já para encontrar a moda, é preciso criar uma tabela de frequência dos resultados da variável para poder identificá-la. Para isso utilizaremos a função `table()`.

Veja algumas situações:

- Para calcular a média de idade dos alunos da turma escreve-se `mean(Idade)`.
- Para calcular a mediana das idades dos alunos da turma escreve-se `median(Idade)`.
- Para identificar a moda das idades dos alunos da turma escreve-se `table(Idade)` e



```
R
RGui (64-bit)
Arquivo  Editar  Visualizar  Misc  Pacotes  Janelas  Ajuda

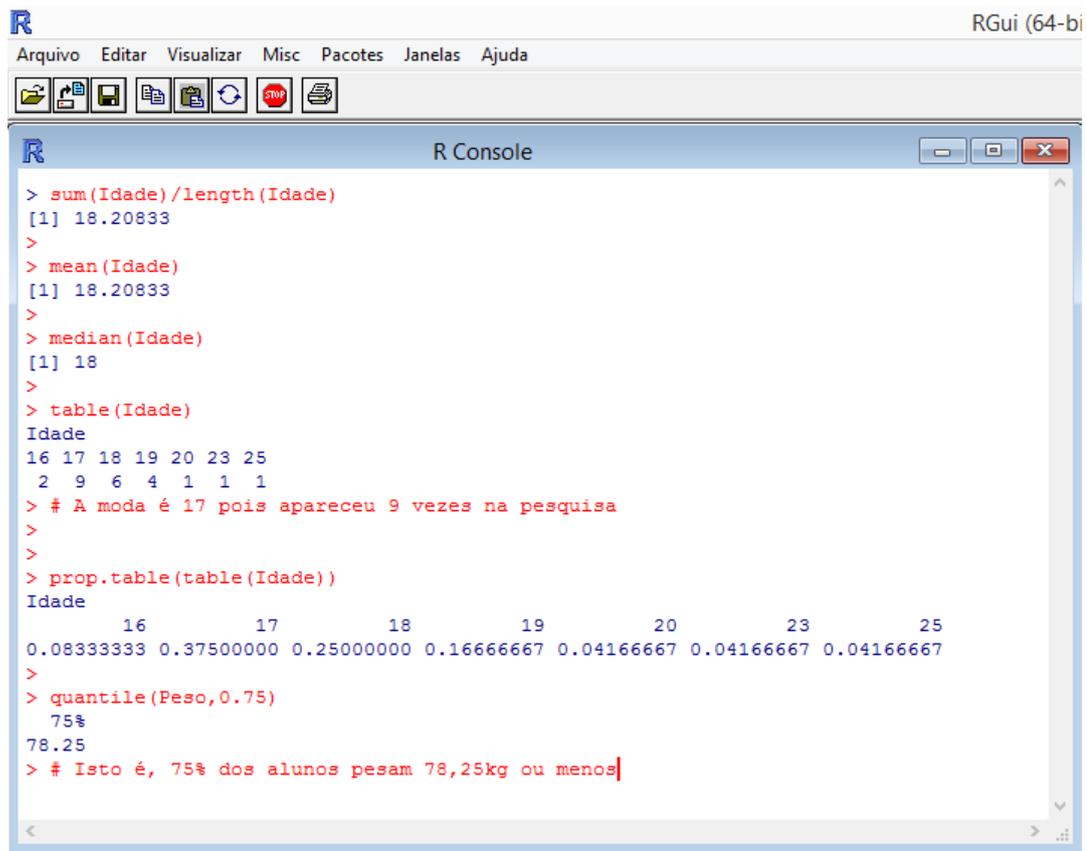
R Console
> length(tabela)
[1] 4
>
> length(Idade)
[1] 24
>
> min(Peso)
[1] 45
>
> max(Peso)
[1] 98
>
> range(Peso)
[1] 45 98
>
> range(Idade)
[1] 16 25
>
> # Amplitude
> max(Peso) - min(Peso)
[1] 53
> |
```

Figura 3.12: Explorando os dados da tabela utilizando alguns comandos básicos

observa-se o resultado com maior frequência.

Aproveitando o uso da função `table()`, aproveita-se para falar de uma informação que muitas vezes demanda um pouco de tempo para ser calculada no papel ou até mesmo com o auxílio de uma calculadora, trata-se da **frequência relativa**, mas com o software R é possível montar a tabela da frequência relativa também através de um único comando. Utiliza-se a função `prop.table()`, ou seja, se quer encontrar a frequência relativa das idades dos alunos da turma escreve-se `prop.table(table(Idade))`.

Para encerrar as medidas de posição, para calcular o percentil de uma certa variável utiliza-se a função `quantile(x,p)` onde `x` é o nome da variável e `p` a porcentagem. Então caso queira calcular o terceiro quartil (75%) da variável `Peso` usa-se `quantile(Peso, 0.75)`.



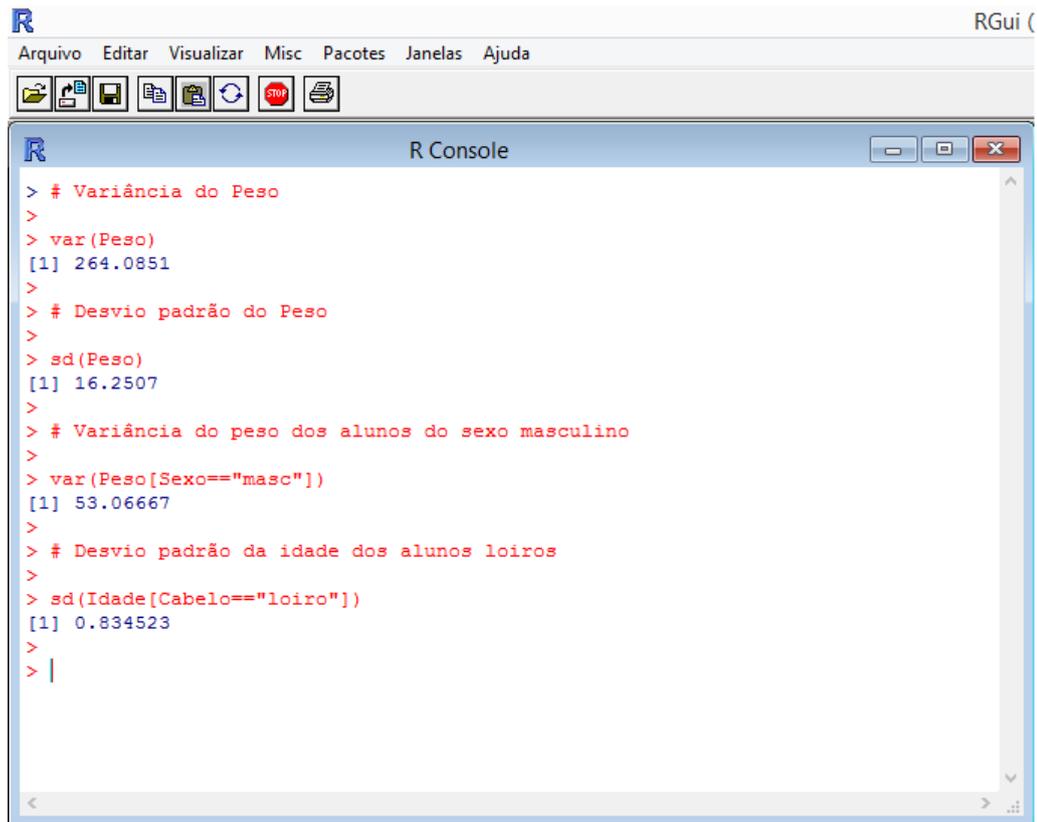
```
R
RGui (64-bit)
Arquivo  Editar  Visualizar  Misc  Pacotes  Janelas  Ajuda

R Console
> sum(Idade)/length(Idade)
[1] 18.20833
>
> mean(Idade)
[1] 18.20833
>
> median(Idade)
[1] 18
>
> table(Idade)
Idade
16 17 18 19 20 23 25
 2  9  6  4  1  1  1
> # A moda é 17 pois apareceu 9 vezes na pesquisa
>
> prop.table(table(Idade))
Idade
      16      17      18      19      20      23      25
0.08333333 0.37500000 0.25000000 0.16666667 0.04166667 0.04166667 0.04166667
>
> quantile(Peso,0.75)
75%
78.25
> # Isto é, 75% dos alunos pesam 78,25kg ou menos
```

Figura 3.13: Calculando medidas de posição

3.4.3 Medidas de Dispersão

Passando agora para as medidas de dispersão, tem-se a variância e o desvio padrão que são obtidas pelos comandos `var()` e `sd()` respectivamente. Para calcular a variância e o desvio padrão do peso dos alunos da turma deve usar `var(Peso)` e `sd(Peso)`. Caso deseje calcular a variância e/ou desvio padrão de algum valor específico na variável é possível especificar. Por exemplo, para calcular a variância do peso dos alunos do sexo masculino escreve-se `var(Peso[Sexo=="masc"])`, ou ainda para calcular o desvio padrão da idade dos alunos loiros escreve-se `sd(Idade[Cabelo=="loiro"])`.



```
> # Variância do Peso
>
> var(Peso)
[1] 264.0851
>
> # Desvio padrão do Peso
>
> sd(Peso)
[1] 16.2507
>
> # Variância do peso dos alunos do sexo masculino
>
> var(Peso[Sexo=="masc"])
[1] 53.06667
>
> # Desvio padrão da idade dos alunos loiros
>
> sd(Idade[Cabelo=="loiro"])
[1] 0.834523
>
> |
```

Figura 3.14: Calculando medidas de dispersão

3.4.4 Representação Gráfica

Foi visto nas aulas que uma forma muito útil para apresentação e organização dos dados de uma pesquisa é através da representação gráfica desses dados. Foram conhecidos os principais gráficos e estes serão os que serão construídos usando o software R.

Ao usar os comandos para construir os gráficos eles aparecerão em uma outra janela, e não mais na própria janela em que foi digitado o comando.

Começando com a função **plot(x,y)** que dará o *gráfico de dispersão* relacionando as variáveis “x” e “y” que devem ser quantitativas. Pode-se editar e/ou inserir informações no gráfico adicionando alguns parâmetros, por exemplo:

- **main= "Título"** : adiciona um título ao gráfico.
- **pch= número** : muda o tipo de ponto no gráfico de dispersão.
- **col= "nome" ou número da cor** : coloca cor no gráfico.
- **type= "l"** (letra L minúscula) : liga os pontos tornando-o um gráfico de linha.
- **lwd= número** e **lty= número** : mudam largura e estilo da linha respectivamente.

Continuando a utilizar como referência para a análise a tabela construída através da

pesquisa realizada com a turma (Figura 3.11).

Para construir o gráfico de dispersão relacionando Idade e Peso deve-se executar o comando `plot(Idade,Peso)`, daí surgirá uma outra janela com o gráfico a seguir:

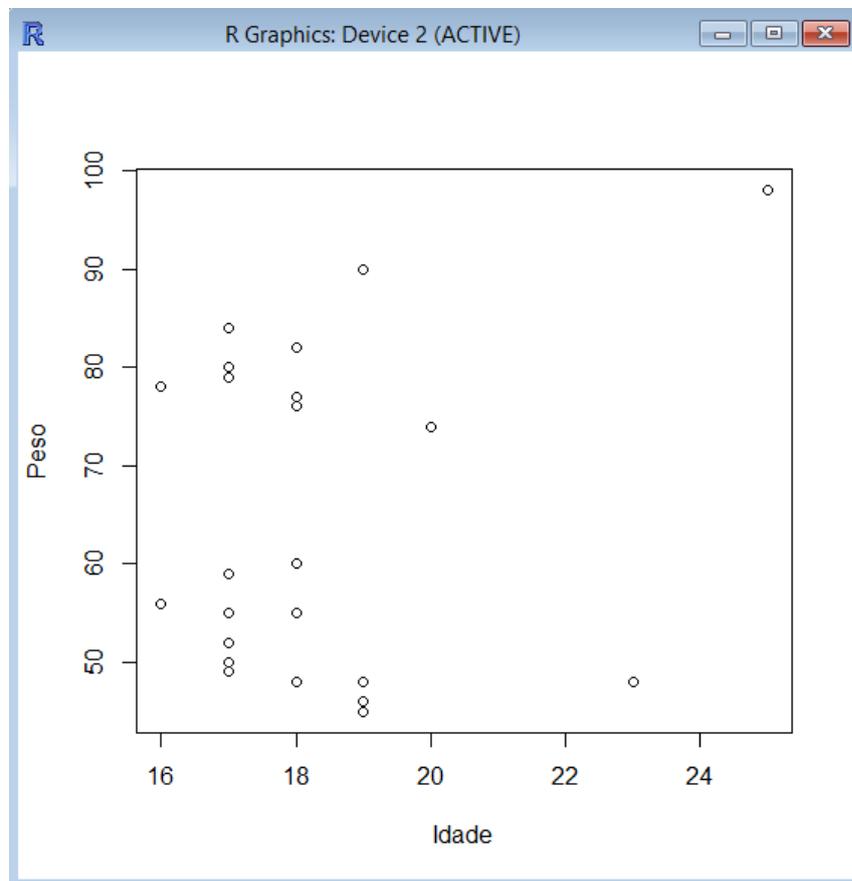


Figura 3.15: Gráfico de dispersão relacionando idade e peso

Deve-se agora propor uma formatação ao gráfico colocando um título, mudando o formato e colorindo o ponto. Pode-se escrever duas formatações para servirem como exemplo e logo depois os alunos ficam a vontade para mudarem o seu gráfico. Digita-se: `plot(Idade, Peso, main = "Gráfico de dispersão 1", pch = 4, col = "blue")` e `plot(Idade, Peso, main = "Gráfico de dispersão 2", pch = 18, col = "violet")`.

Para que os alunos possam comparar os dois gráficos basta usar a função `par(mfrow = c(x,y))` onde x é o número de linhas e y é o número de colunas. Este comando deve ser aplicado antes de escrever os dois comandos citados anteriormente. Assim, para que os dois gráficos apareçam lado a lado usa-se `par(mfrow=c(1,2))` ou, se preferir, um na parte superior e o outro na inferior usa-se `par(mfrow=c(2,1))`. Então as telas a seguir surgirão:

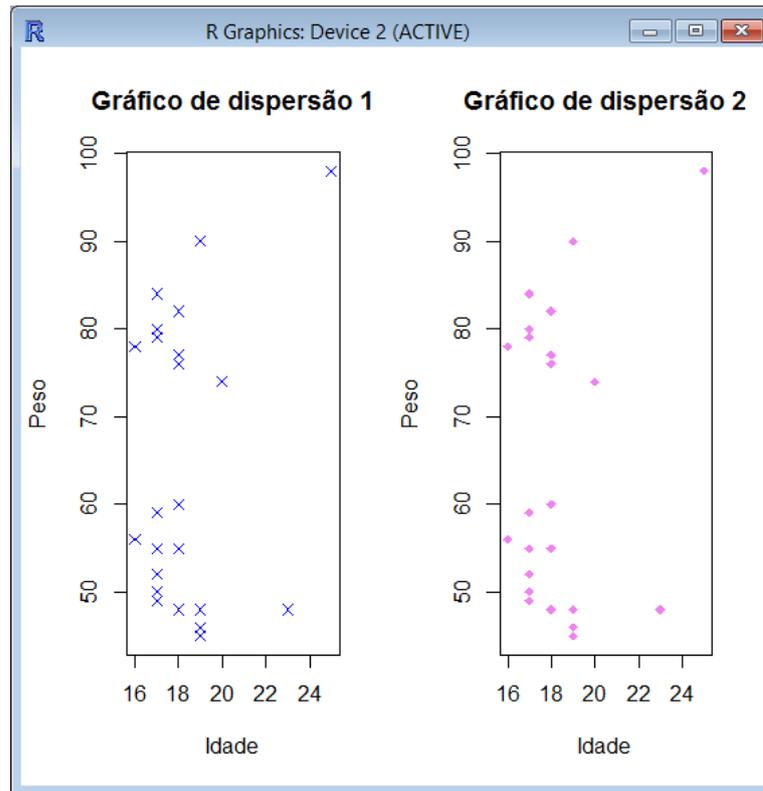


Figura 3.16: Gráfico de dispersão formatado 1

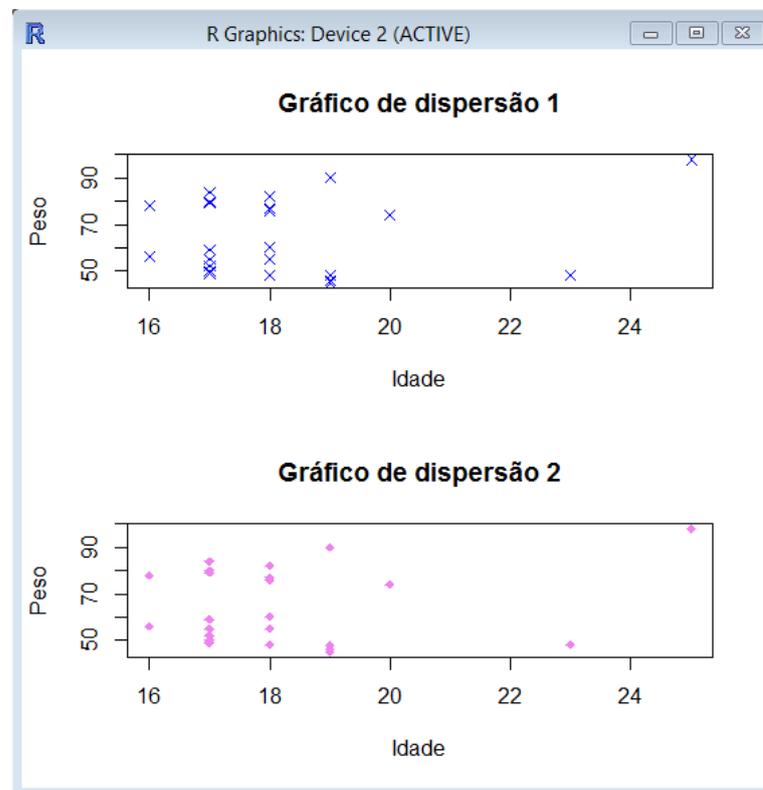


Figura 3.17: Gráfico de dispersão formatado 2

Para construir um **gráfico boxplot** para a idade dos alunos, por exemplo, utiliza-se a função **boxplot(x)** onde x é a variável a ser analisada. Seus principais argumentos são:

- **main=** : título do gráfico.
- **xlab=** : nome do eixo x.
- **ylab=** : nome do eixo y.
- **col=** : cor do preenchimento da caixa.
- **border=** : cor da linha (borda da caixa).
- **horizontal=TRUE** : a caixa aparecerá na horizontal.

O gráfico será obtido escrevendo **boxplot(Idade)**.

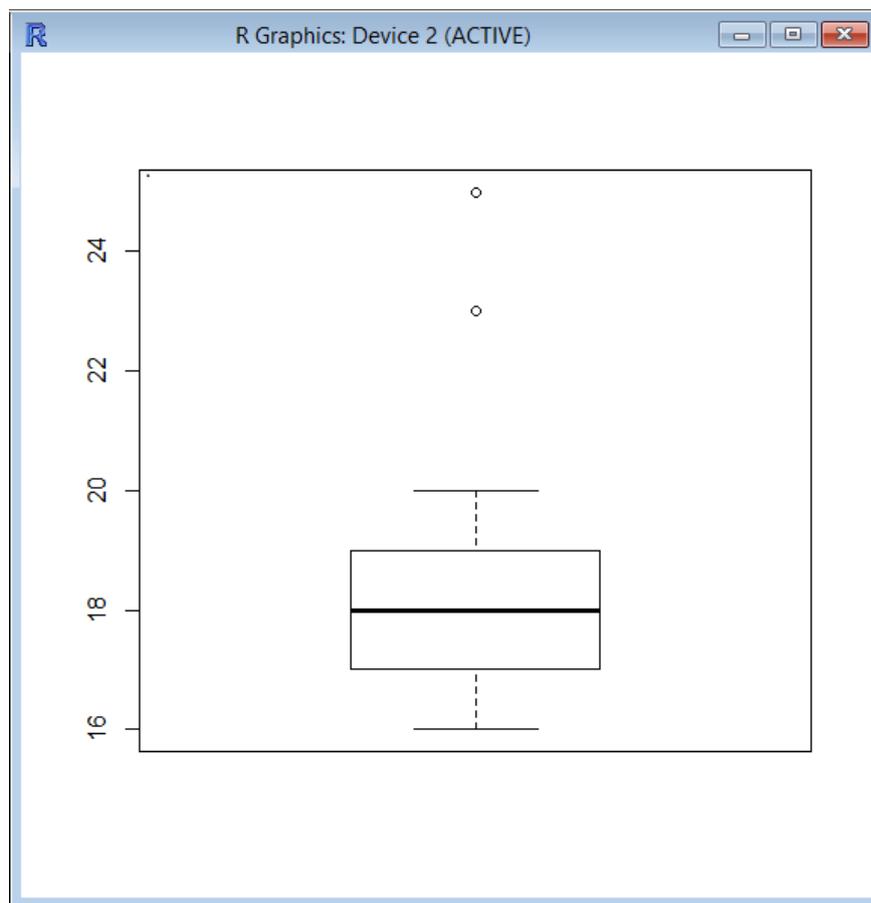


Figura 3.18: Gráfico boxplot da variável idade

Querendo aprimorar o gráfico basta adicionar algumas informações importantes a ele. Coloca-se um título, nome para o eixo y e também uma cor para o interior da caixa e para sua borda. Para isso escreve-se **boxplot(Idade, main= "Boxplot", ylab= "Idade", col= "salmon", border= "blue")** e o gráfico fica assim:

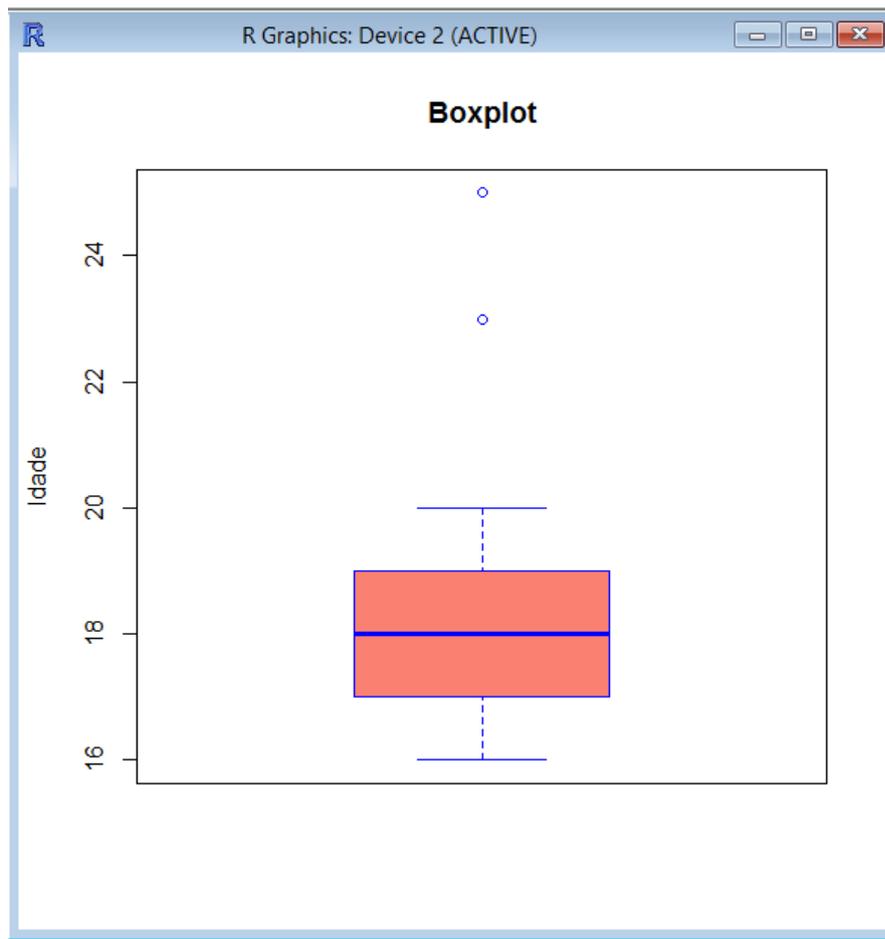


Figura 3.19: Gráfico boxplot da variável idade formatado

Seguindo o conteúdo, ensina-se a construir um histograma no software R em que é utilizada a função **hist(x)** onde x é a variável. Os principais comandos que ajudam a formatar o histograma são:

- **main=** : título do gráfico.
- **xlab=** : nome do eixo x.
- **ylab=** : nome do eixo y.
- **xlim=** : limite do eixo x.
- **ylim=** : limite do eixo y.
- **breaks=** : quantidade de classes (intervalos)
- **col=** : cor do preenchimento do histograma.
- **border=** : cor da linha (borda do histograma).

Construindo o histograma da Idade dos alunos escreve-se **hist(Idade)** e o gráfico a seguir aparece:

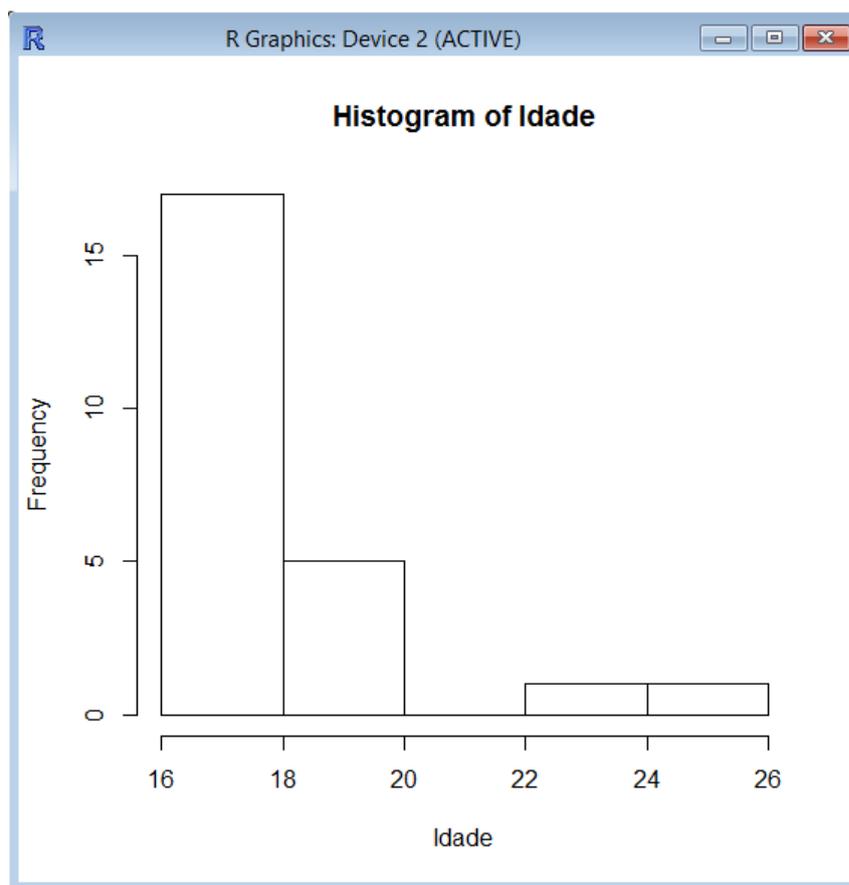


Figura 3.20: Histograma da variável idade

Formatar o histograma é, por exemplo, dar um título a ele, nome aos eixos, uma cor para preencher as colunas e aumentar o limite do eixo y. Para isso, pode-se escrever `hist(Idade, main="Histograma", col="red", xlab="Idade", ylab="Quantidade de alunos", ylim=c(0,20))` e forma o gráfico mostrado na Figura 3.21 mais a frente.

O **gráfico de barras**, é obtido a partir da função `barplot(x)` onde x é o vetor de frequência dos dados de uma variável que já foi visto e é obtido pela função `table()`. Os principais parâmetros para a formatar o gráfico de barras são os mesmos do histograma.

O comando `barplot()` gera um gráfico com as barras na vertical, entretanto, caso queira-se as barras na horizontal, adiciona-se o argumento `hor=TRUE` à função `barplot()`.

Assim, para construir o gráfico de barras referente a cor de cabelo dos alunos da turma escreve-se `barplot(table(Cabelo))` ou `barplot(table(Cabelo),hor=TRUE)` caso queira as barras na vertical ou horizontal respectivamente. Pode-se usar a função `par(mfrow=c(x,y))` para fazer o comparativo dos dois gráficos.

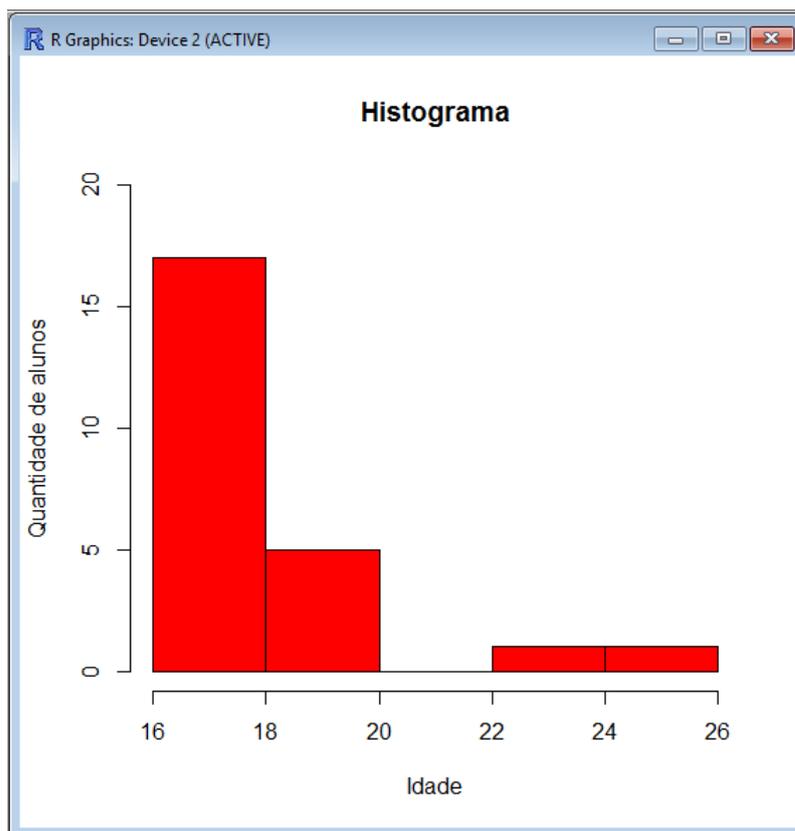


Figura 3.21: Histograma da variável idade formatado

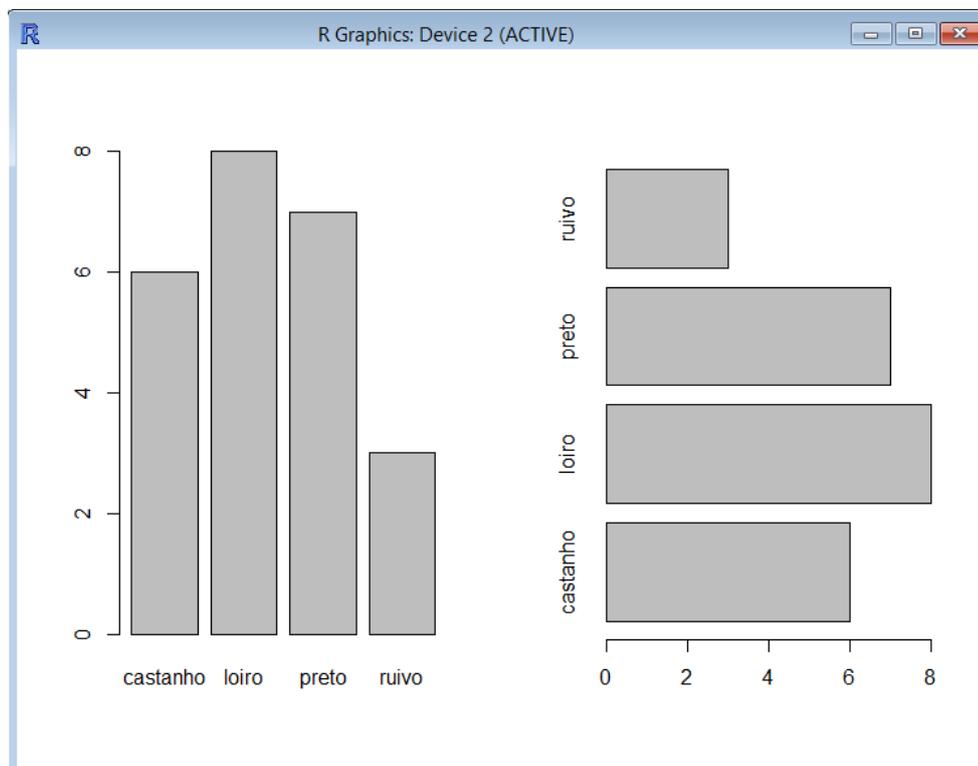


Figura 3.22: Gráfico de Barras da variável cor dos cabelos

Para personalizar o gráfico usando os argumentos conhecidos, pode-se escrever `barplot` (`table(Cabelo)`, `main= "Gráfico de barras"`, `col= 3`, `xlab= "Cor do cabelo"`, `ylab = "Quantidade de alunos"`) e obter:

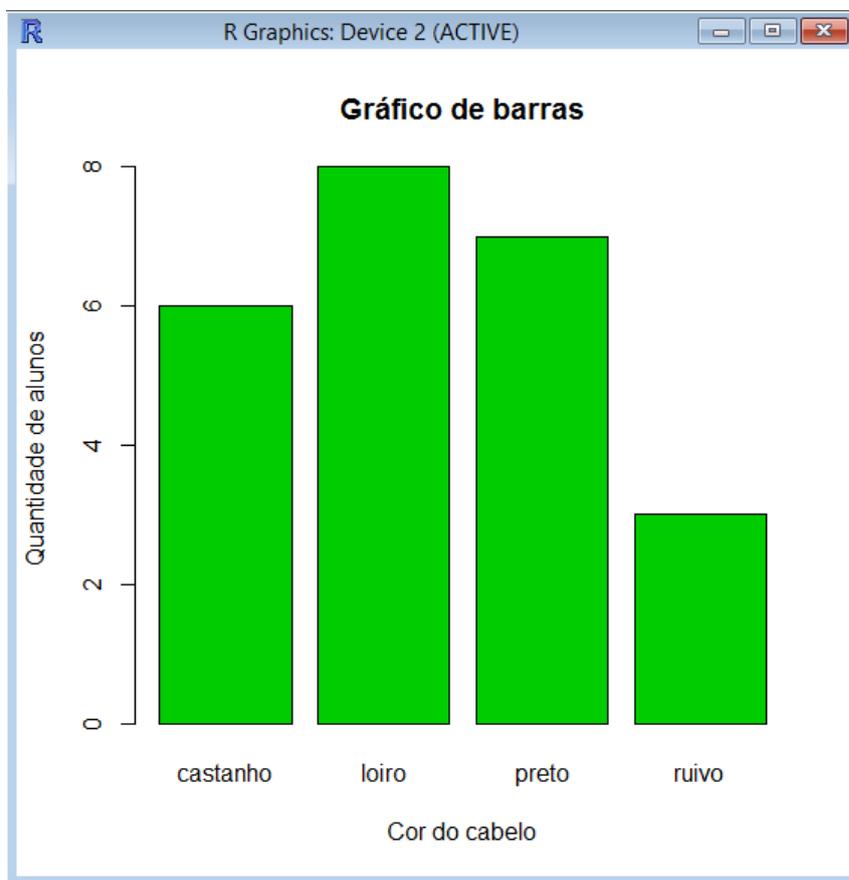


Figura 3.23: Gráfico de Barras da variável cor dos cabelos formatado

Outro tipo de gráfico importante na análise de dados que é possível obter utilizando o R é o **gráfico de setores**. Para isso utiliza-se a função `pie(x)` onde `x` é o vetor com as frequências da variável estudada. Alguns parâmetros são importantes na personalização desse gráfico:

- `main=` : título do gráfico.
- `labels=` : vetor com texto destinado a cada fatia.
- `col=` : vetor com cores para cada fatia.

Para exibir o gráfico de setores que represente a divisão dos alunos por sexo já personalizando pode-se escrever `pie(table(Sexo)`, `main= "Gráfico de Setores"`, `col= c("red", "blue")`) e obter:

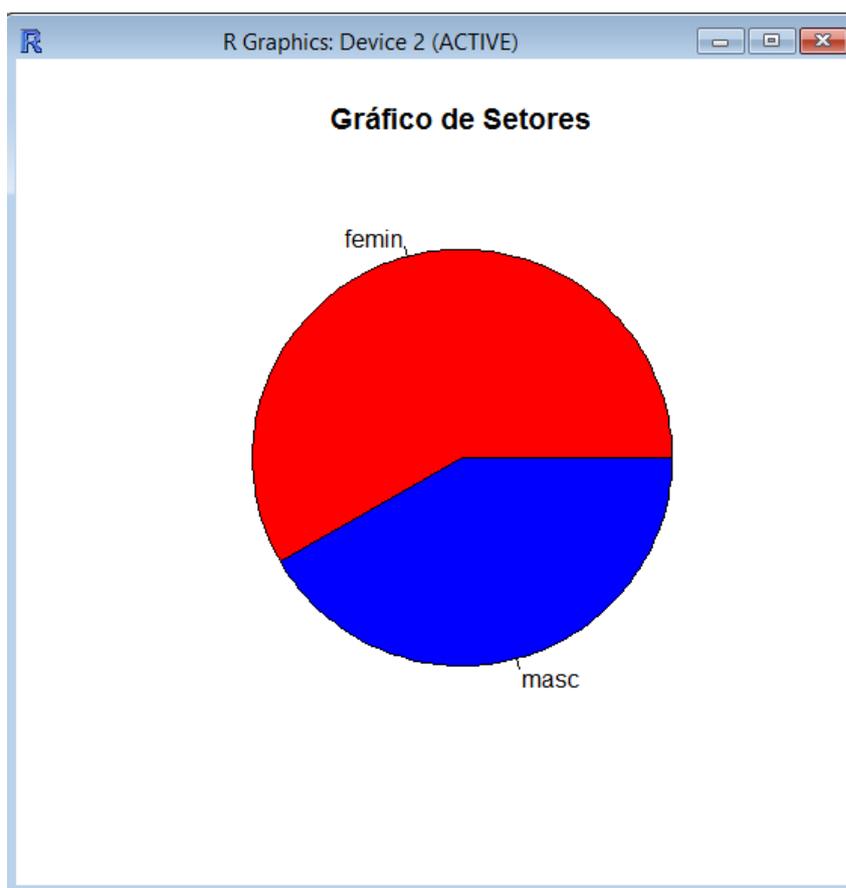


Figura 3.24: Gráfico de Setores da variável sexo

Para editar o texto das fatias utiliza-se o `labels()`, basta adicionar ao comando. Então escreva `pie(table(Sexo), main= "Gráfico de Setores", col= c("red", "blue"), labels= c("FEMININO", "MASCULINO"))` e o gráfico fica como na Figura 3.25.

É possível adicionar uma legenda ao gráfico, usa-se a função `legend(x)` onde `x` é a posição onde estará localizada a legenda que podem ser "bottomright" (embaixo a direita), "bottom" (embaixo), "bottomleft" (embaixo a esquerda), "topright" (em cima a direita), "top" (em cima), "topleft" (em cima a esquerda), "right" (a direita), "center" (centro) e "left" (a esquerda). Os parâmetros mais importantes da função `legend()` são:

- `legend=` : vetor com o nome de cada fatia.
- `fill=` : vetor com as cores de cada fatia.
- `cex=` : tamanho da legenda.

Como exemplo pode-se construir dois gráficos de setores relacionando a cor do cabelo dos alunos onde em um deixa-se o nome nas fatias e em outro usa-se a legenda e assim poder comparar ambos.

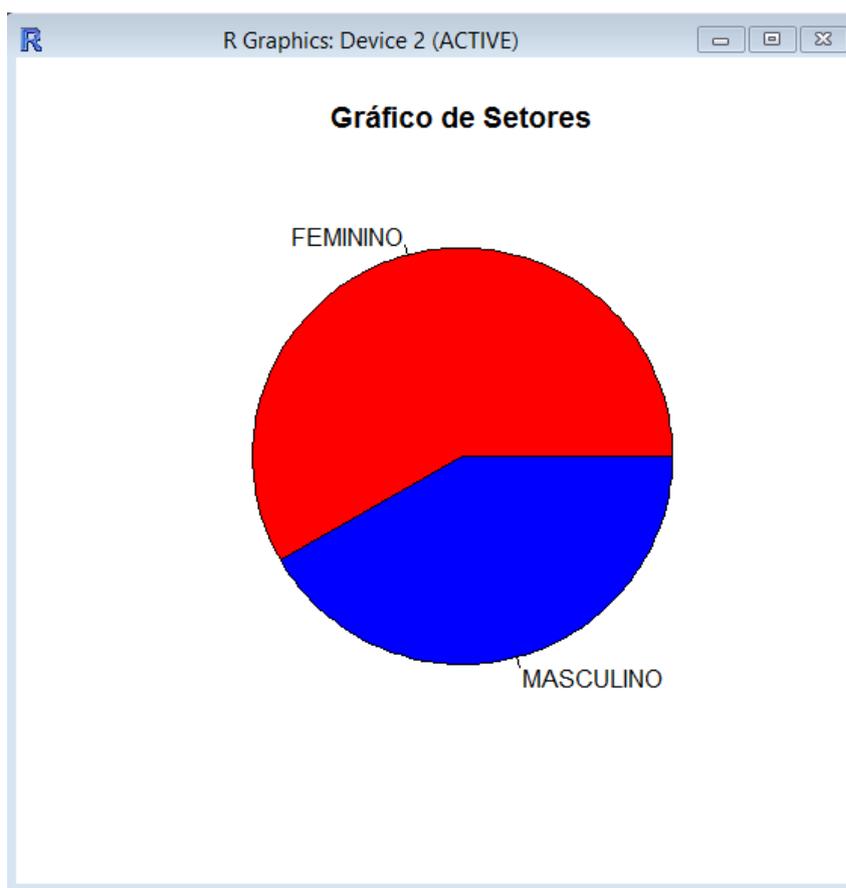


Figura 3.25: Gráfico de Setores da variável sexo editado o texto das fatias

Para o primeiro gráfico escreve-se o comando `pie(table(Cabelo), main= "Gráfico de Setores 1", col= c("red", "blue", "violet", "green"))` o que dá como resultado o gráfico da Figura 3.26.

Para apagar o nome nas fatias, basta deixar vazio os espaços correspondentes aos nomes no parâmetro `labels=` e construir uma legenda para o gráfico utilizando a função `legend()`. Para isso utiliza-se dois comandos:

```
> pie(table(Cabelo), main= "Gráfico de Setores 2", col= c("red","blue",
"violet", "green"), labels= c( , , , ))
```

Com o gráfico aberto clica-se na janela dos comandos e escreve-se:

```
> legend("topright", legend = c( "Castanho", "Loiro", "Preto","Ruivo"),
fill = c( "red", "blue", "violet", "green"))
```

O gráfico está descrito na Figura 3.27.

Se quer melhorar ainda mais a apresentação do gráfico de setores pode-se acrescentar na fatia a frequência relativa referente a cada cor de cabelo. Lembrando que para encontrar

essa informação é utilizada a função `prop.table()`. Então escreve-se:

```
> pie(table(Cabelo), main= "Gráfico de Setores 3", col= c ( "red",  
"blue", "violet", "green"), labels= c("25%", "33,3%", "29,2%", "12,5%"))  
> legend( "bottomright", legend = c ("Castanho", "Loiro", "Preto",  
"Ruivo"), fill = c ("red","blue","violet","green"))
```

E tem-se o gráfico da Figura 3.28.

Ao encerrar os tópicos do curso o professor pode propor uma “atividade final” para os alunos que consiste em uma pesquisa com coleta de dados feita na escola ou na comunidade sobre algum assunto que esteja em evidência no momento.

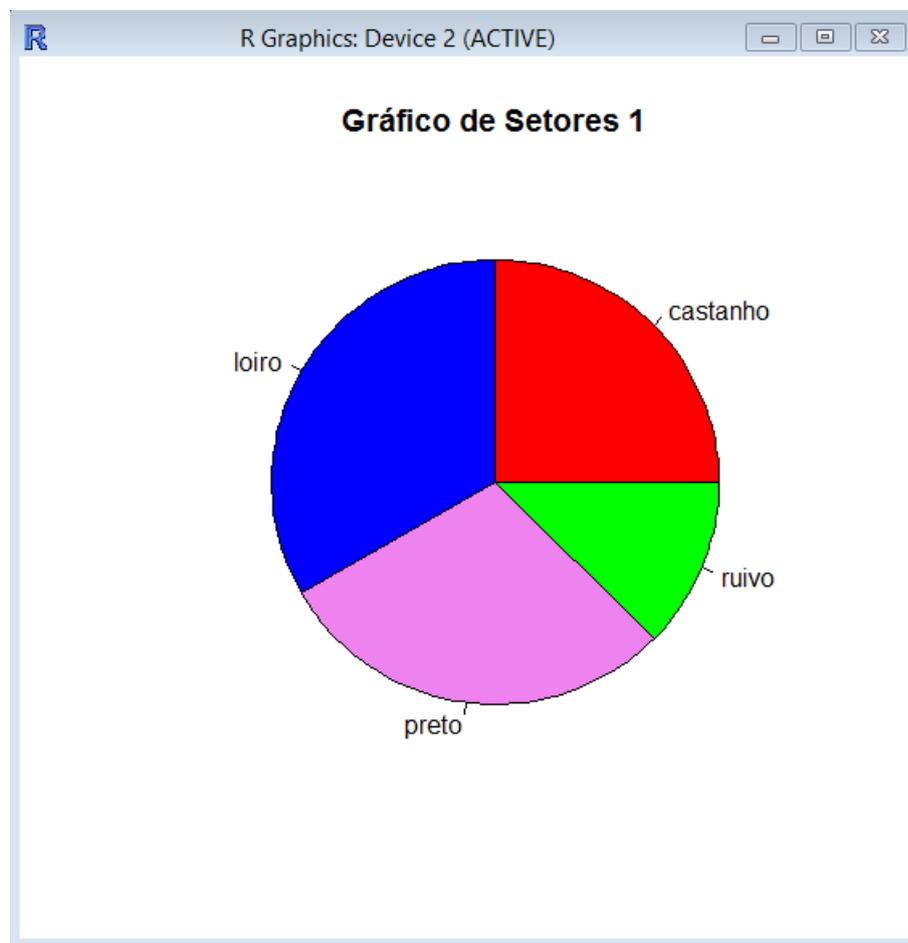


Figura 3.26: Gráfico de Setores da variável cor do cabelo

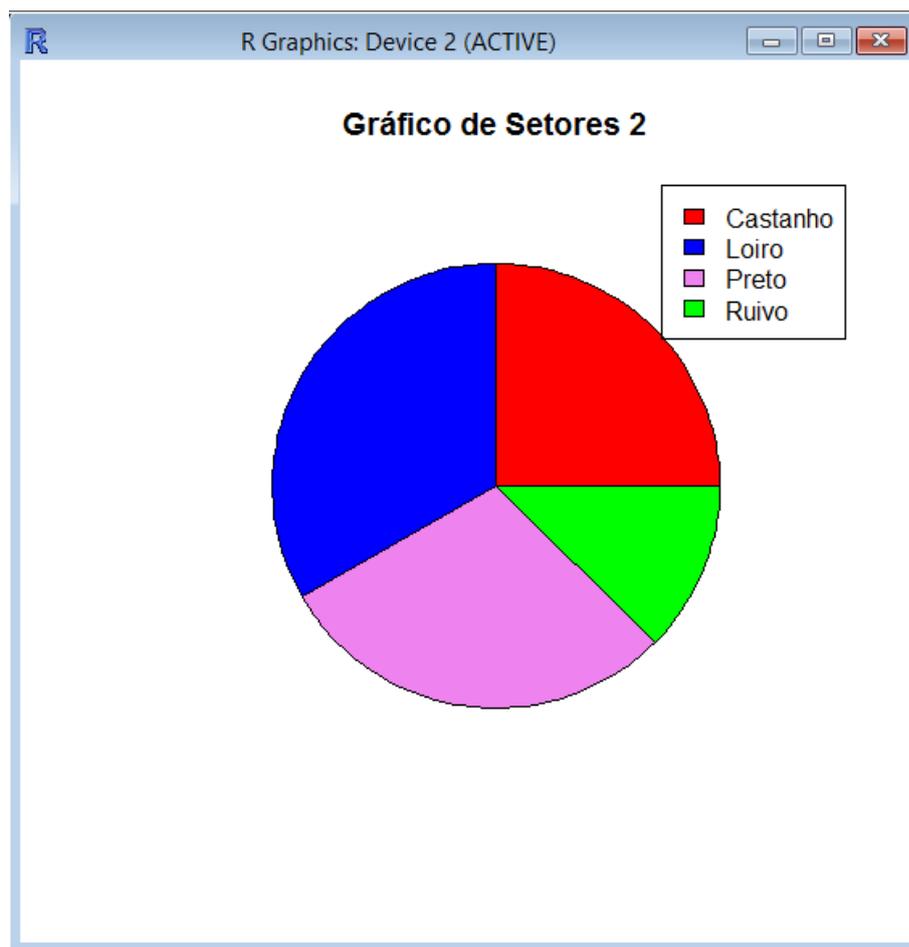


Figura 3.27: Gráfico de Setores da variável cor do cabelo com legenda

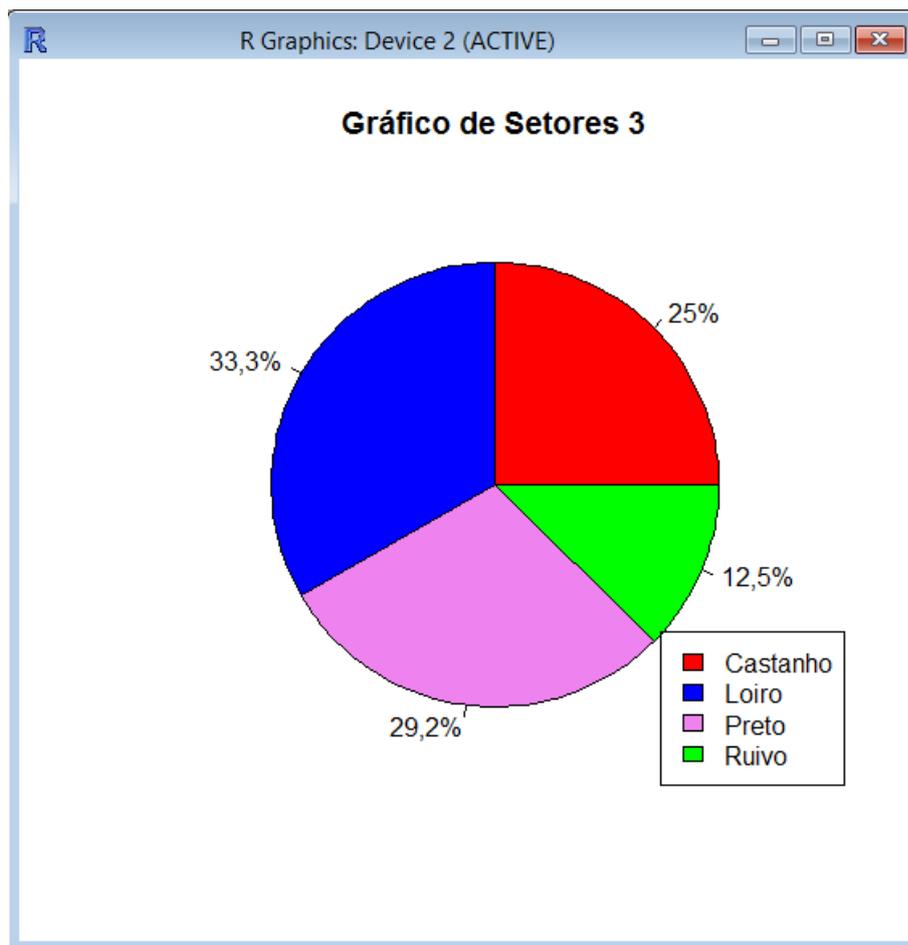


Figura 3.28: Gráfico de Setores da variável cor do cabelo com frequência relativa e legenda

Capítulo 4

Relatório e Resultados

O curso foi ministrado com duas turmas do 3º ano do Ensino Médio em duas escolas estaduais: Unidade Escolar Lucídio Portella (Teresina-PI) e Centro de Ensino Odolfo Medeiros (Caxias-MA) e contou com total apoio por parte da direção de ambas as escolas.

Foi trabalhado um total de 20h/a sendo 10h em sala de aula onde foram passados os conteúdos contidos no Capítulo 2 desta dissertação através de aula expositiva com o auxílio do livro didático e data show, e 10h em laboratório onde foram abordados os conteúdos do Capítulo 3 dessa dissertação em que através dos conhecimentos que os alunos obtiveram em sala puderam aplicar e entender o software R.

Assim que o curso foi apresentado aos alunos gerou de imediato uma certa animação por se tratar de uma novidade e pela utilização de computadores. No Anexo I encontram-se fotos dos alunos no laboratório usando o software R.

A primeira etapa em sala de aula foi bem produtiva, com a exposição dos conteúdos e resolução de problemas contidos principalmente no livro didático adotado.

A segunda etapa, que foi o alvo de grande expectativa tanto dos alunos como do professor começou com um pouco de dificuldade pois chegando ao laboratório nos deparamos com uma realidade da maioria das escolas públicas, que é o pouco número de computadores aptos a serem utilizados o que forçou a utilização compartilhada das máquinas por até 3 alunos. Mas os alunos superaram e foi um período de muita descoberta e consequente solidificação da aprendizagem.

As aulas foram ministradas nos horários destinados a disciplina Matemática que são 4h/a semanais para o 3º ano do ensino médio, divididas em dois dias com 2h/a, ou seja, os alunos estiveram em horário normal de aula evitando evasão por motivos alheios à escola.

No primeiro dia, antes de iniciarmos o manuseio do software foi explicado aos alunos do que se trata o Projeto R e o que é esse software, bem como suas aplicações. Após essa explanação os alunos foram autorizados a ligarem as máquinas, procurarem o ícone do R para abrir o programa e assim iniciar a experiência.

Como proposto no material, o trabalho foi iniciado com as operações conhecidas pelos alunos e que o R pode calcular. Foram propostas questões envolvendo essas operações para que os alunos resolvessem no caderno e “conferissem” utilizando o R.

No segundo dia, foi trabalhado a tópico “Armazenamento de dados” onde os alunos aprenderam a criar objetos, vetores, matrizes e data frames, sempre abrindo espaço para que eles colocassem dados próprios e/ou pessoais nas tabelas.

No terceiro dia, foi feita uma pesquisa entre os próprios alunos e assim foi construída a tabela para posterior análise. Essa etapa foi considerada super interessante por eles pelo fato deles próprios serem “os alvos” de estudo. A partir daí começamos a calcular as medidas de posição e dispersão sempre com a análise do contexto após cada resultado obtido.

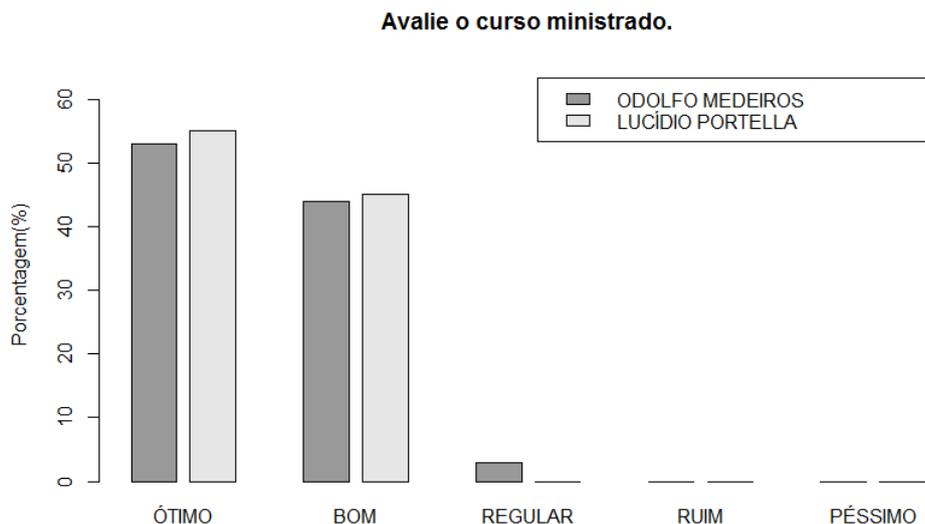
O quarto dia foi destinado à construção de gráficos. Os alunos puderam construir os mais variados tipos de gráficos propostos pelo professor e na hora de formatar tiveram autonomia para aprimorarem seus gráficos como preferissem. Ao final desse dia receberam um questionário para que fossem a campo coletar dados sobre Bullying e Chikungunya, nas escolas Lucídio Portella e Odolfo Medeiros, respectivamente.

No quinto dia os alunos reuniram os dados coletados na pesquisa de campo, produziram a tabela com os resultados e construíram os respectivos gráficos. Após essa apuração debatemos e discutimos a respeito desses assuntos que estão em destaque nessas escolas. Por fim, os alunos responderam a um questionário que tinha como objetivos saber se o curso contribuiu para a aprendizagem deles, saber das perspectivas deles em relação ao uso desse software e obter uma avaliação do curso ministrado.

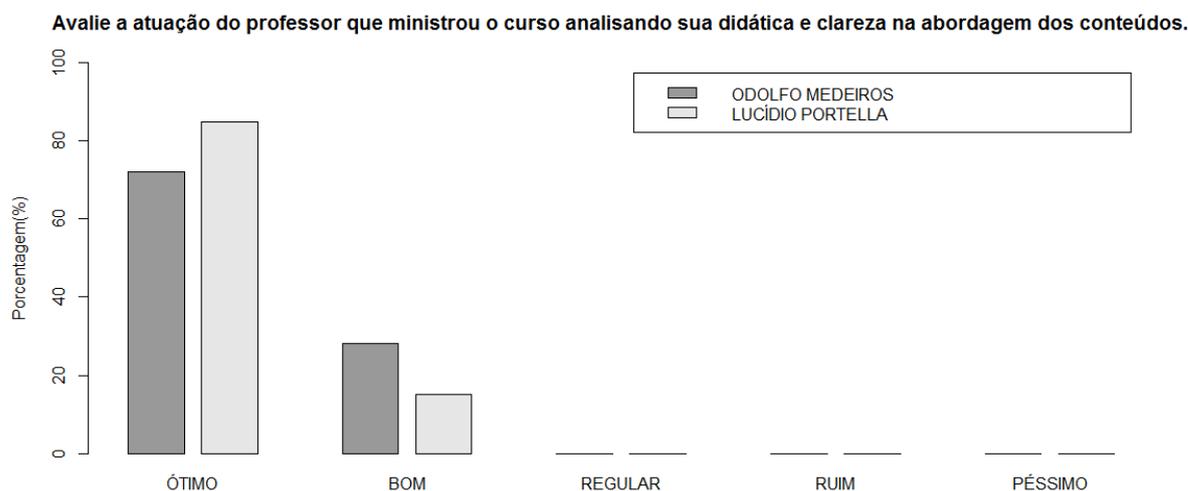
Os resultados do questionário foram, em uma análise geral, parecidos nas duas turmas e mostraram que o curso atingiu seus objetivos, que eram, principalmente, trazer uma dinâmica nova para as aulas de matemática e contribuir para a aprendizagem dos alunos. O único ponto que talvez tenha interferido negativamente foi a já citada precariedade na estrutura dos laboratórios com poucas máquinas em funcionamento.

O curso foi avaliado como BOM ou ÓTIMO por quase a totalidade dos alunos, ape-

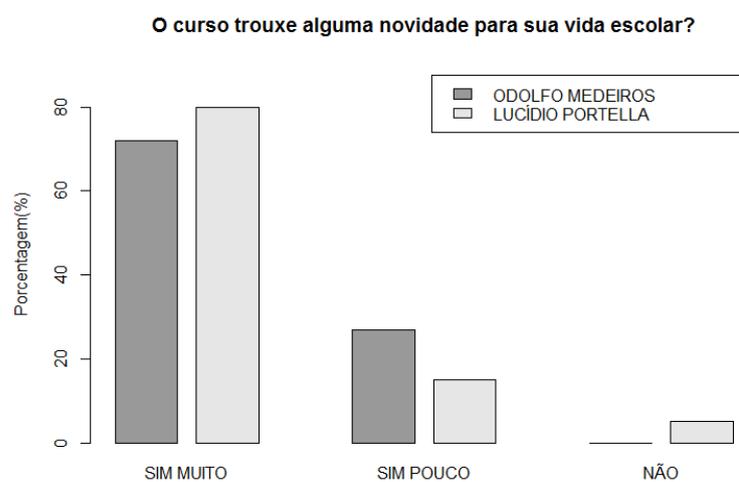
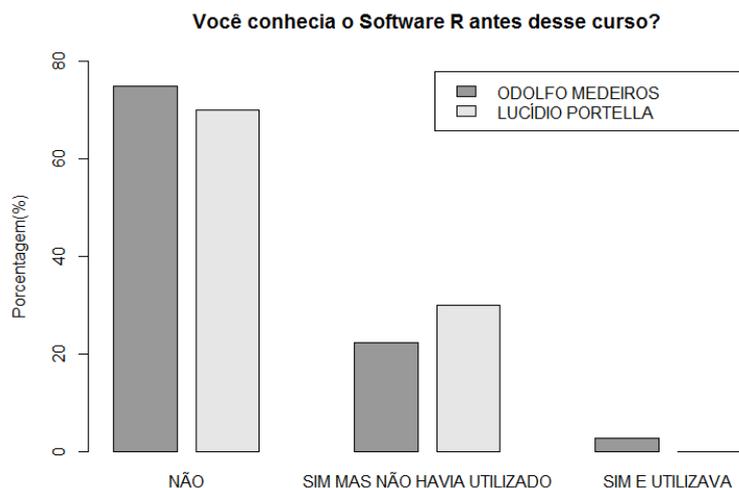
nas um aluno avaliou como REGULAR e nenhum avaliou como RUIM ou PÉSSIMO. Analisando os percentuais graficamente tem-se:



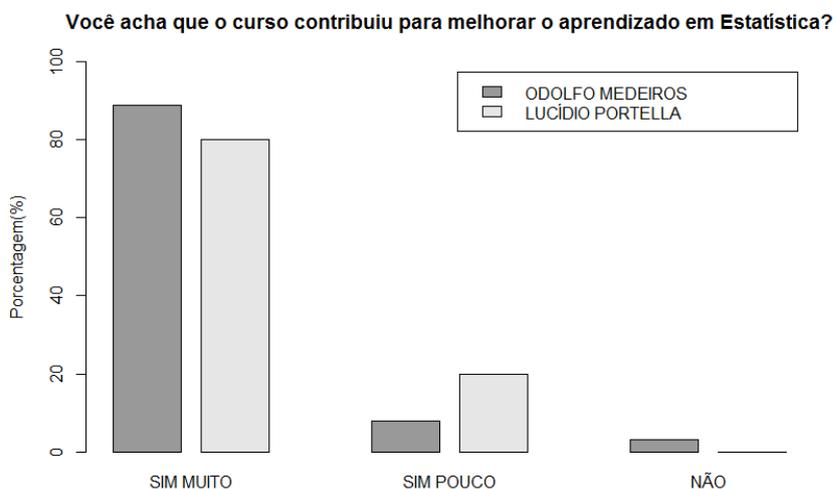
Quando perguntados sobre a didática e a clareza na abordagem dos conteúdos por parte do ministrante o resultado foi melhor, o que deixou a sensação que realmente o quesito estrutura técnica do laboratório não influenciou tanto na avaliação final do curso, pois 100% dos alunos avaliaram este item como ÓTIMO ou BOM, sendo uma grande parcela optando pela nota máxima. Veja o gráfico:



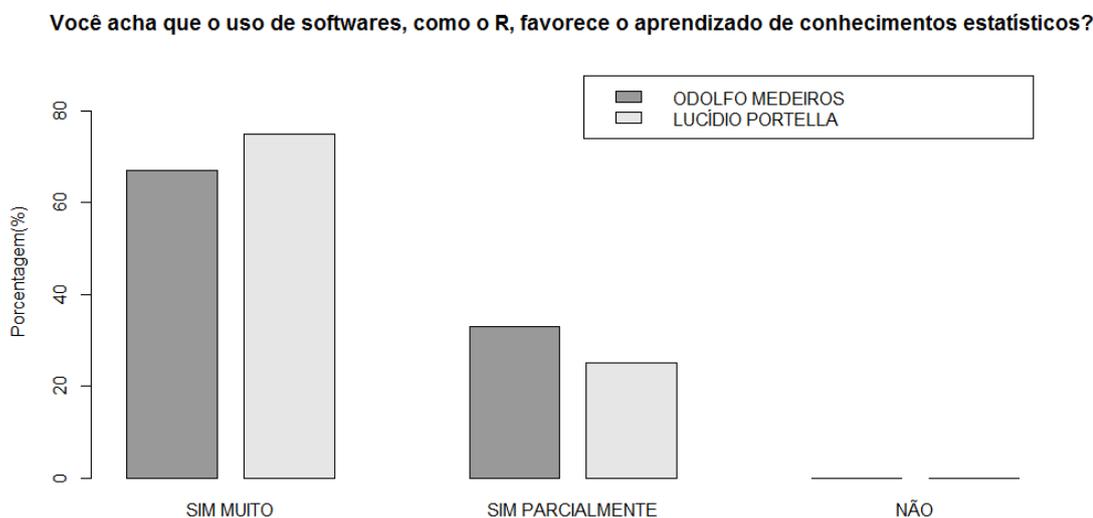
Um ponto que ajudou o curso a obter um resultado final positivo foi também a questão da novidade trazida para os alunos, tanto o fato de trabalhar a matemática em laboratório como o fato de poucos conhecerem o programa antes dessas aulas. Quando perguntados se conheciam o software R antes do curso e se esse curso trouxe alguma novidade para sua vida escolar foram obtidos os seguintes resultados:



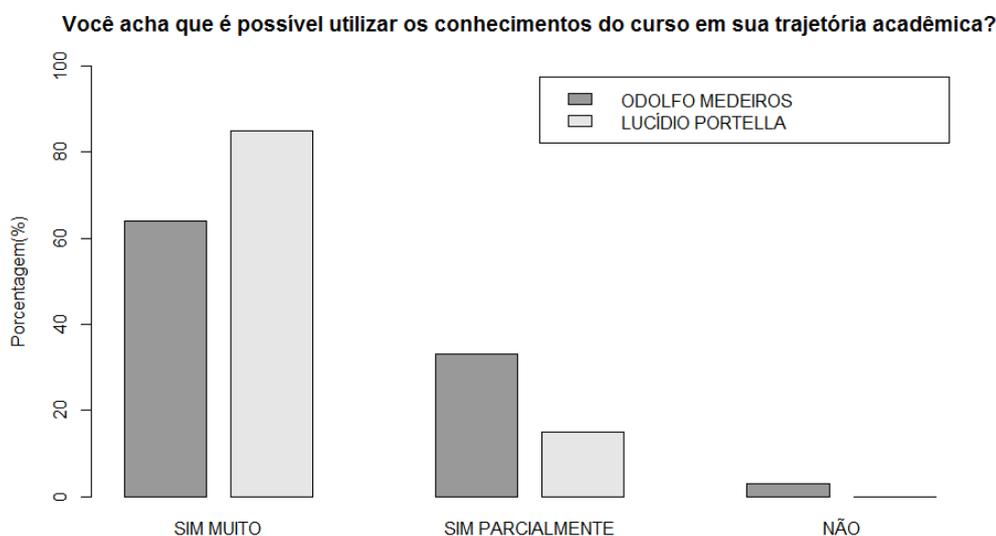
Como o maior objetivo traçado em estratégias como esse curso é a aprendizagem, foi perguntado aos alunos se o curso contribuiu para melhorar o aprendizado em Estatística e o resultado foi satisfatório pois a grande maioria concordou que melhorou muito. Observe o gráfico com o resultado dessa pergunta:



Outro resultado satisfatório foi quando avaliou-se a ajuda que o uso de softwares, como o R, trazem para a compreensão de conteúdos ministrados em sala de aula, nesse caso a Estatística. Todos os alunos acham que o uso de softwares favorecem MUITO ou PARCIALMENTE o aprendizado de Estatística. Veja o resultado no gráfico:



Durante todo o curso e principalmente em sua parte final com a pesquisa de campo que envolveu assuntos como bullying e chikungunya os alunos foram instruídos que a Estatística não está presente apenas nas aulas de matemática, mas sim em todas as disciplinas e ambientes internos ou externos à sala de aula. E quando lhes foi perguntado se consideravam possível utilizar os conhecimentos adquiridos no curso em sua trajetória acadêmica, seja no Ensino Médio ou Superior, somente um aluno respondeu achar que NÃO utilizaria, já o restante considerou que poderia utilizar MUITO ou PARCIALMENTE esses conhecimentos. Veja o resultado no gráfico a seguir:



Considerações Finais

O ensino de Matemática com o auxílio de ferramentas vem sendo amplamente discutido e exigido pelos profissionais em educação, mas não deve ser tratada como mágica e nem trabalhada de forma aleatória. O uso de uma ferramenta deve ser bem planejado e deve-se estudar o seu manuseio e quais resultados se quer obter com o seu uso. Isso ajuda a saber se aquela ferramenta escolhida realmente ajudará no ensino e conseqüente aprendizagem dos alunos.

No caso desse trabalho, analisando os resultados obtidos no curso, percebe-se que o uso do software R como auxílio no processo de ensino-aprendizagem de Estatística acarretou ótimos resultados. Isso ocorreu, principalmente, pelo fato do uso de ferramentas como o R, ultrapassarem as barreiras das paredes da sala de aula e fazem com que os alunos sintam que aqueles conhecimentos podem sim ser utilizados em seu cotidiano.

Vale ressaltar, novamente, que o uso de ferramentas como o software R deve ser compreendido como complemento às aulas expositivas ministradas em sala de aula, onde o aluno aprende os conceitos básicos, calcula as medidas estatísticas e estuda os gráfico, ou seja, o uso do R deve ser tratado como uma forma de concretização do aprendizado desses conteúdos. O R não deve ser utilizado como único meio para se compreender estatística.

Esse curso trouxe novidades para os alunos a partir do momento em que foram utilizados exemplos próximos a eles, analisou-se a própria realidade dos alunos, deixando-os perto do conteúdo estudado e fazendo com que a assimilação e compreensão dos tópicos estatísticos estudados fossem efetivadas com mais rigidez.

Além do êxito com os alunos, o curso também foi reconhecido pela direção de ambas as escolas trabalhadas como um projeto promissor. Na escola Odolfo Medeiros o laboratório já estava há mais de um ano sem uso e a diretora viu com muito bons olhos essa revitalização do laboratório. Na escola Lucídio Portella o curso foi incluído como um projeto da escola no programa *Jovem de Futuro* que trata-se de um programa criado pelo

Instituto Unibanco e é definido pelo Governo do Estado do Piauí, em seu site, como uma ação de gestão escolar para resultados, que oferece às escolas públicas do Ensino Médio apoio técnico e financeiro para que seu desempenho seja melhorado substancialmente e visa melhorar o IDEB (Índice de Desenvolvimento da Educação Básica) das escolas.

Portanto o uso de ferramentas, tecnológicas ou nao, deve ser estimulado em todas as áreas do conhecimento para obter uma formação mais completa e sólida dos jovens encaminhando-os mais capacitados para o Ensino Superior e assim tornarem-se profissionais bem sucedidos.

Referências Bibliográficas

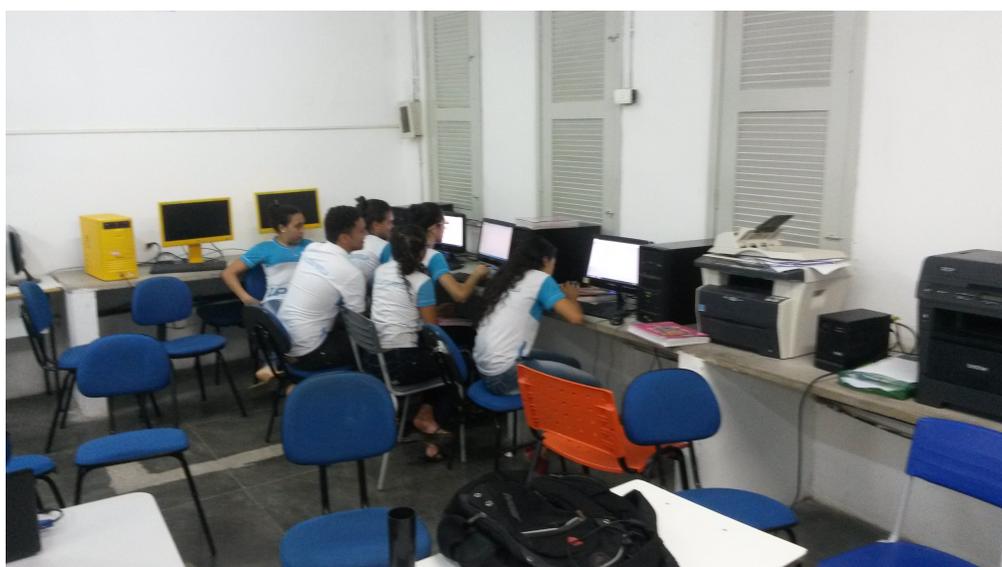
- [1] CRESPO, Antônio Arnot. **Estatística fácil** / Antônio Arnot Crespo. - 19.ed. atual. - São Paulo: Saraiva, 2009.
- [2] DANTE, Luiz Roberto. **Matemática: contexto e aplicações, vol 3** / Luiz Roberto Dante. – São Paulo: Ática, 2010.
- [3] GIOVANNI, José Ruy. **Matemática completa, 2ª série** / José Ruy Giovanni, José Roberto Bonjorno. - 2. ed. renov. - São Paulo: FTD, 2005. - (Coleção matemática completa)
- [4] MORETTIN, Pedro Alberto. **Estatística Básica**/ Pedro A. Morettin, Wilton O. Bussab. - 6. ed. - São Paulo: Saraiva, 2010.
- [5] PAIVA, Manoel. **Matemática: Paiva, 3** / Manoel Paiva. - 2. ed. - São Paulo: Moderna, 2013.
- [6] **The R Project for Statistical Computing**. Disponível em: <<https://www.r-project.org/>>.
- [7] **FGV/IBRE. Estatística com R**. Disponível em: <https://www.youtube.com/watch?v=theKHZaX60w&index=2&list=PLuQWWXrHQLiejECp_ldRPEjuVnJrG7fVS>.
- [8] **Parâmetros Curriculares Nacionais Ensino Médio**. Disponível em <<http://portal.mec.gov.br/seb/arquivos/pdf/ciencian.pdf> >.
- [9] **PCN+ Ensino Médio - Ciências da Natureza, Matemática e suas Tecnologias**. Disponível em <<http://portal.mec.gov.br/seb/arquivos/pdf/CienciasNatureza.pdf>>.

-
- [10] **Minicurso de Estatística Básica: Introdução ao software R**. Disponível em:
<http://www.uft.edu.br/engambiental/prof/catalunha/arquivos/r/r_bruno.pdf>.
- [11] **Introdução ao uso do programa R**. Disponível em:
<<http://peld.inpa.gov.br/sites/default/files/ApostilaR.pdf>>.
- [12] **Utilização do R no Ensino da Estatística Básica I**. Disponível em:
<http://www.professores.uff.br/luciane/images/stories/Arquivos/ap_mci2010.pdf>.

Anexo I



Lucídio Portella 01



Lucídio Portella 02



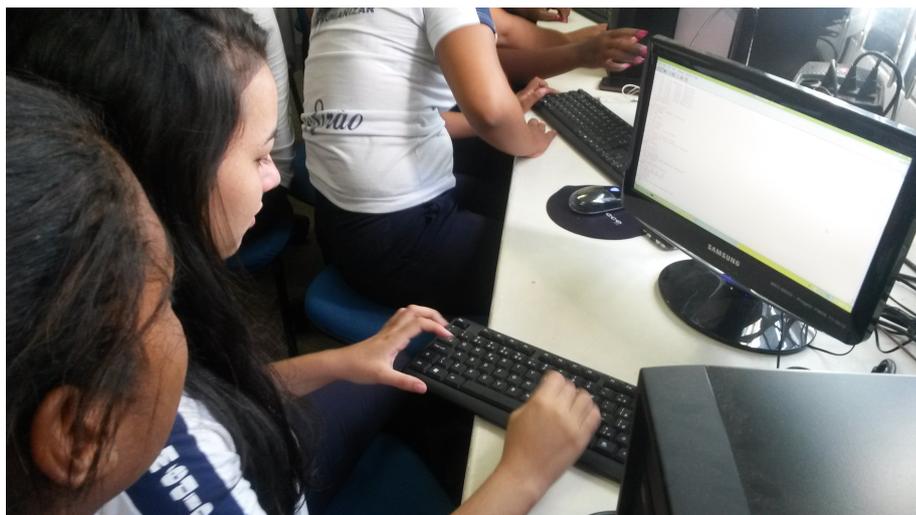
Lucídio Portella 03



Odolfo Medeiros 01



Odolfo Medeiros 02



Odolfo Medeiros 03



Odolfo Medeiros 04



Odolfo Medeiros 05