UNIVERSIDADE FEDERAL DO MARANHÃO

DEPARTAMENTO DE MATEMÁTICA

MESTRADO PROFISSIONAL EM MATEMÁTICA

EM REDE NACIONAL

PROFMAT

IMPLEMENTAÇÃO DA LINGUAGEM R NO ENSINO DE ESTATÍSTICA UTILIZANDO DADOS DE SAÚDE PÚBLICA

MARCOS AURELIO DE SOUZA ALVES MATOS

São Luís - MA

2023

UNIVERSIDADE FEDERAL DO MARANHÃO

DEPARTAMENTO DE MATEMÁTICA

MESTRADO PROFISSIONAL EM MATEMÁTICA

EM REDE NACIONAL

PROFMAT

IMPLEMENTAÇÃO DA LINGUAGEM R NO ENSINO DE ESTATÍSTICA UTILIZANDO DADOS DE SAÚDE PÚBLICA

MARCOS AURELIO DE SOUZA ALVES MATOS

Dissertação apresentada ao Programa de Pós-Graduação em Matemática em Rede Nacional do Departamento de Matemática da Universidade Federal do Maranhão como parte dos requisitos para obtenção do título de Mestre.

Orientador: Prof. Dr. Flausino Lucas Neves Spindola

São Luís - MA

2023

UNIVERSIDADE FEDERAL DO MARANHÃO

DEPARTAMENTO DE MATEMÁTICA

MESTRADO PROFISSIONAL EM MATEMÁTICA

EM REDE NACIONAL

PROFMAT

IMPLEMENTAÇÃO DA LINGUAGEM R NO ENSINO DE ESTATÍSTICA UTILIZANDO DADOS DE SAÚDE PÚBLICA

MARCOS AURELIO DE SOUZA ALVES MATOS

Dissertação apresentada ao Programa de Pós-Graduação em Matemática em Rede Nacional do Departamento de Matemática da Universidade Federal do Maranhão como parte dos requisitos para obtenção do título de Mestre.

Orientador: Prof. Dr. Flausino Lucas Neves Spindola

Aprovado pela Banca Examinadora:

FLAUSINO LUCAS NEVES SPINDOLA:31173439897
Assinado de forma digital por FLAUSINO LUCAS NEVES SPINDOLA:31173439897
Dados: 2023.10.10 21:38:55 -03'00'

Prof. Dr. Flausino Lucas Neves Spindola

gov.br
Documento assinado digitalmente
JOSENILDO DE SOUZA CHAVES
Data: 11/10/2023 14:05:36-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Josenildo de Souza Chaves

gov.br
Documento assinado digitalmente
FABIO NOGUEIRA DA SILVA
Data: 12/10/2023 23:14:43-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Fábio Nogueira da Silva

São Luís - MA

2023

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Matos, Marcos Aurelio de Souza Alves.
Implementação da linguagem R no ensino de Estatística utilizando dados de Saúde Pública / Marcos Aurelio de Souza Alves Matos. - 2023.
45 p.

Orientador(a): Prof. Dr. Flausino Lucas Neves Spindola.
Dissertação (Mestrado) - Programa de Pós-graduação em Rede - Matemática em Rede Nacional/ccet, Universidade Federal do Maranhão, São Luis - MA, 2023.

1. Ensino de Estatística. 2. Estatística Descritiva. 3. Linguagem R. 4. Saúde Pública. I. Spindola, Prof. Dr. Flausino Lucas Neves. II. Título.

AGRADECIMENTOS

Primeiramente a Deus, por me dar mais essa oportunidade e por me conduzir até este momento.

A minha esposa e filha por ter dado todo apoio e me ajudar nos momentos mais difíceis e ter me apoiado para não desistir.

Aos meus pais que sempre me dão todo apoio em tudo que pretendo fazer e sempre me incentivam na busca de sempre encontrar o melhor caminho e não o mais fácil.

A toda minha família em geral que sempre acreditou no meu potencial e sempre me ajuda a conquistar meus objetivos.

Aos meus amigos e colegas de trabalho pela ausência em alguns momentos por não estar com os mesmos em certos momentos.

Aos meus colegas de turma do profmat que me deram apoio nos momentos mais difíceis que enfrentamos e me fizeram persistir quando tudo parecia que não daria certo.

Aos meus professores pelos conselhos e aulas que me ajudaram a chegar até aqui, em especial ao professor orientador que ministra uma área que eu mais me identifico e me ajudou bastante nessa defesa.

“A vida é uma grande universidade, mas pouco ensina a quem não sabe ser um bom aluno…”

Augusto Cury

Resumo

Este trabalho apresenta uma forma prática de abordar os conceitos da Estatística Descritiva no Ensino Médio com utilização da linguagem R, levando o educando ao uso de computadores, tablets, etc, facilitando o processo de ensino aprendizagem. A motivação será por meio de dados retirados da base de dados do Ministério da Saúde, o Vigitel, que trata da vigilância de doenças crônicas não-transmissíveis por meio de inquérito telefônico. Aqui utilizaremos a base de 2015, criando situações-problema em que o uso do R se faz necessário para a determinação das medidas de tendência central, dispersão, gráficos e análise de dados.

Palavras-Chave: Ensino de Estatística, Linguagem R, Saúde Pública, Estatística Descritiva.

Abstract

We present a practical way to approach the concepts in Descriptive Statistics in High School, using programming language R. The purpose is to take easy the teaching-learning process. The motivation will be through data taken from the Ministry of Health's database, Vigitel, which deals with the surveillance of non-transmissible chronic diseases through telephone survey. We will use the 2015 database, creating problem situations in which the use of R is necessary for the determination of measures of central tendency, dispersion, graphs and data analysis.

Keywords: Statistics Teaching, R Software, Public Health, Descriptive Statistics.

SUMÁRIO

Introdução

Em Estatística as medidas de tendência central mais utilizadas são: média, mediana e moda. Outras medidas importantes para a análise de uma variável são as medidas de dispersão: variância, desvio padrão, coeficiente de variação, etc. Esses conceitos são abordados em turmas da educação básica do Ensino Médio, e podem ser abordados de diversas maneiras com o uso de exemplos ou pesquisas de opinião entre os próprios alunos da escola.

O desempenho dos estudantes em disciplinas de matemática no ensino médio é, em média, insatisfatório. Perante esta realidade do ensino de matemática, este trabalho se apresenta com o objetivo de propor uma abordagem computacional para o ensino de medidas de tendência central e medidas de dispersão bem como confecção de gráficos estatísticos, motivado por tratamento de dados de saúde pública.

Utilizaremos a linguagem R que tem boa capacidade na análise de dados estatísticos. Ao tratarmos de uma base de dados de doenças crônicas não transmissíveis, o VIGITEL, motivaremos a discussão sobre tópicos interdisciplinares que podem ser abordados com a disciplina de Biologia, de Educação Física, de Geografia.

O resumo de dados que os alunos poderão desenvolver está relacionado a dados de saúde pública, visto que o tema ainda segue atormentado a todos devido à pandemia recente de COVID-19 que trouxe a todos interesse com relação a temas de saúde e prevenção. A coleta de dados será realizada pelos alunos e vai requerer muito treino e logística, e após a coleta desses dados os mesmos serão organizados em tabelas de frequência construídas pelos educandos. Finalmente os educandos poderão efetuar as construções gráficas para enfim realizarem as análises da pesquisa, sendo que poderão utilizar o tipo de gráfico mais adequado a determinada variável.

Os alunos farão uma apresentação dos resultados obtidos para toda comunidade envolvida, levantando idéias para possíveis soluções aos problemas levantados e demonstrar que a matemática pode sim ser muito significativa.

1 - O que é Estatística?

Estatística é um ramo da matemática que trata de um conjunto de processos que tem por objetivo a coleta de informações sobre determinado assunto, organização dos dados coletados, apresentação e interpretação desses dados, assim como tirar conclusões sobre as características das fontes de onde estes foram retirados, para melhor compreender as situações.

As práticas estatísticas incluem o planejamento e a interpretação de observações. Sabendo que seu objetivo principal é a produção da melhor informação possível a partir dos dados disponíveis, citamos:

> *A Estatística é uma ciência que se dedica ao desenvolvimento e ao uso de métodos para a coleta, resumo, organização, apresentação e análise de dados. (Farias, Soares & César, 2003)*

A palavra *estatística* tem origem na palavra em latim status, traduzida como o estudo do Estado e significava, originalmente, uma coleção de informação de interesse para o estado sobre população e economia. Todo pesquisador precisa dos conceitos de estatística ao realizar uma pesquisa, independentemente do seu objeto de estudo. Ele precisa dominar todos esses conceitos e saber como utilizá-los na busca de soluções para os problemas estabelecidos.

1.1 - Algumas definições de estatística

- A Estatística é uma parte da Matemática Aplicada que fornece métodos para coleta, organização, análise e interpretação de dados e para serem utilizados na tomada de decisões. (CRESPO, 2000).

- Estatística é um conjunto de métodos e processos quantitativos que servem para estudar e medir fenômenos coletivos. (SILVA, 1999).

- De acordo com o site InfoEscola, Estatística é um conjunto de métodos especialmente apropriados à coleta, à apresentação (organização, resumo e descrição), à análise e à interpretação de dados de observação, tendo como objetivo a compreensão de uma realidade específica para a tomada de decisão. Desta forma a Estatística se preocupa com:

  A coleta, a organização, a sintetização e a apresentação de dados;

  A medição da variação nos dados e levantamento de dados;

  Estimativa de parâmetros da população e a determinação da precisão dessas estimativas;

  A aplicação dos testes de hipótese sobre características da população;

A análise da relação entre duas ou mais variáveis.

1.2 - Estatística Descritiva

A Estatística Descritiva é um ramo da estatística que utiliza técnicas para descrever e resumir um conjunto de dados. Para os pesquisadores descreverem os resultados de suas pesquisas de campo, os mesmos criam tabelas e gráficos, colocando seus dados de forma que possam ser interpretados mais facilmente. Conforme Bruni (2008), os gráficos representam um poderoso instrumento para análise e interpretação de um conjunto de dados. Eles podem ser apresentados nos mais diversos veículos de comunicação. O resumo de dados assume um papel fundamental para explicar o comportamento do objeto de estudo.

> *Os mais importantes recursos fornecidos pelos gráficos são a facilidade e a rapidez na absorção e interpretação dos resultados por parte do leitor (Bruni, 2008).*

1.3 - Por que utilizar os conceitos de Estatística no Ensino Médio?

O ensino da matemática, a cada ano, se torna algo cada vez mais desafiador aos professores do ensino básico, além da falta de dedicação e o desinteresse pela maioria dos educandos. Passamos por momentos de medo e incertezas sobre nosso futuro, quando vivenciamos o período marcado pela disseminação da pandemia de COVID-19 no mundo. Todos, nas diversas esferas do ensino, tiveram que se adaptar às novas formas e métodos de aprendizagem e dar o seu melhor para superar tais dificuldades.

Após todo esse processo e o surgimento da vacina, surgiu com ela a esperança de recomeçar, porém as escolas receberam alunos desestimulados, desinteressados, com problemas psicológicos e com outros objetivos que encontraram no mundo virtual e que os distanciaram cada vez mais da escola. Aqui abordaremos uma proposta para tentar aproximar nossos alunos a se interessarem mais na área da pesquisa e perceber a importância da matemática em várias situações do nosso cotidiano. Na Base Nacional Comum Curricular (BRASIL, 2017), encontramos competências específicas e habilidades que abordam temas e tópicos de Estatística. Abaixo listamos as competências e habilidades sobre esse tema:

1.3.1 - Competência específica 1

“Utilizar estratégias, conceitos e procedimentos matemáticos para interpretar situações em diversos contextos, sejam atividades cotidianas, sejam fatos das Ciências da Natureza e Humanas, ou ainda questões econômicas ou tecnológicas, divulgados por diferentes meios, de modo a consolidar uma formação científica geral.”

“O desenvolvimento da competência, que é bastante ampla, pressupõe habilidades que podem favorecer a interpretação e compreensão da realidade pelos estudantes, utilizando conceitos de diferentes campos da Matemática para fazer julgamentos bem fundamentados. Essa competência específica contribui não apenas para a formação de cidadãos críticos e reflexivos, mas também para formação científica geral dos estudantes, uma vez que lhes é proposta a interpretação de situações das Ciências da Natureza ou Humanas. Os estudantes deverão, por exemplo, ser capazes de analisar criticamente o que é produzido e divulgado nos meios de comunicação (livros, jornais, revistas, internet, televisão, rádio etc.), muitas vezes de forma imprópria, dada por generalizações equivocadas de resultados de pesquisa, o que pode ocorrer tanto pelo uso inadequado da amostragem, quanto pela não divulgação de como os dados foram obtidos.”(BRASIL, 2017, pg. 532).

Para tais competências, estão relacionadas as seguintes habilidades:

HABILIDADES

(EM13MAT101) Interpretar criticamente situações econômicas, sociais e fatos relativos às Ciências da Natureza que envolvam a variação de grandezas, pela análise dos gráficos das funções representadas e das taxas de variação, com ou sem apoio de tecnologias digitais.

(EM13MAT102) Analisar tabelas, gráficos e amostras de pesquisas estatísticas apresentadas em relatórios divulgados por diferentes meios de comunicação, identificando, quando for o caso, inadequações que possam induzir a erros de interpretação, como escalas e amostras não apropriadas.

1.3.2 - Competência específica 2

Propor ou participar de ações para investigar desafios do mundo contemporâneo e tomar decisões éticas e socialmente responsáveis, com base na análise de problemas sociais, como os voltados a situações de saúde, sustentabilidade, das implicações da tecnologia no mundo do trabalho, entre outros, mobilizando e articulando conceitos, procedimentos e linguagens próprios da Matemática.

“Essa competência específica amplia a anterior por colocar os estudantes em situações nas quais precisam investigar questões de impacto social que os mobilizem a propor ou participar de ações individuais ou coletivas que visem solucionar eventuais problemas. O desenvolvimento dessa competência específica prevê ainda que os estudantes possam identificar aspectos consensuais ou não na discussão tanto dos problemas investigados como das intervenções propostas, com base em princípios solidários, éticos e sustentáveis, valorizando a diversidade de opiniões de grupos sociais e de indivíduos e sem quaisquer preconceitos. Nesse sentido, favorece a interação entre os estudantes, de forma cooperativa, para aprender e ensinar Matemática de forma significativa. Para o desenvolvimento dessa competência, deve-se também considerar a reflexão sobre os distintos papéis que a educação matemática pode desempenhar em diferentes contextos sociopolíticos e culturais, como em relação aos povos e comunidades tradicionais do Brasil, articulando esses saberes construídos nas práticas sociais e educativas.” (BRASIL, 2017, pg. 534)

HABILIDADE

(EM13MAT202) Planejar e executar pesquisa amostral sobre questões relevantes, usando dados coletados diretamente ou em diferentes fontes, e comunicar os resultados por meio de relatório contendo gráficos e interpretação das medidas de tendência central e das medidas de dispersão (amplitude e desvio padrão), utilizando ou não recursos tecnológicos.

1.3.3 - Competência específica 3

Utilizar estratégias, conceitos, definições e procedimentos matemáticos para interpretar, construir modelos e resolver problemas em diversos contextos, analisando a plausibilidade dos resultados e a adequação das soluções propostas, de modo a construir argumentação consistente.

“As habilidades indicadas para o desenvolvimento dessa competência específica estão relacionadas à interpretação, construção de modelos, resolução e formulação de problemas matemáticos envolvendo noções, conceitos e procedimentos quantitativos, geométricos, estatísticos, probabilísticos, entre outros.

No caso da resolução e formulação de problemas, é importante contemplar contextos diversos (relativos tanto à própria Matemática, incluindo os oriundos do desenvolvimento tecnológico, como às outras áreas do conhecimento). Não é demais destacar que, também no Ensino Médio, os estudantes devem desenvolver e mobilizar habilidades que servirão para resolver problemas ao longo de sua vida – por isso, as situações propostas devem ter significado real para eles. Nesse sentido, os problemas cotidianos têm papel fundamental na escola para o aprendizado e a aplicação de conceitos matemáticos, considerando que o cotidiano não se refere apenas às atividades do dia a dia dos estudantes, mas também às questões da comunidade mais ampla e do mundo do trabalho.

Deve-se ainda ressaltar que os estudantes também precisam construir significados para os problemas próprios da Matemática.

Para resolver problemas, os estudantes podem, no início, identificar os conceitos e procedimentos matemáticos necessários ou os que possam ser utilizados na chamada formulação matemática do problema. Depois disso, eles precisam aplicar esses conceitos, executar procedimentos e, ao final, compatibilizar os resultados com o problema original, comunicando a solução aos colegas por meio de argumentação consistente e linguagem adequada.

No entanto, a resolução de problemas pode exigir processos cognitivos diferentes. Há problemas nos quais os estudantes deverão aplicar de imediato um conceito ou um procedimento, tendo em vista que a tarefa solicitada está explícita. Há outras situações nas quais, embora essa tarefa esteja contida no enunciado, os estudantes deverão fazer algumas

adaptações antes de aplicar o conceito que foi explicitado, exigindo, portanto, maior grau de interpretação.

Há, ainda, problemas cujas tarefas não estão explícitas e para as quais os estudantes deverão mobilizar seus conhecimentos e habilidades a fim de identificar conceitos e conceber um processo de resolução. Em alguns desses problemas, os estudantes precisam identificar ou construir um modelo para que possam gerar respostas adequadas. Esse processo envolve analisar os fundamentos e propriedades de modelos existentes, avaliando seu alcance e validade para o problema em foco. Essa competência 536 BASE NACIONAL COMUM CURRICULAR específica considera esses diferentes tipos de problemas, incluindo a construção e o reconhecimento de modelos que podem ser aplicados.

Convém reiterar a justificativa do uso na BNCC de "Resolver e Elaborar Problemas" em lugar de "Resolver Problemas". Essa opção amplia e aprofunda o significado dado à resolução de problemas: a elaboração pressupõe que os estudantes investiguem outros problemas que envolvem os conceitos tratados; sua finalidade é também promover a reflexão e o questionamento sobre o que ocorreria se algum dado fosse alterado ou se alguma condição fosse acrescentada ou retirada.

Cabe ainda destacar que o uso de tecnologias possibilita aos estudantes alternativas de experiências variadas e facilitadoras de aprendizagens que reforçam a capacidade de raciocinar logicamente, formular e testar conjecturas, avaliar a validade de raciocínios e construir argumentações."(BRASIL, 2017, pg. 535)

HABILIDADE

(EM13MAT316) Resolver e elaborar problemas, em diferentes contextos, que envolvem cálculo e interpretação das medidas de tendência central (média, moda, mediana) e das medidas de dispersão (amplitude, variância e desvio padrão).

1.3.4 - Competência específica 4

Compreender e utilizar, com flexibilidade e precisão, diferentes registros de representação matemáticos (algébrico, geométrico, estatístico, computacional etc.), na busca de solução e comunicação de resultados de problemas.

"As habilidades vinculadas a essa competência específica tratam da utilização das diferentes representações de um mesmo objeto matemático na resolução de problemas em vários contextos, como os socioambientais e da vida cotidiana, tendo em vista que elas têm um papel decisivo na aprendizagem dos estudantes. Ao conseguirem utilizar as representações matemáticas, compreender as ideias que elas expressam e, quando possível, fazer a conversão entre elas, os estudantes passam a dominar um conjunto de ferramentas que potencializa de forma significativa sua capacidade de resolver problemas, comunicar e argumentar; enfim, ampliam sua capacidade de pensar matematicamente. Além disso, a análise das

representações utilizadas pelos estudantes para resolver um problema permite compreender os modos como o interpretaram e como raciocinaram para resolvê-lo.

Portanto, para as aprendizagens dos conceitos e procedimentos matemáticos, é fundamental que os estudantes sejam estimulados a explorar mais de um registro de representação sempre que possível. Eles precisam escolher as representações mais convenientes a cada situação, convertendo-as sempre que necessário. A conversão de um registro para outro nem sempre é simples, apesar de, muitas vezes, ser necessária para uma adequada compreensão do objeto matemático em questão, pois uma representação pode facilitar a compreensão de um aspecto que outra não favorece."(BRASIL, 2017,pg. 538)

HABILIDADES

(EM13MAT406) Utilizar os conceitos básicos de uma linguagem de programação na implementação de algoritmos escritos em linguagem corrente e/ou matemática.

(EM13MAT408) Construir e interpretar tabelas e gráficos de frequências com base em dados obtidos em pesquisas por amostras estatísticas, incluindo ou não o uso de softwares que inter-relacionem estatística, geometria e álgebra.

(EM13MAT409) Interpretar e comparar conjuntos de dados estatísticos por meio de diferentes diagramas e gráficos (histograma, de caixa (box-plot), de ramos e folhas, entre outros), reconhecendo os mais eficientes para sua análise.

Ao observar essas competências e habilidades, vemos que a Estatística é um conteúdo altamente coerente por está relacionada com vários temas e assuntos em muitas áreas do conhecimento. No cotidiano é fácil observar e utilizá-la, isso torna os conteúdos de Estatística um dos mais aptos a serem abordados no ensino médio.

2 - Fundamentação teórica

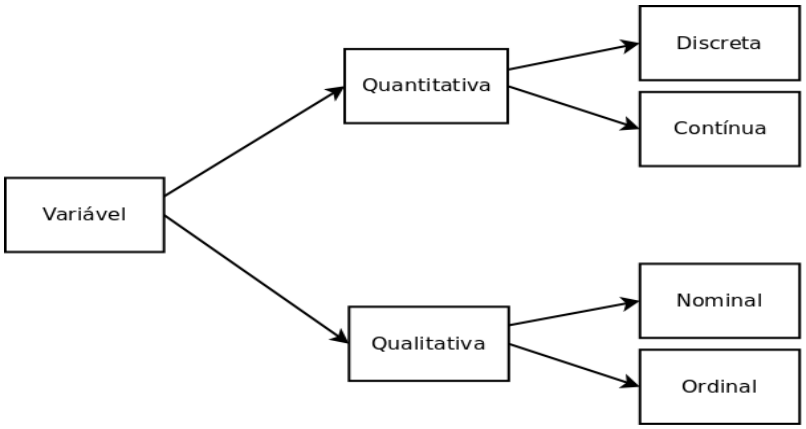
2.1 - Variáveis

Um dos primeiros conceitos a ser trabalhado em uma sala de aula é o de variável. Esta pode ser a idade dos alunos de uma sala de aula, o peso dos alunos, o gênero, ou então nacionalidade de pessoas em um aeroporto ou o grau de instrução de candidatos a uma vaga de trabalho, etc.
Podemos classificar as variáveis em qualitativas e quantitativas.

- As variáveis **qualitativas** - são aquelas que podem ser classificadas em categorias que se diferenciam por características não numéricas. Por exemplo: cor dos olhos (preto, azul, castanho claro ou escuro), sexo (masculino e feminino). As variáveis qualitativas são subdivididas em ordinais e nominais.
  - Variáveis qualitativas ordinais - são aquelas em que seus valores seguem uma certa ordem. Por exemplo: desempenho de um aluno em determinada atividade regular, bom, ótimo.
  - Variáveis qualitativas nominais - são aquelas em que seus valores não podem ser estabelecidos em uma certa ordem, como masculino e feminino, nacionalidade, time de futebol para qual o sujeito torce.
- As variáveis quantitativas - são aquelas em que seus valores são expressos por números e se subdivide em:
  - Variáveis quantitativas discretas - são aquelas que resultam de contagens. Exemplos: número de alunos acima da média, número de faltas, idade.
  - Variáveis quantitativas contínuas - são aquelas que em determinado intervalo, podem assumir qualquer valor no conjunto dos números reais. Exemplos: altura e massa corporal dos alunos de uma escola, taxa de natalidade em um município.

Figura 2.1 destaca a subdivisão das variáveis.

Figura 2.1: Diagrama sobre a classificação de variáveis.



Fonte: https://www.inf.ufsc.br/~andre.zibetti/probabilidade/aed.html

Definição 2.1: População o conjunto de todos os resultados possíveis de um experimento que temos interesse em estudar. Pode ser finito, infinito enumerável ou infinito não-enumerável.

Definição 2.2: Amostra um subconjunto da população.

Em estatística, dificilmente se consegue obter informações de todos os elementos da população. Citamos o caso em que se queira saber, dos brasileiros que viajaram ao Catar para assistir a copa do mundo em novembro de 2022, ou a opinião sobre a segurança (ruim, boa ou excelente) durante os jogos nos estádios. É quase impossível entrevistar todos os brasileiros lá presentes. Portanto, deve-se fazer uma pesquisa por amostragem, ou seja, selecionar uma certa quantidade de brasileiros (amostra) e a partir desses dados específicos generalizar os resultados. Cada elemento que compõe a amostra é chamado de unidade amostral. No exemplo descrito, os indivíduos são os brasileiros entrevistados.

A primeira constatação é que temos n = 20. Podemos colocar os dados em ordem crescente, assim:
65,3 - 68,5 - 68,9 - 69,2 - 70,6 - 70,7 - 72,1 - 72,2 - 72,3 - 73,1 - 75,2 - 75,5 - 75,8 - 78,4 - 83,2 - 86,3 - 86,3 - 87,8 - 88,1 - 90,3.

Outra estatística importante se chama amplitude amostral, definida pela diferença entre o valor máximo e o valor mínimo:

$$Amp = x_{max} - x_{min}$$
$$Amp = 90,3 - 65,3 = 25\,kg$$

Portanto a amplitude amostral é 25 Kg.

2.2 - Construção de Tabela de Frequência

Considere a pesquisa sobre a segurança (ruim, boa ou excelente) e a massa média dos brasileiros presentes na Copa do mundo da FIFA do ano de 2022. Chamamos a qualidade da segurança e a massa de variáveis aleatórias, ou seja, são os atributos pesquisados de cada indivíduo. No exemplo, a qualidade da segurança nos estádios (ruim, boa ou excelente) é uma variável qualitativa nominal e a massa dos brasileiros no Catar é uma variável quantitativa contínua.

Cada resultado diferente de uma amostra é denominado classe e a quantidade de vezes que esta classe se repete chamamos de 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 voltemos ao exemplo citado anteriormente sobre a pesquisa realizada no

Catar e vamos supor que em um grupo de 200 brasileiros entrevistados, obtivemos os seguintes dados obtidos e organizados na tabela de frequência abaixo.

Tabela 2.1: Tabela de Frequência da opinião sobre a segurança dos estádios

Segurança	Frequência Absoluta	Frequência Relativa	Percentual
Ruim	20	$\frac{20}{200} = 0,10$	10%
Boa	60	$\frac{60}{200} = 0,30$	30%
Excelente	120	$\frac{120}{200} = 0,60$	60%
Total	200	1,00	100%

A tabela de frequência da variável massa, 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 sobre a massa dos brasileiros no Catar, vamos supor que a amostra foi dividida em intervalos de 5 kg.

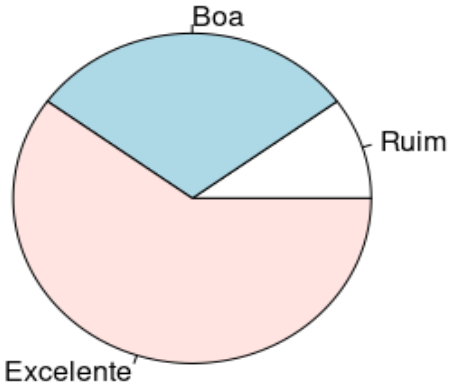
Tabela 2.2: Tabela de Frequência da variável massa.

Massa	Frequência Absoluta	Frequência Relativa	Percentual
65 |-------> 70	4	$\frac{4}{20} = 0,20$	20%
70 |-------> 75	6	$\frac{6}{20} = 0,30$	30%
75 |-------> 80	4	$\frac{4}{20} = 0,20$	20%
80 |--------> 85	1	$\frac{1}{20} = 0,05$	5%
85 |--------> 90	4	$\frac{4}{20} = 0,20$	20%
90 |--------> 95	1	$\frac{1}{20} = 0,05$	5%
Total	20	1,00	100%

2.3 - Representações gráficas

De acordo com as pesquisas acima, podemos representar graficamente os dados obtidos em tabelas de frequência. Usaremos o gráfico de pizza para representar a variável segurança nos estádios na Copa do Mundo no Catar do ano de 2022, pois é comumente utilizado para representar parte de um todo, geralmente em porcentagem, e é bastante apropriado para mostrar frequências de ocorrências de variáveis qualitativas.

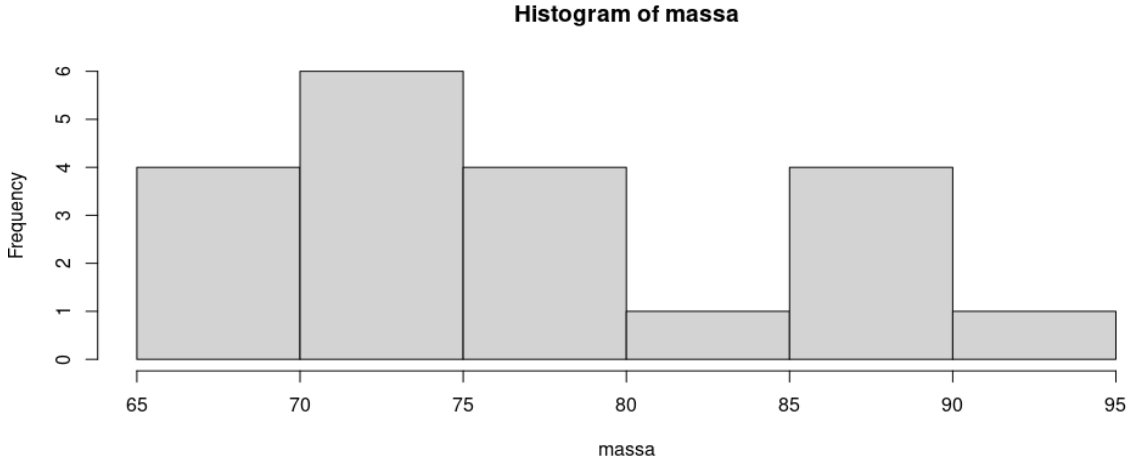
Figura 2.2 - Gráfico em setores da variável segurança nos estádios da Copa do Mundo da FIFA no Catar, 2022.



Fonte: Acervo do autor

O histograma é um gráfico representado em coluna ou barras e é dividido em classes. Essas colunas ou barras são retangulares e suas bases representam uma classe. A altura ou comprimento desses retângulos representam a frequência absoluta com que o valor de cada classe ocorre. Neste caso estamos usando essa representação gráfica por ser adequado para a variável quantitativa.

Figura 2.3 - Histograma da variável massa



Fonte: Acervo do autor

2.4 - Medidas de tendência central

A medida mais comum de tendência central é a **média aritmética**, ou simplesmente média. Segundo Pollatsek, Lima e Well (1981), a média aritmética não é só o conceito mais básico da Estatística e da ciência experimental, é também o mais utilizado na vida cotidiana das pessoas. Em geral, ao fazermos inferências tanto no campo acadêmico como na vida cotidiana, utilizamos a média ou a comparação entre médias.

A média fornece um indicador que pode ser interpretado como um valor típico e que pode representar, em certas circunstâncias, um conjunto de dados. Além disso, é a base para o cálculo de outras medidas tais como variância, desvio padrão, entre outras.

Para dados não agrupados, a média é calculada como o quociente entre a soma de todos os valores observados da variável e o tamanho da amostra. Para dados agrupados, os valores observados da variável devem ser ponderados pelos seus respectivos pesos ou frequências - nesse caso, a média é chamada de ponderada.

Outro conceito para a tendência central é a **mediana**. Dada a amostra $(x_1, x_2, \ldots, x_n)$ de uma variável X, a mediana é o valor central dessa amostra ordenada. Mediana (Md) é calculada da seguinte forma: se n é o tamanho da amostra, caso n seja par,

$$Md = \frac{x\left(\frac{n}{2}\right)+x\left(\frac{n+2}{2}\right)}{2}$$

Caso n seja ímpar,

$$Md = x\left(\frac{n}{2}\right).$$

onde $x(i)$ denota x na posição i depois que os dados quantitativos foram colocados em ordem crescente.

A **moda** é outra medida de posição e ela é a observação que ocorre com maior frequência. Às vezes, num conjunto de dados, podemos ou não, ter moda. Havendo mais de uma moda ela será multimodal - bimodal para duas modas, trimodal para três modas.

A **variância**, no caso populacional, pode ser calculada pela expressão:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n},$$

onde:

$\sum$: o símbolo de somatório, indica que temos que somar todos os termos, desde a primeira posição (i=1) até a posição n;

x_i: valor na posição i no conjunto de dados

$\bar{x}$: média aritmética dos dados

n: quantidade de dados

No caso amostral, divide-se por $n - 1$.

Exemplo: No aniversário de Marcelo, os 10 primeiros convidados a chegar tinham, respectivamente, 4, 5, 7, 8, 10, 11, 12, 15, 19 e 20 anos. Vamos determinar a variância das idades desses 10 convidados. veja:

1º) Determinamos a média ($\bar{x}$).

$$\bar{x} = \frac{4+5+7+8+10+11+12+15+19+20}{10} = 11{,}1 \text{ anos}$$

2º) Fazemos o somatório do quadrado das diferenças $(x_i - \bar{x})^2$.

$$(4 - 11,1)^2 + (5 - 11,1)^2 + \ldots + (19 - 11,1)^2 + (20 - 11,1)^2 = 272,9$$

3º) Dividimos o valor desse somatório por (n).

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{272{,}9}{10} = 27{,}29 \text{ anos}$$

O **desvio padrão**, no caso populacional, representado por σ.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

Analogamente, no caso amostral, divide-se por $n - 1$.

Quanto menor for o desvio padrão, mais concentrados os dados estão em torno da média. E quanto maior, indica que os dados estão mais dispersos.

No exemplo citado anteriormente, temos um desvio dado por:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} = \sqrt{\frac{272{,}9}{10}} = \sqrt{27,29} = 5{,}2240 \text{ anos}$$

Se quisermos saber se a dispersão é muito grande em relação à média, podemos calcular o coeficiente de variação (CV):

$$CV = \frac{desvio\ padrão}{média} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}}{\bar{x}} \cdot 100\%$$

No exemplo acima, temos que o coeficiente de variação será:

$$CV = \frac{5{,}2240}{11{,}1} \cdot 100\% \simeq 0,4706 \cdot 100\% \simeq 47,1\%$$

Caso o valor do coeficiente de variação seja inferior a 15% os dados são homogêneos e a média terá uma grande representatividade, porém se for superior a 30%, a distribuição é esparsa ou heterogênea e indica alto grau de dispersão.

3 - A linguagem R

3.1 - Sobre o R

“ (R core Team, 2023) é uma linguagem de programação multi-paradigma orientada a objetos, programação funcional, dinâmica, fracamente tipada, voltada à manipulação, análise e visualização de dados. Foi criado originalmente por Ross Ihaka e por Robert Gentleman no departamento de Estatística da Universidade de Auckland, Nova Zelândia. A linguagem R é largamente usada entre estatísticos e analistas de dados. Pesquisas e levantamentos com profissionais da área mostram que a popularidade do R aumentou substancialmente nos últimos anos.

A linguagem R (R core Team, 2023) disponibiliza uma ampla variedade de técnicas estatísticas e gráficas, incluindo modelação linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, agrupamento e outras. A linguagem R é facilmente extensível através de funções e extensões, e a comunidade R é reconhecida pelas suas contribuições ativas em termos de pacotes.

A linguagem R é fortemente extensível através do uso de pacotes enviados pelo utilizador para funções específicas ou áreas específicas de estudo. Devido à sua herança do S, o R possui fortes recursos de programação orientada por objetos, mais que a maioria das linguagens de computação estatística. Ampliar o R também é facilitado pelas suas regras de contexto lexical.

Outra força do R são os gráficos estáticos, que podem produzir imagens com qualidade para publicação, incluindo símbolos matemáticos. Gráficos dinâmicos e interativos estão disponíveis através de pacotes adicionais.

O R tem a sua própria documentação em formato LaTeX, a qual é usada para fornecer documentação de fácil compreensão, simultaneamente on-line em diversos formatos e em papel.”

https://pt.wikipedia.org/wiki/R_(linguagem_de_programa%C3%A7%C3%A3o)

O R é uma ferramenta muito útil e pode ser um fator a mais no auxílio do ensino aprendizado dos alunos, além de ser uma ferramenta tecnológica e prática que com certeza chamará a atenção dos alunos podendo assim aprimorar os conceitos vistos em sala de aula. Bonilla (1995) nos diz:

> “... para que um software promova realmente a aprendizagem deve estar integrado ao currículo e às atividades de sala de aula, estar relacionado àquilo que o aluno já sabe e ser bem explorado pelo professor. O computador não atua diretamente sobre os processos de aprendizagem, mas apenas fornece ao aluno um ambiente simbólico onde este pode raciocinar, elaborar conceitos e estruturas mentais, derivando novas descobertas daquilo que já sabia” (BONILLA, p. 68).

Portanto, devemos utilizar o R numa forma de tentar atrair nossos alunos a uma nova realidade de estudo e podendo assim despertá-los a buscar novas formas de pesquisas e estudos.

3.2 - Aplicando o R

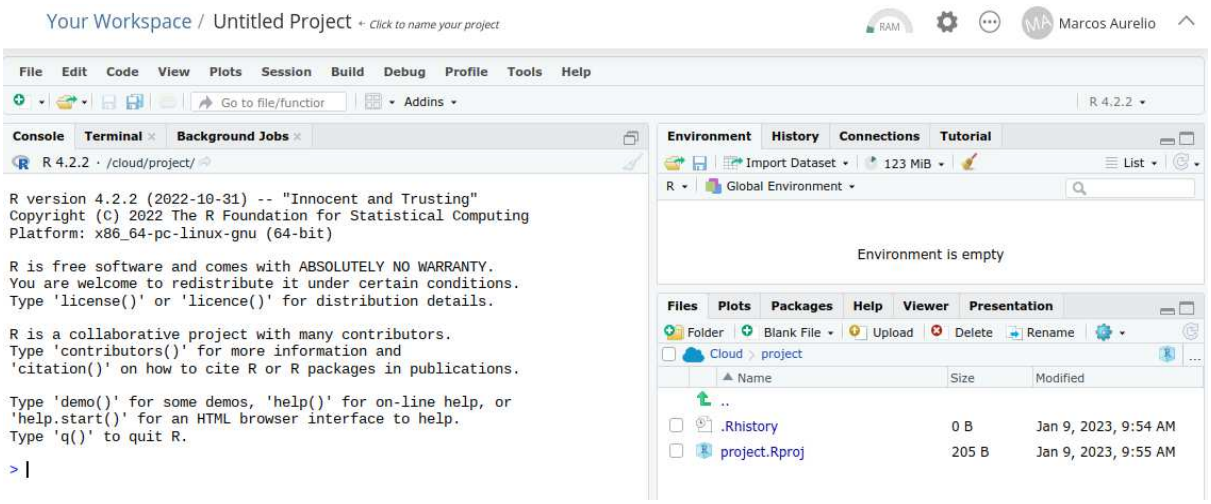
Inicialmente, ao fazer o download do R nos computadores, tablets ou navegando online em uma conta do google qualquer pessoa poderá criar sua conta no R. Quando abrimos o aplicativo teremos a seguinte tela inicial:

Figura 3.1: Tela inicial no R Posit Cloud

Fonte: Acervo do autor

Ao clicar em New Project, aparecerá na tela:

Figura 3.2: Tela inicial do seu projeto



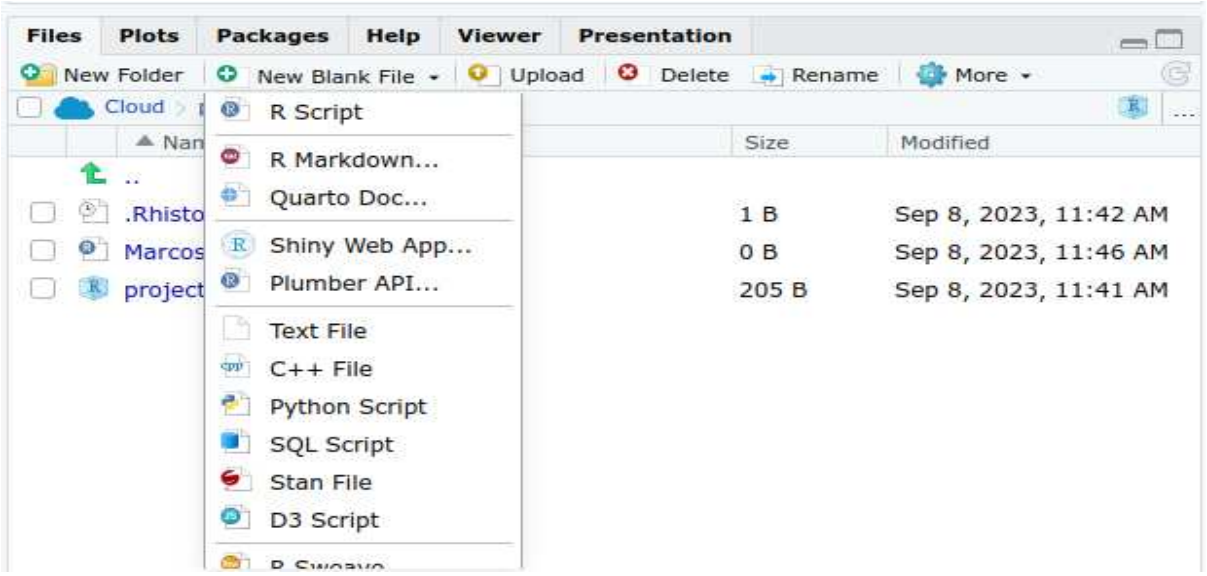
Fonte: Acervo do autor

Para limpar a tela basta pressionar CTRL + L e assim iniciar as atividades.

3.2.1 - Criando um script

Em um script você ordena a execução de uma sequência de comandos, escritos previamente, um seguido do outro. Esses scripts são escritos no editor de códigos do RStudio. Para criá-lo vamos em **New Blank File** e clicamos na opção **R Script.**

Figura 3.3: Criando um script



Fonte: Acervo do autor

3.3 - Operações básicas

Para a familiarização dos alunos com o R, sugerimos iniciar com operações básicas como soma (+), subtração (-), multiplicação (*), divisão (/) e potência (^).

Figura 3.4: Operações básicas no R

Fonte: Acervo do autor

Observação: O símbolo # refere-se a um comentário.

Através de comandos simples, podemos calcular expressões e encontrar informações diretas de uma divisão como o quociente e o resto.

- 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).
- Para calcular a raiz de outra ordem podemos utilizar a potenciação com expoente fracionário. A raiz quarta de 16, por exemplo, calcula-se com 16^(¼). Devemos colocar o parênteses no expoente fracionário pois caso contrário o comando calcularia primeiro a potência 16^1 e em seguida faria a divisão por 4.

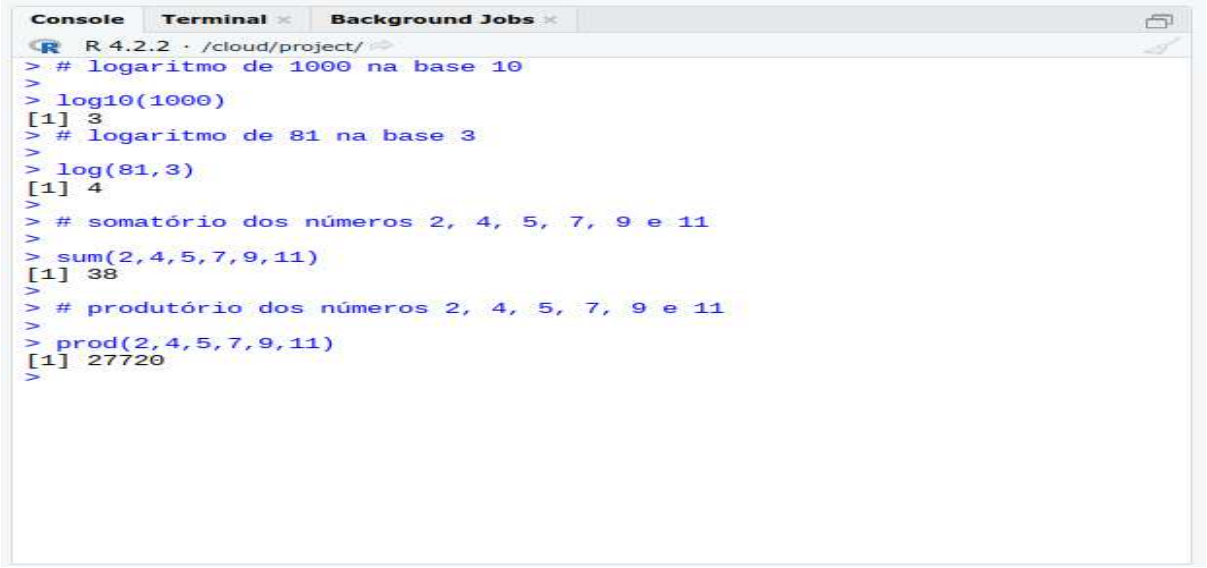
Figura 3.5: Calculando quociente e resto de uma divisão e radiciação

Fonte: Acervo do autor

Podemos calcular algumas funções já conhecidas dos alunos como logaritmo, somatório e produtório.

- Logaritmo na base 10 : log10( )
- Logaritmo de a na base b : log(a,b)
- Somatório : sum( )
- Produtório : prod( )

Figura 3.6: Algumas funções e operações matemáticas



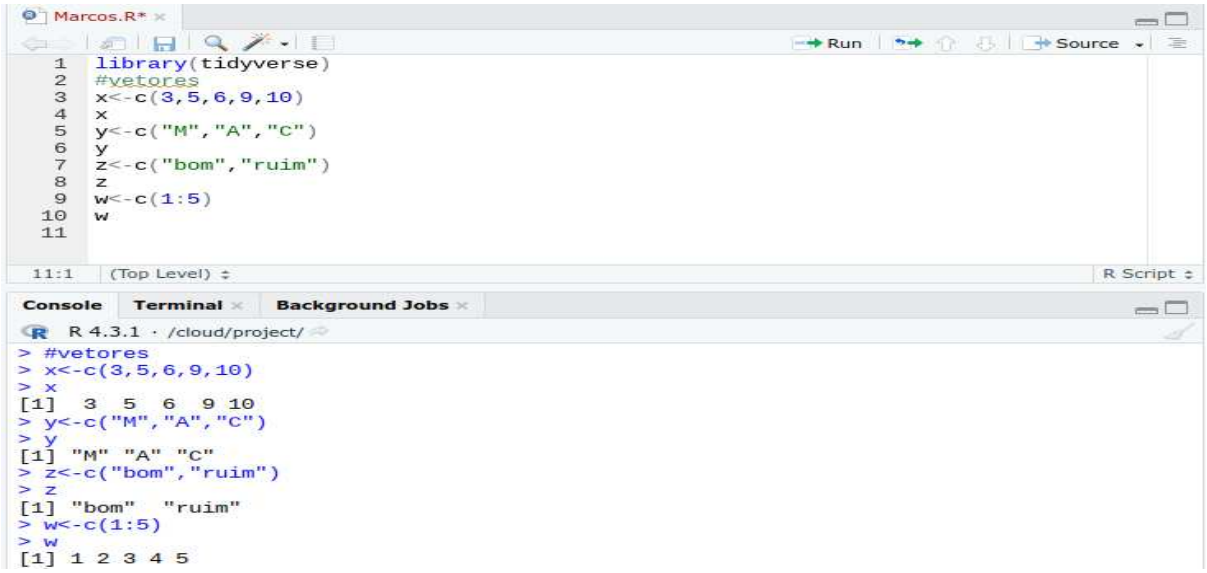
Fonte: Acervo do autor

3.4 - 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 colocá-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(3, 5, 6, 9, 10) : vetor de nome x com os valores numéricos 3, 5, 6, 9 e 10.
- y <- c("M","A","C") : vetor de nome y com os caracteres M, A e C.
- z <- c("bom","ruim") : vetor de nome z com os nomes bom e ruim.
- w <- c(1:5) : vetor de nome w com os números de 1 a 5.

Figura 3.7: Criando vetores



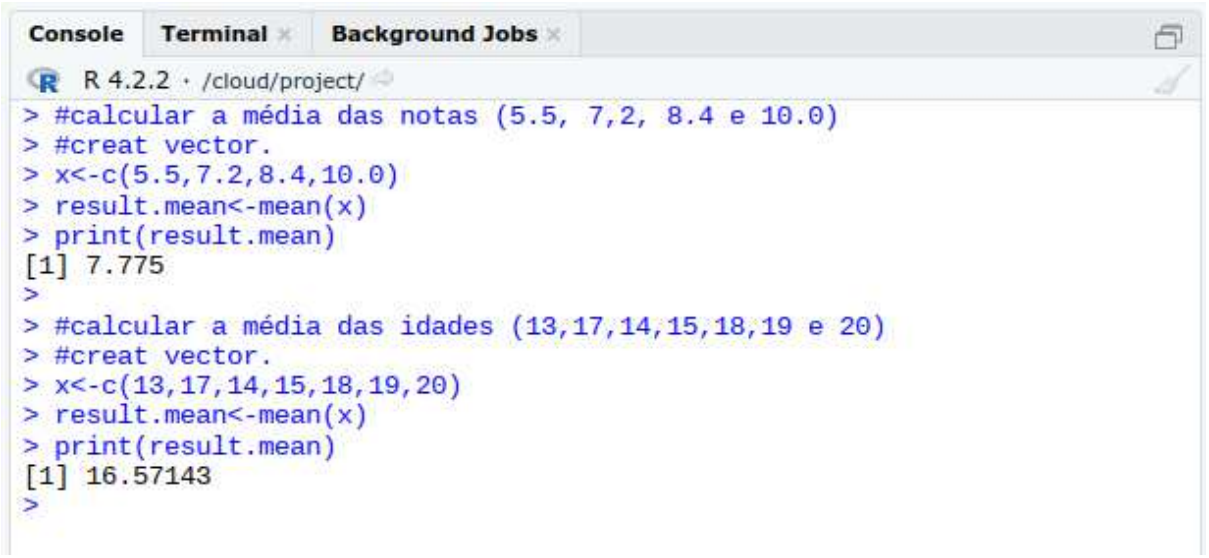
Fonte: Acervo do autor

3.5 - Calculando a média, a mediana e a moda

A partir dos comandos anteriores vamos aplicar o cálculo da média aritmética, da mediana e da moda de um conjunto de dados. A sintaxe básica para calcular a média em R é: mean(x, trim = 0, na.rm = FALSE, …)

Exemplo: Calcular a média aritmética das notas (5.5, 7.2, 8.4 e 10.0) e das idades (13, 17, 14, 15, 18, 19 e 20).

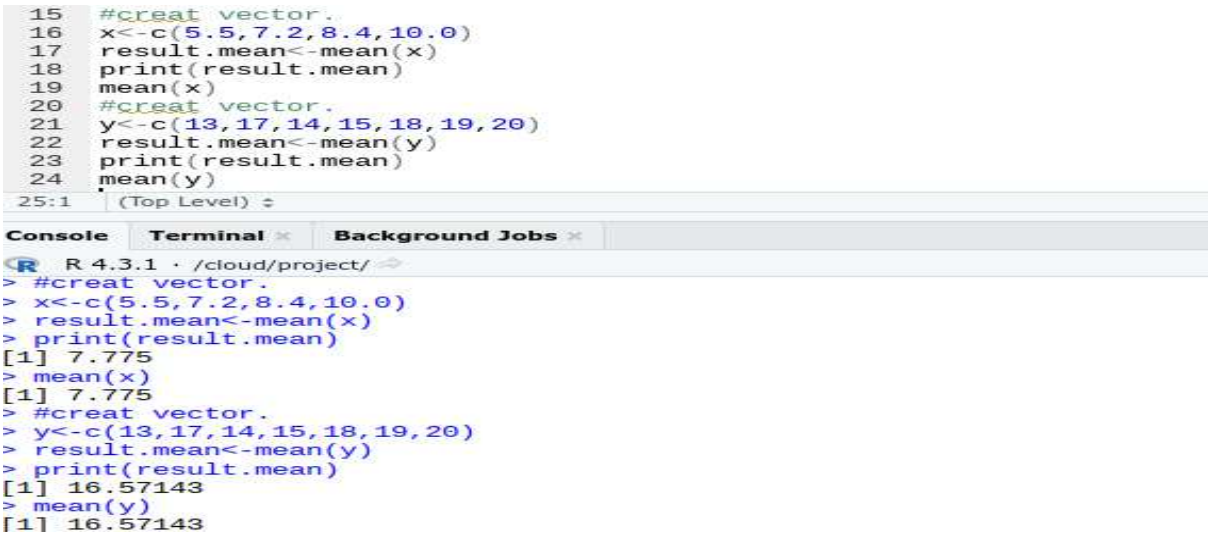
Figura 3.8: Cálculo da média aritmética no R



Fonte: Acervo do autor

Observe na figura 3.9 que a média também pode ser obtida só digitando mean(x) ou mean(y), visto que você já criou o vetor dos dados.

Figura 3.9: Cálculo da média aritmética no R

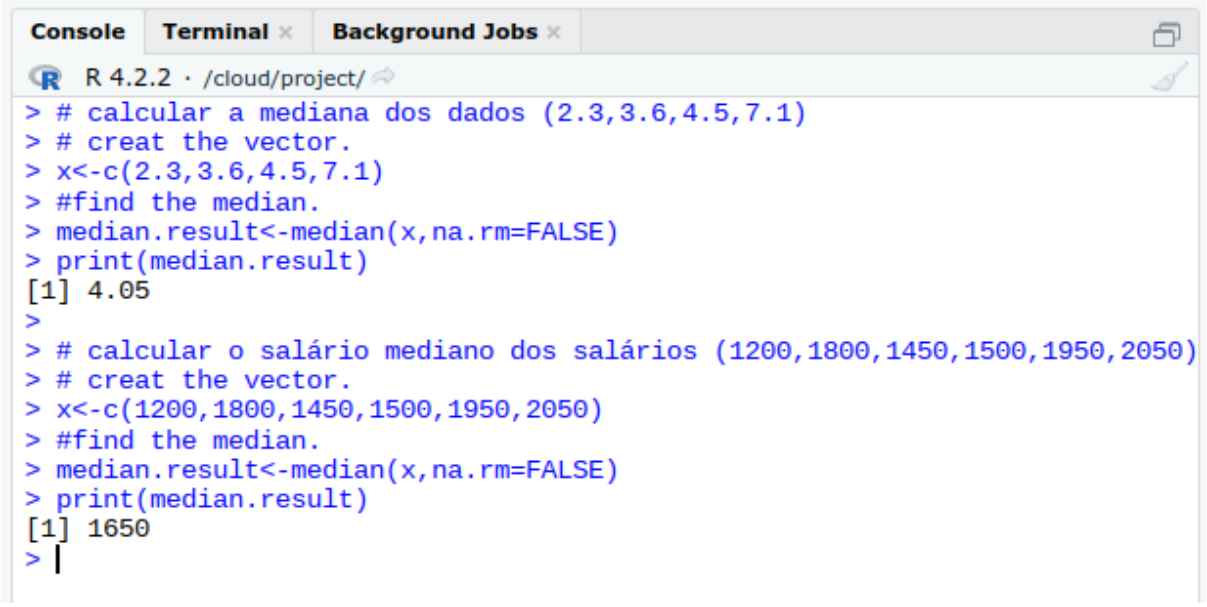


Fonte: Acervo do autor

A sintaxe básica para calcular a mediana em R é: median (x, na.rm = FALSE)

Exemplo: Calcular a mediana do conjunto de dados (2.3, 3.6, 4.5 e 7.1) e o salário mediano entre os dados (1200, 1800, 1450, 1500, 1950 e 2050).

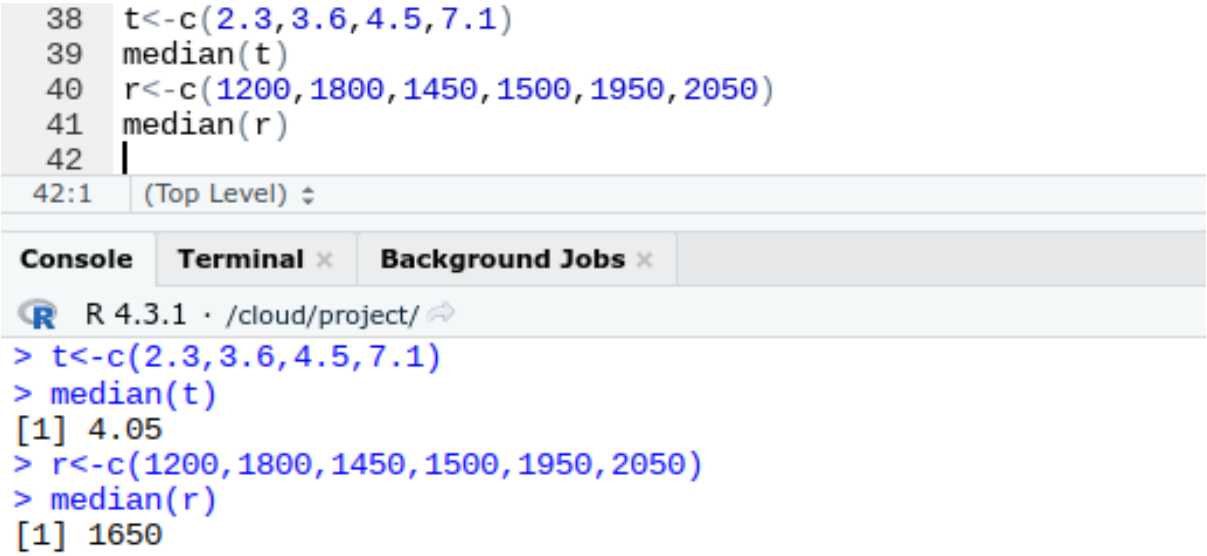
Figura 3.10: Cálculo da mediana no R



Fonte: Acervo do autor

Observe na figura 3.11 que a mediana também pode ser obtida só digitando median(t) ou median(r), visto que você já criou o vetor desses dados.

Figura 3.11: Cálculo da mediana no R

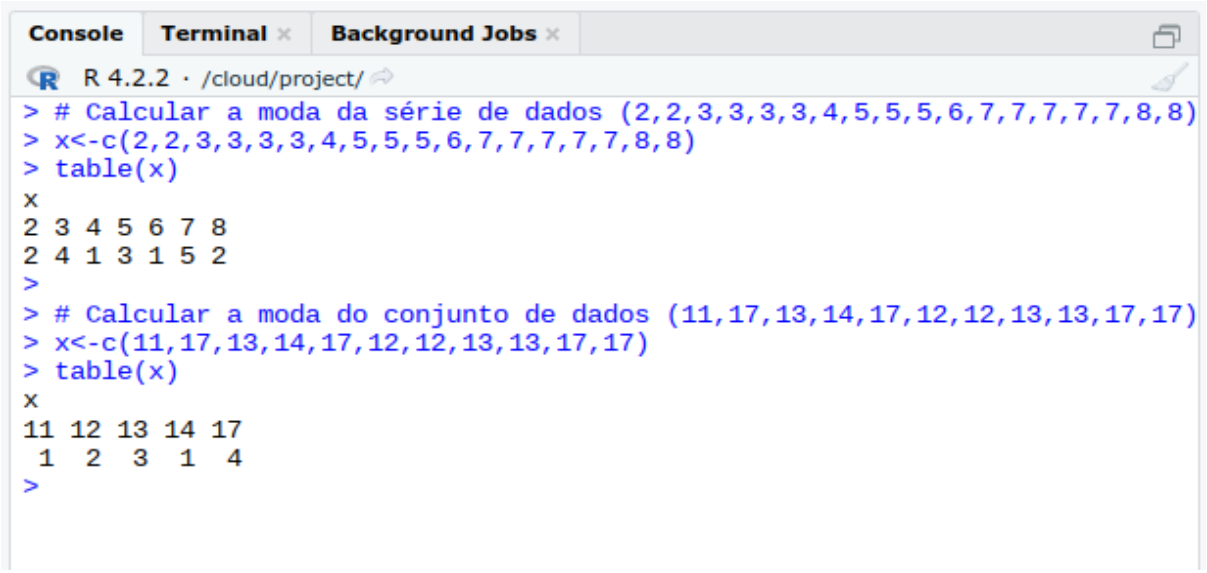


Fonte: Acervo do autor

No R existem duas formas para calcular a moda. Se a série for pequena, facilitando a identificação visual da moda, usa-se a sintaxe **table (series name)** que reportará os elementos da série e, abaixo deles, mostrará com que frequência cada um deles ocorre. Mas para casos em que a série é muito grande, o que tornará difícil a visualização, a sintaxe **subset (table (series names), table (series name) == max (table (series name)))** é utilizada e reportará o elemento modal e abaixo dele a frequência com que ocorre.

Exemplo: Calcular a moda da série de dados (2,2,3,3,3,3,4,5,5,5,6,7,7,7,7,7,8,8) e a moda do conjunto de dados (11,17,13,14,17,12,12,13,13,17,17).

Figura 3.12: Cálculo da moda no R

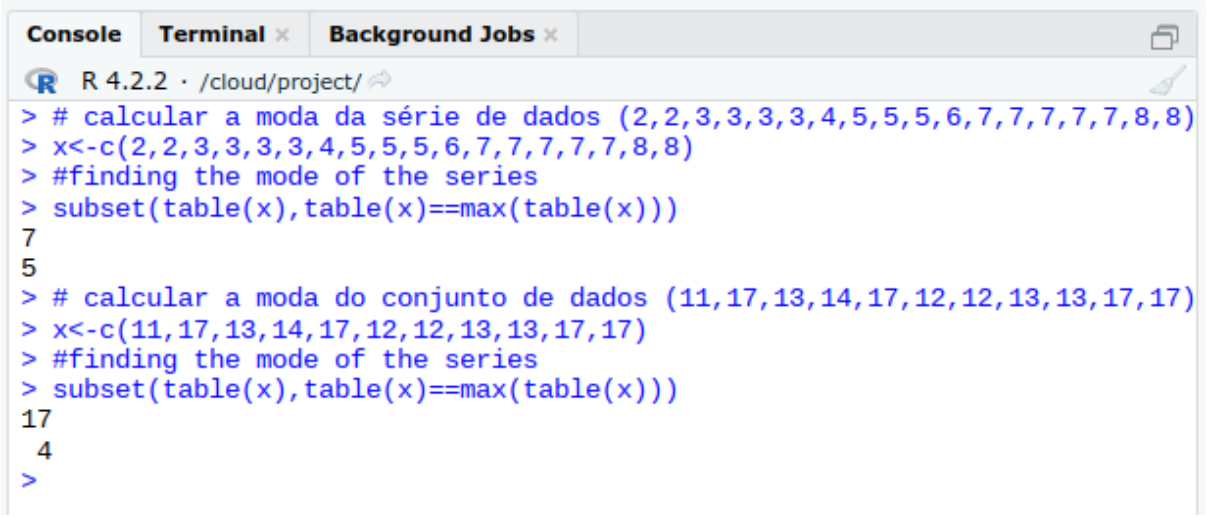


Fonte: Acervo do autor

Pela figura, observe que essa função nos proporciona a frequência de cada termo da série, e vemos notoriamente que a maior frequência determina a moda em cada exemplo em que a série de dados citada. A frequência 5 refere-se a moda 7, e no conjunto de dados a frequência 4 refere-se à moda 17. Porém, podemos usar uma função que reportará o valor da moda e não a frequência para cada um dos elementos da série como visto na figura 3.13.

Exemplo: Calcular a moda da série de dados (2,2,3,3,3,3,4,5,5,5,6,7,7,7,7,7,8,8) e a moda do conjunto de dados (11,17,13,14,17,12,12,13,13,17,17).

Figura 3.13: Cálculo da moda no R



Fonte: Acervo do autor

3.6 - Análise Estatística de dados de Saúde Pública

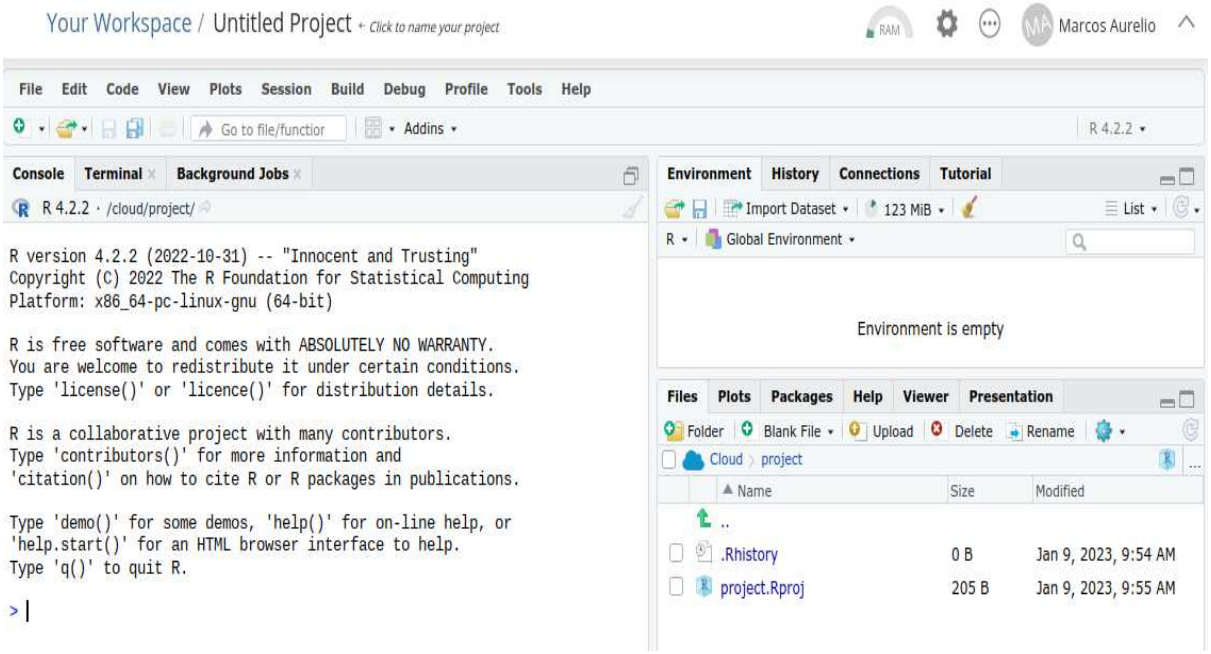
Vamos agora analisar um arquivo retirado da base de dados do Ministério da Saúde que trata da vigilância de doenças crônicas não transmissíveis no ano de 2015 por inquérito telefônico. Este arquivo está no formato .cvs, e possui 54174 observações (ou seja, coleta de dados de 54174 pessoas) e 12 variáveis, dentre elas: cidade, região, estado civil, raça, sexo, peso, altura, e respostas do tipo sim ou não como: se bebe?, se fuma?, é hipertensa? ou se tem diabetes. Vemos que há variáveis qualitativas e quantitativas.

Pelo número de observações, vemos que seria humanamente impossível realizar os cálculos de medidas estatísticas a mão, por isso a relevância do uso de softwares computacionais já no ensino básico, a fim de preparar o estudante para os desafios dos novos tempos no que tange a ciência dos dados.

Com o R, poderemos importar a tabela da base de dados do Ministério da Saúde e, a partir dela, efetuar o cálculo das medidas estatísticas, bem como esboçar os gráficos e diagramas que permitem interpretação e entendimento da disposição dos dados.

As imagens expostas neste texto são de trabalho realizado na Posit Cloud, ambientação de R em nuvem que pode ser realizado em laboratório escolar sem a necessidade de instalação do software R e RStudio, apenas com o uso do navegador de internet.

Figura 3.14: Tela inicial do R na Posit Cloud

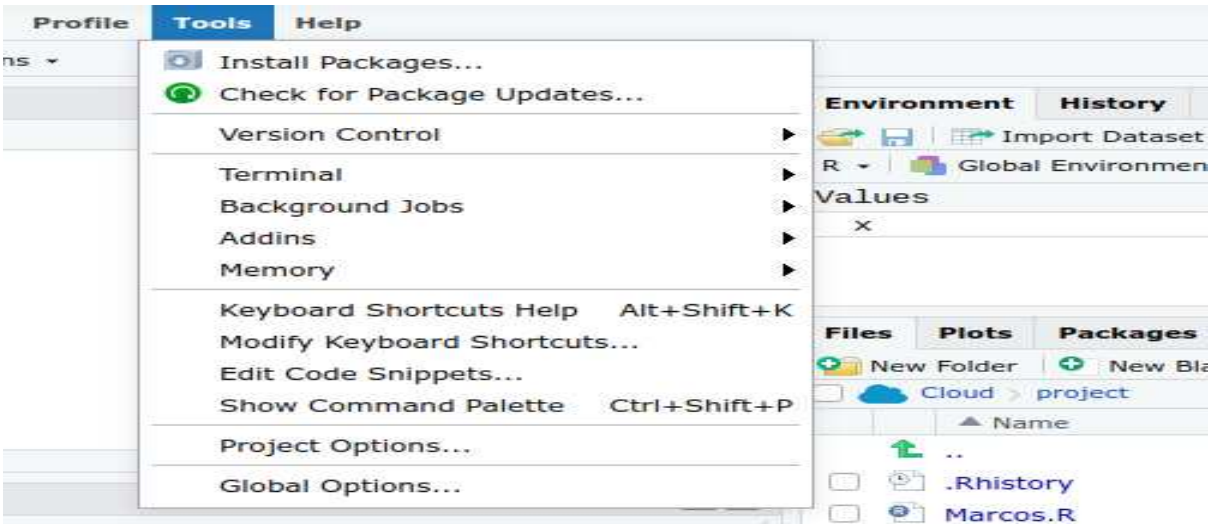


Fonte: Acervo do autor

3.6.1 - Instalando pacotes

Basicamente um pacote do R é uma convenção para organizar e padronizar a distribuição de funções extras do R. E precisamos instalar esses pacotes para podermos executar os comandos. Inicialmente clicamos no ícone **Tools** e em seguida clicamos na opção Install Packages…, e aguardamos a instalação dos mesmos seguindo os passos indicados nas figuras seguintes.

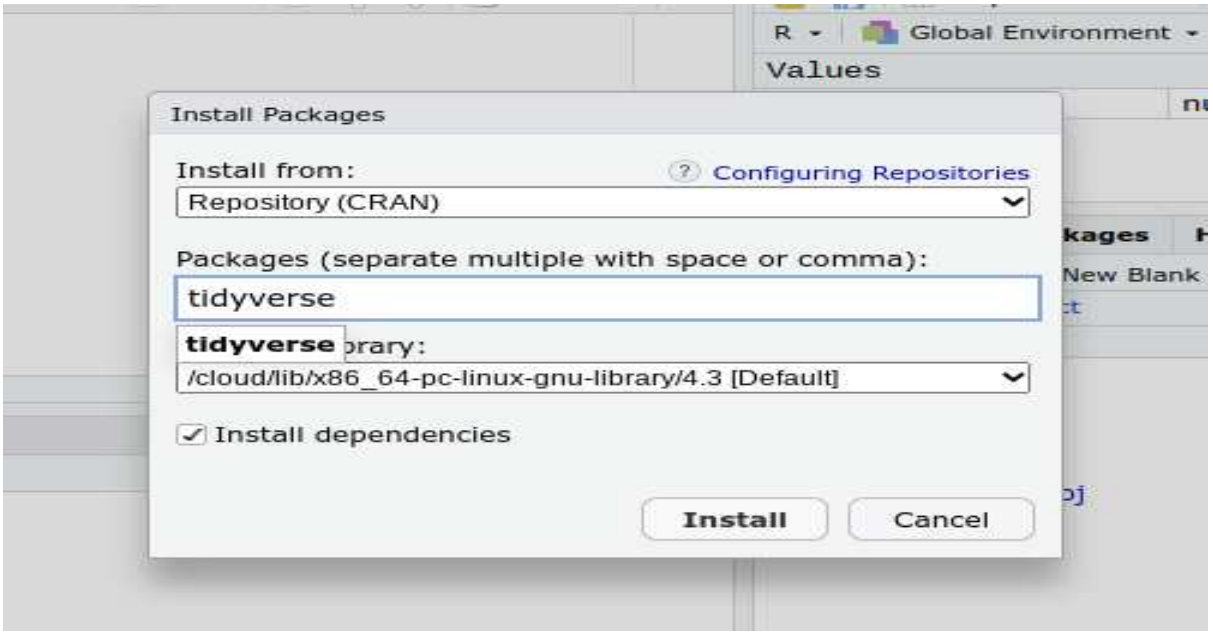
Figura 3.15: Instalando pacotes



Fonte: Acervo do autor

Após clicarmos em Install Packages… aparecerá a imagem da figura 3.16.

Figura 3.16: Instalando pacotes



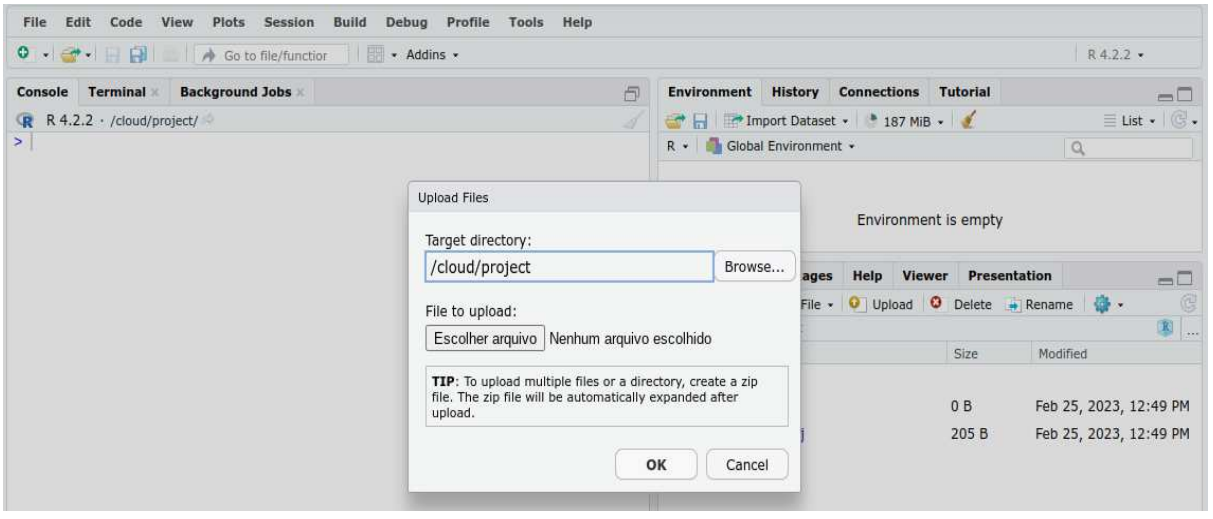
Fonte: Acervo do autor

Dois pacotes que utilizaremos no tratamento de dados que faremos serão o “tidyverse” e o “foreign”.

3.6.2 - Importando os Dados

O próximo passo é localizar o arquivo no formato .csv já disponível no seu computador, indo em **upload**.

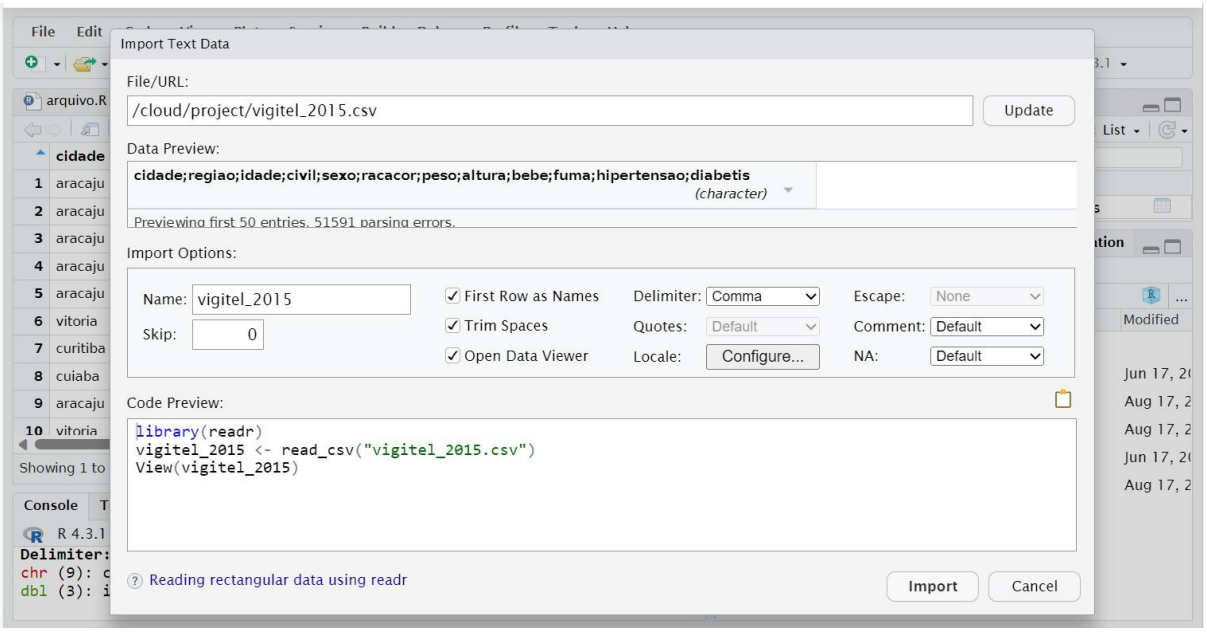
Figura 3.17: Baixando o arquivo para o R



Fonte: Acervo do autor

Em seguida, iremos escolher o arquivo e baixá-lo para o R.

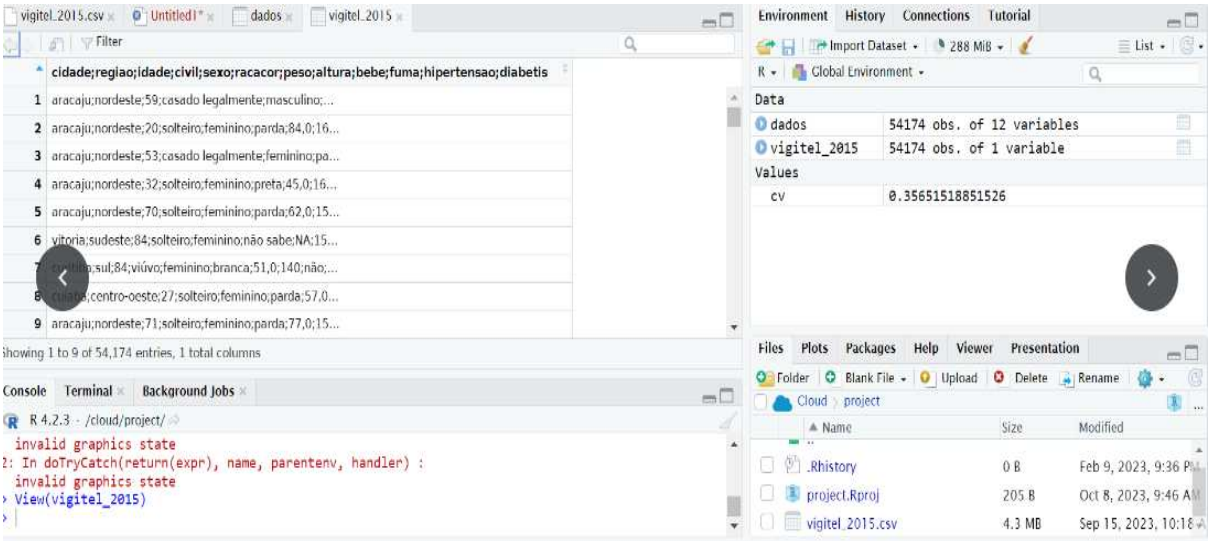
Figura 3.18: Arquivo disponível no R



Fonte: Acervo do autor

Observe que importamos o arquivo chamado "vigitel_2015.csv" e quando importamos o mesmo para o R, ele aparece todo desorganizado com apenas uma coluna (Figura 3.19), e portanto não conseguiremos efetuar o tratamento de dados. precisaremos fazer a importação do vigitel, e passaremos a chamá-lo de "dados", ficando a critério renomear ou não o arquivo.

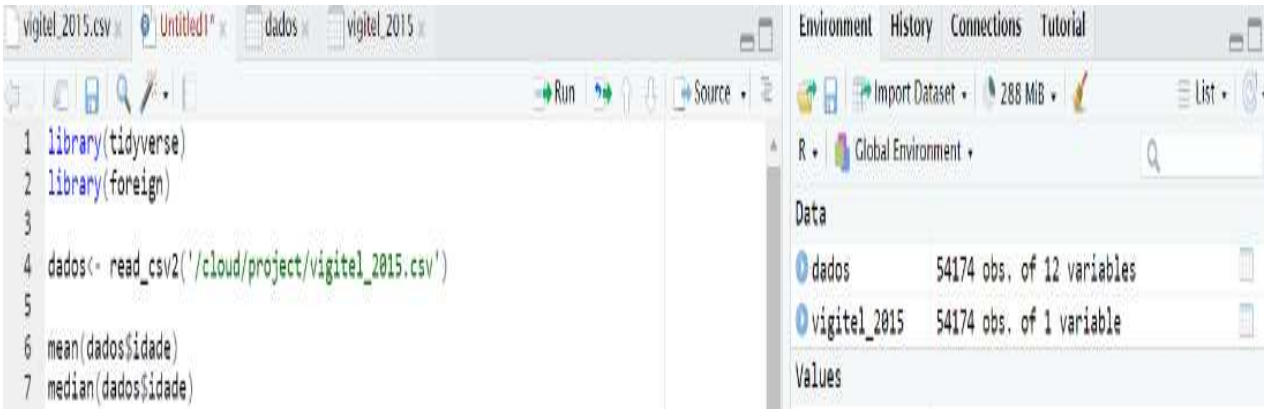
Figura 3.19: Base de dados do arquivo Vigitel no R



Fonte: Acervo do autor

Para realizar a importação utilizaremos o comando “read_csv2”, conforme a Figura 3.20, e a partir daí o R realizará a leitura dos dados do arquivo em 12 colunas e mais de 55 mil linhas, que será utilizada ao longo de todo o tratamento dos dados.

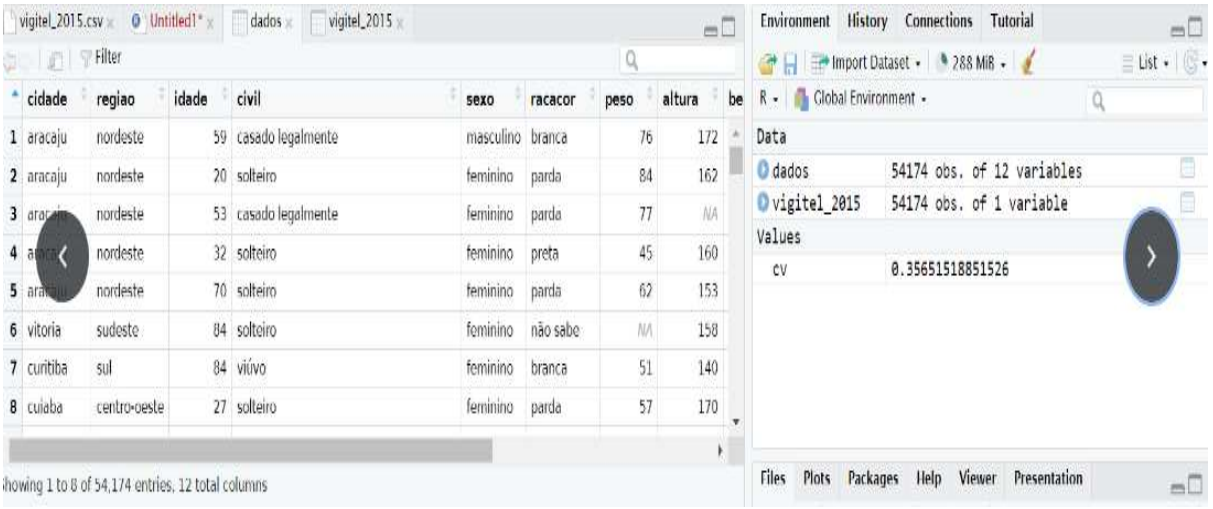
Figura 3.20: Importação do vigitel_2015.csv para dados



Fonte: Acervo do autor

Após esse processo, observe na figura 3.21 que os dados já se encontram organizados e prontos para serem tratados.

Figura 3.21: Base de dados do arquivo “dados”



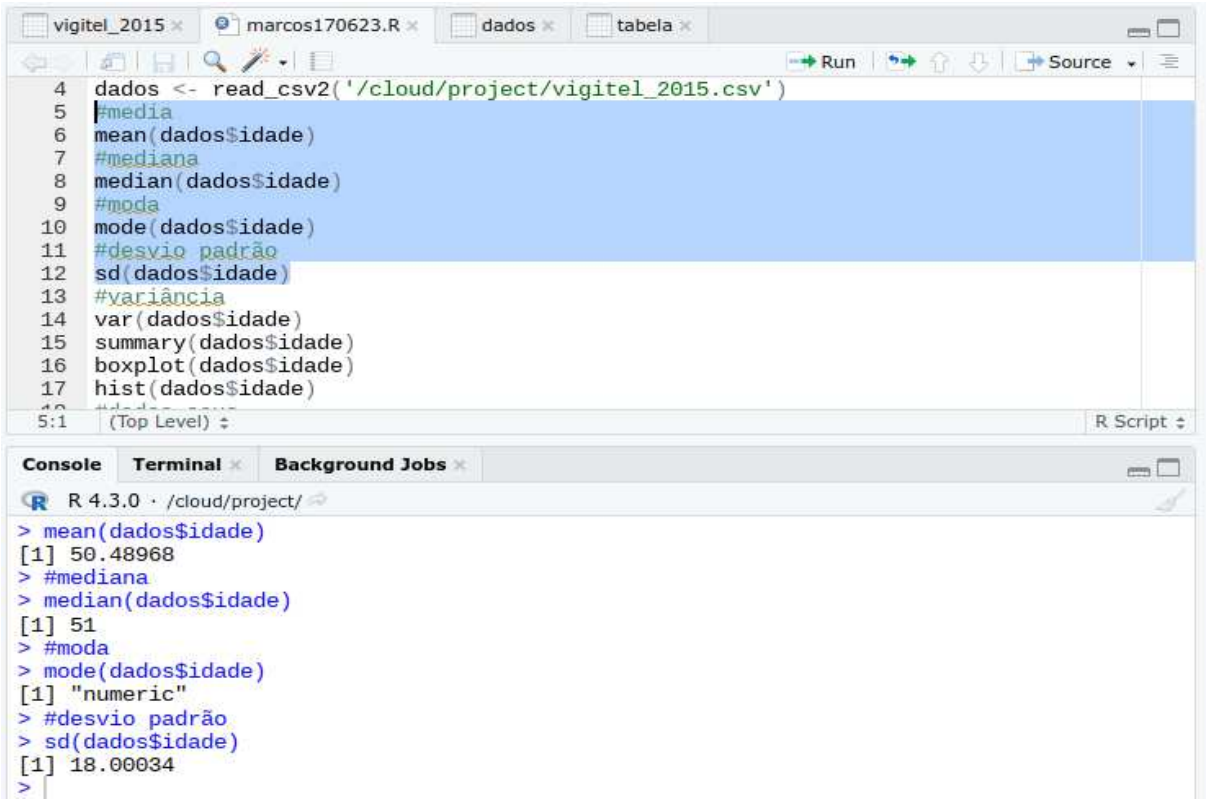
Fonte: Acervo do autor

Observe que importamos o arquivo chamado “vigitel_2015.csv” e passamos a chamar a tabela formada pelas linhas e colunas de “dados”.

3.6.3 - Análise da Variável Idade

A partir daí, já podemos utilizar o arquivo e alguns comandos para calcular as medidas de tendência central, de dispersão e gráficos. Abaixo temos alguns comandos e os valores e gráficos obtidos. Veja o cálculo da média, mediana, moda e desvio padrão da variável idade.

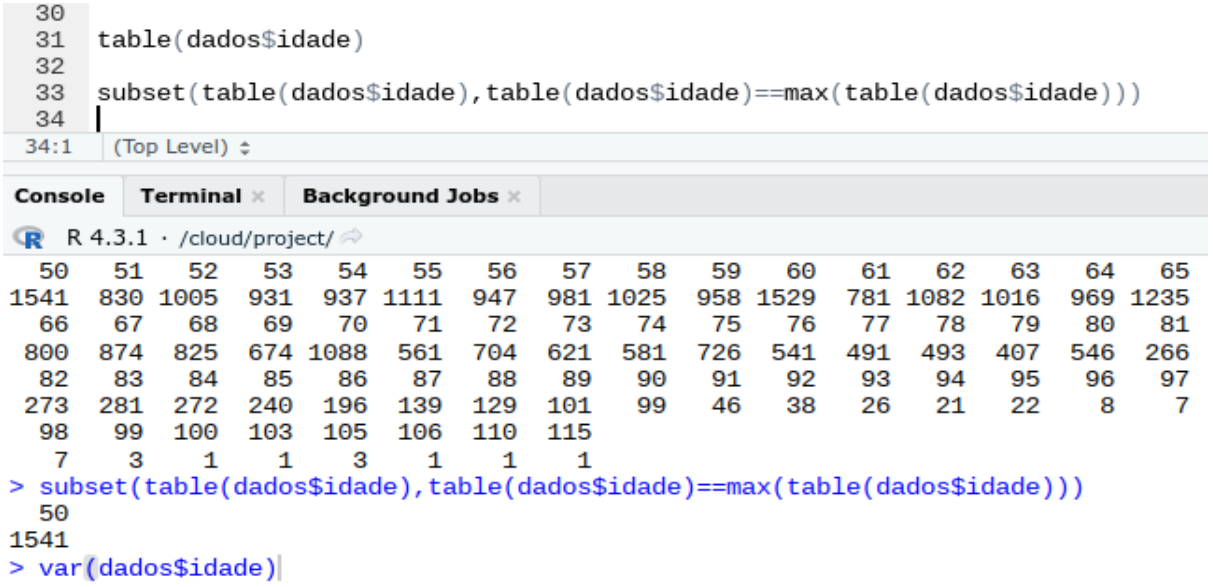
Figura 3.22: Média, mediana, moda e desvio padrão da variável idade



Fonte: Acervo do autor

Observe na figura 3.22, que a moda apareceu a notação “numerc”, isso quer dizer que a mesma é representada por um número. Porém que valor será esse? E como devemos proceder para obtê-la? Devemos utilizar a notação **table(dados$idade)**, porém aparecerá a frequência de cada número e dependendo da quantidade de dados, fica muito difícil de procurar a moda. Portanto, podemos recorrer a notação abaixo que facilitará esse trabalho. **subset(table(dados$idade),table(dados$idade)==max(table(dados$idade)))**

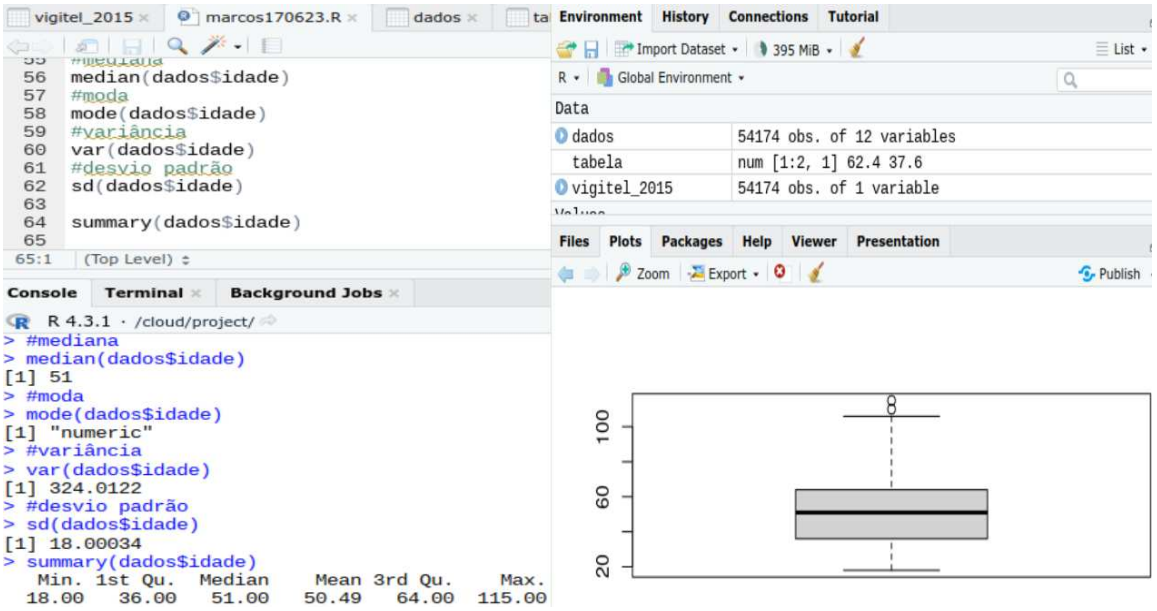
Figura 3.23: Moda da variável idade



Fonte: Acervo do autor

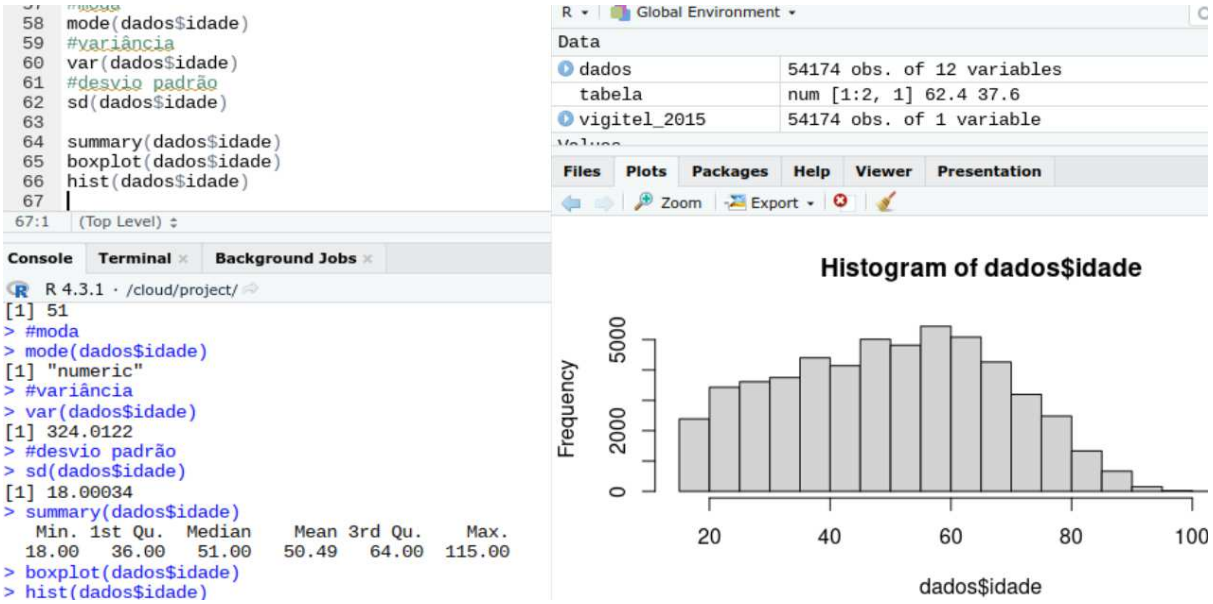
Cálculo da variância, desvio padrão, resumo dos dados da variável idade, boxplot e histograma. Para calcular a variância utilizamos a notação **var(dados$idade)**, o desvio padrão **sd(dados$idade)**, o resumo dos dados **summary(dados$idade)**, o gráfico boxplot **boxplot(dados$idade)** e o histograma **hist(dados$idade)**.

Figura 3.24: Variância, desvio padrão, resumo dos dados da variável idade e boxplot



Fonte: Acervo do autor

Figura 3.25: Variância, desvio padrão, resumo dos dados da variável idade e histograma

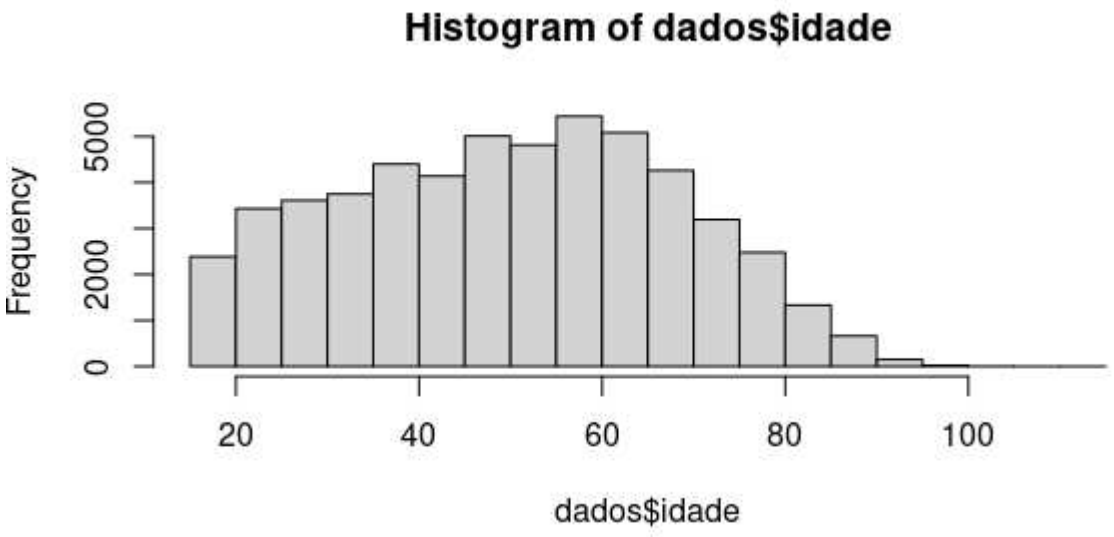


Fonte: Acervo do autor

3.6.4 - Histograma e Boxplot da variável idade

Observe a qualidade do histograma e do boxplot da variável idade.

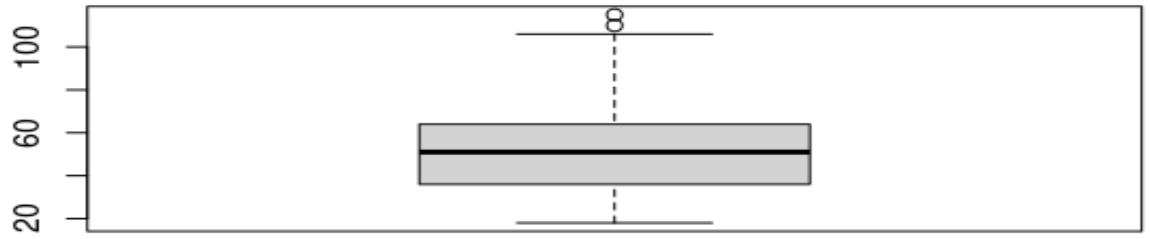
Figura 3.26: Histograma da variável idade



Fonte: Acervo do autor

O boxplot é uma disposição gráfica comparativa e nos fornece um resumo da variabilidade dos valores dos conjuntos de dados em estudo. Nos mostra os valores máximo e mínimo, os valores medianos, os quartis superiores e inferiores e até mesmo os valores atípicos no conjunto de dados.

Figura 3.27: Boxplot da variável idade



Fonte: Acervo do autor

Efetuamos aqui para a análise estatística para a variável idade. Para a variável peso, que também é uma variável quantitativa, o processo é análogo.

3.6.5 - O caso de uma Variável Qualitativa

Vamos expor agora o caso em que precisamos analisar uma variável qualitativa, como sexo, estado civil, raça, ou agravo. Neste caso, precisamos antes criar uma tabela, de modo a quantificar cada categoria.

Trabalhemos com a variável sexo. Criemos um tabela a fim de quantificar quantas pessoas nesta amostra são do sexo feminino e quantos são do sexo masculino.

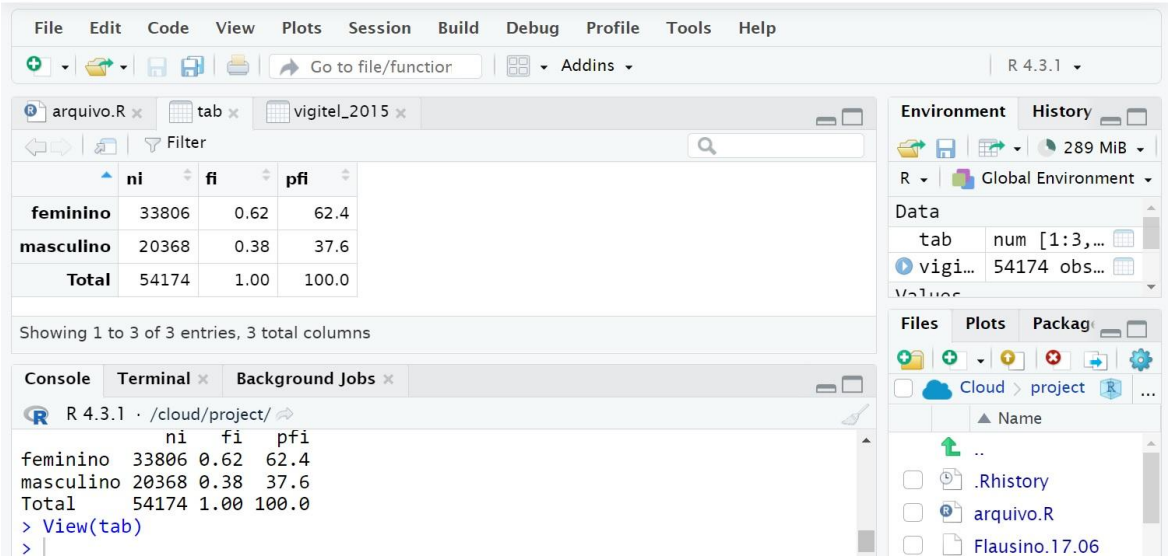
Podemos criar uma tabela de frequência, onde aparece a frequência absoluta, frequência relativa, e a frequência relativa em percentual. Os comandos são:

```
#sexo vigitel
ni<- table(vigitel_2015$sexo)#frequência absoluta
fi<-prop.table(ni)# frequência relativa
pfi<-100*fi#frequencia relativa em porcentagem
# Adiciona linhas de total
ni<-c(ni,sum(ni))
fi<-c(fi,sum(fi))
pfi<-c(pfi,sum(pfi))
names(ni)[3]<-"Total"
```

```
tab<-cbind(ni,fi=round(fi,digits=2),pfi=round(pfi,digits=2))
tab
```

Com isto, aparecerá na tela o seguinte:

Figura 3.28: Tabela de Frequência da Variável Sexo



Fonte: Acervo do Autor

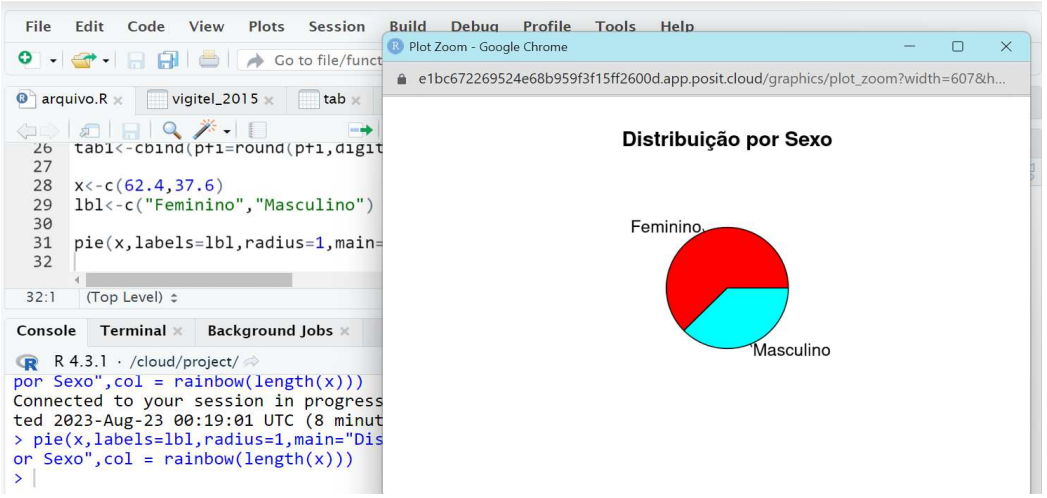
A fim de plotar um gráfico de pizza, que expressa bem uma variável qualitativa, vamos utilizar a quantificação relativa percentual. efetuamos então o seguinte:

```
x<-c(62.4,37.6)
lbl<-c("Feminino","Masculino")

pie(x,labels=lbl,main="Distribuição por Sexo",col = rainbow(length(x)))
```

e obtemos o seguinte gráfico:

Figura 3.29: Gráfico da distribuição da variável sexo



Fonte: Acervo do Autor

4 - Plano de aula

Apresentamos nesta seção um plano de aula a ser realizado com alunos do ensino médio sobre o conteúdo abordado nesta dissertação com o uso de software R. Desta forma, colocamos como proposta metodológica para o ensino de estatística, descrevendo um plano a ser realizado.

PLANEJAMENTO DIDÁTICO	
C. Curricular: Matemática	**Ano/Série: 3ºano**
Professor (a): Marcos Aurelio	**Período**: **1º**

Objetivo Geral: Aplicar os conceitos de Estatística na prática e utilizar a linguagem R para calcular as medidas de tendência central, dispersão e realizar construções gráficas.

Objetivos específicos: Compreender e discutir conceitos de variáveis, população e amostra;

Construir tabelas de frequência e gráficos;

Calcular média aritmética, mediana, moda, variância e desvio-padrão, utilizando a linguagem R; e

Realizar construções gráficas utilizando a linguagem R.

Observação: Cada aula é composta de 2 horários.

Nº DA AULA	CONTEÚDO (OBJETO DE CONHECIMENTO)	METODOLOGIA	RECURSOS	TAREFA DE AVALIAÇÃO
01	Estatística: Conceito, variáveis, população e amostra	Realizar uma pesquisa prévia na própria sala de aula com algumas variáveis como altura, massa e idade.	Balança digital, fita métrica ou trena, data show, pincel e quadro branco.	Realizar uma pesquisa com os familiares.
02	Estatística: Tabela de frequência e gráficos	Organizar os dados coletados em tabelas de frequência e construção de gráficos(barras, colunas, histograma, pizza, linha e boxplot)	Tablet, computador e ou materiais para confecção dos gráficos como carolinas, tintas, pilotos, etc.	Terminar e concluir pendências referente a construções gráficas.
03	Estatística: Média aritmética, mediana, moda, variância e desvio-padrão	Calcular as medidas de tendência central e de dispersão no “R”	Tablet, computador e data show	Calcular as medidas de tendência central e dispersão no “R”.
04	Estatística: Base de dados de Saúde Pública (DATASUS), importação para o “R” e técnicas de análise de dados.	Tratar dados de sistemas de saúde com grande número de observações e variáveis com R.	Tablet, computador e data show	Realizar os cálculos das medidas de tendência central e dispersão utilizando o “R”. Obter conclusões sobre os dados.
05	Estatística: Construir os mais variados tipos de gráfico utilizando o “R”	Realizar a construção dos gráficos das variáveis identificadas na pesquisa realizada	Tablet, computador e data show	Realizar a construção de gráficos utilizando o “R”.

5 - Resultados e Discussão

Este trabalho procurou apresentar uma forma distinta de ensinar o conteúdo de estatística no ensino médio, por meio da qual os alunos seriam agentes de sua própria aprendizagem, pesquisando na internet dados reais disponibilizados no site do Ministério da Saúde.

Com uso do software R, os estudantes fariam tratamento dos dados de saúde pública, e confecção de gráficos, o que possibilitaria o debate em um contexto de interdisciplinaridade.

Este texto iniciou justificando, por meio da BNCC mais recente, que as tecnologias precisam ser implementadas nas escolas de modo que o aluno seja colocado a refletir e propor soluções para problemas de ordem social, "articulando conceitos, procedimentos e linguagens próprios da Matemática."(BRASIL, pg. 526) Assim, trazer o software R, que em geral é utilizado no ambiente acadêmico e nas empresas, para o ambiente do ensino médio, constitui um passo importante, pois capacita o aluno para as linguagens de programação, além de tornar o ensino de estatística mais interessante, visto que esta geração de alunos é muito mais interessada na computabilidade do que na resolução manual.

Foram abordados conceitos elementares: medidas de tendência central, medidas de dispersão, histograma e boxplot e gráfico de pizza, para variáveis quantitativas e qualitativas. De fato, a proposta foi apresentar uma proposta metodológica de aula a ser implementada no ensino médio de uma escola. Para tal, finalizamos com um plano de aula que pode ser executado por qualquer professor que tenha acesso a este texto, e cuja escola possua um laboratório de informática, ou mesmo tenha tablets. Como realizamos a atividade utilizando uma versão do R em nuvem, pode inclusive ser realizado com estudantes a partir de smartphones.

Esperamos ter contribuído com novas estratégias para o ensino de estatística, a fim de aumentar o interesse de nossos estudantes em uma área tão relevante e presente em nossa sociedade.

Referências Bibliográficas

BRASIL. Base Nacional Comum Curricular. Brasília: MEC, 2017. Disponível em: http://basenacionalcomum.mec.gov.br/images/BNC C_20dez_site.pdf. Acesso em: 22 de dezembro de 2017.

BONILLA, Maria Helena S. Concepções do Uso do Computador na Educação. Espaços da Escola, Ano 4, No. 18 (59-68). Ijuí: 1995.

FARIAS A., SOARES, J. & CÉSAR, C.Introdução à Estatística. Rio de Janeiro: Ed. LTC, 2003.

SILVA, Hélio Medeiros et al. Estatística para o curso de economia, administração e ciências contábeis. São Paulo: Atlas, 1999.

CRESPO, Antônio Arnot. **Estatística fácil.** 15. ed. São Paulo: Saraiva, 2000.

BRUNI, A. L. Estatística para Concursos. 1ª Edição. Atlas, 2008.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2023. Disponível em: <https://www.R-project.org/>.