|  |  |
| --- | --- |
|  |  |



# 

# Pós-Graduação em Data Science

**Avaliação de Fundamentos da Linguagem “R”**

# Prof. Gabrielle Gomes dos Santos Ribeiro

# Nome:\_Elvis Felix Dantas\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_RA:\_ 9422410479\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**QUESTÕES**

**QUESTÃO 1**

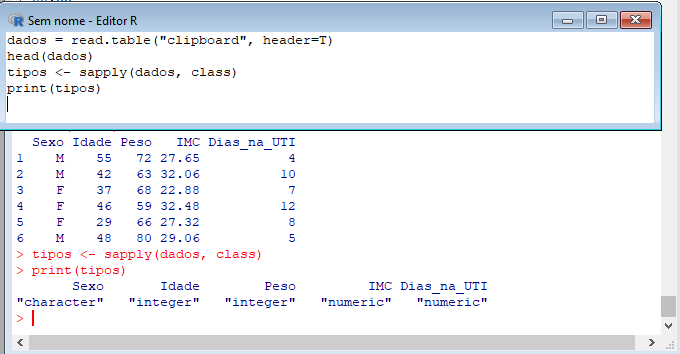
Entre com o seguinte conjunto de dados no R, lendo os dados através da área de transferência. Obs.: atenção ao separador decimal utilizado no IMC.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Paciente | Sexo | Idade | Peso | IMC |
| 1 | M | 55 | 72 | 27,65 |
| 2 | M | 42 | 63 | 32,06 |
| 3 | F | 37 | 68 | 22,88 |
| 4 | F | 46 | 59 | 32,48 |
| 5 | F | 29 | 66 | 27,32 |
| 6 | M | 48 | 80 | 29,06 |
| 7 | F | 50 | 71 | 22,55 |

Através de comandos do R responda as questões abaixo:

1. A partir da função *class*, determine o tipo de cada variável do conjunto de dados.

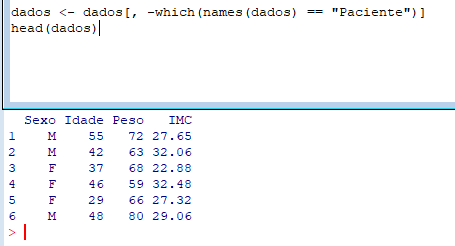
resultado



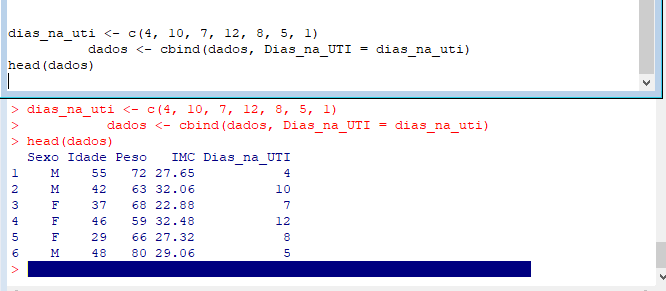
1. Exclua a coluna “Paciente” da tabela e insira a coluna da variável “Dias na UTI”:

Dias na UTI = 4, 10, 7, 12, 8, 5, 13

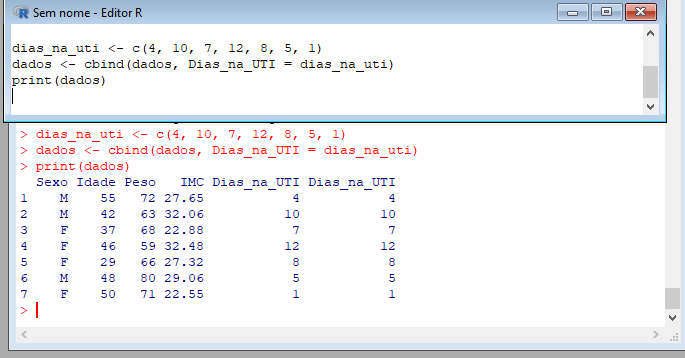
Excluindo



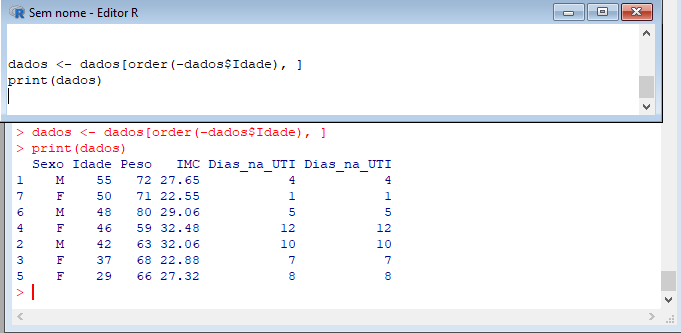
Incluindo



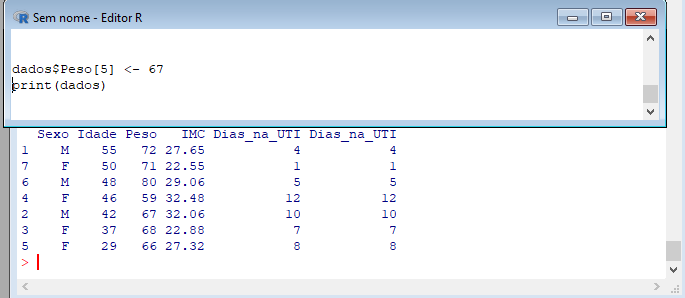
1. Filtre apenas os resultados da coluna Peso e IMC.



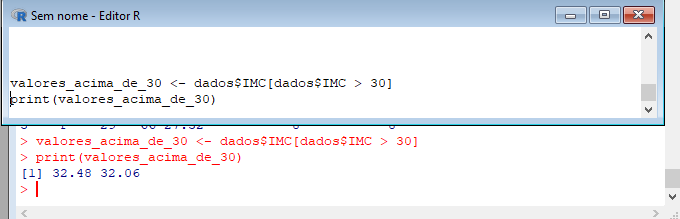
1. Coloque em ordem decrescente a variável Idade.



1. O peso da paciente 5 foi digitado errado, troque o valor por 67 sem digitar tudo novamente.



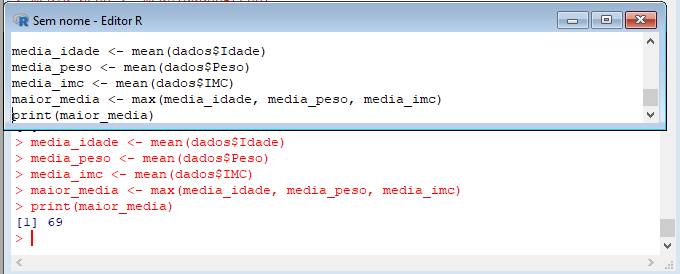
1. Determine quais valores de IMC estão acima de 30.



1. Sabe-se que a equação para o cálculo da média é:

Ou seja, é a soma de todos os valores do vetor, dividida pela quantidade de valores.

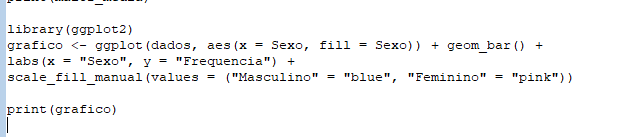
Sendo assim, faça um algoritmo em R que calcula as médias das variáveis Idade, Peso e IMC, e imprime somente o maior resultado entre os três.



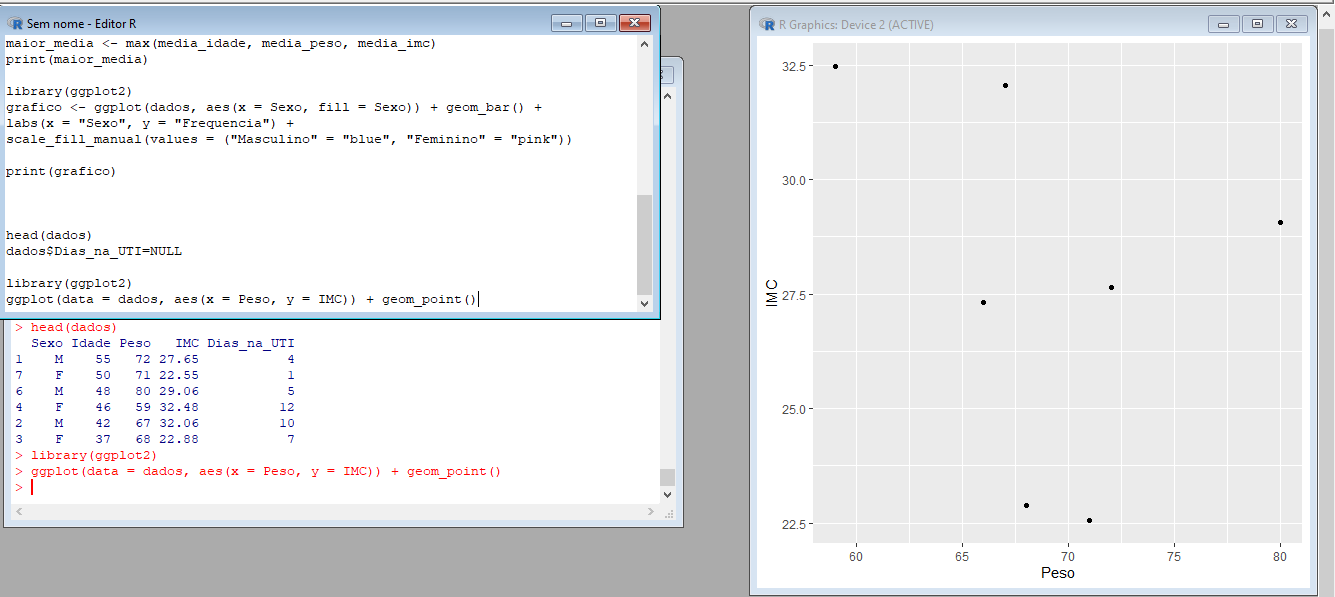
1. Através da função ggplot, construa um gráfico de colunas para a variável Sexo com as seguintes características:

- Coluna do Masculino na cor AZUL e do feminino na cor ROSA

- No eixo X o rótulo SEXO e no eixo Y o rótulo FREQUÊNCIA



1. Construa um gráfico de dispersão para as variáveis peso e IMC.



**QUESTÃO 2**

Aponte os 5 erros do seguinte algoritmo:

2resultados<-matrix(c(2,4,3) , c(“branca”, “parda”, “branca”), c(12,2 , 14,5 , 7,9 ))  
colnames(2resultados)=c(Opção, Cor, Distância)  
rownames(2resultados)=c(1,2,3)

print(2resultados)

opção cor distância

1 2 branca 12.2

2 1 parda 14.5

3 7 branca 7.9

média(2resultados$distância)

[1] 11.53

1ª O nome da Matriz está com número na frente.

2º A forma como os elementos foram separados dentro da matriz.

3º Precisa passar aspas duplas para o nome da coluna.

4º O uso de colnames() está errado

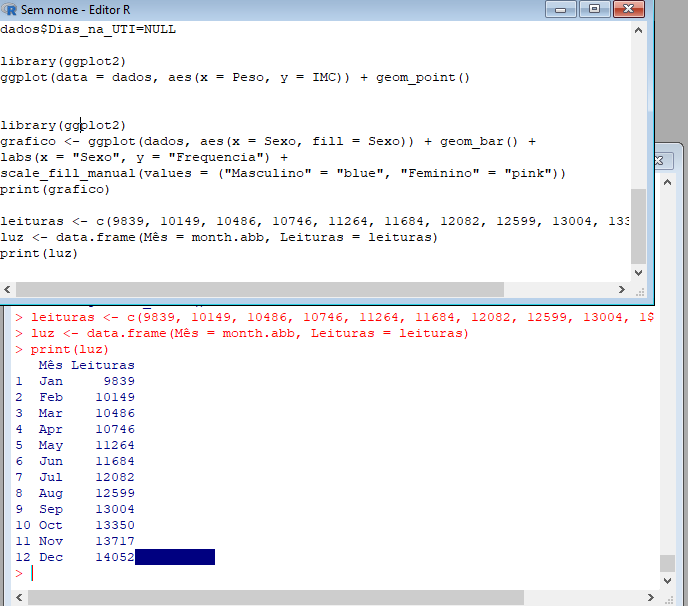
5º A forma de realizar a média precisa ser usado o médoto mean() para calcular a média da distância.

**QUESTÃO 3**

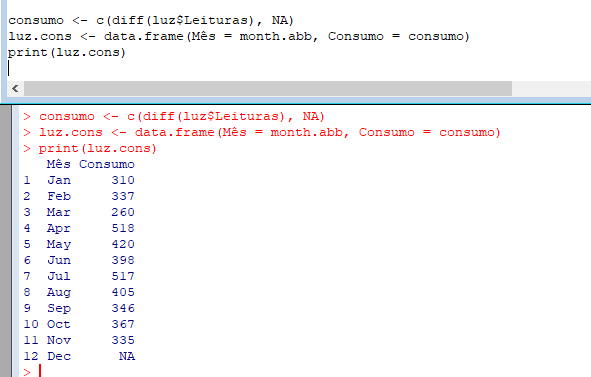
Considere que as leituras mensais do medidor de consumo de eletricidade de uma casa foram:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Jan | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Set | Out | Nov | Dez |
| 9839 | 10149 | 10486 | 10746 | 11264 | 11684 | 12082 | 12599 | 13004 | 13350 | 13717 | 14052 |

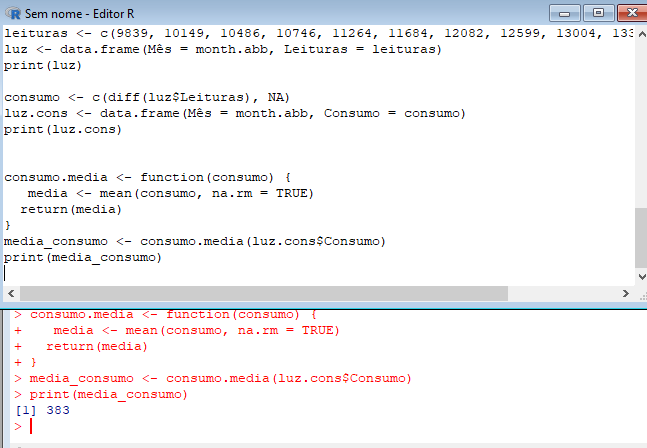
* Crie um objeto chamado *luz* com os valores das leituras, de janeiro a dezembro.



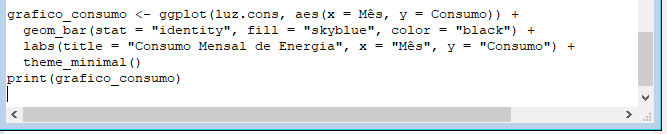
* Sabe-se que para fazer o cálculo do consumo mensal você deve subtrair da leitura atual a leitura do mês anterior. Calcule o consumo de cada mês neste período e guarde o resultado em um objeto chamado *luz.cons*.

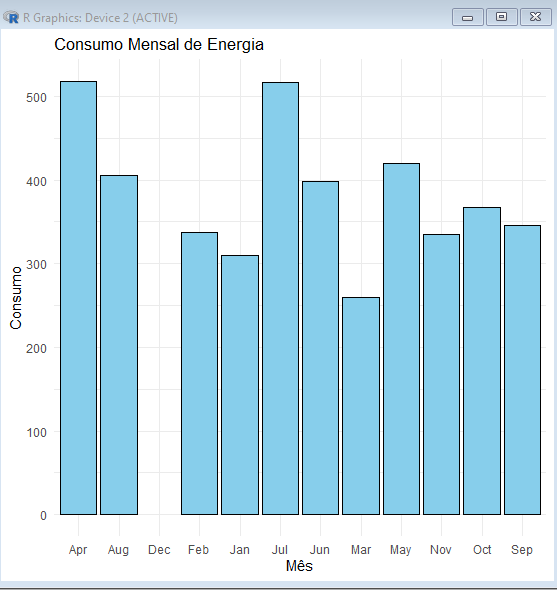


* Crie uma função chamada *consumo.media* e calcule o consumo médio mensal desta casa.
* Calcule o máximo e o mínimo de consumo mensal e guarde os resultados em um objeto chamado *luz.range*. Crie uma função que retorne os nomes dos meses de maior e menor consumo.



* Construa um gráfico que represente o consumo mensal desta casa.

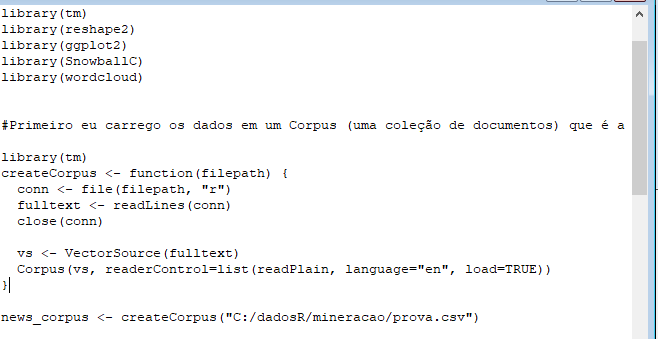


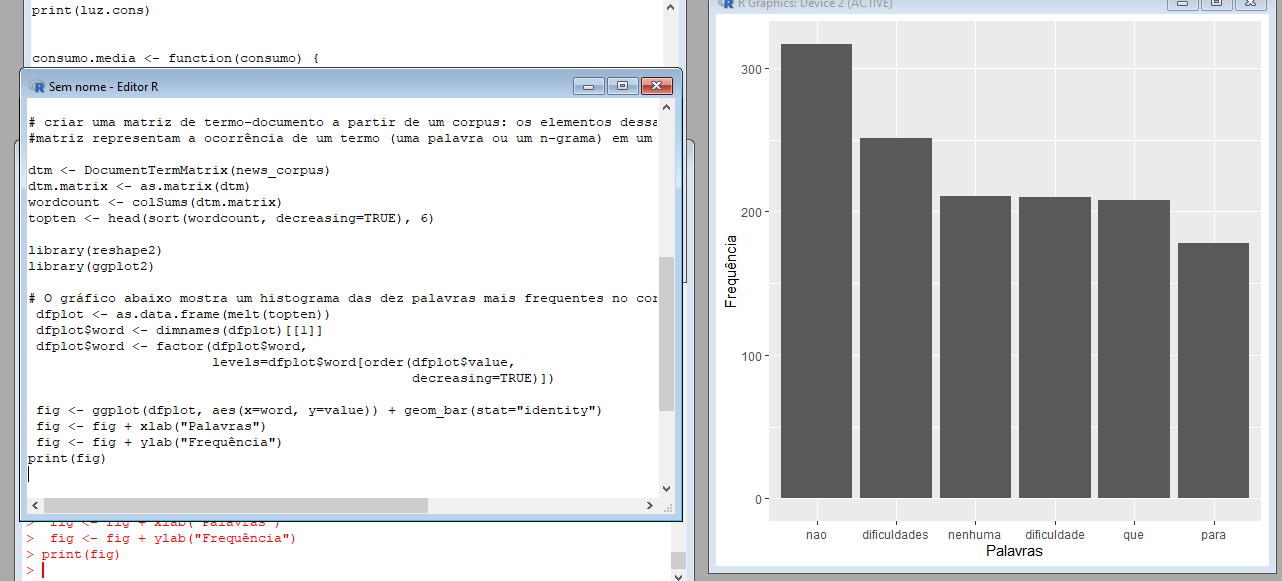


**QUESTÃO 4**

Utilize o conjunto de dados intitulado “prova.csv”, disponibilizado no Aprender, para realizar as seguintes etapas:

1. Carregue os dados em um Corpus.



1. Construa um gráfico histograma que represente as 6 palavras mais frequentes no corpus.
2. 
3. Construa uma “nuvem de palavras” colorida que tenha no máximo 25 palavras.

