

No Moodle, tópico de [material para as aulas](#) veja o arquivo para [instalação do software R](#)

Material para as aulas

- Visão geral da disciplina
- Apostila
- Técnicas de amostragem

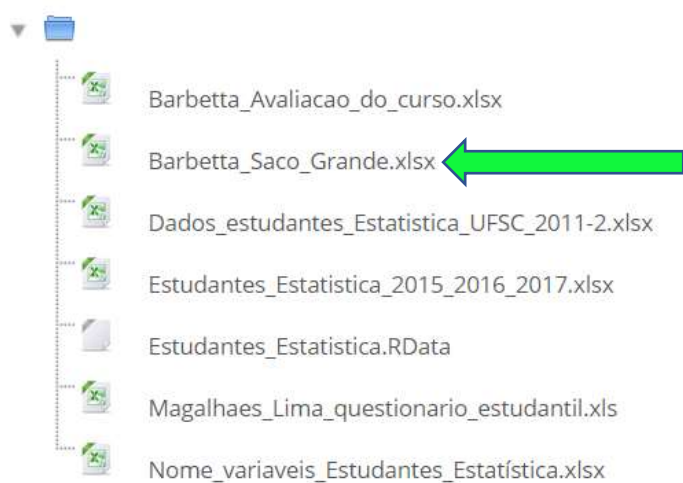
Nesse arquivo encontram-se os exemplos desenvolvidos sobre as técnicas de amostragem

- Instalação do software R
- Importação de arquivo de dados
- Conjuntos de Dados

A pós instalado o software R veja o arquivo que mostra como fazer a [importação de arquivo de dados](#).

Os dados que iremos trabalhar encontram-se na pasta de [conjuntos de dados](#), chamado [Barbetta_Saco_Grande](#)

Conjuntos de Dados



Faça o download do arquivo [Barbetta_Saco_Grande](#) e abra esse arquivo no seu computador. Veja se está aparecendo no [modo de exibição protegido](#). Se sim, você deve [Habilitar Edição](#) e salvar o arquivo onde você possa localizá-lo posteriormente.



Vamos explorar um pouco esse arquivo. Há três planilhas: Dados; Variáveis e Identificação.

Local	P.a.p	Instrucao	Tamanho	Renda
Monte Verde	não usa	segundo grau	4	10,3
Monte Verde	não usa	segundo grau	4	15,4
Monte Verde	usa	primeiro grau	4	9,6
Monte Verde	não usa	primeiro grau	5	5,5
Monte Verde	usa	segundo grau	4	9
Monte Verde	usa	nenhum	1	2,4
Monte Verde	não usa	segundo grau	2	4,1
Monte Verde	usa	segundo grau	4	8,6
Monte Verde	usa	segundo grau	4	12,9
Monte Verde	usa	primeiro grau	4	8,4
Monte Verde	não usa	primeiro grau	6	1,1
Monte Verde	usa	nenhum	4	8,6
Monte Verde	não usa	primeiro grau	4	12,9
Monte Verde	não usa	primeiro grau	6	8,4

Na planilha Dados encontra-se o conjunto de dados que foram levantados na pesquisa e que iremos fazer as análises. Na primeira coluna (A) aparece Local, na segunda coluna (B) P.a.p.; terceira coluna (C) Instrução; quarta coluna (D) Tamanho e quinta coluna (E) Renda. Essas são as características (variáveis) que temos dos respondentes. Rolando a página você pode verificar que os dados terminam na linha 121 (como a primeira linha está informando os nomes das variáveis, os dados dos respondentes iniciam na segunda linha, portanto há 120 respondentes (o tamanho da amostra, $n=120$).

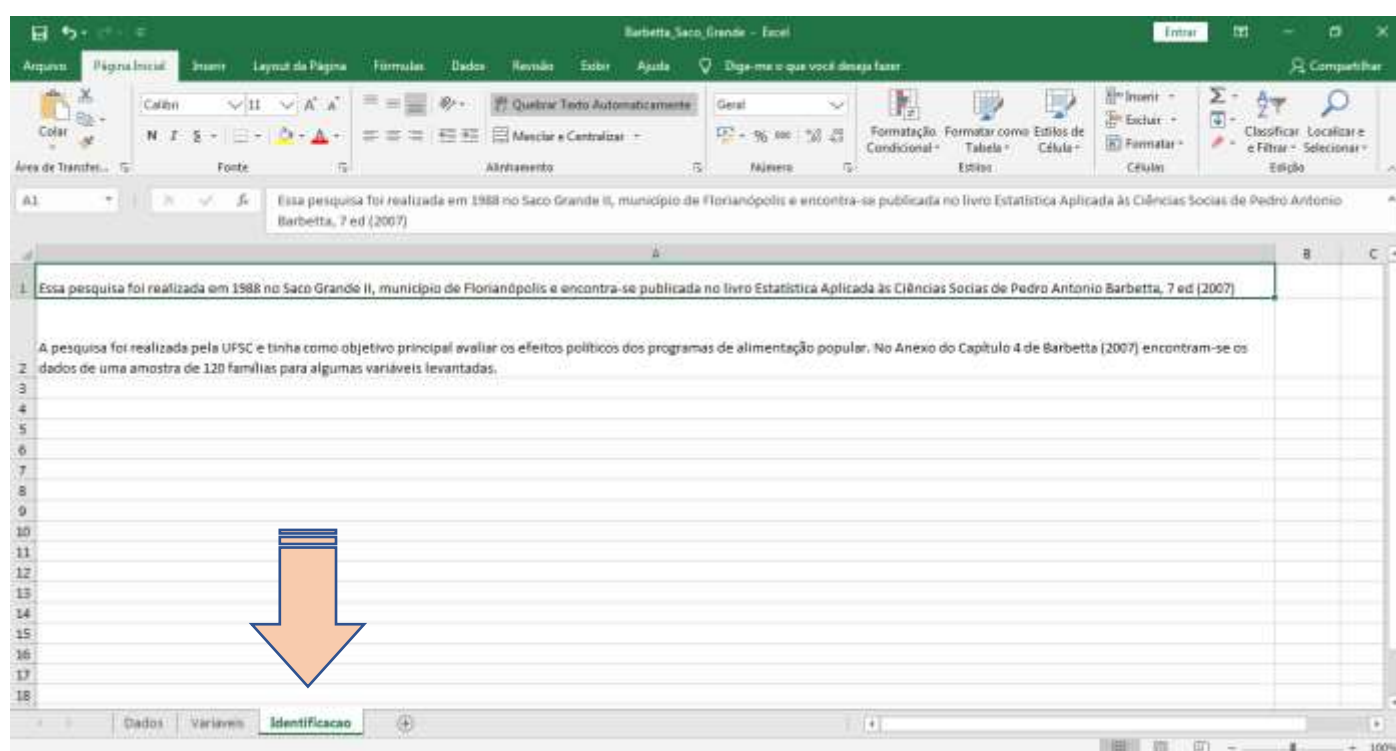
Agora vamos para a planilha denominada **Variáveis**

Local	localidade onde reside o respondente
	Conjunto Residencial Monte Verde,
	Conjunto Residencial Parque da Figueira,
	Encosta do morro
P.a.p	Programa de alimentação popular
	não usa
	usa
Instrucao	Grau de instrução do chefe da casa
	nenhum
	primeiro grau
	segundo grau
Tamanho	Número de pessoas residentes no domicílio
Renda	Renda familiar mensal, em quantidade de salários mínimos

Nessa planilha encontramos as informações do que significam os nomes curtos atribuídos às variáveis do conjunto de dados. Assim:

- Local é a localidade onde reside o respondente. Os respondentes são de três localidades: Conjunto residencial Monte Verde; Conjunto Residencial Parque da Figueira e Encosta do Morro.
- P.a.p. Programa de alimentação popular. Para essa variável as repostas fornecidas são: não usa e usa.
- Instrução: Grau de instrução do chefe da casa. Nesse conjunto de dados os chefes das casas podem ser analfabetos, nenhum grau de instrução, ou com o primeiro grau (corresponde atualmente ao ensino fundamental) e com segundo grau (correspondendo ao ensino médio).
- Tamanho. Número de pessoas residentes no domicílio
- Renda. Renda familiar mensal, em quantidade de salários mínimos. Essa variável nos informa a renda em salários mínimos.

A terceira planilha nos informa a identificação desse conjunto de dados



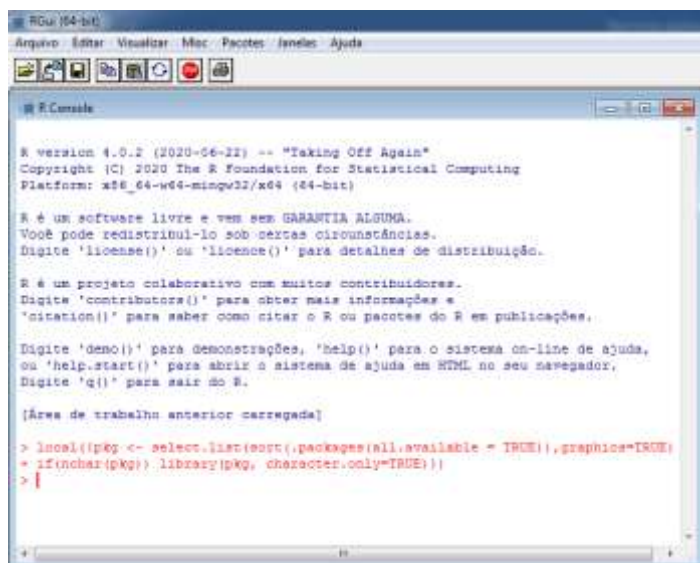
Essa pesquisa foi realizada em 1988 no Saco Grande II, município de Florianópolis e encontra-se publicada no livro Estatística Aplicada às Ciências Sociais de Pedro Antonio Barbetta, 7 ed (2007).

A pesquisa foi realizada pela UFSC e tinha como objetivo principal avaliar os efeitos políticos dos programas de alimentação popular. No Anexo do Capítulo 4 de Barbetta (2007) encontram-se os dados de uma amostra de 120 famílias para algumas variáveis levantadas.

Agora que estamos inteirados a respeito desse conjunto de dados vamos importá-lo para o software R.

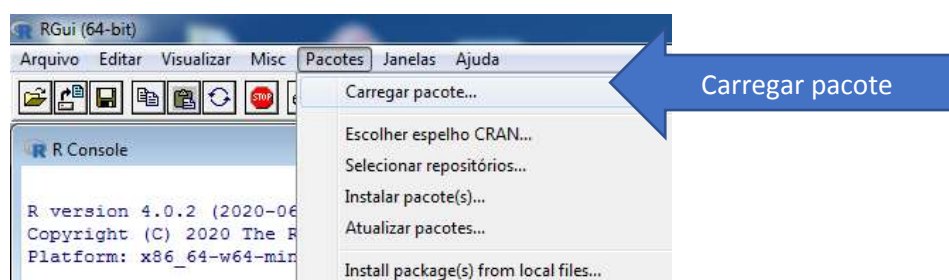
Veja o arquivo que mostra como fazer a **importação de arquivo de dados** que está no Moodle.

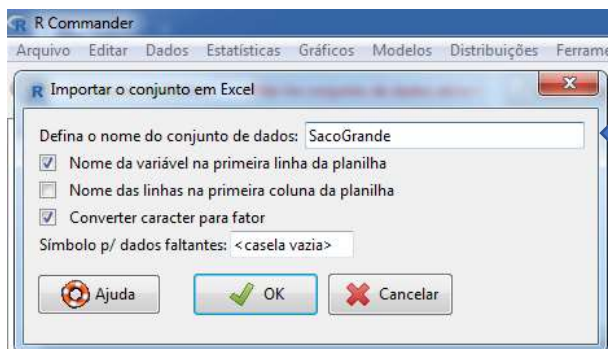
Quando você abre o software R aparece a janela do RGui, que é a área usada pelos programadores ou usuários que saibam empregar as funções do software para executar o que necessitam. O software R é uma linguagem de programação e possibilita fazer muita coisa para quem sabe programar com essa linguagem.



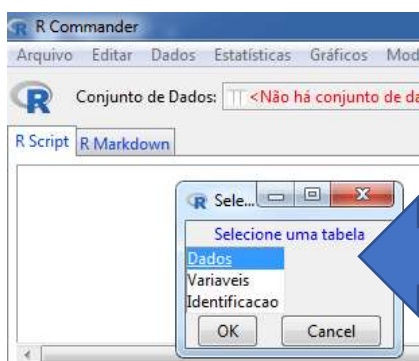
Para essa disciplina usaremos o pacote que traz as funções estatísticas já programadas, facilitando nosso uso. O nome desse pacote é Rcmdr. A primeira vez que você for usar o pacote necessita instalar. Se ainda não fez, veja as instruções no arquivo para **instalação do software R** que está no Moodle.

Revisando... Instala o software R e instala o pacote Rcmdr. Após instalado o pacote Rcmdr sempre que for iniciar alguma atividade quando abrir o software R necessita carregar o pacote Rcmdr.

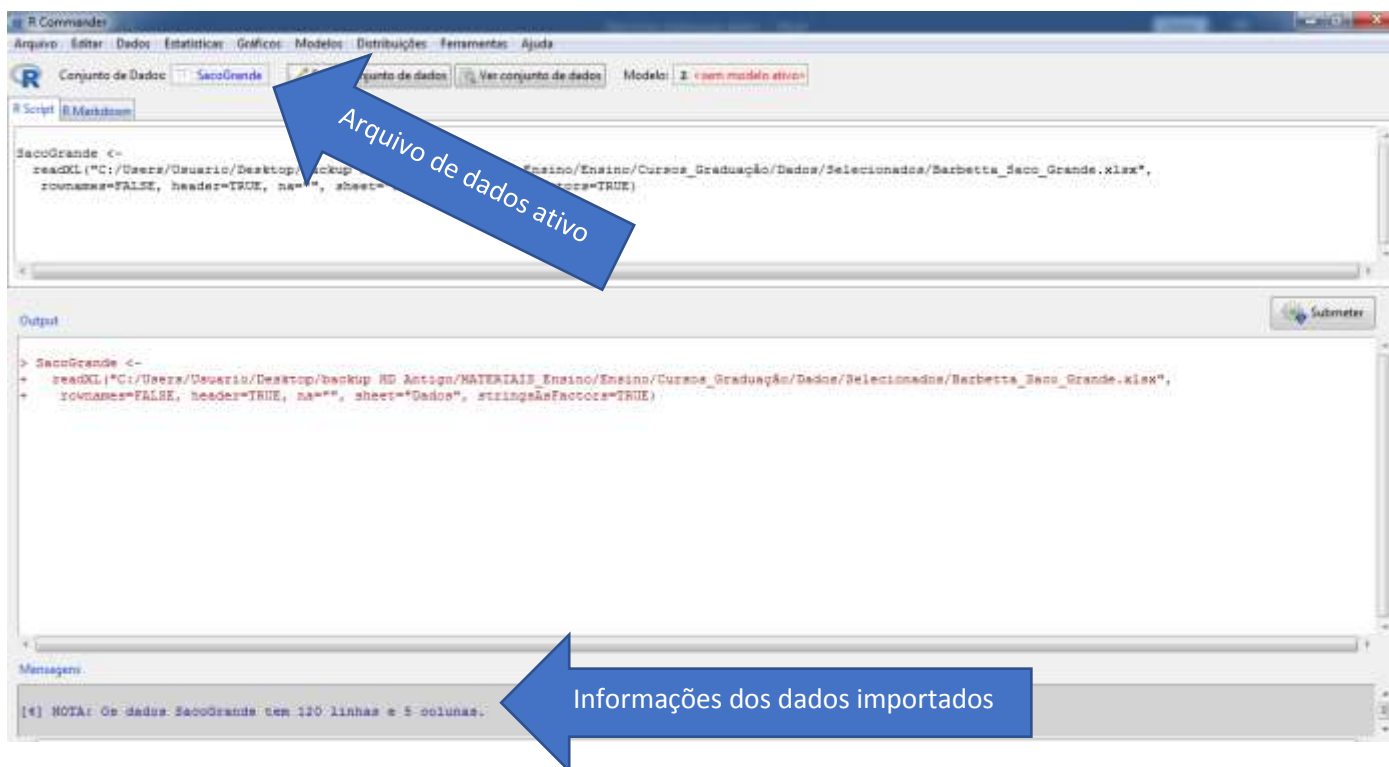




Dê um nome curto sem caracteres especiais



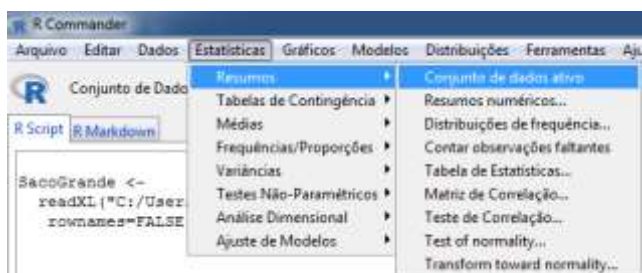
Selecione a planilha Dados



Arquivo de dados ativo

Informações dos dados importados

Agora vamos iniciar as análises exploratórias dos dados (AED).



No menu do R Commander clique em **Estatísticas**, selecione **Resumos** depois selecione **Conjunto de dados ativo**

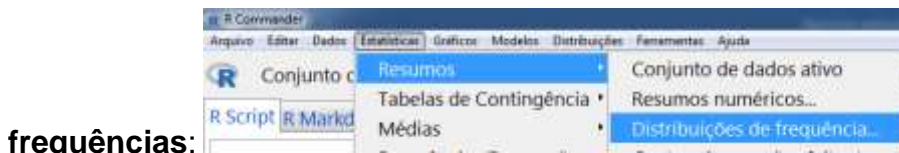
summary(SacoGrande)				
Local	P.a.p	Instrucao	Tamanho	Renda
Encosta do Morro:37	não usa:42	nenhum :38	Min. :1.000	Min. : 0.10
Monte Verde :40	usa :78	primeiro grau:38	1st Qu.:3.750	1st Qu.: 3.90
Pq. Da Figueira :43		segundo grau :44	Median :4.000	Median : 5.40
			Mean :4.492	Mean : 6.34
			3rd Qu.:5.000	3rd Qu.: 8.15
			Max. :9.000	Max. :25.70
				NA's :1

Resumo do conjunto de dados Saco Grande II

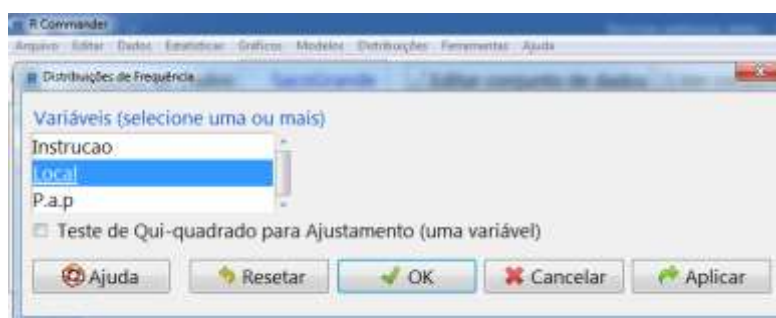
Observe que os resultados para a variável Local, mostram as quantidades de respondentes para cada uma das três localidades: Encosta do Morro: 37; Monte Verde: 40; Parque da Figueira: 43. São as quantidades em valores absolutos.

O que esses resultados nos informam é que nessa pesquisa os respondentes estão distribuídos do seguinte modo nas três localidades: são 37 que moram na Encosta do Morro, 40 residem no Monte Verde e 43 moram no Parque da Figueira. Essas quantidades são valores próximos uns dos outros. Isso é, no Parque da Figueira é onde temos maior número de respondentes, seguido pela Monte Verde. A Encosta do Morro é o local com o menor número de respondentes. Porém, as quantidades em valores absolutos não são muito informativas por não temos noção de quanto isso representa na amostra toda. Para melhorar nossa interpretação podemos calcular os percentuais.

Poderemos obter esses resultados em percentuais, que é uma forma mais fácil de interpretarmos. Para isso vá no menu do R Commander **Estatísticas > Resumos > Distribuições de**



Selecione a variável que estamos analisando e dê OK:



```
counts:
Local
Encosta do Morro      Monte Verde  Pq. Da Figueira
              37              40              43
```

Quantidades em valores absolutos

```
percentages:
Local
Encosta do Morro      Monte Verde  Pq. Da Figueira
              30.83              33.33              35.83
```

Quantidades em percentuais

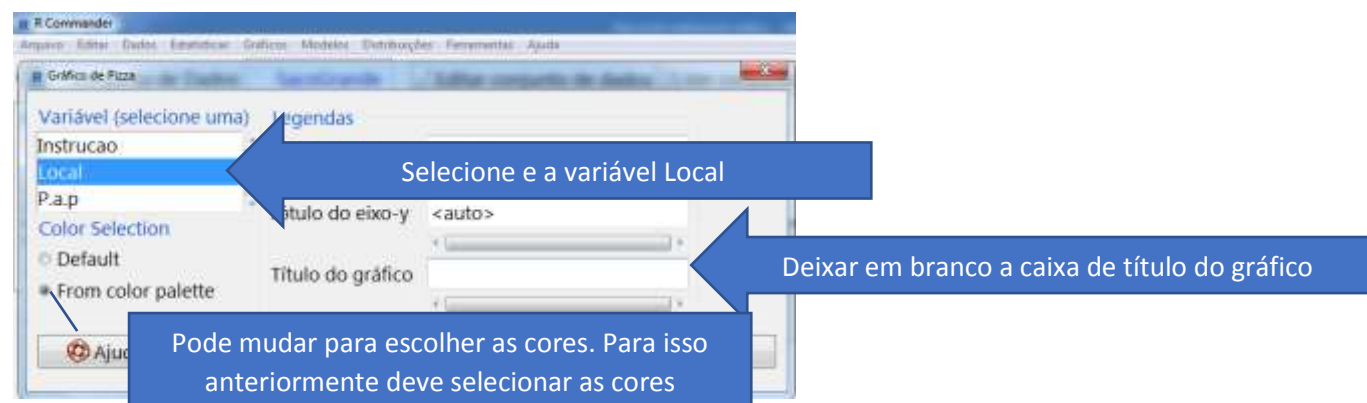
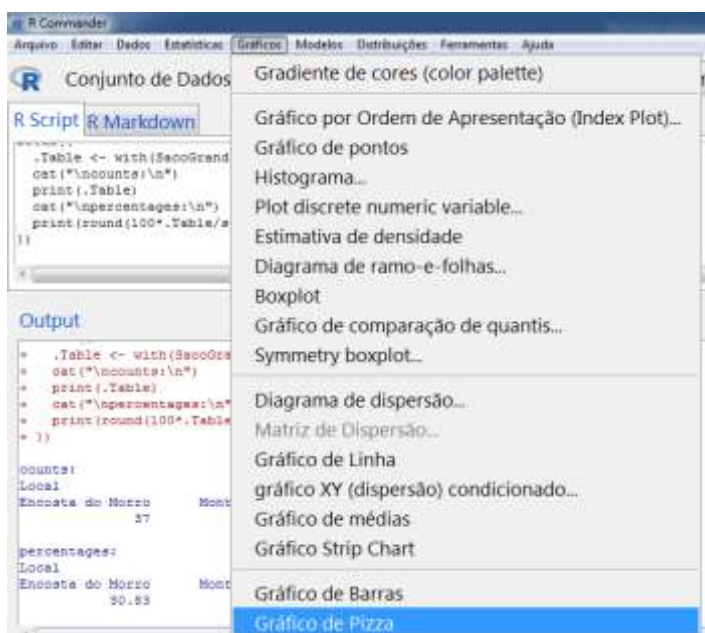
Quando se faz uma pesquisa deseja-se divulgar os resultados em congressos, revistas, jornais ou bancas de avaliação, como TCC, mestrado, doutorado, etc. Para isso se escrevem as análises obtidas dos resultados fornecidos pela pesquisa, que normalmente é feita em forma de texto. Assim, não se apresentam os resultados com um “prints creen” da tela gerada pelo software porque o leitor não tem obrigação de conhecer esse software. Portanto, deve-se apresentar esses resultados em forma de tabela ou na forma de gráfico.

Para gerarmos um gráfico dessa variável.

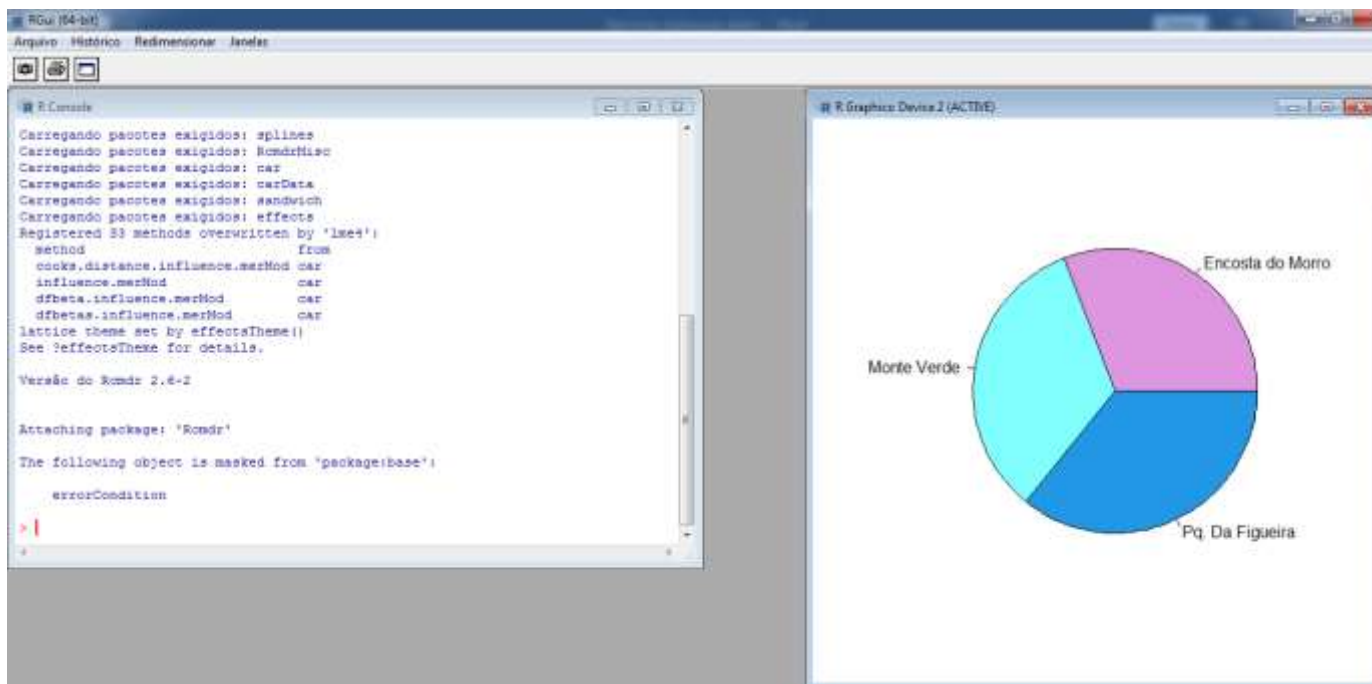
Para escolher as cores para os gráficos. No menu do R Coomander > Gráficos > Gradiente de Cores (color palette):



Para escolher o gráfico a ser criado: No menu do R Commander > Gráficos > Gráfico de Pizza

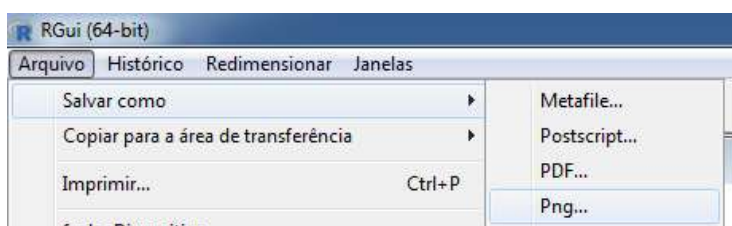


O gráfico irá aparecer na janela do R Gui:



Você pode salvar o gráfico para inserir posteriormente no seu texto.

No menu do RGui > Arquivo > Salva como > Png



escolha a extensão do arquivo .png. Esse é uma forma de salvar a imagem com qualidade sem ser muito pesada.

Assim, temos três opções para apresentarmos os resultados para o leitor:

1) Na forma de texto. Por exemplo:

A pesquisa apresenta três localidades onde residem os respondentes, sendo que a maior quantidade é do Parque da Figueira com 43 respondentes (35,83%). O monte Verde aparece em segundo lugar com 40 respondentes (33,33%) e na Encosta do Morro tem-se a menor quantidade com 37 respondentes (30,83). Observando os resultados em percentuais constata-se a amostra de respondentes está distribuída equilibradamente com representantes das três localidades.

2) Na forma de tabela. Por exemplo:

Tabela 1 – Localidade onde residem os respondente

Localidade	Quantidade	Percentual
Encosta do Morro	37	30,83
Monte Verde	40	33,33
Parque da Figueira	43	35,83
Total	120	99,99

Fonte: Barbetta, 2015

Aberta nas
lateria

Quando se apresentam os resultados na forma de tabela deve-se seguir as regras de construção de tabelas, da ABNT. Apresentar as tabelas numeradas com algarismos arábicos contendo o título (nome da tabela), cabeçalho, corpo e fonte. A numeração é sequencial (Tabela 1, Tabela 2, Tabela 3 ...) O título deve ser completo, informando o leitor do que se tratam os resultados

apresentados na tabela. O cabeçalho é a primeira linha da tabela. Corpo é conteúdo apresentado. Fonte é de onde foram retirados os dados. As tabelas devem ser abertas nas laterais. Se fizer fechada caracteriza “quadro”, que é diferente de “tabela”.

Qual é a diferença entre Quadro e Tabela?

Tabela apresenta resultados numéricos quando tem totalizador. Quadro se apresentam textos, como resumos, esquemas comparativos de conceitos, etc.

3) Na forma de gráfico

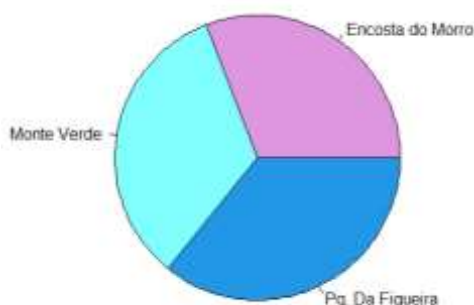


Figura 1 – Localidade onde residem os respondente

Fonte: Barbetta, 2015

Os títulos dos gráficos devem ser colocados na área externa da figura (por isso deixamos o espaço em branco para o título na hora de gerarmos o gráfico no software).

Do mesmo modo que nas tabelas, os gráficos devem ser numerados com algarismos arábicos, sequencialmente (Figura 1, Figura 2, Figura 3) e independente de outras numerações como de Tabelas ou Quadros. O título do gráfico deve ser completo, informando ao leitor do que se trata a ilustração. Também deve apresentar a fonte (de onde foram extraídos os dados).

Portanto, quando se faz a apresentação dos resultados de uma pesquisa opta-se por tabela ou gráfico. Em qualquer uma das modalidades (tabela ou gráfico) deve-se escrever um texto com a interpretação dos resultados para o leitor.

Pode-se mesclar entre usar textos, tabelas e gráficos para cada uma das características (variáveis do conjunto de dados). Você pode optar, por exemplo apresentar a localidade usando uma tabela, o programa de alimentação popular através de texto e o grau de instrução do chefe da casa por meio de um gráfico.