



Universidade Federal de Uberlândia
Faculdade de Engenharia Elétrica
Graduação em Engenharia Biomédica

Roger Amaral Pires

Visualizações de dados no Software R: Um guia prático.

Uberlândia
2017

Roger Amaral Pires

Visualizações de dados no Software R: Um guia prático.

Trabalho apresentado como requisito parcial de avaliação na disciplina Trabalho de Conclusão de Curso de Engenharia Biomédica da Universidade Federal de Uberlândia.

Orientador: Adriano Alves Pereira

Assinatura do Orientador

Uberlândia
2017

Dedico este trabalho à minha mãe e ao meu irmão, pelo estímulo, carinho e compreensão.

Agradecimentos

Agradeço à minha mãe, Euci, ao meu irmão, Ricardo, à minha namorada, Nathália; por todo apoio, carinho, compreensão e paciência, sempre estando presentes e sendo essenciais para esta conquista.

A todos meus amigos, em especial, Eustáquio Fernandes, Homero de Castro, Hugo Morais, Murillo Marcos, Thiago Siqueira, Victor Morales, Yuran Dias e Yuri Cassiolato, sem o apoio diário deles não seria possível chegar até aqui.

A todos professores e funcionários do curso de Engenharia Biomédica e da Faculdade de Engenharia elétrica, em especial, ao Prof. Dr. Adriano Alves Pereira, pela orientação, amizade e generosidade em compartilhar seus conhecimentos para o sucesso deste trabalho.

E agradeço especialmente, ao meu pai, Paulo, que não se encontra mais presente, pelo o seu carinho, dedicação e ensinamentos passados, dedico esta conquista a ele.

RESUMO

No atual contexto em que vivemos, tem-se evidenciado a importância do valor da informação, na maioria das vezes não se sabe por onde começar e qual são a forma correta para demonstração e visualização da informação. O software R é uma importante ferramenta para análise e na manipulação de dados. Este trabalho objetiva servir como um guia prático para usuários que pretendem gerar gráficos no software R. Para isto, são descritas a programação de gráficos desenvolvidos, apresentando as linhas de códigos com as suas devidas explicações. Para ser de fácil compreensão e entendimento todas as linhas são explicadas de forma sucinta e em linguagem simples, todos os resultados gerados são demonstrados nas figuras, assim como, as linhas de códigos. Foram obtidas 19 formas de visualizações de dados, onde todas as estruturas de cada técnica são comentadas e exemplificadas em seguida. As diferentes técnicas e métodos das visualizações podem ser considerados como ferramentas de qualidades. Eles permitem uma melhor compreensão ao público de elementos não gráficos, gerando interesses e criam credibilidade para os pontos que você deseja evidenciar.

Palavras-chave: *Visualizações de dados; Software R; Guia Prático; Gráficos.*

ABSTRACT

In the current context in which we live, the importance of the value of information has been evidenced, most of the time we do not know where to start and what is the correct way for demonstration and visualization of information. R software is an important tool for data analysis and manipulation. This work objective to be a practical guide for users who intends to generate graphics in the R software. For this, the programming of developed graphics is described, presenting the lines of codes with their explications. To be easy to comprehension and understand all the lines of the code are explicate succinctly and in simple language, all the generated results are demonstrated in the figures, as well as the lines of codes. We obtained 19 forms of data visualization, where all the structures of each technique are commented and exemplified next. The different visualization techniques and methods can be considered as tools of qualities. They permit to the public a better understand of non-graphic elements, create interests, and create credibility for the points you want to evidence.

Keywords: *Data visualizations; R; Practical Guide; Charts.*

LISTA DE TABELAS

Tabela 1 - Relação de variáveis e seus valores.....	19
Tabela 2 - Relação de gráficos e dados aleatórios.....	20

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação da formação de um dendrograma.....	14
Figura 2 – Exemplo de gráfico Boxplot.....	21
Figura 3 – Exemplo de gráfico Scatter plot.....	22
Figura 4 – Exemplo de gráfico Hexbin Plot.....	23
Figura 5 – Exemplo de gráfico Heat Map.....	24
Figura 6 – Exemplo de beanplot.....	25
Figura 7 – Exemplo de Forest Plot.....	27
Figura 8 – Exemplo de gráfico de coluna.....	29
Figura 9 – Exemplo de gráfico de Barras.....	30
Figura 10 – Exemplo de gráfico de coluna Empilhado.....	31
Figura 11 – Exemplo de gráfico de barras empilhado.....	32
Figura 12 – Exemplo de gráfico de colunas lado a lado.....	33
Figura 13 – Exemplo de gráfico de barras lado a lado.....	34
Figura 14 – Exemplo de Histograma.....	35
Figura 15 – Exemplo de Gráfico de linha para uma série.....	36
Figura 16 – Exemplo de Gráfico de linha para várias séries.....	37
Figura 17 – Exemplo de Gráfico de área para uma série.....	38
Figura 18 – Exemplo de Gráfico de área para várias séries.....	39
Figura 19 – Exemplo de gráfico de setor.....	40
Figura 20 – Exemplo de gráfico de setor em porcentagem.....	41

Sumário

1 INTRODUÇÃO.....	10
2 DESENVOLVIMENTO.....	11
2.1 Revisão bibliográfica.....	11
2.1.1 Evolução da Visualização de Dados.....	11
2.1.2 Boxplot.....	12
2.1.2 Scatter Plot.....	12
2.1.3 Hexbin Plot.....	13
2.1.4 Heat Map.....	13
2.1.5 Beanplot.....	15
2.1.6 Forest Plot.....	15
2.1.7 Gráfico de coluna e de barra.....	16
2.1.8 Histograma.....	16
2.1.9 Gráfico de Linha.....	17
2.1.10 Gráfico de Área.....	17
2.1.11 Gráfico de Setor ou de Pizza.....	17
3 Materiais e Métodos.....	18
4 Resultados.....	20
4.1 Boxplot.....	20
4.2 Scatter Plot.....	21
4.3 Hexbin Plot.....	22
4.4 Heat Map.....	24
4.5 Beanplot.....	25
4.6 Forest Plot.....	26
4.7 Gráfico de coluna e de barra.....	28
4.8 Histograma.....	35
4.9 Gráfico de Linha.....	36
4.10 Gráfico de Área.....	38
4.11 Gráfico de Setor ou de Pizza.....	40
5 CONCLUSÕES.....	42
6 REFERÊNCIAS.....	43
APÊNDICE.....	48

1 INTRODUÇÃO

As visualizações de dados são uma ferramenta que ampliam a capacidade humana de interpretações e questionamentos de informações tanto quantitativas quanto qualitativas e também as relações entre os dados (MONTEIRO, 1999). No atual contexto em que vivemos, tem-se evidenciado a importância do valor da informação, tais visualizações de informações estão diretamente conectadas com as técnicas e ferramentas que permitem a visualização, interpretação e exploração de uma quantidade imensa de volume de dados (MANSOUR, 1998).

Na maioria das vezes não se sabe por onde começar e qual é a forma correta para demonstração e visualização da informação. Todo o esforço e tempo para a coleta e armazenamento de dados podem ser ineficazes se a forma de visualizar os dados não for de fácil compreensão e interpretação (TRAINA; et al, 2001).

Em síntese, os dados apresentados pelos gráficos podem representar variáveis qualitativas e quantitativas, podendo apresentar valores absolutos ou relativos, o emprego de qual técnica que se enquadra melhor deve ser analisado.

O software R é uma importante ferramenta para análise e na manipulação de dados, possuindo visualizações de testes paramétricos e não paramétricos, apresenta uma grande facilidade para criação de diversos tipos de gráficos, no qual o usuário do software tem plena consciência e controle do gráfico, além disso, possui uma linguagem de fácil aprendizagem (DE SOUZA; PERTENELLI; DE MELLO, 2014).

O software R apresenta uma gama imensa de novos métodos para análise de dados e informações (DA SILVA; DINIZ; BORTOLUZZI, 2009). Nele, há diversas e rápidas atualizações de pacotes novos e já existentes. Estes pacotes possuem explicações das funções, permitindo e facilitando o desenvolvimento, interpretação e realização das análises estatísticas. Além disso, possuem ajuda para suas funções, pois alguns pacotes vêm acompanhados de demonstrações e formas de execuções (DO AMARAL; et al., 2010).

Em suma o software R tem como principais vantagens ser um software gratuito, possuir todos os códigos abertos, sendo eles reproduzíveis, seus códigos são adaptáveis, é desenvolvido para diversas plataformas e possui uma excelente documentação.

O presente trabalho propõe a criação de um guia prático para desenvolvimento de gráficos no Software R, por meio de linguagem sucinta e de fácil entendimento e compreensão. Neste sentido, este trabalho objetiva fornecer uma forma de apoio e esclarecimentos para futuros usuários do software.

2 DESENVOLVIMENTO

2.1 Revisão bibliográfica

2.1.1 Evolução da Visualização de Dados

O conteúdo desta subseção foi extraído de Dulcleri Sternart Alexandre e João Manuel R. S. Tavares (2007). A partir deste estudo é possível observar que os primeiros trabalhos utilizando visualizações de dados são de Willian Playfair (1786), que é o inventor de diversos métodos gráfico: o gráfico de barras, o gráfico de linhas e o gráfico circular. Bertin (1967) publicou uma teoria que identificava os elementos básicos dos diagramas e descrevia as estruturas correspondentes. Em 1983, Tufte (1983) publicou uma teoria em que enfatizava a densidade de informações úteis. Edward Tukey (1977) principiou um movimento que enfatizava a análise através de figuras para facilitar a compreensão de forma prática dos dados.

Cleveland e McGill, em 1988, escreveram o livro *Dynamic Graphics for Statistics*, que esclarecia as novas formas de visualizar as informações na área de estatística. Uma das grandes problemáticas que eles afrontaram foi como observar dados fixos que possuíam várias variáveis. Inselberg e Dimsdale, em 1990, desenvolveu O método de coordenadas paralelas.

Mackinlay (data) criou um sistema que gerava automaticamente ótimas representações de dados, formalizando a teoria de Bertin (1967). Com o avanço exponencial do hardware e do software na área gráfica foi e é possível criar novas interfaces. Recentemente, vários trabalhos começaram a investigar a relação entre fatores humanos e visualização de dados.

As diversas formas de que existem nos fazem nos questionar quais as melhores formas de demonstrar e visualizar dados. Quando empregado de forma correta é uma excelente ferramenta de apoio, para isso é necessário conhece-las.

2.1.2 *Boxplot*

O *boxplot* é uma das técnicas mais utilizadas para visualização de dados univariados (HUBERT; VANDERVIEREN, 2008). Esta técnica foi desenvolvida por John W. Tuckey (1977), por meio dela é possível avaliar simetria dos dados, presença de “pontos fora da curva” (em estatísticas, são dados que estão muito distantes das demais observações), também denominados de *outliers* (CAPELA; CAPELA, 2011).

É uma ferramenta simples, mas muito poderosa, pode-se comparar vários conjuntos de dados referentes à uma mesma variável (BENJAMINI, 1988) (CAPELA; CAPELA, 2011).

A versão mais popular do *boxplot* utiliza as medianas e partes dos quartis porque são de fácil cálculo (GOLDBERG; IGLEWICZ, 1992). Na sua construção, a linha na caixa demonstra a posição de um valor central típico (mediana), os valores adjacentes são ligados à caixa por linhas tracejadas, esses valores devem ser tratados como possíveis *outliers* (FRIGGE; HOAGLIN; IGLEWICZ, 1989).

Schneider e Silva (2014) aplicaram a técnica do *boxplot* com a finalidade de identificar os anos padrão secos, chuvoso e habituais, durante o período de 1980 a 2012 da microrregião de Dourados, localizada em Mato Grosso do sul.

2.1.3 *Scatter Plot*

O gráfico de dispersão (*Scatter Plot*) é o mais utilizado para o estudo da relação entre duas variáveis (ECOR, 2016). É uma representação gráfica entre pares de variáveis, estas variáveis podem ser características de qualidade ou de duas causas (FM2S, 2017).

O diagrama de dispersão permite a visualização dos dados de entrada, é usado para representar conjuntamente as grandezas dos dados em um gráfico de duas dimensões (SHIKAMURA, 2012).

O *scatter plot* possui dois eixos de valores, sendo assim é possível mostrar um conjunto dos dados coletados no eixo horizontal (eixo x) e um outro conjunto de dados na vertical (eixo y) (MICROSOFT, 2017a). Cada elemento do gráfico é relativo à uma intersecção a um par de dados. Por meio deste método é possível a

identificação e intensidade da possível correlação entre as variáveis e também sendo possível a comprovação da relação entre duas causas ou efeitos (MARKETING FUTURO, 2015).

Masunari e Swiech-Ayoub (2003), utilizaram este método para analisar a relação entre a largura da carapaça e o comprimento da maior quela nos machos de *Uca leptodactyla*, e relacionar a largura da carapaça e a largura do abdome em fêmeas de *Uca leptodactyla*.

2.1.4 Hexbin Plot

O *hexagon binning (hexbin)* pode ser visto como um forma alternativa do gráfico de dispersão (PACKTPUB, 2016). O hexbin plot é um histograma bivariado utilizado para visualizar estruturas de dados com um grande número de amostras (LEWIN-KOH, 2016).

Este método possui diversas funções, tais como, suavização bivariada, encontrar uma aproximação bivariada da mediana e identificar a diferença de dois conjuntos de hexágonos na mesma escala (LEWIN-KOH, 2016).

Este método possui um conceito extramente simples, o plano xy é composto por uma grade regular de hexágonos, os números de pontos que cada hexágono possui é contado e armazenado e uma estrutura de dados, a cor da plotagem é propocional ao número de pontos nele contido (LEWIN-KOH, 2016).

Moreira, Simioni e Santana (2016), fizeram uso da técnica de *hexbin plot* para analisar o comportamento de viabilidade econômica em relação às variáveis de risco no regime de manejo para uma e duas rotações, na produção de lenha de eucalipto na região de Itapeva-SP.

2.1.5 Heat Map

O heat map é uma representação bidimensional dos dados, ele substitui os valores por intensidade de cor (ROUSE, 2011) (YAU, 2010). Existem várias maneiras de se mostrar uma *heat map*, mas todos se assemelham em um ponto, utilizam a cor para relacionar valores e dados de difícil compreensão (ROUSE, 2011).

Com um *heat map* simples é possível resumir imediatamente as informações em forma visual, com um *heat map* mais elaborado é possível compreender dados e informações complexas (ROUSE, 2011).

O *Heat Map* é uma imagem que possui uma cor falsa, esta cor representa a intensidade proporcional do valor daquele determinado ponto. Possui um dendrograma em cima e ao lado esquerdo. Normalmente, a reordenação das linhas e colunas são de acordo com as restrições impostas pelo o dendrograma (THE STATS PACKAGE R, 2015). É realizado uma média de todos elementos da coluna, as médias são representadas pela Figura 1 (a), depois realiza-se o agrupamento dos dois elementos mais próximos (figura 1 (b)), após o agrupamento realiza-se a média do agrupamento (Figura 1(c)), após este procedimento, torna-se a realizar o agrupamento dos dois elementos mais próximos (Figura 1(d)), repetindo-se esse processo até forma a árvore do dendrograma (Figura 1(g)). Repete-se este processo para a linha depois.

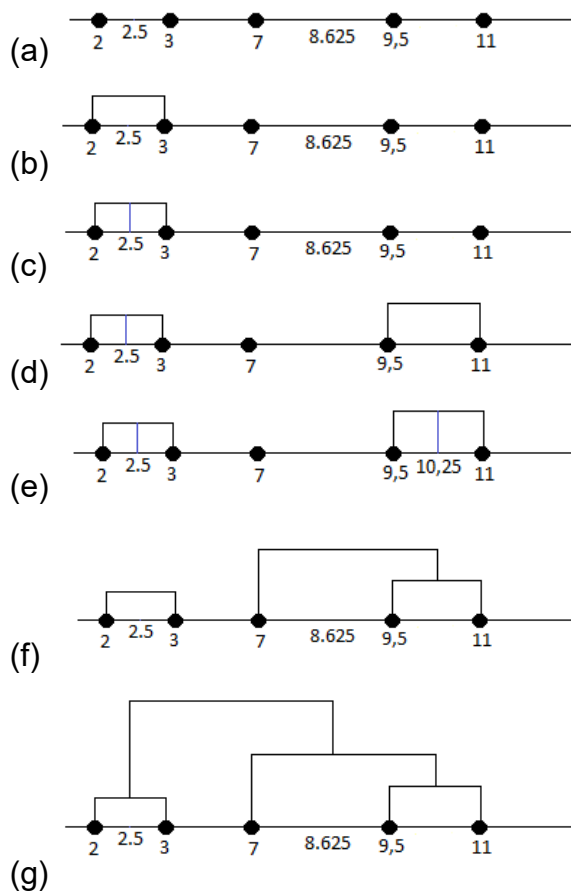


Figura 1 – Representação da formação de um dendrograma

Zacarias, et al. (2015), utilizou a técnica de *Heat Map* para comparar os índices a evolução dos acidentes de trabalhos nas regiões do Brasil no período de 2003 a 2008, para dados abertos.

2.1.6 Beanplot

O *Beanplot*, assim chamado devido à sua forma, é uma técnica implementada e desenvolvida por Kampstra em 2008, o *beanplot* possui comandos muito similares ao *boxplot* (MUTHERS; MATZARAKIS, 2010). O *beanplot* combina a curva de densidade com *scatter plot* (CAMEY; NUNES; CRUZ, 2010).

Por meio do *beanplot* é possível demonstrar a média do conjunto de dados, enquanto outros método exibem a mediana. Nesta técnica existe a possibilidade de evidenciar a média geral de um conjunto de dados para mais de um grupo de sujeitos ou de variáveis (CAMEY; NUNES; CRUZ, 2010).

O *beanplot* é um gráfico onde são mostrados múltiplos lotes, esta técnica é composto por um gráfico de dispersão, que consiste em uma pequena linha para um conjunto de dados, sua distribuição é o corpo da densidade e possui uma linha média para cada distribuição. Ao lado dessa forma, possui uma linha tracejada que indica uma média geral, esta é a forma padrão de um *beanplot* (KAMPSTRA, 2008, 2015).

Molina, et al. (2013), fizeram uso do *beanplot* para demonstrar as estimativas de consumo energético, consumo de carboidrato, lípidios e proteínas, que foram mensuradas pelo Questionário de frequência alimentar, que foi ajustado para energia, e registros alimentares, de um grupo proveniente do Estudo Longitudinal de Saúde do Adulto no Brasil.

2.1.7 Forest Plot

A técnica de *forest plot* foi criada na década de 70 e é frequentemente utilizada em metanálise, não sendo restritas apenas a esta, existem sugestões que esse método foi assim chamado devido à sua forma que se assemelha à uma floresta de linhas (GORDON, 2017)(RODRIGUES, 2010).

Este artifício é muito utilizado quando se deseja comparar múltiplos resultados ou estudos, extramamente utilizado para comparar um grupo controle com um grupo experimental. Outra vantagem é que por meio desse método é possível mostrar estimativas brutas e ajustadas em dados separados (GORDON, 2017).

Godinho, et al. (2012), utilizou a técnica de *forest plot* para análise comparativa de eventos de acidente vascular cerebral, da mortalidade por acidente vascular cerebral, de eventos de complicações renais e eventos de septicemia

2.1.8 Gráfico de coluna e de barra

Esta representação possui várias técnicas para a sua visualização. Esta visualização de dados é a mais utilizada para variáveis qualitativas, tem como objetivo comparar grandezas e/ou alterações durante um determinado período de tempo (GUEDES; MARTINS; ACORSI, 2010) (MICROSOFT, 2016b).

Esta técnica possui uma gama grande de opções, podemos destacar o empilhado e barras lado a lado. Bruschini e Puppini (2004) utilizaram as diversas formas de gráfico de barras e de colunas para demonstrar de mulheres executivas no mercado de trabalho no final do século XX, segregando por sexo e faixa etária.

2.1.9 Histograma

É uma das ferramentas mais utilizada para demonstrar variáveis contínuas, utilizado para representar quando há uma grande quantidade numérica, esta técnica é representação gráfica da distribuição de frequência (CALLEGARI-JACQUES, 2003). Sua análise permite uma interpretação concisa e simples (KUOKAWA, 2002). É muita vezes utilizada como uma ferramenta para análise de qualidade.

Silva, Guimarães e Tavares (2003), fizeram o uso da técnica de histograma para demonstrar a variação de as precipitações anuais e mensais na estação Getúlio Vargas, em Uberaba, durante os anos de 1914 a 2000.

2.1.10 Gráfico de linha

O gráfico de linha é uma técnica eficaz para estabelecer as relações e comparações de uma série de dados, amplamente utilizada para auxiliar à tomada de decisão. Sua análise permite identificar e analisar ao longo do tempo, quais as tomadas de decisões que devem ser tomadas imediatamente, possibilitando uma organização correta e efetiva de recursos humanos e bens materiais (SELEME; STADLER, 2008).

Dorneles e Waechter (2004) fizeram uso do método de gráfico de linhas para relacionar o número cumulativo de espécies arbóreas por número de quadrantes, a distribuição de indivíduos por classe de diâmetro, as alturas estimadas e o número de indivíduos e relação entre as alturas estimadas e o número de *Myrcia multiflora* e *Syagrus romanzoffiana*, estudo realizado no Parque Nacional da Lagoa do Peixe, Rio Grande do Sul.

2.1.11 Gráfico de área

O gráfico de área é uma série de dados como um conjunto de pontos conectados por uma linha e abaixo desta linha toda sua área é preenchida (Microsoft, 2017c).

No gráfico de área empilhado é demonstrada a soma acumulada dos grupos estudados, assim é possível avaliar a contribuição de cada grupo. Nesta técnica as séries são empilhadas no eixo y, e o eixo x compõe normalmente os intervalos de tempo (Minitab, 2017).

Soares (2017) fez uso desta ferramenta para comparar 10 diferentes cenários teóricos, entre 1 bilhão e 10 bilhões de barris, de volume recuperável de petróleo durante um período de 36 anos.

2.1.12 Gráfico de Setor ou Gráfico de pizza

Os gráficos de pizzas normalmente são utilizados para estabelecer uma comparação, harmonização ou contribuição de cada tipo dado para o todo, ele é a representação de uma série estatística em círculo (CORREA, 2003).

O círculo representa o total das contribuições, os setores (ou fatias) são proporcionais à contribuição de cada série, os valores podem ser expressos em números ou em porcentagem (CORREA, 2003) (MARTINS; ALCOFORADO, 2015).

Azzoni (2005) utilizou o gráfico de setores para relacionar a composição setorial do produto interno bruto do estado de São Paulo e demonstrando sua evolução do ano de 1998 em relação ao ano de 1985.

3 Materiais e Métodos

Neste tópico, serão apresentados os materiais e métodos utilizados para a criação das visualizações.

Para as visualizações foi usado um notebook com sistema *Windows*, onde todos os códigos foram programados no *Software R*. Dentre as visualizações estão:

- *Boxplot*;
- *Scatter Plot*;
- *Hexbin Plot*;
- *Heat Map*;
- *Beanplot*;
- *Forest Plot*;
- Gráfico de coluna e de barra;
- Histograma.
- Gráfico de Linha;
- Gráfico de Área;
- Gráfico de Setor ou de Pizza.

Estas visualizações são as mais comumente utilizadas em artigos, trabalhos, palestras e ferramentas de apoio para maior credibilidade e simplificação da visualização dos dados. Após cada figura apresentada, será inserida a linha de código. Nestas linhas de códigos as palavras em cor azul são explicações necessárias para o entendimento, não devem ser utilizadas. Os dados utilizados para representar notas, IMC e os lucros das empresas foram gerados de forma aleatória em um programa de linguagem C, sendo dados fictícios (A programação está contida no APÊNDICE). A área das regiões brasileiras utilizadas são dados da divisão estabelecida do ano de 1970 e a população foram retirados do censo de

2010 do Instituto Brasileiro de Geografia e estatística e população dos países foram retirados da base de dados da Divisão Estatística das Nações Unidas de 2007. Os valores das variáveis são representados na Tabela 1 e o exemplo em que elas foram utilizadas é identificado na Tabela 2.

Tabela 1- Relação de variáveis e seus valores

Variáveis	Valores
Dados 1	64,78,66,82,74,78,86,100,87,73,95,82,89,73,92,85, 80,81,90,78,86,78,100,85,98,75,73,90,86,86,84,86, 76,76,83,100,86,84,85,76,80,92,100,73,87,70,85,7 9,93,82,90,83,81,85,72,81,96,81,85,68,96,86,70,72 ,74,84,99,81,89,71,73,63,100,74,98,78,78,83,96,95 ,94,88,62,91,83,98,93,93,76,94,75,67,95,100,98,71 ,92,72,73,75
Dados 2	62,72,74,78,81,84,86,89,94,98,63,72,74,78,81,84,8 6,90,94,98,64,72,75,78,82,84,86,90,95,99,66,73,75 ,78,82,85,86,90,95,99,66,73,75,78,82,85,86,90,95, 100,67,73,76,79,82,85,86,91,95,100,68,73,76,80,8 3,85,86,92,96,100,70,73,76,80,83,85,87,92,96,100, 70,73,76,81,83,85,87,92,96,100,71,73,78,81,83,85, 88,93,98,100
IMC	22,32,24,38,21,34,26,39,24,38,23,32,24,38,21,34,2 6,30,24,38,24,32,25,38,22,34,26,30,25,39,26,33,25 ,38,22,35,26,30,25,39,26,33,25,38,22,35,26,30,25, 30,27,33,26,39,22,35,26,31,25,30,28,33,26,30,23,2 5,26,32,26,30,20,33,26,30,23,35,29,30,6,32,33,3,2 3,3,24,3,26,3,27,27,7,28,6,29,30,7,32,33,3,24,24,3, 26,27,27,7,28,3,29,3,31,32,8,33,3
Empresa 1	7.5,16.8,20.6,26.4,32.7,18.8,19.9,25.6
Empresa 2	33.3,32.1,36.6,28.4,29.3,29.4,24.9,35.1
Empresa 3	2.3,4.7,6.6,12.8,10.3,15.4,15.1,16.3
Empresa 4	7.3,14.7,16.6,12.6,11.5,13.8,15.1,18.3
Área China	9.59
Área Canadá	9.98
Área Rússia	17.09
Área Estados Unidos	9.37
Área Brasil	8.51
Área da região Norte do Brasil	3.85
Área da região Centro-Oeste do Brasil	1.61
Área da região Nordeste do Brasil	1.55
Área da região Sudeste do Brasil	0.92
Área da região Sul do Brasil	0.58
População da região Norte do Brasil	15.86
População da região Centro-Oeste do Brasil	14.05
População da região Nordeste do Brasil	53.08
População da região Sudeste do Brasil	80.36

Fonte: Autor

Tabela 2 – Relação de gráficos e dados aleatórios

Gráfico	Dados1	Dados2	IMC	Empresa1	Empresa 2	Empresa3	Empresa 4
Boxplot	x						
Scatter Plot	x						
Hexbin Plot	x		x				
Heat Map	x						
Beanplot	x	x					
Histograma	x						
Linha várias séries				x	x	x	x
Área uma série				x			
Área Várias Séries				x	x	x	x

Fonte: Autor

4 Resultados

Neste tópico serão abordados os resultados obtidos nas programações das visualizações de dados. Serão explicados sobre as estruturas de cada resultado obtido.

4.1 BoxPlot

O *boxplot* é formado pelo segundo quartil (linha da caixa), que é a mediana, e sua caixa é formada pelo primeiro e terceiro quartil, os limites superiores são ligados à caixa por linhas tracejadas, para valores maiores que os limites são considerados *outliers*. Um exemplo de *boxplot* é representado na Figura 2.

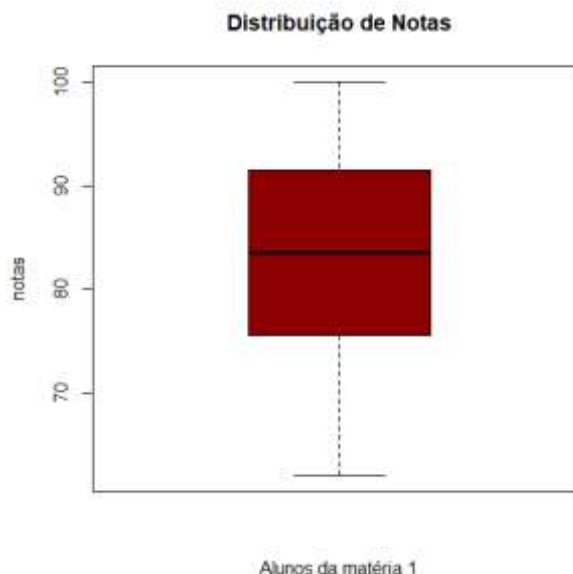


Figura 2 – Exemplo de gráfico *Boxplot*

Para criar o *Boxplot* do exemplo foram utilizadas as linhas de código abaixo:

Alocando valores à variável

`dados1<-`

```
c(64,78,66,82,74,78,86,100,87,73,95,82,89,73,92,85,80,81,90,78,86,78,100,85,98,75,73,90,86,86,84,86,76,76,83,100,86,84,85,76,80,92,100,73,87,70,85,79,93,82,90,83,81,85,72,81,96,81,85,68,96,86,70,72,74,84,99,81,89,71,73,63,100,74,98,78,78,83,96,95,94,88,62,91,83,98,93,93,76,94,75,67,95,100,98,71,92,72,73,75)
```

Definindo características para o gráfico, deve-se primeiro passar o argumento de valor, segundo é a cor da caixa, terceiro argumento o main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x e sub é o subtítulo do gráfico.

```
boxplot(dados1,col="darkred",main="Distribuição de Notas",ylab="notas",
xlab="Alunos da Matéria 1")
```

4.2 Scatter Plot

O gráfico de dispersão faz a correlação entre duas variáveis. É composto por dois eixos, eixo das ordenadas (y) e eixo das abscissa (x). Dentro de seus limites são plotados os pontos que se correlacionam. Um exemplo de *Scatter plot* é representado na Figura 3.

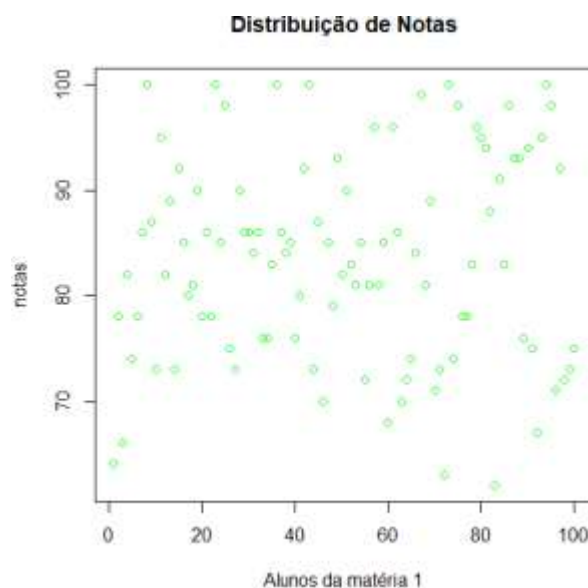


Figura 3 – exemplo de gráfico Scatter plot

Para criar o *Scatter plot* do exemplo foram utilizadas as linhas de código abaixo:

Alocando valores à variável

`dados1<-`

`c(64,78,66,82,74,78,86,100,87,73,95,82,89,73,92,85,80,81,90,78,86,78,100,85,98,75,73,90,86,86,84,86,76,76,83,100,86,84,85,76,80,92,100,73,87,70,85,79,93,82,90,83,81,85,72,81,96,81,85,68,96,86,70,72,74,84,99,81,89,71,73,63,100,74,98,78,78,83,96,95,94,88,62,91,83,98,93,93,76,94,75,67,95,100,98,71,92,72,73,75)`

Definindo características para o gráfico, deve-se primeiro passar o argumento de valor, segundo é a cor da caixa, terceiro argumento o main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x e sub é o subtítulo do gráfico.

`plot(dados1,col="green",main="Distribuição de Notas",ylab="notas", xlab="classe1")`

4.3 Hexbin Plot

O *hexbin plot* é delimitado pelo plano xy, dentro das delimitações é composto por hexágonos, que representam o ponto de relação entre as variáveis. A cor da plotagem é proporcional ao número de repetições contida naquela área. Um exemplo de *Hexbin plot* é representado na Figura 4.

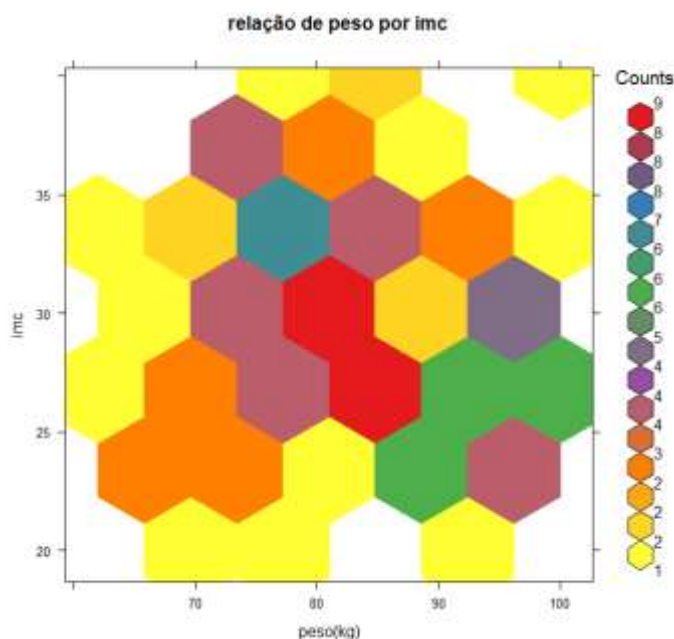


Figura 4 – exemplo de gráfico Hexbin Plot

Para criar o *hexbin* do exemplo foram utilizadas as linhas de códigos abaixo:

Biblioteca do hexbin

```
library(hexbin)
```

Biblioteca para a paleta de cores

```
library(RColorBrewer)
```

Alocando valores à variável

```
dados1<-
```

```
c(64,78,66,82,74,78,86,100,87,73,95,82,89,73,92,85,80,81,90,78,86,78,100,85,98,75,73,90,86,86,84,86,76,76,83,100,86,84,85,76,80,92,100,73,87,70,85,79,93,82,90,83,81,85,72,81,96,81,85,68,96,86,70,72,74,84,99,81,89,71,73,63,100,74,98,78,78,83,96,95,94,88,62,91,83,98,93,93,76,94,75,67,95,100,98,71,92,72,73,75)
```

Alocando valores para a segunda variável

```
dados2<-
```

```
c(22,32,24,38,21,34,26,39,24,38,23,32,24,38,21,34,26,30,24,38,24,32,25,38,22,34,26,30,25,39,26,33,25,38,22,35,26,30,25,39,26,33,25,38,22,35,26,30,25,30,27,33,26,39,22,35,26,31,25,30,28,33,26,30,23,25,26,32,26,30,20,33,26,30,23,35,29,30,6,32,33,3,23,3,24,3,26,3,27,27,7,28,6,29,30,7,32,33,3,24,24,3,26,27,27,7,28,3,29,3,31,32,8,33,3)
```

Definindo qual o set de cores que será utilizado

```
rf <- colorRampPalette(rev(brewer.pal(6,'Set1')))
```

definindo características do Hexbin Plot deve-se primeiro passar os argumentos de

valores, segundo é a cor dos hexagonos., terceiro argumento o main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x e sub é o subtítulo do gráfico.

```
hexbinplot(dados2~dados1, colramp=rf,xbins=5,main="relação de peso por imc",
ylab ="imc", xlab="peso(kg)")
```

4.4 Heatmap

O *Heatmap* é uma imagem que possui uma cor falsa, esta cor representa a intensidade proporcional do valor daquele determinado ponto, ele utiliza os valores extremos como comparações, o maior valor é a cor branca e o menor valor é a cor vermelha escuro, os valores centrais são proporcionais possuem cores de acordo com sua intensidade. Possui um dendrograma em cima e do lado esquerdo. Um exemplo de *Heatmap* é representado na Figura 5.

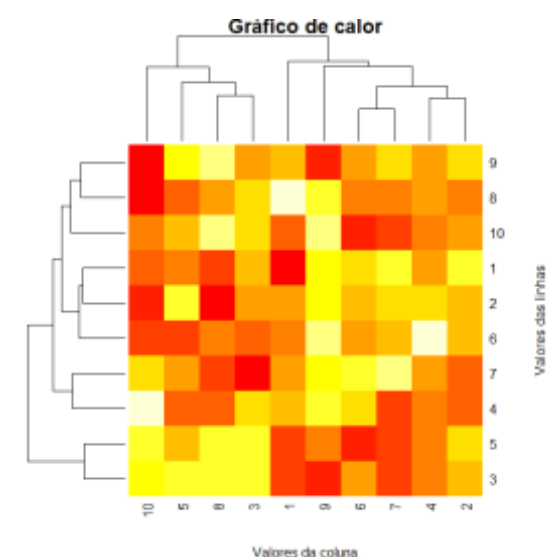


Figura 5 – Exemplo de gráfico Heat Map.

Para criar o *Heat Map* do exemplo foram utilizadas as linhas de código abaixo:

Alocando valores à variável

```
dados1<-
```

```
c(64,78,66,82,74,78,86,100,87,73,95,82,89,73,92,85,80,81,90,78,86,78,100,85,98,7
5,73,90,86,86,84,86,76,76,83,100,86,84,85,76,80,92,100,73,87,70,85,79,93,82,90,8
```

```
3,81,85,72,81,96,81,85,68,96,86,70,72,74,84,99,81,89,71,73,63,100,74,98,78,78,83,
96,95,94,88,62,91,83,98,93,93,76,94,75,67,95,100,98,71,92,72,73,75)
```

Transformando vetor em matriz, primeiro passa o argumento que se deseja transformar em matriz, nrow indica o número de linhas da matriz que deseja-se criar e ncol indica o número de colunas que deseja-se criar

```
minha.matriz <- matrix(data=dados1,nrow=10,ncol=10)
```

Plotando o heat map, main é o título do gráfico, ylab é o título do eixo y e xlab é o título do eixo x

```
heatmap(minha.matriz,main="Gráfico de calor",xlab="Valores da coluna",
ylab="Valores das linhas")
```

4.5 Beanplot

As linhas são a representação de cada dado, quando maior a largura da linha indica que existe mais de um dado, sendo possível identificar anomalias. A linha grossa representa a média aritmética. Um exemplo de *Beanplot* é representado na Figura 6.

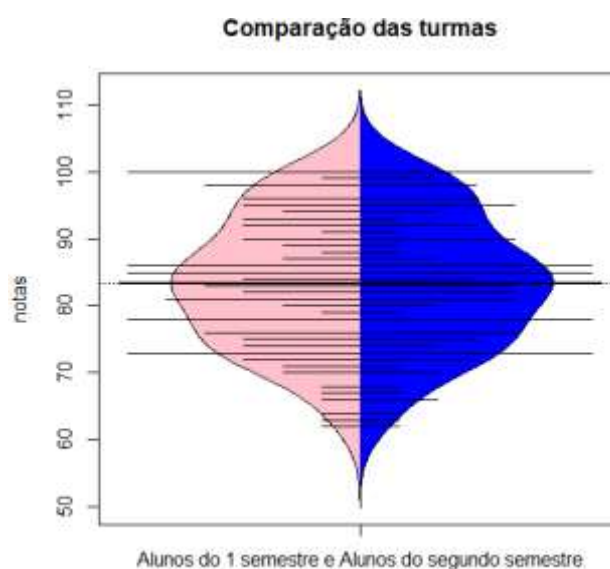


Figura 6 – Exemplo de beanplot

Para criar o *Beanplot* do exemplo foram utilizadas as linhas de código abaixo:

Carregando biblioteca

```
library(beanplot)
```

Alocando dados pra primeira variável

```
dados1<-
```

```
c(64,78,66,82,74,78,86,100,87,73,95,82,89,73,92,85,80,81,90,78,86,78,100,85,98,75,73,90,86,86,84,86,76,76,83,100,86,84,85,76,80,92,100,73,87,70,85,79,93,82,90,83,81,85,72,81,96,81,85,68,96,86,70,72,74,84,99,81,89,71,73,63,100,74,98,78,78,83,96,95,94,88,62,91,83,98,93,93,76,94,75,67,95,100,98,71,92,72,73,75)
```

Alocando dados para segunda variável

```
dados2<-
```

```
c(62,72,74,78,81,84,86,89,94,98,63,72,74,78,81,84,86,90,94,98,64,72,75,78,82,84,86,90,95,99,66,73,75,78,82,85,86,90,95,99,66,73,75,78,82,85,86,90,95,100,67,73,76,79,82,85,86,91,95,100,68,73,76,80,83,85,86,92,96,100,70,73,76,80,83,85,87,92,96,100,70,73,76,81,83,85,87,92,96,100,71,73,78,81,83,85,88,93,98,100)
```

Características do Beanplot, o primeiro e segundo argumentos são passados os argumentos para a plotagem, side indica qual lado será plotado os argumentos (quando apenas um argumetno deve-se utilizar “f” para plotar do lado direito e “s” para o lado esquerdo, se dois argumentos utiliza-se “b”, ambos os lados), col deve ser passado em lista para indicar cada cor de cada gráfico, names é o subtítulo do gráfico, ylab é o texto no eixo y e main é o título do gráfico.

```
beanplot(dados1,dados2, side = "b",col = list("pink", "blue"),names=c("Alunos do 1 semestre e Alunos do segundo semestre"),ylab="notas", main="Comparação das turmas")
```

4.6 Forest Plot

No *Forest Plot* a primeira coluna é reservada para os estudos que foram utilizados, a segunda coluna neste método é reservada para o grupo experimental, a terceira coluna são listados o grupo controle e a quarta coluna é designada para as medidas de efeitos com seus respectivos intervalo de confiança (linha horizontal). A dimensão de cada quadrado é proporcional ao peso que cada estudo possui, quando maior o quadrado maior o peso do estudo. Um exemplo de *Forest plot* é representado na Figura 7.

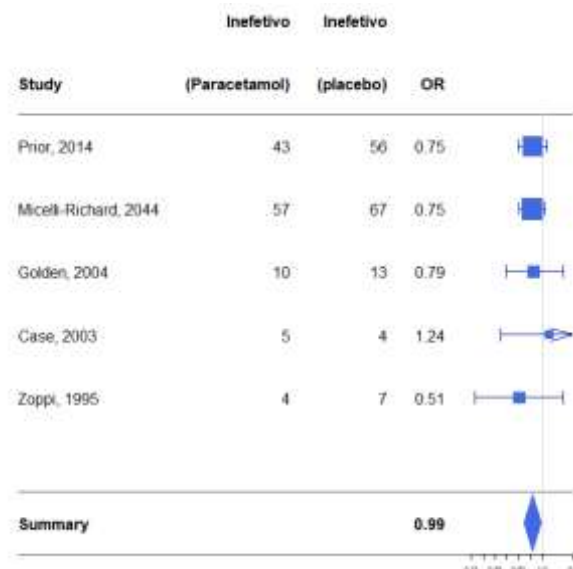


Figura 7 – exemplo de *Forest Plot*.

Para criar o *forest plot* do exemplo foram utilizadas as linhas de código abaixo:

Carregamento da biblioteca

```
library(forestplot)
```

```
# Cochrane data from the 'rmeta'-package
```

```
cochrane_from_rmeta <-
```

```
  structure(list(
```

valores da razão de chance de peto

```
    mean = c(NA, NA, 0.75, 0.75, 0.79, 1.24, 0.51, NA, 0.76),
```

Limite inferior

```
    lower = c(NA, NA, 0.49, 0.51, 0.34, 0.3, 0.14, NA, 0.59)
```

Limite Superior

```
    upper = c(NA, NA, 1.16, 1.1, 1.85, 5.1, 1.9, NA, 0.99)),
```

```
    .Names = c("mean", "lower", "upper"),
```

Em `row.names` o segundo argumento é o numero de linhas utilizados acima

```
    row.names = c(NA, -9L),
```

```
    class = "data.frame")
```

```
tabletext<-cbind(
```

Primeira coluna (coluna dos estudos)

```
  c("", "Study", "Prior, 2014", "Micelli-Richard, 2044",
```

```
    "Golden, 2004", "Case, 2003", "Zoppi, 1995", NA, "Summary"),
```

Segunda coluna (coluna destinada ao grupo experimental)

```
c("Inefetivo", "(Paracetamol)", "43", "57",  
  "10", "5", "4", NA, NA),
```

Terceira coluna (coluna designada ao grupo controle)

```
c("Inefetivo", "(placebo)", "56", "67",  
  "13", "4", "7", NA, NA),
```

Quarta coluna (coluna reservada à medidas de efeito)

```
c("", "OR", "0.75", "0.75",  
  "0.79", "1.24", "0.51", NA, "0.99"))
```

```
forestplot(tabletext,
```

Adicionando linhas horizontais

```
hrzl_lines = gpar(col="#444444"),  
cochrane_from_rmeta,new_page = TRUE,
```

No summary indica o número de linhas entre o study e summary do gráfico

```
is.summary=c(TRUE,TRUE,rep(FALSE,6),TRUE),  
clip=c(0.1,2.5),  
xlog=TRUE,
```

Definindo características do gráfico box define a cor das caixas do gráfico, line define as cores das linhas que passam pelas as caixas, summary define a cor do losango do summary e vertices define a existência de vertices nas linhas, se TRUE são acrescentados vértices se false ou não especificado não há acrescento de vértices nas linhas

```
col=fpColors(box="royalblue",line="darkblue", summary="royalblue"),  
vertices = TRUE)
```

4.7 Gráfico de coluna e de barras

O gráfico de coluna e o de barras apresentam barras retangulares, que possuem comprimento proporcional ao valor da variável que ele representa, um eixo demonstra o que está sendo comparado enquanto o outro apresenta os valores das variáveis. Nas categorias empilhadas e lado a lado demonstram a contribuição de diferentes fatores, sendo que na empilhada demonstra o valor total e no lado a lado

demonstra os valores de cada fator. Exemplos de gráficos de colunas e de barras são representados da Figura 8 a Figura 13.

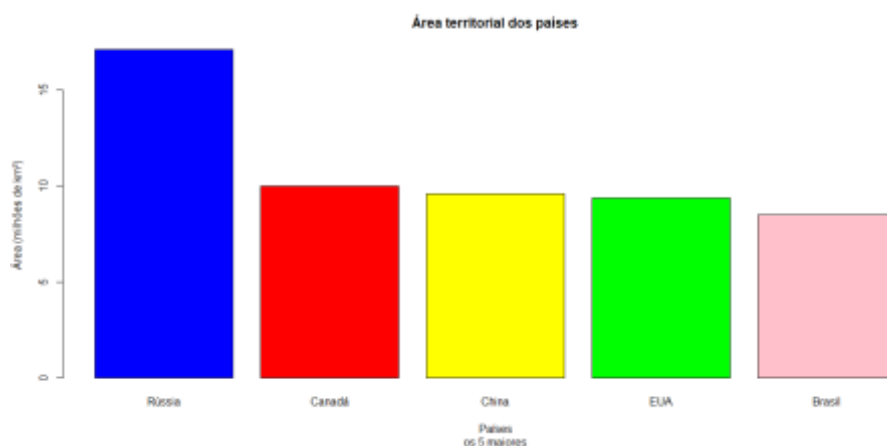


Figura 8 – Exemplo de gráfico de coluna

Para criar o exemplo do gráfico de coluna foram utilizadas as linhas de código abaixo:

Alocando dados para a primeira variável

```
pais<-c("Rússia","Canadá","China","EUA", "Brasil")
```

Alocando dados para a segunda variável

```
area <- c(17.09,9.98,9.59,9.37,8.51)
```

Definindo características para o gráfico, deve-se primeiro passar o argumento de valor, o segundo argumento é o nome variável que deseja analisar, terceiro argumento o main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x, sub é o subtítulo do gráfico e col define as lista de cores das colunas.

```
barplot(area,names.arg=pais,main="Área territorial dos países",ylab="Área (milhões de km²)",xlab="Países",sub="os 5 maiores", col = c("blue", "red", "yellow", "green", "pink"))
```

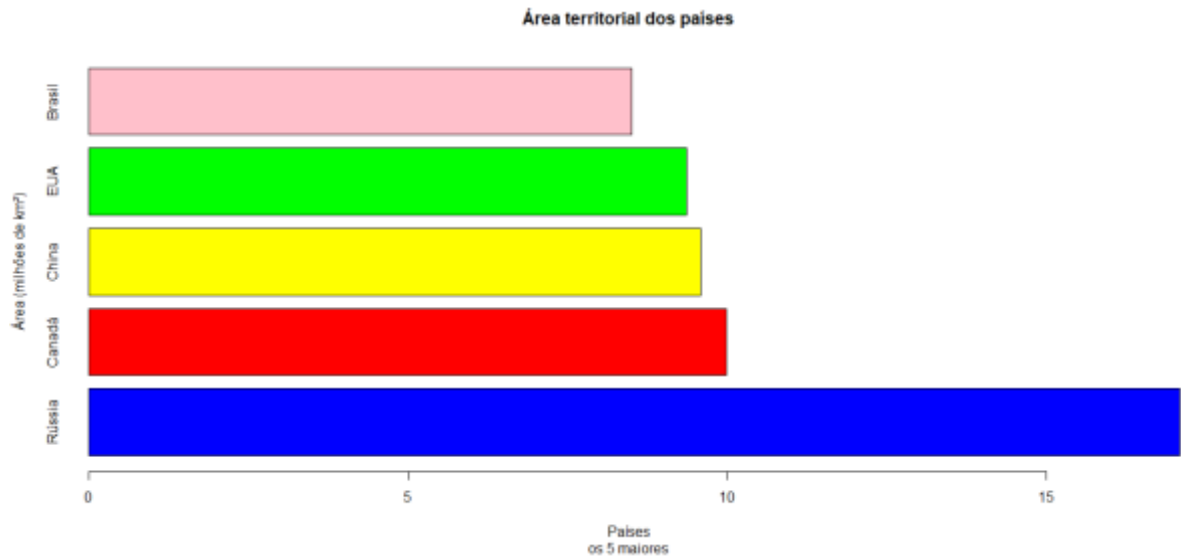


Figura 9 – Exemplo de gráfico de Barras

Para criar o exemplo de gráfico de barras foram utilizadas as linhas de código abaixo:

Alocando dados para a primeira variável

```
pais<-c("Rússia","Canadá","China","EUA", "Brasil")
```

Alocando dados para a segunda variável

```
area <- c(17.09,9.98,9.59,9.37,8.51)
```

Definindo características para o gráfico, deve-se primeiro passar o argumento de valor, o segundo argumento é o nome variável que deseja analisar, terceiro argumento o main é o título do gráfico, xlab é o título do eixo x, ylab é o título do eixo y, sub é o subtítulo do gráfico, horiz define a orientação das barras se TRUE ele será horizontal se FALSE ou não argumentado ele será vertical e col define as lista de cores das colunas.

```
barplot(area,names.arg=pais,main="Área territorial dos países",ylab="Área (milhões de km²)",xlab="Países",sub="os 5 maiores", horiz = TRUE, col = c("blue", "red", "yellow", "green", "pink"))
```

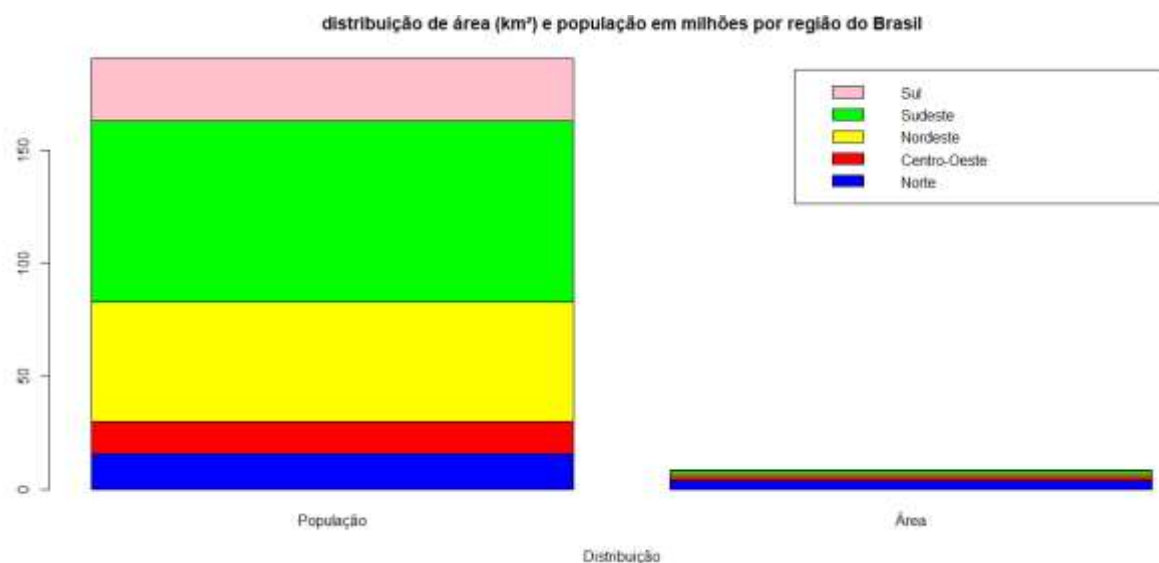


Figura 10 – Exemplo de gráfico de coluna Empilhado

Para criar o exemplo do gráfico de coluna empilhado foram utilizadas as linhas de código abaixo:

Alocando dados, é definido uma matriz, lembre-se que deve-se colocar os valores em ordem (sempre colocando os valores da primeira coluna, depois os da segunda coluna, assim por diante, sempre em sequencia), nrow define o número de linhas da matriz, ncol define o número de colunas, dimnames define as legendas e os subtítulos dos dados, o primeiro vetor é destinado à legenda e o segundo ao subtítulo

```
dados <- matrix(c(3.85,1.61,1.55,0.92,0.58,15.86,14.05,53.08,80.36,27.39), nrow=5,
ncol=2,dimnames=list(c("Norte","Centro-
Oeste","Nordeste","Sudeste","Sul"),c("Área","População")))
```

Definindo características para o gráfico, deve-se primeiro passar o argumento de valor, o segundo argumento é o nome da variável que se encontra as legendas, terceiro argumento o main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x, sub é o subtítulo do gráfico e col define as lista de cores das colunas.

```
barplot(dados[,2:1],legend.text=rownames(dados),main="distribuição de área (km²)
e população em milhões por região do Brasil",ylab=" ", xlab="Distribuição", sub=" ",
col = c("blue", "red", "yellow", "green", "pink"))
```

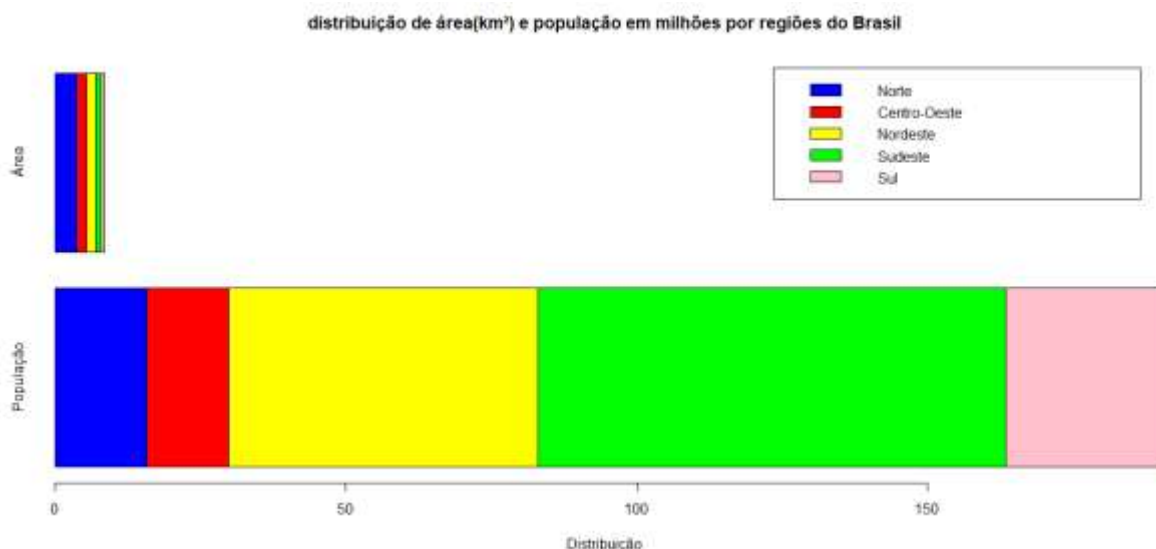


Figura 11 – Exemplo de gráfico de barras empilhado

Para criar o exemplo de gráfico de barras empilhado foram utilizadas as linhas de código abaixo:

Alocando dados, é definido uma matriz, lembre-se que deve-se colocar os valores em ordem (sempre colocando os valores da primeira coluna, depois os da segunda coluna, assim por diante, sempre em sequencia), nrow define o número de linhas da matriz, ncol define o número de colunas, dimnames define as legendas e os subtítulos dos dados, o primeiro vetor é destinado à legenda e o segundo ao subtítulo

```
dados <- matrix(c(3.85,1.61,1.55,0.92,0.58,15.86,14.05,53.08,80.36,27.39), nrow=5,
ncol=2,dimnames=list(c("Norte","Centro-
Oeste","Nordeste","Sudeste","Sul"),c("Área","População")))
```

Definindo características para o gráfico, deve-se primeiro passar o argumento de valor, o segundo argumento é o nome da variável que se encontra as legendas, terceiro argumento o main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x, sub é o subtítulo do gráfico, horiz define a orientação das barras se TRUE ele será horizontal se FALSE ou não argumentado ele será vertical e col é vetor que define as cores das colunas.

```
barplot(dados[,2:1],legend.text=rownames(dados),main="distribuição de área(km²) e
população em milhões por regiões do Brasil",ylab=" ", xlab="Distribuição", sub=" ",
horiz=TRUE, col=c("blue","red","yellow","green","pink"))
```



Figura 12 – exemplo de gráfico de colunas lado a lado

Para criar o exemplo do gráfico de colunas lado a lado foram utilizadas as linhas de código abaixo:

Alocando dados, é definido uma matriz, lembre-se que deve-se colocar os valores em ordem (sempre colocando os valores da primeira coluna, depois os da segunda coluna, assim por diante, sempre em sequencia), nrow define o número de linhas da matriz, ncol define o número de colunas, dimnames define as legendas e os subtítulos dos dados, o primeiro vetor é destinado à legenda e o segundo ao subtítulo

```
dados <- matrix(c(3.85,1.61,1.55,0.92,0.58,15.86,14.05,53.08,80.36,27.39), nrow=5,
ncol=2,dimnames=list(c("Norte","Centro-
Oeste","Nordeste","Sudeste","Sul"),c("Área","População")))
```

Definindo características para o gráfico, deve-se primeiro passar o argumento de valor, o segundo argumento é o nome da variável que se encontra as legendas, terceiro argumento o main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x, sub é o subtítulo do gráfico e beside define se as colunas ficarão lado a lado se TRUE elas ficarão lado a lado se FALSE ou não argumentado ele será empilhado e col é o vetor que define as cores das barras.

```
barplot(dados[,2:1],legend.text=rownames(dados),main="distribuição de área(km²) e
população em milhões por regiões do Brasil",ylab=" ", xlab="Distribuição", sub=" ",
beside=TRUE, col=c("blue","red","yellow","green","pink"))
```

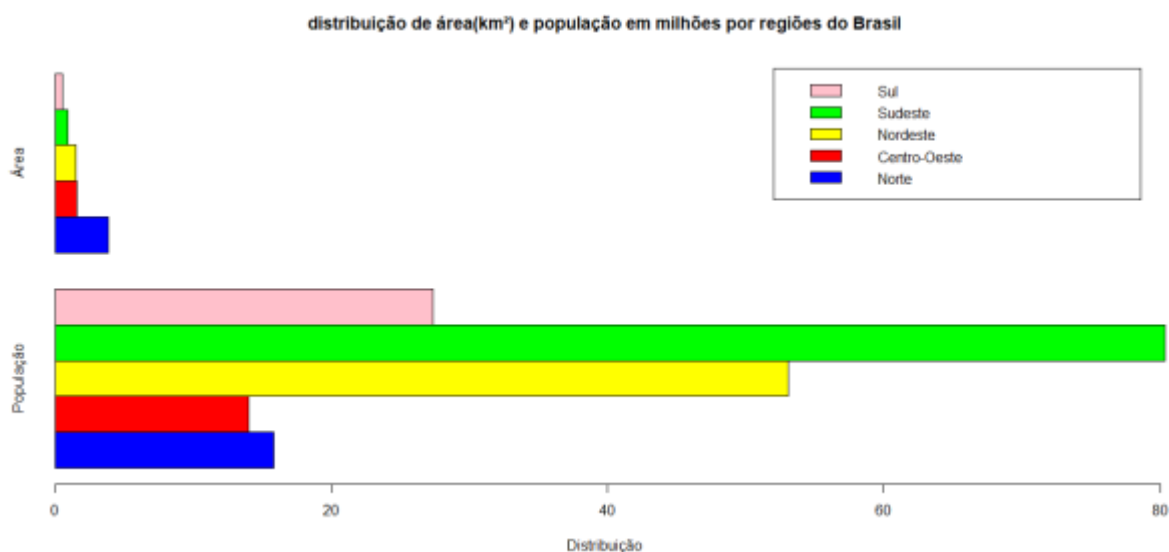


Figura 13 – exemplo de gráfico de barras lado a lado

Para criar o exemplo de gráfico de barras lado a lado foram utilizadas as linhas de código abaixo:

Alocando dados, é definido uma matriz, lembre-se que deve-se colocar os valores em ordem (sempre colocando os valores da primeira coluna, depois os da segunda coluna, assim por diante, sempre em sequencia), nrow define o número de linhas da matriz, ncol define o número de colunas, dimnames define as legendas e os subtítulos dos dados, o primeiro vetor é destinado à legenda e o segundo ao subtítulo

```
dados <- matrix(c(3.85,1.61,1.55,0.92,0.58,15.86,14.05,53.08,80.36,27.39), nrow=5,
ncol=2,dimnames=list(c("Norte","Centro-
Oeste","Nordeste","Sudeste","Sul"),c("Área","População")))
```

Definindo características para o gráfico, deve-se primeiro passar o argumento de valor, o segundo argumento é o nome da variável que se encontra as legendas, terceiro argumento o main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x, sub é o subtítulo do gráfico, horiz define a orientação das barras se TRUE ele será horizontal se FALSE ou não argumentado ele será vertical, beside define se as colunas ficarão lado a lado se TRUE elas ficarão lado a lado se FALSE ou não argumentado ele será empilhado e col é o vetor que define as cores das barras.

```
barplot(dados[,2:1],legend.text=rownames(dados),main="distribuição de área(km²) e
população em milhões por regiões do Brasil",ylab=" ", xlab="Distribuição", sub="
",horiz=TRUE, beside=TRUE, col=c("blue","red","yellow","green","pink"))
```

4.8 Histograma

O histograma consiste de uma sucessão de colunas adjacentes, no eixo das abscissas é demarcado o intervalo das classes e no eixo das ordenadas é composto pela a frequência de repetições do valor. Um exemplo de Histograma é representado na Figura 14.

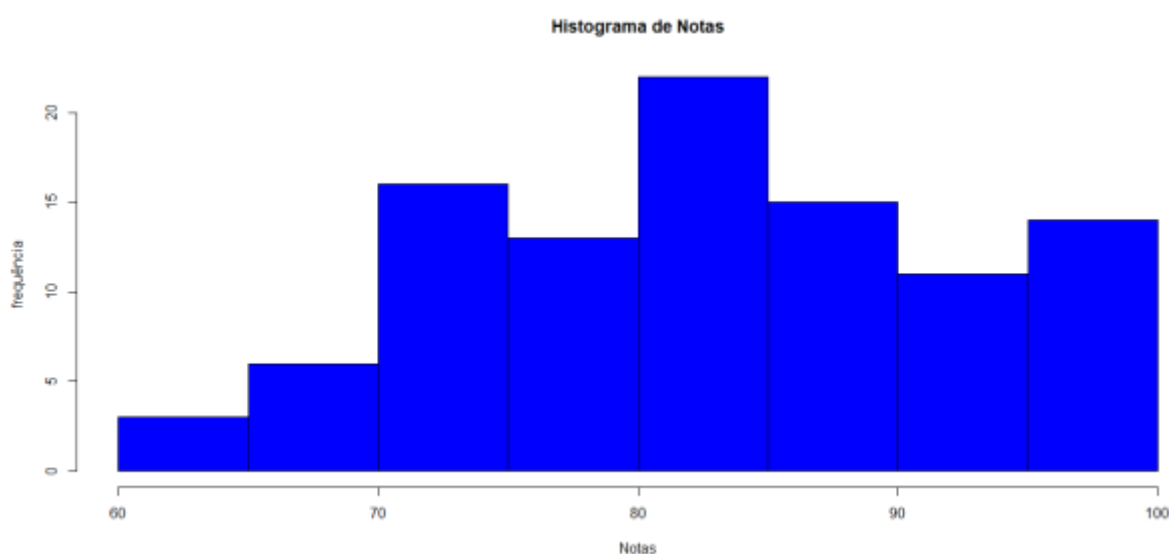


Figura 14 – Exemplo de Histograma

Para criar o exemplo do histograma foram utilizadas as linhas de código abaixo:

Alocando dados, é definido um vetor

`dados<-`

```
c(62,72,74,78,81,84,86,89,94,98,63,72,74,78,81,84,86,90,94,98,64,72,75,78,82,84,86,90,95,99,66,73,75,78,82,85,86,90,95,99,66,73,75,78,82,85,86,90,95,100,67,73,76,79,82,85,86,91,95,100,68,73,76,80,83,85,86,92,96,100,70,73,76,80,83,85,87,92,96,100,70,73,76,81,83,85,87,92,96,100,71,73,78,81,83,85,88,93,98,100)
```

Definindo características para o gráfico, deve-se primeiro passar o argumento de valor, col define a cor das colunas, terceiro argumento o main é o título do gráfico, xlab é o título do eixo x, ylab é o título do eixo y.

```
Hist(dados,col=("blue"),main="Histograma de notas, xlab="Notas", ylab="Frequência)
```

4.9 Gráfico de Linha

Um gráfico de linha é delimitado pelo eixo x, na horizontal, e pelo eixo y, na vertical, dentro desses limites podem haver uma ou mais séries de dados. Exemplos de Gráfico de linha são representado na Figura 15 e Figura 16.

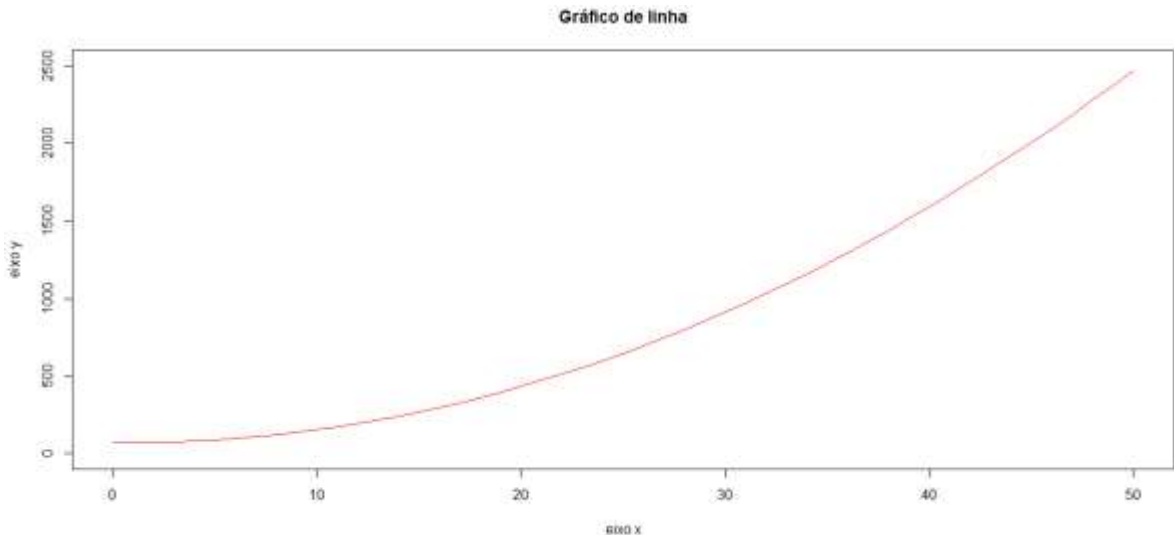


Figura 15 – Exemplo de Gráfico de linha para uma série

Para criar o exemplo do gráfico de linha foram utilizadas as linhas de código abaixo:

Definindo a função

```
x<-0:50
```

```
> y<-(x**2)-(2*x)+72
```

Definindo características para o gráfico, deve-se primeiro passar os argumento de valores, depois definir o intervalo do eixo e depois intervalo do eixo y, type indica que a série será do tipo linha, main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x e col define a cor da linha

```
plot(x,y,c(0,50),c(0,2500),type='l',main="Gráfico de linha",ylab="eixo y",xlab="eixo x",col="red")
```

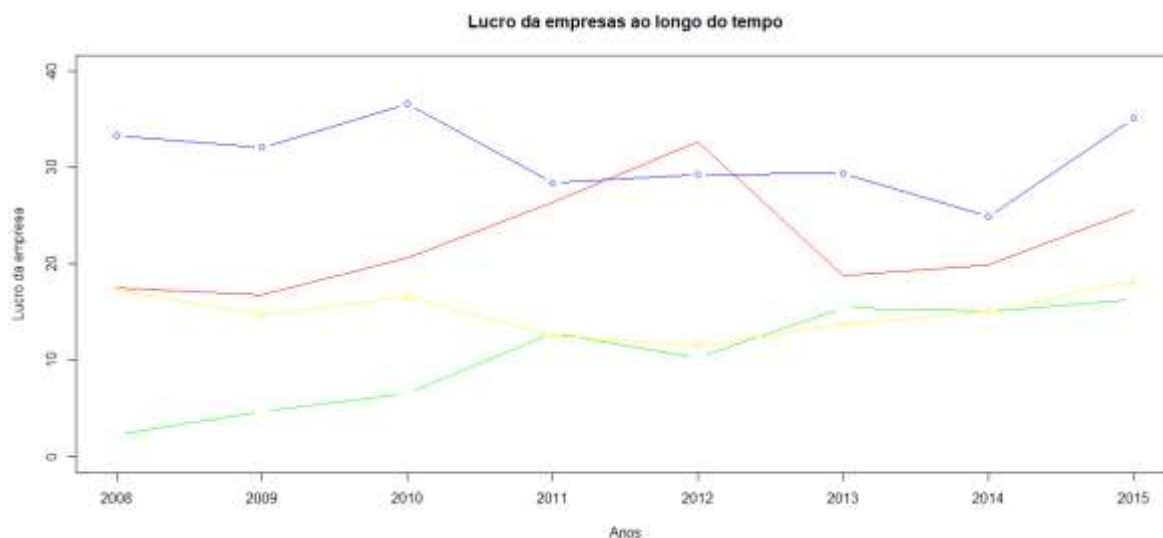


Figura 16 – Exemplo de Gráfico de linha para várias séries

Para criar o exemplo do gráfico de linha para várias séries foram utilizadas as linhas de códigos abaixo, as palavras que possuem a cor azul são explicações necessárias para o entendimento, não devem ser utilizadas nas linhas de código:

Definindo as séries

```
empresa1<-c(17.5,16.8,20.6,26.4,32.7,18.8,19.9,25.6)
```

```
empresa2<-c(33.3,32.1,36.6,28.4,29.3,29.4,24.9,35.1)
```

```
empresa3<-c(2.3,4.7,6.6,12.8,10.3,15.4,15.1,16.3)
```

```
empresa4<-c(17.3,14.7,16.6,12.6,11.5,13.8,15.1,18.3)
```

```
ano<-2008:2015
```

Definindo características para o gráfico, deve-se primeiro passar os argumento de valores, depois definir o intervalo do eixo e depois intervalo do eixo y, type indica que a série será do tipo linha, main é o título do gráfico, ylab é o título do eixo y, xlab é o título do eixo x e col define a cor da linha

```
plot(ano,empresa1,c(2008,2015),c(0,40),type="l",main="Lucro das empresas ao longo do tempo",ylab="anos",xlab="Lucro da empresa",col="red")
```

Adicionando segundo traçado, type “b” indica um traçado com pontos e linhas

```
lines(ano,empresa2,col="blue",type="b")
```

Adicionando terceiro traçado, type “c” indica um traçado para linhas descontínuas nos pontos

```
lines(ano,empresa3,col="green",type="c")
```

Adicionando o quarto traçado, type “o” indica um traçado para pontos sobre a linha

```
lines(ano,empresa4,col="yellow",type="o")
```

4.10 Gráfico de Área

O gráfico de área se assemelha muito com o gráfico de linha, mas sua principal diferença é que a área abaixo desta linha é preenchida, amplamente utilizada para demonstrar a contribuição de um fator durante um período de tempo, já na categoria empilhado, é representado a contribuição de n fatores para uma determinada variável. Exemplos de Gráfico de Área são representados na Figura 17 e Figura 18.

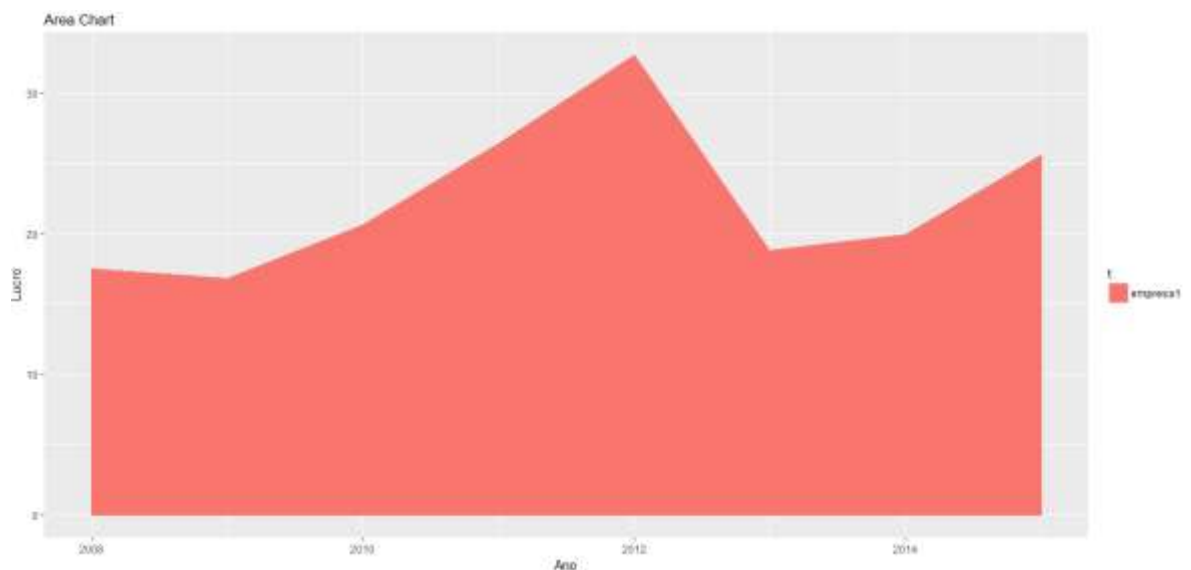


Figura 17 – Exemplo de Gráfico de área para uma série

Para criar o exemplo do gráfico de área para uma série foram utilizadas as linhas de código abaixo:

Definindo a série

```
empresa1<-c(17.5,16.8,20.6,26.4,32.7,18.8,19.9,25.6)
```

Definindo valores para a matriz que será usada para a plotagem, primeiro argumento é o tipo de cada variável, o segundo argumento é o argumento do eixo x e o terceiro é o argumento do eixo y

```
d<-data.frame(t="empresa1",var=2008:2015,val=c(empresa1))
```

Passando os valores que deverão ser plotados, o primeiro é onde se encontra os dados e dentro de aes o que deve ser destinado ao eixo x e depois ao eixo y

```
p<-ggplot(d,aes(var,val))
```

Definindo a área de plotagem, em aes defini-se como deverão ser separados as séries, e em position stack demonstra que será empilhado, labs define o título em

title, em x o texto do eixo x e em y o texto eixo y

```
p+geom_area(aes(colour=t,fill=t),position='stack')+labs(title = "Area Chart", x = "Ano", y = "Lucro")
```

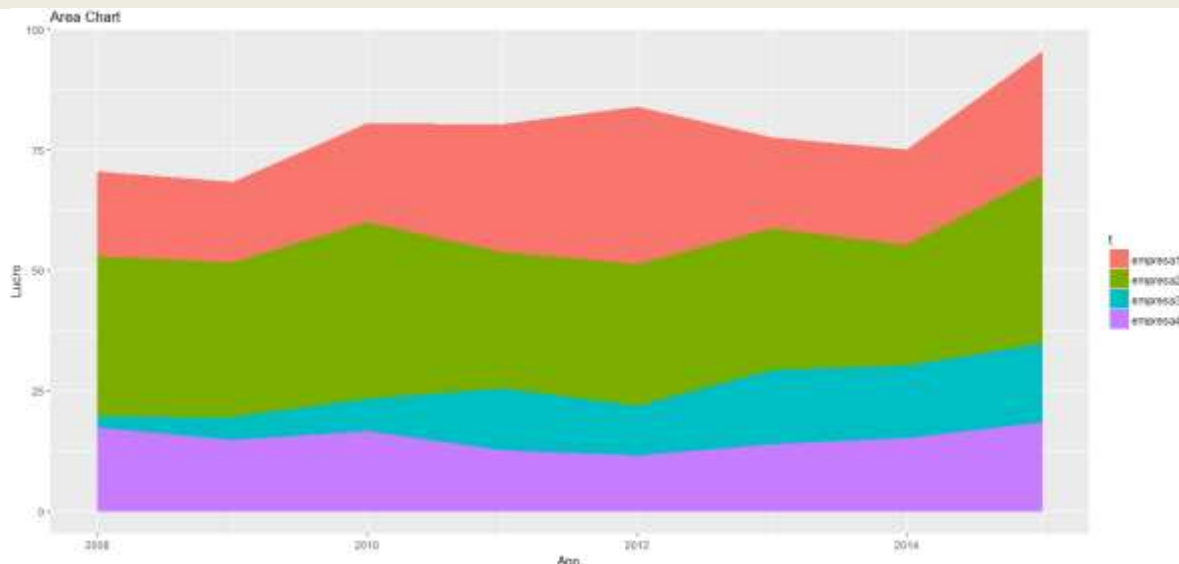


Figura 18 – Exemplo de Gráfico de área para várias séries

Para criar o exemplo do gráfico de área para várias séries foram utilizadas as linhas de código abaixo:

Definindo as séries

```
empresa1<-c(17.5,16.8,20.6,26.4,32.7,18.8,19.9,25.6)
```

```
empresa2<-c(33.3,32.1,36.6,28.4,29.3,29.4,24.9,35.1)
```

```
empresa3<-c(2.3,4.7,6.6,12.8,10.3,15.4,15.1,16.3)
```

```
empresa4<-c(17.3,14.7,16.6,12.6,11.5,13.8,15.1,18.3)
```

Definindo valores para a matriz que será usada para a plotagem, primeiro argumento é o tipo de cada variável, each indica quantas vezes a sequência do tipo da variável deve ser repetida o segundo argumento é o argumento do eixo x e o terceiro é o argumento do eixo y

```
d<-
```

```
data.frame(t=rep(c("empresa1","empresa2","empresa3","empresa4"),each=8),var=rep(2008:2015,2),val=c(empresa1,empresa2,empresa3,empresa4))
```

Passando os valores que deverão ser plotados, o primeiro é onde se encontra os dados e dentro de aes o que deve ser destinado ao eixo x e depois ao eixo y

```
p<-ggplot(d,aes(var,val))
```

Definindo a área de plotagem, em aes defini-se como deverão ser separados as séries, e em position stack demonstra que será empilhado, labs define o título em

title, em x o texto do eixo x e em y o texto eixo y

```
p+geom_area(aes(colour=t,fill=t),position='stack')+labs(title = "Area Chart", x = "Ano",
y = "Lucro")
```

4.11 Gráfico de setor ou gráfico de pizza

Gráfico de setor ou gráfico de pizza é uma visualização circular. Os valores (“fatias” ou setores) são proporcionais à sua contribuição ao total, podendo ser representado em porcentagem a sua contribuição. Exemplos de Gráfico de setor são representados na Figura 19 e Figura 20.



Figura 19 – Exemplo de gráfico de setor

Para criar o exemplo do gráfico de setor foram utilizadas as linhas de código abaixo:

Definindo a série

```
população<-c(15.86,14.05,53.08,80.36,27.39)
```

Definindo os nomes de cada setor do gráfico em sequência

```
names(população)<-c("Norte","Centro-Oeste","Nordeste","Sudeste","Sul")
```

Definindo características do gráfico, primeiro passa-se o argumento de valor para o gráfico, main é o título do gráfico e col define as cores utilizadas

```
pie(população,main="População por região do Brasil",col=rainbow(5))
```



Figura 20 – Exemplo de gráfico de setor em porcentagem

Para criar o exemplo do gráfico de setor em porcentagem foram utilizadas as linhas de código abaixo:

Definindo a série

```
população<-c(15.86,14.05,53.08,80.36,27.39)
```

Definindo os nomes de cada setor do gráfico em sequência

```
names(população)<-c("Norte","Centro-Oeste","Nordeste","Sudeste","Sul")
```

Cálculo da porcentagem, o argumento passado "2" define que será truncado em 2 casa decimais

```
porc<-round(população*100/sum(população),2)
```

Definindo como será escrito no setor

```
legenda<-paste("(",porc,"%)",sep="")
```

Definindo características do gráfico, primeiro passa-se o argumento de valor para o gráfico, main é o título do gráfico, label passa-se o argumento definido acima e col define as cores utilizadas

```
pie(população, main="População por região do Brasil",labels=legenda, col=rainbow(5))
```

Definindo características da legenda, 1,1 indica que será plotado legenda em 1 linha e em 1 coluna, o segundo argumento é o texto que deverá estar contido na legenda, col define as cores das legendas de cada variável e pch monta as bolinhas para a legenda de cada setor

```
legend(1,1,names(população),col = rainbow(5),pch=rep(20,6))
```

5 CONCLUSÕES

Este trabalho abordou de forma prática e fácil compreensão, a programação de gráficos desenvolvidos no software R, apresentando as linhas de códigos com as suas devidas explicações, este trabalho atingiu o objetivo de servir como um guia prático para que futuros usuários possam realizar, compreender as diferentes formas de visualizações gráficas de dados. Outro aspecto importante demonstrar a utilização e o poder do software R.

Pode-se perceber durante o desenvolvimento do guia que as diferentes técnicas e métodos das visualizações podem ser consideradas como ferramentas de qualidades. A implantação dessas gera resultados de alta relevância, propiciando uma visualização agradável, facilitada a análise e sintetização dos dados e possibilitada uma melhor interpretação e identificação de problemas e suas respectivas soluções.

Através das visualizações é possível fazer especificações facilitada de agrupamentos, permitindo comparações e conclusões dos dados. Eles permitem uma melhor compreensão ao público de elementos não gráficos, gerando interesses e criam credibilidade para os pontos que você deseja evidenciar.

Existem diversas outras formas de visualizações de dados que não foram abordadas neste trabalho, por exemplo, o gráfico de bolhas e gráficos de três dimensões.

6 REFERÊNCIAS

- [1] MONTEIRO, Carlos Eduardo Ferreira. **Interpretação de Gráficos: Atividade social e conteúdo de ensino**. ANPED, 22^a, 1999.
- [2] MANSSOUR, Isabel Harb. **Visualização colaborativa de dados científicos com ênfase na área médica**. Porto Alegre: CPGCC, UFRGS. Exame de qualificação, 1998.
- [3] TRAINA, Agma Juci Machado et al. **Visualização de Dados em Sistemas de Bases de Dados Relacionais**. In: SBBD. 2001. p. 95-109.
- [4] DE SOUZA, Emanuel Fernando Maia; PETERNELLI, Luiz Alexandre; DE MELLO, Márcio Pupin. **Software Livre R: aplicação estatística**. 2014
- [5] DA SILVA, Bruno Fontana; DINIZ, Jean; BORTOLUZZI, Matias Américo. **Minicurso de Estatística Básica: Introdução ao software R**. 2009.
- [6] DO AMARAL, Marcelo Rubens dos Santos et al. **Apostila do Curso de Extensão: Software Estatístico Livre R**. 2010.
- [7] ALEXANDRE, Dulclerci Sternadt; TAVARES, João Manuel Ribeiro da Silva. **Factores da percepção visual humana na visualização de dados**. In: CMNE 2007-Congresso de Métodos Numéricos em Engenharia, XXVIII CILAMCE-Congresso Ibero Latino-Americano sobre Métodos Computacionais em Engenharia, Porto, PT. 2007.
- [8] HUBERT, Mia; VANDERVIEREN, Ellen. **An adjusted boxplot for skewed distributions**. Computational statistics & data analysis, v. 52, n. 12, p. 5186-5201, 2008.
- [9] CAPELA, Marisa Veiga; CAPELA, Jorge Manuel Vieira. **Elaboração de gráficos box-plot em planilhas de cálculo**. In: CONGRESSO DE MATEMÁTICA APLICADA E COMPUTACIONAL DA REGIÃO SUDESTE–CNMAC Sudeste. 2011
- [10] BENJAMINI, Yoav. **Opening the Box of a Boxplot**. The American Statistician, v. 42, n. 4, p. 257-262, 1988.
- [11] GOLDBERG, Kenneth Mayer; IGLEWICZ, Boris. **Bivariate extensions of the boxplot**. Technometrics, v. 34, n. 3, p. 307-320, 1992.

- [12] FRIGGE, Michael; HOAGLIN, David Caster; IGLEWICZ, Boris. **Some implementations of the boxplot**. The American Statistician, v. 43, n. 1, p. 50-54, 1989.
- [13] SCHNEIDER, Heverton; DA SILVA, Charlei Aparecido. **O uso do modelo box plot na identificação de anos-padrão secos, chuvosos e habituais na microrregião de Dourados, Mato Grosso do Sul**. Revista do Departamento de Geografia, v. 27, p. 131-146, 2014.
- [14] ECOR. **Análise Exploratória de Dados**. Disponível em: http://ecologia.ib.usp.br/bie5782/doku.php?id=bie5782:03_apostila:05-exploratoria. Acesso em: 03 de out. 2017
- [15] FM2S. **O que é e para que serve gráfico de dispersão?**. Disponível em: <http://www.fm2s.com.br/grafico-de-dispersao/>. Acesso em: 03 de out. 2017.
- [16] SHIKAMURA, Silvia. **O diagrama de dispersão**. 05 mar. 2012, 07 dec. 2012. Notas de Aula.
- [17] OLIVEIRA, Alexandra; DE SÁ, Joaquim Marques. **Optimização da aplicação NNIG-Nets MLP e desenvolvimento da aplicação NNIG-Nets**. 2008.
- [18] Microsoft. **Apresentar os dados em um gráfico de dispersão ou de linhas**. Disponível em: <https://support.office.com/pt-br/article/Apresentar-os-dados-em-um-gr%C3%A1fico-de-dispers%C3%A3o-ou-de-linhas-4570a80f-599a-4d6b-a155-104a9018b86e>. Acesso em: 03 de out. 2017a.
- [19] Marketing Futuro. **Diagrama de dispersão. O que é, como e quando usar?**. Disponível em: <http://marketingfuturo.com/diagrama-de-dispersao-o-que-e-como-e-quando-usar/>. Acesso em: 03 de out. 2017.
- [20] MASUNARI, Setuko; SWIECH-AYOUB, Bianca de Paula. **Relative growth in the fiddler crab *Uca leptodactyla* Rathbun (Crustacea Decapoda Ocypodidae)**. Revista Brasileira de Zoologia, v. 20, n. 3, p. 487-491, 2003.
- [21] Packtpub. **Creating a hexbin plot**. Disponível em: https://www.packtpub.com/mapt/book/big_data_and_business_intelligence/9781783989508/7/ch07lvl1sec70/creating-a-hexbin-plot. Acesso em: 08 de out. 2017.

- [22] LEWIN-KOH, Nicholas. **Hexagon Binning: an Overview**. Disponível em: https://cran.r-project.org/web/packages/hexbin/vignettes/hexagon_binning.pdf. Acesso em: 08 de out. 2017.
- [23] MOREIRA, José Mauro Magalhães Ávila Paz; SIMIONI, Flávio José; DE SANTANA, Lorena Figueira **Impacto do regime de manejo na rentabilidade da produção de lenha de eucalipto na região de Itapeva-SP, sob condições de risco**. In: Embrapa Florestas-Artigo em anais de congresso (ALICE). In: CONGRESSO DA SOCIEDADE BRASILEIRA DE ECONOMIA, ADMINISTRAÇÃO E SOCIOLOGIA RURAL, 54., 2016, Maceió. Desenvolvimento, território e biodiversidade: anais eletrônicos.[SL]: SOBER, 2016.
- [24] ROUSE, Margaret. **Heat map (heatmap)**. Disponível em: <http://searchbusinessanalytics.techtarget.com/definition/heat-map>. Acesso em: 17 de out. 2017.
- [25] YAU, Nathan. **How to Make a Heatmap – a Quickly and Easy Solution**. Disponível em: <https://flowingdata.com/2010/01/21/how-to-make-a-heatmap-a-quick-and-easy-solution/>. Acesso em: 17 de out. 2017.
- [26] The R Stats Package R. **Draw a Heat Map**. Disponível em: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/heatmap.html>. Acesso em: 17 de out. 2017.
- [27] ZACARIAS, Iulio et al. **Análise comparativa dos acidentes de trabalho no Brasil a partir de dados abertos**. Anais do Computer on the Beach, p. 229-238, 2015.
- [28] MUTHERS, Stefan; MATZARAKIS, Andreas. **Use of beanplots in applied climatology–A comparison with boxplots**. Meteorologische Zeitschrift, v. 19, n. 6, p. 641-644, 2010.
- [29] CAMEY, Suzi Alves; NUNES, Luciana Neves; CRUZ, Luciane Nascimento. **Beanplot uma nova ferramenta gráfica**. Revista HCPA. Porto Alegre. Vol. 30, n. 2 (2010), p. 185-191, 2010.
- [30] KAMPSTRA, Peter et al. Beanplot: **A boxplot alternative for visual comparison of distributions**. 2008.

- [31] KAMPSTRA, Peter. **Package ‘beanplot’**. Disponível em: <https://cran.r-project.org/web/packages/beanplot/beanplot.pdf>. Acesso em: 17 de out. 2017.
- [32] MOLINA, Maria del Carmen Bisi et al. **Reprodutibilidade e validade relativa do Questionário de Frequência Alimentar do ELSA-Brasil**. Cadernos de Saude Publica, v. 29, n. 2, p. 379-389, 2013.
- [33] GORDON, Max. **Introduction to forest plots**. 2017.
- [34] RODRIGUES, Caroline Legramanti. **Metanálise: um guia prático**. 2010.
- [35] GORDON, Max. **Package ‘forestplot’**. Disponível em: <https://cran.r-project.org/web/packages/forestplot/forestplot.pdf>. Acesso em: 31 de out. 2017.
- [36] GODINHO, Ana Sofia et al. **Cirurgia de revascularização miocárdica com circulação extracorpórea versus sem circulação extracorpórea: uma metanálise**. Arq Bras Cardiol, v. 98, n. 1, p. 87-94, 2012.
- [37] GUEDES, Terezinha Aparecida; MARTINS, Ana Beatriz Tozzo; ACORSI, Clédina Regina Lonardan. **Projeto de ensino: aprender fazendo estatística**. Disponível em: < <http://www.scribd.com/doc/1880799/Estatistica-Descritiva>>. Acesso em, v. 2, 2010.
- [38] Microsoft. **Gráficos de colunas (Construtor de Relatórios e SSRS)**. Disponível em: [https://msdn.microsoft.com/pt-br/library/dd239318\(v=sql.120\).aspx#Variações de um gráfico de colunas](https://msdn.microsoft.com/pt-br/library/dd239318(v=sql.120).aspx#Variações de um gráfico de colunas). Acesso em: 21 de Nov. 2017b.
- [39] BRUSCHINI, Cristina; PUPPIN, Andrea Brandão. **Trabalho de mulheres executivas no Brasil no final do século XX**. Cadernos de pesquisa, v. 34, n. 121, p. 105-138, 2004.
- [40] CALLEGARI-JACQUES, SIDIA Maria. **Bioestatística: princípios e aplicações**. 2003. Porto Alegre: Editora Artmed. 255p.
- [41] KUROKAWA, Edson et al. **Utilizando o histograma como uma ferramenta estatística de análise da produção de água tratada de Goiânia**. In: XXVIII CONGRESO INTERAMERICANO DE INGENIERÍA SANITARIA Y AMBIENTAL. 2002.

- [42] SILVA, José Waldemar da; GUIMARÃES, Ednaldo Carvalho; TAVARES, Marcelo. **Variabilidade temporal da precipitação mensal e anual na estação climatológica de Uberaba-MG**. Ciência e Agrotecnologia, v. 27, n. 3, p. 665-674, 2003.
- [43] SELEME, Robson; STADLER, Humberto. **Controle da qualidade: as ferramentas essenciais**. Editora Ibpex, 2008.
- [44] DORNELES, Lúcia Patrícia Pereira; WAECHTER, Jorge Luiz. **Fitossociologia do componente arbóreo na floresta turfosa do Parque Nacional da Lagoa do Peixe, Rio Grande do Sul, Brasil**. Acta Botanica Brasilica, v. 18, n. 4, p. 815-824, 2004.
- [45] Microsoft. **Gráfico de área (Construtor de Relatório e SSRS)**. Disponível em: <https://docs.microsoft.com/pt-br/sql/reporting-services/report-design/area-charts-report-builder-and-ssrs>. Acesso em: 28 de Nov. 2017c.
- [46] Minitab. **Visão geral de Gráfico de área**. Disponível em: <https://support.minitab.com/pt-br/minitab/18/help-and-how-to/graphs/how-to/area-graph/overview/>. Acesso em: 28 de Nov. 2017.
- [47] SOARES, Lucas Santana Furtado. **Regimes fiscais na indústria do petróleo: a influência de características contratuais na atratividade econômica de projetos de exploração e produção**. 2017. Tese de Doutorado.
- [48] CORREA, Sonia Maria Barros Barbosa. **Probabilidade e estatística**. 2003
- [49] MARTINS, Paola da Silva; ALCOFORADO, Luciane Ferreira. **Treinando habilidades de elaboração de gráficos com o software R**. 2015.
- [50] AZZONI, Carlos Roberto. **Setor terciário e concentração regional no Brasil**. Economia e Território. Setor terciário e concentração regional no Brasil. Belo Horizonte: Editora UFMG, 2005.

APÊNDICE

```
#include <stdio.h>
#include <conio.h>
#include <stdlib.h>

int main(void)
{
    int i;

    printf("Dados 1:\n\n");

    for (i = 0; i < 100; i++)
    {
        /* gerando valores aleatórios entre zero e 100 */
        printf(" %d", 60+ (rand() % 40));
    }

    printf("\n\nDados 2:\n\n");

    for (i = 0; i < 100; i++)
    {
        /* gerando valores aleatórios entre zero e 100 */
        printf(" %d", 60+ (rand() % 40));
    }

    printf("\n\nIMC:\n\n");

    for (i = 0; i < 100; i++)
    {
        /* gerando valores aleatórios entre 20 e 39 */
        printf(" %.1f", 20.0+ (float)(rand() % 190)/10);
    }

    printf("\n\nEmpresa 1:\n\n");

    for (i = 0; i < 8; i++)
    {
        /* gerando valores aleatórios entre 16 e 33 */
        printf(" %.1f", 16.0+ (float)(rand() % 170)/10);
    }

    printf("\n\nEmpresa 2:\n\n");

    for (i = 0; i < 8; i++)
    {
        /* gerando valores aleatórios entre 28 e 37 */
        printf(" %.1f", 28.0+ (float)(rand() % 110)/10);
    }
}
```

```
    printf("\n\nEmpresa 3:\n\n");

    for (i = 0; i < 8; i++)
    {
        /* gerando valores aleatórios entre 2 e 17 */
        printf(" %.1f", 2.0+ (float)(rand() % 150)/10);
    }

    printf("\n\nEmpresa 4:\n\n");

    for (i = 0; i < 8; i++)
    {
        /* gerando valores aleatórios entre 11 e 19 */
        printf(" %.1f", 11.0+ (float)(rand() % 80)/10);
    }

    getch();
}.
```