

# Como montar um gráfico no R usando o pacote ggplot2

Fernando Basquiroto de Souza

2019-04-05

# 1 Baixando os Dados de Licenciamento

Utilizaremos os dados disponíveis no site do Portal Brasileiro de Dados Abertos (dados.gov.br). Neste, iremos buscar na ferramenta de busca pelo seguinte termo "Licenças Ambientais", conforme figura 1.

The screenshot shows the homepage of dados.gov.br. At the top, there is a navigation bar with links: BRASIL, Serviços, Simplifique!, Participe, Acesso à informação, Legislação, and Canais. Below this is a green header with the dados.gov.br logo and the text 'PORTAL BRASILEIRO DE DADOS ABERTOS'. A secondary navigation bar contains links: Dados, Organizações, Aplicativos, Inventários, Concursos, INDA, Perguntas frequentes, Contato, and Sobre o portal. The main content area features a search bar with the text 'Licenças Ambientais' entered, highlighted with a green box and labeled '1'. To the right of the search bar is a 'PESQUISAR' button, also highlighted with a green box and labeled '2'. Below the search bar, it says 'em 6665 conjuntos de dados com 34549 recursos (o que é isto?)'. On the left side, there are sections for 'Conjuntos de dados em destaque' (highlighting 'Cadastro Nacional de Estabelecimentos de Saúde - CNES' and 'Infopreço') and 'Orçamento Federal - Orçamento Fiscal e da Seguridade Social'. On the right side, there is a section titled 'Publicações mais recentes' with a table of recent publications.

Conjunto de dados	Data	Responsável
Campanhas educativas de trânsito Detran-DF	29/03/2019	Distrito Federal
Cursos oferecidos pelo Detran/DF ao público	29/03/2019	Distrito Federal
Tipos de infrações cometidas nas vias urbanas do Distrito Federal	29/03/2019	Distrito Federal
Número de habilitados nos Distrito Federal nos últimos 10 anos	29/03/2019	Distrito Federal

Figura 1: Digite "Licenças Ambientais" na busca (1) e clique em pesquisar (2).

Em seguida, clique no item "Licenças ambientais de atividades e empreendimentos licenciados pelo IBAMA". Ao abrir essa opção, você poderá acessar os dados nos seguintes formatos: HTML, CSV, XML e JSON. No nosso caso, iremos utilizar o formato CSV. Desta forma, clique sobre o botão "Explorar" do item CSV e selecione "Ir para o recurso". As próximas imagens ilustram esse procedimento (figuras 2 e 3).

Ao baixar o arquivo CSV, iremos manter o mesmo nome dado ao arquivo, ou seja, "sislic-licencas.csv". Agora que já temos nossos dados em mãos, vamos abrir o R para começarmos a trabalhar com eles.



Figura 2: Resultado da busca por "Licenças Ambientais".



Figura 3: Informações sobre o conjunto de dados referentes às licenças ambientais e tipo de dados disponíveis.

## 2 Um gráfico simples no R

Após abrir o R, vamos começar a escrever nosso script. Na lista abaixo, iremos descrever os itens que utilizaremos.

- `install.packages()`: Função usada para instalar pacotes no R, neste caso, usaremos ela para os pacotes `ggplot2` e `lubridate`;
- `library()`: Função utilizada para carregar pacotes não iniciados automaticamente no R;
- `setwd()`: Função para definir o ambiente de trabalho (local onde estão salvos nossos dados);
- `read.csv()`: Função usada para carregar nossa planilha no formato CSV;
- `as.Date()`: Função para converter um conjunto de dados para o formato de data;
- `year()`: Função do pacote `lubridate`, ela extrai o ano de um conjunto de datas;
- `c()`: Função para agrupar um conjunto de dados;
- `ggplot()`: Função utilizada para criar nossos gráficos.

Agora que conhecemos as funções que utilizaremos, vamos aos primeiros passos, ou seja, instalar e carregar os pacotes, determinar o ambiente de trabalho, carregar nossa planilha CSV e realizar algumas modificações nos nossos dados para criar o gráfico.

```
install.packages(lubridate)
install.packages(ggplot2)

library(lubridate)
library(ggplot2)

setwd("C:/Users/ferna/Desktop/")
dados <- read.csv("sislic-licencas.csv", header = TRUE,
                  stringsAsFactors = FALSE, sep=";")
data_ emissao <- as.Date(dados$DAT_EMISSAO, "%d/%m/%Y")

dados$ano_ emissao <- year(data_ emissao)

ggplot(dados, aes(ano_ emissao)) + geom_histogram()
```

Se todo esse procedimento for realizado corretamente, o gráfico apresentado na figura 4 será gerado.

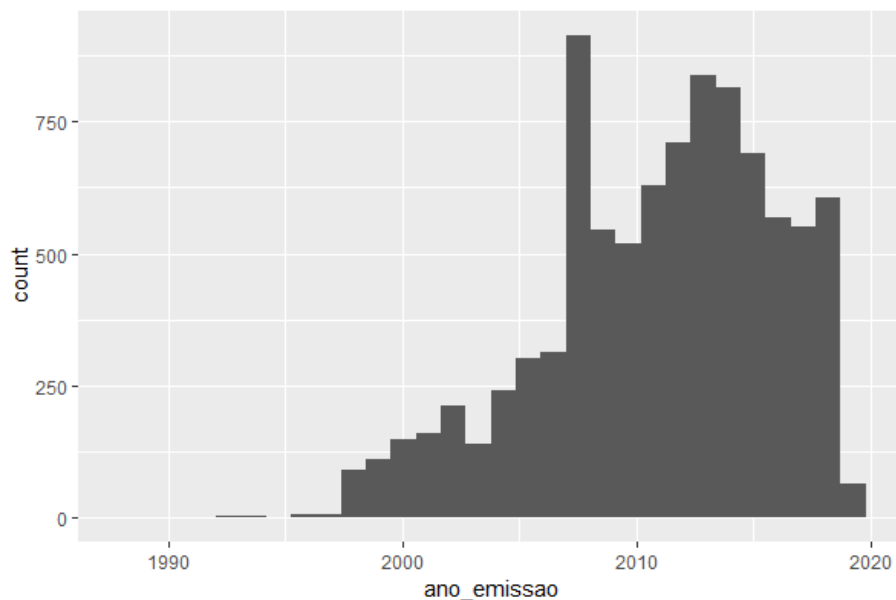


Figura 4: Gráfico gerado com configurações mínimas na função ggplot.

Note que precisamos configurar vários itens do nosso gráfico, tais como os rótulos do eixo x e y, os anos que aparecem no eixo x, os intervalos que se referem a cada barra do histograma e as cores. Vamos ver isso na próxima seção.

### 3 Configurando nosso gráfico no R

Primeiro, iremos criar uma nova variável (chamada de `rotuloX`), a qual irá receber todos os anos do nosso conjunto de dados, para que estes apareçam no eixo x do nosso gráfico. Para configurar quantos itens há em cada intervalo, vamos acrescentar o parâmetro `binwidth` na função `geom_histogram()` (neste caso, 1, ou seja, cada intervalo é representado por um ano).

Também iremos adicionar as funções `ylab()` e `xlab()` para informar quais rótulos deverão aparecer nos eixos y e x, respectivamente.

Confira o código abaixo com as devidas modificações.

```
install.packages(lubridate)
install.packages(ggplot2)

library(lubridate)
library(ggplot2)

setwd("C:/fernando/blog/graph_R_ebook/")
dados <- read.csv("sislic-licencas.csv", header = TRUE,
```

```

stringsAsFactors = FALSE, sep=";")
data_emissao <- as.Date(dados$DAT_EMISSAO, "%d/%m/%Y")

dados$ano_emissao <- year(data_emissao)

rotuloX <- c(1988, 1989, 1990, 1991, 1992, 1993, 1994,
            1995, 1996, 1997, 1998, 1999, 2000, 2001,
            2002, 2003, 2004, 2005, 2006, 2007, 2008,
            2009, 2010, 2011, 2012, 2013, 2014, 2015,
            2016, 2017, 2018, 2019)

ggplot(dados, aes(ano_emissao)) + geom_histogram(binwidth = 1) +
  ylab("Número de Licenças Emitidas pelo IBAMA") + xlab("Ano")

```

A figura 5 nos mostra o resultado dessas modificações. Note que estamos, aos poucos, modificando e dando uma cara nova ao nosso gráfico.

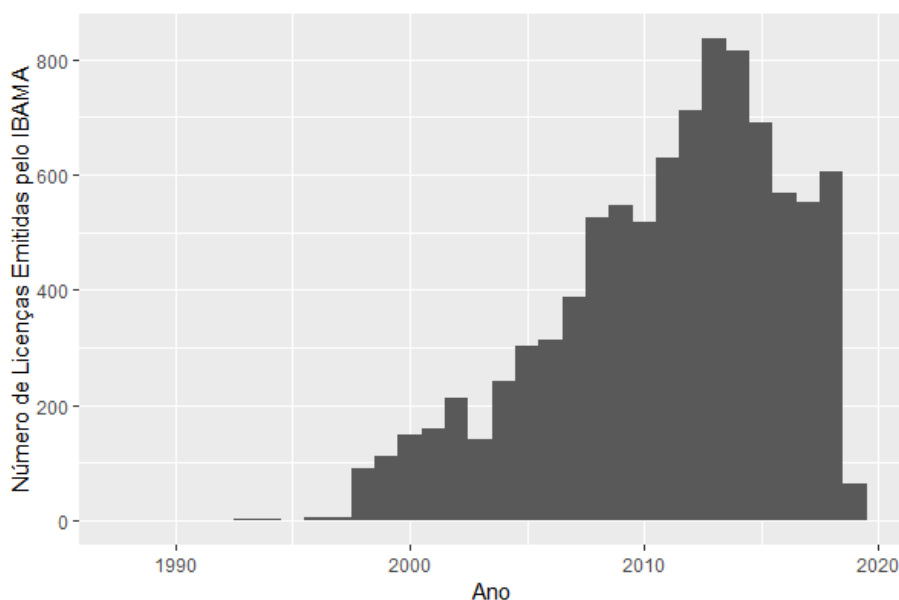


Figura 5: Novo gráfico com eixo x, y e intervalo das barras por ano.

Vamos continuar com os nossos ajustes.

Veja que no eixo x, não são todos os anos que aparecem, e sim, somente os anos 1990, 2000, 2010 e 2020. Como podemos modificar isso? Essa mudança é realizada usando a função `scale_x_continuous()` e para garantir que esses rótulos não fiquem um por cima do outro, vamos usar a função `theme()` para modificar o tema em uso e indicar que este rótulo deverá ficar na vertical.

O código a seguir apresenta essa mudanças, sendo que a figura 6 mostra o resultado.

```

install.packages(lubridate)
install.packages(ggplot2)

library(lubridate)
library(ggplot2)

setwd("C:/fernando/blog/graph_R_ebook/")
dados <- read.csv("sislic-licencas.csv", header = TRUE,
                  stringsAsFactors = FALSE, sep=";")
data_emissao <- as.Date(dados$DAT_EMISSAO, "%d/%m/%Y")

dados$ano_emissao <- year(data_emissao)

rotuloX <- c(1988, 1989, 1990, 1991, 1992, 1993, 1994,
            1995, 1996, 1997, 1998, 1999, 2000, 2001,
            2002, 2003, 2004, 2005, 2006, 2007, 2008,
            2009, 2010, 2011, 2012, 2013, 2014, 2015,
            2016, 2017, 2018, 2019)

ggplot(dados, aes(ano_emissao)) + geom_histogram(binwidth = 1) +
  ylab("Número de Licenças Emitidas pelo IBAMA") + xlab("Ano") +
  scale_x_continuous(breaks = rotuloX) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.3))

```

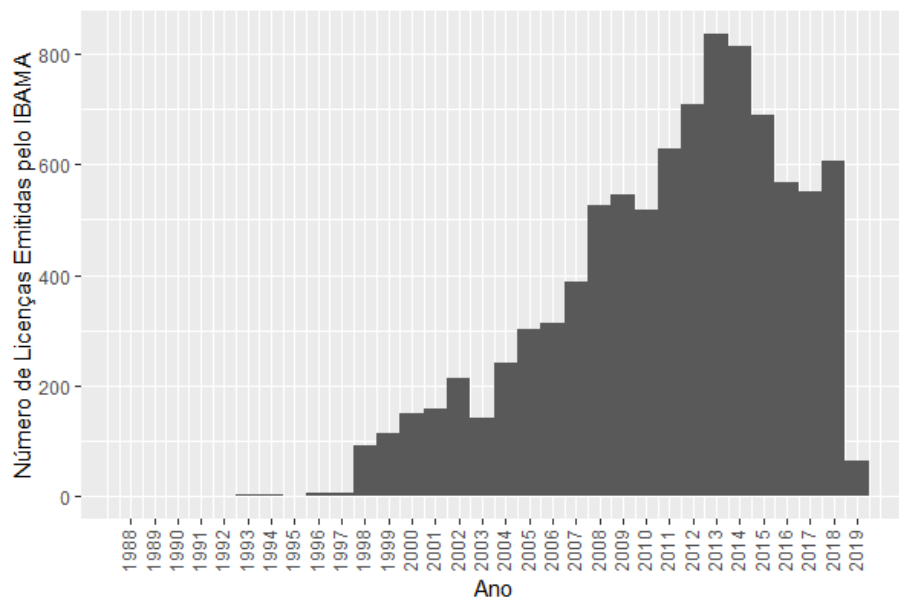


Figura 6: Modificando e rotacionando os rótulos no eixo x.

## 4 Últimas modificações

Agora, para finalizar nosso gráfico, iremos modificar suas cores, assim como o seu tema, adicionar a quantidade de licenças em cada barra e adicionar texto para indicar a referência dos dados.

As cores serão modificadas pelos parâmetros `fill` e `col` questão dentro da função `geom_histogram()`. Já a quantidade de licenças em cada barra será adicionada pela função `stat_bin()` e o tema será alterado usando a função `theme_bw()`. O texto será adicionado usando a função `annotate()`.

O código a seguir demonstra as alterações que acabamos de propor e a figura 7 mostra a gráfico final.

```
install.packages(lubridate)
install.packages(ggplot2)

library(lubridate)
library(ggplot2)

setwd("C:/fernando/blog/graph_R_ebook/")
dados <- read.csv("sislic-licencas.csv", header = TRUE,
                  stringsAsFactors = FALSE, sep=";")
data_ emissao <- as.Date(dados$DAT_EMISSAO, "%d/%m/%Y")

dados$ano_ emissao <- year(data_ emissao)

rotuloX <- c(1988, 1989, 1990, 1991, 1992, 1993, 1994,
            1995, 1996, 1997, 1998, 1999, 2000, 2001,
            2002, 2003, 2004, 2005, 2006, 2007, 2008,
            2009, 2010, 2011, 2012, 2013, 2014, 2015,
            2016, 2017, 2018, 2019)

ggplot(dados, aes(ano_ emissao)) +
  geom_histogram(binwidth = 1, fill = '#32CD32', col="#008000")+
  ylab("Número_de_Licenças_Emitidas_pelo_IBAMA") + xlab("Ano") +
  scale_x_continuous(breaks = rotuloX) +
  stat_bin(binwidth=1, geom="text", colour="black", size=3,
           aes(label=..count..), position="identity", vjust=-0.5) +
  theme_bw()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.3))+
  annotate("text", x = 1994, y = 800,
         label = "Blog_2_Engenheiros", alpha=0.3)+
  annotate("text", x = 1994, y = 760,
         label = "Fonte:_dados.gov.br", alpha=0.3)
```



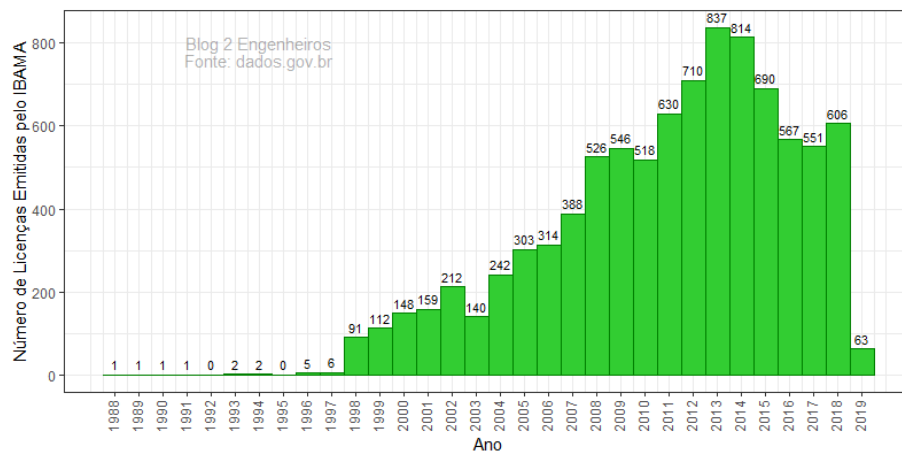


Figura 7: Gráfico finalizado com as ultimas alterações.

Agora você já sabe como construir um gráfico utilizando ggplot2 e R. Caso você queira saber mais sobre licenciamento ambiental, assista o nosso vídeo "Licenciamento Ambiental Auto Declaratório em Santa Catarina" e saiba o que é a Licença Ambiental por Compromisso (LAC).

Visite também o Blog 2 Engenheiros e confira matérias e tutoriais voltados para Engenharia Ambiental.