

R para Economia

Lucas Mendes

24/03/2020

Visualização de dados

Visualização de dados

Uma das coisas mais importantes quando estamos fazendo algum tipo de análise é a visualização dos dados

Caso negligencie essa parte, poderá ter problemas que teriam sido evitados com uma simples olhada em um gráfico.

Por isso é importante sabermos um ferramental básico de construção de gráficos.

Visualização de dados

O R possui funções básicas de visualização já instaladas, como a função `plot()`.

Porém nós iremos aprender utilizando o pacote `ggplot2` que vem instalado junto ao `tidyverse`.

Ele utiliza o conceito de camadas para construir um gráfico, você pode saber mais sobre clicando aqui

Tipos de gráficos

Há diversos tipos de gráficos que servem para variados tipos de finalidade. Hoje nós aprenderemos os mais básicos:

- Gráficos de barras
- Gráficos de linha
- Gráficos de pontos

Montando seu primeiro gráfico

Gráfico de barras

Primeiro iremos criar um tibble, que é o dataframe do universo tidyverse. Não se preocupe, um tibble é quase a mesma coisa que um data frame

```
df1 <- tibble(Notas = c(10,8,4,6),  
              Alunos = c("Jão","Maria","Tony","Sofia"))  
df1
```

Notas	Alunos
10	Jão
8	Maria
4	Tony
6	Sofia

Gráfico de barras

Agora iremos mapear os nossos eixos do gráfico (x,y). Se queremos comparar as notas entre os alunos, o nosso eixo x será o nome dos alunos e o eixo y será a nota de cada.

Para mapear os eixos, usaremos o seguinte código:

```
ggplot(df1, # Base de dados
  aes(
    x = Alunos, # Eixo x
    y = Notas) # Eixo y
)
```


Gráfico de barras

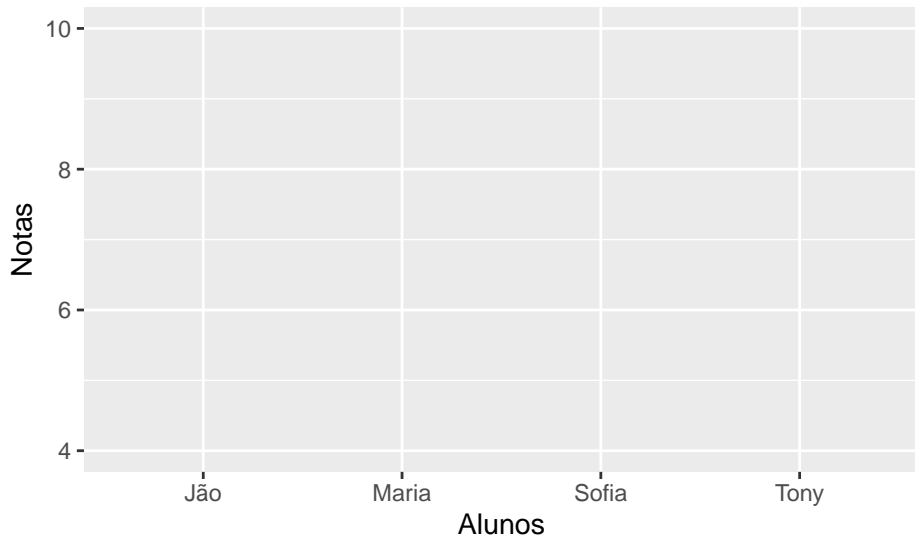


Gráfico de barras

O que acabamos de fazer foi a primeira camada do nosso gráfico, que é simplesmente o mapeamento dos eixos. Agora iremos de fato definir qual o formato do nosso gráfico, que no caso é o de barras.

```
ggplot(df1, # Base de dados
  aes(
    x = Alunos, # Eixo x
    y = Notas) # Eixo y
) +
geom_bar(stat = "identity") # formato do gráfico
```

Gráfico de barras

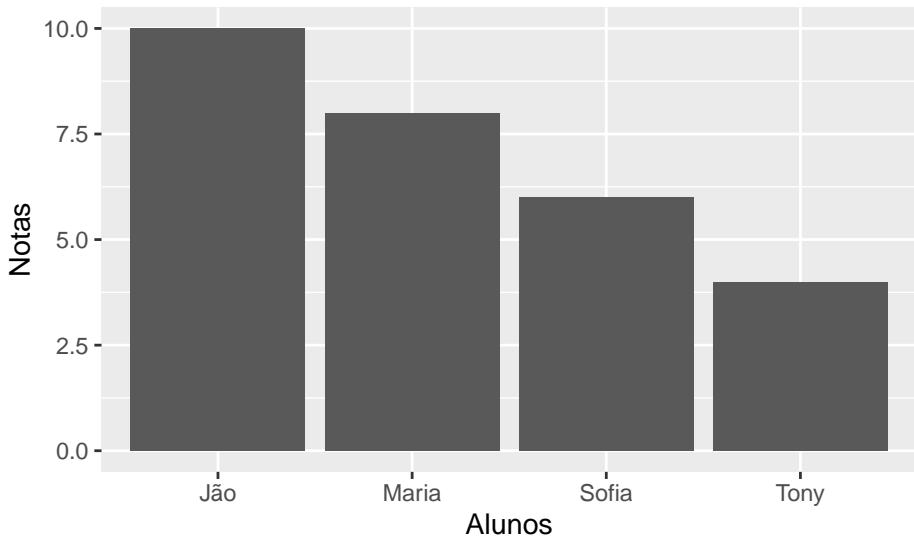


Gráfico de barras

A função `geom_bar` é a responsável pela geração do gráfico de barras.

Essa é a segunda camada do nosso gráfico.

Você pode botar mais camadas, mas aí é opcional. Duas é o mínimo.

Repare também que estamos usando o operador `+` ao invés do `%>%`.

Quando começamos uma função do `ggplot2` passamos a usar `+` para ligar as funções.

Gráfico de linha

Normalmente usamos um gráfico de linha para observar dados no tempo
Vamos usar como base de dados o dataframe `economics` do pacote `tidyr`

```
head(economics)
```

date	pce	pop	psavert	uempmed	unemploy
1967-07-01	506.7	198712	12.6	4.5	2944
1967-08-01	509.8	198911	12.6	4.7	2945
1967-09-01	515.6	199113	11.9	4.6	2958
1967-10-01	512.2	199311	12.9	4.9	3143
1967-11-01	517.4	199498	12.8	4.7	3066
1967-12-01	525.1	199657	11.8	4.8	3018

Gráfico de linha

Iremos elaborar um gráfico de linha que relacione no tempo o número total de desempregados da economia americana em milhares, que é representada pela coluna `unemploy`.

O procedimento da primeira camada é o mesmo do gráfico de barras

```
ggplot(economics, # Base de dados
  aes(
    x = date, # Eixo x
    y = unemploy) # Eixo y
) # formato do gráfico
```

Gráfico de linha

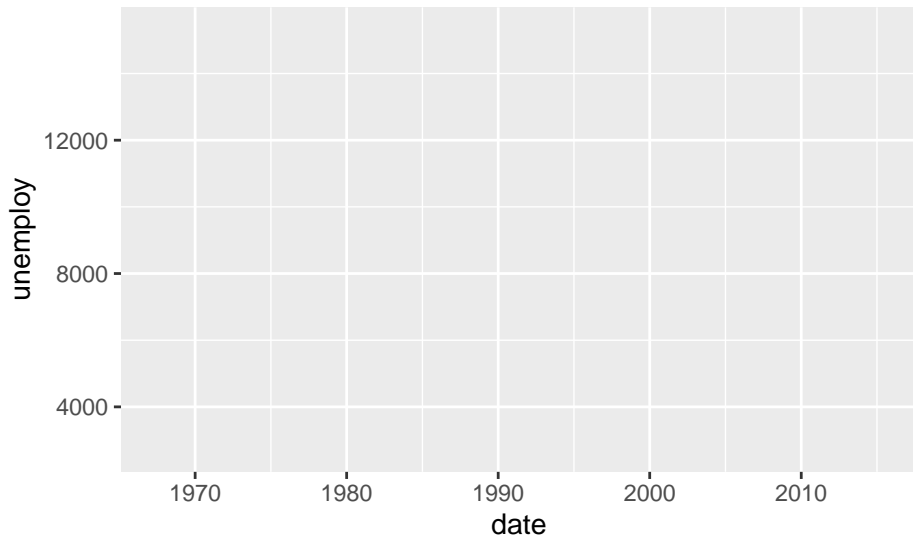


Gráfico de linha

Para contruir a segunda camada usaremos a função `geom_line`

```
ggplot(economics, # Base de dados
  aes(
    x = date, # Eixo x
    y =unemploy) # Eixo y
) +
geom_line() # formato do gráfico
```


Gráfico de linha

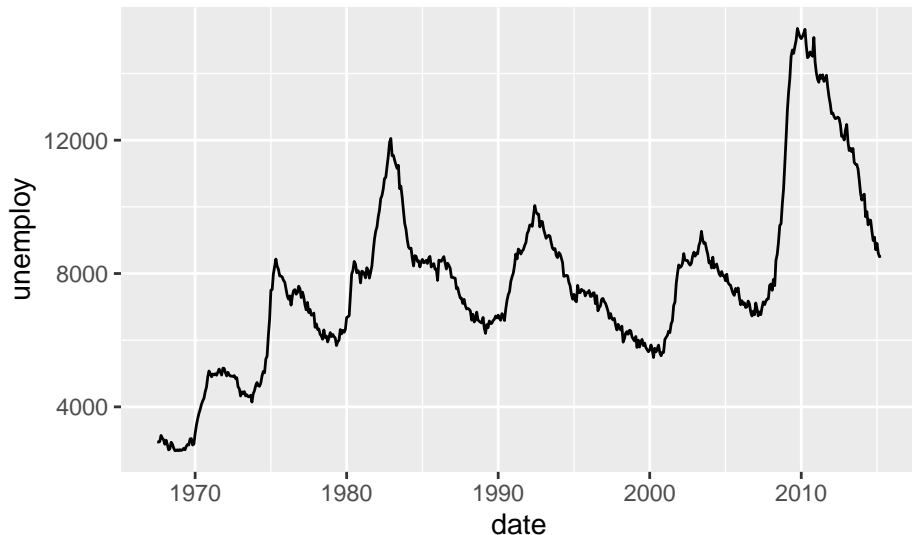


Gráfico de pontos

Um gráfico de pontos, ou gráfico de dispersão, tem a finalidade de mostrar o relacionamento entre duas variáveis. Se elas tem um relacionamento positivo ou negativo ou até se não há relacionamento.

Para elaborar um desses, utilizaremos as colunas `psavert` e `pce` da base de dados `economics`

Gráfico de pontos

Mapeando os eixos

```
ggplot(economics, # Base de dados
  aes(
    x = pce, # Eixo x
    y = psavert) # Eixo y
)
```

Gráfico de pontos

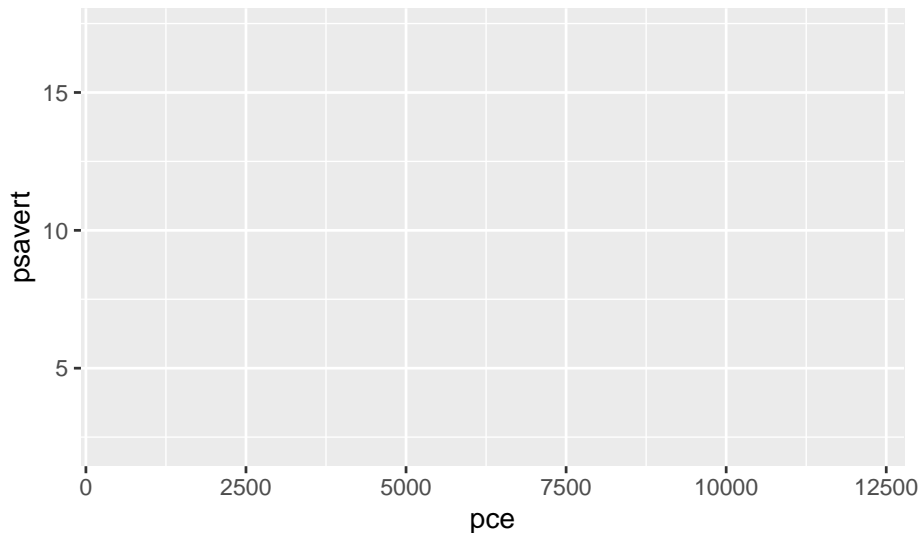
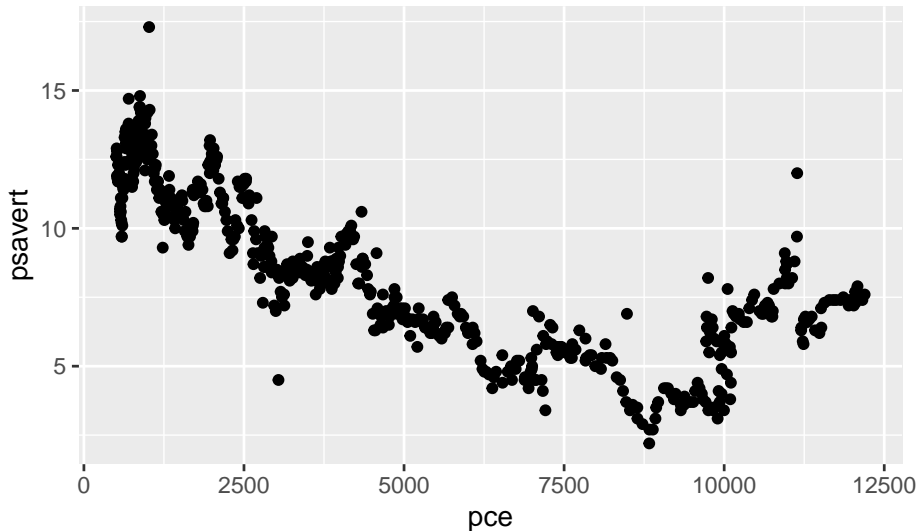


Gráfico de pontos

Utilizaremos agora a função `geom_point` para criar um gráfico de pontos.

```
ggplot(economics, # Base de dados
  aes(
    x = pce, # Eixo x
    y = psavert) # Eixo y
) +
geom_point() # formato do gráfico
```

Gráfico de pontos



Aesthetics

Aesthetics

Os graficos ainda estão muito simples, há diversas funções para melhorar o visual do pacote. Por enquanto nós iremos carregar o pacote `ggthemes` para melhorar o aspecto visual.

```
library(ggthemes)
```


Aesthetics

Primeiro irei usar o grafico sobre desemprego como exemplo. Irei grava - lo numa variavel

```
g1 <- ggplot(economics, # Base de dados
  aes(
    x = date, # Eixo x
    y =unemploy) # Eixo y
) +
geom_line()
```

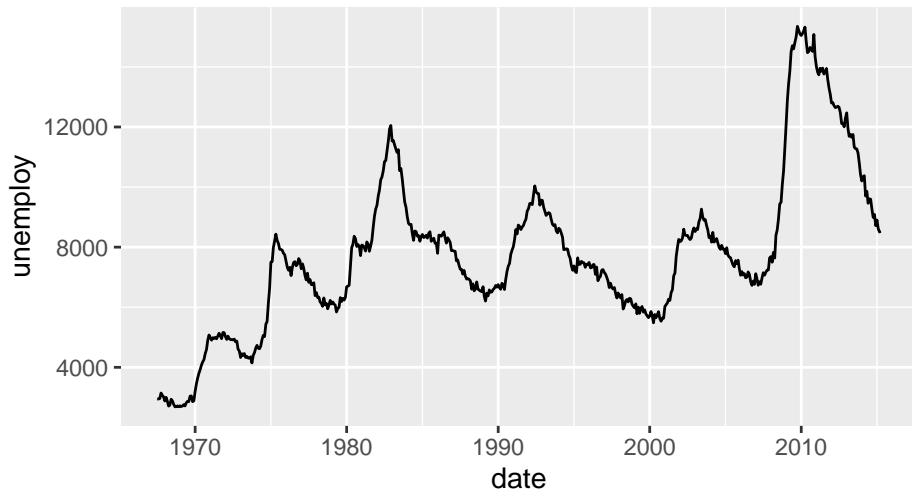
Aesthetics

Agora irei dar um título para o gráfico

```
g1 + labs(title = "Total de Pessoas Desempregadas nos EUA")
```

Aesthetics

Total de Pessoas Desempregadas nos EUA



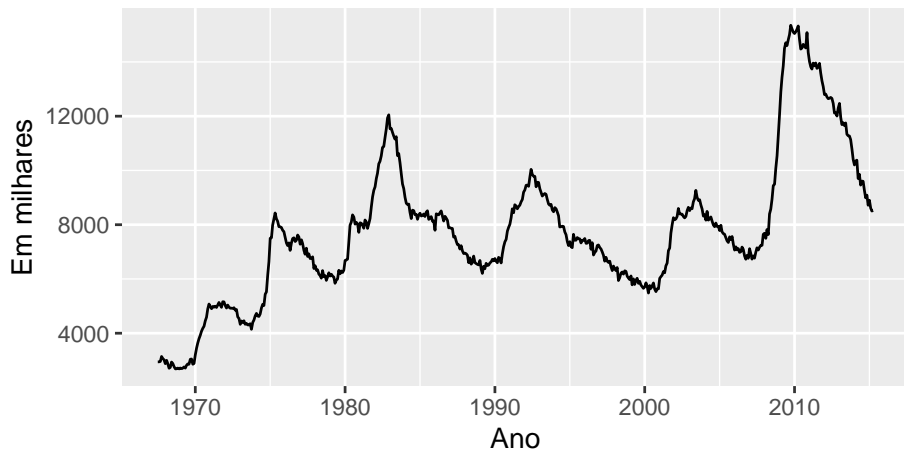
Aesthetics

Agora irei renomear os eixos x e y e a colocar um caption

```
g1 + labs(title = "Total de Pessoas Desempregadas nos EUA",  
          x = "Ano",  
          y = "Em milhares",  
          caption = "Fonte: FED")
```

Aesthetics

Total de Pessoas Desempregadas nos EUA



Fonte: FED

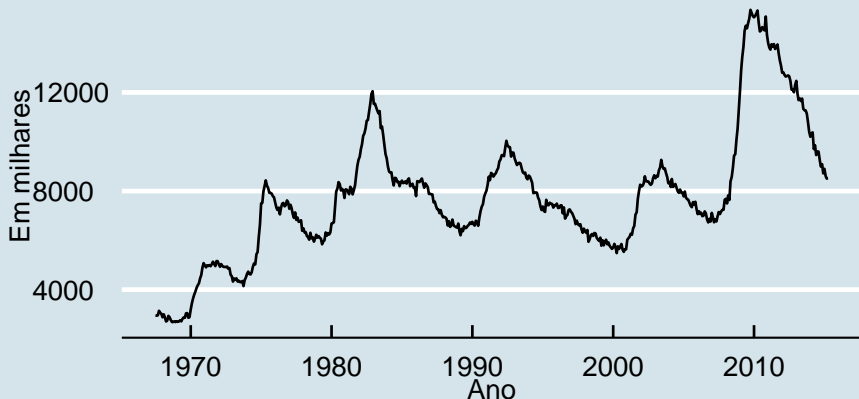
Aesthetics

Agora irei mudar o layout do gráfico, para que fique parecido com o layout do jornal **The Economist**.

```
g1 + labs(title = "Total de Pessoas Desempregadas nos EUA",  
          x = "Ano",  
          y = "Em milhares",  
          caption = "Fonte: FED") +  
theme_economist()
```

Aesthetics

Total de Pessoas Desempregadas nos



Fonte: FED

Aesthetics

Há **N** maneiras de você “embeleazar” o seu gráfico e apresentá-lo. Mas isso é matéria para um curso de storytelling de dados.

Com isso, leia a documentação do pacote ggplot2 e veja esse link para ver como a BBC usa o R para fazer gráficos

Exercícios