



[illegible]

# DashboardTableau

The Tableau logo is centered on the slide. It features a cluster of nine plus signs arranged in a 3x3 grid. The plus signs are colored orange, teal, and dark blue. Below the cluster, the word "tableau" is written in a dark blue, sans-serif font, with a registered trademark symbol (®) to its upper right.

# Text Mining

## Definição

*Text mining, ou mineração de textos, se refere ao processo de recuperar informações relevantes de um texto. Esta informação relevante é derivada de padrões e tendências que sejam relevantes ao negócio e encontradas nos textos.*

**Mas por que Text Mining?  
Para BI?**

# Mas por que Text Mining?

**2 zettabytes de informação em 2013**

**Segundo fontes são informações não estruturadas:**

**80% - Gartner**

**85% - IBM e Teradata**

**Os dados não estruturados possuem informações relevantes para criação de dashboards, relatórios e modelagem**

# Mas por que Text Mining?

Lembrando que o objetivo de BI é:

*“...prover informações para a tomada de decisão...”*

**Usar informações textuais é uma vantagem para tomada de decisão!**

- **Análise de sentimento**
- **Predição do mercado financeiro**
- **CRM – Churn, Retenção e verticalização**

**... IBM – Watson!**

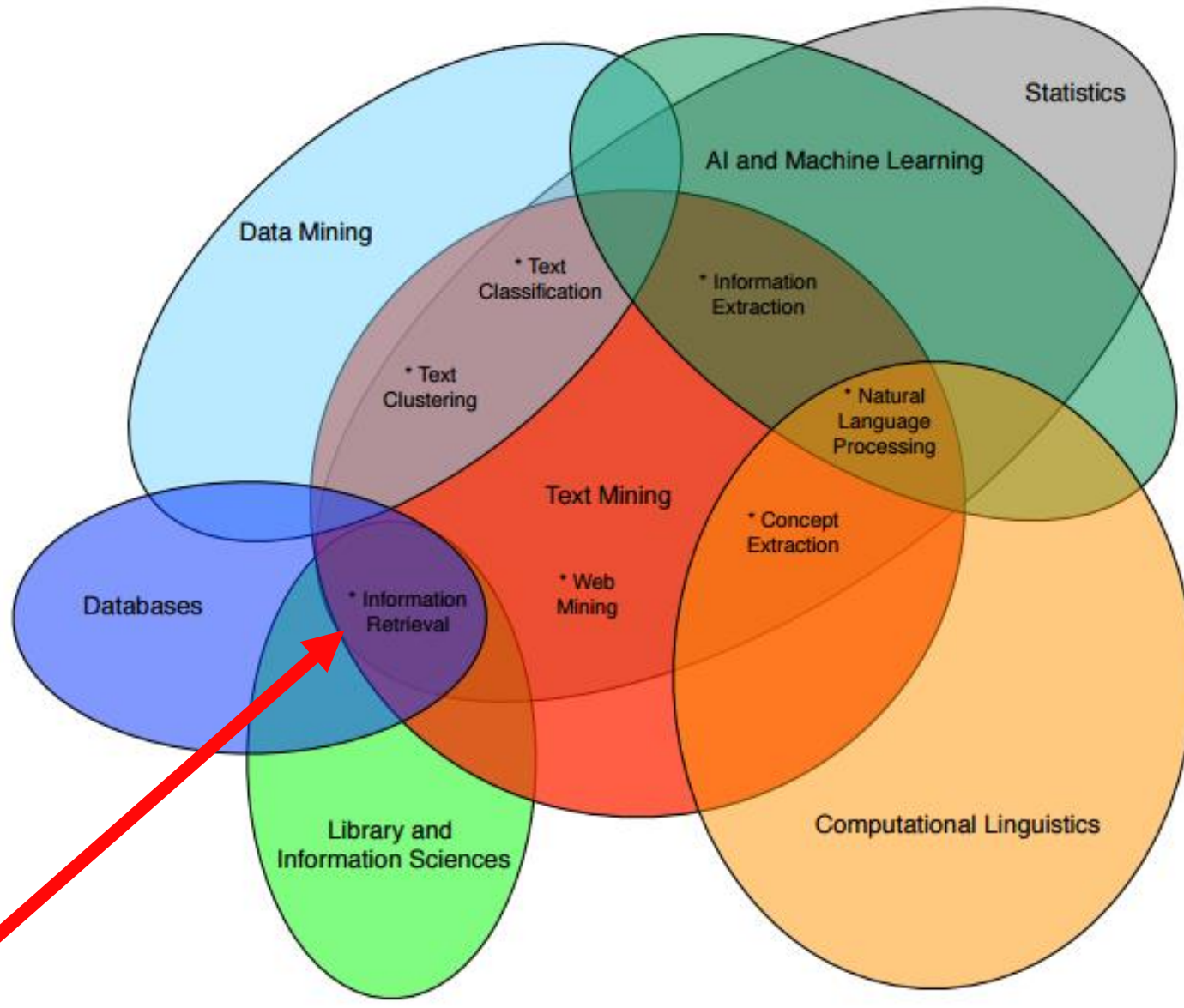
# Mas por que Text Mining?

Text Mining and Text Analytics Compared and Contrasted

	Text Mining	Text Analytics
<b>Approach</b>	Both employ natural language processing (NLP) algorithms to parse text and record entities found there.	
<b>When to use</b>	When you don't know what's in the text or what you need to extract from it.	When you know what's in the text and what you need to extract from it.
<b>Types of sources</b>	With either, text can be anywhere accessible, including various types of files, repositories, databases, or content management systems.	
<b>Quantity of sources</b>	Most configurations operate on large numbers of documents, possibly across many repositories, to understand what topics these documents mention.	Most configurations operate on one or a small number of unstructured data sources, the content of which you understand well.
<b>Output</b>	A taxonomy of entities (with links to their sources). The taxonomy is analytic in nature, since it has classified the content it parsed and even organized the taxonomy by how found entities relate.	A stream of records, each describing an entity instance. These records usually go into a database table or file, less often into a search index.
<b>Use of output</b>	The taxonomy may double as a navigation medium through which users access source documents.	The resulting database or index is itself not analytic per se, but it's used as a source for standard reports or statistical analysis.
<b>Applications</b>	Classifying content in large bodies of documents (possibly in content management systems or portals) in media (especially publishing), research (especially pharma), federal intelligence agencies, etc.	Risk analysis, fraud detection, call center analysis, product service reports, regulatory reporting, customer base segmentation, other customer analytics, etc.
<b>Synonyms</b>	Content classification, topic clustering, taxonomy generation	Entity extraction, concept extraction, concept search

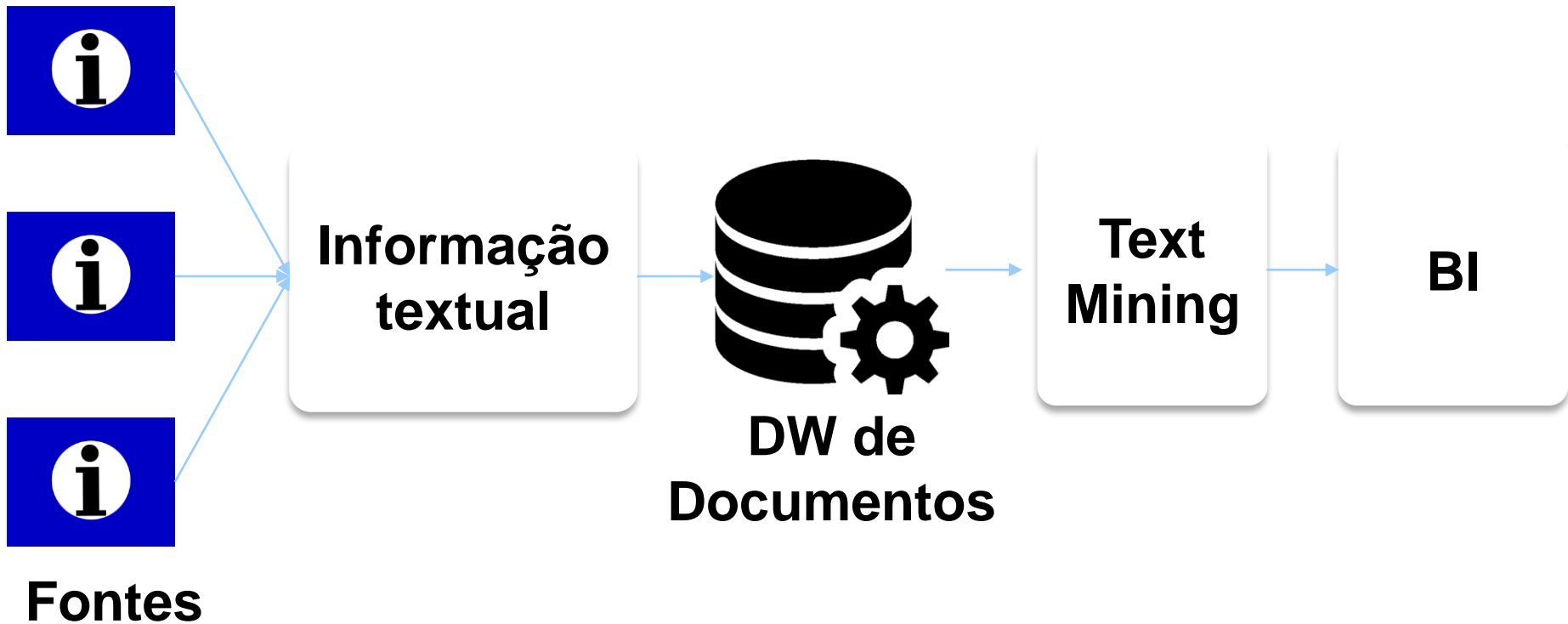


# Para Business Intelligence?



# Para Business Intelligence?

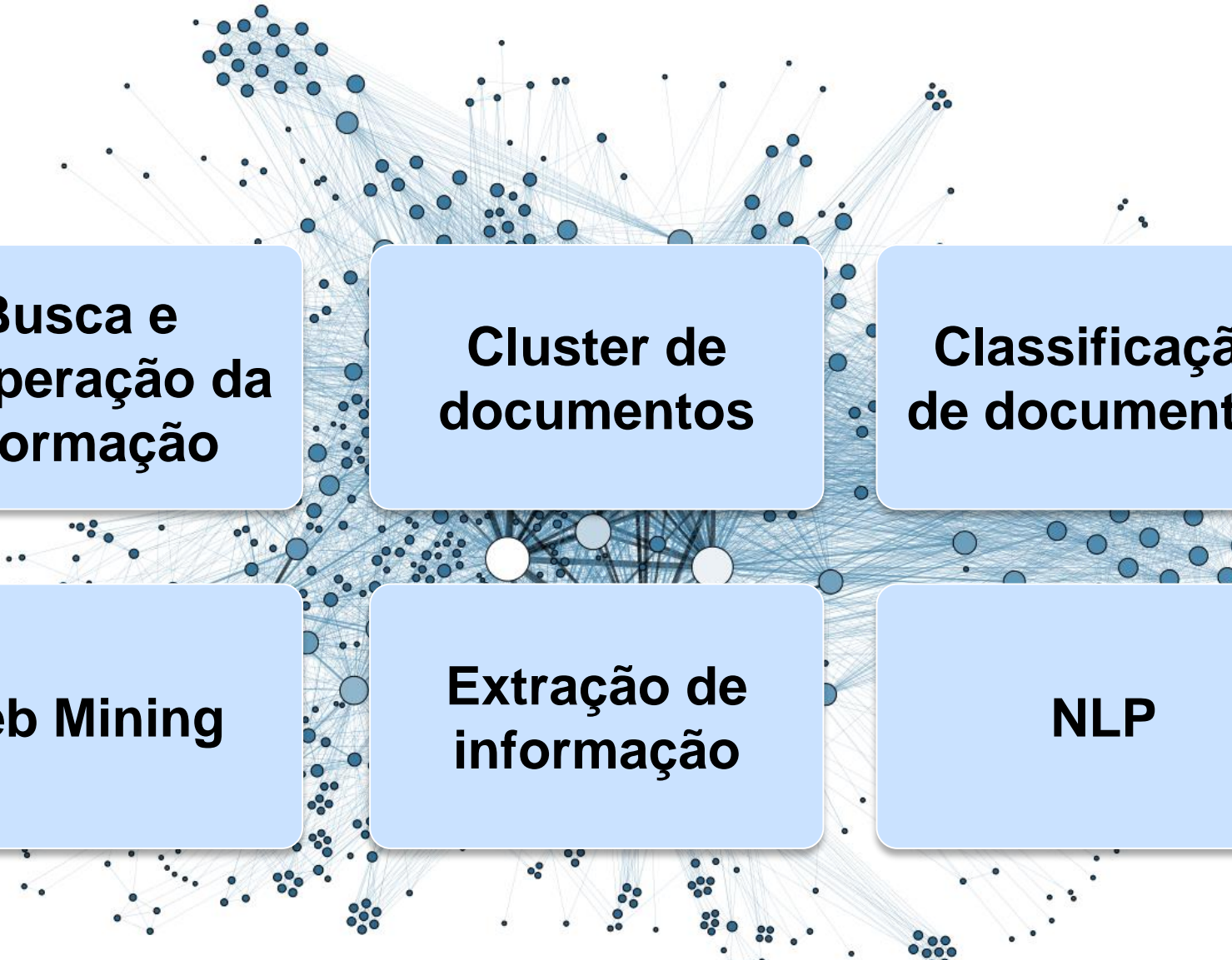
## DW de Documentos





# Para Business Intelligence?

## Tipos de Text Mining



**Busca e  
recuperação da  
informação**

**Cluster de  
documentos**

**Classificação  
de documentos**

**Web Mining**

**Extração de  
informação**

**NLP**

# Aplicações de Text Mining

- **Marketing:** Identificar consumidores em potencial com o uso de informações textuais
- **Indústria:** Web mining para identificação de concorrentes e mapeamento de mercado (preços e produtos)
- **Busca de emprego:** quais são as palavras-chaves mais publicadas em anúncios de emprego?



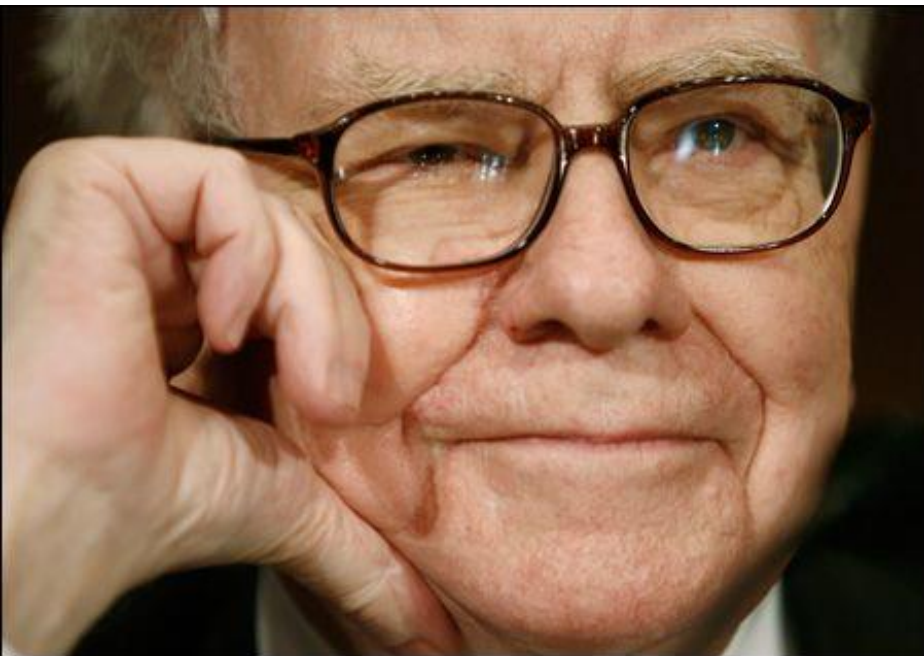
## Aplicações de Text Mining

***Previsão de lucro de companhias listadas no Ibovespa baseado em análise de balanços, indicadores macroeconômicos e monitoramento de mídias eletrônicas***



# Aplicações de Text Mining

***É possível estimar a variação do lucro de uma companhia listada em bolsa na próxima publicação de resultados, para auxiliar na tomada de decisão de investimentos?***

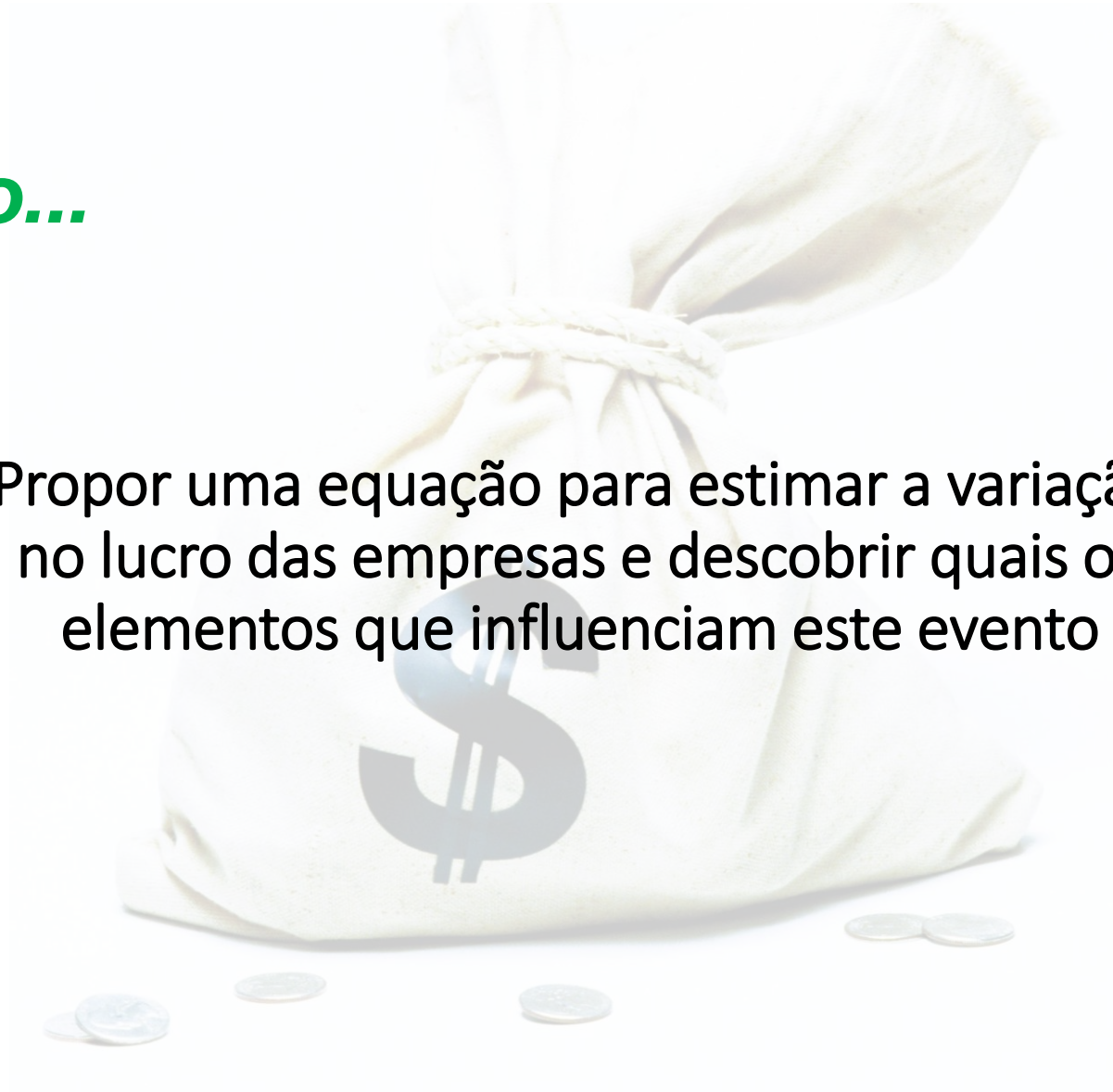




# Aplicações de Text Mining

## Nosso objetivo...

Propor uma equação para estimar a variação  
no lucro das empresas e descobrir quais os  
elementos que influenciam este evento



# Aplicações de Text Mining

## Sendo mais específico...



1

***Relacionar dados sobre custos e receitas publicados nos balanços financeiros com indicadores macroeconômicos e a publicação de notícias em mídias eletrônicas***

2

Extrair informações das notícias nas mídias eletrônicas através de text mining

3

Analisar a flutuação da variação do lucro das empresas em relação às variáveis desenvolvidas

# Aplicações de Text Mining

## E na prática?

***Identificação de empresas lucrativas – maior retorno para o acionista***

***Facilitar a tomada de decisão de investimentos para o pequeno investidor e as pequenas e médias empresas***





# Aplicações de Text Mining

## Como fazer?

***Levantamento dos balanços trimestrais dos últimos 2 anos das empresas listadas no Ibovespa***

Desenvolvimento de um robô para captação das notícias na web relacionadas à estas companhias.

Fontes: Estadão, Folha, Infomoney e Valor Econômico

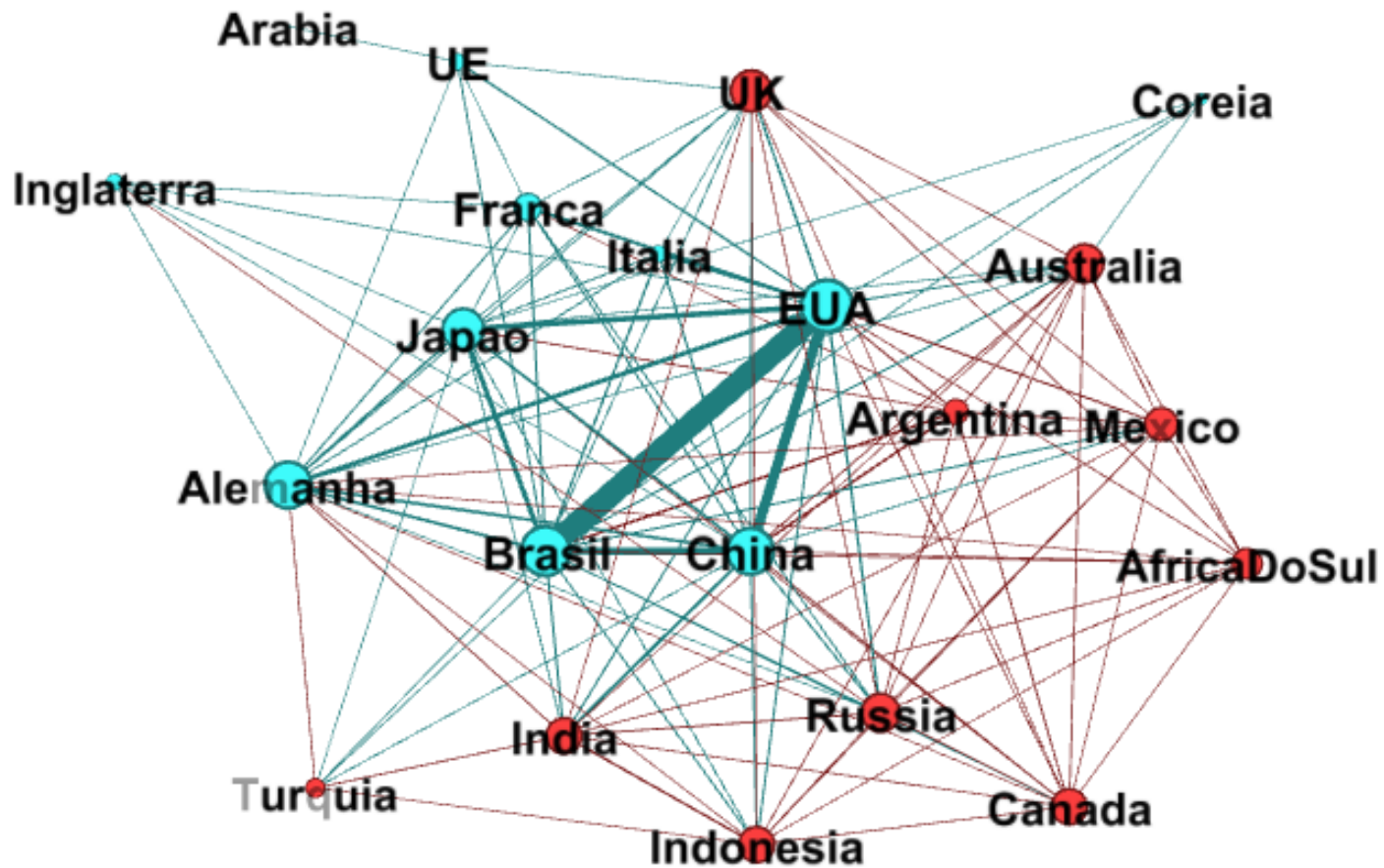
Coleta de informações macroeconômicas: inflação, PIB, câmbio e emprego.

Fontes: IBGE, BACEN e FIPE

Consolidação da base de dados adquiridos e desenvolvimento do modelo utilizando regressão linear múltipla

# Aplicações de Text Mining

## G20 em notícias na WEB

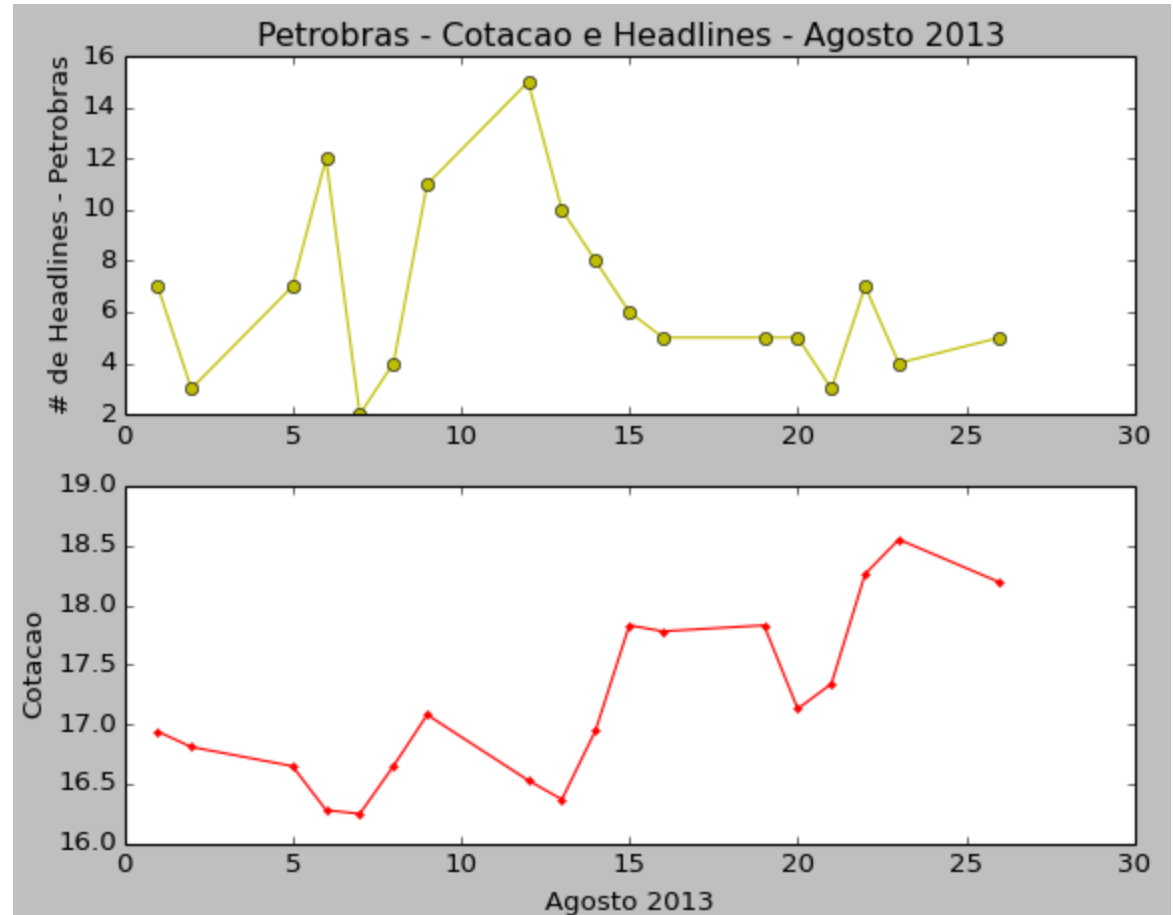


{termo: 'petrobras'}

# Aplicações de Text Mining

## Cotação – PETR4

```
--- CALCULANDO A CORRELACAO ---  
Imediata: -0.347503723053  
Deslocada em 1 dia: -0.428500698149  
Deslocada em 2 dias: -0.184363042478
```

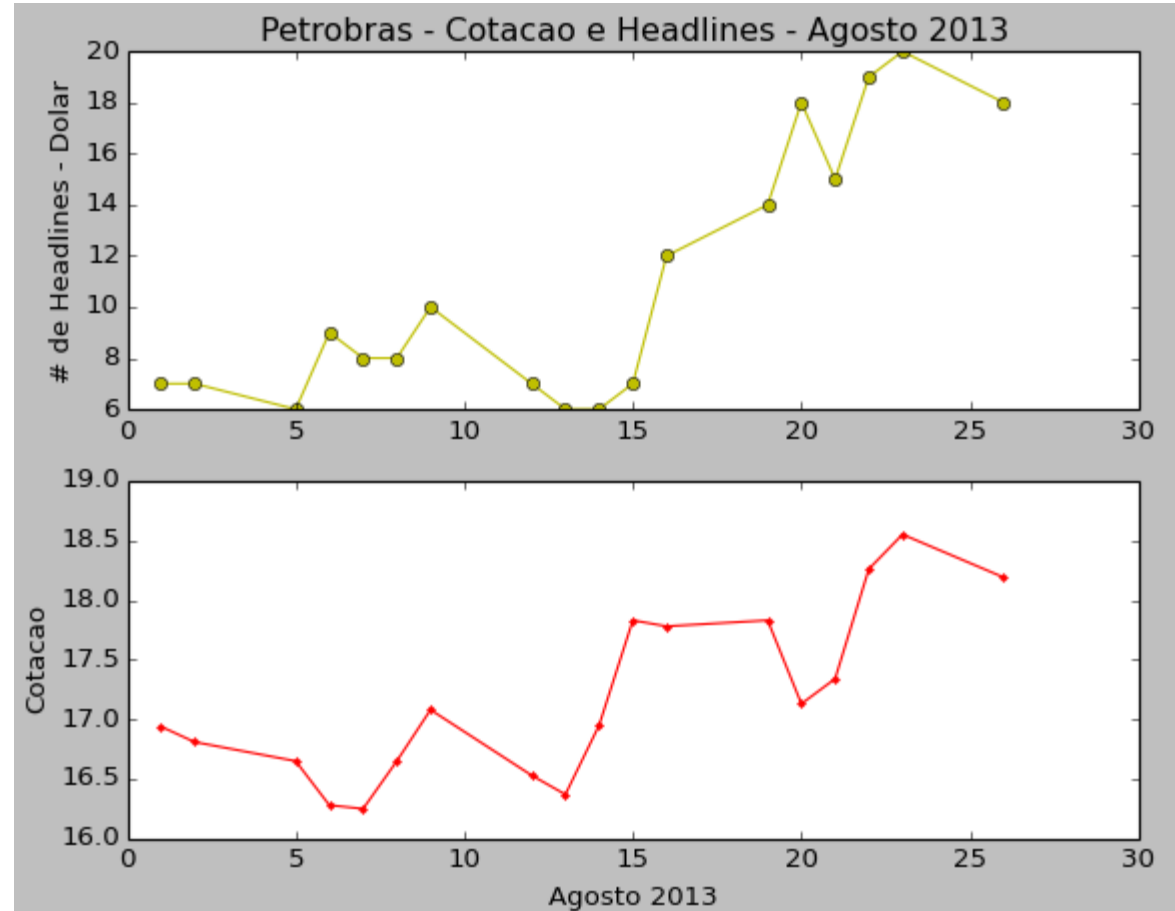


{termo: 'dolar'}

# Aplicações de Text Mining

## Cotação – PETR4

```
--- CALCULANDO A CORRELACAO ---  
Imediata: 0.767748436969  
Deslocada em 1 dia: 0.648694139053  
Deslocada em 2 dias: 0.585659671917
```

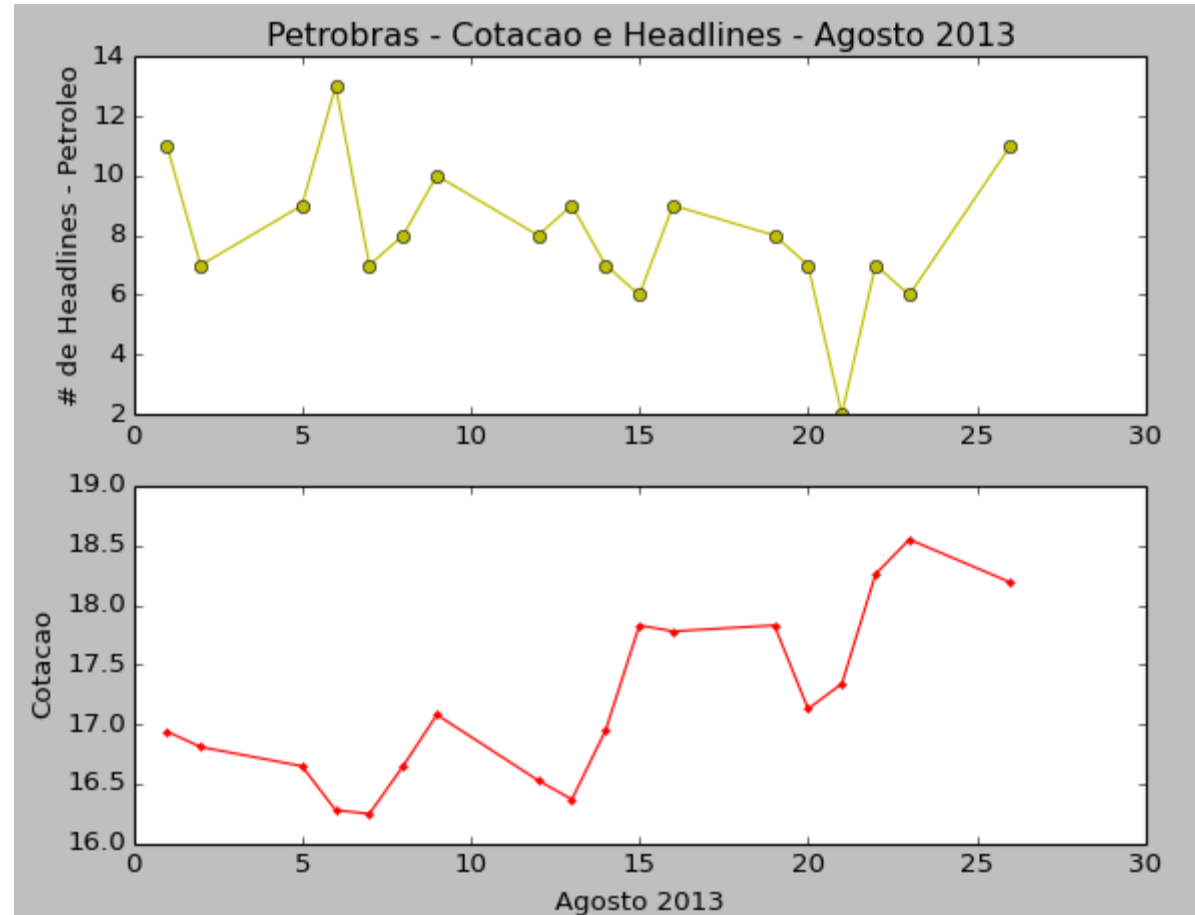


{termo: 'petroleo'}

# Aplicações de Text Mining

## Cotação – PETR4

```
--- CALCULANDO A CORRELACAO ---  
Imediata: -0.257928600463  
Deslocada em 1 dia: -0.651445314419  
Deslocada em 2 dias: -0.638486040372
```

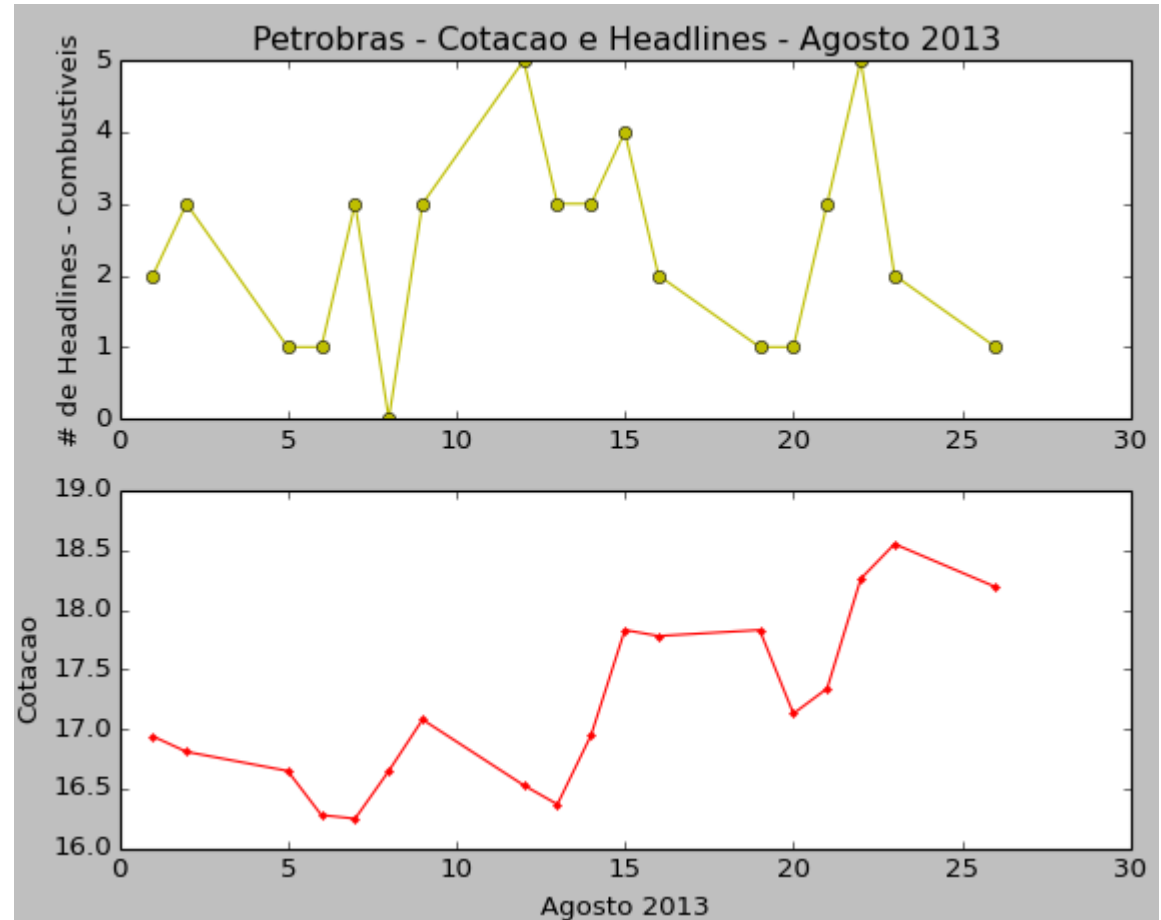


{termo: 'combustíveis'}

# Aplicações de Text Mining

## Cotação – PETR4

```
--- CALCULANDO A CORRELACAO ---  
Imediata: 0.0642411850085  
Deslocada em 1 dia: 0.260771356305  
Deslocada em 2 dias: 0.335367726191
```

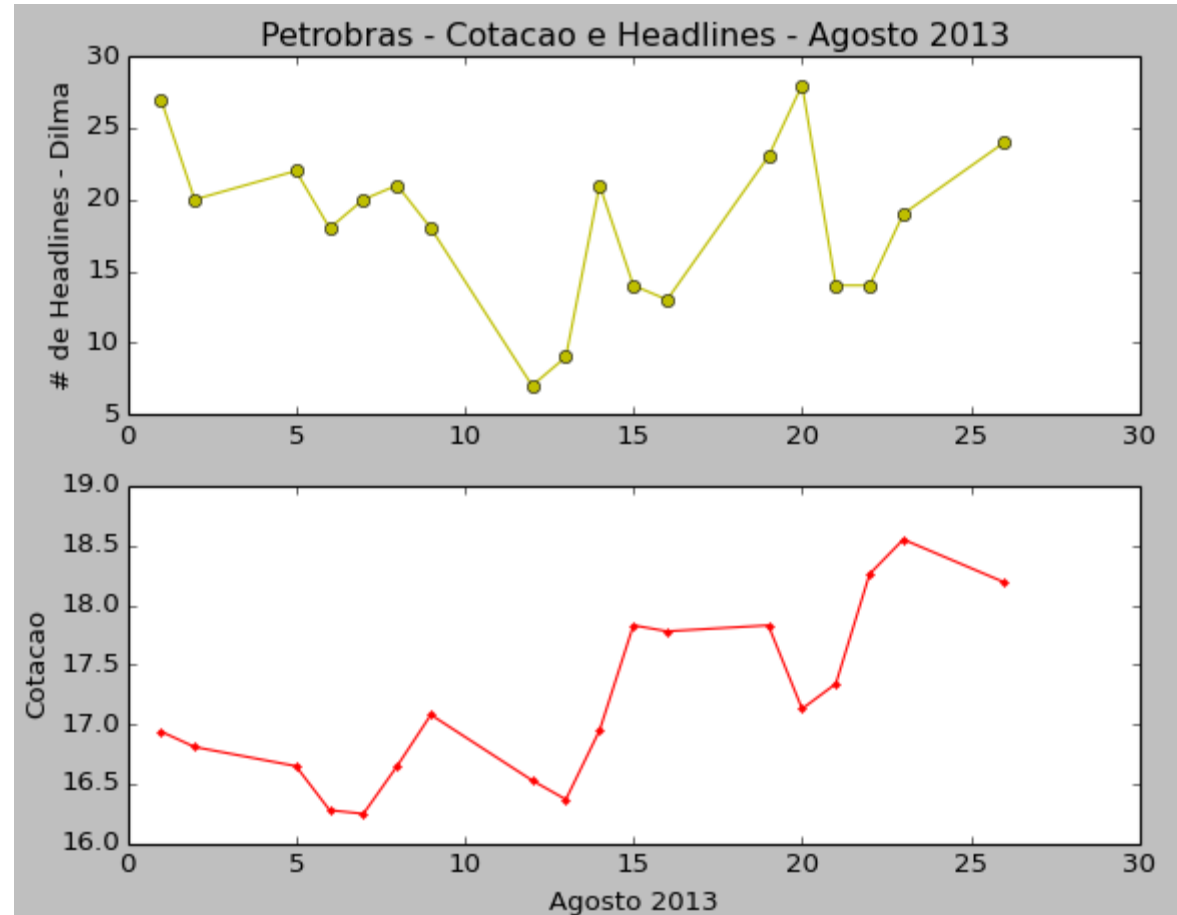


{termo: 'dilma'}

# Aplicações de Text Mining

## Cotação – PETR4

```
--- CALCULANDO A CORRELACAO ---  
Imediata: 0.068617988548  
Deslocada em 1 dia: -0.140796392816  
Deslocada em 2 dias: -0.212488626674
```





# DashboardTableau

The Tableau logo is centered on the slide. It features a circular arrangement of ten plus signs in orange, teal, red, and blue. Below this graphic, the word "tableau" is written in a dark blue, lowercase, sans-serif font, with a registered trademark symbol (®) to the right.

# Tableau

## Primeiro contato com o Tableau Public



### Create a viz now

1. Open your data (.xls or .txt)
2. Drag & Drop to visualize

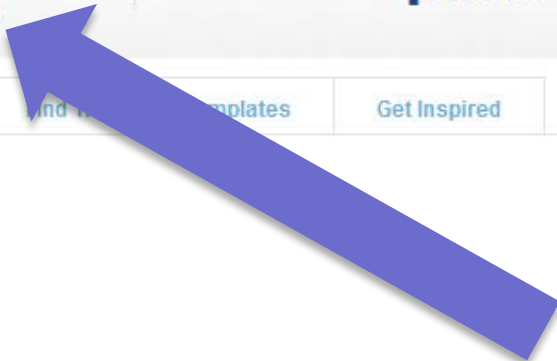
► Open Data

Open. Create. Share.

► Get Started

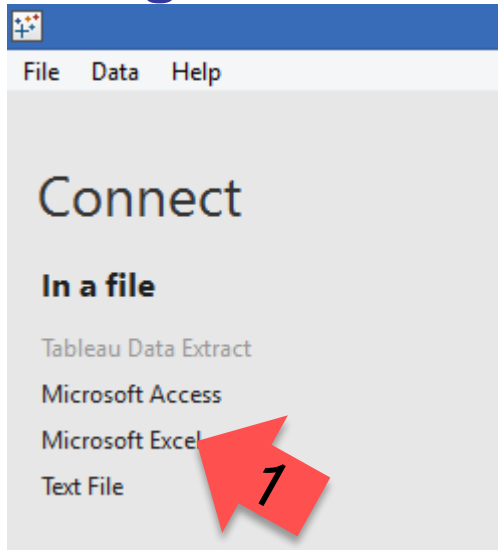
Find a template

Get Inspired



# Tableau

## Carregando a base Adventure Works Cycles




### Adventure\_Works\_Cycles

Connected to Excel

#### Workbook

Adventure\_Works\_Cycles.xlsx

#### Sheets

 SALES\_DATA



*Drag sheets here*

Copy



Go to Worksheet

# Tableau

## Primeira visualização

SALES\_DATA



Go to Worksheet

Copy

☐ Show hidden fields

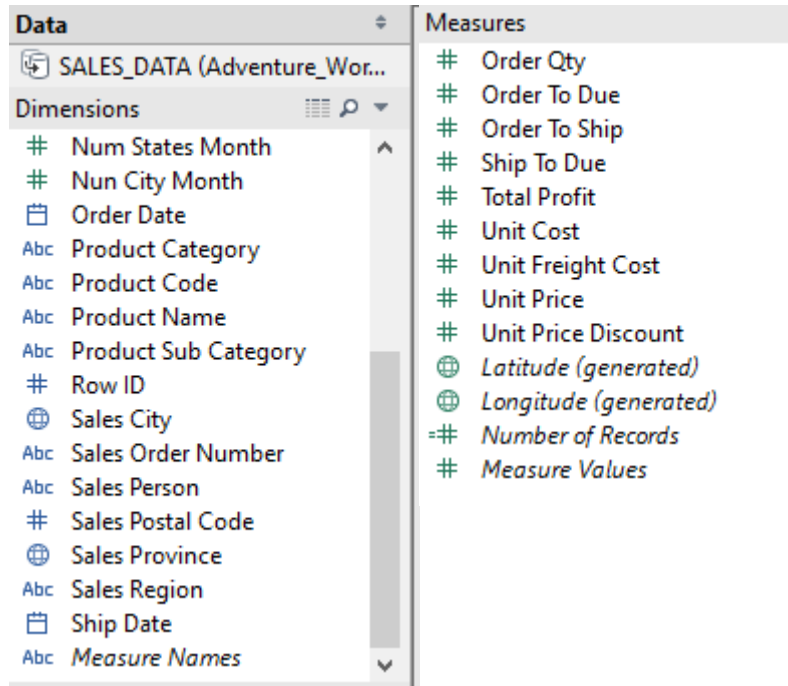
Rows 10,000

Row ID #	Sales Order Number Abc	Order Date 📅	Due Date 📅	Ship Date 📅	Sales Person Abc	Sales Region Abc	Sales Province 🌐	Sales City 🌐
1	SO43671	01/07/2005	13/07/2005	08/07/2005	David Campbell	United States	Washington	Bellevue
2	SO43671	01/07/2005	13/07/2005	08/07/2005	David Campbell	United States	Washington	Bellevue
3	SO43671	01/07/2005	13/07/2005	08/07/2005	David Campbell	United States	Washington	Bellevue
4	SO43671	01/07/2005	13/07/2005	08/07/2005	David Campbell	United States	Washington	Bellevue
5	SO43671	01/07/2005	13/07/2005	08/07/2005	David Campbell	United States	Washington	Bellevue
6	SO43671	01/07/2005	13/07/2005	08/07/2005	David Campbell	United States	Washington	Bellevue
7	SO43671	01/07/2005	13/07/2005	08/07/2005	David Campbell	United States	Washington	Bellevue

# Tableau

## WorkSheet – Price & Profits

Arraste os campos  
para o painel



**Data**

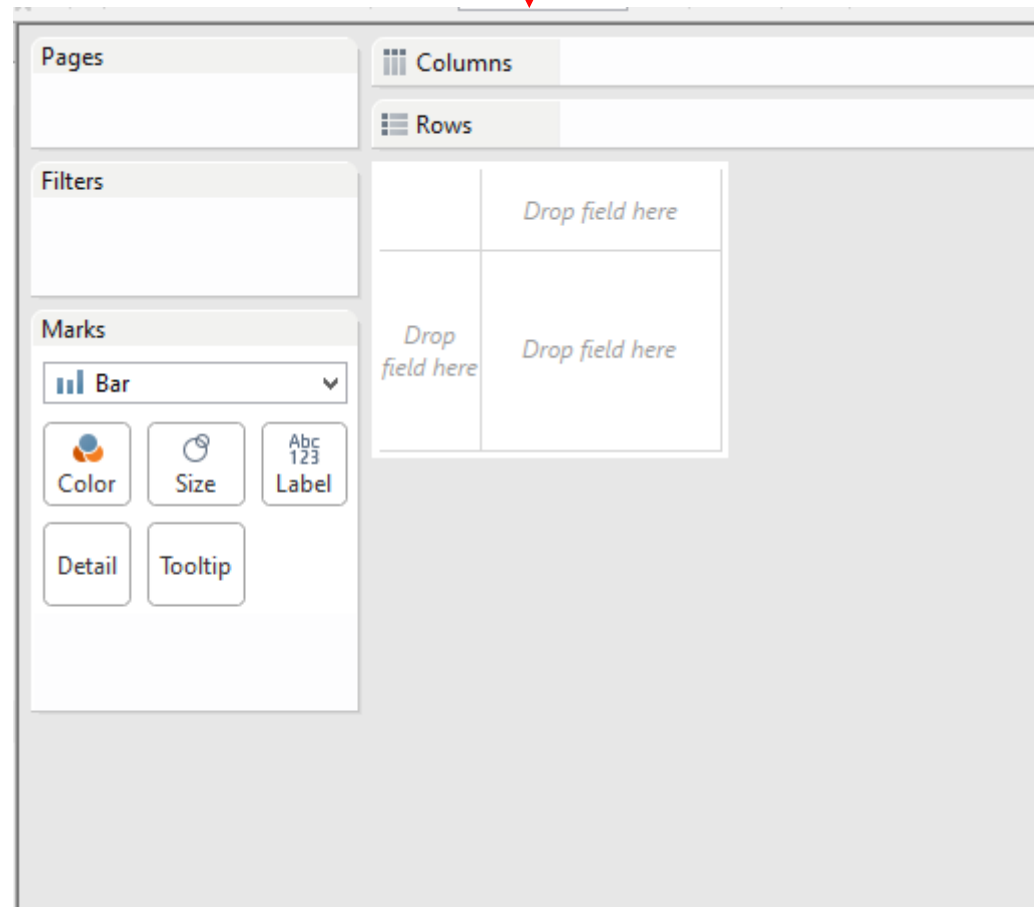
SALES\_DATA (Adventure\_Wor...

**Dimensions**

- # Num States Month
- # Nun City Month
- Order Date
- Product Category
- Product Code
- Product Name
- Product Sub Category
- # Row ID
- Sales City
- Sales Order Number
- Sales Person
- Sales Postal Code
- Sales Province
- Sales Region
- Ship Date
- Measure Names

**Measures**

- # Order Qty
- # Order To Due
- # Order To Ship
- # Ship To Due
- # Total Profit
- # Unit Cost
- # Unit Freight Cost
- # Unit Price
- # Unit Price Discount
- Latitude (generated)
- Longitude (generated)
- =# Number of Records
- # Measure Values



**Pages**

**Columns**

**Rows**

**Filters**

**Marks**

Bar

Color

Size

Label

Detail

Tooltip

Drop field here

Drop field here

Drop field here

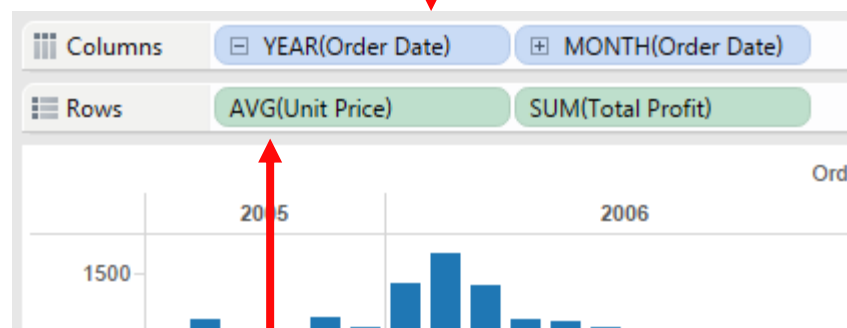
# Tableau

## WorkSheet – Price & Profits

Data	
SALES_DATA (Adventure_Wor...	
Dimensions	
#	Num States Month
#	Nun City Month
📅	Order Date
Abc	Product Category
Abc	Product Code
Abc	Product Name
Abc	Product Sub Category
#	Row ID
🌐	Sales City
Abc	Sales Order Number
Abc	Sales Person
#	Sales Postal Code
🌐	Sales Province
Abc	Sales Region
📅	Ship Date
Abc	Measure Names

Measures	
#	Order Qty
#	Order To Due
#	Order To Ship
#	Ship To Due
#	Total Profit
#	Unit Cost
#	Unit Freight Cost
#	Unit Price
#	Unit Price Discount
🌐	Latitude (generated)
🌐	Longitude (generated)
=#	Number of Records
#	Measure Values

Arraste os campos para o painel



SUM(Unit Price) ▼

Filter...

Show Quick Filter

Format...

✓ Show Header

✓ Include in Tooltip

Dimension

Attribute

• Measure (Sum) ▶

Discrete

• Continuous

Sum

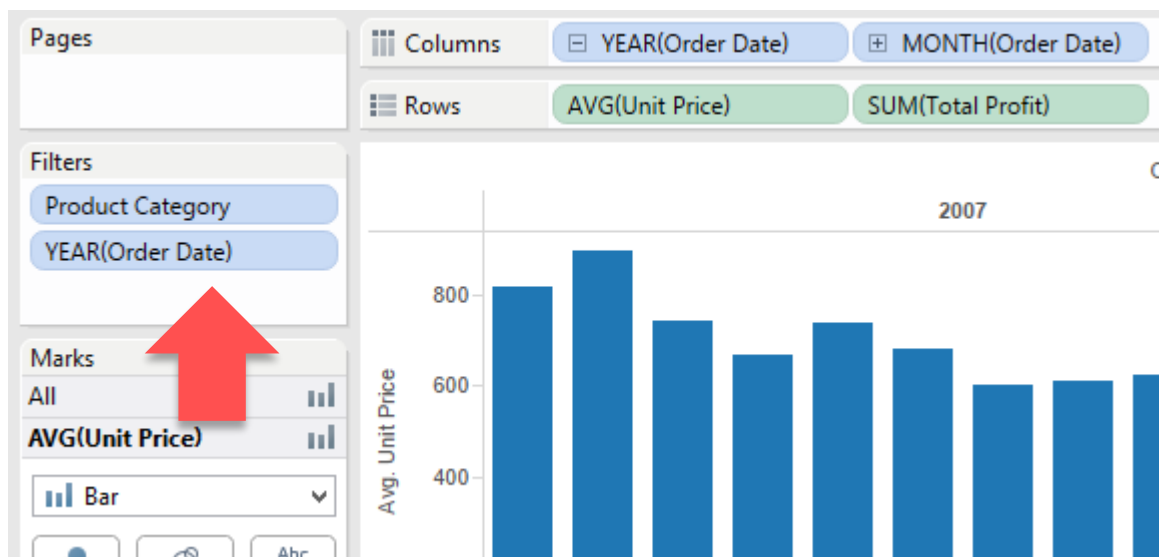
Average

Median

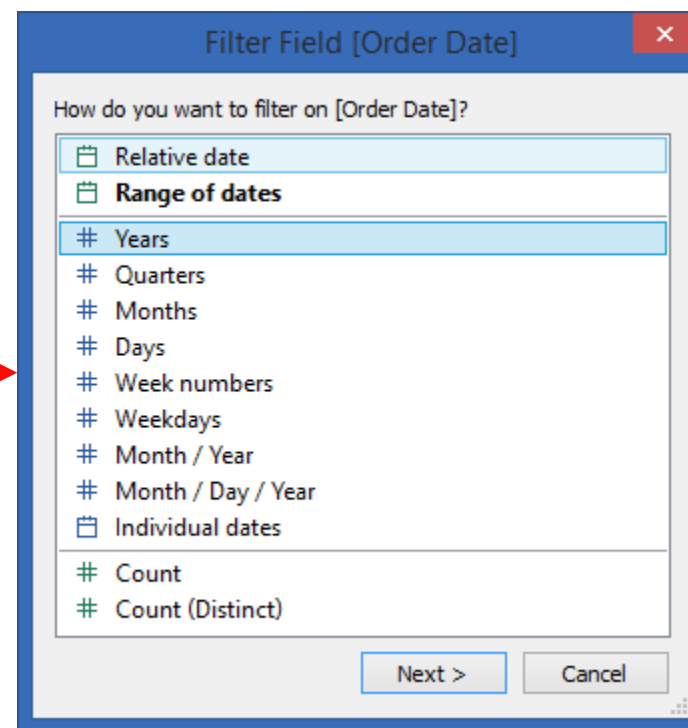
Altere a medida de Unit Price para AVG - Média

# Tableau

## WorkSheet – Price & Profits



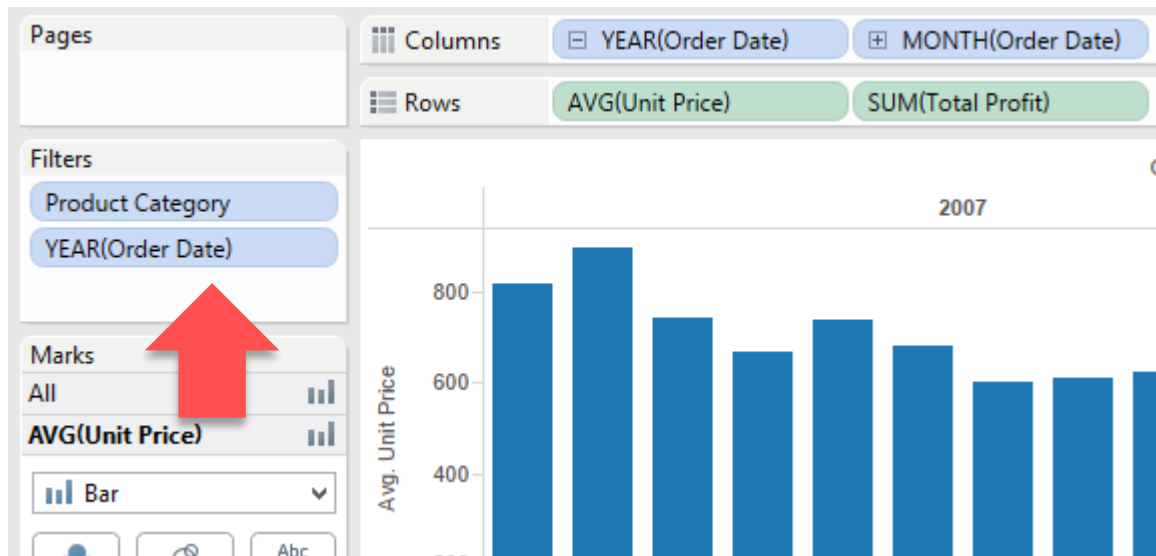
Crie os filtros de categoria de produto e ano (2007 e 2008)



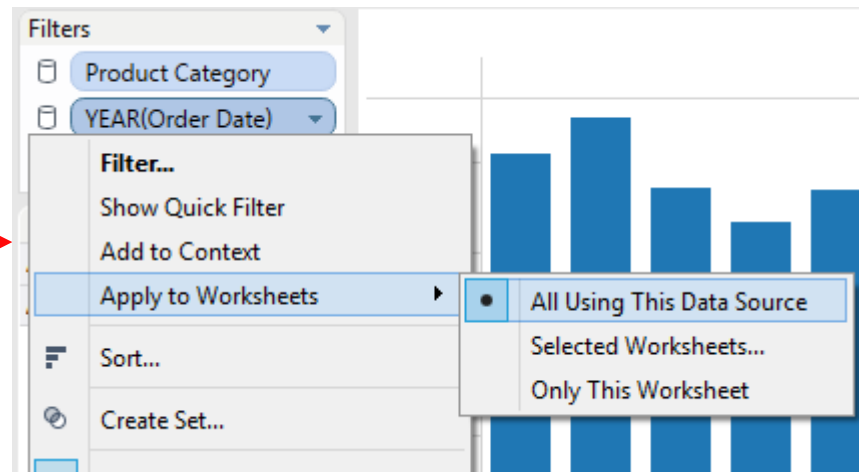


# Tableau

## WorkSheet – Price & Profits

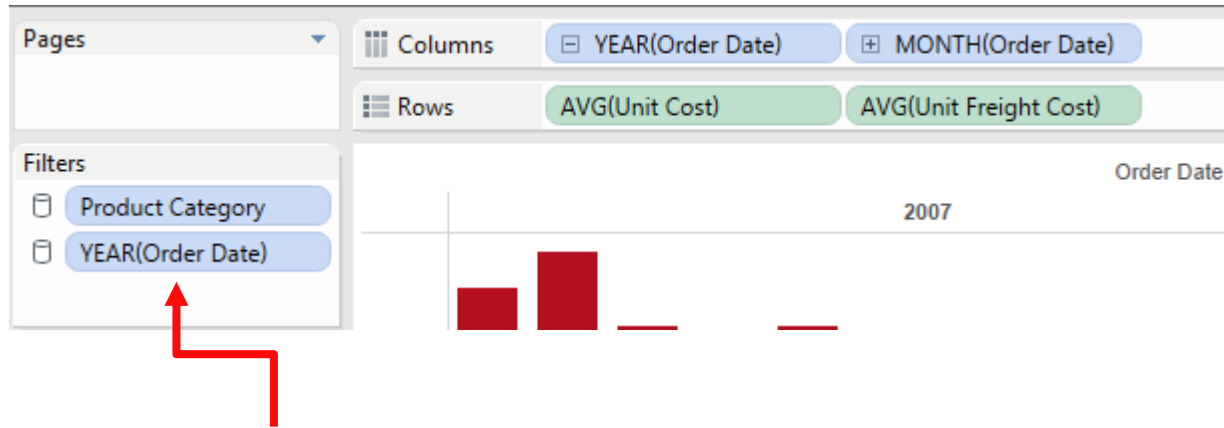


Aplique-os para toda a base



# Tableau

## WorkSheet – Costs



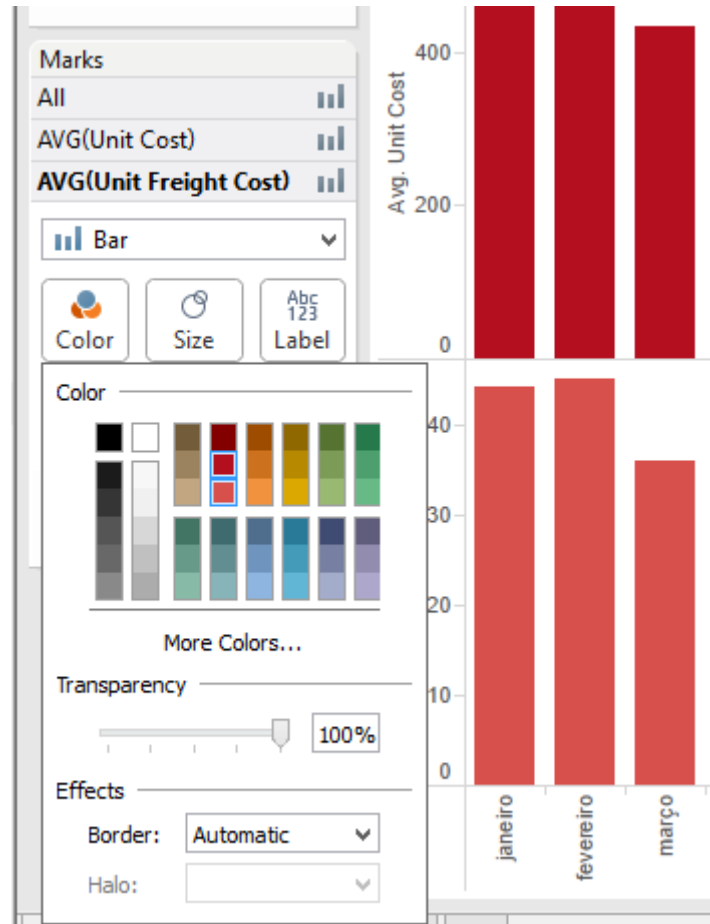
Repare que seus filtros já estão aqui!



Selecione estas dimensões e ajuste Unit Cost e Freight Cost para media (AVG)

# Tableau

## WorkSheet – Costs

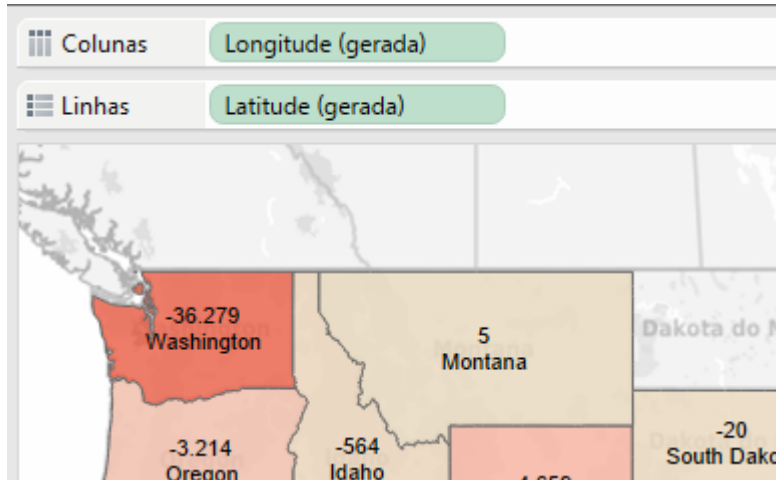


Utilize as configurações em “Marks”

Altere as cores Para vermelho

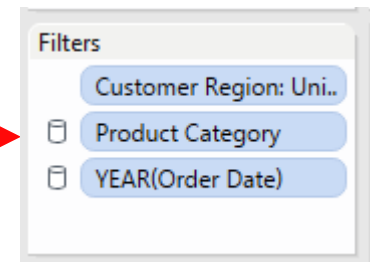
# Tableau

## Worksheet – Geo Profits



**Selecione as  
dimensões de  
latitude e longitude**

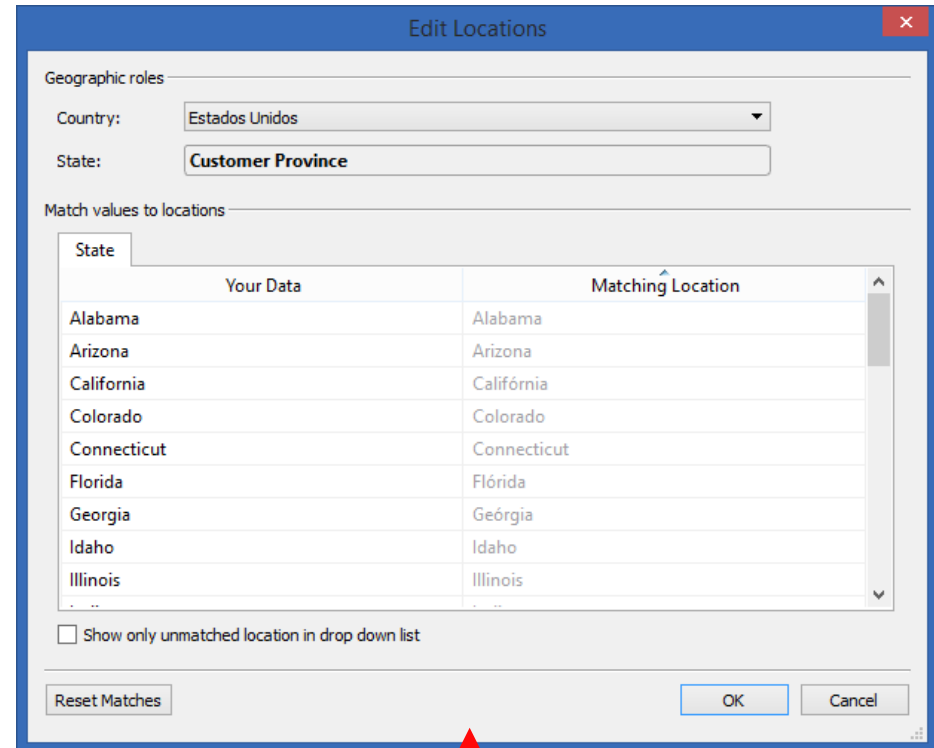
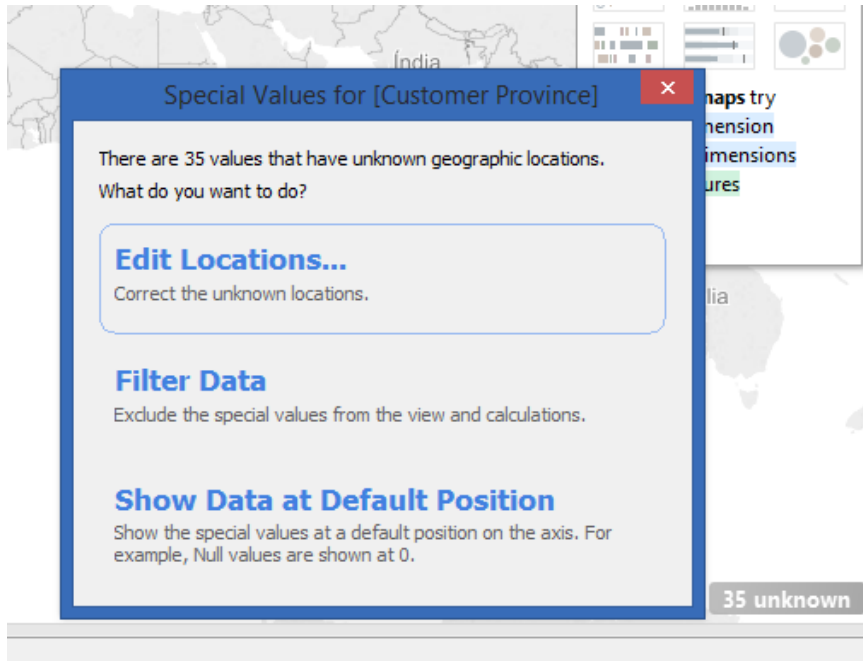
**Acrescente o filtro  
“Customer Region”**



**Selecione apenas  
“United States”**

# Tableau

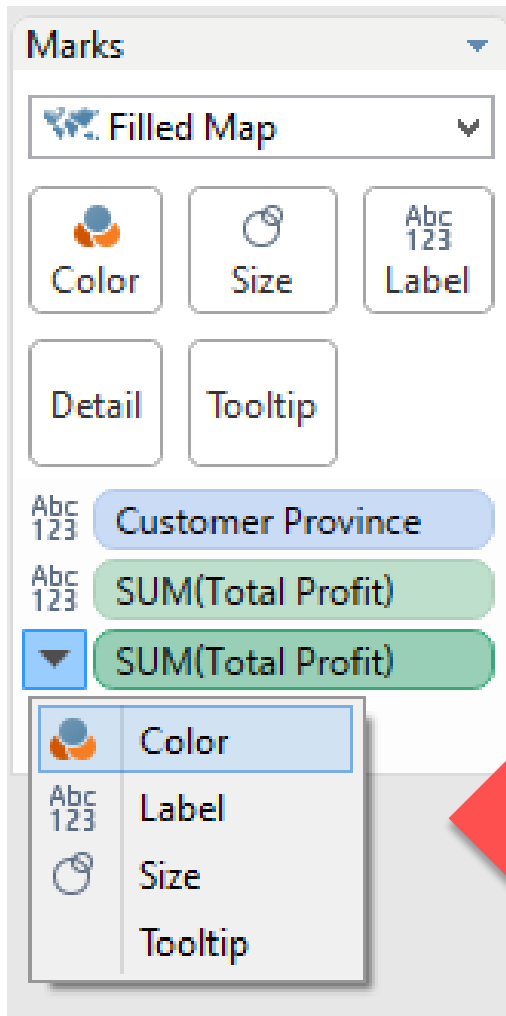
## WorkSheet – Geo Profits



Edite o país da  
localização para EUA

# Tableau

## WorkSheet – Geo Profits

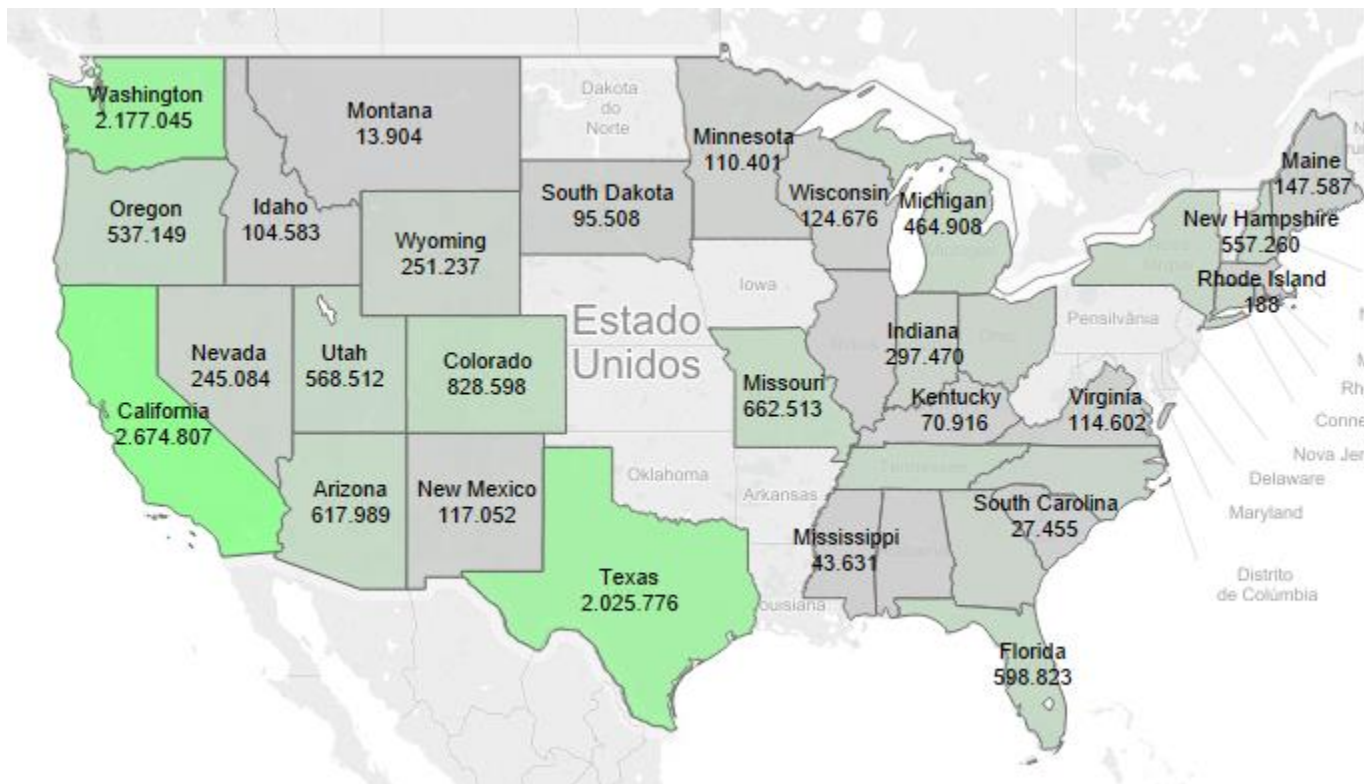


Adicione 3 elementos  
ao card Marks

Configure a exibição  
de rótulos e cores

## WorkSheet – Geo Profits

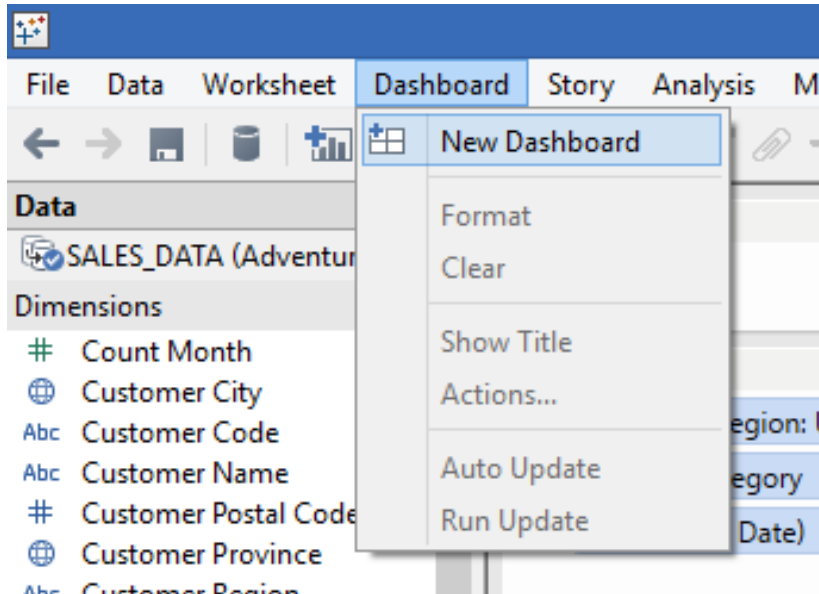
Seu mapa deve estar assim!



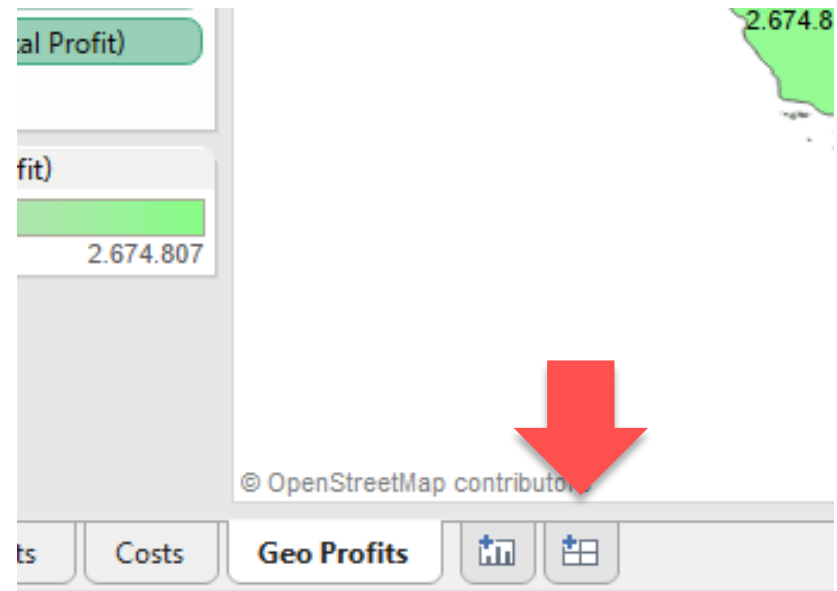


# Tableau

## Criando o dashboard

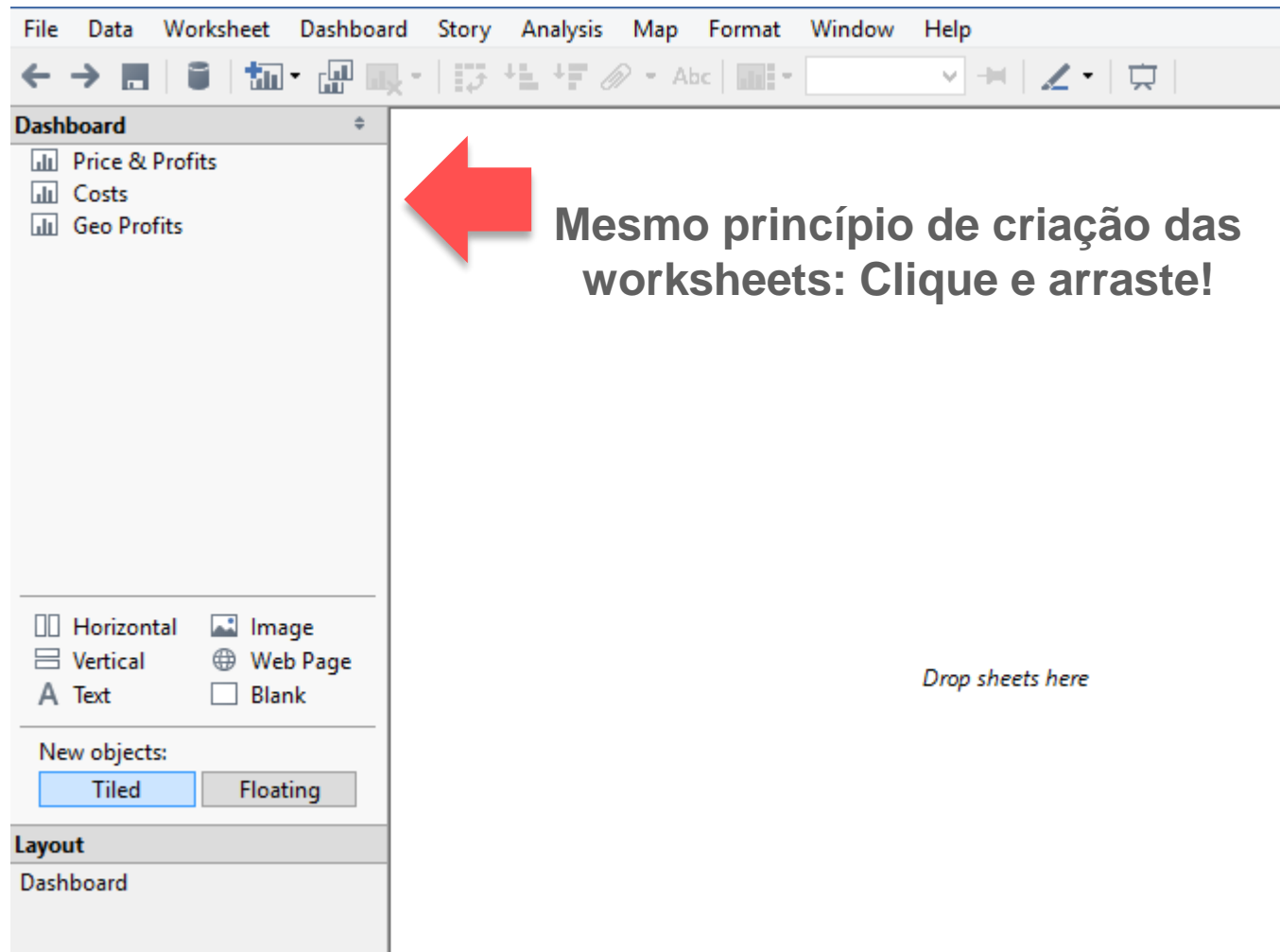


Ou

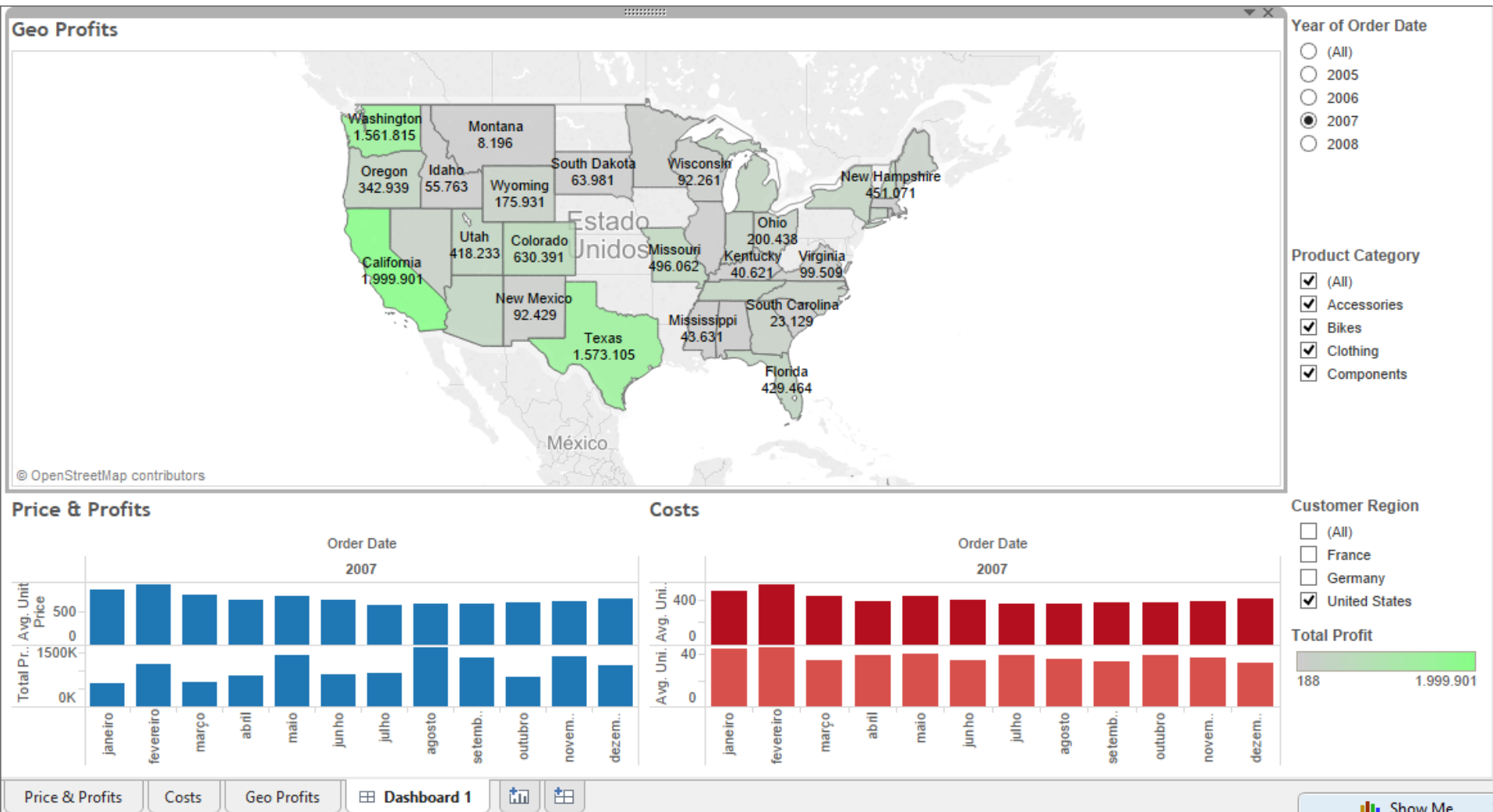


# Tableau

## Criando o dashboard



## Criando o dashboard



# **Business Intelligence**