

Business Intelligence

R, Python, Twitter e Tableau para Business Intelligence

Prof. Leandro Guerra

E-mail: leandro.guerra@outspokenmarket.com.br IG: @leandrowar

R



Python



R

Quem usa R?

The New York Times – visualização de dados

Google – mais de 500 funcionários usam R

O **FDA** – para testes estatísticos de novas drogas.

John Deere - otimização da produção de novos tratores

The National Weather Service - predição dos impactos de enchentes

Zillow - modelagem do preço de imóveis

The **Consumer Financial Protection Bureau** – modelagem financeira

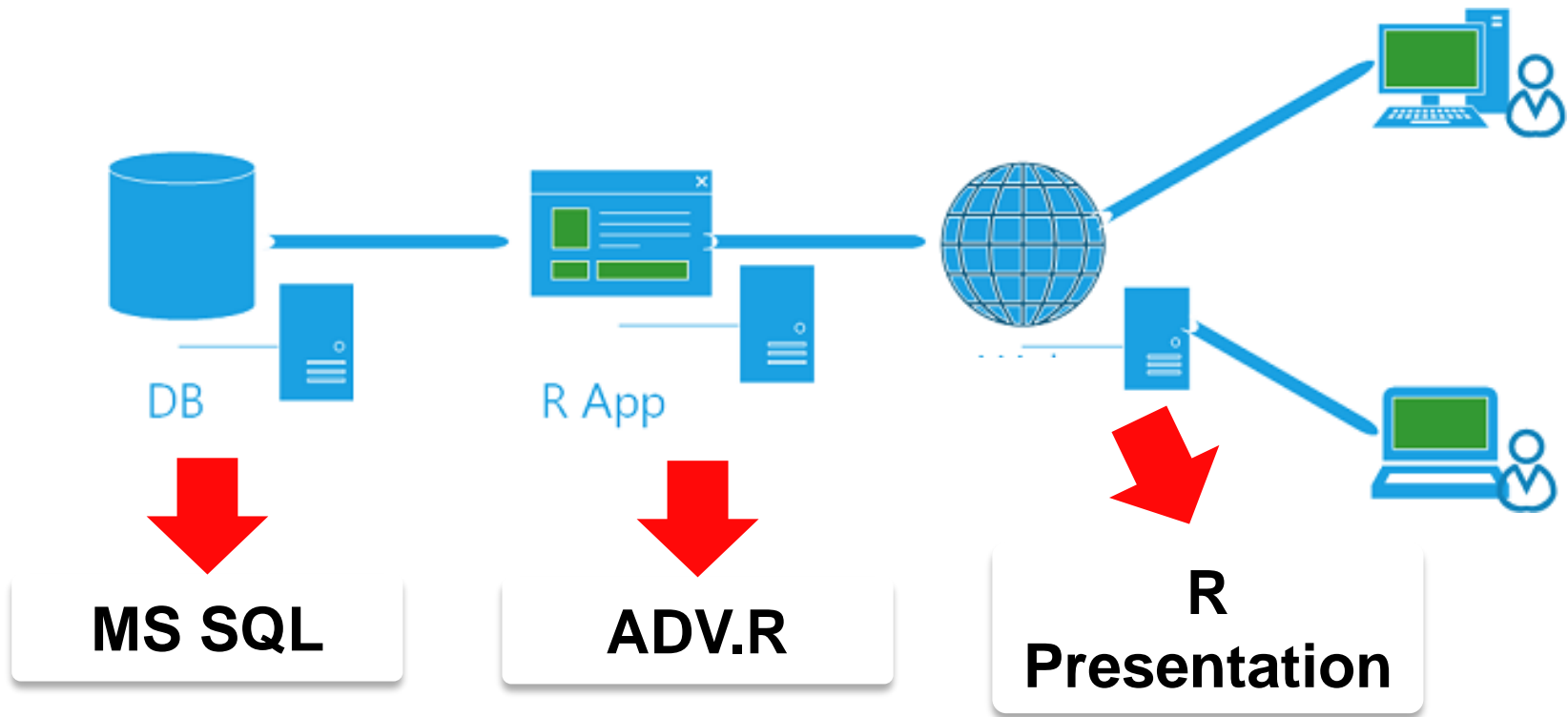
Twitter - data science em suas bases de dados

FourSquare – base de seu sistema de recomendação

Facebook - modelagem do comportamento de seus usuários

R

Arquitetura de extração e transformação dos dados



R

Considerações importantes

Lembrando que R é:

- Uma linguagem de programação
- Utilizada para fins de data mining e estatística

Mas:

- Data mining e estatística fazem parte do universo de BI
- Para a tomada de decisão, é necessária a mineração, manipulação, tratamento e dos dados



R

Engine for Business Intelligence

Extração

- **DBI** - driver nativo para múltiplos fornecedores.
- **RODBC** - Conexão ODBC.
- **RJDBC** - Conexão JDBC.
- Suporte a vários outros formatos: xlsx, xml, json, sas, spss, stata.

Visualização

Indo além de um ETL, o R permite que você crie inúmeras maneiras de divulgação dos dados

- **Shiny** - Web Application Framework para R.
- **R presentation** – Sua apresentação no R

Manipulação

data.table - permita a utilização de uma sintaxe SQL



R

SQL e Sales database

Importando a base SALES para o
MS SQL



MS SQL

Tutorial de importação no SQL

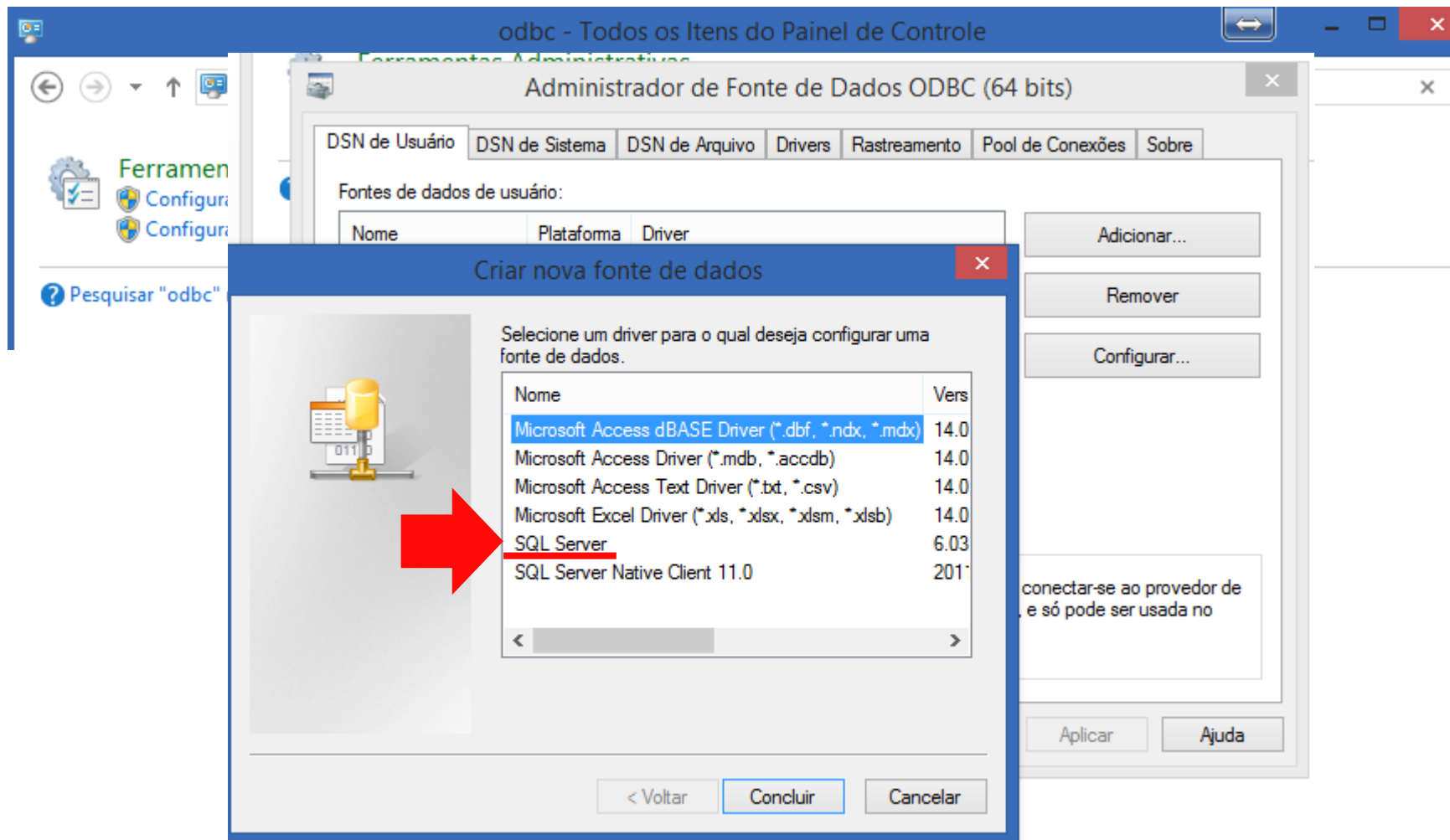
Download da base utilizada no exemplo

Todas as informações estão no
www.artedossdados.com.br

R

Conexão com o MS SQL – R ODBC

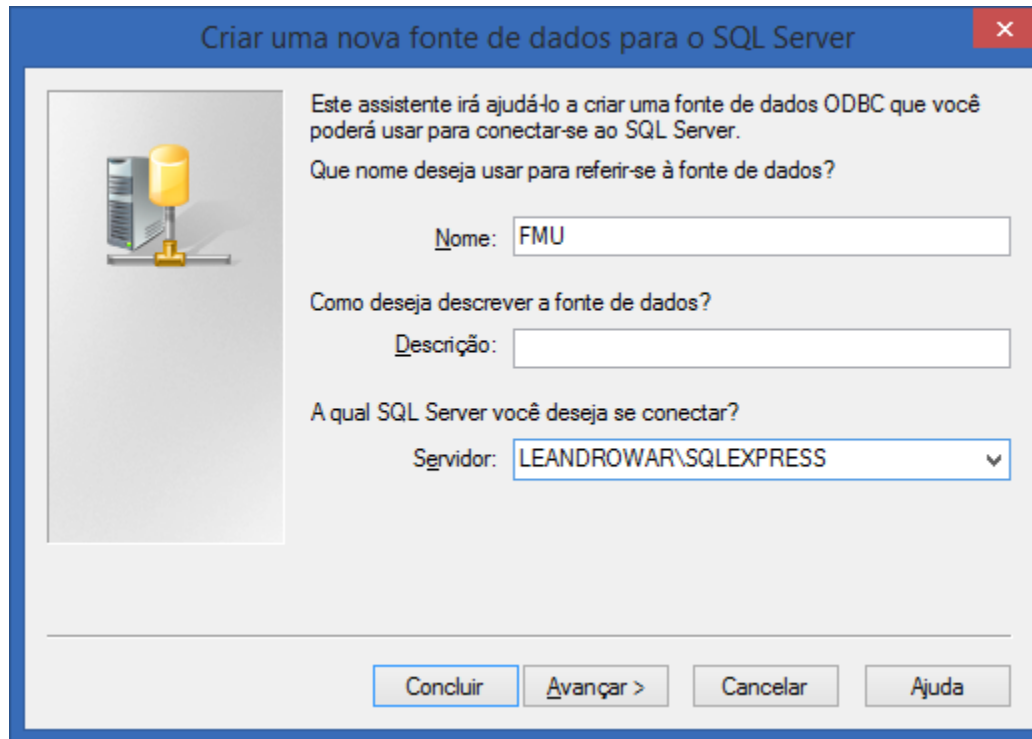
Escolha SQL Server



R

Conexão com o MS SQL – R ODBC

Crie um nome para sua fonte e
selecione o servidor



Criar uma nova fonte de dados para o SQL Server

Este assistente irá ajudá-lo a criar uma fonte de dados ODBC que você poderá usar para conectar-se ao SQL Server.

Que nome deseja usar para referir-se à fonte de dados?

Nome:

Como deseja descrever a fonte de dados?

Descrição:

A qual SQL Server você deseja se conectar?

Servidor:

Concluir Avançar > Cancelar Ajuda

R

Conexão com o MS SQL – R ODBC

Siga os passos

1

Criar uma nova fonte de dados para o SQL Server

Como o SQL Server deve verificar a autenticidade da ID de login?

☒ Com autenticação do Windows NT usando ID de login na rede.

☐ Com autenticação do SQL Server usando ID de login e senha inseridos pelo usuário.

Para alterar a biblioteca de rede usada na comunicação com o SQL Server, clique em 'Configuração do cliente'.

☒ Conectar-se ao SQL Server para obter as configurações padrão das opções de configuração adicionais.

ID de login: Leandro

Senha:

< Voltar Avançar > Cancelar Ajuda

3

Criar uma nova fonte de dados para o SQL Server

☐ Alterar o idioma das mensagens de sistema do SQL Server para: English

☐ Usar criptografia de alta segurança para os dados

☒ Executar conversão dos dados de caracteres

☐ Usar configurações regionais ao exibir unidade monetária, números, datas e hora.

☐ Salvar consultas longas no arquivo de log:

C:\Users\Leandro\AppData\Local\Temp\QUERY.L Procurar...

Tempo de consulta longa (milissegundos): 30000

☐ Registrar estatísticas do driver ODBC no arq. de log:

C:\Users\Leandro\AppData\Local\Temp\STATS.L Procurar...

< Voltar Concluir Cancelar Ajuda

2

Criar uma nova fonte de dados para o SQL Server

☒ Alterar o banco de dados padrão para: FMU

☐ Anexar o nome de arquivo do banco de dados:

☒ Usar identificadores entre aspas ANSI.

☒ Usar nulos, preenchimentos e avisos ANSI.

☐ Usar o SQL Server de reserva se o SQL Server principal não estiver disponível.

4

Instalação do ODBC para Microsoft SQL Server

Uma nova fonte de dados ODBC será criada com a seguinte configuração:

Driver ODBC para Microsoft SQL Server versão 06.03.9600

Nome da fonte de dados: FMU

Descrição da fonte de dados:

Servidor: LEANDROWAR\SQLEXPRESS

Banco de Dados: FMU

Idioma: (Default)

Converter Dados de Caracteres: Yes

Registrar Consultas Longas em Log: No

Registrar Estatísticas do Driver em Log: No

Usar configurações regionais: No

Opção de instruções preparadas: Ignorar procedimentos temporários ao desconectar

Usar servidor de reserva: No

Usar identificadores entre aspas ANSI: Yes

Usar nulos, preenchimentos e avisos ANSI: Yes

Criptografia de dados: No

Testar fonte de dados... OK Cancelar

R

O CÓDIGO NO R



ADV.R

```
1 #Selecionando o diretorio de trabalho
2 setwd("C:/Users/Leandro/Google Drive/FMU")
3
4 #Instalando e carregado a biblioteca RODBC
5 install.packages("RODBC")
6 library(RODBC)
7
8 #Abrindo uma conexao com o banco
9 minha_conexao <- odbcConnect("FMU", uid="LEANDROWAR/Leandro", pwd="")

11 #Escrevendo as consultas
12 #Conhecendo quem são os vendedores
13 vendedores <- sqlQuery(minha_conexao,
14   "
15     SELECT DISTINCT(Sales_Person) AS VENDEDORES
16     FROM SALES_DATA
17   ")
18
19 #Conhecendo quais são os estados com maior volume de venda
20 maior_venda <- sqlQuery(minha_conexao,
21   "
22     SELECT Customer_Province, SUM(Total_Profit) AS LUCRO_TOTAL
23     FROM SALES_DATA
24     GROUP BY Customer_Province
25     ORDER BY LUCRO_TOTAL DESC
26   ")
```

R

O CÓDIGO NO R



ADV.R

```
29 #Conhecendo o histórico de vendas por ano
30 #dos tres maiores estados
31 california <- sqlQuery(minha_conexao,
32     "
33     SELECT YEAR(OrderDate) AS ANO, COUNT(Product_Name) AS QTDE
34     FROM SALES_DATA
35     WHERE Customer_Province = 'California'
36     GROUP BY YEAR(OrderDate), Customer_Province
37     ORDER BY ANO
38     ")
39
40 washington <- sqlQuery(minha_conexao,
41     "
42     SELECT YEAR(OrderDate) AS ANO, COUNT(Product_Name) AS QTDE
43     FROM SALES_DATA
44     WHERE Customer_Province = 'Washington'
45     GROUP BY YEAR(OrderDate), Customer_Province
46     ORDER BY ANO
47     ")
48
49 texas <- sqlQuery(minha_conexao,
50     "
51     SELECT YEAR(OrderDate) AS ANO, COUNT(Product_Name) AS QTDE
52     FROM SALES_DATA
53     WHERE Customer_Province = 'Texas'
54     GROUP BY YEAR(OrderDate)
55     ORDER BY ANO
56     ")
```

R

O CÓDIGO NO R



ADV.R

```
58 #Conhecendo o lucro médio da empresa por ano
59 lucro_medio <-sqlQuery(minha_conexao,
60     "
61     SELECT YEAR(OrderDate) AS ANO, AVG(Total_Profit) AS LUCRO_MEDIO
62     FROM SALES_DATA
63     WHERE OrderDate >= '2005-07-01 00:00:00.000'
64     GROUP BY YEAR(OrderDate)
65     ORDER BY ANO
66     ")
67
68 #Entendendo a participação de cada categoria de produto nas vendas brutas
69 participacao <-sqlQuery(minha_conexao,
70     "
71     SELECT Product_Category, SUM(UnitPrice*OrderQty) as RECEITA_BRUTA
72     FROM SALES_DATA
73     WHERE Product_Category IS NOT NULL
74     GROUP BY Product_Category
75     ")
76
77 #Fechando a conexão
78 close(minha_conexao)
79
80 vendedores
81 maior_venda #Três Maiores - California, Washington, Texas
82 california
83 lucro_medio
84 participacao$RECEITA_BRUTA <- as.numeric(as.character(participacao$RECEITA_BRUTA))
85 participacao$RECEITA_BRUTA
86
```

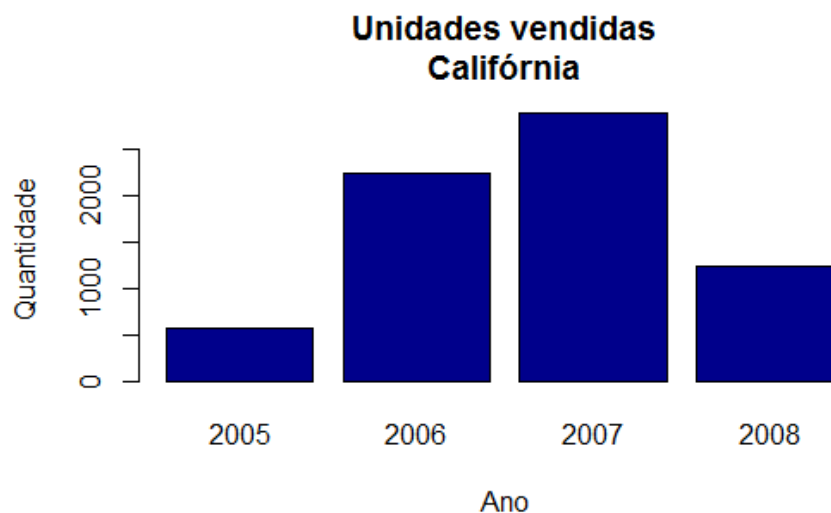
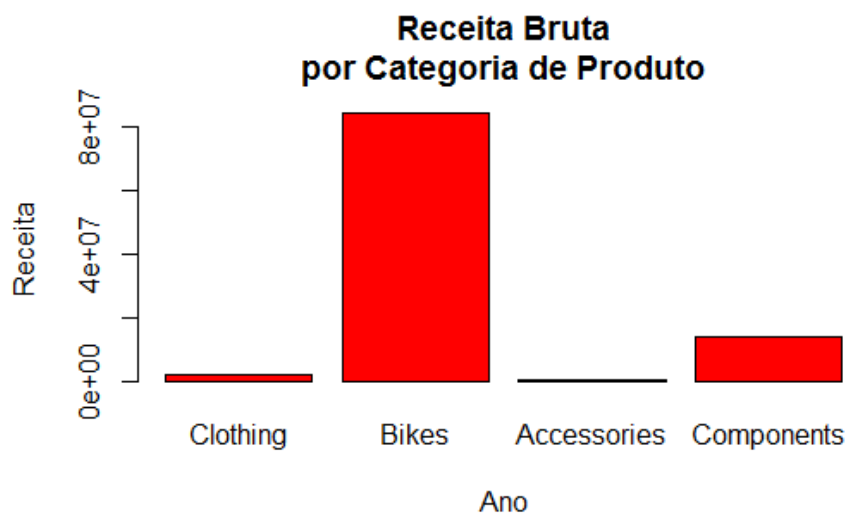
R

O CÓDIGO NO R



ADV.R

```
87 #Algumas visualizações simples
88 barplot(participacao$RECEITA_BRUTA,main="Receita Bruta\npor Categoria de Produto",
89         xlab="Ano",ylab="Receita",col=c("red"),names.arg=participacao$Product_Category)
90 barplot(lucro_medio$LUCRO_MEDIO,main="Lucro Médio\nAnual",
91         xlab="Ano",ylab="Quantidade",col=c("green"),names.arg=lucro_medio$ANO)
92 barplot(california$QTDE,main="Unidades vendidas\nCalifórnia",
93         xlab="Ano",ylab="Quantidade",col=c("darkblue"),names.arg=california$ANO)
94 barplot(washington$QTDE,main="Unidades vendidas\nWashington",
95         xlab="Ano",ylab="Quantidade",col=c("darkblue"),names.arg=washington$ANO)
96 barplot(texas$QTDE,main="Unidades vendidas\nTexas",
97         xlab="Ano",ylab="Quantidade",col=c("darkblue"),names.arg=texas$ANO)
```



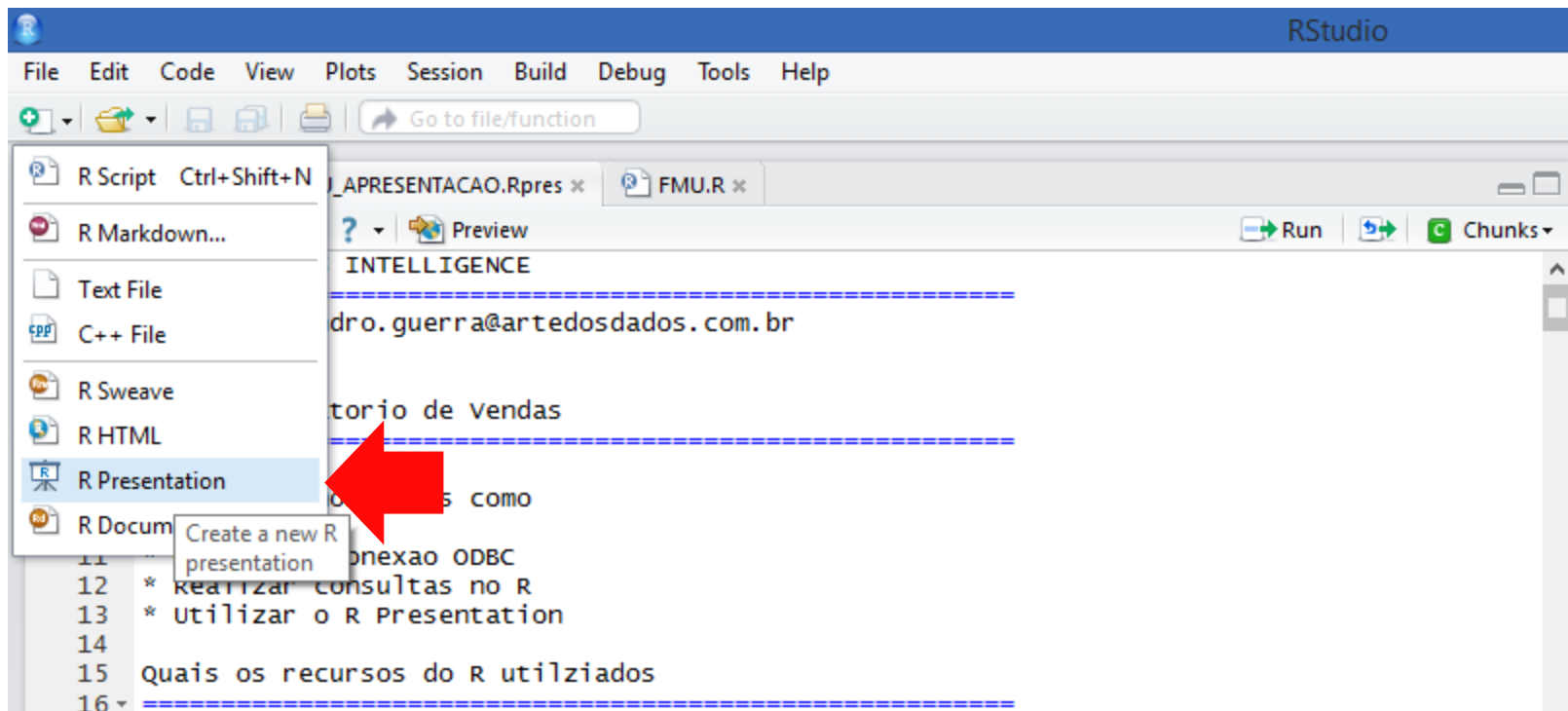
R

R Presentation



R
Presentation

Criando uma nova apresentação



R

R Presentation



R Presentation

R Markdown

```
capstone.R x FMU_APRESENTACAO.Rpres x FMU.R x
← → ABC 🔍 ? Preview Run ↺ Chunks v
1 R & BUSINESS INTELLIGENCE
2 =====
3 author: leandro.guerra@artedosedados.com.br
4 date:
5 |
6 Simples Relatorio de Vendas
7 =====
8
9 Nesta aula mostramos como
10
11 * Criar uma conexao ODBC
12 * Realizar consultas no R
13 * Utilizar o R Presentation
14
15 Quais os recursos do R utilizados
16 =====
17 Alem do R Presentation, estudamos a implantacao de um
18 codigo que utiliza a library RODBC para estabelecer
19 a conexao com um banco MS SQL
20 ```{r}
21 library(RODBC)
22 ```
23
24 Nossa base de vendas
25 =====
26 Neste slide, trazemos a receita bruta por produto
27 ```{r, echo= FALSE, message=FALSE, warning=FALSE}
28 library(RODBC)
29
30 #Abrindo uma conexao com o banco
31 minha_conexao <-odbcConnect("FMU", uid="LEANDROWAR/Leandro", pwd="")
32
```


R R Presentation

Pré-visualização



**R
Presentation**

Ainda é possível:

- Salvar como HTML
- Rpubs
- Visualizar no navegador

R



Python



Python

Quem usa Python?

Yahoo Maps – visualização de dados

Google – spider e o engine de busca

Civilization 4 – AI implementada em Python

Walt Disney Feature Animation - sistemas de animação

ABN AMRO Bank, Treasury Systems – trading system

National Weather Service - mapas e forecasts

Python

Capturando tweets no Python

```
4 import sys
5 import unicodedata
6 from TwitterSearch import *
7 import csv
8 import datetime
9 import time
10 reload(sys)
11 sys.setdefaultencoding("ascii")
12
13 #
14 def coleta_tweets(query, max = 3000):
15     #recebe o termo de consulta e a quantidade máxima de tweets por busca
16     #escreve os resultados em um arquivo csv, sem sobreposição de dados
17
18     #contador
19     i = 0
20     #recebe o termo de busca
21     search = query
22
23     #abre um arquivo csv com o nome do termo de busca
24     with open(search+'.csv', 'a') as outf: # o parametro "a" indica que não haverá sobreposição
25         #cada novo registro será adicionado ao final do arquivo
26         writer = csv.writer(outf,delimiter='|') #utiliza o | como delimitador de campo
27         writer.writerow(['user','time','tweet','latitude','longitude'])
```


Python

Capturando tweets no Python

```
55
56         writer.writerow([user, time, tweet_text, lat, long])
57         i += 1
58     if i > max:
59         return()
60
61     except TwitterSearchException as e: #apenas para mostrar o tipo de erro, caso exista
62         print(e)
63
64     query = raw_input ("Sua busca: ")
65     max_tweets = 3000
66     tempo_espera = 60 # Tempo de espera entre uma busca e outra (em segundos)
67     quantidade_buscas = 10000
68
69     #Loop para busca
70     for l in range(1,quantidade_buscas+1):
71         coleta_tweets(query, max_tweets)
72         time.sleep(tempo_espera)
73         l += 1
```

Tableau

Vamos abrir a base “protestos_tableau”

Objetivo – Exibir num mapa as informações do tweets capturados com o script em Python

+tableau++public

Create a viz now

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

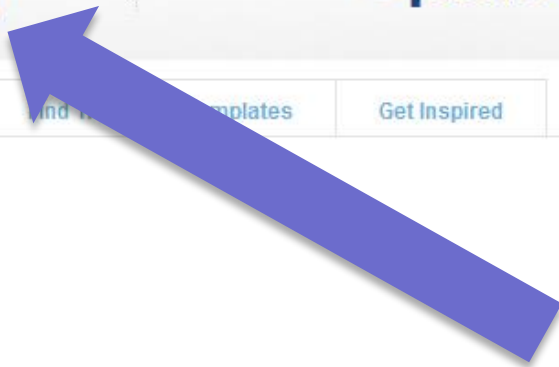
► Open Data

Open. Create. Share.

Get Started

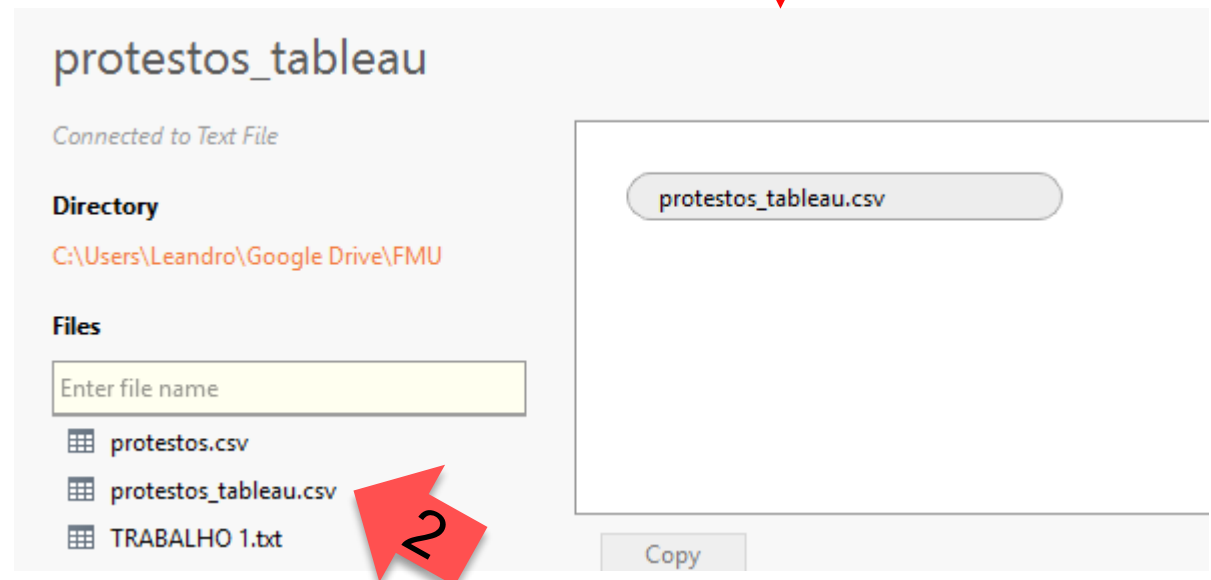
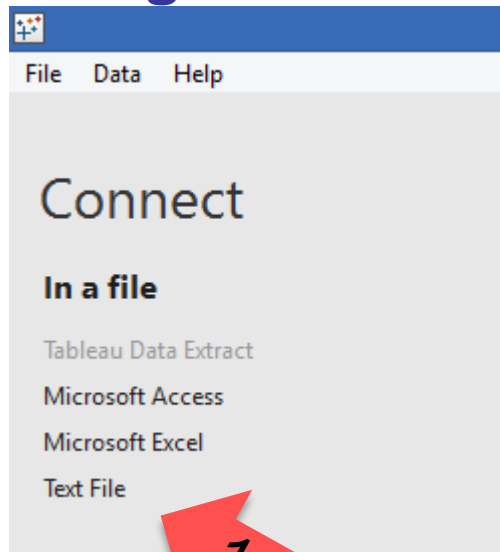
Find a template

Get Inspired



Tableau

Carregando a base “protestos_tableau”



Tableau

Pré-visualização

protestos_tableau.csv



Go to Worksheet

Copy

☐ Show hidden rows and columns

User Abc	Time Abc	Tweet Abc	Latitude Abc	Longitude Abc
pablohein	Sun Mar 15...	RT @Jornal...	null	null
vanessavb...	Sun Mar 15...	RT @HPozz...	null	null
Pedro_Kib...	Sun Mar 15...	#ImaginaSe...	null	null
ricardo__rf	Sun Mar 15...	RT @Jornal...	null	null
ApLeolBDT	Sun Mar 15...	RT @Jornal...	null	null
ravidlimaa	Sun Mar 15...	RT @holiBR...	null	null
QuelRenn...	Sun Mar 15...	J comearam...	null	null
tainaevaristo	Sun Mar 15...	Acordar e v...	-32.56001076	-53.37975668
_GuilhenA...	Sun Mar 15...	RT @MetroJ...	null	null

Tableau

WorkSheet – Mapa_Protestos

The image shows the Tableau interface for a worksheet named 'Mapa_Protestos'. The 'Data' pane on the left lists dimensions: Latitude, Longitude, Time, Tweet, User, and Measure Names. The 'Measures' pane below it lists 'Number of Records' and 'Measure Values'. The 'Columns' and 'Rows' shelves are empty, and the 'Marks' shelf is set to 'Automatic'. The 'Filters' shelf is also empty. The 'Show Me' panel on the right displays various visualization options, including a world map icon. Red arrows indicate the workflow: one arrow points from 'Latitude' and 'Longitude' in the Dimensions list to the 'Show Me' panel, and another arrow points from 'Measure Names' in the Measures list to the 'Show Me' panel.

For **symbol maps** try
1 geo dimension
0 or more dimensions
0 to 2 **measures**

Temos que transformar as informações de latitude e longitude em “Measures” para utilizar os recursos de visualização no mapa

Tableau

WorkSheet – Mapa_Protestos

Arraste os campos de lat e long para “Measures”
Repare que eles estão como strings (Abc)



Measures	
Abc	Latitude (Count (Distinct))
Abc	Longitude (Count (Distinct))
=#	Number of Records
#	Measure Values

Converta-os para tipo decimal



Measures

- Abc Latitude (Count (Distinct))
- Abc Longitude (Count (Distinct))
- =# Number of Records
- # Measure Values

- Convert to Discrete
- Convert to Dimension
- Change Data Type
 - Number (Decimal)
 - Number (Whole)
 - Date & time
 - Date
 - ☒ String
 - Default
- Geographic Role
- Default Properties
- Replace References...
- Describe...

E depois para Latitude e Longitude!



Measures	
🌐	Latitude
🌐	Longitude

Measures

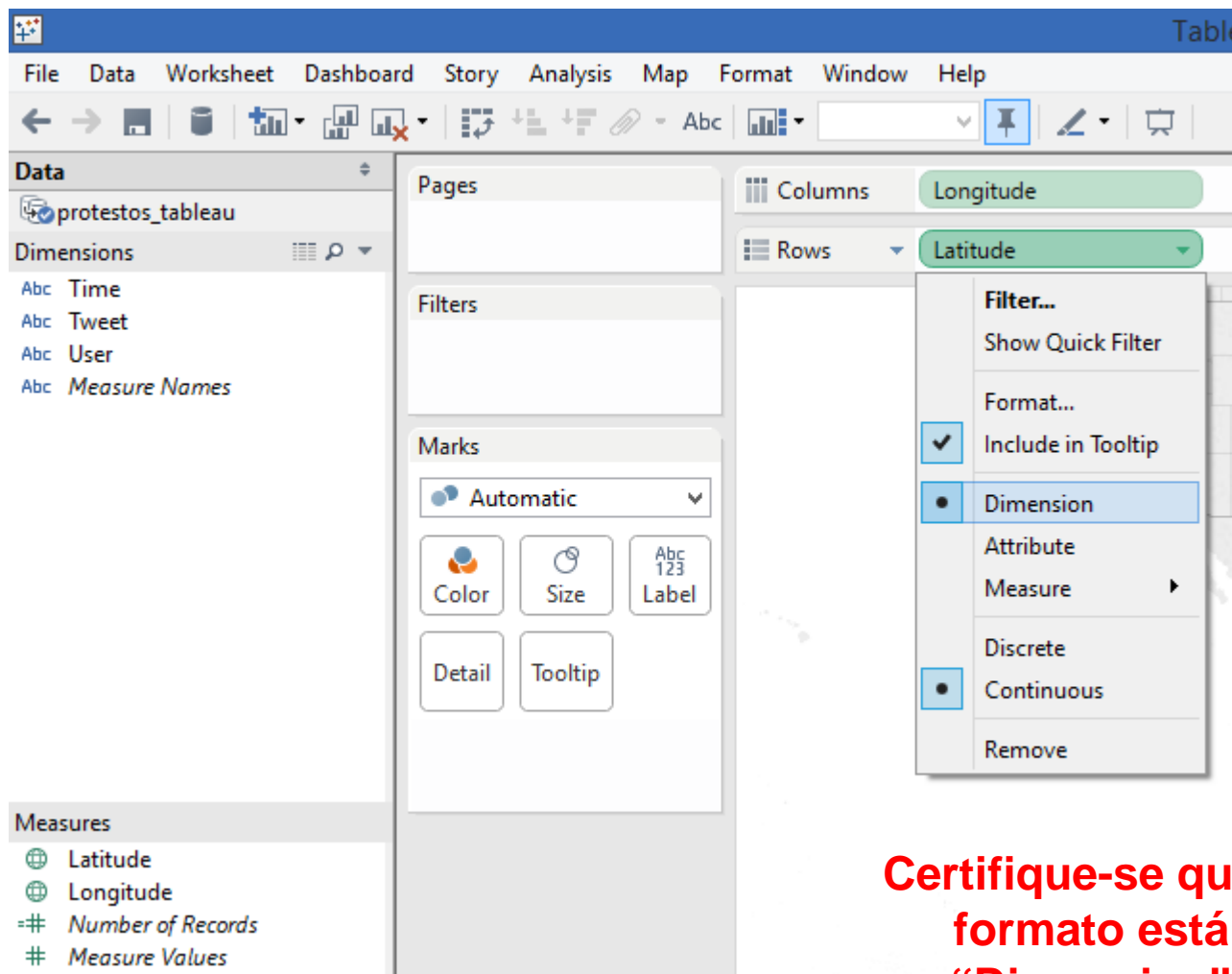
- # Latitude
- # Longitude
- =# Number of Records
- # Measure Values

- Geographic Role
 - ☒ None
 - Código de área
 - CBSA/MSA
 - Cidade
 - Distrito eleitoral
 - País/Região
 - Condado
 - Latitude
 - Longitude
- Default Properties
- Replace References...
- Describe...

Tableau

WorkSheet – Mapa_Protestos

Arraste “Longitude” para Columns e o campo “Latitude” para Rows

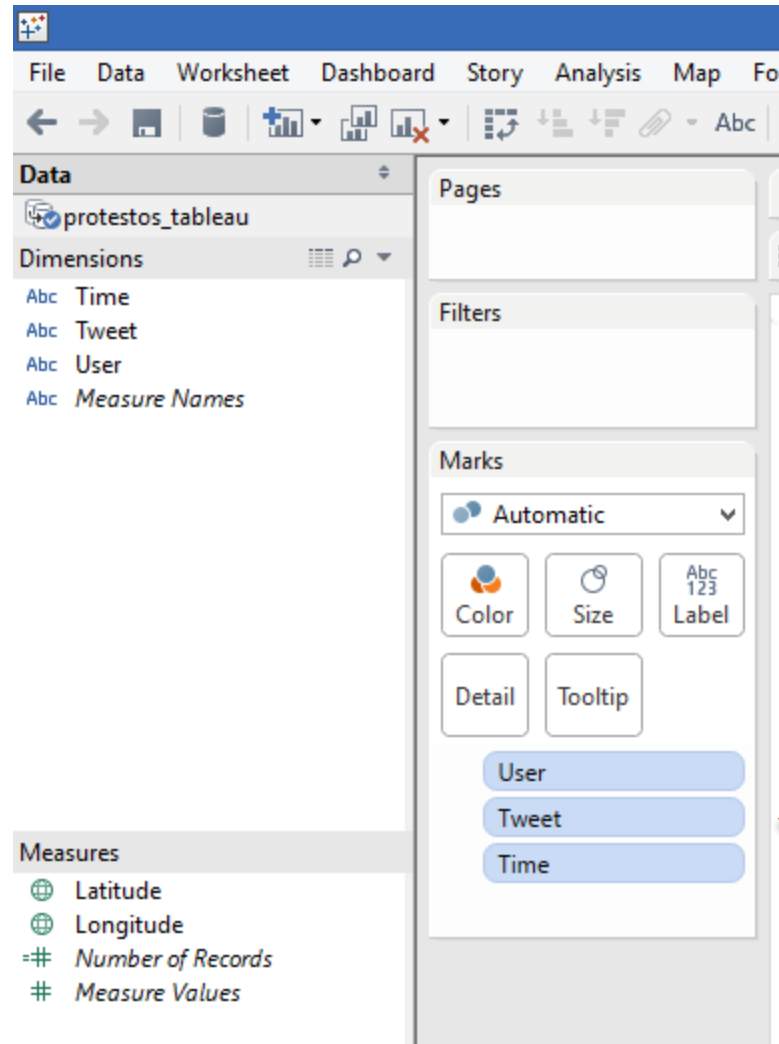


**Certifique-se que o
formato está
“Dimension”**

Tableau

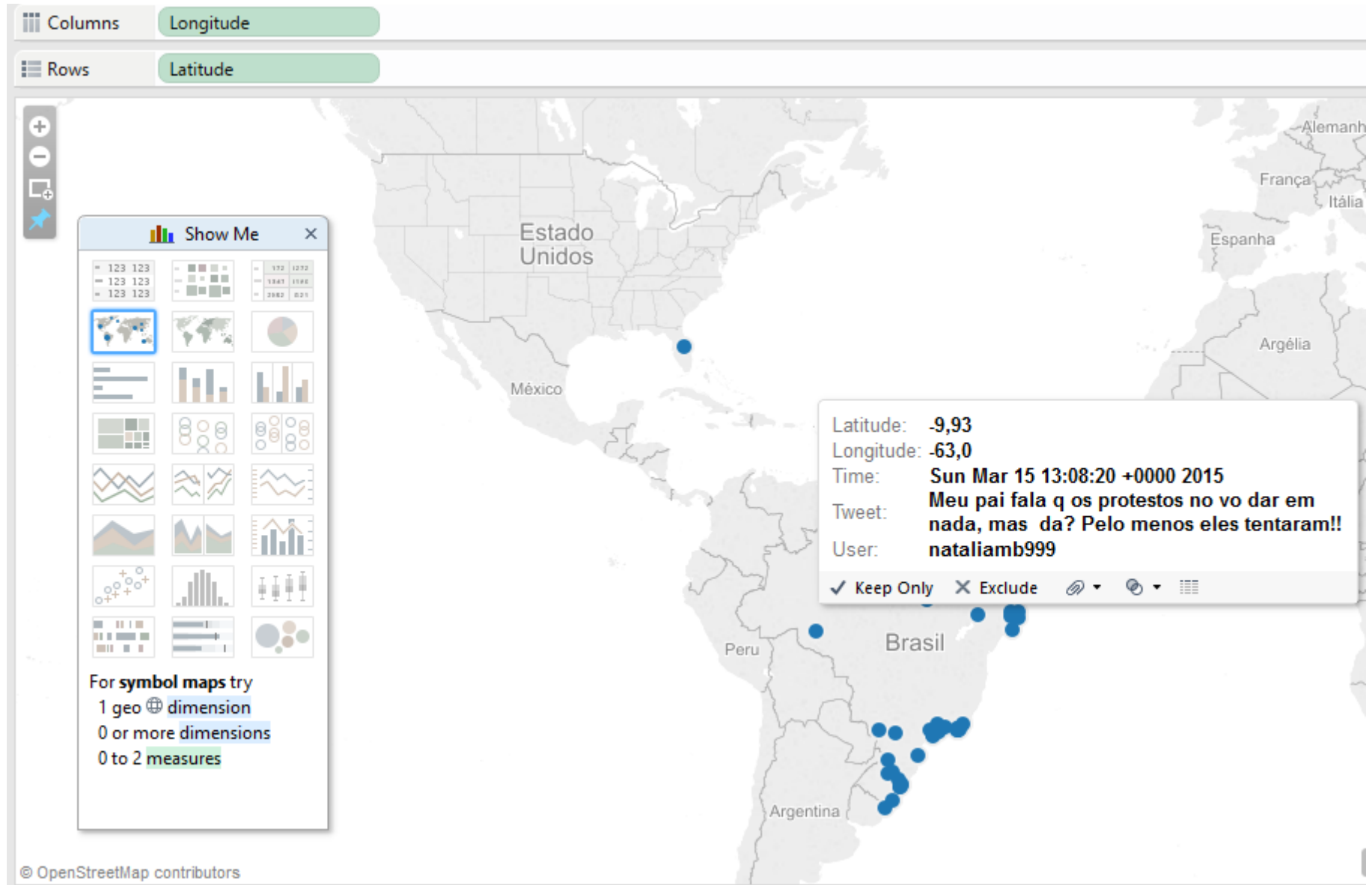
WorkSheet – Mapa_Protestos

Selecione os demais atributos – User, Tweet e Time – para o card “Marks”



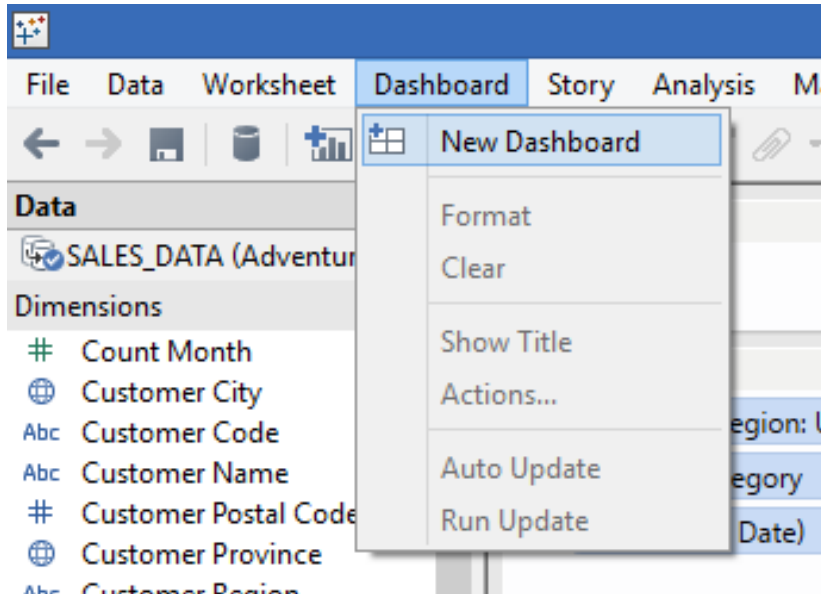
**Estas
informações
serão exibidas
no mapa!**

WorkSheet – Mapa_Protestos

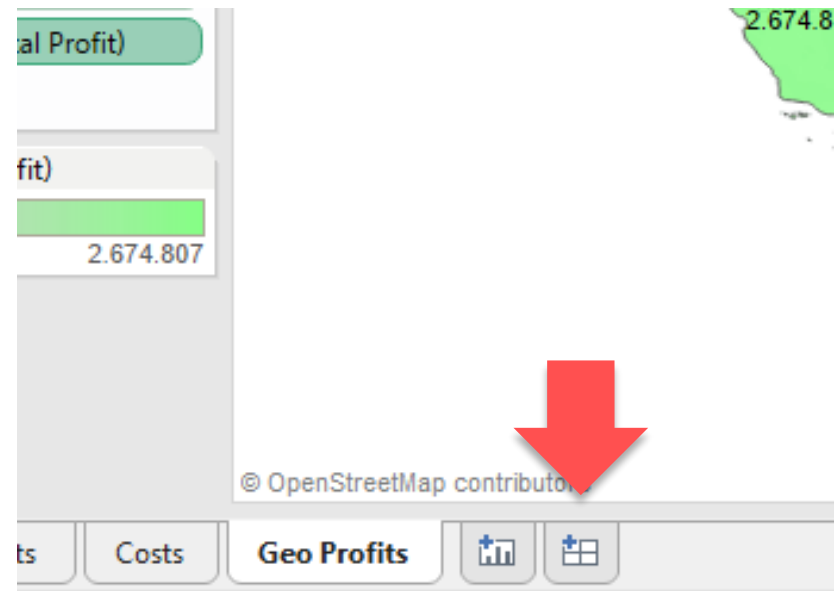


Tableau

Criando o dashboard

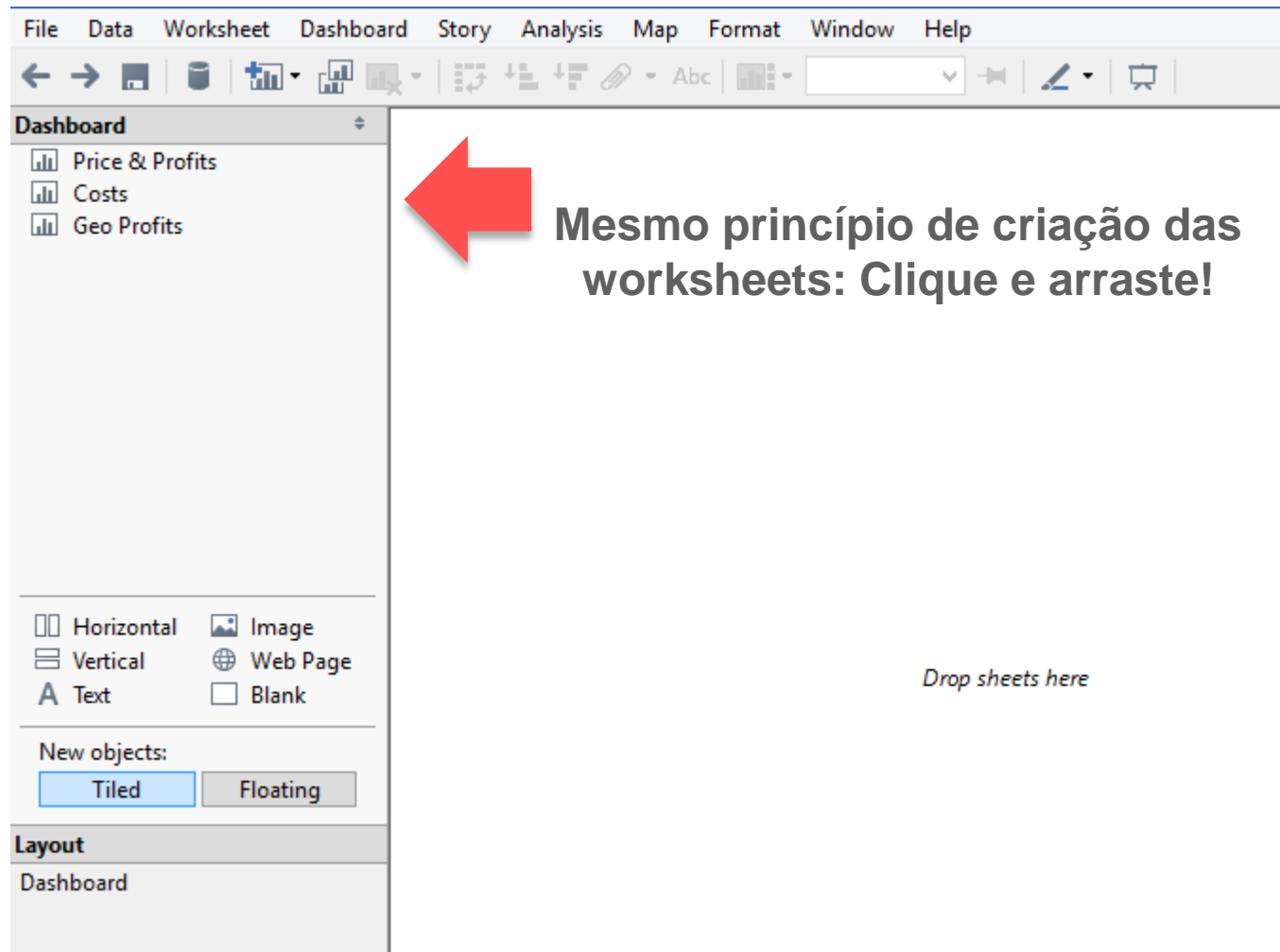


Ou



Tableau

Criando o dashboard



Códigos

Todos os códigos e bases estão disponíveis em

<https://github.com/leandrowar/Business-Intelligence>

Dashboard

<https://public.tableau.com/profile/leandro.guerra#!/>

Business Intelligence