Using Java 8 to Process Government Open Data

About William Antonio Siqueira

- ➔ Graduated at Sao Paulo Technology College (FATEC SJC)
- ➔ JUG Vale Coordinator and Founder
- ➔ Java Programmer for 6 years
- ➔ Transparencia Hacker SJC
- ➔ Works at Red Hat supporting JBoss products
- ➔ Post Graduating at ABC Federal College (UFABC)
- ➔ Blogger, JavaFX & Open Source lover
- ➔ Speaker at Java One, FISL, TDC;

About this presentation

- ★ About Open Data movement
- ★ Programmers and Open data
- ★ How Java 8 can help?
- ★ Some code and small sample applications

The Big Idea

*Work with **government** and non profit organizations to **open their data**, so everyone will be able to handle this data in order to **create applications**, **mix** with other data (mashup) or just create ways to better **visualize the information** behind the data.*

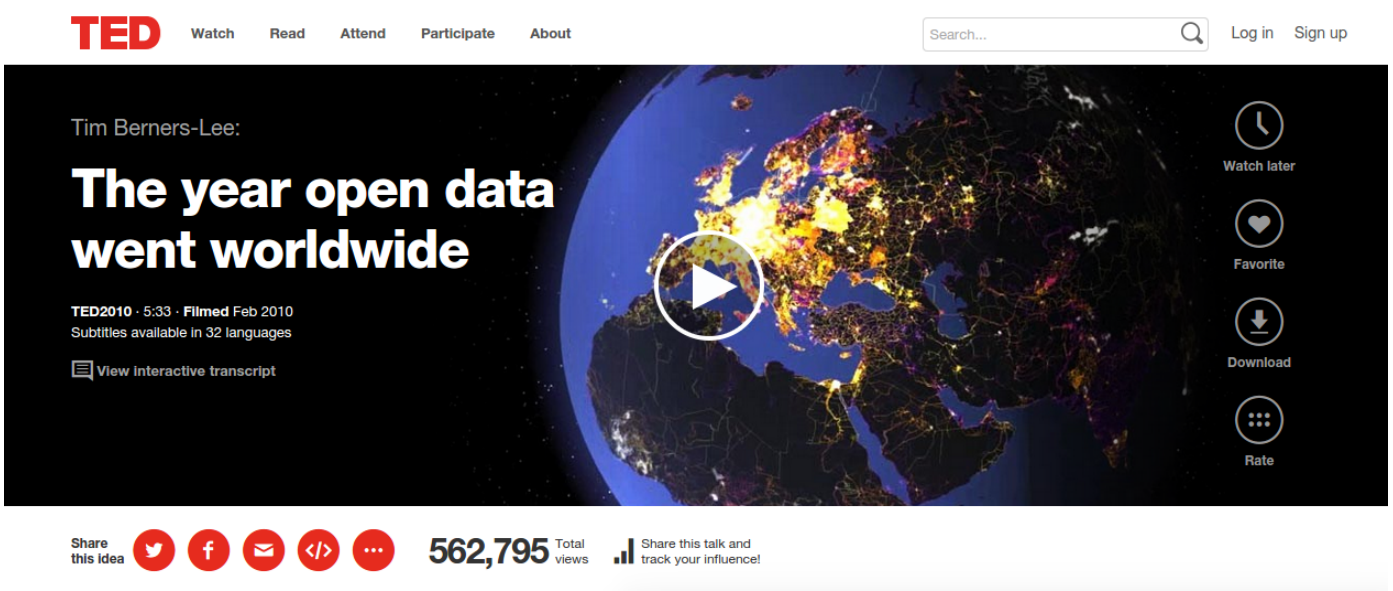
With open data we can improve democracy, create a more efficient government, and smart cities.

What is open data

*“Open data can be freely **used**, **modified**, and **shared** by **anyone** for any purpose”*

http://opendefinition.org/od/

The Open Data movement



Gained strength with Tim Berners-Lee talk at TED:
http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide?language=en

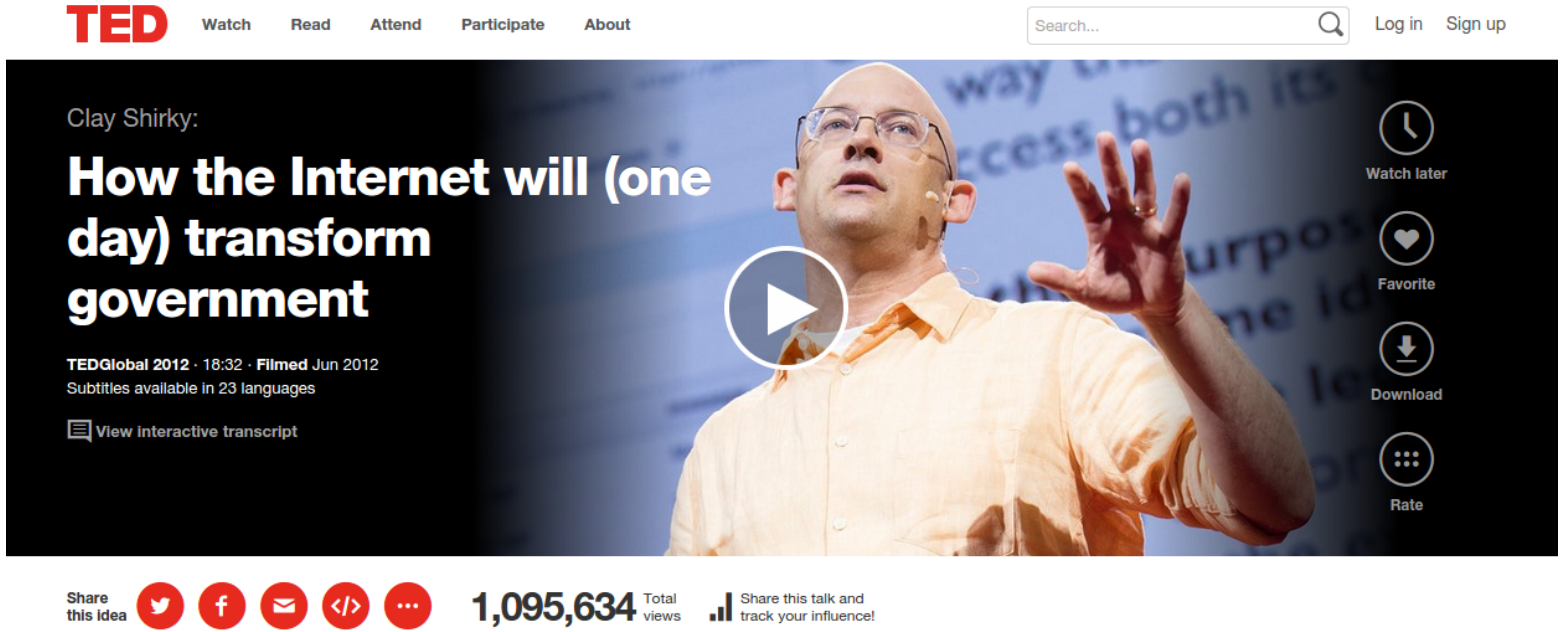
Resistance

Sometimes it’s hard to get data from institutions, mainly from governments. Reasons:

- They don’t understand the benefits of opening their data;
- Political reasons;
- Bureaucracy;

But no one can stop it

Clay Shirky explain why internet will help to create the new government
http://www.ted.com/talks/clay_shirky_how_the_internet_will_one_day_transform_government



In spite of you, tomorrow will be another day. I ask you, where will you hide from the great euphoria? How will you prohibit when the rooster insists on crowing? New water flowing, and our people loving one another, without stopping - rough translation from Chico Buarque's music "Apesar de Voce"

Governments and open data

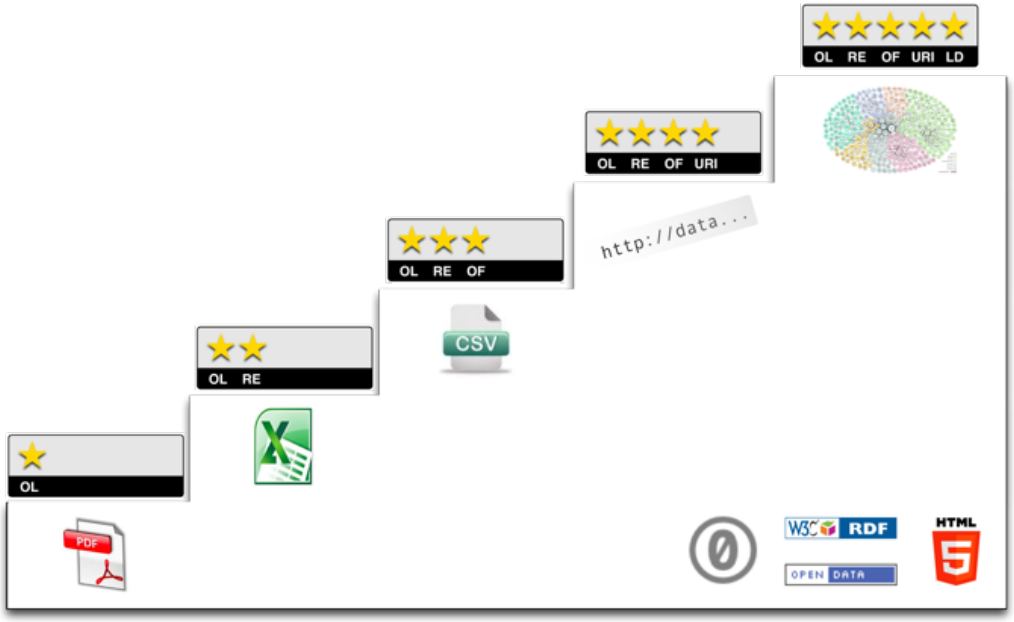
Countries are opening their data (open data portals)

- data.sfgov.org *
- data.gov
- dados.gov.br
- data.gov.uk
- data.gov.au
- data.gov.in
- ....

- *SF was ellected as the best city for open data in US! http://www.governing.com/news/headlines/san-francisco-is-the-best-city-for-open-aata.html*

5 starts

However, a lot still must be done...



http://5stardata.info/

How can programmers help?

Depending on the format of available data, we might need to:

Extract: Retrieve data from web pages, pdfs, etc;

Transform: Make the data available in computer readable formats. Aggregate and reduce operations;

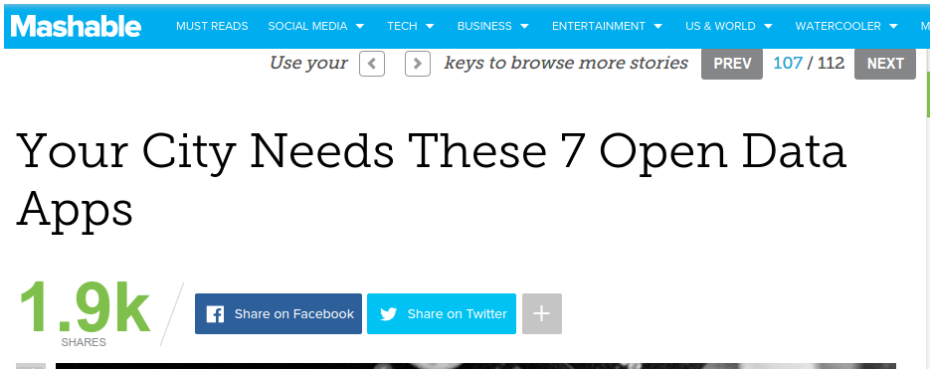
Combine: Take data from various sources and combine them (like security + education information);

Display: Create some way to expose the information behind the data.

Open Data apps

The number of open data apps is growing tremendously! A few years ago we would have only a few samples of open data apps, but today we have infinite apps! Famous open data apps categories:

- Public transportation
- Public services
- Citizens commenting government decisions
- Data about violent and crime ocurrencies



http://mashable.com/2012/11/07/open-data-city-apps/

“Transparencia Hacker” movement

A movement to help to make the open data movement happen, improve transparency and help citizens participation in government by creating applications that explore open data

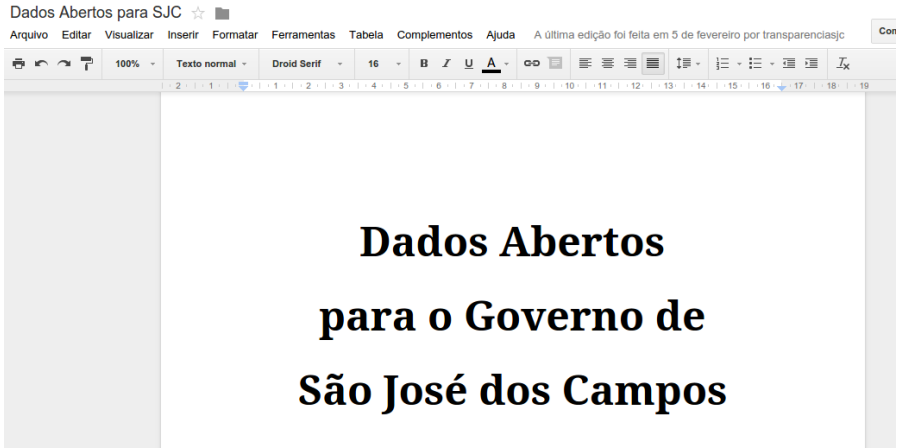
https://speakerdeck.com/alegomes/how-civil-society-is-driving-the-open-data-portal-development-in-brazil

“Transparencia SJC” movement

‘Hackers do bem’ criam grupo para cobrar transparência nas contas da Prefeitura de S. José

Hackers do bem. Foto: Pedro Ivo Prates

Câmara e autarquias também estão na mira do grupo; prefeituras dizem que já seguem as normas federais de publicação de gastos públicos



We have a group at our city to work with government and open the data. One of the members of the group (Paulo) personally delivered a document to the mayor. The document explained the open data movement and how they could get start to make it happen. *More: transparenciasjc.org*

Engaging programmers

Programmers get together to create apps that explores government open data(hackthon)...

But Java was not often used… Until Java 8!

What changes with Java 8?

Java 8 comes with tools that helps programmers to **efficiently** explore open data and create applications.

- ★ Lambda expressions
- ★ Stream
- ★ NIO Improvements
- ★ Nashorn*
- ★ JavaFX*

**** With these tools we can use other popular languages(CSS, Javascript) and tools(Scene Builder) that make it even easier to create applications***

Java 8 tools for Open Data

Extract: NIO with Stream support, Third party APIs;

Transform: Stream API, Collectors;

Combine: Stream API, Collection improvements;

Display: JavaFX, JavaFX + FXML + CSS + JS(Nashorn).

Using Stream

Stream

[filter] [peek] [map] [skip] [distinct] [flatMap] [limit] [sorted] [concat]

{ toArray | reduce | noneMatch | min | max | forEach | forEachOrdered | findAny | findFirst | count | **collect** | anyMatch | allMatch }

➔ Stateless/Stateful intermediate operations
➔ Terminal operations(close the stream)

Operations accept functional interfaces(Predicate, Function, Supplier, Consumer, BiFunction from java.util.function package), which means we can use Lambda expressions.

After you do stuff with the stream, you can collect the data using **Collectors.**

Where do I get the stream from?

- **Generate**
  - java.util.stream.Stream::of
  - java.util.stream.Stream::empty
  - java.util.stream.Stream::generate
- **From any collection**
  - java.util.Collection::stream
  - java.util.Collection::parallelStream
- **Lines from a file**
  - java.nio.file.Files::lines

Using Nashorn

Javascript in Java 8 is easier than ever:

- *jjs* tools comes with Java 8 to directly evaluate JS;
- No need to recompile; Scripting makes easier to quickly create applications
- We can evaluate Javascript from Java. We can create hybrid applications that uses JavaScript and Java;
- No need to parse JSON!
- **jjs** with the flag **-fx** allows you to create a JavaFX application only using JS!

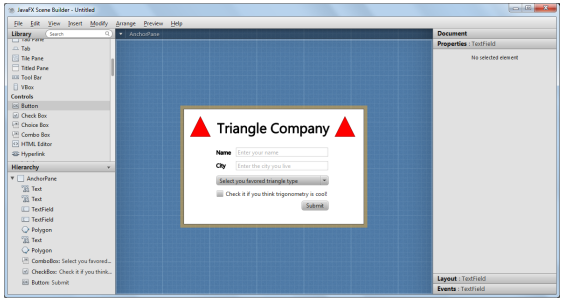
./jjs -fx Hello.js

```
var txtOla = new javafx.scene.control.Label(">> Olá Mundo! <<")
txtOla.font = new javafx.scene.text.Font(50)
stage.scene = new javafx.scene.Scene(txtOla)
stage.width = 500
stage.title = "Olá JavaFX + Javascript"
stage.show()
```

JavaFX

RESOURCES

TOOLS

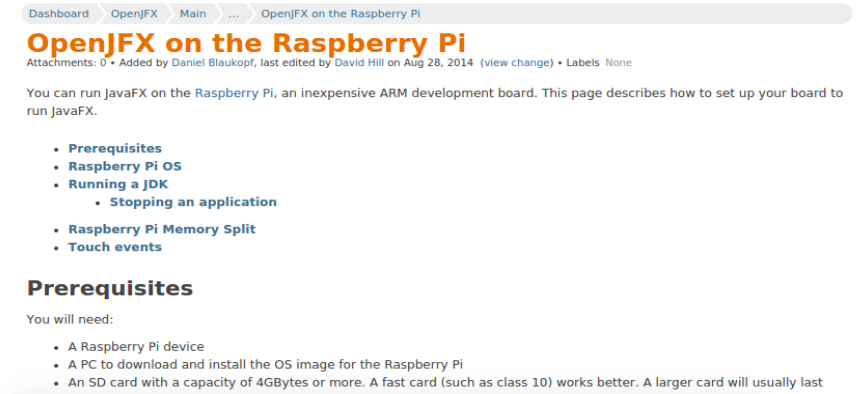


COMMUNITY

- Easy to create charts and visualizations; great resources to present data;
  - See "*JavaFX: A Brilliant Platform for Presenting Data*" youtu.be/XfkLfIGV59s
  - JavaFX Ensemble
- We don't need Java language to build JavaFX apps;
  - We can use **FXML** (with **Scene Builder**), **CSS** and Javascript (with **Nashorn**)
- Modern APIs and active community
  - See **FXtras** and **ControlsFX** projects;
  - Recently released **jfxlab.com**

WORA is back with Java 8

Write Once, Run Anywhere is back with Java 8!



Dashboard › OpenJFX › Main › ... › OpenJFX on the Raspberry Pi

OpenJFX on the Raspberry Pi

Attachments: 0 • Added by Daniel Blaukopf, last edited by David Hill on Aug 28, 2014 (view change) • Labels None

You can run JavaFX on the Raspberry Pi, an inexpensive ARM development board. This page describes how to set up your board to run JavaFX.

- Prerequisites
- Raspberry Pi OS
- Running a JDK
  - Stopping an application
- Raspberry Pi Memory Split
- Touch events

Prerequisites

You will need:

- A Raspberry Pi device
- A PC to download and install the OS image for the Raspberry Pi
- An SD card with a capacity of 4GBytes or more. A fast card (such as class 10) works better. A larger card will usually last

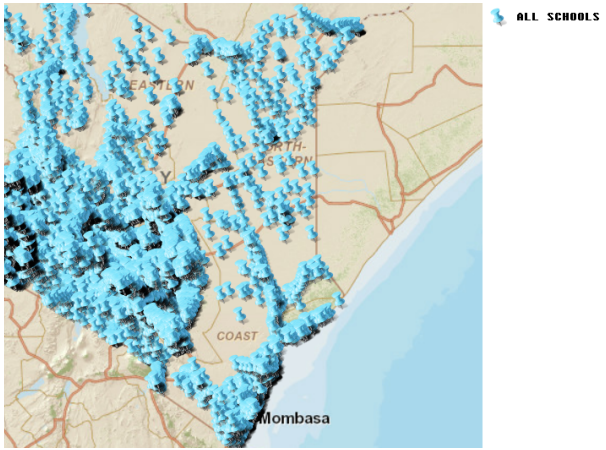
➔ **JavaFXPorts.org**

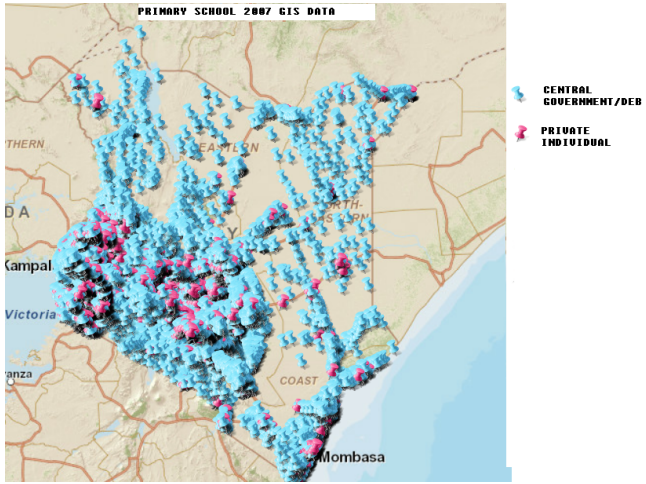
➔ **Java 8 on Raspberry Pi**

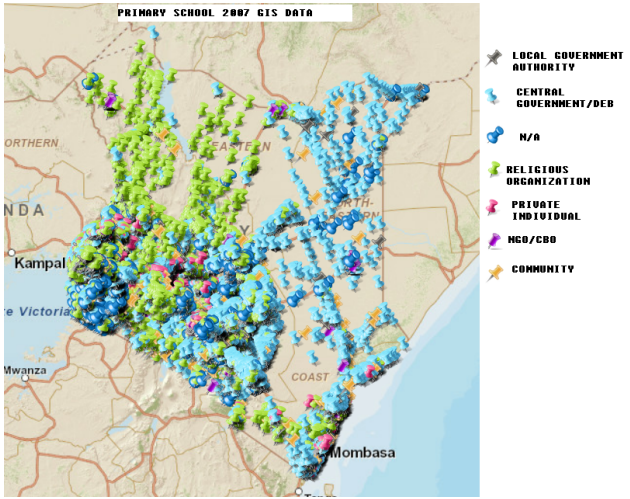
DEMO&CODE TIME!

Kenya GIS App

"I downloaded data from Kenya open data decided to play around with it, this is what i came up in the first 1hr"







http://kenyadevelopers.blogspot.com.br/2014/09/gis-applications-build-with-javajavafx.html

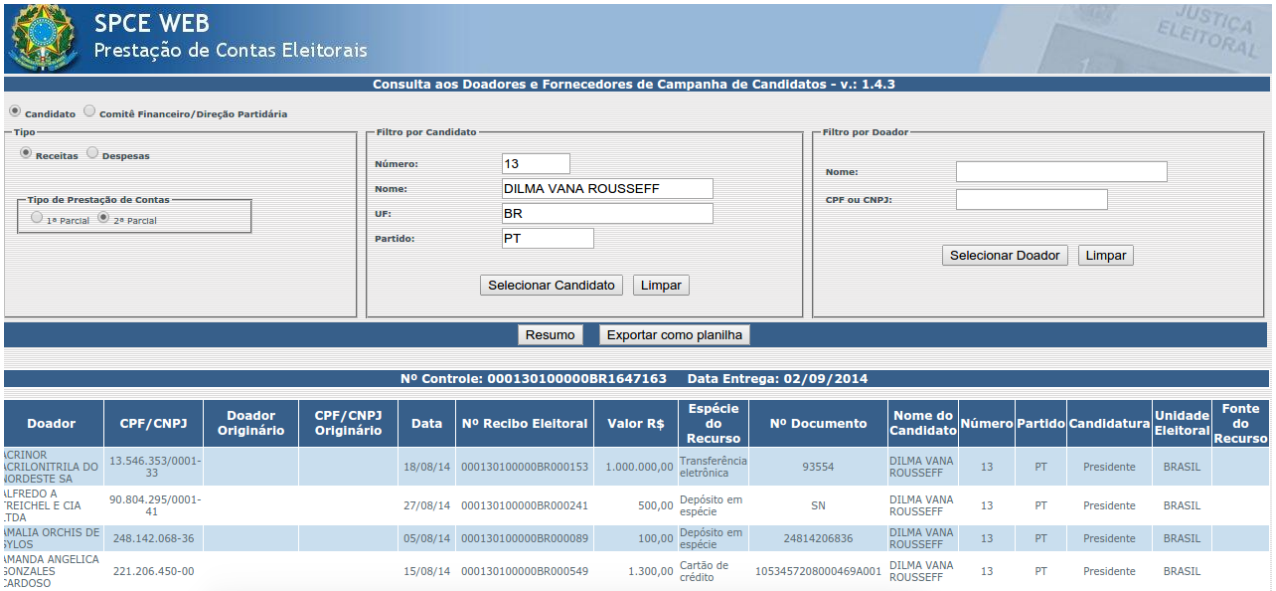
Money Received

One controversial topic about the elections is how much each candidate received from companies to make their political campaign. This sample tries to answer the following question:

“The seven biggest “donors” for each presidential candidate to make their political campaign and how much they donated?”

Money Received

First thing to do is to manually download data for each candidate from TSE site and place it in a given directory. Then we will load the CSVs on the application, aggregate the information and show a simple pie chart.



SPCE WEB
Prestação de Contas Eleitorais

Consulta aos Doadores e Fornecedores de Campanha de Candidatos - v.: 1.4.3

Candidato | Comitê Financeiro/Direção Partidária

Tipo: Receitas | Despesas
Tipo de Prestação de Contas: 1ª Parcial | 2ª Parcial

Filtro por Candidato
Número: 13
Nome: DILMA VANA ROUSSEFF
UF: BR
Partido: PT
Selecionar Candidato | Limpar

Filtro por Doador
Nome:
CPF ou CNPJ:
Selecionar Doador | Limpar

Resumo | Exportar como planilha

Nº Controle: 000130100000BR1647163 Data Entrega: 02/09/2014

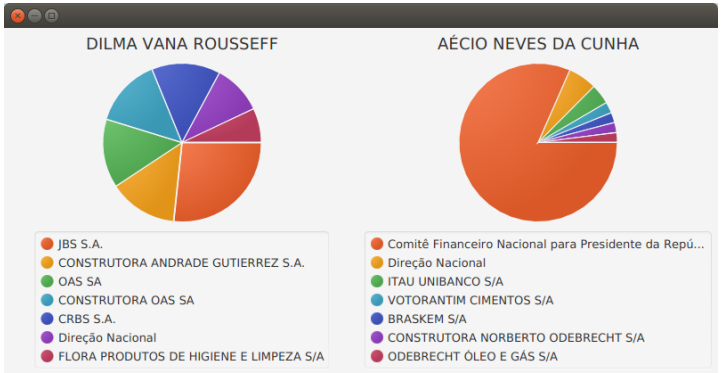
Doador	CPF/CNPJ	Doador Originário	CPF/CNPJ Originário	Data	Nº Recibo Eleitoral	Valor R$	Espécie do Recurso	Nº Documento	Nome do Candidato	Número	Partido	Candidatura	Unidade Eleitoral	Fonte do Recurso
ACRINOR ACRILONITRILA DO NORDESTE SA	13.546.353/0001-33			18/08/14	000130100000BR000153	1.000.000,00	Transferência eletrônica	93554	DILMA VANA ROUSSEFF	13	PT	Presidente	BRASIL	
ALFREDO A REICHEL E CIA LTDA	90.804.295/0001-41			27/08/14	000130100000BR000241	500,00	Depósito em espécie	SN	DILMA VANA ROUSSEFF	13	PT	Presidente	BRASIL	
AMALIA ORCHIS DE SYLOS	248.142.068-36			05/08/14	000130100000BR000089	100,00	Depósito em espécie	24814206836	DILMA VANA ROUSSEFF	13	PT	Presidente	BRASIL	
AMANDA ANGELICA GONZALES CARDOSO	221.206.450-00			15/08/14	000130100000BR000549	1.300,00	Cartão de crédito	1053457208000469A001	DILMA VANA ROUSSEFF	13	PT	Presidente	BRASIL	

Money Received

Scan files in a given directory and aggregate the values on a Map where's the candidate name is the key and the value is the list of donations (with donor name and the donated amount)

```
Map<String, Set<Donation>> donations = new HashMap<>();
Files.list(Paths.get(DATA_DIR))
        .peek(f -> System.out.println("Processing " + f.getFileName()))
        .forEach(p -> {
            try {
                Files.lines(p, StandardCharsets.ISO_8859_1).skip(1).forEach(l -> {
                   // For each Line, retrieve what I want
                });
            } catch (IOException e) {
                System.err.println("Error loading " + p);
            }
        });
```

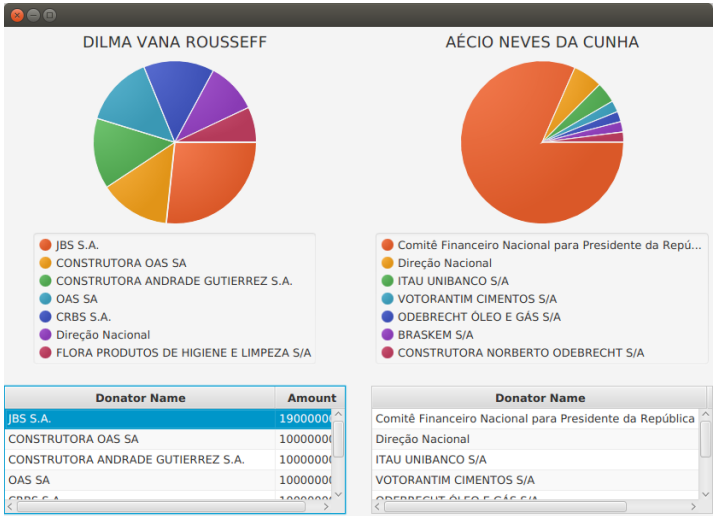
Money Received




```
donations.entrySet().forEach(e -> {
    PieChart chart = new PieChart();
    chart.setPrefSize(400, 300);
    chart.setTitle(e.getKey());
    e.getValue().stream()
            .sorted(Comparator.comparing(Donation::getAmount).reversed())
            .limit(7)
            .map(d -> new PieChart.Data(d.getDonorName(), d.getAmount()))
            .forEach(chart.getData()::add);
    charts.getChildren().add(chart);
});
```

Money Received

```
donations.entrySet().forEach(e -> {
    PieChart chart = new PieChart();
    chart.setPrefSize(300, 400);
    chart.setTitle(e.getKey());
    TableView<Donation> tblDonations = new TableView<>();
    TableColumn<Donation, String> clDonatorName = new TableColumn<>("Donor Name");
    TableColumn<Donation, String> clAmount = new TableColumn<>("Amount");
    clDonatorName.setCellValueFactory(new PropertyValueFactory("donorName"));
    clAmount.setCellValueFactory(new PropertyValueFactory("amount"));
    tblDonations.getColumns().addAll(clDonatorName, clAmount);
    tblDonations.setPrefSize(CHART_WIDTH, 150);
    e.getValue().stream()
            .sorted(Comparator.comparing(Donation::getAmount).reversed())
            .limit(7)
            .peek(tblDonations.getItems()::add)
            .map(d -> new PieChart.Data(d.getDonorName(), d.getAmount()))
            .forEach(chart.getData()::add);
    charts.getChildren().add(new VBox(20, chart, tblDonations));
});
```



Mobile Towers

Goal: Build an app to put together data about mobile coverage in Brazil and try to find why we have such bad services on this area!

App not finished since we are still scrapping data for it.

To engage more programmers, we are moving to use no Java code and create a controller using Javascript.

STOP: Only JS, CSS and FXML Sample

As the mobile towers app is not finished, I thought I would share the world cup app. It was done with no Java code. My girlfriend helped me on get this done in only a few hours!

Advantages of using Nashorn: Scripting (no need to recompile); no JSON parsing;

More information: http://fxapps.blogspot.com/2014/06/another-world-cup-app-using-javafx-fxml.html

http://fxapps.blogspot.com/2014/04/writing-javafx-apps-using-javascript.html

Survey result

A survey was incorrectly released with wrong charts and caused a lot of polemic in Brazil!

According to the wrong results, about 60% of the participants responded that women deserve be attacked according to their clothes! Fortunately it was wrong, only 26% think like that(which still too bad)... Fortunately the data is open and let's see how can we easily handle this using Java 8 features.

Survey result

The survey results are on a spreadsheet where each line is a response and each column represents a question.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
identificad	região onde foi realizada a	idade: qual a sua idade	área	sexo	cor ou rac	última sér	até que se	até que se	renda tota	renda tota	número de	no último	my world	my world	my world	my w
4128	NORTE	34	urbano	feminino	parda	quinta sér	analfabeto	quarta sér	1000	1490	9	1	prioritário	prioritário	secundári	priori
4130	NORTE	53	urbano	masculino	branca	médio cor	oitava sér	oitava sér	4800	4900	3	1	prioritário	secundári	secundári	priori
4134	NORTE	31	urbano	masculino	preta	analfabeto	analfabeto	analfabeto	678	678	5	1	prioritário	secundári	secundári	priori
4135	NORDESTE	38	urbano	feminino	parda	sétima sér	cursou sup	segunda s	1000	1000	4	2	prioritário	secundári	prioritário	priori
4138	NORDESTE	21	urbano	feminino	parda	primeira s	analfabeto	analfabeto	678	678	2	2	secundári	prioritário	prioritário	priori
4139	NORTE	24	urbano	feminino	parda	primeira s	não conhe	quinta sér	678		5	2	prioritário	secundári	secundári	priori
4140	NORDESTE	65	urbano	feminino	parda	superior c	analfabeto	analfabeto	678	1200	7	2	prioritário	secundári	secundári	priori
4142	NORTE	47	urbano	feminino	parda	médio cor	quarta sér	analfabeto	700	700	4	2	prioritário	secundári	secundári	priori
4147	NORDESTE	40	urbano	masculino	parda	fundamen	segunda s	sétima sér	678	3000	6	2	prioritário	secundári	secundári	priori
4148	NORTE	29	urbano	masculino	parda	médio cor	quinta sér	médio cor	678	800	3	1	prioritário	secundári	secundári	priori
4149	NORDESTE	49	urbano	feminino	amarela	médio cor	primeira s	analfabeto	2000	2000	2	2	prioritário	prioritário	secundári	priori
4150	NORDESTE	27	urbano	feminino	branca	segunda s	não sabe	não sabe	300	3000	3	2	secundári	secundári	secundári	secu
4155	NORDESTE	18	urbano	masculino	branca	terceira se	superior c	mestrado		6000	4	2	secundári	secundári	secundári	priori

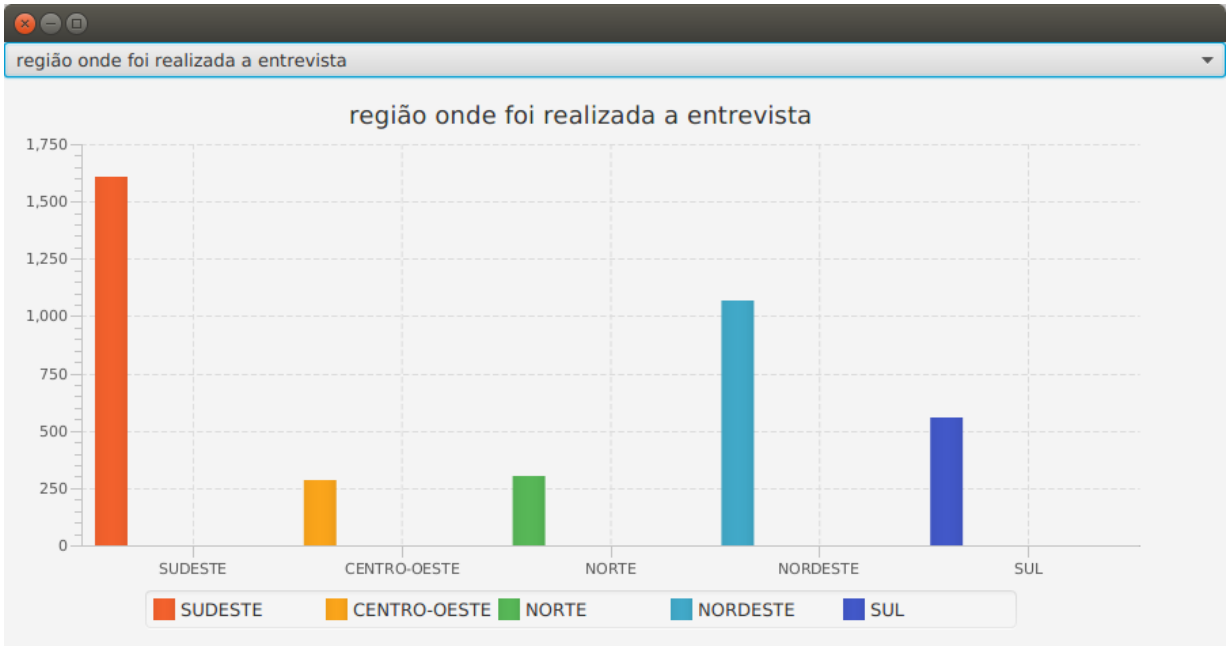
We saved this spreasheet as CSV (columns separated by semicolon) and we will open if from Java to transform it into a data structure where a Map will contain the question, and for each question we will have the possible answers and how much participants choose that answer, so: **Map<String, Map<String, Integer>>;**

Survey result

```
Map<String, Map<String, Integer>> results = new HashMap<>();
Path filePath = Paths.get(FILE_SURVEY);
String[] questions = Files.lines(filePath).findFirst().get().split(SEPARATOR);
Stream.of(questions).forEach(c -> results.put(c, new HashMap<>()));
Files.lines(filePath).skip(1).forEach(l -> {
    String[] anwers = l.split(SEPARATOR);
    for (int i = 1; i < anwers.length; i++) {
        results.get(questions[i]).compute(anwers[i], (o, n) -> {
            return n == null ? 1 : ++n;
        });
    }
});
```

Survey result

Now we have the information translated to Java, we can use JavaFX charts library to visualize it



Survey result

```
Map<String, Map<String, Integer>> results = SurveyData.load();
ComboBox<String> cmbQuestions = new ComboBox<>(FXCollections.observableArrayList(results.
keySet()));
cmbQuestions.getSelectionModel().selectedItemProperty().addListener((t, o, n) -> {
    final BarChart<String, Number> chart = new BarChart<>(new CategoryAxis(), new NumberAxis());
    chart.setTitle(n);
    results.get(n).entrySet().forEach(r -> {
        XYChart.Series resultSeries = new XYChart.Series();
        resultSeries.getData().add(new XYChart.Data(r.getKey(), r.getValue()));
        resultSeries.setName(r.getKey());
        chart.getData().add(resultSeries);
    });
    chart.setPrefWidth(850);
    chartPane.getChildren().setAll(chart);
});
```

Sample: Elections app

Elections are going to happen in Brazil at October 4. A simple application to explore candidates information was created in about 3 hours using JavaFX, FXML and a Java API to access candidates data.

Federal Government Tranfers

CSV that shows money that federal government transfers to city and states to be applied in education, health, security and other areas.

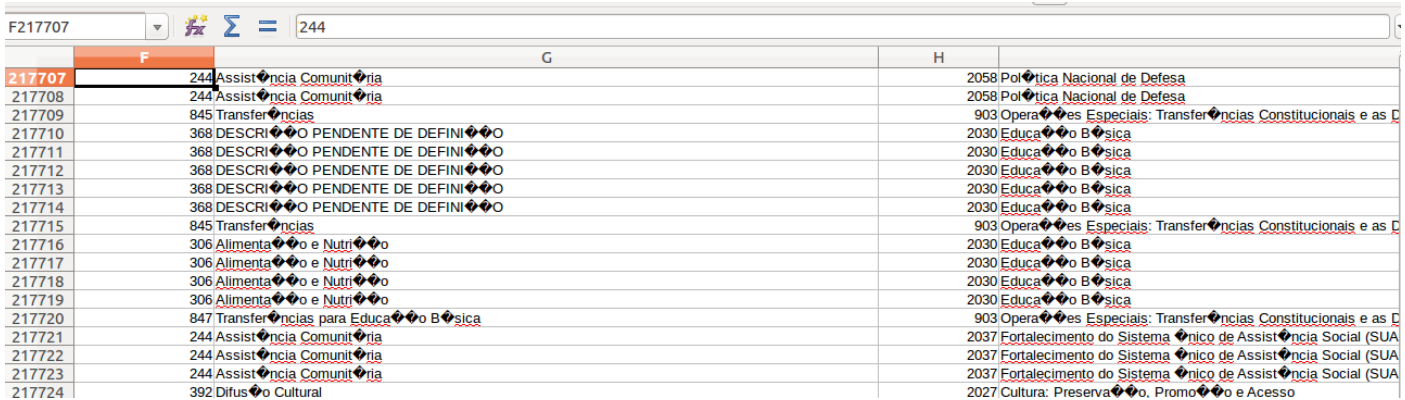
The app should help us to respond the following question:

“What is the HDI at a given city and how is federal government helping by tranferring resources to a given area?”

So we need to get indicator of each area and see the amount of money was transferred to that area from federal government.

Federal Government Tranfers

Transfer Data: CSV with about 220k lines!



F217707 | 244

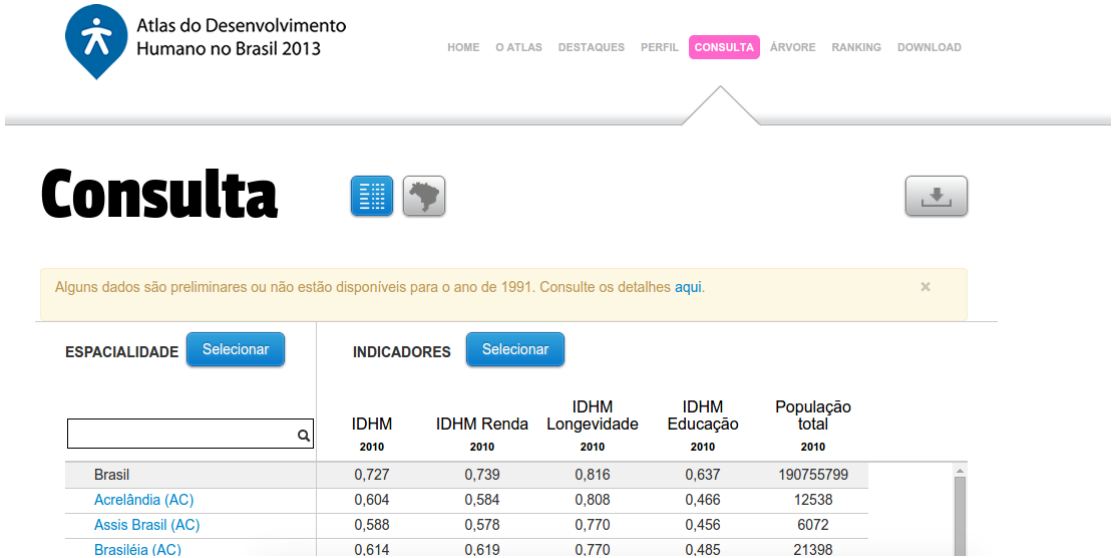
	F	G	H	
217707	244	Assist�ncia Comunit�ria	2058	Pol�tica Nacional de Defesa
217708	244	Assist�ncia Comunit�ria	2058	Pol�tica Nacional de Defesa
217709	845	Transfer�ncias	903	Opera��es Especiais: Transfer�ncias Constitucionais e as D
217710	368	DESCRI��O PENDENTE DE DEFINI��O	2030	Educa��o B�sica
217711	368	DESCRI��O PENDENTE DE DEFINI��O	2030	Educa��o B�sica
217712	368	DESCRI��O PENDENTE DE DEFINI��O	2030	Educa��o B�sica
217713	368	DESCRI��O PENDENTE DE DEFINI��O	2030	Educa��o B�sica
217714	368	DESCRI��O PENDENTE DE DEFINI��O	2030	Educa��o B�sica
217715	845	Transfer�ncias	903	Opera��es Especiais: Transfer�ncias Constitucionais e as D
217716	306	Alimenta��o e Nutri��o	2030	Educa��o B�sica
217717	306	Alimenta��o e Nutri��o	2030	Educa��o B�sica
217718	306	Alimenta��o e Nutri��o	2030	Educa��o B�sica
217719	306	Alimenta��o e Nutri��o	2030	Educa��o B�sica
217720	847	Transfer�ncias para Educa��o B�sica	903	Opera��es Especiais: Transfer�ncias Constitucionais e as D
217721	244	Assist�ncia Comunit�ria	2037	Fortalecimento do Sistema �nico de Assist�ncia Social (SUA
217722	244	Assist�ncia Comunit�ria	2037	Fortalecimento do Sistema �nico de Assist�ncia Social (SUA
217723	244	Assist�ncia Comunit�ria	2037	Fortalecimento do Sistema �nico de Assist�ncia Social (SUA
217724	392	Difus�o Cultural	2027	Cultura: Preserva��o, Promo��o e Acesso

- ➔ For each month a CSV with data of **all** tranfers.
- ➔ We will use **6 files**;
- ➔ Each file is name with its year and month

Data from: transparencia.gov.br

Federal Government Tranfers

Cities Data



2010; Issues matching cities name..

Data from: http://www.atlasbrasil.org.br/2013/pt/consulta/

Federal Government Transfers

To match information between files, normalize Text is needed:

- For example: “SAO JOSE DOS CAMPOS” <-> “São José dos Campos (SP)”

Java Objects were created:

- **Transfer** represents a single transfer;
- **CityInfo** contains the information about a given City
- **TransfersSummary** aggreate and mix information from both files by having the areas, the amount transferred to that area in a given period and the related city information.

Notice that the data is open and it uses an open format(3 stars),but more than ever we need Java and its tool to explore the data.

Federal Government Tranfers

DataLoader: Read the CSV files to load the data into Java objects; Hides extraction details from data transformation;

CityTransferTransformer: Creates a single object(TransfersSummary) with the needed information by aggregating and transforming the data we extracted;

App: Just print the information of the cities on console. We could, of course, create some user interface with JavaFX.

STAHP. ***Let's go to the IDE...***

Conclusion

Conclusion

- ➔ **If you are a Java programmer, you can get start today creating applications to explore open data potential;**

- ➔ **Java 8 became a great alternative to create applications that can process government data. In this version, programmers are able to write much less code to process data and show the information from it;**

- ➔ **This is the future and you can make the difference. Make the future Java, make the future government;**

Questions?

@William_antonio

jugvale.com

fxapps.blogspot.com

github.com/jesuino/j1-open-data

Thanks!