05 - Digital stories

How to get started with the project

TWEB 2017

Olivier Liechti

Weekly menu

Getting started with the project
Data exploration, sketching

Today

- Discuss a method for the project

- Introduce a data exploration tool

- Introduce supporting webcasts

Friday

- Meet with every group

- Review the project idea and initial sketch

- You should have dataset(s), have done some exploration and have a draft plan

Problem

I am a bit lost... I don't know where to start and what I should build.

Forces

- In the first project, we have learned how to fetch data through a REST API. We have also learned how to present it in a web page.

- What is the difference with the new project?

- I know that we also have to visualise in this new project. I know that there are 1 million ways to do that. How do I get started?

- How much text do we have to write?

Solution

- **Step 1:** pick a topic that interests you (bees, beers, bits, beats)

- **Step 2**: identify related dataset**s** (e.g. REST APIs, static csv files, streaming APIs)

- **Step 3**: explore the data (with Excel, Tableau, Kibana, R, Google Charts)

- **Step 4**: sketch the story (write a plan with an introduction, sections and a conclusion). How will you use data visualisation to enrich your story?

- **Step 5**: implement the story (use a combination of D3.js, Leaflet.js and whatever library you want)

Problem

How can I find interesting facts in a given dataset?

Forces

- It is easy to find datasets on the web.

- It is harder to find "interesting" things to tell about it. How can we uncover information in this data?

- Sometimes, you start with a clear goal and a list of questions. Sometimes, you discover facts "by accident".
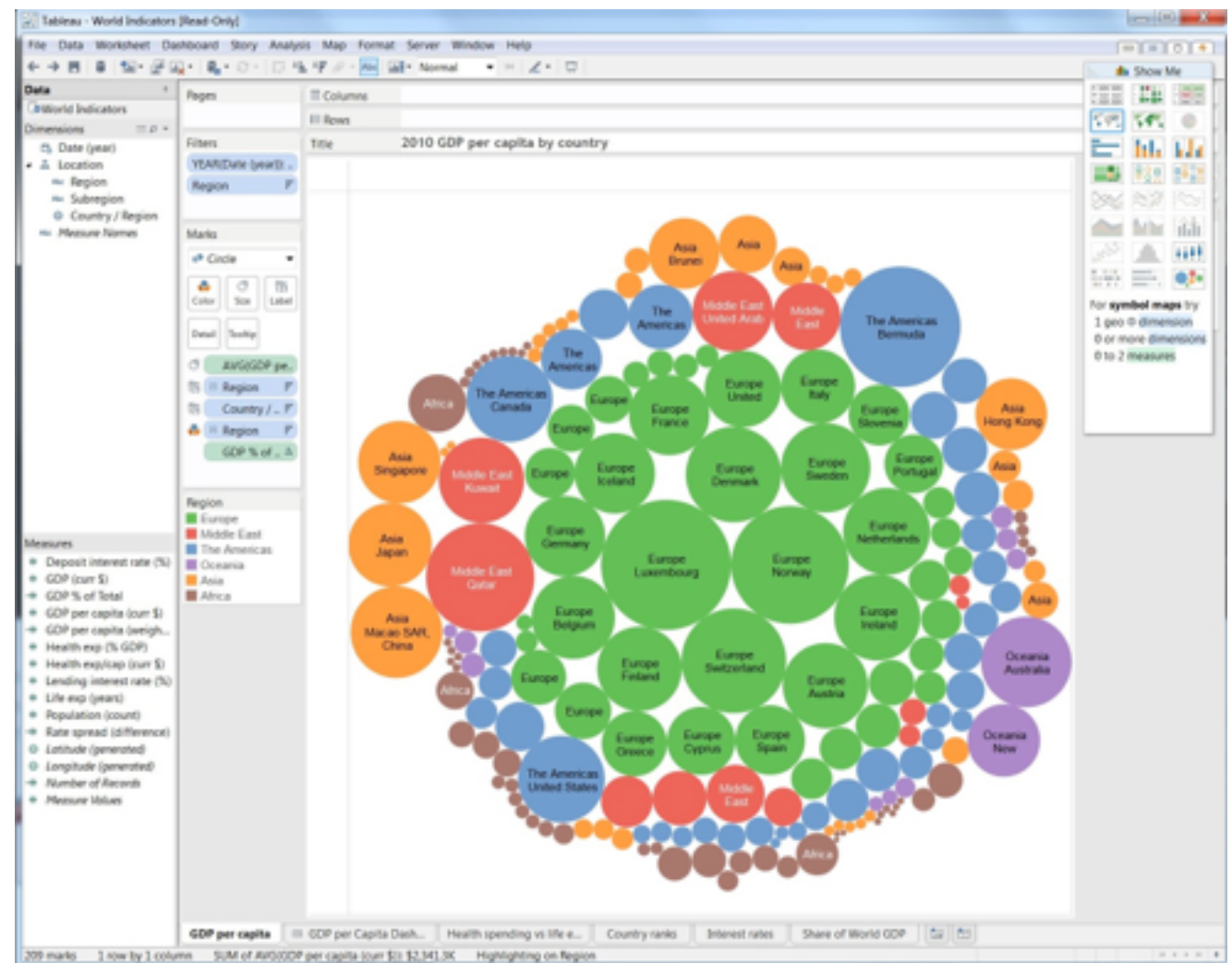
Solution

- Before coding anything, it is very useful to explore the data with tools.

- You all know Excel, and it is one way to do it. There are more powerful tools, such as Tableau, Kibana and Timelion.

- With these tools, you create mockups that you then implement in Javascript (e.g. with D3.js)

Tableau

https://public.tableau.com/s/

How to?

- **Get data**. Tabular format (CSV) is best, even if Tableau can work with nested data structures (JSON).

- Get familiar with the

- With these tools, you create mockups that you then implement in Javascript (e.g. with D3.js)

Dataset

- https://github.com/caesar0301/awesome-public-datasets

Machine Learning

- Context-aware data sets from five domains
- Delve Datasets for classification and regression (Univ. of Toronto)
- Discogs Monthly Data
- eBay Online Auctions (2012)
- IMDb Database
- Keel Repository for classification, regression and time series
- Labeled Faces in the Wild (LFW)
- Lending Club Loan Data
- Machine Learning Data Set Repository
- Free Music Archive
- Million Song Dataset
- More Song Datasets
- MovieLens Data Sets
- New Yorker caption contest ratings
- RDataMining - "R and Data Mining" ebook data
- Registered Meteorites on Earth
- Restaurants Health Score Data in San Francisco
- UCI Machine Learning Repository
- Yahoo! Ratings and Classification Data
- Youtube 8m

Dataset

- https://github.com/caesar0301/awesome-public-datasets

Machine Learning

- Context-aware data sets from five domains
- Delve Datasets for classification and regression (Univ. of Toronto)
- Discogs Monthly Data
- eBay Online Auctions (2012)
- IMDb Database
- Keel Repository for classification, regression and time series
- Labeled Faces in the Wild (LFW)
- Lending Club Loan Data
- Machine Learning Data Set Repository
- Free Music Archive
- Million Song Dataset
- More Song Datasets
- → MovieLens Data Sets

recommended for education and development

MovieLens Latest Datasets

These datasets will change over time, and are not appropriate for reporting research results. We will keep the download links stable for automated downloads. We will not archive or make available previously released versions.

Small: 100,000 ratings and 1,300 tag applications applied to 9,000 movies by 700 users. Last updated 10/2016.

- README.html
- ml-latest-small.zip (size: 1 MB)

Tableau

Option 1: Convert to a Date Field

- Select Analysis > Create Calculated Field.
- Type the following into the Formula field, and then click OK:
  DATE(DATEADD('second', INT([Unix Date]), #1970-01-01#))

Dates Display in Unix Epoch Time Instead of Date or Datetime ...
kb.tableau.com/articles/issue/dates-display-unix-epoch-instead-of-date-datetime

DATE(DATEADD('second', INT([Unix Date]), #1970-01-01#))

Problem

Once we have a draft of our story, how do we implement it?

Forces

- We have a good grip of the basic tools: Javascript, async programming, AJAX.

- However, this will not be enough to produce rich, interactive visualisations (Tableau-like, maps, live data).

- Every group will need some of these tools at different times... when will we learn about them?

Solution

- The training material is available right now, in the form of webcasts.

- You can use them in the order that suits your needs.

- At the end of the project, you are expected to be able to explain what D3.js, Leaflet and Socket.io are and when you would use them.

IoB

HAUTE ÉCOLE D'INGÉNIERIE ET DE GESTION DU CANTON DE VAUD
heig-vd
www.heig-vd.ch

23	**Socket.io (1): description des objectifs** by Olivier Liechti		5:59
24	**Socket.io (2): préparation des 2 projets back-end et front-end** by Olivier Liechti		6:12
25	**Socket.io (3.1): spécification du endpoint /observations** by Olivier Liechti		9:20
26	**Socket.io (3.2): implémentation du schéma Mongoose et du endpoint /observations** by Olivier Liechti		4:42
27	**Socket.io (4.1): test automatisé pour le server Socket.io** by Olivier Liechti		15:35
28	**Socket.io (4.2): le code Socket.io du côté serveur** by Olivier Liechti		6:49
29	**Socket.io (5.1): démo du client et présentation des ressources utilisées** by Olivier Liechti		7:09
30	**Socket.io (5.2): présentation du code client** by Olivier Liechti		15:37

Socket.io   AngularJS   BDD   Leaflet
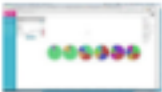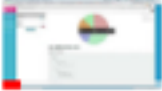
ElasticSearch

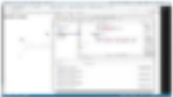| 31 | Intro à Elasticsearch & kibana: présentation des objectifs
by Olivier Liechti | 11:06 |
32	Intro à Elasticsearch & kibana (1): docker by Olivier Liechti	7:24
33	Intro à Elasticsearch & kibana (2): introduction à Elasticsearch by Olivier Liechti	15:54
34	Intro à Elasticsearch & kibana (3): git log et importation des données by Olivier Liechti	18:30
35	Intro à Elasticsearch & kibana (4): exploration des données avec Kibana by Olivier Liechti	18:33
36	Intro à Elasticsearch & kibana (5): implémentation d'une API entre Elasticsearch et nos clients web by Olivier Liechti	11:42

Elasticsearch   Kibana   Timelion

D3.js