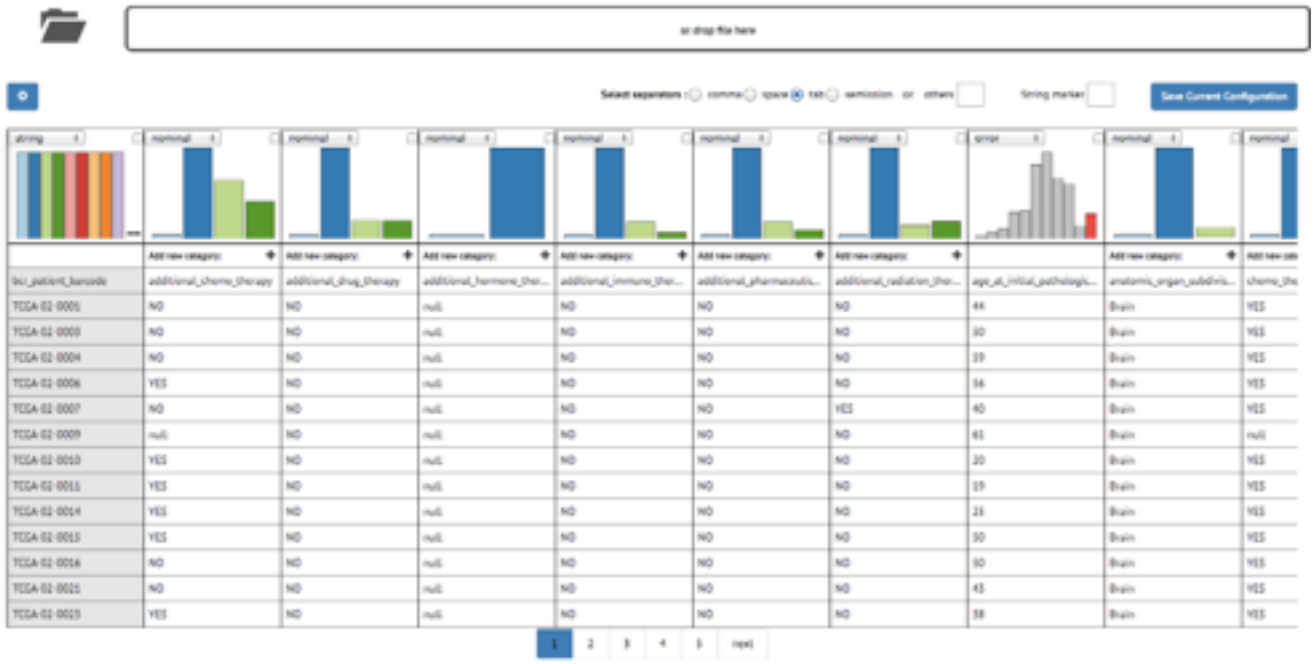


Project Proposal

CS-6630 Visualization

October 23, 2016

The project title: Caleydo Importer

Name: Asmaa M Aljuhani

Email: asmaa.aljuhani@utah.edu

UID: U6002550

Project Repository: https://github.com/aaljuhani/Caleydo_Importer

CALEYDO IMPORTER

Background and Motivation:

We live in the era of Data, in fact Big Data and even bigger analytics. Big Data is now crucial to so many areas as there is so much digital information than ever before and this is because of the digital revolution. In order to understand these data, we need to have a process that convert raw data into a well-structured and usable format to be analyzed and then even be an input to visualization tools.

Project Objectives:

This project will be composed of two parts: the first one is to build a tool (Caleydo Importer) that help users to unlock the potential of their data by providing an intuitive way to explore and prepare the data in a well-structure formate to be used as an input to other Caleydo framework plugins applications. Additionally, *Caleydo Importer* will provide a unified structure of the data that can be used within all Caleydo's applications. Within this project, I would like to work on applying user experience (UX) methods to make the front-end user friendly. I would like also to improve the data wrangler of Caleydo Importer.

The second part will use Caleydo Importer to explore the data of women in computing in Saudi Arabia then build visual elements to answer these topics:

- Which region of Saudi Arabia has more females in tech?

- Compare the number of female vs. male who is studying abroad.

- How number of women in tech changed for individual schools and across the country in the past 10 years.

CALEYDO IMPORTER

Data:

For Part 1(Caleydo Importer):

The input for Caleydo Importer is any text-based data files. For the first version, These data should be well-structured, that is, having a particular and unified data type for each column. This can be expanded later to take semi structure data where data consist of objects.

For Part2 (Saudi women in tech):

My data for this part comes from multiple resources

Saudi data : Ministry of education

http://www.moe.gov.sa/ar/Ministry/Deputy-Ministry-for-Planning-and-Information-affairs/HESC/Ehsaat/Pages/default.aspx

US Data: NSF

https://www.nsf.gov/statistics/2015/nsf15311/tables.cfm

Data Processing:

For part 1: Caleydo Importer should do the clean up for any given data and export the result into JSON object with other Caleydo framework configuration.

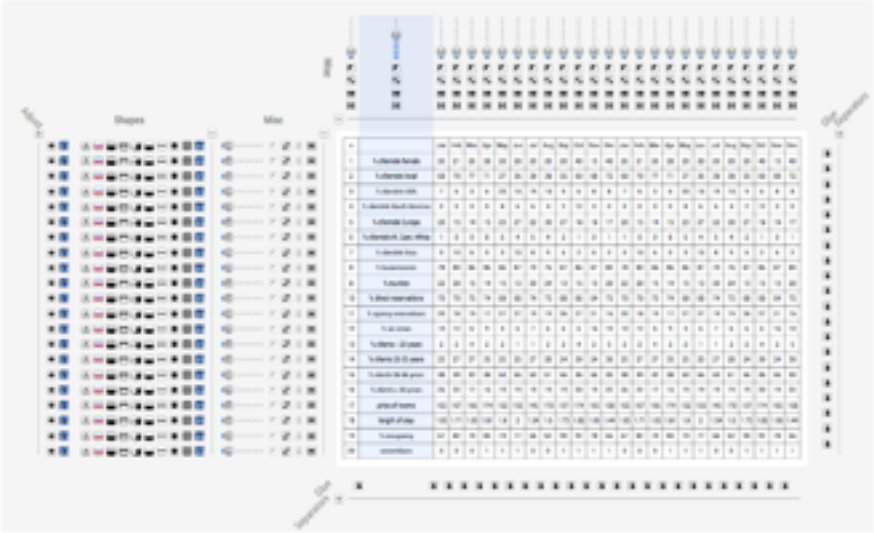
For part 2: Data for Saudi women in tech requires a substantial cleanup process.It is in Arabic and it is not well structured as the data is distributed into different tabs within an excel sheet. Each year has a separate excel sheet. This data will need to be translated to English then cleaned up. In fact this data includes all other majors such as medical, art, science, ..etc. I am still deciding wither to visualize all these data or just focus on tech related field.

Data from the NSF is neatly packed in well documented csv files.

Visualization Design:

I am inspired by many examples provided in the class as well as suggested by Prof.Alex and others that I encountered while looking for different design inspiration.

1) bertifier.com



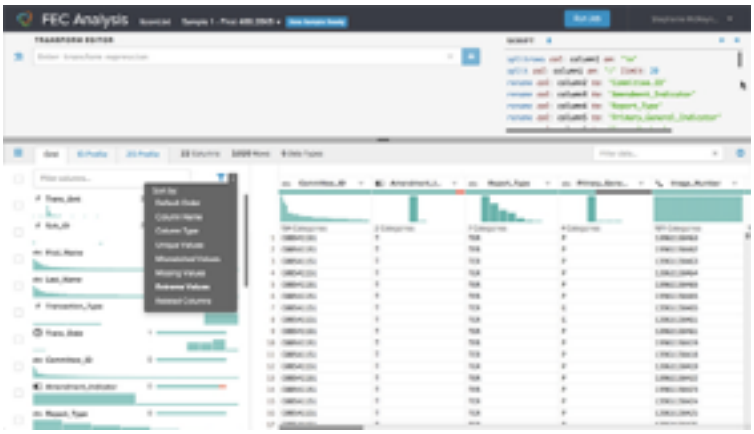
In this example, data is shown in a spreadsheet format. There is top and left table view that allow user to configure the data of the rows and columns. User can specify the headers of the data, show data in visual elements and rearrange columns and rows.

2)upSet

From this example, I liked the circles (bubbles) that is used to filter the data sets. I also want to adopt having a column to show general information about the data such as list of fields and their data type.

3)LineUp

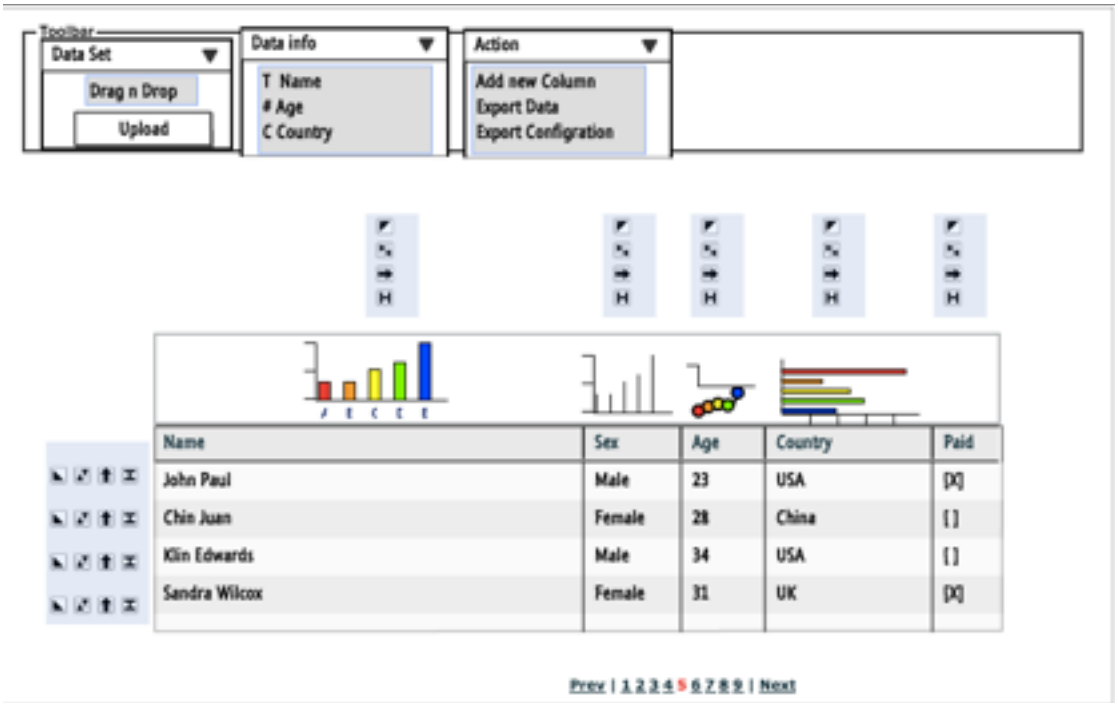
This example sows how data can be visualized as a spreadsheet format a long with a visual element for some of the column. It also show a features to explore the data as cdv file (although the formate is not accurate). There is a feature to add a column that can be used within caleydo importer to create a new field and provide a rule to extract information (e.g categories) from other field.



4)Trifacta

Trifacta is one of the main inspiring example that has provide users with general data information (on the left) and raw data (on the right).

My Final Design for Caleydo Importer:

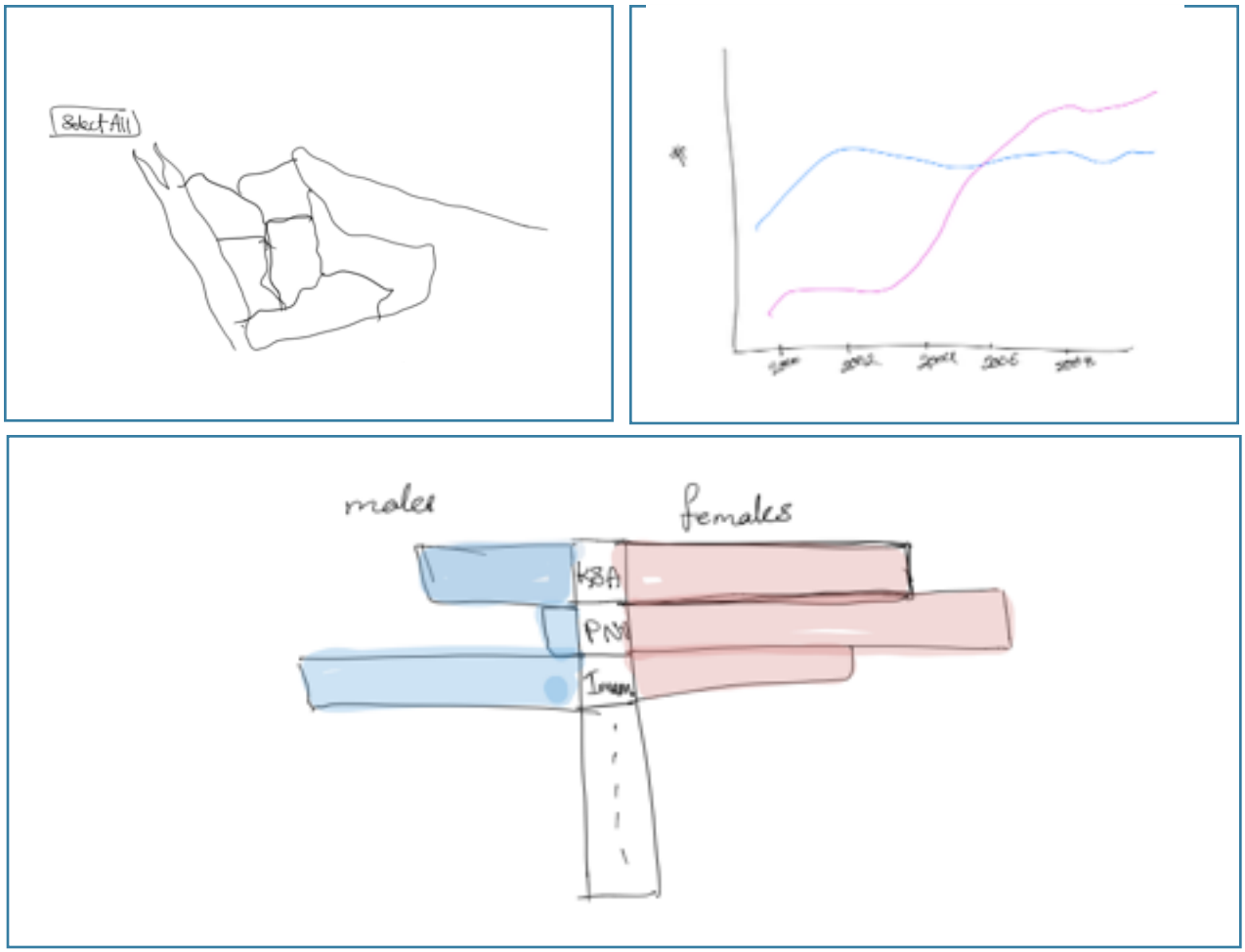


The webpage consist of three parts: a toolbar , a spreadsheet format to show raw data, and a top and left view that provide users with the ability to modify configurations .

The **toolbar** have these features: upload dataset, show Data info (which will show a side widget with data information), Add new Column, Export Data, Export Configuration.

The **spreadsheet** will provide the user with the ability to see the raw data and interact with it using the **top and left view**. A user can specify the headers for the rows and columns, filter data, specify datatype..etc.

For Saudi Women in Tech:



The top left chart shows Saudi map divided into regions, when clicked on a region the other charts will update. There is an option to select all region to see aggregate information.

The top right chart shows the number of females vs males over the past years.

The bottom chart shows the distribution of females vs males in different universities. it can be filtered to show specific universities for specific region as well as in specific year.

I picked this design to make the difference clear as human visual system is sensitive to perceiving symmetry cues.

Must-Have Features.

Caleydo Importer (some of the features are already implemented):

- An intuitive way to drag and drop or upload a data file into the importer

- Parse data and display it in a table format

- Identify the following: Separator, data type for each column, header of the data

- Display proper visual elements for each column depending on its data type

- Allow users to be able to select/deselect data values from the data set to be included/ excluded from the configuration

- Export configuration to JSON file

Saudi Women in Tech:

- Allow user to select region to view its data

- The main view will compare male enrollment to female enrollment

- Filtering data for studying abroad vs local

- Show different category for students (BA, MS, PhD or undergrad / grad)

Optional Features.

- For the importer to be able to extract information from text field.

- Being able to combine data from multiple resources given that it has the same fields.

Project Schedule

Week	Date	Deadlines	Deliverables	notes
11	Nov 4		Caleydo Importer	Importer UI
12	Nov 11	Milestone deadline	Caleydo Importer	Importer Data wrangler
13	Nov 18		Caleydo Importer / translate Saudi Data to English	Importer export configuration
14	Nov 25		Women in tech	build visual elements
15	Dec 2	Final Project	Women in Tech	integration / polishing