Process Book

CS-6630 Visualization

December 2, 2016

The project title: Higher Education in Saudi Arabia

Name: Asmaa M Aljuhani

Email: asmaa.aljuhani@utah.edu

UID: U6002550

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

CONTENTS:

Title:

Higher Education in Saudi Arabia

Project Description:

An application that visualize Saudi education data to allow ward wise analysis and exploration.

Background and Motivation:

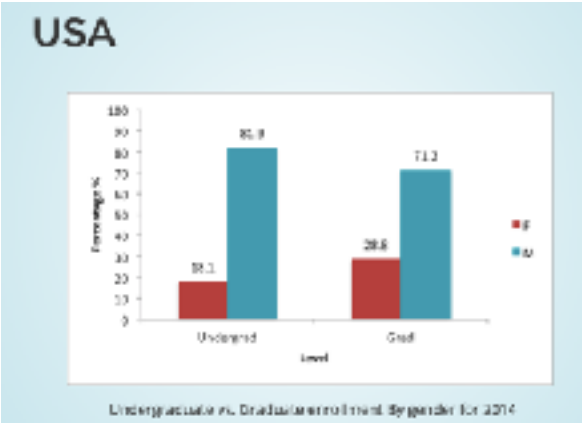
Women are generally under-represented in tertiary computer science education in most countries around the world. The most recent report of the National Science Foundation (NSF) in 2014 shows that only 18% of CS students are women in USA. On the other side, In 2014, 59% of CS students enrolled in government universities in Saudi Arabia were women. Up to my best knowledge, there is no deep researches that studied the phenomena of Saudi women in Tech.

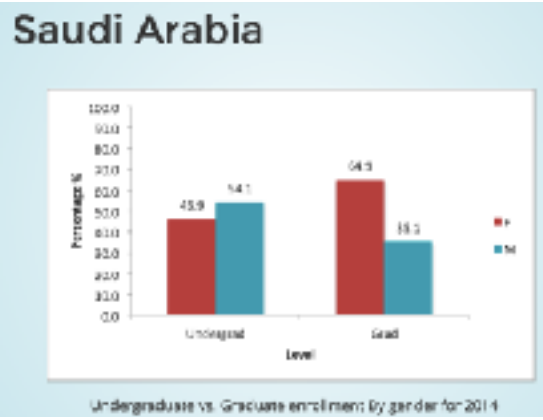
In recent years, many factors influence the enrollment percentage for women in Saudi Arabia. In 2005, the late King Abdullah Bin Abdulaziz Al Saudi responded to concerns for the sustainable development of human resources in Saudi Arabia by launching the King Abdullah Scholarship Program (KASP). Another factor that could impact females studying in CS fields is the establishment of many non-profit organization that empower females in STEM such as ArabWIC (Arab women in computing).

In this project, I would like to explore what subjects are preferred by Saudi female, understand motivation based on momentum factors, and investigate the gap between female and male in STEM fields.

Project Objectives:

At the Rocky Mountain Celebration for women in computing (RCMWIC 2016), I representer ArabWIC organization on behalf of the Saudi chapter. I provided latest facts about female enrollment in CS fields is Saudi and USA. Delivering this information using a simple visualization channel allowed audience to be able to comprehend the given comparison. It also raised a lot of questions about possible factors that affect this phenomena.



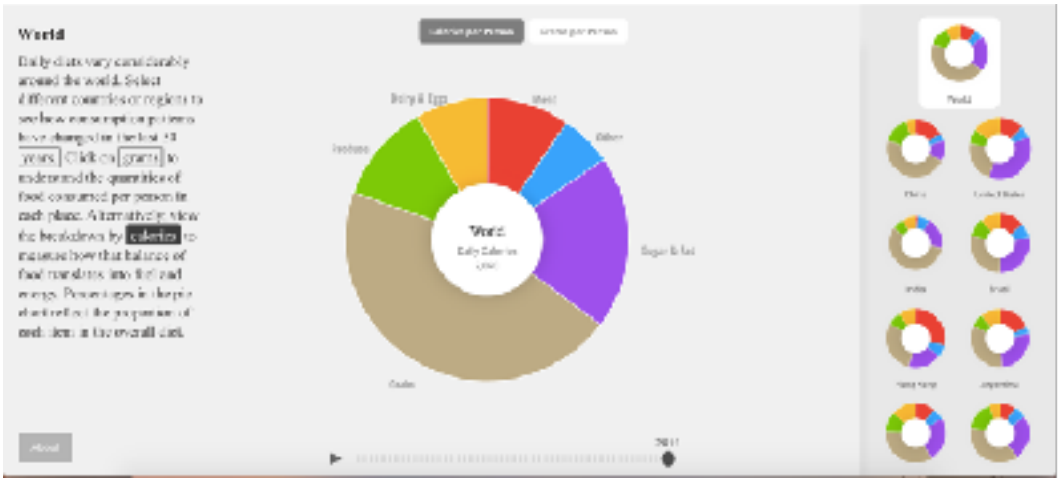


Since most of the researches in this area are still in the preliminary stage, I thought providing data visualizing tool would enable data analysts to effectively discover gender gaps and education trends through graphical means and to represent these findings in a meaningful and effective way.

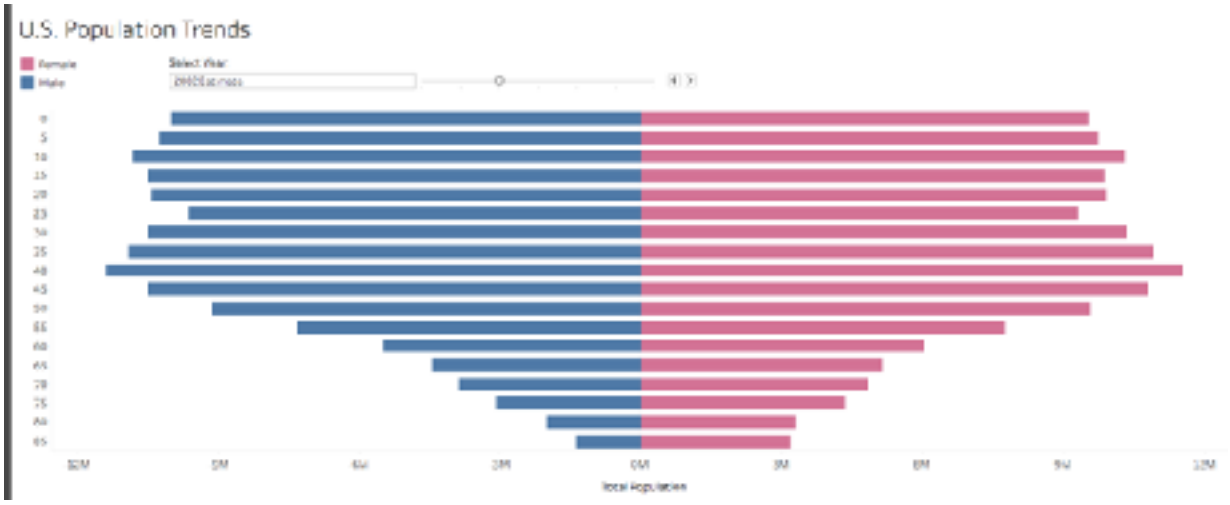
Related Work:

I was inspired by many examples demonstrated in the class. I found most of the related work in the area of comparing gender or education was designed as infographic static images.

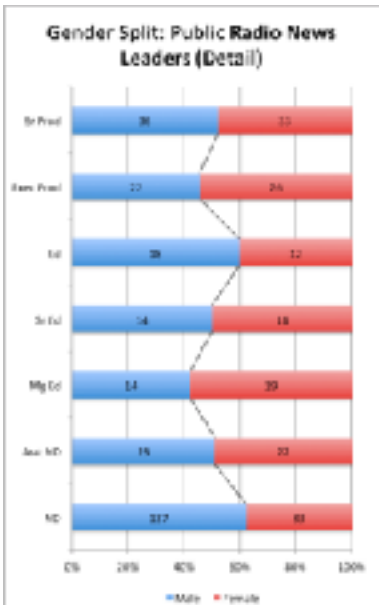
One of the most applications that I based my early sketches on was the fathom.com design. It provide the user with the ability to interact with the small-multiples and provide an overview and a time line.



Another example was one of the Tableau salary examples.



I found a lot o the infographics reports inspiring as it is designed by designers not programmers, hence has a lot of creativity.



Questions:

The goal of this visualization tool is to allow users to explore the data of Saudi higher education, find gender trends in the data, and explore data related to a specific subject. In particular, we would like to answer the following questions:

1) In which year females outperform males in a specific subject?
2) What are the most preferred subjects by females (or males) in specific year (or in range of years)?
3) Compare between local student and students who are studying abroad?
4) What are the trends subjects studied abroad?
5) Filter data based on education level, location, subject, and gender.
6) (A question that I got at the RMCWIC) Why we see a lot of male student in tech but not females?

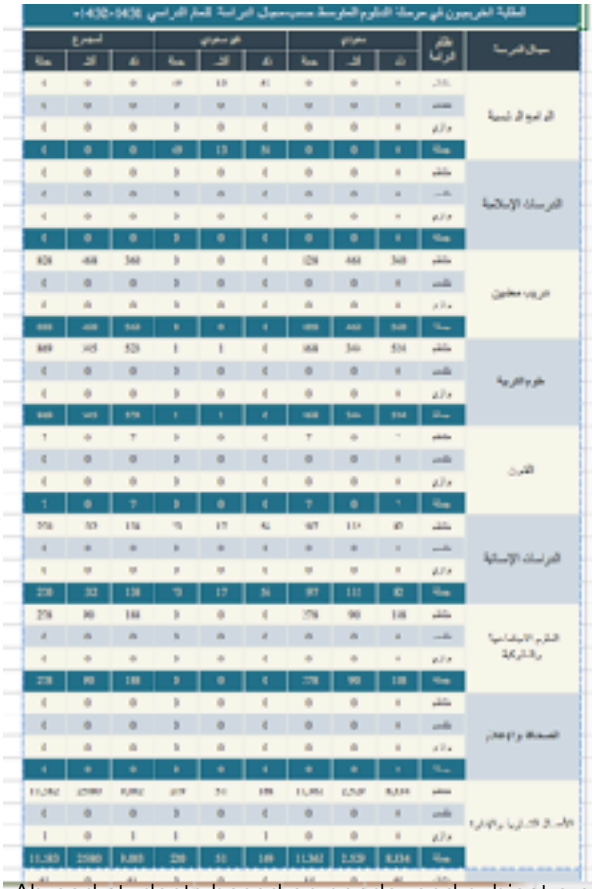
Apart from the above questions, over the course of implementing this tool, I thought we needed a way to filter each value of the data attribute to be able to answer questions that has this

format (I want to know the number of grad female students who are studying Health and Technology subjects abroad). In the future, I'd like to expand this tool by having a dynamic generated overview charts that allow user to compare between different filter results.

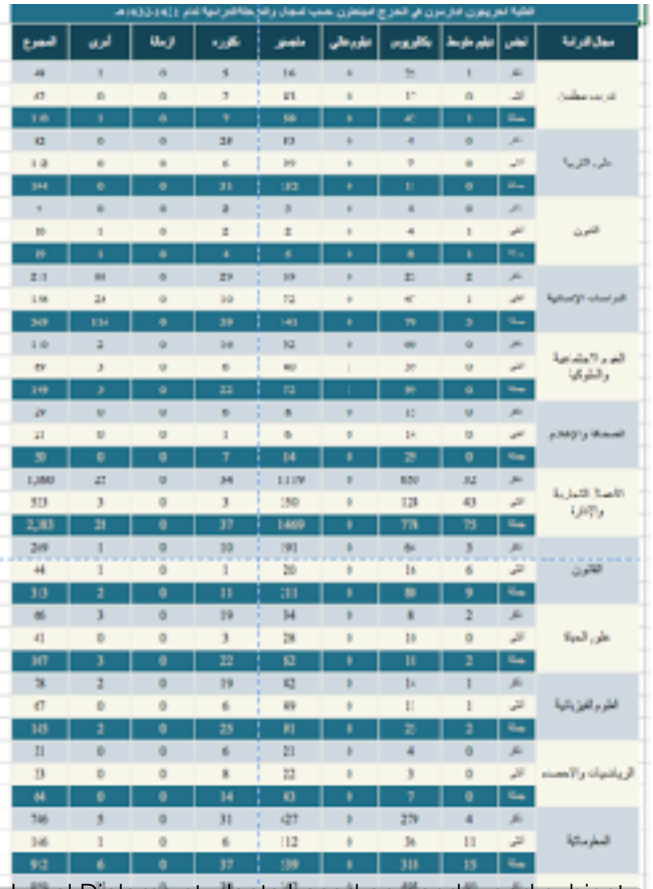
Data:

The main source of the data in this project was the Saudi Ministry of higher Education. It consist of data that goes back to more than 20 years ago (1980). There are separate excel files for each year. These files has several worksheets that aggregate the information based on different factors such as (area of study, level of education, different provinces, local vs domestic students..etc). For the current scope of this project, I focussed on 2 worksheets that aggregate the undergrad and grad of female vs male students based on their study area. The area of study is aggregated in 25 different fields.

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



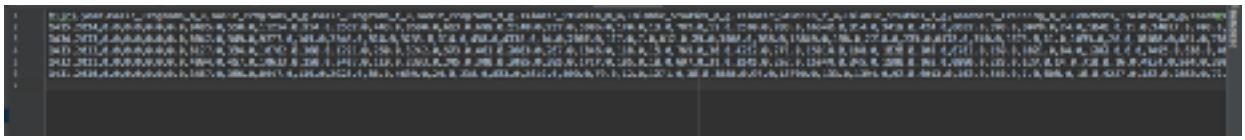
Abroad students based on gender and subject area



Local Diploma students based on gender and subject area

Data Processing:

Data for Saudi women in tech required 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, each group of years have totally different structure and aggregation for the data which made it hard to scrap. Currently, I wrote two scripts to go over these files. the first script worked for the last two years only, the second one worked for 3 years. So I have data for the last 5 years with 25 different subject that shows the number of female vs male and grad vs undergrad.



Data was initially collected in an unstructured wide formate file (that contains years as the first column followed by 4 columns for each subject (subject_male_undergrade, subject_male_grade, subject_female_undergrade , subject_female_grade) that sums up tp 106 columns).

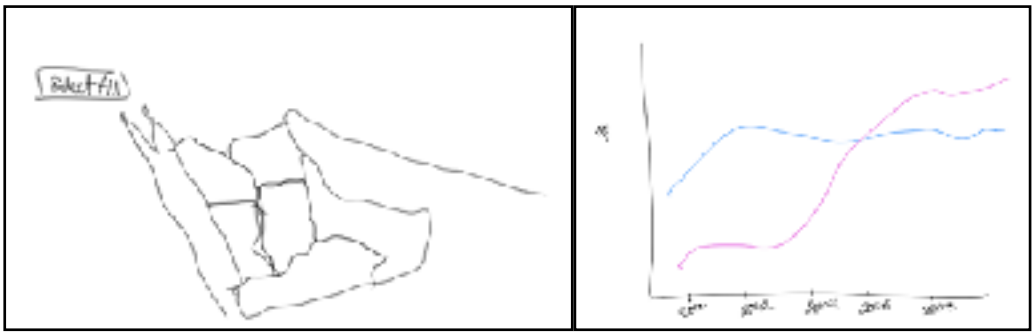
As suggested by a classmate, I restructured the data file to be in a long formate. Here is as maple row of the data

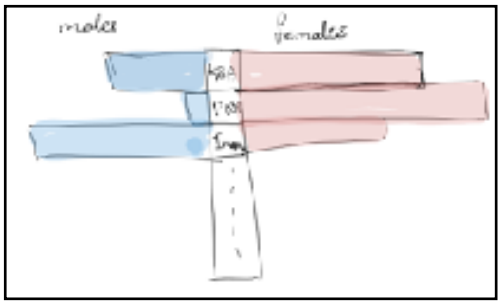
Year	Subject	gender	location	level	value
2010	Art	female	local	grad	5251

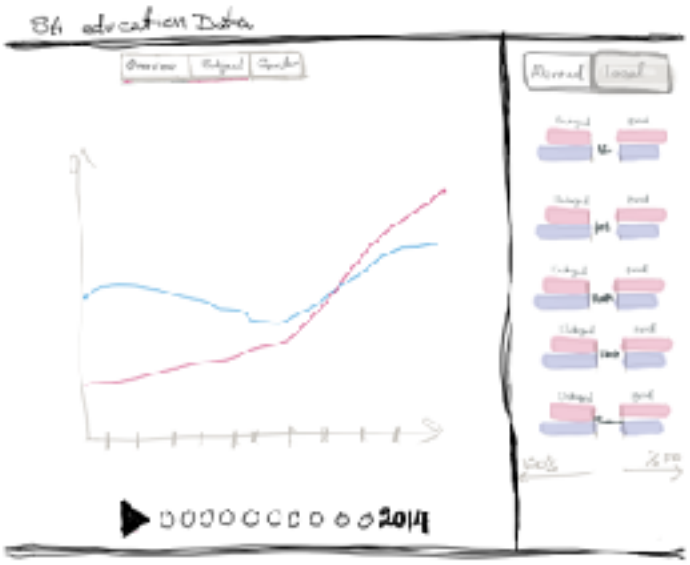
Visualization Design:

I started by sketching different kind of charts. It was hard to define the design requirements. In a time I would focus on uncommon visualization and that would distract me from the main task. I used Sean's design sheets that helped me stay focus on my tasks, and as I sketch a layout, I would revise my main design requirement to check if the sketch match the requirements or not.

My initial sketch to visualize the data had three parts: 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. This sketch was abroad and doesn't focus on the main questions that we are looking to provide an answer to. In addition, location related data is not important in our case.







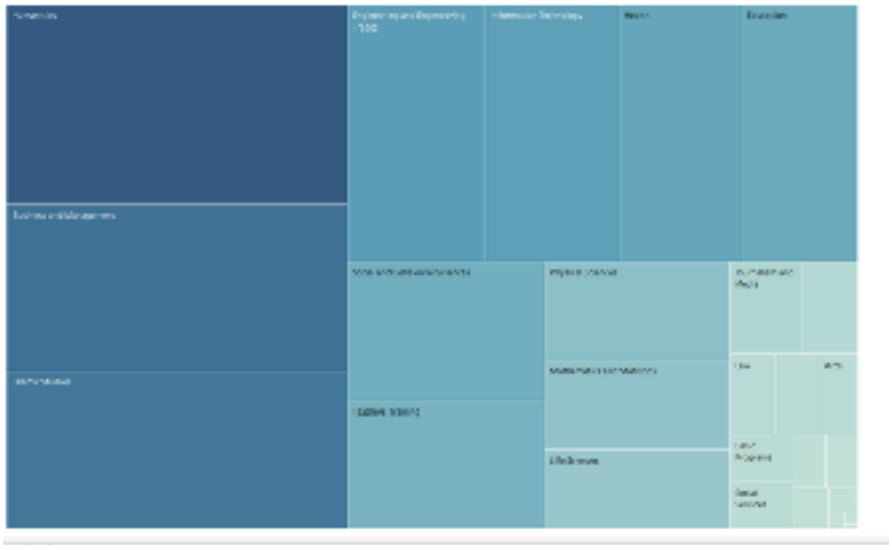
Me second attempt of sketching wasn't that much better as it doesn't show details information about all data attributes.

I then used the design worksheets and discussed different options with Sean and I ended up choosing the final sketch of the design worksheets.

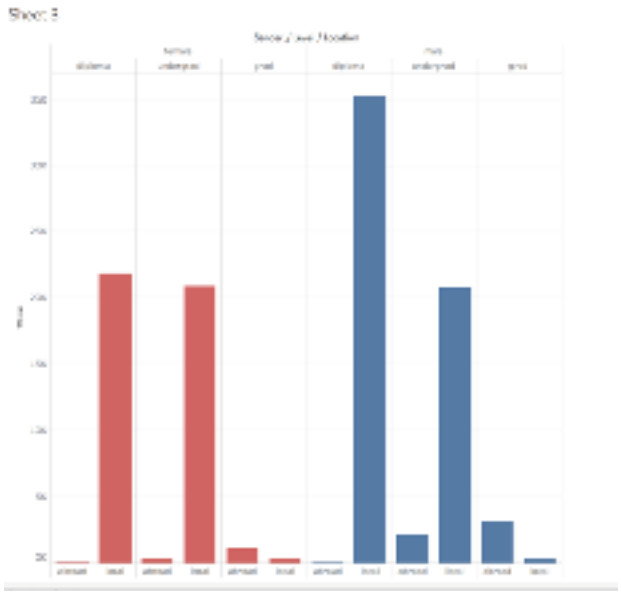
Exploratory Data Analysis:

Because I didn't have any previous knowledge about the data, I didn't know what visual layout would be useful. I needed a way to explore the data in a fast way that doesn't require implementing or coding so I used tableau. Tableau allowed me to bettie understand the data. When I first parsed the data I had two levels (undergrad and grad) where undergrad was the sum for the bachelor degree and the two-year certificate degree (diploma). The result data was showing that the gender gap was huge. so I re parse all my data and had a third level for diploma. Tableau inspire me to choose different valuable layout.

Different charts ideas generated from Tableau:

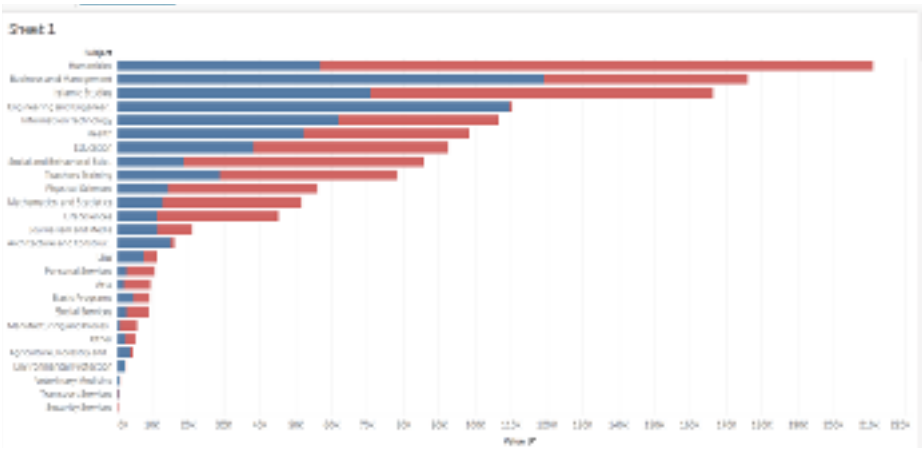


This shows the information for area subjects and the size represent the aggregate number of students.



I tried plotting multiple attributes on one layout and it wan't informative.

Basic charts were mush useful and informative.

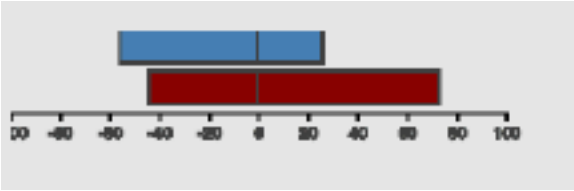


Design Evolution:

As mentioned earlier, I sketched couple layout then used the DAF-worksheets which allow me to put all layout together. Then as I began implementing the design in D3 and html, I realized that most of my layouts are in bars either horizontal or vertical so I redesigned couple of the charts to allow better interactivity and comparison. In my sketches, I evaluated each one based on the design requirements. Some of the sketches required too many clicks from the user, others lacks details of data attributes. I ended up choosing the dash board sketch which provide a viewer with an overall layout as long as the ability to filter each attribute of the data.

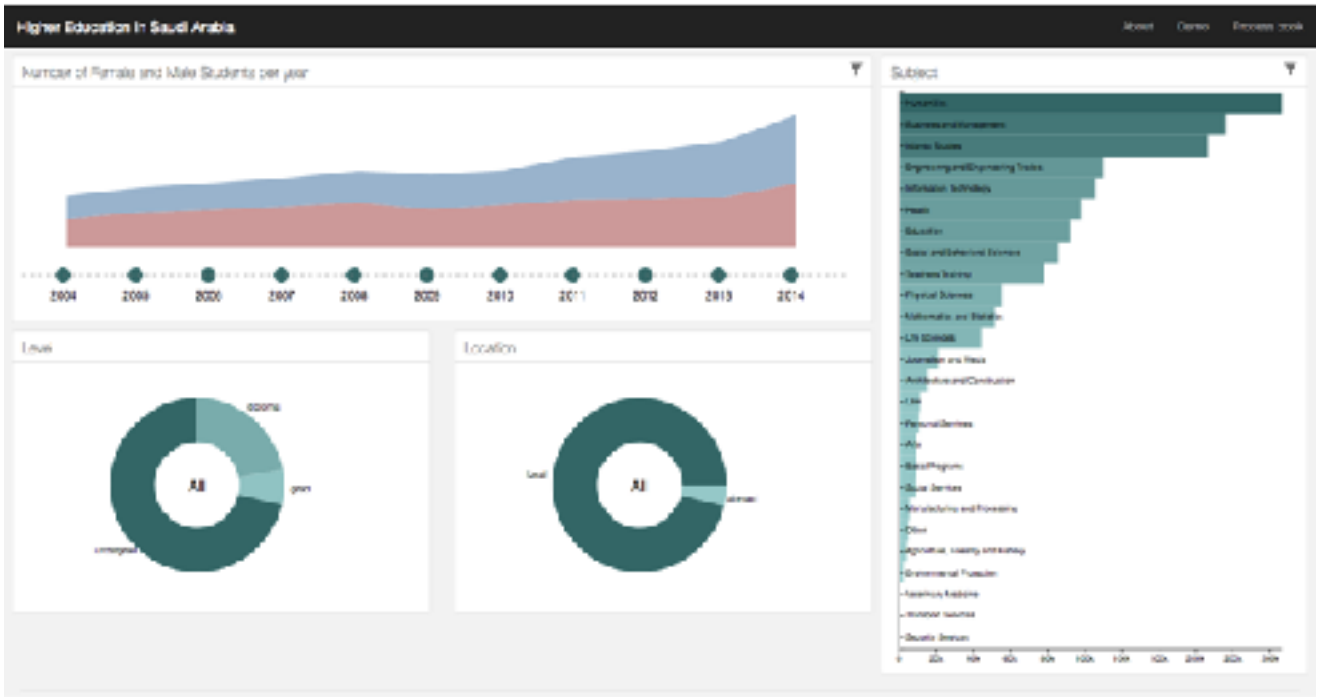
One of the implemented charts that I thought was informative but tend to be more confusing was using the symmetric chart in comparing between levels and gender. My main concern with the current design is the visual elements that is used in the small multiples:

This chart is showing real data for technology field as the following:



	Undergrad	Grad
Female	(44%)2056	(73%)73
Male	(55%)2592	(26%)26
	100%	100%

The final design adopted the dashboard approach in which a viewer is provided with an rich interactivity features that updates all the view and provided interesting findings.



The dashboard is composed of 4 main widgets. The top left is the main view where a viewer can see the gender differences for the years in the dataset. The right widget has a bar chart, each bar represent a subject. It is color coded based on the number of students. the two donut charts are fro filtering level and location. Donut pie was ok to use here since most of the time undergrads are more that grad. Diploma degree might vary depending on the subject. Further explanation about the interactivity between the charts in the implementation section.

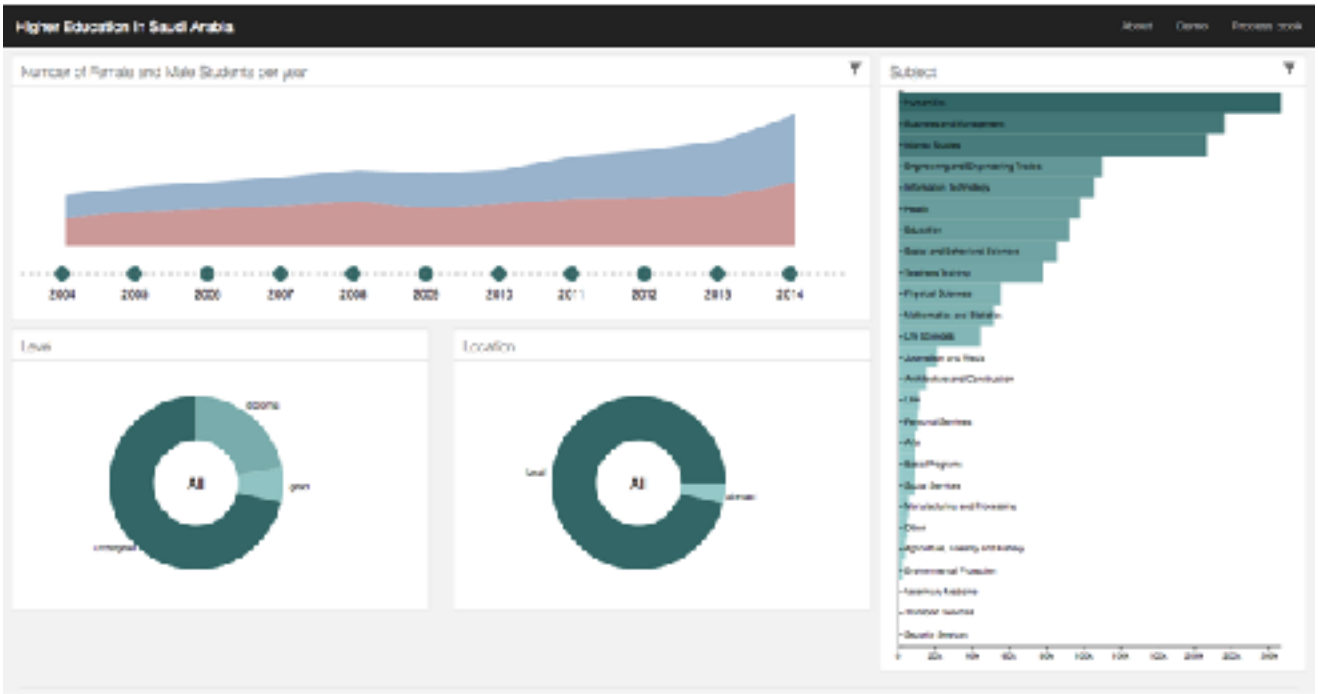
Implementation:

I used the dataset described earlier to build this data visualization dashboard. This dashboard consist of 4 widgets. The first one act as an overview of the total number of student decided in two gender area over the last 10 years. the second widget is for the subjects. it allow the user to select a single or multiple subjects and update all other widgets according to the selected data. The last two are has donut layouts and filter the data based on the level and the location.

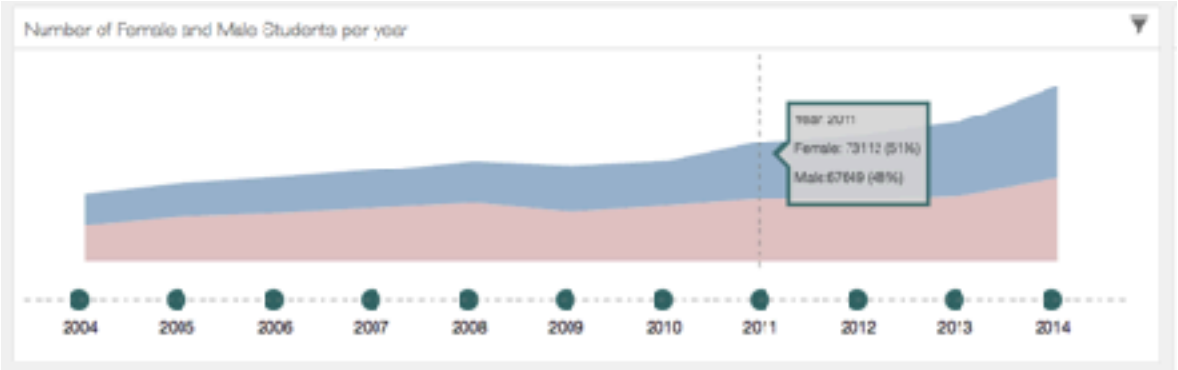
I used D3 javascript library to implement the visual charts along with crossfilter.js library that provides a fast multidimensional filtering for multiple views. And to control the events across the dashboard , I used d3 dispatch javascript library.

For the main layout of the dashboard, I used a template from keen.com.

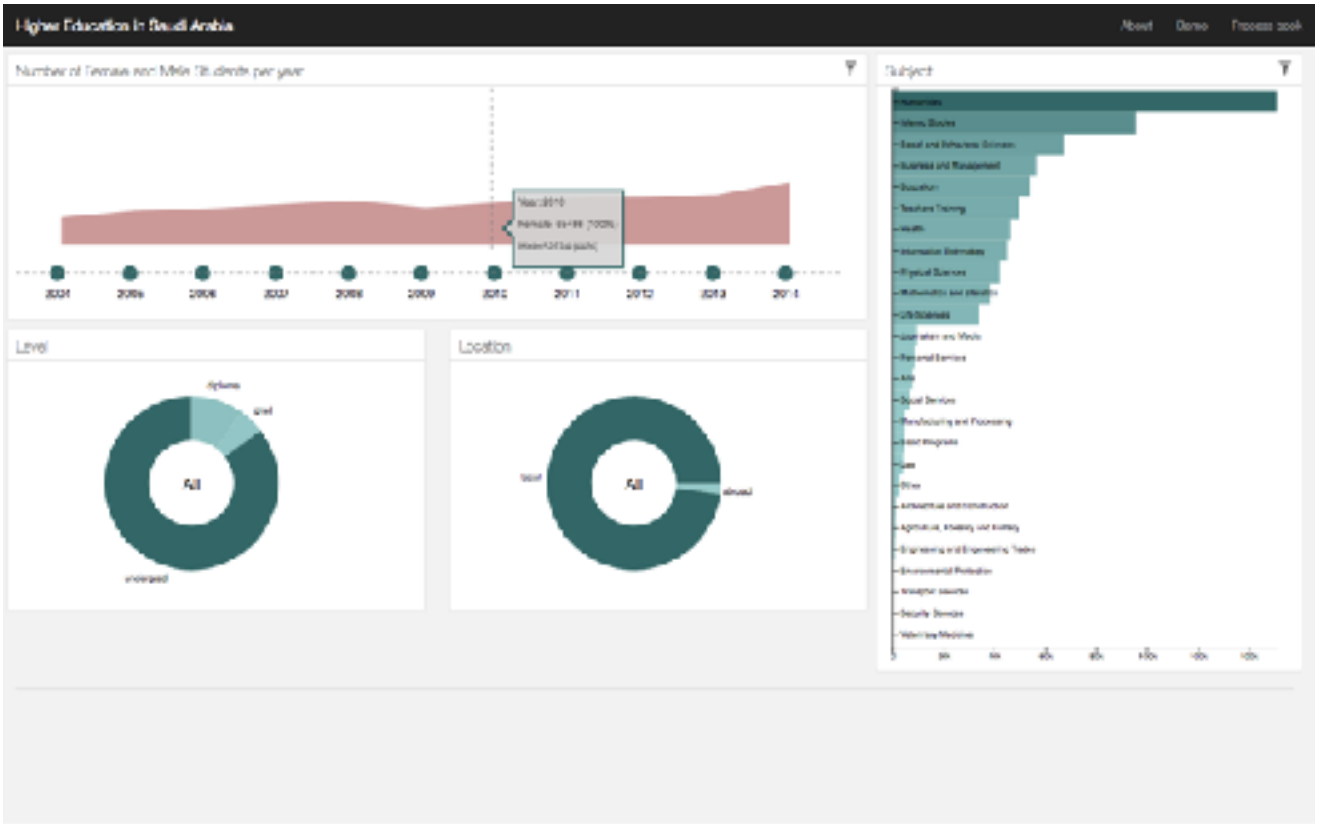
This is a walk through the visualization and the interactivity features of the project. This is the main initial view with no filtering. All data aggregated in a way that provides an overview to the user.



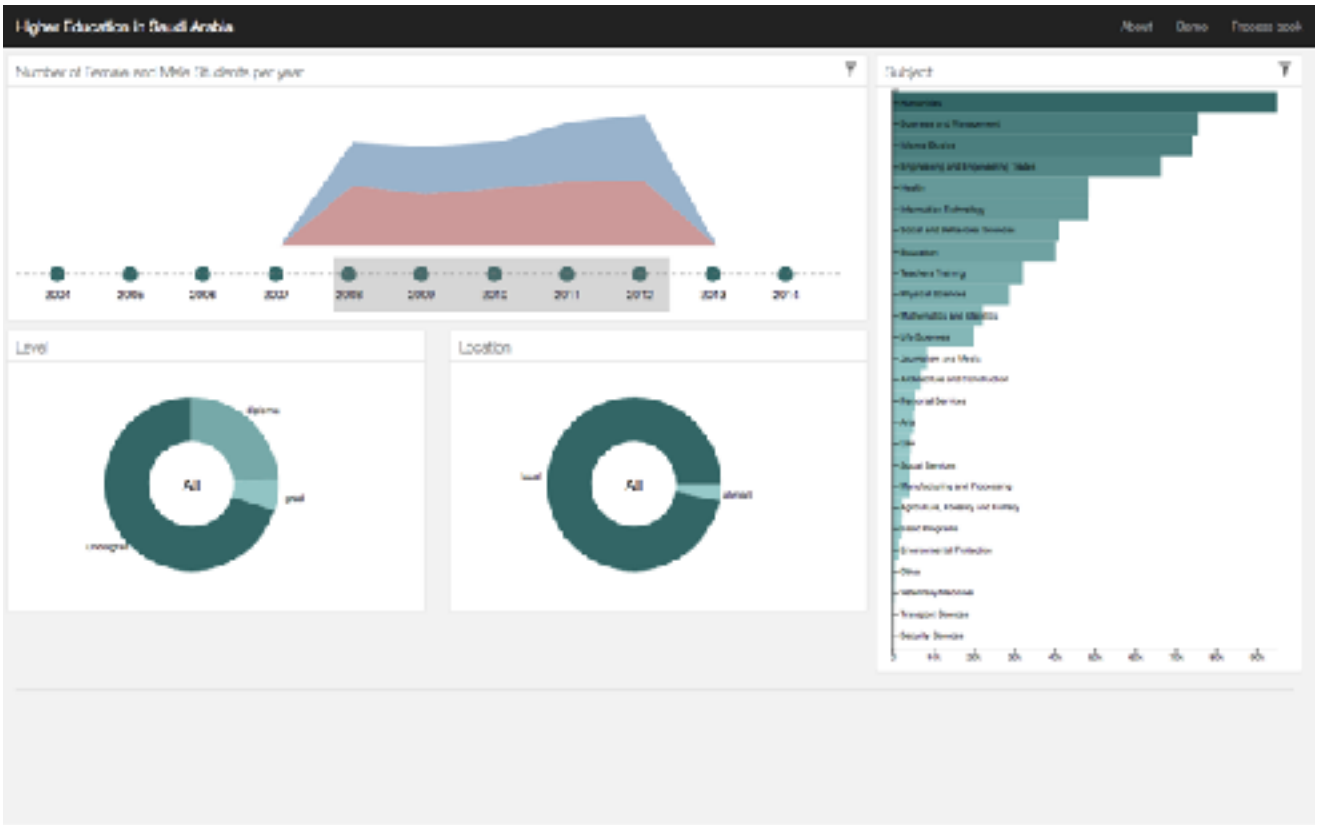
In the gender per year chart a user can hover to see the actual numbers and percentages of each gender for a specific year.



A user can select one gender to filter the other charts.

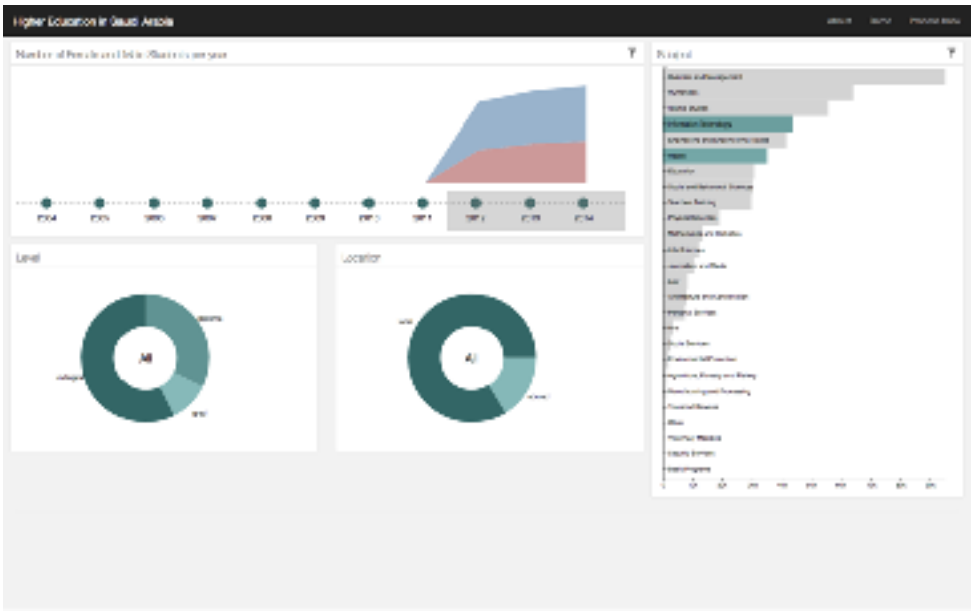


A user can also brush over the year line to filter data related to years.

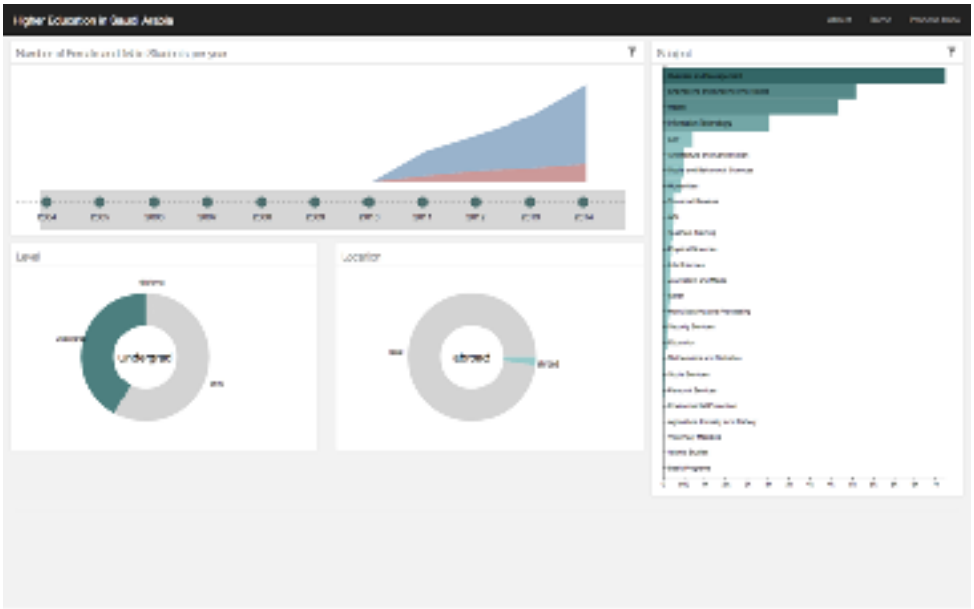


For the subject chart, a user can select one subject to and then brush over the years to see how the order of the subject over the years. In the upper corner, a user can click on the filter button to reset the filter data.

A user can select multiple objects and see the aggregate data on other chars.



The pie carts updates with each filter and a user can select one of the arc to filter the other data. The filtered data is showing that the most common subject for undergrad who are studying abroad in business and Management.



Evaluation:

The main take away from this visualization project is the interactivity between the different visualization components. A typical user can derive interesting findings and result from the project. My aim is to provide this tool to people who are interested in leveraging the gap between gender in STEM fields.

I evaluated the this projects based on the ability to provide a visual answer to the suggested questions. A better improvements would be in usability aspect and providing more story related to each year.

- In which year females outperform males in a specific subject?
  by hovering over the years, a viewer is provided with a tooltip that displays he percentages and the actual numbers of students in each gender.
- What are the most preferred subjects by females (or males) in specific year (or in range of years)?
  The subject charts are orders based on the number of students. the common area will be at the top of the list.