Process Book

CS-6630 Visualization

November 12, 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

HIGHER EDUCATION IN SAUDI ARABIA

Background and Motivation:

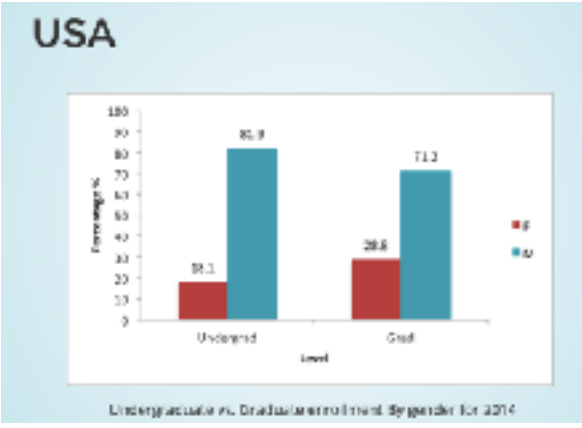
Women are generally under-represented in tertiary computer science education in most countries around the world. The most recent report of the National Science Foundation (NSF) in 2014 shows that only 18% of CS students are women in USA. On the other side, In 2014, 59% of CS students enrolled in government 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.

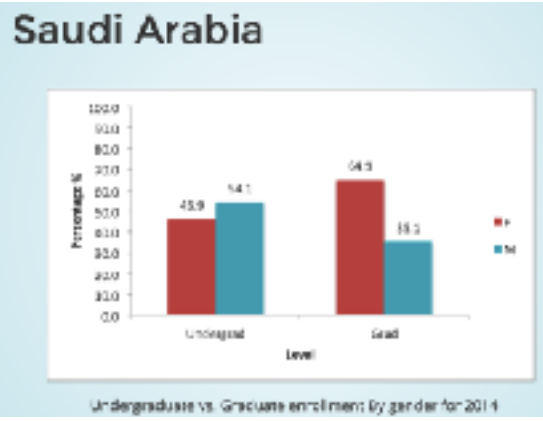
In recent years, many factors influence the enrollment percentage for women in Saudi Arabia. In 2005, the late King Abdullah Bin Abdulaziz Al Saud responded to concerns for the sustainable development of human resources in Saudi Arabia by launching the King Abdullah Scholarship Program (KASP). Another 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. Deliver this information using simple visualization elements allowed audience to be able to comprehend the given comparison. It also raised a lot of questions about possible factors that affect their phenomena.





Since most of the researches in this area are still in the preliminary stage, I thought providing interactive visual information would insight us and allow us as Saudi females to better understand what we are facing.

This visual tool aims to explore the data of women in STEM in Saudi Arabia and try 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.

- What is the influence of the scholarship program on the difference number of female and male in tech.

-

Data:

The main source for the data in this project is 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 am focussing 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

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

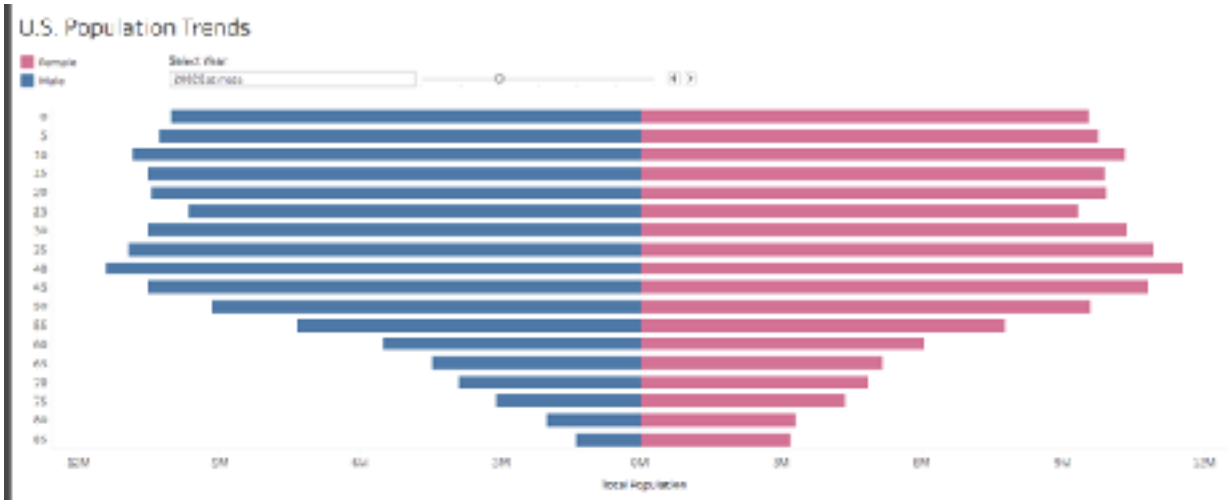
hard to scrap. Currently, I wrote two scripts to go over these files. the first script worked for the last two years only, the second one worked for 3 years. So I have data for the last 5 years with 25 different subject that shows the number of female vs male and grad vs undergrad. By the end of this project, I plan to have data for all the previous years.

I still need to work on the structure of the data. currently I have a file (edu.csv) 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. This didn't work will and wasn't easy to deal with using d3. I want to explore more on how to structure the data. For this reason, I used placeholder data for the visual element provided for this milestone.

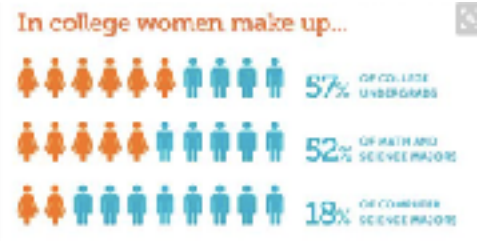
Visualization Design:

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.

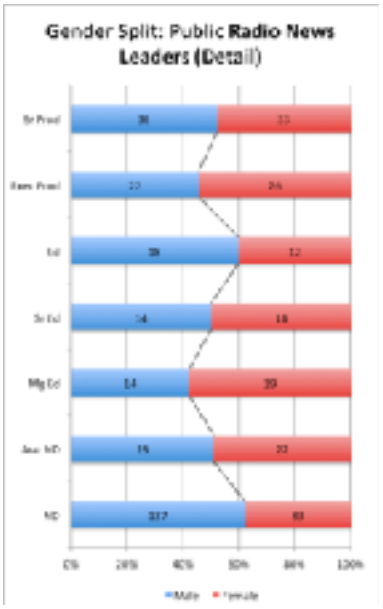
The first visualization is from tableau vis gallery. It shows similar information in which it compares the US population based on gender.

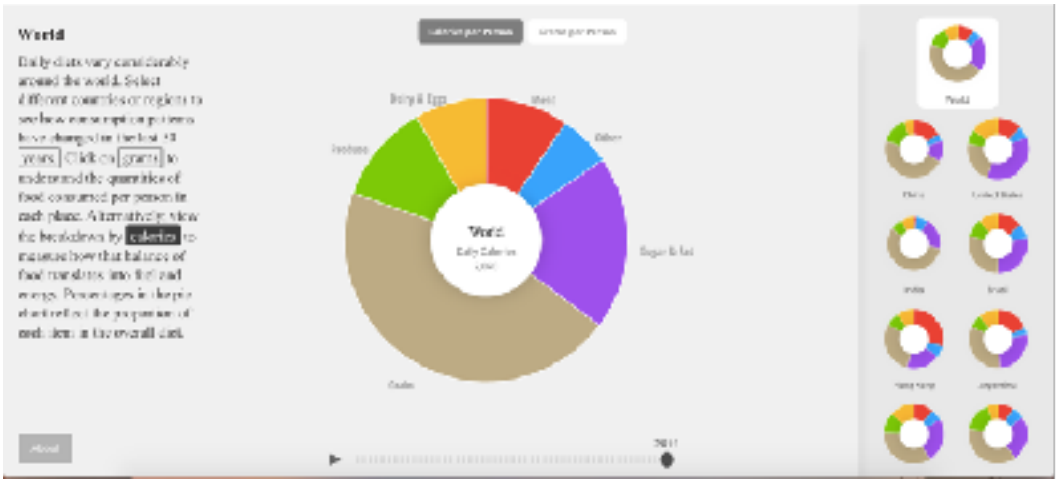


I also liked this part of this infographics, I would probably implement something similar o show that ration between female and male for each subject



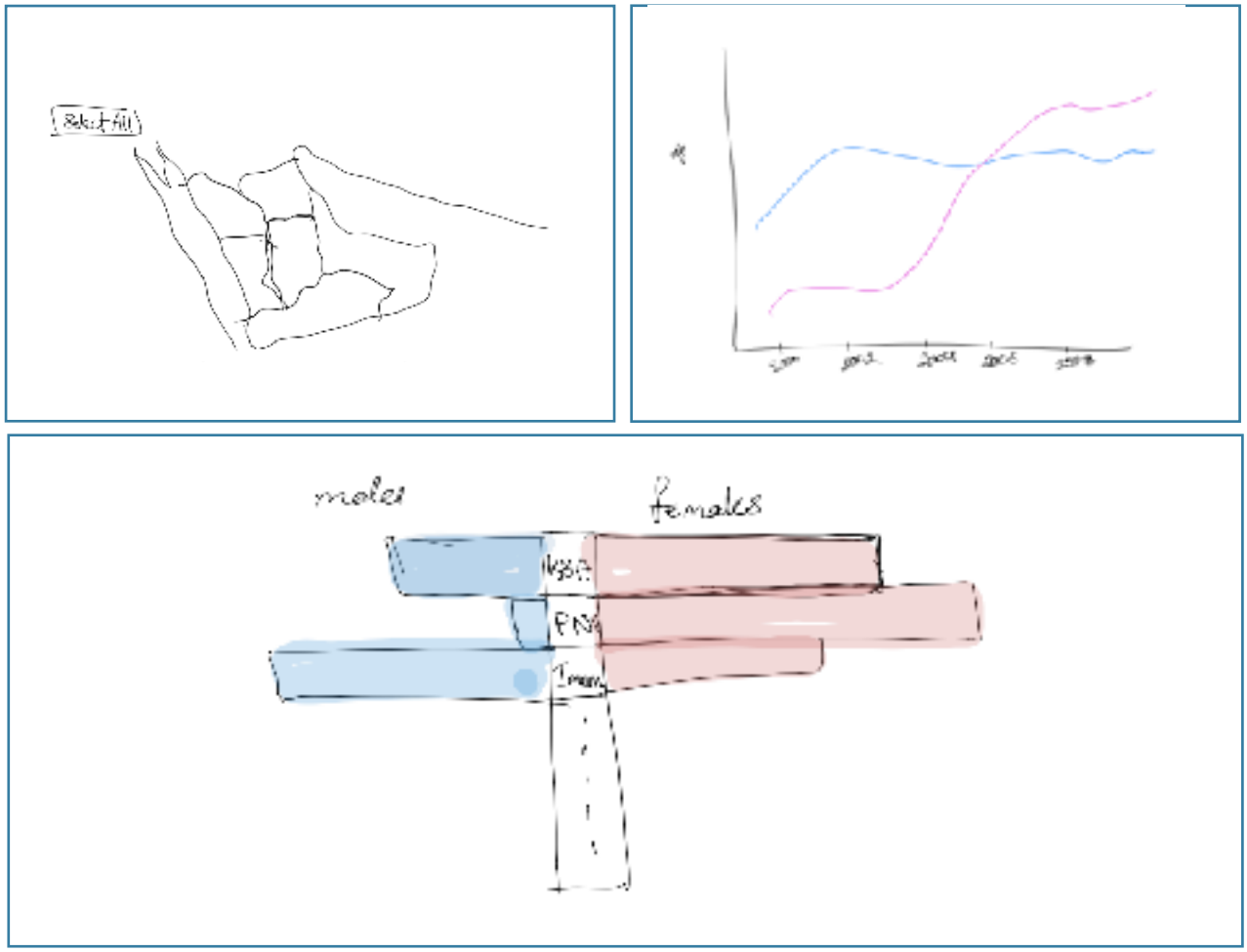
This can be as a small multiple to show the ration for each subject.





Finally, I was inspired by this visualization, I liked the interaction aspect. the interface has 4 components: story (left), main vis (center), years timeline that has smooth animation that goes from the first year to the last one, for each year, all vi element are updated. The small multiples on the right.

My initial prototype 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.

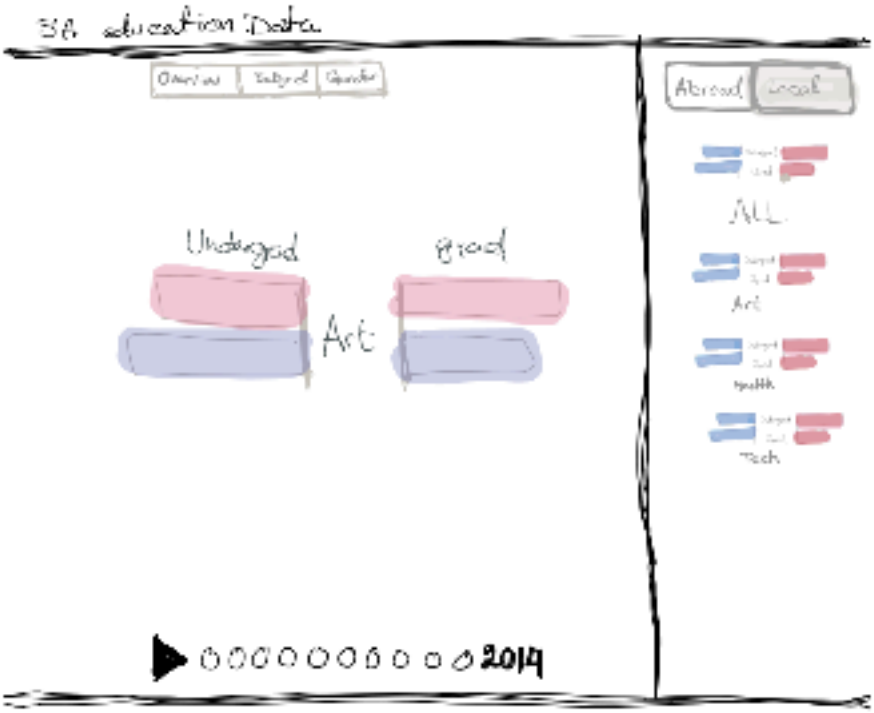
I picked this design to make the difference clear as human visual system is sensitive to perceiving symmetry cues.

Feedback Session:

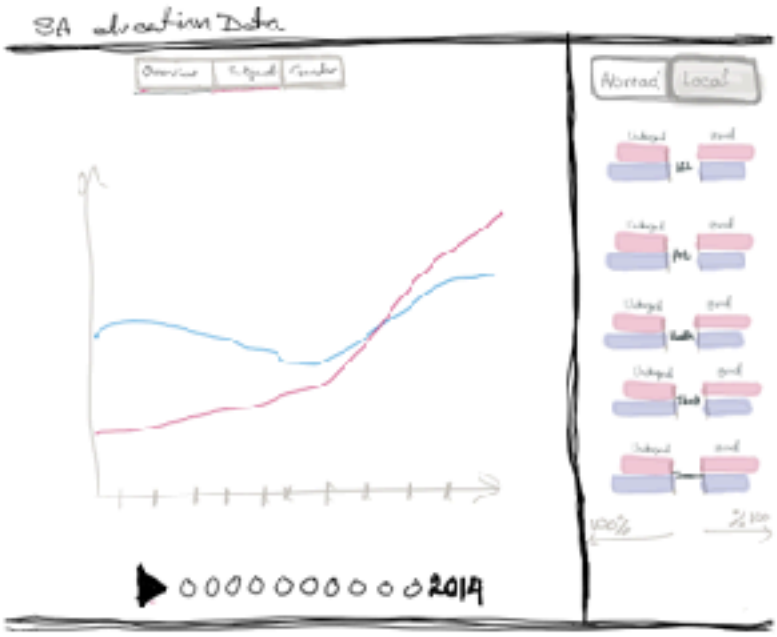
I met with Cameron twice to get feedback on my proposal.

- We discussed the need of the information for other subjects (and not limit it to Tech) so we can explore trends in female numbers.
- Cameron suggested that there was no need for the map as the spacial aspect is not important in this stage.
- He explained the importance of aggregating data and showing percentages instead of actual numbers.
- In the second meeting, we focussed on the visualization
- we came up with specific question that the visualization should answer.

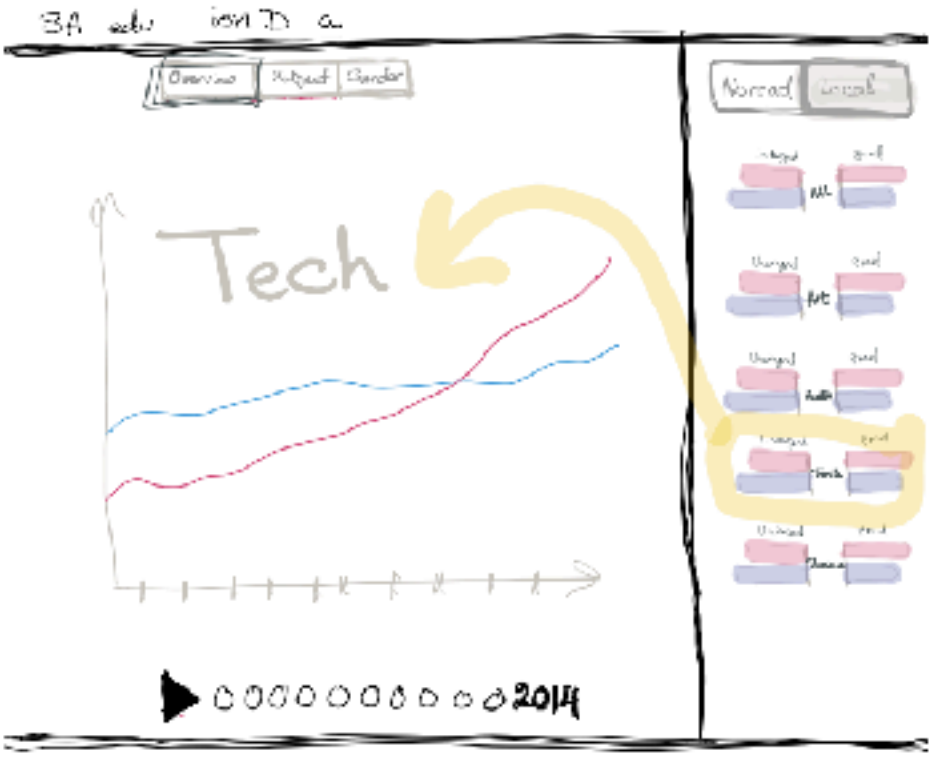
Revised prototype:



This is the main view. it composed of 4sections. the story suction which will update with special events that happen in a specific year. The main chart is for the overview data and a time line for years. User should be able to play/pause the animation and click on any year to highlight it. The right section is the small multiples, this will show data aggregated by subject.
We can show different overview chart, I picked the line chart as it will show clear overview of trend in number of females vs males over the years.

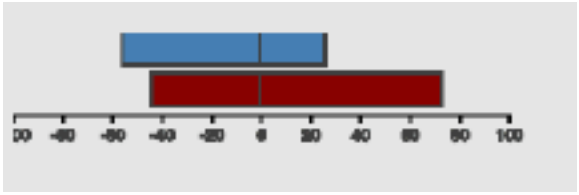


For years time line, these are some ideas.
however due to the limited space I would keep it simple.



When clicking on one of the small multiples, the main vis chart will show the number of female vs male for only the clicked subject.

Evaluation of current design:

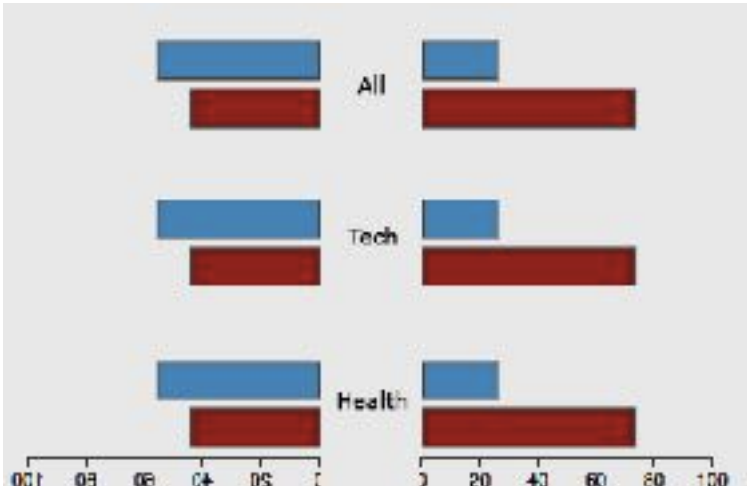


My main concern with the current design is the visual elements that is used in the small multiples:

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%

There is a problem in the current visualization as well as in aggregating the percentage numbers. It is hard to perceive the undergrad and grad bars as two separate bars. a suggestion fix to this chart would be something like this:



This suggestion solution follows the grouping aspect for Gestalt principles. A viewer will compare the female and male bars easily.

With the current design, a viewer might comprehend the given information as the number of grad students exceeds the undergrad and thats not true. so a better way to aggregate the percentages is needed.

What is implemented now:

As mentioned in the data section, data used for the live prototype are random as I need to work on restructure the data to make it easier to aggregate based on different view.

- Data is collected for 5 years
- A line chart for the main overview is implemented.
- Year chart is implemented (animation is yet to come)
- small multiples is partially done, although I need to explore more about different options and visualization.

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		Get the rest of the data/ translate Saudi Data to English	decide on what vis charts will be used for the small multiples
14	Nov 25		Implement the interaction between the different charts	implement the year chart animation
15	Dec 2	Final Project	Women in Tech	integration / polishing