US VISA APPLICATIONS

PROCESS BOOK :CS6630 -DATA VISUALISATION

CONTENTS

1. Project Information
2. Overview and Motivation
3. Related Work
4. Questions
5. Data
6. Design Evolution
7. Implementation
8. Evaluation
9. Resources

PROJECT INFORMATION

TITLE: US VISA APPLICATIONS

PROJECT DESCRIPTION: A visualization design which tells us about how many US Visa Applications over the years.

PROJECT MEMBERS: **Sanket Deshmukh -** u1076955@utah.edu

Ashish Gupta - 1078564@utah.edu

Anmol Ahuja - u1056311@utah.edu

REPOSITORY:https://github.com/sanketd11/dataviscourse-pr-US-VISA-Applications-Visualization

OVERVIEW AND MOTIVATION

In this Project we will be visualizing dataset of US Applications over the period of time. Our motivation behind choosing this project was due to the fact these days US Applications is one of the most debatable topics these days. There are many reasons for choosing this project such as:

- What's the total no of applications being filed over the years.
- Which sector is filing the most visas applications.
- In which category of Visa there has been an increase dramatically over the years.
- What's the success rate of each type of visa getting approved?
- How many H1B visa applications being are filed over the years?

There are personal interest also behind choosing this project as we would also want to be hired on H1B visas in future. So, using this project we can analyse the overall trend and we can get fair amount of information which can be really beneficial for us.

In this project we are going to make visualisations by which we will be able to see a trend in the number of Applications filed over the years. This will include mainly few common visas in which maximum people file such as : F1, H1B, L1 etc so show the trend. The designs will also tell us in which economic sector the most number of visa applications are filed. Overall we can infer from the designs what the trend will be and will be worth reckoning. Our goal is to make visualizations synchronised with each other so that it’s easy to deduce the learning from it.

Related Work

In past also, there have been many visualisations which have shown the data about the applications being filed. In our project we will be focussing on the Economic Sector also which files for the maximum visa applications. We also read a few articles about recent rise in US visa applications. So, we got really interested in doing this project as it fascinated us. After that we saw bloomberg visualisations which seemed very complex but it wasn't conveying that much. Then we decide that we will work on this topic and make simple sober but effective visualizations which can convey a fruitfall message.

Professor Alex also showed few techniques that weren't covered in assignments such as ScrollyTelling and Parallel Plot in of the designs. We found them really interesting and perfect match with our project.

Along with this, our intention behind this project is that we also want to work in US for few years on H1B visa. This is of our personal interest as well so that’s we decided to work on this project. We will be able to find a trend and see what’s the acceptance rate of any particular visa over the years.Lastly, it's a hot debatable topic as President wants to cut down foreign immigrants working in US. This will us to analyse the situation much better.

QUESTIONS

The visualisation project tries to analyze certain statistics of US Visa Applications and will help us giving details to the public and government who is very keen to know what's the current trend.

- In which category maximum visa applications are being filed?
- What's the acceptance rate of any particular visa?
- Which sector is filing the maximum H1b visa?
- What's the salary being offered in different sectors?

So, these are the few questions which maximum Non-US citizens have their in their mind?

DATA

We searched for our dataset on the internet and found various datasets on different websites. We then compared all of the datasets with each other and chose the best dataset which had most of the attributes and more importantly which had more suitable attributes to cater to the needs of the project. Finally, we chose the dataset from www.kaggle.com.

The complete link to our dataset is provided below:

https://www.kaggle.com/ambarish/eda-us-permanent-visas-with-feature-analysis/data

Our Dataset has appropriate data from 2011 to 2016. It also includes information on employer, position, wage offered, job posting history, employee education and past visa history, associated lawyers, and final decision.

DATA PROCESSING

We have done substantial data clean-up on the dataset we got from www.kaggle.com.

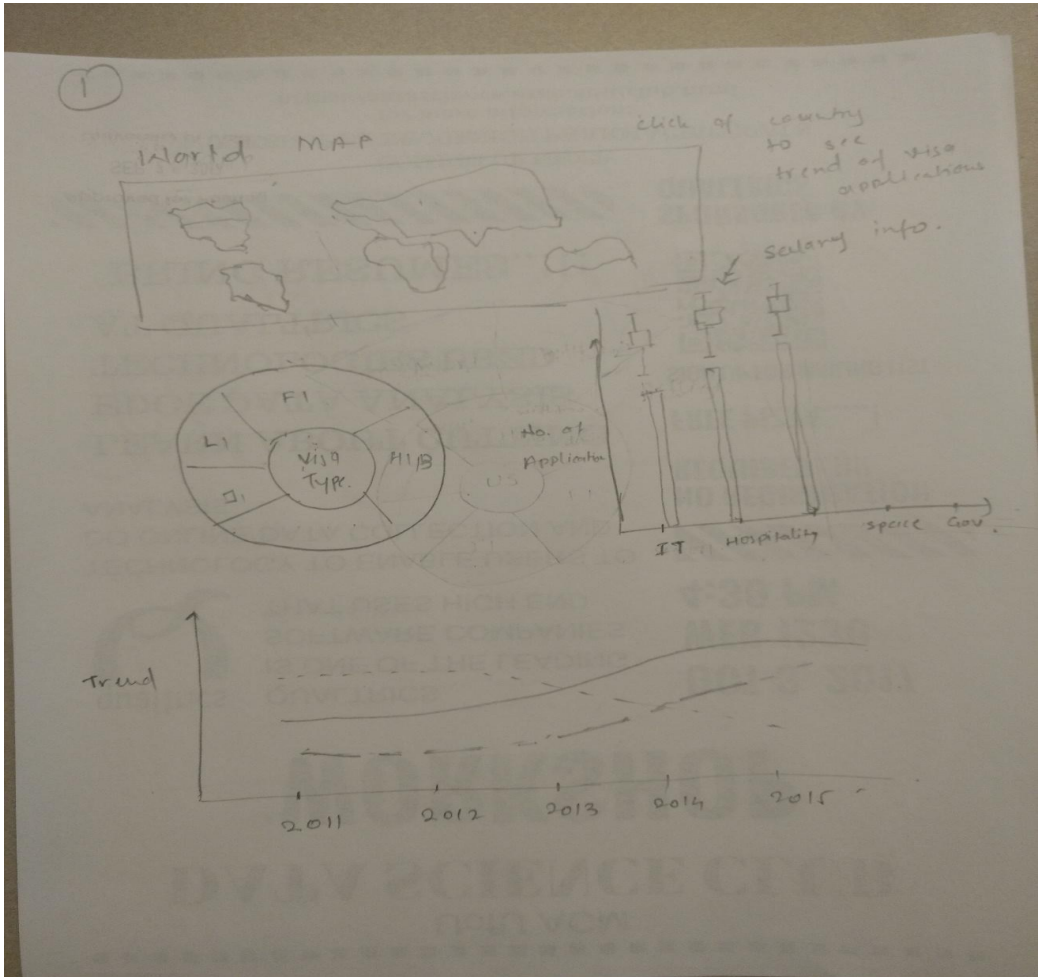
There were many redundant and duplicate columns present in the original dataset. We took only distinct columns in our final dataset which we plan to use towards our project.

Also, there were many rows which had Null values for some of the attributes/columns. We did not include those records in our final dataset as well.

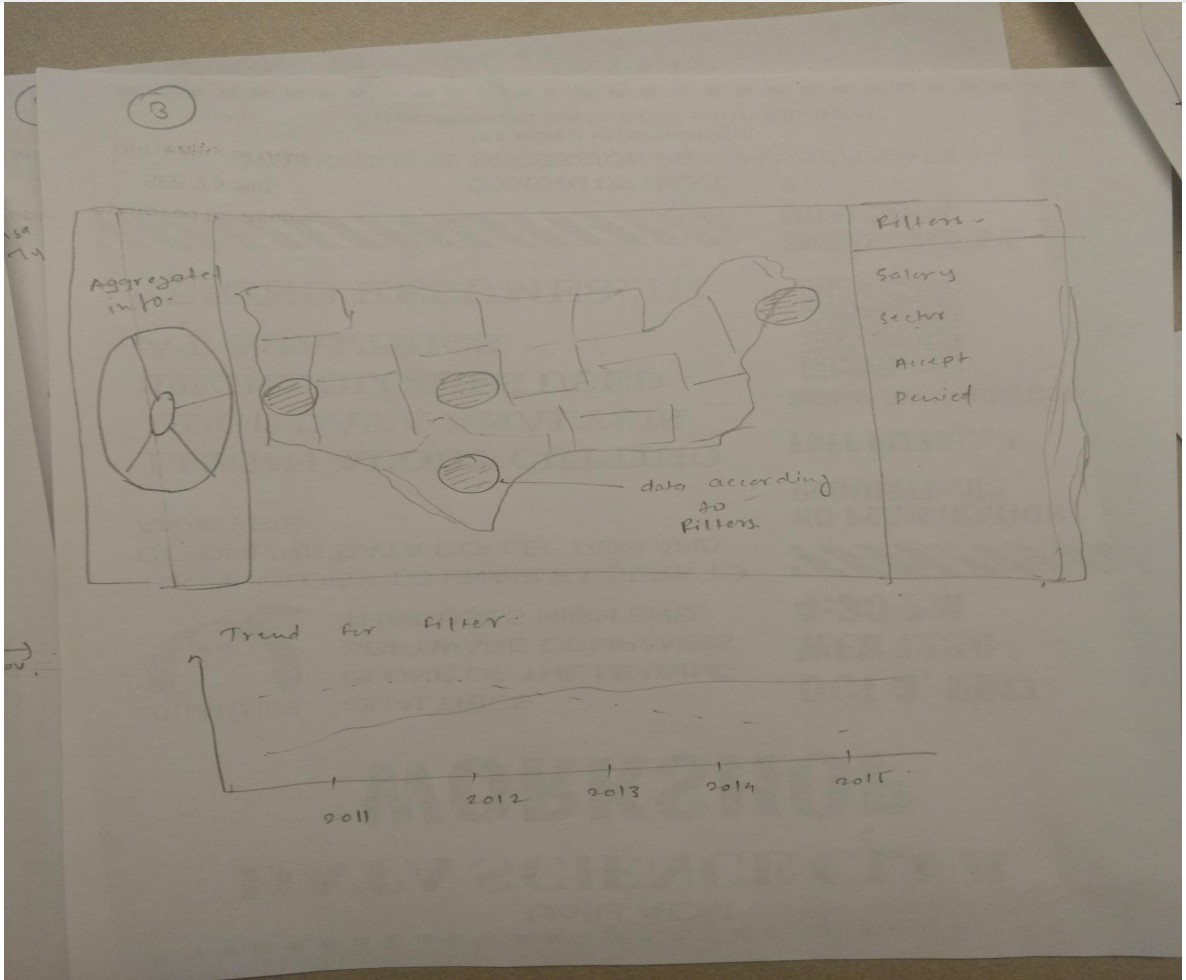
Moreover, there were many columns in the original dataset which had no data. We did not include those columns as well.We did many data validation checks also on our dataset and chose only those records which had correct values for Dates, numbers in Salary and text in Employer_Name, Country_Of_Citizenship etc.

The quantities/attributes which we derived from the original dataset out of 130 columns/ quantities/ attributes include :-Case_Status, Class_Of_Admission, Country_Of_Citzenship, Decision_Date, Employer_Name, Job_State, Job_Title, US_Economic_Sector, Wage_Offer, Wage_Offer_Unit.

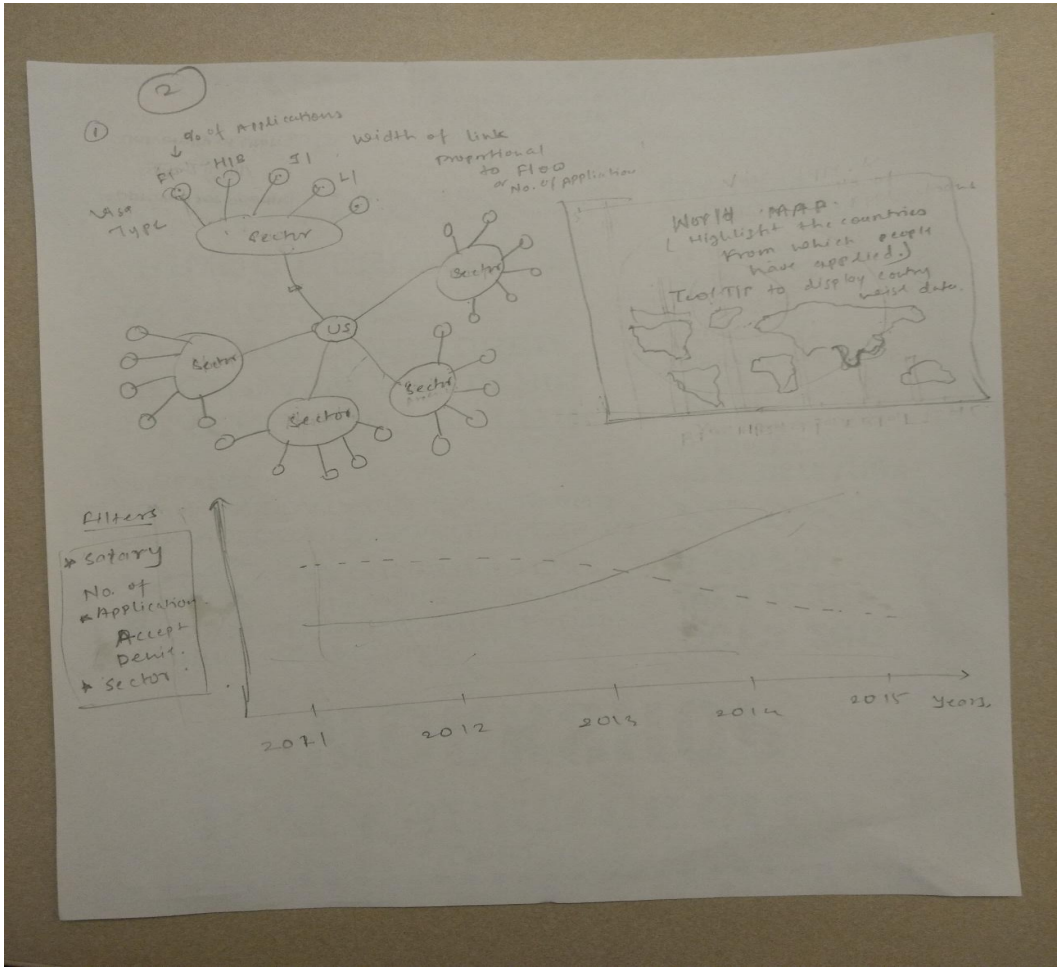
DESIGN 1


1
World MAP
click of country to see trend of visa applications
Salary info.
F1
L1
H1B
Visa Type
No. of Applicants
IT
Hospitality
Trend
2011
2012
2013
2014
2015

DESIGN 2


3
Aggregated info.
Filters -
Salary
Sector
Accept
Denied
data according to Filters.
Trend for filter
2011
2012
2013
2014
2015

DESIGN 3


2
1
No of Applications
Width of link proportional to Flow or No. of Application
F1
H1B
J1
L1
Visa Type
Sector
US
World MAP
(Highlight the countries from which people have applied.)
Tooltip to display country wise data.
Filters
Salary
No. of Application
Accept
Denit
Sector
2011
2012
2013
2014
2015
Years

FEEDBACK

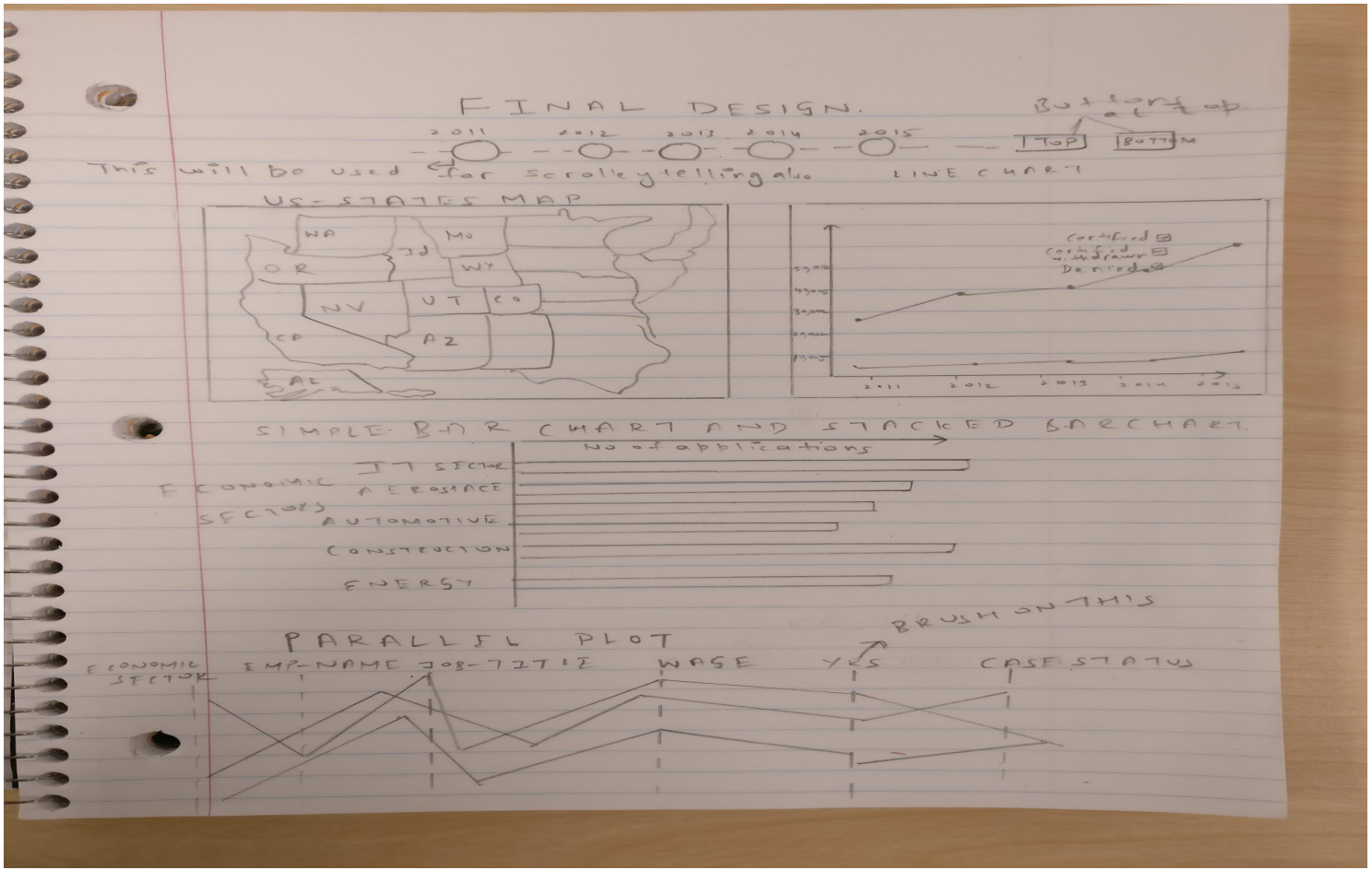
Now, after the feedback which we had got from the other team and plus from TA proved to be really worthy for us. We could come with the final design after incorporating their suggestions which were suiting our project. For example, we decided to include the Parallel Plot instead of the Node-Link Diagram and including the home button and Filters as well to make it look more clean and effective visualization.

There were few imperative points which are really important for us which we found really effective such as we need to make sure that our data doesn't look Biased and the design should look more clean.

So after fetching all the inputs from Feedback Team, Input team and Other classmates we finally decided what all designs we are going to use in our Visualization.

Our main focus is to make Final design elegant and so we derive a learning/inferences from that design. We will be incorporating all the important feedbacks which we have got from our peers and especially the TA's as they have the expertise in that. At this stage feedback played an important role for us as we could figure out what's more important for us. Now, we will start making our visualization step by step and we will make sure it looks coherent and we can derive some significant learning from that.

FINAL DESIGN



Description

In our final design we will be making 4 views namely: US States Map which on hover will the no of applications filed from year 2011-2015. We will be coloring the states depending on the number of applications filed. The we will be using the circular shaped year indicators on clicking that will show the line chart like scrollytelling. They will also be used as Filter as will move on with the next design. After that we will show Line chart as well in which it will be shown no of applications versus no of years showing the case status (Certified, Certified -Withdrawn, Denied). After that Simple Bar Chart and Stacked Bar Chart will be shown having Economy Sector wise Applications filed and Visa Type Wise Applications. In the End we will draw Parallel Line Chart having different Attributes such as Job-Title,Employer Name, Years,Case Status Etc.

IMPLEMENTATION

(FIRST MILESTONE)