EXPLORING
US VISA APPLICATIONS

PROCESS BOOK :CS 6630 -DATA VISUALIZATION

CONTENTS

PROJECT INFORMATION

TITLE: EXPLORING US VISA APPLICATIONS

PROJECT DESCRIPTION: Visualization based on data based on US Visa Applications from 2011-2016.

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

Ashish Gupta - u1078564@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 visualising dataset of US Applications over the period of time. Our motivation behind choosing this project is due to the fact that these days US Applications is one of the most favourite topics. There are many reasons for choosing this project as we want to know:

- What's the total no of applications being filed over the years to observe the trend?
- 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- What's the success rate of a particular visa whether it has been Certified, Denied or Withdrawn ?
- 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 visualizations 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 as to 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 is and will be worth reckoning to know in detail. 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 visualizations which have shown the data about how many applications have been filed. But, in our project we will be focussing on the Economic Sector so as to know which sectors files the most applications. We have also read a few articles about recent rise in US visa applications. So, after that we got really interested in doing this project as it fascinated us in doing more research. After seeing many bloomberg visualizations which seemed very complex on a first view but it wasn't conveying that much for layman man to understand what's happening. Then we decide that we will work on this topic and make simple sober but effective visualizations which can convey a significant message..

Professor Alex also showed few techniques that weren't covered in assignments such as ScrollyTelling and Parallel Plot. We found them really interesting and perfect matched 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-elect wants to cut down foreign immigrants working in US. This will us to analyse the situation much better.

QUESTIONS

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

- In which category maximum visa applications are being filed?
- What's the acceptance rate of any particular visa?
- Which sector is filing the maximum visa?
- Which State is filing maximum visas and the mean salary offered in that state?

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.

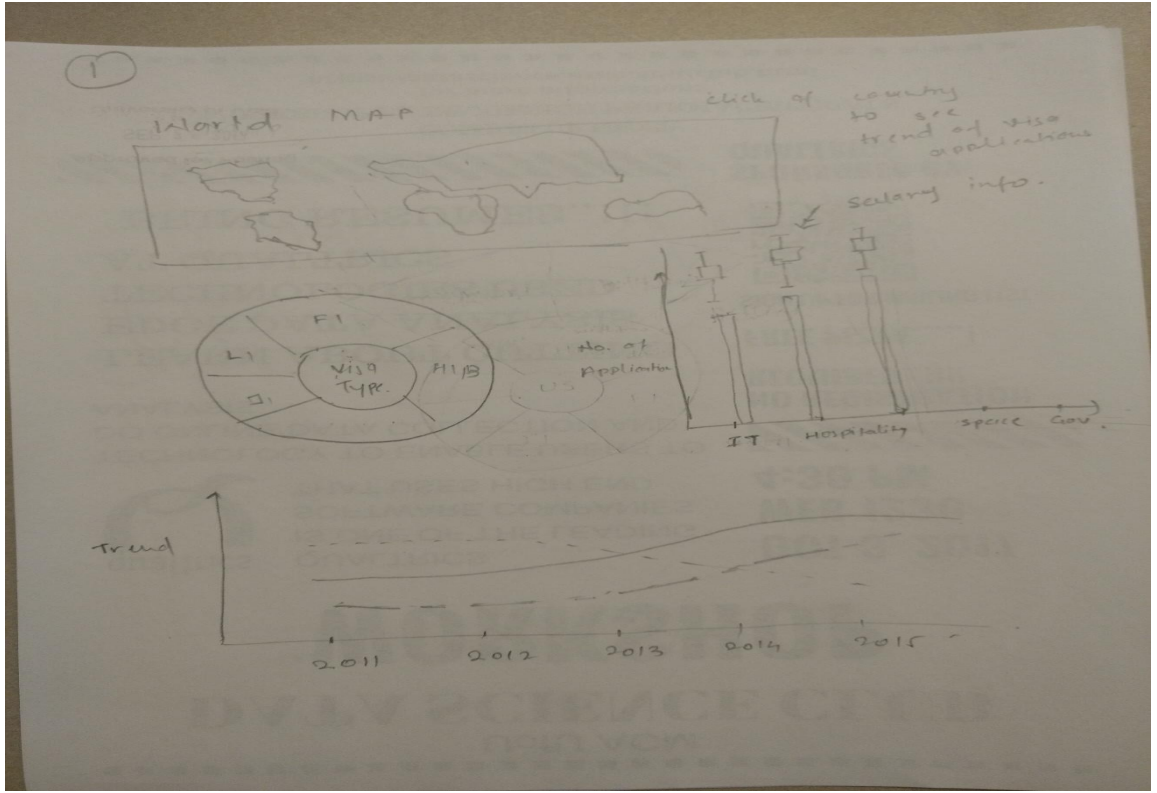
There were many redundant and duplicate columns present in the original dataset. We took only distinct columns in our final dataset which we plan to use towards our project.

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.

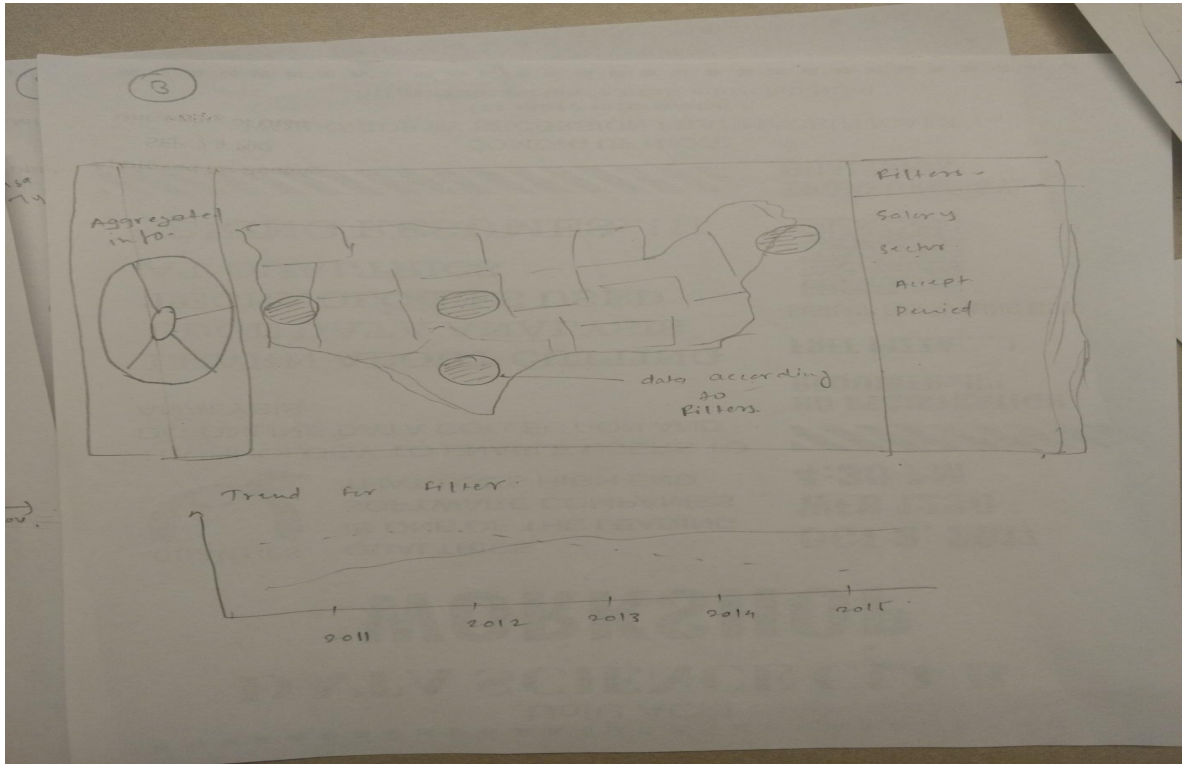
Moreover, there were many columns in the original dataset which had no data. We did not include those columns as well.We did many data validation checks also on our dataset and chose only those records which had correct values for Dates, numbers in Salary and text in Employer_Name, Country_Of_Citizenship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

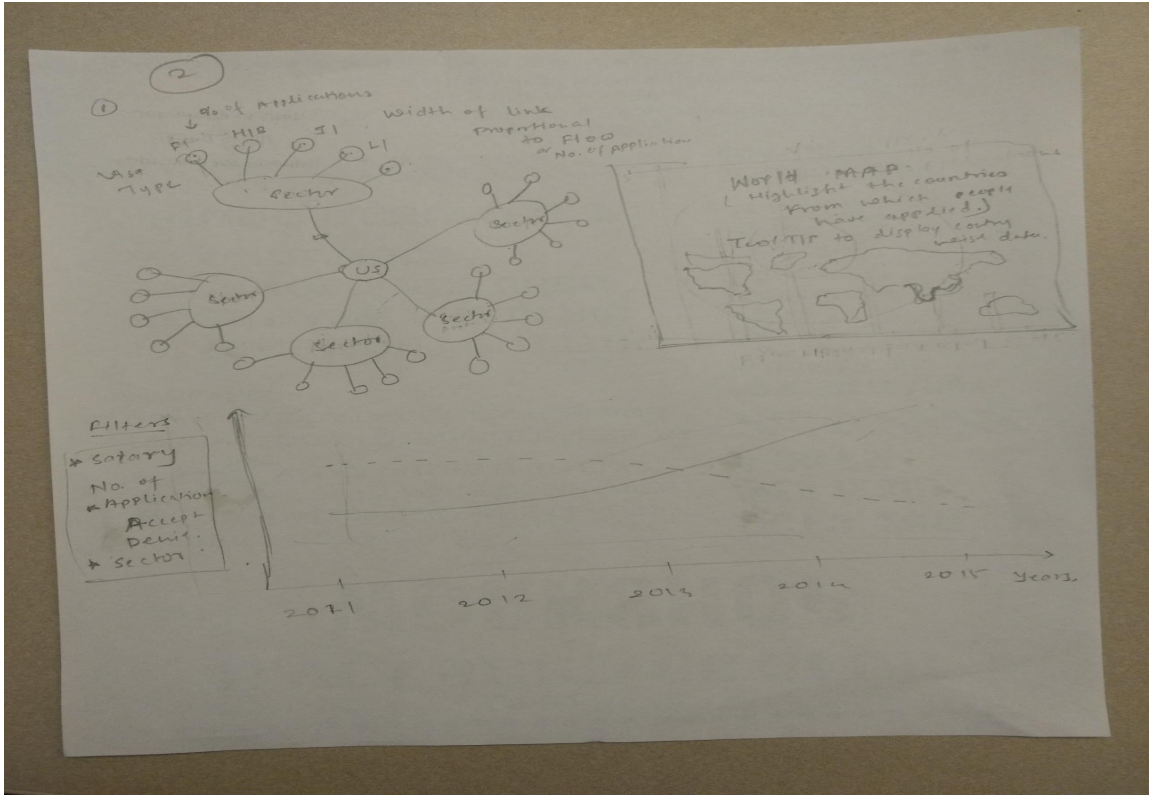
DESIGN 1


1
World MAP
click of country to see trend of visa applications
Salary info.
F1
L1
H1
H1B
Visa Type
No. of Application
IT
Hospitality
Trend
2011
2012
2013
2014
2015

DESIGN 2


B
Aggregated into.
Salary
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 Applications
User Type
HIR
LI
Sector
US
Sector
Sector
Sector
Sector
World MAP
(Highlight the countries
from which people
have applied.)
Tool Tip to display country
wise data.
Filters
Salary
No. of
Applications
Accept
Deny
Sector
2011
2012
2013
2014
2015
Years

FEEDBACK

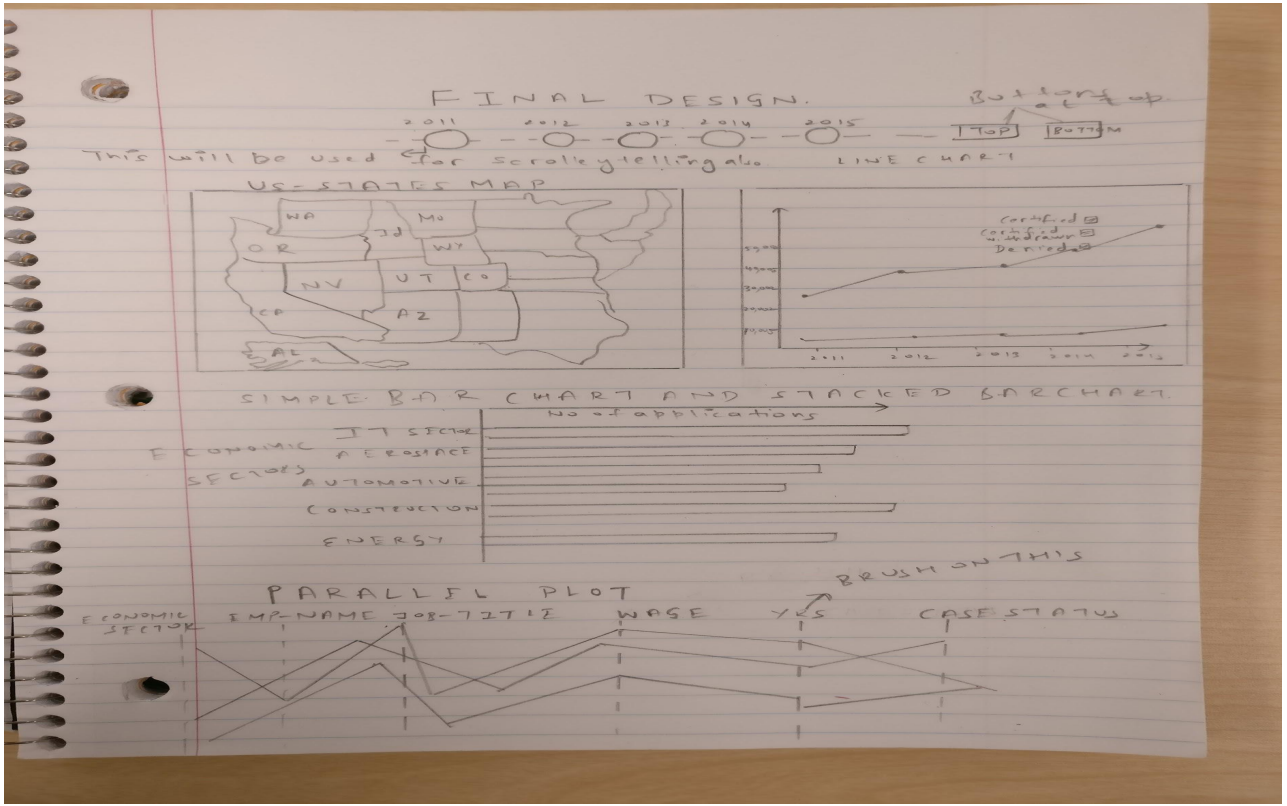
Now, after the feedback which we had got from the other team and plus from TA proved to be really worthy for us. We could come with the final design after incorporating their suggestions which were suiting our project. For example, we decided to include the Parallel Plot instead of the Node-Link Diagram 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.

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.

Design Evolution

In the end , the final four charts which evolved were :

1) US-States Map
2) Line chart
3) Stacked Bar Chart
4) Parallel Coordinate Chart

1) Map Chart - The first chart we will draw is Map Chart. In this ,we are going to show total number of applications statewise displayed by shade of the color. Darker the color it means more is the number of applications in that. Apart from this, we will have a tooltip on each state showing the count of applications shown with a normal bar chart. We can also filter the year using the year wise filter which is placed just above it. Map tells us the number of state-wise applications seen from the first expression.
2) Line Chart : In this chart, we have apply the ScrollyTelling method which is being implemented to show the number of applications of each Case Status such as Certified, Denied , Certified-Expired with respect to the years from 2011-2016. We will make sure that lines are being identified individually so we have given a different color to each line with legend showing the text name of that line.

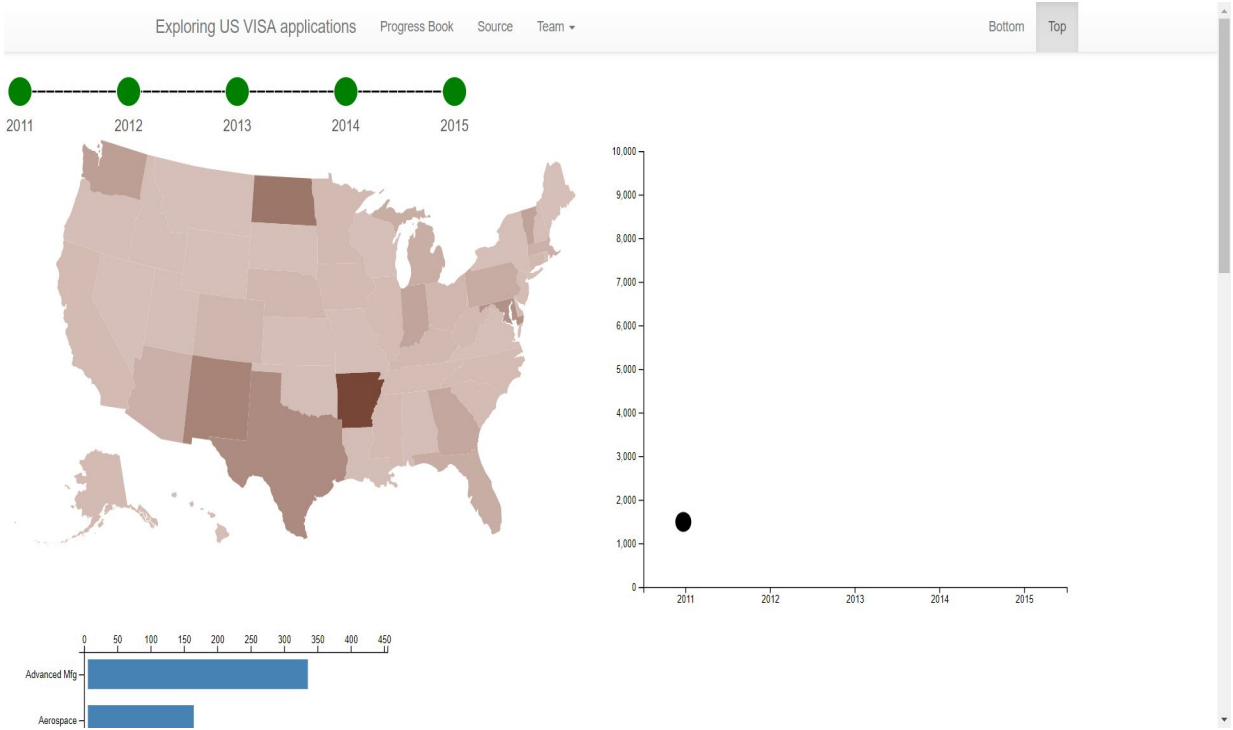
3) Stacked Bar Chart : This is our third design in which we will show the data of different economic sectors over 5 years using the stacked bar chart of each case status (Certified, Denied, Withdrawn and Certified Expired). This helps in showing the data visualization for the different sectors as that's an important attribute to showcase which sector visa is getting Certified ,Denied, Withdrawn and Certified-Expired.

4) Parallel Coordinate Chart:This was the last design which we will use to show the data for different attributes at the same particular time and with the use of brush we can easily see the focus on that particular thing which we want to see.

We chose these 4 as the final designs because of the fact that this suits our project perfectly and helps to showcase our visualization in a much better way which users or viewers will find it easily.

We also took the advice from TA very significantly and had shown what we will be implementing and the TA was satisfied with our project and so we decided we didn’t deviate from our milestone charts and we’re going to implement them only.

MILESTONE


Exploring US VISA applications
Progress Book
Source
Team
Bottom
Top
2011
2012
2013
2014
2015
10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
0
50
100
150
200
250
300
350
400
450
Advanced Mfg
Aerospace

IMPLEMENTATION

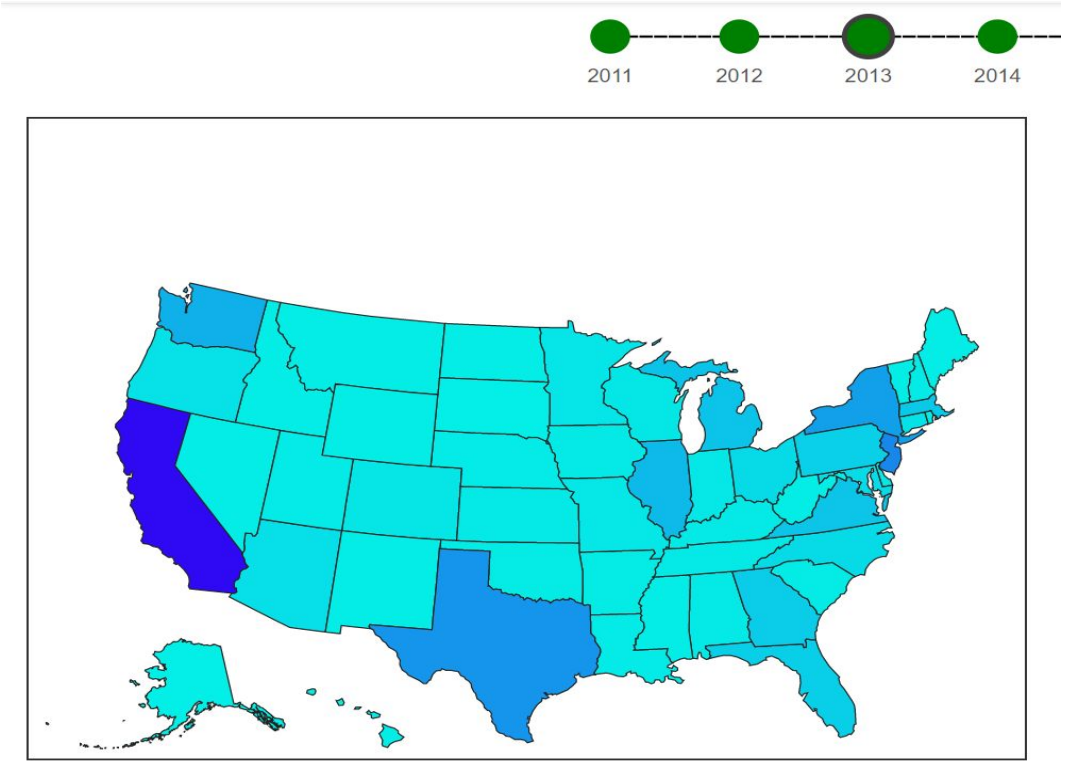
This section discusses the intent and content of each functionality of the interactive visualization components with clear and well-referenced images showing the key design and interaction elements.

MAP CHART

The following are the various functionalities of the Map Chart in our visualization.

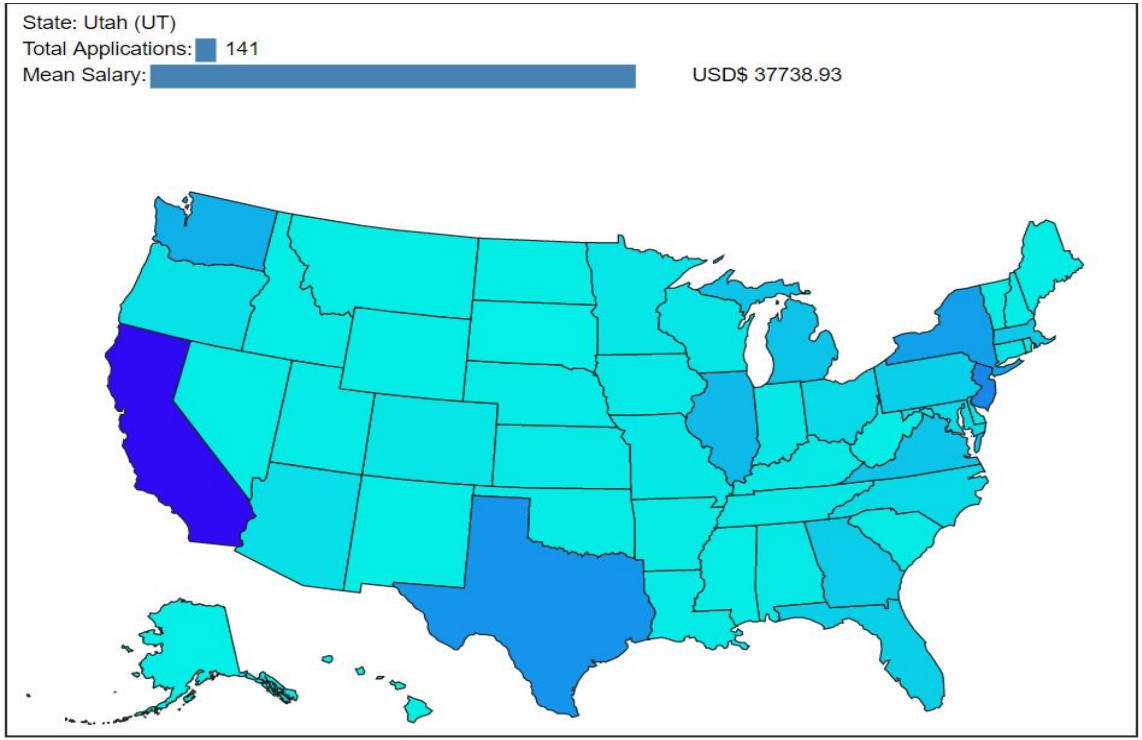
- Basic Map of all US States.
  - This layout shows the map of all US States having a different shade of that particular color from light to dark.
  - The color depends upon the number of applications in that particular year as we can see in the (Fig 1).

FIGURE 1



- Hovering on any State Tool-Tip is shown
- Now, when we hover any state we can see that the name of the state is being shown and the application count of that particular is being displayed(Figure 2)
- User or viewers would find it difficult to recognise name of every state so that’s why we have added the name of state as well.
- Also, to show the count of applications a bar chart will be displayed

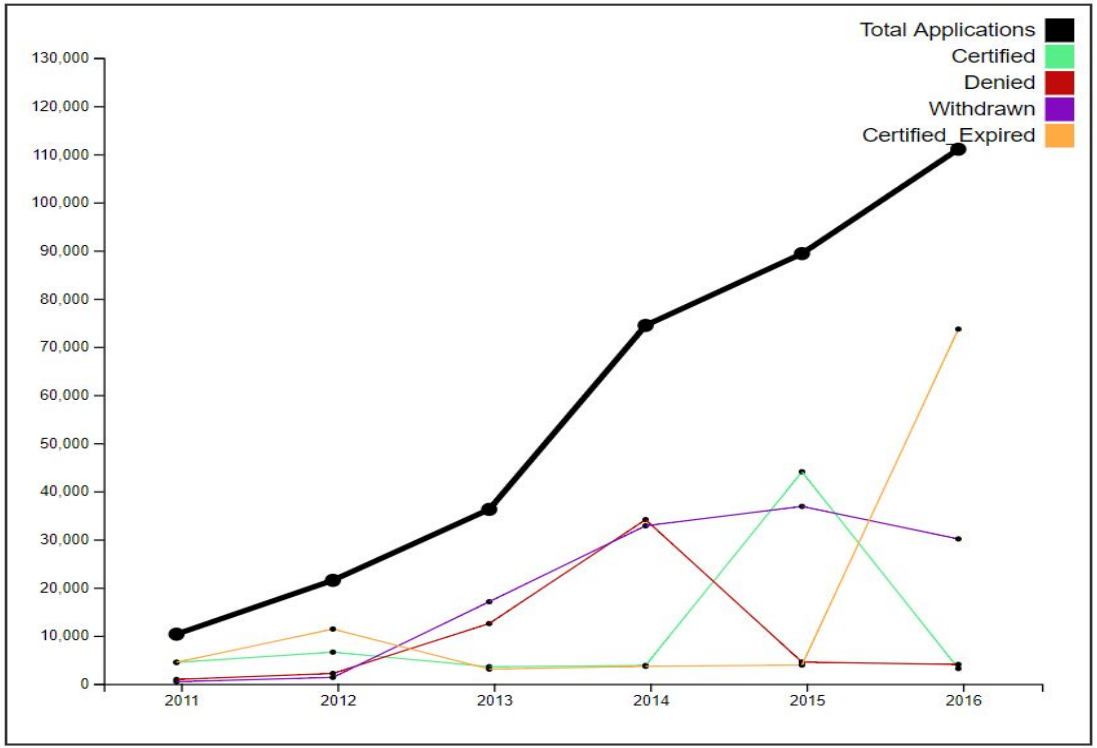
FIGURE 2


State: Utah (UT)
Total Applications: 141
Mean Salary:
USD$ 37738.93

LINE CHART

- On Scrolling Down
- A line chart will generated showing the value of the application count from year 2011-2016.
- In this four different lines will be generated simultaneously each differentiated by a different color.
- Legend is also being implemented to show the name of the lines(Figure 3).

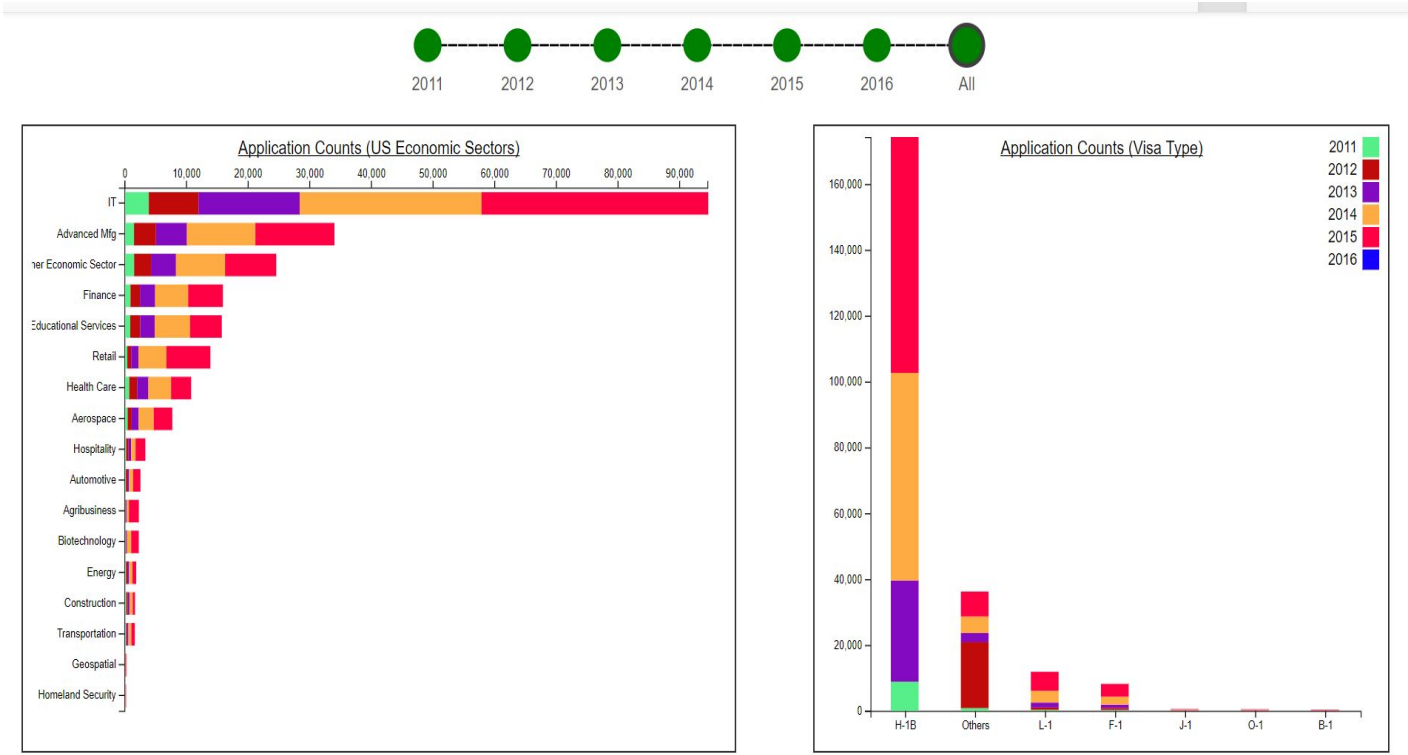
Figure 3


Total Applications
Certified
Denied
Withdrawn
Certified_Expired
130,000
120,000
110,000
100,000
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
2011
2012
2013
2014
2015
2016

STACKED BAR CHART

- On pressing the Button from Yearwise Chart
- When any particular button is being pressed , data for that particular year will be shown with respect to a particular type of case status.
- In this chart, by clicking on “ALL” button, we can access the all years data together.
- Data for different years will be shown in a stacked form having different color for each year.
- In this we have shown data for all years with respect to that Economic Sector in which US Visa Application was being filed.
- Data for different types of Visa Categories is also shown in this chart(Figure 4).

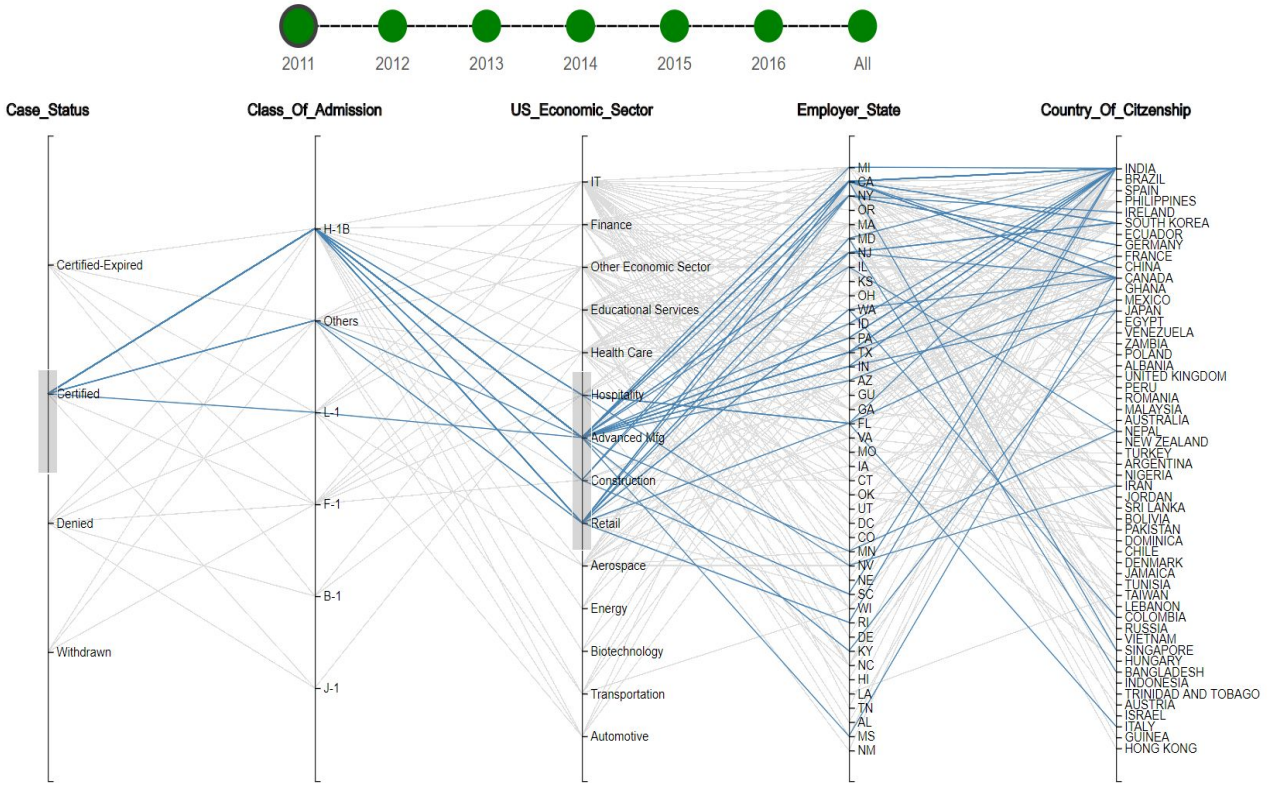
Figure 4


2011
2012
2013
2014
2015
2016
All
Application Counts (US Economic Sectors)
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000
80,000
90,000
IT
Advanced Mfg
ner Economic Sector
Finance
Educational Services
Retail
Health Care
Aerospace
Hospitality
Automotive
Agribusiness
Biotechnology
Energy
Construction
Transportation
Geospatial
Homeland Security
Application Counts (Visa Type)
2011
2012
2013
2014
2015
2016
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
H-1B
Others
L-1
F-1
J-1
O-1
B-1

Parallel Coordinates Chart

- Using the brush to select particular attribute
- Parallel Coordinates Chart is used to show many attributes of a data at one particular time.
- Brush is used to select one particular attribute showing its co-relation between other attributes.
- The connected portion is shown by blue color highlighted focussed lines and the non-connected portion is having gray lines in the background(Figure 5).

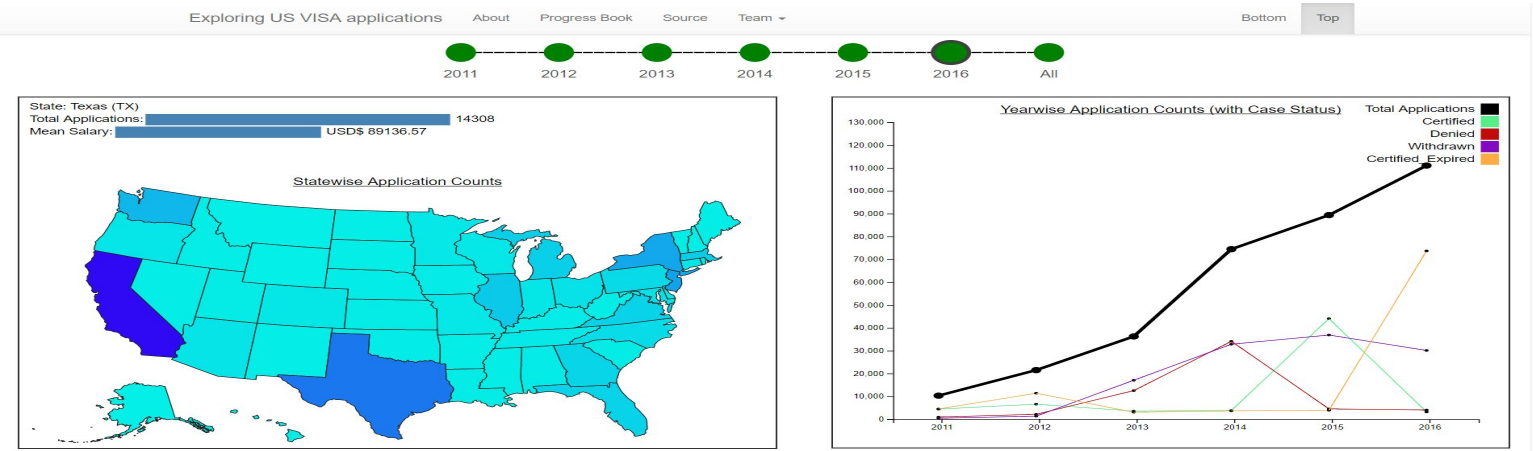
Figure 5


2011
2012
2013
2014
2015
2016
All
Case_Status
Class_Of_Admission
US_Economic_Sector
Employer_State
Country_Of_Citzenship
Certified-Expired
Certified
Denied
Withdrawn
H-1B
Others
L-1
F-1
B-1
J-1
IT
Finance
Other Economic Sector
Educational Services
Health Care
Hospitality
Advanced Mfg
Construction
Retail
Aerospace
Energy
Biotechnology
Transportation
Automotive
MI
CA
NY
OR
MA
MD
NJ
IL
KS
OH
WA
ID
PA
TX
IN
AZ
GU
GA
FL
VA
MO
IA
CT
OK
UT
DC
CO
MN
NV
NE
SC
WI
RI
DE
KY
NC
HI
LA
TN
AL
MS
NM
INDIA
BRAZIL
SPAIN
PHILIPPINES
IRELAND
SOUTH KOREA
ECUADOR
GERMANY
FRANCE
CHINA
CANADA
GHANA
MEXICO
JAPAN
EGYPT
VENEZUELA
ZAMBIA
POLAND
ALBANIA
UNITED KINGDOM
PERU
ROMANIA
MALAYSIA
AUSTRALIA
NEPAL
NEW ZEALAND
TURKEY
ARGENTINA
NIGERIA
IRAN
JORDAN
SRI LANKA
BOLIVIA
PAKISTAN
DOMINICA
CHILE
DENMARK
JAMAICA
TUNISIA
TAIWAN
LEBANON
COLOMBIA
RUSSIA
VIETNAM
SINGAPORE
HUNGARY
BANGLADESH
INDONESIA
TRINIDAD AND TOBAGO
AUSTRIA
ISRAEL
ITALY
GUINEA
HONG KONG

DATA ANALYSIS AND CONCLUSION

The main take away from our visualization projection is the interactivity and visualization of the associated to US VISA applications.

The following is the screenshot of the website of our visualization project



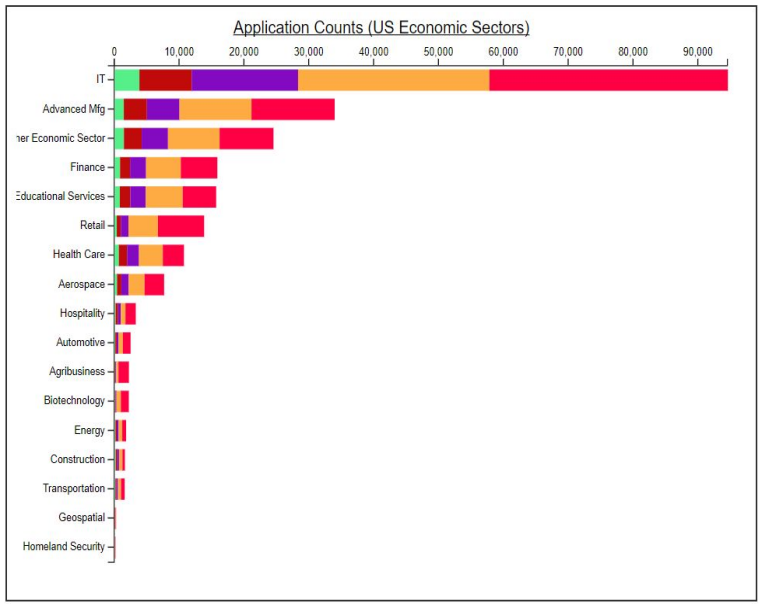
As also mentioned earlier, we wanted to show the main attributes in the project. So, in the Map Chart we showed the state wise application count. The color of the state depends upon the application count as darker the color more is the number of applications in that state.

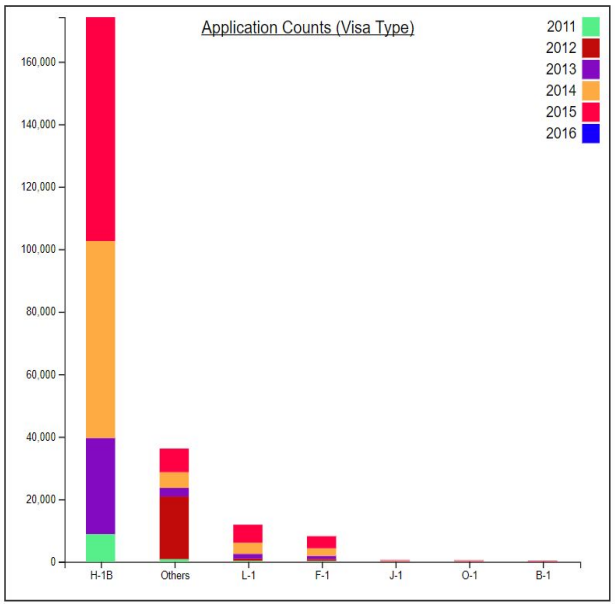
We can change the year of data to be shown by using the filter button from yearwise chart.

The line chart is generated using ScrollyTelling as we scroll down the line chart is being made for years 2011-2015.

The line chart shows the count of applications in each case status of visa and each having a different color line. Data is being shown from 2011-2016


2011
2012
2013
2014
2015
2016
All

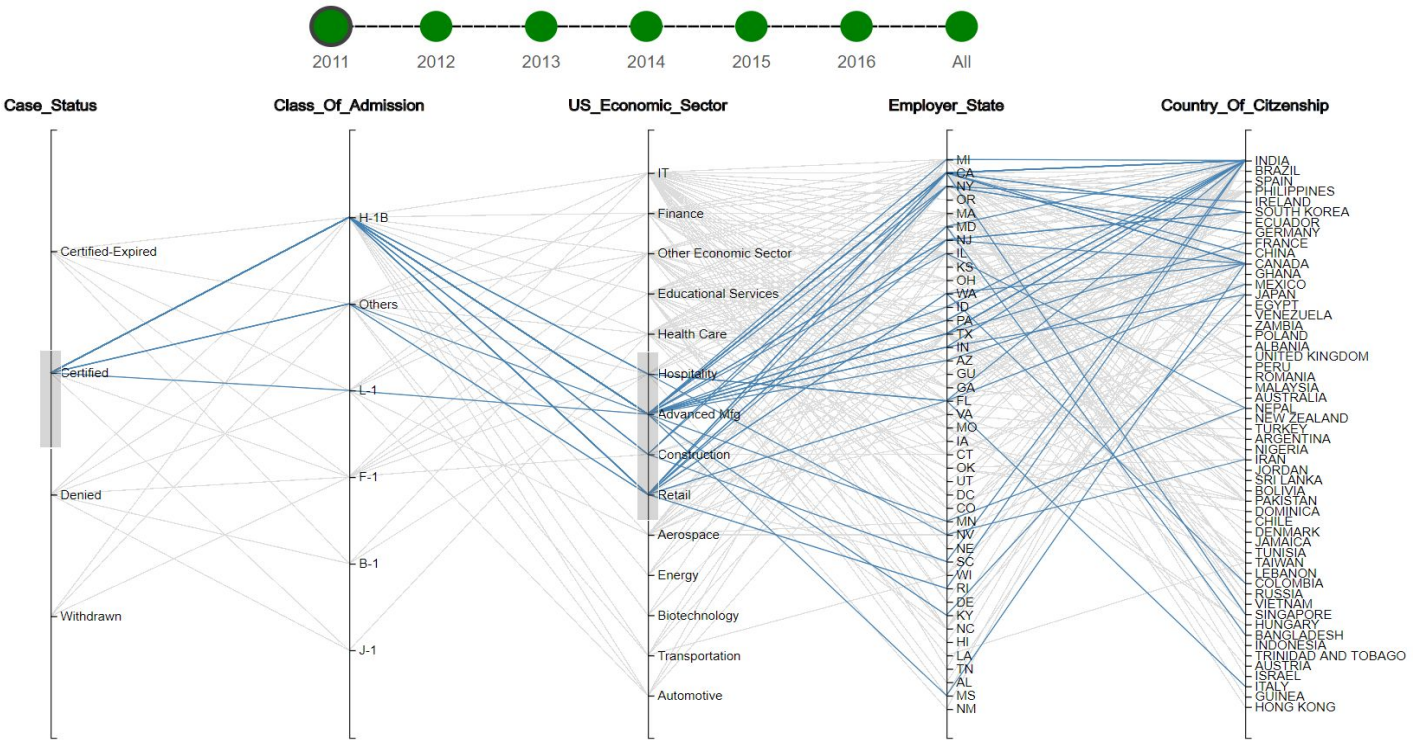

Application Counts (US Economic Sectors)
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000
80,000
90,000
IT
Advanced Mfg
ner Economic Sector
Finance
Educational Services
Retail
Health Care
Aerospace
Hospitality
Automotive
Agribusiness
Biotechnology
Energy
Construction
Transportation
Geospatial
Homeland Security


Application Counts (Visa Type)
2011
2012
2013
2014
2015
2016
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
H-1B
Others
L-1
F-1
J-1
O-1
B-1

As seen in the previous page figure , Stacked Bar Chart tells us about the number of applications in that Economic Sector for that particular year. Apart from this, the applications are stacked by case status so as to make it look more understandable to the reader. The other Stacked Bar Chart is telling the count of the applications for a particular type of visa in that year.

On hovering any stacked portion, we can see the tooltip showing the count of the applications. This comes in handy to know the exact value.

Lastly, the Parallel Coordinates Chart shows the data for many attributes at one time shown(next page) . With the help of brush, we can see only those values which we want. The color of those lines gets changed to blue which are in focus and rest remains in gray in background.We can focus only on those attribute which we want to show.


2011
2012
2013
2014
2015
2016
All
Case_Status
Class_Of_Admission
US_Economic_Sector
Employer_State
Country_Of_Citzenship
Certified-Expired
Certified
Denied
Withdrawn
H-1B
Others
L-1
F-1
B-1
J-1
IT
Finance
Other Economic Sector
Educational Services
Health Care
Hospitality
Advanced Mfg
Construction
Retail
Aerospace
Energy
Biotechnology
Transportation
Automotive
MI
CA
NY
OR
MA
MD
NJ
IL
KS
OH
WA
ID
PA
TX
IN
AZ
GU
GA
FL
VA
MO
IA
CT
OK
UT
DC
CO
MN
NV
NE
SC
WI
RI
DE
KY
NC
HI
LA
TN
AL
MS
NM
INDIA
BRAZIL
SPAIN
PHILIPPINES
IRELAND
SOUTH KOREA
ECUADOR
GERMANY
FRANCE
CHINA
CANADA
GHANA
MEXICO
JAPAN
EGYPT
VENEZUELA
ZAMBIA
POLAND
ALBANIA
UNITED KINGDOM
PERU
ROMANIA
MALAYSIA
AUSTRALIA
NEPAL
NEW ZEALAND
TURKEY
ARGENTINA
NIGERIA
IRAN
JORDAN
SRI LANKA
BOLIVIA
PAKISTAN
DOMINICA
CHILE
DENMARK
JAMAICA
TUNISIA
TAIWAN
LEBANON
COLOMBIA
RUSSIA
VIETNAM
SINGAPORE
HUNGARY
BANGLADESH
INDONESIA
TRINIDAD AND TOBAGO
AUSTRIA
ISRAEL
ITALY
GUINEA
HONG KONG

From our visualization project, we feel that we are successful in answering all the Questions mentioned the process book. There is scope for adding more components into this visualization. In the future, we would like to show more data about the Applicant's country of origin as in from which country the maximum people are coming using interactive World Map. Based on these features which we have used in our project and combining it with other few imperative features out there in the market we can use the Machine Learning Algorithms so as to classify which Category of Visa is applied maximum by Immigrants to US. Plus, we can have useful insight about Countries from which maximum people are applying for US Visas.

REFERENCES

MAP CHART- http://bl.ocks.org/michellechandra/0b2ce4923dc9b5809922

PARALLEL COORDINATE CHART-https://bl.ocks.org/jasondavies/1341281

SCROLLYTELLING- http://vallandingham.me/scroll_talk/examples/

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