Process Book

Wages and Flows

Samik Adihkari

Nicholas Stellitano

CS 171
May 06th 2015

Table of Contents

Overview and Motivation

Standard Metrics

Migration is an actively studied phenomenon within the international development community. Typically, measurement is quite difficult. Census data within wealthy countries (OECD countries) is challenging to come by and nearly impossible to get in developing countries which have very low government capacity. When the numbers are available migration flows, stocks (by gender and by education level), along with remittances are measured. The general theory is that migration allows individuals to earn money to send back to their home country, which can then help improve the lives of the migrant's family members. Other metrics and considerations that are taken into account are regional flows, laws that dictate when and who can migrate into various countries, along with migration patterns.

What's should an OECD Country to do?

So what should an OECD country do to help developing countries increase wealth per capita, improve economic conditions, and decrease poverty? Some theorists argue that globalization can help through increased trade flows, industrialization, and developing young economies. Others see Aid as another avenue to distribute wealth into countries that are dealing with extreme forms of poverty and unrest. Other initiatives involve the creation of major international development banks to try and increase economic development through various financial devices like loans, grants, policy dialogues, technical assistance and equity investments.

Migration and Wage

However, one interesting effect OECD countries have that is misunderstood and overlooked is Migration. One recent theory proposed is that OECD countries can actually have an outsized impact on developing countries through migration policies by opening their borders more. Through the inherent wage differential, migrants of similar skill level can earn orders of magnitude higher wage when moving from a non-OECD to a OECD country.

By analyzing a selected grouping of Non OECD countries and their relationship with OECD countries we hope to be able to show the overall affect migration can have over local

economies. In particular, we are interested in comparing this benefit to migrants with other standard forms of metrics such as remittances and aid.

Related Work

Previous Literature
One large inspiration for this project was the research conducted by Lant Pritchett, Proffesor of Intternational Development at the Harvard Kennedy School. In his book "*Let Their People Come: Breaking the Gridlock on Global Labor Mobility*." Prof Pritchett "cites estimates that if rich countries were to permit a mere 3 percent increase in the size of their labor force by easing restrictions on labor mobility, the benefits to citizens of poor countries would be $305 billion a year--almost twice the combined annual benefits of full trade liberalization ($86 billion); foreign aid ($70 billion) and debt relief (about $3 billion in annual debt service savings)"[1].

Other literature can be found at the Center for Global Development's Migration and Development Initiatives website (http://www.cgdev.org/section/initiatives/_active/migrationanddevelopment)

Questions

How does migration compare to other bilateral indicators?
Globalization is on the rise in the 21st century and labor mobility is a big part of the process. We are interested in comparing the economic benefits people gain when moving from a "developing" to a "developed" country.

In particular, we want to compare the average wages that

---

[1] http://www.cgdev.org/article/migration-and-development-temporary-workers-are-key

people gain from moving into a developed country with other bilateral indicators such as remittances that people send back home, foreign assistance and aid that flows from developed to developing countries and so forth.

Challenges

The biggest challenge with answering this question comes in the form of what data do we have available. Our initial analysis began with comparing just average wages across OECD countries and non-OECD countries but came to appreciate that this rough average did not provide enough of a nuanced view of the type of migrants coming from one country to another. In this vein, we chose to breakout Migrant Stocks/Flows into three education Levels: Level I (less then secondary school education), Level II (secondary school education), Level III (post secondary education). Looking at Migration through the lens of education brings up a few interesting questions:

1. Do different OECD countries tailor to their Migration policies to Migrants of specific education levels?   (It turns out they do: reference Migration Flows by OECD)
2. Do certain OECD countries provide better average wages at particular education levels? (It turns out they do: reference Extra Average Wages Earned)
3. Though the United States has the highest number of total migrants among all OECD countries, are particular countries better in per capita terms?

Data

Data Source:

The required data to build our project consisted of obtaining average wage levels for OECD and non-OECD countries by education level, bi-lateral Migration Stocks from OECD to non-OECD countries broken out across gender, education level, and years (1980-2010), bi-lateral AID and Remittance data, along with TopoJSON data for each country. The following are the data sets utilized:

- Wages data for 33 countries (24 non-OECD and 8 OECD, by education levels): Claudio Montengero, World Bank
- Bilateral migration data by education/gender from 1980-2000: IAB brain drain data (http://www.iab.de/en/daten/iab-brain-drain-data.aspx)

- Bilateral Remittance data for 2010: World Bank (http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTDECPROSPECTS/0,,contentMDK:22759429~pagePK:64165401~piPK:64165026~theSitePK:476883,00.html)
- Bilateral Aid data for 2010: AidData (http://aiddata.org/aiddata-research-releases)
- World 110m TopoJSON data: (http://techslides.com/countries-and-capitals-with-d3-and-natural-earth)

Data Processing:

The Data processing stage involved 4 stages:

1. Cleaning Wage Data:

2. Compiling Map data: to build an interactive map that displays each country with applicable information typically requires either the use of GeoJSON or TopoJSON data. There are many sources of information but we chose to utilize the World 110m TopoJSON data built and provided at techslides.com. The data provided the information on each country and capital to include population, latitude, longitude, etc. Once the data was obtained, building the interactive map is fairly straightforward as explained and demonstrated by Mike Bostock (http://bost.ocks.org/mike/map/).

3. To implement the Wage Comparison visualization we needed to compile and merge data from 4 different sources (Aid, Remittance, Migrant Flow, Wage data). Each data set was in a csv format so we chose to read in the data and build a json file with a hierarchical structure with OECD countries being on the top most level. Subsequent levels included non-OECD countries along with migrant stocks and wages broken down by Education Level and Gender.

    The first of many data design issues came to be with trying to decide what structure we need or wanted our data to be. In the end, we chose to follow a parent, child structure (as detailed in Mike Bostock's example of a collapsible forced layout http://bl.ocks.org/mbostock/1062288) because our initial plan was to utilize the data in some type of tree diagram. In the end, this did not come to be but the overall hierarchical structure proved effective.

4. Finally, separate JSON files needed to be constructed to create the Sankey diagrams that detail the migrant stocks of each country across gender and

education. The particular format required for this diagram involved the
following format:

```
5.  var sankey_fill = {
      "nodes":[
        {"id": 1, "name":"Male"},
        {"id": 2,"name":"Female"},
        {"id": 3,"name":"Ed LVL I"},
        {"id": 4,"name":"Ed LVL II"},
        {"id": 5,"name":"Ed LVL III"},
        {"id": 6,"name": "OECD"}
      ],
      "links":[
        {"id": 1,"source":"Male","target":"Ed LVL I","value":11395189},
        {"id": 2,"source":"Male","target":"Ed LVL II","value":10667654},
        {"id": 3,"source":"Male","target":"Ed LVL III","value":14194359},
        {"id": 4,"source":"Female","target":"Ed LVL I","value":11890368},
        {"id": 5,"source":"Female","target":"Ed LVL II","value":10047154},
        {"id": 6,"source":"Female","target":"Ed LVL III","value":15017105},
        {"id": 7,"source":"Ed LVL I","target":"OECD","value":23285557},
        {"id": 8,"source":"Ed LVL II","target":"OECD","value":20714808},
        {"id": 9,"source":"Ed LVL III","target":"OECD","value":29211464}
      ]}
```
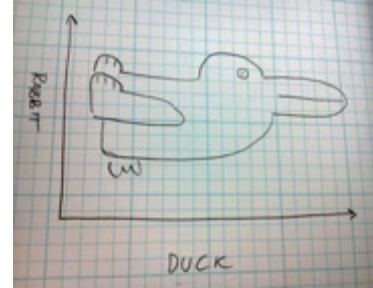
This data set required only minimal data wrangling from the previous
hierarchical JSON files created for the previous section.

Exploratory Data Analysis

Visualizing migration flows is not new. In terms of prior examples that we referenced to get an initial idea of what current visualizations are published and how we could improve upon them we looked a various resources:

1. Center for Data Innovation: (http://www.global-migration.info/)
2. Previous CS171 Project: https://ba66f0012633a2840e66efc393df816fc324c2f4.googledrive.com/host/0B6YljmV-VTJZcElIUWlVZGIzTmM/index.html
3. Facebook: http://www.citylab.com/tech/2013/12/facebook-using-your-profile-track-global-urban-migration-trends/7982/

The challenge with migration data is the desire to want to tie it to a global map. Individuals can visually recognize parts of the world or in the case of Center for Data Innovation, represent it through country names within an enclosed space. Each of these types of visualizations runs the risk of being too cluttered or too broad to where specific information is hard to pull out.

Design Evolution

Form the outset we developed our visualization in terms of changing how users analyze the typical migration data. By allowing users to analyze both existing data along with additional data provided by us, we want the ability to search and compare across countries, genders and education levels.

Our visualization utilizes various marks and channels to help convey the information. In some instances, the information being conveyed is the same but it is important to be able to analyze data through different lenses. In this case,
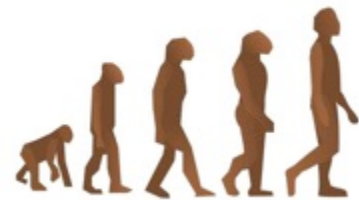
1. Points
2. Lines
3. Areas

Migration across Channels:
1. Shape
2. Color
3. Spatial Position

Various Visualizations:

Our desire was to utilize the map as a starting point to provide a broad overview for the user but then utilize various other visualizations that leveraged different channels to help the user understand the data:

1. GeoJSON Map: Basic starting point providing a frame of reference along with other basic information like wage or migration stock conveyed through color.

2. Scatter Plot: Identification of trends across various countries along with being able to select one or more countries was an important aspect to our visualization. In the context of this type of information you don't always know what countries you want to aggregate or compare. A scatter plot provided two important functions. The first was a way to show migration levels across education levels and countries. Secondly, the scatter plot provide another way for a user to either select one country of interest or aggregate multiple countries together.

3. Bar charts x 2: The utilization of bar charts was important because we need to compare various attributes of a particular element. In this case the element was either OECD or non-OECD country and the attributes ranged from Aid, Remittance, and Wage differential data. The bars could be interconnected and organized in such a way that one bar chart displayed data from the perspective of the 8 OECD countries and the other displayed data from the perspective of the non-OECD. This helped look at the problem but from two different view points.

4. Sankey Diagram: Finally, the use of the Sankey Diagram was a take on displaying migration flows but through a different medium. The advantage of the Sankey

diagram was that it displayed the data in a more visually appealing and understandable format while also conveying more information to include how the flows interconnect across various characteristics (gender, education level etc) along with time.

Unforeseen Problems:

As with most projects, identifying and wrangling the required data proved to be quite challenging. Ensuring the data was comparable across years along with identifying accurate average wages across countries and education levels is still on going. The data we presented is a small subset of a potential larger data set but further time and effort will be required to collect and organize the data.
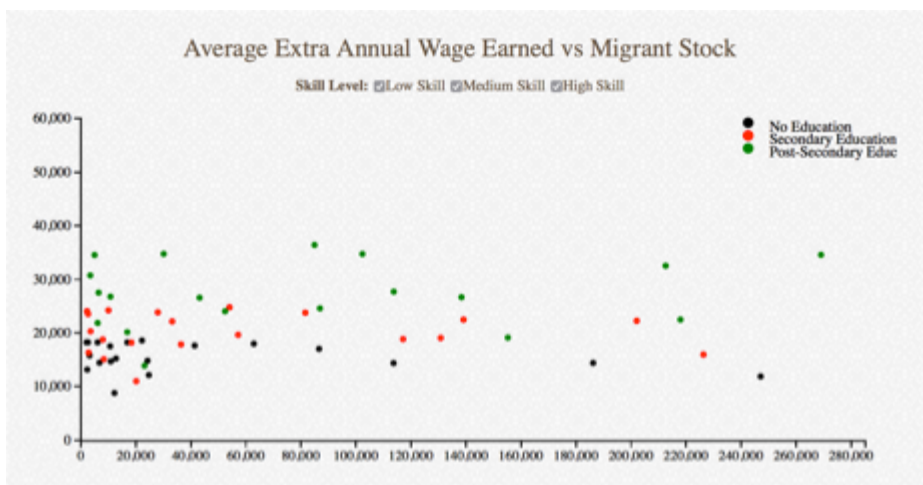
Implementation

Wage Comparison:

The first aspect of our visualization involved comparing average wages across countries. To do this we wanted the user to be able to visually see and compare multiple values from the country in question and from the perspective of the OECD countries. To do this we linked 4 visualizations together. The map of the globe is a heatmap displaying various levels of average wages across countries. This map is then connected with 2 bar charts and a scatter plot.
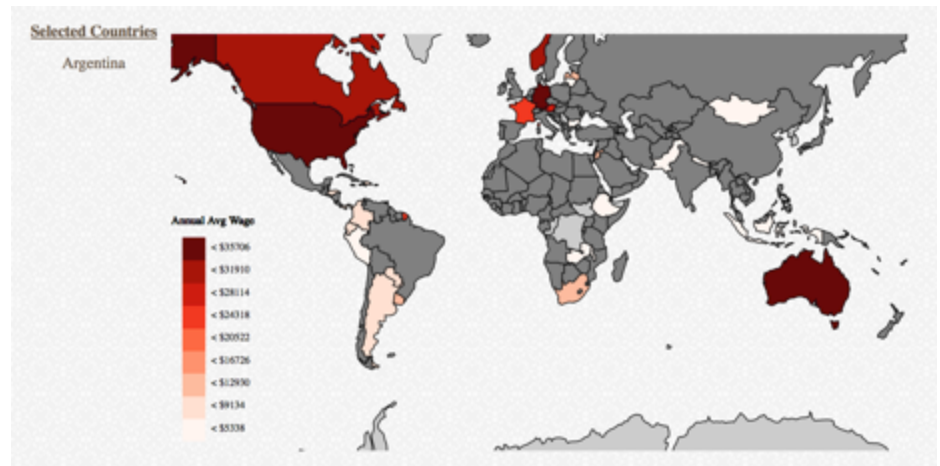
The scatter plot displays the average wage differentials for each non-OECD country across the three education levels relative to the total migrant stock from that country to a OECD country in the year 2010. The three education levels are three separate colors and each country can be identified by placing the cursor over the circle in question.

The scatter plot, like the map, can act as a place to control other visualizations through the selection of an individual country or through the brushing of multiple countries. The selection or brushing will only display the selected list of countries in the other charts.

The two bar charts tell the story from the perspective of non-OECD and OECD countries respectively. In the first chart, each blue bar to shows the total wage differential (average wage differential X total migrant stock) among migrants from a particular non-OECD country to the average wages in the OECD country.

Because total remittances and aid flows are orders of magnitude lower in comparison to wage differentials, we have implemented a slider with a log scale that allows you to see the other two metrics of comparison.

Average wage differentials in comparison to Remittances/Aid

Extra Annual Wage Earned
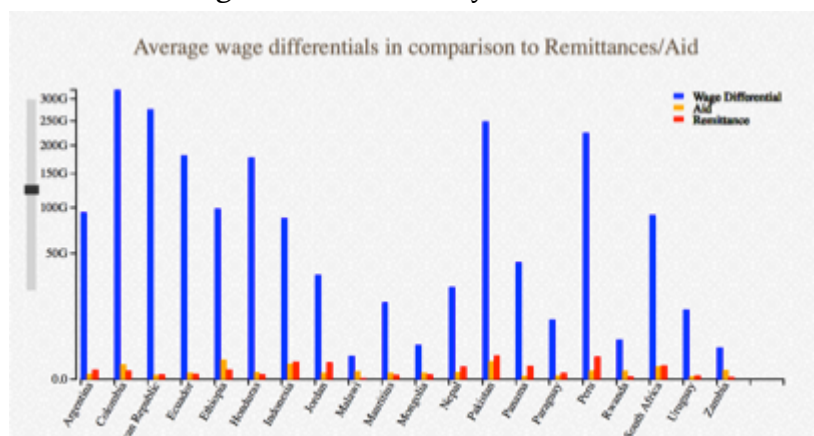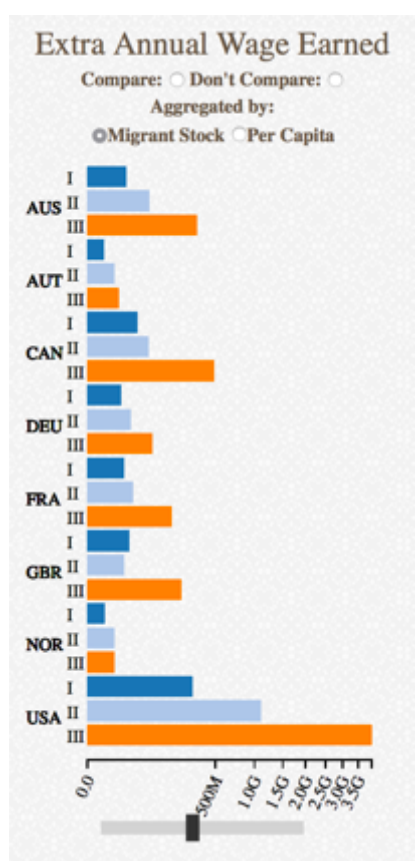
In the second chart, the story is told through the perspective of the OECD country. Depending on a single or a group of non-OECD countries selected, the Extra Annual Wage Chart shows estimated gains in terms of wage differentials for migrants of non-OECD countries. The default layout shows this difference by education levels with radio buttons that allow comparison on a per capita basis as well as an option to compare wage differentials to aid and remittances provided to non-OECD countries.

Migration Flows:

To complete our visualization picture, we tell the story of how migration has evolved over the last 35 years.

Utilizing data from 1980-2015 over a five year interval, we show the flow of male and female migrants from 195 non-OECD countries to 20 OECD countries by gender and education level.

By clicking on the "By OECD" radio button, we can see flows into each individual OECD country.

For example, we can see that in 1980, Canada allowed equal proportions of high-skilled and low-skilled migrants into the country.

Fast forward to 2010, and it is selectively letting in many more high-skilled than low-skilled migrants.
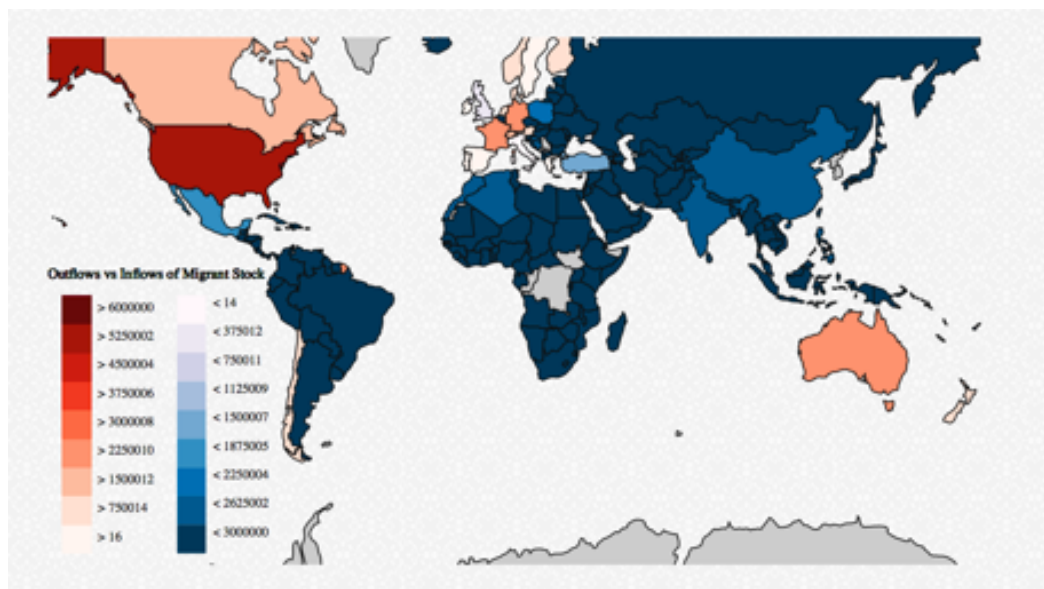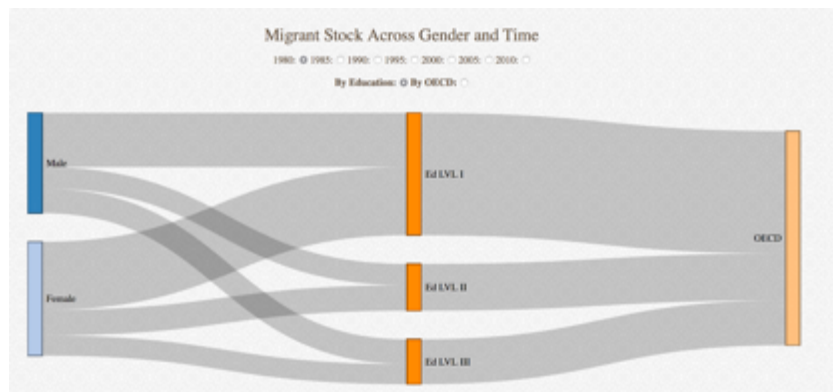
Evaluation

Conclusion:

We were able to tell a comprehensive story regarding the effect that migration to OECD countries has on developing countries through comparison of different measurement yardsticks. In particular, we compared average gain in wage differentials to remittances and aid and found that total wage differentials were orders of magnitudes higher than the later two. We were able to interactively show this effect on a single country, a group of countries, or the entire list of 28 countries in our database. The Sankey diagram helped us dynamically visualize migration from all 195 non-OECD countries to all 20 OECD countries by gender and education level. This helped us understand interesting patterns regarding migration inflows to OECD countries over time from 1980 to 2010.

Functionality and Future Improvements:

There were significant challenges throughout the implementation of the project. Although we addressed most of those challenges, we were still short of addressing some issues which we highlight below.

1. Further distinction for MAP Colors

    Data limitations were a big factor in trying to colorize the map coherently and tell a comprehensive story. A heatmap works best when the distinctions based on colors are clearly demarcated and due to large chunks of missing countries, our heatmap layout for the wage differential map did not turn out to be what we expected at the beginning of the project. Having said that, this has quite a simple fix and it requires collection of data for many more countries.

2. Add more data

    Users trying to replicate this work in the future will have questions about the feasibility of bar charts to display wage differentials of non-OECD countries, especially when trying to fit all 195 OECD countries in the chart. This has quite a simple solution and requires aggregation of countries by continents, income levels, or

population metrics. Similar solutions can be implemented for bar chart of OECD countries and scatter plots.

3. PPP

Economically speaking, our visualization assumes that migrants can make as much as local citizens of OECD countries if they are of the same education level. This is a strong assumption, particularly with low-skilled migrants. While it can be argued that OECD countries have minimum wage standards irrespective of the citizenship status of the workers, a stronger argument would be to convert all wages at PPP (Purchasing Power Parity) levels to allow for better comparison. Time constraints prohibited us from making this adjustment and we leave this issue as a caveat for users of this visualization platform.

Appendix

Initial Project Proposal:

- **Background and Motivation:** As two students of the MPAID program we are very interested in the role Migration plays in International Development. One of our Program Directors is interested in visualizing the overall wage increment of migrants from Non-OECD countries to OECD countries. We hope to visually depict not only the Bilateral flows of Migration and Remittance but the overall changes in Wages by education level and gender when they move from developing countries to non developing countries. Also, showing these in comparison to more standard metrics like Aid and Trade will be beneficial.

- **Project Objectives.** What is the overall increment on wages of migrants of certain skill levels when they move from a developing country to a developed country? How does that increment compare to overall trade flows, remittances, and aid flows.

- **Data.** Our data will be collected in three forms
    - **(Wage/Population/Aid data):** World Development Indicators (http://databank.worldbank.org/data/views/variableSelection/selectvariables.aspx?source=world-development-indicators)
    - **(Migration data by education/gender):** IAB brain drain data (http://www.iab.de/en/daten/iab-brain-drain-data.aspx#Sources)
    - **(Remittance data):** World Bank (http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTDECPROSPECTS/0,,contentMDK:22759429~pagePK:64165401~piPK:64165026~theSitePK:476883,00.html)

- **Data Processing.** We expect to construct some indicators through the aggregation of multiple data sets. We expect this to be somewhat time consuming but doable.

- **Visualization.** Our basic design incorporates two overall products (more if time permits). The idea is to start from a global perspective (product 1) and refine to smaller levels of data by selecting on various types of countries (product 2)
    - The first would be utilizing a map of the world and creating a heat map of migration related indicators. Some of the indicators that are available or could be constructed from the data included are migration flows to and from OECD countries, average wages gained by migrants by moving to a OECD country, remittances, aid flows, and trade flows. The preliminary idea is to use Radio Buttons to switch across different visualizations utilizing the concepts learnt in HW2.
    - The second product utilizes concepts seen in HW3. Along with the world map, we will have various migration related layouts, which can interact with one another through use of sliders, brushes, and point-clickers. The designated task would be quantitative in nature. By selecting either an OECD country or a Non OECD country, various graphs would change to map the top 5 selected "partner" countries along with various characteristics to include overall remittance, migrant stocks, aid, trade, and wage differential.
    - In essence, the idea is to create something similar to one of the previous projects (https://ba66f0012633a2840e66efc393df816fc324c2f4.googledrive.com/host/0B6YljmV-VTJZcElIUWlVZGIzTmM/index.html) but with focus on migration rather than aid:

- **Must-Have Features.** Product 1 (Heat Map), Product 2 (interconnected display of country data)

- **Optional Features.** Extending further, it would be cool to visualize "flows" of migrants or remittance through visual mappings and connected lines. Example: (http://www.citylab.com/tech/2013/12/facebook-using-your-profile-track-global-urban-migration-trends/7982/)

- **Project Schedule.** Our overall timeline is as follows:
    - **Week 0:**

- Setup Github
- Develop Workflow
- Submit Proposal
- Develop Project components
  - **Week 1 (6th-10th):**
    - Create small workable dataset
    - Create Index File
    - Build out Map
    - Identify Graphs to be utilized
  - **Week 2 (13th – 17th):**
    - Complete Data Set
    - Complete Heat MAP
    - Divide up objects (graphs) to be constructed
  - **Week 3 (20th – 24th):**
    - Data crunch required ratios and numbers
    - Complete Object Graphs and interconnect
    - Develop pictures of regional or group flow patterns?
  - **Week 4 (27th- 31st):**
    - Work on Aesthetics of the website (colors, displays, etc…)
    - Loud on Github, test
  - **Week 5 (4th – 7th):**
    - Create Video

Initial Feedback: April 21st 2015

Design Studio Feedback

Group Partner: Michael (MIT)

Feedback 1:

We should use legends and have a really good narrative for the story we want to tell with our visualization.

Comment: Both Michael and us were working with maps. He was working with transportation data from Massachusetts. We are working with the map of the world. And he thought it was more challenging to get the attributes for all countries correct, especially given that both our teams were trying to vary the map by more than one indicator. So, I think this feedback was helpful in making the visualization more compact.

For the narrative part, Michael advised us to use texts and brief paragraphs, either accompanying the maps in the layout or having them pop up as users click different features on the map. We will take that into consideration.

Feedback 2:

Michael felt we had too many graphs in the same layout and he was struggling with the same problem.

Comment: The tradeoff here is telling a comprehensive story vs. having a compelling visual layout. We want to balance the two. We definitely do no want to cram everything together but we do want to say a coherent story given our hypothesis on migration. One optional feature we would like to have is a different layout (from our main layout), where we show animations because we have data that varies by time and we are not using it in our main feature.

Feedback 3:

Michael advised us to use pull information from different smaller json files as opposed to having one large json file.

Comment: The tradeoff here is between having several tractable json files and having one large file that could be used for multiple interactive graphs. Interactivity between several plots and the map is a must for the kind of stories we want to tell and we decided to have one large json file based on which we will create most of our layouts.