Behind the Scenes of “World Bank Lending over Time”

By Vaidehi Dalmia (vd2302) 5/4/2018

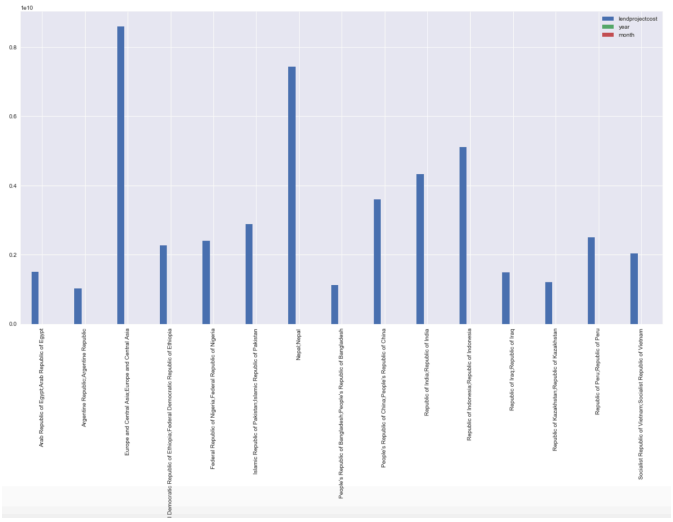
MY IDEA

For my final project I wanted to explore the dataset provided by the World Bank on their lending projects from 1947 to the present. The dataset can be found at: https://datacatalog.worldbank.org/dataset/world-bank-projects-operations. I was interested in seeing whether the regions the world bank focuses on have changed over time.

DATA EXPLORATION

My first step was to understand my data a little better. I used Jupyter notebook to do this. When looking at my dataset, I realized that the dataset had more than 50 columns, however only few of the columns seemed worth investigating as the other columns had a lot of NaNs. I created a subset of the dataset (Image below) with variables such as regionname, countryname, boardapprovaldate, sector, mjsector, lendinginstr.

I then started to dig deeper into each variable by counting the various values and plotting them, hoping to find some interesting patterns. Though I was doing a time series analysis, I found it difficult to aggregate the data based on year as well as region or country. Because of this, I found it hard to find an interesting story and visualize the trends that were there. Never the less, I did some basic analysis.

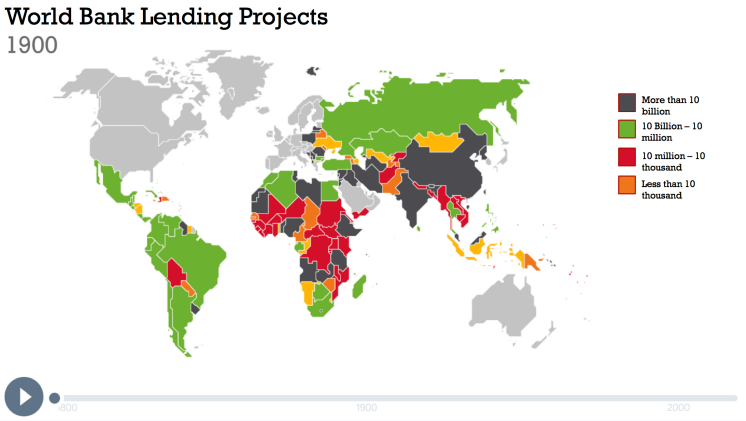


The graph to the left plots the lending costs for countries in 2016 where the cost is above 1B. I initially hypothesized that the lending costs would be correlated with different natural disasters that occurred such as the spike below for Nepal could be due to the earthquake in 2016. However, after looking into the project names further, I realized that this was not always the case and natural disasters is only a small subset of the reason why world bank lends money to different countries. I could have looked at sectors or themes or project names to see the types of projects the world bank was lending to in different companies. However, I discarded the idea as the unique values for these columns were more than 200 and it would have been very time consuming to create broader categories.

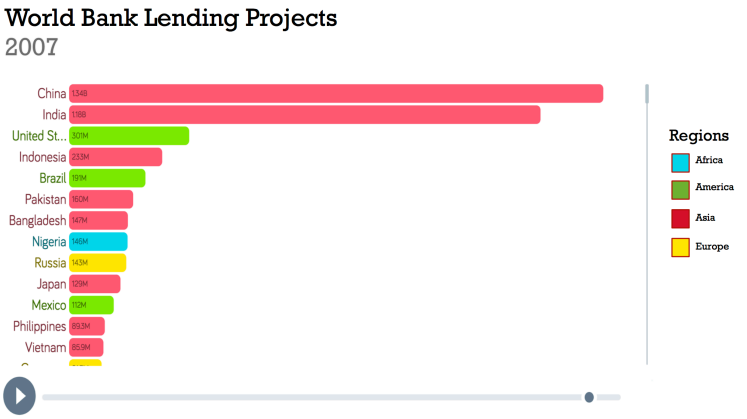
I decided that I wanted to show a bigger picture and wanted to let the user find the trends themselves. I sketched out a few ideas.

INITIAL SKETCHES

My first idea was a heat map with a timeline. The colors would change as the user scrolled through the different years.



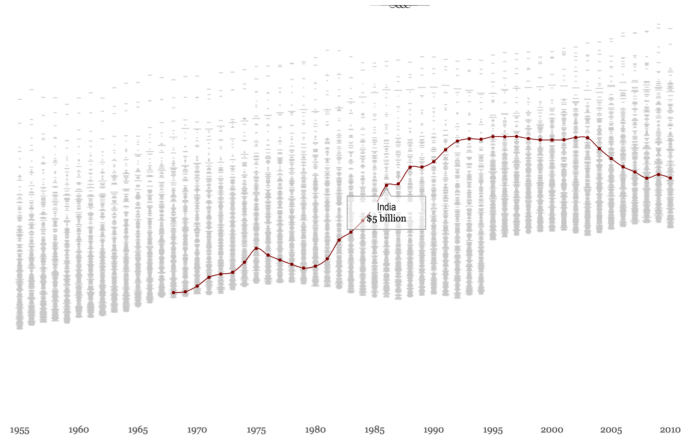
My second idea was different bar graphs colored by region and in descending order of lending cost. The bars would change when the user scrolled through the different years.



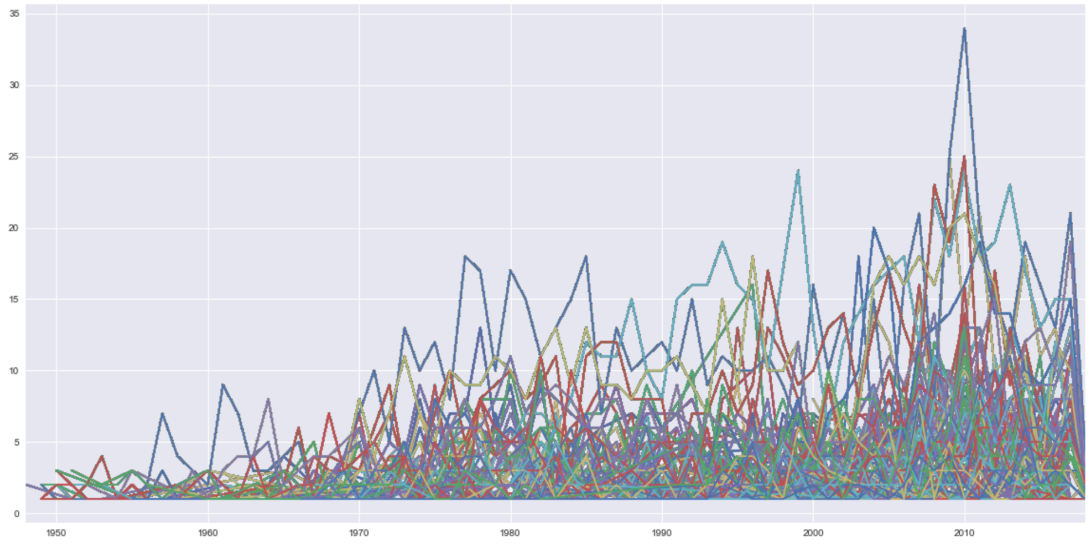
Inspired from https://www.gapminder.org/

My third idea was inspired by a data visualization on the fortune 500s by Fathom.

Inspired from https://fathom.info/fortune500/



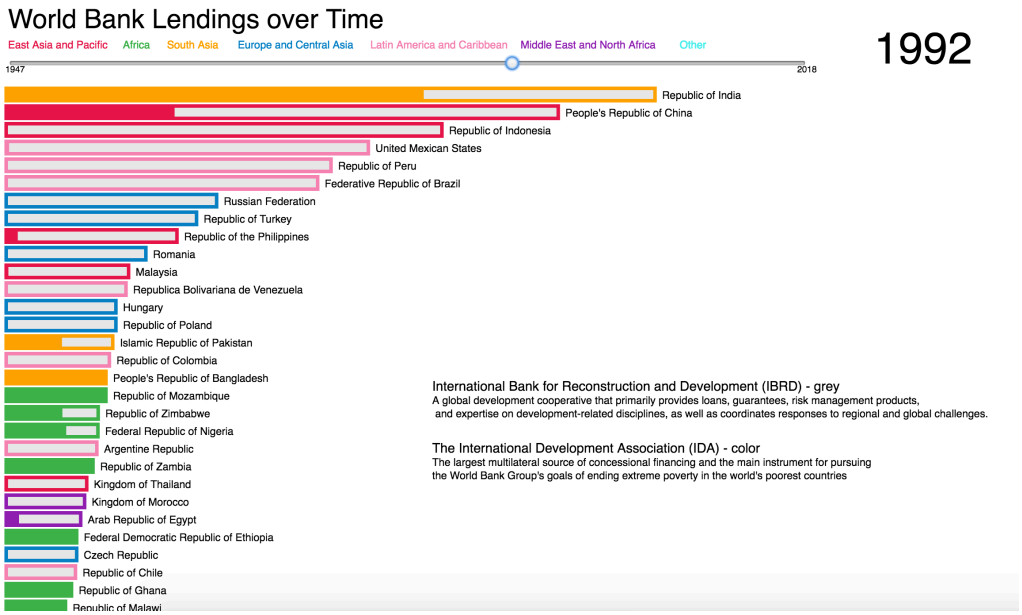
I initially started out by exploring the third idea as I thought it was the most out of the box. However after trying to plot the different countries on the map, the visualization looked messy and was very difficult to understand.



FINAL VISUALIZATION

I thought the bars were more intuitive and easier to compare so I chose my second idea. This way I could capture a lot of data in a simplistic manner. I was still not satisfied with my visualization and was looking to tell a better story. I went back to the data and found three columns I had missed earlier: ibrdcommamt, idacommamt, and grantamt. I was most interested to see the percentage of the total amount that was ida and grant amount. During my research I had found that the World Bank has two types of lendings. The IBRD lendings are loans provided to countries where as IDA lending are grants provided to the world's poorest countries. This seemed like an interesting story to tell!

With this in mind, I decided to split the bar graph into IBRD and IDA lendings. I was struggling to keep the region colors for the graph and simultaneously pick two different colors for IBRD and IDA. I played around with a few different ideas, one of them included coloring the name of the country based on the region and color ibrd and ida neutral shades of grey. Because most of the colors on the page then turned out to be grey, the page looked very ugly. The grey also took the focus away from the region colors. I wanted to focus more on the IDA spendings so I decided to color the bar with the region color for the IDA spending and grey out the spending for IBRD. I have the bar a think border based on the region color so that the user can tell the region even when the IDA spending is 0.



When I presented my visualization in class, I got feedback that it was hard to follow the moving bars to see the change in IDA. Hence, I decided to create a second visualization focusing on all the countries that were initially in IDA status but have graduated from it.This visualization is similar to a line graph and the user can easily see the change in IDA spendings over time for a particular country.

Below, is an image of the IBRD and IDA lendings for Indonesia over time.

I was pretty time constrained for my second visualization. If I had more time I would improve on the aesthetics of that visualization. I would also add reasons why a country might have relapsed or gotten IDA lendings even though they have graduated. I would also better integrate both my visualizations and create a story that flows better.