Harvard Extension School
Biz Viz – Business environment across globe

Fraser Macdonald, Nadiia Novik & Prabhu Akula

Process book

Abstract and Objective

What makes a country "good for business"? The World Bank has been collecting statistics on various indicators of a country's receptiveness to small and medium businesses since 2005. We intend to create a tool for any viewer to investigate how various attributes of a country relate to the state of its commerce.

Business friendliness
User to explore questions about the "ease of doing business" in a country
Explore important indicators which impacts country's growth
How transparency is influencing business friendliness ranking
How various attributes of a country relate to the state of its commerce?

Questions we asked

- What countries are 'good for business'? What countries are not?
- Is there any link between geographic region and ease of doing business?
- Most importantly, however, we want to answer questions of how "ease of doing business" is related to other areas, such as:
    - How is "ease of doing business" related to …
        - Migration?
        - Corruption?
        - Expected length of life?

Data collection

Ease of ***Doing Business***

- Score of ease of business environment from 0 (difficult) to 100 (easy)
- 11 indicators, including “starting a business”, “paying taxes”, “enforcing contracts”
- 189 economies
- 10 years of data
- Source: doingbusiness.org

Net migration

- Overall international net migration per country (all groups: refugees, workforce, students, etc.)
- Average for 5-years periods
- Positive and negative absolute values
- More details available for refugees (destination and origin), personal remittances (received and paid)
- Source: wdi.worldbank.org

Corruption perception index

- Score of corruption perceptions from 0 (very corrupt) to 100 (very clean)
- Data on 168 economies
- Multiple years of data available
- Source: transparency.org

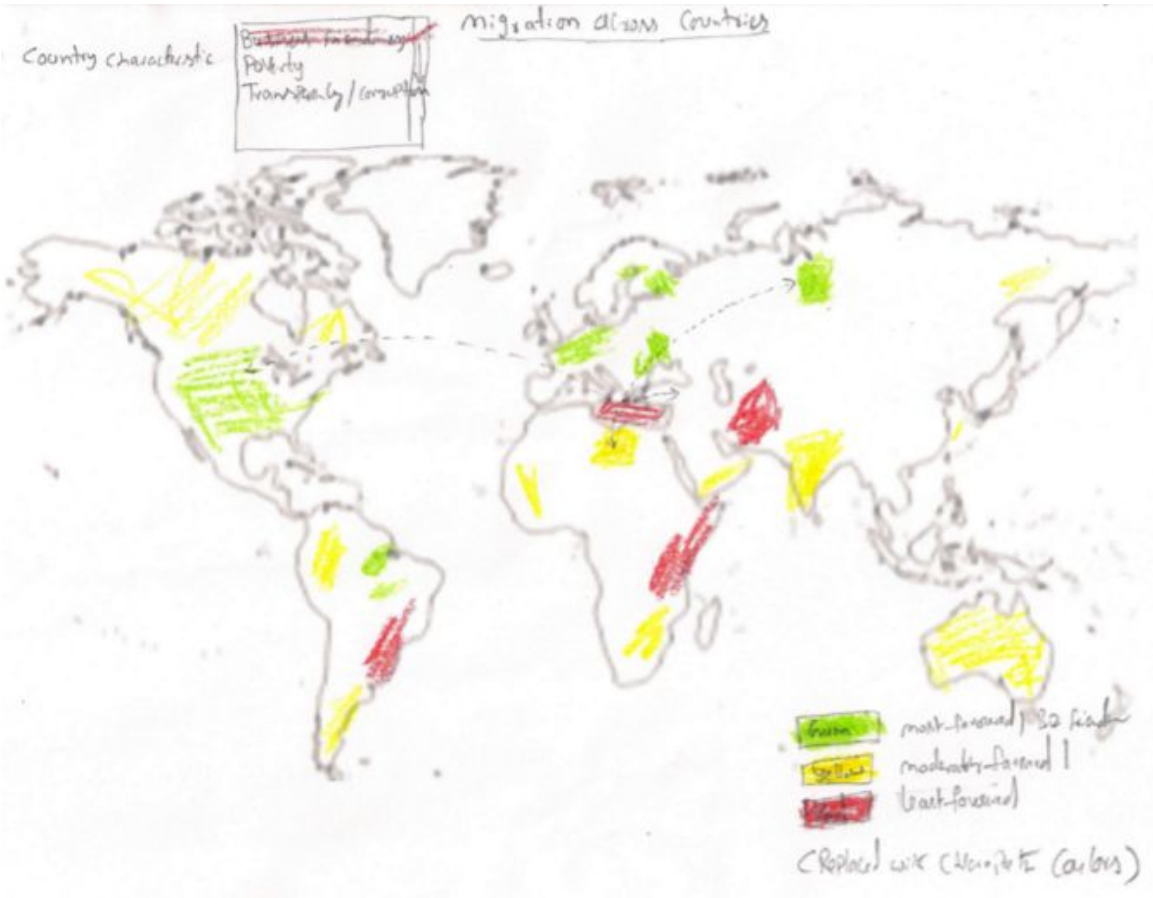
Life expectancy at birth

- Multiple years of data
- Data on 247 economies (including regions)
- Source: data.worldbank.org

Design evolutions

Initial sketches:(03/28/16)

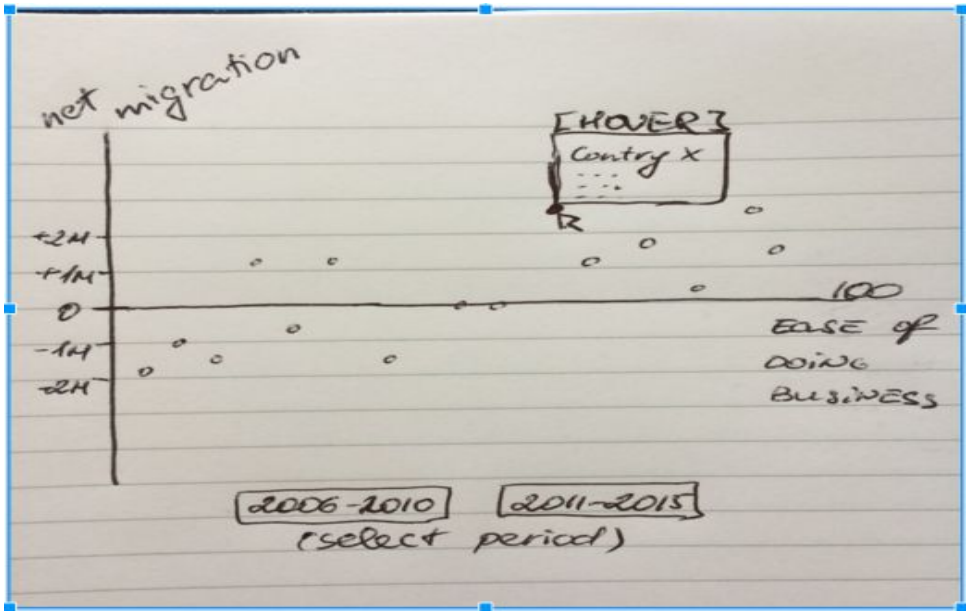
Choropleth map



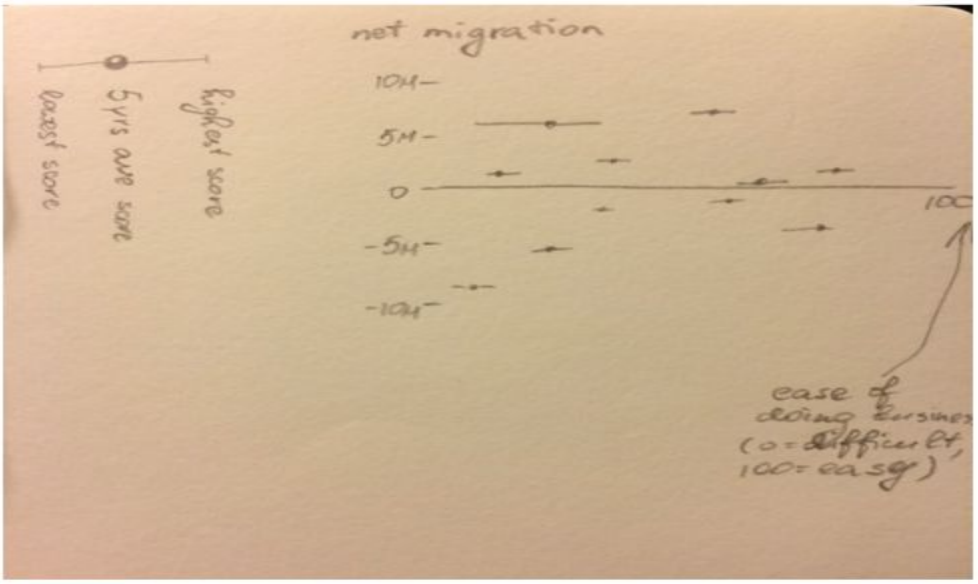
Use dropdown to select development indicators to overlay on world map
Selected indicator will be rendered on choropleth map
Mouse over on a country then it will show all detailed information about that country
Single click on a country it will demonstrates moving arrows/path to a target country (migration). Speed of movement differentiates most favored destination vs moderate vs lowest

Inspired from: http://globe.cid.harvard.edu/?mode=gridmap&id=NG#

Scatter plots: Selection of the region to have all countries from that region be highlighted

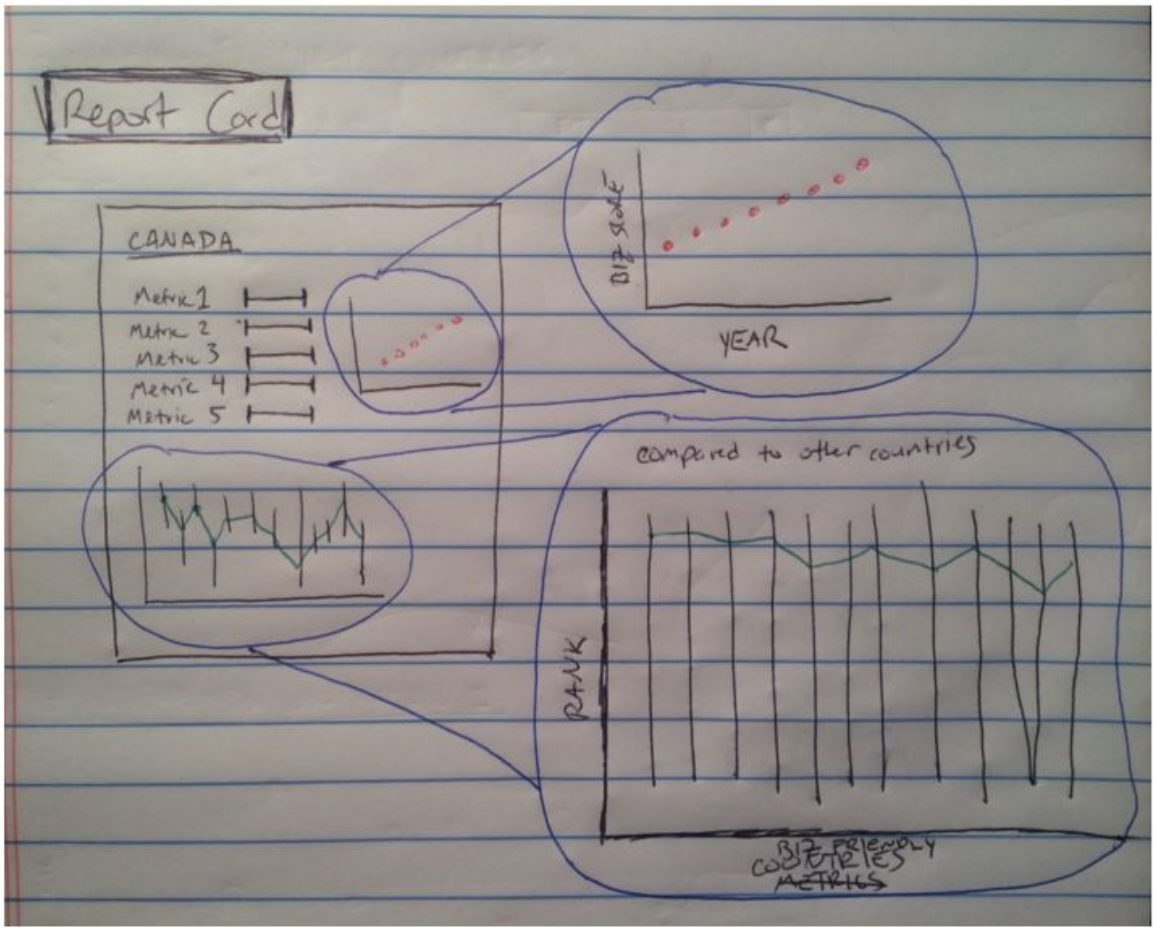


Show the lowest and the highest score on ease of doing business in the 5-years period



Report card

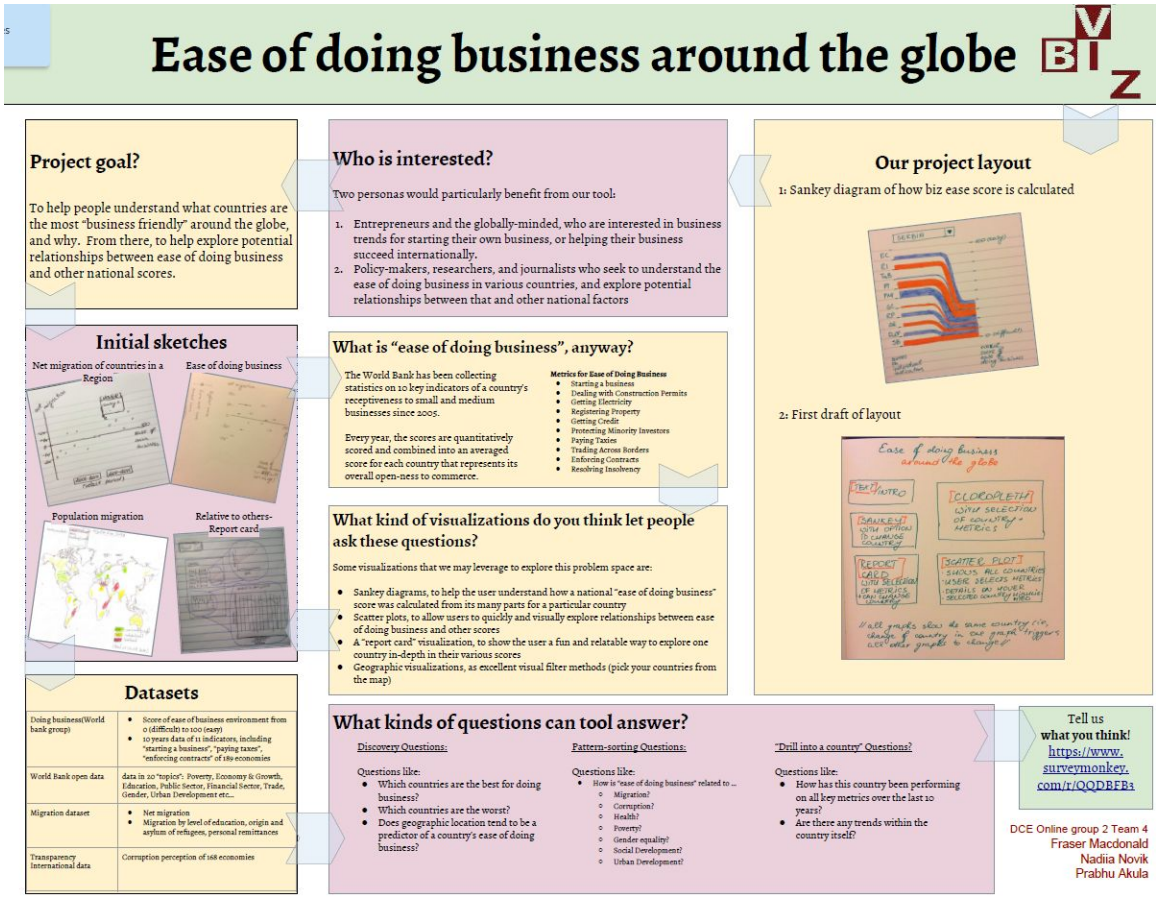
Each country could have a “report card” that shows its ranking relative to others, or how well it has done. This concept is familiar to viewers, and would provide a storytelling format that makes the results more engaging than pure numbers.



Note: At the later stage, we decided not to use the report card for the complexity of the view. This was replaced with basic data in the text format added to the sankey diagram.

Refine designs and create of poster
(04/01/2016)
Based on clarifications received from Zona choose project poster and submitted following poster

https://piazza.com/class/ihv3pvmbyc86g7?cid=616



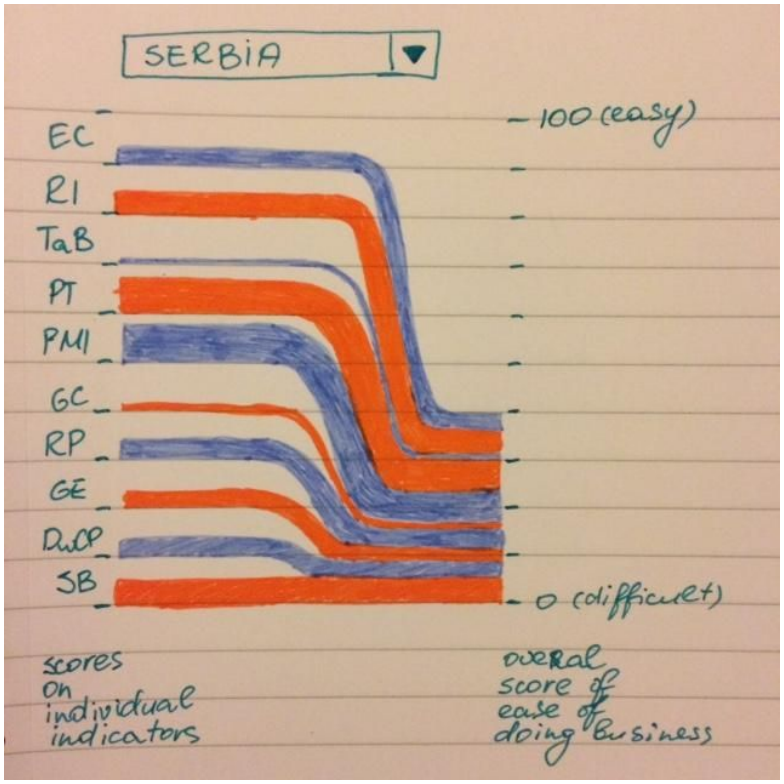
Redesign

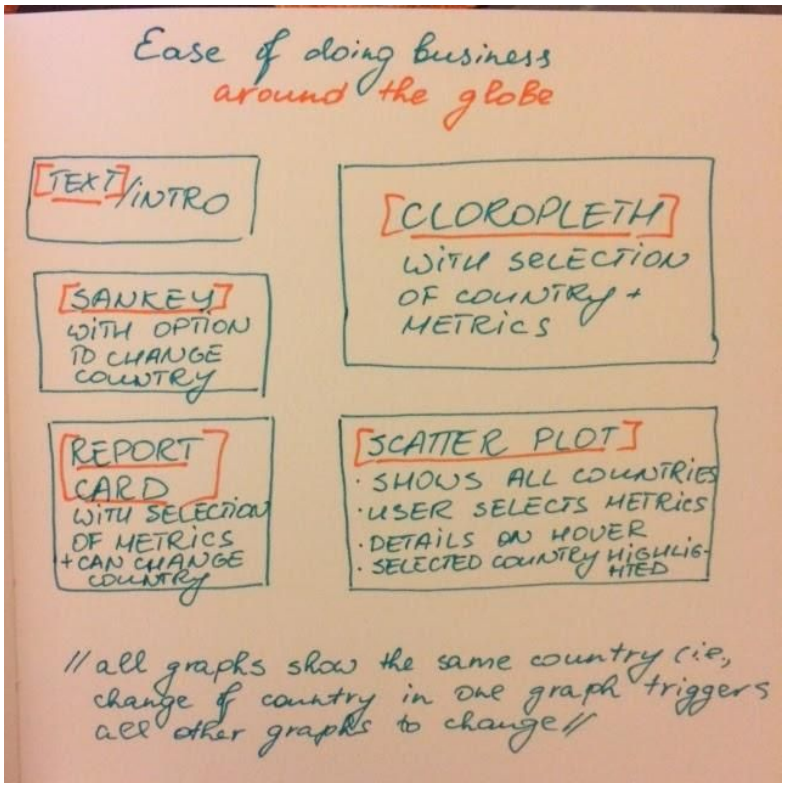
Sankey diagram and Report card

1: Each country could have a "report card" that shows its ranking relative to others, or how well it has done. This concept is familiar to viewers, and would provide a storytelling format that makes the results more engaging than pure numbers.

2: As we learned in class, a scatter plot is the best way to use the "position" channel to encode multi-dimensional information. This scatter plot shows the business friendliness of a country in comparison with the user's selected second metric, letting them explore trends for themselves.

3: Showing the "top X" and "bottom X" countries gives a bird's-eye view of how countries are doing, generally speaking. Summary infographics and visualizations like these are a good framing mechanism to let the viewer put a broad understanding around the problem before diving into details.


SERBIA
EC
RI
TaB
PT
PMI
GC
RP
GE
DwSP
SB
100 (easy)
0 (difficult)
scores on individual indicators
overal score of ease of doing business


Ease of doing business around the globe
TEXT/INTRO
CLOROPLETH
WITH SELECTION OF COUNTRY + METRICS
SANKEY
WITH OPTION TO CHANGE COUNTRY
REPORT CARD
WITH SELECTION OF METRICS + CAN CHANGE COUNTRY
SCATTER PLOT
· SHOWS ALL COUNTRIES
· USER SELECTS METRICS
· DETAILS ON HOVER
· SELECTED COUNTRY HIGHLIGHTED
// all graphs show the same country (ie, change of country in one graph triggers all other graphs to change //

04/06 –
Group met and decided on the visualizations to work toward, splitting up the tasks
Spent a long time evaluating candidate datasets for completeness and relevance
Decided on life expectancy, corruption, and migration

04/08-
Met with group 3 and received feedback on visualization Added legend to choropleth
As indicated in visual will use choropleth colors

04/09 - met with group 5 for their expert take on our prototype

04/13- Team met and divided up visualizations to work on
Fraser to work on scatter plot, Nadiia will be working on Choropleth map and Prabhu to work on Sankey diagram

For V1 implementation we will be using convenient data format or data file, however it was decided to fine tune dataset and web layout subsequently

04/18- Fraser to host GitHub pages and integrate three visualizations

04/18- Working visualization for V1 submission
http://frasermac.github.io/ease_of_business/

04/25 & 05/02- Working visualization for V2 submission
http://frasermac.github.io/ease_of_business/

Storytelling and web layout

Data wrangling and reformatting data

We started with four datasets, which we sourced, cleaned up, imported, and scraped more.

- **Our primary data:** Doing Business dataset
    - Overall score for each country is an average of country's scores on individual indicators (starting a business, paying taxes, enforcing contracts, etc.)
    - For years 2004-2009 not all individual indicators are available (as some were introduced at a later stage). Therefore, we calculated overall scores for those 5-years by averaging available individual indicator scores. We also added a note for the user when selected year falls into the group where overall score is based on a different set of individual indicators

    - We added country codes in order to allow linking data with other data sets

- Migration dataset

    - We focused on the data set which provides net average migration per country over the 5-years periods
    - We added country codes in order to allow linking data with other data sets
    - We manually removed regional data - it's large numbers (given that values are net) were distorting scale used for individual countries
    - We had to address issues related to processing negative numbers, so uploading of this data set is done in a slightly different way compared to other data sets

- Corruption dataset

    - We added country codes in order to allow linking data with other data sets

- Life expectancy at birth dataset

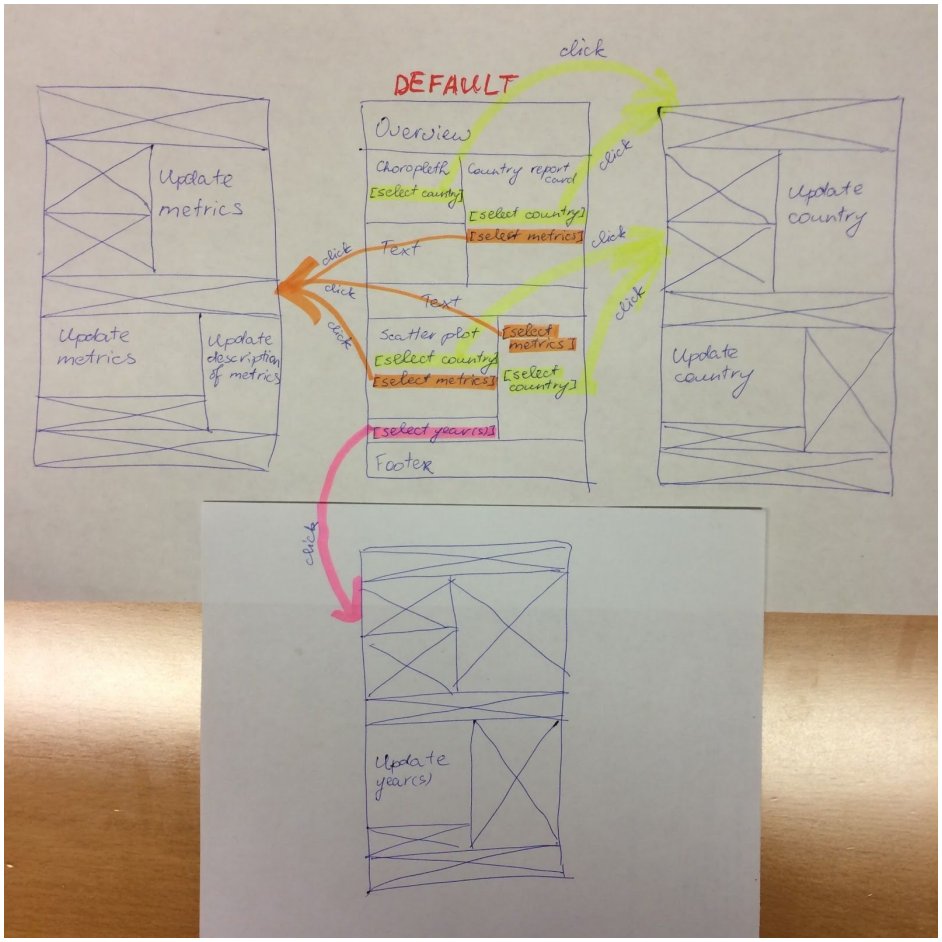
    - We added country codes in order to allow linking data with other data sets

In addition to these primary datasets, we used a dataset of flags (https://github.com/stefangabos/world_countries) to make our country-based visualizations more relatable and appealing to the viewer.

We also used a topojson with all countries (https://gist.github.com/mbostock/4090846), as well as dataset with list of countries with their geographical coordinates and their 3-letter abbreviation (https://d3-geomap.github.io/map/choropleth/world/), which allowed to link the choropleth with available country-level data.

Implementation

Storyboard



During the implementation phase, the features of the storyboard (what is changed by what selection) has slightly changed. But we maintained the general approach of linking the view in different graphs.

Features

Country selection

Country can be selected (1) from drop-down menu and (2) by clicking on the map. Selected country is highlighted on the map with a green border, and details for the country are displayed on the right from the map.

Year selection

Years can be selected for the displayed metric (shows on map and the scatter plot below the map), and separately - for the country details on the right from the map. Years are selected in respective drop-down.

Metric selection
Map and the scatter plot support changing of the displayed metric (for the map) and comparator metric (for the scatter plot; doing business data is always basic data set for scatter plot). Metric can be selected in the drop-down menu, and selection will affect both - map and the scatter plot - simultaneously.

Choropleth
Choropleth supports displaying of all 4 metrics (doing business, net migration, corruption perception, and life expectancy), as well as allows used to select displayed year.

Sankey diagram
Sankey diagram provides details on how the overall score of ease of doing business in composed in each country. Additionally, country's results on net migration, life expectancy and corruption perception are displayed in the text box. The country and the year can be selected by user.

Scatter plot
The basic (primary) data for the scatter plot is score on ease of doing business. To start scatter plot, the comparator (secondary) metric should be selected from the drop-down menu. This metric can be changed at any time. Displayed year of data can be also selected from the drop-down "Metric Year".

Self-evaluation

Overall, we are satisfied with results of our work. The final website of our project answers the questions we initially asked in a clear and comprehensive manner. We implemented several interactive visualizations, which are logically linked between each other. We did not use too much text in order to keep the focus on the visualizations and opportunities for exploring the data.

What did we learn about the data?

Regarding the data, we learned that Africa shows lower performance on ease of doing business as well as on life expectancy at birth. The explanation for this remains outside the scope of our project, but we might assume that the two results are related to some extent. This assumption is further supported by the scatter plot when it is set to showing doing business data compared to life expectancy. That view shows that people in countries with better score of ease of doing business tend to have longer expected life, and *vice versa*. The similar situation is observed with

the corruption perception vs. business environment. However, the geographical view shows that corruption remains of an issue in most countries, irrespectively of the region (with exception of the Northern America). On the other hand, there does not seem to be much dependence between ease of doing business and net migration data. While it is reasonable, that there are many factors affecting people's migration decisions, the results are unexpected as we thought there would be more correlation between the two data sets (metrics).

Lessons learned

1. We learned that raw data requires a lot of cleaning up and formatting in order to use it in the visualizations. Data clean-up is a multistage process, some steps of which happen as a result of code debugging. Careful and thorough design of data requirements help the process.
2. Integrating multiple visualizations in one project is not as easy as it might sound. While we had separate visualizations working well and correctly, it took long time to keep the functionality after merging all the codes in one place. Keeping the consistent variables naming, as well as using one approach to code structure and style helps.
3. Maintaining the single coding style across pieces prepared by different people not only proves to be very useful (especially for debugging purposes), but also shows the entire project as a one-piece rather than a collection of single codes.
4. Although a lot of time and efforts are invested in the preparation stage, the plans can (and most likely will) change at the implementation stage.