PROCESS BOOK

Team name: We D3
Team member 1 (name, email): Phyllis Zhang, phylliszhang@college.harvard.edu
Team member 2: Ethan Lee, ethan_lee@college.harvard.edu
Team member 3: Alma Fredriksson, afredriksson@hsph.harvard.edu

Project Title: *Visualizing Global Health Policy*

Project Abstract. As diseases claim a large portion of human lives, we're particularly interested in noting the differences in government policies and health infrastructure that may be causal in determining the resulting population affected by such diseases. As we are currently in the midst of a pandemic, our team has noted the significance of government policy and health and economic benefits in controlling the disease spread and ultimately the number of people infected or killed. Abstracting this to other diseases over time will give us a more general overview of how countries respond to epidemics, how their health infrastructure is, and how these generally change the number of people affected. With a goal that the governance of other countries following in implementing effective measures, our project allows us to draw conclusions on which health policies/infrastructure components are most important to maintaining a healthy country with adequate resources to respond to an epidemic (or more severe).

We plan on using data from the Global Burden of Disease study, which is available through the Global Health Data Exchange (linked below) to provide an overview of the current global situation. More detailed information on health outcomes, mortality and perceived health status will be obtained from the health data provided by OECD.stat. The same source also provides datasets on government policy and healthcare quality, including variables such as immunization coverage, number of physicians and hospital quality, which we intend to link with the health data. We will potentially also complement this with detailed data on government healthcare spending from the Gapminder Foundation.

One open source dataset published in *Nature* also portrays the different policies put into place by various governments in response to the COVID-19 pandemic. This specialized dataset could be analyzed and visualized to provide the effectiveness of one specific case of health policy for different countries. Additionally, the Quality of Government dataset, collected by researchers at the University of Gothenburg, contains statistics for all countries on various categories relating to the government policies and civilians' wellbeing, such as those relating to corruption, education, health, gender equality, etc. This dataset can be linked with global health datasets to try and identify the primary causes of low health outcomes and quality internationally.

Potential data sources:
http://ghdx.healthdata.org/gbd-2017 (global burden of disease)
https://stats.oecd.org/Index.aspx?ThemeTreeId=9 (Tables for health status of various countries e.g. mortality, perceived health status by age & health care resources e.g. number of physicians, etc & utilization data per country e.g. immunizations)
https://www.gapminder.org/data/ (government health spending)
https://www.nature.com/articles/s41562-020-0909-7 (covid policies)
https://www.gu.se/en/quality-government/qog-data/data-downloads/standard-dataset (qog data)

WEEK 1 - TEAM AGREEMENT & DETAILED PROJECT PLAN

Team Agreement

- *How will you communicate?*
  - We will communicate via a group text message chat.
- *How/when will you meet?*
  - We will meet on Fridays at 5pm ET.
- *How will you collaborate on implementation?*
  - We will keep each other updated via our group messaging and work together on our process book. Additionally, we have created a repository on GitHub with each of us as collaborators.
- *What are your team roles? (decider, GitHub czar...)*
  - Our team roles are equal for now.
- *How will you deal with non-performing members?*
  - We will tell them that they're not performing as expected. We will make expectations clear and be proactive in updating everyone if there are any blockers preventing us from meeting expectations. Additionally, we will ask for help if we are struggling.

Signatures: Alma Fredriksson, Phyllis Zhang, Ethan Lee

Date: October 29, 2020

Detailed Project Plan

Basic Info.

Team name: We D3
Team member 1 (name, email): Phyllis Zhang, phylliszhang@college.harvard.edu

Team member 2: Ethan Lee, ethan_lee@college.harvard.edu
Team member 3: Alma Fredriksson, afredriksson@hsph.harvard.edu

Project Title: *Visualizing Global Health Policy*

Background and Motivation. *Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.*

As diseases claim a large portion of human lives, we're particularly interested in noting the differences in government policies and health infrastructure that may be causal in determining the resulting population affected by such diseases. As we are currently in the midst of a pandemic, our team has noted the significance of government policy and health and economic benefits in controlling the disease spread and ultimately the number of people infected or killed. Abstracting this to other diseases over time will give us a more general overview of how countries respond to epidemics, how their health infrastructure is, and how these generally change the number of people affected.

Related Work. *Anything that inspired you, such as a paper, a website, visualizations we discussed in class, etc.*

One source of inspiration for this project was the Harvard class *GENED 1093: Reimagining Global Health*, which highlights the global inequities in health and emphasizes the effects of social, economic and health policy in shaping the global health situation. This made us curious to further explore and visualize some of these links.

Audience and Questions. *Provide a description of your audience and the primary questions you are trying to answer with your data story. Do you have any overarching goals and objectives that you want to accomplish?*

Our target audience is the general health-interested civilian as well as governments and health-oriented industries. We would like to inform our audience of various diseases and their spread over time and importantly, motivate political and social action within regions to push for administrations (government or industry) to implement effective policies or create infrastructures to reduce the effects of epidemics (or more severe). The ultimate goal is for the governance and industries of countries to follow in implementing effective measures after becoming more aware of how other regions have been handling disease and after urges by their constituent population, as our project allows us to draw conclusions on which health policies/infrastructure components are

most important to maintaining a healthy country with adequate resources to respond to an epidemic (or more severe).

Questions:

1. How does a country's health infrastructure, in terms of vaccination rates, hospitals, medical staff, etc, affect a country's response to a health crisis?
2. What is the impact of spending more of a country's GDP on health related purposes on a country's response to health crises?
3. How do a country's quality of government (i.e. policies on equality of opportunity, corruption levels) affect its ability to minimize the impacts of health crises?
4. How have different country's responses to the COVID-19 pandemic specifically varied across different policy and government systems?
5. Is there an association between the coverage of antiretroviral therapy and the number of annual HIV deaths?
6. Is there an association between the coverage of the World Health Organization's tuberculosis treatment program (DOTS) and the number of TB deaths and new cases?
7. Is there any correlation between the % of births attended by skilled health staff and the maternal mortality ratio?
8. Which are the major causes of death and disability-adjusted life years (DALYs) globally and how has this distribution changed over time?
9. What does the distribution of disease/disability/mortality look like across different continents/regions/countries?

Data. *From where and how are you collecting your data? If appropriate, provide a link to your data sources.*

We plan on using data from the Global Burden of Disease study, which is available through the Global Health Data Exchange (link) to provide an overview of the current global situation. More detailed information on health outcomes, mortality and perceived health status will be obtained from the health data provided by OECD.stat (link). The same source also provides datasets on government policy and healthcare quality, including variables such as immunization coverage, number of physicians and hospital quality, which we intend to link with the health data. We will potentially also complement this with detailed data on government healthcare spending from the Gapminder Foundation (link).

One open source dataset published in *Nature* (link) also portrays the different policies put into place by various governments in response to the COVID-19 pandemic. This specialized dataset could be analyzed and visualized to provide the effectiveness of one specific case of health policy

for different countries. Additionally, the Quality of Government dataset (link), collected by researchers at the University of Gothenburg, contains statistics for all countries on various categories relating to the government policies and civilians' wellbeing, such as those relating to corruption, education, health, gender equality, etc. This dataset can be linked with global health datasets to try and identify the primary causes of low health outcomes and quality internationally.

Data Cleanup. *Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented? Try to minimize the amount of cleanup you have to do by finding cleaned and ready-to-go data sources whenever possible.*

The primary concern for data cleanup will most likely be ensuring that data about countries from different datasets can be easily connected and conclusions can be effectively drawn from the combination of the datasets. Through this process, we plan to derive quantities related to the health policies and statistics of different countries, as well as other aspects of government and legislation that are potentially linked to these factors. Additionally, the data processing will also include the standard process of ensuring that quantitative values are represented as numbers, words are represented as strings, etc. This overall cleaning process will most likely not take a lot of time and should be implementable simply by parsing through each row before working with the data in D3.

More specifically, we expect to extract the following parameters from each of the data sources.

Global Burden of Disease study: *All of the data below are yearly by country.*

- Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories,
- Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017

OECD.stat: *All of the data below are given over time.*

- Percent of GDP spent on Health per country
- Various health care resource numbers per country:
    - Number of physicians per 1000 people
    - Number of hospitals per million population
    - Number of medical graduates per 100000 population
    - Numbers of certain medical technology per 1000000 population
- Various healthcare utilization numbers per country:
    - Vaccination rates (Among children and for the influenza)

Gapminder: *All of the data below are yearly by country.*

- Life expectancy
- Child mortality
- Measures of vaccine perception
- HIV:
    - Annual HIV deaths
    - Newly infected
    - ART coverage
- TB:
    - deaths/100K people
    - new cases/100K people
    - treatment program (DOTS) coverage
- Malaria:
    - deaths/100K people
    - cases/100K people
- Maternal mortality:
    - proportion of births with maternal death
    - % of births attended by skilled health staff
- Government health spending:
    - % of total government spending
    - USD per person
- Overall health spending:
    - share of total health spending provided by government
    - share of total health spending provided by out-of-pocket payments
    - share of total health spending provided by private entities
- Access to at least basic sanitation (%) (overall, urban and rural)
- Access to at least basic water source (%) (overall, urban and rural)
- Measures of income inequality
- Aid received (per person, % of GNI, total)

Nature COVID-19 dataset: *All data below available by country*

- Type of government policy taken to address COVID-19
- Geographical & human target of policy, if applicable
- Mechanism of travel targeted by the policy
- Enforcement level of the policy (i.e. mandatory vs. voluntary)
- Enforcer of the policy
- Timing of the policy (date announced)

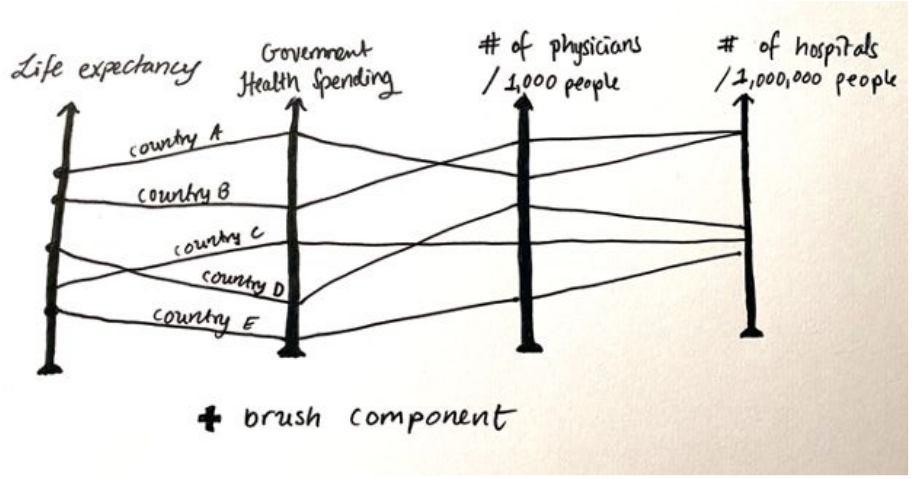
Quality of Government: *All data below available by country*

- Corruption
    - Citizens' perspective of level of corruption in government
    - Policies against corruption in country's legislation
    - Political & Public sector corruption index
    - Citizens' views on whether corruption can be combatted
- Equality
    - Female to male wage ratio
    - Gender Inequality index
    - Level of equality in education across genders
    - Freedom of religion
    - Average schooling years across different demographics
- Health
    - Overall happiness
    - # of doctors per 100,000 inhabitants
    - Life expectancy
    - Adult mortality rate
    - Total govt expenditure on health

WEEK 2 - DATA, SKETCHES, DECIDE AND STORYBOARD

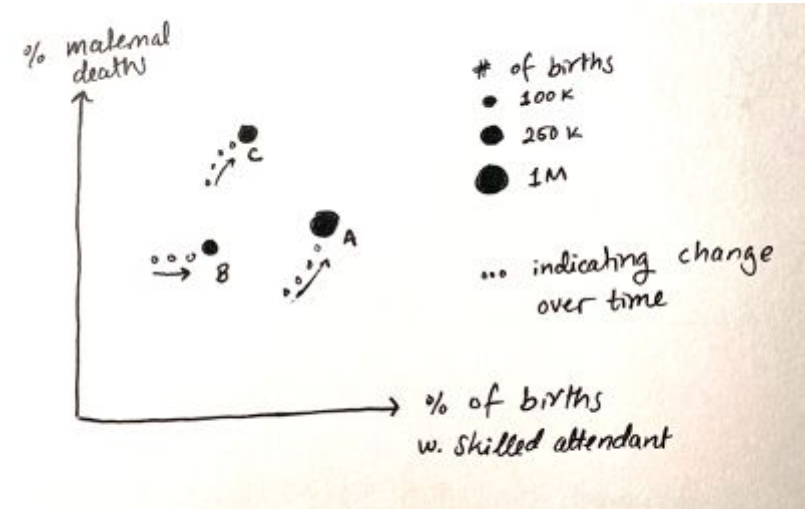
Sketches.

SKETCH # 1 (Alma)

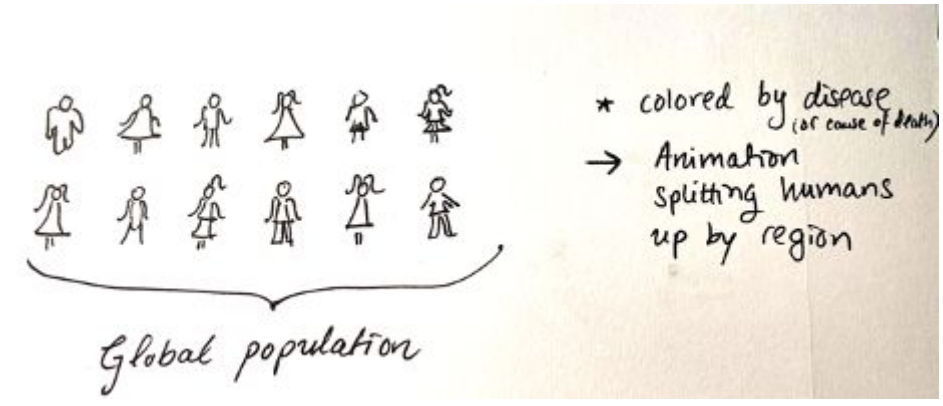


SKETCH # 2 (Alma)

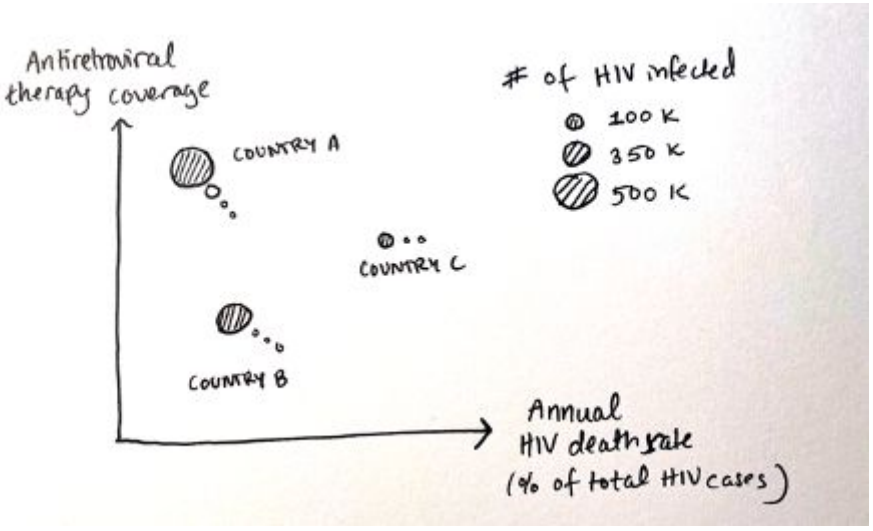
SKETCH # 3 (Alma)



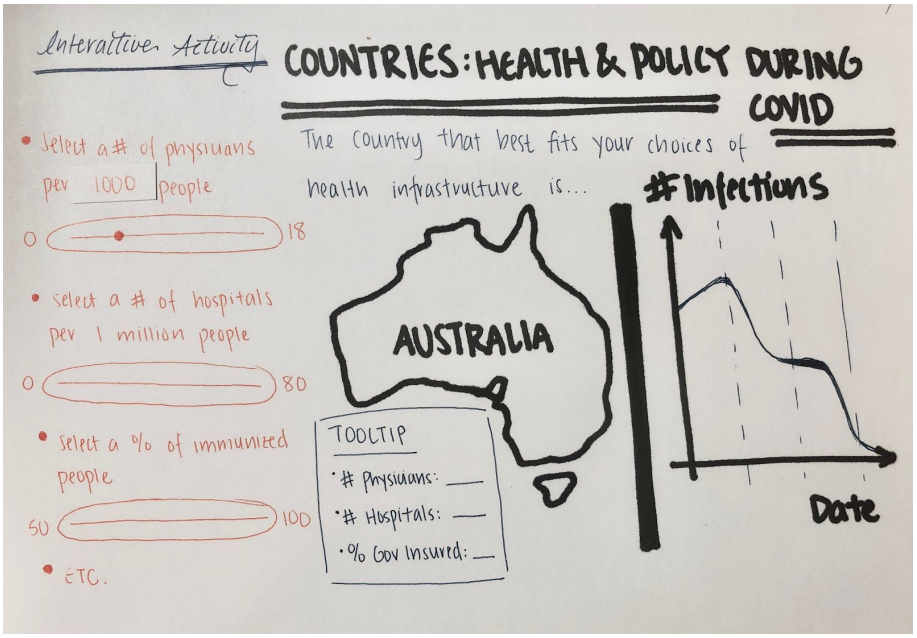
SKETCH # 4 (Alma)



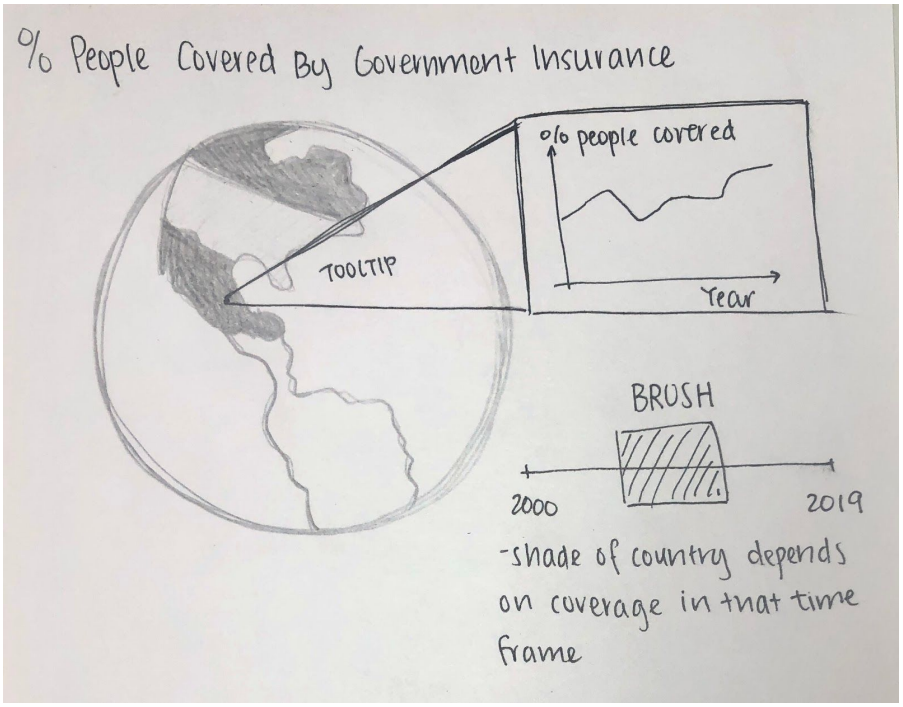
SKETCH # 5 (Alma)



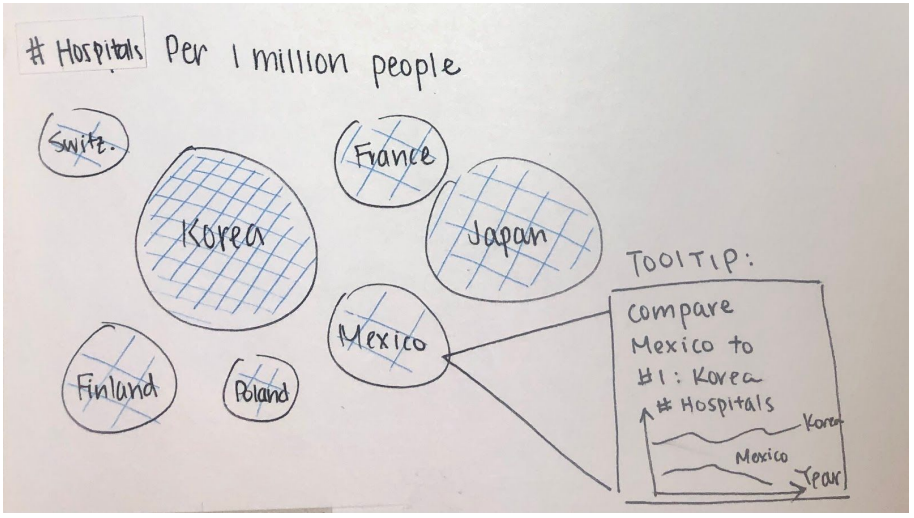
SKETCH # 6 (Phyllis)



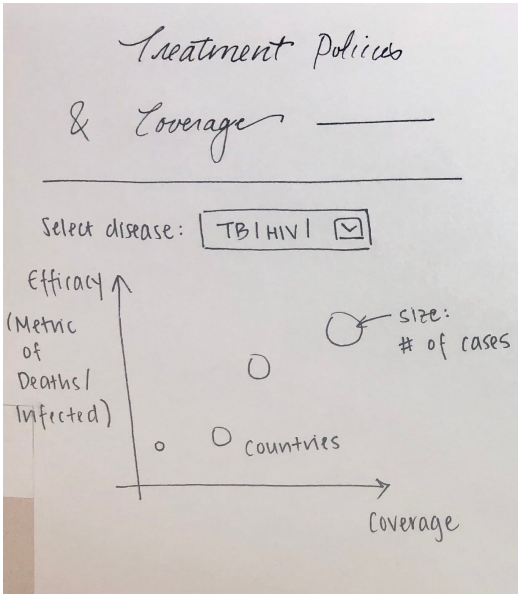
SKETCH # 7 (Phyllis)



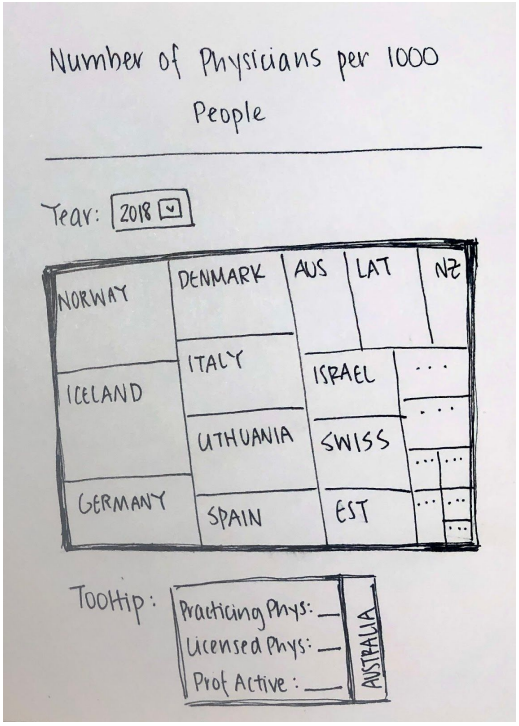
SKETCH # 8 (Phyllis)



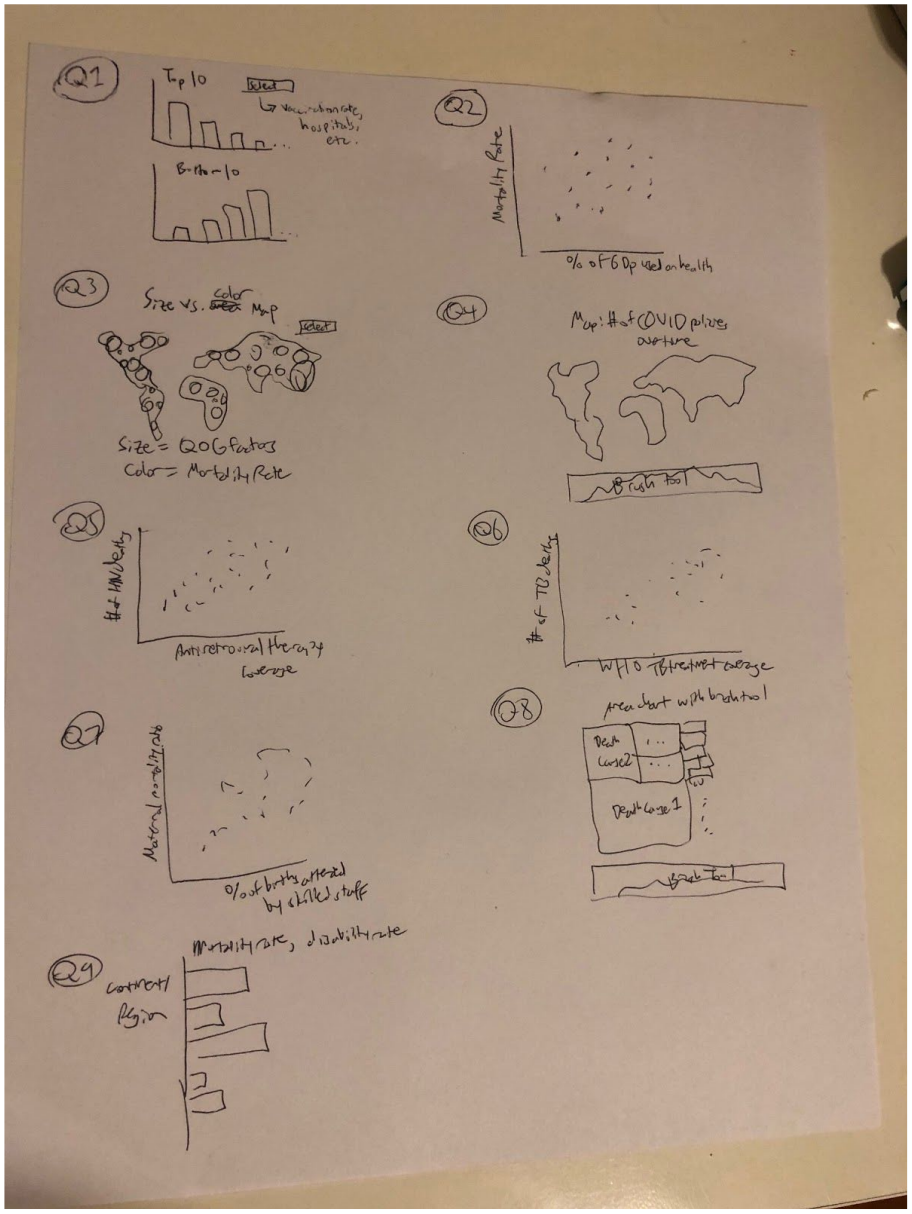
SKETCH # 9 (Phyllis)



SKETCH # 10 (Phyllis)



SKETCHES 11~19 (Ethan)


Q1
Top 10
Select
Bottom 10
Q2
Mortality Rate
% of GDP used on health
Q3
Size vs. color Map
Select
Size = GDP factor
Color = Mortality Rate
Q4
Map: # of COVID patients over time
Brush tool
Q5
# of HIV/AIDS
Antiretroviral therapy coverage
Q6
# of TB deaths
WHO TB treatment coverage
Q7
Maternal mortality rate
% of births attended by skilled staff
Q8
Area chart with brush tool
Death cause 2
Death cause 1
Brush tool
Q9
Mortality rate, disability rate
Continent/ Region

Motivation.

We have selected sketches 1, 2, 5/9 (same visualization), and 6, with sketch 6 being duplicated to allow for users to select policies as well.

We believe that sketch #1 will be able to tell a more general story regarding which policies and sets of health infrastructure are best for a country in terms of disease outcome, as using a coordinate plot will allow the user to draw the general conclusion that higher health spending, higher number of hospitals, etc will result in lesser deaths/infections. Additionally, the user will be able to hover over any specific countries they may be interested in. In terms of interactivity, we will allow for brushing along each policy/health infrastructure pipe, as then users will be able to closely inspect only the countries within a certain range of, for example, health spending. Finally, we'll also allow for the user to pick their own time (in years), as to allow them to see how the effects change over time.

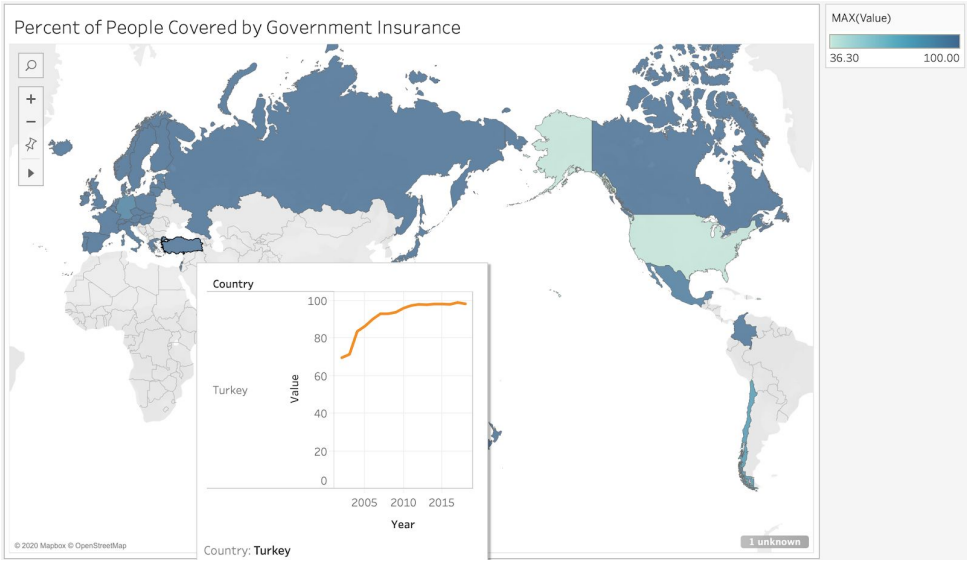
We have additionally selected sketch #2 to be the introduction to our project, as it will be able to effectively provide a distribution regarding how each country fares with the user selected infectious disease. It will allow for both metrics on DALYs and on deaths. We intend for the visualization to be interactive, as it will zoom in to the user's location and provide more information on that specific country before allowing them to zoom out and select another country to investigate.

Sketches 5 and 9 are scatter plots illustrating the relationship between treatment coverage and some metric i.e. death rate. This will be shown when a user selects a particular disease to inspect from sketch #1, and it will provide a simple and meaningful way for the user to see how effective certain programs or treatments were, allowing them to make their own judgement on whether or not the WHO selected treatments are best.

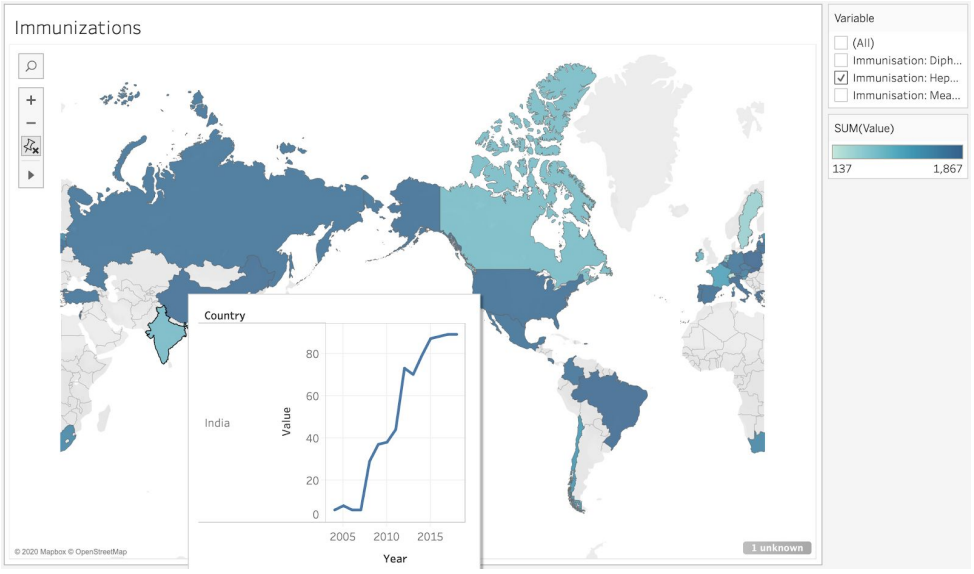
Sketch 6 is an interactive activity for COVID for users to select certain metrics. We will then show them the country that best corresponds to the metrics they just selected within the countries in OECD (Americas, Europe, and a few Asian countries) and show them how such health infrastructure choices affected their change in infected people over time. We have decided to do this for policies as well, allowing users to select the type of policy, the level of enforcement, closing borders, etc, and then a country best fitting their choices will appear showing them how the policies impacted the number of COVID infected individuals. These will allow the users to draw conclusions regarding which policies and health infrastructure are most effective in response to a pandemic.

Tableau Visualizations.

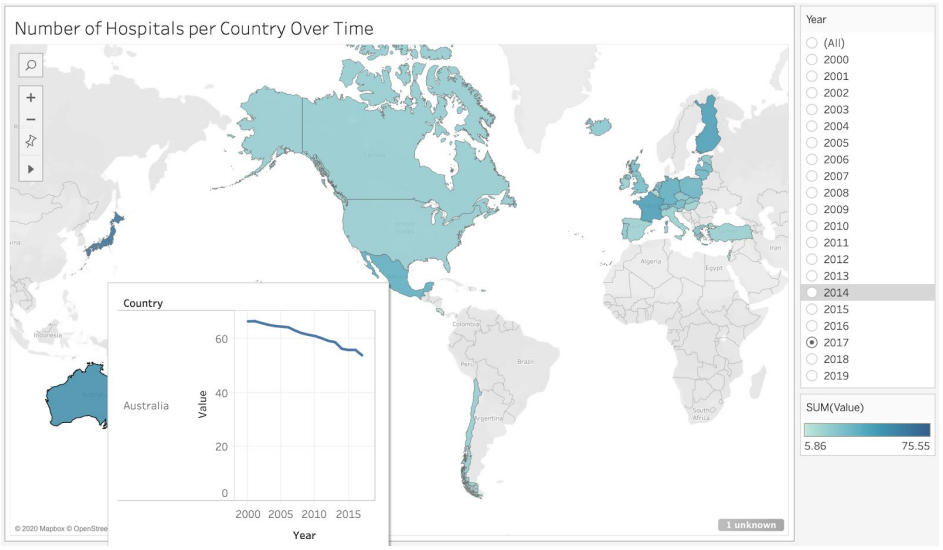
Visual #1 (Phyllis)



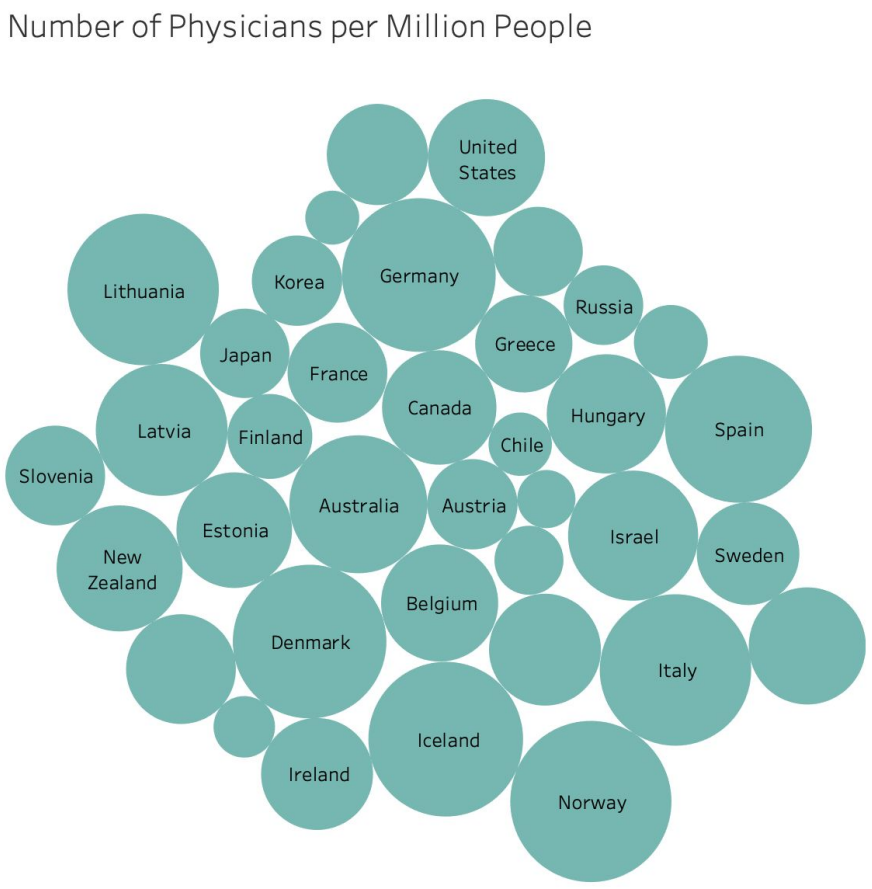
Visual #2 (Phyllis)



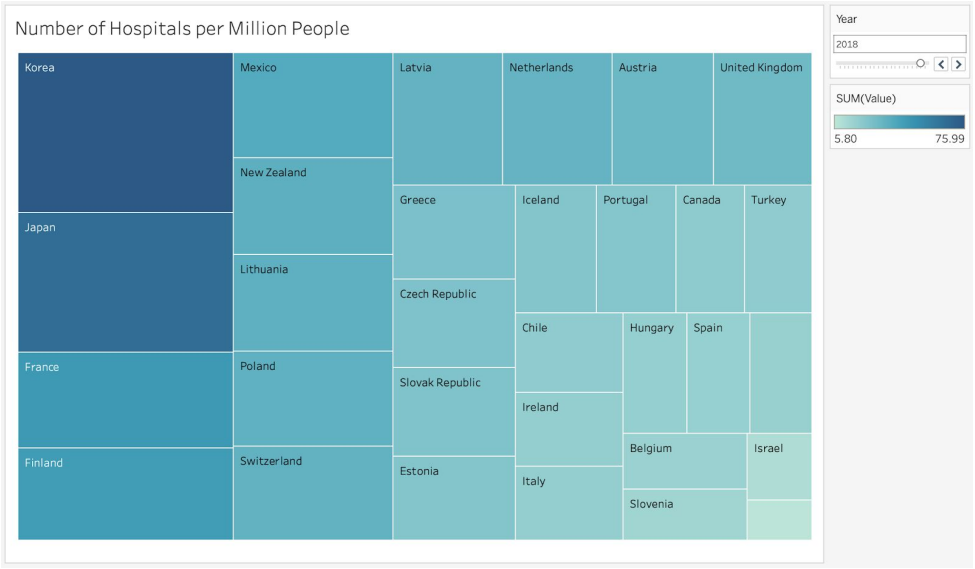
Visual #3 (Phyllis)



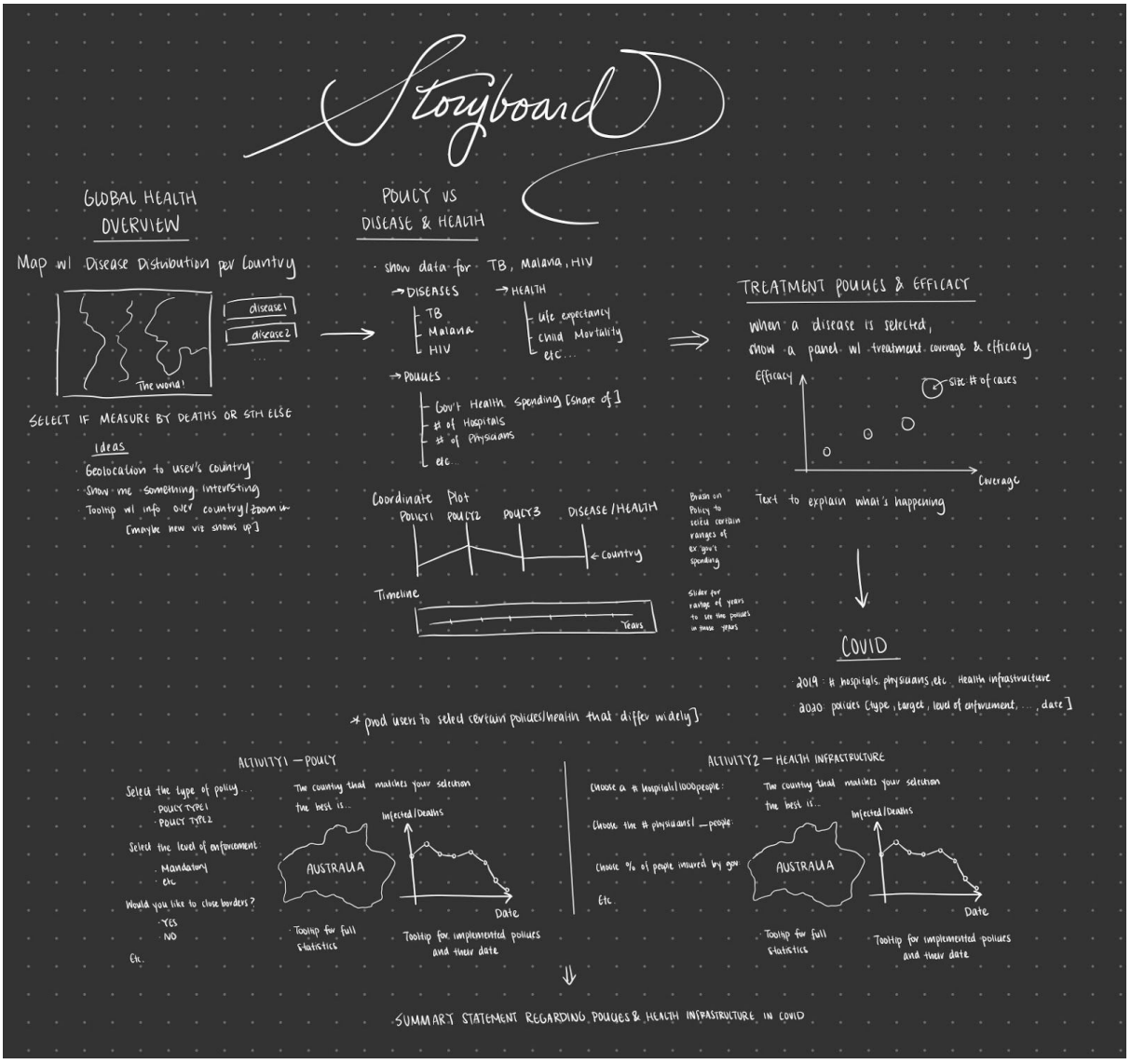
Visual #4 (Phyllis)



Visual #5 (Phyllis)


Number of Hospitals per Million People
Korea
Mexico
Latvia
Netherlands
Austria
United Kingdom
New Zealand
Japan
Greece
Iceland
Portugal
Canada
Turkey
Lithuania
Czech Republic
Chile
Hungary
Spain
France
Poland
Slovak Republic
Ireland
Belgium
Israel
Finland
Switzerland
Estonia
Italy
Slovenia
Year
2018
SUM(Value)
5.80
75.99

Storyboard.


Storyboard
GLOBAL HEALTH OVERVIEW
POLICY VS DISEASE & HEALTH
Map w/ Disease Distribution per Country
disease1
disease2
The world!
SELECT IF MEASURE BY DEATHS OR STH ELSE
Ideas
· Geolocation to user's country
· Show me something interesting
· Tooltip w/ info over country/zoom in
[maybe new viz shows up]
· show data for TB, Malaria, HIV
→DISEASES
TB
Malaria
HIV
→HEALTH
life expectancy
child Mortality
etc...
→POLICIES
Gov't Health Spending [share of]
# of Hospitals
# of Physicians
etc
TREATMENT POLICIES & EFFICACY
when a disease is selected,
show a panel w/ treatment coverage & efficacy
Efficacy
Size: # of cases
Coverage
Coordinate Plot
POLICY1
POLICY2
POLICY3
DISEASE/HEALTH
Country
Timeline
Years
Brush on Policy to select certain ranges of ex: gov't spending
Slider for range of years to see how policies in those years
Text to explain what's happening
COVID
2019: # hospitals, physicians, etc. Health infrastructure
2020 policies [type, target, level of enforcement, ..., date]
* prod users to select certain policies/health that differ widely]
ACTIVITY1 — POLICY
Select the type of policy...
· POLICY TYPE1
· POLICY TYPE2
Select the level of enforcement
· Mandatory
· etc
Would you like to close borders?
· YES
· NO
Etc.
The country that matches your selection the best is...
AUSTRALIA
Infected/Deaths
Date
· Tooltip for full statistics
Tooltip for implemented policies and their date
ACTIVITY2 — HEALTH INFRASTRUCTURE
Choose a # hospitals/1000people:
Choose the # physicians/ —people
Choose % of people insured by gov
Etc.
The country that matches your selection the best is...
AUSTRALIA
Infected/Deaths
Date
· Tooltip for full statistics
Tooltip for implemented policies and their date
SUMMARY STATEMENT REGARDING POLICIES & HEALTH INFRASTRUCTURE IN COVID

WEEK 3 - PROTOTYPE

- Alma, Ethan, and Phyllis have worked on the V1 prototype.
- Interactivity:
    - On the coordinate plot, brushing will allow for certain ranges in health infrastructure to be selected as well as certain time ranges.
    - On the COVID data for health infrastructure, user changeable sliders have been implemented.
    - On the global health overview, users are allowed to select the disease of their choice.
- Non-implemented items will follow the sketches shown above in the storyboard and in the actual sketches.

WEEK 4 - PROTOTYPE

- Alma, Ethan, and Phyllis have worked on the V2 prototype.
- Interactivity:
    - A map has been created that displays values of certain metrics for each country depending on the metric selected.
    - A coordinate plot has been created that changes according to the user selected final metric and allows the user to determine which country corresponds to which line. There is a timeline filter that allows the user to select the years in which they intend to search data from.
    - A scatterplot has been implemented to discuss treatment coverage vs cases for different diseases.
    - A health infrastructure interactive line graph and table have been implemented. Depending on the user's selection of four different metrics, a country will be matched to these values and then plotted against the US in terms of COVID cases. A table will then update displaying the actual values for each country.
- Non-implemented items will follow the sketches shown above in the storyboard and in the actual sketches.
    - The map will have a spider plot adjacent, and this spider plot will cover different diseases in the country highlighted by the user.
    - The coordinate plot will become brushable by each of the metrics.
    - The scatterplot will have a Rosling style animation.
    - Another visualization regarding the government policies implemented by different countries during COVID will be added.

WEEK 5 - TESTING

Tester Name: Austin Hwang
Tester Email: austinhwang@college.harvard.edu

General Observations from the think-aloud study:

- Confused about the coordinate plot and which country corresponded to which

- Thought the map visualization could've used more narration / captions
- Bad data for some of the map visualization (color of the country was just black)
- Generalizing statement for the COVID data would help make a message more clear

What does the tester like about your data story?

- Enjoyed the interactivity of the COVID visualization
- The parallel axis plot was cool even though it doesn't particularly have a message yet

What improvements does the tester point out?

- Have more narration throughout the story
- Improvements on the map and scatterplot to make it more interactive

Was the intended key message clear to the tester? Why or why not?

- The exploration intent was clear, but the tester thinks that it would be best to include a message since it would be "easier" to just write out a message

Did the tester get your next steps or call to action? Why or why not?

- I suppose we didn't particularly have a call to action. We can change that by writing some sort of generalizing statement at the end that encourages governments to see what may be the best set of policies to follow

Tester Name: Gabin Ryu
Tester Email: ryu@college.harvard.edu

General Observations from the think-aloud study:

- Most of the visualizations seemed to be effective (easy to maneuver through, nice interactive parts)
- Parallel coordinates plot especially provides a cool way to look through the data
- Overall needs more text explanation and guides on how each interaction should be used

What does the tester like about your data story?

- Visualizations were solid and are easy to understand in most cases
- COVID policy country-matching visualization was especially interesting, as it allows user to compare their ideas with other countries'

What improvements does the tester point out?

- Adding more text, both in terms of transitioning through the visualizations as well as providing context/explanation on how to use the graphics would be helpful

Was the intended key message clear to the tester? Why or why not?

- Tester wasn't able to discern a clear storyline and message from the visuals as they seemed to be sending different messages from each other. This could be addressed by adding a lot more text, quotes, and simpler, static graphics that lead into each interactive visual that we have

Did the tester get your next steps or call to action? Why or why not?

- As noted before, tester didn't see the storyline clearly, and so next steps and call to action were also not clear; this could be addressed (as noted previously) by adding necessary context/info on next steps in terms of global health policy that countries around the world are taking and how effective they are

Improvements –

- In response to our testers, we plan to tell a more clear data story by adding captions and screens of images and text to introduce users to the following set of visuals and tie different visualizations together.
- We will complete the visualizations and for the ones that are less obvious, write a few tips for users to follow.
- We will provide a final concluding screen that has a more general final message.

WEEK 6 - PROTOTYPE/WRAP-UP

- Alma, Ethan, and Phyllis have worked on the final submission.
- Interactivity:
    - The map has a spider plot adjacent, which covers different diseases in the country highlighted by the user on the map.
    - The coordinate plot is now brushable by certain metrics.
    - Animation has been added to the site – words appear letter by letter, users can go through the content without scrolling via arrows, etc. Words, captions, and pictures have been added throughout the site to allow for contextualization.
    - The scatterplot has a Rosling style animation via a timeline bar on the bottom.
    - A visualization regarding the government policies (COVID) implements a line graph and a bar chart – the line graph allows users to see in detail which policies a particular country implemented, and the bar graph allows users to compare the number of policies and if they were mandatory or voluntary. Users compare two countries at a time.
- Other Documentation:
    - The website has been published at amazhangphyl.github.io/cs171-final-project.
    - A screencast has been recorded and can be found at https://drive.google.com/drive/folders/1_g3xmLSfvcJt_twSIrBejWfwftIr6TK-?usp=sharing.