Coronavirus (COVID-19) Testing Cases & Vaccinations Information Report

Name: Luyang Ye
Student ID: 28049918
Lab Number: 11
Tutor Name: Bruno Luis Mendivez Vasquez

Introduction

This project mainly wants to show the testing cases and vaccination information using various interactive visualization methods. Three core questions need to be answered through the narrative visualization:

1. What are the cumulative testing cases and positive cases increasing trend for each country from last year? And which country has the fast-increasing trend?
2. how many types of vaccines right now?
3. how do cumulative COVID-19 vaccinations administer per country? And how do daily COVID-19 vaccinations administer per country?

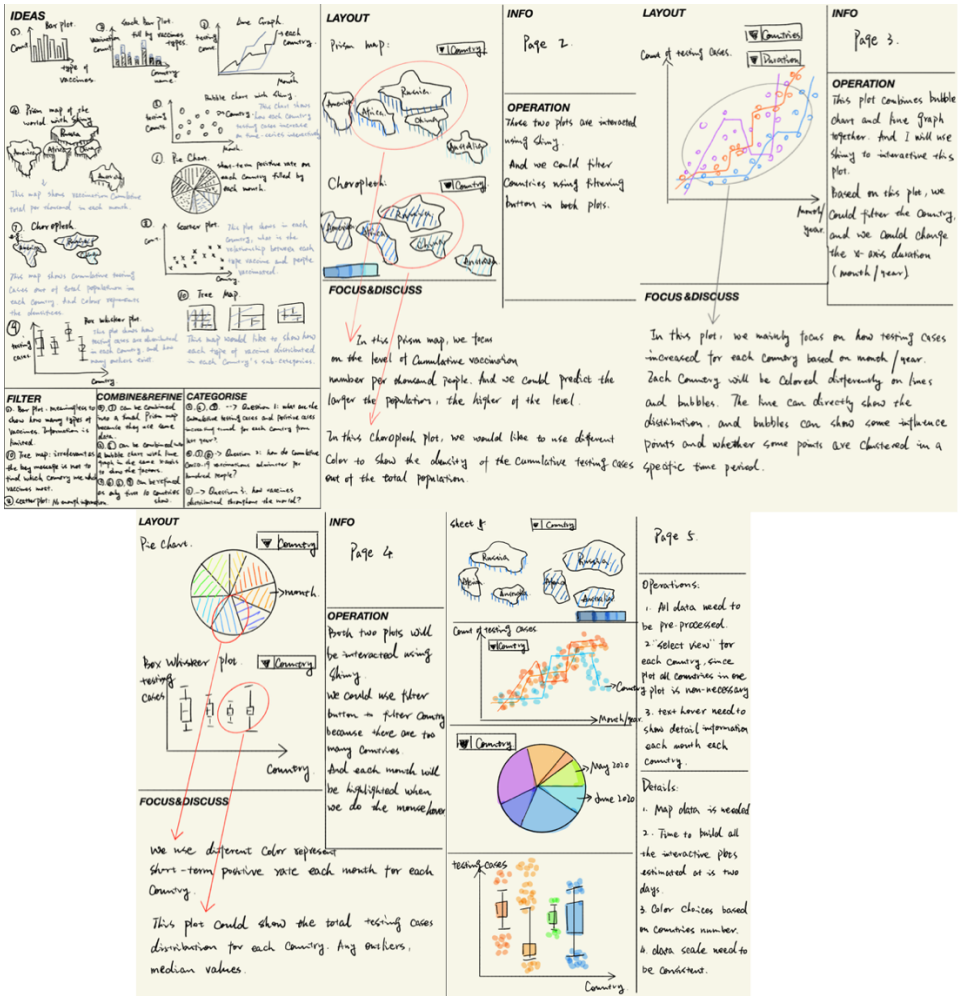
Hence, in my narrative visualization, I would like to show those questions' answers to the audiences interested in the overall Coronavirus situations changing worldwide. And how the vaccination is performed in each country. There are six subsets in my narrative visualization dashboard:

- Overview
- Raw Table
- Testing Cases Information
- Short-term Positive Rate Information
- Monthly Vaccination Information
- Vaccination Map View

The **overview subset** presents the background of this project, and the audiences could have a basic understanding of this project. The **raw table subset** shows the most critical two original databases that we would use to visualize further. And the audiences could check the original data of the country they are most interested in. The **testing cases information subset** uses an animation plot showing how the testing cases amounts and percentages are based on the total population changing for each country each month along the time changing. The **short-term positive rate information subset** uses interactive bar plots showing the distributions of the short-term positive rate for each country. The audiences could realize in which month the country suffered a difficult coronavirus situation. The **Monthly vaccination information subset** would like to present two ideas to our audiences: How many types of vaccines are used in each country? How many vaccinations in each country each month? And we used the data table and bar plot to tell the answers. We could show each country's vaccination amount and vaccination percentage on the map in the **vaccination map view**. And our audiences could directly observe which country had the largest number of vaccinations. And whether the percentage map view would present the same results. Those are the information I would like to show the audiences using my narrative visualization.

Design

At first, we would like to use five design sheet methodology to organize the basic view of the visualizations. In the original, we first brainstorm ten plots for the visualizations. And we need to filter, combine, refine, and categorize three plots as the final view.

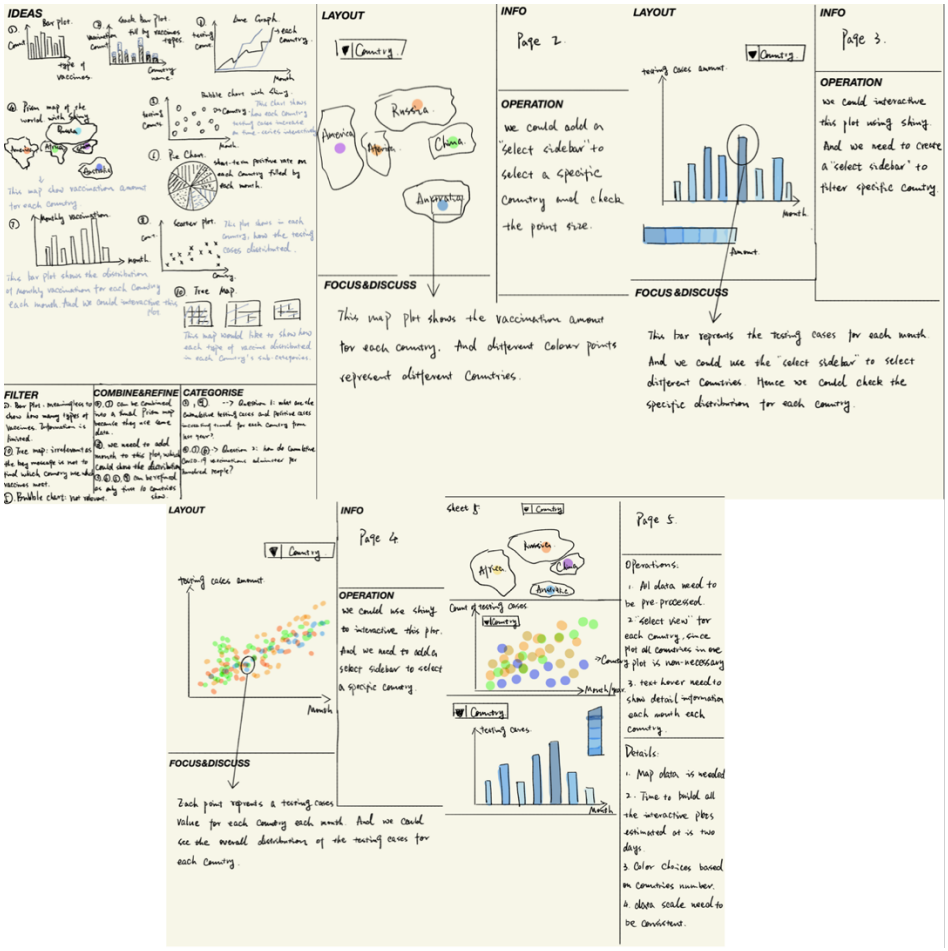


(Appendix: Figure 1)

In my first design, I would like to use the Prism map to show the cumulative vaccination number based on the population for each country and the Choropleth map to show the density of the cumulative testing cases out of the total population. Then in the second plot, I would like to use the combination of scatter plot and line plot showing the changing of the testing cases for each country each month. The line plot could directly show the changes to the audiences for each country, and the scatter plot could show if some influence points exist each month, which could cause the line to increase or decrease. And the third plot, I would like to use the pie chart to present the short-term positive rate for each country each month and the box-whisker plot to show each country's testing cases distribution for each month. However, there are some problems with this five-design sheet.

1. How could we connect the data we collected to the Prism map and Choropleth map?
2. The data we collected are country longitude and latitude data, but the Prism map and Choropleth map may not show it clearly.

3. The combination of the scatter plot and line plot will become quite messy. Since we have lots of countries' data, and in each month, there are many data as well. It isn't easy to plot those things together. And the technological requirement is very high.
4. The box-whisker plot is also not performed well. Because we have a vast number of testing cases for each country each month, we could see lots of outliers that are not useful if we present them separately using a box-whisker plot. And actually, this plot could not tell much information to the audiences.



(Appendix: Figure 2)

Then I prepared an alternative five design sheets to solve the problems in the first five design sheets. In the alternative design, I first change the map to a basic map, and we could plot the vaccination data on the map. Then, I used the bar plot to represent the distribution of the testing cases amount for each country each month instead of combining the scatter plot and the line plot. The bar plot could show the distribution clearer. Finally, I would like to use the scatter plot showing the testing cases changing over time. However, there are still some problems in this alternative five-design sheet:

1. If we only show the vaccination amount for each country on the map view, the information is not enough to let the audiences understand what we really would like to show.
2. The bar plot and the scatter plot using the same data sets, which means we need to add something to identify those two different plots. And we need to use those two plots showing different, various information to our audiences.

3. Each country has more than one type of vaccine, which means it would be tough to plot the types of vaccines for each country. Hence, we need to choose a different way to show this information.

(Appendix: Figure 3)

Now, we need to solve the problems better and produce the final designs. In the final design, we need to show the raw data tables to the audience first. Then, I would like to plot the testing case data using an animation plot. In this animation plot, I choose the testing cases amount as x-axis and percentage of the testing cases as y-axis. And we could play the plot to check the points moving for each country over the timeline automatically. Also, we could adjust the time button by hand to check the points distribution for each country. Next, I would like to plot the short-term positive rate using bar plots. We could use shiny to interact with the bar plots, which means we could select/search a specific country to check the distribution of the short-term positive rate. And from the plot, we could realize which month(s) has the highest positive rate, and we could consider why this could happen. After that, we could move to the vaccination part. I want to show the vaccination amount for each country each month, and we also could use shiny to select a specific country to check the amounts. Meanwhile, I would like to plot the types of vaccines table to directly tell the audiences in each country, how many types of vaccines they used, and those vaccines. Finally, I would like to use the map to plot the vaccination information. And we could adjust the circle radius to present different information. When we select the vaccination amount as radius, the map shows which country has the most significant vaccination amount. Inversely, if we select the vaccination percentage as radius, the map shows which country has the highest percentage

over the total population. And audiences could receive more information and see the difference between those plots.

Before we justification the human perceptual system and human communication assumption of the final design, we need to understand what they mean theoretically. According to Wikipedia (Wikipedia, Perception, n.d.), the human perceptual system organizes, identifies, and interprets sensory information to represent and understand the presented information. Also, according to the Wikipedia (Wikipedia, Human communication, n.d.), the human communication indicate how human communicate. And "Human communication is grounded in cooperative and shared intentions." Since we cannot fully know everything during the communications, we may need to make some assumptions before. In my narrative visualization, I assume the audiences could fully understand what each plot represents, and they could directly and see the difference between each object. And after they see the overall visualizations, they could entirely understand this project, and they could answer the three questions we mentioned before.
For my final design, I first used shiny to interact with all the plots and used different colors to represent each country or each month. The interactive plots within different colors could provide impressive to the audiences straightforward. And I used the animation plot to show the distributions animatedly. Since people are sensitive to the moved plot, this plot would deeply impress the audience. However, in my design, I did not add symbols to represent each country or each month. And I did not add any 3-D plots to let the audience thinking further. Those are the perspectives that I may need to improve later.

Implementation

Now, I will explain the implementation in detail. The interactive method I would like to use is the R Shiny. At first, I will show the libraries that I used to build the shiny application.

```
# Import libraries
library(dplyr)
library(tidyr)
library(shiny) # build shiny app
library(shinydashboard) # create shiny dashboard in the shiny app
library(shinyWidgets) # selectize input function based on this library
library(DT) # show datatable in the shiny app
library(plotly) # create animation plots
library(lubridate) # adjust data format using this library
library(ggplot2) # create static plots
library(leaflet) # show map view in the shiny app
```

(*Figure 4*)

And before we build the shiny application, we had to pre-process the data first. In the following steps, we remove the useless columns for each data table, and I would like to change the date format and filter the not helpful entities. The original data set is daily records for each country. And I would like to group the data by month. I changed the date format because I need to create the animation plot based on the timeline. Hence, I need to

format the date: "Year-Month" and sort the date data ascending.

```
# Remove the non-necessary columns
testdata <- testing_all_observations%>%
  select(-X1,-`ISO code`,-`Short-term tests per case`,-month)
vaccinationdata <- vaccinations%>%
  select(-X1,-X1_1,-iso_code,-people_fully_vaccinated,-people_fully_vaccinated_per_hundred,-daily_vaccinations_raw,-month)

# adjust the date format and remove the countries that contain useless data
pre_processed_testing_cases_data = pre_processed_testing_cases_data%>%
  mutate(Date = format(Month, "%Y-%m"))%>%
  filter(!Entity %in% c("Andorra","Brazil","China","Germany","Hong Kong","Netherlands","Singapore","Spain"))
```

(*Figure 5*)

Since my narrative visualization contains four subsets: testing cases information, short-term positive rate information, monthly vaccination information, and vaccination map view. We need to prepare the relevant data table first. The following tables are what we need to do different visualizations.

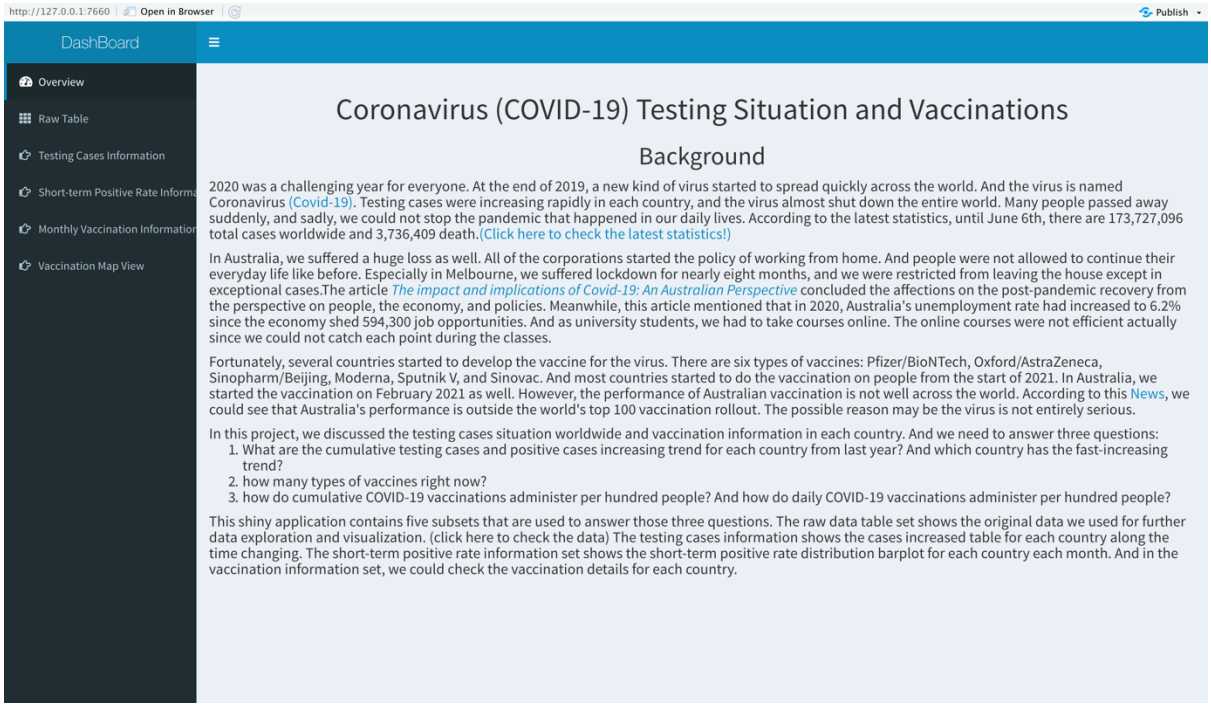
```
# summarize the positive rate data for each month
positive_rate_table <- testdata%>%
  select(Entity,Date,`Short-term positive rate`)%>%
  group_by(Entity,Month=floor_date(Date, "month")) %>%
  summarize(Amount=sum(`Short-term positive rate`))%>%
  mutate(Date = format(Month, "%Y-%m"))

# preprocess the vaccination data
dailyvaccinationdata <- vaccinationdata%>%
  select(-daily_vaccinations_per_million)%>%
  select(location,date,daily_vaccinations)%>%
  mutate(date = format(date, "%Y-%m"))%>%
  group_by(location,date) %>%
  summarize(Amount=sum(daily_vaccinations))

# clean the map data and combine the map data with original data
world_country<-
  world_country_latitude_and_longitude_values%>%
  select(-usa_state_code,-usa_state_latitude,-usa_state_longitude,-usa_state,-country_code)%>%
  rename(location = country)
df_vaccinationdata<-vaccinationdata%>%
  select(location,date,people_vaccinated,people_vaccinated_per_hundred)%>%
  mutate(date = format(date, "%Y-%m"))%>%
  group_by(location,date)%>%
  summarize_all(funs(sum))%>%
  filter(!location %in% c("Algeria","China","Liechtenstein","Maldives","Monaco","Mongolia","Nepal","Panama","Saudi Arabia","
```

(*Figure 6*)

Compared to the final five design sheets, I slightly change the narrative visualization structure. I put the testing cases animation plot first to show a basic overview of the testing cases data changing. Then, I would like to plot the short-term positive rate for each country, which is a further exploration step for the testing cases data. After that, I would like to move to the vaccination data visualizations. I plot the vaccination amounts for each country each month and the types of vaccine table on the same page, which could see the distribution of each country. And I put the map view on the final page, which could show the data word widely.

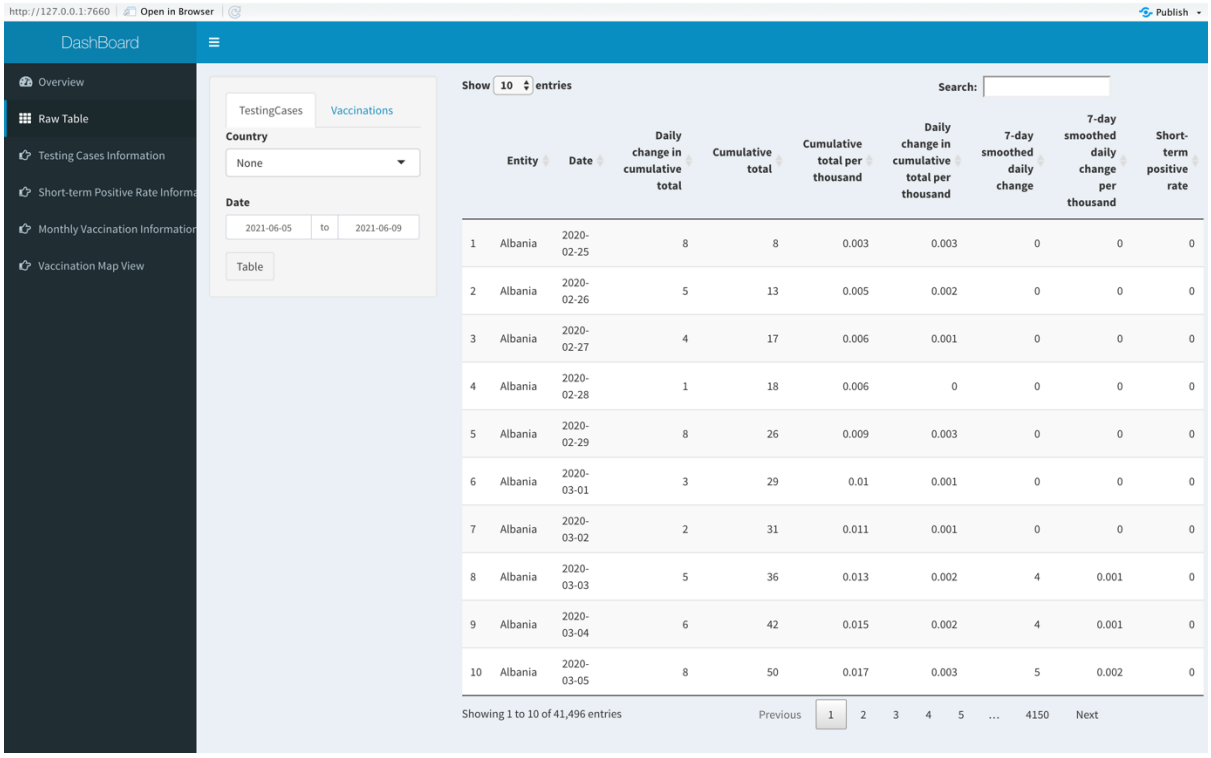
User Guide

Important Hints: It will take a few minutes to start the shiny application. And the data loading may need some time to complete. Since the data set is quite large.



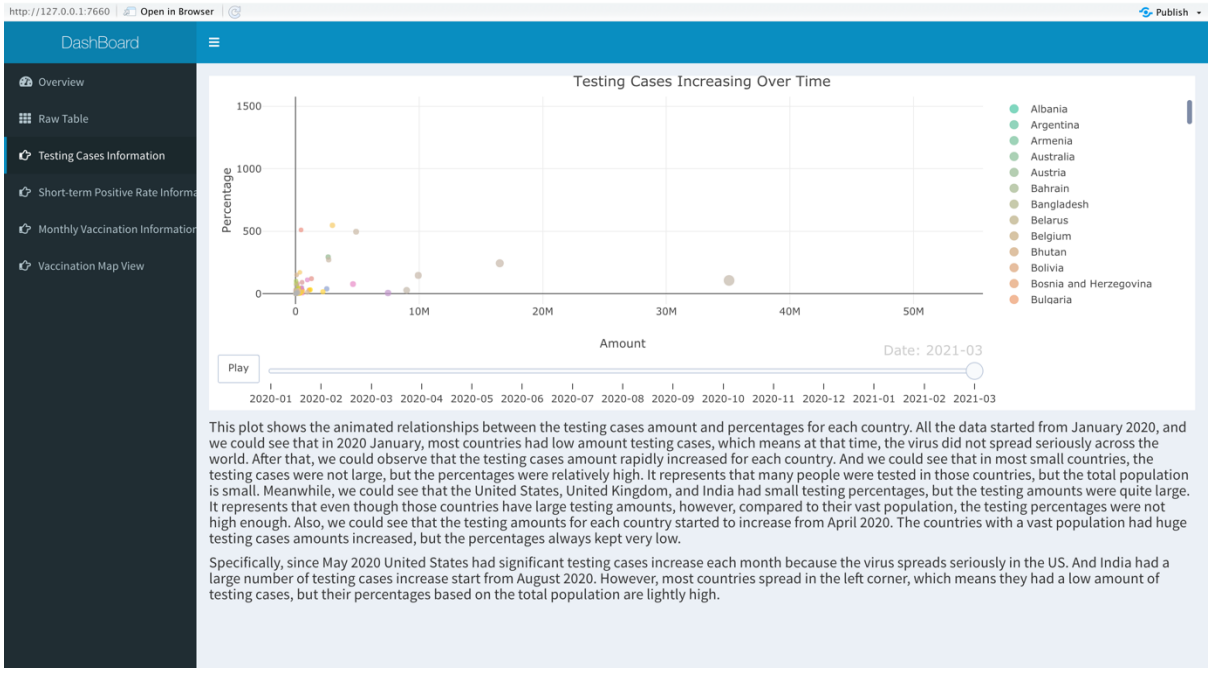
(*Figure 7*)

This is the first page of my shiny application. On this page, I present the background of my project, the coronavirus situation in Australia, the core questions that we need to answer, and the overall shiny application structure. Also, on this page, I add four hyperlinks to connect the website information. If the audience is interested, they could click the hyperlink to check the detailed information. The first hyperlink relates to the definition of the coronavirus, the second hyperlink relates to the latest coronavirus records statistics information, the third hyperlink relates to a formal article that relevant to the coronavirus impact in Australia. The last hyperlink relates to the ABC news about Australia's vaccination performance.



(*Figure 8*)

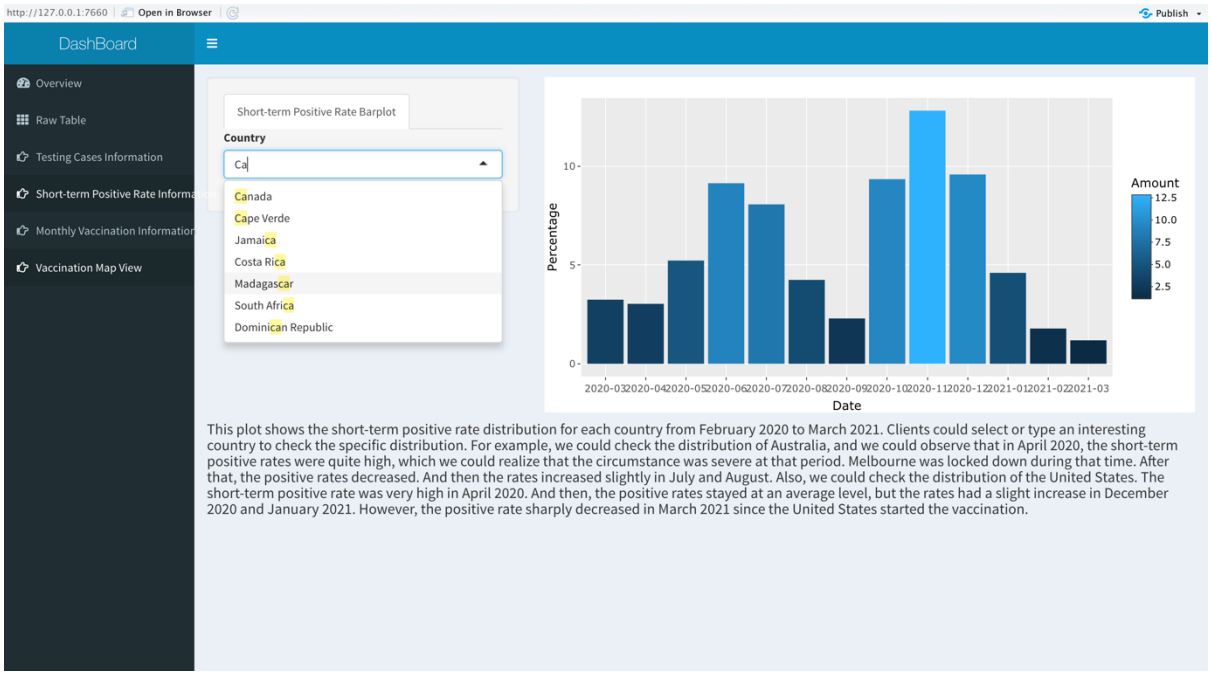
This is the second page of my shiny application. This page presents the two most crucial data tables we used for different visualizations. There are more than 40K data entries. And Clients could use the sidebar to check the data in detail—those two tab buttons link to a different table. The first table mainly focuses on the testing cases data, and the second table mainly focuses on the vaccination data. And we could type or click a specific country or set specific date ranges for each country to have a basic check on the data.



(*Figure 9*)

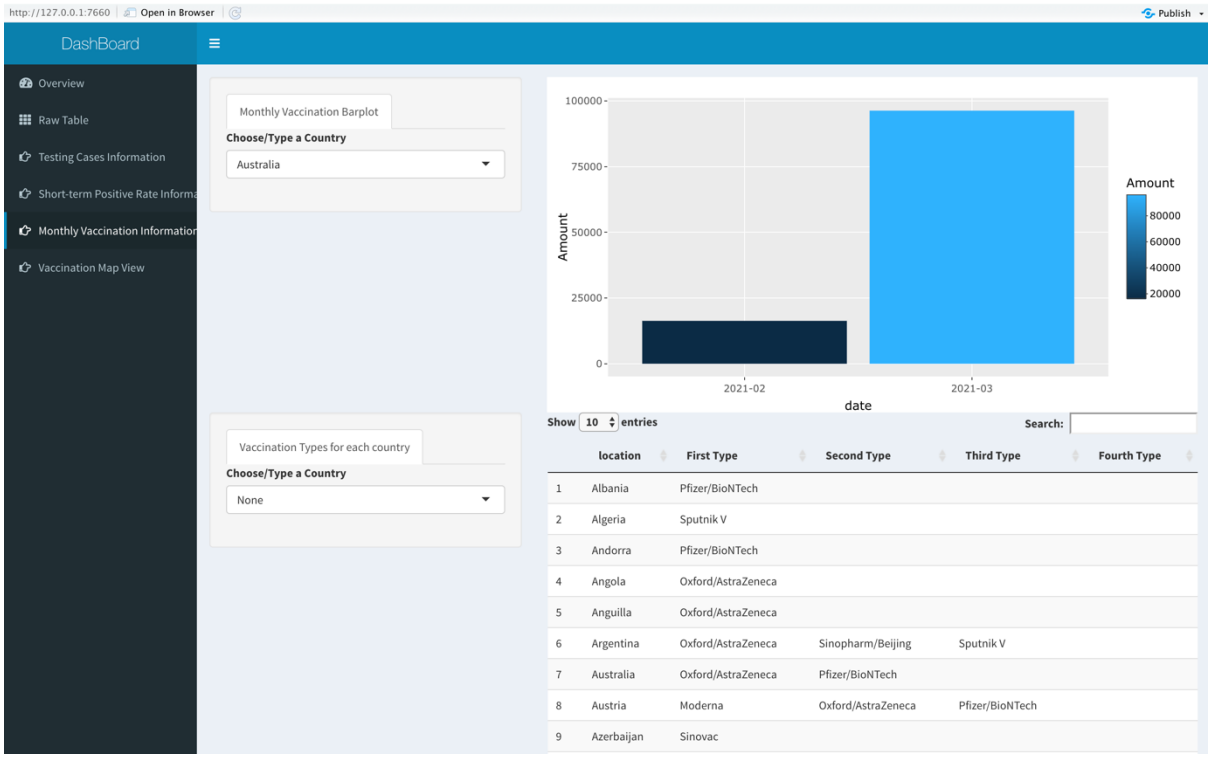
The third page of the application shows the testing cases changing over time in an animated plot. Clients could press the play button to check the points moving status. Also, they could

move the button to a specific month and check the points distribution. The information below the plot discussed the plot performance in detail. Also, clients could click the country name on the right plot side to remove the country point from the plot.



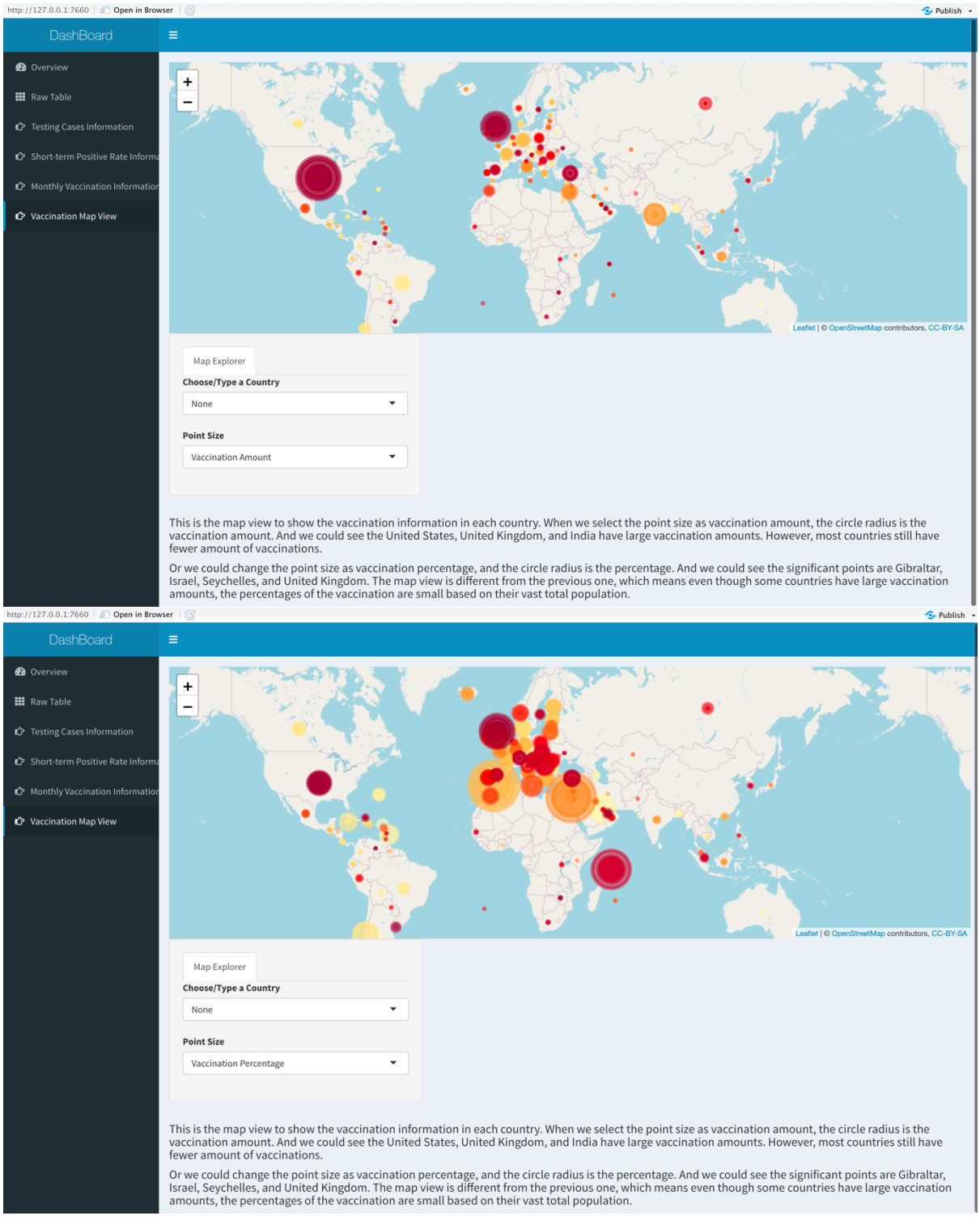
(*Figure 10*)

The fourth page of the application presents the short-term positive rate bar plot for each country. And the x-axis is the date, and the y-axis is the monthly percentage. Clients could type the country name or select the country name from the menu to check the distribution. The information below the plot is the discussion for the plot.



(*Figure 11*)

The fifth page includes two tables. The upper table is the bar plot of the monthly vaccination amount. And the bar plot may look very simple for each country since the vaccination in most countries started from the end of 2020, and the data may be relatively minor. And the second table is the vaccine information for each country. From the table, we could see that more than one type of vaccine they used in some countries. Clients could select the country from the menu or search the country name directly.



(*Figure 12*)

The last page is the map view of the vaccination data. On this page, clients could select different point sizes to check the difference between the map views. And the information below the map discusses the significant map difference and what possible reasons could cause the difference. Clients could also select a specific country, but the zoom size may become huge.

Conclusion

Through the entire the visualization process, we now could answer those three core questions:

1. From the animation plot, we could see the United States had the highest testing cases increasing speed, and the second country is India. And then, we could check the short-term positive rate of the US, and we could observe that the average positive rate is high, and the rate sharply decreased in March 2021, which means when the US started the vaccination, the positive rate decreased obviously.
2. From the types of vaccine table, we could count that now there are six types of vaccine, and they are spread across the world.
3. After we plot the vaccination data, we could see the US, UK and India have the largest vaccination amount. However, the vaccination percentages of those three countries are not high, which means the vaccination amount over the total population is not large.

After doing this project, I learned how to use the shiny to interact with the data plots. And how to build a shiny application combining with some basic HTML. Also, I learned how to complete a whole data visualization project, what steps I need to take, what kind of works I need to prepare, and what results I would like to derive. Meanwhile, I learned how to build animation plots and build map views to show the data distribution in more various ways.

And the adjustment I made is grouping the data by month, and I move the attention on each country. And I calculated the percentage over the total population. The challenge of my project is how to make the data visualization mode human-readable and more reliable. Since there are too many countries' data, which are very tough to plot them in the same plot, the plot will become very messy, and we could not get any helpful information. And the second challenge is how to plot the country data into the world map because initially, we only have the countries data without latitudes and longitudes information. Hence, we need to find a way to solve this problem. The third challenge of my project is the advanced use of shiny to build a beautiful, convenient, and client's friendly shiny application.

Bibliography

Wikipedia, F. (n.d.). *Human communication*. Retrieved from wikipedia: https://en.wikipedia.org/wiki/Human_communication

Wikipedia, F. (n.d.). *Perception*. Retrieved from wikipedia: https://en.wikipedia.org/wiki/Perception

Appendix:

Figure 1-1: Initial FDS

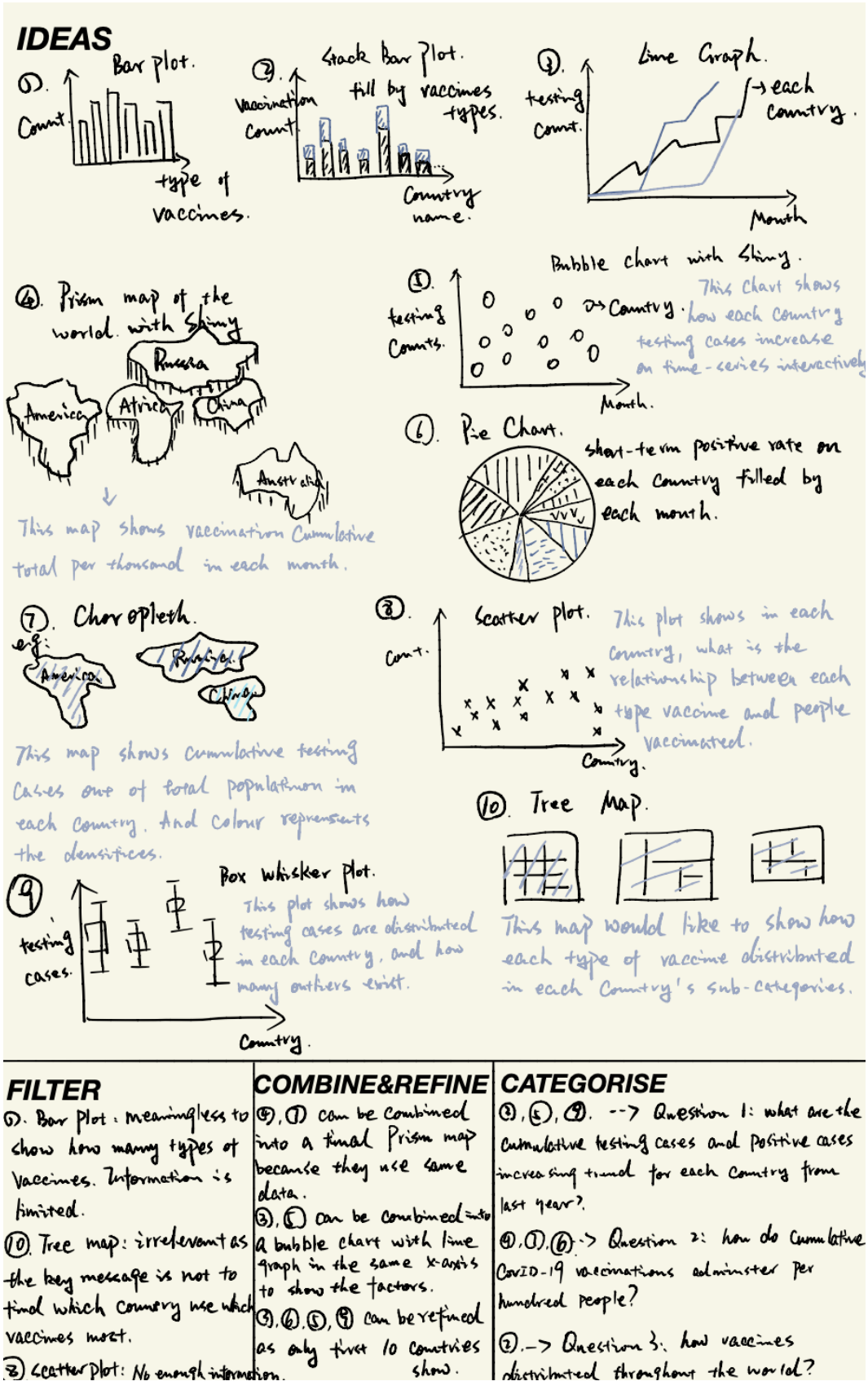


Figure 1-2:

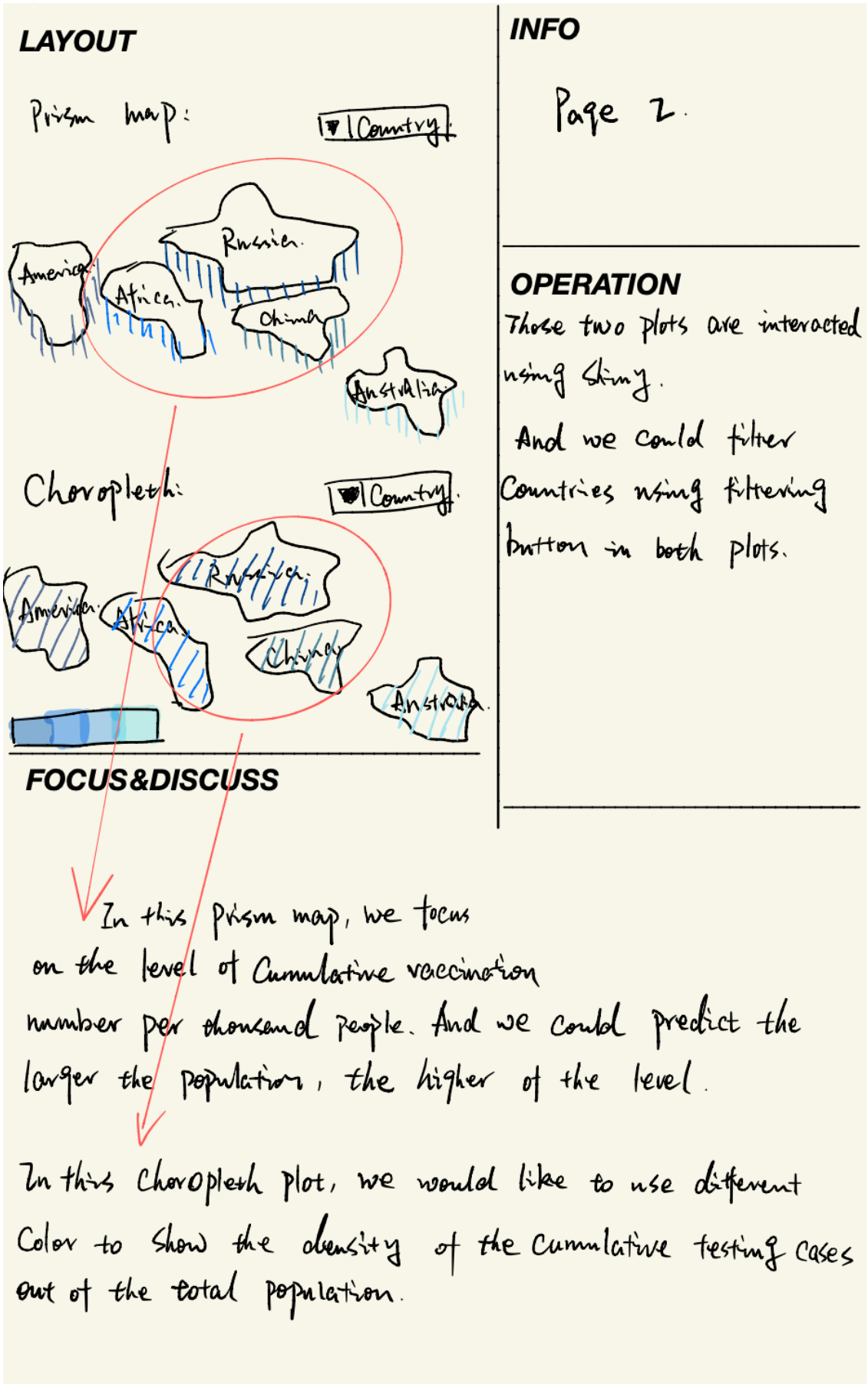

LAYOUT
Prism map:
Country
America
Africa
Russia
China
Australia
Choropleth:
Country
America
Africa
Russia
China
Australia
INFO
Page 2.
OPERATION
Those two plots are interacted using Shiny.
And we could filter Countries using filtering button in both plots.
FOCUS&DISCUSS
In this prism map, we focus on the level of Cumulative vaccination number per thousand people. And we could predict the larger the population, the higher of the level.
In this Choropleth plot, we would like to use different Color to show the density of the Cumulative testing cases out of the total population.

Figure 1-3:

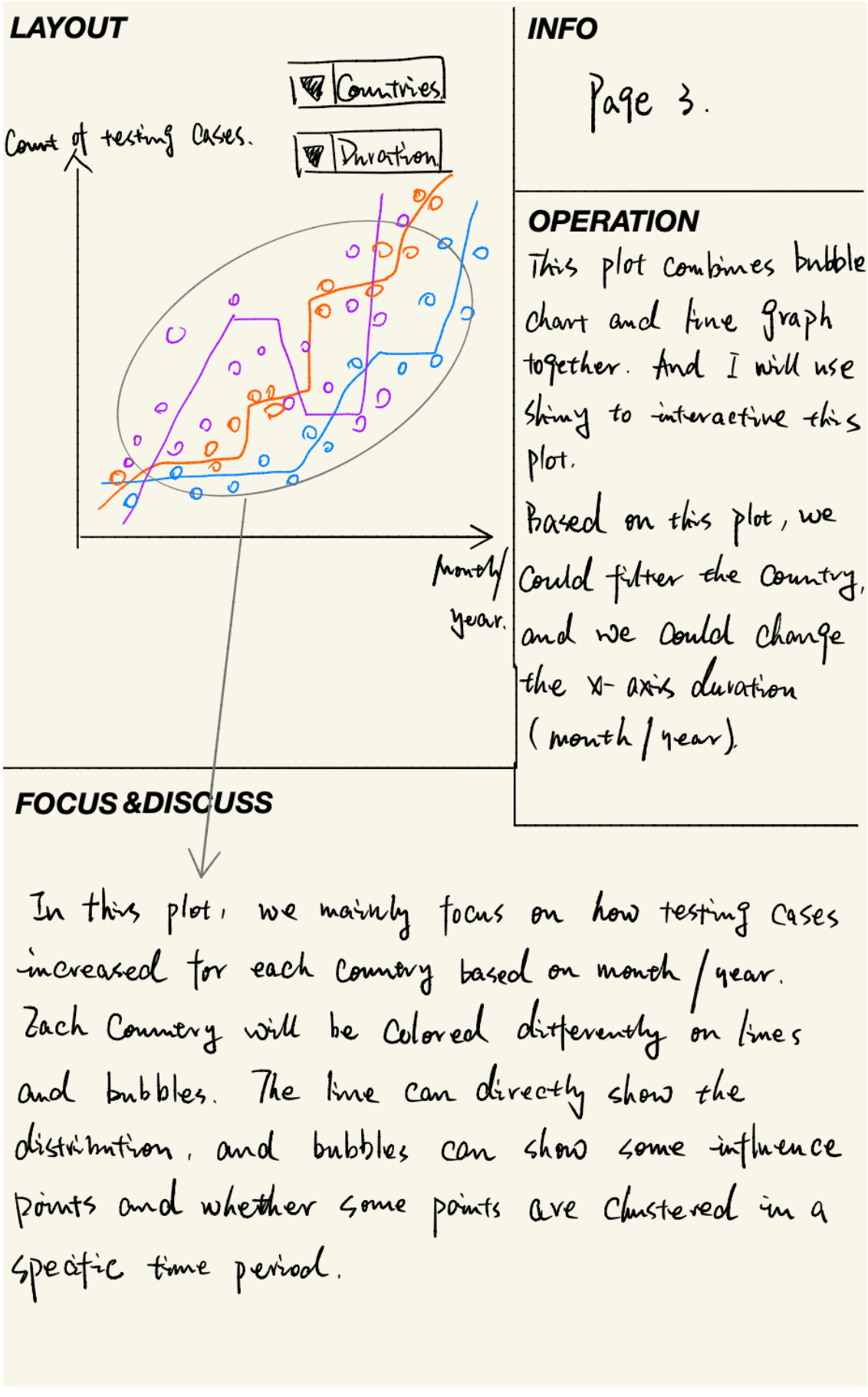

LAYOUT
Countries
Count of testing cases.
Duration
Month/
year.
INFO
Page 3.
OPERATION
This plot combines bubble chart and line graph together. And I will use shiny to interactive this plot.
Based on this plot, we could filter the country, and we could change the x- axis duration (month / year).
FOCUS &DISCUSS
In this plot, we mainly focus on how testing cases increased for each country based on month / year. Each country will be colored differently on lines and bubbles. The line can directly show the distribution, and bubbles can show some influence points and whether some points are clustered in a specific time period.

Figure 1-4:

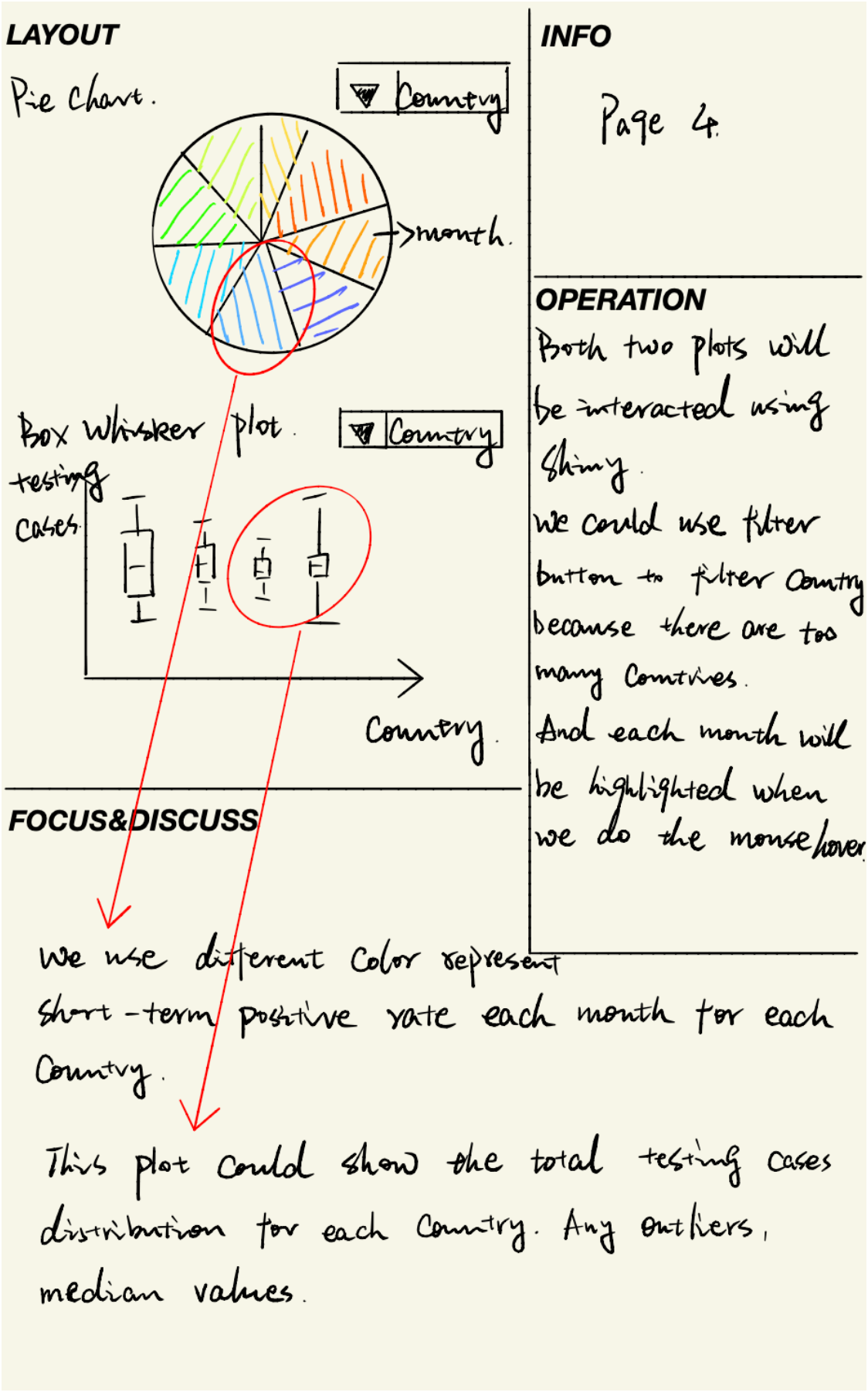

LAYOUT
Pie Chart.
Country
month.
Box Whisker plot.
Country
testing cases
Country.
INFO
Page 4.
OPERATION
Both two plots will be interacted using Shiny.
we could use filter button to filter Country because there are too many Countries.
And each month will be highlighted when we do the mouse hover.
FOCUS&DISCUSS
We use different color represent short-term positive rate each month for each Country.
This plot could show the total testing cases distribution for each Country. Any outliers, median values.

Figure 1-5:

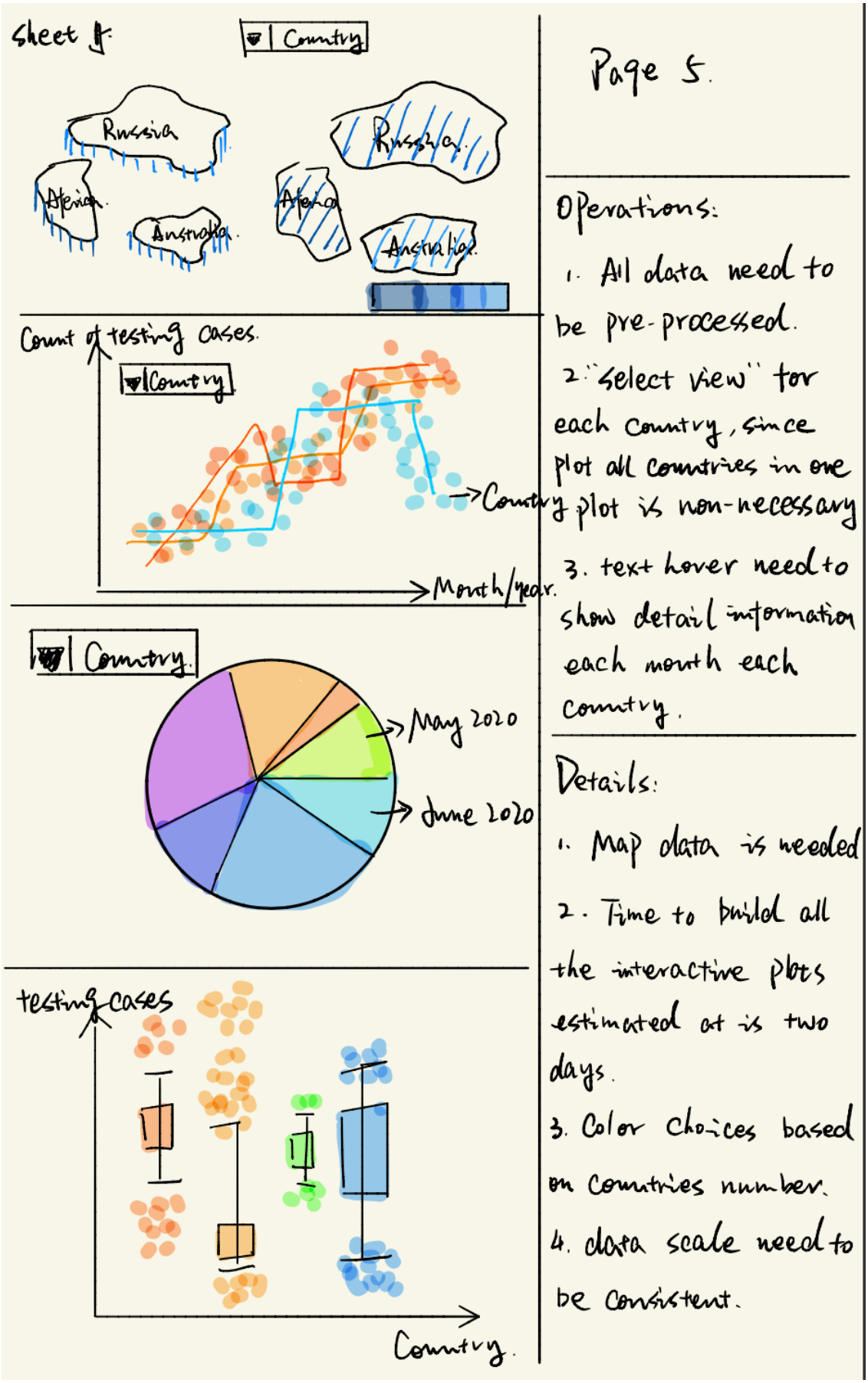
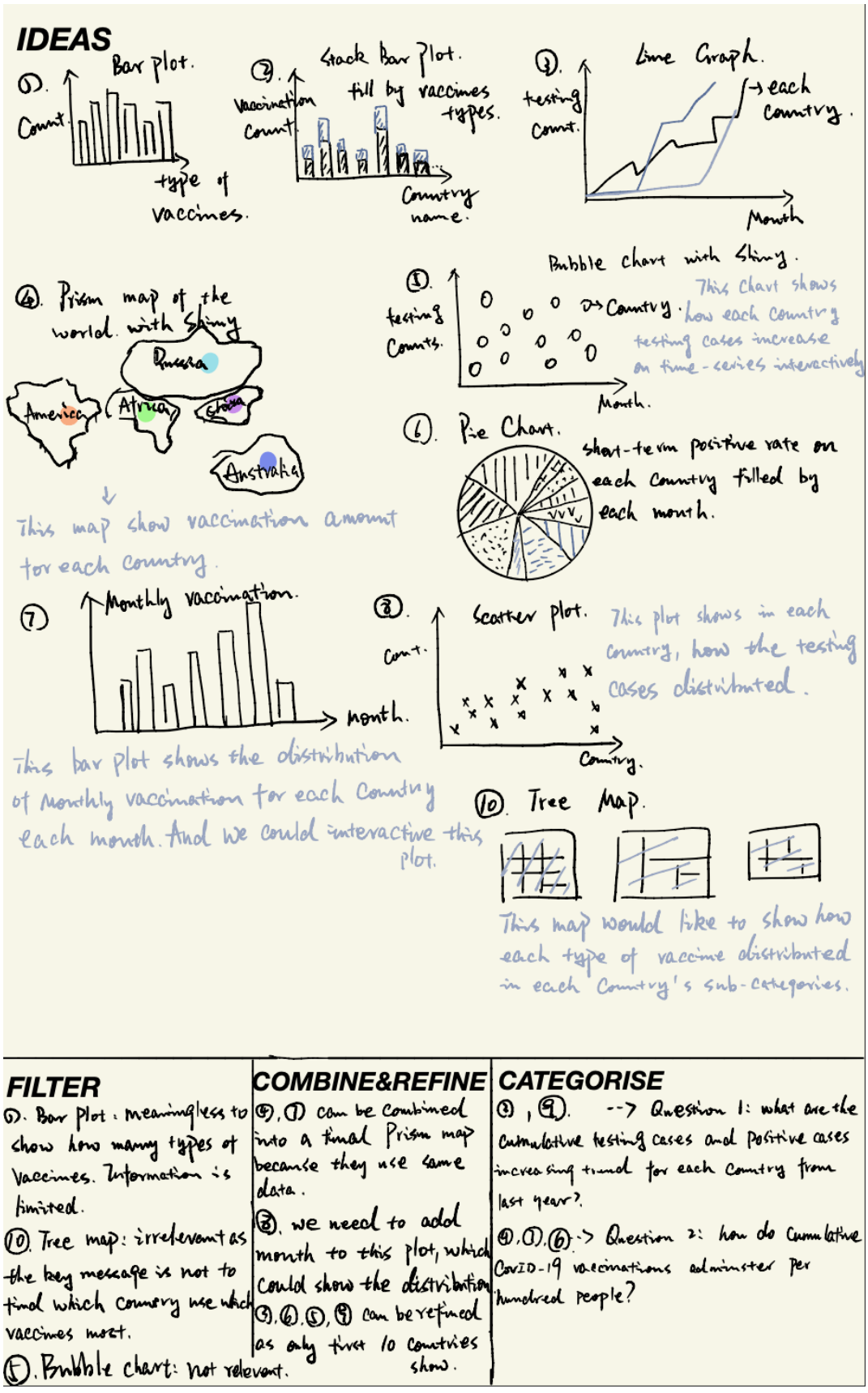

Sheet 1
Country
Russia
Australia
Count of testing cases
Country
Month/year
May 2020
June 2020
testing cases
Country
Page 5
Operations:
1. All data need to be pre-processed.
2. "select view" for each country, since plot all countries in one plot is non-necessary
3. text hover need to show detail information each month each country.
Details:
1. Map data is needed
2. Time to build all the interactive plots estimated at is two days.
3. Color choices based on countries number.
4. data scale need to be consistent.

Figure 2-1: Alternative FDS



FILTER

⑤. Bar Plot: meaningless to show how many types of vaccines. Information is limited.

⑩. Tree map: irrelevant as the key message is not to find which country use which vaccines most.

⑤. Bubble chart: not relevant.

COMBINE&REFINE

④, ⑦ can be combined into a final Prism map because they use same data.

⑧. we need to add month to this plot, which could show the distribution

③, ⑥, ⑤, ⑨ can be refined as only first 10 countries show.

CATEGORISE

③, ⑨. --> Question 1: what are the cumulative testing cases and positive cases increasing trend for each country from last year?

④, ⑦, ⑥ -> Question 2: how do cumulative Covid-19 vaccinations administer per hundred people?

Figure 2-2:

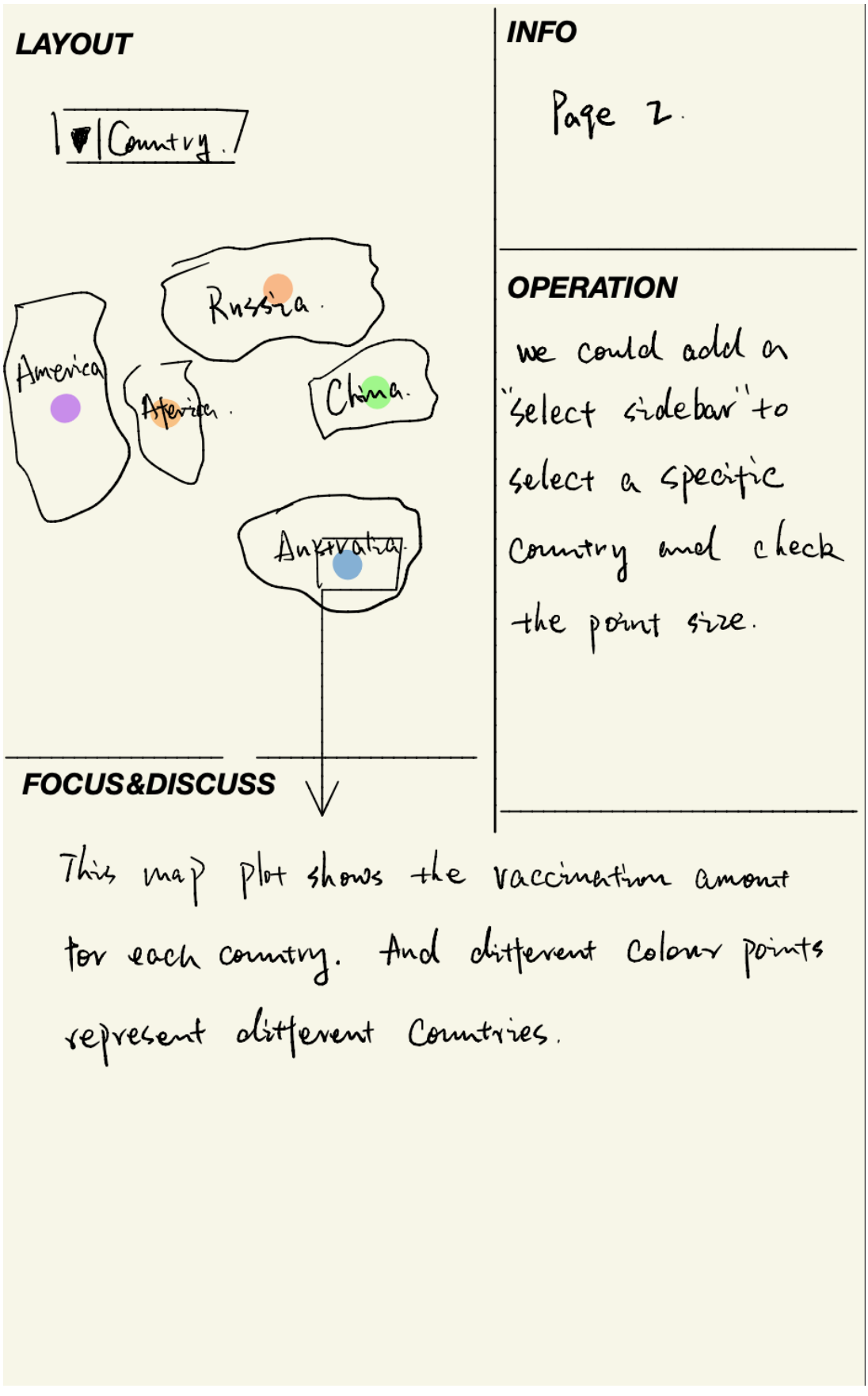

LAYOUT
Country.
Russia.
America
China.
Australia.
INFO
Page 2.
OPERATION
we could add an "select sidebar" to select a specific country and check the point size.
FOCUS&DISCUSS
This map plot shows the vaccination amount for each country. And different colour points represent different Countries.

Figure 2-3:

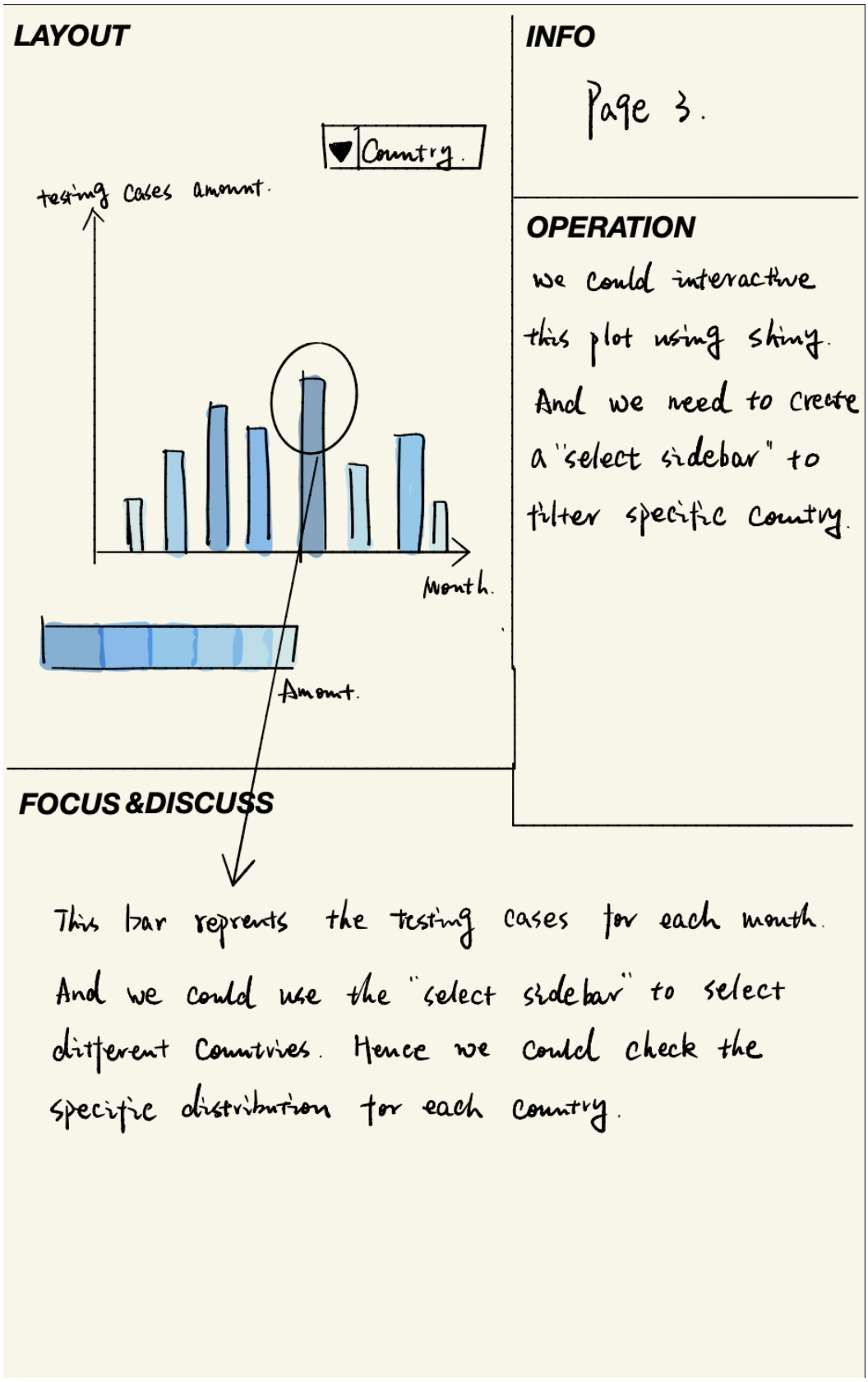

LAYOUT
INFO
Page 3.
Country.
testing cases amount.
month.
Amount.
OPERATION
we could interactive this plot using shiny.
And we need to create a "select sidebar" to filter specific country.
FOCUS &DISCUSS
This bar reprents the testing cases for each month.
And we could use the "select sidebar" to select different Countries. Hence we could check the specific distribution for each country.

Figure 2-4:

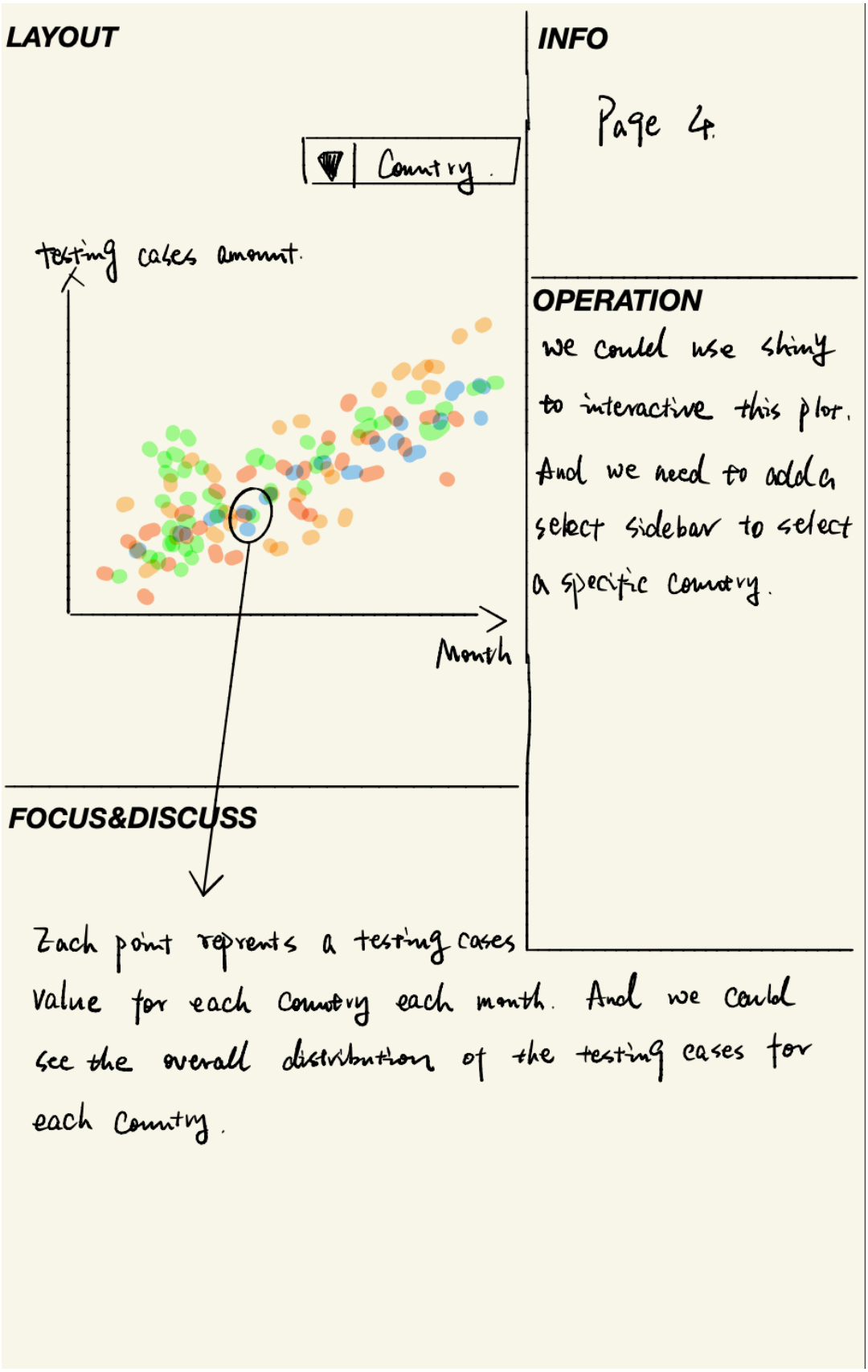

LAYOUT
INFO
Page 4.
Country
testing cases amount.
Month
OPERATION
we could use shiny to interactive this plot. And we need to add a select sidebar to select a specific country.
FOCUS&DISCUSS
Each point reprents a testing cases value for each country each month. And we could see the overall distribution of the testing cases for each country.

Figure 2-5:

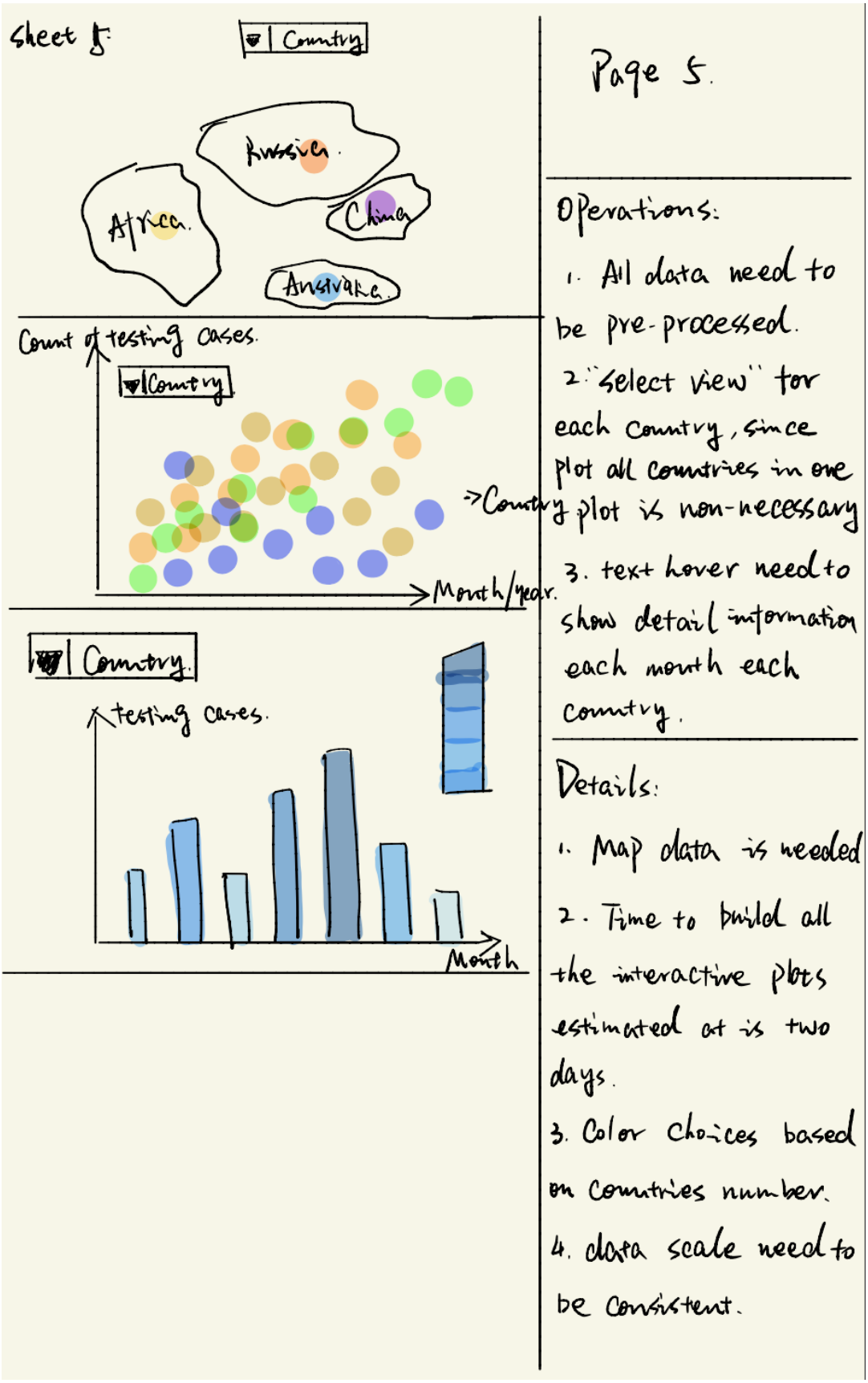
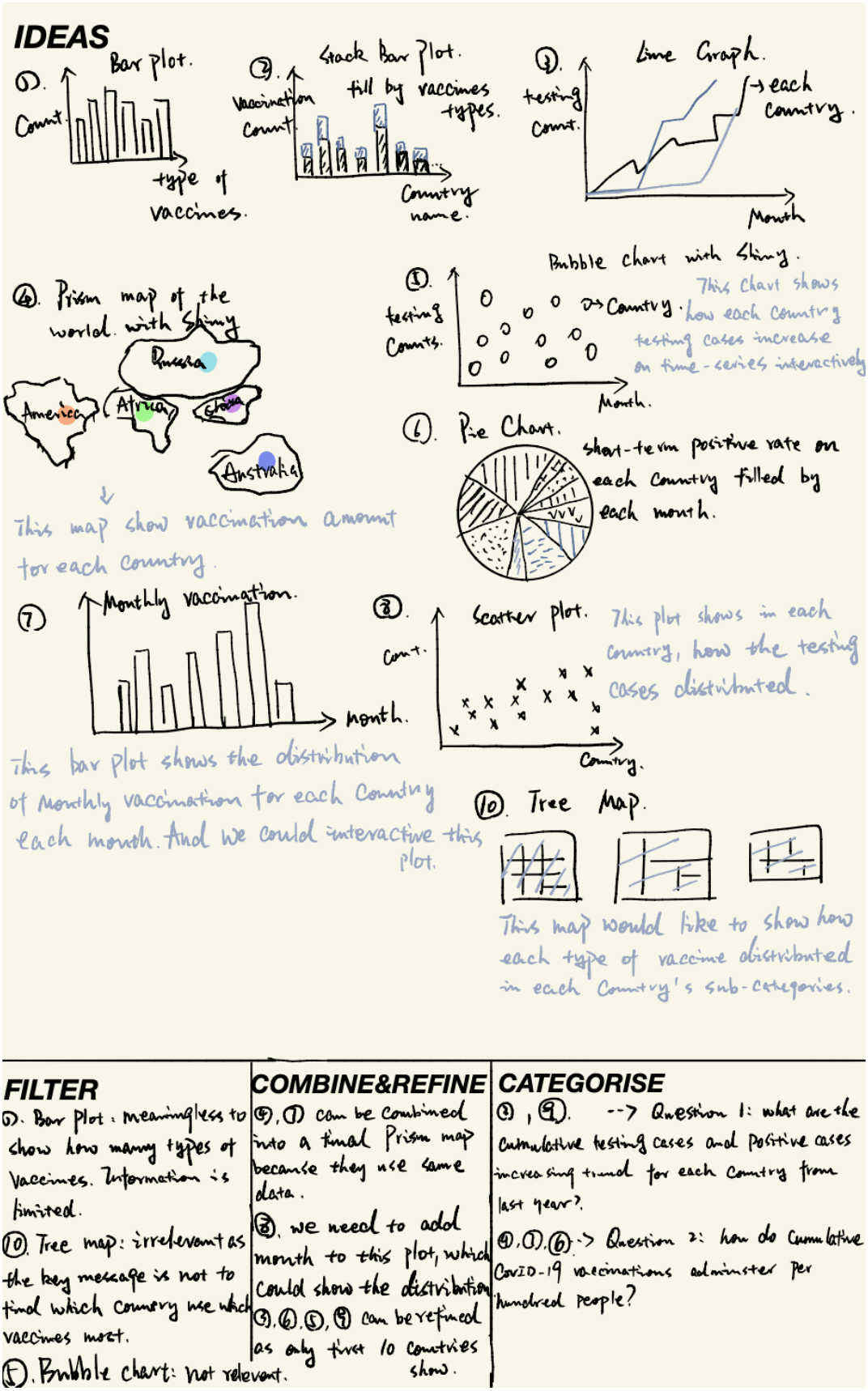

Sheet 5:
Country
Russia
Africa
China
Australia
Count of testing cases.
Country
->Country
Month/year.
Country
testing cases.
Month
Page 5.
Operations:
1. All data need to be pre-processed.
2. "select view" for each country, since plot all countries in one plot is non-necessary
3. text hover need to show detail information each month each country.
Details:
1. Map data is needed
2. Time to build all the interactive plots estimated at is two days.
3. Color choices based on countries number.
4. data scale need to be consistent.

Figure 3-1: FINAL FDS



FILTER

⑤. Bar plot: meaningless to show how many types of vaccines. Information is limited.

⑩ Tree map: irrelevant as the key message is not to find which country use which vaccines most.

⑤. Bubble chart: not relevant.

COMBINE&REFINE

④, ⑦ can be combined into a final Prism map because they use same data.

⑧. we needd to add month to this plot, which could show the distribution

③, ⑥, ⑤, ⑨ can be refined as only first 10 countries show.

CATEGORISE

①, ⑨. --> Question 1: what are the cumulative testing cases and positive cases increasing trend for each country from last year?

④, ⑦, ⑥ -> Question 2: how do Cumulative Covid-19 vaccinations administer per hundred people?

Figure 3-2:

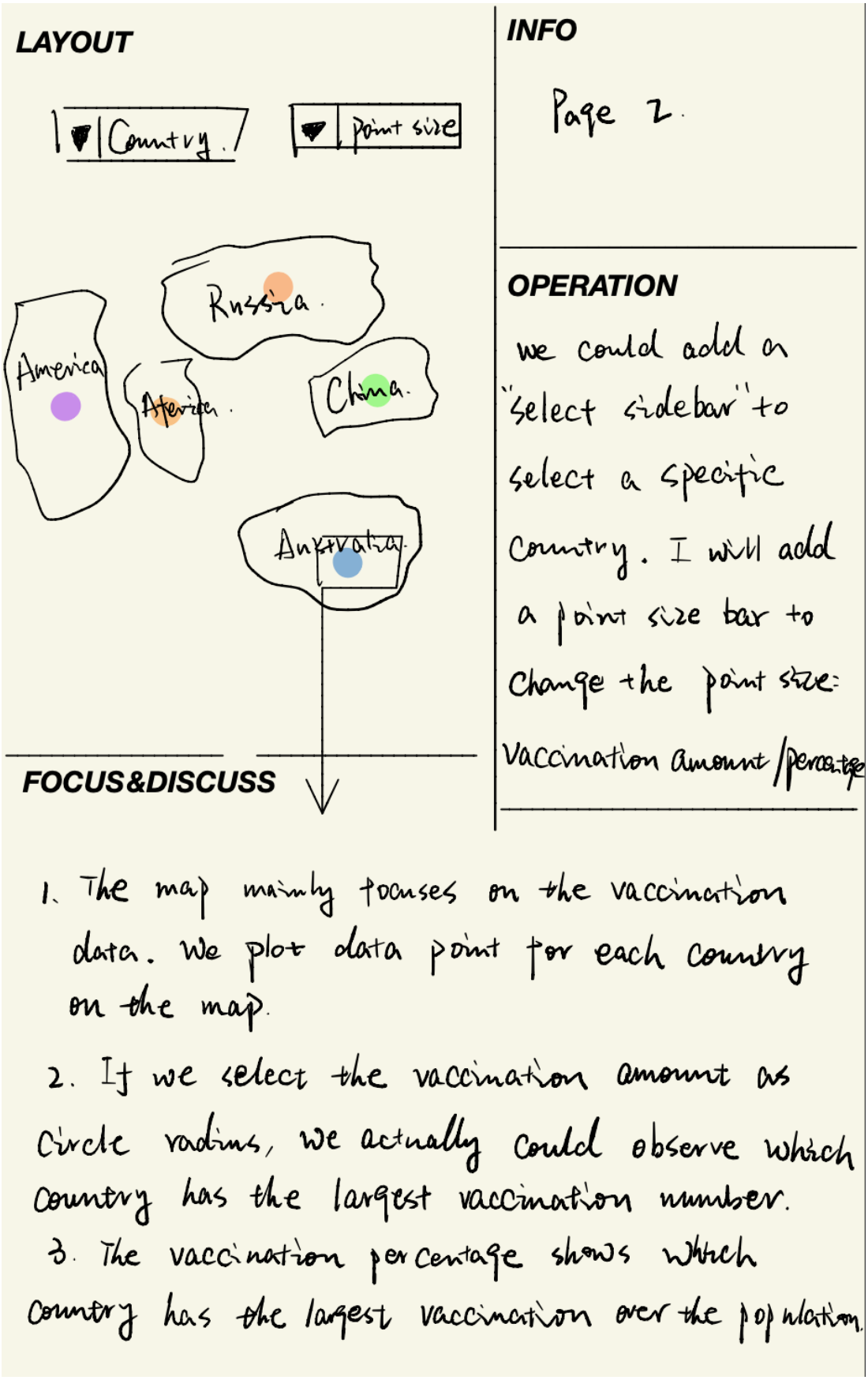

LAYOUT
Country
Point size
Russia
America
Africa
China
Australia
INFO
Page 2.
OPERATION
we could add a "select sidebar" to select a specific country. I will add a point size bar to change the point size: vaccination amount/percentage
FOCUS&DISCUSS
1. The map mainly focuses on the vaccination data. We plot data point for each country on the map.
2. If we select the vaccination amount as circle radius, we actually could observe which country has the largest vaccination number.
3. The vaccination percentage shows which country has the largest vaccination over the population.

Figure 3-3:

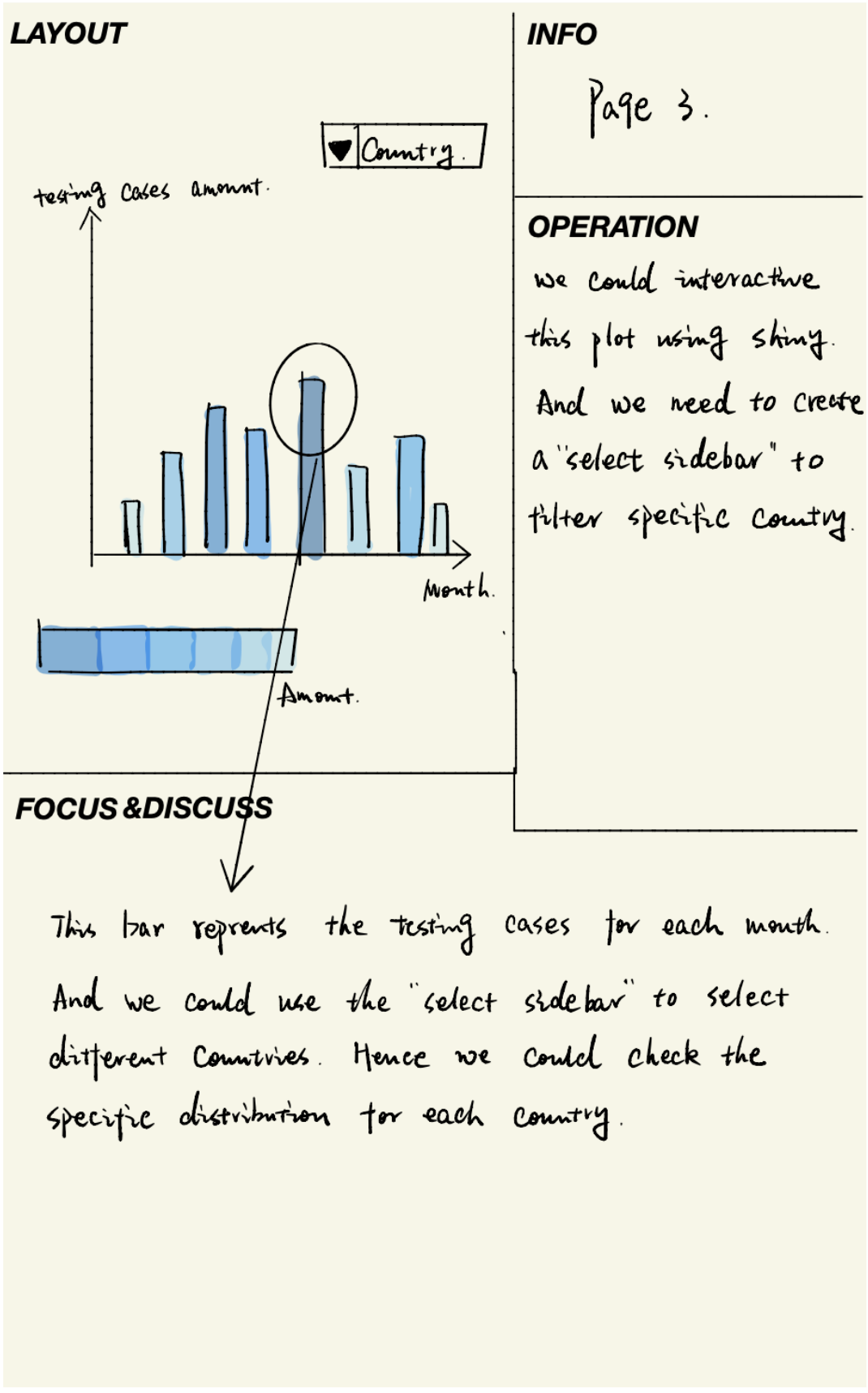

LAYOUT
Country.
testing cases amount.
Month.
Amount.
INFO
Page 3.
OPERATION
we could interactive this plot using shiny.
And we need to create a "select sidebar" to filter specific country.
FOCUS &DISCUSS
This bar reprents the testing cases for each month.
And we could use the "select sidebar" to select different Countries. Hence we could check the specific distribution for each country.

Figure 3-4:

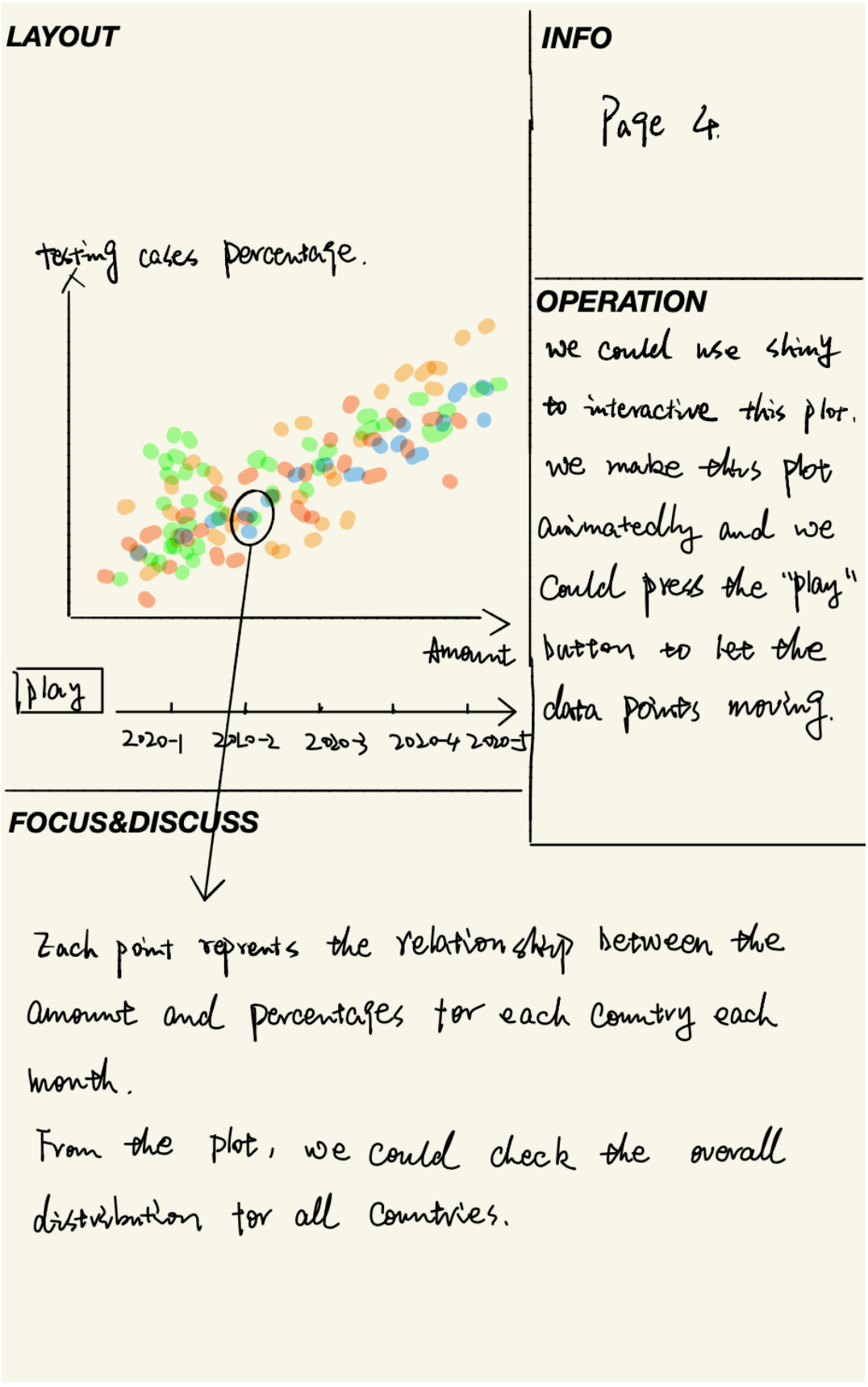
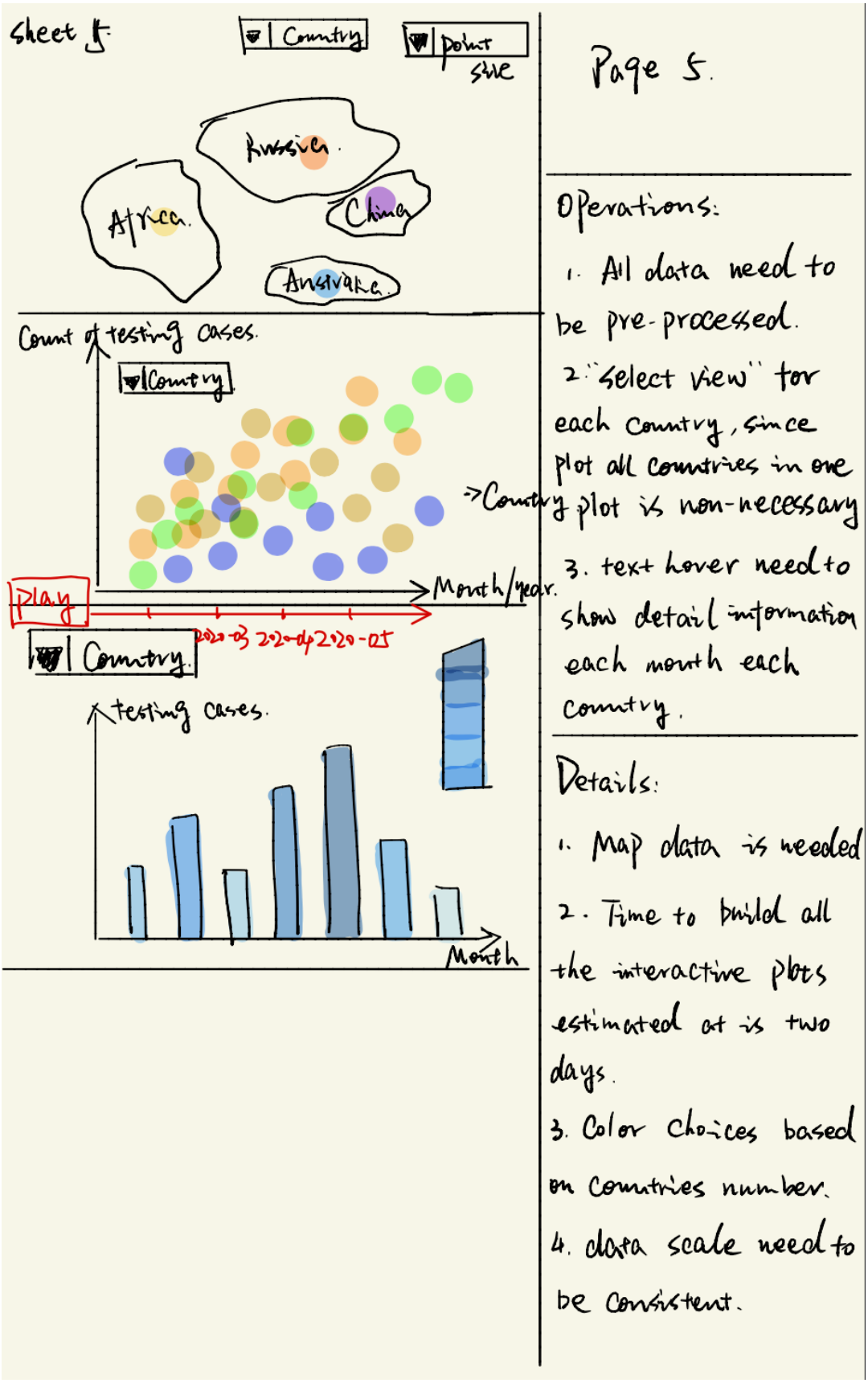

LAYOUT
INFO
Page 4.
testing cases percentage.
Amount
play
2020-1
2020-2
2020-3
2020-4
2020-5
OPERATION
we could use shiny to interactive this plot. we make this plot animatedly and we could press the "play" button to let the data points moving.
FOCUS&DISCUSS
Each point represents the relationship between the amount and percentages for each country each month.
From the plot, we could check the overall distribution for all countries.

Figure 3-5:


Sheet 5
Country
point size
Russia
Africa
China
Australia
Count of testing cases
Country
Month/year
play
2020-03 2020-04 2020-05
Country
testing cases
Month
Page 5.
Operations:
1. All data need to be pre-processed.
2. "select view" for each country, since plot all countries in one plot is non-necessary
3. text hover need to show detail information each month each country.
Details:
1. Map data is needed
2. Time to build all the interactive plots estimated at is two days.
3. Color choices based on countries number.
4. data scale need to be consistent.