

Before Peer Review

GR5063 Final Project Process Book

Idea Explorations.

Topic : Cost Benefit Analysis on Global Tobacco Policy .

Time Chosen : 2008 - 2018 (For every 2 yrs)

Period for Data Exploration : 2008 - 2018 annually.

(Since some datasets are useful but have too many missing values btw 2008-2018) ➔ Add one more "abt US"

Topic Name	Background, Why this Topic, Data Descriptions, etc.
☰ Overview	↳ Interactive Visualizations for each dataset used .
☰ EDAs	↳ Results from Modeling, all datasets are incorporated together in a meaningful way.
☰ Cost Benefit Analysis .	↳ Public Sentiment change by yr towards tobacco control policy.
☰ Sentiment Analysis	
☰ Conclusions	
☰ Appendix	

Tab About Us-

About Us.

Team Intro

Main Objective

Why this topic, what we want to investigate

Contact Us.

Tab Introduction

How to USE THIS Dashboard.

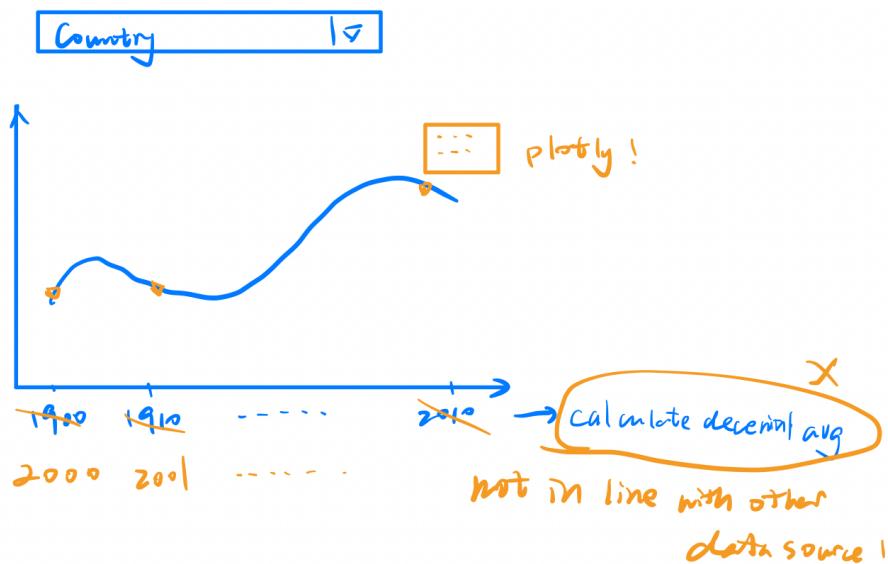
Tobacco Background.

- -----
- -----
- | -----

{ Descriptions on
each Tab and
what users expect
to encounter.

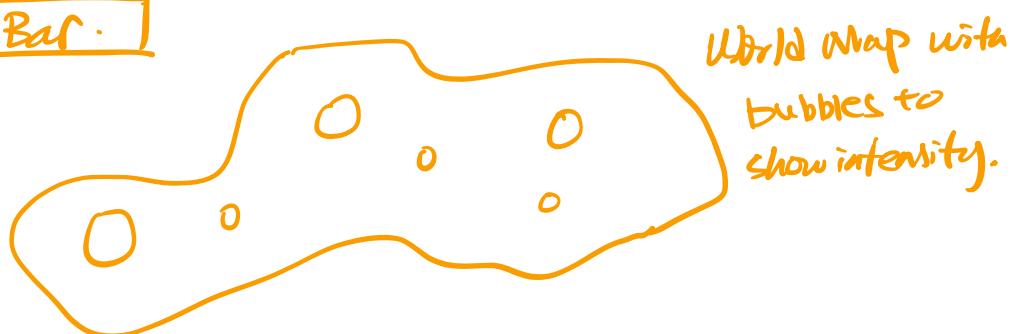
Tab Tobacco Background

Sub Tab 1: Consumption - Cigarette



Sub Tab 2: Consumption - Tobacco Usage Penetration

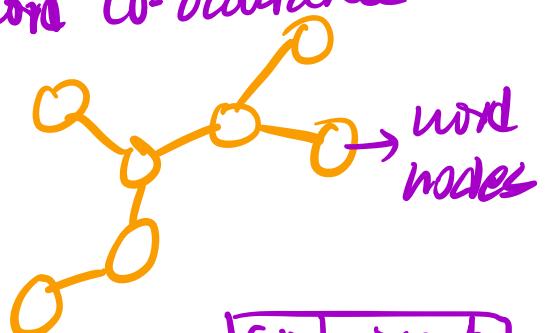
Slider Bar



Sub Tab 3: Death Rate



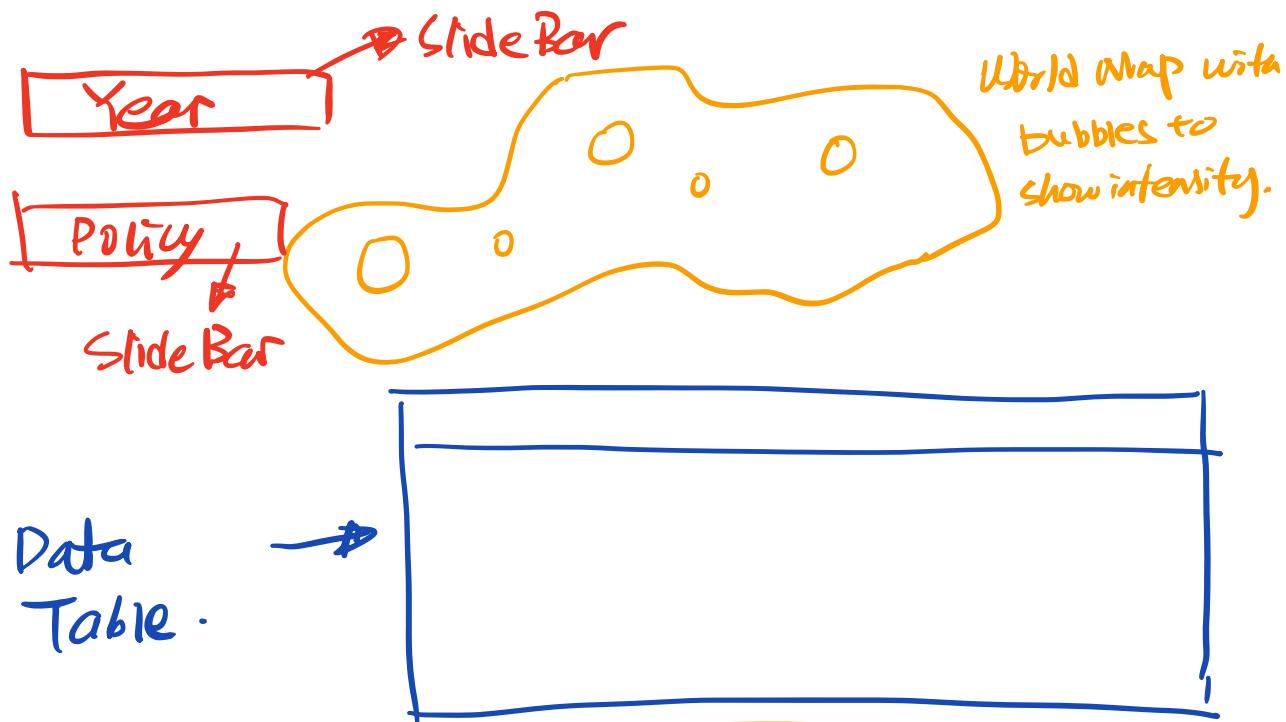
word co-occurrence



Slide Input

Tab Global Tobacco Control Policy.

Descriptions of Different Policies.

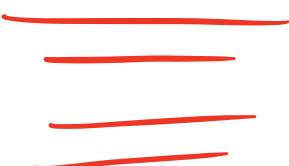


Tab Cost Benefit Analysis.

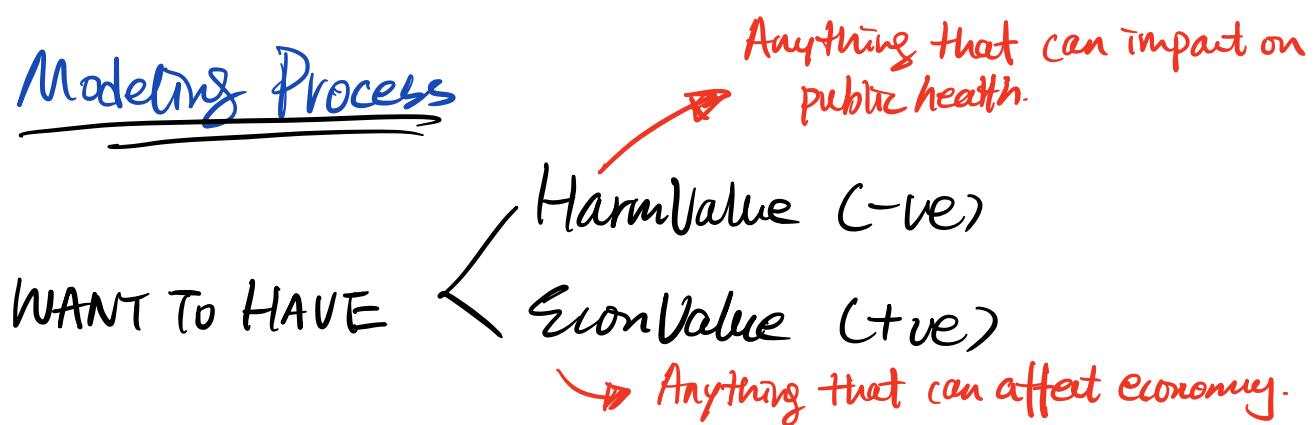
Brief Modeling Process



Results Interpretations



Modeling Process



WHAT WE HAVE NOW.

① Policy Dataset (2008, 10, 12, 14, 16)
(2008, 10, 12, 14, 16)

- ↗ Help To Quit
- ↗ Ads Ban
- ↗ Taxes (\$)
- ↘ Price PPP(\$)

② Death Rate → # of deaths per 100K people.
(2008, 10, 12, 14, 16, 18)

③ Tobacco Use → % of Use by Country by yr.

(2010, 14, 16, 18)

(need to approximate 18 using
2005, & 2013 data)

↗ Missing Data filled by
imputations

④ Sales Per Day → # of cigar consumption
per person per day.
(2008, 10, 12, 14, 16)

(need to approximate 18
using 2017 data)

↗ Missing Data filled by
imputations

Modeling Process

EconValue : Tax, Price, Ads Ban

$$\hookrightarrow = \log_e (\underbrace{\text{Tax Rev} + 1}_{\substack{\rightarrow \text{Tax}(\%) \times \text{Price} \\ \text{PPP}(\$)}}) + \frac{1}{\text{Ads Ban}}$$

so that it follows marginal diminishing returns

adjusted by purchasing power parity

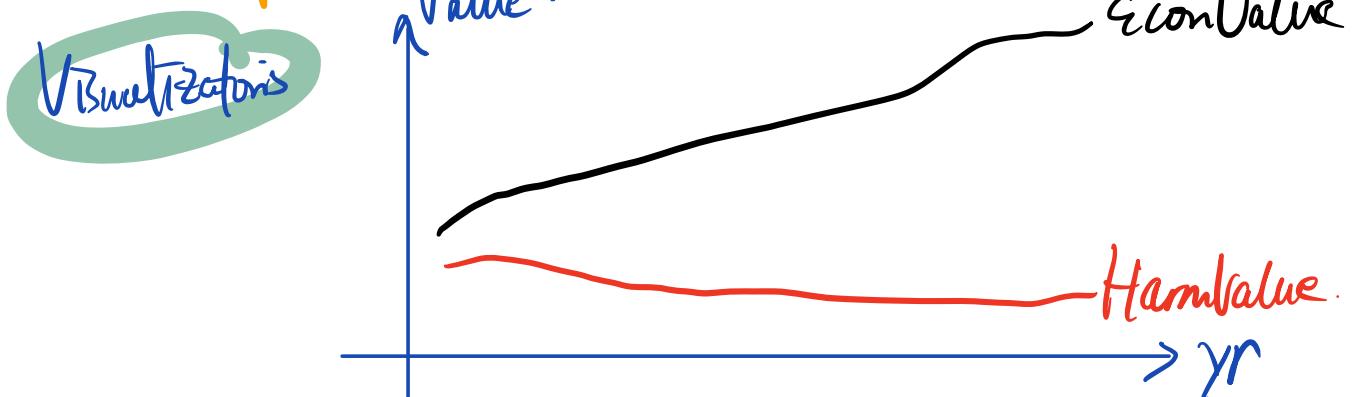
$+ 1$ to make sure the whole term is always greater than 0.

HarmValue : DeathRate, HelpToQuit, Tobacco Use, Sales

$$\hookrightarrow = (\text{Scale Factor} \times \text{Death Rate})^{\text{Power Factor} (>1)} + \text{Scale Factor} \times \text{Tobacco Use (prevalence)}$$

+ Sales (more ppl buy, more harmful to their health)

HelpToQuit \leftarrow Enforcement level \rightarrow + harms



Finalised:

adjusted by PPP(\$)

$$\text{EconValue} = \log_e(\text{tax}(\%) \times \text{price} + 1) + \frac{1}{\text{AdsPer}}$$

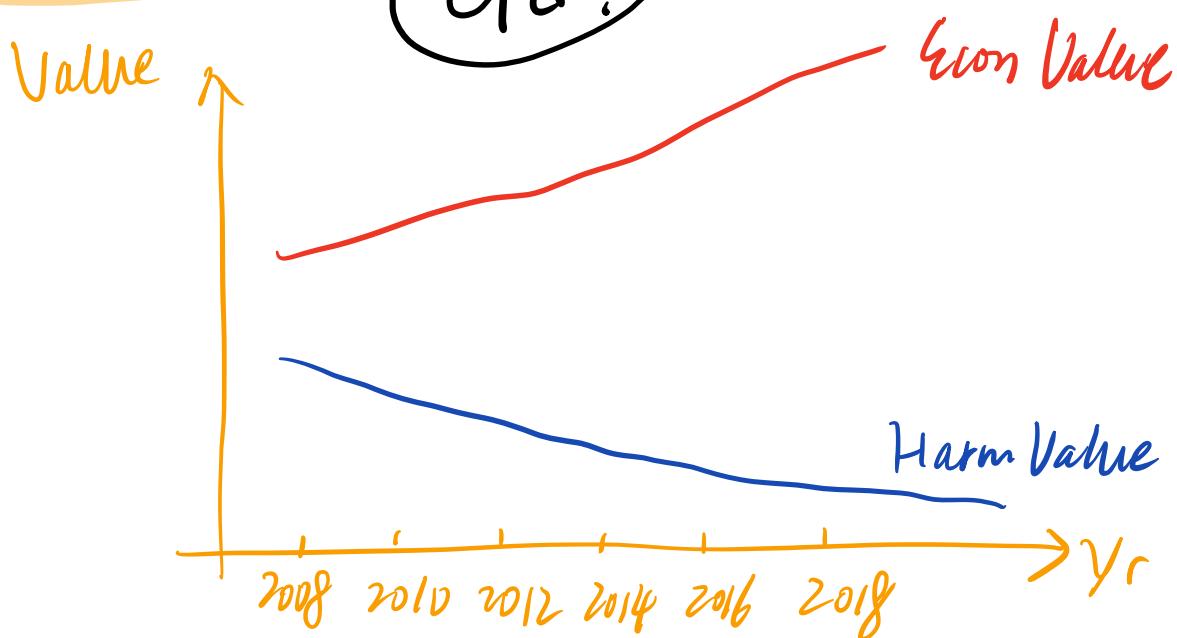
$$\text{HarmValue} = (10^{-5} \times \text{DeathRate} + 1)^{1/3}$$

- $\frac{\max(0, \text{HelpToQuit} - 2)}{2}$

Modeled in this way to make sure EconValue & HarmValue has comparable range

Current Result.

Gd!

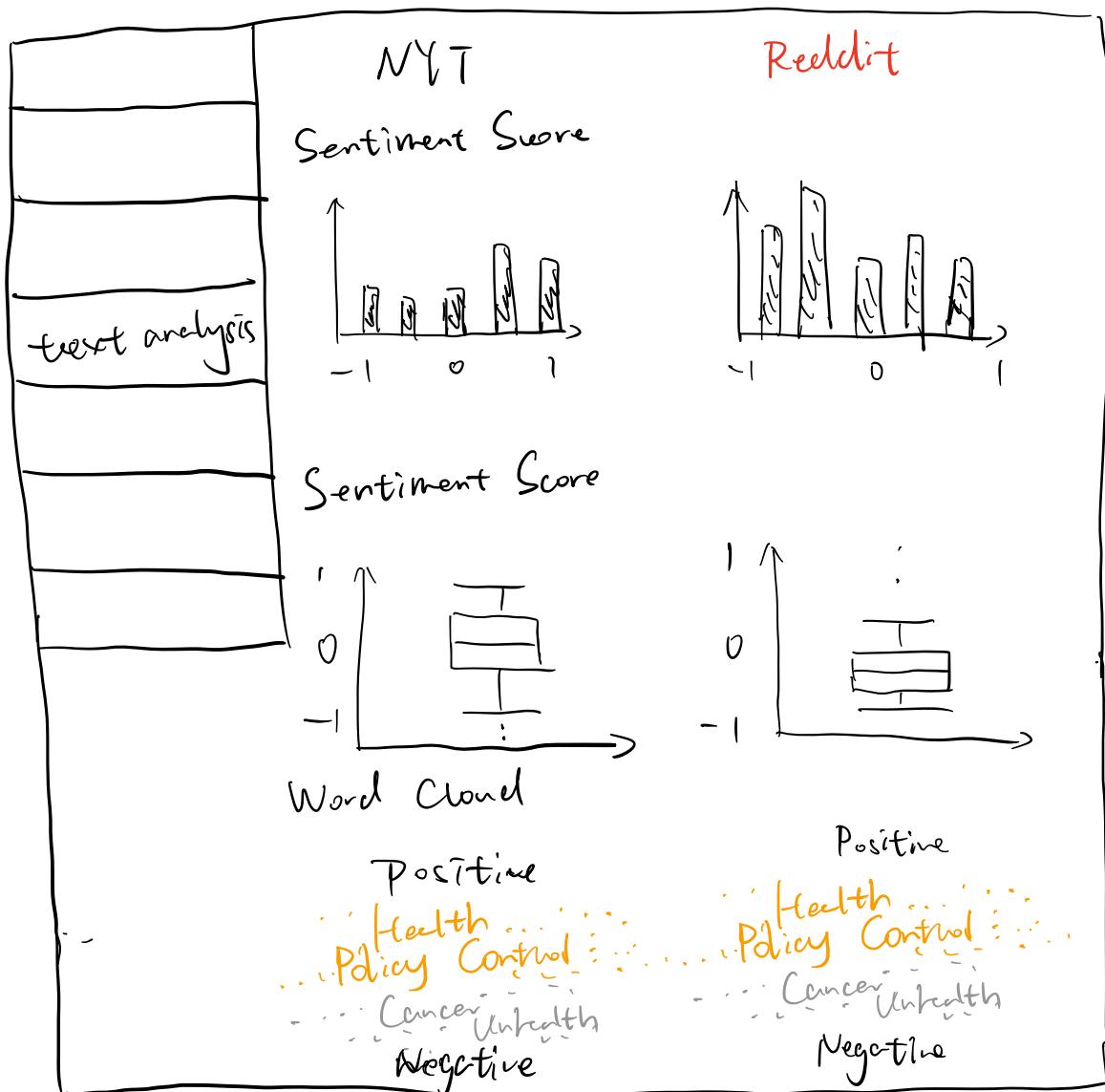


Text analysis from NYT API & Reddit

tone analysis (bar plot) - sentiment score

tone analysis (box plot) - sentiment score

word cloud - comparison cloud

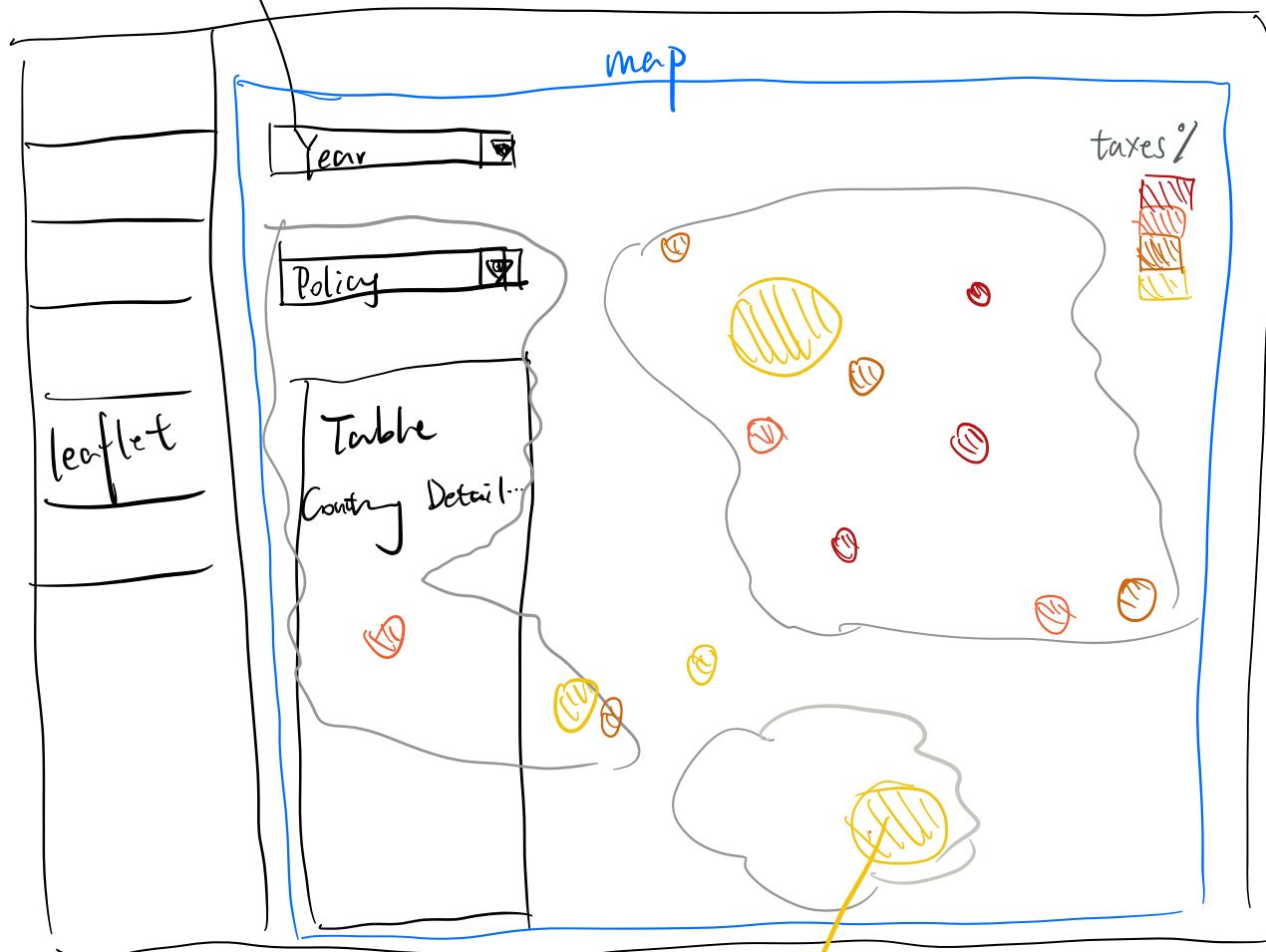


Worldwide Tobacco Control Policy

leaflet map

{ tobacco price (adjusted) \$
tobacco tax %
Help to Quit
Advertisement Policy

drop box



color: tax%

size: price \$

Tab Conclusion

key findings.

(Graph?)

Tab Appendix

Data Sources that we used
(with brief descriptions)

After Receiving Comments

Comments:

1. Add description to each tab

2. Description to each graph

takeaway

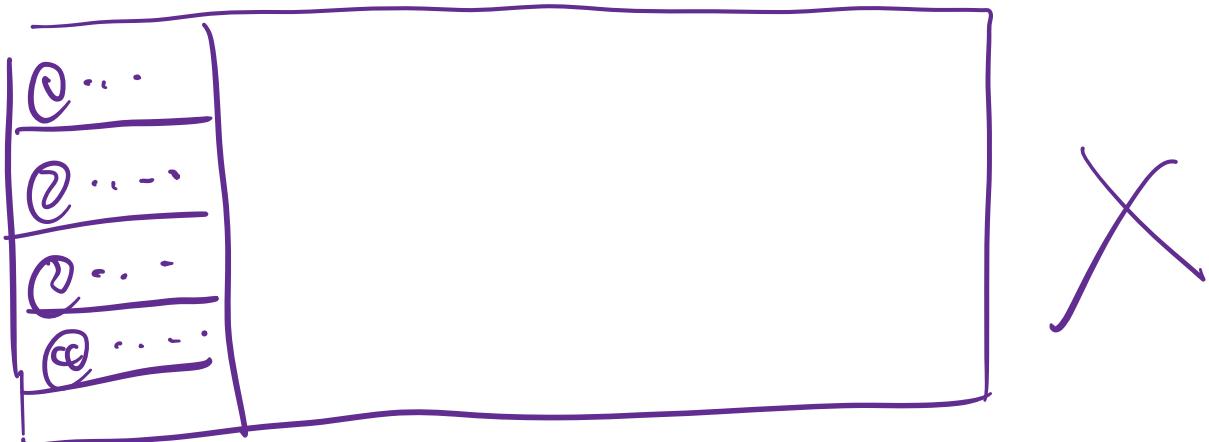
3. Penetration map: bubble too
small, missing
values
4. word cloud: positive /negative
5. control map { pop up title
| policy dropdown
- table → interactive table
6. text : details



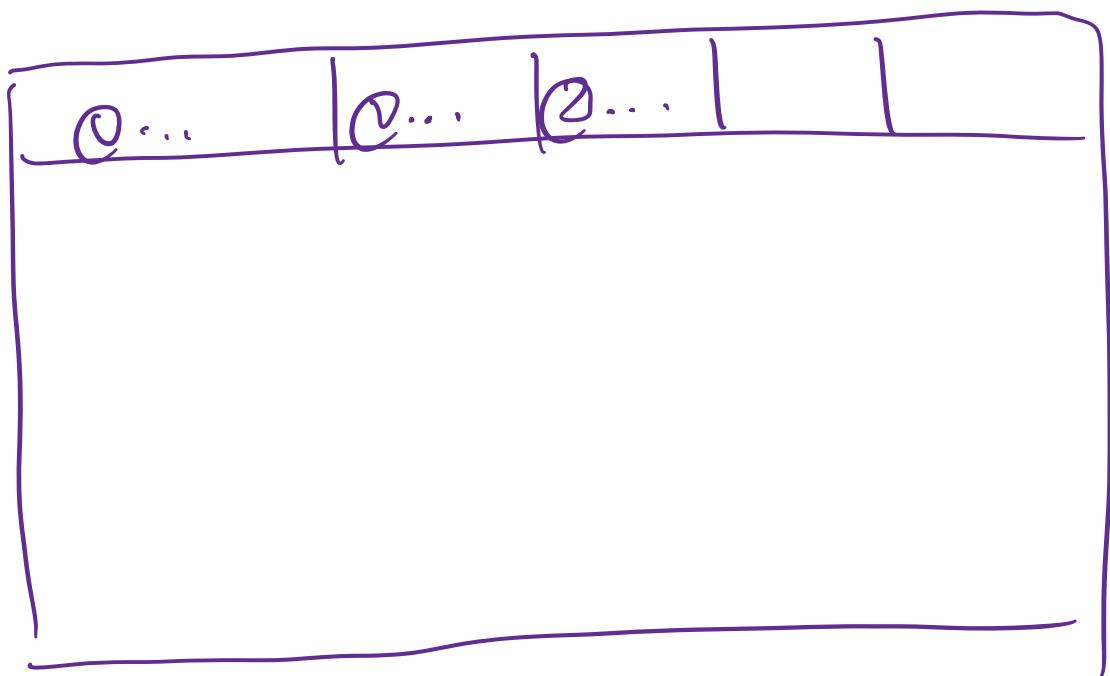
We decided to :

- ① change the overall layout
- ② change the format of penetration map (background tab)
- ③ change the layout of control policy tab
 - + add line plot
 - + change to interactive data table
- ④ text analysis
 - + add negative & positive title
 - + detailed information of the texts
 - + network analysis

① overall layout



⇓ tab layout



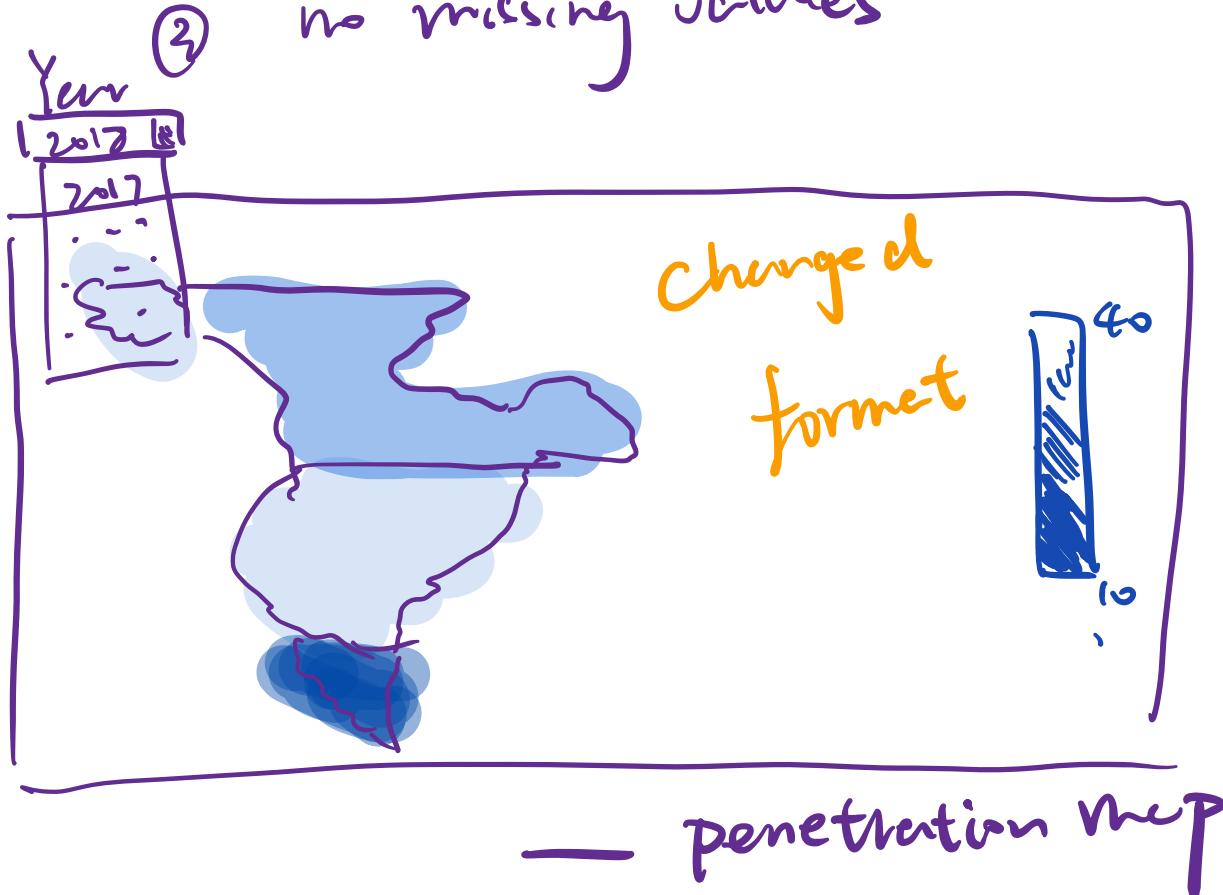
② tobacco background - penetration map

leaflet map → ggplot map

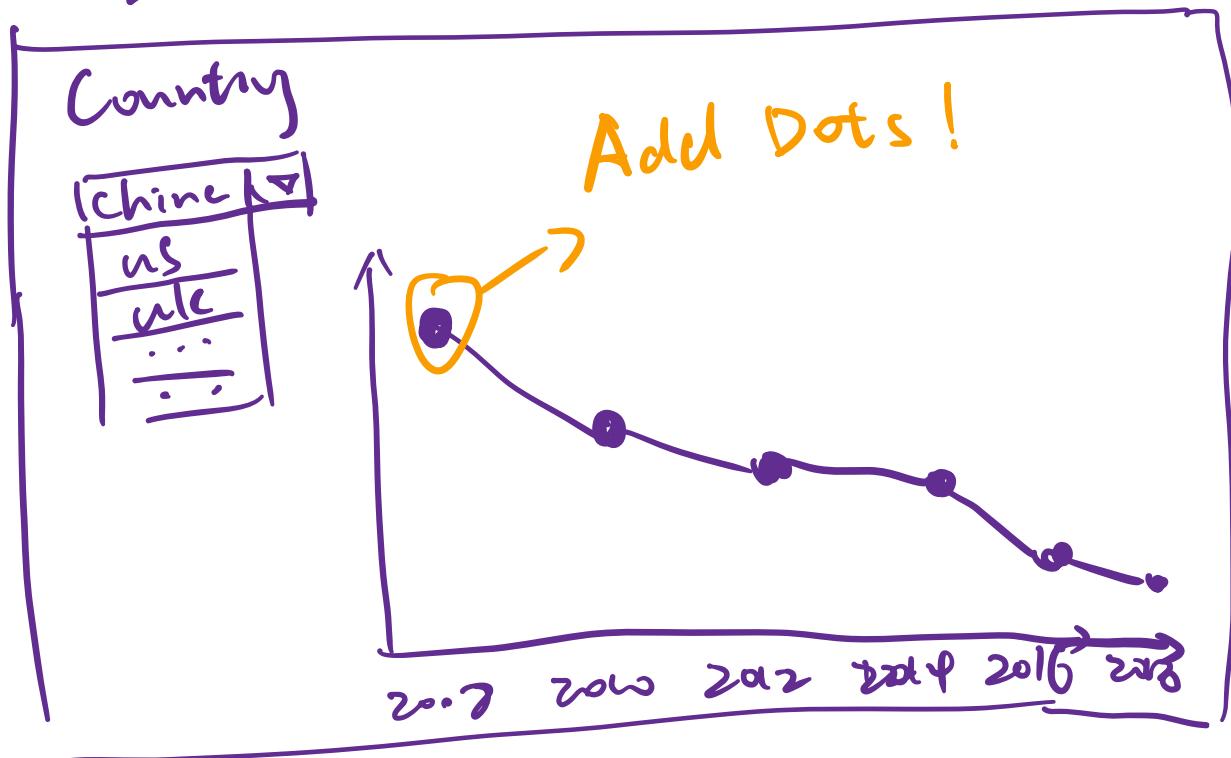
① more direct map

② colors means the penetration

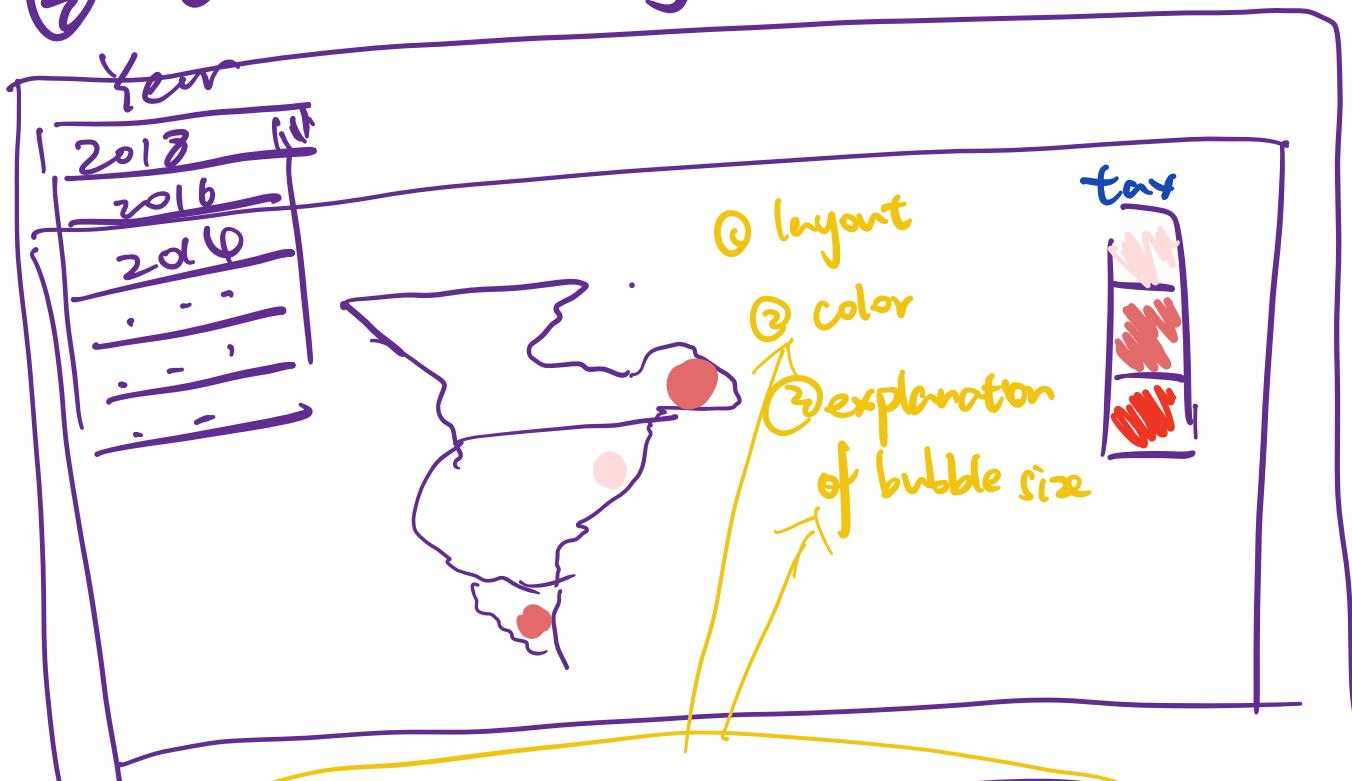
③ no missing values



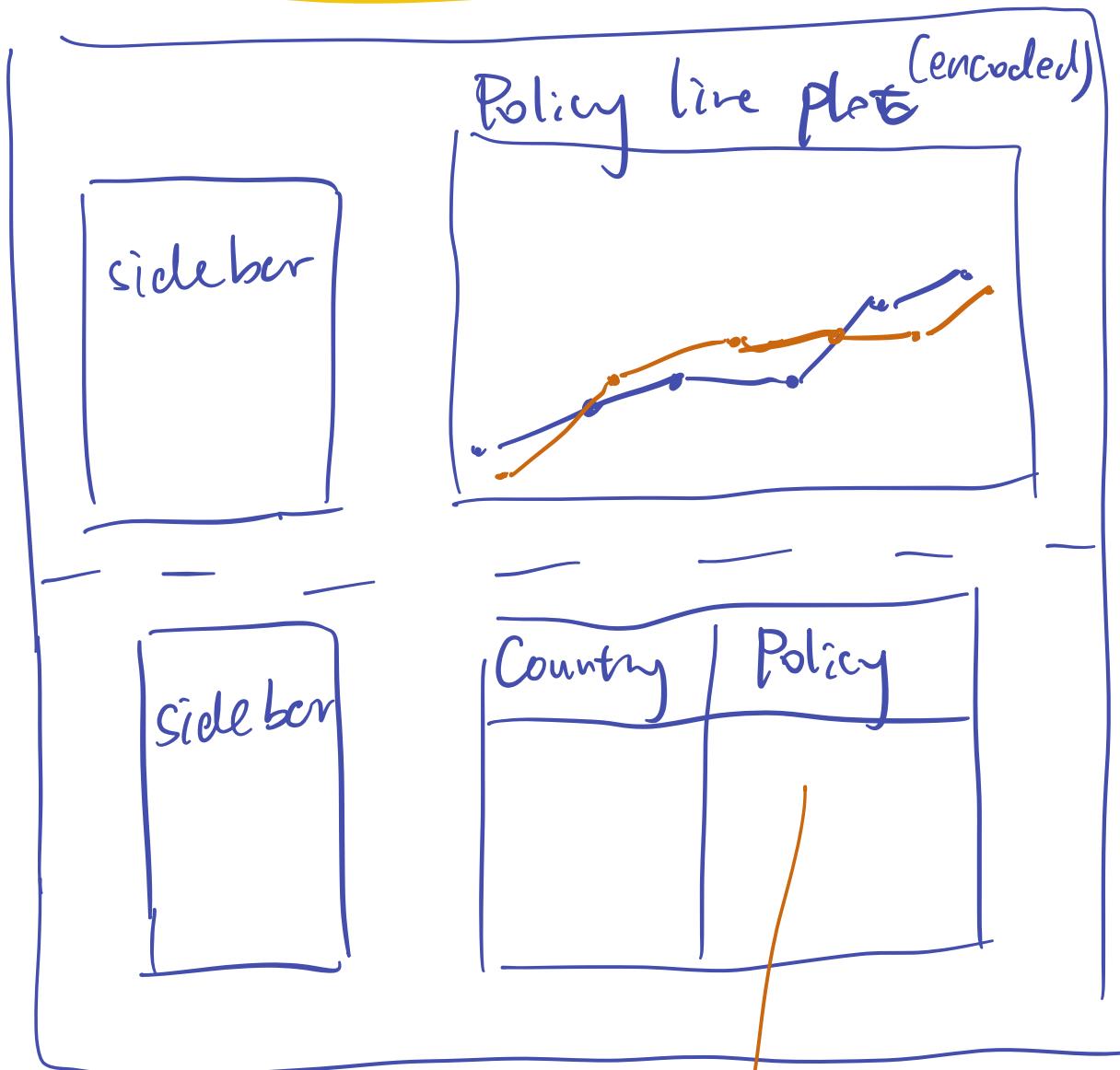
- Death Rate



② Control Policy tab



~~Hi~~ the bubble size means the price,
color means price



④ interactive table output

④ text analysis

the NYTimes : API 500+

① Add detailed articles
information 2008 - 2018

the Reddit : web scraped
of text sources (10+

2008 - 2018

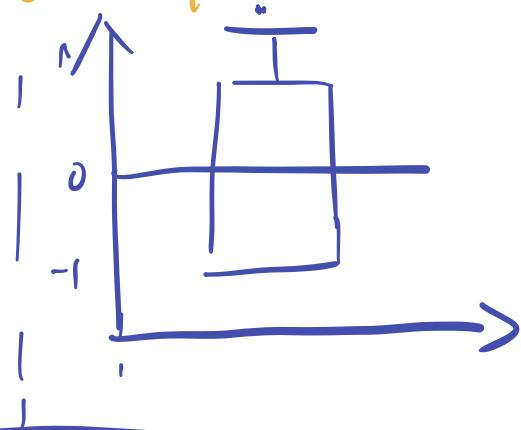
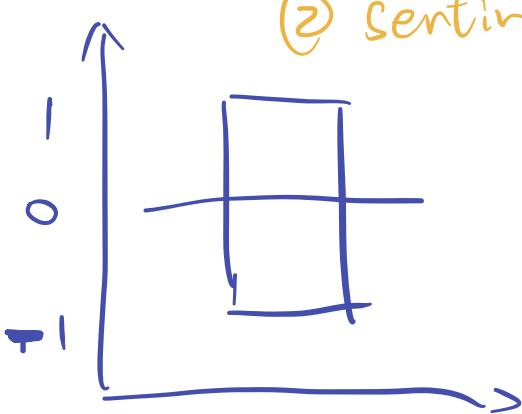
—
New York Times

—
Reddit

Sentiment box

WordCloud Network

② sentiment box ph.



② Add positive & negative

