

al-gore-ithms

CS1951A Final Project

[David Moon, Livia Gimenes, Raymond Dai, Truman Cunningham]



The Goal

All Talk?

Is there a relationship between how much a U.S. President talks about a *certain political topic and how much action they take in regards to that topic?



* Our chosen topic was healthcare

Our Datasources



Speeches

Web scraped annual State of the Union Addresses from The American Presidency Project (UC Santa Barbara)

Budget

Downloaded historical budget data (from 1976 - 2020) from Office of Management and Budget

Our Datasets

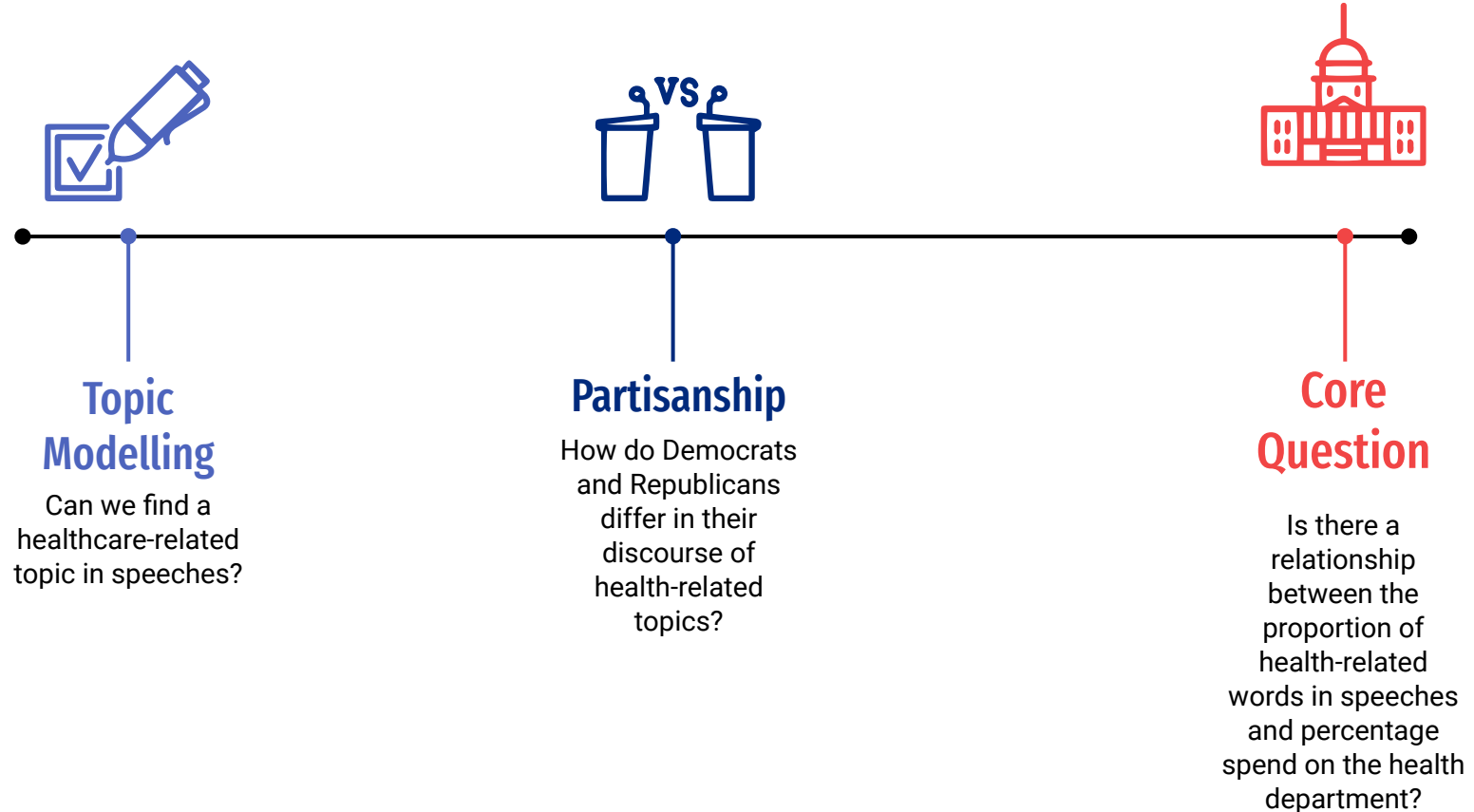
Despite having a State of the Union Address (SOTUA) for every President, we were limited by the **lack of budget data**. Here are some characteristics of our datasets...

- 8 Presidents included (from Jimmy Carter onwards) with 4 or 8 speeches each
- Annual data on percent budget spend on Department of Health and Human Services
- Dictionary of health-related words

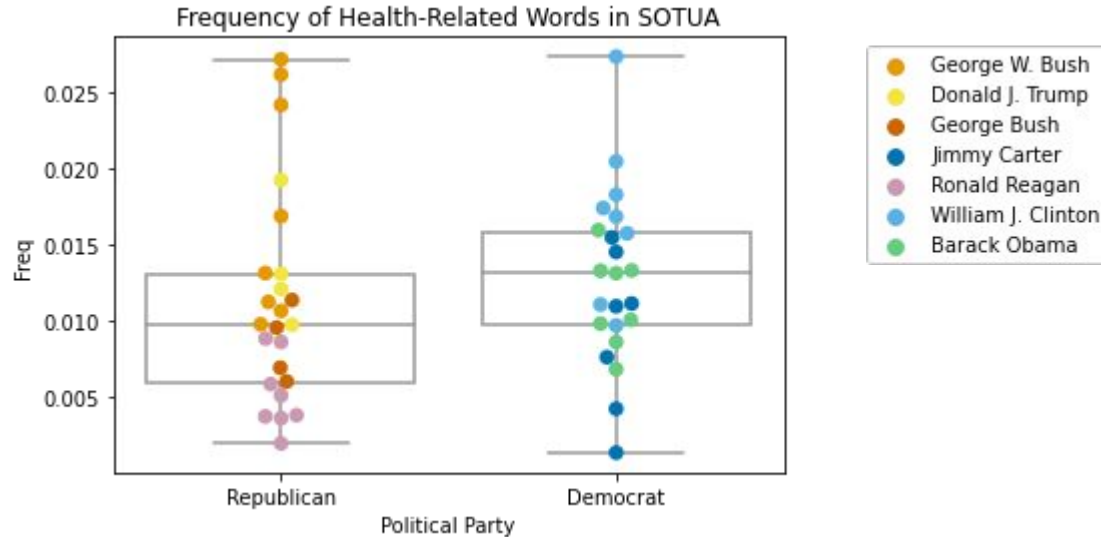
Here is how we prepared our data for analysis...

- Added column for partisanship in speeches dataset
- Tokenized, lemmatized, and removed stop words from speeches
- Split data into 80-20 for train-test, with each President having (at least) 1 speech in test
- Read all of the SOTUA in our dataset to make sure our dictionary of health-related words representative of the speeches.

Leading Up to Our Core Question



Partisanship



Democrats, on average, talk more about health than republicans. However, we are unable to run something like a t-test because frequencies are **not** normally distributed

Partisanship

Because speeches for each president are **not independent** and our distributions are **not normal**, we run Friedman's test (non-parametric two analysis of variance by ranks). Some important notes...

- Groups are represented by which speech a President gives (1st, 2nd, 3rd, or 4th)
- The dependent variable is the proportion of health-related words in a given speech

This null and alternative hypothesis of this test are as follows...

H_0 : there is no difference in the proportion of health-related words (for Republicans and Democrats) between a president's 1st, 2nd, 3rd, and 4th SOTUA

H_a : there is a significant difference in the proportion of health-related words (for Republicans and Democrats) between a president's 1st, 2nd, 3rd, and 4th SOTUA

The results of this test ($\alpha = 0.05$) are as follows...

For **Republicans**, with a p-value of 0.007263, we reject the null hypothesis.

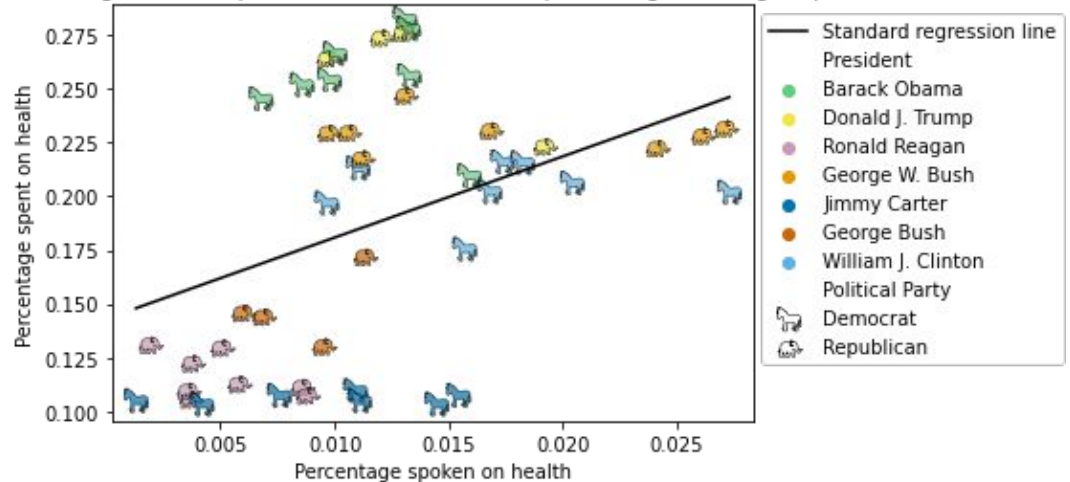
For **Democrats**, with a p-value of 0.0159, we reject the null hypothesis.

Core Research Question

Yes, there does seem to be a **positive correlation** between budget spent on the health department and time spent on health-related topics in SOTUAs...

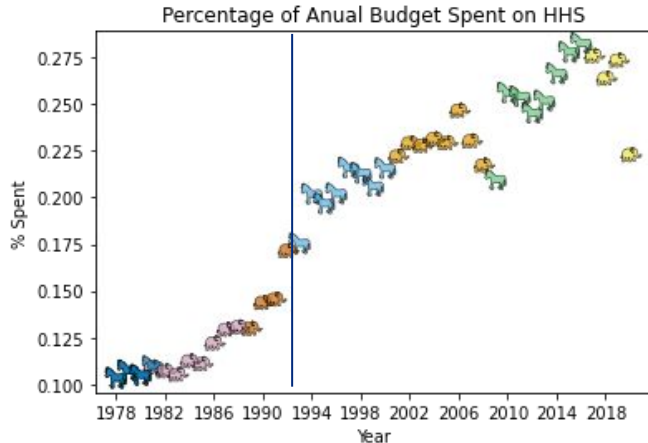
- P-value of **0.004**
- MSE (mean squared error) was 0.002 -> relatively high due to high variance
- 95% confident of a coeff between **1.52 and 7.62**
- As shown in graph to right, distinct clusters may lead to Simpson's paradox*

Percentage of time spoken on health issues vs percentage of budget spent



*Simpson's paradox: when a trend appears in several groups of data but disappears or reverses when the groups are combined.

Pre -> Post Clinton Era Split



Shift in Rhetoric

Clinton's campaign against Bush championed the Health Securities act, one of the **first major pushes for health care reform** in the modern United States

Shift in Budget Priorities

The Clinton Era saw a **massive spike in HHS funding** - the largest acceleration in funding in our time frame

Shift in the Status Quo

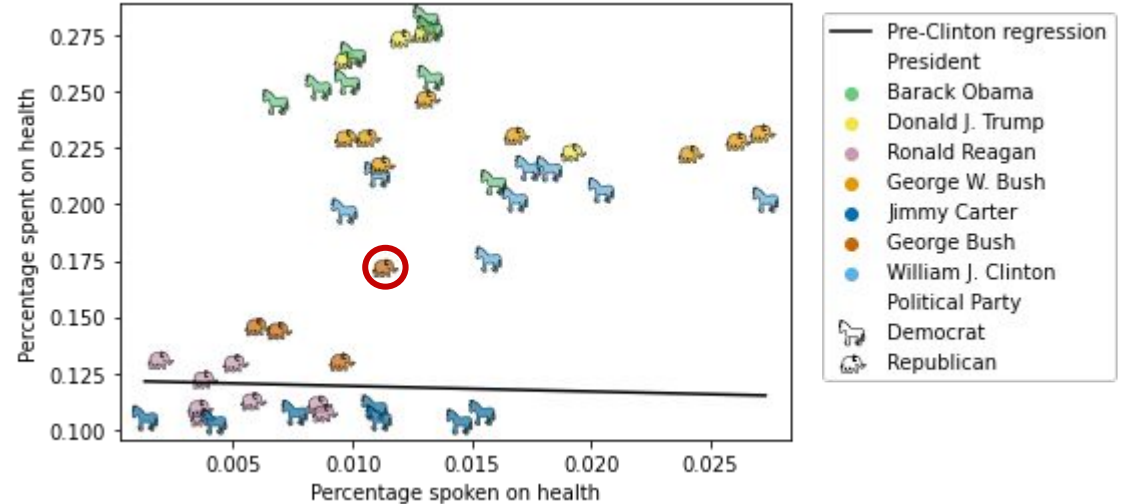
Health care costs grow at double the rate of inflation, making healthcare inaccessible to many. This pushed some of the burden onto the state

Core Research Question - Pre Clinton Era

No, there is not a significant relationship between spending and time spent on health-related topics before Clinton...

- **P-value of 0.603** with a coeff of -0.5659. MSE of was about 0.0008
- High error might be caused by Bush's last year (circled) - right before his campaign against Clinton

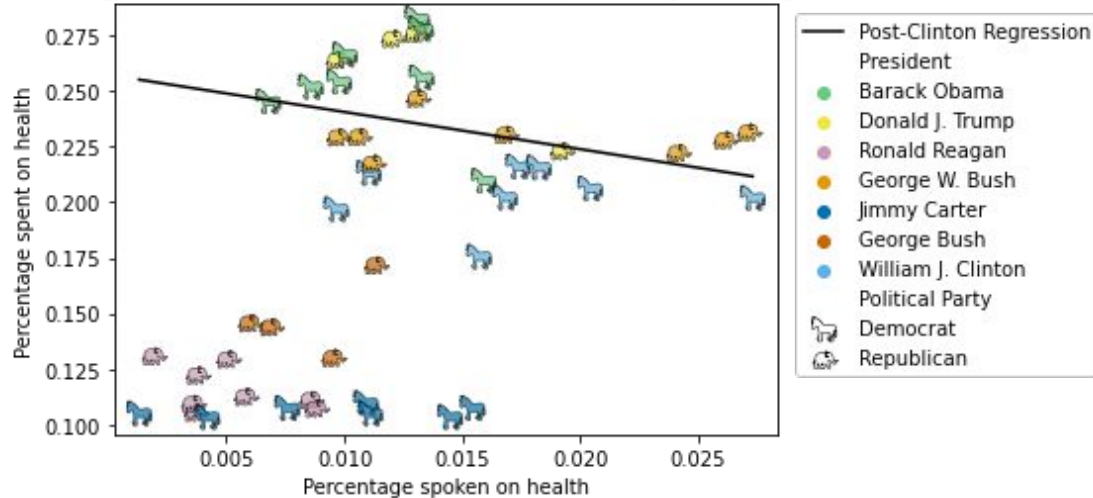
Percentage of time spoken on health issues vs percentage of budget spent



Why the lack of significance? Before the Clinton push we see generally less time spent on healthcare and lower spending; this suggests this was a lower national priority. This would reduce the chances that presidents really watched how much time they spent on the issue. There are also significantly less data points to work with

Core Research Question - Post Clinton Era

Percentage of time spoken on health issues vs percentage of budget spent



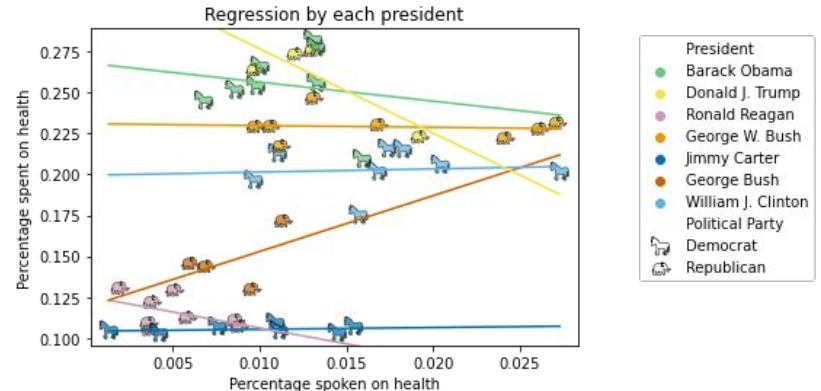
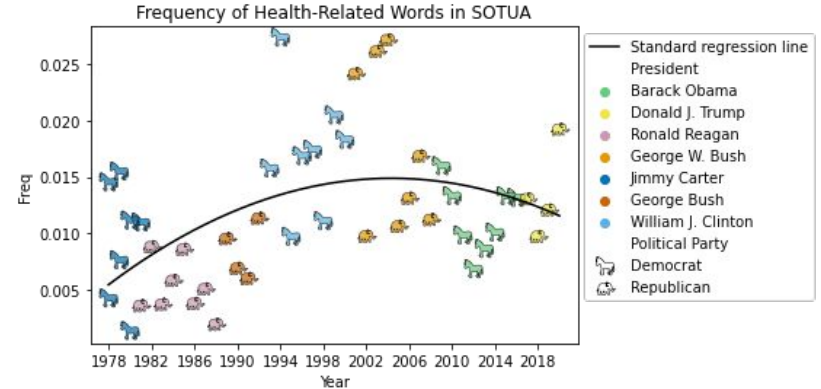
Yes, there does seem to be a **negative correlation** between budget spent on the health department and speech time spent on health-related topics post-Clinton (but it's negative)...

- **P-value of 0.018**
- MSE was **0.0006** - lower than without clustering
- 95% confident of a coeff between **-4.729 and -0.509**; pretty certain of a negative trend

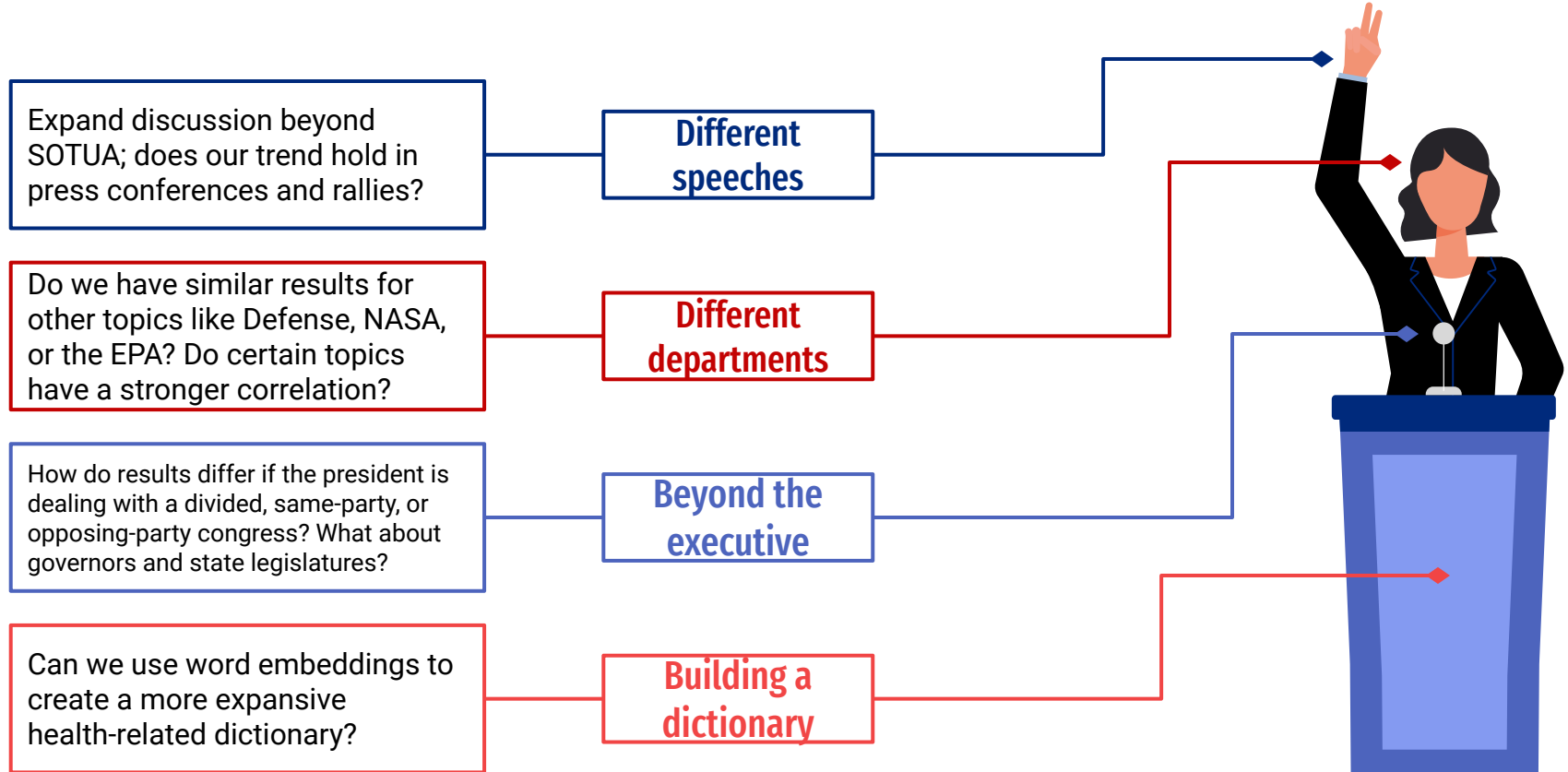
Why the negative trend?

Our regression seems to indicate that **the more a president talks about healthcare, the less budget will be allocated.**

- Recent presidents (Obama and Trump) don't seem to be talking as much about healthcare as Bush and Clinton
 - State of the Unions have been getting longer** as more people can tune in via internet or TV
 - Ideally would be checked by clustering by president to see if trend remains, but each president has at most 8 SOTUA, making any regression insignificant
- Presidents use the SOTUA to **focus on what needs to be done**; if there is a lot of speech time they are **pressuring congress to increase the budget**. If they are not talking about an issue that much, that probably means they are **satisfied with the current level of funding**.



Future analysis



Final Takeaways

Partisanship

There is a shift in how health is discussed over a president's term for both parties.



Core Research Question

After Clinton is a negative statistically significant negative relationship between how much presidents discuss health-related terms in SOTUA and how the health department budget.

Thank You

Please reach out to us if you have any questions!