



Sarah DalleyHood

Angeli Lucila

Dramane Nebie

Perry Reynolds

Keiko Yara



Google Trends & Covid Data Analysis Project



Outline

Motivation

Covid disrupted a lot of industries, and people ended up with lots of free time - so we were interested in what people's interests were during this time.

Project Theme

Exploration of Top Google Searches in the US between 2019-2020.

Key Questions

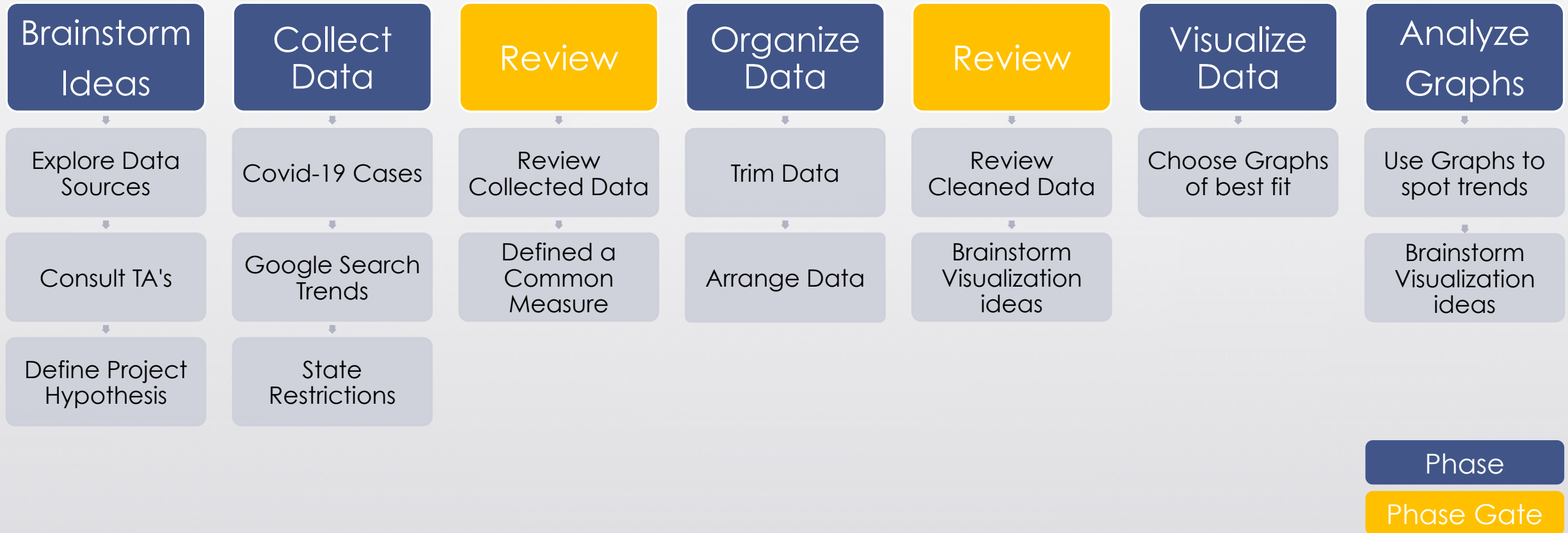
Has Covid-19 impacted Google searches in the US compared to previous years?

Did State Restrictions play a Factor?

Hypothesis

We hypothesize that states will differ in their average search interests between 2019 and 2020, and states will vary amongst themselves in their 2020 search interests.

Summary





Questions & Data

Project Questions

Will states will differ in their average search interests between 2019 and 2020? Will states vary amongst themselves in their 2020 search interests?

Expectations

- 1) The coronavirus had shifted people's google searches, where people would probably search more for entertainment, video online communication, and health information
- 2) A state with strong lockdowns would have more interest in the above categories, due to having more free time

Data Sources

Google Trends API, Covid Tracking Project Data (Atlantic Monthly Group), ESRI Disaster Response Program API for Stay at Home Orders

Assumptions

1. Default search engine in the US is Google
2. States are homogeneous within themselves
3. 2019 represents a typical year



Data Cleanup & Exploration

Step 1: Collecting the Data

- Google API: conduct a Top Trends keyword search on 2018, 2019, 2020, connect those into larger categories for calling to Google Trends Interest Over Time search
- Covid Data: importing CSVs of data for all states over time
- Stay At Home Orders API: calling API for all 50 states (and DoC)


Step 2: Cleanup

- Creating CSVs for each state with Covid case numbers and stay at home orders



Two State Comparison

- **California vs Wyoming**
- Stringent lockdown vs. No lockdown
- Coastal vs. Midwestern
- Large population vs. Small population



Data Analysis - Visualization

Nested Pie Charts

2020 and 2019 groups, 2020 groups and categories

Popularity Graphs

For each state, for each group of categories, line graph of the popularity of each category over time, along with a bar chart of positive Covid cases

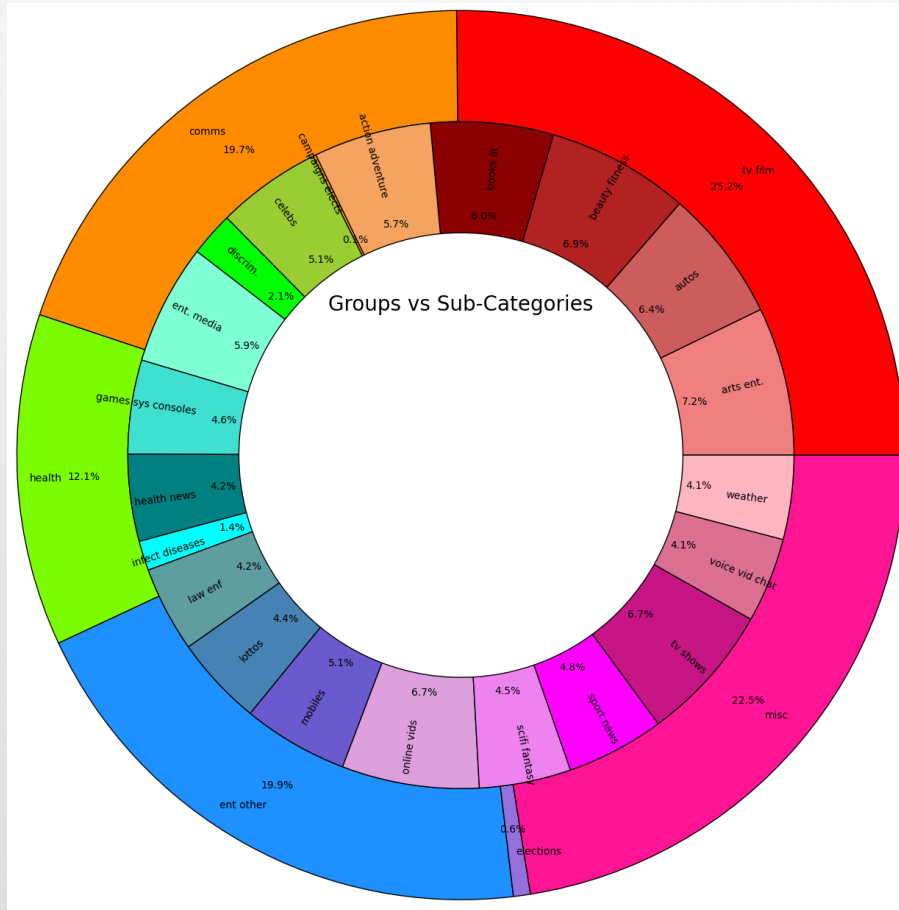
Means Comparisons

The mean percentage popularity for 2020 in that category different to the mean percentage popularity for 2019?

2020-2019 Difference Graphs

Shows difference between the 2019 search percent at any point in time and the 2020 search percent at the same point in time

Groups 2019 vs 2020



Groups v. Categories 2020



Means Comparisons: what is significant?

CA

| | 2019_means | 2020_means | p_vals | stats_signif | 2019_stdev | 2020_stdev | diff_btwn_means |
|------------------------|------------|------------|-------------|--------------|------------|------------|-----------------|
| categories | | | | | | | |
| action_adventure | 66.384615 | 62.450980 | 0.00525988 | Yes | 7.86461 | 5.84208 | -3.933635 |
| scifi_fantasy | 57.230769 | 61.215686 | 0.114452 | No | 14.6663 | 10.0826 | 3.984917 |
| tv_shows | 84.807692 | 81.078431 | 0.00157805 | Yes | 5.19629 | 6.26801 | -3.729261 |
| online_vids | 80.384615 | 78.725490 | 0.250488 | No | 3.42027 | 9.5303 | -1.659125 |
| mobiles | 52.826923 | 48.823529 | 0.0059075 | Yes | 8.8746 | 4.78029 | -4.003394 |
| voice_vid_chat | 58.634615 | 51.784314 | 0.00480011 | Yes | 9.85867 | 13.6171 | -6.850302 |
| infectious_diseases | 9.769231 | 25.980392 | 4.73493e-07 | Yes | 1.01177 | 19.8084 | 16.211161 |
| health_news | 55.903846 | 64.372549 | 7.43562e-05 | Yes | 5.95257 | 12.9402 | 8.468703 |
| games_systems_consoles | 51.423077 | 62.647059 | 2.14724e-06 | Yes | 5.30365 | 14.2812 | 11.223982 |
| arts_entertainment | 91.057692 | 84.803922 | 1.44875e-08 | Yes | 3.35934 | 6.13578 | -6.253771 |
| books_lit | 75.519231 | 68.549020 | 0.00126812 | Yes | 12.1679 | 8.63035 | -6.970211 |
| lottos | 49.057692 | 36.764706 | 1.72109e-10 | Yes | 10.5874 | 5.40003 | -12.292986 |
| celebs | 56.461538 | 60.745098 | 0.0159189 | Yes | 6.10316 | 10.7061 | 4.283560 |
| entertainment_media | 70.384615 | 71.098039 | 0.678035 | No | 4.53683 | 11.2324 | 0.713424 |
| campaigns_elections | 0.192308 | 3.156863 | 0.135203 | No | 0.394113 | 13.8002 | 2.964555 |
| discrimination | 24.115385 | 31.372549 | 0.0081226 | Yes | 8.46634 | 16.8929 | 7.257164 |
| law_enf | 37.519231 | 37.431373 | 0.957739 | No | 3.17739 | 11.2411 | -0.087858 |
| weather | 43.115385 | 33.490196 | 0.00189214 | Yes | 19.3914 | 8.72763 | -9.625189 |
| autos | 82.307692 | 77.941176 | 0.00366652 | Yes | 5.78971 | 8.60554 | -4.366516 |
| beauty_fitness | 86.942308 | 83.372549 | 0.00194262 | Yes | 3.6184 | 6.99288 | -3.569759 |
| sport_news | 77.769231 | 57.392157 | 1.95682e-13 | Yes | 8.23598 | 14.1519 | -20.377074 |

WY

| | 2019_means | 2020_means | p_vals | stats_signif | 2019_stdev | 2020_stdev | diff_btwn_means |
|------------------------|------------|------------|-------------|--------------|------------|------------|-----------------|
| categories | | | | | | | |
| action_adventure | 58.538462 | 58.333333 | 0.92855 | No | 11.9268 | 10.9949 | -0.205128 |
| scifi_fantasy | 45.365385 | 47.529412 | 0.552689 | No | 17.043 | 19.3627 | 2.164027 |
| tv_shows | 81.250000 | 75.588235 | 0.0035644 | Yes | 10.0381 | 9.00583 | -5.661765 |
| online_vids | 78.615385 | 73.941176 | 0.0111616 | Yes | 8.28612 | 9.79478 | -4.674208 |
| mobiles | 53.730769 | 51.078431 | 0.22755 | No | 12.2162 | 9.60156 | -2.652338 |
| voice_vid_chat | 26.519231 | 30.196078 | 0.346462 | No | 16.5314 | 22.0668 | 3.676848 |
| infectious_diseases | 9.173077 | 24.725490 | 1.91166e-05 | Yes | 2.37553 | 23.2076 | 15.552413 |
| health_news | 30.615385 | 34.117647 | 0.471215 | No | 22.8365 | 25.7116 | 3.502262 |
| games_systems_consoles | 46.326923 | 53.843137 | 0.00981761 | Yes | 12.9447 | 15.582 | 7.516214 |
| arts_entertainment | 88.250000 | 84.490196 | 0.000299478 | Yes | 4.58205 | 5.44625 | -3.759804 |
| books_lit | 68.692308 | 62.607843 | 0.00934249 | Yes | 12.9712 | 9.89965 | -6.084465 |
| lottos | 56.884615 | 49.568627 | 0.0351317 | Yes | 17.091 | 17.3321 | -7.315988 |
| celebs | 59.057692 | 65.196078 | 0.00882569 | Yes | 10.0564 | 12.8185 | 6.138386 |
| entertainment_media | 61.019231 | 63.470588 | 0.405385 | No | 14.9428 | 14.5418 | 2.451357 |
| campaigns_elections | 0.076923 | 2.862745 | 0.160403 | No | 0.266469 | 13.8224 | 2.785822 |
| discrimination | 16.576923 | 24.647059 | 0.0142413 | Yes | 12.5627 | 19.1176 | 8.070136 |
| law_enf | 51.076923 | 48.392157 | 0.358082 | No | 15.1097 | 14.1062 | -2.684766 |
| weather | 52.788462 | 46.098039 | 0.0116618 | Yes | 15.2987 | 10.3984 | -6.690422 |
| autos | 75.326923 | 73.647059 | 0.450444 | No | 10.8782 | 11.3928 | -1.679864 |
| beauty_fitness | 86.807692 | 85.098039 | 0.251456 | No | 6.67671 | 8.12948 | -1.709653 |
| sport_news | 44.288462 | 33.215686 | 0.000774639 | Yes | 17.4017 | 14.5728 | -11.072775 |



Category Samples

- Entertainment Media
 - Book retailers
 - CD & Audio Shopping
 - DVD & Video Shopping
 - Entertainment Media Rentals / Video Game Retailers
- Health News
 - All health news
 - Health policy
 - Covid topics



Data Analysis - Means Comparison

Entertainment Media

Average interest in entertainment media *stayed the same* for both states

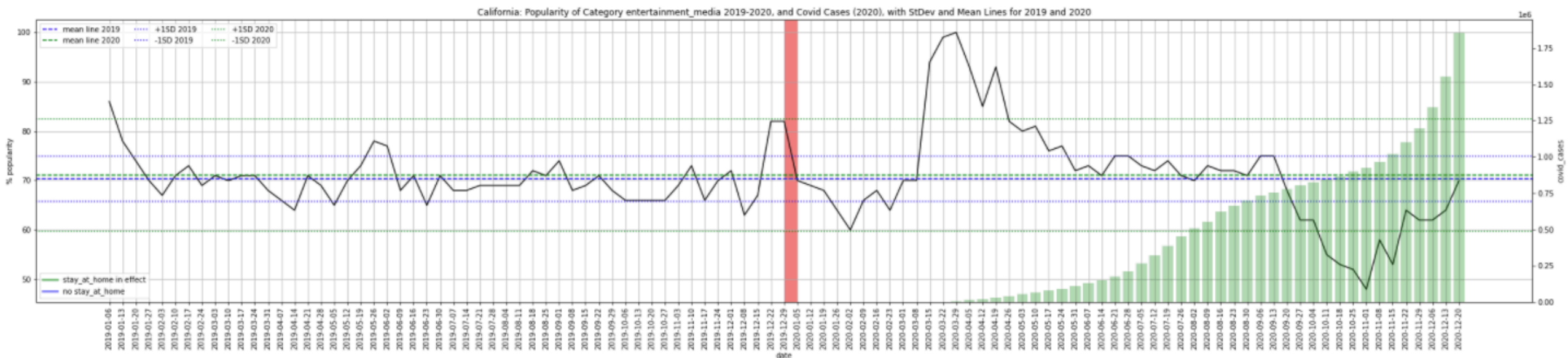
Health News

Average interest in Health News increased for California, but not for Wyoming

Do either of these sound odd? Let's look at the graphs

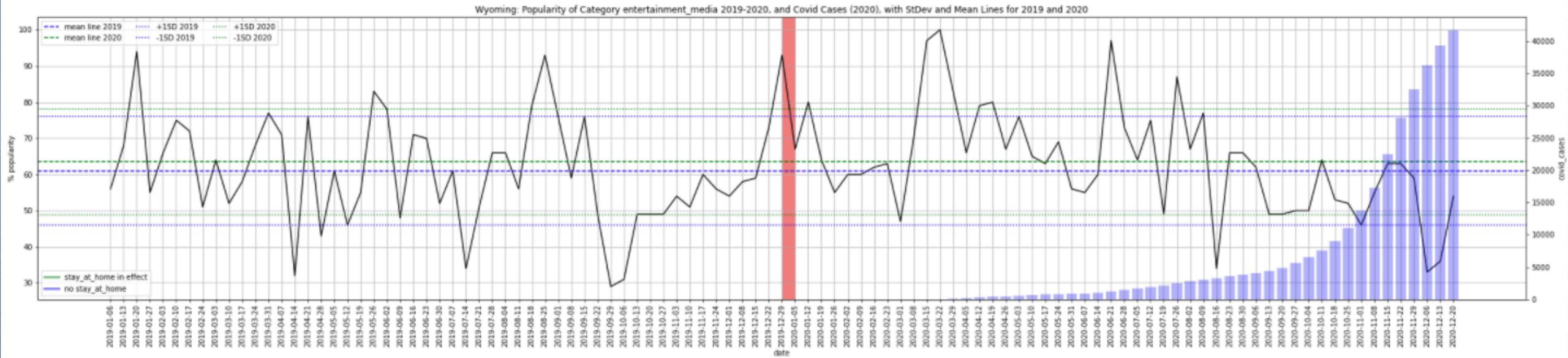
Entertainment Media, California

(YOY by Month)



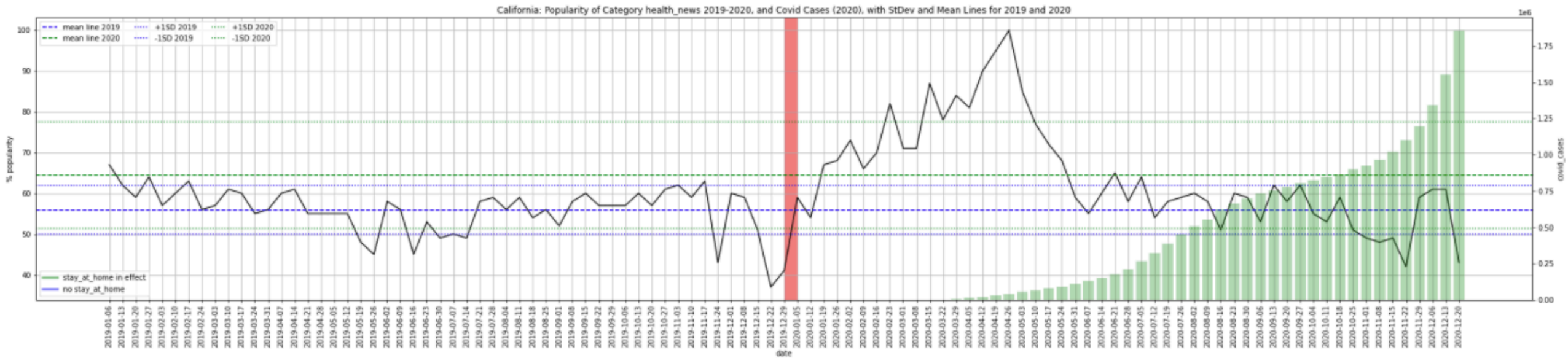
Entertainment Media, Wyoming

(YOY by Month)



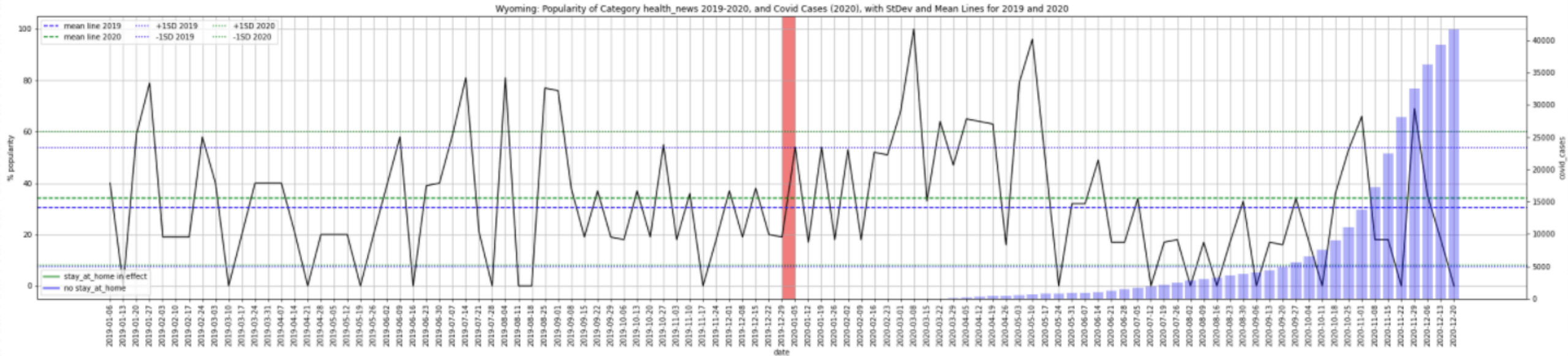
Health News, California

(YOY by Month)



Health News, Wyoming

(YOY by Month)





Discussion

Entertainment Media

- The means test is somewhat misleading
- CA: spike in early months of the lockdown - makes sense, people are looking for more kinds of entertainment to access during their free time
- WY: also a peak in early in the same months, but no lockdown. Perhaps people were in a similar mindset as the rest of country.

Health News

- Both states have spikes in interest when cases rise, but only CA maintains interest throughout the year.



Post Mortem

Conclusions

the lockdown under pandemic led people to be more interested in entertainment and health. However, a simple lockdown/no lockdown dichotomy cannot be drawn.

Problems

Google Trends scope: we wanted to try to analyse +1000 categories. Even using the sleep function, it was taking days to even create a csv for one category. Reduction to 21 categories. Additional issues with some of those categories required changing some specific categories (e.g. errors with specific television categories).

Future Research

Challenging assumption that 2019 is a typical year, comparing 2020 data with more previous years; looking into what employers did despite/in concert with lockdown orders



Questions?