Drive link: <https://drive.google.com/drive/folders/1eD6Z8ejc_DqP8dak-NG56a_hsMUOvyML>

Zoom Link: Stephen Brock is inviting you to a scheduled Zoom meeting.

Topic: Stephen Brock's Personal Meeting Room

Join Zoom Meeting

<https://utoronto.zoom.us/j/6624538516?pwd=aWlxZUJxbGlMam96NWI1VlIzYVlhZz09>

Meeting ID: 662 453 8516

Passcode: 663054

Deadlines/Benchmarks:

- domain selection

- **What are we looking for in a domain? What makes a domain feasible?**

**- How accessible is the data?**

**- What deliverables are possible?**

**- How does it relate to covid?**

To do by next Sunday (October 4, 2020):

* Have 2-3 domain suggestions and consider possible benchmarks/deliverables
* Start Jira

Possible ideas for Sunday (Oct 4th):

* As a group, create a Jira project and plan benchmarks/timelines

Possible ideas:

- Mental health

- Hospital data

- Home network security, connecting remotely to the workplace

- Silverlinings of COVID-19

- Mask and sanitization sales

- lingering effects

- transit effects (commute)

- real estate

- NBA

- Twitch streaming, link to productivity (increased hours during the day?), youtube, netflix, etc.

- UBI

- Esports

- Influence on politics, elections

- scraping reddit and/or other social media

- nature and hiking

- suppression of severity

- covid 19 testing accuracy, significance

Notes from mentor meeting

* Nancy Reid suggestions/thoughts: vaccine trials/community spread/treatments for hospitalized patients/prediction methods for short-term and long-term; demographics of the disease/impacts of lockdown
  + Keep in mind that this is for decision makers - scale and practicality matters (ie. hard to do something for the entire world) - practicality is highly valued by judges
* Datasets to avoid: kaggle competitions (that are nearing completion),
* Start with the UofT COVID dataset library

UofT dataset library: <https://mdl.library.utoronto.ca/covid-19/data>

Google dataset library: <https://console.cloud.google.com/marketplace/browse?filter=solution-type:dataset&filter=category:covid19>

Sports and Politics:

https://projects.fivethirtyeight.com/

|  |  |  |  |
| --- | --- | --- | --- |
| Domain/Idea | Link to dataset (if available) | Tool | Description/Notes (include links) |
| Government Response Tracker - https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker#data | https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker#data | Determine whether the 6 WHO conditions are satisfied to relax lockdown, "lockdown rollback checklist" | The current data is supposedly good enough to determine 4 of the 6 requirements. We can explore verifying the other 2 as well.  Checklist:  https://www.bsg.ox.ac.uk/sites/default/files/2020-06/Lockdown%20Rollback%20Checklist%20v4.pdf  Link to website:  https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker |
| Government Response Tracker | https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker#data | - We could create a dashboard to visualize this data  - determine if the lockdown is working  - Could try to determine the optimal time to have a lockdown  - We could try to anticipate breakouts and closures | The dataset contains government response indicators. |
| Understanding COVID-19 articles (news and research) | Note:  The file size is over 4.5 GB. I tried opening it in Excel and the program crashed.  https://www.semanticscholar.org/cord19/download | Summarize news article into a paragraph | There is a lot of information out there and a lot of articles.  Considerations:   * Summary will need to properly preserve statistical information. Could be difficult to train. |
| COVID-19 Development vs GDP | https://ourworldindata.org/grapher/total-confirmed-cases-of-covid-19-per-million-people-vs-gdp-per-capita | See how COVID-19 has impacted GDP of countries | This might be hard to see as I don't think GDP is recorded very frequently.   * World data is also tricky * Can we make this useful for decision makers |
| Movement of people - how did the number of visitors going to stores change over the course of the pandemic? | Google cloud:  <https://console.cloud.google.com/marketplace/product/geotab-public-data/covid19-mobility-impacts?filter=solution-type:dataset&filter=category:covid19&id=f60d5b5c-8df5-4d07-85fe-5e8a551cede7> | Build a dashboard to see how the number of visitors is changing. I think this would be really cool on a more granular scale (city or state).  https://ourworldindata.org/covid-mobility-trends | Not sure how to get data off of Google cloud. Looks like we need to sign up for a free trial |
| Influence of weather on spread of covid | https://climate.gov/maps-data/primer/what-environmental-data-are-relevant-study-infectious-diseases-covid-19 |  | Professor Patrick Brown - created a model for rise and fall of the disease  - could investigate multiple factors that influence the spread, but start with weather (?)  - temperature and weather conditions |
| Influence of COVID on mental health in Canada | <https://www.delvinia.com/camh-coronavirus-mental-health/?utm_source=CAMH%20Landing%20Page&utm_medium=button&utm_campaign=Camh-Delvinia> |  | Primarily survey data. Need to sign up for a methodify account for access. |
| Looking at the effectiveness of COVID-related policing strategies | <https://dataverse.scholarsportal.info/dataset.xhtml?persistentId=doi:10.5683/SP2/KNJLWS> | Predictive model to determine what types of policing strategies are most effective for preventing recurrent COVID protocol violations |  |
| Another dataset for Govt response | <https://www.acaps.org/covid19-government-measures-dataset> |  | Using this along with the day-by-day general statistics of COVID-19 can make us either laud the efforts or beat a government’s ass. Should certainly useful for decision making. |
| Layoffs in the startup spectrum | <https://airtable.com/shrFrEtQb57krTYGz/tbl8c8kanuNB6bPYr?backgroundColor=green&viewControls=on> |  | Not sure how to extract data from this dataframe but seemed like an interesting data to explore |
| COVID and housing | <https://data.world/abcotvdata/covid-housing-market> |  |  |
| Toronto neighbourhood + minorities data | <https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/> |  |  |
|  |  |  |  |
|  |  |  |  |

High temperature and humidity impact on spread of COVID-19

https://papers.ssrn.com/sol3/Papers.cfm?abstract\_id=3551767

**Stephen Tasks:**

* look into COVID-19 lockdown procedures and their impact on cases
* Plan - (confirmed case count analysis)
  + plot cases vs time for a few cities (e.g. Toronto, Vancouver, Edmonton)
  + plot vertical lines for when lockdown announcements were made
  + investigate the change in cases from there onwards
  + try making a linear model to account for the changes in counts attributed to each lockdown procedure
* possible lockdown procedures:
  + school closures
  + mandatory masks
  + restaurant closures
  + quarantine/self-isolation
  + physical distancing
* links
  + ontario timeline
    - https://covid-19.ontario.ca/data
    - https://globalnews.ca/news/6859636/ontario-coronavirus-timeline/
  + BC timeline
    - https://www.cbc.ca/news/canada/british-columbia/covid-19-bc-timeline-1.5520943
  + Alberta timeline
    - https://calgaryherald.com/news/six-months-a-timeline-of-covid-19-in-alberta
  + Quebec timeline
    - Not sure where downloadable data is
    - radio-canada looks like it has a good dashboard
    - https://montreal.ctvnews.ca/covid-19-in-quebec-a-timeline-of-key-dates-and-events-1.4892912
* GLM or LM
  + explanatory variables: lockdown procedure indicators, holidays
* Interactive plots
* Create plan for moving forward, incorporating Allen's work

**Allen Tasks:**

* Broad question: effectiveness of COVID messaging on lockdowns and cases
  + “Is COVID messaging working?”
* Find data related to COVID messaging (government, media/news, social media)
* Classify data - possibilities:
  + Sentiment analysis (positive vs negative framing)
  + Presence of keywords (ex. Masks, vaccine, conspiracy)
* Effects could be on:
  + Lockdown effectiveness - how many people are attending gatherings and events, going to stores/restaurants/bars/clubs
    - Where to find data for this?
  + Case counts, R-naught values
  + Mental health indicators (survey data, depression and suicide rates, etc.)
* Start with local (Toronto), potentially expand to other cities and provinces after
* Other factors to consider:
  + Varying effects of types of messaging on different demographic populations
* Links:
  + **Global COVID mobility reports**: <https://www.google.com/covid19/mobility/>
    - Really useful resource. Shows daily changes in retail, restaurants, groceries/pharmacies, parks, transit, workplace, etc. Can organize by province, or even county
  + Ontario timeline: <https://www.publichealthontario.ca/-/media/documents/ncov/epi/2020/08/covid-19-wave-1-transmission-patterns-epi-summary.pdf?la=en>
  + COVID Twitter dataset: <https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset>
  + Tutorial on sentiment analysis on news headlines extracted from reddit: <https://www.learndatasci.com/tutorials/sentiment-analysis-reddit-headlines-pythons-nltk/>
  + Swiss dataset on covid news: <https://covidnewsdataset.herokuapp.com/>
  + Daily cases by province dataset (has data on recovery, testing, positivity rate, ICU, etc) <https://resources-covid19canada.hub.arcgis.com/datasets/provincial-daily-totals>

Stats Canada datasets:

https://www150.statcan.gc.ca/n1/en/type/data?subject\_levels=45%2C4510

**We can use these couple of datasets with other problems to make them more interesting**

We can use <https://aqicn.org/data-platform/covid19/report/18344-92a4536b> to add features to any weather related COVID project. Has a world-wide repository of air-quality

US Hospital Beds (ICU and non-ICU etc) <https://coronavirus-resources.esri.com/datasets/definitivehc::definitive-healthcare-usa-hospital-beds/data?geometry=92.988%2C-16.820%2C-117.950%2C72.123>

**Nancy’s readme.rtf file:**

<https://www.stat.cmu.edu/~kass/covid.html>

<https://sciencepolicy.ca/covid-19-resources>

<https://www.statschat.org.nz/2020/04/19/counting-rare-things-is-hard/>

<https://www.significancemagazine.com/business/647>

https://www.vox.com/future-perfect/2020/5/2/21241261/coronavirus-modeling-us-deaths-ihme-pandemic

<https://www.wired.com/story/everything-we-know-about-covid-19-antibody-tests-so-far/#intcid=recommendations_wired-homepage-right-rail-popular_509661cf-fdbd-4fc1-9a77-ecc1b57fdf7d_popular4-1>

<https://threadreaderapp.com/thread/1258607277357006849.html>

https://covid19-projections.com/about/

https://covid-19-canada.uwo.ca/

https://twitter.com/callin\_bull/status/1261855753846927360?s=20

<http://teaching.sociology.ul.ie/bhalpin/wordpress/?p=669>

<https://www.youtube.com/watch?v=GDO72o0Z6-k> Lily wang Iowa State youtube

<https://media.ed.ac.uk/media/Communicating%20statistics,%20risk%20and%20uncertainty%20in%20the%20age%20of%20Covid%20-%20Prof.%20David%20Spiegelhalter/1_y2lw3u6d>

<https://uconn-cmr.webex.com/mw3300/mywebex/default.do?nomenu=true&siteurl=uconn-cmr&service=6&rnd=0.2360978646794718&main_url=https%3A%2F%2Fuconn-cmr.webex.com%2Fec3300%2Feventcenter%2Fevent%2FeventAction.do%3FtheAction%3Ddetail%26%26%26EMK%3D4832534b00000004d595f7174bae84f2fce2a23fe669a78515a3a2a714fb6cfa1003af24f4f2ec25%26siteurl%3Duconn-cmr%26confViewID%3D166533466818642511%26encryptTicket%3DSDJTSwAAAARZz4Y6LZJfI3mdrfhrIfRfQqA28de9StbnWACcUIJfRQ2%26>

(Amherst Reich talk) see also https://arxiv.org/abs/2005.12881

https://www.covid19vaccinetracker.com/. Erica Moodie’s vaccine page