Project Deliverable 1 (Due: 10/15)

Sufficient data should have been collected to perform a preliminary analysis of the data and attempt to answer one question relevant to your project proposal which you will submit as a pull request. If data has already been collected for your project you must answer two questions.

Checklist

- ☑ ~~Collect and pre-process a preliminary batch of data~~
  by zip:
  COVID-data_Massachusetts-vaccines_zipcodes
  by county and city
  COVID-data_Massachusetts-vaccines
- ☑ ~~Perform a preliminary analysis of the data~~
  - Reduce dataset to a select few cities that are representative of the demographics we want to analyze
    1. Revere (same demo as Chelsea without grass roots org)
    2. Chelsea (target city)
    3. Predominantly white middle/upper class city (maybe Wellesley or Newton)
    4. White blue collar city
- ☑ ~~Answer one key question~~
  a. How the population groups of Chelsea and Revere compare in terms of vaccination rates during the rollout
     - Datapoint 1 (4-6-2021)
       - **Revere:** Hispanic 8.4% vaccinated, White 25%
       - **Chelsea:** Hispanic 8.3% vaccinated, White 30%
       - **Newton:** Hispanic 13.24% vaccinated, White: 31.81%
       - **Wellesley:** Hispanic 7.45% vaccinated, White: 23.32%
     - Datapoint 2 (5-18-2021)
       - **Revere:** Hispanic 31% vaccinated, White 45%
       - **Chelsea:** Hispanic 34% vaccinated, White 63%
       - **Newton**: Hispanic 37% vaccinated, White 59%
       - **Wellesley:** Hispanic 26% vaccinated, White 52%
     - Datapoint 3 (7-6-2021)
       - **Revere:** Hispanic 49% vaccinated, White 56%
       - **Chelsea:** Hispanic 51% vaccinated, White 75%
       - **Newton:** Hispanic: 51% vaccinated, White 70%
       - **Wellesley:** Hispanic: 35% vaccinated, White 64%
     - Datapoint 4 (10-12-22)
       - **Revere**: Hispanic 64% vaccinated, White 66%
       - **Chelsea**: Hispanic 68% vaccinated, White 90%
       - **Newton**: Hispanic 64% vaccinated, White 80%
       - **Wellesley**: Hispanic 44% vaccinated, White 73%

Chelsea data reflects that its vaccination roll out matches more with wealthier regions, possibly because of La Colaborativa.

- [x] ~~Refine project scope and list of limitations with data and potential risks of achieving project goal~~
    a. The fully vaccinated population of Asian is greater than the total population of Asian within some time frame
    b. Missing all of other/unknown population data
        i. Missing total population for “other” sex
    c. Hard to tell the actual effects of La Collaborativa without having regions that they supported or timeline to their efforts. - This also applies to other grassroots efforts.

- [ ] Submit a PR with the above report and modifications to original proposal

---

Figure out how to fill out data set

Two groups: earlier time (3-9-21 -> 12-14-21) and older time (12-21-21 -> 10-4-22)

- **Earlier time**: missing 2 booster vaccinations
    - booster is not out yet - so no change as of now, need to consider these time periods as completely separate
    - whether wealthier regions have access **before** poorer regions - discrepancies
- 12-21-21 -> 07-05-22: missing the second booster vaccination
    - probably because second booster is not out - need to look at same discrepancies listed above
- **Older time**: missing partially vaccinated
    - we will fill out partially vaccinated data with **one dose - fully vaccinated**
- Missing infant data for most times
    - we will fill out missing data as **0**
        - most likely will not consider this data because vaccines are not necessarily designed for this population
        -
- In County/City Age, the partially_vaccinated data of Wellesley city from line 2396 to 2403 seems to be shifted up by one row, so we need to shift them down by one row.
- Manually added rows in Revere section are highlighted in red

    - Richer regions might have been able to have access to the booster earlier than the poorer regions
        - need to figure out when the booster actually came out

- what do we need to actually look at in terms of time frame
- normalize data for further analysis
    - vaccination percentage