BlueSkies
Stefani/Galinski
29 March 2020

PM5 Data Warehousing and Analysis

External Data Sources

1. Drug Utilization

    https://data.medicaid.gov/State-Drug-Utilization/State-Drug-Utilization-Data-2019/qnsz-yp89
    a. Information in data: drug utilization for each state, categorized by state and then by each individual drug code.
    b. How to use: we're pulling out only the asthma related drugs (albuterol, advair, atrovent, breo, budesonide, flovent, mometasone, proair, qvar, symbicort, ventolin, xopenex) as a measure of the total amount of asthma existing in each state.

2. CDC Asthma Impact

    https://chronicdata.cdc.gov/Chronic-Disease-Indicators/U-S-Chronic-Disease-Indicators-Asthma/us8e-ubyj
    a. Information in data: Asthma data for each state in the U.S., as measured by several key health indicators such as asthma mortality counts, asthma prevalence in adults, and vaccination percentages. Each indicator has several stratifications available such as gender, age, and race.
    b. How to use: The CDC offers several filters before retrieving the data from the API, so we pulled out only the data pertaining to asthma mortality counts and asthma prevalence in adults, as those indicators tie directly to our hypothesize regarding the public health impact of asthma.
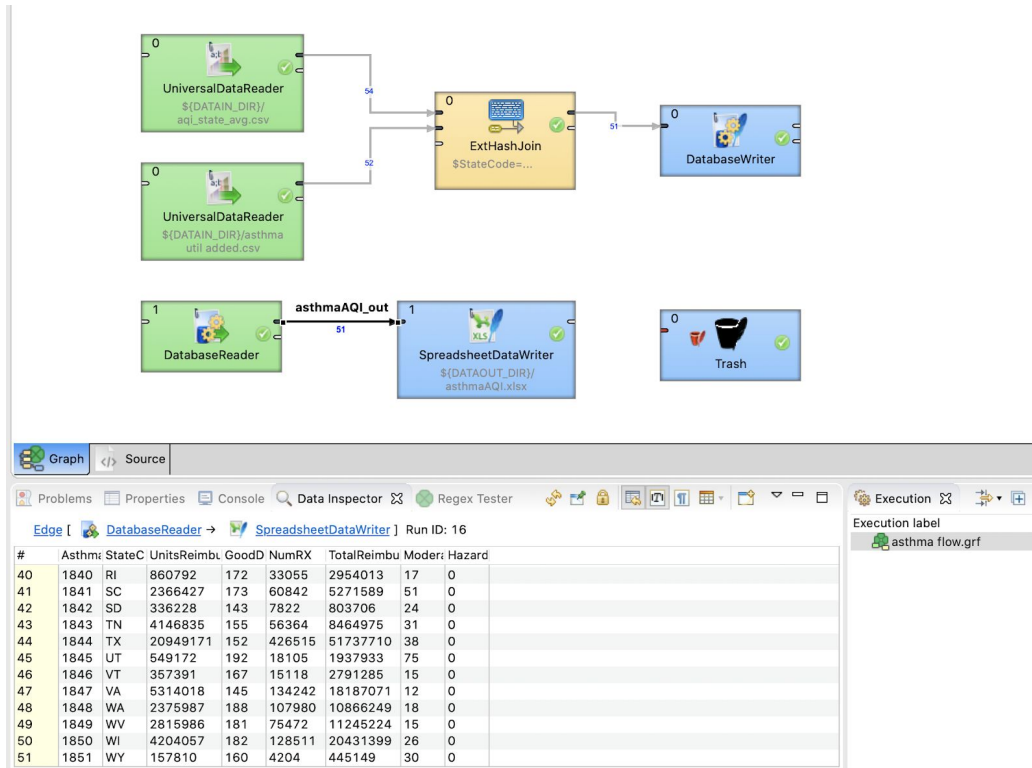
3. Socioeconomic Data by County (Poverty, Unemployment, Education)

    https://www.ers.usda.gov/data-products/county-level-data-sets/
    a. Information in data: Poverty percentages by county, employment and unemployment rates by county, median household income by county, and education levels by county.
    b. How to use: We pulled out the poverty rates by county to compare against asthma trends geographically. Employment rates are already loaded into our database, so we will also use that data to compare against the asthma trends.

ETL Workflows

1. Asthma drug utilization with air quality by state (counties averaged into state)

#	Asthma	StateC	UnitsReimbu	GoodD	NumRX	TotalReimbu	Modera	Hazard
40	1840	RI	860792	172	33055	2954013	17	0
41	1841	SC	2366427	173	60842	5271589	51	0
42	1842	SD	336228	143	7822	803706	24	0
43	1843	TN	4146835	155	56364	8464975	31	0
44	1844	TX	20949171	152	426515	51737710	38	0
45	1845	UT	549172	192	18105	1937933	75	0
46	1846	VT	357391	167	15118	2791285	15	0
47	1847	VA	5314018	145	134242	18187071	12	0
48	1848	WA	2375987	188	107980	10866249	18	0
49	1849	WV	2815986	181	75472	11245224	15	0
50	1850	WI	4204057	182	128511	20431399	26	0
51	1851	WY	157810	160	4204	445149	30	0

a. Phase 0:
   i. Input data (csv files): asthma_total_util.csv, aqi_county.csv
      - Clean aqi_county data to be averaged by state to make it easier to join
      - Clean asthma_total_util to be state numbers to make it easier to join
   ii. Joiner: ExtHashJoin
      - Join aqi_averaged and asthma_state by statecode
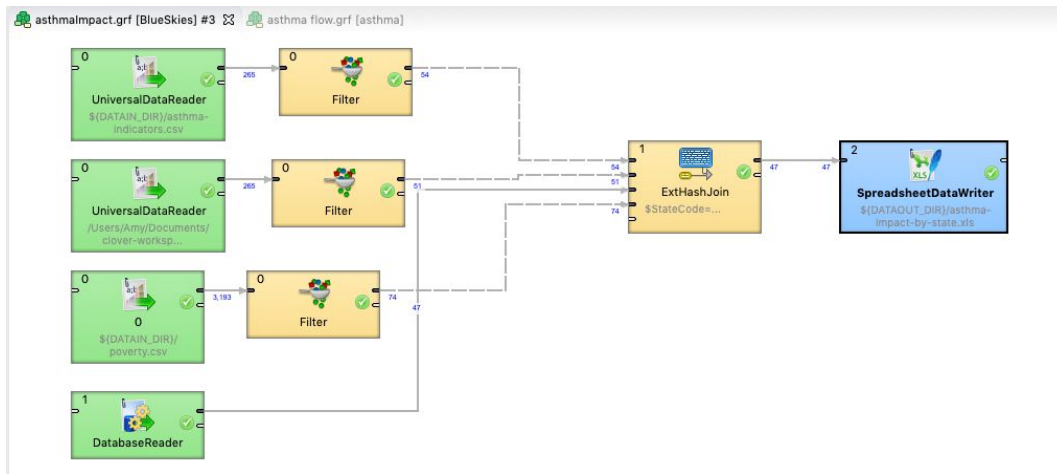   iii. Output: DBOutputTable
      - Output data into SQL table
b. Phase 1:
   i. Input data: DBInputTable
      - Take the data from SQL table
   ii. Writer: SpreadsheetDataWriter
      - Output data as xlsx file

2. Asthma impact against poverty levels, unemployment, and health care spending by state

  a. Phase 0:
    i. Input data: UniversalDataReader (csv files)
      ● asthma_indicators.csv, poverty.csv
    ii. Filter (asthma_prevalence data)
      ● Filter out "US" locations (we only want state-level data) and include only "Crude Prevalence" data.
    iii. Filter (asthma_mortality data)
      ● Filter out "US" locations (we only want state-level data) and include only "Crude Rate" data.
    iv. Filter (poverty data)
      ● Filter out to include only FIPS codes with the substring '000' (To select State aggregated rows, rather than county data).
  b. Phase 1:
    i. Input data: DatabaseReader
      ● UnemploymentRate, TotalHealthCareSpending from Socioeconomic and HealthCareSpending tables.
    ii. Joiner: ExtHashJoin
      ● Joined with asthma_prevalence, asthma_mortality, and poverty on StateCode
    iii. Writer: SpreadsheetDataWriter
      ● Wrote results to spreadsheet for charting trends.

Charts

Hypothesis: The public health burden of asthma varies geographically based on environmental conditions, socioeconomic conditions, and availability of quality healthcare.

Results: The data, represented in the graphs below, validated the above hypothesis. The public health burden of asthma varies based on the environment, the socioeconomic conditions, and the availability of quality healthcare. Further analyses are required to glean additional insights about statistical significance and potential causality.

Chart 1: Total asthma drug* prescriptions vs good AQI days by state

Significance: There is a correlation between the amount of good AQI days and the number of total prescriptions per state, as the number of days goes up, the number of prescriptions decrease (and vice versa).
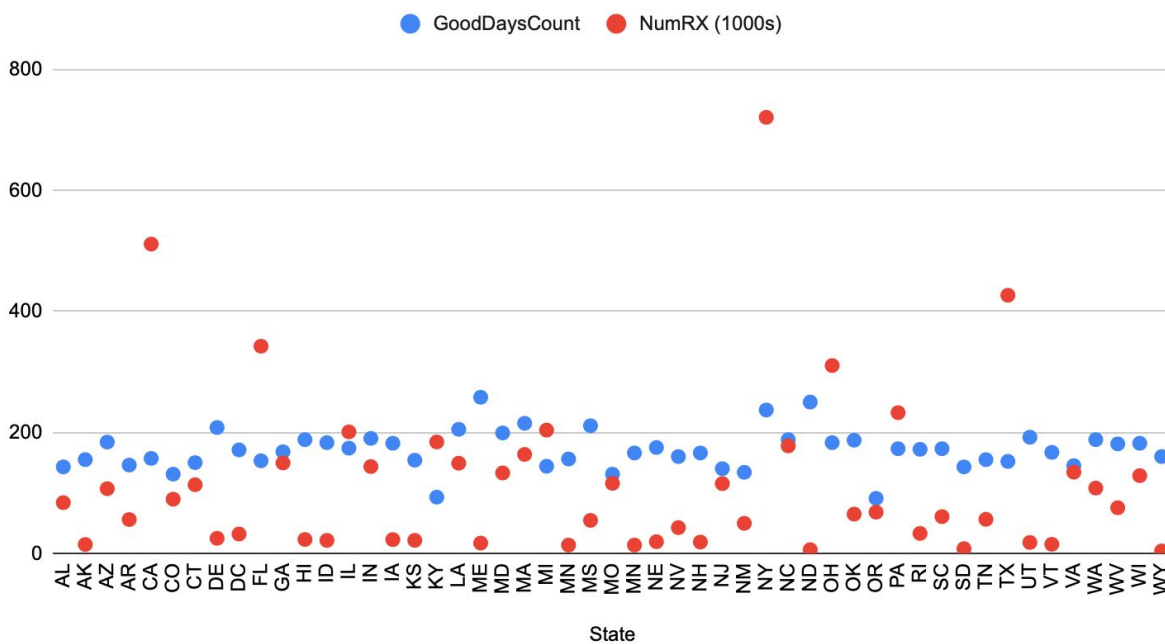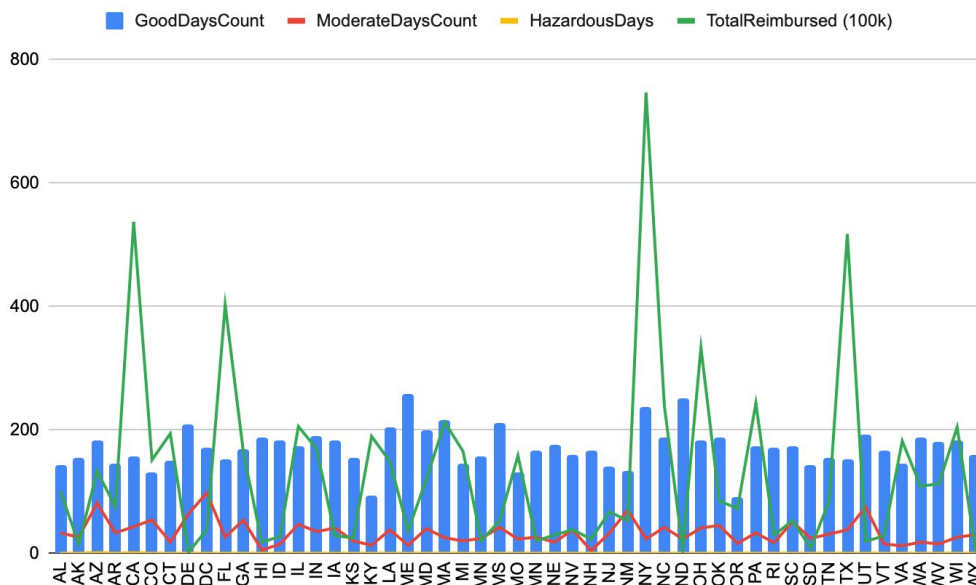

Good AQI Days and Asthma Rx

Chart 2: Asthma drug* utilization vs AQI day counts by state

Significance: There is a correlation that places with less total reimbursement (less utilization) have a higher number of good AQI days.

Common asthma drug treatments: Flovent, Albuterol, Qvar, Advair, Ventolin, ProAir

Chart 3: Asthma Mortality vs Poverty Percentage by state

Significance: There is an overall trend of higher asthma mortality correlated to higher levels of poverty. As poverty levels vary widely across states, additional insights would likely be apparent with future analyses conducted at the county or city level.
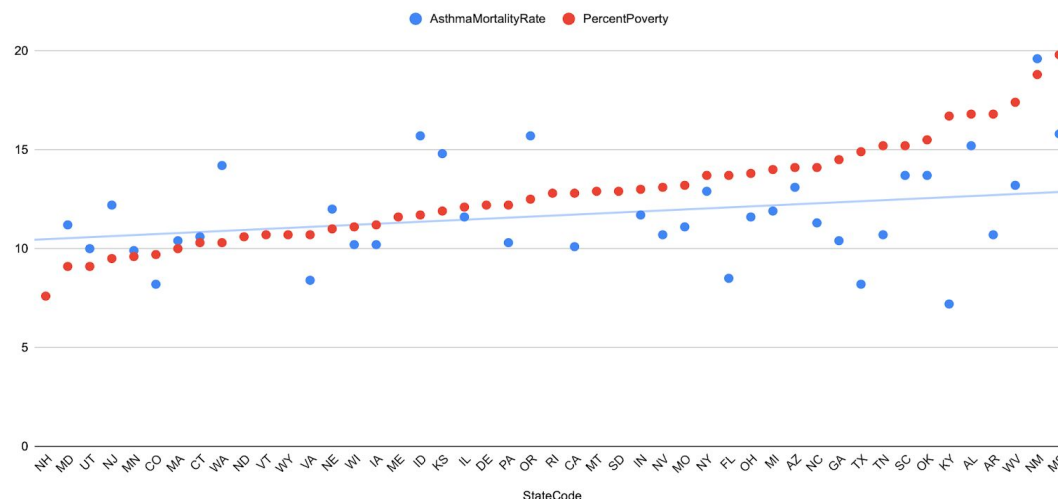
Chart 4: Percentage of adults with asthma vs Unemployment rate by state

Significance: There is a slight overall trend of increased prevalence of adult asthma correlated to increased unemployment rates. Again, unemployment levels vary widely across states, so future analyses should be conducted at the city or county level to further identify trends.
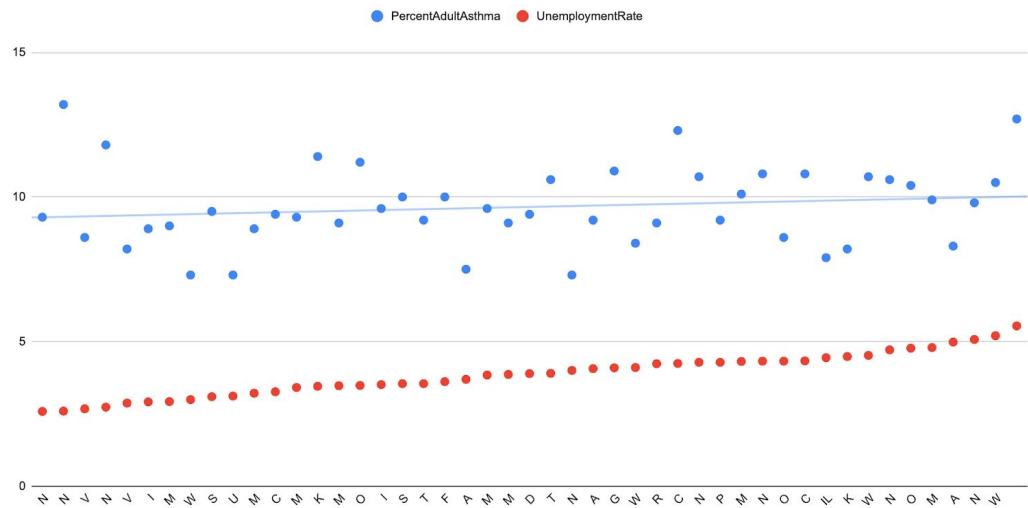

Chart 5: Asthma mortality rate vs Medicaid spending ratio by state

Significance: There is an inverse correlation with total health care spending rates against asthma mortality rates. Future analyses should consider the rates of healthcare spending by category (such as emergency care versus primary care spending) to source additional insights.