Weather Indicators & Storm

WLHA - CS327E

Wenkang Li
Hyejin An

Main dataset & Second dataset

Research Intention: Hope to see any relationship between weather factors such as attributes in Main dataset (which are temperature, precipitation, wind speed, snow) and the number of storm events in the U.S.

Main dataset: Daily Minimum, Maximum Temperatures, Daily Precipitation, Daily SnowFall, Snow Depth, Daily Wind Speed from 2009 - 2013(data source: Enigma, National Oceanic and Atmospheric Administration)

Secondary Dataset: Storm Events 2009 - 2013 (data source: Enigma, National Oceanic and Atmospheric Administration)

minimum_temperatures

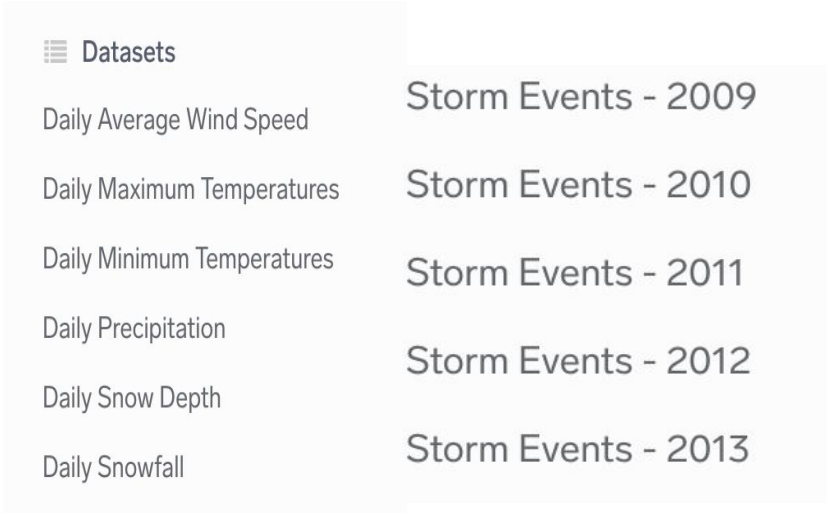
Schema Details Preview

Field name	Type	Mode	Description
id	STRING	NULLABLE	
date	TIMESTAMP	NULLABLE	
latitude	FLOAT	NULLABLE	
longitude	FLOAT	NULLABLE	
state	STRING	NULLABLE	
name	STRING	NULLABLE	
value	INTEGER	NULLABLE	
mflag	STRING	NULLABLE	
qflag	STRING	NULLABLE	
sflag	STRING	NULLABLE	
serialid	INTEGER	NULLABLE	
latitude_longitude_appended	STRING	NULLABLE	

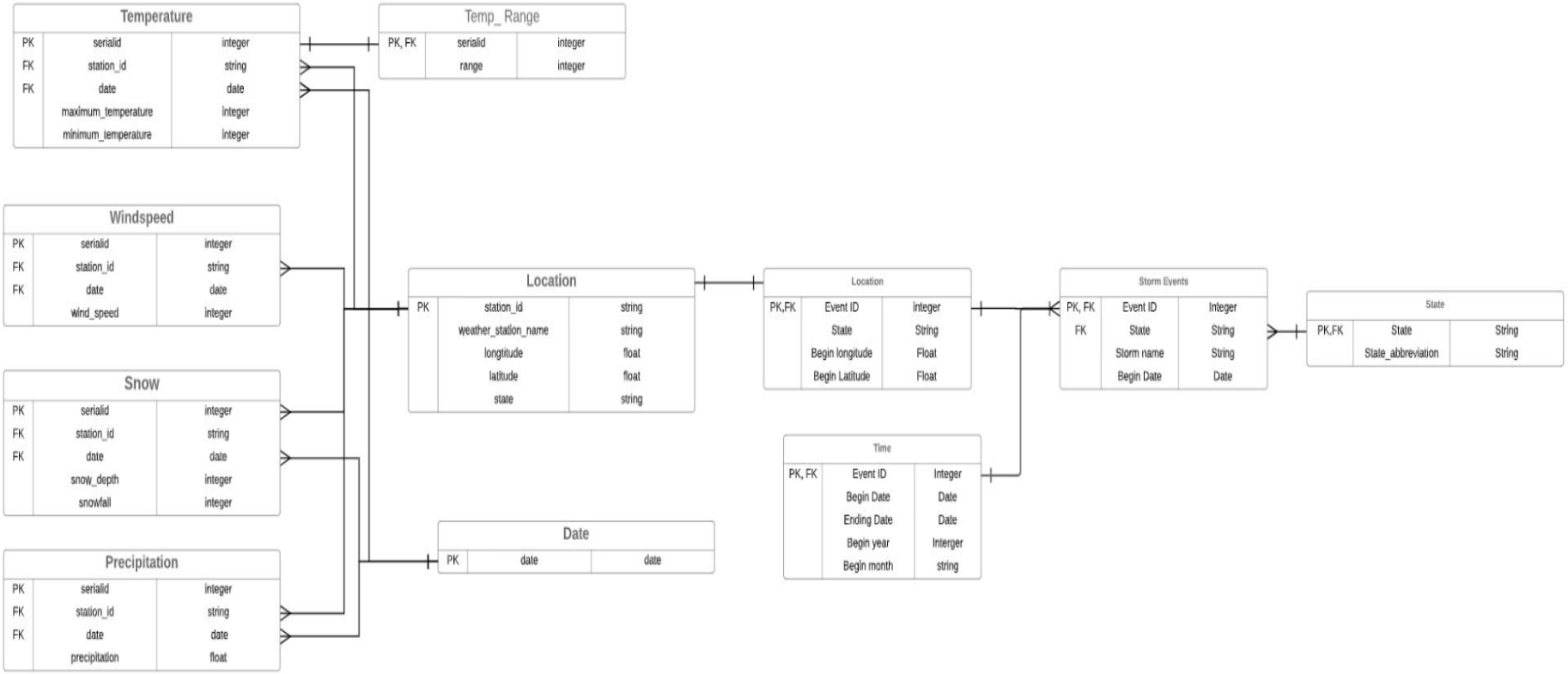
storm2009

Schema Details Preview

Field name	Type	Mode	Description
begin_yearmonth	INTEGER	NULLABLE	
begin_day	INTEGER	NULLABLE	
begin_time	INTEGER	NULLABLE	
end_yearmonth	INTEGER	NULLABLE	
end_day	INTEGER	NULLABLE	
end_time	INTEGER	NULLABLE	
episode_id	INTEGER	NULLABLE	
event_id	INTEGER	NULLABLE	
state	STRING	NULLABLE	
state_fips	INTEGER	NULLABLE	
year	INTEGER	NULLABLE	
month_name	STRING	NULLABLE	
event_type	STRING	NULLABLE	
cz_type	STRING	NULLABLE	



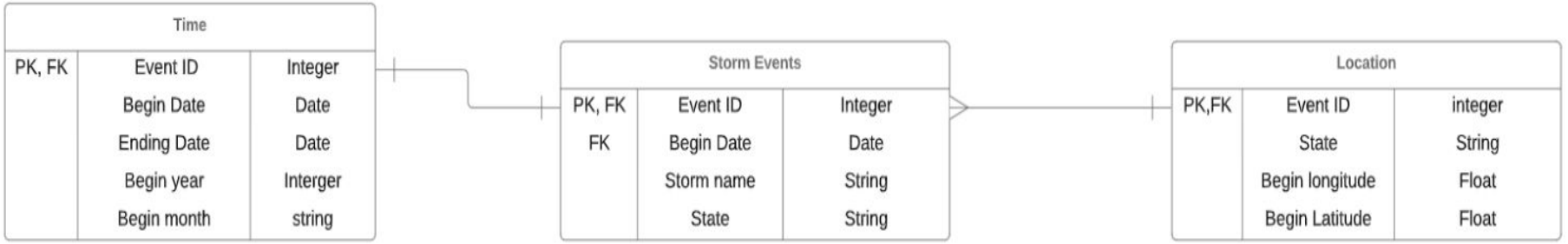
Normalized ERDs (main + secondary)



code: https://github.com/cs327e-spring2019/WLHA/blob/master/join-queries.sql

Normalized ERD

Secondary Dataset



Modeling & Cleansing with Beam

In the **main** dataset, we didn't have bad data with our tables

- Perform pardo to sort out non-null values in a state column in location table
- Perform pardo to create a new table with range between max. temperature & min. temperature

For the **secondary** dataset,

- Cleanse the time table for data attributes from timestamp to date.
- Transform our state attribute in the location table to abbreviation form.

Code:

https://github.com/cs327e-spring2019/WLHA/blob/master/temperature_single.py

https://github.com/cs327e-spring2019/WLHA/blob/master/transform_state_single.py

Cross Dataset

Secondary dataset links the main dataset by state and date

-- find states that have greater than avg number of storm events listing the wind speed, the avg(minimum temperautre), and the avg(maximum temperature)

-- list states and storm names with the over average number of storm events with average of precipitation, in order see relationship between precipitation and the type of storm name.

--list states and storm names with the over average number of storm events with average of precipitation, in order to see relationship between precipitation and the type of storm name.

-- queries show the relationship between the number of storm and average max/min temperature/precipitation/windspeed /snow in each state, each year

SQL code examples:

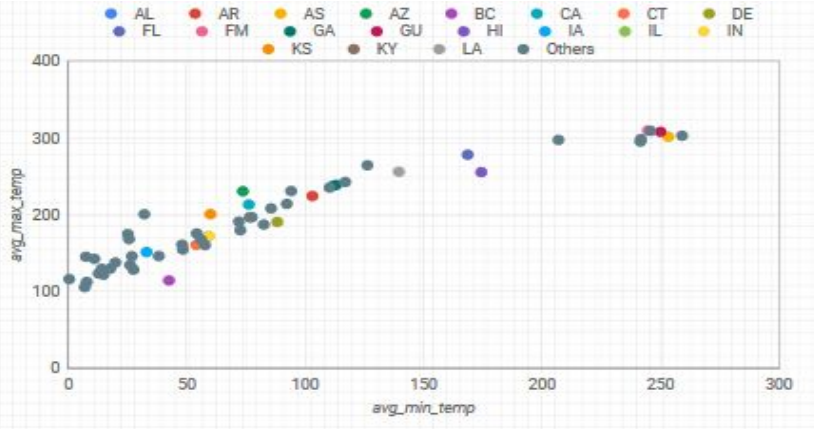
https://github.com/cs327e-spring2019/WLHA/blob/master/cross-dataset-queries.sql

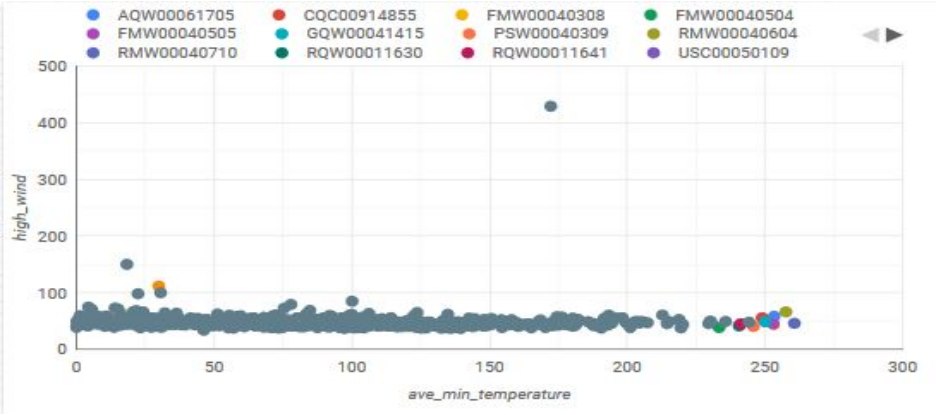
Airflow

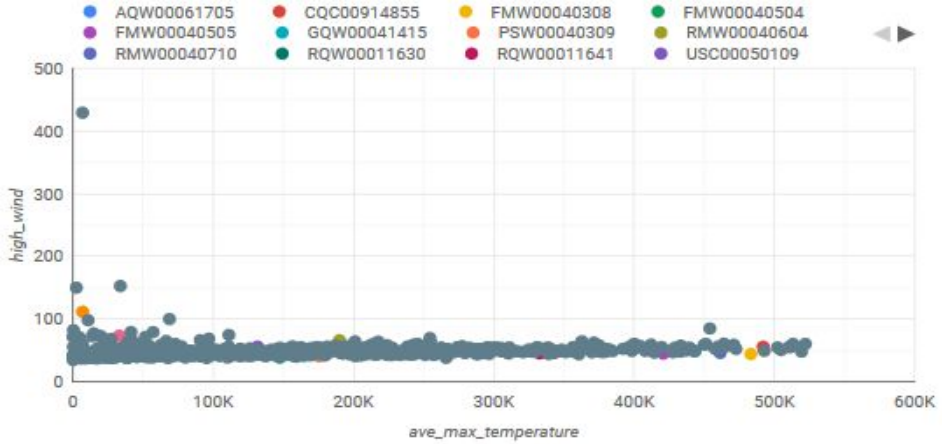
- Idea
  - Create a transformed table for the duration of each storm in workflow dataset
  - Duration = the range between begin date & end date in “time” table of dataset2
- Task Order
  - Create a workflow dataset
  - Create time table (from dataset2)
  - Beam time table (cleansing Time table)
  - Beam duration (calculate range)

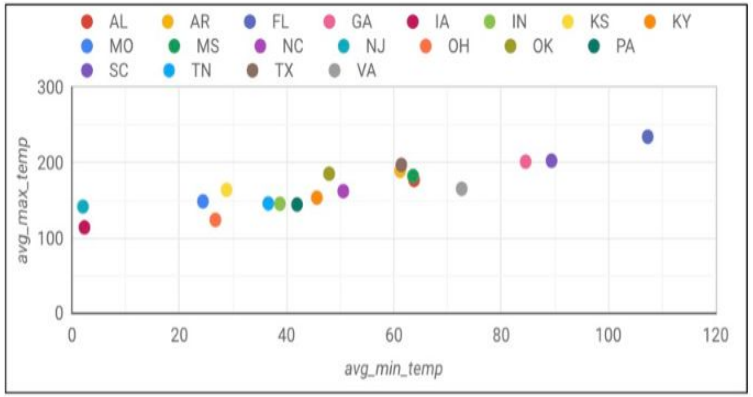
https://github.com/cs327e-spring2019/WLHA/blob/master/workflow.py

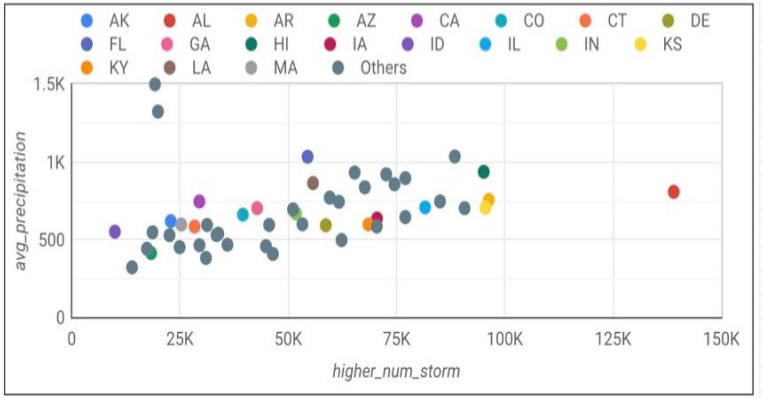
Visualizations

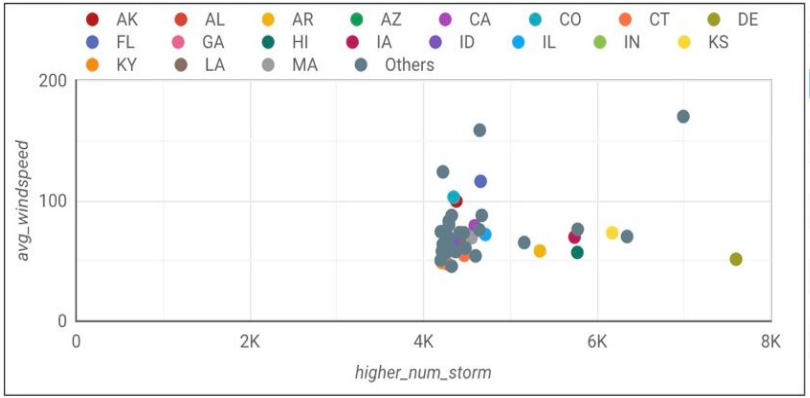







AL AR FL GA IA IN KS KY
MO MS NC NJ OH OK PA
SC TN TX VA
300
200
100
0
avg_max_temp
0 20 40 60 80 100 120
avg_min_temp


AK AL AR AZ CA CO CT DE
FL GA HI IA ID IL IN KS
KY LA MA Others
1.5K
1K
500
0
avg_precipitation
0 25K 50K 75K 100K 125K 150K
higher_num_storm


AK AL AR AZ CA CO CT DE
FL GA HI IA ID IL IN KS
KY LA MA Others
200
100
0
avg_windspeed
0 2K 4K 6K 8K
higher_num_storm

Future Improvements

- Analyze weather indicators of each storm for more specific relationship
  - Focus on each storm event, not on each state/ year
- Minimize the locations (within the U.S)
  - Major locations: the U.S
  - Some other locations out of the U.S.
  - Hard to compare locations
  - Beam transform to filter them
- Convert the units of weather indicators
  - The units were different than we usually know
  - Hard to interpret easily (big number values)

Thanks to all of you !