BigQuery Mini-Project

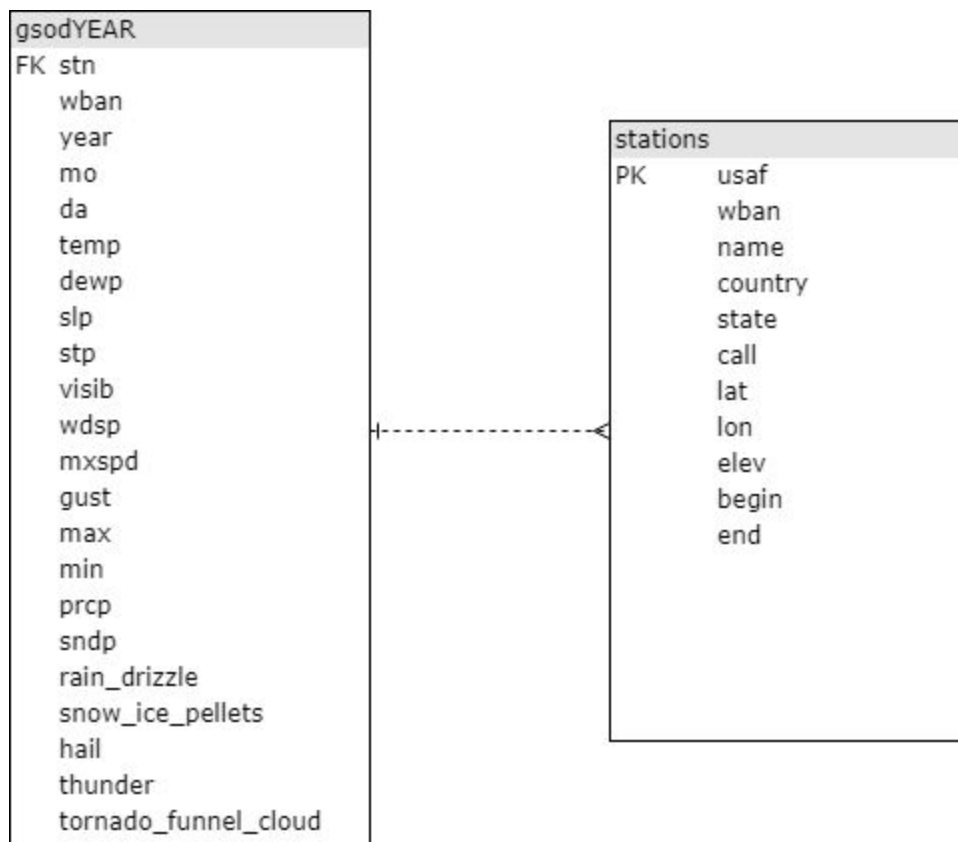*Marcellus Mohanna*

Dylan Williams

Data Set

NOAA - GSOD

The dataset describes weather observations observed at multiple NOAA weather stations across all continents of the world for every day from 1929 to 2019. There are many stations each with a physical location, name, and elevation.
Each data entry contains the station that reported it, the date reported, temperature, wind, and boolean values for reports of rain, hail, tornados, etc.

Schema

Note: 'gsodYEAR' represents 'gsod2001', 'gsod2002, 'gsod2003…'

Data Demand:

"The top 25 states for the average number of tornadoes per year from 2008-2018 ordered by descending frequency."

Query Results

Row	tornados_per_year	state
1	41.5	FL
2	6.3	CA
3	5.9	TX
4	4.8	LA
5	4.1	CO
6	4.1	HI
7	3.2	MS
8	2.5	KS
9	2.2	SC
10	2.1	OK
11	2.0	AL
12	1.6	AK
13	1.4	MA
14	1.4	NC
15	1.3	ND
16	1.2	NM
17	1.2	GA
18	1.1	MO
19	1.1	AR
20	1.0	VA
21	0.9	KY
22	0.7	IL
23	0.7	WY
24	0.6	CT
25	0.5	OR

(Query complete (1.126 sec elapsed, 370.76 MB processed)

This data does not accurately represent total tornado activity within the United States, and this can be explained by two main factors:

1. Tornado data is only recorded when initially reported by a station nearby. It is impossible for all stations to report all weather activity.
2. The number of stations in the United States are limited, and not even dispersed. Data on this can be shown below

Row	state	station_count
1	*null*	650
2	AK	564
3	CA	506
4	TX	500
5	FL	319
6	MI	232
7	NC	209
8	MN	205
9	LA	194
10	CO	170
11	VA	161
12	GA	160
13	WA	159
14	IL	150
15	WI	149
16	NY	137
17	AZ	128
18	AL	127
19	IA	125
20	NM	125
21	OK	118
22	SC	118
23	OR	115
24	NE	111
25	OH	108
26	PA	104
27	MO	102
28	MT	95
29	ND	95
30	KS	93
31	UT	91
32	WY	89
33	MS	88
34	AR	84
35	IN	80
36	ME	78
37	MA	78
38	MD	72
39	NV	70
40	ID	68
41	HI	66
42	TN	62
43	NJ	62
44	KY	62
45	SD	57
46	NH	41
47	CT	39
48	WV	39
49	RI	24
50	VT	23
51	DE	16
52	PR	4
53	DC	2

Our Full Query

```
SELECT count(state) / 10 as tornados_per_year, state
FROM (
        SELECT g08.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2008` g08
          JOIN `bigquery-public-data.noaa_gsod.stations` stn
          ON g08.stn  = stn.usaf
         WHERE stn.country = 'US' and g08.tornado_funnel_cloud = '1'
       UNION ALL

        SELECT g09.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2009` g09
          JOIN `bigquery-public-data.noaa_gsod.stations` stn
          ON g09.stn  = stn.usaf
          WHERE stn.country = 'US' and g09.tornado_funnel_cloud = '1'
           UNION ALL

        SELECT g10.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2010` g10
          JOIN `bigquery-public-data.noaa_gsod.stations` stn
          ON g10.stn  = stn.usaf
          WHERE stn.country = 'US' and g10.tornado_funnel_cloud = '1'
       UNION ALL

        SELECT g11.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2011` g11
          JOIN `bigquery-public-data.noaa_gsod.stations` stn
          ON g11.stn  = stn.usaf
          WHERE stn.country = 'US' and g11.tornado_funnel_cloud = '1'
       UNION ALL

        SELECT g12.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2011` g12
          JOIN `bigquery-public-data.noaa_gsod.stations` stn
          ON g12.stn  = stn.usaf
          WHERE stn.country = 'US' and g12.tornado_funnel_cloud = '1'
       UNION ALL
        SELECT g13.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2013` g13
          JOIN `bigquery-public-data.noaa_gsod.stations` stn
          ON g13.stn  = stn.usaf
          WHERE stn.country = 'US' and g13.tornado_funnel_cloud = '1'
       UNION ALL

        SELECT g14.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2014` g14
          JOIN `bigquery-public-data.noaa_gsod.stations` stn
```

```
            ON g14.stn  = stn.usaf
            WHERE stn.country = 'US' and g14.tornado_funnel_cloud = '1'
        UNION ALL

          SELECT g15.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2015` g15
            JOIN `bigquery-public-data.noaa_gsod.stations` stn
            ON g15.stn  = stn.usaf
            WHERE stn.country = 'US' and g15.tornado_funnel_cloud = '1'
        UNION ALL

          SELECT g16.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2016` g16
            JOIN `bigquery-public-data.noaa_gsod.stations` stn
            ON g16.stn  = stn.usaf
            WHERE stn.country = 'US' and g16.tornado_funnel_cloud = '1'
        UNION ALL

          SELECT g17.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2011` g17
            JOIN `bigquery-public-data.noaa_gsod.stations` stn
            ON g17.stn  = stn.usaf
            WHERE stn.country = 'US' and g17.tornado_funnel_cloud = '1'
        UNION ALL

          SELECT g18.stn, stn.state FROM `bigquery-public-data.noaa_gsod.gsod2018` g18
            JOIN `bigquery-public-data.noaa_gsod.stations` stn
            ON g18.stn  = stn.usaf
            WHERE stn.country = 'US' and g18.tornado_funnel_cloud = '1'
)
 GROUP BY state
 ORDER BY tornados_per_year DESC
 LIMIT 25
```