

Eliza Baier
u1559546
IS 6420: Database Theory and Design
October 25, 2024

Utah Air Quality Study Findings and Recommendations

In analyzing Air Quality Data from 2003, 2013, and 2023, we found that the overall highest average AQI numbers, or worse air quality, tended to occur in the summer months. This holds true in our Utah-specific findings as well. Utah counties did not appear in the top 10 worst average air quality for 2003, 2013, or 2023, indicating that other locations, mostly in California, have larger problems with air quality. However, while California had the month number of counties appearing in the worst air quality table, they also dominated the most improved table. Unfortunately, 2 UT counties - Uintah and Duchesne - appeared near the top of the 'Worst Decline in Air Quality' table. Perhaps there are certain interventions implemented in California that Utah state could also implement to achieve similar results, especially in Uintah and Duchesne counties. Analysis of unhealthy air days in 2003, 2013, and 2023 showed that the number of unhealthy air days almost doubled from 2003 to 2013 but decreased to almost one-third from 2013 to 2023. Overall, 2023 had the lower number of unhealthy air days at 128. There is still much improvement needed but the interventions taken from 2013 to 2023 seem to have worked well. Comparison of unhealthy air days by month in Salt Lake County yielded additional insights. In 2003, the month with the most unhealthy air days was July, fitting with our original analysis, with 19 days of unhealthy (AQI > 100) air. In 2023, the month with the most unhealthy air days was August, with only 7 unhealthy days overall. In 2023, July, the month with the most unhealthy air days in 2003, had dropped to just 4 days of unhealthy air. This result shows a vast improvement in air quality, especially in the summer months. As anyone who has spent a winter in Salt Lake City will know, air quality seems especially bad during the 'inversion' months in the winter. In 2003, these months had a higher incidence of unhealthy air days but by 2023, these months had dropped closer to the bottom of the ranking. This indicates a significant improvement in air quality during inversion months. In conclusion, UT is improving overall air quality. Continued improvement in air quality will become increasingly important given the rapid rate of growth and increasing infrastructure that UT, especially Salt Lake County, has been seeing and will continue to see with the upcoming Olympic Games.

Results Tables

AQI by Year and Season

| | <small>A-Z</small> season | <small>123</small> avg_aqi | <small>A-Z</small> year |
|----|---------------------------|----------------------------|-------------------------|
| 1 | Fall | 43.88 | 2003 |
| 2 | Spring | 48.84 | 2003 |
| 3 | Summer | 57.62 | 2003 |
| 4 | Winter | 38.77 | 2003 |
| 5 | Fall | 40.75 | 2013 |
| 6 | Spring | 45.12 | 2013 |
| 7 | Summer | 46.78 | 2013 |
| 8 | Winter | 43.74 | 2013 |
| 9 | Fall | 40.38 | 2023 |
| 10 | Spring | 45.58 | 2023 |
| 11 | Summer | 52.45 | 2023 |
| 12 | Winter | 39.09 | 2023 |

Top 10 Worst AQI by Year

| | <small>A-Z</small> state | <small>A-Z</small> county | <small>123</small> avg_aqi | <small>A-Z</small> year |
|----|--------------------------|---------------------------|----------------------------|-------------------------|
| 1 | California | Inyo | 196.75 | 2003 |
| 2 | California | Kern | 122.36 | 2003 |
| 3 | California | Fresno | 113.27 | 2003 |
| 4 | California | Riverside | 110.86 | 2003 |
| 5 | California | San Bernardino | 110.13 | 2003 |
| 6 | California | Los Angeles | 109.39 | 2003 |
| 7 | California | Tulare | 108.26 | 2003 |
| 8 | California | Mono | 103.48 | 2003 |
| 9 | Country Of Mexico | BAJA CALIFORNIA | 100.84 | 2003 |
| 10 | California | Merced | 83.98 | 2003 |
| 11 | California | Riverside | 96.07 | 2013 |
| 12 | California | Kern | 94.89 | 2013 |
| 13 | California | San Bernardino | 94.35 | 2013 |
| 14 | Nevada | Washoe | 93.68 | 2013 |
| 15 | California | Fresno | 93.6 | 2013 |
| 16 | California | Tulare | 91.72 | 2013 |
| 17 | California | Los Angeles | 89.21 | 2013 |
| 18 | California | Kings | 82.88 | 2013 |
| 19 | Arizona | Pinal | 80.94 | 2013 |
| 20 | California | Madera | 79.74 | 2013 |
| 21 | California | Riverside | 93.01 | 2023 |
| 22 | California | San Bernardino | 89.88 | 2023 |
| 23 | California | Los Angeles | 83.22 | 2023 |
| 24 | Idaho | Bannock | 79.05 | 2023 |
| 25 | Arizona | Maricopa | 78.19 | 2023 |
| 26 | California | Tulare | 76.68 | 2023 |
| 27 | California | Kern | 76.18 | 2023 |
| 28 | California | Inyo | 72.7 | 2023 |
| 29 | California | Imperial | 72.47 | 2023 |
| 30 | Texas | Harris | 72.47 | 2023 |

Top 10 Improved AQI

| | A-Z state | A-Z county | 123 avg_aqi_2003 | 123 avg_aqi_2023 | 123 aqi_change |
|----|------------|-------------|------------------|------------------|----------------|
| 1 | California | Inyo | 196.75 | 72.7 | 124.05 |
| 2 | California | Mono | 103.48 | 39.02 | 64.46 |
| 3 | California | Kern | 122.36 | 76.18 | 46.18 |
| 4 | California | Fresno | 113.27 | 70.94 | 42.33 |
| 5 | Michigan | Monroe | 59.71 | 27.18 | 32.53 |
| 6 | California | Tulare | 108.26 | 76.68 | 31.58 |
| 7 | California | Merced | 83.98 | 56.5 | 27.48 |
| 8 | California | El Dorado | 74.46 | 47.28 | 27.18 |
| 9 | California | Los Angeles | 109.39 | 83.22 | 26.17 |
| 10 | California | Sacramento | 81.24 | 55.66 | 25.58 |

Top 10 Declined AQI

| | A-Z state | A-Z county | 123 avg_aqi_2003 | 123 avg_aqi_2023 | 123 aqi_change |
|----|--------------|------------|------------------|------------------|----------------|
| 1 | Idaho | Bannock | 23.09 | 79.05 | 55.96 |
| 2 | Utah | Uintah | 7.02 | 62.06 | 55.04 |
| 3 | Utah | Duchesne | 8.58 | 60.78 | 52.2 |
| 4 | Colorado | Garfield | 16.74 | 57.48 | 40.74 |
| 5 | Wisconsin | Forest | 10.12 | 45.31 | 35.19 |
| 6 | Wisconsin | Eau Claire | 15.15 | 49.21 | 34.06 |
| 7 | Colorado | Jackson | 11.88 | 45.46 | 33.58 |
| 8 | North Dakota | Burke | 18.33 | 51.49 | 33.16 |
| 9 | Montana | Cascade | 17.58 | 50.6 | 33.02 |
| 10 | Kansas | Neosho | 22.26 | 53.55 | 31.29 |

UT Unhealthy Air Days by Year

| | 123 days_of_unhealthy_air | A-Z year |
|---|---------------------------|----------|
| 1 | 128 | 2023 |
| 2 | 356 | 2013 |
| 3 | 184 | 2003 |

UT Salt Lake County Unhealthy Air Days by Month 2003

| | 123 month | 123 unhealthy_days |
|----|-----------|--------------------|
| 1 | 7 | 19 |
| 2 | 6 | 7 |
| 3 | 5 | 4 |
| 4 | 12 | 3 |
| 5 | 8 | 3 |
| 6 | 4 | 2 |
| 7 | 1 | 2 |
| 8 | 11 | 1 |
| 9 | 2 | 1 |
| 10 | 10 | 1 |
| 11 | 9 | 0 |
| 12 | 3 | 0 |

UT Salt Lake County Unhealthy Air Days by Month 2023

| | 123 month | 123 unhealthy_days |
|----|-----------|--------------------|
| 1 | 8 | 7 |
| 2 | 7 | 4 |
| 3 | 2 | 3 |
| 4 | 6 | 2 |
| 5 | 5 | 2 |
| 6 | 12 | 1 |
| 7 | 9 | 1 |
| 8 | 3 | 0 |
| 9 | 1 | 0 |
| 10 | 10 | 0 |
| 11 | 11 | 0 |
| 12 | 4 | 0 |

SQL Code

```

-- create daily_aqi_2023 table
drop table if exists daily_aqi_2023;
create table daily_aqi_2023 (
  state_name varchar(100),
  county_name varchar(100),
  state_code char(2),
  county_code char(3),
  date date,
  aqi integer,
  category varchar(100),
  defining_parameter varchar(25),
  defining_site varchar(20),
  number_of_sites_reporting integer
);

-- create daily_aqi_2013 table
drop table if exists daily_aqi_2013;
create table daily_aqi_2013 (
  state_name varchar(100),
  county_name varchar(100),
  state_code char(2),
  county_code char(3),
  date date,
  aqi integer,
  category varchar(100),
  defining_parameter varchar(25),
  defining_site varchar(20),
  number_of_sites_reporting integer
);

-- create daily_aqi_2003 table
drop table if exists daily_aqi_2003;
create table daily_aqi_2003 (
  state_name varchar(100),
  county_name varchar(100),
  state_code char(2),
  county_code char(3),
  date date,
  aqi integer,
  category varchar(100),
  defining_parameter varchar(25),
  defining_site varchar(20),
  number_of_sites_reporting integer
);

--create season function
create or replace function get_season(date_input date)
returns text as $$
begin

```

```

    return case
        when extract(month from date_input) in (12, 1, 2) then 'Winter'
        when extract(month from date_input) in (3, 4, 5) then 'Spring'
        when extract(month from date_input) in (6, 7, 8) then 'Summer'
        when extract(month from date_input) in (9, 10, 11) then 'Fall'
    end;
end;
$$ language plpgsql;
-- test season function
select
    date,
    get_season(date) as season
from
    daily_aqi_2023;
--avg aqi in 2023 by season
select
    get_season(date) as season,
    round(avg(aqi), 2) as avg_aqi_2023
from
    daily_aqi_2023
group by
    season;
-- avg aqi in 2013 by season
select
    get_season(date) as season,
    round(avg(aqi), 2) as avg_aqi_2013
from
    daily_aqi_2013
group by
    season;
-- avg aqi in 2003 by season
select
    get_season(date) as season,
    round(avg(aqi), 2) as avg_aqi_2003
from
    daily_aqi_2003
group by
    season;
-- combine all queries with union all
--avg aqi in 2023 by season
select
    get_season(date) as season,
    round(avg(aqi), 2) as avg_aqi,
    '2023' as year
from
    daily_aqi_2023

```


group by

season

union all

-- avg aqi in 2013 by season

select

get_season(**date**) **as** season,
round(avg(aqi), 2) **as** avg_aqi,
'2013' **as** year

from

daily_aqi_2013

group by

season

union all

-- avg aqi in 2003 by season

select

get_season(**date**) **as** season,
round(avg(aqi), 2) **as** avg_aqi,
'2003' **as** year

from

daily_aqi_2003

group by

season

order by

year,
season;

-- top 10 locations with worst (highest) aqi in each year

with ranked_aqi as

(**select**

state_name **as** state,
county_name **as** county,
aqi,
'2023' **as** year,
row_number() **over** (**partition by** '2023' **order by** aqi **desc**) **as** rank

from

daily_aqi_2023

union all

select

state_name **as** state,
county_name **as** county,
aqi,
'2013' **as** year,
row_number() **over** (**partition by** '2013' **order by** aqi **desc**) **as** rank

from

daily_aqi_2013

union all

select

```

state_name as state,
county_name as county,
aqi,
'2003' as year,
row_number() over (partition by '2003' order by aqi desc) as rank
from
daily_aqi_2003)

select
state,
county,
aqi,
year
from
ranked_aqi
where
rank <= 10
order by
year asc,
aqi desc;
-- top 10 locations with worst (highest) avg aqi in each year
with avg_aqi as
(select
state_name as state,
county_name as county,
round(avg(aqi), 2) as avg_aqi,
'2023' as year
from
daily_aqi_2023
group by
state_name,
county_name
union all
select
state_name as state,
county_name as county,
round(avg(aqi), 2) as avg_aqi,
'2013' as year
from
daily_aqi_2013
group by
state_name,
county_name
union all
select
state_name as state,

```

```

        county_name as county,
        round(avg(aqi), 2) as avg_aqi,
        '2003' as year
    from
        daily_aqi_2003
    group by
        state_name,
        county_name),
ranked_results as
    (select
        state,
        county,
        avg_aqi,
        year,
        row_number() over (partition by year order by avg_aqi desc) as rank

    from
        avg_aqi)
select
    state,
    county,
    avg_aqi,
    year
from ranked_results
where rank <= 10
order by
    year asc,
    avg_aqi desc;

```

-- top 10 locations with greatest improvement of average AQI from 2003 to 2023

```

with aqi_summary as
    (select
        yr2003.state_name,
        yr2003.county_name,
        round(avg(yr2003.aqi), 2) as avg_aqi_2003,
        round(avg(yr2023.aqi), 2) as avg_aqi_2023
    from
        daily_aqi_2003 yr2003
    join
        daily_aqi_2023 yr2023
    on
        yr2003.state_name = yr2023.state_name
        and yr2003.county_name = yr2023.county_name
    group by
        yr2003.state_name,
        yr2003.county_name),
aqi_change as

```

```

(select
    state_name,
    county_name,
    avg_aqi_2003,
    avg_aqi_2023,
    (avg_aqi_2003 - avg_aqi_2023) AS aqi_change
from
    aqi_summary)
select
    state_name as state,
    county_name as county,
    avg_aqi_2003,
    avg_aqi_2023,
    aqi_change
from
    aqi_change
where
    aqi_change > 0
order by
    aqi_change desc
limit 10;
-- top 10 locations with worst decline (increase) in average AQI from 2003 to 2023
with aqi_summary as
(select
    yr2003.state_name,
    yr2003.county_name,
    round(avg(yr2003.aqi), 2) AS avg_aqi_2003,
    round(avg(yr2023.aqi), 2) AS avg_aqi_2023
from
    daily_aqi_2003 yr2003
join
    daily_aqi_2023 yr2023
on
    yr2003.state_name = yr2023.state_name
    and yr2003.county_name = yr2023.county_name
group by
    yr2003.state_name,
    yr2003.county_name),
aqi_change as
(select
    state_name,
    county_name,
    avg_aqi_2003,
    avg_aqi_2023,
    (avg_aqi_2023 - avg_aqi_2003) as aqi_change
from

```

```

    aqi_summary)
select
    state_name as state,
    county_name as county,
    avg_aqi_2003,
    avg_aqi_2023,
    aqi_change
from
    aqi_change
where
    aqi_change > 0
order by
    aqi_change desc
limit 10;

```

-- In Utah counties, how many days of "Unhealthy" air did we have in each year? Is it improving?

```

select
    count(date) as days_of_unhealthy_air,
    '2003' as year
from
    daily_aqi_2003
where
    aqi >= 100 and
    state_name = 'Utah'
union all
select
    count(date) as days_of_unhealthy_air,
    '2013' as year
from
    daily_aqi_2013
where
    aqi >= 100 and
    state_name = 'Utah'
union all
select
    count(date) as days_of_unhealthy_air,
    '2023' as year
from
    daily_aqi_2023
where
    aqi >= 100 and
    state_name = 'Utah';

```

-- In Salt Lake County, which months have the most "Unhealthy" days? Has that changed in 20 years?

-- Unhealthiest months in 2023

```

select
    extract(month from date) as month,
    coalesce(count(date), 0) as unhealthy_days

```

```

from daily_aqi_2023
where
    aqi >= 100 and
    county_name = 'Salt Lake'
group by
    extract(month from date)
order by
    unhealthy_days desc;
with all_months as (
    select generate_series(1, 12) as month -- This generates numbers 1 through 12 (for each month)
)
select
    m.month,
    coalesce(count(d.date), 0) as unhealthy_days
from
    all_months m
left join
    daily_aqi_2023 d
    on extract(month from d.date) = m.month
    and d.aqi >= 100
    and d.county_name = 'Salt Lake'
group by
    m.month
order by
    unhealthy_days desc;

```

-- Unhealthiest months from 2003

```

select
    extract(month from date) as month,
    coalesce(count(date), 0) as unhealthy_days
from daily_aqi_2003
where
    aqi >= 100 and
    county_name = 'Salt Lake'
group by
    extract(month from date)
order by
    unhealthy_days desc;
with all_months as (
    select generate_series(1, 12) as month -- This generates numbers 1 through 12 (for each month)
)
select
    m.month,
    coalesce(count(d.date), 0) as unhealthy_days
from
    all_months m

```

left join

daily_aqi_2003 d

on **extract**(**month from** *d.date*) = *m.month*

and *d.aqi* >= 100

and *d.county_name* = 'Salt Lake'

group by

m.month

order by

unhealthy_days **desc**;