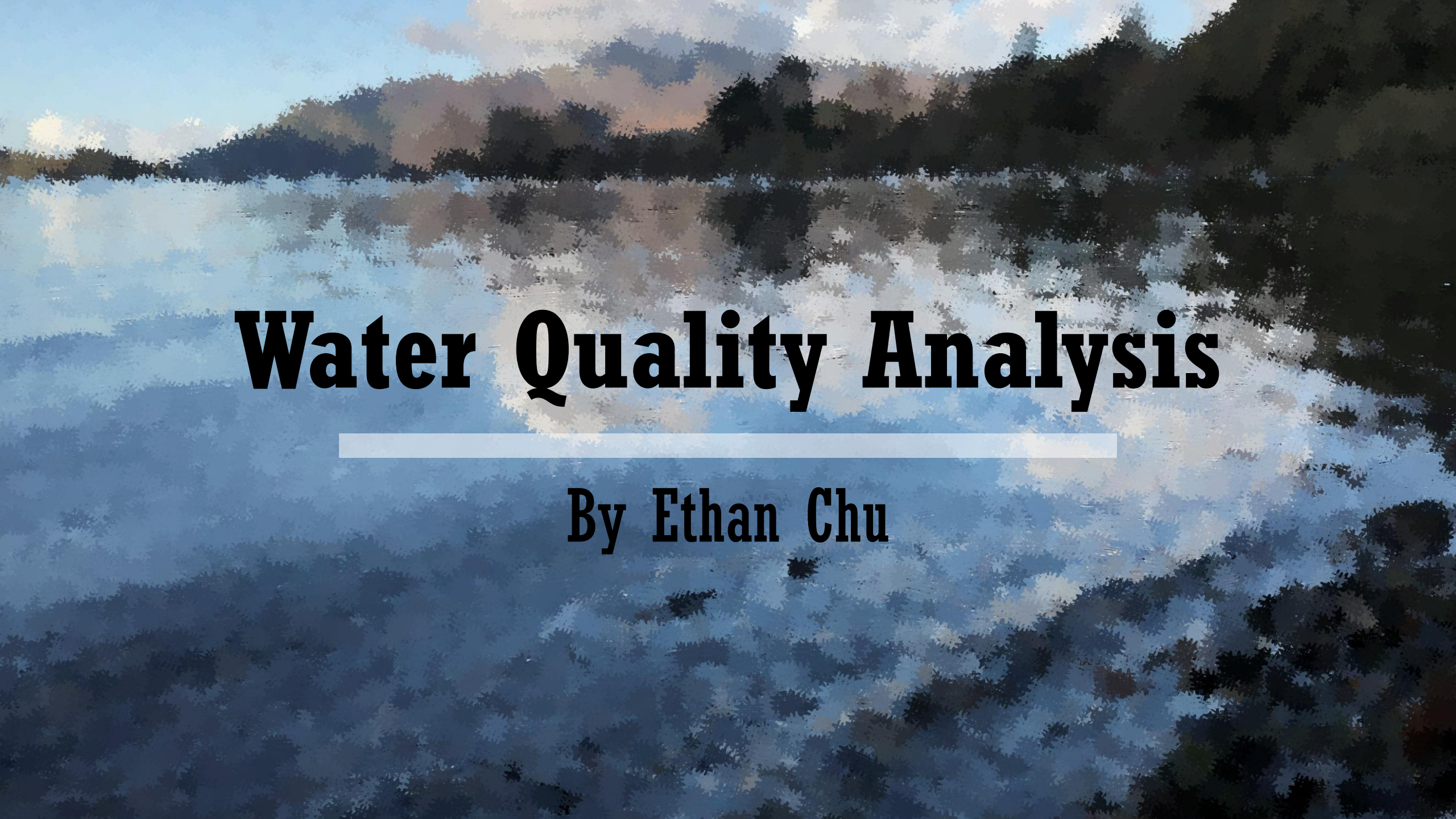

Water Quality Analysis

By Ethan Chu

Creating the Table

```sql
CREATE TABLE water_csv (
    site_id VARCHAR(5),
    time_24 TIME,
    year VARCHAR(4),
    read_date DATE,
    salinity_ppt FLOAT,
    dissolved_oxygen_mg_l FLOAT,
    pH FLOAT,
    secchi_depth_m FLOAT,
    water_depth_m FLOAT,
    water_temp_c FLOAT,
    air_temp_c FLOAT,
    field_tech VARCHAR(50)
)
```

How Many Readings Were There Per Site?

```sql
SELECT site_id, COUNT(read_date) AS num_measurements
FROM water_csv
GROUP BY site_id
ORDER BY num_measurements DESC;
```

	site_id character varying (5)	num_measurements bigint
1	Bay	775
2	B	428
3	D	426
4	A	420
5	C	254

What Was The Average Water Temperature Per Site?

```sql
SELECT site_id, ROUND(AVG(water_temp_c) :: NUMERIC, 2) AS average_water_temp_c
FROM water_csv
GROUP BY site_id
ORDER BY average_water_temp_c DESC;
```

	site_id character varying (5)	average_water_temp_c numeric
1	A	18.66
2	B	18.35
3	D	18.24
4	C	18.18
5	Bay	17.51

What Was The Average Water Temperature Per Site Per Year?

```sql
SELECT site_id, year, ROUND(AVG(water_temp_c) :: NUMERIC, 2) AS average_water_temperature_c
FROM water_csv
GROUP BY site_id, year
ORDER BY site_id, year;
```

	site_id character varying (5)	year character varying (4)	average_water_temp_c numeric
1	A	1999	21.43
2	A	2000	20.18
3	A	2001	19.13
4	A	2002	17.71
5	A	2003	18.89
6	A	2004	20.29
7	A	2005	17.79
8	A	2006	18.30
9	A	2007	18.55
10	A	2008	17.24

- Only the first 10 entries shown!

What Year Was The Average Water Temperature Highest Per Site?

```sql
WITH average_water_temp_per_site_per_year AS (
    SELECT site_id, year, ROUND(AVG(water_temp_c) :: NUMERIC, 2) AS average_water_temp_c
    FROM water_csv
    GROUP BY site_id, year
    ORDER BY site_id, year
)

SELECT site_id, year, average_water_temp_c
FROM (
    SELECT *, RANK() OVER(PARTITION BY site_id ORDER BY average_water_temp_c DESC) AS rank
    FROM average_water_temp_per_site_per_year
)
WHERE rank = 1;
```

	site_id character varying (5)	year character varying (4)	average_water_temp_c numeric
1	A	2019	21.73
2	B	2019	22.75
3	Bay	1989	20.82
4	C	2014	27.00
5	D	1999	21.00

What Year Was The Average Water Temperature Lowest Per Site?

```sql
WITH average_water_temp_per_site_per_year AS (
    SELECT site_id, year, ROUND(AVG(water_temp_c) :: NUMERIC, 2) AS average_water_temp_c
    FROM water_csv
    GROUP BY site_id, year
    ORDER BY site_id, year
)

SELECT site_id, year, average_water_temp_c
FROM (
    SELECT *, RANK() OVER(PARTITION BY site_id ORDER BY average_water_temp_c ASC) AS rank
    FROM average_water_temp_per_site_per_year
)
WHERE rank = 1;
```

	site_id character varying (5)	year character varying (4)	average_water_temp_c numeric
1	A	2016	15.40
2	B	2016	14.60
3	Bay	2016	14.13
4	C	2011	13.75
5	D	2016	14.14

What Year Did Each Site Record Its Highest Water Temperature?

```sql
SELECT site_id, year, water_temp_c
FROM (
    SELECT *, RANK() OVER(PARTITION BY site_id ORDER BY water_temp_c DESC) AS rank
    FROM water_csv
)
WHERE rank = 1;
```

	site_id character varying (5)	year character varying (4)	water_temp_c double precision
1	A	2004	60
2	B	2005	35
3	Bay	2005	74
4	C	1999	43
5	D	2007	54

Which Field Tech (or group of field techs) Had The Most Readings?

```sql
SELECT field_tech, COUNT(read_date) AS num_entries
FROM water_csv
GROUP BY field_tech;
```

	field_tech character varying (50)	num_entries bigint
1	Phillips	17
2	Phillips, Feldman	37
3	Poe	768
4	Feldman	167
5	Strader, Poe	10
6	Pease, Strader	31
7	Strader, Pease, Feldman	33
8	Strader	15
9	No Field Tech Recorded	1225

What Was The Highest Water Temperature Each Field Tech Recorded?

```sql
SELECT field_tech, water_temp_c
FROM(
    SELECT field_tech, water_temp_c,
    RANK() OVER(PARTITION BY field_tech ORDER BY water_temp_c DESC) AS rank
    FROM water_csv
    GROUP BY field_tech, water_temp_c
)
WHERE rank = 1;
```

	field_tech character varying (50)	water_temp_c double precision
1	Feldman	32
2	No Field Tech Recorded	60
3	Pease, Strader	29
4	Phillips	30
5	Phillips, Feldman	27
6	Poe	74
7	Strader	29
8	Strader, Pease, Feldman	28.5
9	Strader, Poe	29

What Was The Greatest Change In Water Temperature Per Site And During What Years Did It Happen?

```sql
WITH yearly_avg_water_temps AS (
    SELECT site_id,
    year::NUMERIC - 1 AS previous_year,
    year, AVG(water_temp_c) AS avg_water_temp
    FROM water_csv
    GROUP BY site_id, year
),
avg_temp_diffs AS (
    SELECT site_id, previous_year || '-' || year AS timeframe,
    LAG(avg_water_temp) OVER(PARTITION BY site_id ORDER BY year) AS prev_water_temp,
    avg_water_temp - LAG(avg_water_temp) OVER(PARTITION BY site_id ORDER BY year)
    AS temp_difference
    FROM yearly_avg_water_temps
)

SELECT site_id, timeframe, temp_difference
FROM (
    SELECT *,
    RANK() OVER(PARTITION BY site_id ORDER BY ABS(temp_difference) DESC) AS rank
    FROM avg_temp_diffs
    WHERE temp_difference IS NOT NULL
)
WHERE rank = 1;
```

What Was The Greatest Change In Water Temperature Per Site And During What Years Did It Happen? Cont.

	site_id character varying (5)	timeframe text	temp_difference double precision
1	A	2016-2017	4.15
2	B	2018-2019	4.645
3	Bay	1989-1990	-4.282988871224166
4	C	2013-2014	13.25
5	D	2016-2017	4.428571428571431

Which Site Was The Hottest On Average Each Year?

```sql
WITH yearly_avg_water_temps AS (
    SELECT site_id, year, AVG(water_temp_c) AS avg_water_temp
    FROM water_csv
    GROUP BY site_id, year
),

yearly_highs_lows AS (
    SELECT *,
    RANK() OVER(PARTITION BY year ORDER BY avg_water_temp DESC) AS high_rank,
    RANK() OVER(PARTITION BY year ORDER BY avg_water_temp ASC) AS low_rank
    FROM yearly_avg_water_temps
)

SELECT year, site_id AS hottest_site
FROM yearly_highs_lows
WHERE high_rank = 1
ORDER BY year;
```

Which Site Was The Hottest On Average Each Year? Cont.

	year character varying (4)	hottest_site character varying (5)
1	1989	Bay
2	1990	Bay
3	1991	Bay
4	1992	Bay
5	1993	Bay
6	1994	Bay
7	1995	Bay
8	1996	Bay
9	1997	Bay
10	1998	Bay
11	1999	C
12	2000	A
13	2001	C
14	2002	Bay
15	2003	A
16	2004	A
17	2005	B
18	2006	B
19	2007	D

- Only the first 19 entries shown!

What About The Coldest Site For Each Year?

```sql
WITH yearly_avg_water_temps AS (
    SELECT site_id, year, AVG(water_temp_c) AS avg_water_temp
    FROM water_csv
    GROUP BY site_id, year
),

yearly_highs_lows AS (
    SELECT *,
    RANK() OVER(PARTITION BY year ORDER BY avg_water_temp DESC) AS high_rank,
    RANK() OVER(PARTITION BY year ORDER BY avg_water_temp ASC) AS low_rank
    FROM yearly_avg_water_temps
)

SELECT year, site_id AS coldest_site
FROM yearly_highs_lows
WHERE low_rank = 1
ORDER BY year;
```

What About The Coldest Site For Each Year? Cont.

	year character varying (4)	coldest_site character varying (5)
1	1989	Bay
2	1990	Bay
3	1991	Bay
4	1992	Bay
5	1993	Bay
6	1994	Bay
7	1995	Bay
8	1996	Bay
9	1997	Bay
10	1998	Bay
11	1999	Bay
12	2000	D
13	2001	Bay
14	2002	C
15	2003	Bay
16	2004	C
17	2005	A
18	2006	Bay
19	2007	Bay

- Only the first 19 entries shown!

How Many Days Was The Water "Cold" In Each Site?

```sql
SELECT site_id,
SUM(CASE
    WHEN water_temp_c < 10 THEN 1
    ELSE 0
END) AS cold_days
FROM water_csv
GROUP BY site_id
ORDER BY cold_days DESC;
```

	site_id character varying (5)	cold_days bigint
1	Bay	149
2	D	76
3	A	72
4	B	71
5	C	46

Which Site Had The Longest Streak Of Cold Water?

```sql
WITH cold_days AS (
    SELECT site_id, read_date,
    CASE
        WHEN water_temp_c < 10 THEN 1
        ELSE 0
    END AS is_cold
    FROM water_csv
),

cold_count AS (
    SELECT *,
    LAG(is_cold) OVER(PARTITION BY site_id ORDER BY read_date) AS was_cold_before
    FROM cold_days
),
```

Which Site Had The Longest Streak Of Cold Water? Cont.

```
16  cold_streak_changed AS (
17      SELECT *,
18      CASE
19          WHEN is_cold <> was_cold_before THEN 1
20          ELSE 0
21          END AS cold_streak_change
22      FROM cold_count
23  ),
24
25  cold_streak_identified AS (
26      SELECT *,
27      SUM(cold_streak_change) OVER(PARTITION BY site_id ORDER BY read_date)
28      AS cold_streak_identifier
29      FROM cold_streak_changed
30  ),
31
32  cold_streak_counts AS (
33      SELECT *,
34      ROW_NUMBER() OVER(PARTITION BY site_id, cold_streak_identifier ORDER BY read_date)
35      AS cold_streak_length
36      FROM cold_streak_identified
37  ),
```

Which Site Had The Longest Streak Of Cold Water? Cont.

```
39  cold_streaks_ranked AS (
40      SELECT *,
41      RANK() OVER(PARTITION BY site_id ORDER BY cold_streak_length DESC) AS rank
42      FROM cold_streak_counts
43  )
44
45  SELECT site_id, cold_streak_length
46  FROM cold_streaks_ranked
47  WHERE rank = 1;
```

	site_id character varying (5)	cold_streak_length bigint
1	A	27
2	B	21
3	Bay	39
4	C	20
5	D	32