## FunSQL A library for compositional construction

of SQL queries

https://github.com/MechanicalRabbit/FunSQL.jl

Clark C. Evans, Kyrylo Simonov

## FunSQL? Who Needs It?

Query Algebra

Correlated Queries

Aggregate & Window Functions

Conclusion

#### Find all patients born in or after 1970.



SELECT p.person\_id
FROM person p
WHERE p.year\_of\_birth >= 1970



```
function find_patients(conn)
    sql = """
    SELECT p.person_id
    FROM person p
    WHERE p.year_of_birth >= 1970
    """
    DBInterface.execute(conn, sql)
end
```

and

```
function find_patients(conn; start_year = nothing, end_year = nothing)
    sql =
    SELECT p.person_id
    FROM person p
    predicates = String[]
    if start year !== nothing
        push!(predicates, "p.year_of_birth >= $start_year")
    end
    if end_year !== nothing
        push!(predicates, "p.year_of_birth <= $end_year")</pre>
    end
    if !isempty(predicates)
        sql *= "\nWHERE " * join(predicates, " AND ")
    end
    DBInterface.execute(conn, sql)
end
```



## FunSQL? Who Needs It?

Query Algebra

Correlated Queries

Aggregate & Window Functions

Conclusion



```
location
using FunSQL: SQLTable
                                                                                      person
const person =
    SQLTable(name = :person,
                                                                            visit_occurrence
             columns = [:person_id, :year_of_birth, :location_id])
const location =
                                                                          condition occurrence
    SQLTable(name = :location,
             columns = [:location id, :city, :state, :zip])
const visit_occurrence =
    SQLTable(name = :visit_occurrence,
             columns = [:visit_occurrence_id, :person_id, :visit_concept_id,
                        :visit start date, :visit end date])
const condition_occurrence =
    SQLTable(name = :condition_occurrence,
             columns = [:condition_occurrence_id, :person_id, :condition_concept_id,
                        :condition_start_date, :condition_end_date])
```

#### Find all patients born in or after 1970.

using FunSQL: From, Get, Select, Where, render

FROM person p



FROM person p

WHERE p.year\_of\_birth >= 1970



SELECT p.person\_id
FROM person p
WHERE p.year\_of\_birth >= 1970

q = From(person)



q = From(person) |>
 Where(Get.year\_of\_birth .>= 1970)



q = From(person) |>
 Where(Get.year\_of\_birth .>= 1970) |>
 Select(Get.person\_id)

sql = render(q, dialect = :postgresql)

	person
PK	person_id
	year_of_birth
FK	location_id

```
q = From(person) |>
    Where(Get.year_of_birth .>= 1970) |>
    Select(Get.person_id)
```



unbound references

BornInOrAfter(Y) = Get.year\_of\_birth .>= Y

#### using FunSQL: Agg, Fun

```
SELECT p.person_id
FROM person p
WHERE p.year_of_birth >= 1970
```

#### Show patients with their state of residence.

using FunSQL: Join

PK person\_id
year\_of\_birth
FK location\_id

### location

PK location\_id city

state





FROM person p

JOIN location l

ON (p.location\_id = l.location\_id)





SELECT p.person\_id, l.state
FROM person p
JOIN location l
ON (p.location\_id = l.location\_id)







#### Find patients

- born in or after 1970
- living in Illinois





q<sub>p</sub> |> Join(q<sub>l</sub>, q<sub>p</sub>.location\_id .== q<sub>l</sub>.location\_id) |>
 Select(q<sub>p</sub>.person\_id)



#### FROM person p



FROM person p
WHERE p.year\_of\_birth >= 1970



FROM person p
WHERE p.year\_of\_birth >= 1970
JOIN location l
ON (p.location\_id = l.location\_id)

From(person)



From(person) |>
Where(Get.year\_of\_birth .>= 1970)



From(person) |>
Where(Get.year\_of\_birth .>= 1970) |>
Join(:location => From(location),
 Get.location\_id .==
 Get.location.location\_id)

			FROM (
		SELECT	SELECT
FROM	FROM	FROM	FROM
	JOIN	JOIN	JOIN
	WHERE	WHERE	WHERE
	GROUP BY	GROUP BY	GROUP BY
	HAVING	HAVING	HAVING
	ORDER BY	ORDER BY	ORDER BY)

SELECT ??? From(table) FROM \$table [ | | | **SELECT** ??? Where(condition) FROM ( ) WHERE \$condition |> Join( , *on*) **SELECT** ??? FROM ( ) JOIN ( ) ON \$on **SELECT** \$(list...) Select(list...) FROM ( )

```
SELECT ???
                                                                SELECT ???
                      FROM person
                                                                FROM location
                      SELECT ???
                                                                SELECT ???
                      FROM ( ) p
                                                                FROM ( ) l
                      WHERE p.year_of_birth >= 1970
                                                                WHERE l.state = 'IL'
q_1 = From(person)
                                                SELECT ???
                                                FROM (
q_2 = q_1 \mid > Where(q_1.year_of_birth .>= 1970)
                                                JOIN (
q_3 = From(location)
                                                  ON p.location_id = l.location_id
q_4 = q_3 \mid > Where(q_3.state .== "IL")
q_5 = q_2 \mid > Join(q_4, q_2.location_id .==
                     q<sub>4</sub>.location_id)
                                                SELECT p.person_id
q_6 = q_5 \mid > Select(q_5.person_id)
                                                FROM ( ) p
```

```
SELECT person_id, year_of_birth, location_id
                                                              SELECT location_id, state
      FROM person
                                                              FROM location
                 SELECT p.person_id, p.location_id
                                                              SELECT l.location_id
                  FROM ( ) p
                                                              FROM ( ) l
                 WHERE p.year_of_birth >= 1970
                                                              WHERE l.state = 'IL'
q_1 = From(person)
                                               SELECT p.person_id
                                               FROM ( ) p
q_2 = q_1 \mid > Where(q_1.year_of_birth .>= 1970)
                                               JOIN (
q_3 = From(location)
                                                ON p.location_id = l.location_id
q_4 = q_3 \mid > Where(q_3.state .== "IL")
q_5 = q_2 \mid > Join(q_4, q_2.location_id .==
                    q_4.location_id)
                                               SELECT p.person_id
q_6 = q_5 \mid > Select(q_5.person_id)
                                               FROM ( ) p
```

```
SELECT person_id, year_of_birth, location_id
FROM person
```

```
SELECT location_id, state
FROM location
```

```
SELECT p.person_id, p.location_id
FROM person p
WHERE p.year_of_birth >= 1970
```

SELECT l.location\_id
FROM location l
WHERE l.state = 'IL'

```
q_1 = From(person)

q_2 = q_1 |> Where(q_1.year_of_birth .>= 1970)

q_3 = From(location)

q_4 = q_3 |> Where(q_3.state .== "IL")

q_5 = q_2 |> Join(q_4, q_2.location_id .== q_4.location_id)

q_6 = q_5 |> Select(q_5.person_id)
```

```
SELECT p.person_id

FROM ( ) p

JOIN ( ) l

ON p.location_id = l.location_id
```

```
SELECT p.person_id
FROM ( ) p
```

#### Find patients

- born in or after 1970
- living in Illinois

```
q_1 = From(person)

q_2 = q_1 |> Where(q_1.year_of_birth .>= 1970)

q_3 = From(location)

q_4 = q_3 |> Where(q_3.state .== "IL")

q_5 = q_2 |> Join(q_4, q_2.location_id .== q_4.location_id)

q_6 = q_5 |> Select(q_5.person_id)
```

```
using FunSQL: AS, FROM, JOIN, OP, SELECT, WHERE
FROM(
  FROM(:person |> AS(:p)) |>
  WHERE(OP(">=", (:p, :year_of_birth), 1970)) |>
  SELECT((:p, :person_id),
         (:p, :location id)) |>
  AS(:p)) |>
JOIN(
  FROM(:location |> AS(:l)) |>
  WHERE(OP("=", (:l, :state), "IL")) |>
  SELECT((:l, :location_id)) |>
 AS(:1),
 OP("=", (:p, :location_id),
         (:l, :location_id))) |>
SELECT((:p, :person_id))
```

#### Find patients

- born in or after 1970
- living in Illinois

```
using FunSQL: Define
const ObservationYear = 2000
FromPerson() =
    From(person) |>
    Define(:approx_age => ObservationYear .- Get.year_of_birth)
FromAdult() =
    FromPerson() |>
    Where(Get.approx_age .>= 18)
julia> q = FromAdult() |> Select(Get.person_id)
```

## FunSQL? Who Needs It?

Query Algebra

Correlated Queries

Aggregate & Window Functions

Conclusion

#### Find all patients born in or after \$YEAR.

```
using FunSQL: Var, pack
```

```
sql = """
SELECT p.person_id
FROM person p
WHERE p.year_of_birth >= :YEAR
"""
params = (YEAR = 1970,)
```

```
q = From(person) |>
    Where(Get.year_of_birth .>= Var.YEAR) |>
    Select(Get.person_id)

sql = render(q, dialect = :sqlite)

params = pack(sql, (YEAR = 1970,))
```

DBInterface.execute(conn, sql, params)

Find patients with at least one medical condition.



condition_occurrence		
PK	condition_occurrence_id	
FK	person_id	
	condition_concept_id	
	condition_start_date	
	condition_end_date	

Find patients with at least one medical condition.

```
SELECT p.*
FROM person p
WHERE EXISTS (SELECT NULL
                 FROM condition_occurrence c
                 WHERE c.person_id = p.person_id)
q_p = From(person)
q<sub>c</sub> = From(condition_occurrence)
q_{corr} = q_c \mid > Where(q_c.person_id .== q_p.person_id)
q = q<sub>D</sub> |> Where(Fun.exists(q<sub>COFF</sub>))
```

# PK person\_id year\_of\_birth FK location\_id

condition_occurrence		
PK	condition_occurrence_id	
FK	person_id	
	condition_concept_id	
	condition_start_date	
	condition_end_date	

ERROR: Cannot find person\_id

#### Find patients with at least one medical condition.

SELECT NULL

CorrelatedCondition:  $X \mapsto FROM$  condition\_occurrence c

WHERE c.person\_id = X

SELECT p.\*
FROM person p
WHERE EXISTS CorrelatedCondition(p.person\_id)



condition_occurrence		
PK	condition_occurrence_id	
FK	person_id	
	condition_concept_id	
	condition_start_date	
	condition_end_date	

From(condition\_occurrence) |>
Where(Get.person\_id .== Var.X)



using FunSQL: Bind

CorrelatedCondition(X) =
 From(condition\_occurrence) |>
 Where(Get.person\_id .== Var.X) |>
 Bind(:X => X)

CorrelatedCondition(6)

SELECT c.\*
FROM condition\_occurrence c
WHERE c.person\_id = :X

SELECT c.\*
FROM condition\_occurrence c
WHERE c.person\_id = 6

## FunSQL? Who Needs It?

Query Algebra

Correlated Queries

Aggregate & Window Functions

Conclusion

#### Number of patients by the year of birth.

using FunSQL: Agg, Group

FROM person p



FROM person p
GROUP BY p.year\_of\_birth

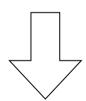


SELECT p.year\_of\_birth, COUNT(\*)
FROM person p
GROUP BY p.year\_of\_birth

From(person)



From(person) |>
Group(Get.year\_of\_birth)



From(person) |>
Group(Get.year\_of\_birth) |>
Select(Get.year\_of\_birth, Agg.count())

#### Average year of birth.

FROM person p



SELECT AVG(p.year\_of\_birth)
FROM person p

From(person)



From(person) |>
Group()



From(person) |>
Group() |>
Select(Agg.avg(Get.year\_of\_birth))

#### Patients who saw a doctor within the last 12 months.

FROM visit\_occurrence v



FROM visit\_occurrence v
GROUP BY v.person\_id



 From(visit\_occurrence)



From(visit\_occurrence) |>
Group(Get.person\_id)



visit_occurrence		
PK	visit_occurrence_id	
FK	person_id	
	visit_concept_id	
	visit_start_date	
	visit_end_date	

Patients who saw a doctor within the last 12 months.

```
SELECT person_id, visit_start_date
FROM visit occurrence
SELECT v.person_id,
      MAX(v.visit_end_date) AS max
FROM ( ) v
GROUP BY v.person_id
SELECT g.person_id
FROM ( ) g
WHERE CURRENT_DATE - g.max <= 365</pre>
```

#### Patients who saw a doctor within the last 12 months.

```
SELECT person_id, visit_end_date
FROM visit_occurrence
```

For each visit, show the time passed since the previous visit.

FROM visit\_occurrence v



FROM visit\_occurrence v
WINDOW w AS (
 PARTITION BY v.person\_id
 ORDER BY v.visit\_start\_date
 ROWS BETWEEN UNBOUNDED PRECEEDING AND
 1 PRECEEDING)

using FunSQL: Partition

From(visit\_occurrence)



finish = -1)

PK visit\_occurrence\_id

FK person\_id
 visit\_concept\_id
 visit\_start\_date

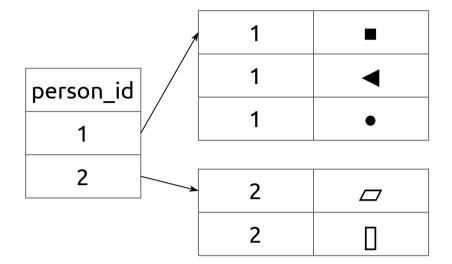
For each visit, show the time passed since the previous visit.

visit_occurrence		
PK	visit_occurrence_id	
FK	person_id	
	visit_concept_id	
	visit_start_date	
	visit_end_date	

```
From(visit_occurrence) |>
Group(Get.person_id) |>
Where(Fun.current_date() .-
      Agg.max(Get.visit_end_date) .<= 365)</pre>
From(visit_occurrence) |>
Partition(Get.person_id,
          order_by = [Get.visit_start_date],
          frame = (mode = :rows,
                   start = -Inf,
                   finish = -1)) |>
Select(Get.visit_start_date,
       :gap => Get.visit_start_date .-
               Agg.max(Get.visit end date))
```

From(visit\_occurrence)

#### Group(Get.person\_id)



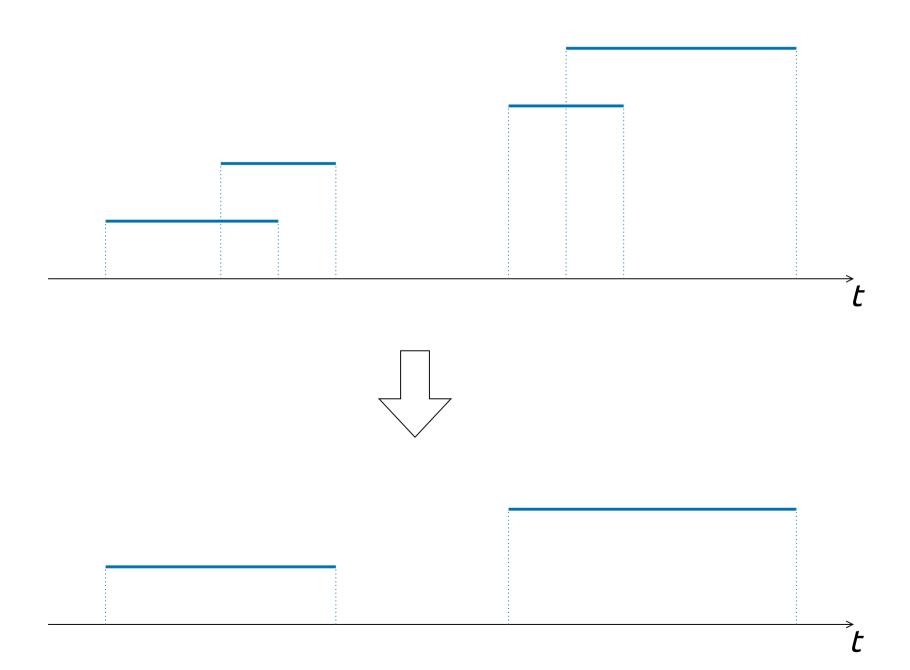
#### From(visit\_occurrence)

person_id	
1	
1	•
1	•
2	
2	

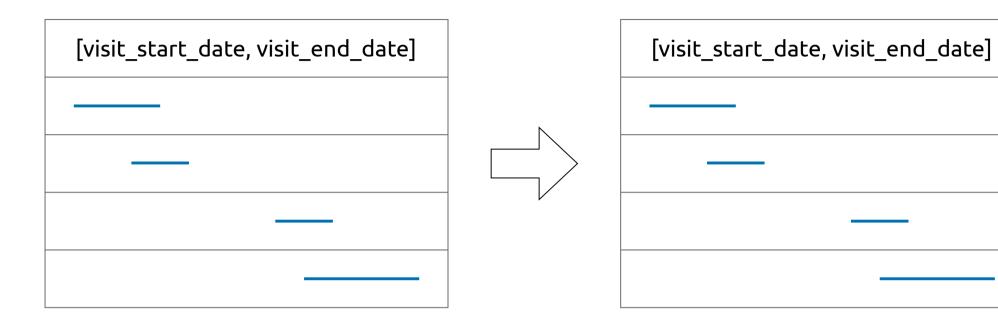




person_id			
1		1	
1	◀	1	
1	•	 1	_
2			•
2		 2	



vi	sit_occurrence
PK	visit_occurrence_id
FK	person_id
	visit_concept_id
	visit_start_date
	visit_end_date

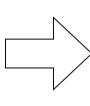


new

0

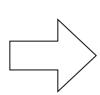
0

[visit_start_date, visit_end_date]	new
	1
	0
	1
	0



[visit_start_date, visit_end_date]	new	group
	1	1
	0	1
	1	2
	0	2

[visit_start_date, visit_end_date]	new	group
	1	1
	0	1
	1	2
	0	2



[start_date, end_date]	group
	1
	2

```
CollapseIntervals(start date, end date) =
    Partition(Get.person_id, order_by = [start_date],
              frame = (mode = :rows, start = -Inf, finish = -1)) |>
   Define(:boundary => Agg.max(end_date)) |>
    Define(:gap => start date .- Get.boundary) |>
    Define(:new => Fun.case(Get.gap .<= 0, 0, 1)) |>
    Partition(Get.person_id, order_by = [start_date, .- Get.new],
              frame = :rows) |>
    Define(:group => Agg.sum(Get.new)) |>
    Group(Get.person_id, Get.group) |>
    Define(:start_date => Agg.min(start_date),
           :end date => Agg.max(end date))
q = From(visit occurrence) |>
    CollapseIntervals(Get.visit_start_date, Get.visit_end_date) |>
    Select(Get.person_id, Get.start_date, Get.end_date)
```

vi	sit_occurrence
PK	visit_occurrence_id
FK	person_id
	visit_concept_id
	visit_start_date
	visit_end_date

# FunSQL? Who Needs It?

Query Algebra

Correlated Queries

Aggregate & Window Functions

Conclusion





```
FilterByYearOfBirth(; start_year, end_year) =
    if start_year !== nothing && end_year !== nothing
        Where(Fun.between(Get.year_of_birth, start_year, end_year))
   elseif start year !== nothing
        Where(Get.year_of_birth .>= start_year)
    elseif end_year !== nothing
        Where(Get.year of birth .<= end year)
    else
        identity
    end
FilterByState(; state) =
    if state !== nothing
        Join(:location => From(location) |>
                          Where(Get.state .== state),
             Get.location_id .== Get.location.location_id)
    else
        identity
    end
```

### DONE

- Select, Where, Join, Define, Group, Partition, Append (UNION ALL), Parameterized and Correlated Queries
- Tested with SQLite, PostgresSQL, RedShift, Microsoft SQL Server

# TODO (June 2021)

- ORDER BY and LIMIT, CTE and WITH clause, WITH RECURSIVE, CUBE/ROLLUP and more
- Documentation and examples
- Comprehensive support for SQL dialects
- Tracking column and expression types
- CRUD (INSERT, UPDATE, DELETE)
- DDL (CREATE TABLE) and schema introspection

## Example

https://github.com/MechanicalRabbit/OHDSICohortExpressions.jl

### Contact Us

julialang.zulipchat.com: funsql.jl