

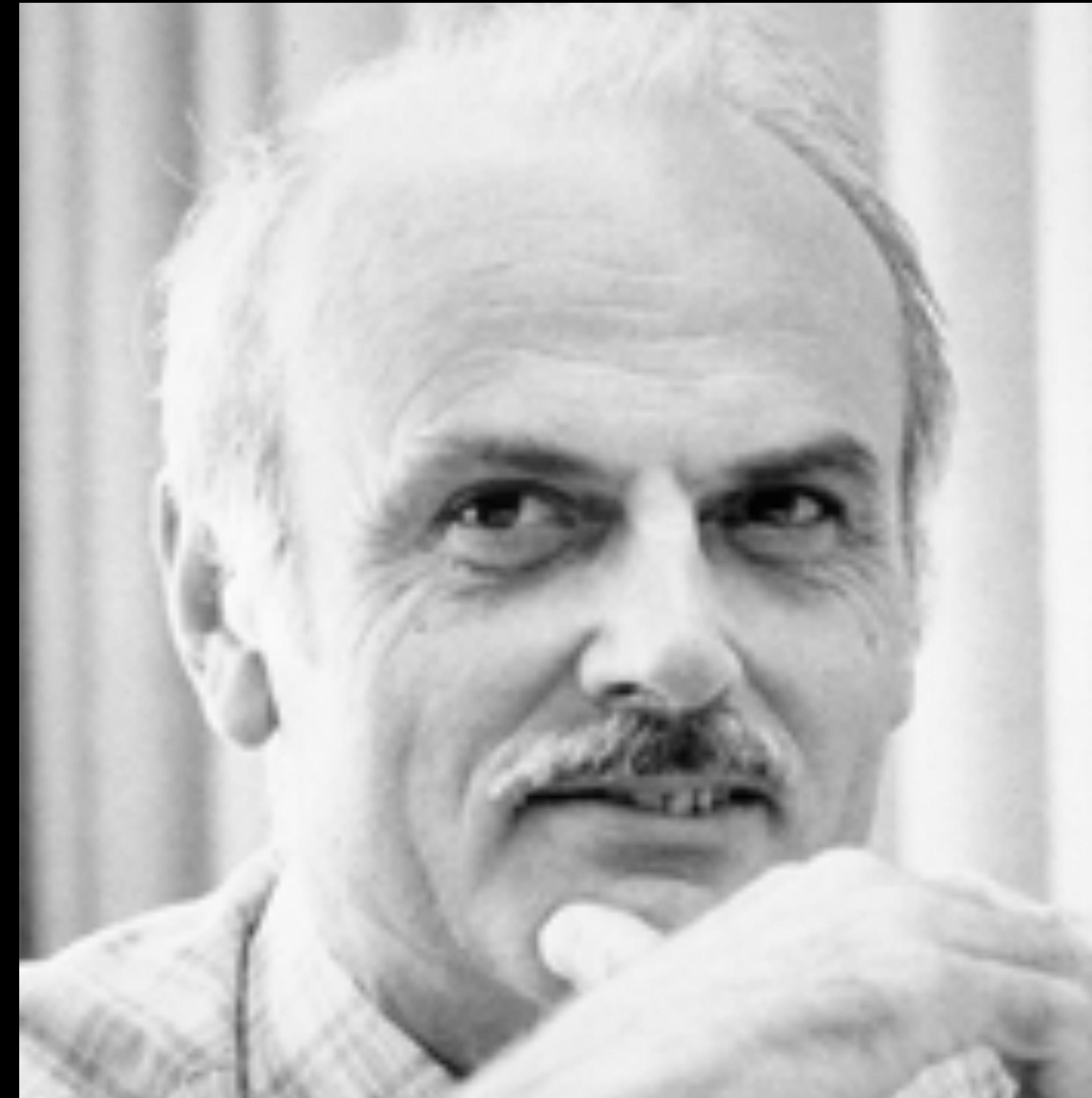
dm: Analyze, build and deploy relational data models

tinyurl.com/dm-readme



Multiple tables!

- 1951: first computer for sale
- 1960: databases
- 1970: Edgar F. Codd:
“A relational model of data for
large shared data banks”
- 2022: Used everywhere



Edgar F. Codd. Source: Wikipedia

Filter

Daten

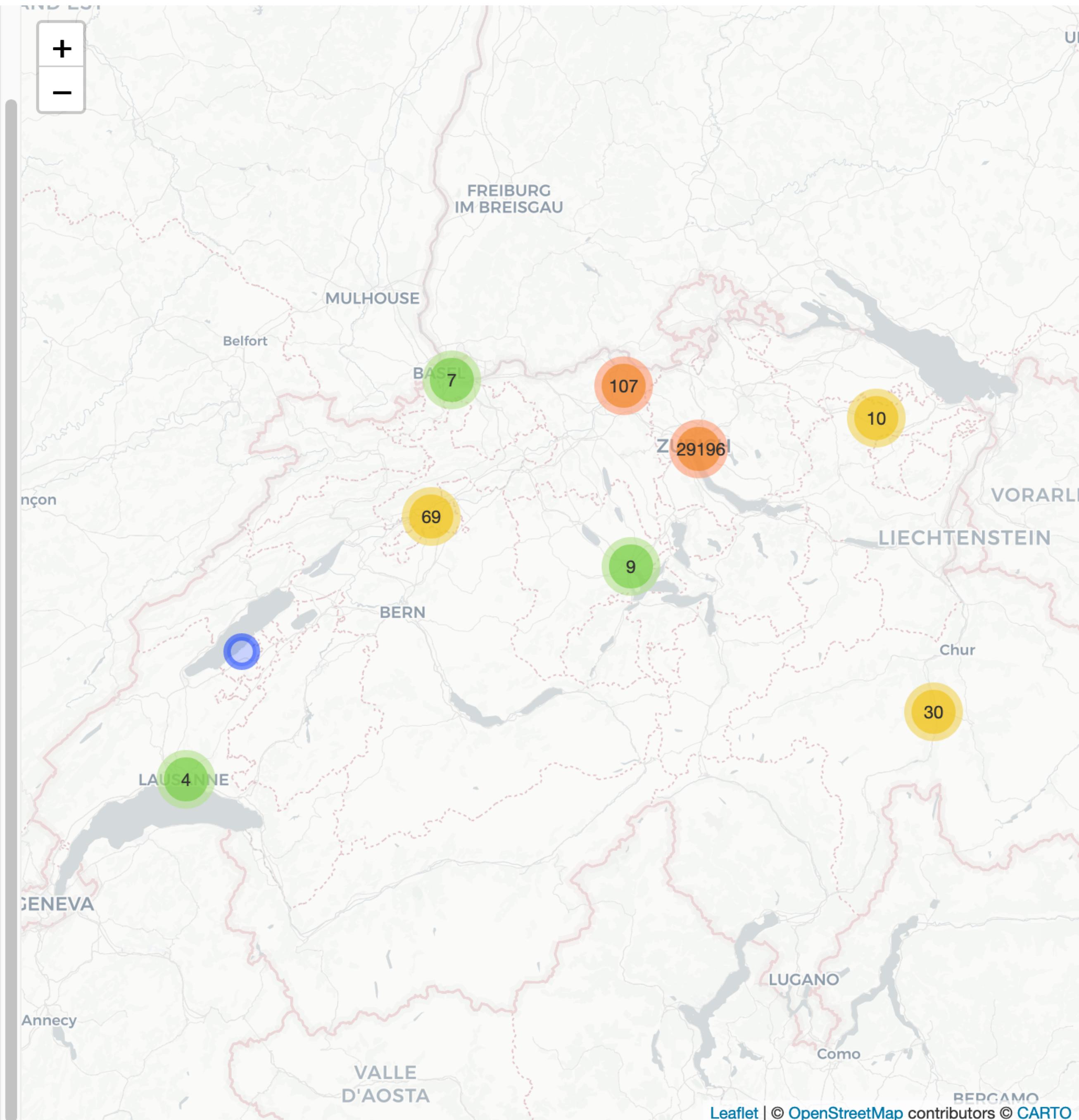
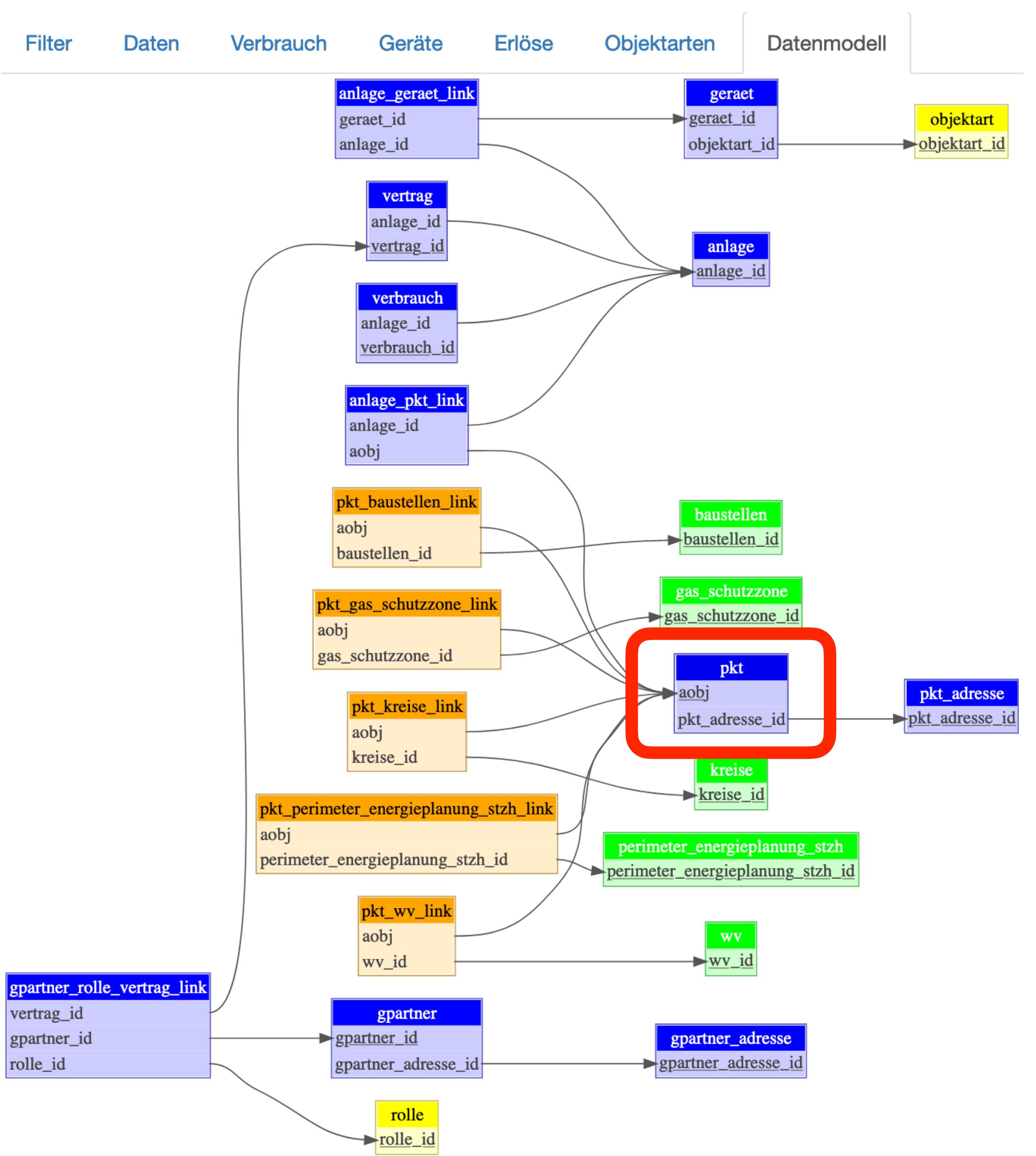
Verbrauch

Geräte

Erlöse

Objektarten

Datenmodell



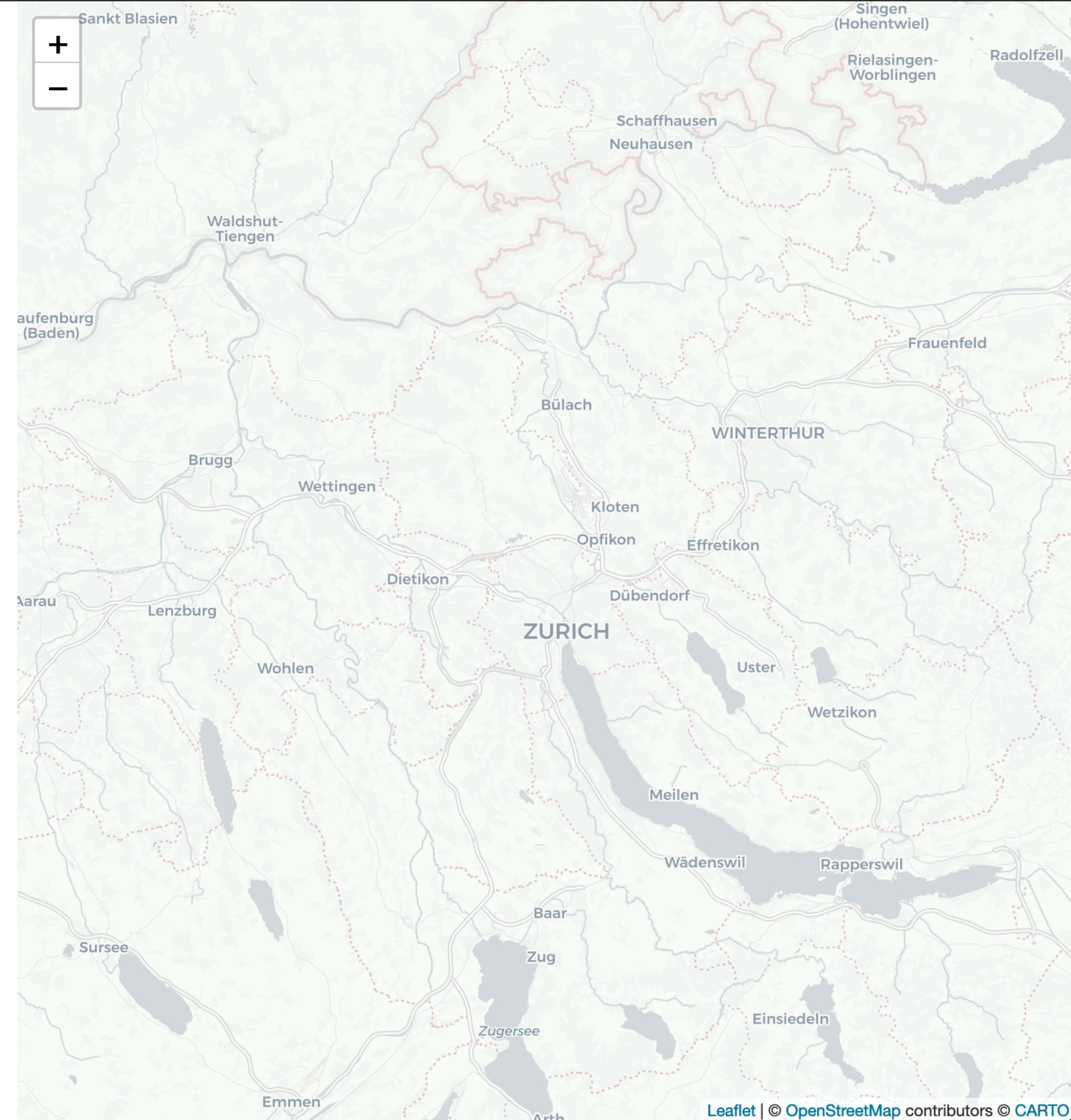
Dynamischer Filter für Gasmarkt

Filter Daten Verbrauch Geräte Erlöse Objektarten Datenmodell

[Remove all filters](#)

[Add new filter](#)

▼



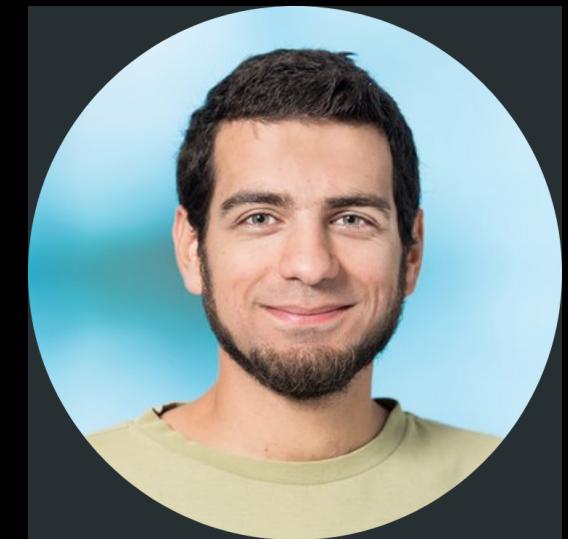
+6 -869 

```
data_model >
dm_filter(
  geraet      = (eq_typ == "A"),
  objektart  = (name == "Friteuse"),
  WV         = (name %in% WV_names)
)
```

What others are saying

- “the {dm} package is amazing!”

Nils Ratnaweera, Twitter



- “a treasure of an R package”

Garrick Aden-Buie, Twitter



- “a life saver”

Hadley Wickham, “R for Data Science” (second edition)



dm 1.0.0

on CRAN!



Dependencies? This is fine!

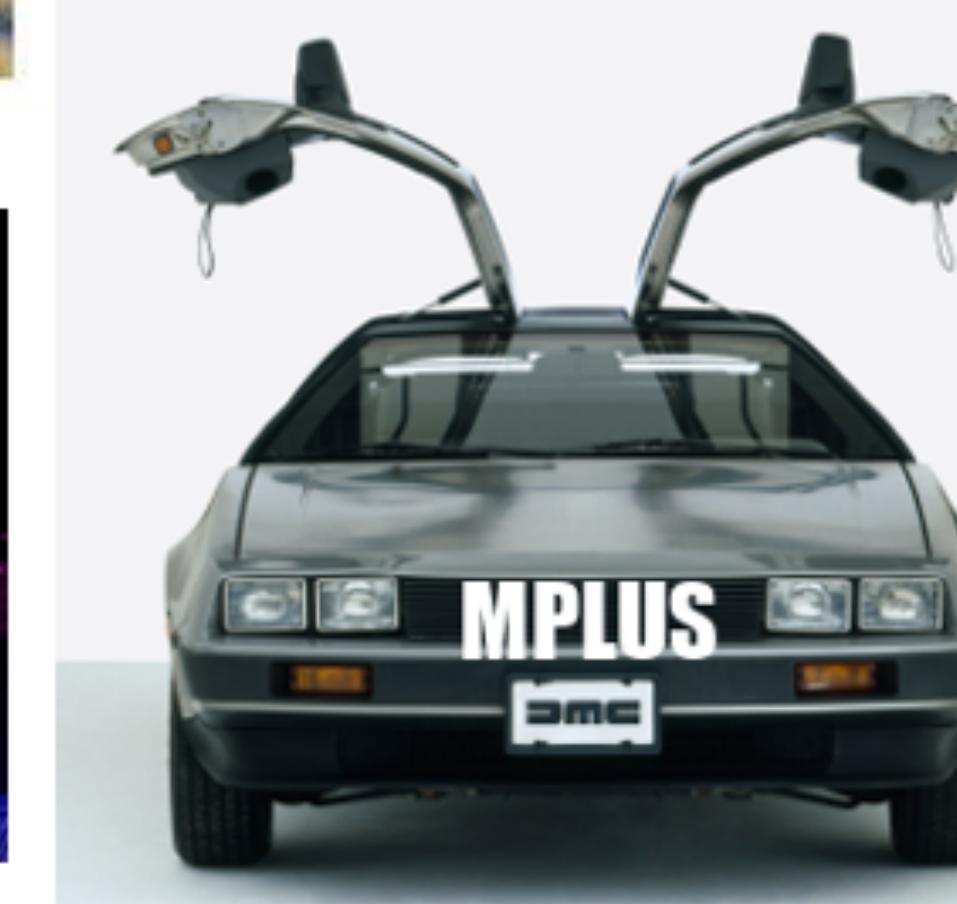


```
tools::package_dependencies("dm", which = "most")[[1]] |>  
  purrr::map_chr(utils::maintainer) |>  
  unique() |>  
  sort() |>  
  as.person() |>  
  format(c("given", "family"))  
#> [1] "Brodie Gaslam"      "Carson Sievert"      "Dean Attali"        "Dirk Eddelbuettel"  
#> [5] "Eric Leung"         "Gábor Csárdi"       "Greg Lin"          "Hadley Wickham"  
#> [9] "Hannes Mühleisen"   "Jennifer Bryan"    "Jeroen Ooms"       "Jim Hester"  
#> [13] "Joe Cheng"         "Kevin Ushey"       "Kirill Müller"     "Lionel Henry"  
#> [17] "Michel Lang"        "R Core Team"       "Richard Iannone"   "Tamás Nepusz"  
#> [21] "Vincent Nijs"       "Winston Chang"     "Yihui Xie"
```

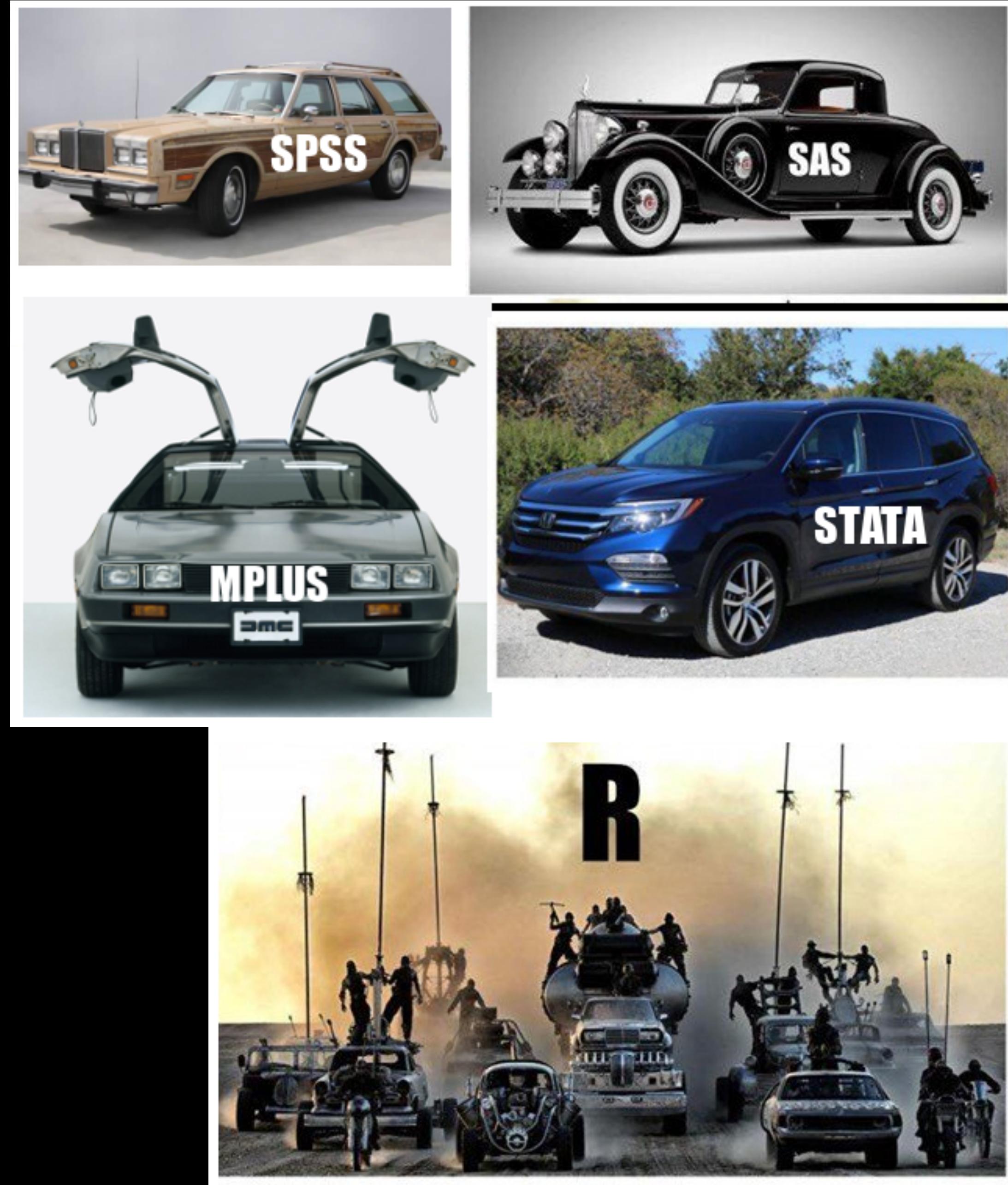
R



If statistics programs/languages were cars...



Multiple tables?



A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	fem	ed	blk		t		exp	wks	occ	ind	lwage
2	1	0	9	0		1		3	32	0	0	5.561
3	1	0	9	0		2		4	43	0	0	5.72
4	1	0	9	0		3		5	40	0	0	5.996
5	1	0	9	0		4		6	39	0	0	5.996
6	1	0	9	0		5		7	42	0	1	6.061
7	1	0	9	0		6		8	35	0	1	6.174
8	1	0	9	0		7		9	32	0	1	6.244
9												
10	2	0	11	0		1		30	34	1	0	6.163
11	2	0	11	0		2		31	27	1	0	6.215
12	2	0	11	0		3		32	33	1	1	6.263
13	2	0	11	0		4		33	30	1	1	6.544
14	2	0	11	0		5		34	30	1	1	6.697

A	B	C	D	E	F	G	H	I	J	K	
1	id	fem	ed	blk		exp_1	wks_1	occ_1	ind_1	lwage_1	
2	1	0	9	0		3	32	0	0	5.561	
3	2	0	11	0		30	34	1	0	6.163	
4	3	0	12	0		6	50	1	1	5.652	
5	4	1	10	1		31	52	1	0	6.157	
6	5	0	16	0		10	50	1	0	6.438	
7	6	0	12	0		26	44	1	1	6.906	
8	7	0	12	0		15	46	1	0	6.133	
9	8	0	10	0		23	51	1	1	6.332	
10	9	0	16	0		3	50	0	0	6.551	
11	10	0	16	0		3	49	0	0	6.397	
12	11	0	12	0		24	47	1	0	6.658	
13	12	0	12	0		21	47	1	0	6.551	
14	13	0	12	0		26	48	1	0	6.906	

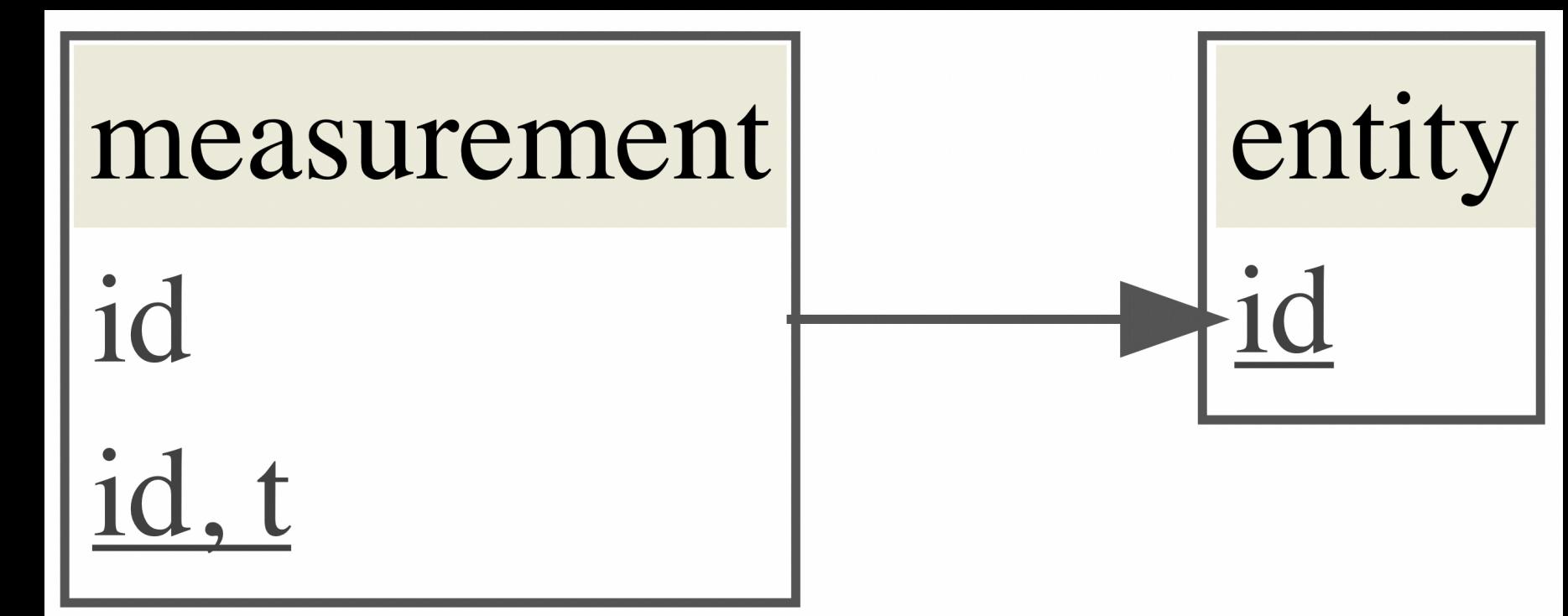
One table
is not enough

...not always

```
if (length( tables ) > 1) {  
  library( dm )  
}
```

Primary and foreign keys

Entities and measurements



	A	B	C	D	E	F	G	H	I	J	K	L
1	id	fem	ed	blk		t		exp	wks	occ	ind	lwage
2	1	0	9	0		1		3	32	0	0	5.561
3	1	0	9	0		2		4	43	0	0	5.72
4	1	0	9	0		3		5	40	0	0	5.996
5	1	0	9	0		4		6	39	0	0	5.996
6	1	0	9	0		5		7	42	0	1	6.061
7	1	0	9	0		6		8	35	0	1	6.174
8	1	0	9	0		7		9	32	0	1	6.244
9												
10	2	0	11	0		1		30	34	1	0	6.163
11	2	0	11	0		2		31	27	1	0	6.215
12	2	0	11	0		3		32	33	1	1	6.263
13	2	0	11	0		4		33	30	1	1	6.544
14	2	0	11	0		5		34	30	1	1	6.697

	A	B	C	D	E
1	id	fem	ed	blk	
2	1	0	9	0	
3	2	0	11	0	
4	3	0	12	0	
5	4	1	10	1	
6	5	0	16	0	
7	6	0	12	0	
8	7	0	12	0	
9	8	0	10	0	
10	9	0	16	0	
11	10	0	16	0	
12	11	0	12	0	
13	12	0	12	0	
14	13	0	12	0	

~git/R/dm - Shiny

http://127.0.0.1:7527 | Open in Browser | C

DM R DATA MODELS

DM Call

1 dm

Choose color

white

Tools:

- Zoom
- Delete table
- Select table
- Remove foreign key
- Remove cycles

http://127.0.0.1:3300 | Open in Browser | ~git/R/dm - Shiny

DM R DATA MODELS

To Console 4

DM Call

```
1 dm %>%
2   dm_select_tbl(BusinessEntity, PhoneNumberType, PersonPh3
```

Employee BusinessEntityID → Person BusinessEntityID → BusinessEntity BusinessEntityID

PersonPhone BusinessEntityID, PhoneNumberTypeID, BusinessEntityID, PhoneNumber, PhoneNumberTypeID

PhoneNumberType PhoneNumberTypeID

1

2

Zoom Delete table Select table Remove foreign key Remove cycles Choose color white

The screenshot shows the DM R DATA MODELS application running in RStudio. On the left, a data model diagram is displayed with tables: Employee, Person, BusinessEntity, PersonPhone, and PhoneNumberType. A red box labeled '1' highlights the PersonPhone table, which has columns BusinessEntityID, PhoneNumberTypeID, BusinessEntityID, PhoneNumber, and PhoneNumberTypeID. A red box labeled '2' highlights the 'Select table' button at the bottom of the interface. On the right, the 'DM Call' panel shows R code: '1 dm %>%' and '2 dm_select_tbl(BusinessEntity, PhoneNumberType, PersonPh3'. A red box labeled '3' highlights the second line of the code. A red box labeled '4' highlights the 'To Console' button at the top right of the panel.

~git/R/dm - Shiny

http://127.0.0.1:4804 | Open in Browser | C

DM R DATA MODELS

Select or search one or several tables

flights

airports

planes

airlines

Choose color
white

Zoom Delete table Select table Remove foreign key Remove cycles

DM Call

1 dm

http://127.0.0.1:4804 | Open in Browser | ~git/R/dm - Shiny | Publish | To Console

DM R DATA MODELS

airports x

flights 4

airlines

planes

airports faa 1

Choose color white

Zoom Delete table Select table Remove foreign key Remove cycles Add primary key 3 Delete column

DM Call

```
1 dm %>%
2   dm_add_pk(airports, faa)
```

Edit Table

	airports	chr
<input type="checkbox"/>	faa 2	chr
<input type="checkbox"/>	name	chr
<input type="checkbox"/>	lat	dbl
<input type="checkbox"/>	lon	dbl
<input type="checkbox"/>	alt	dbl
<input type="checkbox"/>	tz	dbl
<input type="checkbox"/>	dst	chr

Database

Metadata

Quote Information

Coverages

Quote Summary

Projects

Name ↗

 Test Custo... Test Custo... a x a a aaaaaa a a

11–19 of 19 rows

Previous 1 2

Next



Quote Family

Quote ID Country ↗

Quote name

 161 UK Standard Disability 2022-06-27 15:49 160 UK q 2022-06-27 15:47 158 UK a 2022-06-27 15:45

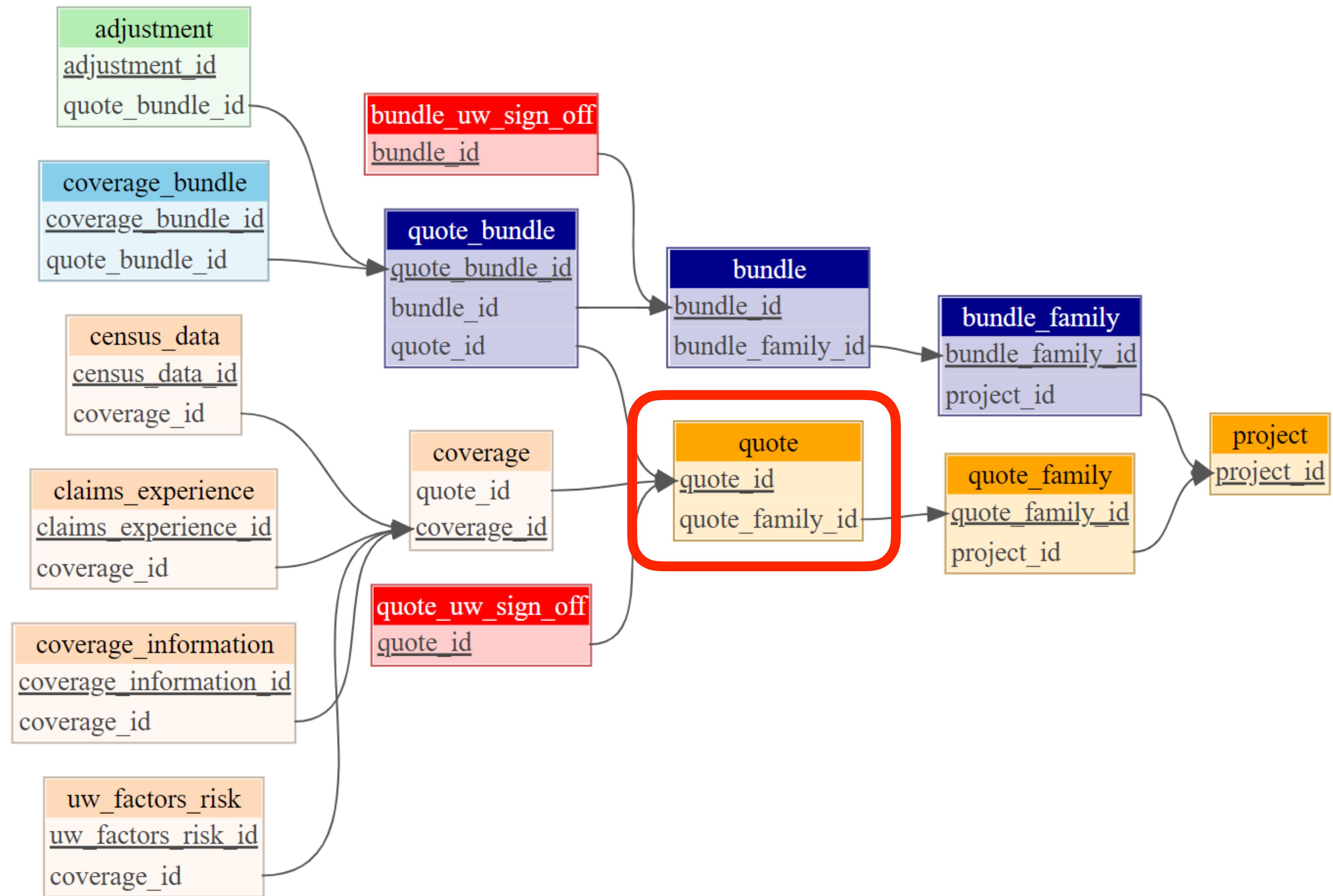
+ New Quote from Excel

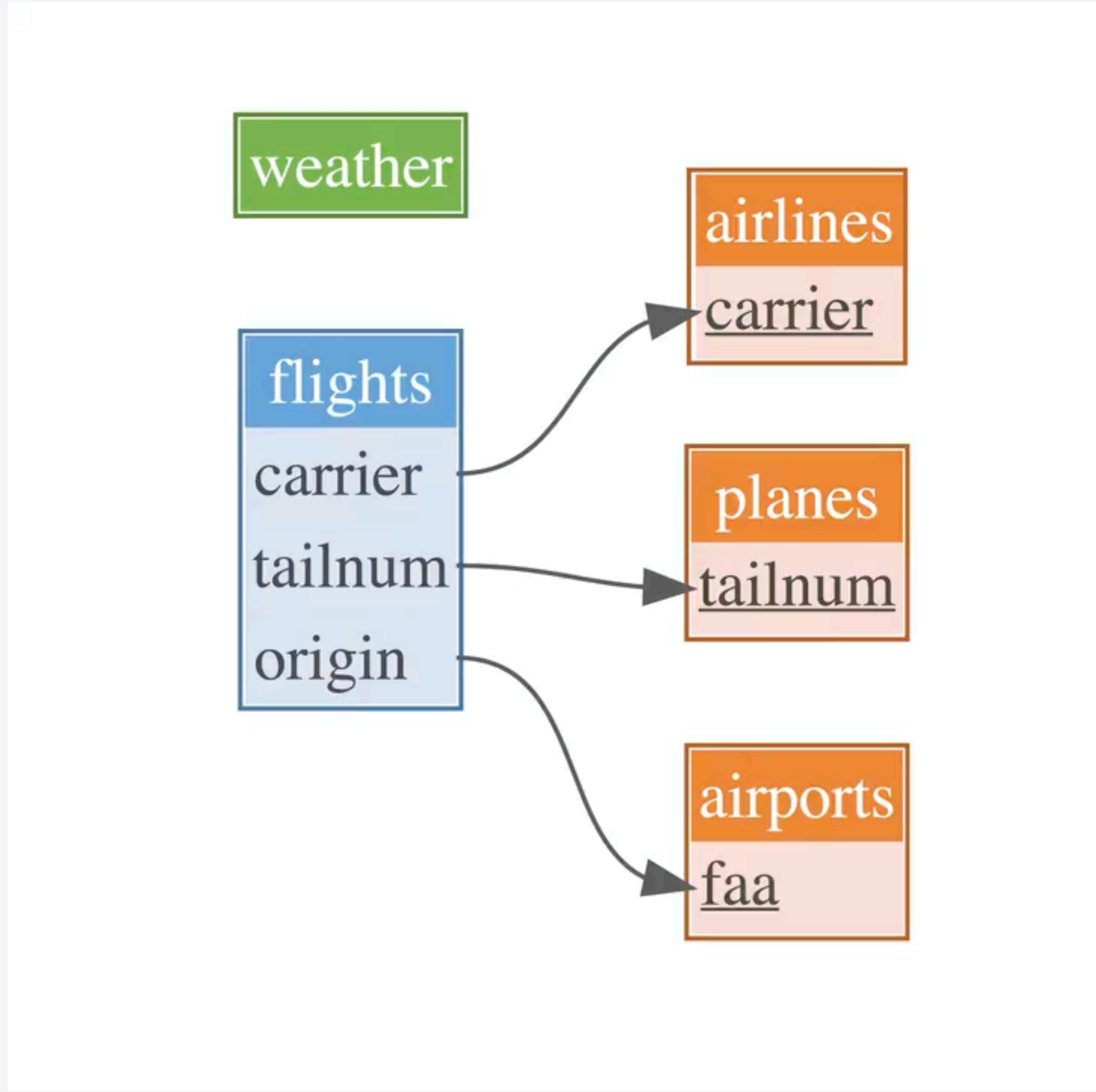
+ New Quote without Excel

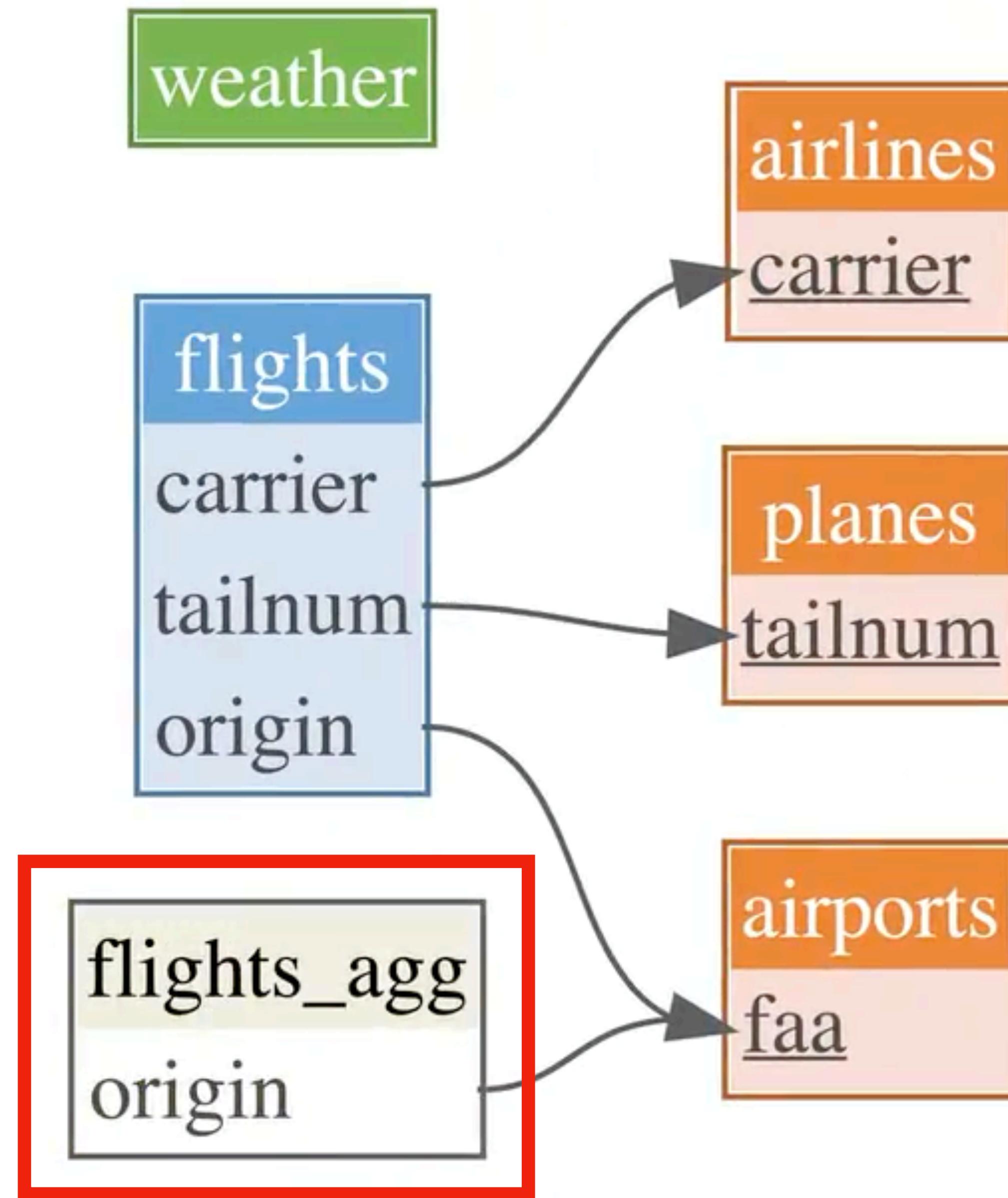
Quotes

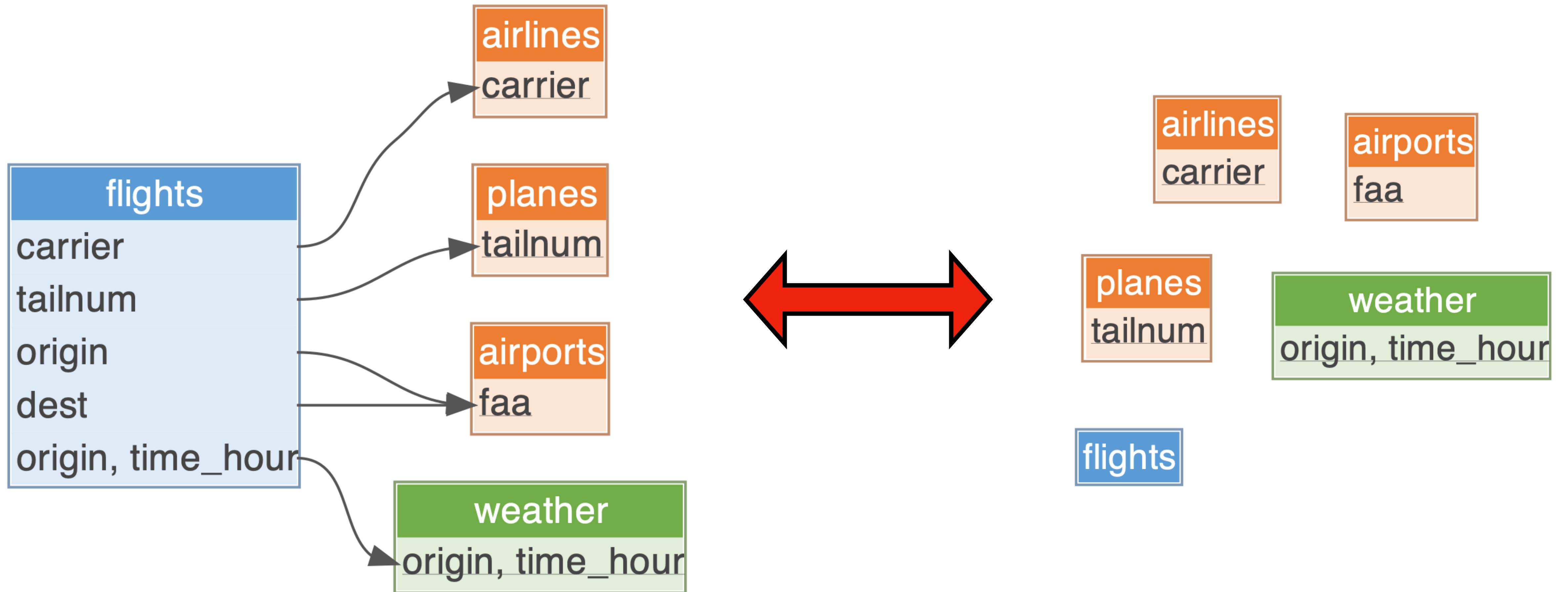
Quote ID Country ↗ Version Comment ↗

 161 UK Created from: Base Quote Template_LCIS_First_Stage_Tool_Disability_Census.xlsx











```
dm <- dm_nycflights13(cycle = TRUE)
dm
#> — Metadata —
#> Tables: `airlines`, `airports`, `flights`, `planes`, `weather`
#> Columns: 53
#> Primary keys: 4
#> Foreign keys: 5
```

```
dm_deconstruct(dm)
#> airlines <- pull_tbl(dm, "airlines", keyed = TRUE)
#> airports <- pull_tbl(dm, "airports", keyed = TRUE)
#> flights <- pull_tbl(dm, "flights", keyed = TRUE)
#> planes <- pull_tbl(dm, "planes", keyed = TRUE)
#> weather <- pull_tbl(dm, "weather", keyed = TRUE)
```



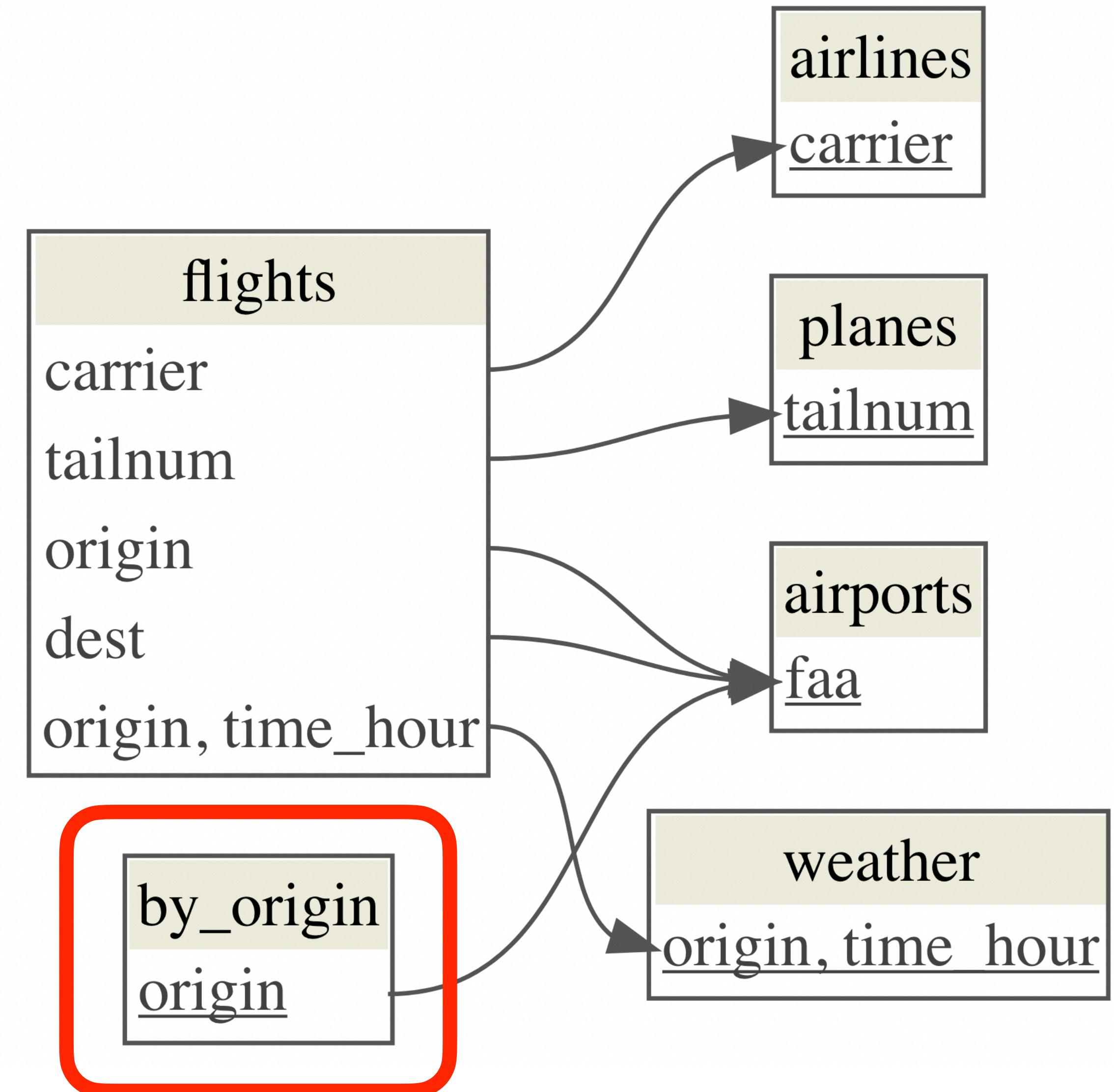
```
airlines <- pull_tbl(dm, "airlines", keyed = TRUE)
airports <- pull_tbl(dm, "airports", keyed = TRUE)
flights <- pull_tbl(dm, "flights", keyed = TRUE)
planes <- pull_tbl(dm, "planes", keyed = TRUE)
weather <- pull_tbl(dm, "weather", keyed = TRUE)

dm(airlines, airports, flights, planes, weather)
#> └─ Metadata ─────────────────────────────────────────────────────────────────
#>   Tables: `airlines`, `airports`, `flights`, `planes`, `weather`
#>   Columns: 53
#>   Primary keys: 4
#>   Foreign keys: 5
```



```
by_origin <-
  flights %>%
  group_by(origin) %>%
  summarize(n = n(), mean_arr_delay = mean(arr_delay)) %>%
  ungroup()

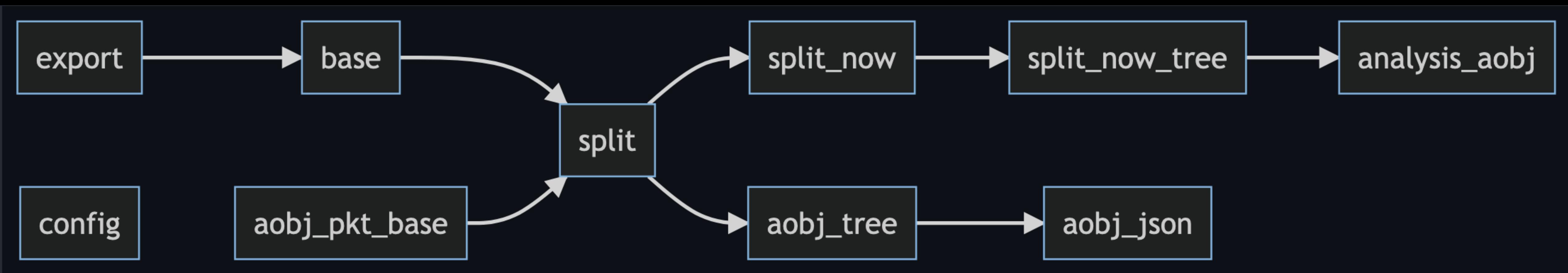
dm(airlines, airports, flights, planes, weather, by_origin)
#> — Metadata —
#> Tables: `airlines`, `airports`, `flights`, `planes`, `weather`, `by_origin`
#> Columns: 56
#> Primary keys: 5
#> Foreign keys: 6
```



Multiple data models?



ETL diagram



Is it all the same?

- Normalized data (databases) `dm`
- Nested tables `tidyverse`
- Deeply nested lists (JSON) `tibblify`

cynkra .

tinyurl.com/dm-readme



Big dreams, dm 2.0.0

- Wrapping and unwrapping
- Disentangling and reentangling
- JSON + nested on the database