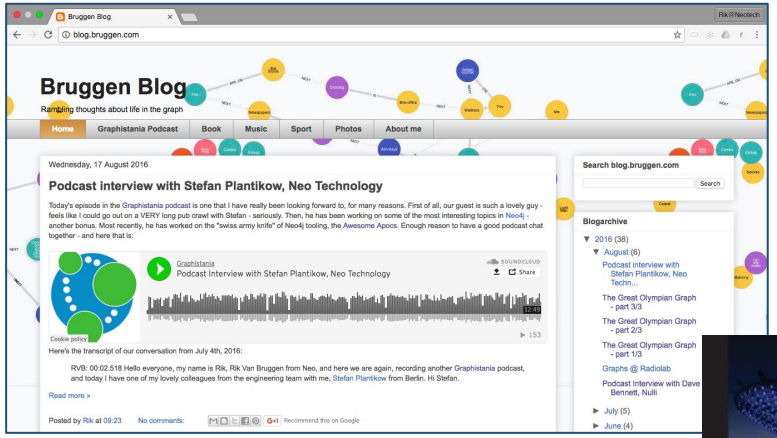
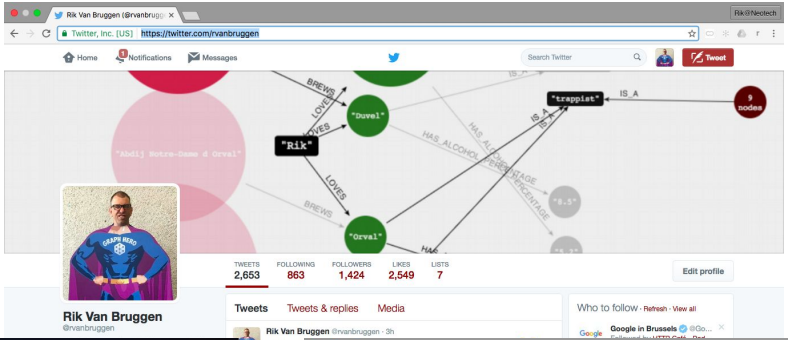
The Great Olympian Graph

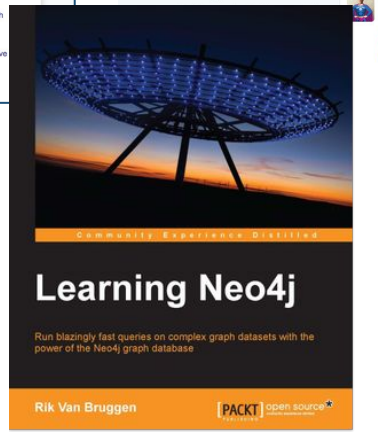
Live Webinar

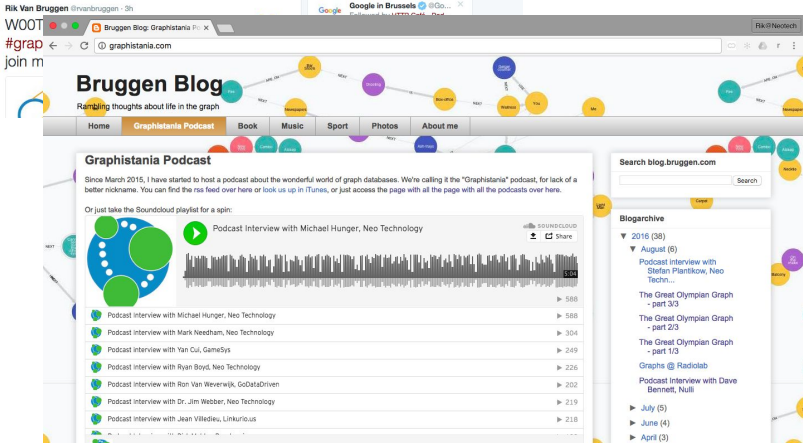
August 18th, 2016

I am @rvanbruggen









Blog.bruggen.com
Graphistania.com
Learningneo4j.net
rik@neotechnology.com

The Background

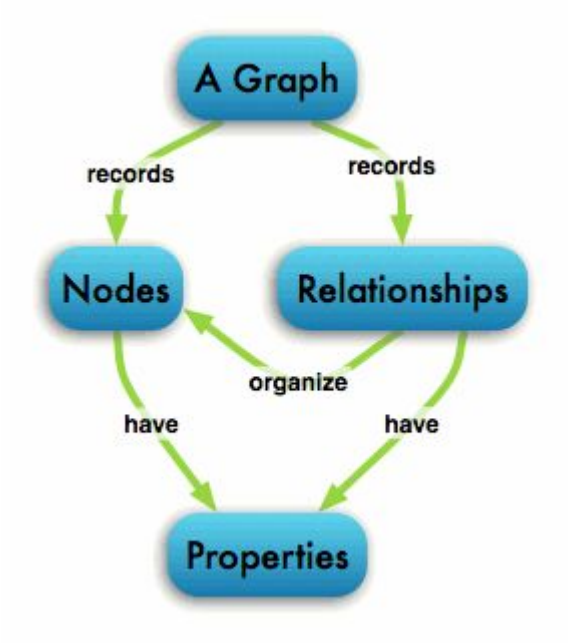
Always on the lookout for interesting datasets

In Love with Graphs since knowing Neo4j because

1. MODEL

2. PERFORMANCE

The Property Graph Model



Whiteboard (& Business) Friendly

Rich, Multi-purpose

HiFi representation of reality

Low Impedance Mismatch

Great for AGILE projects

Performance matters

Traditional DBs have a real problem with JOINS

A Graph RELATIONSHIP ~
The Pre-calculated, Explicity-stored Join-operation

Making JOINS ~ Pointer-chasing = easy!

- Index-free adjacency
- Graph locality

Over the years: MANY Examples

Corporate networks, Terrorism database, Social Networks, Music recommendations, Clickstream analysis, Friends and Foes in the Middle East, Access and Identity Management, Tour de France, Flanders Classics Cycling, Twitter communities, VAT registries, Rail networks, GTFS feeds, Bus / Metro networks, Electricity networks, Process flows, Orienteering maps, PanamaPapers, Wikipedia, Conference Schedules, GraphConnect, Oredev, Qcon, Business Continuity Management, Recommendation Engines, Knowledge Graph, Podcast database, Family Tree, Genealogy, Food Recipes, Food Composition, Belgian Beer Graph, GOTO Conference, Dutch Beer Graph, Single Malt Graph, Conceptnet, Trophic Cascade: how wolves change rivers, Swearwords, World Cup Football, Belgian TV Sitcom show “Thuis”, Email traffic analysis, Telco Churn, Doctor Who, LinkedIn…

http://blog.bruggen.com

Yet another graph...

The Great Olympic Graph

Dataset >> Model >> Import >> Query

http://blog.bruggen.com/search/label/olympics for more details

Starting out with a dataset

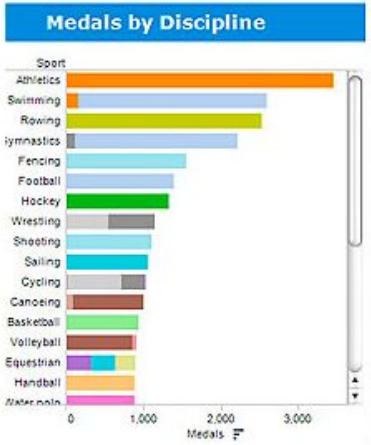
https://www.theguardian.com/sport/datablog/2012/jun/25/olympic-medal-winner-list-data

Summer Olympic medallists 1896 to 2008

List of medallists at the Games of the Olympiad per edition, sport, discipline, gender and event

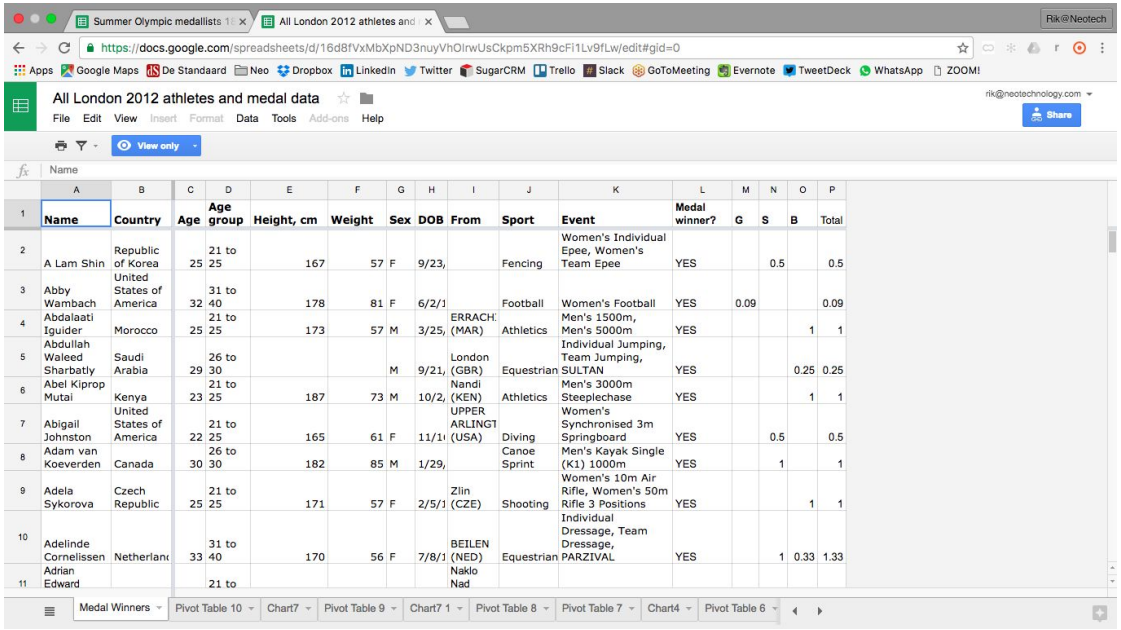
DISCLAIMER: The IOC Research and Reference Service endeavours to provide you with accurate and up-to-date information. However, it offers no guarantees, express or implied, as to the accuracy or completeness of the information provid

City	Edition	Sport	Discipline	Athlete	NOC	Gender	Event	Event_gender
Athens	1896	Aquatics	Swimming	HAJOS, Alfred	HUN	Men	100m freestyle	M
Athens	1896	Aquatics	Swimming	HERSCHMANN, Otto	AUT	Men	100m freestyle	M
Athens	1896	Aquatics	Swimming	DRIVAS, Dimitrios	GRE	Men	100m freestyle for sailors	M
Athens	1896	Aquatics	Swimming	MALOKINIS, Ioannis	GRE	Men	100m freestyle for sailors	M
Athens	1896	Aquatics	Swimming	CHASAPIS, Spiridon	GRE	Men	100m freestyle for sailors	M
Athens	1896	Aquatics	Swimming	CHOROPHAS, Efstathios	GRE	Men	1200m freestyle	M
Athens	1896	Aquatics	Swimming	HAJOS, Alfred	HUN	Men	1200m freestyle	M
Athens	1896	Aquatics	Swimming	ANDREOU, Joannis	GRE	Men	1200m freestyle	M
Athens	1896	Aquatics	Swimming	CHOROPHAS, Efstathios	GRE	Men	400m freestyle	M
Athens	1896	Aquatics	Swimming	NEUMANN, Paul	AUT	Men	400m freestyle	M
Athens	1896	Aquatics	Swimming	PEPANOS, Antonios	GRE	Men	400m freestyle	M
Athens	1896	Athletics	Athletics	LANE, Francis	USA	Men	100m	M
Athens	1896	Athletics	Athletics	SZOKOLYI, Alajos	HUN	Men	100m	M
Athens	1896	Athletics	Athletics	BURKE, Thomas	USA	Men	100m	M
Athens	1896	Athletics	Athletics	HOFMANN, Fritz	GER	Men	100m	M
Athens	1896	Athletics	Athletics	CURTIS, Thomas	USA	Men	110m hurdles	M
Athens	1896	Athletics	Athletics	GOULDING, Grantley	GBR	Men	110m hurdles	M
Athens	1896	Athletics	Athletics	LERMUSIAUX, Albin	FRA	Men	1500m	M
Athens	1896	Athletics	Athletics	FLACK, Edwin	AUS	Men	1500m	M
Athens	1896	Athletics	Athletics	BLAKE, Arthur	USA	Men	1500m	M
Athens	1896	Athletics	Athletics	GMELIN, Charles	GBR	Men	400m	M



Adding the 2012 data

https://www.theguardian.com/sport/datablog/2012/aug/10/olympics-2012-list-medal-winners#data

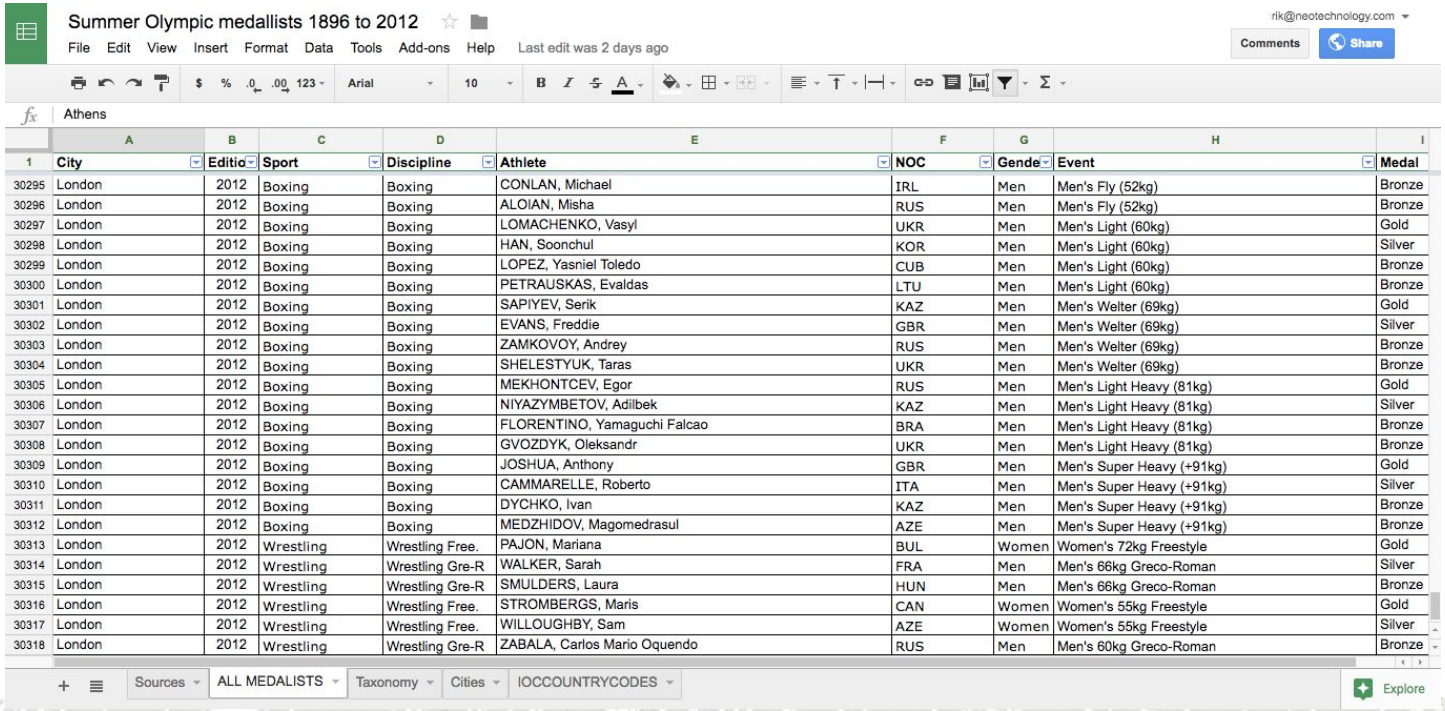


All London 2012 athletes and medal data

Name	Country	Age	Age group	Height, cm	Weight	Sex	DOB	From	Sport	Event	Medal winner?	G	S	B	Total
A Lam Shin	Republic of Korea	25	21 to 25	167	57	F	9/23,		Fencing	Women's Individual Epee, Women's Team Epee	YES		0.5		0.5
Abby Wambach	United States of America	32	31 to 40	178	81	F	6/2/1		Football	Women's Football	YES	0.09			0.09
Abdalaati Iguider	Morocco	25	21 to 25	173	57	M	3/25,	ERRACH (MAR)	Athletics	Men's 1500m, Men's 5000m	YES			1	1
Abdullah Waleed Sharbatly	Saudi Arabia	29	26 to 30			M	9/21,	London (GBR)	Equestrian	Individual Jumping, Team Jumping, SULTAN	YES			0.25	0.25
Abel Kiprop Mutai	Kenya	23	21 to 25	187	73	M	10/2,	Nandi (KEN)	Athletics	Men's 3000m Steeplechase	YES			1	1
Abigail Johnston	United States of America	22	21 to 25	165	61	F	11/1(	UPPER ARLINGT (USA)	Diving	Women's Synchronised 3m Springboard	YES		0.5		0.5
Adam van Koeverden	Canada	30	26 to 30	182	85	M	1/29,		Canoe Sprint	Men's Kayak Single (K1) 1000m	YES		1		1
Adela Sykorova	Czech Republic	25	21 to 25	171	57	F	2/5/1	Zlin (CZE)	Shooting	Women's 10m Air Rifle, Women's 50m Rifle 3 Positions	YES			1	1
Adelinde Cornelissen	Netherlan	33	31 to 40	170	56	F	7/8/1	BEILEN (NED)	Equestrian	Individual Dressage, Team Dressage, PARZIVAL	YES		1	0.33	1.33
Adrian Edward			21 to					Naklo Nad							

Merging the two into something new

[A new spreadsheet for 1896-2012]



Summer Olympic medallists 1896 to 2012

File Edit View Insert Format Data Tools Add-ons Help Last edit was 2 days ago

rik@neotechnology.com

Comments Share

Arial 10

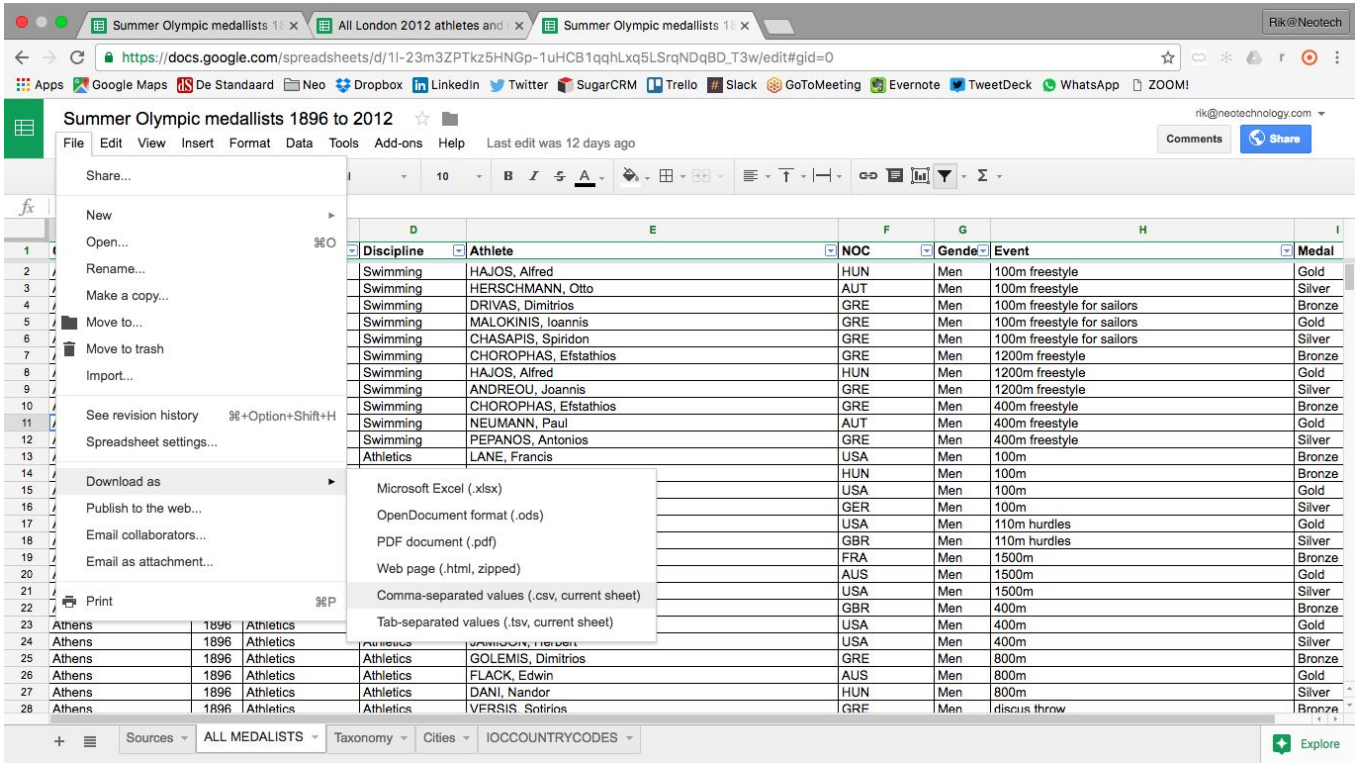
Athens

	A	B	C	D	E	F	G	H	I
1	City	Editio	Sport	Discipline	Athlete	NOC	Gende	Event	Medal
30295	London	2012	Boxing	Boxing	CONLAN, Michael	IRL	Men	Men's Fly (52kg)	Bronze
30296	London	2012	Boxing	Boxing	ALOIAN, Misha	RUS	Men	Men's Fly (52kg)	Bronze
30297	London	2012	Boxing	Boxing	LOMACHENKO, Vasyl	UKR	Men	Men's Light (60kg)	Gold
30298	London	2012	Boxing	Boxing	HAN, Soonchul	KOR	Men	Men's Light (60kg)	Silver
30299	London	2012	Boxing	Boxing	LOPEZ, Yasniel Toledo	CUB	Men	Men's Light (60kg)	Bronze
30300	London	2012	Boxing	Boxing	PETRAUSKAS, Evaldas	LTU	Men	Men's Light (60kg)	Bronze
30301	London	2012	Boxing	Boxing	SAPIYEV, Serik	KAZ	Men	Men's Welter (69kg)	Gold
30302	London	2012	Boxing	Boxing	EVANS, Freddie	GBR	Men	Men's Welter (69kg)	Silver
30303	London	2012	Boxing	Boxing	ZAMKOVOY, Andrey	RUS	Men	Men's Welter (69kg)	Bronze
30304	London	2012	Boxing	Boxing	SHELESTYUK, Taras	UKR	Men	Men's Welter (69kg)	Bronze
30305	London	2012	Boxing	Boxing	MEKHONTCEV, Egor	RUS	Men	Men's Light Heavy (81kg)	Gold
30306	London	2012	Boxing	Boxing	NIYAZYMBETOV, Adilbek	KAZ	Men	Men's Light Heavy (81kg)	Silver
30307	London	2012	Boxing	Boxing	FLORENTINO, Yamaguchi Falcao	BRA	Men	Men's Light Heavy (81kg)	Bronze
30308	London	2012	Boxing	Boxing	GVOZDYK, Oleksandr	UKR	Men	Men's Light Heavy (81kg)	Bronze
30309	London	2012	Boxing	Boxing	JOSHUA, Anthony	GBR	Men	Men's Super Heavy (+91kg)	Gold
30310	London	2012	Boxing	Boxing	CAMMARELLE, Roberto	ITA	Men	Men's Super Heavy (+91kg)	Silver
30311	London	2012	Boxing	Boxing	DYCHKO, Ivan	KAZ	Men	Men's Super Heavy (+91kg)	Bronze
30312	London	2012	Boxing	Boxing	MEDZHIDOV, Magomedrasul	AZE	Men	Men's Super Heavy (+91kg)	Bronze
30313	London	2012	Wrestling	Wrestling Free.	PAJON, Mariana	BUL	Women	Women's 72kg Freestyle	Gold
30314	London	2012	Wrestling	Wrestling Gre-R	WALKER, Sarah	FRA	Men	Men's 66kg Greco-Roman	Silver
30315	London	2012	Wrestling	Wrestling Gre-R	SMULDERS, Laura	HUN	Men	Men's 66kg Greco-Roman	Bronze
30316	London	2012	Wrestling	Wrestling Free.	STROMBERGS, Maris	CAN	Women	Women's 55kg Freestyle	Gold
30317	London	2012	Wrestling	Wrestling Free.	WILLOUGHBY, Sam	AZE	Women	Women's 55kg Freestyle	Silver
30318	London	2012	Wrestling	Wrestling Gre-R	ZABALA, Carlos Mario Oquendo	RUS	Men	Men's 60kg Greco-Roman	Bronze

Sources ALL MEDALISTS Taxonomy Cities IOCCOUNTRYCODES

Explore

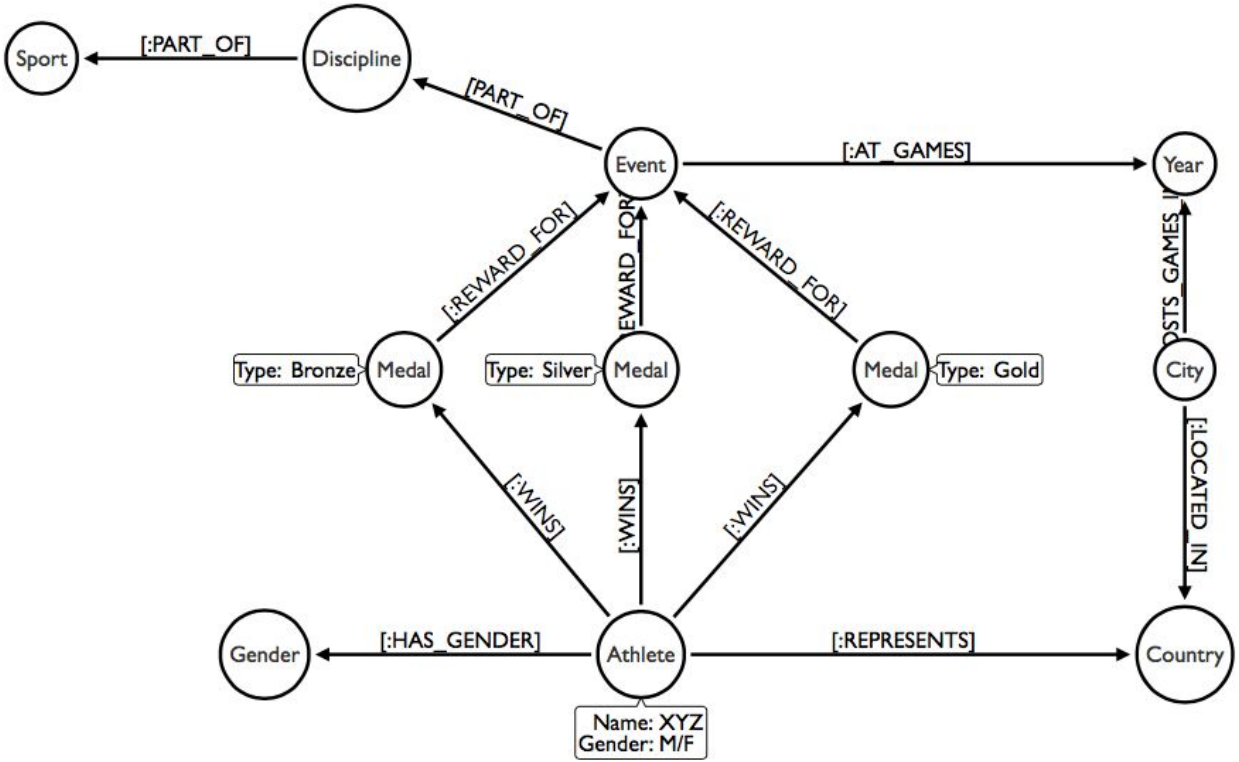
Downloading as a .csv



4 .csv files

1. One for the countries, and their 3-letter country-codes
2. One for the hosting cities, and their mappings to the above countries
3. A two-level categorisation (let's call it a "sports taxonomy") of some sort containing Sports (eg. Aquatics, Cycling,...), and Disciplines (eg. Swimming, Track Cycling, …)
4. A much bigger and longer sheet / .csv file that contains the actual data about each and every of the 30000+ medallists.

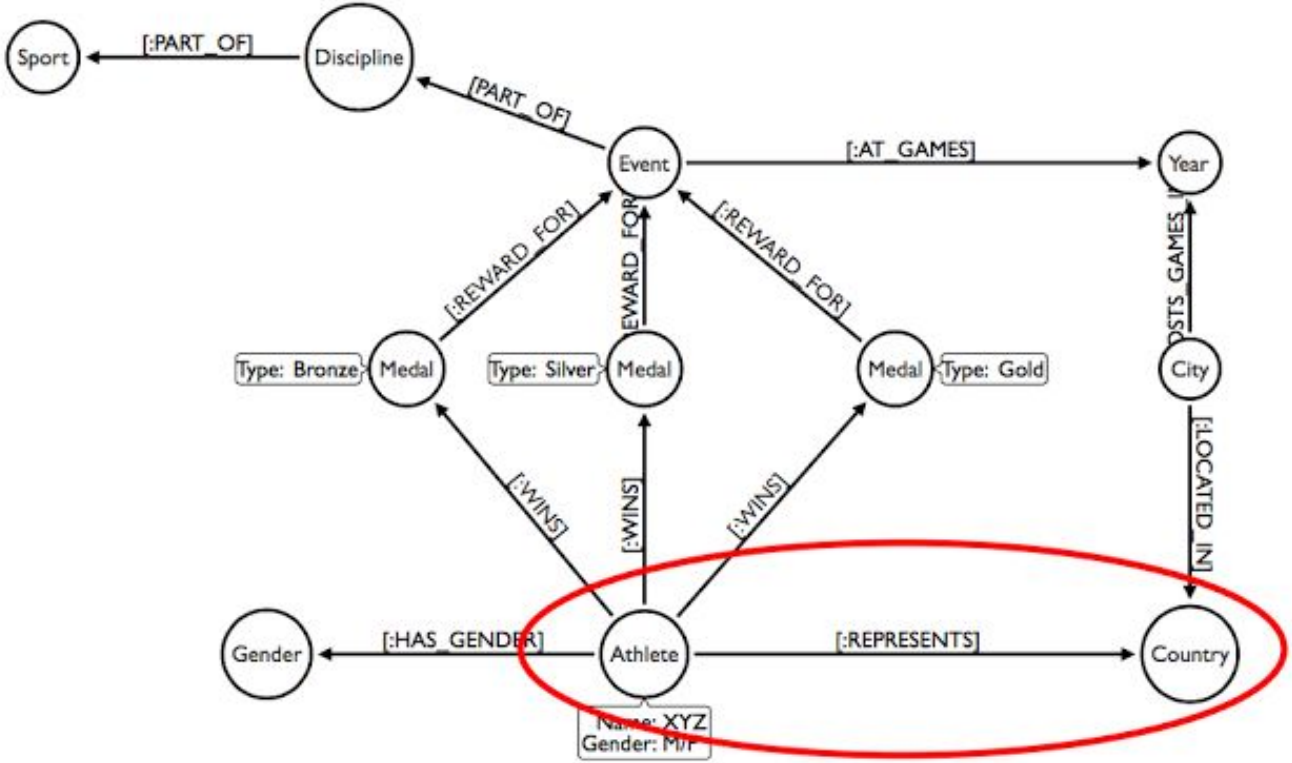
Creating a model



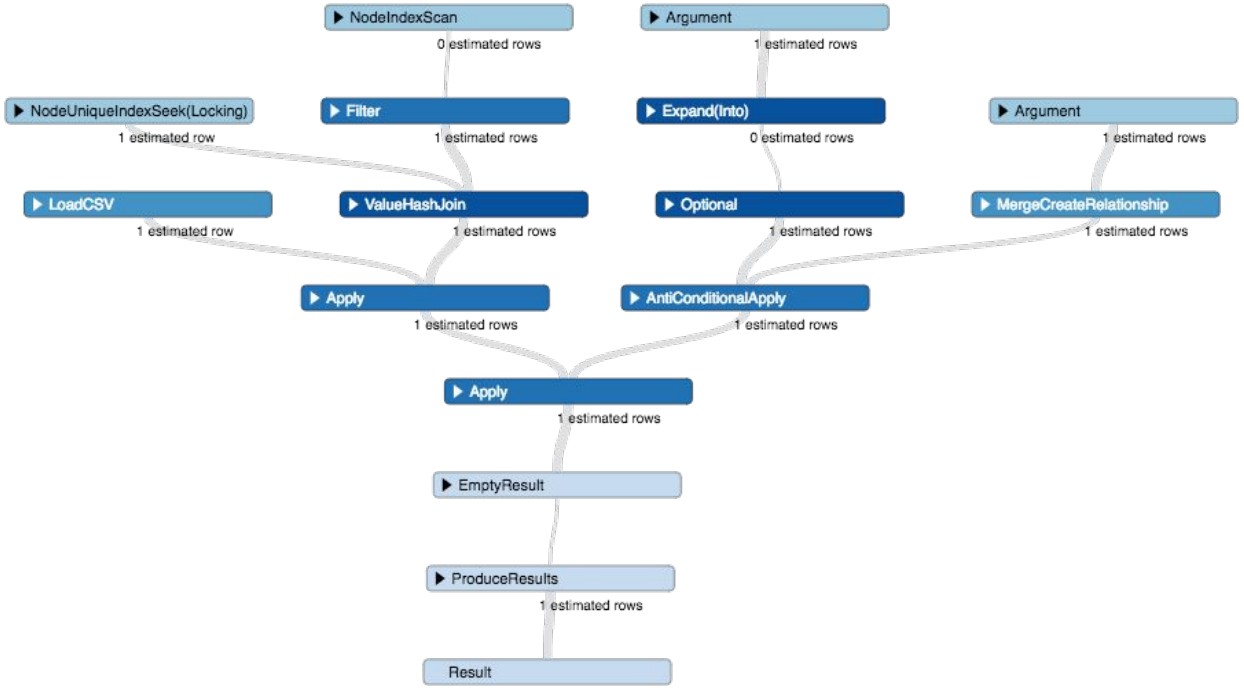
Loading the data

- Loading the Genders of the Athletes
- Loading the Countries
- Loading the Cities in these Countries
- Loading the sports taxonomy
- Loading the different games
- Loading the different events in different disciplines of the taxonomy at the different games
- Loading the different athletes / medallists and their genders

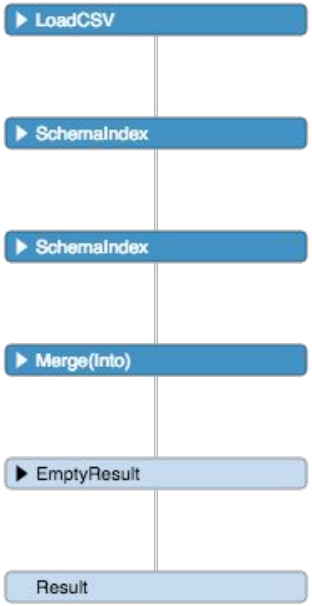
Every import has its challenge :)



Cost-based vs. Rule-based planner: COST

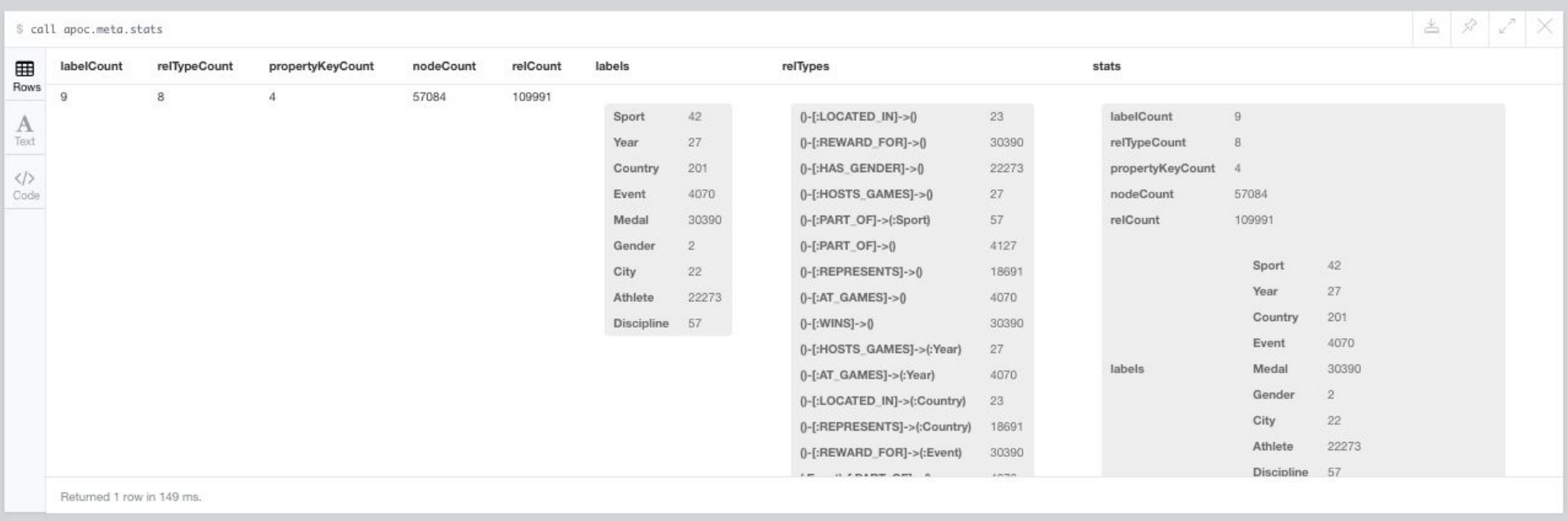


Cost-based vs. Rule-based planner: RULE




```
cypher planner=rule
load csv with headers from
"https://docs.google.com/a/neotechnology.com/spreadsheets/d/1l-23m3ZPTkz5HNGp-1uHCB1qqhLxq5LSrqNDqBD_T3w/export?format=csv&id=1l-23m3ZPTkz5HNGp-1uHCB1qqhLxq5LSrqNDqBD_T3w&gid=0" as csv
match (d:Discipline {name: csv.Discipline})-[:PART_OF]->(s:Sport {name: csv.Sport})match (a:Athlete {name: csv.Athlete})match (e:Event {name: csv.Event})-->(y:Year {name: csv.Edition})create (a)-[:WINS]->(m:Medal {type: csv.Medal})-[:REWARD_FOR]->(e);
```

Load successful!



```
$ call apoc.meta.stats
```

Rows | Text | Code

labelCount	relTypeCount	propertyKeyCount	nodeCount	relCount	labels	relTypes	stats
9	8	4	57084	109991			

labels

Sport	42
Year	27
Country	201
Event	4070
Medal	30390
Gender	2
City	22
Athlete	22273
Discipline	57

relTypes

()-[:LOCATED_IN]->()	23
()-[:REWARD_FOR]->()	30390
()-[:HAS_GENDER]->()	22273
()-[:HOSTS_GAMES]->()	27
()-[:PART_OF]->(:Sport)	57
()-[:PART_OF]->()	4127
()-[:REPRESENTS]->()	18691
()-[:AT_GAMES]->()	4070
()-[:WINS]->()	30390
()-[:HOSTS_GAMES]->(:Year)	27
()-[:AT_GAMES]->(:Year)	4070
()-[:LOCATED_IN]->(:Country)	23
()-[:REPRESENTS]->(:Country)	18691
()-[:REWARD_FOR]->(:Event)	30390

stats

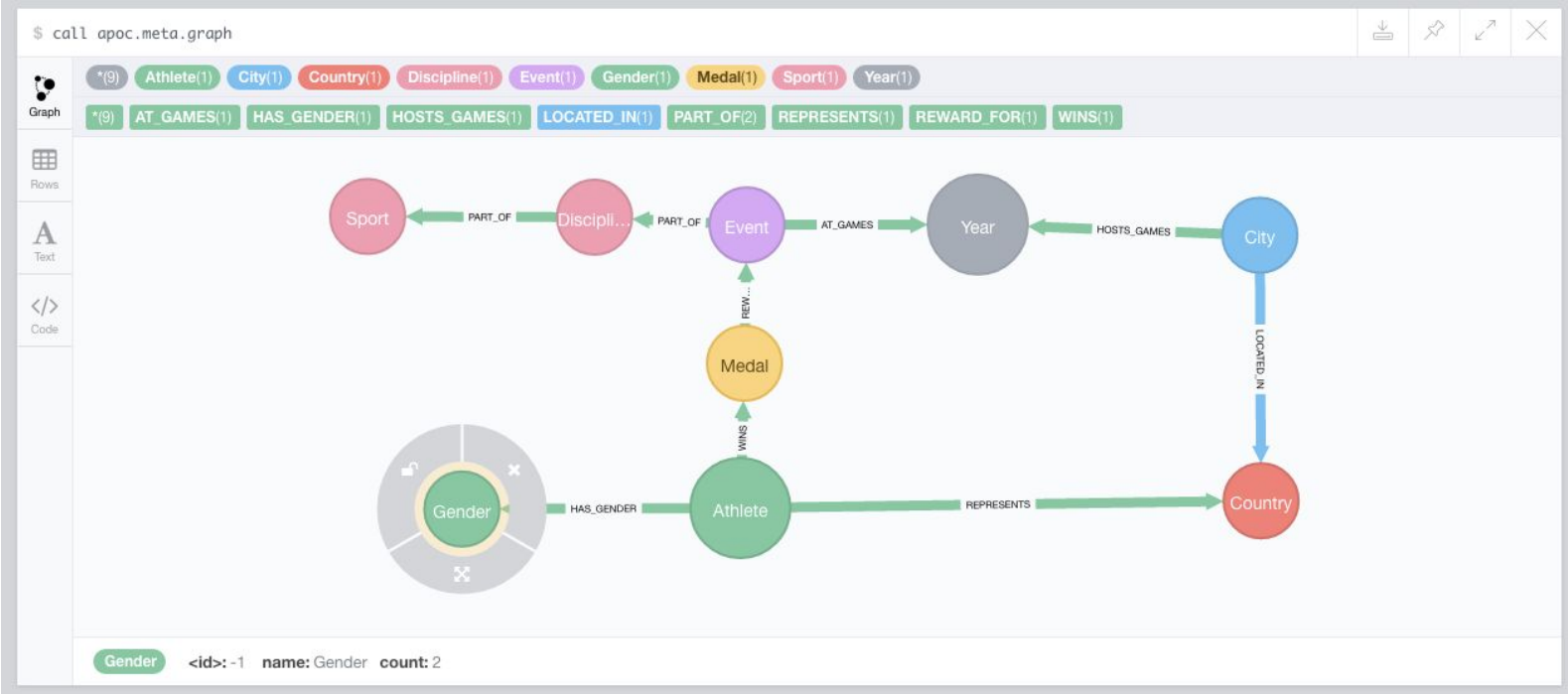
labelCount	9
relTypeCount	8
propertyKeyCount	4
nodeCount	57084
relCount	109991

labels

Sport	42
Year	27
Country	201
Event	4070
Medal	30390
Gender	2
City	22
Athlete	22273
Discipline	57

Returned 1 row in 149 ms.

Load successful!



Querying the Olympic Graph

```
//number of sports per game
match (y:Year)<--(e:Event)-->(d:Discipline)-->(s:Sport)with distinct y.name as game, s.name as sportreturn game, count(sport)order by game ASC
```

```
$ match (y:Year)<--(e:Event)-->(d:Discipline)-->(s:Sport) with distinct y.name as game, s.name as sport return game, count(sport) order by ga...
```

Rows | Text | Code

game	count(sport)
1896	9
1900	19
1904	16
1908	22
1912	14
1920	22
1924	17

Querying the Olympic Graph

```
//number of sports per game
match (y:Year)<--(e:Event)-->(d:Discipline)-->(s:Sport)
with distinct y.name as game, s.name as sport
return game, count(sport), collect(sport)order by game ASC
```

```
$ match (y:Year)<--(e:Event)-->(d:Discipline)-->(s:Sport) with distinct y.name as game, s.name as sport return game, count(sport), collect(sp...
```

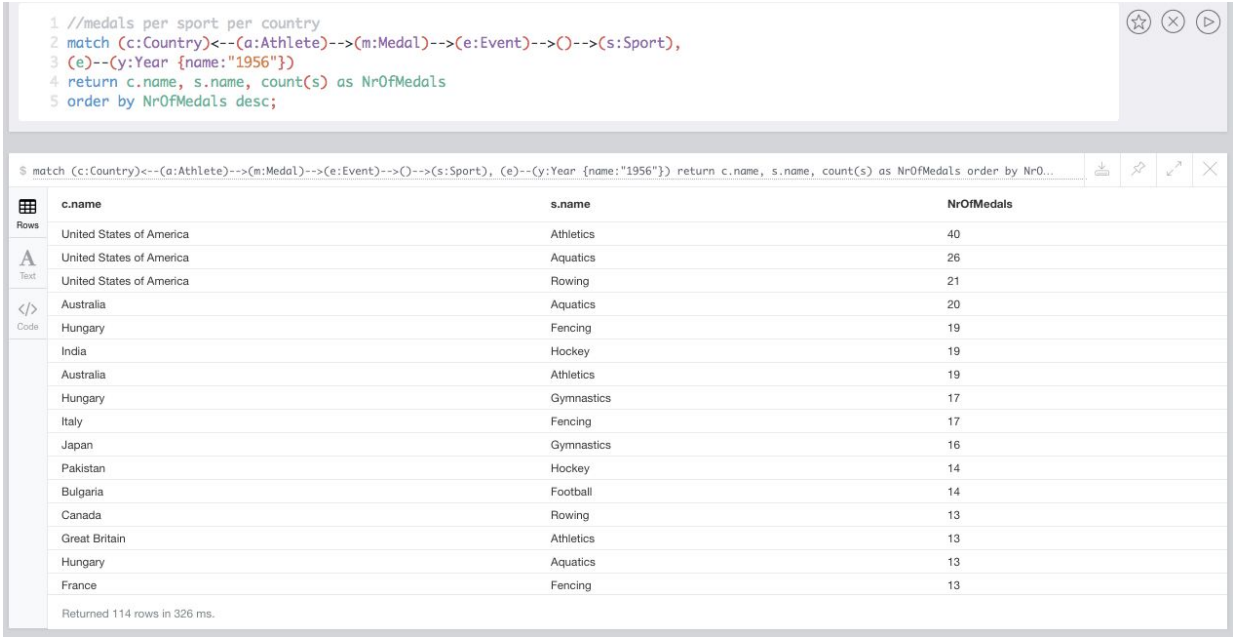
Rows | Text | Code

game	count(sport)	collect(sport)
1896	9	[Aquatics, Athletics, Cycling, Fencing, Gymnastics, Shooting, Tennis, Weightlifting, Wrestling]
1900	19	[Aquatics, Archery, Athletics, Basque Pelota, Cricket, Croquet, Cycling, Equestrian, Fencing, Football, Golf, Gymnastics, Polo, Rowing, Rugby, Sailing, Shooting, Tennis, Tug of War]
1904	16	[Aquatics, Archery, Athletics, Boxing, Cycling, Fencing, Football, Golf, Gymnastics, Lacrosse, Roque, Rowing, Tennis, Tug of War, Weightlifting, Wrestling]
1908	22	[Aquatics, Archery, Athletics, Boxing, Cycling, Fencing, Football, Gymnastics, Hockey, Jeu de paume, Lacrosse, Polo, Rackets, Rowing, Rugby, Sailing, Shooting, Skating, Tennis, Tug of War, Water Motorsports, Wrestling]
1912	14	[Aquatics, Athletics, Cycling, Equestrian, Fencing, Football, Gymnastics, Modern Pentathlon, Rowing, Sailing, Shooting, Tennis, Tug of War, Wrestling]
1920	22	[Aquatics, Archery, Athletics, Boxing, Cycling, Equestrian, Fencing, Football, Gymnastics, Hockey, Ice Hockey, Modern Pentathlon, Polo, Rowing, Rugby, Sailing, Shooting, Skating, Tennis, Tug of War, Weightlifting, Wrestling]
1924	17	[Aquatics, Athletics, Boxing, Cycling, Equestrian, Fencing, Football, Gymnastics, Modern Pentathlon, Polo, Rowing, Rugby, Sailing, Shooting, Tennis, Weightlifting, Wrestling]
1928	14	[Aquatics, Athletics, Boxing, Cycling, Equestrian, Fencing, Football, Gymnastics, Hockey, Modern Pentathlon, Rowing, Sailing, Weightlifting, Wrestling]
1932	14	[Aquatics, Athletics, Boxing, Cycling, Equestrian, Fencing, Gymnastics, Hockey, Modern Pentathlon, Rowing, Sailing, Shooting, Weightlifting, Wrestling]
1936	19	[Aquatics, Athletics, Basketball, Boxing, Canoe / Kayak, Cycling, Equestrian, Fencing, Football, Gymnastics, Handball, Hockey, Modern Pentathlon, Polo, Rowing, Sailing, Shooting, Weightlifting, Wrestling]
1948	17	[Aquatics, Athletics, Basketball, Boxing, Canoe / Kayak, Cycling, Equestrian, Fencing, Football, Gymnastics, Hockey, Modern Pentathlon, Rowing, Sailing, Shooting,

Returned 27 rows in 75 ms.

Querying the Olympic Graph

```
//medals per sport per country
match
(c:Country)<--(a:Athlete)-->(m:Medal
)-->(e:Event)-->()-->(s:Sport),
(e)--(y:Year {name:"1956"})
return c.name, s.name, count(s) as
NrOfMedals
order by NrOfMedals desc;
```

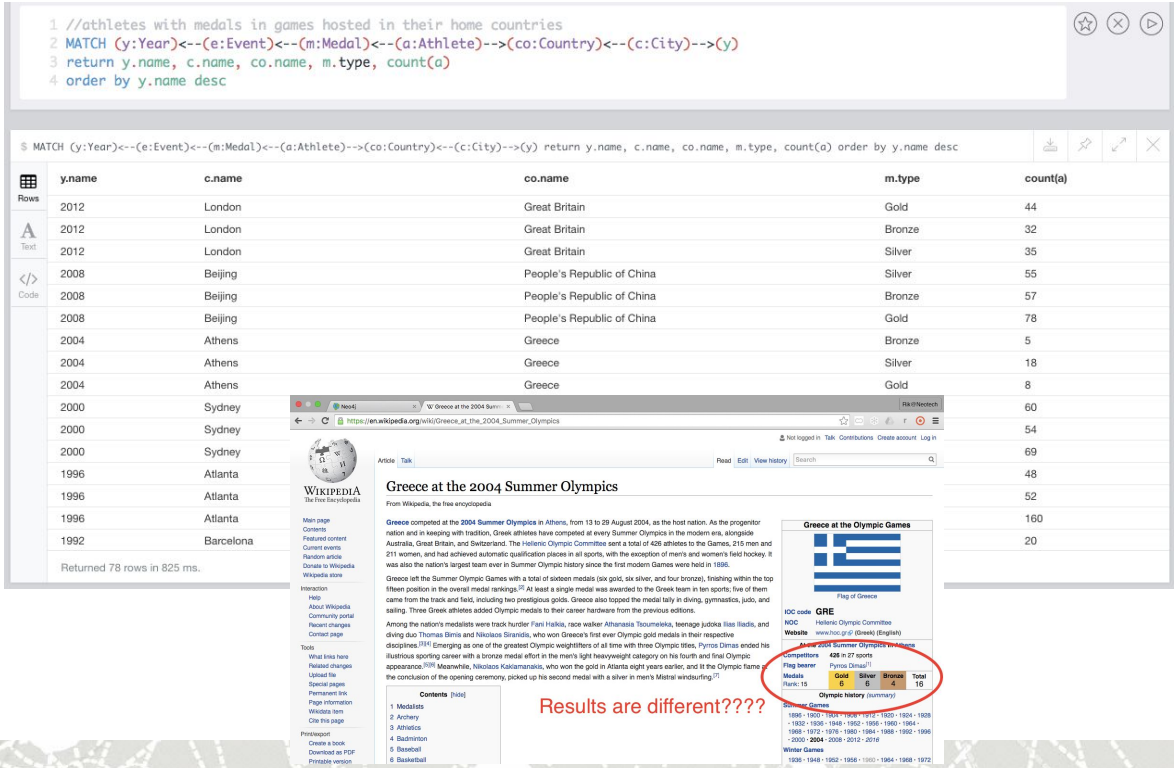


c.name	s.name	NrOfMedals
United States of America	Athletics	40
United States of America	Aquatics	26
United States of America	Rowing	21
Australia	Aquatics	20
Hungary	Fencing	19
India	Hockey	19
Australia	Athletics	19
Hungary	Gymnastics	17
Italy	Fencing	17
Japan	Gymnastics	16
Pakistan	Hockey	14
Bulgaria	Football	14
Canada	Rowing	13
Great Britain	Athletics	13
Hungary	Aquatics	13
France	Fencing	13

Returned 114 rows in 326 ms.

Querying the Olympic Graph

```
//athletes with medals in games hosted
in their home countries
MATCH
(y:Year)<--(e:Event)<--(m:Medal)<--(a:At
hlete)-->(co:Country)<--(c:City)-->(y)
return y.name, c.name, co.name, m.type,
count(a)
order by y.name desc
```



Querying the Olympic Graph

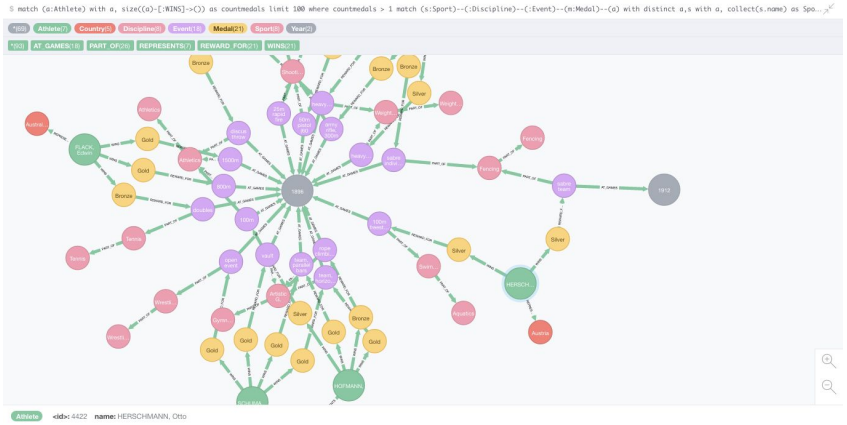
```
//Olympians with medals in different
sports
match (a:Athlete)
with a, size((a)-[:WINS]->()) as
countmedals
limit 100
where countmedals > 1
match
(s:Sport)--(:Discipline)--(:Event)--(m
:Medal)--(a)
with distinct a,s
match (a)-->(c:Country)
with a.name as Athlete, c.name as
Country, collect(s.name) as Sports
where size(Sports)>1
return Athlete, Country, Sports;
```

```
$ match (a:Athlete) with a, size((a)-[:WINS]->()) as countmedals limit 100 where countmedals > 1 match (s:Sport)--(:Discipline)--(:Event)--(m:Medal)--(a) with dis...
```

Rows | Text | Code

Athlete	Country	Sports
HOFMANN, Fritz	Germany	[Athletics, Gymnastics]
NIELSEN, Holger	Denmark	[Fencing, Shooting]
HERSCHMANN, Otto	Austria	[Fencing, Aquatics]
JENSEN, Viggo	Denmark	[Weightlifting, Shooting]
VERSIS, Sotirios	Greece	[Athletics, Weightlifting]
FLACK, Edwin	Australia	[Athletics, Tennis]
SCHUMANN, Carl	Germany	[Gymnastics, Wrestling]

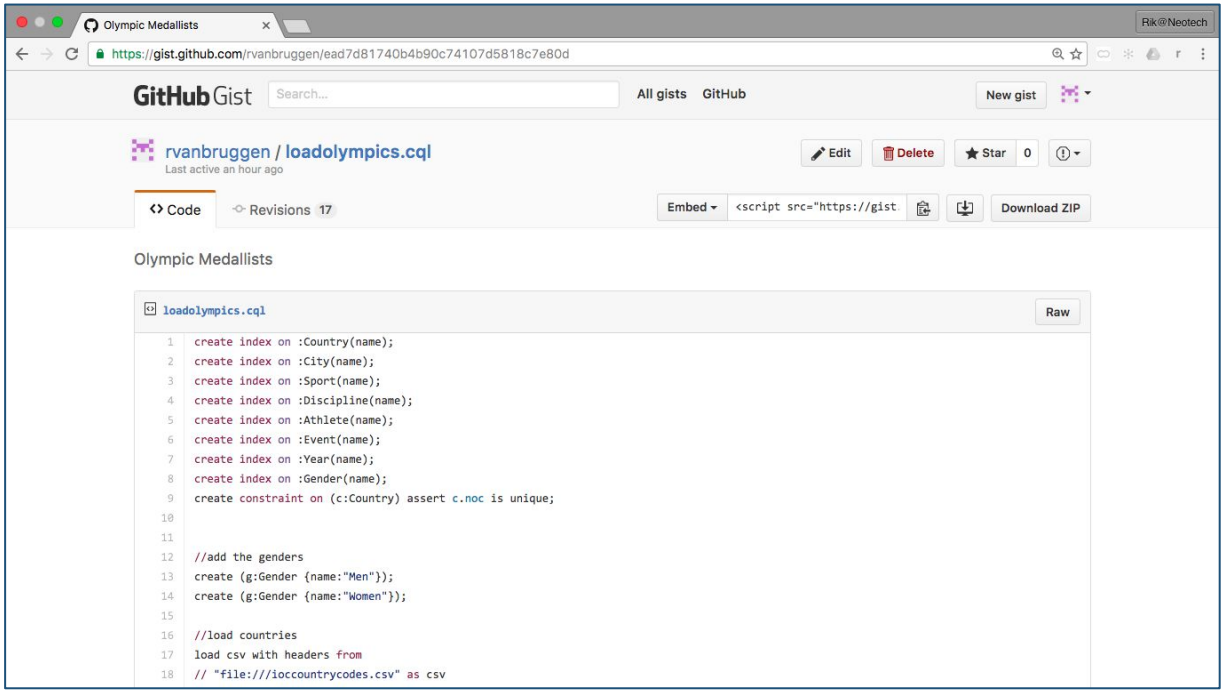
Returned 7 rows in 142 ms.



Try it yourself!

[The Load script]

[The Olympic Queries]



Questions?

OLYMPIC Answers!