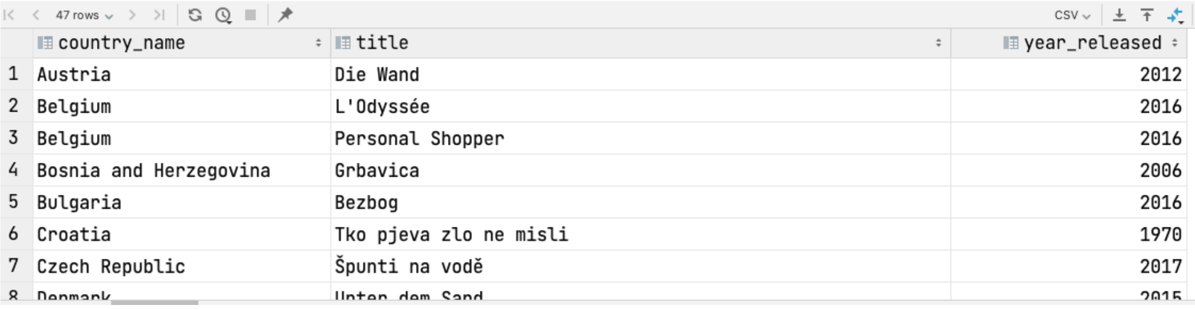
PostgreSQL is recommended to be run in this lab.

Window Functions

Regular Aggregate functions used as Window functions

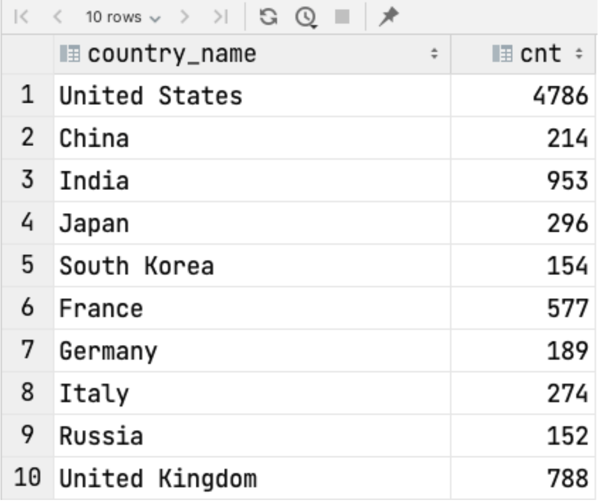
1. **What is the most recent film (title and year) for every Europe country, ordered by country.**



	country_name	title	year_released
1	Austria	Die Wand	2012
2	Belgium	L'Odyssée	2016
3	Belgium	Personal Shopper	2016
4	Bosnia and Herzegovina	Grbavica	2006
5	Bulgaria	Bezbog	2016
6	Croatia	Tko pjeva zlo ne misli	1970
7	Czech Republic	Špunti na vodě	2017
8	Denmark	Unter dem Sand	2015

```
select country_name, title, year_released
from (select c.country_name, m.title, m.year_released, max(m.year_released)
over (partition by c.country_name) as most_recent
      from movies m
           join countries c
             on c.country_code = m.country
      where c.continent = 'EUROPE') x
where most_recent = year_released
order by country_name, title
```

2. Countries in the database with more films than the average, with a common table expression and with a window function

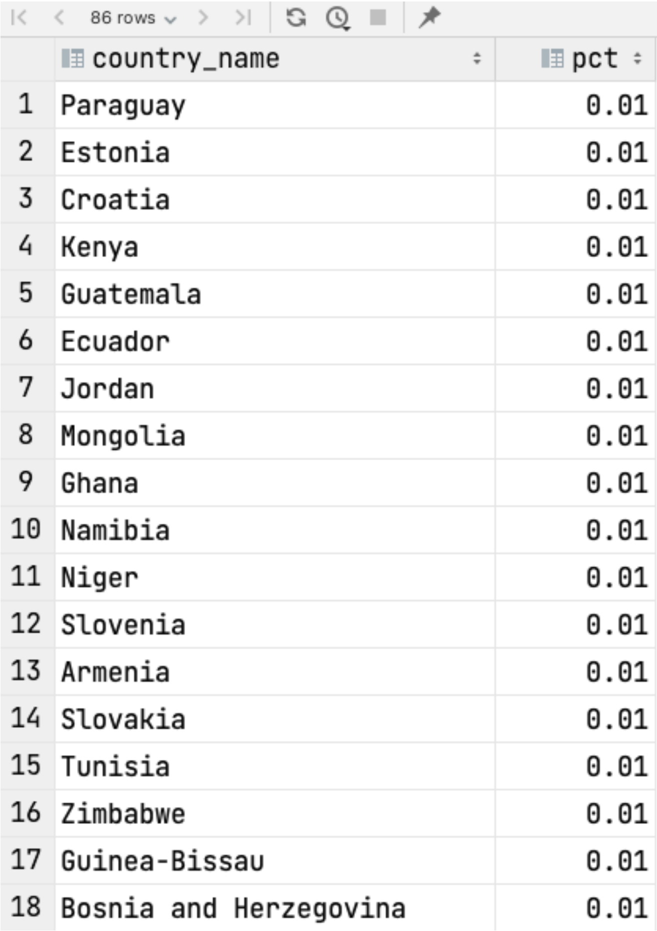


	country_name	cnt
1	United States	4786
2	China	214
3	India	953
4	Japan	296
5	South Korea	154
6	France	577
7	Germany	189
8	Italy	274
9	Russia	152
10	United Kingdom	788

```
with country_films as
   (select country, count(*) as cnt
    from movies
     group by country)
select c.country_name, x.cnt
from (select country, cnt
      from country_films
      where cnt > (select avg(cnt)
                   from country_films)) x
      join countries c
        on c.country_code = x.country
```

```
select c.country_name, cnt from
  ( select country, cnt, avg(cnt) over() as average
   from (select country, count(*) as cnt
       from movies
       group by country) m
       ) x
       join countries c
       on c.country_code=x.country
       where x.cnt>x.average
```

3. **Which percentage of the films of the database does every country represent, ordered by increasing percentage? (use a Window function)**

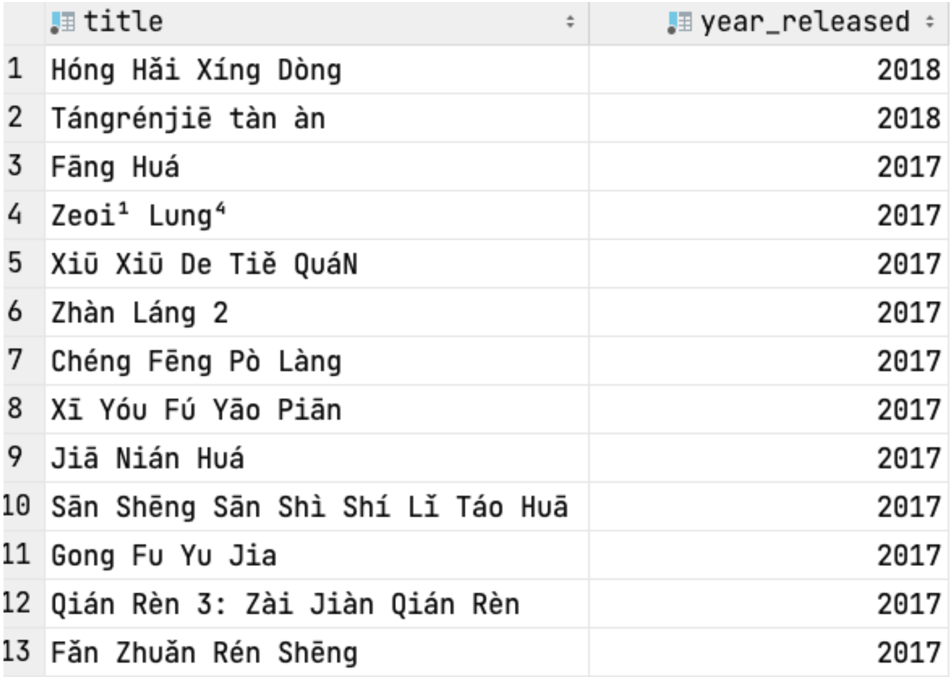


	country_name	pct
1	Paraguay	0.01
2	Estonia	0.01
3	Croatia	0.01
4	Kenya	0.01
5	Guatemala	0.01
6	Ecuador	0.01
7	Jordan	0.01
8	Mongolia	0.01
9	Ghana	0.01
10	Namibia	0.01
11	Niger	0.01
12	Slovenia	0.01
13	Armenia	0.01
14	Slovakia	0.01
15	Tunisia	0.01
16	Zimbabwe	0.01
17	Guinea-Bissau	0.01
18	Bosnia and Herzegovina	0.01

```
select c.country_name, round(100.0 * a.films / sum(a.films) over(), 2) as pct
from (select country, count(*) as films
      from movies
      group by country) a
     join countries c
       on c.country_code = a.country
 order by pct
```

[Ranking](#)

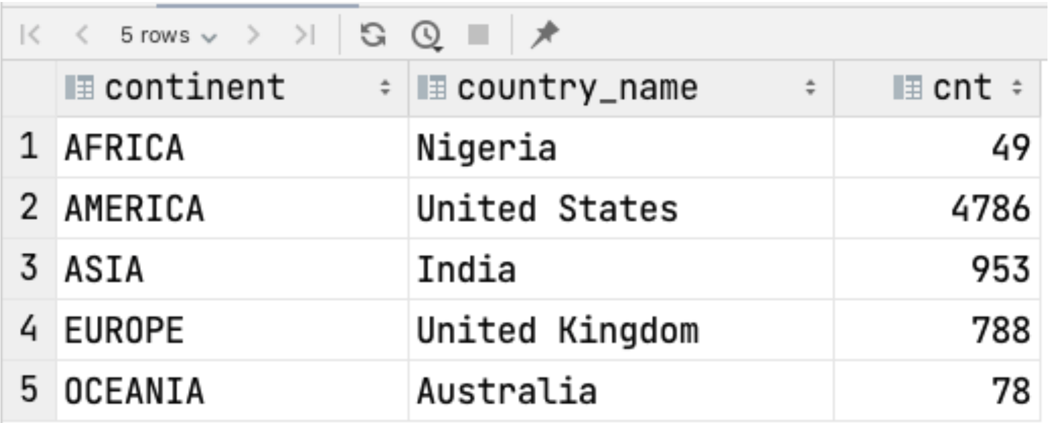
4. What are the title and year of the ten most recent films from China?



	title	year_released
1	Hóng Hǎi Xíng Dòng	2018
2	Tángrénjiē tàn àn	2018
3	Fāng Huá	2017
4	Zeoi¹ Lung⁴	2017
5	Xiū Xiū De Tiě QuáN	2017
6	Zhàn Láng 2	2017
7	Chéng Fēng Pò Làng	2017
8	Xī Yóu Fú Yāo Piān	2017
9	Jiā Nián Huá	2017
10	Sān Shēng Sān Shì Shí Lǐ Táo Huā	2017
11	Gong Fu Yu Jia	2017
12	Qián Rèn 3: Zài Jiàn Qián Rèn	2017
13	Fǎn Zhuǎn Rén Shēng	2017

```
select title, year_released
from (select title, year_released,
             rank() over (order by year_released desc) rnk
      from movies
      where country = 'cn') recent
where rnk <= 10
order by rnk
```

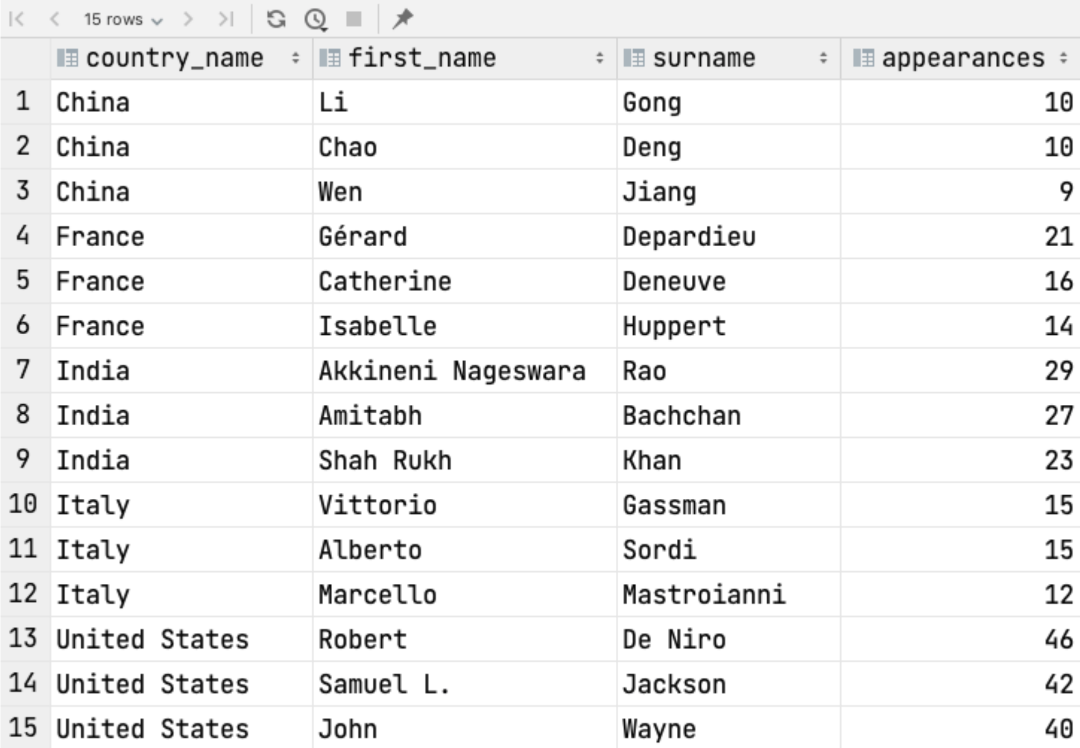
5. **What is by continent the country with the most movies in the database.**



	continent	country_name	cnt
1	AFRICA	Nigeria	49
2	AMERICA	United States	4786
3	ASIA	India	953
4	EUROPE	United Kingdom	788
5	OCEANIA	Australia	78

```
select continent, country_name, cnt
from (select continent,
             country_name,
             cnt,
             rank() over (partition by continent order by cnt desc) as rnk
      from (select c.continent, c.country_name, count(m.movieid) as cnt
            from countries c
                     join movies m
                          on m.country = c.country_code
            group by c.continent, c.country_name) x) y
where rnk = 1
```

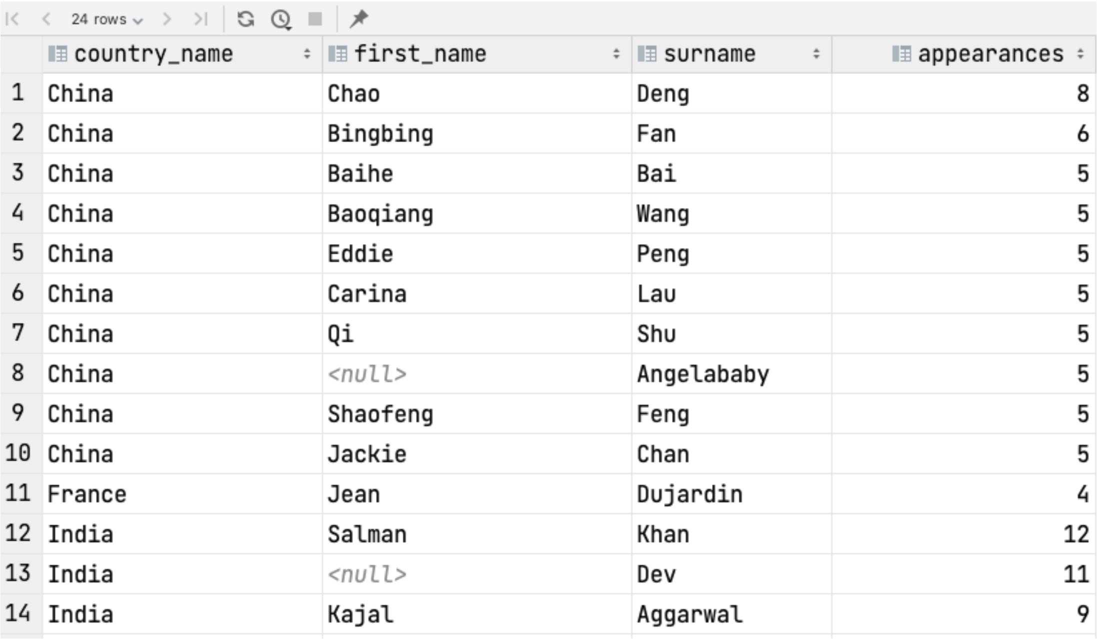
6. **What are, by country, the top three actors that are found most often in films from China, the United States, France, Italy and India**



	country_name	first_name	surname	appearances
1	China	Li	Gong	10
2	China	Chao	Deng	10
3	China	Wen	Jiang	9
4	France	Gérard	Depardieu	21
5	France	Catherine	Deneuve	16
6	France	Isabelle	Huppert	14
7	India	Akkineni Nageswara	Rao	29
8	India	Amitabh	Bachchan	27
9	India	Shah Rukh	Khan	23
10	Italy	Vittorio	Gassman	15
11	Italy	Alberto	Sordi	15
12	Italy	Marcello	Mastroianni	12
13	United States	Robert	De Niro	46
14	United States	Samuel L.	Jackson	42
15	United States	John	Wayne	40

```
select y.country_name, p.first_name, p.surname,  y.cnt as appearances
from (select x.country_name,
             x.peopleid,
             x.cnt,
             rank() over (partition by x.country_name
                          order by x.cnt desc) as rnk
      from (select co.country_name, c.peopleid, count(*) as cnt
            from credits c
                 join movies m
                   on m.movieid = c.movieid
                 join countries co
                   on co.country_code = m.country
            where c.credited_as = 'A'
              and co.country_name in ('China', 'United States',
                                      'France', 'Italy', 'India')
             group by co.country_name, c.peopleid) x) y
      join people p
        on p.peopleid = y.peopleid
 where y.rnk <= 3
 order by country_name, appearances desc
```

7. **Modify the preceding query to get for the same countries the top three actors that have appeared more than 3 times since 2010 (included)**

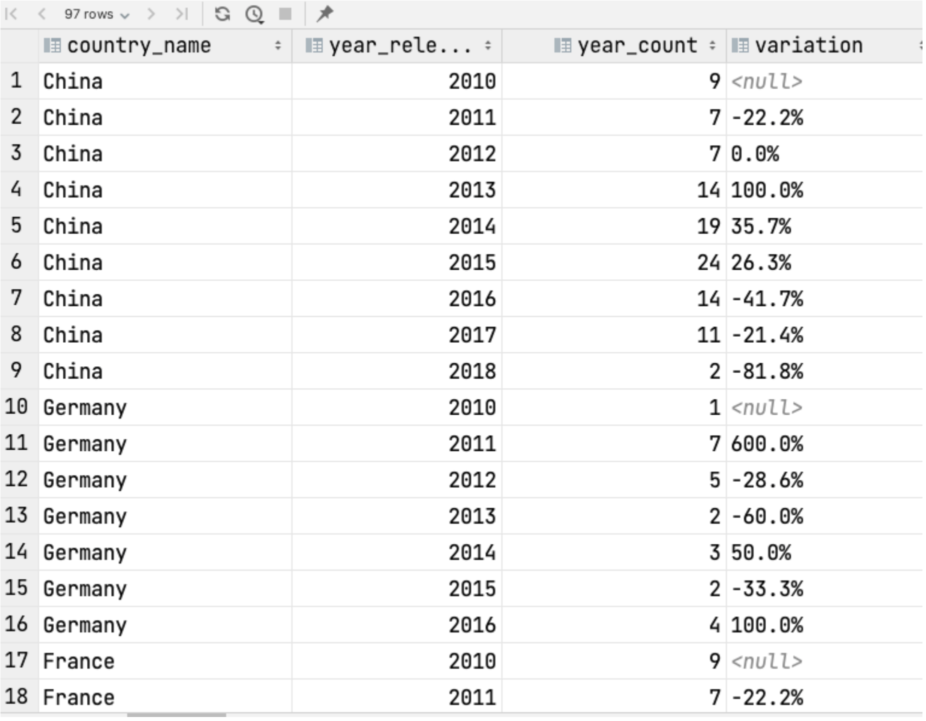


	country_name	first_name	surname	appearances
1	China	Chao	Deng	8
2	China	Bingbing	Fan	6
3	China	Baihe	Bai	5
4	China	Baoqiang	Wang	5
5	China	Eddie	Peng	5
6	China	Carina	Lau	5
7	China	Qi	Shu	5
8	China	<null>	Angelababy	5
9	China	Shaofeng	Feng	5
10	China	Jackie	Chan	5
11	France	Jean	Dujardin	4
12	India	Salman	Khan	12
13	India	<null>	Dev	11
14	India	Kajal	Aggarwal	9

```
select y.country_name, p.first_name, p.surname,  y.cnt as appearances
from (select x.country_name,
             x.peopleid,
             x.cnt,
             dense_rank() over (partition by x.country_name
                                order by x.cnt desc) as rnk
      from (select co.country_name, c.peopleid, count(*) as cnt
            from credits c
                 join movies m
                   on m.movieid = c.movieid
                 join countries co
                   on co.country_code = m.country
            where c.credited_as = 'A'
              and m.year_released >= 2010
              and co.country_name in ('China', 'United States',
                                      'France', 'Italy', 'India')
             group by co.country_name, c.peopleid
             having count(*) > 3) x) y
      join people p
        on p.peopleid = y.peopleid
 where y.rnk <= 3
 order by country_name, appearances desc
```

Other Window functions

8. **For countries for which we have at least 20 films released in 2010 or later, display for each year the year, the name of the country, the number of films, and the percentage variation since the preceding year (use the lag() function)**



	country_name	year_rele...	year_count	variation
1	China	2010	9	<null>
2	China	2011	7	-22.2%
3	China	2012	7	0.0%
4	China	2013	14	100.0%
5	China	2014	19	35.7%
6	China	2015	24	26.3%
7	China	2016	14	-41.7%
8	China	2017	11	-21.4%
9	China	2018	2	-81.8%
10	Germany	2010	1	<null>
11	Germany	2011	7	600.0%
12	Germany	2012	5	-28.6%
13	Germany	2013	2	-60.0%
14	Germany	2014	3	50.0%
15	Germany	2015	2	-33.3%
16	Germany	2016	4	100.0%
17	France	2010	9	<null>
18	France	2011	7	-22.2%

```
select c.country_name,
       year_released,
       year_count,
       round(100.0*(year_count - previous_year_count)
            / case previous_year_count
                 when 0 then null
                 else previous_year_count
               end, 1) || '%' as variation
from (select country, year_released,
             year_count,
             lag(year_count, 1)
                  over (partition by country
                           order by year_released) as previous_year_count,
             sum(year_count)
                  over (partition by country) as country_count
      from (select country, year_released,
                   count(*) as year_count
            from movies
            where year_released >= 2010
            group by country, year_released) a) b
      join countries c
       on c.country_code = b.country
where country_count >= 20
```