



Created by Mathis Dubrul
from Noun Project

Quantifying gender gap among movie directors



Alice Corona

alice-corona.nl

@alice_corona

Data Driven Journalist · Data Skills Trainer · Researcher

Shift_Shape_Mobilize · Odessa, 1-6 August 2016

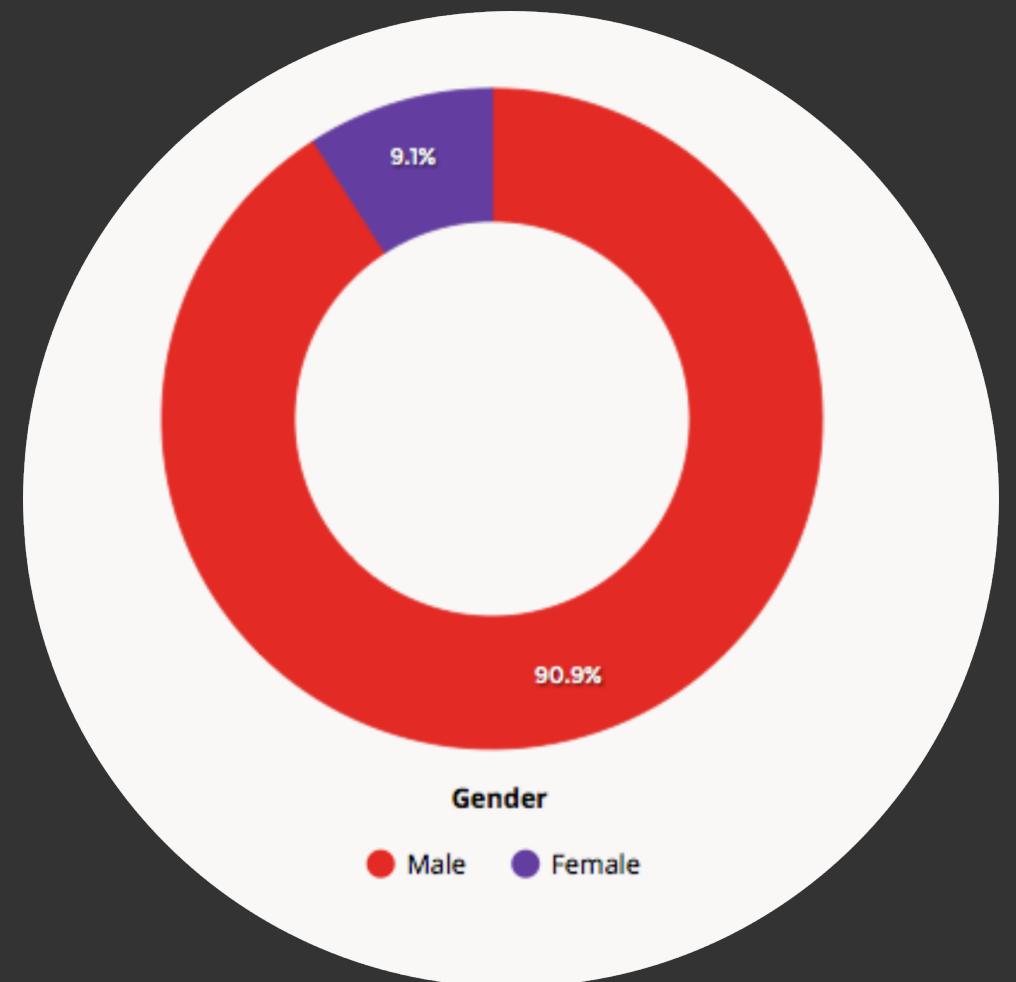
Women at the International Film Festivals of 2015 - 2016

Tracking women's presence at major film competitions

“Overall, only 13% of the movies aired at Venice will be directed by a lady. Good news is that the rate rises 19% for movies that are in competition for an award.

This is mainly due to the movies competing in the Horizon category, the experimental section which groups films representing new aesthetic trends and expressive languages. Here, one movie every four has a women director.

Yet, the movies in the official and most prestigious selection have a much lower female presence. **Only 9% of the movies competing in the main Venice Film Festival selection have a women director. That's less than one every 10.**



Alice Corona

alice-corona.nl

@alice_corona

Data Driven Journalist · Data Skills Trainer · Researcher

Shift_Shape_Mobilize · Odessa, 1-6 August 2016

Women at the International Film Festivals of 2015 - 2016

Tracking women's presence at major film competitions

	A	B	C	D	E	F	G	H
1	Movies	Festivals	In Competition?	In Competition?	In Competition?	In Competition?	Category (Venice)	Category (Toronto)
2	(Otto)	Toronto International Film Festival	Yes					Shortcuts
3	(Self)exhibitions	International Documentary Film Festival (Amsterdam)	No					
4	#100humans	Sundance Film Festival				No		
5	10 Billion - What's on Your Plate	International Documentary Film Festival (Amsterdam)	No					
6	11 Minut (11 Minutes)	Venice Internatio	Yes				Official Selection	Masters
7	11.22.63	Sundance Film Festival				No		
8	25 April	Toronto International Film Festival						Contemporary World C
9	3000 Nights	Toronto International Film Festival						Contemporary World C
10	31	Sundance Film Festival				No		
11	3½ Minutes Ten Bullets	International Documentary Film Festival (Amsterdam)	No					
12	3D Movie	Toronto International Film Festival						Wavelengths
13	4 Quarters	Toronto International Film Festival	Yes					Shortcuts
14	45 Years	Toronto International Film Festival						Special Presentations
15	55 Pastillas	Venice Internatio	Yes				Horizons	
16	6x9: An Immersive Experience	Sundance Film Festival				No		
17	7 sheep	Toronto International Film Festival	Yes					Shortcuts
18	88:88	Toronto International Film Festival						Wavelengths
19	A Bigger Splash	Venice Internatio	Yes				Official Selection	



Alice Corona

alice-corona.nl

@alice_corona

Data Driven Journalist · Data Skills Trainer · Researcher

Shift_Shape_Mobilize · Odessa, 1-6 August 2016

This is a step-by-step tutorial on how to:

- Use **APIs** to add movie information (such as plot, producing country...)
- Use some **Google Sheet advanced functions**
- Handle **JSON data** with **Open Refine** and convert it to a spreadsheet
- Get you data clean through **clustering** and **regex**
- Automatically guess the gender of a name thanks to **APIs**



Alice Corona

alice-corona.nl

@alice_corona

Data Driven Journalist · Data Skills Trainer · Researcher

Shift_Shape_Mobilize · Odessa, 1-6 August 2016

Let's start!



Alice Corona

alice-corona.nl

@alice_corona

Data Driven Journalist · Data Skills Trainer · Researcher

Shift_Shape_Mobilize · Odessa, 1-6 August 2016

1. Install **OpenRefine**
2. Have an empty **Google Drive spreadsheet** ready
3. Create an **IMDb account**
 - Create a list with your favorite movies/movies you watched recently/best sci-fi movies....whatever you like, as long as it has about 20-50 movies (the more, the better).
 - Export the list (go to the list, you'll see the option at the bottom of the page)



Alice Corona

alice-corona.nl

@alice_corona

Data Driven Journalist · Data Skills Trainer · Researcher

Shift_Shape_Mobilize · Odessa, 1-6 August 2016

Some disclaimers:

- This is the simple way. An harder but more efficient way is to **automate** part of the process with a script (I wrote one using Python, if you want to see what I mean)
- Automatic inference of gender from names should be only the first step. To avoid mistakes, I suggest taking additional **manual** precautions. For example:
 - Check a random sample of the data - or all the data when it's feasible.
 - The API is based mostly on American names. I'd manually google for pictures of directors with foreign names.
 - I'd do the same also for directors whose names could be both male or female. You can choose the threshold. For me, anything that has a <100% accuracy or a small sample size was worth double-checking.
 - Alternatively, you could also use different gender-guessing APIs to triangulate results.



Alice Corona

alice-corona.nl

@alice_corona

Data Driven Journalist · Data Skills Trainer · Researcher

Shift_Shape_Mobilize · Odessa, 1-6 August 2016

Want to try again?

alice-corona.nl/tutorials/open-refine-gender-directors



Alice Corona

alice-corona.nl
@alice_corona

Data Driven Journalist · Data Skills Trainer · Researcher

Shift_Shape_Mobilize · Odessa, 1-6 August 2016