

But When Sweden Gets 12 Points From Norway, It's Clearly Just Good Taste: The Determinants of Eurovision Success

Lauren Gilbert

Eurovision is an annual song contest among (mostly) European countries. It has been held since 1956; fifty-one countries have participated at least once, though the number of participants in any given year hovers around 40. Hundreds of millions of people watch every year, and voting patterns are widely interpreted in terms of political trends (e.g. Conchita Wurst's win in 2014). There is also a substantial betting market on Eurovision results.

There is a surprisingly robust literature about voting patterns in Eurovision. Fenn 2006 showed that, considering 1992-2003, countries did form "cliques" that were likely to vote for each other. Yair and Maman 1996 divide Europe into four such voting blocks, but Gatherer 2006 finds five and Clerides and Stengos 2006 find only three.

Ginsburgh and Noury 2008 argue that the blocks arise from linguistic and cultural similarity, rather than political ties. Garcia and Tanase 2014 go one step further and use a Eurovision "friend or foe" index to measure culture. Dekker 2007 and Spierdijk and Vellekoop 2006 also note that immigrant populations can sway national votes toward the countries they emigrated from.

However, as one can see above, this literature is becoming dated. The set of Eurovision Song Contests has increased nearly 20% since the mid-2000s burst of Eurovision scholarship. Most of the above papers include data only through 2003. Since then, Andorra, Armenia, Australia¹, Azerbaijan, Belarus, Bulgaria, the Czech Republic, Georgia, Moldova, Montenegro, San Marino, Serbia and Serbia and Montenegro² have joined the contest for at least one year.

Additionally, the scoring structure has changed substantially since 2003. From 1997 to 1956, points were decided by a panel of judges; from 1998-2003, some countries used judges while some used country-wide televotes from ordinary citizens; from 2003-2008, all countries used televotes; from 2009 onwards, countries have used a combination of jury votes and televotes. Shifting from a unelected (elite) jury to a popular vote could have substantial impact on the voting patterns.

We wish to focus on a few possible explanatory variables for Eurovision performance:

1. Geographic, ethnic, and religious voting - do countries prefer to vote for their neighbors?

Fenn 2006 finds that geographic block voting was alive and well in the 1992-2003 period. Does this persist into the age of the televote era?

In order to differentiate geographic voting from cultural voting (a la Ginsburgh and Noury 2008), we will need to quantify linguistic similarity, geographic distance, religious similarity and cultural similarity. It would also be helpful to have data on the size of immigrant groups in each country.

2. Themes - is it important to submit a love song?

The tendency to submit meaningless pop songs about love to Eurovision is so well-known that it has been parodied at the show itself. But is this actually true? We use dictionary methods and a database of song lyrics to quantify just how true this is.

3. Language - do songs in English do better?

¹No, Australia is not in Europe.

²Serbia and Montenegro appeared as one country in 2004 and 2005; after the country split in two, Serbia and Montenegro have submitted separate entries.

When Eurovision began, most countries submitted entries in their native language. As time has gone on, this has changed; most countries submit songs in English (and it is sometimes perceived as a political statement if they do not; see France). But is this an advantage? Should countries submit in English? We wish to test if there is a relationship between song language and overall score.

(This is differentiated from linguistic similarity because it is based on the language in which the entry is submitted, not the similarity between the two countries' official languages.)

Data

A database of scores 1975-2018 is available on Kaggle. Since there were not semi-finals for much of Eurovision, we consider only votes cast in the final.

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

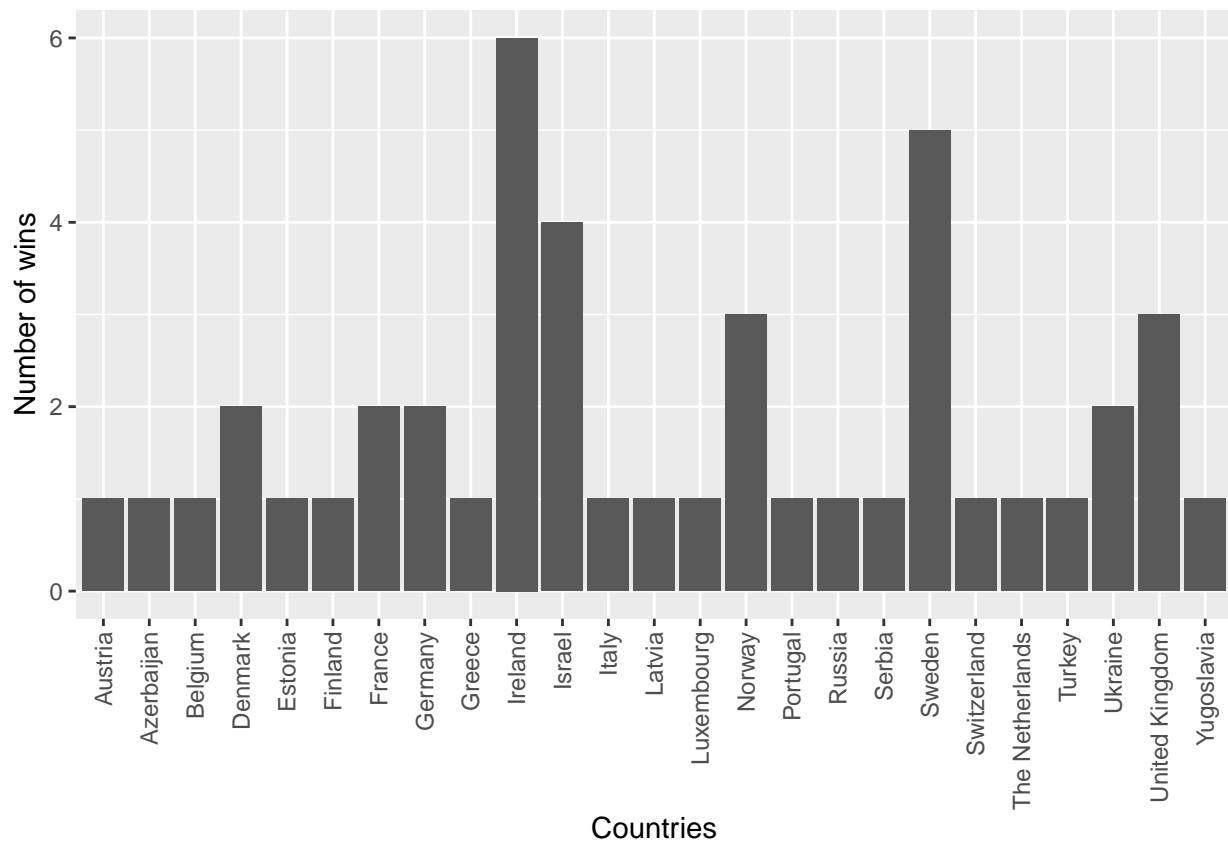
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

For covariates, we will use the following:

1. Proximity: Kaggle dataset on capitals with latitude and longitude data. We can use the geosphere package to calculate capital-to-capital distances.
2. Linguistic/cultural/religious similarity: I am still researching the best index to use for these variables. (For obvious reasons, they are calculated different ways in different papers.)
3. Language: since this is not contained in the dataset with scores, we will have to scrape language data from the Wikipedia page of each year of the ESC
4. Song content: lyrics data will be acquired from this Kaggle dataset. In order to match lyrics with songs, we will need to use song titles scraped from the Wikipedia page for each year of the ESC.

Descriptive statistics

For fun, we create a histogram of the countries that have won most often:



```
## Loading required package: xml2
```

Out-of-sample prediction

Eurovision 2019 will be held on 14 - 18 May. I can test how successful my predictors of Eurovision success are against real (currently unknowable) data.