

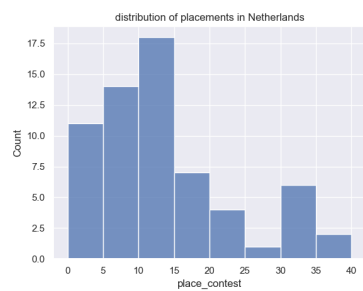
Data visualization assignment 1 – Bas Reterink



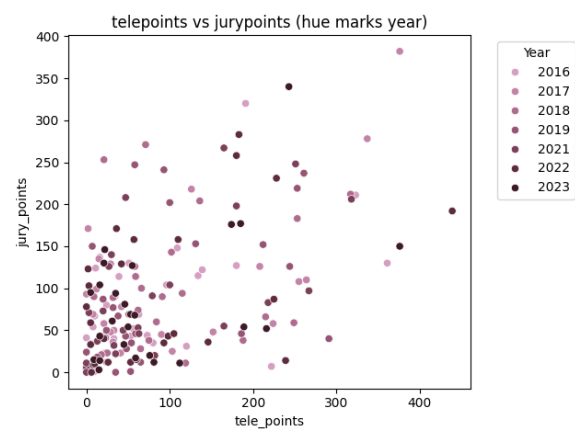
Note

After discussing with Andreas (who was okay with this), I did this assignment in python instead of tableau because I want to learn more about using pandas for data visualization. I included my code as well on github (<https://github.com/basRet/assignment-2-eurosong>). Since I could not create a tableau story for assignment 5, I wrote a story instead.

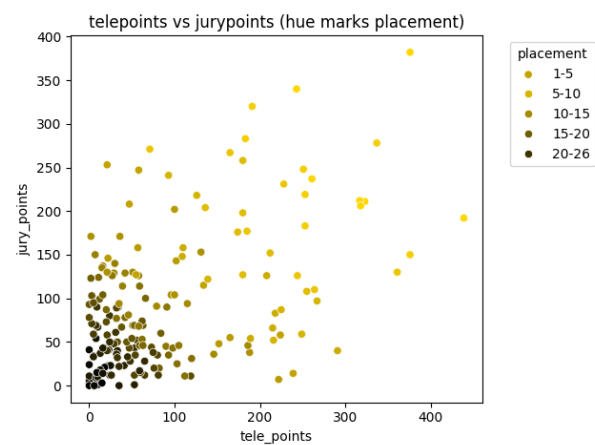
Assignment 1



Assignment 2



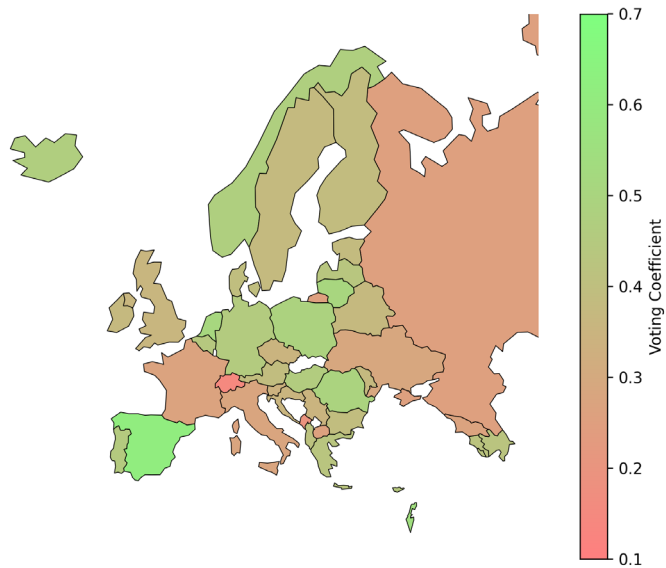
Assignment 3



Assignment 4

I tried to plot how good a country was at voting close to the final placement. In this way, the audience votes of a country that scores highly on this could be used to accurately predict the final placement of other countries for each year. This might give interesting insight into the watching behavior of various countries regarding the Eurovision (do they watch and vote “competitively” or not). I plotted this information on a map of Europe, using a choropleth map with geopandas:

Voting prediction power per European country population in Eurovision (2016-2023)



In order to do this on a technical level, I created a coefficient that measures how similar the ranking of the country’s votes is to the real ranking, where 0 corresponds to an inverse order (the country votes on the last place the most and the first place the least) and 1 corresponds to a completely “correct” prediction (the country votes on the first place the most and the last place the least).

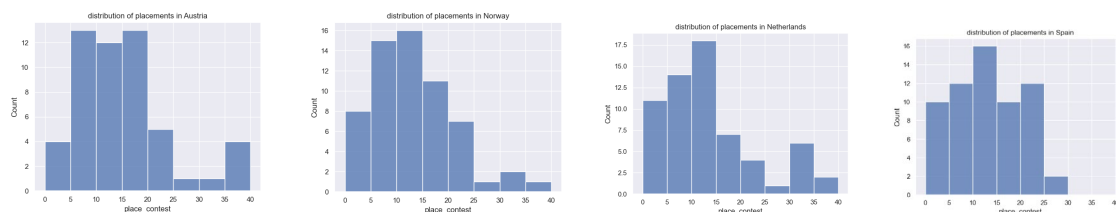
This was done by, for each country A , calculating the difference between 1. each placement of country B and 2. the ranking in terms of how much country A voted on country B . Then, these values are normalized with regards to total amount of countries voted on (minus one) and normalized again to obtain a final coefficient between 0 and 1. This is done for each year in which tele vote information was available and then averaged, to obtain a final coefficient. The more exact formulation can be found in my code.

Assignment 5

For the first assignment, There seems to be a tremendous difference in the placements of countries. Most countries look somewhat similar to Lithuania, seemingly random. However, some countries, such as Sweden, also consistently score higher over the years.

For the second and third assignment, it seems like the jury and televotes disagree quite often. This makes intuitive sense, since the jury is a very small sample of people who work in the same field and know the same people, meaning that they likely have different tastes than the general public. It is not surprising that entries that place higher are also voted on more often, since this is how the contest's placement is decided

Assignment five also gives remarkable results. It seems that some countries are much better at predicting winners than others. Especially Spain is very good and Austria is very bad. By comparing the placements of these countries, it seems like perhaps, highly performing countries such as the Netherlands and Norway are also better at predicting winners. However, Spain is an obvious outlier here.



A major caveat for my interpretation of assignment 5 is that for calculating the coefficient, only the televotes are taken into account, which is data over only a couple of years. To get more reliable information to draw conclusions on, it might be useful to find a way to calculate the coefficient for earlier years as well, using jury vote information instead.

On a personal level, I found that I really enjoyed the process of analyzing data in this way.