Predicting Top Chart Music Consumption Concentration with Spotify Streaming Data

Forecasting and Time Series, Fall 2021

Henry Gu, Yixuan Yu, Kevin Liu, John Warlick, and Jack Sampiere

Motivation

Popular media has historically been described as a "winner-take-all" market in which a small portion of content accounts for the majority of consumption. Unsurprisingly, Spotify streaming volume is no exception. This has strong implications for artists and managers aiming to gain recognition among listeners, who tend to pay almost exclusive attention to tracks in the top charts or in well-known playlists on the platform. As a result, rather than spending time tailoring production to the types of music comprising the top charts and relying on an almost non-existent chance of infiltrating these playlists, artists continually seek placement in less well-known playlists that fit their style and have a growing audience. The extent to which this approach is taken is dictated by the extent of concentration of consumption within the top charts—should this "winner take-all" effect strengthen over a certain time period, artists may react by continuing to disregard top chart ambitions and vice versa. Predicting the extent of this concentration at the top of the market thus has strong implications regarding the optimal path for artists and managers to take in advancing their careers.

Data & Approach

We used daily streaming data from Spotify's top 200 songs across 10 countries—Australia, Brazil, Canada, Denmark, France, Great Britain, Mexico, the Netherlands, Sweden, and the US—from June 29, 2017 through January 26, 2018. To analyze the concentration of music consumption within the top charts, we forecast three quantities: total streams within the top 10 tracks, the top 50 tracks, and the top 200 tracks. By forecasting the number of streams within each of these bins and comparing them, we can evaluate how strong the "winner-take-all" effect is at a given time.

Two weeks worth of data was set aside for a validation set, and another two weeks for a test set. The validation set was used to determine optimal model parameters. Models were then re-trained with a combination of the initial training set and the validation set and evaluated with the test set.

Analyses

We first aggregated streams across each bin to create three time series for each country that would be used for predictions. Initial modeling efforts were heavily inaccurate due to abnormal trends near the end of the time series in the holiday season. Due to a lack of data, this yearly seasonality was problematic and led to stark differences within training, validation, and test sets. We thus chose to use only a subset of the data (before December 1) to address this issue.

We first attempted to use ARIMA models to forecast the top stream volumes for each country. The models for each country were chosen using the auto.arima function. Theoretically, we allowed each country to have its own governing ARIMA model as this would allow country-specific nuances in the data to be accounted for. However, we discovered that using a simple ARIMA model did not appear to have enough information for a good forecast for many countries—often, simply a static value was predicted for all time periods. This seemed to be a poor forecasting model, so we explored other options.

Suspecting a strong correlation between stream volumes in different countries, we decided to implement a vector autoregression (VAR) model. Upon checking for stationarity, we realized that streaming volumes for Australia and France were nonstationary and thus excluded them from the VAR models (Figure 1). We then wanted to investigate which countries had the most heavily correlated stream volumes, which we did by investigating ACFs across the various countries. This analysis showed that there were two primary subsets of correlated countries. The first subset contained Great Britain, Canada, the US, Mexico, and Brazil (Figure 2). These were used to build the first VAR model ("VAR Model 1"), while the second model ("VAR Model 2") contained data from Denmark, the Netherlands, and Great Britain. While streams in Great Britain were forecasted with VAR Model 1, this time series was still heavily correlated with stream volumes in Denmark and the Netherlands and was thus included as an additional predictor in VAR Model 2. Sweden was only minimally correlated with stream volumes from other countries and was thus excluded from both VAR models.

A key step in building each VAR model was to determine how many lag terms to include as well as which lags to include. We were able to use the ACFs in Figures 2 and 3 to get a sense of which lag terms to start with. We also tuned this parameter more robustly for each model by trying different combinations and using the setting that achieved the best performance on the validation set. The optimal lag combinations—which are displayed in Figure 4—demonstrate that, in addition to the expected lag terms (t-1, t-2), weekly seasonality is also strong based on the presence of terms several weeks in the past (t-7, t-14, etc.). These optimal lag terms were then implemented in the final models that were used to make predictions on the test set.

To predict streams in Australia, France, and Sweden, we leveraged a Prophet model due to its ability to deal with nonstationary data as well as to include country-specific holidays, which we suspected would have a strong influence on music consumption. We used the validation set to tune the changepoint.prior.scale parameter, which ended up having little to no effect on the quality of the forecasts.

Some extensions to the model we considered were weekly aggregation of the data and to create forecasts using growth rates. Weekly aggregation of the data was not a viable modeling method for this dataset due to the small number of time periods available—once the data was aggregated, we only had 23 samples to split between the training, validation, and test sets. This meant that we would either have too limited data to form a training set to generate a model, or too few weeks in the validation or test sets to accurately measure each model's performance. Thus, this extension was discarded. An additional extension was to convert the dataset to the growth rate in stream volumes between days and produce a forecast based on this growth rate. However, we found this to have little to no improvement to the quality of the forecasts.

Results

Figure 5 displays a table containing the performance on the test set for all countries for each of the three bins of stream volumes. Note that we use mean absolute percentage error (MAPE) to quantify performance—this is due to the fact that the magnitude of stream volumes varies greatly by country and thus RMSE values can be misleading. We can see from Figure 5 that music consumption is most predictable in Canada, Mexico, and Great Britain, indicating that VAR Model 1 is particularly useful. We also see that Australia is well forecasted on the basis of its own past consumption alone (recall that these forecasts were made with a univariate Prophet model). Additionally, in countries such as Denmark, anomalous data points at the beginning of December are poorly forecasted (Figures 6, 7, and 8), perhaps causing deceivingly high error values in Figure 5. A final noteworthy finding is that music consumption is consistently underestimated in Brazil and Australia (Figures 6, 7, and 8).

Insights

First, the strong performance of the VAR models upholds a premise that makes intuitive sense—that cultural tendencies span across country borders and that this manifests in music consumption. This is apparent in the correlation between streams in countries in the Americas and streams in Great Britain, and can be explained on the basis of proximity in addition to language similarities (Figure 9). For example, the prominence of English in Great Britain, Canada, Mexico, and the US as well as the prominence of Spanish in both Mexico and the US may result in similar music tastes. Additionally, concert venues in border cities like Toronto and Montreal may contribute to this effect. These ideas are upheld by the correlation between music consumption in Denmark and the Netherlands, which are similarly close to one another, as well as by the lack of correlation between other countries and France, Australia, and Sweden. Additionally, in both VAR models, the prominence of weekly lag terms (t-7, t-14, etc.) indicates that music consumption exhibits weekly seasonality. This makes sense as people may consume more music on the weekends during social outings and less during the week.

As the forecasts are most accurate for Canada, Mexico, Great Britain, and Australia, the concentration of music within the top charts is thus most predictable in these countries. As a result, artists and managers could benefit substantially from leveraging music consumption predictions in these countries to inform promotional action on platforms like Spotify to help advance their careers in these markets specifically. This is especially pertinent for artists who have minimal presence in these countries in the first place.

Appendix

Figure 1. Time series of stream volume for Australia (left) and France (right).

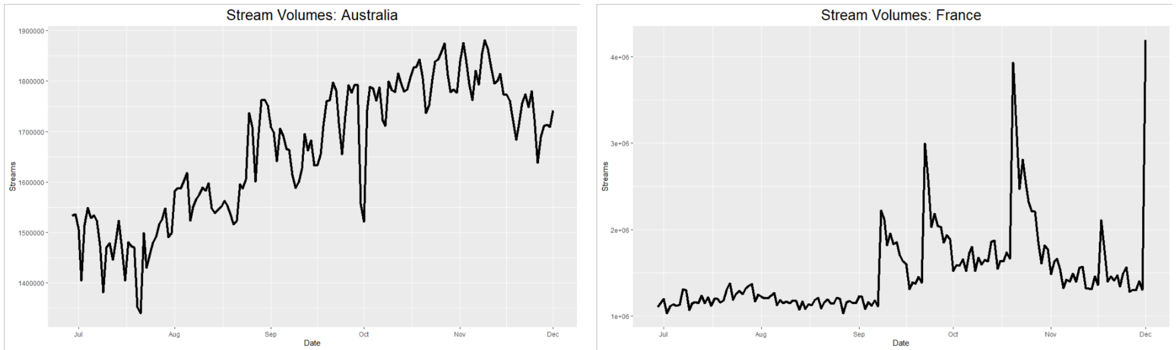


Figure 2. ACF across Great Britain, Canada, the US, Mexico, and Brazil.

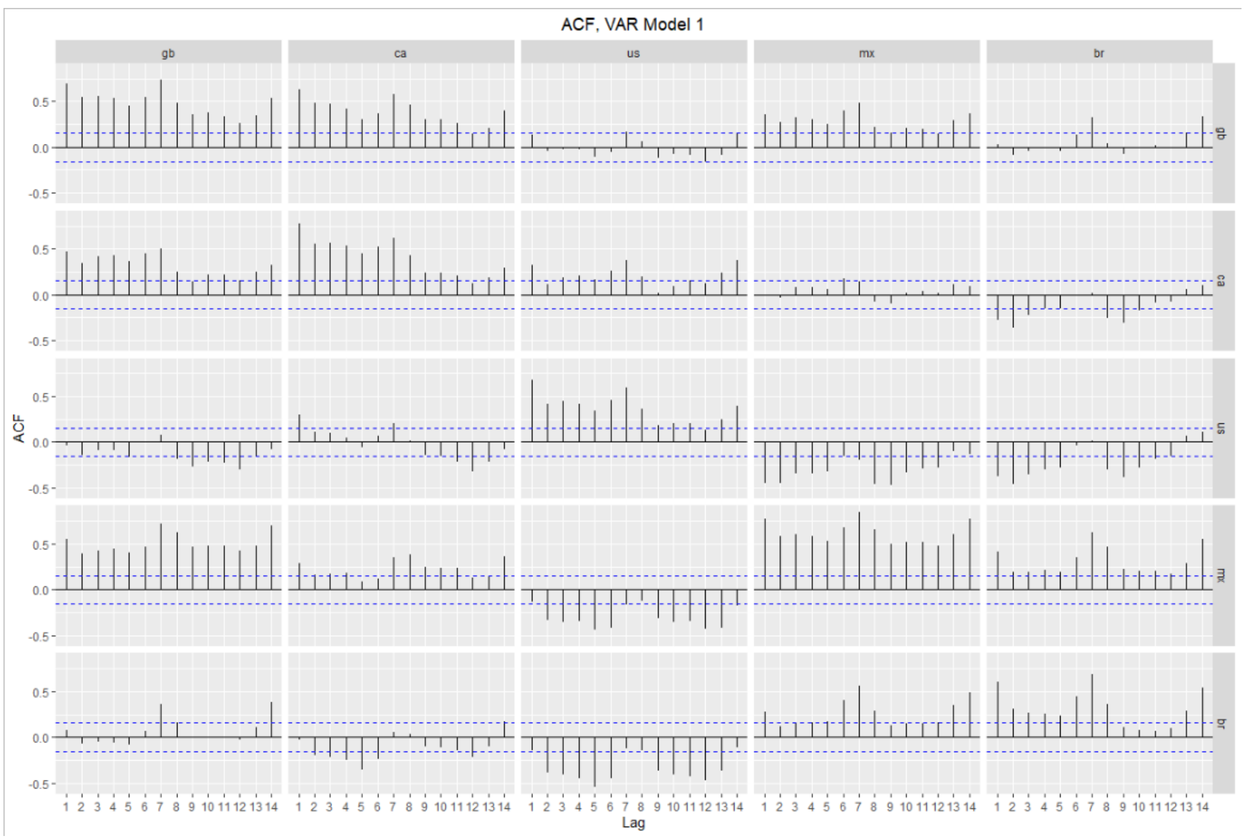


Figure 3. ACF across Denmark, the Netherlands, and Great Britain.

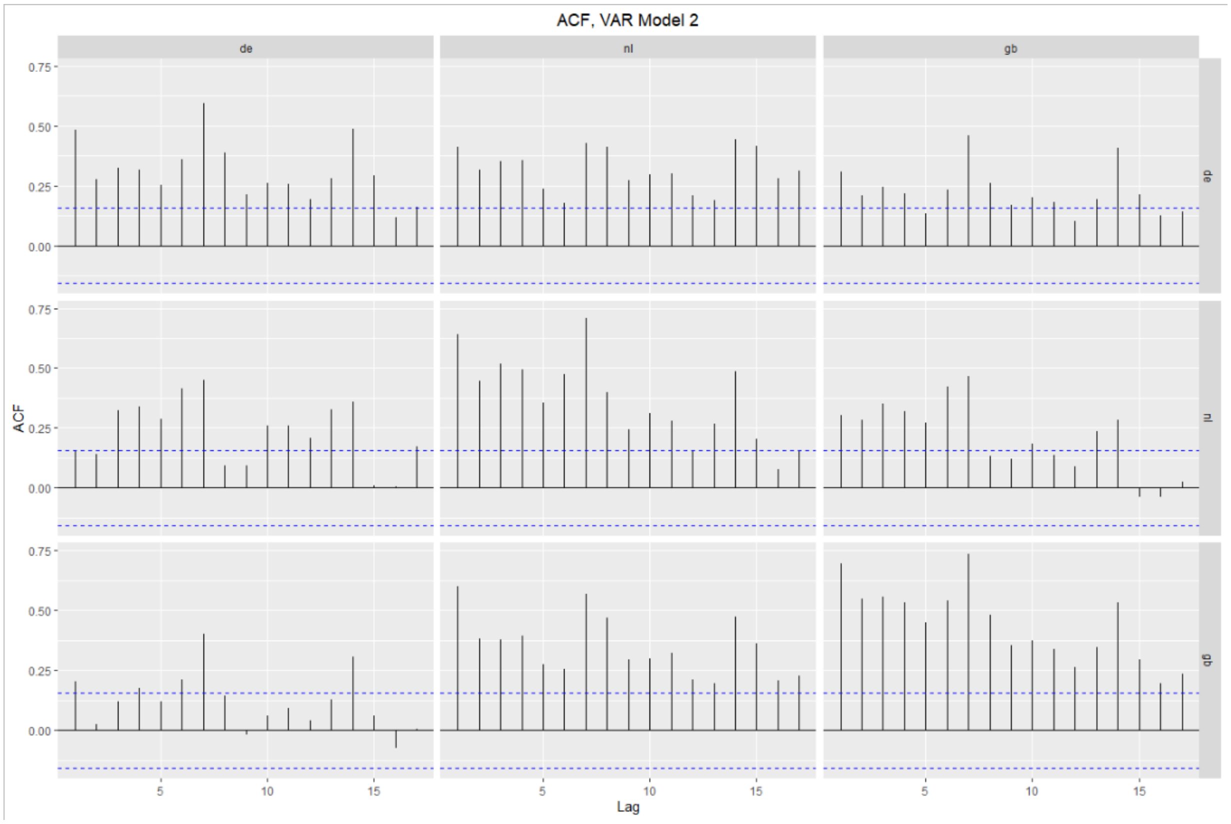


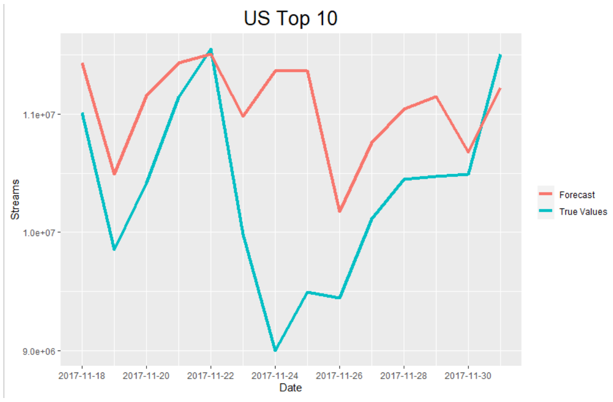
Figure 4. Optimal lag term combinations for each VAR model.

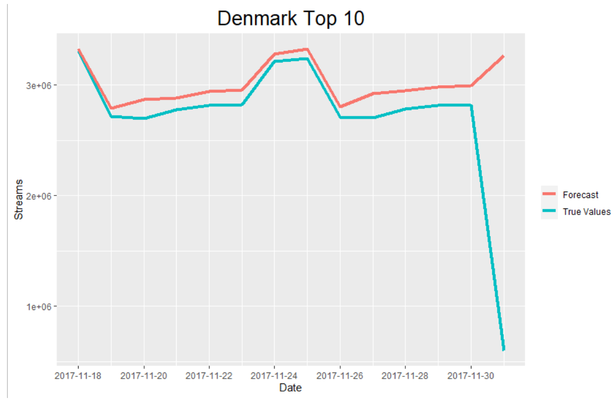
	Top 10	Top 50	Top 200
VAR Model 1	1, 2, 7, 8, 9	1, 2, 7, 8	1, 2, 7, 8, 14
VAR Model 2	1, 2, 7, 8	1, 2, 7, 8, 14, 15	1, 2, 7, 8, 9, 14, 15

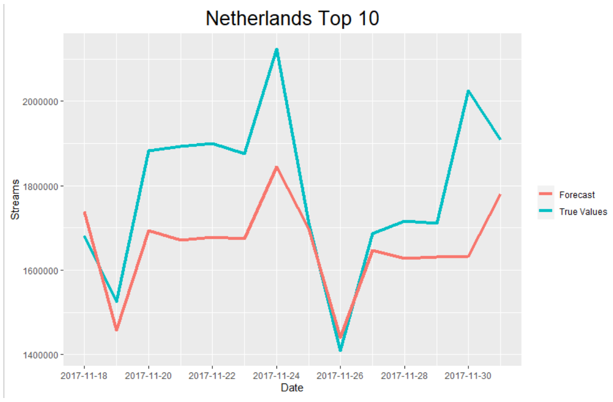
Figure 5. Test set performance (MAPE) of the optimal models for each country in each bin.

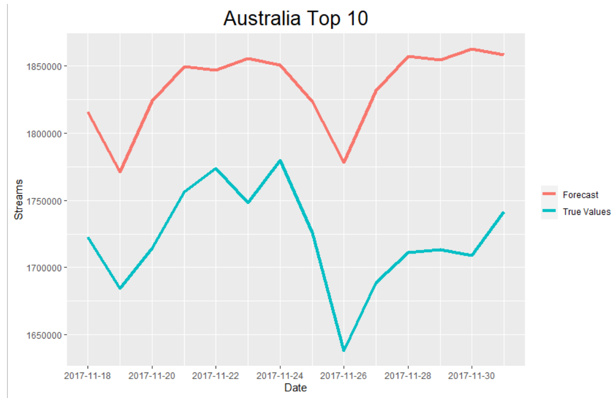
	Top 10	Top 50	Top 200
Australia	6.55	4.18	5.24
Brazil	19.49	6.74	5.55
Canada	2.49	3.40	5.04
Denmark	36.03	5.87	2.35
France	33.12	22.43	5.26
Great Britain	2.60	3.68	3.68
Mexico	2.84	2.65	3.40
Netherlands	7.60	6.11	2.42
Sweden	10.31	3.43	3.15
US	7.63	6.38	7.07

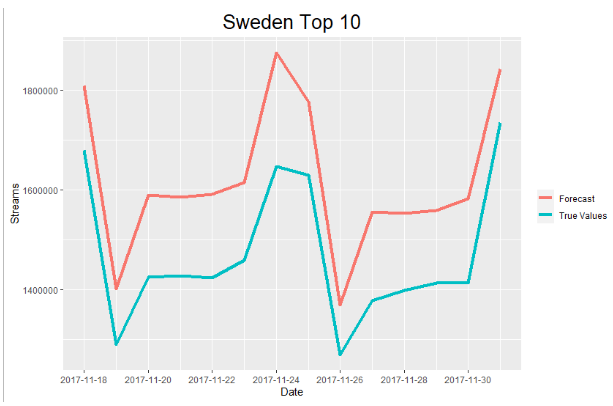
Figure 6. Forecasted top 10 stream volumes in each country.


US Top 10
Streams
1.1e+07
1.0e+07
9.0e+06
2017-11-18
2017-11-20
2017-11-22
2017-11-24
2017-11-26
2017-11-28
2017-11-30
Date
Forecast
True Values


Denmark Top 10
Streams
3e+06
2e+06
1e+06
2017-11-18
2017-11-20
2017-11-22
2017-11-24
2017-11-26
2017-11-28
2017-11-30
Date
Forecast
True Values


Netherlands Top 10
Streams
2000000
1800000
1600000
1400000
2017-11-18
2017-11-20
2017-11-22
2017-11-24
2017-11-26
2017-11-28
2017-11-30
Date
Forecast
True Values


Australia Top 10
Streams
1850000
1800000
1750000
1700000
1650000
2017-11-18
2017-11-20
2017-11-22
2017-11-24
2017-11-26
2017-11-28
2017-11-30
Date
Forecast
True Values


Sweden Top 10
Streams
1800000
1600000
1400000
2017-11-18
2017-11-20
2017-11-22
2017-11-24
2017-11-26
2017-11-28
2017-11-30
Date
Forecast
True Values

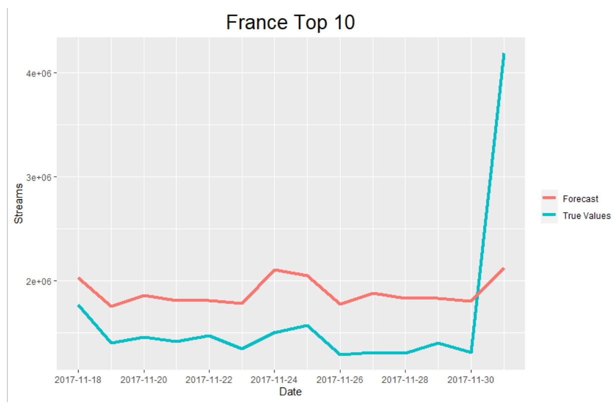
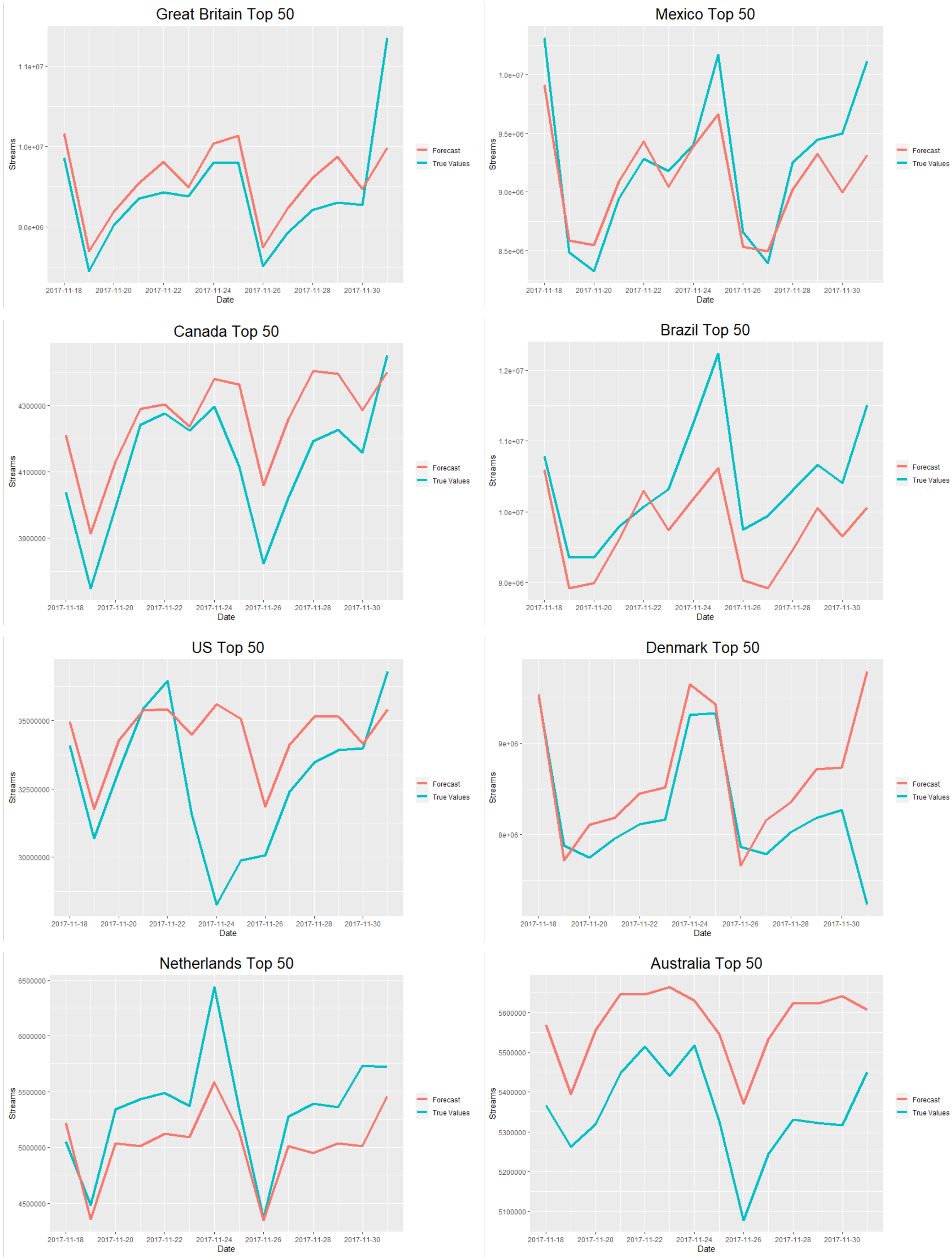

France Top 10
Streams
4e+06
3e+06
2e+06
2017-11-18
2017-11-20
2017-11-22
2017-11-24
2017-11-26
2017-11-28
2017-11-30
Date
Forecast
True Values

Figure 7. Forecasted top 50 stream volumes in each country.



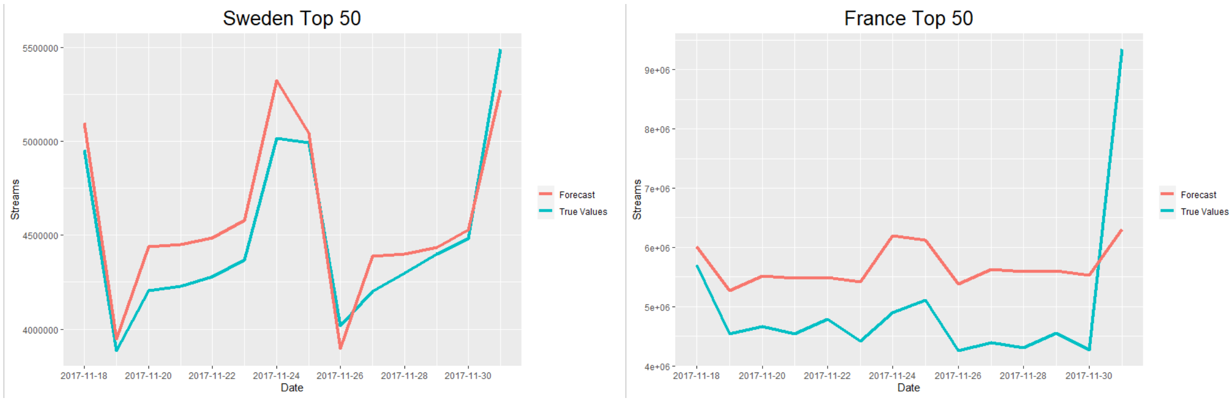
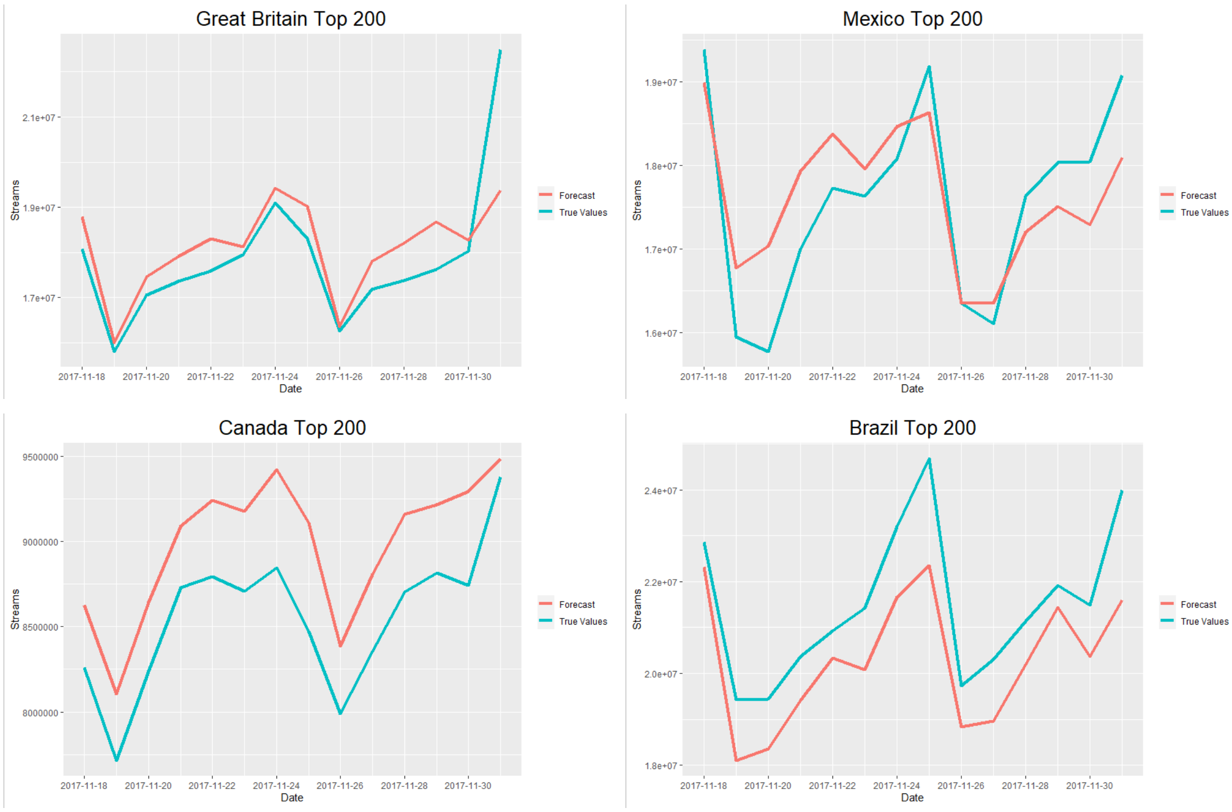


Figure 8. Forecasted top 200 stream volumes in each country.



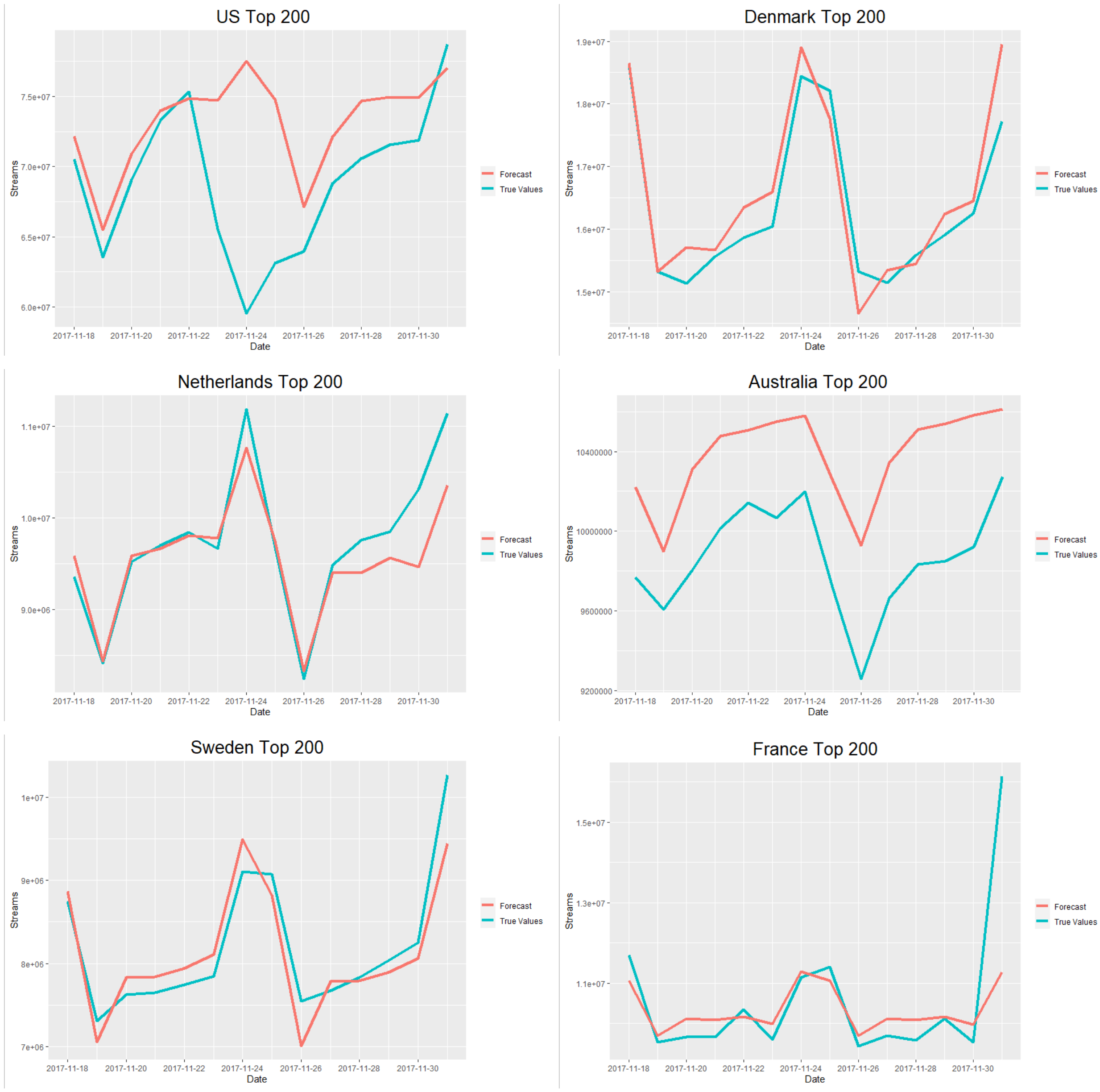


Figure 9. VAR Model 1 correlations visualized.