Executive Summary

This report investigates the factors influencing the Average Shot Length (ASL) in films, focusing on the relationships between genres, directors, countries, and time periods. ASL is a critical element in cinematography that influences the pacing, mood, and storytelling in a movie. We analyze a dataset containing information about 11,001 films, including variables such as the film's ASL, genre, director, country, and release year. Using Bayesian hierarchical modeling, we explore the influence of these factors on ASL. Our findings suggest significant effects of continent and time period on ASL. These results provide insights into how the film industry has evolved and how cinematographic techniques have varied across genres and time.

Introduction

The film industry has undergone significant changes over the past century, with cinematographic techniques, such as Average Shot Length (ASL), playing a crucial role in shaping the mood, pacing, and overall storytelling. ASL, which is the average duration of a shot in a film, has varied over time and across genres. Understanding what factors influence ASL can provide valuable insight into the evolving nature of film and its narrative techniques.

This report aims to explore the factors that affect ASL across a large dataset of 11,001 films. Specifically, the report will examine the role of genre, director, country, and time period in shaping ASL. By fitting a statistical model, we aim to answer the following key question: **How do genres, directors, countries, and time periods influence the ASL of films?**

Data

The dataset used in this analysis includes **11,001 films** with information on:

- **ASL**: Average Shot Length. G**enre**: Primary genre of the film, **Director**: Director's name, C**ountry**: Country where the film was produced, **Year**: Year of release, C**ontinent**: derived from the country, and **Genre (added)**: Scraped from TMDB API

Data Collection

The core dataset was obtained from:
https://users.stat.ufl.edu/~winner/datasets.html

Additional variables were added to enrich the analysis:

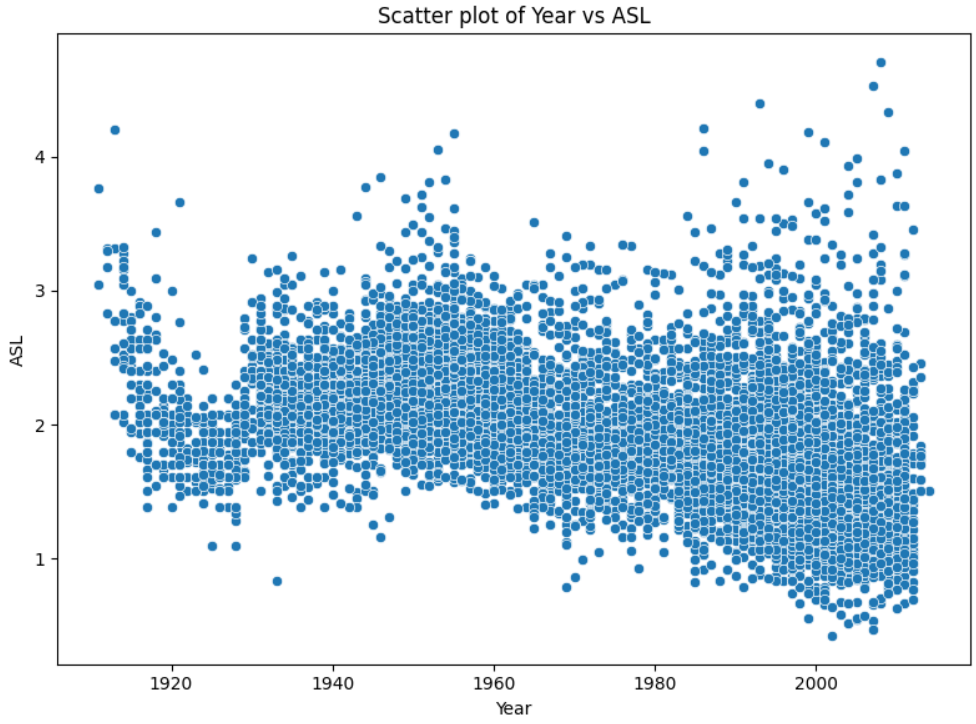
- **Continent**. Created manually based on the country of origin to group countries into broader geographic categories (e.g., Europe, Asia, North America).
- **Silent_era**. Created manually to see if this class has any kind of predictability about ASL.

Data Challenges

- Some records had missing or inconsistent entries for **Year**, **Director**, or **Country**.
- Countries were matched to their corresponding continent using external mapping, and
- Films missing ASL or other core identifiers were excluded from the modeling dataset.

Data Exploration

We performed a preliminary exploratory data analysis (EDA) using visualizations (e.g., histograms, scatter plots) to uncover any apparent patterns. We found that ASL varies across continents, and directors. The most notable relation is a time trend analysis indicated a general decrease in ASL over time, particularly from the 1940s onward.



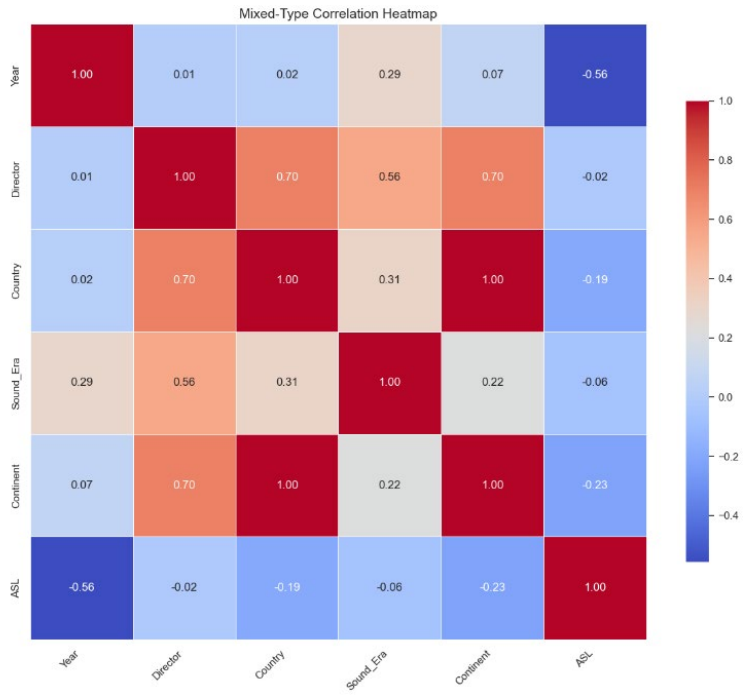
The linear, decreasing trend suggests the use of hierarchical linear models

Model

We model the **average shot length (ASL)** as a function of a **continent-specific intercept** and a **global slope for release year**. Each continent has its own baseline ASL, modeled with a random intercept. These continent intercepts are drawn from a common normal distribution centered at a global mean (**mu_0**) with variance governed by **prec_a**. The effect of year is captured by a shared slope (**b**), which is the same across all continents. The ASL for each film is normally distributed around its predicted value (continent intercept + year slope × standardized year), with residual variance governed by **prec**.

Justification for the Model

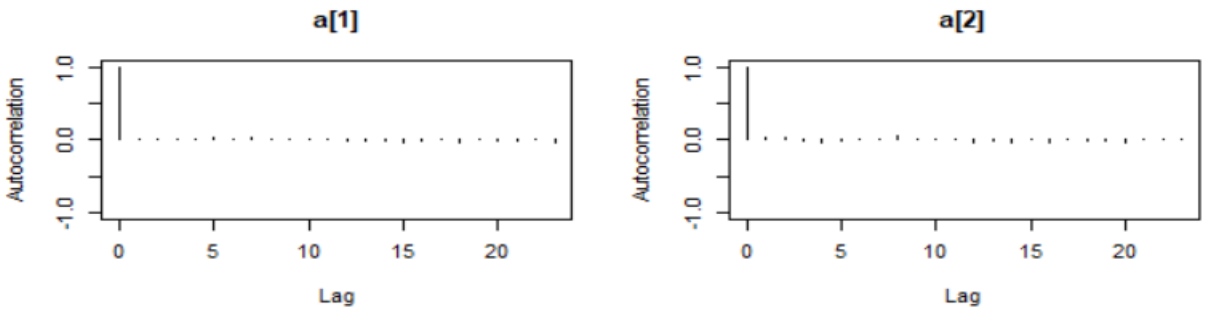
This model is appropriate because it allows us to account for multiple levels of variation (continent, country, year) and assumes that films within each level are correlated. The hierarchical structure enables the pooling of information across films, making the model more robust to variations in the data.



The most correlative features with the response ASL are Continent and Yea

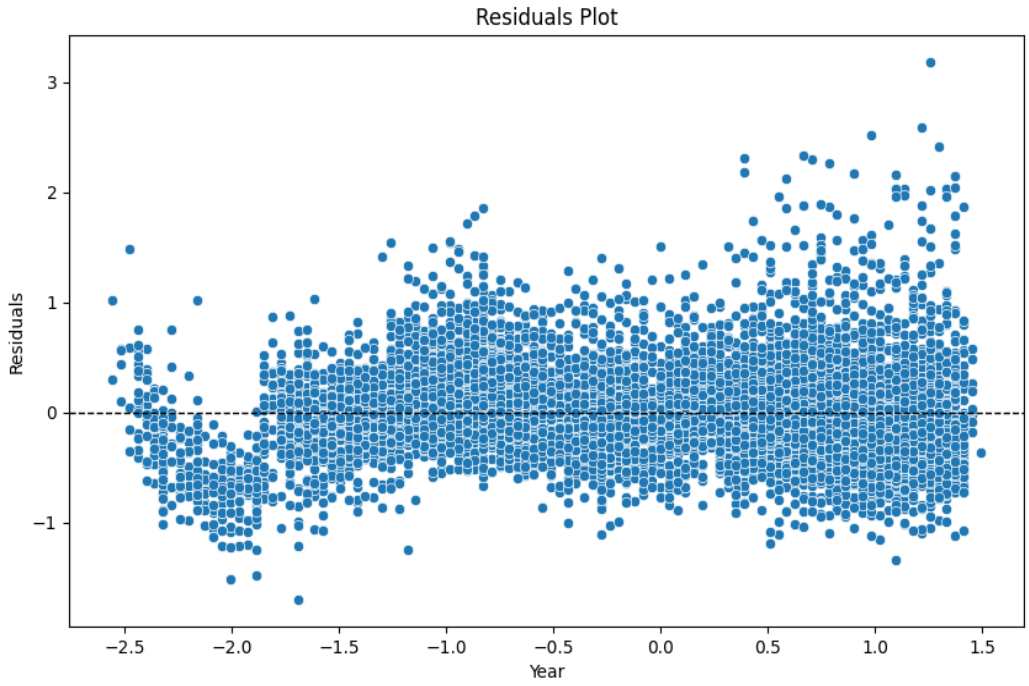
Results

After fitting the model using JAGS and R, we analyzed the posterior distributions for the key parameters. The convergence of the MCMC chains was checked using the Gelman-Rubin diagnostic, being all the point estimates of the parameters almost exactly 1, indicating good convergence.



Correlation plots for the MCMC chains of the first two parameters

The trace plots and autocorrelation plots also confirmed that the chains were mixing well.



Standardize years vs Residuals. The discrepance in the years before 1927 may be caused by the lack of data in those year (lost films), the rapid evolution of cinematographic techniques, the end of the silent era or the adaptation to sound.

We also fit a non-hierarchical model that did not account for continent-level variation. The Deviance Information Criterion (DIC) for this simpler model was higher than that of the hierarchical model, indicating a poorer fit. This suggests that incorporating continent-level random intercepts improves the model's ability to capture the structure in the data and better explains variation in average shot length.

Hierarchical model:

```
Penalized deviance: 10752
```

Non-hierarchical model:

```
Penalized deviance: 11424
```

Conclusions

While the model is robust and explains the data well, some limitations exist. For example, the data may not capture all possible confounding factors, such as budget or production constraints. Additionally, the model assumes that all directors, continents and countries have fixed effects, which may not fully capture the nuances of individual film characteristics.

Overall, this study contributes to our understanding of how ASL is influenced by various factors in the film industry and provides a foundation for further research into cinematographic trends.