



Análise de Dados IMDb: Tendências em Filmes e Séries

Bernardo Abib Juan Gabriel Luiz Felipe Nunes Marcel Hideki Pedro Constantino Yejee Kim



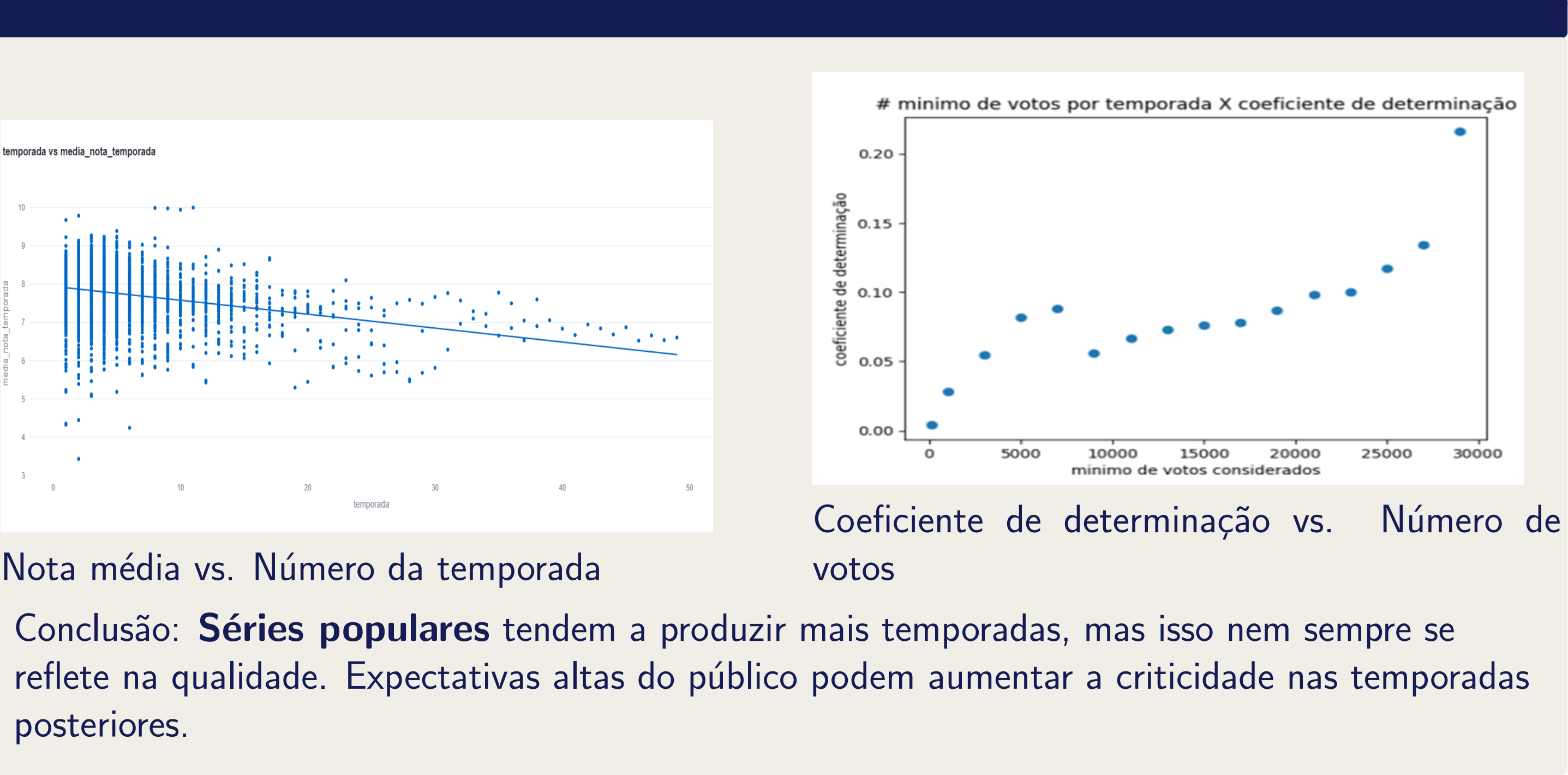
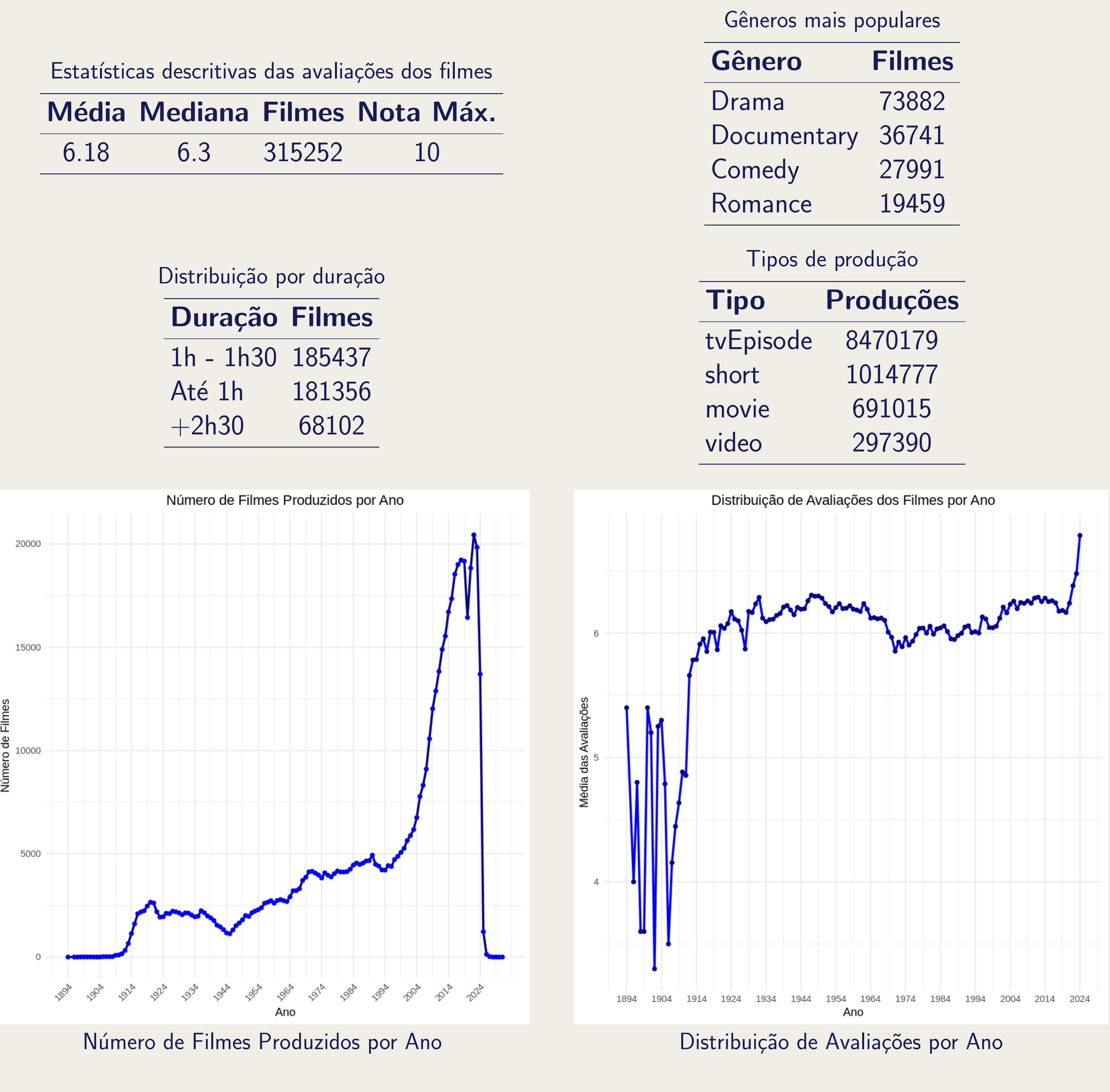
Introdução

Este estudo utiliza dados de produções audiovisuais do IMDb para responder a três seguintes perguntas:

- Existe relação entre o número de temporadas/episódios e a qualidade das séries?
- Quem são os profissionais mais populares em cada área e seus projetos mais conhecidos?
- Quais são as produções mais populares por década?

Os dados foram unificados em um banco de dados SQLite para simplificar a análise.

Análise Descritiva



Pergunta 2: Profissionais mais populares e seus projetos mais conhecidos

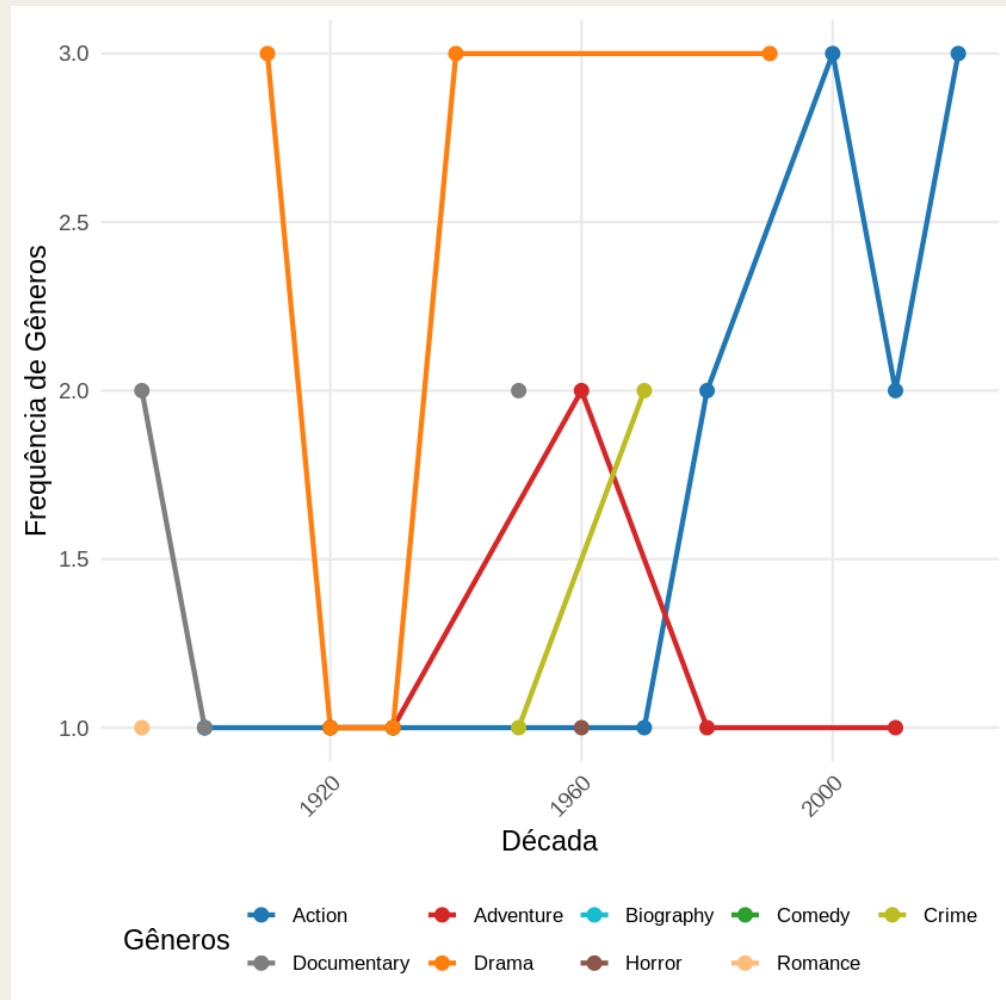
Nesta pergunta, identificamos os profissionais mais populares em diferentes áreas de atuação no cinema e televisão, usando a soma de votos dos projetos como critério de popularidade.

Profissionais mais populares por área e seus Projetos mais conhecidos			
Nome Profissional	Categoria	Projeto Mais Popular	Número de Votos
Samuel L. Jackson	ator	Pulp Fiction	2256119
Scarlett Johansson	atriz	The Avengers	1476184
George W. Bush	imagens de arquivo	Bowling for Columbine	149453
Stan Freberg	som de arquivo	Lady and the Tramp	150788
Sarah Finn	diretor de elenco	The Avengers	1476184
Roger Deakins	diretor de fotografia	The Shawshank Redemption	2937574
Hans Zimmer	compositor	The Dark Knight	2917884
Christopher Nolan	diretor	The Dark Knight	2917884
Lee Smith	editor	The Dark Knight	2917884
Kevin Feige	produtor	The Avengers	1476184
Nathan Crowley	designer de produção	The Dark Knight	2917884
Johnny Knoxville	ele-mesmo	Men in Black II	409403
Christopher Nolan	roteirista	The Dark Knight	2917884

Pergunta 3: Produções Mais Populares por Década

Nesta pergunta, identificamos as séries, filmes e episódios mais populares por década, com base na quantidade de votos recebidos.

Década	Top 3 Filmes	Top 3 Séries	Top 3 Episódios
1940	Casablanca It's a Wonderful Life Citizen Kane	The Lone Ranger	Enter the Lone Ranger The Lone Ranger Fights On The Lone Ranger's Triumph
1950	12 Angry Men I Love Lucy Vertigo	The Twilight Zone Alfred Hitchcock Presents	Time Enough at Last Where Is Everybody? Walking Distance
1960	The Good, the Bad and the Ugly 2001: A Space Odyssey Psycho	Star Trek Monty Python's Flying Circus Scooby Doo, Where Are You!	The City on the Edge of Forever Where No Man Has Gone Before The Cage
1970	The Godfather Star Wars: Episode IV The Godfather Part II	Columbo Fawlty Towers M*A*S*H	Murder by the Book Death Lends a Hand Any Old Port in a Storm
1980	Star Wars: Episode V Back to the Future Star Wars: Episode VI	Twin Peaks The Simpsons Dekalog, jeden	Northwest Passage Simpsons Roasting on an Open Fire Dekalog, jeden
1990	The Shawshank Redemption Fight Club Forrest Gump	Friends The Sopranos One Piece	The One Where Everybody Finds Out I'm Luffy! The Man Who Will Become the Pirate King! The Great Swordsman Appears! Pirate Hunter, Roronoa Zoro
2000	The Dark Knight The Lord of the Rings: The Fellowship of the Ring The Lord of the Rings: The Return of the King	Breaking Bad The Big Bang Theory Dexter	Pilot Crazy Handful of Nothin' Nakama no Itami wa Waga Itami: Zoro Kesshi no Tatakai
2010	Inception Interstellar The Dark Knight Rises	Game of Thrones Stranger Things The Walking Dead	Battle of the Bastards The Iron Throne The Long Night
2020	Spider-Man: No Way Home Dune The Batman	The Queen's Gambit Squid Game Loki	Long, Long Time (The Last of Us) When You're Lost in the Darkness (The Last of Us) Assault (Attack on Titan)



Evolução da Frequência de Gêneros por Década nos Top 3 Filmes

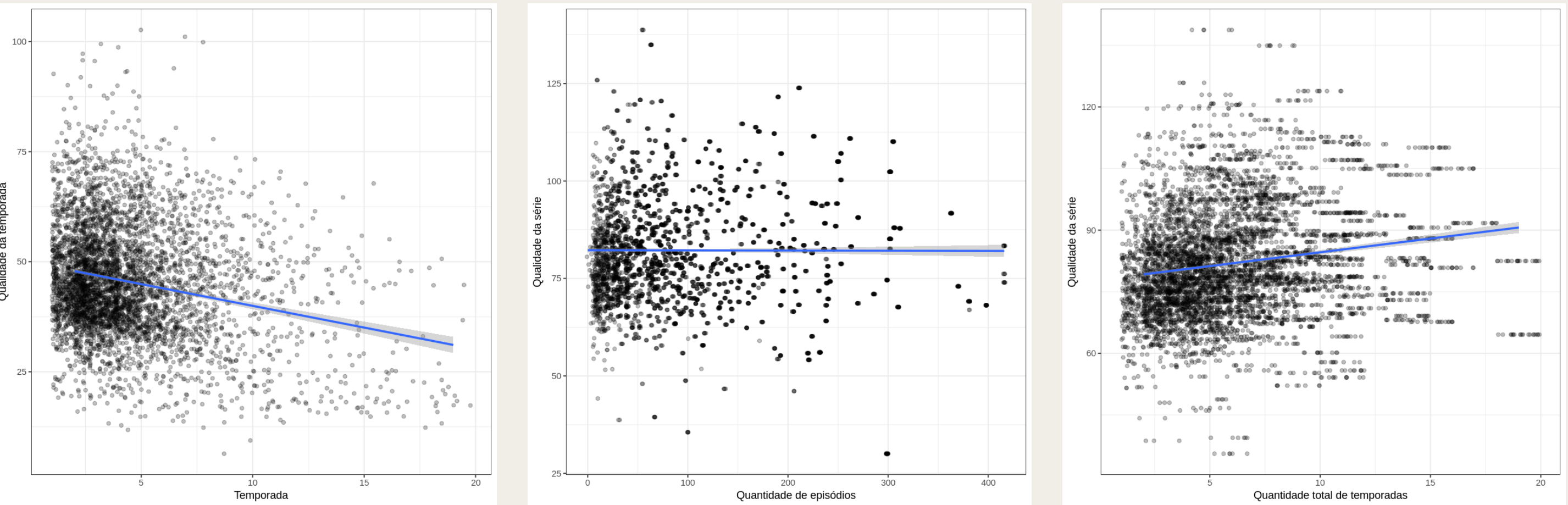
Pergunta 1: Relação entre temporadas/episódios e qualidade das séries

Nesta pergunta, investigamos a relação entre quantidade de temporadas/episódios e a qualidade das séries. A qualidade foi quantificada com o modelo:

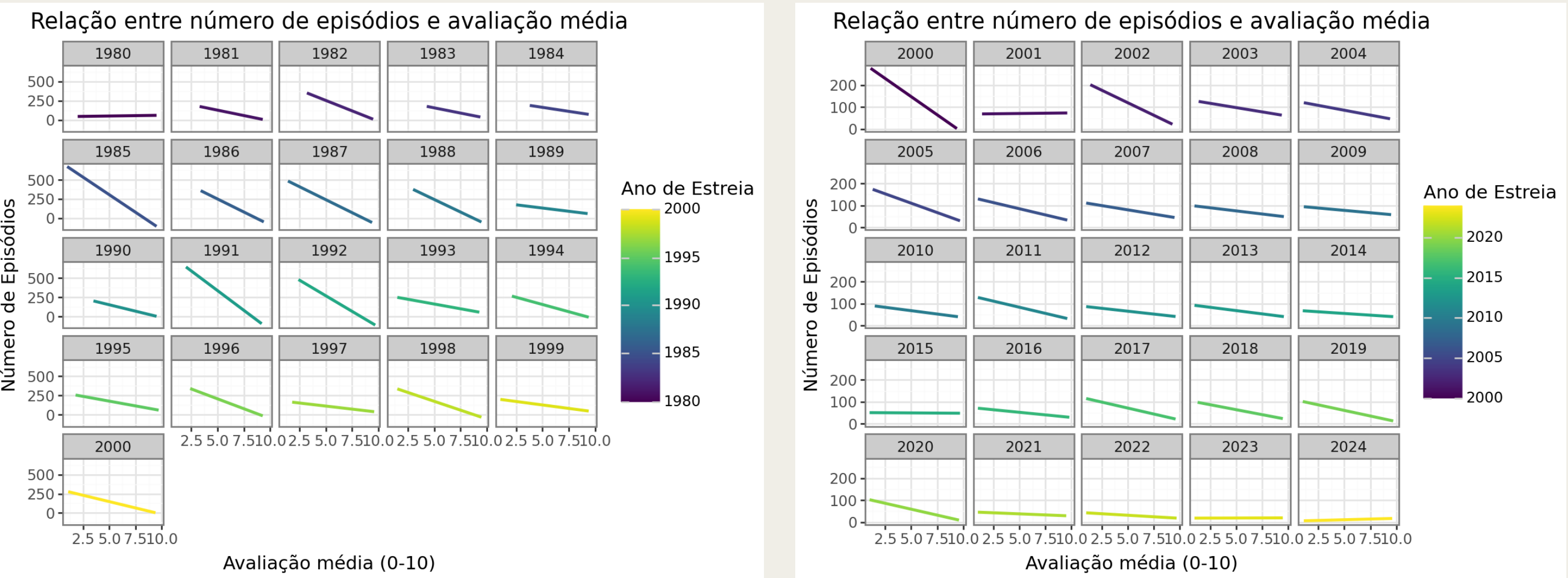
$$\text{Qualidade} = \text{Nota} * \ln(\text{Popularidade})$$

onde "Nota" é a média das avaliações dos usuários e "Popularidade" é o número de votos.

- Temporadas Avançadas:** Quanto maior o número da temporada, menor tende a ser a qualidade.
- Quantidade de Episódios:** Não há uma relação direta com a qualidade.
- Total de Temporadas:** Séries com mais temporadas tendem a ter maior qualidade geral.



Análise em Python mostrou que, entre 1980-2000, **séries com menos episódios tendem a ser mais bem avaliadas**, mas essa diferença diminui a partir de 2000.



Séries entre 1980 - 2000

Séries entre 2000 - 2024

- Declínio da Nota Média com Temporadas:** Regressão mostra uma queda na nota média conforme a série avança em temporadas (p-valor significativo).

Conclusão

Após as análises, as três perguntas por fim foram respondidas. Através da manipulação do banco de dados, e usando a quantidade de votos como critério de popularidade, identificamos as produções mais populares de cada década, e os profissionais mais populares de cada setor. Também mostramos, através de uma análise de regressão linear, que a quantidade de temporadas de uma série tem impacto negativo estatisticamente significativo na nota média das mesmas, indicando que há relação entre número de temporadas e a qualidade das séries.