PREDICTIVE POWER OF FILM ATTRIBUTES WITH RESPECT TO AUDIENCE DEMOGRAPHICS

by

Michael Chambers

A Major Research Paper

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Spatial Analysis (MSA)

Toronto, Ontario, Canada

© Michael Chambers 2013

Author's Declaration for Electronic Submission of a MRP

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions.

I authorize Ryerson University to lend this MRP by photocopying of by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

Abstract

Film attributes were analyzed to investigate if they could be used to predict gender/age demographic proportions of a film's audience. If relationships exist, film attributes could be used to help automate the process of allocating newly released films to theatre markets. Film attribute data were collected for wide release films seen between 2010 and 2012 by loyalty program card holders. Moviegoer demographic data were also gathered through the same program. The data were aggregated into three regions to investigate if there were any obvious spatial patterns regarding gender/age compositions in these motion picture markets. Multiple linear regression was used to determine if any relationships existed between film attribute and gender/age demographic variables for predictive purposes. Results showed that models using selected film variables had a moderate predictive strength for the parent and older cohorts. Although the predictive strengths were not very strong, the results adequately show that film attributes can be used to predict audience gender/age proportions.

Acknowledgements

I greatly appreciate and would like to thank my advisor Dr. Wayne Forsythe and Katherine Goff for their counsel and support.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ANOVA - Analysis of Variance

CCA - Canonical Correlation Analysis

MLR – Multiple Linear Regression

MPAA – Motion Picture Association of America

SWR - Step-Wise Regression

VIF - Variance Inflation Factor

UC – Unstandardized Coefficients

UK – United Kingdom

Chapter 1: Introduction

Due to the fact that people are intrinsically tied to their location, the study of geography and how people interact with it is extremely important. GIScience is, "the development and use of theories, methods, technology, and data for understanding geographic processes, relationships, and patterns" (Mark, 2003; Goodchild, 2010). GIScience and its tools have been adopted by a wide range of disciplines ranging from environmental planning, health studies, urban planning, and retail market analysis (Chen, 2006; Suarez-Vega et al., 2012). With respect to commercial and retail sectors, awareness and application of GIScience has grown immensely with it most commonly used for location and market analysis. For the motion picture industry, movie theatres can benefit greatly from utilizing geotechnology because audience demographic compositions and film interests vary geographically. As a result, geographic processes, relationships and patterns directly affect movie theatres and their operations. GIScience can therefore, be used to analyze these geographic phenomena to make informed decisions regarding the allocation of films to theatre markets.

The movie theatre business is part of the service industry as much as it is the entertainment industry because it does not create nor manufacture products for consumers. Instead of producing products, movie theatres are venues that provide an experience for consuming entertainment.

Over the last five years, box office revenue for the North American market has grown 12% to create a 10.8 billion dollar industry. Of the 12% increase, 6% occurred in 2012 alone. While average admission prices remained constant from the previous year, the 6%

growth in 2012 revenue (reaching 1.36 billion dollars) was due to an equivalent increase in theatre attendance (MPAA, 2012).

Similar to previous years, 225 million people or roughly 68% of North America's population saw at least one movie in 2012. Despite maintaining similar moviegoer populations to previous years, industry growth is fueled by moviegoers who view films at least once a month (MPAA, 2012). Monthly moviegoers represent 13% of the population while accounting for 57% of all movie ticket sales (MPAA, 2012). Therefore, monthly moviegoers are the primary source and generators of profit for movie theatres. As a result, it is extremely important to understand film demands and interests of monthly moviegoers in a theatre's market. This knowledge and understanding will allow newly released films to be appropriately matched to theatre markets that entice more monthly moviegoers and increase overall profits. There is a large potential for further applications of GIScience in the movie theatre business with one application being the study of relationships between film and audience attributes. Specifically, exploring whether film attributes can be used to predict gender/age compositions of a film's potential audiences.

In order to maximize audience draw, theatres must deliver films and other entertainment media that appeal to their market demographics. In other words, the profitability and economic success of the movie theatre business is impacted by the allocation and delivery of entertainment media to diverse markets (Li and Sun, 2011). Newly released films must be allocated to theatre markets in which they will have the greatest appeal, ensuring maximum attendance and revenue. Despite this crucial task, there currently is no scientifically based process for allocating newly released films. They are currently allocated by human intuition based on industry experience and knowledge of what has worked in the past. GIScience has an integral role in this process as film interests vary by market demographic composition and geographic location.

1.1 Research Objectives

The purpose of this research is to determine if film attribute variables can be used to forecast gender and age proportions of a newly released film's potential audience. Film attributes are characteristics of a film relating to genres, ratings and run-time. With relationships established between film attributes and audience gender and age demographics, the process of allocating newly released films to theatres can be more automated and scientifically based to maximize theatre profits.

1.2 Major Research Paper Structure

The structure of this research paper follows a manuscript format in which there are three main chapters. The first chapter is a complete introduction of the research topic and goals. Chapter Two is a comprehensive literature review discussing the current status of research in the field and how techniques and tools have been used and if and why they can be applied to this study. Chapter Three is the manuscript that consists of an abbreviated introduction and literature review along with the research methods, data, results, and conclusion. Chapter Four is the final chapter consisting of recommendations and future research pertaining to this study.

Chapter 2: Literature Review

2.1 Current Motion Picture Research

The breadth of research pertaining to the motion picture industry is very limited, providing a unique and interesting niche for academic research. Even scarcer is research on relationships between films and audience demographics. Currently, most existing research relevant to the motion picture industry is heavily focused on box-office success and investigating which film attributes equal maximum revenue (Holbrook and Redondo, 2010).

Hennig-Thurau et al. (2001) determined that film traits and communication are two controlling factors driving box-office revenue. The authors infer that like any business or product in the service industry, a consumer's preliminary assessment of products or services is crucial in deciding whether or not they purchase or invest. Moviegoers use traits coupled with communication factors to formulate preliminary evaluations of films that in turn dictates their incentive to visit a theatre to view a film (Hennig-Thurau et al., 2001; Elliott and Simmons 2008).

Film traits are attributes that are discernible both before and after viewing a film. For example, genre, rating, language, actors, budget, and movie length are attributes that audiences are able to observe before viewing a film in-theatre (Hennig-Thurau et al., 2001). Attributes that audiences ascertain after viewing a film are those that are associated with the experience of viewing the film. Examples include sound, visual effects, and film format (Hennig-Thurau et al., 2001).

Communication factors as defined by Hennig-Thurau et al. (2001) are “movie advertising as the main company-controlled information source and movie reviews, awards, and word-of-mouth information”. Through various channels, communication factors ensure that moviegoers have an adequate, well-rounded knowledge of film attributes to formulate their preliminary evaluations. Because of this, communication factors are equally important for driving film revenue as the attributes themselves (Hennig-Thurau et al., 2001). Related research by Elberse and Eliashberg (2003) found that advertising is particularly important for opening-week box-office revenue.

Without adequate movie attribute information, moviegoers are hindered in their ability to make informed preliminary evaluations and therefore, incentive to view and potential box-office revenue are reduced for a film (Hennig-Thurau et al., 2001). However, Hennig-Thurau et al. (2006) found that the importance of studio advertising decreases after the opening week because moviegoers utilize opening week performance as a cue towards the quality and attractiveness of a film.

Film critiques also have an important role in determining a film’s fate at the box-office (Gemser et al., 2006). It is because of this that existing research has explored the relationship between critiques and box-office success (Basuroy et al., 2006). A study completed by Elliot and Simmons (2008), focusing on markets in the United Kingdom (UK), investigated how film critiques impact box-office revenue and advertising. The authors found that critical reviews affected box-office revenue directly by 16.2 percent, meaning that a single point increase in critical reviews increases a movie’s box-office revenue by 16.2 percent (Elliot and Simmons, 2008). Evidently, moviegoers incorporate critical reviews in deciding whether or not to see a particular film (Elliot and Simmons, 2008).

When analyzing the relationship between critiques and advertising, Elberse and Eliashberg (2003) found that critical reviews had very little effect. They also uncovered that critical reviews positively impacted opening week box-office revenues but had a negative relation to the number of screens a film is allotted (Elberse and Eliashberg (2003). The authors suggest that the negative relation between critical reviews and the number of screens is the result of theatres showing poorly reviewed films on more screens to increase attendance through availability and not demand (Elberse and Eliashberg, 2003).

Further research on the effects of film critiques was undertaken by Basuroy et al., (2006) and investigated whether or not critiques were biased in their reviews. The authors found that statistical bias towards certain film studios existed, especially towards those based Los Angelis, California (Basuroy et al., 2006). A secondary component to Basuroy et al.'s (2006) research was to determine if film audiences could recognize and distinguish biased and unbiased critiques. Their research indicated that moviegoers did not value one critique over another, implying that audiences do not distinguish biased critiques from other non-biased ones (Basuroy et al., 2006). As a result, biased film critiques appear to have a greater effect on box-office revenue through their ability to influence audience perceptions about a particular film (Basuroy et al., 2006). Despite their ability to sway moviegoers' preliminary perceptions, Gemser et al. (2008) state that moviegoers are more influenced by other means, such as word of mouth, than critical reviews.

Another film attribute that has been analyzed in relation to box-office success is sequel status. A study by Dhar et al. (2011) temporally compared attendance of sequel films to their parent and other non-sequel movies. They focused on the United States (US) film market and chose to measure attendance because they believe that attendance ultimately leads to box-office success. By analyzing week, total, and retention attendance between the first and last week, Dhar et al. (2011) found that both parent and sequel films had higher attendance levels than non-sequel films. This indicates that the parent-film of all sequels is an above average performer at the box-office (Dhar et al., 2001). However, Basuroy and Chatterjee (2008) determined that sequels have a greater decrease in attendance between the first and second week when compared to parent and non-sequel movies. Furthermore, Dhar et al. (2001) also discovered that despite the fact that parent-films always have more attendance, sequels produce greater attendance and therefore revenue in the first week. Regardless of having a better first week performance, Basuroy and Chatterjee (2008) revealed that sequels do not generate as much revenue as their parent-films in the grand scheme. The authors also state that sequels increase their potential revenue the sooner that they are released after the parent-film (Basuroy and Chatterjee, 2008).

Regardless of the improved performance of sequels, Dhar et al. (2001) found that production of sequels has not increased over the twenty-six year period between 1983 and 2008 that their film data spanned. In fact, despite an average growth in film production of 6.68 percent, the number of sequels produced remained constant (Dhar et al., 2001).

The shortcoming of existing research is that it completely ignores the relationship between a film and its audience (Holbrook and Redondo, 2010). Different demographics will have variations in regards to the movies and attributes they prefer. Because of this, a film shown in one demographic market may not perform as well in another. Current research fails to adequately recognize that box-office revenue is a product of audience interest and movie theatre attendance (Elberse and Eliashberg, 2003; Holbrook and Redondo, 2010).

2.2 Audience Demographics

One study that does recognize the importance of a film's audience for box-office revenue is by Holbrook and Redondo (2010), in which they attempted to model the influence of film attributes on the composition of audience demographics. Their study focused on theatre markets in Spain and analyzed the connections between film and demographic group variables. The two main objectives for Holbrook's and Redondo's (2010) study were to identify all interconnections between their chosen film and demographic variables and to uncover audience segments and their associated film preferences. Film variables that the authors analyzed were, "country of origin, genre (drama, comedy, action adventure, thriller, romance, animation, and family), presence of sex/violence, stars' artistic reputation/physical attractiveness, advertising and critical review" (Holbrook and Redondo, 2010). The data were gathered from a variety of sources including internet-based movie databases and websites, film magazines as well as market research companies for 110 films that ran between 1998 and 2008 (Holbrook and Redondo, 2010).

The demographic variables Holbrook and Redondo (2010) chose for their analysis were age cohort, gender, presence of children, education, social class, and size of municipality. Audience demographic data were obtained through surveys conducted by a Spanish media-audience market research firm (Holbrook and Redondo, 2010). To analyze the two distinct sets of variables, they used Canonical Correlation Analysis (CCA) to determine if any linkages existed between their chosen demographic and film attribute variables. Output from their CCA resulted in the formation of ten canonical dimensions with four being statistically significant. The four significant dimensions were able to adequately summarize the linkages between Holbrook's and Redondo's (2010) audience demographic and film attribute variables by representing 44.6% of the 47.4% total variance in audience demographics accounted for by all ten canonical dimensions (Holbrook and Redondo, 2010). Their first dimension represented an audience demographic of people aged 35-44, with children, possessing less than a secondary education level, and living in a non-urban area. Their CAA indicated that this first dimension had a strong linkage to the family movie profile consisting of a variety of film attributes that share a commonality of avoiding sexual and violent content (Holbrook and Redondo, 2010). The second canonical dimension represented men less than 20 years of age and below middle class. This dimension was linked to films that were extensively promoted and more action oriented. Their third dimension represented people living in urban areas and of higher education and social class. This dimension was found to be more influence by film critiques and linked to films that had better reviews. The fourth dimension from Holbrook's and Redondo's (2010) analysis represented people aged 14-24 and was found to be linked to comedic, low violence films.

Contrary to previous research on the effects of critics, Holbrook and Redondo (2010) found that not all moviegoers are influenced by critical reviews equally. Previous research did not account for heterogeneous audiences and assumed all moviegoers were equally universally affected (Holbrook and Redondo, 2010). Their study demonstrates the importance of incorporating audience demographics in film analysis and forecasting.

2.3 Canonical Correlation Analysis

There are a few statistical methods that are possible candidates for this analysis. The first is (CCA), which is a method for identifying relationships between two different sets of variables (independent/predictor and dependent/criterion) (Mach and Grewal, 2008; Lange et al., 2010). CCA is similar to multiple linear regression in that it analyzes a set of independent and dependent variables however, CAA analyses the relationship between the two sets of variables as a whole and not individually as with multiple linear regression (Mach and Grewal, 2008). Furthermore, relationships that appear weak between independent and dependent variables when using multiple linear regression can actually be strengthened using CCA because it incorporates the whole set of variables (Mach and Grewal, 2008).

CCA also reduces and simplifies statistical procedures by eliminating the need for multiple runs of regression (Mach and Grewal, 2008; Lange et al., 2010). For example, Lange et al. (2010) used canonical correlation analysis to identify relationships between genetic markers and expression probes. CCA was extremely helpful for this particular application because there are millions of genetic markers and thousands of single expression probes that could be related (Lange et al., 2010). Trying to analyze relationships for each genetic marker individually using regression would be impractical.

2.4 Multiple Linear Regression

A second statistical method that was considered for the analysis is Multiple Linear Regression (MLR). MLR is a statistical model used to identify relationships between a set of independent and dependent variables and is typically used for either prediction or explanatory purposes (Knofczynski and Mundfrom, 2008; Mach and Grewal, 2008; Mahmoud, 2008).

According to Mach and Grewal (2008), research has determined that regression and CCA produce the same results as long as the individual multiple regression criteria are independent from one another. They analyzed if frequency of transit use was affected by transit performance and the availability of transit system information. The authors found that as the availability of a car decreased, the use of transit increased for various activities (Mach and Grewal, 2008). The criterion variables in this case are all interrelated because the unavailability of a car is a common factor forcing respondents to use public transit (Mach and Grewal, 2008). In their study analyzing number of screens to revenue, Elberse and Eliashberg (2003) encountered the same issue using linear regression with ordinary least squares because number of screens and revenue are bidirectional by having a reciprocal effect on each other (IBM SPSS Statistics, 2001). The number of screens a film has affects revenue and potential revenue for a film affects the number of screens it is allocated. Because of this, they used a three stage least squares procedure. Therefore, criterion variables must not be bidirectional and independent from one another in order for MLR to be an effective statistical method for predictive purposes.

Because of the fact that the majority of independent variables are categorical in this study, a few preliminary procedures must be taken before running an MRL (Stockburger, 1998). The categorical variables must first be converted into dichotomous or "dummy" variables in order for the results of the MLR to be interpretable (Stockburger, 1998). Since all the independent variables have only two categories, their dummy coding involves assigning 0 to instances of one category and 1 to instances of the other. The coded independent dummy variables can then be entered directly into the MLR model and their resulting positive or negative regression weights either added to or subtracted from the predicted "Y" value (Stockburger, 1998). With the help of dummy variable coding, categorical variables can be used like any other variable for predictive analysis (Stockburger, 1998).

2.5 Step-Wise Regression

A third statistical method considered was Step-Wise Regression (SWR). The SWR model is similar to MLR but uses a series of "steps" to determine which selection of independent variables maximize the correlation coefficient with each dependent variable (Afifi and Bendel, 1976; Johnsson, 1992). The procedure first determines the correlation coefficient between the set of independent variables and one dependent. SWR then systematically adds and drops individual independent variables and recalculates the correlation coefficient in steps until the selection of independent variables that best explain or predict the dependent variable is achieved (Breaux, 1968; Afifi and Bendel, 1976; Agostinelli, 2002).

For their study, Holbrook and Redondo (2010) chose to use CCA to analyze the relationships between film attribute and demographic variables. However, Holbrook and Redondo (2010) even state that despite the superiority of CCA in analyzing multidimensional phenomena, very few studies have successfully used this technique due to difficulties in interpreting its results. Furthermore, correlation does not necessarily mean causation and therefore, cannot be used for predictive purposes (Gardner, 2000; Royne, 2008).

Considering this study, CCA and SWR would be appropriate statistical models to use if the goal was to determine which combination of film attributes were most attractive to each gender/age cohort demographic. This application could prove more beneficial for film producers trying to determine what type of movie to produce for a specific demographic market.

However, the purpose of this research is not to determine the best composition of film attributes for each demographic cohort; it is to improve a movie theatre's ability to allocate newly released films. Movie theatres have very little control over the types of films that are produced and as a result, they do not have the ability to decide which attributes a newly released film has. Movie theatres have to invest and allocate new releases with predetermined attributes to their various markets. Therefore, CCA and SWR would not be the best statistical procedures to use for the purpose of this research.

With the research purpose in mind and the exploration of the above statistical methods being considered, MLR is the best model to apply for this research. MLR will identify predictive relationships between film and demographic variables that can then be used to predict demographic compositions of potential audiences.

Chapter 3: Manuscript

PREDICTIVE POWER OF FILM ATTRIBUTES WITH RESPECT TO AUDIENCE DEMOGRAPHICS

3.1 Abstract

Film attributes were analyzed to investigate if they could be used to predict gender/age demographic proportions of a film's audience. If relationships exist, film attributes could be used to help automate the process of allocating newly released films to theatre markets. Film attribute data were collected for wide release films seen between 2010 and 2012 by loyalty program card holders. Moviegoer demographic data were also gathered through the same program. The data were aggregated into three regions to investigate if there were any obvious spatial patterns regarding gender/age compositions in these motion picture markets. Multiple linear regression was used to determine if any relationships existed between film attribute and gender/age demographic variables for predictive purposes. Results showed that models using selected film variables had a moderate predictive strength for the parent and older cohorts. Although the predictive strengths were not very strong, the results adequately show that film attributes can be used to predict audience gender/age proportions.

3.2 Introduction

Over the last five years, box office revenue for the North American market has grown 12% to create a 10.8 billion dollar industry. Of the 12% increase, 6% occurred in 2012 alone. The large growth in 2012 was due to an equivalent increase in theatre attendance, where revenue reached 1.36 billion dollars, as the average admission price remained constant from the previous year (MPAA, 2012).

Similar to previous years, 225 million people or roughly 68% of North America's population saw at least one movie in 2012. Despite maintaining similar moviegoer populations to previous years, industry growth is fueled by moviegoers who view films at least once a month (MPAA, 2012). Monthly moviegoers represent 13% of the population while accounting for 57% of all movie ticket sales (MPAA, 2012). Therefore, monthly moviegoers are the primary source and generators of profit for movie theatres. As a result, it is extremely important to understand film demands and interests of monthly moviegoers in a theatre's market. This knowledge and understanding will allow newly released films to be appropriately matched to theatre markets to increases volume and frequency of monthly moviegoers and ultimately profits.

The breadth of research pertaining to the motion picture industry is very limited, providing a unique and interesting niche for academic research. Even scarcer is research on relationships between films and audience demographics. Currently, most research regarding the motion picture industry is heavily focused on box-office success and investigating which film attributes generate maximum revenue. Film attributes are characteristics relating to genres, rating, and run-time. The shortcoming of existing research is that it completely ignores the relationship between a film and its audience (Holbrook and Redondo, 2010). Different demographics will have variations in regards to the movies and attributes they prefer. Because of this, a film shown in one demographic market may not perform as well in another. Current research fails to adequately recognize that box-office revenue is a product of audience interest and movie theatre attendance (Elberse and Eliashberg, 2003: Holbrook and Redondo, 2010).

Study Area

For this research, Canadian movie theatre markets were analyzed at the provincial level to investigate whether spatial patterns in audience demographics were easily discernible at a coarse scale. Due to the distribution of Canada's population, the Provinces of British Columbia, Alberta, Saskatchewan, and Manitoba were grouped to form a Western Canada study region. Having the largest provincial population, Ontario alone created a second region while Quebec served a third study region (Figure: 3.2.1).

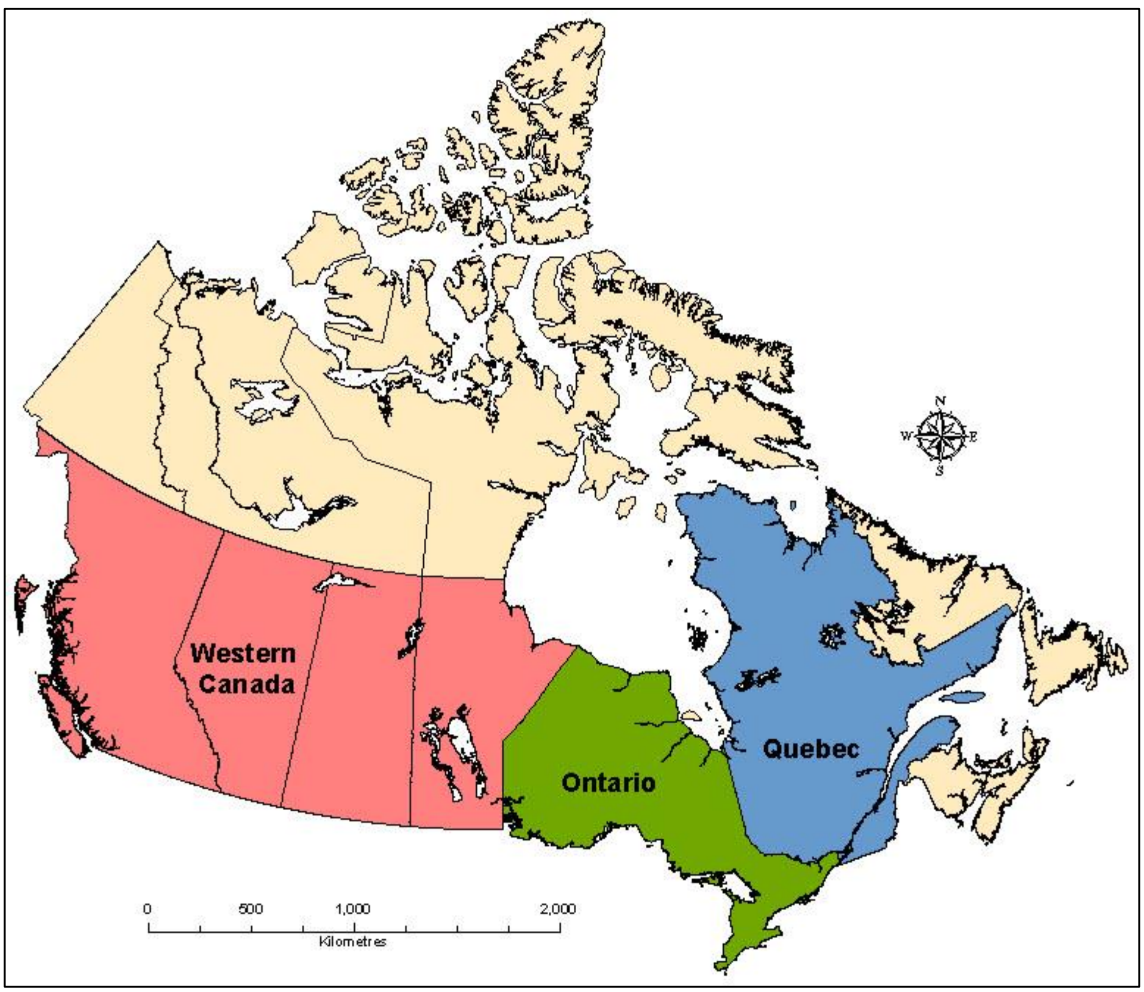

Figure 3.2.1 Three Canadian movie theatre regions being analyzed

Movie theatre markets in the United Kingdom (UK), Spain, and the United States have already been analyzed with respect to film performance research. The analysis of Canadian film markets is a needed addition to current research because Canada's movie-going population is fairly small in comparison to the three aforementioned countries. This research will diversify and expand knowledge regarding motion pictures.

3.3 Data Collection

A Canadian motion picture exhibitor provided data for this study. Moviegoers' demographic data and viewing habits were collected through the exhibitor's loyalty card program. The planned purpose of the card is for it to be used with every transaction a moviegoer makes, creating a database of transaction history and viewing habits of every cardholder. This provides a sound and scientifically collected data source for analysis. Moviegoers voluntarily use the card, which inevitably means that not every transaction at a theatre is recorded due to human forgetfulness and other circumstances. However, the reward incentives that the program offers are sufficient to maintain a fairly high and consistent usage rate of the cards. Regardless of the possibility that not every theatre patron has utilized the program, it still retains a fairly large proportion of moviegoers. As a result, the loyalty card users can be representative of the demographics and consumer habits for each study region. Atlantic Canada was not included in the study because the exhibitor did not include these provinces in the dataset that they provided.

Transaction history has been collected for many years but to capture the most recent movie viewing habits, this study only uses data from 2010 to 2012. Furthermore, the data only used English language films and were aggregated into three areas to analyze if there are any spatial differences between markets. Quebec for instance, has many French language films that are not shown in the other study regions. This would skew the study's results because it would inflate film results for Quebec and not the remaining study regions. Demographic data used for this analysis include gender and age cohorts (F14-18, F19-24, F25-34, F35-49, F50-64, F65+, M14-18, M19-24, M25-34, M35-49, M50-64, M65+, and parent). The *Parent* cohort is composed of loyalty card members who purchased an adult and child ticket together during the same visit. These age breaks were used because they were already established and used by the theatre company providing the data for analysis.

The exhibitor also provided the film attribute data used in this study. They collected the data from film providers and other industry partners. Film attributes that are used for this study include genre (action, adaptation, adventure, drama, comedy, romantic comedy, animated, horror, holiday, romance, sequel, thriller, sci-fi, foreign, fantasy, war, period, family and documentary), Motion Picture Association of America (MPAA) rating (G, PG, PG13, R and None), and film running time. The film genres and MPAA ratings are self-explanatory however, the *None* rating refers to a movie that has no rating.

3.4 Methods

Multiple Linear Regression (MLR) was used to investigate the possible presence of predictive relationships between film attributes and demographic gender/age cohort variables. MLR allows for the inclusion of all relevant elements in one model and is used to identify predictive or explanatory relationships between a dependent variable and a set of independent variables (Marill, 2004; Knofczynski and Mundfrom, 2008; Mach and

Grewal, 2008; Mahmoud, 2008). MLR was the chosen statistical analysis method because it allows for all selected independent variables to be entered into the predictive model. This is an important aspect for this analysis because the interest of the study is to investigate the predictive power of each film variable in relation to the dependents. Other predictive statistical methods like stepwise regression, would only determine the strongest combination of independents without accounting for the possible effects of the remaining independents. The model used film attributes for independent variables while gender/age cohorts were the dependents. By region, the model was run for each dependent gender/age cohort and produced three sets of outputs for each cohort. The results were then analyzed and compared both within and between study regions.

Relationships are determined by measuring the amount of variance each independent/predictor variable accounts for in a single dependent variable (Nathans, Nimon and Oswald, 2012). MLR is a valuable statistical method for investigating relationships between groups of variables as long as the variables are independent from one another (Mach and Grewal, (2008).

Aside from the *Film Runtime* variable, the remaining independent variables are categorical and are required to be converted into dichotomous variables in order for them to be used in a regression model. Also known as “dummy variables,” dichotomous variables are a type of categorical variable with only two possible categories (Salkind, 2010). Dummy variables are created by assigning 0 to instances of one category and 1 to instances of the other. This allows for numeric representation of the two categories that can then be used for meaningful statistical analysis (Stockburger, 1998).

After creating dummy variables for the independents, descriptive statistics were run on each variable to determine if they were normally distributed. One of the main assumptions with statistical hypothesis testing is that the sample data follow a normal distribution so it is extremely important to normalize the data prior to statistical testing (Manikandan, 2010). The skewness and kurtosis values along with histograms were used to assess each variable's distribution. Ideally, normal distributions have skewness and kurtosis values of 0 and therefore, variables with values greater than 0 are not normally distributed (Fink, 2009). Skewed data were then transformed in an effort to normalize the data as much as possible.

There are many different ways to transform data and deciding which one to apply typically involves testing each one (Osborne, 2002). The most commonly used and successful transformation methods are logarithms, square roots, and reciprocal (Manikandan, 2010). Before performing any transformation, a constant had to be applied to certain variables to raise their minimum value to 1. This had to be done because values below 1 behave differently than higher values mathematically when logarithms or square roots are applied (Osborne, 2002). All three transformations were tested but none could normalize the data fully because the raw data were so severely skewed. The square root transformation reduced skewness most effectively and therefore was utilized in this analysis. Despite the transformations, the data were deemed to have permanent skewness. Tables 3.4.1-3.4.3 show the skewness and kurtosis values for all data. Histograms (Figures 3.4.1-3.4.6) for before and after square root transformations are shown using the Females 14-18 data (as an example).

Table 3.4.1 Skewness and kurtosis values for the Western Canada study region

Dependent Variable	Skewness		Kurtosis	
	Raw Value	Square Root Value	Raw Value	Square Root Value
Females 14-18	2.615	.932	8.456	1.133
Females 19-24	2.510	.831	7.726	.958
Females 25-34	3.023	1.034	11.894	1.831
Females 35-49	2.843	.834	10.962	1.613
Females 50-64	2.240	.676	7.465	.544
Females 65+	2.995	1.023	13.471	1.547
Males 14-18	2.770	.983	10.079	1.148
Males 19-24	2.989	.979	12.779	1.286
Males 25-34	3.417	1.167	16.845	2.052
Males 35-49	3.027	1.054	12.579	1.745
Males 50-64	2.732	.877	11.662	1.153
Males 65+	2.860	.892	14.388	1.238
Parents	2.707	1.437	7.720	1.648

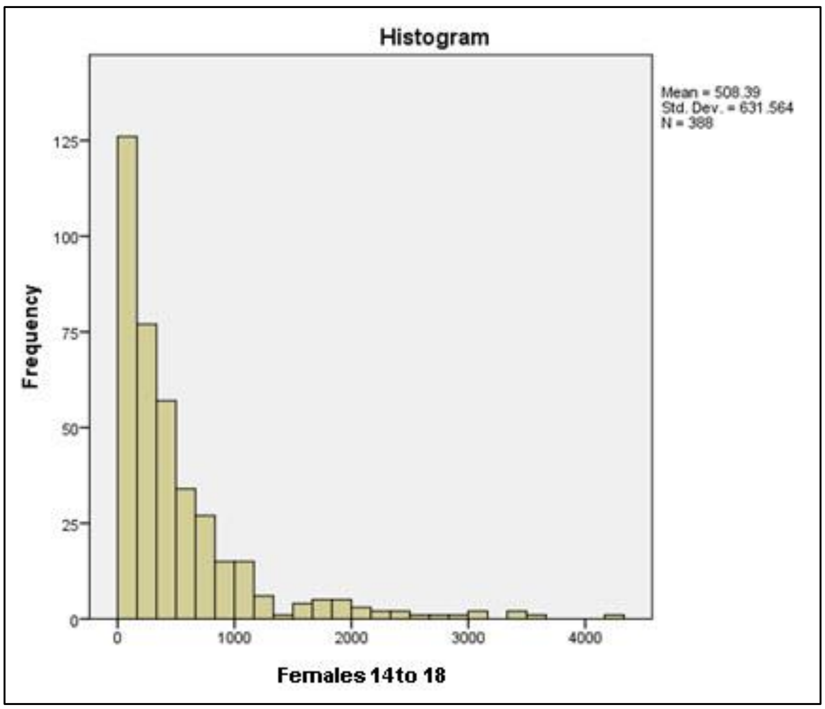

Figure 3.4.1 Western Canada data pre square-root transformation

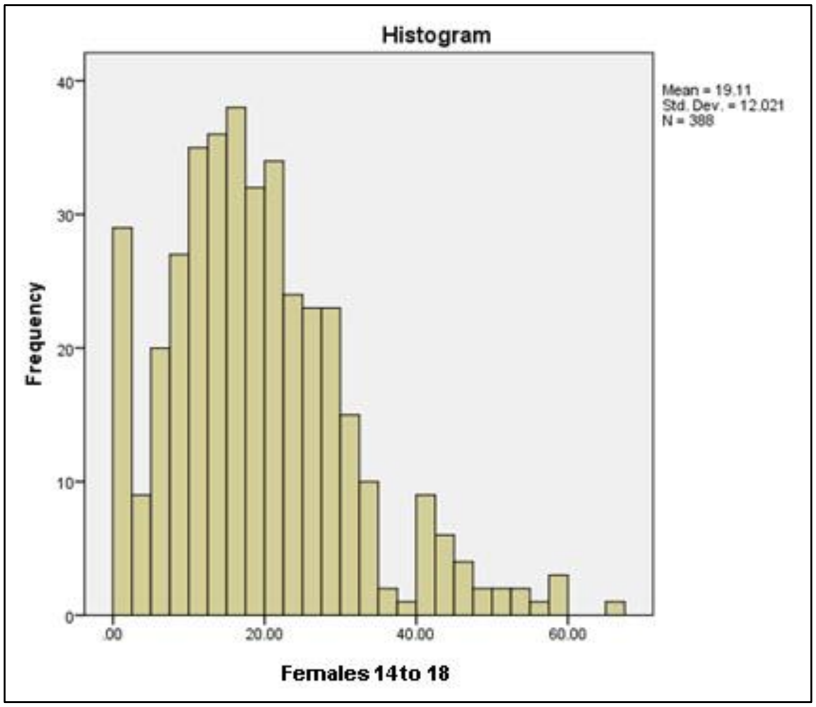

Figure 3.4.2 Western Canada data post square-root transformation

Table 3.4.2 Skewness and kurtosis values for Ontario

Dependent Variable	Skewness		Kurtosis	
	Raw Value	Square Root Value	Raw Value	Square Root Value
Females 14-18	2.585	.893	8.326	.927
Females 19-24	2.487	.800	7.741	.814
Females 25-34	3.001	.994	12.209	1.697
Females 35-49	2.749	.790	10.627	1.374
Females 50-64	2.145	.727	6.494	.427
Females 65+	3.213	1.170	16.023	1.834
Males 14-18	2.746	.956	10.095	.952
Males 19-24	3.129	1.013	14.351	1.371
Males 25-34	3.494	1.154	18.378	2.010
Males 35-49	2.997	1.022	12.715	1.665
Males 50-64	2.591	.864	10.287	.960
Males 65+	2.754	.983	12.472	1.128
Parents	2.516	1.297	6.577	1.156

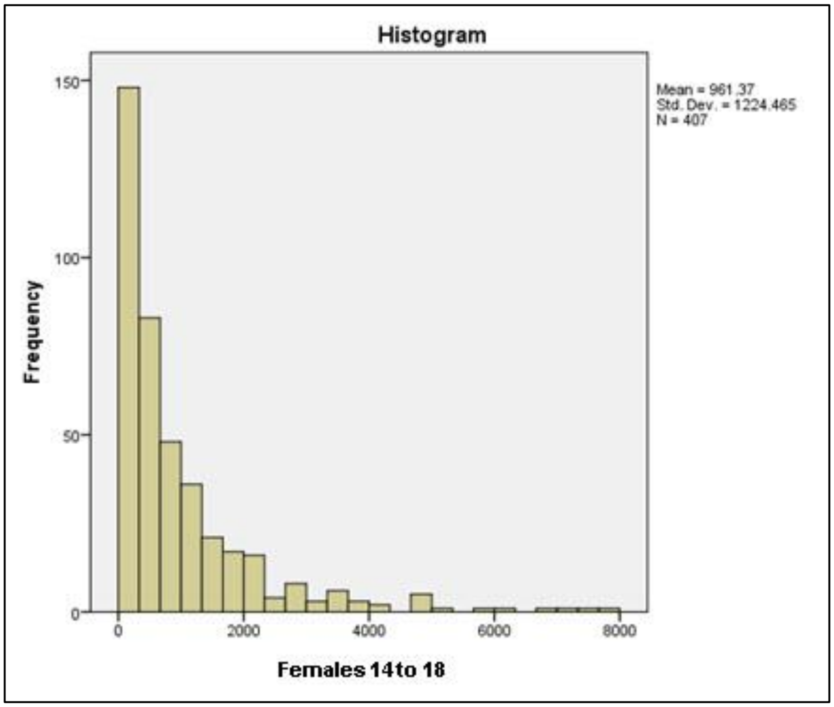

Figure 3.4.3 Ontario data pre square-root transformation

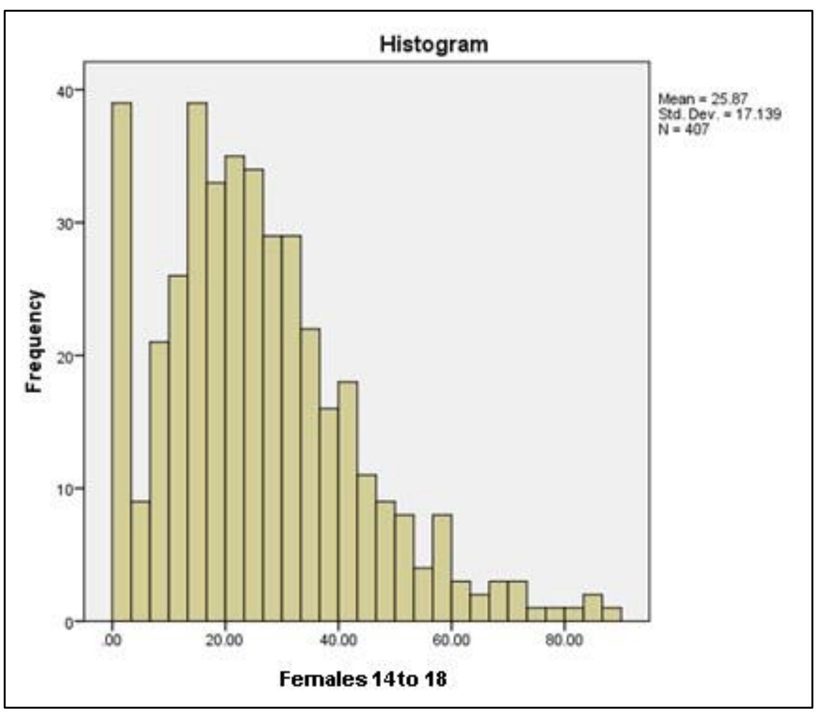

Figure 3.4.4 Ontario data post square-root transformation

Table 3.4.3 Skewness and kurtosis values for Quebec

Dependent Variable	Skewness		Kurtosis	
	Raw Value	Square Root Value	Raw Value	Square Root Value
Females 14-18	3.404	1.389	15.329	2.807
Females 19-24	3.275	1.418	12.980	2.858
Females 25-34	3.541	1.526	15.721	3.414
Females 35-49	3.638	1.457	17.475	3.676
Females 50-64	3.267	1.148	16.776	2.133
Females 65+	4.770	1.750	31.301	5.098
Males 14-18	3.061	1.330	11.919	2.263
Males 19-24	3.602	1.443	17.641	2.947
Males 25-34	3.841	1.579	20.139	3.441
Males 35-49	3.641	1.554	17.657	3.451
Males 50-64	3.279	1.191	16.680	2.232
Males 65+	3.389	1.196	18.274	2.198
Parents	2.861	1.439	10.074	1.863

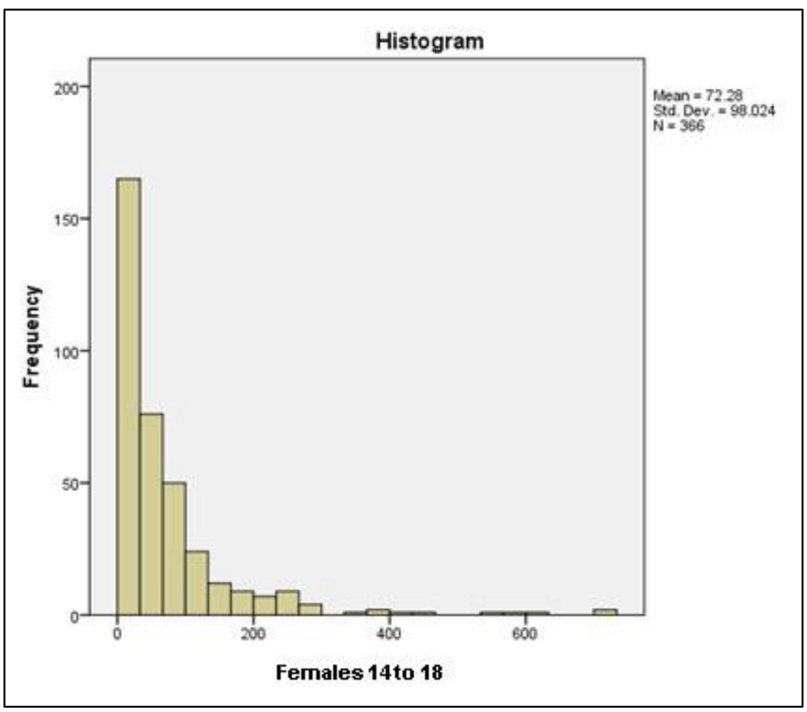

Figure 3.4.5 Quebec data pre square-root transformation

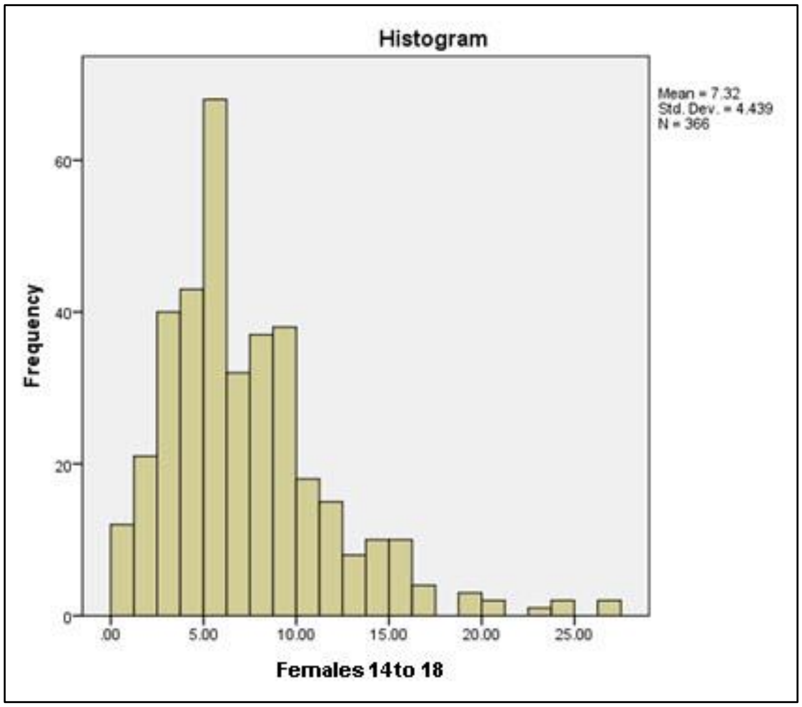

Figure 3.4.6 Quebec data post square-root transformation

After running MLR on the normalized data, there are a number of important statistics needed for interpretation in order to derive meaningful inferences from the model. First, the R^2 value represents the overall regression effect or the proportion of the dependent's variance that is explained by the independent variables (Nathans et al., 2012; UCLA Statistical Consulting Group, 2013). By segmenting the R^2 value by each independent variable, predictive power for each independent can be interpreted. The adjusted R^2 is the same as the R^2 but better for model interpretation because it "penalizes the addition of extraneous predictors to the model" (UCLA Statistical Consulting Group, 2013) and accounts for the degrees of freedom in the data.

The Durbin-Watson statistic is widely accepted as one of the best indicators for autocorrelation in a regression model (Bartels and Goodhew, 1981; Mukhtar, 1987). The statistic has a range of 0-4 with the general level of acceptability being a value of 2.0 (Savin and White, 1977; UCLA Statistical Consulting Group, 2013).

The Sig value from the Analysis of Variance (ANOVA) table is important to assess because it indicates if the model is statistically significant or not. Sig values must be below 0.05 to conclude that the null hypothesis stating that no relationship exists between the dependent and independent variables can be rejected (UCLA Statistical Consulting Group, 2013).

The beta values are the standardized coefficients or regression weights and they can be compared to determine which independent has the greatest effect on the dependent. They indicate the expected change for the dependent as a single independent variable is increased (Nathans et al., 2012; UCLA Statistical Consulting Group, 2013). However, beta weights cannot be used to compare strength of the independents because categorical variables do not have units of measurement and therefore, cannot be standardized. As a result, unstandardized coefficients can only be used to adequately asses the predictive impact that each predictor has on the dependents and not between each independent (UCLA Statistical Consulting Group, 2013).

Tolerance and variance inflation factor (VIF) values from the coefficients output table are the commonly used measures of multi-collinearity in a regression model. They represent the amount of variance that each independent variable shares with the other independents (O'Brien, 2007). A tolerance value less than 0.2 or a VIF of 5 or higher are the standard acceptability values used to indicate collinearity between independent variables (O'Brien, 2007).

A Casewise Diagnostics table presents cases that were unacceptable by having residuals greater than 3 standard deviations. It is an important table to assess because it identifies outliers or abnormalities in the dataset that should be investigated further.

3.5 Results

Spatial comparisons between the three study regions can be made by comparing the adjusted R^2 values for each dependent variable listed in Table 3.5.1 and Figure 3.5.1. Overall, there does not appear to be any substantial difference between the three regions with respect to the overall predictive power of the model. Western Canada and Ontario consistently have very similar adjusted R^2 values while Quebec differs marginally by almost always having a slightly larger value. Therefore, the very similar adjusted R^2 results suggest that at the large provincial scale, there is no distinct spatial variability in the model's predictive power. This infers that at the provincial level for Canada's movie theatre markets, film attributes variables have similar effects on gender/age demographics and therefore, can be used in one predictive that is applied universally across all provinces.

Table 3.5.1 Multiple regression output

Dependent Variable	Western Canada Adjusted R^2 Value	Ontario Adjusted R^2 Value	Quebec Adjusted R^2 Value
Females 14-18	0.260	0.257	0.304
Females 19-24	0.364	0.353	0.430
Females 25-34	0.413	0.400	0.479
Females 35-49	0.383	0.355	0.461
Females 50-64	0.384	0.348	0.439
Females 65+	0.301	0.291	0.334
Males 14-18	0.463	0.446	0.475
Males 19-24	0.528	0.519	0.544
Males 25-34	0.575	0.556	0.575
Males 35-49	0.582	0.557	0.600
Males 50-64	0.532	0.493	0.548
Males 65+	0.416	0.378	0.427
Parents	0.637	0.635	0.601

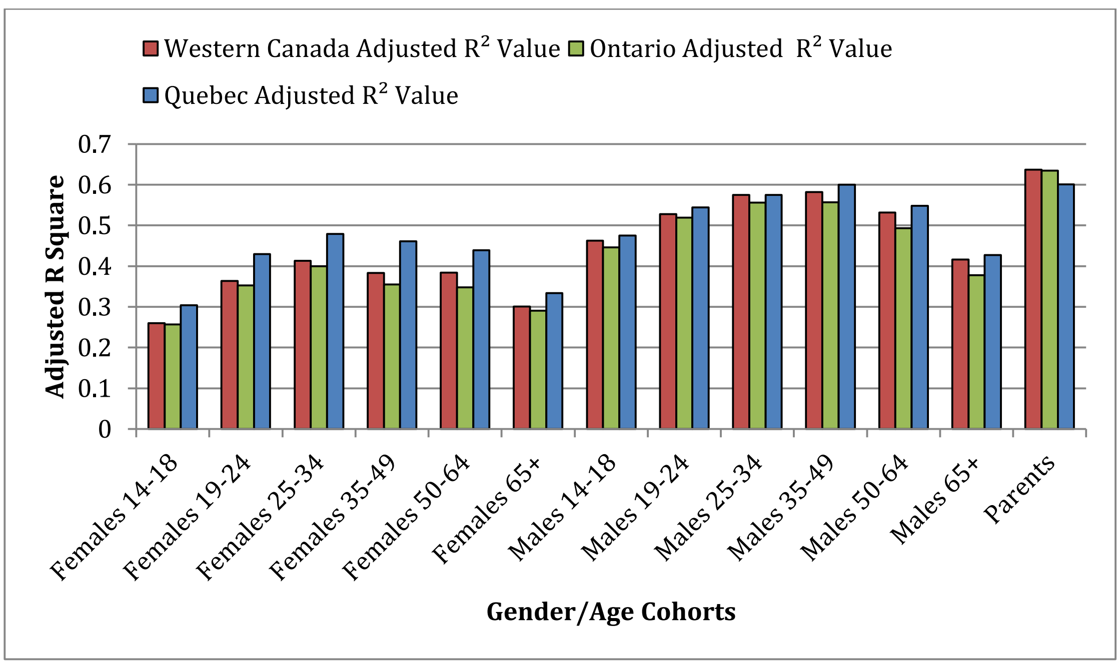

Figure 3.5.1 Adjusted R^2 values by movie theatre region

Analyzing the individual adjusted R^2 values, it is apparent that the model was strongest at predicting the *Parents*, *Males Age 35-49*, *Males Age 25-34*, and *Males Age 50-64* variables for all three regions as all had adjusted R^2 values approximately 0.5 or greater. These values indicate that film attribute variables chosen in this study account for 50 - 60% of the variance in the gender/age demographic variables listed above. In other words, the model was able to predict 60% of the observed values for these gender/age cohort variables.

Conversely, the regression model that had the weakest predictive power was for *Females Age 14-18* accounting for only 26 - 30% of the observed values for all three regions. The remaining eight dependent variables have adjusted R^2 values between 0.3 and 0.49 or 30-50% of their observed value variance is accounted for by the regression model. In general, the model had stronger predictive power for males than it did for females in all three regions.

By analyzing unstandardized coefficients, the impact that each independent variable has in relation to the dependent can be determined. For discussion, only the main contributing variables to the model's predictive power with a significance value between 0.00 and 0.05 will be discussed for each dependent.

Unstandardized Coefficients

Analyzing the weakest to strongest gender/cohorts, (listed below in Table 3.5.2) are unstandardized coefficients (UC) relating to the *Females Age 14-18* dependent variable. The lists of film variables that have the strongest predictive power along with their respective regression weights are almost identical for all three regions. For instance, the

Romance variable has a coefficient of 4.4 for Western Canada, 7.16 for Ontario, and 1.5 for Quebec. This means that the *Romance* variable has a weak positive influence on Females age 14-18 in all three regions.

Interpreting results for Western Canada, it appears that this particular gender/age demographic are disinterested in *Dramas*, *Period* pieces and *Documentaries* since their audience proportion decreases but are attracted to *Sequels*, *Romance*, and *Fantasy* films as shown by the increase in audience proportion. For Ontario, the list of significant film attributes is almost identical to Western Canada with the omission of negatively rated *Period* films. Quebec only has the missing *Documentary* negatively rated attractiveness variable for the demographic.

Table 3.5.2 Significant variables and UC for Females age 14-18

Western Canada F14-18		Ontario F14-18		Quebec F14-18	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Drama	-6.0	*Drama*	-10.08	*Drama*	-1.9
Romance	4.4	*Romance*	7.16	*Romance*	1.5
Sequel	6.89	*Sequel*	9.01	*Sequel*	3.06
Period	-4.83	*Documentary*	-12.10	*Period*	-1.86
Fantasy	4.67	*Fantasy*	6.30	*Fantasy*	1.99
Documentary	-8.19	*Film Runtime*	0.19	*Film Runtime*	0.06
Film Runtime	0.15				

Table 3.5.3 indicates that for Western Canada, Females age 19-24 are similar to the 14-18 cohorts in that their audience proportion decreases with *Drama*, and *Documentary* pieces but increases with *Romance*, *Sequels*, and *Fantasy* films. However, there is the addition of *War* films with negative and *Animated* films with positive impacts on the cohort's audience proportion. Again, Ontario is very similar to this but has added *War* as a third

negatively rated predictor. Quebec also has *Animated* as an additional positive influence on the cohort's audience proportion but is the only region to have a film rating as a significant predictive variable thus far with *PG13*. Unlike Ontario, *Period* pieces as well as *War* films negatively impact Quebec's audience proportion.

Table 3.5.3 Significant variables and UC values for Females age 19-24

Western Canada F19-24		Ontario F19-24		Quebec F19-24	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Drama	-10.51	*Drama*	-16.44	*Drama*	-3.64
Romance	7.88	*Sequel*	16.98	*Sequel*	6.22
Sequel	13.98	*Documentary*	-18.00	*Period*	-3.27
Animated	10.15	*Fantasy*	13.34	*Animated*	3.78
Fantasy	12.12	*Romance*	10.73	*Fantasy*	5.24
War	-39.47	*War*	-50.07	*War*	-15.23
Documentary	-13.12	*Film Runtime*	0.48	*PG13*	12.59
Film Runtime	0.37			*Film Runtime*	0.17

Table 3.5.4 shows that Females age 25-34 in Western Canada are again similar to the previous cohort with respect to film variables that have the strongest relationship in predicting the cohort's audience proportion. It appears that in Western Canada, the negative rating of *Drama* and *Documentary* films decreases slightly compared to Females age 25-34, suggesting that they are not as adverse to them compared to the 19-24 cohort. Ontario's cohort differs from the previous by the addition of positively rated *Action* films. For Quebec, *R* rated and *Horror* films are the only additions from the previous cohort. *R* rated films have a positive relationship while horror movies have a small negative one for

this particular gender/age cohort in Quebec. This infers that Females age 25-34 go to see more *R* rated films compared to the younger cohorts.

Table 3.5.4 Significant variables and UC values for Females age 25-34

Western Canada F25-34		Ontario F25-34		Quebec F25-34	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Drama	-7.60	*Drama*	-11.85	*Drama*	-2.72
Romance	5.53	*Action*	6.24	*Sequel*	5.83
Sequel	13.87	*Romance*	7.55	*Period*	-2.65
Animated	8.79	*Sequel*	15.71	*Animated*	3.35
Fantasy	12.26	*Fantasy*	12.65	*Fantasy*	5.16
War	-36.23	*Documentary*	-13.74	*War*	-12.47
Documentary	-10.71	*War*	-41.76	*PG13*	12.09
Film Runtime	0.40	*Film Runtime*	0.48	*R*	10.14
				Horror	-2.55
				Film Runtime	0.17

Table 3.5.5 displays the regression coefficients relating to the Females age 35-49 cohort for all three study regions. Western Canada's cohort has a few differences with the loss of *Animated* films having a positive relationship compared to the previous cohort. Additionally, *Romance* films are no longer a predictive variable for this female age cohort. Ontario has the addition of *Thriller* films as positive relationships while *Romance* is no longer a positively rated predictor. Quebec loses *Drama* and *Period films* as negative predictors suggesting that this region's cohort is indifferent to them.

Table 3.5.5 Significant variables and UC values for Females age 35-49

Western Canada F35-49		Ontario F35-49		Quebec F35-49	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Drama	-4.54	*Drama*	-8.68	*Sequel*	4.27
Sequel	10.28	*Fantasy*	8.92	*Horror*	-3.00
Fantasy	8.10	*Sequel*	12.60	*Fantasy*	3.49
War	-28.02	*Thriller*	6.00	*PG13*	9.32
Action	4.02	*Documentary*	-12.55	*Film Runtime*	0.13
Documentary	-9.42	*War*	-37.22		
Film Runtime	0.33	*Action*	5.25		
		Film Runtime	0.41		

Analyzing the beta values in Table 3.5.6 below, *Sequel* and *Fantasy* are the only two genres that have a positive predictive rating for Females age 50-64 in Western Canada. *Horror* is the only genre that has a negative relationship with the cohort. Ontario has the same list of variables but with the addition of a positive predictor of *Thriller* films. Quebec is identical to the Western Canada.

Table 3.5.6 Significant variables and UC values for Females age 50-64

Western Canada F50-64		Ontario F50-64		Quebec F50-64	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Sequel	4.60	*Horror*	-6.87	*Sequel*	2.04
Horror	-5.78	*Thriller*	4.58	*Horror*	-3.22
Fantasy	4.33	*Fantasy*	4.99	*Fantasy*	1.96
Film Runtime	0.26	*Sequel*	5.95	*Film Runtime*	0.10
		Film Runtime	0.36		

The list of predictors reduces even more for Females age 65 and older in Table 3.5.7. In the Western Canada, *Drama* is the only positively rated film genre while *Horror* is the only negative. Ontario has even fewer predictors with *Horror* being the sole negatively rated film attribute. Quebec is again identical to Western Canada.

Table 3.5.7 Significant variables and UC values for Females age 65+

Western Canada F65+		Ontario F65+		Quebec F65+	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Drama	1.46	*Horror*	-5.41	*Drama*	1.19
Horror	-3.15	*Film Runtime*	0.19	*Horror*	-2.12
Film Runtime	0.10			*Film Runtime*	0.05

Figures 3.5.2-3.5.7, are stacked bar charts listing the all of the significant film attributes and their respective ratings for each female cohort by region. They provide an overview of the numeric charts above and indicate the cumulative predictive relationships for each female cohort by region. Figure 3.5.8 is a map showing the top three film attributes and their respective unstandardized coefficients for the female cohort having the highest adjusted R^2 value by region.

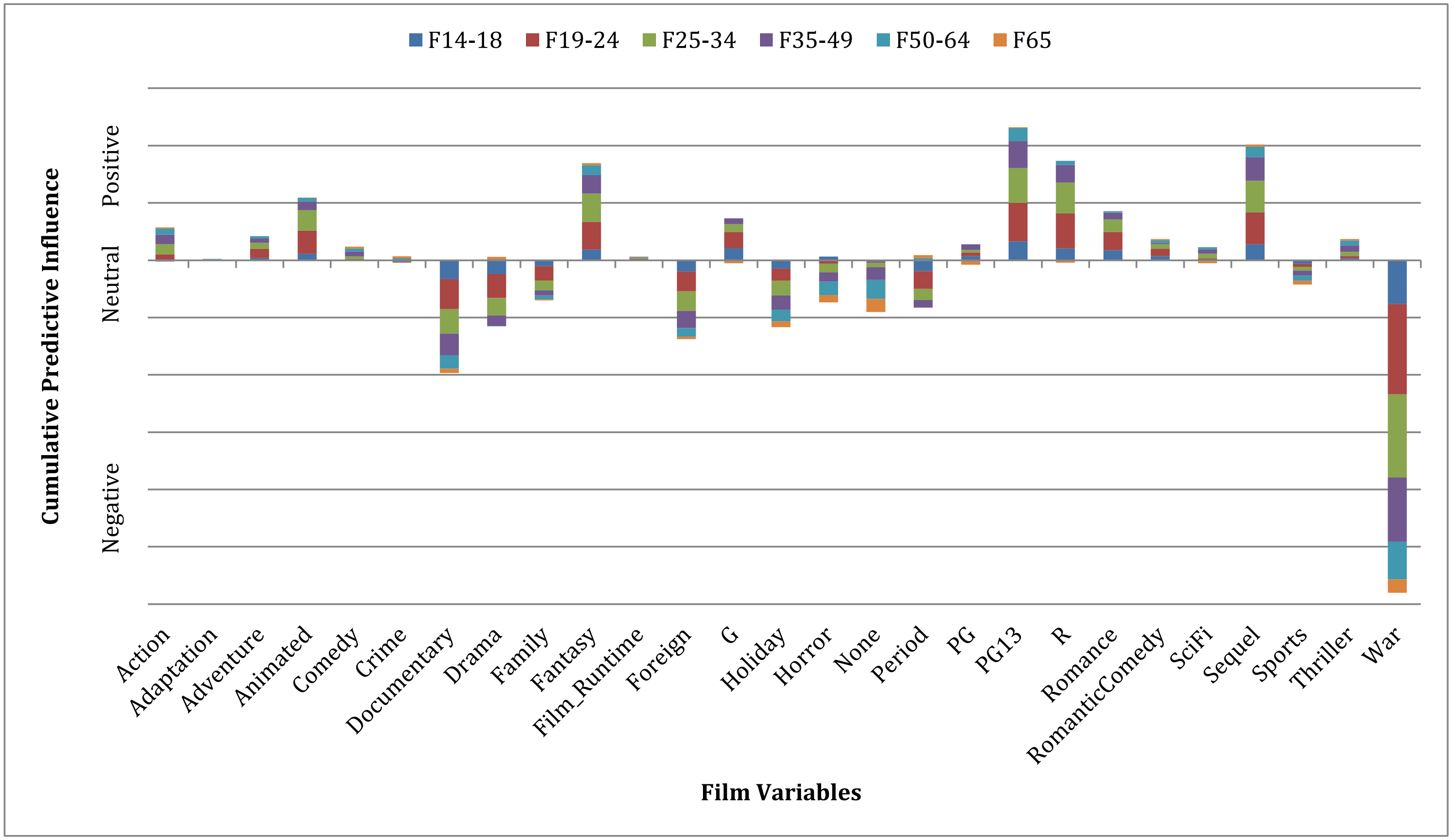

Figure 3.5.2 Predictor Influence for Western Canada Female Cohorts

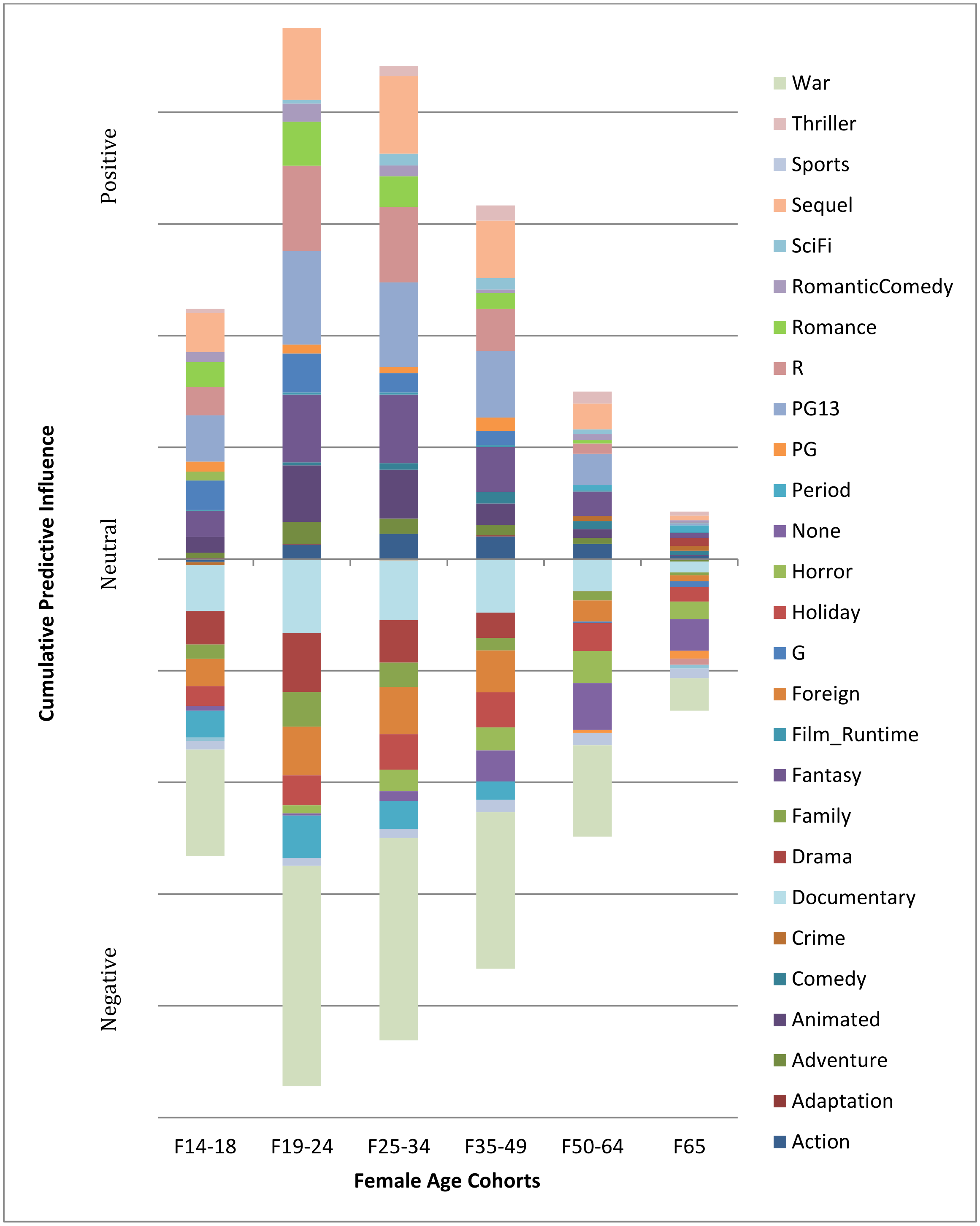

Figure 3.5.3 Predictor Influence for Western Canada Female Cohorts

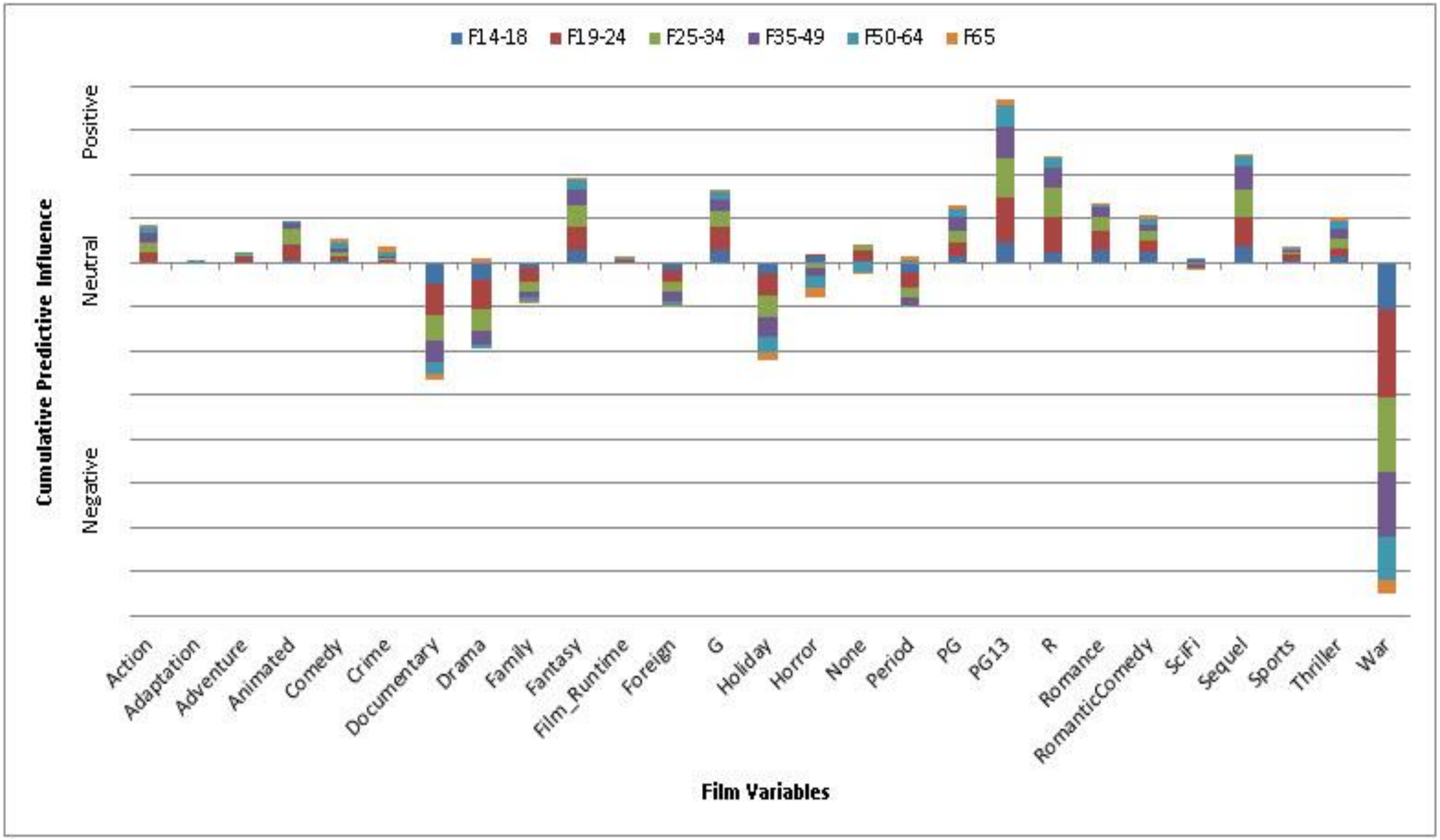

Figure 3.5.4 Predictor Influence for Ontario Female Cohorts

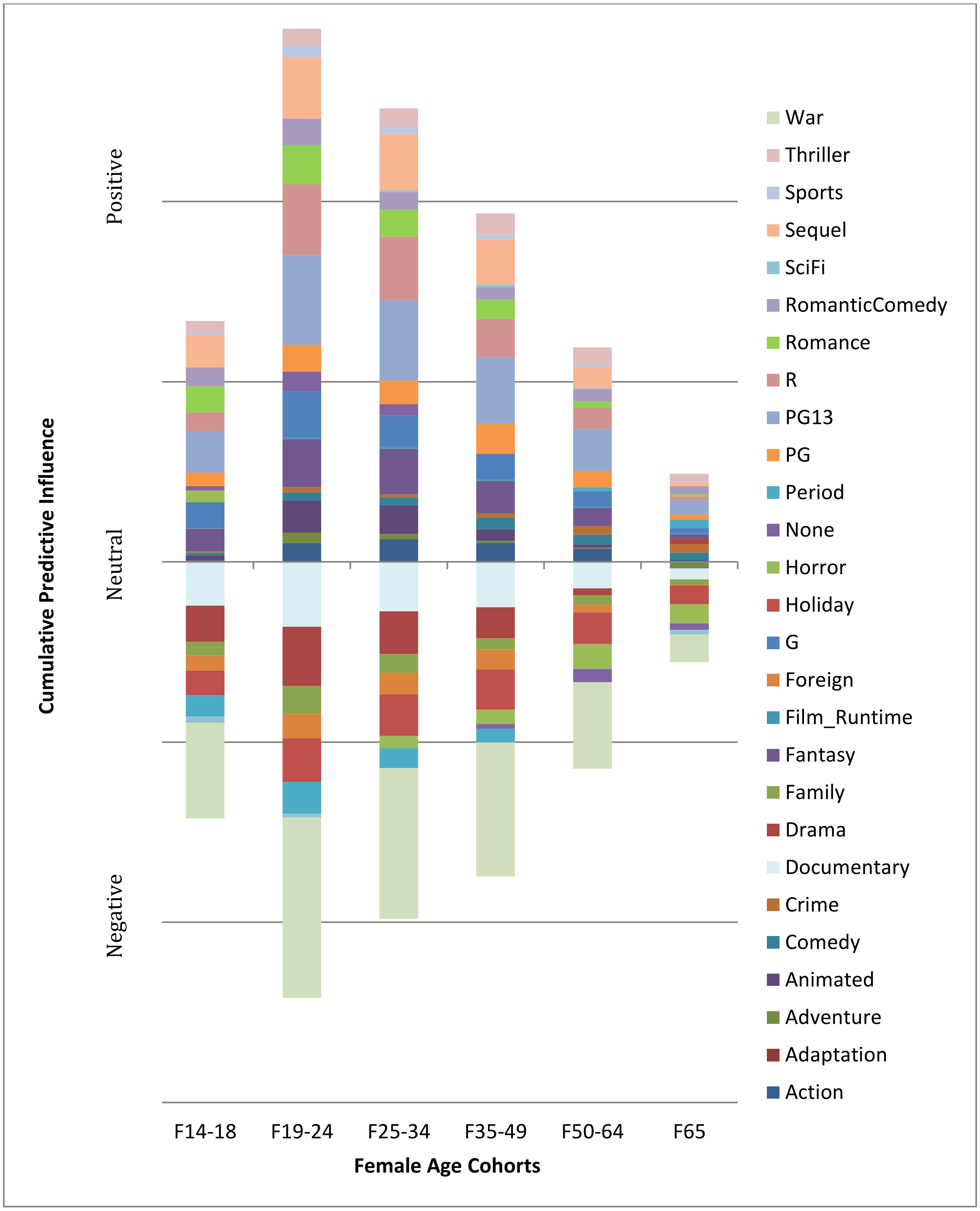

Figure 3.5.5 Predictor Influence for Ontario Female Cohorts

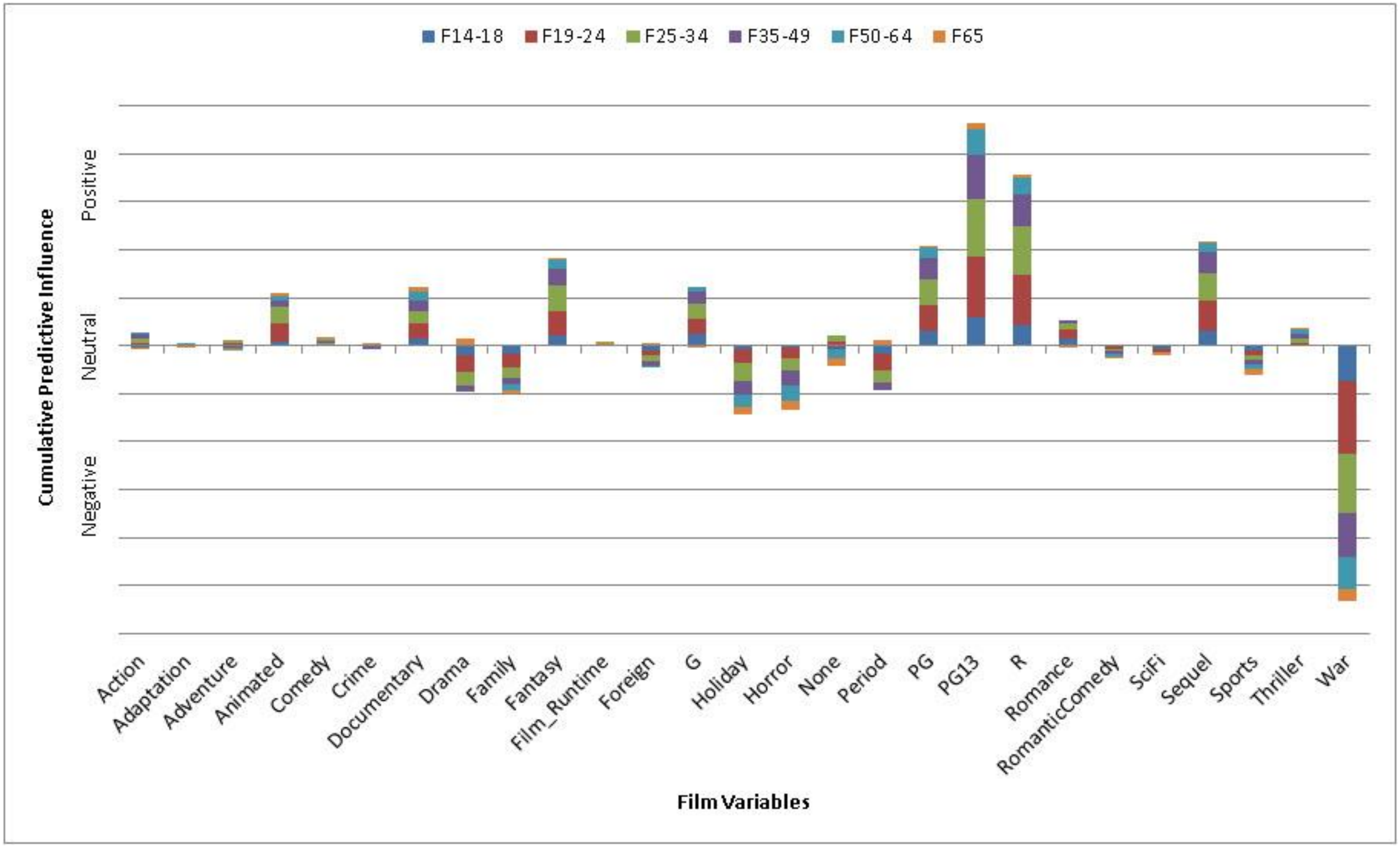

Figure 3.5.6 Predictor Influence for Quebec Female Cohorts

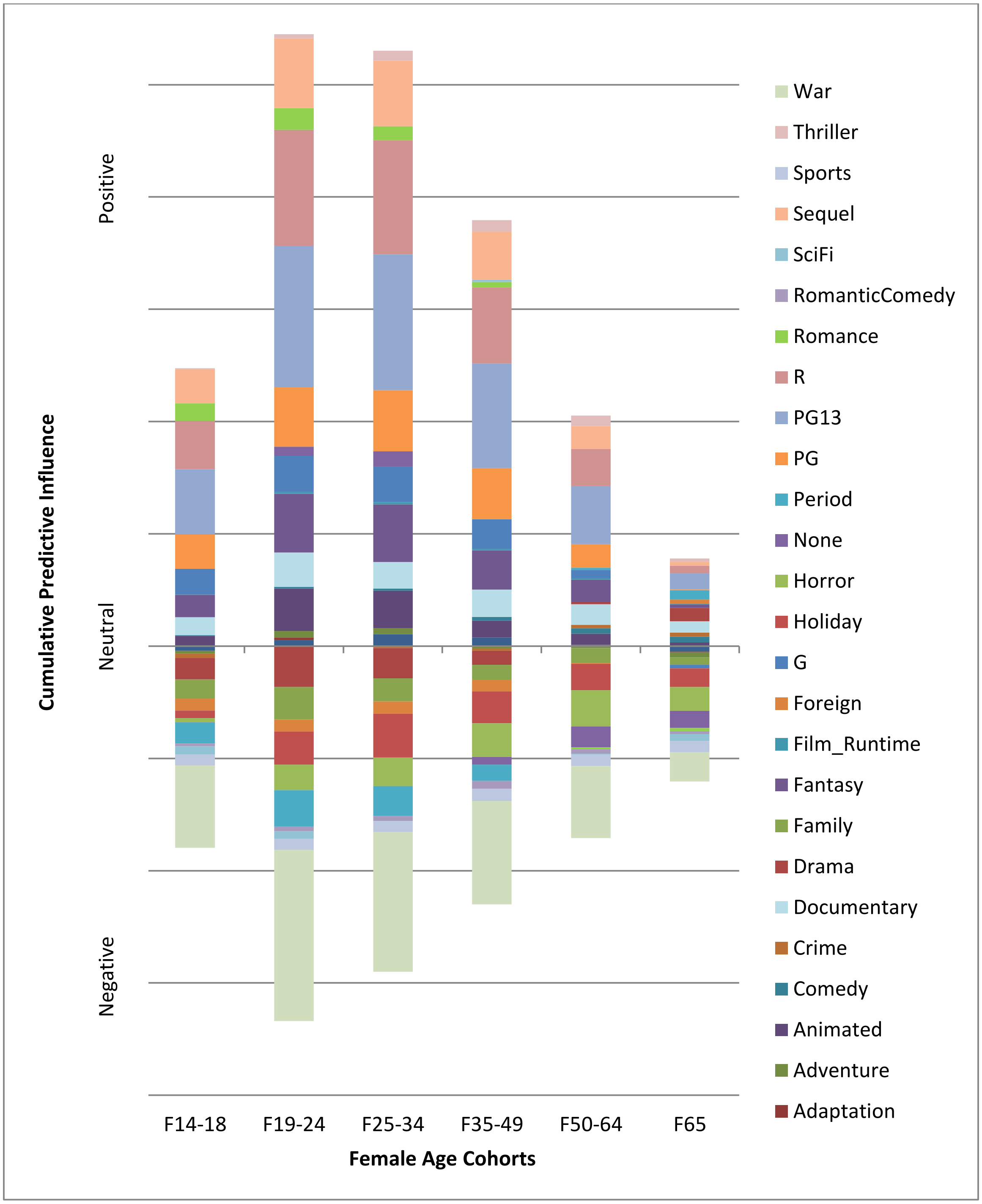

Figure 3.5.7 Predictor Influence for Quebec Female Cohorts

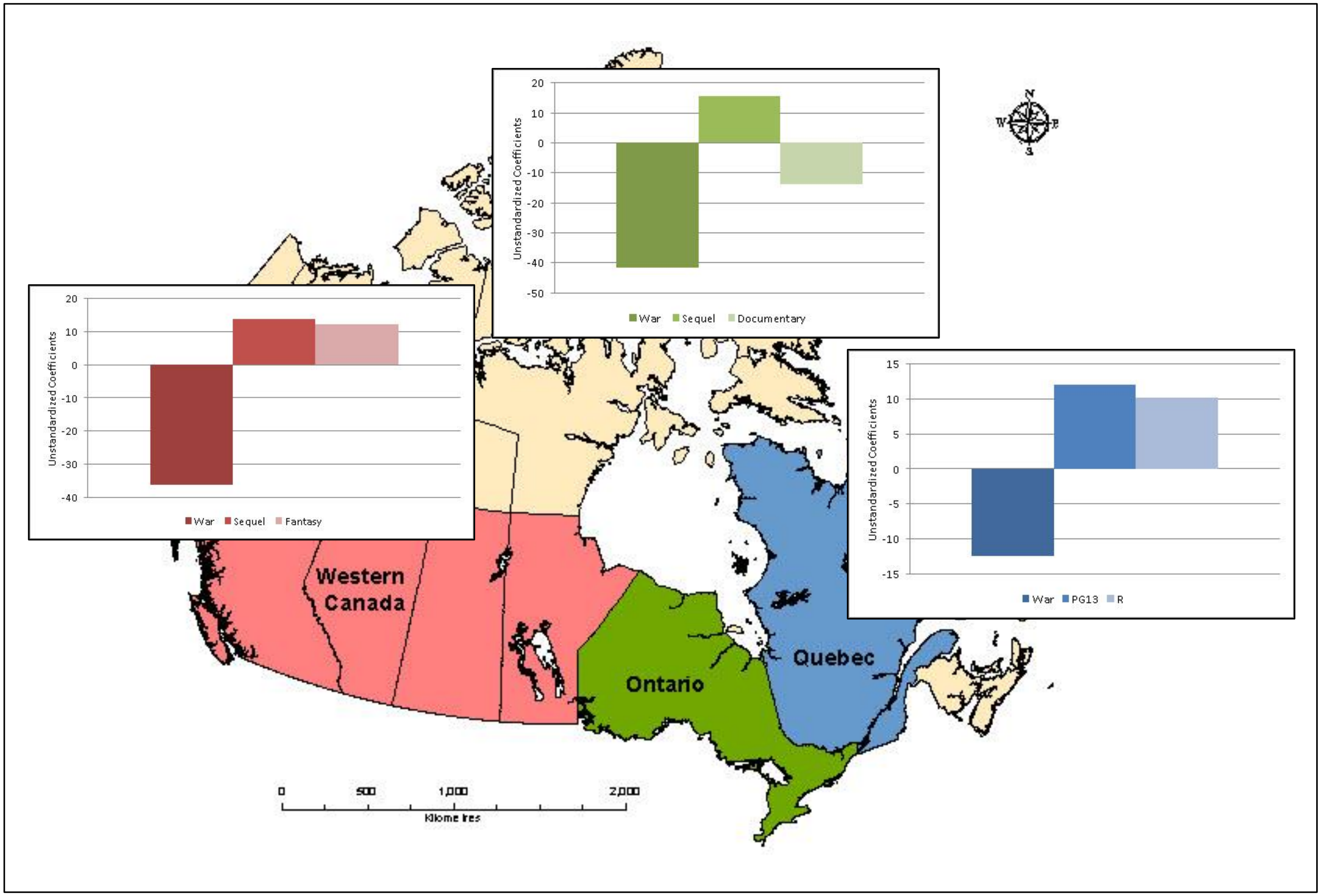

Figure 3.5.8 Top three genres for Females age 25-34 by region

Analyzing the Male age 14-18 beta values in Table 3.5.8, *Action, Sequel, Adventure, PG13 and R* rated films are all positive predictors. *Drama, Period, War, and Documentary* are all negatively rated genres for this cohort in the Western Canada study region. Ontario is very similar to the Western Canada but with the omission of the negatively rated *Period* genre predictor. The Quebec study region differentiates the most with the substitution of *Documentary* for *Romantic Comedy* and added *Family* as negative predictors. *Action* and *Adventure* are both dropped as a positive predictor suggesting that the cohort in the Quebec could be indifferent to these film genres.

Table 3.5.8 Significant variables and UC values for Males age 14-18

Western Canada M14-18		Ontario M14-18		Quebec M14-18	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Action	3.51	*Action*	5.82	*Drama*	-2.74
Drama	-7.89	*Drama*	-12.68	*Sequel*	2.81
Sequel	6.41	*Sequel*	8.73	*Period*	-1.68
Period	-3.82	*Adventure*	4.11	*Romantic Comedy*	-1.69
Adventure	3.49	*Documentary*	-15.16	*Family*	-1.79
War	-19.55	*War*	-30.46	*War*	-6.55
Documentary	-9.99	*PG13*	17.55	*PG13*	6.45
PG13	11.46	*R*	16.17	*R*	6.25
R	11.41	*Film Runtime*	0.17	*Film Runtime*	0.05
Film Runtime	0.13				

The differences between the three regions for Males age 19-24 are shown in Table 3.5.9 are that Western Canada is the only region to have *Sci-Fi* as a positive predictor. It also shares *Animated* as another positive predictor with the Quebec. Ontario shares all of its predictors, both positive and negative, with at least one of the other two regions. Quebec differs by having *Period* films as a negative predictor and has lower coefficient values for both *PG13* and *R* rated films. This suggests that this male cohort in the Quebec has a reduced preference for these higher rated films in comparison to the other two regions.

Table 3.5.9 Significant variables and UC values for Males age 19-24

Western Canada M19-24		Ontario M19-24		Quebec M19-24	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Action	9.08	*Action*	13.16	*Action*	2.94
Drama	-13.04	*Drama*	-19.13	*Drama*	-4.73
Sequel	12.27	*Sequel*	15.30	*Sequel*	5.46
Adventure	6.96	*Adventure*	7.65	*Period*	-3.07
Animated	8.03	*Fantasy*	10.08	*Adventure*	2.15
Sci-Fi	5.77	*Documentary*	-20.25	*Animated*	3.18
Fantasy	8.93	*War*	-50.94	*Fantasy*	3.86
War	-40.30	*PG13*	28.96	*Family*	-3.12
Documentary	-14.99	*R*	30.31	*War*	-16.04
PG13	20.80	*Film Runtime*	0.38	*PG13*	12.96
R	24.15			*R*	13.13
Film Runtime	0.30			*Film Runtime*	0.15

The only regional differences displayed in Table 3.5.10 are that the Quebec has *Period* films as negative predictors and again has lower predictor strength with *PG13 and R* rated films. Again, it appears that the higher rated films are less favourable to males in the Quebec study region. The remaining predictors are all common amongst each other.

Table 3.5.10 Significant variables and UC values for Males age 25-34

Western Canada M25-34		Ontario M25-34		Quebec M25-34	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Action	10.69	*Action*	14.05	*Action*	3.94
Drama	-10.74	*Drama*	-14.64	*Drama*	-4.28
Sequel	12.56	*Sequel*	14.05	*Sequel*	5.57
Adventure	6.47	*Adventure*	5.83	*Period*	-2.95
Animated	8.28	*Fantasy*	11.22	*Adventure*	2.18
Sci-Fi	7.86	*Documentary*	-15.02	*Animated*	3.67
Fantasy	11.07	*Sci-Fi*	6.93	*Fantasy*	4.92
War	-38.92	*War*	-47.15	*War*	-15.40
Documentary	-12.57	*PG13*	25.82	*Romance*	-2.17
PG13	19.72	*R*	26.44	*PG13*	13.55
R	21.92	*Film Runtime*	0.41	*R*	13.31
Film Runtime	0.34			*Film Runtime*	0.18

Regional differences observed from Table 3.5.11 are that no film ratings have any significant predictive power for Western Canada, Quebec is the only one to have *Romance* as a negative predictor and as with the last two cohorts, *PG13 and R* rated films have a positively rated predictive relationship for the Ontario and Quebec regions. The remaining significant predictors are all shared with at least two regions.

Table 3.5.11 Significant variables and UC values for Males age 35-49

Western Canada M35-49		Ontario M35-49		Quebec M35-49	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Action	9.33	*Action*	11.51	*Action*	3.28
Drama	-6.27	*Drama*	-8.94	*Drama*	-2.42
Sequel	8.42	*Sequel*	9.63	*Adventure*	1.44
Adventure	4.51	*Adventure*	4.22	*Romance*	-1.78
Sci-Fi	6.56	*Fantasy*	8.22	*Sequel*	3.84
Fantasy	7.55	*Documentary*	-10.95	*Sci-Fi*	2.12
War	-24.83	*Sci-Fi*	6.26	*Horror*	-2.06
Documentary	-9.07	*War*	-32.89	*Fantasy*	3.43
Film Runtime	0.26	*PG13*	18.76	*War*	-9.94
		Film Runtime	0.30	*PG13*	11.31
				R	10.12
				Film Runtime	0.13

Differences between regions shown in Table 3.5.12 are that *Adventure*, and *Sci-Fi* genres are both positive predictors in the Western Canada. For Ontario, *Drama* and *Documentary* genres are both negative predictors. In Quebec, *Romance* is a negatively rated predictor. Furthermore, it is the only region to have rating attributes as predictors.

Table 3.5.12 Significant variables and UC values for Males age 50-64

Western Canada M50-64		Ontario M50-64		Quebec M50-64	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Action	5.96	*Action*	7.84	*Action*	1.98
Sequel	3.71	*Drama*	-3.33	*Sequel*	1.63
Horror	-4.82	*Sequel*	4.84	*Horror*	-2.66
Adventure	2.44	*Fantasy*	4.22	*Romance*	-1.37
Documentary	-5.43	*Horror*	-5.33	*Fantasy*	1.85
Sci-Fi	3.26	*Thriller*	3.81	*PG13*	6.40
Fantasy	3.63	*War*	-21.10	*R*	5.34
Film Runtime	0.19	*Documentary*	-6.28	*Film Runtime*	0.09
		Film Runtime	0.26		

In Table 3.5.13, the Western Canada and Ontario regions have almost identical predictors with approximately the same values or predictive strength. Quebec however, differs by having *Romance* as a negative predictor.

Table 3.5.13 Significant variables and UC values for Males age 65+

Western Canada M65+		Ontario M65+		Quebec M65+	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Action	2.32	*Action*	3.48	*Action*	0.65
Horror	-3.14	*Horror*	-4.64	*Romance*	-0.81
Film Runtime	0.08	*Thriller*	2.42	*Horror*	-1.95
		Film Runtime	0.14	*Film Runtime*	0.04

As seen in Table 3.5.14, the three regions have very similar lists of predictors for the Parents cohort. In Western Canada, the majority of predictors are shared with Ontario with the exception of the positively rated *Adventure* genre. Quebec differs from the other two by having *Horror* as a negative predictor. Peculiarly, *Family* is not a significant predictor for parents in Quebec. However, *Animated* is the strongest predictor for this cohort in all three regions.

Table 3.5.14 Significant variables and UC values for Parents

Western Canada Parents		Ontario Parents		Quebec Parents	
Film Variables	UC	Film Variables	UC	Film Variables	UC
Action	4.27	*Comedy*	5.57	*Drama*	-2.61
Drama	-7.69	*Action*	5.97	*Sequel*	4.45
Sequel	14.26	*Drama*	-12.08	*Horror*	-1.89
Adventure	4.84	*Sequel*	17.66	*Animated*	6.37
Animated	21.36	*Animated*	21.91	*Fantasy*	2.79
Sci-Fi	5.47	*Sci-Fi*	7.19	*Film Runtime*	0.08
Fantasy	9.85	*Fantasy*	11.16		
Family	9.58	*Family*	13.86		
Film Runtime	0.24	*Film Runtime*	0.30		

Figures 3.5.9-3.5.14, are stacked bar charts listing the all of the significant film attributes and their respective ratings for each male and parent cohort by region. They provide an overview of the numeric charts above and indicate the cumulative predictive relationships for each male/parent cohort by region. Figures 3.5.15-3.5.16 are maps showing the top three film attributes and their respective unstandardized coefficients for the male and parent cohorts having the highest adjusted R^2 value by region.

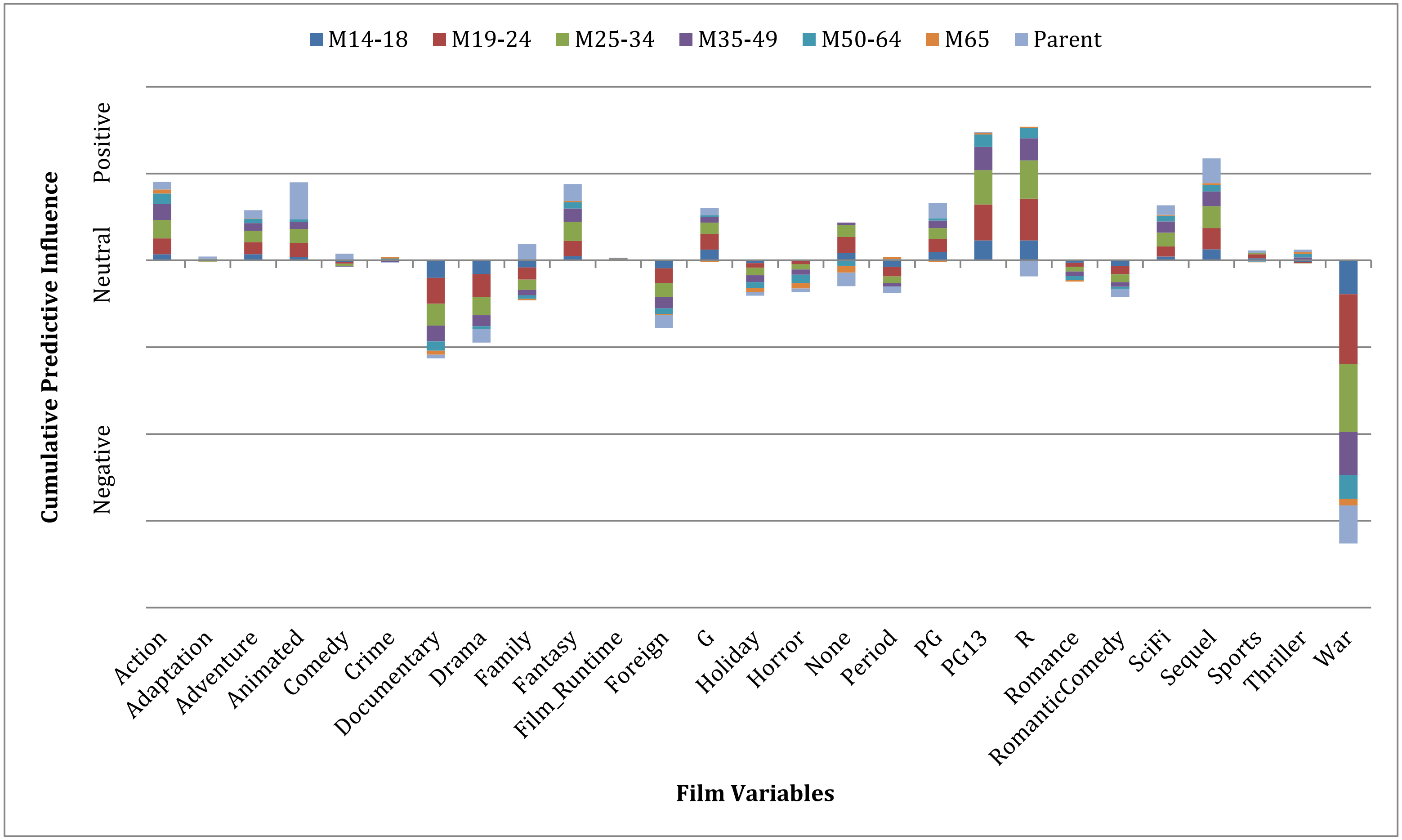

Figure 3.5.9 Predictor Influence for Western Canada Male and Parent Cohorts

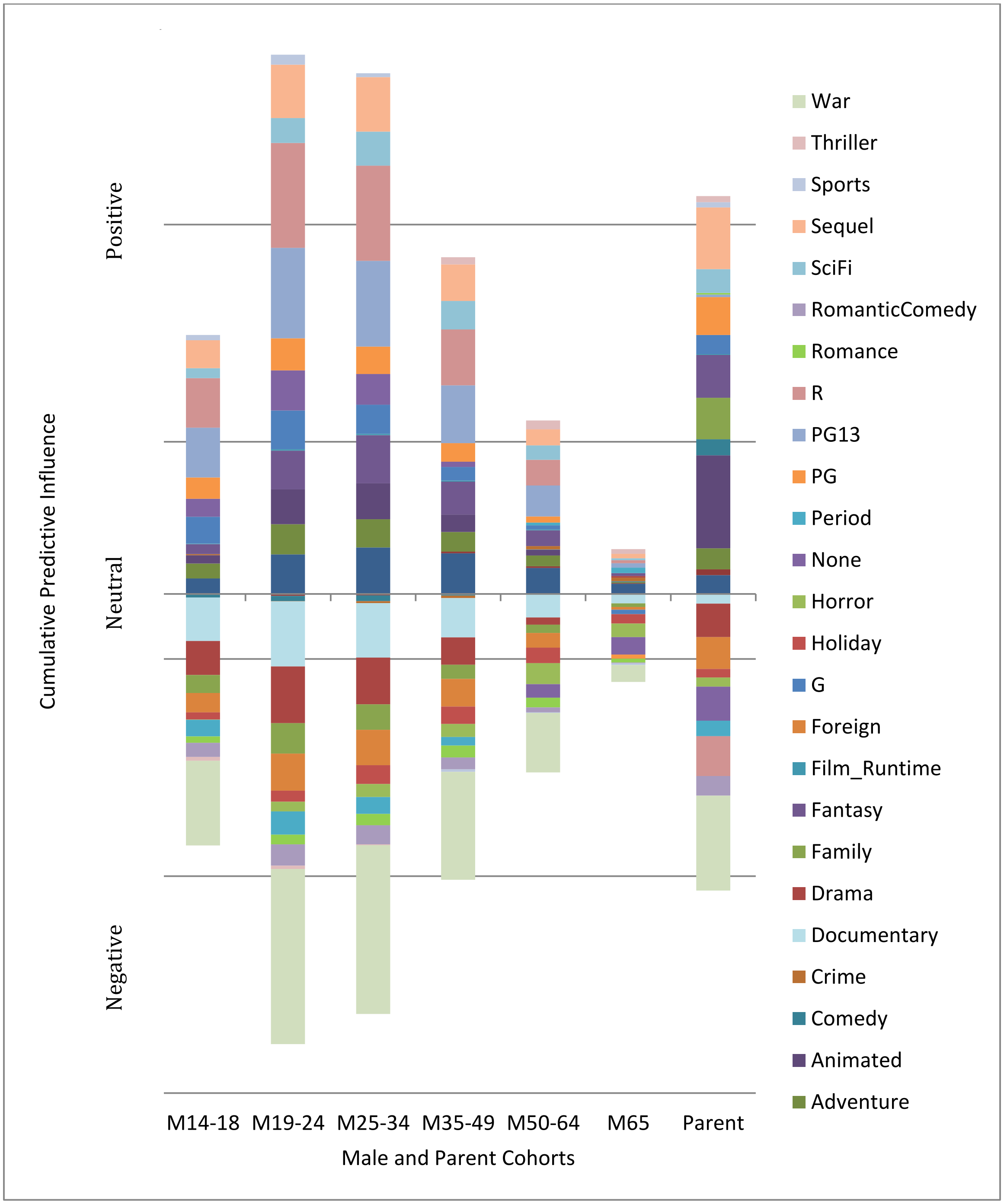

Figure 3.5.10 Predictor Influence for Western Canada Male and Parent Cohorts

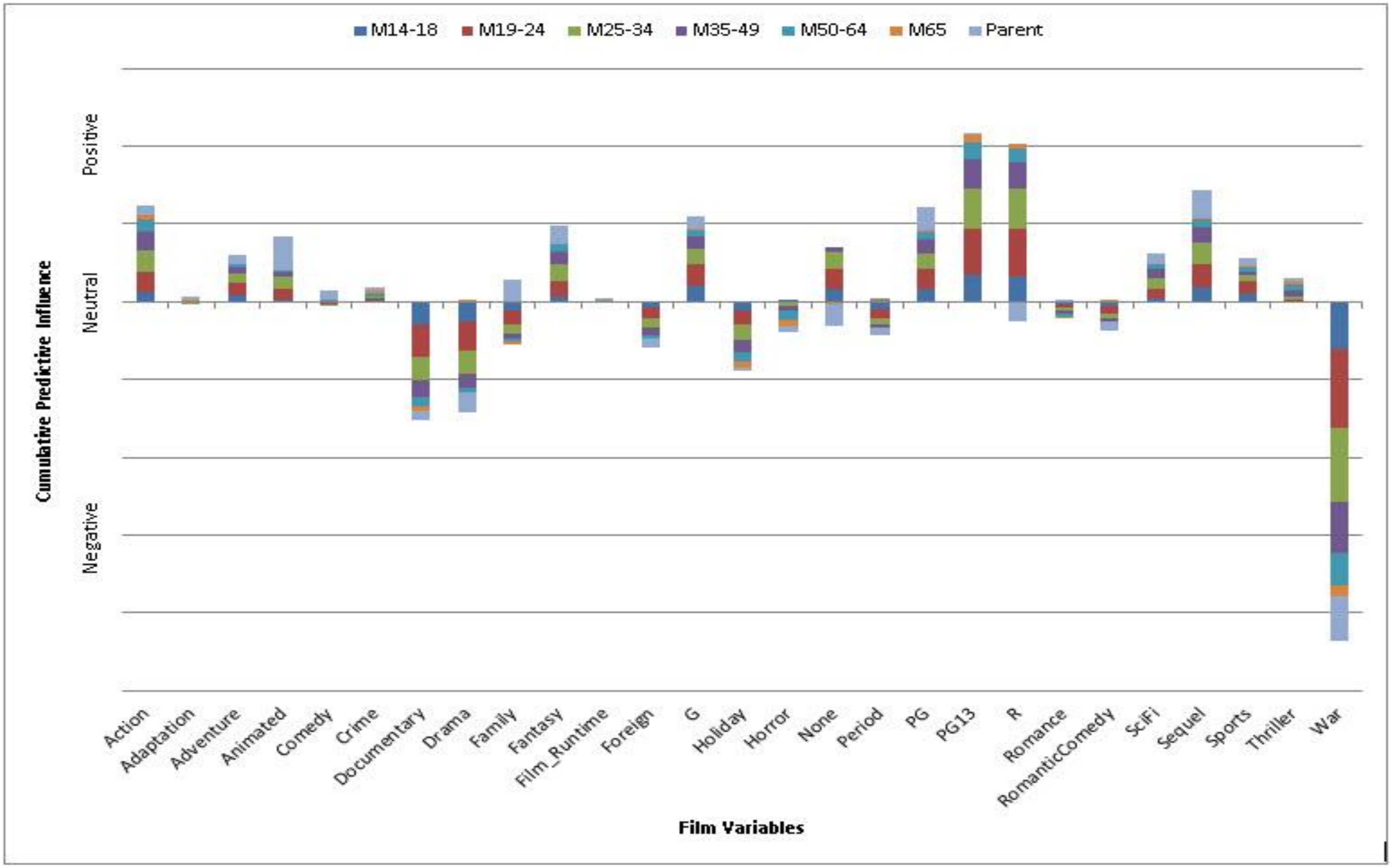

Figure 3.5.11 Predictor Influence for Ontario Male and Parent Cohorts

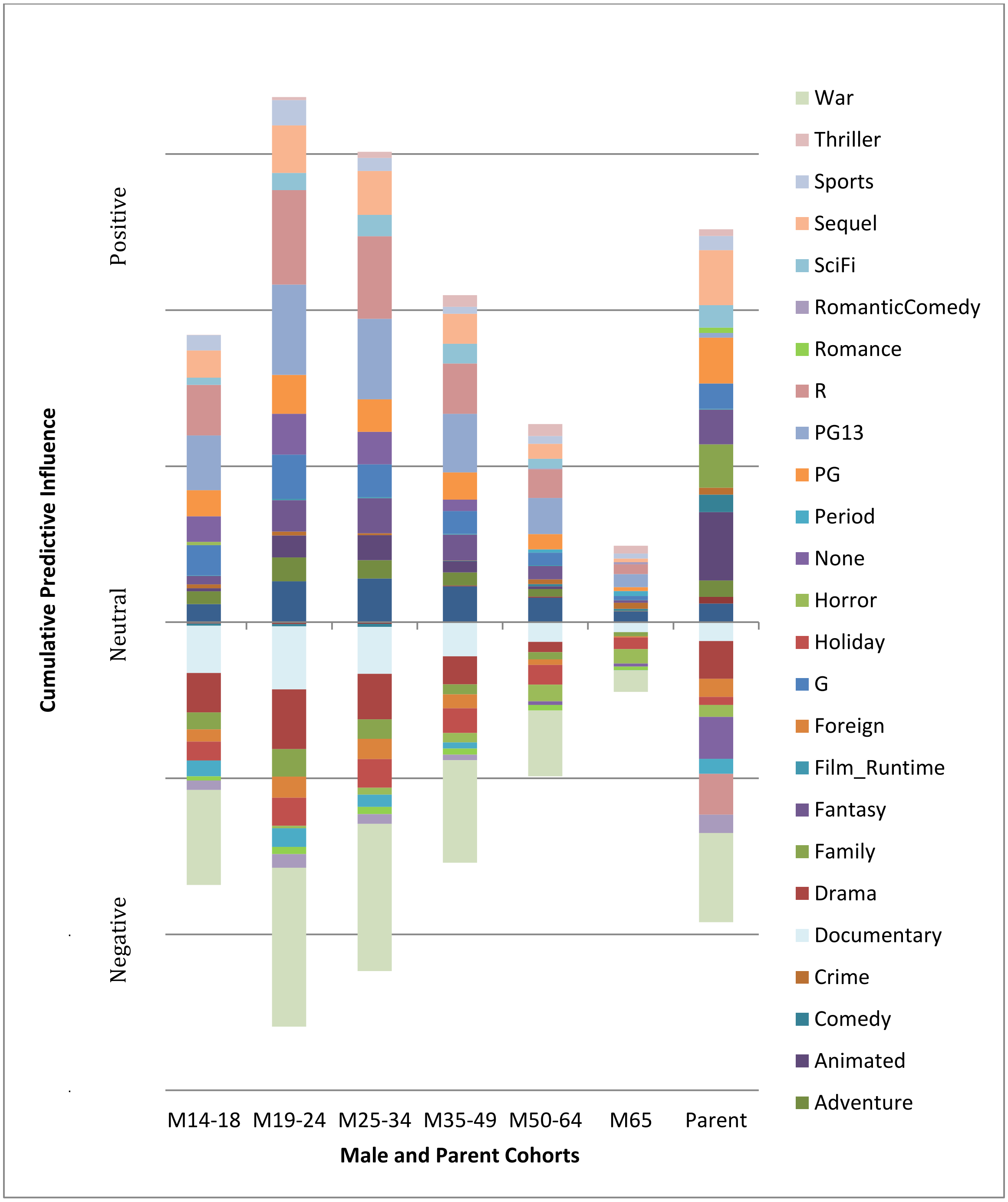

Figure 3.5.12 Predictor Influence for Ontario Male and Parent Cohorts

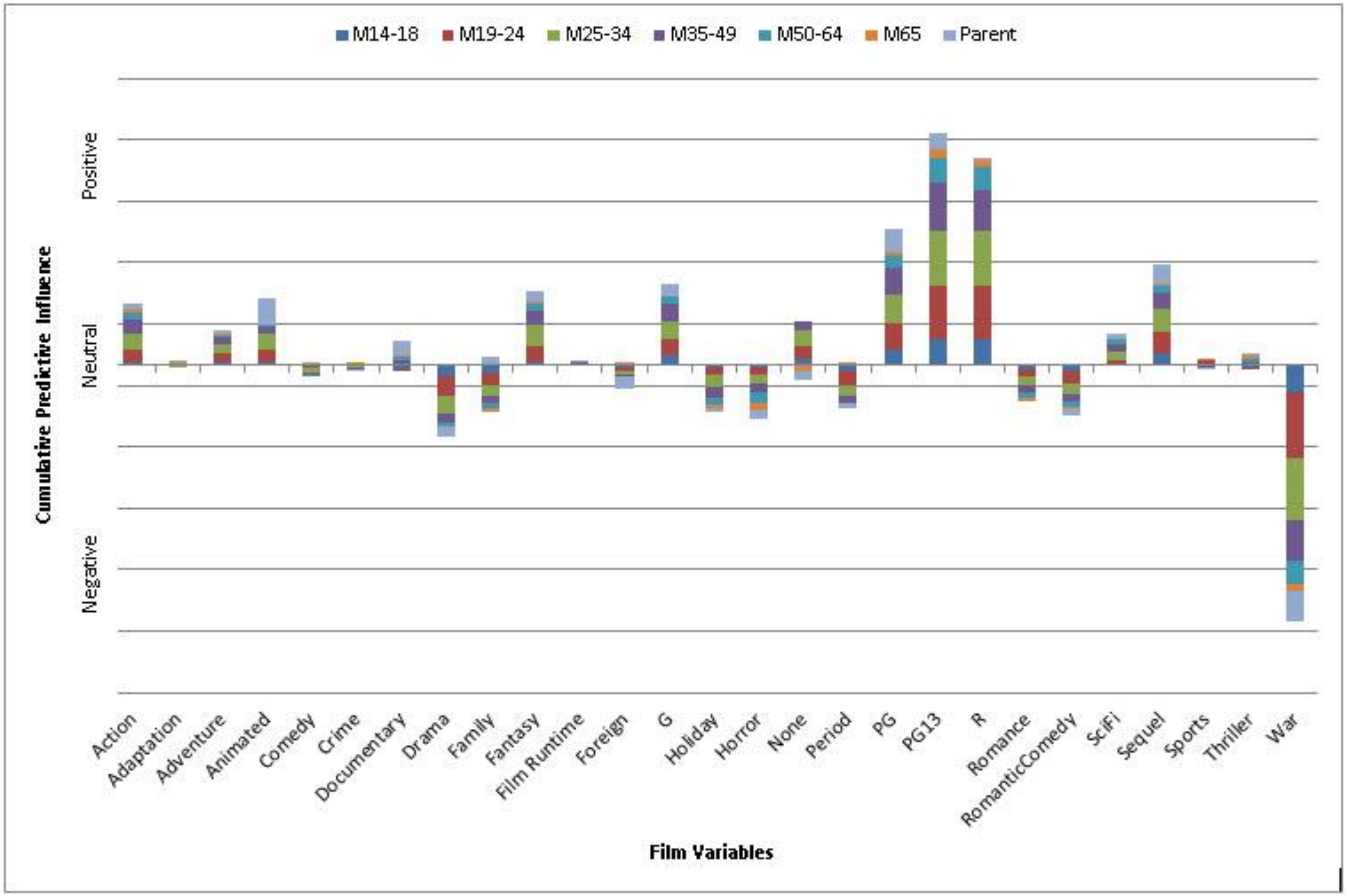

Figure 3.5.13 Predictor Influence for Quebec Male and Parent Cohorts

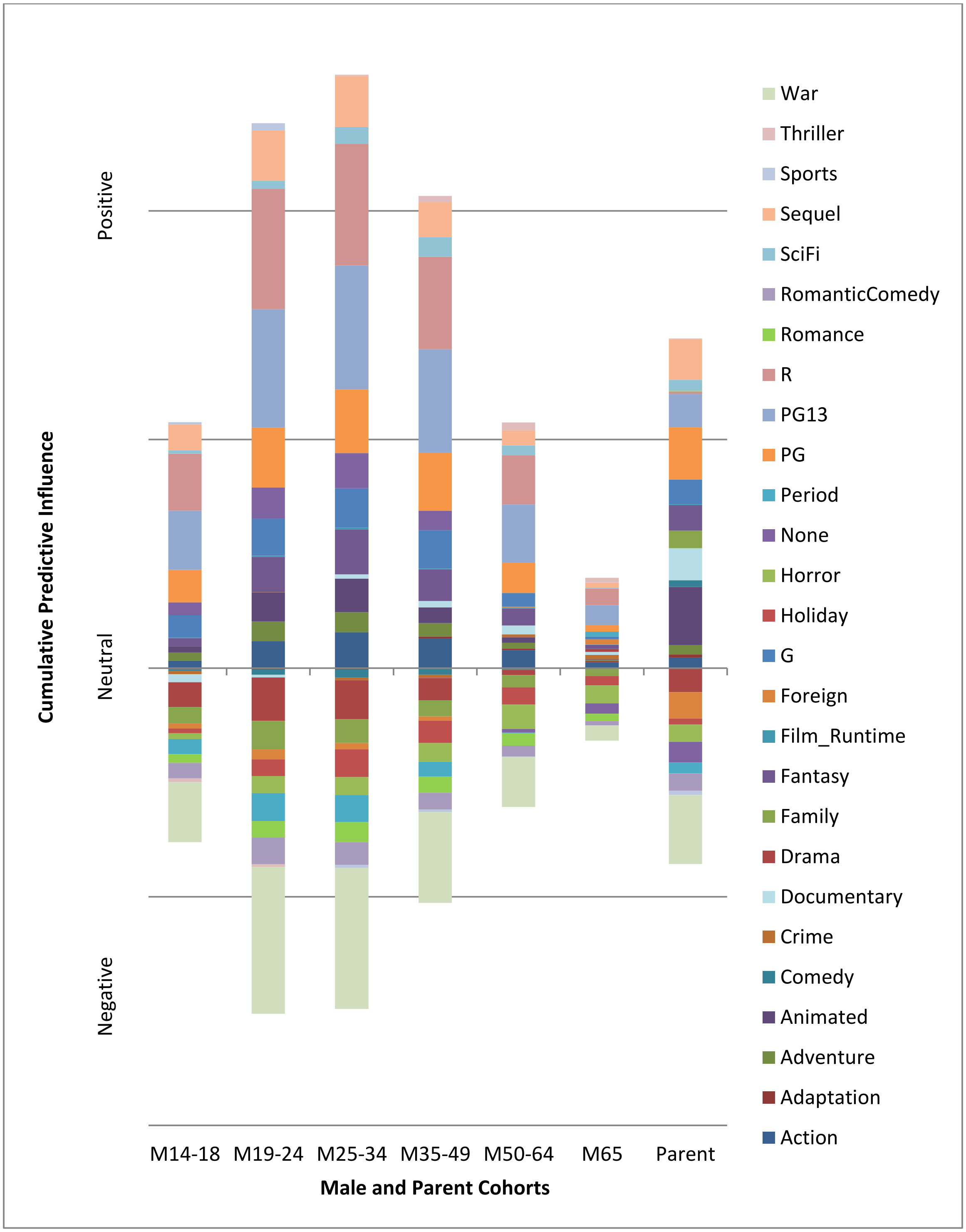

Figure 3.5.14 Predictor Influence for Quebec Male and Parent Cohorts

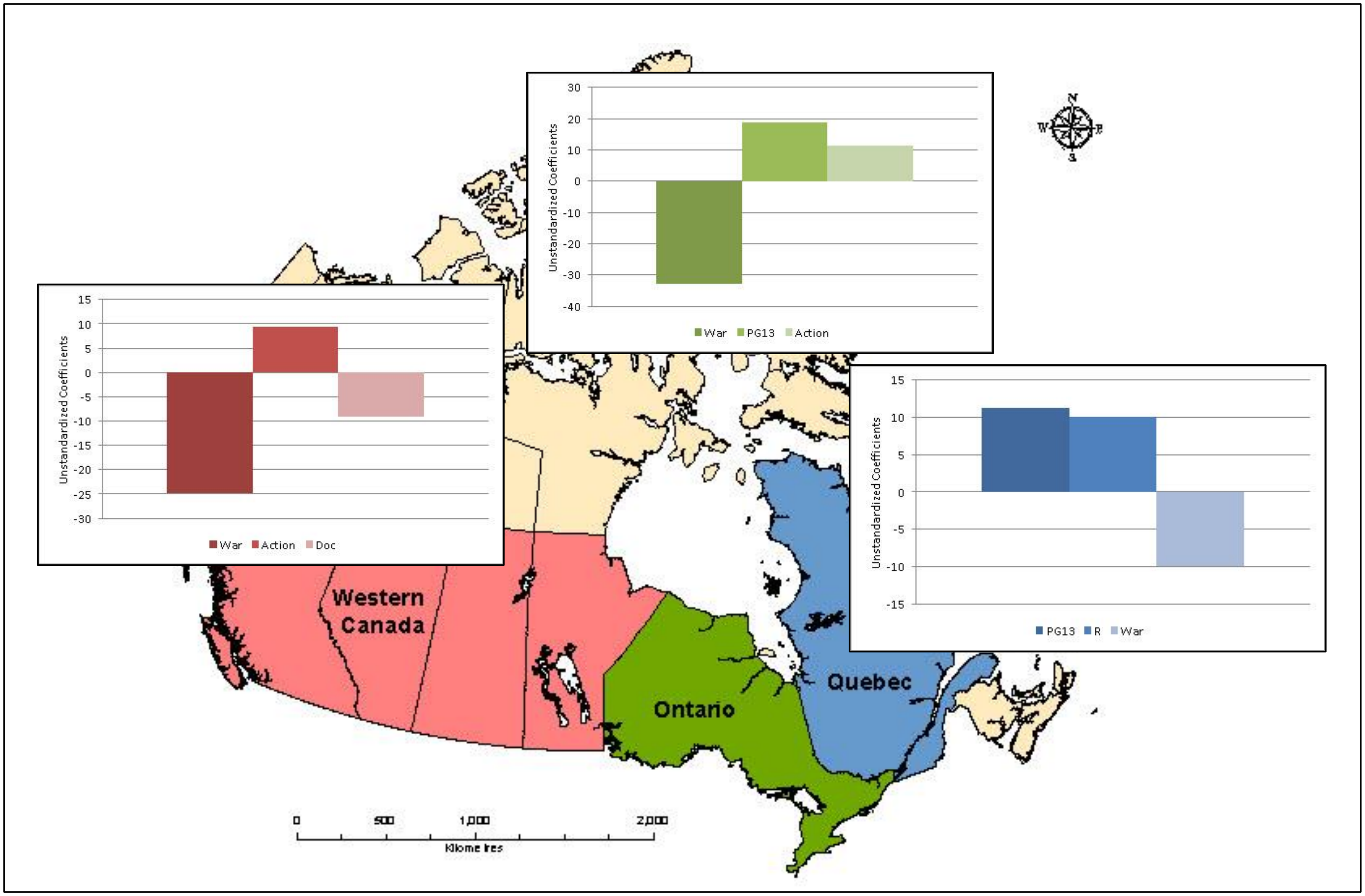

Figure 3.5.15 Top three genres for Males age 35-49

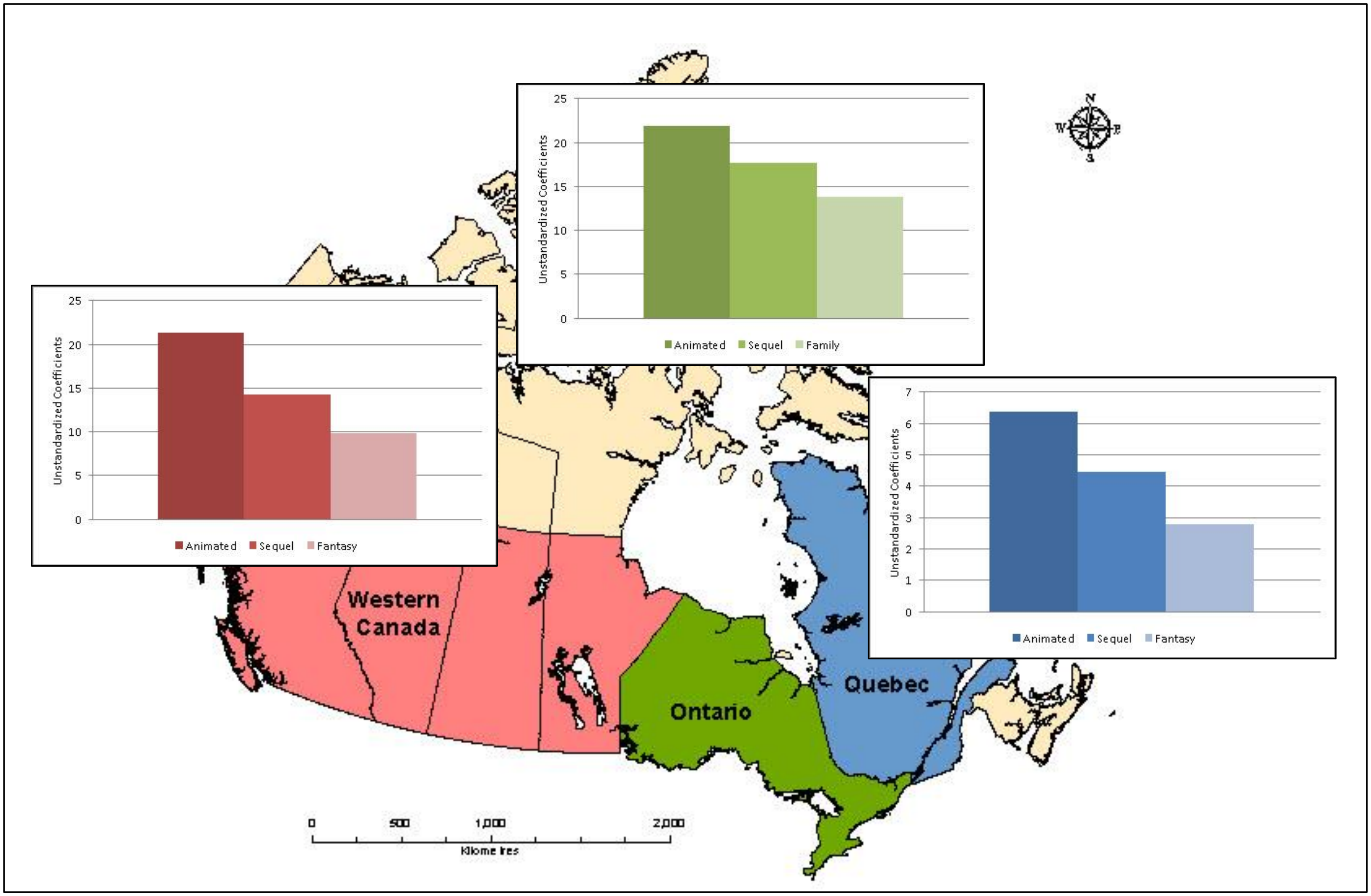

Figure 3.5.16 Top three genres for the Parent cohort

Discussion and Conclusion

After exploring and comparing the significant predictors for each gender age cohort, it is apparent that there are many commonalities among the three study regions. The *War* genre has a fairly universal negative predictor rating for all gender/age cohorts with the exception of the two Female cohorts with ages 50 and above, the Male cohort age 65 and above, and the Parents cohort. This suggests that most cohorts have some form of disinterest with *War* films for it to have a negative relationship. For the older generation however, it appears that the genre does not have any positive or negative relationships suggesting that they are indifferent to the genre.

Film Runtime is the only film attribute that was a significant predictor for all cohorts in all three regions. Furthermore, it always has a positive relationship with all cohorts. However, as the cohorts increased in age, so too does the strength of the relationship with *Film Runtime*. This infers that as people get older, their preference for longer movies increases.

Another observed trend as cohorts get older is that the number of significant predictors decreases. This infers that for the younger cohorts, film attributes have a more important role in attracting these audience members which boosts their audience proportion however, the importance of the basic film attribute variables used in this study decreases as they age. For the older cohorts, it is apparent that other film variables not used in this research are more influential in the viewership and therefore determining the audience proportion of older cohorts. The numbers of awards, actor appeal, film budget and season to name a few are additional film attributes variables that were not included in this study.

Finally, although Quebec tends to have the strongest adjusted R^2 values for all but one cohort, it most often had fewer significant predictor variables compared to the other two regions. Furthermore, having fewer predictors but a greater adjusted R^2 value suggests that of the few film attribute variables that were significant in Quebec, their predictive strength must be greater than any other attribute relationship in Ontario and Western Canada.

Based on the results of this research, it is evident that film attributes can be used to predict gender/age proportions of a film's audience. The overall variance accounted for by the regression model is fairly good considering the limited selection of film attribute variables. Furthermore, the model strength was consistent for all three study regions further implying its reliability at predicting audience demographic proportions. Overall, the results do not produce very strong predictive relationships with the film attributes used in this analysis but they are still significant and demonstrate that film attributes can be used for predictive purposes for a film's audience demographic composition.

Chapter 4: Limitations and Recommendations for Future Research

Due to the coarse scale at which the data were aggregated and the fairly concise and narrow selection of independent variables, this study provided a limited analysis of the relationships between film attributes and gender/age cohort demographics. Analyzing predictive relationships between film attributes and audience gender/age cohort demographics at a smaller scale may produce different results than what was observed in this research. However, the results may not differ and using a coarse provincial scale may be sufficient when predicting gender/age cohort compositions of a newly released film's potential audience.

As previously stated, there are many more film attributes that still need to be investigated in regards to their predictive relationships with audience demographic proportions. Furthermore, there could have been a string of badly performing films for a particular genre in the three year span that the data were gathered from. This would influence the results by causing some genres to appear to have weak or no predictive power. Using a longer time period, these same genres could actually have strong predictive strength.

There were also other temporal factors that limited the scope of research undertaken. Data processing and time restraints dictated the amount and depth of analysis that could be incorporated and undertaken for this study.

Concerning the dataset, a limitation relating to the representation of each cohort in the dataset could have been a factor. The *Parent* cohort could have been underrepresented in the dataset because parents with young children typically see movies less often than individual males and females in older cohorts. This is because older cohorts are independent and are able to visit a movie theatre without a parent accompanying them.

As a result of the limitations and the fact that there has been very little research relating to audience demographics, there is a substantial amount of further research that can be done in the motion picture industry. Some recommendations for future studies include the investigation of the many other film attributes (other film genres, film seasons, competing films, production budget, award nominations, actor appeal) that were not assessed in this research as well as the extension of this study to other theatre markets to determine if relationships are similar with other national and international markets.

References

Agostinelli, C. (2002). Robust stepwise regression. *Journal of Applied Statistics.* 29 (6) 825-840. Retrieved from: doi: 10.1080/02664760220136168

Afifi, A. and Bendel, R. (1976). A Criterion for Stepwise Regression. *The American Statistician.* 30 (2) 85-87. Retrieved from: http://www.jstor.org/stable/2683800

Bartels, R. and Goodhew, J. (1981). The Robustness of the Durbin-Watson Test. *The Review of Economics and Statistics.* 63(1) 136-139. Retrieved from: http://www.jstor.org/stable/1924228

Basuroy, S., Ravid, A., and Wald, J. (2006). Distributors and film critics: does it take two to Tango?. *Journal of Cultural Economics.* 30 (3) 201-218. Retrieved from: http://resolver.scholarsportal.info.ezproxy.lib.ryerson.ca/resolve/08852545/v30i0003/201_dafcditttt

Breaux, H. (1968). A modification of Efroymson's technique for stepwise regression analysis. *Communications of the ACM.* 11 (8) 556-558. Retrieved from: doi:10.1145/363567.363580

Chen, R. (2007). Geographic information systems (GIS) applications in retail tourism and teaching curriculum. *Journal of Retailing and Consumer Services.* 14 (2007) 289-295. Retrieved from: http://dx.doi.org/10.1016/j.jretconser.2006.07.004

Chisholm, D., and Norman, G. (2012). Spatial competition and market share: an application to motion pictures. *Journal of Cultural Economics.* 36, 207-225. Retrieved April 29, 2013 from: http://dx.doi.org/10.1007/s10824-012-9168-4

Dhar, T., Sun, G., and Weinberg, C. (2012). The long-term box office performance of sequel movies. *Marketing Letters.* 23 (1) 13-29. Retrieved from: http://resolver.scholarsportal.info.ezproxy.lib.ryerson.ca/resolve/09230645/v23i0001/13_tlboposm

Dorta-Gonzales, P., Santos-Penate, D., and Suarez-Vega, R. (2012). Location models and GIS tools for retail site location. *Applied Geography.* 35(1-2), 12-22. Retrieved from: http://resolver.scholarsportal.info.ezproxy .lib.ryerson.ca/resolve/014 36228/v35i1-2/12_lmagtfrsl

Elberse, A., and Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science.* 22(3) 329-354. Retrieved from: http://www.jstor.org/stable/4129744

Elliott, C., and Simmons, R. (2008) Determinants of UK Box Office Success: The Impact of Quality Signals. *Review of Industrial Organization.* 33 (2) 93-111. Retrieved from: http://resolver.scholarsportal.info.ezproxy.lib.ryerson.ca/resolve/0889938x/v33i0002/93_doubostioqs

Fink, E. (2009). The FAQs on Data Transformation. *Communication Monographs.* 76(4) 379-397. Retrieved from: http://dx.doi.org/10.1080/03637750903310352

Gardner, R. (2000). Correlation, Causation, Motivation, and Second Language Acquisition. *Canadian Psychology.* 41 (1) 10-24. Retrieved from: http://dx.doi.org/10.1037/h0086854

Gemser, G., Leenders, M., and Oostrum, M. (2007). The impact of film reviews on the box office performance of art house versus mainstream motion pictures. *Journal of Cultural Economics.* 31 (1) 43-63. Retrieved from: http://resolver.scholarsportal.info.ezproxy.lib.ryerson.ca/resolve/08852545/v31i0001/43_tiofroahvmmp

Goodchild, M. (2010). Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science.* 1 (2010) 3-20. Retrieved from: http://www.josis.org/index.php/josis/article/viewArticle/32

Hennig-Thurau, T., Houston, M., and Sridhar, S. (2006). Can good marketing carry a bad product? Evidence from the motion picture industry. *Marketing Letters.* 17(3), 205-219. Retrieved from: http://dx.doi.org/10.1007/s11002-006-7416-0

Hennig-Thurau, T., Walsh, G., and Wruck, O. (2001). An Investigation into the Factors Determining the Success of Service Innovations: The Case of Motion Pictures. *Academy of Marketing Science Review*. 2001, 1-25. Retrieved from: http://ezproxy.lib.ryerson.ca/login?url=http://search.proquest.com.ezproxy.lib.ryerson.ca/docview/200800316?accountid=13631

Holbrook, M., and Redondo, I. (2010). Modeling the appeal of movie features to demographic segments of theatrical demand. *Journal of Cultural Economics*. 34(4), 299-315. Retrieved April 29, 2013 from: http://resolver.scholarsportal.info.ezproxy.lib.ryerson.ca/resolve/08852545/v34i0004/299_mtaomftdsotd

Johnsson, T. (1992). A procedure for stepwise regression analysis. *Statistical Papers.* 33 (1) 21-29. Retrieved from: http://dx.doi.org/10.1007/BF02925308

Knofczynski, G., and Mundfrom, D. (2008). Sample Sizes When Using Multiple Linear Regression for Prediction. *Educational and Psychological Measurement.* 68 (3) 431-442. Retrieved from: http://dx.doi.org/10.1177/0013164407310131

Lange, C., Lin, X., Naylor, M., Raby, B., and Weiss, S. (2010). Using Canonical Correlation Analysis to Discover Genetic Regulatory Variants. *PLoS One.* 5(5), 1-6. Retrieved from: http://dx.doi.org/10.1371/journal.pone.0010395

Li, S. and Sun, B. (2011). Learning and Acting on Customer Information: A Simulation-Based Demonstration on Service Allocation with Offshore Centers. *Journal of Marketing Research.* 48(1) 72-86. Retrieved from: http://dx.doi.org/10.1509/jmkr.48.1.72

Mach, Q. H., and Grewal, R. S. (2008). Using canonical correlation analysis. *ASQ Six Sigma Forum Magazine, 7*(3), 15-18,12. Retrieved from: http://ezproxy.lib.ryerson.ca/login?url=http://search.proquest.com/docview/213849959?accountid=13631

Mahmoud, M. (2008). Phase I Analysis of Multiple Linear Regression Profiles. *Communications in Statistics- Simulation and Computation.* 37 (10) 2106-2130. Retrieved from: http://dx.doi.org/10.1080/03610910802305017

Manikandan, S. (2010). Data transformation. *Journal of Pharmacology and Pharmacotherapeutics, 1*(2), 126-127. Retrieved from: http://dx.doi.org/10.4103/0976-500X.72373

Marill, K. A. (2004), Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression. *Academic Emergency Medicine*. 11: 94–102. Retrieved from: http://dx.doi.org/10.1197/j.aem.2003.09.006

Mark, D. (2003). Geographic Information Science: Defining the Field. *Foundations of Geographic Information Science. (1)* 3-18. Retrieved from: http://dx.doi.org/10.1201/9780203009543.ch1

MPAA (2012). 2012 Theatrical Market Statistics. *Motion Picture Association of America.* Retrieved from: http://www.mpaa.org/policy/industry

Mukhtar, A. (1987). Durbin-Watson and Generalized Durbin-Watson Tests for Autocorrelations and Randomness. *Journal of Business & Economic Statistics.* 5(2) 195-203. Retrieved from: http://www.jstor.org/stable/1391900

Nathans, L., Nimon, K., and Oswald, F. (2012). Interpreting Multiple Linear Regression: A Guidebook of Variable Importance. 17(9) 1-19. Retrieved from: http://pareonline.net/pdf/v17n9.pdf

O'Brien R. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity.* 41 (5), pg. 673-690. Retrieved from: http://journals1.scholarsportal.info/search-advanced.xqy

Osborne, J. (2002). Notes on the use of data transformations. *Practical Assessment, Research and Evaluation.* 8(6). Retrieved from: http://PAREonline.net/getvn.asp?v=8&n=6

Royne, A. (2008). Cautions and Concerns in Experimental Research on the Consumer Interest. *Journal of Consumer Affairs.* 42 (3) 478-483. Retrieved from: http://dx.doi.org/10.1111/j.1745-6606.2008.00120.x

Salkind, N. (2010) Dichotomous variable. *Encyclopedia of research design.* Thousand Oaks, CA: SAGE Publications, Inc. Retrieved from: http://dx.doi.org/10.4135/9781412961288.n112

Savin, N. and White, K. (1977). The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. *Econometrica.* 45(8) 1989-1996. Retrieved from: http://www.jstor.org/stable/1914122

Stockburger, D.(1998). Multivariate Statistics: Concepts, Models, and Applications. *PSYCHSTAT- Missouri State University*. Retrieved from: http://www.psychstat.missouristate.edu/multibook/mlt00.htm

UCLA Statistical Consulting Group (2013). Chapter 2- Regression Diagnostics. *SPSS Web Books: Regression with SPSS.* Retrieved from: http://www.ats.ucla.edu/stat/spss/webbooks/reg/default.htm