



Swiss Federal Institute of Technology Zurich

Seminar for  
Statistics

Department of Mathematics

---

Master Thesis

Autumn 2018

---

Christopher Salahub

## Seen to be Done

A Graphical Investigation of Peremptory Challenge

---

Submission Date: March 3 2019

---

Co-Adviser:

Adviser: Prof. Dr. Marloes Maathuis



# Preface

I would first like to thank Professor Dr. Marloes Maathuis for her guidance in this project, as it provided invaluable clarity and direction to my work. In addition to the time she set aside for our bi-weekly meetings, her attendance at the group meetings organized by her Ph.D. student Marco Eigenmann was greatly appreciated. These meetings were critical in the development of more nuanced analysis and intuitive visualizations through the discussions they facilitated between me and other students. I thank Marco Eigenmann for organizing them, and Jinzhou Li, Armin Fingerle, Sanzio Monti, and Qikun Xiang for attending my presentations and listening attentively. A special thanks is extended to Cédric Bleutler and Leonard Henckel, both of whom were especially engaged and participated in lengthy discussions both during and outside of these group meetings.

I would also like to acknowledge Professor Dr. Tilman Altwicker for his detailed literature suggestions for additional legal context on peremptory challenges and Professor Dr. Samuel Baumgartner for his research suggestions. The suggested literature provided an excellent starting point for a first investigation of the topic. Of course, without the cooperation of Dr. Ronald Wright, Dr. George Woodworth, Dr. Barbara O'Brien, and Dr. Catherine Grosso, all of whom generously provided me with data, the visualizations which form the core of this work would not have been possible. As such, I am exceptionally grateful that they were so enthusiastic to share the fruits of their labour to help cultivate mine, and hope that they find this work a fitting contribution to the discussion of peremptory challenges.



# Abstract

The legal practice of peremptory challenges is described, outlining its past and present racial controversies as well as the modern defences typically provided in its favour. These arguments are analyzed statistically using novel visual tools including the mobile plot and the positional boxplot, which were developed to explore the impact of race on the exercise of peremptory challenges in three data sets ([Wright, Chavis, and Parks \(2018\)](#), [Grosso and O'Brien \(2012\)](#), and [Baldus, Woodworth, Zuckerman, and Weiner \(2001\)](#)). Multinomial regression models motivated by these visualizations are fit and used to generate precise parameter estimates which indicate the dominance of race in peremptory challenge decisions for venire members across all data sets. Trial level summaries of the data from [Wright et al. \(2018\)](#) are produced and discussed in the context of the results from the venire member models.

## Contents

<b>Notation</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 A Note on Palette Choice . . . . .	2
<b>2 Peremptory Challenges</b>	<b>5</b>
2.1 Jury Selection Procedures . . . . .	5
2.2 The Role of the Jury . . . . .	7
2.3 Modern Peremptory Challenge Controversy . . . . .	8
2.4 The Role of the Peremptory Challenge . . . . .	9
2.5 History . . . . .	10
2.5.1 Pre-English History . . . . .	11
2.5.2 In English Law (1066–1988) . . . . .	11
2.5.3 In American Law (ca. 1700–1986) . . . . .	12
2.5.4 In Canadian Law (ca 1800–2018) . . . . .	13
2.6 Summary . . . . .	14
<b>3 Data</b>	<b>15</b>
3.1 Jury Sunshine Project . . . . .	15
3.1.1 Methodology . . . . .	15
3.1.2 Cleaning . . . . .	16
3.1.3 Variable Synthesis . . . . .	20
3.2 Stubborn Legacy Data . . . . .	20
3.2.1 Methodology . . . . .	20
3.2.2 Cleaning . . . . .	21
3.3 Philadelphia Data . . . . .	21
3.3.1 Methodology . . . . .	21
3.3.2 Cleaning . . . . .	21
<b>4 Analysis</b>	<b>23</b>
4.1 Extremes of Partiality . . . . .	23
4.2 The Impact of Race . . . . .	25
4.2.1 In the Stubborn and Philadelphia Data . . . . .	29
4.3 Other Factors . . . . .	30
4.3.1 Political Affiliation in the Sunshine Data . . . . .	31
4.3.2 Gender in the Sunshine Data . . . . .	32
4.3.3 In the Stubborn and Philadelphia Data . . . . .	34
4.4 Modelling . . . . .	35
4.4.1 Multinomial Logistic Regression . . . . .	35
4.4.2 Examining the Coefficients . . . . .	38
4.5 Trial Level Summary . . . . .	42
4.5.1 Estimating Struck Juror Counts . . . . .	42
4.5.2 Visualizing the Racial Trends . . . . .	43
4.6 On Venire Selection . . . . .	45
<b>5 Summary</b>	<b>49</b>
5.1 Future Work . . . . .	50

---

<b>Bibliography</b>	<b>53</b>
<b>A Developing an Effective Visualization of Conditional Probability</b>	<b>59</b>
A.1 The Mobile Plot . . . . .	61
<b>B Complementary information</b>	<b>63</b>
B.1 Jury Sunshine Irregularities . . . . .	63
B.2 Jury Sunshine Charge Classification . . . . .	63
<b>C Mathematical Results</b>	<b>65</b>
C.1 Conditional Distribution of a Poisson Expectation Given Marginal Counts .	65
<b>D Code</b>	<b>67</b>
D.1 Data Processing Code . . . . .	67
D.2 Analysis Code . . . . .	67
<b>Epilogue</b>	<b>69</b>

## List of Figures

3.1	Charge Tree Example . . . . .	19
4.1	The “Mobile Plot” of Strikes by Racial Combination (Sunshine) . . . . .	27
4.2	Strikes by Racial Combination with Confidence Intervals (Sunshine) . . . . .	28
4.3	Strikes by Racial Combination (All Capital Trial Data) . . . . .	30
4.4	Political Affiliation by Race and Gender (Sunshine) . . . . .	31
4.5	Strikes by Political Affiliation, Race, and Defendant Race (Sunshine) . . . . .	33
4.6	Strikes by Race and Gender (Sunshine) . . . . .	34
4.7	Strikes by Gender and Defendant Gender (Sunshine) . . . . .	35
4.8	Strikes by Race and Gender of Venire Member (All Capital Trial Data) . . . . .	36
4.9	All Model Coefficients . . . . .	39
4.10	Select Model Coefficients . . . . .	40
4.11	Prosecution and Defence Strikes by Trial . . . . .	43
4.12	Racial Strike Counts by Party . . . . .	46
4.13	Racial Strike Proportions by Party . . . . .	47
A.1	Mosaic Plot of Defendant and Venire Member Race . . . . .	60
A.2	First Parallel Coordinate Attempt . . . . .	61
B.1	Regular Expression Charge Tree . . . . .	64



## List of Tables

4.1	Chebyshev Strike Boundaries . . . . .	25
4.2	Strike Rate by Race . . . . .	26
4.3	Nested ANOVA Table Demonstrating the Importance of Race . . . . .	38
4.4	Final Model Coefficients and Confidence Intervals . . . . .	38
B.1	Jury Sunshine Irregularities . . . . .	63



# Notation and Terms

## Terms

In order to facilitate clarity and brevity, a list of terms used in this paper is presented here.

**Prosecution/State** The legal representation which presents the case for the guilt of an individual accused of breaking the law.

**Defence** The legal representation which presents the case for the innocence of an individual accused of breaking the law.

**Accused/Defendant** The individual accused of breaking the law.

**Party** One of the prosecution, defence, or judge.

**Court** All of the judge, prosecution, and defence.

**Jury Roll** A list of individuals in a region eligible to serve on a jury, the construction of these lists varies.

**Venire** The population sample generated using the jury roll from which a jury is selected (according to [Mirriam-Webster \(2019a\)](#) derived from the latin *venire facias*: “may you cause to come”).

**Venire Member** An individual in the venire.

**Jury** The final group of (usually) twelve chosen venire members which judge the guilt or innocence of the defendant.

**Voir dire** From old French “to speak the truth” (see [Mirriam-Webster \(2019b\)](#)), this is the questioning process used by the court to assess the suitability of a venire member to sit on the jury.

**Challenge with Cause** An appeal by the prosecution or defence to remove a venire member from the jury selection process due to a bias which is justified to the court and evaluated by the judge. An unlimited number of these challenges can be used.

**Peremptory Challenge** The privileged removal of a venire member from the jury selection process by the prosecution or defence without any reason articulated, these are limited in number in each jury selection.

**Struck** In the context of a venire member being rejected from the jury, removal by either peremptory challenge or challenge with cause.

**Litigants** The accuser and the accused, in trials with juries the accuser is almost always the government or state.

**Disposition** The outcome of a venire member in the jury selection process: either kept, struck with cause, struck by prosecution, or struck by defence.

## Variables

Across data sets and analyses, the variable names and mathematical notation will be as follows. Note that the use of a capital letter indicates a random variable and a lowercase letter a particular realization of a random variable.

- $\mathbf{x}_i = (r_i, e_i, p_i, g_i, s_i)^T$ : the observed explanatory variable combination for venire member  $i$
- $d \in \{1, 2, 3, 4\}$ : indicator of disposition, with the respective levels kept, struck with cause, struck by defence, and struck by prosecution
- $r \in \{1, 2, 3\}$ : indicator of venire member race, with respective levels black, other, and white
- $e \in \{1, 2, 3\}$ : indicator of defendant race, with levels as for the venire member race
- $p \in \{1, 2, 3, 4\}$ : indicator of venire member political affiliation, with respective levels Democrat, Independent, Libertarian, and Republican
- $g \in \{1, 2\}$ : indicator of venire member gender, with respective levels female and male
- $s \in \{1, 2\}$ : indicator of defendant gender, with levels as for the venire member
- $\pi_{d|jklmn} \in [0, 1]$ : the probability of disposition  $d$  given factor levels  $r = j, e = k, p = l, g = m, s = n$ , may be written as  $\pi_d$  for convenience or given a superscript  $(i)$  to indicate this probability for venire member  $i$
- $y_{djklmn} \in \mathbb{N}$ : the count of venire members with  $\mathbf{x}_i = (j, k, l, m, n)^T$  and disposition  $d = d_i$

This work also uses hat notation for estimates (i.e. the estimate for  $\pi$  is  $\hat{\pi}$  and the estimator for  $\pi$  is  $\tilde{\pi}$ ).

# Chapter 1

## Introduction

The Gerald Stanley murder trial, officially *R. v. Stanley*, was noteworthy for all of the wrong reasons. The first reason was the crime itself. The rural region around Biggar, Saskatchewan [Quenneville (2018)] is not known for crime. Indeed, the crime statistics collected by Statistics Canada suggest it is one of the safest in the province [Statistics Canada (2018)]. Any murder at all would be worthy of attention and subject to plenty of drama. But beyond the damage this trial has done to the community, it was noteworthy because it led to a significant re-examination of the legal jurisprudence surrounding the jury selection process in all of Canada. The case’s controversy culminated in the proposition of Bill C-75 by the Canadian government in March of 2018 [42nd Parliament of Canada (2018a)], less than two months after the trial’s verdict [Quenneville and Warick (2018b)].

Bill C-75, in part, aims to ameliorate one of the critical points of contention in the Gerald Stanley case: the use of peremptory challenges in jury selection. The outsized impact of the case was due, in large part, to the case’s racial aspect. Gerald Stanley, a white man, was accused of second degree murder in the killing of Colten Boushie, a First Nations man. Given Canada’s troubled history with First Nations groups, this alone would have been enough to make the trial a flash point for race issues, but that was not the worst aspect of the trial. Rather, the most controversial and influential facet of the entire affair was the alleged use of peremptory challenges to strike five potential jurors who “appeared” to be First Nations, resulting in an all-white jury [Harris (2018), MacLean (2018)].

With Bill C-75 currently moving through the Canadian parliamentary system, having completed its second reading in June 2018 [42nd Parliament of Canada (2018b)], an evaluation of the practice of peremptory challenge is warranted. A great deal of ink has already been spilled on both sides of the debate (see Hasan (2018), Zinchuk (2018), and Roach (2018)), but startlingly little of this discussion has been based on any hard, quantitative evidence on the impact of peremptory challenge in jury selection. This paper aims to provide analysis and evidence to illuminate the topic further by analyzing three separate peremptory challenge data sets collected in the United States, namely the data from Wright et al. (2018), Grosso and O’Brien (2012), and Baldus et al. (2001), henceforth referred to as the “Sunshine,” “Stubborn,” and “Philadelphia” data sets respectively. While this data cannot reveal anything about the alleged racial motivation of peremptory challenge use in *R. v. Stanley*, wider view of the practice is a more sober place to assess its role in modern jury trials than the dissection of a particular controversial case.

Of course, this work is not the first such investigation. Wright et al. (2018), Grosso and

O'Brien (2012), and Baldus et al. (2001) have performed analysis on the factors which impact the use of peremptory challenges in their respective data sets. All of these investigations indicated that race was an important factor in determining if a venire member was struck. Numerous others have performed unique legal, empirical, and analytical analyses of the jury selection process, including Hoffman (1997), Van Dyke (1977), Hans and Vidmar (1986), Brown, McGuire, and Winters (1978), and Ford (2010). Most of the authors which have performed such analysis arrive at similar conclusions on the general importance of race in the exercise of peremptory challenges, and the negative impact this has on the operation and perception of justice in the legal system. Hoffman (1997) gives an exceptionally negative analysis of peremptory challenges from a legal perspective, while the game theory analysis of Ford (2010) suggests that the use of peremptory challenges may even be counter-productive.

What is, perhaps crucially, missing from this rich analysis is an effective method of communicating these results. While the tables generated to summarize the previous analyses certainly contain all the data necessary to evaluate strike patterns, they fail to be accessible to a casual reader, as they require some degree of commitment and focus to interpret and compare. Visual representations of the data which could be used for such quick comparison and interpretation would facilitate dissemination of the empirical results of these analyses to a broader audience, and would make the work of comparing and interpreting data sets far more intuitive than the current table representations. This work endeavours to provide such visual tools.

Consequently, this work proceeds in four parts. Chapter 2 provides the necessary legal context to understand the motivation of the previous investigations. In 2.1, the general jury selection procedure is presented before the modern controversies of this process are outlined in 2.3. Legal arguments for both the jury and the peremptory challenge are provided interspersed in this modern history in 2.2 and 2.4. After the modern description, a brief history of the practice of peremptory challenges in jury trials is presented in 2.5, in particular explaining the original motivation of the practice, its past implementations, and its development in the United States, the England, and Canada.

With the necessary context provided, chapter 3 proceeds to discuss the three data sets obtained, explaining the sources and collection methods before detailing cleaning and pre-processing. Chapter 4 then provides the details and results of the analysis performed on the different data sets. It begins by performing statistical analysis of one common argument in favour of peremptory challenge in 4.1 before visualizing the Sunshine data in 4.2 and 4.3. Mobile plots (see A) are the primary tool used for this visual analysis of the data, and every visualization of the Sunshine data set is compared to analogous visualizations of the Stubborn and Philadelphia data sets. The implications of their similarities for generalization are discussed. These visual analyses are then used to motivate model selection in 4.4 in order to estimate more precisely the impact of race in the Sunshine data. These results and findings are summarized in chapter 5. Recommendations based on the observations obtained are provided alongside suggestions for future work.

## 1.1 A Note on Palette Choice

The analysis and presentation of results in this paper is primarily visual, utilizing graphs and figures rather than tables to communicate patterns and estimates. In order to make

---

these visual presentations of the data as accessible as possible, the colours and palettes used were very deliberately chosen to be distinguishable for as many individuals as possible, including colour blind individuals. In this endeavour, the `RColorBrewer` package in R [Neuwirth (2014)] and Wong (2011) were indispensable, as both provide suggested colour-blind safe palettes and colours. Additionally, most factors encoded by colours are redundantly encoded by position or order where possible.





## Chapter 2

# Peremptory Challenges

As the practice of peremptory challenges in a jury trial system is a highly specific procedure which may be unfamiliar to the reader, a brief exploration of the history, motivation, and current use of peremptory challenges is presented here. It is not exhaustive, but rather explains the terms used and the process of peremptory challenges generally. The references provided throughout are an excellent starting point for interested and motivated readers hoping to learn more.

### 2.1 Jury Selection Procedures

While the process of jury selection varies by jurisdiction and crime severity, the general steps of jury selection shared by the vast majority of jury trials are outlined below. More detail and a discussion of the diversity of jury selection procedures can be found in [Ford \(2010\)](#), [Hans and Vidmar \(1986\)](#), and [Van Dyke \(1977\)](#). To select a jury:

- i.) Eligible individuals are selected at random from the population of the region surrounding the location of the crime using a list called the *jury roll*, the sampled individuals are called the *venire*
- ii.) The venire is presented to the court, either all at once or sequentially (borrowing the names of [Ford \(2010\)](#): the “struck-jury” system and the “sequential-selection” system, respectively)
- iii.) The presented venire member(s) are questioned in a process called *voir dire*, after which there are three possible outcomes for each venire member:
  - (a) The venire member is removed with cause, the cause provided by either the prosecution or defence and admitted by the judge
  - (b) The venire member is removed by a *peremptory challenge* by the prosecution or defence, where no reason need be provided to the court; such privileged rejections of a venire member are limited in number for both lawyers (in Canada a maximum of 20 such challenges per side per defendant are allowed [[Government of Canada \(1985\)](#)])
  - (c) The venire member is accepted into the jury, and so becomes a juror

- iv.) Steps i-iii are repeated until the desired number of venire members have been accepted into the jury, typically 12.

As mentioned above, the details of this process can vary greatly by region. One of the greatest sources of variation is the creation of jury rolls. In the United States the method is somewhat homogeneous, they are typically selected using lists of registered voters (see chapter two of [Van Dyke \(1977\)](#) and page 53 in [Hans and Vidmar \(1986\)](#)), but in Canada their creation is less consistent. Ontario uses a combination of municipal voter lists and First Nations band lists [[Ministry of the Attorney General of Ontario \(2018\)](#)], while in Saskatchewan - the province of *R. v. Stanley* - the jury roll is created from provincial government health insurance data in accordance with the [Government of Saskatchewan \(1998\)](#) guidelines.

Clearly, the variation in these methods will create differences in the coverage of the the population the jury rolls are meant to reflect. Such differences are no doubt important to the composition of resulting juries<sup>1</sup>, but these differences were not the main criticism of *R. v. Stanley*, and are not affected by Bill C-75. As discussed in Chapter 1, peremptory challenges have proven to be of greater interest.

While the exercise of peremptory challenges differs as well, notably between the struck-jury and sequential-selection systems of voir dire, [Ford \(2010\)](#) and [Van Dyke \(1977\)](#) note that the predominant method in the United States and Canada is the sequential-selection system. This is perhaps due to the relative efficiency of the method, as in the sequential system voir dire need not be performed on the entire venire, only for a subset. Contrast this with the struck-jury system, where the entire venire must be reviewed in every trial.

Another source of variation in the exercise of peremptory challenges is the scope of voir dire. The specificity of permitted questions is radically different in the United States and much of the British Commonwealth. [Van Dyke \(1977\)](#) notes on page 143 that Canada and England do not allow questions in areas of “non-specific” bias, or bias which is not directly related to the case before the court. That is to say, while it would be perfectly valid to ask a venire member in a murder case about their work history in the United States for any case, such a question would only be allowed in Canada or England if occupation was specifically related to the crime.

This difference in procedure creates a far greater emphasis on the voir dire process and peremptory challenges in the United States, as noted by [Hans and Vidmar \(1986\)](#). They surmise that the key reason for this marked departure in procedure is a difference in philosophy. To borrow a quote from page 63:

In Canada... the courts have said that we must start with an initial presumption that “a juror will perform his duties in accordance with his oath”

This doctrine places a responsibility on the jurors themselves to overcome their biases and accept arguments in spite of them. This stands in stark contrast to the American attitude implied by expansive voir dire: that certain prejudice cannot be overcome by jurors themselves and thus peremptory challenges are necessary to ensure that biased individuals are not included on the jury. The public statements of the *R. v. Stanley* verdict critics indicate that they subscribe to this viewpoint more than to the guiding Canadian legal philosophy.

---

<sup>1</sup>See [Iacobucci \(2013\)](#) for a detailed report on the implications of these coverage issues for First Nations groups in Canada.

## 2.2 The Role of the Jury

Such a difference in viewpoint is especially relevant given the purpose of the jury. The central function of a jury is to judge the innocence or guilt of an accused in light of the presented evidence, a function which has had drastically different forms throughout history. In the distant past, [von Moschzisker \(1921\)](#) and [Hoffman \(1997\)](#) report that juries primarily acted to collect evidence and evaluate whether it warranted further legal action, essentially assuming the role commonly performed by police departments today. Such a role justified the archaic practice of forming select juries of only the most “trustworthy” individuals.

This is contrasted by the modern jury, which performs no collection of evidence and is representative rather than selective. It is, ideally, a panel of peers or “equals” of the accused taken from the community near the crime, an idea which did not develop until 19th century England (see page 28 of [Hans and Vidmar \(1986\)](#)) and was not applied using random sampling until some time later (see [Hoffman \(1997\)](#), page 29 of [Hans and Vidmar \(1986\)](#), and page 16 of [Van Dyke \(1977\)](#)). The modern jury is meant to apply the law, as told to them by the judge<sup>2</sup>, to the case at hand. Evidence of the guilt of the accused is presented to the jury by the prosecutor, while evidence meant to exonerate is presented by the defence.

The jury listens to the evidence, considers the law as communicated by the judge, and must (typically) reach a unanimous decision of guilt or acquittal. Such a decision cannot be overturned by the judge of the court, and the judge must then determine sentencing based on the decision of the jury and the letter of the law<sup>2</sup>. The jury therefore has tremendous power in the judgement of any case. The philosophical and ethical justification for such power is well explained by [Woolley \(2018\)](#), and best summarized by a quote from the [Supreme Court of Canada \(1991\)](#):

The jury, through its collective decision making, is an excellent fact finder; due to its representative character, it acts as the conscience of the community; the jury can act as the final bulwark against oppressive laws or their enforcement; it provides a means whereby the public increases its knowledge of the criminal justice system and it increases, through the involvement of the public, societal trust in the system as a whole.

While such enthusiastic support for juries has not been explicitly expressed by all countries which practice them, the justification is entirely consistent with the histories and analysis presented by [Hoffman \(1997\)](#), [von Moschzisker \(1921\)](#), [Hans and Vidmar \(1986\)](#), [Van Dyke \(1977\)](#), and others. This suggests that the [Supreme Court of Canada \(1991\)](#) lionization of the jury system is a fair representation of the perceived role of the jury throughout those countries which use them, and motivates the importance of choosing juries which are consistent with these principles through some jury selection process.

---

<sup>2</sup>[Hans and Vidmar \(1986\)](#) note that this system actually varies throughout the United States, though the jury and judge powers described here are consistent across Canada.

## 2.3 Modern Peremptory Challenge Controversy

If the general utility and importance of the jury is clear, the same cannot be said for peremptory challenges. The privileged removal of a venire member<sup>3</sup> without any justification has seen persistent allegations of abuse, often around the use of these challenges by state prosecutors.

In the United States, the criticism has focused on racial discrimination, and has led to significant changes in their allowed use, through cases such as *Swain v. Alabama* (Supreme Court of the United States (1965)) and *Batson v. Kentucky* (Supreme Court of the United States (1986)). The first of these cases, *Swain v. Alabama*, established in 1965 that the systematic exclusion of venire members of a particular race would be unconstitutional discrimination under the Fourteenth Amendment to the United States Constitution, but argued that a *prima facie* (or “based on first impression”) argument of discrimination was not adequate to prove this<sup>4</sup>. This placed a significant burden on the party taking issue with a particular peremptory challenge to demonstrate that the specific challenge had been discriminatory.

However, this ruling was overturned only 21 years later in the 1986 case *Batson v. Kentucky*, which allowed the party objecting to a challenge to use a *prima facie* argument which must be countered by a race-neutral reason that satisfies the judge. If no such reason can be supplied, the challenge would not be allowed. This created a new challenge which could be used to nullify a peremptory challenge: the so-called “Batson Challenge”. While the effectiveness of this system of additional challenges is questionable both practically and in abstract (see Page (2005) and Morehead (1994), and a particularly strong response in Hoffman (1997)), it has only been extended to allow Batson Challenges for both the sex and race of venire members<sup>5</sup>.

Echoes of such racial controversies have also been present in Canada before *R. v. Stanley*. Racial bias against First Nations venire members in Manitoba was alleged in 1991 in a report produced after an inquiry by the provincial government [Roach (2018)]. More damning still was the Iacobucci Report on First Nations representation in juries. This report proposed an explicit restriction to the practice when it recommended:

an amendment to the Criminal Code that would prevent the use of peremptory challenges to discriminate against First Nations people serving on juries.

These controversies also led to a great deal of academic investigation of the practice of peremptories. Legal analyses have been presented by many, including Hoffman (1997), Broderick (1992), and Nunn (1993), and the large majority of these analyses take a negative view of the peremptory challenge as it currently stands. They typically either recommend large modifications to the system beyond the Batson Challenge or the abolition of the practice altogether in the selection of juries.

<sup>3</sup>To be replaced by another randomly selected venire member.

<sup>4</sup>In the actual case, not a single black juror had sat on a jury in Kentucky in the previous 15 years, despite composing 26% of the jury-eligible population. In Swain’s trial, six of the eight black venire members were rejected by state prosecutor peremptory challenges, and the other two removed for cause, leaving not a single black juror to judge Swain, a black man. This was the *prima facie* argument presented by Swain’s defence team against the state prosecutors of Alabama, and it was rejected as insufficient to prove discrimination.

<sup>5</sup>The use of Batson Challenges for sex was established in *J.E.B. v. Alabama* [Supreme Court of the United States (1993)].

These legal analyses have been complemented by theoretical explorations by [Ford \(2010\)](#) and [Flanagan \(2015\)](#) using game theory. Both of these studies indicate that the current system of peremptory challenges may produce juries which are biased towards conviction or acquittal and may include a higher proportion of extremely biased members of the population. The implication is that the current system is more useful for the purpose of “stacking” a jury to be favourable to one side, that is increasing the proportion of jurors sympathetic to defence or prosecution arguments<sup>6</sup>.

Even more relevantly to this work are the empirical analyses performed in [Baldus et al. \(2001\)](#), [Wright et al. \(2018\)](#), [Grosso and O’Brien \(2012\)](#), [Baldus, Grosso, Dunham, Woodworth, and Newell \(2012\)](#), and many others. These have universally found illicit factors such as race to be significant in the exercise of peremptory challenges. This is both in aggregate and when possible confounding factors are controlled using logistic regression or contingency tables. Such findings lend credence to those who view the controversies surrounding the peremptory challenge as justified responses to a broken system.

Despite the preponderance of negative analysis, there is no large political movement in the United States to remove the practice. Furthermore, there had not been a significant political effort to reform the Canadian peremptory challenge system until the furore around *R. v. Stanley* culminated in the tabling of Bill C-75 [[42nd Parliament of Canada \(2018b\)](#)], which would abolish the peremptory challenge in Canada outright. As of the time of writing, the bill has not been approved by the Government of Canada, but it seems likely to become law in the near future. In doing so Canada would join England, which abolished the practice in the Criminal Justice Act of 1988 after the controversial Cyprus spy case in the late 1970s, which led to a “sustained campaign in Parliament and in the press alleging that defence counsel were systematically abusing it” [[Hoffman \(1997\)](#)]<sup>7</sup>.

## 2.4 The Role of the Peremptory Challenge

Despite the legal changes, recommendations, and a great deal of articles providing analysis against the practice, the topic of the peremptory challenge remains controversial in the United States and Canada, and is defended as a key component of the jury selection process by some. The modern defence is perhaps best described by Justice Byron R. White in [Supreme Court of the United States \(1965\)](#):

The function of the challenge is not only to eliminate extremes of partiality on both sides, but to assure the parties that the jurors before whom they try the case will decide on the basis of the evidence placed before them, and not otherwise. In this way, the peremptory satisfies the rule that, “to perform its high function in the best way, justice must satisfy the appearance of justice.”

Such a justification is reminiscent of the now famous words of Lord Chief Justice Hewart in *R. v. Sussex Justices* in 1924: “Justice should not only be done, but should manifestly and undoubtedly be seen to be done” (as reported in [Richardson Oakes and Davies \(2016\)](#)).

<sup>6</sup>In chapter 6 of [Hans and Vidmar \(1986\)](#), the “science” of using peremptory challenges to construct a biased jury is described in great detail for the case of *M.C.I. Communications v. American Telephone and Telegraph*.

<sup>7</sup>It should be noted that this did not abolish the use of “standing-aside” by the Crown, although the practice was restricted to national security trials and heavily curtailed, with strict guidelines to its use outlined by the [Attorney General’s Office of the United Kingdom \(2012\)](#).

While these words originally only referred to the pecuniary interest of court staff involved in the case, they have since come to express the idealized expectation that both the defence and prosecution find the judge and jury acceptable, as explored by [Richardson Oakes and Davies \(2016\)](#)<sup>8</sup>.

This defence suggests two modern justifications for the peremptory challenge. The first is that of removing venire members with “extreme” bias, and the second is the creation of a jury which is composed of jurors mutually acceptable to both the defence and the prosecution. Those who defended the practice of peremptory challenges in Canada after *R. v. Stanley*, including [Hasan \(2018\)](#) and [Macnab \(2018\)](#), seem to use this defence or some variant of it to argue in favour of keeping the practice.

That these articles were written in response to the upset which followed *R. v. Stanley* serves as a counter-argument to the assertion that the exercise of peremptory challenges creates an acceptable jury. Such reasoning fails to account for the impact of removing an unbiased juror to both the perception of justice and the composition of the final jury. Rather, it focuses singularly on the inclusion of a biased juror as the only possible cause of an unacceptable jury. Such a narrow view cannot realistically be held in light of the decisions of *Batson v. Kentucky* and *J.E.B. v. Alabama*, which implicitly acknowledge the corrosive nature of unjustified strikes to the core principles of an unbiased jury of peers.

Additionally, as the purpose of challenges with cause is to remove jurors with a bias that can be articulated, one is left to wonder what exactly forms the basis of the prosecution and defence exercise of peremptories. Investigations by [Baldus et al. \(2001\)](#), [Wright et al. \(2018\)](#), [Grosso and O’Brien \(2012\)](#), and others have all found that there are significant racial differences between venire members removed by peremptory challenges and those kept, even when other possible confounders are controlled. It is possible this observed aggregate discrimination is a manifestation of the inability of lawyers to articulate the specific biases they detect<sup>9</sup>, and so perhaps a comparison of the use of peremptory challenges to challenges with cause, a topic not addressed in detail by [Baldus et al. \(2001\)](#), [Wright et al. \(2018\)](#), or [Grosso and O’Brien \(2012\)](#), is also warranted.

## 2.5 History

An analysis of peremptory challenges most appropriately begins with a historical exploration of the peremptory challenge. Roughly, the presentation of the history of jury trials here follows the comprehensive and exhaustively referenced description provided by [Hoffman \(1997\)](#). Two of the references [Hoffman](#) uses extensively, [Hans and Vidmar \(1986\)](#) and [Van Dyke \(1977\)](#), provided useful context while specific details provided by [von Moschzisker \(1921\)](#), [Forsyth \(1994\)](#), [Brown et al. \(1978\)](#), and [Brown \(2000\)](#) helped to create a clearer picture of particular periods of jury history. Information regarding the history of the Canadian system was provided by [Brown \(2000\)](#) and [Petersen \(1993\)](#). For an excellent exploration of the nineteenth century, a formative time for the development of challenges in case law, see [Brown \(2000\)](#).

<sup>8</sup>Such grand generalizations and myth-making can also be seen in the common belief that the right to a trial by jury was originally established in the Magna Carta, an idea which is not supported by the relevant historical evidence (see [Hoffman \(1997\)](#) and [Van Dyke \(1977\)](#) for a detailed discussion and more accurate history).

<sup>9</sup>A weak argument given that articulation is the speciality of the legal profession.



### 2.5.1 Pre-English History

Although precise timelines are hard to establish, there is evidence that jury trials have occurred in some form or another since antiquity. The concept, that of judgement by a group of peers, is so ancient that it is prevalent not only in historical records, but in myth. As [Hoffman \(1997\)](#) indicates, both Norse and Greek mythology feature groups of individuals assessing the guilt or collecting evidence about the actions of a peer.

Outside of the realm of myth, [Hoffman \(1997\)](#) reports that there is evidence of the use of juries in Ancient Egypt, Mycenae, Druid England, Greece, Rome, Viking Scandanavia, the Holy Roman Empire, and Saracen Jerusalem. It should be noted that in none of these areas was the jury trial the primary form of conflict resolution practiced. Nonetheless, it is clear the jury trial has a broad and long history of use.

Something similar to the modern peremptory challenge does not appear until Rome, however. The Roman *Judices* were groups of senators selected to judge the guilt of the accused in a legal case. According to [Hoffman \(1997\)](#), 81 Senators would be chosen to sit on one of these *Judices*, after which the litigants were permitted to remove 15 of these Senators each. This egalitarian reduction of the jury size seems analogous to the modern peremptory challenge system, as it places the power of removal with the litigant and suggests no justification is necessary for their removal.

### 2.5.2 In English Law (1066–1988)

The peremptory challenge did not reach its modern form, as outlined in [2.1](#), until it was established in the English legal system. Despite some previous debate on the topic, the most modern historical evidence suggests that the basis of the English practice was not related to the system used in the selection of *Judices* in Rome.

Rather, the dominant historical interpretation is presented by [von Moschzisker \(1921\)](#) and [Hoffman \(1997\)](#): that the jury system was introduced to England during the Norman conquest of 1066 by William the Conqueror. The practice, however, was not made official until the Assize of Clarendon in 1166 by Henry II, and it was not until the abolition of trials by ordeal<sup>10</sup> in 1215, that peremptory challenges began to appear in England. These challenges were officially recognized in 1305 when Parliament outlawed their use by the Crown, only to replace them with an analogous system of so-called “standing-aside”<sup>11</sup>.

It should be noted here that although the challenges issued between the Assize of Clarendon and this 1305 act are called “peremptory,” they may not have served the same purpose, nor shared the same justification, as the modern challenges. As [Hoffman \(1997\)](#) argues convincingly, these challenges may have been closer to modern challenges with cause. The argument hinges on the paradigm of royal infallibility and absolutism which was present in the late medieval period when the peremptory challenge first appeared [[Burgess \(1992\)](#)].

Under royal absolutism and infallibility the argument for peremptory challenges is quite simple. If the king cannot be wrong in his judgement and he has some reason to believe that a venire member cannot serve on the jury, then he need not say why he thinks that is so, as his judgement is correct in any case. Indeed, asking for an explanation would be

---

<sup>10</sup>The most common method of trial at that time.

<sup>11</sup>For a detailed explanation of this system see [Hoffman \(1997\)](#) and [Brown \(2000\)](#).

disrespectful and providing one undignified. The Crown prosecutors, as representatives of the king, would be similarly shielded from criticism.

Such an argument is further supported by the abolition of their royal use in 1305, the language of which suggests that peremptory challenges were originally the privilege of the Crown (see [Hoffman \(1997\)](#) and page 147 in [Van Dyke \(1977\)](#)), with none being granted to the defence. [Hoffman \(1997\)](#) suggests that as royal infallibility grew out of favour, the desire to make the legal process more equitable resulted in the granting of peremptory challenges to the defence rather than their removal from the jury selection system.

Whatever the original logic of the expansion of these challenges to the defence, their legal limits are recorded more precisely<sup>12</sup>. From a maximum of 35 challenges allowed at their peak in the fourteenth century, the number of challenges allowed only decreased over time until their abolition in 1988 (discussed in [2.4](#)).

### 2.5.3 In American Law (ca. 1700–1986)

[von Moschzisker \(1921\)](#), [Hoffman \(1997\)](#), and [Van Dyke \(1977\)](#) all agree that the early English colonists that came to North America accepted the jury system with peremptory challenges as common law well before the establishment of the United States of America. [Hans and Vidmar \(1986\)](#) note, however, that the difficulty of ocean travel and the overall indifference of appointed Crown representatives in the colonies led to an increased importance of the jury trial and the role of challenges to these early colonists as a way to exercise some degree of community control in the face of laws drafted in a distant country and implemented by unsympathetic authorities<sup>13</sup>.

It is somewhat interesting, then, that the United States constitution makes no mention of the practice of peremptory challenges. The Sixth and Seventh Amendments specify a great deal of the jury system, including the right to public defence and an impartial jury drawn from the district of the crime, but make no mention of a right to the exercise of peremptory challenges, or any challenges whatsoever (see [Constitution of the United States \(1788\)](#)).

As [Hans and Vidmar \(1986\)](#) report on page 37, an original draft of the Sixth Amendment expressly included challenges for cause, but the debate around their inclusion resulted in the removal of this clause. They continue to say that at the time, even some proponents of the challenge considered the reference unnecessary, as the practice was implied by the text which remained, referring to a trial by an “impartial” jury. Another result of these debates was the adoption of the extensive voir dire process which allows questions of general bias<sup>14</sup>.

Critically, there appears to have been no discussion around the inclusion of peremptory challenges (see page 37 of [Hans and Vidmar \(1986\)](#) and [Hoffman \(1997\)](#)). Despite the clear importance of the jury trial to the drafters of these amendments, it would seem the

<sup>12</sup>See [Brown \(2000\)](#) for a detailed examination of the case law developing around challenges in the nineteenth century.

<sup>13</sup>For more detail on this development among the early colonists, it is instructive to read about the Zenger trial of 1734 (described on pages 33-35 of [Hans and Vidmar \(1986\)](#)). Not only does this trial reveal a great deal about the attitudes of the colonists at the time, but it also presents the idea of a jury assessing guilt and “wrongness” using their own conscience rather than just settling fact. The precept of the modern jury trial in Canada (see [Woolley \(2018\)](#)) is based on this very idea.

<sup>14</sup>This is described on page 37-38 of [Hans and Vidmar \(1986\)](#), though [Brown \(2000\)](#) notes that the 1807 Burr trial was also highly significant in the development of general voir dire in the United States.



peremptory challenge was not considered to have anywhere near the same significance as judgement by an impartial jury of local peers<sup>15</sup>.

Regardless of this, as [Brown \(2000\)](#) notes, the importance and use of challenges increased in the United States in the nineteenth century following American independence due to a desire to prevent the tyranny of the state. This desire also led to the adoption of a limited number of peremptory challenges for the prosecution, rather than the possibly unlimited stand-asides that were allowed under British law to prosecutors (see [Van Dyke \(1977\)](#), page 150).

While the specific numbers of peremptory challenges allowed to both sides and the required motivation of challenges for cause have varied over time [[Hoffman \(1997\)](#), [Brown \(2000\)](#)], they have remained a feature of the American legal system, and numerous Supreme court cases have merely served to make the use of challenges more specific and codified [[Hoffman \(1997\)](#)]. It was not until *Batson v. Kentucky* in 1986 that this system of challenges was drastically changed with the introduction of Batson challenges (described in [2.3](#)).

#### 2.5.4 In Canadian Law (ca 1800–2018)

Canadian law, inspired by a close relationship to both the British Crown and the United States, seems to have adopted elements of both legal systems in its development of peremptory challenges in the nineteenth century. As discussed by [Brown \(2000\)](#), Canada adopted the American practice of replacing prosecutorial stand-asides in favour of a more egalitarian limited number of peremptory challenges to both sides. Despite this, the Canadian voir dire process remains limited and much more similar to the British one, as does the system of challenges for cause (see page 48 of [Hans and Vidmar \(1986\)](#)).

One perfect demonstration of this departure is the Canadian constitution. As in the United States, the Canadian constitution fails to mention challenges. The British North America Act of 1867 [[Constitution of Canada \(1982\)](#)], which established Canada's independence from England, makes no mention of legal rights of the accused, indicating a deference to legal precedent in England. It is not until the Charter of Rights and Freedoms in 1982<sup>16</sup> that such rights were guaranteed in a legal Canadian document. Notably, its language is considerably more vague than the United States Sixth and Seventh Amendments, guaranteeing only “the benefit of trial by jury” [[Constitution of Canada \(1982\)](#)].

This “eclectic” incorporation of both American and English case law, to borrow the term used by [Brown \(2000\)](#), led to a system somewhere between the English and American systems, but decidedly closer in operation to the English system. It should be noted, however, that as Canada grew more populous in the twentieth century and developed a greater legal precedent and more experienced judges of its own, this reliance upon its former colonial master and its more powerful southern neighbour seems to have diminished in importance.

---

<sup>15</sup>Indeed, as *Batson v. Kentucky* and *Swain v. Alabama* have both shown ([Supreme Court of the United States \(1986\)](#) and [Supreme Court of the United States \(1965\)](#)), the modern interpretation of “impartial” may preclude the use of peremptory challenges altogether.

<sup>16</sup>This was the year of patriation of the Canadian constitution. As independence was granted by the British Parliament, the British North America Act outlining Canada's laws was a British law and changing it was the prerogative of the British Parliament rather than the Canadian one. It was not until the Constitution Act of 1982 that the Canadian constitution became a Canadian law. For a more detailed history see [Sheppard \(2018\)](#).

As a result, the mechanics of the peremptory challenge have not changed despite the abolition of the practice in England and the introduction of the Batson Challenge in the United States.

## 2.6 Summary

The peremptory challenge, a practice of much controversy in the English-speaking world, seems to have started in its modern form as a privilege of the King of England in the thirteenth century. After its conception, it spread with English conquest and colonization, with new colonies and local governments accepting the practice based primarily on the adoption of English legal precedent. Though it was abolished in England in 1988, it remains a fixture of American jury trials, and is accompanied there by a thorough and invasive voir dire process which is not seen in Canada nor England.

Though the practice has historical longevity, it is not guaranteed by the constitutions of Canada or the United States, and has been a practice of considerable legal debate and significant change throughout its history. In England this culminated in the Cyprus spy trial, in the United States in *Batson v. Kentucky* and *Swain v. Alabama*, and in Canada in *R. v. Stanley*: the Gerald Stanley murder trial. As a consequence, the broad agreement of the importance and propriety of a jury has conferred little consensus on the place of peremptory challenges in the selection of juries.

Indeed, it seems increasingly impossible for the jury to function in a way consistent with its demanding ideals with the peremptory challenge still present. Its spotted history and abuse to exclude certain minorities may undermine its purported use as a tool to ensure the acceptance of a trial's outcome by both litigants. The three court cases mentioned above are a demonstration of how the peremptory challenge can be used to create a jury which is unacceptable to one litigant in a case. This suggests that any argument which relies upon the mutual acceptance of a jury by all parties in the court is fundamentally flawed, as it fails to account for the removal of venire members as a source of contention equal in measure to that of keeping a biased juror.

The second argument in favour of the peremptory challenge, that of removing the extremely biased jurors, fares little better in light of the controversies, legal analyses, theoretical modelling, and empirical studies outlined in 2.3. That the practice has been gradually curtailed in the countries which practice it or removed entirely suggests that it may not be functioning to remove only biased venire members. Rather, there is some possibility that it is also removing potentially fair jurors.

## Chapter 3

# Data

Without data, performing an analysis that incorporated more than the history and legal argumentation presented in Chapter 2 is impossible. This proved problematic. While the motivation of this text was a Canadian case, no comprehensive data sets which examined jury selection in Canada could be found. The increased prominence of the jury selection process in the United States garnered a more fruitful search.

The author is heavily indebted to [Wright et al.](#); [Grosso and O'Brien](#); and [Baldus et al.](#). These authors shared their data freely with the author, providing him with a wealth of data to analyse empirically. As a consequence of the multiple separate data sets, however, care must be taken to describe each of the data sets separately in order to capture adequately the different methodologies and sources they represent. As [Wright et al. \(2018\)](#) notes:

limited public access to court data reinforces the single-case focus of the legal doctrines related to jury selection. Poor access to records is the single largest reason why jury selection cannot ... become a normal topic for political debate

Currently, the collection of jury data is difficult. Many courtrooms have not digitized past records and concerns over privacy limit the release of those records, which are stored as paper documents in the case file (see [Wright et al. \(2018\)](#)). This limits the ability of investigators to ask for summaries across numerous trials or to view the jury selection process on a scale beyond the basis of one case. Thus, to gather aggregate data the authors of these papers necessarily used different collection techniques dictated by the scope of collection desired and the procedures of the court systems from which data was collected.

### 3.1 Jury Sunshine Project

#### 3.1.1 Methodology

The Jury Sunshine Project [[Wright et al. \(2018\)](#)], so named as it was carried out in order to shed light on the jury selection process, is the most extensive data set which was provided to the author. It endeavoured to collect jury data for all felony trial cases in North Carolina in the year 2011, which ultimately resulted in a data set that detailed the

simple demographic characteristics and trial information of 29,624 individuals summoned for jury duty in 1,306 trials. Note that not all entries were complete.

Due to the scope of the project, there are a number of problems which had to be solved by the authors. The first of these was simply identifying which court cases went to trial in 2011, in order to direct resources effectively. This was accomplished by downloading publicly available case data from the North Carolina Administrative Office of the Courts (NCAOC)<sup>1</sup> and determining the case numbers and counties of cases which went to trial. Wright et al. state that this likely missed some cases, but that they were confident that a “strong majority” of trials was collected, which did not systematically differ from those excluded.

This list was then used to perform a pilot study to refine recording practices before undertaking a more general survey where “law students, law librarians, and undergraduate students” (called *collectors* for convenience) visited court clerk offices to collect the relevant case data, including the presiding judge, prosecutor, defence lawyer, defendant, venire members, charges, verdict, and sentence [Wright et al. (2018)]. The case files also included data about whether a venire member was removed by cause or peremptorily, and the party which challenged in the peremptory case. Using public voter databases, bar admission records, and judge appointment records, these collectors were able to determine demographic (race, gender, and date of birth) and political affiliation data for the venire members, lawyers, defendants, and judges. This data set was stored in a relational database provided to the author by Dr. Ronald Wright.

The analysis of the data provided in Wright et al. (2018) was limited to aggregate summaries of the trends at the venire member level. That is to say, they examined the strike trends for both the defence and the prosecution, conditioning on some additional variables. There was also spatial analysis performed, where different urban counties were directly compared. These analyses were also displayed using contingency tables. The stark differences between prosecution and defence strike patterns for venire members of different races was a key finding when the aggregate data was analyzed.

### 3.1.2 Cleaning

#### Flattening the Data

For greater expediency of analysis, the relational database of the Jury Sunshine Data was first flattened. The relational database was read into Microsoft Excel and the `readxl` package [Wickham and Bryan (2018)] was used to read the excel file into the programming language R. A wrapper for the `merge` function was developed which provided a simple output detailing the failed matches of an outer join in order to ensure that the flattening of the data into a matrix did not miss important data due to partial incompleteness. The code for this wrapper can be seen on the author’s GitHub at [Salahub/peremptory\\_challenges](https://github.com/Salahub/peremptory_challenges). The full GitHub url is provided in D.1.

---

<sup>1</sup>The link provided in the Jury Sunshine Paper to the specific source ([http://www.nccourts.org/Citizens/SRPlanning/Statistics/CARports\\_fy16-17.asp](http://www.nccourts.org/Citizens/SRPlanning/Statistics/CARports_fy16-17.asp)) does not appear to be working as of January 2019, however the NCAOC provides an API functionality at <https://data.nccourts.gov/api/v1/console/datasets/1.0/search/> which may provide the same data.

This wrapper revealed only a small number of irregularities in the data, which are detailed in [B.1](#):

- i.) Twenty-nine charges missing trial information such as the presiding judge (all of trials with IDs of the form 710-0XX)
- ii.) Twenty-six prosecutors not associated with any trials and missing demographic data
- iii.) One trial missing charge information (ID 710-01)

Ultimately, the jurors for trial ID 710-01 were included in the analysis as their records were complete otherwise. The prosecutors and charges which could not be joined were excluded, as they could have easily been included by collectors accidentally. Due to the small size of these inconsistencies relative to the size of the data set, they were not a cause for concern.

### Uninformative Columns

Of course there were other irregularities in the data than the obvious ones that arose in the flattening process. There are a handful of likely sources for these errors. The first of these is the anonymization of the data for public use. The private data includes a wealth of privileged data such as juror name and address, and these were removed in the data given to the author.

As a consequence of this anonymization as well as the inclusion of rarely used columns such as those for additional notes, some columns of the data contained only missing values. Most baffling of these was the `BirthDate` variable in the `Jurors` table, as there was no clear reason for this data to be missing. Thankfully, none of the missing columns were relevant to the joins performed in flattening, and they would have been only secondary in data analysis. As a consequence, these uninformative columns were simply removed from the data.

### Coding Inconsistencies

Related to this problem was the issue of inconsistently coded variable levels. An example of these inconsistencies would be levels recorded as both lower and upper case letters, or the presence of `?` instead of `U` for unknown values. It is very likely this inconsistency was a direct result of the data collection method which used many data collectors working independently in different places at different times. Thankfully, [Wright et al.](#) provided the codebook used by data collectors, which served as the authoritative reference for the admissible factor levels of all variables. Rectifying these inconsistencies was as simple as setting all demographic variable levels to be uppercase and replacing obviously mis-specified levels.

One specific inconsistency which should be noted is that of the outcome, which had a handful of entries recorded as `HC`, an inadmissible level not defined by the codebook. It is likely that this level represented a typo, as the “H” and “G” keys are adjacent on the American QWERTY keyboard layout, and `GC` was the code for ‘guilty as charged’, and so every occurrence of `HC` was replaced with `GC`. Additionally, the inadmissible level `G` was replaced by `GC`.

## Swaps

A more difficult level misspecification problem was the presence of what appeared to be columns with swapped values, frequently occurring with the gender column (the admissible levels of which are M, F, and U) and the political affiliation column (the admissible levels of which are D, L, R, I, or U). The aforementioned “swaps” appeared as records in which, for example, the gender was recorded as R and political affiliation as M. More complicated swaps of three columns also occurred. To address this problem, the `IdentifySwap` function was written.

The `IdentifySwap` function accepts two arguments: a data frame with named columns and a named list of vectors of the acceptable levels for some of the column names. It then performs vectorized checks of the specified column names and presents any rows which may have swaps or errors interactively to the user, along with a suggested reordering to “un-swap” the row. The user can press enter to accept the suggested reordering, enter some other reordering, or enter 0 to indicate that the row was not a true swap, but simply an error. The un-swapped entries are then returned to the data, and the rows with errors have the erroneous values replaced by U, the universal code for unknown in all data variables<sup>2</sup>.

The source of these swaps is also most likely the data collection method. The codebook provided specifically notes that the data collection was meant to record the race (R), gender (G), and political affiliation (P) data in the form RGP, but it is not inconceivable that it would occasionally have been recorded in some other ordering in the tedium of data entry. In any case, this problem affected only 431 records of the nearly 30,000, suggesting that the recorded error rate was not unacceptably large.

## Charge Classification

Perhaps the least regular data in this data set was that of the charge text. Due to the lack of any codebook guidance about the standard way of recording a charge in a trial, identical charges were recorded in numerous ways. The first method used to combat this was removing non-alphanumeric characters, extra spaces, and converting all charges to lower case. This still left considerable variation, however. Consider the charge of breaking and entering, for example. Even after this simple preprocessing the entries varied significantly (e.g. “break or enter”, “breaking andor entering”, “breaking and or entering”, etc.).

As a consequence, the processing was more involved. First, the most common versions of the charge text for the charges were all regularized to be identical (see `StringReg` in the code). Next, a regular expression classification tree was developed, which would also account for specific features of a charge. When identifying murder, for example, it seemed important to ensure attempted murder was separated from murder itself, and separating first and second degree was also desired. This tree would, when presented with a charge, apply the regular expressions at each node to the charge. If the charge matched the expression at a node, the regular expressions of that node’s children were applied to the charge until it was classified to some leaf node, each of which had a standardized value which replaced the charge. A small example of this structure is displayed in Figure 3.1, and the full tree is visualized in B in Figure B.1.

<sup>2</sup>The notable exception to this insertion of U was the case of the judge Arnold O. Jones II, whose gender was not recorded in the data, but who was identifiable as a man using a quick Google search of his unique name.

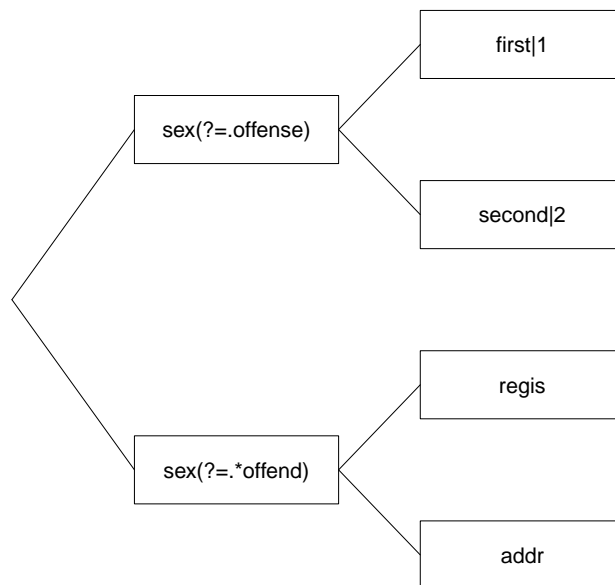


Figure 3.1: An example of a simple charge classification tree to separate the sexual offences from charges levelled against previously known sex offenders. A charge would be classified from most general on the left to most specific on the right.

By performing regularization using this charge tree, regularized charges were guaranteed. The cost of this regularization was the inability to classify all crimes, however. Of the 1407 charges present in the data, the tree provides regularization for 1209. With additional time and inspection of the failed matches, the tree could conceivably be expanded to regularize all charges. As the charges were not the primary feature of interest, however, such effort was not expended.

Instead, a number of helpful aggregation and extraction functions were developed to further simplify the charges. To start, they have been aggregated by intuitive classes: sex-based offences, thefts, murders, drug charges, violent offences not otherwise classified, and driving charges. Other classes, such as the North Carolina felony classes themselves (as provided by [North Carolina Sentencing and Policy Advisory Commission \(2017\)](#)), may provide a more informative classification rationale.

### Variable Level Renaming

The final step of the data cleaning process was to convert the uninformative codes used to indicate variable values to more intuitive and clear names (for example to convert **I** in the political affiliation variable to **Ind**, a clearer indication of independent). Certain variables which were already clear, such as gender (codes **M**, **F**, **U**), were not renamed due to the clarity of the one letter representations.



### 3.1.3 Variable Synthesis

In order to expand the analysis and visualization potential, a number of variables were synthesized from the Jury Sunshine data set. They are detailed below.

**Race Match** A logical variable which is true for a venire member if they are the same race as the defendant, and false otherwise. This variable was motivated in particular by *R. v. Stanley*, implicit contention of which was that the First Nations venire members were struck by the defence because their race did not match that of Stanley.

**Guilty** Logical indicator indicating whether the trial verdict was guilty or not.

**Racial Minority** Logical indicator of non-white venire member race.

**Race of Striking Party** Factor variable which gives the race of the prosecution if the venire member was struck by the prosecution, the race of the defence if the venire member was struck by the defence.

**Simplified Race** Due to the scarcity of the other minority races, this variable simplified the race provided to **White**, **Black**, or **Other** for the venire member.

**Simplified Defendant Race** The same as the simplified race for the defendant races.

**Simplified Disposition** This variable combined the levels **Foreman** and **Kept** in the original disposition variable into the level **Kept**.

## 3.2 Stubborn Legacy Data

### 3.2.1 Methodology

[Grosso and O'Brien \(2012\)](#) also provided data to the author, albeit a more limited set. This study, also based in North Carolina, focused on the trials of inmates on death row as of July 1, 2010, yielding a total of 173 cases. In each proceeding, the study examined only those venire members not excluded for cause, and critically the analysis of the study focused only on prosecutorial peremptory challenges. Besides collecting demographic data as in the Jury Sunshine Case, this study also collected attitudinal data for the venire members. This attitudinal data for the venire members is somewhat more detailed than the political affiliation data provided in the Sunshine data, including attitudes about the death penalty, employment information, and opinions on the trustworthiness of law enforcement.

Staff attorneys from the Michigan State University College of Law were responsible for the data collection in this study. The work was performed similarly to the Jury Sunshine Data, using case files to collect information about the court proceedings such as the peremptory challenges used, presiding judge, prosecutor, and defence lawyer. Detailed verdict and charge information was not collected, as the pre-selection criteria of death row inmates made the verdict clear, and the death penalty can only be applied for serious crimes.

To collect demographic and attitudinal data, the juror questionnaire sheets were consulted. These sheets are typically used as a component of voir dire, in order to make the process more efficient and determine venire members categorically ineligible for jury duty in advance. As a result, they inquire about opinions on the death penalty, for example, as well



as demographic questions. As not all jury questionnaires were available, additional information was collected from jury roll lists to determine the races of the final jury members. It should be noted that this collection was done blind and to high standards of proof, and a reliability study carried out in [Grosso and O'Brien \(2012\)](#) indicated that under this system the race coding was 97.9% accurate when the standards were met. Those for whom the standards were not met were marked as "Unknown."

The lack of an examination of political affiliation by this study, instead choosing to input far more detailed data on venire member viewpoints, serves as a barrier to the comparison of this data to the Sunshine data on an identical basis. However, the racial data for the two is recorded in a very similar way, so this variable can, at least, be compared.

### 3.2.2 Cleaning

The data provided to the author was already exceptionally clean, and so no cleaning was required. There was no variable synthesis performed on the data, rather variables were transformed and combined to generate analogous measures to those recorded by the Sunshine data. These combinations included using indicators of strike status to make a disposition variable analogous to that provided in the Sunshine data, and combining certain racial indicators into the more universal `White/Black/Other` coding.

## 3.3 Philadelphia Data

### 3.3.1 Methodology

[Baldus et al. \(2001\)](#) presents a similar data set to [Grosso and O'Brien \(2012\)](#) collected using similar means. Court files such as the juror questionnaire, voter registration, and census data were all used to complete juror demographic information for 317 venires consisting of 14,532 venire members in Philadelphia capital murder cases between 1981 and 1997<sup>3</sup>. It should be noted that this data included only those jurors kept or peremptorily struck, venire members struck for cause were not included. The procedure used to determine race using the census and voter registration polls was quite complicated, but was rigorously performed using accepted census methods to a standard of 98% reliability<sup>4</sup>.

This data had a number of departures from the Sunshine and Stubborn data. It lacked racial information as detailed as either, and collected detailed attitudinal variables as in the Stubborn data as opposed to the simple political affiliation reported by the Sunshine data. These limitations are quite clear when comparisons between all three data sets are made.

### 3.3.2 Cleaning

One interesting quirk of the Philadelphia data set was missing values. The codebook describing the data explicitly stated a number of variables should be recorded as binary

---

<sup>3</sup>This study took into account the sampling error by reweighting venires based on the year of the trial and the defendant race, as court records showed that the sample coverage varied over these factors.

<sup>4</sup>Additionally, imputation was only performed in a small minority of cases.

values. In the provided data files, however, these variables were missing for a majority of the observations. In the case of the `FINLJURY` variable<sup>5</sup>, for example, there were 4626 records with a value of 1, 3 with a value of 0, and 12890 missing values. These missing values were assumed to be zero, as using this assumption created a data set which was consistent with that reported in [Baldus et al. \(2001\)](#). As with the Stubborn data, the only variable synthesis performed was completed to create analogous variables to the Sunshine data.

---

<sup>5</sup>An indicator of whether the jury member was included in the final jury.

## Chapter 4

# Analysis

With the data cleaned and processed, questions could now be posed and addressed through analysis. As noted in [2.3](#), extensive empirical analysis of this subject and theoretical explorations have answered a great deal of questions already. It seems clear that race is an important factor in the exercise of peremptory challenges, and this can detrimentally affect jury composition. Theoretical investigations have suggested that the exercise of peremptories may increase the proportion of extreme jurors rather than reduce them.

In light of this, it is natural to wonder whether there is any evidence that the most common arguments posed in favour of peremptory challenge are satisfied in this data. As discussed in [2.4](#), there are two primary arguments. The first is the argument that the peremptory challenge is necessary to remove the “extremes of partiality” present in the venire for both sides; that is to remove the most extremely biased jurors. This goal is complemented by the ability of the judge to remove jurors with cause, which is also designed to remove those jurors with extreme bias. The second argument is the creation of a jury which is mutually acceptable to both parties in the trial.

### 4.1 Extremes of Partiality

Unfortunately, not much can be said about the argument of extreme bias. Suppose the bias of the population for a particular trial is modelled using a beta distribution, where each individual has some bias between 0 and 1 which represents that individual’s subjective probability that the accused is guilty before seeing any evidence. A perfectly unbiased individual would have a bias of 0.5: equally likely to say the accused is innocent or guilty and so equally receptive to the arguments of the prosecution and defence. These are the ideal jurors without extreme bias.

This simple bias model is that used in [Ford \(2010\)](#), and is appropriate for the domain of this problem. Whether the bias of members of a population really follows such a pattern is not known, as this definition of “bias” is somewhat difficult to measure. [Flanagan \(2015\)](#), for example, attempts to avoid this issue by instead simply modelling the votes of the venire members conditional on all of the features and evidence of a trial.

Under the simplistic beta model, the diverse possible centres, spreads, and shapes of a beta distribution make statements about the proportion of points around 0.5 impossible.

Consider the case of a beta distribution with a probability density function parameterized as  $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$   $x \in [0, 1]$ , if  $\alpha \rightarrow 0$  and  $\beta \rightarrow 0$  this distribution becomes entirely polarized, with all of its probability mass shifted to its boundaries at 0 and 1. Conversely, as  $\alpha \rightarrow \infty$  and  $\beta \rightarrow \infty$  the density becomes degenerate at 0.5. Clearly the beta assumption permits no conclusive statement on the extremity of observations without additionally assuming some parameterization.

The assumptions of [Flanagan \(2015\)](#) do not allow much more to be determined. Using this conceptualization, the best characterization of an individual with extreme bias is an individual that would vote to convict or not convict regardless of the evidence presented. Once again, knowing the proportion of such individuals present in the broader population is necessary to make statements of the validity of a certain proportion of the venire being struck. Here the relationship is more direct, as the appropriate proportion of struck venire members would simply be the proportion of such certain convictions or acquittals in the venire.

Loosening the idea of bias does permit stronger statements. While it is by no means objectively valid, another definition of “bias” might simply be the distance of a particular venire member’s opinion from the mean. Under this definition, the location of the distribution of venire bias values is no longer relevant. This comes at a cost of meaning, however. If the entire population believes a particular individual to be guilty because of the bad reputation of that individual, the implication that the population is “unbiased” simply because they are in perfect agreement does not satisfy “bias” in its typical meaning.

Granting this definition for a moment, mathematical constraints such as the Chebyshev inequality,  $P(|X - E[X]| > k\sigma) \leq \frac{1}{k^2}$  for  $X$  a random variable with  $Var(X) = \sigma^2$  and  $k \geq 0$  [[Weisstein \(2018\)](#)], can be used to provide an upper limit to the dispersion of any distribution from its mean. The Chebyshev limit gives an estimate of the proportion of observations above a given distance from the mean, measured in standard deviations, given a maximally dispersed distribution. That is, the maximum proportion of individuals of a given extremity relative to the mean. In some sense, it gives the expected rejection distances for symmetric rejection given a maximally extreme population. If given a proportion, one can use this limit to determine the corresponding maximal distance from the mean that could produce such a proportion in a valid distribution.

Therefore, if “bias” is allowed to mean the distance of a particular individual’s voting preferences from the mean in standard deviations, the proportion of rejected venire members allows for the calculation of the distance beyond which a venire member is considered too extreme to be allowed onto a jury. Of course, this calculation assumes that the defence and prosecution are truly acting to remove any venire member that is too extreme, rather than trying to bias the jury towards their case.

Table 4.1 provides a summary of the strike rates of the different data sets and the implied distances from the mean, in standard deviations, that these give for symmetric rejection using the Chebyshev inequality. Note that two rows are provided for the Sunshine data, the first for the entire data set and the second for first degree murder trials only<sup>1</sup>. This was done to facilitate comparison, as both the Stubborn Legacy and Philadelphia dataset only addressed capital cases<sup>2</sup>.

<sup>1</sup>The only sentence with the potential of a death penalty, as informed by the sentencing guidelines provided by the [North Carolina Sentencing and Policy Advisory Commission \(2017\)](#).

<sup>2</sup>It should be noted that there is still some difference in studied populations, as the Stubborn data covers only individuals sentenced to death, while no individuals in the Sunshine data set were sentenced

Table 4.1: The implied statistical extremity bound for symmetric rejection in the datasets under the Chebyshev limit

Data	Strike Rate	Chebyshev Limit
Sunshine	0.434	1.517
Sunshine Capital	0.639	1.251
Stubborn	0.659	1.232
Philadelphia	0.736	1.166

As the rejection bounds are typically near one standard deviation from the mean, this analysis suggests that the average individual sampled from the population is “extremely partial” in the eyes of the court across these data sets. Such rejection is inconsistent with the concept of a jury as the “conscience of the community,” one of the core philosophies of a jury trial, and implies the overuse of rejections by the court in some way.

Of course, as stressed above, it is not possible to comment with authority on the presence of partiality in the population. Critically, if the population has a tendency to be biased as a whole, *i.e.* the “unbiased” opinion is not the mean of the population distribution, the above arguments on symmetric extremity break down, and such high strike rates may very well be appropriate to find the unbiased jurors.

Such criticism may be supported by the increased use of peremptory challenges for the capital cases in the data. As is clear viewing historical Gallup polls on the topic [[Gallup \(2018\)](#)], opinions on the death penalty and murder trials are far more polarized and have varied greatly over time in the United States. At times the population has been biased as a whole towards the death penalty for murder rather than life imprisonment, regardless of the law around the subject. In the case of murder trials, then, perhaps only a minority of the population holds unbiased views on the subject.

Consequently, the strike rates are suggestive, but may be misleading. Clearly strikes remove a large proportion of venire members, but it is impossible to conclusively state whether more individuals are being rejected from serving on the jury than the proper amount. Indeed, the concept of the “appropriate” rate of strikes is unclear without any detailed knowledge of the population bias as a whole. However, such a crude aggregate measure would not resolve all of the controversy of peremptory challenges in any case, and so this inconclusive answer should not dissuade further investigation.

## 4.2 The Impact of Race

The racial controversy surrounding peremptory challenges provides one hypothesis about the pattern of venire member rejection which may warrant such further investigation. To begin, a simple marginal investigation was performed to explore the impact of the simplified venire member race on the peremptory strike probability. The result of this investigation is displayed in Table 4.2. Of particular interest is whether any race is far more likely to be struck by peremptory challenge than the others, as this would suggest that race is the target of an undue rate of strikes.

Table 4.2: The conditional probability of a venire member being struck peremptorily by the simplified venire member race across data sets. These values are smaller than the values presented in the extremity analysis as only the individuals which were identifiably removed by peremptory challenge are counted in this table. Regardless, the comparisons remain similar even if the unattributed removals are included. Note that the Philadelphia trial data only indicated black and non-black venire members and so only two numbers can be reported.

Data	Black	Other	White
Sunshine	0.23	0.24	0.25
Sunshine Capital	0.62	0.76	0.63
Stubborn	0.65	0.36	0.66
Philadelphia	0.67	0.68	

These probabilities are different, but not greatly so. Indeed, the trend of higher probabilities for the removal of white jurors across all data sets is perhaps counter-intuitive given the history of controversy in the United States. In any case, the small magnitude of these differences seems to suggest that there is no strong racial bias at the aggregate level, whether or not the results are statistically significant<sup>3</sup>.

This table also demonstrates some of the drawbacks of tables, the dominant method used to display the data throughout Wright et al. (2018), Grosso and O’Brien (2012), and Baldus et al. (2001). The table, while excellent at communicating specific values, does not provide a great sense of trends or patterns without careful engagement by the reader. A critical component of the communication and comparison of any analysis to others is the ability to quickly and effectively compare trends and patterns in the data. Consequently, the “mobile plot” for visualizing the three way relationships of categorical variables was developed. A detailed description of this plot and its development which includes a discussion of the principles of graphics and perception which were used to devise its form is presented in A<sup>4</sup>. This visual tool was then combined with inspiration from *Swain v. Alabama*, *Batson v. Kentucky*, and *R. v. Stanley* to generate 4.1, which displays the relationship between venire member race and defendant race.

First, a small explanation of this mobile plot. This mobile plot displays the relationship between three categorical variables: venire member race, defendant race, and disposition (whether a venire member is struck and by whom). The vertical axis corresponds to the conditional probability of a particular disposition given a race and defendant race combination<sup>5</sup>. Racial combinations are placed along the horizontal axis, and each combination corresponds to one horizontal black line in the plotting area. The length of these lines is proportional to the number of venire members in the data with the corresponding racial combination, and their vertical positions are the mean conditional probability of a venire member being struck for that particular combination. The dashed vertical lines, coloured by disposition, start at this mean line and extend to the observed conditional probability of the corresponding disposition for the relevant racial combination. As a consequence,

<sup>3</sup>Consider the numerical impact of a statistically significant difference of a few percent when the jury size is 12 for each trial.

<sup>4</sup>Here it suffices to mention that much of its design was motivated by the philosophy of Tufte (2001) and the results of Cleveland and McGill (1987) on the accuracy of visual perception.

<sup>5</sup>Generally, any three categorical variables can be displayed using a mobile plot, as the mobile plot is used to display the distribution of a categorical variable given the combinations of the other two. Mathematically it displays  $X_3|X_1, X_2$ .

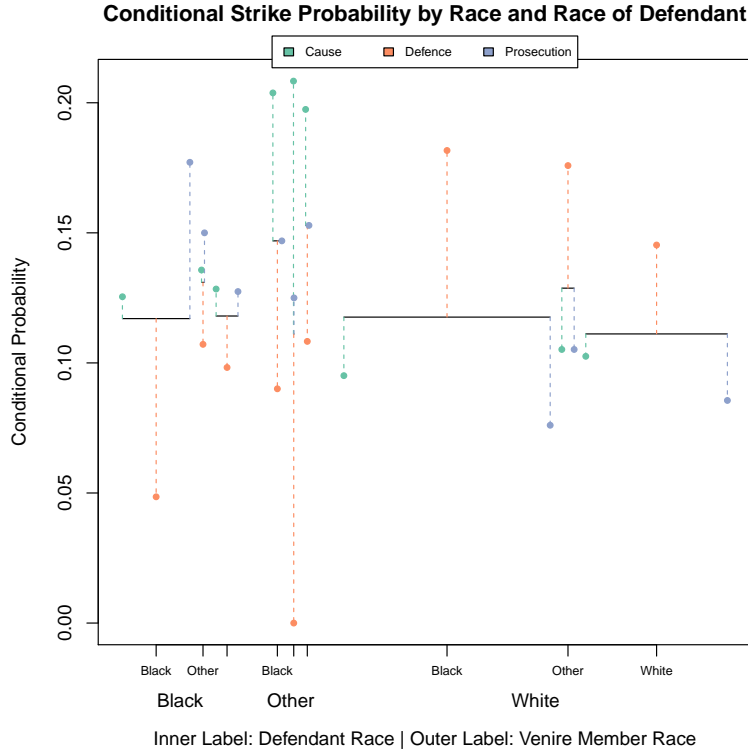


Figure 4.1: The conditional probability of the strike dispositions given the venire member and defendant race, with the expected value represented by the horizontal black lines, and the observed values represented by the points at the end of the dotted lines. Each horizontal black line corresponds to a particular venire member and defendant race combination, with a length proportional to the number of venire members with that combination. The dashed vertical lines, coloured by challenge source, start at these horizontal lines and end at points which show the observed probability of a venire member being struck by the source for the given racial combination.

this plot can be viewed as a visualization of the test of a specific hypothesis:

$$D|D \in \{2, 3, 4\}, R, E \sim Unif(\{2, 3, 4\}) \quad (4.2.0.1)$$

Where  $D, R, E$  are random variables representing the disposition, venire member race, and defendant race respectively as outlined in the Notation section. In words: the conditional distribution of the disposition given both that a venire member is struck and the racial combination is uniform. This implies that causal challenges, defence strikes, and prosecution strikes occur with the same probability for each racial combination, though the rate may differ between racial combinations. Such a hypothesis allows for certain racial combinations to experience a higher strike rate generally, but constrains the strike rate to be the same for all parties, which would imply that all parties in the court pursue an identical strike strategy across all venire member and defendant race combinations.

Clearly, Figure 4.1 casts some doubt on this hypothesis. While the horizontal black lines tell a very similar story to Table 4.2, with little variation between them except for in small population subsets, a number of other striking patterns are visible. The first, and most obvious of these, is the main effect of venire member race. While the aggregate removal rates do not seem to depend on the race of the venire member, it is clear that the defence

and prosecution pursue radically different strategies. The defence seems biased towards a jury with more venire members from racial minorities. All orange points are below the horizontal lines for the black and other venire members, indicating these groups are less likely to be struck by the defence than expected, while the points are above the lines for the white venire members, indicating a higher than expected probability of defence strike for white venire members. The prosecution seems to mirror this tendency, striking the white venire members at a lower rate than expected and the black venire members more often than expected. Strikes with cause seem to show less deviation from expectation for the black and white venire members, and always deviate from expectation in the same direction as the prosecution.

The addition of defendant races shows another interesting trend. It would seem that the aforementioned tendencies of the prosecution and defence are strongest for black defendants, which have the greatest departures of the conditional probabilities from the expectation. The defense and prosecution seem to have slightly more similar habits when the defendant is white, despite their opposite tendencies in all cases. Finally, it would seem that the removals with cause have tendencies similar to the prosecution, as the points representing the conditional probability of a venire member being removed with cause are always on the same side of the expected line, an event which would occur with probability  $2^{-9} \approx 0.002$  under the hypothesis of independent uniform strike rates. Further discussion of the agreement of these two strike tendencies can be found in 4.4.

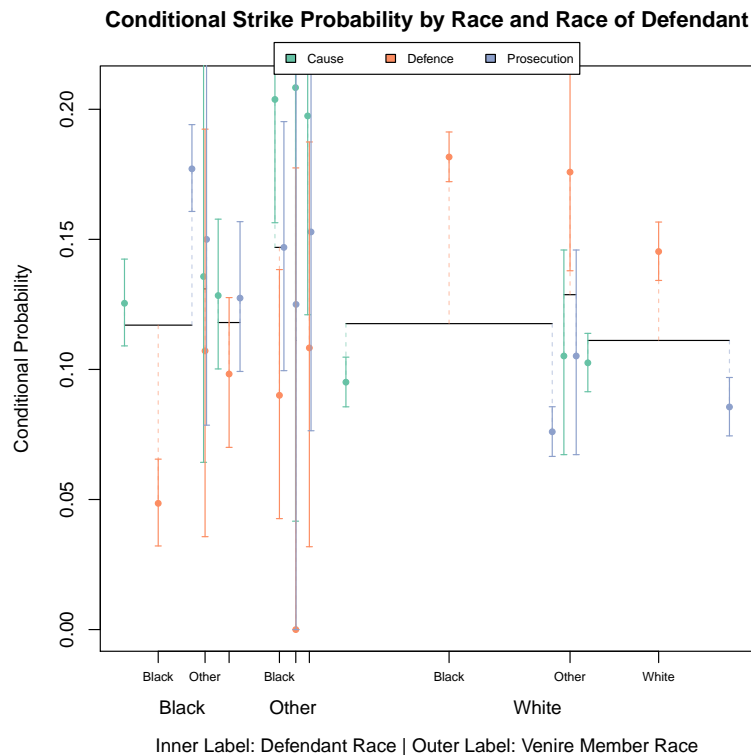


Figure 4.2: The plot of conditional strike probability by racial combination from 4.1 with confidence intervals added. Note that many of the seemingly striking departures seen are insignificant when these confidence intervals are applied, especially for races other than black and white.

While Figure 4.1 is quite suggestive, the widths of certain horizontal black lines, in particular those for venire members with a race other than white or black, suggest that perhaps



some of the more extreme tendencies are simply a result of the well-known higher variation of samples with small sizes. In order to see the true nature of the noted departures some incorporation of the variation one expects from each observed value is required. This is accomplished by the addition of approximate 95% simultaneous multinomial confidence intervals using the `MultinomialCI` package in R, which implements simultaneous confidence intervals for multinomial proportions following the method presented in [Sison and Glaz \(1995\)](#). These confidence intervals can be seen in Figure 4.2.

As suspected, some of the results for the smaller sample sizes do not seem to be significant. The results for the larger groups, in particular for white venire members or black defendants, are significant, however. It should be noted that these simultaneous confidence intervals do not constitute a proper statistical test of the impact of race, they are rather a way of visually providing a viewer some sense of the expected variability in the data over repeated sampling. More rigorous testing requires controlling the impact of confounding factors, as done by the modelling in 4.4.

#### 4.2.1 In the Stubborn and Philadelphia Data

Already the utility of the mobile plot, and visualizations of the data in general, should be clear. A wealth of information is displayed very simply in Figure 4.2. However, the real power of this plot comes with the ability to quickly compare different data sets. Consider, for example, comparing the results of [Baldus et al. \(2001\)](#) and [Grosso and O'Brien \(2012\)](#) using Table 4.2. In order to complete this comparison, the rows for each set must be viewed and the numbers committed to memory before the reader moves to the appropriate row to compare values. While the simple four row and three column structure of this particular table make this rather straightforward, it becomes more difficult and less clear as these tables grow in complexity. Compare this with a cursory glance at Figure 4.3, which makes the similarities of the data sets immediately clear<sup>6</sup>.

Despite the very different study sample universes of these three data sets, all display identical patterns, with only the magnitudes of the different strike rates differing. The mobile plot format makes several interesting aspects immediately clear. The similar level of all black lines within each plot shows that in each data set, the aggregate probability of removal is similar across racial minority combinations, as was implied by Table 4.2. However, these aggregate similarities hide the vastly different strategies of the defence and prosecution which are consistent across all data sets. The defence has a tendency to retain black venire members, striking them at a lower rate than the other<sup>7</sup> venire members, while the prosecution shows a pattern which mirrors that of the defence, removing more black venire members and fewer other venire members. In all data sets the gap between these probabilities and the expected strike rates is greatest for the black venire members in cases with black defendants.

<sup>6</sup>Three important differences between Figure 4.3 and Figure 4.2 must be noted when interpreting these visualizations. The first of these is that data differences have limited the scope of the comparison to simply the prosecution and defence strike rates, as the Stubborn data does not include any information on strikes with cause. Second, the Sunshine data used to generate the mobile plot in Figure 4.3 is filtered to only the first degree murder trials, as the other two data sets only addressed capital trials (*i.e.* those with the death penalty), while the Sunshine data had a broader scope. Finally, the race and defendant race were further simplified to logical indicators of whether the race variable was black or not.

<sup>7</sup>One must keep in mind that these non-black venire members are predominantly white in the data sets for which detailed race information was recorded.

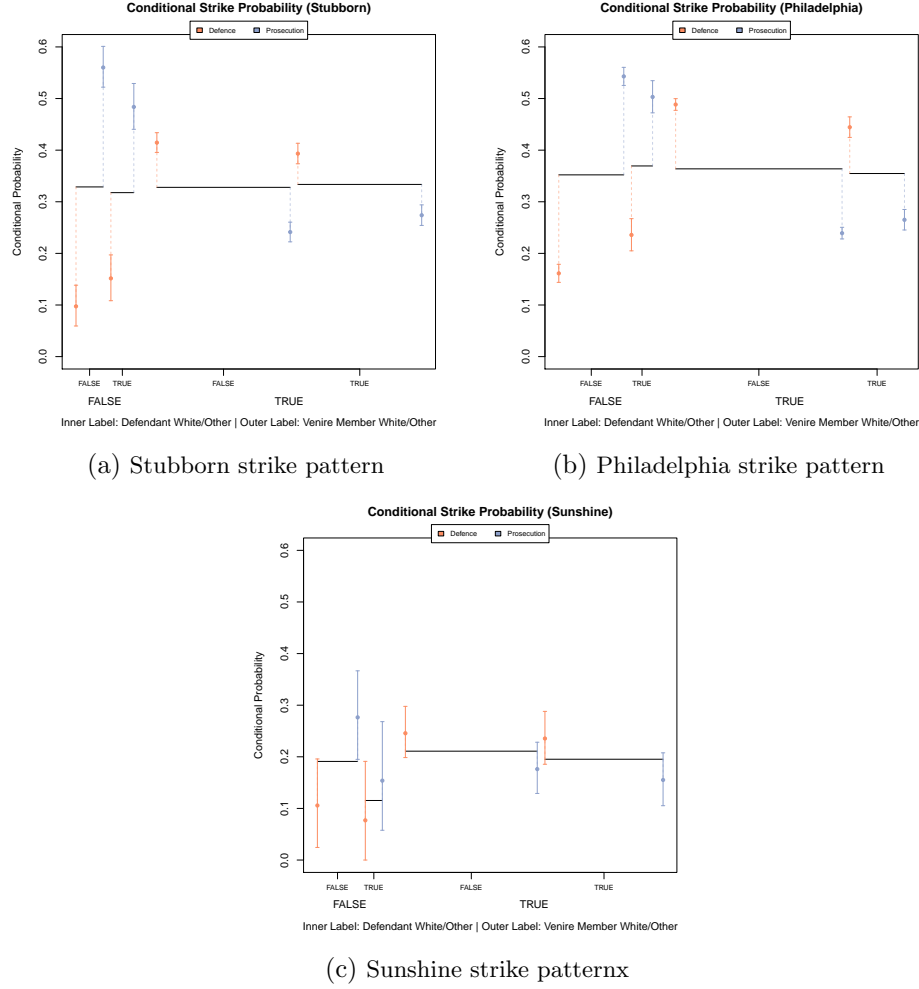


Figure 4.3: The conditional probability of defence and prosecution peremptory challenge by race across all capital trials in all data sets. The pattern, though sometimes different in magnitude, is quite consistent across the three, despite the significant differences in the respective study sample universes.

It should be noted that the Sunshine data set looks most unique of the three, and this may be a result of the sample size. While the Philadelphia data and Stubborn data both collected data which included multiple years, the Sunshine data was restricted to trials which occurred in 2011. This small sample is the reason for the large confidence intervals present in Figure 4.3c. This does not explain the overall lower strike rate observed in the capital trials in this data, which is also visible in Table 4.2. This departure may be of interest for further investigation.

### 4.3 Other Factors

Of course, it would be incorrect to conclude immediately that the cause of the racial patterns observed across these data sets is race itself. There may be a plethora of attitudes associated with race that could serve as legitimate cause for a peremptory challenge. As noted by Justice Byron R. White in the majority opinion in [Supreme Court of the United States \(1965\)](#)

[The peremptory challenge] is no less frequently exercised on grounds normally thought irrelevant to legal proceedings or official action, namely, the race, religion, nationality, occupation or affiliations of people summoned for jury duty. For the question a prosecutor or defense counsel must decide is not whether a juror of a particular race or nationality is in fact partial, but whether one from a different group is less likely to be.

This quote leads directly to the heart of the problem. Without detailed transcripts indicating how the venire members were questioned, it cannot be known if the aggregate pattern of removal is the result of racially based strikes, or whether the lawyers determined valid reasons for a peremptory challenge during the voir dire process which are simply related to race. For example, if defence attorneys reasonably assumed that deference to authority would make a venire member insurmountably predisposed to reject their arguments about mishandling of evidence by police, this would be reasonable grounds for peremptory challenge. If such opinions were distributed heterogeneously by race, the aggregate pattern may appear to reflect racially-based decision making by the defence attorneys despite the true explanation being legitimate and non-racial<sup>8</sup>.

#### 4.3.1 Political Affiliation in the Sunshine Data

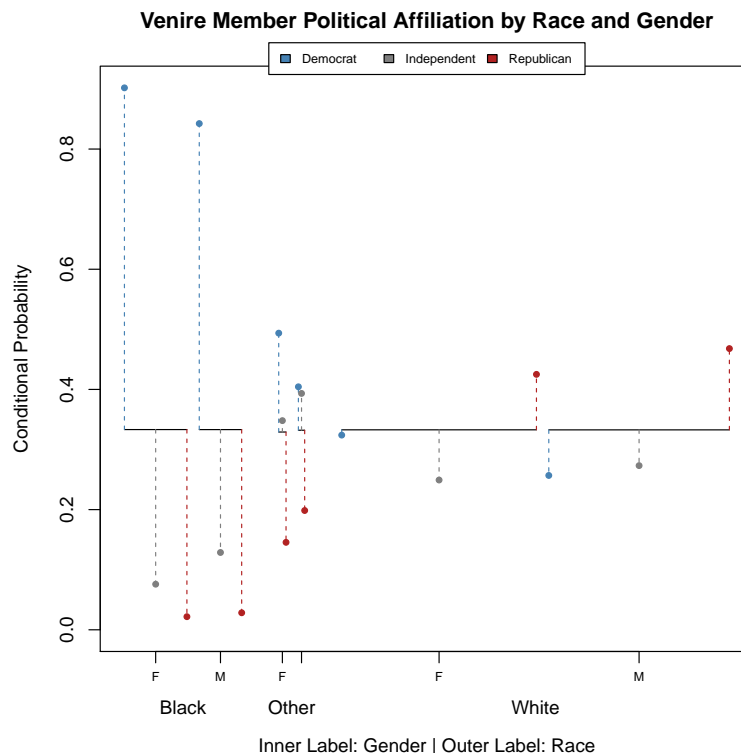


Figure 4.4: Conditional probabilities of political affiliation by race and gender. In this plot, it is clear the black venire members are far more homogeneous than the white venire members for both genders, with the vast majority voting Democrat.

<sup>8</sup>The Stubborn and Philadelphia data both provide detailed attitudinal survey data including questions related to trust of law enforcement, but how this data interacted with the questions asked during voir dire is unclear. Additionally, it is missing in the Sunshine data.

Revesz (2016) provides, inadvertently, data which might support the above defence of peremptory challenges on the basis of confounding variables. He notes that the distribution of political affiliation in the United States is not consistent across races, with black voters far more likely to vote for the Democratic Party and far less likely to vote for the Republican Party, while white voters are more uniformly distributed and likely to vote independently. If political affiliation is used as a surrogate for perspective, this suggests that the observed pattern in the Sunshine data could be the result of the defence removing conservative venire members and the prosecution removing liberal ones. As the Sunshine data has political affiliation, the results of Revesz can be compared to this data. The resulting figure, Figure 4.4, displays the conditional probability of political affiliation across races and genders.

What is immediately apparent viewing this plot and the data in Revesz (2016) is how closely the two data sets agree. Black voters vote Democratic a vast majority of the time, while the political leanings of white voters are less cohesive. Note additionally that the horizontal lines in this plot are all very close to 0.33, as almost all voters are Democrat, Republican, or independent. The “ideological imbalance” of racial voting tendencies, as Revesz aptly calls it, is a clear confounding factor and a possible source of a legitimate cause for an initially suspect overall trend. As such, the mobile plot was used to investigate the relationship between race, political affiliation, and disposition.

To control for the defendant race as well, which appears important in Figure 4.1, the venire members were split into the simplified racial groups black, white, and other. Then mobile plots of the conditional strike probabilities for the different venire member races given the defendant race and political affiliation were generated. Figures 4.5a, 4.5c, and 4.5b display these mobile plots.

This sequence of three plots immediately suggests that a political argument is insufficient for this data. For each venire member race and defendant race, the political affiliation of the venire member does not radically change the pattern of strikes for any party in the court. Rather, the court tendencies for each political affiliation, venire member race, and defendant race seem to follow the pattern seen in 4.1 for all political affiliations with the exception of some very small subgroups of black venire members<sup>9</sup>. Many of the other, less noticeable subgroups represented by this particular set of plots will be too small to provide statistically significant results, as is immediately obvious from the confidence intervals. Regardless of these difficulties with sample size, the consistency of the pattern of strikes across political affiliations provides rather stark evidence against a hypothesis of political affiliation as a primary cause of the strikes in the data.

### 4.3.2 Gender in the Sunshine Data

Another factor which may have an impact is that of gender. While *J.E.B. v. Alabama* (Supreme Court of the United States (1993)) has also ruled peremptory challenges for this reason alone unconstitutional in the United States, it is noted in Van Dyke (1977) on page 152-153 that prosecutor’s guidelines have, in the past, recommended using peremptory challenges to remove female venire members<sup>10</sup>, and so perhaps it is a relevant factor.

<sup>9</sup>The subgroups, black republican venire members for white or other race defendants, have sizes 1 and 10, respectively. The study examines a total of 29634 venire members.

<sup>10</sup>Van Dyke provides specific guidelines for sex written by Jon Sparling, an assistant district attorney in Dallas, which read:

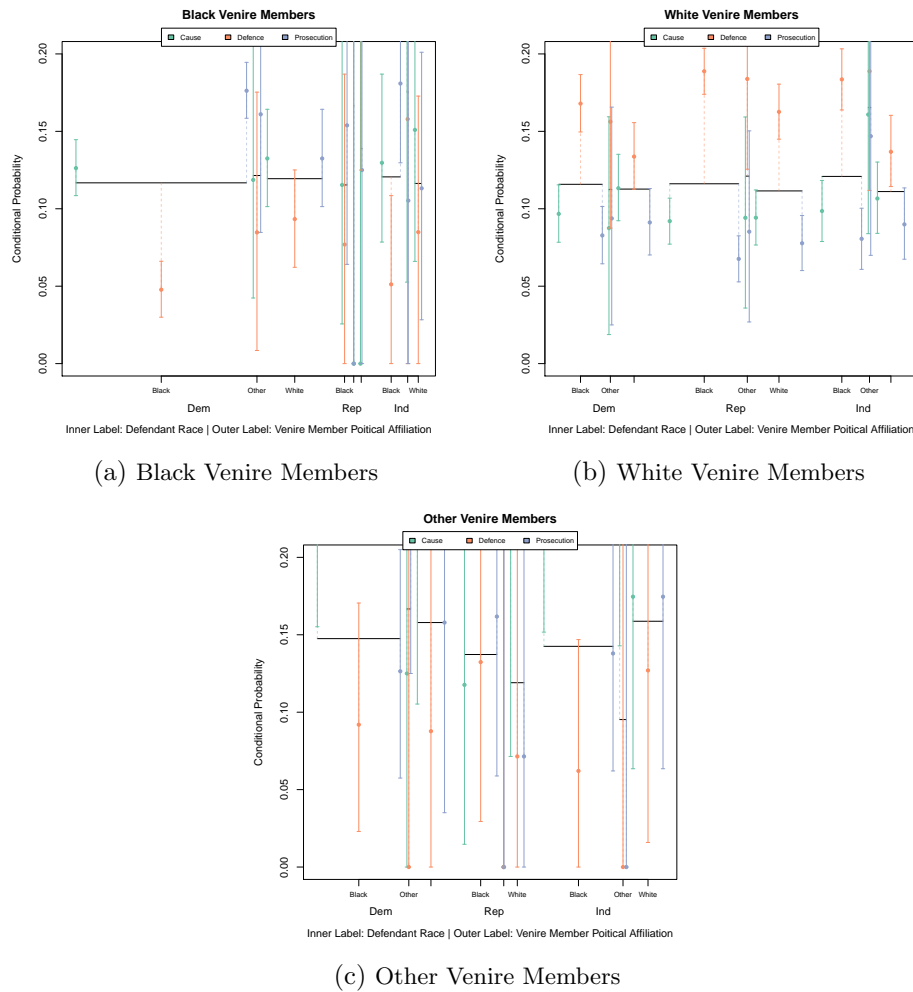


Figure 4.5: Conditional probability of venire member strike by defendant race and political affiliation, split by race. Note how the pattern of conditional probabilities is the same across political affiliations for the same defendant race within a particular venire member race, and this pattern is the aggregate pattern seen in Figure 4.1

Additionally, there is a relationship between gender and race in the data, as noted in [Wright et al. \(2018\)](#): black males are highly under-represented relative to black females. Luckily, the relationship between race, gender, and peremptory challenges can be visualized quite neatly without the need to split the plots as was done in Figure 4.5. Figure 4.6 displays the mobile plot for race, gender, and striking party.

The same pattern is seen here for the most part, that of a dearth of defence strikes against visible minorities and a surplus against white venire members, with a mirrored tendency

- i.) I don't like women jurors because I don't trust them.
- ii.) They do, however, make the best jurors in cases involving crimes against children.
- iii.) It is possible that their "women's intuition"[sic] can help you if you can't win your case with the facts.
- iv.) Young women too often sympathize with the Defendant; old women wearing too much make-up are usually unstable, and therefore are bad State's jurors.

If data on make-up use and jury outcome is ever collected perhaps Mr. Sparling's bold claim can be tested, but in light of his first point it is better treated as prejudice.

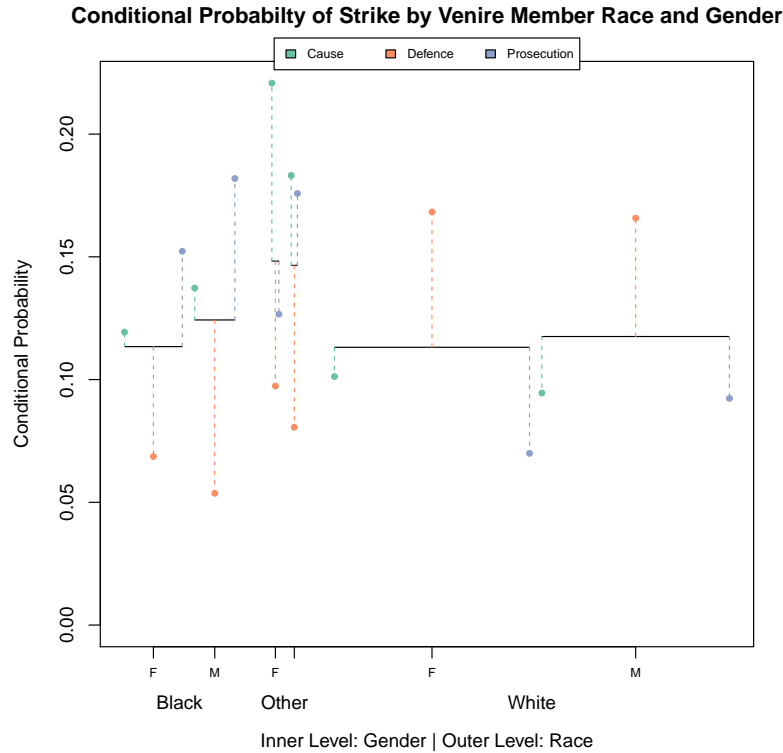


Figure 4.6: Conditional probabilities of strike source by race and gender. Note that the pattern is nearly identical for the genders within each racial group.

for both the prosecution and challenges with cause. One may be motivated next to ask whether such oppositional patterns are just a quirk of the data or mobile plots. Perhaps conditioning on gender alone, for example, would create a similar pattern of tendencies of the court. Figure 4.7 shows a plot of the court's strike tendencies given the gender of venire member and defendant gender to investigate this possibility.

The characteristic pattern seen throughout the other plots is not present. At this point the message from these additional mobile plots should be clear. That is, the dominant determinant of the strike probability for a venire member is likely their race. The race of the defendant does impact this somewhat, but plots across gender and political affiliation for both venire members and defendants in numerous combinations suggest that race is the dominant factor in determining the probability a venire member will be struck by the prosecution, defence, or removed by cause. This hypothesis, motivated by these plots, is examined more rigorously through the construction and testing of specific models in 4.4.

### 4.3.3 In the Stubborn and Philadelphia Data

Grosso and O'Brien (2012) and Baldus et al. (2001) did not collect data on the political affiliation of the venire members<sup>11</sup>, and Grosso and O'Brien (2012) did not record the

<sup>11</sup>While it is perhaps possible that the different combinations of the attitudinal data collected by these groups could be used with data on the voter profiles most typically associated with the political affiliations of different groups in the United States, such an investigation is beyond the scope of this work. Lacking such a predictor, it is also unclear which attitudinal variables would be most comparable to the political

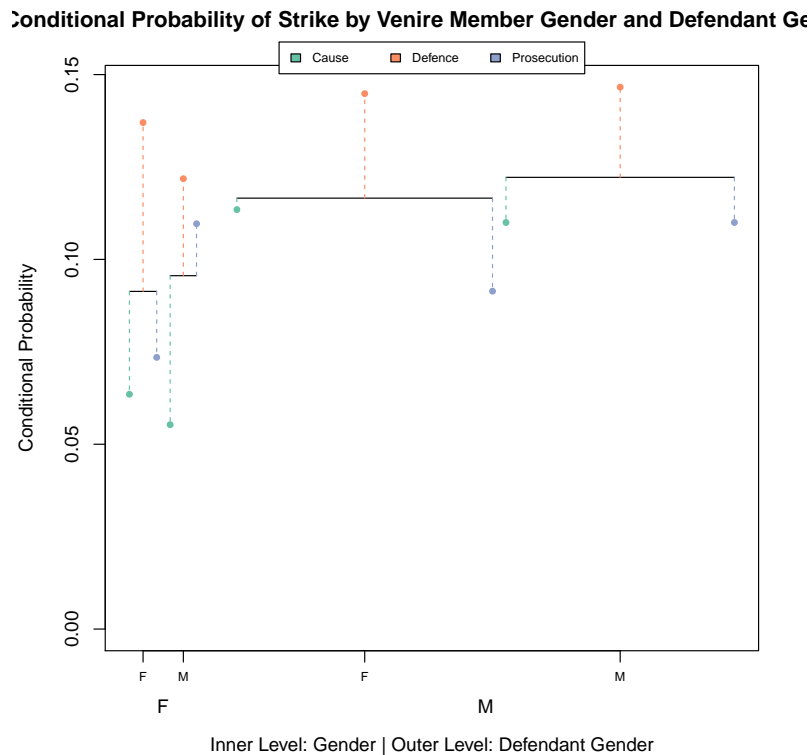


Figure 4.7: Conditional probabilities of striking party by venire member gender and defendant gender. Rather than the changing pattern seen with race, where the defence and prosecution show a preference for one race over others, the pattern here is rather constant, with the defence striking more than expected and the prosecution and strikes with cause generally occurring less.

gender of the defendant. Consequently, of the above plots, only Figure 4.6 can be compared to these two data sets. Figure 4.8 displays the plots. Note that as before, the data has been restricted to ensure analogous study universes across the data sets.

This, again, shows a pattern very similar to the overall Sunshine data, and the differences in pattern between these data sets are minor. The race of the venire member seems to be analogously important across all data sets as it was in Figure 4.3. Such similarity is somewhat surprising, as these data sets have considerably different sample universes both spatially and temporally, as discussed in Chapter 3. When displayed visually, the similarities are nothing short of striking. They suggest that the results seen in the Sunshine data are not a fluke, but reflect a greater pattern of prosecution and defence dynamics throughout the United States over time.

## 4.4 Modelling

### 4.4.1 Multinomial Logistic Regression

Though the above results in the Sunshine data set may be suggestive, refining the estimates for the impact of race on the probability of rejection and controlling for the myriad of

---

affiliation, and so attempting to compare this pattern would be futile.

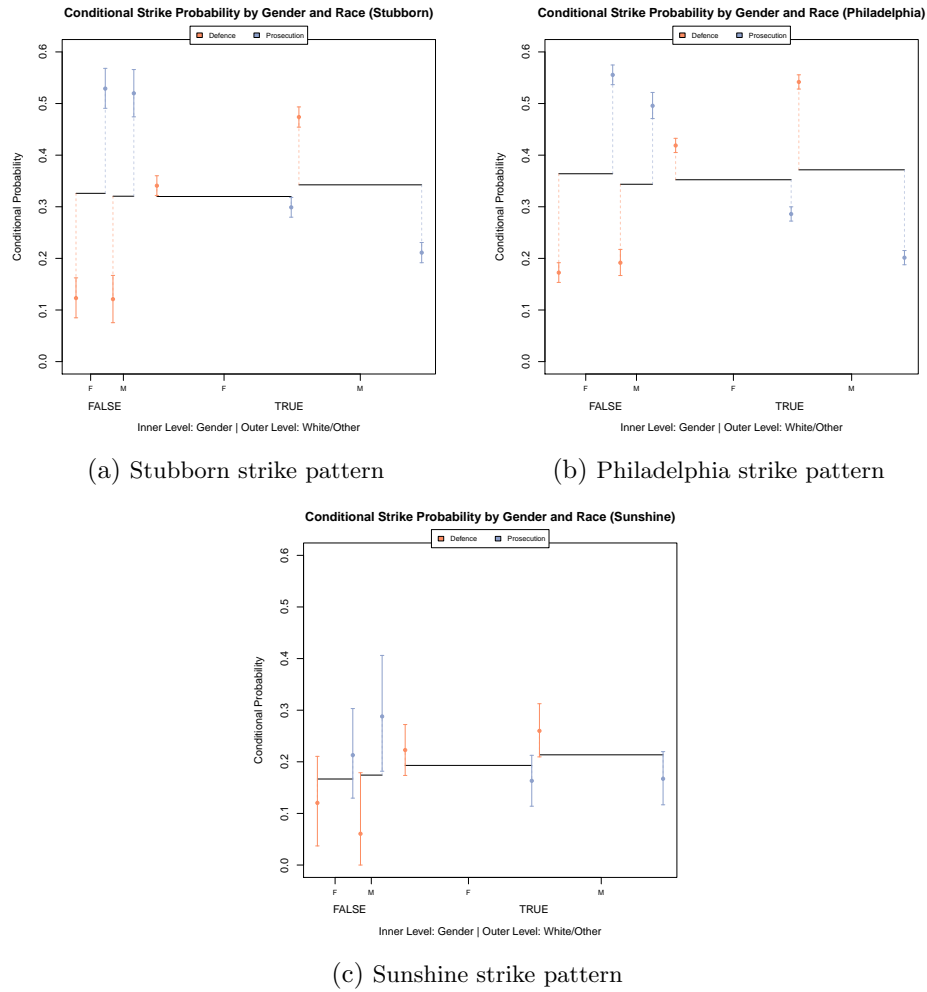


Figure 4.8: The conditional probability of defence and prosecution peremptory challenge by venire member race and gender for all capital trials in all data sets. The pattern, though sometimes different in magnitude, is quite consistent across the three examined data sets, despite the significant differences in the respective study sample universes.

possible confounding factors motivated the fitting of multiple models. This is helped by the nature of the mobile plots used in Figure 4.1 and elsewhere. The mobile plot displays the conditional distribution of a categorical variable with multiple levels, which evokes ideas of a conditional multinomial distribution to the statistically-minded. Consequently, fitting multinomial log-linear regression seemed a natural choice.

It should be noted that early modelling instead used Poisson regression due to the difference in implementations of multinomial regression (based on neural networks in `nnet`) and Poisson regression (based on iteratively-reweighted least squares in `glm`). The implementation difference, however, is not due to a difference in theory. As the conditional distribution of a particular outcome given some marginal count is multinomial for Poisson distributed random variables (as shown in C), these two models are essentially equivalent. There are known transforms to move between the models [Baker (1994)] and their equivalencies are well documented [Lang (1996)]. Multinomial regression was chosen for greater interpretability, as in a Poisson model the parameters of interest are all interactions as opposed to main effects. The specific implementation of multinomial regression utilized is



the `multinom` function in the `nnet` package in R, which implements a method of fitting multinomial models discussed in [Venables and Ripley \(2002\)](#).

In multinomial regression, the coefficients for each possible outcome generated for each factor of interest are log odds ratios of that outcome relative to some reference level, called a pivot. Mathematically, the coefficients give how a change in the independent variables changes  $\frac{\pi_i}{\pi_{pivot}}$ . For all models, the pivot outcome chosen was the probability of a venire member sitting on the final jury, in other words the probability their disposition was coded as `Kept`. Coefficients were then estimated using treatment contrasts with a black female venire member with Democrat voting tendencies in a case with a black female defendant used as the reference treatment (using the notation from the Notation section: the reference venire member combination was  $\mathbf{x} = (1, 1, 1, 1, 1, 1)^T$ ).

While the choice of race, gender, and political affiliation for the reference level was not deliberate, the choice of pivot outcome, that of a venire member being kept, was made in order to make the visualizations constructed using the coefficients clearer and easier to compare to previous visualizations. Previous visualizations, including the mobile plots, have displayed the conditional probabilities of removal with cause and strikes by defence and prosecution, and if any were used as the pivot their coefficient estimates would be hidden as the intercept.

The mobile plots created in [4.3](#) and [4.2](#) suggest, primarily, that an interaction between race and defendant race is relevant in modelling the conditional probability of venire member rejection. These plots do not suggest that any other interactions are likely to be significant. This led to a model:

$$\begin{aligned} \log \frac{\pi_d^{(i)}}{\pi_1^{(i)}} &= a_d + \mathbf{x}_i^T \beta + (re)_i \beta_{re} \\ &= a_d + r_i \beta_r + e_i \beta_e + p_i \beta_p + g_i \beta_g + s_i \beta_s + (re)_i \beta_{re} \end{aligned} \quad (4.4.1.1)$$

The fundamental purpose of fitting this model is to test two effects. The first is whether the interaction between defendant race and venire member race is significant, as has been suggested by the controversial trials named in Chapter 2, and the second is whether the venire member race is a significant factor in the presence of the other controlled factors. These two assumptions can be testing by fitting models nested in [4.4.1.1](#):

$$\log \frac{\pi_d^{(i)}}{\pi_1^{(i)}} = a_d + \mathbf{x}_i^T \beta \quad (4.4.1.2)$$

$$\log \frac{\pi_d^{(i)}}{\pi_1^{(i)}} = a_d + \mathbf{x}_i^T \beta' \quad (4.4.1.3)$$

Where  $\beta'_r = 0$ . Comparing [4.4.1.2](#) and [4.4.1.3](#) to [4.4.1.1](#) gives striking results. These can be seen in Table [4.3](#), which compares the deviances of the different models sequentially.

Even when controlling for defendant characteristics and the venire member's political affiliation and sex, the race of the venire member and its interaction with the defendant race are both highly significant at the  $\alpha = 0.05$  level. This suggests that the rejection of

Table 4.3: Comparison of models 4.4.1.3 and 4.4.1.2 to 4.4.1.1, displaying the residual deviance, residual degrees of freedom, differences, and p-values of these differences for adjacent models.

Model	Residual df	Residual Deviance	Difference	$P(\chi^2)$
4.4.1.3	55527	39496		
4.4.1.2	55521	39087	405	0
4.4.1.1	55509	39023	67	1.4e-9

venire members is, at least in part, based on their racial characteristics. Whether this is due to opinions associated with race or not is impossible to say from this data, but the control of political affiliation provides evidence that commonly held opinions within racial groups are not the cause. Note that the residual deviance values indicate that this model is underdispersed, implying that the significant test results gained are conservative.

Due to its significance compared to the other models, 4.4.1.1 was taken as the final model to estimate the race effects with precision. The estimated coefficients for the different effects and their approximate 95% confidence intervals, calculated using the standard errors of the coefficients and using a normal assumption<sup>12</sup>, are displayed in Table 4.4.

Table 4.4: The coefficients of 4.4.1.1 and approximate 95% confidence intervals. The features of venire members are given no indication, those of the defendant are clearly indicated as such.

Coefficient	Cause	Defence	Prosecution
(Intercept)	-2.18 (-2.4,-1.96)	-2.66 (-2.88,-2.44)	-1.45 (-1.64,-1.26)
Other	0.59 (0.3,0.89)	0.63 (0.22,1.04)	-0.19 (-0.52,0.14)
White	-0.25 (-0.4,-0.1)	1.21 (1.02,1.4)	-0.8 (-0.95,-0.66)
Defendant Other	0.02 (-0.5,0.55)	0.69 (0.08,1.29)	-0.21 (-0.71,0.3)
Defendant White	0.03 (-0.19,0.26)	0.67 (0.39,0.95)	-0.33 (-0.55,-0.11)
Independent	0 (-0.13,0.13)	0.06 (-0.05,0.18)	-0.03 (-0.17,0.1)
Libertarian	-0.5 (-1.74,0.74)	-0.7 (-1.93,0.54)	-0.28 (-1.51,0.96)
Republican	-0.14 (-0.27,-0.02)	0.14 (0.04,0.25)	-0.22 (-0.35,-0.08)
Male	0.03 (-0.07,0.12)	-0.01 (-0.09,0.08)	0.26 (0.16,0.36)
Defendant Male	0.74 (0.55,0.94)	0.25 (0.1,0.39)	0.21 (0.04,0.38)
Other & Def. Other	-0.08 (-1.29,1.13)	-10.68 (-10.68,-10.68)	0.12 (-1.29,1.53)
White & Def. Other	0.18 (-0.42,0.78)	-0.65 (-1.3,0)	0.62 (0.03,1.21)
Other & Def. White	0.02 (-0.55,0.6)	-0.46 (-1.21,0.29)	0.61 (-0.02,1.23)
White & Def. White	0.03 (-0.23,0.28)	-0.91 (-1.2,-0.61)	0.42 (0.16,0.67)

#### 4.4.2 Examining the Coefficients

This table is somewhat daunting and difficult to interpret, as predicted by 4.2. No sense for the relative magnitudes of the coefficients or their significance is offered by the precise recording of all values. This motivated further visualization of these parameter estimates in an intuitive and simple way. Inspired by Solt and Hu (2018), a custom dot and whisker plot was generated. This plot displays the coefficient estimates as points in the centre of a line, the endpoints of which give the associated confidence intervals. A vertical line is placed at zero to provide a reference for the sign and significance of different parameters.

<sup>12</sup>This assumption is only valid asymptotically, but for the large demographic groups the estimates are still a useful guide.

Modifying this concept to suit the multinomial regression data was simple, as it only required grouping the different possible outcomes for each parameter spatially. Figure 4.9 displays these parameter estimates in the dot-whisker format.

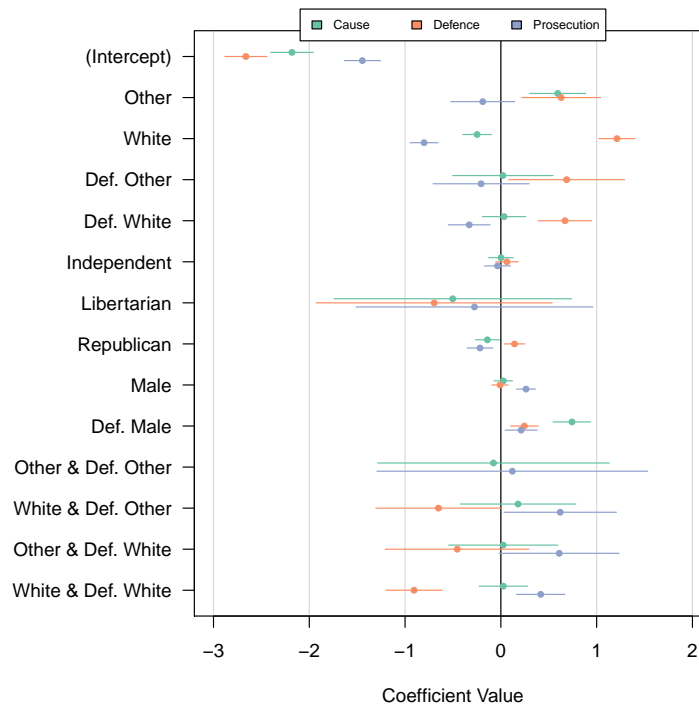


Figure 4.9: Model coefficients from the full model, 4.4.1.1, displayed using a dot-whisker plot. The lines indicate the confidence intervals while the central points indicate the point estimates of coefficients.

This plot is considerably less daunting than the table, and provides a much clearer picture of the magnitudes, signs, and significances of the variables. However, there is still considerable visual clutter. As was noted as early as Figure 4.2, the sample sizes in the Sunshine data set corresponding to racial groups other than white or black are too small to make statements of significance. This same pattern is observed here, where all coefficients estimating the effects of other racial groups have wider confidence intervals and are rarely significant. Additionally, the lack of libertarian voters creates large uncertainty about the estimates of this parameter as well. To provide a clearer picture of the racial coefficient estimates and the effect of the other controls, a second dot-whisker plot was generated excluding these parameter estimates in Figure 4.10.

The first feature of this plot to notice is the position of the coefficient estimates relative to the black line at zero. This indicates the sign of the coefficient, positive if the point is to the right and negative if it is to the left. A positive sign indicates that the factor corresponding to the coefficient increases the probability of a venire member being struck by a particular party and a negative sign indicates that the presence of said factor decreases the probability of a venire member being struck. The pattern of positive and negative values suggests the uncontrolled patterns noticed in the mobile plots were not the result of confounding.

First, take a broad overview of the patterns. The variables which show the greatest significance and the characteristic oppositional tendencies of the prosecution and defence are

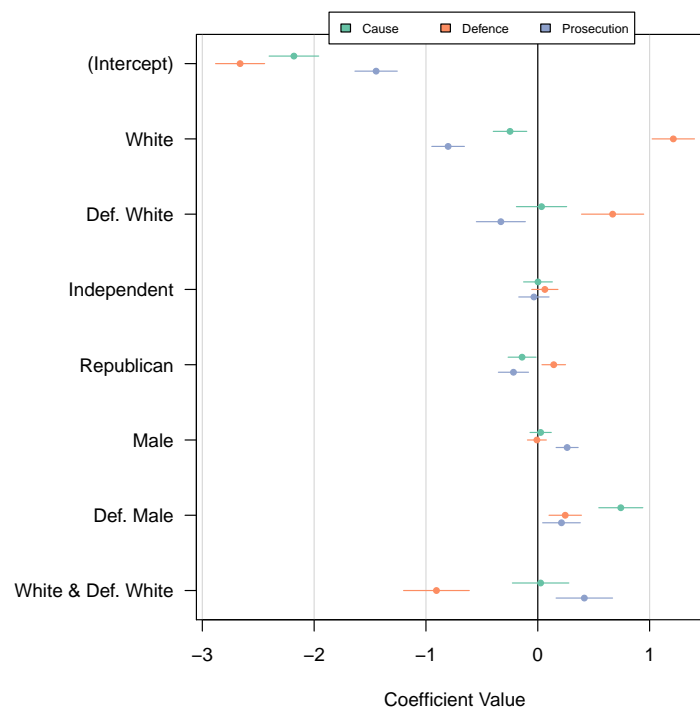


Figure 4.10: Select model coefficients from the full model displayed using a dot-whisker plot. The lines indicate the confidence intervals while the central points indicate the point estimates of coefficients.

all race variables. The effect of a white venire member, white defendant, and their interaction all display significant, and opposite, tendencies for these two oppositional parties. Moreover, the magnitude of the defence coefficients is always greater than the prosecution for the race variables. This suggests that the defence probability to strike is affected far more by race variables than the prosecution probability to strike. To put this more simply, the defence is more sensitive to the racial aspects of the trial in general.

A quick survey of the prosecution and cause coefficients shows far less agreement than seemed apparent in the mobile plots. Scanning from the top to the bottom, the cause coefficients match the prosecution for a few groups, but are generally quite different and are often more similar to the defence than the prosecution. The suggestive pattern of matching prosecution strikes and causal strikes viewed in 4.2 vanishes when controls are placed on possible confounders<sup>13</sup>. In general, the cause strike coefficients are lower in magnitude than both the prosecution and defence, with the notable exception of the effect for male defendants, where it is significantly greater than both.

In order to examine patterns in more detail some important differentiation is necessary between the intercept values and the other coefficients. As with all linear models, the intercept values provide a locational measure. That is to say, the intercepts provide the base level upon which the coefficients act. Consequently, the negative values of the intercept for all striking parties is to be expected. In the Sunshine data more venire members were kept than rejected, as can be seen in Tables 4.2 and 4.1, and so the intercept, which

<sup>13</sup>This serves as an excellent demonstration of how complex multiple confounders can create spurious patterns.

gives the log odds ratio of each strike outcome to the venire member being kept, would naturally be negative for all parties.

The more important feature to note for this intercept is the large difference between the different striking parties, without overlapping confidence intervals. These suggest that all three parties behave differently in the reference group, which consists of a female, black, Democrat venire member in a case with a female, black defendant. Crucially, the defence intercept is far lower than the prosecution, matching what was observed in Figure 4.1, where the lowest defence strike probability occurred for black venire members in cases with black defendants. Also visible in this figure is an increase in defence strike probability and decrease in prosecution strike probability for a white defendant with a black or other venire member. This pattern is reflected perfectly in the positive defence coefficient and negative prosecution coefficient for white defendants.

The coefficients for white venire members also match the expectations of Figure 4.1. The defence attains its largest positive coefficient for this group and the prosecution is largest negative value. The dominant effect of venire member race visible throughout the mobile plots in 4.3 and 4.2 is reflected by the dominance of the magnitudes of the coefficients for white venire members relative to the reference black venire members. The defence is much more likely to reject a white venire member and the prosecution much less likely.

Slightly attenuating this pattern is the interaction effect between a white defendant and a white venire member. For this specific combination, the defence is less likely to reject than for a white venire member with a black or other defendant, while the prosecution is more likely to reject. Both of these trends seem to be significant based on the plotted confidence intervals. Combine this observation with the almost mirrored pattern for a white defendant with a black venire member (given by the white defendant coefficient), and a more nuanced strategy becomes clear. While the prosecution dominantly rejects black venire members and the defence dominantly rejects white venire members, they are also sensitive to possible racial matches. The defence pattern is consistent with a strategy which aims to partially select for venire members which have the same race as the defendant, while the prosecution has the opposite pattern.

This pattern is precisely the problematic behaviour which was implied by the individuals that viewed *R. v. Stanley* as a travesty. While the data here *cannot* reveal anything about that *specific* trial, or indeed about the exercise of peremptory challenges in Canada generally. That such a pattern is clearly present in the data will no doubt feel vindicating for those still musing on *R. v. Stanley*'s perceived injustice.

Such racial patterns in the prosecution and the defence dominate all other effects. The political effect is minor, though consistent with the hypothesis put forward by Revesz (2016), with a preference for Republicans by the prosecution and a preference for Democrats by the defence. Interestingly, all parties seem to have an increased probability for male venire members, suggesting a universal preference for female venire members on the jury. One would hope that this pattern is for some other reason than anything listed on pages 152-153 of Van Dyke (1977) and reproduced here in 4.3.1. While no detailed explanation is forthcoming for this pattern in the presence of racial and political affiliation controls, it would certainly be an interesting avenue for further research, as the exercise of peremptory challenges based purely on gender is also protected since *J.E.B. v. Alabama* [Supreme Court of the United States (1993)].

A lack of consistent information between data sets and of time made fitting additional

models using the other data sets more difficult. Regardless of the lack of specific coefficient comparisons due to these difficulties, the above findings are consistent with the modelling results in [Grosso and O'Brien \(2012\)](#) and [Baldus et al. \(2001\)](#). Both of these studies performed logistic regression analysis of the prosecution strike patterns, controlling for attitudinal variables. In both cases, the venire member race remained a significant predictor of the venire member removal. Critically, [Baldus et al.](#) fit analogous models for the defence strike use and found the race remained significant even when the attitudinal variables were controlled. Additionally, these models found the prosecution struck black venire members at a higher rate and the defence at a lower rate, exactly as found here. The results of the modelling performed here add evidence to the mounting case of racial importance in the execution of peremptory challenges.

## 4.5 Trial Level Summary

While [Wright et al. \(2018\)](#) reported a great deal of aggregate statistics about the venire members themselves, one piece of investigation which was lacking was an analysis which viewed the trends for cases, rather than simply for individual venire members. As it cannot be known why a particular venire member is struck, and viewing their aggregate statistics tells us nothing about how different strikes relate to each other, it is possible the aggregate patterns are the result of some effect which is not due to persistent bias across trials.

By aggregating the venire members by trial and viewing the demographic trends in strikes and behaviour at this level, detailed insight into the impact of challenges at a more relevant scale is gleaned. Additionally, such aggregation allows for the synthesis of certain measures, such as a disitributional difference via the Kullback-Leibler divergence [[Kullback and Leibler \(1951\)](#)], which would otherwise not be well defined. This particular perspective of the data has also not been explored by any other studies known to the author.

### 4.5.1 Estimating Struck Juror Counts

One gap which was present in the Sunshine data was the total count of defence and prosecution removals for each trial. This variable is of minor importance for modelling the individual venire members, but it is of major interest when viewing the trials themselves. While counts of these strikes were provided in the data, there were many missing values, or values inconsistent with the number of associated venire members in the data. For example, many of the recorded values in these columns were zero even when venire members associated with that trial were marked as struck.

To remedy this, a number of variables were generated for each trial. Primarily, these variables were counts of the number of venire members with particular characteristics which were struck by both the prosecution and defence in the trial. For example the variable `DefRem.Race.Black` provided the count of black venire members struck by the defence for a specific trial. To provide an estimate of the number of venire members struck by each party, these counts were summed for one particular variable, typically gender. The sum was then compared to the recorded value for that party's removed count. The greater of the two values was kept as an estimated count to be used in analysis. For both the prosecution and the defence, about 80% of the sum and recorded values matched

exactly and about 18% of the recorded values were less than the sum. This suggests some incompleteness for the remaining 2–4% of the data.

### 4.5.2 Visualizing the Racial Trends

#### The Positional Boxplot

Motivated by the plots in the `extracat` package in R [Pilhöfer and Unwin (2013)], in particular the `rmb` plot, a positional proportion plot, called here the “positional boxplot,” was developed to display the strike tendencies across trials. The data for each trial consists of categorical variables and two integer counts corresponding to the estimated defence and prosecution strike counts described in 4.5.1. The positional boxplot was developed to investigate the relationship between the strike patterns of both parties and the defendant race.

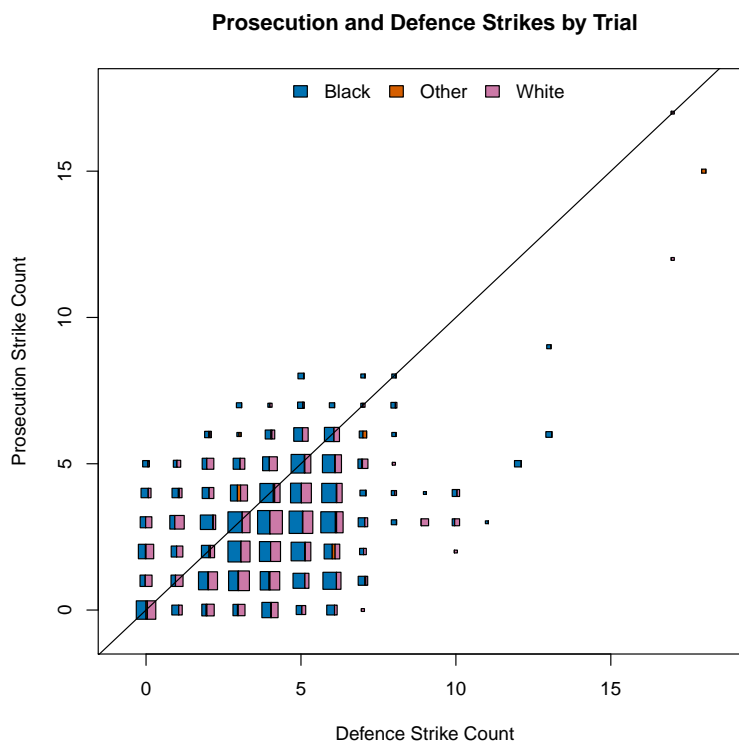


Figure 4.11: The positional boxplot of strikes by race of defendant for the Sunshine data. There does not seem to be a dependence of strike counts on the defendant race, the boxes look similar across the entire plot.

Therefore, the positional boxplot is designed to visualize the conditional distribution of a categorical variable given two integer count variables. Each observation consists of a level of a categorical variable and two counts. Across the whole data, there may be occurrences of identical values for the two counts. At each unique combination of integer counts observed, a box is placed with an area proportional to the number of observations with that specific combination. Each box is then subdivided horizontally such that the width of each subdivision is proportional to the corresponding count of each level of the categorical



variable that occurs with that specific integer combination.

Figure 4.11 displays a positional boxplot of defendant race to prosecution and defence estimated strike counts. While the internal box patterns look rather similar everywhere, split somewhat evenly between black and white defendants, their areas are quite revealing. In the aggregate examination of the data, it was clear that the defence had an overall higher strike rate than the prosecution, this effect can be viewed clearly in Figure 4.7 where the defence has a higher strike rate across all gender combinations. What was not clear was whether this pattern reflected the typical pattern at the individual trial level or was rather due outlier cases with abundant defence strikes. The areas of the boxes in Figure 4.11 indicate clearly that the defence typically uses more peremptory challenges than the prosecution in each trial.

Of course, the most striking result of the modelling from 4.4 seen clearly in Figure 4.10 was the dominance of the venire member race in determining their strike probabilities. Viewing the defendant race may therefore miss interesting patterns. A pair of positional boxplots to display the racial tendencies across trials for the defence and prosecution is shown in Figure 4.12.

Figure 4.12a shows the strong preference of the defence to strike white jurors regardless of the defendant race. The tendency of the defence to strike jurors different than the defendant noted in 4.4.2 is also visible in this plot. Note the generally greater proportion of black defendants for higher white venire strike counts. In contrast, the opposite trend is shown in Figure 4.12b, where the boxes further to the left have a greater proportion of black defendants and those moving up have a greater proportion of white defendants.

Interestingly, the prosecution seems less biased than the defence at the trial level. The distribution of strikes appears symmetric for different venire member races. Such an appearance, however, is misleading, because it fails to account for the number of black and white venire members presented to the court. These numbers are very different, as can be seen by the lengths of the horizontal black lines in Figure 4.1. A sizeable majority of the venire members are white. Luckily, with trial aggregated data, the raw strike counts can be normalized into proportions. The resulting scatterplots of these proportions by trial are displayed in Figure 4.13.

Here, the prosecution and defence biases are much more clear. The prosecution never strikes more than 40% of white venire members presented, and on average strikes a greater proportion of the black venire than the white venire. The defence, in contrast, regularly strikes more than 40% of the white venire, and on average strikes a greater proportion of the white venire members than the black venire members. This reinforces the aggregate observations made in 4.4, indicating that these mechanics operate on the individual trial level, and not just over all trials. Additionally, the high variation visible in Figure 4.13 suggests that the aggregate patterns described in 4.4 are not followed by all defence or prosecution lawyers, merely on average.

While these differences are intriguing, the far more interesting observation that can be gleaned from these plots is the fundamental difference in inclusion of minority and majority groups in jury formation. The aggregate statistics indicate that the black venire members are a minority, and Figure 4.13 suggests that as a consequence, it is common that a majority or all of the potential black jurors will be removed by peremptory challenge while such an occurrence is incredibly rare for the majority white jurors, if not impossible. While in many trials all black venire members were struck, in not a single trial was every white



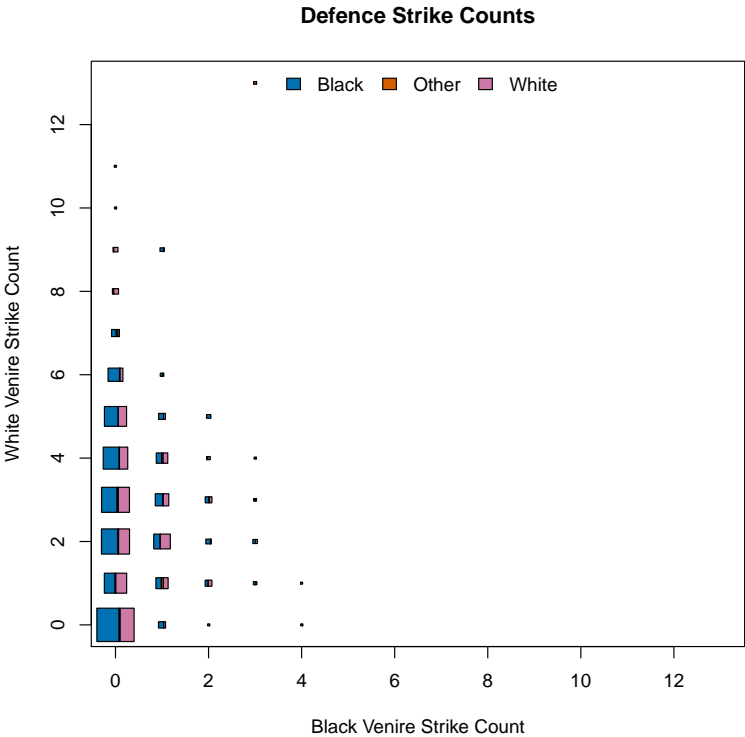
venire member struck. This suggests that if one had a strategy of keeping minorities off of juries, the peremptory challenge system would make this task easier. Such an observation is obvious in theory, but it is no less impactful to see it emerge from the data.

## 4.6 On Venire Selection

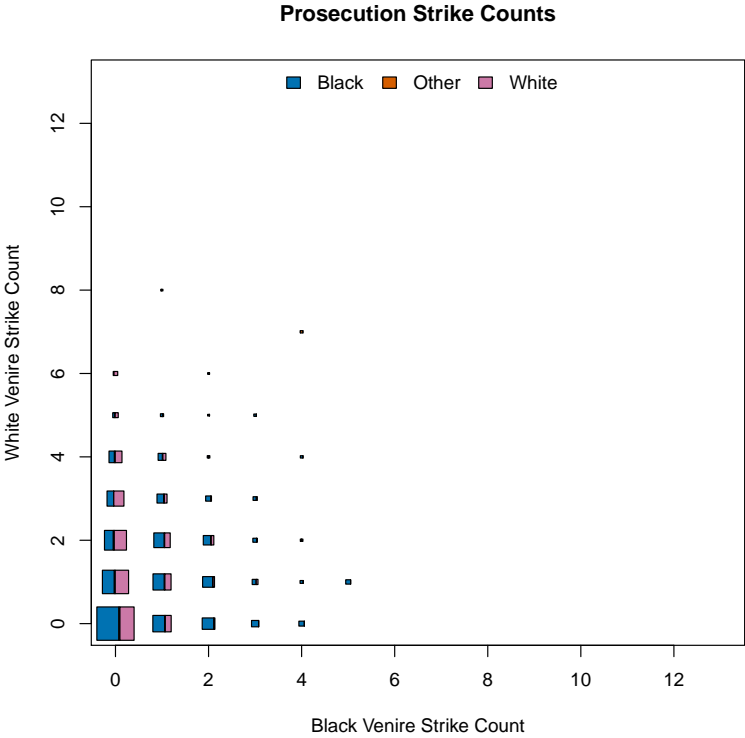
[Van Dyke \(1977\)](#) spends much of chapters four and five exploring the causes for the underrepresentation of certain groups in jury venires, and his analysis suggests that underrepresentation starts at the jury selection stage. Fewer non-white individuals register to vote generally (page 89). Additionally, excusal from jury duty, primarily granted for economic hardship, impacts disadvantaged economic groups to a greater extent (pages 113-120).

Such explanations provide a plausible reason why black males would be most underrepresented in venires, and are also possibly contributing factors to the white majority of the venire despite the majority of defendants being black. Such issues with the jury selection process will not, and cannot, be solved by simply removing the peremptory challenge. They have much more to do with the relationship between minority groups and wider society, and so require more comprehensive and complex solutions.

*R. v. Stanley* serves as an excellent demonstration of this problem. As [Quenneville and Warick \(2018a\)](#) report, 750 individuals on the jury roll were sent court summons, but only 204 of these individuals appeared on the day of the trial. This large non-response population introduces the potential for incredible bias in the venire. If there is some relationship between demographics and an inability to serve on a jury or a lack of will to respond to a court summons, as [Van Dyke \(1977\)](#) suggests with ample evidence, then problems with minority representation may be due in larger part to the method of jury summons than the exercise of challenges in the courtroom. Removing the peremptory challenge may serve as a useful step in the right direction, but it cannot be expected to “fix” minority representation alone.

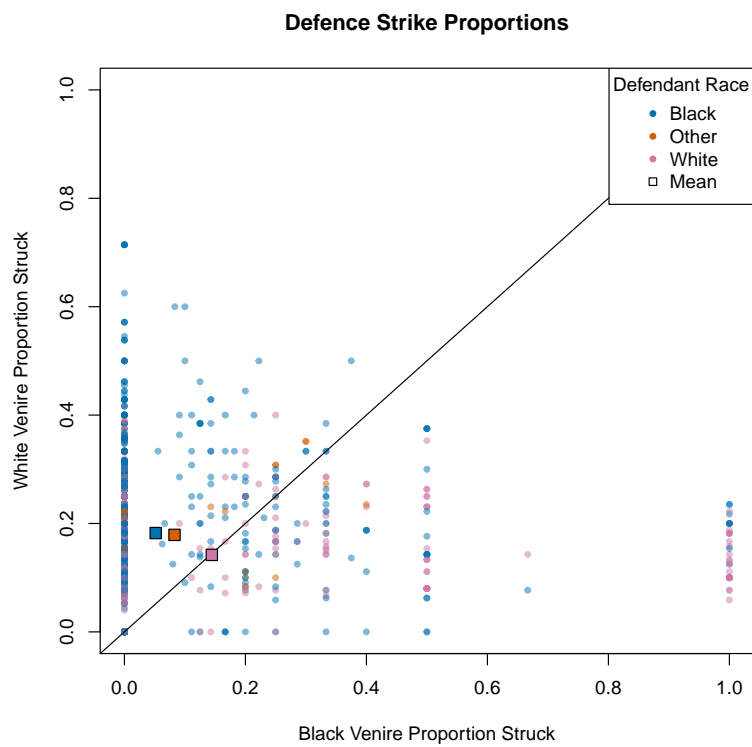


(a) Defence count removed

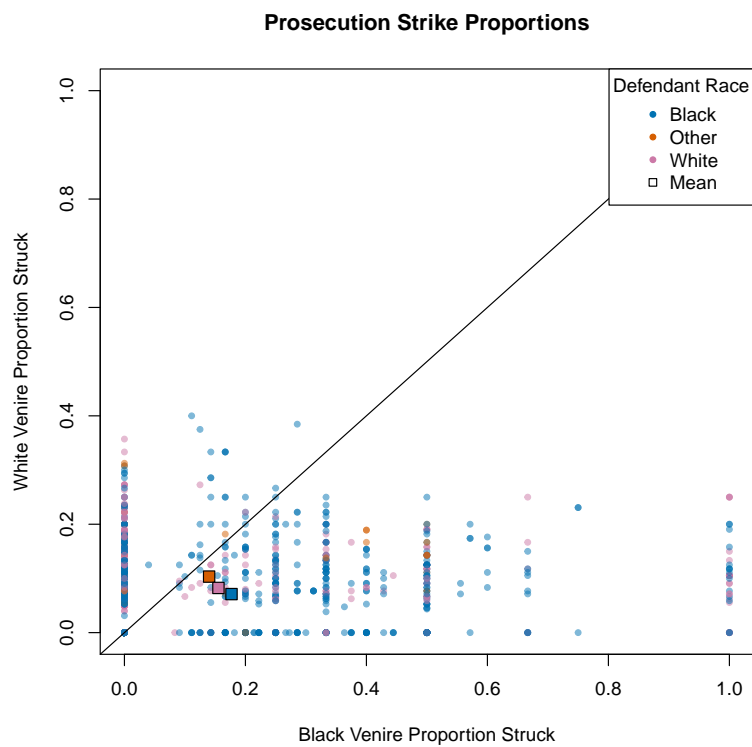


(b) Prosecution count removed

Figure 4.12: Positional boxplot of racial strikes for both the defence and prosecution.



(a) Defence proportion venire removed



(b) Prosecution proportion venire removed

Figure 4.13: Scatterplot of venire proportion removed by race for both the defence and prosecution.



## Chapter 5

# Summary

The visual tools and models presented here support the dominant analysis in the literature. The primary determinant in the exercise of peremptory challenges is race in the Sunshine data set. The prosecution tends to remove more racial minority venire members than expected and fewer white venire members than expected. The defence tends to have the opposite strategy. This pattern is not only seen in aggregate in [4.2](#), but is visible in the trial summaries presented in [4.5](#). The impact of race remains apparent in the mobile plots even when other legitimate factors such as political affiliation are controlled.

Beyond detecting the patterns in one data set, this work demonstrates other strengths of the mobile plot and visual analysis. The first of these is the utility of the mobile plot to compare the strike patterns across multiple data sets. The similarities between the Stubborn, Philadelphia, and Sunshine data sets are immediately clear when visualized appropriately. This is critical in examining the practice of peremptory challenges, as it allows for a comparison of their use across studies with radically different study populations so long as analogous data is collected.

In this case, the strength of the similarities observed between these data sets when visualized with the mobile plot suggests the pattern of racial preferences is not a local phenomenon in location or time, but is a reflection of a strategy utilized by the prosecution and defence in jury trials generally. As more data and investigation take place, further visual comparisons can motivate more analysis of the similarities between these patterns in different studies. Based on the findings here and in the other empirical analyses, the critical data which should be collected for analysis is the disposition of the venire members and the demographic characteristics of both the venire member and defendant. Attitudinal data, which was not standardized in this study, should also be collected and analyzed in some standard way to augment the utility of the mobile plot for comparison.

Another strength of the mobile plot is the motivation of model building. The multinomial regression models from [4.4](#) were created to fit models analogous to the displayed patterns of the mobile plots generated in advance. That the findings of the models matched the analysis of the mobile plots almost exactly justifies these plots as informative analytical tools. The models allowed for the estimation of effects in the Sunshine data controlling for possible legitimate confounders, giving a table of coefficient estimates consistent with those generated previously for other data sets, such as the Stubborn and Philadelphia data.

It was not until these coefficients were visualized with the dot whisker plot, however, that a number of more nuanced patterns became obvious. The first of these is the greater sensitivity of the defence to the racial aspects of a trial than the prosecution. That is, the race of the venire member has a greater impact on the defence's probability of rejection than the prosecution's. The second pattern is the tendency of race matching by the defence and race contrasting by the prosecution. This aggregate pattern also seems to be reflected in the trial level summary of the data, which suggests that this trend is not a quirk of aggregation, but a reflection of individual lawyer decision making.

Of course, as suggestive as these patterns are, and as spotted as the history of peremptory challenges is with controversy, none of this can say definitively whether the individuals rejected from the Sunshine venires were rejected appropriately due to their "extreme" bias. Without detailed descriptions of the bias of the population as a whole, such judgements on the propriety of strikes simply cannot be made, and whether these racial strike patterns are simply the result of legitimate strikes issued for reasons related strongly to race remains unknown.

Despite this limitation, the final scatterplots suggest a criticism of peremptory challenges independent of these concerns which is consistent with the source of controversy for *Batson v. Kentucky*, *Swain v. Alabama*, and *R. v. Stanley*. Peremptory challenges frequently remove all representatives of minority groups from the venire. This prevents their participation in a panel meant to represent the conscience of their community, corroding a critical function of the jury and creating a group sceptical of the operation of the legal system. Certainly, the smaller sizes of minority groups and their relationship to the majority may lead to their under-representation for reasons other than peremptory challenges, but a graphical exploration shows definitively that a component of their under-representation is peremptory challenges. Minority groups are fully struck from the venire by peremptories far more often than majority groups. Striking all majority members may, in fact, be virtually impossible.

All of this paints a bleak picture of the role of peremptory challenges in the modern jury selection process. Without additional work, it is impossible to say with certainty whether the racial patterns observed here and elsewhere are due to racial prejudice by the court, but one may ask the question of whether that matters. As Lord Chief Justice Hewart said in *R. v. Sussex Justices*

Justice should not only be done, but should manifestly and undoubtedly be seen to be done

The visualizations of this paper have made much seen, but it is doubtful that what has been seen here, and by the critics of the *R. v. Stanley* proceedings, looks like justice.

## 5.1 Future Work

One obvious way to extend the work done here is through more thorough modelling. While the multinomial regression model fit in 4.4 served its purpose, much more precise models could be fit using causal graphs. Such causal modelling has the possibility to extend the observations of the model from the simple pattern identification of the multinomial model and visualizations presented here to precise statements about the magnitude of causal effects between factors. Representing the factors in a causal graph would also be a useful

exercise in making the assumptions of the model abundantly clear. Logistic regression models and multinomial models, which have been the norm for peremptory challenge data so far, are less clear about their assumptions, especially to those not trained to fit and analyze these models.

Other possible models of interest are mixed models. In this work the attempts to fit mixed models were not discussed, but at several points models with random effects for each trial were attempted. Unfortunately, these models failed to converge. Not a great deal of time was spent trying to transform the data to facilitate convergence to a reasonable value, and so no mixed models were fit. Such models are attractive because they have the potential to flexibly control for a host of factors which will vary between trials, and do so in a manner which involves minimal parameters. Estimating a random effect for lawyer, for instance, could shed light on how variable lawyers are in their behaviour. This dimension of individual variability is essentially unaddressed by the aggregate examinations of this work.

Another extension would be further investigation of the Sunshine data. It is an incredibly rich data set and this work only examined one small facet of it. The crime classification outlined in 3.1, for example, was never utilized in the analysis, despite the investment of time and effort in performing this clean up. Perhaps this method could also be applied to other irregular data in court cases or elsewhere to efficiently categorize irregular strings.

Finally, as [Wright et al. \(2018\)](#) notes, more data is needed on this topic generally. Further efforts to collect data and reinforce or refute the findings of this work and previous ones should be undertaken, and efforts to centralize and regularize the data would assist in the ease of analysis. Using the visualizations of this paper, such as the mobile plot and positional boxplot, quick and informative comparisons of new data to older data sets could be performed easily. Increased transparency and centralized data collection additionally have the potential to allow for a greater understanding of which elements of the jury trial system work and which are inappropriate. As [Wright et al.](#) puts it:

The transformative power of data, in our view, is not limited to traffic stops or jury selection. We place our proposal in the larger context of using transparency to change criminal justice practices for the better ... the criminal courts already collect useful facts that remain hidden ... scattered ... or inaccessible ... An effort to assemble these facts in aggregate form could improve the courts' efforts to regulate the work of other criminal justice players, such as police and prosecutors.





# Bibliography

- 42nd Parliament of Canada (2018a, March). Bill C-75: An Act to Amend the Criminal Code, Youth Criminal Justice Act and other Acts and to make consequential amendments to other Acts. <http://www.justice.gc.ca/eng/csj-sjc/pl/charter-charte/c75.html> (Accessed February 2019).
- 42nd Parliament of Canada (2018b, November). Bill C75. LEGISinfo. <http://www.parl.ca/LegisInfo> (Accessed February 2019).
- Attorney General's Office of the United Kingdom (2012, November). Jury vetting right of stand-by guidelines. <https://www.gov.uk/guidance/jury-vetting-right-of-stand-by-guidelines-2> (Accessed February 2019).
- Baker, S. G. (1994). The multinomial-poisson transformation. *Journal of the Royal Statistical Society, Series D (The Statistician)* 43(4), 495–504.
- Baldus, D. C., C. M. Grosso, R. Dunham, G. Woodworth, and R. Newell (2012). Statistical proof of racial discrimination in the use of peremptory challenges: The impact and promise of the miller-el line of cases as reflected in the experience of one philadelphia capital case. *Iowa Law Review* 97.
- Baldus, D. C., G. Woodworth, D. Zuckerman, and N. A. Weiner (2001). The Use of Peremptory Challenges in Capital Murder Trials: A Legal and Empirical Analysis. *University of Pennsylvania Journal of Constitutional Law* 3(1).
- Broderick, R. J. (1992). Why the peremptory challenge should be abolished. *Temple Law Review* 65(2), 369.
- Brown, F. L., F. T. McGuire, and M. S. Winters (1978). The peremptory challenge as a manipulative device in criminal trials: Traditional use or abuse. *New England Law Review* 14, 192.
- Brown, R. B. (2000). Challenges for cause, stand-asides, and peremptory challenges in the nineteenth century. *Osgoode Hall Law Journal* 38(3), 453–494.
- Burgess, G. (1992). The divine right of kings reconsidered. *The English Historical Review* 107(425), 837–861.
- Cleveland, W. S. and R. McGill (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society. Series A (General)* 150(3), 192–229.
- Constitution of Canada (1982). Constitution of Canada. <https://laws-lois.justice.gc.ca/eng/Const/index.html> (Accessed February 2019).

- Constitution of the United States (1788). Constitution of the United States. [https://www.senate.gov/civics/constitution\\_item/constitution.htm](https://www.senate.gov/civics/constitution_item/constitution.htm) (Accessed February 2019).
- Flanagan, F. X. (2015). Peremptory Challenges and Jury Selection. *The Journal of Law and Economics* 58.
- Ford, R. (2010). Modeling the effects of peremptory challenges on jury selection and jury verdicts. *George Mason Law Review* 17, 377.
- Forsyth, W. (1994). *History of Trial by Jury* (2 ed.). Lawbook Exchange.
- Friendly, M. (1994). Mosaic plots for multiway contingency tables. *Journal of the American Statistical Association* 89, 190–200.
- Gallup (2018). In depth: Death penalty. Gallup. <https://news.gallup.com/poll/1606/death-penalty.aspx> (Accessed February 2019).
- Government of Canada (1985). Criminal code section 634. <https://laws-lois.justice.gc.ca/eng/acts/C-46/section-634.html> (Accessed February 2019).
- Government of Saskatchewan (1998). Jury act, 1998. <http://www.qp.gov.sk.ca/documents/English/Statutes/Statutes/J4-2.pdf> (Accessed February 2019).
- Grosso, C. M. and B. O'Brien (2012). A Stubborn Legacy: The Overwhelming Importance of Race in Jury Selection in 173 Post-Batson North Carolina Capital Trials. *Iowa Law Review* 97, 1531.
- Hans, V. P. and N. Vidmar (1986). *Judging the Jury* (1 ed.). Plenum Press.
- Harris, K. (2018, February). Liberals review jury selection process after Boushie case uproar. CBC News. <https://www.cbc.ca/news/politics/jury-selection-diversity-indigenous-1.4531792> (Accessed February 2019).
- Hasan, N. R. (2018, April). Eliminating peremptory challenges makes trials less fair. The Star. <https://www.thestar.com/opinion/contributors/2018/04/10/eliminating-peremptory-challenges-make-trials-less-fair.html> (Accessed February 2019).
- Hoffman, M. B. (1997). Peremptory Challenges Should Be Abolished: A Trial Judge's Perspective. *The University of Chicago Law Review* 64(3), 809.
- Iacobucci, F. (2013). First Nations Representation on Ontario Juries: Report of the Independent Review Conducted by the Honourable Frank Iacobucci. Ministry of the Attorney General. <https://www.attorneygeneral.jus.gov.on.ca/english/about/pubs/> (Accessed February 2019).
- Kullback, S. and R. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics* 22(1), 79–86.
- Lang, J. B. (1996). On the comparison of multinomial and poisson log-linear models. *Journal of the Royal Statistical Society, Series B (Methodological)* 58(1), 253–266.
- MacLean, C. (2018, February). Gerald Stanley acquittal renews calls for justice reform 27 years after Manitoba inquiry. CBC News. <https://www.cbc.ca/news/canada/manitoba/aboriginal-justice-inquiry-colten-boushie-gerald-stanley-jury-1.4532394> (Accessed February 2019).

- Macnab, A. (2018, February). Stanley acquittal should not lead to scrapping peremptory challenges, say criminal lawyers. Canadian Lawyer. <https://www.canadianlawyermag.com/legalfeeds/author/aidan-macnab/stanley-acquittal-should-not-lead-to-scrapping-peremptory-challenges-say-criminal-lawyers-15332/> (Accessed February 2019).
- Ministry of the Attorney General of Ontario (2018). The annual jury selection process. Queen's Printer for Ontario. <https://www.attorneygeneral.jus.gov.on.ca/english/courts/jury/> (Accessed February 2019).
- Miriam-Webster (2019a). Miriam-Webster Dictionary Online. <https://www.merriam-webster.com/dictionary/venire> (Accessed February 2019).
- Miriam-Webster (2019b). Miriam-Webster Dictionary Online. <https://www.merriam-webster.com/dictionary/voir%20dire> (Accessed February 2019).
- Morehead, J. W. (1994). When a peremptory challenge is no longer peremptory: Batson's unfortunate failure to eradicate invidious discrimination from jury selection. *DePaul Law Review* 43, 625.
- Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-2.
- North Carolina Sentencing and Policy Advisory Commission (2017). *Classification of a Sample of Offenses*. North Carolina Judicial Branch. <https://www.nccourts.gov/assets/documents/publications/Sample-list-2017.pdf?MAZluXuS0FWod4nFt7zXecJ8Ifu0qEZx> (Accessed February 2019).
- Nunn, K. B. (1993). Rights held hostage: Race, ideology, and the peremptory challenge. *Harvard Civil Rights-Civil Liberties Law Review* 28, 63.
- Page, A. (2005). Batson's blind spot: Unconscious stereotyping and the peremptory challenge. *Boston University Law Review* 85, 155.
- Petersen, C. (1993). Insitutionalized racism: The need for reform of the criminal jury selection process. *McGill Law Journal* 38(1).
- Pilhöfer, A. and A. Unwin (2013). New approaches in visualization of categorical data: R package extracat. *Journal of Statistical Software* 53(7), 1–25.
- Quenneville, G. (2018, February). What happened on Gerald Stanley's farm the day Colten Boushie was shot, as told by witnesses. CBC News. <https://www.cbc.ca/news/canada/saskatoon/what-happened-stanley-farm-boushie-shot-witnesses-colten-gerald-1.4520214> (Accessed February 2019).
- Quenneville, G. and J. Warick (2018a). 'Deck is stacked against us,' says family of Colten Boushie after jury chosen for Gerald Stanley trial. CBC News. <https://www.cbc.ca/news/canada/saskatoon/colten-boushie-shooting-case-jury-selection-gerald-stanley-1.4506931> (Accessed February 2019).
- Quenneville, G. and J. Warick (2018b, February). Shouts of 'murderer' in courtroom after Gerald Stanley acquitted in Colten Boushie shooting. CBC News. <https://www.cbc.ca/news/canada/saskatoon/gerald-stanley-colten-boushie-verdict-1.4526313> (Accessed February 2019).

- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Revesz, J. (2016). Ideological imbalance and the peremptory challenge. *Yale Law Journal* 125.
- Richardson Oakes, A. and H. Davies (2016). Justice must be seen to be done: a contextual reappraisal. *Adelaide Law Review* 37(2), 461–494.
- Roach, K. (2018, April). Ending peremptory challenges in jury selection is a good first step. The Ottawa Citizen. <https://ottawacitizen.com/opinion/columnists/roach-ending-peremptory-challenges-in-jury-selection-is-a-good-first-step> (Accessed February 2019).
- Sheppard, R. (2018). Patriation of the constitution. The Canadian Encyclopedia: Historica Canada. <https://www.thecanadianencyclopedia.ca/en/article/patriation-of-the-constitution> (Accessed February 2019).
- Sison, C. P. and J. Glaz (1995). Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association* 90(429), 366–369.
- Solt, F. and Y. Hu (2018). *dotwhisker: Dot-and-Whisker Plots of Regression Results*. R package version 0.5.0.
- Statistics Canada (2018, November). Table 35-10-0061-01: Crime severity index and weighted clearance rates, police services in Saskatchewan.
- Supreme Court of Canada (1991). *R. v. sherratt*. Supreme Court Judgements. SCC Case Number: 21501; <https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/734/index.do?q=21501> (Accessed February 2019).
- Supreme Court of the United States (1965). *Swain v. Alabama*. <https://supreme.justia.com/cases/federal/us/380/202/> (Accessed February 2019).
- Supreme Court of the United States (1986). *Batson v. Kentucky*. <https://www.law.cornell.edu/supremecourt/text/476/79> (Accessed February 2019).
- Supreme Court of the United States (1993). *J.E.B. v. Alabama*. <https://supreme.justia.com/cases/federal/us/511/127/> (Accessed February 2019).
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2 ed.). Graphics Press.
- Van Dyke, J. M. (1977). *Jury Selection Procedures: Our Uncertain Commitment to Representative Panels* (1 ed.). Ballinger Publishing.
- Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer. ISBN 0-387-95457-0.
- von Moschzisker, R. (1921). The historic origin of trial by jury. *University of Pennsylvania Law Review* 70(1).
- Wegman, E. J. (1990). Hyperdimensional analysis using parallel coordinates. *Journal of the American Statistical Association* 85(411), 664–675.
- Weisstein, E. W. (2018). Chebyshev inequality. MathWorld - A Wolfram Web Resource. <http://mathworld.wolfram.com/ChebyshevInequality.html> (Accessed February 2019).

- Wickham, H. and J. Bryan (2018). *readxl: Read Excel Files*. R package version 1.1.0.
- Wong, B. (2011). Points of view: Color blindness. *Nature Methods* 8(441).
- Woolley, A. (2018). An Ethical Jury? Reflections on the Acquittal of Gerald Stanley for the Murder/Manslaughter of Colten Boushie. *Slaw: Canada's online legal magazine*. <http://www.slaw.ca/2018/02/20/an-ethical-jury-reflections-on-the-acquittal-of-gerald-stanley-for-the-murder-manslaughter-of-colten-boushie/> (Accessed February 2019).
- Wright, R. F., K. Chavis, and G. S. Parks (2018, October). The Jury Sunshine Project: Jury Selection Data as a Political Issue. *University of Illinois Law Review* 2018(4), 1407.
- Zinchuk, B. (2018, March). Both sides wrong about Stanley trial. Prince George Citizen. <https://www.princegeorgecitizen.com/opinion/editorial/both-sides-wrong-about-stanley-trial-1.23199321> (Accessed February 2019).



## Appendix A

# Developing an Effective Visualization of Conditional Probability

As noted in 1 and 4.2, a deficiency of the previous investigations of peremptory challenges was a failure to generate compelling and effective visualizations of the trends in the data. These investigations instead rely on tables to communicate analysis. While such visualizations are not necessarily critical to analysis, they can often be incredibly useful to not only communicate and compare data, but to motivate further investigations in a way which is clearer and more intuitive than a simple table of values.

The first attempt at such a visualization was the mosaic plot (as discussed by Friendly (1994)) using the `mosaicplot` function in the `graphics` package in R [R Core Team (2018)]. Figure A.1 displays this first approach with disposition related to the simplified races of both the defendant and the venire member.

This visualization suffers from a number of limitations, some of which are obvious, and others of which are best explained by the hierarchy of accuracy of visual perception provided in Cleveland and McGill (1987). The obvious limitations are the lack of ability to perceive the differences for the smallest groups, which are compressed enough that their shading is nearly imperceptible. Additionally, the ordering of the axes is incredibly important in how the different areas appear visually, and comparing the different areas is challenging as a result.

This is rather unsurprising. In ranking visual displays by accuracy of perception, Cleveland and McGill (1987) place area low in the hierarchy. Areas are placed below angles, lengths, and positions along common scales. In *The Visual Display of Quantitative Information*, Tufte gives two more sources of possible criticism of the mosaic plot as displayed in Figure A.1: the concept of data-ink and the dimensionality of visualization.

Of the mosaic plot, one may ask how much of the “ink”, or structure, on the page is necessary to communicate the information present. If one has a desire to “above all else show the data” as Tufte does, then these large shaded rectangles, which are likely not perceived accurately according to Cleveland and McGill, seem unnecessary compared to a simpler visualization. This is the concept of “data-ink,” to reduce the complexity of the structures used to display the data.

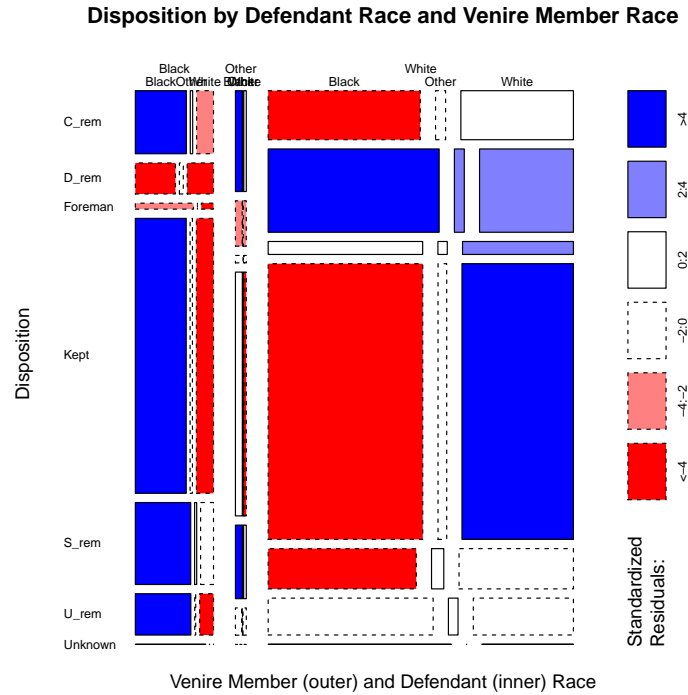


Figure A.1: A mosaic plot of the simplified defendant and venire member race and their relation to the disposition of the venire member.

Hand-in-hand with this concept is [Tuftes](#) rule that the dimensionality of a visualization should not be larger than the data. In the case of the mosaic plot this is not strictly violated, as the marginal lengths used to create the areas reflect a measurement of the data. Nonetheless, the areas of each rectangle correspond to a simple count in a contingency table, and perhaps an area is not the best way to represent such a singular value.

Motivated by these concepts, parallel coordinates (as in [Wegman \(1990\)](#)) were used to visualize the data next, as can be seen in [Figure A.2](#). This attempted visualization is arguably more difficult to interpret than the mosaic plot. It is cluttered by the parallel coordinate lines, the bars emanating from each point obscure the fact that the end point of the bar is the only feature of interest, and the meaning of the black reference line is entirely unclear without extensive explanation. Finally, by viewing the distribution of each disposition, the wrong conditional density is being examined,  $P(Race, Race_{Defendant} | Disposition)$ . Multiple edits and re-conceptualizations of the concept eventually resulted in the “mobile plot,” so-named due to its passing resemblance to the mobiles hung above babies’ cribs.

An example of this plot can be seen in [4.1](#). Note that this plot is less cluttered than either the mosaic plot or the first parallel coordinate plot, despite displaying more information. It is also more efficient with data-ink, avoids displaying data with higher dimensions than the data itself, and uses redundant encoding of information in visual cues which are high in the hierarchy presented by [Cleveland and McGill \(1987\)](#). These redundancies include using both colour and ordering to represent the different dispositions; and using both position along a common scale and length to communicate the conditional probabilities of each disposition.



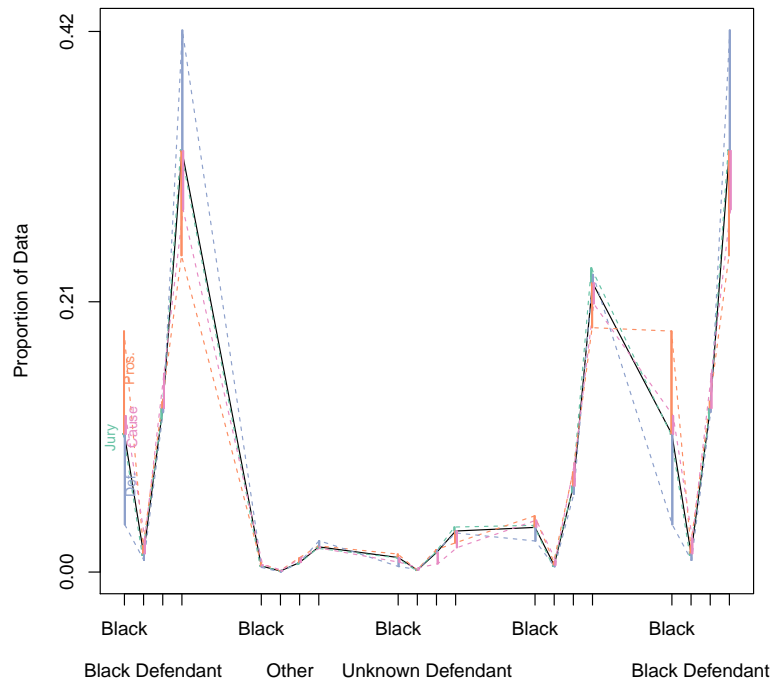


Figure A.2: The first attempt at a parallel coordinate plot. Note that the cramped display and unclear definition of the axis make interpretation even less intuitive than the mosaic plot, suggesting that this first attempt was a decided failure.

An explanation of the features and encoding used in the mobile plot is presented in [A.1](#).

## A.1 The Mobile Plot

The mobile plot consists of multiple grouped vertical lines anchored at one end to horizontal black lines, and at the other to points. Information is encoded using length, colour, and position relative to a common scale. The vertical axis is meant to show the value of a continuous variable, while the horizontal axis shows the value of a, possibly hierarchical, categorical variable. It can be used to display the relationship between three categorical variables in a meaningful two-dimensional plot.

To show the grouping of categories on the horizontal axis, position is used. Those categorical levels which are grouped by some separate categorical variable are placed closer to each other than those which are not in the same group. Each categorical variable combination corresponds to a single horizontal black line, the length of which is proportional to the count of the associated combination in the data being plotted. The vertical position of this line corresponds to the value of the continuous variable expected for that particular combination.

Each of the vertical lines which extend from this horizontal line corresponds to a particular value of a third categorical variable, coloured to show the specific level across the different horizontal lines. The end points of these lines represent the observed distribution of the

third variable for the three way combination of the first two variables associated with the horizontal line to which the points are attached. The lengths of these lines correspond to the deviation of the observation from the expectation. If a different expectation is desired for the different values of the third categorical variable, the horizontal lines can be split evenly and placed vertically at this expectation, to the detriment of grouping clarity.

In the case that such a split is not used and the continuous variable is the probability of a particular value of the third categorical variable given the first two, the plot serves as a visual test of a very specific hypothesis: that of a uniform distribution of the third categorical variable with respect to the two variables represented horizontally. Such a plot is powerful because it allows for the simple detection of main effects and interaction effects over the three categorical variables against this hypothesis.

## Appendix B

# Complementary information

### B.1 Jury Sunshine Irregularities

Table B.1: Jury sunshine data irregularities noted in data flattening

Charges without trial (ACISID)	08CRS50940, 09CRS1106, 10CRS051975, 10CRS51388, 11CRS051642, 11CRS1745, 11CRS51895, 08CRS50113	08CRS52888, 09CRS50752, 10CRS1215, 10CRS51610, 11CRS051795, 11CRS1783, 11CRS52470,	09CRS000305, 10CR52031, 10CRS397, 10CRS52410, 11CRS1577, 11CRS51204, 08CRS54836,
Prosecutors without trials (IDs)	1-000, 11B-000, 12-000, 14-000, 15B-000, 16A-000, 16B-000, 17A-000, 17B-000, 19A-000, 19B-000, 20A-000, 20B-000, 21-000, 22A-000, 22B-000, 24-000, 25-000, 27A-000, 27B-000, 28-000, 29A-000, 29B-000, 30-000, 6-000, 9-000		
Trial missing charge (ID)	710-01		

### B.2 Jury Sunshine Charge Classification

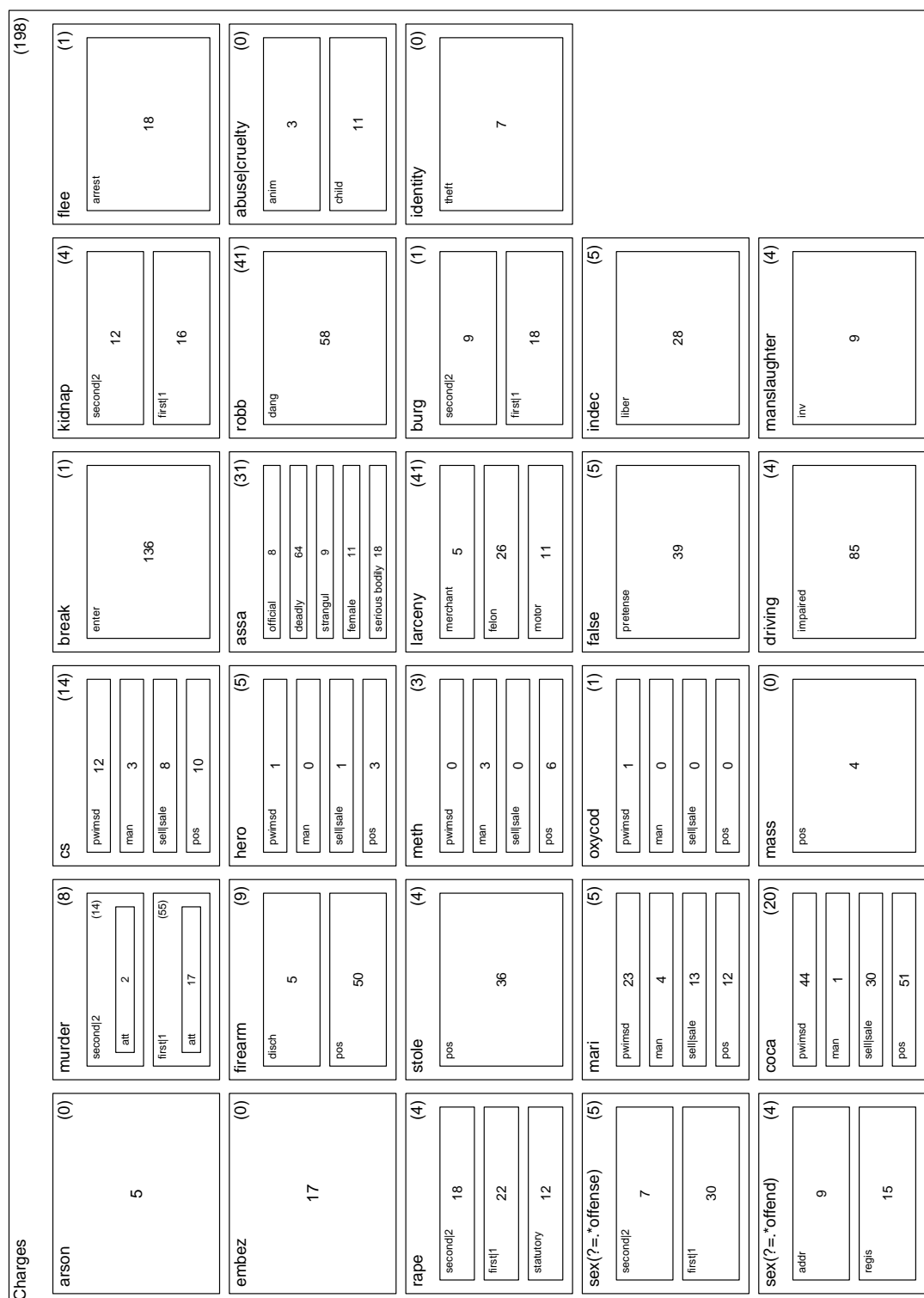


Figure B.1: The regular expression charge tree arranged by hierarchy with counts provided. The counts in brackets indicate the counts of charges which could not be classified to a lower level of the hierarchy

## Appendix C

# Mathematical Results

### C.1 Conditional Distribution of a Poisson Expectation Given Marginal Counts

For simplicity, consider a Poisson random variable  $Y$  with a rate dependent only on one discrete random variable:  $X \in \{1, \dots, m\}$ . Let the count of values observed for  $X = i$  be  $Y_i$  and denote  $E[Y_i|X = i] = \lambda_i$ . Additionally denote the sum of all counts as  $N = \sum_{i=1}^m Y_i$ . The feature of interest is then the distribution of  $Y_1, \dots, Y_m|N = n$ . Or by the definition of the conditional distribution:

$$P(Y_1 = y_1, \dots, Y_m = y_m|N = n) = \frac{P(Y_1 = y_1, \dots, Y_m = y_m, \sum_{i=1}^m Y_i = n)}{P(\sum_{i=1}^m Y_i = n)} \quad (\text{C.1.0.1})$$

Clearly this density is zero if  $\sum_{i=1}^m y_i \neq n$ , so consider its value for  $\sum_{i=1}^m y_i = n$ . Start with the distribution of  $N = \sum_{i=1}^m Y_i$ . Note that for  $A \sim \text{Pois}(\lambda_A)$  and  $B \sim \text{Pois}(\lambda_B)$ , where  $A$  and  $B$  are independent, the distribution of  $A + B$  can be derived quite easily using the characteristic function  $\varphi_{A+B}(t)$ :

$$\varphi_{A+B}(t) = E[e^{it(A+B)}] = E[e^{itA}]E[e^{itB}] = e^{(\lambda_A + \lambda_B)(e^{it} - 1)}$$

This is the characteristic function of a  $\text{Pois}(\lambda_A + \lambda_B)$  variable, and so the sum of two Poisson random variables is a Poisson random variable with a rate corresponding to the sum of the rates of the two variables. Iterating this, then, one obtains  $N \sim \text{Pois}(\sum_{i=1}^m \lambda_i)$ , and so the denominator of C.1.0.1 is:

$$P\left(\sum_{i=1}^m Y_i = n\right) = \frac{e^{-\sum_{i=1}^m \lambda_i} (\sum_{i=1}^m \lambda_i)^n}{n!}$$

Additionally, recognizing that the  $Y_i$  are independent, and considering only the case where  $\sum_{i=1}^m y_i = n$ , as the density is zero otherwise, this can be further simplified, as this joint density can be split into a product of marginal densities:

$$P\left(Y_1 = y_1, \dots, Y_m = y_m, \sum_{i=1}^m y_i = n\right) = P(Y_1 = y_1)P(Y_2 = y_2) \dots P(Y_m = y_m)$$

Now each independent marginal is Poisson distributed, so the product of all of these marginals is:

$$P(Y_1 = y_1)P(Y_2 = y_2) \dots P(Y_m = y_m) = \frac{e^{-\sum_{i=1}^m \lambda_i} \lambda_1^{y_1} \lambda_2^{y_2} \dots \lambda_m^{y_m}}{y_1! y_2! \dots y_m!}$$

And so C.1.0.1 simplifies to

$$\begin{aligned} & \frac{e^{-\sum_{i=1}^m \lambda_i} \lambda_1^{y_1} \lambda_2^{y_2} \dots \lambda_m^{y_m}}{y_1! y_2! \dots y_m!} \cdot \frac{n!}{e^{-\sum_{i=1}^m \lambda_i} (\sum_{i=1}^m \lambda_i)^n} \\ &= \frac{n!}{y_1! y_2! \dots y_m!} \left( \frac{\lambda_1}{\sum_{i=1}^m \lambda_i} \right)^{y_1} \left( \frac{\lambda_2}{\sum_{i=1}^m \lambda_i} \right)^{y_2} \dots \left( \frac{\lambda_m}{\sum_{i=1}^m \lambda_i} \right)^{y_m} \end{aligned} \quad (\text{C.1.0.2})$$

C.1.0.2 is recognizably the multinomial distribution, where the probability of a particular class  $i$  is given by the ratio of  $\lambda_i$  to the sum over all  $\lambda_j$ . ■

# Appendix D

## Code

All code can be found at the author's GitHub repository:

[https://github.com/Salahub/peremptory\\_challenges](https://github.com/Salahub/peremptory_challenges)

### D.1 Data Processing Code

The data processing functions and code are available through the author's GitHub where the processed data is also provided.

### D.2 Analysis Code

As for the data processing code, this code is posted on the author's GitHub.





# Epilogue

I came into this project broadly sceptical of the attempt to change the legal system by the removal of peremptory challenges, fully expecting a thorough analysis to justify my distaste at the politics of the verdict of *R. v. Stanley*. It is easy for a government to take action for political points before considering the consequences. After reading the history of peremptory challenges, studying legal analyses of the practice, and viewing data for myself, however, I am convinced that the abolition of the practice may be the best path forward. [Hoffman \(1997\)](#) was perhaps the best argument I saw from either side of the debate, and [Hoffman](#)'s full-throated support of the complete abolition of peremptory challenges was incredibly influential in developing this view.

That said, as a statistician, I cannot allow myself to say that the empirical analysis in this paper proves anything. As is generally the case with social data, there are so many possible confounders and conflicting factors that it would be easy to mistake a pattern for something it is not. Regardless, my hope is that the visualizations presented are considered useful in the continued investigation of this social phenomenon and others. While the solution may not be clear yet, I believe these displays are useful visual tools to help reach it.



# Declaration of Originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor .

**Title of work** (in block letters):

SEEN TO BE DONE  
A GRAPHICAL EXPLORATION OF PEREMPTORY CHALLENGE

**Authored by** (in block letters):

*For papers written by groups the names of all authors are required.*

**Name(s):**

SALAHUB  
-----  
-----  
-----  
-----

**First name(s):**

CHRISTOPHER  
-----  
-----  
-----  
-----

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the Citation etiquette information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work .
- I am aware that the work may be screened electronically for plagiarism.
- I have understood and followed the guidelines in the document *Scientific Works in Mathematics*.

**Place, date:**

Zürich, Feb. 21  
-----

**Signature(s):**

Christopher Salahub  
-----  
-----  
-----  
-----

*For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.*