

**Technische Universität Dortmund**  
Fakultät Statistik

## **Wissenschaftliches Arbeiten**

### **Bericht über Ergebnisse**

#### **Gruppenmitglieder:**

Daniel Martel  
Grigorii Iakovlev  
Yi Wei Er  
Nur Amirah Khan Binti Rozlan  
Anas Salaheldin

#### **Dozenten:**

Steffan Maletz  
Leia Betting

7. Februar 2025

# Inhaltsverzeichnis

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Einleitung</b>                                 | <b>2</b>  |
| <b>2</b> | <b>Datensatz</b>                                  | <b>2</b>  |
| <b>3</b> | <b>Ergebnis</b>                                   | <b>2</b>  |
| 3.1      | Alter . . . . .                                   | 2         |
| 3.2      | Fahrpreise . . . . .                              | 3         |
| 3.3      | Geschlecht . . . . .                              | 4         |
| 3.4      | Überlebende . . . . .                             | 5         |
| 3.5      | Überlebensrate zwischen Variablen . . . . .       | 5         |
| 3.5.1    | Überlebensrate und Geschlecht . . . . .           | 6         |
| 3.5.2    | Überlebensrate und Fahrklasse . . . . .           | 7         |
| 3.6      | Kombinierte Analyse von 3.5.1 und 3.5.2 . . . . . | 7         |
| 3.7      | Überlebensstatus und Alter . . . . .              | 9         |
| 3.8      | Ticketpreise und Überlebensstatus . . . . .       | 10        |
| <b>4</b> | <b>Fazit</b>                                      | <b>11</b> |

# 1 Einleitung

Der Titanic-Datensatz bietet einen Einblick in die Überlebensmuster während der Katastrophe in 1912. Diese Analyse untersucht den Zusammenhang zwischen dem Überleben und ausgewählten Variablen der Passagiere (Alter, Geschlecht und Ticketpreis) unter Verwendung deskriptiver Statistiken und Visualisierungen.

## 2 Datensatz

Der Datensatz umfasst 891 Passagiere, davon 577 Männer und 314 Frauen auf der Titanic. Es enthält Daten auf individueller Ebene zu demografischen Merkmalen, Reiseinformationen und Überlebensdaten der Passagiere. Genauer zu sagen, der Datensatz umfasst Variablen wie Alter, Geschlecht, Passagierklasse, Ticketpreis, Einschiffungshafen und Überlebensstatus. Die meisten Passagiere reisten ohne Geschwister/Ehepartner ( $SibSp = 0$ ) und ohne Eltern/Kinder ( $Parch = 0$ ). Die Mehrheit ging in Southampton (S) an Bord. Weniger Passagiere gingen in Cherbourg (C) und Queenstown (Q) an Bord. Es wurden fehlende Werte identifiziert. Die meisten fehlenden Daten waren auf die Variable "Cabin".

## 3 Ergebnis

### 3.1 Alter

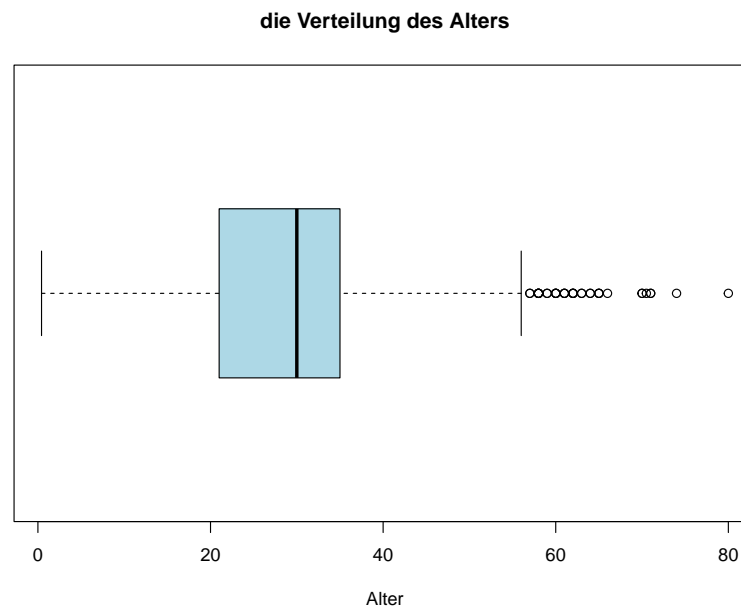


Abbildung 1: Boxplot für die Verteilung des Alters

Basierend auf dem Boxplot zeigt sich, dass das durchschnittliche Alter etwa 29,4 Jahre beträgt, während der Median liegt bei 30 Jahre. Das minimale Alter beträgt 0,42 Jahre, während das maximale Alter 80 Jahre ist. Es gibt eine moderate Variabilität mit Standardabweichung = 13.3. Der Boxplot zeigt eine rechtsschiefe Verteilung, da der rechte Whisker länger ist und mehrere Ausreißer bei hohen Alterswerten auftreten.

Deswegen gibt es mehr jüngere als ältere Passagiere.

### 3.2 Fahrpreise

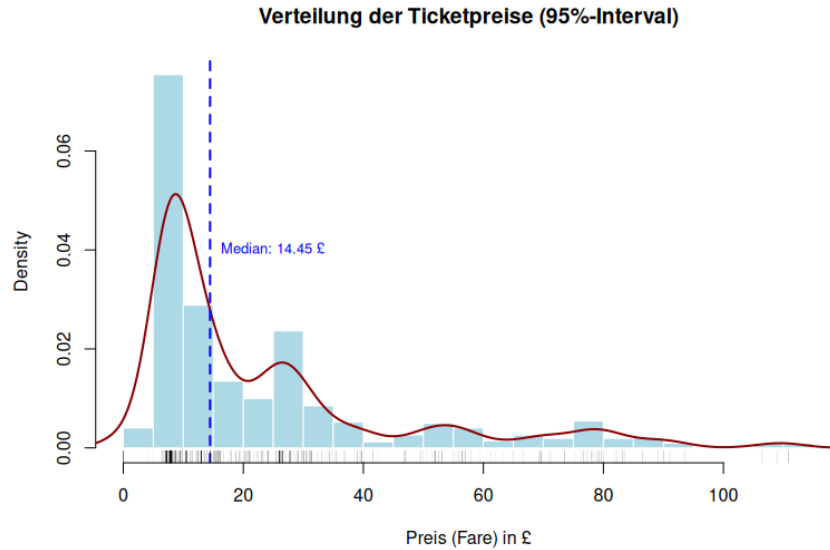


Abbildung 2: Die Verteilung der Ticketpreise

Hier wurden die Fahrpreise unter Berücksichtigung eines 95%-Intervalls analysiert. Der durchschnittliche Ticketpreis beträgt 32,2 Pfund, während der Median bei 14,45 Pfund liegt. Die Abbildung zeigt eine deutlich rechtsschiefe Verteilung, was darauf hinweist, dass der Mittelwert über dem Median liegt. Der Großteil der Ticketpreise befindet sich unterhalb von 31 Pfund, während einzelne extreme Werte bis zu 512,33 Pfund zu einer sehr hohen Streuung beitragen.

### 3.3 Geschlecht

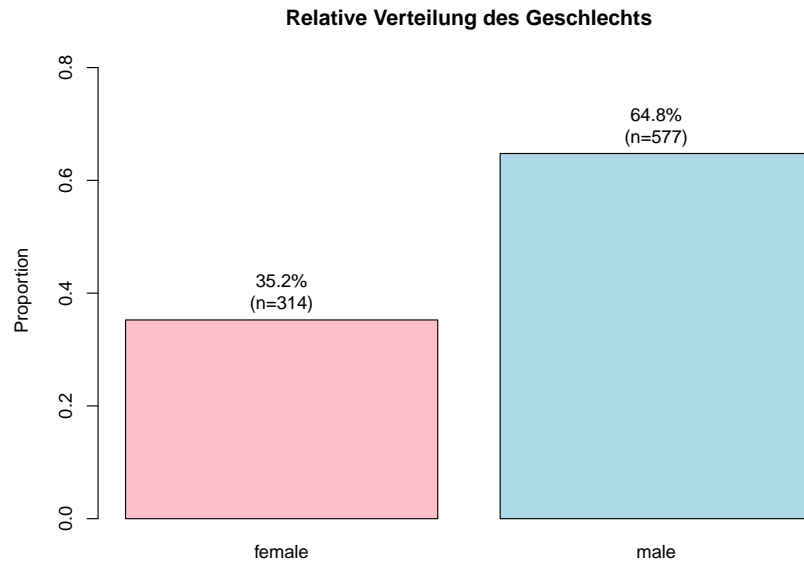


Abbildung 3: Barplot für die Verteilung des Geschlechts

Dieser Abschnitt illustriert die Verteilung des Geschlechts. Die meisten Passagiere waren Männer, mit einem Anteil von 64.8 Prozent. Der Anteil der weiblichen Passagiere betrug jedoch 35.2 Prozent. Deswegen gibt es mehr Männer als Frauen an Bord.

### 3.4 Überlebende

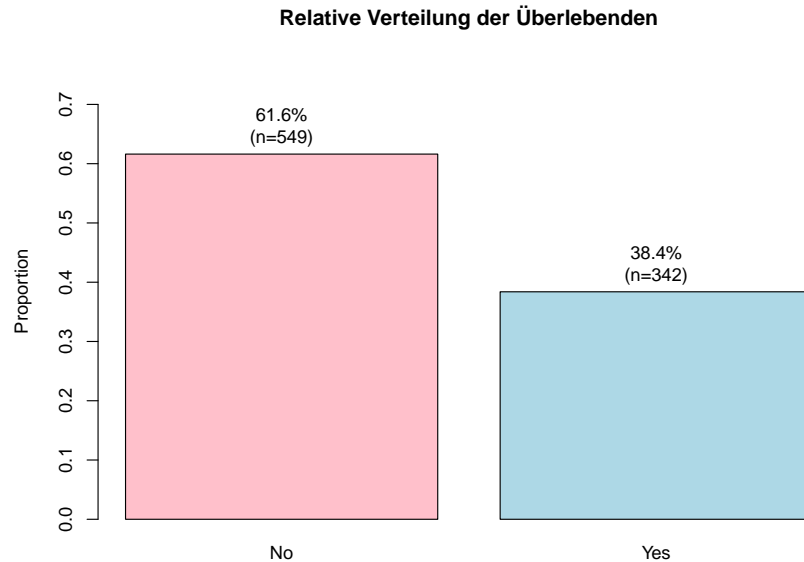


Abbildung 4: Barplot für die Verteilung der Überlebende

Diese Abschnitt zeigt die relative Verteilung der Überlebenden der Titanic. Aus der Abbildung geht hervor, dass etwa 38.38 Prozent der Passagiere den Untergang der Titanic überlebt haben, während rund 61.62 Prozent nicht überlebt haben. Damit stellt die Gruppe der Nicht-Überlebenden die Mehrheit der Passagiere dar.

### 3.5 Überlebensrate zwischen Variablen

In dieser Abschnitt wurden Überlebensrate gegen andere Variablen analysieren. Was besonderes zu betrachten waren das Verhältnis von Überlebensrate gegen Geschlecht und Klasse (bzw. Fahrpreise). Erstens wurde die Analyse über Überlebensrate und Geschlecht diskutiert und anschließend wurde die Untersuchung zwischen Überleben und Fahrklasse beobachtet. Zusätzlich wurde das Muster zwischen Überleben und Alter untersucht. Am Ende wurde Ticketpreise und Überlebensstatus auch betrachtet.

### 3.5.1 Überlebensrate und Geschlecht

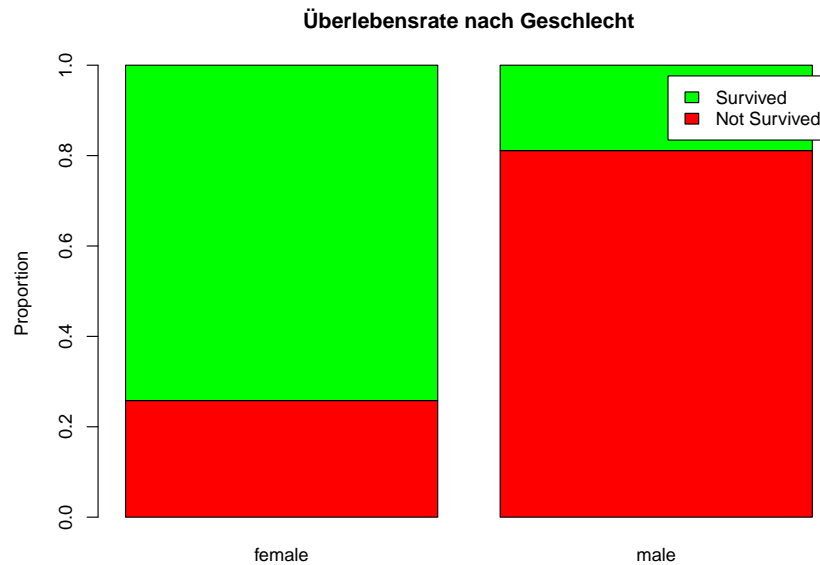


Abbildung 5: Stacked barplot für die Überlebensrate nach Geschlecht

Die Abbildung zeigt die Verteilung von Überlebensrate nach Geschlecht. Frauen wiesen deutlich höhere Überlebensrate als männliche Passagiere auf. Während die Mehrheit der männlichen Passagiere die Katastrophe nicht überlebte, überlebte ein großer Teil der weiblichen Passagiere. Das Geschlecht scheint einer der stärksten Prädiktoren für das Überleben in dem Datensatz zu sein, obwohl es wahrscheinlich mit anderen Faktoren wie der Passagierklasse interagiert.

### 3.5.2 Überlebensrate und Fahrklasse

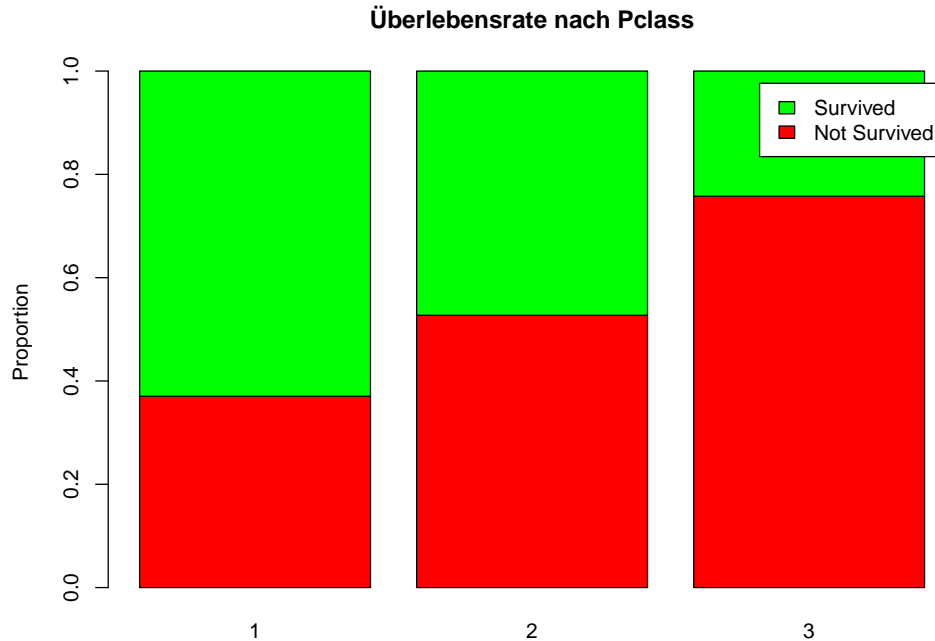


Abbildung 6: Stacked Barplot für die Überlebensrate nach Pclass

Kritisch zu beobachten: Passagiere, die die Katastrophe überlebten, zahlten im Durchschnitt höhere Fahrpreise (in Pclass = 1) als diejenigen, die nicht überlebten. Passagiere der ersten Klasse hatten den höchsten Anteil an Überlebenden, gefolgt von Passagieren der zweiten Klasse, während Passagiere der dritten Klasse die niedrigsten Überlebensraten hatten.

Die Verteilung der Reiseklassen unter den Überlebenden ist zugunsten der höheren Klassen verzerrt, während sich die Nichtüberlebenden stärker auf die niedrigeren Reiseklassen konzentrieren. Dieses Muster deutet darauf hin, dass Überlebende im Durchschnitt höhere Fahrpreise bezahlt haben als Nichtüberlebende. Dies könnte auch darauf hindeuten, dass der sozioökonomische Status, der sich im Ticketpreis und in der Klasse widerspiegelt, eine wichtige Rolle für die Überlebenswahrscheinlichkeit gespielt hat. Höhere Fahrpreise sind eng mit höheren Passagierklassen verbunden, die während der Katastrophe möglicherweise einen besseren Zugang zu Informationen, Ressourcen und Evakuierungsmöglichkeiten geboten haben.

## 3.6 Kombinierte Analyse von 3.5.1 und 3.5.2

Um die ersten beiden Analysen besser zu korrelieren, wurde Geschlecht und Fahrklasse gegen Überlebensstatus in einem Grafik visualisiert.

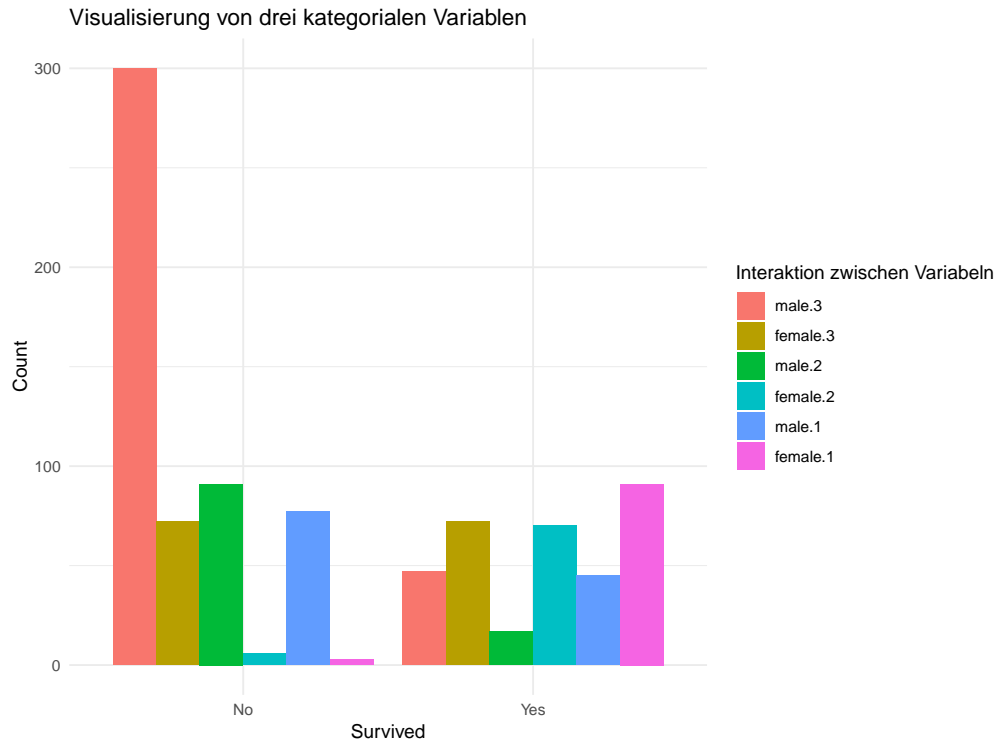


Abbildung 7: Visualisierung von drei kategorialen Variablen

Aus dieser Abbildung ist es klar gezeigt, dass die Frauen aus der erste Klasse die höchsten Überlebenschancen aus die andere hatten. Im Vergleich dazu weisen die Männer aus die dritten Klasse die niedrigsten Überlebenschancen auf. Insgesamt zeigt die Darstellung, dass Geschlecht und Passagierklasse in einem deutlichen Zusammenhang mit dem Überlebensstatus der Passagiere standen.

### 3.7 Überlebensstatus und Alter

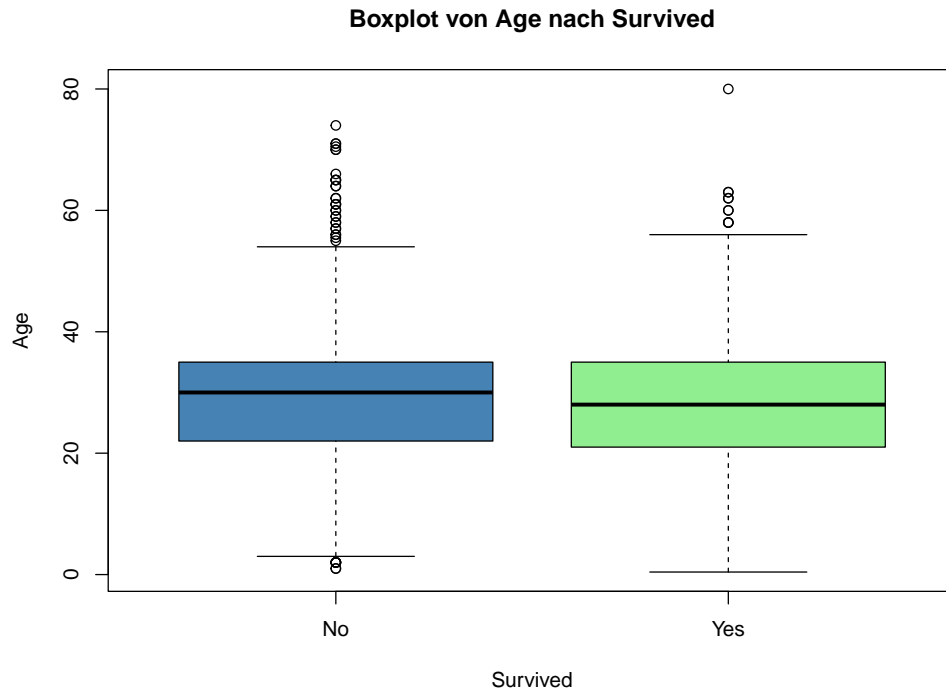


Abbildung 8: Boxplot von Age nach Survived

Dieser Abschnitt analysiert die Zusammenhang zwischen dem Alter der Passagiere und ihrem Überlebensstatus. Der Boxplot zeigt, dass die Überlebende Passagiere im Median jünger waren als nicht überlebende. Die Streuung des Alters ist in beiden Gruppen vergleichbar, wobei in beiden Fällen Ausreißer bei hohen Alterswerten auftreten. Insgesamt ist der Altersunterschied nicht sehr groß, jedoch hatten die jüngere Passagiere tendenziell bessere Überlebenschancen. Dieser Trend steht im Einklang mit dem Protokoll „Frauen und Kinder zuerst“, das bei der Evakuierung der Titanic befolgt wurde.

### 3.8 Ticketpreise und Überlebensstatus

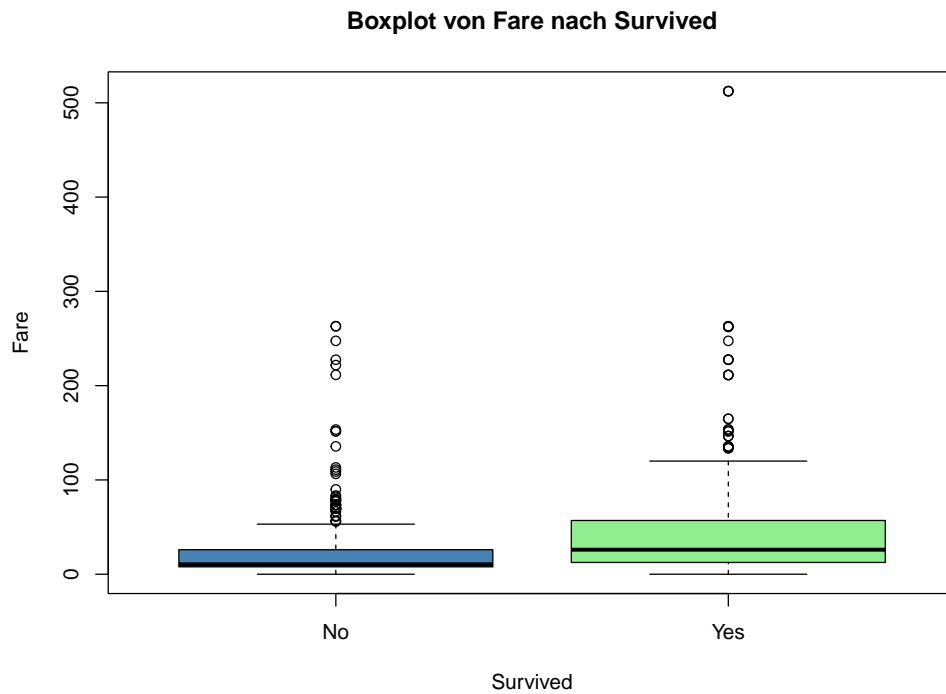


Abbildung 9: Boxplot von Fare nach Survived

Der Boxplot zeigt, dass überlebende Passagiere im Median höhere Ticketpreis zahlten als nicht überlebende. Die ausgeprägte Rechtsschiefe der Verteilung mit mehreren hohen Ausreißern führt zu einer deutlich größeren Streuung bei den Überlebenden. Insgesamt weist die Darstellung auf einen Zusammenhang zwischen Ticketpreis und Überlebenswahrscheinlichkeit hin. Dies deutet darauf hin, dass sie bei der Evakuierung Vorrang hatten. Dies steht auch in Zusammenhang mit der Analyse der Ticketklassen.

## 4 Fazit

Die Gesamtanalyse zeigt einen signifikanten Einfluss von Geschlecht, Alter und wirtschaftlichem Status auf die Überlebenswahrscheinlichkeit. Frauen, insbesondere in jüngeren Altersgruppen, hatten höhere Überlebensraten. Darüber hinaus hatten Passagiere der höheren Klassen bessere Überlebenschancen als diejenigen der unteren Klassen, was den Einfluss der Klasse auf das Überleben während der Titanic-Katastrophe unterstreicht.