

Wissenschaftlicher Bericht: Analyse des Titanic-Datensatzes

Kimia Safaei Jule Dahmke Hanna Engnath
Samuel Kharistafri Charlotte Scholz

8. Februar 2026

Inhaltsverzeichnis

1 Einleitung	2
2 Deskriptive Statistik	2
2.1 Analyse der metrischen Variablen	2
2.2 Analyse der kategorialen Variablen	2
3 Bivariate Statistik: Zusammenhänge	2
3.1 Überlebensrate nach Geschlecht und Klasse	2
3.2 Zusammenhang zwischen Passagierklasse und Überleben	3
4 Visuelle Analyse (Metrisch vs. Dichotom)	3
4.1 Vergleich von SibSp und Fare nach Überlebensstatus	3
5 Multivariate Analyse	5
6 Fazit	5

1 Einleitung

In diesem Bericht wird der bereinigte Titanic-Datensatz mithilfe von verschiedenen R-Funktionen analysiert. Ziel ist es, deskriptive Statistiken zu erstellen und Zusammenhänge zwischen den Variablen, insbesondere in Bezug auf die Überlebensrate, zu untersuchen.

2 Deskriptive Statistik

2.1 Analyse der metrischen Variablen

Die Auswertung der Variablen `Age` und `Fare` zeigt eine diverse Passagierstruktur. Während das Alter mit einem Mittelwert von 29,39 Jahren relativ symmetrisch verteilt ist, weist der Ticketpreis (`Fare`) eine starke Rechtsschiefe auf (Median: 14,45 vs. Mittelwert: 32,20). Dies deutet auf extreme Ausreißer im Bereich der Luxusklassen hin.

2.2 Analyse der kategorialen Variablen

Die Verteilung der kategorialen Merkmale verdeutlicht die Zusammensetzung an Bord:

- **Geschlecht:** Dominanz männlicher Passagiere (64,76 %).
- **Zustiegshafen:** Die Mehrheit (72,28 %) stieg in Southampton zu.
- **Überlebensstatus:** Die Sterberate liegt bei 61,62 %, was die Schwere der Katastrophe unterstreicht.

3 Bivariate Statistik: Zusammenhänge

3.1 Überlebensrate nach Geschlecht und Klasse

Die folgende Tabelle zeigt die Verteilung der Überlebenden nach Geschlecht. Es wird deutlich, dass Frauen eine signifikant höhere Überlebenschance hatten.

Tabelle 1: Kontingenztafel: Überlebensstatus nach Geschlecht

Überlebt	Mann	Frau	Summe
Nein	468	81	549
Ja	109	233	342
Summe	577	314	891

3.2 Zusammenhang zwischen Passagierklasse und Überleben

Neben dem Geschlecht spielte die Klasse eine zentrale Rolle. In der dritten Klasse verstarben ca. 75 % der Passagiere.

Tabelle 2: Kontingenztafel: Überlebensstatus nach Passagierklasse

Überlebt	Klasse 1	Klasse 2	Klasse 3	Summe
Nein	80	97	372	549
Ja	136	87	119	342
Summe	216	184	491	891

4 Visuelle Analyse (Metrisch vs. Dichotom)

4.1 Vergleich von SibSp und Fare nach Überlebensstatus

In diesem Abschnitt werden die metrischen Variablen in Abhängigkeit vom Überlebensstatus visualisiert.

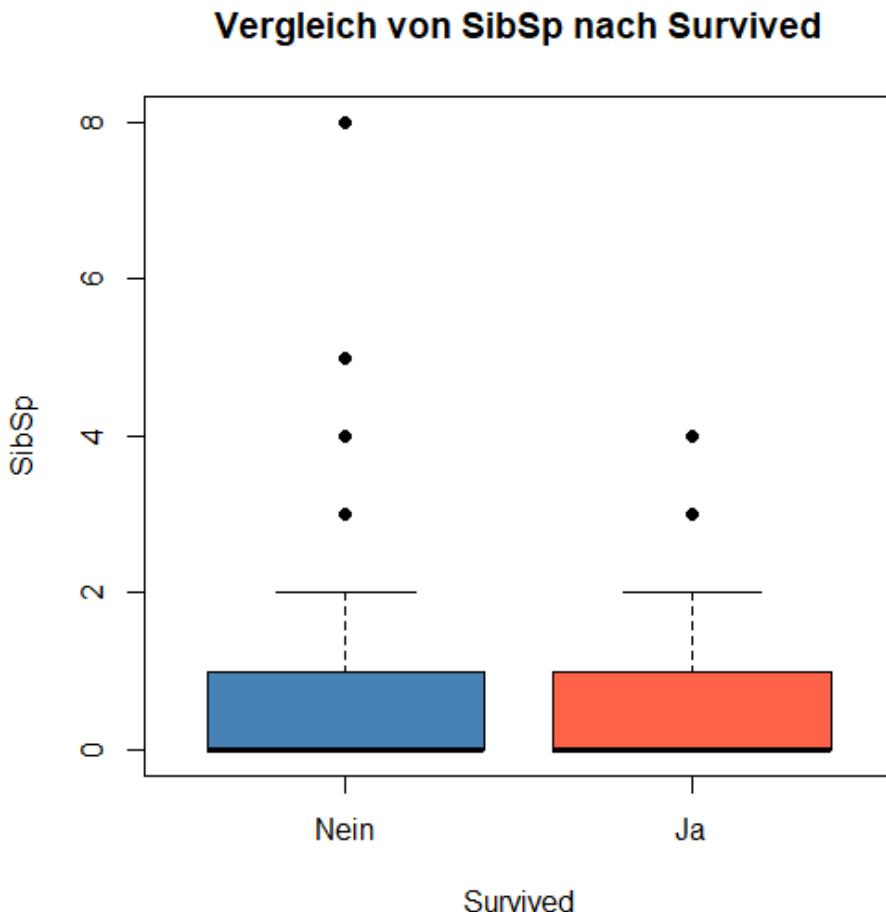


Abbildung 1: Anzahl der Begleiter (SibSp) nach Überlebensstatus

Der Boxplot für SibSp zeigt, dass die Anzahl der Begleiter keinen massiven Einfluss auf die Überlebenschance hatte, obwohl extreme Ausreißer in der Gruppe der Verstorbenen existieren.

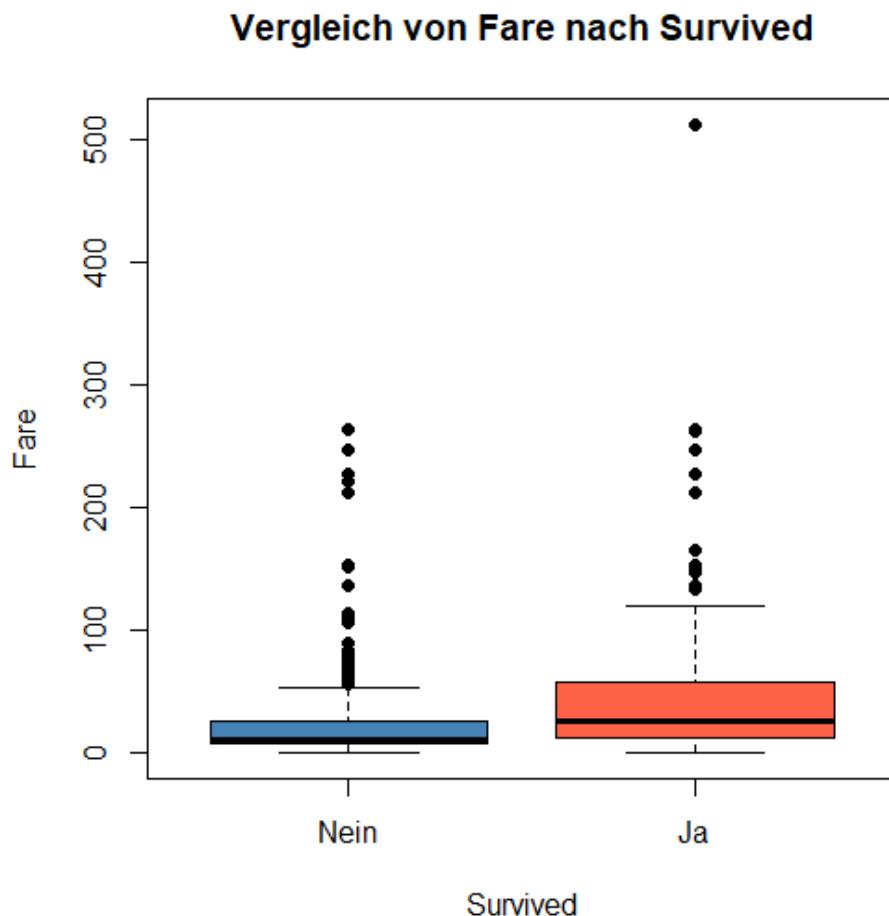


Abbildung 2: Ticketpreise (Fare) nach Überlebensstatus

Im Gegensatz dazu zeigt der Boxplot für Fare, dass Überlebende im Median deutlich teurere Tickets besaßen.

5 Multivariate Analyse

Der Mosaicplot verdeutlicht das Zusammenspiel zwischen Klasse, Hafen und Geschlecht.

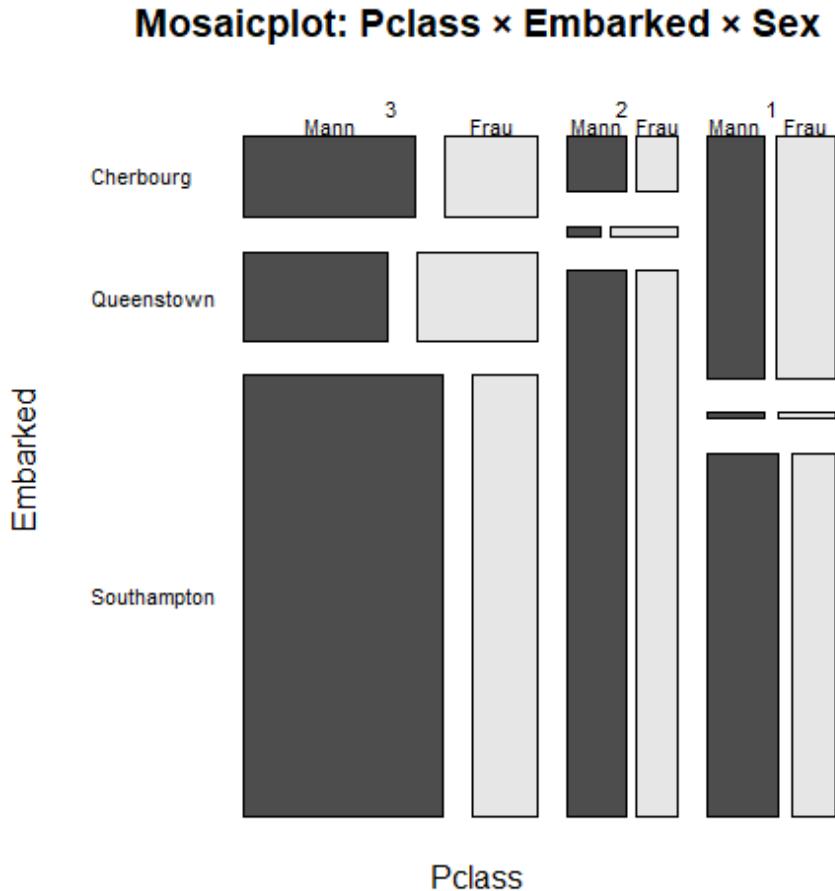


Abbildung 3: Mosaicplot: Pclass, Embarked und Sex

6 Fazit

Die Analyse bestätigt, dass die Überlebenschance auf der Titanic stark von sozioökonomischen Faktoren und dem Geschlecht abhing. Wohlhabendere Passagiere und Frauen wurden bei der Rettung signifikant bevorzugt.