Visualizing Two Categorical Variables

Matthew Renze
SOFTWARE CONSULTANT

@matthewrenze   www.matthewrenze.com

Overview

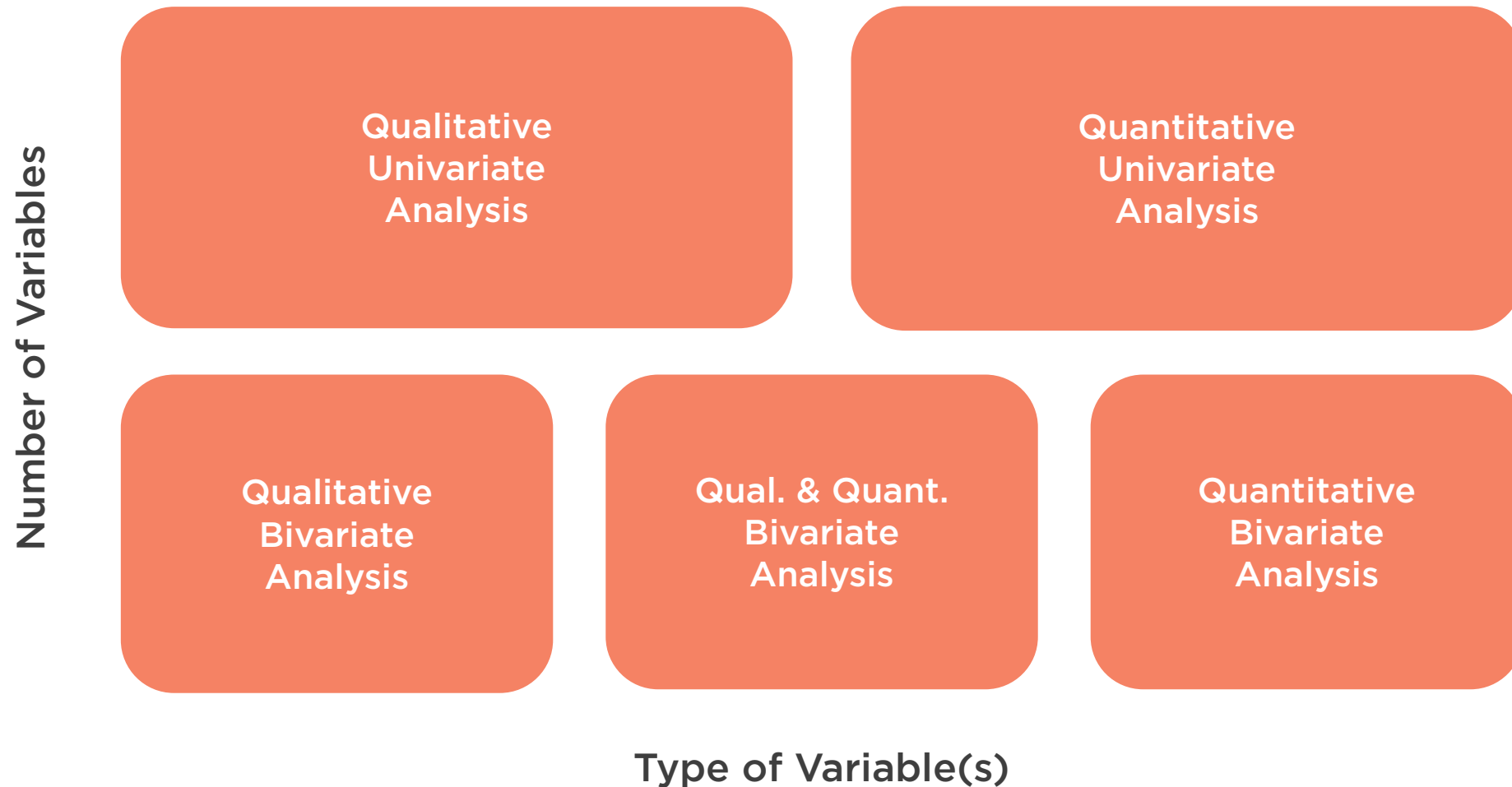**Qualitative Bivariate Analysis**

Demo (Base)

Demo (Lattice)

Demo (ggplot2)

Types of Data Analysis

Number of Variables

Qualitative
Univariate
Analysis

Quantitative
Univariate
Analysis

Qualitative
Bivariate
Analysis

Qual. & Quant.
Bivariate
Analysis

Quantitative
Bivariate
Analysis

Type of Variable(s)

Qualitative Bivariate Analysis

Joint frequency

Marginal frequency

Movies by Genre and Rating					
Genre	G	PG	PG-13	R	Total
Action	2	70	311	229	612
Adventure	44	179	209	64	496
Animation	43	111	8	6	168
Comedy	45	258	472	506	1218
Drama	12	136	586	836	1570
Family	38	181	10	1	230
...	...	...	...	...	...
Total	230	1207	2686	3058	7181

Qualitative Bivariate Analysis

Joint frequency

Marginal frequency

Relative frequency

Movies by Genre and Rating					
Genre	G	PG	PG-13	R	Total
Action	0.001	0.010	0.043	0.032	0.086
Adventure	0.006	0.025	0.029	0.009	0.069
Animation	0.006	0.015	0.001	0.001	0.023
Comedy	0.006	0.036	0.066	0.070	0.170
Drama	0.002	0.019	0.082	0.116	0.219
Family	0.005	0.025	0.001	0.001	0.033
...	...	...	...	...	...
Total	0.032	0.168	0.374	0.426	1.000

Grouped Frequency Bar Chart

Joint frequency

Across categories

Within categories

Count of Movies by Rating and Awards

Awards
■ No
□ Yes

Count of Movies

1500

1000

500

0

G   PG   PG-13   R

Rating
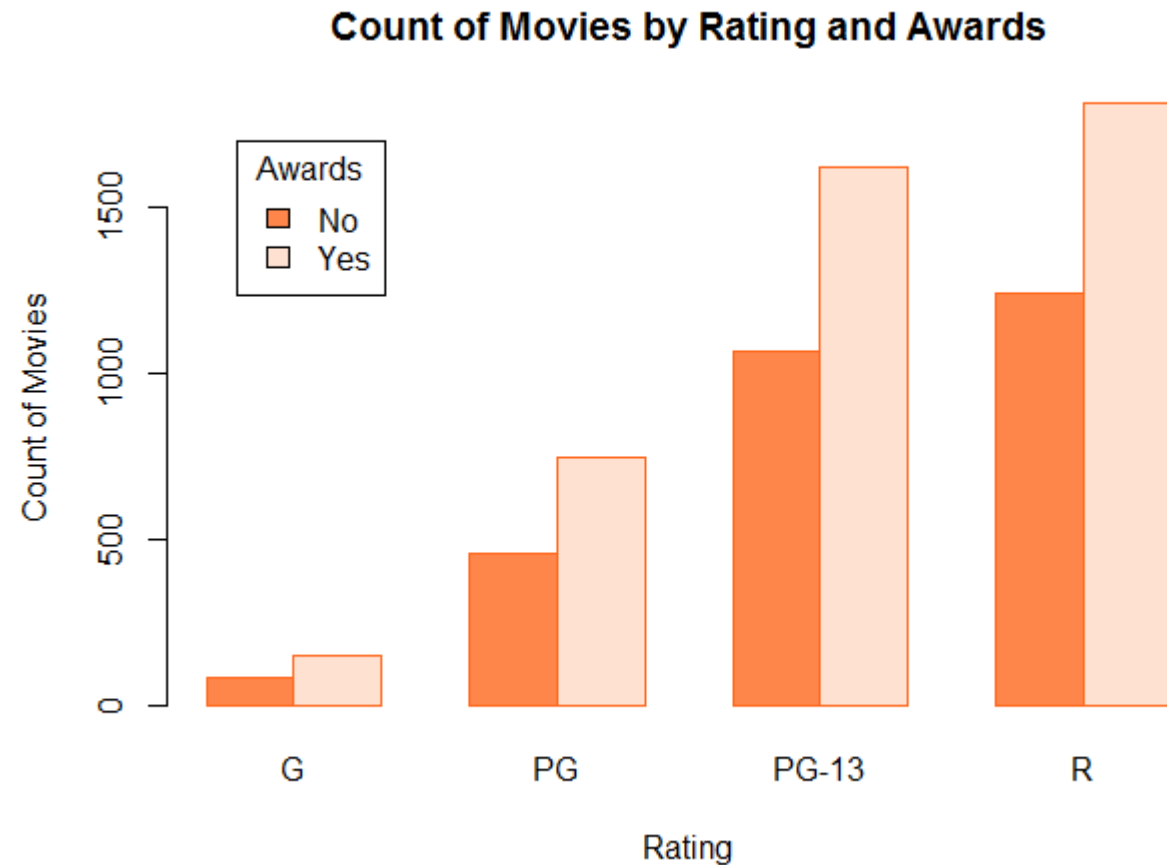
Stacked Frequency Bar Chart

Marginal frequency

Joint frequency

Part of whole

Count of Movies by Rating and Awards
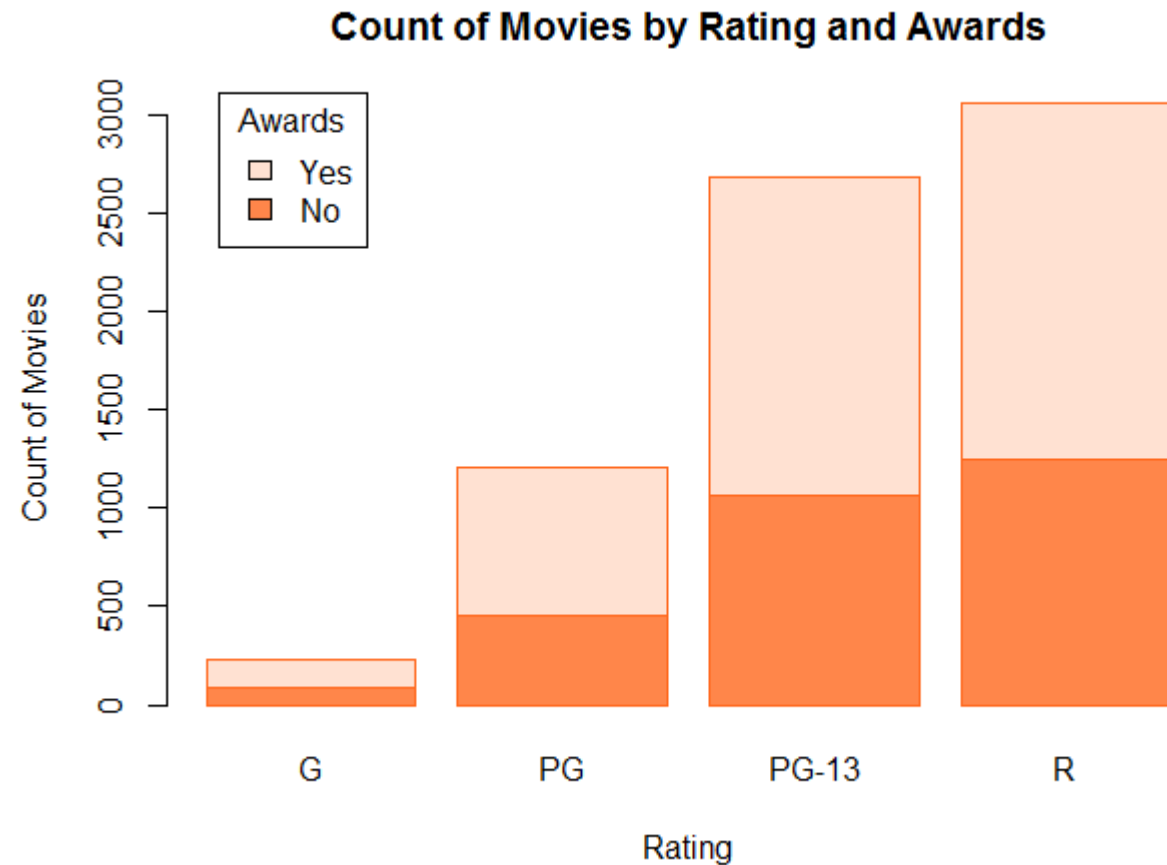
100% Stacked Frequency Bar Chart

Relative frequency

One variable

Proportion of Movies by Rating and Awards
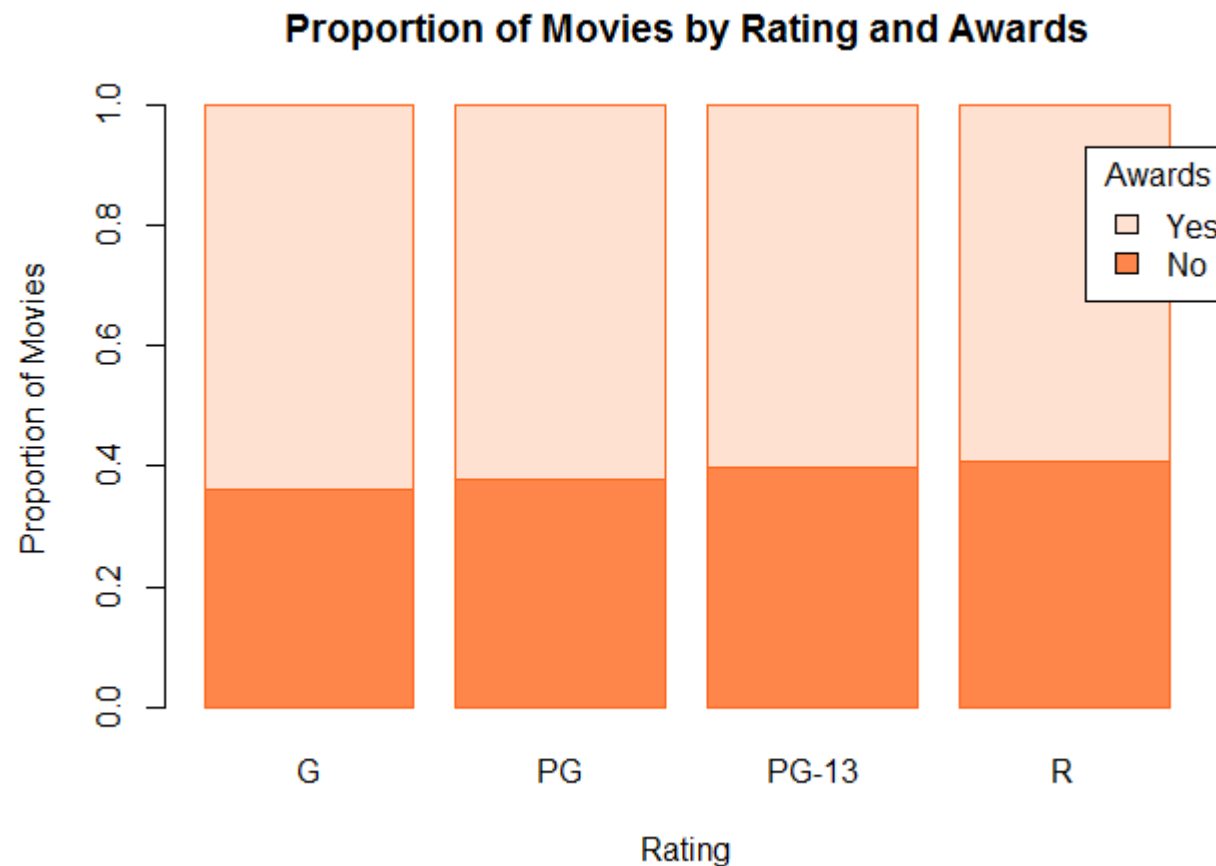
Spine Plot

Relative frequency

Both variables

Proportion of Movies by Rating and Awards
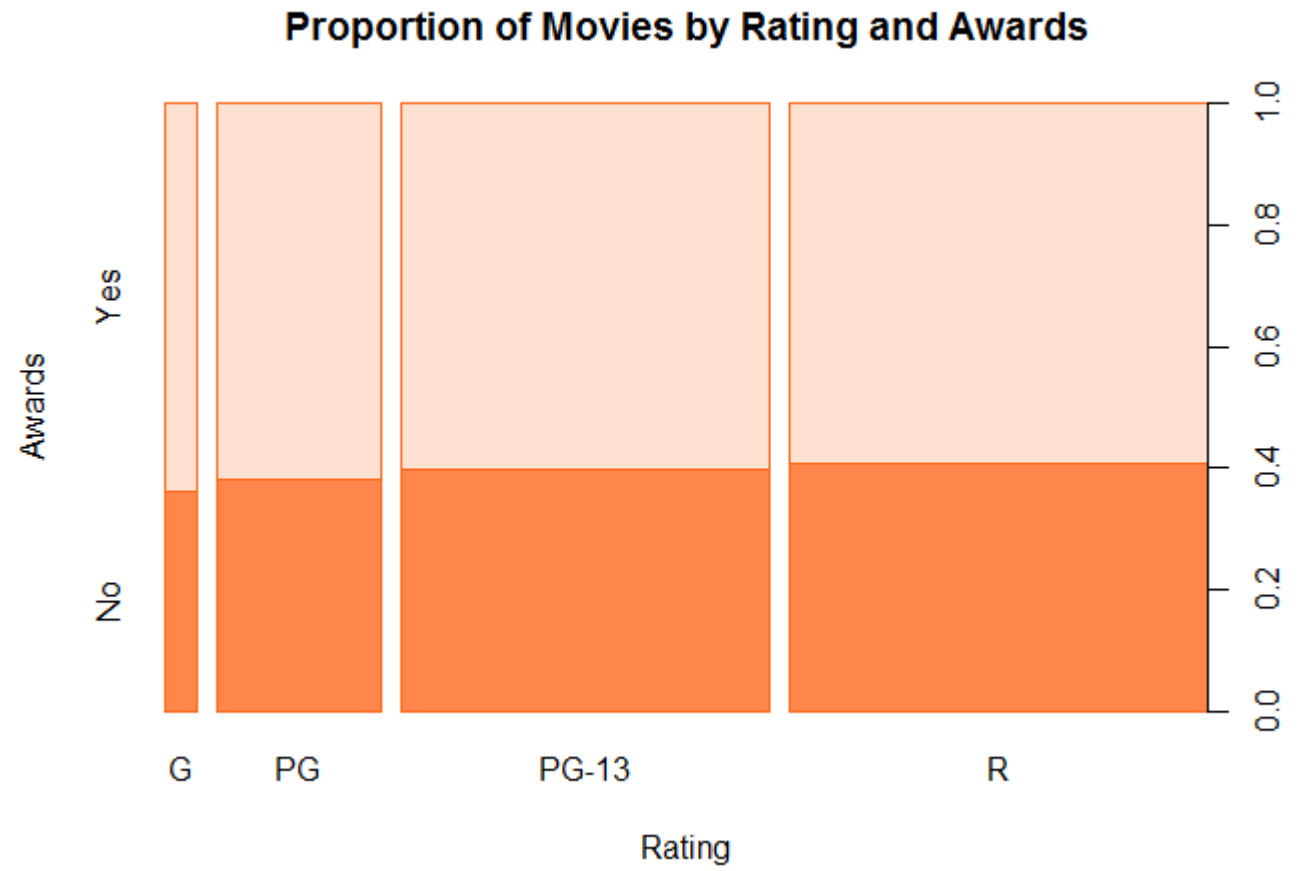
Mosaic Plot

Clean aesthetics

Proportion of Movies by Rating and Awards

Awards: Yes / No

Rating: G, PG, PG-13, R
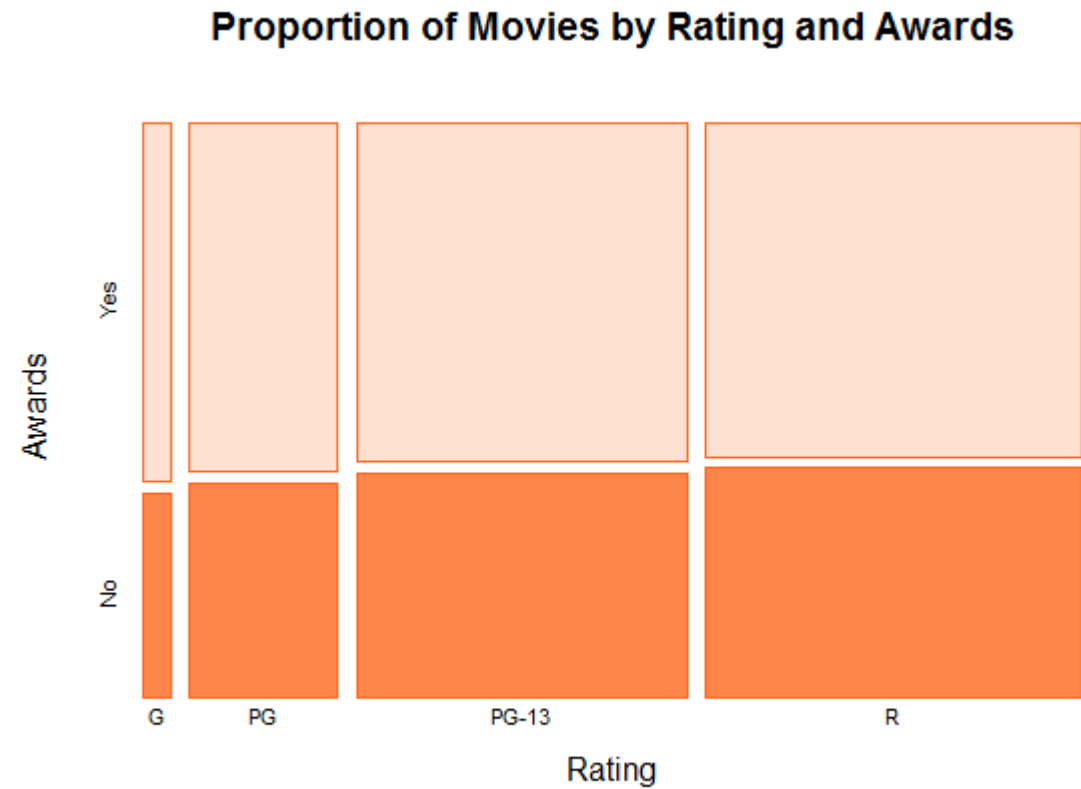
1. How many award-winning movies are in each rating category?

2. What proportion of movies in each rating category have won awards?

Load the Data

Create a Contingency Table

Create a Grouped Frequency Bar Chart

Create a Stacked Frequency Bar Chart

Create a Relative Frequency Table

Create a 100% Stacked Frequency Bar Chart

Create a Second Contingency Table

Create a Spine Plot

Create a Mosaic Plot

Load the Lattice Package

Create a Grouped Frequency Bar Chart

Create a Stacked Frequency Bar Chart

Create a Relative-frequency Contingency Table

Create a 100% Stacked Frequency Bar Chart

No Spine Plot or Mosaic Plot

Load ggplot2 Package

Create a Grouped Frequency Bar Chart

Create a Stacked Frequency Bar Chart

Create a 100% Stacked Frequency Bar Chart

Summary

Qualitative Bivariate Analysis

Demo (Base)

Demo (Lattice)

Demo (ggplot2)