

# Visualizing One Categorical Variable

---



**Matthew Renze**

SOFTWARE CONSULTANT

@matthewrenze [www.matthewrenze.com](http://www.matthewrenze.com)



# Overview



## Qualitative Univariate Analysis

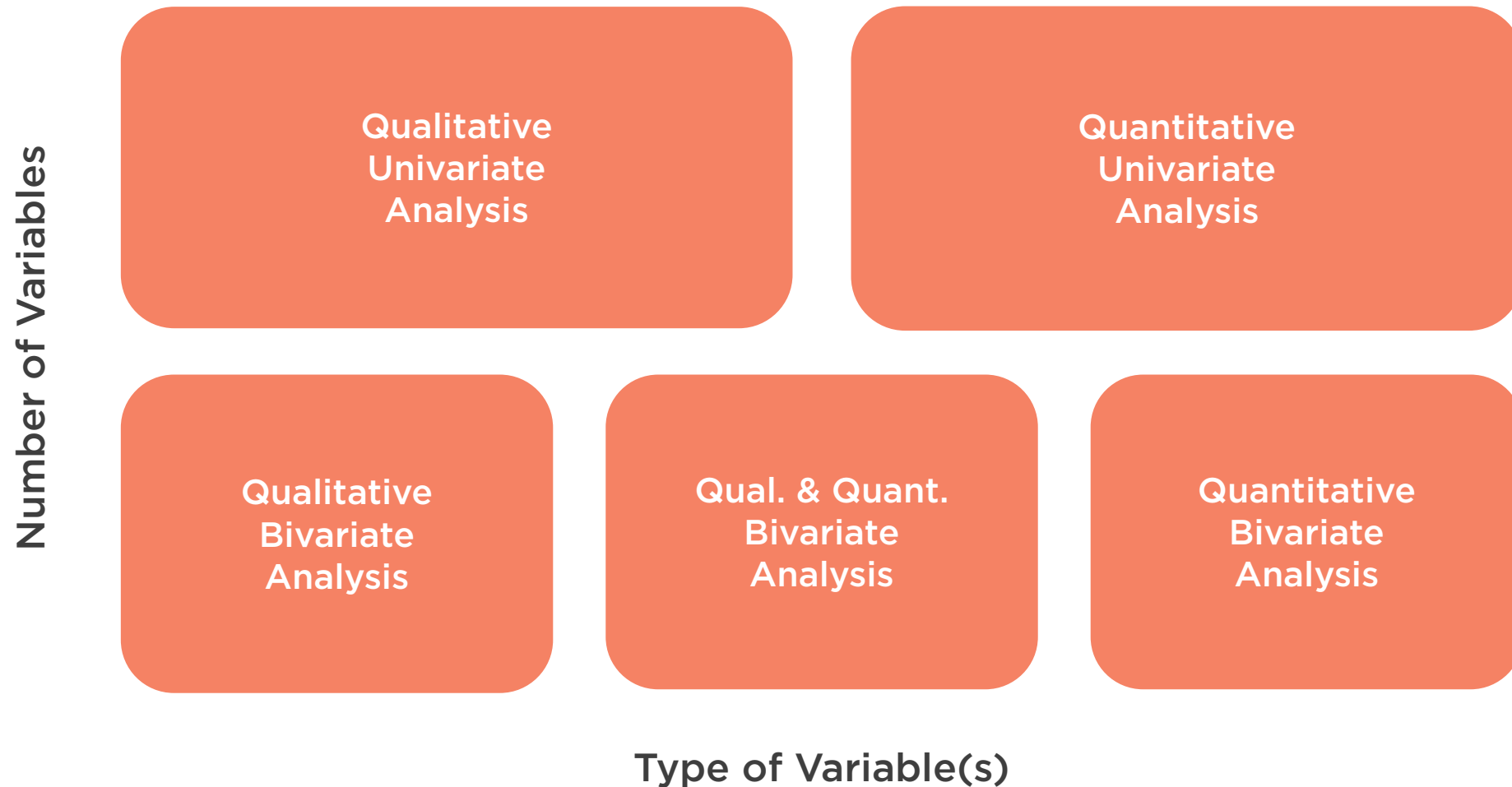
Demo (Base)

Demo (Lattice)

Demo (ggplot2)



# Types of Data Analysis



| Movies               |      |        |                      |        |                 |               |        |                    |
|----------------------|------|--------|----------------------|--------|-----------------|---------------|--------|--------------------|
| Title                | Year | Rating | Runtime<br>(minutes) | Genre  | Critic<br>Score | Box<br>Office | Awards | Inter-<br>national |
| The Whole Nine Yards | 2000 | R      | 98                   | Comedy | 45%             | \$57.3M       | No     | No                 |
| Cirque du Soleil     | 2000 | G      | 39                   | Family | 45%             | \$13.4M       | Yes    | No                 |
| Gladiator            | 2000 | R      | 155                  | Action | 76%             | \$187.3M      | Yes    | Yes                |
| Dinosaur             | 2000 | PG     | 82                   | Family | 65%             | \$135.6M      | Yes    | No                 |
| Big Momma's House    | 2000 | PG-13  | 99                   | Comedy | 30%             | \$0.5M        | Yes    | Yes                |





**PROD. NO.**  
**SCENE**

**TAKE**

**ROLL**

# Qualitative Univariate Analysis

Frequency

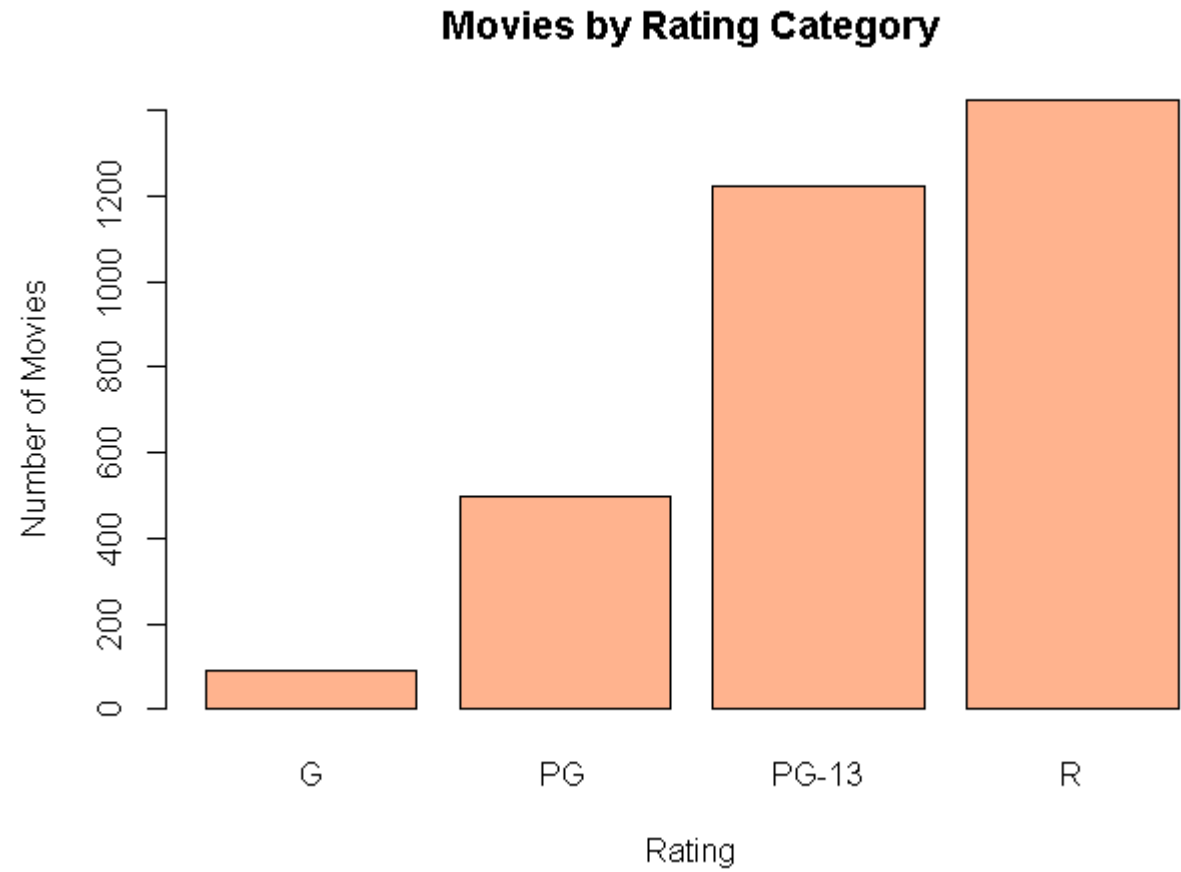
Proportion

| Movies by Genre |           |            |
|-----------------|-----------|------------|
| Genre           | Frequency | Percentage |
| Action          | 612       | 9%         |
| Adventure       | 496       | 7%         |
| Animation       | 168       | 2%         |
| Comedy          | 1281      | 18%        |
| Drama           | 1570      | 22%        |
| Horror          | 269       | 4%         |
| ...             | ...       | ...        |



# Frequency Bar Chart

Frequency  
Comparison  
Min and max

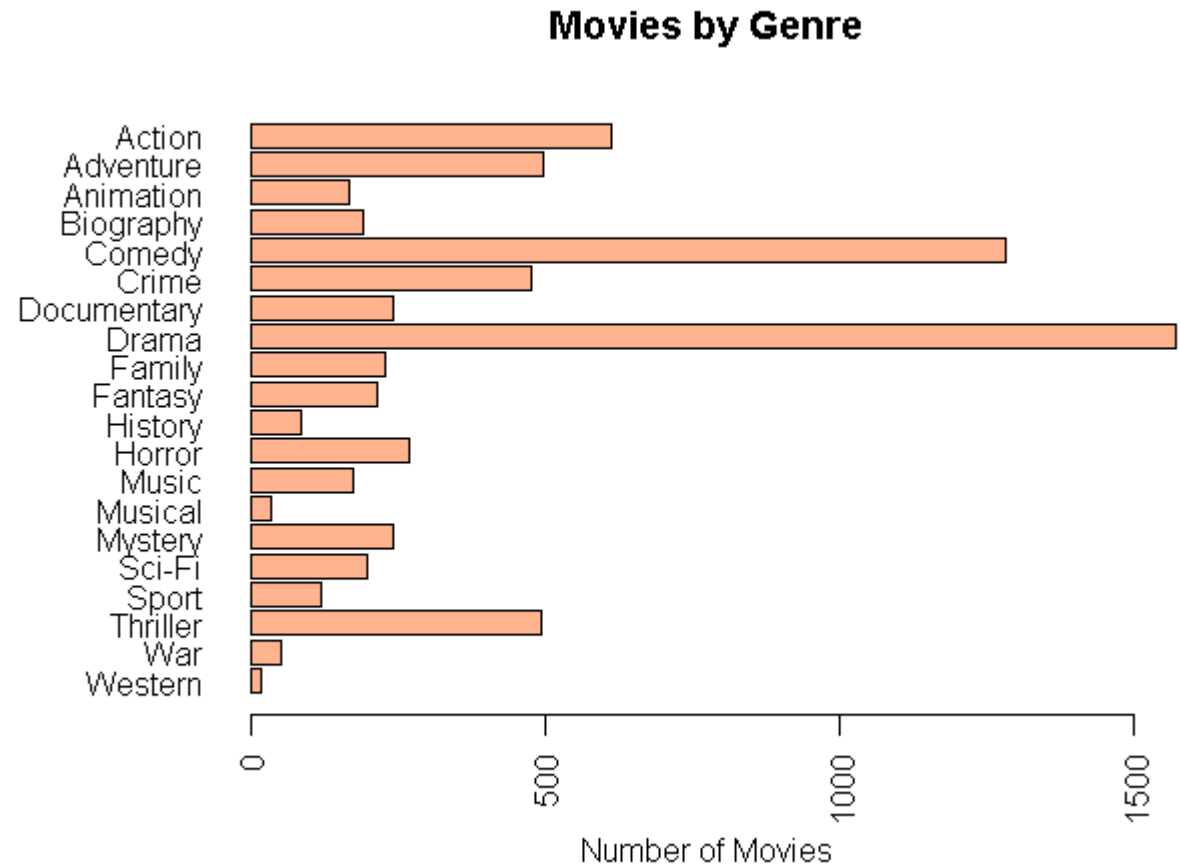


# Horizontal Bar Chart

Many categories

Long names

Easier to read

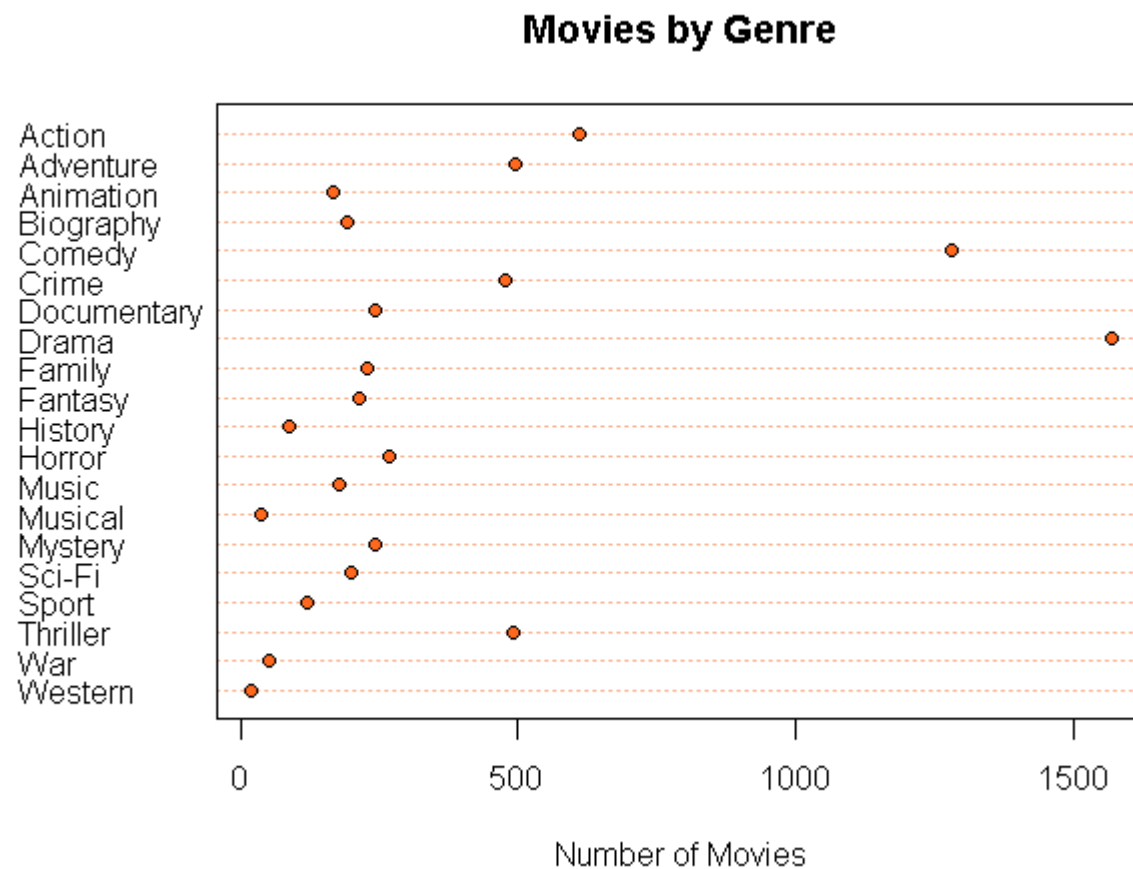




# Cleveland Dot Plot

Data-to-ink ratio

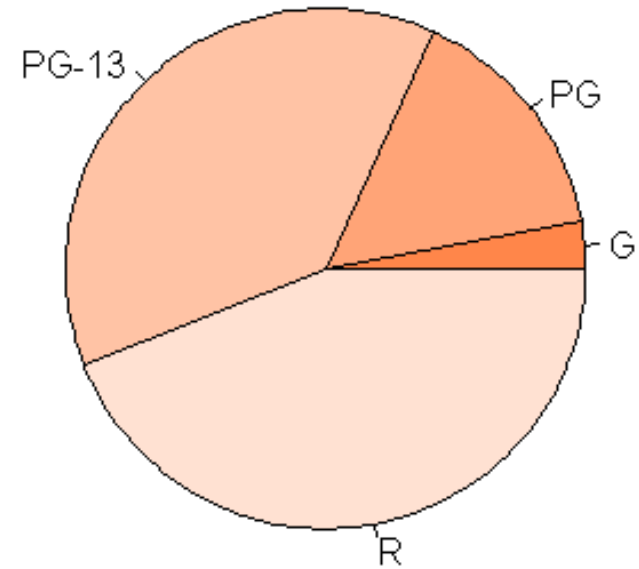
Easier to read



# Pie Chart

Proportion  
Not advised

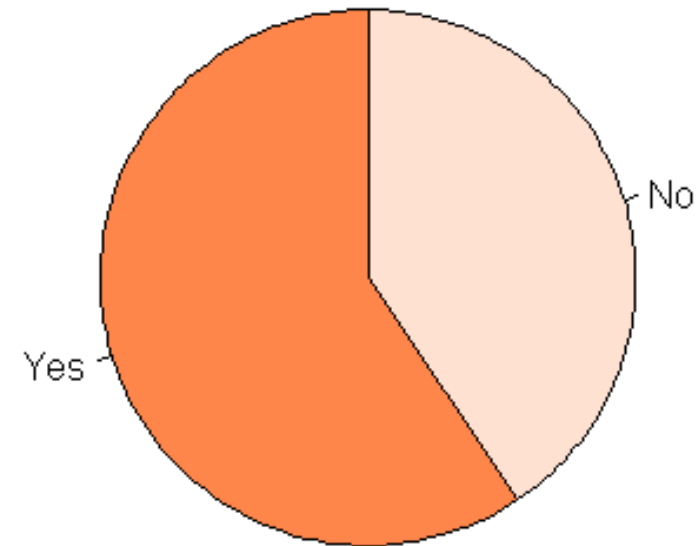
**Movies by Rating Category**



# Pie Chart

Two categories

**Movies by Award Status**







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



# Load the Data



# Create Frequency Bar Chart



# Create Horizontal Bar Chart





# Create a Cleveland Dot Plot



# Create a Pie Chart



# Create a Pie Chart of Awards



# Load the Lattice Package



# Create a Frequency Table



# Create a Frequency Bar Chart



# Create a Horizontal BarChart



# Create a Cleveland Dot Plot





# Create Part of Whole Bar Chart



# Load the ggplot2 Package



# Create a Frequency Bar Chart



# Create a Horizontal Bar Chart



# Cleveland Dot Plot



# Create a Pie Chart



# Create Pie Chart of Awards







# Summary



## Qualitative Univariate Analysis

Demo (Base)

Demo (Lattice)

Demo (ggplot2)