

Data Visualization with R

INTRODUCTION



Matthew Renze

SOFTWARE CONSULTANT

@matthewrenze www.matthewrenze.com











Course Overview



Introduction

Visualizing One Categorical Variable

Visualizing One Numeric Variable

Visualizing Two Categorical Variables

Visualizing Two Numeric Variables

Visualizing Both a Categorical
and a Numeric Variable

Moving Beyond the Basics



Audience

Developers

Data analysts

IT professionals



Prerequisites

**R language basics
(recommended)**

**RStudio basics
(recommended)**



Prerequisites

Exploratory Data Analysis with R

Introduction to R



Matthew Renze

@matthewrenze | www.matthewrenze.com

pluralsight

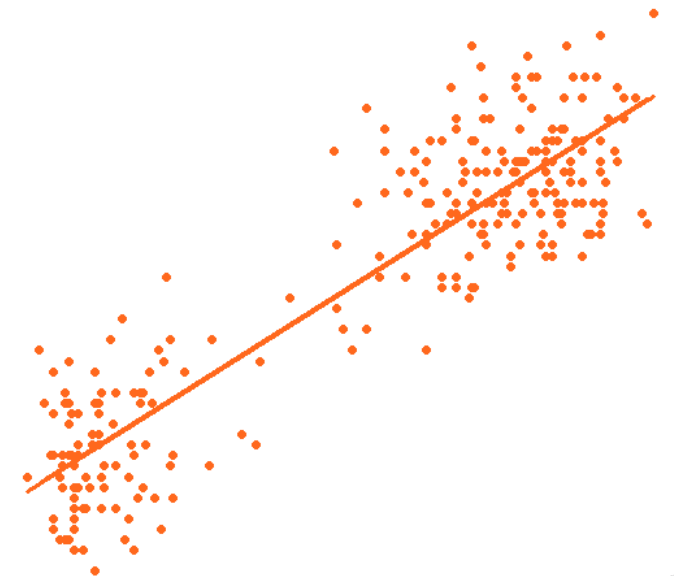
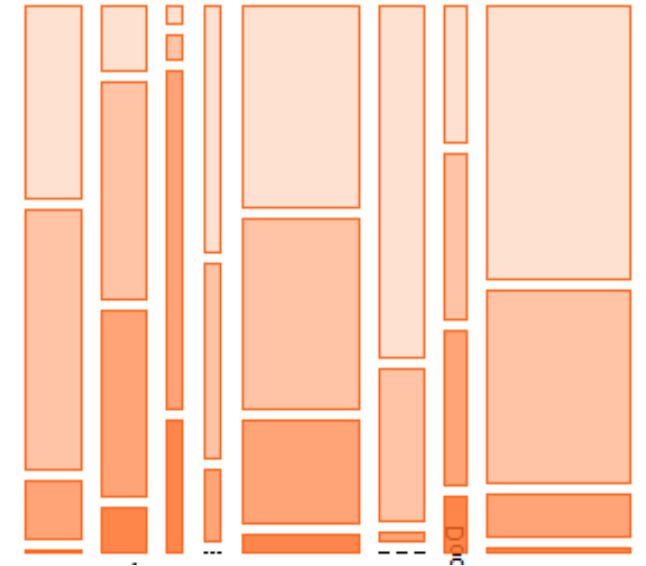
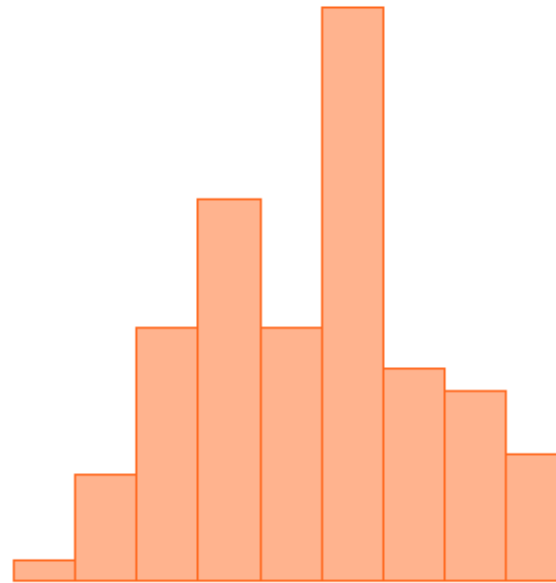
www.pluralsight.com/courses/r-data-analysis



Visual representation

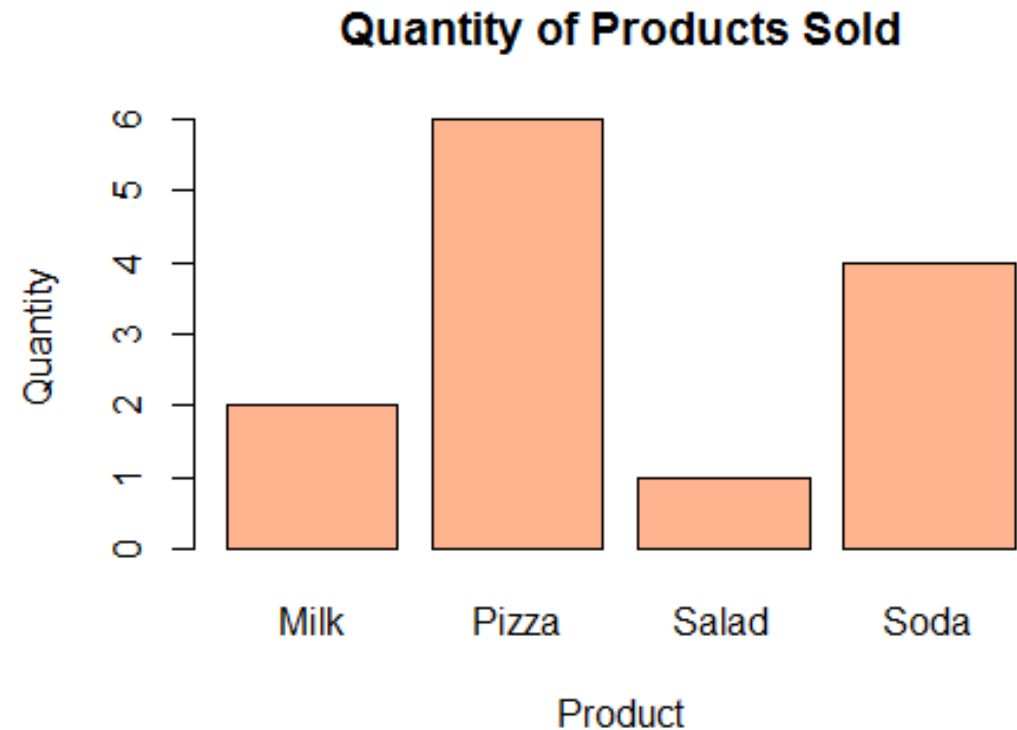
Pattern recognition

Map data to visual

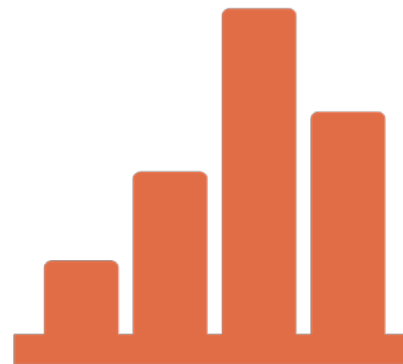
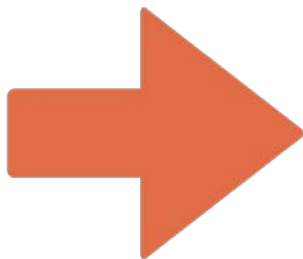


Data Visualization

ID	Date	Customer	Product	Quantity
1	2015-08-27	Jose	Pizza	2
2	2015-08-27	Jose	Soda	2
3	2015-08-27	Jill	Salad	1
4	2015-08-27	Jill	Milk	1
5	2015-08-28	Miko	Pizza	3
6	2015-08-28	Miko	Soda	2
7	2015-08-28	Sam	Pizza	1
8	2015-08-28	Sam	Milk	1



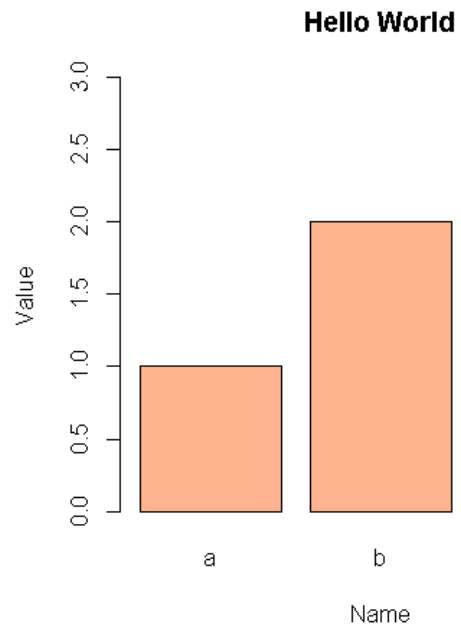
Start with a Question



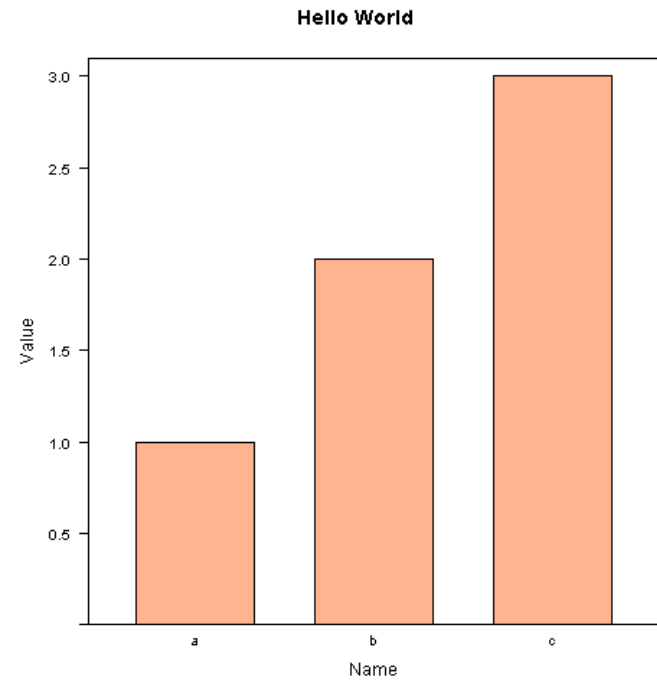
Data Visualization with R



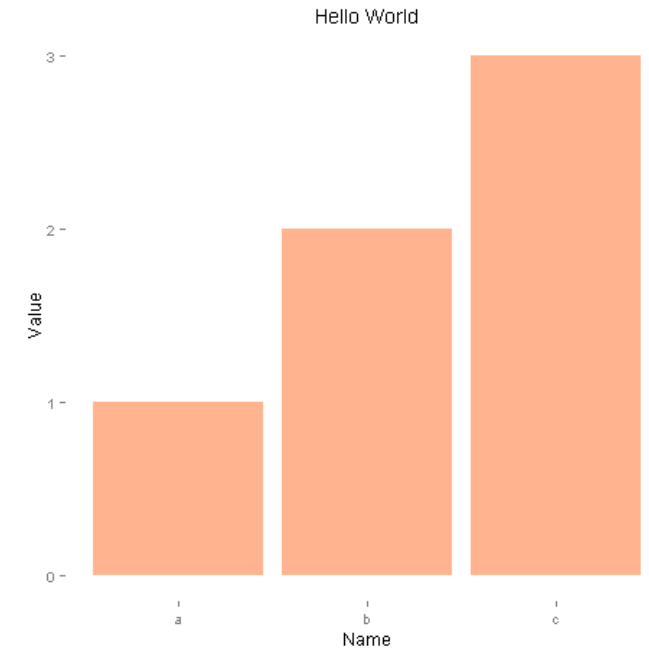
Data Visualization with R



Base



Lattice

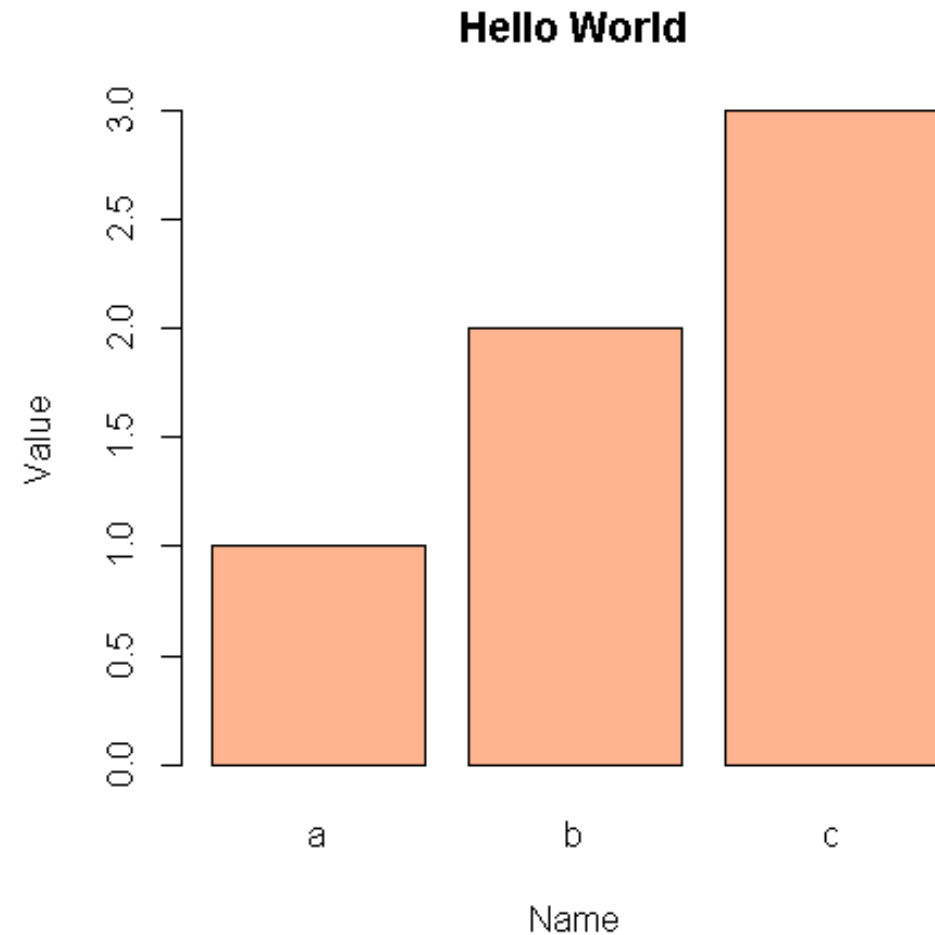


ggplot2



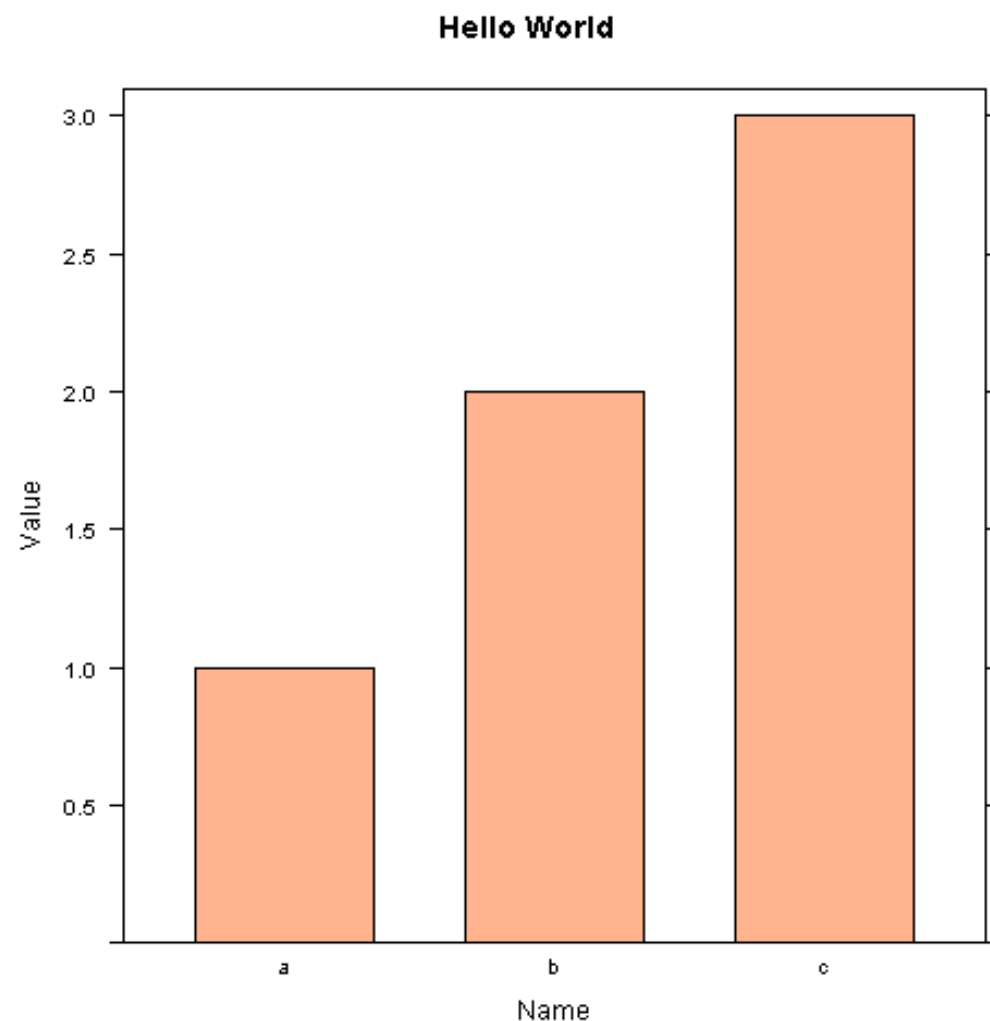
Base Graphics System

```
barplot(  
  names = df$Name,  
  height = df$value,  
  col = alpha("#FF681D", 0.5),  
  main = "Hello world",  
  xlab = "Name",  
  ylab = "value")
```



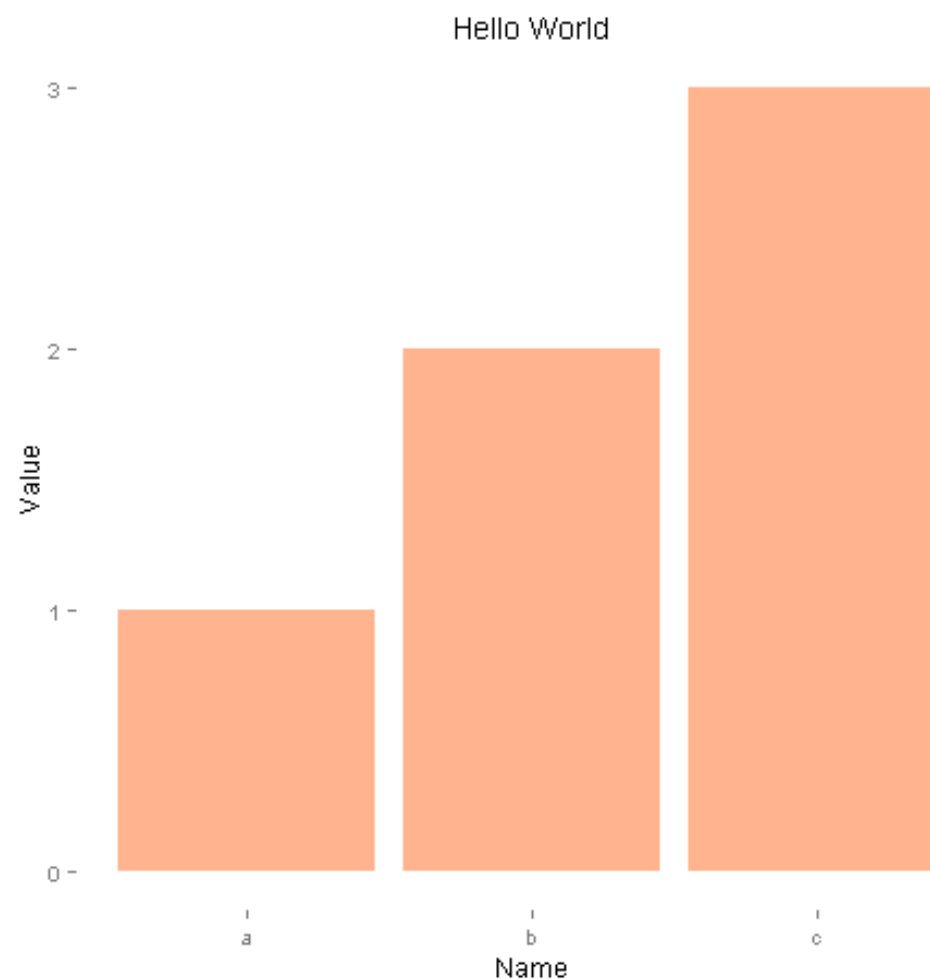
Lattice

```
barchart(  
  x = Value ~ Name,  
  data = df,  
  ylim = c(0, 3.1),  
  col = alpha("#FF681D", 0.5),  
  main = "Hello world",  
  xlab = "Name",  
  ylab = "value")
```

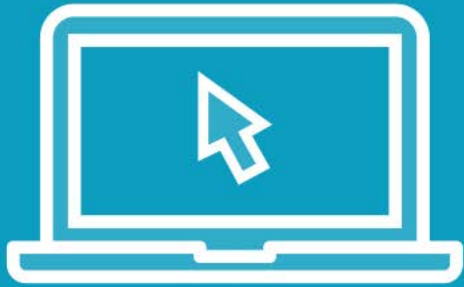


ggplot2

```
ggplot(  
  data = df,  
  aes(x = Name, y = value)) +  
  geom_bar(  
    stat = "identity",  
    fill = alpha("#FF681D", 0.5)) +  
  ggtitle("Hello world") +  
  xlab("Name") +  
  ylab("value")
```



Demo



Base

Lattice

ggplot2



Intro



Create a Data Frame



Display the Data Frame



Plot with Defaults



Plot with Ordered Parameters



Plot with Named Parameters



Create a Bar Chart



View Help Files



Install Lattice



Load Lattice



Create a Data Frame



Plot with Defaults



Plot with Parameters



Create a Bar Chart



View Help Files



Install ggplot2



Load ggplot2



Create a data frame



Plot with Defaults



Plot with Layers



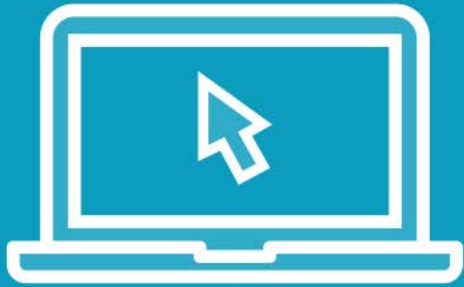
Create a Bar Chart



View Help Files



Demo



Base

Lattice

ggplot2



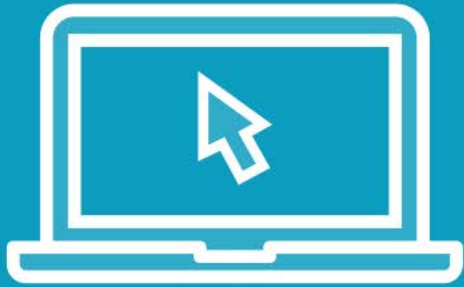
Demo



Pros and cons



Demo

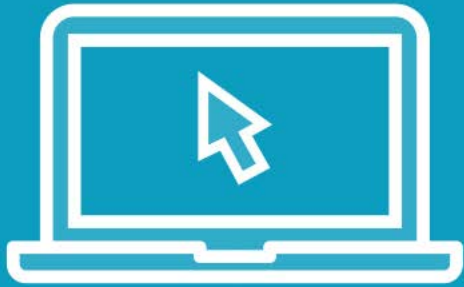


Pros and cons

Repetition



Demo



Pros and cons

Repetition

Perspective



Statistical Terms

Observations

Variables

Qualitative variable

Quantitative variable

ID	Date	Customer	Product	Quantity
1	2015-08-27	Jose	Pizza	2
2	2015-08-27	Jose	Soda	2
3	2015-08-27	Jill	Salad	1
4	2015-08-27	Jill	Milk	1
5	2015-08-28	Miko	Pizza	3
6	2015-08-28	Miko	Soda	2
7	2015-08-28	Sam	Pizza	1
8	2015-08-28	Sam	Milk	1



Types of Data Analysis

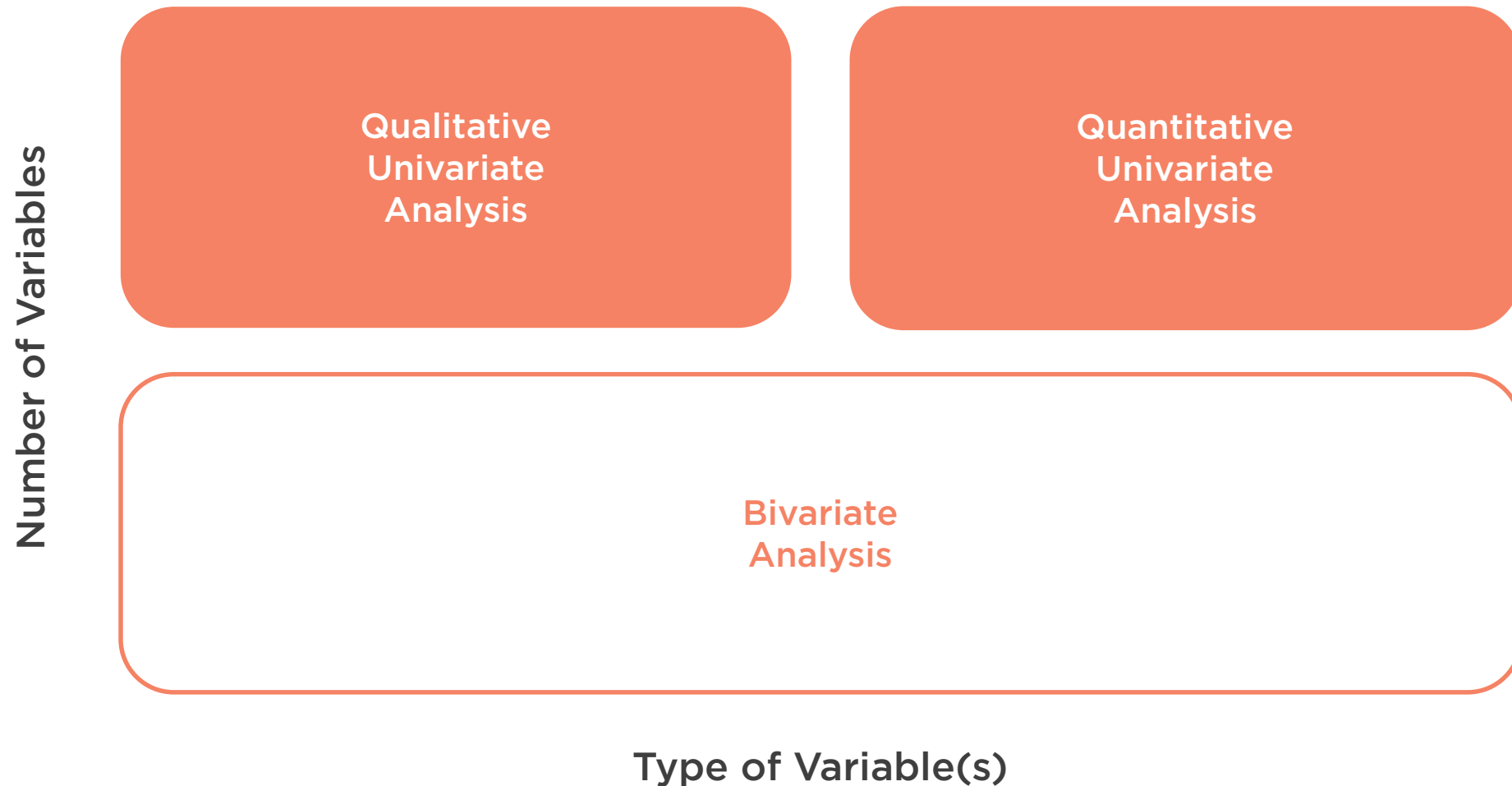
Number of Variables

Univariate
Analysis

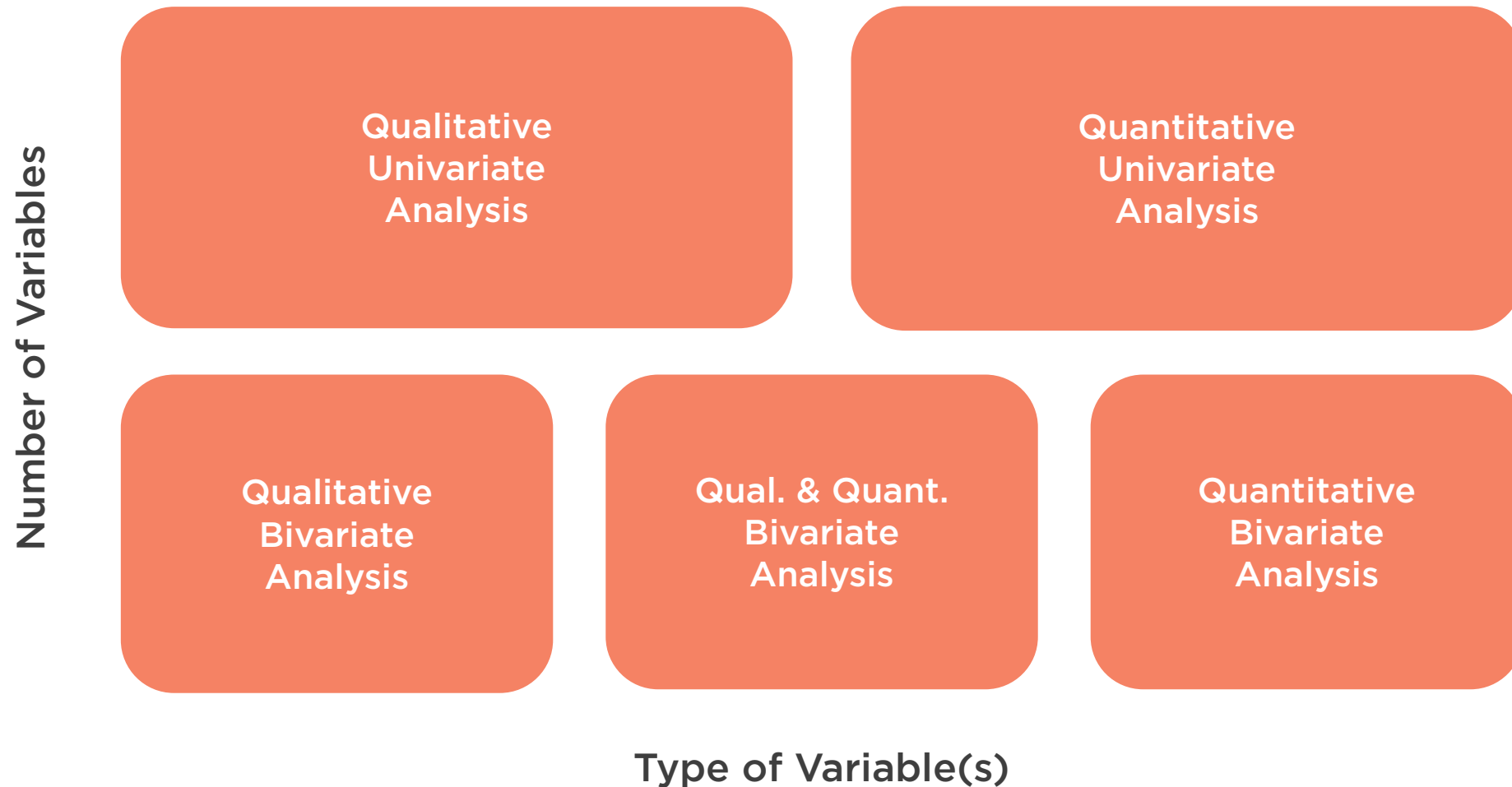
Type of Variable(s)



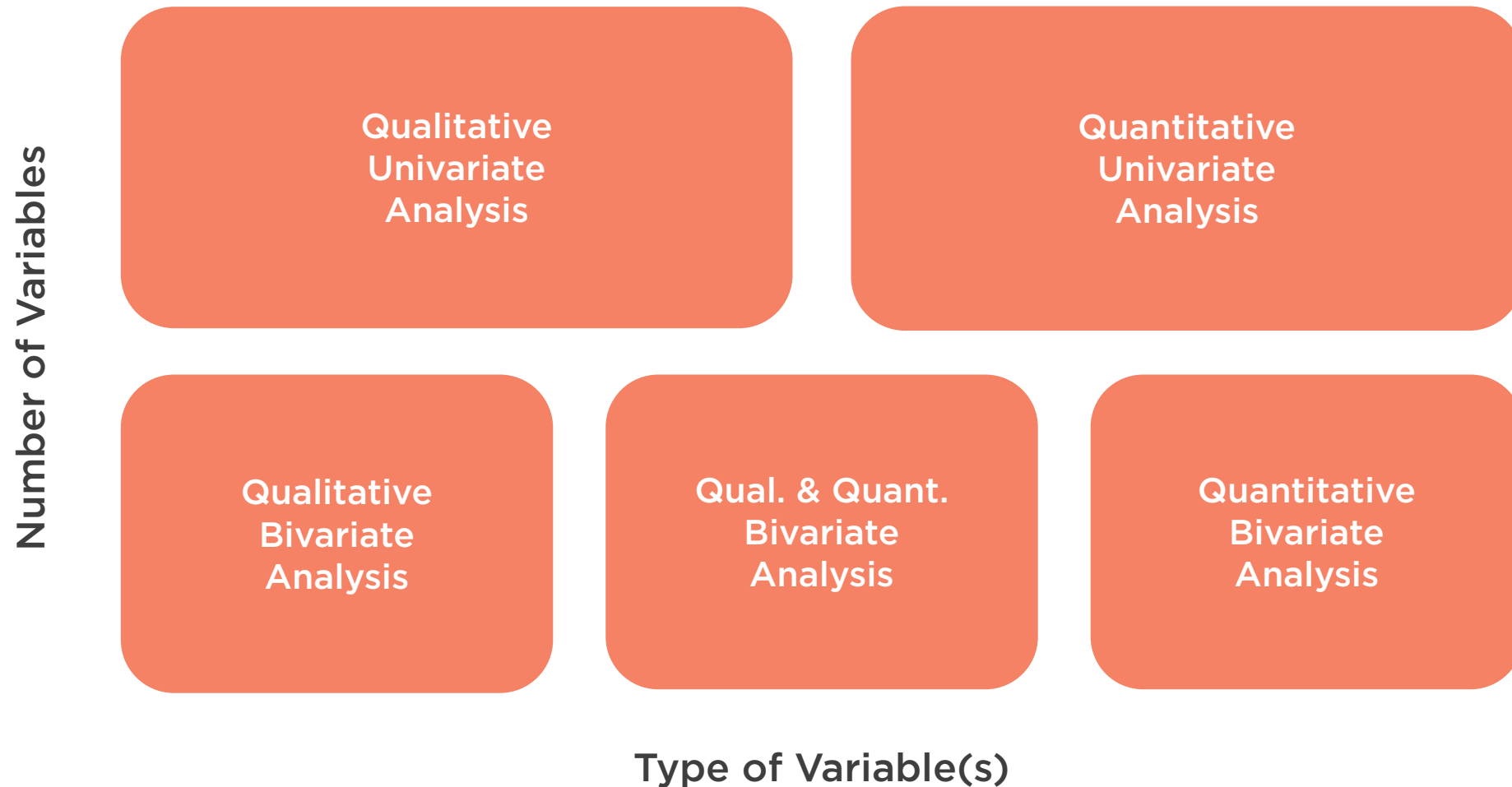
Types of Data Analysis



Types of Data Analysis

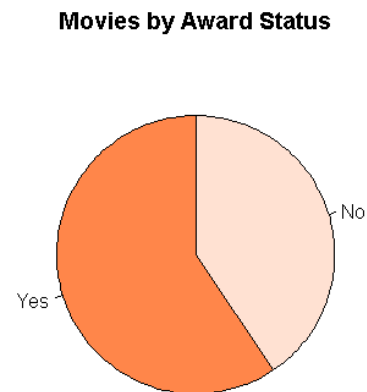
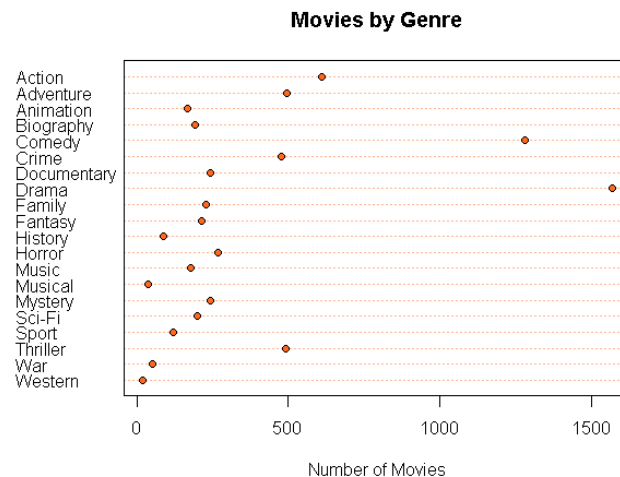
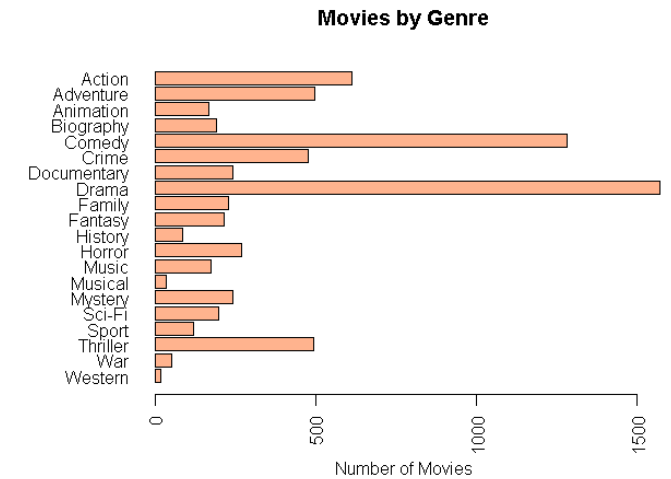
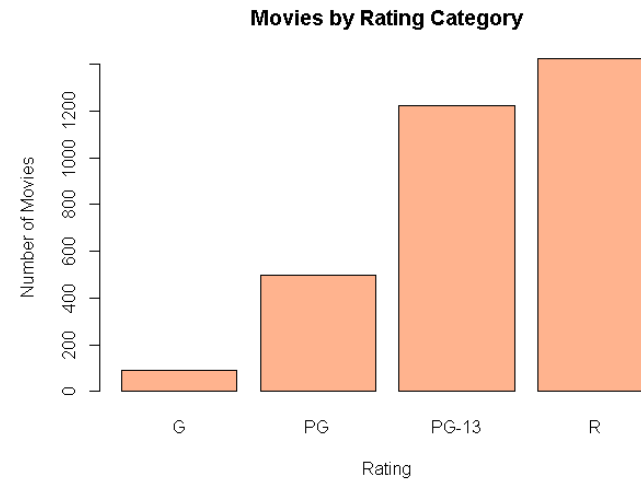


Types of Data Analysis



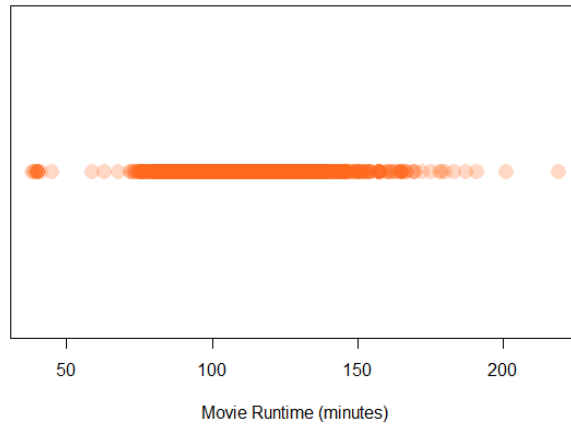
Visualizing One Categorical Variable

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

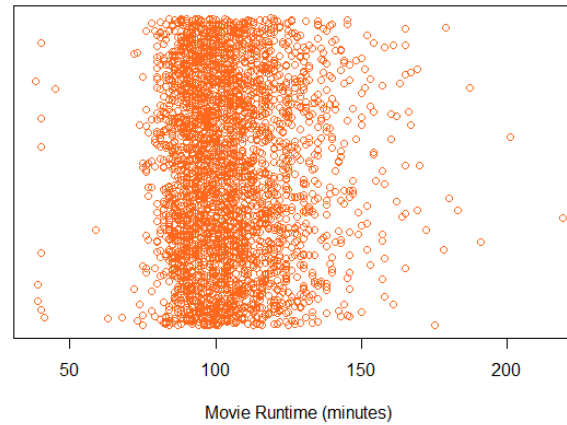


Visualizing One Numeric Variable

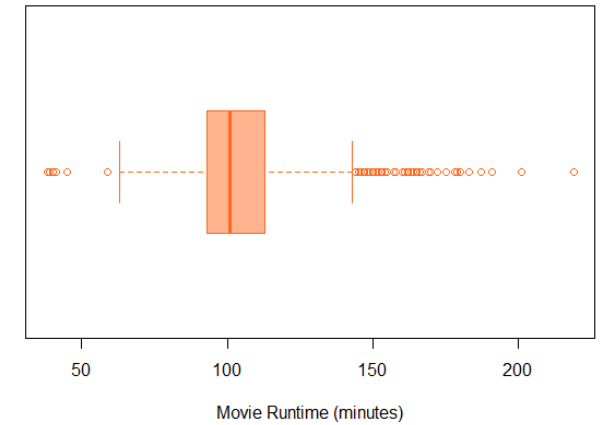
Distribution of Movie Runtimes



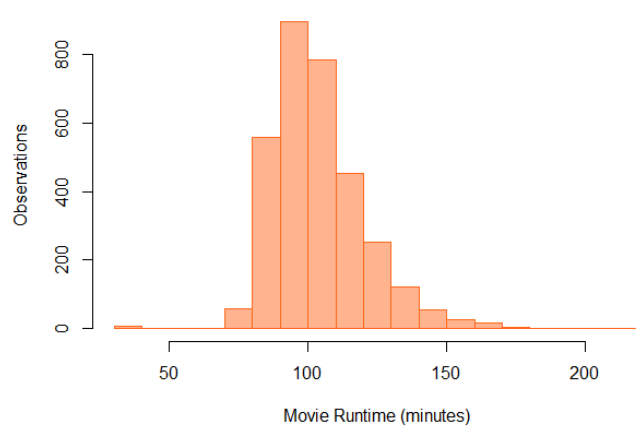
Distribution of Movie Runtimes



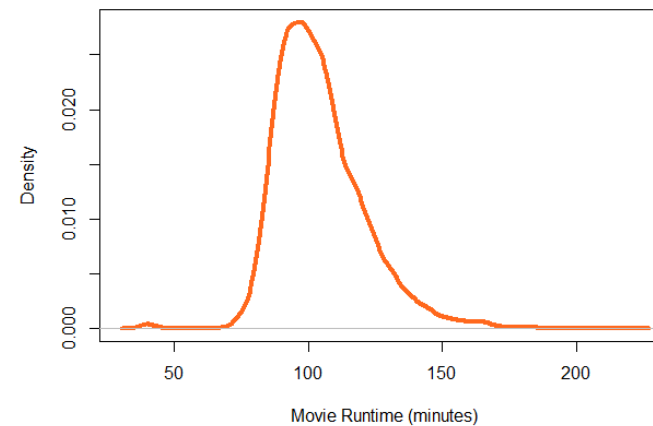
Distribution of Movie Runtimes



Distribution of Movie Runtimes

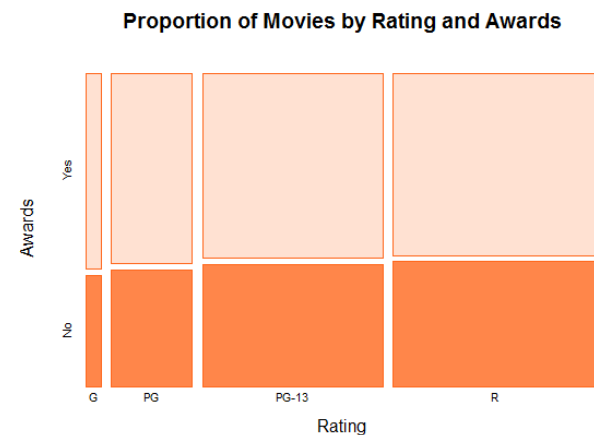
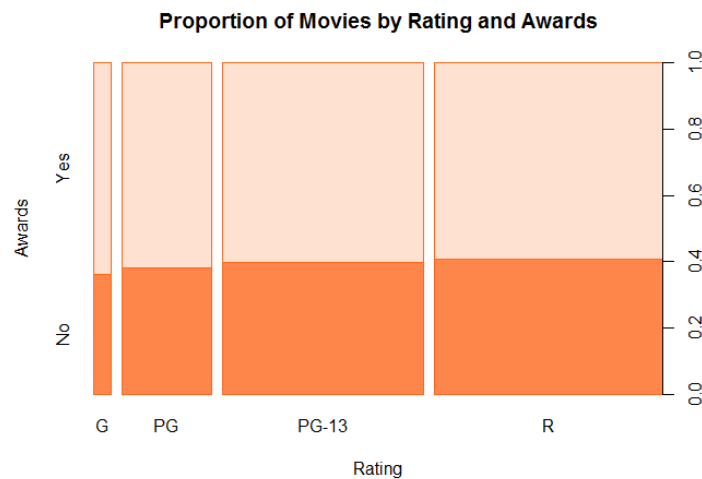
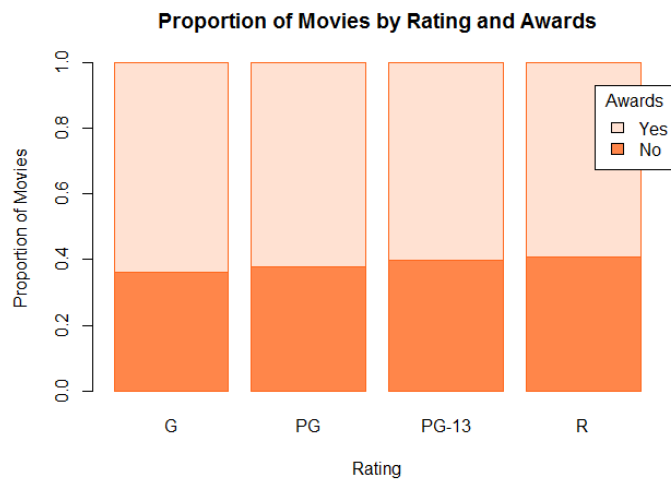
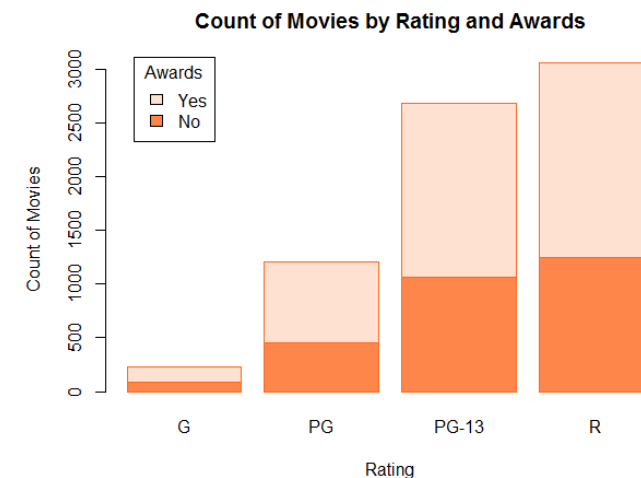
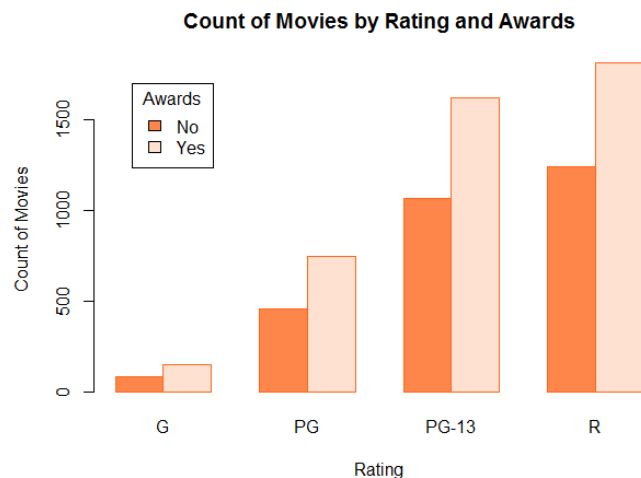


Distribution of Movie Runtimes

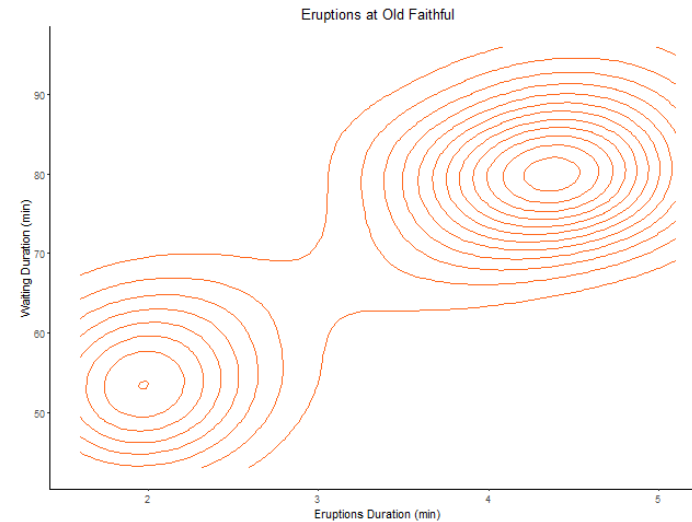
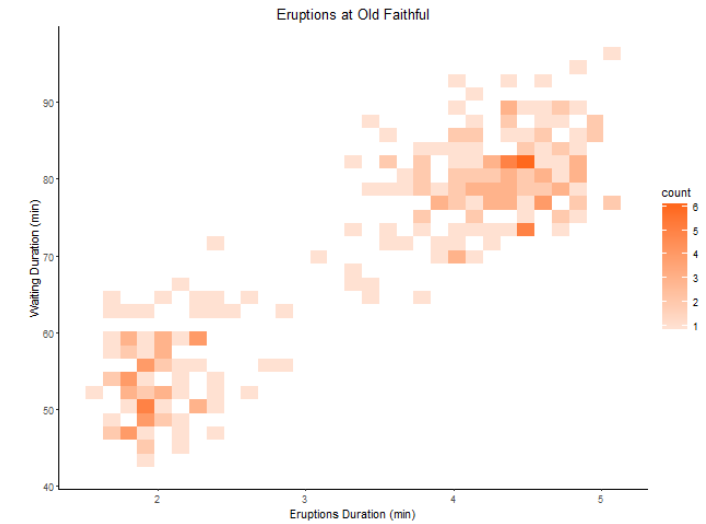
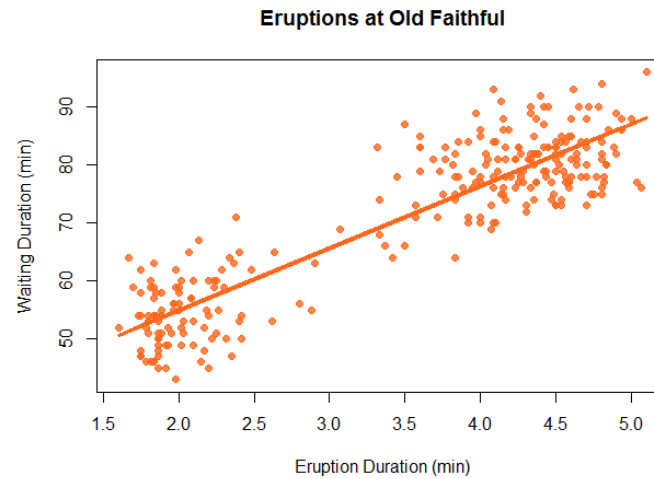
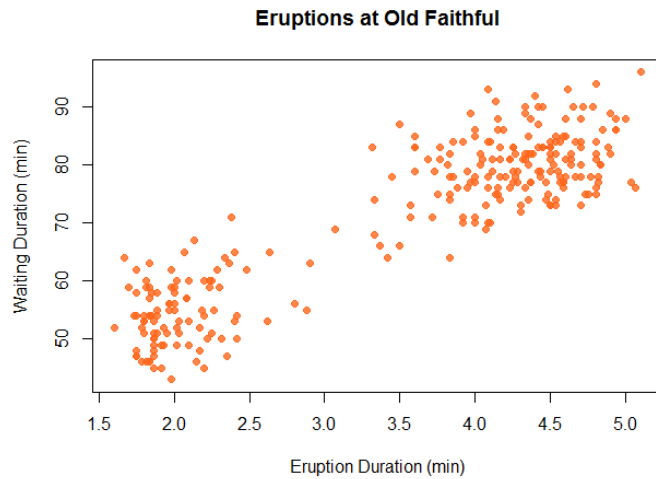


Visualizing Two Categorical Variables

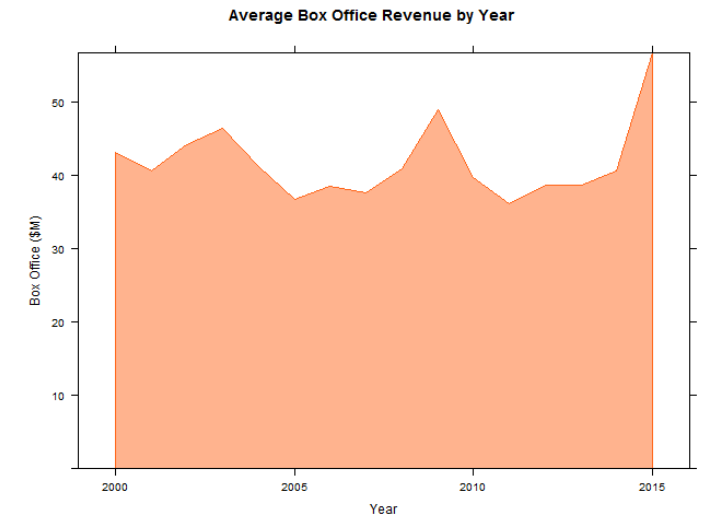
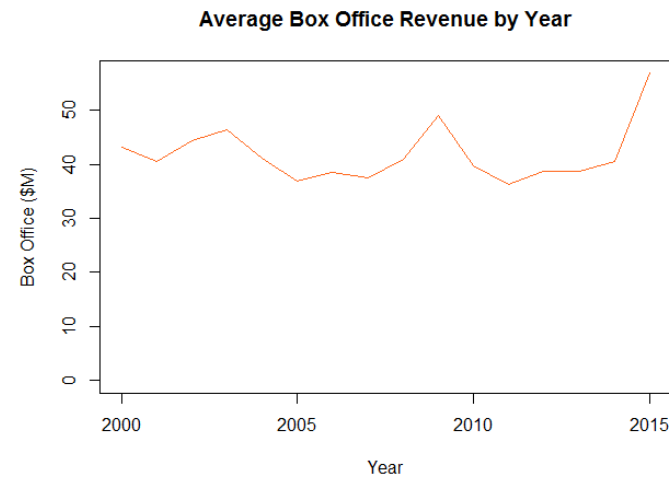
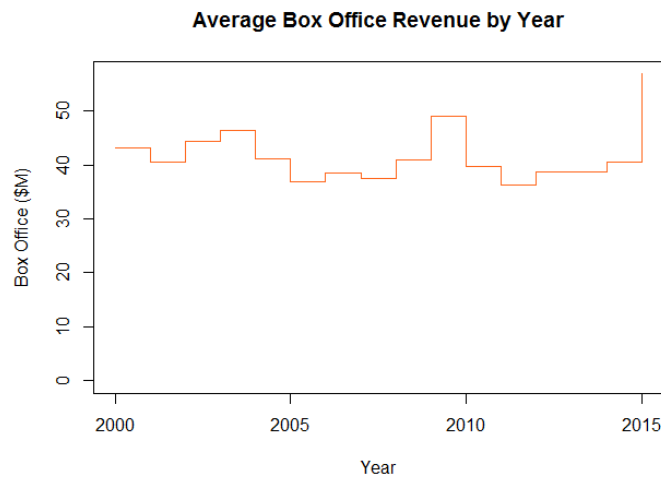
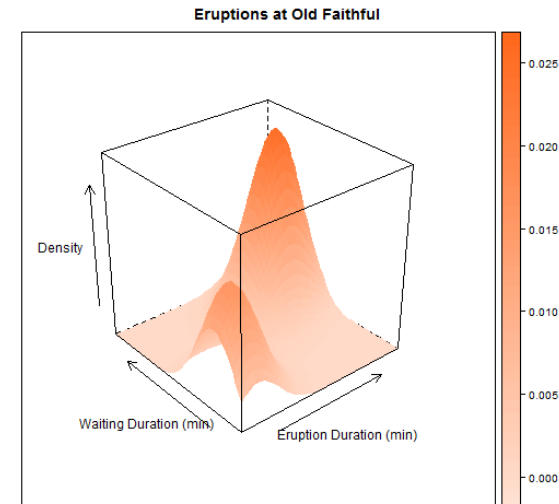
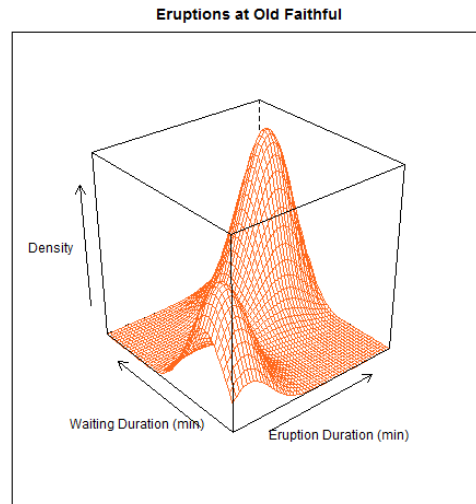
Movies by Genre and Rating					
Genre	G	PG	PG-13	R	Total
Action	2	70	311	229	612
Adventure	44	179	209	64	496
Animation	43	111	8	6	168
Comedy	45	258	472	506	1218
Drama	12	136	586	836	1570
Family	38	181	10	1	230
...
Total	230	1207	2686	3058	7181



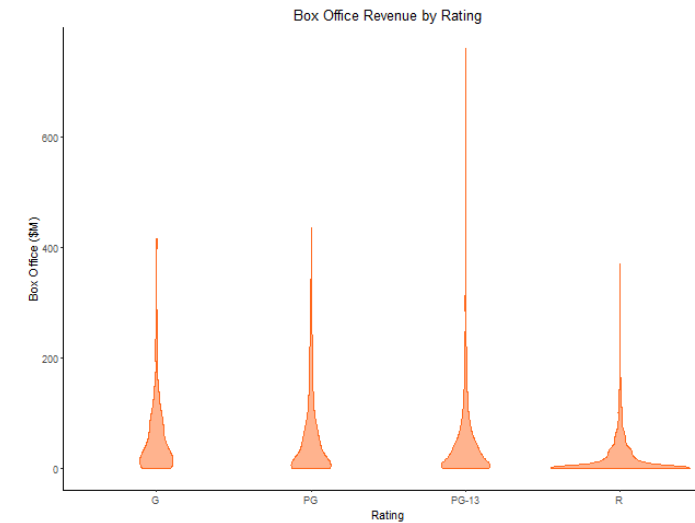
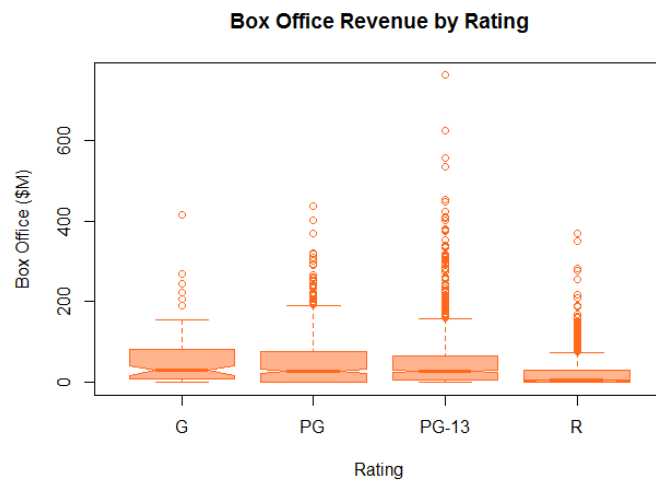
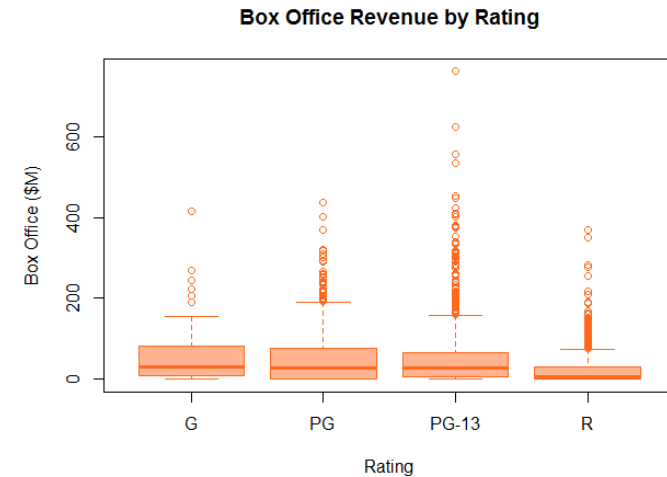
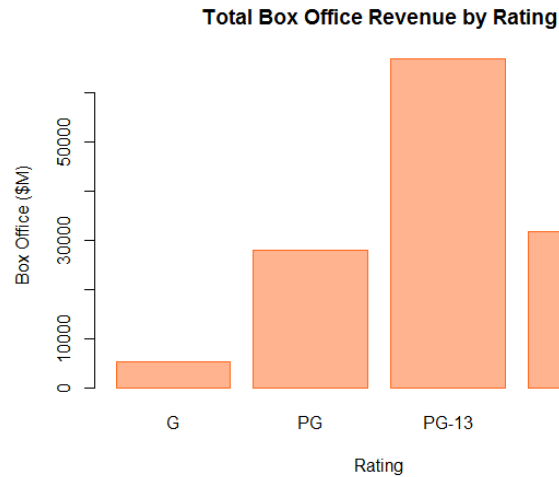
Visualizing Two Numeric Variables



Visualizing Two Numeric Variables



Visualizing Both a Categorical and a Numeric Variable



Summary



Course Introduction

Data Visualization

Demos

Types of Visualizations

