

GÖRAN KIRCHNER

NOTES ON R

1

Packages

2

Visualization

2.1 Data

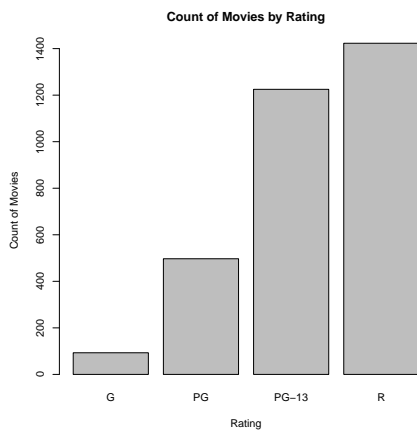
```
movies <- read.csv("data/movies.csv")
head(movies)
```

Title	Year	Rating	Runtime	Critic.Score	Box.Office	Awards	International
The Whole Nine Yards	2000	R	98	45	57.3	FALSE	FALSE
Cirque du Soleil: Journey of Man	2000	G	39	45	13.4	TRUE	FALSE
Gladiator	2000	R	155	76	187.3	TRUE	TRUE
Dinosaur	2000	PG	82	65	135.6	TRUE	FALSE
Big Momma's House	2000	PG-13	99	30	0.5	TRUE	TRUE
Gone in Sixty Seconds	2000	PG-13	118	24	101	TRUE	FALSE

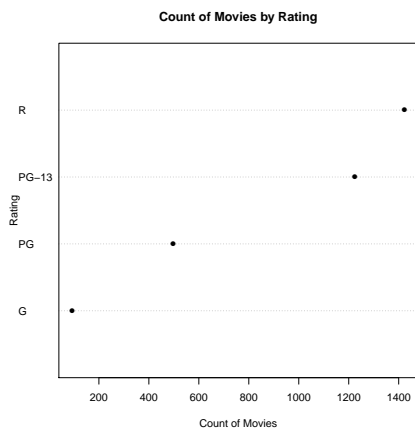
2.2 One Categorical Variable

2.2.1 base

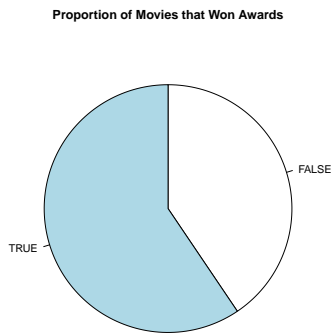
```
movies <- read.csv("data/movies.csv")
plot(
  x = movies$Rating,
  main = "Count of Movies by Rating",
  xlab = "Rating",
  ylab = "Count of Movies")
```



```
movies <- read.csv("data/movies.csv")
dotchart(
  x = table(movies$Rating),
  pch = 16,
  main = "Count of Movies by Rating",
  xlab = "Count of Movies",
  ylab = "Rating")
```



```
movies <- read.csv("data/movies.csv")
pie(
  x = table(movies$Awards),
  clockwise = TRUE,
  main = "Proportion of Movies that Won Awards")
```

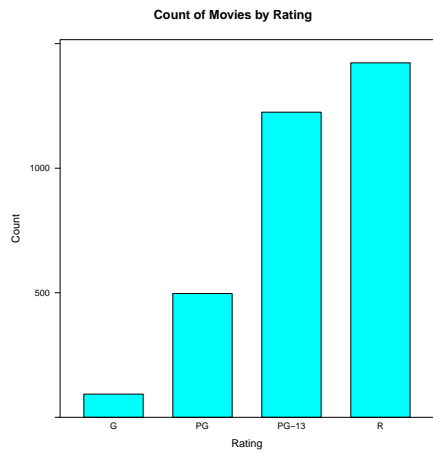


2.2.2 *lattice*

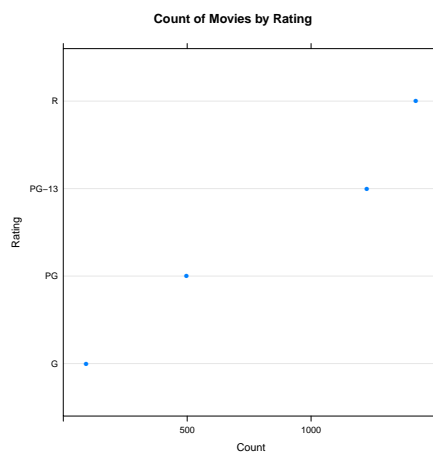
```
library(lattice)
# Create frequency table of ratings
movies <- read.csv("data/movies.csv")
table <- table(movies$Rating)
ratings <- as.data.frame(table)
names(ratings)[1] <- "Rating"
names(ratings)[2] <- "Count"
print(ratings)
```

Rating	Count
G	93
PG	497
PG-13	1225
R	1423

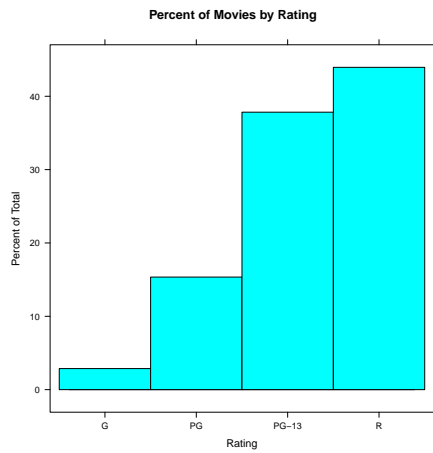
```
library(lattice)
# Create frequency table of ratings
movies <- read.csv("data/movies.csv")
table <- table(movies$Rating)
ratings <- as.data.frame(table)
names(ratings)[1] <- "Rating"
names(ratings)[2] <- "Count"
barchart(
  x = Count ~ Rating,
  data = ratings,
  main = "Count of Movies by Rating",
  xlab = "Rating")
```



```
library(lattice)
# Create frequency table of ratings
movies <- read.csv("data/movies.csv")
table <- table(movies$Rating)
ratings <- as.data.frame(table)
names(ratings)[1] <- "Rating"
names(ratings)[2] <- "Count"
dotplot(
  x = Rating ~ Count,
  data = ratings,
  main = "Count of Movies by Rating",
  ylab = "Rating")
```

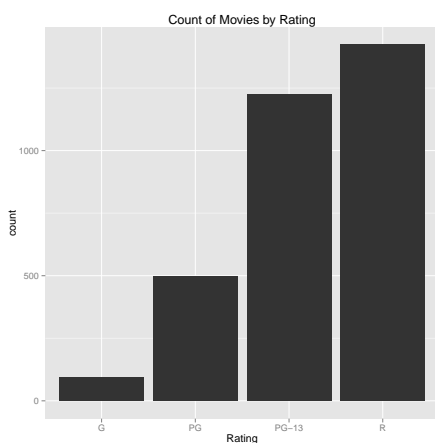


```
library(lattice)
# Create frequency table of ratings
movies <- read.csv("data/movies.csv")
table <- table(movies$Rating)
ratings <- as.data.frame(table)
names(ratings)[1] <- "Rating"
names(ratings)[2] <- "Count"
histogram(
  x = ~Rating,
  data = movies,
  main = "Percent of Movies by Rating")
```

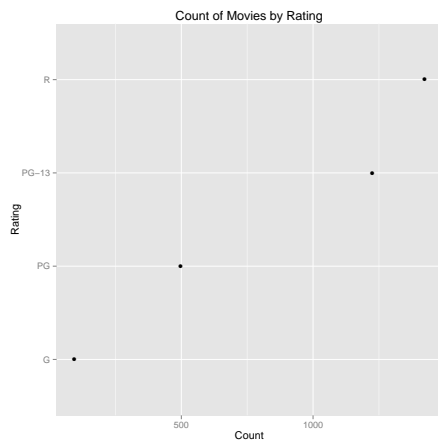



2.2.3 *ggplot2*

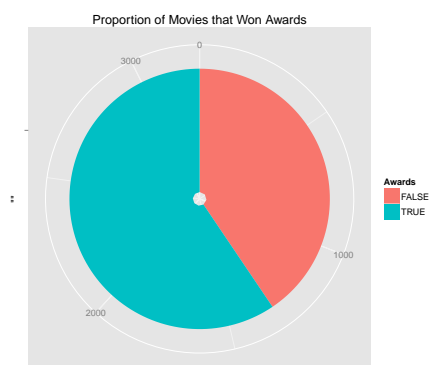
```
library(ggplot2)
movies <- read.csv("data/movies.csv")
ggplot(
  data = movies,
  aes(x = Rating)) +
  geom_bar() +
  ggtitle("Count of Movies by Rating")
```



```
library(ggplot2)
library(lattice)
# Create frequency table of ratings
movies <- read.csv("data/movies.csv")
table <- table(movies$Rating)
ratings <- as.data.frame(table)
names(ratings)[1] <- "Rating"
names(ratings)[2] <- "Count"
ggplot(
  data = ratings,
  aes(x = Rating, y = Count)) +
  geom_point() +
  coord_flip() +
  ggtitle("Count of Movies by Rating")
```



```
library(ggplot2)
movies <- read.csv("data/movies.csv")
ggplot(
  data = movies,
  aes(x = "", fill = Awards)) +
  geom_bar() +
  coord_polar(theta = "y") +
  ggtitle("Proportion of Movies that Won Awards") +
  ylab("")
```



2.3 *One Numeric Variable*

2.4 *Two Categorical Variables*

2.5 *Two Numeric Variables*

2.6 *Both a Categorical and a Numeric Variable*

2.7 *Moving Beyond*