

# DSC 520 Week 2 Exercise 2.1

Dong Woon Kim

2025-12-13

## 1. Bar graph showing product and the sum of sales for each.

```
sales <- read.csv(file = 'sales.csv')

View(sales)

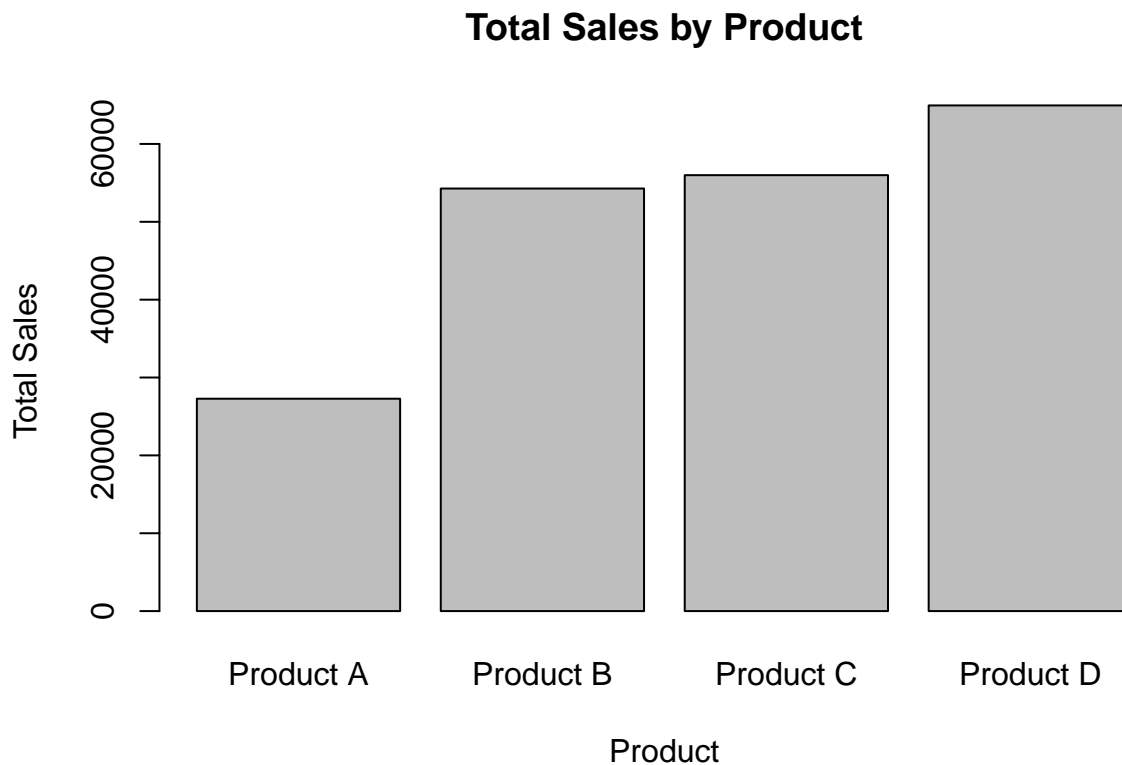
# Bar graph showing product and the sum of sales for each.

dfA <- filter(.data = sales, sales$Product == 'Product_A')
dfB <- filter(.data = sales, sales$Product == 'Product_B')
dfC <- filter(.data = sales, sales$Product == 'Product_C')
dfD <- filter(.data = sales, sales$Product == 'Product_D')

A <- sum(dfA$TotalSales)
B <- sum(dfB$TotalSales)
C <- sum(dfC$TotalSales)
D <- sum(dfD$TotalSales)

sales_by_product <- c(A, B, C, D)

barplot(
  sales_by_product,
  main = "Total Sales by Product",
  xlab = 'Product',
  ylab = 'Total Sales',
  names.arg = c('Product A', 'Product B', 'Product C', 'Product D')
)
```



## 2. Pie chart showing sum of sales by region.

```
help('pie')
```

```
## starting httpd help server ... done
```

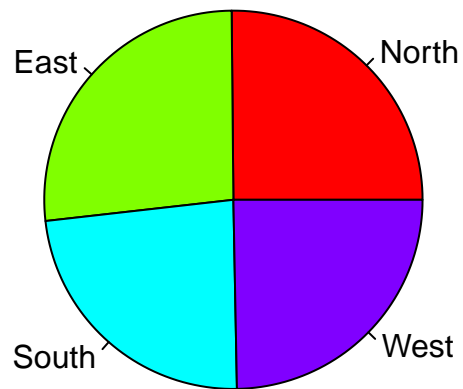
```
dfN <- filter(.data = sales, sales$Region == 'North')
dfE <- filter(.data = sales, sales$Region == 'East')
dfS <- filter(.data = sales, sales$Region == 'South')
dfW <- filter(.data = sales, sales$Region == 'West')
```

```
N <- sum(dfN$TotalSales)
E <- sum(dfE$TotalSales)
S <- sum(dfS$TotalSales)
W <- sum(dfW$TotalSales)
```

```
sales_by_region <- c(N, E, S, W)
names(sales_by_region) = c('North', 'East', 'South', 'West')
```

```
pie(
  x = sales_by_region,
  labels = names(sales_by_region),
  col = rainbow(4),
  main = "Total Sales by Region"
)
```

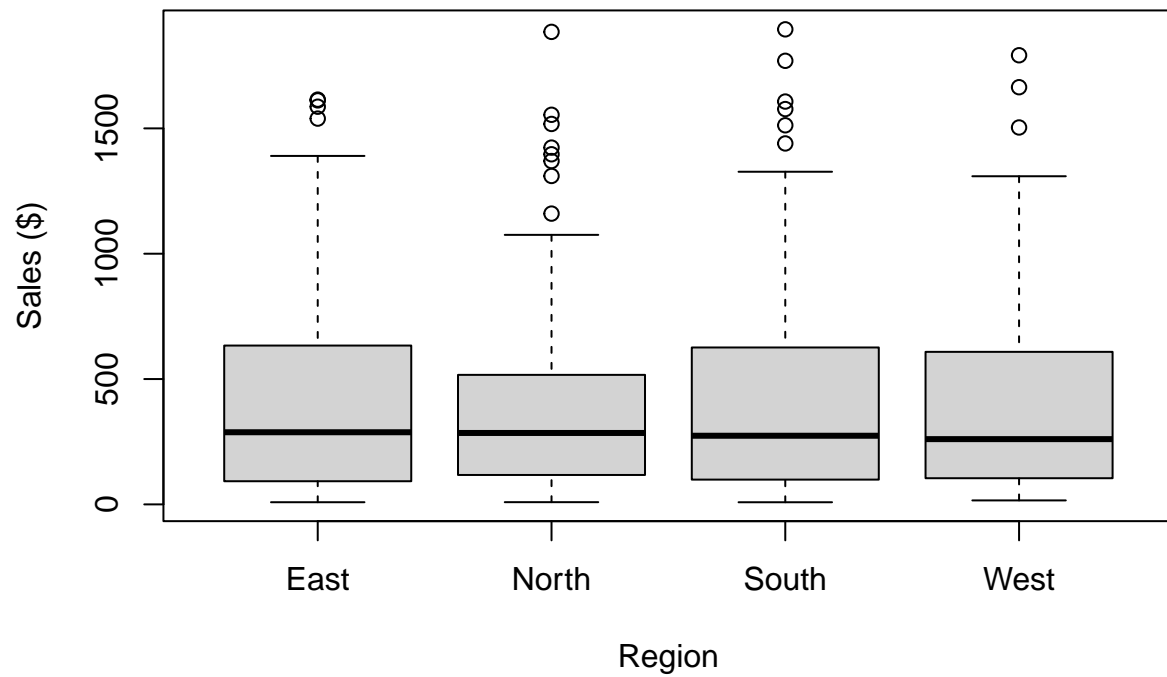
## Total Sales by Region



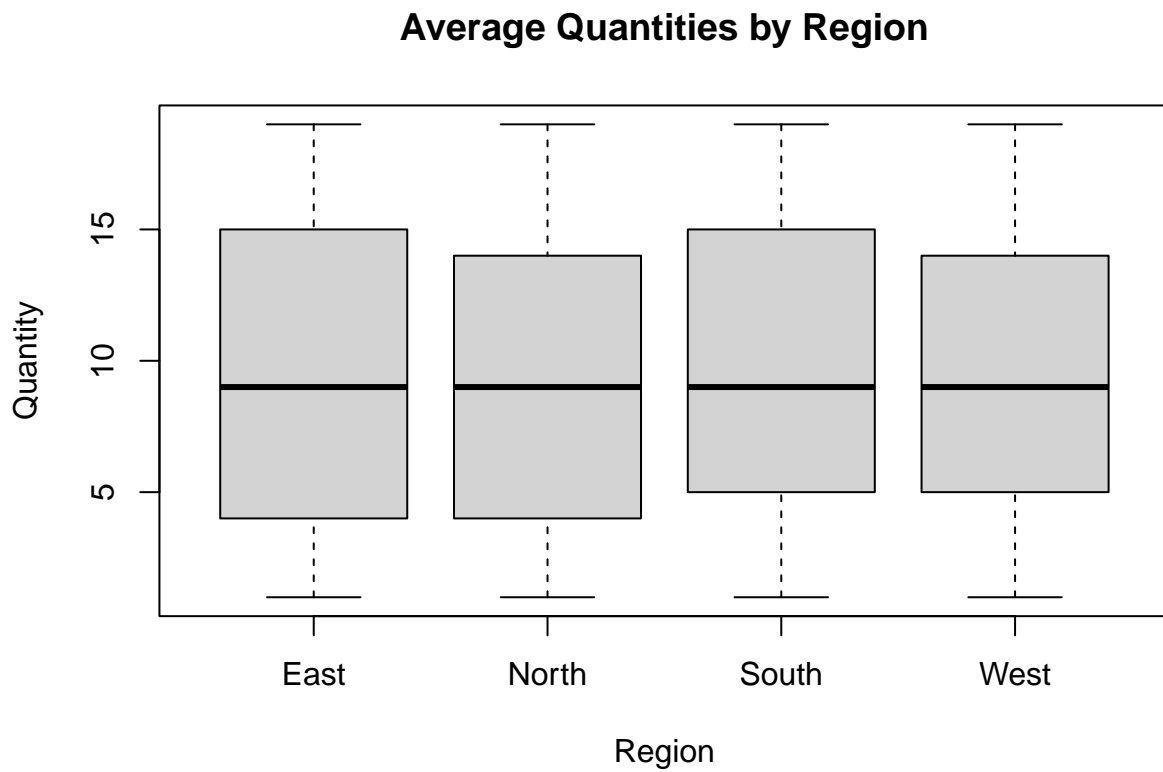
3. Box plots for average sales by region and average quantities by region.

```
boxplot(  
  formula = TotalSales ~ Region,  
  data = sales,  
  main = "Average Sales by Region",  
  ylab = "Sales ($)"  
)
```

## Average Sales by Region



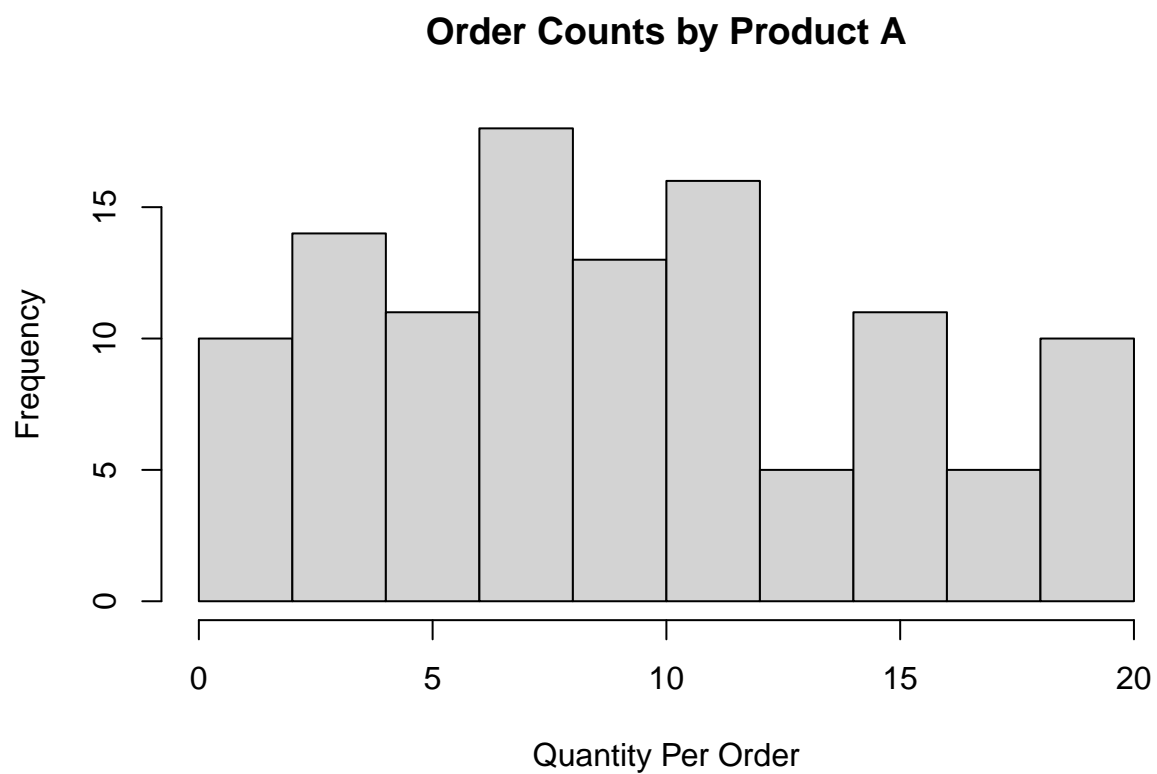
```
boxplot(  
  formula = Quantity ~ Region,  
  data = sales,  
  main = "Average Quantities by Region",  
  ylab = "Quantity"  
)
```



#### 4. Histogram of the count of orders by product.

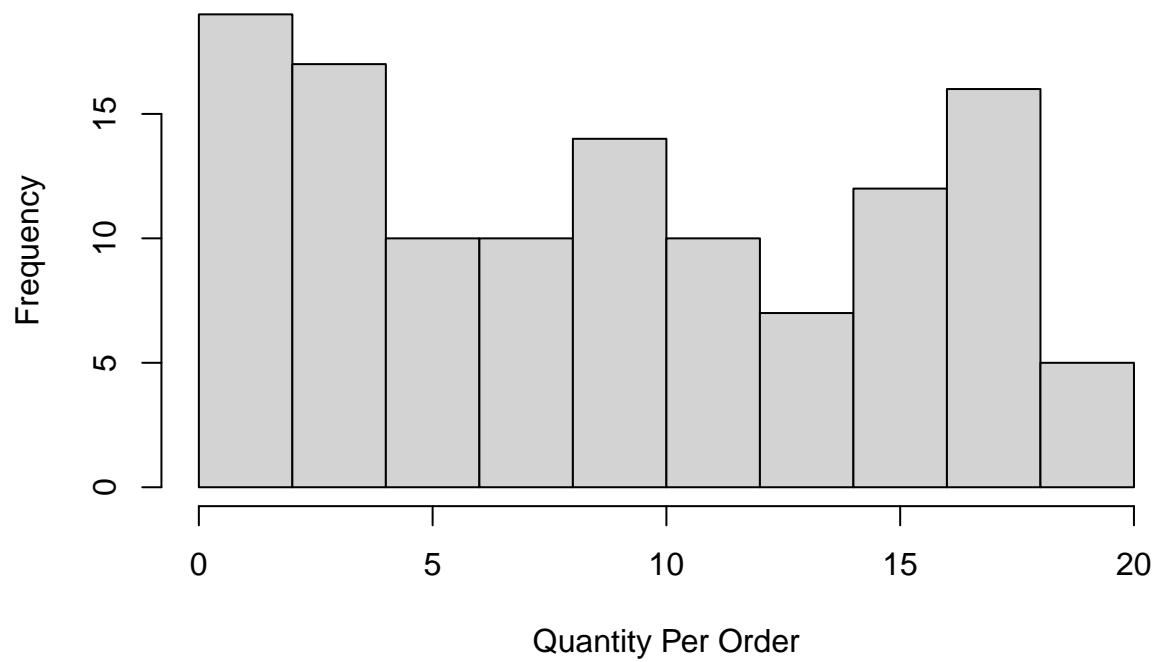
```
help(hist)

hist(
  x = dfA$Quantity,
  main = 'Order Counts by Product A',
  breaks = 'Sturges',
  xlab = 'Quantity Per Order'
)
```



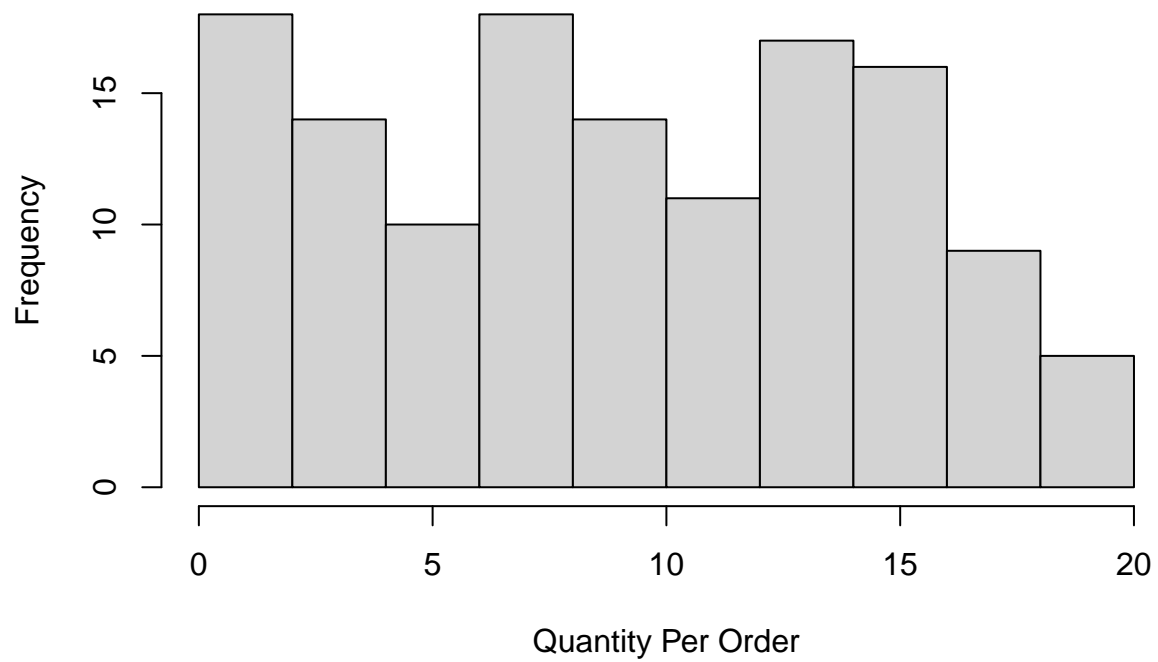
```
hist(  
  x = dfB$Quantity,  
  main = 'Order Counts by Product B',  
  breaks = 'Sturges',  
  xlab = 'Quantity Per Order'  
)
```

## Order Counts by Product B

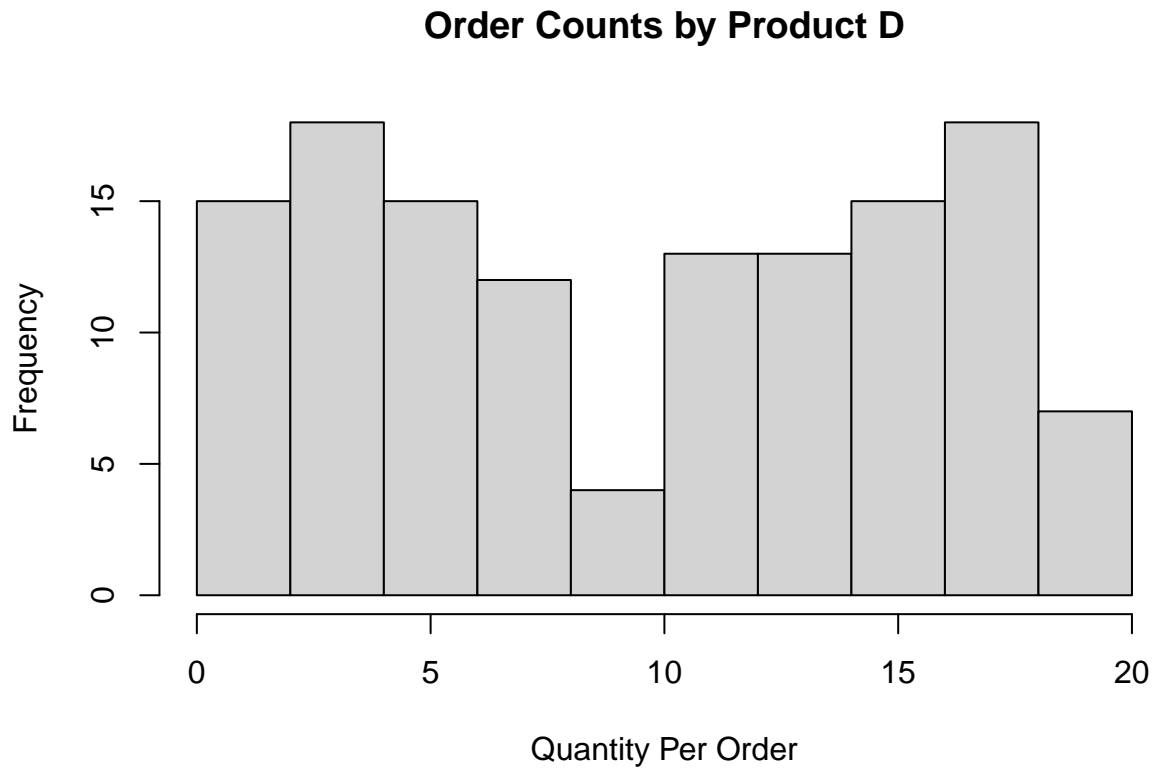


```
hist(  
  x = dfC$Quantity,  
  main = 'Order Counts by Product C',  
  breaks = 'Sturges',  
  xlab = 'Quantity Per Order'  
)
```

### Order Counts by Product C



```
hist(  
  x = dfD$Quantity,  
  main = 'Order Counts by Product D',  
  breaks = 'Sturges',  
  xlab = 'Quantity Per Order'  
)
```



## 5. A radial plot showing total sales (for all products) by month.

```
install.packages('stringr')
```

```
## Warning: package 'stringr' is in use and will not be installed
```

```
library(stringr)
```

```
toString(sales$OrderDate)
```

```
## [1] "1/12/2024, 8/27/2025, 5/8/2024, 4/6/2024, 9/7/2024, 10/23/2023, 8/6/2025, 4/29/2023, 4/25/2025,
```

```
sales$month <- str_extract(sales$OrderDate, regex("(\\d+"))
```

```
View(sales)
```

```
as.integer(sales$month)
```

```
## [1] 1 8 5 4 9 10 8 4 4 5 12 10 3 5 6 2 11 6 7 1 7 1 4 7 2
## [26] 3 5 1 3 9 3 11 9 2 3 1 8 9 5 2 3 5 10 8 2 8 12 1 9 6
## [51] 1 11 10 8 7 8 10 2 1 10 4 4 3 10 2 6 5 8 11 11 2 5 10 11 6
## [76] 9 10 8 12 12 10 9 1 5 12 3 6 4 1 4 10 4 5 11 1 3 11 7 2 5
## [101] 9 9 2 1 12 4 1 2 12 7 8 2 10 5 1 8 7 12 11 5 1 12 5 12 7
## [126] 3 6 2 4 7 5 4 9 6 3 7 4 2 2 9 4 1 12 6 10 6 11 7 12 3
## [151] 4 3 8 7 4 6 7 3 1 10 2 2 10 3 6 9 11 5 3 11 2 10 6 10 12
## [176] 4 7 6 4 6 11 3 10 3 7 7 10 4 12 2 6 1 2 6 8 5 5 6 7 11
## [201] 2 11 3 10 1 8 7 1 9 6 1 7 7 9 6 1 6 7 11 4 7 2 7 12 7
```

```
## [226] 10 2 1 8 4 3 6 5 12 12 2 12 10 7 9 7 10 10 5 12 6 7 5 9 9
## [251] 6 1 5 5 10 2 7 8 10 9 1 3 11 2 11 6 11 5 10 9 10 2 9 10 12
## [276] 3 10 8 10 5 10 2 7 3 1 8 2 1 2 8 3 6 8 5 4 10 10 9 5 6
## [301] 11 5 2 6 7 9 3 12 11 5 3 1 6 8 5 12 12 4 3 11 5 8 1 7 6
## [326] 3 11 11 3 7 2 3 8 10 11 12 2 9 8 8 5 6 8 10 12 9 12 7 2 12
## [351] 9 12 9 7 6 7 10 5 9 4 2 8 12 5 11 12 11 11 5 8 4 4 1 1 6
## [376] 3 2 7 5 3 1 3 2 7 5 7 1 1 10 2 7 8 6 9 12 10 3 3 8 12
## [401] 12 7 12 9 6 10 10 8 5 9 8 1 2 10 11 3 5 10 2 5 1 9 12 12 9
## [426] 8 10 1 10 7 4 1 1 9 10 2 2 11 10 4 10 5 6 9 12 6 3 3 8 3
## [451] 7 4 8 12 7 10 11 8 8 7 1 3 6 3 5 7 5 9 3 7 6 12 6 6 8
## [476] 4 3 12 5 12 10 2 2 5 12 12 7 12 10 8 8 7 11 7 10
```

```
Jan <- filter(.data = sales, sales$month == 1)
Feb <- filter(.data = sales, sales$month == 2)
Mar <- filter(.data = sales, sales$month == 3)
Apr <- filter(.data = sales, sales$month == 4)
May <- filter(.data = sales, sales$month == 5)
Jun <- filter(.data = sales, sales$month == 6)
Jul <- filter(.data = sales, sales$month == 7)
Aug <- filter(.data = sales, sales$month == 8)
Sep <- filter(.data = sales, sales$month == 9)
Oct <- filter(.data = sales, sales$month == 10)
Nov <- filter(.data = sales, sales$month == 11)
Dec <- filter(.data = sales, sales$month == 12)
```

```
months <- c(Jan,Feb,Mar,Apr,May,Jun,Jul,Aug,Sep,Oct,Nov,Dec)
```

```
salesJan <- sum(Jan$TotalSales)
salesFeb <- sum(Feb$TotalSales)
salesMar <- sum(Mar$TotalSales)
salesApr <- sum(Apr$TotalSales)
salesMay <- sum(May$TotalSales)
salesJun <- sum(Jun$TotalSales)
salesJul <- sum(Jul$TotalSales)
salesAug <- sum(Aug$TotalSales)
salesSep <- sum(Sep$TotalSales)
salesOct <- sum(Oct$TotalSales)
salesNov <- sum(Nov$TotalSales)
salesDec <- sum(Dec$TotalSales)
```

```
df <- rbind(salesJan,salesFeb,salesMar,salesApr,salesMay,salesJun,salesJul,salesAug,salesSep,salesOct,salesNov,salesDec)
```

```
View(df)
```

```
radial.plot(
  lengths = df,
  radial.pos = seq(0,22*pi/12, length = 12),
  labels = c('Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec'),
  xlab = '',
  main = 'Total Sales by Month'
)
```

## Total Sales by Month

