Assignment 7

Apurva Hari

August 16, 2016

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```r
library(igraph)
```

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

```r
data("mtcars")
mcor <- cor(mtcars)
# Print mcor and round to 2 digits round(mcor, digits=2)
round(mcor, digits=2)
```

```
##        mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
## mpg   1.00 -0.85 -0.85 -0.78  0.68 -0.87  0.42  0.66  0.60  0.48 -0.55
## cyl  -0.85  1.00  0.90  0.83 -0.70  0.78 -0.59 -0.81 -0.52 -0.49  0.53
## disp -0.85  0.90  1.00  0.79 -0.71  0.89 -0.43 -0.71 -0.59 -0.56  0.39
## hp   -0.78  0.83  0.79  1.00 -0.45  0.66 -0.71 -0.72 -0.24 -0.13  0.75
## drat  0.68 -0.70 -0.71 -0.45  1.00 -0.71  0.09  0.44  0.71  0.70 -0.09
## wt   -0.87  0.78  0.89  0.66 -0.71  1.00 -0.17 -0.55 -0.69 -0.58  0.43
## qsec  0.42 -0.59 -0.43 -0.71  0.09 -0.17  1.00  0.74 -0.23 -0.21 -0.66
## vs    0.66 -0.81 -0.71 -0.72  0.44 -0.55  0.74  1.00  0.17  0.21 -0.57
## am    0.60 -0.52 -0.59 -0.24  0.71 -0.69 -0.23  0.17  1.00  0.79  0.06
## gear  0.48 -0.49 -0.56 -0.13  0.70 -0.58 -0.21  0.21  0.79  1.00  0.27
## carb -0.55  0.53  0.39  0.75 -0.09  0.43 -0.66 -0.57  0.06  0.27  1.00
```
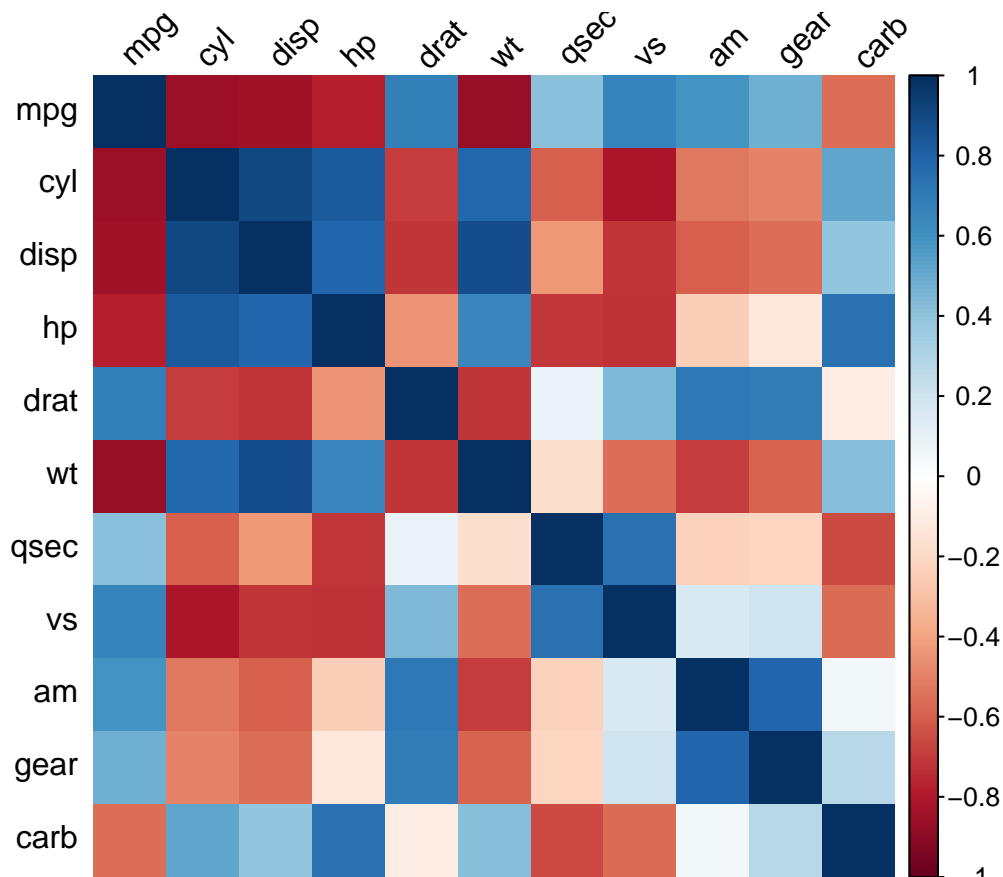
```r
#correlation plot
library(corrplot)
corrplot(mcor)
```
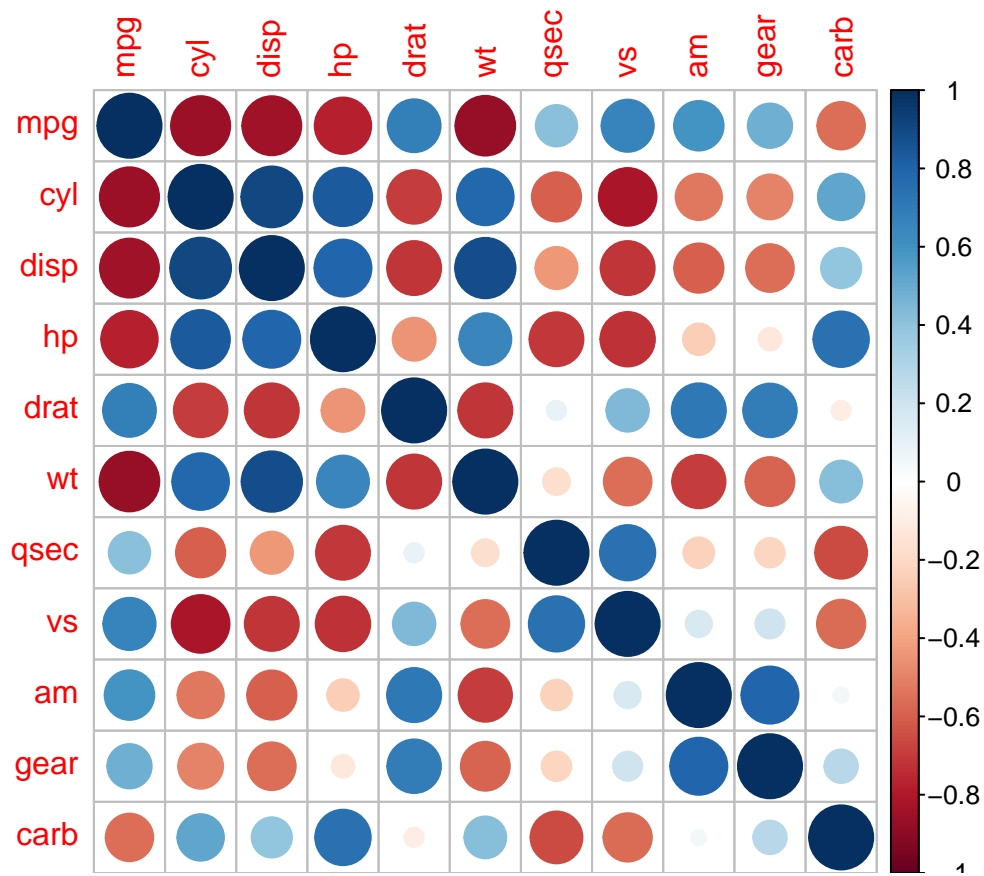
```
#a correlation matrix with colored squares and black labels, rotated 45 degrees along the top
corrplot(mcor, method="shade", shade.col=NA, tl.col="black", tl.srt=45)
```

```
# Generate a lighter palette

col <- colorRampPalette(c("#BB4444", "#EE9988", "#FFFFFF", "#77AADD", "#4477AA"))
corrplot(mcor, method="shade", shade.col=NA, tl.col="black", tl.srt=45,          col=col(200), addCoef.co
```

```
## Warning in text.default(pos.xlabel[, 1], pos.xlabel[, 2], newcolnames, srt
## = tl.srt, : "addcolorlabel" is not a graphical parameter
```
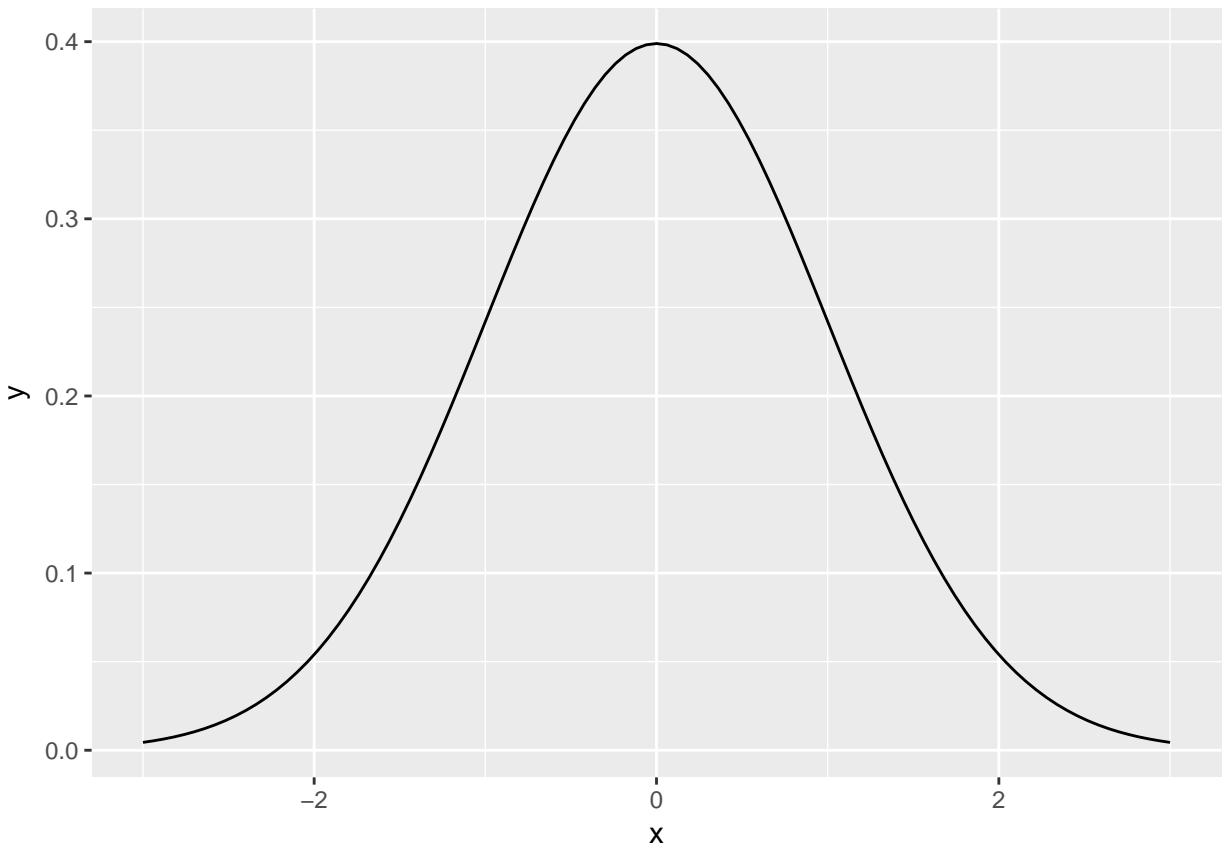
```
## Warning in text.default(pos.ylabel[, 1], pos.ylabel[, 2], newrownames, col
## = tl.col, : "addcolorlabel" is not a graphical parameter
```

```
## Warning in title(title, ...): "addcolorlabel" is not a graphical parameter
```
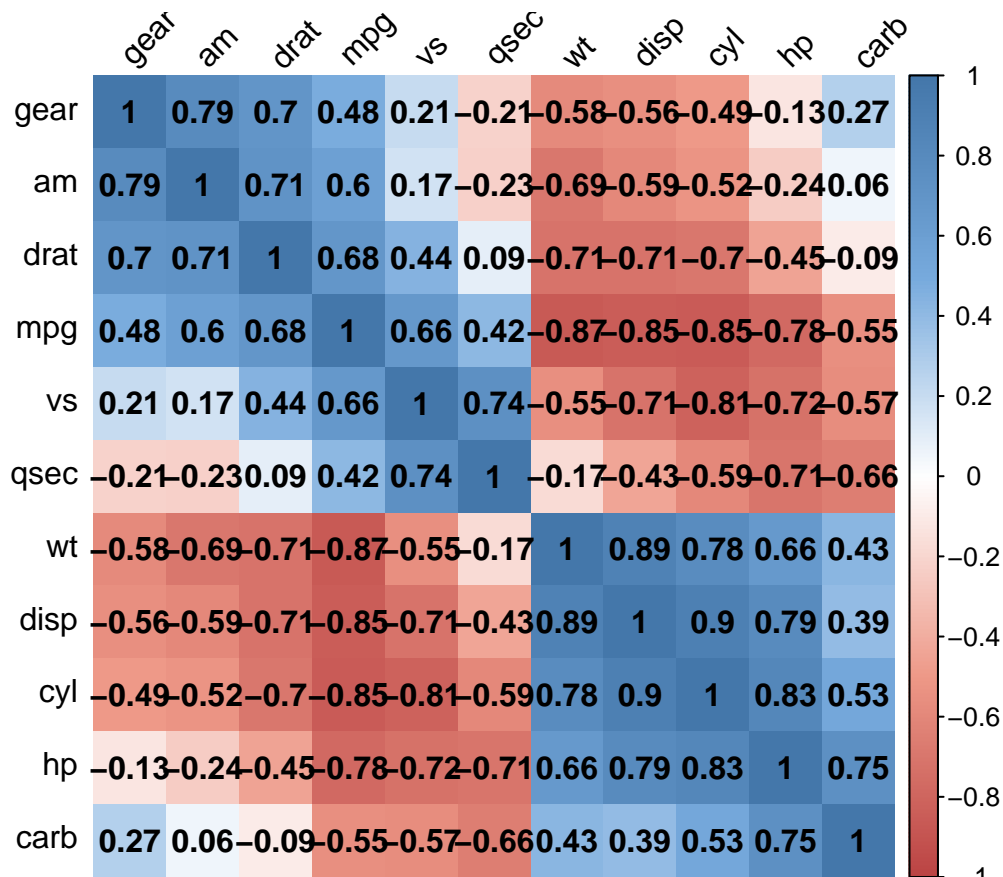
```r
#Plotting a function
# The data frame is only used for setting the range

p <- ggplot(data.frame(x=c(-3,3)), aes(x=x))
p + stat_function(fun = dnorm)
```
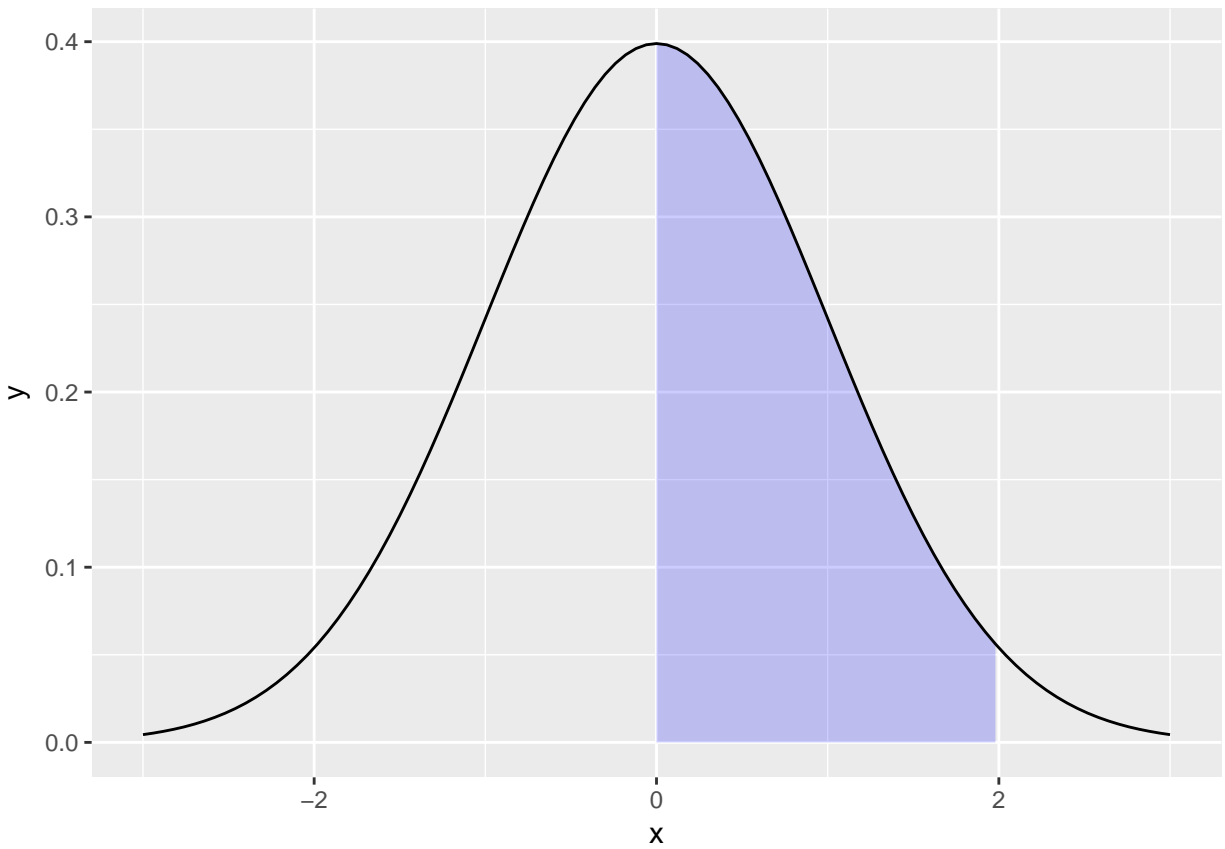
```
#Shading a part of the area under a function curve.

# Return dnorm(x) for 0 < x < 2, and NA for all other x

dnorm_limit <- function(x)
{
  y <- dnorm(x)
  y[x < 0  |  x > 2] <- NA
  return(y)
  }


# ggplot() with dummy data

p <- ggplot(data.frame(x=c(-3, 3)), aes(x=x))
p + stat_function(fun=dnorm_limit, geom="area", fill="blue", alpha=0.2) +    stat_function(fun=dnorm)
```
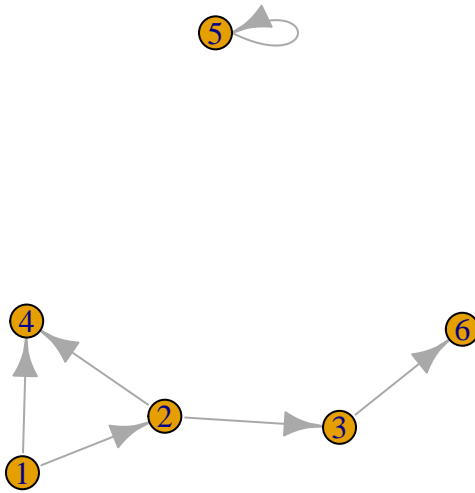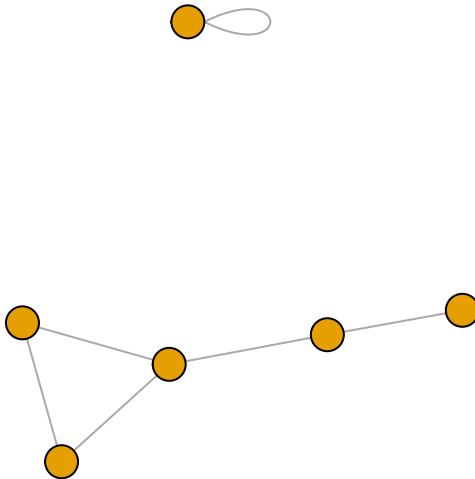
```
#Creating a Network Graph

# Loading the package igraph
library(igraph)
# Specify edges for a directed graph
gd <- graph(c(1,2, 2,3, 2,4, 1,4, 5,5, 3,6))
plot(gd)
```

```r
# For an undirected graph
gu <- graph(c(1,2, 2,3, 2,4, 1,4, 5,5, 3,6), directed=FALSE)

# No labels
plot(gu, vertex.label=NA)
```

```r
#Strcuture of each graph objects
str(gd)
```

```
## IGRAPH D--- 6 6 --
## + edges:
## [1] 1->2 2->3 2->4 1->4 5->5 3->6
```

```r
str(gu)
```

```
## IGRAPH U--- 6 6 --
## + edges:
## [1] 1--2 2--3 2--4 1--4 5--5 3--6
```

```r
#makes the output repeatable
set.seed(229)
plot(gu)
```
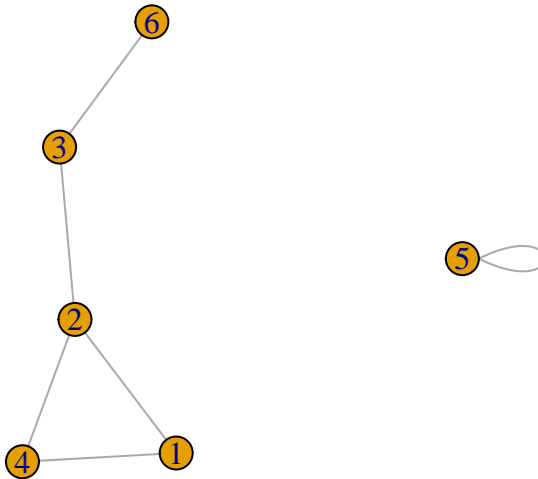
```
#A graph from a data frame
library(gcookbook) # For the data set

# Create a graph object from the data set
g <- graph.data.frame(madmen2, directed=TRUE)
# Remove unnecessary margins par(mar=c(0,0,0,0))
plot(g, layout=layout.fruchterman.reingold, vertex.size=8, edge.arrow.size=0.5,    vertex.label=NA)
```
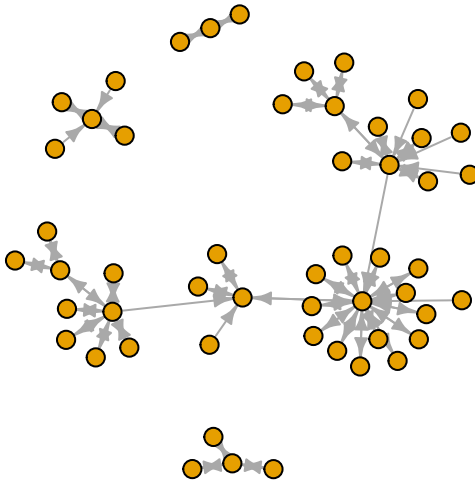
```
#A directed graph from a data frame
g <- graph.data.frame(madmen, directed=FALSE)
par(mar=c(0,0,0,0))
# Remove unnecessary margins
plot(g, layout=layout.circle, vertex.size=8, vertex.label=NA)
```
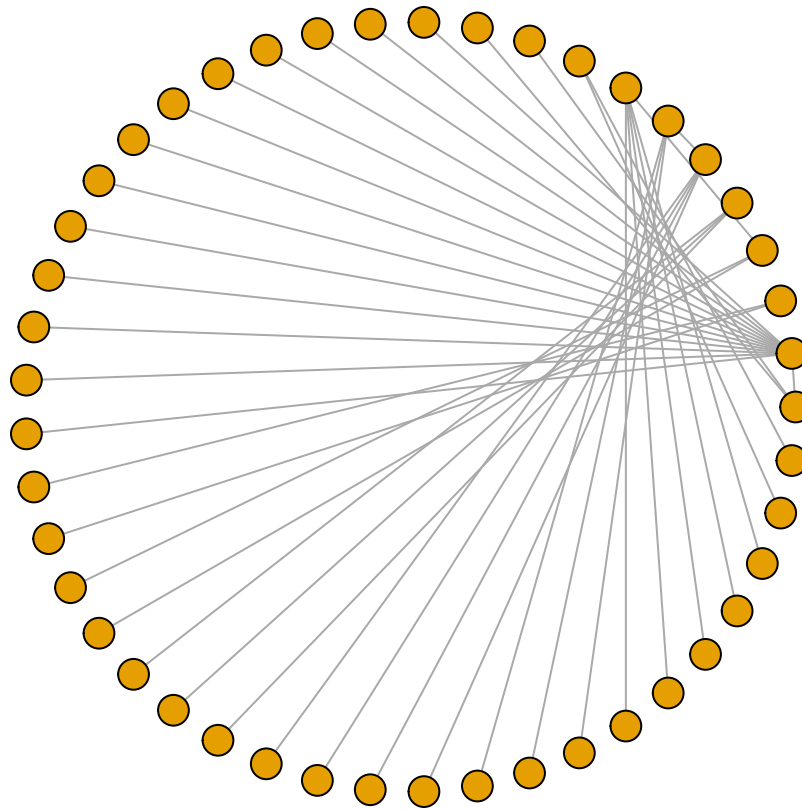
```r
#Text labels in a network graph

library(igraph)
library(gcookbook)
# For the data set
# Copy madmen and drop every other row
m <- madmen[1:nrow(madmen) %% 2 == 1, ]
g <- graph.data.frame(m, directed=FALSE)
# Print out the names of each vertex
V(g)$nam
```

NULL

```r
plot(g, layout=layout.fruchterman.reingold,
      vertex.size       = 4,         # Smaller nodes
    vertex.label       = V(g)$name, # Set the labels
    vertex.label.cex   = 0.8,       # Slightly smaller font
    vertex.label.dist  = 0.4,       # Offset the labels
    vertex.label.color = "black")
```
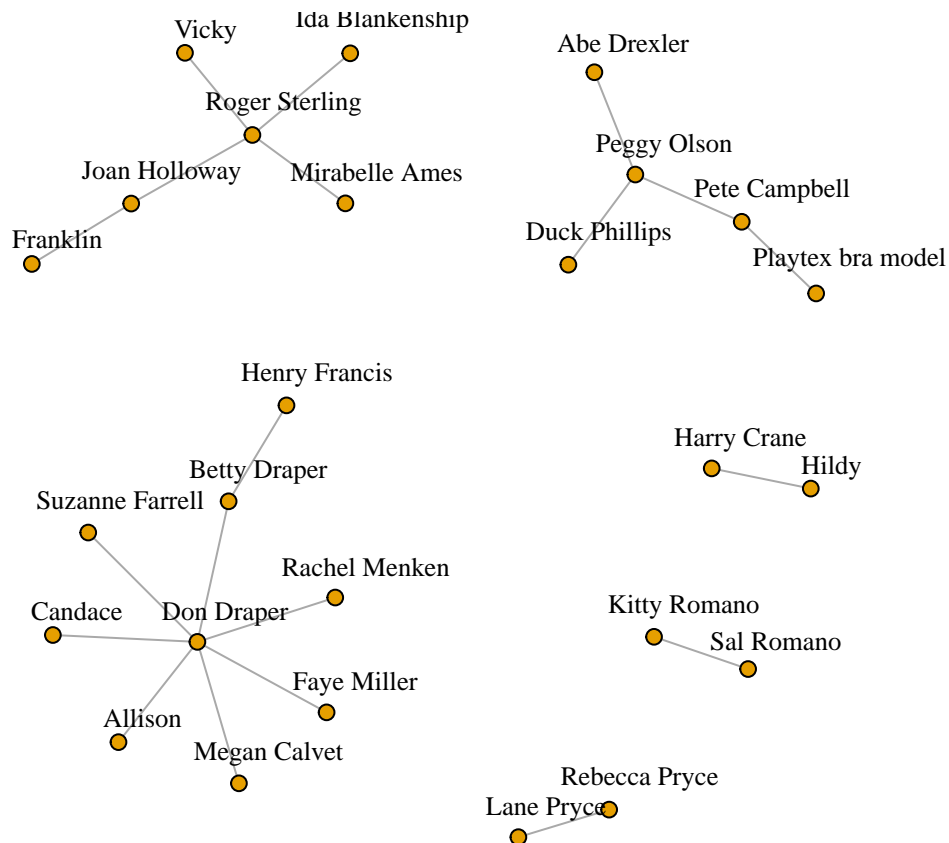
```
# View the edges
E(g)
```

```
## + 20/20 edges (vertex names):
##  [1] Betty Draper  --Henry Francis     Don Draper    --Allison
##  [3] Betty Draper  --Don Draper        Don Draper    --Candace
##  [5] Don Draper    --Faye Miller       Don Draper    --Megan Calvet
##  [7] Don Draper    --Rachel Menken     Don Draper    --Suzanne Farrell
##  [9] Harry Crane   --Hildy             Joan Holloway --Franklin
## [11] Joan Holloway --Roger Sterling    Lane Pryce    --Rebecca Pryce
## [13] Peggy Olson   --Abe Drexler       Peggy Olson   --Duck Phillips
## [15] Peggy Olson   --Pete Campbell     Pete Campbell --Playtex bra model
## [17] Roger Sterling--Ida Blankenship   Roger Sterling--Mirabelle Ames
## [19] Roger Sterling--Vicky             Sal Romano    --Kitty Romano
```

```
# Set some of the labels to "M"
E(g)[c(2,11,19)]$label <- "M"
# Set color of all to grey, and then color a few red
E(g)$color  <- "grey70"
E(g)[c(2,11,19)]$color <- "red"
plot(g)
```
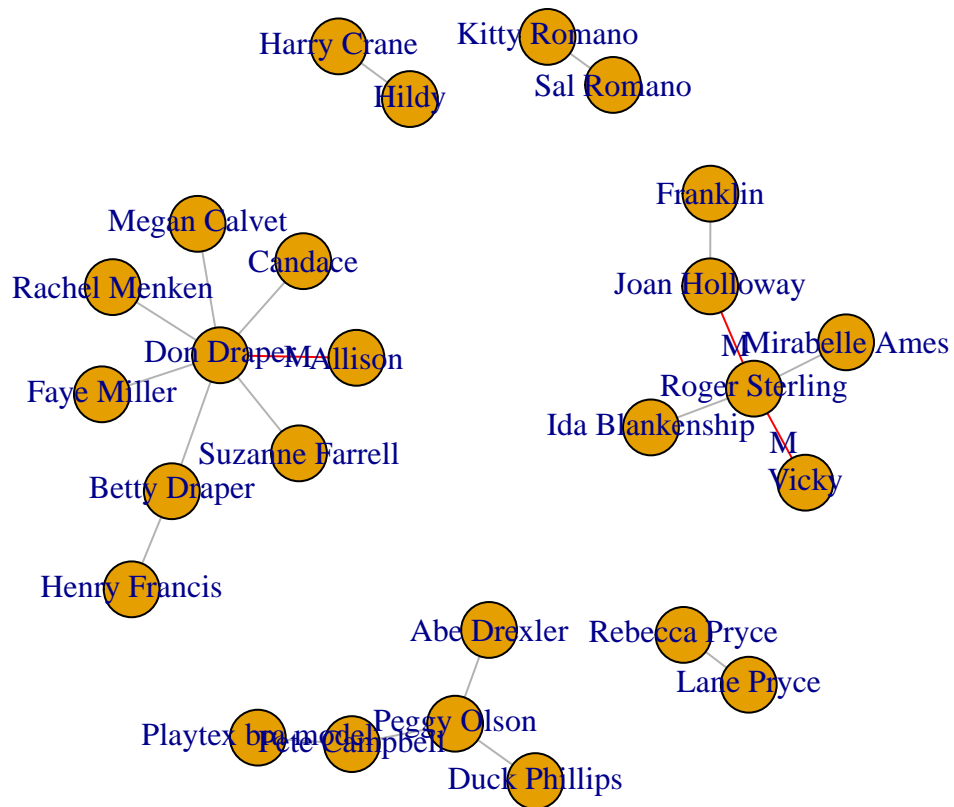
```
#Creating a Heat Map,using the presidents data set

pres_rating <- data.frame(rating  = as.numeric(presidents),
                          year = as.numeric(floor(time(presidents))),
                          quarter = as.numeric(cycle(presidents)) )
pres_rating
```

```
##      rating year quarter
## 1        NA 1945       1
## 2        87 1945       2
## 3        82 1945       3
## 4        75 1945       4
## 5        63 1946       1
## 6        50 1946       2
## 7        43 1946       3
## 8        32 1946       4
## 9        35 1947       1
## 10       60 1947       2
## 11       54 1947       3
## 12       55 1947       4
## 13       36 1948       1
## 14       39 1948       2
## 15       NA 1948       3
## 16       NA 1948       4
## 17       69 1949       1
## 18       57 1949       2
```

```
## 19      57 1949        3
## 20      51 1949        4
## 21      45 1950        1
## 22      37 1950        2
## 23      46 1950        3
## 24      39 1950        4
## 25      36 1951        1
## 26      24 1951        2
## 27      32 1951        3
## 28      23 1951        4
## 29      25 1952        1
## 30      32 1952        2
## 31      NA 1952        3
## 32      32 1952        4
## 33      59 1953        1
## 34      74 1953        2
## 35      75 1953        3
## 36      60 1953        4
## 37      71 1954        1
## 38      61 1954        2
## 39      71 1954        3
## 40      57 1954        4
## 41      71 1955        1
## 42      68 1955        2
## 43      79 1955        3
## 44      73 1955        4
## 45      76 1956        1
## 46      71 1956        2
## 47      67 1956        3
## 48      75 1956        4
## 49      79 1957        1
## 50      62 1957        2
## 51      63 1957        3
## 52      57 1957        4
## 53      60 1958        1
## 54      49 1958        2
## 55      48 1958        3
## 56      52 1958        4
## 57      57 1959        1
## 58      62 1959        2
## 59      61 1959        3
## 60      66 1959        4
## 61      71 1960        1
## 62      62 1960        2
## 63      61 1960        3
## 64      57 1960        4
## 65      72 1961        1
## 66      83 1961        2
## 67      71 1961        3
## 68      78 1961        4
## 69      79 1962        1
## 70      71 1962        2
## 71      62 1962        3
## 72      74 1962        4
```

```
## 73         76 1963          1
## 74         64 1963          2
## 75         62 1963          3
## 76         57 1963          4
## 77         80 1964          1
## 78         73 1964          2
## 79         69 1964          3
## 80         69 1964          4
## 81         71 1965          1
## 82         64 1965          2
## 83         69 1965          3
## 84         62 1965          4
## 85         63 1966          1
## 86         46 1966          2
## 87         56 1966          3
## 88         44 1966          4
## 89         44 1967          1
## 90         52 1967          2
## 91         38 1967          3
## 92         46 1967          4
## 93         36 1968          1
## 94         49 1968          2
## 95         35 1968          3
## 96         44 1968          4
## 97         59 1969          1
## 98         65 1969          2
## 99         65 1969          3
## 100        56 1969          4
## 101        66 1970          1
## 102        53 1970          2
## 103        61 1970          3
## 104        52 1970          4
## 105        51 1971          1
## 106        48 1971          2
## 107        54 1971          3
## 108        49 1971          4
## 109        49 1972          1
## 110        61 1972          2
## 111        NA 1972          3
## 112        NA 1972          4
## 113        68 1973          1
## 114        44 1973          2
## 115        40 1973          3
## 116        27 1973          4
## 117        28 1974          1
## 118        25 1974          2
## 119        24 1974          3
## 120        24 1974          4
```

```r
# Base plot
p <- ggplot(pres_rating, aes(x=year, y=quarter, fill=rating))
# Using geom_tile()
p + geom_tile()
```
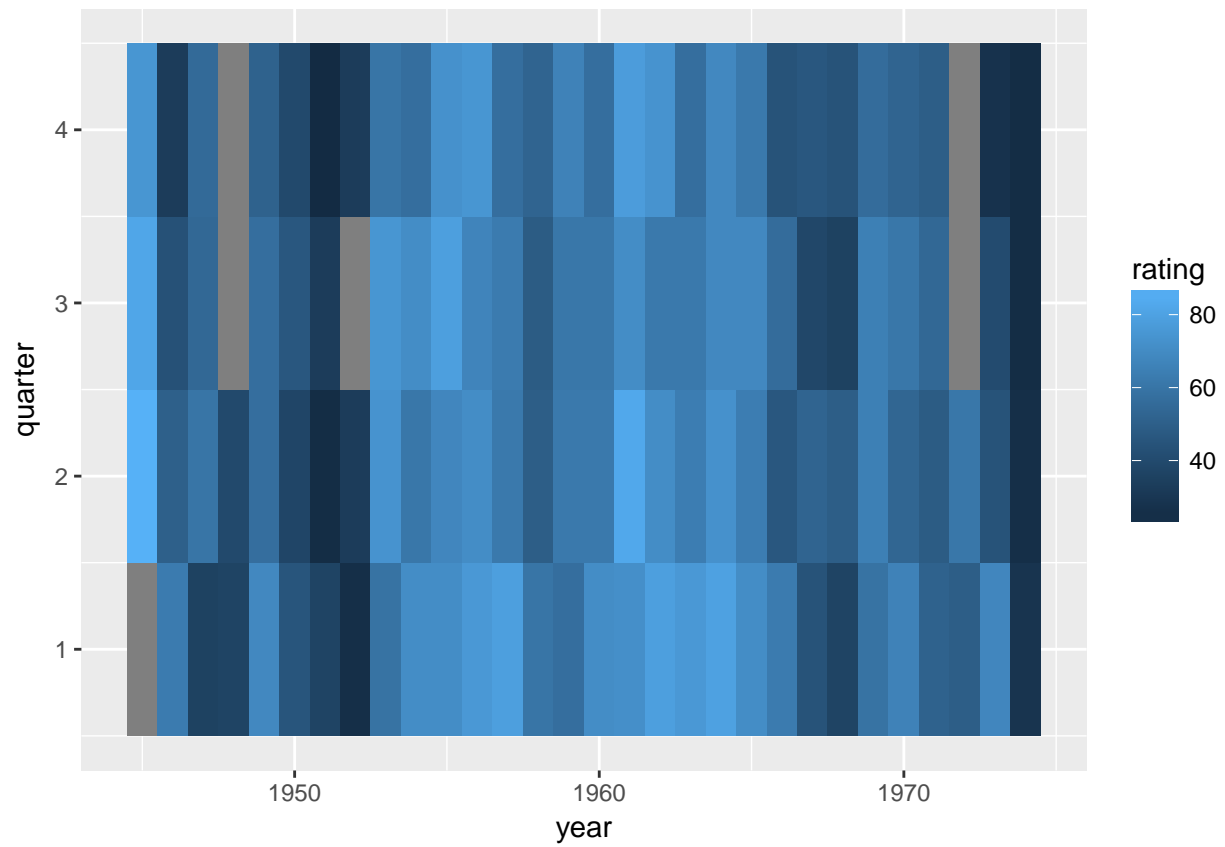
```
# Using geom_raster() - looks the same, but a little more efficient
p + geom_raster()
```
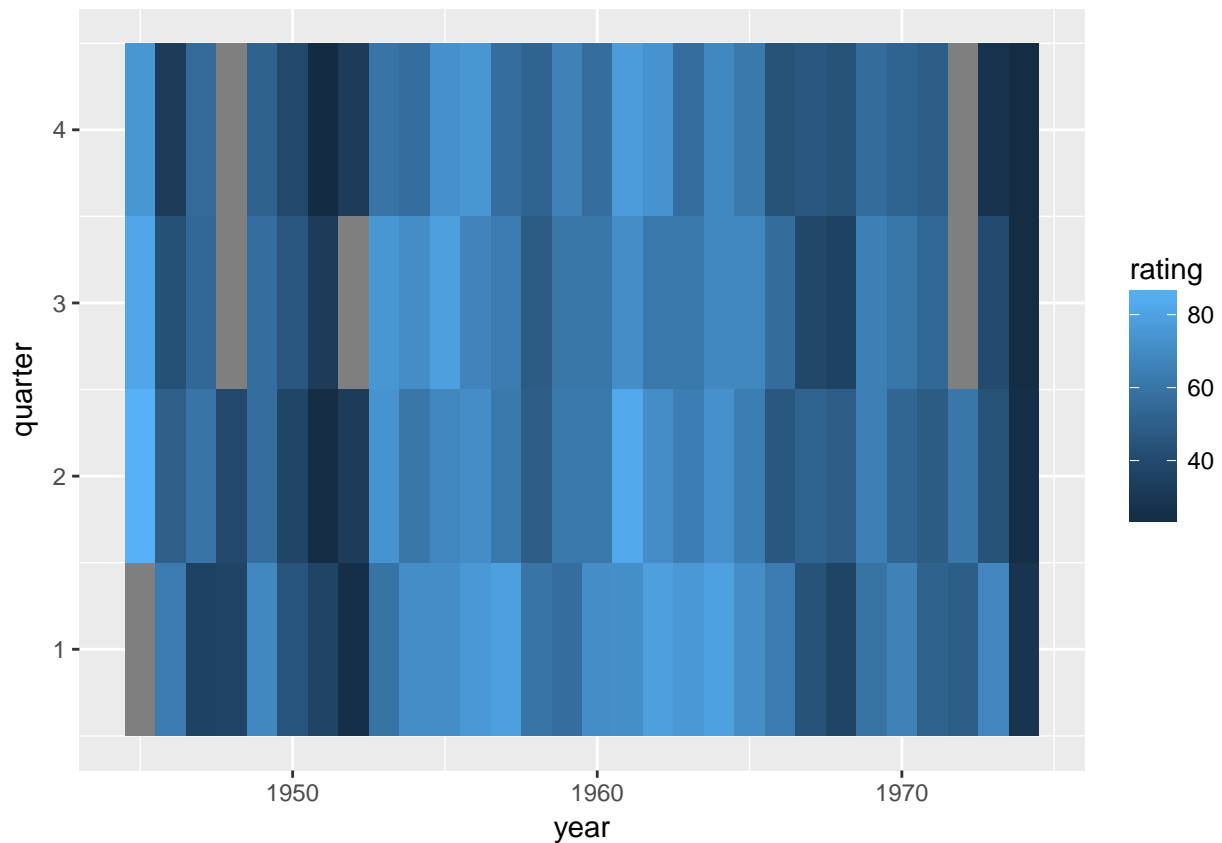
```
#Creating a 3D scatter-plot
library(rgl)
```

```
## Warning: package 'rgl' was built under R version 3.3.3
```

```
##
## Attaching package: 'rgl'
```

```
## The following object is masked from 'package:igraph':
##
##     %>%
```

```
plot3d(mtcars$wt, mtcars$disp, mtcars$mpg, type="s", size=0.75, lit=FALSE)

#Viewing the 3-d plot as a 2-d plot

# Function to interleave the elements of two vectors
interleave <- function(v1, v2)  as.vector(rbind(v1,v2))

# Make plot without axis ticks or labels
plot3d(mtcars$wt, mtcars$disp,
       mtcars$mpg,
       xlab = "",
       ylab = "", zlab = "",
```

```
        axes = FALSE,
        size=.75, type="s", lit=FALSE)

segments3d(interleave(mtcars$wt,   mtcars$wt),
           interleave(mtcars$disp, mtcars$disp),
           interleave(mtcars$mpg,  min(mtcars$mpg)),
           alpha = 0.4, col = "blue")

# Draw the box
rgl.bbox(color="grey50",          # grey60 surface and black text
         emission="grey50",       # emission color is grey50
         xlen=0, ylen=0, zlen=0)  # Don't add tick marks

# Set default color of future objects to black
rgl.material(color="black")


# Add axes to specific sides. Possible values are "x--", "x-+", "x+-", and "x++"

axes3d(edges=c("x--", "y+-", "z--"),
       ntick=6,                       # Attempt 6 tick marks on each side
       cex=.75)                       # Smaller font

# Add axis labels. 'line' specifies how far to set the label from the axis
mtext3d("Weight",       edge="x--", line=2)
mtext3d("Displacement", edge="y+-", line=3)
mtext3d("MPG",          edge="z--", line=3)


#Animating a three-dimensional plot by moving the viewpoint around the plot

# library(rgl)
# plot3d(mtcars$wt, mtcars$disp, mtcars$mpg, type="s", size=0.75, lit=FALSE)
# play3d(spin3d())
#
# # Spin on x-axis, at 4 rpm, for 5 seconds
# play3d(spin3d(axis=c(1,0,0), rpm=4), duration=5)
# # Spin on z axis, at 4 rpm, for 5 seconds
# movie3d(spin3d(axis=c(0,0,1), rpm=4), duration=5, fps=50)


#Creating a dendogram

library(gcookbook) # For the data set
# Get data from year 2009
c2 <- subset(countries, Year==2009)

# Drop rows that have any NA values
c2 <- c2[complete.cases(c2), ]

# Pick out a random 25 countries
# (Set random seed to make this repeatable)
set.seed(201)
c2 <- c2[sample(1:nrow(c2), 25), ]
```
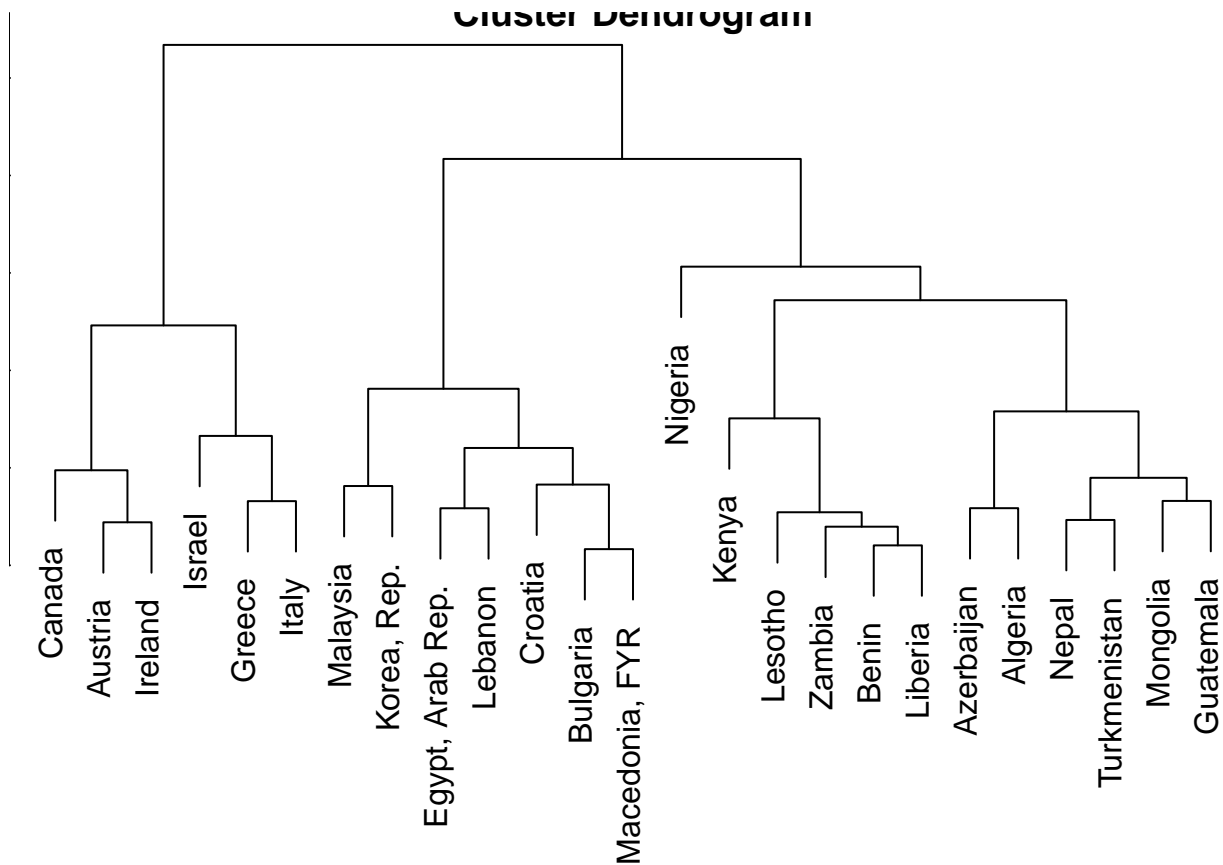
```
rownames(c2) <- c2$Name
c2 <- c2[,4:7]
c3 <- scale(c2)

hc <- hclust(dist(c3))
# Make the dendrogram
plot(hc)
```
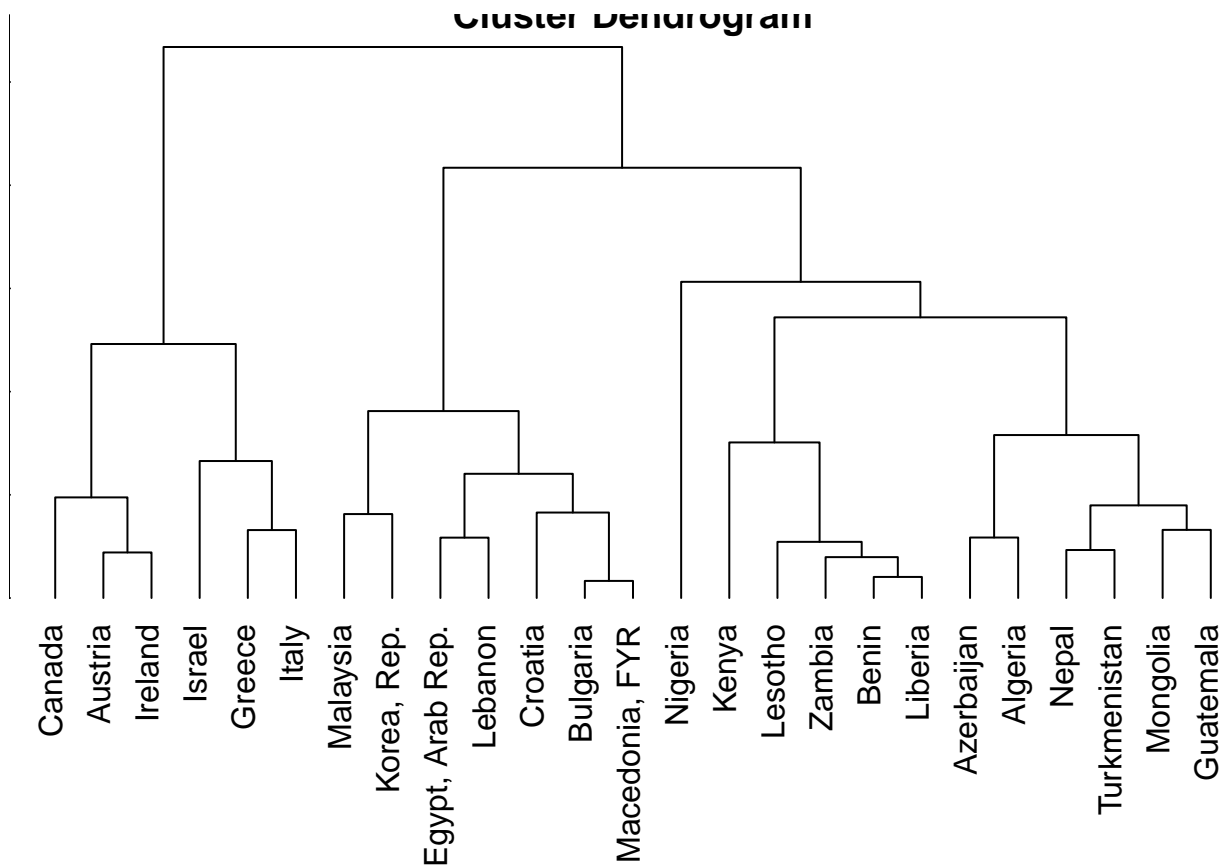

Cluster Dendrogram

```
# With text aligned
plot(hc, hang = -1)
```
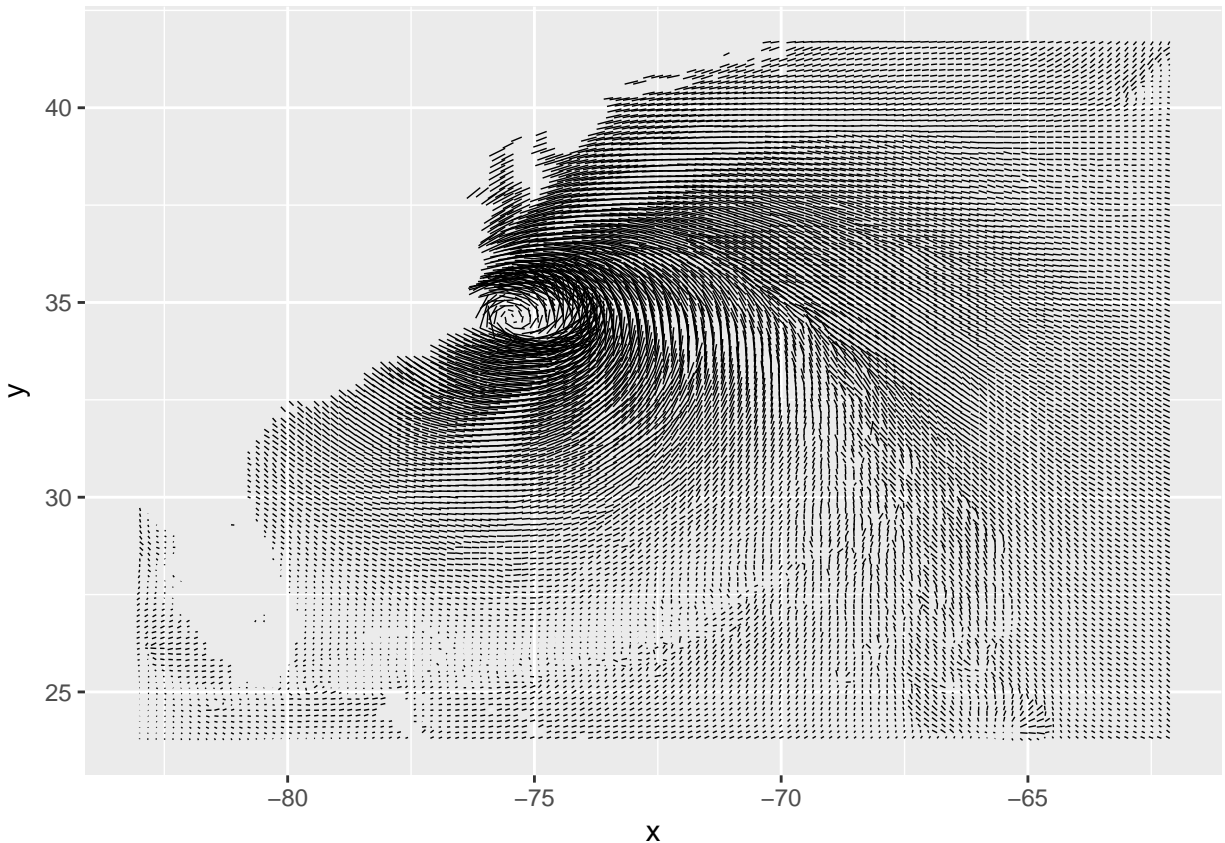
Cluster Dendrogram


```r
#Creating a Vector Field
library(gcookbook) # For the data set

islice <- subset(isabel, z == min(z))
ggplot(islice, aes(x=x, y=y)) + geom_segment(aes(xend = x + vx/50, yend = y + vy/50),
```

```
## Warning: Removed 3745 rows containing missing values (geom_segment).
```
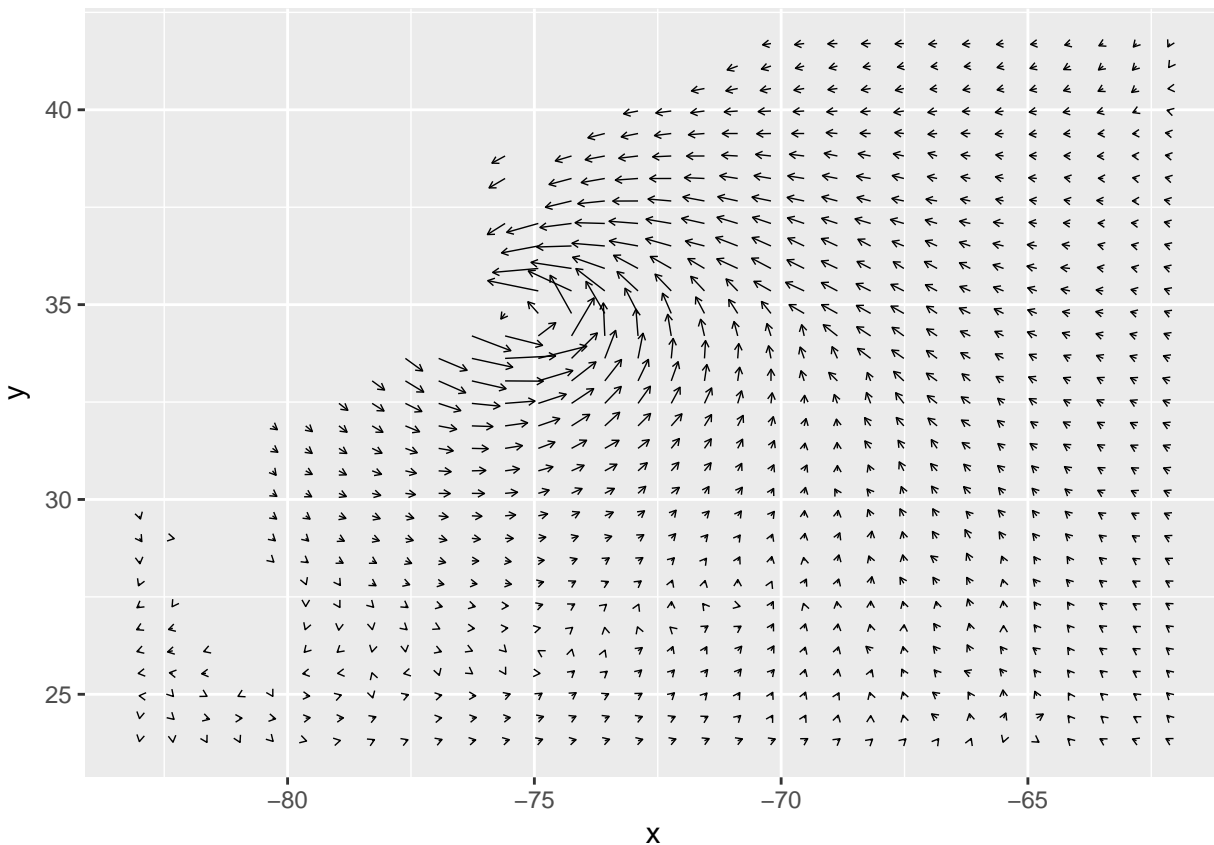
```
# Take a slice where z is equal to the minimum value of z
islice <- subset(isabel, z == min(z))
# Keep 1 out of every 'by' values in vector x
every_n <- function(x, by = 2) {
  x <- sort(x)
  x[seq(1, length(x), by = by)]
  }

# Keep 1 of every 4 values in x and y
keepx <- every_n(unique(isabel$x), by=4)
keepy <- every_n(unique(isabel$y), by=4)

# Keep only those rows where x value is in keepx and y value is in keepy
islicesub <- subset(islice, x %in% keepx  &  y %in% keepy)

# Need to load grid for arrow() function
library(grid)
# Make the plot with the subset, and use an arrowhead 0.1 cm long
ggplot(islicesub, aes(x=x, y=y)) +geom_segment(aes(xend = x+vx/50, yend = y+vy/50),          arro
```

```
## Warning: Removed 248 rows containing missing values (geom_segment).
```

```
# The existing 'speed' column includes the z component. We'll calculate # speedxy, the horizontal speed
islicesub$speedxy <- sqrt(islicesub$vx^2 + islicesub$vy^2)
# Map speed to alpha
ggplot(islicesub, aes(x=x, y=y)) + geom_segment(aes(xend = x+vx/50, yend = y+vy/50, alpha = speed),arro
```
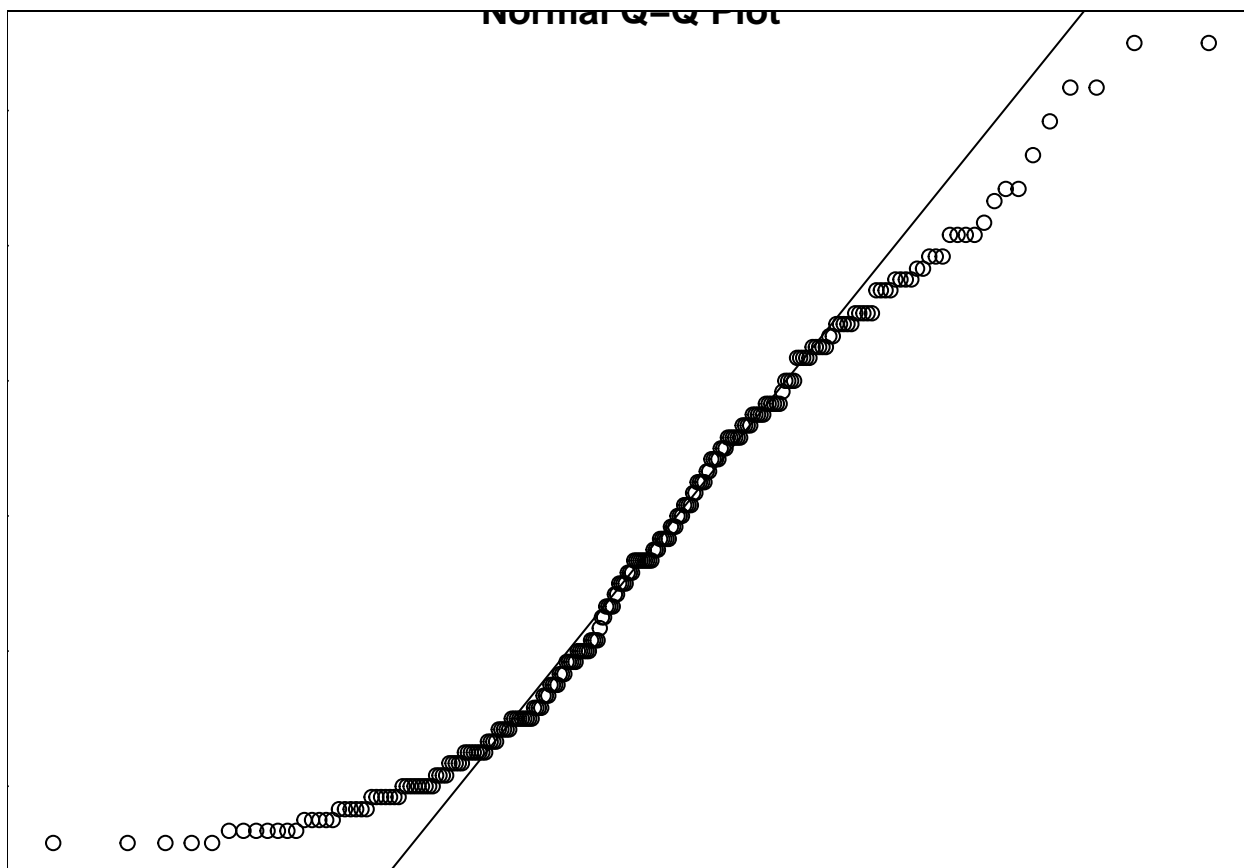
Warning: Removed 248 rows containing missing values (geom_segment).
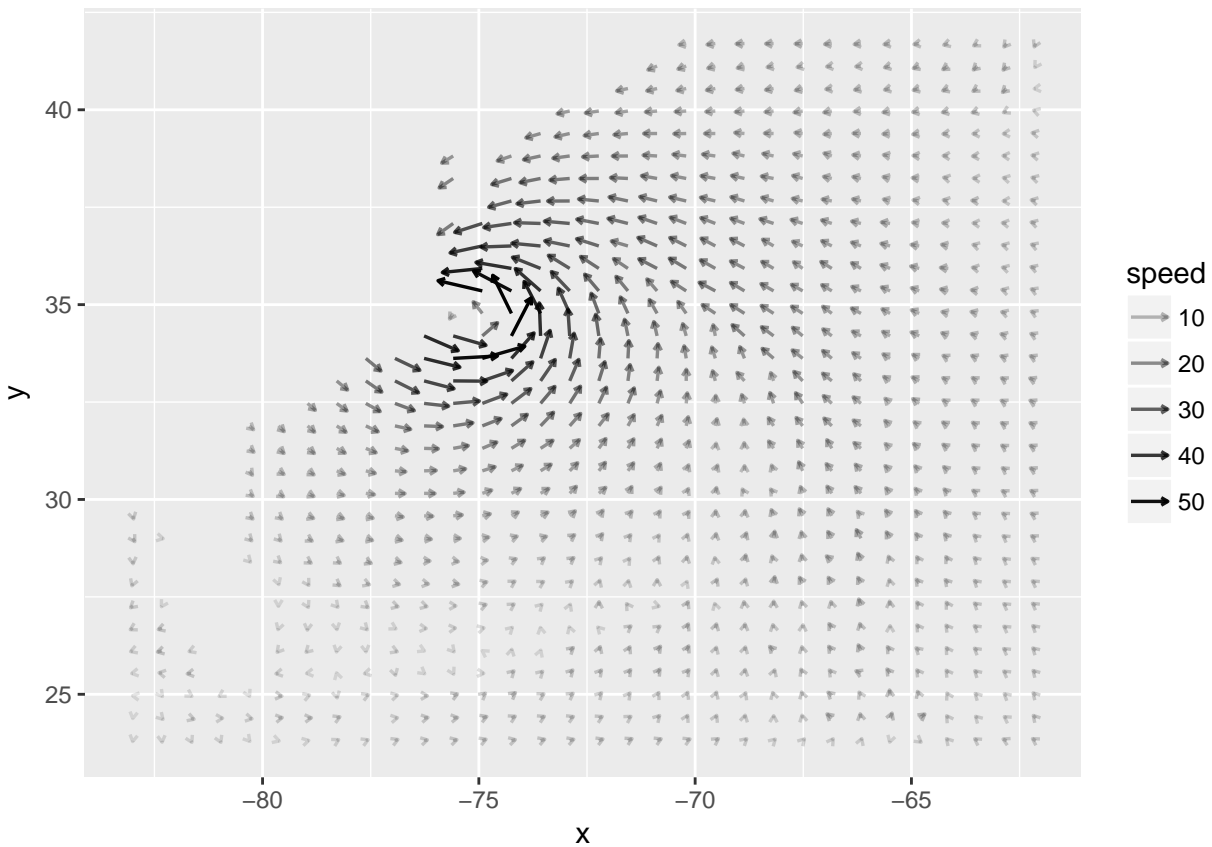
```
#Creating a qqplot to compare empirical to a thoeritical distribution

library(gcookbook) # For the data set
# QQ plot of height
qqnorm(heightweight$heightIn)
qqline(heightweight$heightIn)
```

Normal Q-Q Plot
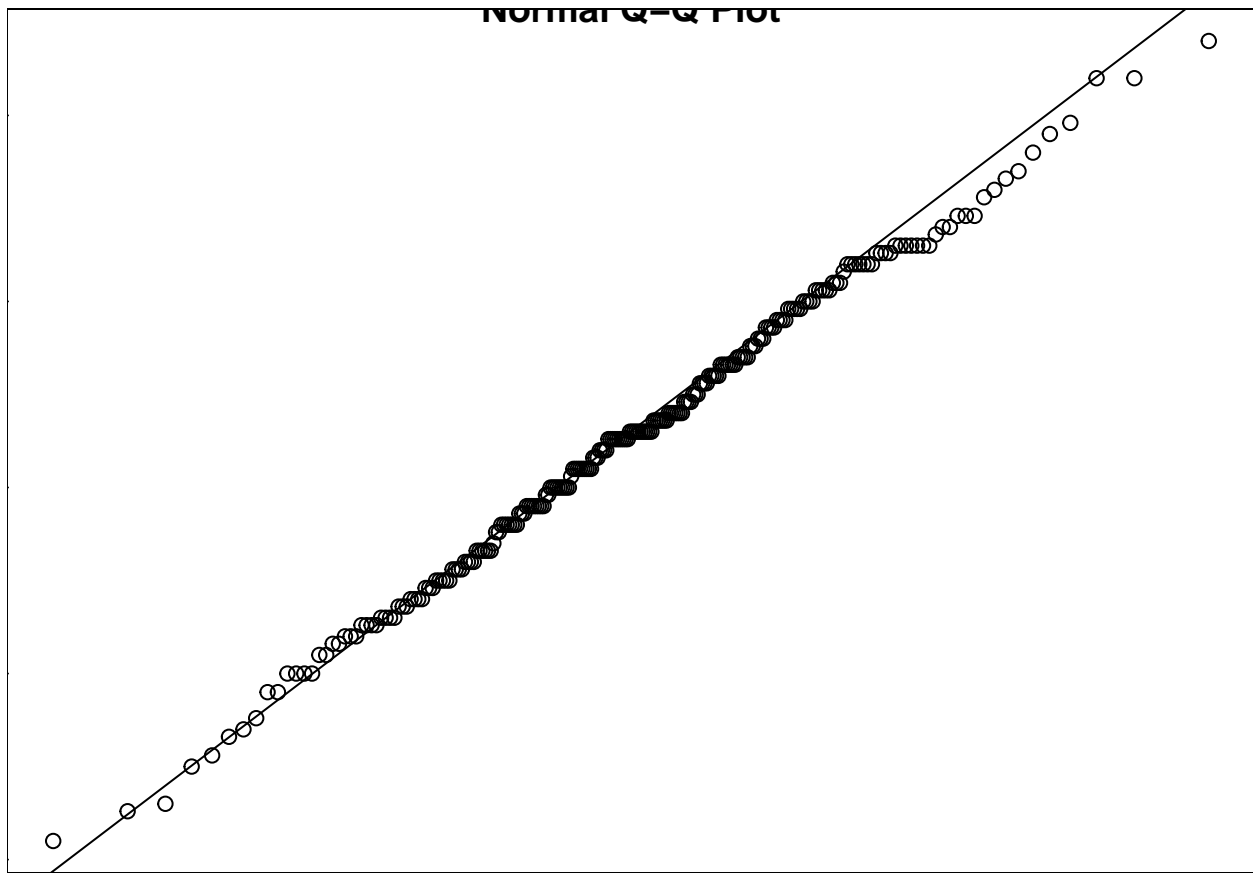

```
# QQ plot of age
qqnorm(heightweight$ageYear)
qqline(heightweight$ageYear)
```

Normal Q–Q Plot


```r
#Creating a Mosaic Plot

# Print a "flat" contingency table
ftable(UCBAdmissions)
```
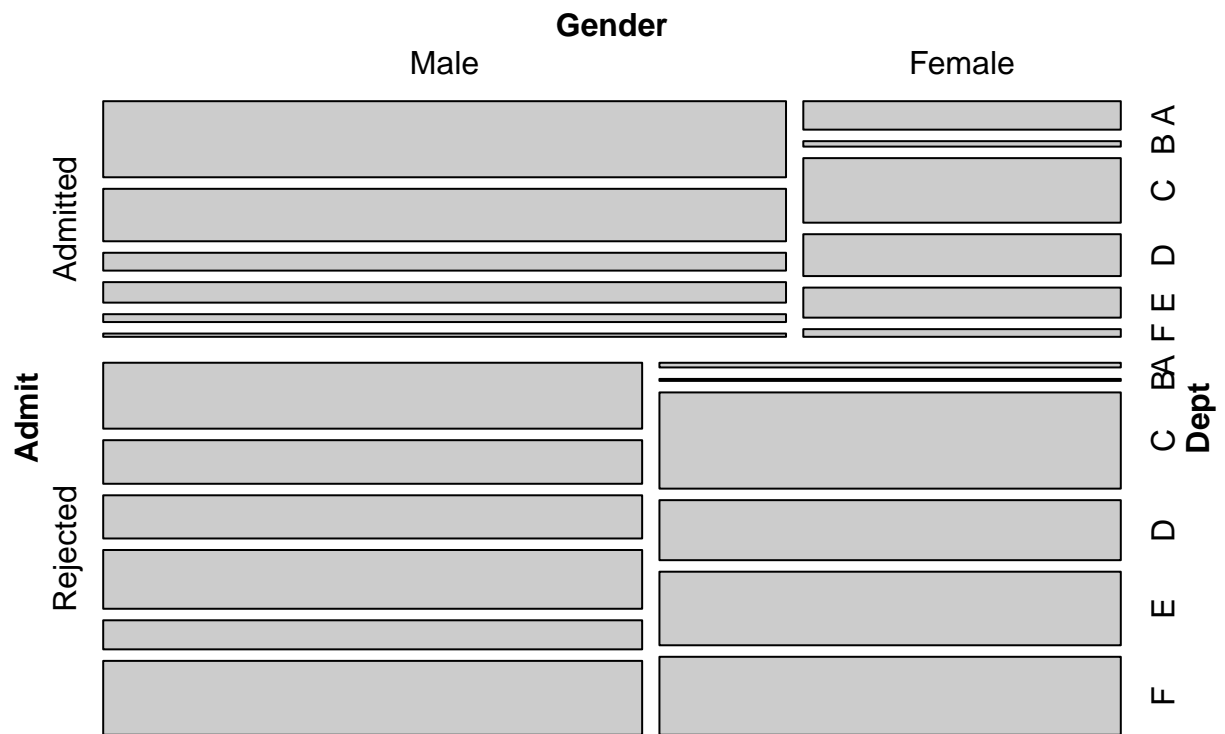
```
##                 Dept   A   B   C   D   E   F
## Admit    Gender
## Admitted Male         512 353 120 138  53  22
##          Female        89  17 202 131  94  24
## Rejected Male         313 207 205 279 138 351
##          Female        19   8 391 244 299 317
```
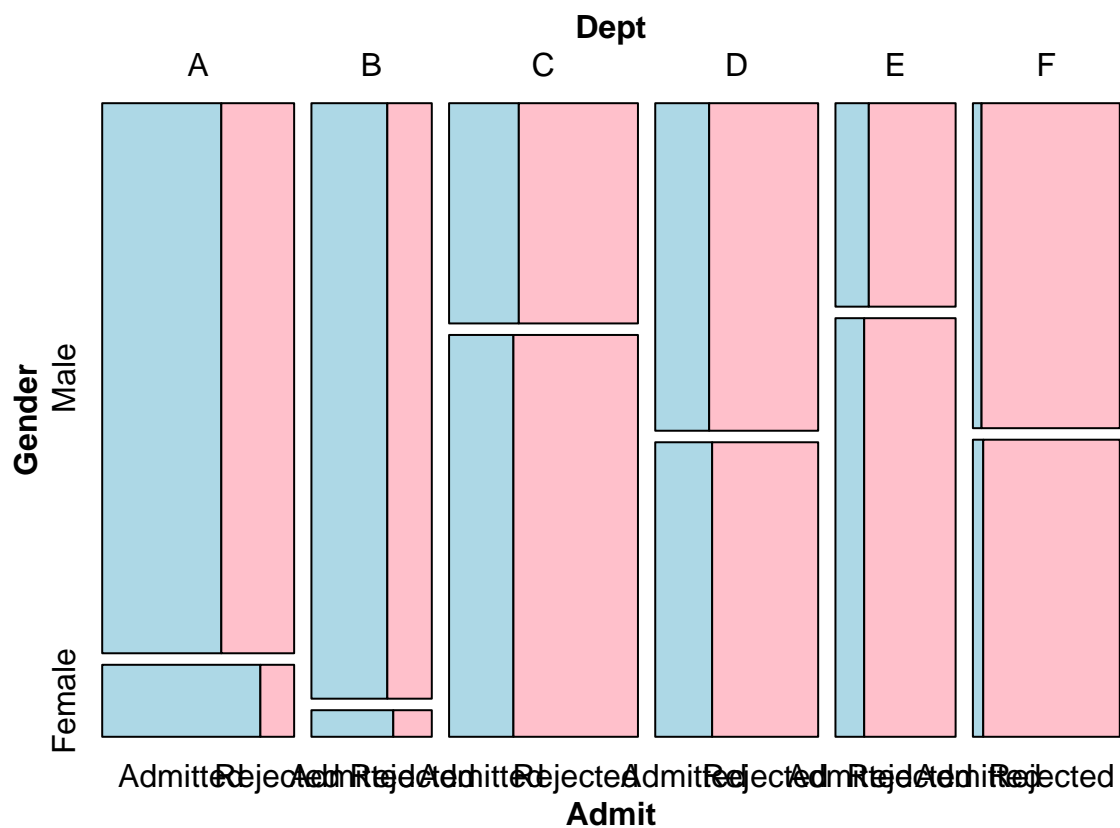
```r
library(vcd)
```

```
## Warning: package 'vcd' was built under R version 3.3.2
```
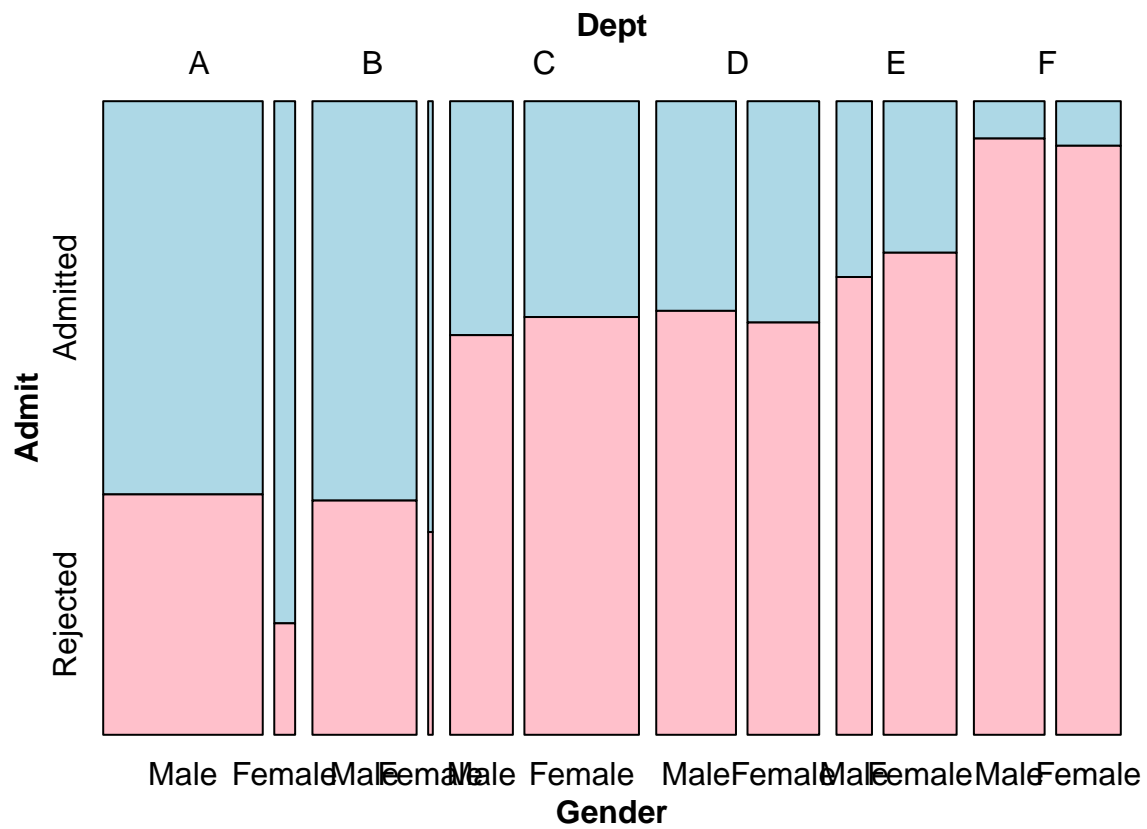
```r
# Split by Admit, then Gender, then Dept
mosaic( ~ Admit + Gender + Dept, data=UCBAdmissions)
```

Gender

Male   Female



```
mosaic( ~ Dept + Gender + Admit, data=UCBAdmissions,    highlighting="Admit", highlighting_fill=c("ligh
```

```
# Another possible set of splitting directions
mosaic( ~ Dept + Gender + Admit, data=UCBAdmissions,    highlighting="Admit", highlighting_fill=c("ligh
```

```
# This order makes it difficult to compare male and female
mosaic( ~ Dept + Gender + Admit, data=UCBAdmissions,     highlighting="Admit", highlighting_fill=c("ligh
```
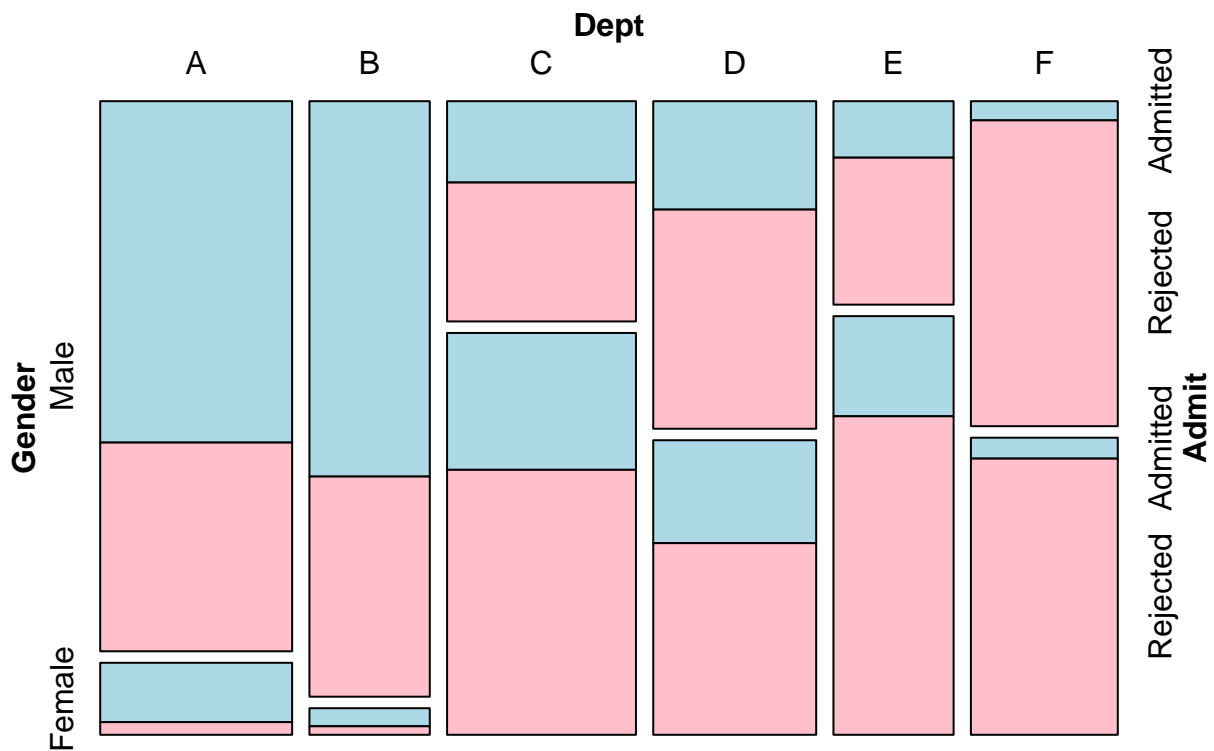
```
#Creating a Pie Chart

library(MASS)   # For the data set
# Get a table of how many cases are in each level of fold
fold <- table(survey$Fold)
fold
```

```
##
##  L on R Neither  R on L
##      99      18     120
```

```
# Make the pie chart
pie(fold)
pie(c(99, 18, 120), labels=c("L on R", "Neither", "R on L"))
```