

# R Section

we are going to be focused on histograms, box plots, and bullet charts and using various tools to create these visualizations

## Data read and preparation

In [8]:

```
library('magrittr')
source("BulletGraph.R", local=TRUE)

# load birth rate data
birthrate <- read.csv('ex6-2/birth-rate.csv')

# load crime data
crime <- read.csv('ex6-2/crimeratesbystate-formatted.csv')

# load education data
education <- read.csv('ex6-2/education.csv')

# check column names
colnames(birthrate)
```

1. 'Country'  
2. 'X1960'  
3. 'X1961'  
4. 'X1962'  
5. 'X1963'  
6. 'X1964'  
7. 'X1965'  
8. 'X1966'  
9. 'X1967'  
10. 'X1968'  
11. 'X1969'  
12. 'X1970'  
13. 'X1971'  
14. 'X1972'  
15. 'X1973'  
16. 'X1974'  
17. 'X1975'  
18. 'X1976'  
19. 'X1977'  
20. 'X1978'  
21. 'X1979'  
22. 'X1980'  
23. 'X1981'  
24. 'X1982'  
25. 'X1983'  
26. 'X1984'  
27. 'X1985'  
28. 'X1986'  
29. 'X1987'  
30. 'X1988'

31. 'X1989'  
32. 'X1990'  
33. 'X1991'  
34. 'X1992'  
35. 'X1993'  
36. 'X1994'  
37. 'X1995'  
38. 'X1996'  
39. 'X1997'  
40. 'X1998'  
41. 'X1999'  
42. 'X2000'  
43. 'X2001'  
44. 'X2002'  
45. 'X2003'  
46. 'X2004'  
47. 'X2005'  
48. 'X2006'  
49. 'X2007'  
50. 'X2008'

In [2]:

```
# format year columns
colnames(birthrate) <- gsub("X", "", colnames(birthrate))

# check column names
colnames(birthrate)
```

1. 'Country'  
2. '1960'  
3. '1961'  
4. '1962'  
5. '1963'  
6. '1964'  
7. '1965'  
8. '1966'  
9. '1967'  
10. '1968'  
11. '1969'  
12. '1970'  
13. '1971'  
14. '1972'  
15. '1973'  
16. '1974'  
17. '1975'  
18. '1976'  
19. '1977'  
20. '1978'  
21. '1979'

22. '1980'  
23. '1981'  
24. '1982'  
25. '1983'  
26. '1984'  
27. '1985'  
28. '1986'  
29. '1987'  
30. '1988'  
31. '1989'  
32. '1990'  
33. '1991'  
34. '1992'  
35. '1993'  
36. '1994'  
37. '1995'  
38. '1996'  
39. '1997'  
40. '1998'  
41. '1999'  
42. '2000'  
43. '2001'  
44. '2002'  
45. '2003'  
46. '2004'  
47. '2005'  
48. '2006'  
49. '2007'  
50. '2008'

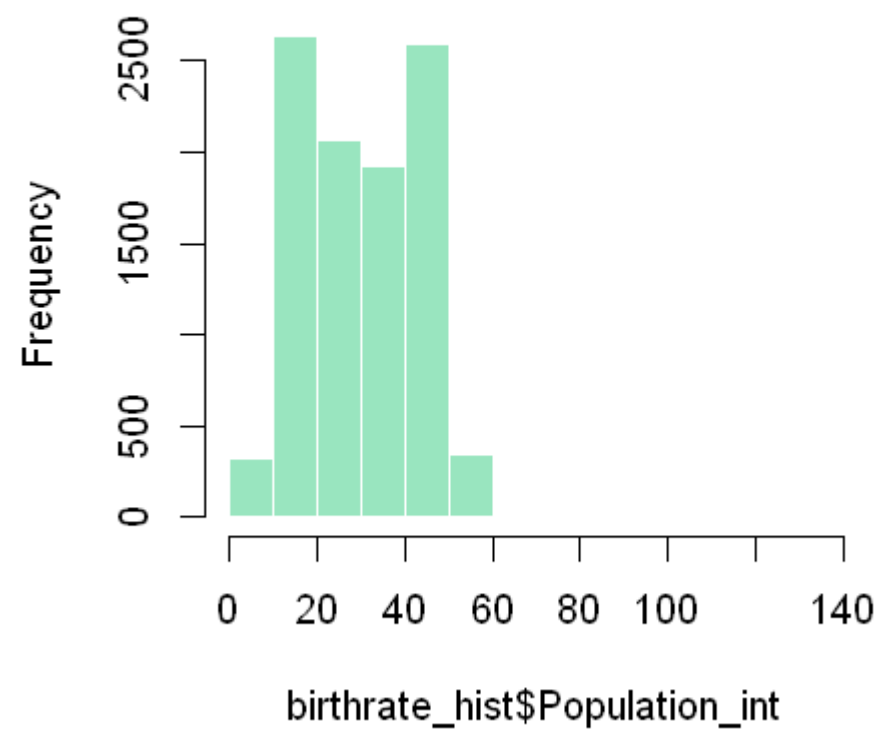
## Histogram

In [5]:

```
options(repr.plot.width = 4, repr.plot.height = 4)

# create pivotted data for plotting
birthrate_hist <- reshape2::melt(birthrate, id=c("Country")) %>%
  dplyr::mutate("Country" = as.character(Country),
               "Year" = as.character(variable),
               "Population" = value,
               "Population_int" = ceiling(value)) %>%
  dplyr::select(c("Country", "Year", "Population", "Population_int"))

# create histogram of population data
hist(birthrate_hist$Population_int, col=rgb(0.2,0.8,0.5,0.5) , border=F , main="")
```

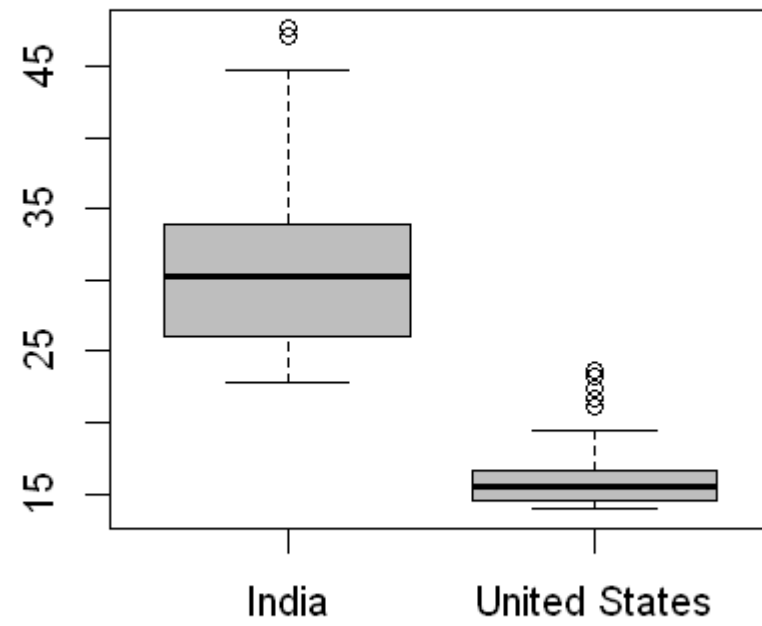


## Box plot

In [6]:

```
# create box plot of population data
birthrate_box <- birthrate_hist %>%
  dplyr::filter(Country %in% c("United States", "India"))

boxplot(birthrate_box$Population ~ birthrate_box$Country , col="grey")
```

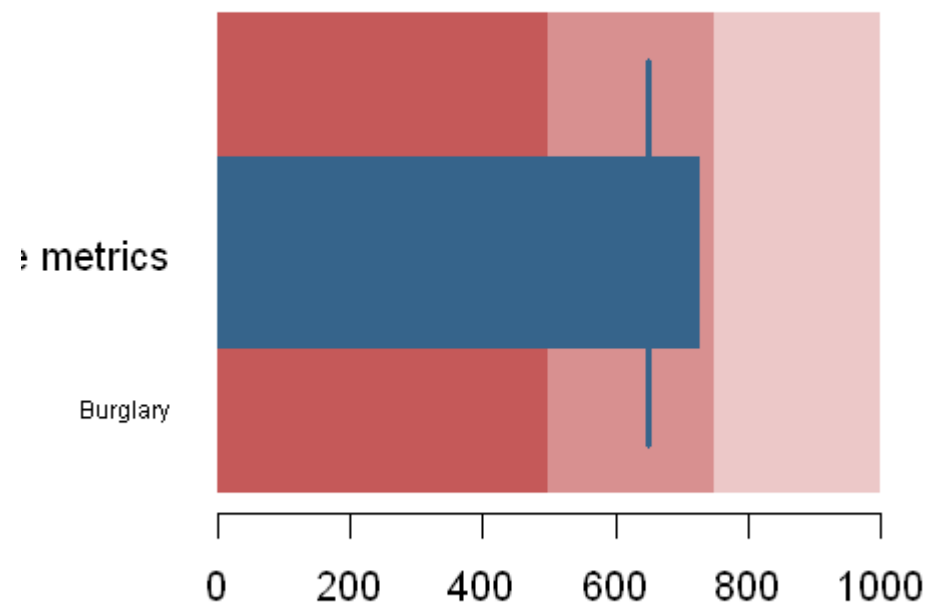


## Bullet chart

In [9]:

```
# create bullet chart with crime data
crime_bullet <- crime %>%
  dplyr::filter(stringr::str_trim(state, "both") == "United States") %>%
  dplyr::select(c(state, burglary))

bulletgraph(x=crime_bullet$burglary,ref=650,limits=c(0,500,750,1000),
  name= "USA Crime metrics",subname="Burglary",
  col="steelblue4",shades="firebrick")
```

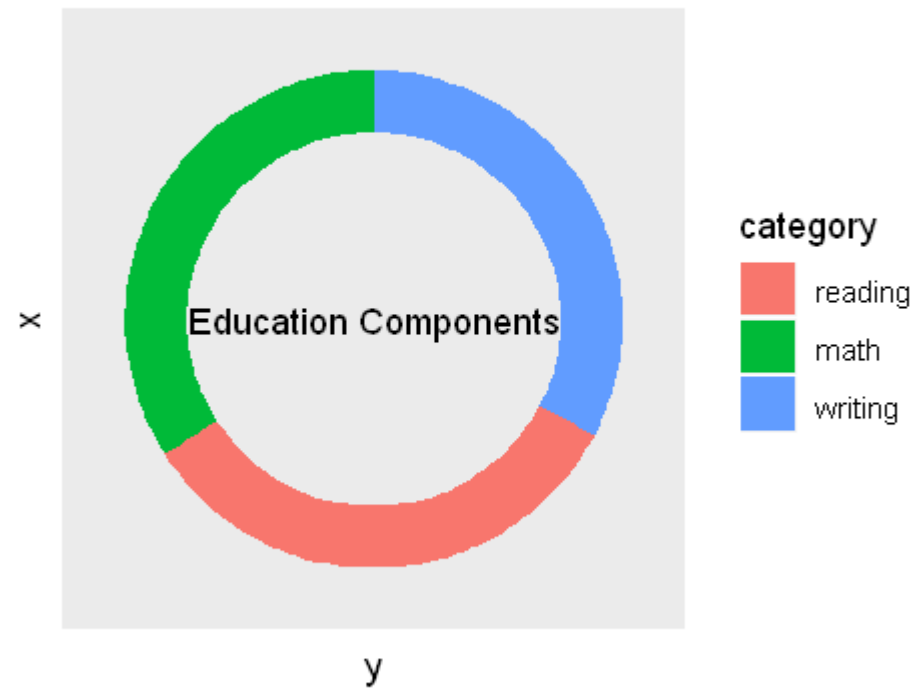


## Donut chart

```
In [10]: # donut chart using USA crime data
education_donut <- education %>%
  dplyr::filter(stringr::str_trim(state, "both") == "United States") %>%
  reshape2::melt(id=c("state")) %>%
  dplyr::rename("category" = variable) %>%
  dplyr::filter(category %in% c("reading", "math", "writing")) %>%
  dplyr::select(-state)

# add addition columns, needed for drawing with geom_rect
education_donut$fraction = education_donut$value / sum(education_donut$value)
education_donut = education_donut[order(education_donut$fraction), ]
education_donut$ymax = cumsum(education_donut$fraction)
education_donut$ymin = c(0, head(education_donut$ymax, n=-1))

# make the plot
ggplot2::ggplot(education_donut, ggplot2::aes(fill=category, ymax=ymax, ymin=ymin, xmax=4, xmin=3)) +
  ggplot2::geom_rect() +
  ggplot2::coord_polar(theta="y") +
  ggplot2::xlim(c(0, 4)) +
  ggplot2::theme(panel.grid=ggplot2::element_blank()) +
  ggplot2::theme(axis.text=ggplot2::element_blank()) +
  ggplot2::theme(axis.ticks=ggplot2::element_blank()) +
  ggplot2::annotate("text", x = 0, y = 0, label = "Education Components") +
  ggplot2::labs(title="")
```

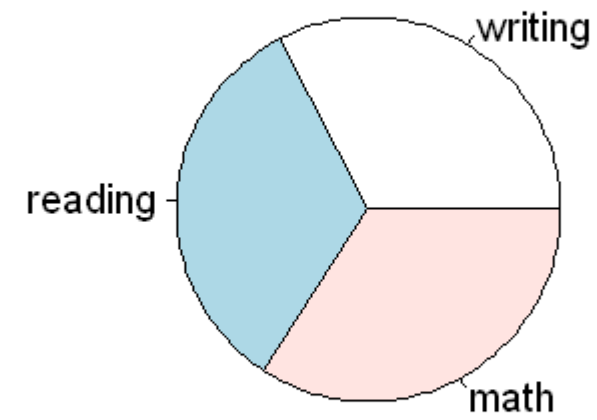


## Pie chart

In [11]:

```
# pie chart
slices <- education_donut$value
lbls <- education_donut$category
pie(slices, labels = lbls, main="Education Components")
```

## Education Components



## Parallel plot

In [12]:

```
# parallel plot
education_parallel <- education %>%
  dplyr::filter(stringr::str_trim(state, "both") != "United States")

# vector color
my_colors=colors()[as.numeric(education_parallel$state)*11]

# make the graph
MASS::parcoord(education_parallel[,c(2:4)] , col= my_colors )
```



