

Graphical Exploration

K.Mwai

Working Directroy

? is it set

Graphical exploration of data.

- To have an understanding of your data we normally conduct exploratory data analysis (EDA) which can be graphical or numerical.
- Primarily EDA is for seeing what the data can tell us before the formal modeling or hypothesis testing task.
- Typical graphical techniques used in EDA are:
 - Scatter plots,
 - Box plots,
 - Bar plots.

Data

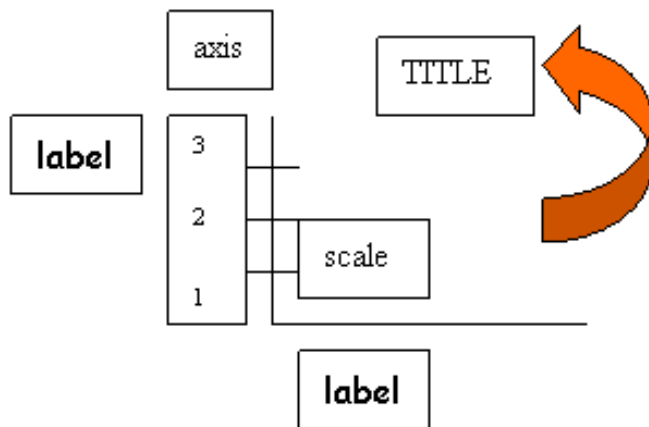
id	matage	ht	gestwks	sex	bweight	lbw	agegrp
1	33	2	38	Female	2410	Weight<2500	30-34 yrs
2	34	2	39	Female	2977	Normal 2500+	30-34 yrs
3	34	2	36	Female	2100	Weight<2500	30-34 yrs
4	30	2	39	Male	3270	Normal 2500+	30-34 yrs
5	35	2	38	Female	2620	Normal 2500+	35-39 yrs
6	37	2	38	Male	3260	Normal 2500+	35-39 yrs

Scatter plot

Scatter plot

- Its a useful summary of a set of bivariate data (two variables)
- It pairs up values of two quantitative variables in a data set with the aim of giving a good visual picture of the relationship between the two variables.
- The resulting pattern indicates the type and strength of the relationship between the two variables.
- Usually drawn before working out a linear correlation coefficient or fitting a regression line.

Parts of a graph



Graphical parameters

- Help in customizing features of your graphs

Text and symbol size.

cex = ;change size relative to **the** (default=1), 1.5 is 50% larger, 0.5 is 50% smaller.

cex.axis, cex.lab, cex.main

Plotting symbols

pch = ;0:25, +, *, -, % (default=1, open circle)

Lines

lty = ;line **type** (default is solid line), can be dashed, dotted, ...

lwd = ;line width. 2 is twice as wide.

Colors

col = ;0:8, col= 1, col="white".

col.axis, col.main, col.lab, bg, fg

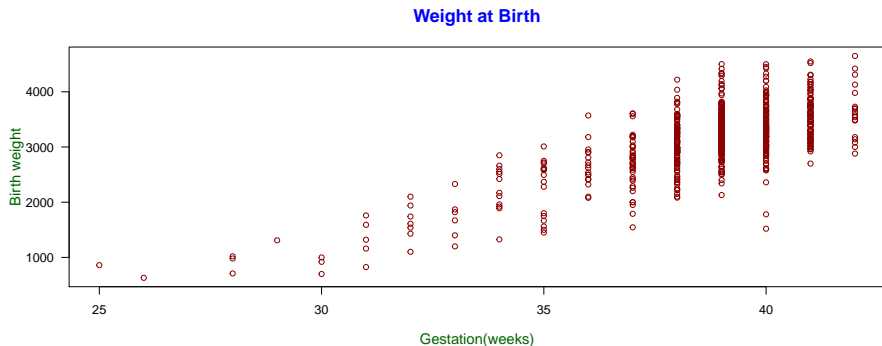
Axes

xlab="X-axis label", main="title", ylab="y-axis label",

xlim=c(xmin, xmax)

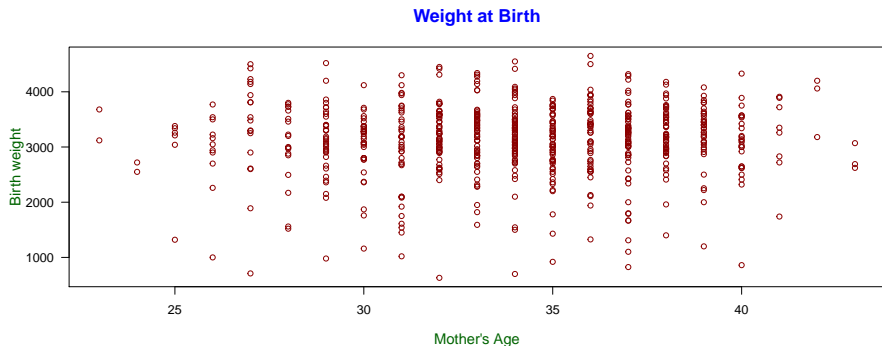
Scatter plot ... R code

```
plot(birth$gestwks,birth$bweight,las=1,xlab="Gestation(weeks)",  
     ylab="Birth weight",col.lab="darkgreen", cex.lab=1.2,  
     main="Weight at Birth",col.main="blue",cex.main=1.4,  
     xlim=c(25,42),col="darkred",type="p", cex=0.9,pch=1)
```



Scatter plot ... R code

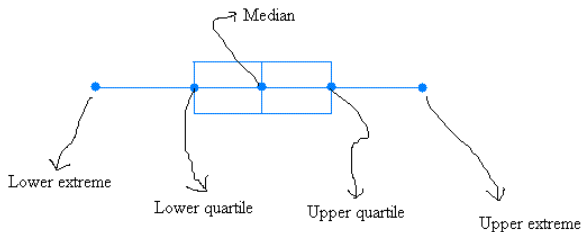
```
plot(birth$matage,birth$bweight,las=1,xlab="Mother's Age",  
     ylab="Birth weight",col.lab="darkgreen", cex.lab=1.2,  
     main="Weight at Birth",col.main="blue",cex.main=1.4,  
     col="darkred",type="p", cex=0.9,pch=1)
```



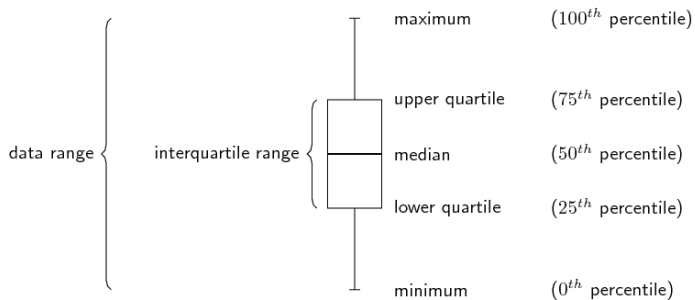
Box plot

Box plot

- Provides a standardized way of displaying the distribution of data.
- It attempts to provide a visual shape of the data distribution.
- This is based on some summary measures: min, 1st quartile, median, 3rd quartile, and max.
- Range, IQR, Outliers= $3 \times IQR$ above 3rd or below 1st quartiles.

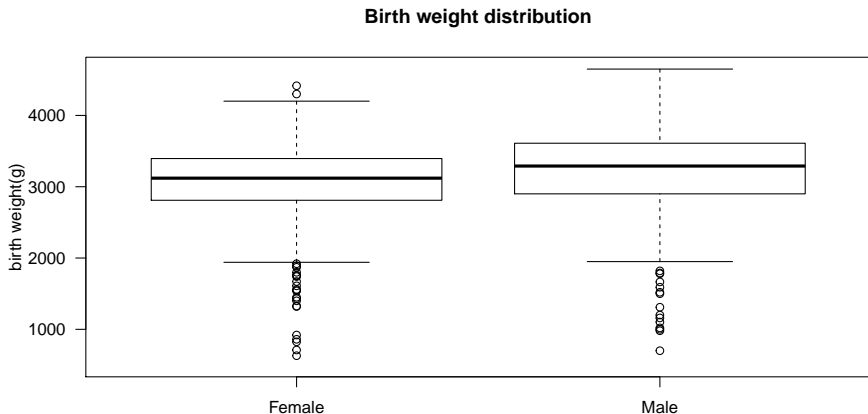


Box plot ...



Box plot ...

```
boxplot(bweight~sex,las=1,main="Birth weight distribution",  
        ylab="birth weight(g)",ylim=c(500,max(bweight)))
```



Bar plot

Bar plot

- Provide a visual presentation of categorical data.
- Present grouped data with rectangular bars with lengths proportional to the values that they represent.
- Two types;
 - Grouped - presents bars clustered in groups
 - Stacked - shows bars divided into subparts to show cumulative effects.

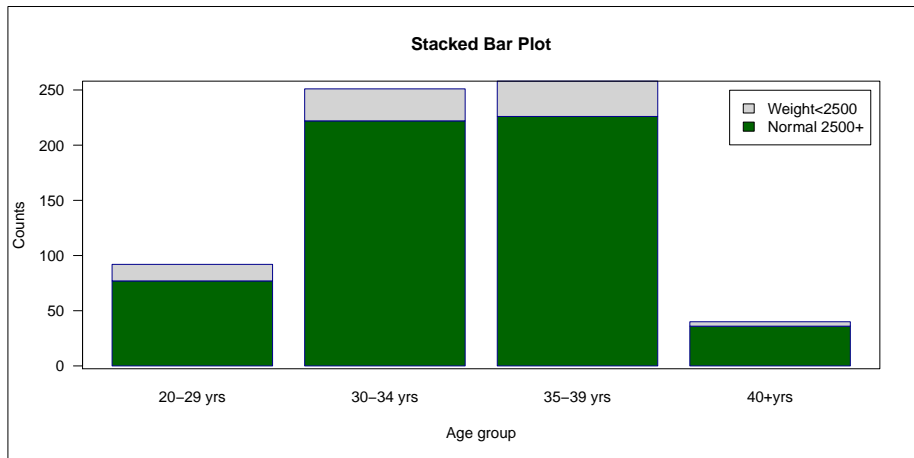
Bar plot ...

e.g. Cross tabulation of mother's age and birth weight.

	20-29 yrs	30-34 yrs	35-39 yrs	40+yrs
Weight <2500	15(0.16)	29(0.12)	32(0.12)	4(0.10)
Normal 2500+	77(0.84)	222(0.88)	226(0.88)	36 (0.90)
Total	92	251	258	40

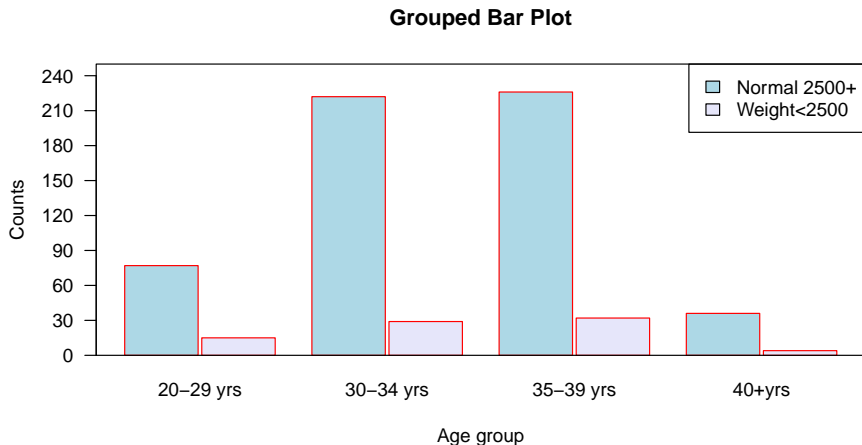
Bar plot ... Stacked Bar Plot

```
barplot(Count, las=1, main="Stacked Bar Plot", border="darkblue", xlab="Age group", ylab="Counts",  
        col=c("darkgreen", "lightgrey"), legend=rownames(Count))  
box(); box("figure")
```



Bar plot ... Grouped Bar Plot

```
barplot(Count, beside = T, main = "Grouped Bar Plot", xlab = "Age group", ylab = "Counts",
        border = "red", yaxt = 'n', col = c("lightblue", "lavender"), ylim = c(0, 250),
        legend = rownames(Count), args.legend = list(x = "topright"), space = c(0.05, 0.5))
axis(2, at = seq(0, 250, by = 30), las = 1); box()
```



Saving plot

Some useful formats.

```
# Start device driver
```

```
pdf("mygraph.pdf")  Produces a PDF file  
jpeg()  jpeg file  
png()   png file  
bmp()   bmp file  
win.metafile() windows metafile  
postscript() creating PostScript graphics files.
```

```
# terminate device  
dev.off()
```

Saving plot ...

```
pdf("barPlot.pdf",width =, height =)

barplot(Count,las=1, main="Stacked Bar Plot",border="darkblue",
        xlab="Age group",ylab="Counts",col=c("darkgreen","lightgrey"),
        legend=rownames(Count))

box()
box("figure")

dev.off()
```

Graphical Parameters

Introduction to Graphics in r R Graphs Gallery

Asante