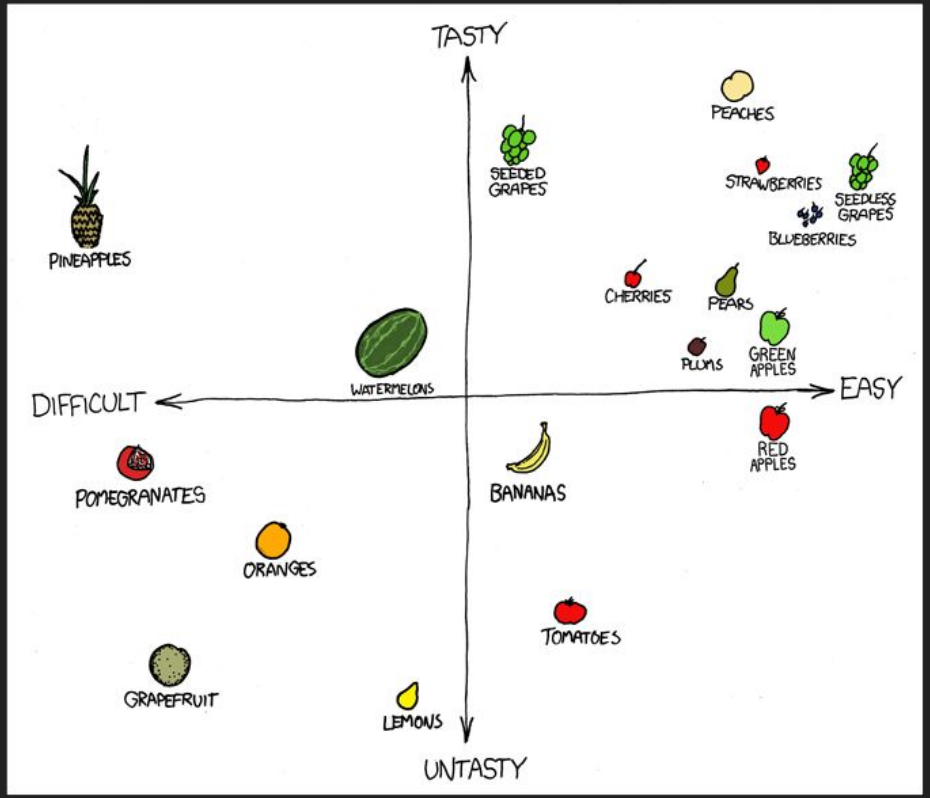
LAB 5: Data Graphics



Elements of data graphics

Visual cues

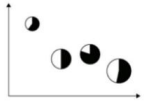
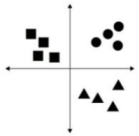
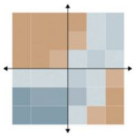
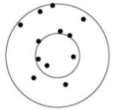
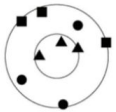
- position, length, area, etc.

Coordinate system

- how are the data points organized?

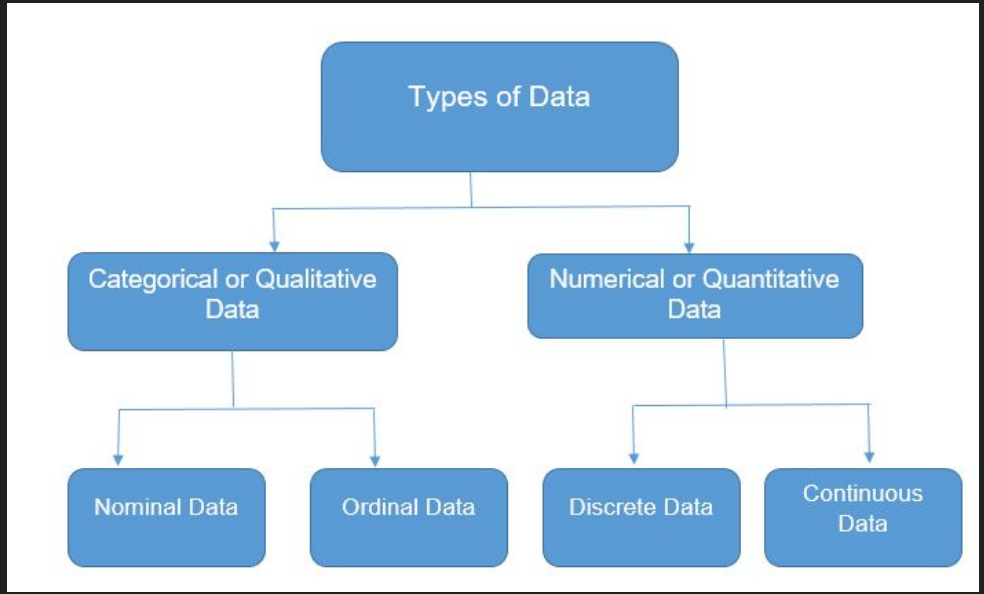
Context

- **in relation to what?**

Coordinate systems	Position	Length	Angle	Direction	Shapes	Area or Volume	Color
Cartesian							
Polar							
Geographic							

What kind of data are we plotting?

The type of variables/data we are trying to visualize will influence the type of graphic we will use.



Example: Type of variables

class(dat$x)
sapply(dat,class)

```r
data("Loblolly")
names(Loblolly)
sapply(Loblolly,class)
```

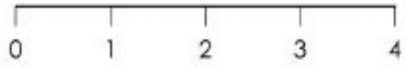
```
[1] "height" "age"    "Seed"
$height
[1] "numeric"

$age
[1] "numeric"

$Seed
[1] "ordered" "factor"
```

Scales

Along with coordinate systems, they dictate where the shapes are placed and how objects are shaded.

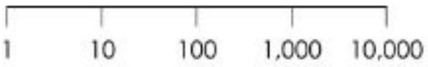
Linear

Values are evenly spaced



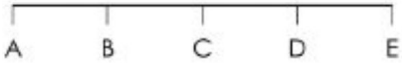
Logarithmic

Focus on percent change



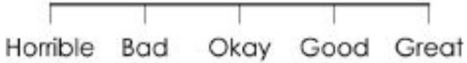
Categorical

Discrete placement in bins



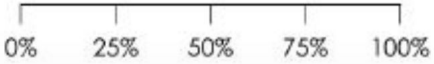
Ordinal

Categories where order matters



Percent

Representing parts of a whole



Time

Units of months, days, or hours

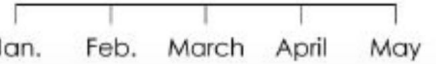
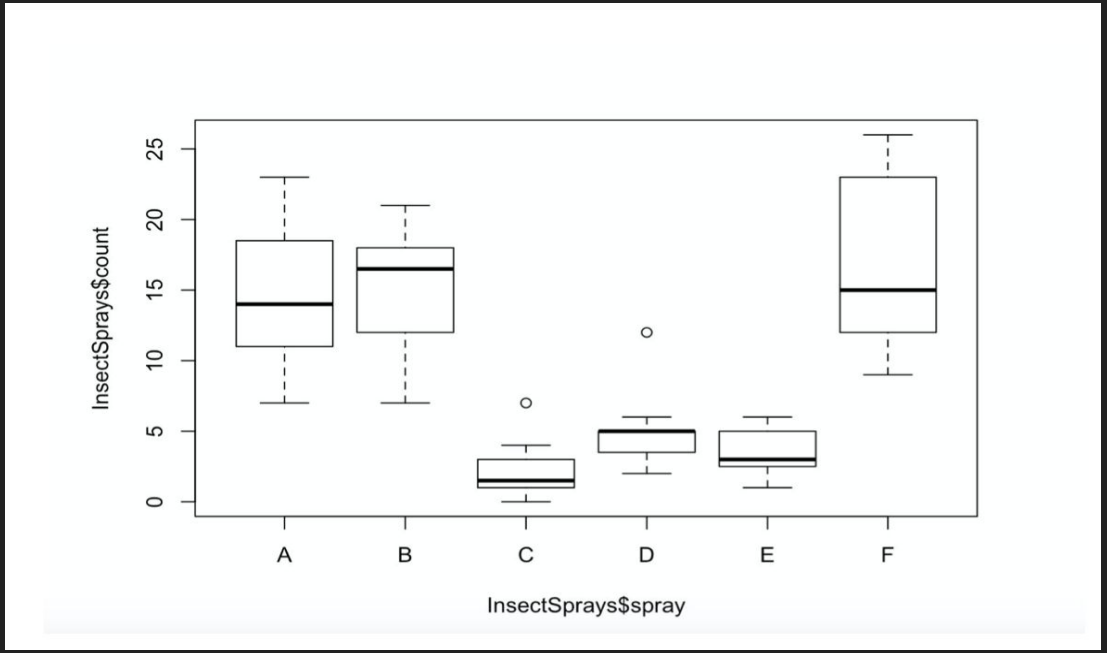


FIGURE 3-15

Boxplot

Boxplot to show the *distribution* of **a numerical variable** split by **a categorical variable.**

R-code:
boxplot(InsectSprays$count ~ InsectSprays$spray)



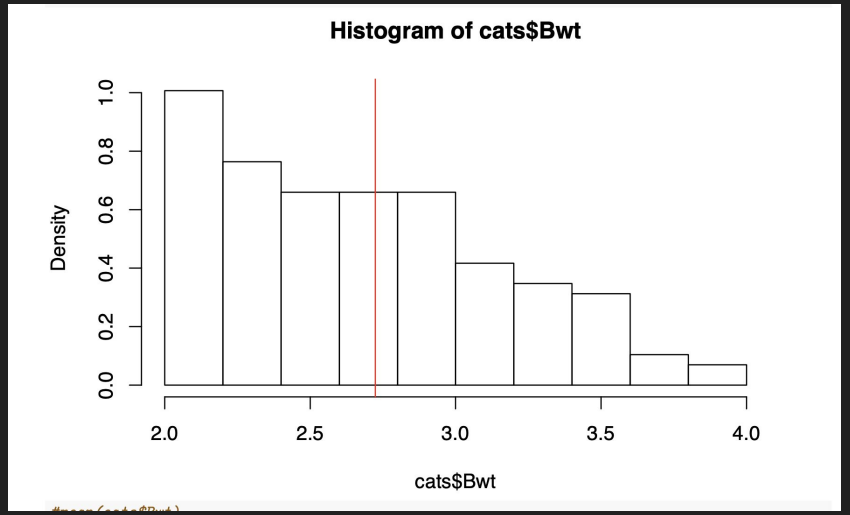
Histogram

Histograms to show *distribution* of a **numerical variable.**

Example R-code:

hist(cats$Bwt,probability=TRUE)

abline(v=**mean**(cats$Bwt),col = "red")



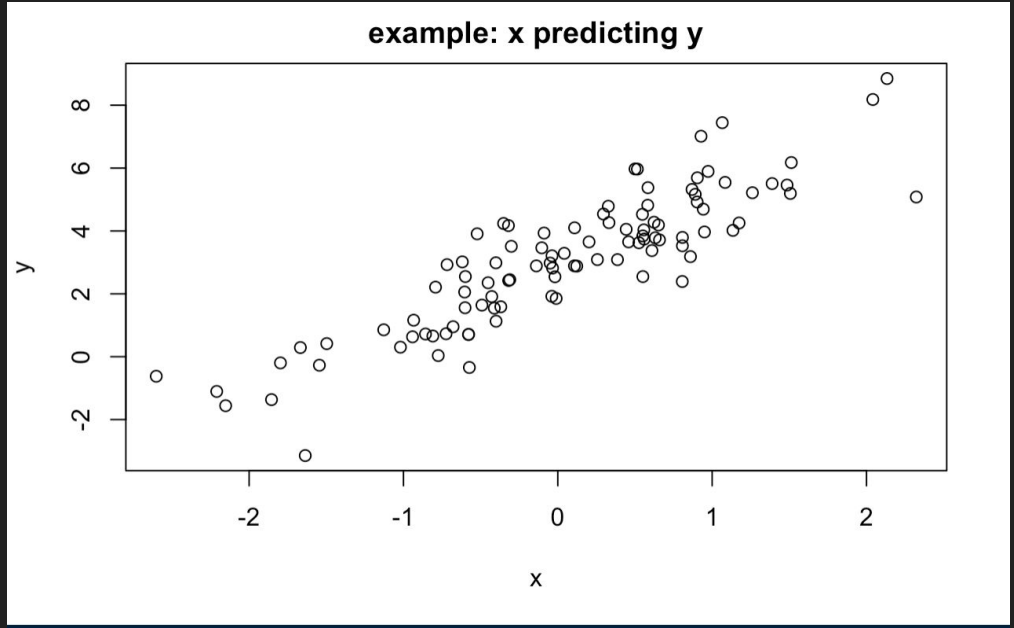
Scatter-plot

Scatter-plots show *the relationship between* ***2*** **numerical variables**

X-axis = explanatory variable

Y-axis = Response Variable

```
plot(x=x, y=y,main= "example: x predicting y")
```



Linear regression

$$y = mx + b + \text{error}$$

Line of best fit (least squares)

$$\hat{y} = mx + b$$

Example

$\hat{y}$= 1.89436x + 3.04541

plot(x,y, main= "example: x predicting y")
abline(3.04541,1.89436, col=”blue”)