Excercise 1

mean: the sum of the items divided by the number of items.
median: the middle item in a sorted list of items
mode: the item with the most frequency
variance: the average sum of the difference from the mean squared.
standard deviation: the square root of the variance.
histogram: a graph where bars represent the frequency of an item.
Normal Distribution: a distribution that forms a bell shape where the center has the most frequency and the tails on both sides are less frequent.
Poisson Distribution: a distribution where one item is more frequent than the rest of the possible outcomes. It defers from a Normal Distribution as it does not always look symmetric.

Excercise 3

In [11]:
```
r_datasets = data()
```

In [12]:
```
r_datasets
```

Data sets

A data.frame: 104 × 3

Package	Item	Title
<chr>	<chr>	<chr>
datasets	AirPassengers	Monthly Airline Passenger Numbers 1949-1960
datasets	BJsales	Sales Data with Leading Indicator
datasets	BJsales.lead (BJsales)	Sales Data with Leading Indicator
datasets	BOD	Biochemical Oxygen Demand
datasets	CO2	Carbon Dioxide Uptake in Grass Plants
datasets	ChickWeight	Weight versus age of chicks on different diets
datasets	DNase	Elisa assay of DNase
datasets	EuStockMarkets	Daily Closing Prices of Major European Stock Indices, 1991-1998
datasets	Formaldehyde	Determination of Formaldehyde
datasets	HairEyeColor	Hair and Eye Color of Statistics Students
datasets	Harman23.cor	Harman Example 2.3
datasets	Harman74.cor	Harman Example 7.4
datasets	Indometh	Pharmacokinetics of Indomethacin
datasets	InsectSprays	Effectiveness of Insect Sprays

Package	Item	Title
<chr>	<chr>	<chr>
datasets	JohnsonJohnson	Quarterly Earnings per Johnson & Johnson Share
datasets	LakeHuron	Level of Lake Huron 1875-1972
datasets	LifeCycleSavings	Intercountry Life-Cycle Savings Data
datasets	Loblolly	Growth of Loblolly pine trees
datasets	Nile	Flow of the River Nile
datasets	Orange	Growth of Orange Trees
datasets	OrchardSprays	Potency of Orchard Sprays
datasets	PlantGrowth	Results from an Experiment on Plant Growth
datasets	Puromycin	Reaction Velocity of an Enzymatic Reaction
datasets	Seatbelts	Road Casualties in Great Britain 1969-84
datasets	Theoph	Pharmacokinetics of Theophylline
datasets	Titanic	Survival of passengers on the Titanic
datasets	ToothGrowth	The Effect of Vitamin C on Tooth Growth in Guinea Pigs
datasets	UCBAdmissions	Student Admissions at UC Berkeley
datasets	UKDriverDeaths	Road Casualties in Great Britain 1969-84
datasets	UKgas	UK Quarterly Gas Consumption
datasets	USAccDeaths	Accidental Deaths in the US 1973-1978
datasets	USArrests	Violent Crime Rates by US State
datasets	USJudgeRatings	Lawyers' Ratings of State Judges in the US Superior Court
datasets	USPersonalExpenditure	Personal Expenditure Data
datasets	UScitiesD	Distances Between European Cities and Between US Cities
datasets	VADeaths	Death Rates in Virginia (1940)
datasets	WWWusage	Internet Usage per Minute
datasets	WorldPhones	The World's Telephones
datasets	ability.cov	Ability and Intelligence Tests
datasets	airmiles	Passenger Miles on Commercial US Airlines, 1937-1960
datasets	airquality	New York Air Quality Measurements
datasets	anscombe	Anscombe's Quartet of 'Identical' Simple Linear Regressions
datasets	attenu	The Joyner-Boore Attenuation Data
datasets	attitude	The Chatterjee-Price Attitude Data
datasets	austres	Quarterly Time Series of the Number of Australian Residents
datasets	beaver1 (beavers)	Body Temperature Series of Two Beavers
datasets	beaver2 (beavers)	Body Temperature Series of Two Beavers
datasets	cars	Speed and Stopping Distances of Cars

Package	Item	Title
<chr>	<chr>	<chr>
datasets	chickwts	Chicken Weights by Feed Type
datasets	co2	Mauna Loa Atmospheric CO2 Concentration
datasets	crimtab	Student's 3000 Criminals Data
datasets	discoveries	Yearly Numbers of Important Discoveries
datasets	esoph	Smoking, Alcohol and (O)esophageal Cancer
datasets	euro	Conversion Rates of Euro Currencies
datasets	euro.cross (euro)	Conversion Rates of Euro Currencies
datasets	eurodist	Distances Between European Cities and Between US Cities
datasets	faithful	Old Faithful Geyser Data
datasets	fdeaths (UKLungDeaths)	Monthly Deaths from Lung Diseases in the UK
datasets	freeny	Freeny's Revenue Data
datasets	freeny.x (freeny)	Freeny's Revenue Data
datasets	freeny.y (freeny)	Freeny's Revenue Data
datasets	infert	Infertility after Spontaneous and Induced Abortion
datasets	iris	Edgar Anderson's Iris Data
datasets	iris3	Edgar Anderson's Iris Data
datasets	islands	Areas of the World's Major Landmasses
datasets	ldeaths (UKLungDeaths)	Monthly Deaths from Lung Diseases in the UK
datasets	lh	Luteinizing Hormone in Blood Samples
datasets	longley	Longley's Economic Regression Data
datasets	lynx	Annual Canadian Lynx trappings 1821-1934
datasets	mdeaths (UKLungDeaths)	Monthly Deaths from Lung Diseases in the UK
datasets	morley	Michelson Speed of Light Data
datasets	mtcars	Motor Trend Car Road Tests
datasets	nhtemp	Average Yearly Temperatures in New Haven
datasets	nottem	Average Monthly Temperatures at Nottingham, 1920-1939
datasets	npk	Classical N, P, K Factorial Experiment
datasets	occupationalStatus	Occupational Status of Fathers and their Sons
datasets	precip	Annual Precipitation in US Cities
datasets	presidents	Quarterly Approval Ratings of US Presidents
datasets	pressure	Vapor Pressure of Mercury as a Function of Temperature
datasets	quakes	Locations of Earthquakes off Fiji
datasets	randu	Random Numbers from Congruential Generator RANDU
datasets	rivers	Lengths of Major North American Rivers

Package	Item	Title
<chr>	<chr>	<chr>
datasets	rock	Measurements on Petroleum Rock Samples
datasets	sleep	Student's Sleep Data
datasets	stack.loss (stackloss)	Brownlee's Stack Loss Plant Data
datasets	stack.x (stackloss)	Brownlee's Stack Loss Plant Data
datasets	stackloss	Brownlee's Stack Loss Plant Data
datasets	state.abb (state)	US State Facts and Figures
datasets	state.area (state)	US State Facts and Figures
datasets	state.center (state)	US State Facts and Figures
datasets	state.division (state)	US State Facts and Figures
datasets	state.name (state)	US State Facts and Figures
datasets	state.region (state)	US State Facts and Figures
datasets	state.x77 (state)	US State Facts and Figures
datasets	sunspot.month	Monthly Sunspot Data, from 1749 to "Present"
datasets	sunspot.year	Yearly Sunspot Data, 1700-1988
datasets	sunspots	Monthly Sunspot Numbers, 1749-1983
datasets	swiss	Swiss Fertility and Socioeconomic Indicators (1888) Data
datasets	treering	Yearly Treering Data, -6000-1979
datasets	trees	Diameter, Height and Volume for Black Cherry Trees
datasets	uspop	Populations Recorded by the US Census
datasets	volcano	Topographic Information on Auckland's Maunga Whau Volcano
datasets	warpbreaks	The Number of Breaks in Yarn during Weaving
datasets	women	Average Heights and Weights for American Women

Use 'data(package = .packages(all.available = TRUE))' to list the data sets in all *available* packages.

In [13]:
```
head(PlantGrowth)
```

A data.frame: 6 × 2

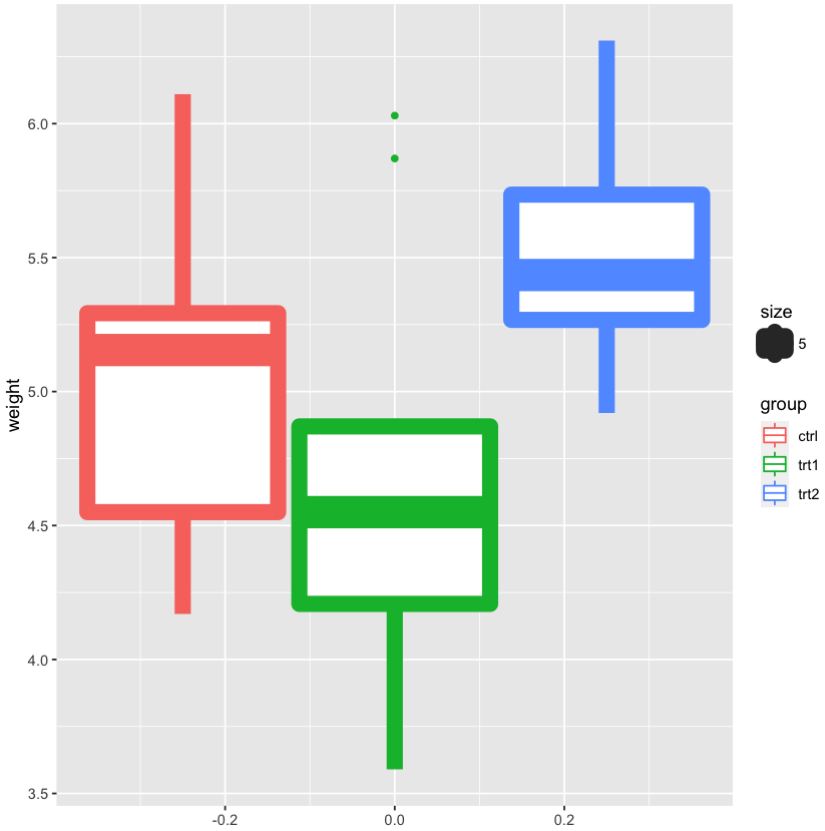
	weight	group
	<dbl>	<fct>
1	4.17	ctrl
2	5.58	ctrl
3	5.18	ctrl
4	6.11	ctrl
5	4.50	ctrl

	weight	group
	<dbl>	<fct>
6	4.61	ctrl

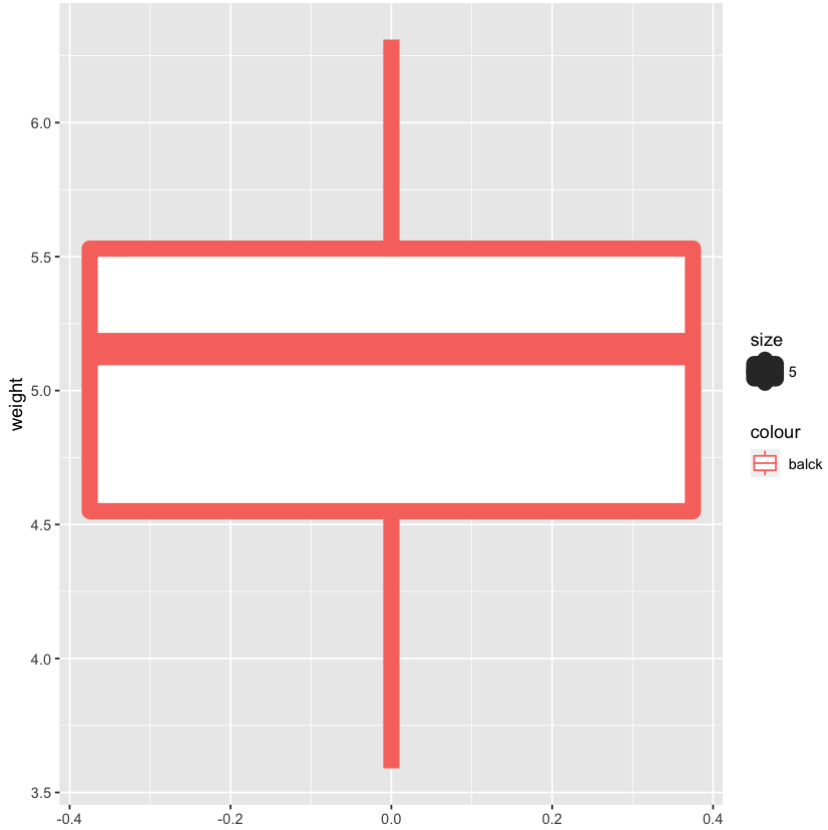
In [14]:
```
summary(PlantGrowth)
```

```
    weight       group   
Min.   :3.590   ctrl:10  
1st Qu.:4.550   trt1:10  
Median :5.155   trt2:10  
Mean   :5.073            
3rd Qu.:5.530            
Max.   :6.310            
```

In [23]:
```
library(ggplot2)
ggplot(PlantGrowth, aes(y=weight)) + geom_boxplot(aes(color = group, size=5))
```



In [22]:
```
ggplot(PlantGrowth, aes(y=weight)) + geom_boxplot(aes(color = 'balck', size=5))
```



The dataset I choose is PlantGrowth. It contains the weight of plants from three different groups. The Average weight of all plants is 5.073 while the median is 5.155. This means that 50% of the plants weight more than 5.155, and the weight for a plant is on average around 5.073.

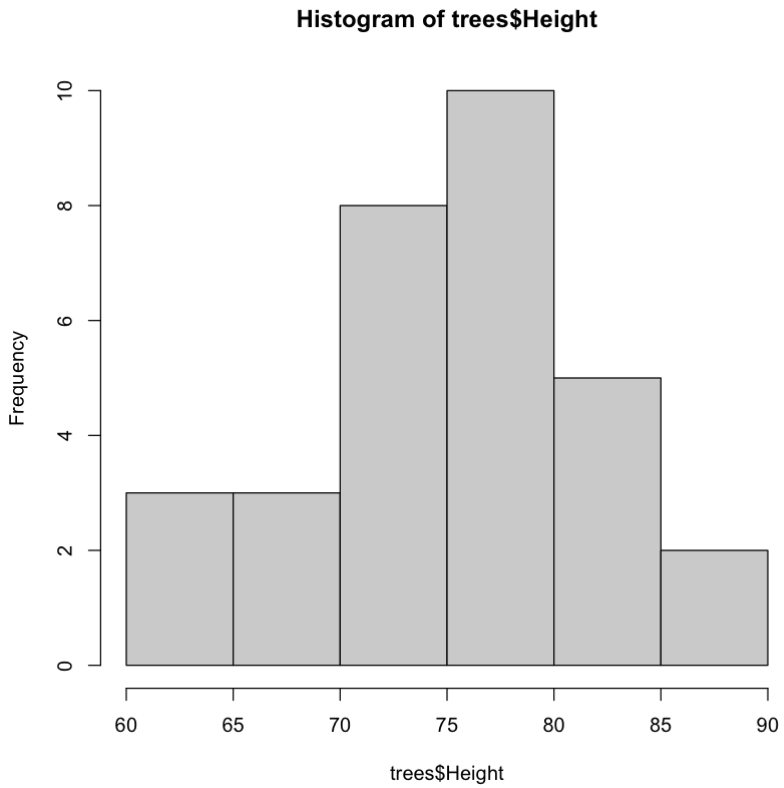
Excercise 4

In [27]:
```
head(trees)
```

A data.frame: 6 × 3

	Girth	Height	Volume
	<dbl>	<dbl>	<dbl>
1	8.3	70	10.3
2	8.6	65	10.3
3	8.8	63	10.2
4	10.5	72	16.4
5	10.7	81	18.8
6	10.8	83	19.7

In [29]:
```
hist(trees$Height)
```



For this exercise I selected the trees dataset. I plotted the histogram of the Height column. Here we can see how the distribution looks normal as most trees fall in the 70-80 height while fewer fall on the tail ends creating a bell curve.